Evidence Based Policy Making in Healthcare using Big Data Analytics

CITY
UNIVERSITY OF LONDON
EST 1894

Marios Prasinos

Department of Computer Science
School of Mathematics,
Computer Science and Engineering
Supervisor: Professor George Spanoudakis

Thesis for the degree of Doctor of Philosophy

May 2019

Table of Contents

List of Figures

List of Tables

Acknowledgements

The development of this thesis has been a long and didactic journey. I would like to thank all the people that supported me, encouraged me and assisted me throughout this chapter of my life.

Firstly, I would like to express my sincere gratitude to my supervisor, Professor George Spanoudakis, for his continuous guidance, support and assistance. I would also like to express my sincere appreciation to my external supervisor, Professor Dimitris Koutsouris, for his continuous encouragement and support.

Furthermore, I would like to extend my sincere appreciation to all the members of the Centre of Adaptive Computing Systems (CeNACS) of City, University of London and the Biomedical Engineering Laboratory of the National Technical University of Athens. I would also like to thank all the EVOTION partners for the great support and collaboration.

In addition to the above people, I would also like to thank all the friends I made during this journey and all my close friends for their support and for the great time we spent together.

Last but not least, I am very grateful to my parents and to my brother for their constant love and support throughout all my life. I would like to dedicate this thesis to my father.

Declaration

The author grants powers of discretion to the University Librarian to allow this thesis to be copied in whole or in part without further reference to him. This permission covers only single copies made for study purposes, subject to normal conditions of acknowledgement.

Abstract

The effective management of various health conditions depends on and requires appropriate public health policies (PHP). Public health policy can affect several aspects of healthcare provision including: (a) prevention and early diagnosis of diseases; (b) early treatment of diagnosed conditions through the provision of appropriate health care devices; (c) longer term treatment of long term disabilities and chronic diseases through systematic checks of the patient's condition and the provision of other vital rehabilitation related services; (d) protection of people with health care devices from the harmful effects of their living environment; (e) setup of standards, services and technology for promoting and ensuring patients' participation and inclusion within various settings (e.g., at work, at school/educational establishments, in everyday life). Although there is a need for evidence based public health policy making, there is currently no computerised tool to enable the process.

The overall aim of our research is to develop an integrated platform by incorporating a big data analytics (BDA) platform that facilitates the collection and analysis of heterogeneous data related to healthcare services, including health care device usage, physiological, cognitive, medical, personal, occupational, behavioural, lifestyle, environmental and open web data. For the purposes of the development of this integrated platform we are introducing a Public Health Policy Decision Making modeling language that allows the specification of models that are executable by the platform.

For the evaluation of the developed platform, we developed a scenario, instantiated the ontology model using Protégé and generated synthetic data. We also ran the scenario using real patient data from EVOTION project. We performed subjective evaluation of the platform as a policy making tool using three questionnaires (one for policy makers, one for clinicians and one for data analysts) and analysed the results.

The novelty of this thesis lies not only in the specification of the PHPDM modeling language, as there is no other ontology on public health policy decision making, but also in the development of the BDA engine and the prototype, as there is no other similar policy making platform to date.

Some open issues regarding the developed platform include (a) further formalization and addition of new constructs to the developed PHPDM specification language to support the full lifecycle of policy formation processes, (b) the provision of templates,

guidelines and supportive material (e.g. tooltips in the interface and tutorial videos) to help policy makers specify data analytics workflows and criteria, (c) interoperability with other data analytics tools and existing health data repositories, (d) the provision of the developed platform as a service, (e) the implementation of more data mining and statistical analysis algorithms and (f) the development of a decision support system that will enable the platform to not only support the execution of big data analytics, but to also directly support the policy making process.

List of Abbreviations

OWL	Web Ontology Language
PHPDM	Public Health Policy Decision Making
RDF	Resource Description Framework
XML	eXtensible Markup Language
XSD	XML Schema Documentation
SQL	Structured Query Language
UML	Unified Modeling Language
BDA	Big Data Analytics
CSV	Comma Separated Values
GUI	Graphical User Interface
HL	Hearing Loss
HAs	Hearing Aids

Chapter 1

Introduction

1.1. Overview

In this section we present the motivation and research challenges, the aims and objectives of this thesis, the research assumptions and contributions, our publications during the period of the thesis and the outline of this document.

For the motivation and research challenges, first, we present the public health policies in healthcare definition and challenges and then the need and role of public policy in healthcare. After that, we present the overall aim of this research and the main objective followed by the sub-objectives. Afterwards, we present the research assumptions and contributions, our publications during the period of the thesis and, finally, the outline of this document.

1.2. Motivation and Research Challenges

1.2.1. Public Health Policies in Healthcare definition and challenges

A governments' health policy is their blueprint for improving citizens' health at a national level [1]. More precisely, health policy is the term used to describe the specific decisions, plans and actions that are made in order to achieve identifiable health targets within a given society [2]. The advantages of an explicit health policy are threefold: first, it explicitly identifies short and medium term goals which can be used as points of reference for evaluating progress; next, it highlights priorities and defines the functional roles of different groups; and finally, it educates people about health care goals within society and consequently helps establish a societal consensus regarding health care [2].

According to Metcalf et al [3], six challenges in modeling for public health policy include (a)communicating the limits of modeling, (b)maintaining the value of models in the face of long time horizons, (c)usefully deploying modeling in the context of 'black swans', (d)integrating modellers and model-building into the policy process, (e)economic analysis and decision support and (f)creating a cycle where results inform decisions and vice versa.

According to a recent review paper [4], rapid changes to the context of public health (i.e. globalisation, migration, demographic changes) have given rise to new challenges that public health policy must address in order to ensure the continued effectiveness of health care systems. These challenges are grouped into ten themes:

1. Human health is dependent upon the planet's health. Consequently, public health policy makers have a responsibility not only to assess the effects of environmental factors upon human health but also to ensure that health policies promote environment sustainability and mitigate the effects of human-caused climate change.
2. Increases in life expectancy are associated with increases in the prevalence of chronic, non-communicable diseases. Public health policy must address these changes in the population. For example, policy makers have a twofold responsibility to promote wellbeing and healthy living/aging and to actively combat the commercial forces driving unhealthy behaviours (i.e. tobacco use, unhealthy diets, physical inactivity, alcohol misuse, and gambling).
3. Given the ease of international travel and increases in international migration, effective public health policy must move beyond the traditional focus on local, regional and national health care systems and show a commitment to improving global health systems and networks. This necessitates strengthening extant international institutions as well as creating new networks of practice to link health care systems in developed and developing countries. This will improve nation-states' ability to respond to global health threats in an organised and efficient manner.
4. Public health officials must continue to advocate for the importance of investing in public health. More specifically, investing in public health entails: developing explicit accreditation, performance, and workforce planning frameworks; delineating how resource use aligns with quality of service; and, developing specific population-level health outcomes.
5. Public health policy has an obligation to address important social issues such as poverty, equity, powerlessness, discrimination, and stigma. Effectively tackling these issues requires a multifaceted approach where public health policy liaises with other sectors, such as the media and advocacy coalitions.
6. National health systems should be affordable and financially sustainable. One way to achieve this goal is by strengthening the ties between public health and clinical care systems.

7. New technologies should be used to inform and facilitate public health policy planning and decision making. For example, information obtained through new technologies can be used to provide context-specific and individualised interventions to target populations.
8. It is important to acknowledge that the public is a vital member of the community of public health policy and practice. Therefore, the public's involvement in public health decision making must be actively pursued (i.e. through the use of social media, online consultation, and open access to government data).
9. The public health policy community must commit to improving their ethical and regulatory framework. Specifically, they must redefine some ambiguous terminology (i.e. precautionary principle and intergenerational equity) and find a balance between the promotion of equity and public good considerations and the maintenance of individual human rights and individual advancement.
10. The contemporary issues challenging public health policy are complex and require holistic, multifaceted solutions. In order to design comprehensive, whole-system solutions, public health policy must be embedded in research and innovation and underpinned by interdisciplinary partnerships between health professionals, economists, social and behavioural scientists.

1.2.2. Need and role of public policy in healthcare

The effective management of various health conditions depends on and requires appropriate public health policies (PHP) as it has been acknowledged in documents by the World Health Organization [5], by governmental institutions and patient associations [6]. Public health policy affects the affordability and hence access to health care devices and ongoing treatment services (e.g., health check-ups, health care device adjustments, provision of related rehabilitation services). Public health policy also has a significant effect on: (a) the prevention and early diagnosis of diseases, (b) the early treatment of diagnosed conditions through the provision of the appropriate health care devices; (c) the longer term treatment of long term disabilities and chronic diseases through systematic checks of the condition of the patient, the provision of other vital related rehabilitation services; (d) the protection of people with health care devices(e.g. hearing impairments) from the harmful effects of their living environment (e.g. loud noise); (e) the early detection, delay or even prevention of the impairment of the patient's condition; (f) the setup of standards, services and technology for promoting and ensuring inclusion of participation of patients with in various settings (e.g., at work, at school/educational establishments, in everyday life) [7]. Ongoing reforms of PHP in this area (e.g., changes in the free provision of HAs for different types of HL in the UK) and the spark of social debate that they have caused demonstrate the importance of PHP in this area. Examples of policy fields that influence big part of the population and mobilize resources encountered in billions of euros yearly are cut off points of hearing aid fitting covered by insurance, decision for unilateral or bilateral HA fitting or cochlear implantation, noise protection measures in working environments, and default maximum dB levels of electronic devices. Managing a health condition can be treated in a personalised level or in a population (public) health level. The relationship between individual and population health is largely relative and dynamic [8]. The management of a health condition and its consequences at a public health policy making level can benefit from the analysis of heterogeneous data, including health care device usage (if applicable), physiological, cognitive, clinical and medication, personal, behavioural, lifestyle data, occupational and environmental data, and more patient specific data (e.g. audiological for hearing loss patients). The analysis of these types of data using big data analytic techniques can enable the investigation of whether the condition relates to other comorbidities and contextual factors and patterns of such relations. The outcomes of such analysis can enable the stratification of related risks [9] and effects to the patients, and – through correlation with other economic, social and physical constraints – help developing a holistic systemic perspective of over interventions regarding the management of the condition and the broader support, social and occupational inclusion and well-being of the patients, exploring missing, under or over-estimated value of specific interventions (e.g. for hearing loss patients: noise protection, visualization of public announcements etc.) and analysing their effectiveness (i.e., understanding the trade-offs between their cost and benefits).

1.3. Aims and Objectives

As there is no computerised tool for the provision of evidence to the policy making processes in healthcare, the overall aim of this research is to enable the provision of such evidence using big data analytics and support evidence-based public health policy formation through the development of an integrated platform incorporating a big data analytics (BDA) engine enabling the collection and analysis of heterogeneous data related to healthcare, including health care device usage, physiological, cognitive, medical, personal, occupational, behavioural, life style, environmental and open web data.

The main objective of this thesis is to develop a high-level language for the specification of evidence-based health policy decision making models based on big data analytics and executable by a developed platform. This language is the core element of the developed platform prototype.

Description: This objective involves the development of a language for specifying health policy decision making models that can be executed by the platform. The existence of an explicit specification of such models enables their reusability against different sets of related data when they become available or against the sets of data that the models have been applied before for verification purposes. Such models are based on the development of an ontology for health policy decision making and the connection of it with abstract declarative specifications of data analysis processes.

Success Indicators: Delivery of an ontology-based language for the specification of evidence-based health policy decision making models and a prototype supporting the specification of models in this language. The language enables the specification of models with verifiable properties (e.g., complete attribution of all possible decisions onto the outcomes of one or more BDA tasks), which can be executed (following some automated transformations) by the platform. This prototype could be a paradigm shift for other policy fields.

In order to achieve the above overall objective, we break it down to the following sub-objectives:

Objective 1: **Literature Review**

To explore and provide a detailed description of existing policy formation processes and guidelines, ontologies in the domains of policy making, data mining and statistics in order to formulate the ontology of our framework, data analytics technologies and big data platforms.

Objective 2: **Development of the public health policy modeling framework**

To develop a modeling framework for the specification of health policy decision making models based on existing ontologies explored (Objective 1).

Objective 3: **Integration with big data mining and statistical analysis frameworks**

To enable the integration of the developed health policy decision making modeling framework with big data mining and statistical analysis frameworks for the execution of the data analytics tasks (data mining or statistical) to provide the evidence for the policy making.

Objective 4: **Proof of concept prototyping**

To formulate a use case scenario for the proof of concept of the use of the developed framework (Objective 2) for the specification of evidence-based health policy decision making models based on big data analytics. This also includes the generation of synthetic data.

To describe the process that needs to be followed in order to make use of the developed framework (create the instances to the ontology according to the use case scenario) in order to facilitate the policy making. This includes the execution of the data analytics tasks that support all the alternatives of the scenario, along with the querying to the ontology for the alternatives with satisfied criteria to be identified.

This also incorporates the development of a tool supporting the specification of models with the use of the developed integrated framework (Objectives 2 and 3). This tool will automate the process described in the proof of concept.

Objective 5: **Experimental evaluation**

To evaluate the proof of concept (Objective 4) including the experimental results extracted during the process of the use of the integrated developed framework (Objectives 2 and 3).

1.4. Research Assumptions

To shape the research, the following assumptions were made giving some starting points and directions to the work:

- The process of the acquisition of the data (health care device usage, physiological, cognitive, medical, personal, occupational, behavioural, lifestyle, environmental and open web data) is out of scope of this research. For the acquisition of data for the purposes of the prototype we used the EVOTION Data Repository and the API to retrieve and store data to it. This repository uses HBASE [10], which is the Hadoop database, a distributed, scalable, big data store.
- For the purposes of the prototype of this research we do not deal with the implementation, monitoring and evaluation of the policy actions of the PHPDM process. We also do not deal with the decision-making process of the stakeholders, although the language covers it.
- The focus of this research is the modeling language for the specification of PHPDM models, as well as the execution of the scenario in Chapter 5 showing the interactions between the components of our platform.

1.5. Research Contributions

This research is aimed at providing a platform for the specification of PHPDM models, the execution of data analytics tasks for the provision of evidence to the stakeholders of our platform and the identification of possible interventions related to the PHPDM models specified. The main contributions of this research are:

- The provision of the ontology-based modeling language for the specification of PHPDM models. This modeling language is novel, as there is no similar and complete approach to PHPDM modeling. The language covers a wide range of policy making processes, data analytics workflows for the provision of evidence to the stakeholders, as well as the decision-making processes. This language is very useful to the policy makers as well as the data analysts, as it simplifies the PHPDM processes and their collaboration.

- The construction of a BDA engine to execute the data mining tasks and provide the evidence needed. The BDA engine we built for the purposes of the prototype that showcases the use of our language is novel, as it can process big datasets and perform data mining tasks fast and accurately, for the provision of the evidence. This is also very useful to the data analysts, as it enables them to explore the available data and perform the required analysis.

1.6. Publications

In this section we list the publications authored during the period of the development of this thesis along with their abstracts. These include conference proceedings, journal publications and book chapters.

Conferences

- Prasinos, M., Spanoudakis, G. & Koutsouris, D. "Towards a Model-Driven Platform for Evidence-based Public Health Policy Making" 29th International Conference on Software Engineering & Knowledge Engineering, 5-7 Jul 2017, Pittsburgh, USA, 2017.
  This conference paper presents the preliminary version of the PHPDM modeling language of this thesis. The language has been enriched with more classes and attributes and has been finalized.
  Abstract: The effective management of various health conditions depends on and requires appropriate public health policies (PHP). Such policies are important for several aspects of healthcare pro-vision, including: (a) screening for prevention of disease; (b) early diagnosis and treatment; (c) long-term management of chronic diseases and disabilities; and (d) setting-up standards. Although it is widely recognised that the PHP life cycle (i.e., the analysis, action plan design, execution, monitoring and evaluation of public health policies) should be evidenced based, current support for it is mainly in the form of guidelines, and is not supported by data analytics and decision making tools tailored to it. In this paper, we present a novel model driven approach to PHP life cycle management and an integrated platform for realising this life cycle. Our approach is based on PHP decision making models. Such models steer the PHP decision making process by defining the data that need to be collected and the ways in which these data should be analysed in

order to produce the evidence required for PHP making. Our work is part of a new research programme on public health policy making for the management of hearing loss, called EVOTION, that is funded by the European Union.

- Themis P Exarchos, George Rigas, Athanasios Bibas, Dimitrios Kikidis, Christos Nikitas, Floris L Wuyts, Berina Ihtijarevic, Leen Maes, Massimo Cenciarini, Christoph Maurer, Nora Macdonald, D-E Bamiou, Linda Luxon, Marios Prasinos, George Spanoudakis, Dimitrios D Koutsouris, Dimitrios I Fotiadis "Multiparametric data analysis for diagnostic decision support in balance disorders." Biomedical and Health Informatics (BHI), 2016 IEEE-EMBS International Conference on. IEEE, 2016.

  Abstract: In this work we present a framework for the analysis and mining of multiparametric data related to balance disorders. The overall concept is to define the schema of the analysis that provides optimal results for diagnostic decision support in balance disorders. The work is part of the integrated EMBalance platform which targets the management of patients with balance disorders, from the diagnosis to treatment and evolution of the disease. The obtained results in four different balance disorders range from 76.4% to 92.1%.

- Amal Anwer, Marios Prasinos, Doris Eva Bamiou, Nora Macdonald, Marousa Pavlou, Themis P Exarchos, George Spanoudakis, Linda Luxon "EMBalance data repository modeling and clinical application." 2015 IEEE 15th International Conference on Bioinformatics and Bioengineering (BIBE).

  Abstract: Dizziness is a common symptom for both benign and life-threatening disorders with subtle distinguishing features. This poses a clinical challenge for physicians dealing with patients suffering from dizziness and vertigo and managing them within primary care. The objective of the EMBalance project is to present a decision support system to assist general practitioners in the diagnosis and management of vestibular disorders. In this work we review the modeling techniques integrated with clinical data to produce a multi-scale, patient-specific balance model that is incorporated in the DSS based on data mining techniques. To understand this we have outlined both technical and clinical aspects to the project. Further we discuss how we intend to test this product in a multicentred, double blind, parallel group randomized controlled trial and the impact we expect the DSS to have both clinically and technologically.

- Themis P Exarchos, Kostas Stefanou, George Rigas, Athanasios Bibas, Dimitris Kikidis, Christos Nikitas, Floris L Wuyts, Berina Ihtijarevic, Leen Maes, Massimo Cenciarini, Christoph Maurer, Dimitra Iliopoulou, Nora Macdonald, Doris Eva Bamiou, Linda Luxon, Marios Prasinos, George Spanoudakis, Dimitrios Koutsouris, Dimitrios I Fotiadis "Diagnosis of balance disorders using decision support systems based on data mining techniques." 2015 IEEE 15th International Conference on Bioinformatics and Bioengineering (BIBE).

Abstract: In this work we present the decision support of the EMBalance platform. EMBalance is a platform for the management of balance disorders in terms of diagnosis, treatment and evolution. The EMBalance platform aims to extend existing but generic and currently uncoupled balance modeling activities leading to a multi-scale and patient-specific balance model, which will be incorporated in a Decision Support System (DSS), towards the early diagnosis, prediction and the efficient treatment planning of balance disorders. The diagnosis part of the decision support system uses various data ranging from demographic characteristics to clinical examinations, auditory and vestibular tests. Currently we present some initial technical choices and indicative results of the decision support system for diagnosing balance disorders, based on data mining techniques and clinical guidelines.

- Nora MacDonald, Marios Prasinos, Themis Exarchos, Dimitris Kikidis, Laura Rammazzo, Amal Anwer, Leen Maes, Laura Celis, Floris Wuyts, Christoph Maurer, Marousa Pavlou, Athanasios Bibas, George Spanoudakis, Linda Luxon, Doris Bamiou (2016). EMBalance data repository modeling and clinical application (Vol. 26, pp. 194–194). Presented at the 29th Bárány Society Meeting.

  Abstract: Dizziness is a common symptom for both benign and life-threatening disorders with subtle distinguishing features. This poses a clinical challenge for physicians when faced with patients suffering from dizziness and vertigo within primary care. Consequently, patients often get misdiagnosed and inappropriately managed which leads to personal burden on both the patient and heath economics. The EU funded EMBalance project aims to develop the first Decision Support System (DSS) that will support not only the clinical decision-making towards accurate and early diagnosis, but also the efficient treatment planning of balance disorders. In this work we present the data mining modeling techniques that have been applied on clinical data to produce a multi-scale, patient-specific balance model that is incorporated in the DSS. Furthermore, we present the clinical proof of concept validation methods and preliminary results by means of multicentred, parallel group, randomised controlled trials. The final outcome of the EMBalance project will be a powerful web-based platform that will be provided to primary and secondary care physicians across specialties, levels of training and geographical boundaries, towards the early diagnostic evaluation, disease course prediction and effective management planning of balance problems.

Journals

- Themis P Exarchos, George Rigas, Athanasios Bibas, Dimitrios Kikidis, Christos Nikitas, Floris L Wuyts, Berina Ihtijarevic, Leen Maes, Massimo Cenciarini, Christoph Maurer, Nora Macdonald, D-E Bamiou, Linda Luxon, Marios Prasinos, George Spanoudakis, Dimitrios D Koutsouris, Dimitrios I Fotiadis "Mining balance disorders' data for the development of diagnostic decision support systems." Computers in Biology and Medicine 77 (2016): 240-248.

  Abstract: In this work we present the methodology for the development of the EMBalance diagnostic Decision Support System (DSS) for balance disorders. Medical data from patients with balance disorders have been analysed using data mining techniques for the development of the diagnostic DSS. The proposed methodology uses various data, ranging from demographic characteristics to clinical examination, auditory and vestibular tests, in order to provide an accurate diagnosis. The system aims to provide decision support for general practitioners (GPs) and experts in the diagnosis of balance disorders as well as to provide recommendations for the appropriate information and data to be requested at each step of the diagnostic process. Detailed results are provided for the diagnosis of 12 balance disorders, both for GPs and experts. Overall, the reported accuracy ranges from 59.3 to 89.8% for GPs and from 74.3 to 92.1% for experts.

Book Chapters

- K Bougoulias, M Prasinos, I Kouris, K Giokas, D Koutsouris. "Ob/Gyn EMR Software: A Solution for Obstetricians." Design, Development, and Integration of Reliable Electronic Healthcare Platforms (2016): 101.

  Abstract: In this chapter, the collection and the analysis for the development of an Ob/Gyn EMR software for small Obstetrics and Gynecology organizations is analysed. The necessary gynecological information was gathered via research concerning the needs of the practice and was organized and categorized according to its importance to the clinicians. The user interface of the developed software provides access to obstetrics, gynecological, surgical, sterilization and PAP test data, along with video and image file storage capabilities. An integrated appointment scheduling module, as well as an expected labor day prediction module, are also part of the application. The developed software is self-contained so that it can be installed on the clinician's computer or accessed within the clinic.

- K Giokas, Y Makris, A Paidi, M Prasinos, D Iliopoulou, D Koutsouris "Global and Local Health Information, Databases, and Networks." Telehealth and Mobile Health. CRC Press, 2015. 233-250.

  Abstract: In this chapter, we discuss health data that are collected and are accessed on local and global levels. Health data on the local level includes all sources that generate, collect, and record health data citywide, region-wide, and nationwide.

1.7. Outline of Thesis

The rest of the thesis is structured as follows. Chapter 2 presents the literature review conducted during this research, 0 presents the platform's information flow and the modeling language developed for the purposes of the platform, Chapter 4 gives some details about the implementation of the platform, Chapter 5 presents the evaluation performed and the experimental results and finally, Chapter 6 includes concluding remarks, limitations and future work.

Chapter 2

Literature Review

2.1. Overview

The purpose of this section is to explore related work in the following areas of literature:

- Public health policy: formation processes and guidelines
- Public health policy making tools
- Ontologies in the domain of public policy making
- Ontologies in the domain of data mining
- Ontologies in the domain of statistical analysis
- Data analytics technologies
- Big data platforms

The Public health policy formation processes and guidelines were reviewed as a basis for the policy making tool prototype we provide as part of this thesis. Also, a search for other existing public health policy tools was conducted, but none was found. In addition, existing ontologies in the domains of public policy making, data mining and statistical analysis were reviewed in order to be based on existing research and create a novel ontology that combines these domains. Finally, data analytics technologies and big data platforms were explored, to develop a big data analytics engine for the execution of data analytics workflows, based on existing technologies.

2.2. Policy Formation processes and/or guidelines

2.2.1. Phases of the policy life cycle

2.2.1.1. Overview

A health policy is qualified as public if it is made by public institutions and for large groups of populations at regional, national or even international level. Public health policy making consists of four key stages:

2.2.1.2. Situational analysis

This stage is concerned with the assessment of the needs and gaps, the resources available, and eventually the gaps and the strengths weaknesses, opportunities and threats (SWOT) arising in connection with a situation that needs to be addressed by health policy.

2.2.1.3. Development of action plan

This stage is concerned with setting the initial aim, objectives, activities and all priorities for implementing a health policy programme, and identifying the resources needed for this implementation.

2.2.1.4. Implementation and monitoring

This stage is concerned with the execution of the action plan for implementing a health programme and monitoring the adherence of implementation activities to the action plan.

2.2.1.5. Programme evaluation

This stage is concerned with the assessment of the effects and other outcomes of implemented health policy programmes.

2.2.2. Canadian Foundation processes for healthcare improvement

2.2.2.1. Overview

The Canadian Foundation for Healthcare Improvement (CFHI) has developed a framework of processes to support Evidence-Informed Health Policy making [11]. These processes are aimed at ensuring that relevant research is identified, appraised and used to inform decisions about the formation of health policies and programs. The processes have been written for people responsible for health policy decision-making (e.g., health system managers and policy-makers) and those who support them. The CHFI framework addresses all four stages of policy formation, albeit to a different extent.

2.2.2.2. Situational analysis

The focus of situational analysis in the CFHI framework is on identifying and analysing the health service needs of a population. In order to identify the population needs CHFI monitors its performance against its population goals and addresses the needs of its patient population by engaging healthcare providers and front-line managers. Their aim is to develop clinical leadership for improvement initiatives.

2.2.2.3. Development of action plan

In the CFHI initiative, the steps that need to be taken in order to develop an action plan are the following:

- Analyse existing improvement frameworks for healthcare organizations in order to identify the key attributes of high-performing such organizations.
- Consider the identified attributes in light of hands-on and practical experience (of CFHI) in animating healthcare improvement in a specific context (i.e., Canada in the case of CFHI).
- Select key targeted levers for healthcare improvement. The framework identifies six levers of improvement and provides questionnaires for making assessments in reference to these six levels. The six levers and questionnaires are related to: (1) focusing on population needs, (2) engaging healthcare providers and front-line managers in creating the improvement initiatives/culture, (3) building organizational capacity, (4) creating supportive policy and incentives, (5) engaging patients and citizens, and (6) promoting evidence-informed decision-making.
- Develop assessment questions per each lever that further clarify the need for and pathway to change.
- Review and revise action plan based on consultation with health service, policy and quality improvement leaders.
- Pilot the action plan with healthcare delivery and policy organizations to test and improve its application.

2.2.2.4. Implementation and monitoring

In the CFHI framework, the monitoring of the realization of an action program is performed by a team of healthcare leaders. Monitoring is based on a rating system using an “1-5” point scale where 1 means “Strongly disagree” and 5 is “Strongly Agree”.

2.2.2.5. Programme evaluation

Policy evaluation in the CFHI framework is the responsibility of healthcare leaders. The framework expects the collection of evidence that can help an organization (or a health system) to

- assess how it performs
- identify the available improvement expertise, assets and strengths (after the completion of assessment a group should spend time to understand the strengths and weaknesses according to their own responses to each of the six layers as they described in implementation and monitoring part).
- develop its improvement capacity, and
- undertake the next step for healthcare improvement.

2.2.2.6. Relation to this thesis

The CHFI framework was useful in the context of this thesis in providing inputs to the formulation of our public health policy decision making modeling language as well as the evaluation scenario.

2.2.3. World Health Organization guide for Health Impact Assessment

2.2.3.1. Overview

WHO offers a number of short guides for health impact assessment (HIA). These support the identification and improvement of consequences of policies or activities on health. HIA was developed for the National Health System (NHS) Executive London [12]. To understand whether to carry out a health impact assessment in a project/policy in Sweden, for example, the authors created a checklist of the items to be considered. These items were [13]:

- the description of the policy
- questioning whether the policy affects selected determinants of health. These determinants are: participation in influence on the society; economic and social security; safe and favourable growing up conditions; healthy working life; sound and safe environments and products; health promoting medical care; physical activity; eating habits and safe food; tobacco, alcohol, illicit drugs, doping and gambling; and prevention of infectious diseases.
- questioning whether the policy affects the whole population or 13 selected vulnerable groups by gender. The 13 vulnerable groups are: children, adults, elderly, chronically ill, people with a handicap/impairment or allergies, people with an addiction, alcohol, drugs etc., unemployed, immigrants, refugees, single-parents, people with low income, homeless people and homosexuals.

The strength of HIA is that its recommendations is connected to a group of stakeholders as participants where they provide a fully considered view issues that affect the health of local communities.

2.2.3.2. Situational analysis

The situational analysis in HIA is based on understanding whether HIA as a tool could be used overall. Assuming that HIA is worth pursuing, situational analysis is based on [12]:

- Screening – HIA may not be possible to use on every project, policy or programme. To determine when to use it, HIA has an initial screening step. Screening involves a quick assessment of the potential of a programme/policy to affect the health of the population. Although screening is the typical first step in deciding whether HIA should be used, there have been cases where it has been skipped, following a decision by key stakeholders (e.g., public health authorities, health planners, managers) and funding is available for it [14].
- Scoping – Scoping is used in cases where screening has identified a positive potential for a programme/policy and involves four key tasks. These are: (a) to establish the boundaries for the appraisal of health impact of the programme, (b) to come to an agreement regarding the way in which the appraisal of the programme/policy will be managed, (c) to decide who will be responsible for decision making, and (d) to agree how to monitor and evaluate the HIA process.

2.2.3.3. Development of action plan

The steps that may be taken to create the action plan in HIA are:

- The appraisal of the potential health effects/impacts. Appraisal is based on analysing the policy, programme or project, profiling the affected population, identifying and characterizing the potential health impacts, and reporting on the impacts and making recommendations for the management of those impacts [12].
- Decision-making. Decision making is about deciding on action plan based on the outcomes of appraisal. The stakeholders, who participate in this stage, are those who have been agreed to do so during scoping.

2.2.3.4. Implementation and monitoring

The implementation of a HIA programme involves the following steps [12]: (1) identifying expertise that already exists within the organization/ partnership and could be deployed in support of HIA; (2) raise awareness about HIA within the organizations involved; and (3) perform a rapid appraisal of possible starting points. HIA does not make special provision for monitoring.

2.2.3.5. Programme evaluation

Whilst HIA recognizes the need for programme evaluation and views this process as a process that should be driven by the impact on health and health services, it does not offer a special process for programme evaluation.

2.2.3.6. Relation to this thesis

HIA was useful for this thesis in making it clearer to understand the key features of health impact assessment.

2.2.4. Applicability and Transferability of Evidence Tool

2.2.4.1. Overview

The applicability and transferability of evidence tool (A&T Tool) was first introduced in 2007 [15]. Subsequently, it was updated by the National Collaborating Centre for Methods and Tools [16] to help public health managers and planners make decisions about local health planning priorities. However, although it is referred to as a "tool" by its creators, for the purposes of our survey will refer to it as a method since it does not constitute a computer based public health policy making tool. The A&T tool can be used by public health decision makers who want to incorporate high-quality evidence in their planning as an aid to determine whether a policy or program is relevant or feasible. Created for public health, A&T Tool includes questions relevant to public health context such as collaboration with stakeholders, needs for local implementation, assessment of the political and organizational climates, and evaluation of the costs related to outcome. This tool can be used by decision-makers in any public health program area.

2.2.4.2. Situational analysis

In A&T Tool, situational analysis is carried out by setting first the users' aims for the method (e.g., to explore whether and how to apply evidence into public health decision making and also policy making). The questions A&T Tool provides are aimed at evaluating political acceptability or leverage, social acceptability, available resources and organizational expertise and capacity. The stakeholders, who should be involved in situational analysis involve inter-sectorial, multidisciplinary and consumer groups. According to NCCMT [17], the preliminary steps that someone should follow are: (i) to generate a question to drive literature search and review process; (ii) to search and retrieve relevant literature, and (iii) to critically appraise the literature. It is also worth mentioning that users of A&T Tool can assign their own scoring system for the tool.

2.2.4.3. Development of action plan

As suggested by NCCMT [18], the development of an action plan for the study of the literature involves 6 steps. These are:

- Establishing a facilitator for the overall process who can act as group leader and maintain timelines.
- Select key stakeholders to form a group that will make use of the A&T Tool and the method that it imposes.
- Select questions for assessing the applicability and transferability of alternative options that are most important for the intervention of interest and local context.
- Identify a scoring system for the assessment questions (e.g. an 1-to-5 point scale where 1 is low level impact and 5 is high level impact). Priority goes to the highest scoring intervention or program.
- Rating the importance of different criteria.
- Document the scoring process used by the group.

Although the above is relevant to the study of the literature it does not constitute public health policy action plan making as such.

2.2.4.4. Implementation and monitoring

The A&T tool does not support public health policy implementation and monitoring.

2.2.4.5. Programme evaluation

The A&T tool does not support programme evaluation.

2.2.4.6. Relation to this thesis

The A&T tool was useful to this thesis as a method for directing investigations of the literature as part of public health policy decision making. Although this process in this thesis is based mainly on analysis of real data, the analysis of relevant literature can also be important for assessing and confirming the outcomes of data analysis.

2.2.5. The Delphi Method

2.2.5.1. Overview

Delphi [19] is a method that is aimed at producing information suitable for decision making. This method is based on a structured process that collects knowledge through an iterative process whereby the knowledge is refined until a consensus is reached amongst a group of experts using questionnaires. The overall aim of this method is to support judgmental or heuristic decision making in the fields of both social policy and public health. The method was created in order to improve the exchange of information, to support social policy and public health-related agencies, as well as other decision-making bodies. Delphi is also a method for improving the generation of critical ideas and process the information collected from experts.

2.2.5.2. Situational analysis

The method comprises a series of questionnaires sent to a group of experts. The questionnaires are designed to develop individual responses to the problems posed and to enable the experts to identify whether the work's progress is going according to the plan or not. There are two phases in the Delphi method. The first one is characterized as the 'exploration phase' and the second one 'evaluation phase'. In the first phase the subject under discussion is explored and then additional information is provided. The second phase involves the process of assessing and gathering the expert's views. If there is a disagreement, then this can be explored further in order to find a solution. The Delphi method has four outcomes [20]. These are the areas of agreement, the areas of disagreement, the areas needing clarification and understanding areas. The overall expectation of the Delphi method is that the judgements achieved within a group through the Delphi method are more reliable than individual judgments.

2.2.5.3. Development of action plan

The development of an action plan for public health policy could be based on the use of Delphi questionnaires.

2.2.5.4. Implementation and monitoring

The Delphi method does not support implementation and monitoring of public health policy explicitly. Delphi questionnaires may, nevertheless, be used for monitoring.

2.2.5.5. Programme evaluation

The Delphi method do not cover health programme evaluation. Delphi questionnaires may, nevertheless, be used for programme evaluation.

2.2.5.6. Relation to this thesis

The Delphi method is useful in devising questionnaires that can aid setting up the objectives and action plans for public health policies, and for monitoring and evaluation of public health policy realization programmes. The Delphi method was also useful for this thesis in formulating the evaluation scenario.

2.2.6. Planning and monitoring of national strategies manual

2.2.6.1. Overview

The World Health Organization has developed a manual for planning and monitoring of national health strategies. The national strategies are professional programmes aiming in the development of a holistic and integrated strategic plan for the provision of effective and sustainable health services. The reason behind the creation of this manual was the need to provide effective and sustainable ear and hearing care services. The manual provides guidance on how to develop and implement such a strategy. The goal of the guidance is to: raise awareness about ear and hearing problems among individuals and communities tailored and targeted separately to the general public, policy-makers, programme managers and funding providers [21]. By doing this it will help the political commitment and also the need for collecting resources in order to develop a plan and a strategy.

2.2.6.2. Situational analysis

To analyse a situation related to the development of a strategy for ear and hearing services, the WHO manual suggests the evaluation of the needs of the population and the resources available for addressing them. To accomplish this evaluation, the manual suggests [21]

- assessing the magnitude and profile (type, causes, age pattern, geographical distribution) of hearing loss and ear diseases;
- obtaining general country information, including population profile, socioeconomic profile and health indicators;
- determining the health system infrastructure and organization;
- assessing the availability of human resources;
- determining what ear and hearing care services are available;
- Performing stakeholder analysis.

The manual also suggests the use of strengths, weakness opportunities and threats (SWOT) analysis. In this context, strengths may be related to availability of trained health workers working in the community, who can be engaged to deliver ear and hearing care services; weaknesses may be related to lack of trained audiologists to provide specialized services; opportunities may be related to increasing engagement in the country of an nongovernmental organization(NGO) working in the field of hearing care, and threats may be related to political unrest. These are just examples provided by the developers of the manual and do not constitute the only SWOT's that should be considered.

2.2.6.3. Development of action plan

The development of an action plan in this approach is based on the overall aim that needs to be achieved. This aim must be specified after taking into account the different views of the stakeholders. It must also be agreed on by all the parties involved and not just the majority of them. The manual provides examples of possible aims such as: "To reduce the overall prevalence of hearing loss in the country by 25%" or "To provide equitable access and coverage of cost-effective, quality health services for ear and hearing care, as close to the people of the country as possible" [21]. The manual introduces also the concept of a "road map" for achieving the aim. The "road map" states the things that need to be done and sets the activities needed to achieve an objective. In the development of an action plan, practical difficulties and the available resources need also to be considered. Lastly, any objectives set towards achieving the overall aim should be [21]:

- Specific, i.e., clearly focused on a particular result
- Measurable, i.e., each objective should have a precise measurable target
- Achievable, i.e., the objective is feasible and can be achieved in the time set
- Realistic, i.e., each objective should be considered with regard to constraints such as resources, personnel, cost, and time frame required for it
- Time-bound, i.e., a timeline should be specified for their achievement.

2.2.6.4. Implementation and monitoring

The implementation of a programme goes through three key phases:

- Pilot phase: In this phase the feasibility and the proposed interventions of the strategy or plan first tested in one or more parts of the country and then in a national level. This happens in order to refine the plan on the basis of feedback. This phase is closely monitored and evaluated.
- Expansion phase: In the pilot phase, due to negative feedbacks, some interventions may have to be dropped. Others may be added if there is supportive feedback from the community.

- Evaluation phase: Evaluation follows the implementation phase and focuses on assessing the overall implementation and impact of the strategy.

Monitoring in WHO's manual is aimed at correcting deviations from objectives and improving performance[21]. Every programme must have a national committee. This committee forms a task force in order to draft sections of the national strategy document. The members of the task force need to have knowledge about the "country's health system and public health approach to ear and hearing care, as well as its medical, surgical, rehabilitative and social aspects" [21]. The task force needs to work with all stakeholders and members of national committee and take in consideration the views and interests of all stakeholders participating in strategy development. The monitoring of the development of strategy should be done by the task force that has been set for the program to check that it is on point. The problems that may occur should be solved before the strategy is finalized. The monitoring of the national strategy is based on:

- an appropriate set of indicators that measures the day-to-day achievements of the strategy or process being monitored
- monitoring tools that allow the systematic collection of relevant information.

2.2.6.5. ***Programme evaluation***

The evaluation of a program can be carried out at several stages in the life cycle of a program [21] According to the stage that it is conducted an evaluation is:

- ongoing evaluation: carried out at the end of a pre-agreed period or midway through a strategy or programme
- terminal evaluation: within 6–12 months after completion of a programme
- ex-post evaluation: after several years, when the full impact could be expected to have been realized.

2.2.6.6. ***Relation to this thesis***

The approach described above can be used as guidance for developing public health policies and it was useful in formulating the public health policy decision making model for the evaluation scenario.

2.2.7. ***Deloitte's framework for assessing hearing services***

2.2.7.1. Overview

Action on Hearing loss (AHL) is a non-profit organization that aims to help patients suffering from hearing loss (HL) in the UK by providing advice, communication services, and day-to-day care. AHL commissioned Deloitte to create a framework on hearing services that could be used to evaluate and compare hearing service providers. Their aim was to help adult HL patients to make informed choices regarding services and service providers, by making it easier to collect report information. The framework that Deloitte created is innovative as it enables the identification of new areas of hearing services and evaluating qualified service providers (referred to as "Any Qualified Provider" or AQP in the framework). The framework also creates a collection of information on important features that will be useful for service users and other stakeholders to consider. Although this framework does not constitute a policy making tool, process or modeling framework, it is relevant to policy making through the provision of a means for evaluating hearing services. To this end, we have included it in our review.

2.2.7.2. Situational analysis

This framework focuses on the analysis and evaluation of hearing services. Hence, situational analysis in the context of this framework is supported in only as far as service situational analysis is concerned. To carry out such analysis, the framework identifies key service performance domains and performance indicators. The performance domains are [22]:

- Accessibility and responsiveness – This performance domain considers whether a service is located near service users, it has flexible appointment times and acceptable waiting/response times.
- Integration with other services – This performance domain considers whether there is a smooth service user transition between services and if the related process is efficient and effective.
- Public health outcomes – This performance domain considers if the service improves service user hearing and quality of life.
- Cost – This performance domain considers if the cost of the service to the service user is affordable.
- Service user focus – This performance domain considers whether service users are satisfied, if there was an individual management plan and if the information provided to service users is helpful and sufficient.

- Quality – This performance domain considers if service providers follow guidelines and professional standards and if their staff are qualified and follows the established processes.
- Safety – This performance domain considers if the service has any related adverse events that have been reported by the provider.
- Innovation – This performance domain considers whether there is a systematic approach to introduce service improvements based on user feedback.

The hearing services targeted by this framework include hearing assessment, hearing aid fitting, follow up visits and aftercare services. Each of these services is examined and assessed in terms of inputs (i.e., the resources invested in undertaking service activities), activities (i.e., the steps taken during the consultation between the service user and hearing service), outputs (i.e., the immediate deliverables of the service) and outcome (i.e., the impact of the service within 6 months (short term) and beyond 6 months (long term)) in order to form assessments with regards to the key performance area.

Evaluation in the context of Deloitte's framework has three different categories: evaluation of the impact of AQP, evaluation of the service users' access to hearing services and lastly, evaluation of the service value. The evaluation of the impact of AQP could be done by both registered and non-registered service providers because the performance indicator outcomes would be compared between registered and non-registered services in the same geographic area. The second category was created in order to determine whether there is equality in access of hearing services for different groups. There is also a way to measure access to hearing services as the total numbers of service users.

2.2.7.3. Development of action plan

This approach does not include the development of an action plan.

2.2.7.4. Implementation and monitoring

This framework does not cover the implementation and monitoring of health policies.

2.2.7.5. Programme evaluation

This framework does not cover the evaluation of health policies. However, the performance indicators that it proposes for the evaluation of hearing services could also be useful as criteria for evaluating certain types of hearing service related health policy programmes.

2.2.7.6. Relation to this thesis

The performance indicators that are proposed by this framework for the evaluation of hearing services could also be useful as criteria for evaluating certain types of hearing service related health policy programmes.

2.2.8. WHO's situational analysis tool for hearing care

2.2.8.1. Overview

WHO developed a tool [23] to support situational analysis for hearing care. Despite having been termed as "tool" by WHO, for the purposes of our survey this tool constitutes a method rather than a computer-based tool. Hence, we will refer to it as "WHO method" in the following. The WHO method includes a questionnaire and an annex providing guidance on how to complete it. The aim of the questionnaire is to collect evidence that is necessary for situational analysis to develop a strategy for national ear and hearing care. This questionnaire can help preventing, identifying and treating ear diseases and hearing loss as well as rehabilitating and supporting people with hearing loss. The questionnaire focuses on the country profile, burden of disease, epidemiology of hearing loss etc. Beyond the questionnaire, the WHO method also suggests collecting information from stakeholders through interviews. After the information is collected and analysed them as documentary evidence and in cases where this is not possible, the source or person that provided the information must be included.

2.2.8.2. Situational Analysis

The WHO method aids the understanding of the epidemiology of hearing loss and the status of they existed systems that support related hearing care. The method has also created an opportunity to reduce the gap between ideal situations and existed ones.

2.2.8.3. Development of Action Plan

WHO has also created an action plan formation manual as a complement to its questionnaire [23]. The aim of this manual has been to provide guidance on action plan formation following a situational analysis. The given manual is an annex that provides guidance on how to complete the questionnaire and on resources that may provide needed information. More specifically, for each stage there is a guidance on how one can find the detailed information needed in order to complete this step and also the resources that are available.

2.2.8.4. <u>Implementation and monitoring</u>

The WHO method does not cover this stage of policy making explicitly.

2.2.8.5. <u>Programme evaluation</u>

The WHO method does not cover this stage of policy making explicitly.

2.2.8.6. <u>Relation to this thesis</u>

The use of the questionnaire developed by WHO could be useful for collecting evidence.

2.2.9. <u>Summary and comparison of policy formation processes</u>

Below in Table 2-1 we summarize the features of the policy formation processes reviewed in Section 2.2. We see that the most complete processes are the Canadian Foundation processes for healthcare improvement and the Planning and monitoring of national strategies manual.

Table 2-1 Policy Formation Processes

Policy formation process name	**Situational Analysis**	**Development of action plan**	**Implementation**	**Monitoring**	**Programme Evaluation**
Canadian Foundation processes for healthcare improvement[11]	✓	✓	✓	✓	✓
World Health Organization guide for Health Impact Assessment [12]	✓	✓	✓	X	X

Policy formation process name	**Situational Analysis**	**Development of action plan**	**Implementation**	**Monitoring**	**Programme Evaluation**
Applicability and Transferability of Evidence Tool[15]	✓	✓	X	X	X
The Delphi Method [19]	✓	✓	X	X	✓
Planning and monitoring of national strategies manual [21]	✓	✓	✓	✓	✓
Deloitte's framework for assessing hearing services [22]	✓	X	X	X	X
WHO's situational analysis tool for hearing care [23]	✓	✓	X	X	X

2.3. Policy Making Tools

None of the existing papers and articles that have been examined would qualify as computer-based tools for public health policy making.

2.4. Ontologies in the domain of Public Policy Making

*2.4.1. **Government (G2G) Collaboration***

Loukis has developed an ontology-based approach for modeling public policies and managing them across their entire life cycle [24]. This approach has been developed with the intention to support policy modeling and management in a collaborative manner involving interactions between different stakeholders involved in such activities, and in particular cases where policy modeling involves collaboration between different government stakeholders (i.e., G2G collaboration).

Loukis has proposed a domain independent (referred to by the author as “horizontal”) ontology for modeling public policy processes, which – according to the author – could be used for governmental policy formation processes in different domains subject to extensions of the core horizontal ontology with domain specific ontologies. The modeling of policies in this approach is based on five core ontological concepts. These are: the *issue* (i.e., the problem to be solved or goal to be achieved by the policy); the *alternatives* (i.e., the alternative directions of action/ways in which the issue(s) can be addressed); the *positions* that different stakeholders may express on different alternatives (positions can be support or object to alternatives); the *preferences* that different stakeholders may express on different positions to indicate their relative importance; and the *criteria* that will be used to reach decisions.

The main concepts – i.e., kinds of elements (or classes) – of the ontology of Loukis for public policy making, implementation and evaluation, and the relations between them, are shown in Figure 1 below using the IDEF5 notation [25]: the kinds of elements (classes) are represented as circles while the kinds of relations (object properties) are represented as arrows. Below we present the ontology description from [24]. The ontology classes and the kinds of relations are shown in capital letters.

The basic elements dealt with in public policy analysis are called ISSUEs. An ISSUE could be a decision to be made, a goal to be achieved, a problem to be solved, a question to be answered, a concern or a basic requirement For example, an ISSUE could be where to build a new hospital. An ISSUE can GENERALIZE, SPECIALIZE, or QUESTION_REPLACE another ISSUE. Two important SUBKINDs of ISSUEs include the CAUSEs and SYMPTOMs. A SYMPTOM can be DUE TO a CAUSE. Four more important SUBKINDs of ISSUEs, which are used in Strategic SWOT (Strengths, Weaknesses, Opportunities and Threats) Analysis (e.g. geographical region, public organizations etc.) are the STRENGTHs, WEAKNESSes, OPPORTUNITYs and THREATs related to the ISSUE.

ALTERNATIVEs, another important element in public policy analysis, are general ways or broad directions of action, that can be used to RESOLVE a specific ISSUE. For example, ALTERNATIVES to the ISSUE of where to build a hospital might be the specific locations under consideration.

Legend
gsqr = generalise, specialise, question-replace
soc = supports, objects-to, comments-on

Figure 1 G2G Ontology adopted from [24]

When stakeholders discuss ALTERNATIVEs, they take POSITIONs with respect to each ALTERNATIVE. For example, a positive POSITION SUPPORTs an ALTERNATIVE, a negative POSITION OBJECTS-TO an ALTERNATIVE, and a neutral POSITION simply COMMENTS-ON an ALTERNATIVE. Additionally, a POSITION can be expressed on any other POSITION and even on the ISSUE itself. Stakeholders determine the relative importance of POSITIONS by expressing their PREFERENCES for each POSITION. A PREFERENCE concerns via the relation REFERS_TO a pair of expressed POSITIONs (p1, p2), and has the form [p1, preference operator, p2], where the preference operator can take values 'more important'(>), 'less important'(<), or 'equally important'(=). POSITIONs (positive, negative or just comments) can be expressed on PREFERENCES as well. In the hospital location example, everyone will state their POSITION regarding a proposed location (e.g. support the location, don't support the location, are neutral to the location) and then they will express their PREFERENCES with regards to each others' POSITIONS.

The stakeholders' discussion regarding ISSUEs, ALTERNATIVEs, POSITIONs and PREFERENCEs is often accompanied by a multicriteria evaluation of the ALTERNATIVEs. Therefore, another important kind of elements in our ontology is evaluation CRITERIONs. Evaluation CRITERIONs can be defined by law, past experience or even by the stakeholders themselves. For example, CRITERIONs might include city planning guidelines that restrict the hospital construction locations. Similarly to the ISSUEs, a CRITERION can GENERALIZE, SPECIALIZE or QUESTION_REPLACE another CRITERION. After the CRITERIONs are finalized, they can be used in the multicriteria evaluation of the ALTERNATIVEs proposed for this ISSUE. VALUEs (ratings) are assigned to ALTERNATIVE with respect to how they fulfil the CRITERIONs. It should be noted that the above kinds of elements and relations can be used both for the design and for the evaluation of public policies, while the subsequent ones described below only concern the implementation of public policies.

Following the multicriteria evaluation, one or more ALTERNATIVEs are selected and inputted to PROGRAMMEs (e.g. programmes, subprogrammes, measures, etc.). A PROGRAMME that directly IMPLEMENTs an ALTERNATIVE is a first level PROGRAMME while one that is PART_OF another PROGRAMME is a lower level PROGRAMME. Similarly, each of the lowest level PROGRAMMEs is analysed into PROJECTs, and again we can have PROJECTs of various levels (e.g. projects, subprojects, etc.). Each of the lowest level PROJECTs is then analysed into TASKs, and we can also have TASKs of various levels (e.g. tasks, subtasks, etc.); finally, for each TASK a number of DELIVERABLEs are defined, and also EXPENSEs and work ASSIGNMENTs are made. These PROGRAMMEs, PROJECTs, TASKs, EXPENSEs and ASSIGNMENTs can be discussed, so POSITIONs (positive, negative or just comments) can be expressed on them as well. Also for each work ASSIGNMENT usually some DOCUMENTs are produced, such as progress reports, etc., for which ISSUEs can be raised.

Our ontology, which is presented in detail in Chapter 3 , is based on Loukis approach for the policy stakeholders and decision-making processes part. We have also been inspired by G2G ontology, but created our own approach for the policy aims, objectives and actions module.

2.5. Ontologies in the domain of data mining

2.5.1. The Data Mining Optimization ontology

Hilario proposed the Data Mining Optimization (DMOP) Ontology with its main goal to support meta-mining (more commonly known as "meta-learning" or "meta-analysis") of data mining experiments to extract workflow patterns [26]. Meta-learning [27] is a subfield of machine learning where automatic learning algorithms are applied on metadata about machine learning experiments. The ontology contains representations of DM tasks, algorithms, models, workflows and experiments, limited to the case of propositional data mining (in a single table). In addition, the authors have developed a knowledge base by populating the ontology with instances. The DMOP ontology version 5.3 [28] started a preliminary alignment of classes and relations with the DOLCE (Descriptive Ontology for Linguistic) [29].

DMOP was constructed to be used for meta-learning [27], so it is not fit for our purpose.

2.5.2. The OntoDM ontology

The domain of data mining (DM) deals with analysing different types of data. The data typically used in data mining is in the format of a single table, with primitive datatypes as attributes. However, structured (complex) data, such as graphs, sequences, networks, text, image, multimedia and relational data, are receiving an increasing amount of interest in data mining [30]. A major challenge for the authors of Onto-DM was to treat and represent the mining of different types of structured data in a uniform fashion.

A reference modular ontology for the domain of data mining OntoDM [31] was proposed by Panov, directly motivated by the need for formalization of the data mining domain. The OntoDM ontology is designed and implemented by following ontology known practices and design principles. Its distinguishing feature is that it uses Basic Formal Ontology [32] (BFO) as an upper-level ontology and a template, a set of formally defined relations from Relational Ontology [33] (RO) and other state-of-the-art ontologies, and reuses classes and relations from the Ontology of Biomedical Investigations [34] (OBI), the Information Artifact Ontology [35] (IAO), and the Software Ontology

[33] (SWO). This ensures compatibility and connections with other ontologies and allow cross-domain reasoning capabilities.

The main ingredient in the process of data mining is the data. In OntoDM-core, they model the data with a data specification entity that describes the datatype of the underlying data. For this purpose, they import the mechanism for representing arbitrarily complex datatypes from OntoDT ontology (Thesis of Panov [36]).

In Figure 2 below, the vital classes of Onto-DM-core are presented.

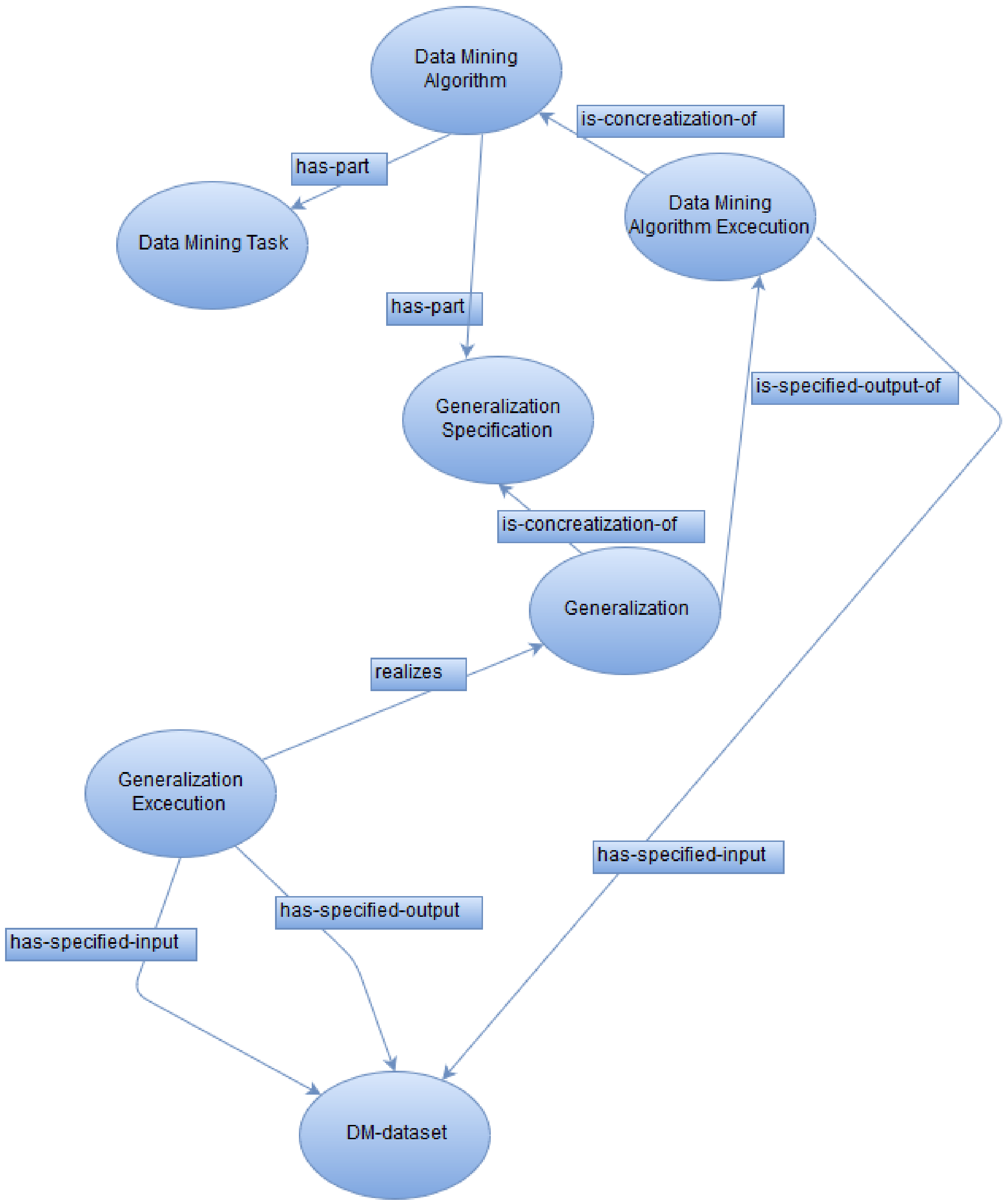


Figure 2 The main classes of Onto-DM core

In OntoDM-core, they distinguish between a descriptive data specification, that specifies the data used for descriptive purposes (e.g., in the clustering and pattern discovery), and output data specification, that specifies the data used for output purposes (e.g., classes/targets in predictive modeling). A tuple of primitives or a graph with boolean edges and discrete nodes are examples of data specified only by a descriptive specification. Feature-based data with primitive output and feature-based data with structured output are examples of data specified by both descriptive and output specifications.

OntoDM imports the IAO class dataset (a new ontology of information entities, originally driven by work by the OBI digital entity and realizable information entity branch, defined as ‘a data item that is an aggregate of other data items of the same type that have something in common’) and extends it by further specifying that a DM dataset has part data examples. OntoDM-core also defines the class dataset specification to enable reasoning about data and datasets. It specifies the type of the dataset based on the type of data it contains. Using data specifications and the taxonomy of datatypes from the OntoDT ontology, in OntoDM-core we build a taxonomy of datasets.

The task of data mining is to produce a generalization from given data. In OntoDM-core, they use the term generalization to denote the outcome of a data mining task. A data mining task is defined as sub-class of the IAO class objective specification. It is an objective specification that specifies the objective that a data mining algorithm needs to achieve when executed on a dataset to produce as output a generalization.

The definition of a data mining task depends directly on the data specification, and indirectly on the datatype of the data at hand. This allows us to form a taxonomy of data mining tasks based on the type of data. Džeroski proposes four basic classes of data mining tasks based on the generalizations that are produced as output: clustering, pattern discovery, probability distribution estimation, and predictive modeling. [37] These classes of tasks are included as the first level of the OntoDM-core data mining task taxonomy. They are fundamental and can be defined on an arbitrary type of data. An exception is the predictive modeling task that is defined on a pair of datatypes (for the descriptive and output data separately).

At the next levels, the taxonomy of data mining task depends on the datatype of the descriptive data (in the case of predictive modeling also on the datatype of the output data). If we focus only on the predictive modeling task and using the output data specification as a criterion, we distinguish between the primitive output prediction task and the structured output prediction task. In the first case, the output datatype is primitive (e.g., discrete, boolean or real); in the second case, it is some structured datatype (such as a tuple, set, sequence or graph).

Primitive output prediction tasks can be feature-based or structure-based, depending on the datatype of the descriptive part. The feature-based primitive output prediction tasks have a tuple of primitives (a set of primitive features) on the description side and a primitive datatype on the output side. This is the most exploited data mining task in traditional single-table data mining, described in all major data mining textbooks [38]. If we specify the output datatype in more detail, we have the binary classification task, the multi-class classification task and the regression task; where the output datatype is boolean, discrete or real, respectively. Structure-based primitive output prediction tasks operate on data that have some structured datatype (other than tuple of primitives) on the description side and a primitive datatype on the output side.

In a similar way, structured output prediction tasks can be feature-based, or structure-based. Feature-based structured output prediction tasks operate on data that have a tuple of primitives on the description side and a structured datatype on the output side. Structure-based structured output prediction tasks operate on data that have structured datatypes both on the description side and the output side.

The authors of Onto-DM focused just on feature-based structured output tasks and further specify a structured output datatype and represented a variety of structured output prediction tasks. For example, they represented the following tasks: multi-target prediction [39] (which has as output datatype tuple of primitives), multi-label classification [40] (having as output datatype set of discrete), time-series prediction [41] (having as output datatype sequence of real) and hierarchical classification [42] (having as output datatype labelled graph with boolean edges and discrete nodes). Multi-target prediction can be further divided into: multi-target binary classification, multi-target multi-class classification [43], and multi-target regression [44].

They take generalization to denote the outcome of a data mining task. In OntoDM-core, they consider and model three different aspects of generalizations, each aligned with a different description layer: the specification of a generalization, a generalization as a realizable entity, and the process of executing a generalization.

Many different types of generalizations have been considered in the data mining literature. The most fundamental types of generalizations, as proposed by Džeroski are in line with the data mining tasks [37]. These include clusterings, patterns, probability distributions, and predictive models.

In OntoDM-core, the generalization specification class is a subclass of the OBI class data representational model. It specifies the type of the generalization and includes as part the data specification for the data used to produce the generalization, and the generalization language, for the language in which the generalization is expressed. Examples of generalization language formalisms for the case of a predictive model include the languages of: trees, rules, Bayesian networks, graphical models, neural networks, etc.

As in the case of datasets and data mining tasks, we can construct a taxonomy of generalizations. In OntoDM-core, at the first level, we distinguish between a single generalization specification and an ensemble specification. Ensembles of generalizations have as parts single generalizations. We can further extend this taxonomy by taking into account the data mining task and the generalization language.

Generalizations have a dual nature [37]. They can be treated as data structures and as such represented, stored and manipulated. On the other hand, they act as functions and are executed, taking as input data examples and giving as output the result of applying the function to a data example. In OntoDM-core, we define a generalization as a sub-class of the BFO class realizable entity. It is an output from a data mining algorithm execution. The dual nature of generalizations in OntoDM-core is represented with two classes that belong to two different description layers: generalization representation, which is a sub-class of information content entity and belongs to the specification layer, and generalization execution, which is a subclass of planned process and belongs to the application layer.

A generalization representation is a sub-class of the IAO class information content entity. It represents a formalized description of the generalization, for instance in the form of a formula or text. For example, the output of a decision tree algorithm execution in any data mining software usually includes a textual representation of the generated decision tree. A generalization execution is a sub-class of the OBI class planned process that has as input a dataset and has as output another dataset. The output dataset is a result of applying the generalization to the examples from the input dataset.

A data mining algorithm is an algorithm (implemented in a computer program), designed to solve a data mining task. It takes as input a dataset of examples of a given datatype and produces as output a generalization (from a given class) on the given datatype. A specific data mining algorithm can typically handle examples of a limited set of datatypes: For example, a rule learning algorithm might handle only tuples of Boolean attributes and a boolean class. In the OntoDM-core ontological framework, we consider three aspects of the DM algorithm entity: a DM algorithm (as a specification), a DM algorithm implementation, and a DM algorithm execution.

Data mining algorithm as a specification is a subclass of the IAO class plan specification having as parts a data mining task, an action specification (reused from IAO), a generalization specification, and a document (reused from IAO). The data mining task defines the objective that the realized plan should fulfil at the end giving as output a generalization, while the action specification describes the actions of the data mining algorithm realized in the process of execution. The generalization specification denotes the type of generalization produced by executing the algorithm. Finally, having a document class as a part allows us to connect the algorithm to the annotations of documents (journal articles, workshop articles, technical reports) that publish knowledge about the algorithm. In analogy with the taxonomy of datasets, data mining tasks and generalizations, in OntoDM-core we also construct a taxonomy of datamining algorithms. As criteria, we use the datamining task and the generalization produced as the output of the execution of the algorithm.

Data mining algorithm implementation is defined as a sub-class of the BFO class realizable entity. It is a concretization of a data mining algorithm, in the form of a runnable computer program, and has as qualities parameters. The parameters of the algorithm affect its behaviour when the algorithm implementation is used as an operator. A parameter itself is specified by a parameter specification that includes its name and description.

In OntoDM-core, they define data mining software as a sub-class of directive information entity (reused from IAO). It represents a specification of a data mining algorithm implementation. It has as parts all the meta-information entities about the software implementation such as: source code, software version specification, programming language, software compiler specification, software manufacturer, the data mining software toolkit it belongs to, etc. Finally, a data mining software toolkit is a specification entity that contains as parts data mining software entities.

Data mining operator is defined as sub-class of the BFO class role. In that context, it is a role of a datamining algorithm implementation that is realized (executed) by a data mining algorithm execution process. The data mining operator has information about the specific parameter setting of the algorithm, in the context of the realization of the operator in the process of execution. The parameter setting is a subclass of data item (reused from IAO), which is a quality specification of a parameter. In OntoDM-core, we define data mining algorithm execution as a sub-class of planned process (reused from the OBI ontology). A data mining algorithm execution realizes (executes) a data mining operator, has as input a dataset, has as output a generalization, has as agent a computer, and achieves as a planned objective a data mining task.

A scenario is “a postulated sequence or development of events”. Therefore, a data mining scenario comprises a logical sequence of actions to infer some type of generalization from a dataset, a sequence of actions for applying a generalization on a new dataset, and a sequence of actions for evaluating the obtained generalizations. OntoDM-core represents a data mining scenario in three different description layers in the ontology: data mining scenario (as a specification), data mining workflow (as an implementation), and data mining workflow execution (as an application).

In OntoDM-core, a data mining scenario is an extension of the OBI class protocol. It includes as parts other information entities such as: title of scenario, scenario description, author of scenario, and document. From the protocol class it also inherits as parts objective specification and action specification. A data mining workflow is a concretization of a data mining scenario and extends the plan entity (defined by OBI). Finally, a data mining workflow is realized (executed) through a data mining workflow execution process.

We were inspired by OntoDM-core for the data mining part of our ontology. OntoDM-core is far more complex than our needs, but it was a good basis for modeling our approach.

2.6. Ontologies in the domain of Statistics

2.6.1. Ontology of Biological and Clinical Statistics (OBCS)

The Ontology of Biological and Clinical Statistics (OBCS) [45] is an ontology in the domain of biological and clinical statistics. OBCS is primarily targeted for statistical term representation in the fields in biological, biomedical, and clinical domains.

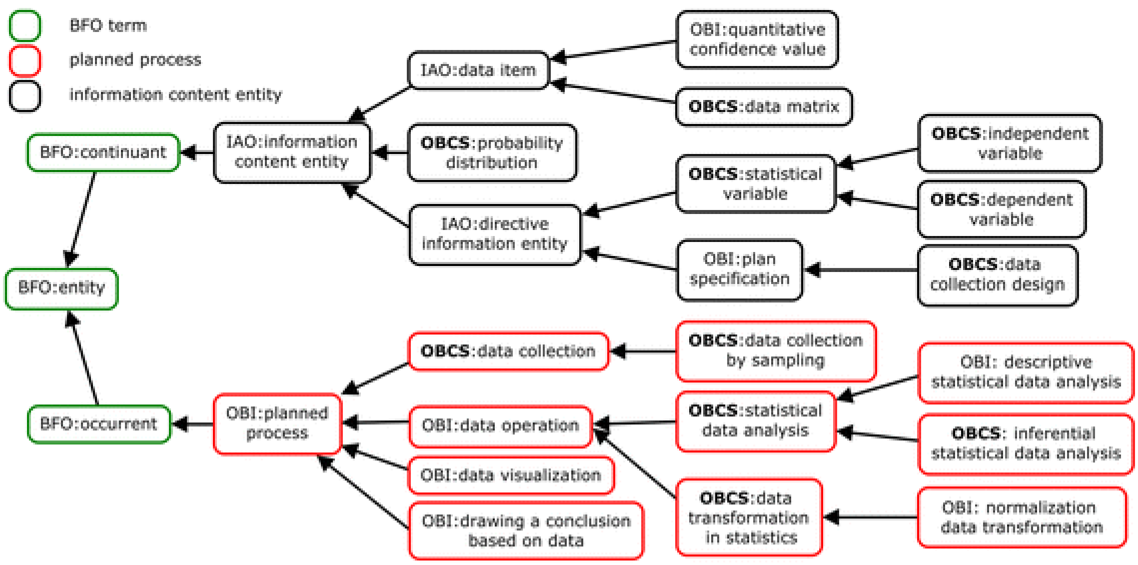


Figure 3 The top level OBCS hierarchical structure and key ontology terms

Figure 3, which was introduced by Zheng et al [46] shows the top level OBCS hierarchical structure and key ontology terms. The terms shown in boxes with the prefix "OBCS:" in bold font are OBCS-specific terms, and the other terms are imported from existing ontologies including BFO, IAO and OBI

OBCS uses the Basic Formal Ontology [32] (BFO) as the upper level ontology. OBCS imports all biostatistics related terms in the Ontology for Biomedical Investigations [34] (OBI) including all logical axioms. 138 new statistics terms have been generated with the "OBCS_" prefix.

OBCS was initiated with an ANOVA meta-analysis of vaccine protection assay, which was presented in the Bio-Ontologies 2010 conference [47]. Many statistics-related terms (e.g., ANOVA, survival rate analysis) were added into OBI at that time. Many more statistics terms have been added into OBCS with original name of "OBIstat" and discussed in the OBI face-to-face workshop in Ann Arbor in May 2012

The OBCS development is driven by many applications including vaccine research statistics study and nursing clinical statistical data collection and analysis.

OBCS forms the statistical domain sufficiently and could be a good basis to form the statistical analysis part of our approach, but it does not include different statistical analysis algorithms.

2.6.2. STATO Ontology

STATO is a general-purpose statistics ontology, whose aim is to provide coverage for statistical processes such as statistical tests, the conditions of their application, and the information needed or resulting from statistical methods, such as probability distributions, variable, spread and variation metrics. STATO also covers aspects of experimental design and description of plots and graphical representations commonly used to provide visual cues of data distribution or layout and to assist review of the results.

STATO has been developed to interoperate with other OBO Foundry ontologies, hence relies on the Basics Formal Ontology [32] (BFO) as a top-level ontology and uses the Ontology for Biomedical Investigations [34] (OBI) as mid-level ontology.

STATO provides textual definitions for all terms, as well as formal definitions for most of the terms allowing automatic classification, for example, categorising the statistical methods depending on the nature of the variables used as input, the conditions and their domain of application.

STATO also benefits from:

1. extensive documentation with the provision of textual and formal definitions;
2. associated R code snippets via a dedicated 'R-command' annotation, to vacillate a 'learn and apply' approach in the popular R environment;
3. query examples documentation, highlighting how the ontology can be harnessed for reviewers/tutors/student alike.

STATO is set to provide:

1. a resource to help in the communication and reporting of scientific results for biologists, scientists using statistical methods. STATO can also currently support Publishers and Journal reviewers by helping reporting guideline compliance and standardizing annotation of result tables
2. a set of core classes for annotating statistical methods used in life, natural and biomedical sciences investigations, but also metrics and estimates generated by those methods and link to the hypothesis being evaluated to allow better representation and data review.
3. formal definitions of most common univariate statistical tests to provide a didactic framework for students and reviewers
4. a formal way of navigating the conditions of application of classic statistical tests and distinguishing them
5. a semantic framework to support the creation of standardized analysis reports to help with review of results
6. a specialized vocabulary enabling text mining of statistical analyses.

In our ontological framework we were based on STATO and used the classes that describe the statistical algorithms of our approach.

2.7. Data Analytics Technologies

2.7.1. The R Project for Statistical Computing

R [48] is a language and environment for statistical computing and graphics. It is a GNU project which is similar to the S language [49] and environment which was developed at Bell Laboratories. R can be considered as a different implementation of S. S is a language for the manipulation of objects. It aims to be both an interactive language (like, for example, a Unix shell language) as well as a complete programming language with some convenient object-oriented features.

R provides a wide variety of statistical (e.g. linear and nonlinear modeling, classical statistical tests, time-series analysis, classification, clustering) and graphical techniques, and is highly extensible through the use of user-submitted packages for specific functions or specific areas of study. The S language is often the vehicle of choice for research in statistical methodology, and R provides an Open Source route to participation in that activity.

We could use R as an alternative to WEKA for the development of our big data analytics engine. Although R offers a JAVA API and is extensible, it is rather complicated to work with R and big data.

2.7.2. RapidMiner

RapidMiner [50] is an open-source software platform for data science teams that unites data preparation, machine learning, and predictive model deployment. It is written in JAVA programming language.

The tool can be used for over a vast range of applications including for business applications, commercial applications, training, education, research, application development, machine learning.

RapidMiner offers the server as both on premise & in public/private cloud infrastructures. It has a client/server model as its base. RapidMiner comes with template-based frameworks that enable speedy delivery with reduced number of errors (which are quite commonly expected in manual code writing process).

Rapid Miner constitutes of three modules, namely

1. Rapid Miner Studio- This module is for workflow design, prototyping, validation etc.
2. Rapid Miner Server- To operate predictive data models created in studio
3. Rapid Miner Radio- Executes processes directly in Hadoop cluster to simplify predictive analysis.

We could use Rapid Miner as an alternative to WEKA for the development of our big data analytics engine. Although it is relatively easy to run analysis in Rapid Miner GUI and it also offers a JAVA API, it is not fit for our purpose, as it is not to be used for big data analytics.

2.7.3. Orange

Orange [51] is an open-source software suite for machine learning & data mining. It best aids the data visualization and is a component-based software. It has been written in Python computing language.

As it is a component-based software, the components of orange are called widgets. These widgets range from data visualization & pre-processing to an evaluation of algorithms and predictive modeling.

Widgets offer major functionalities like

- Showing data table and allowing to select features
- Reading the data
- Training predictors and to compare learning algorithms
- Visualizing data elements etc.

Orange allows users to make smarter decisions in short time by quickly comparing & analysing the data.

We could use Orange as an alternative to WEKA for the development of our big data analytics engine. Although it is easy to run analytics through Orange's GUI, it does not offer a JAVA API, so it is not fit for our purpose.

2.7.4. KNIME

KNIME [52] is an open-source integration platform for data analytics and reporting. It operates on the concept of the modular data pipeline. KNIME constitutes of various machine learning and data mining components embedded together.

KNIME has been used widely for pharmaceutical research[53]. In addition, it performs excellently for customer data analysis, financial data analysis, and business intelligence.

KNIME has some useful features like quick deployment and scaling efficiency. Users get familiar with KNIME in quite lesser time and it has made predictive analysis accessible to even naive users. KNIME utilizes the assembly of nodes to pre-process the data for analytics and visualization.

We could use KNIME as an alternative to WEKA for the development of our big data analytics engine. Although KNIME offers an easy to use interface for running analytics with very useful reporting and visualization capabilities, it is not to be used for big data analytics.

2.7.5. Apache Mahout

Apache Mahout [54] is a project developed by Apache Foundation that serves the primary purpose of creating machine learning algorithms. It focuses mainly on data clustering, classification, and collaborative filtering.

Mahout is written in JAVA and includes JAVA libraries to perform mathematical operations like linear algebra and statistics. Mahout is growing continuously as the algorithms implemented inside Apache Mahout are continuously growing. The algorithms of Mahout have implemented a level above Hadoop through mapping/reducing templates.

To key up, Mahout has following major features

- Extensible programming environment
- Pre-made algorithms
- Math experimentation environment
- GPU computes for performance improvement.

We could use Apache Mahout as an alternative to WEKA for the development of our big data analytics engine. Although Apache mahout is fit for our purpose, as it is built to work with big data and it is written in JAVA, it does not have any user interface to run data analytics with.

2.7.6. Spark MLib

MLlib [55] is Apache Spark's [56] scalable machine learning library. MLib is incorporated into Spark and supports a JAVA API. It is big data enabled as it is compatible with any Hadoop data source (e.g. HBase[10]). "MLlib provides efficient functionality for a wide range of learning settings and includes several underlying statistical, optimization, and linear algebra primitives" [55].

We could use Spark MLib as an alternative to WEKA for the development of our big data analytics engine. Although Spark MLib is fit for our purpose, as it is built to work with big data and it includes a JAVA API, it does not have any user interface to run data analytics with.

2.7.7. Summary and comparison of data analytics technologies

Below in Table 2-2 we summarize and compare the features of the data analytics technologies reviewed in Section 2.7. We use the following criteria for the comparison of the data analytics technologies: whether they incorporate a graphical user interface (GUI), whether they include a JAVA API, whether they are interoperable with big data analytics technologies (Big Data) and whether they enable the user to specify workflows by programmatically (scripting) or with a use of a GUI (Graphics).

Table 2-2 Data Analytics Technologies

Data analytics technology name	GUI	JAVA API	Big Data	Workflows
R	X	✓	X	Scripting
RapidMiner	✓	✓	X	Graphics
Orange	✓	X	X	Graphics
KNIME	✓	✓	X	Graphics
Apache Mahout	X	✓	✓	Scripting
MLib	X	✓	✓	Scripting
WEKA	✓	✓	✓	Graphics

We chose to build our big data analytics engine with WEKA[57], as we had experience with the use of it and with the use of DistributedWekaSpark [58] library we managed to run big data analytics with WEKA. We give more details about WEKA in section 4.3.3.

In Table 2-3 we present the features of the data analytics technologies explored, regarding the workflow specification. In this table we also include our platform

.

Table 2-3 Data Analytics Technologies - Workflows Features

Data analytics technology name	Data Pre-processing	Data Cleaning	Scheduled Executions	Reoccurring Executions	Data Change Driven Executions
R	✓	✓	X	X	X
RapidMiner	✓	✓	X	X	X
Orange	✓	✓	X	X	X
KNIME	✓	✓	X	X	X
Apache Mahout	✓	✓	X	X	X
MLib	✓	✓	X	X	X
Weka	✓	✓	X	X	X

We see that all the explored tools have data pre-processing and data cleaning capabilities, but none of them support scheduled, reoccurring and data change driven executions.

2.8. Big Data Platforms

2.8.1. Stratosphere

Stratosphere [59] is an open-source software stack that can be used for big data analytics. It has a unique set of features (i.e. "in situ" data processing, a declarative query language, treatment of user-defined functions as first-class citizens, automatic program parallelization and optimization, support for iterative programs, and a scalable and efficient execution engine) that efficiently enables the large scale programming of analytical applications. Additionally, Stratosphere offers tools for addressing some "Big Data" use cases, such as data warehousing, information extraction and integration, data cleansing, graph analysis, and statistical analysis applications.

Stratosphere contributed to a platform that became an Apache project in 2014 under the name Apache Flink [60]. Apache Flink is a platform that implements a universal dataflow engine designed to perform both stream and batch analytics.

We could use Stratosphere or Apache Flink when building our big data analytics engine, but it was very complicated to connect to the EVOTION data repository and build a prototype using it.

2.8.2. ASTERIX Big Data Management System (BDMS)

"ASTERIX [61] is a scalable platform that can be used to access, ingest, store, index, query, analyse, and publish very large quantities of semi-structured, complex and flexible data. It was designed as an improvement to the open source Hadoop stack.

ASTERIX became an Apache project under the name Apache AsterixDB™ [62].

We could use ASTERIX as a data management system, but we did not build a data repository from scratch. For the purposes of our prototype we used the EVOTION data repository.

Chapter 3

Specification of PHPDM Models

3.1. Overview of the Approach

The overall information flow, big data analytics, policy making platform is shown in the following figure.

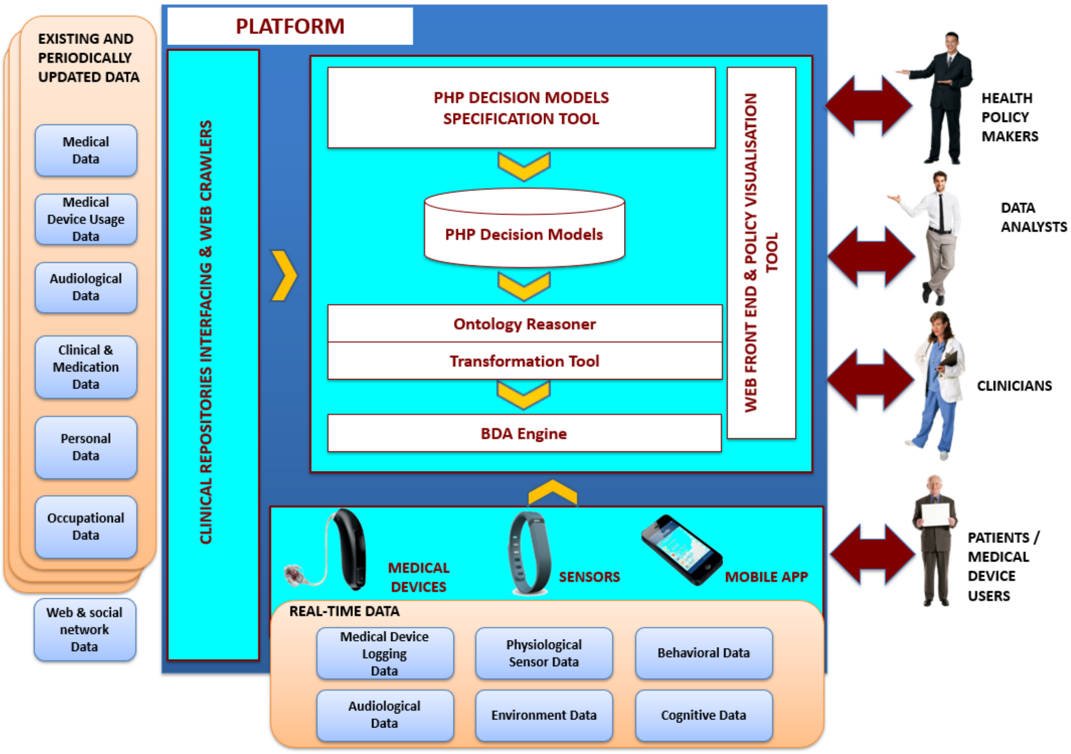


Figure 4 Platform information flows and decision marking

As shown in the figure, our platform uses different types of data to inform its decision making process, including: (a) existing and periodically collected patient data (i.e., audiological, medical, clinical and medication, personal and occupational data), (b) self-auditory and cognitive test data, (c) real time patient data including medical devices usage, audiological, cognitive, behavioural and life style, and environment data (e.g., location of patient, noise environment), and (d) dynamic web and social media data.

The operation of the platform is driven by public health policy decision-making models (**PHPDM models**). These models specify:

(i) the generic goal(s) underpinning the decision to be made (e.g. policies regarding the frequency of follow up care for patients) and the alternative decisions that may be made for this goal (e.g., having no, one or two follow ups within a specific time period)

(ii) the criteria to be used for making such decisions (e.g., whether the difficulties faced by different types of medical device users depend on their condition, their cognitive capabilities, their life style and behaviour, other comorbidities that they may have and/or their overall compliance with medical device usage guidelines given to them by clinicians and whether such difficulties are alleviated depending on the number of follow up treatments and the time that elapses between them)

(iii) the BDA evidence required for applying the criteria (e.g., whether any combination of the factors considered above is a good predictor of the difficulties faced by medical device users as confirmed by specific types of statistical analysis or data mining-based classification) and the BDA process for producing it

(iv) processes to be followed for making specific types of health policies (e.g., what is the threshold of evidence that should be considered sufficient for a particular decision, who are the stakeholders whose views should be considered and recorded prior to reaching a decision, who has responsibility for making the final decision, whether a decision should be continually or periodically reviewed upon the acquisition of new evidence etc.)

Our view is that PHPDM models of this form are essential for realizing BDA evidence-based, scalable, fully dynamic, repeatable and accountable policymaking. This is because PHPDM models covering the aspects identified above could

- be automatically transformed into executable BDA processes and simulation processes whose execution would provide the basic evidence required for making a decision and exploring its consequences
- drive the collaborative stakeholder decision making processes
- provide a structure for organizing the alternatives, arguments and rationale for making decisions in a way that makes them traceable and accountable.

Furthermore, PHPDM models can be (a) repeatedly executed in the same or different policy making settings (e.g., for making policy on the very same issues in different regions) and (b) specified parametrically to make their customization easy in case that this would be required in different policy making settings.

The platform provides a tool supporting the specification of PHPDM models into some high-level language, and their verification and transformation into executable BDA and simulation tasks and decision-making processes that would be passed as inputs to the **BDA platform** and **simulator** to execute them and realize the policy making process specified by them.

As shown in Figure 4, to enable the data collection processes and policy-making process, the platform could also incorporate and integrate:

(a) **Existing repositories of medical data**.

(b) **Enhanced medical devices** enabling the capture and provision of **medical device usage related data** (e.g., rating of device ease or difficulty of use in different listening conditions, frequency and type adjustments of controls).

(c) **Sensors** supporting the collection **real time contextual patient physiological data** (e.g., heart rate, blood pressure, skin conductance)

(d) A third-party **mobile application** with components supporting the acquisition and transmission of **behavioural** (e.g., recording of patient daily activities such as participation in conversations, watching TV)**, contextual** (e.g., patient's location)**, cognitive** (e.g., verbal reaction time) data as well as the notification and acceptance/rejection of decisions by the patient and/or their carers (**decision selection component**); and the execution n of periodic audiological and cognitive to collect the related data (**audiological** and **cognitive test components**). The mobile application is not considered as part of the prototype of this thesis, so an existing application was used, available as part of EVOTION.

3.2. Overview of PHPDM Language

In this document, OWL code is presented using `Consolas` font style.

`Consolas` font style is also used to refer to classes of the language.

3.2.1. Purpose of the Language

The PHPDM language that we introduce in this thesis provides the means for specifying public health decision making models, aka "PHPDM models" in the context of our platform prototype.

The definition of this language has been performed as an iterative process: an initial version of the language was presented to the consortium of EVOTION. After collecting their feedback and producing more intermediate versions of the language, we concluded to the current version, which satisfies all stakeholders' requirements.

PHPDM models do not describe public health policies as such. They describe the process through which decisions regarding the formation of public health policies may be made and the evidence that needs to be taken into account in coming up with such decisions.

To provide a comprehensive specification of PHPDM models, the PHPDM language provides constructs for modeling the following facets of public health policy decision-making process:

(a) the overall goal and the specific objectives that public policy needs to address in a given area of health intervention;
(b) the range of possible actions (interventions) through which the goals and objectives of the policy can be achieved
(c) the evidence that needs to be gathered and analysed in order to make informed and plausible decisions about the actions (interventions) that need to be undertaken (made) as part of the policy;
(d) the processes for analysing and establishing the validity of this evidence;
(e) the stakeholders who will consider the evidence and decide which actions (interventions) should be undertaken (made);
(f) the criteria that should be used to make decisions on the basis of the identified evidence.

The definition of the PHPDM language and the individual PHPDM models specified in it is ontology-based. More specifically, the PHPDM language is defined as a set of classes in the ontology modeling language OWL and PHPDM models are specified as interrelated instances of these classes. This is because OWL provides a modeling framework with clear semantic foundations, providing a solid basis for processing (i.e., querying, drawing inferences, interpreting and executing) the models defined in it.

In the rest of this section, we provide an overview of the top-level OWL classes and structure of the PHPDM language, discuss the different types of users that we envisage for the language, outline the semantic foundations of OWL that apply and give semantics to it, and introduce the syntax that we have used for the language.

3.2.2. Main Modeling Constructs of the Language

The PHPDM language is defined as an OWL ontology, i.e., a set of OWL classes and relationships between them. PHPDM models are defined as instances of this ontology, i.e., by objects which instantiate the classes of the language and are related by instances of the relationships defined in the language.

Conceptually, the PHPDM language can be broken down into 4 modules:

- The *Policy Module*: This module includes classes that specify the overall goal and objectives that a health policy that needs to be formed should address, and the actions (interventions) that will be needed to realise the policy.
- *The Policy Making Module:* This module includes the classes that specify the stakeholders who participate in the decision-making process and the positions that they may express.

- The *Data Analytics and Evidence Module*: This module includes classes that specify the data that will need to be analysed to produce evidence aiding the making of policy decisions, the forms of analysis that should be applied to these data, and the criteria that should be used to assess whether the evidence generated from the data is sufficient in supporting actions.
- The *Policy Implementation and Monitoring and Evaluation Processes Module*: This module includes classes related to the implementation, monitoring and evaluation of a policy that has been decided, and specify how the effects of policy arising from its implementation and monitoring can affect the revision of decision regarding it. This module is not implemented in the PHPDM language presented in this document and is part of our future work.

Figure 5 shows the top-level classes and relationships of the ontology that constitutes the PHPDM language. The figure shows these classes and their relationships as a UML [63] class diagram. It should be noted that the use of UML to present the ontology that defines the PHPDM language has been adopted merely to enable the visual presentation of the language and does not constitute part of the definition of the language.

The class `PolicyModel` in Figure 5 is the class that can be used to specify PHPDM models. As shown in the figure, each `PolicyModel` has a general `Goal`, i.e., a possibly non-measurable target that it aims to address. The goal of a policy is expressed at a generic level and its achievement requires addressing concrete `Objectives`. `Objectives` are measurable policy targets that can be addressed by `PolicyActions`.

A policy action presents a possible way of addressing one of the objectives of the policy. A policy action can act as pre-requisite, as dependency or as dependant to other policy actions. Policy actions may need to be applied as alternative (i.e., mutually exclusive) or complementary means for realising the objectives of a policy. The possible ways of applying actions is specified by the policy model. More specifically, in cases where actions need to be applied as alternatives, the model must describe them as such.

Policy actions reflect the key decisions that may be made in PHPDM process. These decisions need to be explored on the basis of evidence arising from the analysis of data. To express this, in the PHPDM language ontology each policy action is associated with a `Criterion` that determines the circumstances under which the evidence arising from data analytics would support the action. A criterion is specified by a `LogicalExpression` over the outcomes of a `DataAnalyticsWorkflow`.

With regards to the type of processing that they perform upon their input data set(s), data analytics tasks can be distinguished into `StatisticalAnalysisTasks` (i.e., tasks that carry out some statistical analysis upon the data), `DataMiningTasks` (i.e., tasks that carry out some data mining analysis upon the data), SocialMediaAnalyticsTasks (i.e. tasks that carry out some analysis of social media data), SimulationTasks, (i.e. tasks that carry out analysis of simulating-synthetic data), TextMiningTasks (i.e. tasks that perform text mining techniques for analysis of the literature) and (i.e., tasks that perform some pre-processing over the data that is required prior to the analysis such as data cleaning or data joining over correlating factors).

Each data analytics task utilizes a Method, which can be an Algorithm or an Operation. Each algorithm comes with an OutputDataSpecification (i.e. the form of the output data). All types of tasks utilize algorithms, except data processing tasks which utilize operations. A Data processing task can utilize a DataCleaningOperation (i.e. detecting and correcting, or removing corrupt or inaccurate data), a SamplingOperation (i.e. selecting a subset of the data), a FilterOperation (i.e. filtering data), a ProjectOperation (i.e. performing data projection) and JoinOperation (i.e. performing data joins). Statistical analysis tasks and data mining tasks are modeled using special purpose classes of the PHPDM language ontology, which are described in Section 3.3.7 and Section 3.3.8, respectively.

A PHPDM model should also specify the Stakeholders of the policy making process, i.e., the human actors who may participate in it. These participants of the process may express Positions over the different action options that are available in the process. A position expressed by a stakeholder can be a SupportivePosition (i.e., a position that supports the advocation of the action), an OpposingPosition (i.e., a position that is negative to the advocation of the action) or a NeutralPosition (i.e., a decision indicating that the stakeholder neither supports nor objects to the action). A stakeholder may express Supportive, Opposing or Neutral positions for one or more alternative actions but cannot express two different Positions for the same alternative.

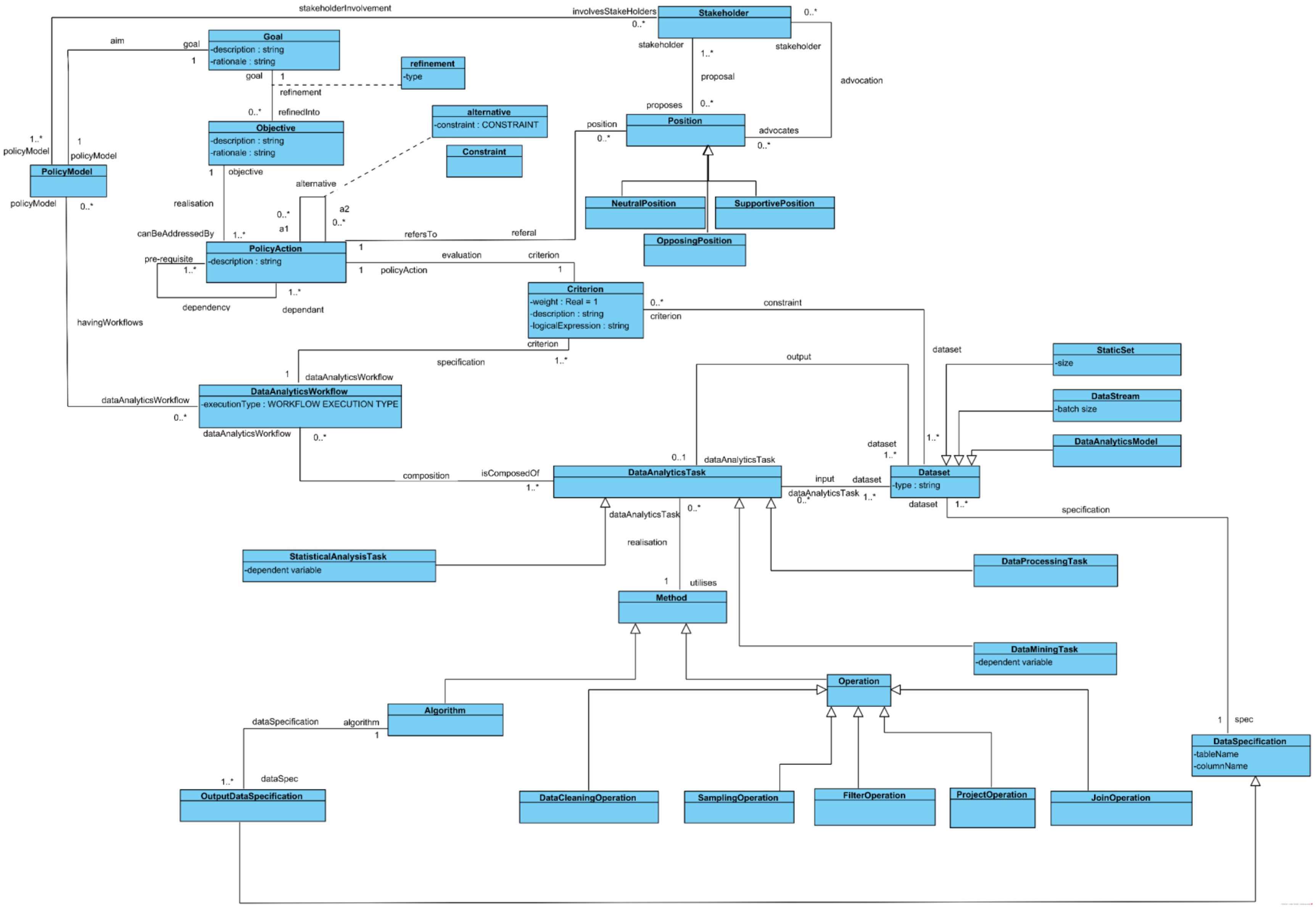


Figure 5 Main Language Classes

3.2.3. Language Users

The PHPDM language is expected to be used by actors in the public health policy modeling ecosystem who have different types of expertise. In particular, it may be used: (i) public health policy makers, (ii) clinicians, and (iii) data scientists.

Public health policy makers are expected to make use of the part of the language that enables the specification of policy goals and objectives (Policy Module), the stakeholders who may be involved in the decision-making process, and the potential policy actions (Policy Making Module). More specifically, the main focus of policy makers is the goal, objectives, policy actions and criteria of the policy model. Criteria are specified using existing data analytics workflows or new ones specified with the help of data analysts.

Clinicians are expected to make use of the part of the language that enables the specification of policy objectives, potential policy actions, and the evidence and criteria required for making decisions regarding the actions (Policy Module). They may also be involved in the identification of the data sets and the analytic processes that need to be analysed for generating the evidence (Data Analytics and Evidence Module). More specifically, the main focus of clinicians is the data entry and the assurance of the validity of the data and the criteria of the policy model. Criteria are specified using existing data analytics workflows or new ones specified with the help of data analysts. Last but not least, clinicians provide their feedback regarding the important features of the available datasets.

Data scientists are expected to make use of the part of the language that enables the specification of the data sets and the analytic processes that need to be analysed for generating the evidence (Data Analytics and Evidence Module), and the criteria for establishing the plausibility of the generated evidence in support of different actions (Policy Module). More specifically, the main focus of data analysts is the creation of the data analytics workflows to support the criteria of the policy model. Data analytics workflows are specified as part of a policy model, to support the criteria of the policy model, or independently, to run an analysis to the data. A data analyst defines data analytics tasks, with their input and output datasets. Then, they create data analytics workflows by choosing a set of previously created tasks. A data analytics workflow can be executed upon user request, periodically (for example, once every year) or when there is a change to the data (for example, when the volume of data received from mobiles changes by 40%). The created tasks are reusable for multiple workflows and the workflows are reusable for multiple policy models.

As an example of the above consider the PHPDM model described in Chapter 5 . This model has been created to aid the definition of policy whose overall goal is to reduce barriers that prevent HA usage. To do so an exploratory process can be initiated to explore whether factors such as the occupation, the education level or the age of a HA user affect the usage of HAs. If following data analysis, such factors are found to have a significant effect on HA usage then related interventions (actions) may be initiated. Such interventions would be to generate appropriate campaigns for specific age group, amend protocols for hearing aid fitting to address more effectively the drop of HA usage in older age groups etc.

In forming a PHPDM model to aid decision making in this area, the overall goal of the model would be set by policy making authorities (e.g., representatives of ministries of health for a policy with national scope). The objectives of the PHPDM model, i.e., to introduce interventions to address prevention of HA usage due to occupation, due to education level, or due to age, could be set following a dialogue between clinicians and policy makers that identifies the particular factors as worth exploring further before making decisions on the relevant interventions.

The specific data analytic procedures that will be used to explore such factors would typically be identified through a dialogue between clinicians and data scientists, through which the most appropriate analytic methods are established following consideration of the types of the data involved (e.g. numeric vs. nominal data) and the conditions that should be satisfied by the available data set in order for an analytics technique to be expected to produce meaningful results. Various forms of statistical analysis, such as regression, could for example be deemed as non-appropriate if the independent variables are themselves linearly independent (i.e., it is not possible to predict any of them through a linear combination of the other). The clinical scientists involved in this process could be professional bodies (e.g., associations of clinical audiologists, ENT doctors). Data scientists may be members of the same bodies which provide the policy authority representatives and the representative of the clinicians or drawn from other organisations with established expertise on the subject (e.g., Universities or specialist research institutes).

It should be noted that whilst the design of the language supports the definition of different stakeholders, it does not provide methodological guidance on how goals, objectives, actions, stakeholders and data analytic workflows should be specified.

3.2.4. Semantic Foundations

The advocation of OWL as the modeling framework for the definition of the PHPDM language and the specification of PHPDM models has been due to its ability to provide a formal foundation for defining the PHPDM language. This is because OWL has a well-defined model theoretic semantics that have been defined as an extension of the semantics of the description logic SROIQ [64].

In a model theoretic semantics, a language like the PHPDM language is defined as a *vocabulary* of lexical terms, which denote classes, individuals (i.e., objects in a given domain), (primitive) data types, literals (i.e., values of different data types such as integers, strings), object properties, data properties, and data type constraining facets. Then, a model theoretic semantics of the language is defined as an *interpretation* (i.e., a mapping) from the vocabulary of the language onto subsets of a given *object domain* Δ_I (i.e., a set of real objects), subsets of a given *data domain* Δ_D and subsets of the cartesian products Δ_I x Δ_I and Δ_I x Δ_D (i.e., unary relations over Δ_I and binary relations over Δ_I x Δ_D. The interpretation of a class C, for example, is a mapping of C onto a subset $(C)_C$ of Δ_I (i.e., $(C)_C \subseteq \Delta_I$), which includes the individual objects in the object domain Δ_I which are instances of the class. Similarly, the interpretation of an object property OP is a mapping of P onto a set $(OP)_{OP} \subseteq \Delta_I$ x Δ_I. $(OP)_{OP}$ is essentially the set of domain object pairs (O_i, O_j) where the object O_j is the value of the property OP of object O_i.

Beyond such mappings, the model theoretic semantics of OWL include the definition of interpretations of OWL expressions and OWL axioms.

Expressions in OWL are used to express "complex notions" in the definitions of data ranges, object properties and classes. OWL expressions can be of three different types, namely: object property expressions (OPE), class expressions (CE), data range expressions (DR). Examples of such expressions and their interpretations are given in the Table 3-1 below. As an example of using expressions in the OWL definition of a language (ontology), consider the class expression `ObjectIntersectionOf(o:Clinician, o:UniversityProfessor)`. This expression may be used to define a class of stakeholders in the decision-making process of a PHPDM model that includes clinicians who are also university professors.

The interpretations of OWL expressions as those listed in Table 3-1 enable the automatic computation of the expressions by reasoning tools and are important for generating "complete" definitions of different types of ontology elements (e.g., classes, properties, data ranges) before attempting any reasoning upon them.

Table 3-1: Examples of OWL object property, data range and class expressions and their semantics

Expression	Type of Expression	Interpretation of Expression
ObjectInverseOf(OP)	Object Property Expression(OPE)	$\{ (x, y) \mid (y, x) \in (OPE)_{OP} \}$
DataIntersectioOf(DR_1 ... DR_n)	Data Range	$(DR_1)_{DT} \cap \ldots \cap (DR_n)_{DT}$

Expression	Type of Expression	Interpretation of Expression
	Expression (DR)	
ObjectIntersectionOf(CE_1 ... CE_n)	Class Expression (CE)	$(CE_1)_C \cap \ldots \cap (CE_n)_C$
ObjectSomeValuesFrom(OPE CE)	Class Expression (CE)	$\{ x \mid \exists y : (x, y) \in (OPE)_{OP}$ and $y \in (CE)_C \}$
ObjectAllValuesFrom(OPE CE)	Class Expression (CE)	$\{ x \mid \forall y : (x, y) \in (OPE)_{OP}$ implies $y \in (CE)_C \}$

Axioms in OWL are statements of an ontology that are asserted to be true in the domain that is being described by it. OWL provides an extensive set of predefined (i.e., "built-in") types of axioms that may be used in the specification of an ontology. These include *Declaration axioms, Class axioms, ObjectProperty axioms, DataProperty axioms, DatatypeDefinition axioms, HasKey axioms, Assertion axioms* and *Annotation axioms*. Axioms define additional conditions over the particular type of ontology elements that they refer to, which must be satisfied by these elements, and therefore restrict the range of possible meanings of the ontology, i.e., the interpretations that are valid models of the ontology. Examples of axioms and their meaning, specified as the conditions that they impose upon ontology interpretations for them to be satisfied, are listed in Table 3-2 below.

Table 3-2: Examples of OWL axioms and the conditions they impose upon OWL models

Axiom	Type of Axiom	Meaning (Conditions)
SymmetricObjectProperty(OPE)	Object Property Axiom	$\forall x, y : (x, y) \in (OPE)_{OP}$ implies $(y, x) \in (OPE)_{OP}$
TransitiveObjectProperty(OPE)	Object Property Axiom	$\forall x, y, z : (x, y) \in (OPE)_{OP}$ and $(y, z) \in (OPE)_{OP}$ imply $(x, z) \in (OPE)_{OP}$
DisjointObjectProperties(OPE_1 ... OPE_n)	Object Property Axiom	$(OPE_j)_{OP} \cap (OPE_k)_{OP} = \varnothing$ for each $1 \le j \le n$ and each $1 \le k \le n$ such that $j \ne k$
SubClassOf(CE_1 CE_2)	Class Axiom	$(CE_1)_C \subseteq (CE_2)_C$
EquivalentClasses(CE_1 ... CE_n)	Class Axiom	$(CE_j)_C = (CE_k)_C$ for each $1 \le j \le n$ and each $1 \le k \le n$
DisjointClasses(CE_1 ... CE_n)	Class Axiom	$(CE_j)_C \cap (CE_k)_C = \varnothing$ for each $1 \le j \le n$ and each $1 \le k \le n$ such that $j \ne k$

As an example of using axioms in the OWL definition of a language (ontology), consider the object property axiom `SymmetricObjectProperty(a:Alternative)`. This axiom may be used for policy actions in the PHPDM language to express that if the object property Alternative of a policy action is symmetric, i.e., if a policy action a_1 is an alternative to a policy action a_2 then a_2 is also an alternative to a_1.

A model for an OWL ontology O is an interpretation that satisfies the axioms of O.

The provision of a model theoretic semantics for OWL provides a non-ambiguous basis for reasoning over OWL ontologies. Some typical inference problems that often arise for OWL ontologies have been defined in [65].

These inference problems are defined in reference to a datatype map D and a vocabulary V over it and include the following inference questions:

- *Ontology Consistency:* O is consistent (or satisfiable) w.r.t. D if a model of O w.r.t. D and V exists.
- *Ontology Entailment:* O entails O_1 w.r.t. D if every model of O w.r.t. D and V is also a model of O_1 w.r.t. D and V.
- *Ontology Equivalence:* O and O_1 are equivalent w.r.t. D if O entails O_1 w.r.t. D and O_1 entails O w.r.t. D.
- *Ontology Equisatisfiability:* O and O_1 are equisatisfiable w.r.t. D if O is satisfiable w.r.t. D if and only if O_1 is satisfiable w.r.t D.
- *Class Expression Satisfiability:* CE is satisfiable w.r.t. O and D if a model I of O w.r.t. D and V exists such that $(CE)_C \neq \varnothing$.
- *Class Expression Subsumption:* CE_1 is subsumed by a class expression CE_2 w.r.t. O and D if $(CE_1)_C \subseteq (CE_2)_C$ for each model I of O w.r.t. D and V.
- *Instance Checking:* a is an instance of CE w.r.t. O and D if $(a)_I \in (CE)_C$ for each model I of O w.r.t. D and V.
- *Boolean Conjunctive Query Answering*: Q is an answer w.r.t. O and D if Q is true in each model of O w.r.t. D and V according to the standard definitions of first-order logic.
- The full description of the model theoretic semantics and the meaning of the full set of standardised object property, data range and class expressions and axioms of OWL is given in [65].

3.2.5. Language Syntax

The PHPDM language has been defined as an ontology expressed in the OWL Web Ontology Language. The syntax that we have used for this purpose is the OWL2/XML syntax [66] Thus, the definition of the PHPDM language is available as an XSD schema, i.e., a document defining the structure of XML documents that express PHPDM models. This XSD document is provided in Appendix B:

OWL XML Definitions of the Language.

This syntax has been defined as a dialect of the OWL abstract syntax and can be automatically transformed into it and/or generated from it. Similarly, the syntax that we use in this document can be automatically generated from and transformed into OWL 2 specifications in other syntaxes defined for OWL 2, namely the RDF, the Manchester syntax. Figure 6 summarises the available syntaxes of OWL 2 and their relationships.

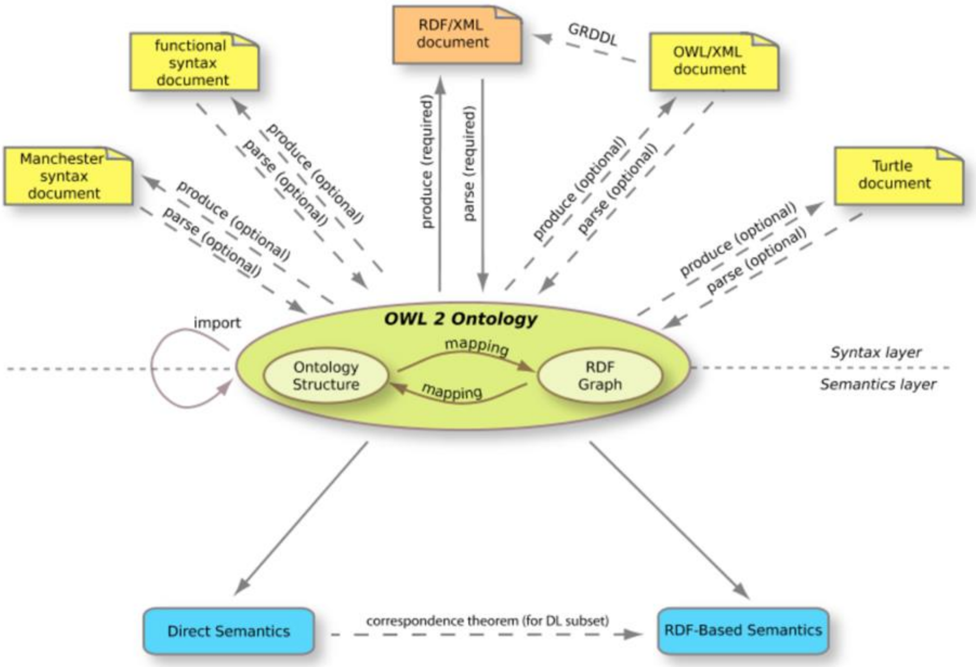


Figure 6 Alternative OWL 2 syntaxes and the mappings between them (taken from (W3C, 2012)[1])

The choice of the XML syntax of OWL for the definition of the PHPDM language and models has been motivated by the ability to process specifications of PHPDM models expressed in this syntax with different tools and exchange them between these tools. Although, the same purpose would also have also been served if we had chosen the RDF syntax of OWL, the fact that in several cases the RDF syntax advocates an asymmetric and more verbose approach than the XML syntax in declaring certain elements of an ontology has made the latter our preferred choice since we wanted to lessen the burden on PHPDM modellers in specifying PHPDM models.

1 https://www.w3.org/2012/pdf/REC-owl2-overview-20121211.pdf

To appreciate the above consider the following two specifications of model elements in an OWL ontology

Case 1 (disjoint classes): RDF/XML Syntax

```
<owl:AllDisjointClasses>
  <owl:members rdf:parseType="Collection">
    <owl:Class rdf:about="Data Mining Task"/>
    <owl:Class rdf:about="Statistical Analysis Task"/>
  </owl:members>
</owl:AllDisjointClasses>
```

Case 1 (disjoint classes): OWL/XML Syntax

```
<DisjointClasses>
    <Class IRI="Woman"/>
    <Class IRI="Man"/>
</DisjointClasses>
```

Case 2 (sub property expression): RDF/XML Syntax

```
<owl:ObjectProperty rdf:about="utilisesStatisticalAlgorithm">
    <rdfs:subPropertyOf rdf:resource="utilises"/>
  </owl:ObjectProperty>
```

Case 2 (sub property expression): OWL/XML Syntax

```
<SubObjectPropertyOf>
   <ObjectProperty IRI="utilises"/>
   <ObjectProperty IRI="utilisesStatisticalAlgorithm"/>
 </SubObjectPropertyOf>
```

The first of these cases (Case 1) provides a declaration of disjoint classes in the PHPDM language, asserting that the classes `Statistical Analysis Task` and `Data Mining Task` cannot have instances in common with the class or in other words that a statistical analysis task cannot be a data mining task and vice versa. The declaration of this restriction in RDF requires the declaration of the two classes as members of a third element (i.e., `Collection`), as well as references to the `rdf:` name space.

The second case (Case 2) asserts that the object property `utilisesStatisticalAlgorithm` between `Statistical Analysis Task` and `Statistical Algorithm` is a sub-property, i.e., a subset of the relation formed by the property `utilises` between `Data Analytics Task` and `Method`. The specification in RDF presents some asymmetry as the super property is identified as resource, which requires the specifier to recall this very concept in creating the declaration.

3.3. Language Modules

3.3.1. Policy Module: Policy Aims, Objectives and Actions

3.3.1.1. Purpose of Module & Overview of Module Concepts

The purpose of this module, presented in Sect. 3.2.2, is to enable the policy makers to define the policy model's general goal and its measurable objectives, as well as the policy actions to be taken for realising each of the defined objectives. A policy action can act as pre-requisite of other policy actions. Policy actions can be alternative to each other, i.e., their undertaking must be mutually exclusive. Alternative actions may arise due to constraints in policy realisation (e.g. resource constraints).

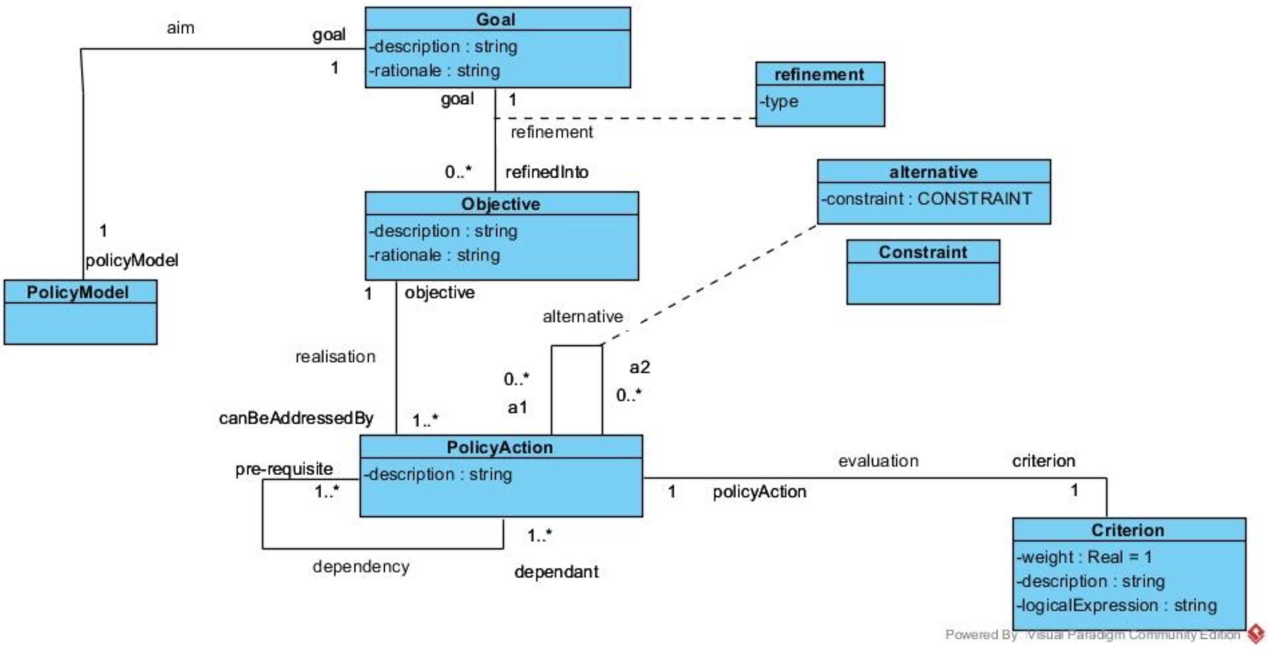


Figure 7 Policy Aims, Objectives and Actions

3.3.1.2. Policy Model

Purpose

The class `PolicyModel` is introduced to express the set of potential policy models that public health policy decision making may arrive at in order to establish public health policies in a particular area and with specific interventions in mind. As said earlier, a policy model is also referred to as a "Public Health Policy Decision Making Model" or shortly "PHPDM model". The binding element of such potential policy models is the overall (public health policy) goal that their models are aimed at achieving.

Concept Definition

A policy model defines the overall goal that a public health policy should address, the concrete objectives through which this goal may be achieved, and the possible actions through which these objectives may be realised. Figure 8 shows a graphical view of policy models and how the other key elements involved in their definition are related to each other.

Each policy model must be associated with only one overall policy goal and there can be no policy model that has more than one goals. Also, as the class `PolicyModel` is meant to represent all the potential policy models that address a particular public health policy goal, a policy goal can only be associated with one policy model. These restrictions are reflected in the cardinalities of the association `aimedAt` between the class `PolicyModel` and the class `Goal` in Figure 8 as well as the OWL specification of the class `PolicyModel` listed below. The objectives and the actions involved in a policy model are indirectly related to the model. In particular, the objectives are related to it through their relation with the overall goal of the policy model, and the actions are related to the model through their relations with the objectives.

Beyond the goal of the potential public health policies which may be formed using it, a policy model defines the stakeholders that may be involved in the decision-making process that will be followed in order to establish public health policies, and the data analytics workflows that will be used in order to derive the evidence that need to be taken into account during the policy formation process. In this sense, a policy model provides the reference context for associating policy goals, stakeholders, data analytics, and decision-making processes with each other.

The relationship between `PolicyModel` and `Stakeholder` (see the association `involvesStakeholders` in the UML diagram of Figure 8 is a many-to-many relationship. This means that a policy model may involve more than one stakeholders who will participate in the decision-making process about the policy to be selected/applied for a given policy goal. It also means that the same stakeholder may be involved in decision making regarding public health policies aimed at addressing different goals and in different areas of intervention. The PHPDM language also requires stakeholders to be specified only if there is a policy model, which they will be associated with, as the lower bound of the multiplicity of the association `involvesStakeholders` has its policy model end.

A policy model may also be associated with zero or more data analytic workflows as indicated by the cardinalities of the association `hasWorkflows` in Figure 8. This allows for using different workflows, if necessary, to analyse data and produce different types of evidence for a given policy model. It also allows to specify policy models with no data analytic workflows. This may be needed as at the time when a policy model is specified, it might not be clear what workflows would be appropriate to use for analysing data to support related decisions. It may also be needed as there could be policy decision-making processes, which cannot involve data analytic workflows as, for example, when there are no available data. Hence, to cover different modeling scenarios and needs, the PHPDM language allows policy models with no associated data workflows. It should be noted, however, that PHPDM models with no analytic workflows won't be executable and no decision-making support could be offered based on them.

Formally, policy models are defined as instances of the OWL class `PolicyModel`. The definition of this class is listed in Appendix B (B.1.1.1) along with the class axioms (B.1.2.1) and examples (B.1.3.1).

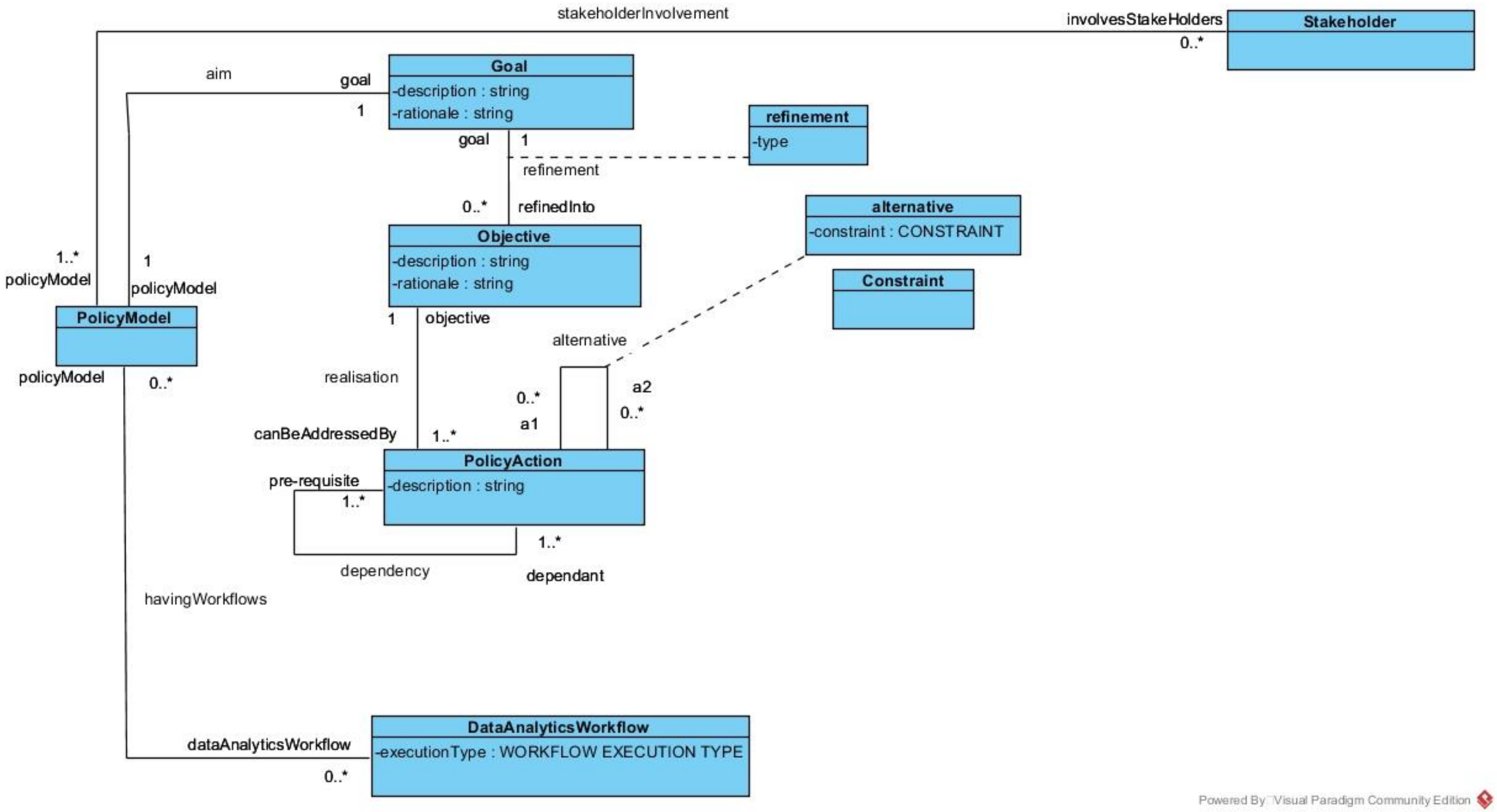


Figure 8 Policy Model and its key relations to other concepts

3.3.1.3. Goal

Purpose

The `Goal` class has been introduced to define the general non-measurable target of the policy. The goal of any policy model should be expressed in a way that expresses at a broad and inclusive level the overall aim of the healthcare interventions targeted by a policy model.

Concept Definition

Each policy model has a single goal expressing at a broad and inclusive level the overall aim of the healthcare interventions targeted by model. This aim should be informally defined in natural language using the `description` attribute of the class `Goal`. Figure 9 shows a graphical view of policy goal and its key relations to other concepts.

In addition to the description of the overall aim of a policy model, the specification of the policy goal should also include a rationale, explaining the reason why the particular goal is important and needs to be tackled by a public health policy (like the one that will be derived from the model). The rationale underpinning a policy model's goal is also expressed in natural language, as the value of the attribute rationale of the class Goal.

As the specification of a policy model's goal is meant to be generic and inclusive of a target area of healthcare interventions, the goal will typically need to be realised through concrete and implementable policy interventions. The latter are expressed by objectives. The intended meaning and specification of policy model objectives are described below.

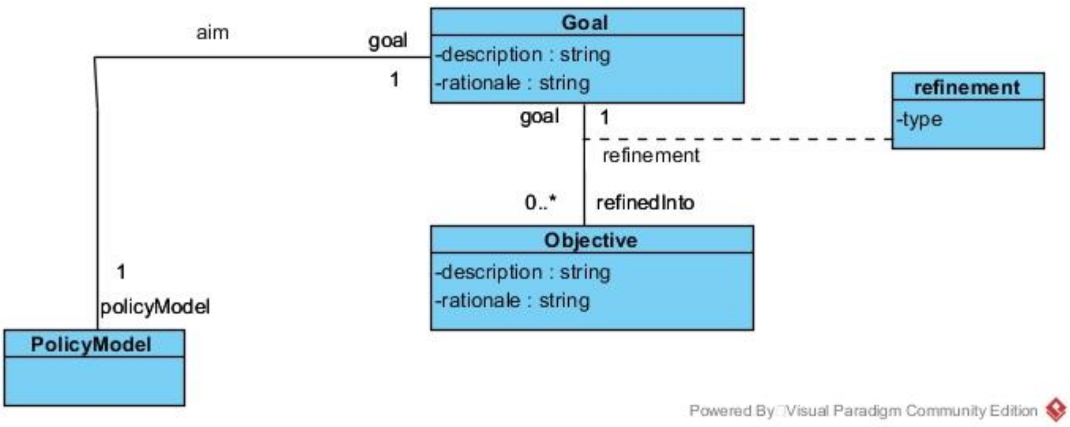


Figure 9 Policy goal and its key relations to other concepts

At the level of policy goal, however, it should be noted that the PHPDM language expects goals to be refined into the concrete objectives that will realise them. This is expressed in a policy model by the association `refinement` between the class `Goal` and the class `Objective`. The multiplicity of this association between goals and objectives is 1 at the side of goal and 0..* at the side of objectives, meaning that each goal must be refined into more than one objectives but also may have no objective associated with it. The latter case, i.e., allowing PHPDM models to have no objective associated with a policy goal for some period of time is necessary in order to accommodate scenarios in which, stakeholders might not have concrete views about possible and/or appropriate objectives for the model as yet. This is likely to arise when PHPDM models are still under development. The PHPDM language allows for this. However, it should be noted that PHPDM models for which no objectives have been specified, will be incomplete it will not be possible to enact them, i.e., to execute data analytics workflows for them or initiate decision-making processes for them.

In cases, where a goal is refined into more than one objectives, the intended meaning of the refinement is declared by the `type` attribute of the refinement association. This attribute can take two possible values: DISJUNCTIVE (aka OR) or CONJUCTIVE (aka AND) refinement. A refinement is disjunctive when achieving any of the refinement objectives would be sufficient for achieving the goal of the policy. A refinement is conjunctive if all of the refinement objectives need to be achieved in order to achieve the target of the policy. With the use of the refinement association, general logical expressions can be supported.

Formally, policy goals are defined as instances of the class `Goal` in OWL. The definition of this class is listed in Appendix B (B.1.1.2) along with the class axioms (B.1.2.2) and examples (B.1.3.2).

3.3.1.4. Objective

Purpose

The purpose of an objective in a policy is to express a concrete and measurable policy intervention, which needs to be addressed in order to realise a policy's goal, either fully or partially.

Concept Definition

Objectives in policy models are specified as instances of the class `Objective` in the PHPDM language. Figure 10 shows a graphical view of the class of policy objectives and its key relations to other classes of the PHPDM language.

Like goals, the specification of objectives includes two attributes: the attribute `description` and the attribute `justification`. These attributes serve the same purpose as in the case of goals, i.e., they enable the provision of a description in natural language of a concrete intervention needed to realise the target of a policy and an explanation of why this intervention is necessary and can achieve this purpose, respectively.

As discussed earlier, an objective must be associated with one (and up to one) policy goal. Also, each objective may be realised through policy actions, modelled as instances of the class `PolicyAction`. The potential realisation of an objective by policy actions is modelled by the association `realisation` between the class `Objective` and the class `PolicyAction`. This cardinality of this association is 1 at the end of `Objective` and 1..* at the side of `PolicyAction`. This means than for an objective to be realised at least one but possibly more than one actions will be required. The execution of all the actions associated with an objective via the `realisation` association will be required to realise the objective, unless some actions have been declared as alternative to each other. In this case, executing just one of the actions declared as alternatives will be required. So if a set of actions {a1, a2, a3, a4} have been associated with an objective o1 via the `realisation` association and the pairs of actions (a1, a2) and (a3, a4) have been declared as alternatives, the meaning of the model is that o1 can be realised through the execution of the following combinations of actions: (a1, a3), (a2, a3), (a1, a4), (a2, a4). Further constraints about the alternative relation between actions are given below in the description of the class `PolicyAction`.

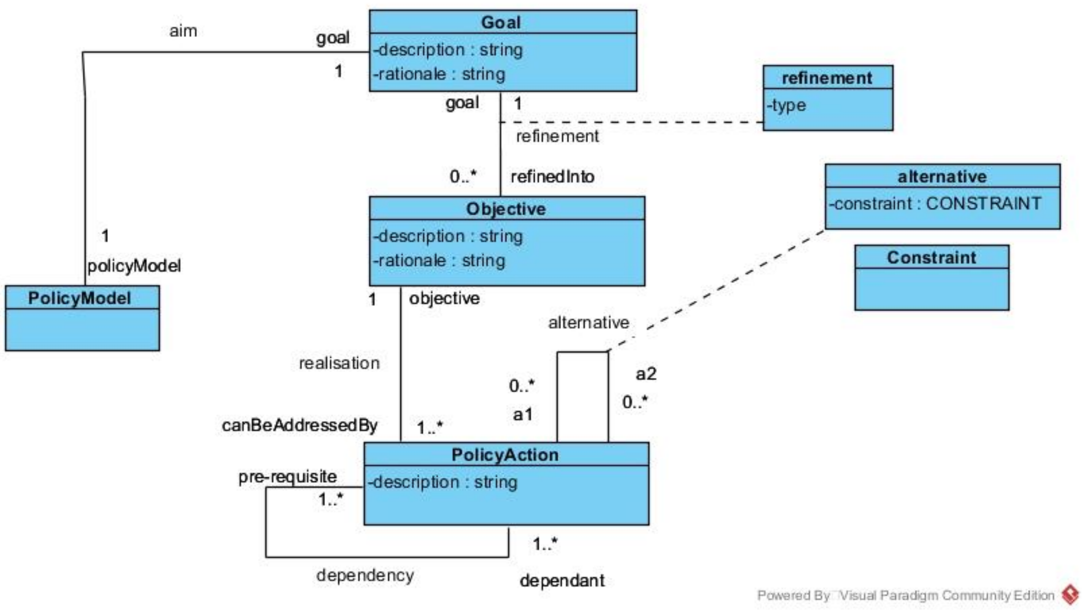


Figure 10 Policy Objective and its key relations to other concepts

Formally, policy objectives are defined as instances of the class `Objective` in OWL. The definition of this class is listed in Appendix B (B.1.1.3) along with the class axioms (B.1.2.3) and examples (B.1.3.3).

3.3.1.5. Policy Action

Purpose

The purpose of actions in a policy model is to define concrete and implementable primitive interventions that may be taken to realise an objective.

Concept Definition

Actions in policy models are specified as instances of the class `PolicyAction` in the PHPDM language. Figure 11 shows a graphical view of the class of policy actions and its key relations to other classes of the PHPDM language.

The class `PolicyAction` has one data attribute, called `description`. The purpose of this attribute is to provide a textual description of the policy action. Actions need always to be specified in the context of single policy objective. This is done by associating them with their objective through the association `realisation`, as discussed earlier. In general, more than one actions may be required in order to realise a single objective. Actions can be taken by health policy experts, health authorities, clinicians etc. An action always models something concrete that can be done to address a policy objective. In the context of an objective's realisation, actions can be mutually exclusive or complementary. Mutually exclusive actions are actions that cannot be executed together and only one of them may be executed. This relation between actions is expressed by the association `alternative`. To understand the use of this relation, suppose that four actions a1, a2, a3 and a4 have been declared as actions that realise an objective o1 and that a3 and a4 have been declared as alternatives to each other. Then o1 may be realised through the execution of the following combinations of actions: (a1, a2, a3) or (a1, a2, a4). In the context of the realisation of the same objective, if both actions a3 and a4 had been declared as alternatives to each other, and actions a1 and a2 had been declared as alternatives to each other, then the combinations that can realise the objective o1 would be: (a1, a3), (a1, a4), (a2, a3) or (a2, a4).

The use of the alternative association between actions is restricted by the constraint that if an action a1 is declared as an alternative to action a2 and a2 is declared as an alternative to a3 then a1 is also an alternative to a3. In other words, the relation `alternative` is a transitive relation. So, if in the previous example action a1 is declared as an alternative to action a2, and a2 is declared as an alternative to a3, then a1 is also an alternative to a3. The consequence of this is that objective o1 would be realisable through the following combinations of actions: (a1, a4), (a2, a4) or (a3, a4).

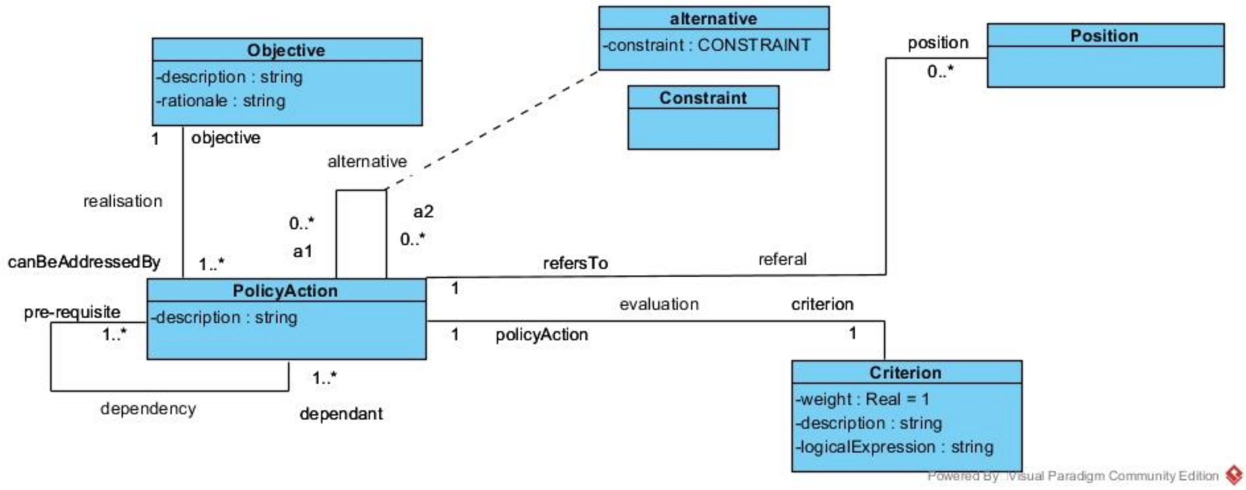


Figure 11 Policy Action and its key relations to other concepts

The need for alternative actions may arise due to a constraint, such as the lack of sufficient monetary or other resources for realising the alternative actions or a constraint requiring that no two actions that have the same effect should be ever executed together. Modeling such constraints is supported in the PHPDM language through the object attribute `constraint` of the `alternative` association.

In the context of an objective's realisation, one action can also be pre-requisite for another action. This can be modelled through the `dependency` association between actions. In our previous example of objective o1 with four realising actions (a1, a2, a3 and a4), if there are no dependency associations between the actions, a1, a2, a3 and a4 may be executed in any possible order. If, however there is a dependency association between a1 and a3 indicating that a3 is a pre-requisite of a1 (or equivalently a1 depends on a3), then the possible orders of action executions would be: a3 -> (a1, a2, a3). A further constraint in the modeling of actions is that if two actions are associated with a `dependency` association, then they cannot be also associated with an `alternative` association and vice versa.

The plausibility of policy actions in the context of policy objective realisations is evaluated by criteria (see association `IsEvaluatedBy` between `PolicyAction` and `Criterion` in Figure 11. Criteria express conditions based upon the outcomes of data analytic workflows (i.e., the evidence arising from the processing of big data). Each policy action may be evaluated by one criterion but there can be actions without an associated criterion. This is to allow a partial specification of policy models containing actions for which the right form of evaluation criteria has not been established yet. A detailed description of how criteria are specified in PHPDM models is given under the class `Criterion` below.

Finally, policy actions may be associated with zero or more positions expressed by different stakeholders. This is expressed through the association `referral` between Policy`Action` and `Position`. In a model there can be action for which no position has been expressed by any stakeholder. Hence, the cardinality of the association `refers`To at the end of the class `Position` is 0..*.

Formally policy actions are defined as instances of the class `PolicyAction` in OWL. The definition of this class is listed in Appendix B (B.1.1.4) along with the class axioms (B.1.2.4) and examples (B.1.3.4).

3.3.1.6. Criterion

Purpose

The purpose of the class Criterion in the PHPDM language is to express conditions for evaluating the plausibility of actions based on the outcomes of the analysis of data (i.e. the outputs of the data analytic workflows in a policy model).

Concept Definition

Criteria are expressed in policy models as instances of the class `Criterion`. The definition of a criterion has three data properties. These are: (i) the attribute `weight`, which is a number that defines how important is the criterion compared to the other criteria, (ii) the attribute `description`, which includes a textual description of the conceptual meaning of the criterion, and (iii) the attribute `logical expression`, which includes the definition of the criterion as a formal logical expression.

As criteria are to be evaluated with respect to the outcomes of data analytic workflows and these outcomes are stored as data sets in the EVOTION repository, the logical expression of criteria is specified as a Phoenix SQL-like expression over these output data sets. These data sets need to be explicitly specified in the specification of the criterion via the association constraints. A criterion may refer to more than one such data sets, as a data analytics workflow may generate more than one output data sets (e.g., a data mining model and the predictions generated by it). A constraint in the specification of criteria is that the Phoenix SQL statement that is used as the value of the logical expression of the criterion must refer to all the data sets and only the data sets associated with it via the constraints association.

One Policy Action is evaluated by one criterion and each criterion specifies a data analytics workflow. Many criteria can specify the same data analytics workflow. Each criterion constraints one or more datasets which are input or output of data analytics tasks that compose the data analytics workflows.

It should be noted that, whilst the PHPDM language provides a means for expressing the weight of different criteria, it does not provide any methodological support for doing so. This is not an omission; it is a choice that reflects the need to be able to accommodate different methodological approaches for specifying weights of criteria in multi criteria decision making (e.g., the use of the Analytic Hierarchy Process [67] or sensitivity analysis [68]).

Formally, criteria are defined as instances of the class `Criterion` in OWL. The definition of this class is listed in Appendix B (B.1.1.5) along with the class axioms (B.1.2.5) and examples (B.1.3.5).

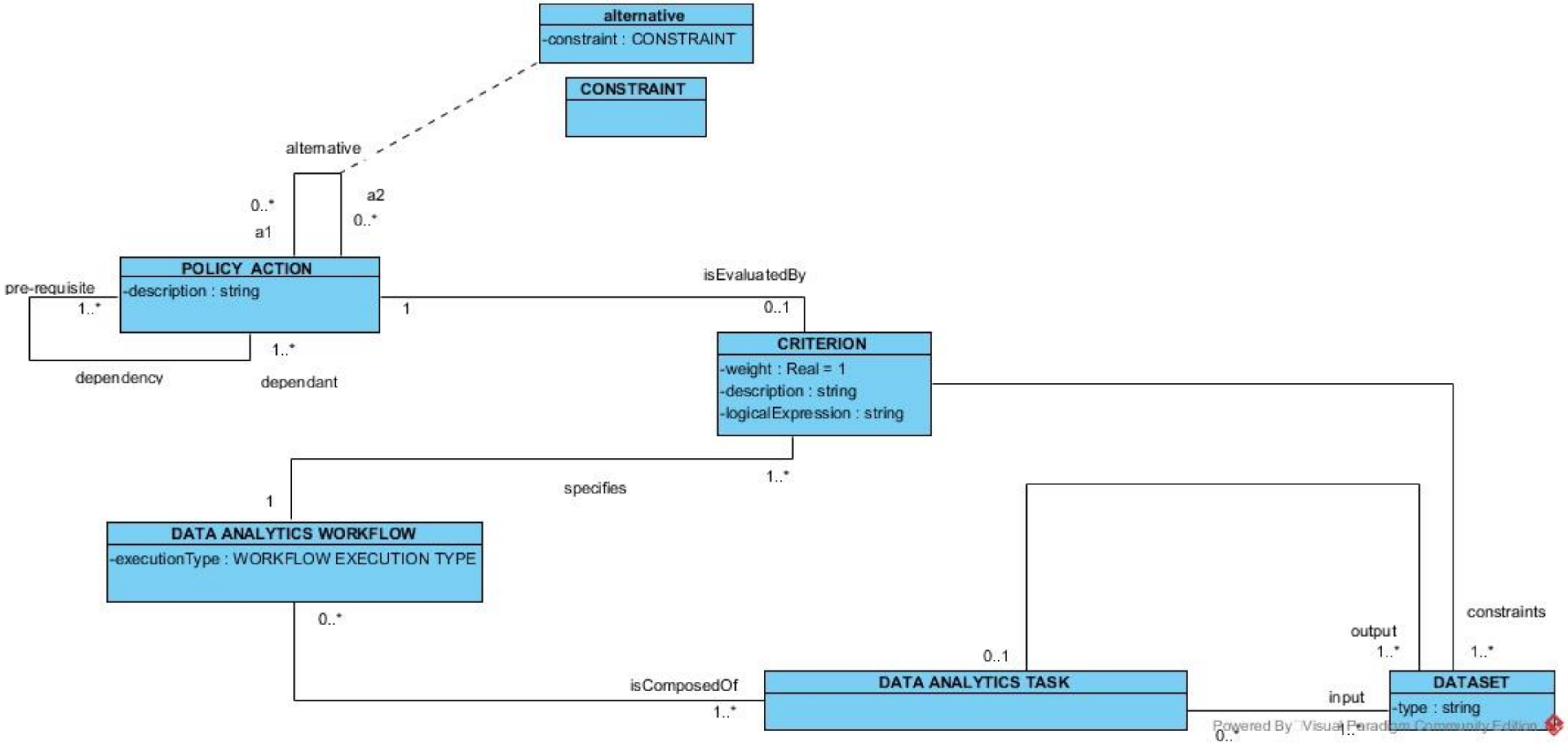


Figure 12 Criterion and its key relations to other concepts

Figure 12 shows a graphical view of `Criterion` and its key relations to other concepts.

3.3.2. Policy Making Module: Stakeholders and Decision-Making Processes

3.3.2.1. Purpose of Module & Overview of Module Concepts

This module of our modeling language enables the model to describe the policy making decision processes. The policy maker specifies the stakeholders of the policy model, the decisions they propose, which can be neutral, opposing and supportive, and connects the specified policy actions with the positions they refer to.

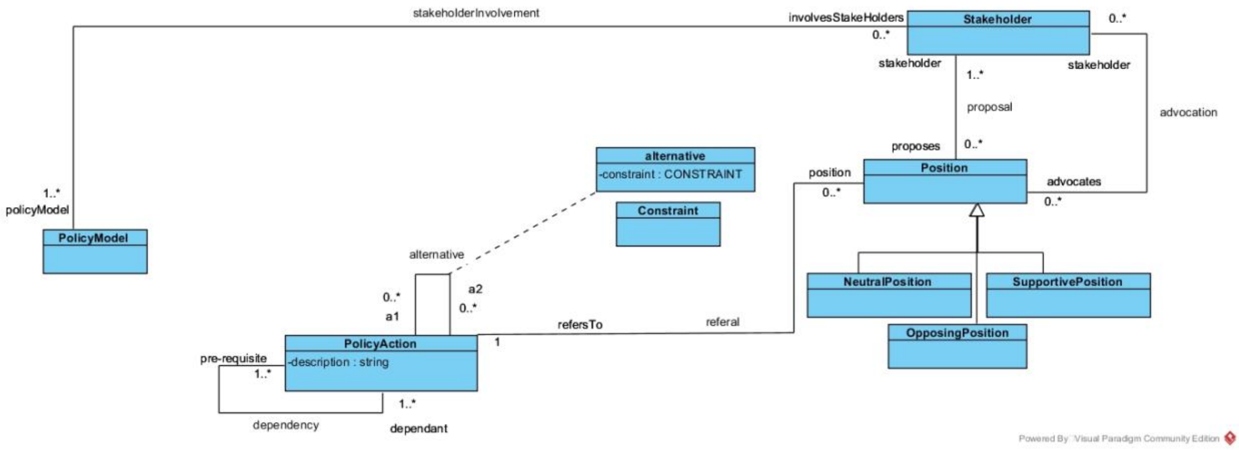


Figure 13 Policy Stakeholders and Decision-Making Processes

An overview of the module concepts is presented in Figure 13.

3.3.2.2. Stakeholder

Purpose

This ontology concept specifies all the different stakeholders who take part in the decision-making process of the policy model instance.

Concept Definition

This ontology concept specifies all the different stakeholders, who take part in the decision-making process of the policy model instance. One Policy Model involves several stakeholders who propose positions or declare that proposed positions advocate them.

Formally, stakeholders are defined as instances of the class `Stakeholder` in OWL. The definition of this class is listed in Appendix B (B.1.1.6) along with the class axioms (B.1.2.6) and examples (B.1.3.6).

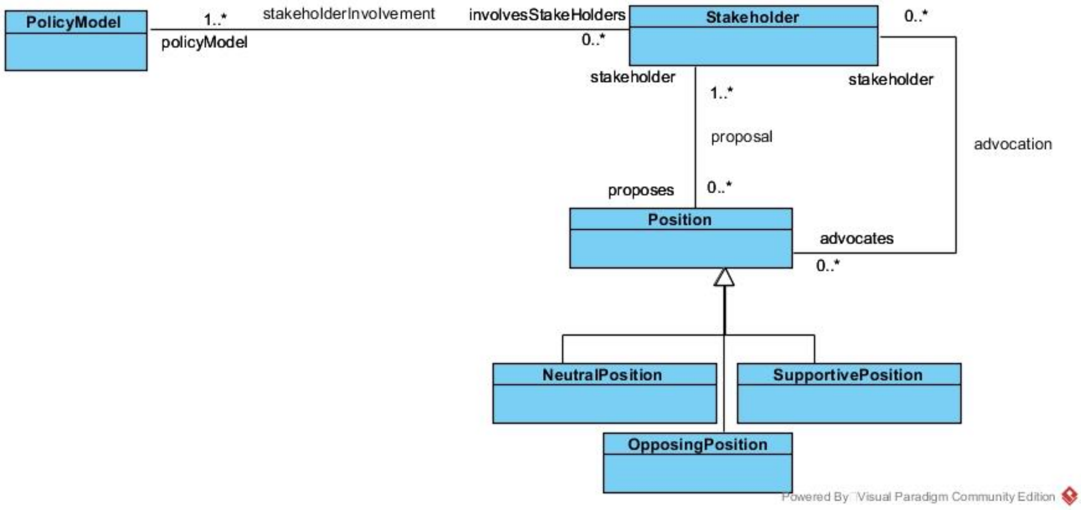


Figure 14 Stakeholder and its key relations to other concepts

Figure 14 shows a graphical view of `Stakeholder` and its key relations to other concepts.

3.3.2.3. <u>Position</u>

Purpose

The concept of `Position` specifies all the different positions that the stakeholders propose or advocate.

Concept Definition

The position concept has as subclasses the following: i) `NeutraPosition`, ii) `OpposingPosition` and iii) `SupportivePosition`, to describe the different types of the stakeholder's positions. The stakeholders propose positions or declare that proposed positions advocate them. In addition, positions refer to specific policy actions.

Formally, positions are defined as instances of the class `Position` in OWL. The definition of this class is listed in Appendix B (B.1.1.7) along with the class axioms (B.1.2.7) and examples (B.1.3.7).

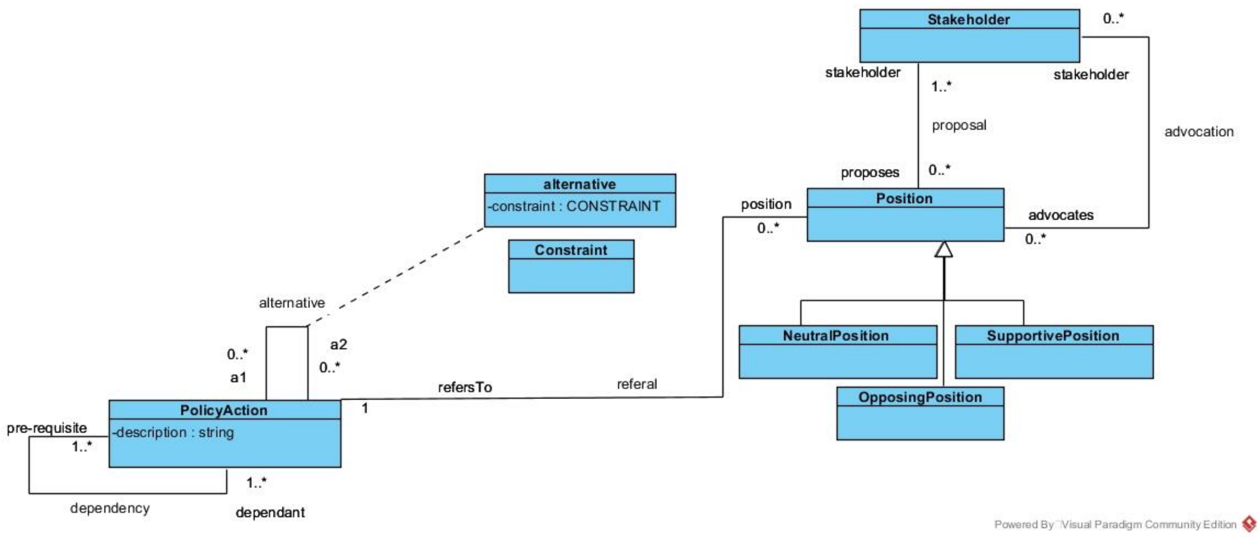


Figure 15 Position and its key relations to other concepts

Figure 15 shows a graphical view of `Position` and its key relations to other concepts.

3.3.3. Data Analytics and Evidence Module: Data Analytics Workflows

3.3.3.1. Purpose of Module & Overview of Module Concepts

The purpose of this core module is to enable the policy makers to collaborate with the data scientists and model the data analytics workflows associated with the policy actions that are required to provide evidence to the policy making process. Each policy action is evaluated against a criterion, which specifies a data analytics workflow. The data analytics workflow is composed of various data analytics tasks, that provide the evidence to the policy model. A data analytics workflow can have as WORKFKLOW EXECUTION TYPE, EXECUTION UPON REQUEST (i.e. to be executed when someone requests it), or PERIODIC EXECUTION (i.e. to be executed periodically from executionStart to executionEnd with a period as specified by periodicity and datasets constrained by dataTiming, e.g., execute workflow from 1/1/17 to 31/12/27 periodically every 6 months), or DATA CHANGE DRIVEN (i.e. to execute the workflow when there is a significant increase in the size of the data set that was previously analysed, e.g., 20% more data in it). Datasets can be: absolute, i.e., sets from a start data to the time point of the execution; and shifting, i.e., sets taken over a shifting time span specified by period and starting from the initial (START) or the last execution (END). An overview of the module concepts is presented in Figure 16.

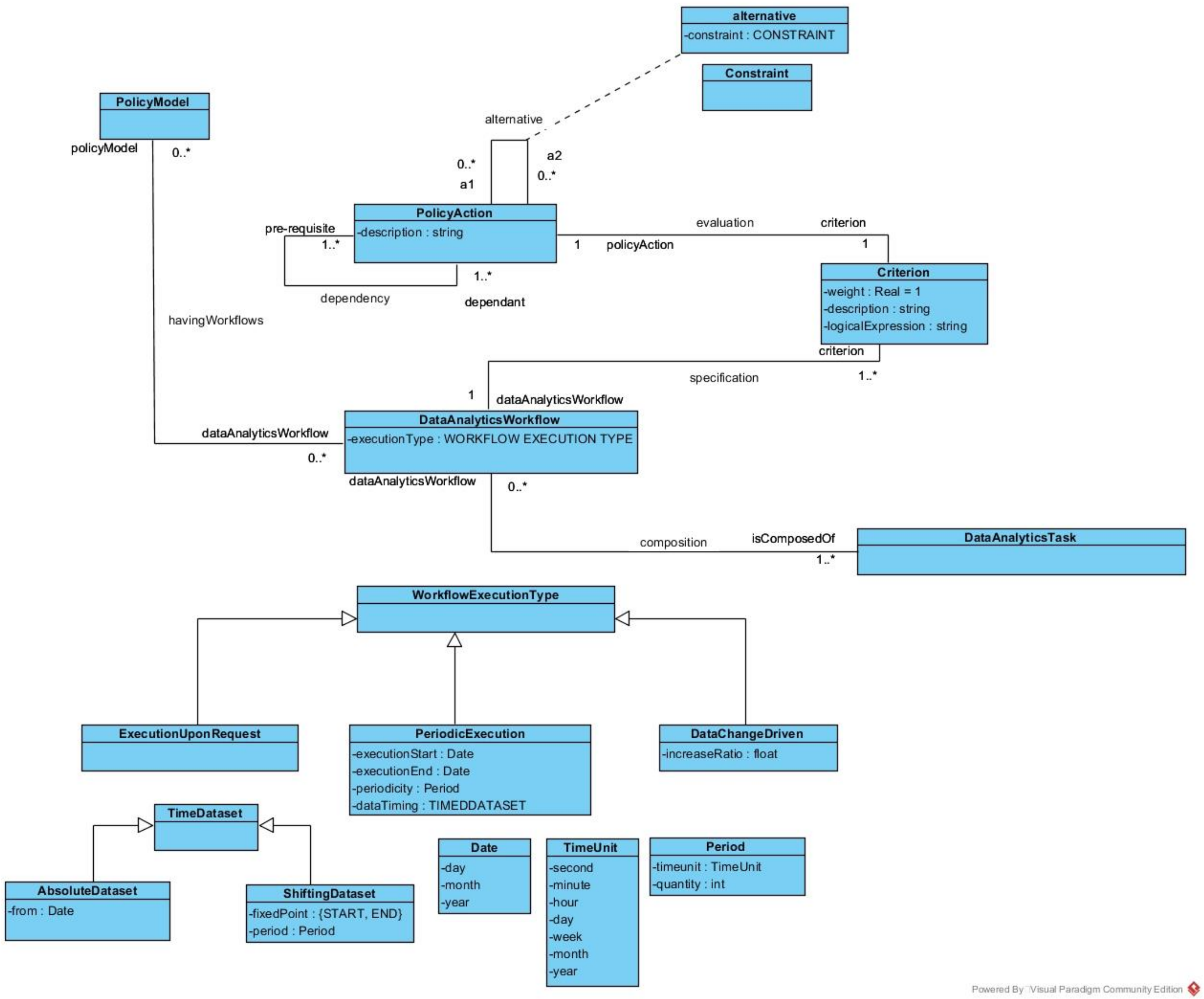


Figure 16 Data Analytics Workflows

3.3.3.2. Data Analytics Workflow

Purpose

The data analytics workflow concept constitutes the specification of an analytics process whose purpose is to obtain evidence for assessing policy actions.

Concept Definition

This core concept has one property: `DataExecutionType`, which is the next concept that is described. As mentioned in the overview of the module, a policy model has one or more data analytics workflows that specify criteria that are used to evaluate policy actions. The workflows are composed of various data analytics tasks. The sequence of data analytics tasks is defined by the datasets that they have as input and output. This will be described in more detail in the Dataset concept. Figure 16 shows a graphical view of `DataAnalyticsWorkflow` and its key relations to other concepts.

Formally, data analytics workflows are defined as instances of the class `DataAnalyticsWorkflow` in OWL. The definition of this class is listed in Appendix B (B.1.1.8) along with the class axioms (0) and examples (B.1.3.8).

3.3.3.3. Workflow Execution Type

Purpose

The workflow execution type concept constitutes the type of execution of the data analytics workflow.

Concept Definition

As mentioned in the module overview, a data analytics workflow can have as `WorkflowExecutionType`, `ExecutionUpponRequest` (i.e. to be executed when someone requests it), or `PeriodicExecution` (i.e. to be executed periodically from `executionStart` to `executionEnd` with a period as specified by periodicity and datasets constrained by `dataTiming`, e.g., execute workflow from 1/1/17 to 31/12/27 periodically every 6 months), or `DataChangeDriven` (i.e. to execute the workflow when there is a significant increase in the size of the data set that was previously analysed, e.g., 20% more data in it). Datasets can be: absolute, i.e., sets from a start date to the time point of the execution; and shifting, i.e., sets taken over a shifting time span specified by period and starting from the initial (START) or the last execution (END).

Formally, workflow execution types are defined as instances of the class `WorkflowExecutionType` in OWL. The definition of this class is listed in Appendix B (B.1.1.9) along with the class axioms (B.1.2.9) and examples (B.1.3.9).

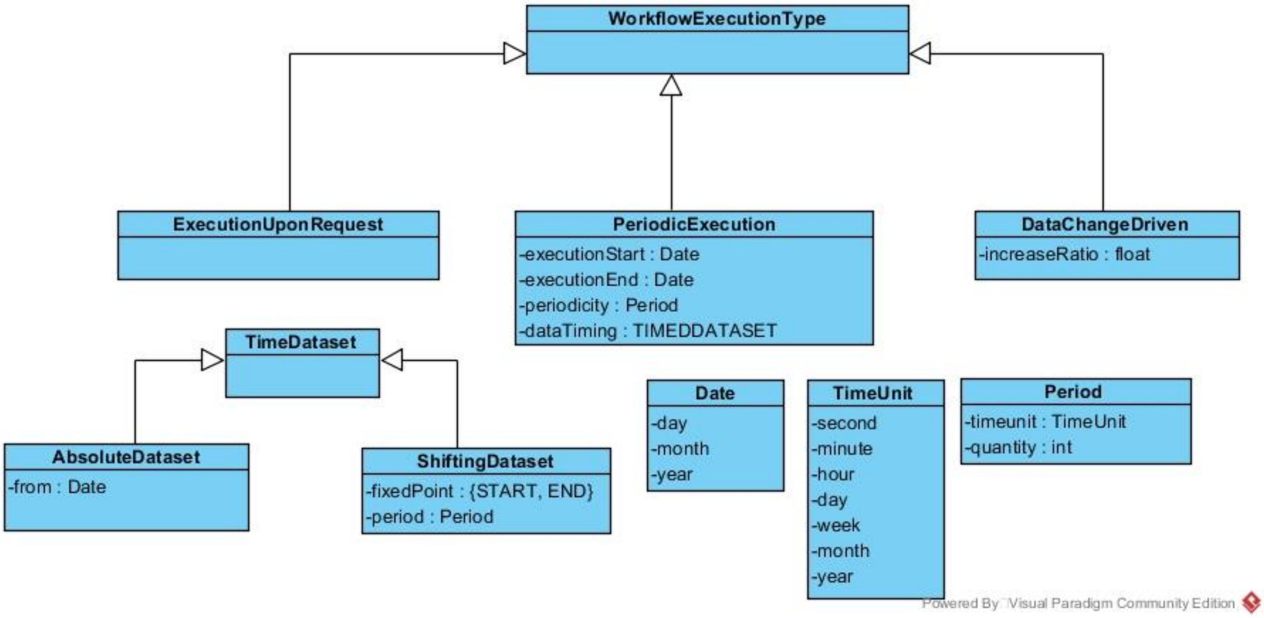


Figure 17 Workflow Execution Type and its subclasses

Figure 17 shows `WorkflowExecutionType` class and its subclasses.

3.3.4. Data Sets

3.3.4.1. Purpose of Module & Overview of Module Concepts

The datasets module of our language is used to connect the PHPDM modeling to the data repository of Evotion. The core concept of the module is the Dataset, which can be a static set, a data stream or a data analytics model.

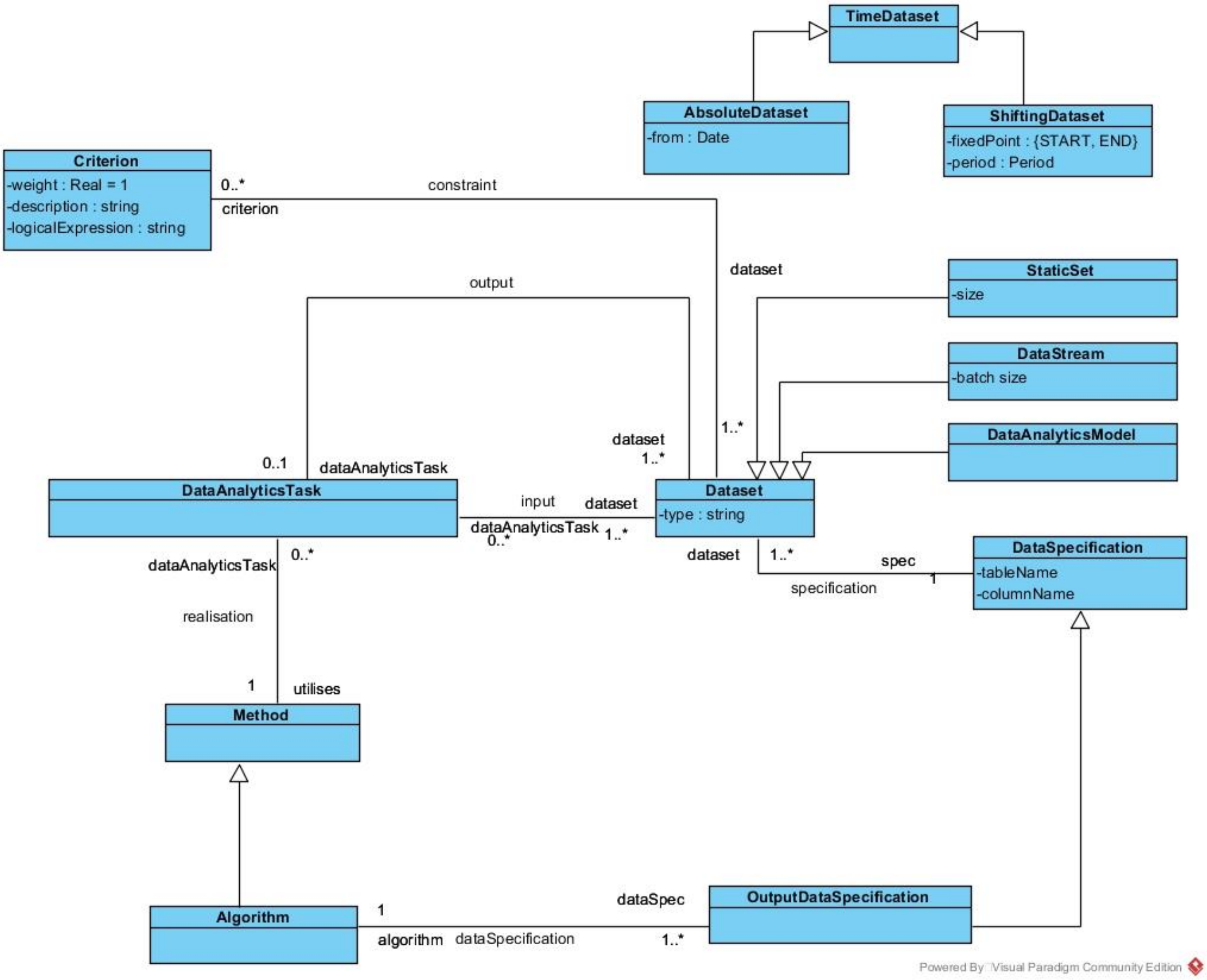


Figure 18 Data Sets

Each criterion can constraint several datasets which are input and output to data analytics tasks.

Each algorithm of the data analytics task has an output data specification for the output dataset of the task. In general, all datasets have a data specification, which is the mapping to Evotion data repository. From the workflow execution type concept, we also see the timed dataset concept which can be absolute or shifting, as we described in that concept description. An overview of this module's concepts and their relations is presented in Figure 18.

3.3.4.2. Data Set

Purpose

The data set concept of our language is used as a bridge between the PHPDM modeling and Evotion data repository. It is used to provide the inputs and outputs to the data analytics tasks for the appropriate evidence to be provided to the decision-making process.

Concept Definition

A data set can be a static set, a data stream, or a data analytics model. It has one object property called type. As mentioned in the overview of the module concepts, each algorithm of the data analytics task has an output data specification for the output dataset of the task. In general, all datasets have a data specification, which is the mapping to Evotion data repository. From the workflow execution type concept, we also see the timed dataset concept which can be absolute or shifting, as we described in that concept description. An overview of this concept's relations to other concepts is presented in Figure 18.

Formally, data sets are defined as instances of the class `Dataset` in OWL. The definition of this class is listed in Appendix B (B.1.1.10) along with the class axioms (B.1.2.10) and examples (B.1.3.10).

3.3.4.3. Data Specification

Purpose

The data specification concept of our language is used to map the datasets of our language with Evotion data repository. This is both for the inputs and the outputs to the data analytics tasks for the appropriate evidence to be provided to the decision-making process.

Concept Definition

Each data set has a data specification, which is the mapping to Evotion data repository. We also introduce the output data specification concept which is a subclass of data specification and constitutes the mapping of the output of each algorithm to the Evotion repository. This concept has two data properties: `tableName`, which is the mapping with a table of Evotion Repository, and `columnName`, which is the mapping with a specific column of the repository. If the data is transient (stored only in the memory), then `tableName` and `columnName` is not set. An overview of this concept's relations to other concepts is presented in Figure 19.

Formally, data specifications are defined as instances of the class `DataSpecification` in OWL. The definition of this class is listed in Appendix B (B.1.1.11) along with the class axioms (B.1.2.11) and examples (B.1.3.11).

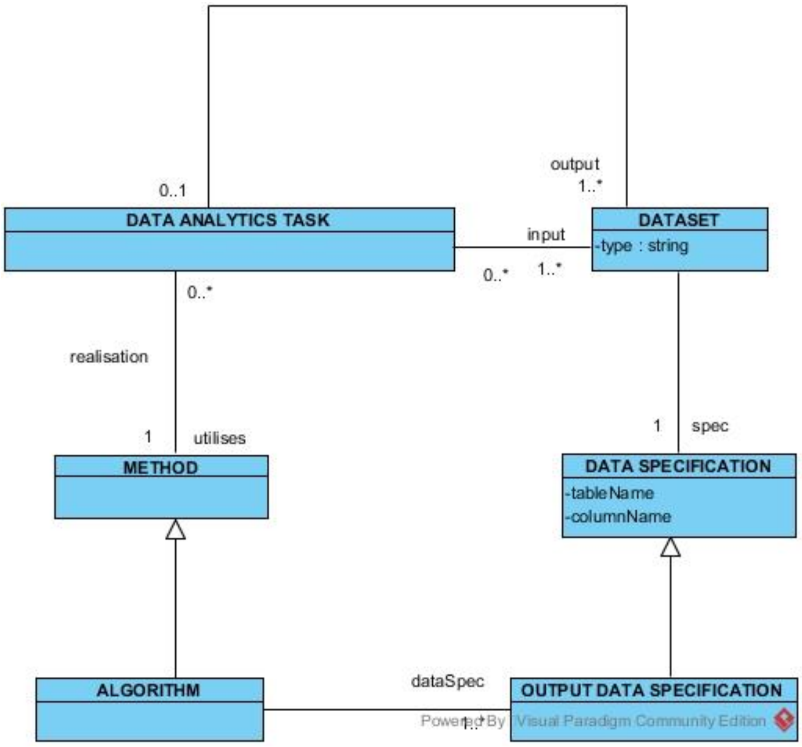


Figure 19 Data Specification and its key relations to other concepts

3.3.5. <u>Data Analytics Tasks</u>

3.3.5.1. <u>Purpose of Module & Overview of Module Concepts</u>

The data analytics tasks module is the core of the data analytics processes of our system that provides the appropriate evidence to the decision-making processes. The core module is the data analytics task which defines an individual analytics task within an analytics workflow. A data analytics task can be one of the following types: (i)a data mining task, (ii)a statistical analysis task, (iii)a simulation task, (iv)a social media analytics task, or (v)a text mining task. Each type of task will be described in detail below in the document. Each task has datasets as inputs and/or outputs. These can be static or dynamic (streams) or may capture models. Datasets are also used to evaluate policy making criteria. Each task utilizes a method, which can be an operation for data processing tasks or an algorithm for data mining, statistical analysis, text mining or simulation tasks. Each algorithm has an output data specification, which is the mapping of its output with Evotion Data Repository. An overview of this module's concepts is presented in Figure 20.

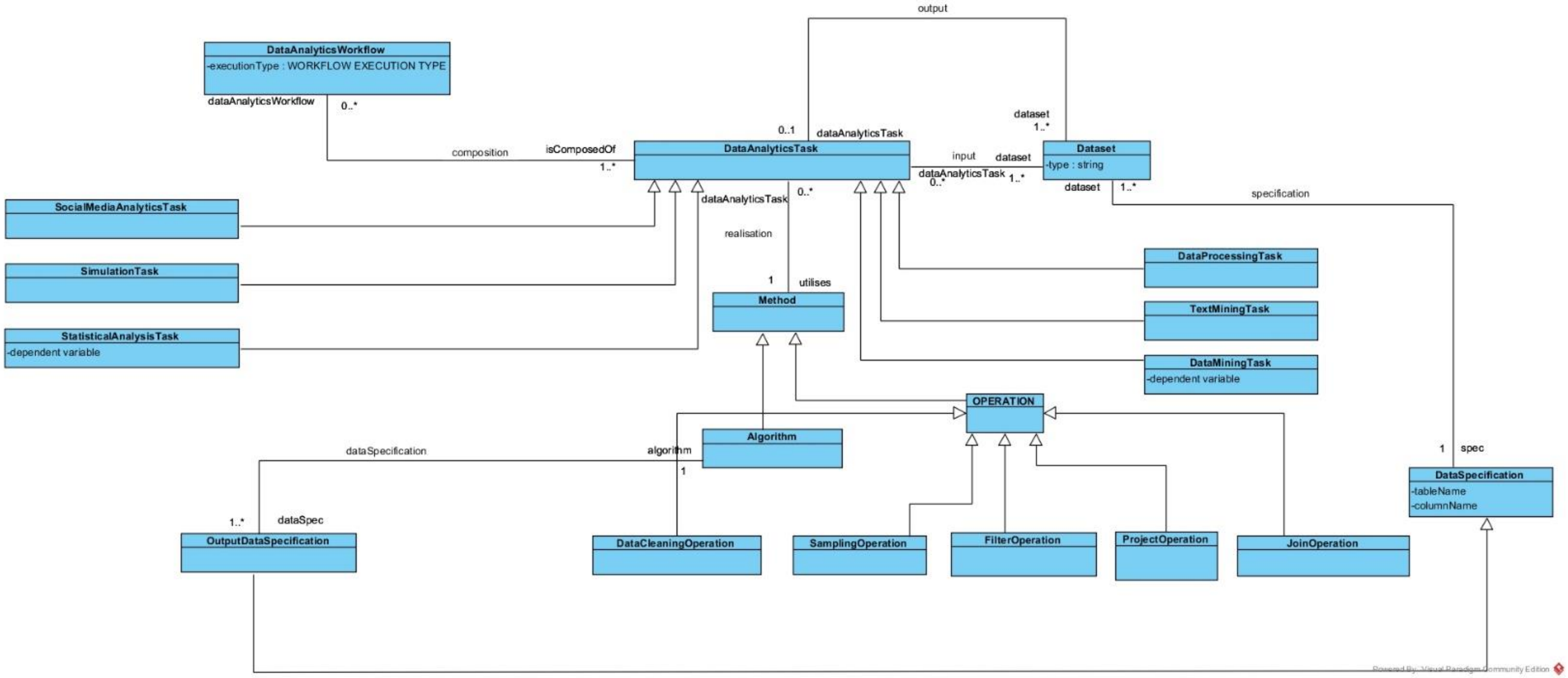


Figure 20 Data Analytics Tasks

3.3.5.2. Data Analytics Task

Purpose

The purpose of this core module of our language is to describe an individual analytics task within an analytics workflow, which is part of the provision of the evidence to the decision-making process.

Concept Definition

As described above, a Data analytics Task is part of a Data Analytics Workflow (the workflow is composed of various Data Analytics Tasks). Each Task has as input one or more Datasets and as output other Datasets. The output of one task can be input to another task of the workflow.

In our language we have the following types (subclasses) of Data Analytics Tasks: Data Processing Tasks (i.e. Basic data processing – e.g. data filtering and merging), Data Mining Tasks (supervised or unsupervised), Statistical Analysis Tasks (e.g. linear regression, ANOVA, statistical testing), Simulation Tasks (i.e. Data Mining or Statistical Analysis Tasks with a hypothesis), Social Media Analytics Tasks (i.e. Twitter or Facebook analytics tasks) and Text Mining Tasks (i.e. analysis of the literature with data mining techniques). Each type of task will be described below in the document. An overview of this concept's relations to other concepts is presented in Figure 21.

Formally, data analytics tasks are defined as instances of the class `DataAnalyticsTask` in OWL. The definitions of this class and its subclasses are listed in Appendix B (0) along with the class axioms (B.1.2.12). This is the abstract class of any data analytics task (e.g., statistical analysis, data mining analysis, date pre-processing tasks etc.). Therefore, we cannot give concrete examples of it. Examples of different, specific types of data analytic tasks are given in the subsequent sections.

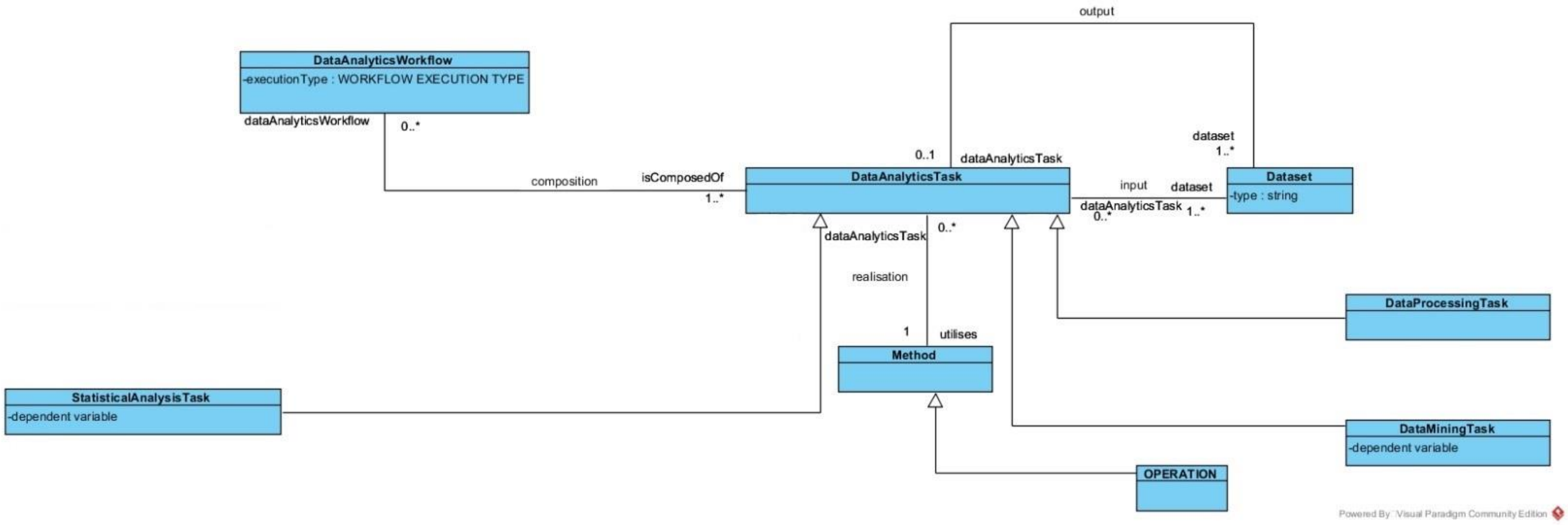


Figure 21 Data Analytics Task and its key relations to other concepts

3.3.5.3. Method

Purpose

The purpose of this module is to express the methods utilized by the data analytics tasks. More specifically, each data processing task utilizes an operation, which is a subclass of method, and each data mining, statistical analysis task, data mining task, simulation task and text mining task utilise algorithms.

Concept Definition

This concept has two subclasses: Algorithm and Operation. The Operation concept has the following subclasses: (i)data cleaning, (ii)sampling, (iii)filter, (iv)project (v)join operations. These operations are going to be described in the Data pre-processing tasks part, below.

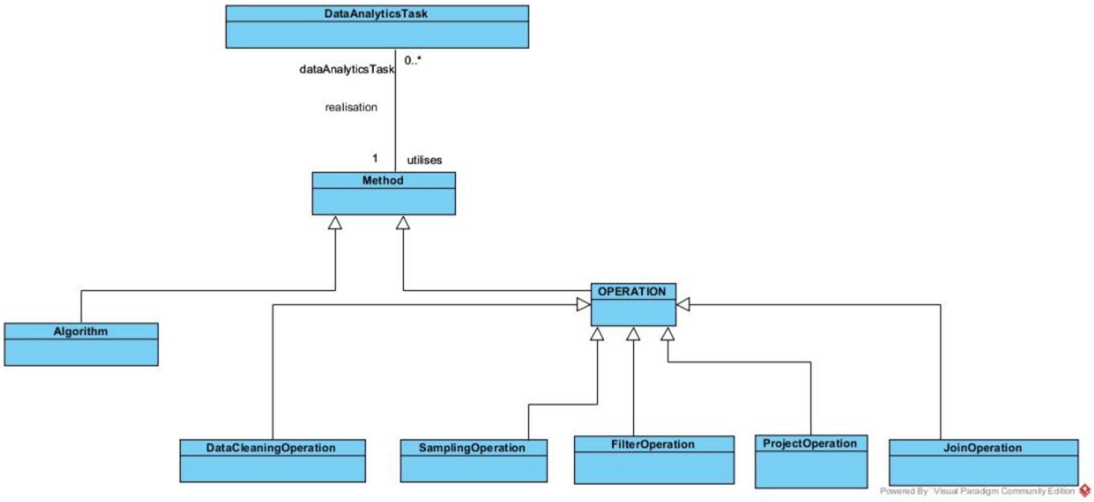


Figure 22 Method and its key relations to other concepts

An overview of this concept's subclasses and relations to other concepts is presented in Figure 22. Formally, methods are defined as instances of the class `Method` in OWL. The definitions of this class and its subclasses, as well as the subclasses of `Operation` are listed in Appendix B (B.1.1.13) along with the class axioms (B.1.2.13). This is the abstract class of all methods used in data analytic tasks (e.g., statistical analysis algorithm, data mining algorithm, data processing operation). Therefore, we cannot give concrete examples of it. Examples of different, specific types of data analytic methods are given in the subsequent sections.

3.3.6. Data Processing Tasks

3.3.6.1. Sampling operation

Purpose

The concept of Sampling operation specifies how to horizontally choose elements from a set, so that they are representative enough for further evaluations. It is required to reduce the complexity of consecutive analytic tasks, which could process a sample. In general, it refers to samples taken randomly, with a specific ordering or following a specific statistical approach.

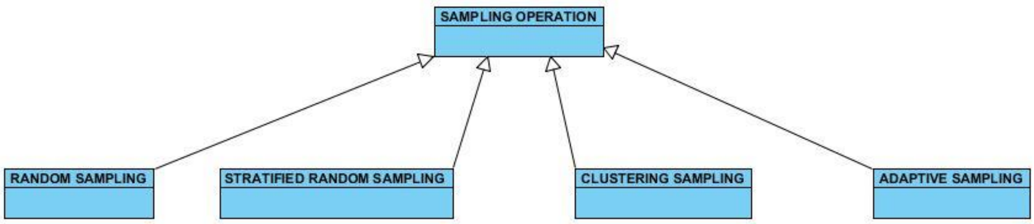


Figure 23 Sampling Operation and its subclasses

Concept Definition

The sampling operation, to preserve representativeness of the subset sampled for the dataset, allows different approaches depending on the nature of the data and of the needs of the subsequent analytic tasks. These approaches are specializations of sampling operation and refer to statistical sampling methods like Random Sampling, Stratified Random Sampling, Clustering Sampling, Adaptive Sampling to name but a few ["Database Sampling for Data Mining"]. Sampling is fundamental while working with Big Data to allow some quick preliminary draft evaluation on the feasibility of executing a specific analytic task. In the process of policy making this is applied for having early feedbacks about the feasibility of some of the possible alternative evaluation approaches before asking a complete Big Data analytic execution that may take some time to be completed. This concept and its subclasses are presented in Figure 23.

Formally, sampling operations are defined as instances of the class `SamplingOperation` in OWL. The definitions of the subclasses of this class are listed in Appendix B (B.1.1.14) along with the class axioms (B.1.2.14) and examples (0).

3.3.6.2. Filter operation

Purpose

The Filter operation allows to specify a subset of a dataset to be returned based on a specified condition.

Figure 24 Filter Operation

Concept Definition

Filter operation returns a subset of the elements in a given dataset with only those elements that match the condition expressed by the filter. In general conditions are expressed in terms of the values of one or more attributes of the dataset. Big Data filtering is applied to streams with the aim of lowering the number of elements to the ones that are really needed for the evaluation. Current Big Data system allows to express filtering using SQL-like language.

The condition is expressed using SQL-like language where the SELECT allows to choose the attributes of the dataset and the WHERE clause is a predicate expressed in terms of attributes and values for selecting the entries of the dataset. The filter operation is presented in Figure 24.

Formally, filter operations are defined as instances of the class `FilterOperation` in OWL. The definition of this class was presented in the `Method` definition. In Appendix B (B.1.1.15) we present the formal definition of the attribute `condition` of this class, which is formally defined as a data property in OWL along with examples (B.1.3.13).

3.3.6.3. Join operation

Purpose

The concept of Join operator specifies how to combine two or more source/dataset of data. This is needed preliminary to an analytic evaluation that requires information coming from different sources.

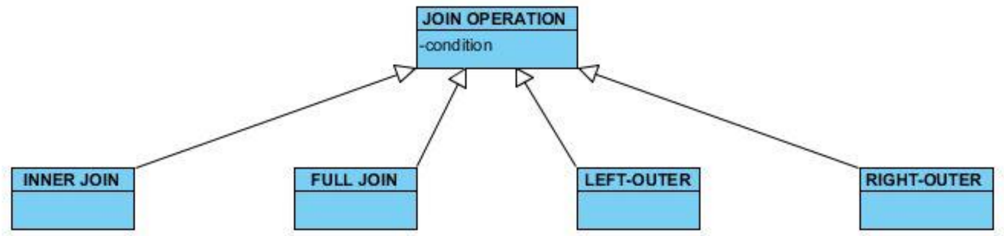


Figure 25 Join Operation and its subclasses

Concept Definition

In Big Data environment join operation is currently mutated from RDBMS approach, due to the diffusion of SQL-like interface for querying a Big Data system. This allows to consider different type of SQL-like join including inner join (intersection), full join (union), left-outer, right-outer, to name but a few. While executing the join, a condition for joining is expressed in terms of common attributes. Even if mutated from RDBMS, they are in general not efficiently implemented, therefore in many cases the source of data is designed in a way that prevent the use of Join as much as possible. For instance, considering a number of streams, one join by design is to make them part of the same queue (e.g. Kafka) and joining them as a union at queue handling level. Another example of simple join is the stream enrichment, which is obtained enriching a stream with data from structured dataset. The condition is expressed using SQL-like language. The subclasses of Join operation are presented in Figure 25.

Formally, join operations are defined as instances of the class `JoinOperation` in OWL. The definition of this class was presented in the `Method` definition. In Appendix B (B.1.1.16) we present the formal definition of the attribute "`condition`" of this class, which is formally defined as a data property in OWL, as well as its subclasses along with the class axioms (B.1.2.15) and examples (B.1.3.14).

3.3.6.4. Project operation

Purpose

The Project operation specifies how to vertically choose elements from a dataset. This operator permits to select fewer yet representative attributes, rather than selecting all attributes of a dataset.

Figure 26 Project Operation

Concept Definition

The Project operation produces a vertical subset of the data, extracting the values of specified attributes for instance eliminating duplicates. This operation is a sort of feature selection, not associated to an analytic evaluation, but performed in the pre-processing step where the space reduction has the goal of eliminating unnecessary attributes. It can be associated to simple aggregation algorithms (i.e., projection rule) for aggregating multiple attributes or values into a new attribute with new values.

While processing Big Data on a batch dataset, the attribute selection by means of projection is quite straightforward, however when processing is done on streams, the projection is often obtained at map stage of a map-reduce approach. `ProjectionRule` is expressed as SQL-like language. Project operation is presented in Figure 26.

Formally, project operations are defined as instances of the class `ProjectOperation` in OWL. The definition of this class was presented in the `Method` definition. In Appendix B (B.1.1.17) we present the formal definition of the attribute `projectRule` of this class, which is formally defined as a data property in OWL along with examples (B.1.3.15).

3.3.6.5. <u>Data Cleaning Operation</u>

Purpose

The concept of Data cleaning specifies the set of operations required to “clean” the dataset, for instance converting unwanted data types or removing unwanted values.

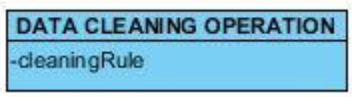


Figure 27 Data Cleaning Operation

Concept Definition

Cleaning refers to pre-processing operations aimed at preparing the dataset for further processing (i.e., cleaning rule). Data sets cleaning can possibly result in inconsistent, invalid, incomplete, inaccurate, non-uniform data sets or having duplicates in data sets. These issues can be caused either by incorrect data entry, system failure, data corruption, to name but a few, and they are severe obstacles for further evaluations. The data cleaning operation is presented in Figure 27.

Formally, data cleaning operations are defined as instances of the class `DataCleaningOperation` in OWL. The definition of this class was presented in the `Method` definition. In Appendix B (B.1.1.18) we present the formal definition of the attribute `cleaningRule` of this class, which is formally defined as a data property in OWL along with examples (B.1.3.16).

3.3.7. Statistical Analysis Tasks

3.3.7.1. Purpose of Module & Overview of Module Concepts

The Statistical Analysis Tasks module is introduced to express and test the statistical analysis of the general policy models with Algorithms through Method (via Data Analytics Task). This entity is also intended to create new scenario hypotheses for statistical analysis tasks in general with complete simulation results also. To be able to complete these complex statistical tasks that the policy makers wants, for the creation of the ontology we were inspired by the STATO ontology[69].

The Statistical Analysis Task entity has the role of all the statistical analysis models of the main models of policy makers through the Data Analytics Workflow. To be more concrete through Data Analytics Task (that follows the workflow), the Statistical Analysis can be achieved for every different hypothesis. Firstly, Data Analytics Task with the help of the Data Processing Task and Data Mining Task could help to define the required data for the (statistical) model that is formulated for every `PolicyModel`. Secondly, through Method, Algorithm (via Data Analytics Task) presents the particular algorithm that is needed. In addition, the second subclass of Method, Operation can be used here for various minor tasks, such as data cleaning operation, sampling operation etc.

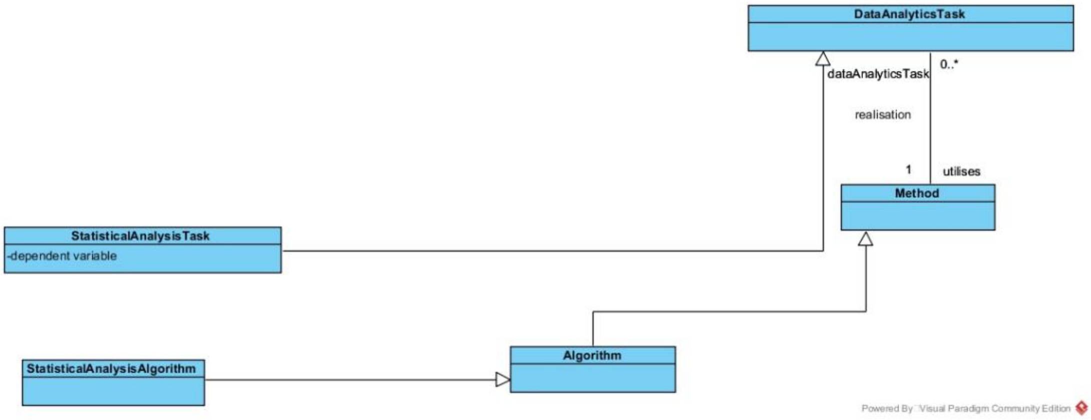


Figure 28 Statistical Analysis Task and its key relations to other concepts

Figure 28 shows a graphical view of Statistical Analysis Task and how the other key elements involved in their definition are related to each other.

3.3.7.2. Statistical Analysis Algorithm

Purpose

The Statistical Analysis Algorithm concept was introduced to group and classify all the available statistical analysis algorithms of our platform.

Concept Definition

In this section we are going we are going to define each algorithm type separately. Formally, statistical analysis algorithms class are defined as subclasses of the general class `StatisticalAnalysisAlgorithm` in OWL. The definition of this class is listed in Appendix B (B.1.1.19) along with the class axioms (B.1.2.16). Examples of different, specific types of statistical analysis algorithms are given in the subsequent sections.

Linear Regression

Purpose

From the Statistical Analysis Task that has been analysed above the Linear Regression is introduced to test the specific statistical model from the ontology we are examining. This entity is intended to create, with the given information data (from every new scenario hypotheses) complete statistical simulation results for the specific linear approach for modeling the relationship between a scalar dependent variable *y* and one or more explanatory variables.

Concept Definition

The Linear Regression entity has the distinct role of the visual embodiment of the model that describes. To be more concrete, linear regression model is a model which attempts to explain data distribution associated with response/dependent variable in terms of values assumed by the independent variable and uses a linear function or linear combination of the regression parameters and the predictor/independent variable(s). Linear regression modeling makes a number of assumptions, which includes homoscedasticity (Constance of variance). Statistical Analysis tasks such as this one here can use the data that has been stored through Dataset to conclude to the best possible model for every case and every scenario the policy maker wants to formulate.

Formally, Linear Regression models of statistical analysis are defined as instances of the class `LinearRegression` in OWL. The definition of this class is listed in Appendix B (B.1.1.20) along with an example (B.1.3.17).

ANOVA

Purpose

From the Statistical Analysis Task that has been analysed above the Analysis of variance (ANOVA) models are introduced to test statistical models for the current ontology. This entity, with all the subclasses that includes, is intended to create, with the given information data (from every new scenario hypotheses), complete statistical simulation results for modeling the differences among group means and their associated procedures (such as "variation" among and between groups).

Concept Definition

The Analysis of variance (ANOVA) collection of models entity has the distinct role of the visual embodiment of the model that describes. To be more concrete the ANOVA models attempt to explain data transformation in which a statistical test is performed to evaluate the null hypothesis that the means computed over the different groups as specified by the investigator do not differ. The test compares an F-statistics (a ratio of means) to an F-distribution and produces a p-value, used to reject or accept the null hypothesis given a false positive rate. The test assumes normality and equivariance of the data. Statistical Analysis tasks such as this one here can use the data that has been stored through Dataset to conclude to the best possible model for every case and every scenario the policy maker wants to formulate.

Formally, the Analysis of variance (ANOVA), collection of models, of statistical analysis is defined as an OWL class with four subclasses. The definitions of these classes are listed in Appendix B (B.1.1.21) along with an example (B.1.3.18).

Breusch-Pagan Test

Purpose

From the Statistical Analysis Task that has been analysed, the Breusch-Pagan Test is actually a chi-squared test, used to test for heteroscedasticity in a linear regression model. This entity is intended to create, with the given information data (from every new scenario hypotheses) complete statistical simulation results. The test statistic is distributed $n\chi_2$ with k degrees of freedom. If the test statistic has a p-value below an appropriate threshold (e.g. $p<0.05$) then the null hypothesis of homoscedasticity is rejected, and heteroscedasticity assumed.

Concept Definition

The Breusch-Pagan Test entity has the distinct role of the visual embodiment of the test model that describes. Moreover, Breusch-Pagan test is a statistical test which computes a score test of the hypothesis of constant error variance against the alternative that the error variance changes with the level of the response (fitted values), or with a linear combination of predictors. Statistical Analysis tasks such as this one here can use the data that has been stored through Dataset to conclude to the best possible statistical model test for every case and every scenario the policy maker wants to formulate, particularly if they want to give a simple test for heteroscedasticity and random coefficient variation.

Formally, the Breusch-Pagan Test model of statistical analysis algorithm is defined as a class in OWL. The definition of this class is listed in Appendix B (B.1.1.22) along with an example (B.1.3.19).

F-test

Purpose

From the Statistical Analysis Task that has been analysed above the F-test is actually a statistical test from the ontology, in which the test statistic has an F-distribution under the null hypothesis. This entity is intended to create, with the given information data (from every new scenario hypotheses) complete statistical simulation results. It is most often used when comparing statistical models that have been fitted to a data set, in order to identify the model that best fits the population from which the data were sampled.

Concept Definition

The F-test entity has the distinct role of the visual embodiment of the model that describes. The F-test checks the difference of two ratios. A key assumption of the test is the normality of the underlying populations from which the compared ratios are calculated. Statistical Analysis tasks such as this one here can use the data that has been stored through Dataset to conclude to the best possible statistical model for every case and every scenario the policy maker wants to formulate.

Formally, the F-test model of statistical analysis task is defined as a class in OWL. The definition of this class is listed in Appendix B (B.1.1.23) along with an example (B.1.3.20).

Fisher's Exact Test

Purpose

From the Statistical Analysis Task that has been analysed above the Fisher's Exact Test is actually a statistical test from the STATO ontology, used in the analysis of contingency tables. This entity is intended to create, with the given information data (from every new scenario hypotheses), complete statistical simulation results. That kind of test is useful for categorical data that result from classifying objects in two different ways and it is used to examine the significance of the association (contingency) between the two kinds of classification.

Concept Definition

The Fisher's Exact Test entity has the distinct role of the visual embodiment of the test model that describes. Additionally, Fisher's exact test is a statistical test, used to determine if there are non-random associations between two categorical variables, that is useful for categorical data that result from classifying objects in two different ways. Statistical Analysis tasks such as this one here can use the data that has been stored through Dataset to conclude to the best possible statistical model for every case and every scenario the policy makers want to formulate, particularly if they want to give in an object/class two different properties or attributes.

Formally, the Fisher's Exact Test model of statistical analysis task is defined as a class in OWL. The definition of this class is listed in Appendix B (B.1.1.24) along with an example (B.1.3.21).

3.3.8. Data Mining Tasks

3.3.8.1. Purpose of Module & Overview of Module Concepts

This module of our language consists of the Data Mining processes for the provision of the evidence to the decision-making processes. The Data Mining Task, which is a subclass of Data Analytics Task inherits all the relationships from it, so, everything already described for the data analytics task applies to it. Data Mining Tasks utilize data mining algorithms, which will be described below. An overview of this module's concepts is presented in Figure 29.

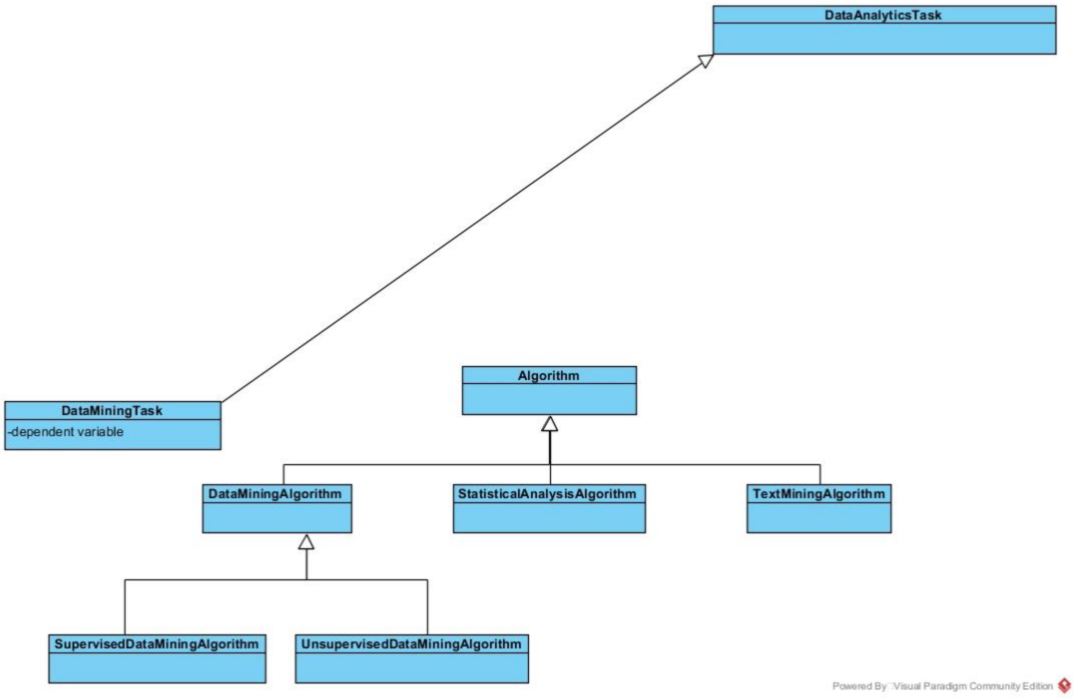


Figure 29 Data Mining Tasks

3.3.8.2. Data Mining Algorithm

Purpose

The Data Mining Algorithm concept was introduced to group and classify all the available data mining algorithms of our platform.

Concept Definition

The algorithms are grouped into supervised and unsupervised. In this section we are going to demonstrate these subclasses and afterwards we are going to define each algorithm type separately. In this class we define two configuration options that are common for all the data mining algorithms: `debug` (if set to true, classifier may output additional info to the console) and `doNotCheckCapabilities` (if set, classifier capabilities are not checked before classifier is built – to be used with caution to reduce runtime).

Formally, data mining algorithms are defined as instances of the class `DataMiningAlgorithm` in OWL. The definitions of this class and its subclasses are listed in Appendix B (B.1.1.25) along with the class axioms (B.1.2.17). Examples of different, specific types of data mining algorithms are given in the subsequent sections.

Supervised Data Mining Algorithms

Naïve Bayes

Overview

Naïve Bayes classifiers [70] are a family of simple probabilistic classifiers based on applying Bayes' theorem with strong (naive) independence assumptions between the features. The featured image is the equation—with P(A|B) is posterior probability, P(B|A) is likelihood, P(A) is class prior probability, and P(B) is predictor prior probability. Numeric estimator precision values are chosen based on analysis of the training data. For this reason, the classifier is not an UpdateableClassifier (which in typical usage are initialized with zero training instances) -- if you need the UpdateableClassifier functionality, use the NaiveBayesUpdateable classifier. The NaiveBayesUpdateable classifier will use a default precision of 0.1 for numeric attributes when buildClassifier is called with zero training instances.

Formally, Naïve Bayes algorithm executions are defined as instances of the class `NaiveBayes` in OWL. The definition of this class and its data properties (configuration options) are listed in Appendix B (B.1.1.26) along with an example (B.1.3.22).

Algorithm configuration options

- useKernelEstimator -- Use a kernel estimator for numeric attributes rather than a normal distribution.
- numDecimalPlaces -- The number of decimal places to be used for the output of numbers in the model.
- batchSize -- The preferred number of instances to process if batch prediction is being performed. More or fewer instances may be provided, but this gives implementations a chance to specify a preferred batch size.
- displayModelInOldFormat -- Use old format for model output. The old format is better when there are many class values. The new format is better when there are fewer classes and many attributes.
- useSupervisedDiscretization -- Use supervised discretization to convert numeric attributes to nominal ones.

Algorithm Capabilities

Class -- Missing class values, Binary class, Nominal class

Attributes -- Numeric attributes, Binary attributes, Empty nominal attributes, Unary attributes, Missing values, Nominal attributes

Additional -- min # of instances: 0

Gaussian Processes

Overview

This concept of our language implements Gaussian processes [71] for regression without hyperparameter-tuning. A Gaussian process uses lazy learning and a measure of the similarity between points (the *kernel function*) to predict the value for an unseen point from training data. The prediction is not just an estimate for that point, but also has uncertainty information—it is a one-dimensional Gaussian distribution (which is the marginal distribution at that point). To make choosing an appropriate noise level easier, this implementation applies normalization/standardization to the target attribute as well as the other attributes (if normalization/standardization is turned on). Missing values are replaced by the global mean/mode. Nominal attributes are converted to binary ones. Note that kernel caching is turned off if the kernel used implements CachedKernel.

Formally, Gaussian Processes algorithm executions are defined as instances of the class `GaussianProcesses` in OWL. The definition of this class and its data properties (configuration options) are listed in Appendix B (0) along with an example (B.1.3.23).

Algorithm configuration options

- seed -- The random number seed to be used.
- numDecimalPlaces -- The number of decimal places to be used for the output of numbers in the model.
- batchSize -- The preferred number of instances to process if batch prediction is being performed. More or fewer instances may be provided, but this gives implementations a chance to specify a preferred batch size.
- kernel -- The kernel to use.
- filterType -- Determines how/if the data will be transformed.
- noise -- The level of Gaussian Noise (added to the diagonal of the Covariance Matrix), after the target has been normalized/standardized/left unchanged).

Algorithm Capabilities

Class -- Date class, Numeric class, Missing class values

Attributes -- Binary attributes, Unary attributes, Numeric attributes, Nominal attributes, Empty nominal attributes, Missing values

Additional -- min # of instances: 1

Linear Regression

Overview

This concept of our language was introduced for using linear regression [72] for prediction. The algorithm uses the Akaike criterion for model selection, and is able to deal with weighted instances.

Formally, Linear Regression algorithm executions are defined as instances of the class `LinearRegression` in OWL. The definition of this class and its data properties (configuration options) are listed in Appendix B (B.1.1.28) along with an example (0).

Algorithm configuration options

- minimal -- If enabled, dataset header, means and stdevs get discarded to conserve memory; also, the model cannot be printed out.
- numDecimalPlaces -- The number of decimal places to be used for the output of numbers in the model.
- batchSize -- The preferred number of instances to process if batch prediction is being performed. More or fewer instances may be provided, but this gives implementations a chance to specify a preferred batch size.
- ridge -- The value of the Ridge parameter.
- attributeSelectionMethod -- Set the method used to select attributes for use in the linear regression. Available methods are: no attribute selection, attribute selection using M5's method (step through the attributes removing the one with the smallest standardised coefficient until no improvement is observed in the estimate of the error given by the Akaike information criterion), and a greedy selection using the Akaike information metric.
- outputAdditionalStats -- Output additional statistics (such as std deviation of coefficients and t-statistics)
- eliminateColinearAttributes -- Eliminate colinear attributes.

Algorithm Capabilities

Class -- Missing class values, Numeric class, Date class

Attributes -- Numeric attributes, Binary attributes, Empty nominal attributes, Unary attributes, Missing values, Nominal attributes

Additional -- min # of instances: 1

Multinomial Logistic Regression

Overview

This concept is a class for building and using a multinomial logistic regression model with a ridge estimator [73].

There are some modifications, however, compared to the paper of leCessie and van Houwelingen(1992):

If there are k classes for n instances with m attributes, the parameter matrix B to be calculated will be an m*(k-1) matrix.

The probability for class j with the exception of the last class is

Pj(Xi) = exp(XiBj)/((sum[j=1..(k-1)]exp(Xi*Bj))+1)

The last class has probability

1-(sum[j=1..(k-1)]Pj(Xi))
    = 1/((sum[j=1..(k-1)]exp(Xi*Bj))+1)

The (negative) multinomial log-likelihood is thus:

L = -sum[i=1..n]{
    sum[j=1..(k-1)](Yij * ln(Pj(Xi)))
    +(1 - (sum[j=1..(k-1)]Yij))
    * ln(1 - sum[j=1..(k-1)]Pj(Xi))
    } + ridge * (B^2)

In order to find the matrix B for which L is minimised, a Quasi-Newton Method is used to search for the optimized values of the m*(k-1) variables. Note that before we use the optimization procedure, we 'squeeze' the matrix B into a m*(k-1) vector. For details of the optimization procedure, please check weka.core.Optimization class.

Although original Logistic Regression does not deal with instance weights, we modify the algorithm a little bit to handle the instance weights.

Note: Missing values are replaced using a ReplaceMissingValuesFilter, and nominal attributes are transformed into numeric attributes using a NominalToBinaryFilter.

Formally, Multinomial Logistic Regression algorithm executions are defined as instances of the class `MultinomialLogisticRegression` in OWL. The definition of this class and its data properties (configuration options) are listed in Appendix B (B.1.1.29) along with an example (B.1.3.25).

Algorithm configuration options

- numDecimalPlaces -- The number of decimal places to be used for the output of numbers in the model.
- batchSize -- The preferred number of instances to process if batch prediction is being performed. More or fewer instances may be provided, but this gives implementations a chance to specify a preferred batch size.
- ridge -- Set the Ridge value in the log-likelihood.
- useConjugateGradientDescent -- Use conjugate gradient descent rather than BFGS updates; faster for problems with many parameters.
- maxIts -- Maximum number of iterations to perform.

Algorithm Capabilities

Class -- Missing class values, Binary class, Nominal class

Attributes -- Date attributes, Numeric attributes, Binary attributes, Empty nominal attributes, Unary attributes, Missing values, Nominal attributes

Additional -- min # of instances: 1

K-nearest neighbours (IBk)

Overview

This concept of our language constitutes the K-nearest neighbours classifier [74]. This class can select appropriate value of K based on cross-validation. It can also do distance weighting.

Formally, K-nearest neighbours algorithm executions are defined as instances of the class `IBk` in OWL. The definition of this class and its data properties (configuration options) are listed in Appendix B (B.1.1.30) along with an example (B.1.3.26).

Algorithm configuration options

- numDecimalPlaces -- The number of decimal places to be used for the output of numbers in the model.
- batchSize -- The preferred number of instances to process if batch prediction is being performed. More or fewer instances may be provided, but this gives implementations a chance to specify a preferred batch size.
- KNN -- The number of neighbours to use.
- distanceWeighting -- Gets the distance weighting method used.
- nearestNeighbourSearchAlgorithm -- The nearest neighbour search algorithm to use (Default: weka.core.neighboursearch.LinearNNSearch).
- windowSize -- Gets the maximum number of instances allowed in the training pool. The addition of new instances above this value will result in old instances

being removed. A value of 0 signifies no limit to the number of training instances.

- meanSquared -- Whether the mean squared error is used rather than mean absolute error when doing cross-validation for regression problems.
- crossValidate -- Whether hold-one-out cross-validation will be used to select the best k value between 1 and the value specified as the KNN parameter.

Algorithm Capabilities

Class -- Missing class values, Numeric class, Binary class, Date class, Nominal class

Attributes -- Date attributes, Numeric attributes, Binary attributes, Empty nominal attributes, Unary attributes, Missing values, Nominal attributes

Additional -- min # of instances: 0

Decision Table

Overview

This concept of our language constitutes the class for building and using a simple decision table majority classifier [75].

Formally, Decision Table algorithm executions are defined as instances of the class `DecisionTable` in OWL. The definition of this class and its data properties (configuration options) are listed in Appendix B (B.1.1.31) along with an example (B.1.3.27).

Algorithm configuration options

- numDecimalPlaces -- The number of decimal places to be used for the output of numbers in the model.
- batchSize -- The preferred number of instances to process if batch prediction is being performed. More or fewer instances may be provided, but this gives implementations a chance to specify a preferred batch size.
- evaluationMeasure -- The measure used to evaluate the performance of attribute combinations used in the decision table.
- search -- The search method used to find good attribute combinations for the decision table.
- displayRules -- Sets whether rules are to be printed.
- useIBk -- Sets whether IBk should be used instead of the majority class.
- crossVal -- Sets the number of folds for cross validation (1 = leave one out).

Algorithm Capabilities

Class -- Nominal class, Binary class, Numeric class, Missing class values, Date class

Attributes -- Binary attributes, Date attributes, Nominal attributes, Unary attributes, Missing values, Numeric attributes, Empty nominal attributes

Additional -- min # of instances: 1

Zero R

Overview

This concept is the class for building and using a 0-R classifier [72]. It predicts the mean (for a numeric class) or the mode (for a nominal class).

Formally, Zero R algorithm executions are defined as instances of the class `ZeroR` in OWL. The definition of this class and its data properties (configuration options) are listed in Appendix B (B.1.1.32) along with an example (B.1.3.28).

Algorithm configuration options

- numDecimalPlaces -- The number of decimal places to be used for the output of numbers in the model.
- batchSize -- The preferred number of instances to process if batch prediction is being performed. More or fewer instances may be provided, but this gives implementations a chance to specify a preferred batch size.

Algorithm Capabilities

Class -- Nominal class, Binary class, Numeric class, Missing class values, Date class

Attributes -- Binary attributes, Relational attributes, String attributes, Date attributes, Nominal attributes, Unary attributes, Missing values, Numeric attributes, Empty nominal attributes

Additional -- min # of instances: 0

J48

Overview

This language concept is the class for generating a pruned or unpruned C4.5 decision tree [76].

Formally, J48 algorithm executions are defined as instances of the class `J48` in OWL. The definition of this class and its data properties (configuration options) are listed in Appendix B (B.1.1.33) along with an example (B.1.3.29).

Algorithm configuration options

- seed -- The seed used for randomizing the data when reduced-error pruning is used.
- unpruned -- Whether pruning is performed.
- confidenceFactor -- The confidence factor used for pruning (smaller values incur more pruning).
- numFolds -- Determines the amount of data used for reduced-error pruning. One fold is used for pruning, the rest for growing the tree.
- numDecimalPlaces -- The number of decimal places to be used for the output of numbers in the model.
- batchSize -- The preferred number of instances to process if batch prediction is being performed. More or fewer instances may be provided, but this gives implementations a chance to specify a preferred batch size.
- reducedErrorPruning -- Whether reduced-error pruning is used instead of C.4.5 pruning.
- useLaplace -- Whether counts at leaves are smoothed based on Laplace.
- doNotMakeSplitPointActualValue -- If true, the split point is not relocated to an actual data value. This can yield substantial speed-ups for large datasets with numeric attributes.
- subtreeRaising -- Whether to consider the subtree raising operation when pruning.
- saveInstanceData -- Whether to save the training data for visualization.
- binarySplits -- Whether to use binary splits on nominal attributes when building the trees.
- minNumObj -- The minimum number of instances per leaf.
- useMDLcorrection -- Whether MDL correction is used when finding splits on numeric attributes.
- collapseTree -- Whether parts are removed that do not reduce training error.

Algorithm Capabilities

Class -- Nominal class, Binary class, Missing class values

Attributes -- Binary attributes, Date attributes, Nominal attributes, Unary attributes, Missing values, Numeric attributes, Empty nominal attributes

Additional -- min # of instances: 0

Random Forest

Overview

This concept of our language constitutes the class for constructing a forest of random trees [77].

Formally, Random Forest algorithm executions are defined as instances of the class `RandomForest` in OWL. The definition of this class and its data properties (configuration options) are listed in Appendix B (B.1.1.34) along with an example (B.1.3.30).

Algorithm configuration options

- seed -- The random number seed to be used.
- representCopiesUsingWeights -- Whether to represent copies of instances using weights rather than explicitly.
- storeOutOfBagPredictions -- Whether to store the out-of-bag predictions.
- numExecutionSlots -- The number of execution slots (threads) to use for constructing the ensemble.
- bagSizePercent -- Size of each bag, as a percentage of the training set size.
- numDecimalPlaces -- The number of decimal places to be used for the output of numbers in the model.
- batchSize -- The preferred number of instances to process if batch prediction is being performed. More or fewer instances may be provided, but this gives implementations a chance to specify a preferred batch size.
- printClassifiers -- Print the individual classifiers in the output
- numIterations -- The number of iterations to be performed.
- outputOutOfBagComplexityStatistics -- Whether to output complexity-based statistics when out-of-bag evaluation is performed.
- classifier -- The base classifier to be used.
- breakTiesRandomly -- Break ties randomly when several attributes look equally good.
- maxDepth -- The maximum depth of the tree, 0 for unlimited.
- computeAttributeImportance -- Compute attribute importance via mean impurity decrease
- calcOutOfBag -- Whether the out-of-bag error is calculated.
- numFeatures -- Sets the number of randomly chosen attributes. If 0, int(log_2(#predictors) + 1) is used.

Algorithm Capabilities

Class -- Nominal class, Binary class, Numeric class, Missing class values

Attributes -- Binary attributes, Date attributes, Nominal attributes, Unary attributes, Missing values, Numeric attributes, Empty nominal attributes

Additional -- min # of instances: 1

Random Tree

Overview

This language concept implements the class for constructing a tree that considers K randomly chosen attributes at each node [72]. It performs no pruning. It also has an option to allow estimation of class probabilities (or target mean in the regression case) based on a hold-out set (backfitting).

Formally, Random Tree algorithm executions are defined as instances of the class `RandomTree` in OWL. The definition of this class and its data properties (configuration options) are listed in Appendix B (B.1.1.35) along with an example (B.1.3.31).

Algorithm configuration options

- seed -- The random number seed used for selecting attributes.
- allowUnclassifiedInstances -- Whether to allow unclassified instances.
- minNum -- The minimum total weight of the instances in a leaf.
- numFolds -- Determines the amount of data used for backfitting. One fold is used for backfitting, the rest for growing the tree. (Default: 0, no backfitting)
- numDecimalPlaces -- The number of decimal places to be used for the output of numbers in the model.
- batchSize -- The preferred number of instances to process if batch prediction is being performed. More or fewer instances may be provided, but this gives implementations a chance to specify a preferred batch size.
- breakTiesRandomly -- Break ties randomly when several attributes look equally good.
- maxDepth -- The maximum depth of the tree, 0 for unlimited.
- minVarianceProp -- The minimum proportion of the variance on all the data that needs to be present at a node in order for splitting to be performed in regression trees.
- KValue -- Sets the number of randomly chosen attributes. If 0, int(log_2(#predictors) + 1) is used.

Algorithm Capabilities

Class -- Nominal class, Binary class, Numeric class, Missing class values

Attributes -- Binary attributes, Date attributes, Nominal attributes, Unary attributes, Missing values, Numeric attributes, Empty nominal attributes

Additional -- min # of instances: 1

Unsupervised Data Mining Algorithm

Canopy

Overview

This concept of our language is used to represent the algorithm that clusters data using the canopy clustering algorithm [78], which requires just one pass over the data. It can run in either batch or incremental mode. Results are generally not as good when running incrementally as the min/max for each numeric attribute is not known in advance. Has a heuristic (based on attribute std. deviations), that can be used in batch mode, for setting the T2 distance. The T2 distance determines how many canopies (clusters) are formed. When the user specifies a specific number (N) of clusters to generate, the algorithm will return the top N canopies (as determined by T2 density) when N < number of canopies (this applies to both batch and incremental learning); when N > number of canopies, the difference is made up by selecting training instances randomly (this can only be done when batch training).

Formally, Canopy algorithm executions are defined as instances of the class `Canopy` in OWL. The definition of this class and its data properties (configuration options) are listed in Appendix B (B.1.1.36) along with an example (B.1.3.32).

Algorithm configuration options

- seed -- The random number seed to be used.
- dontReplaceMissingValues -- Replace missing values globally with mean/mode.
- t2 -- The T2 distance to use. Values < 0 indicate that this should be set using a heuristic based on attribute standard deviation (note that this only works when batch training)
- t1 -- The T1 distance to use. Values < 0 are taken as a positive multiplier for the T2 distance
- numClusters -- Set number of clusters. -1 means number of clusters is determined by T2 distance
- minimumCanopyDensity -- The minimum T2-based density below which a canopy will be pruned during periodic pruning
- periodicPruningRate -- How often to prune low density canopies during training

Algorithm Capabilities

Attributes -- Binary attributes, Nominal attributes, Unary attributes, Missing values, Numeric attributes, Empty nominal attributes

Additional -- min # of instances: 0

Cobweb and Classit

Overview

This language concept constitutes the class implementing the Cobweb [79] and Classit [80] clustering algorithms.

Note: the application of node operators (merging, splitting etc.) in terms of ordering and priority differs (and is somewhat ambiguous) between the original Cobweb and Classit papers. This algorithm always compares the best host, adding a new leaf, merging the two best hosts, and splitting the best host when considering where to place a new instance.

Formally, Coweb-Classit algorithm executions are defined as instances of the class `CowebClassit` in OWL. The definition of this class and its data properties (configuration options) are listed in Appendix B (B.1.1.37) along with an example (B.1.3.33).

Algorithm configuration options

- seed -- The random number seed to be used. Use -1 for no randomization.
- saveInstanceData -- save instance information for visualization purposes
- acuity -- set the minimum standard deviation for numeric attributes
- cutoff -- set the category utility threshold by which to prune nodes

Algorithm Capabilities

Attributes -- Binary attributes, Date attributes, Nominal attributes, Unary attributes, Missing values, Numeric attributes, Empty nominal attributes

Additional -- min # of instances: 0

EM

Overview

This concept of our language is the simple EM (expectation maximisation) [72] algorithm implementation.

EM assigns a probability distribution to each instance which indicates the probability of it belonging to each of the clusters. EM can decide how many clusters to create by cross validation, or you may specify apriori how many clusters to generate.

The cross validation performed to determine the number of clusters is done in the following steps:

1. the number of clusters is set to 1
2. the training set is split randomly into 10 folds.
3. EM is performed 10 times using the 10 folds the usual CV way.
4. the loglikelihood is averaged over all 10 results.
5. if loglikelihood has increased the number of clusters is increased by 1 and the program continues at step 2.

The number of folds is fixed to 10, as long as the number of instances in the training set is not smaller 10. If this is the case the number of folds is set equal to the number of instances.

Missing values are globally replaced with ReplaceMissingValues.

Formally, Expectation Maximisation (EM) algorithm executions are defined as instances of the class `EM` in OWL. The definition of this class and its data properties (configuration options) are listed in Appendix B (B.1.1.38) along with an example (B.1.3.34).

Algorithm configuration options

- seed -- The random number seed to be used.
- numFolds -- The number of folds to use when cross-validating to find the best number of clusters (default = 10)
- numExecutionSlots -- The number of execution slots (threads) to use. Set equal to the number of available CPU/cores
- numKMeansRuns -- The number of runs of k-means to perform.
- displayModelInOldFormat -- Use old format for model output. The old format is better when there are many clusters. The new format is better when there are fewer clusters and many attributes.
- minLogLikelihoodImprovementIterating -- The minimum improvement in log likelihood required to perform another iteration of the E and M steps
- minLogLikelihoodImprovementCV -- The minimum improvement in cross-validated log likelihood required in order to consider increasing the number of clusters when cross-validating to find the best number of clusters
- maximumNumberOfClusters -- The maximum number of clusters to consider during cross-validation to select the best number of clusters
- numClusters -- set number of clusters. -1 to select number of clusters automatically by cross validation.
- maxIterations -- maximum number of iterations

- minStdDev -- set minimum allowable standard deviation

Algorithm Capabilities

Attributes -- Binary attributes, Nominal attributes, Unary attributes, Missing values, Numeric attributes, Empty nominal attributes

Additional -- min # of instances: 1

Farthest First

Overview

This language concept is the algorithm implementation that clusters data using the Farthest First [81], [82] algorithm.

Notes:

- works as a fast simple approximate clusterer

- modelled after SimpleKMeans, might be a useful initializer for it

Formally, Farthest First algorithm executions are defined as instances of the class `FarthestFirst` in OWL. The definition of this class and its data properties (configuration options) are listed in Appendix B (B.1.1.39) along with an example (B.1.3.35).

Algorithm configuration options

- seed -- The random number seed to be used.
- numClusters -- set number of clusters

Algorithm Capabilities

Attributes -- Binary attributes, Date attributes, Nominal attributes, Unary attributes, Missing values, Numeric attributes, Empty nominal attributes

Additional -- min # of instances: 1

Simple K Means

Overview

This language concept is the algorithm implementation that clusters data using the k means [83] algorithm. It can use either the Euclidean distance (default) or the Manhattan distance. If the Manhattan distance is used, then centroids are computed as the component-wise median rather than mean.

Formally, Simple K Means algorithm executions are defined as instances of the class `SimpleKMeans` in OWL. The definition of this class and its data properties (configuration options) are listed in Appendix B (B.1.1.40) along with an example (B.1.3.36).

Algorithm configuration options

seed -- The random number seed to be used.

displayStdDevs -- Display std deviations of numeric attributes and counts of nominal attributes.

numExecutionSlots -- The number of execution slots (threads) to use. Set equal to the number of available CPU/cores

dontReplaceMissingValues -- Replace missing values globally with mean/mode.

canopyMinimumCanopyDensity -- If using canopy clustering for initialization and/or speedup this is the minimum T2-based density below which a canopy will be pruned during periodic pruning

canopyT2 -- The T2 distance to use when using canopy clustering. Values < 0 indicate that this should be set using a heuristic based on attribute standard deviation

numClusters -- set number of clusters

maxIterations -- set maximum number of iterations

preserveInstancesOrder -- Preserve order of instances.

canopyPeriodicPruningRate -- If using canopy clustering for initialization and/or speedup this is how often to prune low density canopies during training

canopyMaxNumCanopiesToHoldInMemory -- If using canopy clustering for initialization and/or speedup this is the maximum number of candidate canopies to retain in main memory during training of the canopy clusterer. T2 distance and data characteristics determine how many candidate canopies are formed before periodic and final pruning are performed. There may not be enough memory available if T2 is set too low.

initializationMethod -- The initialization method to use. Random, k-means++, Canopy or farthest first

distanceFunction -- The distance function to use for instances comparison (default: weka.core.EuclideanDistance).

canopyT1 -- The T1 distance to use when using canopy clustering. Values < 0 are taken as a positive multiplier for the T2 distance

fastDistanceCalc -- Uses cut-off values for speeding up distance calculation, but suppresses also the calculation and output of the within cluster sum of squared errors/sum of distances.

reduceNumberOfDistanceCalcsViaCanopies -- Use canopy clustering to reduce the number of distance calculations performed by k-means

Algorithm Capabilities

Attributes -- Binary attributes, Nominal attributes, Unary attributes, Missing values, Numeric attributes, Empty nominal attributes

Additional -- min # of instances: 1

Chapter 4

Platform Implementation

4.1. Introduction

This chapter describes the prototype implementation of the approach to evidence based public health policy making that is proposed by this thesis, and the technologies used in this implementation. To do so, the chapter first introduces the overall architecture of the prototype, then it reviews the main background technologies used in the implementation, and finally it describes in detail the implementation of the components (tools) of the prototype and how data analytics workflows can be executed.

4.2. Platform Architecture

In Figure 30 we present the platform architecture, our components and their interactions.

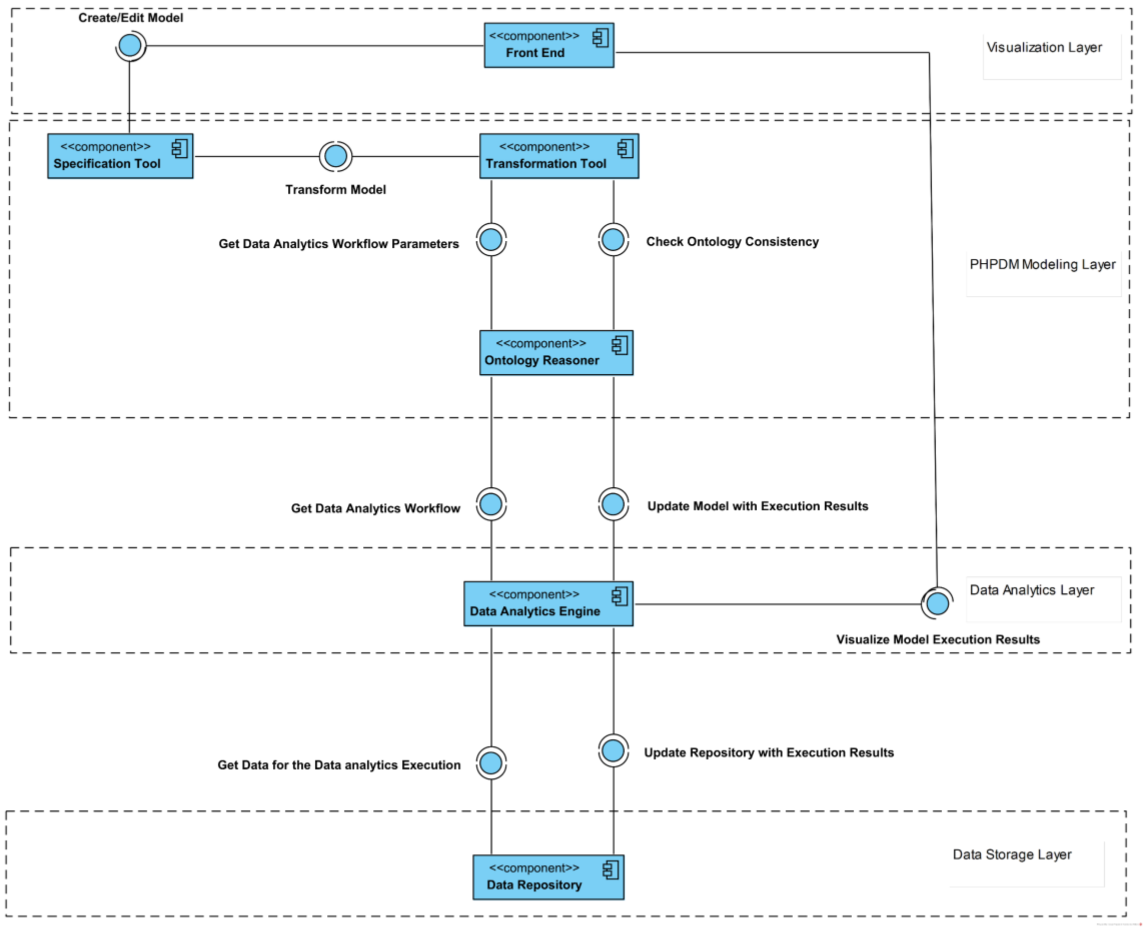


Figure 30 Platform Architecture

The architecture is structured in four layers:

- *Visualization Layer*: This is the front-end layer that is responsible for the creation and editing of the models and the visualization of the execution results. For the purposes of the prototype of this thesis, we used the front-end of the Specification Tool of EVOTION.
- *Data Storage Layer*: This is the layer that is responsible for storing the data collected with the purpose of supporting the analytic processes producing the evidence to be used for public health decision making. It is also the layer, which stores the data generated as a result of executing such processes (e.g., the outcomes of specific analytic tasks that part of a workflow).
- *Data Analytics Layer:* This layer is responsible for (i) ingesting data from Data Storage Layer, and (ii) data processing. The processing is mainly mediated by the PHPD Modeling layer so that all the processing activities are expressed in terms of models and using the modeling language. The principal actor of the data processing is the Data Analytics Engine, which has responsibility for executing the data analytics workflows of a PHPDM model.
- *PHPDM Modeling Layer:* This layer supports the specification of PHPDM models and the transformation of these models into a form that can be directly executed by the data analytics engine. This layer interacts with the Data Analytics Layer for (i) providing an executable data analytics object to the BDA Engine, and (ii) checking the consistency of models from the Ontology Reasoner.

The main interactions between the components in the architecture shown in Figure 30 are summarised below:

- The *front end* interacts with the specification tool for the creation and editing of the PHPDM models It also interacts with the data analytics engine for the visualization of the execution results.
- The *transformation tool* interacts with the *ontology reasoner* to check if the ontology model instance, which expresses the PHPDM model, is consistent with regards to the PHPDM models' specification language introduced in

Chapter 3. The transformation tool also gets the required information for the data analytics workflow from the ontology reasoner.

- After generating an executable data analytics object from a PHPDM model, the transformation tool passes it to the data analytics engine to be executed.
- The data analytics engine interacts with the data repository to get the data required to run the data analytics tasks. It also saves into the repository all the intermediate and final results produced by the execution of the data analytics script in order to provide a full trace and auditability of the analytics process.

4.3. Background Technologies and Tools

4.3.1. RESTful web Services

The programmatic interfaces offered by the different components in the architecture shown in Figure 30 have been realised through RESTful web services [84]. Representational State Transfer (REST) is an architectural style that defines a set of constraints to be used for creating web services. Web services that conform to the REST architectural style, or RESTful web services, provide interoperability between computer systems on the Internet. REST-compliant web services allow the requesting systems to access and manipulate textual representations of web resources by using a uniform and predefined set of stateless operations [85].

We will give more details about the web services we used in the developed tools sections below.

4.3.2. Protégé

Protégé[86] is a free, open-source ontology OWL editor and framework for building intelligent systems. Protégé is supported by a strong community of academic, government, and corporate users, who use Protégé to build knowledge-based solutions in areas as diverse as biomedicine, e-commerce, and organizational modeling. It fully supports the latest OWL 2 Web Ontology Language and RDF specifications from the World Wide Web Consortium. It is based on JAVA, is extensible, and provides a plug-and-play environment that makes it a flexible base for rapid prototyping and application development.

We used Protégé for the development of our ontology. We created all the classes, the object properties, as well as the other axioms using Protégé. We also used Protégé for the instantiation of our use case, namely the creation of all the instances and the data properties of our use case.

4.3.3. Weka

Weka [57], [87] is a collection of machine learning algorithms for data mining tasks. The algorithms can either be applied directly to a dataset or called from your own JAVA code. Weka contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization. It is also well-suited for developing new machine learning schemes.

We used the DistributedWekaSpark package by Mark Hall [58] for the runs of the data mining tasks of our use case. This package enabled us to use Spark, which we describe below, to run our data mining task in distributed local mode using all the cores of our machine.

4.3.4. Spark

Apache Spark [56] is a fast and general engine for large-scale data processing. Apache Spark has an advanced execution engine that supports acyclic data flow and in-memory computing.

Spark offers over 80 high-level operators that make it easy to build parallel apps and offers a JAVA API. Spark incorporates a stack of libraries including SQL and Data Frames, MLlib[55] for machine learning, GraphX, and Spark Streaming. You can combine these libraries seamlessly in the same application. Spark is interoperable with BDA technologies offering access to data in HDFS[88], HBase[89] and any Hadoop data source.

We used Spark via Weka in order to run our data mining tasks in distributed local mode using all the cores of our machine, as we mentioned above in section 4.3.3.

4.3.5. Evotion Data Repository (EDR)

In this section we describe the Evotion Data Repository and more specifically, the types of data held in it, how they are acquired, the technologies used to build the repository and the hospital systems connected to it.

The Evotion Data Repository receives hearing aid usage, biosensor and mobile application data from the Evotion mobile application. It also receives data collected through the filling of different medical questionnaires via the EVOTION dashboard, and daily clinical data collected from the AuditBase system that operates in the UK hospitals and the EVOTION Hospital System that is being developed for the Greek hospitals. More specifically, the questionnaires filled are the following:

- the Montreal Cognitive Assessment (MoCA) [90]
- the Health Utilities Index (HUI) – 3 [91]
- the Glasgow Hearing Aid Benefit Profile (GHAP) [92]
- the Hospital Anxiety and Depression Scale (HADS) [93]

Internally, EDR is implemented using the NoSQL database management system HBase [94] and Phoenix [95], i.e., a middleware that provides access to HBase in an SQL-like manner. The access to the interfaces implemented by EDR is controlled by EVOTION REST Service Layer, which functions as the security mechanism for EDR and acts as a middleware that provides authorised access to all the EVOTION data repository operations, which are provided by EDR.

The data acquisition sources connected to EDR include the following:

- EVOTION Dashboard

  The EVOTION Dashboard (ED) serves a dual purpose. Firstly, it acts as the front end of the EDR enabling the selection of data from different medical questionnaires which have been selected by the clinical partners of EVOTION. Secondly, it helps EVOTION registered end-users to interact with the EVOTION data repository offering access to the data stored in it.

- EVOTION Mobile App

  The EVOTION Mobile Application is a component aimed at providing a user-friendly graphical user interface for patients recruited by EVOTION to access EVOTION's utilities and functionalities. At the same time, the component is handling the communication of data from the peripheral devices (i.e. the wearable sensor device and the hearing aids) to the mobile phone, and from the mobile phone to EDR.

  Part of the EVOTION mobile application’s functionality is going to store sensitive data. All information that is collected by the peripheral devices will be stored by the EVOTION mobile application, before being transferred to

EDR. Periodically, the mobile application transmits to EDR all data collected by the peripheral devices and locally stored.

- Crystal Reporting

  The purpose of Crystal Reporting (CR) is to extract clinical data collected and stored in AuditBase [96], i.e., the system used in the UK hospital as part of the clinical HA fitting, and transmit them to EDR. The report contains all the information a clinician collects after the examination of an EVOTION patient. Crystal Reporting will transmit data of Crystal Reports (for both retrospective and prospective data) to EDR. To do so, it uses an SQL command to extract the data from the AuditBase database. The extracted data is in XML format. These data are initially anonymised by CR using the data hiding and transformation mechanisms and are subsequently encrypted and sent to the EVOTION data repository in JSON format.

- EVOTION Hospital System (EHS)

  EHS is the system that was developed by EVOTION to provide the functionality of the AuditBase system to the Greek hospitals of the EVOTION consortium. EHS did not replicate the full functionality of AuditBase, as doing so was not be feasible given the resources of EVOTION. It only implemented the subset of AuditBase functionality and the relevant data stores that were necessary for collecting core clinical data related to hearing aid fitting, supporting the process of hearing aid fitting and storing the collected data into a local hospital database. EHS also transmits the collected data to EDR following anonymisation and encryption. The latter functions are like those implemented by CR. EHS has a web-accessible front end (dashboard) which supports:

  - Secure registration of end-users (clinicians)
  - Secure login of end-users (clinicians) giving them access to supported functionality via Role–based access control (RBAC)
  - The administration and collection of medical patient data during the clinical pathway of an EVOTION patient that involves :
  - Administration of real patients and their basic information (create/search/edit/delete)
  - Registration of patient' devices

- Administration (per patient) of: – otoscopy and audiometry records
  – hearing history and diagnosis category records
  – tinnitus records
  – lifestyle and communication records
  – outcome information records

It also provides its own local hospital database that is realised using MySQL [97] and communicates with NOAH [98], which is the software used for the actual software fitting, to enable the storage of data collected by NOAH into the local database that it maintains.

The hospital systems connected to the EDR are listed below:

- Athens Medical Center SA
- Otolaryngology Clinic, Ippokration Hospital, University of Athens
- Ear Institute, University College London
- Guy's and St Thomas NHS Foundation Trust, UK
- Otticon A/S

4.4. Developed Tools

4.4.1. Ontology Reasoner

The ontology reasoner tool is responsible for checking the consistency of the ontology model instances, instantiating the ontology model and querying the ontology in order to give the requested information to the transformation tool. The reasoner we used was Hemit Reasoner [99] version 1.3.8.

4.4.1.1. Functional Capabilities

This component is used in the prototype to (A) load the ontology file, (B) check its consistency, (C) get all the data analytics workflow information useful for the transformation of the PHPDM model (i.e. the data analytics tasks and the datasets) and (D) send it as a JAVA object to the transformation tool. A and B are useful, because it is important to check a PHPDM’s consistency before sending the workflow for execution and C and D are the main functional capabilities of this component.

4.4.2. Transformation Tool

The transformation tool is at the heart of the prototype implementation as it takes a PHPDM model specified as an ontology model instance from the ontology reasoner and transforms it to a data analytics script that can be directly executed by the data analytics engine (i.e., the Spark implementation of Weka provided by *DistributedWekaSpark*).

4.4.2.1. Functional Capabilities

The sole functional capability of the Transformation tool is to take a PHPDM model specified as an ontology model instance from the ontology reasoner and transforms it to a data analytics script that can be directly executed by the data analytics engine. This transformation is based on the algorithm *computeExecutionSteps* shown in Table 4-1. This algorithm takes as input the DataAnalyticsWorkflow(namely the data analytics tasks, their input and output datasets, the data specifications and the algorithms of the tasks). The transformation tool defines the order of execution of the tasks, depending on their input and output datasets and uses the methods exposed by the data analytics engine to execute the tasks. The tasks are broken down in execution steps. More specifically, the algorithm works as follows. Initially it puts all the output datasets into a set (lines 3-6 of Table 4-1). Then, while the set of output datasets is not empty, it puts all the tasks that their input datasets are not in the output datasets set in the first execution step (lines 9-13 of Table 4-1). After that, for each data analytics task in the current step tasks, if the output dataset is in the output datasets set, it removes it and it also removes the task from the workflow tasks (lines 15-20 of Table 4-1). It does this recursively until all the output datasets have been removed. Finally, it returns all the data analytics tasks broken down in execution steps in a list of sets. In section 4.5 we describe the data analytics workflow execution summary.

Table 4-1 Transformation Algorithm

```
1    function computeExecutionSteps (input: DataAnalyticsWorkflow daw)
2    output: List<Set> executionSteps{
3       Set outputDatasets;
4       for (DataAnalyticsTask task in daw.tasks){
5           outputDatasets.add(task.outputDataset);
6       }
7       While (not outputDatasets.isEmpty){
```

```
        Set stepTasks;
        for(DataAnalyticsTask task in daw.tasks){
            if(task.iputDatasets not in outputDatasets){
                stepTasks.add(task);
            }
        }
        executionSteps.add(stepTasks);
        for (DataAnalyticsTask stepTask in stepTasks){
            if(stepTask.outputDatasets in outputDatasets){
                outputTasks.remove(stepTask.outputDatasets);
                daw.tasks.remove(stepTask);
            }
        }
    }
    return executionSteps;
}
```

4.4.3. Data Analytics Engine

The data analytics engine is responsible for getting the data analytics workflow from the transformation tool and executing the appropriate data analytics, after getting the relevant data for the data repository (synthetic data for the purposes of our evaluation and experimentation) and persisting the results back to the repository.

4.4.3.1. Functional Capabilities

The Data Analytic Engine principally addresses the functionalities required for the processing of analytics and giving back the processing results.

More specifically, it has the following functional capabilities:

- **prepareDataset:** The purpose of this operation is to prepare the required data from the repository in CSV format to be readable for the data mining processes of distributedWEKA we use.
- **performClassification:** The purpose of this operation is to perform the actual classification task and give the execution results.
- **evaluateClassification:** The purpose of this operation is to perform 10-fold cross validation to the classification previously executed and give the evaluation results.

• **performClustering:** The purpose of this operation is to perform the actual clustering task and give the execution results.

• **performStatisticalAnalysis:** The purpose of this operation is to perform the actual statistical task and give the execution results.

4.4.3.2. Execution Algorithm

The data analytics engine takes as input the data analytics tasks to be executed divided in execution steps and executes them. The execution algorithm that is used by the engine is called executeTask and is specified in Table 4-2 below. More specifically, the algorithm works as follows. For each execution step and for each task in the execution step in executes each task asynchronously.

Table 4-2 Execution Algorithm

```
function executeTasks (input:  List<Set> executionSteps) {
   for (Set executionStep:executionSteps){
       for(DataAnalyticsTask task:executionStep){
           execute asynchronously (using multiple threads){
           if(task isinstance DataProcessingTask){
               prepareDataset(task);
           }
           if(task isinstance StatisticalAnalysisTask){
               performStatisticalAnalysis(task);
           }
           if(task isinstance DataMiningTask){
               if(task.algorithm isinstance
                  SupervisedDataMiningAlgorithm){
                   performClassification(task);
               }
               if(task.algorithm isinstance
                  UnSupervisedDataMiningAlgorithm){
                   performClustering(task);
               }
            }
        }
   }
}
```

4.4.4. *Specification Tool*

For the purposes of this prototype we did not develop a specification tool (for the instantiation of our ontology model). Instead of building a tool, we used Protégé to instantiate our ontology model with the required information for our use case scenario presented in Chapter 5 . For the purposes of the evaluation of our platform prototype, we used the specification tool and the front end of EVOTION.

The PHPDM Specification Tool (PHPDM e-service) is the component that allows end-users of the EVOTION platform (mainly policy makers) to administer decision models and their execution via a web service. This component assists them in defining suitable instances of PHPDM models, in accordance to a predefined template of a particular model. Appropriate functions guide the end-user in defining those PHPDM models by dynamically adapting the possible choices (e.g., of input datasets and parameters, of method to be applied upon them, of thresholds or other execution criteria to be fulfilled) logically defined by the ontology.

4.4.4.1. *Scenarios of use of PHPDM e-service*

As described in the aforementioned high-level description of top-level classes and relationships, the PHPDM e-service supports independent creation of policies, workflows and criteria. A data analyst can create workflows to be used by policy experts to define criteria and policies. Below we present a scenario of the creation of a data analytics workflow by a data analyst and the creation of a policy by a policy by a policy expert, defining criteria based on the previously created workflow.

Creation of a Data Analytics Workflow

In this scenario we assume that a data analyst wants to create a data analytics workflow for the policy expert to use in order to define criteria and create a policy. In order to do that they follow the procedure described below:

1. They go to the Policies tab and push the button to create a Workflow.
2. They name the new workflow “Linear Regression on Average Daily Usage Workflow” and select as execution type: On user action.
3. In the new data analytics workflow, they create a new data analytics task. They select as type Statistical analysis, as Input dataset the age, occupation, educational level, working in noise, working in groups and the computed average daily usage. As Method they select Linear Regression.

4. Next, they choose as dependent variable the computed average daily usage and push the Create button to create the task.

The data analytics workflow is created and ready to be used to define criteria to a policy.

Creation of a Policy

In this scenario we assume that a policy expert wants to create a policy by defining criteria using a previously created data analytics workflow. In order to do that they follow the procedure described below:

1. They go to the Policies tab and push the button to create a Policy.
2. They name the new policy "Policy model of addressing barriers to hearing aid use", input the Goal description as "Addressing barriers to hearing aid use" and Rationale as "Barriers to hearing aid use are a significant public health problem. Barriers occur at all levels of the process of provision of hearing aids including at the level of the HA user. The big data gathered about users through EVOTION would enable policy makers to choose which barriers to address in a population, in order to improve hearing aid use and hence reduce the burden of hearing loss in that population." They select as execution type: On user action and create thee objectives:
    - One with Description "Explore whether the occupation of HA users affects their daily usage" and Rationale "The occupation of HA users could affect their daily usage. If so, we should take the appropriate policy actions targeted to specific occupations."
    - Another with Description "Explore whether the educational level of HA users affects their daily usage" and Rationale "The educational level of HA users could affect their daily usage. If so, we should take the appropriate policy actions targeted to users with specific educational levels."
    - And a third with Description "Explore whether the age of HA users affects their daily usage" and Rationale "The age of HA users could affect their daily usage. If so, we should take the appropriate policy actions targeted to specific age groups."
3. Then, the policy expert defines the policy actions related to each objective. They define one policy action for each objective:
    - For Objective 1 they create a policy action named: "Occupation related action" and select as workflow the previously created workflow "Linear Regression on Average Daily Usage Workflow".
    - For Objective 2 they create a policy action named: "Educational level related action" and select as workflow the same previously created workflow "Linear Regression on Average Daily Usage Workflow".

- For Objective 3 they create a policy action named: “Age related action” and select as workflow the same previously created workflow “Linear Regression on Average Daily Usage Workflow”.

4. Next, they create a criterion for each policy action they have created by pushing the Create criterion button.
    - They create a criterion that is associated with “Occupation related action” Policy action. As weight they do not input anything, as all criteria have the same weight. They create the following criterion regular expression: “Output-LR01.R_square > 0.5 AND Output-LR01.Occupation_P < 0.05”.
    - They create another criterion that is associated with “Educational level related action” Policy action. Again, as weight they do not input anything, as all criteria have the same weight. They create the following criterion regular expression: “Output-LR01.R_square > 0.5 AND Output-LR01.EducationalLevel_P < 0.05”.
    - They create a third criterion that is associated with “Age related action” Policy action. Again, as weight they do not input anything, as all criteria have the same weight. They create the following criterion regular expression: “Output-LR01.R_square > 0.5 AND Output-LR01.Age_P < 0.05”.
5. Finally, they push the Validate button to the policy action to be created, validated and be ready for execution.

After following the above described steps, the policy model created and ready for execution.

4.4.4.2. Web user interface of the Specification Tool

This section describes the implemented features of the initial version (ver 1.94) of the PHPDM E-service front- end, thus the presentation elements implemented according to user input and requirements that have been elicited specifically in the context of the project, and the resulted design (descripted in the previous section). Current version was made available on November 20th 2018, therefore an upcoming evaluation will provide detailed comments to be considered and further enhance the final outcome.

Subsequent sections present basic functionality offered to its end-users. As for the implemented services, for each one of them a screen accompanied with a description is given.

Login

The login page for the EVOTION Dashboard is available at: https://evotion.city.ac.uk/login.php

Upon browsing to this location, a login screen (webpage) is presented. Figure 31 depicts the appearance of the login service where end-users can enter their credentials in order to access the system. PHP htmlspecialchars() build-in function is utilised for sanitizing special characters out of user input. The REST service communicates with the LDAP to verify whether the end-user already exists, and if it does, whether the password is correct or not. If credentials used are the correct ones, a valid TOKEN is generated and used throughout the end-user's session until the web browser is closed or he/she chooses to log out.

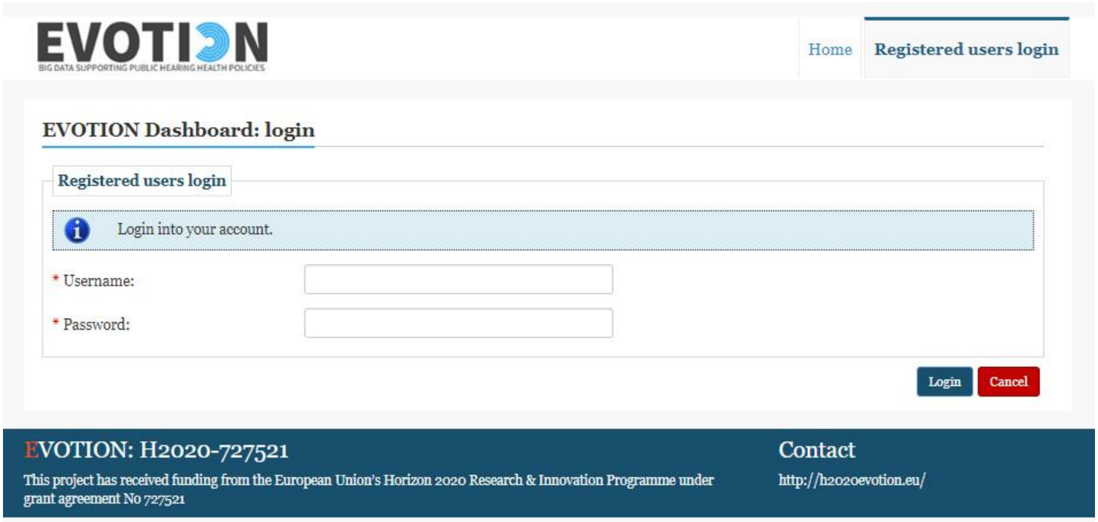


Figure 31: End-user's log in

Policies and Workflows

All previously created policies and workflows presented in Zebra striping tabular form data tables (Figure 32). Selecting the desired link ID from the list will automatically make it the active selection (object's specific information to be displayed). In addition, end-user may initiate the creation of a policy or a workflow.

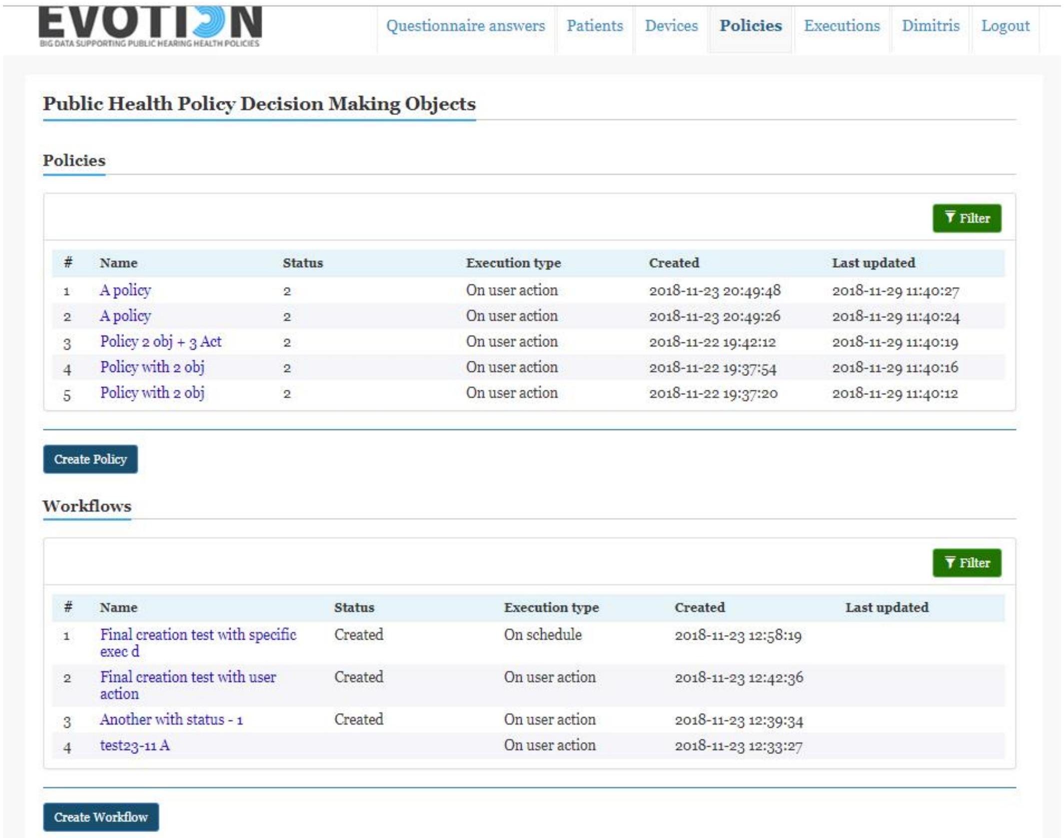


Figure 32: All available policies and workflows presented in Zebra striping tabular form data tables

Workflow creation

To create a new workflow, end-user fills in a name, and corresponding execution type (required fields). As with all forms, after finishing inserting data, he/she clicks on Create button to create the entry, or Cancel to leave without creating the entry (Figure 33).

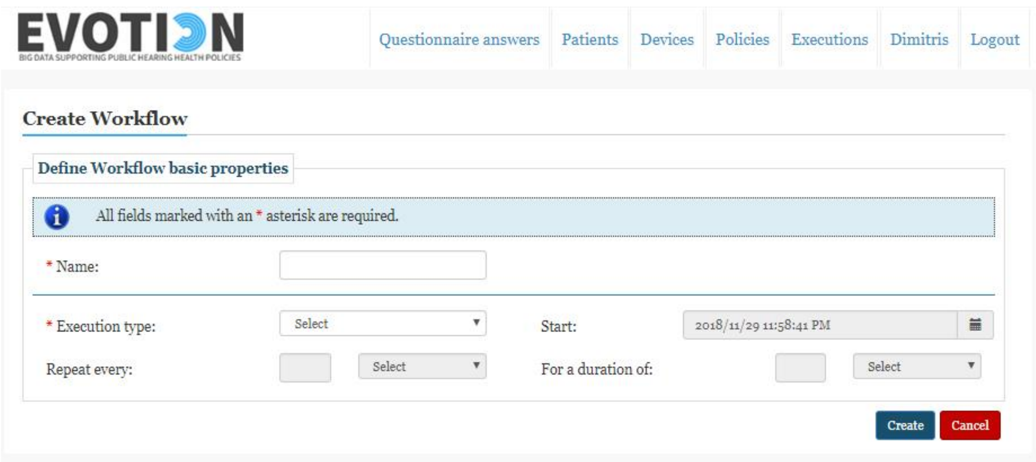
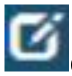


Figure 33: Create a new workflow

Workflow details

Upon successful creation of a workflow, the end-user may review and configure data on specific details (Figure 34) such as the generic info and correlated Data Analytics Task(s). When he/she user clicks on an action button (Validate/Edit/Delete) for current Workflow, or (Create a Data Analytics/ edit/ 

delete) for each associated Data Analytics Task, the system loads the corresponding web-form (e.g., user prompt for deletion, Figure 35).

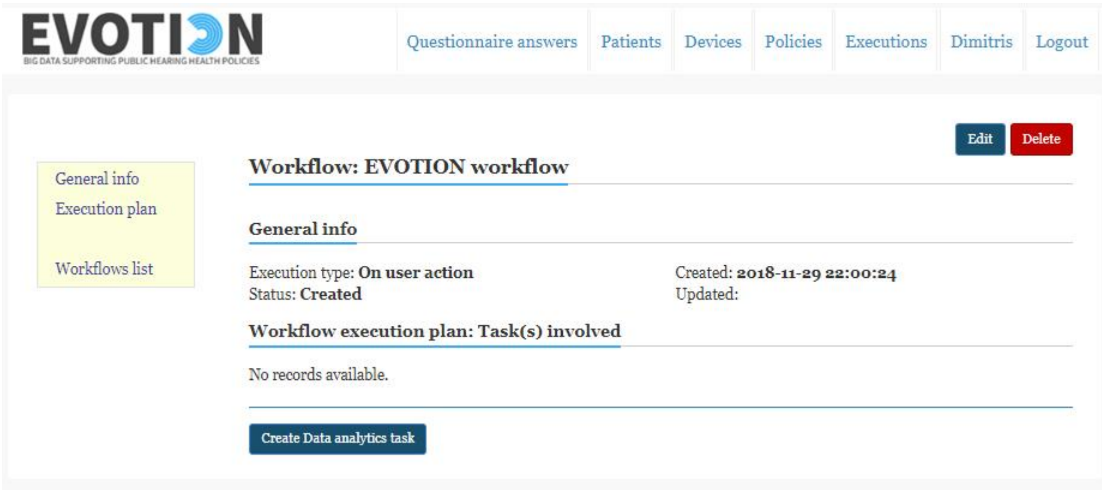


Figure 34: View of current workflow details

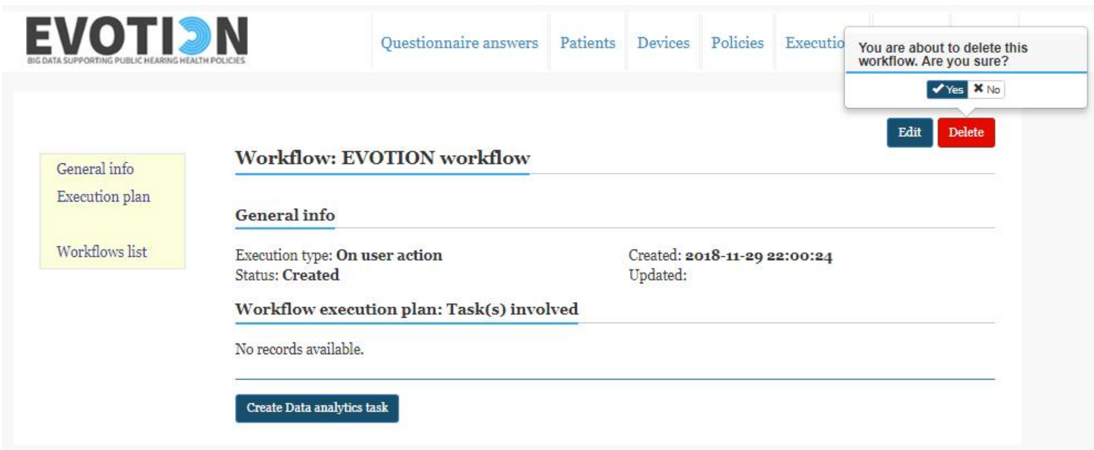


Figure 35: Prompt for deletion of a workflow

Data Analytics Task creation

End-user may define and associate more than one Data Analytics Tasks to a Workflow. To create a new Data Analytics Task, end-user fills in the type, table parameters consist the input data set, and the method to be applied (Figure 36). Later on, during the second step of the wizard, he/she may declare a dependent variable from those previously selected, and method-specific parameters (Figure 37).

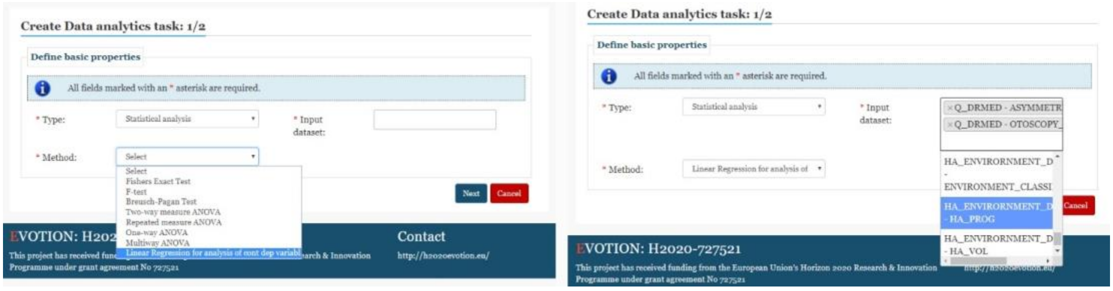


Figure 36: Create a new Data Analytics Task: Step 1/2: end-user selects type, input dataset and associated (to the type) method to be applied

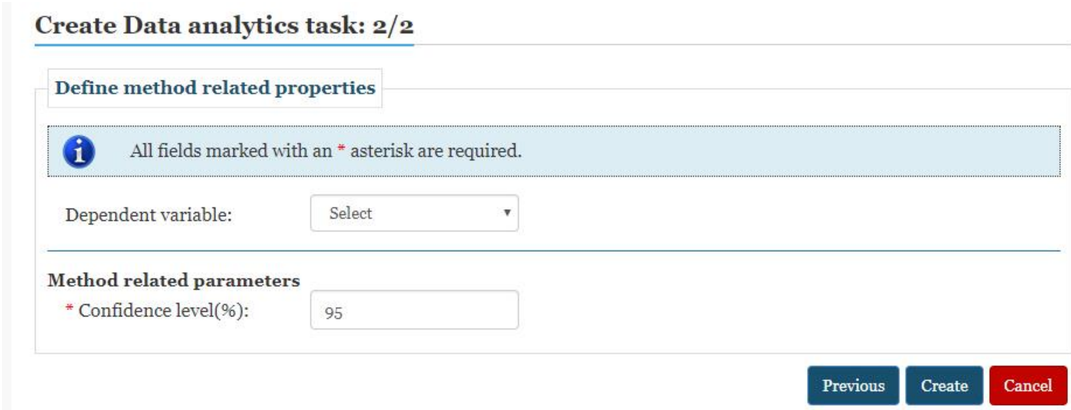


Figure 37: Create a new Data Analytics Task: Step 2/2: end-user may declare a dependent variable (from the ones of input dataset), and value for each method-specific parameter

Workflow validation

A cross-validation can be performed to a Workflow, if and only if at least one Data Analytics Task has been defined. When the end-user presses the Validate button (Figure 38), an internal function of the PHPDM E-service sends all the gathered information to the Ontology Reasoner (workflow). Subsequently, the Ontology Reasoner makes a copy of the latest version of the language OWL file with an instance name, instantiates the ontology, saves it (for backtracking purposes) and returns to the PHPDM E-service whether the ontology instance is valid or not.

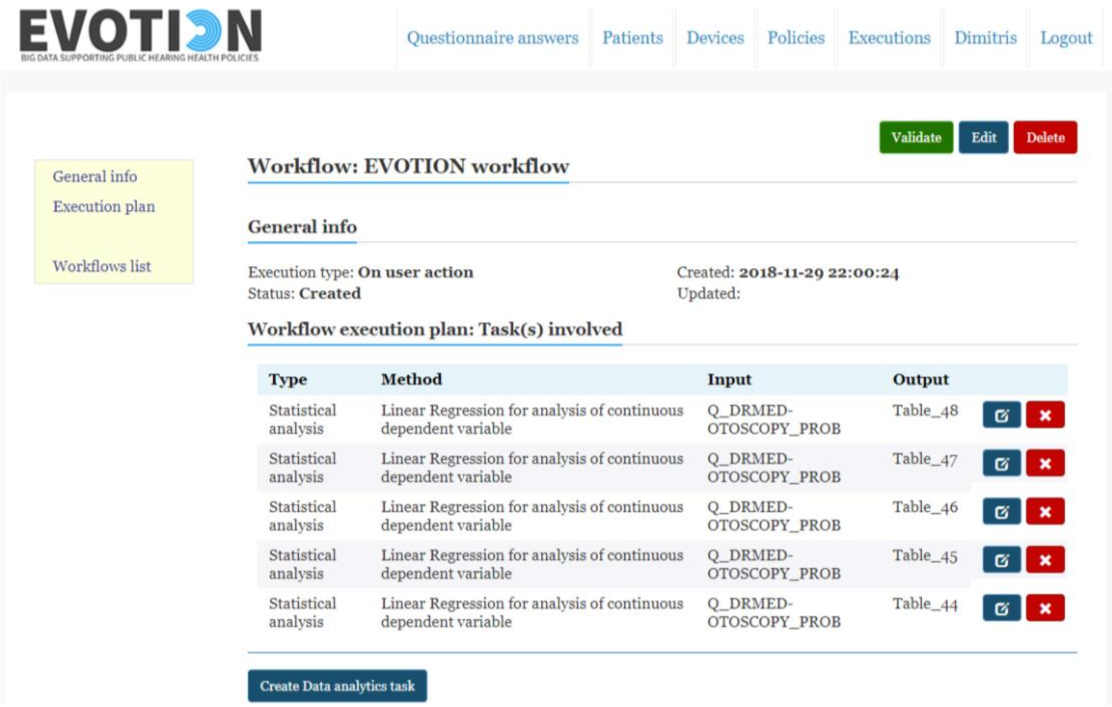


Figure 38: Trigger validation of current workflow

Workflow execution (requires integration to PHPDM model transformation tool)

A previously validated workflow can be executed. When the end-user presses the Execute button, an internal function of the PHPDM E-service sends all the gathered information (including the execution type) to the PHPDM model transformation tool (Figure 39). The latter is responsible for the communication with the BDA engine and the execution of data analytics workflows. The BDA Engine gets all the required datasets from the EVOTION Data Repository and creates relevant tables for the persistence of the execution results. Upon finalisation of the execution (i.e., execution of the workflow (or each task) is completed), PHPDM E-service displays the updated status of the workflow. In addition, the BDA Engine informs the Ontology Reasoner with the output data specifications and the latter updates the ontology accordingly.

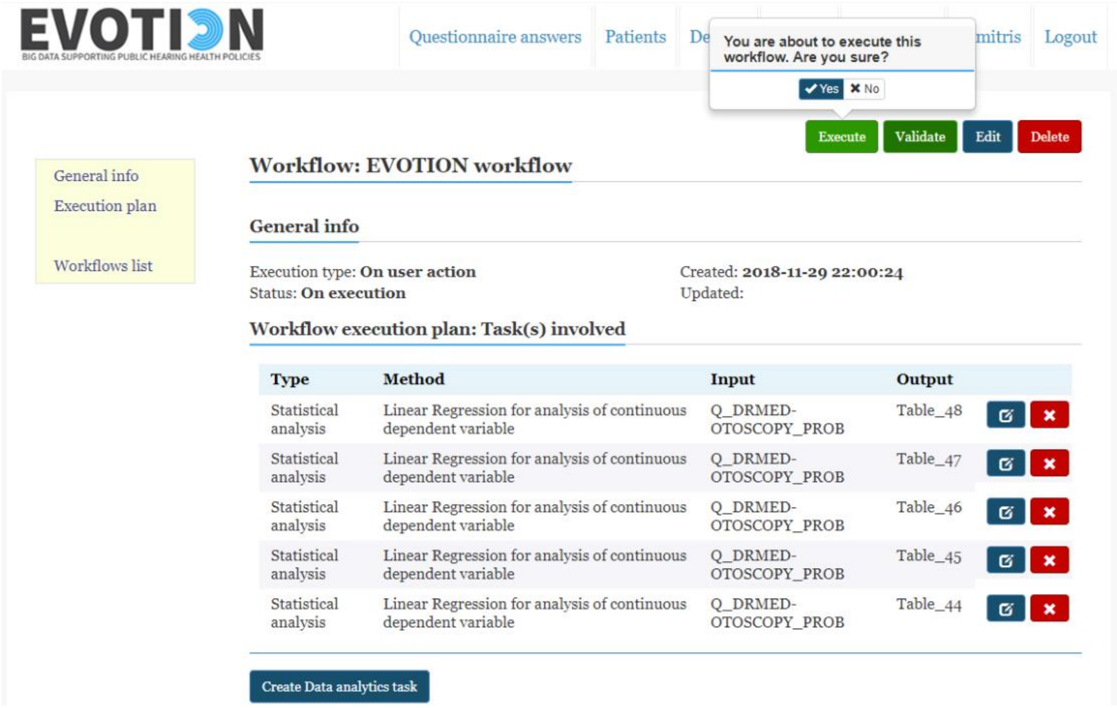


Figure 39: Trigger execution of current workflow

Policy creation

To create a new policy, end-user must fill in a model name, a goal description, an execution type and may enter a rationale. At least one policy objective should be declared by a description (subsequently a rationale can be declared as well), while he/she may add more of them (Figure 40). Later on, during the second step of the wizard, he/she should declare at least one policy action a corresponding workflow for each of the previously declared objectives (Figure 41). This 2-step wizard creates a Policy and associated Objectives, Policy Actions and Workflows (Figure 42).

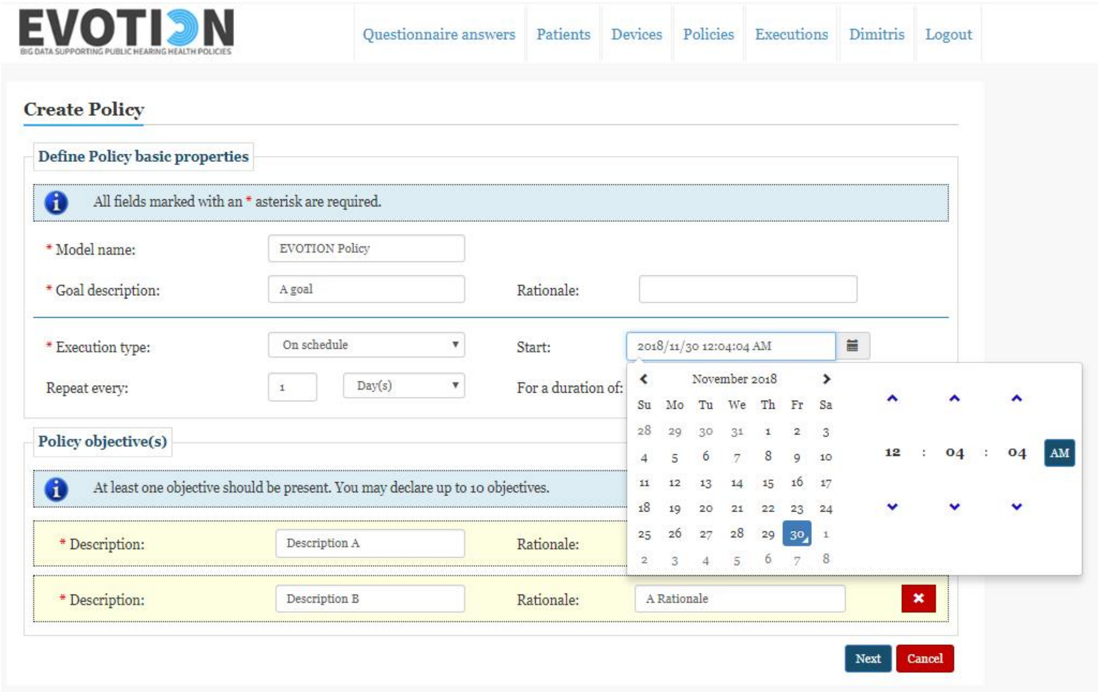


Figure 40: Create a new Policy: Step 1/2: end-user must declare model name, goal description, execution type and (perhaps) rationale. At least one policy objective description should be declared

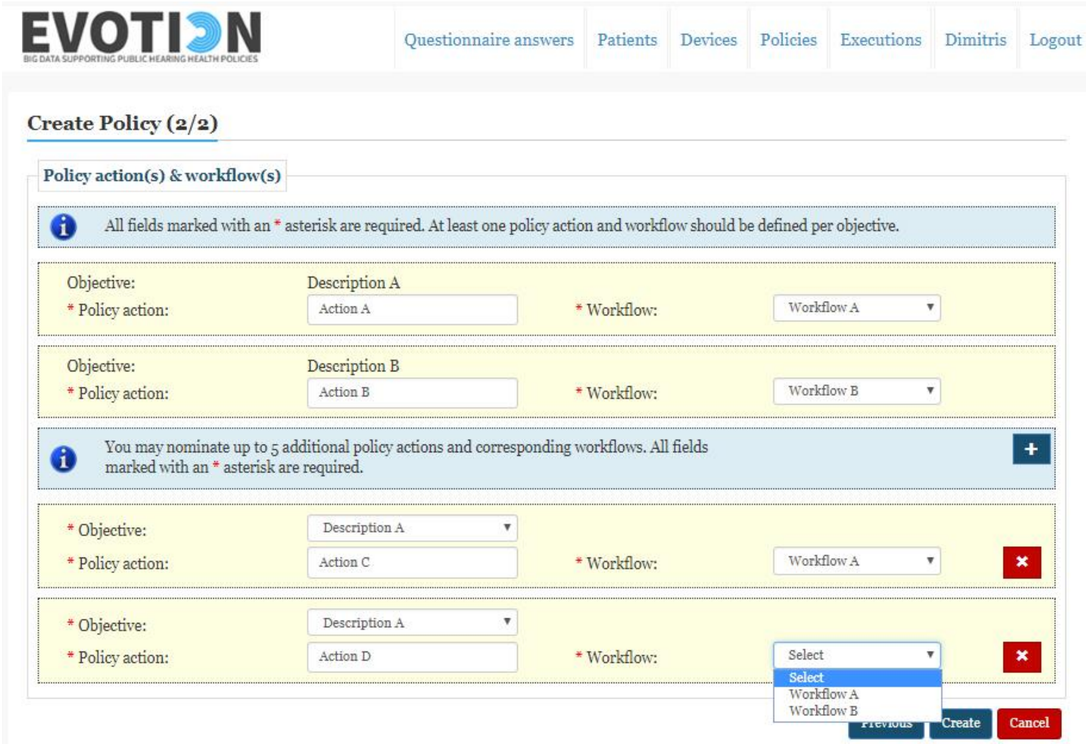


Figure 41: Create a new Policy: Step 2/2: end-user must declare policy action and workflow for each objective. He/she may define more than one policy action – workflow pair for an objective

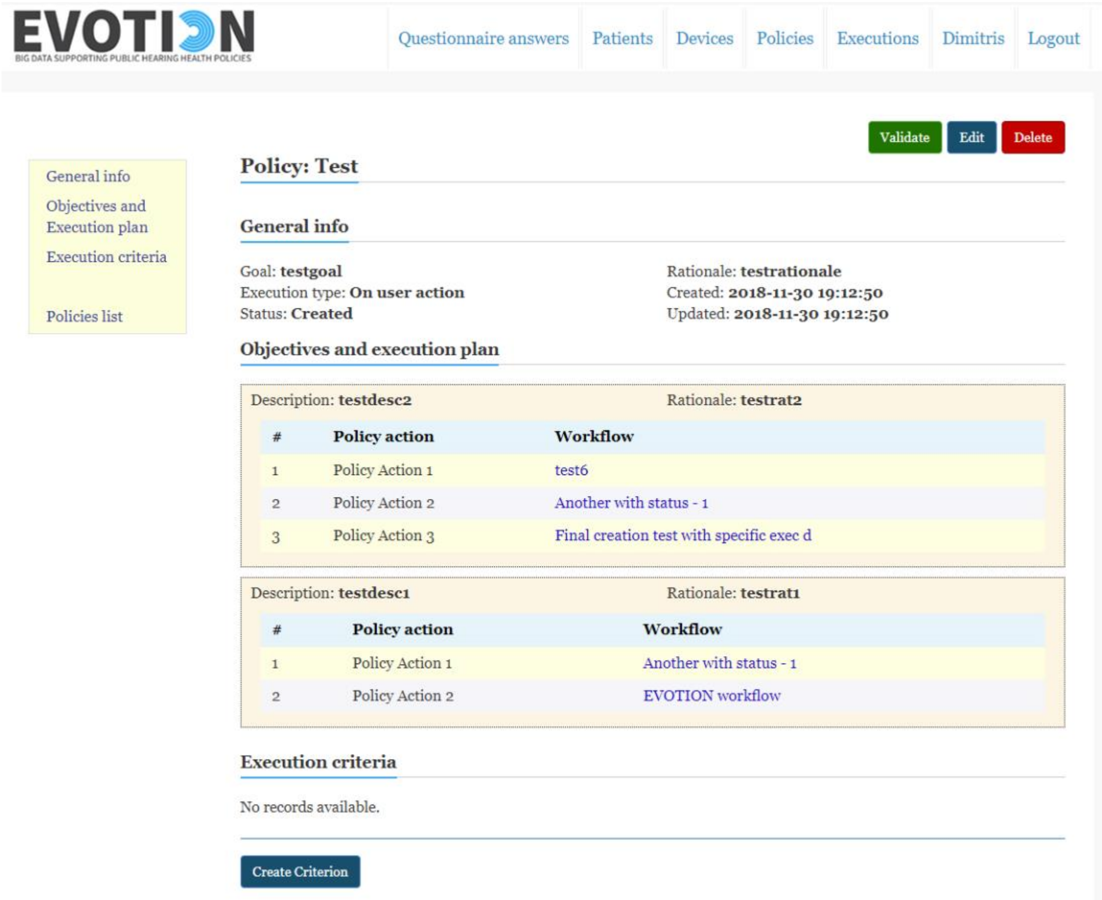


Figure 42: View of current policy details

Execution criterion creation

For a policy action, end-user can define an execution criterion, by selecting the input parameter, operation and value. He/she may define weighted criteria, or a logical combination of more than one (Figure 43).

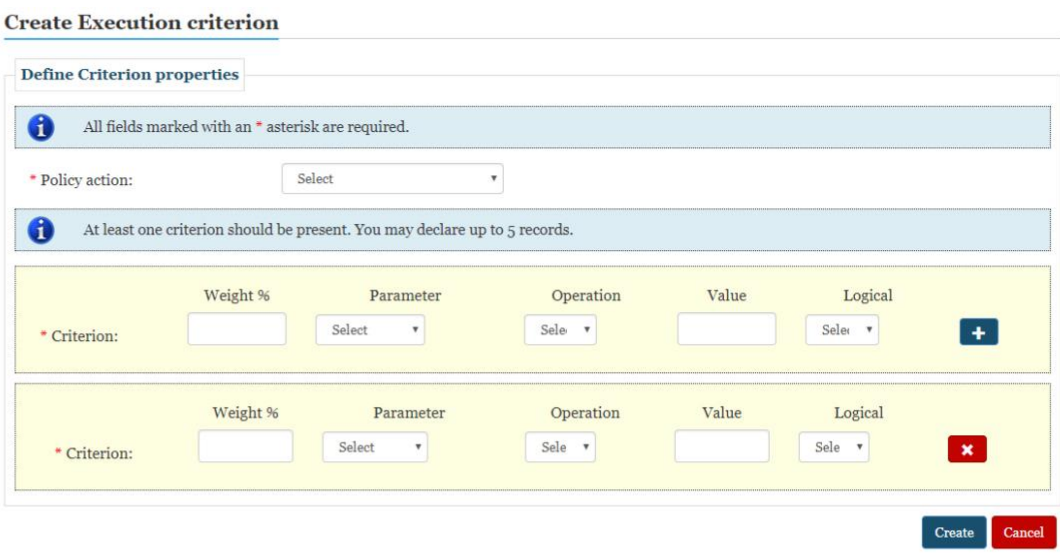


Figure 43: Create an execution criterion

Policy validation

As with workflow validation, whenever the end-user presses the validate button, a cross-validation is preformed to the current policy. In such case, an internal function of the PHPDM E-service sends all the gathered information to the Ontology Reasoner (policy, pObjectivesList, pObjWorkflowList, criteriaList), instantiates the ontology, saves it (for backtracking purposes) and returns to the PHPDM E-service whether the ontology instance is valid or not (Figure 44).

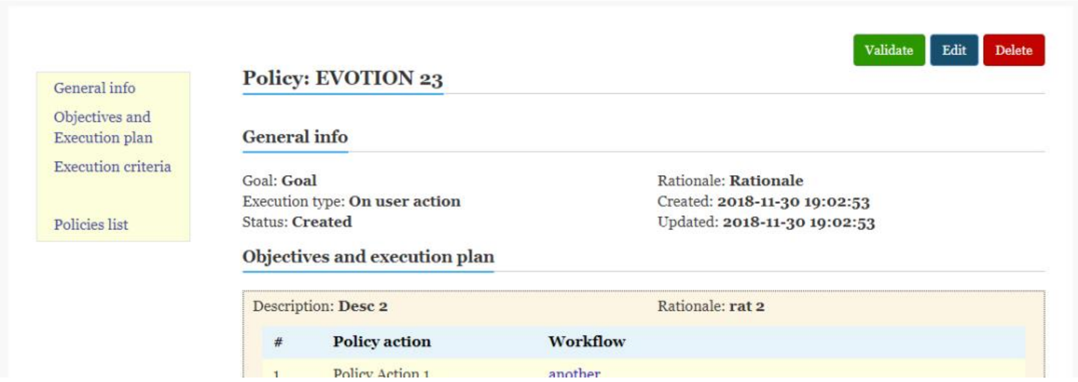


Figure 44: Trigger validation of current policy

Policy execution (requires integration to PHPDM model transformation tool)

When the end-user presses the execute button (Figure 45), all policy related information (including the execution type) is sent to the Transformation Tool. The Transformation Tool is responsible for the communication with the BDA engine and the execution of all data analytics workflows have been associated to the policy. When the execution of the policy is completed, the BDA Engine informs the Ontology Reasoner with the output data specifications and the latter updates the ontology accordingly. Results of an execution retrofit PHPDM e-service.

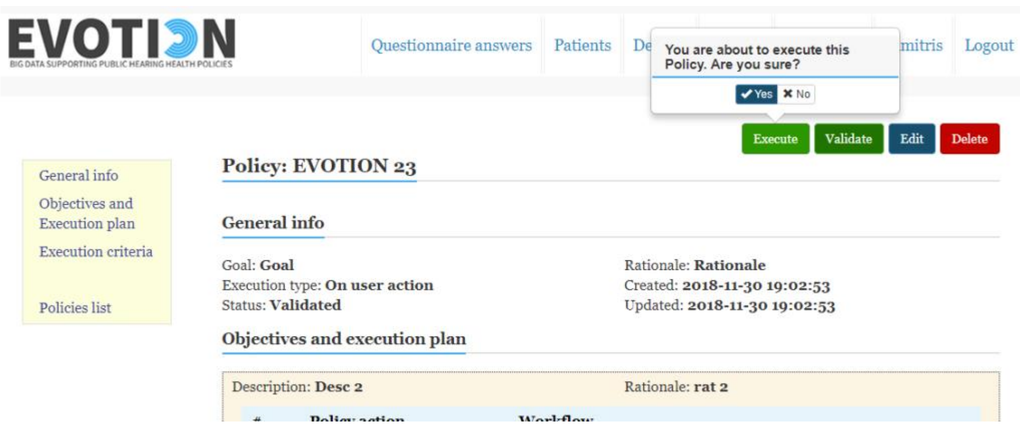


Figure 45: Trigger execution of current policy

4.5. Data Analytics Workflow Execution Summary

When a data analytics workflow is executed by the data analytics engine, the following three things may happen:

(1) The data analytics engine retrieves data from the repository as indicated by the data analytics workflow specified in the policy model. The relevant specification in the policy model includes a Phoenix SQL query that the data analytics engine should retrieve from the ontology reasoner and then execute by calling an appropriate operation in an API of the repository. The retrieved data are passed to the data analytics engine by the repository and stored within it in a table that corresponds to the data set in the policy model. A pointer to the table that has been created (e.g., the exact table name) should be returned to the data analytics engine. Subsequently the data analytics engine should call the ontology reasoner to store the actual table name in the instance of the policy model that describes the execution.

(2) The data analytics engine processes the data retrieved in (1) and produces some outputs. These outputs are determined by an output data set specified in the policy model. When the outputs have been produced, the data analytics engine should call the API of the repository to store the generated data in a table as specified by the policy model. The repository should store the data and return the name of the table where the data were stored. Subsequently the data analytics engine should call the ontology reasoner to store the actual table name in the instance of the policy model that describes the execution.

(3) Following (1) and (2) there will be an instantiation of the policy model within the ontology reasoner that has references to all the actual tables in the repository with the input and output data of the particular execution of the policy model.

Chapter 5

Evaluation

5.1. Overview

In this section we describe (a) the subjective evaluation of the utility of the proposed platform as a policy making tool and (b) the validation of the PHPDM modeling language described in Chapter 3 and (c) the evaluation of the platform's performance as a big data analytics tool. For the purposes of (a), we present the subjective evaluation process, the results and the treats to the validity of the results. For the purposes of (b) we present the validation process, the results and the treats to the validity of the results. For the purposes of (c), we present the evaluation scenario, the evaluation setup, the scenario parameters, the data sources, the synthetic data we used for the purposes of the evaluation of (b), as well as the EVOTION data we used, the evaluation process we followed, the experimental results and threats to validity of the results.

5.2. Evaluation Methodology

In order to evaluate all the aspects of this research, two different activities were performed: (a) a subjective evaluation of the developed policy language presented in Chapter 3 and platform as a policy making tool, and (b) an evaluation of the performance of the platform.

The subjective evaluation of the language and the platform involved three different types of potential users of the language and the platform, namely policy makers, clinicians and data analysts. This evaluation was based on the development of material summarising the platform and its usage and questionnaires to collect feedback on it following the presentation of the material. A separate questionnaire was developed for each of the three types of users to give their feedback. The questionnaires focused on evaluating the complexity and comprehensiveness of the public health policy decision making modeling language and the developed platform prototype as a policy making tool.

The experimental evaluation for the performance of the big data analytics engine at run time was based on executing a specific scenario of data analytics upon synthetic data. This activity was based on the use of synthetic data involving 1 million data points and real data from the EVOTION data repository.

A summary of the evaluation methodology is presented in Figure 46. More details of the methodology used for (a) and (b) are given in the following sub-sections.

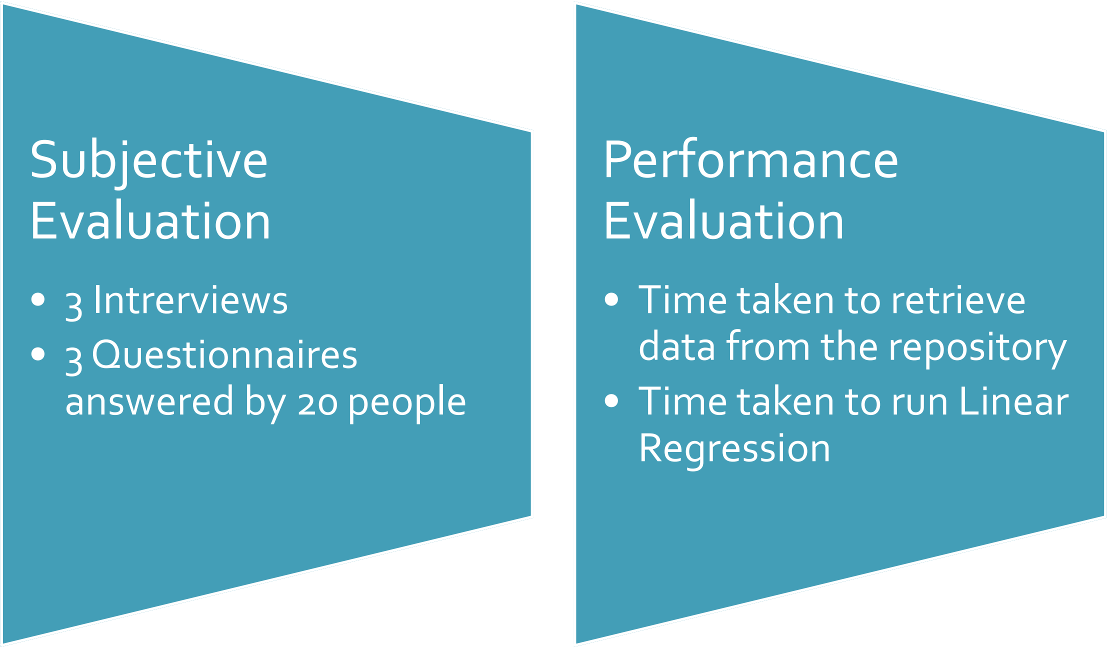


Figure 46 Evaluation Methodology

5.3. Subjective Evaluation

5.3.1. Subjective Evaluation Process

For the subjective evaluation, a set of materials summarising the language and the platform and questionnaires to obtain evaluating feedback for them were developed. Three separate questionnaires were developed, one for each of the three different types of envisaged users of the platform (i.e., policy makers, clinicians and data analysts). These questionnaires included:

(a) A Questionnaire for Policy Makers (available in Appendix A). This questionnaire was used to elicit feedback from policy making experts in Greece.

(b) A Questionnaire for Clinicians (available in Appendix A). This questionnaire was used to elicit feedback from practitioners in Greece.

(c) A Questionnaire for Data Analysts (available in Appendix A). This questionnaire was used to elicit feedback from data scientists in Greece and the UK.

Each of the three questionnaires included an introductory video giving information about the policy making language and the platform. The introductory video provided an example policy model and explained how this model was developed using the language and the tool. Three separate videos were constructed, one for each of the different types of prospective users that were targeted during the evaluation. Each video was tailored to the needs of the particular type of users. The scripts from each video are available in Appendix A.

The usability section was based primarily on Lund [100]. The reason for this was that a community member could declare that the platform did not satisfy him/her for a specific purpose or had a useful generic characteristic, in other words to focus on aspects that should be further improved in order to serve his/her needs. As Lund indicates [100], the first aim was to make the items as simply worded as possible, and as general as possible and constructed as five-point Likert rating scales (1: strongly disagree to 5: strongly agree). Consequently, questions were included in the questionnaires, to assess the following issues:

- Ease of learning (Intuitiveness of user interface): how easy was to learn to use it (user interfaces were easy to understand without training)
- Ease of use: how easy/simple was to use it and complete tasks
- Ensure easy and user-friendly navigation: how clear and consistent was the hierarchical structure
- User confidence and easiness of task completion: if it made the tasks the end-user wants to accomplish easier to get done on every occasion
- Meet end-user needs (Potential of use in normal practice): if it was useful enough to be embedded in normal practice
- Ease of use and learning: how easily was to use it without written instruction
- Ease of use and learning: how much guidance and instructions required for effective usage
- Overall usefulness: how quickly a task can be completed in comparison to normal practice

Usefulness concerns whether the platform could be considered by the users as a tool that would increase their performance, whilst they engage in tasks necessary for achieving their goals [101]. Although a system can be characterised as easy to use and learn, unless it achieves the specific goals of a specific user, it will not be used. Consequently, it is important to assess whether a system works as supposed to (effective and efficient manner) by meeting the user requirements. To address usefulness evaluation, the questionnaires included questions utilising a five-point Likert rating scale, covering the following issues:

- Enhancing effectiveness: how much it helped end-user be more effective
- Enhancing productivity: how much it helped end-user be more productive.
- Minimality of operation A: how effortless and successful was in performing its intended task
- Operating according to expectations: if it worked the way it was expected
- Minimality of operation B: if it required the fewest steps possible to accomplish what the end-user wanted to do with it
- Meeting user's exceptions: if the platform preforms everything end-user would expect from it to do
- Meeting end-user exceptions: if it met end-user needs

Early versions of the developed questionnaires materials were used for applying to obtain ethics approval from the appropriate authority of City, University of London. This was a necessary pre-condition for carrying out the evaluation study. The application for obtaining this approval and the supporting material for it are given in Appendix A.

Following ethics approval, the actual subjective evaluation was carried out. Initially, this process involved, interviewing three senior representatives of policy makers, clinicians and data analysts to establish if the materials that had been developed for the main phase of the evaluation process were fit for purpose. The three experts were given the developed materials and questionnaires and asked to indicate whether these would be effective as drivers of the evaluation process. The senior user representatives included one senior policy maker from the London School of Hygiene & Tropical Medicine, one senior clinician from the Otolaryngology Department of the Hippokration Hospital of the University of Athens, and one senior data analyst from the National Technical University of Athens. These senior representatives were interviewed based on the guide given in Appendix A. The interviews focused on whether the questionnaires that would be used for evaluating the platform would be adequate for this purpose also asking them whether they would add or remove any question. No questions were added or removed as a result of this process. Some changes were made to the wording of the questions and to the video of the questionnaire for clinicians, as described in section 5.3.2.1.

In the main part of the evaluation, all the participants were asked to sign a consent form and received a participant information sheet (available in Appendix A). No personal data was gathered by the questionnaires.

In total, feedback was collected from 20 participants (7 policy makers, 5 clinicians and 8 data analysts).The results from the analysis of this feedback are presented in the following section.

5.3.2. Subjective Evaluation Results

In this section we present the results from the subjective evaluation using the three types of questionnaires described above. First, we present the results from the interviews that evaluate each questionnaire's completeness. Then we present the questionnaire answers for each question in detail.

5.3.2.1. Results from the Interviews

Some important points were extracted from the interview with the senior policy maker for the evaluation of the questionnaire for policy makers. There were some technical issues in the video that introduces the participants to the platform. More specifically, in the opinion of the expert, the speaker in the video sometimes spoke too quickly and joined up sentences so it became difficult to follow. The expert indicated that the usability and usefulness sections in the questionnaire would evaluate the platform adequately and that the general section added important points to the evaluation of the platform. The overall judgement of the expert was the questionnaire and the video are generally very good and only minor issues should be addressed. As a consequence of this interview, some corrections were made to the wording of some questions.

The interview with the senior clinician added some important points to the evaluation of the questionnaire for clinicians. In the opinion of this expert, it was difficult to follow the video sometimes and there was a need to provide more detail about the platform within it. The suggestions of the expert clinician were taken into account and the video for clinicians was edited. Apart from the comments given for the video, the expert clinician found that the questionnaire's general section (i.e., the section that includes the questions about whether the clinical decisions need to be evidence based, whether they would benefit from statistical analysis and machine learning and whether the described platform would enhance the clinical practice) was adequate for the evaluation of the platform and that the other two sections of it were also fit for purpose. As a consequence of this interview, the video of the questionnaire for clinicians was made more descriptive.

Finally, the interview with the senior data analyst confirmed that both the video and the questionnaire for data analysts were very informative. The senior data analyst indicated that the video was descriptive enough, but not tiring and easy to follow for a data analyst. With regards to the questionnaire, the senior data analysts approved all the questions. As a consequence of this interview, no change was made to the video and the questionnaire for the data analysts.

5.3.2.2. Results from the Questionnaire for Policy Makers

The questionnaire for policy makers was answered by 7 experts in Greece. Below we present each question's answers in detail.

General Section

Need for evidence-based decisions: almost all the experts (86%) agree (i.e. rate 4 or 5) that there is a need for evidence based public health policy decisions (Figure 47). There was one participant who does not agree.

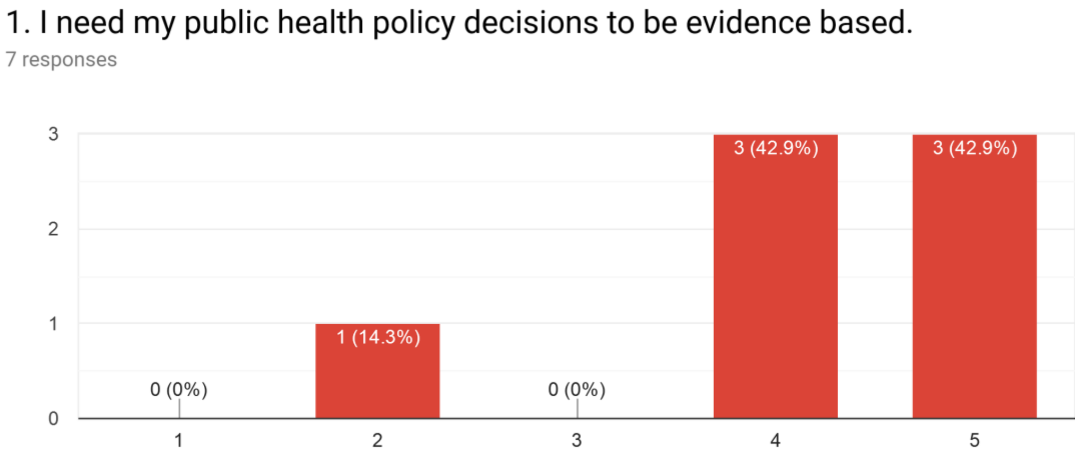


Figure 47 Need for evidence-based decisions

Benefits from statistical analysis: almost all of the participants (86%) indicated that they would benefit from statistical analysis (i.e. rate 4 or 5). There was one participant who does not agree. The answers of this question are shown in Figure 48.

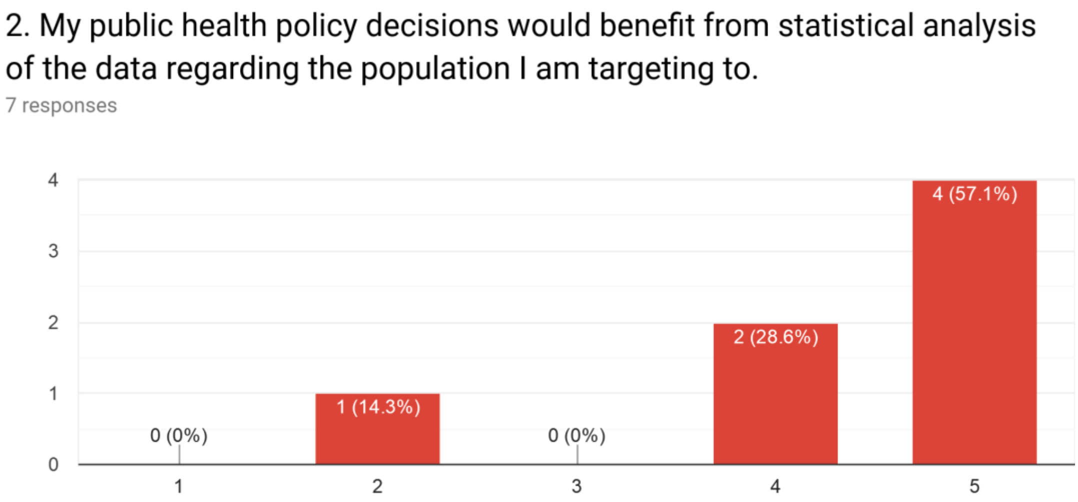


Figure 48 Benefits from statistical analysis

Benefits from machine learning: almost all of the participants (86%) indicated that they would benefit from prediction analysis (i.e. rate 4 or 5). There was one participant who does not agree. The answers of this question are shown in Figure 49.

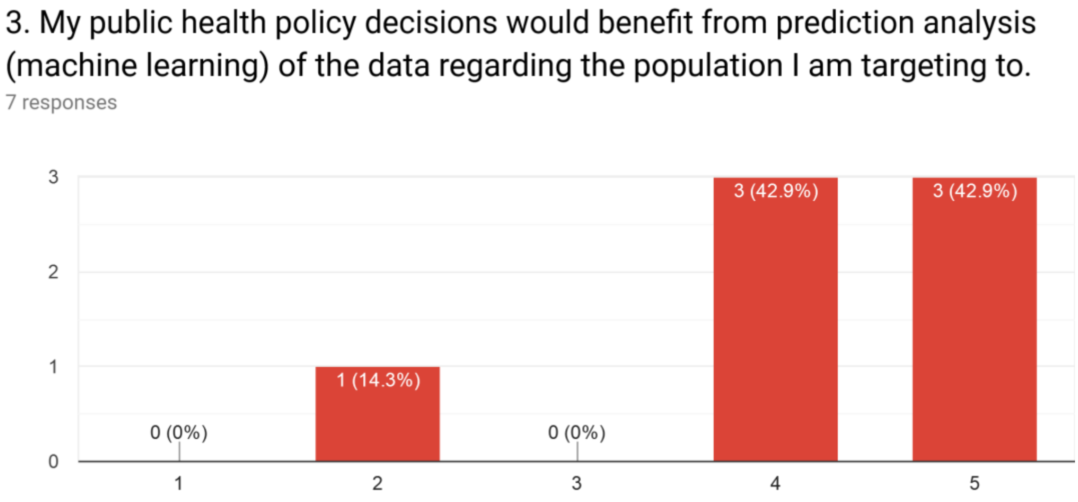


Figure 49 Benefits from machine learning

Enhancement of decision making: most of the participants (86%) answered that the described platform would enhance their decision making (i.e. rate 4 or 5). There was one participant who disagrees. The answers of this question are shown in Figure 50.

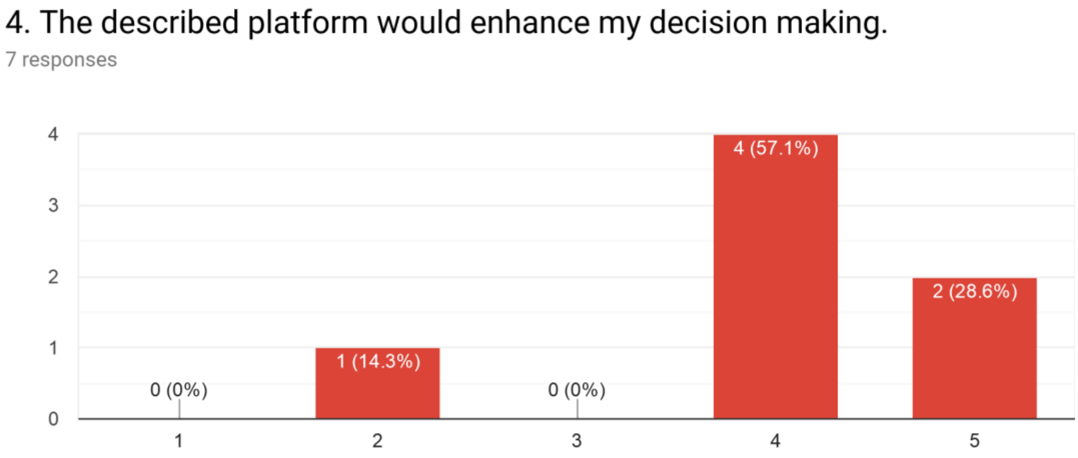


Figure 50 Enhancement of decision making

Usability

Intuitiveness of user interface: most of the participants (86%) answered that they found the user interface intuitive (i.e. rate 4 or 5). There was one participant who disagrees. The answers of this question are shown in Figure 51.

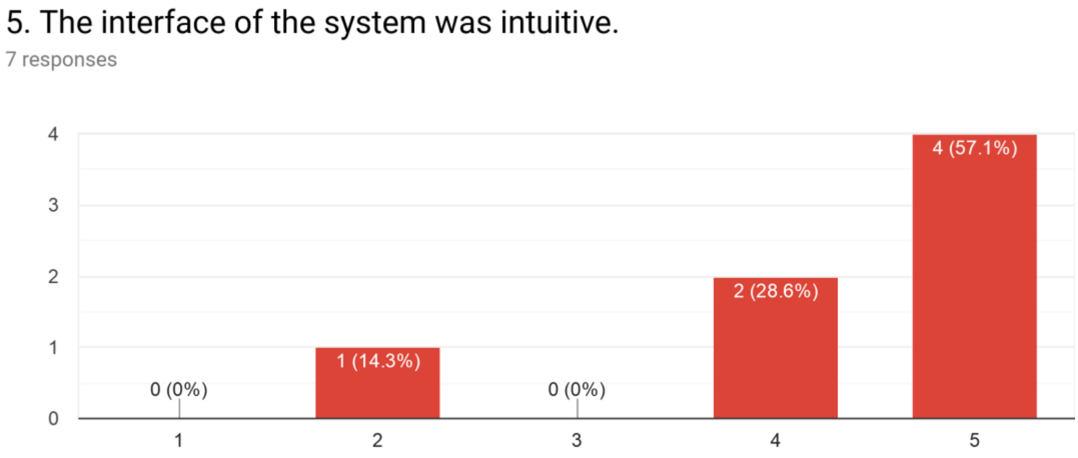


Figure 51 Intuitiveness of user interface

Easiness of task completion: most of the participants (86%) answered that it was easy to complete tasks using the system (i.e. rate 4 or 5). There was one participant who does not agree. The answers of this question are shown in Figure 52.

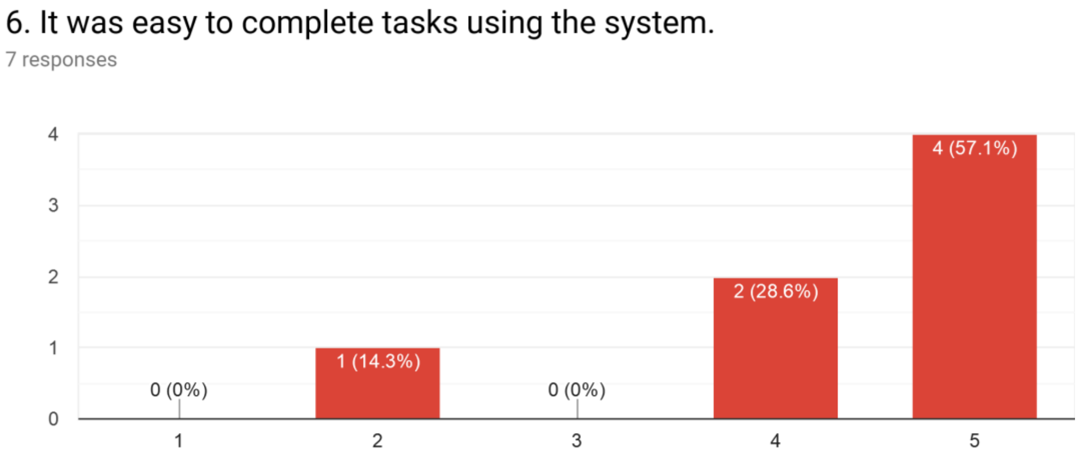


Figure 52 Easiness of task completion

User-friendly navigation: most of the participants (86%) answered that navigation through the different options of the system was effective (i.e. rate 4 or 5). There was one participant who disagrees. The answers of this question are shown in Figure 53.

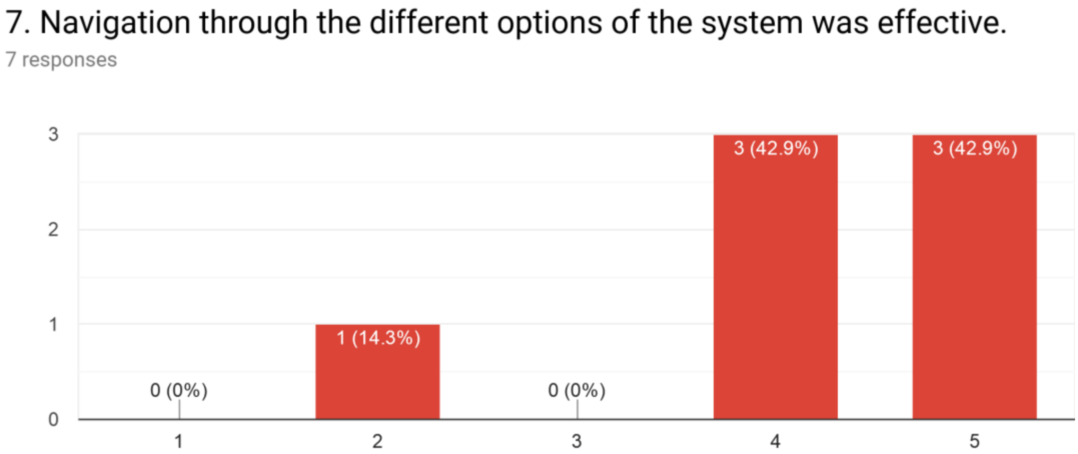


Figure 53 User-friendly navigation

User confidence: most of the participants (86%) answered that they would feel confident using the system (i.e. rate 4 or 5). There was one participant who does not agree. The answers of this question are shown in Figure 54.

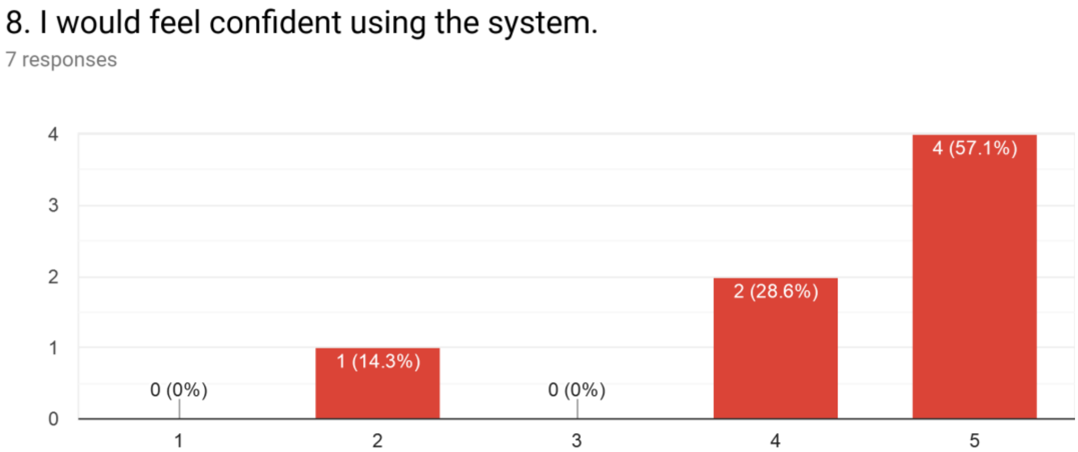


Figure 54 User confidence

Utilisation of the system as it is in normal practice: almost all of the participants (86%) answered that they could use the system frequently in their normal policy making practice (i.e. rate 4 or 5). 71% of them strongly agree. There was one participant who does not agree. The answers of this question are shown in Figure 55.

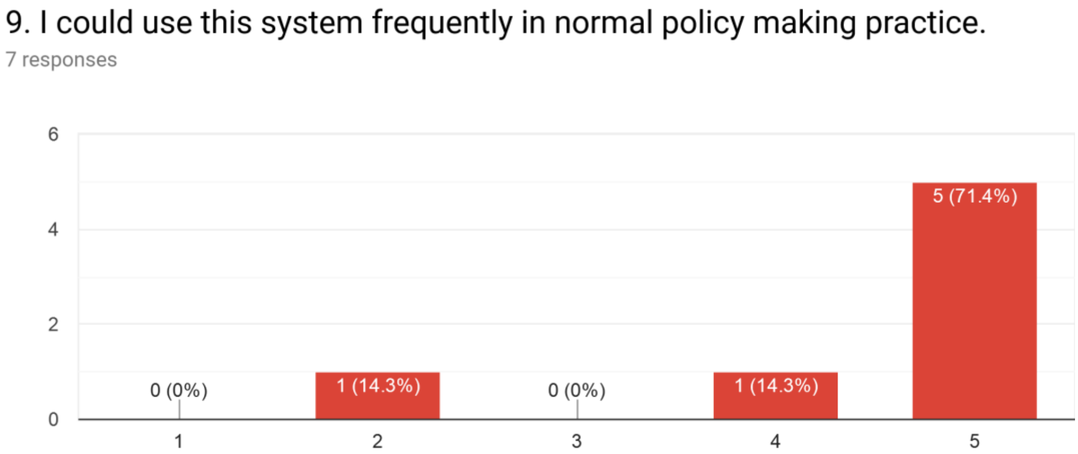


Figure 55 Utilisation of the system as it is in normal practice

Self-explanatory usage: 57% of the participants are neutral on whether they would need technical advice in order to use the system effectively. 29% of the participants disagree and only one participant (14%) agrees. The answers of this question are shown in Figure 56.

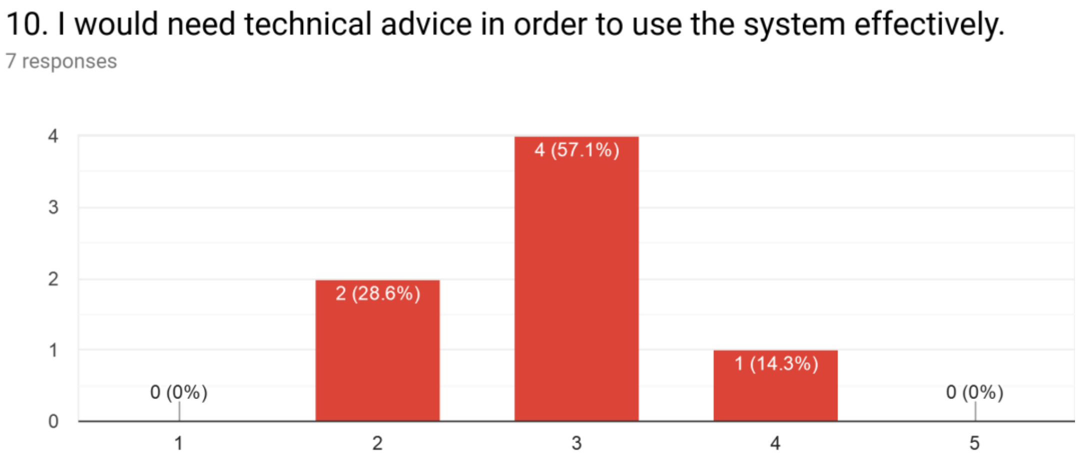


Figure 56 Self-explanatory usage

Guidance and instructions required for effective usage: 71% of the participants are neutral on whether they would need training in order to use the system effectively. One participants (14%) disagrees and one participant (14%) agrees. The answers of this question are shown in Figure 57.

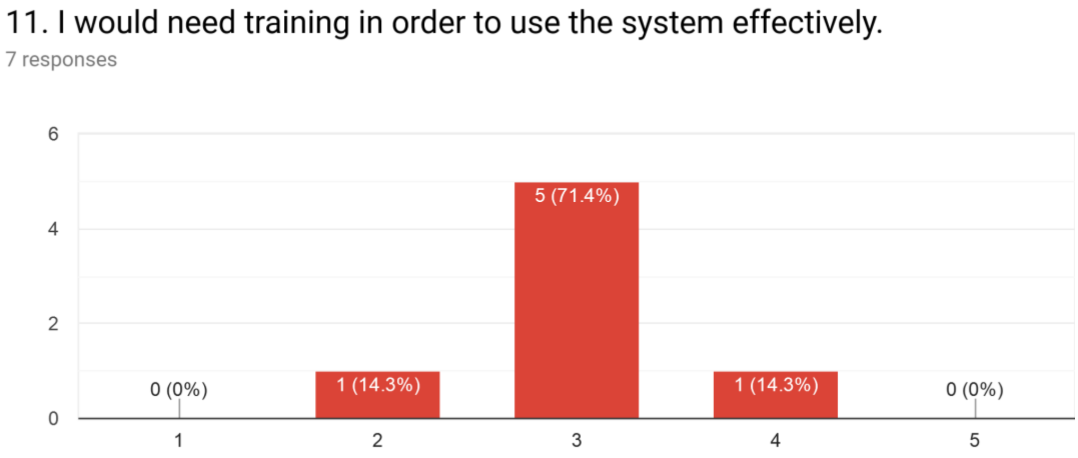


Figure 57 Guidance and instructions required for effective usage

How quickly a task can be completed in comparison to normal practice: most of the participants (86%) answered that they feel satisfied with the amount of time it took to complete a task in comparison to their previous working method (i.e. rate 4 or 5). There was one participant who does not agree. The answers of this question are shown in Figure 58.

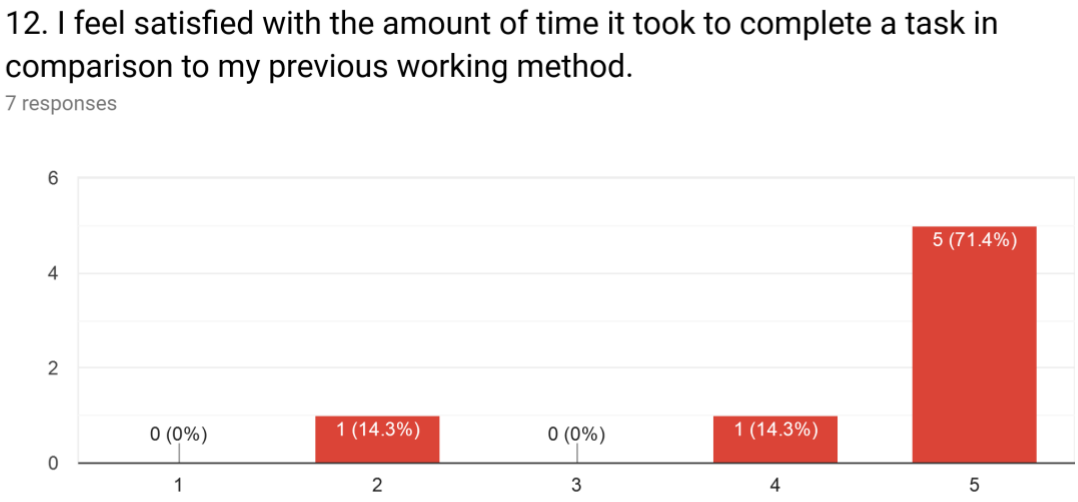


Figure 58 How quickly a task can be completed in comparison to normal practice

Usefulness

Effect of the platform on effectiveness: most of the participants (86%) answered that the use of the system would help them to be more effective in taking policy decisions (i.e. rate 4 or 5). There was one participant who disagrees. The answers of this question are shown in Figure 59.

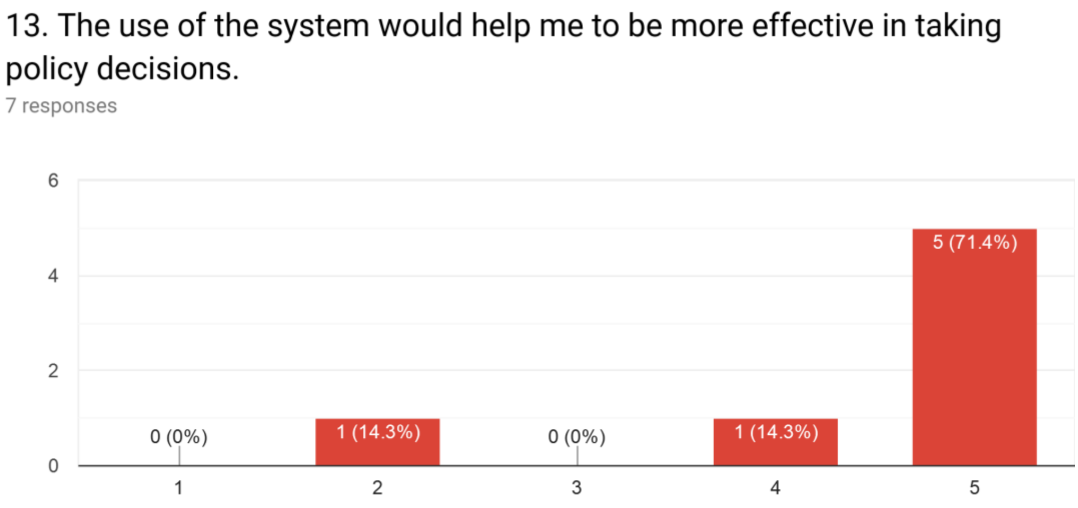


Figure 59 Effect of the platform on effectiveness

Effect of the platform on productivity: most of the participants (86%) answered that the use of the system would help them to be more productive (i.e. rate 4 or 5). There was one participant who does not agree. The answers of this question are shown in Figure 60.

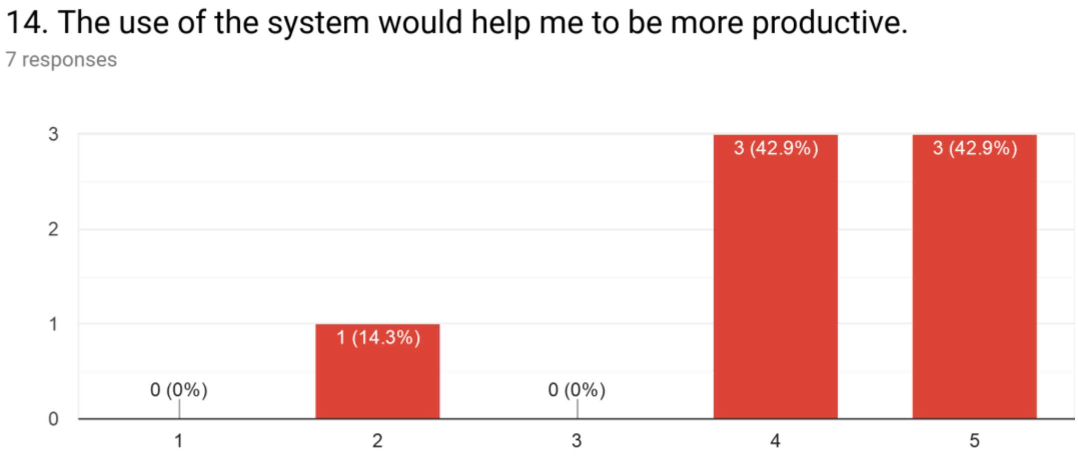


Figure 60 Effect of the platform on productivity

Success of platform in supporting an intended task: almost all of the participants (86%) answered that the system is successful in performing its intended task (i.e. rate 4 or 5). There was one participant who does not agree. The answers of this question are shown in Figure 61.

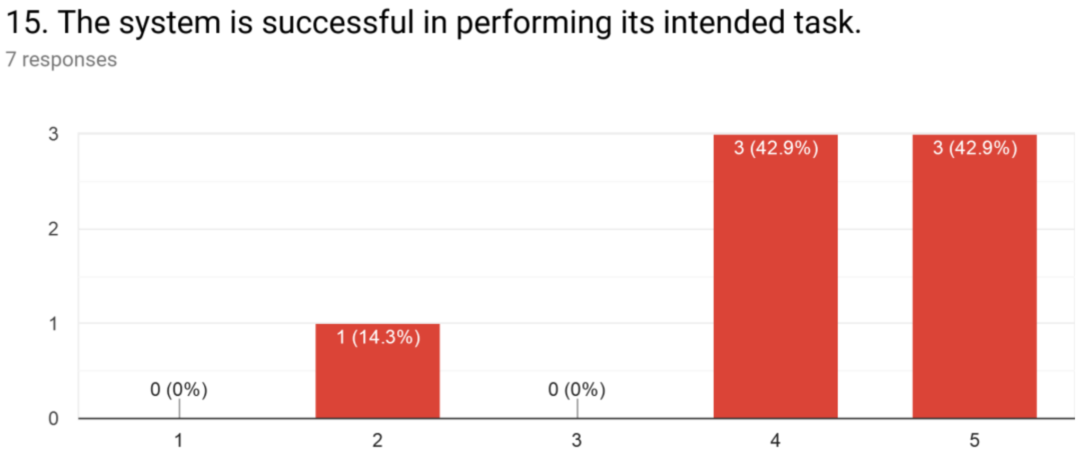


Figure 61 Success of platform in supporting an intended task

Meeting end-user expectations: almost all of the participants (86%) answered that the system works the way they want it to work (i.e. rate 4 or 5). There was one participant who disagrees. The answers of this question are shown in Figure 62.

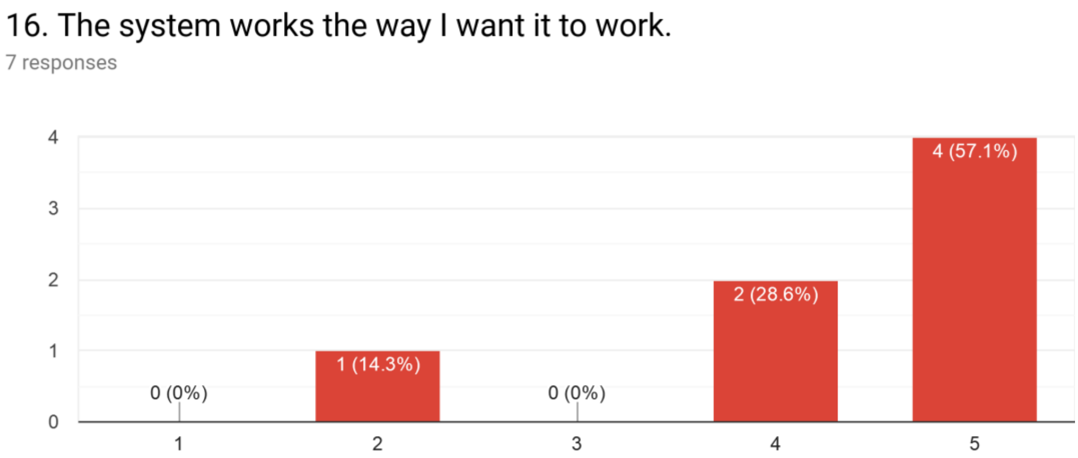


Figure 62 Meeting end-user expectations

Minimality of operation: most of the participants (86%) answered that the system did not require taking unnecessary steps in order to accomplish what they wanted to do with it (i.e. rate 4 or 5). There was one participant who disagrees. The answers of this question are shown in Figure 63.

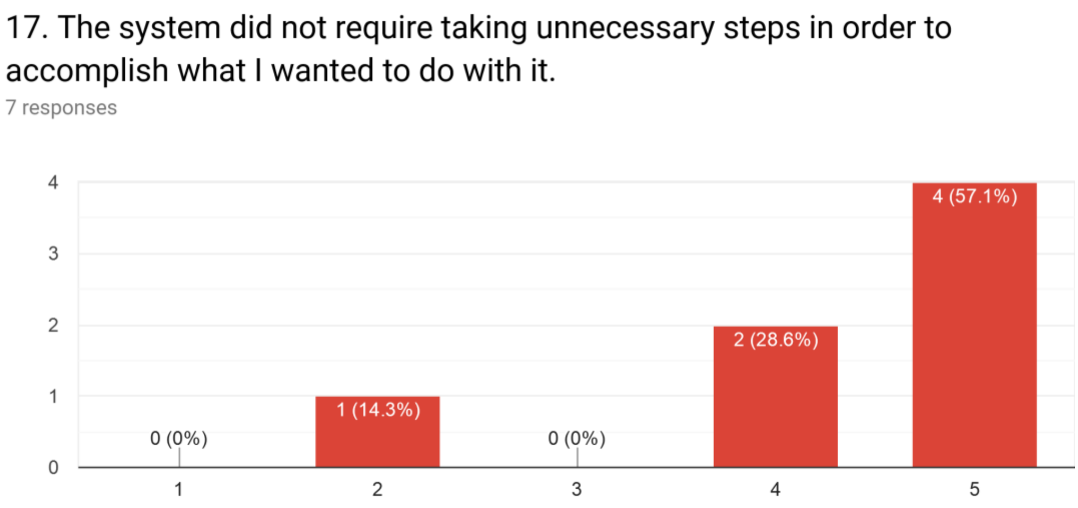


Figure 63 Minimality of operation

Meeting end-user needs: most of the participants (86%) answered that the system meets their needs (i.e. rate 4 or 5). There was one participant who does not agree. The answers of this question are shown in Figure 64.

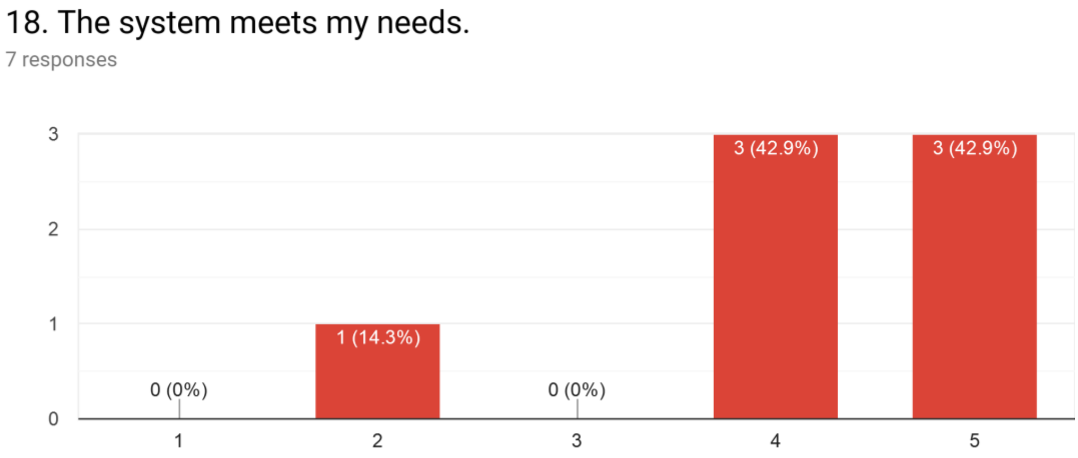


Figure 64 Meeting end-user needs

No comments were left by any participant in the last question.

5.3.2.3. Results from the Questionnaire for Clinicians

General Section

Need for evidence-based decisions: all of the participants (100%) strongly agree that their clinical decisions need to be evidence based. The answers of this question are shown in Figure 65.

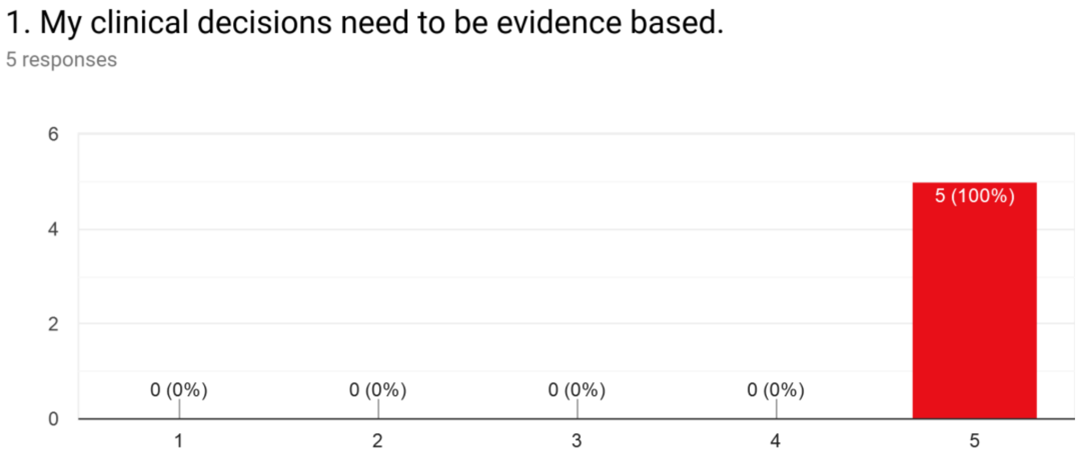


Figure 65 Need for evidence-based decisions

Benefits from statistical analysis: all of the participants (100%) agree that their clinical decisions would benefit from statistical analysis of the data regarding the population they are targeting to (i.e. rate 4 or 5). The answers of this question are shown in Figure 66.

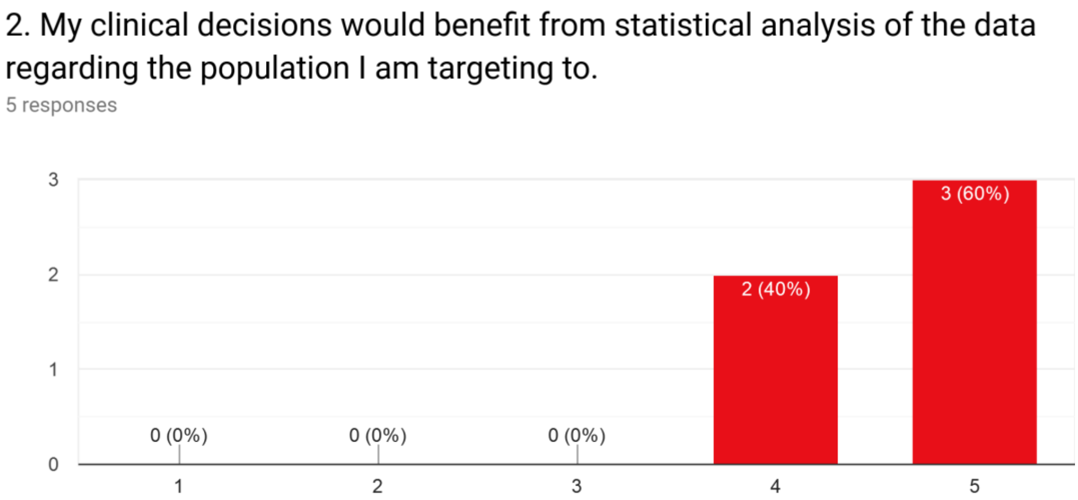


Figure 66 Benefits from statistical analysis

Benefits from machine learning: most of the participants (80%) indicated that their clinical decisions would benefit from prediction analysis (machine learning) of the data regarding the population they are targeting to (i.e. rate 4 or 5). There was one clinician who was neutral. The answers of this question are shown in Figure 67.

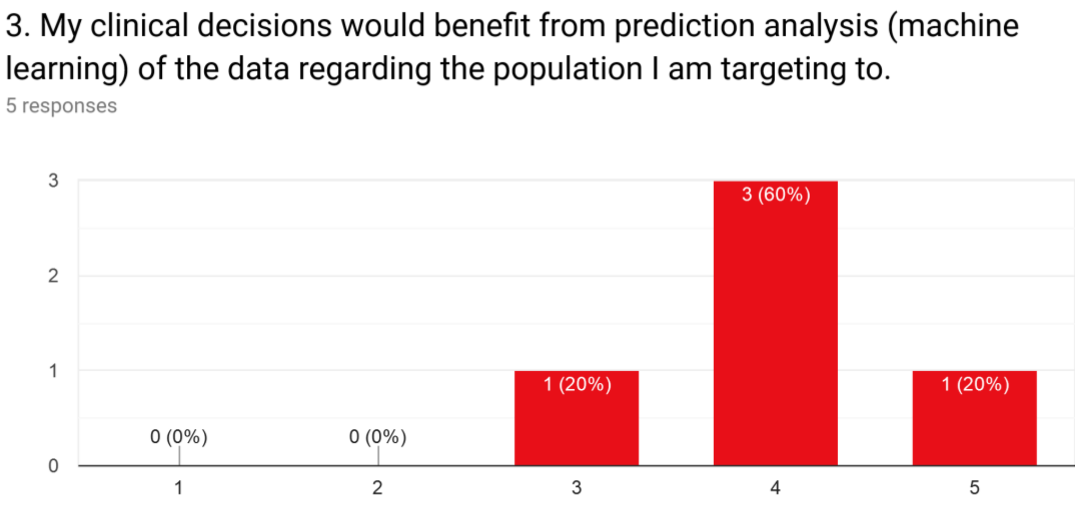


Figure 67 Benefits from machine learning

Enhancement of clinical practice: most of the participants (80%) indicated that the described platform would enhance their clinical practice (i.e. rate 4 or 5). There was one clinician who was neutral. The answers of this question are shown in Figure 68.

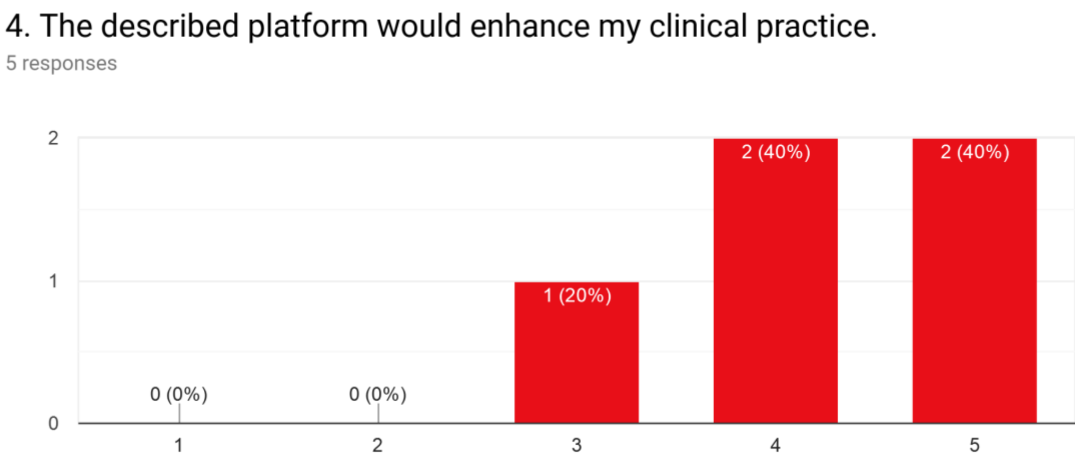


Figure 68 Enhancement of clinical practice

Usability

Intuitiveness of user interface: most of the participants (80%) found the user interface of the system intuitive (i.e. rate 4 or 5). There was one clinician who was neutral. The answers of this question are shown in Figure 69.

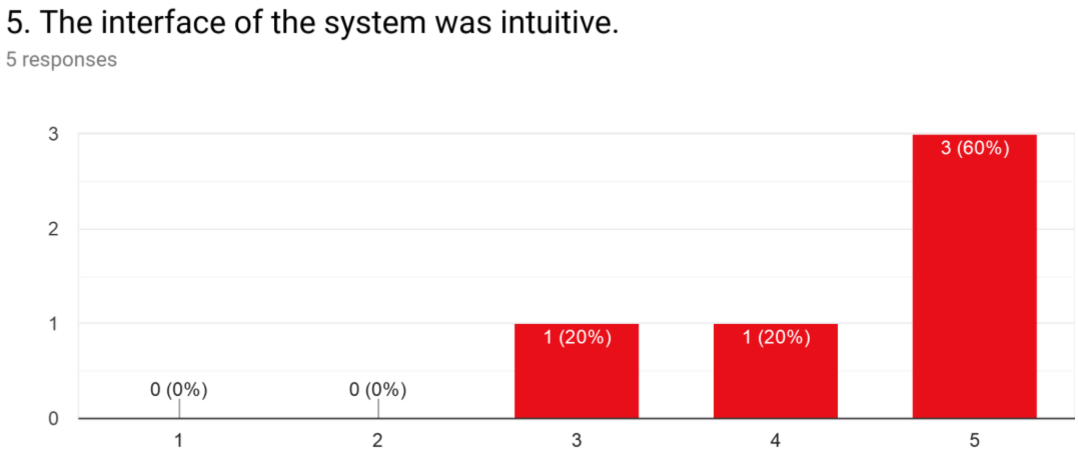


Figure 69 Intuitiveness of user interface

Easiness of task completion: most of the participants (80%) indicated that it was easy to complete tasks using the system (i.e. rate 4 or 5). There was one clinician who was neutral. The answers of this question are shown in Figure 70.

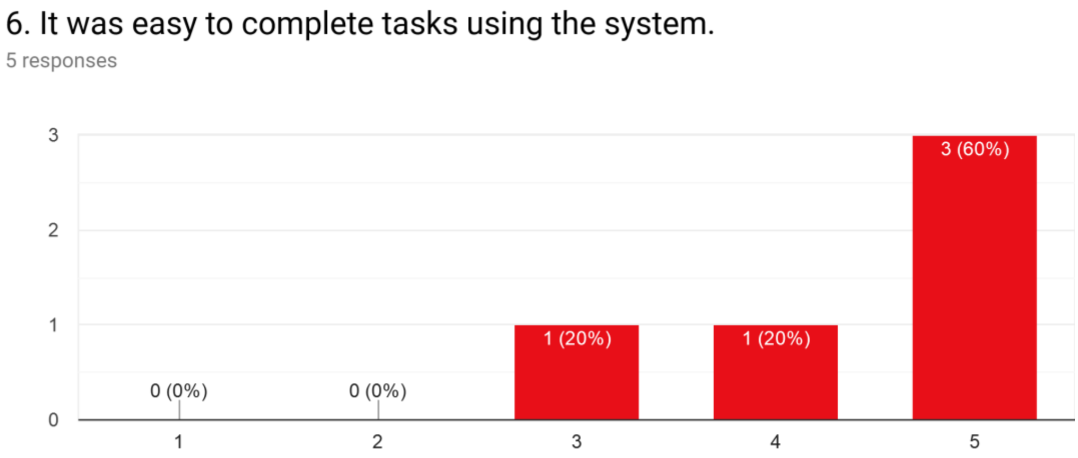


Figure 70 Easiness of task completion

User-friendly navigation: most of the participants (80%) found the navigation through the different options of the system effective (i.e. rate 4 or 5). There was one clinician who disagrees. The answers of this question are shown in Figure 71.

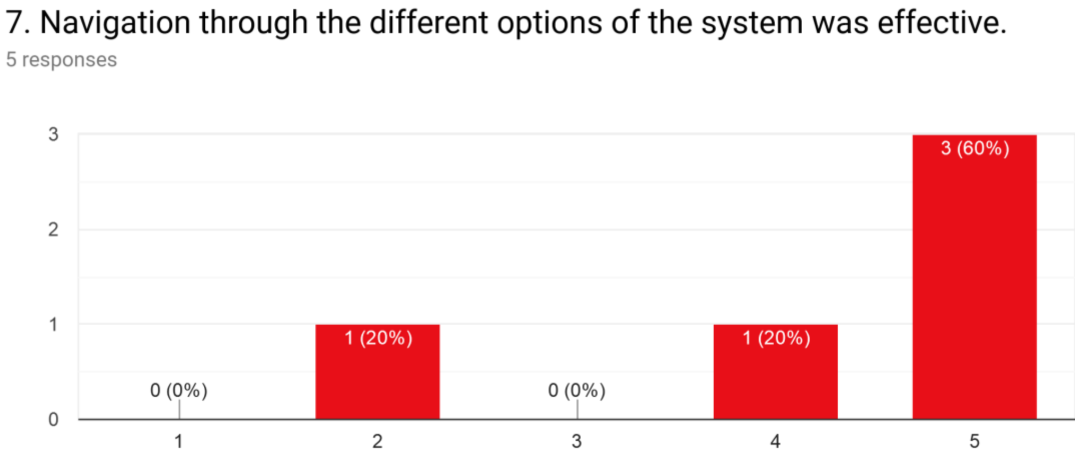


Figure 71 User-friendly navigation

User confidence: most of the participants (80%) indicated that they would feel confident using the system (i.e. rate 4 or 5). There was one clinician who strongly disagrees. The answers of this question are shown in Figure 72.

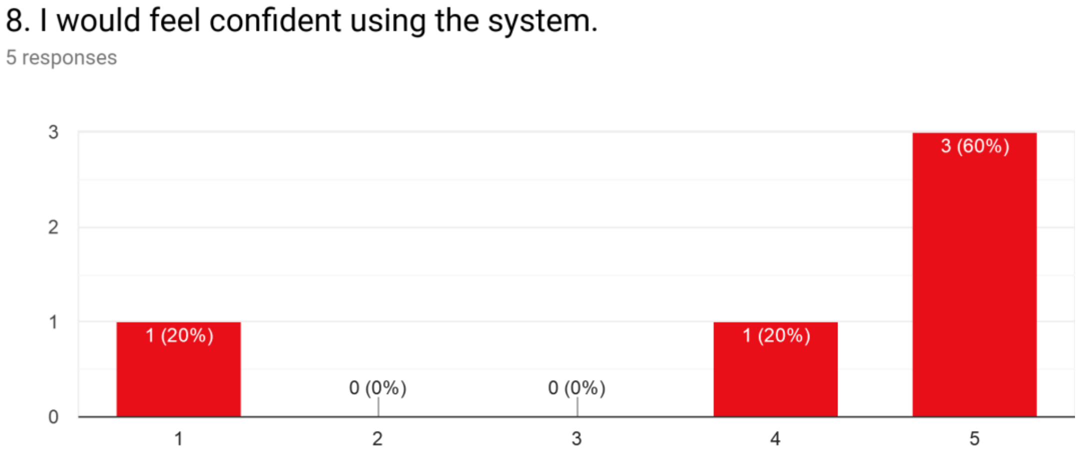


Figure 72 User confidence

Utilisation of the system as it is in normal practice: most of the participants (80%) strongly agree that they could use this system frequently in their normal clinical practice (i.e. rate 5). There was one clinician who strongly disagrees. The answers of this question are shown in Figure 73.

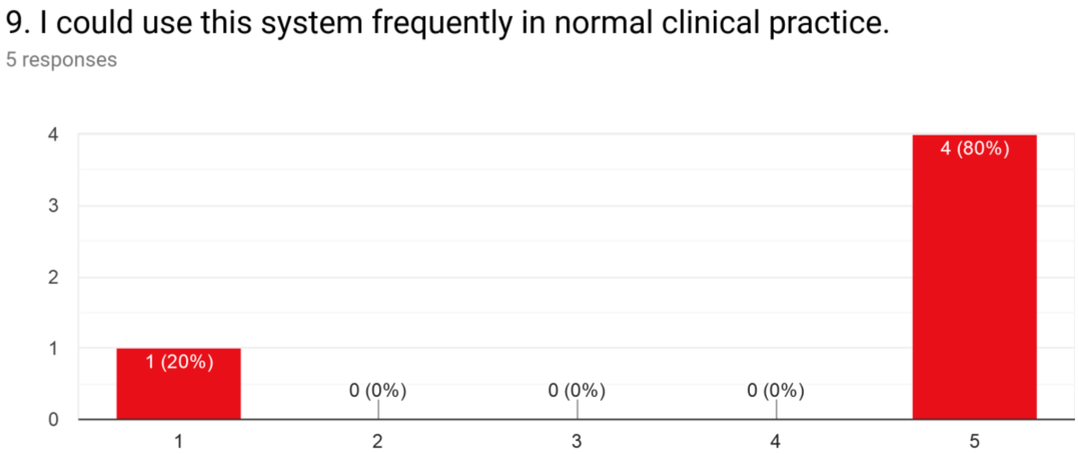


Figure 73 Utilisation of the system as it is in normal practice

Self-explanatory usage: most of the participants (80%) expressed a neutral position on whether they would need technical advice in order to use the system effectively (i.e. rate 3). There was one clinician who strongly agrees. The answers of this question are shown in Figure 74.

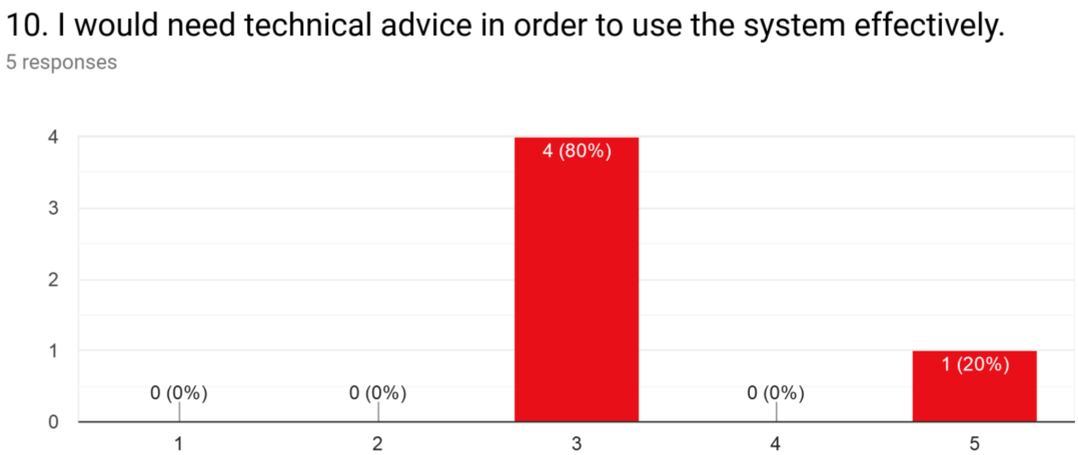


Figure 74 Self-explanatory usage

Guidance and instructions required for effective usage: most of the participants (80%) expressed a neutral position on whether they would need training in order to use the system effectively (i.e. rate 3). There was one clinician who strongly agrees. The answers of this question are shown in Figure 75.

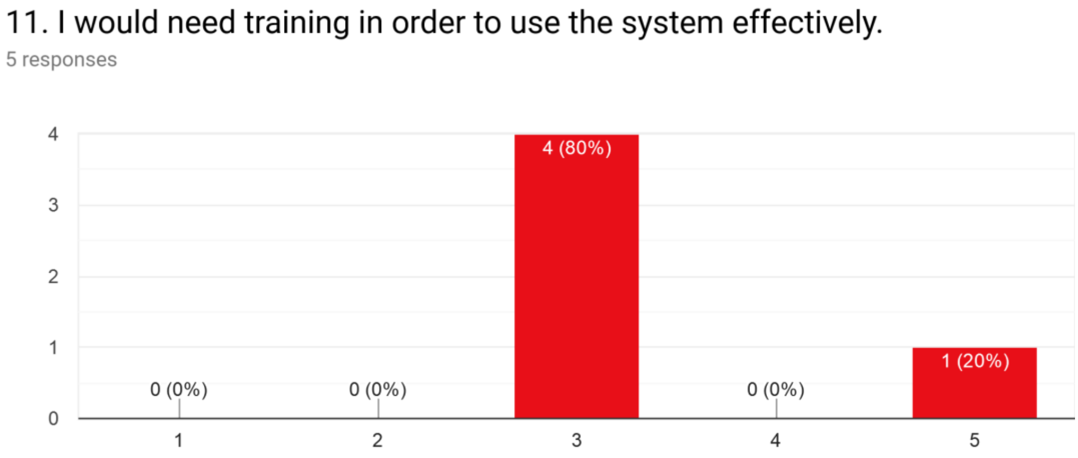


Figure 75 Guidance and instructions required for effective usage

How quickly a task can be completed in comparison to normal practice: most of the participants (80%) indicated that they strongly feel satisfied with the amount of time it took to complete a task in comparison to their previous working method (i.e. rate 5). There was one clinician who expressed a neutral position. The answers of this question are shown in Figure 76.

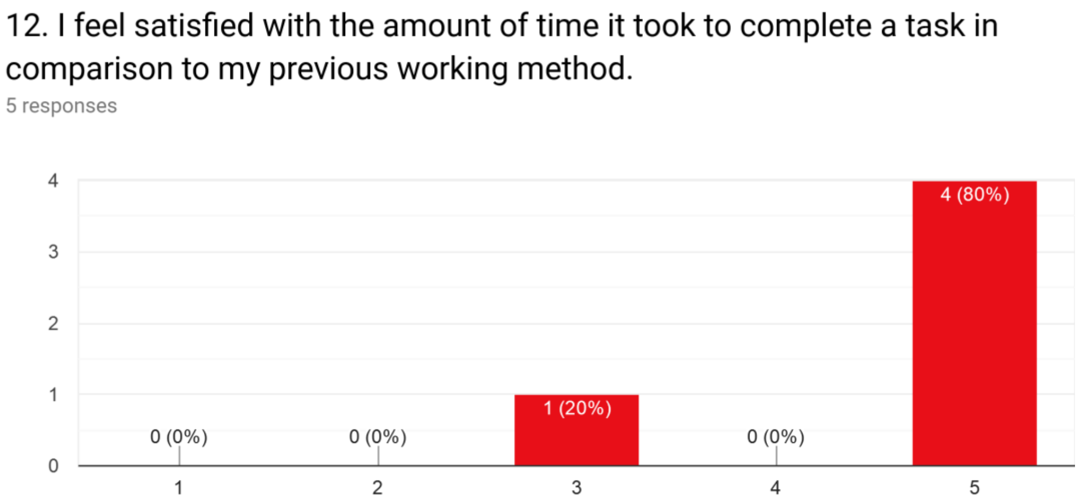


Figure 76 How quickly a task can be completed in comparison to normal practice

Usefulness

Effect of the platform on effectiveness: most of the participants (80%) strongly agree that the use of the system would help them to be more effective in their clinical practice (i.e. rate 5). There was one clinician who expressed a neutral position. The answers of this question are shown in Figure 77.

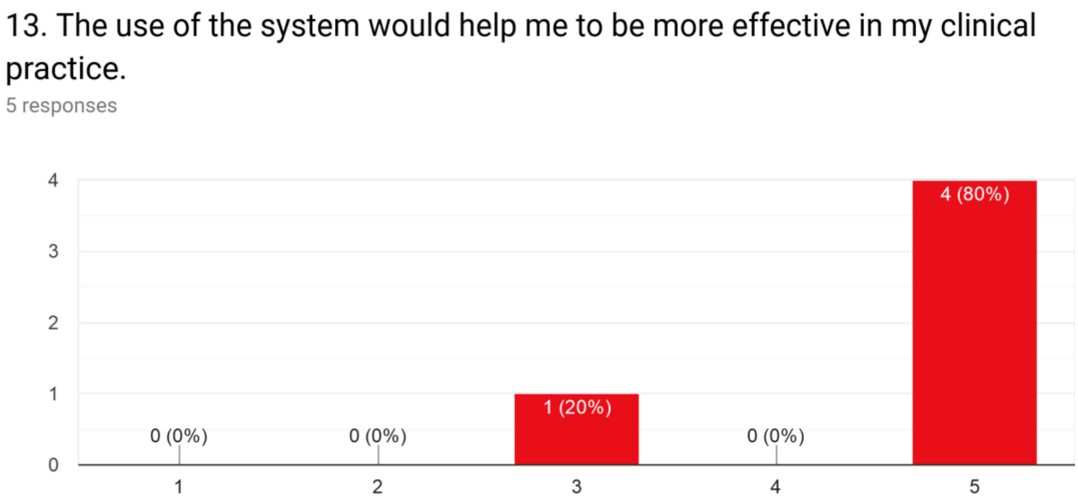


Figure 77 Effect of the platform on effectiveness

Effect of the platform on productivity: most of the participants (80%) strongly agree that the use of the system would help them to be more productive (i.e. rate 5). There was one clinician who expressed a neutral position. The answers of this question are shown in Figure 78.

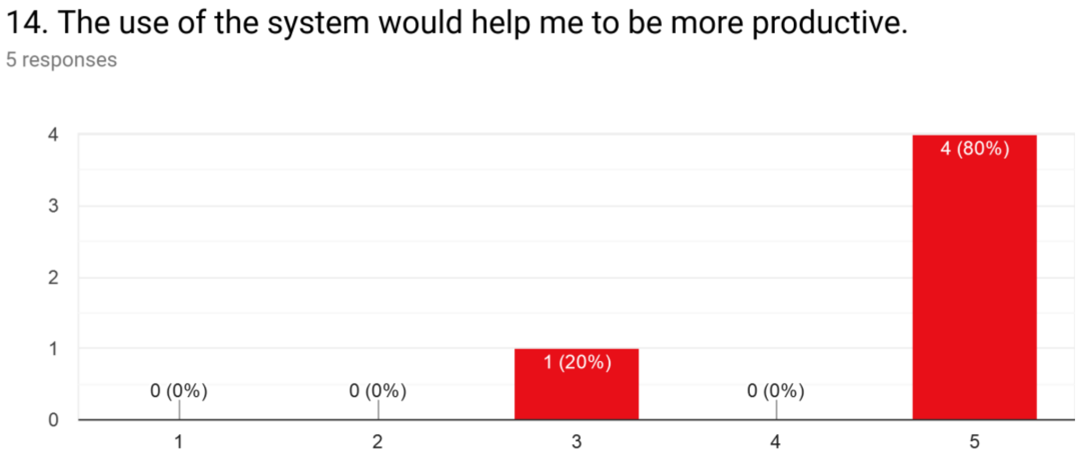


Figure 78 Effect of the platform on productivity

Success of platform in supporting an intended task: most of the participants (80%) indicated that the system is successful in performing its intended task (i.e. rate 4 or 5). There was one clinician who expressed a neutral position. The answers of this question are shown in Figure 79.

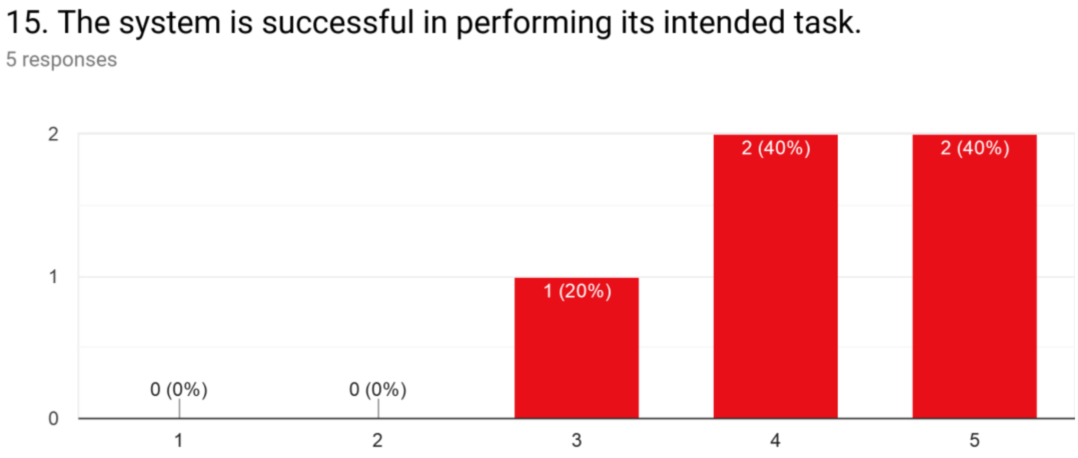


Figure 79 Success of platform in supporting an intended task

Meeting end-user expectations: 60% of the clinicians indicated that the system works the way they want it to work. There was one clinician who also agrees and another one who strongly disagrees. The answers of this question are shown in Figure 80.

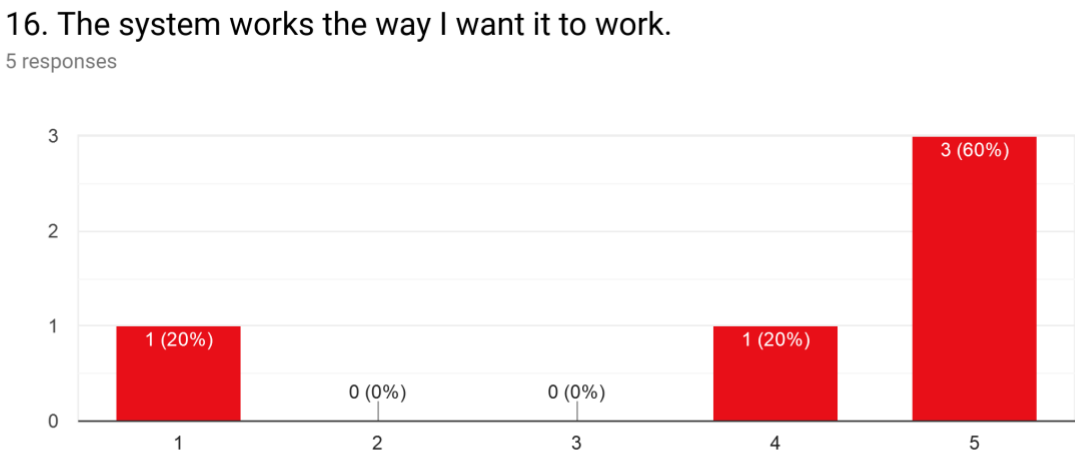


Figure 80 Meeting end-user expectations

Minimality of operation: most of the participants (80%) indicated that the system did not require taking unnecessary steps in order to accomplish what they wanted to do with it(i.e. rate 5). There was one clinician who expressed a neutral position. The answers of this question are shown in Figure 81.

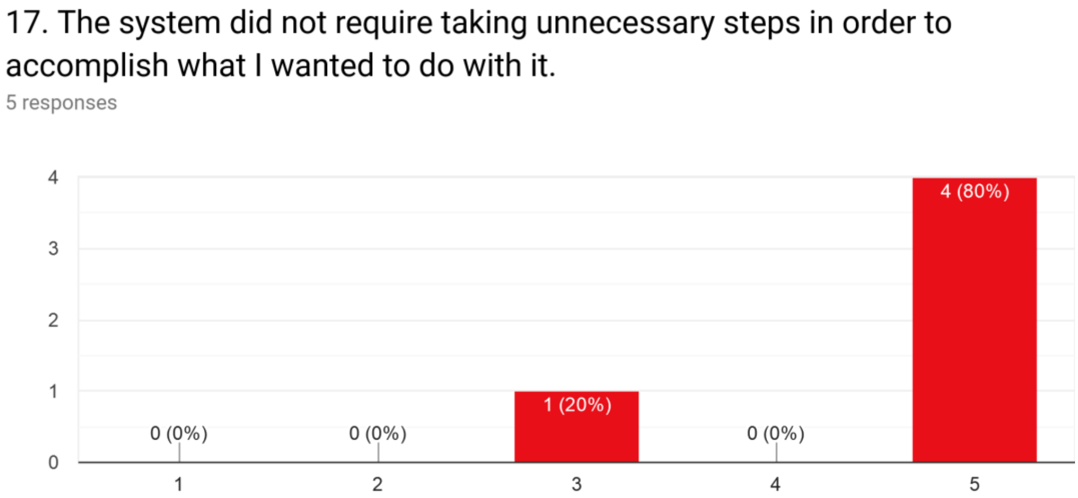


Figure 81 Minimality of operation

Meeting end-user needs: most of the participants (80%) indicated that the system meets their needs (i.e. rate 5). There was one clinician who disagrees. The answers of this question are shown in Figure 82.

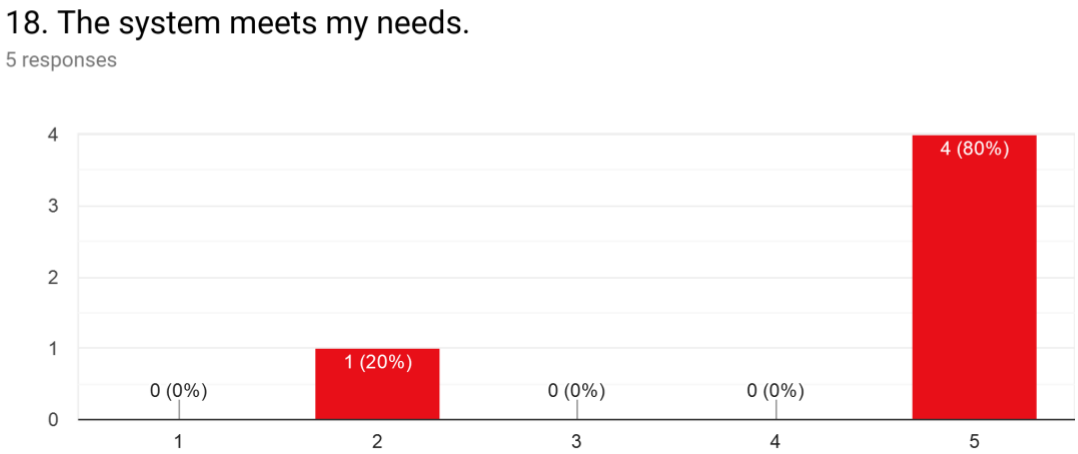


Figure 82 Meeting end-user needs

There was one clinician who left the following comment on the last question of the questionnaire: “The presentation of the system could have been more detailed. The example was not clearly presented neither the interpretation of the result”.

5.3.2.4. Results from the Questionnaire for Data Analysts

General Section

Enhancement of data analysis: all the experts (100%) agree (i.e. rate 4 or 5) that the described platform would enhance their data analysis (Figure 83).

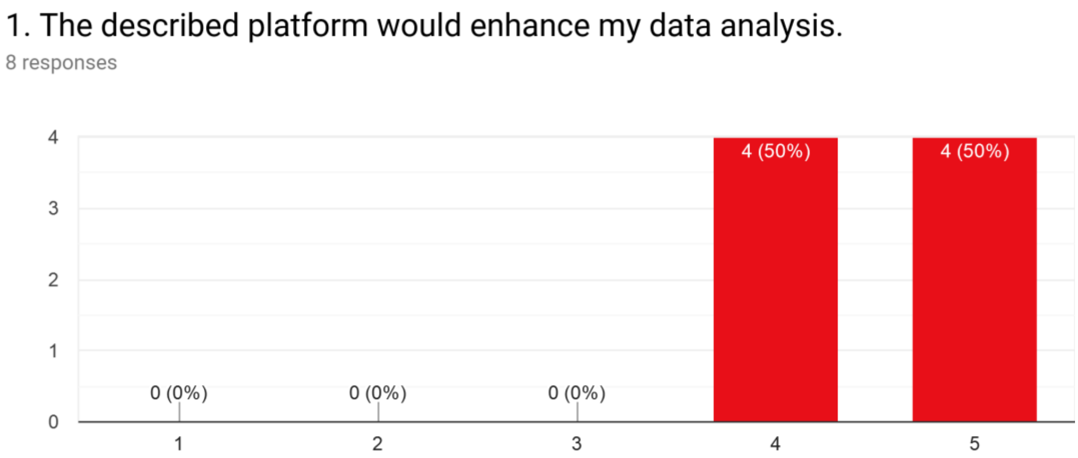


Figure 83 Enhancement of data analysis

Usability

Intuitiveness of user interface: all the experts (100%) agree (i.e. rate 4 or 5) that the user interface is intuitive (Figure 84).

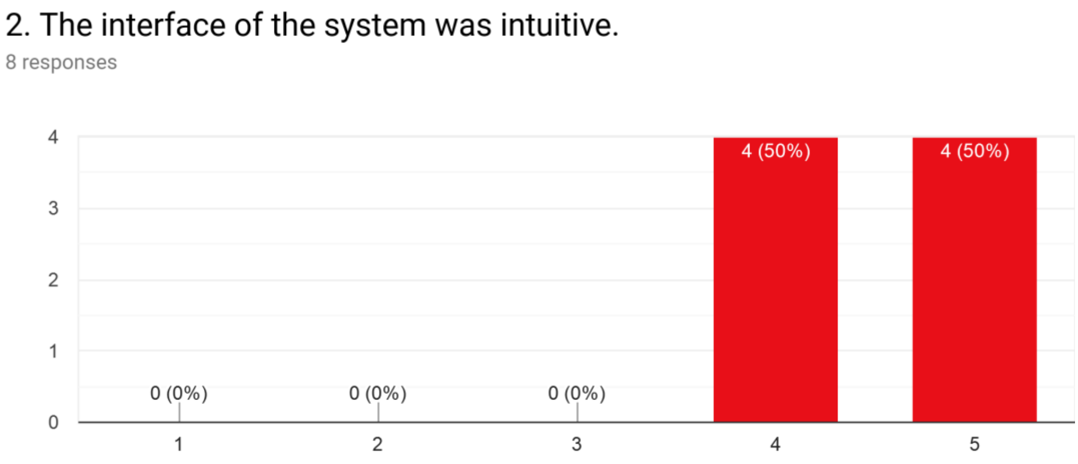


Figure 84 Intuitiveness of user interface

Easiness of task completion: all the participants (100%) agree (i.e. rate 4 or 5) that it was easy to complete tasks using the system (Figure 85).

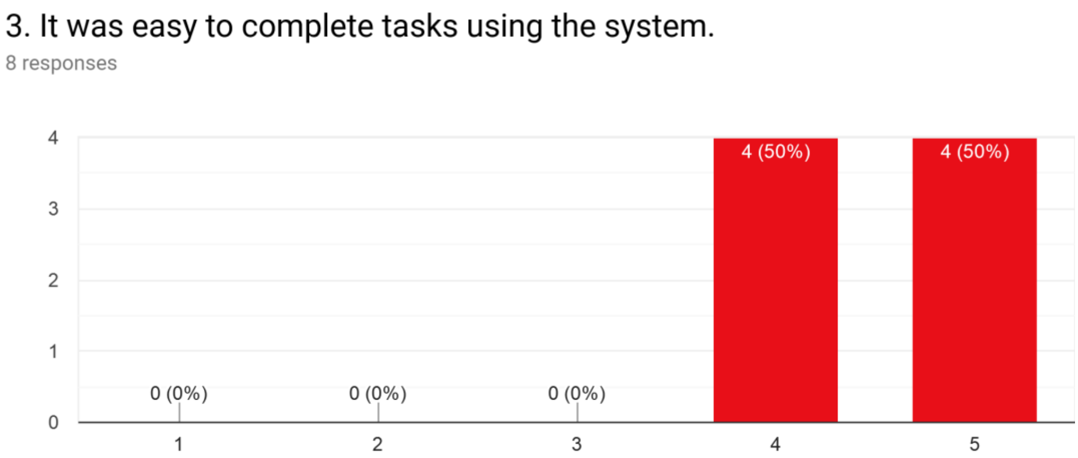


Figure 85 Easiness of task completion

User-friendly navigation: almost all the participants (88%) agree (i.e. rate 4 or 5) that navigation through the different options of the system was effective. There was one participant who expressed a neutral position. The answers of this question are presented in Figure 86.

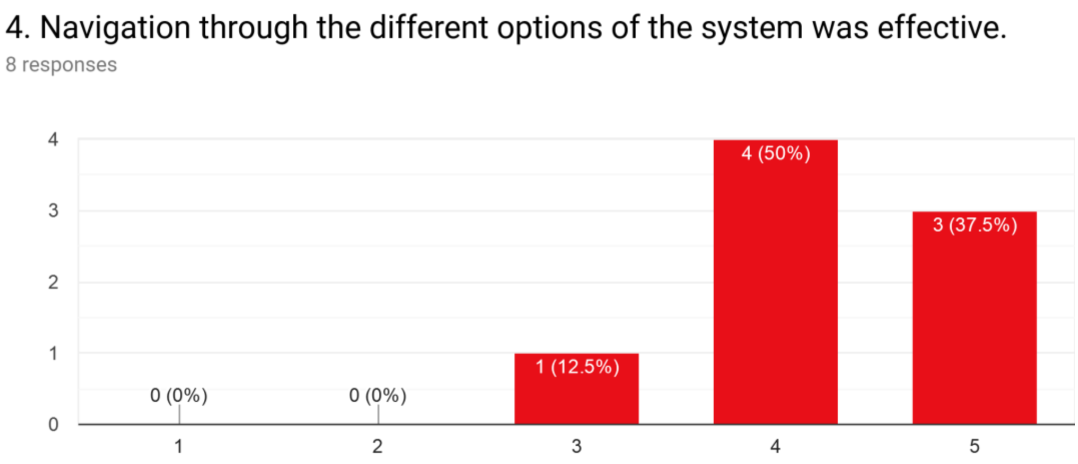


Figure 86 User-friendly navigation

User confidence: all the participants (100%) agree (i.e. rate 4 or 5) that they would feel confident using the system (Figure 87).

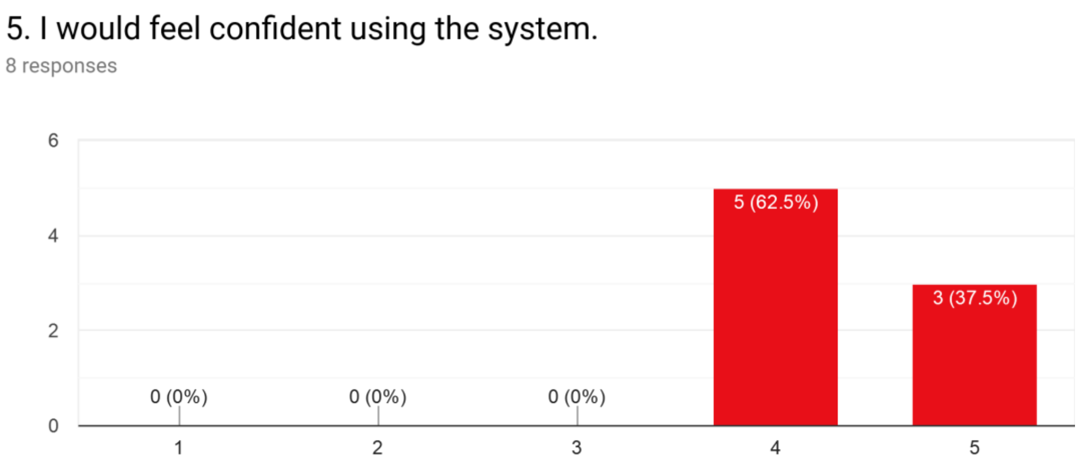


Figure 87 User confidence

Utilisation of the system as it is in normal practice: most of the participants (75%) agree (i.e. rate 4 or 5) that they could use this system frequently in their analyses. There was one expert that expressed a neutral position and another one who disagrees. The answers of this question are presented in Figure 88.

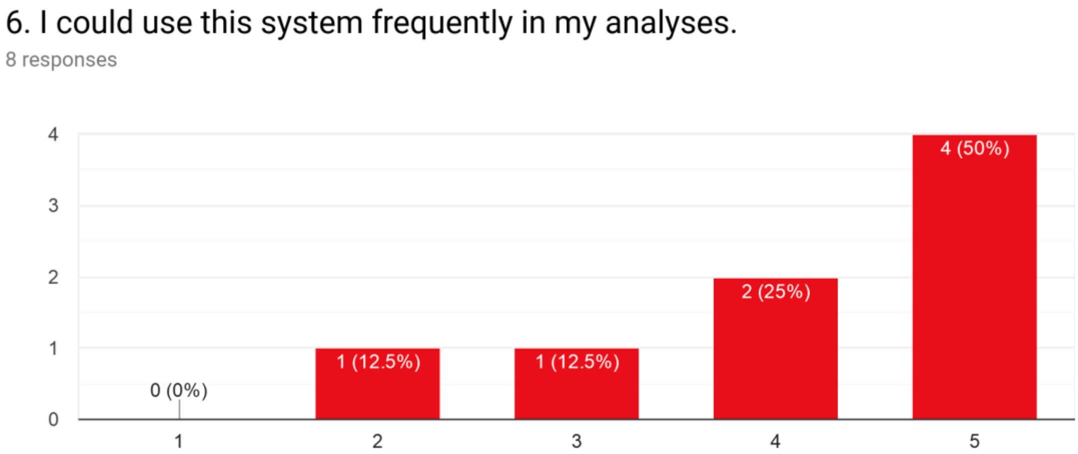


Figure 88 Utilisation of the system as it is in normal practice

Self-explanatory usage: the system was considered as self-explanatory, as it was anticipated, since 25% of the participants indicated that they would not need technical assistance to use it (i.e., rate 2 or less), 62,5% of them expressed a neutral view regarding it (i.e., rate 3) and one participant expressed a less positive view (i.e., rate 4) (Figure 89).

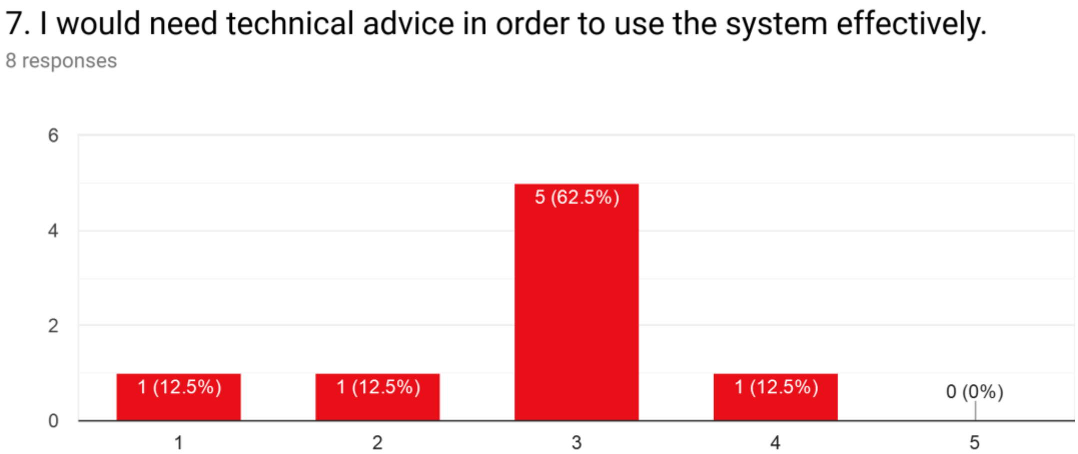


Figure 89 Self-explanatory usage

Guidance and instructions required for effective usage : one expert indicated that they would not need training in order to use the system effectively, 62,5% of them expressed a neutral view regarding it (i.e., rate 3) and 25% of the participants expressed a less positive view (i.e., rate 4 or more) (Figure 90).

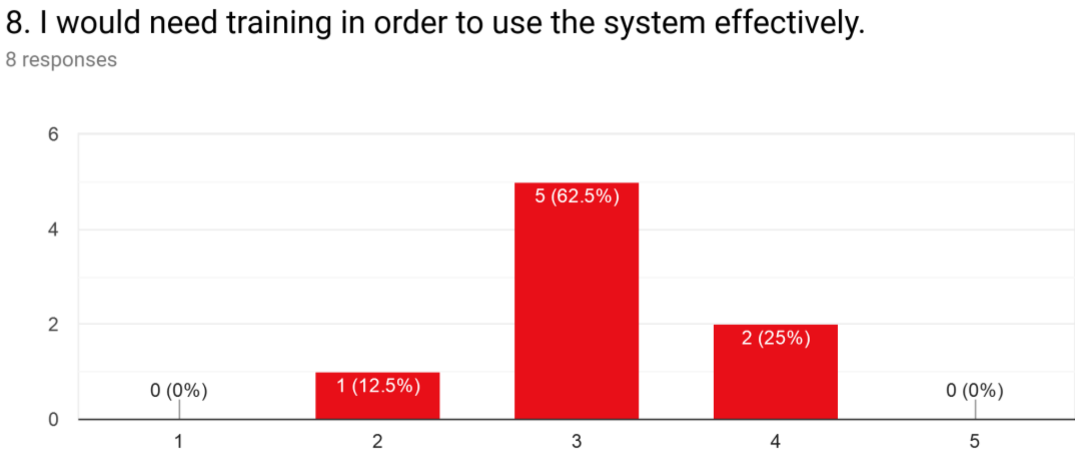


Figure 90 Guidance and instructions required for effective usage

How quickly a task can be completed in comparison to normal practice: all the participants (100%) agree (i.e. rate 4 or 5) that they feel satisfied with the amount of time it took to complete a task in comparison to their previous working method (Figure 91).

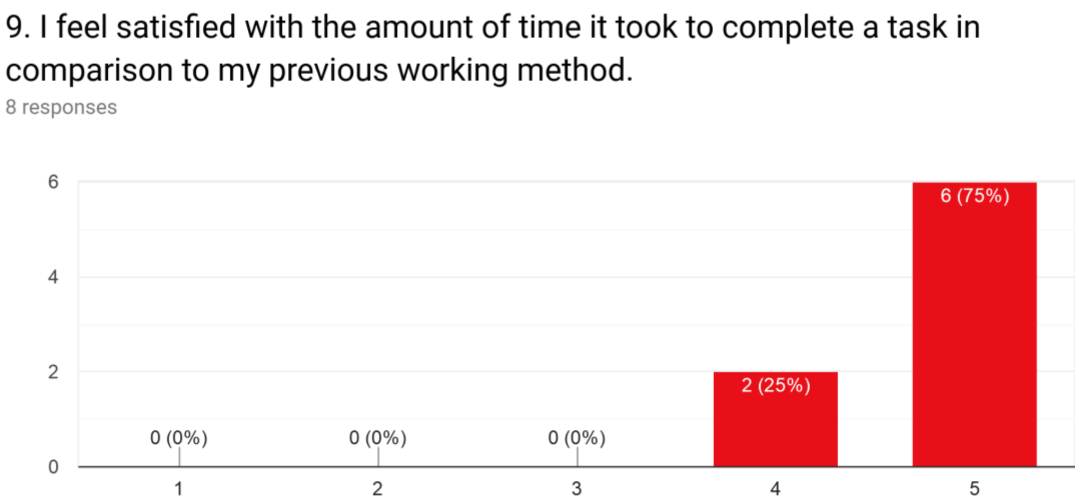


Figure 91 How quickly a task can be completed in comparison to normal practice

Usefulness

Effect of the platform on effectiveness: all the participants (100%) agree (i.e. rate 4 or 5) that the use of the system would help them to be more effective in producing data analytics (Figure 92).

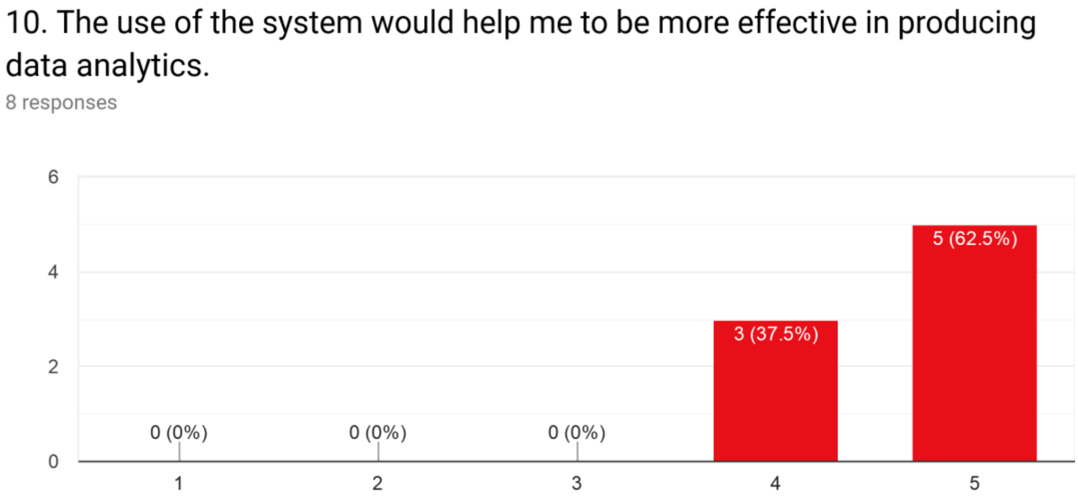


Figure 92 Effect of the platform on effectiveness

Effect of the platform on productivity: all the participants (100%) agree (i.e. rate 4 or 5) that the use of the system would help them to be more productive (Figure 93).

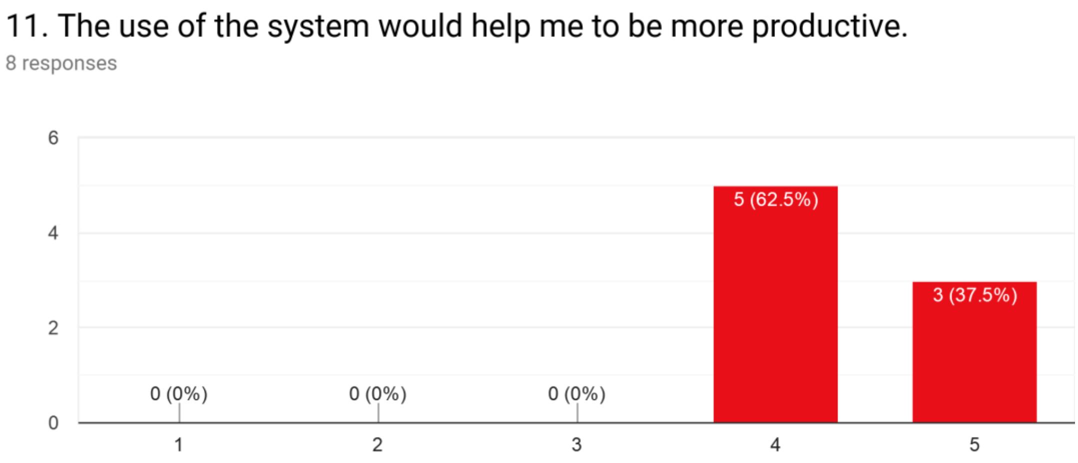


Figure 93 Effect of the platform on productivity

Success of platform in supporting an intended task: most of the participants (75%) agree (i.e. rate 4 or 5) that the use of the system is successful in performing its intended task. 25% of the participants expressed a neutral position. The answers of this question are presented in Figure 94.

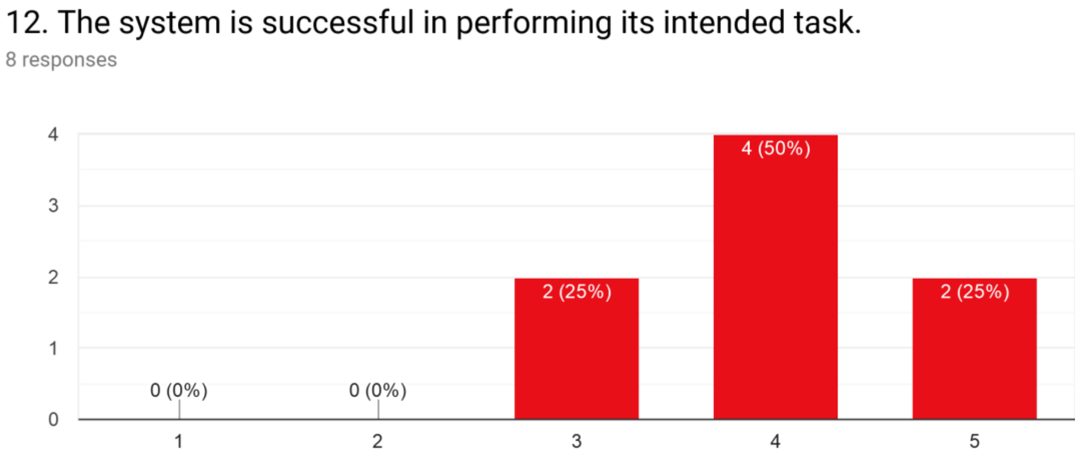


Figure 94 Success of platform in supporting an intended task

Meeting end-user expectations: all the participants (100%) agree (i.e. rate 4 or 5) that the system works the way they want it to work (Figure 95).

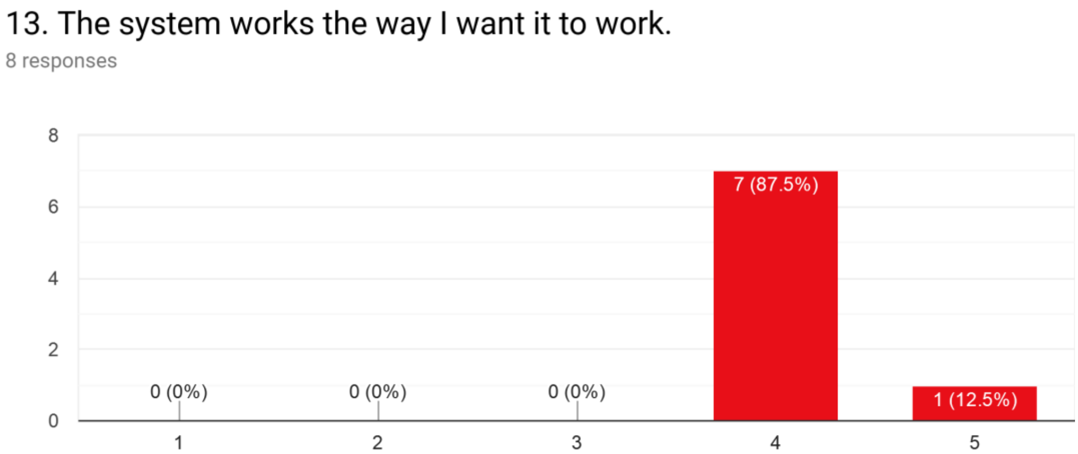


Figure 95 Meeting end-user expectations

Minimality of operation: all the participants (100%) agree (i.e. rate 4 or 5) that the system did not require taking unnecessary steps in order to accomplish what they wanted to do with it (Figure 96).

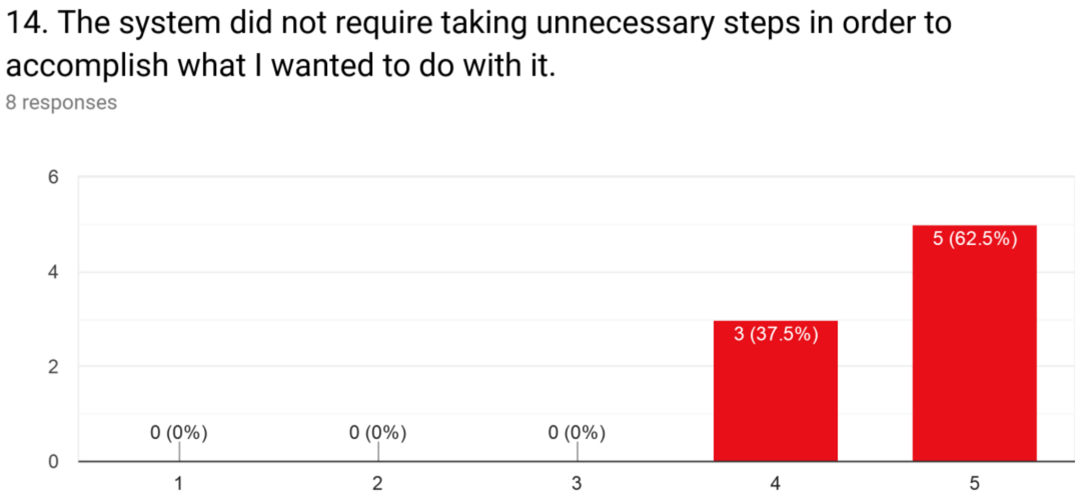


Figure 96 Minimality of operation

Meeting end-user needs: most of the participants (87,5%) agree (i.e. rate 4 or 5) that the system meets their needs. There was one participant who expressed a neutral position regarding this issue. The answers of this question are presented in Figure 97.

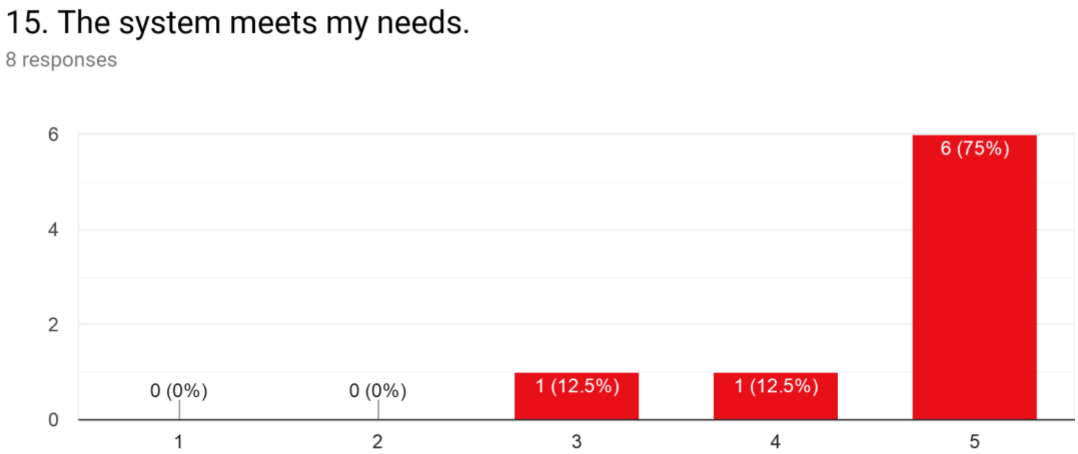


Figure 97 Meeting end-user needs

In the last question that asks for comments there was one answer as follows: "I'm not sure how I could use this tool for my work (for instance how to upload data for analysis)".

5.3.2.5. Comparison of Results

The following tables (Table 5-1, Table 5-2, Table 5-3) present and compare results from the three versions of the questionnaires.

Table 5-1 Comparison of General Sections of the three Questionnaires

General Section				
Question	**Version**	**Agree (5,4)**	**Neutral (3)**	**Disagree (2,1)**
Enhancement of normal practice	Policy Makers	86%	0%	14%
	Clinicians	80%	20%	0%
	Data Analysts	100%	0%	0%

Regarding the enhancement of normal practice, we see that most of the participants expressed positive views, as anticipated. Only 20% of the clinicians were neutral and 14% of the policy makers were negative. This is a very positive aspect for our platform and the language underpinning it.

Table 5-2 Comparison of Usability Sections of the three Questionnaires

Usability Section				
Question	**Version**	**Agree (5,4)**	**Neutral (3)**	**Disagree (2,1)**
Intuitiveness of user interface	Policy Makers	86%	0%	14%
	Clinicians	80%	20%	0%
	Data Analysts	100%	0%	0%
Easiness of task completion	Policy Makers	86%	0%	14%
	Clinicians	80%	20%	0%
	Data Analysts	100%	0%	0%
User-friendly navigation	Policy Makers	86%	0%	14%
	Clinicians	80%	0%	20%
	Data Analysts	87,5%	12,5%	0%
User confidence	Policy Makers	86%	0%	14%
	Clinicians	80%	0%	20%
	Data Analysts	100%	0%	0%
Utilisation of the system as it is in normal practice	Policy Makers	86%	0%	14%
	Clinicians	80%	0%	20%
	Data Analysts	75%	12,5%	12,5%
	Policy Makers	14%	57%	29%
	Clinicians	20%	80%	0%

Usability Section				
Question	**Version**	**Agree (5,4)**	**Neutral (3)**	**Disagree (2,1)**
Self-explanatory usage	Data Analysts	12,5%	62,5%	25%
Guidance and instructions required for effective usage	Policy Makers	14%	72%	14%
	Clinicians	20%	80%	0%
	Data Analysts	25%	62,5%	12,5%
How quickly a task can be completed in comparison to normal practice	Policy Makers	86%	0%	14%
	Clinicians	80%	20%	0%
	Data Analysts	100%	0%	0%

Regarding the usability section of our questionnaire, we see that most of the participants (in most of the questions 86% of the policy makers, 80% of the clinicians and 100% of the data analysts) express positive views. On the other hand, There are a lot of neutral views on the self-explanatory usage and guidance and instructions required for effective usage. This could be addressed with the addition of instructions to the system along with tutorial videos.

Table 5-3 Comparison of Usefulness Sections of the three Questionnaires

Usefulness Section				
Question	**Version**	**Agree (5,4)**	**Neutral (3)**	**Disagree (2,1)**
Effect of the platform on effectiveness	Policy Makers	86%	0%	14%
	Clinicians	80%	20%	0%
	Data Analysts	100%	0%	0%
Effect of the platform on productivity	Policy Makers	86%	0%	14%
	Clinicians	80%	20%	0%
	Data Analysts	100%	0%	0%
	Policy Makers	86%	0%	14%

Usefulness Section				
Question	**Version**	**Agree (5,4)**	**Neutral (3)**	**Disagree (2,1)**
Success of platform in supporting an intended task	Clinicians	80%	20%	0%
	Data Analysts	75%	25%	0%
Meeting end-user expectations	Policy Makers	86%	0%	14%
	Clinicians	80%	0%	20%
	Data Analysts	100%	0%	0%
Minimality of operation	Policy Makers	86%	0%	14%
	Clinicians	80%	20%	0%
	Data Analysts	100%	0%	0%
Meeting end-user needs	Policy Makers	86%	0%	14%
	Clinicians	80%	0%	20%
	Data Analysts	87,5%	12,5%	0%

Regarding the usefulness section of our questionnaires, that also reflects the usefulness of the modeling language underpinning the platform, we see that most of the participants (in most of the questions 86% of the policy makers, 80% of the clinicians and 100% of the data analysts) express positive views. This is a very positive aspect for our platform and the language underpinning it.

5.3.3. *Differences in Views of Different Types of Users*

This section explores whether the observed differences in the opinions of the different types of users for the policy language and the platform have a statistical significance, and if they have, the possible reasons that might have led to them.

Since the questionnaires were constructed as five-point Likert rating scales (1: strongly disagree to 5: strongly agree), the Mann-Whitney test[1] [102] was used to test the statistical significance of the differences. More specifically, the test was used to identify whether the users of different types (i.e., policy makers, clinicians and data analysis) expressed different views with regards to:

- The common questions of the questionnaire for policy makers and the questionnaire for clinicians (Test 1)
- The common questions of the questionnaire for policy makers and the questionnaire for data analysts (Test 2)
- The common questions of the questionnaire for clinicians and the questionnaire for data analysts (Test 3).

The results of these tests are summarised below.

Test 1

As Set A (sample size = 7): Participants of the questionnaire for policy makers.

As Set B (sample size = 5): Participants of the questionnaire for clinicians.

Test: whether two sample means are equal or not

Significance Level: 0,05

Two-tailed

Table 5-4 Mann-Whitney Test 1 Results

Mann-Whitney Test 1	
Question	**Result**
Enhancement of normal practice	The U-value is 16. The critical value of U at $p < .05$ is 5. Therefore, the result is not significant at $p < .05$.
Intuitiveness of user interface	The U-value is 17. The critical value of U at $p < .05$ is 5. Therefore, the result is not significant at $p < .05$.

[1] For calculations the online Mann-Whitney U Test Calculator was used: http://www.socscistatistics.com/tests/mannwhitney/default2.aspx

Mann-Whitney Test 1	
Question	**Result**
Easiness of task completion	The U-value is 17. The critical value of U at p < .05 is 5. Therefore, the result is not significant at p < .05.
User-friendly navigation	The U-value is 15.5. The critical value of U at p < .05 is 5. Therefore, the result is not significant at p < .05.
User confidence	The U-value is 17. The critical value of U at p < .05 is 5. Therefore, the result is not significant at p < .05.
Utilisation of the system as it is in normal practice	The U-value is 17. The critical value of U at p < .05 is 5. Therefore, the result is not significant at p < .05.
Self-explanatory usage	The U-value is 12. The critical value of U at p < .05 is 5. Therefore, the result is not significant at p < .05.
Guidance and instructions required for effective usage	The U-value is 14. The critical value of U at p < .05 is 5. Therefore, the result is not significant at p < .05.
How quickly a task can be completed in comparison to normal practice	The U-value is 16. The critical value of U at p < .05 is 5. Therefore, the result is not significant at p < .05.
Effect of the platform on effectiveness	The U-value is 16. The critical value of U at p < .05 is 5. Therefore, the result is not significant at p < .05.
Effect of the platform on productivity	The U-value is 12. The critical value of U at p < .05 is 5. Therefore, the result is not significant at p < .05.
Success of platform in supporting an intended task	The U-value is 17. The critical value of U at p < .05 is 5. Therefore, the result is not significant at p < .05.
Meeting end-user expectations	The U-value is 17. The critical value of U at p < .05 is 5. Therefore, the result is not significant at p < .05.
Minimality of operation	The U-value is 14. The critical value of U at p < .05 is 5. Therefore, the result is not significant at p < .05.
Meeting end-user needs	The U-value is 12.5. The critical value of U at p < .05 is 5. Therefore, the result is not significant at p < .05.

Test 2

As Set A (sample size = 7): Participants of the questionnaire for policy makers.

As Set B (sample size = 8): Participants of the questionnaire for data analysts.

Test: whether two sample means are equal or not

Significance Level: 0,05

Two-tailed

Table 5-5 Mann-Whitney Test 2 Results

Mann-Whitney Test 2	
Question	**Result**
Enhancement of normal practice	The U-value is 20. The critical value of U at p < .05 is 10. Therefore, the result is not significant at p < .05.
Intuitiveness of user interface	The U-value is 28. The critical value of U at p < .05 is 10. Therefore, the result is not significant at p < .05.
Easiness of task completion	The U-value is 28. The critical value of U at p < .05 is 10. Therefore, the result is not significant at p < .05.
User-friendly navigation	The U-value is 27.5. The critical value of U at p < .05 is 10. Therefore, the result is not significant at p < .05.
User confidence	The U-value is 25. The critical value of U at p < .05 is 10. Therefore, the result is not significant at p < .05.
Utilisation of the system as it is in normal practice	The U-value is 22.5. The critical value of U at p < .05 is 10. Therefore, the result is not significant at p < .05.
Self-explanatory usage	The U-value is 27.5. The critical value of U at p < .05 is 10. Therefore, the result is not significant at p < .05.
Guidance and instructions required for effective usage	The U-value is 25. The critical value of U at p < .05 is 10. Therefore, the result is not significant at p < .05.
How quickly a task can be completed in comparison to normal practice	The U-value is 26. The critical value of U at p < .05 is 10. Therefore, the result is not significant at p < .05.
Effect of the platform on effectiveness	The U-value is 27. The critical value of U at p < .05 is 10. Therefore, the result is not significant at p < .05.
Effect of the platform on productivity	The U-value is 27. The critical value of U at p < .05 is 10. Therefore, the result is not significant at p < .05.
Success of platform in supporting an intended task	The U-value is 23. The critical value of U at p < .05 is 10. Therefore, the result is not significant at p < .05.
Meeting end-user expectations	The U-value is 19. The critical value of U at p < .05 is 10. Therefore, the result is not significant at p < .05.

Mann-Whitney Test 2	
Question	**Result**
Minimality of operation	The U-value is 25. The critical value of U at p < .05 is 10. Therefore, the result is not significant at p < .05.
Meeting end-user needs	The U-value is 19.5. The critical value of U at p < .05 is 10. Therefore, the result is not significant at p < .05.

Test 3

As Set A (sample size = 5): Participants of the questionnaire for clinicians.

As Set B (sample size = 8): Participants of the questionnaire for data analysts.

Test: whether two sample means are equal or not

Significance Level: 0,05

Two-tailed

Table 5-6 Mann-Whitney Test 3 Results

Mann-Whitney Test 3	
Question	**Result**
Enhancement of normal practice	The U-value is 16. The critical value of U at p < .05 is 6. Therefore, the result is not significant at p < .05.
Intuitiveness of user interface	The U-value is 20. The critical value of U at p < .05 is 6. Therefore, the result is not significant at p < .05.
Easiness of task completion	The U-value is 20. The critical value of U at p < .05 is 6. Therefore, the result is not significant at p < .05.
User-friendly navigation	The U-value is 17.5. The critical value of U at p < .05 is 6. Therefore, the result is not significant at p < .05.
User confidence	The U-value is 18. The critical value of U at p < .05 is 6. Therefore, the result is not significant at p < .05.
Utilisation of the system as it is in normal practice	The U-value is 16. The critical value of U at p < .05 is 6. Therefore, the result is not significant at p < .05.
Self-explanatory usage	The U-value is 14. The critical value of U at p < .05 is 6. Therefore, the result is not significant at p < .05.
Guidance and instructions required for effective usage	The U-value is 18. The critical value of U at p < .05 is 6. Therefore, the result is not significant at p < .05.

Mann-Whitney Test 3	
Question	**Result**
How quickly a task can be completed in comparison to normal practice	The U-value is 20. The critical value of U at $p < .05$ is 6. Therefore, the result is not significant at $p < .05$.
Effect of the platform on effectiveness	The U-value is 18. The critical value of U at $p < .05$ is 6. Therefore, the result is not significant at $p < .05$.
Effect of the platform on productivity	The U-value is 14. The critical value of U at $p < .05$ is 6. Therefore, the result is not significant at $p < .05$.
Success of platform in supporting an intended task	The U-value is 17. The critical value of U at $p < .05$ is 6. Therefore, the result is not significant at $p < .05$.
Meeting end-user expectations	The U-value is 14. The critical value of U at $p < .05$ is 6. Therefore, the result is not significant at $p < .05$.
Minimality of operation	The U-value is 18. The critical value of U at $p < .05$ is 6. Therefore, the result is not significant at $p < .05$.
Meeting end-user needs	The U-value is 20. The critical value of U at $p < .05$ is 6. Therefore, the result is not significant at $p < .05$.

The outcome of these three tests provide evidence that the distributions of the three categories of the questionnaires are equal, thus selective distributions do not differentiate results.

5.3.4. Threats to validity

In the subjective evaluation of our platform and language different factors affected the results. The main limitation in this activity was that we did not have the number of results we expected. We sent the questionnaires to more than 10 people per category (policy makers, clinicians and data analysts), but got results only from 20 people totally.

Another issue is that most of our participants where colleagues, so they may have been more positive to the evaluation than we would like them to be. In order to prevent this, we asked them to make sure they fill in the questionnaires subjectively.

One more limitation was that, although we sent the questionnaires to stakeholders based in the UK, Bulgaria, Poland and Greece, we only got answers from people in Greece and one answer from the UK.

5.4. Performance Evaluation

5.4.1. The Overall Performance Evaluation Scenario

The performance evaluation was based on a case study involving the analysis of data for formulating potential public health policy models related to the management of hearing loss.

Hearing Loss is the most frequent sensory deficit and one of the most prevalent chronic disease reported by the elderly (according to the World Health Organisation (WHO), it affects approximately one-third of people over the age of 65 and over 5% of the world's population [103]). The consequences of HL in the overall health of people suffering from it are significant. More specifically, HL increases the risk of cognitive decline/dementia by 20% [104], mental illness [105], depression [105], [106], the risk of mortality [107], and the risk of accidental injury [108]). It is also the 8th most important disability with respect to the Years-Lived-with-Disability (YLD) indicator [109]. The economic consequences of HL are also significant since HL results in reduced productivity, unemployment or early retirement, loss of income and work discrimination [110]. Health economics studies also indicate that the treatment of HL has a significant cost. For example, according to [111], the annual cost of HL in the European Union is €213bn.

Currently, the pre-eminent management strategy for HL is the provision of hearing aids (HAs). Despite the fact that new generation HAs support a wide variety of advanced programming settings, literature suggests that older adults do not use these as they are less able to decide on complex circumstances and alternatives [112]. As a consequence, the majority (80%) of adults aged 55 to 74 years who would benefit from a HA, do not use them [112], and nearly 30% of HA users are dissatisfied with their HAs in noisy situations [113].

Motivated by the above factors, the case study that we used for the formation of example public health policy decision making models and the analysis of data based on them, was related to the management of hearing loss. Within this context, the data analytics workflows that were used for performance evaluation were focused on exploring whether the occupational context (working in noise, working in groups and occupation), educational level and the age of hearing aid (HA) users affects the daily usage of hearing aids.

The outcomes of such analysis could be used to inform policy making involving interventions, targeted to HA users of different occupations, HA users of different educational levels and HA users of different age groups.

5.4.2. Scenario Variables

For the purposes of our evaluation, the following HA user and HA usage variables, were used:

a) Working in noise (HA user variable). This variable indicates whether a HA user works in a noise environment or not. The variable is nominal and has the following possible values:
   - Yes
   - No

   In the case study this was used as an independent variable.

b) Working in groups (HA user attribute). This variable indicates whether a HA user works in groups or not. The variable is nominal and has the following possible values:
   - Yes
   - No

   In the case study this was used as an independent variable.

c) Occupation type. This variable indicates whether a HA user's occupation type is regular or part-time. The variable is nominal and has the following possible values:
   - Regular
   - Part-time

   In the case study this was used as an independent variable.

d) Educational Level. This variable is the educational level of a HA user. The variable is nominal and has the following possible values:

- Level1
- Level2
- Level3

In the case study this was used as an independent variable.

e) Age. This variable is age of a HA user. The variable is nominal and has the following variances:
   - less than 50
   - between 50 and 60
   - between 60 and 70
   - between 70 and 80
   - greater than or equal to 80

   In the case study this was used as an independent variable.

   Average daily usage. This variable contains the average daily usage of a HA user. The computed average daily usage is calculated in seconds. The possible maximum value is 57600 seconds, as we assume that each patient is awake for 16 hours per day. In the case study this was used as the dependent variable.

5.4.3. Data types and structures

The data used in the case study were available as part of a repository with the following tables:

(a) *Average Daily usage table*: This table stores the data stream of the average daily usage. The data are obtained by transforming data obtained directly from hearing aids used by users. The table has the following structure:

Table 5-7 Data Stream: Average Daily Usage

HA1 Id	HA2 id	Day	HA1 daily use	HA2 daily use

The table stores the daily usage of the two different HAs of a particular user. The two HA identifiers in the table are uniquely associated with a particular user (in another table). Column *Day* describes the date of the particular day of the recordings. Column *HA1 daily use* stores the daily use of the left HA of the user measured in seconds. Column *HA2 daily use* stores the daily use of the right HA of the user measured in seconds.

Our system will also be fed with “almost static” data of the hearing aid users. This data is presented in the table below. Column *Patient ID* includes the unique system generated identifier of the patient. Column *Working in Noise* indicates whether the patient is working in a noise environment or not. Column Working in Groups indicates whether the patient is working in groups. Columns *Occupation Type, Educational Level* and *Age* include the patient’s occupation type, educational level and age respectively.

Table 5-8 “Static” Data: Patient Details

Patient ID	Working in Noise	Working in Groups	Occupation Type	Educational Level	Age

This data stream is expected to be updated with lower frequency.

After performing a pre-processing task, we aggregate the data and transform it to a form presented in the table below:

Table 5-9 Data Stream: Average Daily Usage (1 month)

HA1 Id	HA2 id	Ave HA1 duse	Ave HA2 duse	Working in Noise	Working in Groups	Occupation	Edu Level	Age
1	-	322	-	Yes	No	Occup1	Level2	65
1	-	255	-	Yes	No	Occup1	Level2	65
1	-	321	-	Yes	No	Occup1	Level2	65
1	2	301	122	Yes	No	Occup1	Level2	65
1	2	489	184	Yes	No	Occup1	Level2	65
1	2	215	244	Yes	No	Occup1	Level2	65
1	2	211	0	Yes	No	Occup1	Level2	65
1	2	300	285	Yes	No	Occup1	Level2	65
1	2	241	255	Yes	No	Occup1	Level2	65

5.4.4. ***The Public Health Policy Decision Making Model for the evaluation***

The evaluation of performance was based on a public health decision making model, whose full specification in OWL is given in

Appendix C:

Generation of Synthetic Data. The main aspects and the analytics workflow of this model are summarised below.

GOAL: Addressing Barriers to HA Use

- Description: The purpose of this case study is to determine the largest barriers that affect hearing aid use in a population in order to make public health policy decisions to address them.
- Rationale: Barriers to hearing aid use are a significant public health problem. Barriers occur at all levels of the process of provision of hearing aids including at the level of the HA user. The big data gathered about users through EVOTION would enable policy makers to choose which barriers to address in a population, in order to improve hearing aid use and hence reduce the burden of hearing loss in that population.

OBJECTIVES:

- To intervene in order to address prevention of HA usage due to occupation.
- To intervene in order to address prevention of HA usage due to education level.
- To intervene in order to address prevention of HA usage due to age.

POLICY ACTIONS:

- Occupation Related ACTION: A particular occupation has to be addressed with additional measures to improve HA use;
- Educational level related ACTION: Failure to reach a particular educational level has to be addressed to improve HA use.
- Age related ACTION: Age related fitting;

STAKEHOLDERS

representatives of:

- Regional ENT-specialists' Advisory Committee (in their role as prescribing the use of HAs);
- Regional Directorate for Social support (in their role as authorising financial support for purchasing HAs and performing follow-up on administration and use);
- Regional structures of the national Health Insurance Fund (in their role as funding clinical pathways);
- HA vendors/fitting experts (providing follow-up rehab);
- Patients' association – regional repres. of patients

POSITIONS:

- Supportive_1, that refers to PA_1, proposed by Regional ENT-specialists' Advisory Committee and supported by them, as well as HA vendors/fitting experts.
- Opposing_1, that refers to PA_1, proposed by Regional Directorate for Social support and supported by them, as well as the Patients' association.
- Neutral_1, that refers to PA_1, proposed by Regional structures of the national Health Insurance Fund and supported by them,
- Supportive_2, that refers to PA_2, proposed by HA vendors/fitting experts and supported by them, as well as Regional ENT-specialists' Advisory Committee,
- Opposing_2, that refers to PA_2, proposed by Regional structures of the national Health Insurance Fund and supported by them, as well as the Regional Directorate for Social support,
- Neutral_2, that refers to PA_2, proposed by Patients' association and supported by them,
- Supportive_3, that refers to PA_3, proposed by Patients' association and supported by them,
- Opposing_3, that refers to PA_3, proposed by HA vendors/fitting and supported by them, as well as the Regional Directorate for Social support, and
- Neutral_3, that refers to PA_3, proposed by Regional structures of the national Health Insurance Fund and supported by them.

CRITERIA

Policy making criteria:

- Model wide
  - Statistically significant model overall
  - Homoscedasticity of prediction errors
- Variable specific
  - Effect of particular variable is statistically significant

- CR1: R2 square > 0.5 (constraints.spec.Reg_Overall_Stats.R Square>0.5)
- CR2: Edu_Level P < 0.05 (constraints.spec.Reg_Detailed_Stats [Factor = Edu_Level].P-value < 0.05)
- CR3: Age P < 0.05 (constraints.spec.Reg_Detailed_Stats [Factor = Age.P-value < 0.05)
- CR4: Occup P < 0.05 (constraints.spec.Reg_Detailed_Stats [Factor = Occup].P-value < 0.05)

Logical expressions of Criteria:

- EDU_CRIT: CR2 and CR1
- AGE_CRIT: CR3 and CR1
- OCCUP_CRIT: CR4 and CR1

WORKFLOW

The Data Analytics Workflow WF1 is composed of three `DataProcessingTasks`: two data filtering (select) tasks and one merge task and one statistical analysis task. Each of the two select produces as output a corresponding `DataStream`. Data streams are then merged by a different `DataProcessingTask`, utilizing a `JoinOperation`. The Join operation takes as parameters the data used to define policies (EDUCATION, AGE), and DAILY USAGE.

The Data Processing Task produces a Data stream, which is the input to the `StatisticalAnalysisTask`. The algorithm used for the `StatisticalAnalysisTask` is the `StatisticalRegressionAlgorithm`, which has as dataspecs the following `OutputDataSpecifications`: a PredictedValues Spec, a Regression Detailed Stats spec and a Regression overall stats spec. The Task produces as output four datasets with the above output data specifications. These datasets are constrained by the three criteria mentioned above in the scenario.

5.4.5. ***Synthetic Data***

To carry out the analytic process required for the performance evaluation, we generated random data from 1000000 binaural hearing aid users. We generated SQL statements for the insertion of 1000000 entries to the Patient table and the Q_DRMED table and the RETRO_HA table of the EVOTION Data Repository using python. The code is presented in

Appendix C:

Generation of Synthetic Data (Chapter 6 C.1). A sample of the SQL file generated is presented in

Appendix C:

Generation of Synthetic Data (Chapter 6 C.2). The nominal data that we generated follow a normal distribution and the numeric data a uniform distribution. The purpose of the data generation is to measure the time taken to retrieve the data from the EVOTION Data Repository and to run the linear regression to build the models of our evaluation scenario, so the random data fit our purposes. We decided to generate test data, because at the time of the experiments the EVOTION Data Repository did not have enough patient entries for the purposes of the evaluation of our tool as a big data analytics tool. The synthetic data may not have the same statistical characteristics as the real data, but they act as a good data set for our experiments, as what we are measuring is the time taken to retrieve them from the repository and the time taken to run the linear regression.

5.4.6. EVOTION Data

The analytics workflow specified in section 5.4.3.1 was also executed against real data obtained from the EVOTION project. At the time we ran our experiments the repository had 12967 patients and 31010 entries in TIME_PERIOD table. The patients that had both patient details and average daily usage data that we used were 238.

5.4.7. Execution

In our scenario, we execute linear regression to the dataset described in Section 5.2.

Execution Summary

Exploring the issue of our use case could be based on a process involving the following steps:

(i) The specification of PHP decision making model as an instance of the ontology introduced in Section IV to identify: the (policy) issue, the policy actions for addressing it, the data and the data analytics tasks that will be used to produce evidence from these data in order to explore each policy action, and the criteria for selecting amongst the different alternatives.

(ii) The execution of the data analytics tasks of the model specified in (i) and the recording of its outcomes as instances of the PHP decision making model.

This process is shown in Figure 98.

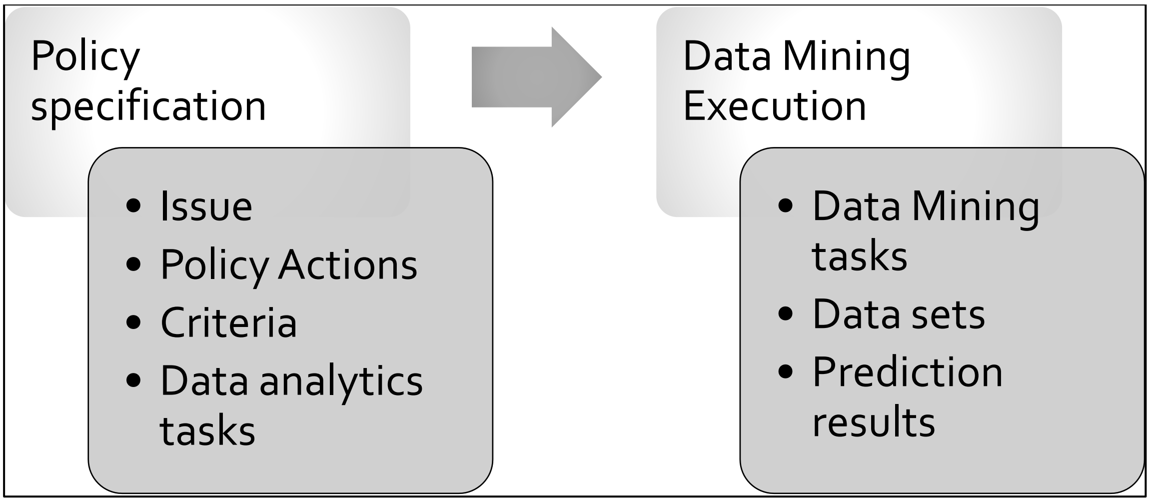


Figure 98 Execution Summary

5.4.8. Evaluation Process

During the evaluation, we measure (i) the time taken to retrieve the data from the database and to merge them and create the csv file that is the input of the data mining task and (ii) the time taken to run the linear regression task.

These two metrics help us evaluate the platform's performance as a big data analytics tool. It is important to evaluate the platform's efficiency in processing data analytics tasks. The choice of metrics is based on our capability to separate these two tasks and measure the performance of the processing of the data and execution of the data analytics task independently. Different volumes of data and different complexities of the data analytics task are investigated. We tenfold the volume of data from 1000 to 1000000 instances. We also include the case of 500000 instances. We change the complexity of the data analytics task of the experiment by changing the number of independent variables.

The two metrics are measured in the following cases:

(a) 1000 instances with one independent variable
(b) 1000 instances with two independent variables
(c) 1000 instances with three independent variables
(d) 1000 instances with four independent variables
(e) 1000 instances with five independent variables
(f) 10000 instances with one independent variable
(g) 10000 instances with two independent variables
(h) 10000 instances with three independent variables
(i) 10000 instances with four independent variables
(j) 10000 instances with five independent variables
(k) 100000 instances with one independent variable
(l) 100000 instances with two independent variables
(m) 100000 instances with three independent variables
(n) 100000 instances with four independent variables
(o) 100000 instances with five independent variables
(p) 500000 instances with one independent variable
(q) 500000 instances with two independent variables
(r) 500000 instances with three independent variables
(s) 500000 instances with four independent variables

(t) 500000 instances with five independent variables

(u) 1000000 instances with one independent variable

(v) 1000000 instances with two independent variables

(w) 1000000 instances with three independent variables

(x) 1000000 instances with four independent variables

(y) 1000000 instances with five independent variables

In section 5.4.9 we present two figures with the results of the above metrics and the experimental results of the above experiments.

In order to run our experiments, we used test04 server of the City, University of London Cluster. This server has the specs presented in Table 5-10.

Table 5-10: Server Specifications

Server Specifications	
Architecture	x86_64
CPU op-mode(s)	32-bit, 64-bit
Byte Order	Little Endian
CPU(s)	8
On-line CPU(s) list	0-7
Thread(s) per core	2
Core(s) per socket	4
Socket(s)	1
NUMA node(s)	1
Vendor ID	GenuineIntel
CPU family	6
Model	62
Stepping	4
CPU MHz	3500.050
BogoMIPS	7000.10
Virtualisation	VT-x
L1d cache	32K
L1i cache	32K
L2 cache	256K

Server Specifications	
L3 cache	15360K
NUMA node0 CPU(s)	0-7
Memory	16382400k
OS version	Ubuntu 12.04.4 LTS

5.4.9. Performance Results

In Table 5-11 we present the time taken to retrieve data from the repository and construct the intermediate CSV file for 1K, 10K, 100K, 500K and 1M rows of data

Table 5-11 Time taken to retrieve data from the repository

Experiment	Time taken in s
1K	2,6167
10K	2,5586
100K	4,2056
500K	10,4445
1M	16,7925

In Figure 99 we present the chart of the time taken to retrieve data from the repository and construct the intermediate CSV file for 1K, 10K, 100K, 500K and 1M rows of data.

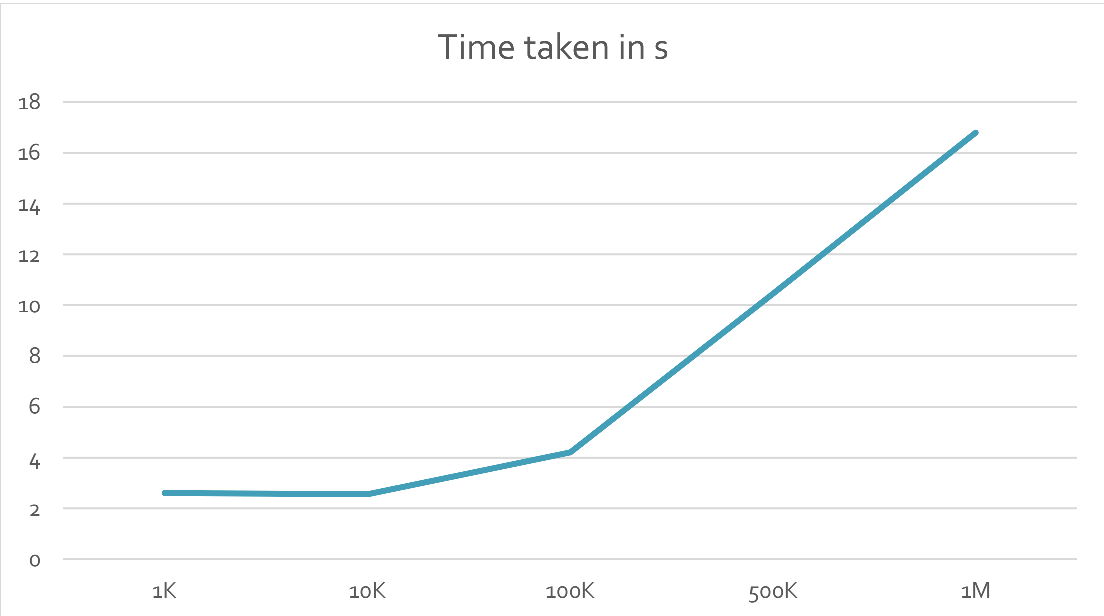


Figure 99 Time taken to retrieve data from the repository

In Table 5-12 we present the time taken to run linear regression to each case described in the Evaluation Process.

Table 5-12 Time taken to run Linear Regression

Experiment	Time taken in s
1K1V	0,4598
1K2V	0,465
1K3V	0,4762
1K4V	0,4937
1K5V	0,5232
10K1V	0,9245
10K2V	0,9906
10K3V	1,0222
10K4V	1,0235
10K5V	1,0687
100K1V	2,487
100K2V	2,555
100K3V	2,7058
100K4V	2,8408
100K5V	2,8845
500K1V	7,8916
500K2V	8,4783
500K3V	9,0195
500K4V	9,5025
500K5V	10,3028
1M1V	16,0161
1M2V	17,3317
1M3V	18,2021
1M4V	19,1872
1M5V	20,0999

In Figure 100 we present the chart of the time taken to run linear regression to each case described in the Evaluation Process. We also show the exponential trendline (dotted line).

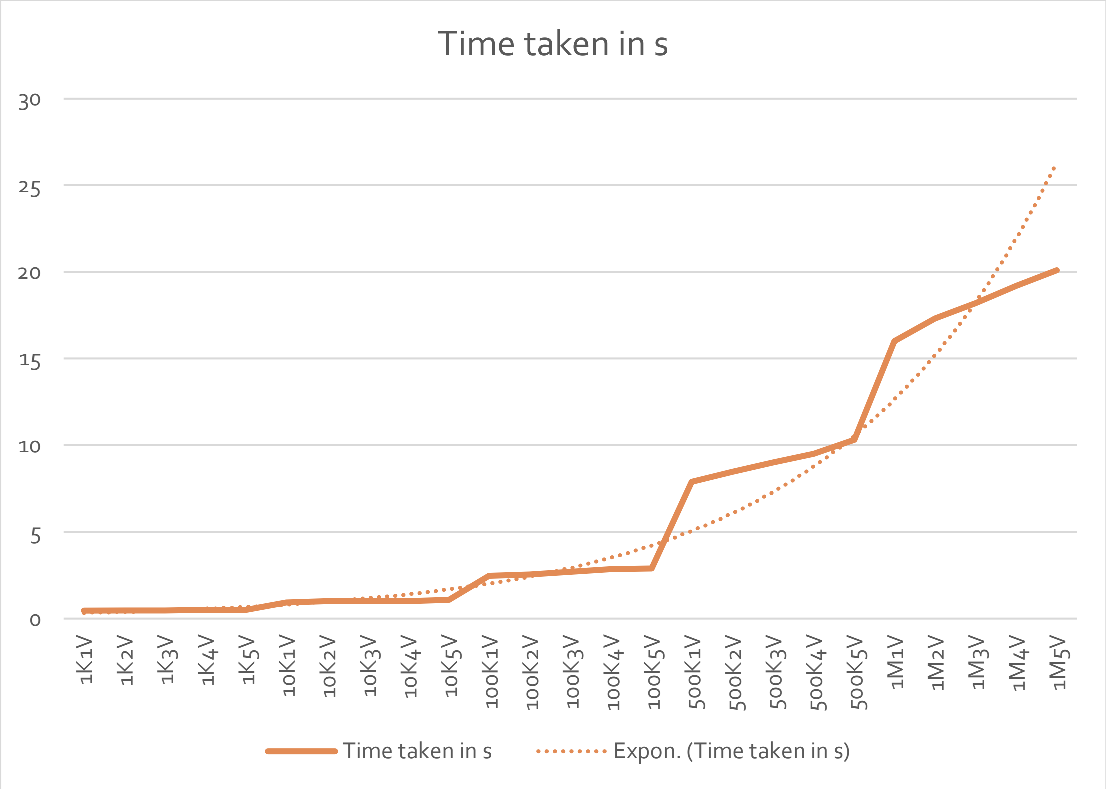


Figure 100 Time taken to run Linear Regression

In Figure 101 to Figure 105 we present the charts of the time taken in seconds to run linear regression to 1K up to 1M rows of data separately. We also show the exponential trendline for these separate charts, which is the orange dotted line.

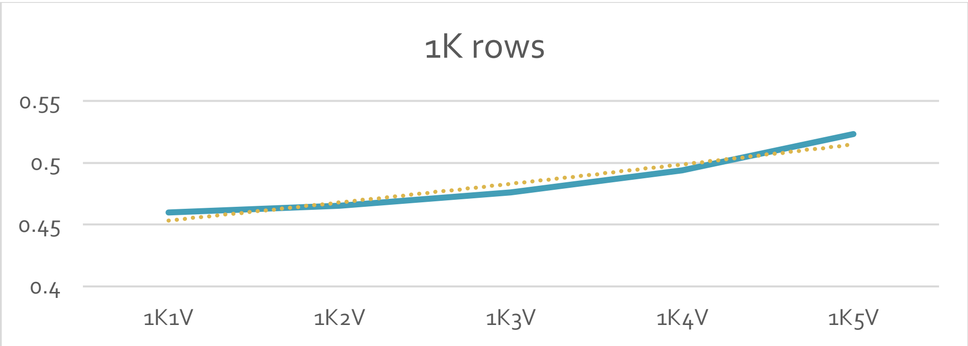


Figure 101 Time taken (sec) to run Linear Regression for 1000 rows

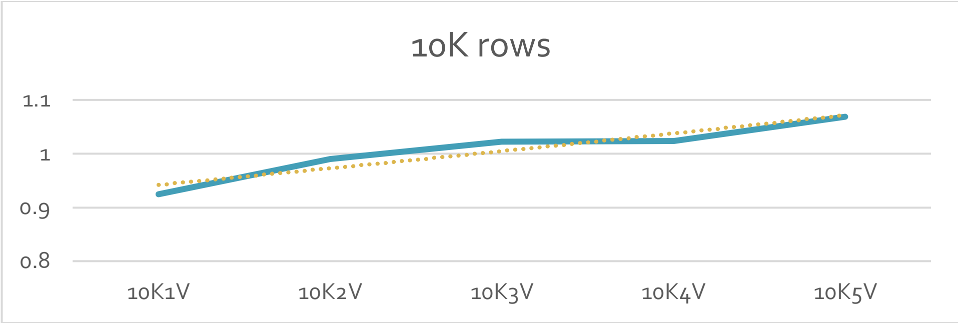


Figure 102 Time taken (sec) to run Linear Regression for 10000 rows

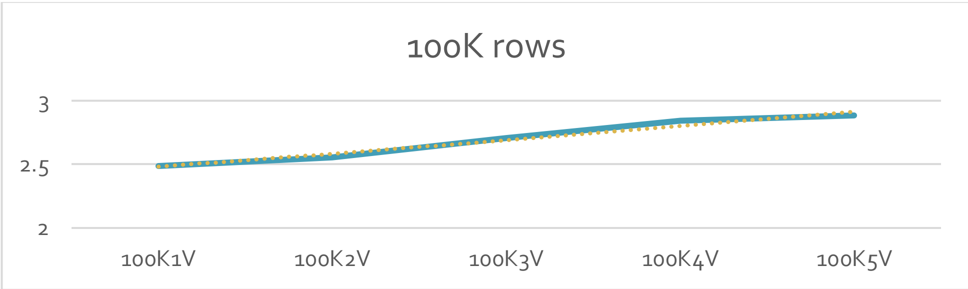


Figure 103 Time taken (sec) to run Linear Regression for 100000 rows

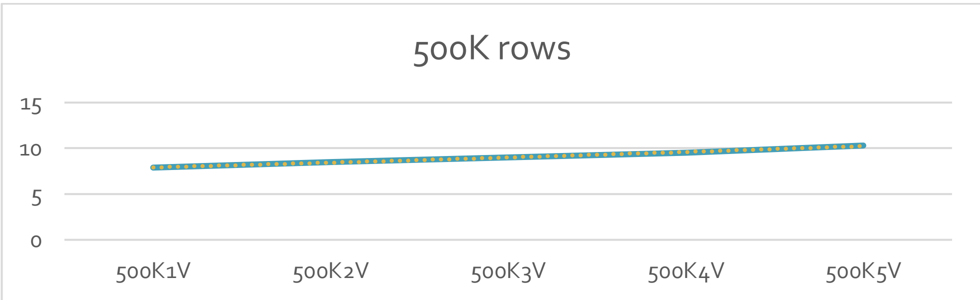


Figure 104 Time taken (sec) to run Linear Regression for 5000000 rows

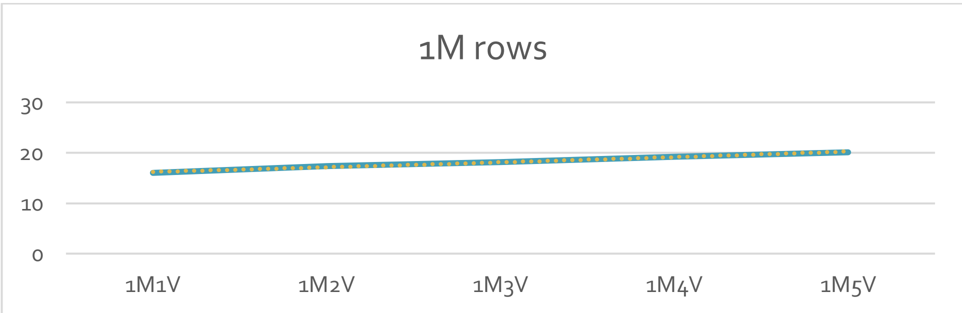


Figure 105 Time taken (sec) to run Linear Regression for 1000000 rows

In the figures above we see that the time taken to run linear regression for each case follows the exponential trendline, as expected.

5.4.10. *Threats to Validity*

In the experimental performance evaluation of our platform different factors affected the results. The main limitation in this activity was that we ran the queries to the EVOTION repository using Phoenix SQL. We figured out that this caused significant delays to the retrieval of the data.

Another limitation in this activity was that we ran the experiments in Spark local mode. In Spark cluster mode we expect to get better execution times for the linear regression execution. These times are dependent on the library we used to develop our BDA engine (DistributedWekaSpark).

Chapter 6

Conclusions

6.1. Overview

In this final chapter of the thesis, we point out the main novelties of the modeling language and the big data analytics engine presented in the earlier chapters and the contributions that our research has made to the state of the art. Moreover, the limitations of our research are also presented and directions for future work are discussed.

6.2. Contributions

6.2.1. Main Contributions

The main contributions of this thesis include the following:

- The provision of the ontology-based modeling language for the specification of PHPDM models. This modeling language is novel, as there is no similar and complete approach to PHPDM modeling. The language covers a wide range of policy making processes, data analytics workflows for the provision of evidence to the stakeholders, as well as the decision-making processes. This language is very useful to the policy makers as well as the data analysts, as it simplifies the PHPDM processes and their collaboration.
- The construction of a BDA engine to execute the data mining tasks and provide the evidence needed. The BDA engine we built for the purposes of the prototype that showcases the use of our language is novel, as it can process big datasets and perform data mining tasks fast and accurately, for the provision of the evidence. This is also very useful to the data analysts, as it enables them to explore the available data and perform the required analysis.

6.2.2. Platform Contributions to each Stakeholder

Below we describe how the proposed platform enhances the processes of each user of the system, namely policy makers, clinicians and data analysts.

6.2.2.1. Policy Makers

The platform enhances the policy making processes of the policy makers by giving them the capability to explore the data, find correlations, calculate important features extracted from the data (for example, the average daily usage per patient of hearing aids from the time periods of use), perform statistical analysis and machine learning algorithms to the data, define crucial criteria for the selection of the alternative policy actions and see which criteria are met, in order to support their decisions. Finally, the platform enables the policy makers to have a historical view of their policy models in order to be able to compare them and take more robust future decisions.

6.2.2.2. Clinicians

The platform enhances the clinical processes of the clinicians by giving them the capability to explore the data, find correlations, calculate important features extracted from the data (for example, the average daily usage per patient of hearing aids from the time periods of use), perform statistical analysis and machine learning algorithms to the data and take evidence based decisions in their clinical practice and explore whether the analysis results are equivalent to the clinical guidelines.

6.2.2.3. Data Analysts

The platform enhances the analyses of the data analysts by giving them the capability to explore the data, find correlations, calculate important features extracted from the data (for ex-ample, the average daily usage per patient of hearing aids from the time periods of use), per-form statistical analysis and machine learning algorithms to the data and view and compare the results of all the analyses performed and create evidence based reports with the analyses of the data.

6.3. Limitations

The platform we have developed for the specification of PHPDM models and the execution of their data analytics workflows has some limitations.

The main limitation is that the platform currently supports the initial stages of the policy formation processes, up to the development of an action plan. It does not support the monitoring of the plan’s implementation and its evaluation. To address that we need to extend the developed PHPDM modeling language and also construct a decision support system that will directly support the policy making process. More details are given in section 6.4 below.

A technical limitation is that for the execution of the workflows, an intermediate CSV file needs to be created. This is not very efficient regarding performance, but it was mandatory for the execution of the linear regression task of our performance evaluation scenario. On the other hand, this intermediate CSV file, makes it easier to connect to any SQL or noSQL database.

Furthermore, our proposed platform requires the collaboration of policy makers with data analysts, for the definition of the data analytics workflows. This could be an issue, because direct collaboration of policy makers with data analysts may be difficult or impossible in some cases. We tried to address this issue by separating the criteria definition from the workflow definition and by enabling the reusability of existing workflows.

Finally, in order to correctly define a complete PHPDM model, some training is needed, especially for the definition of the data analytics workflow and the criteria, as observed during the subjective evaluation of this thesis.

6.4. Future Work

This research proposed a platform that allows the specification of PHPDM models to support the decisions of policy makers, based on collected evidence. However, this approach can be further extended to support the full lifecycle of the policy formation processes. Therefore, some directions for future work are listed below:

- *The Public Health Policy Decision Making Modeling Language*

Although the developed PHPDM modeling language supports the definition and validation of PHPDM models to support the decision making processes of policy makers, there is space for further development and extension. We could add some more axioms to further formalize the language and add more validations for the specified PHPDM models. We could also add some more constructs to support the full life cycle of policy, namely, to include the monitoring of the implementation of the developed action plan and the programme evaluation.

- *The Public Health Policy Decision Making Specification Tool*

Although the existing PHPDM specification tool, developed by EVOTION, enables its stakeholders to specify data analytics workflows, criteria and policy models, it was observed during the subjective evaluation of this thesis, that technical advice and guidance is needed to help the users take advantage of the capabilities of the platform. We could add some template data analytics workflows, template criteria and template policy models, to be used by the stakeholders as predefined examples. We could also add tooltips to the interface, explaining the steps of the specification of data analytics workflows and policy models along with tutorial videos.

- *Interoperability*

Although we have developed a big data analytics engine for the execution of the specified PHPDM models and we have connected the prototype with EVOTION Data Repository, there are further interoperability capabilities. We could interconnect the platform with other existing data analytics tools, to enable the specification of PHPDM models from various sources and the reusability and execution of existing data analytics workflows in our platform. We could also connect the platform to various existing health data repositories, to extend the platform to support the formation of policy models for various other health conditions and patient cohorts. To this direction, we could use some standards for health care data exchange like FHIR[114].

- *Software as a Service*

Although the developed platform prototype serves the purposes of this thesis, some advances could be made for the developed platform to be offered as a service. We could extend the specification tool to have billing capabilities for each policy model specification and each data analytics workflow execution, depending on the complexity of the workflow and the volume of the data used for the analysis. We could also extend the user management and access controls to create user groups for stakeholders from the same organization to view only their common data.

- *The Big Data Analytics Engine*

Although the developed BDA engine serves the purposes of this thesis, we could proceed to the implementation of more data mining and statistical analysis algorithms to support even more types of analyses. Another task of future work would be to continuously update the types of analysis by adding new ones and removing deprecated ones.

- *A Decision Support System*

A last advancement that could be made to the work done for the purposes of this thesis could be the development of a decision support system that will enable the platform to not only support the execution of big data analytics, but to also directly support the policy making process.

Bibliography

[1] R. A. Stevens, C. E. Rosenberg, and L. R. Burns, Eds., *History and Health Policy in the United States*. Rutgers University Press, 2006.

[2] WHO, "Health policy." [Online]. Available: http://www.who.int/topics/health_policy/en/. [Accessed: 10-Sep-2018].

[3] C. J. E. Metcalf, W. J. Edmunds, and J. Lessler, "Six challenges in modelling for public health policy," *Epidemics*, 2015.

[4] T. S. Weeramanthri and R. S. Bailie, "Grand Challenges in Public Health Policy," *Front. Public Heal.*, vol. 3, no. February, pp. 8–11, 2015.

[5] World Health Organization, "Key Policy Issues in Long-Term Care," *World Heal. Organ.*, pp. 139–190, 2003.

[6] M. I. Wallhagen, "Access to care for hearing loss: Policies and stakeholders," *J. Gerontol. Nurs.*, vol. 40, no. 3, pp. 15–19, 2014.

[7] G. Spanoudakis, P. Katrakazas, D. Koutsouris, D. Kikidis, A. Bibas, and N. H. Pontopidan, "Public Health Policy for Management of Hearing Impairments Based on Big Data Analytics: EVOTION at Genesis," *2017 IEEE 17th Int. Conf. Bioinforma. Bioeng.*, pp. 525–530, 2017.

[8] O. A. Arah, "On the relationship between individual and population health," *Med. Heal. Care Philos.*, vol. 12, no. 3, pp. 235–244, 2009.

[9] E. Just, "Understanding Risk Stratification , Comorbidities , and the Future of Healthcare," *Heal. Catal.*, pp. 1–7, 2017.

[10] Apache, "HBase," *The Apache Software Foundation*, 2017. .

[11] Canadian Foundation of Healthcare Improvement, "Accelerating Healthcare Improvement : Canadian Foundation for Healthcare Improvement ' s Assessment Tool (CFHI Assessment Tool TM)," 2014.

[12] NHS Executive London, "A Short Guide to Health Impact Assessment Informing Healthy Decisions," p. , 2000.

[13] B. Nilunger, L., Elinder, L.S. and Pettersson, "HIA: screening of Swedish governmental inquiries. Europe," *Eurohealth (Lond).*, vol. 8, no. 5, 2003.

[14] World Health Organisation, "WHO | Health systems: Governance," *Who*, 2013. [Online]. Available: http://www.who.int/healthsystems/topics/stewardship/en/. [Accessed: 28-Feb-2018].

[15] C. Buffett, D. Ciliska, and H. Thomas, "Can I Use This Evidence in my Program Decision? Assessing Applicability and Transferability of Evidence," no. November, pp. 1–22, 2007.

[16] National Collaborating Centre for Methods and Tools, "Applicability and Transferability of Evidence Tool (A&T Tool)," no. 2007, pp. 1–3, 2009.

[17] M. Dobbins, "Rapid Review Guidebook Steps for conducting a rapid review," p. 25, 2017.

[18] M. Dobbins, "Rapid Review Guidebook Steps for conducting a rapid review," p. 25, 2017.

[19] M. Linstone, Harold A. Turoff, "Delphi Method: Techniques and Applications," *1975*, p. 640, 1975.

[20] A. Delbecq, A. Van de Ven, and D. Gustafson, "The Delphi Technique," in *Group Techniques*

for Program Planning: A Guide to Nominal Group and Delphi Processes, 1975.

[21] WHO, “Planning & monitoring of national strategies: a manual,” 2015.

[22] Deloitte Access Economics, “Evaluation framework for adult hearing services in England,” *Deloitte Access Econ.*, no. January, 2013.

[23] WHO, “Ear and Hearing Care Situation Analysis Tool,” 2015.

[24] E. N. Loukis, “An ontology for G2G collaboration in public policy making, implementation and evaluation,” *Artif. Intell. Law*, vol. 15, no. 1, pp. 19–48, 2007.

[25] M. Perakath, P., Menzel, C., Mayer, R., Fillion, F., Futrell, M., DeWitte, P. and Lingineni, “IDEF5 Method Report.,” 1994.

[26] M. Hilario, P. Nguyen, H. Do, A. Woznica, and A. Kalousis, “Ontology-based meta-mining of knowledge discovery workflows,” *Stud. Comput. Intell.*, vol. 358, pp. 273–315, 2011.

[27] T. Schaul and J. Schmidhuber, “Metalearning,” *Scholarpedia*, 2010.

[28] C. M. Keet, A. Ławrynowicz, C. D’Amato, and M. Hilario, “Modeling issues and choices in the Data Mining OPtimization Ontology,” in *CEUR Workshop Proceedings*, 2013, vol. 1080.

[29] A. Gangemi, N. Guarino, C. Masolo, A. Oltramari, and L. Schneider, “Sweetening ontologies with DOLCE,” *Knowl. Eng. Knowl. Manag. Ontol. Semant. Web, Lect. Notes Comput. Sci. vol. 2473*, pp. 223–233, 2002.

[30] G. H. Bakir, T. Hofmann, B. Schölkopf, A. J. Smola, B. Taskar, and S. V. N. Vishwanathan, “Predicting Structured Data. Neural Information Processing.” The MIT Press Cambridge, 2007.

[31] P. Panov, L. Soldatova, and S. Džeroski, *Ontology of core data mining entities*, vol. 28, no. 5–6. 2014.

[32] B. Smith and P. Grenon, “Basic formal ontology (bfo),” *INFOMIS Reports*, 2006.

[33] B. Smith *et al.*, “The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration,” *Nat. Biotechnol.*, vol. 25, p. 1251, Nov. 2007.

[34] A. Bandrowski *et al.*, “The Ontology for Biomedical Investigations,” *PLoS One*, 2016.

[35] W. Ceusters, “An information artifact ontology perspective on data collections and associated representational artifacts,” in *Studies in Health Technology and Informatics*, 2012.

[36] P. Panov, “A MODULAR ONTOLOGY OF DATA MINING,” *Dr. Diss.*, 2002.

[37] S. Džeroski, “Towards a General Framework for Data Mining,” *Knowl. Discov. Inductive Databases, 5th Int. KDID 2006*, pp. 259–300, 2007.

[38] D. Hand, D. Hand, H. Mannila, H. Mannila, P. Smyth, and P. Smyth, *Principles of data mining*, vol. 30. 2001.

[39] R. Caruana, “Multitask Learning,” *Mach. Learn.*, vol. 28, no. 1, pp. 41–75, 1997.

[40] G. Tsoumakas, I. Katakis, and A. Overview, “Multi-Label Classification : An Overview,” *Int. J. Data Warehous. Min.*, vol. 3, no. September, pp. 1–13, 2007.

[41] I. Slavkov, V. Gjorgjioski, J. Struyf, and S. Dzeroski, “Finding explained groups of time-course gene expression profiles with predictive clustering trees.,” *Mol. Biosyst.*, vol. 6, pp. 729–740, 2010.

[42] C. N. Silla and A. A. Freitas, “A survey of hierarchical classification across different application domains,” *Data Mining and Knowledge Discovery*, vol. 22, no. 1–2. pp. 31–72,

2011.

[43] D. Demšar *et al.*, “Using multi-objective classification to model communities of soil microarthropods,” in *Ecological Modelling*, 2006, vol. 191, no. 1, pp. 131–143.

[44] D. Kocev, S. Džeroski, M. D. White, G. R. Newell, and P. Griffioen, “Using single- and multi-target regression trees and ensembles to model a compound index of vegetation condition,” *Ecol. Modell.*, vol. 220, no. 8, pp. 1159–1168, 2009.

[45] J. Zheng, M. R. Harris, A. M. Masci, Y. Lin, A. Hero, and Y. He, “OBCS : The Ontology of Biological and Clinical Statistics,” p. 2014, 2014.

[46] J. Zheng *et al.*, “The Ontology of Biological and Clinical Statistics (OBCS) for standardized and reproducible statistical analysis,” *J. Biomed. Semantics*, 2016.

[47] Y. He *et al.*, “Ontology representation and ANOVA analysis of vaccine protection investigation,” in *CEUR Workshop Proceedings*, 2010.

[48] R Development Core Team, “R: A Language and Environment for Statistical Computing,” *R Found. Stat. Comput. Vienna Austria*, 2016.

[49] J. M. Chambers, *Programming with data: A guide to the S language*. Springer Science & Business Media, 1998.

[50] M. Hofmann; and R. Klinkenberg;, “RapidMiner: Data Mining Use Cases and Business Analytics Applications,” *Zhurnal Eksp. i Teor. Fiz.*, 2013.

[51] J. Demšar *et al.*, “Orange: Data Mining Toolbox in Python,” *J. Mach. Learn. Res.*, 2013.

[52] M. R. Berthold, N. Cebron, F. Dill, T. R. Gabriel, T. Kötter, and B. Meinl T.... & Wiswedel, “KNIME-the Konstanz information miner: version 2.0 and beyond.,” *AcM SIGKDD Explor. Newsl.*, 2009.

[53] S. O’Hagan, N. Swainston, J. Handl, and D. B. Kell, “Erratum to A ‘rule of 0.5’ for the metabolite-likeness of approved pharmaceutical drugs [Metabolomics, DOI 10.1007/s11306-014-0733-z],” *Metabolomics*. 2015.

[54] T. A. S. Foundation, “Apache Mahout: Scalable machine learning and data mining,” *Apache Mahout*, 2016. .

[55] X. Meng *et al.*, “MLlib: Machine Learning in Apache Spark,” 2015.

[56] M. Zaharia *et al.*, “Apache Spark: a unified engine for big data processing,” *Commun. ACM*, vol. 59, no. 11, pp. 56–65, 2016.

[57] M. A. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, “The WEKA data mining software: an update,” *SIGKDD Explor.*, vol. 11, no. 1, pp. 10–18, 2009.

[58] M. A. Hall, “Distributed Weka Spark.” [Online]. Available: http://weka.sourceforge.net/packageMetaData/distributedWekaSpark/index.html.

[59] A. Alexandrov *et al.*, “The Stratosphere platform for big data analytics,” *VLDB J.*, vol. 23, no. 6, pp. 939–964, 2014.

[60] P. Carbone, S. Ewen, S. Haridi, A. Katsifodimos, V. Markl, and K. Tzoumas, “Apache Flink(TM): Stream and Batch Processing in a Single Engine,” *Bull. Tech. Comm. Data Eng.*, 2015.

[61] C. L. Vinayak R. Borkar, Michael J. Carey, “Big Data Platforms: What’s Next?,” *ACM Crossroads*, vol. 19, no. 19, pp. 44–49, 2012.

[62] S. Alsubaiee *et al.*, “AsterixDB: A Scalable, Open Source BDMS,” 2014.

[63] OMG, "Unified Modelling Language (UML), Version 2.5," 2015. [Online]. Available: http://www.omg.org/spec/UML/2.5/PDF. [Accessed: 08-Nov-2017].

[64] F. Baader, D. Calvanese, D. L. McGuinness, D. Nardi, and P. F. Patel-Schneider, "The Description Logic Handbook: Theory, Implementation, and Applications," *Descr. Log. Handb.*, p. 622, 2003.

[65] B. Motik, I. Horrocks, and U. Sattler, "Bridging the gap between OWL and relational databases," *Web Semant.*, vol. 7, no. 2, pp. 74–89, 2009.

[66] W3C, "OWL2/XML Syntax W3C Website," 2012.

[67] T. L. Saaty, "What is the Analytic Hierarchy Process?," in *Mathematical Models for Decision Support*, 1988, pp. 109–121.

[68] V. Chýna, M. Kuncová, and J. Sekni, "Estimation of weights in multi-criteria decision-making optimization models," *Proc. 30th Int. Conf. Math. Methods Econ.*, pp. 355–360, 2013.

[69] "STATO Ontology." [Online]. Available: http://stato-ontology.org.

[70] G. H. G. John and P. Langley, "Estimating Continuous Distributions in Bayesian Classifiers," *Proc. Elev. Conf. Uncertain. Artif. Intell. Montr. Quebec, Canada*, vol. 1, pp. 338--345, 1995.

[71] D. J. C. Mackay, "Introduction to Gaussian processes," *Neural Networks Mach. Learn.*, vol. 168, no. 1996, pp. 133–165, 1998.

[72] I. H. Witten, E. Frank, and M. a. Hall, *Data Mining: Practical Machine Learning Tools and Techniques, Third Edition*, vol. 54, no. 2. 2011.

[73] S. Le Cessie, J. C. Van Houwelingen, and R. S. Society, "Ridge Estimators in Logistic Regression," *J. R. Stat. Soc. Ser. C (Applied Stat.*, vol. 41, no. 1, pp. 191–201, 1992.

[74] D. W. Aha, D. Kibler, and M. K. Albert, "Instance-Based Learning Algorithms," *Mach. Learn.*, vol. 6, no. 1, pp. 37–66, 1991.

[75] R. Kohavi, "The power of decision tables," *Mach. Learn. ECML-95*, pp. 174–189, 1995.

[76] J. Quinlan, *C4. 5: programs for machine learning*, vol. 240. 1993.

[77] L. Breiman, "Random Forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[78] A. McCallum, K. Nigam, and L. H. Ungar, "Efficient clustering of high-dimensional data sets with application to reference matching," *Proc. sixth ACM SIGKDD Int. Conf. Knowl. Discov. data Min. KDD 00*, pp. 169–178, 2000.

[79] D. Fisher, "Knowledge acquisition via incremental clustering," *Mach. Learn.*, vol. 2, no. 1980, pp. 139–182, 1987.

[80] J. H. Gennari, P. Langley, and D. Fisher, "Models of incremental concept formation," *Artif. Intell.*, vol. 40, no. 1–3, pp. 11–61, 1989.

[81] D. S. Hochbaum and D. B. Shmoys, "A Best Possible Heuristic for the k-Center Problem," *Math. Oper. Res.*, vol. 10, no. 2, pp. 180–184, 1985.

[82] S. Dasgupta and P. M. Long, "Performance guarantees for hierarchical clustering," in *Journal of Computer and System Sciences*, 2005, vol. 70, no. 4, pp. 555–569.

[83] D. Arthur and S. Vassilvitskii, "k-means++: the advantages of careful seeding," in *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*, 2007, pp. 1027–1025.

[84] L. Richardson and S. Ruby, *RESTful Web Services*. 2008.

[85] C. Pautasso, E. Wilde, and R. Alarcon, *REST: Advanced research topics and practical*

applications. 2013.

[86] Stanford Center for Biomedical Informatics Research, “Protégé Website,” 2016. [Online]. Available: http://protege.stanford.edu.

[87] and I. H. W. Eibe Frank, Mark A. Hall, “The WEKA Workbench. Online Appendix,” in *Data Mining: Practical Machine Learning Tools and Techniques*, Fourth Edi., Morgan Kaufmann, 2016.

[88] K. Shvachko, H. Kuang, S. Radia, and R. Chansler, “The Hadoop distributed file system,” in *2010 IEEE 26th Symposium on Mass Storage Systems and Technologies, MSST2010*, 2010.

[89] Apache, “HBase,” *The Apache Software Foundation*, 2013. .

[90] Z. S. Nasreddine *et al.*, “The Montreal Cognitive Assessment, MoCA: A brief screening tool for mild cognitive impairment,” *J. Am. Geriatr. Soc.*, 2005.

[91] J. Horsman, W. Furlong, D. Feeny, and G. Torrance, “The Health Utilities Index (HUI®): Concepts, measurement properties and applications,” *Health and Quality of Life Outcomes*. 2003.

[92] S. Gatehouse, “Glasgow Hearing Aid Benefit Profile: derivation and validation of a client-centered outcome measure for hearing aid services,” *J. Am. Acad. Audiol.*, 1999.

[93] A. S. Zigmond and R. P. Snaith, “The Hospital Anxiety and Depression Scale,” *Acta Psychiatr. Scand.*, 1983.

[94] Apache, “HBase,” *The Apache Software Foundation*, 2013. [Online]. Available: http://hbase.apache.org/.

[95] S. Akhtar and R. Magham, “Using Phoenix,” in *Pro Apache Phoenix*, Springer, 2017, pp. 15–35.

[96] “Auditbase.” .

[97] D. Axmark and M. Widenius, *MySQL 5.6 Reference Manual*. 2016.

[98] “NOAH 4.” .

[99] R. Shearer, B. Motik, and I. Horrocks, “HermiT: A highly-eficient OWL reasoner,” in *CEUR Workshop Proceedings*, 2009.

[100] A. M. Lund, “Measuring usability with the USE questionnaire,” *Soc. Tech. Commun.*, 2001.

[101] J. D. Gould and C. Lewis, “Designing for usability---key principles and what designers think,” in *Proceedings of the SIGCHI conference on Human Factors in Computing Systems - CHI ’83*, 1983.

[102] J. C. F. J.de Winter and D. Dodou, “Five-Point Likert Items : t test versus Mann-Whitney-Wilcoxon,” *Pract. Assessment, Res. Eval.*, 2010.

[103] WHO, “WHO | Deafness and hearing loss,” *Fact sheet N°300*, pp. 1–5, 2014.

[104] F. R. Lin, E. J. Metter, R. J. O’Brien, S. M. Resnick, A. B. Zonderman, and L. Ferrucci, “Hearing loss and incident dementia.,” *Arch. Neurol.*, vol. 68, no. 2, pp. 214–20, Feb. 2011.

[105] S. Mathews, N. Abrahams, and R. Jewkes, “Exploring mental health adjustment of children post sexual assault in South Africa,” *J. Child Sex. Abus.*, vol. 22, no. 6, pp. 639–657, 2013.

[106] R. N. Davis, M. M. Davis, G. L. Freed, and S. J. Clark, “Fathers’ Depression Related to Positive and Negative Parenting Behaviors With 1-Year-Old Children,” *Pediatrics*, vol. 127, no. 4, pp. 612–618, 2011.

[107] T. J. Holwerda *et al.*, “Increased risk of mortality associated with social isolation in older men: only when feeling lonely? Results from the Amsterdam Study of the Elderly (AMSTEL).,” *Psychol. Med.*, vol. 42, no. 4, pp. 843–53, Apr. 2012.

[108] F. R. Lin and L. Ferrucci, “Hearing loss and falls among older adults in the United States,” *Arch. Intern. Med.*, vol. 172, no. 4, pp. 369–371, 2012.

[109] H. A. Whiteford *et al.*, “Global burden of disease attributable to mental and substance use disorders: Findings from the Global Burden of Disease Study 2010,” *Lancet*, vol. 382, no. 9904, pp. 1575–1586, 2013.

[110] S. Kochkin, “MarkeTrak VIII: The efficacy of hearing aids in achieving compensation equity in the workplace,” *Hear. J.*, vol. 63, no. 10, pp. 19–28, 2010.

[111] W. Kaplan, V. J. Wirtz, A. Mantel-Teeuwisse, P. Stolk, B. Duthey, and R. Laing, “Priority Medicines for Europe and the World, 2013 Update,” *WHO 2013*, 2013. .

[112] A. McCormack and H. Fortnum, “Why do people fitted with hearing aids not wear them?,” *Int. J. Audiol.*, vol. 52, no. 5, pp. 360–8, May 2013.

[113] S. Kochkin, “A Comparison of Consumer Satisfaction, Subjective Benefit, and Quality of Life Changes Associated with Traditional and Direct-mail Hearing Aid Use | Hearing Review,” 2014. .

[114] D. Bender and K. Sartipi, “HL7 FHIR: An Agile and RESTful approach to healthcare information exchange,” in *Proceedings - IEEE Symposium on Computer-Based Medical Systems*, 2013.

Appendix A: Questionnaires

In this appendix we present the questionnaires used for the subjective evaluation of this thesis (section 5.3), one for Policy Makers (section A.1), one for Clinician (section A.2) and one for Data Analysts (section A.3). We include screenshots of the forms along with the scripts of the videos that were embedded in the questionnaires.

We also include the consent forms we prepared for the interviews (section A.4) and the questionnaires (section A.5) along with the participant information sheets we distributed to each participant of the interviews (section A.6) and the questionnaires (section A.7).

For the purposes of the interviews we prepared an Interview Topic Guide presented in section A.8.

Finally, we attach the Ethics Application (section A.9) submitted at City, University of London for the ethics approval along with the formal Ethics Application Decision (section A.10).

A.1. Questionnaire for Policy Makers

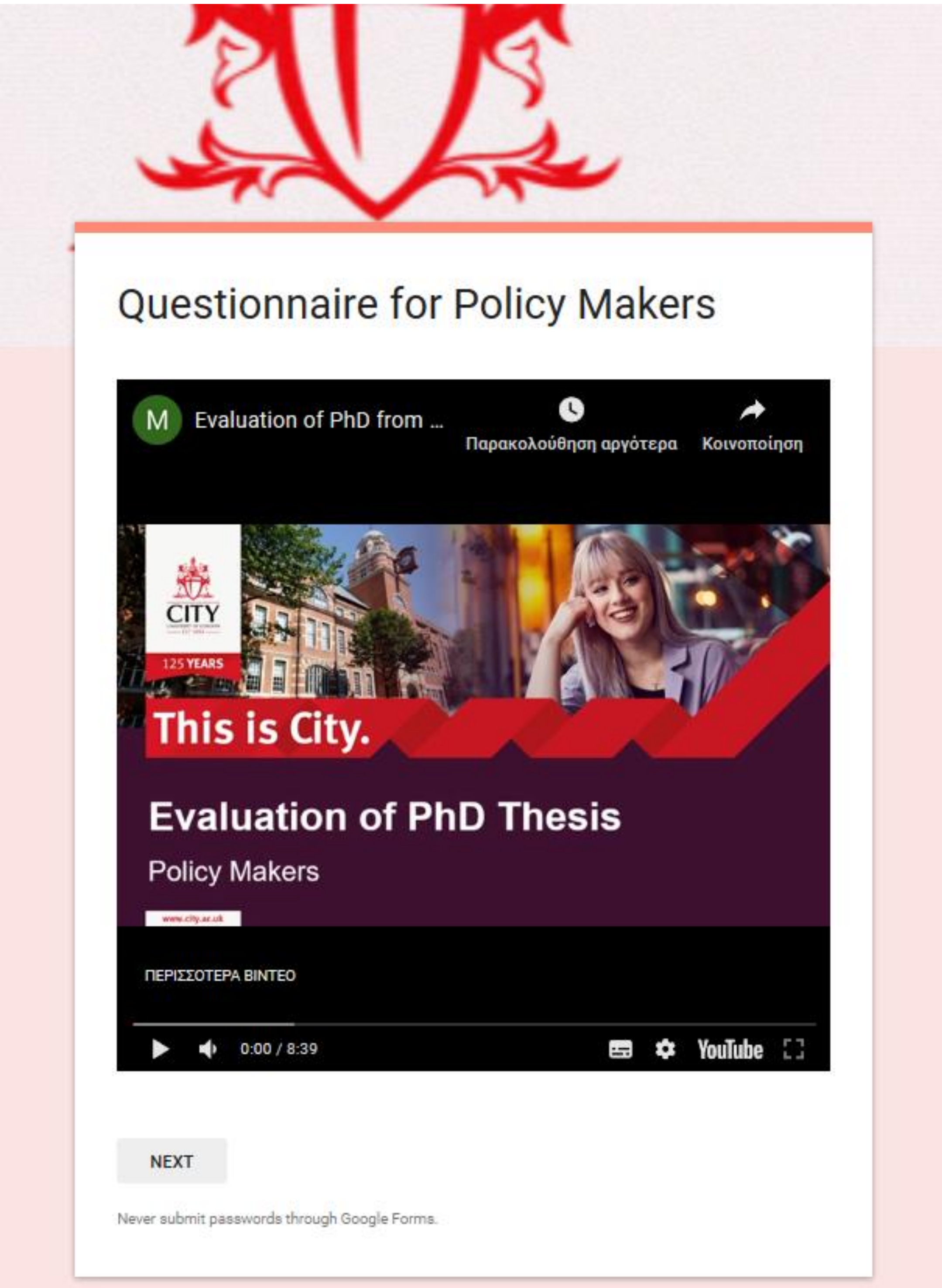
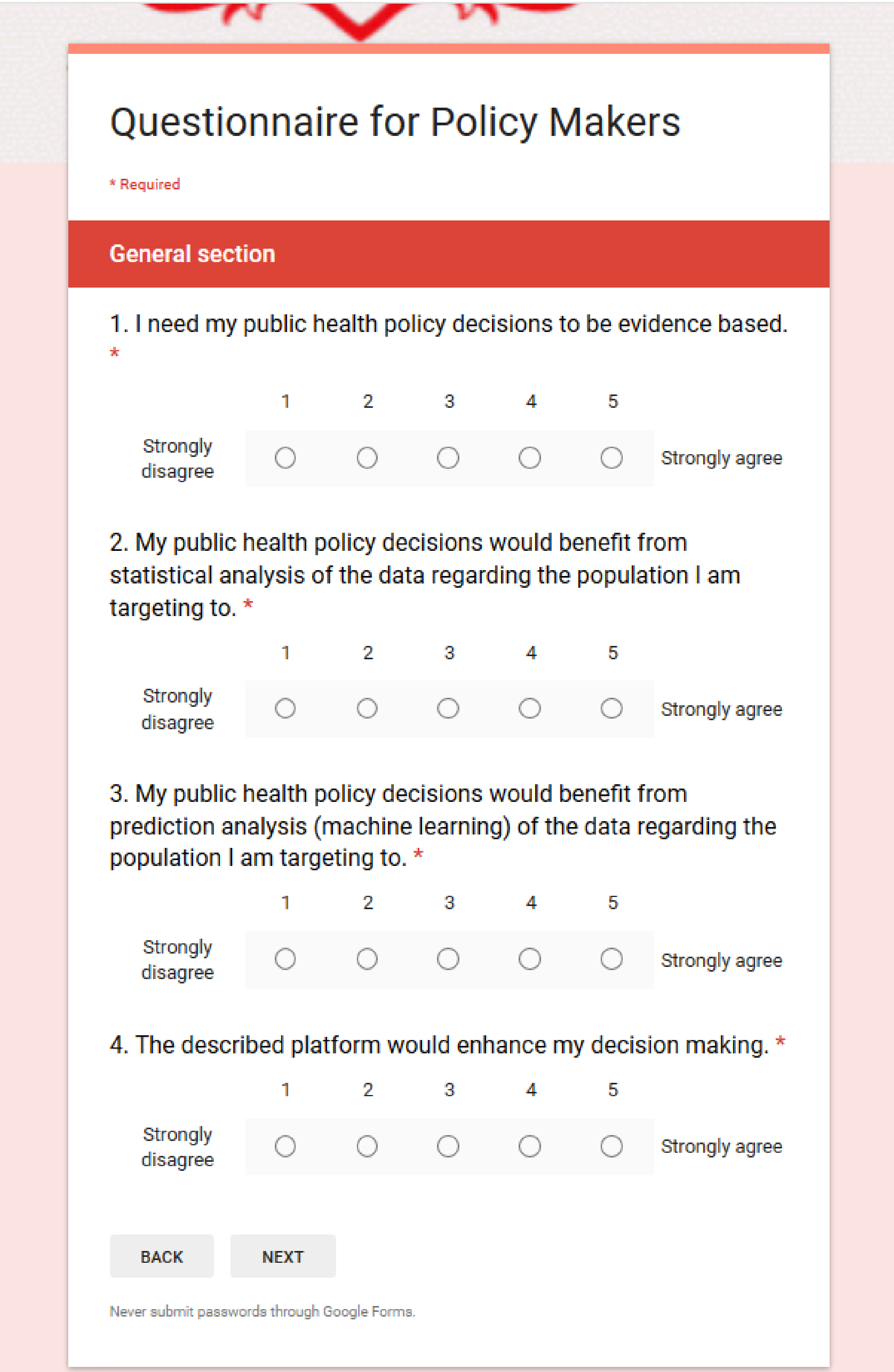


Figure 106 Policy Makers Questionnaire Video



Questionnaire for Policy Makers

* Required

General section

1. I need my public health policy decisions to be evidence based. *

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

2. My public health policy decisions would benefit from statistical analysis of the data regarding the population I am targeting to. *

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

3. My public health policy decisions would benefit from prediction analysis (machine learning) of the data regarding the population I am targeting to. *

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

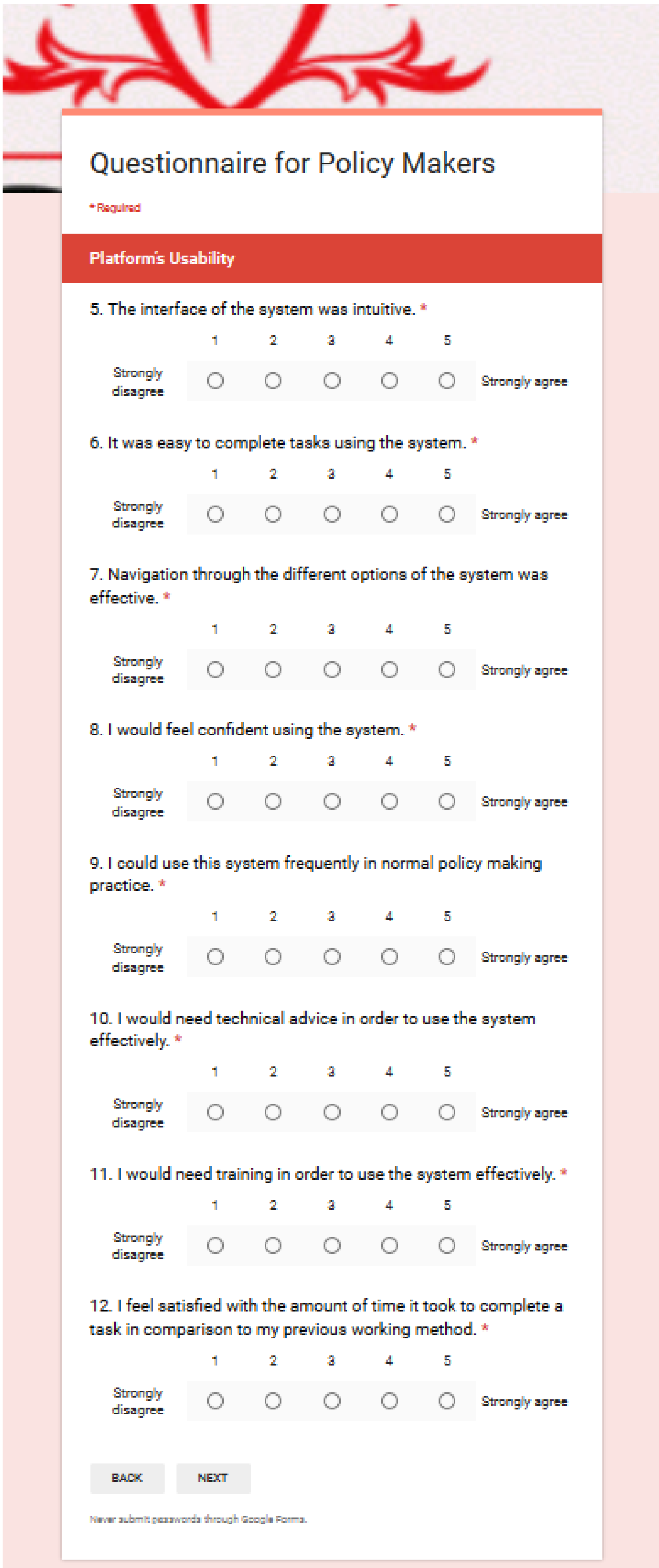
4. The described platform would enhance my decision making. *

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

BACK NEXT

Never submit passwords through Google Forms.

Figure 107 Policy Makers Questionnaire General Section



Questionnaire for Policy Makers

* Required

Platform's Usability

5. The interface of the system was intuitive. *

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

6. It was easy to complete tasks using the system. *

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

7. Navigation through the different options of the system was effective. *

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

8. I would feel confident using the system. *

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

9. I could use this system frequently in normal policy making practice. *

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

10. I would need technical advice in order to use the system effectively. *

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

11. I would need training in order to use the system effectively. *

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

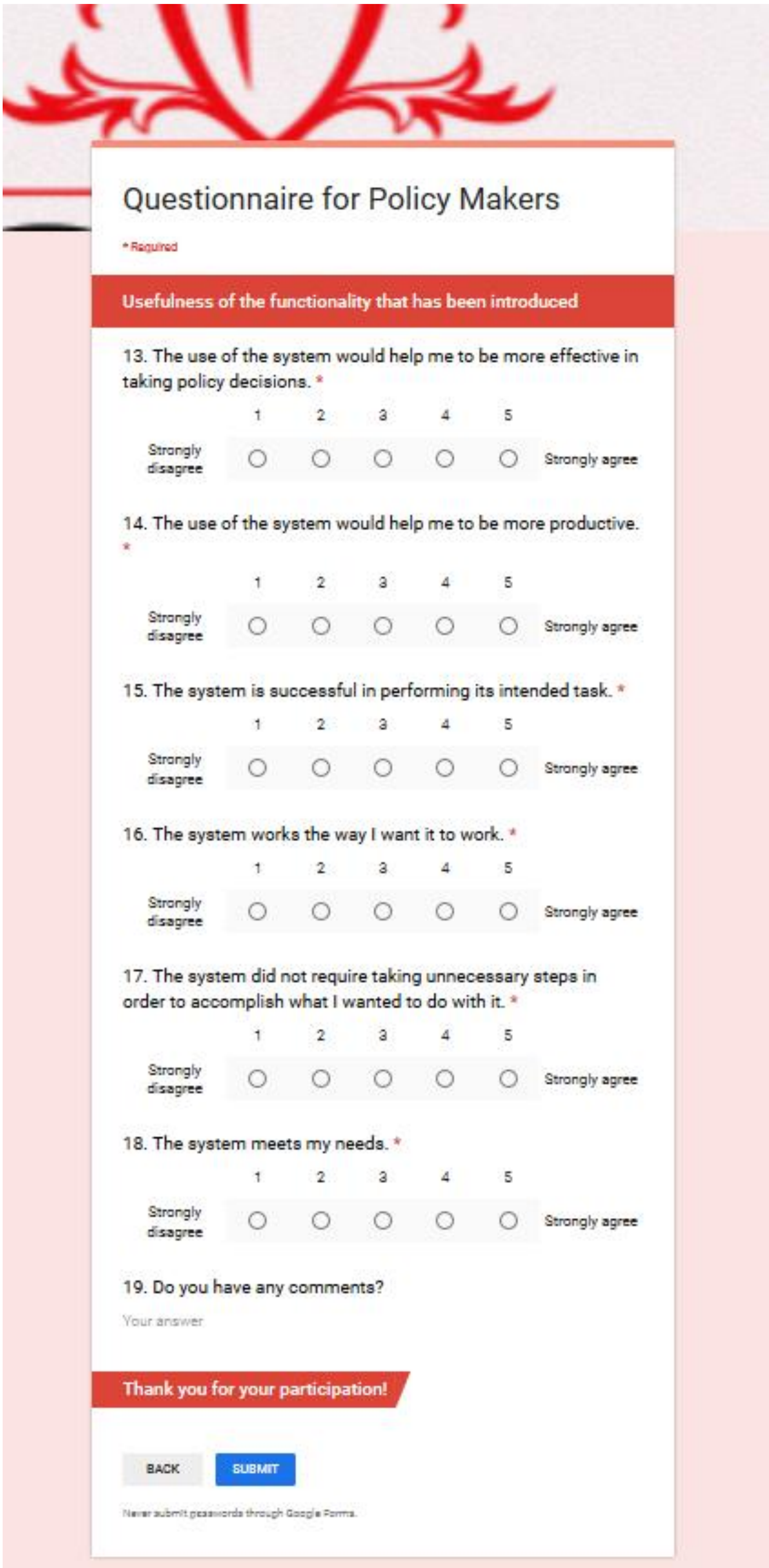
12. I feel satisfied with the amount of time it took to complete a task in comparison to my previous working method. *

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

BACK NEXT

Never submit passwords through Google Forms.

Figure 108 Policy Makers Questionnaire Usability Section



Questionnaire for Policy Makers

* Required

Usefulness of the functionality that has been introduced

13. The use of the system would help me to be more effective in taking policy decisions. *

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

14. The use of the system would help me to be more productive. *

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

15. The system is successful in performing its intended task. *

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

16. The system works the way I want it to work. *

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

17. The system did not require taking unnecessary steps in order to accomplish what I wanted to do with it. *

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

18. The system meets my needs. *

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

19. Do you have any comments?

Your answer

Thank you for your participation!

BACK SUBMIT

Never submit passwords through Google Forms.

Figure 109 Policy Makers Questionnaire Usefulness Section

Script of Questionnaire for Policy Makers

Intro

Welcome to the evaluation of the PhD of Marios Prasinos, titled "Evidence Based Policy Making in Healthcare using Big Data Analytics".
Thanks a lot for your participation as a policy maker.

TOC

For the evaluation of this PhD thesis, first we are going to introduce you to the developed platform.
Then, we are going to give some details about the developed public health policy decision models specification language.
After that we are going to give an example public health policy decision making model.
Afterwards we are going to demonstrate the use of the tool for the specification of models and how to specify the example model.
Finally, we are going to show the results of the execution of the example model.

Platform

We introduce a platform that acts as a tool for evidence based public health policy making.
This platform requires the collaboration of policy makers, clinicians and data analysts.
The platform enables the policy maker to create public health policy decision making models.
The modeling is based on a newly introduced modeling language.

Language

According to the language, each policy model is aimed at one goal and may have multiple data analytics workflows.
The goal has a description and a rationale and is refined into multiple objectives.
Each objective has a description and a rationale and can be addressed by one or more policy actions.
A policy action can be alternative, dependent, or prerequisite to another policy action.
Each policy action is evaluated by one criterion.
The criterion can constraint one or many datasets and specifies a data analytics workflow.

The data analytics workflow is composed of one or many data analytics tasks.
A data analytics task can be a statistical analysis task, a data processing task, or a data mining task.
Each task utilizes a method, which according to the type is an operation (for data processing tasks) or an algorithm (a data mining task utilizes a data mining algorithm, a statistical analysis task utilizes a statistical analysis algorithm).
Each task also has one or many input datasets and one or many output datasets.
Each dataset has a data specification.
Finally, each algorithm has a specific output data specification.

Focus

The main focus of the policy makers is the goal, objectives, policy actions and criteria of the policy model.
The criteria are specified using existing data analytics workflows or new ones specified with the help of data analysts.

Example Slide1

In this case study we are using data analytics to explore whether the (1)Occupational situation (working in noise, working in groups and occupation), (2)Educational level and (3)Age of Hearing Aid users affects their daily usage. This is to inform policy making involving interventions, targeted to (1)Hearing Aid users of different occupations, (2)Hearing Aid users of different educational levels and (3)Hearing Aid users of different age groups.
First, the policy maker has to specify the goal of the policy model, which is "Addressing Barriers to Hearing Aid Use".
The purpose of this case study is to determine the largest barriers that affect hearing aid use in a population in order to make public health policy decisions to address them.
Barriers to hearing aid use are a significant public health problem. Barriers occur at all levels of the process of provision of hearing aids including at the level of the HA user. The big data gathered about users through the platform would enable policy makers to choose which barriers to address in a population, in order to improve hearing aid use and hence reduce the burden of hearing loss in that population.
Then, the objectives need to be specified.

In this example we have three objectives.

One is to intervene in order to address prevention of HA usage due to occupation, the other, due to educational level and the third one, due to age.

Example Slide2

We specify three policy actions, one for each objective.

The first policy action is about that a particular occupation has to be addressed with additional measures to improve hearing aid use.

The second one has to do with the failure to reach a particular educational level has to be addressed to improve HA use.

Finally, the third one has to do with age related fitting.

After that the policy maker has to define criteria.

Demo Policy Model

In this demo, we show the creation of a policy model with three objectives and three policy actions, one for each objective.

Demo Criteria

----

1 In this demo we create three criteria, one for each policy action. The first criterion is that the regression r square must be greater than 0,5 and the p value of the educational placement must be less than 0.05.

2 The second criterion is that the regression r square must be greater than 0,5 and the p value of the age must be less than 0.05.

3 The third criterion is that the regression r square must be greater than 0,5 and the p value of the employment type must be less than 0.05.

4 We have now successfully created the three criteria.

Results

The platform enhances the policy making processes of the policy makers by giving them the capability to explore the data, find correlations, calculate important features extracted from the data (for example, the average daily usage per patient of hearing aids from the time periods of use), perform statistical analysis and machine learning algorithms to the data, define crucial criteria for the selection of the alternative policy actions and see which criteria are met, in order to support their decisions.

Finally, the platform enables the policy makers to have a historical view of their policy models in order to be able to compare them and take more robust future decisions.

A.2. Questionnaire for Clinicians

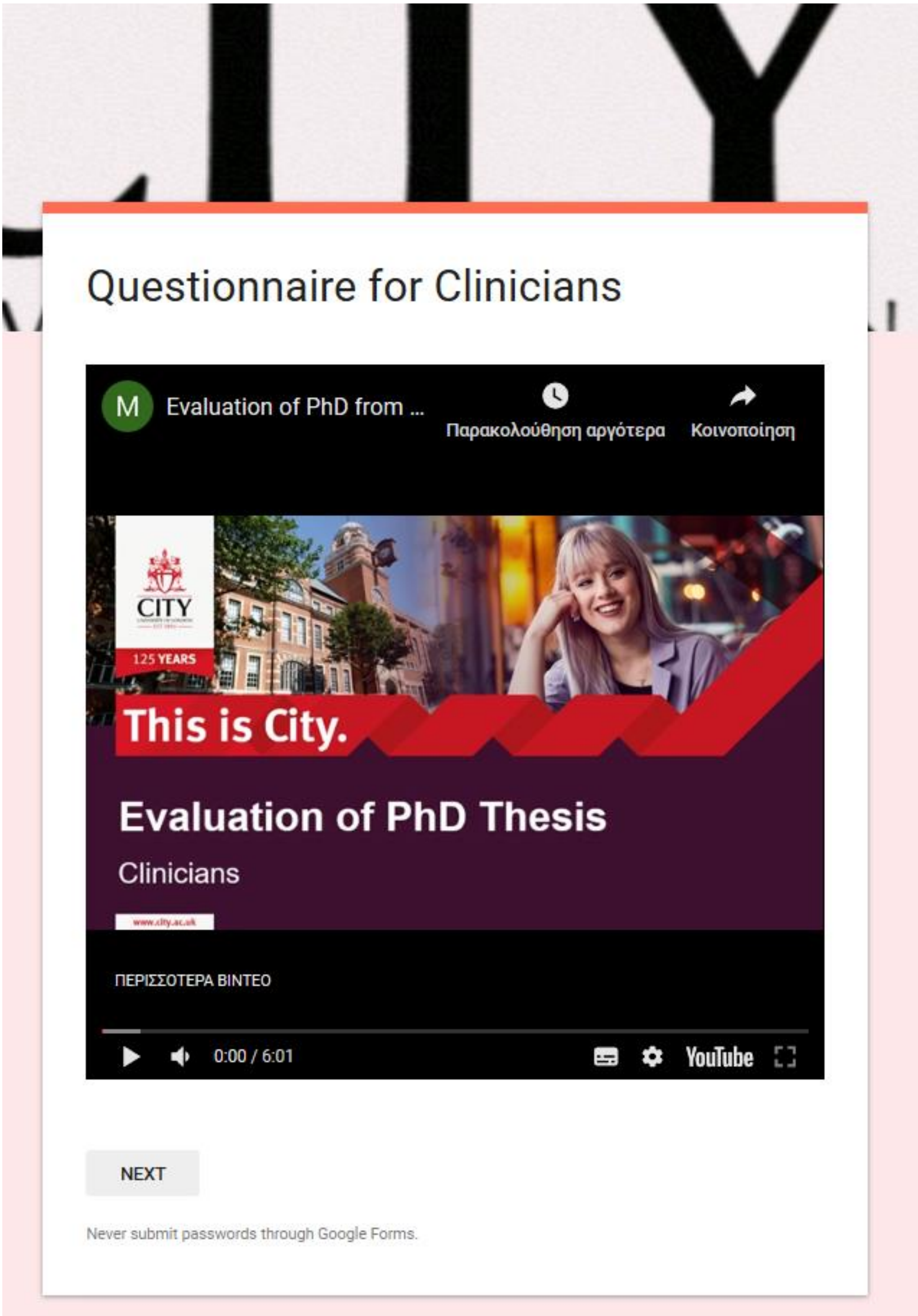
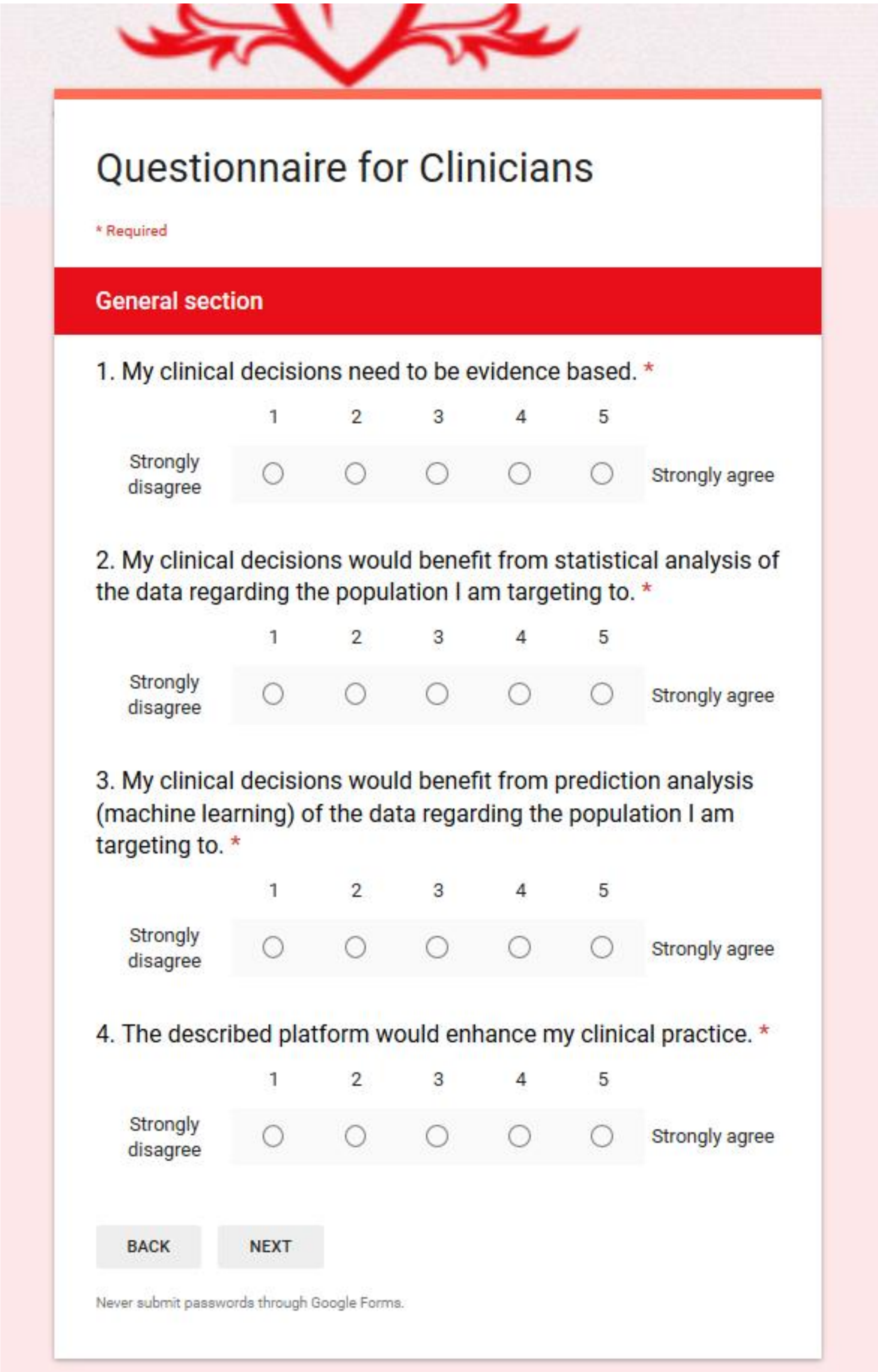


Figure 110 Clinicians Questionnaire Video



Questionnaire for Clinicians

* Required

General section

1. My clinical decisions need to be evidence based. *

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

2. My clinical decisions would benefit from statistical analysis of the data regarding the population I am targeting to. *

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

3. My clinical decisions would benefit from prediction analysis (machine learning) of the data regarding the population I am targeting to. *

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

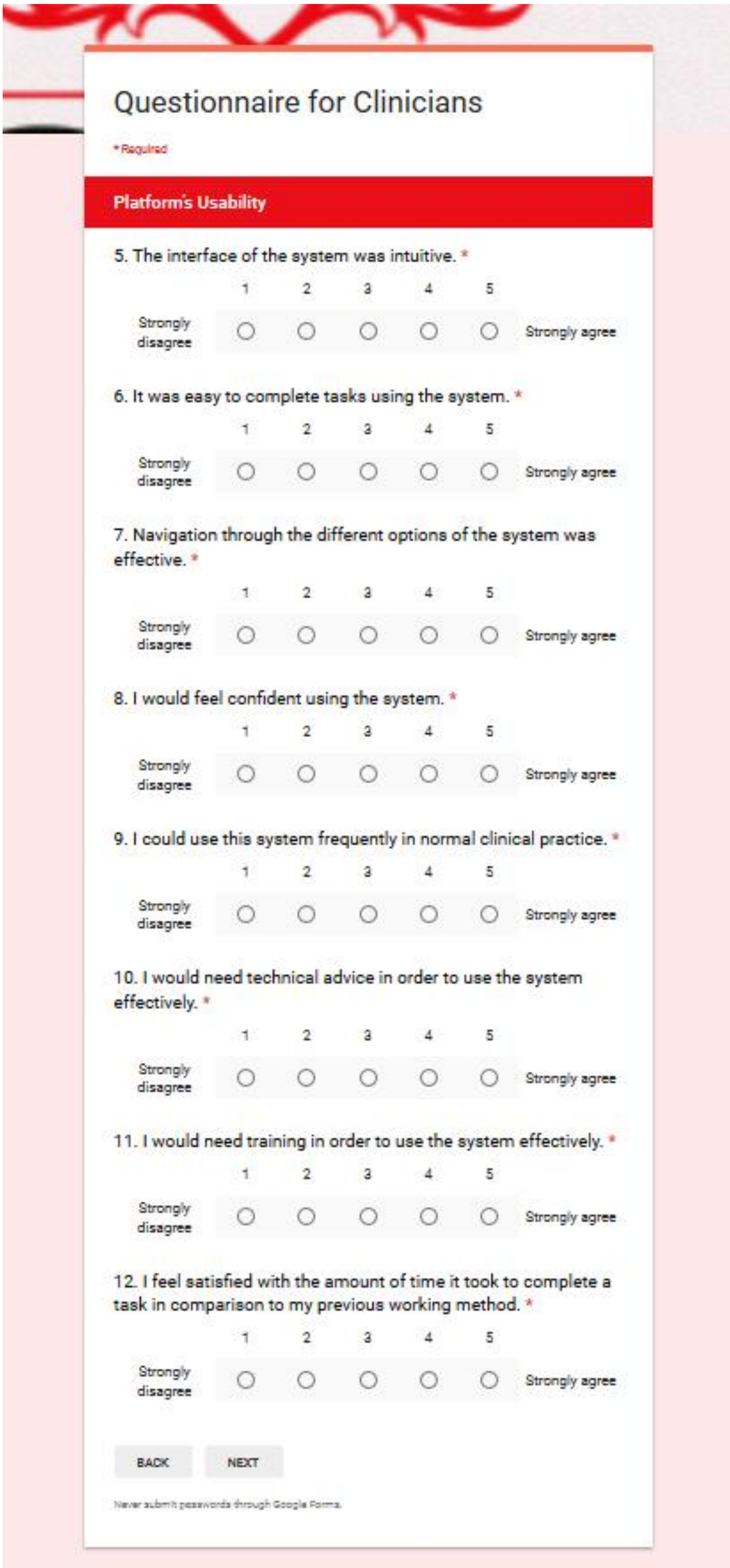
4. The described platform would enhance my clinical practice. *

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

BACK NEXT

Never submit passwords through Google Forms.

Figure 111 Clinicians Questionnaire General Section



Questionnaire for Clinicians

* Required

Platform's Usability

5. The interface of the system was intuitive. *

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

6. It was easy to complete tasks using the system. *

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

7. Navigation through the different options of the system was effective. *

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

8. I would feel confident using the system. *

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

9. I could use this system frequently in normal clinical practice. *

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

10. I would need technical advice in order to use the system effectively. *

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

11. I would need training in order to use the system effectively. *

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

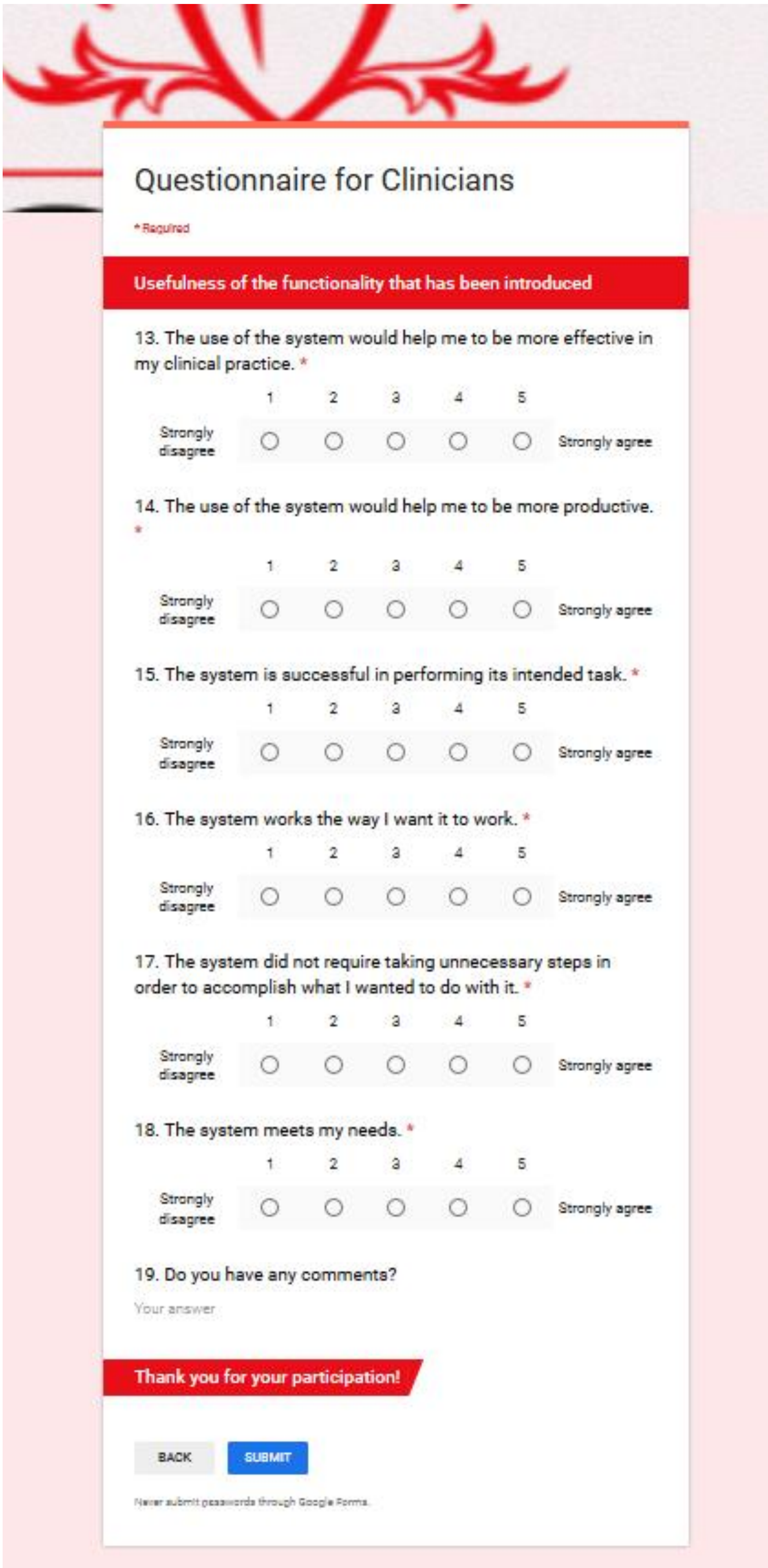
12. I feel satisfied with the amount of time it took to complete a task in comparison to my previous working method. *

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

BACK NEXT

Never submit passwords through Google Forms.

Figure 112 Clinicians Questionnaire Usability Section



Questionnaire for Clinicians

* Required

Usefulness of the functionality that has been introduced

13. The use of the system would help me to be more effective in my clinical practice. *

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

14. The use of the system would help me to be more productive. *

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

15. The system is successful in performing its intended task. *

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

16. The system works the way I want it to work. *

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

17. The system did not require taking unnecessary steps in order to accomplish what I wanted to do with it. *

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

18. The system meets my needs. *

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

19. Do you have any comments?

Your answer

Thank you for your participation!

BACK SUBMIT

Never submit passwords through Google Forms.

Figure 113 Clinicians Questionnaire Usefulness Section

Script of Questionnaire for Clinicians

Intro

Welcome to the evaluation of the PhD of Marios Prasinos, titled "Evidence Based Policy Making in Healthcare using Big Data Analytics".
Thanks a lot for your participation as a clinician.

TOC

For the evaluation of this PhD thesis, first we are going to introduce you to the developed platform.
Then, we are going to give some details about the developed public health policy decision models specification language.
After that we are going to give an example public health policy decision making model.
Afterwards we are going to demonstrate the use of the tool for the specification of models and how to specify the example model.
Finally, we are going to show the results of the execution of the example model.

Platform

We introduce a platform that acts as a tool for evidence based public health policy making.
This platform requires the collaboration of policy makers, clinicians and data analysts.
The platform enables the policy maker to create public health policy decision making models.
The modeling is based on a newly introduced modeling language.

Language

According to the language, each policy model is aimed at one goal and may have multiple data analytics workflows.
The goal has a description and a rationale and is refined into multiple objectives.
Each objective has a description and a rationale and can be addressed by one or more policy actions.
A policy action can be alternative, dependent, or prerequisite to another policy action.
Each policy action is evaluated by one criterion.
The criterion can constraint one or many datasets and specifies a data analytics workflow.

The data analytics workflow is composed of one or many data analytics tasks.
A data analytics task can be a statistical analysis task, a data processing task, or a data mining task.
Each task utilizes a method, which according to the type is an operation (for data processing tasks) or an algorithm (a data mining task utilizes a data mining algorithm, a statistical analysis task utilizes a statistical analysis algorithm).
Each task also has one or many input datasets and one or many output datasets.
Each dataset has a data specification.
Finally, each algorithm has a specific output data specification.

Focus

The main focus of the clinicians is the data entry and the assurance of the validity of the data and the criteria of the policy model.
The criteria are specified using existing data analytics workflows or new ones specified with the help of data analysts.
The clinicians provide their feedback regarding the important features of the available datasets.

Example

In this case study we are using data analytics to explore whether the (1)Occupational situation (working in noise, working in groups and occupation), (2)Educational level and (3)Age of Hearing Aid users affects their daily usage. This is to inform policy making involving interventions, targeted to (1)Hearing Aid users of different occupations, (2)Hearing Aid users of different educational levels and (3)Hearing Aid users of different age groups.
The clinicians can contribute to the specification of three criteria: one for the educational level, one for the age and one for the occupational situation.

Demo Criteria

----

1 In this demo we create three criteria, one for each policy action. The first criterion is that the regression r square must be greater than 0,5 and the p value of the educational placement must be less than 0.05.

2 The second criterion is that the regression r square must be greater than 0,5 and the p value of the age must be less than 0.05.

3 The third criterion is that the regression r square must be greater than 0,5 and the p value of the employment type must be less than 0.05.

4 We have now successfully created the three criteria.

Results

The platform enhances the clinical processes of the clinicians by giving them the capability to explore the data, find correlations, calculate important features extracted from the data (for example, the average daily usage per patient of hearing aids from the time periods of use), perform statistical analysis and machine learning algorithms to the data and take evidence based decisions in their clinical practice and explore whether the analysis results are equivalent to the clinical guidelines.

Outro

Thank you very much for watching this video.

Please fill in the questionnaire for the evaluation of the previously described platform.

A.3. Questionnaire for Data Analysts

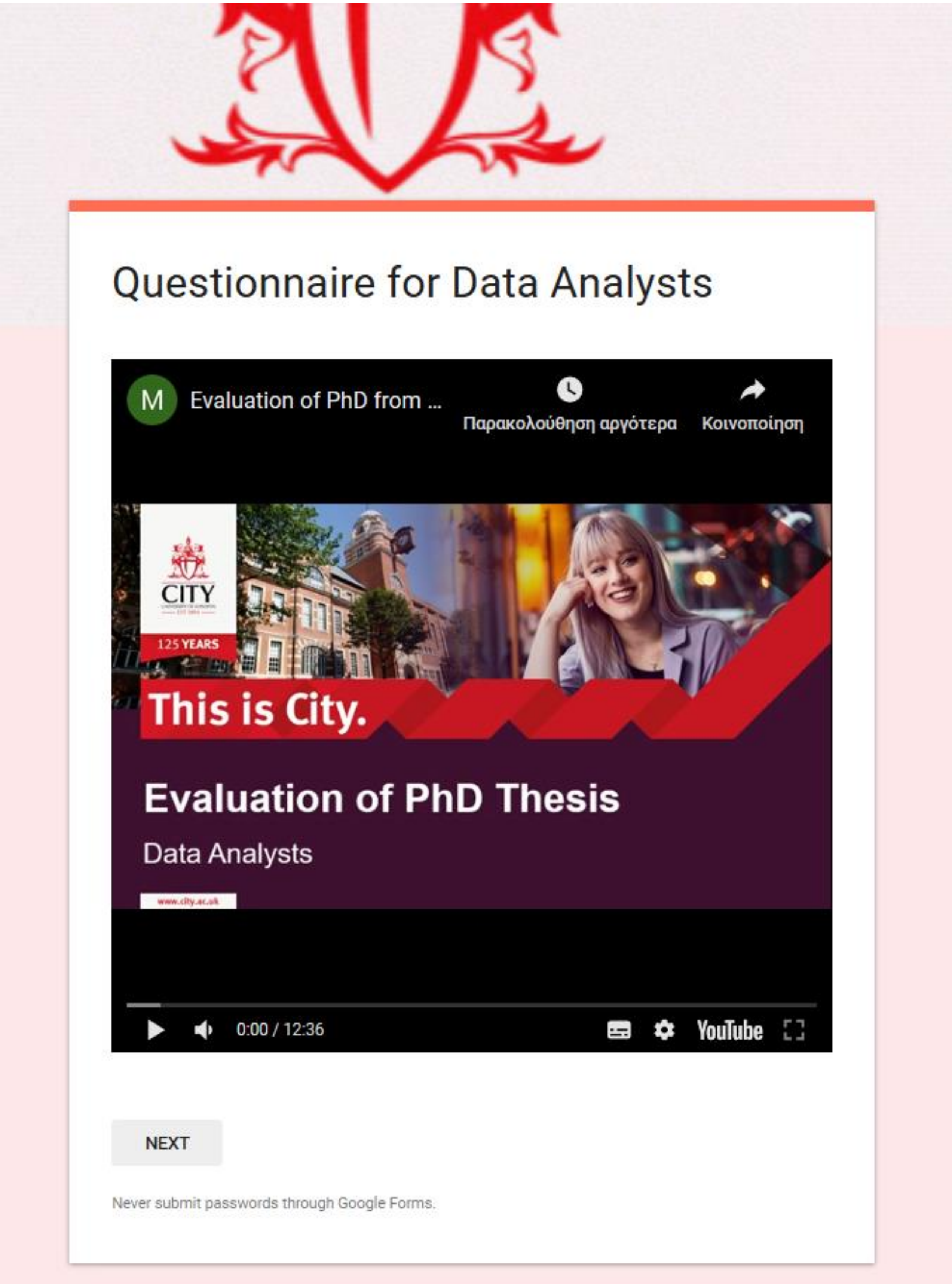


Figure 114 Data Analysts Questionnaire Video

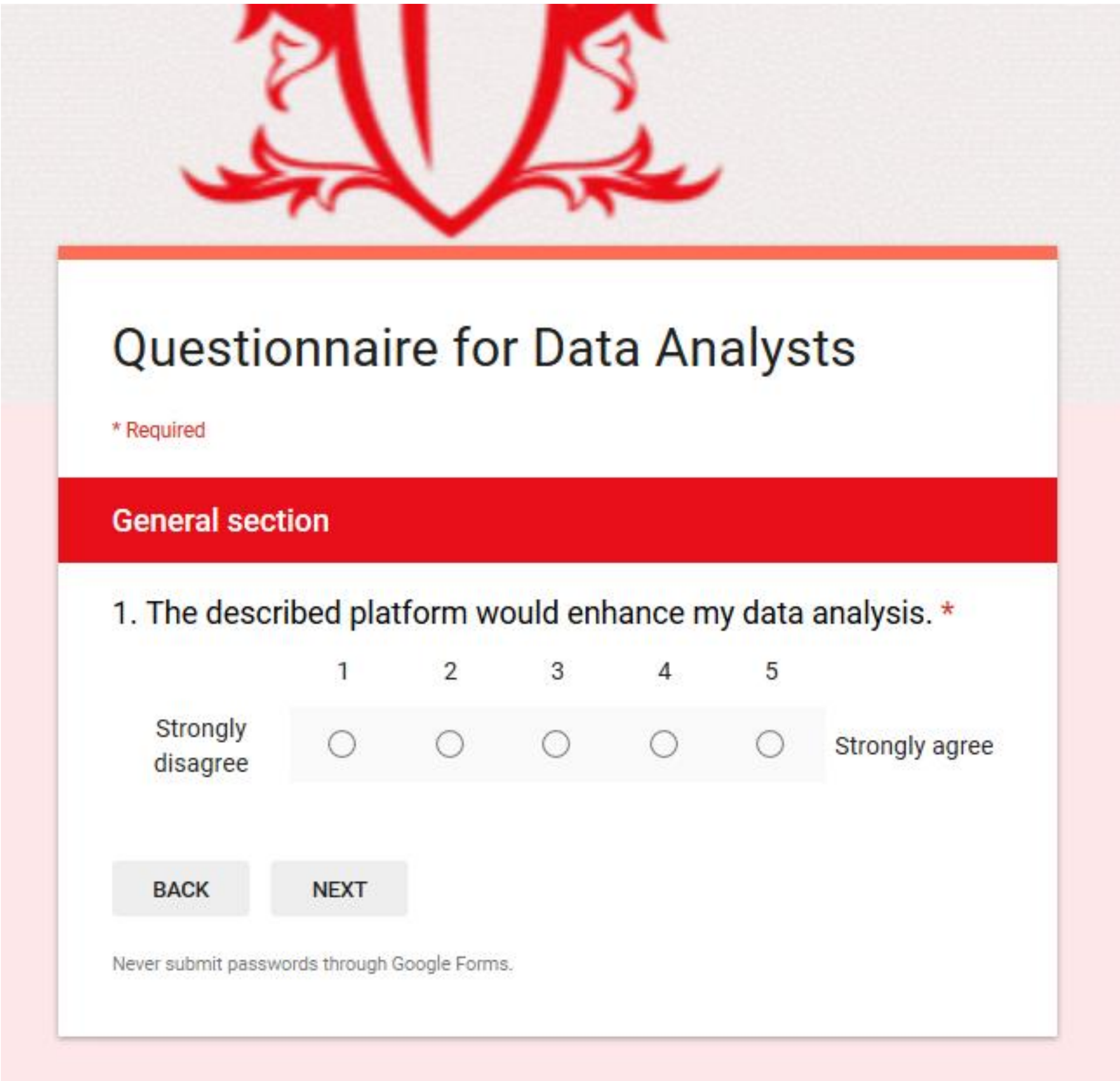
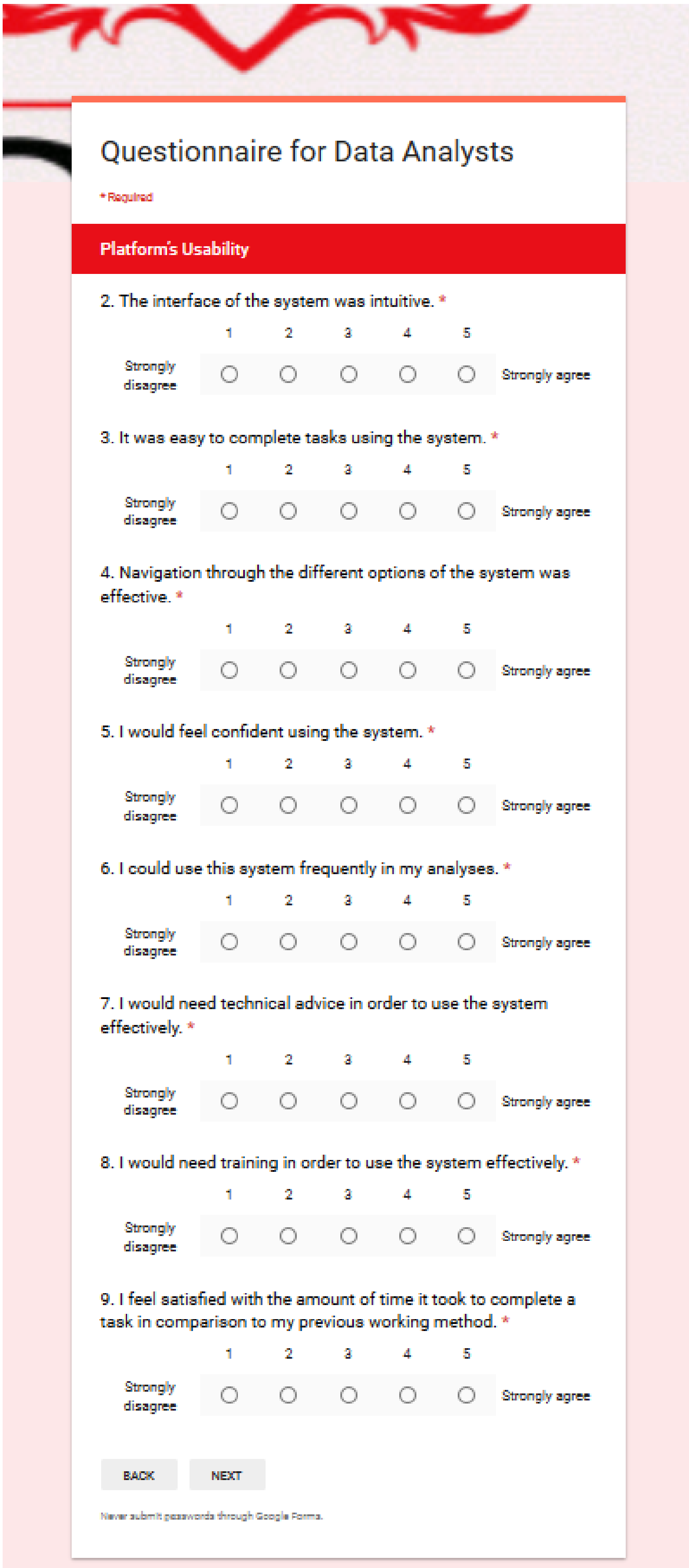


Figure 115 Data Analysts Questionnaire General Section



Questionnaire for Data Analysts

* Required

Platform's Usability

2. The interface of the system was intuitive. *

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

3. It was easy to complete tasks using the system. *

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

4. Navigation through the different options of the system was effective. *

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

5. I would feel confident using the system. *

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

6. I could use this system frequently in my analyses. *

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

7. I would need technical advice in order to use the system effectively. *

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

8. I would need training in order to use the system effectively. *

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

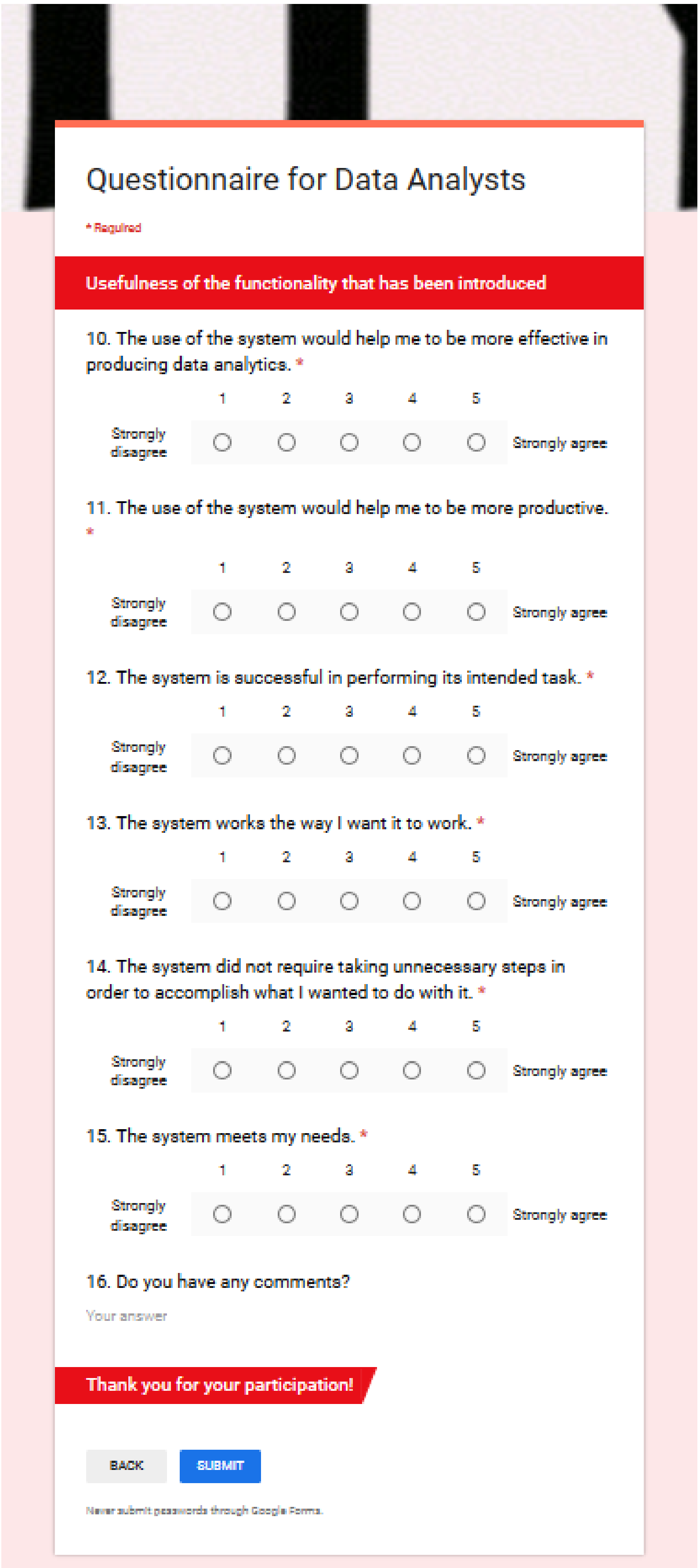
9. I feel satisfied with the amount of time it took to complete a task in comparison to my previous working method. *

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

BACK NEXT

Never submit passwords through Google Forms.

Figure 116 Data Analysts Questionnaire Usability Section



Questionnaire for Data Analysts

* Required

Usefulness of the functionality that has been introduced

10. The use of the system would help me to be more effective in producing data analytics. *

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

11. The use of the system would help me to be more productive. *

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

12. The system is successful in performing its intended task. *

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

13. The system works the way I want it to work. *

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

14. The system did not require taking unnecessary steps in order to accomplish what I wanted to do with it. *

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

15. The system meets my needs. *

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

16. Do you have any comments?

Your answer

Thank you for your participation!

BACK SUBMIT

Never submit passwords through Google Forms.

Figure 117 Data Analysts Questionnaire Usefulness Section

Script of Questionnaire for Data Analysts

Intro

Welcome to the evaluation of the PhD of Marios Prasinos, titled "Evidence Based Policy Making in Healthcare using Big Data Analytics".
Thanks a lot for your participation as a data analyst.

TOC

For the evaluation of this PhD thesis, first we are going to introduce you to the developed platform.
Then, we are going to give some details about the developed public health policy decision models specification language.
After that we are going to give an example public health policy decision making model.
Afterwards we are going to demonstrate the use of the tool for the specification of models and how to specify the example model.
Finally, we are going to show the results of the execution of the example model.

Platform

We introduce a platform that acts as a tool for evidence based public health policy making.
This platform requires the collaboration of policy makers, clinicians and data analysts.
The platform enables the policy maker to create public health policy decision making models.
The modeling is based on a newly introduced modeling language.

Language

According to the language, each policy model is aimed at one goal and may have multiple data analytics workflows.
The goal has a description and a rationale and is refined into multiple objectives.
Each objective has a description and a rationale and can be addressed by one or more policy actions.
A policy action can be alternative, dependent, or prerequisite to another policy action.
Each policy action is evaluated by one criterion.
The criterion can constraint one or many datasets and specifies a data analytics workflow.

The data analytics workflow is composed of one or many data analytics tasks.

A data analytics task can be a statistical analysis task, a data processing task, or a data mining task.

Each task utilizes a method, which according to the type is an operation (for data processing tasks) or an algorithm (a data mining task utilizes a data mining algorithm, a statistical analysis task utilizes a statistical analysis algorithm).

Each task also has one or many input datasets and one or many output datasets.

Each dataset has a data specification.

Finally, each algorithm has a specific output data specification.

Focus

The main focus of the data analysts is the creation of the data analytics workflows to support the criteria of the policy model.

Data analytics workflows are specified as part of a policy model, to support the criteria of the policy model, or independently, to run an analysis to the data.

A data analyst defines data analytics tasks, with their input and output datasets.

Then, they create data analytics workflows by choosing a set of previously created tasks.

A data analytics workflow can be executed upon user request, periodically (for example, once every year) or when there is a change to the data (for example, when the volume of data received from mobiles changes by 40%).

The created tasks are reusable for multiple workflows and the workflows are reusable for multiple policy models.

Example

In this case study we are using data analytics to explore whether the (1)Occupational situation (working in noise, working in groups and occupation), (2)Educational level and (3)Age of Hearing Aid users affects their daily usage.

This is to inform policy making involving interventions, targeted to (1)Hearing Aid users of different occupations, (2)Hearing Aid users of different educational levels and (3)Hearing Aid users of different age groups.

The Data Analytics Workflow WF1 is composed of three DataProcessingTasks: two data filtering tasks and one full join task and one statistical analysis task.

The first filtering task has as input the PATIENT-COMPUTED-AGE and the TIME_PERIOD-COMPUTED-ADU and the second filtering task has as input the Q_DRMED-LS_EDUC_PLACEM, Q_DRMED-LS_EMPL_SIT1, Q_DRMED-LS_EMPL_SIT2 and Q_DRMED-LS_EMPL_TYPE.
Each of the two select produces as output a corresponding DataStream.
Data streams are then merged by a different DataProcessingTask, utilizing a JoinOperation.
The Join operation takes as parameters the data used to define policies (EDUCATION, AGE), and DAILY USAGE.
The Data Processing Task produces a Data stream, which is the input to the StatisticalAnalysisTask.
The algorithm used for the StatisticalAnalysisTask is the StatisticalRegressionAlgorithm.
Then we proceed to the specification of three criteria: one for the educational level, one for the age and one for the occupational situation.

Demo WF1

----

1 In this demo we show the creation of Data Analytics Workflow WF1.

2 This workflow is composed of three data processing tasks and one statistical analysis task.

3 We first create a data filtering task, which has as inputs patient's age and the computed average daily usage of his hearing aids.

4 We then create another data filtering task. This task's inputs are the patient's employment type, whether the patient is working in noise, whether the patient is working in groups and the educational level of the patient.

5 After that, we create a full join task with inputs the outputs of the previously created tasks.

6 Finally, we create a statistical analysis task with method linear regression, and inputs the patients age, the computed average daily usage of his hearing aids, his educational placement and his employment type.

7 We select as dependent variable the average daily usage.

8 We have now successfully created all the tasks of WF1.

Demo Criteria

----

1 In this demo we create three criteria, one for each policy action. The first criterion is that the regression r square must be greater than 0,5 and the p value of the educational placement must be less than 0.05.

2 The second criterion is that the regression r square must be greater than 0,5 and the p value of the age must be less than 0.05.
3 The third criterion is that the regression r square must be greater than 0,5 and the p value of the employment type must be less than 0.05.
4 We have now successfully created the three criteria.

Results

The platform enhances the analyses of the data analysts by giving them the capability to explore the data, find correlations, calculate important features extracted from the data (for example, the average daily usage per patient of hearing aids from the time periods of use), perform statistical analysis and machine learning algorithms to the data and view and compare the results of all the analyses performed and create evidence based reports with the analyses of the data.

Outro

Thank you very much for watching this video.
Please fill in the questionnaire for the evaluation of the previously described platform.

A.4. Consent form for Interviews

CONSENT FORM

Title of Study: Evaluation of the EVOTION public health policy tool supporting policy making decisions in healthcare using big data analytics

1	I confirm that I have had the study explained to me, and I have read the participant information sheet, which I may keep for my records.	
	I understand this will involve:	
	• be interviewed by the researcher	
	• complete questionnaires asking me about the evaluation of the proposed platform	
2	This information will be held by City, as data controller, and processed for the following purpose(s): Public Task: The legal basis for processing your personal data will be that this research is a task in the public interest, i.e., City, University of London considers the lawful basis for processing personal data to fall under Article 6(1)(e) of GDPR (public task) as the processing of research participant data is necessary for learning and teaching purposes and all research with human participants by staff and students has to be scrutinised and approved by one of City's Research Ethics Committees.	
3	I understand that any information I provide is confidential, and that no information that could lead to the identification of any individual will be disclosed in any reports on the project, or to any other party. No identifiable personal data will be published. The identifiable data will not be shared with any other organisation.	
4	I understand that my participation is voluntary, that I can choose not to participate in part or all of the project, and that I can withdraw at any stage of the project without being penalised or disadvantaged in any way.	
5	I agree to City recording and processing this information about me. I understand that this information will be used only for the purpose(s) set out in this statement and my consent is conditional on	

	City complying with its duties and obligations under the General Data Protection Regulation (GDPR).	
6.	I agree to the arrangements for data storage, archiving, sharing.	
7	I agree to the use of anonymised quotes in publication.	
8	I agree to take part in the above study.	

Name of Participant	Signature	Date
Name of Researcher	Signature	Date

A.5. Consent form for Questionnaires

CONSENT FORM

Title of Study: Evaluation of the EVOTION public health policy tool supporting policy making decisions in healthcare using big data analytics

1	I confirm that I have had the study explained to me, and I have read the participant information sheet, which I may keep for my records.	
	I understand this will involve:	
	• complete questionnaires asking me about the evaluation of the proposed platform	
2	This information will be held by City, as data controller, and processed for the following purpose(s): Public Task: The legal basis for processing your personal data will be that this research is a task in the public interest, i.e., City, University of London considers the lawful basis for processing personal data to fall under Article 6(1)(e) of GDPR (public task) as the processing of research participant data is necessary for learning and teaching purposes and all research with human participants by staff and students has to be scrutinised and approved by one of City's Research Ethics Committees.	
3	I understand that any information I provide is confidential, and that no information that could lead to the identification of any individual will be disclosed in any reports on the project, or to any other party. No identifiable personal data will be published. The identifiable data will not be shared with any other organisation.	
4	I understand that my participation is voluntary, that I can choose not to participate in part or all of the project, and that I can withdraw at any stage of the project without being penalised or disadvantaged in any way.	
5	I agree to City recording and processing this information about me. I understand that this information will be used only for the purpose(s) set out in this statement and my consent is conditional on	

	City complying with its duties and obligations under the General Data Protection Regulation (GDPR).	
6.	I agree to the arrangements for data storage, archiving, sharing.	
7	I agree to the use of anonymised quotes in publication.	
8	I agree to take part in the above study.	

______________________ ______________________ ______________

Name of Participant Signature Date

______________________ ______________________ ______________

Name of Researcher Signature Date

A.6. Participant information sheet for Interviews

PARTICIPANT INFORMATION SHEET

(1) Title of study: Evaluation of the EVOTION public health policy tool supporting policy making decisions in healthcare using big data analytics

(2) Name of principal investigator: Marios Prasinos [redacted]

We would like to invite you to take part in a research study. Before you decide whether you would like to take part it is important that you understand why the research is being done and what it would involve for you. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information.

(3) What is the purpose of the study?

The effective management of various health conditions depends on and requires appropriate public health policies. Public health policy can affect several aspects of healthcare provision including: (a) the prevention and early diagnosis of diseases; (b) the early treatment of diagnosed conditions through the provision of the appropriate health care devices; (c) the longer term treatment of long term disabilities and chronic diseases through systematic checks of the condition of the patient, the provision of other vital related rehabilitation services; (d) the protection of people with health care devices from the harmful effects of their living environment; (e) the setup of standards, services and technology for promoting and ensuring inclusion of participation of patients with in various settings (e.g., at work, at school/educational establishments, in everyday life).
Although there is a need for evidence based public health policy making, at present there are no computer tools supporting this in a comprehensive manner, i.e., in a way that integrates data analytics with policy decision making and uses the outcomes of data analysis (aka evidence) to aid stakeholders in making relevant public health policy making decisions.

The overall aim of our research has been to develop a software tool that supports this process. This tool supports: (a) the description of alternative public health policy actions and interventions, (b) associates them with data that should be analysed in order to explore the viability of alternative public health decision making options, (c) describes how the data should be analysed to produce evidence related to the assessment of different options, and (d) supports stakeholders in making decisions based on the generated evidence.

The purpose of this study is to present the developed tool to public health policy makers and clinicians and obtain feedback from them regarding the features of this tool and its effectiveness, and analyse it to provide an evaluation of the developed software tool.

(4) Why have I been invited?

You have been invited to participate in this study and evaluate the proposed platform, because you are an experienced policy maker, clinician or data analyst that can evaluate such a platform independently and subjectively.

(5) Do I have to take part?

Your participation in the study is voluntary. This means that you can choose to participate in all, part of or none of the study and that, even if you have decided to participate in it, you can withdraw at any stage of the study without being penalised or disadvantaged in any way.

If you do decide to take part in the study, you will be asked to sign a consent form. If you decide to take part, you are still free to withdraw at any time and without giving a reason. Also, you will be able to request information about the data that have been collected from you and their deletion provided that by the time that you do so, there has been no publication of research outcomes that have used your data. If such a publication has been made, it will still be possible to request that no further processing is performed upon the data that you provided and that only the portion of your data which contributed to the existing publication will be kept but in a form that will no longer make it possible to attribute them to you.

(6) What will happen if I take part?

Your participation to the study will last up no more than 30 minutes and will involve you answering questions specified in a given questionnaire online and answer some questions to the principal investigator regarding the questionnaire.

This will follow a request to you to view some presentational material (video and/or slides) of the integrated public health policy decision making platform described in (3) above.

The questionnaires are sent to people in the UK, Greece, Bulgaria, Poland and Croatia. We ensure you that the data is processed to the same standards as within European Economic Area (EEA).

(7) What do I have to do?

You have to view some presentational material (video and/or slides) of the integrated public health policy decision making platform described in (3) above and fill in the questionnaire.

(8) What are the possible disadvantages and risks of taking part?

There are no disadvantages and risks of taking part.

(9) What are the possible benefits of taking part?

By taking part to this study, you will help us evaluate the merit of the integrated platform described in (3) above as a public health policy making tool.

(10) What will happen when the research study stops?

When the study is completed, the data collected will be processed and the outcomes of their analysis will be presented in the PhD thesis of the principal investigator and possibly published papers authored/co-authored by the principal investigator and his supervisor.

(11) Will my taking part in the study be kept confidential?

- The data collected from you will be anonymised prior to any further processing.
- To be able to fulfil a future request from your side to delete or stop using your data, an association between the data you provided and an identifier that can implicitly enable the linking of the data to you will be kept separately in an encrypted form. This association will only be available to the principal investigator.
- The supervisor of the study and the examiners of the PhD thesis of the principal investigator will be given access to anonymised data collected through the study if they so wish and solely for the purposes of validating the outcomes of the evaluation analysis carried out by the principal investigator.
- The anonymised data will be kept at City for 10 years. After that, all the acquired data will be deleted.
- The encrypted data enabling the association of the acquired data with the individuals participants of the study will be deleted as soon as the PhD thesis is examined.

(12) What should I do if I want to take part?

In order to take part, you will need to read the platform description and answer the questionnaire sent to you via email.

(13) What will happen to results of the research study?

The results of the research study will be presented in the thesis of the Principal Investigator and possibly published papers authored/co-authored by the principal investigator and his supervisor.

(14) What will happen if I do not want to carry on with the study?

You are free to withdraw from the study without an explanation or penalty at any time.

(15) Who has reviewed the study?

This study has been approved by City, University of London Research Ethics Committee.

(16) Further information and contact details

Marios Prasinos [redacted]

(17) Data Protection Privacy Notice: What are my rights under the data protection legislation?

City, University of London is the data controller for the personal data collected for this research project. Your personal data will be processed for the purposes outlined in this notice. The legal basis for processing your personal data will be that this research is a task in the public interest, that is City, University of London considers the lawful basis for processing personal data to fall under Article 6(1)(e) of GDPR (public task) as the processing of research participant data is necessary for learning and teaching purposes and all research with human participants by staff and students has to be scrutinised and approved by one of City's Research Ethics Committees.

The rights you have under the data protection legislation are listed below, but not all of the rights will be applied to the personal data collected in each research project.

- right to be informed
- right of access
- right to rectification
- right to erasure
- right to restrict processing
- right to object to data processing
- right to data portability
- right to object
- rights in relation to automated decision making and profiling

For more information, please visit www.city.ac.uk/about/city-information/legal

(18) What if I have concerns about how my personal data will be used after I have participated in the research?

In the first instance you should raise any concerns with the research team, but if you are dissatisfied with the response, you may contact the Information Compliance Team at dataprotection@city.ac.uk or phone [redacted], who will liaise with City's Data Protection Officer [redacted] to answer your query.

If you are dissatisfied with City's response you may also complain to the Information Commissioner's Office at www.ico.org.uk

(19) What if there is a problem?

If you have any problems, concerns or questions about this study, you should ask to speak to a member of the research team. If you remain unhappy and wish to complain formally, you can do this through City's complaints procedure. To complain about the study, you need to phone [redacted]. You can then ask to speak to the Secretary to Senate Research Ethics Committee and inform them that the name of the project is: Evidence Based Policy Making in Healthcare using Big Data Analytics

You could also write to the Secretary at:

[redacted]
Research Integrity Manager

Research & Enterprise
City, University of London
Northampton Square
London
EC1V 0HB
[redacted]

City holds insurance policies which apply to this study. If you feel you have been harmed or injured by taking part in this study, you may be eligible to claim compensation. This does not affect your legal rights to seek compensation. If you are harmed due to someone's negligence, then you may have grounds for legal action.

Thank you for taking the time to read this information sheet.

02/02/2019 version 1.0

A.7. Participant information sheet for Questionnaires

PARTICIPANT INFORMATION SHEET

(1) Title of study: Evaluation of the EVOTION public health policy tool supporting policy making decisions in healthcare using big data analytics

(2) Name of principal investigator: Marios Prasinos [redacted]

We would like to invite you to take part in a research study. Before you decide whether you would like to take part it is important that you understand why the research is being done and what it would involve for you. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information.

(3) What is the purpose of the study?

The effective management of various health conditions depends on and requires appropriate public health policies. Public health policy can affect several aspects of healthcare provision including: (a) the prevention and early diagnosis of diseases; (b) the early treatment of diagnosed conditions through the provision of the appropriate health care devices; (c) the longer term treatment of long term disabilities and chronic diseases through systematic checks of the condition of the patient, the provision of other vital related rehabilitation services; (d) the protection of people with health care devices from the harmful effects of their living environment; (e) the setup of standards, services and technology for promoting and ensuring inclusion of participation of patients with in various settings (e.g., at work, at school/educational establishments, in everyday life).
Although there is a need for evidence based public health policy making, at present there are no computer tools supporting this in a comprehensive manner, i.e., in a way that integrates data analytics with policy decision making and uses the outcomes of data analysis (aka evidence) to aid stakeholders in making relevant public health policy making decisions.

The overall aim of our research has been to develop a software tool that supports this process. This tool supports: (a) the description of alternative public health policy actions and interventions, (b) associates them with data that should be analysed in order to explore the viability of alternative public health decision making options, (c) describes how the data should be analysed to produce evidence related to the assessment of different options, and (d) supports stakeholders in making decisions based on the generated evidence.

The purpose of this study is to present the developed tool to public health policy makers and clinicians and obtain feedback from them regarding the features of this tool and its effectiveness, and analyse it to provide an evaluation of the developed software tool.

(4) Why have I been invited?

You have been invited to participate in this study and evaluate the proposed platform, because you are an experienced policy maker, clinician or data analyst that can evaluate such a platform independently and subjectively.

(5) Do I have to take part?

Your participation in the study is voluntary. This means that you can choose to participate in all, part of or none of the study and that, even if you have decided to participate in it, you can withdraw at any stage of the study without being penalised or disadvantaged in any way.

If you do decide to take part in the study, you will be asked to sign a consent form. If you decide to take part, you are still free to withdraw at any time and without giving a reason. Also, you will be able to request information about the data that have been collected from you and their deletion provided that by the time that you do so, there has been no publication of research outcomes that have used your data. If such a publication has been made, it will still be possible to request that no further processing is performed upon the data that you provided and that only the portion of your data which contributed to the existing publication will be kept but in a form that will no longer make it possible to attribute them to you.

(6) What will happen if I take part?

Your participation to the study will last up no more than 30 minutes and will involve you answering questions specified in a given questionnaire online, without any need to have a physical meeting with the principal investigator of the study.

This will follow a request to you to view some presentational material (video and/or slides) of the integrated public health policy decision making platform described in (3) above.

The questionnaires are sent to people in the UK, Greece, Bulgaria, Poland and Croatia. We ensure you that the data is processed to the same standards as within European Economic Area (EEA).

(7) What do I have to do?

You have to view some presentational material (video and/or slides) of the integrated public health policy decision making platform described in (3) above and fill in the questionnaire.

(8) What are the possible disadvantages and risks of taking part?

There are no disadvantages and risks of taking part.

(9) What are the possible benefits of taking part?

By taking part to this study, you will help us evaluate the merit of the integrated platform described in (3) above as a public health policy making tool.

(10) What will happen when the research study stops?

When the study is completed, the data collected will be processed and the outcomes of their analysis will be presented in the PhD thesis of the principal investigator and possibly published papers authored/co-authored by the principal investigator and his supervisor.

(11) Will my taking part in the study be kept confidential?

- The data collected from you will be anonymised prior to any further processing.
- To be able to fulfil a future request from your side to delete or stop using your data, an association between the data you provided and an identifier that can implicitly enable the linking of the data to you will be kept separately in an encrypted form. This association will only be available to the principal investigator.
- The supervisor of the study and the examiners of the PhD thesis of the principal investigator will be given access to anonymised data collected through the study if they so wish and solely for the purposes of validating the outcomes of the evaluation analysis carried out by the principal investigator.
- The anonymised data will be kept at City for 10 years. After that, all the acquired data will be deleted.
- The encrypted data enabling the association of the acquired data with the individuals participants of the study will be deleted as soon as the PhD thesis is examined.

(12) What should I do if I want to take part?

In order to take part, you will need to read the platform description and answer the questionnaire sent to you via email.

(13) What will happen to results of the research study?

The results of the research study will be presented in the thesis of the Principal Investigator and possibly published papers authored/co-authored by the principal investigator and his supervisor.

(14) What will happen if I do not want to carry on with the study?

You are free to withdraw from the study without an explanation or penalty at any time.

(15) Who has reviewed the study?

This study has been approved by City, University of London Research Ethics Committee.

(16) Further information and contact details

Marios Prasinos [redacted]

(17) Data Protection Privacy Notice: What are my rights under the data protection legislation?

City, University of London is the data controller for the personal data collected for this research project. Your personal data will be processed for the purposes outlined in this notice. The legal basis for processing your personal data will be that this research is a task in the public interest, that is City, University of London considers the lawful basis for processing personal data to fall under Article 6(1)(e) of GDPR (public task) as the processing of research participant data is necessary for learning and teaching purposes and all research with human participants by staff and students has to be scrutinised and approved by one of City's Research Ethics Committees.

The rights you have under the data protection legislation are listed below, but not all of the rights will be applied to the personal data collected in each research project.

- right to be informed
- right of access
- right to rectification
- right to erasure
- right to restrict processing
- right to object to data processing
- right to data portability
- right to object
- rights in relation to automated decision making and profiling

For more information, please visit www.city.ac.uk/about/city-information/legal

(18) What if I have concerns about how my personal data will be used after I have participated in the research?

In the first instance you should raise any concerns with the research team, but if you are dissatisfied with the response, you may contact the Information Compliance Team at dataprotection@city.ac.uk or phone [redacted], who will liaise with City's Data Protection Officer [redacted] to answer your query.

If you are dissatisfied with City's response you may also complain to the Information Commissioner's Office at www.ico.org.uk

(19) What if there is a problem?

If you have any problems, concerns or questions about this study, you should ask to speak to a member of the research team. If you remain unhappy and wish to complain formally, you can do this through City's complaints procedure. To complain about the study, you need to phone [redacted] You can then ask to speak to the Secretary to Senate Research Ethics Committee and inform them that the name of the project is: Evidence Based Policy Making in Healthcare using Big Data Analytics

You could also write to the Secretary at:
[redacted]
Research Integrity Manager

Research & Enterprise
City, University of London
Northampton Square
London
EC1V 0HB
Email: [redacted]

City holds insurance policies which apply to this study. If you feel you have been harmed or injured by taking part in this study, you may be eligible to claim compensation. This does not affect your legal rights to seek compensation. If you are harmed due to someone's negligence, then you may have grounds for legal action.

Thank you for taking the time to read this information sheet.

02/02/2019 version 1.0

A.8. Interview Topic Guide

Introduction

- Thank you for seeing me today and offering to take part in this study.
- I would like first to outline the study so that you are able to decide whether you wish to proceed further (recap information sheet).
- Sign consent form × 2 (one for participant and information sheet, one for interviewer).
- I have a list of topics that I want to address.
- Feel free to ask questions at any stage during the interview.
- I might make a few notes in case I want to come back to something later.
- Take your time to read the questionnaire supporting material and fill in the questionnaire.

Topics/questions

1. Does the questionnaire's usability section evaluate the platform's usability adequately?
2. Does the questionnaire's usefulness section evaluate the platform's usefulness adequately?
3. Does the questionnaire's general section boost the platforms evaluation?
4. Would you add/remove any question?
5. Do you have any further comments on any of the questions of the questionnaire?

End of interview – thank you.

A.9. Ethics Application

Ethics ETH1819-0626: Marios Prasinos (Low risk)

Date 28 Jan 2019
Researcher Marios Prasinos
[redacted]
Project Evaluation of the EVOTION public health policy tool supporting policy making decisions in healthcare using big data analytics
School School of Mathematics, Computer Science & Engineering
Department Computer Science

Ethics application

Risks

R1) Does the project have funding?
No
R2) Does the project involve human participants?
Yes
R3) Will the researcher be located outside of the UK during the conduct of the research?
Yes
R4) Will any part of the project be carried out under the auspices of an external organisation, involve collaboration between institutions, or involve data collection at an external organisation?
No
R5) Does your project involve access to, or use of, material that could be classified as security sensitive?
No
R6) Does the project involve the use of live animals?
No
R7) Does the project involve the use of animal tissue?
No
R8) Does the project involve accessing obscene materials?
No
R9) Does the project involve access to confidential business data (e.g. commercially sensitive data, trade secrets, minutes of internal meetings)?
No
R10) Does the project involve access to personal data (e.g. personnel or student records) not in the public domain?
No
R11) Does the project involve deviation from standard or routine clinical practice, outside of current guidelines?
No
R12) Will the project involve the potential for adverse impact on employment, social or financial standing?
No
R13) Will the project involve the potential for psychological distress, anxiety, humiliation or pain greater than that of normal life for the participant?
No
R15) Will the project involve research into illegal or criminal activity where there is a risk that the researcher will be placed in physical danger or in legal jeopardy?
No
R16) Will the project specifically recruit individuals who may be involved in illegal or criminal activity?
No
R17) Will the project involve engaging individuals who may be involved in terrorism, radicalisation, extremism or violent activity and other activity that falls within the Counter-Terrorism and Security Act (2015)?
No

Applicant & research team

T1) Principal Applicant
Name
Marios Prasinos
T2) Co-Applicant(s) at City
Name
[redacted]
T3) External Co-Applicant(s)

T4) Supervisor(s)

[redacted]

T5) Do any of the investigators have direct personal involvement in the organisations sponsoring or funding the research that may give rise to a possible conflict of interest?

No

T6) Will any of the investigators receive any personal benefits or incentives, including payment above normal salary, from undertaking the research or from the results of the research above those normally associated with scholarly activity?

No

T7) List anyone else involved in the project.

Project details

P1) Project title

Evaluation of the EVOTION public health policy tool supporting policy making decisions in healthcare using big data analytics

P1.1) Short project title

P2) Provide a lay summary of the background and aims of the research, including the research questions (max 400 words).

The effective management of various health conditions depends on and requires appropriate public health policies (PHP). Public health policy can affect several aspects of healthcare provision including: (a) the prevention and early diagnosis of diseases; (b) the early treatment of diagnosed conditions through the provision of the appropriate health care devices; (c) the longer term treatment of long term disabilities and chronic diseases through systematic checks of the condition of the patient, the provision of other vital related rehabilitation services; (d) the protection of people with health care devices from the harmful effects of their living environment; (e) the setup of standards, services and technology for promoting and ensuring inclusion of participation of patients with in various settings (e.g., at work, at school/educational establishments, in everyday life). Although there is a need for evidence based public health policy making, there is no computerised tool to enable that.

The overall aim of our research is to develop an integrated platform incorporating a big data analytics (BDA) platform enabling the collection and analysis of heterogeneous data related to healthcare services, including health care device usage, physiological, cognitive, medical, personal, occupational, behavioural, life style, environmental and open web data. For the purposes of the development of this integrated platform we are introducing a Public Health Policy Decision Making (PHPDM) modeling language that allows the specification of models that are executable by the platform.

The research question is the subjective evaluation of the newly introduced PHPDM modeling language and the proposed platform as a policy making tool.

P4) Provide a summary and brief explanation of the research design, method, and data analysis.

This research is aimed at providing a platform for the specification of Public Health Policy Decision Making (PHPDM) models, the execution of data analytics tasks for the provision of evidence to the stakeholders of our platform and the identification of possible interventions related to the PHPDM models specified. The PHPDM models are specified with the use of an ontology based PHPDM modeling language.

We request ethics approval for the subjective evaluation of the comprehensiveness and complexity of the PHPDM modeling language and the developed platform as a policy making tool for (a) policy makers, (b) clinicians and (c) data analysts, and whether the tool covers their needs. This activity will be based on the development of three separate questionnaires for (a) policy makers, (b) clinicians and (c) data analysts to give their feedback about the complexity and comprehensiveness of the public health policy decision making modeling language and the developed platform prototype as a policy making tool. We will also evaluate the three questionnaires based on interviews with one senior stakeholder of each group (i.e. one policy maker, one clinician and one data analyst) to validate whether each questionnaire evaluates the described platform sufficiently. The results from the interviews and the questionnaires will be presented as part of our thesis.

P4.1) If relevant, please upload your research protocol.

P5) What do you consider are the ethical issues associated with conducting this research and how do you propose to address them?

An ethical issue arises as the evaluation of our introduced platform requires interviewing and sending questionnaires to policy makers, clinicians and data analysts. GDPR entails that special precautions should be taken pertinent to handling participant's personal data. The participants will complete a consent form. Additionally, a participant information sheet will be provided to each participant with details about the study.

P6) Project start date

12 Mar 2019

P7) Anticipated project end date

24 Mar 2019

P8) Where will the research take place?
The questionnaires will be dispatched to participants in several European countries.
P10) Is this application or any part of this research project being submitted to another ethics committee, or has it previously been submitted to an ethics committee?
No

Human participants: information and participation

The options for the following question are one or more of:
'Under 18'; 'Adults at risk'; 'Individuals aged 16 and over potentially without the capacity to consent'; 'None of the above'.
H1) Will persons from any of the following groups be participating in the project?
None of the above
H2) How many participants will be recruited?
15
H3) Explain how the sample size has been determined.
We have developed three distinct types of questionnaires, each one targeting a unique professional category: one for policy makers, one for clinicians and one for data analysts. Each questionnaire type will be sent to at least five participants.
H4) What is the age group of the participants?
Lower Upper
25 70
H5) Please specify inclusion and exclusion criteria.
Experienced policy makers, clinicians and data analysts have been selected, who can evaluate such a platform independently and subjectively.
H6) What are the potential risks and burdens for research participants and how will you minimise them?
There are no risks associated with this research. We will make every possible effort to minimise the time required for filling in the questionnaire, as time is the only burden identified from participation.
H7) Will you specifically recruit pregnant women, women in labour, or women who have had a recent stillbirth or miscarriage (within the last 12 months)?
No
H8) Will you directly recruit any staff and/or students at City?
None of the above
H8.1) If you intend to contact staff/students directly for recruitment purpose, please upload a letter of approval from the respective School(s)/Department(s).
H9) How are participants to be identified, approached and recruited, and by whom?
The participants are all members of the EVOTION consortium and will be approached and recruited via email.
H10) Please upload your participant information sheets and consent form, or if they are online (e.g. on Qualtrics) paste the link below.
H11) If appropriate, please upload a copy of the advertisement, including recruitment emails, flyers or letter.
H12) Describe the procedure that will be used when seeking and obtaining consent, including when consent will be obtained.
I will obtain the consent. All arrangements will be done via email. The participants will receive the participant information sheet prior to the completion of the questionnaire. Participants have 15 calendar days to respond.
H13) Are there any pressures that may make it difficult for participants to refuse to take part in the project?
No
H14) Is any part of the research being conducted with participants outside the UK?
Yes

Human participants: method

The options for the following question are one or more of:
'Invasive procedures (for example medical or surgical)'; 'Intrusive procedures (for example psychological or social)'; 'Potentially harmful procedures of any kind'; 'Drugs, placebos, or other substances administered to participants'; 'None of the above'.
M1) Will any of the following methods be involved in the project:
None of the above
M2) Does the project involve any deceptive research practices?
No
M3) Is there a possibility for over-research of participants?
No
M4) Please upload copies of any questionnaires, topic guides for interviews or focus groups, or equivalent research materials.
M5) Will participants be provided with the findings or outcomes of the project?

Yes
M5.1) Explain how this information will be provided.
The information will be provided through my PhD thesis that will be available in public.
M6) If the research is intended to benefit the participants, third parties or the local community, please give details.
M7) Are you offering any incentives for participating?
No
M8) Does the research involve clinical trial/intervention testing that does not require Health Research Authority or MHRA approval?
No
M9) Will the project involve the collection of human tissue or other biological samples that does not fall under the Human Tissue Act (2004) that does not require Health Research Authority Research Ethics Service approval?
No
M10) Will the project involve potentially sensitive topics, such as participants' sexual behaviour, their legal or political behaviour, their experience of violence?
No
M11) Will the project involve activities that may lead to 'labelling' either by the researcher (e.g. categorisation) or by the participant (e.g. 'I'm stupid', 'I'm not normal')?
No

Data

D1) Indicate which of the following you will be using to collect your data.
Questionnaire
Interviews
D2) How will the the privacy of the participants be protected?
Any other method
D2.1) Provide details of 'any other method' used.
The consent forms must include the names of the participants or otherwise we would not be able to prove that consent was sought. However, these names will not be associated with the questionnaires that the participants will fill subsequently. Also the questionnaires do not ask for any personal information of the participants. Hence, the anonymity of the responses of the participants will be preserved.
D3) Will the research involve use of direct quotes?
No
D5) Where/how do you intend to store your data?
Storage at City
D6) Will personal data collected be shared with other organisations?
No
D7) Will the data be accessed by people other than the named researcher, supervisors or examiners?
No
D8) Is the data intended or required (e.g. by funding body) to be published for reuse or to be shared as part of longitudinal research or a different/wider research project now or in the future?
No
D10) How long are you intending to keep the research data generated by the study?
10 years.
D11) How long will personal data be stored or accessed after the study has ended?
No personal data will be stored.
D12) How are you intending to destroy the personal data after this period?
No personal data will be stored.

International research

I1) State the location(s) of your fieldwork.
Region
Europe
Country
Greece
Region
Europe
Country
Bulgaria
Region
Europe
Country
Poland

I2) Will the researcher be travelling to a country outside the UK where the Foreign & Commonwealth Office has issued an orange or red travel advisory?
No
I3) Have you identified and complied with all local requirements concerning ethical approval, research governance and data protection?
No
I4) Will the research be carried out in a country where people will be able to contact City directly using the complaints procedure?
Yes

Health & safety

HS1) Are there any health and safety risks to the researchers over and above that of their normal working life?
No
HS3) Are there hazards associated with undertaking this project where a formal risk assessment would be required?
No

A.10. Ethics Application Decision

Dear Marios
Reference: ETH1819-0626
Project title: Evaluation of the EVOTION public health policy tool supporting policy making decisions in healthcare using big data analytics
Start date: 12 Mar 2019
End date: 24 Mar 2019
I am writing to you to confirm that the research proposal detailed above has been granted formal approval from the CS/LIS Proportionate Review Committee. The Committee's response is based on the protocol described in the application form and supporting documentation. Approval has been given for the submitted application only and the research must be conducted accordingly. You are now free to start recruitment.
The approval was given with the following conditions: none.
While we are happy with your clarifications, and do not make any conditions, may we recommend that in any future ethics applications you use the term 'de-identified' rather than 'anonymous' to describe the kinds of methods used here; this will avoid your being asked for explanations.
Please ensure that you are familiar with City's Framework for Good Practice in Research and any appropriate Departmental/School guidelines, as well as applicable external relevant policies.
Please note the following:
Project amendments/extension
You will need to submit an amendment or request an extension if you wish to make any of the following changes to your research project:

- Change or add a new category of participants;
- Change or add researchers involved in the project, including PI and supervisor;
- Change to the sponsorship/collaboration;
- Add a new or change a territory for international projects;
- Change the procedures undertaken by participants, including any change relating to the safety or physical or mental integrity of research participants, or to the risk/benefit assessment for the project or collecting additional types of data from research participants;
- Change the design and/or methodology of the study, including changing or adding a new research method and/or research instrument;
- Change project documentation such as protocol, participant information sheets, consent forms, questionnaires, letters of invitation, information sheets for relatives or carers;
- Change to the insurance or indemnity arrangements for the project;
- Change the end date of the project.

Adverse events or untoward incidents
You will need to submit an Adverse Events or Untoward Incidents report in the event of any of the following:
a) Adverse events
b) Breaches of confidentiality
c) Safeguarding issues relating to children or vulnerable adults
d) Incidents that affect the personal safety of a participant or researcher
Issues a) and b) should be reported as soon as possible and no later than five days after the event. Issues c) and d) should be reported immediately. Where appropriate, the researcher should also report adverse events to other relevant institutions, such as the police or social services.
Should you have any further queries relating to this matter, please do not hesitate to contact me. On behalf of the CS/LIS Proportionate Review Committee, I do hope that the project meets with success.
Best regards
[redacted]
CS/LIS Proportionate Review Committee
City, University of London

Appendix B:
OWL XML Definitions of the Language

In this section we present the OWL XML definitions of the Public Health Policy Decision Making Modeling Language presented in Chapter 3 . In section B.1 we present the OWL XML Definition of each Language concept. In section B.2 we provide the complete OWL XML definition of the Language.

B.1. OWL XML Definitions of Language Concepts

In section B.1.1 we present the formal definitions of the Language concepts, in section B.1.2 we present the concepts' axioms and in section B.1.3 we provide the OWL XML definition of examples.

B.1.1. Formal Definitions

In this section we present the formal definition of each language concept presented in Chapter 3 .

B.1.1.1. Policy Model

Formally, policy models are defined as instances of the OWL class `PolicyModel`. The definition of this class is listed below.

```
    <Declaration>
            <Class IRI="#PolicyModel"/>
        </Declaration>
    <AnnotationAssertion>
            <AnnotationProperty abbreviatedIRI="rdfs:label"/>
            <IRI>#PolicyModel</IRI>
            <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">Policy Model</Literal>
        </AnnotationAssertion>
    <Declaration>
            <ObjectProperty IRI="#aimedAt"/>
        </Declaration>
    <ObjectPropertyDomain>
            <ObjectProperty IRI="#aimedAt"/>
```

```
        <Class IRI="#PolicyModel"/>
    </ObjectPropertyDomain>
<ObjectPropertyRange>
        <ObjectProperty IRI="#aimedAt"/>
        <Class IRI="#Goal"/>
    </ObjectPropertyRange>
<InverseObjectProperties>
        <ObjectProperty IRI="#aimedAt"/>
        <ObjectProperty IRI="#policyModel"/>
    </InverseObjectProperties>
<Declaration>
        <ObjectProperty IRI="#hasWorkflows"/>
    </Declaration>
<ObjectPropertyDomain>
        <ObjectProperty IRI="#hasWorkflows"/>
        <Class IRI="#PolicyModel"/>
    </ObjectPropertyDomain>
<ObjectPropertyRange>
        <ObjectProperty IRI="#hasWorkflows"/>
        <Class IRI="#DataAnalyticsWorfklow"/>
    </ObjectPropertyRange>
<InverseObjectProperties>
        <ObjectProperty IRI="#hasWorkflows"/>
        <ObjectProperty IRI="#policyModel"/>
    </InverseObjectProperties>
<Declaration>
        <ObjectProperty IRI="#involvesStakeholders"/>
    </Declaration>
<ObjectPropertyDomain>
        <ObjectProperty IRI="#involvesStakeholders"/>
        <Class IRI="#PolicyModel"/>
    </ObjectPropertyDomain>
<ObjectPropertyRange>
        <ObjectProperty IRI="#involvesStakeholders"/>
        <Class IRI="#Stakeholder"/>
    </ObjectPropertyRange>
<InverseObjectProperties>
        <ObjectProperty IRI="#involvesStakeholders"/>
```

```
        <ObjectProperty IRI="#policyModel"/>
    </InverseObjectProperties>
```

Formally, the relations of the `PolicyModel` class are defined as the following object properties: an object property of type `aimedAt` for the relation with the `Goal` class, an object property of type `hasWorkflows` for the relation with the `DataAnalyticsWorkflow` class and another object property of type `involvesStakeholders` for the relation with the `Stakeholder` class. The definition of these axioms in OWL is listed above.

B.1.1.2. <u>Goal</u>

Formally, policy goals are defined as instances of the class `Goal` in OWL. The definition of this class is listed below.

```
<Declaration>
        <Class IRI="#Goal"/>
    </Declaration>
<DataPropertyDomain>
        <DataProperty IRI="#description"/>
        <Class IRI="#Goal"/>
    </DataPropertyDomain>
<DataPropertyDomain>
        <DataProperty IRI="#rationale"/>
        <Class IRI="#Goal"/>
    </DataPropertyDomain>
<Declaration>
        <ObjectProperty IRI="#refinedInto"/>
    </Declaration>
<ObjectPropertyDomain>
        <ObjectProperty IRI="#refinedInto"/>
        <Class IRI="#Goal"/>
    </ObjectPropertyDomain>
<ObjectPropertyRange>
        <ObjectProperty IRI="#refinedInto"/>
        <Class IRI="#Objective"/>
    </ObjectPropertyRange>
<InverseObjectProperties>
        <ObjectProperty IRI="#refinedInto"/>
        <ObjectProperty IRI="#goal"/>
```

```
        </InverseObjectProperties>
    <Declaration>
            <Class IRI="#Remfinement"/>
        </Declaration>
    <DataPropertyDomain>
            <DataProperty IRI="#type"/>
            <Class IRI="#Remfinement"/>
        </DataPropertyDomain>
        <DataPropertyRange>
            <DataProperty IRI="#type"/>
            <DataOneOf>
                <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">DISJUNCTIVE</Literal>
                <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">CONJUCTIVE</Literal>
            </DataOneOf>
        </DataPropertyRange>
    <Declaration>
            <ObjectProperty IRI="#policyModel"/>
        </Declaration>
    <ObjectPropertyDomain>
            <ObjectProperty IRI="#policyModel"/>
            <Class IRI="#Goal"/>
        </ObjectPropertyDomain>
    <ObjectPropertyRange>
            <ObjectProperty IRI="#policyModel"/>
            <Class IRI="#PolicyModel"/>
        </ObjectPropertyRange>
    <InverseObjectProperties>
            <ObjectProperty IRI="#policyModel"/>
            <ObjectProperty IRI="#aimedAt"/>
        </InverseObjectProperties>
```

The above declaration also includes the two data properties, `description` and `rationale` and `type` of the `Refinement` class.

Formally, the relations of the `Goal` class are defined as the following object properties: an object property of type `refinedInto` for the relation with the `Objective` class and an object property of type `policyModel`, which is the inverse object property of `aimedAt`, for the relationship with `PolicyModel` class. The definition of these object properties is presented above.

B.1.1.3. Objective

Formally, policy objectives are defined as instances of the class `Objective` in OWL. The definition of this class is listed below.

```
<Declaration>
        <Class IRI="#Objective"/>
    </Declaration>
<DataPropertyDomain>
        <DataProperty IRI="#description"/>
        <Class IRI="#Objective"/>
    </DataPropertyDomain>
<DataPropertyDomain>
        <DataProperty IRI="#rationale"/>
        <Class IRI="#Objective"/>
    </DataPropertyDomain>
<Declaration>
        <ObjectProperty IRI="#goal"/>
    </Declaration>
<ObjectPropertyDomain>
        <ObjectProperty IRI="#goal"/>
        <Class IRI="#Objective"/>
    </ObjectPropertyDomain>
<ObjectPropertyRange>
        <ObjectProperty IRI="#goal"/>
        <Class IRI="#Goal"/>
    </ObjectPropertyRange>
<InverseObjectProperties>
        <ObjectProperty IRI="#goal"/>
        <ObjectProperty IRI="#refinedInto"/>
    </InverseObjectProperties>
<Declaration>
        <ObjectProperty IRI="#canBeAddressedBy"/>
    </Declaration>
<ObjectPropertyDomain>
        <ObjectProperty IRI="#canBeAddressedBy"/>
```

```
          <Class IRI="#Objective"/>
      </ObjectPropertyDomain>
   <ObjectPropertyRange>
          <ObjectProperty IRI="#canBeAddressedBy"/>
          <Class IRI="#PolicyAction"/>
      </ObjectPropertyRange>
   <InverseObjectProperties>
          <ObjectProperty IRI="#canBeAddressedBy"/>
          <ObjectProperty IRI="#objective"/>
      </InverseObjectProperties>
```

The above declaration also includes the two data properties: `description` and `rationale`.

Formally, the relations of the `Objective` class are defined the following object properties: an object property of type `Goal` for the relation with the `Goal` class and an object property of type `canBeAddressedBy` for the relationship with `PolicyAction` class.

B.1.1.4. Policy Action

Formally, policy actions are defined as instances of the class `PolicyAction` in OWL. The definition of this class is listed below.

```
   <Declaration>
          <Class IRI="#PolicyAction"/>
      </Declaration>
   <AnnotationAssertion>
          <AnnotationProperty abbreviatedIRI="rdfs:label"/>
          <IRI>#PolicyAction</IRI>
          <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">Policy Action</Literal>
      </AnnotationAssertion>
   <DataPropertyDomain>
          <DataProperty IRI="#description"/>
          <Class IRI="#PolicyAction"/>
      </DataPropertyDomain>
   <Declaration>
          <ObjectProperty IRI="#objective"/>
      </Declaration>
   <ObjectPropertyDomain>
          <ObjectProperty IRI="#objective"/>
```

```
        <Class IRI="#PolicyAction"/>
    </ObjectPropertyDomain>
<ObjectPropertyRange>
        <ObjectProperty IRI="#objective"/>
        <Class IRI="#Objective"/>
    </ObjectPropertyRange>
<InverseObjectProperties>
        <ObjectProperty IRI="#objective"/>
        <ObjectProperty IRI="#canBeAddressedBy"/>
  </InverseObjectProperties>
<Declaration>
        <ObjectProperty IRI="#a1"/>
    </Declaration>
<ObjectPropertyDomain>
        <ObjectProperty IRI="#a1"/>
        <Class IRI="#PolicyAction"/>
    </ObjectPropertyDomain>
<ObjectPropertyRange>
        <ObjectProperty IRI="#a1"/>
        <Class IRI="#PolicyAction"/>
    </ObjectPropertyRange>
<Declaration>
        <ObjectProperty IRI="#a2"/>
    </Declaration>
<ObjectPropertyDomain>
        <ObjectProperty IRI="#a2"/>
        <Class IRI="#PolicyAction"/>
    </ObjectPropertyDomain>
<ObjectPropertyRange>
        <ObjectProperty IRI="#a2"/>
        <Class IRI="#PolicyAction"/>
    </ObjectPropertyRange>
<InverseObjectProperties>
        <ObjectProperty IRI="#a1"/>
        <ObjectProperty IRI="#a2"/>
  </InverseObjectProperties>
   <Declaration>
        <ObjectProperty IRI="#dependant"/>
    </Declaration>
```

```
<ObjectPropertyDomain>
        <ObjectProperty IRI="#dependant"/>
        <Class IRI="#PolicyAction"/>
    </ObjectPropertyDomain>
<ObjectPropertyRange>
        <ObjectProperty IRI="#dependant"/>
        <Class IRI="#PolicyAction"/>
</ObjectPropertyRange>
<Declaration>
        <ObjectProperty IRI="#pre-requisite"/>
</Declaration>
<ObjectPropertyDomain>
        <ObjectProperty IRI="#pre-requisite"/>
        <Class IRI="#PolicyAction"/>
    </ObjectPropertyDomain>
<ObjectPropertyRange>
        <ObjectProperty IRI="#pre-requisite"/>
        <Class IRI="#PolicyAction"/>
    </ObjectPropertyRange>
<InverseObjectProperties>
        <ObjectProperty IRI="#dependant"/>
        <ObjectProperty IRI="#pre-requisite"/>
  </InverseObjectProperties>
      <Declaration>
        <ObjectProperty IRI="#position"/>
</Declaration>
<ObjectPropertyDomain>
        <ObjectProperty IRI="#position"/>
        <Class IRI="#PolicyAction"/>
    </ObjectPropertyDomain>
<ObjectPropertyRange>
        <ObjectProperty IRI="#position"/>
        <Class IRI="#Position"/>
    </ObjectPropertyRange>
<InverseObjectProperties>
        <ObjectProperty IRI="#position"/>
        <ObjectProperty IRI="#refersTo"/>
  </InverseObjectProperties>
       <Declaration>
```

```
        <ObjectProperty IRI="#isEvaluatedBy"/>
</Declaration>
<ObjectPropertyDomain>
        <ObjectProperty IRI="#isEvaluatedBy"/>
        <Class IRI="#PolicyAction"/>
    </ObjectPropertyDomain>
<ObjectPropertyRange>
        <ObjectProperty IRI="#isEvaluatedBy"/>
        <Class IRI="#Criterion"/>
    </ObjectPropertyRange>
<InverseObjectProperties>
        <ObjectProperty IRI="#isEvaluatedBy"/>
        <ObjectProperty IRI="#policyAction"/>
  </InverseObjectProperties>
```

The above declaration also includes the data property `description`.

Formally, the relations of the `PolicyAction` class are defined as the following object properties: an object property of type `objective` for the relation with the `Objective` class, four object properties of types: `a1`, a2, `dependent` and `pre-requisite` for the relationships with itself, an object property of type `position` for the relationship with `Position` class and an object property of type `isEvaluatedBy` for the relationship with `Criterion` class. The definition of these object properties is presented above.

B.1.1.5. <u>Criterion</u>

Formally, criteria are defined as instances of the class `Criterion` in OWL. The definition of this class is listed below.

```
<Declaration>
        <Class IRI="#Criterion"/>
    </Declaration>
<DataPropertyDomain>
        <DataProperty IRI="#weight"/>
        <Class IRI="#Criterion"/>
    </DataPropertyDomain>
<DataPropertyDomain>
        <DataProperty IRI="#description"/>
        <Class IRI="#Criterion"/>
    </DataPropertyDomain>
```

```
<DataPropertyDomain>
        <DataProperty IRI="#logicalExpression"/>
        <Class IRI="#Criterion"/>
    </DataPropertyDomain>
<Declaration>
        <ObjectProperty IRI="#constraints"/>
    </Declaration>
<ObjectPropertyDomain>
        <ObjectProperty IRI="#constraints"/>
        <Class IRI="#Criterion"/>
    </ObjectPropertyDomain>
<ObjectPropertyRange>
        <ObjectProperty IRI="#constraints"/>
        <Class IRI="#Dataset"/>
    </ObjectPropertyRange>
<InverseObjectProperties>
        <ObjectProperty IRI="#constraints"/>
        <ObjectProperty IRI="#criterion"/>
  </InverseObjectProperties>
<Declaration>
        <ObjectProperty IRI="#specifies"/>
    </Declaration>
<ObjectPropertyDomain>
        <ObjectProperty IRI="#specifies"/>
        <Class IRI="#Criterion"/>
    </ObjectPropertyDomain>
<ObjectPropertyRange>
        <ObjectProperty IRI="#specifies"/>
        <Class IRI="#DataAnalyticsWorfklow"/>
    </ObjectPropertyRange>
<InverseObjectProperties>
        <ObjectProperty IRI="#specifies"/>
        <ObjectProperty IRI="#criterion"/>
  </InverseObjectProperties>
<Declaration>
        <ObjectProperty IRI="#policyAction"/>
    </Declaration>
<ObjectPropertyDomain>
        <ObjectProperty IRI="#policyAction"/>
```

```
            <Class IRI="#Criterion"/>
        </ObjectPropertyDomain>
    <ObjectPropertyRange>
            <ObjectProperty IRI="#policyAction"/>
            <Class IRI="#PolicyAction"/>
        </ObjectPropertyRange>
```

The above declaration also includes the data properties: `weight`, `description` and `logicalExpression`.

Formally, the relations of the `Criterion` class are defined as the following object properties: an object property of type `constraints` for the relation with the `Dataset` class, an object property of type `specifies` for the relationships with `DataAnalyticsWorkflow` and an object property of type `policyAction` for the relationship with `PolicyAction` class.

B.1.1.6. ***Stakeholder***

Formally, stakeholders are defined as instances of the class `Stakeholder` in OWL. The definition of this class is listed below.

```
    <Declaration>
            <Class IRI="#Stakeholder"/>
        </Declaration>
    <Declaration>
            <ObjectProperty IRI="#policyModel"/>
        </Declaration>
    <ObjectPropertyDomain>
            <ObjectProperty IRI="#policyModel"/>
            <Class IRI="#Stakeholder"/>
        </ObjectPropertyDomain>
    <ObjectPropertyRange>
            <ObjectProperty IRI="#policyModel"/>
            <Class IRI="#PolicyModel"/>
        </ObjectPropertyRange>
    <InverseObjectProperties>
            <ObjectProperty IRI="#policyModel"/>
            <ObjectProperty IRI="#involvesStakeholders"/>
      </InverseObjectProperties>
    <Declaration>
            <ObjectProperty IRI="#advocates"/>
```

```
    </Declaration>
<ObjectPropertyDomain>
        <ObjectProperty IRI="#advocates"/>
        <Class IRI="#Stakeholder"/>
    </ObjectPropertyDomain>
<ObjectPropertyRange>
        <ObjectProperty IRI="#advocates"/>
        <Class IRI="#Position"/>
    </ObjectPropertyRange>
<Declaration>
        <ObjectProperty IRI="#proposes"/>
    </Declaration>
<ObjectPropertyDomain>
        <ObjectProperty IRI="#proposes"/>
        <Class IRI="#Stakeholder"/>
    </ObjectPropertyDomain>
<ObjectPropertyRange>
        <ObjectProperty IRI="#proposes"/>
        <Class IRI="#Position"/>
    </ObjectPropertyRange>
```

Formally, the relations of the `Stakeholder` class are defined as the following object properties: an object property of type `involvesStakeholders` for the relation with the `PolicyModel` class, and two object properties for the relationships with `Position`: one of type `advocates` and one of type `proposes`.

B.1.1.7. Position

Formally, positions are defined as instances of the class `Position` in OWL. The definitions of this class and its subclasses are listed below.

```
<Declaration>
        <Class IRI="#Position"/>
    </Declaration>
<Declaration>
        <Class IRI="#OpposingPosition"/>
    </Declaration>
<Declaration>
        <Class IRI="#NeutralPosition"/>
```

```
        </Declaration>
    <Declaration>
            <Class IRI="#SupportivePosition"/>
        </Declaration>
    <AnnotationAssertion>
            <AnnotationProperty abbreviatedIRI="rdfs:label"/>
            <IRI>#OpposingPosition</IRI>
            <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">OpposingPosition</Literal>
        </AnnotationAssertion>
    <AnnotationAssertion>
            <AnnotationProperty abbreviatedIRI="rdfs:label"/>
            <IRI>#NeutralPosition</IRI>
            <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">Neutral Position</Literal>
        </AnnotationAssertion>
    <AnnotationAssertion>
            <AnnotationProperty abbreviatedIRI="rdfs:label"/>
            <IRI>#SupportivePosition</IRI>
            <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">Supportive Position</Literal>
        </AnnotationAssertion>
    <SubClassOf>
            <Class IRI="#OpposingPosition"/>
            <Class IRI="#Position"/>
        </SubClassOf>
    <SubClassOf>
            <Class IRI="#NeutralPosition"/>
            <Class IRI="#Position"/>
        </SubClassOf>
    <SubClassOf>
            <Class IRI="#SupportivePosition"/>
            <Class IRI="#Position"/>
        </SubClassOf>
    <Declaration>
            <ObjectProperty IRI="#refersTo"/>
        </Declaration>
    <ObjectPropertyDomain>
```

```
        <ObjectProperty IRI="#refersTo"/>
        <Class IRI="#PolicyAction"/>
    </ObjectPropertyDomain>
<ObjectPropertyRange>
        <ObjectProperty IRI="#refersTo"/>
        <Class IRI="#Position"/>
    </ObjectPropertyRange>
<InverseObjectProperties>
        <ObjectProperty IRI="#refersTo"/>
        <ObjectProperty IRI="#position"/>
  </InverseObjectProperties>
<Declaration>
        <ObjectProperty IRI="#stakeholder"/>
    </Declaration>
<ObjectPropertyDomain>
        <ObjectProperty IRI="#stakeholder"/>
        <Class IRI="#Position"/>
    </ObjectPropertyDomain>
<ObjectPropertyRange>
        <ObjectProperty IRI="#stakeholder"/>
        <Class IRI="#Stakeholder"/>
    </ObjectPropertyRange>
<InverseObjectProperties>
        <ObjectProperty IRI="#proposes"/>
        <ObjectProperty IRI="#stakeholder"/>
  </InverseObjectProperties>
<InverseObjectProperties>
        <ObjectProperty IRI="#advocates"/>
        <ObjectProperty IRI="#stakeholder"/>
  </InverseObjectProperties>
```

Formally, the relations of the `Position` class are defined as the following object properties: one object property of type `refersTo` for the relationship with `PolicyAction` and one object property for the relationship with `Stakeholder` of type `stakeholder`.

B.1.1.8. Data Analytics Workflow

Formally, data analytics workflows are defined as instances of the class `DataAnalyticsWorkflow` in OWL. The definition of this is listed below.

```
  <Declaration>
          <Class IRI="#DataAnalyticsWorfklow"/>
      </Declaration>
  <AnnotationAssertion>
          <AnnotationProperty abbreviatedIRI="rdfs:label"/>
          <IRI>#DataAnalyticsWorfklow </IRI>
          <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">DataAnalyticsWorfklow</Literal>
      </AnnotationAssertion>
  <AnnotationAssertion>
  <Declaration>
          <ObjectProperty IRI="#criterion"/>
      </Declaration>
  <ObjectPropertyDomain>
          <ObjectProperty IRI="#criterion"/>
          <Class IRI="#DataAnalyticsWorfklow"/>
      </ObjectPropertyDomain>
  <ObjectPropertyRange>
          <ObjectProperty IRI="#criterion"/>
          <Class IRI="#Criterion"/>
      </ObjectPropertyRange>
  <InverseObjectProperties>
          <ObjectProperty IRI="#specifies"/>
          <ObjectProperty IRI="#criterion"/>
    </InverseObjectProperties>
  <Declaration>
          <ObjectProperty IRI="#isComposedOf"/>
      </Declaration>
  <ObjectPropertyDomain>
          <ObjectProperty IRI="#isComposedOf"/>
          <Class IRI="#DataAnalyticsWorfklow"/>
      </ObjectPropertyDomain>
  <ObjectPropertyRange>
          <ObjectProperty IRI="#isComposedOf"/>
          <Class IRI="#DataAnalyticsTask"/>
      </ObjectPropertyRange>
  <InverseObjectProperties>
          <ObjectProperty IRI="#isComposedOf"/>
```

```
        <ObjectProperty IRI="#dataAnalyticsWorkflow"/>
    </InverseObjectProperties>
    <Declaration>
        <ObjectProperty IRI="#policyModel"/>
    </Declaration>
    <ObjectPropertyDomain>
        <ObjectProperty IRI="#policyModel"/>
        <Class IRI="#DataAnalyticsWorfklow"/>
    </ObjectPropertyDomain>
    <ObjectPropertyRange>
        <ObjectProperty IRI="#policyModel"/>
        <Class IRI="#PolicyModel"/>
    </ObjectPropertyRange>
    <InverseObjectProperties>
        <ObjectProperty IRI="#hasWorkflows"/>
        <ObjectProperty IRI="#policyModel"/>
    </InverseObjectProperties>
```

Formally, the relations of the `DataAnalyticsWorkflow` class are defined as the following object properties: one object property of type `criterion` for the relationship with `Criterion`, one object property of type `isComposedOf` for the relationship with `DataAnalyticsTask` and an object property of type `policyModel` for the relationship with `PolicyModel`.

B.1.1.9. Workflow Execution Type

Formally, workflow execution types are defined as instances of the class `WorkflowExecutionType` in OWL. The definition of this class and its subclasses are listed below.

```
    <Declaration>
        <Class IRI="#WorkflowExecutionType"/>
    </Declaration>
    <Declaration>
        <Class IRI="#ExecutionUponRequest"/>
    </Declaration>
    <Declaration>
        <Class IRI="#PeriodicExecution"/>
    </Declaration>
    <Declaration>
        <Class IRI="#DataChangeDriven"/>
```

```
   </Declaration>
   <AnnotationAssertion>
           <AnnotationProperty abbreviatedIRI="rdfs:label"/>
           <IRI>#WorkflowExecutionType</IRI>
           <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">Workflow Execution Type</Literal>
       </AnnotationAssertion>
   <AnnotationAssertion>
           <AnnotationProperty abbreviatedIRI="rdfs:label"/>
           <IRI>#ExecutionUponRequest</IRI>
           <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">Execution Upon Request</Literal>
       </AnnotationAssertion>
   <AnnotationAssertion>
           <AnnotationProperty abbreviatedIRI="rdfs:label"/>
           <IRI>#PeriodicExecution</IRI>
           <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">Periodic Execution</Literal>
       </AnnotationAssertion>
   <AnnotationAssertion>
           <AnnotationProperty abbreviatedIRI="rdfs:label"/>
           <IRI>#DataChangeDriven</IRI>
           <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">DataChangeDriven</Literal>
       </AnnotationAssertion>
   <SubClassOf>
           <Class IRI="#ExecutionUponRequest"/>
           <Class IRI="#WorkflowExecutionType"/>
       </SubClassOf>
   <SubClassOf>
           <Class IRI="#PeriodicExecution"/>
           <Class IRI="#WorkflowExecutionType"/>
       </SubClassOf>
   <SubClassOf>
           <Class IRI="#DataChangeDriven"/>
           <Class IRI="#WorkflowExecutionType"/>
       </SubClassOf>
```

B.1.1.10. <u>Data Set</u>

Formally, data sets are defined as instances of the class `Dataset` in OWL. The definition of this class and its subclasses, as well as their related classes are listed below.

```
<Declaration>
        <Class IRI="#Dataset"/>
    </Declaration>
<Declaration>
        <Class IRI="#StaticSet"/>
    </Declaration>
<Declaration>
        <Class IRI="#DataStream"/>
    </Declaration>
<Declaration>
        <Class IRI="#DataAnalyticsModel"/>
    </Declaration>
<Declaration>
        <Class IRI="#TimedDataset"/>
    </Declaration>
<Declaration>
        <Class IRI="#AbsoluteDataset"/>
    </Declaration>
<Declaration>
        <Class IRI="#ShiftingDataset"/>
    </Declaration>
<AnnotationAssertion>
        <AnnotationProperty abbreviatedIRI="rdfs:label"/>
        <IRI>#StaticSet</IRI>
        <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">Static Set</Literal>
    </AnnotationAssertion>
<AnnotationAssertion>
        <AnnotationProperty abbreviatedIRI="rdfs:label"/>
        <IRI>#DataStream</IRI>
        <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">DataStream</Literal>
    </AnnotationAssertion>
<AnnotationAssertion>
```

```
            <AnnotationProperty abbreviatedIRI="rdfs:label"/>
            <IRI>#DataAnalyticsModel</IRI>
            <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-ns#PlainLiteral">Data Analytics Model</Literal>
        </AnnotationAssertion>
    <AnnotationAssertion>
            <AnnotationProperty abbreviatedIRI="rdfs:label"/>
            <IRI>#TimedDataset</IRI>
            <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-ns#PlainLiteral">Timed Dataset</Literal>
        </AnnotationAssertion>
    <AnnotationAssertion>
            <AnnotationProperty abbreviatedIRI="rdfs:label"/>
            <IRI>#AbsoluteDataset</IRI>
            <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-ns#PlainLiteral">Absolute Dataset</Literal>
        </AnnotationAssertion>
    <AnnotationAssertion>
            <AnnotationProperty abbreviatedIRI="rdfs:label"/>
            <IRI>#ShiftingDataset</IRI>
            <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-ns#PlainLiteral">Shifting Dataset</Literal>
        </AnnotationAssertion>
    <SubClassOf>
            <Class IRI="#StaticSet"/>
            <Class IRI="#Dataset"/>
        </SubClassOf>
    <SubClassOf>
            <Class IRI="#StaticSet"/>
            <Class IRI="#Dataset"/>
        </SubClassOf>
    <SubClassOf>
            <Class IRI="#DataAnalyticsModel"/>
            <Class IRI="#Dataset"/>
        </SubClassOf>
    <SubClassOf>
            <Class IRI="#AbsoluteDataset"/>
            <Class IRI="#TimedDataset"/>
```

```
    </SubClassOf>
<SubClassOf>
        <Class IRI="#ShiftingDataset"/>
        <Class IRI="#TimedDataset"/>
    </SubClassOf>
```

Formally, the relations of the `Dataset` class are defined as the following object properties: one object property of type `criterion` for the relationship with `Criterion`, one object property of type `dataAnalyticsTask` for the relationships with `DataAnalyticsTask` (one inverse of input and one inverse of output) and one object property of type `spec` for the relationship with `DataSpecification`.

```
<Declaration>
        <ObjectProperty IRI="#criterion"/>
    </Declaration>
<ObjectPropertyDomain>
        <ObjectProperty IRI="#criterion"/>
        <Class IRI="#Dataset"/>
    </ObjectPropertyDomain>
<ObjectPropertyRange>
        <ObjectProperty IRI="#criterion"/>
        <Class IRI="#Criterion"/>
    </ObjectPropertyRange>
<InverseObjectProperties>
        <ObjectProperty IRI="#constraints"/>
        <ObjectProperty IRI="#criterion"/>
    </InverseObjectProperties>
<Declaration>
        <ObjectProperty IRI="#dataAnalyticsTask"/>
    </Declaration>
<ObjectPropertyDomain>
        <ObjectProperty IRI="#dataAnalyticsTask"/>
        <Class IRI="#Dataset"/>
    </ObjectPropertyDomain>
<ObjectPropertyRange>
        <ObjectProperty IRI="#dataAnalyticsTask"/>
        <Class IRI="#DataAnalyticsTask"/>
    </ObjectPropertyRange>
```

```
<InverseObjectProperties>
        <ObjectProperty IRI="#input"/>
        <ObjectProperty IRI="#dataAnalyticsTask"/>
    </InverseObjectProperties>
<InverseObjectProperties>
        <ObjectProperty IRI="#output"/>
        <ObjectProperty IRI="#dataAnalyticsTask"/>
    </InverseObjectProperties>
<Declaration>
        <ObjectProperty IRI="#spec"/>
    </Declaration>
<ObjectPropertyDomain>
        <ObjectProperty IRI="#spec"/>
        <Class IRI="#Dataset"/>
    </ObjectPropertyDomain>
<ObjectPropertyRange>
        <ObjectProperty IRI="#spec"/>
        <Class IRI="#DataSpecification"/>
    </ObjectPropertyRange>
<InverseObjectProperties>
        <ObjectProperty IRI="#dataset"/>
        <ObjectProperty IRI="#spec"/>
    </InverseObjectProperties>
```

B.1.1.11. Data Specification

Formally, data specifications are defined as instances of the class `DataSpecification` in OWL. The definition of this class and its subclass are listed below.

```
<Declaration>
        <Class IRI="#DataSpecification"/>
    </Declaration>
<Declaration>
        <Class IRI="#OutputDataSpecification"/>
    </Declaration>
<AnnotationAssertion>
        <AnnotationProperty abbreviatedIRI="rdfs:label"/>
        <IRI>#DataSpecification</IRI>
```

```
          <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">Data Specification</Literal>
      </AnnotationAssertion>
  <AnnotationAssertion>
          <AnnotationProperty abbreviatedIRI="rdfs:label"/>
          <IRI>#OutputDataSpecification</IRI>
          <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">Output Data Specification</Literal>
      </AnnotationAssertion>
  <SubClassOf>
          <Class IRI="#OutputDataSpecification"/>
          <Class IRI="#Data_Specification"/>
      </SubClassOf>
  <Declaration>
      <DataProperty IRI="#tableName"/>
    </Declaration>
    <DataPropertyDomain>
      <DataProperty IRI="#tableName"/>
      <Class IRI="#DataSpecification"/>
    </DataPropertyDomain>
    <DataPropertyRange>
      <DataProperty IRI="#tableName"/>
      <Datatype abbreviatedIRI="xsd:string"/>
    </DataPropertyRange>
    <Declaration>
      <DataProperty IRI="#columnName"/>
    </Declaration>
    <DataPropertyDomain>
      <DataProperty IRI="#columnName"/>
      <Class IRI="#DataSpecification"/>
    </DataPropertyDomain>
    <DataPropertyRange>
      <DataProperty IRI="#columnName"/>
      <Datatype abbreviatedIRI="xsd:string"/>
    </DataPropertyRange>
```

Formally, the relations of the `DataSpecification` and `OutputDataSpecification` classes are defined as the following object properties: one object property of type `dataset` for the relationship with `DataSet` and another object property of type `algorithm` for the relationship of `Algorithm` with `OutputDataSpecification`.

```
<Declaration>
        <ObjectProperty IRI="#dataset"/>
    </Declaration>
<ObjectPropertyDomain>
        <ObjectProperty IRI="#dataset"/>
        <Class IRI="#DataSpecification"/>
    </ObjectPropertyDomain>
<ObjectPropertyRange>
        <ObjectProperty IRI="#dataset"/>
        <Class IRI="#Dataset"/>
    </ObjectPropertyRange>
<InverseObjectProperties>
        <ObjectProperty IRI="#spec"/>
        <ObjectProperty IRI="#dataset"/>
  </InverseObjectProperties>
<Declaration>
        <ObjectProperty IRI="#algorithm"/>
    </Declaration>
<ObjectPropertyDomain>
        <ObjectProperty IRI="#algorithm"/>
        <Class IRI="#OutputDataSpecification"/>
    </ObjectPropertyDomain>
<ObjectPropertyRange>
        <ObjectProperty IRI="#algorithm"/>
        <Class IRI="#Algorithm"/>
    </ObjectPropertyRange>
<InverseObjectProperties>
        <ObjectProperty IRI="#dataSpec"/>
        <ObjectProperty IRI="#algorithm"/>
  </InverseObjectProperties>
```

B.1.1.12. Data Analytics Task

Formally, data analytics tasks are defined as instances of the class `DataAnalyticsTask` in OWL. The definition of this class and its subclasses are listed below.

```
   <Declaration>
           <Class IRI="#DataAnalyticsTask"/>
       </Declaration>
   <Declaration>
           <Class IRI="#DataProcessingTask"/>
       </Declaration>
    <Declaration>
           <Class IRI="#DataMiningTask"/>
       </Declaration>
   <Declaration>
           <Class IRI="#StatisticalAnalysisTask"/>
       </Declaration>
   <AnnotationAssertion>
           <AnnotationProperty abbreviatedIRI="rdfs:label"/>
           <IRI>#DataAnalyticsTask</IRI>
           <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">Data Analytics Task</Literal>
       </AnnotationAssertion>
   <AnnotationAssertion>
           <AnnotationProperty abbreviatedIRI="rdfs:label"/>
           <IRI>#DataProcessingTask</IRI>
           <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">Data Processing Task</Literal>
       </AnnotationAssertion>
   <AnnotationAssertion>
           <AnnotationProperty abbreviatedIRI="rdfs:label"/>
           <IRI>#DataMiningTask</IRI>
           <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">Data Mining Task</Literal>
       </AnnotationAssertion>
   <AnnotationAssertion>
           <AnnotationProperty abbreviatedIRI="rdfs:label"/>
           <IRI>#StatisticalAnalysisTask</IRI>
```

```
          <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-ns#PlainLiteral">Statistical Analysis Task</Literal>
      </AnnotationAssertion>
   <SubClassOf>
          <Class IRI="#DataProcessingTask"/>
          <Class IRI="#DataAnalyticsTask"/>
      </SubClassOf>
   <SubClassOf>
          <Class IRI="#DataMiningTask"/>
          <Class IRI="#DataAnalyticsTask"/>
   </SubClassOf>
   <SubClassOf>
          <Class IRI="#StatisticalAnalysisTask"/>
          <Class IRI="#DataAnalyticsTask"/>
   </SubClassOf>
```

Formally, the relations of the `DataAnalyticsTask` are defined as the following object properties: one object property of type `input` and one of type `output` for the relationship with `Dataset`, one object property of type `utilizes` for the relationship with `Method` and one of type `isComposedOf` for the relationship with `DataAnalyticsWorkflow`.

```
   <Declaration>
          <ObjectProperty IRI="#input"/>
      </Declaration>
   <ObjectPropertyDomain>
          <ObjectProperty IRI="#input"/>
          <Class IRI="#DataAnalyticsTask"/>
      </ObjectPropertyDomain>
   <ObjectPropertyRange>
          <ObjectProperty IRI="#input"/>
          <Class IRI="#Dataset"/>
      </ObjectPropertyRange>
   <InverseObjectProperties>
          <ObjectProperty IRI="#dataset"/>
          <ObjectProperty IRI="#input"/>
   </InverseObjectProperties>
   <Declaration>
          <ObjectProperty IRI="#output"/>
      </Declaration>
```

```
<ObjectPropertyDomain>
        <ObjectProperty IRI="#output"/>
        <Class IRI="#DataAnalyticsTask"/>
    </ObjectPropertyDomain>
<ObjectPropertyRange>
        <ObjectProperty IRI="#output"/>
        <Class IRI="#Dataset"/>
    </ObjectPropertyRange>
<InverseObjectProperties>
        <ObjectProperty IRI="#dataset"/>
        <ObjectProperty IRI="#output"/>
  </InverseObjectProperties>
<Declaration>
        <ObjectProperty IRI="#utilizes"/>
    </Declaration>
<ObjectPropertyDomain>
        <ObjectProperty IRI="#utilizes"/>
        <Class IRI="#DataAnalyticsTask"/>
    </ObjectPropertyDomain>
<ObjectPropertyRange>
        <ObjectProperty IRI="#utilizes"/>
        <Class IRI="#Method"/>
    </ObjectPropertyRange>
<InverseObjectProperties>
        <ObjectProperty IRI="#utilizes"/>
        <ObjectProperty IRI="#dataAnalyticsTask"/>
  </InverseObjectProperties>
<Declaration>
        <ObjectProperty IRI="#dataAnalyticsWorkflow"/>
    </Declaration>
<ObjectPropertyDomain>
        <ObjectProperty IRI="#dataAnalyticsWorkflow"/>
        <Class IRI="#DataAnalyticsTask"/>
    </ObjectPropertyDomain>
<ObjectPropertyRange>
        <ObjectProperty IRI="#dataAnalyticsWorkflow"/>
        <Class IRI="#DataAnalyticsWorfklow"/>
    </ObjectPropertyRange>
<InverseObjectProperties>
```

```
        <ObjectProperty IRI="#isComposedOf"/>
        <ObjectProperty IRI="#dataAnalyticsWorkflow"/>
    </InverseObjectProperties>
```

B.1.1.13. Method

Formally, methods are defined as instances of the class `Method` in OWL. The definition of this class and its subclasses, as well as the subclasses of `Operation` are listed below.

```
 <Declaration>
        <Class IRI="#Method"/>
    </Declaration>
 <Declaration>
        <Class IRI="#Operation"/>
    </Declaration>
  <Declaration>
        <Class IRI="#Algorithm"/>
    </Declaration>
 <Declaration>
        <Class IRI="#DataCleaningOperation"/>
    </Declaration>
 <Declaration>
        <Class IRI="#SamplingOperation"/>
    </Declaration>
 <Declaration>
        <Class IRI="#FilterOperation"/>
    </Declaration>
 <Declaration>
        <Class IRI="#ProjectOperation"/>
    </Declaration>
 <Declaration>
        <Class IRI="#JoinOperation"/>
    </Declaration>
 <AnnotationAssertion>
        <AnnotationProperty abbreviatedIRI="rdfs:label"/>
        <IRI>#DataCleaningOperation</IRI>
        <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">Data Cleaning Operation</Literal>
    </AnnotationAssertion>
```

```
   <AnnotationAssertion>
           <AnnotationProperty abbreviatedIRI="rdfs:label"/>
           <IRI>#SamplingOperation</IRI>
           <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">Sampling Operation</Literal>
       </AnnotationAssertion>
   <AnnotationAssertion>
           <AnnotationProperty abbreviatedIRI="rdfs:label"/>
           <IRI>#FilterOperation</IRI>
           <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">Filter Operation</Literal>
       </AnnotationAssertion>
   <AnnotationAssertion>
           <AnnotationProperty abbreviatedIRI="rdfs:label"/>
           <IRI>#ProjectOperation</IRI>
           <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">Project Operation</Literal>
       </AnnotationAssertion>
   <AnnotationAssertion>
           <AnnotationProperty abbreviatedIRI="rdfs:label"/>
           <IRI>#JoinOperation</IRI>
           <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">Join Operation</Literal>
       </AnnotationAssertion>
   <SubClassOf>
           <Class IRI="#Operation"/>
           <Class IRI="#Method"/>
       </SubClassOf>
   <SubClassOf>
           <Class IRI="#Algorithm"/>
           <Class IRI="#Method"/>
       </SubClassOf>
   <SubClassOf>
           <Class IRI="#DataCleaningOperation"/>
           <Class IRI="#Operation"/>
       </SubClassOf>
   <SubClassOf>
           <Class IRI="#SamplingOperation"/>
```

```
            <Class IRI="#Operation"/>
        </SubClassOf>
    <SubClassOf>
            <Class IRI="#FilterOperation"/>
            <Class IRI="#Operation"/>
        </SubClassOf>
    <SubClassOf>
            <Class IRI="#ProjectOperation"/>
            <Class IRI="#Operation"/>
        </SubClassOf>
    <SubClassOf>
            <Class IRI="#JoinOperation"/>
            <Class IRI="#Operation"/>
        </SubClassOf>
```

Formally, the relation of the Method class is defined as an object property of type `dataAnalyticsTask` for the relationship with `DataAnalyticsTask` class.

```
      <Declaration>
            <ObjectProperty IRI="#dataAnalyticsTask"/>
        </Declaration>
    <ObjectPropertyDomain>
            <ObjectProperty IRI="#dataAnalyticsTask"/>
            <Class IRI="#Method"/>
        </ObjectPropertyDomain>
    <ObjectPropertyRange>
            <ObjectProperty IRI="#dataAnalyticsTask"/>
            <Class IRI="#DataAnalyticsTask"/>
        </ObjectPropertyRange>
    <InverseObjectProperties>
            <ObjectProperty IRI="#utilizes"/>
            <ObjectProperty IRI="#dataAnalyticsTask"/>
      </InverseObjectProperties>
```

B.1.1.14. <u>Sampling Operation</u>

Formally, sampling operations are defined as instances of the class `SamplingOperation` in OWL. The definitions of the subclasses of this class are listed below.

```
    <Declaration>
```

```
            <Class IRI="#RandomSampling"/>
        </Declaration>
    <Declaration>
            <Class IRI="#StratifiedRandomSampling"/>
        </Declaration>
   <Declaration>
            <Class IRI="#ClusteringSampling"/>
        </Declaration>
   <Declaration>
            <Class IRI="#AdaptiveSampling"/>
        </Declaration>
   <AnnotationAssertion>
            <AnnotationProperty abbreviatedIRI="rdfs:label"/>
            <IRI>#RandomSampling</IRI>
            <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">Random Sampling</Literal>
        </AnnotationAssertion>
   <AnnotationAssertion>
            <AnnotationProperty abbreviatedIRI="rdfs:label"/>
            <IRI>#StratifiedRandomSampling</IRI>
            <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">Stratified Random Sampling</Literal>
        </AnnotationAssertion>
   <AnnotationAssertion>
            <AnnotationProperty abbreviatedIRI="rdfs:label"/>
            <IRI>#ClusteringSampling</IRI>
            <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">Clustering Sampling</Literal>
        </AnnotationAssertion>
   <AnnotationAssertion>
            <AnnotationProperty abbreviatedIRI="rdfs:label"/>
            <IRI>#AdaptiveSampling</IRI>
            <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">Adaptive Sampling</Literal>
        </AnnotationAssertion>
   <SubClassOf>
            <Class IRI="#RandomSampling"/>
            <Class IRI="#SamplingOperation"/>
```

```
        </SubClassOf>
    <SubClassOf>
            <Class IRI="#StratifiedRandomSampling"/>
            <Class IRI="#SamplingOperation"/>
        </SubClassOf>
    <SubClassOf>
            <Class IRI="#ClusteringSampling"/>
            <Class IRI="#SamplingOperation"/>
        </SubClassOf>
    <SubClassOf>
            <Class IRI="#AdaptiveSampling"/>
            <Class IRI="#SamplingOperation"/>
        </SubClassOf>
```

B.1.1.15. ***Filter Operation***

Formally, filter operations are defined as instances of the class `FilterOperation` in OWL. The definition of this class was presented above in the `Method` definition. Below we present the formal definition of the attribute `condition` of this class, which is formally defined as a data property in OWL.

```
    <Declaration>
            <DataProperty IRI="#condition"/>
        </Declaration>
<DataPropertyDomain>
            <DataProperty IRI="#condition"/>
            <Class IRI="#FilterOperation"/>
        </DataPropertyDomain>
    <DataPropertyRange>
            <DataProperty IRI="#condition"/>
            <Datatype abbreviatedIRI="xsd:string"/>
        </DataPropertyRange>
```

B.1.1.16. ***Join Operation***

Formally, join operations are defined as instances of the class `JoinOperation` in OWL. The definition of this class was presented above in the `Method` definition. Below we present the formal definition of the attribute “`condition`” of this class, which is formally defined as a data property in OWL, as well as its subclasses.

```
<DataPropertyDomain>
          <DataProperty IRI="#condition"/>
          <Class IRI="#JoinOperation"/>
   </DataPropertyDomain>
   <Declaration>
          <Class IRI="#InnerJoin"/>
      </Declaration>
    <Declaration>
          <Class IRI="#FullJoin"/>
      </Declaration>
   <Declaration>
          <Class IRI="#LeftOuter"/>
      </Declaration>
   <Declaration>
          <Class IRI="#RightOuter"/>
      </Declaration>
   <AnnotationAssertion>
          <AnnotationProperty abbreviatedIRI="rdfs:label"/>
          <IRI>#InnerJoin</IRI>
          <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">Inner Join</Literal>
      </AnnotationAssertion>
   <AnnotationAssertion>
          <AnnotationProperty abbreviatedIRI="rdfs:label"/>
          <IRI>#FullJoin</IRI>
          <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">Full Join</Literal>
      </AnnotationAssertion>
   <AnnotationAssertion>
          <AnnotationProperty abbreviatedIRI="rdfs:label"/>
          <IRI>#LeftOuter</IRI>
          <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">Left Outer</Literal>
      </AnnotationAssertion>
   <AnnotationAssertion>
          <AnnotationProperty abbreviatedIRI="rdfs:label"/>
```

```
            <IRI>#RightOuter</IRI>
            <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-ns#PlainLiteral">Right Outer</Literal>
        </AnnotationAssertion>
    <SubClassOf>
            <Class IRI="#InnerJoin"/>
            <Class IRI="#JoinOperation"/>
        </SubClassOf>
    <SubClassOf>
            <Class IRI="#FullJoin"/>
            <Class IRI="#JoinOperation"/>
        </SubClassOf>
    <SubClassOf>
            <Class IRI="#LeftOuter"/>
            <Class IRI="#JoinOperation"/>
        </SubClassOf>
    <SubClassOf>
            <Class IRI="#RightOuter"/>
            <Class IRI="#JoinOperation"/>
        </SubClassOf>
```

B.1.1.17. Project Operation

Formally, project operations are defined as instances of the class `ProjectOperation` in OWL. The definition of this class was presented above in the `Method` definition. Below we present the formal definition of the attribute `projectRule` of this class, which is formally defined as a data property in OWL.

```
    <Declaration>
            <DataProperty IRI="#projectRule"/>
    </Declaration>
<DataPropertyDomain>
            <DataProperty IRI="#projectRule"/>
            <Class IRI="#ProjectOperation"/>
        </DataPropertyDomain>
    <DataPropertyRange>
            <DataProperty IRI="#projectRule"/>
            <Datatype abbreviatedIRI="xsd:string"/>
        </DataPropertyRange>
```

B.1.1.18. Data Cleaning Operation

Formally, data cleaning operations are defined as instances of the class `DataCleaningOperation` in OWL. The definition of this class was presented above in the `Method` definition. Below we present the formal definition of the attribute `cleaningRule` of this class, which is formally defined as a data property in OWL.

```
   <Declaration>
           <DataProperty IRI="#cleaningRule"/>
   </Declaration>
<DataPropertyDomain>
           <DataProperty IRI="#cleaningRule"/>
           <Class IRI="#DataCleaningOperation"/>
       </DataPropertyDomain>
   <DataPropertyRange>
           <DataProperty IRI="#cleaningRule"/>
           <Datatype abbreviatedIRI="xsd:string"/>
       </DataPropertyRange>
```

B.1.1.19. Statistical Analysis Algorithm

Formally, statistical analysis algorithms class are defined as subclasses of the general class `StatisticalAnalysisAlgorithm` in OWL. The definition of this class is listed below.

```
   <Declaration>
           <Class IRI="#StatisticalAnalysisAlgorithm"/>
   </Declaration>
   <SubClassOf>
           <Class IRI="#StatisticalAnalysisAlgorithm"/>
           <Class IRI="#Algorithm"/>
       </SubClassOf>
```

B.1.1.20. Linear Regression

Formally, Linear Regression models of statistical analysis are defined as instances of the class `LinearRegression` in OWL. The definition of this class is listed below.

```
   <Declaration>
           <Class IRI="#LinearRegression"/>
       </Declaration>
   <SubClassOf>
```

```
            <Class IRI="#LinearRegression"/>
            <Class IRI="#StatisticalAnalysisAlgorithm"/>
        </SubClassOf>
```

B.1.1.21. <u>ANOVA</u>

Formally, the Analysis of variance (ANOVA), collection of models, of statistical analysis is defined as an OWL class with four subclasses. The Analysis of variance (ANOVA), collection of models has the following four subclasses: multiway ANOVA, one-way ANOVA, repeated measure ANOVA and two-way ANOVA. The definitions of these classes are listed below

```
    <Declaration>
            <Class IRI="#ANOVA"/>
    </Declaration>
    <SubClassOf>
            <Class IRI="#ANOVA"/>
            <Class IRI="#StatisticalAnalysisAlgorithm"/>
        </SubClassOf>
    <SubClassOf>
            <Class IRI="#MultiwayANOVA"/>
            <Class IRI="#ANOVA"/>
    </SubClassOf>
        <SubClassOf>
            <Class IRI="#OneWayANOVA"/>
            <Class IRI="#ANOVA"/>
        </SubClassOf>
        <SubClassOf>
            <Class IRI="#RepeatedMeasureANOVA"/>
            <Class IRI="#ANOVA"/>
        </SubClassOf>
        <SubClassOf>
            <Class IRI="#TwoWayANOVA"/>
            <Class IRI="#ANOVA"/>
        </SubClassOf>
```

B.1.1.22. <u>Breusch-Pagan Test</u>

Formally, the Breusch-Pagan Test model of statistical analysis algorithm is defined as a class in OWL. The definition of this class is listed below.

```
    <Declaration>
            <Class IRI="#BreuschPaganTest"/>
    </Declaration>
    <SubClassOf>
            <Class IRI="# BreuschPaganTest"/>
            <Class IRI="#StatisticalAnalysisAlgorithm"/>
```

```
</SubClassOf>
```

B.1.1.23. F-Test

Formally, the F-test model of statistical analysis task is defined as a class in OWL. The definition of this class is listed below.

```
<Declaration>
   <Class IRI="#FTest"/>
</Declaration>
<SubClassOf>
        <Class IRI="#FTest"/>
        <Class IRI="#StatisticalAnalysisAlgorithm"/>
</SubClassOf>
```

B.1.1.24. Fischer’s Exact Test

Formally, the Fisher’s Exact Test model of statistical analysis task is defined as a class in OWL. The definition of this class is listed below.

```
<Declaration>
   <Class IRI="#FishersExactTest"/>
</Declaration>
<SubClassOf>
        <Class IRI="#FishersExactTest"/>
        <Class IRI="#StatisticalAlgorithm"/>
</SubClassOf>
```

B.1.1.25. Data Mining Algorithm

Formally, data mining algorithms are defined as instances of the class `DataMiningAlgorithm` in OWL. The definition of this class and its subclasses are listed below.

```
<Declaration>
        <Class IRI="#DataMiningAlgorithm"/>
    </Declaration>
<Declaration>
        <Class IRI="#SupervisedDataMiningAlgorithm"/>
    </Declaration>
 <Declaration>
        <Class IRI="#UnsupervisedDataMiningAlgorithm"/>
    </Declaration>
<AnnotationAssertion>
```

```
          <AnnotationProperty abbreviatedIRI="rdfs:label"/>
          <IRI>#DataMiningAlgorithm</IRI>
          <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">Data Mining Algorithm</Literal>
      </AnnotationAssertion>
   <AnnotationAssertion>
          <AnnotationProperty abbreviatedIRI="rdfs:label"/>
          <IRI>#SupervisedDataMiningAlgorithm</IRI>
          <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">Supervised Data Mining Algorithm</Literal>
      </AnnotationAssertion>
   <AnnotationAssertion>
          <AnnotationProperty abbreviatedIRI="rdfs:label"/>
          <IRI>#UnsupervisedDataMiningAlgorithm</IRI>
          <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">Unsupervised Data Mining Algorithm</Literal>
      </AnnotationAssertion>
   <SubClassOf>
          <Class IRI="#DataMiningAlgorithm"/>
          <Class IRI="#Algorithm"/>
      </SubClassOf>
   <SubClassOf>
          <Class IRI="#Supervised"/>
          <Class IRI="#DataMiningAlgorithm"/>
      </SubClassOf>
   <SubClassOf>
          <Class IRI="#Unsupervised"/>
          <Class IRI="#DataMiningAlgorithm"/>
   </SubClassOf>
```

Below we present the formal declaration of the configuration options that are common for each data mining algorithm: `debug` and `doNotCheckCapabilities` and are declared as data properties of the `DataMiningAlgorithm` class in OWL.

```
   <Declaration>
          <DataProperty IRI="#debug"/>
      </Declaration>
   <SubDataPropertyOf>
          <DataProperty IRI="#debug"/>
```

```
        <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
<DataPropertyDomain>
        <DataProperty IRI="#debug"/>
        <Class IRI="#DataMiningAlgorithm"/>
    </DataPropertyDomain>
<DataPropertyRange>
        <DataProperty IRI="#debug"/>
        <Datatype abbreviatedIRI="xsd:boolean"/>
    </DataPropertyRange>
<Declaration>
        <DataProperty IRI="#doNotCheckCapabilities"/>
    </Declaration>
<SubDataPropertyOf>
        <DataProperty IRI="#doNotCheckCapabilities"/>
        <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
<DataPropertyDomain>
        <DataProperty IRI="#doNotCheckCapabilities"/>
        <Class IRI="#DataMiningAlgorithm"/>
    </DataPropertyDomain>
<DataPropertyRange>
        <DataProperty IRI="#doNotCheckCapabilities"/>
        <Datatype abbreviatedIRI="xsd:boolean"/>
    </DataPropertyRange>
```

B.1.1.26. Naïve Bayes

Formally, Naïve Bayes algorithm executions are defined as instances of the class `NaiveBayes` in OWL. The definition of this class and its data properties (configuration options) are listed below.

```
<Declaration>
        <Class IRI="#NaiveBayes"/>
    </Declaration>
<SubClassOf>
        <Class IRI="#NaiveBayes"/>
        <Class IRI="#Supervised"/>
    </SubClassOf>
<Declaration>
```

```
        <DataProperty IRI="#useKernelEstimator"/>
    </Declaration>
<SubDataPropertyOf>
        <DataProperty IRI="#useKernelEstimator"/>
        <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
<DataPropertyDomain>
        <DataProperty IRI="#useKernelEstimator"/>
        <Class IRI="#NaiveBayes"/>
    </DataPropertyDomain>
<DataPropertyRange>
        <DataProperty IRI="#useKernelEstimator"/>
        <Datatype abbreviatedIRI="xsd:boolean"/>
    </DataPropertyRange>
<Declaration>
        <DataProperty IRI="#numDecimalPlaces"/>
    </Declaration>
<SubDataPropertyOf>
        <DataProperty IRI="#numDecimalPlaces"/>
        <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
<DataPropertyDomain>
        <DataProperty IRI="#numDecimalPlaces"/>
        <Class IRI="#NaiveBayes"/>
    </DataPropertyDomain>
<DataPropertyRange>
        <DataProperty IRI="#numDecimalPlaces"/>
        <Datatype abbreviatedIRI="xsd:positiveInteger"/>
    </DataPropertyRange>
<Declaration>
        <DataProperty IRI="#batchSize"/>
    </Declaration>
<SubDataPropertyOf>
        <DataProperty IRI="#batchSize"/>
        <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
<DataPropertyDomain>
        <DataProperty IRI="#batchSize"/>
```

```
        <Class IRI="#NaiveBayes"/>
    </DataPropertyDomain>
<DataPropertyRange>
        <DataProperty IRI="#batchSize"/>
        <Datatype abbreviatedIRI="xsd:positiveInteger"/>
    </DataPropertyRange>
<Declaration>
        <DataProperty IRI="#displayModelInOldFormat"/>
    </Declaration>
<SubDataPropertyOf>
        <DataProperty IRI="#displayModelInOldFormat"/>
        <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
<DataPropertyDomain>
        <DataProperty IRI="#displayModelInOldFormat"/>
        <Class IRI="#NaiveBayes"/>
    </DataPropertyDomain>
<DataPropertyRange>
        <DataProperty IRI="#displayModelInOldFormat"/>
        <Datatype abbreviatedIRI="xsd:boolean"/>
    </DataPropertyRange>
<Declaration>
        <DataProperty IRI="#useSupervisedDiscretization"/>
    </Declaration>
<SubDataPropertyOf>
        <DataProperty IRI="#useSupervisedDiscretization"/>
        <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
<DataPropertyDomain>
        <DataProperty IRI="#useSupervisedDiscretization"/>
        <Class IRI="#NaiveBayes"/>
    </DataPropertyDomain>
<DataPropertyRange>
        <DataProperty IRI="#useSupervisedDiscretization"/>
        <Datatype abbreviatedIRI="xsd:boolean"/>
</DataPropertyRange>
```

B.1.1.27. Gaussian Processes

Formally, GaussianProcesses algorithm executions are defined as instances of the class GaussianProcesses in OWL. The definition of this class and its data properties (configuration options) are listed below.

```
 <Declaration>
     <Class IRI="#GaussianProcesses"/>
   </Declaration>
   <SubClassOf>
     <Class IRI="#GaussianProcesses"/>
     <Class IRI="#Supervised"/>
   </SubClassOf>
   <Declaration>
     <DataProperty IRI="#seed"/>
   </Declaration>
   <SubDataPropertyOf>
     <DataProperty IRI="#seed"/>
     <DataProperty IRI="#algorithmConfigurationOptions"/>
   </SubDataPropertyOf>
   <DataPropertyDomain>
     <DataProperty IRI="#seed"/>
     <Class IRI="#GaussianProcesses"/>
   </DataPropertyDomain>
   <DataPropertyRange>
     <DataProperty IRI="#seed"/>
     <Datatype abbreviatedIRI="xsd:integer"/>
   </DataPropertyRange>
   <Declaration>
     <DataProperty IRI="#numDecimalPlaces"/>
   </Declaration>
   <SubDataPropertyOf>
     <DataProperty IRI="#numDecimalPlaces"/>
     <DataProperty IRI="#algorithmConfigurationOptions"/>
   </SubDataPropertyOf>
   <DataPropertyDomain>
     <DataProperty IRI="#numDecimalPlaces"/>
     <Class IRI="#GaussianProcesses"/>
   </DataPropertyDomain>
```

```
    <DataPropertyRange>
      <DataProperty IRI="#numDecimalPlaces"/>
      <Datatype abbreviatedIRI="xsd:positiveInteger"/>
    </DataPropertyRange>
    <Declaration>
      <DataProperty IRI="#batchSize"/>
    </Declaration>
    <SubDataPropertyOf>
      <DataProperty IRI="#batchSize"/>
      <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <DataPropertyDomain>
      <DataProperty IRI="#batchSize"/>
      <Class IRI="#GaussianProcesses"/>
    </DataPropertyDomain>
    <DataPropertyRange>
      <DataProperty IRI="#batchSize"/>
      <Datatype abbreviatedIRI="xsd:positiveInteger"/>
    </DataPropertyRange>
    <Declaration>
      <DataProperty IRI="#kernel"/>
    </Declaration>
    <SubDataPropertyOf>
      <DataProperty IRI="#kernel"/>
      <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <DataPropertyDomain>
      <DataProperty IRI="#kernel"/>
      <Class IRI="#GaussianProcesses"/>
    </DataPropertyDomain>
    <DataPropertyRange>
      <DataProperty IRI="#kernel"/>
      <Datatype abbreviatedIRI="xsd:string"/>
    </DataPropertyRange>
    <Declaration>
      <DataProperty IRI="#filterType"/>
    </Declaration>
    <SubDataPropertyOf>
```

```
    <DataProperty IRI="#filterType"/>
    <DataProperty IRI="#algorithmConfigurationOptions"/>
  </SubDataPropertyOf>
  <DataPropertyDomain>
    <DataProperty IRI="#filterType"/>
    <Class IRI="#GaussianProcesses"/>
  </DataPropertyDomain>
  <DataPropertyRange>
    <DataProperty IRI="#filterType"/>
    <Datatype abbreviatedIRI="xsd:positiveInteger"/>
  </DataPropertyRange>
  <Declaration>
    <DataProperty IRI="#noise"/>
  </Declaration>
  <SubDataPropertyOf>
    <DataProperty IRI="#noise"/>
    <DataProperty IRI="#algorithmConfigurationOptions"/>
  </SubDataPropertyOf>
  <DataPropertyDomain>
    <DataProperty IRI="#noise"/>
    <Class IRI="#GaussianProcesses"/>
  </DataPropertyDomain>
  <DataPropertyRange>
    <DataProperty IRI="#noise"/>
    <Datatype abbreviatedIRI="xsd:float"/>
 </DataPropertyRange>
```

B.1.1.28. <u>Linear Regression</u>

Formally, Linear Regression algorithm executions are defined as instances of the class `LinearRegression` in OWL. The definition of this class and its data properties (configuration options) are listed below.

```
  <Declaration>
    <Class IRI="#LinearRegression"/>
  </Declaration>
  <SubClassOf>
    <Class IRI="#LinearRegression"/>
    <Class IRI="#SupervisedDataMiningAlgorithm"/>
```

```
    </SubClassOf>
    <Declaration>
      <DataProperty IRI="#minimal"/>
    </Declaration>
    <SubDataPropertyOf>
      <DataProperty IRI="#minimal"/>
      <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <DataPropertyDomain>
      <DataProperty IRI="#minimal"/>
      <Class IRI="#LinearRegression"/>
    </DataPropertyDomain>
    <DataPropertyRange>
      <DataProperty IRI="#minimal"/>
      <Datatype abbreviatedIRI="xsd:boolean"/>
    </DataPropertyRange>
    <Declaration>
      <DataProperty IRI="#numDecimalPlaces"/>
    </Declaration>
    <SubDataPropertyOf>
      <DataProperty IRI="#numDecimalPlaces"/>
      <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <DataPropertyDomain>
      <DataProperty IRI="#numDecimalPlaces"/>
      <Class IRI="#LinearRegression"/>
    </DataPropertyDomain>
    <DataPropertyRange>
      <DataProperty IRI="#numDecimalPlaces"/>
      <Datatype abbreviatedIRI="xsd:positiveInteger"/>
    </DataPropertyRange>
    <Declaration>
      <DataProperty IRI="#batchSize"/>
    </Declaration>
    <SubDataPropertyOf>
      <DataProperty IRI="#batchSize"/>
      <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
```

```
    <DataPropertyDomain>
      <DataProperty IRI="#batchSize"/>
      <Class IRI="#LinearRegression"/>
    </DataPropertyDomain>
    <DataPropertyRange>
      <DataProperty IRI="#batchSize"/>
      <Datatype abbreviatedIRI="xsd:integer"/>
    </DataPropertyRange>
    <Declaration>
      <DataProperty IRI="#ridge"/>
    </Declaration>
    <SubDataPropertyOf>
      <DataProperty IRI="#ridge"/>
      <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <DataPropertyDomain>
      <DataProperty IRI="#ridge"/>
      <Class IRI="#LinearRegression"/>
    </DataPropertyDomain>
    <DataPropertyRange>
      <DataProperty IRI="#ridge"/>
      <Datatype abbreviatedIRI="xsd:string"/>
    </DataPropertyRange>
    <Declaration>
      <DataProperty IRI="#attributeSelectionMethod"/>
    </Declaration>
    <SubDataPropertyOf>
      <DataProperty IRI="#attributeSelectionMethod"/>
      <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <DataPropertyDomain>
      <DataProperty IRI="#attributeSelectionMethod"/>
      <Class IRI="#LinearRegression"/>
    </DataPropertyDomain>
    <DataPropertyRange>
      <DataProperty IRI="#attributeSelectionMethod"/>
      <Datatype abbreviatedIRI="xsd:positiveInteger"/>
    </DataPropertyRange>
```

```
<Declaration>
  <DataProperty IRI="#outputAdditionalStats"/>
</Declaration>
<SubDataPropertyOf>
  <DataProperty IRI="#outputAdditionalStats"/>
  <DataProperty IRI="#algorithmConfigurationOptions"/>
</SubDataPropertyOf>
<DataPropertyDomain>
  <DataProperty IRI="#outputAdditionalStats"/>
  <Class IRI="#LinearRegression"/>
</DataPropertyDomain>
<DataPropertyRange>
  <DataProperty IRI="#outputAdditionalStats"/>
  <Datatype abbreviatedIRI="xsd:boolean"/>
</DataPropertyRange>
<Declaration>
  <DataProperty IRI="#estimateColinearAttributes"/>
</Declaration>
<SubDataPropertyOf>
  <DataProperty IRI="#estimateColinearAttributes"/>
  <DataProperty IRI="#algorithmConfigurationOptions"/>
</SubDataPropertyOf>
<DataPropertyDomain>
  <DataProperty IRI="#estimateColinearAttributes"/>
  <Class IRI="#LinearRegression"/>
</DataPropertyDomain>
<DataPropertyRange>
  <DataProperty IRI="#estimateColinearAttributes"/>
  <Datatype abbreviatedIRI="xsd:boolean"/>
</DataPropertyRange>
```

B.1.1.29. Multinomial Logistic Regression

Formally, Multinomial Logistic Regression algorithm executions are defined as instances of the class `MultinomialLogisticRegression` in OWL. The definition of this class and its data properties (configuration options) are listed below.

```
<Declaration>
<Class IRI="#MultinomialLogisticRegression"/>
```

```
    </Declaration>
    <SubClassOf>
      <Class IRI="#MultinomialLogisticRegression"/>
      <Class IRI="#SupervisedDataMiningAlgorithm"/>
    </SubClassOf>
    <Declaration>
      <DataProperty IRI="#numDecimalPlaces"/>
    </Declaration>
    <SubDataPropertyOf>
      <DataProperty IRI="#numDecimalPlaces"/>
      <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <DataPropertyDomain>
      <DataProperty IRI="#numDecimalPlaces"/>
      <Class IRI="#MultinomialLogisticRegression"/>
    </DataPropertyDomain>
    <DataPropertyRange>
      <DataProperty IRI="#numDecimalPlaces"/>
      <Datatype abbreviatedIRI="xsd:positiveInteger"/>
    </DataPropertyRange>
    <Declaration>
      <DataProperty IRI="#batchSize"/>
    </Declaration>
    <SubDataPropertyOf>
      <DataProperty IRI="#batchSize"/>
      <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <DataPropertyDomain>
      <DataProperty IRI="#batchSize"/>
      <Class IRI="#MultinomialLogisticRegression"/>
    </DataPropertyDomain>
    <DataPropertyRange>
      <DataProperty IRI="#batchSize"/>
      <Datatype abbreviatedIRI="xsd:integer"/>
    </DataPropertyRange>
    <Declaration>
      <DataProperty IRI="#ridge"/>
    </Declaration>
```

```
<SubDataPropertyOf>
  <DataProperty IRI="#ridge"/>
  <DataProperty IRI="#algorithmConfigurationOptions"/>
</SubDataPropertyOf>
<DataPropertyDomain>
  <DataProperty IRI="#ridge"/>
  <Class IRI="#MultinomialLogisticRegression"/>
</DataPropertyDomain>
<DataPropertyRange>
  <DataProperty IRI="#ridge"/>
  <Datatype abbreviatedIRI="xsd:string"/>
</DataPropertyRange>
<Declaration>
  <DataProperty IRI="#useConjugateGradientDescent"/>
</Declaration>
<SubDataPropertyOf>
  <DataProperty IRI="#useConjugateGradientDescent"/>
  <DataProperty IRI="#algorithmConfigurationOptions"/>
</SubDataPropertyOf>
<DataPropertyDomain>
  <DataProperty IRI="#useConjugateGradientDescent"/>
  <Class IRI="#MultinomialLogisticRegression"/>
</DataPropertyDomain>
<DataPropertyRange>
  <DataProperty IRI="#useConjugateGradientDescent"/>
  <Datatype abbreviatedIRI="xsd:boolean"/>
</DataPropertyRange>
<Declaration>
  <DataProperty IRI="#maxIts"/>
</Declaration>
<SubDataPropertyOf>
  <DataProperty IRI="#maxIts"/>
  <DataProperty IRI="#algorithmConfigurationOptions"/>
</SubDataPropertyOf>
<DataPropertyDomain>
  <DataProperty IRI="#maxIts"/>
  <Class IRI="#MultinomialLogisticRegression"/>
</DataPropertyDomain>
```

```
    <DataPropertyRange>
      <DataProperty IRI="#maxIts"/>
      <Datatype abbreviatedIRI="xsd:integer"/>
    </DataPropertyRange>
```

B.1.1.30. K-nearest neighbours (IBk)

Formally, K-nearest neighbours algorithm executions are defined as instances of the class `IBk` in OWL. The definition of this class and its data properties (configuration options) are listed below.

```
        <Declaration>
      <Class IRI="#IBk"/>
    </Declaration>
    <SubClassOf>
      <Class IRI="#IBk"/>
      <Class IRI="#SupervisedDataMiningAlgorithm"/>
    </SubClassOf>
    <Declaration>
      <DataProperty IRI="#numDecimalPlaces"/>
    </Declaration>
    <SubDataPropertyOf>
      <DataProperty IRI="#numDecimalPlaces"/>
      <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <DataPropertyDomain>
      <DataProperty IRI="#numDecimalPlaces"/>
      <Class IRI="#IBk"/>
    </DataPropertyDomain>
    <DataPropertyRange>
      <DataProperty IRI="#numDecimalPlaces"/>
      <Datatype abbreviatedIRI="xsd:positiveInteger"/>
    </DataPropertyRange>
    <Declaration>
      <DataProperty IRI="#batchSize"/>
    </Declaration>
    <SubDataPropertyOf>
      <DataProperty IRI="#batchSize"/>
      <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
```

```
    <DataPropertyDomain>
      <DataProperty IRI="#batchSize"/>
      <Class IRI="#IBk"/>
    </DataPropertyDomain>
    <DataPropertyRange>
      <DataProperty IRI="#batchSize"/>
      <Datatype abbreviatedIRI="xsd:integer"/>
    </DataPropertyRange>
    <Declaration>
      <DataProperty IRI="#KNN"/>
    </Declaration>
    <SubDataPropertyOf>
      <DataProperty IRI="#KNN"/>
      <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <DataPropertyDomain>
      <DataProperty IRI="#KNN"/>
      <Class IRI="#IBk"/>
    </DataPropertyDomain>
    <DataPropertyRange>
      <DataProperty IRI="#KNN"/>
      <Datatype abbreviatedIRI="xsd:positiveInteger"/>
    </DataPropertyRange>
    <Declaration>
      <DataProperty IRI="#distanceWeighting"/>
    </Declaration>
    <SubDataPropertyOf>
      <DataProperty IRI="#distanceWeighting"/>
      <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <DataPropertyDomain>
      <DataProperty IRI="#distanceWeighting"/>
      <Class IRI="#IBk"/>
    </DataPropertyDomain>
    <DataPropertyRange>
      <DataProperty IRI="#distanceWeighting"/>
      <Datatype abbreviatedIRI="xsd:positiveInteger"/>
    </DataPropertyRange>
```

```
<Declaration>
  <DataProperty IRI="#nearestNeighbourSearchAlgorithm"/>
</Declaration>
<SubDataPropertyOf>
  <DataProperty IRI="#nearestNeighbourSearchAlgorithm"/>
  <DataProperty IRI="#algorithmConfigurationOptions"/>
</SubDataPropertyOf>
<DataPropertyDomain>
  <DataProperty IRI="#nearestNeighbourSearchAlgorithm"/>
  <Class IRI="#IBk"/>
</DataPropertyDomain>
<DataPropertyRange>
  <DataProperty IRI="#nearestNeighbourSearchAlgorithm"/>
  <Datatype abbreviatedIRI="xsd:string"/>
</DataPropertyRange>
<Declaration>
  <DataProperty IRI="#windowSize"/>
</Declaration>
<SubDataPropertyOf>
  <DataProperty IRI="#windowSize"/>
  <DataProperty IRI="#algorithmConfigurationOptions"/>
</SubDataPropertyOf>
<DataPropertyDomain>
  <DataProperty IRI="#windowSize"/>
  <Class IRI="#IBk"/>
</DataPropertyDomain>
<DataPropertyRange>
  <DataProperty IRI="#windowSize"/>
  <Datatype abbreviatedIRI="xsd:positiveInteger"/>
</DataPropertyRange>
<Declaration>
  <DataProperty IRI="#meanSquared"/>
</Declaration>
<SubDataPropertyOf>
  <DataProperty IRI="#meanSquared"/>
  <DataProperty IRI="#algorithmConfigurationOptions"/>
</SubDataPropertyOf>
<DataPropertyDomain>
```

```
    <DataProperty IRI="#meanSquared"/>
    <Class IRI="#IBk"/>
  </DataPropertyDomain>
  <DataPropertyRange>
    <DataProperty IRI="#meanSquared"/>
    <Datatype abbreviatedIRI="xsd:boolean"/>
  </DataPropertyRange>
  <Declaration>
    <DataProperty IRI="#crossValidate"/>
  </Declaration>
  <SubDataPropertyOf>
    <DataProperty IRI="#crossValidate"/>
    <DataProperty IRI="#algorithmConfigurationOptions"/>
  </SubDataPropertyOf>
  <DataPropertyDomain>
    <DataProperty IRI="#crossValidate"/>
    <Class IRI="#IBk"/>
  </DataPropertyDomain>
  <DataPropertyRange>
    <DataProperty IRI="#crossValidate"/>
    <Datatype abbreviatedIRI="xsd:boolean"/>
 </DataPropertyRange>
```

B.1.1.31. <u>Decision Table</u>

Formally, Decision Table algorithm executions are defined as instances of the class `DecisionTable` in OWL. The definition of this class and its data properties (configuration options) are listed below.

```
        <Declaration>
    <Class IRI="#DecisionTable"/>
  </Declaration>
  <SubClassOf>
    <Class IRI="#DecisionTable"/>
    <Class IRI="#SupervisedDataMiningAlgorithm"/>
  </SubClassOf>
  <Declaration>
    <DataProperty IRI="#numDecimalPlaces"/>
  </Declaration>
```

```
    <SubDataPropertyOf>
      <DataProperty IRI="#numDecimalPlaces"/>
      <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <DataPropertyDomain>
      <DataProperty IRI="#numDecimalPlaces"/>
      <Class IRI="#DecisionTable"/>
    </DataPropertyDomain>
    <DataPropertyRange>
      <DataProperty IRI="#numDecimalPlaces"/>
      <Datatype abbreviatedIRI="xsd:positiveInteger"/>
    </DataPropertyRange>
    <Declaration>
      <DataProperty IRI="#batchSize"/>
    </Declaration>
    <SubDataPropertyOf>
      <DataProperty IRI="#batchSize"/>
      <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <DataPropertyDomain>
      <DataProperty IRI="#batchSize"/>
      <Class IRI="#DecisionTable"/>
    </DataPropertyDomain>
    <DataPropertyRange>
      <DataProperty IRI="#batchSize"/>
      <Datatype abbreviatedIRI="xsd:integer"/>
    </DataPropertyRange>
    <Declaration>
      <DataProperty IRI="#evaluationMeasure"/>
    </Declaration>
    <SubDataPropertyOf>
      <DataProperty IRI="#evaluationMeasure"/>
      <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <DataPropertyDomain>
      <DataProperty IRI="#evaluationMeasure"/>
      <Class IRI="#DecisionTable"/>
    </DataPropertyDomain>
```

```
    <DataPropertyRange>
      <DataProperty IRI="#evaluationMeasure"/>
      <Datatype abbreviatedIRI="xsd:postiveInteger"/>
    </DataPropertyRange>
    <Declaration>
      <DataProperty IRI="#search"/>
    </Declaration>
    <SubDataPropertyOf>
      <DataProperty IRI="#search"/>
      <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <DataPropertyDomain>
      <DataProperty IRI="#search"/>
      <Class IRI="#DecisionTable"/>
    </DataPropertyDomain>
    <DataPropertyRange>
      <DataProperty IRI="#search"/>
      <Datatype abbreviatedIRI="xsd:string"/>
    </DataPropertyRange>
    <Declaration>
      <DataProperty IRI="#displayRules"/>
    </Declaration>
    <SubDataPropertyOf>
      <DataProperty IRI="#displayRules"/>
      <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <DataPropertyDomain>
      <DataProperty IRI="#displayRules"/>
      <Class IRI="#DecisionTable"/>
    </DataPropertyDomain>
    <DataPropertyRange>
      <DataProperty IRI="#displayRules"/>
      <Datatype abbreviatedIRI="xsd:boolean"/>
    </DataPropertyRange>
    <Declaration>
      <DataProperty IRI="#useIBk"/>
    </Declaration>
    <SubDataPropertyOf>
```

```
    <DataProperty IRI="#useIBk"/>
    <DataProperty IRI="#algorithmConfigurationOptions"/>
  </SubDataPropertyOf>
  <DataPropertyDomain>
    <DataProperty IRI="#useIBk"/>
    <Class IRI="#DecisionTable"/>
  </DataPropertyDomain>
  <DataPropertyRange>
    <DataProperty IRI="#useIBk"/>
    <Datatype abbreviatedIRI="xsd:boolean"/>
  </DataPropertyRange>
  <Declaration>
    <DataProperty IRI="#crossVal"/>
  </Declaration>
  <SubDataPropertyOf>
    <DataProperty IRI="#crossVal"/>
    <DataProperty IRI="#algorithmConfigurationOptions"/>
  </SubDataPropertyOf>
  <DataPropertyDomain>
    <DataProperty IRI="#crossVal"/>
    <Class IRI="#DecisionTable"/>
  </DataPropertyDomain>
  <DataPropertyRange>
    <DataProperty IRI="#crossVal"/>
    <Datatype abbreviatedIRI="xsd:integer"/>
</DataPropertyRange>
```

B.1.1.32. <u>Zero R</u>

Formally, Zero R algorithm executions are defined as instances of the class `ZeroR` in OWL. The definition of this class and its data properties (configuration options) are listed below.

```
  <Declaration>
    <Class IRI="#ZeroR"/>
  </Declaration>
  <SubClassOf>
    <Class IRI="#ZeroR"/>
    <Class IRI="#SupervisedDataMiningAlgorithm"/>
  </SubClassOf>
```

```
<Declaration>
  <DataProperty IRI="#numDecimalPlaces"/>
</Declaration>
<SubDataPropertyOf>
  <DataProperty IRI="#numDecimalPlaces"/>
  <DataProperty IRI="#algorithmConfigurationOptions"/>
</SubDataPropertyOf>
<DataPropertyDomain>
  <DataProperty IRI="#numDecimalPlaces"/>
  <Class IRI="#ZeroR"/>
</DataPropertyDomain>
<DataPropertyRange>
  <DataProperty IRI="#numDecimalPlaces"/>
  <Datatype abbreviatedIRI="xsd:positiveInteger"/>
</DataPropertyRange>
<Declaration>
  <DataProperty IRI="#batchSize"/>
</Declaration>
<SubDataPropertyOf>
  <DataProperty IRI="#batchSize"/>
  <DataProperty IRI="#algorithmConfigurationOptions"/>
</SubDataPropertyOf>
<DataPropertyDomain>
  <DataProperty IRI="#batchSize"/>
  <Class IRI="#ZeroR"/>
</DataPropertyDomain>
<DataPropertyRange>
  <DataProperty IRI="#batchSize"/>
  <Datatype abbreviatedIRI="xsd:integer"/>
</DataPropertyRange>
```

B.1.1.33. J48

Formally, J48 algorithm executions are defined as instances of the class `J48` in OWL. The definition of this class and its data properties (configuration options) are listed below.

```
  <Declaration>
  <Class IRI="#J48"/>
</Declaration>
```

```
<SubClassOf>
  <Class IRI="#J48"/>
  <Class IRI="#SupervisedDataMiningAlgorithm"/>
</SubClassOf>
<Declaration>
  <DataProperty IRI="#seed"/>
</Declaration>
<SubDataPropertyOf>
  <DataProperty IRI="#seed"/>
  <DataProperty IRI="#algorithmConfigurationOptions"/>
</SubDataPropertyOf>
<DataPropertyDomain>
  <DataProperty IRI="#seed"/>
  <Class IRI="#J48"/>
</DataPropertyDomain>
<DataPropertyRange>
  <DataProperty IRI="#seed"/>
  <Datatype abbreviatedIRI="xsd:integer"/>
</DataPropertyRange>
<Declaration>
  <DataProperty IRI="#unpruned"/>
</Declaration>
<SubDataPropertyOf>
  <DataProperty IRI="#unpruned"/>
  <DataProperty IRI="#algorithmConfigurationOptions"/>
</SubDataPropertyOf>
<DataPropertyDomain>
  <DataProperty IRI="#unpruned"/>
  <Class IRI="#J48"/>
</DataPropertyDomain>
<DataPropertyRange>
  <DataProperty IRI="#unpruned"/>
  <Datatype abbreviatedIRI="xsd:boolean"/>
</DataPropertyRange>
<Declaration>
  <DataProperty IRI="#confidenceFactor"/>
</Declaration>
<SubDataPropertyOf>
```

```
    <DataProperty IRI="#confidenceFactor"/>
    <DataProperty IRI="#algorithmConfigurationOptions"/>
  </SubDataPropertyOf>
  <DataPropertyDomain>
    <DataProperty IRI="#confidenceFactor"/>
    <Class IRI="#J48"/>
  </DataPropertyDomain>
  <DataPropertyRange>
    <DataProperty IRI="#confidenceFactor"/>
    <Datatype abbreviatedIRI="xsd:float"/>
  </DataPropertyRange>
  <Declaration>
    <DataProperty IRI="#numFolds"/>
  </Declaration>
  <SubDataPropertyOf>
    <DataProperty IRI="#numFolds"/>
    <DataProperty IRI="#algorithmConfigurationOptions"/>
  </SubDataPropertyOf>
  <DataPropertyDomain>
    <DataProperty IRI="#numFolds"/>
    <Class IRI="#J48"/>
  </DataPropertyDomain>
  <DataPropertyRange>
    <DataProperty IRI="#numFolds"/>
    <Datatype abbreviatedIRI="xsd:positiveInteger"/>
  </DataPropertyRange>
  <Declaration>
    <DataProperty IRI="#numDecimalPlaces"/>
  </Declaration>
  <SubDataPropertyOf>
    <DataProperty IRI="#numDecimalPlaces"/>
    <DataProperty IRI="#algorithmConfigurationOptions"/>
  </SubDataPropertyOf>
  <DataPropertyDomain>
    <DataProperty IRI="#numDecimalPlaces"/>
    <Class IRI="#J48"/>
  </DataPropertyDomain>
  <DataPropertyRange>
```

```
    <DataProperty IRI="#numDecimalPlaces"/>
    <Datatype abbreviatedIRI="xsd:positiveInteger"/>
  </DataPropertyRange>
  <Declaration>
    <DataProperty IRI="#batchSize"/>
  </Declaration>
  <SubDataPropertyOf>
    <DataProperty IRI="#batchSize"/>
    <DataProperty IRI="#algorithmConfigurationOptions"/>
  </SubDataPropertyOf>
  <DataPropertyDomain>
    <DataProperty IRI="#batchSize"/>
    <Class IRI="#J48"/>
  </DataPropertyDomain>
  <DataPropertyRange>
    <DataProperty IRI="#batchSize"/>
    <Datatype abbreviatedIRI="xsd:integer"/>
  </DataPropertyRange>
  <Declaration>
    <DataProperty IRI="#reducedErrorPruning"/>
  </Declaration>
  <SubDataPropertyOf>
    <DataProperty IRI="#reducedErrorPruning"/>
    <DataProperty IRI="#algorithmConfigurationOptions"/>
  </SubDataPropertyOf>
  <DataPropertyDomain>
    <DataProperty IRI="#reducedErrorPruning"/>
    <Class IRI="#J48"/>
  </DataPropertyDomain>
  <DataPropertyRange>
    <DataProperty IRI="#reducedErrorPruning"/>
    <Datatype abbreviatedIRI="xsd:boolean"/>
  </DataPropertyRange>
  <Declaration>
    <DataProperty IRI="#useLaplace"/>
  </Declaration>
  <SubDataPropertyOf>
    <DataProperty IRI="#useLaplace"/>
```

```
      <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <DataPropertyDomain>
      <DataProperty IRI="#useLaplace"/>
      <Class IRI="#J48"/>
    </DataPropertyDomain>
    <DataPropertyRange>
      <DataProperty IRI="#useLaplace"/>
      <Datatype abbreviatedIRI="xsd:boolean"/>
    </DataPropertyRange>
    <Declaration>
      <DataProperty IRI="#doNotMakeSplitPointActualValue"/>
    </Declaration>
    <SubDataPropertyOf>
      <DataProperty IRI="#doNotMakeSplitPointActualValue"/>
      <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <DataPropertyDomain>
      <DataProperty IRI="#doNotMakeSplitPointActualValue"/>
      <Class IRI="#J48"/>
    </DataPropertyDomain>
    <DataPropertyRange>
      <DataProperty IRI="#doNotMakeSplitPointActualValue"/>
      <Datatype abbreviatedIRI="xsd:boolean"/>
    </DataPropertyRange>
    <Declaration>
      <DataProperty IRI="#subtreeRaising"/>
    </Declaration>
    <SubDataPropertyOf>
      <DataProperty IRI="#subtreeRaising"/>
      <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <DataPropertyDomain>
      <DataProperty IRI="#subtreeRaising"/>
      <Class IRI="#J48"/>
    </DataPropertyDomain>
    <DataPropertyRange>
      <DataProperty IRI="#subtreeRaising"/>
```

```
    <Datatype abbreviatedIRI="xsd:boolean"/>
  </DataPropertyRange>
  <Declaration>
    <DataProperty IRI="#saveInstanceData"/>
  </Declaration>
  <SubDataPropertyOf>
    <DataProperty IRI="#saveInstanceData"/>
    <DataProperty IRI="#algorithmConfigurationOptions"/>
  </SubDataPropertyOf>
  <DataPropertyDomain>
    <DataProperty IRI="#saveInstanceData"/>
    <Class IRI="#J48"/>
  </DataPropertyDomain>
  <DataPropertyRange>
    <DataProperty IRI="#saveInstanceData"/>
    <Datatype abbreviatedIRI="xsd:boolean"/>
  </DataPropertyRange>
  <Declaration>
    <DataProperty IRI="#binarySplits"/>
  </Declaration>
  <SubDataPropertyOf>
    <DataProperty IRI="#binarySplits"/>
    <DataProperty IRI="#algorithmConfigurationOptions"/>
  </SubDataPropertyOf>
  <DataPropertyDomain>
    <DataProperty IRI="#binarySplits"/>
    <Class IRI="#J48"/>
  </DataPropertyDomain>
  <DataPropertyRange>
    <DataProperty IRI="#binarySplits"/>
    <Datatype abbreviatedIRI="xsd:boolean"/>
  </DataPropertyRange>
  <Declaration>
    <DataProperty IRI="#minNumObj"/>
  </Declaration>
  <SubDataPropertyOf>
    <DataProperty IRI="#minNumObj"/>
    <DataProperty IRI="#algorithmConfigurationOptions"/>
```

```
    </SubDataPropertyOf>
    <DataPropertyDomain>
      <DataProperty IRI="#minNumObj"/>
      <Class IRI="#J48"/>
    </DataPropertyDomain>
    <DataPropertyRange>
      <DataProperty IRI="#minNumObj"/>
      <Datatype abbreviatedIRI="xsd:integer"/>
    </DataPropertyRange>
    <Declaration>
      <DataProperty IRI="#useMDLcorrection"/>
    </Declaration>
    <SubDataPropertyOf>
      <DataProperty IRI="#useMDLcorrection"/>
      <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <DataPropertyDomain>
      <DataProperty IRI="#useMDLcorrection"/>
      <Class IRI="#J48"/>
    </DataPropertyDomain>
    <DataPropertyRange>
      <DataProperty IRI="#useMDLcorrection"/>
      <Datatype abbreviatedIRI="xsd:boolean"/>
    </DataPropertyRange>
    <Declaration>
      <DataProperty IRI="#collapseTree"/>
    </Declaration>
    <SubDataPropertyOf>
      <DataProperty IRI="#collapseTree"/>
      <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <DataPropertyDomain>
      <DataProperty IRI="#collapseTree"/>
      <Class IRI="#J48"/>
    </DataPropertyDomain>
    <DataPropertyRange>
      <DataProperty IRI="#collapseTree"/>
      <Datatype abbreviatedIRI="xsd:boolean"/>
  </DataPropertyRange>
```

B.1.1.34. Random Forest

Formally, Random Forest algorithm executions are defined as instances of the class `RandomForest` in OWL. The definition of this class and its data properties (configuration options) are listed below.

```
    <Declaration>
    <Class IRI="#RandomForest"/>
  </Declaration>
  <SubClassOf>
    <Class IRI="#RandomForest"/>
    <Class IRI="#SupervisedDataMiningAlgorithm"/>
  </SubClassOf>
  <Declaration>
    <DataProperty IRI="#seed"/>
  </Declaration>
  <SubDataPropertyOf>
    <DataProperty IRI="#seed"/>
    <DataProperty IRI="#algorithmConfigurationOptions"/>
  </SubDataPropertyOf>
  <DataPropertyDomain>
    <DataProperty IRI="#seed"/>
    <Class IRI="#RandomForest"/>
  </DataPropertyDomain>
  <DataPropertyRange>
    <DataProperty IRI="#seed"/>
    <Datatype abbreviatedIRI="xsd:integer"/>
  </DataPropertyRange>
  <Declaration>
    <DataProperty IRI="#representCopiesUsingWeights"/>
  </Declaration>
  <SubDataPropertyOf>
    <DataProperty IRI="#representCopiesUsingWeights"/>
    <DataProperty IRI="#algorithmConfigurationOptions"/>
  </SubDataPropertyOf>
  <DataPropertyDomain>
    <DataProperty IRI="#representCopiesUsingWeights"/>
    <Class IRI="#RandomForest"/>
  </DataPropertyDomain>
```

```
    <DataPropertyRange>
      <DataProperty IRI="#representCopiesUsingWeights"/>
      <Datatype abbreviatedIRI="xsd:boolean"/>
    </DataPropertyRange>
    <Declaration>
      <DataProperty IRI="#storeOutOfBagPredictions"/>
    </Declaration>
    <SubDataPropertyOf>
      <DataProperty IRI="#storeOutOfBagPredictions"/>
      <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <DataPropertyDomain>
      <DataProperty IRI="#storeOutOfBagPredictions"/>
      <Class IRI="#RandomForest"/>
    </DataPropertyDomain>
    <DataPropertyRange>
      <DataProperty IRI="#storeOutOfBagPredictions"/>
      <Datatype abbreviatedIRI="xsd:boolean"/>
    </DataPropertyRange>
    <Declaration>
      <DataProperty IRI="#numExecutionSlots"/>
    </Declaration>
    <SubDataPropertyOf>
      <DataProperty IRI="#numExecutionSlots"/>
      <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <DataPropertyDomain>
      <DataProperty IRI="#numExecutionSlots"/>
      <Class IRI="#RandomForest"/>
    </DataPropertyDomain>
    <DataPropertyRange>
      <DataProperty IRI="#numExecutionSlots"/>
      <Datatype abbreviatedIRI="xsd:positiveInteger"/>
    </DataPropertyRange>
    <Declaration>
      <DataProperty IRI="#bagSizePercent"/>
    </Declaration>
    <SubDataPropertyOf>
```

```
    <DataProperty IRI="#bagSizePercent"/>
    <DataProperty IRI="#algorithmConfigurationOptions"/>
  </SubDataPropertyOf>
  <DataPropertyDomain>
    <DataProperty IRI="#bagSizePercent"/>
    <Class IRI="#RandomForest"/>
  </DataPropertyDomain>
  <DataPropertyRange>
    <DataProperty IRI="#bagSizePercent"/>
    <Datatype abbreviatedIRI="xsd:positiveInteger"/>
  </DataPropertyRange>
  <Declaration>
    <DataProperty IRI="#batchSize"/>
  </Declaration>
  <SubDataPropertyOf>
    <DataProperty IRI="#batchSize"/>
    <DataProperty IRI="#algorithmConfigurationOptions"/>
  </SubDataPropertyOf>
  <DataPropertyDomain>
    <DataProperty IRI="#batchSize"/>
    <Class IRI="#RandomForest"/>
  </DataPropertyDomain>
  <DataPropertyRange>
    <DataProperty IRI="#batchSize"/>
    <Datatype abbreviatedIRI="xsd:integer"/>
  </DataPropertyRange>
  <Declaration>
    <DataProperty IRI="#printClassifiers"/>
  </Declaration>
  <SubDataPropertyOf>
    <DataProperty IRI="#printClassifiers"/>
    <DataProperty IRI="#algorithmConfigurationOptions"/>
  </SubDataPropertyOf>
  <DataPropertyDomain>
    <DataProperty IRI="#printClassifiers"/>
    <Class IRI="#RandomForest"/>
  </DataPropertyDomain>
  <DataPropertyRange>
```

```
    <DataProperty IRI="#printClassifiers"/>
    <Datatype abbreviatedIRI="xsd:boolean"/>
  </DataPropertyRange>
  <Declaration>
    <DataProperty IRI="#numIterations"/>
  </Declaration>
  <SubDataPropertyOf>
    <DataProperty IRI="#numIterations"/>
    <DataProperty IRI="#algorithmConfigurationOptions"/>
  </SubDataPropertyOf>
  <DataPropertyDomain>
    <DataProperty IRI="#numIterations"/>
    <Class IRI="#RandomForest"/>
  </DataPropertyDomain>
  <DataPropertyRange>
    <DataProperty IRI="#numIterations"/>
    <Datatype abbreviatedIRI="xsd:positiveInteger"/>
  </DataPropertyRange>
  <Declaration>
    <DataProperty IRI="#outputOutOfBagComplexityStatistics"/>
  </Declaration>
  <SubDataPropertyOf>
    <DataProperty IRI="#outputOutOfBagComplexityStatistics"/>
    <DataProperty IRI="#algorithmConfigurationOptions"/>
  </SubDataPropertyOf>
  <DataPropertyDomain>
    <DataProperty IRI="#outputOutOfBagComplexityStatistics"/>
    <Class IRI="#RandomForest"/>
  </DataPropertyDomain>
  <DataPropertyRange>
    <DataProperty IRI="#outputOutOfBagComplexityStatistics"/>
    <Datatype abbreviatedIRI="xsd:boolean"/>
  </DataPropertyRange>
  <Declaration>
    <DataProperty IRI="#classifier"/>
  </Declaration>
  <SubDataPropertyOf>
    <DataProperty IRI="#classifier"/>
```

```
    <DataProperty IRI="#algorithmConfigurationOptions"/>
  </SubDataPropertyOf>
  <DataPropertyDomain>
    <DataProperty IRI="#classifier"/>
    <Class IRI="#RandomForest"/>
  </DataPropertyDomain>
  <DataPropertyRange>
    <DataProperty IRI="#classifier"/>
    <Datatype abbreviatedIRI="xsd:string"/>
  </DataPropertyRange>
  <Declaration>
    <DataProperty IRI="#breakTiesRandomly"/>
  </Declaration>
  <SubDataPropertyOf>
    <DataProperty IRI="#breakTiesRandomly"/>
    <DataProperty IRI="#algorithmConfigurationOptions"/>
  </SubDataPropertyOf>
  <DataPropertyDomain>
    <DataProperty IRI="#breakTiesRandomly"/>
    <Class IRI="#RandomForest"/>
  </DataPropertyDomain>
  <DataPropertyRange>
    <DataProperty IRI="#breakTiesRandomly"/>
    <Datatype abbreviatedIRI="xsd:boolean"/>
  </DataPropertyRange>
  <Declaration>
    <DataProperty IRI="#maxDepth"/>
  </Declaration>
  <SubDataPropertyOf>
    <DataProperty IRI="#maxDepth"/>
    <DataProperty IRI="#algorithmConfigurationOptions"/>
  </SubDataPropertyOf>
  <DataPropertyDomain>
    <DataProperty IRI="#maxDepth"/>
    <Class IRI="#RandomForest"/>
  </DataPropertyDomain>
  <DataPropertyRange>
    <DataProperty IRI="#maxDepth"/>
```

```
      <Datatype abbreviatedIRI="xsd:positiveInteger"/>
    </DataPropertyRange>
    <Declaration>
      <DataProperty IRI="#computeAttributeImportance"/>
    </Declaration>
    <SubDataPropertyOf>
      <DataProperty IRI="#computeAttributeImportance"/>
      <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <DataPropertyDomain>
      <DataProperty IRI="#computeAttributeImportance"/>
      <Class IRI="#RandomForest"/>
    </DataPropertyDomain>
    <DataPropertyRange>
      <DataProperty IRI="#computeAttributeImportance"/>
      <Datatype abbreviatedIRI="xsd:boolean"/>
    </DataPropertyRange>
    <Declaration>
      <DataProperty IRI="#calcOutOfBag"/>
    </Declaration>
    <SubDataPropertyOf>
      <DataProperty IRI="#calcOutOfBag"/>
      <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <DataPropertyDomain>
      <DataProperty IRI="#calcOutOfBag"/>
      <Class IRI="#RandomForest"/>
    </DataPropertyDomain>
    <DataPropertyRange>
      <DataProperty IRI="#calcOutOfBag"/>
      <Datatype abbreviatedIRI="xsd:boolean"/>
    </DataPropertyRange>
    <Declaration>
      <DataProperty IRI="#numFeatures"/>
    </Declaration>
    <SubDataPropertyOf>
      <DataProperty IRI="#numFeatures"/>
      <DataProperty IRI="#algorithmConfigurationOptions"/>
```

```
    </SubDataPropertyOf>
    <DataPropertyDomain>
      <DataProperty IRI="#numFeatures"/>
      <Class IRI="#RandomForest"/>
    </DataPropertyDomain>
    <DataPropertyRange>
      <DataProperty IRI="#numFeatures"/>
      <Datatype abbreviatedIRI="xsd:positiveInteger"/>
   </DataPropertyRange>
```

B.1.1.35. Random Tree

Formally, Random Tree algorithm executions are defined as instances of the class `RandomTree` in OWL. The definition of this class and its data properties (configuration options) are listed below.

```
        <Declaration>
      <Class IRI="#RandomTree"/>
    </Declaration>
    <SubClassOf>
      <Class IRI="#RandomTree"/>
      <Class IRI="#SupervisedDataMiningAlgorithm"/>
    </SubClassOf>
    <Declaration>
      <DataProperty IRI="#seed"/>
    </Declaration>
    <SubDataPropertyOf>
      <DataProperty IRI="#seed"/>
      <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <DataPropertyDomain>
      <DataProperty IRI="#seed"/>
      <Class IRI="#RandomTree"/>
    </DataPropertyDomain>
    <DataPropertyRange>
      <DataProperty IRI="#seed"/>
      <Datatype abbreviatedIRI="xsd:integer"/>
    </DataPropertyRange>
    <Declaration>
      <DataProperty IRI="#allowUnclassifiedInstances"/>
```

```
    </Declaration>
    <SubDataPropertyOf>
      <DataProperty IRI="#allowUnclassifiedInstances"/>
      <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <DataPropertyDomain>
      <DataProperty IRI="#allowUnclassifiedInstances"/>
      <Class IRI="#RandomTree"/>
    </DataPropertyDomain>
    <DataPropertyRange>
      <DataProperty IRI="#allowUnclassifiedInstances"/>
      <Datatype abbreviatedIRI="xsd:boolean"/>
    </DataPropertyRange>
    <Declaration>
      <DataProperty IRI="#minNums"/>
    </Declaration>
    <SubDataPropertyOf>
      <DataProperty IRI="#minNums"/>
      <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <DataPropertyDomain>
      <DataProperty IRI="#minNums"/>
      <Class IRI="#RandomTree"/>
    </DataPropertyDomain>
    <DataPropertyRange>
      <DataProperty IRI="#minNums"/>
      <Datatype abbreviatedIRI="xsd:float"/>
    </DataPropertyRange>
    <Declaration>
      <DataProperty IRI="#numFolds"/>
    </Declaration>
    <SubDataPropertyOf>
      <DataProperty IRI="#numFolds"/>
      <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <DataPropertyDomain>
      <DataProperty IRI="#numFolds"/>
      <Class IRI="#RandomTree"/>
```

```
    </DataPropertyDomain>
    <DataPropertyRange>
      <DataProperty IRI="#numFolds"/>
      <Datatype abbreviatedIRI="xsd:positiveInteger"/>
    </DataPropertyRange>
    <Declaration>
      <DataProperty IRI="#numDecimalPlaces"/>
    </Declaration>
    <SubDataPropertyOf>
      <DataProperty IRI="#numDecimalPlaces"/>
      <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <DataPropertyDomain>
      <DataProperty IRI="#numDecimalPlaces"/>
      <Class IRI="#RandomTree"/>
    </DataPropertyDomain>
    <DataPropertyRange>
      <DataProperty IRI="#numDecimalPlaces"/>
      <Datatype abbreviatedIRI="xsd:positiveInteger"/>
    </DataPropertyRange>
    <Declaration>
      <DataProperty IRI="#batchSize"/>
    </Declaration>
    <SubDataPropertyOf>
      <DataProperty IRI="#batchSize"/>
      <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <DataPropertyDomain>
      <DataProperty IRI="#batchSize"/>
      <Class IRI="#RandomTree"/>
    </DataPropertyDomain>
    <DataPropertyRange>
      <DataProperty IRI="#batchSize"/>
      <Datatype abbreviatedIRI="xsd:integer"/>
    </DataPropertyRange>
    <Declaration>
      <DataProperty IRI="#breakTiesRandomly"/>
    </Declaration>
```

```
    <SubDataPropertyOf>
      <DataProperty IRI="#breakTiesRandomly"/>
      <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <DataPropertyDomain>
      <DataProperty IRI="#breakTiesRandomly"/>
      <Class IRI="#RandomTree"/>
    </DataPropertyDomain>
    <DataPropertyRange>
      <DataProperty IRI="#breakTiesRandomly"/>
      <Datatype abbreviatedIRI="xsd:boolean"/>
    </DataPropertyRange>
    <Declaration>
      <DataProperty IRI="#maxDepth"/>
    </Declaration>
    <SubDataPropertyOf>
      <DataProperty IRI="#maxDepth"/>
      <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <DataPropertyDomain>
      <DataProperty IRI="#maxDepth"/>
      <Class IRI="#RandomTree"/>
    </DataPropertyDomain>
    <DataPropertyRange>
      <DataProperty IRI="#maxDepth"/>
      <Datatype abbreviatedIRI="xsd:positiveInteger"/>
    </DataPropertyRange>
    <Declaration>
      <DataProperty IRI="#minVarianceProp"/>
    </Declaration>
    <SubDataPropertyOf>
      <DataProperty IRI="#minVarianceProp"/>
      <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <DataPropertyDomain>
      <DataProperty IRI="#minVarianceProp"/>
      <Class IRI="#RandomTree"/>
    </DataPropertyDomain>
```

```
<DataPropertyRange>
  <DataProperty IRI="#minVarianceProp"/>
  <Datatype abbreviatedIRI="xsd:float"/>
</DataPropertyRange>
<Declaration>
  <DataProperty IRI="#KValue"/>
</Declaration>
<SubDataPropertyOf>
  <DataProperty IRI="#KValue"/>
  <DataProperty IRI="#algorithmConfigurationOptions"/>
</SubDataPropertyOf>
<DataPropertyDomain>
  <DataProperty IRI="#KValue"/>
  <Class IRI="#RandomTree"/>
</DataPropertyDomain>
<DataPropertyRange>
  <DataProperty IRI="#KValue"/>
  <Datatype abbreviatedIRI="xsd:integer"/>
</DataPropertyRange>
```

B.1.1.36. <u>Canopy</u>

Formally, Canopy algorithm executions are defined as instances of the class `Canopy` in OWL. The definition of this class and its data properties (configuration options) are listed below.

```
<Declaration>
  <Class IRI="#Canopy"/>
</Declaration>
<SubClassOf>
  <Class IRI="#Canopy"/>
  <Class IRI="#UnsupervisedDataMiningAlgorithm"/>
</SubClassOf>
<Declaration>
  <DataProperty IRI="#seed"/>
</Declaration>
<SubDataPropertyOf>
  <DataProperty IRI="#seed"/>
  <DataProperty IRI="#algorithmConfigurationOptions"/>
</SubDataPropertyOf>
```

```
<DataPropertyDomain>
  <DataProperty IRI="#seed"/>
  <Class IRI="#Canopy"/>
</DataPropertyDomain>
<DataPropertyRange>
  <DataProperty IRI="#seed"/>
  <Datatype abbreviatedIRI="xsd:integer"/>
</DataPropertyRange>
<Declaration>
  <DataProperty IRI="#dontReplaceMissingValues"/>
</Declaration>
<SubDataPropertyOf>
  <DataProperty IRI="#dontReplaceMissingValues"/>
  <DataProperty IRI="#algorithmConfigurationOptions"/>
</SubDataPropertyOf>
<DataPropertyDomain>
  <DataProperty IRI="#dontReplaceMissingValues"/>
  <Class IRI="#Canopy"/>
</DataPropertyDomain>
<DataPropertyRange>
  <DataProperty IRI="#dontReplaceMissingValues"/>
  <Datatype abbreviatedIRI="xsd:boolean"/>
</DataPropertyRange>
<Declaration>
  <DataProperty IRI="#t2"/>
</Declaration>
<SubDataPropertyOf>
  <DataProperty IRI="#t2"/>
  <DataProperty IRI="#algorithmConfigurationOptions"/>
</SubDataPropertyOf>
<DataPropertyDomain>
  <DataProperty IRI="#t2"/>
  <Class IRI="#Canopy"/>
</DataPropertyDomain>
<DataPropertyRange>
  <DataProperty IRI="#t2"/>
  <Datatype abbreviatedIRI="xsd:float"/>
</DataPropertyRange>
```

```
    <Declaration>
      <DataProperty IRI="#t1"/>
    </Declaration>
    <SubDataPropertyOf>
      <DataProperty IRI="#t1"/>
      <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <DataPropertyDomain>
      <DataProperty IRI="#t1"/>
      <Class IRI="#Canopy"/>
    </DataPropertyDomain>
    <DataPropertyRange>
      <DataProperty IRI="#t1"/>
      <Datatype abbreviatedIRI="xsd:float"/>
    </DataPropertyRange>
    <Declaration>
      <DataProperty IRI="#numClusters"/>
    </Declaration>
    <SubDataPropertyOf>
      <DataProperty IRI="#numClusters"/>
      <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <DataPropertyDomain>
      <DataProperty IRI="#numClusters"/>
      <Class IRI="#Canopy"/>
    </DataPropertyDomain>
    <DataPropertyRange>
      <DataProperty IRI="#numClusters"/>
      <Datatype abbreviatedIRI="xsd:integer"/>
    </DataPropertyRange>
    <Declaration>
      <DataProperty IRI="#minimumCanopyDensity"/>
    </Declaration>
    <SubDataPropertyOf>
      <DataProperty IRI="#minimumCanopyDensity"/>
      <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <DataPropertyDomain>
```

```
    <DataProperty IRI="#minimumCanopyDensity"/>
    <Class IRI="#Canopy"/>
  </DataPropertyDomain>
  <DataPropertyRange>
    <DataProperty IRI="#minimumCanopyDensity"/>
    <Datatype abbreviatedIRI="xsd:float"/>
  </DataPropertyRange>
  <Declaration>
    <DataProperty IRI="#periodicPruningRate"/>
  </Declaration>
  <SubDataPropertyOf>
    <DataProperty IRI="#periodicPruningRate"/>
    <DataProperty IRI="#algorithmConfigurationOptions"/>
  </SubDataPropertyOf>
  <DataPropertyDomain>
    <DataProperty IRI="#periodicPruningRate"/>
    <Class IRI="#Canopy"/>
  </DataPropertyDomain>
  <DataPropertyRange>
    <DataProperty IRI="#periodicPruningRate"/>
    <Datatype abbreviatedIRI="xsd:integer"/>
</DataPropertyRange>
```

B.1.1.37. Cobweb and Classit

Formally, Coweb-Classit algorithm executions are defined as instances of the class `CowebClassit` in OWL. The definition of this class and its data properties (configuration options) are listed below.

```
    <Declaration>
    <Class IRI="#CowebClassit"/>
  </Declaration>
  <SubClassOf>
    <Class IRI="#CowebClassit"/>
    <Class IRI="#UnsupervisedDataMiningAlgorithm"/>
  </SubClassOf>
  <Declaration>
    <DataProperty IRI="#seed"/>
  </Declaration>
```

```
    <SubDataPropertyOf>
      <DataProperty IRI="#seed"/>
      <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <DataPropertyDomain>
      <DataProperty IRI="#seed"/>
      <Class IRI="#CowebClassit"/>
    </DataPropertyDomain>
    <DataPropertyRange>
      <DataProperty IRI="#seed"/>
      <Datatype abbreviatedIRI="xsd:integer"/>
    </DataPropertyRange>
    <Declaration>
      <DataProperty IRI="#saveInstanceData"/>
    </Declaration>
    <SubDataPropertyOf>
      <DataProperty IRI="#saveInstanceData"/>
      <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <DataPropertyDomain>
      <DataProperty IRI="#saveInstanceData"/>
      <Class IRI="#CowebClassit"/>
    </DataPropertyDomain>
    <DataPropertyRange>
      <DataProperty IRI="#saveInstanceData"/>
      <Datatype abbreviatedIRI="xsd:boolean"/>
    </DataPropertyRange>
    <Declaration>
      <DataProperty IRI="#acuity"/>
    </Declaration>
    <SubDataPropertyOf>
      <DataProperty IRI="#acuity"/>
      <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <DataPropertyDomain>
      <DataProperty IRI="#acuity"/>
      <Class IRI="#CowebClassit"/>
    </DataPropertyDomain>
```

```
    <DataPropertyRange>
      <DataProperty IRI="#acuity"/>
      <Datatype abbreviatedIRI="xsd:float"/>
    </DataPropertyRange>
    <Declaration>
      <DataProperty IRI="#cutoff"/>
    </Declaration>
    <SubDataPropertyOf>
      <DataProperty IRI="#cutoff"/>
      <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <DataPropertyDomain>
      <DataProperty IRI="#cutoff"/>
      <Class IRI="#CowebClassit"/>
    </DataPropertyDomain>
    <DataPropertyRange>
      <DataProperty IRI="#cutoff"/>
      <Datatype abbreviatedIRI="xsd:double"/>
   </DataPropertyRange>
```

B.1.1.38. EM

Formally, EM (expectation maximisation) algorithm executions are defined as instances of the class `EM` in OWL. The definition of this class and its data properties (configuration options) are listed below.

```
    <Declaration>
      <Class IRI="#EM"/>
    </Declaration>
    <SubClassOf>
      <Class IRI="#EM"/>
      <Class IRI="#UnsupervisedDataMiningAlgorithm"/>
    </SubClassOf>
    <Declaration>
      <DataProperty IRI="#seed"/>
    </Declaration>
    <SubDataPropertyOf>
      <DataProperty IRI="#seed"/>
      <DataProperty IRI="#algorithmConfigurationOptions"/>
```

```
    </SubDataPropertyOf>
    <DataPropertyDomain>
      <DataProperty IRI="#seed"/>
      <Class IRI="#EM"/>
    </DataPropertyDomain>
    <DataPropertyRange>
      <DataProperty IRI="#seed"/>
      <Datatype abbreviatedIRI="xsd:integer"/>
    </DataPropertyRange>
    <Declaration>
      <DataProperty IRI="#numFolds"/>
    </Declaration>
    <SubDataPropertyOf>
      <DataProperty IRI="#numFolds"/>
      <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <DataPropertyDomain>
      <DataProperty IRI="#numFolds"/>
      <Class IRI="#EM"/>
    </DataPropertyDomain>
    <DataPropertyRange>
      <DataProperty IRI="#numFolds"/>
      <Datatype abbreviatedIRI="xsd:integer"/>
    </DataPropertyRange>
    <Declaration>
      <DataProperty IRI="#numExecutionSlots"/>
    </Declaration>
    <SubDataPropertyOf>
      <DataProperty IRI="#numExecutionSlots"/>
      <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <DataPropertyDomain>
      <DataProperty IRI="#numExecutionSlots"/>
      <Class IRI="#EM"/>
    </DataPropertyDomain>
    <DataPropertyRange>
      <DataProperty IRI="#numExecutionSlots"/>
      <Datatype abbreviatedIRI="xsd:integer"/>
```

```
    </DataPropertyRange>
    <Declaration>
      <DataProperty IRI="#numKMeansRuns"/>
    </Declaration>
    <SubDataPropertyOf>
      <DataProperty IRI="#numKMeansRuns"/>
      <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <DataPropertyDomain>
      <DataProperty IRI="#numKMeansRuns"/>
      <Class IRI="#EM"/>
    </DataPropertyDomain>
    <DataPropertyRange>
      <DataProperty IRI="#numKMeansRuns"/>
      <Datatype abbreviatedIRI="xsd:integer"/>
    </DataPropertyRange>
    <Declaration>
      <DataProperty IRI="#displayModeInOldFormat"/>
    </Declaration>
    <SubDataPropertyOf>
      <DataProperty IRI="#displayModeInOldFormat"/>
      <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <DataPropertyDomain>
      <DataProperty IRI="#displayModeInOldFormat"/>
      <Class IRI="#EM"/>
    </DataPropertyDomain>
    <DataPropertyRange>
      <DataProperty IRI="#displayModeInOldFormat"/>
      <Datatype abbreviatedIRI="xsd:boolean"/>
    </DataPropertyRange>
    <Declaration>
      <DataProperty IRI="#minLogLikelihoodImprovementIterating"/>
    </Declaration>
    <SubDataPropertyOf>
      <DataProperty IRI="#minLogLikelihoodImprovementIterating"/>
      <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
```

```
    <DataPropertyDomain>
      <DataProperty IRI="#minLogLikelihoodImprovementIterating"/>
      <Class IRI="#EM"/>
    </DataPropertyDomain>
    <DataPropertyRange>
      <DataProperty IRI="#minLogLikelihoodImprovementIterating"/>
      <Datatype abbreviatedIRI="xsd:string"/>
    </DataPropertyRange>
    <Declaration>
      <DataProperty IRI="#minLogLikelihoodImprovementCV"/>
    </Declaration>
    <SubDataPropertyOf>
      <DataProperty IRI="#minLogLikelihoodImprovementCV"/>
      <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <DataPropertyDomain>
      <DataProperty IRI="#minLogLikelihoodImprovementCV"/>
      <Class IRI="#EM"/>
    </DataPropertyDomain>
    <DataPropertyRange>
      <DataProperty IRI="#minLogLikelihoodImprovementCV"/>
      <Datatype abbreviatedIRI="xsd:string"/>
    </DataPropertyRange>
    <Declaration>
      <DataProperty IRI="#maximumNumberOfClusters"/>
    </Declaration>
    <SubDataPropertyOf>
      <DataProperty IRI="#maximumNumberOfClusters"/>
      <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <DataPropertyDomain>
      <DataProperty IRI="#maximumNumberOfClusters"/>
      <Class IRI="#EM"/>
    </DataPropertyDomain>
    <DataPropertyRange>
      <DataProperty IRI="#maximumNumberOfClusters"/>
      <Datatype abbreviatedIRI="xsd:integer"/>
    </DataPropertyRange>
```

```
    <Declaration>
      <DataProperty IRI="#numClusters"/>
    </Declaration>
    <SubDataPropertyOf>
      <DataProperty IRI="#numClusters"/>
      <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <DataPropertyDomain>
      <DataProperty IRI="#numClusters"/>
      <Class IRI="#EM"/>
    </DataPropertyDomain>
    <DataPropertyRange>
      <DataProperty IRI="#numClusters"/>
      <Datatype abbreviatedIRI="xsd:integer"/>
    </DataPropertyRange>
    <Declaration>
      <DataProperty IRI="#maxIterations"/>
    </Declaration>
    <SubDataPropertyOf>
      <DataProperty IRI="#maxIterations"/>
      <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <DataPropertyDomain>
      <DataProperty IRI="#maxIterations"/>
      <Class IRI="#EM"/>
    </DataPropertyDomain>
    <DataPropertyRange>
      <DataProperty IRI="#maxIterations"/>
      <Datatype abbreviatedIRI="xsd:integer"/>
    </DataPropertyRange>
    <Declaration>
      <DataProperty IRI="#minStdDev"/>
    </Declaration>
    <SubDataPropertyOf>
      <DataProperty IRI="#minStdDev"/>
      <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <DataPropertyDomain>
```

```
    <DataProperty IRI="#minStdDev"/>
    <Class IRI="#EM"/>
  </DataPropertyDomain>
  <DataPropertyRange>
    <DataProperty IRI="#minStdDev"/>
    <Datatype abbreviatedIRI="xsd:string"/>
  </DataPropertyRange>
```

B.1.1.39. *Farthest First*

Formally, Farthest First algorithm executions are defined as instances of the class `FarthestFirst` in OWL. The definition of this class and its data properties (configuration options) are listed below.

```
    <Declaration>
    <Class IRI="#FarthestFirst"/>
  </Declaration>
  <SubClassOf>
    <Class IRI="#FarthestFirst"/>
    <Class IRI="#UnsupervisedDataMiningAlgorithm"/>
  </SubClassOf>
  <Declaration>
    <DataProperty IRI="#seed"/>
  </Declaration>
  <SubDataPropertyOf>
    <DataProperty IRI="#seed"/>
    <DataProperty IRI="#algorithmConfigurationOptions"/>
  </SubDataPropertyOf>
  <DataPropertyDomain>
    <DataProperty IRI="#seed"/>
    <Class IRI="#FarthestFirst"/>
  </DataPropertyDomain>
  <DataPropertyRange>
    <DataProperty IRI="#seed"/>
    <Datatype abbreviatedIRI="xsd:integer"/>
  </DataPropertyRange>
  <Declaration>
    <DataProperty IRI="#numClusters"/>
  </Declaration>
```

```
<SubDataPropertyOf>
  <DataProperty IRI="#numClusters"/>
  <DataProperty IRI="#algorithmConfigurationOptions"/>
</SubDataPropertyOf>
<DataPropertyDomain>
  <DataProperty IRI="#numClusters"/>
  <Class IRI="#FarthestFirst"/>
</DataPropertyDomain>
<DataPropertyRange>
  <DataProperty IRI="#numClusters"/>
  <Datatype abbreviatedIRI="xsd:integer"/>
</DataPropertyRange>
```

B.1.1.40. Simple K Means

Formally, Simple K Means algorithm executions are defined as instances of the class `SimpleKMeans` in OWL. The definition of this class and its data properties (configuration options) are listed below.

```
<Declaration>
  <Class IRI="#SimpleKMeans"/>
</Declaration>
<SubClassOf>
  <Class IRI="#SimpleKMeans"/>
  <Class IRI="#UnsupervisedDataMiningAlgorithm"/>
</SubClassOf>
<Declaration>
  <DataProperty IRI="#seed"/>
</Declaration>
<SubDataPropertyOf>
  <DataProperty IRI="#seed"/>
  <DataProperty IRI="#algorithmConfigurationOptions"/>
</SubDataPropertyOf>
<DataPropertyDomain>
  <DataProperty IRI="#seed"/>
  <Class IRI="#SimpleKMeans"/>
</DataPropertyDomain>
<DataPropertyRange>
  <DataProperty IRI="#seed"/>
```

```
      <Datatype abbreviatedIRI="xsd:integer"/>
    </DataPropertyRange>
    <Declaration>
      <DataProperty IRI="#displayStdDevs"/>
    </Declaration>
    <SubDataPropertyOf>
      <DataProperty IRI="#displayStdDevs"/>
      <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <DataPropertyDomain>
      <DataProperty IRI="#displayStdDevs"/>
      <Class IRI="#SimpleKMeans"/>
    </DataPropertyDomain>
    <DataPropertyRange>
      <DataProperty IRI="#displayStdDevs"/>
      <Datatype abbreviatedIRI="xsd:boolean"/>
    </DataPropertyRange>
    <Declaration>
      <DataProperty IRI="#numExecutionSlots"/>
    </Declaration>
    <SubDataPropertyOf>
      <DataProperty IRI="#numExecutionSlots"/>
      <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <DataPropertyDomain>
      <DataProperty IRI="#numExecutionSlots"/>
      <Class IRI="#SimpleKMeans"/>
    </DataPropertyDomain>
    <DataPropertyRange>
      <DataProperty IRI="#numExecutionSlots"/>
      <Datatype abbreviatedIRI="xsd:integer"/>
    </DataPropertyRange>
    <Declaration>
      <DataProperty IRI="#dontReplaceMissingValues"/>
    </Declaration>
    <SubDataPropertyOf>
      <DataProperty IRI="#dontReplaceMissingValues"/>
      <DataProperty IRI="#algorithmConfigurationOptions"/>
```

```
    </SubDataPropertyOf>
    <DataPropertyDomain>
      <DataProperty IRI="#dontReplaceMissingValues"/>
      <Class IRI="#SimpleKMeans"/>
    </DataPropertyDomain>
    <DataPropertyRange>
      <DataProperty IRI="#dontReplaceMissingValues"/>
      <Datatype abbreviatedIRI="xsd:boolean"/>
    </DataPropertyRange>
    <Declaration>
      <DataProperty IRI="#canopyMinimumCanopyDensity"/>
    </Declaration>
    <SubDataPropertyOf>
      <DataProperty IRI="#canopyMinimumCanopyDensity"/>
      <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <DataPropertyDomain>
      <DataProperty IRI="#canopyMinimumCanopyDensity"/>
      <Class IRI="#SimpleKMeans"/>
    </DataPropertyDomain>
    <DataPropertyRange>
      <DataProperty IRI="#canopyMinimumCanopyDensity"/>
      <Datatype abbreviatedIRI="xsd:float"/>
    </DataPropertyRange>
    <Declaration>
      <DataProperty IRI="#canopyT2"/>
    </Declaration>
    <SubDataPropertyOf>
      <DataProperty IRI="#canopyT2"/>
      <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <DataPropertyDomain>
      <DataProperty IRI="#canopyT2"/>
      <Class IRI="#SimpleKMeans"/>
    </DataPropertyDomain>
    <DataPropertyRange>
      <DataProperty IRI="#canopyT2"/>
      <Datatype abbreviatedIRI="xsd:float"/>
```

```
    </DataPropertyRange>
    <Declaration>
      <DataProperty IRI="#numClusters"/>
    </Declaration>
    <SubDataPropertyOf>
      <DataProperty IRI="#numClusters"/>
      <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <DataPropertyDomain>
      <DataProperty IRI="#numClusters"/>
      <Class IRI="#SimpleKMeans"/>
    </DataPropertyDomain>
    <DataPropertyRange>
      <DataProperty IRI="#numClusters"/>
      <Datatype abbreviatedIRI="xsd:integer"/>
    </DataPropertyRange>
    <Declaration>
      <DataProperty IRI="#maxIterations"/>
    </Declaration>
    <SubDataPropertyOf>
      <DataProperty IRI="#maxIterations"/>
      <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <DataPropertyDomain>
      <DataProperty IRI="#maxIterations"/>
      <Class IRI="#SimpleKMeans"/>
    </DataPropertyDomain>
    <DataPropertyRange>
      <DataProperty IRI="#maxIterations"/>
      <Datatype abbreviatedIRI="xsd:integer"/>
    </DataPropertyRange>
    <Declaration>
      <DataProperty IRI="#preserveInstancesOrder"/>
    </Declaration>
    <SubDataPropertyOf>
      <DataProperty IRI="#preserveInstancesOrder"/>
      <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
```

```
    <DataPropertyDomain>
      <DataProperty IRI="#preserveInstancesOrder"/>
      <Class IRI="#SimpleKMeans"/>
    </DataPropertyDomain>
    <DataPropertyRange>
      <DataProperty IRI="#preserveInstancesOrder"/>
      <Datatype abbreviatedIRI="xsd:boolean"/>
    </DataPropertyRange>
    <Declaration>
      <DataProperty IRI="#canopyPeriodicPruningRate"/>
    </Declaration>
    <SubDataPropertyOf>
      <DataProperty IRI="#canopyPeriodicPruningRate"/>
      <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <DataPropertyDomain>
      <DataProperty IRI="#canopyPeriodicPruningRate"/>
      <Class IRI="#SimpleKMeans"/>
    </DataPropertyDomain>
    <DataPropertyRange>
      <DataProperty IRI="#canopyPeriodicPruningRate"/>
      <Datatype abbreviatedIRI="xsd:integer"/>
    </DataPropertyRange>
    <Declaration>
      <DataProperty IRI="#canopyMaxNumCanopiesToHoldInMemory"/>
    </Declaration>
    <SubDataPropertyOf>
      <DataProperty IRI="#canopyMaxNumCanopiesToHoldInMemory"/>
      <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <DataPropertyDomain>
      <DataProperty IRI="#canopyMaxNumCanopiesToHoldInMemory"/>
      <Class IRI="#SimpleKMeans"/>
    </DataPropertyDomain>
    <DataPropertyRange>
      <DataProperty IRI="#canopyMaxNumCanopiesToHoldInMemory"/>
      <Datatype abbreviatedIRI="xsd:integer"/>
    </DataPropertyRange>
```

```
<Declaration>
  <DataProperty IRI="#initializationMethod"/>
</Declaration>
<SubDataPropertyOf>
  <DataProperty IRI="#initializationMethod"/>
  <DataProperty IRI="#algorithmConfigurationOptions"/>
</SubDataPropertyOf>
<DataPropertyDomain>
  <DataProperty IRI="#initializationMethod"/>
  <Class IRI="#SimpleKMeans"/>
</DataPropertyDomain>
<DataPropertyRange>
  <DataProperty IRI="#initializationMethod"/>
  <Datatype abbreviatedIRI="xsd:string"/>
</DataPropertyRange>
<Declaration>
  <DataProperty IRI="#distanceFunction"/>
</Declaration>
<SubDataPropertyOf>
  <DataProperty IRI="#distanceFunction"/>
  <DataProperty IRI="#algorithmConfigurationOptions"/>
</SubDataPropertyOf>
<DataPropertyDomain>
  <DataProperty IRI="#distanceFunction"/>
  <Class IRI="#SimpleKMeans"/>
</DataPropertyDomain>
<DataPropertyRange>
  <DataProperty IRI="#distanceFunction"/>
  <Datatype abbreviatedIRI="xsd:string"/>
</DataPropertyRange>
<Declaration>
  <DataProperty IRI="#canopyT1"/>
</Declaration>
<SubDataPropertyOf>
  <DataProperty IRI="#canopyT1"/>
  <DataProperty IRI="#algorithmConfigurationOptions"/>
</SubDataPropertyOf>
<DataPropertyDomain>
```

```
    <DataProperty IRI="#canopyT1"/>
    <Class IRI="#SimpleKMeans"/>
  </DataPropertyDomain>
  <DataPropertyRange>
    <DataProperty IRI="#canopyT1"/>
    <Datatype abbreviatedIRI="xsd:float"/>
  </DataPropertyRange>
  <Declaration>
    <DataProperty IRI="#fastDistanceCalc"/>
  </Declaration>
  <SubDataPropertyOf>
    <DataProperty IRI="#fastDistanceCalc"/>
    <DataProperty IRI="#algorithmConfigurationOptions"/>
  </SubDataPropertyOf>
  <DataPropertyDomain>
    <DataProperty IRI="#fastDistanceCalc"/>
    <Class IRI="#SimpleKMeans"/>
  </DataPropertyDomain>
  <DataPropertyRange>
    <DataProperty IRI="#fastDistanceCalc"/>
    <Datatype abbreviatedIRI="xsd:boolean"/>
  </DataPropertyRange>
  <Declaration>
    <DataProperty IRI="#reduceNumberOfDistanceCalcsViaCanopies"/>
  </Declaration>
  <SubDataPropertyOf>
    <DataProperty IRI="#reduceNumberOfDistanceCalcsViaCanopies"/>
    <DataProperty IRI="#algorithmConfigurationOptions"/>
  </SubDataPropertyOf>
  <DataPropertyDomain>
    <DataProperty IRI="#reduceNumberOfDistanceCalcsViaCanopies"/>
    <Class IRI="#SimpleKMeans"/>
  </DataPropertyDomain>
  <DataPropertyRange>
    <DataProperty IRI="#reduceNumberOfDistanceCalcsViaCanopies"/>
    <Datatype abbreviatedIRI="xsd:boolean"/>
</DataPropertyRange>
```

B.1.2. Axioms

In this section we present the axiom definitions for each language concept presented in Chapter 3 .

B.1.2.1. Policy Model

In order to express the cardinalities of the relations of the `PolicyModel` class we use the following axioms of type EquivalentTo: one axiom for the relationship of type `aimedAt` with `Goal` with cardinality exactly 1, another for the relationship of type `involvesStakeholders` with `Stakeholder` with cardinality min 0 and one for the relationship of type hasWorkflows with DataAnalyticsWorkflow with cardinality min 0. The formal definition of these axioms is presented below.

```
<EquivalentClasses>
        <Class IRI="#PolicyModel"/>
        <ObjectExactCardinality cardinality="1">
            <ObjectProperty IRI="#aimedAt"/>
            <Class IRI="#Goal"/>
        </ObjectExactCardinality>
    </EquivalentClasses>
<EquivalentClasses>
        <Class IRI="#PolicyModel"/>
        <ObjectMinCardinality cardinality="0">
            <ObjectProperty IRI="#involvesStakeholders"/>
            <Class IRI="#Stakeholder"/>
        </ObjectMinCardinality>
    </EquivalentClasses>
<EquivalentClasses>
        <Class IRI="#PolicyModel"/>
        <ObjectMinCardinality cardinality="0">
            <ObjectProperty IRI="#hasWorkflows"/>
            <Class IRI="#DataAnalyticsWorfklow"/>
        </ObjectMinCardinality>
    </EquivalentClasses>
```

B.1.2.2. Goal

In order to express the cardinalities of the relations of the `Goal` class we use the following axioms of type `EquivalentTo`: one axiom for the relationship of type `refinedInto` with `Objective` with cardinality minimum 0, another for the relationship of type `policyModel` with `PolicyModel` with cardinality exactly 1. The formal definition of these axioms is presented below.

```
<EquivalentClasses>
        <Class IRI="#Goal"/>
        <ObjectMinCardinality cardinality="0">
            <ObjectProperty IRI="#refinedInto"/>
            <Class IRI="#Objective"/>
        </ObjectMinCardinality>
    </EquivalentClasses>
    <EquivalentClasses>
        <Class IRI="#Goal"/>
        <ObjectExactCardinality cardinality="1">
            <ObjectProperty IRI="#policyModel"/>
            <Class IRI="#PolicyModel"/>
        </ObjectExactCardinality>
    </EquivalentClasses>
```

B.1.2.3. Objective

In order to express the cardinalities of the relations of the `Objective` class we use the following axioms of type `EquivalentTo`: one axiom for the relationship of type `goal` with `Goal` with cardinality exactly 1 and another for the relationship of type `canBeAddressedBy` with `PolicyAction` with cardinality minimum 1. The formal definition of these axioms is presented below.

```
<EquivalentClasses>
        <Class IRI="#Objective"/>
        <ObjectExactCardinality cardinality="1">
            <ObjectProperty IRI="#goal"/>
            <Class IRI="#Goal"/>
        </ObjectExactCardinality>
    </EquivalentClasses>
<EquivalentClasses>
        <Class IRI="#Objective"/>
        <ObjectMinCardinality cardinality="1">
            <ObjectProperty IRI="#canBeAddressedBy"/>
```

```
            <Class IRI="#PolicyAction"/>
        </ObjectMinCardinality>
    </EquivalentClasses>
```

B.1.2.4. Policy Action

In order to express the cardinalities of the relations of the `PolicyAction` class we use the following axioms of type EquivalentTo: one axiom for the relationship of type `objective` with `Objective` with cardinality exactly 1, four axioms for the relationships with itself of type a1 and a2 with cardinality minimum 0 and of type dependant and pre-requisite with cardinalities minimum 1, one for the relationship of type position with Position with cardinality minimum 0 and another for the relationship of type `isEvaluatedBy` with `Criterion` with cardinality exactly 1. The formal definition of these axioms is presented below.

```
<EquivalentClasses>
        <Class IRI="#PolicyAction"/>
        <ObjectMinCardinality cardinality="0">
            <ObjectProperty IRI="#a1"/>
            <Class IRI="#PolicyAction"/>
        </ObjectMinCardinality>
    </EquivalentClasses>
    <EquivalentClasses>
        <Class IRI="#PolicyAction"/>
        <ObjectMinCardinality cardinality="0">
            <ObjectProperty IRI="#a2"/>
            <Class IRI="#PolicyAction"/>
        </ObjectMinCardinality>
    </EquivalentClasses>
    <EquivalentClasses>
        <Class IRI="#PolicyAction"/>
        <ObjectMinCardinality cardinality="1">
            <ObjectProperty IRI="#dependant"/>
            <Class IRI="#PolicyAction"/>
        </ObjectMinCardinality>
    </EquivalentClasses>
    <EquivalentClasses>
        <Class IRI="#PolicyAction"/>
        <ObjectMinCardinality cardinality="1">
```

```
                <ObjectProperty IRI="#pre-requisite"/>
                <Class IRI="#PolicyAction"/>
            </ObjectMinCardinality>
        </EquivalentClasses>
        <EquivalentClasses>
            <Class IRI="#PolicyAction"/>
            <ObjectMinCardinality cardinality="0">
                <ObjectProperty IRI="#position"/>
                <Class IRI="#Position"/>
            </ObjectMinCardinality>
        </EquivalentClasses>
        <EquivalentClasses>
            <Class IRI="#PolicyAction"/>
            <ObjectExactCardinality cardinality="1">
                <ObjectProperty IRI="#isEvaluatedBy"/>
                <Class IRI="#Criterion"/>
            </ObjectExactCardinality>
        </EquivalentClasses>
```

B.1.2.5. ***Criterion***

In order to express the cardinalities of the relations of the `Criterion` class we use the following axioms of type EquivalentTo: one axiom for the relationship of type `constraints` with `Dataset` with cardinality minimum 1, another for the relationship of type `policyAction` with `PolicyAction` with cardinality exactly 1 and a third one for the relationship of type `specifies` with `DataAnalyticsWorkflow` with cardinality exactly 1. The formal definition of these axioms is presented below.

```
    <EquivalentClasses>
            <Class IRI="#Criterion"/>
            <ObjectMinCardinality cardinality="1">
                <ObjectProperty IRI="#constraints"/>
                <Class IRI="#Dataset"/>
            </ObjectMinCardinality>
        </EquivalentClasses>
        <EquivalentClasses>
            <Class IRI="#Criterion"/>
            <ObjectExactCardinality cardinality="1">
```

```
            <ObjectProperty IRI="#policyAction"/>
            <Class IRI="#PolicyAction"/>
        </ObjectExactCardinality>
    </EquivalentClasses>
    <EquivalentClasses>
        <Class IRI="#Criterion"/>
        <ObjectExactCardinality cardinality="1">
            <ObjectProperty IRI="#specifies"/>
            <Class IRI="#DataAnalyticsWorfklow"/>
        </ObjectExactCardinality>
    </EquivalentClasses>
```

B.1.2.6. Stakeholder

In order to express the cardinalities of the relations of the `Stakeholder` class we use the following axioms of type EquivalentTo: one axiom for the relationship of type `policyModel` with `policyModel` with cardinality minimum 1 and another two for the relationships of type `advocates` and `proposes` with `Position` with cardinality minimum 0 and minimum 1 respectively. The formal definition of these axioms is presented below.

```
<EquivalentClasses>
        <Class IRI="#Stakeholder"/>
        <ObjectMinCardinality cardinality="1">
            <ObjectProperty IRI="#policyModel"/>
            <Class IRI="#PolicyModel"/>
        </ObjectMinCardinality>
    </EquivalentClasses>
<EquivalentClasses>
        <Class IRI="#Stakeholder"/>
        <ObjectMinCardinality cardinality="0">
            <ObjectProperty IRI="#advocates"/>
            <Class IRI="#Position"/>
        </ObjectMinCardinality>
    </EquivalentClasses>
    <EquivalentClasses>
        <Class IRI="#Stakeholder"/>
        <ObjectMinCardinality cardinality="1">
            <ObjectProperty IRI="#proposes"/>
```

```
            <Class IRI="#Position"/>
        </ObjectMinCardinality>
    </EquivalentClasses>
```

B.1.2.7. Position

In order to express the cardinalities of the relations of the `Position` class we use the following axioms of type `EquivalentTo`: one axiom for the relationship of type `refersTo` with `PolicyAction` with cardinality exactly 1 and another two for the relationship of type `stakeholder` with `Stakeholder: one inverse of advocates` with cardinality minimum 0 and one inverse of proposes with cardinality minimum 1. We also introduce an axiom of type DisjointClasses to express that a position can be one of the following: NeutralPosition, OpposingPosition or SupportivePosition. The formal definition of these axioms is presented below.

```
<EquivalentClasses>
        <Class IRI="#Position"/>
        <ObjectExactCardinality cardinality="1">
            <ObjectProperty IRI="#refersTo"/>
            <Class IRI="#PolicyAction"/>
        </ObjectExactCardinality>
    </EquivalentClasses>
<EquivalentClasses>
        <Class IRI="#Position"/>
        <ObjectMinCardinality cardinality="0">
            <ObjectProperty IRI="#stakeholder"/>
            <Class IRI="#Stakeholder"/>
        </ObjectMinCardinality>
    </EquivalentClasses>
    <EquivalentClasses>
        <Class IRI="#Position"/>
        <ObjectMinCardinality cardinality="1">
            <ObjectProperty IRI="#stakeholder"/>
            <Class IRI="#Stakeholder"/>
        </ObjectMinCardinality>
    </EquivalentClasses>
<DisjointClasses>
        <Class IRI="#NeutralPosition"/>
        <Class IRI="#OpposingPosition"/>
```

```
        <Class IRI="#SupportivePosition"/>
    </DisjointClasses>
```

B.1.2.8. Data Analytics Workflow

In order to express the cardinalities of the relations of the DataAnalyticsWorkflow class we use the following axioms of type EquivalentTo: one axiom for the relationship of type criterion with Criterion with cardinality minimum 0, one for the relationship of type isComposedOf with DataAnalyticsTask with cardinality minimum 1 and another for the relationship of type policyModel with policyModel with cardinality minimum 0. The formal definition of these axioms is presented below.

```
    <EquivalentClasses>
            <Class IRI="#DataAnalyticsWorfklow"/>
            <ObjectMinCardinality cardinality="0">
                <ObjectProperty IRI="#criterion"/>
                <Class IRI="#Criterion"/>
            </ObjectMinCardinality>
        </EquivalentClasses>
        <EquivalentClasses>
            <Class IRI="#DataAnalyticsWorfklow"/>
            <ObjectMinCardinality cardinality="1">
                <ObjectProperty IRI="#isComposedOf"/>
                <Class IRI="#DataAnalyticsTask"/>
            </ObjectMinCardinality>
        </EquivalentClasses>
        <EquivalentClasses>
            <Class IRI="#DataAnalyticsWorfklow"/>
            <ObjectMinCardinality cardinality="0">
                <ObjectProperty IRI="#policyModel"/>
                <Class IRI="#PolicyModel"/>
            </ObjectMinCardinality>
        </EquivalentClasses>
```

B.1.2.9. Workflow Execution Type

We introduce an axiom of type DisjointClasses in order to express that a workflow execution type can be one of the following: DataChangeDriven, ExecutionUponRequest or PeriodicExecution. The formal definition of these axioms is presented below.

```
<DisjointClasses>
        <Class IRI="#DataChangeDriven"/>
```

```
        <Class IRI="#ExecutionUponRequest"/>
        <Class IRI="#PeriodicExecution"/>
    </DisjointClasses>
```

B.1.2.10. Data Set

In order to express the cardinalities of the relations of the `Dataset` class we use the following axioms of type EquivalentTo: one axiom for the relationship of type `criterion` with `Criterion` with cardinality minimum 1, one axiom for the relationship of type `spec` with `DataSpecification` with cardinality exactly 1 and two axioms for the relationship of type `dataAnalyticsTask` with `DataAnalyticsTask:` one inverse of `input` with cardinality minimum 0 and another inverse of `output` with cardinality maximum 1. We also introduce one axiom of type DisjointClasses in order to express that a dataset can be `DataAnalyticsModel, DataStream,` or `StaticSet` and another axiom of the same type to show that a timed dataset can be either an `AbsoluteDataset` or a `ShiftingDataset`. The formal definition of these axioms is presented below.

```
<EquivalentClasses>
        <Class IRI="#Dataset"/>
        <ObjectMinCardinality cardinality="0">
            <ObjectProperty IRI="#criterion"/>
            <Class IRI="#Criterion"/>
        </ObjectMinCardinality>
    </EquivalentClasses>
    <EquivalentClasses>
        <Class IRI="#Dataset"/>
        <ObjectExactCardinality cardinality="1">
            <ObjectProperty IRI="#spec"/>
            <Class IRI="#DataSpecification"/>
        </ObjectExactCardinality>
    </EquivalentClasses>
<EquivalentClasses>
        <Class IRI="#Dataset"/>
        <ObjectMinCardinality cardinality="0">
            <ObjectProperty IRI="#dataAnalyticsTask"/>
            <Class IRI="#DataAnalyticsTask"/>
        </ObjectMinCardinality>
    </EquivalentClasses>
```

```
    <EquivalentClasses>
        <Class IRI="#Dataset"/>
        <ObjectMaxCardinality cardinality="1">
            <ObjectProperty IRI="#dataAnalyticsTask"/>
            <Class IRI="#DataAnalyticsTask"/>
        </ObjectMaxCardinality>
    </EquivalentClasses>
<DisjointClasses>
        <Class IRI="#DataAnalyticsModel"/>
        <Class IRI="#DataStream"/>
        <Class IRI="#StaticSet"/>
    </DisjointClasses>
<DisjointClasses>
        <Class IRI="#AbsoluteDataset"/>
        <Class IRI="#ShiftingDataset"/>
    </DisjointClasses>
```

B.1.2.11. <u>Data Specification</u>

In order to express the cardinalities of the relations of the `DataSpecification` and `OutputDataSpecification` classes we use the following axioms of type EquivalentTo: one axiom for the relationship of type `dataset` with `Dataset` with cardinality minimum 1 and another for the relationship of `OutputDataSpecification` of type `algorithm` with `Algorithm` with cardinality exactly 1. The formal definition of these axioms is presented below.

```
<EquivalentClasses>
        <Class IRI="#DataSpecification"/>
        <ObjectMinCardinality cardinality="1">
            <ObjectProperty IRI="#dataset"/>
            <Class IRI="#Dataset"/>
        </ObjectExactCardinality>
    </EquivalentClasses>
<EquivalentClasses>
        <Class IRI="#OutputDataSpecification"/>
        <ObjectExactCardinality cardinality="1">
            <ObjectProperty IRI="#algorithm"/>
            <Class IRI="#Algorithm"/>
        </ObjectExactCardinality>
```

```
        </EquivalentClasses>
```

B.1.2.12. Data Analytics Task

In order to express the cardinalities of the relations of the `DataAnalyticsTask` class we use the following axioms of type EquivalentTo: one axiom for the relationship of type `dataAnalyticsWorkflow` with `dataAnalyticsWorkflow` with cardinality minimum 1, two axioms for the relationships with Dataset with cardinality minimum 1: one of type input and one of type output and another for the relationship of type `utilizes` with `Method` with cardinality exactly 1. We also introduce another axiom of type `DisjointClasses` to express that a data analytics task can only be one of the below: a data mining task, a statistical analysis task of a data processing task. The formal definition of these axioms is presented below.

```
    <EquivalentClasses>
            <Class IRI="#DataAnalyticsTask"/>
            <ObjectMinCardinality cardinality="0">
                <ObjectProperty IRI="#dataAnalyticsWorkflow"/>
                <Class IRI="#DataAnalyticsWorfklow"/>
            </ObjectMinCardinality>
        </EquivalentClasses>
        <EquivalentClasses>
            <Class IRI="#DataAnalyticsTask"/>
            <ObjectMinCardinality cardinality="1">
                <ObjectProperty IRI="#input"/>
                <Class IRI="#Dataset"/>
            </ObjectMinCardinality>
        </EquivalentClasses>
        <EquivalentClasses>
            <Class IRI="#DataAnalyticsTask"/>
            <ObjectMinCardinality cardinality="1">
                <ObjectProperty IRI="#output"/>
                <Class IRI="#Dataset"/>
            </ObjectMinCardinality>
        </EquivalentClasses>
        <EquivalentClasses>
            <Class IRI="#DataAnalyticsTask"/>
            <ObjectExactCardinality cardinality="1">
                <ObjectProperty IRI="#utilizes"/>
```

```
            <Class IRI="#Method"/>
        </ObjectExactCardinality>
    </EquivalentClasses>
<DisjointClasses>
        <Class IRI="#DataMiningTask"/>
        <Class IRI="#DataProcessingTask"/>
        <Class IRI="#StatisticalAnalysisTask"/>
    </DisjointClasses>
```

B.1.2.13. Method

In order to express the cardinality of the relation of the `Method` class with `DataAnalyticsTask` we use the following axiom of type `EquivalentTo`: one axiom for the relationship of type `dataAnalyticsTask` with cardinality minimum 0. We also introduce an axiom of type `DisjointClasses` in order to express that a method can be either an algorithm or an operation and another axiom of the same type to express that an operation can be one of the following: `DataCleaningOperation`, `FilterOperation`, `JoinOperation`, `ProjectOperation`, or `SamplingOperation`. The formal definition of these axioms is presented below.

```
<EquivalentClasses>
        <Class IRI="#Method"/>
        <ObjectMinCardinality cardinality="0">
            <ObjectProperty IRI="#dataAnalyticsTask"/>
            <Class IRI="#DataAnalyticsTask"/>
        </ObjectMinCardinality>
</EquivalentClasses>
<DisjointClasses>
        <Class IRI="#Algorithm"/>
        <Class IRI="#Operation"/>
</DisjointClasses>
<DisjointClasses>
        <Class IRI="#DataCleaningOperation"/>
        <Class IRI="#FilterOperation"/>
        <Class IRI="#JoinOperation"/>
        <Class IRI="#ProjectOperation"/>
        <Class IRI="#SamplingOperation"/>
</DisjointClasses>
```

B.1.2.14. Sampling Operation

We introduce an axiom of type `DisjointClasses` in order to express that a sampling operation can be one of the following: `AdaptiveSampling`, `ClusteringSampling`, `RandomSampling` or `StratifiedRandomSampling`. The formal definition of these axioms is presented below.

```
<DisjointClasses>
        <Class IRI="#AdaptiveSampling"/>
        <Class IRI="#ClusteringSampling"/>
        <Class IRI="#RandomSampling"/>
        <Class IRI="#StratifiedRandomSampling"/>
    </DisjointClasses>
```

B.1.2.15. Join Operation

We introduce an axiom of type `DisjointClasses` in order to express that a join operation can be one of the following: `FullJoin`, `InnerJoin`, `LeftOuter` or `RightOuter`. The formal definition of these axioms is presented below.

```
<DisjointClasses>
        <Class IRI="#FullJoin"/>
        <Class IRI="#InnerJoin"/>
        <Class IRI="#LeftOuter"/>
        <Class IRI="#RightOuter"/>
    </DisjointClasses>
```

B.1.2.16. Statistical Analysis Algorithm

We introduce an axiom of type `DisjointClasses` in order to express that a statistical analysis algorithm can be one of the following: `ANOVA`, `Breusch-Pagan_Test`, `F-test`, `LinearRegression` or `Fisher's_Exact_Test`. To express the cardinality of the relation of the `StatisticalAnalysisTask` class with `StatisticalAnalysisAlgorithm` we use the following axiom of type EquivalentTo of type `utilizes` with cardinality exactly 1. The formal definition of these axioms is presented below.

```
  <DisjointClasses>
          <Class IRI="#ANOVA"/>
          <Class IRI="#Breusch-Pagan_Test"/>
          <Class IRI="#F-test"/>
```

```
        <Class IRI="#LinearRegression"/>
        <Class IRI="#Fisher&apos;s_Exact_Test"/>
</DisjointClasses>
<EquivalentClasses>
        <Class IRI="#DataMiningTask"/>
        <ObjectExactCardinality cardinality="1">
            <ObjectProperty IRI="#utilizes"/>
            <Class IRI="#DataMiningAlgorithm"/>
        </ObjectExactCardinality>
    </EquivalentClasses>
```

B.1.2.17. Data Mining Algorithm

To express the cardinality of the relation of the `DataMiningTask` class with `DataMiningAlgorithm` we use the following axiom of type EquivalentTo of type `utilizes` with cardinality exactly 1. We also introduce one axiom of type `DisjointClasses` to express that a data mining algorithm can be either supervised or unsupervised and two other axioms of the same type to express the different types of supervised and unsupervised data mining algorithm types. The formal definitions of these axioms are presented below.

```
    <EquivalentClasses>
        <Class IRI="#DataMiningTask"/>
        <ObjectExactCardinality cardinality="1">
            <ObjectProperty IRI="#utilizes"/>
            <Class IRI="#DataMiningAlgorithm"/>
        </ObjectExactCardinality>
    </EquivalentClasses>
<DisjointClasses>
        <Class IRI="#SupervisedDataMiningAlgorithm"/>
        <Class IRI="#UnsupervisedDataMiningAlgorithm"/>
    </DisjointClasses>
<DisjointClasses>
        <Class IRI="#DecisionTable"/>
        <Class IRI="#GaussianProcesses"/>
        <Class IRI="#IBk"/>
        <Class IRI="#J48"/>
        <Class IRI="#LinearRegression"/>
        <Class IRI="#MultinomialLogisticRegression"/>
```

```
            <Class IRI="#NaiveBayes"/>
            <Class IRI="#RandomForest"/>
            <Class IRI="#RandomTree"/>
            <Class IRI="#ZeroR"/>
        </DisjointClasses>
    <DisjointClasses>
            <Class IRI="#Canopy"/>
            <Class IRI="#CowebClassit"/>
            <Class IRI="#EM"/>
            <Class IRI="#FarthestFirst"/>
            <Class IRI="#SimpleKMeans"/>
    </DisjointClasses>
```

B.1.3. Examples

In this section we present the examples for each language concept presented in Chapter 3 .

B.1.3.1. Policy Model

As an example of the definition of instances of the classes in OWL we have created an instance of the `PolicyModel` class named “PM_1” with label “Policy Model of Addressing Barriers to HA Use”. The declaration of this instance in OWL is presented below.

```
   <Declaration>
           <NamedIndividual IRI="#PM_1"/>
       </Declaration>
       <ClassAssertion>
           <Class IRI="#PolicyModel"/>
           <NamedIndividual IRI="#PM_1"/>
        </ClassAssertion>
        <AnnotationAssertion>
           <AnnotationProperty abbreviatedIRI="rdfs:label"/>
           <IRI>#PM_1</IRI>
           <LiteraldatatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax
ns#PlainLiteral">Policy Model of Addressing Barriers to HA Use</Literal>
       </AnnotationAssertion>
```

For the purposes of the example, we have used DAW1, which is an instance of the `DataAnalyticsWorkflow` class, `Stakeholder_1`, `Stakeholder_2`, `Stakeholder_3`, `Stakeholder_4` and `Stakeholder_5` which are instances of the `Stakeholder` class and `Goal_1` which is an instance of `Goal` class.

Below we present the formal definition of the relations of `PM_1`: that `PM_1` has the relation of type `hasWorkflows` with `DAW1`, the relation of type `involvesStakeholders` with `Stakeholder_1`, `Stakeholder_2`, `Stakeholder_3`, `Stakeholder_4` and `Stakeholder_5` and the relation of type `aimedAt` with `Goal_1`.

```
    <ObjectPropertyAssertion>
        <ObjectProperty IRI="#hasWworkflows"/>
        <NamedIndividual IRI="#PM_1"/>
        <NamedIndividual IRI="#DAW1"/>
    </ObjectPropertyAssertion>
    <ObjectPropertyAssertion>
        <ObjectProperty IRI="#involvesStakeholders"/>
        <NamedIndividual IRI="#PM_1"/>
        <NamedIndividual IRI="#Stakeholder_1"/>
    </ObjectPropertyAssertion>
    <ObjectPropertyAssertion>
        <ObjectProperty IRI="#involvesStakeholders"/>
        <NamedIndividual IRI="#PM_1"/>
        <NamedIndividual IRI="#Stakeholder_2"/>
    </ObjectPropertyAssertion>
    <ObjectPropertyAssertion>
        <ObjectProperty IRI="#involvesStakeholders"/>
        <NamedIndividual IRI="#PM_1"/>
        <NamedIndividual IRI="#Stakeholder_3"/>
       </ObjectPropertyAssertion>
    <ObjectPropertyAssertion>
        <ObjectProperty IRI="#involvesStakeholders"/>
        <NamedIndividual IRI="#PM_1"/>
        <NamedIndividual IRI="#Stakeholder_4"/>
       </ObjectPropertyAssertion>
    <ObjectPropertyAssertion>
        <ObjectProperty IRI="#involvesStakeholders"/>
        <NamedIndividual IRI="#PM_1"/>
        <NamedIndividual IRI="#Stakeholder_5"/>
       </ObjectPropertyAssertion>
<ObjectPropertyAssertion>
```

```
        <ObjectProperty IRI="#aimedAt"/>
        <NamedIndividual IRI="#PM_1"/>
        <NamedIndividual IRI="#Goal_1"/>
    </ObjectPropertyAssertion>
```

B.1.3.2. Goal

As an example of the definition of instances of the classes in OWL we have created an instance of the `Goal` class named “Goal_1” with label “Policy Model of Addressing Barriers to HA Use”. The declaration of this instance in OWL is presented below.

```
   <Declaration>
           <NamedIndividual IRI="#Goal_1"/>
       </Declaration>
   <ClassAssertion>
           <Class IRI="#Goal"/>
           <NamedIndividual IRI="#Goal_1"/>
       </ClassAssertion>
   <AnnotationAssertion>
           <AnnotationProperty abbreviatedIRI="rdfs:label"/>
           <IRI>#Goal_1</IRI>
           <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">Addressing Barriers to HA Use</Literal>
       </AnnotationAssertion>
   <DataPropertyAssertion>
           <DataProperty IRI="#description"/>
           <NamedIndividual IRI="#Goal_1"/>
           <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">The purpose of this case study is to determine the largest
barriers that affect hearing aid use in a population in order to make public
health policy decisions to address them.</Literal>
       </DataPropertyAssertion>
       <DataPropertyAssertion>
           <DataProperty IRI="#rationale"/>
           <NamedIndividual IRI="#Goal_1"/>
           <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">Barriers to hearing aid use are a significant public health
problem. Barriers occur at all levels of the process of provision of hearing
aids including at the level of the HA user. The big data gathered about users
```

```
through EVOTION would enable policy makers to choose which barriers to address
in a population, in order to improve hearing aid use and hence reduce the
burden of hearing loss in that population.</Literal>
      </DataPropertyAssertion>
```

For the purposes of the example, we have used `Obj_1`, `Obj_2` and `Obj_3` which are an instance of the `Objective` class.

Below we present the formal definition of the relations of `Goal_1`: `Goal_1` has the relation of type `refinedInto` with `Obj_1`, `Obj_2` and `Obj_3`.

```
  <ObjectPropertyAssertion>
          <ObjectProperty IRI="#refinedInto"/>
          <NamedIndividual IRI="#Goal_1"/>
          <NamedIndividual IRI="#Obj_1"/>
      </ObjectPropertyAssertion>
      <ObjectPropertyAssertion>
          <ObjectProperty IRI="#refinedInto"/>
          <NamedIndividual IRI="#Goal_1"/>
          <NamedIndividual IRI="#Obj_2"/>
      </ObjectPropertyAssertion>
      <ObjectPropertyAssertion>
          <ObjectProperty IRI="#refinedInto"/>
          <NamedIndividual IRI="#Goal_1"/>
          <NamedIndividual IRI="#Obj_3"/>
      </ObjectPropertyAssertion>
```

B.1.3.3. Objective

As examples of the definition of instances of the classes in OWL we have created three instances of the `Objective` class: one named “Obj_1” with label “Explore whether the occupation of HA users affects their daily usage”, one named “Obj_2” with label “Explore whether the educational level of HA users affects their daily usage” and one named “obj_3” with label “Explore whether the age of HA users affects their daily usage”. The declaration of these instances in OWL are presented below.

```
  <Declaration>
          <NamedIndividual IRI="#Obj_1"/>
      </Declaration>
  <Declaration>
          <NamedIndividual IRI="#Obj_2"/>
```

```
        </Declaration>
    <Declaration>
            <NamedIndividual IRI="#Obj_3"/>
        </Declaration>
    <ClassAssertion>
            <Class IRI="#Objective"/>
            <NamedIndividual IRI="#Obj_1"/>
        </ClassAssertion>
        <ClassAssertion>
            <Class IRI="#Objective"/>
            <NamedIndividual IRI="#Obj_2"/>
        </ClassAssertion>
        <ClassAssertion>
            <Class IRI="#Objective"/>
            <NamedIndividual IRI="#Obj_3"/>
        </ClassAssertion>
    <AnnotationAssertion>
            <AnnotationProperty abbreviatedIRI="rdfs:label"/>
            <IRI>#Obj_1</IRI>
            <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">Explore whether the occupation of HA users affects their daily
usage</Literal>
        </AnnotationAssertion>
        <AnnotationAssertion>
            <AnnotationProperty abbreviatedIRI="rdfs:label"/>
            <IRI>#Obj_2</IRI>
            <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">Explore whether the educational level of HA users affects
their daily usage</Literal>
        </AnnotationAssertion>
        <AnnotationAssertion>
            <AnnotationProperty abbreviatedIRI="rdfs:label"/>
            <IRI>#Obj_3</IRI>
            <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">Explore whether the age of HA users affects their daily
usage</Literal>
        </AnnotationAssertion>
```

Below we present the formal definition of the relations of `Obj_1`, `Obj_2` and `Obj_3` of type `refinedInto` with `Goal_1`.

```
<ObjectPropertyAssertion>
        <ObjectProperty IRI="#refinedInto"/>
        <NamedIndividual IRI="#Goal_1"/>
        <NamedIndividual IRI="#Obj_1"/>
    </ObjectPropertyAssertion>
    <ObjectPropertyAssertion>
        <ObjectProperty IRI="#refinedInto"/>
        <NamedIndividual IRI="#Goal_1"/>
        <NamedIndividual IRI="#Obj_2"/>
    </ObjectPropertyAssertion>
    <ObjectPropertyAssertion>
        <ObjectProperty IRI="#refinedInto"/>
        <NamedIndividual IRI="#Goal_1"/>
        <NamedIndividual IRI="#Obj_3"/>
    </ObjectPropertyAssertion>
```

B.1.3.4. Policy Action

As examples of the definition of instances of the classes in OWL we have created three instances of the `PolicyAction` class: one named "`PA_1`" with label "Occupation Related ACTION", one named "`PA_2`" with label "Explore whether the educational level of HA users affects their daily usage" and one named "`PA_3`" with label "Explore whether the age of HA users affects their daily usage". The declaration of these instances in OWL are presented below.

```
<Declaration>
        <NamedIndividual IRI="#PA_1"/>
    </Declaration>
<Declaration>
        <NamedIndividual IRI="#PA_2"/>
    </Declaration>
<Declaration>
        <NamedIndividual IRI="#PA_3"/>
    </Declaration>
<ClassAssertion>
        <Class IRI="#PolicyAction"/>
```

```
        <NamedIndividual IRI="#PA_1"/>
    </ClassAssertion>
    <ClassAssertion>
        <Class IRI="#PolicyAction"/>
        <NamedIndividual IRI="#PA_2"/>
    </ClassAssertion>
    <ClassAssertion>
        <Class IRI="#PolicyAction"/>
        <NamedIndividual IRI="#PA_3"/>
    </ClassAssertion>
  <AnnotationAssertion>
        <AnnotationProperty abbreviatedIRI="rdfs:label"/>
        <IRI>#PA_1</IRI>
        <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">Occupation Related ACTION</Literal>
    </AnnotationAssertion>
    <AnnotationAssertion>
        <AnnotationProperty abbreviatedIRI="rdfs:label"/>
        <IRI>#PA_2</IRI>
        <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">Age related ACTION</Literal>
    </AnnotationAssertion>
    <AnnotationAssertion>
        <AnnotationProperty abbreviatedIRI="rdfs:label"/>
        <IRI>#PA_3</IRI>
        <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">Educational level related ACTION</Literal>
    </AnnotationAssertion>
  <DataPropertyAssertion>
        <DataProperty IRI="#description"/>
        <NamedIndividual IRI="#PA_1"/>
        <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">A particular occupation has to be addressed with additional
measures to improve HA use</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
        <DataProperty IRI="#description"/>
        <NamedIndividual IRI="#PA_2"/>
```

```
          <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-ns#PlainLiteral">Age related fitting</Literal>
      </DataPropertyAssertion>
      <DataPropertyAssertion>
          <DataProperty IRI="#description"/>
          <NamedIndividual IRI="#PA_3"/>
          <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-ns#PlainLiteral">Failure to reach a particular educational level has to be addressed to improve HA use</Literal>
      </DataPropertyAssertion>
```

Below we present the formal definition of the relations of `PA_1` with `Obj_1`, `PA_2` with `Obj_2` and `PA_3` with `Obj_3` of type `objective` and the relations of `PA_1` with `OccupationCriterion`, `PA_2` with `EducationalLevelCriterion` and `PA_3` with `AgeCriterion` of type `isEvaluatedBy`.

```
  <ObjectPropertyAssertion>
          <ObjectProperty IRI="#objective"/>
          <NamedIndividual IRI="#Obj_1"/>
          <NamedIndividual IRI="#Goal_1"/>
      </ObjectPropertyAssertion>
      <ObjectPropertyAssertion>
          <ObjectProperty IRI="#objective"/>
          <NamedIndividual IRI="#Obj_2"/>
          <NamedIndividual IRI="#Goal_1"/>
      </ObjectPropertyAssertion>
      <ObjectPropertyAssertion>
          <ObjectProperty IRI="#objective"/>
          <NamedIndividual IRI="#Obj_3"/>
          <NamedIndividual IRI="#Goal_1"/>
      </ObjectPropertyAssertion>
  <ObjectPropertyAssertion>
          <ObjectProperty IRI="#isEvaluatedBy"/>
          <NamedIndividual IRI="#PA_1"/>
          <NamedIndividual IRI="#OccupationCriterion"/>
      </ObjectPropertyAssertion>
      <ObjectPropertyAssertion>
          <ObjectProperty IRI="#isEvaluatedBy"/>
          <NamedIndividual IRI="#PA_2"/>
```

```
        <NamedIndividual IRI="#EducationalLevelCriterion"/>
    </ObjectPropertyAssertion>
    <ObjectPropertyAssertion>
        <ObjectProperty IRI="#isEvaluatedBy"/>
        <NamedIndividual IRI="#PA_3"/>
        <NamedIndividual IRI="#AgeCriterion"/>
    </ObjectPropertyAssertion>
```

B.1.3.5. Criterion

As examples of the definition of instances of the classes in OWL we have created three instances of the `Criterion` class: one named "`OccupationCriterion`", one named "`EducationalLevelCriterion`" and one named "`AgeCriterion`". Since all the criteria have the same weight, we have left the weight data property empty. We have inserted the criteria logical expressions to each criterion. The declaration of these instances in OWL are presented below.

```
<Declaration>
        <NamedIndividual IRI="#OccupationCriterion"/>
    </Declaration>
<Declaration>
        <NamedIndividual IRI="#EducationalLevelCriterion"/>
    </Declaration>
<Declaration>
        <NamedIndividual IRI="#AgeCriterion"/>
    </Declaration>
<ClassAssertion>
        <Class IRI="#Criterion"/>
        <NamedIndividual IRI="#OccupationCriterion"/>
    </ClassAssertion>
<ClassAssertion>
        <Class IRI="#Criterion"/>
        <NamedIndividual IRI="#EducationalLevelCriterion"/>
    </ClassAssertion>
<ClassAssertion>
        <Class IRI="#Criterion"/>
        <NamedIndividual IRI="#AgeCriterion"/>
    </ClassAssertion>
<DataPropertyAssertion>
```

```
        <DataProperty IRI="#logicalExpression"/>
        <NamedIndividual IRI="#OccupationCriterion"/>
        <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">constraints.spec.Reg_Detailed_Stats [Factor = Occup].P-value
&lt; 0.05
  AND
  constraints.spec.Reg_Overall_Stats.R Square&gt;0.5
  AND
  constraints.spec.Reg_ANOVA.Significance F&lt;0.05
  AND
  constraints.spec.HET_ANOVA.Significance F&gt;=0.05</Literal>
    </DataPropertyAssertion>
  <DataPropertyAssertion>
        <DataProperty IRI="#logicalExpression"/>
        <NamedIndividual IRI="#EducationalLevelCriterion"/>
        <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">constraints.spec.Reg_Detailed_Stats [Factor = Edu_Level].P-
value &lt; 0.05
  AND
  constraints.spec.Reg_Overall_Stats.R Square&gt;0.5
  AND
  constraints.spec.Reg_ANOVA.Significance F&lt;0.05
  AND
  constraints.spec.HET_ANOVA.Significance F&gt;=0.05</Literal>
    </DataPropertyAssertion>
  <DataPropertyAssertion>
        <DataProperty IRI="#logicalExpression"/>
        <NamedIndividual IRI="#AgeCriterion"/>
        <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">constraints.spec.Reg_Detailed_Stats [Factor = Age].P-value
&lt; 0.05
  AND
  constraints.spec.Reg_Overall_Stats.R Square&gt;0.5
  AND
  constraints.spec.Reg_ANOVA.Significance F&lt;0.05
  AND
  constraints.spec.HET_ANOVA.Significance F&gt;=0.05</Literal>
    </DataPropertyAssertion>
```

Below we present the formal definition of relations of `PA_1` with `OccupationCriterion`, `PA_2` with `EducationalLevelCriterion` and `PA_3` with `AgeCriterion` of type `isEvaluatedBy`.

```
<ObjectPropertyAssertion>
        <ObjectProperty IRI="#isEvaluatedBy"/>
        <NamedIndividual IRI="#PA_1"/>
        <NamedIndividual IRI="#OccupationCriterion"/>
    </ObjectPropertyAssertion>
    <ObjectPropertyAssertion>
        <ObjectProperty IRI="#isEvaluatedBy"/>
        <NamedIndividual IRI="#PA_2"/>
        <NamedIndividual IRI="#EducationalLevelCriterion"/>
    </ObjectPropertyAssertion>
    <ObjectPropertyAssertion>
        <ObjectProperty IRI="#isEvaluatedBy"/>
        <NamedIndividual IRI="#PA_3"/>
        <NamedIndividual IRI="#AgeCriterion"/>
    </ObjectPropertyAssertion>
```

B.1.3.6. Stakeholder

As examples of the definition of instances of the classes in OWL we have created five instances of the `Stakeholder` class: one named “`Stakeholder_1`” with label “Regional ENT-specialists’ Advisory Committee”, one named “`Stakeholder_2`” with label “Regional Directorate for Social support”, one named “`Stakeholder_3`” with label “Regional structures of the national Health Insurance Fund”, one named “`Stakeholder_4`” with label “HA vendors/fitting experts” and one named “`Stakeholder_5`” with label “Regional repres. of patients”. The declaration of these instances in OWL are presented below.

```
<Declaration>
        <NamedIndividual IRI="#Stakeholder_1"/>
    </Declaration>
<Declaration>
        <NamedIndividual IRI="#Stakeholder_2"/>
    </Declaration>
<Declaration>
        <NamedIndividual IRI="#Stakeholder_3"/>
    </Declaration>
```

```
   <Declaration>
           <NamedIndividual IRI="#Stakeholder_4"/>
       </Declaration>
   <Declaration>
           <NamedIndividual IRI="#Stakeholder_5"/>
       </Declaration>
   <AnnotationAssertion>
           <AnnotationProperty abbreviatedIRI="rdfs:comment"/>
           <IRI>#Stakeholder_1</IRI>
           <Literal       datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">in their role as prescribing the use of HAs</Literal>
       </AnnotationAssertion>
       <AnnotationAssertion>
           <AnnotationProperty abbreviatedIRI="rdfs:label"/>
           <IRI>#Stakeholder_1</IRI>
           <Literal       datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">Regional ENT-specialists’ Advisory Committee</Literal>
       </AnnotationAssertion>
       <AnnotationAssertion>
           <AnnotationProperty abbreviatedIRI="rdfs:comment"/>
           <IRI>#Stakeholder_2</IRI>
           <Literal       datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">in their role as authorising financial support for purchasing
HAs and performing follow-up on ad-ministration and use</Literal>
       </AnnotationAssertion>
       <AnnotationAssertion>
           <AnnotationProperty abbreviatedIRI="rdfs:label"/>
           <IRI>#Stakeholder_2</IRI>
           <Literal       datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">Regional Directorate for Social support</Literal>
       </AnnotationAssertion>
       <AnnotationAssertion>
           <AnnotationProperty abbreviatedIRI="rdfs:comment"/>
           <IRI>#Stakeholder_3</IRI>
           <Literal       datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">in their role as funding clinical pathways</Literal>
       </AnnotationAssertion>
       <AnnotationAssertion>
```

```
            <AnnotationProperty abbreviatedIRI="rdfs:label"/>
            <IRI>#Stakeholder_3</IRI>
            <Literal        datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">Regional   structures   of   the   national   Health   Insurance
Fund</Literal>
        </AnnotationAssertion>
        <AnnotationAssertion>
            <AnnotationProperty abbreviatedIRI="rdfs:comment"/>
            <IRI>#Stakeholder_4</IRI>
            <Literal        datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">providing follow-up rehab</Literal>
        </AnnotationAssertion>
        <AnnotationAssertion>
            <AnnotationProperty abbreviatedIRI="rdfs:label"/>
            <IRI>#Stakeholder_4</IRI>
            <Literal        datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">HA vendors/fitting experts</Literal>
        </AnnotationAssertion>
        <AnnotationAssertion>
            <AnnotationProperty abbreviatedIRI="rdfs:comment"/>
            <IRI>#Stakeholder_5</IRI>
            <Literal        datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">regional repres. of patients</Literal>
        </AnnotationAssertion>
        <AnnotationAssertion>
            <AnnotationProperty abbreviatedIRI="rdfs:label"/>
            <IRI>#Stakeholder_5</IRI>
            <Literal        datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">Patients' association</Literal>
        </AnnotationAssertion>
```

Below we present the formal definition of the relations of PM_1: that PM_1 has the relation of type `policyModel` with `Stakeholder_1`, `Stakeholder_2` `Stakeholder_3`, `Stakeholder_4` and `Stakeholder_5`.

```
    <ObjectPropertyAssertion>
            <ObjectProperty IRI="#policyModel"/>
            <NamedIndividual IRI="#Stakeholder_1"/>
            <NamedIndividual IRI="#PM_1"/>
```

```
    </ObjectPropertyAssertion>
    <ObjectPropertyAssertion>
        <ObjectProperty IRI="#policyModel"/>
        <NamedIndividual IRI="#Stakeholder_2"/>
        <NamedIndividual IRI="#PM_1"/>
    </ObjectPropertyAssertion>
    <ObjectPropertyAssertion>
        <ObjectProperty IRI="#policyModel"/>
        <NamedIndividual IRI="#Stakeholder_3"/>
        <NamedIndividual IRI="#PM_1"/>
       </ObjectPropertyAssertion>
    <ObjectPropertyAssertion>
        <ObjectProperty IRI="#policyModel"/>
        <NamedIndividual IRI="#Stakeholder_4"/>
        <NamedIndividual IRI="#PM_1"/>
       </ObjectPropertyAssertion>
    <ObjectPropertyAssertion>
        <ObjectProperty IRI="#policyModel"/>
        <NamedIndividual IRI="#Stakeholder_5"/>
        <NamedIndividual IRI="#PM_1"/>
       </ObjectPropertyAssertion>
```

B.1.3.7. ***Position***

As examples of the definition of instances of the classes in OWL we have created three instances of the subclasses of the Position class: one named “`Supportive_1`”, instance of the `SupportivePosition` class, one named “`Opposing_1`”, instance of the `OpposingPosition` class and one named “`Neutral_1`”, instance of the `NeutralPosition` class. The declaration of these instances in OWL are presented below.

```
<Declaration>
        <NamedIndividual IRI="#Supportive_1"/>
    </Declaration>
<ClassAssertion>
        <Class IRI="#SupportivePosition"/>
        <NamedIndividual IRI="#Supportive_1"/>
    </ClassAssertion>
<Declaration>
```

```
        <NamedIndividual IRI="#Opposing_1"/>
    </Declaration>
<ClassAssertion>
        <Class IRI="#OpposingPosition"/>
        <NamedIndividual IRI="#Opposing_1"/>
  </ClassAssertion>
<Declaration>
        <NamedIndividual IRI="#Neutral_1"/>
    </Declaration>
<ClassAssertion>
        <Class IRI="#NeutralPosition"/>
        <NamedIndividual IRI="#Neutral_1"/>
  </ClassAssertion>
```

Below we present the formal definition of the relations of the above declared instances: that `Suppotive_1`, `Opposing_1` and `Neutral_1` have the relation of type `refersTo` with Policy Action `PA_1`, that `Suppotive_1` has the relation of type `proposes` with `Stakeholder_1` and the relationship of type `advocates` with `Stakeholder_1` and `Stakeholder_4`, `Opposing_1` has the relationship of type `proposes` with `Stakeholder_2` and the relationship of type `advocates` with `Stakeholder_2` and `Stakeholder_5` and `Neutral_1` has the relationship of type `proposes` with `Stakeholder_3` and the relationship of type `advocates` with `Stakeholder_3`.

```
    <ObjectPropertyAssertion>
        <ObjectProperty IRI="#refersTo"/>
        <NamedIndividual IRI="#PA_1"/>
        <NamedIndividual IRI="#Neutral_1"/>
    </ObjectPropertyAssertion>
    <ObjectPropertyAssertion>
        <ObjectProperty IRI="#refersTo"/>
        <NamedIndividual IRI="#PA_1"/>
        <NamedIndividual IRI="#Opposing_1"/>
    </ObjectPropertyAssertion>
    <ObjectPropertyAssertion>
        <ObjectProperty IRI="#refersTo"/>
        <NamedIndividual IRI="#PA_1"/>
        <NamedIndividual IRI="#Supportive_1"/>
    </ObjectPropertyAssertion>
     <ObjectPropertyAssertion>
```

```
        <ObjectProperty IRI="#proposes"/>
        <NamedIndividual IRI="#Stakeholder_1"/>
        <NamedIndividual IRI="#Supportive_1"/>
    </ObjectPropertyAssertion>
    <ObjectPropertyAssertion>
        <ObjectProperty IRI="#advocates"/>
        <NamedIndividual IRI="#Stakeholder_1"/>
        <NamedIndividual IRI="#Supportive_1"/>
    </ObjectPropertyAssertion>
    <ObjectPropertyAssertion>
        <ObjectProperty IRI="#advocates"/>
        <NamedIndividual IRI="#Stakeholder_4"/>
        <NamedIndividual IRI="#Supportive_1"/>
    </ObjectPropertyAssertion>
    <ObjectPropertyAssertion>
        <ObjectProperty IRI="#proposes"/>
        <NamedIndividual IRI="#Stakeholder_2"/>
        <NamedIndividual IRI="#Opposing_1"/>
    </ObjectPropertyAssertion>
    <ObjectPropertyAssertion>
        <ObjectProperty IRI="#advocates"/>
        <NamedIndividual IRI="#Stakeholder_2"/>
        <NamedIndividual IRI="#Opposing_1"/>
    </ObjectPropertyAssertion>
    <ObjectPropertyAssertion>
        <ObjectProperty IRI="#advocates"/>
        <NamedIndividual IRI="#Stakeholder_5"/>
        <NamedIndividual IRI="#Opposing_1"/>
    </ObjectPropertyAssertion>
    <ObjectPropertyAssertion>
        <ObjectProperty IRI="#proposes"/>
        <NamedIndividual IRI="#Stakeholder_3"/>
        <NamedIndividual IRI="#Neutral_1"/>
    </ObjectPropertyAssertion>
    <ObjectPropertyAssertion>
        <ObjectProperty IRI="#advocates"/>
        <NamedIndividual IRI="#Stakeholder_3"/>
        <NamedIndividual IRI="#Neutral_1"/>
```

```
        </ObjectPropertyAssertion>
```

B.1.3.8. Data Analytics Workflow

As an example of the definition of instances of the classes in OWL we have created one instance of the DataAnalyticsWorkflow class named “DAW_1” with label “Data Analytics Workflow of Addressing Barriers to HA Use”. The declaration of this instance in OWL are presented below.

```
   <Declaration>
           <NamedIndividual IRI="#DAW_1"/>
       </Declaration>
   <AnnotationAssertion>
           <AnnotationProperty abbreviatedIRI="rdfs:label"/>
           <IRI>#DAW_1</IRI>
           <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">Data Analytics Workflow of Addressing Barriers to HA
Use</Literal>
       </AnnotationAssertion>
   <ClassAssertion>
           <Class IRI="#DataAnalyticsWorkflow"/>
           <NamedIndividual IRI="#DAW_1"/>
       </ClassAssertion>
```

Below we present the formal definition of relations of `DAW_1` with `OccupationCriterion`, `EducationalLevelCriterion` and `AgeCriterion` of type `specifies`.

```
   <ObjectPropertyAssertion>
           <ObjectProperty IRI="#criterion"/>
           <NamedIndividual IRI="#DAW_1"/>
           <NamedIndividual IRI="#OccupationCriterion"/>
       </ObjectPropertyAssertion>
       <ObjectPropertyAssertion>
           <ObjectProperty IRI="#criterion"/>
           <NamedIndividual IRI="#DAW_1"/>
           <NamedIndividual IRI="#EducationalLevelCriterion"/>
       </ObjectPropertyAssertion>
       <ObjectPropertyAssertion>
           <ObjectProperty IRI="#criterion"/>
           <NamedIndividual IRI="#DAW_1"/>
```

```
        <NamedIndividual IRI="#AgeCriterion"/>
    </ObjectPropertyAssertion>
```

B.1.3.9. Workflow Execution Type

As an example of the definition of instances of the classes in OWL we have created one instance of the `ExecutionUponRequest` class named “`EUR_1`”. The declaration of this instance in OWL are presented below.

```
<Declaration>
        <NamedIndividual IRI="#EUR_1"/>
    </Declaration>
<ClassAssertion>
        <Class IRI="#EUR_1"/>
        <NamedIndividual IRI="#ExecutionUponRequest"/>
    </ClassAssertion>
```

Below we present the formal definition of the relation of `EUR_1` with `DAW_1`, of type `executionType`.

```
<ObjectPropertyAssertion>
        <ObjectProperty IRI="#executionType"/>
        <NamedIndividual IRI="#DAW_1"/>
        <NamedIndividual IRI="#EUR_1"/>
    </ObjectPropertyAssertion>
```

B.1.3.10. Data Set

As an example of the definition of instances of the classes in OWL we have created two instances of the `StaticSet` class: one named “`AgeData`” and one named “`EducationalData`”, one instance of the `DataStream` class named “`DailyUsageData`” and one instance of the `DataAnalyticsModel` class named “`GaussianProcessesModel_1`”. The declaration of this instance in OWL are presented below.

```
<Declaration>
        <NamedIndividual IRI="#AgeData"/>
    </Declaration>
<ClassAssertion>
        <Class IRI="#StaticSet"/>
```

```
        <NamedIndividual IRI="#AgeData"/>
    </ClassAssertion>
<Declaration>
        <NamedIndividual IRI="#EducationalData"/>
    </Declaration>
<ClassAssertion>
        <Class IRI="# StaticSet "/>
        <NamedIndividual IRI="#EducationalData"/>
    </ClassAssertion>
<Declaration>
        <NamedIndividual IRI="#DailyUsageData"/>
    </Declaration>
<ClassAssertion>
        <Class IRI="#DataStream"/>
        <NamedIndividual IRI="#DailyUsageData"/>
    </ClassAssertion>
<Declaration>
        <NamedIndividual IRI="#GaussianProcessesModel_1"/>
    </Declaration>
<ClassAssertion>
        <Class IRI="#DataAnalyticsModel"/>
        <NamedIndividual IRI="#GaussianProcessesModel_1"/>
    </ClassAssertion>
```

B.1.3.11. <u>Data Specification</u>

As an example of the definition of instances of the classes in OWL we have created three instances of the `DataSpecification` class: one named “`AgeDS`”, one named “`EducationalDS`”, one named “`DailyUsageDS`”; and one instance of the `OutputDataSpecification` class named “`GaussianProcessesModelDS`”. The declaration of these instances in OWL are presented below.

```
<Declaration>
        <NamedIndividual IRI="#AgeDS"/>
    </Declaration>
<ClassAssertion>
        <Class IRI="DataSpecification"/>
        <NamedIndividual IRI="#AgeDS"/>
    </ClassAssertion>
<DataPropertyAssertion>
```

```
            <DataProperty IRI="#columnName"/>
            <NamedIndividual IRI="#AgeDS"/>
            <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">Age</Literal>
        </DataPropertyAssertion>
        <DataPropertyAssertion>
            <DataProperty IRI="#tableName"/>
            <NamedIndividual IRI="#AgeDS"/>
            <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">PATIENT</Literal>
        </DataPropertyAssertion>
    <Declaration>
            <NamedIndividual IRI="#EducationalDS"/>
        </Declaration>
    <ClassAssertion>
            <Class IRI="DataSpecification"/>
            <NamedIndividual IRI="#EducationalDS"/>
        </ClassAssertion>
    <DataPropertyAssertion>
            <DataProperty IRI="#columnName"/>
            <NamedIndividual IRI="#EducationalDS"/>
            <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">EducationalLevel</Literal>
        </DataPropertyAssertion>
        <DataPropertyAssertion>
            <DataProperty IRI="#tableName"/>
            <NamedIndividual IRI="#EducationalDS"/>
            <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">PATIENT</Literal>
        </DataPropertyAssertion>
    <Declaration>
            <NamedIndividual IRI="#DailyUsageDS"/>
        </Declaration>
    <ClassAssertion>
            <Class IRI="DataSpecification"/>
            <NamedIndividual IRI="#DailyUsageDS"/>
        </ClassAssertion>
    <DataPropertyAssertion>
```

```
        <DataProperty IRI="#tableName"/>
        <NamedIndividual IRI="#DailyUsageDS"/>
        <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">RETRO_HA</Literal>
    </DataPropertyAssertion>
  <Declaration>
        <NamedIndividual IRI="#GaussianProcessesModelDS"/>
    </Declaration>
  <ClassAssertion>
        <Class IRI="OutputDataSpecification"/>
        <NamedIndividual IRI="#GaussianProcessesModelDS"/>
    </ClassAssertion>
```

Below we present the formal definition of relations of `AgeDS` with `AgeData`, `EducationalDS` with `EducationalData`, `DailyUsageDS` with `DailyUsageData` of type `spec` and the relation of `GaussianProcessesModelDS` with `GaussianProcessesModel_1` of type `dataSpec`.

```
  <ObjectPropertyAssertion>
        <ObjectProperty IRI="#spec"/>
        <NamedIndividual IRI="#AgeData"/>
        <NamedIndividual IRI="#AgeDS"/>
    </ObjectPropertyAssertion>
  <ObjectPropertyAssertion>
        <ObjectProperty IRI="#spec"/>
        <NamedIndividual IRI="#EducationalData"/>
        <NamedIndividual IRI="#EducationalDS"/>
    </ObjectPropertyAssertion>
  <ObjectPropertyAssertion>
        <ObjectProperty IRI="#spec"/>
        <NamedIndividual IRI="#DailyUsageData"/>
        <NamedIndividual IRI="#DailyUsageDS"/>
    </ObjectPropertyAssertion>
  <ObjectPropertyAssertion>
        <ObjectProperty IRI="#dataSpec"/>
        <NamedIndividual IRI="#GaussianProcessesModel_1"/>
        <NamedIndividual IRI="#GaussianProcessesModelDS"/>
    </ObjectPropertyAssertion>
```

B.1.3.12. Sampling Operation

For example, sampling on patients' data taking a subset randomly and performing a preliminary analysis of clusters to evaluate how many clusters are feasible to obtain from the entire dataset.

As an example of the definition of instances of the classes in OWL we have created one instance of `RandomSampling` named "`RSO1`, one instance of the `StratifiedRandomSampling` class named "`SRSO1`", one instance of the `ClusteringSampling` class named "`CSO1`" and one instance of the `AdaptiveSampling` class named "`ASO1`". The declaration of these instances in OWL are presented below.

```
<Declaration>
        <NamedIndividual IRI="#RSO1"/>
    </Declaration>
<ClassAssertion>
        <Class IRI="#RandomSampling "/>
        <NamedIndividual IRI="#RSO1"/>
    </ClassAssertion>
<Declaration>
        <NamedIndividual IRI="#SRSO1"/>
    </Declaration>
<ClassAssertion>
        <Class IRI="#StratifiedRandomSampling"/>
        <NamedIndividual IRI="#SRSO1"/>
    </ClassAssertion>
<Declaration>
        <NamedIndividual IRI="#CSO1"/>
    </Declaration>
<ClassAssertion>
        <Class IRI="#ClusteringSampling"/>
        <NamedIndividual IRI="#CSO1"/>
    </ClassAssertion>
<Declaration>
        <NamedIndividual IRI="ASO1"/>
    </Declaration>
<ClassAssertion>
        <Class IRI="#AdaptiveSampling"/>
        <NamedIndividual IRI="#ASO1"/>
    </ClassAssertion>
```

B.1.3.13. Filter Operation

An example of simple filtering could be for selecting information related to the average usage of hearing devices.

In this case, we could select all information from retrospective data under the condition that the average usage time is greater than a certain threshold that domain experts (e.g., clinicians) consider relevant. For instance, that could be done to analyse the characteristics of patients that had an intensive usage of hearing devices, in order to study how to increase the average usage time of the whole patient population.

The SQL query would have a form like: select * from RETRO_HA where AV_USE_P>time

As an example of the definition of instances of the classes in OWL we have created one instance of `FilterOperation` named “`FO1`, with condition “select * from RETRO_HA where AV_USE_P>time”. The declaration of this instance in OWL are presented below.

```
    <Declaration>
        <NamedIndividual IRI="#FO1"/>
      </Declaration>
    <ClassAssertion>
        <Class IRI="#FilterOperation "/>
        <NamedIndividual IRI="#FO1"/>
      </ClassAssertion>
    <DataPropertyAssertion>
        <DataProperty IRI="#condition"/>
        <NamedIndividual IRI="#FO1"/>
        <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-ns#PlainLiteral">select * from
RETRO_HA where AV_USE_P&gt;time</Literal>
      </DataPropertyAssertion>
```

B.1.3.14. Join Operation

For example, join information coming from personal data, HA data and HA utilization data to analyse the engagement

```
select  PATIENT.*,  HEARING_COACH_TRAINING.*, RETRO_HA .*
from PATIENT
inner join  HEARING_COACH_TRAINING
on PATIENT.PATIENT_ID = HEARING_COACH_TRAINING.PATIENT_ID;
inner join RETRO_HA
```

on RETRO_HA.PATIENT_ID = HEARING_COACH_TRAINING.PATIENT_ID;

where HEARING_COACH_TRAINING< VALUE

As an example of the definition of instances of the classes in OWL we have created one instance of `InnerJoin` named “`IJO1`”, with condition “select PATIENT.*, HEARING_COACH_TRAINING.*, RETRO_HA .* from PATIENT inner join HEARING_COACH_TRAINING on PATIENT.PATIENT_ID = HEARING_COACH_TRAINING.PATIENT_ID; inner join RETRO_HA on RETRO_HA.PATIENT_ID = HEARING_COACH_TRAINING.PATIENT_ID; where HEARING_COACH_TRAINING< VALUE”. The declaration of this instance in OWL is presented below.

```
<Declaration>
    <NamedIndividual IRI="#IJO1"/>
     </Declaration>
<ClassAssertion>
        <Class IRI="#InnerJoin "/>
        <NamedIndividual IRI="#IJO1"/>
     </ClassAssertion>
<DataPropertyAssertion>
        <DataProperty IRI="#condition"/>
        <NamedIndividual IRI="#FO1"/>
        <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">select  PATIENT.*,  HEARING_COACH_TRAINING.*, RETRO_HA .* from
PATIENT inner join  HEAR-ING_COACH_TRAINING on PATIENT.PATIENT_ID =
HEARING_COACH_TRAINING.PATIENT_ID; inner join RETRO_HA  on RETRO_HA.PATIENT_ID
= HEARING_COACH_TRAINING.PATIENT_ID; where HEAR-ING_COACH_TRAINING<
VALUE</Literal>
     </DataPropertyAssertion>
```

B.1.3.15. <u>Project Operation</u>

For example, projecting the patients’ dataset on a subset of attributes where the age attribute is projected.

select PATIENT_ID, DATEOFBIRTH, CITY, CIVILSTATUS from PATIENT where DATEOFBIRTH between ’date1′ AND ’date2′

As an example of the definition of instances of the classes in OWL we have created one instance of ProjectOperation named “`PO1`”, with `projectRule` “select PATIENT_ID, DATEOFBIRTH, CITY, CIVILSTATUS from PATIENT where DATEOFBIRTH between ’date1′ AND ’date2′”. The declaration of this instance in OWL is presented below.

```
   <Declaration>
        <NamedIndividual IRI="#PO1"/>
       </Declaration>
   <ClassAssertion>
           <Class IRI="#ProjectOperation "/>
           <NamedIndividual IRI="#PO1"/>
       </ClassAssertion>
   <DataPropertyAssertion>
           <DataProperty IRI="#projectRule"/>
           <NamedIndividual IRI="#PO1"/>
           <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral"> select  PATIENT_ID,  DATEOFBIRTH, CITY,  CIVILSTATUS  from
PATIENT  where DATEOFBIRTH  between  ’date1’ AND ’date2’</Literal>
       </DataPropertyAssertion>
```

B.1.3.16. Data Cleaning Operation

For example, on a specific attribute, values could be normalized with respect to a given scale.

As an example of the definition of instances of the classes in OWL we have created one instance of `DataCleaningOperation` named “`DCO1`”. The declaration of this instance in OWL is presented below.

```
   <Declaration>
       <NamedIndividual IRI="#DCO1"/>
     </Declaration>
   <ClassAssertion>
       <Class IRI="#DataCleaningOperation "/>
       <NamedIndividual IRI="#DCO1"/>
     </ClassAssertion>
```

B.1.3.17. Linear Regression

As an example of the definition of instances of the classes in OWL we have created one instance of `LinearRegression` named “`LinearRegressionExecution`”. The declaration of this instance in OWL is presented below.

```
<Declaration>
        <NamedIndividual IRI="#LinearRegressionExecution"/>
    </Declaration>
<ClassAssertion>
        <Class IRI="#LinearRegression"/>
        <NamedIndividual IRI="#LinearRegressionExecution"/>
    </ClassAssertion>
```

B.1.3.18. ANOVA

As an example of the definition of instances of the classes in OWL we have created one instance of `MultiwayANNOVA` named “`multiwayExecution`”. The declaration of this instance in OWL is presented below.

```
<Declaration>
        <NamedIndividual IRI="#multiwayExecution"/>
    </Declaration>
<ClassAssertion>
        <Class IRI="#multiway_ANOVA"/>
        <NamedIndividual IRI="#multiwayExecution"/>
    </ClassAssertion>
```

B.1.3.19. Breusch-Pagan Test

As an example of the definition of instances of the classes in OWL we have created one instance of `BreuschPaganTest` named “`breuschExecution`”. The declaration of this instance in OWL is presented below.

```
<Declaration>
        <NamedIndividual IRI="#breuschExecution"/>
    </Declaration>
<ClassAssertion>
        <Class IRI="#BreuschPaganTest"/>
        <NamedIndividual IRI="#breuschExecution"/>
    </ClassAssertion>
```

B.1.3.20. F-Test

As an example of the definition of instances of the classes in OWL we have created one instance of `FTest` named “`fExecution`”. The declaration of this instance in OWL is presented below.

```
<Declaration>
    <NamedIndividual IRI="#fExecution"/>
</Declaration>
<ClassAssertion>
```

```
        <Class IRI="#FTest"/>
        <NamedIndividual IRI="#fExecution"/>
    </ClassAssertion>
```

B.1.3.21. Fischer's Exact Test

As an example of the definition of instances of the classes in OWL we have created one instance of FTest named "fisherExecution". The declaration of this instance in OWL is presented below.

```
    <Declaration>
        <NamedIndividual IRI="#fisherExecution"/>
    </Declaration>
    <ClassAssertion>
        <Class IRI="#FishersTest"/>
        <NamedIndividual IRI="#fisherExecution"/>
    </ClassAssertion>
```

B.1.3.22. Naïve Bayes

Below we present the definition in OWL of the default configuration of NaiveBayes algorithm in Weka.

```
    <Declaration>
        <NamedIndividual IRI="#NBExecution1"/>
    </Declaration>
    <ClassAssertion>
        <Class IRI="#NaiveBayes"/>
        <NamedIndividual IRI="#NBExecution1"/>
    </ClassAssertion>
    <DataPropertyAssertion>
        <DataProperty IRI="#batchSize"/>
        <NamedIndividual IRI="#NBExecution1"/>
        <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#positiveInteger">100</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
        <DataProperty IRI="#debug"/>
        <NamedIndividual IRI="#NBExecution1"/>
        <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#boolean">false</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
```

```
            <DataProperty IRI="#displayModeInOldFormat"/>
            <NamedIndividual IRI="#NBExecution1"/>
            <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#boolean">false</Literal>
        </DataPropertyAssertion>
        <DataPropertyAssertion>
            <DataProperty IRI="#doNotCheckCapabilities"/>
            <NamedIndividual IRI="#NBExecution1"/>
            <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#boolean">false</Literal>
        </DataPropertyAssertion>
        <DataPropertyAssertion>
            <DataProperty IRI="#numDecimalPlaces"/>
            <NamedIndividual IRI="#NBExecution1"/>
            <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#positiveInteger">2</Literal>
        </DataPropertyAssertion>
        <DataPropertyAssertion>
            <DataProperty IRI="#useKernelEstimator"/>
            <NamedIndividual IRI="#NBExecution1"/>
            <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#boolean">false</Literal>
        </DataPropertyAssertion>
        <DataPropertyAssertion>
            <DataProperty IRI="#useSupervisedDiscretization"/>
            <NamedIndividual IRI="#NBExecution1"/>
            <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#boolean">false</Literal>
        </DataPropertyAssertion>
```

B.1.3.23. Gaussian Processes

Below we present the definition in OWL of the default configuration of Gaussian Processes algorithm in Weka.

```
    <Declaration>
        <NamedIndividual IRI="#GPExecution1"/>
      </Declaration>
      <ClassAssertion>
```

```
      <Class IRI="#GaussianProcesses"/>
      <NamedIndividual IRI="#GPExecution1"/>
    </ClassAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#seed"/>
      <NamedIndividual IRI="#GPExecution1"/>
      <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#integer">1</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#numDecimalPlaces"/>
      <NamedIndividual IRI="#GPExecution1"/>
      <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#positiveInteger">2</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#batchSize"/>
      <NamedIndividual IRI="#GPExecution1"/>
      <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#positiveInteger">100</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#kernel"/>
      <NamedIndividual IRI="#GPExecution1"/>
      <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#string">PolyKernel</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#filterType"/>
      <NamedIndividual IRI="#GPExecution1"/>
      <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#positiveInteger">1</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#noise"/>
      <NamedIndividual IRI="#GPExecution1"/>
      <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#float">1.0</Literal>
```

```
    </DataPropertyAssertion>
```

B.1.3.24. Linear Regression

Below we present the definition in OWL of the default configuration of Linear Regression algorithm in Weka.

```
    <Declaration>
      <NamedIndividual IRI="#LRExecution1"/>
    </Declaration>
    <ClassAssertion>
      <Class IRI="#LinearRegression"/>
      <NamedIndividual IRI="#LRExecution1"/>
    </ClassAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#minimal"/>
      <NamedIndividual IRI="#LRExecution1"/>
      <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#boolean">false</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#numDecimalPlaces"/>
      <NamedIndividual IRI="#LRExecution1"/>
      <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#positiveInteger">4</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#batchSize"/>
      <NamedIndividual IRI="#LRExecution1"/>
      <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#integer">100</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#ridge"/>
      <NamedIndividual IRI="#LRExecution1"/>
      <Literal datatypeIRI="http://www.w3.org/2001/XMLSchema#string">1.0E-
8ri</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#attributeSelectionMethod"/>
      <NamedIndividual IRI="#LRExecution1"/>
```

```
      <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#positiveInteger">1</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#outputAdditionalStats"/>
      <NamedIndividual IRI="#LRExecution1"/>
      <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#boolean">false</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#estimateColinearAttributes"/>
      <NamedIndividual IRI="#LRExecution1"/>
      <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#boolean">true</Literal>
    </DataPropertyAssertion>
```

B.1.3.25. <u>Multinomial Logistic Regression</u>

Below we present the definition in OWL of the default configuration of MultinomialLogisticRegression algorithm in Weka.

```
      <Declaration>
      <NamedIndividual IRI="#MLRExecution1"/>
    </Declaration>
    <ClassAssertion>
      <Class IRI="#MultinomialLogisticRegression"/>
      <NamedIndividual IRI="#MLRExecution1"/>
    </ClassAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#numDecimalPlaces"/>
      <NamedIndividual IRI="#MLRExecution1"/>
      <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#positiveInteger">4</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#batchSize"/>
      <NamedIndividual IRI="#MLRExecution1"/>
      <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#integer">100</Literal>
```

```
    </DataPropertyAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#ridge"/>
      <NamedIndividual IRI="#MLRExecution1"/>
      <Literal datatypeIRI="http://www.w3.org/2001/XMLSchema#string">1.0E-
8</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#useConjugateGradientDescent"/>
      <NamedIndividual IRI="#MLRExecution1"/>
      <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#boolean">false</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#maxIts"/>
      <NamedIndividual IRI="#MLRExecution1"/>
      <Literal datatypeIRI="http://www.w3.org/2001/XMLSchema#integer">-
1</Literal>
    </DataPropertyAssertion>
```

B.1.3.26. K-nearest neighbours (IBk)

Below we present the definition in OWL of the default configuration of K-nearest neighbours algorithm in Weka.

```
        <Declaration>
      <NamedIndividual IRI="#IBkExecution1"/>
    </Declaration>
    <ClassAssertion>
      <Class IRI="#IBk"/>
      <NamedIndividual IRI="#IBkExecution1"/>
    </ClassAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#numDecimalPlaces"/>
      <NamedIndividual IRI="#IBkExecution1"/>
      <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#positiveInteger">2</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
```

```
      <DataProperty IRI="#batchSize"/>
      <NamedIndividual IRI="#IBkExecution1"/>
      <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#integer">100</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#KNN"/>
      <NamedIndividual IRI="#IBkExecution1"/>
      <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#positiveInteger">1</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#distanceWeighting"/>
      <NamedIndividual IRI="#IBkExecution1"/>
      <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#positiveInteger">1</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#nearestNeighbourSearchAlgorithm"/>
      <NamedIndividual IRI="#IBkExecution1"/>
      <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#string">LinearNNSearch</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#windowSize"/>
      <NamedIndividual IRI="#IBkExecution1"/>
      <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#positiveInteger">0</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#meanSquared"/>
      <NamedIndividual IRI="#IBkExecution1"/>
      <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#boolean">false</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#crossValidate"/>
      <NamedIndividual IRI="#IBkExecution1"/>
```

```
       <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#boolean">false</Literal>
     </DataPropertyAssertion>
```

B.1.3.27. Decision Table

Below we present the definition in OWL of the default configuration of Decision Table algorithm in Weka.

```
           <Declaration>
       <NamedIndividual IRI="#DTExecution1"/>
     </Declaration>
     <ClassAssertion>
       <Class IRI="#DecisionTable"/>
       <NamedIndividual IRI="#DTExecution1"/>
     </ClassAssertion>
     <DataPropertyAssertion>
       <DataProperty IRI="#numDecimalPlaces"/>
       <NamedIndividual IRI="#DTExecution1"/>
       <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#positiveInteger">2</Literal>
     </DataPropertyAssertion>
     <DataPropertyAssertion>
       <DataProperty IRI="#batchSize"/>
       <NamedIndividual IRI="#DTExecution1"/>
       <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#integer">100</Literal>
     </DataPropertyAssertion>
     <DataPropertyAssertion>
       <DataProperty IRI="#evaluationMeasure"/>
       <NamedIndividual IRI="#DTExecution1"/>
       <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#postiveInteger">1</Literal>
     </DataPropertyAssertion>
     <DataPropertyAssertion>
       <DataProperty IRI="#search"/>
       <NamedIndividual IRI="#DTExecution1"/>
       <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#string">BestFirst</Literal>
```

```
    </DataPropertyAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#displayRules"/>
      <NamedIndividual IRI="#DTExecution1"/>
      <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#boolean">false</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#useIBk"/>
      <NamedIndividual IRI="#DTExecution1"/>
      <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#boolean">false</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#crossVal"/>
      <NamedIndividual IRI="#DTExecution1"/>
      <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#integer">1</Literal>
    </DataPropertyAssertion>
```

B.1.3.28. Zero R

Below we present the definition in OWL of the default configuration of ZeroR algorithm in Weka.

```
    <Declaration>
      <NamedIndividual IRI="#ZRExecution1"/>
    </Declaration>
    <ClassAssertion>
      <Class IRI="#ZeroR"/>
      <NamedIndividual IRI="#ZRExecution1"/>
    </ClassAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#numDecimalPlaces"/>
      <NamedIndividual IRI="#ZRExecution1"/>
      <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#positiveInteger">2</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#batchSize"/>
```

```
      <NamedIndividual IRI="#ZRExecution1"/>
      <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#integer">100</Literal>
    </DataPropertyAssertion>
```

B.1.3.29. J48

Below we present the definition in OWL of the default configuration of J48 algorithm in Weka.

```
      <Declaration>
      <NamedIndividual IRI="#J48Execution1"/>
    </Declaration>
    <ClassAssertion>
      <Class IRI="#J48"/>
      <NamedIndividual IRI="#J48Execution1"/>
    </ClassAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#seed"/>
      <NamedIndividual IRI="#J48Execution1"/>
      <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#integer">1</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#unpruned"/>
      <NamedIndividual IRI="#J48Execution1"/>
      <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#boolean">false</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#confidenceFactor"/>
      <NamedIndividual IRI="#J48Execution1"/>
      <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#float">0.25</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#numFolds"/>
      <NamedIndividual IRI="#J48Execution1"/>
      <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#positiveInteger">3</Literal>
```

```
    </DataPropertyAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#numDecimalPlaces"/>
      <NamedIndividual IRI="#J48Execution1"/>
      <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#positiveInteger">2</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#batchSize"/>
      <NamedIndividual IRI="#J48Execution1"/>
      <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#integer">100</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#reducedErrorPruning"/>
      <NamedIndividual IRI="#J48Execution1"/>
      <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#boolean">false</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#useLaplace"/>
      <NamedIndividual IRI="#J48Execution1"/>
      <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#boolean">false</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#doNotMakeSplitPointActualValue"/>
      <NamedIndividual IRI="#J48Execution1"/>
      <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#boolean">false</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#subtreeRaising"/>
      <NamedIndividual IRI="#J48Execution1"/>
      <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#boolean">true</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
```

```
        <DataProperty IRI="#saveInstanceData"/>
        <NamedIndividual IRI="#J48Execution1"/>
        <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#boolean">false</Literal>
      </DataPropertyAssertion>
      <DataPropertyAssertion>
        <DataProperty IRI="#binarySplits"/>
        <NamedIndividual IRI="#J48Execution1"/>
        <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#boolean">false</Literal>
      </DataPropertyAssertion>
      <DataPropertyAssertion>
        <DataProperty IRI="#minNumObj"/>
        <NamedIndividual IRI="#J48Execution1"/>
        <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#integer">2</Literal>
      </DataPropertyAssertion>
      <DataPropertyAssertion>
        <DataProperty IRI="#useMDLcorrection"/>
        <NamedIndividual IRI="#J48Execution1"/>
        <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#boolean">true</Literal>
      </DataPropertyAssertion>
      <DataPropertyAssertion>
        <DataProperty IRI="#collapseTree"/>
        <NamedIndividual IRI="#J48Execution1"/>
        <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#boolean">true</Literal>
      </DataPropertyAssertion>
```

B.1.3.30. __Random Forest__

Below we present the definition in OWL of the default configuration of Random Forest algorithm in Weka.

```
          <Declaration>
        <NamedIndividual IRI="#RFExecution1"/>
      </Declaration>
      <ClassAssertion>
```

```
      <Class IRI="#RandomForest"/>
      <NamedIndividual IRI="#RFExecution1"/>
    </ClassAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#seed"/>
      <NamedIndividual IRI="#RFExecution1"/>
      <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#integer">1</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#representCopiesUsingWeights"/>
      <NamedIndividual IRI="#RFExecution1"/>
      <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#boolean">false</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#storeOutOfBagPredictions"/>
      <NamedIndividual IRI="#RFExecution1"/>
      <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#boolean">false</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#numExecutionSlots"/>
      <NamedIndividual IRI="#RFExecution1"/>
      <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#positiveInteger">1</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#bagSizePercent"/>
      <NamedIndividual IRI="#RFExecution1"/>
      <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#positiveInteger">100</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#batchSize"/>
      <NamedIndividual IRI="#RFExecution1"/>
      <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#integer">100</Literal>
```

```
    </DataPropertyAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#printClassifiers"/>
      <NamedIndividual IRI="#RFExecution1"/>
      <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#boolean">false</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#numIterations"/>
      <NamedIndividual IRI="#RFExecution1"/>
      <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#positiveInteger">100</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#outputOutOfBagComplexityStatistics"/>
      <NamedIndividual IRI="#RFExecution1"/>
      <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#boolean">false</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#classifier"/>
      <NamedIndividual IRI="#RFExecution1"/>
      <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#string">DefaultClassifier</Litera
l>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#breakTiesRandomly"/>
      <NamedIndividual IRI="#RFExecution1"/>
      <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#boolean">false</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#maxDepth"/>
      <NamedIndividual IRI="#RFExecution1"/>
      <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#positiveInteger">0</Literal>
    </DataPropertyAssertion>
```

```
    <DataPropertyAssertion>
      <DataProperty IRI="#computeAttributeImportance"/>
      <NamedIndividual IRI="#RFExecution1"/>
      <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#boolean">false</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#calcOutOfBag"/>
      <NamedIndividual IRI="#RFExecution1"/>
      <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#boolean">false</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#numFeatures"/>
      <NamedIndividual IRI="#RFExecution1"/>
      <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#positiveInteger">0</Literal>
    </DataPropertyAssertion>
```

B.1.3.31. <u>Random Tree</u>

Below we present the definition in OWL of the default configuration of Random Tree algorithm in Weka.

```
        <Declaration>
      <NamedIndividual IRI="#RTExecution1"/>
    </Declaration>
    <ClassAssertion>
      <Class IRI="#RandomTree"/>
      <NamedIndividual IRI="#RTExecution1"/>
    </ClassAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#seed"/>
      <NamedIndividual IRI="#RTExecution1"/>
      <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#integer">1</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#allowUnclassifiedInstances"/>
```

```
      <NamedIndividual IRI="#RTExecution1"/>
      <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#boolean">false</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#minNums"/>
      <NamedIndividual IRI="#RTExecution1"/>
      <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#float">1.0</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#numFolds"/>
      <NamedIndividual IRI="#RTExecution1"/>
      <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#positiveInteger">0</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#numDecimalPlaces"/>
      <NamedIndividual IRI="#RTExecution1"/>
      <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#positiveInteger">2</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#batchSize"/>
      <NamedIndividual IRI="#RTExecution1"/>
      <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#integer">100</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#breakTiesRandomly"/>
      <NamedIndividual IRI="#RTExecution1"/>
      <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#boolean">false</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#maxDepth"/>
      <NamedIndividual IRI="#RTExecution1"/>
```

```
      <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#positiveInteger">0</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#minVarianceProp"/>
      <NamedIndividual IRI="#RTExecution1"/>
      <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#float">0.001</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#KValue"/>
      <NamedIndividual IRI="#RTExecution1"/>
      <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#integer">0</Literal>
    </DataPropertyAssertion>
```

B.1.3.32. Canopy

Below we present the definition in OWL of the default configuration of Random Tree algorithm in Weka.

```
    <Declaration>
      <NamedIndividual IRI="#CExecution1"/>
    </Declaration>
    <ClassAssertion>
      <Class IRI="#Canopy"/>
      <NamedIndividual IRI="#CExecution1"/>
    </ClassAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#seed"/>
      <NamedIndividual IRI="#CExecution1"/>
      <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#integer">1</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#dontReplaceMissingValues"/>
      <NamedIndividual IRI="#CExecution1"/>
      <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#boolean">false</Literal>
```

```
    </DataPropertyAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#t2"/>
      <NamedIndividual IRI="#CExecution1"/>
      <Literal datatypeIRI="http://www.w3.org/2001/XMLSchema#float">-
1.0</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#t1"/>
      <NamedIndividual IRI="#CExecution1"/>
      <Literal datatypeIRI="http://www.w3.org/2001/XMLSchema#float">-
1.25</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#numClusters"/>
      <NamedIndividual IRI="#CExecution1"/>
      <Literal datatypeIRI="http://www.w3.org/2001/XMLSchema#integer">-
1</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#minimumCanopyDensity"/>
      <NamedIndividual IRI="#CExecution1"/>
      <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#float">2.0</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#periodicPruningRate"/>
      <NamedIndividual IRI="#CExecution1"/>
      <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#integer">10000</Literal>
    </DataPropertyAssertion>
```

B.1.3.33. Cobweb and Classit

Below we present the definition in OWL of the default configuration of Coweb-Classit algorithm in Weka.

```
      <Declaration>
      <NamedIndividual IRI="#CCExecution1"/>
```

```
    </Declaration>
    <ClassAssertion>
      <Class IRI="#CowebClassit"/>
      <NamedIndividual IRI="#CCExecution1"/>
    </ClassAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#seed"/>
      <NamedIndividual IRI="#CCExecution1"/>
      <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#integer">42</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#saveInstanceData"/>
      <NamedIndividual IRI="#CCExecution1"/>
      <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#boolean">false</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#acuity"/>
      <NamedIndividual IRI="#CCExecution1"/>
      <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#float">1.0</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#cutoff"/>
      <NamedIndividual IRI="#CCExecution1"/>
      <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#double">0.0028209479177387815</Li
teral>
    </DataPropertyAssertion>
```

B.1.3.34. EM

Below we present the definition in OWL of the default configuration of EM (expectation maximisation) algorithm in Weka.

```
    <Declaration>
      <NamedIndividual IRI="#EMExecution1"/>
    </Declaration>
```

```
    <ClassAssertion>
      <Class IRI="#EM"/>
      <NamedIndividual IRI="#EMExecution1"/>
    </ClassAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#seed"/>
      <NamedIndividual IRI="#EMExecution1"/>
      <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#integer">100</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#numFolds"/>
      <NamedIndividual IRI="#EMExecution1"/>
      <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#integer">10</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#numExecutionSlots"/>
      <NamedIndividual IRI="#EMExecution1"/>
      <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#integer">1</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#numKMeansRuns"/>
      <NamedIndividual IRI="#EMExecution1"/>
      <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#integer">10</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#displayModeInOldFormat"/>
      <NamedIndividual IRI="#EMExecution1"/>
      <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#boolean">false</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#minLogLikelihoodImprovementIterating"/>
      <NamedIndividual IRI="#EMExecution1"/>
```

```
      <Literal datatypeIRI="http://www.w3.org/2001/XMLSchema#string">1.0E-
6</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#minLogLikelihoodImprovementCV"/>
      <NamedIndividual IRI="#EMExecution1"/>
      <Literal datatypeIRI="http://www.w3.org/2001/XMLSchema#string">1.0E-
6</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#maximumNumberOfClusters"/>
      <NamedIndividual IRI="#EMExecution1"/>
      <Literal datatypeIRI="http://www.w3.org/2001/XMLSchema#integer">-
1</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#numClusters"/>
      <NamedIndividual IRI="#EMExecution1"/>
      <Literal datatypeIRI="http://www.w3.org/2001/XMLSchema#integer">-
1</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#maxIterations"/>
      <NamedIndividual IRI="#EMExecution1"/>
      <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#integer">100</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#minStdDev"/>
      <NamedIndividual IRI="#EMExecution1"/>
      <Literal datatypeIRI="http://www.w3.org/2001/XMLSchema#string">1st.0E-
6</Literal>
    </DataPropertyAssertion>
```

B.1.3.35. <u>Farthest First</u>

Below we present the definition in OWL of the default configuration of FarthestFirst algorithm in Weka.

```
      <Declaration>
      <NamedIndividual IRI="#FFExecution1"/>
    </Declaration>
    <ClassAssertion>
      <Class IRI="#FarthestFirst"/>
      <NamedIndividual IRI="#FFExecution1"/>
    </ClassAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#seed"/>
      <NamedIndividual IRI="#FFExecution1"/>
      <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#integer">1</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#numClusters"/>
      <NamedIndividual IRI="#FFExecution1"/>
      <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#integer">2</Literal>
    </DataPropertyAssertion>
```

B.1.3.36. Simple K Means

Below we present the definition in OWL of the default configuration of Simple K Means algorithm in Weka.

```
        <Declaration>
      <NamedIndividual IRI="#SKMExecution1"/>
    </Declaration>
    <ClassAssertion>
      <Class IRI="#SimpleKMeans"/>
      <NamedIndividual IRI="#SKMExecution1"/>
    </ClassAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#seed"/>
      <NamedIndividual IRI="#SKMExecution1"/>
      <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#integer">10</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
```

```
      <DataProperty IRI="#displayStdDevs"/>
      <NamedIndividual IRI="#SKMExecution1"/>
      <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#boolean">false</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#numExecutionSlots"/>
      <NamedIndividual IRI="#SKMExecution1"/>
      <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#integer">1</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#dontReplaceMissingValues"/>
      <NamedIndividual IRI="#SKMExecution1"/>
      <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#boolean">false</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#canopyMinimumCanopyDensity"/>
      <NamedIndividual IRI="#SKMExecution1"/>
      <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#float">2.0</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#canopyT2"/>
      <NamedIndividual IRI="#SKMExecution1"/>
      <Literal datatypeIRI="http://www.w3.org/2001/XMLSchema#float">-
1.0</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#numClusters"/>
      <NamedIndividual IRI="#SKMExecution1"/>
      <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#integer">2</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#maxIterations"/>
      <NamedIndividual IRI="#SKMExecution1"/>
```

```
      <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#integer">500</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#preserveInstancesOrder"/>
      <NamedIndividual IRI="#SKMExecution1"/>
      <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#boolean">false</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#canopyPeriodicPruningRate"/>
      <NamedIndividual IRI="#SKMExecution1"/>
      <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#10000">10000</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#canopyMaxNumCanopiesToHoldInMemory"/>
      <NamedIndividual IRI="#SKMExecution1"/>
      <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#integer">100</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#initializationMethod"/>
      <NamedIndividual IRI="#SKMExecution1"/>
      <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#string">Canopy</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#distanceFunction"/>
      <NamedIndividual IRI="#SKMExecution1"/>
      <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#string">EuclideanDistance</Litera
l>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#canopyT1"/>
      <NamedIndividual IRI="#SKMExecution1"/>
```

```
      <Literal datatypeIRI="http://www.w3.org/2001/XMLSchema#float">-
1.25</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#fastDistanceCalc"/>
      <NamedIndividual IRI="#SKMExecution1"/>
      <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#boolean">false</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
      <DataProperty IRI="#reduceNumberOfDistanceCalcsViaCanopies"/>
      <NamedIndividual IRI="#SKMExecution1"/>
      <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#boolean">false</Literal>
    </DataPropertyAssertion>
```

B.2. Complete OWL XML Definition of the Language

Below we present the complete definition of the Public Health Policy Decision Making Modeling Language presented in Chapter 3 in OWL XML.

```
  <?xml version="1.0"?>
  <Ontology xmlns="http://www.w3.org/2002/07/owl#"
       xml:base="http://www.semanticweb.org/mprasinos/ontologies/2017/8/evotion-
ontology"
       xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
       xmlns:xml="http://www.w3.org/XML/1998/namespace"
       xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
       xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
       ontologyIRI="http://www.semanticweb.org/mprasinos/ontologies/2017/8/evotion-
ontology">
      <Prefix                                                              name=""
IRI="http://www.semanticweb.org/mprasinos/ontologies/2017/8/evotion-ontology#"/>
      <Prefix name="owl" IRI="http://www.w3.org/2002/07/owl#"/>
      <Prefix name="rdf" IRI="http://www.w3.org/1999/02/22-rdf-syntax-ns#"/>
      <Prefix name="xml" IRI="http://www.w3.org/XML/1998/namespace"/>
      <Prefix name="xsd" IRI="http://www.w3.org/2001/XMLSchema#"/>
      <Prefix name="rdfs" IRI="http://www.w3.org/2000/01/rdf-schema#"/>
      <Annotation>
          <AnnotationProperty IRI="#Contributor"/>
          <Literal            datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">George Spanoudakis</Literal>
      </Annotation>
      <Annotation>
          <AnnotationProperty IRI="#Creator"/>
          <Literal            datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">Marios Prasinos</Literal>
      </Annotation>
      <Annotation>
          <AnnotationProperty IRI="#Title"/>
          <Literal datatypeIRI="http://www.w3.org/2001/XMLSchema#string">Ontology of
Evidence Based Policy Making in
              Healthcare
          </Literal>
      </Annotation>
```

```
        <Declaration>
            <ObjectProperty IRI="#input"/>
        </Declaration>
        <Declaration>
            <DataProperty IRI="#kernel"/>
        </Declaration>
        <Declaration>
            <DataProperty IRI="#numKMeansRuns"/>
        </Declaration>
        <Declaration>
            <Class IRI="#DataProcessingTask"/>
        </Declaration>
        <Declaration>
            <DataProperty IRI="#canopyMaxNumCanopiesToHoldInMemory"/>
        </Declaration>
        <Declaration>
            <DataProperty IRI="#useKernelEstimator"/>
        </Declaration>
        <Declaration>
            <DataProperty IRI="#useLaplace"/>
        </Declaration>
        <Declaration>
            <Class IRI="#CowebClassit"/>
        </Declaration>
        <Declaration>
            <ObjectProperty IRI="#executionType"/>
        </Declaration>
        <Declaration>
            <DataProperty IRI="#tableName"/>
        </Declaration>
        <Declaration>
            <ObjectProperty IRI="#a1"/>
        </Declaration>
        <Declaration>
            <DataProperty IRI="#periodicPruningRate"/>
        </Declaration>
        <Declaration>
            <AnnotationProperty IRI="#Creator"/>
```

```
        </Declaration>
        <Declaration>
            <Class IRI="#TimedDataset"/>
        </Declaration>
        <Declaration>
            <DataProperty IRI="#minStdDev"/>
        </Declaration>
        <Declaration>
            <DataProperty IRI="#cutoff"/>
        </Declaration>
        <Declaration>
            <ObjectProperty IRI="#criterion"/>
        </Declaration>
        <Declaration>
            <DataProperty IRI="#canopyMinimumCanopyDensity"/>
        </Declaration>
        <Declaration>
            <Class IRI="#Algorithm"/>
        </Declaration>
        <Declaration>
            <Class IRI="#JoinOperation"/>
        </Declaration>
        <Declaration>
            <DataProperty IRI="#ridge"/>
        </Declaration>
        <Declaration>
            <Class IRI="#DataMiningAlgorithm"/>
        </Declaration>
        <Declaration>
            <ObjectProperty IRI="#a2"/>
        </Declaration>
        <Declaration>
            <Class IRI="#ANOVA"/>
        </Declaration>
        <Declaration>
            <Class IRI="#RandomSampling"/>
        </Declaration>
        <Declaration>
```

```
        <Class IRI="#Dataset"/>
    </Declaration>
    <Declaration>
        <Class IRI="#AdaptiveSampling"/>
    </Declaration>
    <Declaration>
        <ObjectProperty IRI="#position"/>
    </Declaration>
    <Declaration>
        <ObjectProperty IRI="#dataset"/>
    </Declaration>
    <Declaration>
        <Class IRI="#DataMiningTask"/>
    </Declaration>
    <Declaration>
        <ObjectProperty IRI="#advocates"/>
    </Declaration>
    <Declaration>
        <ObjectProperty IRI="#periodicity"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="#useMDLcorrection"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="#minLogLikelihoodImprovementIterating"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="#windowSize"/>
    </Declaration>
    <Declaration>
        <Class IRI="#SupervisedDataMiningAlgorithm"/>
    </Declaration>
    <Declaration>
        <AnnotationProperty IRI="#Title"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="#bagSizePercent"/>
    </Declaration>
```

```
    <Declaration>
        <DataProperty IRI="#noise"/>
    </Declaration>
    <Declaration>
        <Class IRI="#Constraint"/>
    </Declaration>
    <Declaration>
        <Class IRI="#DataStream"/>
    </Declaration>
    <Declaration>
        <Class IRI="#Alternative"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="#unpruned"/>
    </Declaration>
    <Declaration>
        <Class IRI="#EM"/>
    </Declaration>
    <Declaration>
        <Class IRI="#Project_Operation"/>
    </Declaration>
    <Declaration>
        <Class IRI="#Stakeholder"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="#canopyT1"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="#executionEnd"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="#distanceWeighting"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="#columnName"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="#outputOutOfBagComplexityStatistics"/>
```

```
    </Declaration>
    <Declaration>
        <ObjectProperty IRI="#involvesStakeholders"/>
    </Declaration>
    <Declaration>
        <Class IRI="#Breusch-Pagan_Test"/>
    </Declaration>
    <Declaration>
        <Class IRI="#Position"/>
    </Declaration>
    <Declaration>
        <Class IRI="#DataAnalyticsTask"/>
    </Declaration>
    <Declaration>
        <Class IRI="#DataSpecification"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="#confidenceFactor"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="#fastDistanceCalc"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="#canopyT2"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="#numFeatures"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="#breakTiesRandomly"/>
    </Declaration>
    <Declaration>
        <Class IRI="#PeriodicExecution"/>
    </Declaration>
    <Declaration>
        <Datatype IRI="http://www.w3.org/2001/XMLSchema#10000"/>
    </Declaration>
    <Declaration>
```

```
        <Class IRI="#One_way_ANOVA"/>
    </Declaration>
    <Declaration>
        <ObjectProperty IRI="#isComposedOf"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="#algorithmConfigurationOptions"/>
    </Declaration>
    <Declaration>
        <Class IRI="#Operation"/>
    </Declaration>
    <Declaration>
        <Class IRI="#DecisionTable"/>
    </Declaration>
    <Declaration>
        <Class IRI="#WorkflowExecutionType"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="#crossVal"/>
    </Declaration>
    <Declaration>
        <ObjectProperty IRI="#policyModel"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="#fixedPoint"/>
    </Declaration>
    <Declaration>
        <ObjectProperty IRI="#isEvaluatedBy"/>
    </Declaration>
    <Declaration>
        <Class IRI="#Repeated_Measure_ANOVA"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="#size"/>
    </Declaration>
    <Declaration>
        <Class IRI="#FarthestFirst"/>
    </Declaration>
```

```
    <Declaration>
        <Class IRI="#DataAnalyticsModel"/>
    </Declaration>
    <Declaration>
        <Class IRI="#FilterOperation"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="#search"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="#description"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="#numDecimalPlaces"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="#useIBk"/>
    </Declaration>
    <Declaration>
        <Class IRI="#Two_Way_ANOVA"/>
    </Declaration>
    <Declaration>
        <ObjectProperty IRI="#has_workflows"/>
    </Declaration>
    <Declaration>
        <ObjectProperty IRI="#dataAnalyticsTask"/>
    </Declaration>
    <Declaration>
        <Class IRI="#Period"/>
    </Declaration>
    <Declaration>
        <Class IRI="#StatisticalAnalysisAlgorithm"/>
    </Declaration>
    <Declaration>
        <Class IRI="#AbsoluteDataset"/>
    </Declaration>
    <Declaration>
        <ObjectProperty IRI="#proposes"/>
```

```
    </Declaration>
    <Declaration>
        <DataProperty IRI="#displayModeInOldFormat"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="#computeAttributeImportance"/>
    </Declaration>
    <Declaration>
        <Class IRI="#FullJoin"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="#nearestNeighbourSearchAlgorithm"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="#subtreeRaising"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="#condition"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="#cleaningRule"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="#numIterations"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="#maxIterations"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="#representCopiesUsingWeights"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="#storeOutOfBagPredictions"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="#acuity"/>
    </Declaration>
    <Declaration>
```

```
        <ObjectProperty IRI="#output"/>
    </Declaration>
    <Declaration>
        <ObjectProperty IRI="#refersTo"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="#t1"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="#canopyPeriodicPruningRate"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="#cleaningRule"/>
    </Declaration>
    <Declaration>
        <ObjectProperty IRI="#spec"/>
    </Declaration>
    <Declaration>
        <Class IRI="#OpposingPosition"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="#binarySplits"/>
    </Declaration>
    <Declaration>
        <Class IRI="#Fisher&apos;s_Exact_Test"/>
    </Declaration>
    <Declaration>
        <ObjectProperty IRI="#policyAction"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="#minLogLikelihoodImprovementCV"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="#numClusters"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="#displayStdDevs"/>
    </Declaration>
```

```
    <Declaration>
        <Class IRI="#F-test"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="#type"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="#collapseTree"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="#t2"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="#rationale"/>
    </Declaration>
    <Declaration>
        <Class IRI="#J48"/>
    </Declaration>
    <Declaration>
        <ObjectProperty IRI="#dependant"/>
    </Declaration>
    <Declaration>
        <Class IRI="#DataCleaningOperation"/>
    </Declaration>
    <Declaration>
        <ObjectProperty IRI="#canBeAddressedBy"/>
    </Declaration>
    <Declaration>
        <Class IRI="#UnsupervisedDataMiningAlgorithm"/>
    </Declaration>
    <Declaration>
        <Class IRI="#ClusteringSampling"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="#KValue"/>
    </Declaration>
    <Declaration>
        <ObjectProperty IRI="#pre-requisite"/>
```

```xml
    </Declaration>
    <Declaration>
        <ObjectProperty IRI="#dataAnalyticsWorkflow"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="#numExecutionSlots"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="#saveInstanceData"/>
    </Declaration>
    <Declaration>
        <Class IRI="#Method"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="#increaseRatio"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="#maxIts"/>
    </Declaration>
    <Declaration>
        <ObjectProperty IRI="#algorithm"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="#printClassifiers"/>
    </Declaration>
    <Declaration>
        <ObjectProperty IRI="#stakeholder"/>
    </Declaration>
    <Declaration>
        <ObjectProperty IRI="#period"/>
    </Declaration>
    <Declaration>
        <ObjectProperty IRI="#specifies"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="#minimal"/>
    </Declaration>
    <Declaration>
```

```
        <DataProperty IRI="#debug"/>
    </Declaration>
    <Declaration>
        <Class IRI="#PolicyAction"/>
    </Declaration>
    <Declaration>
        <Class IRI="#MultinomialLogisticRegression"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="#seed"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="#minVarianceProp"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="#projectRule"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="#quantity"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="#crossValidate"/>
    </Declaration>
    <Declaration>
        <Class IRI="#DataChangeDriven"/>
    </Declaration>
    <Declaration>
        <Class IRI="#Goal"/>
    </Declaration>
    <Declaration>
        <Class IRI="#OutputDataSpecification"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="#maxDepth"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="#distanceFunction"/>
    </Declaration>
```

```
        <Declaration>
            <Class IRI="#Objective"/>
        </Declaration>
        <Declaration>
            <DataProperty IRI="#minNums"/>
        </Declaration>
        <Declaration>
            <DataProperty IRI="#calcOutOfBag"/>
        </Declaration>
        <Declaration>
            <Class IRI="#ShiftingDataset"/>
        </Declaration>
        <Declaration>
            <Class IRI="#DataAnalyticsWorfklow"/>
        </Declaration>
        <Declaration>
            <Class IRI="#SamplingOperation"/>
        </Declaration>
        <Declaration>
            <DataProperty IRI="#doNotMakeSplitPointActualValue"/>
        </Declaration>
        <Declaration>
            <DataProperty IRI="#minimumCanopyDensity"/>
        </Declaration>
        <Declaration>
            <DataProperty IRI="#batchSize"/>
        </Declaration>
        <Declaration>
            <AnnotationProperty IRI="#Description"/>
        </Declaration>
        <Declaration>
            <Class IRI="#LinearRegressionforAnalysisofcontinuousdependentvariable"/>
        </Declaration>
        <Declaration>
            <DataProperty IRI="#initializationMethod"/>
        </Declaration>
        <Declaration>
            <ObjectProperty IRI="#constraint"/>
```

```
        </Declaration>
        <Declaration>
            <DataProperty IRI="#useConjugateGradientDescent"/>
        </Declaration>
        <Declaration>
            <DataProperty IRI="#reduceNumberOfDistanceCalcsViaCanopies"/>
        </Declaration>
        <Declaration>
            <Class IRI="#IBk"/>
        </Declaration>
        <Declaration>
            <DataProperty IRI="#displayRules"/>
        </Declaration>
        <Declaration>
            <Class IRI="#StratifiedRandomSampling"/>
        </Declaration>
        <Declaration>
            <Class IRI="#Multiway_ANOVA"/>
        </Declaration>
        <Declaration>
            <DataProperty IRI="#minNumObj"/>
        </Declaration>
        <Declaration>
            <ObjectProperty IRI="#dataTiming"/>
        </Declaration>
        <Declaration>
            <DataProperty IRI="#preserveInstancesOrder"/>
        </Declaration>
        <Declaration>
            <Class IRI="#LeftOuter"/>
        </Declaration>
        <Declaration>
            <DataProperty IRI="#from"/>
        </Declaration>
        <Declaration>
            <Class IRI="#InnerJoin"/>
        </Declaration>
        <Declaration>
```

```
        <Class IRI="#GaussianProcesses"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="#timeunit"/>
    </Declaration>
    <Declaration>
        <ObjectProperty IRI="#goal"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="#attributeSelectionMethod"/>
    </Declaration>
    <Declaration>
        <ObjectProperty IRI="#aimedAt"/>
    </Declaration>
    <Declaration>
        <Class IRI="#Refinement"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="#KNN"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="#dontReplaceMissingValues"/>
    </Declaration>
    <Declaration>
        <Datatype abbreviatedIRI="xsd:postiveInteger"/>
    </Declaration>
    <Declaration>
        <Class IRI="#RightOuter"/>
    </Declaration>
    <Declaration>
        <Class IRI="#PolicyModel"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="#maximumNumberOfClusters"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="#outputAdditionalStats"/>
    </Declaration>
```

```
        <Declaration>
            <Class IRI="#RandomTree"/>
        </Declaration>
        <Declaration>
            <DataProperty IRI="#evaluationMeasure"/>
        </Declaration>
        <Declaration>
            <ObjectProperty IRI="#dataSpec"/>
        </Declaration>
        <Declaration>
            <Class IRI="#NaiveBayes"/>
        </Declaration>
        <Declaration>
            <ObjectProperty IRI="#hasWorkflows"/>
        </Declaration>
        <Declaration>
            <ObjectProperty IRI="#constraints"/>
        </Declaration>
        <Declaration>
            <DataProperty IRI="#classifier"/>
        </Declaration>
        <Declaration>
            <Class IRI="#Canopy"/>
        </Declaration>
        <Declaration>
            <ObjectProperty IRI="#objective"/>
        </Declaration>
        <Declaration>
            <Class IRI="#LinearRegression"/>
        </Declaration>
        <Declaration>
            <Class IRI="#Criterion"/>
        </Declaration>
        <Declaration>
            <DataProperty IRI="#filterType"/>
        </Declaration>
        <Declaration>
            <Class IRI="#SimpleKMeans"/>
```

```
    </Declaration>
    <Declaration>
        <DataProperty IRI="#reducedErrorPruning"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="#executionStart"/>
    </Declaration>
    <Declaration>
        <Class IRI="#SupportivePosition"/>
    </Declaration>
    <Declaration>
        <Class IRI="#ZeroR"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="#useSupervisedDiscretization"/>
    </Declaration>
    <Declaration>
        <Class IRI="#NeutralPosition"/>
    </Declaration>
    <Declaration>
        <Class IRI="#RandomForest"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="#logicalExpression"/>
    </Declaration>
    <Declaration>
        <ObjectProperty IRI="#refinedInto"/>
    </Declaration>
    <Declaration>
        <Class IRI="#StaticSet"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="#weight"/>
    </Declaration>
    <Declaration>
        <Class IRI="#StatisticalAnalysisTask"/>
    </Declaration>
    <Declaration>
```

```
        <DataProperty IRI="#meanSquared"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="#numFolds"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="#dependentVariable"/>
    </Declaration>
    <Declaration>
        <ObjectProperty IRI="#utilizes"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="#allowUnclassifiedInstances"/>
    </Declaration>
    <Declaration>
        <AnnotationProperty IRI="#Contributor"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="#doNotCheckCapabilities"/>
    </Declaration>
    <Declaration>
        <Class IRI="#ExecutionUponRequest"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="#estimateColinearAttributes"/>
    </Declaration>
    <EquivalentClasses>
        <Class IRI="#Algorithm"/>
        <ObjectMinCardinality cardinality="1">
            <ObjectProperty IRI="#dataSpec"/>
            <Class IRI="#OutputDataSpecification"/>
        </ObjectMinCardinality>
    </EquivalentClasses>
    <EquivalentClasses>
        <Class IRI="#Criterion"/>
        <ObjectMinCardinality cardinality="1">
            <ObjectProperty IRI="#constraints"/>
            <Class IRI="#Dataset"/>
```

```
        </ObjectMinCardinality>
    </EquivalentClasses>
    <EquivalentClasses>
        <Class IRI="#Criterion"/>
        <ObjectExactCardinality cardinality="1">
            <ObjectProperty IRI="#policyAction"/>
            <Class IRI="#PolicyAction"/>
        </ObjectExactCardinality>
    </EquivalentClasses>
    <EquivalentClasses>
        <Class IRI="#Criterion"/>
        <ObjectExactCardinality cardinality="1">
            <ObjectProperty IRI="#specifies"/>
            <Class IRI="#DataAnalyticsWorfklow"/>
        </ObjectExactCardinality>
    </EquivalentClasses>
    <EquivalentClasses>
        <Class IRI="#DataAnalyticsTask"/>
        <ObjectMinCardinality cardinality="0">
            <ObjectProperty IRI="#dataAnalyticsWorkflow"/>
            <Class IRI="#DataAnalyticsWorfklow"/>
        </ObjectMinCardinality>
    </EquivalentClasses>
    <EquivalentClasses>
        <Class IRI="#DataAnalyticsTask"/>
        <ObjectMinCardinality cardinality="1">
            <ObjectProperty IRI="#input"/>
            <Class IRI="#Dataset"/>
        </ObjectMinCardinality>
    </EquivalentClasses>
    <EquivalentClasses>
        <Class IRI="#DataAnalyticsTask"/>
        <ObjectMinCardinality cardinality="1">
            <ObjectProperty IRI="#output"/>
            <Class IRI="#Dataset"/>
        </ObjectMinCardinality>
    </EquivalentClasses>
    <EquivalentClasses>
```

```
        <Class IRI="#DataAnalyticsTask"/>
        <ObjectExactCardinality cardinality="1">
            <ObjectProperty IRI="#utilizes"/>
            <Class IRI="#Method"/>
        </ObjectExactCardinality>
    </EquivalentClasses>
    <EquivalentClasses>
        <Class IRI="#DataAnalyticsWorfklow"/>
        <ObjectMinCardinality cardinality="0">
            <ObjectProperty IRI="#criterion"/>
            <Class IRI="#Criterion"/>
        </ObjectMinCardinality>
    </EquivalentClasses>
    <EquivalentClasses>
        <Class IRI="#DataAnalyticsWorfklow"/>
        <ObjectMinCardinality cardinality="1">
            <ObjectProperty IRI="#isComposedOf"/>
            <Class IRI="#DataAnalyticsTask"/>
        </ObjectMinCardinality>
    </EquivalentClasses>
    <EquivalentClasses>
        <Class IRI="#DataAnalyticsWorfklow"/>
        <ObjectMinCardinality cardinality="0">
            <ObjectProperty IRI="#policyModel"/>
            <Class IRI="#PolicyModel"/>
        </ObjectMinCardinality>
    </EquivalentClasses>
    <EquivalentClasses>
        <Class IRI="#DataMiningTask"/>
        <ObjectExactCardinality cardinality="1">
            <ObjectProperty IRI="#utilizes"/>
            <Class IRI="#DataMiningAlgorithm"/>
        </ObjectExactCardinality>
    </EquivalentClasses>
    <EquivalentClasses>
        <Class IRI="#DataProcessingTask"/>
        <ObjectExactCardinality cardinality="1">
            <ObjectProperty IRI="#utilizes"/>
```

```
            <Class IRI="#Operation"/>
        </ObjectExactCardinality>
    </EquivalentClasses>
    <EquivalentClasses>
        <Class IRI="#DataSpecification"/>
        <ObjectMinCardinality cardinality="1">
            <ObjectProperty IRI="#dataset"/>
            <Class IRI="#Dataset"/>
        </ObjectMinCardinality>
    </EquivalentClasses>
    <EquivalentClasses>
        <Class IRI="#Dataset"/>
        <ObjectMinCardinality cardinality="0">
            <ObjectProperty IRI="#criterion"/>
            <Class IRI="#Criterion"/>
        </ObjectMinCardinality>
    </EquivalentClasses>
    <EquivalentClasses>
        <Class IRI="#Dataset"/>
        <ObjectMinCardinality cardinality="0">
            <ObjectProperty IRI="#dataAnalyticsTask"/>
            <Class IRI="#DataAnalyticsTask"/>
        </ObjectMinCardinality>
    </EquivalentClasses>
    <EquivalentClasses>
        <Class IRI="#Dataset"/>
        <ObjectExactCardinality cardinality="1">
            <ObjectProperty IRI="#spec"/>
            <Class IRI="#DataSpecification"/>
        </ObjectExactCardinality>
    </EquivalentClasses>
    <EquivalentClasses>
        <Class IRI="#Dataset"/>
        <ObjectMaxCardinality cardinality="1">
            <ObjectProperty IRI="#dataAnalyticsTask"/>
            <Class IRI="#DataAnalyticsTask"/>
        </ObjectMaxCardinality>
    </EquivalentClasses>
```

```
        <EquivalentClasses>
            <Class IRI="#Goal"/>
            <ObjectMinCardinality cardinality="0">
                <ObjectProperty IRI="#refinedInto"/>
                <Class IRI="#Objective"/>
            </ObjectMinCardinality>
        </EquivalentClasses>
        <EquivalentClasses>
            <Class IRI="#Goal"/>
            <ObjectExactCardinality cardinality="1">
                <ObjectProperty IRI="#policyModel"/>
                <Class IRI="#PolicyModel"/>
            </ObjectExactCardinality>
        </EquivalentClasses>
        <EquivalentClasses>
            <Class IRI="#Method"/>
            <ObjectMinCardinality cardinality="0">
                <ObjectProperty IRI="#dataAnalyticsTask"/>
                <Class IRI="#DataAnalyticsTask"/>
            </ObjectMinCardinality>
        </EquivalentClasses>
        <EquivalentClasses>
            <Class IRI="#Objective"/>
            <ObjectMinCardinality cardinality="1">
                <ObjectProperty IRI="#canBeAddressedBy"/>
                <Class IRI="#PolicyAction"/>
            </ObjectMinCardinality>
        </EquivalentClasses>
        <EquivalentClasses>
            <Class IRI="#Objective"/>
            <ObjectExactCardinality cardinality="1">
                <ObjectProperty IRI="#goal"/>
                <Class IRI="#Goal"/>
            </ObjectExactCardinality>
        </EquivalentClasses>
        <EquivalentClasses>
            <Class IRI="#OutputDataSpecification"/>
            <ObjectExactCardinality cardinality="1">
```

```
                <ObjectProperty IRI="#algorithm"/>
                <Class IRI="#Algorithm"/>
            </ObjectExactCardinality>
        </EquivalentClasses>
        <EquivalentClasses>
            <Class IRI="#PolicyAction"/>
            <ObjectMinCardinality cardinality="0">
                <ObjectProperty IRI="#a1"/>
                <Class IRI="#PolicyAction"/>
            </ObjectMinCardinality>
        </EquivalentClasses>
        <EquivalentClasses>
            <Class IRI="#PolicyAction"/>
            <ObjectMinCardinality cardinality="0">
                <ObjectProperty IRI="#a2"/>
                <Class IRI="#PolicyAction"/>
            </ObjectMinCardinality>
        </EquivalentClasses>
        <EquivalentClasses>
            <Class IRI="#PolicyAction"/>
            <ObjectMinCardinality cardinality="1">
                <ObjectProperty IRI="#dependant"/>
                <Class IRI="#PolicyAction"/>
            </ObjectMinCardinality>
        </EquivalentClasses>
        <EquivalentClasses>
            <Class IRI="#PolicyAction"/>
            <ObjectMinCardinality cardinality="0">
                <ObjectProperty IRI="#position"/>
                <Class IRI="#Position"/>
            </ObjectMinCardinality>
        </EquivalentClasses>
        <EquivalentClasses>
            <Class IRI="#PolicyAction"/>
            <ObjectMinCardinality cardinality="1">
                <ObjectProperty IRI="#pre-requisite"/>
                <Class IRI="#PolicyAction"/>
            </ObjectMinCardinality>
```

```
    </EquivalentClasses>
    <EquivalentClasses>
        <Class IRI="#PolicyAction"/>
        <ObjectExactCardinality cardinality="1">
            <ObjectProperty IRI="#isEvaluatedBy"/>
            <Class IRI="#Criterion"/>
        </ObjectExactCardinality>
    </EquivalentClasses>
    <EquivalentClasses>
        <Class IRI="#PolicyModel"/>
        <ObjectMinCardinality cardinality="0">
            <ObjectProperty IRI="#hasWorkflows"/>
            <Class IRI="#DataAnalyticsWorfklow"/>
        </ObjectMinCardinality>
    </EquivalentClasses>
    <EquivalentClasses>
        <Class IRI="#PolicyModel"/>
        <ObjectMinCardinality cardinality="0">
            <ObjectProperty IRI="#involvesStakeholders"/>
            <Class IRI="#Stakeholder"/>
        </ObjectMinCardinality>
    </EquivalentClasses>
    <EquivalentClasses>
        <Class IRI="#PolicyModel"/>
        <ObjectExactCardinality cardinality="1">
            <ObjectProperty IRI="#aimedAt"/>
            <Class IRI="#Goal"/>
        </ObjectExactCardinality>
    </EquivalentClasses>
    <EquivalentClasses>
        <Class IRI="#Position"/>
        <ObjectMinCardinality cardinality="0">
            <ObjectProperty IRI="#stakeholder"/>
            <Class IRI="#Stakeholder"/>
        </ObjectMinCardinality>
    </EquivalentClasses>
    <EquivalentClasses>
        <Class IRI="#Position"/>
```

```
        <ObjectMinCardinality cardinality="1">
            <ObjectProperty IRI="#stakeholder"/>
            <Class IRI="#Stakeholder"/>
        </ObjectMinCardinality>
    </EquivalentClasses>
    <EquivalentClasses>
        <Class IRI="#Position"/>
        <ObjectExactCardinality cardinality="1">
            <ObjectProperty IRI="#refersTo"/>
            <Class IRI="#PolicyAction"/>
        </ObjectExactCardinality>
    </EquivalentClasses>
    <EquivalentClasses>
        <Class IRI="#Stakeholder"/>
        <ObjectMinCardinality cardinality="0">
            <ObjectProperty IRI="#advocates"/>
            <Class IRI="#Position"/>
        </ObjectMinCardinality>
    </EquivalentClasses>
    <EquivalentClasses>
        <Class IRI="#Stakeholder"/>
        <ObjectMinCardinality cardinality="1">
            <ObjectProperty IRI="#policyModel"/>
            <Class IRI="#PolicyModel"/>
        </ObjectMinCardinality>
    </EquivalentClasses>
    <EquivalentClasses>
        <Class IRI="#Stakeholder"/>
        <ObjectMinCardinality cardinality="1">
            <ObjectProperty IRI="#proposes"/>
            <Class IRI="#Position"/>
        </ObjectMinCardinality>
    </EquivalentClasses>
    <EquivalentClasses>
        <Class IRI="#StatisticalAnalysisTask"/>
        <ObjectExactCardinality cardinality="1">
            <ObjectProperty IRI="#utilizes"/>
            <Class IRI="#StatisticalAnalysisAlgorithm"/>
```

```
        </ObjectExactCardinality>
    </EquivalentClasses>
    <SubClassOf>
        <Class IRI="#ANOVA"/>
        <Class IRI="#StatisticalAnalysisAlgorithm"/>
    </SubClassOf>
    <SubClassOf>
        <Class IRI="#AbsoluteDataset"/>
        <Class IRI="#TimedDataset"/>
    </SubClassOf>
    <SubClassOf>
        <Class IRI="#AdaptiveSampling"/>
        <Class IRI="#SamplingOperation"/>
    </SubClassOf>
    <SubClassOf>
        <Class IRI="#Algorithm"/>
        <Class IRI="#Method"/>
    </SubClassOf>
    <SubClassOf>
        <Class IRI="#Breusch-Pagan_Test"/>
        <Class IRI="#StatisticalAnalysisAlgorithm"/>
    </SubClassOf>
    <SubClassOf>
        <Class IRI="#Canopy"/>
        <Class IRI="#UnsupervisedDataMiningAlgorithm"/>
    </SubClassOf>
    <SubClassOf>
        <Class IRI="#ClusteringSampling"/>
        <Class IRI="#SamplingOperation"/>
    </SubClassOf>
    <SubClassOf>
        <Class IRI="#CowebClassit"/>
        <Class IRI="#UnsupervisedDataMiningAlgorithm"/>
    </SubClassOf>
    <SubClassOf>
        <Class IRI="#DataAnalyticsModel"/>
        <Class IRI="#Dataset"/>
    </SubClassOf>
```

```
    <SubClassOf>
        <Class IRI="#DataChangeDriven"/>
        <Class IRI="#WorkflowExecutionType"/>
    </SubClassOf>
    <SubClassOf>
        <Class IRI="#DataCleaningOperation"/>
        <Class IRI="#Operation"/>
    </SubClassOf>
    <SubClassOf>
        <Class IRI="#DataMiningAlgorithm"/>
        <Class IRI="#Algorithm"/>
    </SubClassOf>
    <SubClassOf>
        <Class IRI="#DataMiningTask"/>
        <Class IRI="#DataAnalyticsTask"/>
    </SubClassOf>
    <SubClassOf>
        <Class IRI="#DataProcessingTask"/>
        <Class IRI="#DataAnalyticsTask"/>
    </SubClassOf>
    <SubClassOf>
        <Class IRI="#DataStream"/>
        <Class IRI="#Dataset"/>
    </SubClassOf>
    <SubClassOf>
        <Class IRI="#DecisionTable"/>
        <Class IRI="#SupervisedDataMiningAlgorithm"/>
    </SubClassOf>
    <SubClassOf>
        <Class IRI="#EM"/>
        <Class IRI="#UnsupervisedDataMiningAlgorithm"/>
    </SubClassOf>
    <SubClassOf>
        <Class IRI="#ExecutionUponRequest"/>
        <Class IRI="#WorkflowExecutionType"/>
    </SubClassOf>
    <SubClassOf>
        <Class IRI="#F-test"/>
```

```
        <Class IRI="#StatisticalAnalysisAlgorithm"/>
    </SubClassOf>
    <SubClassOf>
        <Class IRI="#FarthestFirst"/>
        <Class IRI="#UnsupervisedDataMiningAlgorithm"/>
    </SubClassOf>
    <SubClassOf>
        <Class IRI="#FilterOperation"/>
        <Class IRI="#Operation"/>
    </SubClassOf>
    <SubClassOf>
        <Class IRI="#FullJoin"/>
        <Class IRI="#JoinOperation"/>
    </SubClassOf>
    <SubClassOf>
        <Class IRI="#GaussianProcesses"/>
        <Class IRI="#SupervisedDataMiningAlgorithm"/>
    </SubClassOf>
    <SubClassOf>
        <Class IRI="#IBk"/>
        <Class IRI="#SupervisedDataMiningAlgorithm"/>
    </SubClassOf>
    <SubClassOf>
        <Class IRI="#InnerJoin"/>
        <Class IRI="#JoinOperation"/>
    </SubClassOf>
    <SubClassOf>
        <Class IRI="#J48"/>
        <Class IRI="#SupervisedDataMiningAlgorithm"/>
    </SubClassOf>
    <SubClassOf>
        <Class IRI="#JoinOperation"/>
        <Class IRI="#Operation"/>
    </SubClassOf>
    <SubClassOf>
        <Class IRI="#LeftOuter"/>
        <Class IRI="#JoinOperation"/>
    </SubClassOf>
```

```
        <SubClassOf>
            <Class IRI="#LinearRegression"/>
            <Class IRI="#SupervisedDataMiningAlgorithm"/>
        </SubClassOf>
        <SubClassOf>
            <Class IRI="#LinearRegression"/>
            <Class IRI="#StatisticalAnalysisAlgorithm"/>
        </SubClassOf>
        <SubClassOf>
            <Class IRI="#MultinomialLogisticRegression"/>
            <Class IRI="#SupervisedDataMiningAlgorithm"/>
        </SubClassOf>
        <SubClassOf>
            <Class IRI="#Multiway_ANOVA"/>
            <Class IRI="#ANOVA"/>
        </SubClassOf>
        <SubClassOf>
            <Class IRI="#NaiveBayes"/>
            <Class IRI="#SupervisedDataMiningAlgorithm"/>
        </SubClassOf>
        <SubClassOf>
            <Class IRI="#NeutralPosition"/>
            <Class IRI="#Position"/>
        </SubClassOf>
        <SubClassOf>
            <Class IRI="#One_way_ANOVA"/>
            <Class IRI="#ANOVA"/>
        </SubClassOf>
        <SubClassOf>
            <Class IRI="#Operation"/>
            <Class IRI="#Method"/>
        </SubClassOf>
        <SubClassOf>
            <Class IRI="#OpposingPosition"/>
            <Class IRI="#Position"/>
        </SubClassOf>
        <SubClassOf>
            <Class IRI="#OutputDataSpecification"/>
```

```
        <Class IRI="#DataSpecification"/>
    </SubClassOf>
    <SubClassOf>
        <Class IRI="#PeriodicExecution"/>
        <Class IRI="#WorkflowExecutionType"/>
    </SubClassOf>
    <SubClassOf>
        <Class IRI="#Project_Operation"/>
        <Class IRI="#Operation"/>
    </SubClassOf>
    <SubClassOf>
        <Class IRI="#RandomForest"/>
        <Class IRI="#SupervisedDataMiningAlgorithm"/>
    </SubClassOf>
    <SubClassOf>
        <Class IRI="#RandomSampling"/>
        <Class IRI="#SamplingOperation"/>
    </SubClassOf>
    <SubClassOf>
        <Class IRI="#RandomTree"/>
        <Class IRI="#SupervisedDataMiningAlgorithm"/>
    </SubClassOf>
    <SubClassOf>
        <Class IRI="#Repeated_Measure_ANOVA"/>
        <Class IRI="#ANOVA"/>
    </SubClassOf>
    <SubClassOf>
        <Class IRI="#RightOuter"/>
        <Class IRI="#JoinOperation"/>
    </SubClassOf>
    <SubClassOf>
        <Class IRI="#SamplingOperation"/>
        <Class IRI="#Operation"/>
    </SubClassOf>
    <SubClassOf>
        <Class IRI="#ShiftingDataset"/>
        <Class IRI="#TimedDataset"/>
    </SubClassOf>
```

```
    <SubClassOf>
        <Class IRI="#SimpleKMeans"/>
        <Class IRI="#UnsupervisedDataMiningAlgorithm"/>
    </SubClassOf>
    <SubClassOf>
        <Class IRI="#StaticSet"/>
        <Class IRI="#Dataset"/>
    </SubClassOf>
    <SubClassOf>
        <Class IRI="#StatisticalAnalysisAlgorithm"/>
        <Class IRI="#Algorithm"/>
    </SubClassOf>
    <SubClassOf>
        <Class IRI="#StatisticalAnalysisTask"/>
        <Class IRI="#DataAnalyticsTask"/>
    </SubClassOf>
    <SubClassOf>
        <Class IRI="#StratifiedRandomSampling"/>
        <Class IRI="#SamplingOperation"/>
    </SubClassOf>
    <SubClassOf>
        <Class IRI="#SupervisedDataMiningAlgorithm"/>
        <Class IRI="#DataMiningAlgorithm"/>
    </SubClassOf>
    <SubClassOf>
        <Class IRI="#SupportivePosition"/>
        <Class IRI="#Position"/>
    </SubClassOf>
    <SubClassOf>
        <Class IRI="#Two_Way_ANOVA"/>
        <Class IRI="#ANOVA"/>
    </SubClassOf>
    <SubClassOf>
        <Class IRI="#UnsupervisedDataMiningAlgorithm"/>
        <Class IRI="#DataMiningAlgorithm"/>
    </SubClassOf>
    <SubClassOf>
        <Class IRI="#ZeroR"/>
```

```
            <Class IRI="#SupervisedDataMiningAlgorithm"/>
        </SubClassOf>
        <SubClassOf>
            <Class IRI="#Fisher&apos;s_Exact_Test"/>
            <Class IRI="#StatisticalAnalysisAlgorithm"/>
        </SubClassOf>
        <DisjointClasses>
            <Class IRI="#ANOVA"/>
            <Class IRI="#Breusch-Pagan_Test"/>
            <Class IRI="#F-test"/>
            <Class IRI="#LinearRegression"/>
            <Class IRI="#Fisher&apos;s_Exact_Test"/>
        </DisjointClasses>
        <DisjointClasses>
            <Class IRI="#AbsoluteDataset"/>
            <Class IRI="#ShiftingDataset"/>
        </DisjointClasses>
        <DisjointClasses>
            <Class IRI="#AdaptiveSampling"/>
            <Class IRI="#ClusteringSampling"/>
            <Class IRI="#RandomSampling"/>
            <Class IRI="#StratifiedRandomSampling"/>
        </DisjointClasses>
        <DisjointClasses>
            <Class IRI="#Algorithm"/>
            <Class IRI="#Operation"/>
        </DisjointClasses>
        <DisjointClasses>
            <Class IRI="#Canopy"/>
            <Class IRI="#CowebClassit"/>
            <Class IRI="#EM"/>
            <Class IRI="#FarthestFirst"/>
            <Class IRI="#SimpleKMeans"/>
        </DisjointClasses>
        <DisjointClasses>
            <Class IRI="#DataAnalyticsModel"/>
            <Class IRI="#DataStream"/>
            <Class IRI="#StaticSet"/>
```

```
        </DisjointClasses>
        <DisjointClasses>
            <Class IRI="#DataChangeDriven"/>
            <Class IRI="#ExecutionUponRequest"/>
            <Class IRI="#PeriodicExecution"/>
        </DisjointClasses>
        <DisjointClasses>
            <Class IRI="#DataCleaningOperation"/>
            <Class IRI="#FilterOperation"/>
            <Class IRI="#JoinOperation"/>
            <Class IRI="#Project_Operation"/>
            <Class IRI="#SamplingOperation"/>
        </DisjointClasses>
        <DisjointClasses>
            <Class IRI="#DataMiningAlgorithm"/>
            <Class IRI="#StatisticalAnalysisAlgorithm"/>
        </DisjointClasses>
        <DisjointClasses>
            <Class IRI="#DataMiningTask"/>
            <Class IRI="#DataProcessingTask"/>
            <Class IRI="#StatisticalAnalysisTask"/>
        </DisjointClasses>
        <DisjointClasses>
            <Class IRI="#DecisionTable"/>
            <Class IRI="#GaussianProcesses"/>
            <Class IRI="#IBk"/>
            <Class IRI="#J48"/>
            <Class IRI="#LinearRegression"/>
            <Class IRI="#MultinomialLogisticRegression"/>
            <Class IRI="#NaiveBayes"/>
            <Class IRI="#RandomForest"/>
            <Class IRI="#RandomTree"/>
            <Class IRI="#ZeroR"/>
        </DisjointClasses>
        <DisjointClasses>
            <Class IRI="#FullJoin"/>
            <Class IRI="#InnerJoin"/>
            <Class IRI="#LeftOuter"/>
```

```
        <Class IRI="#RightOuter"/>
    </DisjointClasses>
    <DisjointClasses>
        <Class IRI="#NeutralPosition"/>
        <Class IRI="#OpposingPosition"/>
        <Class IRI="#SupportivePosition"/>
    </DisjointClasses>
    <DisjointClasses>
        <Class IRI="#SupervisedDataMiningAlgorithm"/>
        <Class IRI="#UnsupervisedDataMiningAlgorithm"/>
    </DisjointClasses>
    <InverseObjectProperties>
        <ObjectProperty IRI="#a1"/>
        <ObjectProperty IRI="#a2"/>
    </InverseObjectProperties>
    <InverseObjectProperties>
        <ObjectProperty IRI="#advocates"/>
        <ObjectProperty IRI="#stakeholder"/>
    </InverseObjectProperties>
    <InverseObjectProperties>
        <ObjectProperty IRI="#aimedAt"/>
        <ObjectProperty IRI="#policyModel"/>
    </InverseObjectProperties>
    <InverseObjectProperties>
        <ObjectProperty IRI="#algorithm"/>
        <ObjectProperty IRI="#dataSpec"/>
    </InverseObjectProperties>
    <InverseObjectProperties>
        <ObjectProperty IRI="#canBeAddressedBy"/>
        <ObjectProperty IRI="#objective"/>
    </InverseObjectProperties>
    <InverseObjectProperties>
        <ObjectProperty IRI="#constraints"/>
        <ObjectProperty IRI="#criterion"/>
    </InverseObjectProperties>
    <InverseObjectProperties>
        <ObjectProperty IRI="#criterion"/>
        <ObjectProperty IRI="#specifies"/>
```

```
        </InverseObjectProperties>
        <InverseObjectProperties>
            <ObjectProperty IRI="#dataAnalyticsTask"/>
            <ObjectProperty IRI="#input"/>
        </InverseObjectProperties>
        <InverseObjectProperties>
            <ObjectProperty IRI="#dataAnalyticsTask"/>
            <ObjectProperty IRI="#output"/>
        </InverseObjectProperties>
        <InverseObjectProperties>
            <ObjectProperty IRI="#dataAnalyticsTask"/>
            <ObjectProperty IRI="#utilizes"/>
        </InverseObjectProperties>
        <InverseObjectProperties>
            <ObjectProperty IRI="#dataAnalyticsWorkflow"/>
            <ObjectProperty IRI="#isComposedOf"/>
        </InverseObjectProperties>
        <InverseObjectProperties>
            <ObjectProperty IRI="#dataset"/>
            <ObjectProperty IRI="#spec"/>
        </InverseObjectProperties>
        <InverseObjectProperties>
            <ObjectProperty IRI="#dependant"/>
            <ObjectProperty IRI="#pre-requisite"/>
        </InverseObjectProperties>
        <InverseObjectProperties>
            <ObjectProperty IRI="#goal"/>
            <ObjectProperty IRI="#refinedInto"/>
        </InverseObjectProperties>
        <InverseObjectProperties>
            <ObjectProperty IRI="#hasWorkflows"/>
            <ObjectProperty IRI="#policyModel"/>
        </InverseObjectProperties>
        <InverseObjectProperties>
            <ObjectProperty IRI="#involvesStakeholders"/>
            <ObjectProperty IRI="#policyModel"/>
        </InverseObjectProperties>
        <InverseObjectProperties>
```

```
        <ObjectProperty IRI="#isEvaluatedBy"/>
        <ObjectProperty IRI="#policyAction"/>
    </InverseObjectProperties>
    <InverseObjectProperties>
        <ObjectProperty IRI="#position"/>
        <ObjectProperty IRI="#refersTo"/>
    </InverseObjectProperties>
    <InverseObjectProperties>
        <ObjectProperty IRI="#proposes"/>
        <ObjectProperty IRI="#stakeholder"/>
    </InverseObjectProperties>
    <SymmetricObjectProperty>
        <ObjectProperty IRI="#aimedAt"/>
    </SymmetricObjectProperty>
    <ObjectPropertyDomain>
        <ObjectProperty IRI="#a1"/>
        <Class IRI="#PolicyAction"/>
    </ObjectPropertyDomain>
    <ObjectPropertyDomain>
        <ObjectProperty IRI="#a2"/>
        <Class IRI="#PolicyAction"/>
    </ObjectPropertyDomain>
    <ObjectPropertyDomain>
        <ObjectProperty IRI="#advocates"/>
        <Class IRI="#Stakeholder"/>
    </ObjectPropertyDomain>
    <ObjectPropertyDomain>
        <ObjectProperty IRI="#aimedAt"/>
        <Class IRI="#PolicyModel"/>
    </ObjectPropertyDomain>
    <ObjectPropertyDomain>
        <ObjectProperty IRI="#algorithm"/>
        <Class IRI="#OutputDataSpecification"/>
    </ObjectPropertyDomain>
    <ObjectPropertyDomain>
        <ObjectProperty IRI="#canBeAddressedBy"/>
        <Class IRI="#Objective"/>
    </ObjectPropertyDomain>
```

```
    <ObjectPropertyDomain>
        <ObjectProperty IRI="#constraint"/>
        <Class IRI="#Alternative"/>
    </ObjectPropertyDomain>
    <ObjectPropertyDomain>
        <ObjectProperty IRI="#constraints"/>
        <Class IRI="#Criterion"/>
    </ObjectPropertyDomain>
    <ObjectPropertyDomain>
        <ObjectProperty IRI="#criterion"/>
        <Class IRI="#DataAnalyticsWorfklow"/>
    </ObjectPropertyDomain>
    <ObjectPropertyDomain>
        <ObjectProperty IRI="#criterion"/>
        <Class IRI="#Dataset"/>
    </ObjectPropertyDomain>
    <ObjectPropertyDomain>
        <ObjectProperty IRI="#dataAnalyticsTask"/>
        <Class IRI="#Dataset"/>
    </ObjectPropertyDomain>
    <ObjectPropertyDomain>
        <ObjectProperty IRI="#dataAnalyticsTask"/>
        <Class IRI="#Method"/>
    </ObjectPropertyDomain>
    <ObjectPropertyDomain>
        <ObjectProperty IRI="#dataAnalyticsWorkflow"/>
        <Class IRI="#DataAnalyticsTask"/>
    </ObjectPropertyDomain>
    <ObjectPropertyDomain>
        <ObjectProperty IRI="#dataSpec"/>
        <Class IRI="#Algorithm"/>
    </ObjectPropertyDomain>
    <ObjectPropertyDomain>
        <ObjectProperty IRI="#dataTiming"/>
        <Class IRI="#PeriodicExecution"/>
    </ObjectPropertyDomain>
    <ObjectPropertyDomain>
        <ObjectProperty IRI="#dataset"/>
```

```
        <Class IRI="#DataSpecification"/>
    </ObjectPropertyDomain>
    <ObjectPropertyDomain>
        <ObjectProperty IRI="#dependant"/>
        <Class IRI="#PolicyAction"/>
    </ObjectPropertyDomain>
    <ObjectPropertyDomain>
        <ObjectProperty IRI="#executionType"/>
        <Class IRI="#DataAnalyticsWorfklow"/>
    </ObjectPropertyDomain>
    <ObjectPropertyDomain>
        <ObjectProperty IRI="#goal"/>
        <Class IRI="#Objective"/>
    </ObjectPropertyDomain>
    <ObjectPropertyDomain>
        <ObjectProperty IRI="#hasWorkflows"/>
        <Class IRI="#PolicyModel"/>
    </ObjectPropertyDomain>
    <ObjectPropertyDomain>
        <ObjectProperty IRI="#involvesStakeholders"/>
        <Class IRI="#PolicyModel"/>
    </ObjectPropertyDomain>
    <ObjectPropertyDomain>
        <ObjectProperty IRI="#isComposedOf"/>
        <Class IRI="#DataAnalyticsWorfklow"/>
    </ObjectPropertyDomain>
    <ObjectPropertyDomain>
        <ObjectProperty IRI="#isEvaluatedBy"/>
        <Class IRI="#PolicyAction"/>
    </ObjectPropertyDomain>
    <ObjectPropertyDomain>
        <ObjectProperty IRI="#objective"/>
        <Class IRI="#PolicyAction"/>
    </ObjectPropertyDomain>
    <ObjectPropertyDomain>
        <ObjectProperty IRI="#period"/>
        <Class IRI="#ShiftingDataset"/>
    </ObjectPropertyDomain>
```

```
        <ObjectPropertyDomain>
            <ObjectProperty IRI="#periodicity"/>
            <Class IRI="#PeriodicExecution"/>
        </ObjectPropertyDomain>
        <ObjectPropertyDomain>
            <ObjectProperty IRI="#policyAction"/>
            <Class IRI="#Criterion"/>
        </ObjectPropertyDomain>
        <ObjectPropertyDomain>
            <ObjectProperty IRI="#policyModel"/>
            <Class IRI="#DataAnalyticsWorfklow"/>
        </ObjectPropertyDomain>
        <ObjectPropertyDomain>
            <ObjectProperty IRI="#policyModel"/>
            <Class IRI="#Goal"/>
        </ObjectPropertyDomain>
        <ObjectPropertyDomain>
            <ObjectProperty IRI="#policyModel"/>
            <Class IRI="#Stakeholder"/>
        </ObjectPropertyDomain>
        <ObjectPropertyDomain>
            <ObjectProperty IRI="#position"/>
            <Class IRI="#PolicyAction"/>
        </ObjectPropertyDomain>
        <ObjectPropertyDomain>
            <ObjectProperty IRI="#pre-requisite"/>
            <Class IRI="#PolicyAction"/>
        </ObjectPropertyDomain>
        <ObjectPropertyDomain>
            <ObjectProperty IRI="#proposes"/>
            <Class IRI="#Stakeholder"/>
        </ObjectPropertyDomain>
        <ObjectPropertyDomain>
            <ObjectProperty IRI="#refersTo"/>
            <Class IRI="#Position"/>
        </ObjectPropertyDomain>
        <ObjectPropertyDomain>
            <ObjectProperty IRI="#refinedInto"/>
```

```
        <Class IRI="#Goal"/>
    </ObjectPropertyDomain>
    <ObjectPropertyDomain>
        <ObjectProperty IRI="#spec"/>
        <Class IRI="#Dataset"/>
    </ObjectPropertyDomain>
    <ObjectPropertyDomain>
        <ObjectProperty IRI="#specifies"/>
        <Class IRI="#Criterion"/>
    </ObjectPropertyDomain>
    <ObjectPropertyDomain>
        <ObjectProperty IRI="#stakeholder"/>
        <Class IRI="#Position"/>
    </ObjectPropertyDomain>
    <ObjectPropertyDomain>
        <ObjectProperty IRI="#utilizes"/>
        <Class IRI="#DataAnalyticsTask"/>
    </ObjectPropertyDomain>
    <ObjectPropertyRange>
        <ObjectProperty IRI="#a1"/>
        <Class IRI="#Alternative"/>
    </ObjectPropertyRange>
    <ObjectPropertyRange>
        <ObjectProperty IRI="#a1"/>
        <Class IRI="#PolicyAction"/>
    </ObjectPropertyRange>
    <ObjectPropertyRange>
        <ObjectProperty IRI="#a2"/>
        <Class IRI="#Alternative"/>
    </ObjectPropertyRange>
    <ObjectPropertyRange>
        <ObjectProperty IRI="#a2"/>
        <Class IRI="#PolicyAction"/>
    </ObjectPropertyRange>
    <ObjectPropertyRange>
        <ObjectProperty IRI="#advocates"/>
        <Class IRI="#Position"/>
    </ObjectPropertyRange>
```

```
        <ObjectPropertyRange>
            <ObjectProperty IRI="#aimedAt"/>
            <Class IRI="#Goal"/>
        </ObjectPropertyRange>
        <ObjectPropertyRange>
            <ObjectProperty IRI="#algorithm"/>
            <Class IRI="#Algorithm"/>
        </ObjectPropertyRange>
        <ObjectPropertyRange>
            <ObjectProperty IRI="#canBeAddressedBy"/>
            <Class IRI="#PolicyAction"/>
        </ObjectPropertyRange>
        <ObjectPropertyRange>
            <ObjectProperty IRI="#constraint"/>
            <Class IRI="#Constraint"/>
        </ObjectPropertyRange>
        <ObjectPropertyRange>
            <ObjectProperty IRI="#constraints"/>
            <Class IRI="#Dataset"/>
        </ObjectPropertyRange>
        <ObjectPropertyRange>
            <ObjectProperty IRI="#criterion"/>
            <Class IRI="#Criterion"/>
        </ObjectPropertyRange>
        <ObjectPropertyRange>
            <ObjectProperty IRI="#dataAnalyticsTask"/>
            <Class IRI="#DataAnalyticsTask"/>
        </ObjectPropertyRange>
        <ObjectPropertyRange>
            <ObjectProperty IRI="#dataAnalyticsWorkflow"/>
            <Class IRI="#DataAnalyticsWorfklow"/>
        </ObjectPropertyRange>
        <ObjectPropertyRange>
            <ObjectProperty IRI="#dataSpec"/>
            <Class IRI="#OutputDataSpecification"/>
        </ObjectPropertyRange>
        <ObjectPropertyRange>
            <ObjectProperty IRI="#dataTiming"/>
```

```
        <Class IRI="#TimedDataset"/>
    </ObjectPropertyRange>
    <ObjectPropertyRange>
        <ObjectProperty IRI="#dataset"/>
        <Class IRI="#Dataset"/>
    </ObjectPropertyRange>
    <ObjectPropertyRange>
        <ObjectProperty IRI="#dependant"/>
        <Class IRI="#PolicyAction"/>
    </ObjectPropertyRange>
    <ObjectPropertyRange>
        <ObjectProperty IRI="#executionType"/>
        <Class IRI="#WorkflowExecutionType"/>
    </ObjectPropertyRange>
    <ObjectPropertyRange>
        <ObjectProperty IRI="#goal"/>
        <Class IRI="#Goal"/>
    </ObjectPropertyRange>
    <ObjectPropertyRange>
        <ObjectProperty IRI="#hasWorkflows"/>
        <Class IRI="#DataAnalyticsWorfklow"/>
    </ObjectPropertyRange>
    <ObjectPropertyRange>
        <ObjectProperty IRI="#involvesStakeholders"/>
        <Class IRI="#Stakeholder"/>
    </ObjectPropertyRange>
    <ObjectPropertyRange>
        <ObjectProperty IRI="#isComposedOf"/>
        <Class IRI="#DataAnalyticsTask"/>
    </ObjectPropertyRange>
    <ObjectPropertyRange>
        <ObjectProperty IRI="#isEvaluatedBy"/>
        <Class IRI="#Criterion"/>
    </ObjectPropertyRange>
    <ObjectPropertyRange>
        <ObjectProperty IRI="#objective"/>
        <Class IRI="#Objective"/>
    </ObjectPropertyRange>
```

```
        <ObjectPropertyRange>
            <ObjectProperty IRI="#period"/>
            <Class IRI="#Period"/>
        </ObjectPropertyRange>
        <ObjectPropertyRange>
            <ObjectProperty IRI="#periodicity"/>
            <Class IRI="#Period"/>
        </ObjectPropertyRange>
        <ObjectPropertyRange>
            <ObjectProperty IRI="#policyAction"/>
            <Class IRI="#PolicyAction"/>
        </ObjectPropertyRange>
        <ObjectPropertyRange>
            <ObjectProperty IRI="#policyModel"/>
            <Class IRI="#PolicyModel"/>
        </ObjectPropertyRange>
        <ObjectPropertyRange>
            <ObjectProperty IRI="#position"/>
            <Class IRI="#Position"/>
        </ObjectPropertyRange>
        <ObjectPropertyRange>
            <ObjectProperty IRI="#pre-requisite"/>
            <Class IRI="#PolicyAction"/>
        </ObjectPropertyRange>
        <ObjectPropertyRange>
            <ObjectProperty IRI="#proposes"/>
            <Class IRI="#Position"/>
        </ObjectPropertyRange>
        <ObjectPropertyRange>
            <ObjectProperty IRI="#refersTo"/>
            <Class IRI="#PolicyAction"/>
        </ObjectPropertyRange>
        <ObjectPropertyRange>
            <ObjectProperty IRI="#refinedInto"/>
            <Class IRI="#Objective"/>
        </ObjectPropertyRange>
        <ObjectPropertyRange>
            <ObjectProperty IRI="#spec"/>
```

```
        <Class IRI="#DataSpecification"/>
    </ObjectPropertyRange>
    <ObjectPropertyRange>
        <ObjectProperty IRI="#specifies"/>
        <Class IRI="#DataAnalyticsWorfklow"/>
    </ObjectPropertyRange>
    <ObjectPropertyRange>
        <ObjectProperty IRI="#stakeholder"/>
        <Class IRI="#Stakeholder"/>
    </ObjectPropertyRange>
    <ObjectPropertyRange>
        <ObjectProperty IRI="#utilizes"/>
        <Class IRI="#Method"/>
    </ObjectPropertyRange>
    <SubDataPropertyOf>
        <DataProperty IRI="#KNN"/>
        <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <SubDataPropertyOf>
        <DataProperty IRI="#KValue"/>
        <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <SubDataPropertyOf>
        <DataProperty IRI="#acuity"/>
        <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <SubDataPropertyOf>
        <DataProperty IRI="#allowUnclassifiedInstances"/>
        <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <SubDataPropertyOf>
        <DataProperty IRI="#attributeSelectionMethod"/>
        <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <SubDataPropertyOf>
        <DataProperty IRI="#bagSizePercent"/>
        <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
```

```
        <SubDataPropertyOf>
            <DataProperty IRI="#batchSize"/>
            <DataProperty IRI="#algorithmConfigurationOptions"/>
        </SubDataPropertyOf>
        <SubDataPropertyOf>
            <DataProperty IRI="#binarySplits"/>
            <DataProperty IRI="#algorithmConfigurationOptions"/>
        </SubDataPropertyOf>
        <SubDataPropertyOf>
            <DataProperty IRI="#breakTiesRandomly"/>
            <DataProperty IRI="#algorithmConfigurationOptions"/>
        </SubDataPropertyOf>
        <SubDataPropertyOf>
            <DataProperty IRI="#calcOutOfBag"/>
            <DataProperty IRI="#algorithmConfigurationOptions"/>
        </SubDataPropertyOf>
        <SubDataPropertyOf>
            <DataProperty IRI="#canopyMaxNumCanopiesToHoldInMemory"/>
            <DataProperty IRI="#algorithmConfigurationOptions"/>
        </SubDataPropertyOf>
        <SubDataPropertyOf>
            <DataProperty IRI="#canopyMinimumCanopyDensity"/>
            <DataProperty IRI="#algorithmConfigurationOptions"/>
        </SubDataPropertyOf>
        <SubDataPropertyOf>
            <DataProperty IRI="#canopyPeriodicPruningRate"/>
            <DataProperty IRI="#algorithmConfigurationOptions"/>
        </SubDataPropertyOf>
        <SubDataPropertyOf>
            <DataProperty IRI="#canopyT1"/>
            <DataProperty IRI="#algorithmConfigurationOptions"/>
        </SubDataPropertyOf>
        <SubDataPropertyOf>
            <DataProperty IRI="#canopyT2"/>
            <DataProperty IRI="#algorithmConfigurationOptions"/>
        </SubDataPropertyOf>
        <SubDataPropertyOf>
            <DataProperty IRI="#classifier"/>
```

```
        <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <SubDataPropertyOf>
        <DataProperty IRI="#collapseTree"/>
        <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <SubDataPropertyOf>
        <DataProperty IRI="#computeAttributeImportance"/>
        <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <SubDataPropertyOf>
        <DataProperty IRI="#confidenceFactor"/>
        <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <SubDataPropertyOf>
        <DataProperty IRI="#crossVal"/>
        <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <SubDataPropertyOf>
        <DataProperty IRI="#crossValidate"/>
        <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <SubDataPropertyOf>
        <DataProperty IRI="#cutoff"/>
        <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <SubDataPropertyOf>
        <DataProperty IRI="#debug"/>
        <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <SubDataPropertyOf>
        <DataProperty IRI="#displayModeInOldFormat"/>
        <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <SubDataPropertyOf>
        <DataProperty IRI="#displayRules"/>
        <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
```

```
        <SubDataPropertyOf>
            <DataProperty IRI="#displayStdDevs"/>
            <DataProperty IRI="#algorithmConfigurationOptions"/>
        </SubDataPropertyOf>
        <SubDataPropertyOf>
            <DataProperty IRI="#distanceFunction"/>
            <DataProperty IRI="#algorithmConfigurationOptions"/>
        </SubDataPropertyOf>
        <SubDataPropertyOf>
            <DataProperty IRI="#distanceWeighting"/>
            <DataProperty IRI="#algorithmConfigurationOptions"/>
        </SubDataPropertyOf>
        <SubDataPropertyOf>
            <DataProperty IRI="#doNotCheckCapabilities"/>
            <DataProperty IRI="#algorithmConfigurationOptions"/>
        </SubDataPropertyOf>
        <SubDataPropertyOf>
            <DataProperty IRI="#doNotMakeSplitPointActualValue"/>
            <DataProperty IRI="#algorithmConfigurationOptions"/>
        </SubDataPropertyOf>
        <SubDataPropertyOf>
            <DataProperty IRI="#dontReplaceMissingValues"/>
            <DataProperty IRI="#algorithmConfigurationOptions"/>
        </SubDataPropertyOf>
        <SubDataPropertyOf>
            <DataProperty IRI="#estimateColinearAttributes"/>
            <DataProperty IRI="#algorithmConfigurationOptions"/>
        </SubDataPropertyOf>
        <SubDataPropertyOf>
            <DataProperty IRI="#evaluationMeasure"/>
            <DataProperty IRI="#algorithmConfigurationOptions"/>
        </SubDataPropertyOf>
        <SubDataPropertyOf>
            <DataProperty IRI="#fastDistanceCalc"/>
            <DataProperty IRI="#algorithmConfigurationOptions"/>
        </SubDataPropertyOf>
        <SubDataPropertyOf>
            <DataProperty IRI="#filterType"/>
```

```
        <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <SubDataPropertyOf>
        <DataProperty IRI="#initializationMethod"/>
        <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <SubDataPropertyOf>
        <DataProperty IRI="#kernel"/>
        <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <SubDataPropertyOf>
        <DataProperty IRI="#maxDepth"/>
        <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <SubDataPropertyOf>
        <DataProperty IRI="#maxIterations"/>
        <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <SubDataPropertyOf>
        <DataProperty IRI="#maxIts"/>
        <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <SubDataPropertyOf>
        <DataProperty IRI="#maximumNumberOfClusters"/>
        <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <SubDataPropertyOf>
        <DataProperty IRI="#meanSquared"/>
        <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <SubDataPropertyOf>
        <DataProperty IRI="#minLogLikelihoodImprovementCV"/>
        <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <SubDataPropertyOf>
        <DataProperty IRI="#minLogLikelihoodImprovementIterating"/>
        <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
```

```
        <SubDataPropertyOf>
            <DataProperty IRI="#minNumObj"/>
            <DataProperty IRI="#algorithmConfigurationOptions"/>
        </SubDataPropertyOf>
        <SubDataPropertyOf>
            <DataProperty IRI="#minNums"/>
            <DataProperty IRI="#algorithmConfigurationOptions"/>
        </SubDataPropertyOf>
        <SubDataPropertyOf>
            <DataProperty IRI="#minStdDev"/>
            <DataProperty IRI="#algorithmConfigurationOptions"/>
        </SubDataPropertyOf>
        <SubDataPropertyOf>
            <DataProperty IRI="#minVarianceProp"/>
            <DataProperty IRI="#algorithmConfigurationOptions"/>
        </SubDataPropertyOf>
        <SubDataPropertyOf>
            <DataProperty IRI="#minimal"/>
            <DataProperty IRI="#algorithmConfigurationOptions"/>
        </SubDataPropertyOf>
        <SubDataPropertyOf>
            <DataProperty IRI="#minimumCanopyDensity"/>
            <DataProperty IRI="#algorithmConfigurationOptions"/>
        </SubDataPropertyOf>
        <SubDataPropertyOf>
            <DataProperty IRI="#nearestNeighbourSearchAlgorithm"/>
            <DataProperty IRI="#algorithmConfigurationOptions"/>
        </SubDataPropertyOf>
        <SubDataPropertyOf>
            <DataProperty IRI="#noise"/>
            <DataProperty IRI="#algorithmConfigurationOptions"/>
        </SubDataPropertyOf>
        <SubDataPropertyOf>
            <DataProperty IRI="#numClusters"/>
            <DataProperty IRI="#algorithmConfigurationOptions"/>
        </SubDataPropertyOf>
        <SubDataPropertyOf>
            <DataProperty IRI="#numDecimalPlaces"/>
```

```
        <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <SubDataPropertyOf>
        <DataProperty IRI="#numExecutionSlots"/>
        <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <SubDataPropertyOf>
        <DataProperty IRI="#numFeatures"/>
        <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <SubDataPropertyOf>
        <DataProperty IRI="#numFolds"/>
        <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <SubDataPropertyOf>
        <DataProperty IRI="#numIterations"/>
        <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <SubDataPropertyOf>
        <DataProperty IRI="#numKMeansRuns"/>
        <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <SubDataPropertyOf>
        <DataProperty IRI="#outputAdditionalStats"/>
        <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <SubDataPropertyOf>
        <DataProperty IRI="#outputOutOfBagComplexityStatistics"/>
        <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <SubDataPropertyOf>
        <DataProperty IRI="#periodicPruningRate"/>
        <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <SubDataPropertyOf>
        <DataProperty IRI="#preserveInstancesOrder"/>
        <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
```

```
    <SubDataPropertyOf>
        <DataProperty IRI="#printClassifiers"/>
        <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <SubDataPropertyOf>
        <DataProperty IRI="#reduceNumberOfDistanceCalcsViaCanopies"/>
        <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <SubDataPropertyOf>
        <DataProperty IRI="#reducedErrorPruning"/>
        <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <SubDataPropertyOf>
        <DataProperty IRI="#representCopiesUsingWeights"/>
        <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <SubDataPropertyOf>
        <DataProperty IRI="#ridge"/>
        <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <SubDataPropertyOf>
        <DataProperty IRI="#saveInstanceData"/>
        <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <SubDataPropertyOf>
        <DataProperty IRI="#search"/>
        <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <SubDataPropertyOf>
        <DataProperty IRI="#seed"/>
        <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <SubDataPropertyOf>
        <DataProperty IRI="#storeOutOfBagPredictions"/>
        <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <SubDataPropertyOf>
        <DataProperty IRI="#subtreeRaising"/>
```

```
        <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <SubDataPropertyOf>
        <DataProperty IRI="#t1"/>
        <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <SubDataPropertyOf>
        <DataProperty IRI="#t2"/>
        <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <SubDataPropertyOf>
        <DataProperty IRI="#unpruned"/>
        <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <SubDataPropertyOf>
        <DataProperty IRI="#useConjugateGradientDescent"/>
        <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <SubDataPropertyOf>
        <DataProperty IRI="#useIBk"/>
        <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <SubDataPropertyOf>
        <DataProperty IRI="#useKernelEstimator"/>
        <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <SubDataPropertyOf>
        <DataProperty IRI="#useLaplace"/>
        <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <SubDataPropertyOf>
        <DataProperty IRI="#useMDLcorrection"/>
        <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <SubDataPropertyOf>
        <DataProperty IRI="#useSupervisedDiscretization"/>
        <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
```

```
    <SubDataPropertyOf>
        <DataProperty IRI="#windowSize"/>
        <DataProperty IRI="#algorithmConfigurationOptions"/>
    </SubDataPropertyOf>
    <DataPropertyDomain>
        <DataProperty IRI="#KNN"/>
        <Class IRI="#IBk"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#KValue"/>
        <Class IRI="#RandomTree"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#acuity"/>
        <Class IRI="#CowebClassit"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#allowUnclassifiedInstances"/>
        <Class IRI="#RandomTree"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#attributeSelectionMethod"/>
        <Class IRI="#LinearRegression"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#bagSizePercent"/>
        <Class IRI="#RandomForest"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#batchSize"/>
        <Class IRI="#DataStream"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#batchSize"/>
        <Class IRI="#DecisionTable"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#batchSize"/>
```

```
        <Class IRI="#GaussianProcesses"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#batchSize"/>
        <Class IRI="#IBk"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#batchSize"/>
        <Class IRI="#J48"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#batchSize"/>
        <Class IRI="#LinearRegression"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#batchSize"/>
        <Class IRI="#MultinomialLogisticRegression"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#batchSize"/>
        <Class IRI="#NaiveBayes"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#batchSize"/>
        <Class IRI="#RandomForest"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#batchSize"/>
        <Class IRI="#RandomTree"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#batchSize"/>
        <Class IRI="#ZeroR"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#binarySplits"/>
        <Class IRI="#J48"/>
    </DataPropertyDomain>
```

```
        <DataPropertyDomain>
            <DataProperty IRI="#breakTiesRandomly"/>
            <Class IRI="#RandomForest"/>
        </DataPropertyDomain>
        <DataPropertyDomain>
            <DataProperty IRI="#breakTiesRandomly"/>
            <Class IRI="#RandomTree"/>
        </DataPropertyDomain>
        <DataPropertyDomain>
            <DataProperty IRI="#calcOutOfBag"/>
            <Class IRI="#RandomForest"/>
        </DataPropertyDomain>
        <DataPropertyDomain>
            <DataProperty IRI="#canopyMaxNumCanopiesToHoldInMemory"/>
            <Class IRI="#SimpleKMeans"/>
        </DataPropertyDomain>
        <DataPropertyDomain>
            <DataProperty IRI="#canopyMinimumCanopyDensity"/>
            <Class IRI="#SimpleKMeans"/>
        </DataPropertyDomain>
        <DataPropertyDomain>
            <DataProperty IRI="#canopyPeriodicPruningRate"/>
            <Class IRI="#SimpleKMeans"/>
        </DataPropertyDomain>
        <DataPropertyDomain>
            <DataProperty IRI="#canopyT1"/>
            <Class IRI="#SimpleKMeans"/>
        </DataPropertyDomain>
        <DataPropertyDomain>
            <DataProperty IRI="#canopyT2"/>
            <Class IRI="#SimpleKMeans"/>
        </DataPropertyDomain>
        <DataPropertyDomain>
            <DataProperty IRI="#classifier"/>
            <Class IRI="#RandomForest"/>
        </DataPropertyDomain>
        <DataPropertyDomain>
            <DataProperty IRI="#cleaninRule"/>
```

```
        <Class IRI="#DataCleaningOperation"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#collapseTree"/>
        <Class IRI="#J48"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#computeAttributeImportance"/>
        <Class IRI="#RandomForest"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#condition"/>
        <Class IRI="#FilterOperation"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#condition"/>
        <Class IRI="#JoinOperation"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#confidenceFactor"/>
        <Class IRI="#J48"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#crossVal"/>
        <Class IRI="#DecisionTable"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#crossValidate"/>
        <Class IRI="#IBk"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#cutoff"/>
        <Class IRI="#CowebClassit"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#debug"/>
        <Class IRI="#DataMiningAlgorithm"/>
    </DataPropertyDomain>
```

```
    <DataPropertyDomain>
        <DataProperty IRI="#dependentVariable"/>
        <Class IRI="#DataMiningTask"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#dependentVariable"/>
        <Class IRI="#StatisticalAnalysisTask"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#description"/>
        <Class IRI="#Criterion"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#description"/>
        <Class IRI="#Goal"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#description"/>
        <Class IRI="#Objective"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#description"/>
        <Class IRI="#PolicyAction"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#displayModeInOldFormat"/>
        <Class IRI="#EM"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#displayModeInOldFormat"/>
        <Class IRI="#NaiveBayes"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#displayRules"/>
        <Class IRI="#DecisionTable"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#displayStdDevs"/>
```

```
        <Class IRI="#SimpleKMeans"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#distanceFunction"/>
        <Class IRI="#SimpleKMeans"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#distanceWeighting"/>
        <Class IRI="#IBk"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#doNotCheckCapabilities"/>
        <Class IRI="#DataMiningAlgorithm"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#doNotMakeSplitPointActualValue"/>
        <Class IRI="#J48"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#dontReplaceMissingValues"/>
        <Class IRI="#Canopy"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#dontReplaceMissingValues"/>
        <Class IRI="#SimpleKMeans"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#estimateColinearAttributes"/>
        <Class IRI="#LinearRegression"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#evaluationMeasure"/>
        <Class IRI="#DecisionTable"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#executionEnd"/>
        <Class IRI="#PeriodicExecution"/>
    </DataPropertyDomain>
```

```
        <DataPropertyDomain>
            <DataProperty IRI="#executionStart"/>
            <Class IRI="#PeriodicExecution"/>
        </DataPropertyDomain>
        <DataPropertyDomain>
            <DataProperty IRI="#fastDistanceCalc"/>
            <Class IRI="#SimpleKMeans"/>
        </DataPropertyDomain>
        <DataPropertyDomain>
            <DataProperty IRI="#filterType"/>
            <Class IRI="#GaussianProcesses"/>
        </DataPropertyDomain>
        <DataPropertyDomain>
            <DataProperty IRI="#fixedPoint"/>
            <Class IRI="#ShiftingDataset"/>
        </DataPropertyDomain>
        <DataPropertyDomain>
            <DataProperty IRI="#from"/>
            <Class IRI="#AbsoluteDataset"/>
        </DataPropertyDomain>
        <DataPropertyDomain>
            <DataProperty IRI="#increaseRatio"/>
            <Class IRI="#DataChangeDriven"/>
        </DataPropertyDomain>
        <DataPropertyDomain>
            <DataProperty IRI="#initializationMethod"/>
            <Class IRI="#SimpleKMeans"/>
        </DataPropertyDomain>
        <DataPropertyDomain>
            <DataProperty IRI="#kernel"/>
            <Class IRI="#GaussianProcesses"/>
        </DataPropertyDomain>
        <DataPropertyDomain>
            <DataProperty IRI="#logicalExpression"/>
            <Class IRI="#Criterion"/>
        </DataPropertyDomain>
        <DataPropertyDomain>
            <DataProperty IRI="#maxDepth"/>
```

```
        <Class IRI="#RandomForest"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#maxDepth"/>
        <Class IRI="#RandomTree"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#maxIterations"/>
        <Class IRI="#EM"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#maxIterations"/>
        <Class IRI="#SimpleKMeans"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#maxIts"/>
        <Class IRI="#MultinomialLogisticRegression"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#maximumNumberOfClusters"/>
        <Class IRI="#EM"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#meanSquared"/>
        <Class IRI="#IBk"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#minLogLikelihoodImprovementCV"/>
        <Class IRI="#EM"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#minLogLikelihoodImprovementIterating"/>
        <Class IRI="#EM"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#minNumObj"/>
        <Class IRI="#J48"/>
    </DataPropertyDomain>
```

```
        <DataPropertyDomain>
            <DataProperty IRI="#minNums"/>
            <Class IRI="#RandomTree"/>
        </DataPropertyDomain>
        <DataPropertyDomain>
            <DataProperty IRI="#minStdDev"/>
            <Class IRI="#EM"/>
        </DataPropertyDomain>
        <DataPropertyDomain>
            <DataProperty IRI="#minVarianceProp"/>
            <Class IRI="#RandomTree"/>
        </DataPropertyDomain>
        <DataPropertyDomain>
            <DataProperty IRI="#minimal"/>
            <Class IRI="#LinearRegression"/>
        </DataPropertyDomain>
        <DataPropertyDomain>
            <DataProperty IRI="#minimumCanopyDensity"/>
            <Class IRI="#Canopy"/>
        </DataPropertyDomain>
        <DataPropertyDomain>
            <DataProperty IRI="#nearestNeighbourSearchAlgorithm"/>
            <Class IRI="#IBk"/>
        </DataPropertyDomain>
        <DataPropertyDomain>
            <DataProperty IRI="#noise"/>
            <Class IRI="#GaussianProcesses"/>
        </DataPropertyDomain>
        <DataPropertyDomain>
            <DataProperty IRI="#numClusters"/>
            <Class IRI="#Canopy"/>
        </DataPropertyDomain>
        <DataPropertyDomain>
            <DataProperty IRI="#numClusters"/>
            <Class IRI="#EM"/>
        </DataPropertyDomain>
        <DataPropertyDomain>
            <DataProperty IRI="#numClusters"/>
```

```
        <Class IRI="#FarthestFirst"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#numClusters"/>
        <Class IRI="#SimpleKMeans"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#numDecimalPlaces"/>
        <Class IRI="#DecisionTable"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#numDecimalPlaces"/>
        <Class IRI="#GaussianProcesses"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#numDecimalPlaces"/>
        <Class IRI="#IBk"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#numDecimalPlaces"/>
        <Class IRI="#J48"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#numDecimalPlaces"/>
        <Class IRI="#LinearRegression"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#numDecimalPlaces"/>
        <Class IRI="#MultinomialLogisticRegression"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#numDecimalPlaces"/>
        <Class IRI="#NaiveBayes"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#numDecimalPlaces"/>
        <Class IRI="#RandomTree"/>
    </DataPropertyDomain>
```

```
    <DataPropertyDomain>
        <DataProperty IRI="#numDecimalPlaces"/>
        <Class IRI="#ZeroR"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#numExecutionSlots"/>
        <Class IRI="#EM"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#numExecutionSlots"/>
        <Class IRI="#RandomForest"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#numExecutionSlots"/>
        <Class IRI="#SimpleKMeans"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#numFeatures"/>
        <Class IRI="#RandomForest"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#numFolds"/>
        <Class IRI="#EM"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#numFolds"/>
        <Class IRI="#J48"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#numFolds"/>
        <Class IRI="#RandomTree"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#numIterations"/>
        <Class IRI="#RandomForest"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#numKMeansRuns"/>
```

```
        <Class IRI="#EM"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#outputAdditionalStats"/>
        <Class IRI="#LinearRegression"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#outputOutOfBagComplexityStatistics"/>
        <Class IRI="#RandomForest"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#periodicPruningRate"/>
        <Class IRI="#Canopy"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#preserveInstancesOrder"/>
        <Class IRI="#SimpleKMeans"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#printClassifiers"/>
        <Class IRI="#RandomForest"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#quantity"/>
        <Class IRI="#Period"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#rationale"/>
        <Class IRI="#Goal"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#rationale"/>
        <Class IRI="#Objective"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#reduceNumberOfDistanceCalcsViaCanopies"/>
        <Class IRI="#SimpleKMeans"/>
    </DataPropertyDomain>
```

```
        <DataPropertyDomain>
            <DataProperty IRI="#reducedErrorPruning"/>
            <Class IRI="#J48"/>
        </DataPropertyDomain>
        <DataPropertyDomain>
            <DataProperty IRI="#representCopiesUsingWeights"/>
            <Class IRI="#RandomForest"/>
        </DataPropertyDomain>
        <DataPropertyDomain>
            <DataProperty IRI="#ridge"/>
            <Class IRI="#LinearRegression"/>
        </DataPropertyDomain>
        <DataPropertyDomain>
            <DataProperty IRI="#ridge"/>
            <Class IRI="#MultinomialLogisticRegression"/>
        </DataPropertyDomain>
        <DataPropertyDomain>
            <DataProperty IRI="#saveInstanceData"/>
            <Class IRI="#CowebClassit"/>
        </DataPropertyDomain>
        <DataPropertyDomain>
            <DataProperty IRI="#saveInstanceData"/>
            <Class IRI="#J48"/>
        </DataPropertyDomain>
        <DataPropertyDomain>
            <DataProperty IRI="#search"/>
            <Class IRI="#DecisionTable"/>
        </DataPropertyDomain>
        <DataPropertyDomain>
            <DataProperty IRI="#seed"/>
            <Class IRI="#Canopy"/>
        </DataPropertyDomain>
        <DataPropertyDomain>
            <DataProperty IRI="#seed"/>
            <Class IRI="#CowebClassit"/>
        </DataPropertyDomain>
        <DataPropertyDomain>
            <DataProperty IRI="#seed"/>
```

```
        <Class IRI="#EM"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#seed"/>
        <Class IRI="#FarthestFirst"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#seed"/>
        <Class IRI="#GaussianProcesses"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#seed"/>
        <Class IRI="#J48"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#seed"/>
        <Class IRI="#RandomForest"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#seed"/>
        <Class IRI="#RandomTree"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#seed"/>
        <Class IRI="#SimpleKMeans"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#size"/>
        <Class IRI="#StaticSet"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#storeOutOfBagPredictions"/>
        <Class IRI="#RandomForest"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#subtreeRaising"/>
        <Class IRI="#J48"/>
    </DataPropertyDomain>
```

```
    <DataPropertyDomain>
        <DataProperty IRI="#t1"/>
        <Class IRI="#Canopy"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#t2"/>
        <Class IRI="#Canopy"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#timeunit"/>
        <Class IRI="#Period"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#type"/>
        <Class IRI="#Dataset"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#unpruned"/>
        <Class IRI="#J48"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#useConjugateGradientDescent"/>
        <Class IRI="#MultinomialLogisticRegression"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#useIBk"/>
        <Class IRI="#DecisionTable"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#useKernelEstimator"/>
        <Class IRI="#NaiveBayes"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#useLaplace"/>
        <Class IRI="#J48"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#useMDLcorrection"/>
```

```
        <Class IRI="#J48"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#useSupervisedDiscretization"/>
        <Class IRI="#NaiveBayes"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#weight"/>
        <Class IRI="#Criterion"/>
    </DataPropertyDomain>
    <DataPropertyDomain>
        <DataProperty IRI="#windowSize"/>
        <Class IRI="#IBk"/>
    </DataPropertyDomain>
    <DataPropertyRange>
        <DataProperty IRI="#KNN"/>
        <Datatype abbreviatedIRI="xsd:positiveInteger"/>
    </DataPropertyRange>
    <DataPropertyRange>
        <DataProperty IRI="#KValue"/>
        <Datatype abbreviatedIRI="xsd:integer"/>
    </DataPropertyRange>
    <DataPropertyRange>
        <DataProperty IRI="#acuity"/>
        <Datatype abbreviatedIRI="xsd:float"/>
    </DataPropertyRange>
    <DataPropertyRange>
        <DataProperty IRI="#allowUnclassifiedInstances"/>
        <Datatype abbreviatedIRI="xsd:boolean"/>
    </DataPropertyRange>
    <DataPropertyRange>
        <DataProperty IRI="#attributeSelectionMethod"/>
        <Datatype abbreviatedIRI="xsd:positiveInteger"/>
    </DataPropertyRange>
    <DataPropertyRange>
        <DataProperty IRI="#bagSizePercent"/>
        <Datatype abbreviatedIRI="xsd:positiveInteger"/>
    </DataPropertyRange>
```

```
        <DataPropertyRange>
            <DataProperty IRI="#batchSize"/>
            <Datatype abbreviatedIRI="xsd:integer"/>
        </DataPropertyRange>
        <DataPropertyRange>
            <DataProperty IRI="#batchSize"/>
            <Datatype abbreviatedIRI="xsd:positiveInteger"/>
        </DataPropertyRange>
        <DataPropertyRange>
            <DataProperty IRI="#binarySplits"/>
            <Datatype abbreviatedIRI="xsd:boolean"/>
        </DataPropertyRange>
        <DataPropertyRange>
            <DataProperty IRI="#breakTiesRandomly"/>
            <Datatype abbreviatedIRI="xsd:boolean"/>
        </DataPropertyRange>
        <DataPropertyRange>
            <DataProperty IRI="#calcOutOfBag"/>
            <Datatype abbreviatedIRI="xsd:boolean"/>
        </DataPropertyRange>
        <DataPropertyRange>
            <DataProperty IRI="#canopyMaxNumCanopiesToHoldInMemory"/>
            <Datatype abbreviatedIRI="xsd:integer"/>
        </DataPropertyRange>
        <DataPropertyRange>
            <DataProperty IRI="#canopyMinimumCanopyDensity"/>
            <Datatype abbreviatedIRI="xsd:float"/>
        </DataPropertyRange>
        <DataPropertyRange>
            <DataProperty IRI="#canopyPeriodicPruningRate"/>
            <Datatype abbreviatedIRI="xsd:integer"/>
        </DataPropertyRange>
        <DataPropertyRange>
            <DataProperty IRI="#canopyT1"/>
            <Datatype abbreviatedIRI="xsd:float"/>
        </DataPropertyRange>
        <DataPropertyRange>
            <DataProperty IRI="#canopyT2"/>
```

```
        <Datatype abbreviatedIRI="xsd:float"/>
    </DataPropertyRange>
    <DataPropertyRange>
        <DataProperty IRI="#classifier"/>
        <Datatype abbreviatedIRI="xsd:string"/>
    </DataPropertyRange>
    <DataPropertyRange>
        <DataProperty IRI="#cleaningRule"/>
        <Datatype abbreviatedIRI="xsd:string"/>
    </DataPropertyRange>
    <DataPropertyRange>
        <DataProperty IRI="#collapseTree"/>
        <Datatype abbreviatedIRI="xsd:boolean"/>
    </DataPropertyRange>
    <DataPropertyRange>
        <DataProperty IRI="#columnName"/>
        <Datatype abbreviatedIRI="xsd:string"/>
    </DataPropertyRange>
    <DataPropertyRange>
        <DataProperty IRI="#computeAttributeImportance"/>
        <Datatype abbreviatedIRI="xsd:boolean"/>
    </DataPropertyRange>
    <DataPropertyRange>
        <DataProperty IRI="#condition"/>
        <Datatype abbreviatedIRI="xsd:string"/>
    </DataPropertyRange>
    <DataPropertyRange>
        <DataProperty IRI="#confidenceFactor"/>
        <Datatype abbreviatedIRI="xsd:float"/>
    </DataPropertyRange>
    <DataPropertyRange>
        <DataProperty IRI="#crossVal"/>
        <Datatype abbreviatedIRI="xsd:integer"/>
    </DataPropertyRange>
    <DataPropertyRange>
        <DataProperty IRI="#crossValidate"/>
        <Datatype abbreviatedIRI="xsd:boolean"/>
    </DataPropertyRange>
```

```
        <DataPropertyRange>
            <DataProperty IRI="#cutoff"/>
            <Datatype abbreviatedIRI="xsd:double"/>
        </DataPropertyRange>
        <DataPropertyRange>
            <DataProperty IRI="#debug"/>
            <Datatype abbreviatedIRI="xsd:boolean"/>
        </DataPropertyRange>
        <DataPropertyRange>
            <DataProperty IRI="#description"/>
            <Datatype abbreviatedIRI="xsd:string"/>
        </DataPropertyRange>
        <DataPropertyRange>
            <DataProperty IRI="#displayModeInOldFormat"/>
            <Datatype abbreviatedIRI="xsd:boolean"/>
        </DataPropertyRange>
        <DataPropertyRange>
            <DataProperty IRI="#displayRules"/>
            <Datatype abbreviatedIRI="xsd:boolean"/>
        </DataPropertyRange>
        <DataPropertyRange>
            <DataProperty IRI="#displayStdDevs"/>
            <Datatype abbreviatedIRI="xsd:boolean"/>
        </DataPropertyRange>
        <DataPropertyRange>
            <DataProperty IRI="#distanceFunction"/>
            <Datatype abbreviatedIRI="xsd:string"/>
        </DataPropertyRange>
        <DataPropertyRange>
            <DataProperty IRI="#distanceWeighting"/>
            <Datatype abbreviatedIRI="xsd:positiveInteger"/>
        </DataPropertyRange>
        <DataPropertyRange>
            <DataProperty IRI="#doNotCheckCapabilities"/>
            <Datatype abbreviatedIRI="xsd:boolean"/>
        </DataPropertyRange>
        <DataPropertyRange>
            <DataProperty IRI="#doNotMakeSplitPointActualValue"/>
```

```
        <Datatype abbreviatedIRI="xsd:boolean"/>
    </DataPropertyRange>
    <DataPropertyRange>
        <DataProperty IRI="#dontReplaceMissingValues"/>
        <Datatype abbreviatedIRI="xsd:boolean"/>
    </DataPropertyRange>
    <DataPropertyRange>
        <DataProperty IRI="#estimateColinearAttributes"/>
        <Datatype abbreviatedIRI="xsd:boolean"/>
    </DataPropertyRange>
    <DataPropertyRange>
        <DataProperty IRI="#evaluationMeasure"/>
        <Datatype abbreviatedIRI="xsd:postiveInteger"/>
    </DataPropertyRange>
    <DataPropertyRange>
        <DataProperty IRI="#executionEnd"/>
        <Datatype abbreviatedIRI="xsd:dateTime"/>
    </DataPropertyRange>
    <DataPropertyRange>
        <DataProperty IRI="#executionStart"/>
        <Datatype abbreviatedIRI="xsd:dateTime"/>
    </DataPropertyRange>
    <DataPropertyRange>
        <DataProperty IRI="#fastDistanceCalc"/>
        <Datatype abbreviatedIRI="xsd:boolean"/>
    </DataPropertyRange>
    <DataPropertyRange>
        <DataProperty IRI="#filterType"/>
        <Datatype abbreviatedIRI="xsd:positiveInteger"/>
    </DataPropertyRange>
    <DataPropertyRange>
        <DataProperty IRI="#fixedPoint"/>
        <Datatype abbreviatedIRI="xsd:string"/>
    </DataPropertyRange>
    <DataPropertyRange>
        <DataProperty IRI="#from"/>
        <Datatype abbreviatedIRI="xsd:dateTime"/>
    </DataPropertyRange>
```

```xml
    <DataPropertyRange>
        <DataProperty IRI="#increaseRatio"/>
        <Datatype abbreviatedIRI="xsd:float"/>
    </DataPropertyRange>
    <DataPropertyRange>
        <DataProperty IRI="#initializationMethod"/>
        <Datatype abbreviatedIRI="xsd:string"/>
    </DataPropertyRange>
    <DataPropertyRange>
        <DataProperty IRI="#kernel"/>
        <Datatype abbreviatedIRI="xsd:string"/>
    </DataPropertyRange>
    <DataPropertyRange>
        <DataProperty IRI="#logicalExpression"/>
        <Datatype abbreviatedIRI="xsd:string"/>
    </DataPropertyRange>
    <DataPropertyRange>
        <DataProperty IRI="#maxDepth"/>
        <Datatype abbreviatedIRI="xsd:positiveInteger"/>
    </DataPropertyRange>
    <DataPropertyRange>
        <DataProperty IRI="#maxIterations"/>
        <Datatype abbreviatedIRI="xsd:integer"/>
    </DataPropertyRange>
    <DataPropertyRange>
        <DataProperty IRI="#maxIts"/>
        <Datatype abbreviatedIRI="xsd:integer"/>
    </DataPropertyRange>
    <DataPropertyRange>
        <DataProperty IRI="#maximumNumberOfClusters"/>
        <Datatype abbreviatedIRI="xsd:integer"/>
    </DataPropertyRange>
    <DataPropertyRange>
        <DataProperty IRI="#meanSquared"/>
        <Datatype abbreviatedIRI="xsd:boolean"/>
    </DataPropertyRange>
    <DataPropertyRange>
        <DataProperty IRI="#minLogLikelihoodImprovementCV"/>
```

```
        <Datatype abbreviatedIRI="xsd:string"/>
    </DataPropertyRange>
    <DataPropertyRange>
        <DataProperty IRI="#minLogLikelihoodImprovementIterating"/>
        <Datatype abbreviatedIRI="xsd:string"/>
    </DataPropertyRange>
    <DataPropertyRange>
        <DataProperty IRI="#minNumObj"/>
        <Datatype abbreviatedIRI="xsd:integer"/>
    </DataPropertyRange>
    <DataPropertyRange>
        <DataProperty IRI="#minNums"/>
        <Datatype abbreviatedIRI="xsd:float"/>
    </DataPropertyRange>
    <DataPropertyRange>
        <DataProperty IRI="#minStdDev"/>
        <Datatype abbreviatedIRI="xsd:string"/>
    </DataPropertyRange>
    <DataPropertyRange>
        <DataProperty IRI="#minVarianceProp"/>
        <Datatype abbreviatedIRI="xsd:float"/>
    </DataPropertyRange>
    <DataPropertyRange>
        <DataProperty IRI="#minimal"/>
        <Datatype abbreviatedIRI="xsd:boolean"/>
    </DataPropertyRange>
    <DataPropertyRange>
        <DataProperty IRI="#minimumCanopyDensity"/>
        <Datatype abbreviatedIRI="xsd:float"/>
    </DataPropertyRange>
    <DataPropertyRange>
        <DataProperty IRI="#nearestNeighbourSearchAlgorithm"/>
        <Datatype abbreviatedIRI="xsd:string"/>
    </DataPropertyRange>
    <DataPropertyRange>
        <DataProperty IRI="#noise"/>
        <Datatype abbreviatedIRI="xsd:float"/>
    </DataPropertyRange>
```

```
        <DataPropertyRange>
            <DataProperty IRI="#numClusters"/>
            <Datatype abbreviatedIRI="xsd:integer"/>
        </DataPropertyRange>
        <DataPropertyRange>
            <DataProperty IRI="#numDecimalPlaces"/>
            <Datatype abbreviatedIRI="xsd:positiveInteger"/>
        </DataPropertyRange>
        <DataPropertyRange>
            <DataProperty IRI="#numExecutionSlots"/>
            <Datatype abbreviatedIRI="xsd:integer"/>
        </DataPropertyRange>
        <DataPropertyRange>
            <DataProperty IRI="#numExecutionSlots"/>
            <Datatype abbreviatedIRI="xsd:positiveInteger"/>
        </DataPropertyRange>
        <DataPropertyRange>
            <DataProperty IRI="#numFeatures"/>
            <Datatype abbreviatedIRI="xsd:positiveInteger"/>
        </DataPropertyRange>
        <DataPropertyRange>
            <DataProperty IRI="#numFolds"/>
            <Datatype abbreviatedIRI="xsd:integer"/>
        </DataPropertyRange>
        <DataPropertyRange>
            <DataProperty IRI="#numFolds"/>
            <Datatype abbreviatedIRI="xsd:positiveInteger"/>
        </DataPropertyRange>
        <DataPropertyRange>
            <DataProperty IRI="#numIterations"/>
            <Datatype abbreviatedIRI="xsd:positiveInteger"/>
        </DataPropertyRange>
        <DataPropertyRange>
            <DataProperty IRI="#numKMeansRuns"/>
            <Datatype abbreviatedIRI="xsd:integer"/>
        </DataPropertyRange>
        <DataPropertyRange>
            <DataProperty IRI="#outputAdditionalStats"/>
```

```
            <Datatype abbreviatedIRI="xsd:boolean"/>
        </DataPropertyRange>
        <DataPropertyRange>
            <DataProperty IRI="#outputOutOfBagComplexityStatistics"/>
            <Datatype abbreviatedIRI="xsd:boolean"/>
        </DataPropertyRange>
        <DataPropertyRange>
            <DataProperty IRI="#periodicPruningRate"/>
            <Datatype abbreviatedIRI="xsd:integer"/>
        </DataPropertyRange>
        <DataPropertyRange>
            <DataProperty IRI="#preserveInstancesOrder"/>
            <Datatype abbreviatedIRI="xsd:boolean"/>
        </DataPropertyRange>
        <DataPropertyRange>
            <DataProperty IRI="#printClassifiers"/>
            <Datatype abbreviatedIRI="xsd:boolean"/>
        </DataPropertyRange>
        <DataPropertyRange>
            <DataProperty IRI="#projectRule"/>
            <Datatype abbreviatedIRI="xsd:string"/>
        </DataPropertyRange>
        <DataPropertyRange>
            <DataProperty IRI="#quantity"/>
            <DatatypeRestriction>
                <Datatype abbreviatedIRI="xsd:int"/>
                <FacetRestriction
facet="http://www.w3.org/2001/XMLSchema#minExclusive">
                    <Literal
datatypeIRI="http://www.w3.org/2001/XMLSchema#int">0</Literal>
                </FacetRestriction>
            </DatatypeRestriction>
        </DataPropertyRange>
        <DataPropertyRange>
            <DataProperty IRI="#rationale"/>
            <Datatype abbreviatedIRI="xsd:string"/>
        </DataPropertyRange>
        <DataPropertyRange>
```

```
        <DataProperty IRI="#reduceNumberOfDistanceCalcsViaCanopies"/>
        <Datatype abbreviatedIRI="xsd:boolean"/>
    </DataPropertyRange>
    <DataPropertyRange>
        <DataProperty IRI="#reducedErrorPruning"/>
        <Datatype abbreviatedIRI="xsd:boolean"/>
    </DataPropertyRange>
    <DataPropertyRange>
        <DataProperty IRI="#representCopiesUsingWeights"/>
        <Datatype abbreviatedIRI="xsd:boolean"/>
    </DataPropertyRange>
    <DataPropertyRange>
        <DataProperty IRI="#ridge"/>
        <Datatype abbreviatedIRI="xsd:string"/>
    </DataPropertyRange>
    <DataPropertyRange>
        <DataProperty IRI="#saveInstanceData"/>
        <Datatype abbreviatedIRI="xsd:boolean"/>
    </DataPropertyRange>
    <DataPropertyRange>
        <DataProperty IRI="#search"/>
        <Datatype abbreviatedIRI="xsd:string"/>
    </DataPropertyRange>
    <DataPropertyRange>
        <DataProperty IRI="#seed"/>
        <Datatype abbreviatedIRI="xsd:integer"/>
    </DataPropertyRange>
    <DataPropertyRange>
        <DataProperty IRI="#storeOutOfBagPredictions"/>
        <Datatype abbreviatedIRI="xsd:boolean"/>
    </DataPropertyRange>
    <DataPropertyRange>
        <DataProperty IRI="#subtreeRaising"/>
        <Datatype abbreviatedIRI="xsd:boolean"/>
    </DataPropertyRange>
    <DataPropertyRange>
        <DataProperty IRI="#t1"/>
        <Datatype abbreviatedIRI="xsd:float"/>
```

```
        </DataPropertyRange>
        <DataPropertyRange>
            <DataProperty IRI="#t2"/>
            <Datatype abbreviatedIRI="xsd:float"/>
        </DataPropertyRange>
        <DataPropertyRange>
            <DataProperty IRI="#tableName"/>
            <Datatype abbreviatedIRI="xsd:string"/>
        </DataPropertyRange>
        <DataPropertyRange>
            <DataProperty IRI="#timeunit"/>
            <DataOneOf>
                <Literal         datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">day</Literal>
                <Literal         datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">hour</Literal>
                <Literal         datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">minute</Literal>
                <Literal         datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">month</Literal>
                <Literal         datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">second</Literal>
                <Literal         datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">week</Literal>
                <Literal         datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">year</Literal>
            </DataOneOf>
        </DataPropertyRange>
        <DataPropertyRange>
            <DataProperty IRI="#type"/>
            <Datatype abbreviatedIRI="xsd:string"/>
        </DataPropertyRange>
        <DataPropertyRange>
            <DataProperty IRI="#type"/>
            <DataOneOf>
                <Literal         datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">CONJUCTIVE</Literal>
```

```
                <Literal         datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-ns#PlainLiteral">DISJUNCTIVE</Literal>
            </DataOneOf>
        </DataPropertyRange>
        <DataPropertyRange>
            <DataProperty IRI="#unpruned"/>
            <Datatype abbreviatedIRI="xsd:boolean"/>
        </DataPropertyRange>
        <DataPropertyRange>
            <DataProperty IRI="#useConjugateGradientDescent"/>
            <Datatype abbreviatedIRI="xsd:boolean"/>
        </DataPropertyRange>
        <DataPropertyRange>
            <DataProperty IRI="#useIBk"/>
            <Datatype abbreviatedIRI="xsd:boolean"/>
        </DataPropertyRange>
        <DataPropertyRange>
            <DataProperty IRI="#useKernelEstimator"/>
            <Datatype abbreviatedIRI="xsd:boolean"/>
        </DataPropertyRange>
        <DataPropertyRange>
            <DataProperty IRI="#useLaplace"/>
            <Datatype abbreviatedIRI="xsd:boolean"/>
        </DataPropertyRange>
        <DataPropertyRange>
            <DataProperty IRI="#useMDLcorrection"/>
            <Datatype abbreviatedIRI="xsd:boolean"/>
        </DataPropertyRange>
        <DataPropertyRange>
            <DataProperty IRI="#useSupervisedDiscretization"/>
            <Datatype abbreviatedIRI="xsd:boolean"/>
        </DataPropertyRange>
        <DataPropertyRange>
            <DataProperty IRI="#weight"/>
            <Datatype abbreviatedIRI="owl:real"/>
        </DataPropertyRange>
        <DataPropertyRange>
            <DataProperty IRI="#windowSize"/>
```

```
            <Datatype abbreviatedIRI="xsd:positiveInteger"/>
        </DataPropertyRange>
        <AnnotationAssertion>
            <AnnotationProperty abbreviatedIRI="rdfs:label"/>
            <IRI>#AbsoluteDataset</IRI>
            <Literal             datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">Absolute Dataset</Literal>
        </AnnotationAssertion>
        <AnnotationAssertion>
            <AnnotationProperty abbreviatedIRI="rdfs:label"/>
            <IRI>#AdaptiveSampling</IRI>
            <Literal             datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">Adaptive Sampling</Literal>
        </AnnotationAssertion>
        <AnnotationAssertion>
            <AnnotationProperty abbreviatedIRI="rdfs:label"/>
            <IRI>#Breusch-Pagan_Test</IRI>
            <Literal             datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">Breusch-Pagan Test</Literal>
        </AnnotationAssertion>
        <AnnotationAssertion>
            <AnnotationProperty abbreviatedIRI="rdfs:label"/>
            <IRI>#ClusteringSampling</IRI>
            <Literal             datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">Clustering Sampling</Literal>
        </AnnotationAssertion>
        <AnnotationAssertion>
            <AnnotationProperty abbreviatedIRI="rdfs:label"/>
            <IRI>#DataAnalyticsModel</IRI>
            <Literal             datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">Data Analytics Model</Literal>
        </AnnotationAssertion>
        <AnnotationAssertion>
            <AnnotationProperty abbreviatedIRI="rdfs:label"/>
            <IRI>#DataAnalyticsTask</IRI>
            <Literal             datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">Data Analytics Task</Literal>
        </AnnotationAssertion>
```

```
        <AnnotationAssertion>
            <AnnotationProperty abbreviatedIRI="rdfs:label"/>
            <IRI>#DataAnalyticsWorfklow</IRI>
            <Literal             datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">Data Analytics Worfklow</Literal>
        </AnnotationAssertion>
        <AnnotationAssertion>
            <AnnotationProperty abbreviatedIRI="rdfs:label"/>
            <IRI>#DataChangeDriven</IRI>
            <Literal             datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">Data Change Driven</Literal>
        </AnnotationAssertion>
        <AnnotationAssertion>
            <AnnotationProperty abbreviatedIRI="rdfs:label"/>
            <IRI>#DataCleaningOperation</IRI>
            <Literal             datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">Data Cleaning Operation</Literal>
        </AnnotationAssertion>
        <AnnotationAssertion>
            <AnnotationProperty abbreviatedIRI="rdfs:label"/>
            <IRI>#DataMiningAlgorithm</IRI>
            <Literal             datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">Data Mining Algorithm</Literal>
        </AnnotationAssertion>
        <AnnotationAssertion>
            <AnnotationProperty abbreviatedIRI="rdfs:label"/>
            <IRI>#DataMiningTask</IRI>
            <Literal             datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">Data Mining Task</Literal>
        </AnnotationAssertion>
        <AnnotationAssertion>
            <AnnotationProperty abbreviatedIRI="rdfs:label"/>
            <IRI>#DataProcessingTask</IRI>
            <Literal             datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">Data Processing Task</Literal>
        </AnnotationAssertion>
        <AnnotationAssertion>
            <AnnotationProperty abbreviatedIRI="rdfs:label"/>
```

```
            <IRI>#DataSpecification</IRI>
            <Literal             datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-ns#PlainLiteral">Data Specification</Literal>
        </AnnotationAssertion>
        <AnnotationAssertion>
            <AnnotationProperty abbreviatedIRI="rdfs:label"/>
            <IRI>#DataStream</IRI>
            <Literal             datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-ns#PlainLiteral">Data Stream</Literal>
        </AnnotationAssertion>
        <AnnotationAssertion>
            <AnnotationProperty abbreviatedIRI="rdfs:label"/>
            <IRI>#ExecutionUponRequest</IRI>
            <Literal             datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-ns#PlainLiteral">Execution Upon Request</Literal>
        </AnnotationAssertion>
        <AnnotationAssertion>
            <AnnotationProperty abbreviatedIRI="rdfs:label"/>
            <IRI>#FilterOperation</IRI>
            <Literal             datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-ns#PlainLiteral">Filter Operation</Literal>
        </AnnotationAssertion>
        <AnnotationAssertion>
            <AnnotationProperty abbreviatedIRI="rdfs:label"/>
            <IRI>#FullJoin</IRI>
            <Literal             datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-ns#PlainLiteral">Full Join</Literal>
        </AnnotationAssertion>
        <AnnotationAssertion>
            <AnnotationProperty abbreviatedIRI="rdfs:label"/>
            <IRI>#InnerJoin</IRI>
            <Literal             datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-ns#PlainLiteral">Inner Join</Literal>
        </AnnotationAssertion>
        <AnnotationAssertion>
            <AnnotationProperty abbreviatedIRI="rdfs:label"/>
            <IRI>#JoinOperation</IRI>
```

```
            <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">Join Operation</Literal>
        </AnnotationAssertion>
        <AnnotationAssertion>
            <AnnotationProperty abbreviatedIRI="rdfs:label"/>
            <IRI>#LeftOuter</IRI>
            <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">Left Outer</Literal>
        </AnnotationAssertion>
        <AnnotationAssertion>
            <AnnotationProperty abbreviatedIRI="rdfs:label"/>
            <IRI>#LinearRegression</IRI>
            <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">Linear Regression</Literal>
        </AnnotationAssertion>
        <AnnotationAssertion>
            <AnnotationProperty abbreviatedIRI="rdfs:label"/>
            <IRI>#Multiway_ANOVA</IRI>
            <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">Multiway ANOVA</Literal>
        </AnnotationAssertion>
        <AnnotationAssertion>
            <AnnotationProperty abbreviatedIRI="rdfs:label"/>
            <IRI>#NeutralPosition</IRI>
            <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">Neutral Position</Literal>
        </AnnotationAssertion>
        <AnnotationAssertion>
            <AnnotationProperty abbreviatedIRI="rdfs:label"/>
            <IRI>#One_way_ANOVA</IRI>
            <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">One way ANOVA</Literal>
        </AnnotationAssertion>
        <AnnotationAssertion>
            <AnnotationProperty abbreviatedIRI="rdfs:label"/>
            <IRI>#OpposingPosition</IRI>
            <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">Opposing Position</Literal>
```

```
        </AnnotationAssertion>
        <AnnotationAssertion>
            <AnnotationProperty abbreviatedIRI="rdfs:label"/>
            <IRI>#OutputDataSpecification</IRI>
            <Literal             datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">Output Data Specification
            </Literal>
        </AnnotationAssertion>
        <AnnotationAssertion>
            <AnnotationProperty abbreviatedIRI="rdfs:label"/>
            <IRI>#PeriodicExecution</IRI>
            <Literal             datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">Periodic Execution</Literal>
        </AnnotationAssertion>
        <AnnotationAssertion>
            <AnnotationProperty abbreviatedIRI="rdfs:label"/>
            <IRI>#PolicyAction</IRI>
            <Literal             datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">Policy Action</Literal>
        </AnnotationAssertion>
        <AnnotationAssertion>
            <AnnotationProperty abbreviatedIRI="rdfs:label"/>
            <IRI>#PolicyModel</IRI>
            <Literal             datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">Policy Model</Literal>
        </AnnotationAssertion>
        <AnnotationAssertion>
            <AnnotationProperty abbreviatedIRI="rdfs:label"/>
            <IRI>#Project_Operation</IRI>
            <Literal             datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">Project Operation</Literal>
        </AnnotationAssertion>
        <AnnotationAssertion>
            <AnnotationProperty abbreviatedIRI="rdfs:label"/>
            <IRI>#RandomSampling</IRI>
            <Literal             datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">Random Sampling</Literal>
        </AnnotationAssertion>
```

```
        <AnnotationAssertion>
            <AnnotationProperty abbreviatedIRI="rdfs:label"/>
            <IRI>#Repeated_Measure_ANOVA</IRI>
            <Literal             datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">Repeated Measure ANOVA</Literal>
        </AnnotationAssertion>
        <AnnotationAssertion>
            <AnnotationProperty abbreviatedIRI="rdfs:label"/>
            <IRI>#RightOuter</IRI>
            <Literal             datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">Right Outer</Literal>
        </AnnotationAssertion>
        <AnnotationAssertion>
            <AnnotationProperty abbreviatedIRI="rdfs:label"/>
            <IRI>#SamplingOperation</IRI>
            <Literal             datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">Sampling Operation</Literal>
        </AnnotationAssertion>
        <AnnotationAssertion>
            <AnnotationProperty abbreviatedIRI="rdfs:label"/>
            <IRI>#ShiftingDataset</IRI>
            <Literal             datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">Shifting Dataset</Literal>
        </AnnotationAssertion>
        <AnnotationAssertion>
            <AnnotationProperty abbreviatedIRI="rdfs:label"/>
            <IRI>#StaticSet</IRI>
            <Literal             datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">Static Set</Literal>
        </AnnotationAssertion>
        <AnnotationAssertion>
            <AnnotationProperty abbreviatedIRI="rdfs:label"/>
            <IRI>#StatisticalAnalysisAlgorithm</IRI>
            <Literal             datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">Statistical Analysis Algorithm</Literal>
        </AnnotationAssertion>
        <AnnotationAssertion>
            <AnnotationProperty abbreviatedIRI="rdfs:label"/>
```

```
        <IRI>#StatisticalAnalysisTask</IRI>
        <Literal             datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">Statistical Analysis Task</Literal>
    </AnnotationAssertion>
    <AnnotationAssertion>
        <AnnotationProperty abbreviatedIRI="rdfs:label"/>
        <IRI>#StratifiedRandomSampling</IRI>
        <Literal             datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">Stratified Random Sampling</Literal>
    </AnnotationAssertion>
    <AnnotationAssertion>
        <AnnotationProperty abbreviatedIRI="rdfs:label"/>
        <IRI>#SupervisedDataMiningAlgorithm</IRI>
        <Literal             datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">Supervised Data Mining Algorithm</Literal>
    </AnnotationAssertion>
    <AnnotationAssertion>
        <AnnotationProperty abbreviatedIRI="rdfs:label"/>
        <IRI>#SupportivePosition</IRI>
        <Literal             datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">Supportive Position</Literal>
    </AnnotationAssertion>
    <AnnotationAssertion>
        <AnnotationProperty abbreviatedIRI="rdfs:label"/>
        <IRI>#TimedDataset</IRI>
        <Literal             datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">Timed Dataset</Literal>
    </AnnotationAssertion>
    <AnnotationAssertion>
        <AnnotationProperty abbreviatedIRI="rdfs:label"/>
        <IRI>#Two_Way_ANOVA</IRI>
        <Literal             datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">Two Way ANOVA</Literal>
    </AnnotationAssertion>
    <AnnotationAssertion>
        <AnnotationProperty abbreviatedIRI="rdfs:label"/>
        <IRI>#UnsupervisedDataMiningAlgorithm</IRI>
```

```
        <Literal            datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-ns#PlainLiteral">Unsupervised Data Mining Algorithm</Literal>
    </AnnotationAssertion>
    <AnnotationAssertion>
        <AnnotationProperty abbreviatedIRI="rdfs:label"/>
        <IRI>#WorkflowExecutionType</IRI>
        <Literal            datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-ns#PlainLiteral">Workflow Execution Type</Literal>
    </AnnotationAssertion>
    <AnnotationAssertion>
        <AnnotationProperty abbreviatedIRI="rdfs:label"/>
        <IRI>#Fisher&apos;s_Exact_Test</IRI>
        <Literal            datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-ns#PlainLiteral">Fisher&apos;s Exact Test</Literal>
    </AnnotationAssertion>
</Ontology>



<!-- Generated by the OWL API (version 4.2.8.20170104-2310) https://github.com/owlcs/owlapi -->
```

Appendix C:

Generation of Synthetic Data

In this appendix we include the Python Code for the generation of synthetic data. We have also copied a sample of the generated SQL file.

C.1. Python Code for the Generation of Synthetic Data SQL file

The following code generates SQL statements for the insertion of entries to the Patient table and the Q_DRMED table and the RETRO_HA table of the EVOTION Data Repository. The nominal data generated follow a normal distribution and the numeric data a uniform distribution.

```
import sys
import random
from random import randint
import datetime

def generateSQL():

    employmentTypeChoices=['Regular','Part-time']
    educationalLevelChoices=['Level1','Level2','Level3']
    ageChoices=['<50','[50-60)','[60-70)','[70-80)','>=80']
    print('Creating the sql file...')

    try:
            lines=int(input('Number of lines:'))
            file = open('thesis_draft_data_new.sql','w')   # Trying to create a new
file or open one
            count=1
            while(count<=lines):
                    PATIENT_ID='TEST_'+str(count)
                    DATE_OF_BIRTH=str(randint(1918,1978))+'-'+str(randint(1,12))+'-
01'
                    EDU_LEVEL=random.choice(educationalLevelChoices)
                    WORKING_IN_NOISE=str(random.randint(0,1))
                    WORKING_IN_GROUPS=str(random.randint(0,1))
                    EMPLOYMENT_TYPE=random.choice(employmentTypeChoices)
                    AVERAGE_HA_DAILY_USAGE=str(randint(0,57600))
```

```
                    file.write('UPSERT INTO PATIENT (PATIENT_ID, DATEOFBIRTH) VALUES
(\''+PATIENT_ID+'\',\''+DATE_OF_BIRTH+'\');\n')
                    file.write('UPSERT  INTO  Q_DRMED  (ID,  CREATE_DATE,  PATIENT_ID,
LS_EDUC_PLACEM,      LS_EMPL_SIT1,      LS_EMPL_SIT2,      LS_EMPL_TYPE)      VALUES
('+str(count)+',\''+str(datetime.datetime.now())+'\',\''+PATIENT_ID+'\','+'\''+EDU_L
EVEL+'\','+WORKING_IN_NOISE+','+WORKING_IN_GROUPS+',\''+EMPLOYMENT_TYPE+'\');\n')
                    file.write('UPSERT  INTO  RETRO_HA  (PATIENT_ID,  TOTAL_USE)  VALUES
(\''+PATIENT_ID+'\','+AVERAGE_HA_DAILY_USAGE+');\n')
                    count+=1
             file.close()
             print('The sql file was successfully created.')

      except Exception as e:
             print(e)
             sys.exit(0) # quit Python

  generateSQL()
```

C.2. Generated SQL file Sample

The following sample includes 3 SQL UPSERT statements, one to the PATIENT table, one to the Q_RMED table and one to the RETRO_HA table of the EVOTION Data Repository. We generated 1000000 entries to each table (3000000 UPSERT statements in total) for the insertion of random synthetic data to the EVOTION Data Repository.

```
  UPSERT INTO PATIENT (PATIENT_ID, DATEOFBIRTH) VALUES ('TEST_1','1952-12-01');
  UPSERT INTO Q_DRMED (ID, CREATE_DATE, PATIENT_ID, LS_EDUC_PLACEM, LS_EMPL_SIT1,
LS_EMPL_SIT2, LS_EMPL_TYPE) VALUES (1,'2018-03-21
19:27:23.644000','TEST_1','Level3',1,1,'Part-time');
  UPSERT INTO RETRO_HA (PATIENT_ID, TOTAL_USE) VALUES ('TEST_1',43807);
```

Appendix D:

OWL XML Definition of instances for the Evaluation Scenario

In this appendix we present the formal definition of the Public Health Policy Decision Making Model used in the performance evaluation (section 5.4).

GOAL: Addressing Barriers to HA Use

- Description: The purpose of this case study is to determine the largest barriers that affect hearing aid use in a population in order to make public health policy decisions to address them.
- Rationale: Barriers to hearing aid use are a significant public health problem. Barriers occur at all levels of the process of provision of hearing aids including at the level of the HA user. The big data gathered about users through EVOTION would enable policy makers to choose which barriers to address in a population, in order to improve hearing aid use and hence reduce the burden of hearing loss in that population.

OBJECTIVES:

- To intervene in order to address prevention of HA usage due to occupation.
- To intervene in order to address prevention of HA usage due to education level.
- To intervene in order to address prevention of HA usage due to age.

POLICY ACTIONS:

- Occupation Related ACTION: A particular occupation has to be addressed with additional measures to improve HA use;
- Educational level related ACTION: Failure to reach a particular educational level has to be addressed to improve HA use.
- Age related ACTION: Age related fitting;

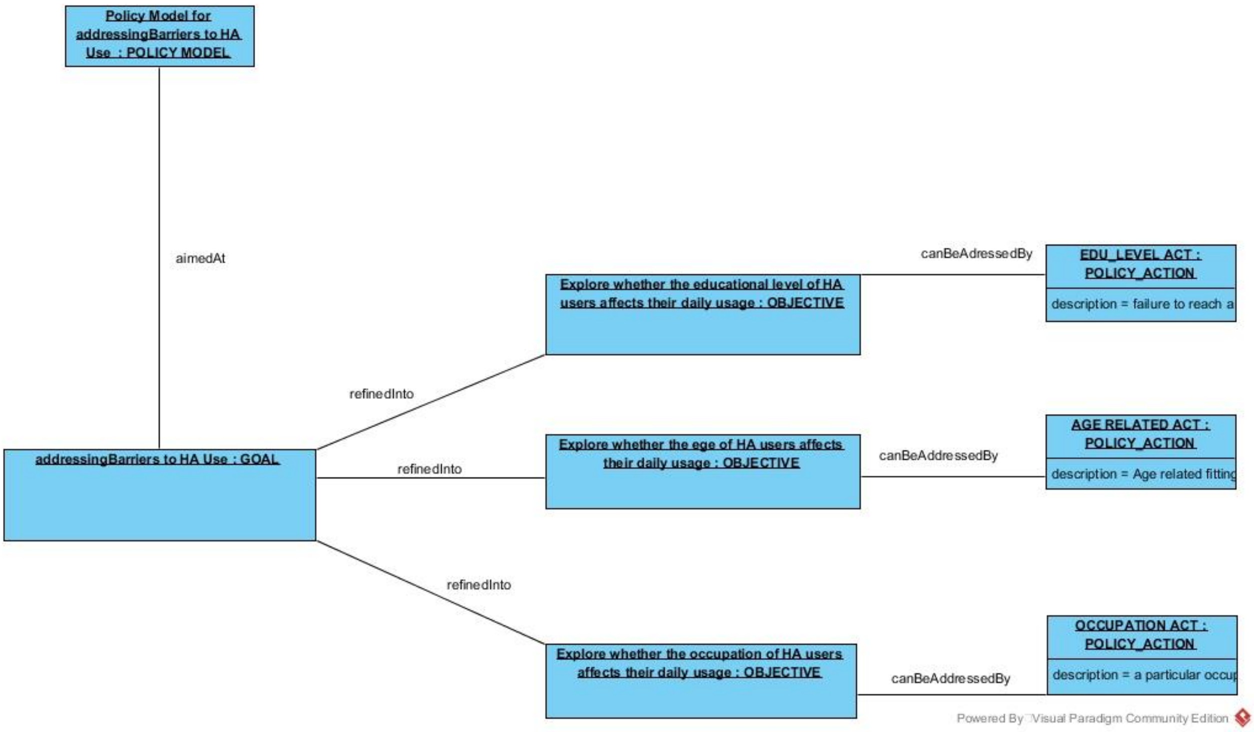


Figure 118 Case Study: Addressing Barriers to HA use: Policy Model, Goal, Objectives and Policy Actions

For the definition of these instances of the classes in OWL we have created the following:

- an instance of the `PolicyModel` class named “PM_1” with label “Policy Model of Addressing Barriers to HA Use”,
- an instance of the `Goal` class named “Goal_1” with label “Addressing Barriers to HA Use”,
- three instances of the `Objective` class: one named “Obj_1” with label “Explore whether the occupation of HA users affects their daily usage”, one named “Obj_2” with label “Explore whether the educational level of HA users affects their daily usage” and one named “obj_3” with label “Explore whether the age of HA users affects their daily usage” and
- three instances of the `PolicyAction` class: one named “PA_1” with label “Occupation Related ACTION”, one named “PA_2” with label “Explore whether the educational level of HA users affects their daily usage” and one named “PA_3” with label “Explore whether the age of HA users affects their daily usage”.

```
<Declaration>
        <NamedIndividual IRI="#PM_1"/>
    </Declaration>
    <ClassAssertion>
        <Class IRI="#PolicyModel"/>
        <NamedIndividual IRI="#PM_1"/>
     </ClassAssertion>
```

```xml
     <AnnotationAssertion>
        <AnnotationProperty abbreviatedIRI="rdfs:label"/>
        <IRI>#PM_1</IRI>
        <LiteraldatatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax
ns#PlainLiteral">Policy Model of Addressing Barriers to HA Use</Literal>
    </AnnotationAssertion>
<Declaration>
        <NamedIndividual IRI="#Goal_1"/>
    </Declaration>
<ClassAssertion>
        <Class IRI="#Goal"/>
        <NamedIndividual IRI="#Goal_1"/>
    </ClassAssertion>
<AnnotationAssertion>
        <AnnotationProperty abbreviatedIRI="rdfs:label"/>
        <IRI>#Goal_1</IRI>
        <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">Addressing Barriers to HA Use</Literal>
    </AnnotationAssertion>
<DataPropertyAssertion>
        <DataProperty IRI="#description"/>
        <NamedIndividual IRI="#Goal_1"/>
        <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">The purpose of this case study is to determine the largest
barriers that affect hearing aid use in a population in order to make public
health policy decisions to address them.</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
        <DataProperty IRI="#rationale"/>
        <NamedIndividual IRI="#Goal_1"/>
        <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">Barriers to hearing aid use are a significant public health
problem. Barriers occur at all levels of the process of provision of hearing
aids including at the level of the HA user. The big data gathered about users
through EVOTION would enable policy makers to choose which barriers to address
in a population, in order to improve hearing aid use and hence reduce the burden
of hearing loss in that population.</Literal>
    </DataPropertyAssertion>
```

```
<Declaration>
        <NamedIndividual IRI="#Obj_1"/>
    </Declaration>
<Declaration>
        <NamedIndividual IRI="#Obj_2"/>
    </Declaration>
<Declaration>
        <NamedIndividual IRI="#Obj_3"/>
    </Declaration>
<ClassAssertion>
        <Class IRI="#Objective"/>
        <NamedIndividual IRI="#Obj_1"/>
    </ClassAssertion>
    <ClassAssertion>
        <Class IRI="#Objective"/>
        <NamedIndividual IRI="#Obj_2"/>
    </ClassAssertion>
    <ClassAssertion>
        <Class IRI="#Objective"/>
        <NamedIndividual IRI="#Obj_3"/>
    </ClassAssertion>
<AnnotationAssertion>
        <AnnotationProperty abbreviatedIRI="rdfs:label"/>
        <IRI>#Obj_1</IRI>
        <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">Explore whether the occupation of HA users affects their daily
usage</Literal>
    </AnnotationAssertion>
    <AnnotationAssertion>
        <AnnotationProperty abbreviatedIRI="rdfs:label"/>
        <IRI>#Obj_2</IRI>
        <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">Explore whether the educational level of HA users affects their
daily usage</Literal>
    </AnnotationAssertion>
    <AnnotationAssertion>
        <AnnotationProperty abbreviatedIRI="rdfs:label"/>
        <IRI>#Obj_3</IRI>
```

```
        <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">Explore whether the age of HA users affects their daily
usage</Literal>
    </AnnotationAssertion>
<Declaration>
        <NamedIndividual IRI="#PA_1"/>
    </Declaration>
<Declaration>
        <NamedIndividual IRI="#PA_2"/>
    </Declaration>
<Declaration>
        <NamedIndividual IRI="#PA_3"/>
    </Declaration>
<ClassAssertion>
        <Class IRI="#PolicyAction"/>
        <NamedIndividual IRI="#PA_1"/>
    </ClassAssertion>
    <ClassAssertion>
        <Class IRI="#PolicyAction"/>
        <NamedIndividual IRI="#PA_2"/>
    </ClassAssertion>
    <ClassAssertion>
        <Class IRI="#PolicyAction"/>
        <NamedIndividual IRI="#PA_3"/>
    </ClassAssertion>
<AnnotationAssertion>
        <AnnotationProperty abbreviatedIRI="rdfs:label"/>
        <IRI>#PA_1</IRI>
        <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">Occupation Related ACTION</Literal>
    </AnnotationAssertion>
    <AnnotationAssertion>
        <AnnotationProperty abbreviatedIRI="rdfs:label"/>
        <IRI>#PA_2</IRI>
        <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">Age related ACTION</Literal>
    </AnnotationAssertion>
    <AnnotationAssertion>
```

```
        <AnnotationProperty abbreviatedIRI="rdfs:label"/>
        <IRI>#PA_3</IRI>
        <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">Educational level related ACTION</Literal>
    </AnnotationAssertion>
<DataPropertyAssertion>
        <DataProperty IRI="#description"/>
        <NamedIndividual IRI="#PA_1"/>
        <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">A particular occupation has to be addressed with additional
measures to improve HA use</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
        <DataProperty IRI="#description"/>
        <NamedIndividual IRI="#PA_2"/>
        <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">Age related fitting</Literal>
    </DataPropertyAssertion>
    <DataPropertyAssertion>
        <DataProperty IRI="#description"/>
        <NamedIndividual IRI="#PA_3"/>
        <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">Failure to reach a particular educational level has to be
addressed to improve HA use</Literal>
    </DataPropertyAssertion>
```

Below we present the formal definition of the following:

- the relations of PM_1: that PM_1 has the relation of type `hasWorkflows` with DAW1, and the relation of type `aimedAt` with Goal_1,
- the relations of Goal_1: Goal_1 has the relation of type `refinedInto` with Obj_1, Obj_2 and Obj_3
- the relations of PA_1 with Obj_1, PA_2 with Obj_2 and PA_3 with Obj_3 of type `canBeAddressedBy` and the relations of PA_1 with OccupationCriterion, PA_2 with EducationalLevelCriterion and PA_3 with AgeCriterion of type `isEvaluatedBy`

```
    <ObjectPropertyAssertion>
        <ObjectProperty IRI="#hasWorkflows"/>
        <NamedIndividual IRI="#PM_1"/>
        <NamedIndividual IRI="#DAW1"/>
```

```
        </ObjectPropertyAssertion>
            <ObjectProperty IRI="#aimedAt"/>
            <NamedIndividual IRI="#PM_1"/>
            <NamedIndividual IRI="#Goal_1"/>
        </ObjectPropertyAssertion>
    <ObjectPropertyAssertion>
            <ObjectProperty IRI="#refinedInto"/>
            <NamedIndividual IRI="#Goal_1"/>
            <NamedIndividual IRI="#Obj_1"/>
        </ObjectPropertyAssertion>
        <ObjectPropertyAssertion>
            <ObjectProperty IRI="#refinedInto"/>
            <NamedIndividual IRI="#Goal_1"/>
            <NamedIndividual IRI="#Obj_2"/>
        </ObjectPropertyAssertion>
        <ObjectPropertyAssertion>
            <ObjectProperty IRI="#refinedInto"/>
            <NamedIndividual IRI="#Goal_1"/>
            <NamedIndividual IRI="#Obj_3"/>
      </ObjectPropertyAssertion>
    <ObjectPropertyAssertion>
            <ObjectProperty IRI="#refinedInto"/>
            <NamedIndividual IRI="#Goal_1"/>
            <NamedIndividual IRI="#Obj_1"/>
        </ObjectPropertyAssertion>
        <ObjectPropertyAssertion>
            <ObjectProperty IRI="#refinedInto"/>
            <NamedIndividual IRI="#Goal_1"/>
            <NamedIndividual IRI="#Obj_2"/>
        </ObjectPropertyAssertion>
        <ObjectPropertyAssertion>
            <ObjectProperty IRI="#refinedInto"/>
            <NamedIndividual IRI="#Goal_1"/>
            <NamedIndividual IRI="#Obj_3"/>
        </ObjectPropertyAssertion>
    <ObjectPropertyAssertion>
            <ObjectProperty IRI="#isEvaluatedBy"/>
            <NamedIndividual IRI="#PA_1"/>
            <NamedIndividual IRI="#OccupationCriterion"/>
```

```
    </ObjectPropertyAssertion>
    <ObjectPropertyAssertion>
        <ObjectProperty IRI="#isEvaluatedBy"/>
        <NamedIndividual IRI="#PA_2"/>
        <NamedIndividual IRI="#EducationalLevelCriterion"/>
    </ObjectPropertyAssertion>
    <ObjectPropertyAssertion>
        <ObjectProperty IRI="#isEvaluatedBy"/>
        <NamedIndividual IRI="#PA_3"/>
        <NamedIndividual IRI="#AgeCriterion"/>
  </ObjectPropertyAssertion>
```

STAKEHOLDERS

Representatives of:

- Regional ENT-specialists' Advisory Committee (in their role as prescribing the use of HAs);
- Regional Directorate for Social support (in their role as authorising financial support for purchasing HAs and performing follow-up on administration and use);
- Regional structures of the national Health Insurance Fund (in their role as funding clinical pathways);
- HA vendors/fitting experts (providing follow-up rehab);
- Patients' association – regional repres. of patients

POSITIONS:

- Supportive_1, that refers to PA_1, proposed by Regional ENT-specialists' Advisory Committee and supported by them, as well as HA vendors/fitting experts.
- Opposing_1, that refers to PA_1, proposed by Regional Directorate for Social support and supported by them, as well as the Patients' association.
- Neutral_1, that refers to PA_1, proposed by Regional structures of the national Health Insurance Fund and supported by them,
- Supportive_2, that refers to PA_2, proposed by HA vendors/fitting experts and supported by them, as well as Regional ENT-specialists' Advisory Committee,
- Opposing_2, that refers to PA_2, proposed by Regional structures of the national Health Insurance Fund and supported by them, as well as the Regional Directorate for Social support,
- Neutral_2, that refers to PA_2, proposed by Patients' association and supported by them,
- Supportive_3, that refers to PA_3, proposed by Patients' association and supported by them,
- Opposing_3, that refers to PA_3, proposed by HA vendors/fitting and supported by them, as well as the Regional Directorate for Social support, and
- Neutral_3, that refers to PA_3, proposed by Regional structures of the national Health Insurance Fund and supported by them.

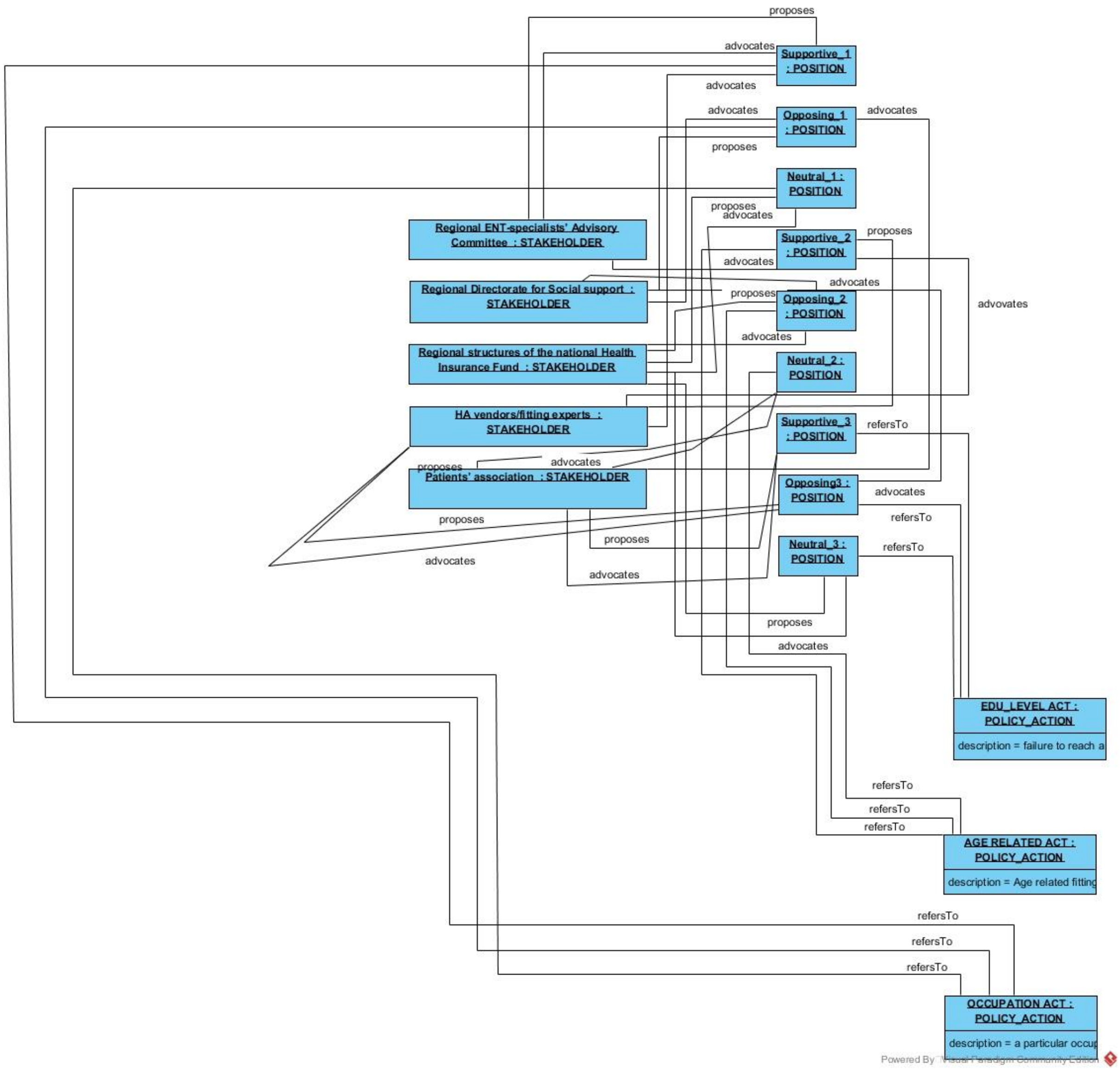


Figure 119 Case Study: Addressing Barriers to HA use: Stakeholders, Positions and Policy Actions

For the definition of these instances of the classes in OWL we have created the following:

- five instances of the `Stakeholder` class: one named "Stakeholder_1" with label "Regional ENT-specialists' Advisory Committee", one named "Stakeholder_2" with label "Regional Directorate for Social support", one named "Stakeholder_3" with label "Regional structures of the national Health Insurance Fund", one named "Stakeholder_4" with label "`HA vendors/fitting experts`" and one named "Stakeholder_5" with label "Regional repres. of patients", and
- nine instances of the subclasses of the `Position` class: three named "Supportive_1", "Supportive_2" and "Supportive_3", instances of the SupportivePosition class, three named "Opposing_1", "Opposing_2" and "Opposing_3", instances of the OpposingPosition class and three named "Neutral_1", "Neutral_2" and "Neutral_3",instances of the NeutralPosition class.

```
<Declaration>
        <NamedIndividual IRI="#Stakeholder_1"/>
    </Declaration>
<Declaration>
        <NamedIndividual IRI="#Stakeholder_2"/>
    </Declaration>
<Declaration>
        <NamedIndividual IRI="#Stakeholder_3"/>
    </Declaration>
<Declaration>
        <NamedIndividual IRI="#Stakeholder_4"/>
    </Declaration>
<Declaration>
        <NamedIndividual IRI="#Stakeholder_5"/>
    </Declaration>
<ClassAssertion>
        <Class IRI="#Stakeholder"/>
        <NamedIndividual IRI="#Stakeholder_1"/>
    </ClassAssertion>
<ClassAssertion>
        <Class IRI="#Stakeholder"/>
        <NamedIndividual IRI="#Stakeholder_2"/>
    </ClassAssertion>
<ClassAssertion>
        <Class IRI="#Stakeholder"/>
        <NamedIndividual IRI="#Stakeholder_3"/>
    </ClassAssertion>
<ClassAssertion>
        <Class IRI="#Stakeholder"/>
        <NamedIndividual IRI="#Stakeholder_4"/>
    </ClassAssertion>
<ClassAssertion>
        <Class IRI="#Stakeholder"/>
        <NamedIndividual IRI="#Stakeholder_5"/>
    </ClassAssertion>
<AnnotationAssertion>
        <AnnotationProperty abbreviatedIRI="rdfs:comment"/>
```

```
          <IRI>#Stakeholder_1</IRI>
          <Literal         datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-ns#PlainLiteral">in their role as prescribing the use of HAs</Literal>
      </AnnotationAssertion>
      <AnnotationAssertion>
          <AnnotationProperty abbreviatedIRI="rdfs:label"/>
          <IRI>#Stakeholder_1</IRI>
          <Literal         datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-ns#PlainLiteral">Regional ENT-specialists’ Advisory Committee</Literal>
      </AnnotationAssertion>
      <AnnotationAssertion>
          <AnnotationProperty abbreviatedIRI="rdfs:comment"/>
          <IRI>#Stakeholder_2</IRI>
          <Literal         datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-ns#PlainLiteral">in their role as authorising financial support for purchasing HAs and performing follow-up on ad-ministration and use</Literal>
      </AnnotationAssertion>
      <AnnotationAssertion>
          <AnnotationProperty abbreviatedIRI="rdfs:label"/>
          <IRI>#Stakeholder_2</IRI>
          <Literal         datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-ns#PlainLiteral">Regional Directorate for Social support</Literal>
      </AnnotationAssertion>
      <AnnotationAssertion>
          <AnnotationProperty abbreviatedIRI="rdfs:comment"/>
          <IRI>#Stakeholder_3</IRI>
          <Literal         datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-ns#PlainLiteral">in their role as funding clinical pathways</Literal>
      </AnnotationAssertion>
      <AnnotationAssertion>
          <AnnotationProperty abbreviatedIRI="rdfs:label"/>
          <IRI>#Stakeholder_3</IRI>
          <Literal         datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-ns#PlainLiteral">Regional structures of the national Health Insurance Fund</Literal>
      </AnnotationAssertion>
      <AnnotationAssertion>
          <AnnotationProperty abbreviatedIRI="rdfs:comment"/>
```

```
            <IRI>#Stakeholder_4</IRI>
            <Literal         datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-ns#PlainLiteral">providing follow-up rehab</Literal>
        </AnnotationAssertion>
        <AnnotationAssertion>
            <AnnotationProperty abbreviatedIRI="rdfs:label"/>
            <IRI>#Stakeholder_4</IRI>
            <Literal         datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-ns#PlainLiteral">HA vendors/fitting experts</Literal>
        </AnnotationAssertion>
        <AnnotationAssertion>
            <AnnotationProperty abbreviatedIRI="rdfs:comment"/>
            <IRI>#Stakeholder_5</IRI>
            <Literal         datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-ns#PlainLiteral">regional repres. of patients</Literal>
        </AnnotationAssertion>
        <AnnotationAssertion>
            <AnnotationProperty abbreviatedIRI="rdfs:label"/>
            <IRI>#Stakeholder_5</IRI>
            <Literal         datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-ns#PlainLiteral">Patients' association</Literal>
        </AnnotationAssertion>
    <Declaration>
            <NamedIndividual IRI="#Supportive_1"/>
        </Declaration>
    <ClassAssertion>
            <Class IRI="#SupportivePosition"/>
            <NamedIndividual IRI="#Supportive_1"/>
        </ClassAssertion>
    <Declaration>
            <NamedIndividual IRI="#Opposing_1"/>
        </Declaration>
    <ClassAssertion>
            <Class IRI="#OpposingPosition"/>
            <NamedIndividual IRI="#Opposing_1"/>
        </ClassAssertion>
    <Declaration>
            <NamedIndividual IRI="#Neutral_1"/>
        </Declaration>
```

```
<ClassAssertion>
        <Class IRI="#NeutralPosition"/>
        <NamedIndividual IRI="#Neutral_1"/>
  </ClassAssertion>
<Declaration>
        <NamedIndividual IRI="#Supportive_2"/>
    </Declaration>
<ClassAssertion>
        <Class IRI="#SupportivePosition"/>
        <NamedIndividual IRI="#Supportive_2"/>
    </ClassAssertion>
<Declaration>
        <NamedIndividual IRI="#Opposing_2"/>
    </Declaration>
<ClassAssertion>
        <Class IRI="#OpposingPosition"/>
        <NamedIndividual IRI="#Opposing_2"/>
  </ClassAssertion>
<Declaration>
        <NamedIndividual IRI="#Neutral_2"/>
    </Declaration>
<ClassAssertion>
        <Class IRI="#NeutralPosition"/>
        <NamedIndividual IRI="#Neutral_2"/>
  </ClassAssertion>
<Declaration>
        <NamedIndividual IRI="#Supportive_3"/>
    </Declaration>
<ClassAssertion>
        <Class IRI="#SupportivePosition"/>
        <NamedIndividual IRI="#Supportive_3"/>
    </ClassAssertion>
<Declaration>
        <NamedIndividual IRI="#Opposing_3"/>
    </Declaration>
<ClassAssertion>
        <Class IRI="#OpposingPosition"/>
        <NamedIndividual IRI="#Opposing_3"/>
  </ClassAssertion>
```

```
<Declaration>
        <NamedIndividual IRI="#Neutral_3"/>
    </Declaration>
<ClassAssertion>
        <Class IRI="#NeutralPosition"/>
        <NamedIndividual IRI="#Neutral_3"/>
  </ClassAssertion>
```

Below we present the formal definition of the following:

- the relations of PM_1: that PM_1 has the relation of type `involvesStakeholders` with Stakeholder_1, Stakeholder_2 Stakeholder_3, Stakeholder_4 and Stakeholder_5,
- the relations that Suppotive_1, Opposing_1 and Neutral_1 have the relation of type `refersTo` with Policy Action PA_1, that Suppotive_1 has the relation of type `proposes` with Stakeholder_1 and the relationship of type `advocates` with Stakeholder_1 and Stakeholder_4, Opposing_1 has the relationship of type `proposes` with Stakeholder_2 and the relationship of type `advocates` with Stakeholder_2 and Stakeholder_5 and Neutral_1 has the relationship of type `proposes` with Stakeholder_3 and the relationship of type `advocates` with Stakeholder_3
- the relations that Suppotive_2, Opposing_2 and Neutral_2 have the relation of type `refersTo` with Policy Action PA_2, that Suppotive_1 has the relation of type `proposes` with Stakeholder_2 and the relationship of type `advocates` with Stakeholder_2 and Stakeholder_3, Opposing_2 has the relationship of type `proposes` with Stakeholder_1 and the relationship of type `advocates` with Stakeholder_1 and Stakeholder_3 and Neutral_2 has the relationship of type `proposes` with Stakeholder_5 and the relationship of type `advocates` with Stakeholder_5
- the relations that Suppotive_1, Opposing_1 and Neutral_1 have the relation of type `refersTo` with Policy Action PA_1, that Suppotive_1 has the relation of type `proposes` with Stakeholder_1 and the relationship of type `advocates` with Stakeholder_1 and Stakeholder_4, Opposing_1 has the relationship of type `proposes` with Stakeholder_2 and the relationship of type `advocates` with Stakeholder_2 and Stakeholder_5 and Neutral_1 has the relationship of type `proposes` with Stakeholder_3 and the relationship of type `advocates` with Stakeholder_3

```
<ObjectPropertyAssertion>
        <ObjectProperty IRI="#involvesStakeholders"/>
        <NamedIndividual IRI="#PM_1"/>
        <NamedIndividual IRI="#Stakeholder_1"/>
    </ObjectPropertyAssertion>
    <ObjectPropertyAssertion>
        <ObjectProperty IRI="#involvesStakeholders"/>
```

```
        <NamedIndividual IRI="#PM_1"/>
        <NamedIndividual IRI="#Stakeholder_2"/>
    </ObjectPropertyAssertion>
    <ObjectPropertyAssertion>
        <ObjectProperty IRI="#involvesStakeholders"/>
        <NamedIndividual IRI="#PM_1"/>
        <NamedIndividual IRI="#Stakeholder_3"/>
       </ObjectPropertyAssertion>
    <ObjectPropertyAssertion>
        <ObjectProperty IRI="#involvesStakeholders"/>
        <NamedIndividual IRI="#PM_1"/>
        <NamedIndividual IRI="#Stakeholder_4"/>
       </ObjectPropertyAssertion>
    <ObjectPropertyAssertion>
        <ObjectProperty IRI="#involvesStakeholders"/>
        <NamedIndividual IRI="#PM_1"/>
        <NamedIndividual IRI="#Stakeholder_5"/>
       </ObjectPropertyAssertion>
    <ObjectPropertyAssertion>
        <ObjectProperty IRI="#refersTo"/>
        <NamedIndividual IRI="#PA_1"/>
        <NamedIndividual IRI="#Neutral_1"/>
    </ObjectPropertyAssertion>
    <ObjectPropertyAssertion>
        <ObjectProperty IRI="#refersTo"/>
        <NamedIndividual IRI="#PA_1"/>
        <NamedIndividual IRI="#Opposing_1"/>
    </ObjectPropertyAssertion>
    <ObjectPropertyAssertion>
        <ObjectProperty IRI="#refersTo"/>
        <NamedIndividual IRI="#PA_1"/>
        <NamedIndividual IRI="#Supportive_1"/>
    </ObjectPropertyAssertion>
     <ObjectPropertyAssertion>
         <ObjectProperty IRI="#proposes"/>
         <NamedIndividual IRI="#Stakeholder_1"/>
         <NamedIndividual IRI="#Supportive_1"/>
     </ObjectPropertyAssertion>
```

```
    <ObjectPropertyAssertion>
        <ObjectProperty IRI="#advocates"/>
        <NamedIndividual IRI="#Stakeholder_1"/>
        <NamedIndividual IRI="#Supportive_1"/>
    </ObjectPropertyAssertion>
    <ObjectPropertyAssertion>
        <ObjectProperty IRI="#advocates"/>
        <NamedIndividual IRI="#Stakeholder_4"/>
        <NamedIndividual IRI="#Supportive_1"/>
    </ObjectPropertyAssertion>
    <ObjectPropertyAssertion>
        <ObjectProperty IRI="#proposes"/>
        <NamedIndividual IRI="#Stakeholder_2"/>
        <NamedIndividual IRI="#Opposing_1"/>
    </ObjectPropertyAssertion>
    <ObjectPropertyAssertion>
        <ObjectProperty IRI="#advocates"/>
        <NamedIndividual IRI="#Stakeholder_2"/>
        <NamedIndividual IRI="#Opposing_1"/>
    </ObjectPropertyAssertion>
    <ObjectPropertyAssertion>
        <ObjectProperty IRI="#advocates"/>
        <NamedIndividual IRI="#Stakeholder_5"/>
        <NamedIndividual IRI="#Opposing_1"/>
    </ObjectPropertyAssertion>
    <ObjectPropertyAssertion>
        <ObjectProperty IRI="#proposes"/>
        <NamedIndividual IRI="#Stakeholder_3"/>
        <NamedIndividual IRI="#Neutral_1"/>
    </ObjectPropertyAssertion>
    <ObjectPropertyAssertion>
        <ObjectProperty IRI="#advocates"/>
        <NamedIndividual IRI="#Stakeholder_3"/>
        <NamedIndividual IRI="#Neutral_1"/>
</ObjectPropertyAssertion>
    <ObjectPropertyAssertion>
       <ObjectProperty IRI="#refersTo"/>
       <NamedIndividual IRI="#PA_2"/>
       <NamedIndividual IRI="#Neutral_2"/>
```

```
    </ObjectPropertyAssertion>
    <ObjectPropertyAssertion>
        <ObjectProperty IRI="#refersTo"/>
        <NamedIndividual IRI="#PA_2"/>
        <NamedIndividual IRI="#Opposing_2"/>
    </ObjectPropertyAssertion>
    <ObjectPropertyAssertion>
        <ObjectProperty IRI="#refersTo"/>
        <NamedIndividual IRI="#PA_2"/>
        <NamedIndividual IRI="#Supportive_2"/>
    </ObjectPropertyAssertion>
     <ObjectPropertyAssertion>
         <ObjectProperty IRI="#proposes"/>
         <NamedIndividual IRI="#Stakeholder_4"/>
         <NamedIndividual IRI="#Supportive_2"/>
     </ObjectPropertyAssertion>
     <ObjectPropertyAssertion>
         <ObjectProperty IRI="#advocates"/>
         <NamedIndividual IRI="#Stakeholder_4"/>
         <NamedIndividual IRI="#Supportive_2"/>
     </ObjectPropertyAssertion>
     <ObjectPropertyAssertion>
         <ObjectProperty IRI="#advocates"/>
         <NamedIndividual IRI="#Stakeholder_1"/>
         <NamedIndividual IRI="#Supportive_2"/>
     </ObjectPropertyAssertion>
     <ObjectPropertyAssertion>
         <ObjectProperty IRI="#proposes"/>
         <NamedIndividual IRI="#Stakeholder_3"/>
         <NamedIndividual IRI="#Opposing_2"/>
     </ObjectPropertyAssertion>
     <ObjectPropertyAssertion>
         <ObjectProperty IRI="#advocates"/>
         <NamedIndividual IRI="#Stakeholder_3"/>
         <NamedIndividual IRI="#Opposing_2"/>
     </ObjectPropertyAssertion>
     <ObjectPropertyAssertion>
         <ObjectProperty IRI="#advocates"/>
```

```
            <NamedIndividual IRI="#Stakeholder_2"/>
            <NamedIndividual IRI="#Opposing_2"/>
        </ObjectPropertyAssertion>
        <ObjectPropertyAssertion>
            <ObjectProperty IRI="#proposes"/>
            <NamedIndividual IRI="#Stakeholder_5"/>
            <NamedIndividual IRI="#Neutral_2"/>
        </ObjectPropertyAssertion>
        <ObjectPropertyAssertion>
            <ObjectProperty IRI="#advocates"/>
            <NamedIndividual IRI="#Stakeholder_5"/>
            <NamedIndividual IRI="#Neutral_2"/>
        </ObjectPropertyAssertion>
       <ObjectPropertyAssertion>
           <ObjectProperty IRI="#refersTo"/>
           <NamedIndividual IRI="#PA_3"/>
           <NamedIndividual IRI="#Neutral_3"/>
       </ObjectPropertyAssertion>
       <ObjectPropertyAssertion>
           <ObjectProperty IRI="#refersTo"/>
           <NamedIndividual IRI="#PA_3"/>
           <NamedIndividual IRI="#Opposing_3"/>
       </ObjectPropertyAssertion>
       <ObjectPropertyAssertion>
           <ObjectProperty IRI="#refersTo"/>
           <NamedIndividual IRI="#PA_3"/>
           <NamedIndividual IRI="#Supportive_3"/>
       </ObjectPropertyAssertion>
        <ObjectPropertyAssertion>
            <ObjectProperty IRI="#proposes"/>
            <NamedIndividual IRI="#Stakeholder_5"/>
            <NamedIndividual IRI="#Supportive_3"/>
        </ObjectPropertyAssertion>
        <ObjectPropertyAssertion>
            <ObjectProperty IRI="#advocates"/>
            <NamedIndividual IRI="#Stakeholder_5"/>
            <NamedIndividual IRI="#Supportive_3"/>
        </ObjectPropertyAssertion>
```

```
        <ObjectPropertyAssertion>
            <ObjectProperty IRI="#proposes"/>
            <NamedIndividual IRI="#Stakeholder_4"/>
            <NamedIndividual IRI="#Opposing_3"/>
        </ObjectPropertyAssertion>
        <ObjectPropertyAssertion>
            <ObjectProperty IRI="#advocates"/>
            <NamedIndividual IRI="#Stakeholder_4"/>
            <NamedIndividual IRI="#Opposing_3"/>
        </ObjectPropertyAssertion>
        <ObjectPropertyAssertion>
            <ObjectProperty IRI="#advocates"/>
            <NamedIndividual IRI="#Stakeholder_2"/>
            <NamedIndividual IRI="#Opposing_3"/>
        </ObjectPropertyAssertion>
        <ObjectPropertyAssertion>
            <ObjectProperty IRI="#proposes"/>
            <NamedIndividual IRI="#Stakeholder_3"/>
            <NamedIndividual IRI="#Neutral_3"/>
        </ObjectPropertyAssertion>
        <ObjectPropertyAssertion>
            <ObjectProperty IRI="#advocates"/>
            <NamedIndividual IRI="#Stakeholder_3"/>
            <NamedIndividual IRI="#Neutral_3"/>
        </ObjectPropertyAssertion>
```

CRITERIA

Policy making criteria:

- Model wide
    - Statistically significant model overall
    - Homoscedasticity of prediction errors
- Variable specific
    - Effect of particular variable is statistically significant

- CR1: R2 square > 0.5 (constraints.spec.Reg_Overall_Stats.R Square>0.5)
- CR2: Edu_Level P < 0.05 (constraints.spec.Reg_Detailed_Stats [Factor = Edu_Level].P-value < 0.05)
- CR3: Age P < 0.05 (constraints.spec.Reg_Detailed_Stats [Factor = Age.P-value < 0.05)

- CR4: Occup P < 0.05 (constraints.spec.Reg_Detailed_Stats [Factor = Occup].P-value < 0.05)

Logical expressions of Criteria:

- EDU_CRIT: CR2 and CR1
- AGE_CRIT: CR3 and CR1
- OCCUP_CRIT: CR4 and CR1

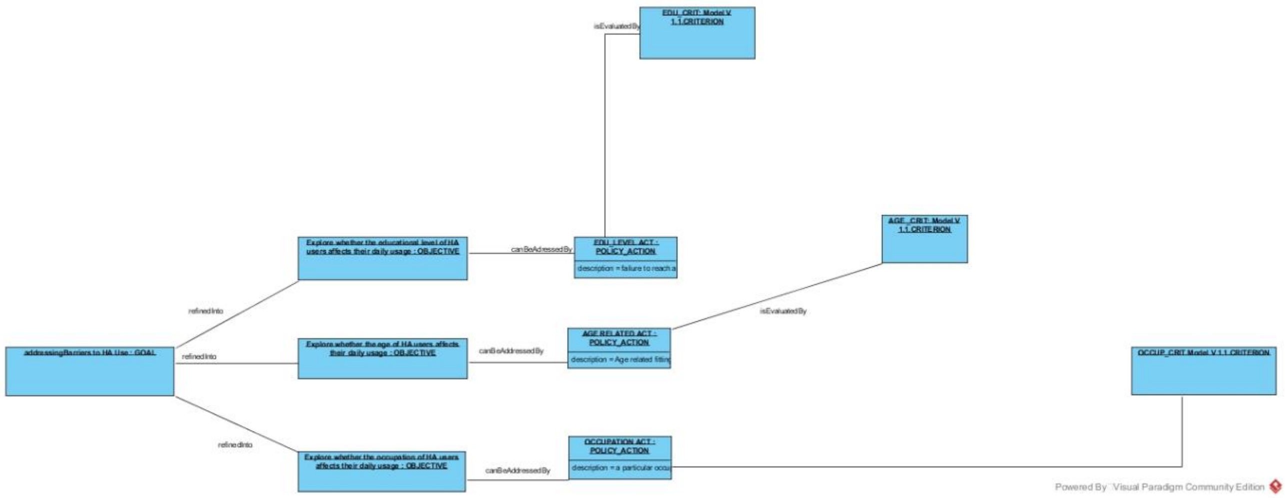


Figure 120 Case Study: Addressing Barriers to HA use: Criteria

For the definition of these instances of the classes in OWL we have created three instances of the `Criterion` class: one named “OccupationCriterion”, one named “EducationalLevelCriterion” and one named “AgeCriterion”. Since all the criteria have the same weight, we have left the weight data property empty. We have inserted the criteria logical expressions to each criterion. The declaration of these instances in OWL are presented below.

```
<Declaration>
        <NamedIndividual IRI="#OccupationCriterion"/>
    </Declaration>
<Declaration>
        <NamedIndividual IRI="#EducationalLevelCriterion"/>
    </Declaration>
<Declaration>
        <NamedIndividual IRI="#AgeCriterion"/>
    </Declaration>
<ClassAssertion>
        <Class IRI="#Criterion"/>
        <NamedIndividual IRI="#OccupationCriterion"/>
    </ClassAssertion>
```

```
   <ClassAssertion>
           <Class IRI="#Criterion"/>
           <NamedIndividual IRI="#EducationalLevelCriterion"/>
       </ClassAssertion>
   <ClassAssertion>
           <Class IRI="#Criterion"/>
           <NamedIndividual IRI="#AgeCriterion"/>
       </ClassAssertion>
   <DataPropertyAssertion>
           <DataProperty IRI="#logicalExpression"/>
           <NamedIndividual IRI="#OccupationCriterion"/>
           <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">constraints.spec.Reg_Detailed_Stats [Factor = Occup].P-value
&lt; 0.05
   AND
   constraints.spec.Reg_Overall_Stats.R Square&gt;0.5</Literal>
       </DataPropertyAssertion>

   <DataPropertyAssertion>
           <DataProperty IRI="#logicalExpression"/>
           <NamedIndividual IRI="#EducationalLevelCriterion"/>
           <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">constraints.spec.Reg_Detailed_Stats [Factor = Edu_Level].P-
value &lt; 0.05
   AND
   constraints.spec.Reg_Overall_Stats.R Square&gt;0.5</Literal>
       </DataPropertyAssertion>
   <DataPropertyAssertion>
           <DataProperty IRI="#logicalExpression"/>
           <NamedIndividual IRI="#AgeCriterion"/>
           <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-
ns#PlainLiteral">constraints.spec.Reg_Detailed_Stats [Factor = Age].P-value
&lt; 0.05
   AND
   constraints.spec.Reg_Overall_Stats.R Square&gt;0.5</Literal>
       </DataPropertyAssertion>
```

Below we present the formal definition of relations of PA_1 with OccupationCriterion, PA_2 with EducationalLevelCriterion and PA_3 with AgeCriterion of type isEvaluatedBy.

```
<ObjectPropertyAssertion>
        <ObjectProperty IRI="#isEvaluatedBy"/>
        <NamedIndividual IRI="#PA_1"/>
        <NamedIndividual IRI="#OccupationCriterion"/>
    </ObjectPropertyAssertion>
    <ObjectPropertyAssertion>
        <ObjectProperty IRI="#isEvaluatedBy"/>
        <NamedIndividual IRI="#PA_2"/>
        <NamedIndividual IRI="#EducationalLevelCriterion"/>
    </ObjectPropertyAssertion>
    <ObjectPropertyAssertion>
        <ObjectProperty IRI="#isEvaluatedBy"/>
        <NamedIndividual IRI="#PA_3"/>
        <NamedIndividual IRI="#AgeCriterion"/>
    </ObjectPropertyAssertion>
```

WORKFLOW

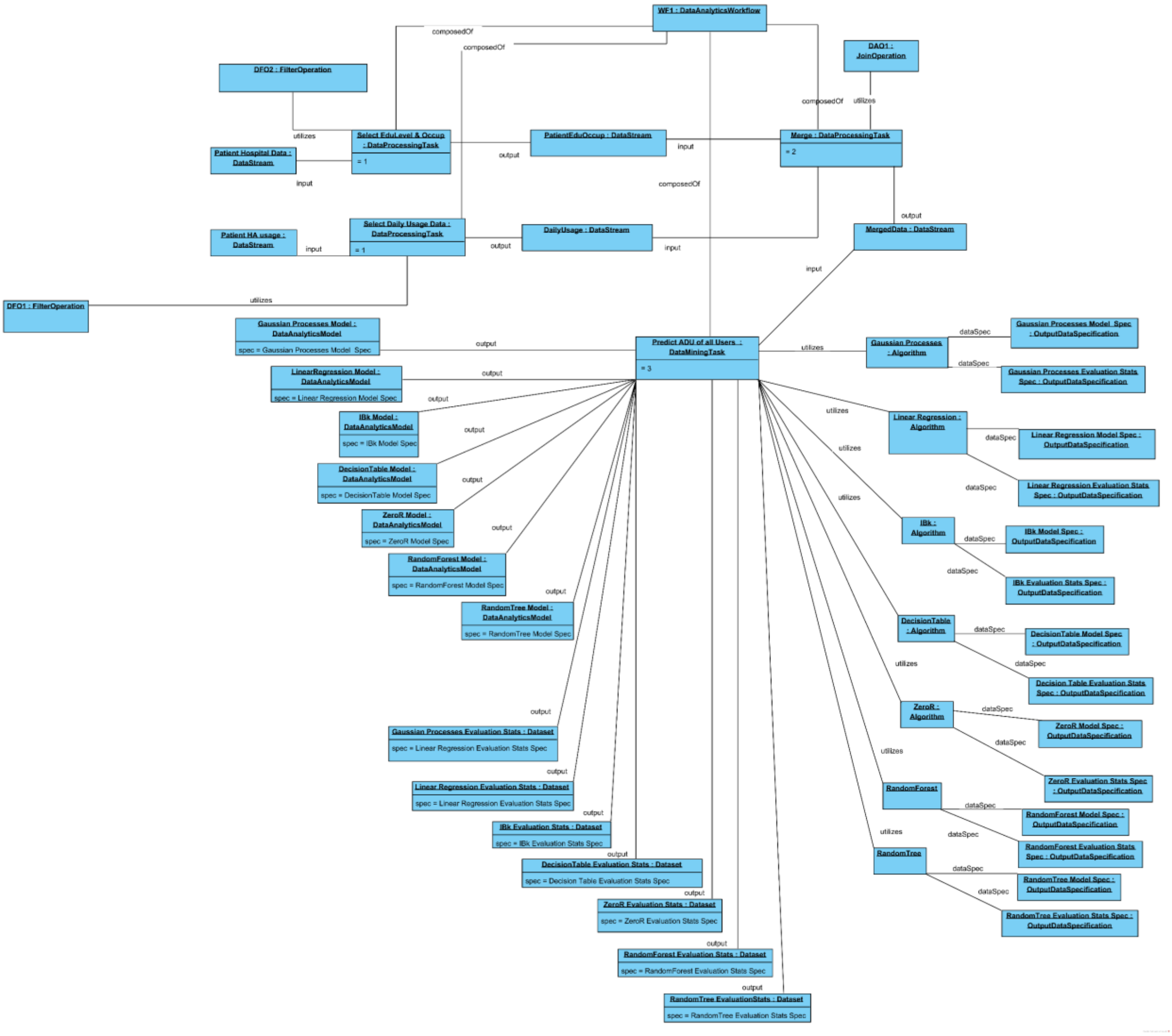


Figure 121 Case Study: Addressing Barriers to HA use: Workflow

Description

The Data Analytics Workflow WF1 is composed of three `DataProcessingTasks`: two data filtering (select) tasks and one merge task and one statistical analysis task. Each of the two select produces as output a corresponding `DataStream`. Data streams are then merged by a different `DataProcessingTask`, utilizing a `JoinOperation`. The Join operation takes as parameters the data used to define policies (EDUCATION, AGE), and DAILY USAGE.

The Data Processing Task produces a Data stream, which is the input to the `StatisticalAnalysisTask`. The algorithm used for the `StatisticalAnalysisTask` is the `StatisticalRegressionAlgorithm`, which has as dataspecs the following `OutputDataSpecifications`: a PredictedValues Spec, a Regression Detailed Stats spec and a Regression overall stats spec. The Task produces as output four datasets with the above output data specifications. These datasets are constrained by the three criteria mentioned above in the scenario.

For the definition of these instances of the classes in OWL we have created:

- One instance of `DataAnalyticsWorfklow` (WF1)
- Three instances of `DataProcessingTask` (SelectEduLevelAndOccup, SelectDailyUsage and Merge)
- One instance of `DataMiningTask` (PredictADUofAllUsers)
- Seven instances of `Algorithm` (GaussianProcessesExecution1, LinearRegressionExecution1, IBkExecution1, DecisionTableExecution1, ZeroRExecution1, RandomForestExecution1, RandomTreeExecution1)
- Two instances of `FilterOperation` (DF01 and DF02)
- One instance of `JoinOperation` (JO01)
- Fourteen instances of `OutputDataSpecification` (GaussianProcesses Model Spec, GaussianProcesses Evaluation Stats Spec, LinearRegression Model Spec, LinearRegression Evaluation Stats Spec, IBk Model Spec, IBk Evaluation Stats Spec, DecisionTable Model Spec, DecisionTable Evaluation Stats Spec, ZeroR Model Spec, ZeroR Evaluation Stats Spec, RandomForest Model Spec, RandomForest Evaluation Stats Spec, RandomTree Model Spec, RandomTree Evaluation Stats Spec)
- Five instances of `DataStream` (PatientHospitalData, PatientHAUsage, PatientEduoccup, DailyUsage, MergedData)

- Fourteen instances of `StaticSet` (GaussianProcesses Model, GaussianProcesses Evaluation Stats, LinearRegression Model, LinearRegression Evaluation Stats, IBk Model, IBk Evaluation Stats, DecisionTable Model, DecisionTable Evaluation Stats, ZeroR Model, ZeroR Evaluation Stats, RandomForest Model, RandomForest Evaluation Stats, RandomTree Model, RandomTree Evaluation Stats)


```
<Declaration>
        <NamedIndividual IRI="#WF1"/>
    </Declaration>
<ClassAssertion>
        <Class IRI="#DataAnalyticsWorfklow"/>
        <NamedIndividual IRI="#WF1"/>
    </ClassAssertion>
<Declaration>
        <NamedIndividual IRI="#SelectEduLevelAndOccup"/>
    </Declaration>
<Declaration>
        <NamedIndividual IRI="#SelectDailyUsage"/>
    </Declaration>
<Declaration>
        <NamedIndividual IRI="#Merge"/>
    </Declaration>
<ClassAssertion>
        <Class IRI="#DataProcessingTask"/>
        <NamedIndividual IRI="#SelectEduLevelAndOccup"/>
    </ClassAssertion>
<ClassAssertion>
        <Class IRI="#DataProcessingTask"/>
        <NamedIndividual IRI="#SelectDailyUsage"/>
    </ClassAssertion>
<ClassAssertion>
        <Class IRI="#DataProcessingTask"/>
        <NamedIndividual IRI="#Merge"/>
    </ClassAssertion>
<Declaration>
        <NamedIndividual IRI="#PredictADUofAllUsers"/>
    </Declaration>
<ClassAssertion>
```

```
        <Class IRI="#DataMiningTask"/>
        <NamedIndividual IRI="#PredictADUofAllUsers"/>
    </ClassAssertion>
<Declaration>
        <NamedIndividual IRI="#GaussianProcessesExecution1"/>
    </Declaration>
<ClassAssertion>
        <Class IRI="#GaussianProcesses"/>
        <NamedIndividual IRI="#GaussianProcesses1"/>
    </ClassAssertion>
<Declaration>
        <NamedIndividual IRI="#LinearRegressionExecution1"/>
    </Declaration>
<ClassAssertion>
        <Class IRI="#LinearRegression"/>
        <NamedIndividual IRI="#LinearRegressionExecution1"/>
    </ClassAssertion>
<Declaration>
        <NamedIndividual IRI="#IBkExecution1"/>
    </Declaration>
<ClassAssertion>
        <Class IRI="#IBk"/>
        <NamedIndividual IRI="#IBkExecution1"/>
    </ClassAssertion>
<Declaration>
        <NamedIndividual IRI="#DecisionTableExecution1"/>
    </Declaration>
<ClassAssertion>
        <Class IRI="#DecisionTable"/>
        <NamedIndividual IRI="#DecisionTableExecution1"/>
    </ClassAssertion>
<Declaration>
        <NamedIndividual IRI="#ZeroRExecution1"/>
    </Declaration>
<ClassAssertion>
        <Class IRI="#ZeroR"/>
        <NamedIndividual IRI="#ZeroRExecution1"/>
    </ClassAssertion>
<Declaration>
```

```
        <NamedIndividual IRI="#RandomForestExecution1"/>
    </Declaration>
<ClassAssertion>
        <Class IRI="#RandomForest"/>
        <NamedIndividual IRI="#RandomForestExecution1"/>
    </ClassAssertion>
<Declaration>
        <NamedIndividual IRI="#RandomTreeExecution1"/>
    </Declaration>
<ClassAssertion>
        <Class IRI="#RandomTree"/>
        <NamedIndividual IRI="#RandomTreeExecution1"/>
    </ClassAssertion>
<Declaration>
        <NamedIndividual IRI="#DF01"/>
    </Declaration>
<ClassAssertion>
        <Class IRI="#FilterOperation"/>
        <NamedIndividual IRI="#DF01"/>
    </ClassAssertion>
<Declaration>
        <NamedIndividual IRI="#DF02"/>
    </Declaration>
<ClassAssertion>
        <Class IRI="#FilterOperation"/>
        <NamedIndividual IRI="#DF02"/>
    </ClassAssertion>
<Declaration>
        <NamedIndividual IRI="#JO01"/>
    </Declaration>
<ClassAssertion>
        <Class IRI="#JoinOperation"/>
        <NamedIndividual IRI="#JO01"/>
    </ClassAssertion>
<Declaration>
        <NamedIndividual IRI="#GaussianProcessesModelSpec"/>
    </Declaration>
<ClassAssertion>
        <Class IRI="#OutputDataSpecification"/>
```

```
        <NamedIndividual IRI="#GaussianProcessesModelSpec"/>
    </ClassAssertion>
<Declaration>
        <NamedIndividual IRI="#GaussianProcessesEvaluationStatsSpec"/>
    </Declaration>
<ClassAssertion>
        <Class IRI="#OutputDataSpecification"/>
        <NamedIndividual IRI="#GaussianProcessesEvaluationStatsSpec"/>
    </ClassAssertion>
<Declaration>
        <NamedIndividual IRI="#LinearRegressionModelSpec"/>
    </Declaration>
<ClassAssertion>
        <Class IRI="#OutputDataSpecification"/>
        <NamedIndividual IRI="#LinearRegressionModelSpec"/>
    </ClassAssertion>
<Declaration>
        <NamedIndividual IRI="#LinearRegressionEvaluationStatsSpec"/>
    </Declaration>
<ClassAssertion>
        <Class IRI="#OutputDataSpecification"/>
        <NamedIndividual IRI="#LinearRegressionEvaluationStatsSpec"/>
    </ClassAssertion>
<Declaration>
        <NamedIndividual IRI="#IBkModelSpec"/>
    </Declaration>
<ClassAssertion>
        <Class IRI="#OutputDataSpecification"/>
        <NamedIndividual IRI="#IBkModelSpec"/>
    </ClassAssertion>
<Declaration>
        <NamedIndividual IRI="#IBkEvaluationStatsSpec"/>
    </Declaration>
<ClassAssertion>
        <Class IRI="#OutputDataSpecification"/>
        <NamedIndividual IRI="#IBkEvaluationStatsSpec"/>
    </ClassAssertion>
<Declaration>
        <NamedIndividual IRI="#DecisionTableModelSpec"/>
```

```
    </Declaration>
<ClassAssertion>
        <Class IRI="#OutputDataSpecification"/>
        <NamedIndividual IRI="#DecisionTableModelSpec"/>
    </ClassAssertion>
<Declaration>
        <NamedIndividual IRI="#DecisionTableEvaluationStatsSpec"/>
    </Declaration>
<ClassAssertion>
        <Class IRI="#OutputDataSpecification"/>
        <NamedIndividual IRI="#DecisionTableEvaluationStatsSpec"/>
    </ClassAssertion>
<Declaration>
        <NamedIndividual IRI="#ZeroRModelSpec"/>
    </Declaration>
<ClassAssertion>
        <Class IRI="#OutputDataSpecification"/>
        <NamedIndividual IRI="#ZeroRModelSpec"/>
    </ClassAssertion>
<Declaration>
        <NamedIndividual IRI="#ZeroREvaluationStatsSpec"/>
    </Declaration>
<ClassAssertion>
        <Class IRI="#OutputDataSpecification"/>
        <NamedIndividual IRI="#ZeroREvaluationStatsSpec"/>
    </ClassAssertion>
<Declaration>
        <NamedIndividual IRI="#RandomForestModelSpec"/>
    </Declaration>
<ClassAssertion>
        <Class IRI="#OutputDataSpecification"/>
        <NamedIndividual IRI="#RandomForestModelSpec"/>
    </ClassAssertion>
<Declaration>
        <NamedIndividual IRI="#RandomForestEvaluationStatsSpec"/>
    </Declaration>
<ClassAssertion>
        <Class IRI="#OutputDataSpecification"/>
        <NamedIndividual IRI="#RandomForestEvaluationStatsSpec"/>
```

```
    </ClassAssertion>
<Declaration>
        <NamedIndividual IRI="#RandomTreeModelSpec"/>
    </Declaration>
<ClassAssertion>
        <Class IRI="#OutputDataSpecification"/>
        <NamedIndividual IRI="#RandomTreeModelSpec"/>
    </ClassAssertion>
<Declaration>
        <NamedIndividual IRI="#RandomTreeEvaluationStatsSpec"/>
    </Declaration>
<ClassAssertion>
        <Class IRI="#OutputDataSpecification"/>
        <NamedIndividual IRI="#RandomTreeEvaluationStatsSpec"/>
    </ClassAssertion>
<Declaration>
        <NamedIndividual IRI="#PatientHospitalData"/>
    </Declaration>
<ClassAssertion>
        <Class IRI="#DataStream"/>
        <NamedIndividual IRI="#PatientHospitalData"/>
    </ClassAssertion>
<Declaration>
        <NamedIndividual IRI="#PatientHAUsage"/>
    </Declaration>
<ClassAssertion>
        <Class IRI="#DataStream"/>
        <NamedIndividual IRI="#PatientHAUsage"/>
    </ClassAssertion>
<Declaration>
        <NamedIndividual IRI="#PatientEduoccup"/>
    </Declaration>
<ClassAssertion>
        <Class IRI="#DataStream"/>
        <NamedIndividual IRI="#PatientEduoccup"/>
    </ClassAssertion>
<Declaration>
        <NamedIndividual IRI="#DailyUsage"/>
    </Declaration>
```

```
<ClassAssertion>
        <Class IRI="#DataStream"/>
        <NamedIndividual IRI="#DailyUsage"/>
    </ClassAssertion>
<Declaration>
        <NamedIndividual IRI="#MergedData"/>
    </Declaration>
<ClassAssertion>
        <Class IRI="#DataStream"/>
        <NamedIndividual IRI="#MergedData"/>
    </ClassAssertion>
<Declaration>
        <NamedIndividual IRI="#GaussianProcessesModel"/>
    </Declaration>
<ClassAssertion>
        <Class IRI="#StaticSet"/>
        <NamedIndividual IRI="#GaussianProcessesModel"/>
    </ClassAssertion>
<Declaration>
        <NamedIndividual IRI="#GaussianProcessesEvaluationStats"/>
    </Declaration>
<ClassAssertion>
        <Class IRI="#StaticSet"/>
        <NamedIndividual IRI="#GaussianProcessesEvaluationStats"/>
    </ClassAssertion>
<Declaration>
        <NamedIndividual IRI="#LinearRegressionModel"/>
    </Declaration>
<ClassAssertion>
        <Class IRI="#StaticSet"/>
        <NamedIndividual IRI="#LinearRegressionModel"/>
    </ClassAssertion>
<Declaration>
        <NamedIndividual IRI="#LinearRegressionEvaluationStats"/>
    </Declaration>
<ClassAssertion>
        <Class IRI="#StaticSet"/>
        <NamedIndividual IRI="#LinearRegressionEvaluationStats"/>
    </ClassAssertion>
```

```
<Declaration>
        <NamedIndividual IRI="#IBkModel"/>
    </Declaration>
<ClassAssertion>
        <Class IRI="#StaticSet"/>
        <NamedIndividual IRI="#IBkModel"/>
    </ClassAssertion>
<Declaration>
        <NamedIndividual IRI="#IBkEvaluationStats"/>
    </Declaration>
<ClassAssertion>
        <Class IRI="#StaticSet"/>
        <NamedIndividual IRI="#IBkEvaluationStats"/>
    </ClassAssertion>
<Declaration>
        <NamedIndividual IRI="#DecisionTableModel"/>
    </Declaration>
<ClassAssertion>
        <Class IRI="#StaticSet"/>
        <NamedIndividual IRI="#DecisionTableModel"/>
    </ClassAssertion>
<Declaration>
        <NamedIndividual IRI="#DecisionTableEvaluationStats"/>
    </Declaration>
<ClassAssertion>
        <Class IRI="#StaticSet"/>
        <NamedIndividual IRI="#DecisionTableEvaluationStats"/>
    </ClassAssertion>
<Declaration>
        <NamedIndividual IRI="#ZeroRModel"/>
    </Declaration>
<ClassAssertion>
        <Class IRI="#StaticSet"/>
        <NamedIndividual IRI="#ZeroRModel"/>
    </ClassAssertion>
<Declaration>
        <NamedIndividual IRI="#ZeroREvaluationStats"/>
    </Declaration>
<ClassAssertion>
```

```
        <Class IRI="#StaticSet"/>
        <NamedIndividual IRI="#ZeroREvaluationStats"/>
    </ClassAssertion>
<Declaration>
        <NamedIndividual IRI="#RandomForestModel"/>
    </Declaration>
<ClassAssertion>
        <Class IRI="#StaticSet"/>
        <NamedIndividual IRI="#RandomForestModel"/>
    </ClassAssertion>
<Declaration>
        <NamedIndividual IRI="#RandomForestEvaluationStats"/>
    </Declaration>
<ClassAssertion>
        <Class IRI="#StaticSet"/>
        <NamedIndividual IRI="#RandomForestEvaluationStats"/>
    </ClassAssertion>
<Declaration>
        <NamedIndividual IRI="#RandomTreeModel"/>
    </Declaration>
<ClassAssertion>
        <Class IRI="#StaticSet"/>
        <NamedIndividual IRI="#RandomTreeModel"/>
    </ClassAssertion>
<Declaration>
        <NamedIndividual IRI="#RandomTreeEvaluationStats"/>
    </Declaration>
<ClassAssertion>
        <Class IRI="#StaticSet"/>
        <NamedIndividual IRI="#RandomTreeEvaluationStats"/>
    </ClassAssertion>
```

Below we present the formal definition of the following:

- the relations of WF1: that WF1 has the relation `specifies` with OccupationCriterion EducationalLevelCriterion and AgeCriterion and that WF1 `isComposedOf` the tasks SelectEduLevelAndOccup, SelectDailyUsage, Merge and PredictADUofAllUsers.
- The relations of type `utilizes` of the task PredictADUofAllUsers with the algorithms GaussianProcessesExecution1, LinearRegressionExecution1, IBkExecution1, DecisionTableExecution1, ZeroRExecution1, RandomForestExecution1,

RandomTreeExecution1, of the task SelectEduLevelAndOccup with the operation DF01, of the task SelectDailyUsage with the operation DF02 and of the task Merge with the operation JO01.

- The relations of type `dataSpec` of the algorithm execution GaussianProcessesExecution1 with the `OutputDataSpecifications` GaussianProcessesModelSpec and GaussianProcessesEvaluationStatsSpec, the algorithm execution LinearRegressionExecution1 with the `OutputDataSpecifications` LinearRegressionModelSpec and LinearRegressionEvaluationStatsSpec, the algorithm execution IBkExecution1 with the `OutputDataSpecifications` IBkModelSpec, IBkEvaluationStatsSpec, the algorithm execution DecisionTableExecution1 with the `OutputDataSpecifications` DecisionTableModelSpec and DecisionTableEvaluationStatsSpec, the algorithm execution ZeroRExecution1 with the `OutputDataSpecifications` ZeroRModelSpec and ZeroREvaluationStatsSpec, the algorithm execution RandomForestExecution1 with the `OutputDataSpecifications` RandomForestModelSpec and RandomForestEvaluationStatsSpec and the algorithm execution RandomTreeExecution1 with the `OutputDataSpecifications` RandomTreeModelSpec, RandomTreeEvaluationStatsSpec.
- The relations of the following data streams: that PatientHospitalData is `input` to SelectEduLevelAndOccup task and PatientEduOccup is `output` from this task, that PatientHAUSage is `input` to SelectDailyUsageData task and DailyUsage is `output` from this task, that the previously referred data streams are `input` to the Merge task and MergedData is `output` of the Merge task and `input` to PredictADUofAllUsers task.
- The relations of the following static sets: that GaussianProcessesModel `haSpec` GaussianProcessesModelSpec, that GaussianProcessesEvaluationStats `haSpec` GaussianProcessesEvaluationStatsSpec, that LinearRegressionModel `haSpec` LinearRegressionModelSpec, that LinearRegressionEvaluationStats `haSpec` LinearRegressionEvaluationStatsSpec, that IBkModel `haSpec` IBkModelSpec, that IBkEvaluationStats `haSpec` IBkEvaluationStatsSpec, that DecisionTableModel `haSpec` DecisionTableModelSpec, that DecisionTableEvaluationStats `haSpec` DecisionTableEvaluationStatsSpec, that ZeroRModel `haSpec` ZeroRModelSpec, that ZeroREvaluationStats `haSpec` ZeroREvaluationStatsSpec, that RandomForestModel `haSpec` RandomForestModelSpec, that RandomForestEvaluationStats `haSpec` RandomForestEvaluationStatsSpec, that RandomTreeModel `haSpec` RandomTreeModelSpec, that RandomTreeEvaluationStats `haSpec` RandomTreeEvaluationStats Spec and that all the above are `outputs` of PredictADUofAllUsers task.

```
<ObjectProperty IRI="#specifies"/>
```

```
        <NamedIndividual IRI="#OccupationCriterion"/>
        <NamedIndividual IRI="#WF1"/>
    </ObjectPropertyAssertion>
<ObjectPropertyAssertion>
        <ObjectProperty IRI="#specifies"/>
        <NamedIndividual IRI="#AgeCriterion"/>
        <NamedIndividual IRI="#WF1"/>
    </ObjectPropertyAssertion>
<ObjectPropertyAssertion>
        <ObjectProperty IRI="#specifies"/>
        <NamedIndividual IRI="#EducationalLevelCriterion"/>
        <NamedIndividual IRI="#WF1"/>
    </ObjectPropertyAssertion>
<ObjectPropertyAssertion>
        <ObjectProperty IRI="#isComposedOf"/>
        <NamedIndividual IRI="#WF1"/>
        <NamedIndividual IRI="#SelectEduLevelAndOccup"/>
    </ObjectPropertyAssertion>
<ObjectPropertyAssertion>
        <ObjectProperty IRI="#isComposedOf"/>
        <NamedIndividual IRI="#WF1"/>
        <NamedIndividual IRI="#SelectDailyUsage"/>
    </ObjectPropertyAssertion>
<ObjectPropertyAssertion>
        <ObjectProperty IRI="#isComposedOf"/>
        <NamedIndividual IRI="#WF1"/>
        <NamedIndividual IRI="#Merge"/>
    </ObjectPropertyAssertion>
<ObjectPropertyAssertion>
        <ObjectProperty IRI="#isComposedOf"/>
        <NamedIndividual IRI="#WF1"/>
        <NamedIndividual IRI="#PredictADUofAllUsers"/>
    </ObjectPropertyAssertion>
<ObjectPropertyAssertion>
        <ObjectProperty IRI="#utilizes"/>
        <NamedIndividual IRI="#PredictADUofAllUsers"/>
        <NamedIndividual IRI="#GaussianProcessesExecution1"/>
    </ObjectPropertyAssertion>
<ObjectPropertyAssertion>
```

```
        <ObjectProperty IRI="#utilizes"/>
        <NamedIndividual IRI="#PredictADUofAllUsers"/>
        <NamedIndividual IRI="#LinearRegressionExecution1"/>
    </ObjectPropertyAssertion>
<ObjectPropertyAssertion>
        <ObjectProperty IRI="#utilizes"/>
        <NamedIndividual IRI="#PredictADUofAllUsers"/>
        <NamedIndividual IRI="#IBkExecution1"/>
    </ObjectPropertyAssertion>
<ObjectPropertyAssertion>
        <ObjectProperty IRI="#utilizes"/>
        <NamedIndividual IRI="#PredictADUofAllUsers"/>
        <NamedIndividual IRI="#DecisionTableExecution1"/>
    </ObjectPropertyAssertion>
<ObjectPropertyAssertion>
        <ObjectProperty IRI="#utilizes"/>
        <NamedIndividual IRI="#PredictADUofAllUsers"/>
        <NamedIndividual IRI="#ZeroRExecution1"/>
    </ObjectPropertyAssertion>
<ObjectPropertyAssertion>
        <ObjectProperty IRI="#utilizes"/>
        <NamedIndividual IRI="#PredictADUofAllUsers"/>
        <NamedIndividual IRI="#RandomForestExecution1"/>
    </ObjectPropertyAssertion>
<ObjectPropertyAssertion>
        <ObjectProperty IRI="#utilizes"/>
        <NamedIndividual IRI="#PredictADUofAllUsers"/>
        <NamedIndividual IRI="#RandomTreeExecution1"/>
    </ObjectPropertyAssertion>
<ObjectPropertyAssertion>
        <ObjectProperty IRI="#utilizes"/>
        <NamedIndividual IRI="#SelectEduLevelAndOccup"/>
        <NamedIndividual IRI="#DF01"/>
    </ObjectPropertyAssertion>
<ObjectPropertyAssertion>
        <ObjectProperty IRI="#utilizes"/>
        <NamedIndividual IRI="#SelectDailyUsage"/>
        <NamedIndividual IRI="#DF02"/>
    </ObjectPropertyAssertion>
```

```
<ObjectPropertyAssertion>
        <ObjectProperty IRI="#utilizes"/>
        <NamedIndividual IRI="#Merge"/>
        <NamedIndividual IRI="#JO01"/>
    </ObjectPropertyAssertion>
<ObjectPropertyAssertion>
        <ObjectProperty IRI="#dataSpec"/>
        <NamedIndividual IRI="#GaussianProcessesExecution1"/>
        <NamedIndividual IRI="#GaussianProcessesModelSpec"/>
    </ObjectPropertyAssertion>
<ObjectPropertyAssertion>
        <ObjectProperty IRI="#dataSpec"/>
        <NamedIndividual IRI="#GaussianProcessesExecution1"/>
        <NamedIndividual IRI="#GaussianProcessesEvaluationStatsSpec"/>
    </ObjectPropertyAssertion>
<ObjectPropertyAssertion>
        <ObjectProperty IRI="#dataSpec"/>
        <NamedIndividual IRI="#LinearRegressionExecution1"/>
        <NamedIndividual IRI="#LinearRegressionModelSpec"/>
    </ObjectPropertyAssertion>
<ObjectPropertyAssertion>
        <ObjectProperty IRI="#dataSpec"/>
        <NamedIndividual IRI="#LinearRegressionExecution1"/>
        <NamedIndividual IRI="#LinearRegressionEvaluationStatsSpec"/>
    </ObjectPropertyAssertion>
<ObjectPropertyAssertion>
        <ObjectProperty IRI="#dataSpec"/>
        <NamedIndividual IRI="#IBkExecution1"/>
        <NamedIndividual IRI="#IBkModelSpec"/>
    </ObjectPropertyAssertion>
<ObjectPropertyAssertion>
        <ObjectProperty IRI="#dataSpec"/>
        <NamedIndividual IRI="#IBkExecution1"/>
        <NamedIndividual IRI="#IBkEvaluationStatsSpec"/>
    </ObjectPropertyAssertion>
<ObjectPropertyAssertion>
        <ObjectProperty IRI="#dataSpec"/>
        <NamedIndividual IRI="#DecisionTableExecution1"/>
        <NamedIndividual IRI="#DecisionTableModelSpec"/>
```

```
    </ObjectPropertyAssertion>
  <ObjectPropertyAssertion>
        <ObjectProperty IRI="#dataSpec"/>
        <NamedIndividual IRI="#DecisionTableExecution1"/>
        <NamedIndividual IRI="#DecisionTableEvaluationStatsSpec"/>
    </ObjectPropertyAssertion>
  <ObjectPropertyAssertion>
        <ObjectProperty IRI="#dataSpec"/>
        <NamedIndividual IRI="#ZeroRExecution1"/>
        <NamedIndividual IRI="#ZeroRModelSpec"/>
    </ObjectPropertyAssertion>
  <ObjectPropertyAssertion>
        <ObjectProperty IRI="#dataSpec"/>
        <NamedIndividual IRI="#ZeroRExecution1"/>
        <NamedIndividual IRI="#ZeroREvaluationStatsSpec"/>
    </ObjectPropertyAssertion>
  <ObjectPropertyAssertion>
        <ObjectProperty IRI="#dataSpec"/>
        <NamedIndividual IRI="#RandomForestExecution1"/>
        <NamedIndividual IRI="#RandomForestModelSpec"/>
    </ObjectPropertyAssertion>
  <ObjectPropertyAssertion>
        <ObjectProperty IRI="#dataSpec"/>
        <NamedIndividual IRI="#RandomForestExecution1"/>
        <NamedIndividual IRI="#RandomForestEvaluationStatsSpec"/>
    </ObjectPropertyAssertion>
  <ObjectPropertyAssertion>
        <ObjectProperty IRI="#dataSpec"/>
        <NamedIndividual IRI="#RandomTreeExecution1"/>
        <NamedIndividual IRI="#RandomTreeModelSpec"/>
    </ObjectPropertyAssertion>
  <ObjectPropertyAssertion>
        <ObjectProperty IRI="#dataSpec"/>
        <NamedIndividual IRI="#RandomTreeExecution1"/>
        <NamedIndividual IRI="#RandomTreeEvaluationStatsSpec"/>
    </ObjectPropertyAssertion>

  <ObjectPropertyAssertion>
        <ObjectProperty IRI="#input"/>
```

```
        <NamedIndividual IRI="#SelectEduLevelAndOccup"/>
        <NamedIndividual IRI="#PatientHospitalData"/>
    </ObjectPropertyAssertion>
<ObjectPropertyAssertion>
        <ObjectProperty IRI="#input"/>
        <NamedIndividual IRI="#SelectDailyUsageData"/>
        <NamedIndividual IRI="#PatientHAUsage"/>
    </ObjectPropertyAssertion>
<ObjectPropertyAssertion>
        <ObjectProperty IRI="#output"/>
        <NamedIndividual IRI="#SelectEduLevelAndOccup"/>
        <NamedIndividual IRI="#PatientEduOccup"/>
    </ObjectPropertyAssertion>
<ObjectPropertyAssertion>
        <ObjectProperty IRI="#output"/>
        <NamedIndividual IRI="#SelectDailyUsageData"/>
        <NamedIndividual IRI="#DailyUsage"/>
    </ObjectPropertyAssertion>
<ObjectPropertyAssertion>
        <ObjectProperty IRI="#input"/>
        <NamedIndividual IRI="#Merge"/>
        <NamedIndividual IRI="#PatientEduOccup"/>
    </ObjectPropertyAssertion>
<ObjectPropertyAssertion>
        <ObjectProperty IRI="#input"/>
        <NamedIndividual IRI="#Merge"/>
        <NamedIndividual IRI="#DailyUsage"/>
    </ObjectPropertyAssertion>
<ObjectPropertyAssertion>
        <ObjectProperty IRI="#output"/>
        <NamedIndividual IRI="#Merge"/>
        <NamedIndividual IRI="#MergedData"/>
    </ObjectPropertyAssertion>
<ObjectPropertyAssertion>
        <ObjectProperty IRI="#input"/>
        <NamedIndividual IRI="#PredictADUofAllUsers"/>
        <NamedIndividual IRI="#MergedData"/>
    </ObjectPropertyAssertion>
<ObjectPropertyAssertion>
```

```
        <ObjectProperty IRI="#output"/>
        <NamedIndividual IRI="#PredictADUofAllUsers"/>
        <NamedIndividual IRI="#GaussianProcessesModel"/>
    </ObjectPropertyAssertion>
<ObjectPropertyAssertion>
        <ObjectProperty IRI="#spec"/>
        <NamedIndividual IRI="#GaussianProcessesModel"/>
        <NamedIndividual IRI="#GaussianProcessesModelSpec"/>
    </ObjectPropertyAssertion>
<ObjectPropertyAssertion>
        <ObjectProperty IRI="#output"/>
        <NamedIndividual IRI="#PredictADUofAllUsers"/>
        <NamedIndividual IRI="#GaussianProcessesEvaluationStats"/>
    </ObjectPropertyAssertion>
<ObjectPropertyAssertion>
        <ObjectProperty IRI="#spec"/>
        <NamedIndividual IRI="#GaussianProcessesEvaluationStats"/>
        <NamedIndividual IRI="#GaussianProcessesEvaluationStatsSpec"/>
    </ObjectPropertyAssertion>
<ObjectPropertyAssertion>
        <ObjectProperty IRI="#output"/>
        <NamedIndividual IRI="#PredictADUofAllUsers"/>
        <NamedIndividual IRI="#LinearRegressionModel"/>
    </ObjectPropertyAssertion>
<ObjectPropertyAssertion>
        <ObjectProperty IRI="#spec"/>
        <NamedIndividual IRI="#LinearRegressionModel"/>
        <NamedIndividual IRI="#LinearRegressionModelSpec"/>
    </ObjectPropertyAssertion>
<ObjectPropertyAssertion>
        <ObjectProperty IRI="#output"/>
        <NamedIndividual IRI="#PredictADUofAllUsers"/>
        <NamedIndividual IRI="#LinearRegressionEvaluationStats"/>
    </ObjectPropertyAssertion>
<ObjectPropertyAssertion>
        <ObjectProperty IRI="#spec"/>
        <NamedIndividual IRI="#LinearRegressionEvaluationStats"/>
        <NamedIndividual IRI="#LinearRegressionEvaluationStatsSpec"/>
    </ObjectPropertyAssertion>
```

```
  <ObjectPropertyAssertion>
        <ObjectProperty IRI="#output"/>
        <NamedIndividual IRI="#PredictADUofAllUsers"/>
        <NamedIndividual IRI="#IBkModel"/>
    </ObjectPropertyAssertion>
  <ObjectPropertyAssertion>
        <ObjectProperty IRI="#spec"/>
        <NamedIndividual IRI="#IBkModel"/>
        <NamedIndividual IRI="#IBkModelSpec"/>
    </ObjectPropertyAssertion>
  <ObjectPropertyAssertion>
        <ObjectProperty IRI="#output"/>
        <NamedIndividual IRI="#PredictADUofAllUsers"/>
        <NamedIndividual IRI="#IBkEvaluationStats"/>
    </ObjectPropertyAssertion>
  <ObjectPropertyAssertion>
        <ObjectProperty IRI="#spec"/>
        <NamedIndividual IRI="#IBkEvaluationStats"/>
        <NamedIndividual IRI="#IBkEvaluationStatsSpec"/>
    </ObjectPropertyAssertion>
  <ObjectPropertyAssertion>
        <ObjectProperty IRI="#output"/>
        <NamedIndividual IRI="#PredictADUofAllUsers"/>
        <NamedIndividual IRI="#DecisionTableModel"/>
    </ObjectPropertyAssertion>
  <ObjectPropertyAssertion>
        <ObjectProperty IRI="#spec"/>
        <NamedIndividual IRI="#DecisionTableModel"/>
        <NamedIndividual IRI="#DecisionTableModelSpec"/>
    </ObjectPropertyAssertion>
  <ObjectPropertyAssertion>
        <ObjectProperty IRI="#output"/>
        <NamedIndividual IRI="#PredictADUofAllUsers"/>
        <NamedIndividual IRI="#DecisionTableEvaluationStats"/>
    </ObjectPropertyAssertion>
  <ObjectPropertyAssertion>
        <ObjectProperty IRI="#spec"/>
        <NamedIndividual IRI="#DecisionTableEvaluationStats"/>
        <NamedIndividual IRI="#DecisionTableEvaluationStatsSpec"/>
```

```
    </ObjectPropertyAssertion>
<ObjectPropertyAssertion>
        <ObjectProperty IRI="#output"/>
        <NamedIndividual IRI="#PredictADUofAllUsers"/>
        <NamedIndividual IRI="#ZeroRModel"/>
    </ObjectPropertyAssertion>
<ObjectPropertyAssertion>
        <ObjectProperty IRI="#spec"/>
        <NamedIndividual IRI="#ZeroRModel"/>
        <NamedIndividual IRI="#ZeroRModelSpec"/>
    </ObjectPropertyAssertion>
<ObjectPropertyAssertion>
        <ObjectProperty IRI="#output"/>
        <NamedIndividual IRI="#PredictADUofAllUsers"/>
        <NamedIndividual IRI="#ZeroREvaluationStats"/>
    </ObjectPropertyAssertion>
<ObjectPropertyAssertion>
        <ObjectProperty IRI="#spec"/>
        <NamedIndividual IRI="#ZeroREvaluationStats"/>
        <NamedIndividual IRI="#ZeroREvaluationStatsSpec"/>
    </ObjectPropertyAssertion>
<ObjectPropertyAssertion>
        <ObjectProperty IRI="#output"/>
        <NamedIndividual IRI="#PredictADUofAllUsers"/>
        <NamedIndividual IRI="#RandomForestModel"/>
    </ObjectPropertyAssertion>
<ObjectPropertyAssertion>
        <ObjectProperty IRI="#spec"/>
        <NamedIndividual IRI="#RandomForestModel"/>
        <NamedIndividual IRI="#RandomForestModelSpec"/>
    </ObjectPropertyAssertion>
<ObjectPropertyAssertion>
        <ObjectProperty IRI="#output"/>
        <NamedIndividual IRI="#PredictADUofAllUsers"/>
        <NamedIndividual IRI="#RandomForestEvaluationStats"/>
    </ObjectPropertyAssertion>
<ObjectPropertyAssertion>
        <ObjectProperty IRI="#spec"/>
        <NamedIndividual IRI="#RandomForestEvaluationStats"/>
```

```
        <NamedIndividual IRI="#RandomForestEvaluationStatsSpec"/>
    </ObjectPropertyAssertion>
<ObjectPropertyAssertion>
        <ObjectProperty IRI="#output"/>
        <NamedIndividual IRI="#PredictADUofAllUsers"/>
        <NamedIndividual IRI="#RandomTreeModel"/>
    </ObjectPropertyAssertion>
<ObjectPropertyAssertion>
        <ObjectProperty IRI="#spec"/>
        <NamedIndividual IRI="#RandomTreeModel"/>
        <NamedIndividual IRI="#RandomTreeModelSpec"/>
    </ObjectPropertyAssertion>
<ObjectPropertyAssertion>
        <ObjectProperty IRI="#output"/>
        <NamedIndividual IRI="#PredictADUofAllUsers"/>
        <NamedIndividual IRI="#RandomTreeEvaluationStats"/>
    </ObjectPropertyAssertion>
<ObjectPropertyAssertion>
        <ObjectProperty IRI="#spec"/>
        <NamedIndividual IRI="#RandomTreeEvaluationStats"/>
        <NamedIndividual IRI="#RandomTreeEvaluationStatsSpec"/>
    </ObjectPropertyAssertion>
```