CRiB: A service oriented architecture for digital preservation outsourcing

Miguel Ferreira[1], Ana Alice Baptista[1], José Carlos Ramalho[2]

[1] Department of Information Systems,
University of Minho, Guimarães, Portugal
{mferreira, analice}@dsi.uminho.pt
[2] Department of Informatics,
University of Minho, Braga, Portugal
jcr@di.uminho.pt

Abstract. This paper identifies some of the most prominent issues present in today's digital repository systems, which hinder the long-term preservation of digital materials. In order to address some of those issues, we propose the CRiB system, a service oriented architecture (SOA) supported by Web services' technology, which will enable institutions to outsource part of the functionality that is necessary to carry out effective long-term digital preservation. The proposed system delivers a set of services that client applications will be able to invoke in order to perform complex format migrations; evaluate the outcome of those migrations according to multiple criteria (e.g. data loss and performance); and obtain detailed migration reports for documenting the preservation intervention.

1 Introduction

The number of archives and libraries responsible for managing and safeguarding large collections of digital materials is growing at a startling rate [1]. Several reasons can be outlined which may explain this phenomenon: most information items created today are crafted with the help of digital authoring tools; the physical space required to store those items is nearly insignificant when compared with the requirements for storing conventional analogue-based materials and the possibility of disseminating such content over the Internet entitles these institutions with a whole new branch of business opportunities. Moreover, a large part of such materials can not be adequately represented in conventional analogue media like paper or microfilm (e.g. 3D model, Web page). As a result, classic techniques for preserving information can not be applied in the digital domain.

In addition, digital archiving technology is now affordable to most institutions. Several products have been developed which are both reliable and free of charge (e.g. DSpace [2, 3], Eprints [4], Fedora [5], Greenstone [6]). Most of these products incorporate, off-the-shelf, an assortment of standards developed mostly by the library and archival communities which promote trust and facilitate interoperability between these systems. Among these are the Open Archival Information System Reference

This work was supported by the FCT under the grant SFRH/BD/17334/2004.

Model (OAIS) [7, 8]; standards for describing and structuring information items (e.g. Dublin Core [9], EAD [10], MARC [11], METS [12]); protocols for disseminating metadata (e.g. OAI-PMH [13]) and standards to produce identifiers that are more reliable and persistent than the traditional URL (e.g. CNRI Handle System [14], PURL [15]).

Although current repository software[1] performs a remarkable job at storing, managing and disseminating digital materials, they are not truly capable of assuring the long-term preservation of those materials. The key problem in the design of those systems is that the period of time that the materials are expected to be interpreted is much longer than the lifetime of individual storage media, hardware and software components, as well as the formats in which the information is encoded [16]. As hardware and software turn obsolete, digital materials become prisoners of their own encodings.

In this context, digital preservation is defined as the set of processes and activities that ensure the continued access to information and all kinds of cultural heritage existing in digital formats [17]. A digital object is "(…) an information object, of any type of information or any format, that is expressed in digital form" [18]. Text documents, digital photographs, databases, virtual reality models, Web pages and computer games are just a few examples of digital objects.

This paper introduces a Service Oriented Architecture (SOA) [19] which will enable institutions, as well as individuals, to preserve collections of digital materials without having to go through complex development projects in order to implement all the necessary functionality. The proposed system will deliver a set of services that institutions will be able to invoke in order to convert their digital materials from near obsolete formats to newer encodings that most users will be able to interpret. The proposed system will also be capable of providing suggestions of best suitable preservation alternatives by taking into consideration the individual requirements of each client institution.

This paper is organised as follows: section 2 provides background on the most significant migration-based preservation strategies; section 3 outlines a few of the most prominent issues in today's digital repository software; section 4 describes the proposed system; and section 4 summarises the ongoing research.

2 Format migration as a preservation strategy

Over the last decade, the research community has come up with a considerable number of strategies aiming at solving the problem of digital preservation and technological obsolescence. Among these is migration[2].

Migration consists of a "(…) set of organized tasks designed to achieve the periodic transfer of digital materials from one hardware/software configuration to another or from one generation of computer technology to a subsequent generation." [20]. Contrary to other preservation strategies (e.g. emulation, encapsulation, etc.), migration techniques do not attempt to preserve digital objects in their original

[1] Also known as Content Management Systems (CMS).
[2] Also known as conversion.

formats. Alternatively, they intentionally transform objects from near obsolete formats into up-to-date encodings that most users are able to interpret on their personal computers.

The main disadvantage of this approach is that when a digital object is migrated, there is a high probability that of some of its inner properties will not be correctly transferred to the target format (i.e. some data loss is expected to take place). The reason for this is twofold: there may be structural incompatibilities between the source and the target formats or the converter may be faulty and incapable of performing its tasks appropriately. Nevertheless, migration is by far the most widely used preservation strategy and the only one that has actually worked to date [21].

The most advanced endeavours in the field of migration are based on networks of conversion services [22-26]. In such strategies, a set of well known protocols, such as the ones associated with Web Services technology [27], are used to support the discovery and invocation of procedures capable of carrying out format migrations.

This type of migration entails several advantages over more traditional solutions: the use of Web Services hides the complexity of the conversion software that is being used underneath and promotes interoperability by cloaking the peculiarities of its supporting platform; the development of redundant services insures that the network remains functional during situations of partial break down, while at the same time, facilitates distribution of the workload; ultimately, the possibility of having multiple migration paths enables this solution to cope with the gradual disappearing of converters. In addition, this approach is compatible with several variants of migration, such as migration on-request [28] and normalisation [18, 21, 29-33].

3. Missing elements in today's digital repository systems

Although advances in digital repository software have come a long way there are still a reasonable amount of issues regarding the long-term preservation of digital materials which demand full attention from the R&D community. Some of these issues are outlined next.

Limited preservation functionality. Digital repositories perform an outstanding job ingesting, storing and disseminating materials over the Internet. However, most digital repository systems are not yet capable of effectively performing long-term preservation of those materials. A few repository systems already incorporate some preservation functionality, but in most cases this is restricted to the definition of ingest policies such as the limitation of accepted formats or the normalisation of received objects to formats more suitable for preservation [32-34]. However, when formats in the data store become obsolete, an assortment of more sophisticated techniques are usually required. In such cases, it is necessary to develop specialized migration or emulation tools which often require significant human intervention in order to be fully exploited [35].

Authenticity. Another important issue has to do with authenticity. Authenticity is defined as the quality of a digital object being what it purports to be [29, 36-41]. In a preservation environment, where tampering is admissible as a form of preservation (e.g. when migration strategies are in place), the issue of authenticity assumes even greater importance. Although most requirements for ensuring authenticity and trustworthiness in a preservation environment have already been identified [37, 40-43], most digital repository systems do not yet implement all the necessary functionality to support them. Technical approaches for assuring authenticity involve maintaining, within the archival system, detailed information about the provenance of the digital object, the context in which it was created, the identification of all agents in the chain of custody and the continuous logging of activities that take place within archive which may affect the preserved object in a significant way, e.g. a migration, updates on the metadata, object accesses, etc. [36, 42]. The management of all these metadata combined with the preservation of the bit-wise original object constitute a handful of mechanisms for ensuring authenticity [41].

Cost management. Effective digital preservation is not cheap. It requires investment in a robust technological infrastructure capable of coping with preservation issues at all possible levels failure. At the physical level, storage media is expected to degrade with time, hardware is subject to failures and the network infrastructure may loose its connectivity. Replication of these components reduces the risk of a single point of failure [16]. At the logical level, software components are also expected to become obsolete as hardware and operating systems are replaced by newer versions. Although the bits in which data are encoded may still be accessible, if the necessary decoding software is no longer available the information is bound to become prisoner of its own encodings [16]. There are also concerns at the social level. Operators are expected to commit errors. Functionality which limits the number of irrecoverable human errors is also an important requirement for digital repository software [16].

Furthermore, preservation requires planning and pre-emptive decision making. Monitoring the external environment for new standards or product updates is quintessential to avoid unexpected changes in the technology market [44]. Additionally, preservation activities should be automated as much as possible in order to reduce costs and liberate human resources. Examples of such functionality could be: an autonomous service that monitors the Web and detects changes in the general public behaviour towards a given format or a repository system that automatically updates formats in its data store whenever a new version of format is released.

4 The CRiB system

This paper proposes a service oriented architecture (SOA) [19] to address the array of issues outlined in the previous section that are currently hindering digital repository software from carrying out effective digital preservation. The proposed system, recently named CRiB[3], will provide a set of Web methods that will entitle client applications to carry out the following activities:

a) Perform complex format migrations;

[3] CRiB stands for Conversion and Recommendation of Digital Object Formats.

b) Determine the amount of data loss resulting from those migrations;
c) Document preservation interventions;
d) Obtain suggestions of migration alternatives to adequately preserve collections of digital objects.

The general architecture of the proposed system is depicted in Fig. 1. The application layer illustrates client applications that may take advantage of the services provided by the CRiB system. Examples of such applications are: digital repository systems (e.g. DSpace, Fedora or Eprints) and custom applications developed by individual users.

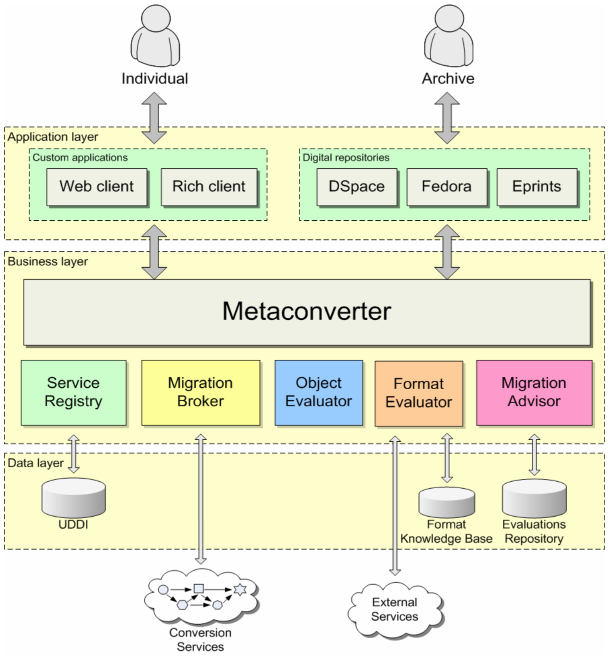

Fig. 1 - Overview of the CRiB architecture.

The middle layer illustrates the whole set of components that constitute the actual CRiB system. The Metaconverter acts as the mediator between client applications and the rest of the system. Furthermore, it is responsible for generating and coordinating all the messages within the system that are necessary to adequately carry out its activities. The Service Registry comprises information to support the discovery of conversion services. The Migration Broker handles invocations to local and remote conversion services. The Object Evaluator is responsible for detecting any loss of information that may take place during the migration process. The Format Evaluator provides information about the current status of digital formats. Finally, the Migration Advisor combines all that information to generate suggestions of migration alternatives that maximise users' satisfaction.

The data layer outlines the sources of information that support the CRiB system (e.g. UDDI server, Format Knowledge Base, Evaluations' Repository). Some external sources of information may also be used, such as remote conversion services or live sources of information, such as Google's own set of Web services.

A possible use scenario where the CRiB system may play an important role is described next:

A large company decides that all technical reports produced in the course of its activity should be available to every employee at the distance of a mouse-click. For that reason, a person was hired to set up and maintain a digital repository system for safekeeping and providing access to those reports. At that time, all existing technical reports were produced using Microsoft Word 95 but, as time went by, new versions of Word begun to be exploited within the company. At the same time, some employees, more fond of the whole open-source movement, preferred to use OpenOffice to produce their technical reports. As a consequence, the number of formats available in the repository became so heterogeneous that barely anyone in the company was able to read a single project's collection of technical reports without having to install additional software in their personal computers. Meanwhile, Microsoft announces that the next version of Office will not support Word 95.

The person responsible for managing the digital repository decides that something had to be done, so he launches an application that is capable of communicating with the CRiB system. First, he informs the CRiB of the format that he soon expects to become obsolete – i.e. the Word 95. The CRiB system responds with a list of criteria that it is capable of evaluating for assessing the quality of the preservation intervention. The manager is expected to weight, according to what he feels is more important for the company, all the criteria that the CRiB has provided. Among these criteria are items such as: textual content preservation, layout, cost, conversion throughput (Kb/s), etc.

The manager decides that the textual content and the page layout are very important items and should be preserved at all cost. After informing the CRiB of the requirements, the system responds with a list of possible formats to which Word 95 files could be converted to. Among these are PDF, Word 2003 and OpenOffice 2, being PDF the top choice suggested by the system.

The manager decides to follow the system's suggestion and requests a list of possible migration services. Based on the weights previously assigned by the manager, the system suggests a rather expensive conversion service but which is able to perform very high quality migrations from Word 95 to PDF. The manager sends its Word 95 files to the CRiB system and the migration process is initiated.

After each conversion the manager receives a PDF version of the technical report and a metadata record that describes the preservation intervention. In that metadata record is included information such as: a description of the conversion services involved in the migration, the date and time of the conversion and other interesting information such as the list of properties of the original Word file that were not properly preserved when converted to PDF. He then uses these reports do document the preservation intervention and validate the conversion.

After running the same procedure for all other formats available in the repository, the manager realises that PDF is often suggested as the preferred format. He then decides to write a ingest policy stating that all technical reports should be converted to PDF before entering the repository.

The following sections describe the inner workings of the components that constitute the CRiB system.

4.1 Service Registry

The Service Registry is responsible for managing all the metadata, necessary to support the discovery of conversion services. Additionally it facilitates the calculation of composite migrations (i.e. conversions that involve more than one migration service). The metadata elements used within this component are based on the Universal Description, Discovery and Integration (UDDI) standard [45]. The UDDI outlines three basic entities to describe Web services: a service entity that contains information about the service it self (e.g. name, description, etc.); a business entity to describe the producer/developer of the service (e.g. name, description and contacts); and a set of binding templates which comprise information about how the service may be invoked by a client application (e.g. URL).

In order to facilitate the identification of composite migrations, two new metadata elements were introduced, i.e. the source and target formats. Possible values for these elements are obtained from a controlled vocabulary – the PRONOM Registry, an initiative from the National Archives of the UK which aims at building a registry of information about every existing file format [46].

A third metadata element is included in the service description: the cost. The cost refers to the fee that each client is expected to pay in order to use the conversion service. The rationale for this is to stimulate the development of new conversion services which may be published and sold on CRiB platform. Although this economic model is probably too simplistic to be put into practice it constitutes an important tool for assessing how preservation costs may influence decisions in favour, or against, certain migration alternatives (see Migration Advisor).

4.2 Migration Broker

The Migration Broker is responsible for generating and coordinating all SOAP messages, necessary to correctly perform object migrations. In practice, this component will make sure that composite conversions are carried out atomically from the CRiB's point of view.

Additionally, this component is responsible for measuring the performance of each migration service. Performance will be measured according to the following criteria: availability [47], stability [48], scalability [49-51] and throughput [52]. Afterwards, these evaluations will be forwarded to the Migration Advisor in order to be archived. Later on, they will serve as a basis for ranking migration alternatives (see Migration Advisor).

4.3 Object Evaluator

The Object Evaluator is in charge of judging the quality of the migration's outcome. It accomplishes this by comparing the objects submitted to migration with its converted counterparts. Evaluations will be performed according to multiple criteria. These criteria, also known as significant properties, constitute the set of attributes of an object that should be maintained intact during a preservation intervention [53]. They constitute the range of attributes that characterise an object as a unique intellectual entity, independently of the encoding in which it is being represented. The Bible for example, may exist in many different formats and media, e.g. ASCII text, PDF, written on paper or carved on stone, and still be regarded as the Holy Bible. Considering text documents as an example, some relevant significant properties could be: the textual content, the page size, the number of pages, the graphical layout, the number of characters, the order of those characters, the font type and size, etc.

Work is underway as to produce a general taxonomy of significant properties for various classes of digital objects. A good foundation for this, is the work developed by Rauch and Rauber [54-56]. They have conducted a series of workshops from which they have devised a practical list of significant properties for text documents, audio and video objects.

The evaluations performed by the Object Evaluator will be returned to the client user and stored in the Evaluations Repository. The report sent to the client userwill follow the structure of the Event Entity included in the PREMIS Data Dictionary [41], i.e. a dictionary of metadata elements specifically designed for documenting preservation activities within an archival environment. The Event Entity includes elements for describing the type of event (e.g. Migration), the date and time of its occurrence, the agent that carried out the event and detailed information about the outcome of the event (e.g. the amount of data loss that resulted from the migration).

4.4 Format Evaluator

The Format Evaluator provides information about the current status of digital formats. This information will enable the Migration Advisor to determine the formats to which a given object should be converted to. Examples of such information are: the format's market share, its level of support, if it has an open specification, etc. The Format Evaluator will be able to determine this information by questioning the Format Knowledge Base, i.e. a data store of known facts about digital formats; or resorting to external sources of information such as the PRONOM Registry or Google's Web services.

4.5 Migration Advisor

The Migration Advisor is responsible for generating suggestions of migration alternatives. It accomplishes this by confronting the preservation requirements outlined by the user, with the accumulated knowledge about the performance of each individual conversion service. The quality of each migration service will be measured according to multiple criteria: expected data loss, current status of involved formats and computational performance. Using this information, the Migration Advisor will be able to rank all possible alternatives and produce an appropriate suggestion for the invoking client.

	Evaluation criteria	WS1	WS2	Weights	WS1	WS2		Weights		WS1	WS2
Process	Availability	95%	90%	40%	5	5	x	0.4	=	2	2
Process	Stability	80%	90%	30%	5	5		0.3		1.5	1.5
Process	Scalability	1	1.5	0%	1	5		0		0	0
Process	Throughput	1 Kb/s	2 Kb/s	20%	1	5		0.2		0.2	1
Process	Cost	0.01 €	0.05 €	10%	5	1		0.1		0.5	0.1
Object	Number of chars	70%	95%	40%	4	5		0.4		1.6	2
Object	Ordering of chars	60%	85%	40%	4	5		0.4		1.6	2
Object	Page layout	80%	20%	20%	5	2		0.2		1	0.4
Format	Market share	30%	70%	30%	2	4		0.3		0.6	1.2
Format	Support level	5	4	30%	5	4		0.3		1.5	1.2
Format	Open standard	yes	no	40%	5	1		0.4		2	0.4
										12.5 ★	11.8
	(1) Average evaluations per service			(2) User assigned weights of importance	(3) Normalised evaluations [1-5]			(4) Ranking process		(5) Overall rating	

Fig. 2 – Steps involved in the process of ranking migration alternatives. WS1 and WS2 represent two different conversion services.

The ranking process of the alternatives is based on the work developed by Rauch and Rauber [54-57]. They have built a framework which allows users to evaluate, compare and select preservation alternatives according to their individual requirements [54-56]. The process of ranking the alternatives is composed by the following steps (Fig. 2):

1. It begins with the assembly of the evaluation criteria which will be used for assessing the quality of each migration alternative. This set of criteria is regarded as the evaluation taxonomy. All of these criteria fall into one of the following top level categories: object characteristics (handled by the Object Evaluator), format characteristics (handled by the Format Evaluator) and process characteristics (handled by the Migration Broker). For each of the evaluation criteria present in the taxonomy, the system will devise an average performance rating based on previous logged evaluations (Fig. 2, step 1).
2. In order to obtain an adequate suggestion, the user will have to inform the system about which format he/she wishes to migrate from. For that format, the Migration Advisor will return the collection of criteria that it is capable of evaluating (i.e. the evaluation taxonomy). The user manifests its preservation preferences by assigning weights to all the criteria included in the collection. For example, a user might feel that preserving both textual content and graphic layout of a document is fundamental for the success of the preservation intervention. He/she might not care about how much it will cost in terms of money or execution time as long as those characteristics are well preserved. Another user, with a more limited budged, might give up on the graphic layout in order to reduce preservation costs (Fig. 2, step 2).

3. After receiving the weighted list of criteria the Advisor will normalise the average performance of each alternative to a scale of 0 to 5. Different scales are also possible, e.g. 0-100. The important factor here is that all evaluations are converted to comparable values (Fig. 2, step 3).
4. The ranking process works by multiplying the user assigned weights by the normalised values of each evaluation criterion (Fig. 2, step 4).
5. In the end, an overall rating is calculated for each of the migration alternatives (WS1 and WS2 in Fig. 2) by summing up all individual ratings (Fig. 2, step 5).

5 Conclusions

This paper begins by identifying some of the issues that are currently hindering digital repository systems from performing effective long-term preservation of digital materials. To address those issues, we propose a system supported by Web services' technology which enables client institutions to carry out the following activities:

1. Convert digital objects from near obsolete formats to up-to-date encodings that most users will be able to interpret;
2. Evaluate the outcome of a migration by comparing the original digital object with its converted counterpart and identifying the significant properties that have not been adequately preserved;
3. Obtain migration reports in appropriate forms for inclusion in the preservation metadata of the migrated object;
4. Request suggestions of best suitable migration alternatives taking into consideration the preservation requirements of the client institution;

The proposed system addresses the problem of limited preservation functionality in today's digital repository systems by enabling any client application capable of invoking Web services to perform complex format migrations. It addresses authenticity by delivering detailed migration reports that fully document the preservation intervention. It reduces preservation costs by providing suggestion mechanisms that automate preservation activities such as planning while at the same time, facilitating other preservation processes such as documentation and process evaluation.

Parallel contributions are also expected from this research. Developers will have the possibility of publishing and selling their conversion applications. Conversion software published on the CRiB platform will be automatically compared and benchmarked according to multiple quality criteria.

Additionally, sharing experiences of practical use of recently created metadata schemas such as the PREMIS Data Dictionary [41] may contribute to increase its adoption, while the same time, push creators to improve future versions and accelerate the development of XML bindings.

The project presented in this paper is now entering its first stages of development. A group of students is currently working on the development of several migration services for text documents and digital images. Future work will be centred in the construction of a complete prototype of the proposed architecture for proof-of-concept.

References

1. Brody, T., *Analysis*, from http://archives.eprints.org/index.php?action=analysis.
2. Smith, M. *DSpace: An Institutional Repository from the MIT Libraries and Hewlett Packard Laboratories*. in *6th European Conference on Research and Advanced Technology for Digital Libraries*. 2002. Roma, Italy: Springer-Verlag.
3. Hewlett-Packard Company and MIT Libraries, *DSpace Web site*, from http://dspace.org.
4. *EPrints Web site*, from http://www.eprints.org/.
5. University of Virginia and Cornell University, *Fedora Web site*, from http://www.fedora.info/.
6. New Zealand Digital Library Project, *Greenstone Web site*, from http://www.greenstone.org.
7. Consultative Committee for Space Data Systems, *Reference Model for an Open Archival Information System (OAIS) - Blue Book*. 2002, Washington: National Aeronautics and Space Administration.
8. Lavoie, B.F., *The Open Archival Information System Reference Model: Introductory Guide*. 2004, Digital Preservation Coalition: Dublin, USA.
9. Dublin Core Metadata Initiative, *Dublin Core Metadata Initiative*, from http://dublincore.org/.
10. Library of Congress, *EAD - Encoded Archival Description*, Library of Congress, from http://www.loc.gov/ead/.
11. Library of Congress, *MARC Standards*, from http://www.loc.gov/marc/.
12. Library of Congress, *METS - Metadata Encoding & Transmission Standard*, from http://www.loc.gov/standards/mets/.
13. Open Archives Initiative, *The Open Archives Initiative Protocol for Metadata Harvesting*. 2002, from http://www.openarchives.org/OAI/openarchivesprotocol.html.
14. Corporation for National Research Initiatives, *CNRI Handle System*, Corporation for National Research Initiatives, from http://www.handle.net/.
15. Online Computer Library Center, *PURL - Persistent Uniform Resource Locator*, from http://purl.oclc.org/.
16. Rosenthal, D.S.H., et al., *Requirements for Digital Preservation Systems.* D-Lib Magazine, 2005. **11**(11).
17. Webb, C., *Guidelines for the Preservation of Digital Heritage*. 2003, United Nations Educational Scientific and Cultural Organization - Information Society Division.
18. Thibodeau, K. *Overview of Technological Approaches to Digital Preservation and Challenges in Coming Years*. in *The State of Digital Preservation: An International Perspective*. 2002. Washington D.C.: Documentation Abstracts, Inc. - Institutes for Information Science.
19. OASIS SOA Reference Model TC, *OASIS Reference Model for Service Oriented Architectures (Working Draft 10)*. 2005: OASIS.
20. Task Force on Archiving of Digital Information, Commission on Preservation and Access, and Research Libraries Group, *Preserving digital information: report of the Task Force on Archiving of Digital Information*. 1996, Washington, D.C.: Commission on Preservation and Access. 59.
21. Lee, K.-H., et al., *The State of the Art and Practice in Digital Preservation.* Journal of Research of the National Institute of Standards and Technology, 2002. **107**(1): p. 93-106.
22. Ockerbloom, J.M., *Mediating Among Diverse Data Formats*, in *School of Computer Science*. 1998, Carnegie Mellon University: Pittsburg. p. 164, from
23. Walker, F.L. and G.R. Thoma. *A Web-Based Paradigm for File Migration*. in *IS&T's 2004 Archiving Conference*. 2004. San Antonio, Texas, USA.
24. Hunter, J. and S. Choudhury. *A Semi-Automated Digital Preservation System based on Semantic Web Services*. in *Joint ACM/IEEE Conference on Digital Libraries (JCDL'04)*. 2004: ACM.
25. Hunter, J. and S. Choudhury, *Preservation webservices Architecture for Newmedia and Interactive Collections (PANIC)*. 2005, from http://metadata.net/newmedia/.
26. Wing, J.M. and J. Ockerbloom, *Respectful Type Converters.* IEEE Transactions on Software Engeneering, 2000. **26**(7).
27. Graham, S., et al., *Building Web Services with Java: Making Sense of XML, SOAP, WSDL and UDDI.* 2002: Sams Publishing.
28. Mellor, P., P. Wheatley, and D.M. Sergeant. *Migration on Request, a Practical Technique for Preservation*. in *ECDL '02: 6th European Conference on Research and Advanced Technology for Digital Libraries*. 2002. London, UK: Springer-Verlag.
29. Hofman, H. *How to keep digital records understandable and usable through time?* in *Long-Term Preservation of Electronic Records*. 2001. Paris, France.

30. Heslop, H., S. Davis, and A. Wilson, *An Approach to the Preservation of Digital Records*. 2002, National Archives of Australia: Camberra, Australia, from
31. Howel, A.G., *Preserving Digital Information: Challenges and Solutions*. 2004, Cooperative Action by Victorian Academic Libraries, Victorian university libraries, State Library of Victoria.
32. Hodge, G. and E. Frangakis, *Digital Preservation and Permanent Access to Scientific Information: The State of the Practice*. 2004, International Council for Scientific and Technical Information & CENDI.
33. Hedstrom, M., *Digital Preservation: A time bomb for digital libraries.* Computers and the Humanities, 1998. **31**: p. 189-202.
34. Han, Y., *Digital content management: the search for a content management system.* Library Hi Tech, 2004. **22**(4): p. 355-365.
35. Ross, S. and M. Hedstrom, *Preservation research and sustainable digital libraries.* Springer-Verlag, 2005.
36. Lavoie, B. and R. Gartner, *Technology Watch Report - Preservation Metadata*. 2005, Online Computer Library Center Inc., Oxford University Library Services and Digital Preservation Coalition.
37. MacNeil, H., et al., *Authenticity Task Force Report*. 2001, InterPARES Project: Vancouver, Canada.
38. Millar, L., *Authenticity of electronic records: a report prepared for UNESCO and the International Council on Archives*. 2004, International Council on Archives: London, UK.
39. Lynch, C., *Authenticity and Integrity in the Digital Environment: An Exploratory Analysis of the Central Role of Trust*, in *Authenticity in a Digital Environment*. 2000, Council on Library and Information Resources: Washington, DC.
40. Authenticity Task Force, *Requirements for Assessing and Maintaining the Authenticity of Electronic Records*. 2002, InterPARES Project: Vancouver, Canada.
41. PREMIS Working Group, *Data dictionary for preservation metadata: final report of the PREMIS Working Group*. 2005, OCLC Online Computer Library Center & Research Libraries Group: Dublin, Ohio, USA.
42. Diessen, R.J.v. and T.v.d. Werf-Davelaar, *Authenticity in a digital environment*. 2002, Koninklijke Bibliotheek & IBM: Amesterdam, The Netherlands.
43. Cullen, C.T., et al., *Authenticity in a Digital Environment*. 2000, Washington, DC: Council on Library and Information Resources.
44. Anderson, C., *Digital Preservation: Will Your Files Stand the Test of Time?* Library Hi Tech News, 2005. **6**: p. 9-10.
45. OASIS, *Universal Description, Discovery and Integration (UDDI)*. 2005, from http://www.uddi.org/.
46. Darlington, J., *PRONOM - A Practical Online Compendium of File Formats.* RLG DigiNews, 2003. **7**(5).
47. Wikipedia, *Availability*. 2005, from http://en.wikipedia.org/wiki/Availability.
48. Wikipedia, *Stability*. 2005, from http://en.wikipedia.org/wiki/Stability.
49. Deters, R. *Scalability and Multi-Agent Systems*. in *Workshop "Infrastructure for Scalable Multi-Agent Systems" at Agents 2001*. 2001. Montreal.
50. Luke, E.A. *Defining and measuring scalability*. in *Scalable Parallel Libraries Conference*. 1993. Mississippi State, USA.
51. Wikipedia, *Scalability*. 2005, from http://en.wikipedia.org/wiki/Scalability.
52. Wikipedia, *Throughput*. 2005, from http://en.wikipedia.org/wiki/Throughput.
53. Rusbridge, A., *Migration on Request*. 2003, University of Edinburgh - Division of Informatics.
54. Rauch, C. and A. Rauber. *Preserving Digital Media: Towards a Preservation Solution Evaluation Metric*. in *International Conference on Asian Digital Libraries*. 2004. Shanghai, China: Springer.
55. Rauch, C., et al. *Evaluating preservation strategies for audio and video files*. in *DELOS Digital Repositories Workshop*. 2005. Heraklion, Crete.
56. Rauch, C., *Preserving Digital Entities - A Framework for Choosing and Testing Preservation Strategies*, in *Institute for Software Technology and Interactive Systems*. 2004, Vienna University of Technology: Vienna, from
57. Rauch, C., et al., *A Framework for Documenting the Behaviour and Funcionality of Digital Objects and Preservation Strategies*. 2005, DELOS Network of Excellence: Glasgow.