City Research Online

City St George's, University of London

Citation: Polydoratou, P. (2006). Use and functionality of metadata registry systems. (Unpublished Doctoral thesis, City, University of London)

This is the accepted version of the paper.

This version of the publication may differ from the final published version. To cite this item please consult the publisher's version.

Permanent repository link: https://openaccess.city.ac.uk/id/eprint/30748/

Copyright and Reuse: Copyright and Moral Rights remain with the author(s) and/or copyright holders. Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge, unless otherwise indicated, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way. For full details of reuse please refer to City Research Online policy.

City Research Online: http://openaccess.city.ac.uk/ publications@citystgeorges.ac.uk

Use and functionality of metadata registry systems

Panayiota Polydoratou

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

City University, London

Department of Information Science

July 2006

Contents

LIST OF FIGURES — **IV**

LIST OF TABLES — **VI**

ACKNOWLEDGEMENTS — **VII**

DECLARATION — **VIII**

ABSTRACT — **IX**

CHAPTER 1 – INTRODUCTION — **1**

1.1 AIMS AND OBJECTIVES — 2

1.2 DEFINITIONS — 3

1.3 SCOPE — 4

1.3.1 Users of metadata registry systems — *5*

1.3.2 Metadata registry systems and other cases studied — *6*

1.3.2.1 SCHEMAS project, UK — 6

1.3.2.2 MetaForm registry, Germany — 8

1.3.2.3 The Environmental Data Registry (EDR)/System of Registries, USA — 9

1.4 OUTLINE OF THE THESIS — 13

CHAPTER 2 - LITERATURE REVIEW — **14**

2.1 ORGANISATION OF INFORMATION — 14

2.2 METADATA — 17

2.3 OVERVIEW OF METADATA RESEARCH AND IMPLEMENTATION — 17

2.4 METADATA REGISTRY SYSTEMS — 21

2.5 OVERVIEW OF METADATA REGISTRY SYSTEMS RESEARCH AND IMPLEMENTATION — 24

2.5.1 Users of metadata registry systems — *29*

2.5.2 Role and functionality of metadata registry systems — *30*

2.5.3 Use of metadata registry systems — *31*

2.5.4 Metadata registry systems and interoperability — *32*

CHAPTER 3 – METHODOLOGY — **34**

3.1 INTRODUCTION — 34

3.2 INTERNET-BASED RESEARCH STUDIES — 35

3.3 MIXED METHODS RESEARCH — 38

3.4 RESEARCH METHODS USED — 39

3.4.1 Web log transactions — *41*

3.4.1.1 Sampling and recruiting — 43

3.4.1.2 Metrics used — 46

3.4.2 Questionnaire surveys — *52*

3.4.2.1 Sampling and recruiting — 53

3.4.2.2 Questionnaire questions — 58

3.4.3 Email interviews — *69*

3.4.3.1 Sampling and recruiting — 69

3.4.3.2 Interviews questions — 71

CHAPTER 4 – RESULTS — **73**

RECEIVED RESPONSE BY METHOD USED — 74

PART A – CHARACTERISTICS OF THE SAMPLE POPULATION 77

A.1 Response by age group 77
A.2 Response by occupation 78
A.3 Response by field of interest 79
A.6 Response by familiarity with using information resources 83
A.6 Response by frequency and longevity of use 84
A.4 Response by domain (Domain reports) 85
A.5 Response by referrer (Referrer reports) 89

PART B – USE OF METADATA REGISTRY SYSTEMS 94

B.1 Metadata terminology 94
B.2 Use of metadata element sets 96
B.3 Applications of metadata element sets 98
B.4 Use of metadata registry systems 100
B.4.1 Use by day of the week (Daily summary) 101
B.4.2 Monthly use (successful requests for pages) 106
B.4.3 Yearly use (successful page requests) 110
B.4.4 Increase/Decrease of use (successful requests for pages per day) 112
B.4.5 Use by type of information (Directory requests) 114
B.4.6 Use by type of information (Requests report) 115
B.4.7 Use of features/services 116
B.5 Reasons for using metadata registry systems 122
B.7 Requirement for use by the organisation 126
B.9 Metadata registry systems applications 127
B.10 Metadata registry systems – applications in use 131

PART C – FUNCTIONALITY OF METADATA REGISTRY SYSTEMS 133

C.1 Essential features/services of metadata registry systems 133
C.2 Secondary services of metadata registry systems 136
C.3 Functionality of metadata registry systems 136
C.4 Interoperability 139
C.5 Barriers to use 142
C.6 Failed requests 144
C.7 Non use 148

PART D – COMMENTS 150

D.1 Definitions of metadata registry systems 150
D.2 General comments 151

CHAPTER 5 – DISCUSSION AND CONCLUSIONS 158

5.1 DISCUSSION AND CONCLUSIONS 158

5.1.1 Characteristics of metadata registry systems users 159
5.1.2 Metadata use 161
5.1.3 Metadata registry systems use 162
5.1.4 Metadata registry systems functionality 165
5.1.5 User expectations 167
5.1.6 XML and metadata registry systems 168
5.2 Limitations of the study 168
5.3 Future work 169

BIBLIOGRAPHY 170

APPENDICES 186

A1-PUBLICATIONS 186

A2- QUESTIONNAIRES 187

A3A – DISCUSSION LISTS 200

List of figures

Figure 1: Introduction - Graphical representation of the EPA's metadata registries (*Source: Standard Update: Better Data through standards, Vol. 2, No. 3, Winter 2000)* 10
Figure 2: Literature Review - Graphical representation of the EPA's metadata registries (*Source: Standard Update: Better Data through standards, Vol. 2, No. 3, Winter 2000)* 27
Figure 3: Methodology - Environmental Data Registry - Example of log file, March 1996 45
Figure 4: Methodology - System of Registries - Example of log file, January 2004 46
Figure 5: Results - Response by age groups 77
Figure 6: Results - Response by post of the respondents 79
Figure 7: Results - Discussion Lists, SCHEMAS workshop and MetaForm surveys - Response by domains 86
Figure 8: Results - MetaForm - Web log transactions - Access representation by domain (1998-2004) 87
Figure 9: Results - MetaForm - Web log transactions - Domain requests by year 87
Figure 10: Results - EDR/SoR - Web log transactions - Access representation by domain (1998-2004) 88
Figure 11: Results - EDR/SoR - Web log transactions - Domain requests by year 89
Figure 12: Results - System of Registries - Web log transactions - Example of referrer report 90
Figure 13: Results - System of Registries - Web log transactions - Referrers access by year 91
Figure 14: Results - System of Registries - Comparison of referrers' access by year 93
Figure 15: Results - SCHEMAS workshop and Discussion Lists surveys - Metadata terms 95
Figure 16: Results - Discussion Lists and SCHEMAS workshop surveys - Most popular metadata element sets 97
Figure 17: Results - SCHEMAS workshop and Discussion Lists surveys - Metadata element sets use in domains 97
Figure 18: Results - SCHEMAS workshop and Discussion Lists surveys - Reasons for using metadata element sets 99
Figure 19: Results - Discussion Lists and SCHEMAS workshop surveys - Most popular metadata registry systems 101
Figure 20: Results - MetaForm - Web log transactions - Use by day of the week (July 1998-November 2004) 102
Figure 21: Results - MetaForm - Web log transactions - Daily summary - Yearly comparison 103
Figure 22: Results - System of Registries - Web log transactions - Use by day of the week (1998-October 2004) 104
Figure 23: Results - System of Registries - Web log transactions - Use by day of the week - Yearly comparison (1998 – October 2004) 105
Figure 24: Results - MetaForm - Web log transactions - Monthly representation of successful page requests 1998-2004 108
Figure 25: Results - System of Registries - Web log transactions - Monthly use - Successful page requests for the period of 1998-2004 110
Figure 26: Results - MetaForm - Web log transactions - Yearly use 111

Figure 27: Results - System of Registries - Web log transactions - Yearly use - Successful page requests 1998 - October 2004 111
Figure 28: Results - MetaForm - Web log transactions - Daily use 112
Figure 29: Results - MetaForm - Web log transactions - Comparison of the daily use by year 113
Figure 30: Results - System of Registries - Web log transactions - Increase/decrease in use by month (October 2001 - October 2004) 113
Figure 31: Results - System of Registries - Increase/decrease in use (October 1998-October 2004) 114
Figure 32: Results - MetaForm - Directory accesses 1998-2004 115
Figure 33: Results - EDR/SoR - Access by requests report (1998-2004) 116
Figure 34: Results - Environmental Data Registry - Questionnaire survey A - Services used 118
Figure 35: Results - Environmental Data Registry - Questionnaire Survey A - Ranking of website features 119
Figure 36: Results - System of Registries - Questionnaire survey B - Most used System of Registries applications 120
Figure 37: Results - System of Registries - Questionnaire survey B - Type of information most sought on System of Registries 121
Figure 38: Results - System of Registries - Questionnaire Survey B - Information retrieved 122
Figure 39: Results - Discussion Lists, SCHEMAS workshop and MetaForm registry surveys - Reasons for using metadata registry systems 123
Figure 40: Results - MetaForm - Questionnaire survey - Rating of features 137
Figure 41: Results - MetaForm - Desired features 139
Figure 42: Results - MetaForm - Web log transactions - Comparison of failed requests by year 144
Figure 43: Results - MetaForm - Web log transactions - Comparison of failed requests by month 145
Figure 44: Results - MetaForm - Web log transactions - Failed requests - Comparison of yearly representation 146
Figure 45: Results - System of Registries - Web log transactions - Comparison of failed requests by year 147
Figure 46: Results - System of Registries - Web log transactions - Comparison of failed requests by month 147
Figure 47: Results - System of Registries - Web log transactions - Comparison of failed requests by year 148

List of tables

Table 1: Methodology - Research methods used by registry system and date 39
Table 2: Methodology - Transaction Logs - Metrics used 46
Table 3: Methodology - Questionnaire surveys by type 53
Table 4: Methodology - Range of questions used in the questionnaire surveys and interviews 59
Table 5: Methodology - MetaForm questionnaire – List of future services 63
Table 6: Methodology - Environmental Data Registry questionnaire - List of difficulties encountered during the searching process 64
Table 7: Methodology - Environmental Data Registry questionnaire – List of services 65
Table 8: Methodology - Email Interviews – List of discussion lists and contacts 70
Table 9: Results - SCHEMAS workshop and Discussion Lists surveys - Response 75
Table 10: Results - Email Interviews - Response rates 75
Table 11: Results - Email Interviews - Non use 76
Table 12: Results - Response by method 76
Table 13: Results - Response by age groups - Category *"Other"* analysis 78
Table 14: Results - Email interviews - Response by field of interest 82
Table 15: Results - Difficulties encountered during the information seeking process 83
Table 16: Results - MetaForm questionnaire survey - Longevity of use 84
Table 17: Results - EDR questionnaire survey - Longevity of use 84
Table 18: Results - MetaForm questionnaire survey - Frequency of use 85
Table 19: Results - EDR questionnaire survey - Frequency of use 85
Table 20: Results - System of Registries - Web log transactions - Referrers access by month 92
Table 21: Results - SCHEMAS workshop and Discussion Lists surveys - Category *"Other"* metadata element sets used 98
Table 22: Results - MetaForm - Web log transactions - Daily summary - Increase 104
Table 23: Results - System of Registries - Web log transactions - Daily summary – Increase 106
Table 24: Results - MetaForm - Web log transactions - Monthly 107
Table 25: Results - System of Registries - Web log transactions - Monthly use - Thumbnails by year 109
Table 26: Results - Email interviews - Projects 130
Table 27: Results - EDR - Applications of retrieved information from the EDR 131
Table 28: Results - Email Interviews - Applications in use of metadata registry systems views 132
Table 29: Results - MetaForm - Web log transactions - Failed requests - Error codes 146
Table 30: Results - EDR - Questionnaire survey A - Definition of EDR 153
Table 31: Results - EDR - Questionnaire survey A - Use of EDR in the EPA 154
Table 32: Appendix 3b - SCHEMAS workshop and Discussion Lists surveys – Response by discussion list 202

Acknowledgements

I would like to thank my supervisors, Professor David Nicholas and Dr Tamara Eisenschitz. Professor Nicholas for providing me the opportunity to conduct this research study, gain valuable experience as a new researcher through several research projects that he was engaged in and for his guidance and valuable feedback throughout the whole period of my studies. I would like to thank Dr Eisenschitz for her patience and kindness.

During this study there have been several people that provided valuable help and support. I would like to thank Dr. Aida Slavic, Mr Michael Pendleton, Dr. Carola Wessel, Dr. Manjula Patel, Dr. Andrew McFarlane and Mr Anthony Watkinson for their help.

Several fellow students and friends made this experience memorable. I would like to thank Fabio, Maria, Cristina, Χριστίνα, Κρυσταλλία, Eddy, Andrea and all the IS and DOC students that I shared the several offices and discussions with over the years.

Also, I would like to thank Antony for his love, patience and understanding during the last few years of my PhD studies.

Last but not least I would like to thank my family for loving me, for believing in me and for their patience. In particular, I would like to dedicate this PhD to my father and to my two nieces, Εύη and Σαλώμη.

This study has been partially sponsored by the State Scholarships Foundation, Greece.

Declaration

I grant powers of discretion the University Librarian to allow this thesis to be copied in whole or in part without further reference to me. These permission covers only single copies make for single study purposes, subject to normal conditions of acknowledgement.

Abstract

Metadata registries are systems that store authoritative information about the exact meaning and use of metadata elements and metadata terms. They also record the semantic relationships between the different elements and terms. They are a vital part in the process of data standardisation and metadata management and consequently in the organisation of digital information. This study investigates the use and functionality of two active metadata registry systems, as current users of those systems perceive them. A combination of methods such as questionnaire surveys, web log transactions analysis and email interviews was employed to assess them. Users of two metadata registry systems, the MetaForm registry and the Environmental Data Registry (EDR) – currently known as System of Registries (SoR) were surveyed about their satisfaction towards services and products that the registries provided, the efficiency of the support services within the metadata registries and they also recorded feedback about their future expectations of such systems. Results indicated that metadata registry systems are important components in the process of data standardisation and data management. The development of software to support mappings between the elements of different metadata element sets is expected to trigger future research in the evaluation of metadata registry systems. User needs assessment at the initial stages of the set up of metadata registry systems helps to build systems closer to the user needs. Understanding about the context of use of the retrieved information helps to improve the provision of metadata registry system services.

Chapter 1 – Introduction

A fundamental characteristic of the "*information revolution*" is the availability of information to anyone with a computer, a network connection and some basic computer literacy skills. The downside of such availability is information overload, extraneous information and the ability to find information or interest. It was soon recognised that in order to be able to locate, access and retrieve information of interest from the exponentially growing information pile that is the World Wide Web it would have to be organised and managed.

Libraries are the institutions that have been traditionally associated with the organisation of information. Library practices such as cataloguing, indexing, abstracting, and classifying have been applied to large data sets for years. These practices are based upon agreed rules and codes that refer to the syntax, semantics and the structural form of the resource described. In time, they have grown to sophisticated standards and the library community has succeeded in dealing with the problems associated with preservation, maintenance, and exchange of information. Dempsey (1989) noted that libraries have a longer tradition in the production and exchange of information in electronic form than any other organisation in the bibliographic information chain. But of course nothing has been tackled on the scale of the web and of course they have played only a minor role in the web's actual development. Nevertheless Baker (1996) acknowledges *that "the Web must become more like a well-organised library*". And Lynch (1997) also quotes that "*In short, the Net is not a digital library. But if it is to continue to grow and thrive as a new means of communication, something very much like traditional library services will be needed to organise, access and preserve networked information*".

The application and use of metadata standards has been suggested as a means for the description and efficient retrieval of information. "*The effective management of networked digital information...will increasingly rely on the effective development and use of systems that can collect and use appropriate metadata*" (Day, 1999). Since the first Dublin Core Workshop that initiated the "*metadata movement*" (Baker 1999) in 1995, rapid advances have been made. Implementations carry on side by side with research. Metadata appears to provide an effective answer to discovering and retrieving networked information (Dempsey, 1996, 2000; Dillon, 2000).

The idea for this research was initiated from the importance that metadata and standards in general could play in the field of data standardisation and data management in a digital

environment. Research in the field of digital libraries addressed several issues associated with metadata. The emergence of metadata registry systems as authoritative sources for the management of metadata posed as the opportunity to conduct user evaluations about the use and the applications of metadata. An examination of the literature indicates that the assessment of use and application of metadata is vital for its implementation. Re use of metadata spurred the development of metadata registry systems. Metadata registry systems are defined as registration authorities associated with the description, discovery, storage and exchange of metadata, and as such they address data sharing and standardisation problems often associated with networked resources. Data about the use and functionality of such services is important for the implementation and deployment of metadata applications and the support of data exchange and management. In particular, studies about how such systems are used by their own users would provide helpful information to system developers and system implementers in order to improve current services.

1.1 Aims and objectives

This study aims to identify the role of metadata registry systems in the area of data standardisation and management, in particular with relevance to digital information. It investigates how metadata registry systems are used to organise, describe and retrieve digital information. In particular, it aims to detect currently active metadata registry systems, define their functions and assess whether they meet their aims of use. Also, to evaluate their functionality as this is perceived by actual users of the metadata registry systems. This data will be valuable to determine whether their use conforms to previous findings in the literature and whether this complements or differs from what has been discovered. The aims of this study are:

- To identify how users of different domains perceived metadata and their use.
- To find out about the intentional use of metadata and metadata registry systems.
- To specify whether there will be an increase in the use of metadata registry systems in the future.
- To evaluate the role of metadata registry systems in data standardisation and management by assessing their use and functionality.
- To evaluate the primary services/features of metadata registry systems in order to find out about users expectations of such system and their prescribed role in the fields of data management and data standardisation.
- To specify factors that they might prevent them from using them.

In particular, the objectives were:

- To draw of an indicative profile of metadata registry systems users by gathering information about their age, sex, occupation, field of interest and occupation. Additionally, to establish use by specific domains and referrers as this had been recorded in the web log transaction reports. And to find out about the metadata registry systems users' familiarity with information resources use in general.
- To explore the users' understanding of metadata in general and the use of metadata registry systems in particular.
- To establish frequency and duration of metadata registry systems use.
- To investigate use by day, month and year and question any particular patterns of use deriving from this.
- To find out about the particular directories and/or types of information that users are interested in metadata registry systems.
- To investigate how an organisation's requirement for a metadata registry system could affect the system's use.
- To present current implementations of metadata registry systems.
- To assess how satisfactory the design and the content of the specific metadata registry systems studied had been.
- To assess metadata registry systems' functionality by asking users to indicate the services they had used and deem how satisfactory they were.
- To discover if there is a requirement for metadata registry systems' use by different domains and if so to see what are the primary features/services that users are interested in.
- To identify the most popular metadata element sets in terms of publication and use and to investigate the reasons why these element sets were used.
- To identify the most popular metadata registries in terms of publication and use across different domains and find out the reasons why they were used.

1.2 Definitions

Throughout this thesis several terms related to metadata and digital libraries research are repeated. In order to avoid confusion and repetition of explanation, the most important terms are defined in the following section:

- ***Digital Library.*** *"Digital libraries are organizations that provide the resources, including the specialized staff, to select, structure, offer intellectual access to, interpret, distribute, preserve the integrity of, and ensure the persistence over time of collections of digital works so that they are readily and economically available for use by a defined community or set of communities".* Digital Library Federation. (1999). Available at: http://www.clir.org/diglib/dldefinition.htm.

- **Metadata**. "Structured data about data" (Weibel, 1997)
- **Metadata format**. *"The format is a set of rules that govern the data structure and content"* (Hopkinson, 1998)
- **Metadata schema or scheme**. *"A structured set of attributes with associated semantics and name elements"* (UKOLN, a combined definition from RLSP project and SCHEMAS glossary. Available at: http://www.uklon.ac.uk).
- **Metadata registry system**. *"A system to provide management of metadata elements. Metadata registries are formal systems that provide authoritative information about the semantics and structure of data elements. Each element will include the definition of the element, the qualifiers associated with it, mappings to multilingual versions and elements in other schema".* Dublin Core Metadata Glossary. Available at http://dublincore.org/documents/usageguide/glossary.shtml
- **Interoperability.** *"The ability of different types of computers, networks, operating systems, and applications to work together effectively, without prior communication, in order to exchange information in a useful and meaningful manner. There are three aspects of interoperability: semantic, structural and syntactical."* Dublin Core Metadata Glossary. Available at http://dublincore.org/documents/usageguide/glossary.shtml
- **Usability.** "*The extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use*". ISO 9241-11: Guidance on Usability (1998)
- **Functionality.** "*In information technology or computing circles, functionality is a term that describes the capacity of a system (software application or computing device) to perform the functions required by the user. It can be used to identify the features, capabilities or behaviours of a system – otherwise known as the things it can do*". StORe project definitions. Available at: http://jiscstore.jot.com/WikiHome/OrganisationAndManagement/Definitions.doc/!converted/index.html

1.3 Scope

Research in the wider context of digital libraries addressed issues such as digitisation, preservation, maintenance and access to information. Some of the challenges in creating and using digital libraries include technical issues such as high speed networks, fast connections to the Internet, databases that support a variety of digital formats and aids to the management of digital resources (Cleveland, 1998). Other issues included collections' content, user communities and their needs, skills that staff should have, and copyright/rights management.

Metadata is part of this wider context of digital libraries' use. It has been reported in the literature (please see Chapter 2 – Literature Review) that the application and use of metadata could improve the discovery and exchange of information and therefore promote education and knowledge. Metadata registries as mentioned earlier are systems that store authoritative information about the structure and the semantics of the metadata elements and therefore are believed to be a vital part in the process of data standardisation and metadata management and consequently in the organisation of digital information.

The intention of this study is not to attend all of these issues, as the field is large. The concentration is on a user centred approach of the evaluation of metadata registry systems. In particular the study addresses the following:

- An understanding of how actual users of metadata registry systems assess their usability and functionality.
- Identification of metadata systems users' characteristics, denote services that registries provide and rank their usefulness based on "the *extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use", and*
- To obtain users' feedback in order to find out how to establish better services in data standardisation and management.

1.3.1 Users of metadata registry systems

For the purpose of this study, it was aimed that the target group would consist of current users of two active metadata registry systems, the MetaForm registry and the Environmental Data Registry/ System of Registries as well as researchers in the area of metadata in general and metadata registry systems in particular. At the beginning of this study the following metadata registry systems were identified:

- The Australian Institute of Health and Welfare (AIHW) registry system
- The CORES registry system
- The DESIRE registry system
- The Environmental Data Registry (EDR) and
- The MetaForm registry system

Relevant members of staff at all the metadata registry systems were contacted and enquired about the prospect of conducting user surveys with the systems' users. The AIHW registry system refused the invitation on the grounds of confidentiality issues that their content abided to. The DESIRE and the CORES projects were funded projects that

were developed for future use and implementations were not available at the time that this study commenced. The EDR and the MetaForm registry systems were those that agreed to the research invitation and therefore used in this study. As this has been a first of its kind assessment of metadata registry systems as their users valued them, it was not possible to identify similar user studies in the area of metadata registry systems in order to compare the findings. Furthermore, at the beginning of this study, research in the area of metadata in general, was in its formative stages. Metadata registry system implementations were few around the world but innovative in concept in what they aimed to achieve. The majority of user studies in the broader area of digital libraries use tend to be in the academic domain (Monopoli, 2005) and it has been argued that academic services users form the largest group of digital library users and therefore their feedback could be indicative for other user groups as well. The users of metadata registry systems though were not limited in any one scientific domain; metadata registry system applications had been found in many domains such as the governmental and the academic. This fact notwithstanding, it is believed that those individuals surveyed are representative of the user base of metadata registry system.

1.3.2 Metadata registry systems and other cases studied

This following section presents the two metadata registry systems that were surveyed and the SCHEMAS project that was used as a third case study at it dealt with issues relevant to metadata registry systems' research and implementations and its members were relevant to this research study.

1.3.2.1 SCHEMAS project, UK

The "SCHEMAS – A Forum for Metadata Schema Implementors" project was a European Union funded project that run between January 2000 and December 2001. It was a collaboration between the UK Office for Library Networking, the PricewaterhouseCoopers Consultants and the German National Research Centre for Information Technology. Its objectives were to "*provide information for schema implementors about the status and proper use of new and emerging metadata standards, as well as promoting good-practice guidelines for adapting multiple standards or metadata modules for local use in customised schemas*"[1]. The target audience of this project were individuals and institutions that were interested in the implementation and use of metadata standards. The project provided with three deliverables listed below:

- A series of workshops that aimed to explore issues about the management of currently implemented metadata schemas and encouraged the participation of users

[1] SCHEMAS project. Available at: http://www.schemas-forum.org/ (Last accessed 12/06/2006)

from all scientific domains. It was expected that by the completion of the workshops the organisers and the workshop attendees would have addressed research issues such as the registration of diverse metadata element sets, suggestions for the structure of a metadata registry system and what were the best practices for the use of metadata.

- Four metadata watch reports reviewing the use of several metadata schemas as recognised standards in the audiovisual, cultural heritage, educational, publishing, and the Governmental domains. It was expected that by the completion of the metadata watch reports the SCHEMAS researchers would have drawn the landscape of metadata research in the domains of academia, audio visual, cultural heritage, education, geographic information, industry, publishing and rights management, research and other.
- The design of a metadata registry system. The SCHEMAS project used the expertise of the UKOLN staff and built the SCHEMAS registry based on the framework of the previously successful DESIRE registry and ROADS templates. It was expected that by the completion of the project the SCHEMAS researcher would be able to provide with a registration authority that would hold the elements and the definitions of metadata element sets. Furthermore, as the SCHEMAS registry would be based on the previously successful prototype of the UKOLN hosted DESIRE registry it would serve as a follow up on the use of those systems.

The opportunity to collaborate with the SCHEMAS project arose in the light of the second SCHEMAS workshop in November 2000. The theme of the workshop was "Publishing and sharing your application profile" and its aim was "*to present the state of the art in constructing and publishing an application profile and how it may be declared in XML/RDF, especially in light of new metadata harvesters that support the indexing and browsing of standards and application profiles located on multiple Web servers*"[2] The intended audience of the workshop were people involved or with an interest in the implementation of metadata standards. The number of attendees was limited to 50 people as this was the number of people that could be accommodated by the venue's size. The project officer was contacted regarding the prospect of having a questionnaire survey run at the workshop. After exchanging a few emails explaining the purposes of the survey and its intentional use the workshop's organisers granted their permission to conduct the survey. After the permission was granted, the questionnaire was drawn. It was agreed that the questionnaire would be included in the workshop delegates' packs and that they would be advised by the organisers to complete it and return it to the

[2] SCHEMAS Second Workshop: Publishing and sharing your metadata application profile. Information available at the following URL: http://www.schemas-forum.org/workshops/ (Last accessed, 12/06/2006)

registration point of the workshop before they leave the venue. The questionnaire was preceded by a letter providing the contact details of the researcher, explaining the purpose of this survey and guarantying confidentiality of any given response (please see also Chapter 3 – Methodology).

1.3.2.2 MetaForm registry, Germany

The MetaForm registry (http://www2.sub.uni-goettingen.de/metaform/index.html) and the MetaGuide (http://www2.sub.uni-goettingen.de/metaguide/index.html) form the Metadata Server at SUB Göttingen, Germany. They were both outcomes of the Meta-Lib project, an initiative of the *National Library of Germany (*Die Deutsche Bibliothek) and the State University of *Göttingen (SUB Göttingen)*, which was funded by the German Research Foundation and aimed to address issues and challenges that networked resources, their storage, preservation and retrieval posed upon the German library community. The Meta-Lib project run from 1998 to 2003. The objectives of the Meta-Lib project and therefore the MetaForm and MetaGuide were:

- To monitor the international metadata development (particularly Dublin Core),
- To analyse the conception and application of various metadata formats,
- To specify the basic requirements for the description of all types of digital objects and
- To exchange knowledge and experience (Schimmer, 1997)

MetaForm was developed in order to act as an authoritative source of information that would provide guidelines on the use of metadata. It is a "*...database for metadata formats with a special emphasis on the Dublin Core and its manifestations as they are expressed in various implementations. As a project deliverable, the idea behind this database is to identify the core elements that are used for the description of networked resources*" (MetaForm, http://www2.sub.uni-goettingen.de/metaform/index.html).

MetaForm comprised of the following three services:

- Crosswalks. This section of MetaForm stores the Dublin Core metadata element set and other application profiles that they were formed by using the Dublin Core as their base. It also provides the relative documentation for each element set or application profile such as a description of the element and how can be used in context. In some cases examples of use in practice are also provided. Furthermore, crosswalks list associated DTD schemas, and assorted organisations that they have implemented the relevant application profiles or element sets.

- Crosscuts. The second of the three MetaForm services describes how Dublin Core elements are used in various implementations.[3] Crosscuts are presented in the form of a table with two columns. In the first column there are listed the variations that an element of a metadata element set or application profile can take and in the second column are listed the descriptions of this element.
- Mappings. The last of the three MetaForm services lists mappings of "*DC applications and other formats with each other*," [4] which are dynamically generated between two element sets at a time. A typical representation of a mapping includes data presented in a table consisting of four columns: (i) the first column cover the elements of the first set (ii) the second column presents the description of each element, the third column covers the equivalent (or otherwise blank) element of the second set, and finally the fourth column describes the elements of the second set.

The opportunity to collaborate with the MetaForm registry arose after personal communication between the researcher and the principle investigator of the MetaLib project. The project investigator was contacted regarding the prospect of having a questionnaire survey run on the MetaForm registry's website. After exchanging a few emails explaining the purposes of the survey and its intentional use the MetaForm registry staff gave their permission to conduct the survey. After the permission was granted, the questionnaire was drawn. It was agreed that the questionnaire would be made available at the MetaForm registry's website under the title MetaQuest and the responses would be delivered directly to the researcher's mailbox. The questionnaire was linked to an explanatory note providing the contact details of the researcher, explaining the purpose of this survey and guarantying confidentiality of any given response (please see also Chapter 3 – Methodology).

1.3.2.3 The Environmental Data Registry (EDR)/System of Registries, USA

The Environmental Protection Agency (EPA, http://www.epa.gov/sor) defines the Environmental Data Registry (EDR) as "*...a comprehensive, authoritative source of reference information about the definition, source, and uses of environmental data. The EDR catalogs the Environmental Protection Agency's (EPA) major data collections and helps locate environmental information of interest*".

[3] About the server. MetaForm, http://www2.sub.uni-goettingen.de/metaform/index.html (Last accessed: 12/06/2006)

The development of the EDR began in 1993, and made significant strides under the auspices of the EPA's Reinventing Environmental Information (REI) Initiative that began in the late 1990's. The EDR is EPA's primary resource for metadata pertaining to data within the Agency's major information systems. The EDR also serves as a clearinghouse for EPA's data standards. Between 1993 and 2002 the EDR was the main registry system of the EPA but it was closely allied with several other registry systems including the Terminology Reference System, the Substance Registry System, the Chemical Registry System and the Biology Registry System (Figure 2). Those systems provide search tools for retrieving information about how environmental terminology and physical, chemical and biological substances are represented in the EPA's regulations and data systems. Analytically the following registries were hosted by EDR:

- Substance Registry Systems. The Substance Registry System (SRS), the Chemical Registry System (CRS), and the Biology Registry System (BioRS) that were the gateways to information on chemicals, biological, and other substances. They also held information about how they were represented in EPA regulations and information systems.
- Terminology Registry System (TRS). The TRS was a compilation of environmental terms, definitions and information sources relevant to the mission of EPA and other environmentally oriented entities.
- MetaPro. MetaPro was a tool that could be used to create a metadata registry (EPA Environmental Data Registry Users Conference Agenda, 2002).

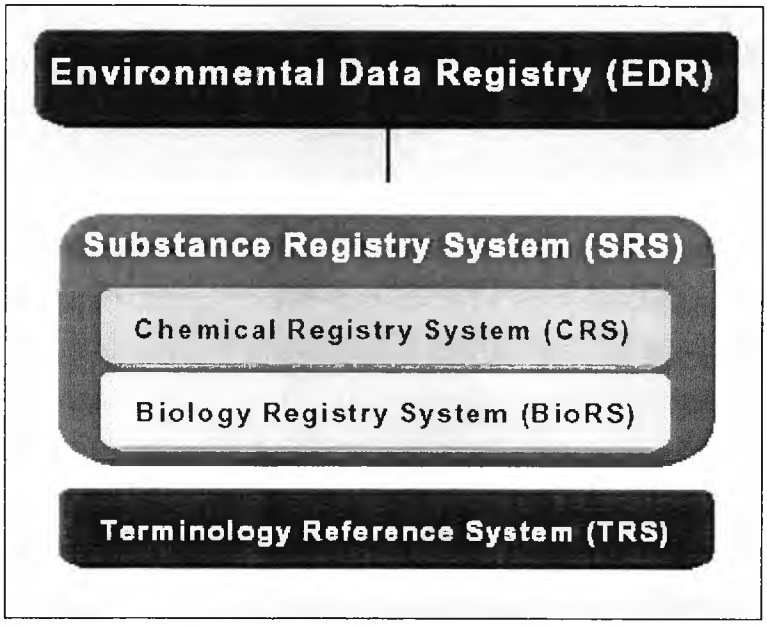


Figure 1: Introduction - Graphical representation of the EPA's metadata registries (*Source: Standard Update: Better Data through standards, Vol. 2, No. 3, Winter 2000)*

Since its inception, the EDR has continually evolved in order to serve the needs of its users. The users' conference held in January 2002 contributed to continued improvements, and additional functionality has been added as a direct result of input provided by conference participants. As described in the winter 2002 EDR newsletter *"Users recommended improvements to the application interface, data content, software functionality, and the process for how data standards are developed and implemented."* The EPA implemented those suggestions and an improved version of the EDR based on the suggestions of its users was launched in spring 2002. The newly introduced service was named as the System of Registries (SoR) and some of the changes that were implemented included enhanced search performance for the Chemical Substance Registry (CRS); site redesign to improve access to downloadable code sets; improvements to the Compare Tool (a tool designed to support data harmonization); and inclusion of XML tags associated with standard data elements.

The system of Registries (SoR) is defined as a web-based collection of metadata registries and repositories residing in the EPA's Office of Environmental Information (OEI). The registries that comprise the SoR provide identification information for objects of interest to EPA trading partners, including states and tribal entities, and the public. They described objects consist of data elements, XML tags, data standards, substances (chemicals, biological organisms, and physical properties), terms and definitions, facilities, regulations, and data sets that the Agency uses in its core business processes. These registries comprise a critical link in EPA's information architecture and are vital components to the Exchange Network being developed to facilitate data exchange with stakeholders through network nodes. In fall 2002, the System or Registries consisted of the following six registries:

- Environmental Data Registry (EDR). The EDR provided access to EPA data standards, XLM tags and application metadata such as data elements.
- Facility Registry System (FRS). The FRS was a centrally managed database that identified facilities, sites, or places subject to environmental regulations or of environmental interest.
- Information Resource Registry System (IRRS). The IRRS provided information about the EPA's application inventory, organisation hierarchy, and other information resources (adopted from the fall 2002 Newsletter).
- Substance Registry System (SRS). The SRS acted as the gateway to information about chemical, biological and miscellaneous organisms as they are represented in the EPA's regulations, data systems, and other information resources.
- Environmental Information Management System (EIMS). The EIMS was a single source of information to spatial data. It hosted information such as metadata on

remote sensing data, Geographical Information System (GIS) coverages, databases, documents, models, and multimedia.

- Terminology Reference System (TRS). The TRS provided access to environmental terms and definitions that were compiled from a wide range of EPA program offices, information systems, and state collections. This registry could be searched by keyword, information resource, and organisation.

The SoR continued to evolve over the last years and some of the significant changes of the recent years included the development and implementation of two additional registries. Those are described hereafter:

- Web Registry (WR). WR is described as a centrally managed database used to collect metadata for EPA's priority web materials. This registry enables the combination of metadata with the full text of corresponding pages and documents, in order to produce more relevant search results.
- XML Registry. The XML registry holds information such as XML Data Exchange Template (DETs), XML Schemas, Namespaces, WSDL files, and other supporting files needed to map data flows between partners.

The maturity of the EDR in relation to other existing metadata registry systems made it a suitable candidate for evaluating how people use these systems; it also allowed to explore how such systems might be enhanced to better serve their customers.

The opportunity to survey the EDR users arose in the light of the 1st users' conference, which was held at the U.S. Bureau of Labor Statistics' Conference Center on January 24, 2002 in Washington, DC and then again at the follow up users' conference the next year. Following personal communication between the researcher and the organisers of the conference it was agreed to conduct a questionnaire survey in the context of the conference. The questionnaire was distributed to the conference delegates during the hands-on session of the conference and they would be advised by the organisers to complete it and return it to the registration point of the conference before they leave the venue. A letter providing the contact details of the researcher, explaining the purpose of this survey and guarantying confidentiality of the respondents, preceded the questionnaire. Furthermore, an additional note that the EPA held no responsibility for the content of the questionnaire and the participation to the survey was at the conference attendees' own will, was also included at the request of the organisers (please see also Chapter 3 – Methodology).

1.4 Outline of the thesis

This thesis is divided into five chapters and includes appendices. The current chapter gives an introduction to the research study, provides the context and scope of the research and presents the metadata registry systems and other projects that were studied. Chapter 2 provides an overview of the literature; gives some background on the evolution of the metadata registry systems and discusses their use and functionality as this has been recorded in the literature. Chapter 3 discusses the various methods available to conduct user oriented research and presents the methods that apply to Internet based studies, such as those conducted for this study. Furthermore it describes in detail the questions included in all questionnaire surveys by study, the reports selected for the web log analysis and the type of calculations that needed to be conducted to process the web log data as well as the questions that were used in the email interviews. Chapter 4 discusses the results of the questionnaire surveys, the web log analysis and the email interviews conducted with users of two active metadata registry systems and metadata researchers. Chapter 5 discusses the results, lists the conclusions of the study and recommends future work in the area of metadata registry systems research.

Chapter 2 - Literature review

This chapter is divided in two sections. The first section provides a brief overview of the role of catalogues and standards in the organisation of information particularly as this is recorded in the library and information science. It discusses the concept of metadata, how it evolved and its role in the organisation of networked information. Furthermore, it provides an overview of the research about metadata and metadata applications and inscribes the impact that metadata applications had on the description, search, access, retrieval and exchange of networked information. The second part shows how the need for metadata registry systems is documented in the literature and provides an overview of metadata registry systems developments around the world. Finally, it discusses metadata registry systems' use and functionality from a user's point of view as this is reported in the literature particularly in studies of information use.

2.1 Organisation of information

Knowing facts about organising information date back to the Library of Alexandria. Callimachous of Cyrene (c.305-c.240B.C), one of the first librarians at the Library of Alexandria, compiled what is considered the first form of a library catalogue, the *Pinakes ton en pase paideia dialampsanton kai hon synegrapsan*[5]. The Pinakes formed 120 scrolls listing knowledge under subject areas. Each subject area listed in alphabetical order the works of a given author. Pinakes, although it had not been preserved, provided access to information about the intellectual property of an organisation.

Although the evolution of the catalogue in Western Europe as a means to finding and managing information can be traced in the ecclesiastic environment (Haynes (2004), the focus during the 19th and 20th century was placed on the development of cataloguing codes, the rules for describing, accessing and guiding cataloguing practice (Fatahhi, 1997). This aspect is very important in data standardisation and data management as it provides the context for data and metadata use. The development of the modern cataloguing codes is attributed to Sir Anthony Panizzi, a librarian at the British Museum, who compiled the *Ninety-One* rules for the printed book catalogue at the British Museum. The significance of the *Ninety-One* rules is attributed to the fact that for the first time it was attempted to compile guidelines for author catalogues, introducing the notion of a heading accompanied by guidelines for cross references. Influenced by Sir Anthony Panizzi, Charles Jewett and Charles Cutter modified the rules to make them applicable to American libraries and laid the ground for an Anglo- American collaboration that

[5] Books and writers. Available at : http://www.kirjasto.sci.fi/callimac.htm (Last accessed: 12/06/2006)

produced the Anglo-American Code, published in 1908. C. Jewett brought up the importance of uniformity in the use of cataloguing principles in particular if data was going to be accessed by other institutions. That was important as it raised concern about standardisation and how this could be addressed by cataloguing codes. The Lubetzky's code, presented at the International Conference on Cataloguing Principles (ICCP) in Paris during 1961, was considered one of the most significant events in the history of descriptive cataloguing. It emphasised the principles of main and added entries in a catalogue and the function of collocating information. It pointed to the fact that now it was possible to gather all the works of an author or corporate body together irrespectively of edition, translation, publication type together (Gorman, 2000; Fatahhi, 1997). Summarising, the cataloguing rules provided the context for the use of the described information. They facilitated the semantic representation of the information and therefore give meaning to the information they describe.

In parallel to the cataloguing codes' advances, changes to the physical form of catalogues, moving from printed books to the card catalogue and the development of computer systems in the 1960s and 1970s influenced the automation and production of bibliographic data. The Machine Readable Cataloguing (MARC) standard that was established in 1969 provided the framework for the production of automated catalogue records. Although its initial use was for the production and storage of bibliographic information in magnetic tapes, it got increasingly used for the exchange of information among different systems as it conformed to the ISO 2709: Information and documentation – Format for Information Exchange. Furthermore, the establishment of the International Standard Bibliographic Descriptions (ISBDs) in the early 1970s and their incorporation in the Anglo American Cataloguing Rules enhanced the framework for a standardised approach in the description and organisation of information. The use of the MARC standard facilitated the reduction of cataloguing effort, ensured a standardised approached to the exchanged of bibliographic records and minimised duplication of effort in the production of bibliographic records. Moreover, the output of that standardisation effort was accessible to users via online catalogues. The combination of the availability of standards, technology and production of information contributed to the development of large databases of published material that now required new skills and "*contributed to the development of information science as a discipline*" (Hayes, 2004). In summary, the codes, standards and the formats that were the carriers of information facilitated the framework for the transferring and exchange of information. That standardised approach to using and managing information made its storing in different formats and across different systems and therefore making interoperable systems feasible.

The use of computers has affected the way information was produced and managed not only in the library and information community but also the publishing domain, academic institutions, research and governmental organisations. In 1989 Tim Berners - Lee wrote a proposal regarding managing information in order to address the problem of information loss and information access in the Conseil Européen pour la Recherche Nucléaire (CERN). That proposal led to the development of a program that grew to become the World Wide Web. The incentive behind the development of such a system was to facilitate the effective communication between a web of people in an organisation that used different information systems, different tools and different software to communicate. The aim was to avoid misunderstandings, duplication of effort deriving from lack of access to information and provide access to information that until that time was getting lost (Berners - Lee, 1989; W3C, 2000).

CERN endorsed the proposal and by 1992 the World Wide Web is released and the Internet society is chartered. By 1995 the National Science Foundation, the backbone network for academic institutions in the USA since 1985, announced that they would no longer allow direct access to the NSF network and they contracted companies, individual service providers (ISPs) to control the traffic to the network. At the same time the number of people connected to the Internet and the number of information made available had increased dramatically. *"What had been doubling each year now doubles in three months"* (Computer History Museum, http://www.computerhistory.org/). The number of available hosts rose from 1,313,000 in 1993 to 9,472,000 in January 1996. The most resent survey by the Internet Systems Consortium (http://www.isc.org/index.pl?/ops/ds/) listed 394,991,609 hosts in January 2006.

It was prominent that the vast amount of information now available to anyone in the world from any place providing access to a computer and a connection to the Internet needed to be organised. It was essential that in order to be able to find useful information it needed to be organised in an effective manner. In one of the first conferences about metadata, organised by the Institute of Electrical and Electronics Engineers (IEEE), Smith and colleagues (1996) noted *"...the most important issue facing researchers in the area of digital libraries (DLs) is to discover mechanisms that support efficient access to appropriate information"*. One of these mechanisms to support efficient access to information appeared to be metadata.

2.2 Metadata

The term metadata meant different things to different scientific communities. In the literature, the term metadata originates in computing science. Caplan (2003) noted that the term METADATA had been invented by Jack Myers who later on registered the term as trademark in 1986 for a company providing metadata solutions and support. Haynes (2004) also reported that metadata dates as back at the 1960s but it *"became established in the context of Database Management Systems in the 1970s"*. Heery, Powell and Day (1997) best describe the different forms that metadata would assume for different scientific communities in the following quote:

> *"the unit being described would be a data element in computer science, and a resource in the information world. In the information world metadata may consist of an agreed set of data elements with agreed semantics, agreed syntax and agreed rules for formulating the content of the elements".*

NISO (2004) defines metadata by its functions. "Metadata is structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource". Hunter (2002) also argues that as metadata means different things to different people, its effectiveness is based on several criteria including the identification of best metadata models, the selection of the level of granularity to satisfy specific user needs, the balance of costs and ensuring the management and long term maintenance and quality control in its use. Traditionally, libraries had been the organisations that were associated with the organisation of information. In a library context, metadata had been primarily associated with the use of standards such as the International Standards for Bibliographic Description (ISBDs), the Anglo American Cataloguing Rules (AACR) and the MARC format for the description and exchange of information. With the rise of computing in the 1990s they came to the realisation that practices applied in the traditional library world could be similarly applied in the electronic environment. In the library and information science field metadata was primarily associated with the description of information.

2.3 Overview of metadata research and implementation

Over the last decade, metadata attracted the attention of a wide range of research communities at both domestic and international level. To this day, it still remains a subject of widespread interest. Studies about how metadata can be used to describe networked information resources and facilitate efficient information retrieval have been the focus of many Research and Development projects and conferences. A simple search on the EU CORDIS database in June 2001 revealed 47 research projects alone on metadata

associated with the description, organisation and retrieval of information. The same search performed in November 2002 and in 2005 revealed more than twice that number (97 and 128 projects respectively). Professional organisations such as the Institute of Electronic and Electrical Engineers (IEEE) and the Association for Computing Machinery (ACM) were among the first to organise conferences and workshops to discuss research around metadata and its applications. The International Federation of Library Associations and Institutions (IFLA) maintains an extensive list to metadata related issues, among them conferences and workshops that have taken place since 1996, including those organised by the IEEE and ACM on Digital Libraries.

Research in metadata is interlinked with the research and implementations of digital libraries. The European Research Consortium for Informatics and Mathematics (ERCIM) organised a series of workshops and conferences about metadata dating as early as 1996. A collaboration with the National Science Foundation (NSF) addressed areas of research relevant to metadata and digital libraries. Those addressed issues such as the following:

- Metadata and interoperability in Digital Library related fields,
- Emerging technologies in the Digital Libraries domain,
- Metadata for Web databases,
- The Dublin Core metadata workshop and conference series,
- Metadata for resource description,
- Metadata for information retrieval,
- Metadata for intellectual rights and property,
- Metadata for interoperability.

The association of metadata with resources appears to provide an effective answer to discover and retrieve networked information (Dempsey, 1996, 2000; Dillon, 2000). Woodward (1996) and Vellucci (1998) have conducted detailed literature reviews on the evolution and use of metadata formats; Milstead and Feldman (1999) bolstered those studies by over viewing emerging metadata projects and standards. Depmsey and Heery (1996), within the requirements of the DESIRE project to create a generic format for use by Internet Based Subject Gateways, have produced a comparative description of several metadata formats. They defined a typology of metadata initiatives using as criterion the level of complexity that is involved in the use of a particular metadata element set. Caplan (2000) further discussed the challenges for metadata schema implementers with reference to some metadata element sets. Dempsey and Heery (1998) identified areas of metadata research with applications to html pages, internet emails, subject-based information gateways, archives and records management, statistical data sets, geographic information systems, metadata registries. They further

discussed the potential user interest in association with metadata in general and that focused in information description, information discovery, associated intellectual property rights and conditions of use, technical information about the storage and maintenance of the metadata and administrative information about the time and author of its creation. The all had concluded in the important role that metadata was meant to play in the description, organisation and discovery of information.

Metadata since 1995 has been primarily associated with the Dublin Core metadata initiative. This is discussed in the following section.

The Dublin Core Metadata Initiative (DCMI)

The "metadata movement" originates in the Dublin Core workshop series. The first Dublin Core workshop took place in Dublin, Ohio in 1995. It was attended by researchers and professionals from a wide range of scientific disciplines such as computing, information science, publishing, and the commercial domain. This first workshop aimed to address the need of finding a mechanism to organise and be able to retrieve web based resources "*without having to undergo the extensive training required to create records conforming to established standards*"6. Proposing elements that would describe accurately and in depth the resource as well as fulfilling all possible retrieval options proved problematic. Instead, the proposal of a core set of elements that would form the base for any minimum level of description was formed.

The idea of describing web based resources using a minimum set of elements led to the creation of the metadata element sets. The aim was to produce a core set of elements that is applicable to all disciplines and all languages and it is easily accessible and usable by everyone. That led to the development of Dublin Core Metadata Element Set (http://dublincore.org) in 1995. The Dublin Core metadata initiative defined a core set of elements that can be used as the carriers of the information that is associated with a resource. This information is then embedded in the resource so that is accessible to the systems and users that search for this resource. "*Using metadata to record information sources allows an initial assessment of compatibility and provides an avenue for merging information or for exchanging information between systems*" (Haynes 2004)

The metadata research was representative of the scientific communities that were involved in the organisation of networked information. Although everybody was interested in the development of standards for managing sharable data some were interested in the

---

6 Weibel, S., et al. (1995). OCLC/NCSA Metadata Workshop Report. Available at: http://dublincore.org/workshops/dc1/report.shtml (last accessed on the 12/06/2006)

structural side of metadata that defined the computing infrastructure and others in the semantic side of metadata that defined the content and facilitated consistency and common understanding.

Functional use of metadata

The purposes of metadata include a range of functions such as resource description, resource discovery, administration and management of resources, record of intellectual property rights, documenting software and hardware environments, preservation management of digital resources and providing information on context and authenticity (Large et al, 1999; Day, 2001). Eden (2002) added to the above the role of metadata in bringing similar items together, allow resources to be found based on relevant criteria and it also gives the location information. Burnett and Park (1999) argued that depending on the community that discusses the metadata use, the emphasis is placed upon different functions. Caplan (2003) identified three broad categories of metadata based on their functional use. Those were descriptive, administrative or structural.

Descriptive metadata is usually associated with resource discovery. It facilitates the discovery, identification, selection of resources and makes it possible to evaluate, link and assess usability of services (Caplan, 2003). Haynes (2004) argued that metadata "*can improve retrieval by establishing a context for individual descriptors*". The definition of elements and the rules/guiding instructions that accompany their use can prove important tools for information system managers as they can help to improve indexing mechanisms and improve retrieval of information.

Administrative metadata is associated with the management of resources. This type of metadata provides information on when the object's record was created, who (in the form of an individual or institution) created it, when it was modified, etc. Many content management systems use metadata to track any changes in the history of an online resource. Also, metadata provide an authenticity factor as they enable us to know the creator, the host organisation, the date and time that any modification occurred to a given resource.

Structural metadata is associated with the relations of objects. It is used to provide an indication of how compound objects are put in order. Example of such structural representation is the order of chapters to form a book (NISO, 2004). Lagoze (1996), defines structural metadata as "*...the data defining the logical components of complex or compound objects and how to access those components*". The structural metadata define the whole metadata framework particularly in an information system. Often structural

metadata is known as metadata schemes and consist of sets of metadata elements. The definitions of the meaning for each of the elements are commonly known as the semantics of the metadata schema. The values that are assigned to each element are the content of the schema and some times it can include content rules. In some metadata registry systems those are known as business rules. The business rules provide information about the context in which elements can be used. Furthermore, the schemes may include syntax as well that defines how elements are ordered and their relations with other elements and associated objects. Summarising, structural metadata is very useful for information retrieval purposes and it is usually maintained and used by the particular system in which structural metadata is defined for use.

2.4 Metadata registry systems

The ISO/IEC 11179 Metadata Registry Implementation Coalition defined metadata registry systems as "*Data semantics management systems, or data element concept metadata registries as they are called – the term can be used interchangeably – are automated databases that contain all the information that defines the exact meaning of the individual terms and metadata elements, including the semantic relationships between different data elements and different terms*".

During the 1970s and 1980s, the evolution of computer networks emphasised the need for interoperability across heterogeneous and distributed systems. The explosion of the World Wide Web highlighted the same need for interoperable systems but this time at the level of the semantic representation of the information exchanged (Cordeiro & Slavic, 2002). Blanchi and Petrone (2001) emphasised that the main problem of interacting across digital library applications did not lie in any underlying network infrastructure but in the ability to be able to characterise the information they contain in a consistent manner. Although different levels of interoperability such as technical, semantic, political/human, inter-community and international raised different issues for consideration (Johnston & Miller, 2001; Johnston, 2000), the question, as set by Baker (1996), was "*...how to integrate access to a broad variety of Web resources without assuming that they sacrifice their customized catalogs, format preferences, or institutional autonomy*". Interest in metadata registry systems arose from that exact need to be able to search across diverse metadata elements sets that were handled by different systems. It was essential to develop registration authorities that would support the registration, maintenance, cross searching of metadata elements and exchange of metadata information across heterogeneous systems.

Leviston (2001) distinguished between two different types of metadata registry system prototypes:

- Systems that usually served as reference points by listing URLs that point to Web sites of metadata initiatives and projects and
- Systems, which were concerned with the management of the evolution of metadata vocabularies over time and provide with mappings between metadata schemes.

The latter were usually developed to satisfy more sophisticated needs (p.2) and they have been based on the application of international standards such as the ISO/IEC 11179 standard that refers to the Information Technology: Specification and Standardization of Data Elements and the Dublin Core metadata standard. Baker and colleagues (2001) noted that the term registry "*covers a broad range of databases, documentation services or web-based portals providing access to schemas*". Heery and Wagner (2002) discussed how metadata registries "*essentially provide an index of terms*" and complementary to Baker and colleagues they described metadata registries that provide links to externally maintained terms and definitions of schemas such as the ROADS software templates and metadata registries that could harvest metadata schemes from where they were stored and maintained. Day (2003), also differentiated between registries that take the form *of "a database or [metadata registries] encoded in a structured syntax like RDF"*.

Irrespectively of the type of metadata registry systems they all comprise of data elements accompanied by their assorted semantic and syntactic representations. Duval and colleagues (2002) encapsulated the meaning of semantics and syntax in the following phrase,

"Semantics is about meaning; syntax is about form"

Mayes (2000) also differentiated between the syntax that he defines as the representational aspects of data and semantics, the conceptual content of the data. The semantics that accompany a data element describe the meaning that the data element can adapt with precise definitions.

One of the earliest efforts to address the issues associated with the exchange of information across heterogeneous systems in a networked environment was the establishment of the ISO/IEC 11179 standard. This is discussed in the following section.

The ISO/IEC 11179 Information Technology – Specification and Standardisation of Data Elements

The ISO/IEC 11179 Information Technology – Specification and Standardisation of Data Elements is an international standard that consists of six (6) parts and aims to support data standardisation and interoperability by providing guidelines on how to specify, standardise and register data elements. The six parts are described hereafter:

Part 1: Framework for the specification and standardisation of data elements.

Part 2: Classification for data elements.

Part 3: Basic attributes of data elements.

Part 4: Rules and guidelines for the formulation of data definitions.

Part 5: Naming and identification principles for data elements.

Part 6: Registration of data elements.

Part 3, which refers to the basic attributes of data elements, was the first part of the standard to be created in 1994. Currently, all six parts are at 60.60 stages, which refer to international standards that have been published.[7] The problems that the ISO/IEC 11179 standard addressed were those of data sharing and standardization. Its purpose was to make data elements sharable and understandable. It denoted that for data to be shareable it must be based on a common understanding among those who create it and those who use it. It proposed the establishment of a data element registry and gave guidance on how to classify, describe, name, identify and maintain both the data element descriptions and the metadata that were intended to be used to configure those data elements. Some of the advantages of using an ISO/IEC 11179 based metadata registry system referred to the ability to group similar data elements together under a shared concept, have access the all data elements that were linked together due to the sharing of a concept and showing the relations between them, group together data elements that share the same values and use linked concepts to access data elements and vice versa (Bargmeyer, 2000). Furthermore, more general benefits such as having a single point of reference for data harmonisation in an organisation, establishing an authoritative source of information and making data more understandable to people, acting as a central point for managing metadata, increasing the feasibility of exchanging data with other agencies and/or organisations were also acknowledged (Bargmeyer, 2000).

Although when it originated, it aimed to address problems associated with the exchange of information among US governmental organisations, the ISO/IEC 11179 standard grew out to an international standard. A few representative examples of the ISO/IEC 11179 implementations can be found at the following organisations:

[7] Information available at: http://www.iso.ch/iso/en/widepages/stagetable.html (last accessed 12/06/2006)

- Australian Institute of Health and Welfare - Metadata Online Registry (METeOR). (http://meteor.aihw.gov.au/content/index.phtml/itemId/181162). Previously known as Knowledgebase
- US Environmental Protection Agency – System of Registries. (http://www.epa.gov/sor). Previously known as the Environmental Data Registry.
- US National Cancer Institute - Cancer Data Standards Repository (caDSR) (http://ncicb.nci.nih.gov/NCICB/infrastructure/cacore_overview/cadsr)

2.5 Overview of metadata registry systems research and implementation

Research about metadata registry systems took off in the USA and Australia during the 1990s. The research was driven by governmental organisations that with the explosion of the World Wide Web sought to identify standardised approaches for the organisation of federal information across the USA. In Europe the pioneer in the research and development of metadata registry systems was the United Kingdom for Library Networking centre (UKOLN). Under funding from the European Union, UKOLN developed the DESIRE metadata registry. The DESIRE metadata registry was the first in Europe to address issues of interoperability and a single source to register elements of metadata element sets. Implementations and follow ups to the DESIRE project led to the development of ROADS software, which has been implemented to many Internet-based subject gateways to support resource description and to facilitate cross searching and interoperability among resources. Among the most recent metadata registry projects in Europe were the "SCHEMAS: A Forum for Metadata Schema Implementors" and the CORES projects. The SCHEMAS project was a collaboration between the UK Office for Library Networking, the PricewaterhouseCoopers Consultants and the German National Research Centre for Information Technology. Its objectives were to "*provide information for schema implementors about the status and proper use of new and emerging metadata standards, as well as promoting good-practice guidelines for adapting multiple standards or metadata modules for local use in customised schemas*"[8] and it aimed to reach individuals and institutions that were interested in the implementation and use of metadata standards. The project delivered a series of workshops that explored issues about the management of currently implemented metadata schemas and encouraged the participation of users from all scientific domains. It was expected that by the completion of the workshops the organisers and the workshop attendees would have addressed research issues such as the registration of diverse metadata element sets, suggestions

[8] SCHEMAS project. Available at: http://www.schemas-forum.org/ (Last accessed 12/06/2006)

for the structure of a metadata registry system and what were the best practices for the use of metadata. Furthermore four metadata watch reports reviewing the use of several metadata element sets as recognised standards in the audiovisual, cultural heritage, educational, publishing, and the Governmental domains. It was expected that by the completion of the metadata watch reports the SCHEMAS researchers would have drawn the landscape of metadata research in the domains of academia, audio visual, cultural heritage, education, geographic information, industry, publishing and rights management, research and other.

Another metadata registry system in Europe was the MetaForm. MetaForm (http://www2.sub.uni-goettingen.de/metaform/index.html) was an outcome of the Meta-Lib project, an initiative of the *National Library of Germany (*Die Deutsche Bibliothek) and *SUB Göttingen*, which was funded by the German Research Foundation and aimed to address issues and challenges that networked resources, their storage, preservation and retrieval posed upon the German library community. MetaForm is defined as a "*...database for metadata formats with a special emphasis on the Dublin Core and its manifestations as they are expressed in various implementations. As a project deliverable, the idea behind this database is to identify the core elements that are used for the description of networked resources*" (MetaForm, http://www2.sub.uni-goettingen.de/metaform/index.html). It comprised of the following three services:

- Crosswalks. This section of MetaForm stores the Dublin Core metadata element set and other application profiles that were formed by using the Dublin Core as their base. It also provides the relative documentation for each element set or application profile such as a description of the element and how can be used in context. In some cases examples of use in practice are also provided. Furthermore, crosswalks list associated DTD schemas, and assorted organisations that they have implemented the relevant application profiles or element sets.
- Crosscuts. The second of the three MetaForm services describes how Dublin Core elements are used in various implementations.[9] Crosscuts are presented in the form of a table with two columns. In the first column there are listed the variations that an element of a metadata element set or application profile can take and in the second column are listed the descriptions of this element.
- Mappings. The last of the three MetaForm services lists mappings of "*DC applications and other formats with each other*," [10] which are dynamically generated between two element sets at a time. A typical representation of a mapping includes data presented in a table consisting of four columns: (i) the first column cover the elements of the first set (ii) the second column presents the description of each element, the third column

[9] About the server. MetaForm, http://www2.sub.uni-goettingen.de/metaform/index.html (Last accessed: 12/06/2006)
[10] ibid.

covers the equivalent (or otherwise blank) element of the second set, and finally the fourth column describes the elements of the second set.

In Australia, the Institute of Health and Welfare's Knowledgebase is described as "*...an electronic register of Australian health, community services, housing and related data definitions and standards. It includes the relevant National Data Dictionaries, national minimum data sets, and the National Health Information Model.*" Knowledgebase incorporates for the first time in electronic version the National Health Data Dictionary and the National Community Services Data Dictionary, which had been respectively published in hard copies since 1989 and 1998. It bases the data element definitions used from the previous resources on the ISO/IEC 11179 standard. Additionally it supported the use of national minimum data sets and facilitated links across information agreements on collection and provision of data and data elements in Australia. In May 2005 Knowledgebase has been replaced by METeOR. METeOR has incorporated feedback from users and improved the search options of the metadata registry by being able to search across multiple metadata elements and registration authorities. Furthermore, it enhanced the feature for downloading relevant information, yield lists in alphabetical order and maintains a feature called "my items" where users can store their searched and retrieved information to download.

The Dublin Core Metadata Registry has been defined as "*... a collection of RDF schemas, application profiles and related semantics belonging to various resource communities. Its goal is to promote the discovery, reuse and extension of existing semantics, and to facilitate the creation of new vocabularies*" (Dublin Core Metatata Registry, http://dublincore.org). Heery & Wagner (2002) described the working progress of three prototypes of the DCMI metadata registry and they saw the primary role of the DCMI registry as to facilitate semantics by agreeing on using the same rules for describing Internet resources and having the respecting tools to support them. Additionally, another important aspect of any registry system was pointed out to be the stressed the importance of the re-use of metadata elements in different languages (Nagamori and colleagues, 2001).

The pioneer of metadata registry systems in the USA was the Environmental Data Registry (EDR). The Environmental Protection Agency (EPA, http://www.epa.gov/sor) defined the EDR as "*...a comprehensive, authoritative source of reference information about the definition, source, and uses of environmental data. The EDR catalogs the Environmental Protection Agency's (EPA) major data collections and helps locate environmental information of interest*". The development of the EDR began in 1993, and made significant strides under the auspices of the EPA's Reinventing Environmental

Information (REI) Initiative that began in the late 1990's. The EDR was EPA's primary resource for metadata pertaining to data within the Agency's major information systems and also served as a clearinghouse for EPA's data standards. Between 1993 and 2002 the EDR was the main registry system of the EPA but it was closely allied with several other registry systems including the Terminology Reference System, the Substance Registry System, the Chemical Registry System and the Biology Registry System (Figure 2). Those systems provided search tools for retrieving information about how environmental terminology and physical, chemical and biological substances were represented in the EPA's regulations and data systems. Analytically the following registries were hosted by EDR:

- Substance Registry Systems. The Substance Registry System (SRS), the Chemical Registry System (CRS), and the Biology Registry System (BioRS) that were the gateways to information on chemicals, biological, and other substances. They also held information about how they were represented in EPA regulations and information systems.
- Terminology Registry System (TRS). The TRS was a compilation of environmental terms, definitions and information sources relevant to the mission of EPA and other environmentally oriented entities.
- MetaPro. MetaPro was a tool that could be used to create a metadata registry (EPA Environmental Data Registry Users Conference Agenda, 2002).

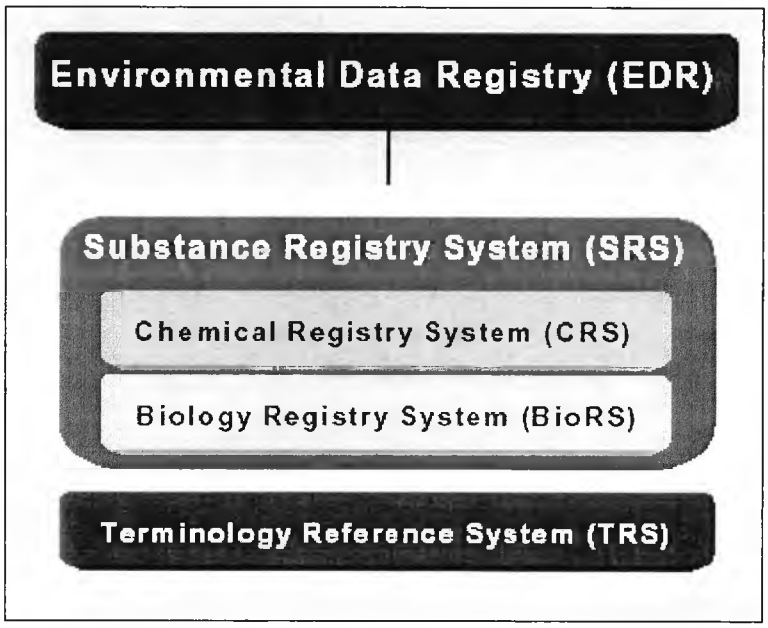


Figure 2: Literature Review - Graphical representation of the EPA's metadata registries (*Source: Standard Update: Better Data through standards, Vol. 2, No. 3, Winter 2000)*

Since its inception, the EDR had continually evolved in order to serve the needs of its users. The users' conference held in January 2002 contributed to continued improvements, and additional functionality has been added as a direct result of input

provided by conference participants. As described in the winter 2002 EDR newsletter "*Users recommended improvements to the application interface, data content, software functionality, and the process for how data standards are developed and implemented*". The EPA implemented those suggestions and an improved version of the EDR based on the suggestions of its users was launched in spring 2002. The newly introduced service was named as the System of Registries (SoR) and some of the changes that were implemented included enhanced search performance for the Chemical Substance Registry (CRS); site redesign to improve access to downloadable code sets; improvements to the Compare Tool (a tool designed to support data harmonization); and inclusion of XML tags associated with standard data elements.

The system of Registries (SoR) was defined as a web-based collection of metadata registries and repositories residing in the EPA's Office of Environmental Information (OEI). The registries that comprised the SoR provided identification information for objects of interest to EPA trading partners, including states and tribal entities, and the public. They described objects consisted of data elements, XML tags, data standards, substances (chemicals, biological organisms, and physical properties), terms and definitions, facilities, regulations, and data sets that the Agency used in its core business processes. These registries comprised a critical link in EPA's information architecture and were considered vital components to the Exchange Network that was developed to facilitate data exchange with stakeholders through network nodes. In fall 2002, the System or Registries consisted of the following six registries:

- Environmental Data Registry (EDR). The EDR provided access to EPA data standards, XLM tags and application metadata such as data elements.
- Facility Registry System (FRS). The FRS was a centrally managed database that identified facilities, sites, or places subject to environmental regulations or of environmental interest.
- Information Resource Registry System (IRRS). The IRRS provided information about the EPA's application inventory, organisation hierarchy, and other information resources (adopted from the fall 2002 Newsletter).
- Substance Registry System (SRS). The SRS acted as the gateway to information about chemical, biological and miscellaneous organisms as they are represented in the EPA's regulations, data systems, and other information resources.
- Environmental Information Management System (EIMS). The EIMS was a single source of information to spatial data. It hosted information such as metadata on remote sensing data, Geographical Information System (GIS) coverages, databases, documents, models, and multimedia.

- Terminology Reference System (TRS). The TRS provided access to environmental terms and definitions that were compiled from a wide range of EPA program offices, information systems, and state collections. This registry could be searched by keyword, information resource, and organisation.

The SoR continued to evolve over the last years and some of the significant changes of the recent years included the development and implementation of two additional registries. Those are described hereafter:

- Web Registry (WR). WR is described as a centrally managed database used to collect metadata for EPA's priority web materials. This registry enables the combination of metadata with the full text of corresponding pages and documents, in order to produce more relevant search results.
- XML Registry. The XML registry holds information such as XML Data Exchange Template (DETs), XML Schemas, Namespaces, WSDL files, and other supporting files needed to map data flows between partners.

2.5.1 Users of metadata registry systems

To the researcher's knowledge there are no other studies in the literature that investigate how the users of metadata registry systems make use of them and what their expectations of their services are. Nevertheless, the majority of those systems have established reports about their functional use and the intended use of their systems. The Metadata for Education Group (MEG) registry that focused on the use of metadata for education and learning defined the purpose of the MEG registry as a publication environment for intended developers, implementers and researchers to publish their metadata schemes and disclose information on their usage. Johnston (2002) identified 6 types of users of a metadata registry. Those were:

- Those who create metadata schemes and want to publish them,
- Those who implement metadata schemes and want to publish application profiles,
- Those who search for appropriate application profiles being either implementers or developers,
- Those who are interested in finding guidelines to use metadata element sets,
- Researchers who are interested in the use of metadata schemes and
- Developers studying tools to support schema usage.

Heery and Patel (2000) argued that people who create and manage metadata are usually those who also create standards and implement them. Therefore, these users would require facilities that would allow them to register both the combination and mixed elements and metadata element sets. For example, the CORES metadata registry, declared its role of more than a "dictionary" of metadata element sets. The users of the

registry were expected to register pragmatic metadata element sets and/or application profiles and although it was built around the Dublin Core element set it welcomed the expansion to more than one metadata element set (Heery, 2000). The METoR registry in Australia also differentiated between general metadata users and metadata developers. General users included anyone who would visit the METeOR registry system and search and download metadata information irrespectively of type and format. The metadata developers have been identified as those users who would user the METeOR development tools additionally to features available to all other users. They would usually review and suggest metadata information to be included to the registry system and they were assigned a predefined user name and login information to access and use the METeOR registry. The Dublin Core metadata registry defined four categories of intended users based on their search interests. Those were information seekers, computer specialists, applications and administrators. The differentiation in the DCMI registry was the option for a searchable interface by computers and that was encapsulated among the potential users. Duval and colleagues (2002) also identified in the intended audience of metadata registry users application designers, creators and managers of metadata, applications and finally end users.

2.5.2 Role and functionality of metadata registry systems

Duval et al (2002) argues that metadata registry systems are a key component in digital libraries research because as the use and development of element sets from different scientific communities will grow, the need for authoritative systems that will facilitate the maintenance, storage and reference to element sets and the guidelines for their use will increase, too. Current metadata registry systems indicate a variation in functionality. Belkin (2002) denoted that system's functionality should not be seen separately from the interface to a digital library or an information system. The StORe project (http://jiscstore.jot.com/WikiHome) defined functionality in the domain of information technology and computing as *"... a term that describes the capacity of a system (software application or computing device) to perform the functions required by the user. It can be used to identify the features, capabilities or behaviours of a system – otherwise known as the things it can do".* Functionality in metadata registry systems is closely associated with their role as tools that facilitate the management of metadata by specifying data elements and describing their structure and meaning (DCMI Open Metadata Registry Functional Requirements, 2001[11]; Baker et al, 2001; Nagamori et al, 2001). The role of metadata registry systems is prescribed hereafter:

[11] This document is a DCMI working draft. Available at: http://dublincore.org/groups/registry/fun_req_ph1.shtml (last accessed on the 12/06/2006)

- To support data standardisation and data documentation by allowing registration of metadata elements.
- To support the management of basic semantics by clearly defining the context of use.
- To function as an authoritative source of reference information about data. One of the most important characteristics of a metadata registry system is to act as an authoritative source about the management, development and evolution of the data pertaining in such a system.
- To act as medium for recording and disseminating data standards by providing free access to publicly and freely available standards for the description, exchange and access information.
- To function as a "mediator" to the actual source of information.

Although each metadata registry system prioritised different functions, primarily those that were considered as most important for their intended users, there is a core set that are considered essential of the role and requirements of all metadata registry systems. These are:

- The description of data elements,
- Provision of guidelines for their use,
- Mappings across elements of different metadata sets and,
- For the case of DC registry, facilitation of multilingual searching.

The consistent update of any metadata registry system's content is also essential for validity and credibility. A description of the purpose and scope of DC with a particular emphasis on multilingual support is provided on the DC Registry Web site. Information on minimum levels of metadata description to resources hosted on Internet based subject gateways is included in the DESIRE II Handbook.

Duval et al (2002) similarly to Heery (2002) noted that metadata registry systems distinct between those that act as pointers to authoritative lists of rules and those that could harvest those rules across many such dictionaries. A similar approach is discussed in Leviston (2001) where he discussed the role and functional requirements for a metadata registry system with application to record keeping.

2.5.3 Use of metadata registry systems

A combination of human, social and technical factors affects the way users navigate across an information system. Hsieh-Yee (2001) listed among the factors that affect the information seeking process *"...a user's background and experience with computers, the*

Web, and other information retrieval tools can affect how he or see seeks information. Information need, domain knowledge, cognitive abilities, affecting states, demographics and the environment of the information need also contribute to the way in which the seeker seeks information". Furthermore, issues such as how the information system organises and presents the information to its users and what services provides to support their searches contributes to both users' experience and expectations of the system. Additionally, the quality and relevance of the retrieved information directs the user in specific features in a system, like using any help or support services that are provided, re structuring their search or even use a different information system.

Bargmyer (2003) identified two purposes for statistical metadata registry systems. Those are a) user oriented purposes and b) production oriented purposes. In detail they address issues about the re use of metadata and reduction in the duplication of effort in the production and management of metadata. Xu (1997) argued that the very fact that metadata element sets can be mapped shows that they share commonalities across their elements, their semantics and their syntax. As described by Bargmeyer (2003), Xu also noted that metadata registry systems should be able to collocate all metadata elements and metadata element sets that are linked to a concept and maintain all the linkages across both element and element sets and concepts. Standards for managing sharable data focus on better defining the semantic layer of the description and develop structures that help to manage the registry.

2.5.4 Metadata registry systems and interoperability

There are many definitions about interoperability. Johnston (2001) and Miller & Johnston (2000) distinguished between technical, semantic, political/human, inter-community and international interoperability. The IEEE (1990) defined interoperability as the "*the ability of two or more systems or components to exchange information and to use the information that has been exchanged*". The Dublin Core metadata initiative expanded the previous definition by noting "*the ability of different types of computers, networks, operating systems, and applications to work together effectively, without prior communication, in order to exchange information in a useful and meaningful manner. There are three aspects of interoperability: semantic, structural and syntactical.*"[12] Furthermore, Caplan (2003), associated interoperability with metadata, and defined it as "*the ability to perform a search over diverse sets of metadata records and obtain meaningful results*". Blanchi and Petrone (2001) defined metadata interoperability "as a measure of the compatibility

[12] Dublin Core Metadata Glossary. Available at: *http://dublincore.org/documents/usageguide/glossary.shtml*, *last accessed: 12/06/2006)*.

of two metadata sets...metadata interoperability represents the ability of a system to cross-walk from the conceptual space of one metadata set to the other".

Preston and Lin (2002) differentiated between semantic and syntactic interoperability in digital libraries. They define semantic interoperability as the ability of the system domains to understand and share both the meaning and use of terminology from different domains and the mappings across diverse metadata elements. By syntactic interoperability they note the ability of different software components of systems to be able to exchange meaningful information. Under the auspice of the European Research Consortium for Informatics and Mathematics (ERCIM) a series of workshops addressed how metadata could support interoperability in the Digital Libraries research. Interoperability among systems that use a common metadata standard is clearly easier to support by current technology than to those systems that use diverge metadata standards. There have been two approaches to the exchange of metadata so far. The first one is based on the Z39.50 protocol used for harvesting. The philosophy behind this approach is that systems do not share their data but they map their metadata elements to common set of search attributes. The second approach, endorsed by the Open Access Initiative has been for all data providers to translate their metadata to a common metadata element set (in most cases the Dublin Core metadata element set) and expose this for harvesting (NISO, 2004).

Metadata element sets provide an essential basis for the exchange of data between systems as they define a core set of elements and the rules that accompany them and define their use. Furthermore, they can be understood by both humans and machines (NISO, 2004). Haynes (2004) argued that "*metadata fulfils an important role in enabling this to work, by establishing standards for data elements and by providing information about the data on one system so that is can be processed and used by other systems or departments*". Larsen (2002) noted that metadata interoperability is essential to support the following:

- Search and retrieval,
- Intellectual property and rights management,
- Administration and preservation and
- Evaluation and use.

Eden (2002) discussed how metadata can be used to promote interoperability. He argues that if it is "*accompanied by careful mapping of data elements and crosswalking of standards. Interoperability allows multiple systems to exchange data with minimal loss of content and functionality, regardless of different hardware and software platforms, data structures, and interfaces*".

Chapter 3 – Methodology

This chapter provides an overview of Internet-based studies and discusses the advantages and disadvantages of mixed methods research. It also presents the methods that were used for this research study. Those were: A) questionnaire surveys, B) web logs analysis and C) email interviews. Furthermore it discusses the sampling and recruiting process and also presents the metrics and questions used within each method and what they aimed to achieve.

3.1 Introduction

In the computing science, user studies are placed in the area of information retrieval (IR) and systems' evaluation (Jansen & Pooch, 2001). The use of quantitative methods to describe, observe, obtain and analyse data in user behavioural sciences is well documented as well (Wang, 1999). In the library and Information science (LIS) field though, the focus is placed on the results of the research rather than the methods used (Wang, 1999; Nicholas, 2000). The metrics that can be used in user studies can are usually examined from two different standpoints. Either from a specific view of identifying individual variables to measure user behaviour or from a broader view of trying to understand how a specific metric can determine what can be measured (Tenopir, 2003a; Wang, 1999).

Furthermore, although Internet based research studies had been thriving since the middle of the 1990s, established methodologies for them are still under discussion (Whitmer et al., 1999). Often the combination of two or more methods is needed to be able to draw sustainable conclusions from this type of research (Nicholas, 1999, 2001, Tenopir, 2003b).

For this study a combination of methods was employed such as questionnaire and online questionnaire surveys, transaction logs analysis and interviews. Combining quantitative and qualitative methods can enhance our understanding of how people do searches and why they do them (Tenopir, 2003; Bishop and colleagues, 2000; Fox, 1999). The methodology used for this study is based on models defined by the ciber group at University College London that employ the combination of more than one methods to ensure validity and quality of data. Wang (1999) and Ingwerson & Jarvelin (2002) also discussed the advantages of using a combination of quantitative and qualitative methods to conduct user studies. Human interaction with computer systems is a multi-dimensional

process and requires a variety and combination of methods to allow us to have an in depth understanding of how users search and why they conduct their searches in the particular way they do.

3.2 Internet-based research studies

During the 21st century the proliferation of the Internet had an effect to the way research had been conducted until then. Researchers primarily from the social sciences (psychology, behavioural sciences, and information sciences) investigated the use of the Internet to reach large numbers of people and collect valuable data. In the literature there are three main areas of Internet research. Those are use of systems or resources, completion of demographic surveys and use of empirical investigations. Internet based research methods include access to discussion groups and mailing lists, website visitor's tracking systems, online surveys (including e-mail and web-based surveys), online focus groups and transaction log analysis (Montaya – Weiss and colleagues, 1998; Haag Granello and Wheaton, 2004). Miller (2001) also discussed the way the World Wide Web chanced considerably the way research is conducted and he notes that in order to be able to understand the data that are collected and interpret them adequately, it is important to understand the context of online research. He argues that it is important to understand how people interact with technology and what experiences and attitudes they bring into it. Couper (2001) discussed the issues and approaches to web surveys and notes that the advantages of web based surveys outweigh the possible limitations. Among the advantages of using Internet based methods to conduct research are the reduced costs, the ease of administration and facilitation of data entry/process, the ability to provide for confidentiality and anonymity which can boost response, the ability to reach large numbers of people from diverse disciplines and irrespectively of their geographic location, allowing for flexibility in the format of the survey. Furthermore, discussion groups and mailing lists in particular allow us to have unobtrusive observation of conversations of a specialised group and to reach specialist groups. Visitor tracking systems provide with a *"traffic pattern"* of a website's use. Linking tracking systems data with user behaviour surveys allow us to have a rounder view of users' requirements and search patterns. Online surveys are means of gathering data from population that is difficult to locate. Online focus groups allow us to have real time and interactive communication with disperse groups. Advantages associated with the use of the World Wide Web to conduct research are discussed hereafter:

- Reduced costs. The cost of running a survey is usually divided between materials/consumables and the wages for the people working on the research study. The cost of creating a web based survey is usually minimal for an academic

institution that has adequate Internet presence and technical support. Creating and uploading an HTML form is not difficult but it may be needed to create a CGI script to associate with the web page so that it records the responses of the respondents. This way it can be ensured that data will be returned in an electronic form and therefore facilitate the ease of data entry and processing after the survey is completed. In the case though, that a researcher needs to set up a server, design and upload a web based questionnaire, write a CGI script, collect, print and analyse the results of the respondents the cost could increase dramatically.

- Ease of administration. Compared to traditional, paper based surveys web based and email surveys have the advantage that they record users' responses in electronic form. This advantage eliminates transcription mistakes that could occur during data entry and makes the process and the analysis of data easier to manipulate. Time saved from data entry though should be put into the initial construction of the questionnaire and the posting of the survey request message to the discussion and/or mailing lists. Explaining the aims of the study, providing with contact details for further elaboration of questions that respondents may have with the set up, structure and the aims of the questionnaire are important to ensure higher response rates. Designing the questionnaire in a way that allows for enough virtual space for the respondents to answer open – ended questions, providing adequate options to chose from in closed – ended questions and its structure reflects to the aims and intentions of the study contribute to earning credibility and attracting more response. After the end of the survey, thanking the respondents for their contribution to our research and providing an initial descriptive analysis of the responses is considered good practice and most respondents' value and appreciate.
- Confidentiality and anonymity. An ethical consideration for respondents' identity is an important aspect for any type of survey, traditional or Internet based. Guaranties for the anonymity of the response should be provided and respected. It has been reported that respondents provide with lengthier and richer answers when they feel that they are not going to be judged about them and that they are free from any social criticism.
- Reach/Recruitment of respondents. It has been argued that the Internet makes it possible to reach people that otherwise we could not have access to due to factors such as geographic location or listing. The ease of use and the speed in which web surveys can be set up and made available has been one of the characteristics of internet based studies in general (Couper, 2001). Furthermore the nature of discussion and mailing lists group together people that share the same interest in a subject and therefore making it easier to reach more people that could be representative of a user population.

Disadvantages and/or limitations of Internet based methods for data collection include sampling problems, response rates including non-response, measurement problems and technical errors (Couper, 2001; Grandcolas and colleagues, 2003; Haag Granello and Wheaton, 2004).

- Sampling. Although the Internet can be used to reach a large number of people, there has been criticism in the literature that it is difficult to include people from the target population, as it is difficult to identify users. People that do not have access to Internet access are by default excluded from any survey. The whole of the Internet population remains unknown and as such we can not predict the bias that the existing users can measure up to. Kayne and Johnson (1999) pointed out that *"the difficulties of online sampling stem from the lack of a central registry of Web users"*.
- Response rates/non-response. Non-users of the Internet do not have access to the surveys. Although it could be argued that *"frequently web recruited surveys are used to research attitudes to the Internet or attitudes to web sites"* (Grandcolas and colleagues, 2003). It has been very difficult to be able to measure non response to a web survey when the total number of those who could participate to the survey is not known. In some cases, discussion and mailing lists provide the number of their current subscribers at a given moment but in the cases that this information is unavailable, non response can be very difficult if not impossible to be measured. Web log transactions in such cases could prove a helpful tool as they can provide an indication about accesses to a particular system at a given time and therefore provide some measurement of potential users.
- Measurement. Measurement errors can arise from both the researcher and the respondent's end. Formulation of survey questions, survey's design are important for the respondent to understand and be able to provide with answers. Lack of motivation, problems comprehending the questions, not feeling reassured regarding confidentiality and anonymity of the response can result in measurement errors. The design of the questionnaire and the restrictions that it can impose such as forcing respondents to select one of the available responses can eliminate mistakes that can arise from misunderstands from the wording of the questions. This though should be outbalanced with providing enough options to cover all potential answers.
- Technical errors. The appearance of a web survey on the respondent's screen can vary significantly due to software, browser settings, screen display. Bosnjak (2001) discussed how the graphic designs included in questionnaire surveys, tables, lack of navigational tools can affect the response rate of any web based survey.

3.3 Mixed methods research

In areas of research that are new and innovative such as those that deal with developments, implementations and applications of digital information, the concerns when choosing the appropriate research methods lie in the validity of those methods and their applicability to the specific area of research. Research in digital libraries had, mainly, been divided between systems and services. In an attempt to define metrics for digital library evaluation, the DLib metrics working group was put together with the objective to "*define a set of scientifically rigorous metrics and measures that would enable comparative evaluation of information discovery techniques and algorithms that yielded repeatable results over multiple experiments*" [13]. They considered evaluation methods for three areas, that of systems, users and content. They concluded that for the three areas we need identify metrics that would allow us to assess:

- Systems: Interoperability, scalability, heterogeneity, reliability and integration.
- Users: Relevance, specificity, timeliness, effort vs. effect and usability.
- Content: Sufficiency, currency, and quality.

Metadata research studies find application in all three areas. But assessing systems requires a different methodological approach than assessing users (Boyce, 1994). A combination of both quantitative and qualitative methods would be necessary to provide reliable results. Kaske (1993) argued that *"human/computer interaction is just too complex to be studied by only a qualitative or quantitative method".* Researchers have argued that the use of more than one method can enhance and enrich the understanding of research findings. It can be done in a sequential mode gathering data from one method (e.g. survey) and use it to design the next one (e.g., in depth interviews) or in parallel. Also, the order, the level, the type and the proportion of different methods is not important although research has shown that is usually a combination of three methods that is mostly used. Roco et al (2003) listed five purposes for using mixed methods for a research study. Those were triangulation, complementarily, development, initiation and expansion. They discussed examples of possible ways of mixing methods such as the combination of a qualitative interview and a quantitative questionnaire to enhance and complement data about users' perceptions (p.22). Among the advantages of using a combination of methods are the validation of research data, encouragement of creativity that could stimulate further work and expansion of the scope of the study (Tashakkori and Teddie, 1998). Greene and colleagues (1989) added also development and initiation. Rocco and colleagues (2003) noted that "*mixing makes room for both the exploratory inductive process that begins with empirical evidence of the particular and proceeds with*

[13] Larsen, Ronald L. (2002). The DLib test suite and metrics working group: harvesting the experience from the digital library initiative, pp. 14. In, Borgman, C....et al (eds.). Evaluation of digital libraries: testbeds, measurement, and metrics. Budapest: ERCIM.

to a level of abstracting/theorizing/generalizing and the confirmatory deductive process of hypothesis testing of theories".

On the downside, triangulation can be time consuming and it employs the use of different levels of analysis as this is imposed by the different methods. Expertise in both qualitative and quantitative methods is essential in order to be able to adequately process and analyse data.

3.4 Research methods used

The methods that were used for this study were three: questionnaire surveys, email interviews and web log transactions. The analysis of more than 10.000.000 successful requests of which 2625672 referred to the activity recorded by the MetaForm registry and 9932969 referred to the activity recorded by the Environmental Data Registry and the System of Registries and the email interviews provided rich information about how users interact with metadata registry systems and their expectations of such systems. The questionnaire surveys conducted with users of the two active metadata registry systems also highlighted attributes of specific features in each of the metadata registry systems. The email interviews provided more in depth information about users' expectations of metadata registry systems, current implementations and applications of metadata in use and views regarding the role of metadata registry systems in data management and exchange of information. The following table (Table 1) presents the type of methods used by registry system and date.

Methods	Web log transactions	Print/Online questionnaire surveys	Email interviews
Discussion lists		November 2000	
SCHEMAS		November 2000	
MetaForm Metadata registry	1998 -2004	April 2001	
EDR Metadata registry	1998 - 2004	January 2002	
SoR Metadata registry	1998 - 2004	March 2003	
Various			October 2004 – November 2004, May 2005

Table 1: Methodology - Research methods used by registry system and date

It has to be noted that additional methods such as observation of how users interact with a system would have provided useful information but at the time that this study commenced it was not possible to implement a prototype registry system as the open access software was not available.

The questionnaire surveys and the email interviews provided useful qualitative data to the survey. The web log transactions for the two metadata registry systems provided much of the quantitative data that are presented and discussed in Chapter 4.

3.4.1 Web log transactions

Peters and colleagues (1993) defined transaction log analysis as *"the study of electronically recorded interactions between online information retrieval systems and the persons who search for the information found on those systems"*. Transaction logs record the interactions that users conduct with information systems. Peters and colleagues (1993) notes that *"...researchers most often use transaction log data with the intention of improving an information retrieval system, human utilization of the system and human (and also perhaps system) understanding of how the system is used by information seekers"*. Kaske (1993) complemented this statement by discussing that the aim of analysing data from transactional logs is to acquire knowledge that will contribute to the understanding of how users use information systems. Burton (2001) also pointed out that the way users interact with a World Wide Web site provides valuable data about both the usefulness and the effectiveness of the web site. Web log transactions facilitate a better understanding of how users work through a system and which services and/or individuals features they had used. They record the overall traffic of a system by counting page requests that the users had made. They provide some information on the origins and the ID of the system's users as they note the geographic location and IP address of the access point. Furthermore, they enable system administrators and service providers to have an indication of the users' interest in particular resources as they count the files that they have downloaded and record the use of the system by peak hours and days of the week.

Although transaction logs can provide information about how users move around a system they provide limited empirical information about their search behaviour. Kurth (1993) discussed the limits and limitations of transaction logs and distinguishes between system factors; user/search processes factors, data analysis and ethical and legal factors that limit this method. Those include the fact that users are unaware that they are monitored during their search and their consent to the use of such data is not sought as their identities are linked to IP addresses of computer systems and not to person's details. What is recorded in the access log is not the personal details of a user but and IP address of a computer that most of the times is impossible to link directly to a person. Furthermore, it is essential to be able to differentiate between a system's performance and a user's performance. It is vital that the researcher is able to read and understand the access logs and where applicable to employ statistical methods to complement primary data. The difference with web log transactions though is down to its sheer volume, detail of information and the potential that they encompass. Although they provide rich information they are not reliable for precision and attribution (Nicholas and colleagues, 1999). To counterbalance proprietary log analysis software it is essential to

employ statistical methods to validate results. When this is done though, the richness of data and its sheer volume make it an invaluable source of information to assess use and in some cases functionality of systems.

The main points of the transaction logs critique include the following:

- Number of "hits" versus the number of accesses. The number of "hits" reflects the number of items a user downloads when they access a particular page of the system. That could be images, text files, actual web pages, etc. The number of accesses reflects the actual pages of the system that are being downloaded regardless of the number of files that this page entails and represent a more accurate representation of the page's overall traffic (Bertot, 1997).
- The format of the file that the transactional logs are available. Most of the web based systems provide statistical information about the use of their website. Sometimes this type of data is available in an open access web page such as the case of the System of Registries and sometimes it can be obtained after direct communication with the web system's server administrator such as the case of the MetaForm registry system. In most cases the web log transaction data that are available on the World Wide Web come in the form of an html file. This format poses some limitations as it requires transferring this data to a relational database (such as MS Access) and process data manually. The option of accessing the complete data set by the organisation is not always applicable as in some cases this type of statistical process of data is assigned to external contractors to a system and issues with ownership of data could prevent access to it.
- Domain reports and domain name server (DNS) lookup. In order to tackle problems associated with slow connections local files are stored on the user's browser. When a user requests a page from a website, the system first checks its cache memory of the page. If the requested page is available in the cache then it is retrieved from there and nothing is recorded by the server. Furthermore, accessing a web page through a corporate, departmental or campus firewall that stores its own cache for all their users does not link the access to an individual user but the organisation they represent. Furthermore, Internet service providers (ISPs) also maintain their own-shared caches which link the accesses to the ISP rather than the user. Hence, a true story of how end-users search the system is not recorded (Nicholas, 2000; Dowling, 2001).

Some of the reports that are produced by transaction logs are the following:

- Access Log (server's audience and usage patterns). This type of log records the IP address of the computer that accessed the web site and how users behave when they are in the system or at the site. This could include records of times spend at one

particular page, navigation across different directories stored in a website server and patterns of accesses of individual pages.

- Date, Time, IP address and user action, e.g. which files the users downloaded. Bertot (1997), suggested that if the above are considered as variables then it is feasible to generate information regarding the percentage of users from a particular domain, the use that a particular domain is making of a web site, the use of the site during a particular time and day and the navigating behaviour of the user while in the system or at the site. Dowling (2001) explained that the log of a *"user session"* can be defined as *" a series of hits from a specific address with no gaps of more than a specified amount of time between them"*. Usually this type of information was made available from system transactional logs and it is not common in web sites transactional logs.
- Agent Log. This type of report gathers data that can be important for the design and development of web sites. They include recording of the users' browser type, browser version and operating system.
- Error Log. This type of report gathers data about missing files, erroneous links and aborted downloads. Error logs provide information of erroneous files along with the time and domain name of the user and therefore can be very useful to system administrators to establish problems in specific directories of the web site and its design.
- Referrer Log. This type of report records data about the web sites that directed the users to the first web site. It provides useful information about which domains had demonstrated an interest in a web site and therefore help managers of the web sites to target user groups.

3.4.1.1 Sampling and recruiting

The web log transactions of the Metadata Server's usage were provided by members of staff at the University Library of Lower Saxony (SUB) Göttingen, the host institute of the MetaForm registry. SUB Göttingen used the Analog 3.2 software (http://www.analog.cx) to capture and analyse their web log transactions. The data referred to 2625672 successful page requests, covered the period starting from July 1998 to the 10th of November 2004[14] and they were provided in three separate files. The files were split into three date ranges:

- July 1998 – April 2001,
- May 2001- April 5th 2004,
- April 6th – November 10th 2004.

[14] The researcher is grateful to Dr Carola Wessel who provided the initial log files (covering monthly statistics of the months December 2000 and January, February, March and April 2001) in 2001.

The web log transactions listed the accesses of all services that the Metadata Server acted as a gateway to, including both MetaForm and MetaGuide. As a consequence, the data that referred to MetaForm needed to be isolated. The directory report that recorded the number of accesses by the metric of requests for pages for each of the Metadata Server services was used as the basis to work out averaged statistics for the MetaForm registry. This fact had advantages and disadvantages. Among the advantages were the presentation of the total number of page requests on a weekly basis and the ability to verify total number of accesses over a given period. It complicated the analysis though, as it was not possible to verify weekly and monthly accesses on a yearly basis. For this reason, it was needed to calculate accessed to the closest proximity and present results as averaged statistics. The calculations that were conducted for each metric are explained under the relevant metric sections.

Web log transactions for the Environmental Data Registry were freely available at the following URL http://www.epa.gov/reports/objects/idmdssc1/ and dated back to the use of the system in 1996. Similarly to the MetaForm, EDR used the Analog software (http://www.analog.cx) to capture and analyse their web log transactions. Since 1996, the metrics that were used had been adjusted to reflect both changes in the Environmental Data Registry's structure and adhere to improvements in log measurements. Figure 3 presents an example of the first file of EDR web logs, dating the 6th of March in 1996 and reflecting logs transaction during the month of January 1996. EDR was identified as an "Untitled Object on EPA Web Server" and there were no counts reported. The metrics that were used at the time counted the numbers of total completed requests, total failed requests, total redirected requests, average requests per day, corrupt log file entries and total and average number of transferred bytes.

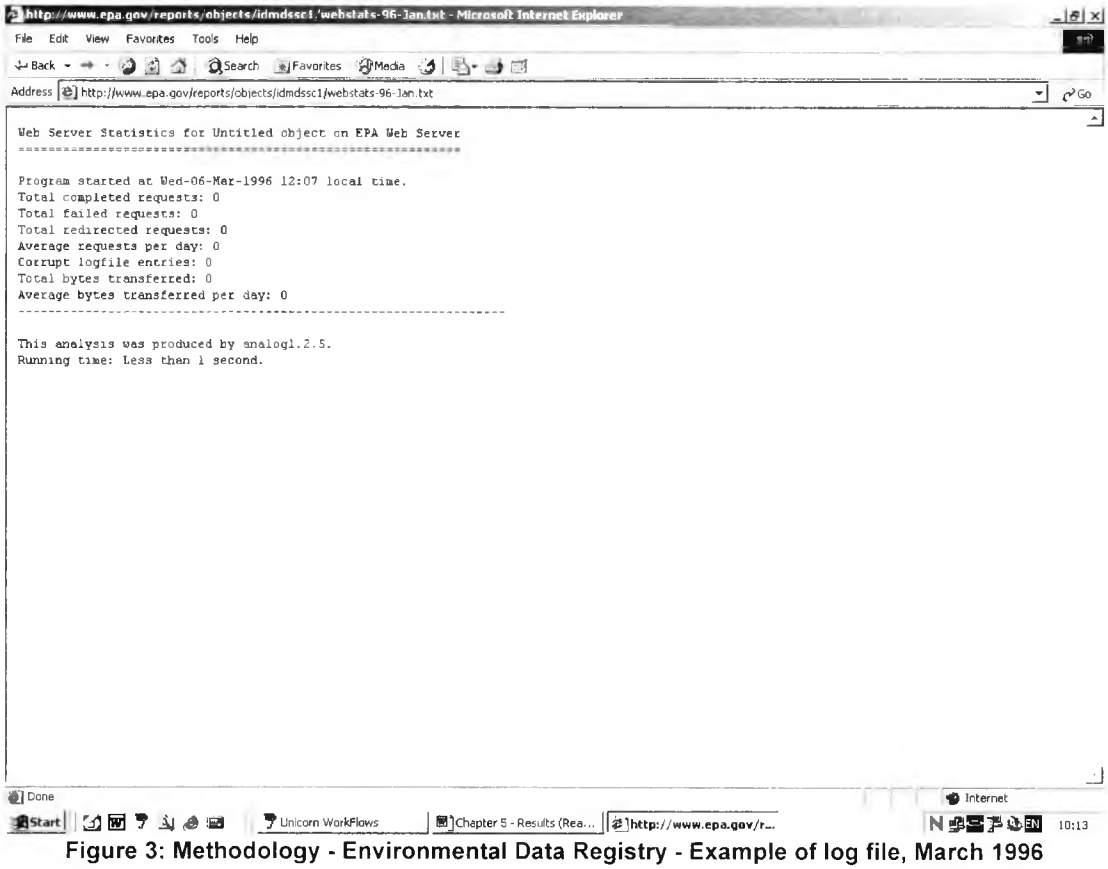


Figure 3: Methodology - Environmental Data Registry - Example of log file, March 1996

Compared to a recent file of web logs it can be seen that web log transactions provide rich data and facilitate the measurement of hourly, daily and weekly use of the registries. Additionally, they collect information about the domains (the countries), the referrers (the sites that directed computers to our services) and the requested files that were downloaded. They enable the monitoring of the use of the individual directories (specific registries and/or services) from which the requested files were downloaded. Furthermore, they give access to technical information of the computers that accessed services such as browser types, operating systems and file types downloaded (Figure 4). The data analysed referred to 9932969 successful requests for the Environmental Data Registry and the System of Registries and covered a period of six years and eight months dating from January 1998 to October 2004.

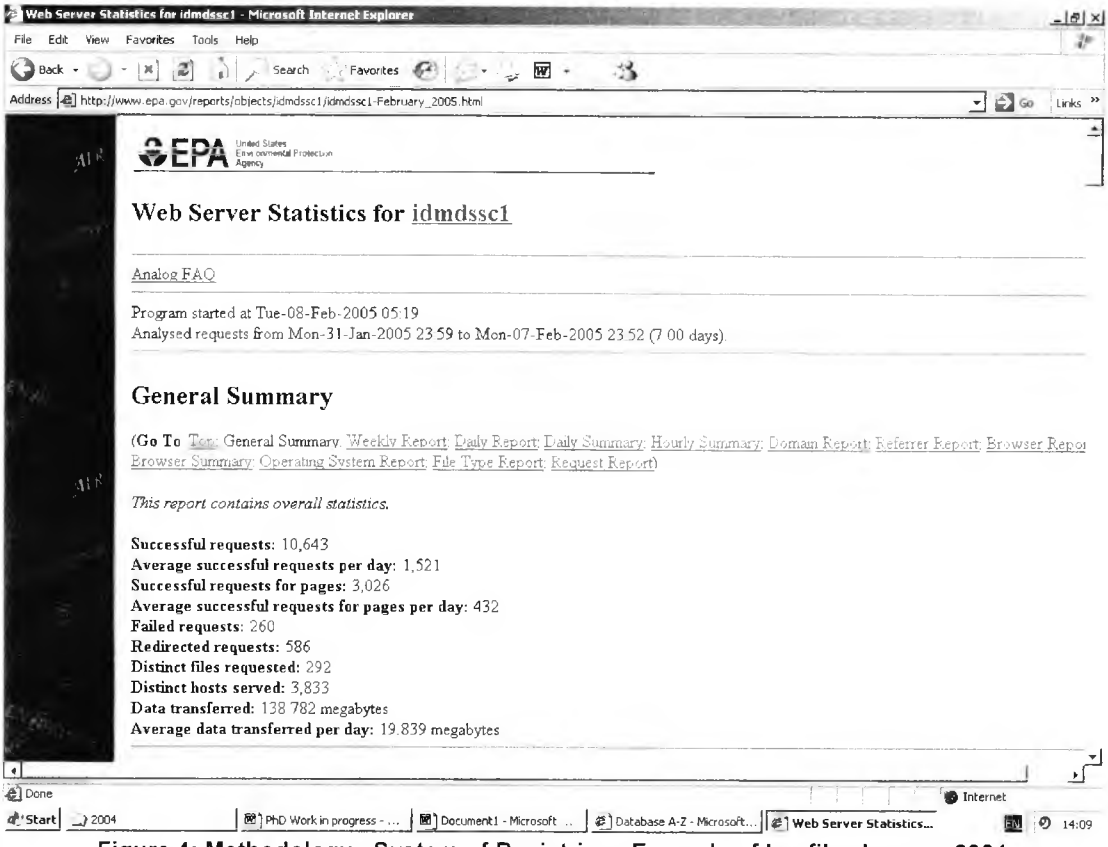


Figure 4: Methodology - System of Registries - Example of log file, January 2004

3.4.1.2 Metrics used

Although both metadata registry systems used the Analog software (http://www.analog.cx) to capture the users' interactions with the systems, each system produced different reports. The type of data in the provided reports imposed certain restrictions in the analyses that could be conducted. This subsequently resulted in having to perform additional calculations to the data in order to facilitate comparison between the two metadata registry systems. The metrics that were used and the calculations performed, where necessary, are presented in Table 2 and described in the sections hereafter.

Transaction loMetricsgs metrics	MetaForm	EDR/SoR
Daily summary	√	√
Increase/decrease	√	√
Domain	√	√
Failed	√	√
Directory	√	x
Requests	X	√
Referrer	X	√
Monthly	√	√
Yearly	√	√

Table 2: Methodology - Transaction Logs - Metrics used

3.4.1.2.1 Daily summary

In any time summary report, the Analog software gathers the total number of requests for pages in each time measure, e.g. day, hour, etc. over a period of time. The daily summary enables the assessment of which day or days during the week the metadata registry system is mostly accessed. The metric that Analog (version 5.22 and 3.2) uses to capture the daily use is that of *"successful requests for pages"*. Successful requests for pages reflect the number of actual file downloads so we can say that this metric provides with the representative daily use.

Regarding the MetaForm registry system, due to the type of web log transactions that were provided (data grouped over three date ranges) the daily summaries for each year were not available. In order to present the closest representation of the daily summaries it was essential to calculate an estimated analogy of the daily use over the period of 1998-2004 and then multiply it with the total number of page requests during each year. The estimated analogy of daily use was calculated by dividing the total of each day's page requests by the total of all days' page requests. Then, the daily summary during each year was produced by multiplying this average estimate of each day with the total number of page requests during each year. [15]

3.4.1.2.2 Increase/decrease in use

Both metadata registry systems included in their web log measurements the metric of *"average successful page requests per day"*. The metric of average successful page requests per day reflects the actual page downloads. Average successful page requests per day reflect the actual page downloads and facilitate the monitoring of the increase and decrease of the system's daily use. In the case of MetaForm in order to facilitate the comparison of the daily activity across months for the period of 1998-2004, the ten days of the November 2004 use, have been excluded from the counts in order to provide with a more accurate monthly representation. In the case of the Environmental Data Registry and the System of Registries, the metric of *"average* successful page requests per day" had been added to the web logs transactions since October 2001. This metric reflects the actual page downloads and facilitates the monitoring of the increase and decrease of SoR's daily use. Data were analysed for the period of October 2001 to October 2004 only in order to provide more accurate monthly representation of the daily use.

[15] The number of page requests for each month during every year was available in the web log files.

3.4.1.2.3 Domain requests

In the domain summaries, Analog lists the countries of the computers that requested files from the metadata registry systems. As the list with the countries can be extensive in presentation, it was decided to group them into continents and along with the domains from which the users requested files to facilitate the analysis of the domain requests. It was decided to use the MapMachine, the online atlas from National Geographic available at the following URL: http://plasma.nationalgeographic.com/mapmachine/ to facilitate this grouping. The analysis of the domain requests lists the following categories: Asia, European Union (EU), North America (NA), South America (SA), Australia and the Pacific Islands (OC). Furthermore the following domains were included:

- Commercial,
- Educational,
- EPA (for the case of the EDR/SoR),
- Governmental,
- Network,
- Unresolved and
- All other. All other includes domains that accounted for less than 10000 requests for pages such as governmental (.gov), international (.int), military (.mil), organisational (.org), accesses from South and North America, Australia, unknown and not listed domains.

As with the daily summary, the yearly accesses by domain for the MetaForm registry needed to be calculated. In order to provide the closest representation of the yearly domain accesses, it was essential to calculate an estimated analogy of the domain use over the period of 1998-2004 and multiply it with the total number of page requests during each year. The estimated analogy of the domains' use was calculated by dividing the total of each domain's page requests by the total of all domains' page requests. Then, the report for the domains accesses for each year was calculated by multiplying this average estimate of each domain with the total number of page requests during each year.

Finally, one of the disadvantages associated with web log transaction software is the problem of unresolved domains. Getting an unresolved domain message means that for some reason the Domain Name System (DNS) server that hosts the IP address for the computer, which accessed the metadata registry system, is unreachable or it does not have a Fully Qualified Domain Name (FQDN) associated in the Domain Name System (DNS). This data has been excluded from the analysis.

3.4.1.2.4 Failed requests

The failed requests reports list the number of files that caused errors. Analog defines failed requests as those *"...which return a web server status code in the 400s (error in request) or 500s (server error)"*. The most common reasons for a failed request are when the requested file is not found or is read-protected. This type of report is useful to system administrators as they can identify the files that are problematic and correct them. For the case of the MetaForm registry system, failed requests were counted against requests sent to the metadata server rather than the actual downloads. Therefore, as with the daily summary and the domain reports the monthly statistics the failed requests for MetaForm needed to be calculated. In order to provide the closest representation of the monthly failed requests over the period of 1998-2004 it was essential to calculate an estimated analogy of the monthly failed requests over the period of 1998-2004 and multiply it with the total number of page requests during each year. The estimated analogy of the failed requests was calculated by dividing the total number of requests for each year by the number of requests each month. Then, the monthly report of failed requests during each year was produced by multiplying this average estimate of each month with the total number of requests.

3.4.1.2.5 Directory requests

In the directory report the Analog software lists the directories from which the requested files were downloaded. In the case of MetaForm, the web log transactions provided did not list subdirectories individually, but the numbers for each directory included the accesses for all of its subdirectories. They also referred to the whole of the Metadata Server accesses. To facilitate the analysis, additionally to the main directories such as MetaGuide, MetaLib, main metadata server, similar directories were grouped together. For example, accesses to individual metadata projects were grouped together under the category projects. Category *"Other"* included graphics and/or diagrams, the glossary, and files of the new metadata server.

This data was not available for the Environmental Data Registry and the System of Registries. The requests report presents relevant usage data instead.

3.4.1.2.6 Requests report

The requests report records, which files, the SoR users had downloaded. Complementary to the directory report that records the directories where those files were stored, the requests report provides useful information to any system administrator as it

shows which parts of the website are mostly used. The launch of the SoR in spring 2002 replaced the Environmental Data Registry as the gate to the EPA's environmental information and standards. Although the EDR remained as one of the registries and therefore maintaining the majority of its files stored in one directory, the launch of the SoR introduced changes in the web site directories and file storage. This change obviously affected the web log statistics and posed implications to the analysis. In order to identify which registries of the SoR were mostly used it was decided to refer to the URLs structure and use the root to identify and group together downloaded files from the same registry. Each file (URL) was checked on a web browser to identify the page to what kind of resource it referred to. Files accessed before 2002 listing old or outdated URLs were the root of each URL was used as the basis to identify the registry to which it referred. Besides the main SoR registries identified such as the EDR, the CRS, the SRS, the TRS,

This data was not available for the MetaForm metadata registry system. The directory report presents relevant usage data instead.

3.4.1.2.7 Referrer reports (SoR)

The Analog software documentation (http://www.analog.cx) notes that the referrer report lists in the form of URLs *"...where people followed links from, or pages which included this site's images"*. In order to be able to present the kind of services and organisations that directed users to the SoR services, the referrer URLs were grouped under three categories. Those are explained hereafter:

- Referrers from the Environmental Protection Agency (EPA). This category included all URLs which their first part of the file path was www.epa.gov.
- Referrers from Search Engines (SE). This category included URLs that referred to major search engines such as Google, Yahoo, Lycos, Altavista, MSN, AskGeeves and AOL.
- Other. The category other covered those URLs that did not fall in the first two categories. It included links from governmental and commercial organisations, universities, professional bodies and World Wide Web directories.

This data was not available for the MetaForm metadata registry system.

3.4.1.2.8 Monthly use

The metric that the Analog software uses to capture the monthly use is that of successful requests for pages counting actual file downloads each month. Similarly to the daily use this report allows system administrators to identify the peak times of use and to associate with particular events that occurred during those times.

3.4.1.2.9 Yearly use

The Analog software uses the metric of successful page requests to record the yearly use. Complementary to the daily and monthly use, the yearly reports provide information about the trends of use in a given year and facilitate the comparison of use across years. This type of data can be useful to monitor how changes in a web site have affected its use over a given period.

3.4.2 Questionnaire surveys

The design of a questionnaire is a meticulous process that is very important for the quality of the data that is going to be processed and analysed. Oppenheim (1992) discussed how the questionnaire is more than a set of questions; in fact, it is an instrument that facilitates accurate data collection about a population sample by using scientific methods to measure attitudes, trends and use. Hague (1994), identified four purposes for questionnaires: to be designed in a manner that will obtain accurate information from the respondents, to have a predefined structure that is reflective of the research study aims, to be designed in a standard format on which facts, attitudes and trends can be recorded and measured and to facilitate data processing.

There are different types of questionnaires depending on the type of questions they entail and the aims they try to achieve. Questionnaires can be structured, semi-structured and unstructured. Structured questionnaires are mainly used to obtain response from large sample populations and their characteristic is that questions are in an order, they have predefined answers and users can rarely give answers outside of those predefined questions. Semi-structured questionnaires comprise of structured questions with predefined answers and open-ended questions that allow the respondents to express their opinions in free text fields. They are more flexible and they can be used to obtain both quantitative data (such as measurable variables) and qualitative data (free text fields). This type of approach was employed for this research study. Finally, unstructured questionnaires usually comprise of a checklist of questions and they are used to obtain purely qualitative data. They tend to be used for interviewing (Hague, 1994).

Furthermore, the structuring and layout of the questionnaire is a major significance. Questionnaire surveys rely on the individual's comprehension of the written text so questions should be formed in a way that will enable the recipients to fully understand their aims. It should have a title, it should provide introductory information about the study and the intentional use of the collected data, and it should denote the code of conduct for the collection and process of data and provide the contact details of the research for further clarifications if those are needed. It should be designed in a way that it will enable easy and quick process of data. Furthermore, the length and the time that they require for completion can affect the response rate.

Questionnaire surveys can be conducted in various manners. They can take the form of paper based surveys, delivered and collected via postal services. They can also be electronic and attached in an email send to the recipient or even to discussion and/or mailing lists. Alternatively the questions can be embedded in an email or in an electronic

form accessible via a Uniform Resource Locator (URL). Although during the middle and late 1990s electronic surveys took the form of attached files to emails today this option is advised to be avoided as the generation of viruses show email attachments as alerts rather that research calls. It is frequently the case that questionnaires are designed with the aid of a software program and they can be made available in a web form. It is common that instead of email attachments a link to a URL is provided where users could complete the online questionnaire and by clicking the submit button at the end of the questionnaire the data is transferred to the server that hosts the URL. The advantages and disadvantages of online and offline studies are discussed earlier in this chapter.

The use of email and the World Wide Web to conduct questionnaire surveys prerequisites their use by the sample population. It is assumed that the target group is computer literate and familiar with the use of email programs and web browsers. Therefore this approach excludes users of systems that they do not have access to email and the World Wide Web. The very nature of metadata registry systems, being based on an open access initiative and only accessible online does not probe this limitation as their users would have to be computer literate and have access to a computer and an Internet connection in order to be able to find them in the first place.

The questionnaire surveys for this research study were conducted between 2000 and 2003. The number of questionnaire surveys conducted were five, two of them online (Table 3).

Studies	Online Questionnaire	Paper based Questionnaire
Discussion lists	√	
SCHEMAS workshop		√
MetaForm metadata registry system	√	
EDR metadata registry system		√
SoR metadata registry system	√	

Table 3: Methodology - Questionnaire surveys by type

3.4.2.1 Sampling and recruiting

The sampling and recruitment processes for the questionnaire surveys conducted with attendees to the 2nd SCHEMAS workshop in Bonn, subscribers to several discussion lists, users of the MetaForm registry system, the Environmental Data Registry and the System or Registries are presented in the following sections.

SCHEMAS workshop and discussion lists surveys

The idea for conducting a questionnaire survey arose when the opportunity presented itself in the form of the 2nd SCHEMAS workshop in Bonn, Germany (http://www.schemas-forum.org/workshops/). The 2nd SCHEMAS workshop was conducted as part of the European Union funded SCHEMAS project. Its aim was to inform people involved in metadata implementations about evolving metadata standards by providing guidelines and training material on their use. Also, to investigate how the use of metadata registry systems could support those aims. The target group for this project and therefore at an extent the attendees of the workshop were people and institutions *"that [were] participating in projects under the IST Programme and in related initiatives on a national level in Europe"* [16] pertaining to the development and maintenance of metadata standards. In parallel with the 2nd SCHEMAS workshop the questionnaire was posted to four discussion lists (Appendix 3) between the period of the 12th and the 15th of November 2000. It aimed to complement results from the SCHEMAS workshop by addressing an international audience from different scientific communities. The lists were chosen on the criteria of relevance to metadata and/or digital libraries research, the number of the members and the worldwide coverage of the lists. Information regarding the lists can be found in Appendix 3.

In parallel with the SCHEMAS workshop the same questionnaire was posted to four discussion lists between the period of the 12th and the 15th of November 2000. It aimed to complement results from the SCHEMAS workshop by addressing an international audience from different scientific communities. A two weeks period was provided to respondents for completing and emailing the questionnaire back to the researcher. The lists were chosen on the criteria of relevance to metadata and/or digital libraries research, the number of the members and the worldwide coverage of the lists. The four (4) chosen lists were given hereafter:

- ***11179 Metadata Registries Coalition Discussion list.*** A United States of America based list which its members were people interested in the implementation of ISO/IEC 11179 metadata registries. The list could be found at: http://hmrha.hirs.osd.mil/mrc/ (Last accessed 20/01/2002). It was formed in 1998 and it aimed to *"provide a forum and source of practical experience for mutual cooperation among organizations that are introducing metadata registries into their information systems asset base in order to be able to manage the semantics of the data elements in their databases."* Members included people from diverse communities, mainly from US federal agencies, with an interest in ISO/IEC 11179 standard implementation approaches, developments and support. The list moderator was contacted regarding

[16] SCHEMAS project objectives. Available at: http://www.schemas-forum.org/project-info/objectives.htm#purpose (Last accessed: 12/06/2006)

the number of subscribers to the list but a response failed to reach the researcher. It is believed that at the time of the survey the number of subscribers were around 30 people.

- ***Diglib.*** A European Union based list that was hosted by the International Federation for Library Association. Discussion on this list covered issues and technology pertaining to digital libraries research. The list can be found at: http://www.ifla.org/II/lists/diglib.htm (last accessed 27/11/2005). It was formed in 1995 and it is *"...a mailing list is for librarians, information scientists, and other information professionals to share information about the many issues and technologies pertaining to the creation of "digital libraries".* Members included both individuals and organisations *"...from around the world who [were] creating or providing electronic access to digital collections to participate in knowledge sharing about current developments in digital library research."* The list moderator was contacted regarding the number of subscribers to the list but a response failed to reach the researcher.
- ***Interoperability.*** A United Kingdom JISCMAIL list that hosted discussion about metadata and interoperability. The list can be found at: http://www.jiscmail.ac.uk/lists/interoperability.html (last accessed 27/11/2005). It was formed in January 1999 and aimed to address issues such as*"...metadata, distributed library systems and public library networking. Interoperability Focus also has a special interest in moving beyond the library sphere, specifically encompassing museums, archives, and other aspects of the cultural heritage, as well as Government and community information." Members included "Projects and individuals with experiences to share, working implementations to show, or core issues to resolve which might usefully be addressed by Interoperability Focus...."* The interoperability list counted 385 members on the 14/11/2000.
- ***UK-Meg.*** A United Kingdom JISCMAIL list that hosted discussion about metadata for education. The list can be found at: http://www.jiscmail.ac.uk/lists/uk-meg.html (last accessed 27/11/2005). Formed in June 2000, it aimed to *"...support the work of the Metadata for Education Group (MEG), which seeks common approaches to the description and exchange of educational content across all levels of the UK's educational system..." Members included "...a number of the current players in this field, drawn from primary, secondary, tertiary and continuing education, as well as from relevant standards initiatives and the museum and library sectors, which have valuable content to offer."* The UK-Meg list counted 90 members on the 14/11/2000.

MetaForm questionnaire survey

The discussion lists surveys and SCHEMAS workshop survey provided useful information about trends in metadata research and the use of metadata element sets during 2000. It was decided that the next step for the study would be to build upon those findings by approaching current implementations of metadata registry systems and survey their users regarding their use and functionality. A current implementation at the time of study was the MetaForm registry system (http://www2.sub.uni-goettingen.de/metaform/index.html), based at the State and University Library (SUB) Göttingen, Germany and funded by the German Research Foundation. A questionnaire was designed and made available online at the MetaForm registry's website. The questionnaire survey run under the name of MetaQuest and was launched onto the MetaForm's website on the 15th of March 2001. A two weeks period was provided to the MetaForm users to participate to the survey. Due to an initial slow response it was decided to extend the completion period to a month. Furthermore, to increase interest in the survey and boost response, several discussion and mailing lists in Germany were contacted. The lists were provided by the liaison contact in MetaForm based on the criterion of relevance to metadata. Information about the lists can be found in Appendix 3.

EDR questionnaire survey

The opportunity to survey the EDR users arose in the light of the 1st users' conference, which was held at the U.S. Bureau of Labor Statistics' Conference Center on January 24, 2002 in Washington, DC. Following the exchange of several emails and telephone discussions explaining the purpose of the survey and the intentional use of the survey data, the organisers granted the permission to conduct the survey at the users' conference. It was agreed that the questionnaire would be distributed to the conference delegates during the hands-on session of the conference and they would be advised by the organisers to complete it and return it to the registration point of the conference before they leave the venue. A letter providing the contact details of the researcher, explaining the purpose of this survey and guarantying confidentiality of the respondents, preceded the questionnaire. Furthermore, an additional note that the EPA held no responsibility for the content of the questionnaire and the participation to the survey was at the conference attendees' own will, was also included at the request of the organisers.

The users' conference was publicised among the registered parties through the Standard Update Newsletter and on the EDR's website for a period of three weeks between the 1st and 23rd of January, 2002. The intended audience of the users' conference were the registered parties to the EPA's Standard Update Newsletter. Although confidentiality reasons prevented EPA staff from disclosing information about

the scientific background and domains that they represented they were all described as people involved or with an interest in the implementation of metadata standards. Some examples include people with interest in data harmonisation, data standardisation and implementation, systems developers, metadata managers. The number of the conference delegates was restricted to 50 people due to restrictions imposed by the venue's size.

SoR questionnaire survey

Following the success of the 1st users' conference in 2002, the Office for the Environmental Information (OEI) at the EPA repeated the event in March 2003. The users' conference was held at the National Center for Supercomputing Applications (NCSA) in Virginia on the 20th of March, 2003. Similarly to the previous conference the event included presentations of the several registry applications and hands on sessions. The 2003 SoR conference provided the opportunity to follow up on last year's findings. Similarly to the previous conference the organisers granted the permission to conduct the survey. A letter providing the contact details of the researcher, explaining the purpose of this survey and guarantying confidentiality of the respondents, preceded the questionnaire. Furthermore, an additional note that the EPA held no responsibility for the content of the questionnaire and the participation to the survey was at the conference attendees' own will, was also included at the request of the organisers. Differently to the previous survey, it was decided to administer the questionnaire online in order to facilitate the quicker processing of the results and to fit in with the conference's context. The questionnaire was made available online on the day of the conference and it was agreed that the organisers would ask the SoR users to complete the questionnaire at the end of the hands on session and submit the results directly to the researcher.

The users' conference was publicised on the SoR's website. The intended audience of the users' conference were the registered parties to the EPA's Standard Update Newsletter. Although confidentiality reasons prevented EPA staff from disclosing information about the scientific background and domains that they represented they were all described as people involved or with an interest in the implementation of metadata standards. Some examples include people with interest in data harmonisation, data standardisation and implementation, systems developers, metadata managers. Although sixty-five people attended the conference, only forty-nine could have completed the questionnaire due to the limited access to the computers in the venue.

3.4.2.2 Questionnaire questions

The design of all the questionnaire surveys followed the same principles and similar structure. All of the questionnaires comprised of three or four sections. The first section aimed to identify demographic or other user characteristics such as scientific domain they represented and post or position they held. The second section of the questionnaire usually included questions relevant to the use of metadata, metadata applications and metadata registry systems. The third part of the questionnaire usually aimed to assess the functionality of metadata registry systems and to identify the expectations of the users for such systems. The last section usually invited respondents to provide any feedback or additional comments that were not addressed by any of previous sections. Questions grouped by section and addressed in all questionnaire surveys are presented in Table 4 and described in the sections that follow.

	E- lists	CHEMAS	MetaForm	EDR	SoR	Interviews
User characteristics						
Age, sex, domain they represent and/or position held	√	√	√	√	√	√
Familiarity with electronic systems (including resources, databases, etc.)			√	√		
Frequency of using electronic systems (including resources, databases, etc.)			√	√	√	
Difficulties encountered during the search process				√	√	√
Use						
Use of metadata	√	√	√			
Use of metadata registries	√	√	√			√
Registry systems that they have used/developed/are developing	√	√	√	√	√	√
Assessment of usefulness			√	√		
Reasons for using the registry			√	√		√
Metadata applications	√	√	√	√	√	
Frequency of use			√	√		
Longevity of use			√	√	√	
Future use			√	√	√	√
Organisation/Institute requirement/role	√	√	√	√		√
Functionality						
Services/Features that they are looking for in a registry system						√
Services/Features that every registry should have						√
Services/Features used			√	√	√	√
Rank services/features used			√	√		
Content/coverage of element sets			√	√		
Additions/Improvements			√	√		√
Describe search				√		
Strongest features			√	√		√
Weakest features			√	√		√
Rank info retrieved				√		
Barriers in development/implementation/use of registry systems						√
Support services						
Comments			√	√	√	√
Why use registry systems in the future						√
Metadata registry systems definition						√
How can metadata registry systems support your work				√	√	

Table 4: Methodology - Range of questions used in the questionnaire surveys and interviews

3.4.2.2.1 Discussion Lists and SCHEMAS workshop surveys

Both the discussion lists and the SCHEMAS workshop questionnaires consisted of three parts. The first part aimed to gather information regarding demographic characteristics of the target population, the second part aimed to find out about metadata element sets use among the respondents and the third part aimed to gather information about metadata registry systems use among the respondents. The areas that each part dealt with are described hereafter:

Demographic characteristics (Part 1, questions 1-4)

The respondents were invited to provide information regarding their age, sex, scientific domain they represented and the post at their current employment. They were provided with age range options from 17 to 64+ years old. The domains were drawn based on the SCHEMAS first metadata watch report and included the options of academic, audio-visual, cultural heritage, educational, geographic information systems, industrial, publishing, research and other. Regarding the post at their current employment the respondents were given the options of computer scientist, information scientist, researcher, consultant and other. Category other in both questions invited respondents to provide any additional information that was not included in the options provided.

Use of metadata element sets (Part 2, questions 1-5)

Respondents were asked to denote whether they had used a metadata element set in the past. They were asked to indicate from a list of terms which of those they had associated with metadata. The terms were: element sets, formats, schemas, standards, systems, catalogues and other. As certain terms had been associated with specific scientific domains it was hoped that response to this question would enable a cross tabulation of the results with questions from part 1 of the questionnaire. Unfortunately, due to an error of the SCHEMAS workshop's organisers, a previous version of the questionnaire was distributed to the workshop's participants that did not include that question. All respondents were further asked to select from a list of metadata element sets those that they were aware of and those that they had used. The options included the Dublin Core metadata element set, the IEEE LOM, MARC, IMS, GILS and other. Category other invited the respondents to provide additional information about any metadata element set that was not included in the options provided.

Use of metadata registry systems (Part 3, questions 1-5)

Similarly to the section for metadata element sets, respondents were invited to denote whether they had used a metadata registry system in the past. All the respondents were asked to select from a list of metadata registry systems those that they were aware of

and those that they had used. The options included the indecs registry project, the DESIRE metadata registry system, the ROADS templates, the MetaForm metadata registry system, the Environmental Data Registry, the Basic Semantics Register, the Knowledgebase metadata registry system and other. Category other invited the respondents to provide additional information about any metadata registry system that was not included in the options provided.

Reasons for use and applications (Part 2, question 5; part 3, questions 5-6)

Respondents were invited to specify the reasons for which they had used metadata element sets and metadata registry systems. They were provided with a list of potential reasons. This list of reasons for using metadata element sets included the following options: to use for/on my personal web page, to user for/on my organisation's web page, to use for library resources, to use for/on the project that I am currently working on and other. The list of reasons for using metadata registry systems included the following options: for resource discovery/find out about relevant information, for data exchange, to find definitions of elements, for the description of text resources, for mapping of elements, to see dictionary structures, and other. Furthermore for metadata registry systems in particular, respondents were asked to indicate if their organisation had a requirement for using such systems.

Future use (Part 2, question 7; part 3, question 7)

In order to have even a small indicator of the respondents intentional use of metadata element sets and registry systems, they were asked to report whether they were interested in finding out more about both in the future. Respondents who gave a negative answer were invited to specify their reasons.

3.4.2.2.2 MetaForm questionnaire survey

The MetaForm questionnaire consisted of three parts. The first part aimed to draw some demographic characteristics of the MetaForm users. The second part aimed to assess the MetaForm's use and the third part of the questionnaire aimed to gather information about MetaForm's functionality. The areas that each part dealt with are described hereafter:

Demographic characteristics (Part 1, questions 1-4)

The respondents were invited to provide information regarding their age, sex, scientific domain they represented and the post at their current employment. They were provided with age range options from 17 to 64+ years old. The grouping of the domains was based

on the SCHEMAS First Metadata Watch Report (May 2000) but different to our previous surveys it provided examples of what areas each domain represented. For example, next to the Cultural Heritage there was the explanation *"(e.g., Libraries, Archives, Museums)"*. The list of posts was adjusted to the findings of the SCHEMAS workshop and Discussion Lists surveys and now listed computer scientists, consultants, information scientists, researchers and other. Category other invited the MetaForm users to provide any additional information that was not included in the options provided.

Use of the MetaForm registry system (Part 2, questions1-5)

The respondents were invited to provide information regarding their use of the MetaForm registry. Questions aimed to find out about the frequency and longevity of the MetaForm's use and whether the respondents' organisation had a requirement for employing a metadata registry. Users were asked to choose from a frequency of use by week, month, occasionally, hardly ever, first time users and of a period of use of one month, three months, six months or more than that. Those who selected the option more than six months were asked to specify for how long they had been using the MetaForm registry. Furthermore, those who noted that their organisation had a requirement for a metadata registry system where asked to elaborate in what way. It was hoped that response would provide a small indicator how organisations envisioned the harmonisation of such a system in their operations. Options included whether the organisation was a contributor to the design, implementation, population and maintenance of the registry or simply a user.

Reasons for use and applications (Part 2, questions 6-7)

Respondents were invited to specify the reasons for which they had used the MetaForm registry system. Additionally, this section of the questionnaire aimed to identify the reasons why the respondents used MetaForm and how they had used the information they retrieved in order to identify some indication of what the applications of metadata registry system could be. They were provided with a list of potential reasons for use that included the following options: for resource discovery/find out about relevant information, for data exchange, to find definitions of elements, for the description of text resources, for mapping of elements, to see dictionary structures, and other. The list of potential implementations of the retrieved information included the following options: to use for/on my personal web page, to user for/on my organisation's web page, to use for library resources, to use for/on the project that I am currently working on and other. Those who had selected the option other were invited to specify how they had used the information they had retrieved from the MetaForm.

Use of other metadata registry systems (Part 2, questions 8-10)

The MetaForm registry users were asked to provide information about other metadata registry systems that they had used and to denote those that they had found more useful. The list of metadata registry systems was adjusted to the findings of the SCHEMAS workshop and Discussion Lists surveys and now included one more metadata registry system, the Dublin Core.

Functionality of the MetaForm use (Part 3, questions 1-5)

The respondents were invited to provide information regarding the funtionality of the MetaForm registry's services. The respondents were asked to indicate which of the three MetaForm's features, crosscuts, crossroads and mappings, they had used and to express their preference of them. They were provided a scale of 1 to 5, where 1 denoted the least satisfactory service and 5 the most satisfactory service. Similarly to the previous question, they were also asked to indicate their satisfaction or not towards the MetaForm's coverage of metadata element sets and application profiles and to rate the functionality within services using the same scale of 1 to 5. The functions that were asked to assess were: breadth of information, coverage of metadata element sets, linking of different metadata element sets, mappings of metadata element sets, relevance of retrieved information and availability of retrieving information in different formats.

Future services (Part 3, questions 6-7)

Respondents were asked if they would welcome additions in the MetaForm services and those that replied that they would were asked to prioritise from a list of possible future additions. Those are presented in the following table:

Please select additional features that you would like MetaForm to add:
Advance linking facility
Advance mapping facilities
Automatic translation of metadata formats into other languages
Automatic validation of information
Larger coverage of information resources
Metadata conversion Tools
Metadata creation Tools
Metadata formats in several languages
Search facilities (index, browse, keyword searching)
Vocabulary facility
More metadata formats
Other (please specify)

Table 5: Methodology - MetaForm questionnaire – List of future services

Comments (Part 3, question 8)

Respondents were invited to provide any additional comments regarding the use and functionality of the MetaForm registry system that were not address in the previous questions.

3.4.2.2.3 Environmental Data Registry

Similarly to the previous questionnaires, the Environmental Data Registry questionnaire also consisted of three parts. The first part aimed to gather information regarding demographic characteristics of the target population. The second part aimed to assess the Environmental Data Registry's use and the third part of the questionnaire aimed to gather information about Environmental Data Registry's functionality. The areas that each part dealt with are described hereafter:

User characteristics (Part 1, questions 1-5)

The respondents were invited to provide information regarding their field of interest. They were provided with a list of options that was reflecting the areas of metadata registry research as this appeared on the Environmental Data Registry's website. Those were: data harmonisation, data standardisation and implementation, systems development, all of the previous categories and other. Respondents were also asked about their familiarity with electronic resources (including use of the Internet, electronic journals and online databases) and to denote their level of computer literacy by identifying themselves as advanced, novice or beginner users. Finally, in order to identify potential search behaviour the EDR users were asked to denote how easy or difficult they have encountered the process of obtaining useful information in the past. All respondents were provided with a list of difficulties that could have encountered during the search process and were asked to select those if any that they had experienced. The list is presented in the following table:

Difficulties during the searching process
Badly designed sites that I find difficult to navigate
Cost of services/information to obtain
Lack of guidance tools
Lack of online help facilities
Lack of supportive services
Lack of time required for the search
Other (please specify)
Too much information
Too much unorganised information
Unfamiliar with searching methods

Table 6: Methodology - Environmental Data Registry questionnaire - List of difficulties encountered during the searching process

Use of the Environmental Data Registry (Part 2, questions 1-2, 7)

Similarly to the MetaForm questionnaire, the respondents were invited to provide information regarding their use of the Environmental Data Registry. Questions aimed to find out about the frequency and longevity of the Environmental Data Registry's use and whether the respondents' organisation had a requirement for employing a metadata registry. Users were asked to choose from a frequency of use by week, month, occasionally, hardly ever, first time users and of a period of use of one month, three months, six months or more than that. Those who selected the option more than six months were asked to specify for how long they had been using the Environmental Data Registry. Furthermore, those who noted that their organisation had a requirement for a metadata registry system where asked to elaborate in what way. It was hoped that response would provide a small indicator how organisations envisioned the harmonisation of such a system in their operations.

Functionality of the Environmental Data Registry (Part 2, questions 3-4; part 3, questions 1-3)

The respondents were invited to select from a list of services that the Environmental Data Registry provided and assess how useful they considered them to be using a scale of 1 to 5, where 1 was poor and 5 was excellent. The list of the provided services is presented in the following table:

Usefulness of services and products	1	2	3	4	5	n/a
Comments						
Download						
FAQ						
Glossary						
Help						
How to...facility						
MetaPro						
Newsletter						
Registration Facility						
Search Facility						
Subscription Service						
Thesaurus						
What's new...link						
Workshops/Conferences information and announcements						
XML						

Table 7: Methodology - Environmental Data Registry questionnaire – List of services

Furthermore, the respondents were asked to describe how easy or not they had encountered the search process in the EDR selecting from three options. Those were: it was easy and straightforward, I found what I wanted but I would appreciate more guidance, I have not been able to find what I needed. A field where they could expand on any of those options was provided for any of the respondents that wished to elaborate

with comments. Additionally, they were also asked to assess the relevance of the retrieved information.

Type of information sought (Part 2, questions 5-6)

The EDR users were asked to denote their satisfaction about the breadth of information resources covered by the EDR and to indicate what type of information that they were looking for in the EDR. They were provided the following options: data standards, data elements, business rules documents, environmental related organisations, metadata registry architecture information, definitions of elements and other. Other invited users to provide with any other type of information that they were interested in and was not included in the list of options provided.

Applications of the retrieved information (Part 2, questions 7-8)

The respondents were invited to denote where they had used the information they had retrieved from the Environmental Data Registry. They were provided with the following options: I have been testing metadata registry systems and how they work; it is within my organisational commitments, for my own research and other. Other invited users to provide with any additional information about metadata implementation that was not included in the list of options provided. They had also been asked whether their organisation (in this case, the Environmental Protection Agency) had a requirement for a metadata registry system and if so to describe in what way.

Support services and future additions (Part 3, questions 3-4)

The EDR users were asked to assess the help and site map facilities and to indicate their satisfaction or dissatisfaction for the support services that the Environmental Data Registry provided to during the search process. They were provided with a scale of 1 to 3 where 1 was not satisfactory and 3 satisfactory. Furthermore, they were asked to indicate from a list of planned future additions those that they would mostly welcome. Those were: improved online harmonization, XML, MetaPro and the facility of multiple files download.

3.4.2.2.4 System of Registries

Differently to the previous questionnaires, the System of Registries questionnaire consisted of two parts. The first part aimed to gather information regarding some demographic characteristics of the target population, and about the use and functionality of the System of Registries. The second part invited the respondents to indicate what they considered the weakest and strongest features of the SoR and to suggest how the

SoR could support their work. The areas that each part dealt with are described hereafter:

User characteristics (Part 1, questions 1, 4)

The respondents were invited to provide information regarding their field of interest. They were provided a list of options that was adjusted to findings from the Environmental Data Registry questionnaire survey. The list of options included: Data standards developer, data standards implementer, exchange network participant, registry developer, someone interested in substance identification, systems developer, and other. Category other invited respondents to provide any additional information that was not included in the list of options provided. Furthermore, in order to identify first time users and those who had used the SoR in the past, respondents were asked to select from the following options those that applied in their case: I have searched for information on SoR, I have downloaded information that I have found on SoR, I have applied download information from SoR to my work, I have used the service to register an information resource, and none of the above, I am here today to find out more on SoR.

Use of the System of Registries (Part 1, question 2)

The respondents were invited to select from the SoR components those that they had used in the past. The list with the System or Registries components included the following: EDR (Environmental Data Registry), EIMS (Environmental Information Management System), EMG (Environmental Metadata Gateway), FRS (Facility Registry System), IRRS (Information Resources Registry System), SoR quick search, SRS (Substance Registry System), TRS (Terminology Registry System) and the option None of the above, it's my first time using the SoR. The question aimed to identify first time users and to have some indication about the use by particular type of information. In cross tabulation with questions 1 and 4 it was expected to obtain some indication about what information SoR users were interested in and how they applied it, if so, in their work.

Type of information sought (Part 1, question 3)

The SoR users were asked to denote their satisfaction about the breadth of information resources covered by the EDR and to indicate what type of information that they were looking for in the EDR. They were provided the following options: business rules documents, code sets, data elements, data standards, facilities, organisations, regulations, resources in general (IRRS material), substances (chemical/biological), environmental terminology, XML tags and other. Other invited users to provide with any other type of information that they were interested in and was not included in the list of options provided. The list of different types of information was expanded compared to

that of the EDR questionnaire in order to reflect the changes in the development of the System of Registries.

Applications of the retrieved information (Part 1, question 5)

The respondents were invited to denote where they had used the information they had retrieved from the System of Registries. They were provided with the following options: to keep up to date with information resources within the EPA, to select and/or review information resources (e.g., data sets, databases, code sets, etc) that my office must comply with, in my work with data standards (search for, review, amend), I wanted to download multiple data standards information at once, and for other reasons. Those who had indicated other reasons were invited to provide any additional information about metadata implementation that was not included in the list of options provided.

Strongest and weakest features of the System of Registries and comments and feedback (Part 2, questions 1-3)

The respondents were invited to denote the SoR features that they considered the weakest and those that they considered the strongest. Furthermore they were invited to indicate how the System of Registries could support their everyday work. Finally, they were provided with free text space to make any additional comments or expand on any of the answers given. They were also asked whether they would like to contribute further to this study.

3.4.3 Email interviews

Interviewing can be conducted in various modes. Examples include, face to face, focus groups interviews, computer-mediated interviews, survey interviewing, qualitative and in depth interviewing. Deciding which the most appropriate interviewing method is has been influenced by several factors that describe metadata registry systems. Those were their geographic disperse, their accessibility by almost everyone around the world and the financial constrains that telephone or face to face interviews would impose on the researcher. The mode chosen was that of email interviewing.

3.4.3.1 Sampling and recruiting

The process of recruiting the interviewees has been ongoing throughout the research study. Recruiting interviewees was conducted in parallel with the traditional and online questionnaires the invitations to participate to the email interviews targeted three groups of people:

- Individuals that had participated in previous surveys conducted, such as the questionnaire surveys, and expressed interest in participating further to the research,
- Subscribers to selected discussion lists of relevance to metadata registries research and
- Individuals identified in the literature as people that were primarily involved in the development, implementation and/or use of metadata registry systems. Furthermore, the emails were sent to attendees of metadata registries conferences, such as the Open Forum on Metadata Registries.

The email invitations were sent out to the first group during October 2004, gradually from the 5th until the end of October 2004. The recipients of the emails were presented with two weeks to respond. A further week's time was given to those people who were away from their desk for reasons such as leave or attending conferences and therefore unavailable to respond straight away. Despite the provision of an additional week, the response from this initial posting was not the anticipated. Among the problems encountered were mail postage failures due to wrong, expired and outdated email addresses.

During the same time, October 2004 and in order to boost the response rate it was decided to post the invitation to more discussion lists. The following discussion lists were contacted DC-ANZ, DC – Registry, ALIA, USHIK and xml4lib. DC-ANZ and ALIA were suggested from a respondent as representative to reach the Australian audience that was

engaged in metadata research. The DC-Registry list was set up to address research needs of people engaged in development, implementation and use of the Dublin Core metadata registry. The United States Health Information Knowledgebase (USHIK) discussion list is an internal list for those developing, implementing and using the USHIK metadata registry. Finally, the xml4lib list aimed to reach audiences that use the Extensible Markup Language in, by and for libraries.

Additionally to the discussion lists' posting, the attendees of the Open Forums on Metadata Registries (2000 and 2003) were contacted. The attendees' names of the Open Forums on Metadata Registries were publicised on the web sites of the Forums. In order to find their contact addresses several searches were conducted. Mostly the searching required a combination of the name and the institution or organisation that each person represented and/or visiting their organisational web sites and contacting the administrators. Duplications of names were excluded and each individual was contacted on a personal basis. The email invitations that were sent explained how their contact details were found, the aims of the email interview and provided the contact details of the researcher should the recipients had any questions regarding the research. During May 2005 the same process was followed for the participants of the Open Forum on Metadata Registries 2005. The interview questions were slightly amended to address expectations of future use rather than current use and emailed the 23 participants of the Open Forum 2005 that we had not contacted before. Again, response was requested within two weeks' time and the extension of a further week was provided to those who requested it.

During November 2004 the last targeted group was contacted. Those were the individuals that have been identified in the literature as people primarily involved in the area of metadata registries research. Those included people from the DELOS network, SoR, UKOLN and MetaForm. The same explanations regarding how their contact details were obtained and the aims of the research were provided and the same time was given for the recipients of the emails to respond. Please see Table 8 for a list of the discussion lists and other contacts and the period over which the email invitations were sent.

	October 2004	November 2004	May 2005
Alia – catlibs	√		
DC-anz	√		
DC-Registry	√		
Individuals (DELOS, MetaForm, SoR, Questionnaire respondents, UKOLN)	√	√	
Open Forum on Metadata Registries	√		√
USHIK	√		
Xml4lib	√		

Table 8: Methodology - Email Interviews – List of discussion lists and contacts

Although the use of additional methods that facilitated synchronous communication, such as the conduction of online focus groups was considered, that approach was not adopted in the end as it required that several of the interviewees would have to dedicate time in familiarising themselves with the relevant technologies and software. Furthermore, it also required that interviewees that were located in different places around the world would be available at a given day and time which proved not to be possible.

3.4.3.2 Interviews questions

Thirty seven (37) people were interviewed via email. The estimated time required to respond to the interview questions was 30 minutes. The email invitation that was sent explained the purpose of the research study and the intended use of the responses. It also provided the researcher's contact details in case the interviewees wished to discuss the study. The questions that the interviewees were invited to answer were all open ended and were accompanied with examples in all instances. On some occasions the researcher exchanged more than one email with the interviewees asking them to expand or elaborate on particular answers they provided. In particular they concentrated on the following areas:

Job/Post responsibilities and features/services that the interviewees were interested in a registry. Interviewees were asked to provide information about their scientific background, their current post responsibilities relevant to metadata registries research and indicate the features and/or services they were interested in a metadata registry system deriving directly from those. The intention of this question was to find out about the interviewees field of interest and occupation were associated and/or interlinked with their metadata research activities and use of metadata registry systems.

Metadata registry systems that the interviewees had used, developed or were developing at the time of the email interview. Interviewees were asked to note the metadata registry systems that they had used, developed or were currently in development at the time of the email interview. It was hoped that response to this question would provide information regarding metadata registry system implementations that were used or under development and list the breadth of such applications if that was the case.

Reasons for use and perceived value of metadata registry systems. Interviewees were invited to specify the reasons for which they had used metadata registry systems in the past but also to denote why they thought that metadata registries should be used, if they believed that was the case. This was an open ended question and interviewees

were invited to respond based on their own experiences of metadata registry systems use and future expectations.

Essential features of metadata registry systems. Interviewees were asked to list the three main features that they considered essential for any metadata registry system. They were provided with examples but this was again an open ended question. It was hoped that this question would provide with what the interviewees considered as essential functional requirements in a metadata registry system.

Secondary features of metadata registry systems. Interviewees were asked to list the features that they considered to be of secondary importance to them in metadata registry systems. As with the previous question they were provided examples but this was also an open ended question. It was hoped that complementary to essential functional requirements, feedback from the interviewees would give some indications of what the future expectations or possible overstated features of metadata registry systems were.

Barriers of use. Interviewees were asked to denote any barriers they had experienced in the use, development or implementation of metadata registry systems, depending of the degree of their involvement in any of those processes. The aim of this question was to point out any drawbacks in the use and development of such systems that could be useful to those engaged in the development and management of metadata registry systems.

Comments. The interviewees were asked to comment on the following two statements:

> *"Metadata registries provide the effective solution to the interoperability problem on the web"*
>
> *"Metadata registries can be used everywhere, from governmental and standardisation organisations to libraries and web based search engines"*

The aim of the invitation to commentaries on those two statements was to gather the interviewees' views on the role of metadata registry system in the domain of data management and whether they were considered as an effective means to address interoperability problems of networked information and their potential applications. Interviewees were also invited to make any additional comments that they wished for regarding metadata registry systems.

Chapter 4 – Results

This chapter presents the results about the use and functionality of two active metadata registry systems, the MetaForm registry, and the Environmental Data Registry/System of Registries, as these had been recorded by web log transactions of the metadata registry systems. The web log transactions data was complemented with findings from four questionnaire surveys and email interviews with users of the metadata registry systems. The chapter is divided into four parts that are preceded by the presentation of the response rates that each method attracted. ***Part A - Characteristics of the users*** presents the response pertaining to the demographic and other characteristics of the respondents such as age, sex, domain representation, their occupation and field of interest. Additionally, information about the respondents' familiarity with the use of online resources in general, as well as the frequency and longevity of their use of the MetaForm and the EDR/SoR are also discussed. Furthermore, data from the web log transactions that recorded access by domains and the referrers to the metadata registry systems was also included to provide a more comprehensive profile of the metadata registry systems users. ***Part B - Metadata registry systems use*** presents results referring to the respondents' understanding and use of metadata in general and the use of metadata registry systems in particular. Use of the MetaForm registry and the EDR/SoR is presented by day of the week, month and year and over the six year period of web log data that was analysed. Additionally to this, users' access to specific directories and/or types of information as well as particular features and/or services of the metadata system registries. This part also discusses how an organisation's requirement for a metadata registry system can or cannot increase the demand for such services and finally, presents current implementations of metadata registry systems as the respondents denoted those. ***Part C - Metadata registry systems functionality*** discusses the features of metadata registry systems that the respondents considered essential in such systems, their preference and assessment over their usefulness, and the barriers that prevented them from using them. This section discusses non use and complements findings from the respondents' feedback with results from web log transactions that recorded the number and type of failed requests of the registry systems. The notion of interoperability is also discussed by the respondents in particular the email interviewees. ***Part D - General comments*** presents the respondents views about what is a metadata registry system as well as their general comments regarding the role of metadata registry systems in data standardisation and sharing and their intended use by diverse organisations and in different information environments.

Received response by method used

Web log transactions of two active metadata registry systems covering a period of six years, from 1998 to 2004, provided rich information about the use of two active metadata registry systems, the MetaForm and the Environmental Data Registry (EDR) which since 2002 is known as the System of Registries (SoR). The analysis involved more than 10.000.000 successful requests of which 2625672 referred to the activity recorded by the MetaForm registry and 9932969 referred to activity recorded by the EDR/SoR.[17] Additionally to the information that was revealed by the transactional logs, response from four questionnaire surveys conducted across several discussion lists and with users of the metadata registry systems at dedicated conferences and workshops complemented the research findings. One hundred and forty six (146) people responded to the discussion lists, the SCHEMAS workshop, the MetaForm, and EDR/SoR questionnaire surveys and highlighted the current use and their future expectations of metadata registry systems (Table 12). The response rate was calculated based on the number of the conference/workshops attendees for the SCHEMAS workshop and EDR and SoR user conferences. In two cases, where the number of the target population was partially known or unknown, the response rate was marked as not applicable. Those cases were:

- The discussion lists. The number of the subscribers to some of the discussion lists was not made available to the researcher despite several efforts of communication with the administrators of the lists. The information in the table below shows the response rate by discussion list based on the received response. For information regarding the scope and the subscription rates of each discussion list, please see Appendix 3b. Within the received responses, the higher response came from the Diglib list (32% of overall response) followed by the SCHEMAS delegates (20%). Almost 17% of the response was received from people who did not state in which list they came across the questionnaire (Table 9). Many people subscribed to more than one discussion list and not indicating in which one they had seen the questionnaire could be explained also as having seen it in multiple lists. Although the overlap of the subscribers to the lists has been identified as one of the disadvantages of discussion lists as a medium to conduct research (Chapter 3 – Methodology) the response shed some light in the trends of metadata research at the time and provides with an adequate insight on the familiarity of the respondents with the use and interest in metadata and metadata registries.

[17] Results list successful page requests which reflect actual page downloads. Successful requests are included in brackets next to the successful page requests throughout the chapter.

Discussion Lists	Response(%)
11179 Metadata Registry Coalition Forum	9.1
Diglib (IFLA)	31.8
Interoperability (UKOLN)	12.1
SCHEMAS	19.7
UK-meg (UKOLN)	10.6
No list	16.7
Total	100

Table 9: Results - SCHEMAS workshop and Discussion Lists surveys - Response

- The MetaForm, which did not employ a registration facility for its users and therefore the number of the users, was unknown. Although findings from the web log transactions provided an indication of the number of accesses on the MetaForm website it was not an absolute for the exact number of users as what was recorded was actual page downloads that could have been performed by the same user.

- Regarding the email interviews, there were two hundred and fifty one (251) email invitations sent out to individuals on a personalised base to invite them to participate to this study. For information regarding the timeline and the selection criteria for those targeted for the email interviews please see Chapter 3 – Methodology. Fifty four (54) emails failed to get delivered due to "undelivered", "unrecognised" or "out of the office" responses. Six of those 54 returned messages were "out of the office" automatic replies set by the individuals when they are unavailable to respond to their email. Those six people were contacted again after two weeks' time but they failed to respond to our invitation. Thirty - nine (39) people responded accounting for 15.5% of the overall response. A hundred and fifty eight (158) people did not reply to the email invitation accounting for 63% of non response. (62.9%, Table 10).

Email interview invitations	Messages	%
Response	39	15.5
Undelivered	54	21.5
Non – response	158	62.9
Total	251	100

Table 10: Results - Email Interviews - Response rates

From the thirty nine (39) responses to the email invitations, two got discarded due to irrelevance to the study. Two people replied that they did not use metadata registries and therefore they did not feel that they were suitable to participate to the interviews. Those two people were contacted again explaining that results from previous studies showed that those identified as users of metadata registry systems were not only the end-users of the systems but also people involved in their design, development and implementation. Furthermore to this explanation, they were invited again to take part to the email

interview. The other remaining nine responses were either directional to other people due to lack of availability or to other resources for information (Table 11).

Email interviews response	Number of messages
Response	26
Directional to others/resources	9
Non use	2
Discarded	2
Total	39

Table 11: Results - Email Interviews - Non use

The response rates by all different methods that were used for this study are presented in the following table:

Individual studies	Response	%
Discussion lists	53	n/a
SCHEMAS	13	32%
MetaForm Metadata registry	8	n/a
EDR Metadata registry	19	38%
SoR Metadata registry	16	33%
Email interviews	37	15%
Web log transactions	2625672 (MetaForm) 9932969 (SoR)	n/a

Table 12: Results - Response by method

PART A – Characteristics of the sample population

This section presents the response pertaining to the demographic and other characteristics of the respondents such as age, sex, domain representation, their occupation and field of interest. Additionally, information about the respondents' familiarity with the use of online resources in general, as well as the frequency and longevity of their use of the MetaForm and the EDR/SoR are also discussed. Furthermore, data from the web log transactions that recorded access by domains and the referrers to the metadata registry systems was also included to provide a more comprehensive profile of the metadata registry systems users.

A.1 Response by age group

The response to the discussion lists, the SCHEMAS workshop and the MetaForm registry surveys was by users from all age groups. People aged between 25 and 34 years (38%) and 45 and 54 years (26%) were the primer respondents to the surveys (Figure 5). Although the received response was equally divided between men and women, the majority of people in the groups from 17-24 and 45-54 were women, while, it was the opposite for the group of people aged between 55-64. The academic domain represented more than half of the response in the age group of 25-34 and more than three quarters in the age group of 65+. The majority of the responses (37%) in the second higher response group, that of people aged between 45 and 54, indicated *"Other"* domains as their main scientific domain. Those included references to the governmental, consulting and natural resources information management domains (Table 13).

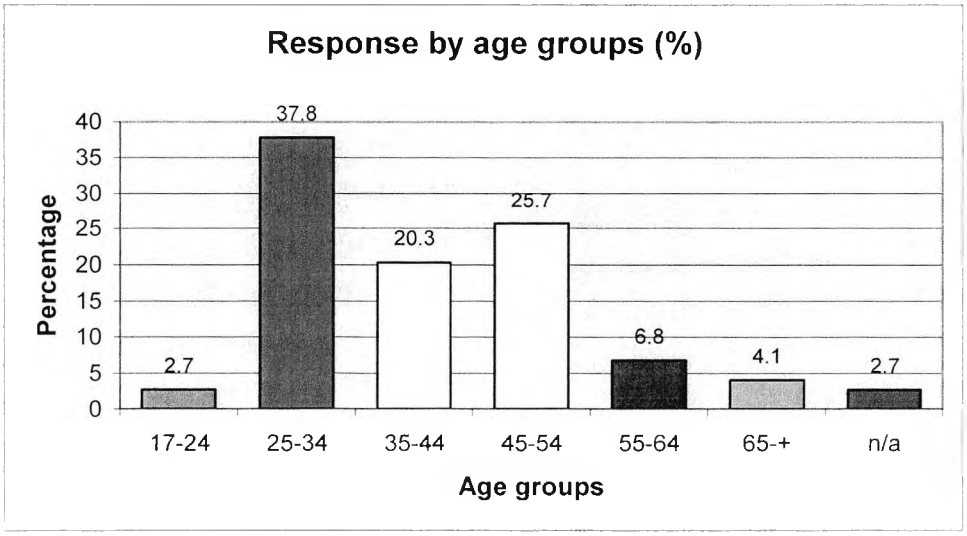


Figure 5: Results - Response by age groups

More than half of the respondents in the group from 25-34 indicated that their job was in the information science (61%). In fact, respondents in all age groups declared their profession in the information science. The only other group that more than half of its members (60%) represented one domain in particular was that of the 55 to 64 years olds. Those respondents identified themselves as computer scientists. The respondents who have indicated the option *"Other"* to denote their work field, fall into the age groups of 65+ (33%), 45-54 (26%), 35-44 (20%) and 25-34 (7%). The category *"Other"* included occupations such as: a director of an organisation, a person working in library relations and training, a private contractor to public land agencies, three project managers (industrial, cultural heritage and research domains), staff development unit project officer, two professors, someone who identified themselves as library faculty, an information executive and a senior projects officer in the area of records management. Table 13 shows the occupations that the respondents indicated in the option field *"Other"* grouped under age groups.

Age group	Other categories
25-34	Developer (Cultural Heritage Domain)
	Information executive
	Library faculty
35-44	Professor
	Senior projects officer, records management project
	Project Manager/ Researcher (Research Domain)
45-54	Director
	Library Relations and Training
	Project manager
	Private contractor to public land agencies
	Staff development unit project officer
65-+	Professor

Table 13: Results - Response by age groups - Category *"Other"* analysis

A.2 Response by occupation

The respondents were invited to provide some information about their occupation. Responses to this question were grouped into the categories of Information Scientists, Computer Scientists, Consultants and *"Other"* using as criterion their professional background as most respondents indicated some involvement in the areas of digital libraries, metadata, and project management. Half of the respondents to the surveys identified themselves as information scientists. A quarter of the remaining respondents (15%) chose the option *"Other"* to indicate the post in their current employment. Category *"Other"* consisted of a director of an organisation, a person working in library relations and training, a private contractor to public land agencies, three project managers (industrial, cultural heritage and research domains), staff development unit project officer, two professors, someone who identified themselves as library faculty, an information executive and a senior projects officer in the area of records management. Computer

Scientists represented the 17.6% of the overall response to the discussion lists, the MetaForm registry and the SCHEMAS workshop surveys. Consultants and people who did not answer this question represented equally 5.4% of the overall response (Figure 6). Respondents to the Environmental Data Registry and the System of Registries surveys as well as the email interviewees were asked to denote their field of interest instead. Results from their responses are presented in the following section.

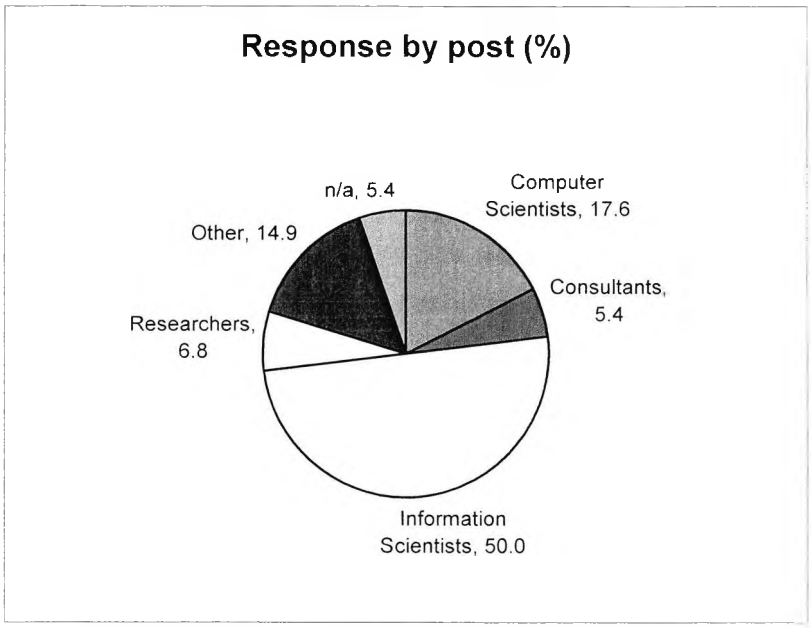


Figure 6: Results - Response by post of the respondents

A.3 Response by field of interest

Response to the EDR/SoR questionnaire surveys represented a range of U.S. state and federal government organisations including the Maine Department of Environmental Protection, the Bureau of Labor Statistics, the U.S. Federal Aviation Administration, the Veterans Health Administration, the Defence Information Systems Agency and Canada such as Environment Canada. Also, it hosted representatives from the Chemical Abstracts Service, numerous consulting firms and EPA contractors and other employees with an interest in the development and implementation of data standards. The first part of the EDR/SoR questionnaire aimed to monitor user interest in specific EDR areas in order to provide baseline for evaluating current services as well as to identify how the EDR could provide a forum for networking for parties interested in environmental information management. The Environmental Data Registry website was designed in order to provide useful information to people interested in systems development, standards implementation and data harmonization. Information for these three areas was accessible via a single access point, named the "How to...Facility". All three services shared the same interface while providing links to different information, relevant to each

of the three areas. The conference attendees were asked to identify the area that posed more interest for them. More than half of the respondents (63%) declared their interest in all three predefined areas rather than any one in particular. This finding was emphasised through comments that the respondents made stressing the importance of understanding how metadata registry systems work, network with other interested parties in the area of metadata standardisation, and promote the use of the EDR. One of the attendees noted how the use of the EDR was directly linked with their work at the Environmental Protection Agency. They noted that the Environmental Protection Agency's policy about evaluating data before re-use in order to review use requirements, entailed keeping up to date with advances in data standardisation and management and using the EDR in order to produce "good metadata". International and intergovernmental liaising were also denoted as important reasons for attending the users' conference and use the Environmental Data Registry. In particular, the users provided the following information regarding their interest in the use of the Environmental Data Registry:

- [I am interested in the] strengths and weaknesses of EDR, how to use, meet others in user community.
- I work with the Canadian Government [organisation omitted] and we are searching for chemical data on 23.000 substances. This is a good forum to learn what the EPA is building and fasten relations between international partners.
- To help promote the use of EDR in my office of EPA among regulation, systems developments and others; and to ensure that EDR is plugged into the Enterprise Architecture effort to gain information to ensure that EA effort will make full use of the EDR

Respondents that indicated that they were interested in *"Other"* that the three predefined areas of information in the EDR, specified that they were interested in policy and technical issues, such as to understand how the EDR fits into the agency's enterprise architecture and to look at the "technical process and progress to learn for the department of Defence". This finding indicated that users were trying to understand how the EDR could be used in compliance with its foster organisation policies for data harmonisation and standardisation rather than already be familiar with a system and use or be interested in specific functions that it performs. Therefore, one of the characteristics of the users of metadata registry systems at that point, were not necessarily the end-users of information systems but those involved into their deployment that they were interested in understanding the context and requirements for their use.

The attendees to the second users' conference and respondents to the System of Registries questionnaire survey that was conducted a year after the EDR survey were

asked again to identify the area that posed the most interest for them. This survey aimed to find out if the changes that occurred in the EDR's structure and the launch of more registry systems, in the form of the System of Registries within the Environmental Projection Agency had affected users' viewings about the registries' use. The categories that the respondents were invited to choose from were adjusted to the previous year's conference response and were now defined in more specification. They included the following options: data standards developers, data standards implementers, exchange network participants, registry developers, someone interested in substances, systems developers and other. This list was also comprised with the conference programme and aims in mind. Surprisingly almost half of the respondents indicated that they were interested in other areas than those provided. Category *"Other"* included the following specifications regarding the respondents' field of interest: evaluation and/or technical support for the overall data sharing and standardisation project in the Environmental Protection Agency such as an ICIS developer, an Information systems analyst, and ISB technical contact, and IT auditor, the manager of data standards implementers, the program evaluator and a project manager. Representatives from the EPA's integrated compliance information system (ICIS) and the information services (ISB) indicated the attendance of senior level managers to the users' conference and demonstrated the significance that the EPA attributed to the role of the System of Registries in the organisation's data standardisation programme. It appears that in spite of the effort that had been made to design specialised metadata registry systems, an effort that lead to the launch of the System of Registries, the publicising of those applications and the endorsement of the EPA policy about data standardisation and harmonisation across its various offices were not equally successful. This finding had been emphasised by comments that the respondents made about the lack of publicity about the launch of SoR and the users conference and the general *"inability of the System of Registries to make its contents understandable to the public"*.

Finally, response from the email interviews highlighted once more the fact that those identified as the users of the metadata registry systems were not restricted to one scientific domain but included a variety of people involved in the deployment of metadata registries such as systems developers, implementers of registries, people who create metadata element sets, people who create combination of elements from different metadata sets (also known as application profiles) and those who search or browse those registries for research purposes.

Interviewees were asked to describe their job responsibilities and to indicate the features and/or services that they were interested in a metadata registry system. Based on their job requirements and responsibilities, the respondents were grouped in three categories:

1. Those engaged in metadata management and/or metadata implementation (MM),
2. Those with a background and interest in data standardisation (DS) and
3. Those in software development (SD) (Table 14).

Field of interest	Number of respondents
SD (Software Developer)	8
MM (Metadata Management)	11
DS (Data Standardisation)	7
Total	26

Table 14: Results - Email interviews - Response by field of interest

One of the findings form the questionnaire surveys was that the respondents' interest in metadata registry systems was not associated with only one area in particular. The majority of the people engaged in metadata registries research tend to get involved in all processes of a registry's deployment from the development to the implementation and advocacy and management. The interviewees' response showed that people involved in ***metadata management*** were those that coordinated the development and implementation of the systems. They did not necessarily use metadata themselves but they acted as the mediators to the data. They ensured that it is available for use and support services such as help guides and relevant documentation were in place to help those who accessed the information. A federal government employee specified *"...I don't "consume" metadata, I make it available. Who is using it is a difficult question to answer. People that are interested / passionate about metadata are rarely customers..."*

Those in ***data standardisation*** tend to work more on data specific projects or in national and/or international standardisation bodies and were those who were more interested in the semantic aspects of metadata registry systems. These people were usually involved in the process of drawing the policies and strategies behind the development and implementation of such systems and they had been interested in promoting the understanding of why registries are important and how they can be used. One of the interviewees with background in data standardisation specified, *"Too often we assume the recipient understands our messages. We have concentrated on the messaging and transport of the messages without the conveyance of understanding."*

Respondents that their main responsibilities lied in ***software development*** were primarily involved in performance compliance and functionality of the systems, developing the tools to facilitate searching and training users of the systems. They tend to use metadata

registry systems mainly for testing rather than information consumption. On of the respondents with background in software development noted, *"My role is as a software developer and my interest in registries is in the machine to machine interfaces. I therefore do not use any metadata registries with any regularity..."*

A.6 Response by familiarity with using information resources

More than three quarters of the respondents (84%) to the EDR survey replied that they access and use online information on a daily basis, although only about half considered themselves to be advanced users. The majority (74%) of them regarded searching and retrieving online information as a relatively easy process that they were familiar with. Despite that, they had pointed out restrains they encountered (Table 15) from time to time during their information seeking process, identifying the *"Large amount of unorganised information"* and *"Badly designed sites that are difficult to navigate"* as the two most common reasons that prevented them from easily accessing the information they sought. Other significant reasons were the lack of services to guide users during their search such as help guides, glossary and site maps. Although the respondents noted the difficulties they encountered when they searched for online information sources in general, the respondents that selected the option "other reasons" referred specifically to the Environmental Data Registry. Those reasons were: too much irrelevant information, confusing search interface, lack of information and limitations imposed by the organisation's firewall. The search feature in the Environmental Data Registry probed many comments and suggestions for improvement. Also, the assessment of the content in the EDR/SoR is discussed further in the section Part B, Use. The reasons that the respondents noted as restricting during their information seeking process are listed in the table below:

DIFFICULTIES EXPERIENCED WHILE TRYING TO OBTAIN INFORMATION FROM ONLINE SOURCES	RESPONSE [%]
Badly designed sites that I find difficult to navigate	61.1
Cost of services/information to obtain	11.1
Lack of guidance tools	5.6
Lack of online help facilities	11.1
Lack of supportive services	33.3
Lack of time required for the search	16.7
Too much information	27.8
Too much unorganized information	61.1
Unfamiliar with searching methods	11.1
Other (please specify): Too many irrelevant sites returned on a fairly general search Confusing search interface Info[rmation] lacking EPA firewall limiting (e.g. access to US NRC ADAMS system or US DOE directives)	22.2

Table 15: Results - Difficulties encountered during the information seeking process

A.6 Response by frequency and longevity of use

Respondents to the MetaForm registry survey were asked to denote the frequency and the period over time that had been using the system. They were given the options of one month, three months, six months and more than that, to select from. More than half of the response to the MetaForm survey indicated that the respondents were either first time or novice users of the system. They specified a usage period of one to three months. Less than a quarter of the response came from respondents that declared they had been using the MetaForm registry for more than six months (12.5%) or for a period longer than that (Table 16). Response regarding the longevity of use is similar. The respondents were provided the options of weekly, monthly, occasionally, hardly ever and first time user to select from. Resembling to the response regarding the frequency of the MetaForm's use, less than half of the respondents noted that they were occasional users and only 12.5% had been using the system on a weekly or monthly basis (Table 18). Although respondents to the MetaForm questionnaire survey were first time or novice users of the registry, web log transactions demonstrated a stable use of the service throughout the six years for which data was analysed. Specifically, the web log transactions showed that use during 2001 recorded an increment of 1.41% compared to that of the previous year. Furthermore, an additional 3.91% was recorded in 2002 compared to 2001 (Part B, Table 22). Regarding the Environmental Data Registry, almost half of the respondents to the survey indicated that they had been long-term users of the EDR (Table 17). They noted that they had been using the EDR services for a period of more than six months ranging from one to five years between sessions. As the Environmental Data Registry had been one of the few registry systems that their parent organisation made it a requirement for the employees to develop and use such a system, the use of the system reflects this organisational policy.

Longevity of use MetaForm	%
1 month	37.5
3 months	37.5
6 months	12.5
More than that (please specify)	12.5
n/a	0
Total	100

Table 16: Results - MetaForm questionnaire survey - Longevity of use

Longevity of use Environmental Data Registry	%
1 month	5.26
3 months	5.26
6 months	10.52
More than that (please specify)	47.36
n/a	31.57
Total	100

Table 17: Results - EDR questionnaire survey - Longevity of use

One third of the conference attendees were first time users of the EDR services and 10.5% of the response indicated that the respondents had been accessing the EDR on a weekly and/or monthly basis. A good one quarter of the response (26.3%) declared an occasional use of the EDR services. Occasional was defined as a period greater than

one month between sessions (Table 19). The contrast of the frequency and longevity of the EDR's use reflected the event that hosted the questionnaire survey, the first users' conference, which, its programme's structure aimed to address a broad range of user familiarity and experience. Furthermore, as metadata registry systems were still at their formative stages of deployment at the time of the survey and events such as the first EDR users' conference attracted interest from many parties across the government and other domains in the USA (use by domains is discussed in the next section). Additionally, web log transactions of the EDR/SoR demonstrated that more than 8500 requests for pages had been made every month (averaged number) throughout the six years for which data was analysed. In particular, web log transactions showed that use during 2002, when the users' conference took place, recorded an increment of 0.9% compared to that of the previous year. It should be noted that use during 2002 recorded its highest point in all years since EDR's inception in 1996. Therefore, re use of the system by either new or returned users is reflected in the web log transactions. The variations in the increase/decrease of the EDR use between 1998 and 2004 are discussed in Part B, Table 23.

Frequency of use MetaForm	%
Weekly basis	12.5
Monthly basis	12.5
Occasionally	37.5
Hardly ever	0
Never used it before, this is the first time	37.5
Total	100

Table 18: Results - MetaForm questionnaire survey - Frequency of use

Frequency of use Environmental Data Registry	%
Weekly basis	10.5
Monthly basis	10.5
Occasionally	26.3
Hardly ever	21
Never used it before, this is the first time	31.6
Total	100

Table 19: Results - EDR questionnaire survey - Frequency of use

A.4 Response by domain (Domain reports)

Comparing the results by different scientific domains from respondents of the discussion lists, the SCHEMAS workshop and the MetaForm registry surveys, the highest response was received from people in the academic domain (39% of overall response) followed by those who selected the option *"Other"* (17.6%) and those in the research domain (16.2%,Figure 7). Category *"Other"* included people working in the governmental, consulting and natural resource information management. There was no response to the surveys from the audio-visual and geographic information domains and only one person represented the publishing domain. Although this could be considered a limitation of the surveys, the selection of the discussion lists and registries was conducted on the criteria of relevance to metadata and metadata research rather than scientific domains that were represented. The scope of this study was not to provide an exhaustive and in depth

review of metadata and metadata registries use in each scientific domain but to highlight general use and in particular by users of active metadata registry systems. Furthermore, data from the transactional logs highlights the actual metadata registry systems use by domain and therefore complemented gaps in this part of the study.

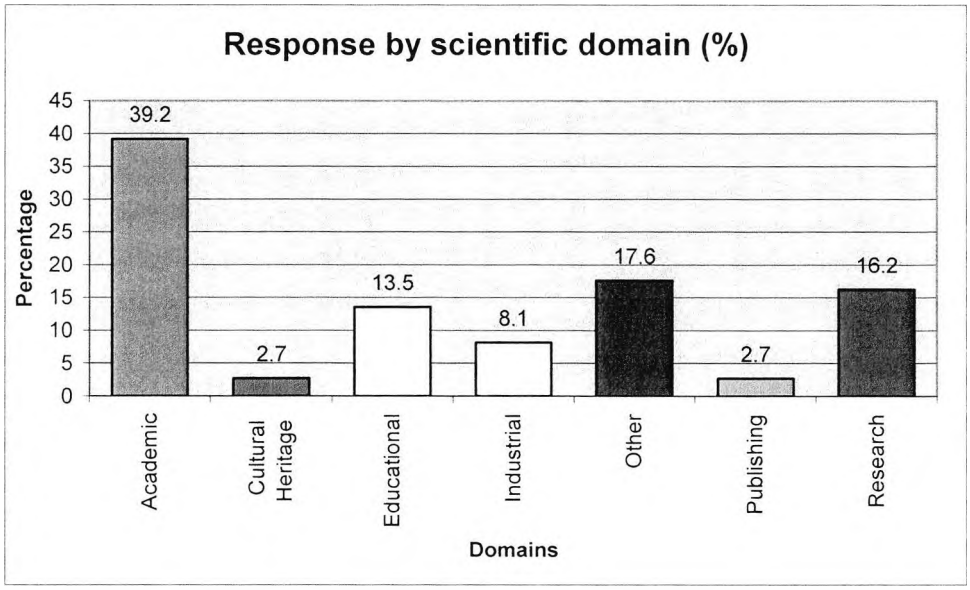


Figure 7: Results - Discussion Lists, SCHEMAS workshop and MetaForm surveys - Response by domains

Data from the transactional log analysis revealed that the European Union countries represented the largest domain to request files from the MetaForm registry, at a total of 3882 (293114) page requests for the period of 1998-2004. The network and commercial domains followed closely with 1689 (146439) and 1189 (103047) page requests respectively. Although there are no restrictions as to who can register a .net or .com domain name these have been primarily associated with telecommunications and network providers and private and commercial entities including companies. As the MetaForm registry was developed to help the library community address issues that network information posed to the standardisation and management of data (see, Chapter 1 - Introduction) it was expected that the educational and academic domains would have developed a primer interest in such services as the MetaForm. On the contrary though, the domains that accessed the MetaForm registry besides network and commercial sites, were the educational (USA), Asian, North American, Australian and organisational domains. Each domain counted less than 200 (17.500) page requests (averaged numbers, Figure 8).

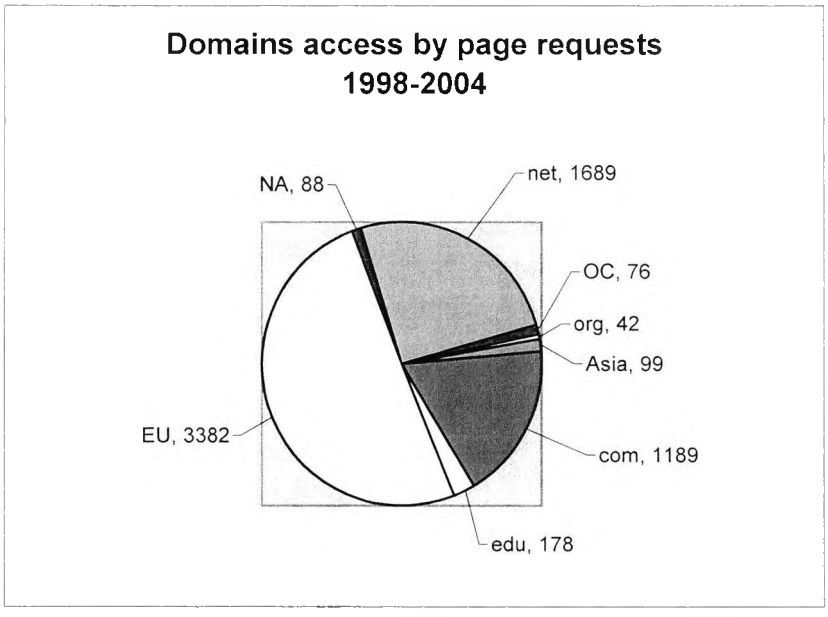


Figure 8: Results - MetaForm - Web log transactions - Access representation by domain (1998-2004)

The constant and stable increase in the access of the MetaForm registry across the years and by several domains is shown in Figure 9. The increase in the domain use has, obviously, been closely related with the general increase in the use of MetaForm and it is discussed in PART B, use of metadata registry systems. The European Union countries represented the domain that showed the highest and most stable use during the six year period of the transaction logs analysed. Since 2002 the use by American educational institutes is also demonstrated in the domain reports.

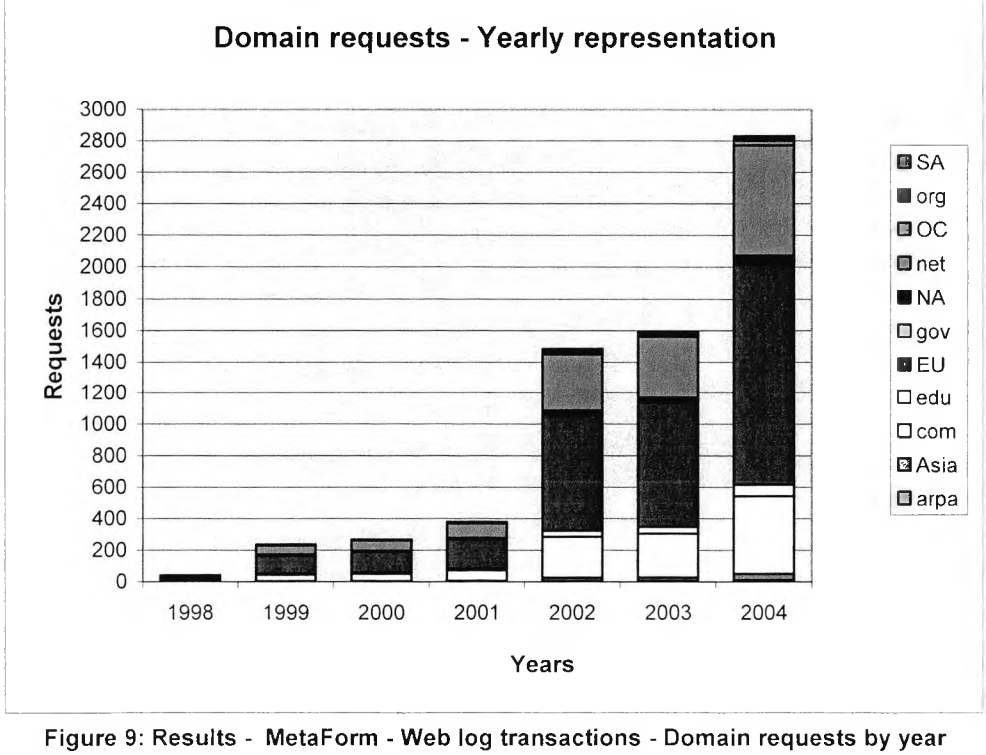


Figure 9: Results - MetaForm - Web log transactions - Domain requests by year

The commercial domain represented the largest domain (205523) to request files from the EDR/SoR. This has been a surprising finding as the System of Registries was primarily developed for the use of a governmental organisation in the USA, the Environmental Protection Agency. Similarly to what had been noted for the case of the MetaForm, there are no restrictions as to who can register a .net or .com domain name these have been primarily associated with telecommunications and network providers and private and commercial entities including companies. One potential explanation for the interest in the EDR/SoR information could be the nature of the data itself. Vast amounts of chemical research data such as crystallographic never become available in either the literature or any databases (Duke and colleagues, 2005). Therefore open access to such type of information could have yielded the interest of commercial domains. This is shown in the directory requests (PART B, section B4.5) that showed that the Chemical Registry System had been the most used registry behind the EDR. Following the use of the commercial domain is the governmental domain at 92078 requests for files during the six year period analysed. Internally at EPA, 81823 files were requested. Another surprising finding was the use of System of Registries by academic institutions. Examples of universities that pointed their students to the information in EDR/SoR included the University of Virginia (USA) and the University of Liverpool (UK).

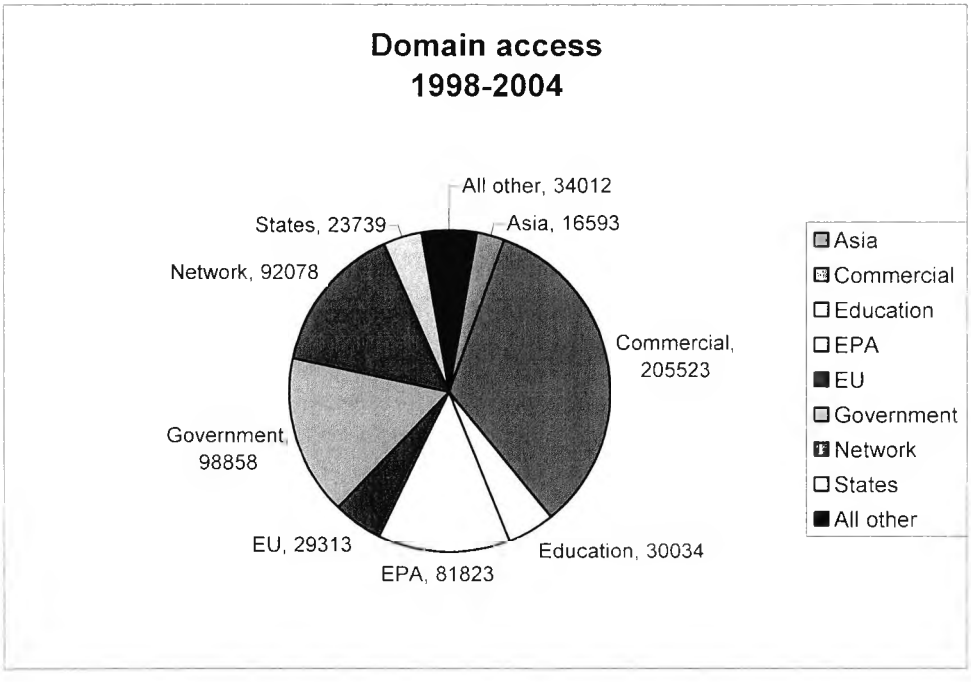


Figure 10: Results - EDR/SoR - Web log transactions - Access representation by domain (1998-2004)

Access by commercial domains remained stable throughout the six years period. The highest number of downloads by the commercial domain was recorded during 2001 and

although it decreased in volume since then it still remained the domain that recorded the highest activity. Contradictory, both governmental domains and the EPA showed a decrease in the System of Registries' use which had been introduced during 2002. It is not a surprise that the highest number of failed requests, which denote erroneous file problems in finding and/or accessing files in general, was recorded in the year 2002 as well (Part B, Figure 45). All domains accesses recorded a dramatic fall during 1999 which had been directly associated with the overall fall in use during 1999. The governmental domain in particular recorded the biggest decrease in use by a domain counting from 26780 to 4093 requests.

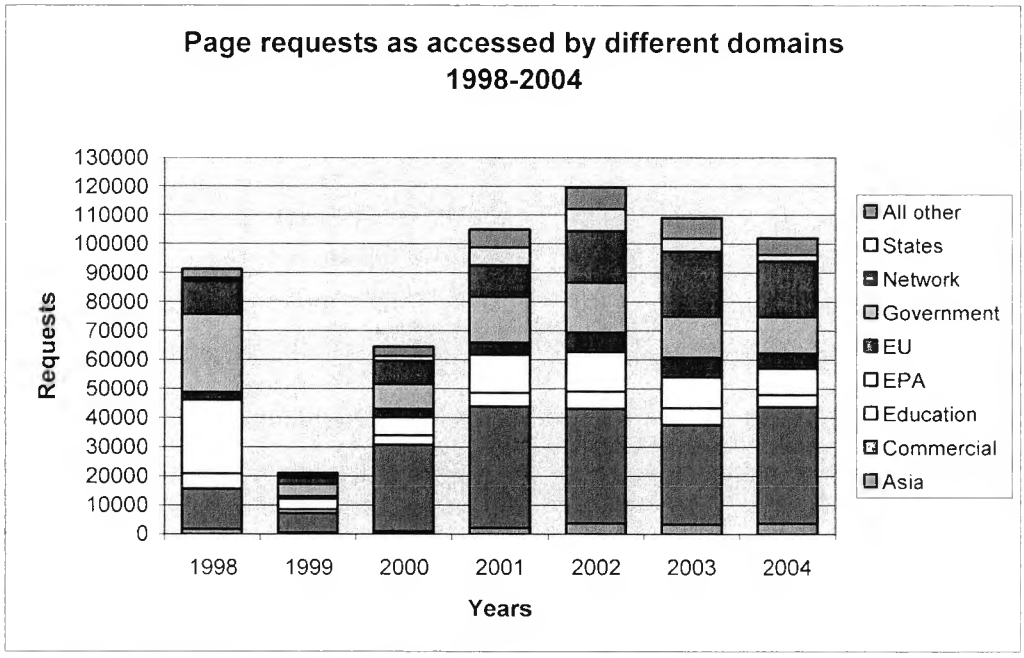


Figure 11: Results - EDR/SoR - Web log transactions - Domain requests by year

A.5 Response by referrer (Referrer reports)

The Analog software documentation (http://www.analog.cx) notes that the referrer report lists in the form of URLs *"...where people followed links from, or pages which included this site's images"*. An example of a referrer report with URLs that linked to EDR/System of Registries resources is shown in Figure 12.

In order to be able to present the type of services and organisations that directed users to the System of Registries services, the referrer URLs were grouped under three categories. Those are explained hereafter:

- Referrers from the Environmental Protection Agency (EPA). This category included all URLs which their first part of the file path was www.epa.gov and were considered as internal referrers.

- Referrers from Search Engines (SE). This category included URLs that referred to major search engines such as Google, Yahoo, Lycos, Altavista, MSN, AskGeeves and AOL. Those were considered external referrers.
- Other. The category other covered those URLs that did not fall in the first two categories. It included links from governmental and commercial organisations, universities, professional bodies and World Wide Web directories. Those were considered external referrers.

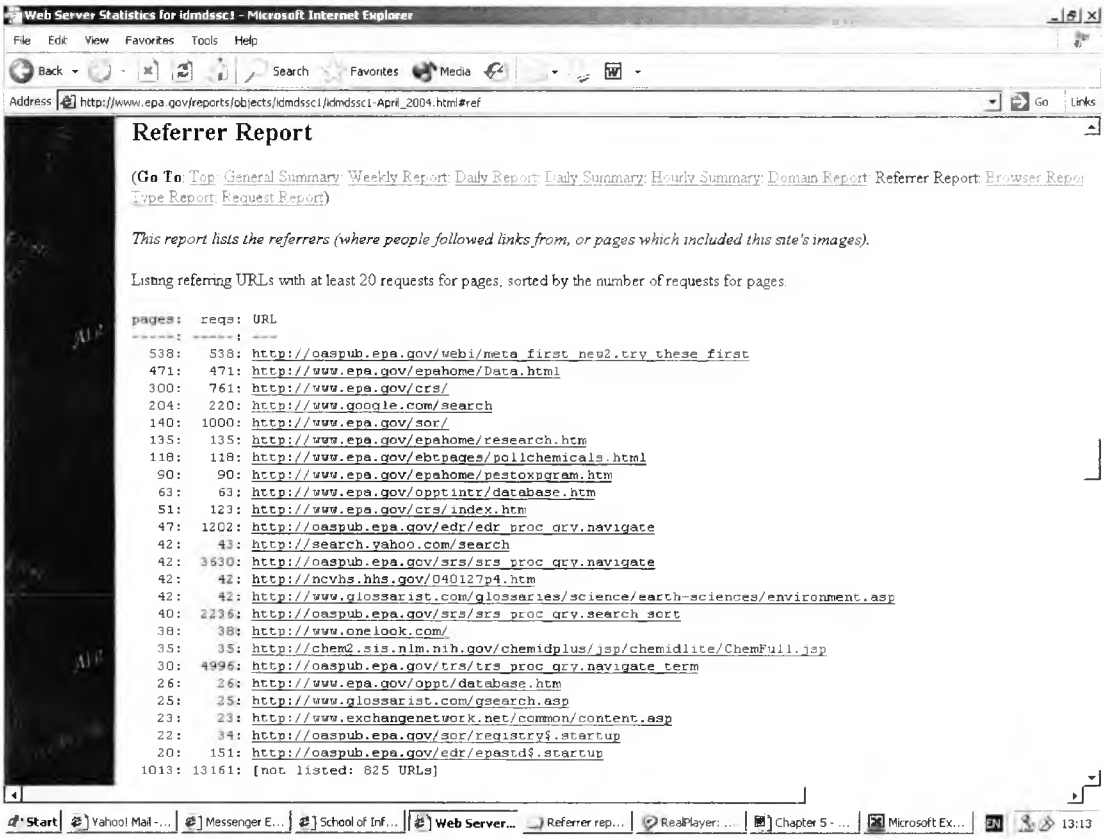


Figure 12: Results - System of Registries - Web log transactions - Example of referrer report

The referrer reports showed that the majority of users that linked to the System of Registries came from the Environmental Protection Agency. Although, as mentioned earlier in this chapter, the System of Registries was developed primarily for the use of EPA staff, the recorded use by the Environmental Protection Agency had been unstable during the six years period of web logs analysed. On the contrary, search engines recorded a more stable increase in their use of the system. There was a small fall in directed access by search engines during the year 2003 but in the next year the use was increasing again. Organisations in category other had been the most stable directors of users to the EDR/System of registries. Although use was first recorded during the year 2000, there has been a constant and stable increase ever since (Figure 13). As mentioned in the section above, referrers in category other included governmental and

commercial organisations, universities, professional bodies and World Wide Web directories.

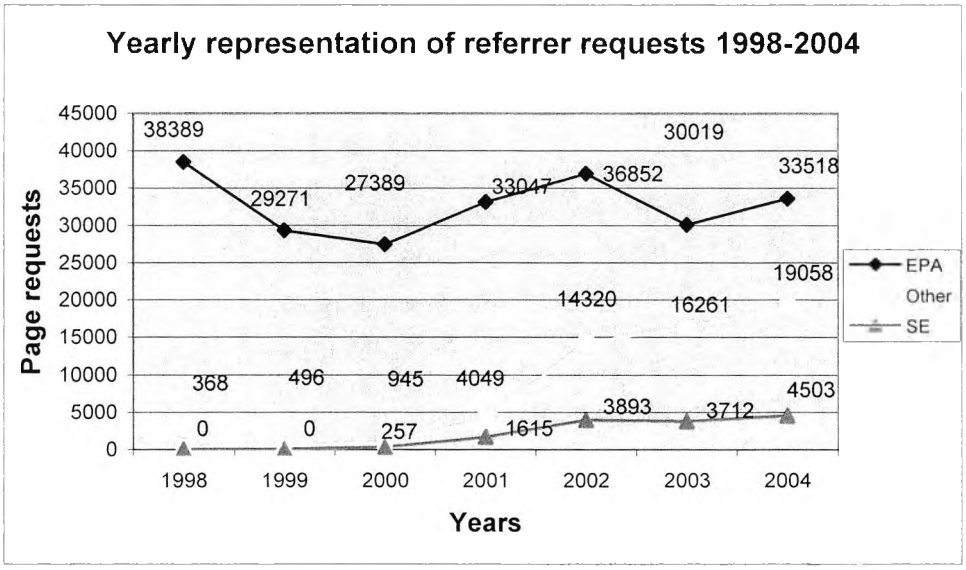
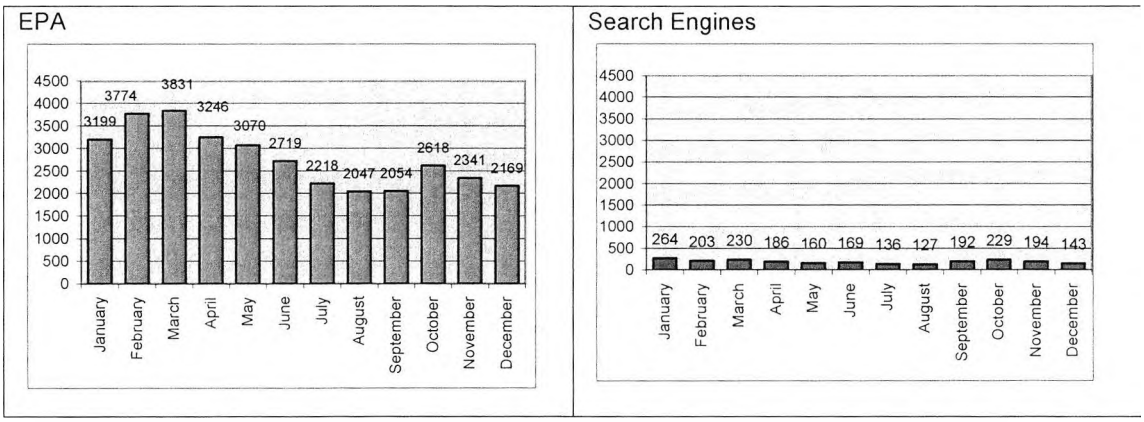


Figure 13: Results - System of Registries - Web log transactions - Referrers access by year

Comparing the monthly accesses by the different referrers, July and August appear the months that all referrers recorded the lowest number of directions to the EDR/SoR website. The EPA referrers made use of the system mainly during February and March recording more than 3500 requests at each of those months. Search engines recorded the highest numbers of directing users to the EDR/SoR in January and October. Finally, other referrers directed most of the users to the system during the months of March and October.



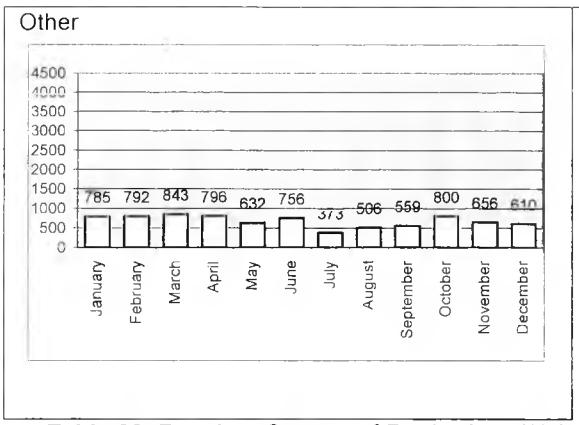


Table 20: Results - System of Registries - Web log transactions - Referrers access by month

Data from the web logs transactions showed that the users of the System of Registries visited the services after the EPA, search engines or other sources directed them there. Other referrals include academic institutions, commercial websites and health and/or environmental organizations. The peak year for Environmental Protection Agency referrals was 2002 when 36852 users where directed via the EPA to the EDR/SoR services. Search engines did not start directing users to the EDR/SoR until the year 2000 but since then they had been noting a constant and stable increase, recording 4503 referrers during the year 2004 after a small drop in access number during the previous year. Organisations and institutions such as universities and other governmental organisations had recorded the highest increase in referrers' accesses during the year 2002 counting an almost four times more to the 4049 recorded during the year 2001. It is interesting to note that governmental and internal EPA use had fallen during the EPA so the use when the System of Registries was introduced it was mainly by external visitors directed there either via search engines, academic institutions and other governmental sites. This is a significant finding for the EDR/SoR as it shows that the work that had been done with metadata and the facilitating of standardisation of information has been successful as it was sought out by users other than those for which the system was initially developed for. *"Other"* organisations and institutions had been noting a constant increase in directing users to the EDR/SoR during the period of 1998-2004 (Figure 14).

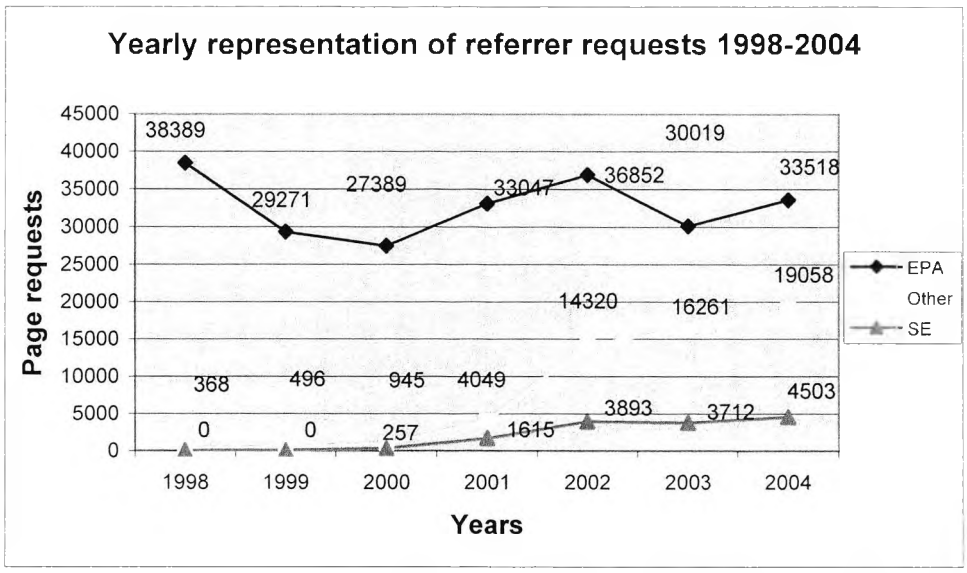


Figure 14: Results - System of Registries - Comparison of referrers' access by year

PART B – Use of metadata registry systems

The following section presents results referring to the respondents' understanding and use of metadata in general and the use of metadata registry systems in particular. Use of the MetaForm registry and the EDR/SoR is presented by day of the week, month and year and over the six year period of web log data that was analysed. Additionally, users' access to specific directories and/or types of information as well as particular features and/or services of the metadata system registries is also presented. This part also discusses how an organisation's requirement for a metadata registry system can or cannot increase the demand for such services and finally, presents current implementations of metadata registry systems as the respondents denoted those.

B.1 Metadata terminology

The respondents to the discussion lists questionnaire surveys were invited to indicate the terms that they associated mostly with metadata. They were asked to select from a list of terms such as catalogue, element set, format, schema, standard, system and to indicate any other terms that were not included in the list provided[18]. The terms that they were asked to choose from were not accompanied with any examples of use or clarification on definitions. That was intentional in order to find out about the respondents' perception of the above mentioned terms with minimum intervention. It is representative of the growing interest in metadata research at the time of the surveys that all of the respondents had heard about metadata prior to participating to the questionnaire surveys. The most popular terms that the respondents had associated with metadata were standards (94%) followed by schema (91%), element sets (87%) and formats (76%) (Figure 15).

[18] Although the intention was to use exactly the same questionnaire for both the SCHEMAS and the Discussion Lists' surveys in the end it was not possible. As a result of a mistake in the distribution at SCHEMAS workshop, a previous version of the questionnaire was included in the delegates' pack. That version of the questionnaire excluded part 2, question 2 that refers to metadata terms. The version that was posted to the Discussion Lists was the fullest one. Therefore the response discussed in this section refers to response received by the discussion lists respondents only.

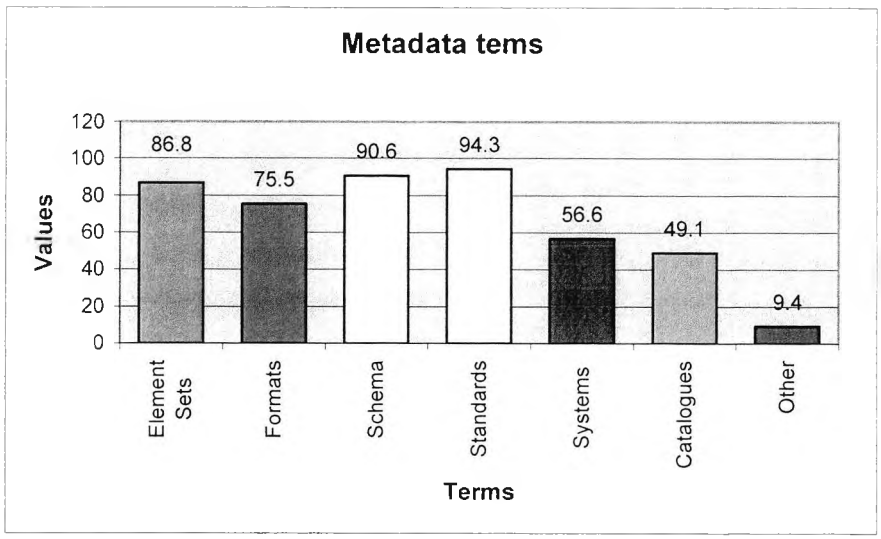


Figure 15: Results - SCHEMAS workshop and Discussion Lists surveys - Metadata terms

A **"standard"** bares the notions of safeguarding and raising levels of quality, safety, reliability and efficiency (International Organisation for Standardisation, http://www.iso.ch). Therefore, its meaning has references to and from all scientific domains. During the late 1990s and the beginning of the new decade, a **"schema"** was mostly associated with the computing domain and denoted *"...the representation of a vocabulary in a particular machine-processable form, such as an RDF or relational-database schema (a "semantic schema"). Most specifically, "schema" may refer to a file describing the tag structure of an XML-encoded document, as in an XML Document Type Definition (a "document schema")"*[19]. The **"element set"** as a term is primarily associated with the Dublin Core and has its roots in the "metadata movement" as this was defined by Thomas Baker in 1995. Finally, **"format"** was primarily associated with the library community. References include the MARC, UNIMARC and domain specific (e.g., locally created) formats. Taking into consideration the above elaborations, it can bee acclaimed that interest in metadata research have been primarily associated with the provision of quality, reliability and efficiency. Furthermore, the computing and networking infrastructure that would support the use and management of metadata were also strongly emphasised. Taking into consideration though that **"schema"** has also been used to *"...designate a set of semantic units (i.e., metadata elements or subject headings) along with their attributes,*

[19] The SCHEMAS Forum - a Retrospective Glossary. Available at: http://www.schemas-forum.org/info-services/d74.htm#8 (Last accessed, 22/11/2005) "In current usage, the term "schema" can refer to a wide range of things from the abstract and general to the very specific. In the abstract, "schema" is sometimes used to designate a set of semantic units (ie, metadata elements or subject headings) along with their attributes, such as name, identifier, definition, or relationship to other semantic units. In the SCHEMAS Project, we have avoided referring to such concept sets generically as "schemas" and prefer the popular term "vocabulary" (see Vocabulary). More narrowly, "schema" can refer to the representation of a vocabulary in a particular machine-processable form, such as an RDF or relational-database schema (a "semantic schema"). Most specifically, "schema" may refer to a file describing the tag structure of an XML-encoded document, as in an XML Document Type Definition (a "document schema")". Information from the SCHEMAS project glossary. Available at: http://www.schemas-forum.org/info-services/d74.htm (Last accessed, 12/06/2006)

other semantic units"[20] the contribution of multiple domains in the semantic aspects of metadata is also strong. Finally, the library domain's presence is strong as well providing the links to past use of metadata for the description and discovery of mainly text resources.

The respondents that selected the option *"Other"* indicated additional terms that they had primarily associate with metadata. Those were the following: retrieval facetted classification, metadata registries, metadata repositories and open archival information systems reference model. One of the respondents commented that they have come across numerous terms about metadata in the literature and when they were asked to elaborate they noted:

"Not sure where you want me to stop"

B.2 Use of metadata element sets

The respondents to the discussion lists and SCHEMAS workshop surveys were invited to indicate which metadata element sets they had come across either in the literature, their work or in discussions with colleagues. They were also asked to indicate those that they had used and to provide examples of how they had used them. The aim of this question was to gather some information about the use and the popularity of specific metadata element sets in each scientific domain and to list current metadata research activity during the time that the surveys were conducted. All of the respondents to the discussion lists and the SCHEMAS workshop surveys had heard about metadata element sets in the past and more than 70% had also used one. The metadata element sets that the respondents had most frequently come across in the literature and/or at their workplace were the Dublin Core (84%) and the MARC standard (74%) followed by the GILS (47%) and those noted in category *"Other"* (39%, Figure 16). The respondents indicated a range of "other metadata element sets", a few of which were confused with funded metadata research projects and an electronic journal. The list of "other metadata element sets", grouped by scientific domain, is presented in Table 21.

[20] ibid.

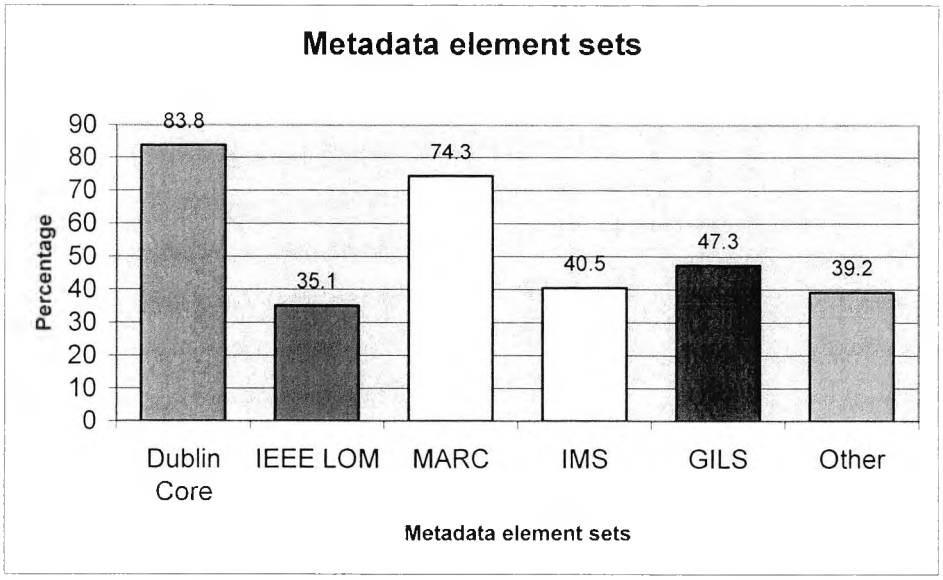


Figure 16: Results - Discussion Lists and SCHEMAS workshop surveys - Most popular metadata element sets

Both the MARC standard and the Dublin Core metadata element set had been used by respondents from all domains and both recorded their highest use by respondents in the academic domain. The Dublin Core metadata element set has been used primarily by the academic (95%), the educational (88%) and the research domains (71%). The MARC standard was indicated as the standard mostly used in the academic domain (76%) and surprisingly by respondents from the industrial (50%) and publishing domains (50%, Figure 17).

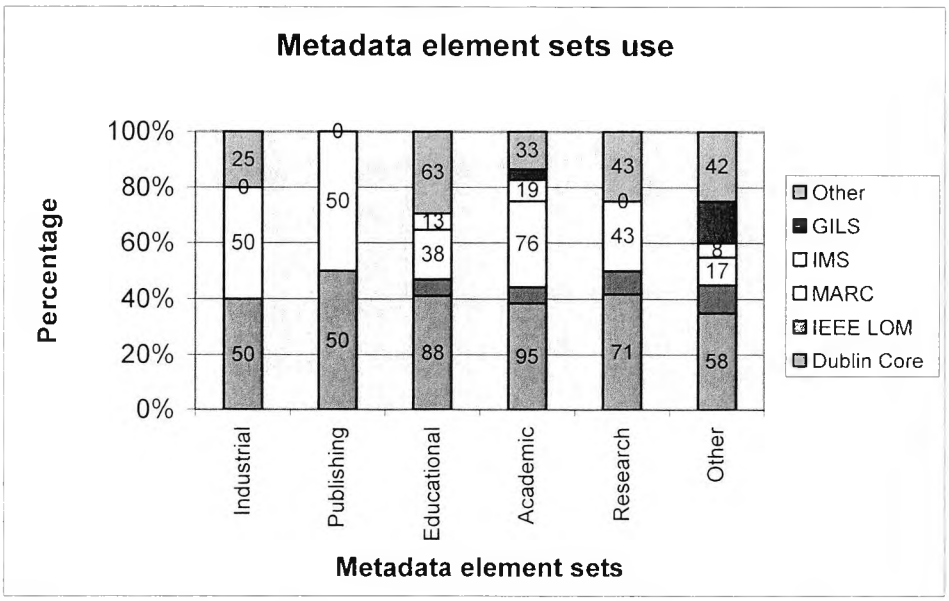


Figure 17: Results - SCHEMAS workshop and Discussion Lists surveys - Metadata element sets use in domains

The respondents from all scientific domains, with the exception of Publishing, had noted a variety of *"Other"* element sets that they had used. These are listed in Table 21. The metadata element sets that were noted did not always refer to an actual element set but to metadata research projects, frameworks, project outcomes, periodicals, locally created application profiles and metadata registry systems software. Results from this question are indicative of the different domains contribution to metadata research. The range of the noted "other element sets" and their international breadth including references to research undertaken in Australia, the USA and Europe is fascinating. It shows that metadata research is conducted at an international level with a contribution from many and different scientific communities. The potential and the opportunities for collaboration as an outcome of such interest are intriguing.

Domain	*"Other"* metadata element sets used
Academic	Ariadne CCF CERIF EAD, TEI, FGDC INIS ("pre MARC" library format) SGML "Various others, especially from archival community - can't remember the exact names. TEI. Collection level descriptions. Z39.50 profiles"
Educational	AGLS, BEP, EdNA, South austalian government metadata standard EOS,TasDiscorver, RDS, AGLS, EdNA, GEM NC (National Curriculum) RSLP (& other collection description schemas), ISAD (G) TEI Lite
Industrial	Whitemarsh Metabase schema
Research	Browsable Corpus-MPI ESAD (g)/Technical metadata for digital still images (NISO draft standard), EAD Own schema
Other	AGLS AGLS, Acore Environmental Data Registry; Australian Knowledgebase; USHIK (US Health Information KB) FGDC Schemas in EPA and HCFA Beta MetaPro

Table 21: Results - SCHEMAS workshop and Discussion Lists surveys - Category *"Other"* metadata element sets used

B.3 Applications of metadata element sets

The two most cited reasons for using metadata were: a) the project the respondents were working for required the use of a metadata element set (70%) and b) to describe their Library's Resources (41%, Figure 18). As discussed in Chapter 2 – Literature Review, research in the area of metadata at the beginning of the 21st century was thriving counting 47 research projects alone on metadata associated with the description, organisation and retrieval of information. Having the same search performed in later years (November 2002 and 2005) revealed more than twice that number (97 and 128

projects respectively). Taking into account conferences and workshops organised by international associations such as Institute of Electrical and Electronics Engineers (IEEE) and the Association for Computing Machinery (ACM) since 1997 it can be said that the above finding is representative of the metadata research that was taking place at the time of the survey and indicative of the 50% of the information scientists that responded to the surveys. The description of library resources is primarily associated with text resources and it shows the active role of libraries and archives in metadata research. This finding is also supported by the choice of the metadata element sets that the respondents indicated such as the Encoded Archival Description (EAD) standard that began as early as 1993 from a collaboration of the Library of Congress and University of California in order to support encoding archival finding aids using the Extensible Markup Language (XML). Also, the ISAD (G) that is the default standard for archival description; the RLSP schema that was again, a UKOLN based project for the description of collections rather than single resources in the research domain and ran from 1999 to 2002; the TEI and TEI Lite that was first published in 1994 and provided with guidelines for electronic text encoding and interchange. They all aimed to provide guidelines for the description and facilitation of exchange of resources in a networked environment.

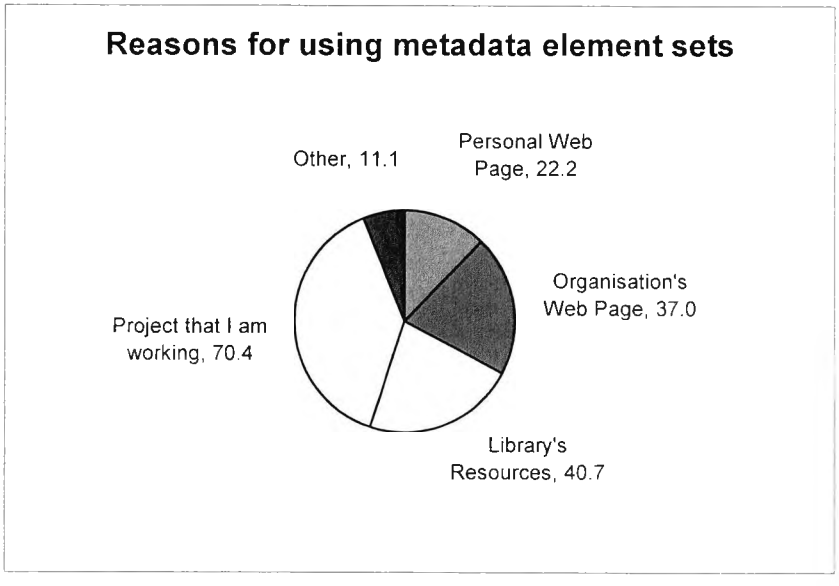


Figure 18: Results - SCHEMAS workshop and Discussion Lists surveys - Reasons for using metadata element sets

The respondents were invited to describe the projects that they were working for and that they required the use of metadata element sets. They provided a fascinating list of projects that included the description of digital image collections, the collection of art images in a digital library, the digitisation of the Southeast Native American documents 1730-1842, the compilation of a metadata encyclopaedia, a training village database of learning resources and description of resources for a Subject Based Information Gateway.

B.4 Use of metadata registry systems

More than half of the respondents (60%) to the discussion lists and SCHEMAS workshop surveys noted that they have heard about metadata registry systems before they participated to the questionnaire surveys but less than one quarter (22%) of them had actually used one. The SCHEMAS workshop aimed to look at the diverse and emerging metadata standards from the viewpoint of the metadata implementers and it was anticipated that the workshop attendees would have some experience and/or understanding of metadata standards use and implementation. The fact that from the 60% of respondents who have admitted that they were familiar with metadata registry systems less than one quarter (22%) had used one indicates that metadata registry systems were not as popular or as deployed as metadata and metadata element sets at the time.

The most popular registries among the respondents were the DESIRE registry (73%), the ROADS templates (64%) followed by the indecs framework (43%) and the Knowledgebase (39%)[21]. The dominance of European metadata registry systems over those based in the USA and Australia can be explained by the fact that the SCHEMAS project was funded by the European Union and run by the UK Office for Library Networking and the focus was in European initiatives. Therefore, the attendees to the workshop represented the European metadata research scene. Additionally to the DESIRE registry and the ROADS templates more than one third of the respondents (43%) responded that they were familiar with the "indecs metadata registry". The indecs project was an initiative of the publishing community to address issues pertaining to metadata and e-commerce. One of the project deliverables was the analysis of requirements for a metadata framework that could be applied in the area of digital information and e-commerce.

[21] The Knowledgebase since 2005 is known as METeOR. Information available at: http://meteor.aihw.gov.au/content/index.phtml/itemId/181162 (Last accessed: 12/006/2006)

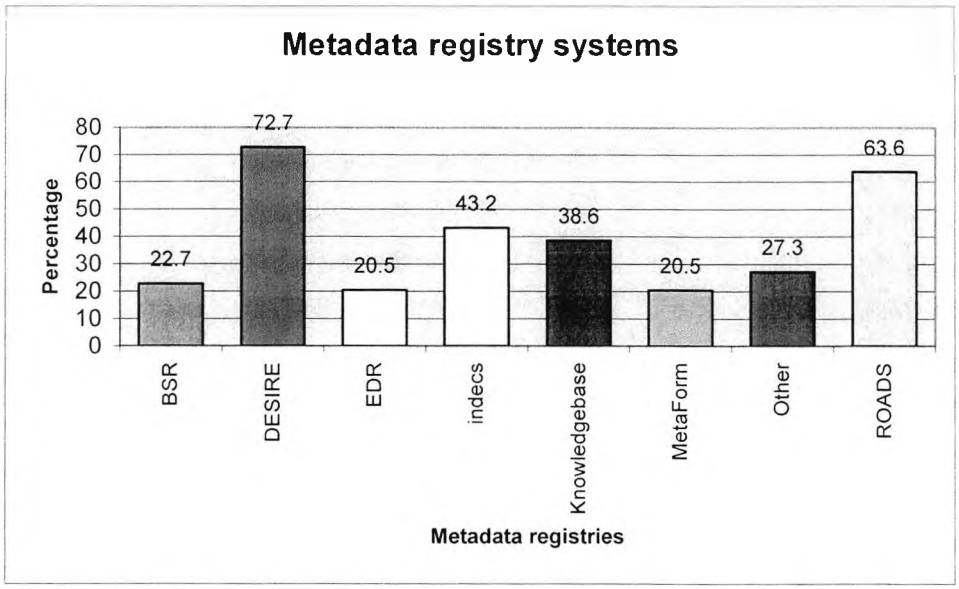


Figure 19: Results - Discussion Lists and SCHEMAS workshop surveys - Most popular metadata registry systems

The 22% of the response that indicated use of a metadata registry system referred to the use of the UKOLN designed and hosted ROADS templates and the MetaForm registry. Although respondents equally noted the DESIRE and the EDR registries as the systems that they had mostly used, they also indicated *"Other"* metadata registry systems. Among actual metadata registry systems, the respondents had also, confusingly, noted workshops, metadata standards and commercial software for automatic creation of metadata. Those were: BSI and govtalk, SMMS, USHIK, US's EPA registry, US's Department of Census registry, and Whitemarsh's, CEN workshop; the DCMI Registry, the SCHEMAS Registry[22], and the MMI-DC Observatory. The MMI-DC Observatory was an initiative by the European Committee for Standardization (CEN) and the Dublin Core community to collect and maintain information regarding metadata activity in the European Union that used the Dublin Core standard[23].

B.4.1 Use by day of the week (Daily summary)

In the MetaForm and EDR/SoR time summary reports, the Analog software, gathers the total number of requests for pages in each time measure, e.g. day, hour, etc. over a period of time. The daily summary presents the total number of "successful requests for pages" of days during the week. This report enables the systems managers to assess which days during the week that the MetaForm and EDR/SoR registry systems were

[22] The SCHEMAS metadata registry was under development at the time and it was not included as an option in the questionnaire as it was not officially publicised as an active metadata registry system.

[23] The Dublin Core metadata element set has been an accredited international standard with ISO number 15836 since 2003. For information please see the International Organization for Standardization website at the following URL: http://www.iso.ch.

mostly accessed by their users. The following figure (Figure 20) presents the average number of successful requests for pages for MetaForm during the period of July 1998 to the 10th of November 2004.

On average, Monday (678) appears to be the busiest day during the week, followed closely by Thursday and Wednesday, both recording 664 requests for pages. Although the weekend, compared to the week days, is the least busy time at a respective average of 537 page requests on Saturdays and 529 on Sundays, it is still an active time for the MetaForm registry. This is one of the characteristics and advantages that the accessibility of open access, online systems such as metadata registries provides to users all over the world. This is also reflected in the domains reports (see Part A) where besides European Union countries, other major users of the MetaForm were the educational (USA), Asian, North American, Australian and organisational domains.

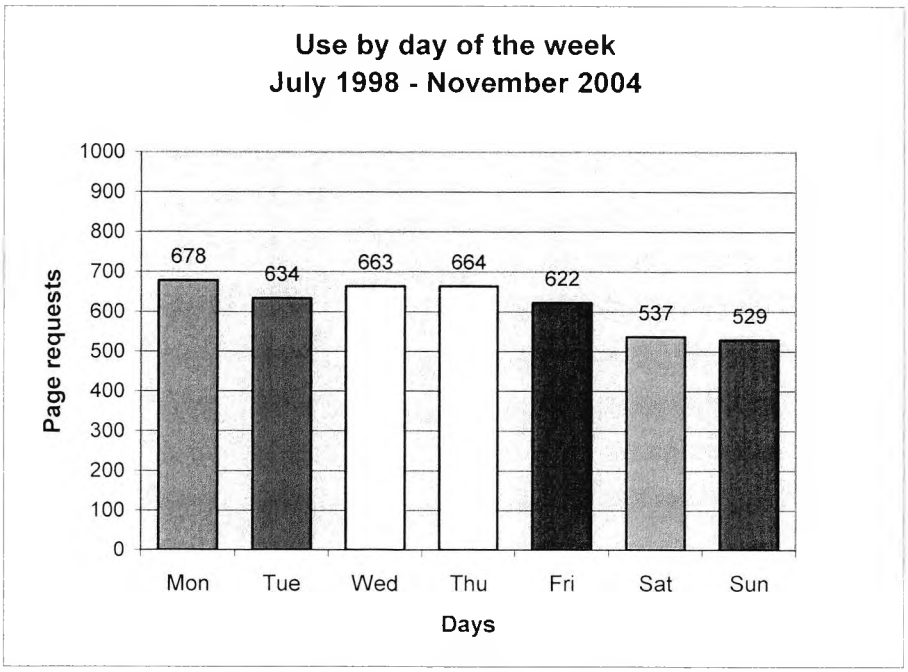
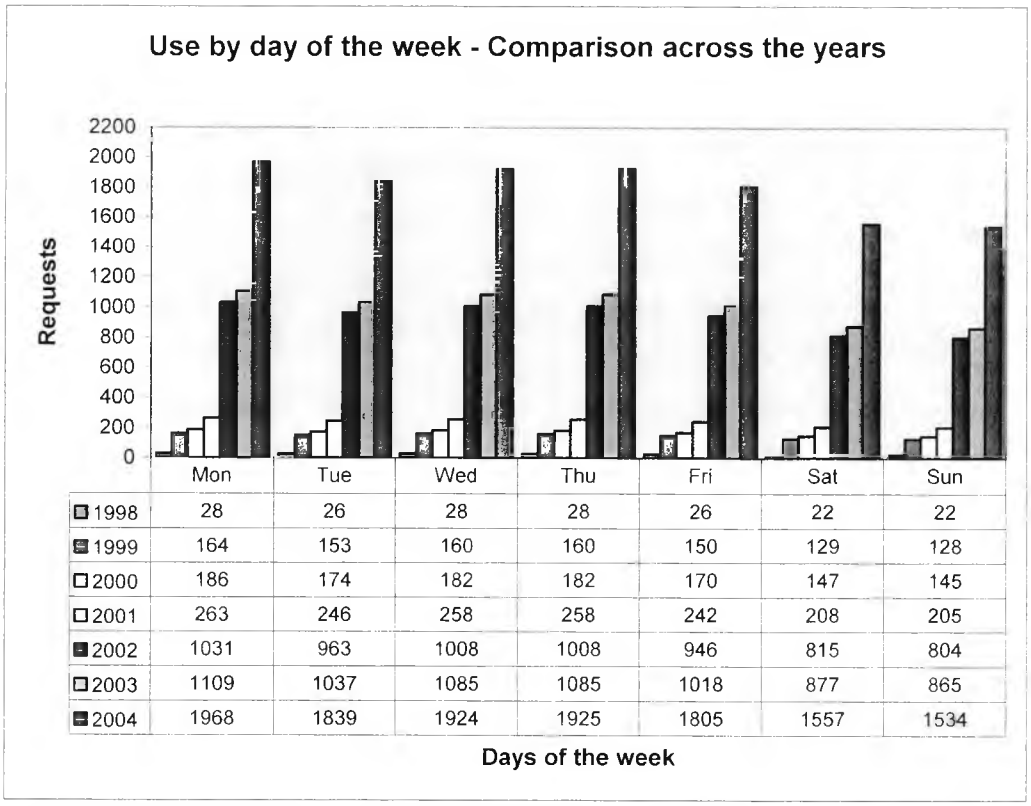


Figure 20: Results - MetaForm - Web log transactions - Use by day of the week (July 1998-November 2004)

Comparing the daily use over the six years, Monday appears to be the day of highest activity counting at the peak of its use in 2004, 1968 page requests. Thursday and Wednesday follow close as days of high activity, recording respectively at their peak during 2004, 1925 and 1924 page requests. On average, the daily summary of all days from 1998 to 2004 has seen an increase of almost 70 times.[24] Obviously the increase in the use every year had a direct effect on the daily use as well.

[24] Please note that data for 1998 represent use of a six months period (July – December) and for 2004 from January – the 10th of November.



	Mon	Tue	Wed	Thu	Fri	Sat	Sun
1998	28	26	28	28	26	22	22
1999	164	153	160	160	150	129	128
2000	186	174	182	182	170	147	145
2001	263	246	258	258	242	208	205
2002	1031	963	1008	1008	946	815	804
2003	1109	1037	1085	1085	1018	877	865
2004	1968	1839	1924	1925	1805	1557	1534

Figure 21: Results - MetaForm - Web log transactions - Daily summary - Yearly comparison

The daily summary has recorded a constant increase of the MetaForm registry's use throughout the six years period of web log transactions analysed. The biggest increase was noted in 1999 compared to the previous year, counting on average daily summary from 28 to 164 page requests, almost six times higher.[25] The second biggest increase was noted from 2001 to 2002, moving from 263 page requests to 1031, almost four times higher. The increase all years from 1998 to 2004 is shown in Table 22. As the Dublin Core was the base for the services that the MetaForm provided, an increase or decrease in the implementation and use of the Dublin Core element set would directly affect the use of the MetaForm as well. The continuous increase in the use of the MetaForm can be better understood when the following events in the Dublin Core deployment are taken into consideration. At the beginning of the 21st century the Dublin Core counted 33 separate implementations in libraries only (Guinchard, 2002) and it can be acclaimed that it was among the most popular metadata element sets in use. Furthermore, in the year 2003 the International Standardization Organisation proclaimed it to an official standard under publication contributing further to its establishment.

[25] Ibid.

Years	Increase in daily summary
1998-1999	5.79
1999-2000	1.13
2000-2001	1.41
2001-2002	3.91
2002-2003	1.07
2003-2004	1.90

Table 22: Results - MetaForm - Web log transactions - Daily summary - Increase

For the EDR/SoR systems the recorded use had been even higher. During the period of the six years and eight months that data was analysed, Wednesday and Tuesday were the days during the week when EDR/SoR were mostly accessed. On Wednesdays, 22242 pages (averaged number) were downloaded, making it the busiest day for the EDR/SoR. Wednesdays were followed closely by Tuesdays and Thursdays as the next most active days recording 22153 and 21324 page downloads respectively. In general, days during the week, comparing to the weekend, appeared significantly more active although some activity had still been recorded on the weekends (Figure 22).

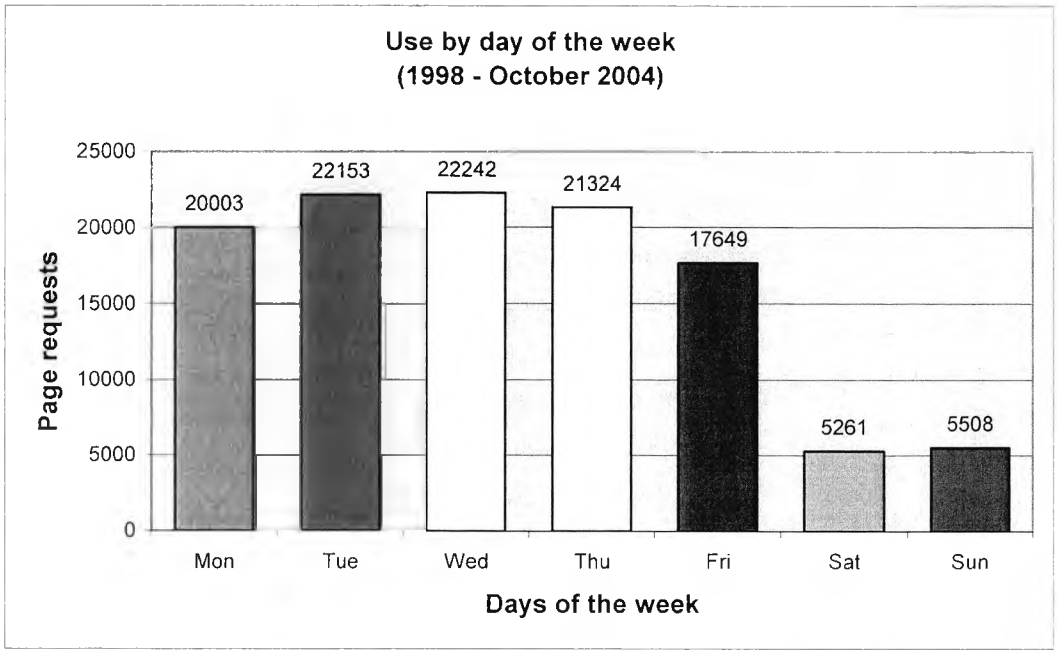
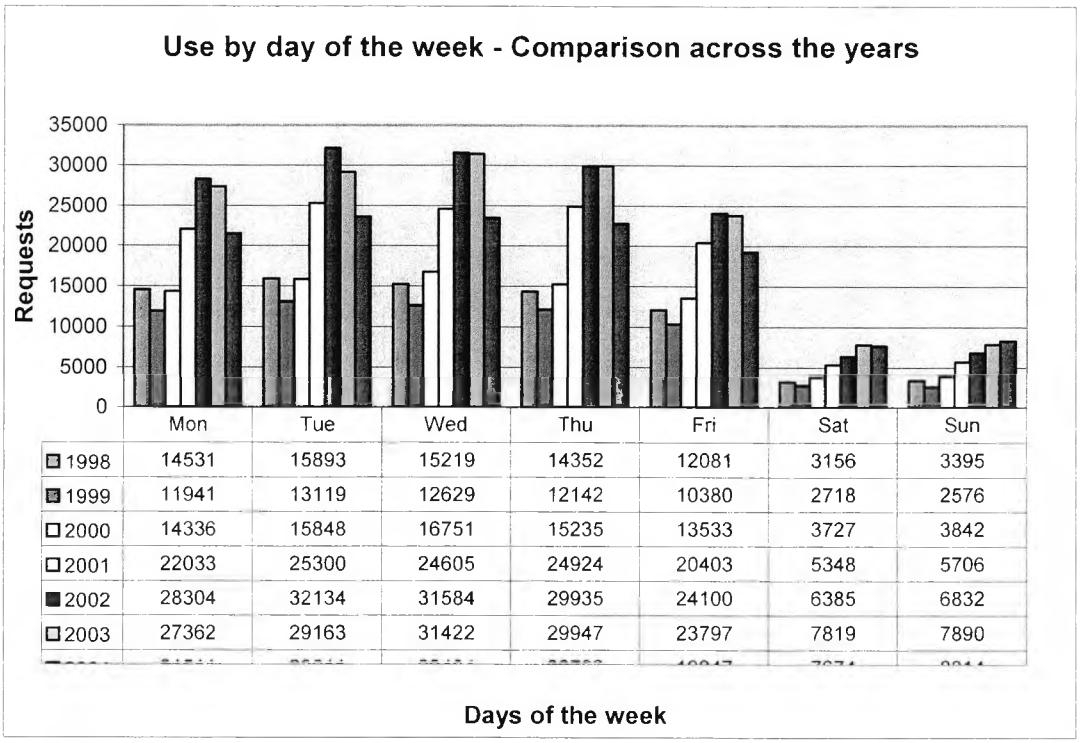


Figure 22: Results - System of Registries - Web log transactions - Use by day of the week (1998-October 2004)

Tuesday, Wednesday and Thursday were the days during the week when Analog recorded the highest user activity. Although the number of downloaded pages appeared almost the same for the three days, the highest activity was recorded on Tuesdays during the year 2002. The System of Registries users had downloaded some 32134 pages (averaged number). The least active day was Sunday during 1999 when only 2575 page requests were noted (Figure 23). On average, the daily summary for all days for the period between 1998 and 2004 has seen an increase of 1.8 times (averaged number).



	Mon	Tue	Wed	Thu	Fri	Sat	Sun
1998	14531	15893	15219	14352	12081	3156	3395
1999	11941	13119	12629	12142	10380	2718	2576
2000	14336	15848	16751	15235	13533	3727	3842
2001	22033	25300	24605	24924	20403	5348	5706
2002	28304	32134	31584	29935	24100	6385	6832
2003	27362	29163	31422	29947	23797	7819	7890

Figure 23: Results - System of Registries - Web log transactions - Use by day of the week - Yearly comparison (1998 – October 2004)

From 1998 to 2002 the Environmental Data Registry recorded a constant increase in use. The biggest increase was noted in 2001 compared to the previous year, counting from 83272 to 128319 page requests (averaged number), a 1.5 times higher. The second biggest increase was noted from 1999 to 2000, moving from 65505 page requests to 83272, almost 1.3 times higher compared to the previous year. The increase in all years from 1998 to 2004 is shown in Table 23. This is a significant finding as in the year 2002 the Environmental Protection Agency introduced the new System or Registries that replaced the Environmental Data Registry as it was previously known (please see, Chapter 1 – Introduction). Although it could also be interpreted as a preference of the users to the previous metadata registry system, cross tabulation of the daily summary with failed requests on the year 2002 reports (see, Part C) the highest ever number recorded for the System of Registries. That means that changes in the System of Registries website structure affected the system's use but not necessarily because of lack of interest but because files were not found. It is surprising though, that failed requests had fallen in the years 2003 and 2004 but the use had not increased. Unfortunately, the System of Registries did not provide with the errors that caused the failed requests therefore it can only be generally said that failed requests reflect problems of accessibility to information.

Years	Increase in daily summary
1998-1999	0.83
1999-2000	1.27
2000-2001	1.54
2001-2002	1.24
2002-2003	0.98
2003-2004	0.80

Table 23: Results - System of Registries - Web log transactions - Daily summary – Increase

B.4.2 Monthly use (successful requests for pages)

The metric that Analog uses to measure the monthly use for the MetaForm and the EDR/SoR systems is that of successful requests for pages. During the first four years that the data was analysed, the monthly use for the MetaForm registry appeared to increase slowly but at a stable rate. Use during 2003 and 2004 was more unstable, with what appeared as constant ups and downs in the recorded accesses. The biggest increase is recorded from September to October during the year 2002. As the published literature about the MetaForm registry is not strong, it can be assumed that events that spurred that increased use were associated with adjoining projects such as the Meta-Lib project (see Chapter 1 – Introduction). During October 2002 the Meta-Lib project run its conclusive workshop which was hosted by the same institute that hosts the MetaForm registry, too and therefore workshop attendees could have shown an interest in the use of the MetaForm registry. The biggest fall in monthly use is noted from September to October 2004 (Table 24). As the MetaForm was concluded as a funded project in 2003 and as there was no registration facility for its users it can not be said what caused this decline in use. The monthly use for all years is presented in Table 24.

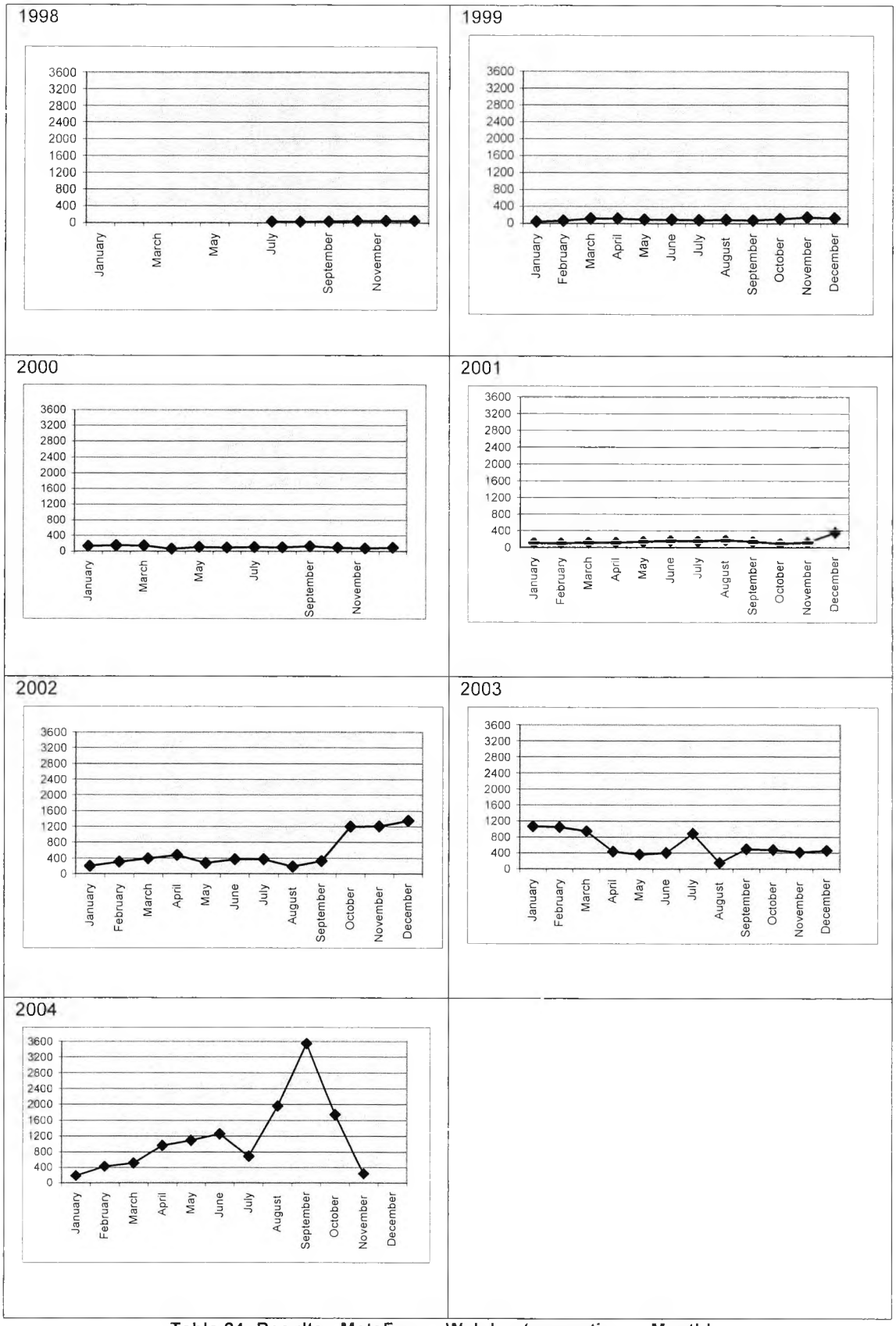


Table 24: Results - MetaForm - Web log transactions - Monthly

Comparing the monthly use over the six years period of web log transactions analysed, the months of September (669) and October (531) appear to be those during which the highest activity had been recorded. The least active months were January (279) and November (316). [26]

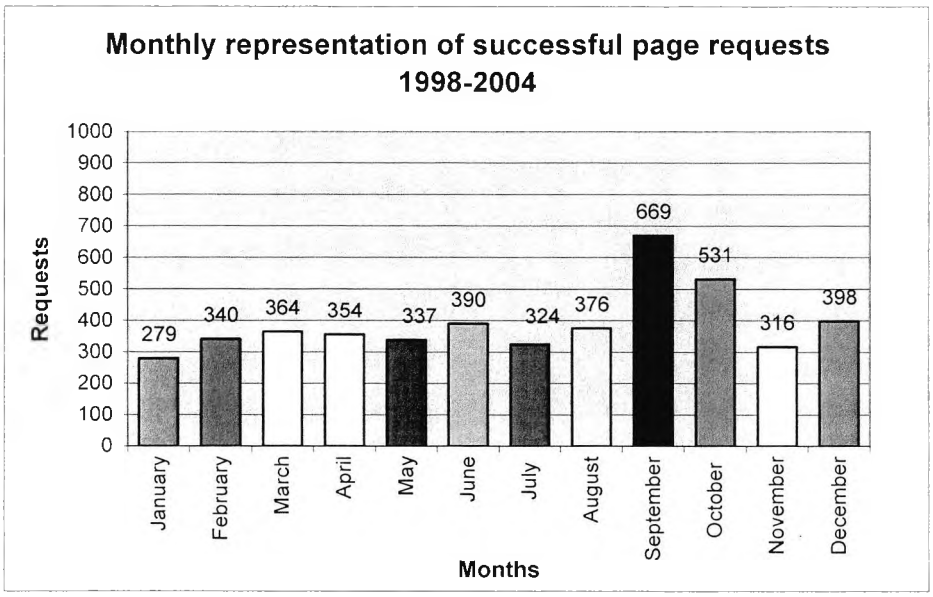


Figure 24: Results - MetaForm - Web log transactions - Monthly representation of successful page requests 1998-2004

The recorded use of the EDR during the first three months of 1998 appeared to be on the increase counting at the point of its highest use during March 9723 successful requests for pages. The lowest point of the recorded use was in August (4259). During 1999 all months recorded similar numbers of successful page requests with an average of 5000 each month. The most unstable year in the use o fthe systems had been the year 2001 when use from September to October mountain from 7536 to 23771 page requests recording 68% increase. It is believed that the increase in the use had been triggered by the Environmental Protection Agency's decision to form a committee to implement the Chemical Identification Data Standard throughout the EPA. Although a considerable drop in use was recorded with the launch of the SoR during 2002 and similarly a smaller decrease during the years2003 and 2004 it had been stabilised above 10000 (averaged number) requests per month.

[26] Averaged number of successful page requests.

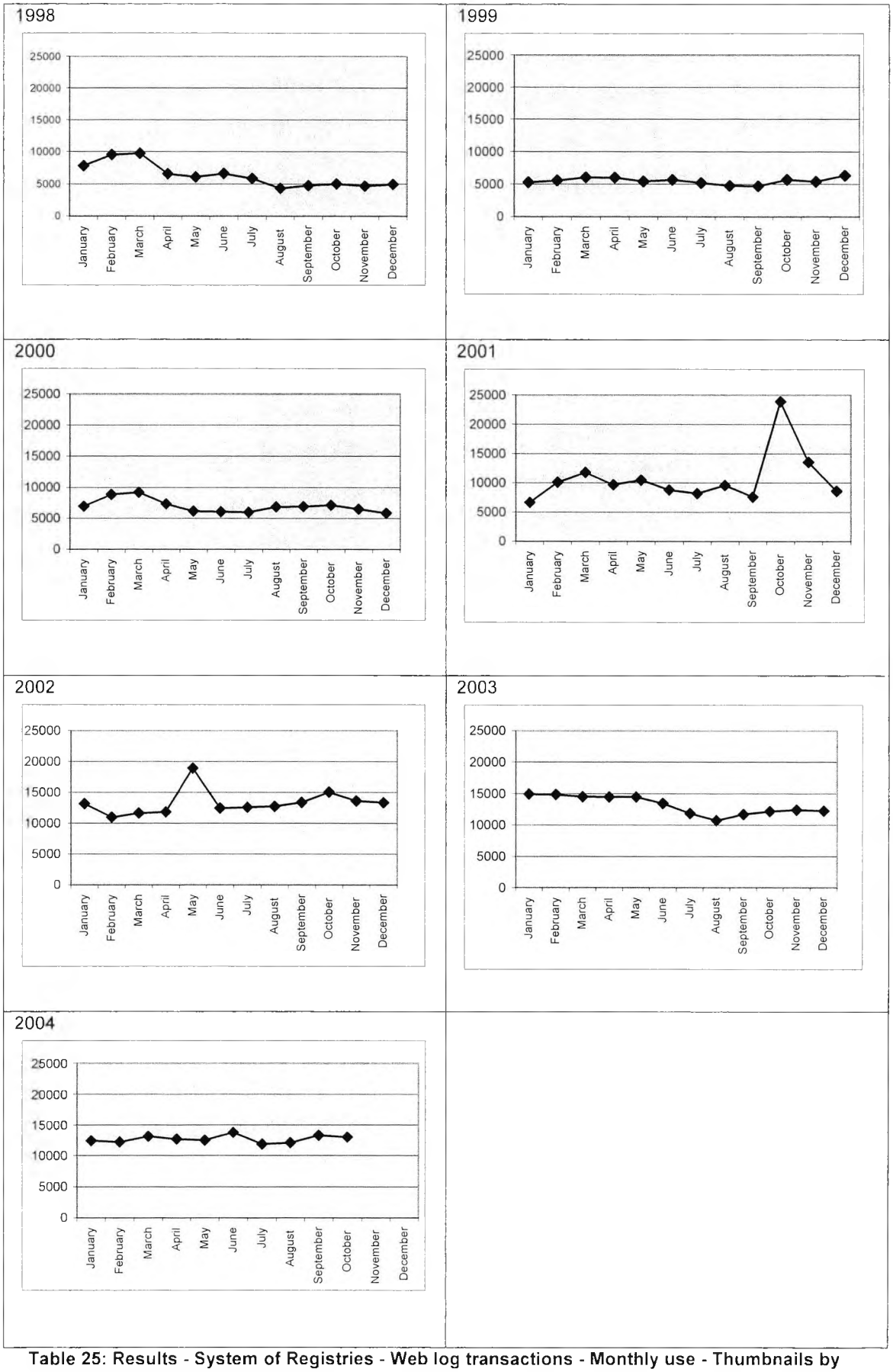


Table 25: Results - System of Registries - Web log transactions - Monthly use - Thumbnails by year

Throughout the six years and eight months from 1998 to October 2004 the EDR/SoR users had requested some 795801 files. Most of the requests were made during October (11666, averaged number). The other most active months had been March at 10830 page requests and May at 10538 (Figure 25). The least active month was December that recorded 8510 page requests.

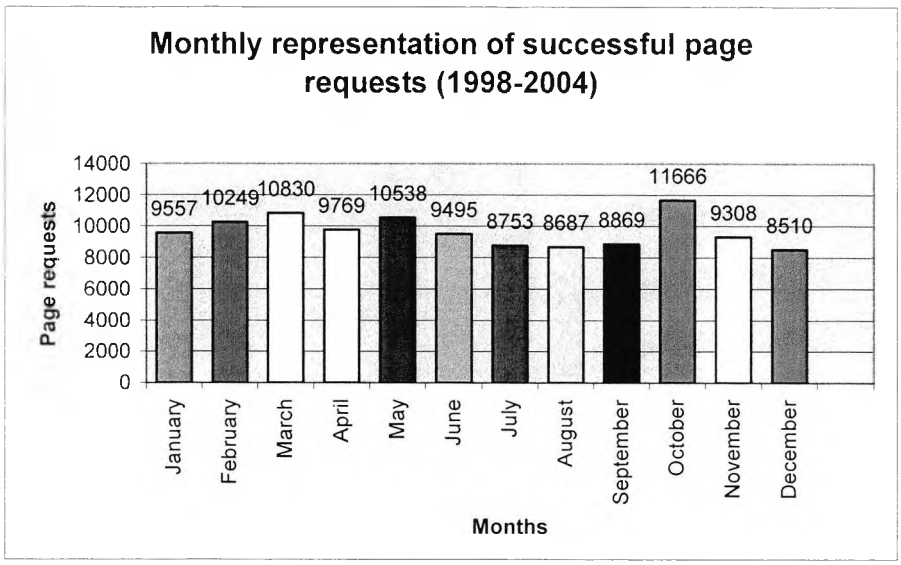


Figure 25: Results - System of Registries - Web log transactions - Monthly use - Successful page requests for the period of 1998-2004

B.4.3 Yearly use (successful page requests)

Web log transactions showed that the MetaForm registry had recorded a constant increase between 1998 and 2004. The biggest increase was recorded in 2002 counting almost four times more accesses compared to the previous year. Although the MetaForm registry as a funded project was completed in 2003 there had been a continuous growth in its use even after then. The MetaForm registry has been hosted by the same server that hosted the MetaGuide, which appears to be the resource that the users of the metadata server in SUB Göttingen were primarily interested in. This association is thought to have increased the use of the MetaForm registry.

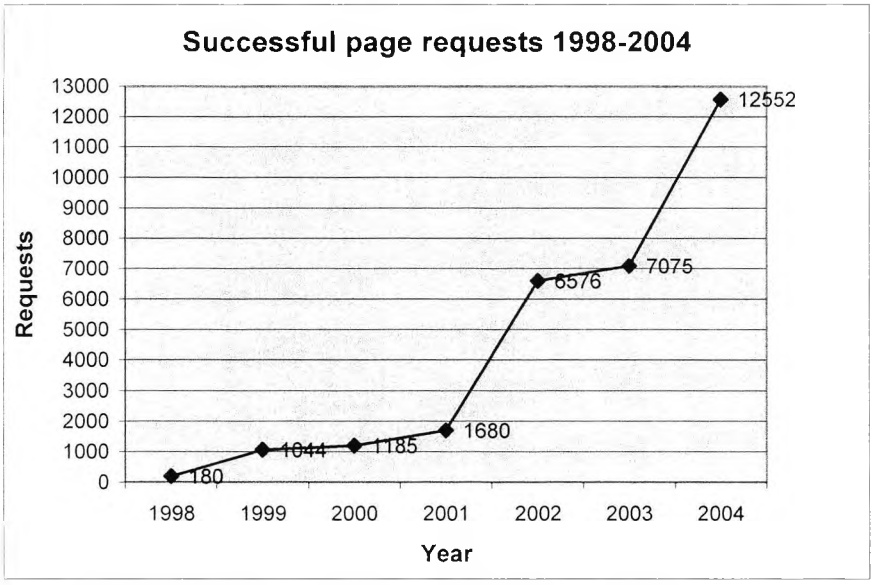


Figure 26: Results - MetaForm - Web log transactions - Yearly use

The launch of the System of Registries in 2002 which replaced the Environmental Data Registry at its current state had an effect in the yearly use of the registries. Between 1999 and 2002 there had been a constant increase in the use of the EDR recording an impressing more than double page requests. Since mid 2002 the launch of System of Registries meant more and new metadata registry systems and consequently an increase in its use was expected. Surprisingly, web log transactions recorded the exact opposite outcome. The use had fallen from 159274 page requests to 157400 in the year 2003 and further to 126573 in 2004.

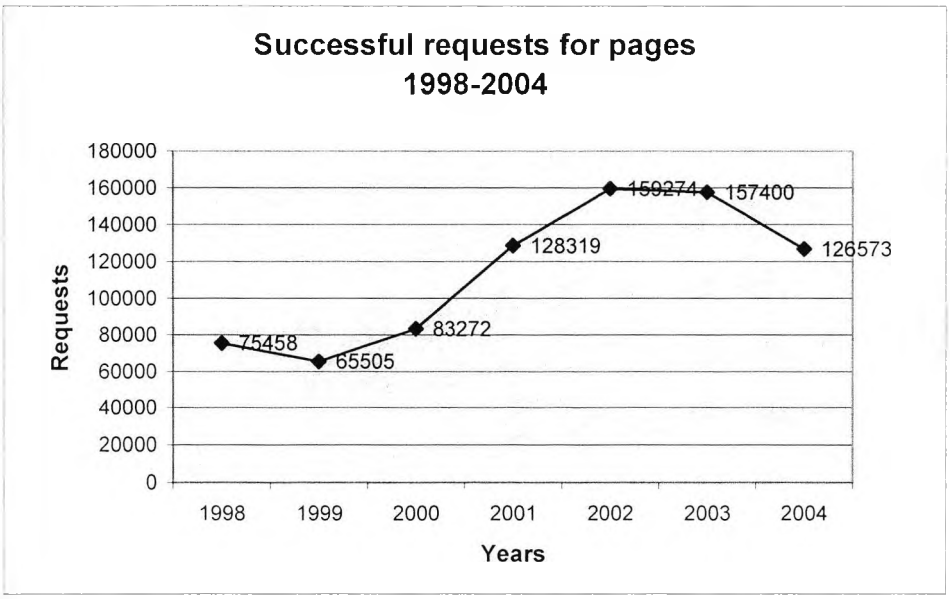


Figure 27: Results - System of Registries - Web log transactions - Yearly use - Successful page requests 1998 - October 2004

B.4.4 Increase/Decrease of use (successful requests for pages per day)

The metric of average successful page requests per day reflects the actual page downloads that users had made. This metric facilitates the monitoring of the increase and/or decrease of the MetaForm and EDR/SoR systems daily use. During the comparison of the daily activity across months for the period of 1998-2004, the ten days of the November 2004 use, have been excluded from the counts in order to provide a more accurate monthly representation of the daily use.

The MetaForm had been mostly accessed during September, the peak month for all years, recording the highest daily activity counting 91 average successful page requests. October is the second busiest month for all years and the only other month that counts more than 70 average successful page requests (73). The use is noted at its lowest during January adding 38 average successful page requests per day which is less than double compared to the peak month September (Figure 28).

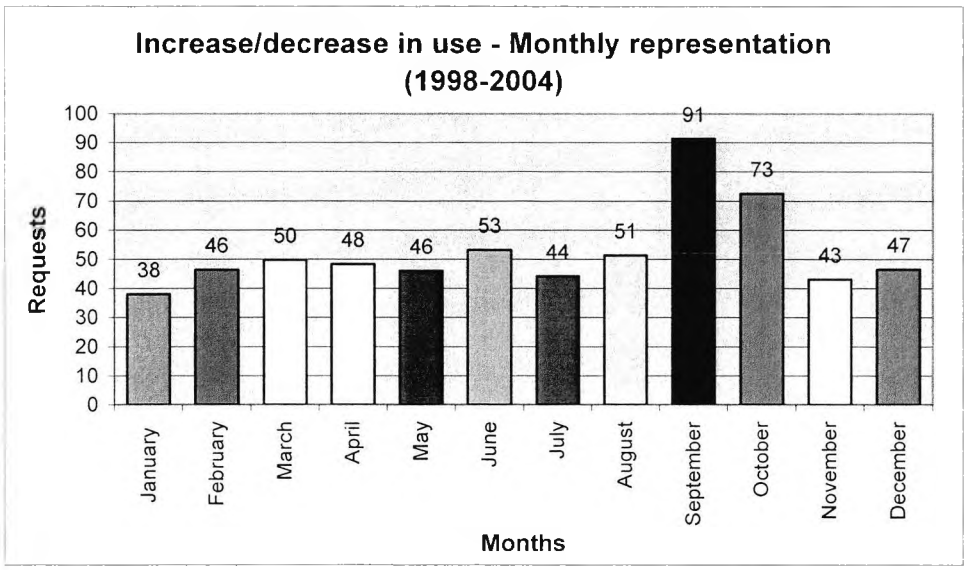


Figure 28: Results - MetaForm - Web log transactions - Daily use

The highest level of use in a year was recorded during 2004 when 1713 files were downloaded from the MetaForm registry. The biggest increase in use though it was recorded between the years 2001 and 2002 when an almost four times increase in the files that were downloaded was recorded. This finding is encouraging for the use of the MetaForm registry and metadata registry systems in general if it is taken into consideration that it had been as a funded project that ran between 1998 and 2003. The number of downloaded files by year is shown in.

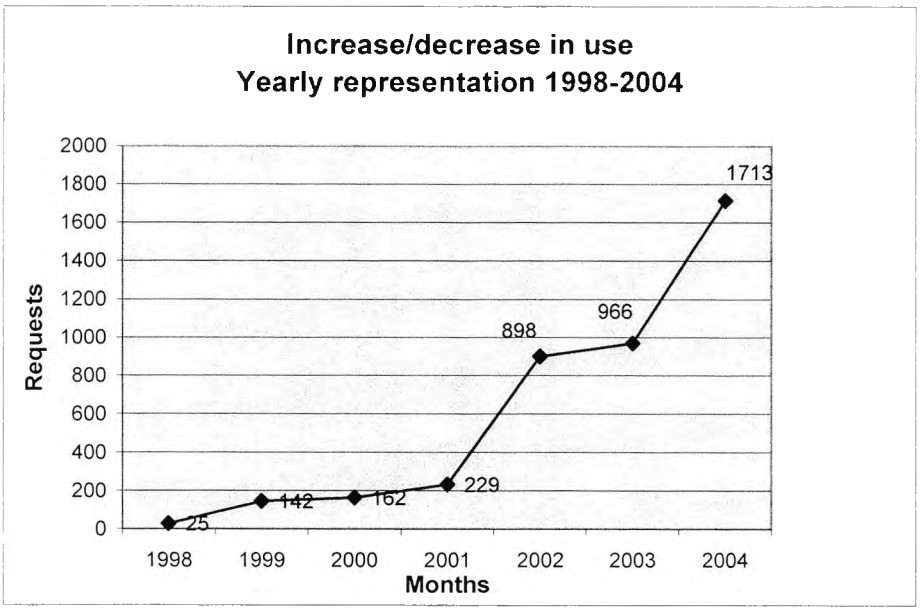


Figure 29: Results - MetaForm - Web log transactions - Comparison of the daily use by year

Since October 2001 the EDR/SoR added to its web logs transactions the metric of average successful page requests per day. This metric reflects the actual page downloads and facilitates the monitoring of the increase and decrease of EDR/SoR's daily use. Data were analysed for the period of October 2001 to October 2004 only in order to provide more accurate monthly representation of the daily use. The highest and lowest points in daily use had been noted in the year 2001. October 2001 had been the peak month for the daily activity in all years from 2001-2004 recording 767 average successful page requests at its highest point and December of the same year, the month with the least average successful page requests at 276 counts (Figure 30).

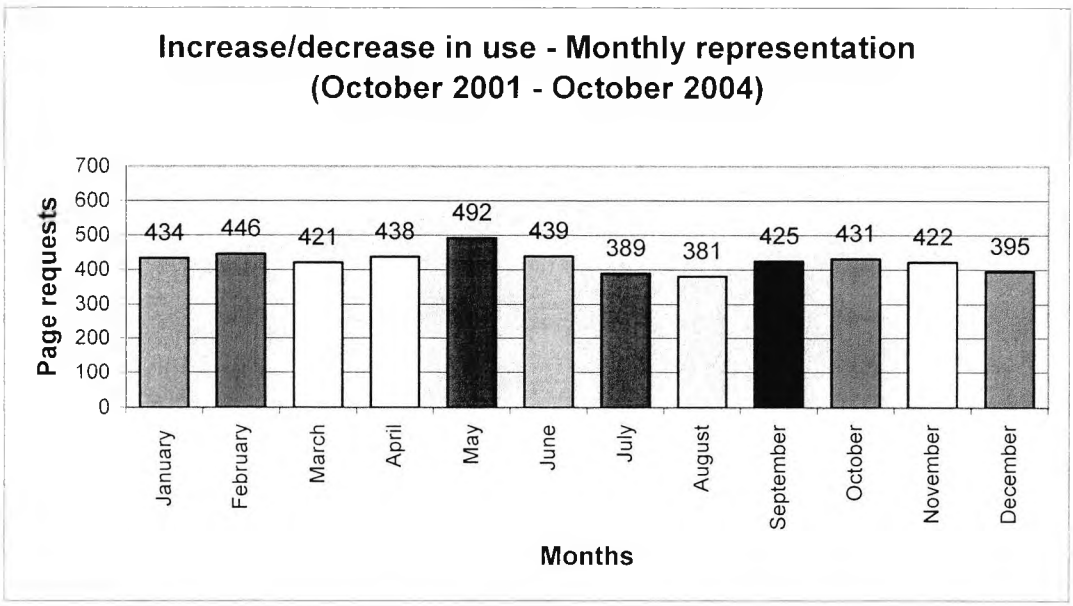


Figure 30: Results - System of Registries - Web log transactions - Increase/decrease in use by month (October 2001 - October 2004)

In order to show an accurate representation of the daily use during the years of 2002 to 2004 the three months at the end of 2001 were excluded. The daily use during the year 2002 had been stable with the exception of the month of May that a rise in use of almost one and a half times the average 430 counts that had been recorded during the rest of the months. For the years 2003 and 2004, a small but stable decrease in use had been recorded. As discussed above, in the year 2002 the Environmental Protection Agency introduced the new System of Registries that replaced the Environmental Data Registry. The decrease in use is believed to be associated with the changes in the EPA's website.

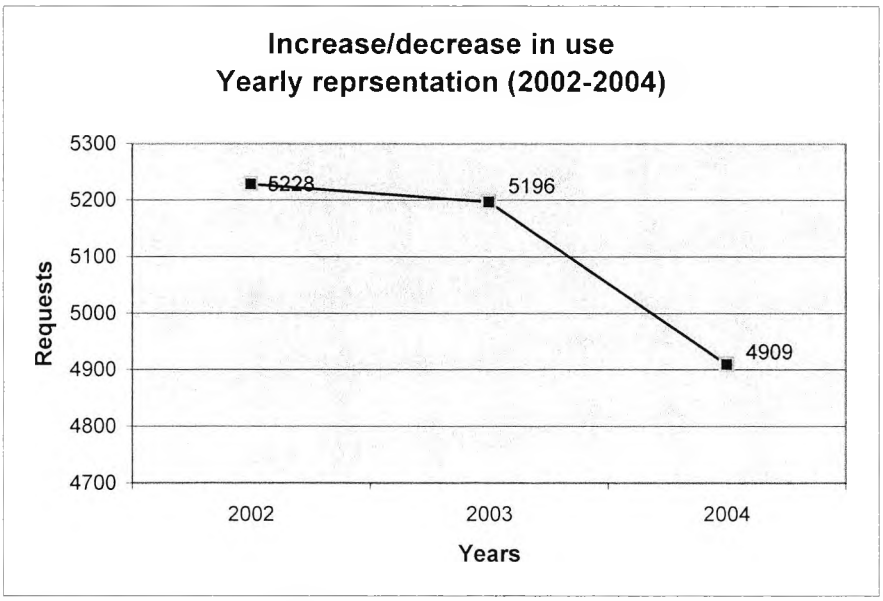


Figure 31: Results - System of Registries - Increase/decrease in use (October 1998-October 2004)

B.4.5 Use by type of information (Directory requests)

In the directory report the Analog software lists the directories from which the requested files were downloaded. In the case of the MetaForm registry, the web log transactions that were provided did not list subdirectories individually, but the numbers for each directory included all of its subdirectories. They also referred to the whole of the Metadata Server accesses. To facilitate the analysis, additionally to the main directories such as MetaGuide, MetaLib, main metadata server, similar directories were grouped together. For example, accesses to individual metadata projects were grouped together under the category projects. Category *"Other"* included graphics and/or diagrams, the glossary, and files of the new metadata server.

The largest number of requests targeted the Subject catalogue, which listed links to full text resources related to metadata such as formats, projects, evaluations and reports,

software, etc. The MetaForm registry, including all three services (crosswalks, crosscuts, mappings) accounted for 30293 requests for the whole six year period.

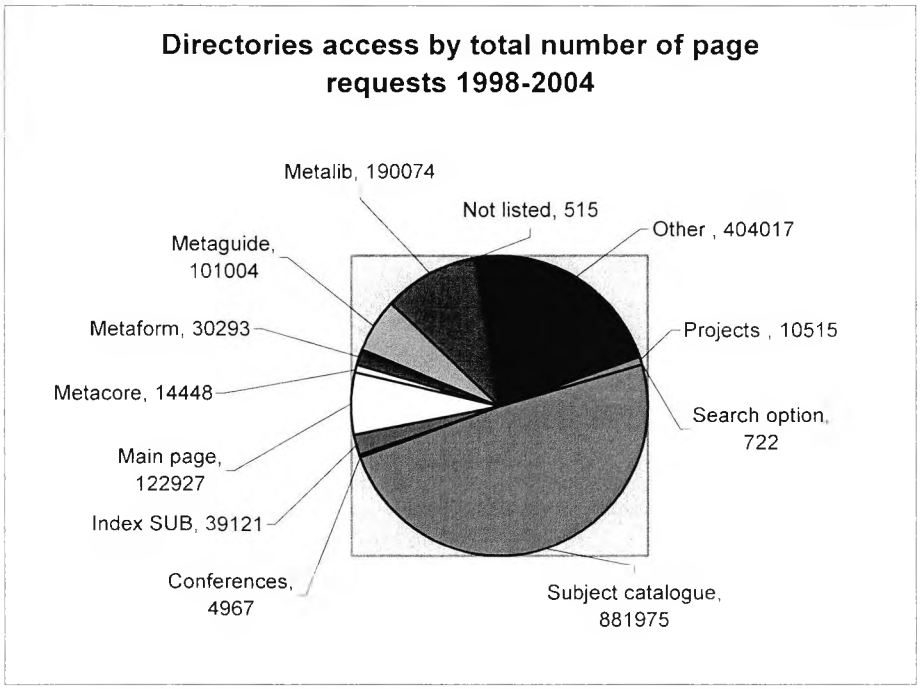


Figure 32: Results - MetaForm - Directory accesses 1998-2004

B.4.6 Use by type of information (Requests report)

The requests report records, which files, the System of Registries users had downloaded. Complementary to the where those files were stored, the requests report provides useful information to any system administrator as it shows which parts of the website are mostly accessed. The launch of the System of Registries in spring 2002 replaced the Environmental Data Registry as the gate to the EPA's environmental information and standards. Although the EDR remained as one of the registries and therefore maintaining the majority of its files stored in one directory, the launch of the System of Registries introduced changes in the web site directories and file storage. This change obviously affected the web log statistics and posed implications to the analysis.

In order to identify which registries of the System of Registries were mostly used it was decided to refer to the Uniform Resource Locators (URLs) structure and use the root of the URL to identify and group together downloaded files from the same registry. Each file (URL) was checked on a web browser to identify the page to what kind of resource it referred to. Files accessed before 2002 listing old or outdated URLs were the root of each URL was used as the basis to identify the registry to which it referred.

As well as the EDR, the CRS, the SRS and the TRS, the accesses that were clearly definable were measured separately. Those included use of the sitemap, search option,

the usage statistics; individual business rules documents, the download feature and other. Other included fact sheets, memorandum, the newsletters and information hosted by the Environmental Data Standards Council (EDSC). The most used registry appeared to be the EDR recording 327440 requests during the period from 1998 to 2004. The second busiest registry is the CRS (139583). As discussed earlier in the Part B, vast amounts of chemical research data such as crystallographic never become available in either the literature or any databases (Duke and colleagues, 2005) and therefore it could be of interest to relevant researchers, professional bodies, and commercial vendors. The next mostly used registries or resources in the System of Registries were the SRS (101098), those in category other (62409) and the TRS (56262).

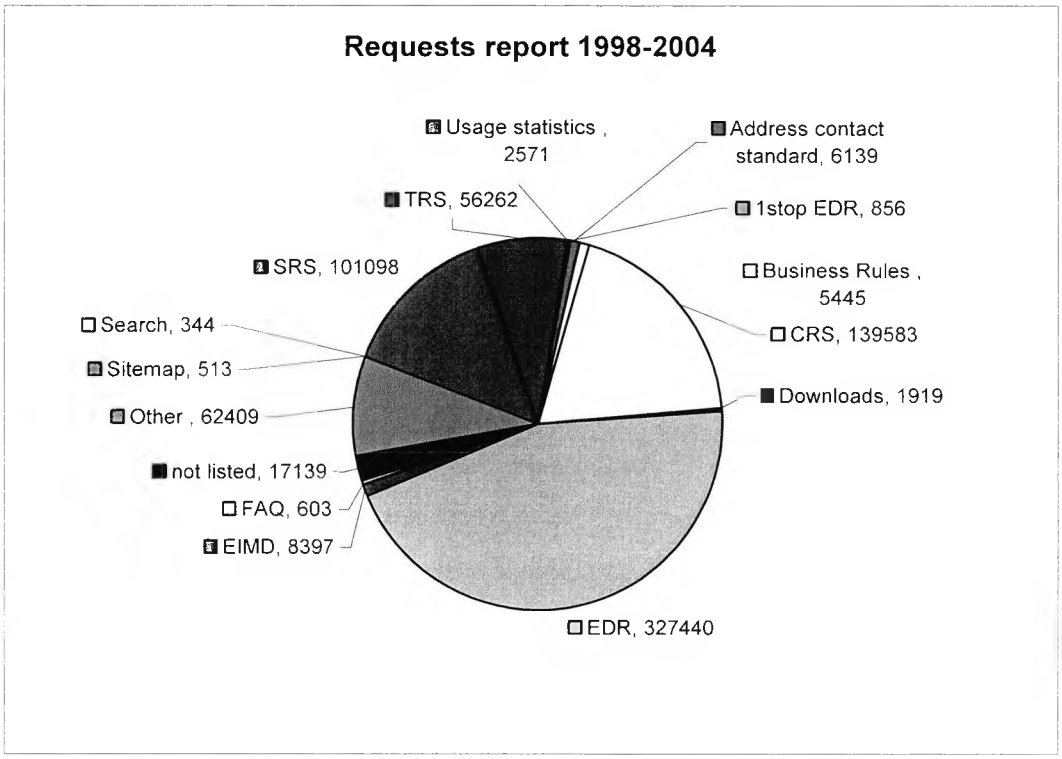


Figure 33: Results - EDR/SoR - Access by requests report (1998-2004)

B.4.7 Use of features/services

The respondents to the EDR questionnaire survey were asked to denote which of the EDR features and/or services they had used and to assess their usefulness. Usefulness was defined as the ability of the service to retrieve/provide effective information to the initial user query. A scale of five points was used in the assessment. The points from one to five corresponded to the following attributions:

Point 1 = Poor service

Point 2 = Adequate service

Point 3 = Good service

Point 4 = Very good service

Point 5 = Excellent service.

During 2002, the EDR listed on its web site the following features:

- Comments. Via this feature, the users were able to provide feedback about the EDR. It comprised from fields that users could enter their personal details such as name, email, organisation they represented and contact numbers and the comments field where they could enter their feedback about the EDR.
- Download. This feature enabled the EDR users to download and save the information they had retrieved in three different formats. Those were a) text report, b) Oracle (SQL Loader) and c) Comma separated (for MS office applications such as Access and Excel).
- FAQ. A section where the EDR provided replies to the most frequently asked questions.
- Glossary. This feature of the EDR listed in alphabetical order the terms and their respective definitions that were used in the EDR, the Substance Registry System and the Terminology Reference System.
- Help. Users could find relevant information about how to navigate through the many registries at SoR and how to conduct efficient searches.
- How to...facility. This feature provided information to system developers, standards implementers and people involved in data harmonisation about how the EDR could best help their work.
- MetaPro. MetaPro is licensed software developed by members of staff at the EPA and it is based on the ISO/IEC 11179, part 3 standard. It enables users to create their own metadata registries and is available for free distribution among governmental organisations on the provision that the copyright will be acknowledged every time the software is used. At the time of the 1st users' conference, MetaPro was still in production.
- Newsletter. The website of past and current newsletters of the EPA that they covered topics related to metadata, information registries, and data standards.
- Registration. This feature provided information about how and what kind of information users of the EDR could register to the metadata registry system.
- Search. The main search engine of the EDR.
- Subscription. This feature provided the option to register with the EDR in order to receive useful information about updates, upcoming events and other EDR news.
- Thesaurus. Internally developed list of acronyms and terms used in the EPA.
- What's new...link. This section of the EDR hosted news about the most recent developments and additions to the EDR.

- Workshops/Conferences information and announcements. Information about upcoming and past workshops and conferences of interest to the EDR users.
- XML. It provided information about the XML tags that could be used with the EPA data standards.

The most popular feature among the respondents was the Search Facility (63%). The least popular feature, indicated by only one of the respondents, was the MetaPro. The unpopularity of the MetaPro can be explained by taking into consideration that at the time that the questionnaire survey took place MetaPro was still under development and available for distribution to any other governmental organisations. Other preferred features included the Download, the How to... (42%) and the Help (37%) sections (Figure 34).

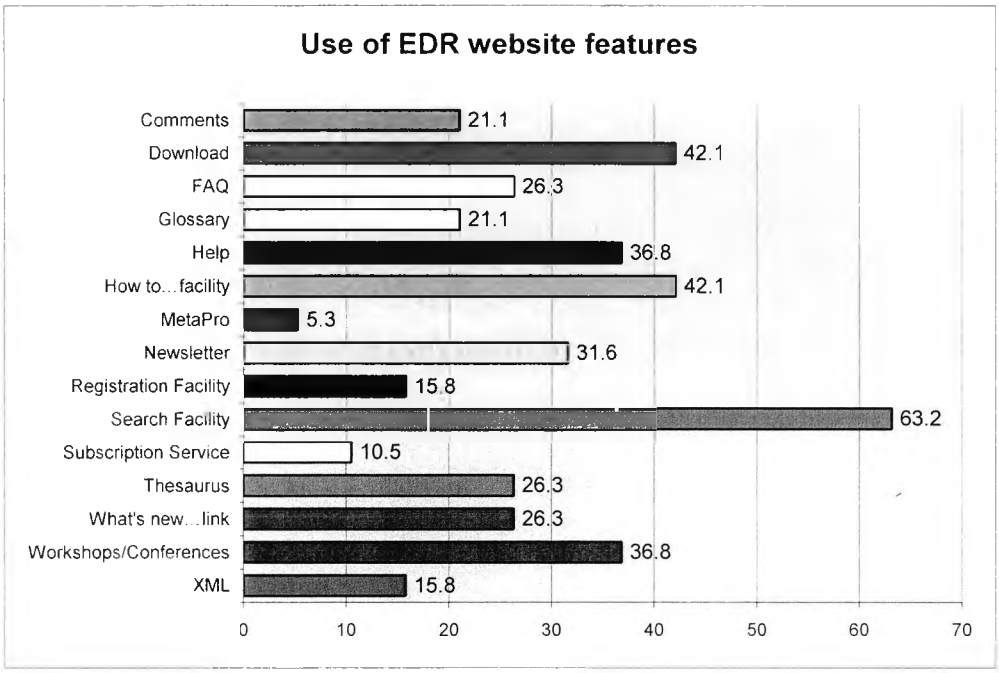


Figure 34: Results - Environmental Data Registry - Questionnaire survey A - Services used

Overall, the entire EDR website features were ranked above average with the exception of the search feature. The search facility, which had been the most popular EDR feature among the conference attendees, had been ranked by half of the respondents as a poor service. The EDR users commented the main reasons this feature has been ranked as poor were the lack of a text search option for the entire EDR website, Boolean support, a confusing interface and that the service's speed was slower than desired. Conversely, the download feature was considered to be a very good service by almost half of the respondents. The output and saving file options were thought to be satisfactory and only one of the respondents regarded it as poor. Overall, the organisation of the users'

conference and the issues that were addressed were considered to be an excellent event by the majority of respondents. The ranking of all EDR features is presented in Figure 35.

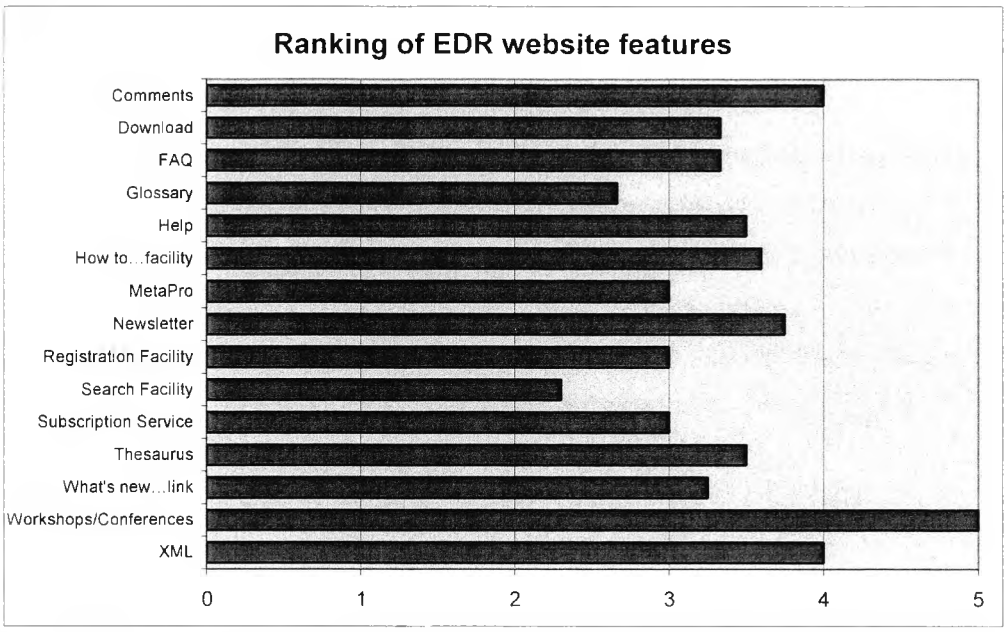


Figure 35: Results - Environmental Data Registry - Questionnaire Survey A - Ranking of website features

Additionally to the assessment of the usefulness of the EDR features, the users were asked to indicate if they were satisfied with the coverage of the information provided by the EDR. The majority of respondents (73%) replied that although they were satisfied with the coverage of the information in the EDR, they would welcome additions. Most appealing resources for future inclusion were data standards in general, data elements and the business rules documents. The business rules documents explain the context of use for each data standard and therefore are essential for the implementation of data standards. One of the respondents commented on general organisation policies and noted that they would be interested to see how *"How the Environmental Data Registry will integrate EPA's ability to enforce regulation[s]"* and how new services dedicated to certain areas of interest such as data harmonisation across the EPA, would be achieved. Furthermore, they declared an interest in *"... "future" services for harmonisation and XML e.g. list management"*.

The second System of Registries users' conference provided the opportunity to ask the System of Registries users if their preferences in the use of the services of the System of Registres provided had remained the same. The Environmental Data Registry, which in 2003 was a separate registry system (see Chapter 1 – Introduction), was identified as the most popular of the System of Registries applications by half of the respondents. One third (31%) of the respondents also declared their preference for the Facility Registry

System (FRS) and the Substances Registry System (SRS). The EDR is the main registry system in the System of Registries to search for data standards and their assorted business rules documents and also to download them (Figure 36). Surprisingly, more than a third of the respondents noted that it was their first time to use any of the System of Registries applications. Although last year's similar finding was justified by the conference's structure and theme that aimed to address a wide range of user familiarity and experience it can not be justified this year. The emphasis of the conference was placed on metadata standards implementation and the exchange information network that aimed to endorse the use of System of Registries even more. It appears that comments made from the respondents about lack of publicity for the conference and the remaining confusion about how data standardisation could be achieved is reflected in this finding.

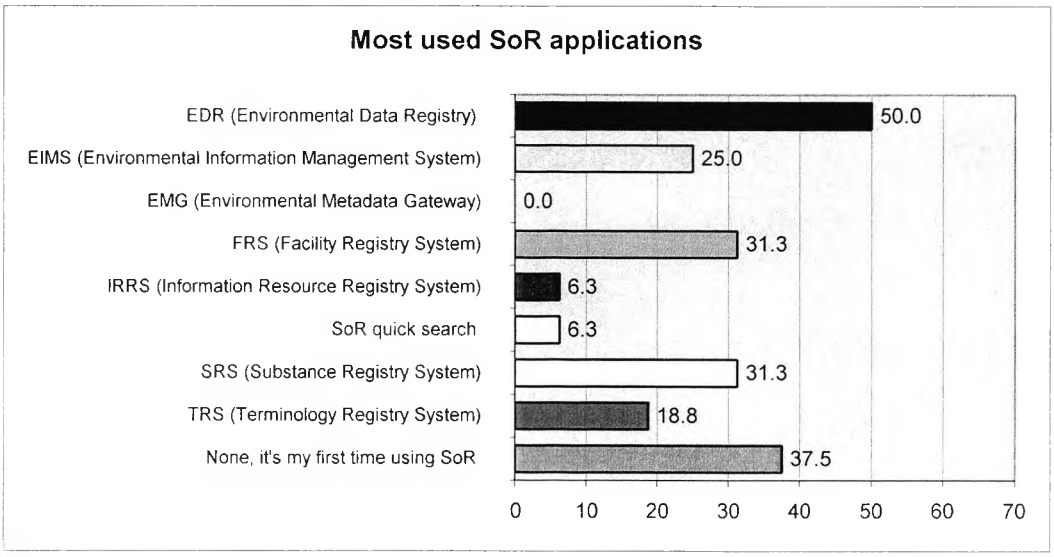


Figure 36: Results - System of Registries - Questionnaire survey B - Most used System of Registries applications

Similarly to the results from the previous year's survey, data standards (56%), data elements (44%) and their associated business rules (38%) represented the most sought after information on the System of Registries. The newly available XML tags were also quite popular (Figure 37). The respondents were given the option to provide any other information that was not included in the list of options provided. Those who noted *"Other"* in their response and similarly to the discussion lists, the SCHEMAS workshop and the MetaForm registry surveys, specified that they were interested in finding out about the relationships of data across different systems. The role of mapping across metadata element sets and how these affect the exchange of information and allows for interoperable systems to act was also noted.

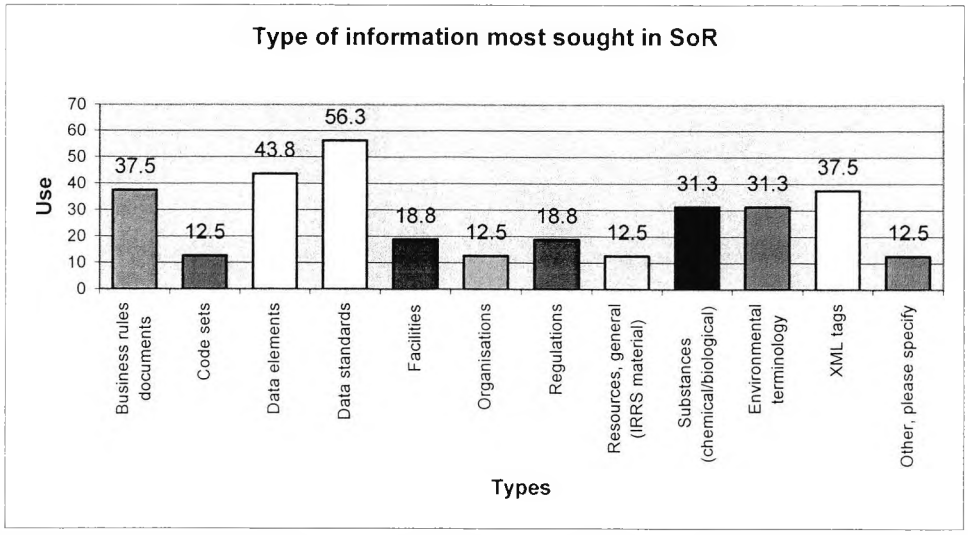


Figure 37: Results - System of Registries - Questionnaire survey B - Type of information most sought on System of Registries

The System of Registries users were asked to denote how they had used the information they retrieved from the System of Registries applications. Half of the respondents replied that the main reason they had used the System of Registries was to keep up to date with information resources in the EPA. Few people have used the service to register an information resource, download information and use information obtained in their work. Those who have searched and/or downloaded information from the System of Registries indicated a need to stay current with EPA information resources (50%) or to download information on multiple standards at once (38%). *"Other"* reasons included "to retrieve information on chemicals for use in risk assessment and regulatory compliance work" and "Audit, verification, meeting requirements". It appears that compared to the last year's replies the users were more aware of the EPA's policies about data standardisation and they seemed to access the System of Registries applications to identify relevant information. Results are presented in Figure 38.

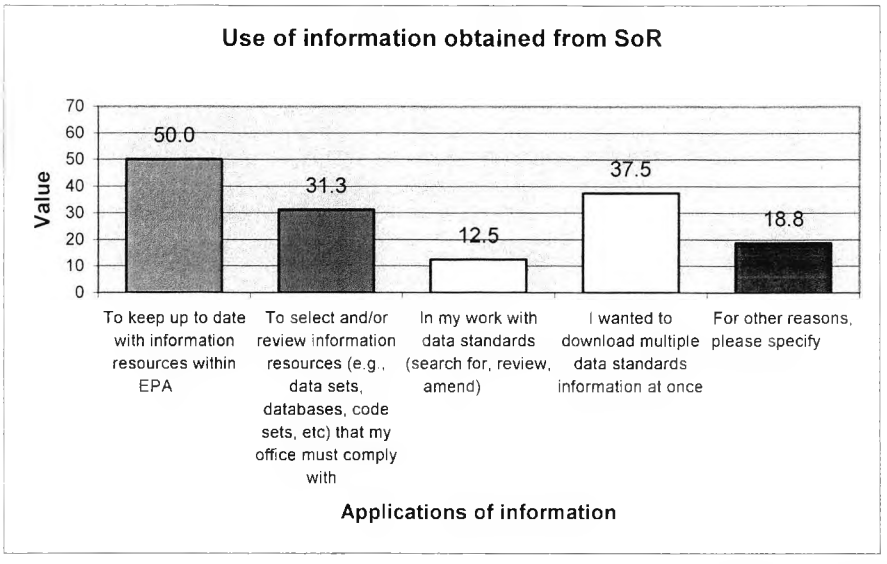


Figure 38: Results - System of Registries - Questionnaire Survey B - Information retrieved

B.5 Reasons for using metadata registry systems

The respondents to the discussion lists, the SCHEMAS workshop and the MetaForm registry surveys were invited to select from a list of option the reasons for which they had visited and used a metadata registry system. The list included the following options: for resource discovery/find out about relevant information, for data exchange, to find definitions of elements, for the description of text resources, for mapping of elements, to see dictionary structures (in particular for ISO/IEC 11179 implementations), and other. Almost three quarters of the respondents (73%) indicated that they visited metadata registry systems to look for the best applicable metadata for their needs and to see examples of how others have applied them in practice. Furthermore, more than half of the respondents who have used a metadata registry system indicated that they search for mappings of elements across different metadata element sets (64%). Crosswalks, crosscuts or mappings are some of the different terms that the connections between elements of different metadata element sets are known. These mappings enable people who use metadata element sets to make the best possible decisions about what element sets are best to use for their own applications and also to use combination of elements of different metadata element sets. Comparison of elements across diverse element sets is important to see how other element sets have been used in practice. The third most noted reason was to find out about definitions of elements (59%). One of the main functions of any metadata registry system is to support data standardisation and data documentation by allowing registration of metadata elements. Finding out about the structure of a metadata registry was noted as another reason for using metadata registry systems by one third of the respondents (32%). This reason was mostly cited by people who were involved in metadata registries implementations. People who indicated that

they use metadata registry systems for describing text resources value the role of metadata registries as authoritative system for storing and maintaining metadata information (36%). Other reasons stated were "*to check on progress of registries based on ISO/IEC 11179*". This comment was made by one of the respondents who were involved in the implementation of a metadata registry system. Discovery and exchange of metadata were also noted as reasons for using registries but not as the primer ones.

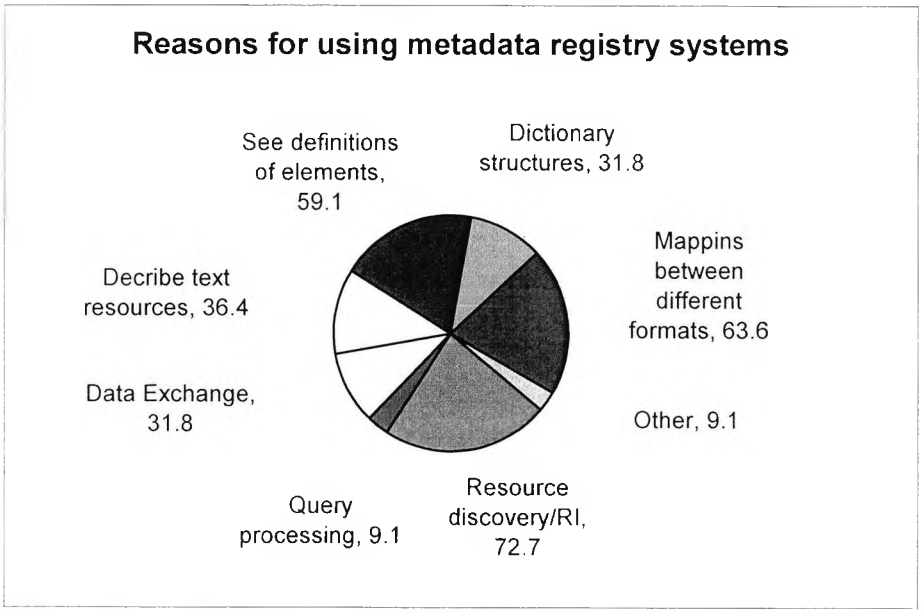


Figure 39: Results - Discussion Lists, SCHEMAS workshop and MetaForm registry surveys - Reasons for using metadata registry systems

Response from the interviewees provided rich information regarding the reasons people use metadata registry systems. As discussed in Chapter 2 – Literature, metadata registry systems are tools that enable the management of metadata by specifying data elements and describing their structure and associated meaning (semantics). Interviewees were asked why they had used metadata registry systems and also why, if that was the case, they thought people should use them in order to identify the potential benefits. Their response met the above prescribed role of those systems as this is stated in the literature. The reasons that the respondents expressed showed that they considered metadata registry systems as essential authoritative sources of information that ensure a standardised approach to the use, management and exchange of information. They emphasised their role in avoiding duplication of this effort and being a means to achieve interoperable information systems. In particular the following responses were noted:

- **Authoritative source of information**. Response from the interviewees emphasised the advantages associated with having one single source of registration that eliminates duplication of effort at any institution and ensures quality control of the information that is stored there. Functionality of the system that permits the

registration of other than known and used metadata element sets such as locally developed and used data that meet specific needs in an institution, is considered important by the users. For example, one of the interviewees noted that such a system *"...allows us to have central register of variables – not to ask the same things in different researchers"* and also providing the flexibility to register not only certain metadata element sets but to be able to *"Document[...] specific application profiles"*. The use of international standards in the development of metadata registries also ensures that they facilitate the registration of elements of different metadata element sets (also known as application profiles). The use of metadata for the description of digital information had been largely characterised by this trend (Heery & Patel, 2000) and is considered common practice when describing resources that usually aim to meet needs of interdisciplinary users. As another interviewee said *"It is the only way to standardize things (statistical process, vocabulary, variables, questionnaires..."*. Furthermore, response also stressed the importance of having one single point of reference for all employees of an organisation that would enhance a common standardised approach in the maintenance, use and exchange of metadata information. It ensures enforcement of policies as everyone knows where to refer for authoritative information. In particular, one of the respondents with a background in data standardisation explains how metadata registry systems could support their work "our main interest in metadata registries is in the future development, rather than in use of particular registries. The principal application that I see over the field is that users will have a single convenient source for metadata elements. The existence of a register of metadata will help ensure that different suppliers of metadata use the same elements to represent the same concepts. Those benefits will out weight the costs over the user and supplier communities."

- **Management of semantics.** The availability of an authoritative, centrally administered system that is developed based on international standards that ensure the quality and long term preservation of information by the use of controlled lists, guarantee the quality and added value to the information. It was suggested that the use of *"...data category identifiers as reference points to ensure semantic interoperability across applications"* is among the most important factors for use across respondents with an interest in data standardisation. Having access to administrative metadata and therefore being able to monitor the day, time, creator, version of resource, etc provides access to information that is difficult and expensive to maintain with large organisations that share data and rely on their quality and value for their every day work. As one of the respondents working with statistical information noted *"it is the only way to do document our statistical process and check where is it, where it stopped and found out who is responsible for this step"*.

- **Record and disseminate data standards.** The benefits that any organisation could achieve by recording and disseminating the data standards they use in their every day work had been noted by responded as the reasons why they had used or would encourage the use of metadata registries. One of the respondents emphasised "*The main reason for using the [organisation omitted] registry is the re-use of data-definitions and models...Large scale re-use of data models could potentially save months of effort at integration time.*" And another stressed the importance of cost effectiveness in the use and re use of data that can be accomplished with the use of metadata registries and the benefits deriving from good organisation. "*...a better self-organisation and information exchange with public authorities to increase awareness and reduce cost for reusing information and data already available.*" Also, such systems could ensure that they present a record of the work that each department or employee is conducting and their given expertise in a subject area and act as one of the solutions regarding interoperability and exchange of information that could "*...provide tools for researchers to search information and to disseminate information of their own work*". Respondents with a background in data standardisation emphasised their interest in future development and population of metadata registries. Also, maintenance of registered metadata and accessibility to information would ensure the "Preserv[ation] of data after end of project and to make publications more accessible." Finally, an enthusiastic user of metadata registries noted only one reason for using such systems. In their own words "*...registries are the means of discovery, promoting re-use and integration of work between related projects*".
- **Other.** The category other listed reasons, also confirmed in the questionnaire surveys, such as exploring or identify functions and services of metadata registries already implemented in other institutions. They usually represent system developers and/or data implementers that they are seeking for more information about how they could use metadata registries to best support their own institutional needs, research interests and work requirements. For example, one respondent noted that they "*...use the registry not for actual applications, but to understand the functionalities of registry and help other Federal Government in the use of the registry*" and another respondent explained that their organisation did not have requirement for a metadata registry system at the time but as "*one of the new European Law to provide information on the environment to the public and to report regularly to the European Commission*" it was essential to understand how they could be used to support their work. Also, for some of the respondents it was their organisation's requirement to employ a metadata registry system to facilitate the use of data standards across that organisation so it was essential for them to use them. "*The agency decided that it*

required a data standards program and as a result we determined that in order to manage this program a registry was deemed required. The driving force for the registry was the data standards program." While for others the reasons to use metadata registry systems is simply because they believe in the role they could play in addressing data sharing and quality at affordable costs. In their own words, *"The cumulative effect of a working registry is to support seamless door-to-door services though integration of open systems. Without a registry this goal will be achieved later and at greater cost, as various organisations slowly find out how to integrate fragments of the overall service".*

B.7 Requirement for use by the organisation

The respondents to the discussion lists and the SCHEMAS workshop surveys were asked to indicate if their organisation had a requirement for a metadata registry system and if so to describe in what way. This question aimed to gather some information about how organisations envisioned the harmonisation of a metadata registry system in their operations. One third of the respondents replied that their organisation did not have a requirement to use a metadata registry system. More than one quarter of the of the response (27%) indicated that the users did not know whether there was such a requirement set by their organisation and all of the respondents who had used metadata registry systems said that it was their organisation's requirement. In particular, they provided the following comments:

- For reasons of standardisation that an authenticated source of information provides such as "...[to] ensure controlled vocabulary lists are use".
- Also, because their organisation had decided on the use of a registry, e.g. *"BSR will be use in the [project omitted] project" and*
- *"Yes, as organisations may want to register their additions to LOM".*
- Furthermore, for reasons pertaining to data preservation "We need to use our own metadata registry. This is to describe material, which we are collecting and archiving as part of a "proof of concept" for Digital preservation" and data quality "For projects in which several organisations are involved (also in other countries) with the ambition to improve and extend quality of metadata, the value of a metadata registry is obvious".
- And last to "to assist in metadata creation."

Half of the respondents to the MetaForm registry survey noted that their organisation did not have a requirement for a metadata registry system but they still accessed and used a registry because the projects they were working on required that they used metadata.

Although there was no specification as of what project the second most noted application where metadata had been applied was in the description of library resources. Therefore it can be assumed that there must have been a direct association in the use of metadata and metadata registry systems with digital resources and their management. Contradictory to this statement of no organisational requirement for a metadata registry system, one third of the respondents indicated that their organisations contributed to the development and the population of the MetaForm registry by adding content. In particular, they specified that they contributed in the development and population of the registry by registering metadata elements. Those who replied that their organisation had a requirement for a metadata registry specified as in the form of *"the SUB Göttingen"* and *"yes, I have built one [metadata registry system]"*.

The majority (74%) of the respondents to the EDR survey replied that the organisation they represented had a requirement for a metadata registry system. In particular they specified:

- *As part of the Ent(erprise's) Architecture*
- *EPA (YES) for office work (NO)*
- *[I am an] EPA Employee*
- *EPA Systems*
- *Evaluation of DoD (Department of Defence) data standardisation (8320)*
- *I am on the Agency's Quality Staff that develop policy and procedures for Data re-use (e.g. EPA Order 5360.1). Data needs context and to be evaluated for re-use. Registries are needed to access metadata needed for data evaluation and to comply with content standards like FBDC.*
- *Metadata for OW Info resources; Metadata for OW Datasets*
- *Only to use it for audit*
- *System must comply with EPA standards*
- *Under Canadian Environment Protect Act, the Government of Canada must build and maintain a registry (EPA registry)*

B.9 Metadata registry systems applications

Metadata research applications can be found in many and diverse scientific domains (SCHEMAS Standards Framework 2000, 2001, 2002). Consequently people involved in metadata use and application represent a wide range of scientific background. This finding is encouraging for the deployment of metadata registries and the interest that had been generated in them. Web log transactions had also confirm this finding in the form of

the different domains that had used metadata registries (,). Furthermore, referrer services to metadata registry systems as identified in the web log statistics also note an increase in use after the year 2000. In particular access via search engines and other referrers such as academic institutions, commercial websites and health and/or environmental organizations has risen even though the general use of the two registries that were analysed had dropped (Figure 12).

Interviewees represented a wide range of scientific domains. They ranged from a meteorologist to software engineer, database systems developer, statistician, US federal government employees, web-site developer, chemist, head of academic library services, registrar, researchers in the area of stellar structure and evolution and members of the International Data Standardisation Organisation. [27] Therefore, it was not surprising to see that they noted the application of metadata registry systems in the following domains:

- The academic environment, e.g. Libraries.
- The Government domain, e.g. Federal government, Statistical Offices.
- The research domain, e.g. National Scientific Institute and
- International organisations.

Deriving from their job responsibilities, interviewees provided a wealth of information about the various tasks that they had dealt when working with metadata registry systems. Such feedback is representative the potential applications and use of metadata registry systems. Response was grouped in three categories similarly to those describing the job responsibilities of the respondents.

People engaged in ***metadata management and/or metadata implementation*** noted working generally with the development of "*...a corporate information portal that provides integrated access to disparate and distributed sources of data and metadata*" and "*...information dissemination tools and portals within the forest domain*" and "*...d*eveloping metadata registry systems for the [country omitted] Statistical Office Development of data management systems for the support of data grids, digital libraries and persistent archives". But they also noted specific tasks such as the development of collection policies for the content of the metadata registry systems and creating the cataloguing practices and/or business rules documents for the use of the data standards. Developing the collection policy for a metadata registry system addressed issues such as the number and the type of information that would be included in a registry system and the access rights to this content. Furthermore other tasks were associated the quality

[27] Although some of the respondents provided information about their professional background, for those who did not the researcher conducted a search either in their home institute and/or the World Wide Web to find out more information.

control and assessment of the registry services, ensuring that users have access to information that meets their requirements. In particular it was noted, *"...management and evaluation of quality of services (both collection and technical)"* and *"metadata management for websites usability and metadata issues of organisation's website"*

Those with a background and interest in ***data standardisation (DS***) declared tasks associated with the semantics aspects of metadata and its use. The focus was mainly in the development of tools that support the role of metadata registries as authoritative sources of information such as ontologies, mappings of different elements of metadata element sets and compliance to international standards. They noted tasks such as the creation and management of controlled vocabularies, development of crosswalks and ensuring *"the establishment of data quality and a data standards program within a government agency"*. Other noted tasks included *"the design and/or application of existing metadata frameworks to a specific project, e.g. application of XML for metadata design"*

Those in **software development (SD)** noted the most tangible tasks such as the *"design and/or application of existing metadata frameworks to a specific project, e.g. applications of ML for metadata design"*. Also tasks associated with the support of the semantics of the registries such as *"developing applications for the management of classification [schemes]"* and *"Implementation of a registry to facilitate discovery and re-use of ontologies and sources of metadata within an organisation"*. But also a wide range of tasks dedicated to the development and deployment of metadata registries such as *"...development of environmental registry systems and in particular searching technologies, via full text option, controlled vocabulary, time restrictions and geographic regions, as well as editor technologies and organisational aspects of metadata registries"*. Other tasks included:

- Interoperability testing technology
- M2M interfaces
- Smart card technology
- XML registry framework and standards

In Table 26, grouped under the four categories of technical aspects, semantics, user support services and evaluation/quality assurance are the tasks that the interviewees had worked on.

Categories	Applications
Technical aspects	Design and/or application of existing metadata frameworks to a specific project, e.g. applications of ML for metadata design Developing applications for the management of classification [schemes] Implementation of a registry to facilitate discovery and re-use of ontologies and sources of metadata within an organisation Interoperability testing technology M2m interfaces Smart card technology XML registry framework and standards Developing metadata registry systems for the [country omitted] Statistical Office Development of data management systems for the support of data grids, digital libraries and persistent archives
Semantics	Creation and management of controlled vocabularies Design and/or application of existing metadata frameworks to a specific project, e.g. application of XML for metadata design Development of crosswalks Metadata management for websites
User support services	Creating collection policies Developing a corporate information portal that provides integrated access to disparate and distributed sources of data and metadata Development of environmental registry systems and in particular searching technologies, via full text option, controlled vocabulary, time restrictions and geographic regions, as well as editor technologies and organisational aspects of metadata registries Writing of cataloguing practices and /or business rules
Evaluation/Quality assurance	Establishment of a data quality and a data standards program within a government agency. Management and evaluation of quality of services (both collection and technical) Usability and metadata issues of organisation's website. Also, development of information dissemination tools and portals within the forest domain.

Table 26: Results - Email interviews - Projects

Almost half of the respondents to the EDR survey had used the retrieved information to meet needs of their programs. Some of the specifications highlighted the organisation's policy by comments as specific as "*I use the EDR information to ensure the systems I develop conform to EPA standards" and* "the *application of data standards to conform to systems development".* The EPA's developments in data standardisation attracted interest from other governmental organisations such as the Environment Canada. Other governmental organisation expressed interest in the "[*use of the EDR] services, as Environment Canada is categorising 23.000 substances on persistence, bioaccumulation and toxicity*". Three respondents noted that they have used information from the EDR for their systems development work, or for [enterprise] architecture research. Category *"Other"* (21%) included responses such as:

- Audit of data
- Incorporated into system development facility registry

- To check details of data standards as needed for meetings on systems development, standard implementation and data integration
- For data standards implementation in NARSTO and Environment

The applications of the EDR retrieved information are presented in Table 27.

Application of retrieved information
I have been testing registries and how they work CRS System development Just getting started One did not specify
Within my organisational commitments Limited, not that should change I use the EDR information to ensure the systems I develop conform to EPA standards Applying data standards to system development Just learning about these Environment Canada is categorising 23.000 substances on persistence, bioaccumulation and toxicity Systems development and standards implementation Developing systems PPT systems development Two people did not specified
For my research System development Audit of data Agency architecture research
Other (please specify) Audit of data Incorporated into system development facility registry To check details of data standards as needed for meetings on systems development, standard implementation and data integration For data standards implementation in NARSTO and Environment

Table 27: Results - EDR - Applications of retrieved information from the EDR

B.10 Metadata registry systems – applications in use

Respondents to the email interview invitation provided with examples of metadata registry systems that they have either used, designed, developed or currently developing. They provided of the most interesting findings of this research that were the identification of current implementations in use (Table 26). The responses indicated a variety of different applications in different domains. Responses were grouped in three categories:

- Business. The business category listed references software and/or data management middleware for the support and management of distributed data collections such as the Electronic Business eXtensible Markup Language (EbXLM) and its implementation Yellow Dragon software (electronic business XML) and the Storage Resource Broker.
- Governmental. The governmental domain included references to many well known implementations of the ISO/IEC 11179 standard for data managements such as the EDR/SoR, the Unites States Health Information Database (USHIK) that is *"...a metadata registry established to assist in cataloging and harmonizing data elements*

across healthcare Standards Development Organizations (SDO) and other interested healthcare organizations"[28] and the Federal Aviation Data Registry. Also, one of the first ISO/IEC 11179 standard implementations, the Australian Institute of Health and Welfare Knowledebase. Further more, the UK Highways Agency data registry that presents a case of metadata registry use for transportation, the Swiss Federal Statistics Office metadata registry and finally the US National Cancer Institute metadata registry were also among the metadata registry systems that the respondents either used or developed. Also, there was a reference to Umweltdatenkatalog UDK, an information system for locating environmental data and its assorted metadata held by public authorities across 15 states in Germany.

- Other. Category other lists international standard that are the back bone of the metadata registry implementations such as the ISO morpho-syntactic data registry, the Dublin Core metadata registry, NOKIS (North Sea and Coastal Sea information system), ISO/TC 37 data category registry, IEEE-ITS and the Global Forest Information Service (GFIS)

Categories	Applications
Business	EbXLM registry Yellow Dragon (electronic business XML) Storage Resource Broker CCLRC
Governmental	USEPA-EDR-SoR USHIK FAA-FDR AIHW Knowledebase Highways Agency data registry (HA) SFSO National Cancer Institute
Other	ISO morpho-syntactic data registry Dublin Core NOKIS METIS ISO/TC 37 data category registry IEEE-ITS Global Forest Information Service (GFIS)

Table 28: Results - Email Interviews - Applications in use of metadata registry systems views

[28] USHIK. Information available at: http:// www.ushik.org/ (Last visited, 12/06/2006)

PART C – Functionality of metadata registry systems

The following section discusses the features of metadata registry systems that the respondents considered essential in such systems, their preference and assessment over their usefulness, and the barriers that prevented them from using them. Furthermore, it discusses the respondents' recommendations for future metadata registry systems services. This section discusses non use and complements findings from the respondents' feedback with results from web log transactions that recorded the number and type of failed requests of the registry systems. The notion of interoperability is also discussed by the respondents in particular the email interviewees.

C.1 Essential features/services of metadata registry systems

Metadata registry systems are registration authorities that aim to support the maintenance, the cross searching of metadata elements and the exchange of metadata information between different networked systems.

Interviewees were asked to indicate what they considered to be the most important features in a metadata registry system. Examples included xml validation; advanced search options, open access software, implementation guides, terminology from an authoritative source and discussion forums. The features and/or services that respondents indicated as most desired reflected the very notions that metadata registry systems by definition support. In particular, they all stressed the importance of having an authority control agency that looks after the management and maintenance of metadata. One of the respondents wrote *"For each of those systems, the consistent management of metadata about the registered material is crucial"*.

Maintenance of the registered metadata should be applicable to all types of metadata such as*"...system attributes (....[e.g.]where the data is stored, audit trails, checksums, access controls) and descriptive metadata (such as Dublin core, user-defined attributes)"*. Facilitating cross searching of metadata elements requires the consistent assignment of control elements. One of the respondents in the description of the services that their institutional metadata registry will support noted, "*We are now working on ways to dynamically assign consistency and control constraints. This requires building a persistent name space for the constraints, evaluating levels of aggregation of constraints, and dynamic application of constraints to the data collections. Applications include constraints on data placement, views on metadata, transformations on data, access*

controls, etc." Exchange of metadata information requires a combination of technical and non technical support systems. The respondents to the email invitations also emphasised the importance of being able to import and export metadata information by use of XML, under an RDF framework and by the establishment of a semantic framework.

Response to this question has been grouped in three categories:

1. User support services
2. Search options/functionality and
3. Interoperability support

User support services referred to user guides, business rules or user documentation and site maps. The application of metadata triggered the need for practice examples. Some people were interested in adapting the same practices as other institutes in their everyday work, for example check for cataloguing practices in use. User guides or business rules are the accompanying documentation that describes how to use specific data elements. They define the context of their use and give explicit examples of applied practice. Other users tend to visit registries to see how a system is set up and what attributes would suit their own organisational needs. One of the respondents with experience in the use of the Dublin Core and the features that this metadata registry provides noted that it was important that the registry has the "*ability to provide attribution to definitions (for example if the definition came from a professional society or book)*" and "*possibly access to various cataloguing best practices so that other groups know what other groups do with a particular element*". Governmental organisations rate the "business rules" documents as very important and response to questionnaire surveys confirmed this finding. Certain groups of users emphasise the guidance that is more applicable to their own needs. For example system implementers stressed the importance of implementation guides and support provided throughout the implementation process.

Search options/functionality *referred to the search and/or browse features and the availability of text search. The nature of the organisation and the services that each organisation wants to provide influence the significance attributed to each feature. One of the respondents that their registry is supported by the Storage Resource Broker, a software application primarily used in business and engineering applications, noted the features that their registry includes "...metadata extraction from files, bulk load of metadata, import and export of metadata from XLM files, automated SQL generation, schema extension, access controls on metadata, federation of registries. The services are designed to work through firewalls" though rated the search/browse availability and "business rules" as the most important features that a metadata registry system should*

have. Suggestions regarding the search/browse facilities and their associated role with functionality of systems included the differentiation but support for both humans and machines information retrieval. Additionally, the content and its structure which at some extent is dependent on the software that enables the population of the metadata registry, determines the kind of searches that can be performed. Also, it was suggested that the implementation of "classification schema/ontologies" will improve the search/browse facilities of any metadata registry and provide the basis for interoperable systems. Furthermore, accessibility options with particular emphasis on the extension of the search/browse features by "...link[ing] your own registry/database to a search engine is a great plus to be interested in a registry." Ease of use is also noted as an essential feature. Data and data management services are crucial.

Interoperability support referred to features that enable systems to exchange information and communication in a useful and effective manner. Those features included the facilitation of import and export of metadata, availability of mappings, and use of XML for the exportation of data. One of the respondents specified, *"I am interested in the development of standards for metadata registries that manage information that documents the semantics of data and concepts for use by business, not just the information technology sector...I liken semantic metadata...where information is organized by concepts rather than terms...the concepts may have sets of permitted meanings; the concepts may have multiple representations; the sets of valid meanings may have multiple presentation forms..."*

Another respondent noted how their organisation's metadata registry was made conformant to ISO 11179 in order to provide features such as *"...online browsing and querying access, private area to select subsets, import facilities (XML, HTML), online submission facilities, online voting facilities for appointed experts..."* Depending on the domain that each registry aims to serve there are different requirements and emphasis is placed in different features. For example, the [organisation omitted] registry is dedicated to the language resource domains, with basically three types of data categories:

- linguistic descriptors that will directly refer to language related
- concepts: e.g. partOfSpeech/, etc.
- metadata descriptors to identify and document language resources or tools: e.g. /date Inputted/ /role/, /translator/
- administrative data categories: /projectSubset/, /lastModified/, etc.

Conformance to standards, merging lexicons (for metadata registry systems with particular applications in linguistics), improvement of interoperability between different metadata registries, "provenance information" and administration of data elements were all deemed essential by one respondent.

C.2 Secondary services of metadata registry systems

The interviewees were invited to denote those features and/or services of metadata registry systems that they considered of being overstated when it came to the use of metadata registry systems. Their response noted the omnipotence of XML and the "hype" of publishing data. In particular, an interviewee noted that what seems to have been overstated about metadata registry systems is that they *"...solve all the problems related to metadata, e.g. data element definitions, taxonomy, XML schema registry and management."* In fact, current implementations have proved that the whole process of implementing a metadata registry system is not an easy thing to do. It is not the case of *"you buy and install a registry and everyone uses it" or that the "registry "stands "alone". Without a formalized data standards process that is mandated, the registry is just another reference database of information that falls into disuse."* Support from the organisation and a policy that mandates the use of a standardised process to transfer and exchange information is essential.

C.3 Functionality of metadata registry systems

This section discusses results pertaining to the functionality of the MetaForm registry. Respondents were asked to indicate which of the three MetaForm services they have used and to assess how satisfactory they considered them to be based on the criterion of finding the information they needed in an efficient and effective manner.

Response showed that the respondents used all three features that the MetaForm provides. Amongst the three services, the mappings proved the most popular among respondents, followed by the crosswalks and the crosscuts (see Chapter 1 – Introduction). The popularity of mappings had also been confirmed via findings from the discussion lists and the SCHEMAS workshop surveys where respondents listed that finding mappings across metadata element sets was the second most popular reason for using a metadata registry system. Respondents were asked to rate all three MetaForm services using a scale of one to five. One (1) represented the least satisfactory service and five (5) represented the most satisfactory service. The respondents rated all three services with above average rates. They assigned the mappings with 3.6 (average number), the crosscuts with 3.4 and the crosswalks 3.3.

Furthermore, they were also invited to indicate how satisfied they were with the interface of the MetaForm registry and the way the information was stored, presented and made

available to its users. The response dropped to half when they were asked about the content and the semantic representation of the information in store.

Respondents were invited to denote how satisfied they were with the following six features related to the MetaForm registry's interface and the functions performed across several of its services. Those were: coverage of information in general, coverage of metadata element sets, linking between different metadata element sets, mappings between metadata element sets, relevance of retrieved information and feasibility of retrieving information in more than one formats. Similarly to the assessment of the three main MetaForm services, the respondents were asked to use a scale of one to five to indicate their satisfaction. One (1) represented the least satisfactory service and five (5) represented the most satisfactory service. All services were rated as above average which indicates that users were generally happy with the layout of information, the user interface, the linking and mapping facilities and the relevance of the retrieved information. In fact, the highest rated features at 3.5 rate were the coverage of information and the relevance of the information the users had retrieved from the MetaForm registry (Figure 40).

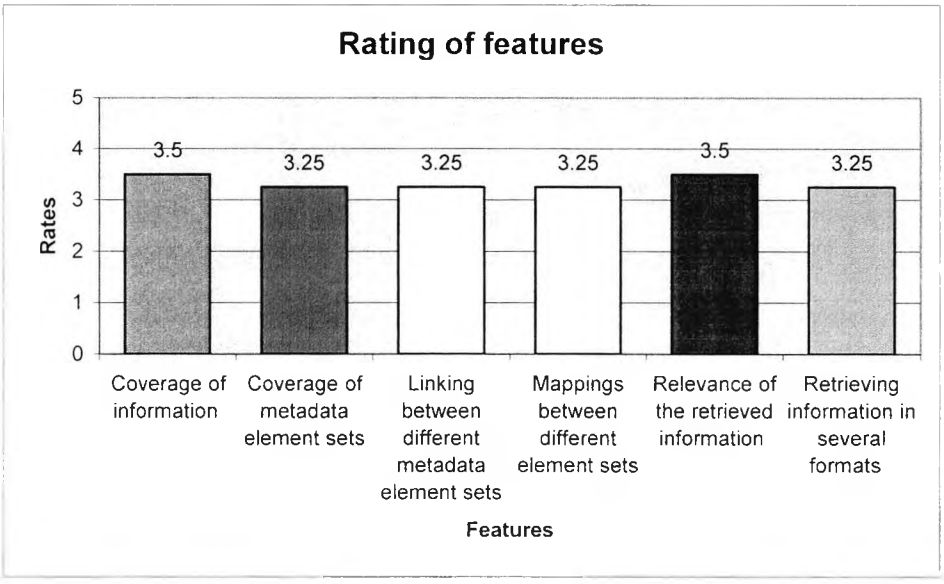


Figure 40: Results - MetaForm - Questionnaire survey - Rating of features

Similarly to the MetaForm registry's users, the Environmental Data Registry users were also invited to assess both the content and the interface of the EDR. They were asked to denote how easy or difficult they encountered the process of searching for information in the EDR. Also, to assess the relevance of the information they retrieved and manifest their satisfaction about the help guides and other support documentation in EDR such as the glossary and the site map. Furthermore, they were provided with a list of features that EDR was planning to implement in the near future and asked to note their preferences.

Almost half of the respondents found and downloaded the information they sought relatively easy but it was noted that more guidance would be helpful. This finding is consistent with the ranking results pertaining to the search facility (see previous section in this chapter), which were found to be poor by half of the respondents. The respondents commented that it was easy and straightforward to find useful information *"...except when the search did not work"* and that more guidance would be appreciated as there were *"too many links that made it difficult to grasp the various features"*. Three respondents indicated that they were unable to find what they wanted and they commented that more help features and events such as the user conference would make it easier to understand and use the EDR.

More than two-thirds of respondents though were satisfied with the relevance of the information they retrieved from the Environmental Data Registry specifying that it was either relevant or adequately satisfying their query. Respondents were asked to rate the content of the mostly sought resources on EDR; the business rules documents, data standards, and other information resources held in the EDR such as internal memorandum, fact sheets and policy documents. Surprisingly, almost half of the respondents did not rank the content of the Business Rules Documents, even though it was one information resource that they indicated strong interest in. In general, they commented that inclusion of more data standards and information would be welcome. The EDR users were also asked to assess the interface of the Comparison Matrix and to note their satisfaction or otherwise towards the services provided to help them through their searching. They were called supportive to emphasise their role in providing help and guidance while using the EDR. Those services included the glossary, the help guide, the search facility, and the site map. All of the respondents considered the comparison matrix as a satisfactory or adequate feature and a full two-thirds (68%) replied that while they were generally satisfied with the services provided, they would welcome additional improvements. Again, the search facility was identified as the service that needed improvement by almost half the respondents, followed by the site map and the glossary. Recommendations from respondents included:

- Needs to standardise terms
- [Incorporate] Text search for entire site
- Organisation search should yield ordered (alphabetical?) results
- Add systematic diagram
- Need more context sub settings help- e.g., click on a term & get the definition, an example, etc.
- [Improve] Navigation - Hard to get back to where you came from

- Improved search function; acronyms
- Duplicate both commercial and free websites, most of which are faster.
- CRS "not responsive" / Union intersection; public is aware of this?

All of the respondents noted that they were happy with the services that MetaForm provided but they would welcome additions such as metadata creation tools and metadata conversion tools. Also, larger coverage of information whether that being more metadata element sets, guidelines for their use and suggested external sites were also indicated as desired features to add.

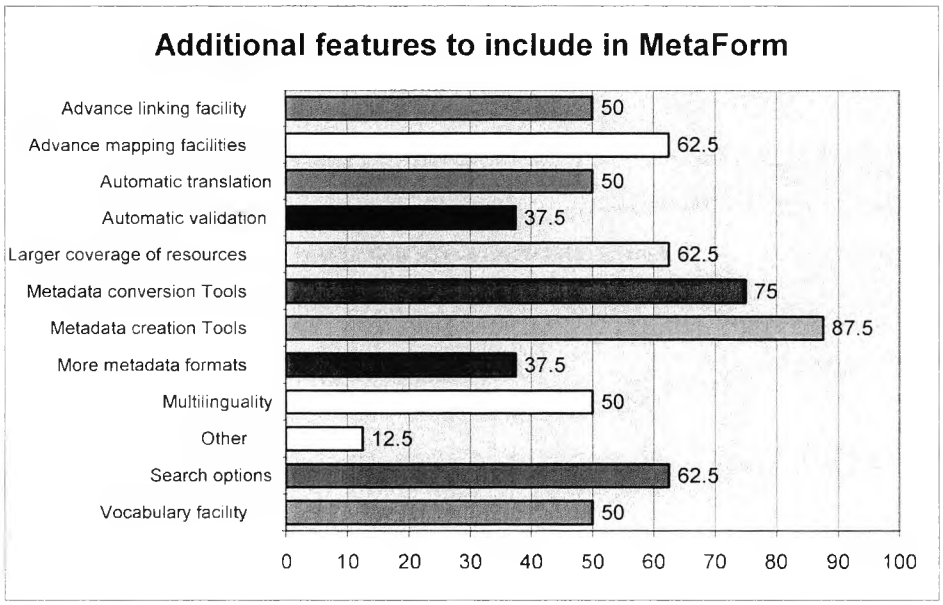


Figure 41: Results - MetaForm - Desired features

During 2002, the EDR was planning to add four more services to their current ones. Those included improved online harmonisation, XML, MetaPro and multiple download of files. The respondents were asked to note their preference as to which of those future additions they considered more important. The majority of the respondents (12 people) indicated that it would be more important to see information about the harmonisation of data across the EPA. Information and guidance on how this could be done and tools that would support this were the respondents' priority. XML and multiple files download were the next preferred additions in order of preference.

C.4 Interoperability

Interoperability problems are caused by the different requirements and expectations that the scientific communities have about the use of metadata. There is not one single metadata element set that can fit the requirements of all different scientific communities and usually the choice of the one being used is based on familiarity with the element set,

its use by other fellow communities and local systems compatibility (Palmer & Knutson, 2004). Interoperability is a complex issue to address and use of metadata registries is primarily associated with a common agreement of standards. One of the interviewees stressed the impact of social and political interoperability on the use of metadata registry systems; they noted that *"If re-use and sharing of schemas or ontologies was the solution to interoperability problems, then registries could have an important role, but I am sceptic that re-use will grow to that extent, due to the non-technical issues of interoperability".* Furthermore, organisations are reluctant to invest money and effort in projects that are not clear yet what the associated benefits could be *"...there has to be a benefit in making something interoperable. It's expensive to add metadata. I'm not going to make the effort to make my metadata interoperable with someone else's metadata unless there's a clear benefit to doing so."*

Another interviewee noted that in general, metadata registry systems can address interoperability problems"*...although interoperability is much more than creating registries...."*

> *"Interoperability is solved by providing uniform name spaces for storage resources users files file attributes control and consistency constraints. Sharing becomes possible when people have common reference systems (naming conventions)."*

> *"[I agree]...only partially. There are many metadata standards. It is a huge task within an organization to start implementing a metadata standard. You would have to decide at which level you want to go in the descriptions and which levels you show publicly and which only internally. The information systems have to be designed to comply with this."*

> *"I do wish to note though that I do not believe that metadata registries, as they are currently used, are very effective methods for achieving interoperability."*

The interviewees were asked to comment about the intending use of metadata registries in various information environments. Several respondents replied in an enthusiastic manner stating that yes, it is true that *"metadata registries can be used everywhere, from governmental and standardisation organisations to libraries and web based search engines" and*

The enthusiasm of the comments that interviewees made stresses again the importance that a standardisation process entails. Furthermore the role of the interaction with other

institutions/organisations and the exchange of information require a dynamic and active metadata registry system.

"There are numerous iterations of metadata registries to be found. From 11179, ebXML etc. each attempts to solve one piece of the puzzle relating to data reuse, data definition and data exchange. As a result, implementing a registry is a time consuming and challenging process. There are benefits to having a registry but the main benefit is to establish a repeatable process for data standards that the organization adopts. The registry is just the tool that facilitates a data standards process."

Can be used everywhere "This is a very broad statement. If the agency does not interact significantly with other external agencies or those interactions are static (i.e. the messages do not change dynamically), a registry maybe overkill!"

Furthermore, by defining the context of their use and presenting the benefits that could be gained by the re-use of metadata can be used to find information on implementation, examples in practice, authoritative information, and information on how to apply retrieved information (business rules) and eliminate one off use such as described below.

"I think that a registry is not something that you return to very often. After you've decided on your metadata elements and schemes for your project, you don't really need to look at the registry again until you decide to review your metadata."

"I don't think interoperability is an end in itself. You need to define what kind of information needs to be interoperable"

"Change "provide" with "will provide";-)

"Disagree. There are an attempt to get metadata without having to do the same kind of work that cataloguers traditionally put into their work. Interoperability is not really achieved only by making it possible to search across databases. It needs more than that. So I would say *"Metadata registries provide an attempt to alleviate the problem of interoperability on the web"*

Some of the interviewees commented on the success that their organisational metadata registry systems had. They emphasised the impact that such a program had in the organisation's culture and voiced their concerns regarding the associated costs and benefits that such a program had for their work. In particular they noted, "*In my opinion, we've had mixed success with Registries at EPA. The facility registry sys*tem has been a

very effective means of normalizing the identification of facilities regulated by EPA and has helped ensure that we aren't double counting the same facility. Similarly, the Substance Registry System has been a significant help in standardizing the names of chemicals and providing links to systems using synonyms of the same chemical. Some of the other registries, such as the Environmental Data Registry have shown to be of dubious value. Attempting to maintain up-to-date data element dictionaries of all systems at the Agency is resource consuming and I don't think the benefits have outweighed the significant costs."

C.5 Barriers to use

The respondents were asked to note any reasons that prevented them from using metadata registry systems. The stated reasons were grouped in three categories: A) Those imposed by the individual, B) Those imposed by the organisation and C) Reasons outside of their control. Barriers imposed by the individual referred to the amount of time that would be required to learn how to use the system and the amount of extra work that this would entail. In particular, an interviewee noted "*...potential barrier for usage of any registry is the amount of work required by a submitter of potential material to a registry. With the* [registry omitted] much has been made to try to minimise this (at the expense of making more work for the registrar and stewards). This has the potential benefit of making it more attractive to submit models – since the amount of extra work required to prepare a submission would be minimal." Also, the way someone is looking for information is affected by their prior knowledge of the system and the information that they are looking for. For example,

> *"If I'm looking for particular information I always use "Search" – this has become the most common access method due to Google and other search engines. If I'm looking for information in a certain topic, I would "Browse" like I would browse a library shelf. Usually an interface is designed in away that I browse to a category/facet and then by clicking a link I would commit a search to list all records within that section."*

Reasons stated under those imposed by the organisation referred to lack of the implementation of a policy that would encourage people to use standardised approaches for the description, storage, maintenance and transfer of metadata information. Lack of information about the advantages and benefits deriving from the use of metadata registry systems also act as barriers particularly to those who are enthusiastic about metadata use but finding it difficult to understand them. It was noted, "*My office is responsible for*

the development and management of the [organisation omitted] metadata registries. The primary barrier to my work is the (expected) slow movement within the [organisation omitted] toward wide-spread acceptance of the registries. This is an issue that should resolve itself as the agency continues down the path toward an enterprise perspective."

Another issue that was considered a barrier was the usability and functionality of certain metadata registries. The degree of user friendliness of a metadata registry system and the "*easiness of metadata entry (includes both manual entry one by one and in machine readable format (csv, xml or similar)*" was crucial. Also, *"easiness of updating the information – long term maintenance. If either of these two aspects [is] missing, the service will not be used in the long term."* Having adequate support systems such as help guides, implementation manuals, "business rules" or other associated documentation to describe how to use both the system and the content that can be found is important. Lack of such information can become a huge barrier both for the people who work on the aspect of populating the metadata registries but also for those who visit them to retrieve information. One of the respondents noted, "*Barriers would be having incomplete information about the registries and how to get things into them. Yes, there are probably ways that registries could be my work or other people's work*". Also, the role of IT in general, including support, network platforms and software limitations can affect the degree of services that can be provided.

Reasons grouped as those outside of the users control referred to the way people respond to change and the gradual adaptation to a new system for doing things that have conducted in a different manner in the past. Three of the respondents pointed out what could become the major constrain for using metadata registries:

"Statistical process is changing – people do not like changes"

"People do not understand that they have to work hard at the beginning, but that there will be less work in the future"

> *"There are no specific barriers at the current time. We anticipate that the greatest risk is one of adoption by all developers etc. to use the registry in their day-to-day work. It may also be viewed as another obstacle to getting the development work done quickly. By avoiding the data standard process and avoiding the use of the registry, developers could claim it would be faster to develop business solutions more quickly".*

One interesting comment that indicates the importance of IT support services when it comes to the use of new technologies referred to the pre conception that metadata can or

should only be used within certain domain specific communities. One of the respondents noted that one of the barriers that restricted their use of metadata registries was *"The old adage "You know what I mean." "It is only for the IT* community". Training in the use and clear, specific policy guidelines about the role of such systems can have in the data standardisation process of an organisation are important to understand the benefits that could be obtained.

C.6 Failed requests

The failed requests reports list the number of files that caused errors. Analog defines failed requests as those that return a web server status code in the 400s (error in request) or 500s (server error). The most common reasons for a failed request are when the requested file is not found or is read-protected. This type of report is useful to system administrators as they can identify the files that are problematic and correct them. Contrary to the other sections in the web log transactions analysis, failed requests are counted against requests sent to the Metadata server rather than the actual downloads.

With the exemption of 1998 to 1999 when failed requests were of almost double in size, it appears that MetaForm have made an effort to minimise problems of failed access. From 1999 to 2003 there is a constant decrease in failed accesses and although there has been a small increase from 2003 to 2004 it is minimal considering the big increase of the general MetaForm's use between those two years (Figure 42).

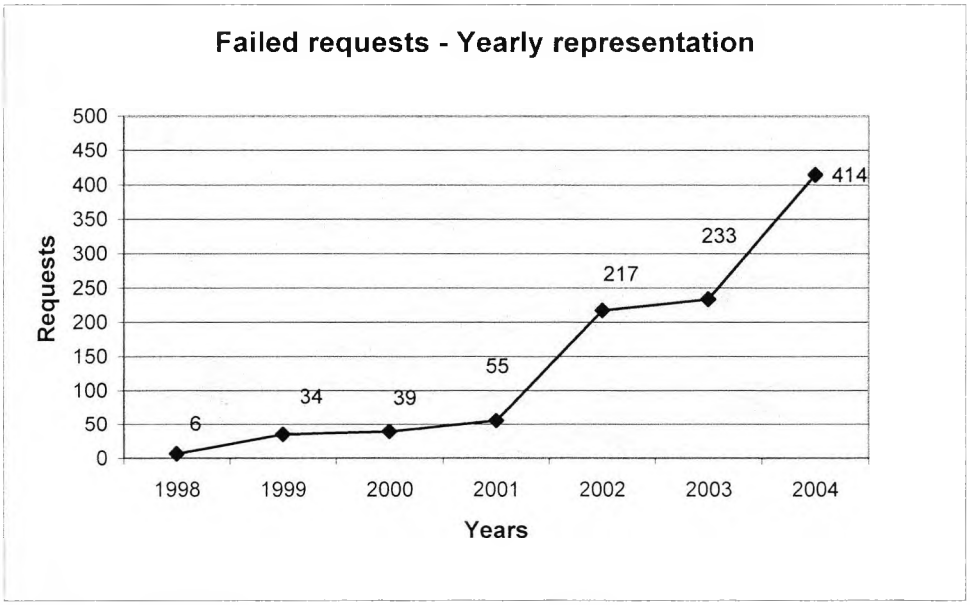


Figure 42: Results - MetaForm - Web log transactions - Comparison of failed requests by year

Again, as with the daily summary and the domain reports the monthly statistics for the failed requests needed to be calculated. In order to provide the closest representation of

the monthly failed requests over the period of 1998-2004 it was essential to calculate an estimated analogy of the monthly failed requests over the period of 1998-2004 and multiply it with the total number of page requests during each year. The estimated analogy of the failed requests was calculated by dividing the total number of requests for each year by the number of requests each month. Then, the monthly report of failed requests during each year (Figure 43) was produced by multiplying this average estimate of each month with the total number of requests. Failed requests remain unstable throughout all months for the period of 1998-2004. September and October counting respectively 28 and 24 averaged failed requests are the months when the most recorded failed requests took place.

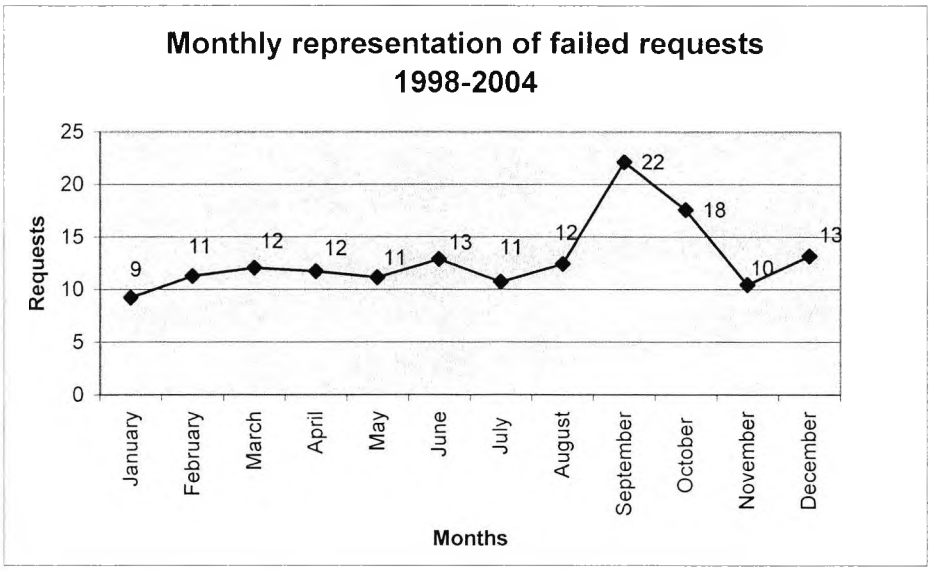


Figure 43: Results - MetaForm - Web log transactions - Comparison of failed requests by month

Throughout the six year period analysed the metadata server at SUB Göttingen encountered 86587 failed requests. Similarly to calculations described previously in this chapter the directory report that stated the number of requests for MetaForm was used again to calculate the average number of failed requests. The total number of failed requests for MetaForm was 1457 over the six years analysed. From those failed requests the vast majority (94%) returned an error code of 404 which is assigned when a document is not found. In particular, Fielding et al (1999) specify that the 404 error is displayed when the server has not found something matching the requests URI. The second most frequent error recorded on MetaForm was the "Bad request" error 400. This message is displayed when the server does not understand the request due to a malformed syntax. Other common errors recorded by the MetaForm registry were when users tried to access files that were forbidden (error 403) or while an internal server error occurred (error 500). The lists of errors that were recorded in the MetaForm's failed requests report is presented in Table 29.

Code	Definition	Requests	%
400	Bad request	34.15	2.34
401	Authentication	0.24	0.02
403	Access forbidden	26.77	1.84
404	Document not found	1368.79	93.95
405	Method not allowed	5.25	0.36
408	Request timeout	3.60	0.25
414	Requested filename too long	4.29	0.29
416	Requested filename too long	0.03	0.00
500	Internal server error	9.24	0.63
501	Request type not supported	4.37	0.30
	Total	1457	100

Table 29: Results - MetaForm - Web log transactions - Failed requests - Error codes

Failed requests are associated with changes and/or upgrade in web content software. Increase in the use a service is also bound to reveal any problems associated with file management and record keeping. The highest number of failed requests for MetaForm was recorded in September 2004. 2004 was the busiest in all six years analysed. From 1998 to 2001 the number of recorded failed requests appears to be stable. The highest increase and the lowest fall of failed requests were both recorded in the year 2004. From August to September 2004 the number of failed requests had dropped to almost half but so has the overall use of the MetaForm's registry.

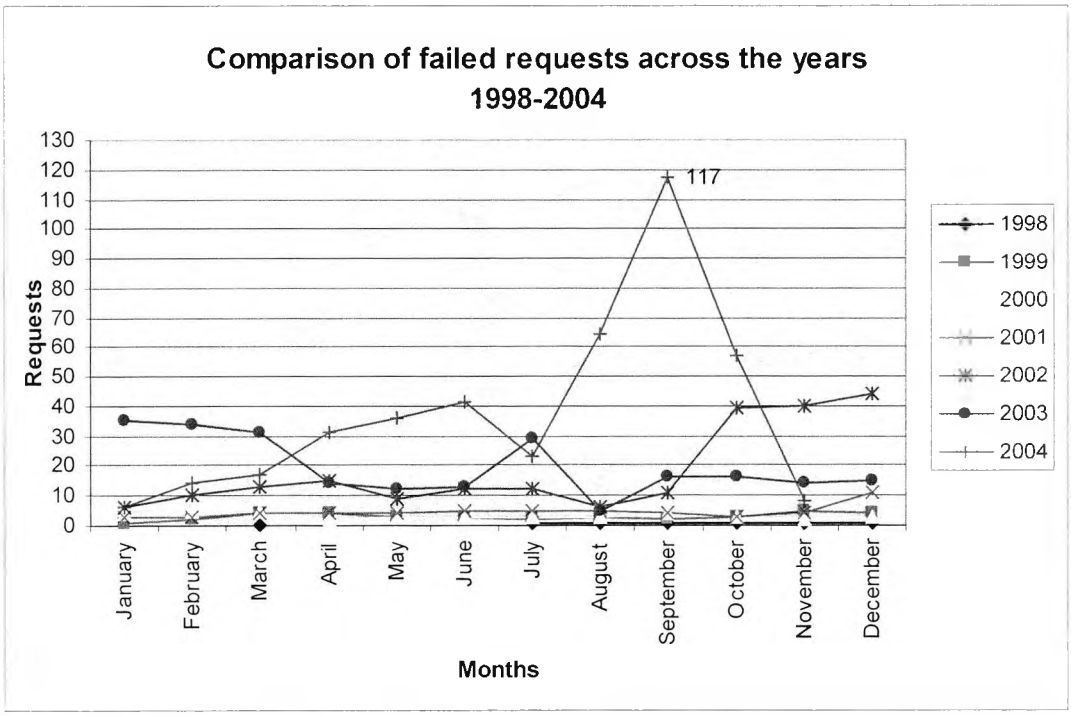


Figure 44: Results - MetaForm - Web log transactions - Failed requests - Comparison of yearly representation

Regarding the Environmental Data Registry, the year when most failed requests occurred is 2002 at a 32389 counts and the year with the least failed requests is 2000 at 4108 counts (Figure 45). Between 1998 and 1999 the number of failed requests appears to be equally distributed.

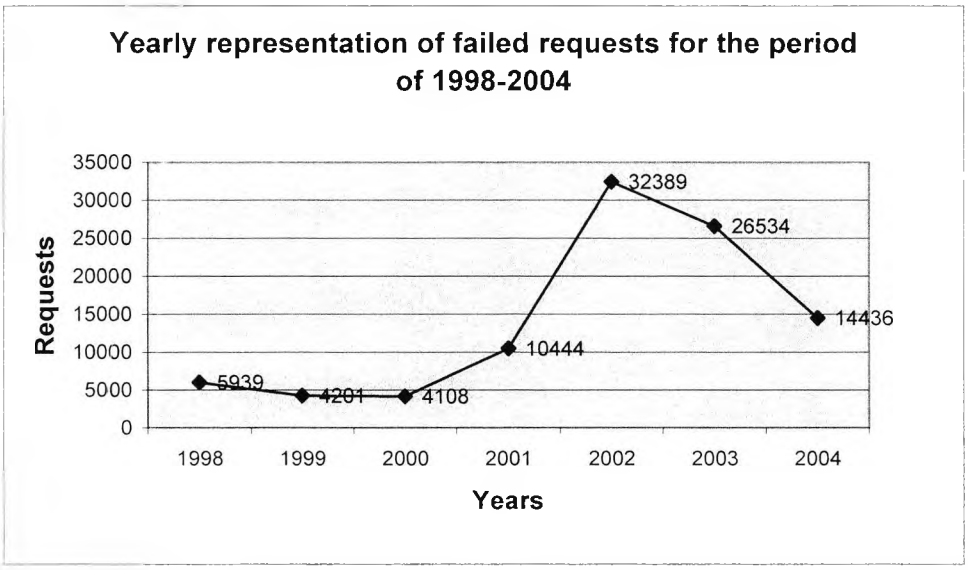


Figure 45: Results - System of Registries - Web log transactions - Comparison of failed requests by year

October (1628) and May (1462) are the months with the higher number of failed requests for the whole period of six years and eight months from 1998 to June 2004. August (899) and September (894) are the months when the least failed requests occurred (Figure 46).

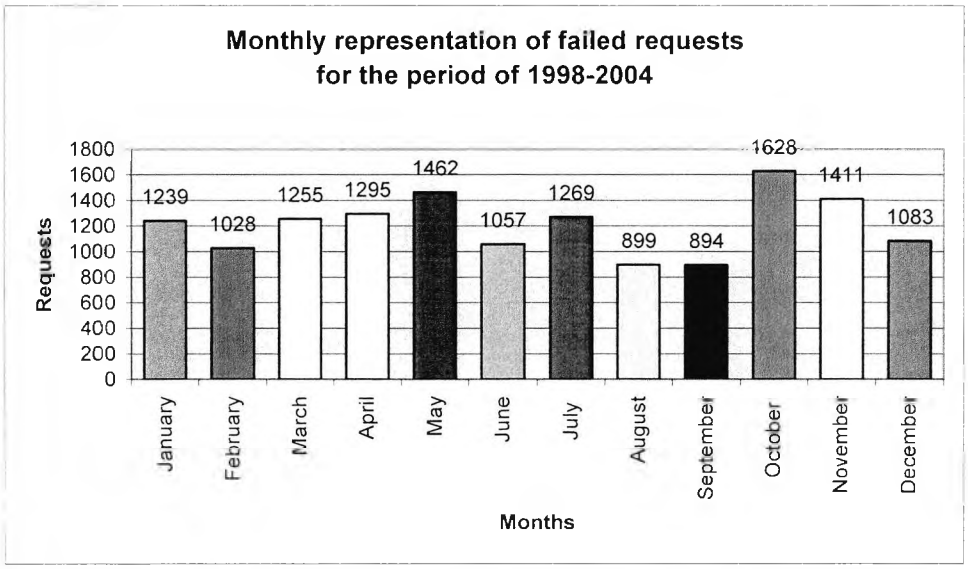


Figure 46: Results - System of Registries - Web log transactions - Comparison of failed requests by month

The failed requests appear to be more unstable in the recent years of the data analysed. In particular, 2002 and 2003 appear to be the years with the highest and lowest marks of failed requests. The earlier years of 1998, 1999 and 2000 appear to be more stable. As it

has been noted earlier failed requests can be associated with the re design of a web site and the re arrangement of the directories and files that can not be found.

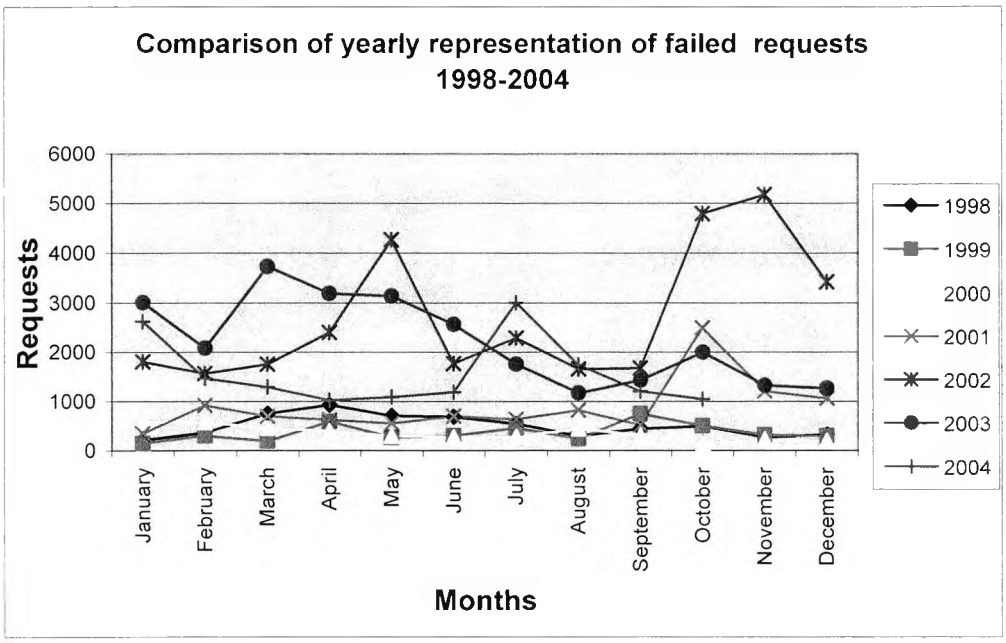


Figure 47: Results - System of Registries - Web log transactions - Comparison of failed requests by year

C.7 Non use

Users of metadata registry systems are difficult to locate due to several reasons including confidentiality restrictions imposed by the organisation that hosts the registry, the nature of online distance services that tend to de personalise users, and the level of deployment of such systems at the moment of the current research study. Those who have responded to surveys enquiring about metadata registry system users often distanced themselves from the perceived end-user and focused in the main area of their job responsibilities such as development or implementation or management or evaluation of systems only. Additionally, the lack of understanding of the role and uses of metadata registry systems resolved in non use as well. For example, one of the respondents noted, "*once you define your metadata schema there is no need to visit the registry unless you are revising your metadata.*" Therefore, this person, primarily involved in the drawing of metadata guidelines for use did not feel that people who use metadata may need to visit the registries to see how one element is applied in use in different organisations/institutes, retrieve mappings of their institutional element set or application profiles and maybe cross check for the case that they are downloading records described in different element set or even see actual examples of use.

One other respondent, who was also involved in the development of their organisation's metadata registry, noted that they did not need to use a registry because they were familiar with the use of metadata information and they knew how to implement it.

> *"The main reason I do not use registries is because [organisation's registry] metadata information has been developed in house and I know how to translate it to Dublin Core and other frameworks like IEEE-LOM."*

Furthermore, "not being required to use" metadata registry systems as part of their every day job was another reason for non use. In fact, the respondent who noted this reason did say that browsed through metadata registry systems to see how they work. In particular it was stated "*I do not use any particular registry on an every day basis but I need to have an eye on many systems to peak up ideas for our own website and registry development.*"

Finally a respondent noted that their primary job responsibility was in the population of the metadata registry systems and the development of the associated rules for their use. "*We are not making wide use of registers. Our main interest in registries is in the development of rules for registration and in the creation and population of registers for areas of interest to us, such as metadata elements, coordinate reference systems, and sensor types and properties.*"

PART D – Comments

The following section discusses metadata registry systems definitions that the respondents had made and the general comments that had been recorded via the questionnaire surveys and the email interviews. Response provided useful information about the respondents understanding of what metadata registry systems are and what they can be used for and their role in facilitating data standardisation and exchange of information across diverse networked systems. The respondents provided useful information about how they envision the use of a metadata registry system within an organisation with their responses regarding how metadata registry systems can support their work.

D.1 Definitions of metadata registry systems

Response from the email interviews showed that different scientific communities perceive metadata registry systems in a domain specific manner and the views are to a great extent dependent upon the projects that the people had been involved in. For example, interviewees with a scientific background in data standardisation and linguistics defined a metadata registry as an "*ISO morpho-syntactic data categories registry...based on EAGLES works*". The emphasis was placed upon expressions of linguistic nature and the grammatical functions of words and metadata registry systems were described as dictionaries that describe the meaning of the words and the relations between them.

Respondents that their primary interest in metadata registries laid in the area of data standardisation placed the emphasis on using registries for sharing information by several people particularly in the same organisation via a standardised approach. They defined a registry as, "*...a data base, in which tables metadata are saved to [be] used by different users – metadata* used in organization" and as "A standards-based application for the management of metadata."

Respondents that were primarily engaged in metadata management and implementation stressed the role of managing registered metadata within an organisation as well as exchanging information and liaising with associated parties of the parent organisation or other organisations. In particular they referred to a metadata registry as, "*...a system to administer data standards and store standards data definitions; a repo*sitory of standard data definitions that can be shared with business partners."

And finally interviewees with a primer interest in software development stressed the use of such a system in the resource discovery process by defining it as *"...an ontology discovery service."*

D.2 General comments

The respondents to the email interviews were asked to comment on two statements about:

- the role of metadata registry systems in data standardisation and sharing and
- their intended use by diverse organisations and in different information environments.

Their response provided feedback about the beliefs, understanding and expectations of metadata registry systems' developers, implementers and users. The response provided useful feedback about what constitutes a useful metadata registry system and how users see their role in supporting data sharing and standardisation. The statements that the respondents were asked to comment are listed below:

"Metadata registries provide the effective solution to the interoperability problem on the web" and

"Metadata registries can be used everywhere, from governmental and standardisation organisations to libraries and web based search engines"

Understanding how metadata registry systems can be used and what are the potential benefits for any organisation, its employees and/or other users, and its workflow can be critical for their deployment. The comments the respondents made emphasised the need of having clear policies and contextual frameworks that require the use of metadata registry systems before commencing on their development and implementation. One of the comments that have been made from a respondent that their job was about the management of metadata was that emphasis is often placed on perfecting a model out of context rather than providing the context of use beforehand. People would find it easier and more helpful to use any system that there is a requirement for its use. For example, one of the respondents noted the importance of having completed applications running that require the use of metadata registry systems beforehand and then commence the projects that use them. By doing so, people would see the benefits of using a standardise approach to describe, retrieve and exchange metadata information and request to implement the systems that support such functions. For example,

"I think they help, but I think the way it's done is a bit backwards. I think you need to fund projects which use metadata first, so that the benefits of using it are there before you start applying it. i.e., you need to set up a portal/search engine that

uses metadata, then put the details of what metadata is required for you to become part of the portal in to the registry..."

Another of the respondents discussing the importance of an underlying standardisation framework for the use of any metadata registry system noted that metadata registry systems are one of the components in the wider area of metadata standardisation and their successful implementation and use is dependent upon the underlying standardisation framework in which they operate and are used.

"Er...they provide an effective solution to the problem of **** discovery **** *of metadata standards and profiles within web communities. The problem is that without an underlying model, any set of "metadata terms" is essentially meaningless... It is the* **** model **** *that provides the basis for solving the problem of interoperability, not the registry."*

Furthermore, the usefulness of any information system is very much dependent upon the level of functionality that its services can be performed. Support services such as help guides are considered very important for any information system as they can provide help and guidance when human support is not available. One of the respondents noted that among the most useful features of the metadata registry system in their organisation they counted the thesaurus, the subject search and the browse options. Also, it was suggested that *"...metadata registries combined with other tools like indexing and data mining are probable a better solution."* The business rules documents received particular mention,

Without a deep understanding of the business rules associated with the data that is being administered in the metadata registry, partners cannot conduct business. Without unambiguous understanding of the data or metadata, business partners cannot effectively conduct business. Metadata registries are just part of the solution.

Also, another significant area that respondents commented upon was the potential benefits from the use of metadata registries. Those included the availability and ease of access to information and the awareness that can be achieved by providing access to the information. One respondent noted the short term benefits that can be accomplished *"As more important than interoperability I see the aspect of awareness. Many organisations do not know what data they have, what data is available on partner institutions or what data is available on the Web. Good environmental registries help here to create awareness and for self-management of public organisations."* Response ranged from those who were enthusiastic about the use of metadata registry systems and those who were still sceptical about their use. The importance of international collaboration and the

effective role that metadata registries can have in the exchange of information has also been noted.

> *"They provide A solution –potentially the best solution we have based on available technology. Care has to be taken though, since currently there are many differing standards and versions of standards – which may cause problems at a future date if (when) it becomes more of a requirement that disparate data registries have to work together, not in isolation (e.g. linking transport registries in UK, Europe, USA, and Australia)".*

The last part of the questionnaire invited the conference attendees to define what the Environmental Data Registry was and to let the organisers of the users' conference know how it could be best help their work. Comments and suggestions from this section aimed to provide a better understanding to the organisers of what their users understood by and desired from the EDR services.

In general, the users' replies met the EPA's definition of the EDR as "*...a comprehensive, authoritative source of reference information about the definition, source, and uses of environmental data. The EDR catalogs the Environmental Protection Agency's (EPA) major data collections and helps locate environmental information of interest*". They seemed to acknowledge its role as a means to support data standardisation across EPA. Some though commented on the structure and the management of the information in the EDR. A user in particular noted that *"[It] should be a registry of data elements that should permit a larger system to link to appropriate data and documents. It has stretched beyond that"*.The users' definitions of the EDR are presented in Table 30.

Please let us know in your own words, what do you think EDR is?
A library of EDR data standards.
A lot of info that needs a wider use base and enforcement to use process.
A registry of data elements used in EPA systems.
A registry of information about environmental data.
A site to access information on EPA data elements and tools to analyze this information with respect to existing & planned datasets and systems.
A very useful source of information on the process and status of data standards development. I have not had much experience with harmonisations tools, yet.
Compendium of EPA's metadata.
Compilation of registries used by the Agency for standards and metadata.
Comprehensive EPA metadata repository.
Data Information about data.
Data standards for the U.S. Fed Government.
Information tool that directs you to data information for the environment.
Should be a registry of data elements that should permit a larger system to link to appropriate data and documents. It has stretched beyond that.
Tool to understand data.

Table 30: Results - EDR - Questionnaire survey A - Definition of EDR

The conference attendees were asked to indicate how the EDR could support their work. The aim of this question was to find out how the EDR users intended to use the registry system in practice but also to gather suggestions and recommendations about how the EDR could be used. The majority of the users noted the importance of data standardization and harmonisation across the EPA. The role of the EDR in this effort was highlighted in the responses. Additionally external visitors also replied that they would use the EDR for their work, in particular one respondent noted *"[to] adopt metadata and data standards for NARSTO & Canadian Atmospheric Measurements"* and also to use it as a database for retrieving information about chemical substances e.g. *"It could help facilitate some of the searching we need to do for chemicals we need to collect information on (23.000 substances)"*. Suggestions such as the use of EDR as a quality assurance tool and as the authoritative tool to manage future integration of systems and also as a means to standardise internal systems life cycle were made. The users' suggestions are grouped in Table 31.

How do you or could you use EDR to support the work that you do?
Adopt metadata and data standards for NARSTO & Canadian Atmospheric Measurements
Currently harmonise data elements/ XML tag developments
data element standardisation
Data harmonisation with online tools; standards compliance; systems integration/consolidation
Data standardisation
Ensure program office complies with data standards
I use the EDR to ensure the systems I develop conform to EPA standards
Identify standards to use in won systems. To assess where common elements are being used and thus where integration could occur
It could help facilitate some of the searching we need to do for chemicals we need to collect information on (23.000 substances)
It will become part of the overall system life cycle and budget process (CPIC)
To develop fields using EDR; TO develop allowable values
To do looking for adding chemical resources to databases
Use it for data quality /Audit

Table 31: Results - EDR - Questionnaire survey A - Use of EDR in the EPA

Finally the users were asked to make any additional comments that they wished regarding the EDR and its use. Again, the search function attracted criticism and one of the respondents noted specifically noted *"Search function needs work; hard enter gives oracle error; add acronyms list; CRS is slow"*. Another user touched upon the organisation's contribution and support of the deployment of the EDR across EPA. Although the compliance of the structure and intended use to an international standard the population of the registry and its organisation was criticised. In particular it was noted, *"While the EDR has apparently been constructed in line with ISO 11179, the data elements contained within it have not gone through a recognised standards development process"*. Another suggestion was referred to additional services for system developers. A respondent suggested that the EDR should become the gateway to resources by area of interest. Additional information such as implementation guides and policy documents were considered important. A user explicitly suggested *"[The] site needs a document for*

developers which tells them which attributes of a data element are required (by OEI) for development of appropriate metadata and which are recommended (e.g. How to build your DED)".

The second part of the questionnaire aimed at gathering feedback about the strengths of the System of Registries and the applications that needed to be improved to meet the aims of the EPA for data standardisation, as its users perceived those. Also, it asked the users to denote how the System of Registries could help them in their work. Respondents felt that one of the strengths of the System of Registries was the enhanced search capability, which had been a by product of feedback from last year's conference participants. Comments from last year's conference urged for a redesign of the search function with an improved site map, search across all registry systems and support of Boolean operators. This year the comments reflected the change. A few of the respondents noted in particular that they considered an advantage the *"ease of access [and] search capability", "the provision of information resources, search capability".* Other respondents emphasised the diversity of the System of Registries applications and the association of their use with their specific needs. For example, the IRRS and FRS were praised as systems that their comment specialised user needs and another respondent that used the SRS commented on the ease of access to it and the wealth of the information that was included there. In particular, they noted *"It is easy to find by CAS RN, all the programs and regulations that are associated with a particular chemical. The link to the HDSB is particularly useful for us. Any links to the databases with toxicological or environmental information are/would be useful for our work in risk assessment and regulatory compliance for chemicals".* Other respondents placed the emphasis on the role of the System of Registries in the EPA's policy about data standardisation and harmonisation across the various systems and sections in the EPA. In particular their comments reflected on the role of the System of Registries:

- "The ability to integrate and harmonize data sets. To be able to have all of the data sets in one place".
- "The definition of a standard set of information for exchange and the documentation of those standards."
- "The standardization of data and the publication and exchange of the information."

The comments that the respondents made did not only reflect on the strengths of the System of Registries applications but also on what they considered as less strong features. This year the suggestions for improvement revolved around the substances registry system. Although it had been praised as one of the most useful System of Registries applications, a few of the respondents suggested that improvements in the

database response time would increase its use. Also reviewing of the content and additions to certain lists and information in the SRS would make it even better. In particular it was noted:

- *"There is no information on some chemicals. Chemical-structure searching capability would be nice to have, so that one could look for \"similar\" chemicals or look for chemicals for which one does not have the name or the CAS number. There is no information on certain chemicals; more information is always better"*
- *"In the Substance Registry, it takes a very long time to get results from the substance name or CASRN sorting option for some lists"*
- *"The still incomplete domain lists of substances and physical measures"*

Further issues of concern drawn by the respondents were again about the search facility, which for some features remained slow. According to one respondent, there is a need to provide *"better/simpler search and download"*. Also, concerns were raised about the appeal of the System of Registries to the general public and reviewing of its current content. One of the respondents suggested expanding the collaboration with other US organizations and even other countries. In particular, there was a suggestion to enhance information in the System of Registries with data sources from *"the European Chemicals Bureau or international organizations such as OECD"* in exchange of estimation techniques and QSARs.

- *"Expand into non-environmental reporting to include general business report e.g. xbrl"*
- *"The ability to make it more understandable to the public"*.

The conference attendees were asked to let the System of Registries managers know how the metadata registry systems could support their work. The replies provided a wealth of input for System of Registries to consider. Surprisingly most replies came from the external users of System of Registries and they referred to how they could gain experience from using the System of Registries applications and attend their conferences so that they could implement similar approaches to their own organisations. For example one of them noted *"I am looking to transfer some of the ideas and processes in SOR for use in the US Federal Aviation Administration (FAA)"*, *""[By]... Retrieving data standards, XML tags for the in house development of our Facility dB"* and *"Enhance the software for my organization sells"*. Other suggestions referred to implementation of the EPA policies for data standardisation across its different offices and to the improvement of current System of Registries applications. In detail some of the suggestions are listed below:

- Can help with one stop shopping with different registries.
- Complete the still incomplete domain lists of substances and physical measures.

- Help with data validation, data requirements, required standards and data relationships.
- Prevent duplication of effort.

One of the suggestions that deserves separate mention as it raised the issue of publicising resources and collaboration with international parties noted "...for work for other geographies it would be useful to bring in data sources from other countries (such as from ECB – the European Chemicals Bureau, or international organizations such as OECD. Also the cooperation with the US EPA's OPPT (Office for Pollution Prevention and Toxics) does not seem to be present; OPPT would most likely be a useful – and probably willing – "trading partner".

Chapter 5 – Discussion and conclusions

One of the proposed solutions to establish data standardisation and management and to address interoperability problems associated with the exchange of information in a networked environment is forming authoritative registration spaces. Initiatives towards that direction included the establishment of the ISO/IEC 11179 standard to guide implementation of metadata registry systems along with other international standardisation approaches such as the use of a common element set such as the Dublin Core.

Data from web log transactions of two active metadata registry systems covering a period of more than six years and accounting for more than 10.000.000 accesses provided useful information about the users of those registry systems, how they interact with the metadata registries, the use they make of certain features within each system, their preference in particular types of information and barriers to their use. Additionally to this information, response from four questionnaire studies both with people engaged in metadata research and users of metadata registry systems provided a basis for the email interviews that were conducted with developers, implementers and managers of such systems.

Questionnaire studies conducted with people engaged in metadata research via discussion lists and attendees to the SCHEMAS project workshop as well as with users of two active metadata registry systems provided useful information about the metadata registry systems' role in data standardisation and management.

Furthermore email interviews conducted with people engaged in the development, implementation and management of metadata registry systems provided in depth information about future expectations of such systems and their role data standardisation and management in particular how they can address the problem of information interoperability and scope of domain specific applications.

5.1 Discussion and conclusions

The received response from two questionnaire surveys that were conducted with four discussion lists subscribers and the 2nd SCHEMAS workshop delegates provided useful information about the views of people regarding metadata and metadata registries. It shed some light on the demographic characteristics of people involved in metadata

research at the time of the surveys and it provided with indicative feedback on metadata research trends. Some of the primary questions asked were about what metadata element sets respondents were familiar with and used. They were asked for what purposes they have used metadata in order to see what metadata applications existed at the time of the survey. Furthermore, respondents provided feedback on what metadata registry systems they were familiar with, if and which of those systems they have used and for what purposes. Their response set the current trends in metadata research of the time. The following can be acclaimed:

5.1.1 Characteristics of metadata registry systems users

Sixty six (people) from different scientific backgrounds responded to a four discussion lists survey and the SCHEMAS workshop survey during November 2001. The scientific domains that were represented in the surveys were the Academic, Educational, Industrial, Publishing and Research domains. As research by the Audio visual and the Geographic information domains is well documented in the literature, non response in the surveys is attributed to the subscriptions overlap of the discussion lists selected and the fact that the lists were selected with the criteria of relevance to metadata research, number of members and international coverage and not the domains represented in each list. The majority of the respondents (39%) represented the academic domain, "other" domains (18%) such as governmental, consulting and natural resources information management and the research domain (16%).

Data from the web log transactions revealed that the domains that made the most use of the MetaForm registry and the EDR/SoR were European Union countries and the commercial and network domains. "Other" users included academic institutions and governmental organisations. Also, systems that directed users to the EDR/SoR were also search engines, the Environmental Protection Agency and web directories. Surprisingly, the EDR/SoR was mostly used by the commercial sector that could be interpreted as an indication by commercial organisations to have access to standardised, quality, authoritative information such as that that a governmental organisation produces. Furthermore, the lists of services that referred users to the registries included the Environmental Protection Agency, several search engines such as Google, Yahoo, AltaVista, MSN, AskGeeves and AOL and "other" referrers such as governmental and commercial organisations, academic institutions, professional bodies and World Wide Web directories.

Response was equally received by both men and women. Although people in all age ranges responded to the research study, the groups that provided the highest response rate were the 24-35s and 44-54s. Men were primarily represented in the age group of 24-35 and women were more in the group of 44-54s.

Information and computer scientists, researchers, developers and implementers of metadata registries, consultants and educators were some of the occupational posts that the questionnaires' respondents indicated as their profession. Information scientists represented half (50%) of the response. The EDR respondents indicated a variety of areas that they were interested in such as standards development, implementation, auditing, and systems analysis. This was indicative of the EDR conference participants who were not limited to EPA employees but included other governmental organisations employees, commercial parties, a research company and the Chemical Abstracts Service. Almost two-thirds of respondents expressed their interest in all three predefined areas (develop systems, implement standards and harmonise data) as opposed to focussing on one particular service. This could be interpreted along with the number of first time users that EDR users are still more interested in learning about how a metadata registry system works. Nevertheless, the expressed interest in more than one specific area is in line with Johnston (2002) and Duval and colleagues (2002) that listed six types of users for metadata registry systems including developers, implementers, researchers and metadata managers. The development of an internal gateway – as we see the How to...Facility within EDR - to resources for specific areas of interest represents EDR's ongoing interest and research in the area of metadata standardisation and its user satisfaction. Such findings are also supported by similar studies in the literatures (Monopoli, 2005; Monopoli, 2001) that showed that use of subject specific portals increase the use of digital libraries in general.

Findings from a questionnaire survey that was conducted after one year revealed that the users of the EDR/SoR similarly expressed interest in more than one area of metadata research. Equal interest was expressed in data standards development, implementation, and exchange of metadata, system development and evaluation. A factor that was thought to have increased the use in specific areas was the advocacy and publication of the changes that occurred in the SoR and the endorsement of the parent organisation, the EPA, of the data standardisation program. The role of an organisational requirement for the use of metadata registry systems had been denoted by the interviewees as a primer factor for their future use.

The response that was received by the questionnaire surveys and the web log transactions data enabled the formation of an indicative profile of the EDR users' information seeking behaviour. That showed that in their majority they were people familiar with the use and searching of electronic resources on a daily basis. More than three quarters (84%) of the users noted that they make use of electronic information on a daily basis. Almost half described themselves as advanced users who, although they find the process of obtaining information of interest relatively easy, also encountered some difficulties, which they attributed mostly to, unorganised information and poor web site design and resultant lack of supporting services. Approximately one-third of respondents were first time users, and all with the exception of two people declared a relation with the Environmental Protection Agency, which developed and maintains the Environmental Data Registry. One quarter of those who have used EDR in the past were seen as regular users of the system, and about 26.1% had been using EDR for a period of time that ranged between one and five years. It was the contrary for the MetaForm registry that more than two thirds of the respondents indicated that they were either first time or new users of the system. The frequency of the system's use also reflected that fact. Less than a quarter or the users replied that they were occasional users of the system. The web log transaction data though had recorded an constant use of the system over the six years of data analysed and that could be interpreted as having many different users that visited the MetaForm registry occasionally.

5.1.2 Metadata use

All respondents from all scientific domains were familiar with the term metadata. The engagement of the different scientific domains in metadata research has been noted in the literature. The projects DESIRE, SCHEMAS and CORES has recorded the evolution of metadata elements in several scientific domains and Vellucci (1998) and Milstead and Feldman (1999) also reported on metadata element sets and projects evolution. This study invited respondents to note the terms with which they had mostly associated metadata in order to find out about the influence of certain domains in metadata research. The respondents noted that they had mostly associated metadata with the following terms: standards (94%), schema (91%), element sets (87%) and formats (76%). The different notions that each of those terms carries such as quality and efficiency for standards, computing and data processing for schema, uniformity and simplicity for element set and experience and reliability for format represent the contribution of the different scientific domains in metadata research.

The Dublin Core and the MARC standards were the most cited metadata element sets in terms of familiarity and use among the respondents. A survey of the DCMI Libraries Working Group (Guinchard, 2001) on the use of Dublin Core in libraries indicated that Dublin Core was the most used format in University Libraries because of 'its international acceptance as a de facto standard and its "flexibility' (as most popular reasons for it use). 'The most frequently reported Dublin Core implementations were related to subject gateways, and the management of electronic publications, often including theses and dissertations. Almost half of the response pointed in the use of *"Other"* element sets. The element sets that were noted did not always refer to an actual element set but to metadata research projects, frameworks, project outcomes, periodicals, locally created schemes, metadata registries software. This is indicative of the lack of information and/or confusion as to what a metadata element set is and how and where it can be used.

The most cited reasons for using metadata element sets were the projects that the respondents were working on and to use it to describe library resources. The numerous conferences and workshops about metadata and an increasing number of metadata related projects at the beginning of the 21st century (see Chapter 2 – Literature) justifies the reasons that the respondents noted. Some examples of the projects that used metadata at the time included digital image collections, collection of art images in a digital library, Southeast Native American documents 1730-1842, a metadata encyclopaedia, training village database of learning resources and description of resources for a Subject Based Information Gateway.

5.1.3 Metadata registry systems use

Although metadata registry systems were popular among the respondents of the questionnaire surveys and the SCHEMAS workshop attendees less than a quarter of them (22%) had actually used one at the time of the survey. Those respondents who had used a metadata registry system, noted as the primer reasons for using it to find out about definitions of elements and to see how a metadata registry system is structured. The only other reason that has been stated was to *"check on progress of registries based on ISO/IEC 11179"*. This finding complements what has been recorded in the literature as the main functions of a metadata registry system. Baker et al (2001), Nagamori et al (2001) and Heery (2002) all listed the description of metadata elements and the guidelines for their use as some of the core functions of a metadata registry system. The most popular registry systems among the questionnaire surveys' respondents were the UKOLN based ROADS templates and the MetaForm registry.

The daily, monthly and yearly web log reports showed the use of the metadata registry systems over the period of six years. The MetaForm registry's use appeared to be almost equally spread throughout all days of the week while the EDR/SoR's use was significantly lower during the weekend. Monthly and yearly use showed that use can be associated with particular events that occur at a given time. The months that the users' conferences took place indicated a higher use of the system at the EDR/SoR and it was similar with the final workshop at the MetaLib project in SUB Göttingen, which was the parent project of the MetaForm registry system. Additionally, the use of the Dublin Core standard as the basis for the MetaForm registry contributed to added interest in its services. As Guinchard (2002) reported in her survey of the Dublin Core use, it counted more than 33 implantations in libraries alone during 2002. Its proclaiming to an international standard in 2003 also contributed to its establishment as the main metadata element set for describing electronic resources in the Internet.

The use by type of information indicated an interest to links in full text documents and published research outcomes about use of metadata in general. In the case of the MetaForm registry, use was significantly increased by the placement of the MetaForm's website next to the subject gateway of the MetaLib server. That way the users navigated from one system to the other to gain access to full text information of interest. For the EDR the most popular registries were the Environmental Data Registry in itself as it contained information about the majority of the data standards used in the EPA and the Chemical Substances Registry. This registry was of primer interest to external users such as those coming from the commercial sector. The introduction of the new SoR did not change much in the use of the individual registry systems. Again, those mostly accessed by the SoR users were the EDR and the SRS. The newly introduced XML tags – available through EDR – appeared to be popular among the respondents. They denoted that the information they tend to search for is data standards, elements and the business rules documents that explain their usage. XML tags were also rated among those most appealing. The main reason that the respondents accessed the System of Registries was to search for information. Web usage statistics identified that the main referrer to System of Registries pages is the EPA. The respondents that indicated that the main reason for accessing System of Registries was to search for information they have done so to keep up to date with information resources within EPA and have also downloaded information for multiple standards. Another pointer from the EDR/SoR users was the publicity of events such as the users' conference that would increase the use of the metadata registry systems.

The services that the EDR users accessed mostly were the Search Facility, the Download and the "How to..." features. Half of the respondents ranked the Search Facility as a poor or fair service, while Download was considered to be a very good service. The majority of respondents are satisfied with the EDR's coverage of resources, but they would welcome future additions. Of particular interest were the Environmental Protection Agency's data standards, data elements and business rules documents. This again, corresponds to the reporting of the core functions of metadata registry systems in the literature. The description of metadata elements and the associated rules that define their context of use are considered to form the basic functions of a metadata registry system (Duval et al, 2002; Heery, 2002). Again, the majority of respondents replied that there is a requirement for a metadata registry system within their organisation and 42% of respondents have used the information they retrieved from the EDR within their organisation. In particular, applications have been associated with systems development and data standardisation for purposes of ensuring data quality and system interoperability.

Less than half of respondents considered the process of obtaining information from the EDR as relatively straightforward, but they would replied they would appreciate more guidance during their search. More than one-third either did not answer the question, or haven't been able to find what they were looking for. Suggestions referred mainly to the improvement of Search Facility and the Site Map. Most respondents noted that the information they retrieved from the EDR has been either relevant or adequately covering their initial query, which suggests that the EDR provides users with valuable information, and is an appropriate means for information retrieval needs, in spite of findings that suggest that continued improvement is needed. One key improvement identified by respondents was the need for inclusion of "*Boolean, acronyms and text search for entire site*". Studies that investigated the use of subject gateway from a user perspective (Monopoli, 2005; Monopoli, 2001) also reported that users prefer to use the search option rather than browsing and it is considered the vital point for any such system that provides access to information.

The reasons for using metadata registry systems as those have been reported in the literature tend to take two approaches. The user oriented approach and the production oriented approach (Bargmyer, 2003). They tend to focus in explaining the benefits from re use of metadata and avoiding duplication of effort in its production and management. The response to this study brings valuable information and complements those findings already reported in the literature. The reasons that the users noted were divided in four categories: authoritative use or information, management of semantics, recording and

disseminating data standards and other. All stressed the importance of having a single source of registration that ensures quality control and eliminates duplication of any production effort. Being able to address issues of long term preservation to access of information as well as managing and exchanging valuable administrative metadata was also noted by the users. The recording and dissemination of metadata was noted as an important factor for using metadata registries for addressing issues such as cost effectiveness, monitoring organisations expertise in specific areas and making information available to relevant parties. As one of the interviewees noted "...registries are the means of discovery, promoting re-use and integration of work between related projects". Other reasons included exploration and identification of metadata registry systems functions, assessment and evaluation of systems and an organisational requirement that entails the use of such a system. One third of the respondents noted that their organisation has a requirement to use a metadata registry system for reasons such as quality control, to establish data standardisation and uniformity, for reasons of digital data preservation and to create metadata. Such a requirement could significantly increase the use of a metadata registry system.

5.1.4 Metadata registry systems functionality

Different metadata registry systems entail a variety of functionality. All functions though are considered essential of the role and the requirements of metadata registry systems. Those include the description of data elements, provision of guidelines for their use, mappings across elements of different metadata schemas and, for the case of DC registry, facilitation of multilingual searching. Consistent update of registry content is essential for validity and credibility. Metadata registry systems applications, even at their formative stages, can be found in the governmental sector (EDR/SoR, AIHKW/METeOR), in the research sector (SCHEMAS, CORES), in the library community (MetaForm, DCMI) and in the educational sector (MEG).Their role as authoritative systems of information that describes digital objects and facilitates their retrieval is vital in the process of data standardisation and metadata management and consequently in the organisation of digital information. Feedback from the end users provided some of the most interesting results of this study. Building on the SCHEMAS Standards Framework (2000, 2001, 2002) current metadata registry applications can be found in areas such as interoperability testing technology, M2M interfaces, smart card technology, XML registry framework and standards and statistical office development programs.

Feedback from the EDR users was valuable to the improvement of System of Registries. This was demonstrated in the SoR questionnaire survey were most of the respondents

deemed the search option the strongest feature of the System of Registries although they have pointed out that in some occasions it remained slow. The new integrated search option (both for the SRS and the main search option of System of Registries) had been one of the improvements based on the previous year's conference feedback. The System of Registries in itself had been noted by one of the respondents as an important feature as it facilitates the exchange and documentation of information. The service that needed improvement as respondents indicated is the SRS. Update and upload of more information, the improvement of the domain lists of substances and physical measures and the search option within SRS have been some of the suggestions. Better and simpler publicising of the services and the expansion into non-environmental reporting were others. There have been plenty of comments on how the System of Registries can improve current work of its users. Those included expansion of collaboration with other governmental organisations and countries, transfer of ideas and System of Registries processes into other organisations, data validation, and exchange of information and avoidance of duplication of efforts.

The interviewees denoted that among the most important features of any metadata registry system were those interlinked with having an authority control agency that looks after the management and maintenance of metadata. In particular one of the interviewees noted, *"...for each of those systems, the consistent management of metadata about the registered material is crucial"*.

In particular the stressed the importance of user support services such as user guides, business rules documents or other user documentation and functional site maps. The ability to navigate in a quick and effective way was considered important by all users. The search options attracted particular mentions as well as they provide the main point to determine accessibility to useful information. Last interoperability support functions were also high in the users list as those interested in metadata use were also interested in exchange of information and enhancing research in general.

The users provided useful feedback about the functions and features of metadata registry systems that they considered overstated. They pointed out that misconceptions about what metadata registry systems are able to do could prove misleading. The use of XML as a "one – solves – all" solution had also been pointed out by the users as an overstated feature in a metadata registry system. The users suggested that it is more important to have ensured policies that mandate a formalised data process in order to sustain use and exchange of information. The same point was made for interoperable systems. In order to be able to facilitate efficient exchange of information across networked systems it is

important to define the context of use, to have policies that mandate the deposition and re use of metadata and having common systems that store that information.

Barriers to the use of metadata registry systems have been identified as those imposed by the individual, those imposed by the organisation and reasons outside of the users' control. The main barriers have been the lack of time and limitations imposed by previous knowledge in the use of a system. Hsieh-Yee (2001) lists among the factors that affect the way a user's information seeking behaviour *"...a user's background and experience with computers, the Web, and other information retrieval* tools can affect how he or see seeks information. Information need, domain knowledge, cognitive abilities, affecting states, demographics and the environment of the information need also contribute to the way in which the seeker seeks information". The major barriers imposed by an organisation have been identified as the lack of formal policies that define the data process, the lack of information regarding the benefits of using metadata registry systems and the long time and/or slow movement that sometimes is associated with the acceptance of introducing new systems. Other barriers were pointed out to be the user friendliness of a system, the usability and its functionality, the lack of supportive services in such a system and the reluctance of people to respond to change. Furthermore, changes in a system's structure or the re-design of a web site could prove to be a barrier in the use of the system. The failed requests reports recorded as the most common reasons for not accessing information in a system the unavailability of the document, badly send requests to the server and forbidden access to the document.

Non use of metadata registry systems has been associated with not always been able to identify the users of metadata registry systems due to several reasons including confidentiality restrictions imposed by the organisation that hosts the registry, the nature of online distance services that tend to de personalise users, and the level of deployment of such systems at the moment of the current research study. Furthermore, the lack of advocacy regarding the benefits that could be obtained from the use of metadata registry systems and the unavailability of an organisational requirement for their use result in non use as well.

5.1.5 User expectations

Users of the SoR searched for data standards, XML tags, associated business rules documents, and metadata pertaining to individual data elements that explained their usage and new features such as XML tags – available through the EDR – appeared to be popular as well. Comments and suggestions on the System of Registries services were

rich and representative of the interest in System of Registries. Expanding collaboration with other governmental organizations and countries, transfer of ideas and System of Registries processes into other organizations, data validation, exchange of information and avoidance of duplication of efforts were among those issue identified by users as key next steps.

5.1.6 XML and metadata registry systems

The use of XML was considered one of the desired features in a metadata registry system as the users of the EDR noted in the first EDR questionnaire survey in 2001. It had been noted that XML was one of the possible options that would enable the exchange of metadata across heterogeneous systems. Although the use of XML tags had been added as a new feature in the SoR the second questionnaire survey and response from the email interviews showed that users of the registry systems did not make extensive use of this feature. One of the interviewees in particular noted that they considered the omnipotence of XML as an overstated feature in a metadata registry system. In fact, it was stressed that it was considered crucial to have an underlying framework that mandates and supports the use of metadata for the description, use and management of information.

5.2 Limitations of the study

A limitation of this study is to generalise its results and conclude that the findings of this study show how the majority of end-users use and perceive metadata registry systems. In order to support this, a greater number of metadata registry systems and end-users should be studied.

As discussed earlier, users of metadata registry systems are difficult to locate due to several reasons including confidentiality restrictions imposed by the organisation that hosts the registry, the nature of online distance services that tend to de personalise users, and the level of deployment of such systems at the moment of the current research study. Those who have responded to surveys enquiring about metadata registry system users often distanced themselves from the perceived end-user and focused in the main area of their job responsibilities such as development or implementation or management or evaluation of systems only. The MetaForm registry and the Environmental Data Registry and the System of Registries did not employ a registration system for their users. Therefore, it was not possible to know the exact number of the end-users and demonstrate how the majority of metadata registry systems' users use

and perceive them. Furthermore as the only data that is known about the number of accesses to metadata registry systems are provided by the web log transactions, there is not information about the actual number of users who accessed the systems every day, month or year. The only data that is known is the number of accesses/ file requests and the IPs numbers that specify the location that those accesses had taken place from.

For example, regarding the Environmental Data Registry and the System of Registries surveys in the space of a month when the research took place there were approximately 10.000 requests for files. But, it is not known the number of users who carried out these requests. Regarding, the IPs number it is possible to reveal that these requests occurred from different domains, such as government organisations or profession bodies, or academic institutions but, it is not possible to figure out the number of individual users who accessed the Environmental Data Registry/System of Registries.

5.3 Future work

This study investigated the use and functionality of metadata registry systems. Results indicated that such systems are a vital part in the process of data standardisation and metadata management and it is believed that that research in the area is going to grow in the future, particularly after the establishment of the software to support the mappings and interoperability among different schemas.

This has been a first of its kind assessment of the use and functionality of two active metadata registry systems based on the views of their users. The combination of other research methods such as focus groups with the systems' users would enhance our insight of metadata registry systems usage. Also, additional methods to identify metadata registry systems' users such as a registration facility to those systems would prove beneficial to the identification of the systems' users. It is essential though to maintain the open access availability as this has been found to be a restricting factor in the use of similar services such as subject gateways (Monopoli, 2005).

Bibliography

Acherman, M. S. and Fielding, R. T. (1995). Collection maintenance in the digital library, in, *Digital Libraries '95: The Second Annual Conference on the Theory and Practice of Digital Libraries.* Texas, USA. Available at: http://www.csdl.tamu.edu/DL95/papers/ackerman/ackerman.html (Last accessed: 12/06/2006)

Ahronheim, J. R. (1998). Descriptive metadata: emerging standards. *The Journal of Academic Librarianship* 24(5), pp. 395-403.

Allard, S. (2000). *Digital Libraries: A frontier for LIS education. ALISE Annual Conference.* San Antonia, Texas. Available at: http://www.alise.org/conferences/conf00_Allard_Digital_Libraries.htm (Last accessed: 12/06/2006)

Antelman, K. (1999). Web lists and the decline of the library catalog. *Library Computing* 18(3), pp. 189-195.

Arms, C. R. (2000). Some observations on metadata and digital libraries, in *Conference on Bibliographic Contron in the New Millennium (Library of Congress).* Washington: Library of Congress. Available at: http://lcweb.loc.gov/catdir/bibcontrol/arms_paper.html (Last accessed: 12/06/2006)

Armstrong, C, Barker, A, Everitt, J, Fenton, R, Lonsdale, R, Stoker, D, Thomas R & Urquhart, C. (2001). The JISC usage surveys: trends in electronic information services (JUSTEIS) project - Supply and demand in higher education. *Library & Information Briefings* 106/107. London: Library Information Technology Centre, South Bank University, July 2001, 18 pages.

Ayres, M.-L., Kilner, K., Fitch, K., Scarvell, A. (2002). Report on the successful AustLit: Australian literature gateway implementation of the FRBR and INDECS event models, and implications for other FRBR implementation. 68th IFLA conference and general conference. IFLA. Available at: http://www.ifla.org/IV/ifla68/papers/054-133e.pdf (Last accessed: 12/06/2006)

Baker, T. (1996). Introduction in 2nd *DELOS Workshop: Metadata and interoperability in digital library related fields.* Bonn, 7-8 October 1996.

Baker, T. (1999). TIAC *White Paper on appropriate technology for digital libraries.* Technical Information Access Center. Available at: http://server.tiac.or.th/tiacweb/Baker/Title.html (Last accessed: 12/06/2006)

Baker, T. (2000). A grammar for Dublin Core. *D-Lib Magazine* 6(10). Available at: http://dlib.org/dlib/october00/baker/10baker.html (Last accessed: 12/06/2006)

Baker,T. (2001). What terms does your metadata use? Application profiles as machine - understandable narratives. DC-2001, International Conference on Dublin Core and

Metadata Applications. Japan. Available at: http://www.nii.ac.jp/dc2001/proceedings/product/paper-25.pdf (Last accessed: 12/06/2006)

Bargemeyer, B. E. and Gillman, D. W. (2000). Metadata Standards and Metadata Registries: An Overview, *in International Conference on Establishment Surveys II. Buffalo, New York*. Available at: http://www.bls.gov/ore/pdf/st000010.pdf (Last accessed: 12/06/2006)

Bates, M. (1998). Indexing and access for digital libraries and the Internet: Human, database and domain factors. *Journal of the American Society for Information Science* 49(13), pp. 1185-1205

Bawden, D. and Robinson, L. (1999). Internet subject gateways (Review). *International Journal of Information Management* 19 pp. 511-522.

Bawden, D. and Robinson, L. (2002). Internet subject gateways revisited (Review). *International Journal of Information Management* 22 pp. 157-162.

Bawden, D. and Rowlands, I. (1999). *Understanding digital libraries: towards a conceptual framework*. London: City University, pp. 40.

Bearman, D. (1999). A common model to support interoperable metadata: Progress report on rconcilling metadata requirements from the Dublin Core and INDECS/DOI communities. *D-Lib Magazine*. Available at: http://www.dlib.org/dlib/january99/bearman/01bearman.html (Last accessed: 12/06/2006)

Belkin, N. J. (2002). A framework for criteria and measures for evaluation of user interfaces in digital libraries. In, Borgman, C. (ed.) et al, Evaluation of digital libraries: testbeds, measurements, and metrics. Fourth DELOS workshop, Budabest, 6-7 June, 2002, pp.33-36.

Berg, B. L. (1998). *Qualitative research methods for the social sciences*. Boston, Allyn and Bacon.

Berners-Lee, T. (1998). Web design issues: What a semantic web can represent. Available at: http://www.w3.org/DesignIssues/RDFnot.html (Last accessed: 12/06/2006)

Berners-Lee, T., Hender, J., Lassila, O. (2001). The semantic web. *The Scientific American*, 284 (5), pp. 34-43.

Bertot, J. C., McClure, R., Moen, W.E., Rubin, J. (1997). Web usage statistics: Measurement issues and analytical techniques. *Government Information Quarterly* 14(4), pp. 373-395.

Bishoff, L. (2000). Metadata, cataloging, digitization and retrieval: Who's doing what to whom: The Colorado digitization project experience, in *Bicentennial Conference on Bibliographic Control for the New Millennium: Confronting the challenges of networked resources and the Web*. Washington: Library of Congress. Available at: http://lcweb.loc.gov/catdir/bibcontrol/bishoff_paper.html

Bishop, A. P., Neumann, L.J., Leigh-Starr, S., Merkel, C., Ignacio, E., Sandusky, R.J. (2000). Digital libraries: Situating use in changing information infrastructure. *Journal of the American Society for Information Science* 51(4), pp. 394.

Blanchi, C. and Petrone, J. (2001). Distributed interoperable metadata registry. *D-Lib Magazine*. Available at: http://www.dlib.org/dlib/december01/blanchi/12blanchi.html, (Last accessed: 12/06/2006)

Borgman, C. L. (1996). Why are online catalogs still hard to use? *Journal of the American Society for Information Science* 47(7), pp. 493-503.

Bosnjak, M. (2001). Classifying response behaviors in web-based survyes. *Journal of Computer-Mediated Communication*, 6 (3). Available at: http://www.ascusc.org/jcmc/vol6/issue3/burton.html (Last accessed 12/06/2006)

Bowers, S. and Delcambre, L. (2000). Representing and transforming model-based information. ECDL 2000: Workshop on the semantic web. Lisbon, Portugal: ERCIM. Available at: http://www.ics.forth.gr/isl/SemWeb/proceedings/session1-1/paper.pdf (Last accessed: 12/06/2006)

Boyce, B. R., Meadow, C.T., Kraft, D.H. (1994). *Measurement in information science*, Academic Press.

Branin, J., Groen, F., Thorin, S. (2000). The changing nature of collection management in research libraries. *Library Resources & Technical Services* 44(1), pp.23-32.

Brasethvik, T. (1998). A semantic modeling approach to metadata. *Internet Research: Electronic Networking Applications and Policy* 8(5), pp. 377-386.

Brisson, R. (1998). The world discovers cataloging: a conceptual introduction to digital libraries, metadata and the implications for library administrations. *Journal of Internet Cataloguing* 1(4), pp. 1998

Burnett, K. and Park, B. N. (1999). A comparison of the two traditions of metadata development. *Journal of the American Society for Information Science* 50(13), pp. 1209-1217.

Burton, M. (2001). The value of web log data in use-based design and testing. *Journal of Computer-Mediated Communication*, 6 (3). Available at: http://www.ascusc.org/jcmc/vol6/issue3/burton.html (Last accessed 12/06/2006)

Busha, C. H. (1980). *Research methods in librarianship: Techniques and interpretation*, Academic Press.

Butterfield, K. L. (1997). Cataloger's and the creation of metadata systems. Available at: http://www.oclc.org/oclc/man/colloq/butter.htm (Last accessed: 12/06/2006)

Caplan, P. (2003). *Metadata fundamentals for all librarians*. Chicago: American Library Association, 2003.

Caplan, P. (1995). You call it corn, we call it syntax-independent metadata for document like objects. *The Public - Access Computer Systems Review*. Available at: info.lib.uh.edu/pr/v6/n4/capl6n4.html (Last accessed: 12/06/2006)

Caplan, P. (2000). International metadata initiatives: Lessons in bibliographic control. Library of Congress, in, *Bicentennial Conference on Bibliographic Control for the new Milennium: Confronting the challenges of networked resources and the web. Washington: Library of Congress.* Available at: http://lcweb.loc.gov/catdir/bibcontrol/caplan_paper.html (Last accessed: 12/06/2006)

Carvalho, J. d. and Cordeiro, M. I. (2002). XML and bibliographic data: the TVS (Transport, Validation and Services) model. 68th IFLA General Conference and Council. Libraries for Life: Democracy, Diversity, Delivery. Glasgow: IFLA. Available at: http://www.ifla.org/IV/ifla68/papers/075-095e.pdf (Last accessed: 12/06/2006)

Cathro, W. S. (1999). Digital Libraries: A national perspective. Information Online & On Disc Conference. Sydney. Available at: http://www.nla.gov.au/nla/staffpaper/cathro4.html (Last accessed: 12/06/2006)

Chowdhury, G. G. (2002). Digital libraries and reference services: Present and future. *Journal of Documentation* 58(3), pp. 258-283.

Couper, M. (2001). Web surveys: a review of issues and approaches. *Public opinion quarterly*, 64, pp.464-494.

Covi, L. and Kling, R. (1996). Organizational dimensions of effective digital library use: closed rational and open natural systems models. *Journal of the American Society for Information Science* 47(9), pp. 672-689.

Dalrymple, P. W. (2001). A quarter century of user-centered study. The impact of Zweizig and Dervin on LIS research. *Library & Information Science Research* 23 pp. 155-165.

Day, M. (1999). The metadata challenge for libraries: A view from Europe. In Kaser, R. T. and Cox Kaser, V. Metadiversity: responding to the grand challenge for biodiversity information management through metadata. Philadelphia, Penn: National Federation of Abstracting and Information Services, pp. 131-140.

Day, M., Heery, R., Powell, A. (1999). National bibliographic records in the digital information environment: Metadata, links and standards. *Journal of Documentation* 55(1), pp. 16-32.

Day, M. (2001). Metadata in a nutshell. Available at: http://www.ukoln.ac.uk/metadata/publications/nutshell/ (Last accessed: 12/06/2006)

Deacon, P. (2001). Using metadata to create navigation paths in the Healthnsite Internet gateway. *Health Information and Libraries Journal* 18 pp. 20-29.

Debowski, S. (2000). The hidden user: providing an effective service to users of electronic information sources. *OCLC Systems & Services* 16(4), pp. 175-180.

Dekkers, M. and Weibel, S. (2003). State of the Dublin Core metadata initiative, April 2003. *D-Lib Magazine*. Available at: http://www.dlib.org/dlib/april03/weibel/04weibel.html, (Last accessed: 12/06/2006)

Dempsey, L. (1989) as cited by R. Heery, M. Day and A. Powel in National Bibliographic records in the digital information environment: metadata, links and standards. *Journal of Documentation*, Vol.55 (1), p.21

Dempsey, L. (1996). ROADS to DESIRE: Some UK and other European metadata and resource discovery projects. *D-Lib Magazine*. Available at: www.dlib.org/dlib/july96/07dempsey.html, (Last accessed: 12/06/2006)

Dempsey, L. (2000). The subject gateway: Experiences and issues based on the emergence of the Resource Discovery Network. *Online Information Review* 24(1), pp. 8-23.

Dempsey, L. and Heery, R. (1998). Metadata: A current view of practice and issues. *Journal of Documentation* 54(2), pp. 145-172.

Dillon, M. (2000). Metadata for web resources: How metadata works on the web. *Library of Congress Bicentennial Conference on Bibliographic Control for the New Milennium: Confronting the challenges of networked resources and the web. Washington: Library of Congress.* Available at: http://lcweb.loc.gov/catdir/bibcontrol/dillon_paper.html (Last accessed: 12/06/2006)

Dowling, T. (2001). Lies, damned lies, and web logs. *Library Journal / Net Connect* 126(7), pp. 34-35

Drewry, M., et al. (1997). Metadata: Quality vs. Quantity. The Second IEEE Metadata Conference. Maryland, USA: IEEE. Available at: http://www.llnl.gov/liv_comp/metadata/md97.html (Last accessed 12/06/2006)

Duke, M., et al. (2005). Enhancing access to research data: the challenge for crystallography, in, Proceedings of the *5th ACM/IEEE-CS joint conference on Digital libraries,* Denver, CO, USA June 07 - 11, 2005. Available at: http://delivery.acm.org/10.1145/1070000/1065397/p46-duke.pdf?key1=1065397&key2=3173447411&coll=portal&dl=ACM&CFID=46117110&CFTOKEN=8390583 (Last accessed: 12/06/2006)

Dunham, B. (2002). Different formats: Linking serial titles for display through bibliographic relationships. Is it possible? *Library Collections, Acquititions & Technical Services* 26 pp. 3-17.

Duval, E., et al. (2002). Metadata principles and practicalities. *D-Lib Magazine*. Available at: Available at: http://www.dlib.org/dlib/april02/weibel/04weibel.html, (Last accessed: 12/06/2006)

Eden, B. (2002). Metadata and its application. *Library Technology Reports*, 38 (5), 62p.

El- Sherbini, M. (2001). Metadata and the future of cataloguing. *Library Review* 50(1), pp. 16-27.

Farber, M. and Shoham, S. (2002). Users, end-users, and end-user searchers of online information: A historical overview. *Online Information Review* 26(2), pp. 92-100.

Fattahi, R. (1995). Relevance of cataloguing principles to the online environment: A historical and analytical study. Sydney, Australia: A Ph. D. Thesis submitted to the University of New South Wales, Sydney, Australia. Available at: http://www.um.ac.ir/~fattahi/thesis1.htm (Last accessed: 12/06/2006)

Fattahi, R. (1997). AACR2 and catalogue production technology. International conference on the principles and future development of AACR. Toronto, Canada. Available at: http://collection.nlc-bnc.ca/100/200/300/jsc_aacr/aacr_cat/r-aacr2.pdf (Last accessed: 12/06/2006)

Feeney, M. (1999). Digital culture: maximising the nation's investment: a synthesis of JISC/NPO studies of the preservation of electronic materials. London, The British Library.

Feldman, S. (1995). Internet stabilizers. *Searcher* 3(9), pp. 32-41.

Feldman, S. (1997). Advances in digital libraries 1997. *Information Today* 14(7), pp. 12-15.

Fielding, R. ...et al. (1999). RFC 2616. Hypertext Transfer Protocol -- HTTP/1.1. Available at: http://www.ietf.org/rfc/rfc2616.txt (Last accessed: 12/06/2006)

Fink, A. and Kosecoff, J. (1998). *How to conduct surveys: A step by step guide*. London, Sage.

Flaherty, P. (1993). Transaction logging systems: A descriptive summary. *Library Hi Tech* 11(2), pp. 67-77.

Ford, N., et al. (2001). Web search strategies and retrieval effectiveness: An empirical study. *Journal of Documentation* 58(1), pp. 30-48.

Fox, J., et al. (2002). Conducting research using web-based questionnaires: practical, methodological, and ethical considerations. *International Journal of Social Research Methodology* 6(2), pp.167-180.

Gaiser, T. J. (1997). Conducting on-line focus groups: a methodological discussion. *Social Science Computer Review* 15(2), pp. 135-144.

Gallagher, L. and Carnahan, L. (2000). A general purpose registry/repository information model. National Institute of Standards and Technology, pp. 32. Available at: http://lists.oasis-open.org/archives/regrep/200010/pdf00000.pdf (Last accessed: 12/06/2006)

Gardner, T. and Iannella, R. (2000). Architecture and software solutions. *Online Information Review* 24(1), pp. 35-39.

Gill, T., et al. (1998). *Introduction to metadata: Pathways to digital information*. Version 2. Getty Institution. Available at:

www.getty.edu/research/institute/standards/intrometadata/index.html, (Last accessed: 12/06/2006)

Gilleson, M. L. and Frost, R. D. (1993). The evolution of the met-data concept: Dictionaries, catalogs and repositories. *Journal of Database Management* 4(3), pp. 17-26.

Gleason, D. (1999). Business metadata. *Enterprise Systems Journal* 58(1), pp. 14-19.

Gorman, M. (1999). Metadata or cataloguing? A false choice. *Journal of Internet Cataloguing* 2(1), pp. 5-22.

Gorman, M. (2000). From card catalogues to WEBPACs: Celebrating cataloging in the 20th century. *Library of Congress Bicentennial Conference on Bibliographic control for the New Milennium*. Washington: Library of Congress. Available at: lcweb.loc.gov/catdir/bibcontrol/gorman_paper.html (Last accessed: 12/06/2006)

Gradmann, S. (1998). Cataloguing vs. metadata: old wine in new bottles? 64th IFLA General Conference. Amsterdam: IFLA. Available at: http://www.ifla.org/IV/ifla64/007-126e.htm

Graham, P. S. (1995). The digital research library: tasks and commitments. *Digital Libraries '95: The Second Annual Conference on the Theory and Practice of Digital Libraries*. Texas, USA. Available at: http://www.csdl.tamu.edu/csdl/DL95/papers/graham/graham.html (Last accessed: 12/06/2006)

Grandcolas, Ursula, et al. (2003). Web survey bias: sample or mode effect? *Journal of Marketing Management*_, 19, pp.541-561

Gredley, E. and Hopkinson, A. (1990). *Exchanging bibliographic data: MARC and other formats*. Washington, American Library Association.

Greenberg, J., et al. (2003). Metadata: A fundamental component of the semantic web. *Bulletin of the American Society for Information Science* (April/May), pp. 16-18.

Greene, S., et al. (2000). Previews and overviews in digital libraries: designing surrogates to support visual information seeking. *Journal of the American Society for Information Science* 51(4), pp. 380-393.

Groetschel, M. and Luegger, J. (1998). *Scientific information systems and metadata. Classification in the Information Age. Proceedings of the 22nd Annual GfKl Conference*. Dresden: Springer. Available at: http://elib.zib.de/ftp/pub/UserHome/Luegger/Dresden/Metadata.htm (Last accessed: 12/06/2006)

Guinchard, C. (2002). Dublin Core use in libraries: a survey. *OCLC Systems & Services* 18(1), pp. 40-50.

Haag Granello, D. (2004). Online data collection. Journal of Counselling & Development 82, pp.387-393.

Hakala, J. (1998). Internet metadata and library cataloguing. "The Function of Bibliographic Control in the Global Information Infrastructure". Vilnius, Lithuania: IFLA. Available at: http://www.lnb.lt/events/ifla/hakala.html (Last accessed: 12/06/2006)

Hannabuss, S. (1995). Approaches to research. *Aslib Proceedings* 47(1), pp. 3-11.

Haynes, D. (2004). Metadata for information management and retrieval. London: Facet publishing, 2004.

Heaney, M. (1995). Object - oriented cataloguing. *Information Technology and Libraries* 14(3), pp. 135-153.

Heery, R. (1996). Review of metadata formats. *Program* 30(4), pp. 345-373.

Heery, R. (1997). Metadata corner: naming names: metadata registries. *Ariadne* 11. Available at: http://www.ariadne.ac.uk/issue11/**metadata**/ (Last accessed: 12/06/2006)

Heery, R. (2000). Information gateways: Collaboration on content. *Online Information Review* 24(1), pp. 40-45.

Heery, R. (2001). Draft DCMI Open Metadata Registry Functional Requirements. Available at: http://dublincore.org/groups/registry/fun_req_ph1-20011031.shtml (Last accessed: 12/06/2006)

Heery, R. (2002). Functional Requirements for CORES schema creation and registration tool. UKOLN - CORES. Available at: http://www.cores-eu.net/registry/d22/funcreq.html, (Last accessed: 12/06/2006)

Heery, R. and Dempsey, L. (1998). Metadata: A current view of practices and issues. *Journal of Documentation* 54(2), pp. 145-172.

Heery, R. and Patel, M. (2000), Application profiles: mixing and matching metadata schemas. Ariadne, http://www.ariadne.ac.uk/issue25/app-profiles/ (Last accessed: 12/06/2006)

Heery, R., Powell, A., Day, M. (1997). Metadata. *Library & Information Briefings* (75), 19 pages.

Heery, R. and Wagner, H. (2002). A metadata registry for the semantic Web. *D-Lib Magazine*. Available at: http://www.dlib.org/dlib/may02/wagner/05wagner.html (Last accessed: 27/11/2003)

Hernon, P. (2002). Quality: New directions in the research. *The Journal of Academic Librarianship* 28(4), pp. 224-231.

Hjorland, B. (2000). Library and Information science: practice, theory, and philosophical basis. *Information Processing & Management* 36(3), pp. 501-531.

Hopkinson, A. (1998). Traditional communication formats: MARC is far from dead. *The function of Bibligraphic Control in the Global Information Infrastructure*. Vinius, Lithuania: IFLA. Available at: http://www.lnb.lt/events/ifla/hopkinson.html (Last accessed: 12/06/2006)

Hsieh-Yee, I. (2000). Organizing Internet resources: teaching cataloging standards and beyond. *OCLC Systems & Services* 16(3), pp. 130-143.

Hsieh-Yee, I. (2001). Research on web searh behavior. *Library & Information Science Research* 23 , pp. 167-185.

Hunter, J. (2003). A survey of metadata research for organizing the web. *Library trends*, 52(2), pp.318-344.

Hunter, J. and Armstrong, L. (1999). *A comparison of schemas for video metadata representation. The Eighth International World Wide Web Conference. Toronto, Canada.* Available at: http://www8.org/w8-papers/3c-hypermedia-video/comparison/comparison.html (Last accessed: 12/06/2006)

Hyun - Hee, K. and Chang - Seok, C. (2000). XML: how it will be applied to digital library systems. *The Electronic Library* 18(3), pp. 183-189.

IFLA (1998). Functional Requirements for bibliographic records: a final report. Munchen: IFLA, pp. 144. Available at: http://www.ifla.org/VII/s13/frbr/frbr.pdf and http://www.ifla.org/VII/s13/frbr/frbr.htm (Last accessed: 12/06/2006)

Igbaria, M. (1999). The driving forces in the virtual society. *Communications of the ACM* 42(12), pp. 64-70.

ISO9241-11 (1998). *Ergonomic requirements for office work with visual display terminals (VDTs) – Part 11: Guidance on usability.* Available at http: www.iso.ch

ISO/IEC (1995-2003). *ISO/IEC 11179 Standard. Information technology -- Specification and standardization of data elements.* Parts 1-6. Available at: http://www.iso.ch

Im, Eun-Ok and Chee, Wonshik (2003). Issues in Internet research. *Nursing Outlook*, 51, pp.6-12

Jansen, B. J. and Pooch, U. (2001). A review of web searching studies and a framework for future research. *Journal of the American Society for Information Science* 52(3), pp. 235-246.

Jansen, B. J. and Spink, A. (2000). Methodological approach in discovering user search patterns through web log analysis. *Bulletin of the American Society for Information Science* (October/November), pp. 15-17.

Johnston, P. (2001). Interoperability: supporting effective access to information resources. *Library & Information Briefings* 108 pp. 10

Johnston, P. (2002). Functional Requirements for MEG registry. UKOLN - MEG. Available at: http://www.ukoln.ac.uk/metadata/education/regproj/funcreq/20020519.html, (Last accessed: 12/06/2006)

Jones, S., Gatford, M. , Do, T. (1997). Transaction logging. *Journal of Documentation* 53(1), pp. 35-50.

Jul, E. (1997). Cataloging Internet resources: survey and prospectus. *Bulletin of the American Society for Information Science* 24(1),

Kaske, N. K. (1993). Research methodologies and transaction log analysis: Issues, questions, and a proposed model. *Library Hi Tech* 11(2), pp. 79-85.

Kassam, S. (2002). Freedom of information. *Intelligent Enterprise* pp. 49-51.

Kaye Gapen, D. (1993). The virtual library: knowledge, society, and the librarian. In Saunders, L. M. The virtual library: visions and realities. London: Meckler.

Kebede, G. (2002). The changing information needs of uses in electronic information environment. *The Electronic Library* 20(1), pp. 14-21.

Kikki, J. (2000). A new model for electronic recordkeeping in the Finnish defence forces. *Records Management Quarterly* 10(3), pp. 150-160.

King, D. W. and Tenopir, C. (1999). Using and reading scholarly literature. *Annual Review of Information Science and Technology* 34 pp. 423-477.

Kirriemuir, J. (1999). A brief survey of quality resource discovery systems. Commissioned by the JISC-funded Resource Discovery Network Centre, Available at: http://www.rdn.ac.uk/publications/studies/survey/ (Last accessed: 12/06/2006)

Koch, T. (2000). Quality controlled gateways: Definitions, typologies, empirical view. *Online Information Review* 24(1), pp. 24-34.

Kuny, T. and Cleveland, G. (1998). The digital library: myths and challenges. *IFLA Journal* 24(2), pp. 107-113.

Kurth, M. (1993). The limits and limitations of transaction log analysis. *Library Hi Tech* 11(2), pp. 99-103

Lagoze, C. (1996). The Warwick framework: A container architecture for diverse sets of metadata. *D-Lib Magazine*. Available at: www.dlib.org/dlib/july96/lagoze/07lagoze.html, (Last accessed: 12/06/2006)

Lagoze, C. (2000). An event- aware model for metadata interoperability. European Conference on Research and Advanced Technology for Digital Libraries (ECDL). Lisbon, Portugal. Available at: http://www.cs.cornell.edu/lagoze/papers/ev.pdf (Last accessed: 12/06/2006)

Landsbergen, D. J. (2001). Realizing the promise: government information systems and the fourth generation of information technology. *Public Administration Review* 61(2), pp. 206-218.

Large, A., Tedd, L., Hartley, R.J. (1999). Information seeking in the online age: Principles and practice, Bowker-Saur.

Lee, H.-L. (2000). What is a collection? *Journal of the American Society for Information Science* 51(12), pp. 1106-1113.

Lee, T. B (1989). Information management: A proposal. Available at: http://www.w3.org/History/1989/proposal.html (Last accessed: 12/06/2006)

Lee-Smeltzer, K.-H. (2000). Finding the needle: controlled vocabularies, resource discovery, and Dublin Core. *Library Collections, Acquititions & Technical Services* 24 pp. 205-215.

LeFurgy, W. G. (2002). Levels of services for digital repositories. *D-Lib Magazine* 8(5),

Leviston, T. (2001). Describing metadata registry requirements and realities. Australia: Distributed Systems Technology Centre, pp. 23. Available at: http://www.dstc.edu.au/Research/Projects/Infoeco/publications/registry-discussion.pdf (Last accessed: 12/06/2006)

Levy, D. M. (1995). Cataloguing in the digital order. Digital Libraries '95: The Second Annual Conference on the Theory and Practice of Digital Libraries. Austin, Texas, USA. Available at: http://csdl.tamu.edu/DL95/papers/levy/levy.html (Last accessed: 12/06/2006)

Levy, D. M. and Marshal, C. C. (1995). Going digital: a look at assumptions underlying digital libraries. *Communications of the ACM* 38(4), pp. 78-84.

Levy, D. M. and Marshall, C. C. (1994). Washington's White horse? A look at assumptions underlying digital libraries. Digital Libraries 94: Proceedings of the First Annual Conference on the Theory and Practice of Digital Libraries. Texas, USA. Available at: http://www.csdl.tamu.edu/csdl/DL94/paper/levy.html (Last accessed: 12/06/2006)

Lynch, C. (1997). Searching the Internet. *Scientific American* 397

Madison Dean, O. (1998). Standards in Light of New Technologies Functional Requirements For Bibliographic Records. The Function of Bibliographic Control in the Global Information Infrastructure". Vilnius. Available at: http://www.lnb.lt/events/ifla/dean.html (Last accessed: 12/06/2006)

Maguire, M. (2001). Context of Use within usability activities. *International Journal of Human-Computer Studies* 55(4), pp. 453-483.

Marchionini, G. (2002). Co-Evolution of user and organizational interfaces: A longitudinal case study of WWW dissemination of national statistics. *Journal of the American Society for Information Science* 53(14), pp. 1192-1209.

Mariner, V. (2002). Logging usability. *Library Journal / Net Connect* (Winter 2002), pp. 30-31.

Marshall, C. C. Making metadata: a study of metadata creation for a mixed physical - digital collection. Digital Libraries '98: Third ACM Conference on Digital Libraries. Pittsburgh, USA. Available at: http://www.csdl.tamu.edu/~marshall/dl98-making-metadata.html (Last accessed: 12/06/2006)

Marshall, C. C. (1998). The Future of Annotation in a Digital (Paper) World. 35th Annual GSLIS Clinic: Successes and Failures of Digital Libraries. Available at: http://csdl.tamu.edu/~marshall/uiuc-paper-complete.pdf (Last accessed: 12/06/2006)

Miksa, F. L. and Doty, P. (1994). Intellectual realities and the digital library. Digital Libraries 1994: Proceedings of the First Annual Conference on the Theory and Practice of Digital Libraries. Texas, USA. Available at: http://www.csdl.tamu.edu/DL94/paper/miksa.html (Last accessed: 12/06/2006)

Miller, P. (2000). Interoperability. What is it and Why should I want it? Available at: http://www.ariadne.ac.uk/issue24/interoperability/intro.html (Last accessed: 12/06/2006)

Miller, T. (2001). Can we trust the data of online research? Marketing Research, 13, pp. 26-32.

Milstead, J. and Feldman, S. (1999). Metadata: Cataloguing by any other name. *Online*. Available at: http://www.infotoday.com/online/OL1999/milstead1.html, (Last accessed: 12/06/2006)

Moen, W. (2001). The metadata approach to accessing government information. *Government Information Quarterly* 18(3), pp. 155-166.

Moen, W. (2001). Assessing interoperability in the networked environment: Standards, evaluation, and testbeds in the context of Z39.50. In McClure, C. R. and Bertot, J. C. Evaluating networked information services: Techniques, policy and issues. Silver Spring, MD: American Society of Information Science and Technology.

Moen, W. (2001). Mapping the interoperability landscape for networked information retrieval. Proceedings of the first ACM/IEEE-CS joint conference on Digital libraries. ACM. Available at: http://doi.acm.org/10.1145/379437.379447 or http://www.unt.edu/wmoen/publications/MapInteropJCDLFinal.pdf (Last accessed: 12/06/2006)

Monopoli, M. (2005). User-Based Evaluation of Academic Digital Libraries: case studies Social Science Information Gateway (SOSIG), Art Design Architecture & Media Gateway (ADAM) and the Electronic Journals Service of the University of Patras, Greece. PhD thesis. City University.

Monopoli, M. and Nicholas, D. (2001). A user evaluation of Subject Based Information Gateways: case study ADAM, *Aslib Proceedings*, 53 (1), pp. 39-52

Montoya - Weiss, M. M., et al. (1998). Online focus groups: conceptual issues and a research tool. European Journal of Marketing 32(7/8), pp. 713-723.

Morgan, D. L. (1998). Focus groups as qualitative research. Qualitative Research Methods. London, Sage.

Moyo, L. M. (2002). Collections on the Web: some access and navigation issues. *Library Collections, Acquitions & Technical Services* 26 pp. 47-59.

Mullen, A. (2001). GILS metadata initiatives at the state level. *Government Information Quarterly* 18 pp. 167-180.

Nagamori, M., Baker, T. , Sakaguchi, T., Sugimoto, S. , Tabata. K. (2001). A multilingual metadata schema registry based on RDF schema. DC-2001, International Conference on

Dublin Core and Metadata Applications. Japan. Available at: http://www.nii.ac.jp/dc2001/proceedings/product/paper-31.pdf (Last accessed: 12/06/2006)

Newton, J. (1996). Applications of metadata standards. IEEE Meta96. Available at: http://www.computer.org/conferences/meta96/newton/paper.html, (Last accessed: 12/06/2006)

Newton, J. (1998). Using levels of astraction to name data elements. *The Data Administration Newsletter* 7

Nicholas, D. (1996). An assessment of the online searching behaviour of practitioner end users. The *Journal of Documentation* 52(3), pp. 227-251.

Nicholas, D., Huntington, P., Williams, P., Lievesley, N. Dobrowolski, T., Withey, R. (1999). Cracking the code: Web log analysis. *Online & CD-ROM Review* 23(5), pp. 263-269.

Nicholas, D., Huntington, P., Williams, P., Lievesley, N., Dobrowolski, T., Withey, R (1999). Developing and testing methods to determine the use of web sites: case study newspapers. *Aslib Proceedings* 51(5), pp. 144-154.

Nicholas, D. (2000). Assessing information needs: tools, techniques and concepts for the Internet age. *2nd ed.* London: The Association for Information Management, 2000.

Nicholas, D., Huntington, P., Williams, P. (2001). Evaluating metrics for comparing the use of web sites: A case study of two consumer health web sites. *Journal of Information Science* 28(1), pp. 63-75.

National Information Standards Organisation (2004). Understanding metadata. Available at: http://www.niso.org/standards/resources/UnderstandingMetadata.pdf (Last accessed: 12/06/2006)

Nuernberg, P. J., et al. (1995). Digital Libraries: Issues and architectures. Digital Libraries '95: The Second Annual Conference on the Theory and Practice of Digital Libraries. Austin, Texas, USA. Available at: http://www.csdl.tamu.edu/DL95/papers/nuernberg/nuernberg.html (Last accessed: 12/06/2006)

Oppenheim, A.N. (1992), Questionnaire design, interviewing and attitude measurement, London, Pinter Publishers.

Palmer, C. L. and Knutson, E. M. (2004). Metadata practices and implications for federated collections. Proceedings of the 67th ASIS&T Annual Meeting, 41, pp. 456-462.

Payette, S. and Lagoze, C. (2000). Value-added surrogates for distributed content. *D-Lib Magazine* 6(6),

Peters, T. A., et al. (1993). Transaction log analysis. *Library Hi Tech* 11(2), pp. 37-65.

Peterson, R. A. (2000). Constructing effective questionnaires. London, Sage.

Pierre, M. and LaPlant, P. (1998). Issues in crosswalking: Content metadata standards. NISO, Available at: http://www.niso.org/press/whitepapers/crsswalk.html (Last accessed: 12/06/2006)

Pouchard, L. (1998). Cataloging for digital libraries: the TEI scheme and the TEI header. *Katharine Sharp Review* 6

Powell, R. R. (1999). Recent trends in research: A methodological essay. *Library & Information Science Research* 21(1), pp. 91-119.

Preston, C. and Lin, B. (2002). Database technology in digital libraries. *Information Services & Use* 22 pp. 9-17.

PricewaterhouseCoopers (2000). Study on the Impact of the Telematics for Libraries FP4 Programme. *Exploit Interactive* 5 pp. 34.

Qin, J. (2000). Representation and organization of information in the web space: From MARC to XML. *Information Science* 3(2), pp. 83-87.

Quam, E. (2001). Informing and evaluating a metadata initiative: Usability and metadata studies in Minnesota's Foundations Project. *Government Information Quarterly* 18(3), pp. 181-194

Rocco, T.S., Bliss, L.A., Gallagher, S., Perez-Prado, A. Taking the next step: mixed methods reasearch in organisational systems. *Infomration Technology, Learning, and Performance Journal*, 21 (1), pp.19-29.

Rowlands, I. and Bawden, D. (1999). Building the digital library on solid research foundations. *Aslib Proceedings* 51(8), pp. 275-282.

Rushbridge, C. (1998). Towards the hybrid library. *D-Lib Magazine* July/August

Sandore, B. (1993). Applying the results of transaction log analysis. *Library Hi Tech* 11(2), pp. 87-97.

Sandore, B., Flaherty, P., Kaske, N.K. (1993). A manifesto: Regarding the future of transaction log analysis. *Library Hi Tech* 11(2), pp. 105-106.

Schatz, B. and Chen, H. (1999). Digital Libraries: technological advances and social impacts. *IEEE Computer* 32(2), pp. 45-50.

SCHEMAS Metadata watch reports. Available at: http://www.schemas-forum.org/stds-framework/ (Last accessed: 12/06/2006)

Schimmer, R. (1997). Metadata and META-LIB. The Metadata Initiative of German Libraries. Available at: http://www.mathematik.uni-osnabrueck.de/projects/workshop97/papers/schimmer10.12.97.html (Last accessed: 12/06/2006)

Sherman, R. P. (1997). Metadata: the missing link. DBMS 10(9), pp. 73-79.

Shuler, J. A. (2002). Information policy of web portals, E-Gov, and the public's prints. *The Journal of Academic Librarianship* 28(6), pp. 410-413.

Smith, T. R. (1996). The meta-information environment of digital libraries. *D-Lib Magazine* July/August

Smith, T. R., et al. (1996). A general framework for the meta-information and catalogs in digital libraries. First IEEE Metadata Conference. Siver Spring, Maryland: IEEE. Available at: http://www.computer.org/conferences/meta96/smith/ieee.html (Last accessed: 12/06/2006)

Steele, C. (1995). New Romances or pulp fiction? Do libraries and librarians have an Internet future? Available at: http://www.ukoln.ac.uk/services/papers/follett/steele/paper.html, (Last accessed: 12/06/2006)

Sturdy, D. (2001). Squirrels and nuts: metadata and knowledge management. *Business Information Review* 18(4), pp. 34-42.

Sugimoto, S., Gotou,S., Zhao,Y., Sakaguchi, T., Tabata, K. (1995). Enhancing usability of network-based library information system - experimental studies of a user interface of OPAC and of a collaboration tool for library services, in, *Digital Libraries '95: The Second Annual Conference on the Theory and Practice of Digital Libraries*. Austin, Texas, USA. Available at: http://www.csdl.tamu.edu/DL95/papers/sugimoto/sugimoto.html (Last accessed: 12/06/2006)

Sullenger, P. (1997). A Serials transaction log analysis. *Serials Review* 23(3), pp. 21-26.

Sumpter, R. M. (1994). White Paper on Data Management. Lawrence Livermore National Laboratory, Available at: http://www.llnl.gov/liv_comp/metadata/papers/whitepaper-draft.html (Last accessed: 12/06/2006)

Tenopir, C. (2003a). Information metrics and user studies. *Aslib Proceedings* 55(1/2), pp. 13-17.

Tenopir, C. (2003b). *Use and users of electronic library resources: an overview and analysis of recent research studies*. Report for the Council on Library and Information Resources. August 2003. Available at: http://www.clir.org/pubs/reports/pub120/pub120.pdf (Last accessed: 12/06/2006)

Thelwall, M. (2002). Subject gateway sites and search engine ranking. *Online Information Review* 26(2), pp. 101-107.

Thiele, H. (1998). The Dublin Core and the Warwich Framework: A review of the literature. *D-Lib Magazine*. Available at: www.dlib.org/dlib/january98/01thiele.html, (Last accessed: 27/11/2003).

Vellucci, S. L. (1997). Options for organizing electronic resources: The coexistence of metadata. *Bulletin of the American Society for Information Science* (Oct/Nov),

Vellucci, S. L. (1998). Metadata. Annual *Review of Information Science and Technology* 33 pp. 187-220

Vellucci, S. L. (2001). Music metadata and authority control in an international context. *Notes* 57(3), pp. 541-554.

W3C. (2000) *A Little History of the World Wide Web*. Available at: http://www.w3.org/History.html (Last accessed: 12/06/2006)

Waugh, A. (1998). Specifying metadata standards for metadata tool configuration. *Computer Networks and ISDN Systems* 30 pp. 23-32.

Weibel, S. (1995). Metadata: The foundations of resource description. *D-Lib Magazine*. Available at: www.dlib.org/dlib/July95/07weibel.html, (Last accessed: 27/11/2003).

Weibel, S. (1997). The metadata landscape: Conventions for semantics, syntax, and structure in the internet commons. *NFAIS*. Available at: http://www.nfais.org/Metadiversity1.asp?WhichMeta1=s1-weibel. (Last accessed: 12/06/2006)

Wessel, C. and B. Weib. (2001). META-LIB: Die Metadaten-Initiative deutscher Bibliotheken. *Bibliothek* 25, (3), pp.301-305.

Wiederhold, G. (1995). Digital libraries, value and productivity. *Communications of the ACM* 38(4), pp. 85-96.

Wilson, P. (1996). Interdisciplinary research and information overload. *Library Trends* 45(2), pp. 192-203

Witmer, D. F., Robert W. Colman and Sandra Lee Katzman. From paper-and-pencil to screen-and-keybord, *in, Jones, S (ed.) Doing Internet Research*. London: Sage, 1999.

Wolfram, D. and Xie, H. I. (2002). Traditional IR for web users: a context for general audience digital libraries. *Information Processing & Management* 38 pp. 627-648.

Woodward, J. (1996). Cataloguing and classifying information res.ources on the Internet. *Annual Review of Information Science and Technology* 31 pp. 189-220.

Xu, A. (1996). Accessing Information on the Internet: Feasibility Study of USMARC Format and AACR2, in, *Proceedings of the OCLC Internet Cataloging Colloquium*. Available at: http://www.oclc.org/oclc/man/colloq/xu.htm (Last accessed: 12/06/2006)

Xu, A. (1997). Metadata conversion and the library OPAC. *The Serials Librarian* 33(1-4),

Younger, J. A. (1997). Resources description in the digital age (cataloging and indexing; resource sharing in a changing environment). *Library Trends* 45(3), pp. 462-481.

Appendices

A1-Publications

1. Polydoratou, P., Pendelton, M. and Nicholas, D. (2006). Using web logs transactions to assess a metadata registry system's use: the case of the Environmental Data Registry and the System of Registries. Submitted for consideration to the 9th International Conference of Asian Digital Libraries (ICADL2006)
2. Polydoratou, P. (2006) Using web logs transactions to assess a metadata registry system's use: the case of MetaForm. OCLC Systems and Services, 22(1), pp. 67-79.
3. Polydoratou, P., Pendleton, M. and Nicholas, D. (2003). The System of Registries: An Evaluation of User Feedback...A Year Later. Accepted for presentation at the 6th International Conference of Asian Digital Libraries (ICADL2003); 8-11 December, Kuala Lumpur, 2003.
4. Polydoratou, Panayiota, Michael Pendleton and David Nicholas. (2002). The Use and Functionality of the Environmental Data Registry: An Evaluation of User Feedback in E.-P. Lim, S. Foo, C. Khoo, H. Chen, E. Fox, S. Urs, T. Costantino (Eds.): Proceedings of the 5th International Conference on Asian Digital Libraries (ICADL2002); 11-14 December, Singapore. Abstract available at: Springer Lecture Notes in Computer Science
5. Polydoratou, Panayiota and David Nicholas. (2001) Familiarity with and use of metadata formats and metadata registries amongst those working in diverse professional communities within the information sector. Aslib Proceedings. Vol. 53 (8), pp.309-324.
6. Polydoratou, Panayiota. (2001) A Discussion Lists Survey on Awareness of Metadata Formats and Metadata Registries among Diverse Scientific Communities in Pedro Isais (ed.) Proceedings of the First International Workshop on New Developments in Digital Libraries (NDDL 2001), 6-7 July, Setubal, Portugal.

A2- Questionnaires

SCHEMAS workshop and Discussion Lists Questionnaire

The purpose of this survey is to find out how familiar people are with metadata schemas and metadata registries. I am carrying out this research as part of my studies on metadata at City University, Department of Information Science. Contributing to this survey by taking the time to complete this questionnaire will be highly appreciated. The answers will be treated anonymously and your personal details, if filled in for further participation in the survey, will be used for the purpose of contacting you only.

I would be grateful if you complete and return this questionnaire no later than **Monday 27th of November 2000**, as an attached document to the following email address: p.polydoratou@city.ac.uk .

Please underline, use **bold** or highlight your answers.

Part 1- Personal details

1. Age range

   17-24
   25-34
   35-44
   45-54
   55-64
   65-+

2. Sex

   Male
   Female

3. Please indicate the domain (sector) that you represent:

   Industrial Sector
   Publishing Sector
   Audio - Visual Sector
   Educational Sector
   Academic Sector
   Research Sector
   Geographic Information Sector
   Other Sectors (please specify)
   .....

4. Please specify the post you hold
   .....

Part 2 - Metadata Schemas

1. Have you ever heard of metadata schemas before?

   Yes
   No
   If no, please go to question 7

2. Please indicate any of the terms you have come across before

   Metadata Element Sets
   Metadata Formats
   Metadata Schemas
   Metadata Standards
   Metadata Systems
   Metadata Catalogues
   Other (please specify)
   .....

3. Please indicate the metadata schemas that you have heard about

   Dublin Core Metadata Element Set
   IEEE LOM
   MARC
   IMS
   GILS
   Other (please specify)
   .....

4. Have you ever used a metadata schema?

   Yes
   If yes, please go to question 5
   No
   If no, please go to question 7

5. Please indicate the metadata schemas that you have used

   Dublin Core Metadata Element Set
   IEEE LOM
   MARC
   IMS
   GILS
   Other (please specify)
   .....

6. Please specify where you used that metadata schema

   To describe my own personal web page
   To describe my organisation's web page
   To describe the Library's resources
   For the needs of project that I am working
   (please indicate which project)
   .....

Other (please specify)

.....

7. Would you be interested in finding out more about metadata schemas in the future?
   Yes
   No

Part 3 - Metadata Registries

1. Have you heard of metadata registries before?

   Yes
   If yes, please go to question 2
   No
   If no, please go to question 6

2. Please indicate the metadata registries that you have heard about

   <indecs> metadata registry
   DESIRE metadata registry
   ROADS metadata registry
   German Metadata Registry
   Environmental Data Registry
   BSR (Basic Semantic Register)
   Australian Knowledgebase (Australia's health, community services and housing metadata registry
   Other (please specify)
   .....

3. Have you used a metadata registry before?

   Yes
   If yes, please go to question 4
   No
   If no, please go to question 6

4. Please indicate the metadata registries that you have used

   <indecs> metadata registry
   DESIRE metadata registry
   ROADS metadata registry
   German Metadata Registry
   Environmental Data Registry
   BSR (Basic Semantic Register)
   Australian Knowledgebase (Australia's health, community services and housing metadata registry
   Other (please specify)
   .....

5. Please indicate the reasons for using the metadata registry:

   For resource discovery-Locate relevant information
   Query processing

Data exchange
To describe text resources
To see definitions of elements
To find out about dictionaries structures
To find out about mapping between different metadata schemas
Other (please specify)
.....
.....
.....

6. Does your organisation have a requirement for using a metadata registry?

   Yes
   No
   I don't know

7. Would you be interested in finding out more about metadata registries?
   Yes
   No

If you are interested in participating further to this survey please provide your personal details so that I can get in touch with you. Thank you very much for taking the time to complete this questionnaire.

Name:

Address:

Email:

Tel.:

Fax:

MetaForm Questionnaire

The purpose of this survey is to find out what people think of the services MetaForm provides and evaluate their functionality and use. I am carrying out this research as part of my studies on metadata at City University, Department of Information Science. Contributing to this survey by taking the time to complete this questionnaire will be highly appreciated. The answers will be treated anonymously and your personal details, if filled in for further participation in the survey, will be used for the purpose of contacting you only.

Please underline, **bold** or highlight your answers.

Part 1 - Personal details

5. Age range

a) 17-24 ( )
b) 25-34 ( )
c) 35-44 ( )
d) 45-54 ( )
e) 55-64 ( )
f) 65-+ ( )

6. Gender

a) Male ( )
b) Female ( )

7. Please indicate the domain (sector) that you represent:

a) Academic Sector ( )
b) Audio - Visual Sector ( )
c) Cultural Heritage Sector (e.g., Libraries, Archives, Museums) ( )
d) Educational Sector ( )
e) Geographic Information Sector ( )
f) Industrial Sector ( )
g) Publishing Sector ( )
h) Research Sector ( )
i) Other Sectors ( )
(please specify)
.....
.....

8. Please specify the post you hold

a) Computer Scientist ( )
b) Consultant ( )
c) Information Scientist ( )
d) Researcher ( )
e) Other (please specify) ( )
.....

Part 2 - Use

1. **How long have you been using MetaForm?**
   a) 1 month ( )
   b) 3 months ( )
   c) 6 months ( )
   d) More than that (please specify) ( )
   .....

2. **How often do you use MetaForm?**
   a) On a weekly basis ( )
   b) On a monthly basis ( )
   c) Occasionally ( )
   d) Hardly ever ( )
   e) This is my first time ( )

3. **Does your organisation have a requirement of a metadata registry?**
   a) Yes (please describe)
   .....
   .....
   .....
   b) No, I am using MetaForm out of personal interest ( )
   c) I don't know, I am using MetaForm out of personal interest ( )

4. **Please, select those that apply.**
   The organisation that I represent/Personally, has/have been a:
   a) Contributor to the registry ( )
   Please, go to question 5
   b) User of the registry ( )
   Please, go to question 6
   c) Both ( )
   Please, go to question 5
   d) Other (please specify) ( )
   .....
   .....
   .....

5. **Please describe the content of information that you provide:**
   a) Data elements ( )
   b) Entire dataset, dB ( )
   c) Schemas ( )
   d) Other (please specify)
   .....
   .....
   .....

6. **Please indicate the reasons for using MetaForm:**
   a) Data exchange ( )
   b) For resource discovery-Locate relevant information ( )
   c) To describe text resources ( )
   d) To find out about mapping between different metadata formats ( )
   e) To see definitions of elements ( )
   f) To search for metadata formats ( )
   g) Other (please specify)

.....
.....
.....

7. **Please specify where you used information (metadata formats/application profiles, etc) obtained from MetaForm**
   a) To describe my organisation's web page
   b) To describe my own personal web page
   c) To describe the Library's resources
   d) For the needs of project that I am working
      (please indicate which project)
      .....
      .....
      .....
   e) Other (please specify)
      .....
      .....
      .....

8. **Have you used other metadata registries in the past?**
   a) Yes ( )
   b) No ( )
      Please, go to Part 3

9. **Please indicate those metadata registries that you have used in the past**
   a) **<indecs>** metadata registry ( )
   b) BSR (Basic Semantic Register) ( )
   c) DESIRE metadata registry ( )
   d) DCMI Open Metadata Registry ( )
   e) Environmental Data Registry ( )
   f) Knowledgebase ( )
   g) MetaForm ( )
   h) ROADS metadata registry ( )
      Other (please specify)
      .....
      .....
      .....

10. **From the registries that you have used, please select those that you found useful**
    a) **<indecs>** metadata registry ( )
    b) BSR (Basic Semantic Register) ( )
    c) DESIRE metadata registry ( )
    d) DCMI Open Metadata Registry ( )
    e) Environmental Data Registry ( )
    f) Knowledgebase ( )
    g) MetaForm ( )
    h) ROADS metadata registry ( )
    i) Other (please specify) ( )
       .....
       .....
       .....

Part 3 - Functionality

1. **Which of the following features of MetaForm have you used?**
   a) Crosswalks ( )
   b) Crosscuts ( )
   c) Mappings ( )

2. **On a rate of 1-5, where 1 is the least satisfactory level and 5 the most satisfactory level, how would you rate the crosswalks, crosscuts and mappings facilities of MetaForm?**

Crosswalks

1		2		3		4		5	

Crosscuts

1		2		3		4		5	

Mappings

1		2		3		4		5	

3. **How would you describe the structure of the metadata formats covered?**
   a) Satisfactory defined and described ( )
   b) Adequate defined and described ( )
   c) Not satisfactory defined and described ( )

4. **How would you describe the semantics of the metadata formats covered?**
   a) Satisfactory defined and described ( )
   b) Adequate defined and described ( )
   c) Not satisfactory defined and described ( )

5. **On a rate of 1-5, where 1 is the least satisfactory level and 5 the most satisfactory level, how would you rate the following facilities provided by MetaForm?**

	1	2	3	4	5
a) Coverage of information resources					
b) Coverage of metadata formats					
c) Linking between different metadata formats					
d) Mappings between different formats					
e) Relevance of the information that you retrieve					
f) Retrieving information in several formats					

6. **Would you want MetaForm to provide with additional facilities in the future?**
   a) Yes, I find the service satisfactory but I would welcome additional services ( )
   b) No, I find the service satisfactory as it is

Please, go to question 8 ( )

7. **Please select additional features that you would like MetaForm to add:**
   a) Advance linking facility ( )
   b) Advance mapping facilities ( )
   c) Automatic translation of metadata formats into other languages ( )
   d) Automatic validation of information ( )
   e) Larger coverage of information resources ( )
   f) Metadata conversion Tools ( )
   g) Metadata creation Tools ( )
   h) Metadata formats in several languages ( )
   i) Search facilities (index, browse, keyword searching) ( )
   j) Vocabulary facility ( )
   k) More metadata formats ( )
   l) Other (please specify)
   .....
   .....
   .....

8. **Additional comments that you would like to make**
   .....
   .....
   .....

If you are interested in participating further to this survey please provide your personal details so that I can get in touch with you. Thank you very much for taking the time to complete this questionnaire.

Name:

Address:

Email:

Tel.:

Fax:

Panayiota Polydoratou
Research Student
City University
Department of Information Science
Northampton Square
EC1V OHB
London Room A420a

email: P.Polydoratou@city.ac.uk
tel.: 0044 020 7477 8000 ext. 3905
fax: 0044 020 7477 8574

SoR Questionnaire

Dear Sir/Madame,
the purpose of this questionnaire survey is find out what people think of the System of Registries, attempt an initial evaluation of some of the functions provided and ask you to contribute with your input to what are the future expectations regarding registry systems. This questionnaire is part of my Ph. D. studies and your answers will be treated anonymously and your personal details, if filled in for further participation in the survey, will be used for the purpose of contacting you only.

In some cases more than one answer is possible (rectangular boxes).

The Environmental Protection Agency has not asked me to conduct this research on their behalf and this questionnaire does not constitute an expression of their views or interests. They agreed only to distribute this questionnaire and allow me to collect valuable data for my research. You are not obliged to fill in this questionnaire but I would be grateful if you did.

Part 1

1. **Would you identify yourself as:**

- Data Standards Developer ( )
- Data Standards Implementer ( )
- Exchange Network Participant ( )
- Registry Developer ( )
- Someone interested in Substance Identification ( )
- Systems Developer ( )

Other (please specify)
.....
.....
.....

2. **Please indicate the SoR components that you have used.**

- EDR (Environmental Data Registry) ( )
- EIMS (Environmental Information Management System) ( )
- EMG (Environmental Metadata Gateway) ( )
- FRS (Facility Registry System) ( )
- IRRS (Information Resource Registry System) ( )
- SoR quick search ( )
- SRS (Substance Registry System) ( )
- TRS (Terminology Registry System) ( )
- None, it's my first time using SoR ( )

3. **When accessing the SoR, what types of information do you usually search for?**

- Business rules documents ( )
- Code sets ( )
- Data elements ( )
- Data standards ( )
- Facilities ( )
- Organisations ( )
- Regulations ( )
- Resources, general (IRRS material) ( )
- Substances (chemical/biological) ( )
- Environmental terminology ( )

- XML tags ( )
- Other, please specify ( )

.....
.....
.....

4. **Please select from the following those that apply in your case:**
   - I have searched for information on SoR ( )
   - I have downloaded information that I have found on SoR ( )
   - I have applied download information from SoR to my work ( )
   - I have used the service to register an information resource (e.g., data sets) ( )
   - None of the above, I am here today to find out more on SoR ( )

5. **Please specify how you have used the information obtained from the SoR:**
   - To keep up to date with information resources within EPA ( )
   - To select and/or review information resources (e.g., data sets, databases, code sets, etc) that my office must comply with ( )
   - In my work with data standards (search for, review, amend) ( )
   - I wanted to download multiple data standards information at once (COMPARE TOOL) ( )
   - For other reasons, please specify ( )

.....
.....
.....
.....

Part 2

1. **Please indicate which features of the SoR you consider the strongest**

2. **Please indicate which features of the SoR need improvement and how**

3. **Please indicate how the SoR can support or improve your work**

Thank you for your time and your input!

If you are interested in participating further to this survey please provide with your email address or other best contact so I can get in touch with you. Thank you very much for taking the time to complete this questionnaire.

Name:

Email address or other best contact details:

Panayiota Polydoratou
Research Student
City University
Department of Information Science
Northampton Square
EC1V OHB
London
email: P.Polydoratou@city.ac.uk
tel.: 0044 (0)20 7040 5060 ext. 3905
fax: 0044 020 7477 8574

A3a – Discussion Lists

- 11179 Metadata Registries Coalition Discussion list. A United States of America based list which its members were people interested in the implementation of ISO/IEC 11179 metadata registries. The list could be found at: http://hmrha.hirs.osd.mil/mrc/. It was formed in 1998 and it aimed to *"provide a forum and source of practical experience for mutual cooperation among organizations that are introducing metadata registries into their information systems asset base in order to be able to manage the semantics of the data elements in their databases."* Members included people from diverse communities, mainly from US federal agencies, with an interest in ISO/IEC 11179 standard implementation approaches, developments and support. The list moderator was contacted regarding the number of subscribers to the list but a response failed to reach the researcher. It is believed that at the time of the survey the number of subscribers were around 30 people.
- AG-Metadaten@mail.sub.uni-goettingen.de . A discussion list for people at SUB Göttingen working with metadata. There was no information to the researcher's disposal regarding the subscribers and establishment of this list.
- Diglib. A European Union based list that was hosted by the International Federation for Library Association. Discussion on this list covered issues and technology pertaining to digital libraries research. The list can be found at: http://www.ifla.org/II/lists/diglib.htm (last accessed 27/11/2005). It was formed in 1995 and it is *"...a mailing list is for librarians, information scientists, and other information professionals to share information about the many issues and technologies pertaining to the creation of "digital libraries"*. Members included both individuals and organisations *"...from around the world who [were] creating or providing electronic access to digital collections to participate in knowledge sharing about current developments in digital library research."* The list moderator was contacted regarding the number of subscribers to the list but a response failed to reach the researcher.
- EULER@zblmath.fiz-karlsruhe.de There was no information to the researcher's disposal regarding the subscribers and establishment of this list.
- EULER-Consortium@zblmath.fiz-karlsruhe.de . A discussion list for people working with mathematical subject gateways. There was no information to the researcher's disposal regarding the subscribers and establishment of this list.
- Interoperability. A United Kingdom JISCMAIL list that hosted discussion about metadata and interoperability. The list can be found at: http://www.jiscmail.ac.uk/lists/interoperability.html (last accessed 27/11/2005). It was formed in January 1999 and aimed to address issues such as*"...metadata, distributed library systems and public library networking. Interoperability Focus also has a*

special interest in moving beyond the library sphere, specifically encompassing museums, archives, and other aspects of the cultural heritage, as well as Government and community information." Members included *"Projects and individuals with experiences to share, working implementations to show, or core issues to resolve which might usefully be addressed by Interoperability Focus...."* The interoperability list counted 385 members on the 14/11/2000.

- Lis-elib). A United Kingdom JISCMAIL list that hosted discussion about metadata and interoperability. The list can be found at: http://www.jiscmail.ac.uk/lists/lis-elib.html (last accessed 27/11/2005). It was formed in 1996 and aimed to address issues *"on the JISC Electronic Libraries Programme (eLib) and its projects. It also carries more general discussion of the changing IT environment in which libraries operate and requests for help and advice on specific topics related to eLib Programme interests."* The list moderator was contacted regarding the number of subscribers to the list but a response failed to reach the researcher.
- mb-net@www.SUB.Uni-goettingen.de (Math-Bib-Net). There was no information to the researcher's disposal regarding the subscribers and establishment of this list. In 2004, MetaForm staffs were contacted for a follow up to this initial survey, this list was no longer operational.
- meta-bib@ddb.de . A discussion list for people in Germany working with metadata. There was no information to the researcher's disposal regarding the subscribers and establishment of this list. In 2004, MetaForm staffs were contacted for a follow up to this initial survey; this list was no longer operational.
- reynard@nic.surfnet.nl . A discussion list addressing issues in European subject gateways. There was no information to the researcher's disposal regarding the subscribers and establishment of this list.
- UK-Meg. A United Kingdom JISCMAIL list that hosted discussion about metadata for education. The list can be found at: http://www.jiscmail.ac.uk/lists/uk-meg.html (last accessed 27/11/2005). Formed in June 2000, it aimed to *"...support the work of the Metadata for Education Group (MEG), which seeks common approaches to the description and exchange of educational content across all levels of the UK's educational system..."* Members included *"...a number of the current players in this field, drawn from primary, secondary, tertiary and continuing education, as well as from relevant standards initiatives and the museum and library sectors, which have valuable content to offer."* The UK-Meg list counted 90 members on the 14/11/2000.
- vlib@hub26.tib.uni-hannover.de . A discussion list addressing issues in German virtual libraries. There was no information to the researcher's disposal regarding the subscribers and establishment of this list.

A3b – Response by discussion list

Discussion Lists	Response	%
11179 Metadata Registry Coalition Forum	6	9.1
Diglib (IFLA)	21	31.8
Interoperability (UKOLN)	8	12.1
SCHEMAS	13	19.7
UK-meg (UKOLN)	7	10.6
No list	11	16.7
Total	66	100

Table 32: Appendix 3b - SCHEMAS workshop and Discussion Lists surveys – Response by discussion list