DL.org

Digital Library Technology & Methodology Cookbook

An Interoperability Framework, Best Practices & Solutions

www.dlorg.eu

COOPERATION

DL.org Digital Library Technology & Methodology Cookbook

This booklet is abstracted and abridged from “The Digital Technology and Methodology Cookbook”, D3.4 DL.org Project Deliverable, April 2011

Authors

G. Athanasopoulos[1], L. Candela[2], D. Castelli[2], K. El Raheb[1], P. Innocenti[3], Y. Ioannidis[1], A. Katifori[1], A. Nika[1], S. Ross[4], A. Tani[2], C. Thanos[2], E. Toli[1], G. Vullo[3]

[1] Department of Informatics and Telecommunications - National and Kapodistrian University of Athens

[2] Istituto di Scienza e Tecnologie dell’Informazione - National Research Council of Italy

[3] Humanities Advanced Technology and Information Institute - University of Glasgow

[4] Faculty of Information – University of Toronto

Contributing Authors

K. Ashley (DCC), P. Burnhill (University of Edinburg), T. Catarci (University of Rome “La Sapienza“), G. Clavel-Merrin (Swiss National Library), P. De Castro (Carlos III University Madrid), A. De Robbio (University of Padua), J. Faundeen (USGS), N. Ferro (University of Padua), E. Fox (Virginia Tech), S. Higgins (DCC), R. van Horik (DANS), W. Horstmann (Bielefeld University Library), R. Jones (Symplectic Ltd), G. Kakaletris (NKUA), S. Kapidakis (Ionian University of Corfu), G. Koutrika (Stanford University), P. Manghi (CNR), N. Manola (NKUA), C. Meghini (CNR), R. W. Moore (University of North Carolina at Chapel Hill), L. Moreau (University of Southampton), A. Nürnberger (University Magdeburg), P. Pagano (CNR), H. Pfeiffenberger (Alfred Wegener Institute), A. Rauber (TU-- Wien), M. Smith (MIT), D. Soergel (University of Buffalo), M. Thaller (University of Cologne)

Acknowledgements

This work has been partially supported by DL.org (December 2008-February 2011), a Coordination and support action, received funding from the Commission of the European Union (EC) under the 7th Framework Programme ICT Thematic Area “Digital libraries and technology-enhanced learning” through the EC’s Cultural Heritage and Technology Enhanced Learning Unit.

Designed and editorial work by Trust-IT Services Printed by Osmotica

Contents

1. Introduction

The needs for '*building by re-use*' and '*sharing*' have grown out of the demand for powerful and rich Digital Libraries supporting a large variety of interdisciplinary activities coupled with the data deluge which the information society is now facing. Interoperability at a technical, semantic and organisational level is a central issue to satisfy these needs. Interoperable systems broaden choice and open up new perspectives for researchers, governments and citizens across a spectrum of disciplines and domains. Interoperability is key to improve Digital Libraries, enabling wider collaborations and ensuring that a broader spectrum of resources are available to a wider range of people whether for simple consumption or to enhance research activities. Although the importance of interoperability is well known and many attempts have been made in the past to give solutions to interoperability problems, there still is a lack of systematic approaches, and, on average, a scarce knowledge of existing solutions which remain confined to the systems they have been designed for.

The need for interoperability goes actually well beyond the digital library domain. Interoperability is among the most critical issues to be faced when building systems as "collections" of independently developed constituents (systems on their own) that should co-operate and rely on each other to accomplish larger tasks. The "Digital Agenda for Europe" (European Commission, May 2010), one of the seven flagship initiatives of Europe's 2020 Strategy, outlines seven priority areas for actions; the second one concerns "*improving the framework conditions for interoperability between ICT products and services*". This key priority foresees that it is essential to enhance interoperability between devices, applications, data repositories, services and networks inside a framework where the conditions for interoperability can be improved in various ways. One important means to that end is to ensure that good ICT standards are available and used, notably in public procurement and legislation.

Interoperability is actually a multi-layered and context-specific concept, which encompasses different levels along a multi-dimensional spectrum ranging from organisational to semantic and technological aspects. DL.org has investigated interoperability from multiple perspectives: *content*, *user*, *functionality*, *policy*, *quality*, and *architecture*. It has also examined interoperability at *technical*, *semantic* and *organisational* levels, all central to powerful Digital Libraries needed in today's context. DL.org is the first initiative to examine interoperability from an all-encompassing perspective by harnessing leading figures in the Digital Library space globally. The output is an innovative *Digital Library Technological and Methodological Cookbook* with a portfolio of best practices and pattern solutions to common issues faced when developing interoperable digital library systems. A key facet of the Cookbook is the interoperability framework that can be used to systematically characterise diverse facets linked to the interoperability challenge as well as current and emerging solutions and approaches. The Cookbook is designed to facilitate the assessment and selection of the solutions presented, enabling professionals working towards interoperability to define and pursue the different steps involved. This publication presents the Interoperability Framework and discusses interoperability from the perspectives of the content, user, functionality, policy, quality and architecture domains.

2. The DL.org Interoperability Framework

One of the major difficulties surrounding Digital Library interoperability is the lack of a common model on the one hand and the lack of access to existing and emerging solutions on the other. There is no single interoperability solution or approach that is generic and powerful enough to serve all the needs of digital library organisations and digital library systems. Actually, there is no single definition of interoperability which is accepted in the Digital Library community or by other communities facing this kind of problem.

To deal with this lack, DL.org has defined an Interoperability Framework serving as a blueprint in the drive towards interoperable digital library systems.

The IEEE Glossary defines interoperability as "*the ability of two or more systems or components to exchange information and to use the information that has been exchanged*". This definition highlights that to achieve interoperability between two entities (provider, consumer) two conditions must be satisfied: (*i*) the two entities must be able to exchange information; (*ii*) the consumer entity must be able to effectively use the exchanged information, that is, the consumer must be able to perform the tasks it is willing to do by relying on the exchanged information.

Taking this definition as a starting point, DL.org has identified three concepts:

1. *Interoperability scenario*, i.e., the settings where interoperability takes place;

2. *Interoperability issue*, i.e., a problem hindering an interoperability scenario;

3. *Interoperability solution*, i.e., an approach aiming at removing an interoperability issue to achieve an interoperability scenario.

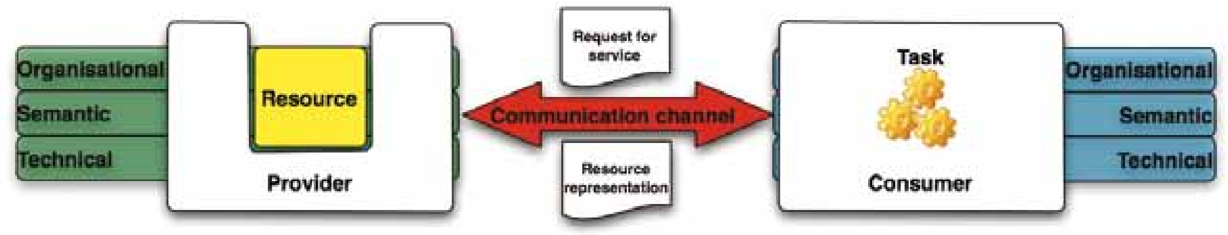

Figure 1. Interoperability Scenario

An ***interoperability scenario*** occurs whenever the following conditions are manifest:

- There are at least two entities that have to co-operate in the context of the scenario. One of the entities is playing the role of *Provider* while the other one is playing the role of *Consumer*.

- The co-operation consists in a Consumer willing to exploit a certain *Resource* – owned by the Provider – to perform a certain *Task* – the work the Consumer is willing to do by relying on that third party Resource.
- To make the scenario feasible the two entities should be able to exchange "meaningful" information. There can be no exchange of information without a communication channel and a protocol regulating the channel functioning, that is, a medium enabling information exchange and some rules governing its effective use to pass information between entities. There can be no information without some form of representation, that is, information is "carried by" or "arises from" a representation. The meaningfulness of the information depends on the Resource and the Task characterising the scenario, that is, the Resource should satisfy the Consumer needs and the Consumer should acquire the information on the Resource that is required to perform the Task (*Task preconditions*).
- The operation of each entity, either Provider or Consumer, depends on *Organisational*, *Semantic* and *Technical* aspects. Organisational aspects capture characteristics of business goals and processes of the institution operating the entity. Examples of organisational aspects are the type of policies governing Information Objects consumption, the type of functionality to be exposed to Consumers, the quality of service to be supported with respect to a specific functionality. Semantic aspects capture characteristics of the meaning of the exchanged digital library resource as well as of the rest of the information exchanged through the communication channel. Examples of semantic aspects are the meaning assigned to a certain policy, the meaning assigned to a certain quality parameter, the meaning assigned to a certain value in a metadata record. Technical aspects capture characteristics of the technology supporting the operation of the entity as well as of the communication channel and the information exchanged through it. Examples of technical aspects are the Digital Library Management Systems (DLMS) used to implement the Digital Library, the protocol used to expose a certain function, the encoding format of an Information Object. It is important to note that these three levels influence each other in a top-down fashion, i.e., organisational aspects set the scene for the entire domain characterising its scope and its overall functioning, semantic aspects define the meaning of the entities involved in the domain according to the organisational aspects, technical aspects have to put in place or implement the organisational and semantic aspects.

An ***interoperability issue*** occurs whenever the Task preconditions are not met. Task preconditions are not met whenever Consumers' expectations about the Provider's Resource in the context of the Task to be performed are not satisfied by the settings of the scenario. In other words, the technical, semantic and/or organisational aspects characterising the Provider and the Consumer regarding the Resource and the Task are not compatible. Examples of interoperability issues include: the format used by the Provider to represent an Information Object differs from the format expected by the Consumer to support a processing activity; the interface through which the Information Object access function is supported by the Provider differs from the one the Consumer is expected to use for content fetching; the semantic of the search function implemented by the Provider is different from the semantic the Consumer aims at relying on to support a cross system search; the Policy governing Information Object consumption supported by the Provider is different from the Policy expected by the Consumer.

An ***interoperability solution*** is an approach reconciling the differences captured by an interoperability issue. It is based on a generic transformation function that conceptually acts at any of the levels characterising Provider and Consumer interaction – organisational, semantic and technical – to make Provider characteristics and

Consumer needs uniform. Such a transformation function may act on Provider characteristics or on Consumer needs as well as on both. Examples of interoperability solutions include: the transformation and exposure of metadata objects through the harvesting protocol and format expected by the Consumer, the implementation of a search client based on a search interface specification implemented by the Provider, the implementation of policies client-side and server-side to guarantee the agreed quality of service on a distributed search operation.

2.1 Interoperability Patterns

All of the heterogeneous interoperability scenarios and related issues existing in the Digital Library domain can be resolved by relying on two classes of solutions independently of their distinguishing characteristics: 'Agreement-based' approaches and 'Mediator-based' approaches. In practice, interoperability scenarios and issues are complex and require the combination of multiple solutions to be resolved. Even in this case, the constituent solutions are either agreement-based or mediator-based. In some cases agreement-based and mediator-based approaches blend into each other, e.g., a mediator-service is actually implementing part of its mediation function according to the agreement settings and rules.

Agreement-based approaches are the traditional way to achieve interoperability, i.e., agreeing on a set of principles that achieve a limited amount of homogeneity among heterogeneous entities is one of the most effective approaches to reach interoperability. Standards belong to this category and the value of standards is clearly demonstrable. The major drawbacks of these solutions reside in the fact that standards and agreements are challenging to be agreed by different organisations and digital libraries. They often end up being complex combinations of features reflecting the interests of many disparate parties. Moreover, by nature they infringe autonomy of the entities adopting them.

Mediators-based approaches have been proposed to resolve scenarios where there is the need to guarantee an high level of autonomy among the partaking entities. These approaches consist in isolating the interoperability machinery and implementing it in components specifically conceived to link the entities partaking to the scenario. These solutions have been initially conceived in the Information Systems domain (Wiederhold & Genesereth, 1997) and are nowadays used in many cases and realised in many ways. The most important part of such kind of approaches is represented by the 'mediation function', i.e., the interoperability machinery they implement. Primary functions are transformation of data formats and interaction modes. In the majority of cases, developing a mediation function is very demanding and time consuming (e.g., in the case of non collaborative scenarios, it is the developer of the mediation function that should take care of acquiring the knowledge needed to link Provider and Consumer and implement it) while in others it might be semi-automatic (e.g., in the case of collaborative scenarios, the entities involved expose data characterising them according to certain rules and the developer of the mediation function might rely on these characterisations to link them).

With respect to Standards, Mediators are strong in supporting the criteria of autonomy. However, their effectiveness depends on the dynamicity of the parties they are going to mediate, i.e., every time changes occur in the interacting entities there is the need for changes in the interoperability machinery implemented by the mediator.

However, the two classes of approaches described above are not mutually exclusive, they can be combined with each other in concrete interoperability scenarios leading to ***Blending approaches***. The need to combine them arises because of the peculiarities that each scenario or partaking entity has. Thus it may happen either that agreements or standards are not sufficient to satisfy the interoperability needs and they have to be complemented with specific interoperability machinery implemented by a mediator, or that a mediator relies on one or more standards to regulate the interaction with either the Provider or the Consumer.

3 Organisational, Semantic and Technical Interoperability: Best practices and Solutions

In addition to the above framework, the DL.org *Digital Library Technological and Methodological Cookbook* contains a rich array of best practices and pattern solutions to common interoperability issues faced when building interoperable Digital Libraries. These solutions are described as to highlight the following aspects: *(i) overview*: a description of the context of the proposed item including a characterisation in terms of the Interoperability Framework and providing the reader with pointers to extensive descriptions of it; (ii) *requirements*: a description of which settings for Organisational, Semantic and/or Technical aspects should occur in order to make it possible to use the solution; (iii) *results*: a description of the changes resulting from the exploitation of the solution in Organisational, Semantic and/or Technical aspects; (iv) *implementation guidelines*: a description of how the solution has to be implemented; (v) *assessment*: an evaluation of the quality of the proposed approach including an estimation of its implementation cost and effectiveness. A brief overview of the analysed issues and the proposed practices and solutions is described in the remainder of this booklet.

3.1 Content Domain Interoperability, Best practices and Solutions

Content Domain interoperability is the problem arising whenever two or more Digital Library "systems" are willing to interoperate by exploiting each other's content resources. The cookbook investigates approaches dedicated to (i) exposing a characterisation of a Provider's Information Object to allow Consumers to render services by relying on such a characterisation (Information Object Description Publishing/Presentation); (ii) agreement oriented approaches focused on reaching a common understanding on Information Object characterisations (Standards for Information Objects / Metadata); (iii) agreement oriented approaches dedicated to arriving at common understanding on schemas for Information Object characterisation (Application Profile); (iv) approaches aimed at mediating between different Information Object characterisations (Mapping/ Crosswalks); and (v) approaches centred on a common understanding on tokens allowing different resources to be distinguished (Information Object / Resource Identifier).

Approaches related to "Information Object Description Publishing/Presentation" are based on mechanisms to expose metadata of information objects, including compound information objects, so that they can be shared by other systems. Concrete exemplars of this kind of interoperability solutions are: (i) Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH), a protocol that allows Providers to expose, and Consumers to harvest, metadata records (the Resource) in the Dublin Core format, while sharing a common understanding of the notions of repository item, metadata record, and metadata format; (ii) Open Archives Initiative Object

Reuse and Exchange (OAI-ORE), an approach through which a Provider exposes compound information objects (Resource) on the Web so that they can be consumed by any Consumer. Provider and Consumer should share a common understanding of the semantic of the entities forming the ORE model, i.e., Aggregation, Resource Map, and Proxies, and their relationships; (iii) Linked Data, a set of best practices for publishing and connecting structured data on the Web. It allows Providers and Consumers to share any kind of information about an arbitrary Resource, e.g., the description of the resource structure, metadata or contextual relationships with other resources. Provider and Consumer should share a common understanding of the semantics of the RDF terms, as well as the meaning of the additional vocabularies; (iv) Open Data Protocol, an open Web protocol for data sharing and modification that allows the Provider to expose to Consumers any kind of structured and unstructured data (the Resource), including associated metadata and available operations. Provider and Consumer should have a common knowledge of the semantics of the elements forming the abstract data model (Entity Data Model).

Approaches related to "Standards for Information Objects/Metadata" are based on agreement (shared formats) for representing Information Objects and Metadata. Concrete exemplars of this kind of interoperability solution are: (i) Dublin Core, a metadata standard providing a simple set of elements for the description of any kind of resource. It allows Provider and Consumer to share metadata describing arbitrary Resources according to the Dublin Core Metadata Scheme; (ii) Europeana Data Model (EDM), a new proposal for structuring objects and metadata coming from Europeana, the European Digital Library, Museum and Archive. Through EDM, Europeana supports the integration of the various models used in cultural heritage data, retaining the original data while still enabling semantic interoperability. In this context, the Provider is any data provider willing to expose digital objects and related metadata (the Resource) according to the EDM; such objects can be exploited by any Consumer that is able to comply with such a model; (iii) CERIF (the Common European Research Information Format), a formal model to setup Research Information Systems and to enable their interoperation. In this context, the Provider is any data provider willing to expose digital objects and related metadata according to the CERIF; such objects can be exploited by any Consumer that is able to comply with such a model.

Approaches related to "Application Profiles" are based on "a type of metadata schema that consists of data elements drawn from one or more namespaces, combined together by implementers, and optimised for a particular local application". Application Profiles adapt or combine existing schemas into a package tailored to the functional requirements of a particular application, while retaining interoperability with the original base schemas. For this, a Provider exposes metadata describing arbitrary Resources in conformity with an agreed Application Profile. Such metadata can be exploited by any Consumer complying with such an Application Profile. Widely known application profiles are Scholarly Works Application Profile (SWAP), Education Application Profile, Dublin Core Collections Application Profile, DC-Library Application Profile, the AGRIS Application Profile, Biological Data Profile, Darwin Core, DCMI Government Application Profile (DC-Gov).

Approaches related to "Mapping/Crosswalks" consist in defining a mapping of the elements, semantics, and syntax from one metadata scheme to those of another. Crosswalks are commonly used to enable interoperability between and among different metadata schemes/ formats. There are a plenty of existing mappings between metadata formats and schemas (e.g. MARC21 to Dublin Core) and many others will come. Currently, crosswalks are by far the most commonly used approach to enable interoperability between and among metadata schemes. Different approaches can be put in place including the usage of switching schema/ pivot schema or lingua franca. In addition to that, mapping and crosswalks can be used to transform Information Objects from their native format into others, e.g. OAI-ORE.

Approaches related to "Information Object (Resource) Identifier" are usually based on agreements for referring to Resources. Exemplars of this kind of interoperability solution include: (i) Uniform Resource Identifier (URI), that provides Provider and Consumer with a common way to identify any kind of Resource, either abstract or physical; (ii) The Handle System, according to which the Provider is the Handle system itself implementing the Handle protocol, the Consumer is any client that wishes to identify the web resource associated to a handle, while the shared Resource is a persistent identifier (handle) associated with any type of web resource. The largest and best-known implementation of the Handle System is that of the International DOI Foundation (IDF).

3.2 User Domain Interoperability, Best practices and Solutions

Interoperability of DLs in the context of the User Domain is the ability of two or more "DLs" to exchange information about the same user and to use the information that has been exchanged meaningfully and accurately in order to produce useful results as defined by the users of these systems. User-level interoperability of DLs arises with respect to (i) sharing representations of user models (User Model Shared Format); (ii) transforming user models and profiles (User Models and Profiles Conversion); and (iii) cross system authentication and authorisation (Authentication and Authorisation).

Approach related to "User Model Shared Format" enforces the use of a shared syntax and semantics to represent user models. The user modelling community has recently focused on ontology based approaches. The General User Model Ontology (GUMO) is based on OWL and is used for the representation of user model characteristics and their interrelationships. The construction of GUMO is based on the thought of dividing the descriptions of user model characteristics into three elements: auxiliary, predicate and range. This description is called a situational statement. In this framework, two kinds of actors are involved: the user model service (the Provider) and applications (the Consumer). The Resource the two entities are willing to share is a situational statement. The user model service is an application-independent server for accessing and storing user information and for exchanging this information between different applications. An application may add or request information that is stored into the u2m.org server.

Approaches related to "User Models and Profiles Conversion" employ appropriate methods to transform the syntax and semantics of the user model used in one system into those of another system. The Generic User model Component (GUC) applies Semantic Web technologies to provide user model server functionalities. There are two kinds of actors involved in the proposed architecture: GUC (the Provider) and UM-based applications (the Consumer). GUC is a generic component that offers functionalities to store schema models for applications and to exchange user information (user profile) between these models. A UM-based application is an entity outside the GUC that uses GUC to store user information. If an application schema is stored in GUC, the application can upload instances of that schema for particular users, i.e., for every user that uses an application, an instance of this schema (the Resource) is stored in GUC. Such an instance is called a user application-view (UAV). For every user, GUC keeps a UAV repository that stores a UAV for every application the user uses.

Approaches related to "Authorisation and Authentication" are best based on the notion of "federated identity". Two very important interoperability approaches are: (i) *OpenID* – there are two kinds of actors involved in the framework: OpenID Providers (the Provider) and Relying Parties (the Consumer). An OpenID Provider is an OpenID Authentication server on which a Relying Party relies for an assertion that a user controls an Identifier (the Resource). A Relying Party is a Web application that requires proof that the user controls an Identifier; and (ii) *Security Assertion Markup Language* (SAML) – this standard defines an XML-based framework for describing and exchanging security information between on-line business partners. This security information is specified in the form of portable SAML assertions that applications working across security domain boundaries can trust. There are two kinds of actors that are involved in the framework: SAML Asserting Party (the Provider) and SAML Relying Party (the Consumer). An Asserting Party is a system entity that makes SAML assertions. A Relying Party is a system entity that uses the received assertions (the Resource). At the heart of most SAML assertions is a subject (an entity that can be authenticated within the context of a particular security domain) about which something is being asserted. The subject might be a human but might also be some other kind of entity, such as a company or a computer.

3.3 Functionality Domain Interoperability, Best practices and Solutions

Function Interoperability approaches can be divided into three main classes: (i) approaches oriented to resolving interoperability issues at the level of function interface, (ii) approaches oriented to resolving interoperability issues at the level of function behaviour, and (iii) approaches oriented to resolving interoperability issues at the level of function. In addition to that, there are protocols and standards that have been conceived to capture service behaviour.

Approaches related to "Function Interface Reconciliation" rely on the use of function descriptions (usually semantically enhanced ones) for the specification of important properties. There are several approaches with varying degree of automation that can be divided in (a) standard-based and (b) dynamic/mediation-based approaches. As for the standard-based approaches, they propose a relatively static approach consisting in the specification of predefined interfaces for certain types of services and include Function Interface Specification Primitives (e.g. SWDL, SAWDL, OWL-S, WSMO), RosettaNet and e-Framework. As for the dynamic/mediation-based approaches, they are essentially based on the (either semi-automated or fully automated) utilisation of Adapters which can be provided in either an automated or manual way. All these approaches are mainly based on the use of appropriate function (or service) specification. It is important to note that all these approaches are mainly research outcomes that have not been tested on a product or industrial scale. Yellin and Storm[1] propose an approach that facilitates the interoperation of components at interface and protocol level. Based on the use of appropriate semantics and Finite State Machine model, they provide appropriate mechanisms that are able to (semi)automatically synthesise component adapters. Benatallah et al.[2] present a semi-automated approach that leverages manually defined templates for accommodating both interface and behavioural incompatibilities. Differences between services are captured using mismatch patterns, which also help in analysing and resolving

1 Yellin, D., & Strom, R. E. (1997). Protocol Specifications and Component Adaptors. ACM Transactions on Programming Languages and Systems (TOPLAS) , 19 (2), 292–333.

2 Benatallah, B., Casati, F., Grigori, D., Nezhad, H. R., & Toumani, F. (2005). Developing Adapters for Web Services Integration. CAiSE 2005, The 17th Conference on Advanced Information Systems Engineering. Porto, Portugal.

them. Bordeaux et al.[3] provide a formal-based approach for evaluating and accommodating the compatibility of services with respect to interface and behaviour aspects. They use π-calculus to formally represent properties of services conveyed in service description protocols and matchmaking algorithms to evaluate the interoperation of services. Ponnekanti and Fox[4] have also presented an approach that leverages static and dynamic analysis tools to evaluate the replaceability of services.

Approaches related to "Function behaviour reconciliation" are actually complementary to the above ones, i.e. function interoperability is closely related to the anticipation of interface, pre/post condition and behaviour concerns. Therefore, both the above-mentioned static and dynamic approaches resolve behavioural concerns as well. Moreover there are standards for representing functional behaviour, e.g. WS-CDL, WSCL, WS-BPEL. Mediator-based approaches rely on the use of formal behaviour specification mechanisms, e.g., Finite State Models and π-calculus, and appropriate algorithms. These include: (i) AI-based, which use Finite State Machine Models and appropriate AI planning techniques. State Transition Systems are extracted out of service descriptions, which are usually semantically enhanced, and, depending on the constraints enforced by each approach, deterministic or non-deterministic planning algorithms are utilised to assert and provide service integrations; (ii) Deng et al.[5] utilize π-calculus to model the service behaviour and interaction in a formal way. They also propose a method based on the operational semantics of the π-calculus to automate the verification of compatibility between two services and an algorithm to measure the compatibility degree quantitatively; (iii) Peng et al.[6] utilize a model of service behaviour based on Petri-nets with weight. They also propose a formal method to verify and compute service behaviour compatibility; (iv) Stollberg et al.[7] build upon the WSMO to create a mediation model able to handle and resolve heterogeneity that may occur in the Semantic Web Service domain.

Approaches related to "Function Conditions Modelling" are based on specifying the conditions that hold for a function to be called (pre-condition) and the ones that hold after its execution (post-condition), respectively. Mechanisms and approaches that have been used for the description of such conditions include: (i) Function Pre/Post Condition Specification Primitives. Such conditions enable the provision of formal methods and tools that are able to validate systems either at design or execution time. Assertions constitute the most common approach towards the implementation of such mechanisms in several programming languages e.g., Eiffel, Java and C/C++, and (ii) Function Pre/Post Condition Reconciliation Approaches. Pre/Post condition issues have always been confronted along with interface and behavioural issues. Thus, the approaches presented above address this problem as well. A prerequisite in these approaches is the use of formalised representations of pre/post conditions such as the ones used by WSMO and OWL-S.

3 Bordeaux, L., Salaün, G., Berardi, D., & Mecella, M. (2004). When are two Web Services Compatible? VLDB TES'04. Toronto, Canada.
4 Ponnekanti, S. R., & Fox, A. (2004). Interoperability among Independently Evolving Web Services. Middleware'04. Toronto, Canada.
5 Deng, S., Wu, Z., Zhou, M., Li, Y., & Wu, J. (2006). Modeling Service Compatibility with Pi-calculus for Choreography. Proceedings of 25th International Conference on Conceptual Modeling. Tucson, AZ, USA.
6 Peng, Y., Zheng, Z., Xiang, J., Gao, J., Ai, J., Lu, Z., et al. (2009). A Model of Service Behavior Based on Petri Nets with Weights. World Congress on Software Engineering, (pp. 3-6).
7 Stollberg, M., Cimpian, E., Mocan, A., & Fensel, D. (2006). A Semantic Web Mediation Architecture. Proceedings of 1st Canadian Semantic Web Working Symposium (CSWWS 2006).

3.4 Policy Domain Interoperability, Best practices and Solutions

Policy interoperability is virtually an uncharted territory. Hence, for policy interoperability it is more appropriate to talk about a "future" state as opposed to "solutions". Thus the cookbook is not listing best practices only but also a wish-list that digital library stakeholders will try to implement. The following cluster of approaches supporting policy interoperability have been identified: (i) Approaches for access policy interoperability – XML, EML-Election Markup Language, METS, DOI, COUNTER 3 Code of Practice, OpenURL Framework Standard, W3C WAI WCAG - Web Content Accessibility Guidelines, W3C Markup Validation Service, US Federal Government Section 508 Guidelines, DLF ERM Initiative; (ii) Approaches for preservation policy interoperability – PRONOM, DROID, JHOVE, UDFR, Global GDFR, Planets Testbed Beta, OAIS, TRAC, DRAMBORA Interactive toolkit, LOCKSS, Portico's Digital Preservation Service, EAD, METS, OAI-PMH, XML, PREMIS, DIDL, DCMI, MARC, ONIX; (iii) Approaches for Network policy interoperability – iRODs, WSDL, XACML; (iv) Approaches for Intellectual property policy interoperability – METS, NLM XML DTDs for Journal Publishing, Archiving and Interchange, PREMIS, CC-Creative Commons licences; Approaches for authentication policy interoperability – XACML, Shibboleth, Athens; (v) Approaches for evaluation and assessment policy interoperability – DRIVER Guidelines 2.0 for content providers, SHAMAN Assessment Framework; Approaches for Policy Representation and Enforcement for policy interoperability – PLEDGE project, AIR Policy Language, iRODS rules, SWRL, Turtle RDF Triples, REWERSE Policy Language, OWL, KAoS, Web Services Policy Framework (WS-Policy), Web Services Policy 1.5, WSPL, XACML, Rei.

3.5 Quality Domain Interoperability, Best practices and Solutions

One of the main obstacles towards the identification of quality interoperability solutions within the DL field is that often quality is not formally described but implied or "hidden" in the background as a mark of excellence, compliance to standards, effectiveness, performance, and so forth. Approaches and best practices are organised in (i) data quality interoperability frameworks, (ii) web solutions and (iii) guidelines and checklists.

In the data quality research field, specific interoperability frameworks have been built for co-operative information systems (CIS). Co-operative information systems are large scale information systems interconnecting diverse systems of autonomous organisations that share common objectives. These frameworks include: (i) DaQuinCIS Framework – which covers a large set of issues in the areas of assessment, data correction, object identification, source trustworthiness and data integration by offering a suite of data quality oriented services; and (ii) Fusionplex Framework - a system for integrating multiple heterogeneous and autonomous information sources that uses data fusion to resolve factual inconsistencies among the individual sources. It characterises every data source and uses their characterisations to resolve conflicts.

Web based approaches are based on ontologies. Several ontologies have been developed for Web services in the field of Quality of Service (QoS) and SLA (Service Level Agreement), which define acceptable levels of service to be met by factors such as availability, accessibility, integrity, performance, and so forth.

Within the DL field, several guidelines, checklists, and certification methodologies have been produced to solve heterogeneity issues affecting DLs interoperation. The establishment of common rules, standards application and best practices can have very different scopes (data integration, long-term digital preservation, etc.) and focus (content, functionalities, policies, etc). However, the general aim of these tools is to facilitate co-operation and development within DL networks and infrastructures. The following items have been considered: (i) The DRIVER Guidelines 2.0 – constituting a powerful tool to map/translate the metadata used in the repository to the Dublin Core metadata as harvested by DRIVER, but providing orientation for managers of new repositories to define their local data management policies, for managers of existing repositories to take steps towards improved services and for developers of repository platforms to add supportive functionalities in future versions; (ii) The DINI Certification – which has been developed in the context of the German Initiative for Networked Information (DINI). Among its aims is providing users and operators with documentary evidence of a repository's compliance with standards and recommendations; (iii) Data Seal of Approved Guidelines – that was established by a number of institutions committed to durability in the archiving of research data. By assigning the seal, they not only wish to guarantee the durability of the data concerned, but also to promote the goal of durable archiving in general; (iv) DRAMBORA (Digital Repository Audit Method Based on Risk Assessment) – which is a digital repository audit methodology for self-assessment, encouraging organisations to establish a comprehensive self-awareness of their objectives, activities and assets before identifying, assessing and managing the risks implicit within their organisation; (v) Trustworthy Repositories Audit & Certification (TRAC) – which lays the groundwork for international collaboration on digital repository audit and certification between the DCC, RLG (now OCLC-RLG Programs), NARA, NESTOR, and the US Center for Research Libraries.

3.6. Architecture Domain Interoperability, Best practices and Solutions

Architecture Domain interoperability is a multifaceted yet very concrete issue arising whenever two entities, actually two software systems, playing the role of Provider and Consumer are willing to share *Architectural Components* initially owned by the Provider only. Although this problem seems to be well known, in practice this is not the case. It is often mixed and confused with all the others Interoperability problems. Moreover, the problem is well beyond the digital library domain, for instance approaches based on programming practices and methodologies include aspect-oriented programming, subject-oriented programming, multidimensional separation of concerns to cite a few.

In the digital library domain, interoperability approaches falling in this domain are dedicated to (i) expose a characterisation of an architectural component (component profile) to allow a Consumer to use it; (ii) exploit a third party facility by relying on standards; and (iii) exploit a third party component through a mediator service.

Approaches related to "component profile" are aiming at reaching a common understanding of the component features as to provide the provider with the characterisation needed to exploit the component. Currently, there are two major approaches for Architectural Components Profile Modeling. The first one relates to the definition of a profile that is specific (proprietary) to the system which defines it (a sort of proprietary solution). The second one is aimed to model, by defining a general profile structure, possible interoperability scenarios related to service oriented architectures. This latter approach can be exemplified by the profile defined in the context of the eFramework initiative, that seeks to promote the use of the service-oriented approach in the analysis and design of software for use within education and research. A concrete exemplar of this kind of interoperability solution is the WS-I Basic Profile, that is a set of non-proprietary Web services specifications, along with clarifications, refinements, interpretations and amplifications of those specifications which promote interoperability among Web services.

Approaches related to "Standard-based Exploitation of third party Architectural Component" include SRU (Search/Retrieval via URL), OpenSearch, and SWORD (Simple Web service Offering Repository Deposit). These are standards implemented by certain architectural components to make available (provider point of view) and/or exploit (consumer point of view) a well known facility, e.g. a search service in the case of SRU and OpenSearch, a deposit service of a repository in the case of SWORD.

Approaches related to "mediator services" are complementary to standard based ones. In fact, standards represent ideally the best interoperability solution when addressing implementation, but they are not always possible or desirable. This because they do not support existing systems that were not built to the standards, and because they may preclude some optimized implementations. Besides that, the standardization process itself often takes more time than even the implementation of the standards. Mediators are components specifically conceived to host the interoperability machinery. These components realise patterns of various genre including: *(i) Blackboard-based Approaches* – based on components that allow asynchronous communication between components in a system; *(ii) Connector / Adaptor-based Approaches* – based on components that translate one interface for a component into a compatible interface; *(iii) Proxy-based Approaches* – based on components that provides an interface to another component; *(iv) Mediator-based Approaches* – based on components that provide a unified interface to a set of other component interfaces and encapsulate how this set of objects interact; *(v) Broker-based Approaches* – based on components that are responsible for coordinating communication, such as forwarding requests, transmitting results and exceptions; *(vi) Registry-based Approaches* – components used to grant access to other components in a system.

3.7. Cross-domain Interoperability, Best practices and Solutions

In addition to domain oriented solutions, there are some involving concepts that are cross-domain and are gaining a lot of importance in the digital library domain like *provenance*. As stated by the W3C Provenance Incubator Group: "*Provenance of a resource is a record that describes entities and processes involved in producing and delivering or otherwise influencing that resource. Provenance provides a critical foundation for assessing authenticity, enabling trust, and allowing reproducibility. Provenance assertions are a form of contextual metadata and can themselves become important records with their own provenance.*" In its essence, provenance refers to the sources of information, such as entities and processes, involved in producing or delivering an artefact. A proposed solution for provenance interoperability is the *Open Provenance Model (OPM)*, which is designed to allow provenance information about arbitrary resources to be represented in a technology-independent manner, and to be exchanged between systems by means of a compatibility layer based on a shared provenance model. Through this model, a Provider exposes provenance information about arbitrary Resources and such provenance information can be exploited by any Consumer that is able to comply with the OPM.

4. Concluding Notes

Interoperability is one of the challenges faced when building modern Information Systems and Digital Libraries that have to be built as "collections" of independently developed constituents (systems on their own) and that should cooperate and rely on each other to accomplish larger tasks. The DL.org project has addressed the interoperability challenge by bringing together over 50 international experts active in the Digital Library domain through thematic groups investigating the problem from diverse yet complementary perspectives.

However, interoperability is clearly a "hot topic", and new solutions and approaches continue to be designed and developed. Hence, at any point in time the Cookbook should be considered a dynamic artefact that is expected to expand and evolve thanks to the support of the Digital Library community at large.

In order to contact the keepers of this artefact and to contribute to it, please send an email to cookbook@dlorg.eu.

5. Bibliography

Athanasopoulos, G.; Candela, L.; Castelli, D.; El Raheb, K.; Innocenti, P.; Ioannidis, Y.; Katifori, A.; Nika, A.; Ross, S.; Tani, A.; Thanos, C.; Toli, E.; Vullo, G. (2011). The Digital Library Technology and Methodology Cookbook. DL.org Project Deliverable.

Candela, L.; Athanasopoulos, G.; Castelli, D.; El Raheb, K.; Innocenti, P.; Ioannidis, Y.; Katifori, A.; Nika, A.; Vullo, G.; & Ross, S. (2011). The Digital Library Reference Model. DL.org Project Deliverable.

European Commission (2010). A Digital Agenda for Europe – Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. Brussels.

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

Haslhofer, B. & Klas, W. (2010). A survey of techniques for achieving Metadata Interoperability. ACM Computing Survey, 42 (2), Article 7, pp. 1-37

Heiler, S. (1995). ACM Computing Survey. Semantic interoperability , 27, 271-273.

Ioannidis, Y.; Maier, D.; Abiteboul, S.; Buneman, P.; Davidson, S.; Fox, E.; Halevy, A.; Knoblock, C.; Rabitti, F.; Schek, H.; & Weikum, G. (2005). Digital library information-technology infrastructures. International Journal of Digital Libraries , 5 (4), 266-274.

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

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.

Lagoze, C. (2010). Lost Identity: The Assimilation of Digital Libraries into the Web. Cornell University.

Paepcke, A., Chang, C. K., Winograd, T., & García-Molina, H. (1998). Interoperability for Digital Libraries Worldwide. Communications of the ACM , 41, 33-42.

Park, J., & Ram, S. (2004). Information Systems Interoperability: What Lies Beneath? ACM Transactions on Information Systems , 22, 595-632.

Wegner, P. (1996). Interoperability. ACM Computing Survey , 28, 285-287.

Wiederhold, G., & Genesereth, M. (1997). The Conceptual Basis for Mediation Services. IEEE Expert: Intelligent Systems and Their Applications , 12 (5), 38-47.

DL.org (www.dlorg.eu) has mobilised professionals, educationalists and students at various stages in their academic careers mainly from Computer Science and Library and Information Science domains, to promote knowledge in digital library interoperability, best practices and modelling foundations.

DL.org Experts

Kevin Ashley (Digital Curation Centre, UK), Detlev Balzer (Independent Consultant, Germany), Tiziana Catarci (University of Rome “La Sapienza”, Italy), Vassilis Christophides (University of Crete, Greece), Genevieve Clavel-Merrin (Swiss National Library, Switzerland), Antonella De Robbio (University of Padua, Italy), John Faundeen (U.S. Geological Survey Centre, U.S.), Nicola Ferro (University of Padua, Italy), Ed Fox (Virginia Tech, U.S.), Stefan Gradmann (Humboldt University, Germany), C.H.J.P. (Kees) Hendrick (Naturalis - National Museum of Natural History, The Netherlands), Sarah Higgins (Aberystwyth University, Wales, UK), René van Horik (Data Archiving and Networked Services, The Netherlands), Wolfram Horstmann (Bielefeld University Library, Germany), George Kakaletris (University of Athens, Greece), Sarantos Kapidakis (Ionian University of Corfu, Greece), Georgia Koutrika (Stanford University, U.S.), Paolo Manghi (National Research Council, Italy), Natalia Manola (University of Athens, Greece), Carlo Meghini (National Research Council, Italy), Jan Molendijk (National Library of the Netherlands, The Netherlands), Luc Moreau (University of Southampton, UK), Andreas Nürnberger (University Magdeburg, Germany), Pasquale Pagano (National Research Council, Italy), Hans Pfeiffenberger (Alfred Wegener Institute, Germany), Axel Poigné (Fraunhofer Institute Intelligent Analysis and Information Systems, Germany), Pavlos Polydoras (UBITECH, Greece), Andreas Rauber (TU-Wien, Austria), Dirk Roorda (Royal Netherlands Academy of Science, Netherlands), Robert Sanderson (Los Alamos National Laboratory, U.S.), Tefko Saracevic (Rutgers University, U.S.), MacKenzie Smith (MIT Libraries, U.S.), Thornton Staples (Smithsonian Institution, U.S.), Dagobert Soergel (University at Buffalo, U.S.), Manfred Thaller (University of Cologne, Germany), Bram van der Werf (Independent Consultant, Netherlands), René van Horik (Royal Netherlands Academy of Science, Netherlands).

DL.org Liaison Group Members

Tobias Blanke (King’s College London, UK), Peter Brantley (Access for the Internet Archive, U.S.), Schubert Foo (Nanyang Technological University, Singapore), Jane Hunter (University of Queensland, Australia), Joan K. Lippincott (Coalition for Networked Information, U.S.), Clifford A. Lynch (Coalition for Networked Information, U.S.), Leonid Kalinichenko (Russian Academy of Science, Russia), Dean B. Krafft (Cornell University Library, U.S.), Carl Lagoze (Cornell University, U.S.), Ronald L. Larsen (University of Pittsburgh, U.S.), Andrea Mulrenin (Salzburg Research, Austria), Erich J. Neuhold (University of Vienna, Austria), Areti Ramachandra Durga Prasad (Indian Statistical Institute Bangalore, India), Jela Steinerová (Comenius University Bratislava, Slovakia), Shigeo Sugimoto (University of Tsukuba, Japan), Herbert Van de Sompel (Los Alamos National Laboratory, U.S.), Jens Vigen (European Organization for Nuclear Research, Switzerland), Andrew Wilson (National Archives of Australia, Australia).

DL.org Advisory Group Members

Marianne Backes (Virtual Resource Centre for Knowledge about Europe, Luxembourg), Stephen Griffin (National Science Foundation, U.S.), Geneva Henry (Rice University, U.S.).

Italian National Research Council

University of Athens

University of Glasgow | Humanities Advanced Technology & Information Institute

Trust-IT Services

DL.org 231551

ISBN: 978 889 553 4084