TOWARDS THE DEVELOPMENT OF A GENERAL-PURPOSE DIGITAL REPOSITORY

Dimitrios A. Koutsomitropoulos, Anastasia A. Tsakou, Dimitrios K. Tsolis and Theodore S. Papatheodorou

High Performance Information Systems Laboratory, Department of Computer Engineering and Informatics, University of Patras, Rio 26500, Patras, Greece

Email: {kotsomit,tsakou,dkt,tsp}@hpclab.upatras.gr

Keywords: Repositories, Functional Model, Preservation, Interoperability, OAIS.

Abstract: The need for efficient storage, management and consumption of digital resources has long been a puzzle for institutions and organizations that aim at preserving digital information. Numerous standards, specifications and practices have been proposed so far, in order to alleviate the problems that come in to surface when setting up a digital repository. Nevertheless, only few of these initiatives have enjoyed wide spread adoption, possibly because they are bound to specific content, functionality or implementation needs. This paper aims to contribute towards the development of a general-purpose digital repository by first providing a common definition, independent of specific needs and thus widely applicable. It also presents a set of state-of-the-art requirements that are of key importance for the success of any similar task. Finally, it proposes a Functional Model so as to demonstrate how the above specifications can actually be realized.

1 INTRODUCTION

The maintenance and preservation of digital resources is an important challenge and a significant matter that has so far concerned several organizations and institutes. Although several standards, criteria and mechanisms have been used for an efficient and flexible way of creating and managing digital objects, there is still the need for a complete and efficient model that will facilitate the establishment of digital repositories.

One important step towards this direction has been accomplished by the joint RLG-Commission on Preservation and Access Task Force on Archiving Digital Information in 1994. The CPA/RLG report (RLG-OCLG 2001) proposes the definition of a trusted digital repository, identifies the primary attributes of this repository, articulates a framework for the development of a certification program and makes several recommendations for future work. Work in these areas has been advanced by the Consultative Committee for Space Data Systems in its Reference Model for an Open Archival Information System (OAIS) (CCSDS 2002) and by many groups and individual institutions that are designing their own digital repository systems. Fedora (Payette and Staples 2002; Payette and Lagoze 1998), Dspace (Bass et al. 2002; MacKenzie 2002), Dienst (CDRLG 2000) and GreenStone (NZDL 2003) are some examples of such systems and architectures of digital repositories. However, this broad interest in digital repositories has revealed that there are still certain issues under concern. More specifically, a clear, independent and widely acceptable definition for a digital repository does not exist. In addition, critical requirements from a variety of perspectives need to be identified, in order to provide a basis for future initiatives and assist to the construction of a "best-practices" framework. Finally, the proposal of a robust and sound functional model can demonstrate the possibility of integrating various technologies, standards and architectural principles in order to reap the benefits of a universal, general-purpose digital repository.

The rest of this paper is organized as follows; Section 2 attempts to provide definitions for a general-purpose digital repository; Section 3 describes a series of key requirements that any repository implementation has to satisfy; Section 4 describes a functional model so as to illustrate how these requirements can actually take effect. Finally, Section 5 summarizes our conclusions and future work.

2 DEFINING THE DIGITAL REPOSITORY

The strict definition of the term "digital repository" is an issue that generates a lot of discussion and it is almost certain that unanimity for a common definition would not be achieved. One of the main reasons for this ambiguity is that a digital repository is mainly characterized by the type of content that makes available and by the reasons that necessitate its creation and functionality. However, technical knowledge is not required in order to understand that the following intuitional definition is correct:

Definition 1: A Digital Repository is a collection of digital entities that are subject to the following three operations: insertion, deletion and retrieval.

This abstractive and simple definition obviously provides a minimal set of requirements for any digital repository. One more refined definition is the following:

Definition 2 (RLG): A Digital Repository is an organization that has the responsibility for long-term preservation[1] of digital resources, as well as for making them available to communities agreed on by the producer and the management authority of the repository.

The need for long-term preservation has long been recognized by important institutions that implement digital repositories (libraries, digitization organizations, standardization consortia etc), as well as by the European Union regarding digitization of cultural heritage (e.g. Spanish Presidency resolution (European Council 2002)). Further analysis of the importance of long-term preservation is outside the scope of this paper.

Let us consider the digital repository as a closed system. Then the digital repository interacts with three entities of the external environment as shown in Figure 1.

Producer: Is the role that corresponds to those people or client systems which provide the information to the repository.

Management: Is the role that corresponds to those people who set the overall policy of the repository, as one component in a broader policy domain. In other words, the management control of the repository reflects only one of Management's responsibilities.

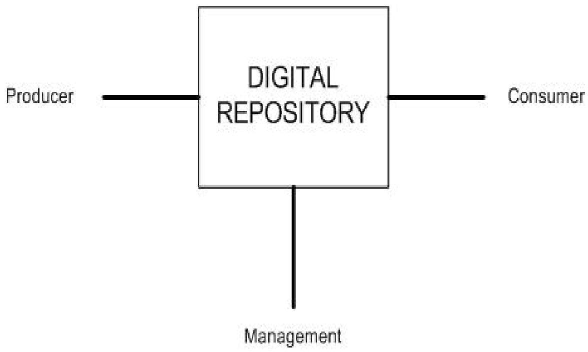

Figure 1: The interaction model between the Digital Repository and the Environment

Consumer: Is the role that corresponds to the people or client systems that interact with the repository's services in order to find and retrieve archival information of their interest. One special class of consumers is the Designated Community.

We have now set the stage for the following final definition that derives from OAIS:

Definition 3 (OAIS): A Digital Repository is an archive that aims at the preservation of digital information for access and use by a Designated Community and satisfies specific requirements[2].

Notice that none of the above definitions is biased towards a specific design or implementation.

3 REQUIREMENTS FOR A DIGITAL REPOSITORY

Having in mind the above digital repository definitions and taking into account current research in the field of digital repositories and digital libraries we recognize a series of six requirements that every

[1] The action of preserving information in a correct and "independently understandable", long-term form. For more information please refer to (CCSDS 2002).

[2] These requirements are described in detail in (CCSDS 2002). They pertain mainly to the quality of the repository services and to IPR issues.

digital repository has to satisfy. In the following, each of these requirements is explained and justified in brief.

Long Term Preservation / Access to the repository's content: A significant step towards preserving digital assets and ensuring their accessibility in the long term is to develop a digital repository conformant to the OAIS specification. Content should also be characterized by the ability of permanent access (e.g. by persistent URI's). At the same time the repository should maintain multiple versions of the content, because digital objects may be modified / updated from time to time (persistency).

Metadata: Metadata should be used during the whole life-circle of the digital content. The main objectives are description of digital content, support for its management and facilitation of access to it, even in the long term (descriptive, administrative, preservation metadata). However, it is important for the metadata to follow a widely adopted standard (Dublin Core (DCMI 2003) can be used in general, MPEG-7 (MPEG 2003) for multimedia content, DIG-35 (DIG 2000) for digital images and METS (DLF 2003) for wrapping and encoding all the above). Furthermore, there are cases where metadata should not only be used for describing individual repository objects but should also support a higher level of abstraction, i.e. the collection level (Dublin Core-Collection Level Description).

Interoperability/Import-Export Capability:
Interoperability can be achieved by adopting well-known standards during the repository's development. One of them is the platform-independent language XML (and XML Schema). Implementation of the OAI-PMH protocol (Lagoze and Van de Sompel 2001), (Van de Sompel and Lagoze 2002) is highly recommended in order to accommodate mass metadata import/export to and from the repository. Support for the Z39.50 (ANSI/NISO 1995) protocol is also of crucial importance, especially for transparent and remote search in a huge amount of documents. Interoperability and accessibility of the digital repository are enhanced by exposing its services as Web Services. Practically, this means that the services will be described using the WSDL language (W3C 2003) and registered with some UDDI registry (OASIS 2002). The major benefit of UDDI is that it enables the automate discovery (and possibly utilization) of a Web Service by the machine, similar to the way that physical users use search engines. Recently, attention seems to draw the ZING Initiative ("Z39.50 International: Next Generation") (Z39.50 IMA 2003) and especially its SRW ("Search/Retrieve for the Web") part. SRW is a web-service-based protocol which aims to integrate access across networked resources, and to promote interoperability between distributed databases by providing a common platform. It features XML and SOAP and thus it is able to integrate more tightly with XML-based infrastructures.
Security/ User Certification: It is clear that none but the Designated Community will be allowed to access the repository's content. A practical way to achieve this is to establish a set of access policies for each Consumer or Consumers' Community, to support their authentication using login/password pairs and/or digital certificates and to cipher access to the repository's services (e.g. SSL).

Intellectual Property Rights Management: The need for copyrighting original content and for economic exploitation of the repository necessitates the management and encoding of IPR information into the content. Watermarking not only for digital images but also for any type of multimedia content is widely used. At a metadata level, we indicatively mention the XML-based MPEG-21, Part 5: Rights Expression Language (MPEG 2002) and the W3C's XML security suite (XML Encryption, XML Key Management and XML Signature).

Knowledge Representation / Management: Repository's content will not be restricted within only one thematic domain, but it may also span over several domains or their combinations. Consequently, it is convenient to describe the content in a semantically hierarchical and structural way. In other words, the establishment of ontologies for each content domain is proposed. For example, the CIDOC ontology (CRM-Conceptual Reference Model (Crofts et al. 2001) can be used for the cultural heritage domain. An ontology-enabled system can assist the user in his search by supporting automated reasoning, even if the information being sought is not explicitly defined in the metadata. Ontologies can also be used for the management of a digital repository; e.g. ABC (Lagoze and Hunter 2001) is capable of organizing events that occurred in the repository at any moment. Traditionally, RDF is used for the development of ontologies; however the DAML+OIL (McGuinness et al. 2002) and the more recent OWL language (W3C 2003) are recommended, as they are specifically designed for ontologies.

4 DIGITAL REPOSITORY FUNCTIONAL MODEL

Implementing the above requirements results to the following *Functional Model of the digital repository* (Figure 2). This functional model is in fact a refinement of the digital repository definition as a closed system, and describes the details of its internal organization as well as its interaction with the external entities Producer, Consumer and Management. It is worthy noting that the functional model proposed here does not necessarily conform to the OAIS reference model. Instead, the conformance to the OAIS specification is left as a designer's choice together with other design choices (see Section 4.3).

The functional model is divided in to four major layers: Insertion, Repository[3] (Actual Storage), Management and Consumption.

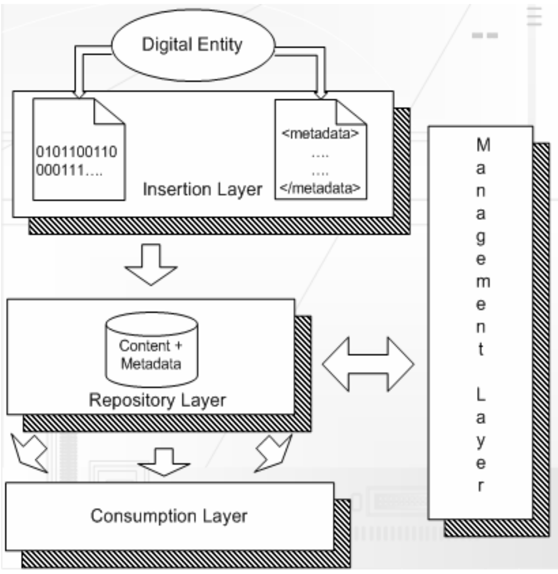

Figure 2: Digital Repository General Functional Model

4.1 Insertion

The Producer has to provide the entity to be preserved in digital format. It is also the Producer's responsibility to provide this entity in acceptable by the repository, accessible and error-free formats (see OAIS - Submission Agreement). Since a digital object is made available, it is inserted in the repository in two parallel components (Figure 3):

- *Binary Upload:* The digital (binary) representation of the object is stored in the repository. This is really an interface through which the Producer can upload files to the repository. The binary file, after being properly watermarked, is stored in the repository's database.
- *Metadata attachment:* Depending on content type, the metadata schemata to be followed are decided in advance. Based on these schemata, the Producer will be provided with a series of forms consisting of the necessary metadata fields that have to be completed (*Manual Insertion*), while efforts will be made to enhance automated metadata extraction (*Automated Extraction*), whenever possible (e.g. the dominant colors of a picture). This is also the phase where the metadata pertaining to the intellectual property rights of the content are being inserted. Before the metadata are stored in the repository, they are properly encoded and, possibly, ciphered.

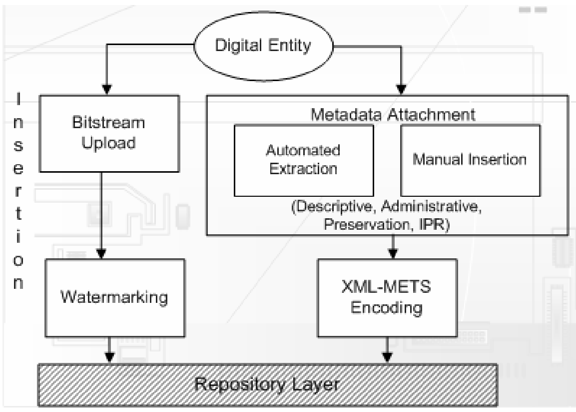

Figure 3: Digital Repository Functional Model in detail: The Insertion Layer

4.2 Repository (Actual Storage)

This is the layer where the content is actually being stored. Figure 4 shows the internal organization of the Repository layer as well as its interaction with the other layers. The Repository layer takes as input the products of the Insertion layer, feeds the Consumption layer and is externally affected by the Management layer choices.

Since the insertion of a digital entity is being conducted in two separate components (file + metadata), actual storage follows a similar manner. We have therefore to deal with the problem of

[3] In order to solve ambiguity between the "digital repository" as a whole and the Repository layer, we refer to the latter with a capital "R".

storing two conceptually separated components: the binary file, where the digital entity is binary encoded, and its accompanying metadata. Depending on the extent to which these components are also physically separated, we come up with the following design choices:

- *Relationally biased solution*: Store files as well as metadata in a multimedia data base. The files are stored as BLOBS and an Entity-Relationship (ER) schema is being designed, which will guide the construction of the corresponding tables in the DB. Metadata are stored in the fields of the aforementioned tables. The problem of disseminating XML information that naturally arises, since XML documents are banned from this solution, can be met by dynamic, on-demand XML document generation from the DB fields. The advantage of this solution is that metadata are directly related to the content they describe, through the ER model relations.
- *Document-biased solution*: In this solution, metadata are also physically separated from the content. Digital files are stored in the multimedia DB, possibly along with some other information (like labels, pointers, identifiers etc), while metadata are stored separately in an XML Document Repository. The latter may be either proprietarily developed or it can be based on some commercial native XML DBMS. The burden of linking together the objects that are stored in the DB with their metadata is placed over business logic, as there is no other away to associate the two components. In other words, there is a need for much more code to be developed as well as for unique identifiers to be used, not only in the DB, but also in the XML documents, in order to maintain consistency; that is, the XML documents point to those objects that they actually describe. Efficient information retrieval from an XML document collection is achieved using the SQL-like XML query languages: XQuery and XQL.
- *Unbiased solution*: Finally, a more balanced solution is to use a relational DB with native XML support (so far Oracle 9i Release 2 and, to a lesser extent, IBM DB2 with XML Extenders). Thus, linkage between content and metadata is more imminent than in the previous solution, for both are virtually managed by the same DMBS. At the same time, metadata remain satisfyingly separated from the content, as they are independently stored in XML documents and follow their own schema.

In any case, digital objects are related to XML documents with a 1:n relationship. An XML document can therefore correspond to multiple digital objects, thus fulfilling the requirement to describe the content also in the collection level.

Separating content from its description, even physically, seems to be the dominant practice in related academic initiatives for repository development (e.g. in (Payette and Staples 2002; McKenzie 2002)). The main advantages of such an approach are summarized below:

- *Content information autonomy*: Since the accompanying information of a digital object,(that is, its metadata) are separated from the object itself, they are completely autonomous. In this way, even in the case of DB failure or break-down, the digital repository remains capable of disseminating object information, despite the fact that the object itself is not accessible.

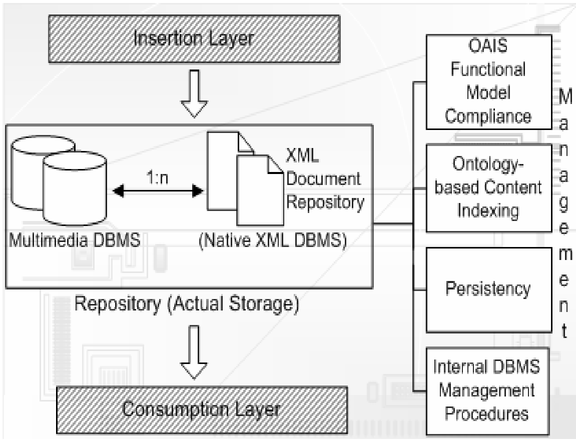

Figure 4**:** Digital Repository Functional Model in detail: The Repository (Actual Storage) and the Management Layers

- *Support for XML-oriented Standards and Interoperability*: Many consumption standards that enjoy great popularity nowadays (like, for example, OAI), demand to be provided with information in XML format. In addition, separate storage of the XML documents seems to be a more natural and, most important, a more efficient solution, especially in the case of mass consumption, compared to dynamic XML document generation from a DB. In addition, the XML's platform–independent nature, its human / machine readable characteristic and its wide-spread adoption in the previous years render it ideal for achieving interoperability with almost every up-to-date information system.
- *Straightforward modification of the metadata schema*: XML Schema language is designed

and proved to be a powerful tool for designing schemata for XML documents and establishing the necessary constraints. By using XML Schema, the process of modifying or merging metadata schemata is significantly simplified, in contrast to modifying a relational DB schema. It consists also the only efficient solution in the case of an evolving, open (semi-structured) schema, as it is frequently the case with metadata schemata.

On the other hand, a complete physical separation between content and metadata has the side-effect of manually maintaining consistency between content and its description and incurs the corresponding programming burden. Finally, it is still open whether a relational DB is more efficient, as a whole, than a native XML DB (e.g. are queries actually answered faster? (Bourret 2003; Florescu and Kossmann 1999)).

4.3 Management

Management here has a meaning similar to the one given in the OAIS definition, that is, management is external to the repository and has nothing to do with the day-to-day maintenance / administration of its content. According to the OAIS specification, management typically is responsible for negotiating financial resources, conducting some regular review process for progress evaluation, determining pricing policies for OAIS services and resolving conflicts involving the external entities that interact with the repository. Effective management should also "provide support for the OAIS by establishing procedures that assure OAIS utilization within the repository" (CCSDS 2002). In the functional model proposed here, Management is also considered to include, in addition to the above, the series of the optional design choices which, if implemented, can affect the repository from the inside. In particular, these choices include (see Figure 4):

- *Conformance to the OAIS reference model*: It involves the establishment of entities and procedures in order for the digital repository to conform with the OAIS specifications as a whole.
- *Ontology-based content indexing*: Digital objects that are ingested by the repository should be indexed based on the ontology describing the domain they belong to. More specifically, every repository object will be related to its ontology class and, subsequently, to its corresponding properties. Typically, these ontologies will be expressed in OWL or DAML+OIL format. Maintenance and evolution of these ontologies will be driven by the metadata produced during Insertion, with little manual intervention or even in a fully automatic manner (Alani et al. 2003). Ontology-based indexing will later allow for content understandability by intelligent agents and for "intelligent queries" submission.
- *Persistency*: This is about ensuring persistent access to the content and keeping multiple versions of the digital objects. It may be considered internal responsibility of the DBMS (at least for storing and keeping track of multiple versions).
- *Internal DBMS Administration Procedures*: These are the common internal procedures of a database, including for example indexing, organization of the information on the physical medium (disk), ER relationship model etc. It may be considered as a design choice, in the sense of the appropriate DB system selection. Part of the day-to-day repository administration is conducted by these functions.

4.4 Consumption and Access Policies

The Consumption layer provides the appropriate interfaces that allow the Consumer Communities to consume the content and access the services of the repository. Between the Repository and the Consumption layer lays the intermediate sub-layer of Access Policies.

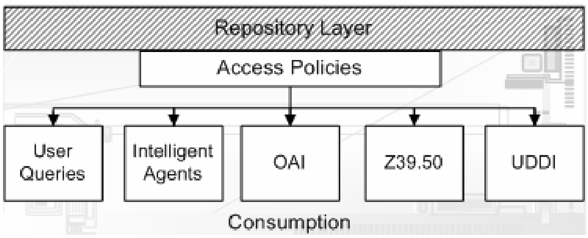

Figure 5: Digital Repository Functional Model in detail: The Consumption Layer and the Access Policies sub-Layer

In this functional model, Access Policies include the implementation of the Security / User Authentication requirements for the digital repository. This sub-layer therefore contains the part of business logic that ensures effective control of access to the repository content and services by Consumers. Access policies must provide for smooth and unobstructed access to the repository by the Designated Community, according to the *Order*

Agreement (see (CCSDS 2002) & Definition 2). In other words, Access Policies make available whole or part of the content to all or specific users or user communities, check whether the user who requests consumption actually fulfills the necessary requirements in order to be served (e.g. billing) and determine whether whole or part of the content will be available through all or specific Consumption interfaces.

The following interfaces are proposed for the Consumption layer:

- *User Queries*: The designated user of the repository should be able to submit queries to the repository databases, as a means to retrieve information and content from it. In fact, the user queries interface is implemented by providing the ability to submit SQL / XQL queries in a user-friendly manner. As a result, the user will be presented with the digital objects as well as (all or part of) their accompanying metadata. In addition, the user should be able to exploit the repository ontologies and submit *intelligent queries*: He should be able to retrieve information, even if this information is not explicitly stored in the repository. Except of the query submission capability, the user should also be able to navigate through the repository, using a hierarchical (ontological or other) organization of the content.
- *Intelligent Agents*: The ontologies constructed for the content and services of the repository will allow for its access and consumption by intelligent agents. Agents should be able to submit intelligent queries to the repository and negotiate content and information provision in an automatic and transparent way. Agents will also contribute in augmenting or modifying the repository ontologies.
- *OAI*: The digital repository can function as *Data Provider* making available its metadata according to the OAI-PMH protocol. Apart from the Consumption layer, the repository may support OAI as *Service Provider* as well: It will receive metadata by other Data Providers (Insertion layer) and make them available for consumption.
- *Z39.50*: The repository can function as a server providing its information, by supporting the Z39.50 client / server communication protocol for information retrieval. Support for Z39.50 will allow remote client systems to access the repository; these systems in turn will make the retrieved information available to their end users. Since Z39.50 is an internationally standardized protocol for information retrieval (ANSI / NISO), its support is of critical importance for the repository interoperability. As in the case of OAI, the repository can support Z39.50 as a client, at the Insertion layer.
- *UDDI*: The digital repository may conform to the UDDI specification, in order to be discoverable, accessible and finally consumable as a Web Service. Although the use of the WSDL language for describing the repository as a Web Service is not mandated by the UDDI specification (version 3), it is recommended as a best practice. Depending on its development policy, the repository can also function as a UDDI Node, and collect Web Services information from other repositories. As a UDDI Node, the repository can be part of a larger UDDI Business Registry (UBR) or even function as the Root Node of a domain-specific registry.

5 CONCLUSIONS AND FUTURE WORK

In this paper we tried to capture the meaning and importance of a digital repository, and provided a series of definitions that are independent of specific content, functionality or implementation needs. We also reviewed a series of key requirements for a general purpose digital repository and argued that they are of crucial importance for any relevant initiative. Based on these requirements, we developed a flexible, scalable and extensible functional model, trying at the same time to maintain a (thin really) equilibrium between abstract functional design and detailed technical dictation of implementation. Our previous discussion has also shown that, in order to achieve elementary functionality, it is not necessary to implement all the interfaces proposed by the model; We are currently following a process of incremental implementation of the various repository's components, deployed on top of an existing open-source repository framework. Future work will be focused on migrating existing systems to the proposed functional model and fine-tuning its components in order to achieve seamless integration of heterogeneous standards and maximum efficiency.

REFERENCES

Harith Alani, Sanghee Kim, David E. Millard, Mark J. Weal, Wendy Hall, Paul H. Lewis and Nigel R. Shadbolt, 2003. Automated Ontology-Based Knowledge Extraction from Web Documents. IEEE Intelligent Systems 18 (1).

ANSI/NISO, 1995. Z39.50-1995: Information Retrieval (Z39.50) Application Service Definition and Protocol Specification. http://lcweb.loc.gov/z3950/agency/markup/markup.html

Michael J. Bass et al, 2002. DSPACE Internal Reference Specification. Technology and Architecture http://www.dspace.org/technology/architecture.pdf

Ronald Bourret, 2003. XML and Databases. http://www.rpbourret.com/xml/XMLDatabases.htm

Consultative Committee for Space Data Systems, 2002. CCSDS.650.0-B-1Q Reference Model for an Open Archival Information System (OAIS). Blue Book. Issue 1. http://wwwclassic.ccsds.org/documents/pdf/CCSDS-650.0-B-1.pdf

Cornell Digital Library Research Group (CDLRG), 2000. Dienst: Overview and Introduction. http://www.cs.cornell.edu/cdlrg/dienst/DienstOverview.htm

Nick Crofts, Ifigenia Dionissiadou, Martin Doerr, Matthew Stiff, 2001. Definition of the CIDOC object-oriented Conceptual Reference Model (version 3.2.1). http://cidoc.ics.forth.gr/docs/cidoc_crm_version_3.2.1.rtf

Digital Imaging Group. Inc, 2000. DIG-35 Metadata Specification for Digital Images, Ver1. http://www.i3a.org/i_dig35.html

Digital Library Federation, 2003. METS-Metadata Encoding and Transmission Standard. http://www.loc.gov/standards/mets/

Dublin Core Metadata Initiative, 2003. The Dublin Core Metadata Element Set Version 1.1: Reference Description. http:///www.dublincore.org/documents/2003/02/04/dces/

European Council, 2002. Resolution on preserving tomorrow's memory – preserving digital content for future generations. Official Journal no 2002/C 162/02. http://europa.eu.int/eur-lex/

D. Florescu and D. Kossmann, 1999. Storing and querying XML data using an RDMBS. IEEE Data Eng. Bull. 22(3): 27-34.

Carl Lagoze, Herbert Van de Sompel, 2001. The Open Archive Initiative. Building a low-barrier interoperability framework. Proceedings of the first ACM/IEEE-CS joint conference on Digital Libraries: 54-62.

Carl Lagoze, Jane Hunter, 2001. The ABC ontology and model. Journal of Digital Information, Special Issue - selected papers from Dublin Core 2001 Conference.

Smith MacKenzie, 2002. DSpace: An Institutional Repository from the MIT Libraries and Hewlett Packard Laboratories. In: M.Agosti, C. Thanos (Eds.): Research and Advanced Technology for Digital Libraries. Lecture Notes in Computer Science 2458: 543-549.

Deborah L.McGuinness, Richard Fikes, James Hendler and Lynn Andrea Stein, 2002. DAML+OIL: An ontology language for the Semantic Web. IEEE Intelligent Systems 17(5).

Moving Picture Experts Group, 2003. MPEG-7 Ver. 8. http://mpeg.telecomitalialab.com/standards/mpeg-7/mpeg-7.htm

Moving Picture Experts Group, 2002. MPEG-21 Overview ver 5. http://mpeg.telecomitalialab.com/standards/mpeg-21/mpeg-21.htm

New Zealand Digital Library, 2004. GreenStone Digital Library Software. http://www.greenstone.org/english/home.html

OASIS UDDI Specifications TC, 2002. UDDI Version 3.0 Published Specification. http://uddi.org/pubs/uddi_v3.htm20

Sandra Payette and Thornton Staples, 2002. The Mellon Fedora Project. In: M.Agosti, C. Thanos (Eds.): Research and Advanced Technology for Digital Libraries. Lecture Notes in Computer Science 2458: 406-421.

Sandra Payette and Carl Lagoze, 1998. Flexible and Extensible Digital Object and Repository architecture. Second European Conference on Research and Advanced Technology for Digital Libraries. Lecture Notes in Computer Science 1513: 41-59.

RLG-OCLC, 2001. Meeting the Needs of Research Resources: Attributes of a Trusted Digital Repository. http://www.rlg.org/longterm/attributes01.pdf

Herbert Van de Sompel, Carl Lagoze, 2002. Notes from the Interoperability Front: A progress Report on the Open Archives Initiative. In: M.Agosti, C. Thanos (Eds.): Research and Advanced Technology for Digital Libraries. Lecture Notes in Computer Science 2458: 144-157.

World Wide Web Consortium, 2003. Web Services Definition Language (WSDL) Version 1.2. W3C Working Draft. http://www.w3.org/ TR/wsdl12/

World Wide Web Consortium, 2003. Web Ontology Language (OWL) Guide Version 1.0. W3C Working Draft. http://www.w3.org/TR/owl-guide/

Z39.50 International Maintenance Agency, 2003. Z39.50 International: Next Generation (ZING). http://www.loc.gov/z3950/agency/zing/zing-home.html