The SMART BEAR Project: an overview of its infrastructure

Qiqi Su[1], Vadim Peretokin [2], Ioannis Basdekis[3], Ioannis Kouris[4], Jonatan Maggesi[5], Mario Sicuranza[6], Alberto Acebes[7], Anca Bucur[2], Vinod Jaswanth Roy Mukkala[5], Konstantin Pozdniakov[1], Christos Kloukinas[1], Dimitrios D. Koutsouris[4], Elefteria Iliadou[8], Ioannis Leontsinis[9], Luigi Gallo[6], Giuseppe De Pietro[6] and George Spanoudakis[3]

[1] Department of Computer Science, City, University of London, London, U.K.
[2] Philips Research, Eindhoven, The Netherlands.
[3]SPHYNX Technology Solutions AG, Zug, CH, Switzerland.
[4]Biomedical Engineering Laboratory, School of Electrical and Computer Engineering, National Technical University of Athens, Athens, Greece.
[5]Computer Science Department, Università degli Studi di Milano, Milan, Italy.
[6]Institute for High-Performance Computing and Networking, National Research Council of ITALY, ICAR - CNR, Naples, Italy.
[7]Atos Research and Innovation. Madrid, Spain.
[8]1st Otolaryngology University Department, National and Kapodistrian University of Athens, Athens, Greece
[9]1st Cardiology Clinic, Medical school, National and Kapodistrian University of Athens, Athens Greece

{Qiqi.Su, Konstantin.Pozdniakov, C.Kloukinas}@city.ac.uk, {vadim.peretokin, anca.bucur}@philips.com, maggesi@di.unimi.it, vinod.mukkala@unimi.it, {mario.sicuranza, luigi.gallo, giuseppe.depietro}@icar.cnr.it

Abstract. The paper describes a cloud-based platform that utilizes Artificial Intelligence (AI) and Explainable AI techniques to deliver evidence-based, personalized interventions to individuals over 65 suffering or at risk of hearing loss, cardiovascular disease, cognitive impairments, balance disorders, or mental health issues, while supporting efficient remote monitoring and clinician-driven guidance. As part of the SMART BEAR integrated project, this platform has been developed to support its large-scale clinical trials. The platform consists of a standards-based data harmonization and management layer, as well as a security component, a Big Data Analytics system, a Clinical Decision Support system, and a dashboard component to facilitate efficient data collection across pilot sites.

Keywords: Cloud, AI, semantic interoperability, HL7 FHIR, Healthcare, GDPR, Evidence-based, Ageing, Hearing Loss, Cardiovascular Disease, Balance disorder

1 Introduction

The EU-funded SMART BEAR project[1] develops an integrated platform to provide evidence-based personalized support for several pressing healthcare issues faced by the aging EU societies, including hearing loss, cardiovascular disease, cognitive disease, and balance disorders. Providing support for these health-related issues to the ageing population, in order to promote healthy and independent living, is particularly important in the EU societies since ageing can have a significant social and financial impact due to a higher incidence of these issues.

In SMART BEAR, continuously collected data from a variety of sensors, assistive medical and mobile devices will be harmonized and analyzed in order to provide effective recommendations and personalized interventions. The developed SMART BEAR platform will be tested with five thousand elderly participants from six EU countries: France, Greece, Italy, Romania, Portugal, and Spain. The large-scaled project is scheduled to commence in autumn 2022 and run for 24 months. Prior to this, a smaller-scaled pilot study, named Pilot-of-Pilots (PoP), with 100 participants is already underway in the island of Madeira, Portugal since June 2022.

There has been an increased interest in e-health monitoring systems situated at homes in recent years, leading to the creation of Health Smart Homes. Such technologies can facilitate monitoring patients' activities, in order to improve the quality of care for the elderly and increase their well-being in a non-obtrusive way. Health Smart Homes can also enable efficient and decentralized healthcare services at home, which allows for greater independence and empowerment, preventing social isolation for the individuals, and maintaining good health longer. Furthermore, elderly individuals can avoid being placed in institutions such as nursing homes and hospitals for as long as possible, thus reducing the burden on the healthcare system (Mshali et al., 2018).

Health Smart Homes are powered by the Internet of Things (IoT), and more specifically, Medical IoT, which refers to the increasing range of applications of IoT in the medical domain (Akyildiz et al., 2015). Major advancements in wireless technology and computing power have enabled the current wide use of IoT and has led to the proliferation of specialized and diverse Medical IoT devices that can generate and transmit data through an open protocol, which can then be analyzed subsequently. Among the benefits of Medical IoT are ease of service delivery, early diagnosis, improved patient management, and reduced manual errors (Adhikary et al., 2020).

The growth of Medical IoT monitoring devices for medical and well-being measurements is not the only factor that is changing the landscape in consumer health and personalized medicine. Through a connected infrastructure of medical devices, software applications, health systems and services, as well as the data generated at an accelerated rate, are transforming the delivery of healthcare. Today, e-health systems equipped with Big Data Analytics (BDA) capabilities enable the provision of high-quality decision support, thus improving the quality of care. Information exchange and data reusability, combined with the application of data mining and machine learning (ML) analytics, can facilitate the conversion of information into knowledge (Dash et al., 2019).

[1] https://www.smart-bear.eu/

Despite significant progress in this domain, challenges remain. As the scientific community does not have a commonly accepted method of systematically evaluating the captured information and derived knowledge, the challenge remains in determining how these resources can be utilized productively without being exploited commercially. A well-known specification for the representation of clinical data is the HL7 (Health Level Seven) FHIR (Fast Healthcare Interoperability Resources) standard – and we use it as the underlying basis for our data harmonization solution. HL7 FHIR also incorporates a well-defined semantics which is captured using widely accepted ontologies such as LOINC[2] and SNOMED-CT[3]. Standardizing the data representations will facilitate the development of analytics and decision models, with the potential to provide accurate, personalized interventions.

Data protection must also be adequately addressed in addition to knowledge production. All applicable legal requirements and privacy obligations must be met when processing sensitive personal data, including those imposed by the General Data Protection Regulation (GDPR), which is an EU legal framework that fundamentally changed how personal data is managed lawfully in the European Union. In this context, it is not sufficient just to have implemented organizational procedures and IT-enabled processes for exercising certain GDPR rights. Vulnerabilities do occur even in the most well-designed and well-coded IT applications. Furthermore, 82% of the healthcare providers have reported to experience attacks against their Medical IoT according to the Health Insurance Portability and Accountability Act statistics[4]. Therefore, continuous security and privacy assurance measures must be implemented, to ensure the security and privacy of the stored data, as well as the integrity of any platform on which they are stored and managed (integrity, confidentiality, and availability of data at rest, in transit, and during processing for data flows). In light of the legal obligations imposed by the GDPR and the state-of-the-art guidelines, such as the NIST encryption guidelines[5], data minimization, pseudo-anonymization, transparency in the processing of personal data, and audit support are among the appropriate technical (and organizational) measures that need to be considered, preferably at an early stage, to ensure that all legal requirements are met.

Last but not least, BDA systems for healthcare decision-making must not only focus on the production of ML knowledge but also convey it in an easy-to-use way. Currently, e-health systems do not appear to be rated satisfactorily in terms of their usability (Basdekis et al., 2012), while understanding ML models still remains an open question (Liao et al., 2020). Furthermore, the integration of ML models in the healthcare field continues to be criticized for not adhering to high standards of accountability, reliability, and transparency (Anderson, 2018). These limitations can be addressed by utilizing Explainable Artificial Intelligence (XAI) techniques, which aims to make ML results more understandable to humans, to increase the trust of end-users in the ML algorithms

[2]https://loinc.org/

[3]https://www.snomed.org/

[4]https://www.hipaajournal.com/82-of-healthcare-organizations-have-experienced-a-cyberattack-on-their-iot-devices/

[5]https://csrc.nist.gov/Projects/cryptographic-standards-and-guidelines

that produced them, and eventually their confidence in applying ML algorithms in sensitive domains. These systems, in particular, are being used within a high-stress environment, by non-technical end-users, and perhaps with time constraints that made the situation even worse. Thus, the acceptance and usability by the involved end-users of such functionality is a critical factor for its success and a key requirement in the SMART BEAR project.

The presented paper is an extended version of an earlier research paper (Peretokin et al., 2022). The discussion of the components in the SMART BEAR infrastructure has been greatly expanded here.

2 The SMART BEAR Architecture

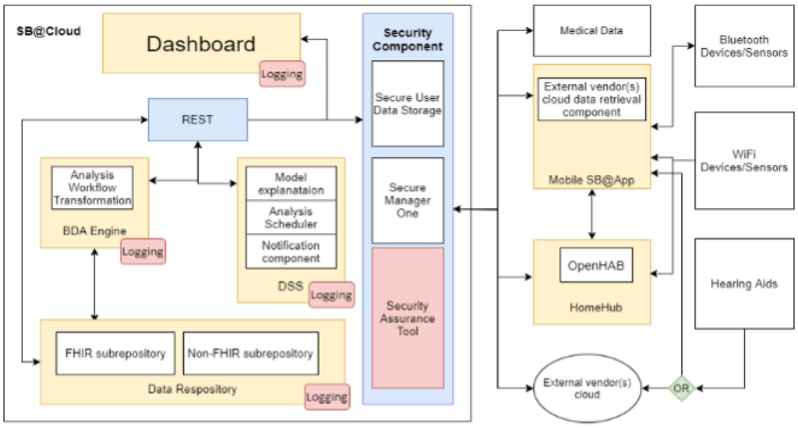

Fig. 1. The SMART BEAR Architecture (Peretokin et al., 2022)

There are three main systems in the SMART BEAR architecture as shown in Figure 1, namely the mobile phone application (SB@App), the SMART BEAR Home Hub (HomeHub), and the SMART BEAR Cloud (SB@Cloud). Data are continuously collected from all linked portable devices via the Mobile SB@App (e.g., hearing aid program, steps measurement, and blood pressure), and through the Mobile SB@App itself (e.g., questionaries about the diet, mood, sleep quality, or medication adherence). The HomeHub then accumulates data from different home-based sensors, such as movement sensors and weight scales. Finally, SB@Cloud is the core system that is responsible for the secure storage and model-driven big data analysis of the collected data, as well as personalized decision-making. There are several components within SB@Cloud, namely the BDA Engine, the Data Repository and its underlying Information Model, the Synthetic Data Generator, Decision Support System (DSS), and the Dashboard.

SB@Cloud interacts with the SB@App and the HomeHub, as well as the external medical system and device vendor clouds through Representational State Transfer

(RESTful) interfaces. All collected data from the SB@App and the HomeHub reaching the SB@Cloud are already anonymized before the transmission and are then stored in the Data Repository in compliance with the GDPR rules.

All components inside the SB@Cloud are also interconnected with the RESTful interfaces. The REST layer in the SB@Cloud is used for retrieving, saving, or analyzing the data that are store in the database, as well as provides the interfaces to both the Dashboard and external components to SB@Cloud. The REST services also allow interactions between the BDA Engine and the Data Repository, so that the end-users can perform a data analytics workflow with multiple tasks to the BDA Engine for analysis. Communications between components, as well as authentication at the Dashboard, are secured according to GDPR through the security component, which also ensures all security mechanisms are functioning correctly. Moreover, the security component facilitates interoperability with external platforms that represent medical/usage data using the FHIR standard. Finally, the Dashboard implements the user interface, allowing users to interact with the project infrastructure such as enter data, set up data Analysis Workflow models, validate and execute these models, register/unregister external data sources, and retrieve/visualize execution results. A secure, privacy-preserving, machine-to-machine bridge between two platforms, which was developed within the Smart4Health[6] and Holobalance[7] EU-funded projects, is currently being tested in the PoP.

3 The SMART BEAR Mobile Phone Applications

The SB@App component serves as a backbone for integration between the devices that are supplied to the recruited participants and the SB@Cloud system. These devices, depending on the comorbidities of the participants, are hearing aids, smart watch, smart blood pressure tracker, smart weight scale, smart glucometer, and smart phone. Therefore, SB@App is the main point of interaction between participants and the SMART BEAR platform. The SB@App interface aims to be user-friendly and contains functionalities that target the six medical comorbidities targeted by SMART BEAR, namely MyHeart, MyBalance, MyMood, MyDiary, MyDiet, MyHearing, MyMemory, MyMedication, MySmartBear and MyAppointments.

SB@App is responsible for sending all collected data to SB@Cloud, as well as receiving informational material and data analysis results performed by the platform. Furthermore, the functionalities of SB@App also include notifications/alerts management, calendar-based appointment setup with clinicians, questionnaire and surveys, and reporting on participants' interaction with the SMART BEART platform. SB@App connects to the available devices using either their Application Programming Interface (API) or their respective Software Development Kit, depending on what is available for each.

[6]Smart4Health: Citizen-centered EU-HER exchange for personalized health. https://smart4health.eu/

[7]Holobalance: Holograms for personalized virtual coaching and motivation in an ageing population with balance disorders. https://holobalance.eu/

4 The SMART BEAR Home Hub

The SMART BEAR HomeHub component is based on the openHAB platform[8] – an open-source implementation towards a common approach in addressing security/software development and protocol connectivity concerns of Smart IoT. In SMART BEAR, the HomeHub monitors the use of light sources, temperature, humidity, and movement inside a patient's home. Another reason why openHAB was chosen as the HomeHub solution is because it allows sensors or devices from different vendors to be integrated in a single solution.

5 The SMART BEAR Cloud Components

A detailed discussion of each SB@Cloud component is presented in this section to allow a fuller picture of their utility and how they fit to the overall architecture.

5.1 Data Repository

Database Implementation

The data repository component of SB@Cloud contains a combination of FHIR-compliant and non-FHIR databases. The FHIR database stores those data that represent medical entities, whereas the non-FHIR database stores data related to non-medical entities. The non-FHIR database contains data that are not mapped to FHIR models, such as dashboard user settings or intermediate results of the analytics models when these are applied to FHIR data. Data transmitted by the HomeHub can also be stored in the non-FHIR database. Finally, intervention, notifications, and alerts generated by the DSS are also stored in the non-FHIR database.

Data Model Specification Compliant with FHIR

HL7 FHIR is the latest standard from HL7, an international standards development organization that has been publishing healthcare interoperability standards since 1989. The FHIR standard incorporates the best of and builds upon the lessons learned from the different approaches taken previously by HL V2 and HL7 V3, while simultaneously using well-known, modern technologies such as REST and JSON. In addition to providing out-of-the-box tooling, the standard is published for free and is open source. Therefore, FHIR was chosen to be used within SMART BEAR as the standard for clinical

[8] https://www.openhab.org/

data for its speed and ease of implementation, as well as the fact that REST and JSON are an especially good fit for mobile applications, which the project makes use of.

The FHIR standard is also used by a number of leading international organizations that provide solutions to specific healthcare problems. Among these organizations is Integrating the Healthcare Enterprise (IHE)[9], which is an initiative by healthcare professionals and industry to improve the way computer systems share health information, as well as the Personal Connected Health Alliance (PCHA)[10], which is a membership-based Healthcare Information and Management Systems Society Innovation Company that develops the Continua Design Guidelines (CDG) in order to advance patient-centered health, wellness, and disease prevention. IHE and PCHA are updating their technical specifications to incorporate FHIR. Furthermore, FHIR is also used nationally in The Netherlands as part of the MedMij project[11], and is implemented in Estonia's national electronic health record system as well. Several countries, including the Netherlands, Switzerland, and Belgium, have established national core profiles for FHIR that standardize clinical information relevant to the respective countries. A standard that is gaining such strong acceptance in Europe will make it easier to support future needs in data exchange.

Due to the nature of the data treated in the project, and in accordance with the FHIR standard, an Implementation Guide (IG) was required. For this reason, an analysis of the IGs published on the FHIR registries was carried out. Among these, particular attention was paid the Personal Health Device (PHD)[12] and International Patient Summary (IPS)[13] IGs.

The PHD IG adapts FHIR resources to transmit measurements and supporting data from PHDs to different types of systems, such as electronic medical records and clinical decision support platforms. This IG is of particular interest due to the fact that it is based on the CDG as well as the ISO/IEEE 11073 PHD Domain Information Model (Huang et al., 2020). In spite of this, given that many health data gathered in SMART BEAR are questionnaires rather than PHDs, this IG was not considered appropriate for the SMART BEAR project.

The IPS IG defines the rules to produce a document containing the essential healthcare information about a subject of care. IPS is designed for, but not limited to, supporting unplanned, cross-border care. Although this IG provides an important contribution to identify a minimal, specialty-agnostic, condition-independent, clinically relevant dataset for a patient, it was also not considered relevant for the SMART BEAR project.

For these reasons, the project defined a dedicated SMART BEAR IG in compliance with the FHIR standard and in line with the choices adopted in many European projects. A set of identified FHIR resources is used to profile the SMART BEAR IG, along with the terminologies individuated from the international standard code systems, as well as internal value sets. The tool chosen for modelling the FHIR information model is

[9] https://www.ihe.net/
[10] https://www.pchalliance.org/
[11] https://medmij.nl/en/home/
[12] http://hl7.org/fhir/uv/phd/
[13] https://hl7.org/fhir/uv/ips/

SUSHI[14], considering that it integrates well with the IG publisher which is an official tool provided by HL7. Currently, the published IG consists of 84 profiles (of type Observation, Condition, Questionnaires, Bundle, Patient, DeviceUseStatement, FamilyMemberHistory, MedicationStatement, ResearchSubject), 2 extensions, 33 value Sets, and 133 examples.

The Clinical Data Repository

The SMART-BEAR Clinical Data Repository (CDR) is based on the Health Data Hub, which is built around the HL7 FHIR standard. The CDR is also able to structure and dispose of clinical information using the FHIR standard as the specification. Thus, SMART BEAR CDR is capable of storing and serving clinical information in a secure, scalable, and HL7 standardized manner. Furthermore, this allows the BDA and DSS developers to focus on developing algorithms and applications appropriate to the requirements of the SMART BEAR pilot program, enabling them to create a common set of products and solutions that are seamlessly connected using standardized information. SNOMED-CT will be used to annotate medical terminology that is not fully covered by FHIR. By adapting the Atos Terminology Server (ATS), some of the different clinical terminologies commonly used across the healthcare industry, such as ICD9 and LOINC, will become interoperable. The implementation and customization of the ATS will occur in the second phase after the finalization of the PoP, and a RESTful API will also be provided. As a result of this API, clinical information may be accessed safely through interaction with the FHIR database for terminology purposes.

5.2 Synthetic Data Generator

In order to ensure fitness for purpose of a system of such complexity during its development, it is essential to test it with realistic data and use this information to guide its design and development. Therefore, we have adopted Synthea[15], a synthetic patient generator that generates realistic patient records pertaining to the entire life of a patient, including condition onset, encounters with physicians, observations, and prescriptions.

5.3 Security

Data protection is considered a critical issue, especially when dealing with special categories of personal data (Article 9, GDPR). In this context, SB@Cloud, by virtue of its design, supports privacy. In particular, the Security Component provides mechanisms that handle data minimization, authentication, and other security and privacy aspects through pseudonymization and resource identifier reassociation (Basdekis et al., 2019). In order to protect the transmission of any (sensitive or not) data, this component supports Role-based Access Control authentication and authorization of all RESTful API

[14] https://fshschool.org
[15] https://synthea.mitre.org/

endpoints and introduces services for the management of privacy-related requests in order to demonstrate compliance with GDPR. More specifically, the RESTful API implements token-based access via encrypted HTTPS connections. Due to the fact that the data is stored in two separate repositories, where pseudonymized medical and usage data are stored in the CDR and personal data and Personalized Identifiable Information are stored in a separate encrypted repository, it is possible to continue analyzing fully anonymized data after the project has concluded, provided all personal information has been deleted. Therefore, after the completion of the SB project, the Security Component data will no longer be needed for research purposes (e.g., analytics and interventions) and will be disposed of.

In parallel, The Security Component is also responsible for monitoring, testing, and assessing the security and privacy of all platform operations. A comprehensive audit of key infrastructure components and processes will also be performed, as well as leveraging monitoring mechanisms developed in the context of the project to provide an evidence-based, certifiable assessment of the platform's security posture, along with accountability provisions for changes in the security posture and analyses of the cascading effects of those changes. In addition to several built-in security assessments addressing Confidentiality, Integrity, and Availability principles, custom metrics related to the platform's components will be used, utilizing an evidence-based approach to provide security and privacy assurance assessments with certifiable results.

5.4 Information Model

As described above, data is partitioned across two databases in SMART BEAR – clinical data in the FHIR database, and non-clinical or private information that is not exposed to analytics in a non-FHIR one.

Due to the fact that FHIR is a platform specification meant to be confined to a specific use case, we have profiled various resources in the FHIR database according to our requirements (Figure 2). Basic demographic information such as name, age, and ethnicity are stored in the Patient resource, whereas most of the clinical information is stored in Conditions and Observations, which are tied to the Encounter resource.

Each assessment is represented by an Encounter resource instance, since patient assessments are performed by clinicians in SMART BEAR. This Encounter resource is central to the information model, as all other resources either link to or from it, creating a graph in which all relevant nodes (resources) can be reached. The Conditions resource records any clinical issues that were noted during an assessment. FHIR Observations contain issues of lesser importance, as well as 'negations' - issues that a clinician has determined that the patient does not have. As a result of this fine but important distinction between a lack of data (unknown value) and a refuting observation (known negative), we can develop more accurate analytics algorithms. Furthermore, a significant part of the data acquired by the clinical assessments comes in a form of over 20 Questionnaires; these are internationally recognized; standard data collection points whose outcome scores will be used for analytical purposes.

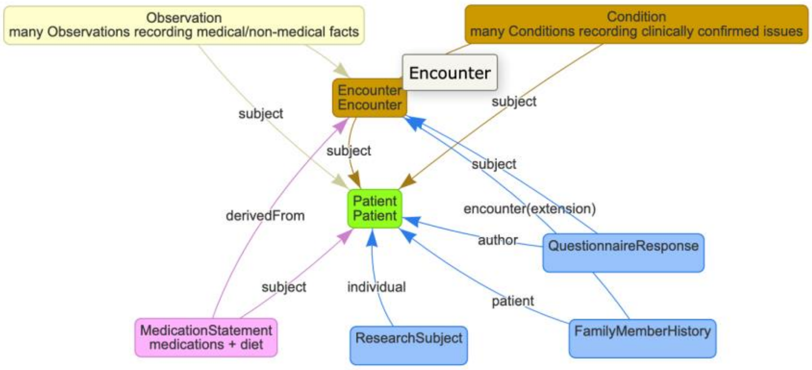

Fig. 2. Information Model of FHIR Resources in Use (Peretokin et al., 2022)

Most Observations follow a simple 'key-value' pattern, where Observation.code identifies the type of measurement, or type of condition in case of Conditions, and Observation.value[x] records the measurement value. As an example, in case of the patient having anxiety, Condition.code will be populated with “197480006 |Anxiety disorder|” from SNOMED. Should they not be affected by anxiety, Observation.code will have the same terminology code but Observation.valueCodeableConcept will be populated with “260385009 |Negative|”.

FHIR Vital Signs standard profiles is also adhered to where possible. For example, in the case of blood pressure, the Observation.component is utilized to record systolic/diastolic measurements, BodySite for the left or right arm, and LOINC codes are used to indicate the patient's standing/supine position. SMART BEAR IG contains a wealth of Conditions, as well as positive and negative Observation examples to assist users in understanding the FHIR database.

Furthermore, specialized resources are used where appropriate. FamilyMemberHistory for example is used to record the family history of hearing loss and ResearchSubject is used to record the source of referral to our clinical study. MedicationStatement records both the list of medications the patient is taking using the Anatomic Therapeutic Chemical value set endorsed by the World Health Organization (WHO), and the diet they are prescribed.

Previously mentioned Conditions and Observations relied on over 120 terminology mappings, with most codes coming from SNOMED, to link the semantic meaning within. Codes from LOINC and MeSH[16] complement the rest of the mappings. In order to prevent the development of new medical knowledge, the creation of custom codes was avoided as much as possible. Only four new codes have been introduced in

[16] https://www.nlm.nih.gov/mesh/meshhome.html

SMART BEAR, which have no equivalents in any of the searched code systems. Several food/diet-related concepts that are not available in the SNOMED international core, but are available in the Australian edition, were used for this reason. We have verified that this does not impose any additional licensing constraints on SNOMED.

To adhere to the principle that introducing new codes should be avoided as much as possible, two SNOMED post-coordinated expressions were crafted to accurately represent very specific concepts: "number of non-scheduled visits due to volume overload in subjects with heart failure" as:

> 4525004 |emergency department patient visit| :362981000 |qualifier value| = 260299005 |number|, 42752001 |due to| = 21639008 |hypervolemia|

and “number of Visits to the Emergency Room due to Hypertension peak” as:

> 4525004 |emergency department patient visit| :362981000 |qualifier value| = 260299005 |number|, 42752001 |due to| = 38341003 |hypertension|

We chose to re-use the FHIR extension and value set published by the German Corona Consensus Data Set project (Sass et al., 2020), which is partially based on WHO ISARIC eCRF value set when it came to recording the patient's ethnicity. As a result of the reuse of existing knowledge, long-term interoperability is enhanced.

Analytics are a crucial aspect of the system, as it provides the necessary intelligence for the task at hand. They are driven by the BDA Engine, which has several requirements placed upon it – raw data processing, incremental updates, and scalability. Clinical data are stored in the system in a FHIR repository, as previously mentioned. Despite its advantages as an excellent interface for clinical data, FHIR interfaces make compromises when processing bulk data. It is for this reason that the BDA engine requires the capability of converting and flattening the hierarchical format of FHIR into a relational format that is more appropriate for bulk data processing. In order to run analytics continuously, this conversion should be possible to do incrementally as new data is received in the clinical repository, as well as being able to scale to large data volumes.

5.5 BDA Engine

The BDA Engine mainly addresses the functionalities required for processing Data Analysis Workflows, as well as providing and storing the execution results. A set of APIs is provided to perform analysis on raw data. In terms of ML, a preliminary extraction of data analytics - which will be carried out on the pre-processed datasets - are going to indicate variables or combinations of variables for the feature selection approach. It is important to note that all ML methods and techniques are data-driven, and the "best" method will be determined after its application. A longer, more detailed, discussion of how the BDA component's AI & XAI capabilities are planned to be used, in particular in the setting of the Hearing Loss comorbidity, is presented elsewhere (Iliadou et al., 2022).

The preliminary extraction of data analytics is performed by the following subcomponents featured in the BDA Engine architecture: Delta Lake [17], Spark[18], Trino[19], and Airflow[20]. A bottom-up approach will be used to describe the components, with the layers at the bottom being closest to the data repositories. Figure 3 illustrates its architecture, which is an expanded version of the architecture presented in (Anisetti et al., 2021).

A cloud object store, Delta Lake, provides ACID[21] table storage and is the closest component to the data repositories. With Delta Lake, a Lakehouse Architecture can be built using existing storage systems, including Amazon S3, Azure Data Lake Storage, Google Cloud Storage, and Hadoop Distributed File System (HDFS)[22] (Armbrust et al., 2020). The Lakehouse Architecture also enables Business Intelligence and ML analysis on all data. In the case of SMART BEAR, the adopted standard is HDFS.

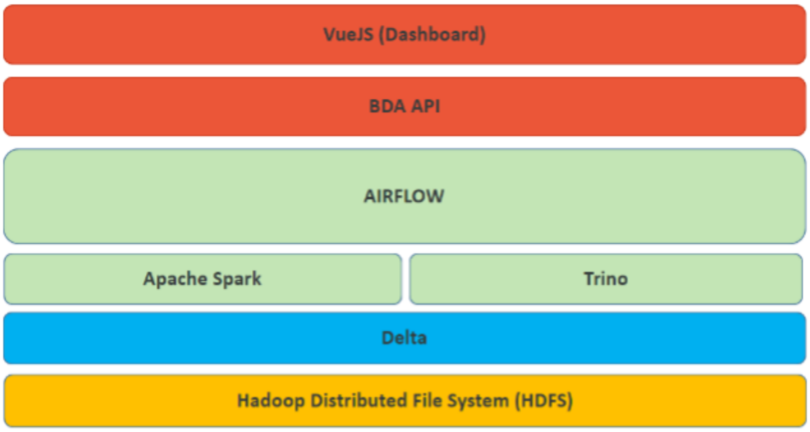

Fig. 3. The BDA Architecture (Peretokin et al., 2022)

On the third layer from the bottom, Spark and Trino are collocated together, providing the capability to access data and perform queries on those datasets. Spark is a multi-language engine supporting data engineering, data science, and ML on both single-node machines and clusters. Spark was chosen due to its capabilities of processing tasks encompassing custom analytics on large data volumes, as well as the fact that it features many bindings with other commonly used Data Science and ML libraries. Additionally, Spark is capable of handling batch and streaming data. Trino on the other hand provides

[17] https://delta.io/
[18] https://spark.apache.org/
[19] https://trino.io/
[20] https://airflow.apache.org/
[21] ACID is an acronym refers the four properties that define a transaction: Atomicity, Consistency, Isolation, and Durability.
[22] https://hadoop.apache.org/docs/r1.2.1/hdfs_design.html

the capability of accessing and processing data from multiple systems in a highly parallel and distributed manner. In addition to supporting HDFS data, Trino also provides the BDA Engine with the ability to manage On-Line Analytical Processing queries and data warehousing tasks.

Airflow is located on the fourth layer from the bottom and allows programmatic authoring, scheduling, and monitoring of workflows written in Python.

5.6 Decision Support System

The DSS is intended to aid clinicians in assessing every patient in terms of the optimal assessments that must be completed to assess the patient, and to provide them with the optimal combination of devices to monitor their health during the pilot study. As a result of the continuous collection and analysis of data that will be digested into the platform, this component is designed to evolve throughout the project. Initially, the DSS available for the PoP was developed in accordance with the rules and medical guidelines provided by the clinicians so as to establish a ground truth system that is based on the most current medical knowledge. For each of the monitoring conditions of the SMART BEAR project (Hearing Loss, Cardiovascular Diseases, Mild Cognitive Impairment, Mild Depression, Balance Disorders, and Frailty), the medical teams provide the rules-based scenarios and relevant interventions that should be administered to the participants. Rules-based algorithms take into account the personalized thresholds that are set for each patient individually. Cardiovascular Diseases, for example, have optimal and extreme cut-off values for blood pressure, which trigger notifications and alerts to the patient and clinical care team.

Data collected at the PoP will be used to develop BDA engine models, and the output from the analytics will be analyzed in conjunction with the measured parameters in order to determine the extent to which patients are satisfied and what adjustments need to be made to the personalized thresholds. It is possible for the DSS to be extended to support all the new interventions that will be provided by clinicians if the results of the analytics provide insights that lead to new interventions. It must be noted that any new intervention must first be validated by the clinicians before it is included in the interventions provided to the patient.

5.7 Dashboard

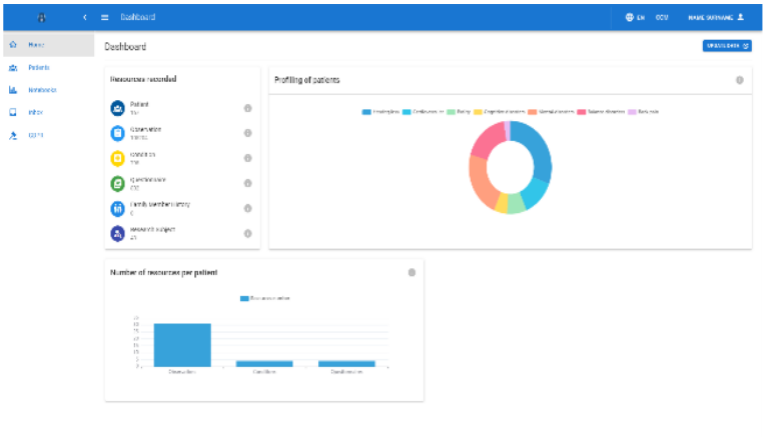

Fig. 4. Dashboard Homepage (Peretokin et al., 2022)

The SMART BEAR Dashboard is a component aimed at providing clinicians with a user-friendly graphical user interface. The Dashboard home page is shown in Figure 4. Clinicians can utilize the Dashboard to create and manage patients, taking into account their devices and medications, conducting first visits and check-ups, performing analytics on data, and developing interventions. All collected data are stored in FHIR and non-FHIR repositories depending on their clinical value: the first collection takes place during the Baseline Assessment of a patient, which includes medical history, physical examinations, and questionnaire responses. By analyzing the information provided, the dashboard visualizes suggestions about the eligibility of prospective participants for the SMART BEAR pilot studies. According to the conditions detected, the profiling functionality suggests specific tabs and questionnaires that should be activated by the clinician. Although the patient's profile is ultimately selected by a clinician, the profiling functionality redirects users to the clinical tools and devices that are required to match a patient's profile consistently with the SMART BEAR protocol, regardless of which clinical tool is chosen. Upon creation and eligibility determination of a patient, the Dashboard displays specific tabs that enable patient management and provide information regarding demographics, such as living situation and ethnicity, participation in synergies, and type and status of devices provided.

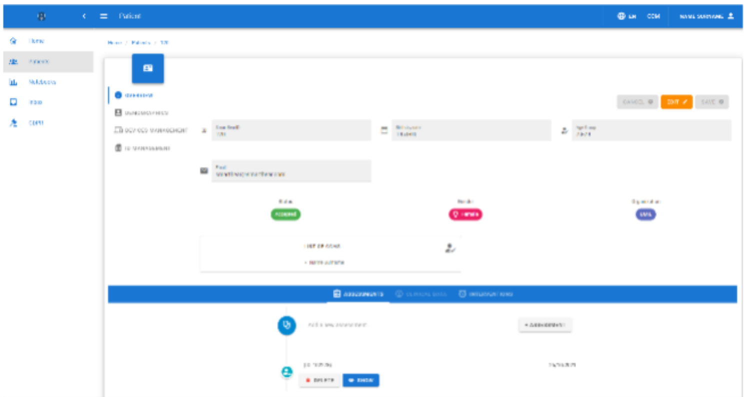

Fig. 5. Patient Management Page (Peretokin et al., 2022)

As shown in Figure 5, the patient management tab is another featured functionality that allows users to visualize the delivered notifications. Currently, the analytics and intervention mechanism are still in the development phase. These mechanisms aim to assist the clinicians to perform analytics on the collected data either targeting all patients or only a specific subgroup defined by certain parameters, in order to monitor the patients in the future with a determined condition. Based on the outcome of the analytics and with support from the DSS, the Dashboard visualizes suggestions for clinicians on the interventions to be delivered. The final choice of the intervention is still to be made by clinicians, and they will also be able to monitor the intervention outcome. Examples of analytics to be made available in the Dashboard are discussed in (Bellandi et al., 2021).

6 Interaction Specifications

Having introduced each component in the SMART BEAR architecture, this section presents an example of how these cooperate, though some interaction specification diagrams. These diagrams are representative of the main data flows in SMART BEAR (Kloukinas et al., 2020). The example described here is the MyDiet functionality of the SB@App.

In order for the MyDiet functionality to suggest appropriate dietary recommendations to a participant, their current weight needs to be collected first by using the smart weight scale. If the vendor of the smart weight scale has its own mobile application, then this vendor-provided application will be available to the participant and will transfer the new measurements to the vendor cloud, from where the Data Repository of SB@Cloud will retrieve this information periodically. This smart weight scale – vendor application – vendor cloud – Data Repository data flow is shown in Figure 6.

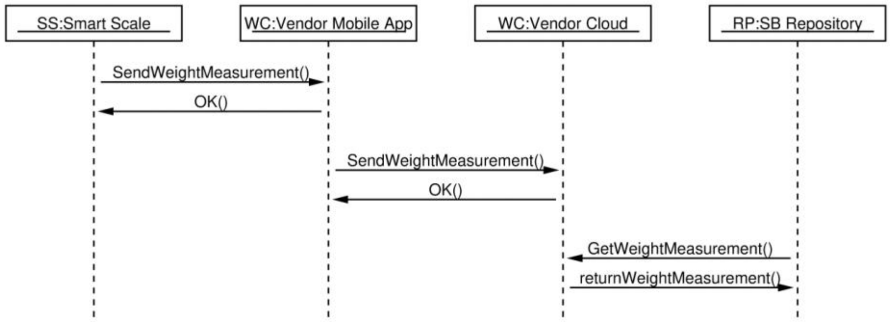

Fig. 6. Data flow of smart weight scale to the SMART BEAR Data Repository (Kloukinas et al., 2020)

In the case the vendor of the smart weight scale does not have its own mobile application, Figure 7 shows the alternative data flow where the new weight measurements will be collected by the SB@App and transmit them to the HomeHub, which then transmits them to the Data Repository.

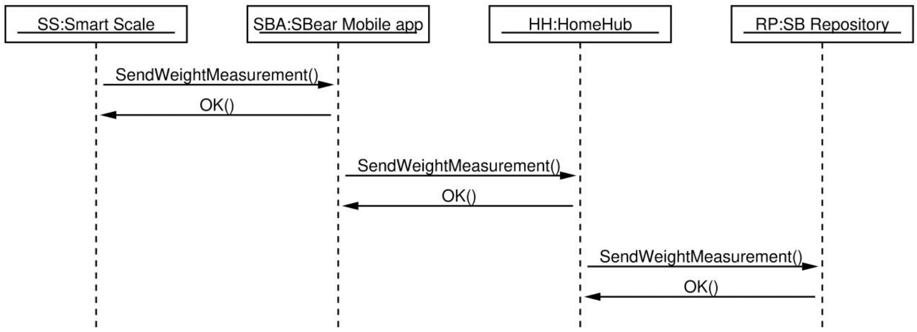

Fig. 7. Alternative data flow of smart weight scale to the SMART BEAR Data Repository – in the case of absence of vendor-provided mobile application (Kloukinas et al., 2020)

Figure 8 shows an alternative data flow of Figure 7 if the HomeHub is not available. In this case, SB@App anonymizes its own data and connects directly to the SB@Cloud to in order to transfer the data to the Data Repository.

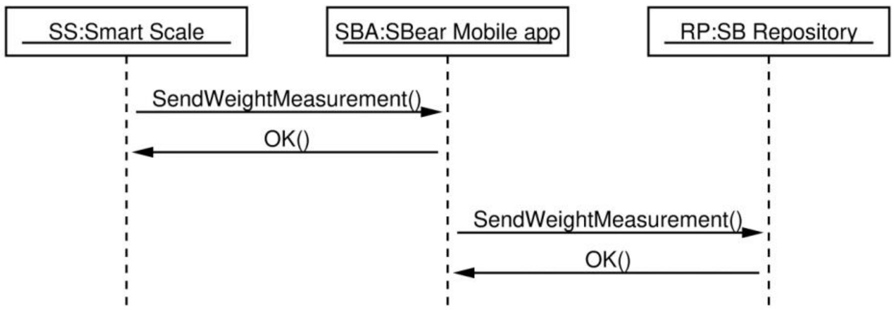

Fig. 8. Alternative data flow of smart weight scale to the SMART BEAR Data Repository – in the case of absence of the HomeHub (Kloukinas et al., 2020)

Finally, Figure 9 shows how the DSS takes into account the measurements stored in the Data Repository, along with other data such as particular dietary requirements from the participant's profile, to form a set of recommended recipes for the participant. This recommendation is then transmitted to the SB@App and notes the choices made by the participant through the MyDiet functionality. The choice is then transmitted back to the Data Repository to allow future analysis of appropriate recipes and uptake of the suggestions of the MyDiet functionality.

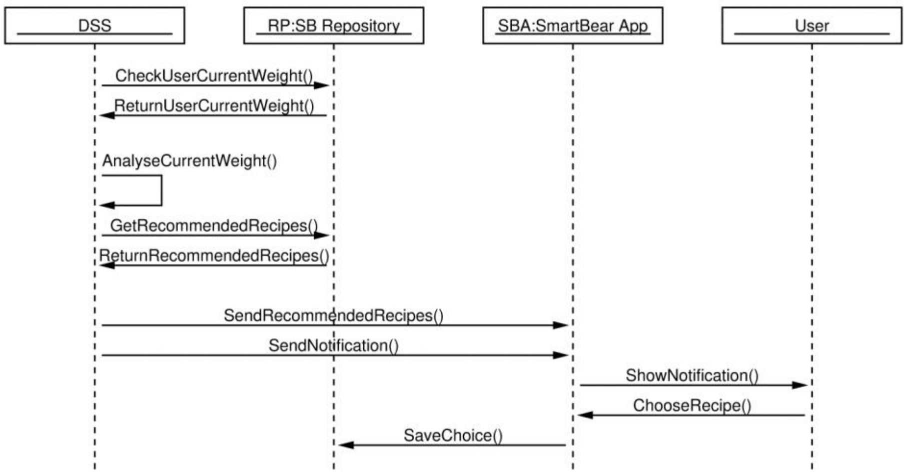

Fig. 9. DSS to MyDiet recommendation use case

7 Future Work

In order to demonstrate the efficacy, extensibility, sustainability, and cost-effectiveness of SB@Cloud, it will operate for at least three years where its solution will be tested and validated through five large-scale pilots involving 5000 elderly individuals living at home in Greece, Italy, France, Spain-Portugal, and Romania. It is expected to generate useful evidence during this period, such as metrics and observational evidence base, from analysis of the collected data that is driven by high-level BDA and decision models for offering personalized healthcare and medicine solutions in clinical practice. Since the pseudonymization mechanism is in place, SMART BEAR intends to develop a data sharing and valorization model (DSVM) that will support further analysis using anonymous data even after the lifecycle of the project. Through the integration of new data providers and open sources, the DSVM will identify methods for extending the data collected by SMART BEAR on both a technical and organizational level. Using the outcomes of data analysis, we will be able to enhance the platform's performance, personalize its relationship with its end-users further, develop new services, and monetize data-intensive services.

8 Conclusion

This paper provides an overview of the cloud-enabled, standards-based integrated system developed by the SMART BEAR project. This system allows for recording assessments and monitoring, as well as delivering clinician-vetted interventions to facilitate monitoring, empowering, and promoting healthy living at home for senior citizens. Based on widely accepted standards such as HL7 FHIR and advanced analytics, the system is supported by an underlying semantic interoperability solution. It is intended to leverage the platform during the SMART BEAR PoP, and further refine it to support the planned large-scale pilots in all participating countries.

9 Acknowledgement

This work was supported by the European Commission’s Horizon 2020 research and innovation program under the SMART-BEAR project, grant agreement No 857172.

References

Adhikary, T., Jana, A. D., Chakrabarty, A., & Jana, S. K. (2020). The Internet of Things (IoT) Augmentation in Healthcare: An Application Analytics. In *ICICCT 2019 – System Reliability, Quality Control, Safety, Maintenance and Management* (pp. 576–583). Springer Singapore. https://doi.org/10.1007/978-981-13-8461-5_66

Akyildiz, I., Pierobon, M., Balasubramaniam, S., & Koucheryavy, Y. (2015). The internet of Bio-Nano things. *IEEE Communications Magazine*, *53*(3), 32–40. https://doi.org/10.1109/MCOM.2015.7060516

Anderson, C. (2018). Ready for Prime Time?: AI Influencing Precision Medicine but May Not Match the Hype. *Clinical OMICs*, *5*(3), 44–46. https://doi.org/10.1089/clinomi.05.03.26

Anisetti, M., Ardagna, C. A., Braghin, C., Damiani, E., Polimeno, A., & Balestrucci, A. (2021). Dynamic and Scalable Enforcement of Access Control Policies for Big Data. *Proceedings of the 13th International Conference on Management of Digital EcoSystems*, 71–78. https://doi.org/10.1145/3444757.3485107

Armbrust, M., Das, T., Sun, L., Yavuz, B., Zhu, S., Murthy, M., Torres, J., van Hovell, H., Ionescu, A., Łuszczak, A., Świtakowski, M., Szafrański, M., Li, X., Ueshin, T., Mokhtar, M., Boncz, P., Ghodsi, A., Paranjpye, S., Senster, P., … Zaharia, M. (2020). Delta lake. *Proceedings of the VLDB Endowment*, *13*(12), 3411–3424. https://doi.org/10.14778/3415478.3415560

Article 9, General Data Protection Regulation (GDPR). 2018. *General Data Protection Regulation (GDPR) – Final text neatly arranged*. [online] Available at: <https://gdpr-info.eu/> [Accessed 20 August 2022].

Basdekis, I., Pozdniakov, K., Prasinos, M., & Koloutsou, K. (2019). Evidence Based Public Health Policy Making: Tool Support. *2019 IEEE World Congress on Services (SERVICES)*, 272–277. https://doi.org/10.1109/SERVICES.2019.00080

Basdekis, I., Sakkalis, V., & Stephanidis, C. (2012). Towards an Accessible Personal Health Record. In *Wireless Mobile Communication and Healthcare* (pp. 61–68). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-29734-2_9

Bellandi, V., Basdekis, I., Ceravolo, P., Cesari, M., Damiani, E., Iliadou, E., Dan Marzan, M., & Maghool, S. (2021). Engineering Continuous Monitoring of Intrinsic Capacity for Elderly People. *2021 IEEE International Conference on Digital Health (ICDH)*, 166–171. https://doi.org/10.1109/ICDH52753.2021.00030

Dash, S., Shakyawar, S. K., Sharma, M., & Kaushik, S. (2019). Big data in healthcare: management, analysis and future prospects. *Journal of Big Data*, *6*(1), 54. https://doi.org/10.1186/s40537-019-0217-0

Huang, Z. Y., Wang, Y., & Wang, L. (2020). ISO/IEEE 11073 Treadmill Interoperability Framework and its Test Method: Design and Implementation. *JMIR Medical Informatics*, *8*(12), e22000. https://doi.org/10.2196/22000

Iliadou, E., Su, Q., Kikidis, D., Bibas, T., & Kloukinas, C. (2022). Profiling Hearing Aid Users through Big Data Explainable Artificial Intelligence Techniques. *Frontiers in Neurology*.

Kloukinas, C., Basdekis, I., Bucur, A., Cavero, C., Giotis, G., Kalatzis, F., Kouris, I., Maggesi, J., & Pozdniakov, K. (2020). *Smart Big Data Platform to Offer Evidence-based Personalised Support for Healthy and Independent Living at Home*. Deliverable to the SMART BEAR (857172) Project funded by the European Union D6 – SMART BEAR Architecture (Public), City, University of London, UK. [online] <https://www.smart-bear.eu/wp-content/uploads/2021/06/D6-D2.2.pdf > [Accessed 21 August 2022].

Liao, Q. V., Gruen, D., & Miller, S. (2020). Questioning the AI: Informing Design Practices for Explainable AI User Experiences. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–15. https://doi.org/10.1145/3313831.3376590

Mshali, H., Lemlouma, T., Moloney, M., & Magoni, D. (2018). A survey on health monitoring systems for health smart homes. *International Journal of Industrial Ergonomics*, *66*, 26–56. https://doi.org/10.1016/j.ergon.2018.02.002

Peretokin, V., Basdekis, I., Kouris, I., Maggesi, J., Sicuranza, M., Su, Q., Acebes, A., Bucur, A., Mukkala, V., Pozdniakov, K., Kloukinas, C., Koutsouris, D., Iliadou, E., Leontsinis, I., Gallo, L., de Pietro, G., & Spanoudakis, G. (2022). Overview of the SMART-BEAR Technical Infrastructure. *Proceedings of the 8th International Conference on Information and Communication Technologies for Ageing Well and E-Health*, 117–125. https://doi.org/10.5220/0011082700003188

Sass, J., Bartschke, A., Lehne, M., Essenwanger, A., Rinaldi, E., Rudolph, S., Heitmann, K. U., Vehreschild, J. J., von Kalle, C., & Thun, S. (2020). The German Corona Consensus Dataset (GECCO): a standardized dataset for COVID-19 research in university medicine and beyond. *BMC Medical Informatics and Decision Making*, *20*(1), 341. https://doi.org/10.1186/s12911-020-01374-w