DL.org: Coordination Action on Digital Library Interoperability, Best Practices and Modelling Foundations

Funded under the Seventh Framework Programme, ICT Programme – "Cultural Heritage and Technology Enhanced Learning"

Project Number: 231551

Deliverable Title: *D4.2 Second International Workshop Proceedings*

Submission Due Date: *30 November 2010*

Actual Submission Date: *04 February 2011*

Work Package: *WP4 DL.org Events*

Responsible Partner: *NKUA*

Deliverable Status: *Final*

Document Information

Project	
Project acronym:	DL.org
Project full title:	Coordination Action on Digital Library Interoperability, Best Practices & Modelling Foundations
Project start:	1 December 2008
Project duration:	27 months
Call:	ICT CALL 3, FP7-ICT-2007-3
Grant agreement no.:	231551

Document	
Deliverable number:	D4.2
Deliverable title:	Second International Workshop Proceedings
Editor(s):	Anna Nika, Katerina El Raheb, Eleni Toli NKUA
Author(s):	Anna Nika, Katerina El Raheb, Eleni Toli NKUA
Reviewer(s):	Costantino Thanos
Contributor(s):	
Participant(s):	NKUA; CNR; UG; TRUST-IT
Work package no.:	WP4
Work package title:	DL.org Events
Work package leader:	NKUA
Work package participants:	NKUA; CNR; UG; TRUST-IT
Est. Person-months:	2 months
Distribution:	Public
Nature:	Report
Version/Revision:	V0.3
Draft/Final	Final
Total number of pages: (including cover)	24
Keywords:	workshop, brainstorming, position statements, linked data

Disclaimer

This document contains information on the core activities, findings, and outcomes of the EC-funded project, DL.org, and in some instances, distinguished experts forming part of the project's Liaison Group, six Thematic Working Groups and External Advisory Board. The document may contain references to content in the DELOS Digital Library Reference Model, which is under copyright. Any references to content herein should clearly indicate the authors, source, organisation and date of publication.

This publication has been produced with the funding of the European Commission. The content of this publication is the sole responsibility of the DL.org consortium and cannot be considered to reflect the views of the European Commission.

European Commission
Information Society and Media

The European Union was established in accordance with the Treaty on European Union (Maastricht). There are currently 27 Member States of the Union. It is based on the European Communities and member states cooperation in the fields of Common Foreign and Security Policy and Justice and Home Affairs. The five main institutions of the European Union are the European Parliament, the Council of Ministers, the European Commission, the Court of Justice and the Court of Auditors. (http://europa.eu.int/)

DL.org is funded by the European Commission under the 7th Framework Programme (FP7).

Glossary of Terms

DL.org	Coordination Action on Digital Library Interoperability, Best Practices and Modelling Foundations.
ECDL	European Conference on Digital Libraries, annual conferences located in EU cities, proposed as the backdrop for DL.org Workshops. The 2nd Workshop is set within ECDL2010, September 6-10, Glasgow, Scotland, UK

Table of Contents

Summary

Attended by over 40 participants from the DL community, the 2nd DL.org Workshop brought into sharp relief Digital Library (DL) interoperability issues from multiple perspectives based on the 7 presentations that have been selected by the international program committee out of 12 submissions.

Alex Wade, the Director for Scholarly Communication at Microsoft Research, delivered a keynote on "An Industrial Perspective on Digital Library Interoperability", with particular reference to Microsoft's solutions and approaches to interoperability. There were seven additional presentations, the four were delivered during the first day of the Workshop, and the final three papers were presented during the second day of the Workshop on September 10, 2010.

The four papers presented during the first day of the workshop were (1) The gCube Interoperability Framework by Leonardo Candela, George Kakaletris, Pasquale Pagano, Giorgios Papanikos, and Fabio Simeoni, (2) Handling Repository-Related Interoperability Issues: the SONEX Workgroup by Peter Burnhill, Pablo de Castro, Jim Downing, Richard Jones, and Mogens Sandfær, (3) Epidemiology Experiment and Simulation Management through Schema-Based Digital Libraries by Jonathan Leidig, Edward A. Fox, Madhav Marathe, and Henning Mortveit, and (4) Interoperability Patterns in Digital Library Systems Federations by Paolo Manghi, Leonardo Candela, and Pasquale Pagano.

The final three papers presented on September 10, 2010, were (1) A Framework for Digital Library Function Description, Publication, and Discovery: A prerequisite for interoperable digital libraries by George Athanasopoulos, Katerina El Raheb, Edward Fox, George Kakaletris, Natalia Manola, Carlo Meghini, Andreas Rauber, and Dagobert Soergel, (2) Quality interoperability within digital libraries: the DL.org perspective by Giuseppina Vullo, Genevieve Clavel, Nicola Ferro, Sarah Higgins, René van Horik, Wolfram Horstmann, Sarantos Kapidakis, and Seamus Ross, and (3) Modeling User and Context in Digital Libraries: Interoperability Issues by Anna Nika, Tiziana Catarci, Akrivi Katifori, Georgia Koutrika, Natalia Manola, Andreas Nürnberger, and Manfred Thaller.

The final part of the Workshop opened the floor to participants offering a springboard for further discussions and exchanges of ideas on the relation between Linked Data and Digital Libraries. The Workshop concluded with participants volunteering their own position statements on DL interoperability challenges and DL.org's approach to this highly complex issue.

This Deliverable reports on the focus of the presentations and pursuing discussion points at the 2nd DL.org Workshop. Revised versions of the papers along with the papers presented during the "Very Large Digital Libraries" workshop, which was also organised within ECDL2010, will be published. Publishing companies have already been contacted and final arrangements are pending.

1. Deliverable Outline

The purpose of this Deliverable is to provide a comprehensive report of the 2nd DL.org Workshop held on 9 and 10 September in Glasgow, Scotland, within the 14th European Conference on Digital Libraries (ECDL2010). The deliverable opens with the Welcome Message from the Workshop Organisers: Donatella Castelli, CNR-ISTI, Italy; Yannis Ioannidis, University of Athens, Greece; Seamus Ross, University of Toronto, Canada. The Deliverable is then divided into three parts. The first part includes a summary of the invited talk as well as the description of the papers that were presented during the first day of the workshop. The second part includes the abstracts of the papers and the main discussion points of the second day of the workshop. The third part contains the Position Statements received from leading figures in the Digital Library space. Finally, Annex 1 provides the agenda of the workshop and Annex 2 the call for papers.

2. Welcome Message from the Workshop Organisers

The central theme of the 2nd DL.org Workshop is "Digital Library Interoperability". Interoperability is a multifaceted and very context-specific concept, encompassing different levels along a multi-dimensional spectrum ranging from Organisational to technological aspects. These multidimensional aspects make it hard to find generic and fully-comprehensive solutions, thus requiring an approach that investigates interoperability from multiple perspectives.

The workshop builds on the successful outcomes of the 1st DL.org Workshop held in September 2009 during the 13th European Conference on Research and Advanced Technology for Digital Libraries (ECDL2009), which raised many new questions related to interoperability scientific challenges and to its conceptualization. This year's workshop has sought to continue the path initiated during the 1st Workshop by soliciting and gathering research paper submissions that analyse the different aspects involved in achieving interoperability, from conceptualization at a high Organisational level to instantiation at process level, as well as to modelling techniques for representing and enabling interoperability between heterogeneous digital libraries, mediation approaches, methods, and supporting systems.

The research illustrated by the papers accepted at the workshop provides researchers, practitioners and digital library developers with novel solutions for improving digital library interoperability, thus, contributing an effective cross-exploitation of the resources managed by these systems.

All these papers have been peer reviewed by at least three reviewers. We would like to thank the members of the international program committee and the authors of the papers submitted for their contribution to the quality and success of the Workshop.

Donatella Castelli, Yannis Ioannidis, Seamus Ross, Worksop Organisers, 2nd DL.org Workshop

3. Workshop Program - Thursday, 9 September 2010

3.1 Session 1: Invited Talk

"Digital Library Interoperability: An Industrial Perspective" by Alex Wade, Director for Scholarly Communication, Microsoft External Research

In today's interconnected heterogeneous computing environments, and with the emergence of Web services and service-oriented architectures, the need for interoperable solutions to connect workflow across systems has become critical. Microsoft supports interoperability through a holistic approach, including design and development of interoperable products, industry collaboration, technology access, and support for industry standards. Microsoft designs its products to work with third-party software, as well as its own software. Interoperability drives innovation within a thriving IT industry, creating technologies that improve citizen services and government efficiency. Governments and businesses alike are looking to new technologies, like cloud services, to enable innovative offerings. When the IT community collaborates and develops new technologies that work together, new opportunities are created. Companies across the IT sector are working with partners, customers, and competitors to deliver the levels of interoperability customers want. Interoperability has been an emphasis from the early days of the Windows Azure software platform, Microsoft's cloud platform. As an example, Microsoft and IBM have both joined an open source project initiated by Zend Technologies to create basic interfaces that open source developers can use to work on top of any cloud platform. Microsoft has made a major commitment to document interoperability with Open XML. Open XML is a standards-based specification that enables documents to function across numerous systems and platforms. At the enterprise level, the Open XML SDK enables Open XML documents to be processed by large-scale document management systems, such as those used in the insurance or government sectors. Digital Library environments will follow similar trends to the commercial sector and will leverage computing and data storage in the cloud as scientists already experimenting with Amazon S3 and EC2 services. Jim Gray's key areas for action related to the future of Scholarly Communication and Libraries included the establishment of Digital Libraries that support other sciences, the funding of development of new authoring tools and publication models, and the

exploration of development of digital data libraries that contain scientific data (not just the metadata) and support integration with published literature. Issues and challenges of economically sustainable digital preservation include migration to standards that are likely to be supported in the future, and creation of emulators of hardware and simulators of software systems to enable old programmes to run and old data to be used.

Discussion

After the end of the keynote talk, Peter Burnhill opened the discussion by focusing on the issue of extended metadata. Alex Wade pointed out that the problem lies in getting the information out of the authors without being owners of this information. Microsoft is using a number of things such as entity extraction data mining and linguistics analysis tools in order to provide the authors with the tools that facilitate their work. Yannis Ioannidis stressed the importance of easily deposited workflow results to Digital Library repositories. He asked if Microsoft deals with the issue that might be more needed, that is, the use of a particular workflow and the specification of the sources of inputs or the software representing a specific workflow model. Alex Wade replied that the scenarios Microsoft manages include the things projects run on a daily basis and the information about the software is stored within the history of a job. Yannis Ioannidis mentioned that Microsoft puts the greatest emphasis on getting one tool to work with another in terms of software calls and in some cases the exchange of data. He stated that DL.org deals with other types of interoperability such as the interoperability of users or the interoperability of policies and asked if Microsoft pays attention to any of these types of interoperability. Alex Wade replied that one important thing for Microsoft is federated identity management with the use of web services.

3.2   Session 2: DL Interoperability Principles and Practice

3.2.1   The gCube Interoperability Framework

Authors: *Leonardo Candela, Pasquale Pagano – Institute of Information Science and Technologies, National Research Council of Italy; George Kakaletris, Giorgios Papanikos - University of Athens, Greece; Fabio Simeoni - University of Strathclyde, Scotland, UK*

Presented by: *Pasquale Pagano*

 Abstract: Interoperability is one of the most challenging issues in the design and interoperation of large-scale computing infrastructures for multidisciplinary research. Interoperability requires solutions that can "embrace" heterogeneity, that is, accommodate multiple incarnations of similar resources, as well as hide heterogeneity, that is, provide consumers with a

homogeneous view of the different resources. The authors look at the measures put in place by the D4Science infrastructure to address a range of interoperability issues.

Discussion

The main issues the discussion addressed focused on the automatic generation of workflows that are important to support the interaction between the systems of the gCube infrastructure. Pasquale Pagano reported that the standards used in web services are not enough to automatically generate workflows that combine services to achieve a goal. However, the authors of this paper are enriching the specification in order to extend it with additional information for publishing the profile. The profile is stored in a common place that all services of the infrastructure can access with a simple query. Based on this information, two examples of automatically generated workflows exist. The first example is the search engine: one specific query on a set of collections automatically identifies which are the operators that are involved to perform a query and it can include the transformation of metadata if they are needed to execute the query. The second example is based on the data transformation service: if there is an input format, e.g., "mytype" and the user wishes to generate another output format in a different "mytype" he has to submit this request to the system. The system will automatically check if there are certain services that can be combined to go from the input "mytype" to the output "mytype" and, if possible, it will combine several executions. If there are many transformation parts, the system will try to optimise which of the transformation parts have to be executed. In some cases, by using the standard definition of the web service the user can automatically combine and generate workflows and then he can execute these workflows in other cases. This is not possible for generic services and the user has to define manually the workflows that can be executed.

3.2.2   Handling Repository-Related Interoperability Issues: the SONEX Workgroup

Authors: *Peter Burnhill - EDINA National Data Centre, University of Edinburgh, Scotland, UK; Pablo de Castro - Carlos III University Madrid, Spain; Jim Downing - University of Cambridge, UK; Richard Jones - Symplectic Ltd, UK; Mogens Sandfær - Technical University of Denmark*

Presented by: *Peter Burnhill and Pablo de Castro*
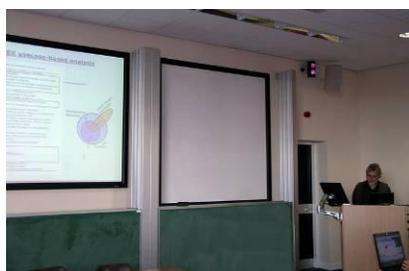
Abstract: The sharing of scholarly content through a network of Open Access repositories is becoming commonplace but there is still a need for systematic attention into ways to increase the rate of deposit into, and transfer of content across, the OA repository space. This is a report of a small international group supported by the Joint Information Steering Committee (JISC) with remit to describe, analyse, and make recommendations on deposit opportunities and use cases that might provide a framework for project activity geared to the ingest of

research papers and other scholarly works. The multi-authored, multi-institutional work is put forward as the default and nine use case actors are listed, as deposit agents, with four main use case scenarios. There is also some comment and pointers to projects in Europe which address some of these use case scenarios.

Discussion

Donatella Castelli opened the discussion which revolved around the issue of representation of policies and policy interoperability that the SONEX workgroup has faced. Peter Burnhill reported that SONEX as a group didn't focus on policy interoperability, but the projects that they collaborate with are doing some things for this issue. He presented the example of the RoMEO database that contains publishers' general policies on self-archiving of journal and conference articles. Each entry provides a summary of the publisher's policy, including what version of an article can be deposited, where it can be deposited, and any conditions that are attached to that deposit. By using the author's request button, someone can make a request to the author and the author can directly supply the article without concerns about publisher's policies. Giuseppina Vullo asked whether the SONEX workgroup focuses on technical, semantic, and organisational interoperability. Peter Burnhill and Pablo de Castro replied that they don't focus on the different kinds of interoperability levels but they investigate the purpose of interworking.

3.2.3 Epidemiology Experiment and Simulation Management through Schema-Based Digital Libraries

Authors: *Jonathan Leidig, Madhav Marathe and Henning Mortveit - Network Dynamics and Simulation Science Laboratory (NDSSL), Virginia Tech, U.S.; Edward A. Fox - Department of Computer Science, Virginia Tech, U.S.*

Presented by: *Jonathan Leidig*

Abstract: Digital Libraries that maintain inputs, result datasets, and the documentation of analyses would be beneficial for users of simulation and experimentation systems. A schema describing the simulation model may be used by a generated digital library for semantic integration with a simulation infrastructure. Based on a provided schema, the digital library may generate a generic user interface layout, database schema, and processes for launching new simulations. User roles are supported through a generic interface capable of performing a variety of user tasks. Formal descriptions of simulation content enable provenance investigations and domain-specific search. Simulation infrastructures with digital libraries are provided interoperability through managed datasets and interconnected software components. The authors present a generic, component-based simulation

supporting a digital library that is customised through provided schemas and extensible through the addition of components.

Discussion

George Athanasopoulos opened the discussion by focusing on the use of the Google platform as well as on the type of information that is used for the description of services. Jonathan Leidig pointed out that they use the Google web toolkit that lays out different widgets and components to allow a generic schema to be displayed to the users and also takes care of harvesting things that the user types in different labels and components and bring them back to the Digital Library. He continued by mentioning that, at the minimum, their schema has to list all the parameters so that the Digital Library can harvest all the required information from the interface. In addition, that schema could contain a structure as expected from the simulation system, so that the user can expect the production of a generic set of analysis scripts. The user can perform a standard set of analysis scripts, any time that the simulation model runs to conduct a new study. Donatella Castelli asked if there are other similar attempts of generic simulations. Jonathan Leidig mentioned that they differ from other general experiment management systems, as they assume a set of scientific workflows that automate the process of connecting simulation software.

3.2.4   Interoperability Patterns in Digital Library Systems Federations

Authors: *Leonardo Candela, Paolo Manghi, Pasquale Pagano – Institute of Information Science and Technologies, National Research Council of Italy*

Presented by: *Paolo Manghi*

Abstract: Service providers willing to offer functionality over an aggregation of information objects are becoming key actors in the digital library world. This aggregation is usually performed by collecting and processing objects from a federation of data providers, for example, institutional repositories, data archives, through well-known standard protocols for data exchange. Higher level interoperability issues affect these scenarios and determine the quality of the service offered and the success of these provider initiatives. Interoperability issues are mainly due to the impedance mismatch occurring between the data models of the objects to be exported by data providers and the data model of the service provider and can be measured in terms of "structural heterogeneity", "semantic heterogeneity" and "granularity-of-representation heterogeneity". The authors define a basic architecture through which to describe interoperability patterns arising when

constructing such a federation and present the D-NET software Toolkit as a general-purpose practical solution.

Discussion

The main issue the discussion addressed was the schema mapping issue. Ed Fox asked about the role of experts on the architecture proposed in this paper. Paolo Manghi pointed out that they tried to create a solution that is generic enough to solve a number of common scenarios on schema mapping and values transformation. They have created end-user interfaces to define mappings between different schemas but those can be used only if the schemas are of certain kind. The complexity of the schema is not challenging to make the interface unusable. Sarantos Kapidakis stressed the difficulty on providing solutions for schema mapping. He mentioned another common case where many Digital Libraries are supposed to have the same schema but actually they have different quality in their implementation of the schema because every schema has a lot of optional parts on its cataloguing rules. He asked about the way the proposed architecture handles the different quality of the records on the same schema. Paolo Manghi replied that this is a kind of mapping values into other values and transforming them from one domain to another. Depending on the different solutions, information can be lost or can be enriched. Kevin Ashley pointed out that the work of this paper sounds quite valuable. He asked about the result when two Organisations try to merge their collections. Paolo Manghi replied that the merging of collections comes after the schema mapping and there are certain services that deal with this problem.

4. Workshop Program - Friday, 10 September 2010

4.1 Session 3: The DL.org Approach to DL Interoperability

4.1.1 Framework for Digital Library Function Description, Publication, and Discovery: A prerequisite for interoperable digital libraries

Authors: *George Athanasopoulos, Katerina El Raheb, George Kakaletris and Natalia Manola - University of Athens, Greece; Edward Fox - Virginia Tech, U.S.; Carlo Meghini - Institute of Information Science and Technologies, National Research Council of Italy; Andreas Rauber - Vienna University of Technology, Austria; Dagobert Soergel - University at Buffalo, U.S*

Presented by: *Katerina El Raheb*

Abstract: Digital Library interoperability, as reported in the DL.org Digital Library Reference Model, is intimately bound up with function interoperability. A prerequisite for the latter is an appropriate function description, publication and discovery mechanisms. The importance of a framework that accommodates the specification of key digital library function characteristics, such as interface, behaviour, dependencies and semantics, has been underscored by the DL.org Functionality Working Group. Such a framework should be used in appropriate registries that cater for the publication and discovery of DL functionality. The authors report the findings of the DL.org Functionality Working Group in terms of a DL function description framework and a set of contemporary registries that can serve as the basis for the provision of a DL function interoperability enabling registry.

Discussion

The discussion opened by addressing the issue of pre and post conditions of functions. Katerina El Raheb pointed out that these conditions are used to ensure the quality of functions. She gave the example of invoking a function where there are things that the user would like to do, but he wouldn't want to change some other things. George Athanasopoulos added that the framework proposed in this paper is at the moment a conceptual idea and it hasn't been implemented yet. The pre and post conditions of a function can be user to express what needs to exist before calling a function and what the result will eventually be after calling the function. This can be used to perform certain advanced operations such as to compose functions or to assert whether some functions can be composed. The idea is to use some kind of formalism but this is something that will be done in a future work. Donatella Castelli asked about the scenarios that have been identified for testing the framework. Katerina El Raheb mentioned that these scenarios were included in the survey of existing applications that can be adapted for the framework. George Athanasopoulos added that based on the requirements for implementing this framework, the authors tried to identify what is available in the literature that can be used in order to satisfy these requirements. He, finally, pointed out that the authors decided to use the Service-oriented Computing domain because the notion of function is related to the notion of service.

4.1.2   Quality interoperability within digital libraries: the DL.org perspective

Authors: *Giuseppina Vullo - University of Glasgow, Scotland, UK; Genevieve Clavel - Swiss National Library; Nicola Ferro - University of Padua, Italy; René van Horik - Data Archiving and Networked Services, Netherlands; Wolfram Horstmann - University of Bielefled, Germany; Sarantos Kapidakis - Ionian University; Seamus Ross - University of Toronto*

Presented by: *Giuseppina Vullo*

Abstract: Quality is the most dynamic aspect of digital libraries and becomes even more complex when it comes to interoperability. The authors formalize the research motivations and hypotheses on quality interoperability conducted by the DL.org Quality Working Group. The authors start by providing a multi-level interoperability framework adopted by DL.org and go on to illustrate key research points and approaches connected with the interoperability of DL quality, grounding them in the DL.org Reference Model, which stems from the DELOS Network of Excellence. By applying the Reference Model concept map to their interoperability motivating scenario, the authors subsequently present the two main research outcomes of their investigation: the Quality Core Model and the Quality Interoperability Survey.

Discussion

Geneva Henry opened the discussion that revolved around the quality interoperability survey and the targeted digital libraries. Giuseppina Vullo pointed out that they decided to target three kinds of digital libraries: very large and general digital libraries like Europeana, digital repositories of scientific work like university digital repositories and specialized digital libraries. They decided not to include commercial digital libraries. The data that they received were interpreted according to the kind of organisation that provided these data. Ed Fox mentioned that the Organisational issue that was included in the quality survey is really important. He asked why user, functionality, and architecture quality parameters are not addressed in the Quality Core Model. Giuseppina Vullo replied that, indeed, the Quality group has to deal with functionality quality parameters before finalizing the model. At the beginning, they faced about 40 quality parameters. The experts of the Quality Working Group decided that the quality of policies on the Organisational level and the quality of content from the semantic and technical level were really crucial. Based on this selection, the group decided to structure its survey. The group is aware that there is a limitation in their work because the quality domain is really too broad, but it wasn't possible to deal also with user aspects in only one year.

4.1.3   Modeling Users and Context in Digital Libraries: Interoperability Issues

Authors: *Anna Nika, Akrivi Katifori and Natalia Manola - University of Athens, Greece; Tiziana Catarci - Sapienza University of Rome, Italy; Georgia Koutrika - Stanford University, California, U.S.; Andreas Nürnberger - Otto-von-Guericke University Magdeburg, Germany; Manfred Thaller - University of Cologne, Germany*

Presented by: *Anna Nika*

Abstract: The modeling of user characteristics is an important mechanism for the support and enhancement of personalization in Digital Libraries (DLs). Contextual information is sometimes considered in the literature as part of the user model, although it in fact influences user-DL interaction. However, context modeling and user modeling are strongly interrelated. Similar methods are employed for the representation of user context and user model and the use of ontologies constitutes a promising approach for such a representation because they facilitate the sharing of information and reinforce interoperability. In this paper, we define the user model and user context for Digital Libraries, we examine identified context dimensions as well as context-independent and context-dependent user model attributes, we proposes the use of an ontology for user representation in DLs, and we discuss the advantages of such an approach for augmenting personalization and achieving user interoperability.

Discussion

After the end of the last presentation, Ed Fox opened the discussion and addressed the issue of ontologies. Specifically, he pointed out that ontologies are becoming a fundamental part of digital libraries for all the activities. But, the use of ontologies is not simple because they require ontology mapping, ontology conversion and other things. The Functionality Working Group also addressed the issue of ontologies, but this paper should be the basis of discussions on the use of ontologies in digital libraries. Donatella Castelli mentioned that because digital libraries are so broad, a core should be defined first. Ed Fox replied that the use of concept maps in the Reference Model could be enriched to some extent by using ontologies. Donatella Castelli asked if the authors of this paper tried to validate the ontology in real cases. Anna Nika replied that they tried to examine user models of existing projects. It wasn't possible to validate the ontology because current projects like DRIVER use very simple user models. It would be a good future work to check other systems and try to validate the ontology.

## 4.2   Session 4: Digital Libraries and Linked Data

Following the presentations, all participants joined a discussion, chaired by Seamus Ross on the topic

"Digital Libraries and Linked Data". The discussion opened by Yannis Ioannidis, defining what are Linked Data (http://linkeddata.org/), and underlining the connection but also the differences between Linked Data and Digital Libraries, as the latter can be considered a lose form of Organised connected data, whereas Linked Data is web-based amorphous huge amount of data semantically connected. The first question raised was "What is the difference between the terms Linked and Interoperable", and as Costantino

Thanos explained the difference is that if two actors are linking their data this linking does not guarantee that these actors can interoperate, i.e., work together, collaborate and exchange content and services. Fact is Interoperability is a wider concept than Linked Data, i.e., Linking is a subset of Digital Library Interoperability, which involves syntactic, semantic and functionality interoperability, although semantics and tools of the Semantic Web are very important in connecting knowledge and operations in a way that draws the path towards interoperability.

Most participants agreed that there is a need to connect the data of a Digital Library with what is outside, as it is essential for Digital Libraries to connect, translate and communicate their information. Answering the question "How Linked Data can contribute to Digital Library Interoperability?", a couple of examples where given where syntactic search, or connecting data statically through keywords is not enough, underlining the importance of semantic enhancement in advanced search of the content.
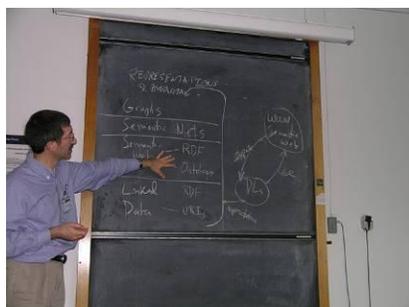
Nonetheless, semantics are needed in module architecture, if we need to achieve interoperability.

Ed Fox contributed to clarifying the role that Linked Data as a tool of the Semantic Web Technologies should play by giving a schema which shows this relation. According to this schema: 1) all of Graphs, Semantic Nets, Semantic Web providing RDF and Ontologies, and Linked Data including RDF and URIs are all tools of Knowledge Representation that a Digital Library can use and 2) Digital Libraries use and at the same time are exposed through the WWW (Semantic or not).

Another question raised during the Brainstorming is how Links are going to be created and if these links are going to be the result of random, dynamic creation or of an automated process designed by human intellectual process. Apparently, we need both ways for creating links but what is more important is to retain the transparency of this linking processes, as at this point raises the issue of provenance of content and as a result the issue of trust and quality. Obviously there is a need to check the quality and trustworthiness of this kind of links, in relation with the creator of the data and/or the links.

The next question raised was "What is the relation of Lined Data and Annotations? Are Linked Data just another way of annotating content?" An answer to this were some examples of the Europeana project e.g. using thesaurus for connecting objects that they don't share the same key-words in their description but are conceptually linked. These examples show that using the concept of semantically Linked Data in Digital Libraries is not new, but Linked Data is much more than annotating, as the hot question is not just creating or adding metadata, but also how to create these metadata in order to allow syntactic and semantic connection. Another point highlighted is that by using unique URI's we define the identity of datasets, including their connections to other concepts, so actually we have a kind of context-free connections between data.

After a comment about Digital Libraries being between Databases and the Web in terms of Restricted Vs Open Access and quality evaluation of content, the need to establish connections among Digital Libraries but also between Digital Libraries and the (Semantic) Web was expressed and, a discussion has followed on if and under what conditions Digital Libraries should expose their content to the web without affecting their quality and policy. Nevertheless, semantics and Linked Data should be

investigated in terms of to improving user profiles and addressing user interoperability issues. Additionally, Digital Libraries interoperability is not only about annotating content but also about annotating and linking information about services. Even if we agree that these content links can be static, when talking about functionality these links should be dynamic. In Digital Libraries, functions must be first class citizens of this domain as moving from one thing (content, data, metadata, function) to another requires appropriate handling, i.e., description and discovery of functions.

The conclusion of this Brainstorming session can be summarized into these main statements:

1) There is a huge difference between the web (Semantic or not) and Digital Libraries issues in terms of addressing issues such as open access, content quality evaluation, specifics policy applied, and services that are implemented.

2) Semantics provide a useful means to achieve different kind of interoperability and Linked Data as part of the Semantic Web Technologies should be investigated by Digital Libraries under this context.

5. Position Statements

Several key experts attending the workshop provided their position statements on the current landscape of digital libraries, feedback on the workshop and key points to be addressed for DL.org.

Pablo de Castro – Carlos III University Madrid, Spain

Interesting discussion on linked data! Suggested ORE-OAI would be worth taking a look at as precedent specification for digital object linking (and particularly research datasets). If this standard was widely-enough implemented, it might make up for a good starting point for further work.

Great workshop! Looking forward to having future opportunities for cooperation with DL.org from the Sonex Workgroup! Thank you!

Geneva Henry – Rice University, USA

The discussions throughout the workshop were interesting but sometimes seemed to focus on concepts tangential to the key topic of interoperability. It would be good to continually tie the discussions to the main theme so that the workshop is more coherent.

Perla Innocenti – University of Glasgow, UK

Re: Digital Libraries and Linked Data Session

I think it is useful that we address linked data from a DL perspective, in particular regarding the ever increasing importance of URIs and the implications for their reference and accessibility; the formalisms that will inform a new generic metadata-driven structure and interfaces for linked data; responsibilities and validation in a web-centric environment; the possibilities opened by e.g. the open annotation

framework and memento for persistent web centric annotations and analysis of time series referencing a single URI; and finally sustainability of linked data.

Sarantos Kapidakis – Ionian University, Greece

Interoperability has many aspects, the data and controlled vocabularies interoperability is attached with linked-data. Linked-data is crucial for all next generation DL. There is a debate on simplicity vs. functionality and linked data serves the functionality perspective. The linking is not simple, as the context has to be considered.

Jonathan Leidig – Virginia Tech, USA

The work in functionality descriptions should be leveraged to produce a collection/repository/registry of DL functions for developer discovery and reuse.
The topic of linked data internal and external to a DL is worth perusing in the DL community.

Lukasz Mesek – Jagiellonian University, Poland

The issue of linked data should be discussed in the context of practical solutions.
It would be useful to try to apply the linked data model to some of existing and working digital libraries and find out how the model works in reality.

Edwin Montoya – EAFIT University, Colombia

Discussions about the relation between interoperability and lack of standards!
What about the standardization in DL?
How to build best practices in DL in order to apply those?

Giuseppina Vullo – University of Glasgow, UK

I really trust on the feedback DL research community can have from the community of practice to face the interoperability challenges.
As for the linked data issue, I think DLs can benefit by the advances of Semantic Web without forgetting the active role of users and society and the values implicit in the open access to information.

6. Conclusions

The 2nd DL.org workshop on Digital Library interoperability issues successfully brought together researchers and practitioners from the broad Digital Library community that discussed the current approaches towards resolving practical interoperability issues and shaped the important questions that need to be answered in order to achieve interoperability among the future Digital Library systems.

The workshop has been a big success in particular when considering the quality of the selected papers and their presentations. This was achieved due to the excellent work of authors and reviewers. The workshop participants took an enthusiastic approach to each presentation asking various questions and making useful comments that the authors could use for the revised version of their papers.

The different presentations of the workshop provided insights into the current solutions towards resolving different levels of interoperability issues and raised a number of important questions for future interoperable Digital Libraries. In particular, Alex's Wade presentation about Microsoft's approach to interoperability stressed the importance of exploration of development of digital data libraries that contain scientific data and support integration with published literature. Pasquale Pagano presented the D4Science interoperability approach that raised questions about the automatic generation of workflows that are important to support the interaction between the interoperable systems. Peter Burnhill and Pablo de Castro stressed the significance of policy interoperability and the importance of systematic attention of the sharing of scholarly content through a network of Open Access repositories. Jonathan Leidig presented a generic, component-based simulation supporting a digital library that is customised through provided schemas and extensible through the addition of components. Paolo Manghi raised important questions about the mismatch occurring between the different data models that can be measured in terms of "structural heterogeneity", "semantic heterogeneity" and "granularity-of-representation heterogeneity". Katerina El Raheb reported the future significant aspects of function interoperability in terms of a DL function description framework and a set of contemporary registries that can serve as the basis for the provision of a DL function interoperability enabling registry. Giuseppina Vullo presented the multi-level interoperability framework adopted by DL.org and illustrated key research points and approaches connected with the interoperability of DL quality. Anna Nika presented the User WG approach for the representation of users in Digital Libraries through the use of ontologies that support user model interoperability. Ed Fox supported the fact that onologies are becoming a fundamental part in Digital Libraries and should be further discussed in the future. Finally, the interactive discussion among the participants of the workshop about Linked Data and Digital Libraries provided significant questions associated with the relation between Linked and Interoperable, the contribution of Linked Data to Digital Library Interoperability, and the relation between Linked Data and Annotations. The Brainstorming session reached two important conclusions, i.e., (a) the existence of great difference between the web (Semantic or not) and Digital Libraries issues in terms of open access, content quality evaluation, specific policies applied, and services that are implemented, and (b) Semantics provide a useful means to achieve different kind of interoperability and Linked Data as part of the Semantic Web Technologies should be investigated by Digital Libraries under this context.

During the project's lifespan, DL.org organised two workshops, both of them in the frame of ECDL. It has to be noticed that both events have been evaluated very positively by the participants. By almost all of them it has been highlighted that there is a need for similar events, as they offer a springboard for

further discussions and exchanges of ideas on effective methods for DL interoperability and best practices.

7. Annexes

7.1 Workshop Agenda

Date: Thursday and Friday, 9-10 September 2010, **Venue:** University of Glasgow, Scotland, UK, **Host Event:** ECDL2010 - The European Conference on Digital Libraries

Thursday, 9-9-2010			
From	To		Duration
2:30 PM	3:00 PM	**Welcome and introduction** *Donatella Castelli, Yannis Ioannidis, Seamus Ross*	0:30
		Session #1: Invited Talk Chair: Donatella Castelli	
3:00 PM	4:00 PM	**Digital Library Interoperability: An Industrial Perspective** *Alex Wade*	1:00
4:00 PM	4:30 PM	**Coffee Break**	0:30
		Session #2: DL Interoperability Principles and Practice Chair: Yannis Ioannidis	
4:30 PM	5:00 PM	**The gCube Interoperability Framework** *Leonardo Candela, George Kakaletris, Pasquale Pagano, Giorgos Papanikos and Fabio Simeoni*	0:30
5:00 PM	5:30 PM	**Handling Repository-Related Interoperability Issues: the SONEX Workgroup** *Peter Burnhill, Pablo deCastro, Jim Downing, Richard Jones and Mogens Sandfaer*	0:30
5:30 PM	6:00 PM	**Epidemiology Experiment and Simulation Management through Schema-Based Digital Libraries** *Jonathan Leidig, Edward A. Fox, Madhav Marathe and Henning Mortveit*	0:30
6:00 PM	6:30 PM	**Interoperability Patterns in Digital Library Systems Federations** *Paolo Manghi, Leonardo Candela and Pasquale Pagano*	0:30

Friday, 10-9-2010			
Session #3: The DL.org Approach to DL Interoperability Chair: Costantino Thanos			
9:00 AM	9:30 AM	**Framework for Digital Library Function Description, Publication, and Discovery: A prerequisite for interoperable digital libraries** *George Athanasopoulos, Katerina El Raheb, Edward Fox, George Kakaletris, Natalia Manola, Carlo Meghini, Andreas Rauber and Dagobert Soergel*	0:30
9:30 AM	10:00 AM	**Quality interoperability within digital libraries: the DL.org perspective** *Giuseppina Vullo, Genevieve Clavel, Nicola Ferro, Sarah Higgins, René van Horik, Wolfram Horstmann, Sarantos Kapidakis*	0:30
10:00 AM	10:30 AM	**Modeling Users and Context in Digital Libraries: Interoperability Issues** *Anna Nika, Tiziana Catarci, Akrivi Katifori, Georgia Koutrika, Natalia Manola, Andreas Nürnberger and Manfred Thaller*	0:30
10:30 AM	11:00 AM	**Coffee Break**	0:30
Session #4: Digital Libraries and Linked Data Chair: Seamus Ross			
11:00 AM	12:30 PM	**Brainstorming** *ALL*	1:30
12:30 PM	1:00 PM	**Conclusions** *Donatella Castelli, Yannis Ioannidis, Seamus Ross*	0:30

7.2 Call for Papers for the 2nd DL.org Workshop

Making Digital Libraries Interoperable, 9-10 September 2010, during the 14th European Conference on Digital Libraries (ECDL2010), Glasgow, Scotland, UK
http://www.dlorg.eu/index.php/dl-org-events/2nd-dl-org-workshop

Workshop Theme
The central theme of the 2nd DL.org Workshop will be "Digital Library Interoperability". Interoperability is a multi-layered and very context-specific concept. It encompasses different levels along a multidimensional spectrum. The Workshop will address this difficult problem from

several perspectives: content, user, functionality, policy, quality, and architecture. Many of the above facets include open research issues. This Workshop will highlight the several dimensions of the problem and indicate new research directions.

Paper Submission

We are soliciting two types of contributions: Research Papers and Project Papers.

Research Papers

Authors are invited to submit original research papers addressing current approaches and new research directions for tackling the multi-faceted digital library interoperability issues.

Topics of interest include, but are not limited to:

- Theoretical Foundations of Interoperability in Digital Library
- Emerging interoperability issues
- Standard-based approaches toward interoperability
- Technologies promoting interoperability
- Mediator-based approaches toward interoperability
- Concrete implementations and exploitations of interoperability solution
- Evaluation of interoperability approaches and solutions
- Interoperability Models
- Ontology driven interoperability
- Metadata interoperability
- Interoperability levels

Research papers will have a full-length oral presentation and will be published in a high-quality proceedings volume. Each submitted paper must not exceed 10 pages in total.

Project Papers

Authors are invited to submit papers focusing on DL interoperability approaches and solutions adopted and lessons learned when implementing interoperable DL systems in the context of European, international and national projects.

These papers will have a short oral presentation and will be included in the "Project Papers" section of the Workshop proceedings. Each submitted paper must not exceed 8 pages in total.

Formatting

All contributions must be written in English. They must follow the formatting guidelines of Springer's Lecture Notes in Computer Science (LNCS - http://www.springer.com/computer/lncs?SGWID=0-164-6-793341-0) and submitted via the workshop submission system (http://www.easychair.org/conferences/?conf=dlorg2010).

Workshop Reviewing Process

The reviewing process will be carried out by the members of the Workshop International Programme Committee. Two referees will review each paper.

Programme Committee

- Tiziana Catarci, Professor of Computer Science, Sapienza University of Rome, Italy - Member of DL.org's User Working Group
- Jane Hunter, Professor at the School of Information Technology & Electrical Engineering, University of Queensland, Australia - Member of DL.org's Liaison Group
- Ronald Larsen, Dean of the School of Information Sciences, University of Pittsburgh, U.S. - Member of DL.org's Liaison Group
- John Mylopoulos, Professor of Computer Science, University of Toronto, Canada
- Fabio Simeoni, Research Assistant, University of Strathclyde, Scotland, UK
- Dagobert Soergel, Dean of the Library & Information Studies, University at Buffalo, U.S. - Scientific Chair of DL.org's Functionality Working Group

Workshop Organisers

- Donatella Castelli, Senior Researcher, Institute of Information & Technologies, National Research Council of Italy - DL.org Co-ordinator
- Yannis Ioannidis, Professsor at the Department of Informatics and Telecommunications, University of Athens, Greece - Scientific Chair of the User Working Group
- Seamus Ross, Dean of the Faculty of Information, University of Toronto, Canada - Member of DL.org's Working Groups on Policy & Quality

Important Dates

Paper Submission: 20 June 2010
Notification of Acceptance: 16 July 2010
Camera ready Papers: 30 July 2010