Interoperability for digital repositories: towards a policy and quality framework

Giuseppina Vullo[1], Perla Innocenti[2], Seamus Ross[3]

[1]Humanities Advanced Technology & Information Institute (HATII), University of Glasgow; [2]Faculty of Information, University of Toronto

{g.vullo, p.innocenti}@hatii.arts.gla.ac.uk; seamus.ross@utoronto.ca

Abstract

Interoperability is a property referring to the ability of diverse systems and organisations to work together. Today interoperability is considered a key-step to move from isolated digital repositories towards a common information space that allow users to browse through different resources within a single integrated environment. In this conference we describe the multi-level challenges that digital repositories face towards policy and quality interoperability, presenting the approaches and the *interim* outcomes of the Policy and Quality Working Groups within the EU-funded project DL.org (http://www.dlorg.eu/).

1. Introduction

At the heart of digital libraries lays a digital repository. As outlined in the DELOS Digital Library Manifesto (Candela et al. 2008), the digital library universe is a complex framework in which at least three types of conceptually different "systems" can be identified, namely, digital libraries (DLs), digital library systems (DLSs) and digital library management systems (DLMSs). Architecture, personalization, quality, policy and usability are essential to the design and deployment of digital libraries.

Digital libraries represent the confluence of many interdisciplinary fields, from data management, information retrieval, library sciences, document management to web services, information systems, image processing, artificial intelligence, human-computer interaction, and digital curation. Their multi-faceted nature has led researchers to offer a variety of definitions as to what a digital library is, often reflecting on different disciplinary perspectives (Borgman, 2000; Fox, Akscyn, Furuta, & Leggett, 1995; Fox & Marchionini, 1998; Ioannidis Y., 2001; Ioannidis Y. , 2005; Ioannidis, et al., 2005; Lagoze C. , 2010).

As (Gonçalves, Fox, Watson, & Kipp, 204) have explained the lack of unambiguous clarity on the boundaries of the term digital library arise because they are essentially complex multi-

dimensional applications. Ross (Ross S., 2003) pinpointed those aspects by characterizing a digital library as "*the infrastructure, policies and procedures, and organisational, political and economic mechanisms necessary to enable access to and preservation of digital content*" (p. 5).

Among the current crop of digital libraries, there is a variety in character and type of content, with some being homogeneous collections on particular topics or media whereas others have a heterogeneous character (Ross S. , 2003). The lack of agreement on the best design of digital library systems reflects, in part, a lack of agreement on the nature, functionality, and architecture of such applications. The DELOS Digital Library Manifesto (Candela, et al., 2006) and the DELOS Digital Library Reference Model (Candela, et al., 2008) aimed to address these *lacunae*. Starting with the DELOS Digital Library Reference Model as its conceptual framework, the EU funded project "DL.org Digital Library Interoperability, Best Practices and Modeling Foundations" investigates interoperability issues in the context of digital libraries as part of larger ecosystems. The DL.org project addresses digital library interoperability issues from the perspective of six core constituent parts characterizing the digital library according to the Reference Model: Content, User, Functionality, Policy, Quality, and Architecture.

So far there has been a lack of reference models against which to assess their designs. The DELOS Digital Library Reference Model (Candela et al., 2008) aimed to address these lacunae. Starting with the DELOS Reference Model as its conceptual framework, the DL.org project (http://www.dlorg.eu/) investigates interoperability issues in the context of digital libraries and digital repositories. DL.org addresses digital library interoperability issues from the DELOS Digital Library Reference Model six core domains (Architecture, Content, Functionality, Policy, Quality and User) perspectives, through the creation and the operation of six dedicated Working Groups, composed by international experts on each domain.

This paper focuses on the research hypothesis and *ad interim* outcomes on policy and quality interoperability developed within the DL.org Policy and Quality Working Groups upon their liaising with real life current international digital libraries, repositories, archives and initiatives.

2. Policy and quality interoperability within digital repositories: a multi-level approach

The DL.org Policy (https://workinggroups.wiki.dlorg.eu/index.php/Policy_Working_Group) and Quality (https://workinggroups.wiki.dlorg.eu/index.php/Quality_Working_Group) Working Groups are gathering both the expertise of international experts and the experiences within

current digital libraries and repositories in order to investigate policy and quality interoperability issues between distributed environments that give access to large-scale digital collections.

Interoperability is among the most critical issues to be faced when building systems as "collections" of independently developed constituents (systems on its own) that should co-operate and rely on each other to accomplish larger tasks. There is not yet a full interoperability solution or approach that is sufficient to serve the overall needs of digital library organisations and digital library systems. In fact, there is no yet single definition of interoperability which is accepted neither by the research and professional communities. The IEEE definition considers interoperability as "the ability of two or more systems or components to exchange information and to use the information that has been exchanged" (IEEE, 1991); the ISO/IEC 2382-2001 Information Technology Vocabulary, Fundamental Terms defines interoperability as "the capability to communicate, execute programs, or transfer data among various functional units in a manner that requires minimal knowledge of the unique characteristics of those units" (ISO, 2001).

In order to address the interoperability challenge exhaustively, the DL.org project is adopting a multi-level approach, along the classification of the European Interoperability Framework for eGovernment services (IDABC, 2004):

- Organisational interoperability: refers to cooperation between and within digital library organisations, business goals and process modeling. This is the most challenging level of interoperability, especially at a machine-readable and automation level.

- Semantic interoperability: refers to understanding the meaning of information in digital libraries.

- Technical interoperability: refers to interconnection, presentation and exchange of digital objects within digital library, accessibility and security issues.

At an organisational level, interoperability is a property referring to the ability of diverse systems and organisations to work together. Today organisational interoperability is considered a key step towards a common information space that allows users to browse through different resources within a single integrated environment (Fox, 1995; Miller, 2000; Borgman, 2000; Ross, 2008; Lagoze, 2010).

Organisational interoperability for digital libraries and digital repositories is a challenging and almost uncharted research area. The DL.org Policy and Quality Working Groups are currently investigating policy and quality interoperability focusing on the organisational level (Innocenti, Vullo, Ross, 2010; Vullo, Innocenti, Ross, 2010). The rationale behind this is that no technical interoperability is possible in the fields of policy and quality without taking into

3

account the over-arching policies established by the organisation which is behind a digital library or a digital repository.

Digital repositories are the newest and most dynamic example of information systems: they are evolving rapidly as technologies develop and as the ways in which researchers and learners – and administrators – accommodate to the digital age and its opportunities (Weenink, Waaijers, Godtsenhoven, 2007). Much has been learned already about how best to develop successful repositories through the intense collaborative effort of the professional community. In this context, the policy and quality interoperability organisational challenges that are under the DL.org Policy and Quality Working Groups' investigation are key-factors towards the cooperation and the development of successful collaborative services and infrastructures between digital repositories.

3. Towards Policy Interoperability

Digital libraries (together with digital repositories and data centres) represent the confluence of vision, mandate and the imagined possibility of content and services constructed around the opportunity of use. Underpinning every digital library is a policy framework. It is the policy framework that makes them viable - without a policy framework a digital library is little more than a container for content. Also the mechanisms for structuring the content within a traditional library building as container (e.g. deciding what will be on what shelves where) are based upon policy. Policy governs how a digital library is instantiated and run.

The policy domain is therefore a meta-domain which is situated both outside the digital library and any technologies used to deliver it, and with in the digital library itself. That is, policy exists as an intellectual construct, that is deployed to frame the construction the digital library and its external relationships, and then these and other more operational policies are represented in the functional elements of the digital library. Policy permeates the digital library from conceptualisation through to operation and needs to be so represented at these various levels.

The DL.org Policy Working Group (https://workinggroups.wiki.dlorg.eu/index.php/Policy_Working_Group) is addressing this almost unexplored territory of digital library policy at a holistic organisational rather than only technical level (Innocenti, Vullo, Ross, 2010; Vullo, Innocenti, Ross, 2010). The objective of the Policy Working Group is to define a Policy Interoperability Framework that will be included both in the forthcoming new enhanced version of the DELOS Digital Library Reference Model being produced within DL.org, and in the DL.org Digital Library Coobook.

The following areas were identified as particularly relevant for the goals of DL.org: policy classification (focusing on the Policy by scope section of the DELOS Reference Model); manual vs. automated policies (and in particular how to encode those policies for machine discovery, and which languages can be used to represent policies and make them functional, with particular attention to semantic web technologies); policy management (in particular how policies are appraised and enforced); policies evolution over time; and interconnectedness of policy and quality.

For example, there is a lack of policy formalisation and representation methods covering each one of the organisational, semantic and technical interoperability levels in current digital libraries. So far formally-encoded policies haven't been implemented in actual digital libraries - even the ones that interoperate with iRODS - , and there is yet no standard policy language for the Web. Another overlooked issue is the handling of policy drift over time, and of real-time vs. asynchronous policy interoperability.

The DL.org Policy Working Group defined Policy Interoperability as Business Level Interoperability, which allows comparing values and goals of organisations in order to 'make business' with them. This kind of interoperability takes places at a high (organisational) level, and it is then instantiated at process level -whether those processes are being handled by human or machine.

For expressing policy in a machine-encoded way at the organisational, semantic and technical interoperability levels, we are looking at the policy categorization of the MIT PLEDGE project MIT (http://pledge.mit.edu/index.php/Main_Page), which focused on the determination of a set of policies that affect operational digital preservation archives, with the goal of developing standardized means of recording and enforcing them using rules engines (Smith, Moore, 2006). In general, it seems that it's too early to expect formally-encoded digital library policies in actual digital libraries and there are no formal standard policy languages for the Web as yet, although there are ongoing efforts to map research languages like AIR (Kagal, 2009) to the new W3C recommendation standards for the Rules Interchange Format (http://www.w3.org/2005/rules/wiki/RIF_Working_Group). So we are analyzing the potential of the following: AIR Policy Language iRODS rules, SWRL, Turtle RDF Triples, REWERSE Policy Language, OWL, KAoS, WSPF-WS, WSPF, WSPL, XACML, Rei.

Policy user scenarios based upon real-life case studies are being produced to support investigations in this field and the collection and definition of best practices for developing a Policy Interoperability Framework in the digital library domain.

The Policy Working Group team has also been surveying policy interoperability examples from a selected representative sample of about twenty cross-domain international large/medium scale public and commercial digital libraries, digital repositories, digital archives, and federated digital library services. This study aimed to gain insight into these areas to underpin other aspects of interoperability (such as those around metadata and network protocols). The survey asked about written and publicly available policies within current institutions, as they may affect the development of digital library/digital repository within them. It asked in particular about:

1. Any policies, strategies, frameworks, programs, plans, or statements that have been prepared at current institutions to guide how it develops and exploits aspects of its digital library/digital repository's information management.

2. How these policies, strategies, frameworks, programs, plans, or statements affects or are affected by interoperability. There has been some work already looking at policies in this domain – who has them, and what issues they cover. But there has been little, if any, work on how these policies support interoperability between collections, and how the policies themselves can interoperate or be reused and assessed, possibly in a machine-encoded way.

The results of the survey are going to be disseminated in further scientific publications and included in the enhanced version of the DELOS Digital Library Reference Model and in the DL.org Digital Library Cookbook.

4. Towards Quality Interoperability

A small fraction of works on digital libraries and digital repositories is dedicated to quality: they often focus on the establishment, adoption and measurement of quality requirements and indicators. However, the manner that these quality indicators can interoperate is still under-researched.

The investigation of the DL.org Quality Working Group ((https://workinggroups.wiki.dlorg.eu/index.php/Quality_Working_Group) aims to gain insight into these areas, underpinning work on other aspects of interoperability addressed by DL.org (Content, Architecture, Policy, Quality, Functionality, User), according to the DELOS Reference Model (Candela et al., 2008).

As quality is still a low-prioritized aspect of digital libraries and digital repositories, the Quality Working Group is investigating both the research areas and the real-world cases in which quality issues have been developed. The research investigation is taking into account the

definitions of quality, i.e. what and how to measure, the digital library theoretical models, and the advances on data quality, digital libraries evaluation and quality parameters.

The DL.org Quality Working Group defined quality interoperability as "the possibility for digital libraries [*and therefore for digital repositories*] to share a common quality framework". The Quality Working Group also agreed to adopt the DELOS Reference Model as a conceptual framework, addressing to the project partners to consider an additional level of its "Three-tier Framework" (Candela et al., 2008, p. 17) that is termed "Organisation", over-arching the existing levels of Digital Library, Digital Library System and Digital Library Management System. The underlying rationale of this extension is that the concept Digital Library itself might not be sufficient to address all interoperability issues that are under investigation in DL.org. It is considered likely that there is an organisation beyond a digital library or a digital repository which defines the policy of the overall system in which a digital library or a digital repository is operating.

Upon the agreement that the core business of digital libraries resides in the management of their collections, the Quality Working Group identified a quality pattern that is thought to be most characteristic for digital libraries and that shall help digital libraries to interoperate in the quality domain. This pattern focuses on content and policy parameters, which have been considered crucial to allow interoperability, which has been called the "Quality Core Model" (Fig. 1). The Quality Working Group investigated the Quality Core Model parameters' definitions and relationships, and produced related examples and user scenarios.

Fig. 1 - The DL.org Quality Core Model[1]

Aiming to test the feasibility of the "Quality Core Model" and to produce a Quality Interoperability Checklist addressed to digital library and repository managers, the Quality

---

[1] The mind map in Figure 1 was prepared by Giuseppina Vullo (28 July 2009) to capture the discussions DL.org Quality Working Group members had at a meeting in Tirrenia (Italy) on 2-3 July 2009. Participants contributing to defining the model were Genevieve Clavel-Merrin, Nicola Ferro, Sarah Higgins, Wolfram Horstmann, Sarantos Kapidakis, Dirk Roorda, Seamus Ross and Giuseppina Vullo.

7

Vullo, Innocenti, Ross, Interoperability for digital repositories: towards a policy and quality framework
OR2010 – 5th International Conference on Open Repositories
Madrid, Spain, July 6th- 9th 2010

Working Group is currently working to test its feasibility. A survey on existing approaches and common practices on quality aspects has been conducting within a selection of targeted digital libraries and digital repositories. The results will help to understand how the professional community is facing quality interoperability issues and to identify best practices in this area.

The Quality Working Group successfully completed the survey pilot in May 2010 and is currently working on its official version. The survey pilot involved some representative international digital libraries and digital repositories. The Quality Working Group organized the survey pilot in order to gather information on quality requirements regarding the different quality facets and asked specific questions on quality interoperability, focusing on the Quality Core Model parameters. This first collection of inputs and feedbacks showed the importance of quality guidelines and certifications coming from the digital repositories world, such as the DRIVER guidelines (DRIVER, 2008) and the DINI certification (http://www.dini.de/).

One of the main results of the survey pilot is that "quality" is considered as a subjective and dynamic entity, and that a common understanding even on the basic terms is needed. In order to satisfy this need, the Quality Working Group has been involved in the preparation of a glossary to be integrated with the survey, improving in parallel the questions structure.

The Quality Interoperability Survey results will be collected as best practices to be included in the Digital Library Technology & Methodology Cookbook - one of the core cross-domains outputs of DL.org (http://www.dlorg.eu/index.php/outcomes) - with a portfolio of best practices, outlining patterns and solutions to common issues faced when developing large-scale interoperable digital library systems. The Cookbook will also provide guidelines for selecting appropriate interoperability standards and approaches when implementing interoperable federated systems.

The Quality Working Group contribution will consider the following quality topics: Guidelines, Certifications, Checklists, Validators, Best Practices, Ontologies for Quality of Service, Web accessibility issues, Harvestability via OAI-PMH. The following case studies will be considered: Document repositories, Research data archives, Digital Preservation systems.

The Cookbook will also include a Quality Interoperability Checklist, a light-weight document with practical recommendations based on the Quality Core Model parameters and the Quality Interoperability Survey results.

5. Ongoing results and next steps

The evaluation of existing approaches and best practices in relation to policy and quality interoperability within digital libraries and digital repositories, and the investigation within the DL.org Policy and Quality Working Groups allowed reaching the following preliminary results:

1. *Expansion of the broadness of the DELOS Digital Library Reference Model for policy.* The Policy Working Group agreed that the policy domain is broader than how it is currently represented in the Reference Model and its description needs to be updated.

2. *Highlight of the organisational interoperability level for both policy and quality.* The current DELOS Digital Library Reference Model heavily focused on system architecture and doesn't clearly address the issue of the context of digital library systems, which is crucial for both policy and quality issues.

3. *Possible repositioning of Policy within the DELOS Digital Library Reference Model.* A proposal for the repositioning of Policy in the overall Reference Model, as a metalayer above and beyond the digital library, and also a operational layer in the digital library, has been produced. The rationale for this, as previously argued, is that in real life the policy domain is a meta-domain which is situated both outside the digital library and any technologies used to deliver it and with in the digital library.

4. *Lack of formalization for expressing organisational, semantic and technical interoperability within digital libraries.* The PLEDGE project provided evidence that very few current digital libraries have formal policies in place. They do when there are business concerns (e.g. the commercial digital libraries) and they do for example for access control, but for many types of policies there is very little written down and none of it is machine-readable. Formalized digital library quality frameworks are equally rare and normally focus only on specific facets of the library (such as data or standards); quality changes over time and is still a low priority issue within digital libraries. As for quality, the spread of quality indicators to evaluate information systems needs indeed an upper framework to foster cooperation and exchange of quality data, and quality is a dynamic and complex dimension to assess within digital information systems.

5. *Looking at real-life current digital libraries.* While investigating the state of the art on interoperability issues, both the Policy and Quality Working Groups have adopted a bottom up approach and are producing user scenarios and conducting surveys with real life digital libraries, repositories, archives and initiatives. The fundamental

assumption behind this is that digital libraries stakeholders are first of all librarians, and decisions makers.

In the next months both the DL.org Policy and the Quality Working Groups will finalise their contribution to the DL.org scientific research on policy and quality interoperability, including the survey results, the Policy Interoperability Framework, the Quality Core Model and Quality Interoperability Checklist in the enhanced version of the DELOS Digital Library Reference Model and in the DL.org Digital Library Technology & Methodology Cookbook. By presenting our activities and research outcomes on policy and quality, we expect to contribute to the work of other digital library interoperability efforts and offer the community an example of a comprehensive approach to policy and quality interoperability.

Acknowledgements

Bibliography

Borgman, C. L. *From Gutenberg to the global information infrastructure: access to information in the networked world*, Cambridge, MA: MIT Press, 2000.

Candela, L., et al. *The DELOS Digital Library Reference Model - Foundations for Digital Libraries, Version 0.98*, DELOS Network of Excellence - Project no. 507618, 2008. Web. Feb 2010, <http://www.delos.info/files/pdf/ReferenceModel/DELOSDLReferenceModel_0.98.pdf>.

Candela, L., Castelli, D., Ioannidis, Y., Koutrika, G., Pagano, P., Ross, S., et al. (2006). *The Digital Library Manifesto.* DELOS.

Fox, E. A., et al. "Digital Libraries." Commun ACM, 38.4, 23–28 (1995).

DRIVER, *DRIVER Guidelines 2.0. Guidelines for content providers - Exposing textual resources with OAI-PMH*, [November 2008] <http://www.driver-support.eu/documents/DRIVER_Guidelines_v2_Final_2008-11-13.pdf>.

Fox, E., & Marchionini, G. (1998). Toward a Worldwide Digital Library. *Communications of the ACM , 41* (4), 29-32.

Gonçalves, M.A., et al., "A quality model for digital libraries", Information processing and management, 43, 1416-1437 (2007).

Gonçalves, M.A., et al., "Streams, structures, spaces, scenarios, societies (5s): A formal model for digital libraries", ACM Trans. Inf. Syst., 22, 270-312 (2004).

Hofman P., et al., *Selection Criteria for Quality Controlled Information Gateways*, Version 1.1 (DESIRE, 1997) pg. 33, <http://www.ukoln.ac.uk/metadata/desire/quality/>.

IDABC. (2004). *European Interoperability Framework for pan-European eGovernment Services. Luxembourg*.

Innocenti, P., Vullo, G., Ross, S., Towards a Digital Library Policy and Quality Interoperability Framework: the DL.org Project, *NRIN*, 15, 1–25 (2010).

Institute of Electrical and Electronics Engineers, *IEEE Standard Computer Dictionary. A Compilation of IEEE Standard Computer Glossaries : 610*, New York : Institute of Electrical and Electronics Engineers, 1991, p. 114, <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=182763&isnumber=4683>.

ISO 8402:1994. Quality management and quality assurance – Vocabulary.

ISO/IEC 2382:2001. Interoperability - Information Technology Vocabulary, Fundamental Terms.

Lagoze, C., *Lost Identity: The Assimilation of Digital Libraries into the Web*, PhD thesis, Information Science, Cornell University, 2010, <http://www.cs.cornell.edu/lagoze/dissertation/dissertation.html>.

Ioannidis, Y. (2005). *Digital libraries at a crossroads. International Journal of Digital Libraries , 5* (4), 255 –265.

Ioannidis, Y. (2001). *Digital Libraries: Future Directions for a European Research Programme.* San Cassiano, Italy: DELOS.

Lagoze, C., & Van de Sompel, H. (2001). *The open archives initiative: building a low-barrier interoperability framework. Proceedings of the 1st ACM/IEEE-CS Joint Conference on Digital Libraries, Roanoke, Virginia, United States, JCDL '01* (pp. 54–62). New York, NY: ACM.

Miller, P., *Interoperability: What it is and why should I want it.* Ariadne, 24, 2000, <http://www.ariadne.ac.uk/issue24/interoperability/intro.html>.

Ross, S. (2003). *Digital library development review.* National Library of New Zealand.

Ross, S. *Preservation of interoperability and interoperability of preservation.* Third Workshop on Foundations of Digital Libraries,ECDL2008, Aarhus, 2008, <http://www.delos.info/files/pdf/DLFoundations2008/10_RossDLFoundations08.pdf>.

Smith, M., and R. Moore."Digital Archive Policies and Trusted Digital Repositories." International Journal of Digital Curation 2.1: 92–101.

Weenink K., Leo Waaijers L., Godtsenhoven K. van, *A DRIVER's Guide to European Repositories : Five studies of important Digital Repository related issues and good Practices*, Amsterdam, Amsterdam University Press, 2007, <http://dare.uva.nl/document/93898>.

Vullo, G. Innocenti, P. Ross, S. *Towards Policy and Quality Interoperability: Challenges and Approaches for Digital Libraries*, IS&T Archiving 2010 Preservation Strategies and Imaging Technologies for Cultural Heritage Institutions and Memory Organisations, Conference Proceedings,

1-4 June 2010, The Hague, The Netherlands, Society for Imaging Science and Technology, 2010, pp. 33-38

Giuseppina VULLO

Dr. Giuseppina Vullo is Co-Principal Investigator in the EU-funded project Digital Library Interoperability, Best Practices and Modelling Foundations (DL.org). She's a researcher at the Humanities Advanced Technology Institute (HATII), at the University of Glasgow. Her research interests range from quality to contextualisation in digital libraries and enhancement of special collections within digital environments. She has been a DigitalPreservationEurope Exchange (DPEX) fellow at HATII in 2008, where she worked on digital collections assessment, applying Digital Repository Audit Method Based on Risk Assessment (DRAMBORA) and InterPARES 3 methodologies. She completed in 2009 a PhD in Library Science and has worked in university libraries and international institutes in Italy and Switzerland.

Perla INNOCENTI

Perla Innocenti is Research Associate at HATII, University of Glasgow and co-Principal Investigator in the EU-funded projects projects Sustaining Heritage Access through Multivalent ArchiviNg (SHAMAN) and Digital Library Interoperability, Best Practices and Modelling Foundations (DL.org, in which she is leading the Policy Working Group (https://workinggroups.wiki.dlorg.eu/index.php/Policy_Working_Group). Her research interests include digital preservation methodologies, preservation of media art, audit and risk assessment for digital repositories, digital libraries design and usage models, digitization methodologies. She was involved in repository design, audit research as part of DigitalPreservationEurope (DPE) and Digital Curation Center (DCC), co-ordinating activities and development for the DRAMBORA) Toolkit, usage models research within the EU-funded project Planets, and the investigation of the potential application of the DRAMBORA toolkit in digital libraries within the DELOS project.

Seamus ROSS

Prof. Seamus Ross is Dean of the Faculty of Information at the University of Toronto. Seamus Ross, was Professor of Humanities Informatics and Digital Curation, and founding Director of the Humanities Advanced Technology and Information Institute (HATII) at the University of Glasgow from 1997 through 2008. He was Associate Director of the Digital Curation Centre in the UK , 2004-8), Principal Director of DigitalPreservationEurope DPE and a partner in Preservation and Long-term Access through NETworked Services (Planets). He was a co-principal investigator in the DELOS Digital Libraries Network of Excellence (2002-8). He was Principal Director of ERPANET, a European Commission activity to enhance the preservation of cultural heritage and scientific digital objects. His research focuses on digital preservation including work on preservation, repository design, digital library design and services, ingest, and semantic metadata extraction.