The Role of Digital Libraries in Moving Toward Knowledge Environments

Edward A. Fox, Marcos André Gonçalves, and Rao Shen

Virginia Tech Dept. of Computer Science, 660 McBryde Hall, Blacksburg, VA 24061 USA
{fox, mgoncalv, rshen}@vt.edu
http://www.dlib.vt.edu

Abstract. For thousands of years, libraries have allowed humanity to collect and organize data and information, and to support the discovery and communication of knowledge, across time and space. Coming together in this Internet Age, the world's societies have extended this process to span from the personal to the global, as the concepts, practices, systems, and services related to Library and Information Science unfold through digital libraries. Scientists, scholars, teachers, learners, and practitioners of all kinds benefit from the distributed and collaborative knowledge environments that are at the heart of the digital library movement. Digital libraries thus encompass the dimensions in the 5S Framework: Societies, Scenarios, Spaces, Streams, and Structures. To clarify this approach, we explain the role of meta-models, such as of a minimal digital library (DL), and of more specialized (discipline-oriented) DLs, such as archeological DLs. We illustrate how suitable knowledge environments can be more easily prepared as instances of these meta-models, resulting in usable and useful DLs, including for education, computing, and archaeology.

1 Introduction

People need information. Libraries help us satisfy this requirement, playing a key role in the Information Life Cycle [3]. Today, Digital Libraries (DLs [2, 6, 18]) provide essential cyberinfrastructure, moving us toward knowledge environments geared toward individual needs, as well as helping us address global concerns.

Figure 1 provides perspective on our work. DLs typically focus on some domain, so that interested patrons (actors) interact with a running system, built according to some DL architectural approach. Underlying all this is some model of the "real world", which fits into a meta-model tuned to the domain of interest. "5S" (see Table 1) provides fundamental abstractions to make this whole process straightforward.

Thus, Section 2, on 5S, helps situate the field of DLs in the world of (computer/ library/information) science, considering formalisms, models, minimalist meta-models, and (semi)automatic approaches. Section 3 briefly illustrates how our framework for DLs can be applied to representative domains. Section 4 summarizes and concludes our discussion.
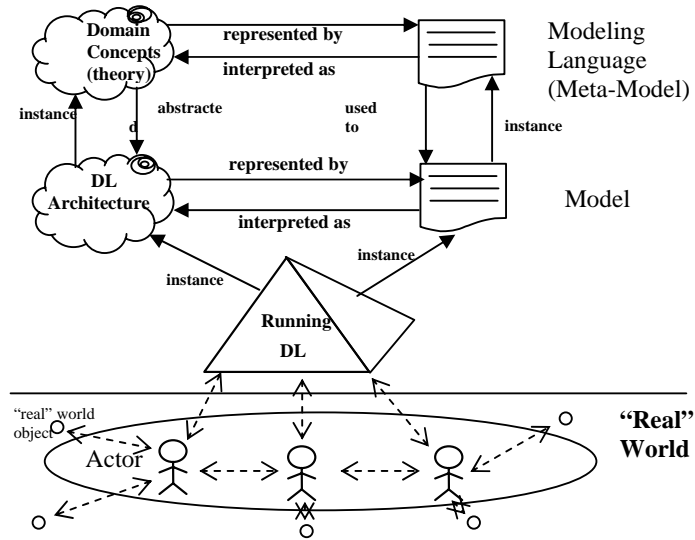
Figure 1. Perspective underlying the 5S framework.

Table 1. 5S fundamental abstractions.

Models	Examples	Objectives
Stream	Text; video; audio; image	Describes properties of the DL content such as encoding and language for textual material or particular forms of multimedia data
Structures	Collection; catalog; hypertext; document; metadata; organization tools	Specifies organizational aspects of the DL content
Spaces	Measure; measurable, topological, vector, probabilistic	Defines logical and presentational views of several DL components
Scenarios	Searching, browsing, recommending,	Details the behavior of DL services
Societies	Service managers, learners, teachers, etc.	Defines managers responsible for running DL services; actors that use those; and relationships among them

2 5S Framework

Building upon the 5 constructs explained in Table 1, we have developed a formal framework for the DL field [9, 10], summarized in Figure 2. Figure 3 illustrates how we define additional elements, e.g., for an archaeological DL [20].
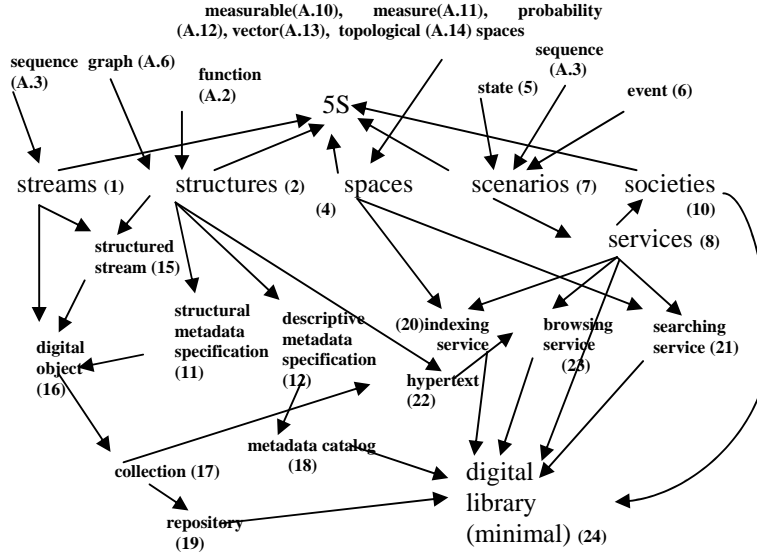

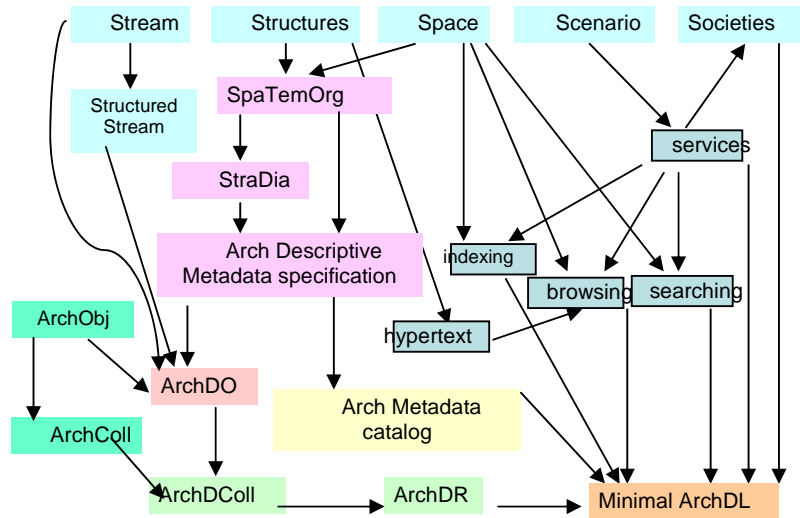**Figure 2. 5S definitional structure.**

Figure 3. Minimal archaeological DL in the 5S framework.

In particular, we go beyond the foundation of *digital object (do)* and include *ArchDO* that refers to a real world *ArchObj*. Our minimalist *ArchDL* includes a digital collection *(ArchDColl)* and digital repository *(ArchDR)*. The metadata in our catalog makes use of a stratigraphic diagram *(StraDia)* which builds upon space-time organization *(SpaTemOrg)*. While Figure 3 conveys the essence of a meta-model for archaeological DLs, Figure 2 characterizes a minimal (typical) DL.

Moving toward a theory, we developed a minimal DL ontology, summarized in Figure 4 [10]. Here Ac=actor, C=collection, DM_c=metadata catalog, e=event, I_c=index, ms=descriptive metadata specification, mss=set of ms, op=operation, R=repository, Sc=scenario, Se=service, and SM=service manager. An application of the ontology is a taxonomy of DL services and activities (Figure 5). This formalism (where e=employs, p=produces) enables reasoning about composition of services (Figure 6).

3   Applications of 5S

The 5S framework allows a new approach to DL development (Figure 7). 5SGraph [23] supports analysis and specification, while 5SLGen [15] melds together suitable components from a large software pool to yield a running system. Figure 8 illustrates use of 5SGraph to specify the CITIDEL system (see collections in top right); the minimal DL meta-model is shown in the bottom portion. Figure 9 shows this for the Nimrin archaeological site, focusing on space, drawing upon a meta-model for archaeology, built for ETANA-DL [20]. The current interface for the ETANA system is shown in Figure 11. Further, we argue that 5S allows formalization of quality issues [10] in the Information Life Cycle [3]. In addition to ongoing work on 5S in the context of archaeology, we have applied it to education and computing.

The **C**omputing and **I**nformation **T**echnology **I**nteractive **D**igital **E**ducational **L**ibrary [4] was launched in 2001 through support from the US National Science Foundation as one of the National Science Digital Library (NSDL) collection projects. Included are almost a million metadata records. We balanced including broad digital libraries from professional societies (ACM [1] and IEEE-CS [13]), large collections obtained by web crawling (CiteSeer [8]), and curated bibliographies covering key conferences and journals (DBLP). Using the Open Archives Initiative Protocol for Metadata Harvesting [17], we were able to gather the subset of records for electronic theses and dissertations (ETDs) relating to computing [22] or PlanetMath (a community-build online math encyclopedia.

Another goal of NSDL collection efforts has been to provide improved support for targeted communities, such as through portals. In the latter regard, the VIADUCT portion of CITIDEL [5] helps with the creation of lesson plans built around learning resources, extending earlier NSDL-funded work on the Instructional Architect [21] and on Walden's Paths [7]. With enhanced usability, such systems can become easier to use [19]. With visualization support, users can more quickly learn about the

collection as well as manage large result sets [14]. Further, to provide enhanced browsing support, four different category systems relevant to computing have been mapped by hand so that an automatic process has categorized works, previously indexed in only one scheme, into all applicable schemes [16].

Assessment of CITIDEL is ongoing, to guide further enhancements. Our main emphasis has been on using logs. We proposed [11] and refined [12] a suggested standard for DL logging, based on 5S, covering all important behaviors and activities. Indeed, 5S has guided much of the work on CITIDEL as well as ETANA-DL.
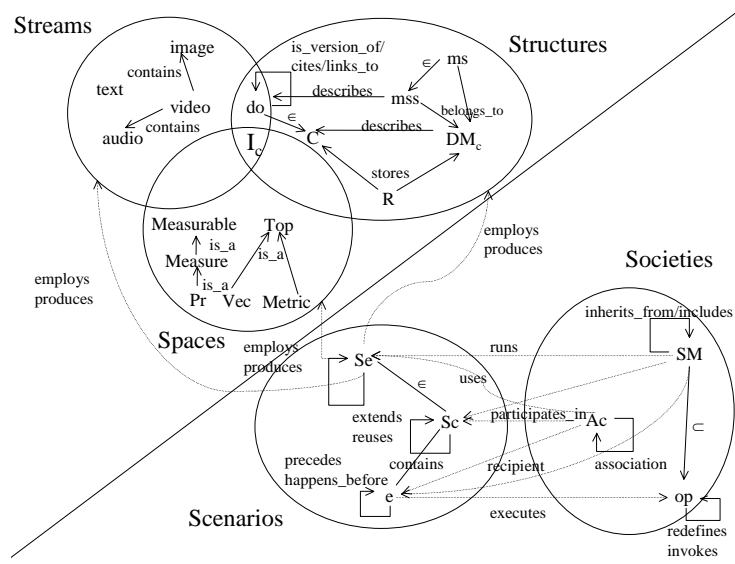

Figure 4. DL ontology.

Table 2. DL services/activities taxonomy.

| Infrastructure Services			Information Satisfaction Services
Repository-Building		*Add Value*	
Creational	Preservational		
Acquiring Cataloging Crawling (focused) Describing Digitizing Federating Harvesting Purchasing Submitting	Conserving Converting Copying/Replicating Emulating Renewing Translating (format)	Annotating Classifying Clustering Evaluating Extracting Indexing Measuring Publicizing Rating Reviewing (peer) Surveying Translating (language)	Browsing Collaborating Customizing Filtering Providing access Recommending Requesting Searching Visualizing
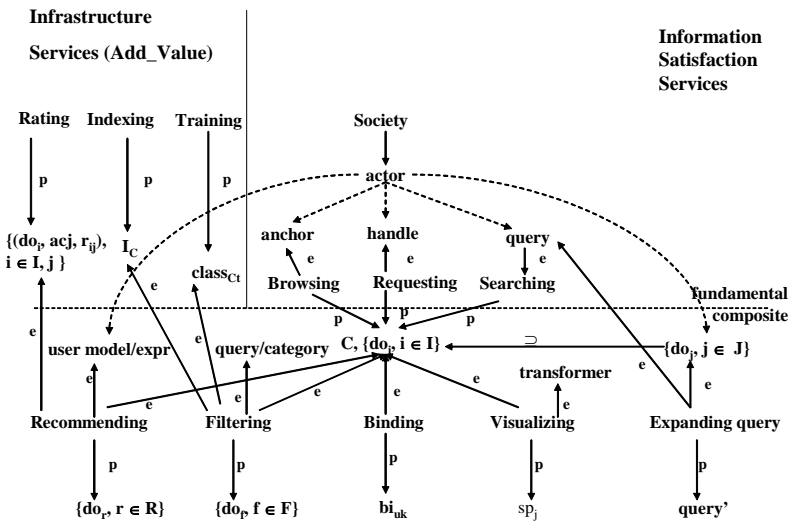
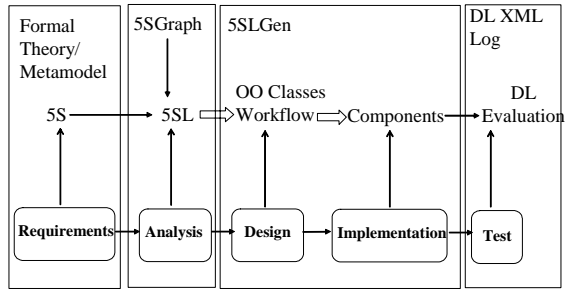**Figure 5. Example of composition of DL services.**
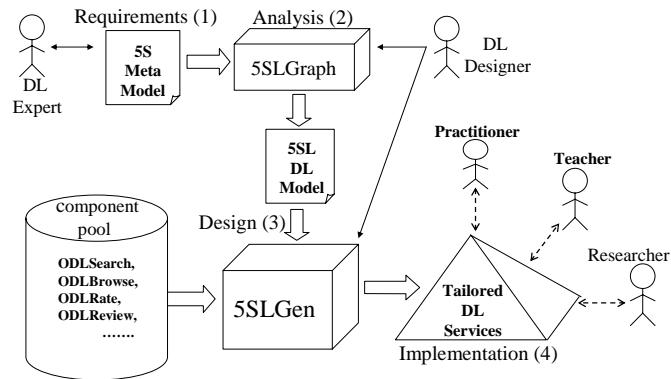
Figure 6. 5S framework and DL development.

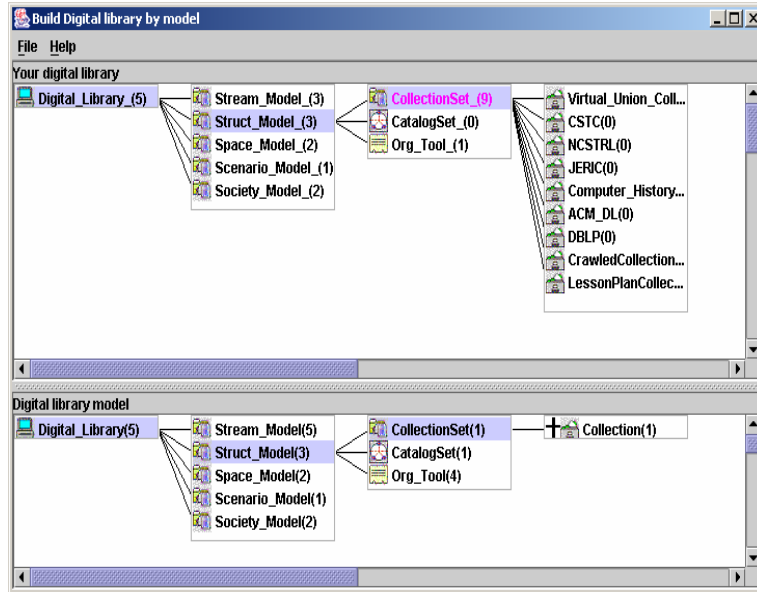**Figure 7. 5SLGen: automatic DL generation.**

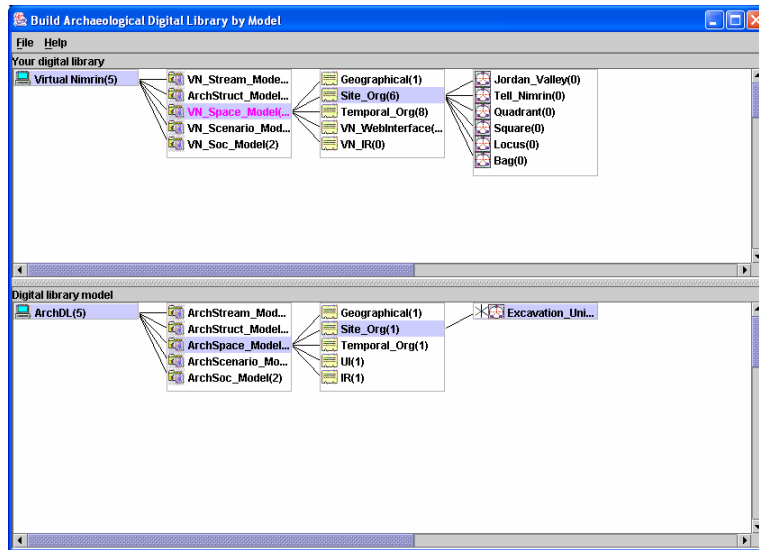**Figure 8. 5SGraph, minimal DL, CITIDEL model.**

Figure 9. 5SGraph, arch. DL, Space model for Nimrin.
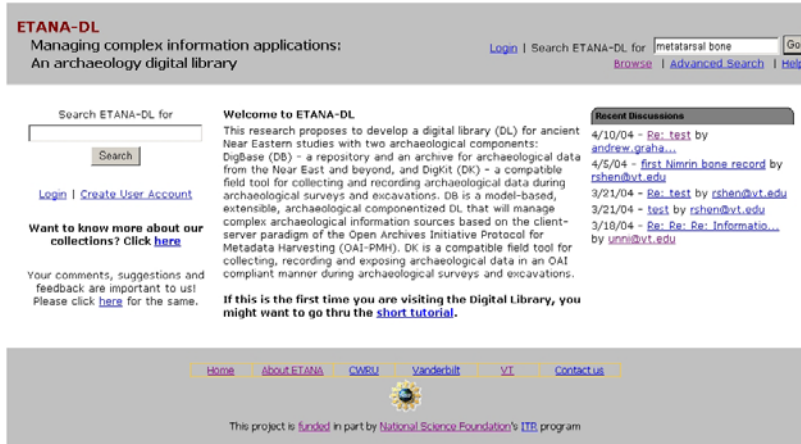
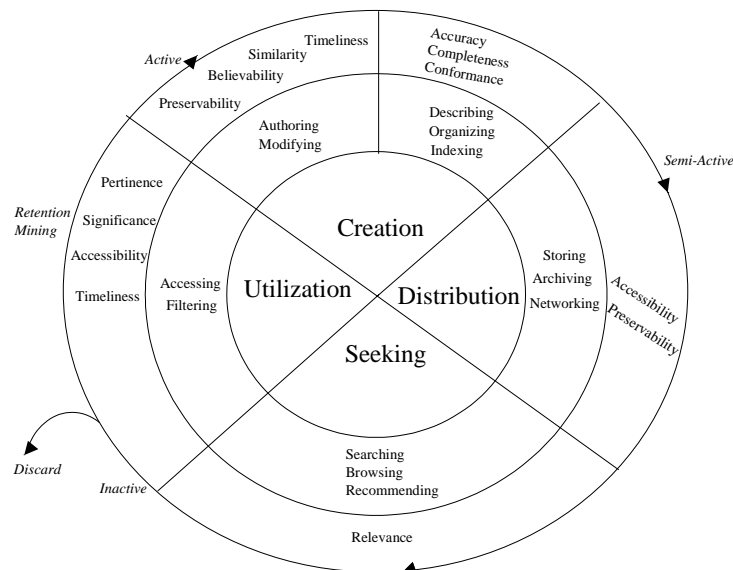**Figure 10. ETANA-DL searching service.**

Figure 11. Quality and the Information Life Cycle.

6 Conclusion

Digital libraries have evolved since their inception in the early 1990s. Our work on the 5S framework aims to guide a movement toward knowledge environments, as has been illustrated by our applications of 5S to education, computing, and archaeology. We believe that the 5S framework may provide a firm foundation for advanced DLs.

7 Acknowledgements

References

[1]   ACM. "ACM Digital Library", 2000. *http://www.acm.org/dl/*.
[2]   Arms, W. Y. *Digital Libraries*. Cambridge, MA, MIT Press, 2000.
[3]   Borgman, C. "Social Aspects of Digital Libraries", UCLA, Los Angeles, NSF Workshop Report, Feb. 16-17 1996. *http://www-lis.gseis.ucla.edu/DL/*.
[4]   CITIDEL. "CITIDEL: Computing and Information Technology Interactive Digital Educational Library", E. A. Fox, D. Knox, L. Cassel, J. A. N. Lee, M. Pérez-Quiñones, J. Impagliazzo, and C. L. Giles, Eds. Blacksburg, VA: Virginia Tech, 2002. *http://www.citidel.org*.
[5]   CITIDEL. Virginia Instructional Architect for Digital Undergraduate Computing Teaching (VIADUCT), 2004. *http://www.citidel.org/?op=viaduct_front*.
[6]   Fox, E. A. and Urs, S. "Digital Libraries", in *Annual Review of Information Science and Technology*, vol. 36, Ch. 12, B. Cronin, Ed., 2002, pp. 503-589.
[7]   Furuta, R. Walden's Paths, 2004. *http://csdl.cs.tamu.edu/walden/*.
[8]   Giles, C. L., Bollacker, K., and Lawrence, S. "CiteSeer: An Automatic Citation Indexing System", in *Proc. Third ACM Conf. Digital Libraries, DL'98 (Pittsburgh)*, I. Witten, R. Akscyn, and F. M. Shipman III, Eds. ACM Press: ACM, 1998, pp. 89-98. *http://www.neci.nj.nec.com/homepages/lawrence/papers/cs-dl98/*.
[9]   Gonçalves, M., Fox, E., Watson, L., and Kipp, N. *Streams, Structures, Spaces, Scenarios, Societies (5S): A Formal Model for Digital Libraries*. ACM Transactions on Information Systems, vol. 22(2), pp. 270-312, April, 2004.
[10]   Gonçalves, M. A. *Streams, Structures, Spaces, Scenarios, Societies (5S): A Formal Digital Library Framework and Its Applications*. Doctora dissertation. Computer Science, Virginia Tech, Blacksburg, VA, 2004.
[11]   Gonçalves, M. A., Luo, M., Shen, R., Farooq, M., and Fox, E. A. *An XML Log Standard and Tool for Digital Library Logging Analysis*. Presented at Sixth European Conference on Research and Advanced Technology for Digital Libraries, Rome, Italy, 2002.

[12]   Gonçalves, M. A., Panchanathan, G., Ravindranathan, U., Krowne, A., Fox, E., Jagodzinski, F., and Cassel, L. *Standards, mark-up, and metadata: The XML log standard for digital libraries: analysis, evolution, and deployment*. Presented at 3rd ACM/IEEE-CS joint conference on digital libraries, IEEE Computer Society, 2003.

[13]   IEEE-CS. IEEE Computer Society Digital Library, 2004. *http://www.computer.org/publications/dlib/*.

[14]   Kampanya, N., Shen, R., Kim, S., North, C., and Fox, E. A. "Citiviz: A Visual User Interface to the CITIDEL System", in *Proc. European Conference on Digital Libraries (ECDL) 2004, September 12-17, University of Bath, UK*, 2004.

[15]   Kelapure, R. *Scenario-Based Generation of Digital Library Services*. MS thesis. Computer Science, Virginia Tech, Blacksburg, VA, 2003. *http://scholar.lib.vt.edu/theses/available/etd-06182003-055012/unrestricted/Thesis_etd_changes.pdf*.

[16]   Krowne, A. and Fox, E. A. "An Architecture for Multischeming in Digital Libraries", in *Proceedings 6th International Conference on Asian Digital Libraries, ICADL 2003, Digital Libraries: Technology and Management of Indigenous Knowledge for Global Access;Kuala Lumpur, Malaysia, Dec.; Springer, Lecture Notes in Computer Science 2911*, T. Mohd, T. Sembok, H. B. Zaman, H. Chen, S. R. Urs, and S. H. Myaeng, Eds., 2003, pp. 563-577.

[17]   Lagoze, C., Van de Sompel, H., Nelson, M., and Warner, S. "The Open Archives Initiative Protocol for Metadata Harvesting - Version 2.0, Open Archives Initiative", 2002. *http://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm*.

[18]   Lesk, M. *Practical Digital Libraries: Books, Bytes and Bucks*. San Francisco, Morgan Kaufmann Publishers, 1997.

[19]   Perugini, S., McDevitt, K., Richardson, R., Perez-Quinones, M., Shen, R., Ramakrishnan, N., Williams, C., and Fox, E. A. "Enhancing Usability in CITIDEL: Multimodal, Multilingual, and Interactive Visualization Interfaces", in *Proceedings Fourth ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL2004), Tucson, AZ, June 7-11*, 2004, pp. 315-324.

[20]   Ravindranathan, U., Shen, R., Gonçalves, M. A., Fan, W., Fox, E. A., and Flanagan, J. W. "Prototyping Digital Libraries Handling Heterogeneous Data Sources – The ETANA-DL Case Study", in *Proc. European Conference on Digital Libraries (ECDL) 2004, ECDL2004, September 12-17, 2004, U. Bath, UK*, 2004, pp. 186-197.

[21]   Rucker, M. Instructional Architect, 2004. *http://ia.usu.edu/*.

[22]   Zhang, B., Gonçalves, M. A., and Fox, E. A. "An OAI-Based Filtering Service for CITIDEL from NDLTD", in *Proceedings 6th International Conference on Asian Digital Libraries, ICADL 2003, Digital Libraries: Technology and Management of Indigenous Knowledge for Global Access; Kuala Lumpur, Malaysia, Dec.; Springer, Lecture Notes in Computer Science 2911*, T. Mohd, T. Sembok, H. B. Zaman, H. Chen, S. R. Urs, and S. H. Myaeng, Eds., 2003, pp. 590-601.

[23]   Zhu, Q. *5SGraph: A Modeling Tool for Digital Libraries*. Masters thesis. Department of Computer Science, Virginia Tech, Blacksburg, 2002. *http://scholar.lib.vt.edu/theses/available/etd-11272002-210531/*.