Digital Preservation and Workflow Process

Su-Shing Chen

Department of Computer and Information Science and Engineering,
University of Florida, Gainesville, Florida 32611, USA
suchen@cise.ufl.edu

Abstract. The digital societies of E-government, E-learning, and E-business have grown by leaps and bounds worldwide during the last several years. While we have invested significant time and effort to create and maintain those workflow processes, we do not have the ability to make digital objects generated by the processes all available across generations of information technology, making it accessible with future technology and enabling people to determine whether it is authentic and reliable. This is a very serious problem for which no complete solutions have been devised yet. This paper discusses three important factors - archival stability, organizational process, and technology continuity – for digital preservation to succeed, and describes a general framework of digital libraries (or the life cycle of information) to address this important problem so that we may find reasonable ways to preserve digital objects that can be analyzed and evaluated in quantitative measures and incremental manners.

1 Introduction

Although digital societies have emerged and digital communities have formed of E-government, E-learning, and E-business, we are still facing a fundamental paradox in digital preservation: On the one hand, we want to maintain digital information intact as it was created, but on the other we want it to be accessible in a dynamic context of use [2]. Why the rapid progress being made in information technology today to create, capture, process and communicate information in the digital form threatens the accessibility in the near future? This is because of two reasons: First digital information has mushroomed, and secondly hardware and software products are being upgraded and replaced roughly every eighteen months. Companies in the information technology sector have reported that the majority of products and services they offer did not exist five years ago. For cost-effectiveness, we have to change hardware and software products from generation to generation. The digital library community, which is a part of digital societies, needs to pay attention to digital preservation.

The digital environment has fundamentally changed the concept of preservation. Traditionally preservation means keeping things unchanged. For example, we can still read the Rosetta Stone of Ptolemy V in hieroglyphic, demotic, and Greek today. However if we could succeed in holding on to digital information without any change, the information would become increasingly harder, if not impossible, to access. Even

Z. Chen et al. (Eds.): ICADL 2004, LNCS 3334, pp. 61–72, 2004.

if a physical medium could hold digital information intact, the formats in which information is recorded digitally do change and the hardware and software to retrieve the information from the medium often become obsolete. In summary, we can not predict the future of information technology, and thus can not plan our digital preservation well in the present.

In [4], we have emphasized how important it is to preserve digital objects as the essential information memory for our society, and introduced the life cycle of information for addressing the requirements and strategies of preserving digital objects. By establishing the long-term requirements and strategies, we can stabilize the development of digital objects and archives/libraries. In this paper, we argue that *archival stability* must be coupled with *organizational process*, and *technology continuity* in digital preservation, and propose a new approach beyond traditional digital preservation that these factors are interrelated and must be considered together. We introduce the new technical concept of "dynamic" objects which are digital objects associated with workflow processes, and the idea of process repositories (to preserve workflow processes) in addition to the traditional object libraries/archives. Therefore librarians/archivists cannot achieve *archival stability* without the coupling with *organizational workflow processes* and *technology implementations*. Likewise organizations and technology companies must seek librarians/archivists to participate in their planning.

2 A New Framework for Digital Preservation

Digital societies are concerned with the participants - human citizens, software agents, and robots (e.g., ATM and vending machines), information (knowledge bases, databases, and repositories of records), and workflow processes (E-government, E-business, E-learning and etc.). It has become evident that objects must be "dynamic", that is they are not only traditional objects but also associated with appropriate workflow processes. Thus not only objects must be preserved, so are their workflow processes. The traditional archival requirements alone may not work well, because workflow processes are dynamic and subject to change due to administrative and legislative changes.

We believe that this new framework will need the research, design, and implementation of IT technologies – secure networking, distributed computing, information fusion, dynamic records, process repositories, and intelligent record archives. These technologies are necessary for the economic vibrant and civil digital societies. In this paper, we can only introduce briefly some concepts and argue for their importance. The technologies will be developed in subsequent work. The framework consists of five components:

(1) A digital society has a governance model with the organizational structure that consists of a role hierarchy of participants carrying out various collaborative workflow processes (e.g., accessing and processing information and other activities). The maintaining and updating of role hierarchy databases assures the democracy of the society [26].

(2) *An object/process management server that manages workflow processes and provides core workflow services and supporting services (e.g., ontology services synchronize semantics from heterogeneous dynamic objects into a consistent dynamic object for relevant processes. It has a user interface that visualizes the logic of the workflow and dependencies of dynamic objects and their fusion. It associates each dynamic object with the associated workflow process.*

(3) *An archive server of societal databases that provides reliable access for participants to all dynamic objects based on the processes and the role hierarchy. Here we are concerned about the preservation of a digital society's memory. In a digital society, digital objects proliferate and propagate. How to assure that digital objects are appraised, authenticated, and preserved in a societal memory for the long-term? We must develop a digital preservation management model for a digital society, by examining archival stability, organizational process, and technical continuity of the society.*

(4) *Information (objects) fusion and workflow process integration have broad impacts on digital societies of E-government, E-learning and E-business. Practical case studies must be made to study the dynamics and to understand the advantages and disadvantages to societies. Case studies should not be carried out by librarians/archivists alone. A multidisciplinary team of social scientists is needed to study the social impacts, process models, and object keeping.*

(5) *The development of consistent hardware, software and applications sustaining the technology continuity is the responsibility of the IT industry. The IT industry must realize that technology continuity means good business. Although disruptive technologies are the driving force of IT advancement, the technology continuity makes good business sense by keeping customers in the long-term. The industry norm should not be chaotic new products and services that win competition in the IT industry. Recently governments around the world have introduced funding opportunities to the IT industry that will influence this outcome (e.g., [18-20]). The industry has also developed future visions that potentially include digital preservation (e.g., [17]).*

A general digital society can be modeled as workflow systems on the Internet. This digital society model supports the five components of the new framework. Such model consists of three major parts: user interfaces, workflow dynamics and societal databases. Users interface through the use of portals and clients to the web-based system. The workflow dynamics is supported by a collection of services, including the core workflow processes and affiliated digital library services: ontology, administration, evaluation, discussion and visualization. In our proposed framework, the core workflow processes access the workflow template depositories and the subject role hierarchy databases, which represent the organizational structure of workflow participants. Workflow information sources are in the societal databases, whose objects are indexed for multiple contexts and services to generate dynamic objects.

One of the most important issues of digital preservation is to know what to preserve and how to preserve? The new framework places emphasis first on the *organizational process* to preservation and clarifies the overall picture of digital preservation. The responsibility of the holding organizations - government offices, companies, hospitals, and institutions - will affect the ultimate outcome. In addition to *organizational process*, the information science community (e.g., librarians and archivists) must provide *archival stability*, and the computer industry must develop *technical continuity* for *archival stability* in synchrony. For this purpose, we will formulate the digital preservation problem within a life cycle of information. The life cycle of information spans acquisition, preservation, collection, indexing, accessing, and utilization in a dynamic manner [3]. If preservation is missed, then the life cycle is broken and will be disrupted. *Thus organizations should design their preservation of digital libraries/ archives seamlessly in the life cycle of information so that technology can provide the necessary continuity.*

3 Organizational Process

Different organizations have very different requirements and implementations of their archives/libraries, which depend much on the nature of organizations. For example, hospitals will maintain their patient records, school systems their student records, and companies their financial records. In general, there is no standard on whether organizations should preserve, how they preserve and what they preserve? In the more focused context of digital societies, organizations' process in relevance to digital preservation must be evident for their sustainability. This section examines various issues of this kind of processes. One must factor this process as an archival variable into the future organizational management equation. Digital preservation will assume a variety of storage and preservation functions. Traditionally, the preserved objects have been in the forms of books, monographs, reports, maps, photographs, analog sound tracks and films, which are readable, listenable and viewable directly by humans with the aid of magnification, scanning, playing and projection devices. The preservation of physical and analog media has to ensure long-term stability and accessibility. The preservation of digital objects takes on a somewhat different direction, because the technology advances so rapidly that hardware and software products are being upgraded and replaced constantly. Companies in the information technology sector report that the majority of their products and services they offer did not exist 5 years ago. Simultaneously, the explosive growth of information in digital forms has posed a severe challenge for organizations and their information providers because the digital information can be easily lost or corrupted. The pace of technology evolution is further causing severe pressure on the ability of existing data structures or formats to represent information for the organizations in the future. The supporting information necessary to preserve the digital information is available or only available at the time when the original digital information is produced. Usually after the information is produced, the running software may be updated and its version may be changed. Organizations must start to preserve in long-term, otherwise their information will be lost forever.

There are several functions of organizations in preservation. Except some national archives and research groups, these functions are generally ignored by the IT people. We emphasize three of them here. The organizations must monitor community needs, interact with consumers and producers to track changes in their service requirements and available product technologies. Such requirements might include data formats, media choices, software packages, computing platforms, and mechanisms for communicating with the digital libraries/archives. This first function may be accomplished via surveys, a periodic formal review process, community workshops where feedback is solicited, or by individual interactions. It provides reports, requirements alerts and emerging standards for developing future preservation strategies and standards. It sends preservation requirements to the digital library/archive developers and managers. The second function is the responsibility of tracking emerging digital technologies, information standards and computing platforms (i.e., hardware and software) to identify technologies which could cause obsolescence in the archiving computing environment and could prevent access to some of the archives' current holdings. This function may contain a prototyping capability for better evaluation of emerging technologies and receive prototype requests from the digital archive developers. This function is also responsible for developing and recommending strategies and standards to enable the digital archives to better anticipate future changes in the community service requirements or technology trends that would require migration of some current holdings or new acquisitions. The third function approves standards and migration goals from the digital library/archive managers. The standards include format standards, metadata standards and documentation standards. It applies these standards to preservation requirements. The migration goals received by this function involve transformations of the preservation package, including transformations of the content to avoid loss of access due to technology obsolescence. The response to the migration goals may involve the development of new preservation designs, prototype software, test plans, community review plans and implementation plans.

Facing the ever-increasing cost of preservation, digital libraries/archives need sound policies and strategies to preserve the essential information in the long-term. To develop policies and strategies digital libraries/archives need a generally accepted framework or a life cycle of information. A life cycle of information is not only for preservation and access, but rather for the full business model of digital libraries/archives. In the following we discuss the important aspect of the organizational structure – workflow process in general. The specific life cycle of information will be discussed in the next section.

A workflow process is the computerized representation of a business process. It specifies the various activities of a business process that have to be executed in some order, the flow of data between activities and the multiple collaborating agents that execute activities to carry out a common objective. A workflow management system is a software system for defining, instantiating and executing workflows and is currently the leading technology for supporting business processes (e.g., financial markets, banks, retailing stores, transportation and others).

Existing workflow solutions focus on the capability of representing project information and information exchange between applications. However, the ever-changing nature of organizations requires solutions equipped with facilities that are able to treat information (e.g., records) as a dynamic entity. The information changes either because of the normal progress of the business from preplanned activities, or because of events, which occur due to uncertainty. In both cases, they are the results of the life cycle of information!

We believe that the design of a flexible and expressive workflow system can and should be tackled through an interoperable and dynamic information process model. This architecture allows participants and information objects flow in a flexible workflow process architecture, where data and service providers will interact through the Internet in the workflow system. For such system architecture, there is an open harvesting protocol of metadata of data providers by service providers proposed by the Open Archives Initiative (OAI) [23-25]. By exposing a sufficient amount of metadata, data providers openly advertise their content for retrieval and usage by other data providers and service providers. This open architecture must certainly be protected by security mechanisms. In this paper, we will not discuss this issue. In [26], we have focused on this access control issue in the framework of life cycle of information.

The database system is further interfaced to the workflow system through certain service protocols (depending on the functionality of organizations). The workflow engine executes and manages workflow processes with the underlying information model. Applications (e.g., digital society services) reside in the domain-specific middleware layer which is also not discussed in this paper.

Now some discussions on the workflow process repository. The flexibility of workflow architecture is accomplished by the construction of basic workflow building blocks as directed graphs, and indexing of basic workflow processes into a repository from which several processes can be composed into more complex processes, as graph rewriting. During workflow enactment, a workflow can be modified due to on-line constraints and subsequently ingested into the repository as a new entry for future use. The idea of treating workflows as an archival component greatly improves the preservation principle in the long-term. Apparently the design of a workflow system for preservation needs much further research. Our effort has been to investigate the applicability of formal models such as graph-rewriting, Petri-net, and extended transaction for the flexible and dynamic configuration of workflows. More detailed results will be reported elsewhere. Such an interoperable and dynamic information model consists of networked application services with distributed users, objects, workflow processes, and their repositories. This kind of application services can be implemented based on the dynamic object model which originates from our research (described in the next section). This section has discussed the overall system architecture only, while the details at the object level must be considered in the next section.

4 Archival Stability of Objects

The object-oriented methodology has emerged to be a standard representation scheme of information technology, preservation and thus archival science [12]. The information encapsulation principle provides a representation of digital objects. It wraps the information content by its accompanying procedures, which are applied whenever necessary. Processes in the life cycle of information may be expressed as complex objects of objects [3]. In each organizational infrastructure, we will use the object-oriented framework of the life cycle of information. The life cycle consists of acquisition, preservation, collection, indexing, accessing, and utilization (including various workflow processes discussed in the previous section), where preservation is an important component of the life cycle.

The information encapsulation principle provides a general representation scheme of various multimedia data content of the life cycle. The multimedia data content is wrapped by its representation information and accompanying software programs, which are applied whenever necessary. The representation information maps the bit streams of data content into understandable information of certain format, structure, and type. Thus data content, representation information, and software programs become modular and reusable. Digital objects are convertible and transformable, and operable under processes of the life cycle of information. Since digital objects are encapsulated, they are active, dynamic, and extensible. They are active, because software agents may be embedded in objects so that activities may be initiated by objects. They are dynamic under any process that is associated with accompanying software programs of objects. They are extensible in the sense of multimedia content and networked sources. Any object can be augmented by other objects of multimedia content and from networked sources. It is highly plausible that a digital object pulls several networked records into itself. The multimedia digital content is composed of one or more bit sequences. The purpose of the representation information is to convert the bit sequences into more meaningful information. It does this by describing the format, or data structure concepts, which are to be applied to the bit sequences and that in turn result in more meaningful values such as characters, numbers, pixels, arrays, tables, etc. For simplicity, we will call such active, dynamic and extensible objects by a single adjective: "dynamic". Thus in this paper, "dynamic" has the characteristics of active, dynamic and extensible. In our subsequent work, we will provide more details on the development of dynamic objects (DO's) in this convention.

How can we develop a preservation strategy for these objects? The data types, their aggregations, and mapping rules which map from the underlying data types to the higher level concepts are referred to as the structure information component of the representation information. These structures are commonly identified by name or by relative position within the associated bit sequences. The representation information provided by the structure information component is usually insufficient to understand the digital content. The additional required information is referred to as the semantic information. Semantic information may be quite complex. It may include special meanings associated with all the elements of the structural information, processes that may be performed on each data type, and their inter-relationships. Moreover representation information may contain further associative references to other representation information.

In order to preserve a digital object, its representation information, both structural and semantic, must also be preserved. This is commonly accomplished when the representation information is expressed in text descriptions that use widely supported standards such as ASCII characters for digital versions. If text descriptions are ambiguous, we should use standardized, formal description languages (e.g., XML markup languages) containing well-defined constructs with which to describe data structures. These markup languages will augment text descriptions to fully convey the semantics of the representation information.

Software programs associated with digital objects commonly are representation rendering software and access software. Representation rendering software is able to display the representation information in human-readable forms, such as the PDF display software to render the record human-readable. Access software presents some or all of the information content of a digital object in forms understandable to humans or systems. It may also provide some types of access service, such as displaying, manipulating, processing, to another object (e.g., scientific visualization systems supporting time series or multidimensional array). Again its future existence and migration depend highly on the *technology continuity*, which is a very difficult prediction to make. Since representation rendering software and access software are provided at the desktop, their preservation is not necessary at each object level, but only at the environmental level.

It is tempting to use Internet-based access software to incorporate some of the representation information as a cost-effective means. Many web-based services actually do use web access software as the full representation information, and the WWW consortium is doing an excellent job for this effort. Access software source code becomes at least the partial representation information of those digital objects. First such information may be mixed with various other processing and display algorithms, and may be incomplete since the code assumes an underlying operating environment. Secondly, if executables of access software are used, without the source code, such archives have great risks for loss of representation information. It is more difficult to maintain an operating environment for software than to migrate over time. If the organizational computing environment supports the software, it is not difficult to access the preserved package. The environment consists of the underlying hardware and operating system, various utilities that effectively augment the operating system, and storage and display devices and their drivers. A change to any of these will cause the software no longer function properly. This is why preservation of software is complex and complicated. To push one more level, representation information may need to include dictionary and grammar of any natural language (e.g., English) used in expressing the digital content. Over long time periods the meaning of natural language expressions can evolve significantly in both general and specific disciplines.

An important step of preservation is the *bundling* of necessary preservation information to a digital object so that we can still access and retrieve the content whatever the hardware, software, and media migration may advance. A preservation package is a conceptual container of these two types of information, content information and preservation information [13]. The content information and preservation information are encapsulated and identifiable by the package information. The resulting package is accessible by the descriptive information of the preservation package. The content information is the original target information of preservation. It consists of the digital content and its associated representation information and software programs. The preservation information applies to the content information and is needed to preserve the content information, to ensure it is clearly identified, and to understand the environment in which it was created.

The preservation information is divided into four types of preserving information called provenance, context, reference, and fixity. Briefly, they are described in the following four categories:

1. Provenance describes the history and source of the content information: who has had custody of it since its origination, and its history (including processing history). This gives future users some assurance as to the likely reliability of the content. Provenance can be viewed as a special type of context information.

2. Context describes how the content information relates to other information outside the information package. For example, it would describe why the content information was produced, and it may include a description of how it relates to another content information object that is available.

3. Reference provides one or more identifiers, or systems of identifiers, by which the content information may be uniquely identified. Examples include an ISBN number for a book, or a set of attributes that distinguish one instance of content information from another. Further examples include taxonomic systems, reference systems and registration systems.

4. Fixity provides a wrapper, or protective shield, that protects the content information from undocumented alteration. It provides the data integrity checks or validation/verification keys used to ensure that the particular content has not been altered in an undocumented manner.

The packaging information is that information which binds, identifies, and relates the content information and preservation information either actually or logically. The descriptive information of the package is the information, which is used to discover which package has the content information of interest. It may be a full set of attributes or metadata that are searchable in a catalog service. In OAIS [1], the total archival information over an indefinite period of time is called preservation description information (PDI). The packaging information does not necessarily need to be preserved since it does not contribute to the content information or the PDI. The preservation should also avoid holding PDI or content information only in the naming conventions of directory or file name structures. These structures are most likely to be used as packaging information. Packaging information is not preserved by migration. Any information saved in file names or directory structures may be lost when the packaging information is altered.

The life cycle of information is also represented in an object-oriented framework [3]. The life cycle consists of at least the following processes of information: acquisition, preservation, collection, indexing, accessing, and utilization. In particular, utilization may further include many other workflow processes in organizations. In any organization, information is received from external transactional sources, generated by its own workflow processes, or accessed from information resources (e.g., libraries and archives). The life cycle starts, only after an object is acquired (ingested or initiated) by an organization. An object acquired will be preserved for long-term use. All preserved objects are stored into appropriate collections, each of which is properly indexed for future accessing and utilization. The life cycle captures a spiral (rather than a linear) cycle, because acquisition follows utilization in a repetitive and iterative pattern. Information generated by utilization is often ingested into the life cycle!

There are also secondary processes. It is impossible to have a complete listing of secondary processes. These secondary processes may be embedded in various stages of the life cycle. In this paper, workflow processes are assumed in general forms and not described in details. Workflow processes may include secondary processes. To illustrate, we discuss a few secondary processes. Conversion and transformation of formats and structures of digital objects permits the interchange among them. They are used in preservation for example. Communication and transmission sends digital objects from computer hardware and storage systems through communications networks. They are needed perhaps in every process of the life cycle. Brokerage and integration mediates query results from networked sources into unified objects for users. They are essential to the accessing process. Delivery and presentation brings information in useful manners to users in access and utilization.

Since digital objects are associated with processes in the life cycle of information, it is natural to develop the definition of dynamic objects as those with associated processes. Let us envision the following scenario: On the Internet, dynamic objects automatically transmit in the network from one stage to the next stage according to some associated processes, supporting dynamic horizontal and vertical information flows of the life cycle. For each stage of the process, dynamic objects will do the following: build their metadata, join a group of objects of the same purpose, and offer-of-access-of-metadata to processes and members of the organization through some automatic indexing and clustering algorithms. To accomplish this scenario, some sophisticated techniques must be employed. For examples, knowledge discovery and data mining must be used. Knowledge discovery has been used to extract useful knowledge from large volumes of data and it is one of the many potential objectives of information fusion. Data mining is defined as the nontrivial extraction of implicit, previously unknown, and potentially useful information from data and can be used as a means of accomplishing the objectives of knowledge discovery. Dynamic objects can be constructed by simultaneously indexing objects with their workflow processes and dynamically organize and search the resource space by constructing links among the objects based on the metadata that describes their contents, types, context, and workflow processes. The offer-of-access-of-metadata is specified by the user profiles, and the semantics of the type of dynamic objects. The links will be used to generate a virtual graph, with a flexible set of multiple hierarchies to provide searching and browsing facilities in the organized concept space. We may also build an ontology space by using the graph constructed. The ontology space is an explicit specification of a conceptualization of objects. The use of ontology provides an effective way to describe objects and their relationships to other objects.

5 Conclusions

We have proposed a potential solution to the digital preservation problem in digital libraries/archives. The interplay of organizational process, archival stability, and technology continuity is evident in our discussion that more innovative research should be conducted in this direction. Our solution has also developed a new model of digital libraries that is dynamic and associated with workflow processes.

References

1. CCSDS and ISO TC20/SC13, "Reference Model for an Open Archival Information System", July 2001, http://www.ccsds.org/documents/pdf/CCSDS-650.0-R-2.pdf.
2. S. Chen, The paradox of digital preservation, IEEE Computer, March 2001.
3. S. Chen, Digital Libraries: The Life Cycle of Information, BE Publisher, 1998.
4. S. Chen, Digital preservation and the life cycle of information, Advances in Computers, M. Zelkowitz (ed.), volume 57, Academic Press/Elsevier Science, 2003.
5. S. Chen, C. Choo, and R. Y. Chow, Internet security: A novel role/object-based access control for information domains, under revision by Journal of Organizational Computing and Electronic Commerce, http://phoenix.lite.cise.ufl.edu:8080/dllsl/papers.html
6. C. M. Dollar, Archival Theory and Information Technologies: The Impact of Information Technologies on Archival Principles and Methods, Macerata: University of Macerata Press, 1992.
7. L. Duranti, Diplomatics: New uses for an old science (Part V), Archivaria, 32 1991, pp. 6–24.
8. J. Garrett and D. Waters, Preserving digital information, Report of the Task Force on Archiving of Digital Information, May 1996, <http://www.rlg.org/ArchTF/>.
9. M. Hedstrom, Descriptive practices for electronic records: Deciding what is essential and imaging what is possible, Archivaria, 36, 1993, pp. 53–63.
10. InterPARES, http://www.InterPARES.org/.
11. J. Rothenberg, Ensuring the longevity of digital documents, Scientific American, 272, 1995, pp. 42–47.
12. J. Rumbaugh, M. Blaha, W. Premerlani, F. Eddy, and W. Lorensen, Object-Oriented Modeling and Design, Prentice Hall, 1991.
13. T. Shepard, UPF User Requirements, http://info.wgbh.org/upf/.
14. Victorian Electronic Records Strategy Final Report, Public Record Office Victoria 1999, http://www.prov.vic.gov.au/vers/
15. National Digital Information Infrastructure and Preservation Program, http://www.digitalpreservation.gov/.
16. OCLC, http://www.oclc.org/services/preservation/default.htm.
17. IBM Vision of Autonomic Computing (IEEE Computer January 2003), www.ibm.com/ research/ autonomic/.
18. NARA ERA, http://www.archives.gov/electronic_records_archives/research/research.html.
19. Australian Standards and Framework: Records Management and Metatagging of Web Pages, http://www.lester.boisestate.edu/metatags/Australian_Standards_and_Framework1.htm.
20. Sue McKemmish Describing Records in Context in the Continuum: the Australian Recordkeeping Metadata Schema, http://www.sims.monash.edu/research/rcrg/publications/archiv01.htm.

21. H. Kim and S. Chen, Ontology Search and Text Mining of MEDLINE Database, Conference on "Data Mining in Biomedicine", February 16-18, 2004, University of Florida, Gainesville, FL; Book published by Kluwer.
22. S. Shi, O. Rodriguez, S. Chen and Y. Shang, Open learning objects as an intelligent way of organizing educational material, International Journal on E-Learning, Vol. 3 No. 2, 2004, pp. 51–63.
23. H. Kim, C. Choo, and S. Chen, An Integrated Digital Library Server with OAI and Self-Organizing Capabilities. In *Proceedings of the 7th European Conference on Research and Advanced Technology for Digital Libraries (ECDL 2003)*, Trondheim, Norway, August (2003)
24. C. Lagoze and H. Van de Sompel, The Open Archives Initiative: Building a low-barrier interoperability framework. In Proceedings of the First ACM/IEEE Joint Conference on Digital Libraries, Roanoke, VA, 2001. Pages 54–62.
25. C. Lagoze, H. Van de Sompel, M. Nelson and S. Warner, The Open Archives Initiative Protocol for Metadata Harvesting. Open Archives Initiative, 2001. http://www.openarchives.org/OAI/openarchivesprotocol.htm.
26. S. Chen, C. Choo, and Y. Chow, Internet Security: A Novel Role/Object-Based Access Control for Digital Libraries, Journal of Organizational Computing and Electronic Commerce, Lawrence Erlbaum Publisher, under revision.
27. InterPARES Project, http://www.interpares.org/rws/login.cfm?accessdenied=%2Frws%2Findex%2Ecfm.