



Digital Libraries for the Next Millennium: Challenges and Research Directions

Sudha Ram, Jinsoo Park and Dongwon Lee

Department of Management Information Systems, College of Business and Public Administration, The University of Arizona, Tucson, AZ 85721,

E-mail: {ram, jpark, dlee}@bpa.arizona.edu URL: http://vishnu.bpa.arizona.edu

Abstract. The unprecedented growth of Internet technologies has made resources on the World Wide Web instantly accessible to various user communities through digital libraries. Since the early 1990s, there have been several digital library initiatives sponsored by government agencies and/or private organizations all over the world. A digital library is a networked system environment that provides diverse user communities with coherent, seamless and transparent access to large, organized, and digitized information resources. This article provides a comprehensive overview of major digital library projects that are currently being undertaken across the globe. We also identify and discuss major challenges and research issues to be addressed in the design and implementation of digital libraries for the next millennium. We believe that digital libraries are ripe with research opportunities, offer many challenges, and will continue to grow in the next several years.

Key Words. digital library, metadata management, resource sharing, resource harvesting, software agent, socioeconomic context of digital library

1. Introduction

The unprecedented growth of Internet technologies has made resources on the World Wide Web (WWW) instantly accessible to various user communities. The business community has utilized this new world with innovative strategies to expand their customer base (e.g., Internet marketing, electronic commerce). The general public including scholars and students use the Internet to access and share information stored throughout the world. The Internet and related information technologies (e.g., WWW) have also facilitated information sharing worldwide. Before this

new trend, the majority of information access and dissemination activities were carried out through printed materials (e.g., journals, newspapers, magazines, etc.), mass media or libraries. Although the library's role as information facilitator will not fade in the foreseeable future, new technologies will almost inevitably require different arrangements among them with far-reaching implications for our society and economy. For instance, while characteristics of libraries have been determined by print technology in the past (Wiederhold, 1995), it is certain that the emergence of digital libraries due to the proliferation of digital technologies and the WWW will transform traditional libraries. Such technological advance is not the only contributing factor in this new movement. Economic factors, such as diminishing purchase power due to price increases in printed materials and overhead cost for maintaining old materials, are also culprits. The objectives of this article are to provide a comprehensive overview of major digital library initiatives that are currently being undertaken across the globe, and to identify the major challenges and research issues to be addressed in the design and implementation of digital libraries for the next millennium.

Traditionally, libraries have served as repositories of information. Although libraries of different types have served different user communities, their fundamental roles remain the same; that is, collecting, selecting, filtering, storing, and classifying information sources to make them easily accessible to the user community, and preserving them for future generations. Libraries have also served critical roles in teaching and learning. Marchionini and Maurer

(1995) describe three roles of libraries in teaching and learning: (1) *a practical role* in sharing resources, (2) *a cultural role* in preserving and organizing information resources, and (3) *social and intellectual roles* by bringing together people and ideas. These traditional roles still remain primary characteristics of digital libraries. However, digital libraries can provide enhanced and value-added services that traditional libraries could not. For example, information resources are open 24 hours a day, and are accessible via a network that allows users to access materials no matter where they live. In this sense, digital libraries are global. Digital libraries are also dynamic in that materials can be added and updated by many contributors from any place at any time, although quality control procedures for resource collection that the digital library enforces may limit the dynamics of digital libraries.

There are many definitions of a "digital library." Terms such as "electronic library," "virtual library," "paperless library," "networked library," "library without walls," and "multimedia library" are often used synonymously. The term "digital library" has, however, become the preferred term due to growing interest in digital networks, digital audio, and digital video (Fox, et al., 1995). The definition of a "digital library" ranges from merely a collection of electronic documentation accessible through the Internet, to a collection of all types of digital information, including text, images, audio, video and scientific data (both raw and processed) along with value-added services to all possible users. In this paper, we define a digital library as a networked system environment that provides diverse user communities with coherent, seamless and transparent access to large, organized, and digitized information resources. The user community is defined as a group of people who have common interests (e.g., database, childcare, hiking, fishing, etc.) or common characteristics (e.g., age, culture, etc.). Digital libraries are a federated network system environment since many different types of information resources that have specialized collections will be inter-linked together to provide user communities with value-added services. In this context, we do not view the digital library as a single entity.

Although the scope of a digital library may vary, we can briefly explicate a common set of desirable characteristics of a digital library. A digital library:

- has an open architecture built on a collection of distributed information repositories;
- stores multiple formats;
- contains both metadata and data;
- provides integrated functions of searching and presenting;
- allows seamless and transparent access to the multiple heterogeneous information sources;
- facilitates interoperability;
- purveys user-friendly easy-to-use interfaces and query facilities;
- and serves users (both information seekers and providers).

Since the early 1990s, there have been several digital library initiatives all over the world. These initiatives are sponsored by government agencies and/or private organizations. In Section 2, we survey and discuss major digital library development efforts that are being undertaken in North America, Europe, and other parts of the world including the Pacific Region. Section 3 explains the major components of digital libraries based on an analysis from Section 2, and discusses major challenges and research issues that should be addressed in the development and deployment of digital libraries. Our concluding remarks are summarized in Section 4.

2. On-Going Digital Library Projects

2.1. Major digital libraries initiative in the U.S.

Recognizing the role of digital libraries, several federal agencies and private organizations in the United States have launched initiatives to spur the development of digital libraries. Among these initiatives, the Digital Libraries Initiative (DLI) is jointly sponsored by the National Science Foundation (NSF), the Department of Defense Advanced Research Projects Agency (DARPA), and the National Aeronautics and Space Administration (NASA). Phase-1 of this initiative consists of six 4-year research projects funded from September 1994 to 1998. The common objectives of these projects are: (1) to advance the means of collection, storage, searching, retrieval and processing in digital form via wide-area communication networks (Cole and Harum, 1995); (2) to improve practices of communities so they are more effective, efficient, and productive

(Atkins, 1997); (3) to create a large testbed and acquire substantial cost sharing; and (4) to facilitate the emergence of new communities of discourse, research, and learning (Atkins, 1997).

The six DLI projects are centered at Carnegie Mellon University, the University of California, Berkeley, the University of Michigan, the University of Illinois, the University of California, Santa Barbara, and Stanford University. Each effort brings together researchers and users from the university as well as from many other organizations including other academic institutions, libraries, museums, publishers, government and state agencies, secondary schools, and business community (Cole and Harum, 1995).

Informedia Digital Video Library Project. The Informedia project, led by Carnegie Mellon University, has focused on the development of intelligent, automatic mechanisms to populate the library through integrated speech, image and language understanding (Christel et al., 1995), as well as full-content, knowledge-based information retrieval via desktop computer and metropolitan area networks (Cole and Harum, 1995). The distinguishing features of this project include: speech recognition for automated transcript generation (Hauptmann, 1995), network billing server (NetBill) (Cox et al., 1995), color image similarity matching and indexing, video skimming and characterization (Smith and Kanade, 1998), human face detection and name association (Rowley et al., 1995), and natural language processing for information retrieval (Wactlar et al., 1996). Industry partners for this project include Bell Atlantic, Boeing, DEC, Intel, Microsoft, and Motorola.

Environmental Electronic Library Project. The focus of the Environmental Electronic Library Project at the University of California, Berkeley, is on environmental information that can be accessed using advanced geographical information retrieval interfaces. The project supports work-centered digital information services that address a work group's information retrieval needs. To realize work-centered digital information systems, the library focuses on (1) document image analysis, natural language analysis, and computer vision analysis for effective information extraction, (2) new user interface paradigms and authoring tools for easily accessing multimedia information, and (3) improved protocols for client program interaction with repositories (Wilensky, 1996). In addition, the digital library allows untrained

users to contribute and find relevant information (House, 1996). Participating organizations range from federal and state agencies (e.g., US Geological Survey, California Department of Water Resources, California Department of Fish and Game, California Land Use Planning Network, California State Library, etc.) to industry partners (e.g., HP, Informix, IBM, Philips Research, Sun Microsystems, Xerox PARC, etc.).

University of Michigan Digital Library Project.

The University of Michigan Digital Library (UMDL) project has focused on creating and evaluating a testbed of a large-scale, continually evolving multimedia digital library. The core part of the research is to build a system that enables users to more easily access and customize their information inquiries through a cooperating set of three types of software agents; user interface agents, mediator agents, and collection interface agents (Cole and Harum, 1995). In order to fulfill the needs of a variety of users, the collections emphasize earth and space sciences. Although publishers will provide collection material of such type, in time, users will have the ability to publish their own material as well. In a related endeavor called the Journal Storage Project (JSTOR), ten academic journals focusing on economics are being digitized from their initial printing up through 1990. JSTOR is available to the NSF-UMDL. Industry partners for this project include Apple Computer, Eastman-Kodak, Hewlett-Packard, IBM, Sybase, McGraw-Hill, University Microfilm (UMI), etc.

University of Illinois Digital Library Project.

The University of Illinois (VIVC) project is a multifaceted effort that has attempted to build a large testbed of science and engineering literature. The project has been evaluated for effectiveness of use with many collections, and has resulted in enhanced search techniques (Schatz, 1996). The testbed, dubbed DeLiver (Desktop Link to Virtual Engineering Resources), uses an SGML document framework to expedite searching of full text articles from several engineering journals. Currently, only UIUC faculty, students, and staff have access to DeLiver. The testbed includes a compilation of articles with figures, tables, images, and equations directly from publishers in SGML. Furthermore, complex services are provided over several networks through client-server systems that rely on server software (user services), client software (user interface), gateways (relays and translators), and back-end services. Some of the

industry partners include HP, IBM, Academic Press, John Wiley & Sons, and IEEE Computer Society.

Alexandria Digital Library Project. The Alexandria Digital Library (ADL) Project, developed by the University of California, Santa Barbara, provides easy access to geographically referenced collections of maps, satellite images and digitized aerial photographs, specialized textual material and their associated metadata (Smith, 1996a). An important focus for ADL's collection is on information supporting basic science, including the earth and social sciences. The architecture of the testbed consists of a set of Internet nodes implementing combinations of the four primary components—storage, catalogue, ingest, and user-interface facilities (Smith and Frew, 1995). Each of the major components has a network interface subcomponent that provides protocols for communication over a wide area network. Active participants include DEC, ESRI, Excalibur/Conquest, Hughes, Informix/Illustra, Microsoft, Oracle, SPOT Image, and several government agencies, such as California Environmental Resources Evaluation System, Library of Congress, National Imagery and Mapping Agency, US Geological Survey and US Navy.

Stanford Integrated Digital Library Project. The Stanford Integrated Digital Library Project has sought to develop the enabling technologies for a single, integrated and universal “virtual” library that can provide uniform access to the large number of emerging networked information sources and collections (Cole and Harum, 1995). The primary focus of this project is “interoperability”, which is very important because standardization efforts are lagging behind the development of digital library services. The project uses CORBA (Common Object Request Broker Architecture) technology to implement information access and payment protocols (Paepcke et al., 1996). Information Bus (InfoBus), an abstraction layer, mediates communication between clients, sources, and services so that users can navigate and manage the “information space” in a consistent and unified way (Stanford DL Group, 1995). Some of the major partners are ACM, Bell, Knight-Ridder Information, Enterprise Integration Technologies, HP Labs, Hitachi Corp., Hughes Research Laboratory, Interconnect Technologies Corporation, NASA Ames Advanced Interaction Media Group, WAIS Inc., and Xerox PARC.

Table 1 summarizes the project foci, architecture,

contents, users, and related projects for the six DLI projects.

All of these projects, as intended by DLI, have tried to advance technologies and related research for collecting, searching and accessing digital materials. Throughout this initiative, participants have learned several lessons, which include the following (Atkins, 1997):

- There is a need for building a large and dynamic enough testbed to evaluate the operational effectiveness of the services.
- Intuitive user interface design (e.g., efficient navigation, multiple levels of abstraction and summarization, etc.) is critical.
- Design of digital libraries should be user-centric and not just technology-driven.
- Development of usability evaluation methodologies and research in economic and social implications are very important.

A new initiative, called Digital Library Initiative—Phase 2 (DLI-2), will be jointly supported by NSF, DARPA, NASA, the National Library of Medicine (NLM), the Library of Congress (LoC), the National Endowment for the Humanities (NEH) and others. Based on the lessons learned from DLI, DLI-2 intends to accelerate development, management, preservation, and accessibility of digital contents and collections. While focusing on operational digital libraries, DLI-2 encourages research on interactions between humans and digital libraries in various social and organizational contexts.

2.2. Other digital libraries in the U.S.

National Digital Library Program (NDLP) (<http://lcweb2.loc.gov/ammem/>): The NDLP is a five-year project at the Library of Congress to build an introductory core of American history and culture in digitized form. The initial material will be selected from the Library's large collection of print and non-print items (Arms, 1996). Among the items available in the Library's collection are manuscripts, maps, rare books, audio and video material, as well as photographs. One of the Library's main efforts is to provide access to its collections in digitized form via networks. Other on-going projects include an effort (called THOMAS) to make legislative information available to the public, and a prototype system (called

Table 1. Summary of DLI phase-1 projects

Project foci	Architecture	Contents	Users	Related projects
Informedia (CMU): http://informedia.cs.cmu.edu : Integrated Speech, Images and Language Understanding for Creation and Exploration of Digital Libraries				
<ul style="list-style-type: none"> • Automatic speech recognition (Hauptmann, 1995) • Network billing (NetBill) (Cox et al., 1995; Sirbu and Tygar, 1995) • Combining speech, image and natural language (Christel et al., 1995; Smith and Kanade, 1998) • Video skimming and characterization (Smith and Kanade, 1998) • Human face detection and name association (Rowley et al., 1995; Sato and Kanade, 1997) 	<ul style="list-style-type: none"> • Annotations • Client Software • Spoken query interface 	Digital video, audio, images, and text <ul style="list-style-type: none"> • WQED Pittsburgh • Fairfax Co. • VA School's Electronic Field Trips • The British Open University's BBC-produced video courses 	<ul style="list-style-type: none"> • College students • Faculty • Students from lower, middle and upper schools 	<ul style="list-style-type: none"> • NetBill Project • VASC • SPHINX /ARPA
Environmental EL (UC Berkeley): http://elib.cs.berkeley.edu : A Prototype of a Scalable, Intelligent, Distributed Electronic Library				
<ul style="list-style-type: none"> • Automated indexing & Intelligent retrieval (Wilensky, 1995; Ogle and Stonebraker, 1995; Schatz and Chen, 1996) • Developing database technology to support DL applications (Ogle and Stonebraker, 1995) • User-Centered Iterative Design (House et al., 1996) • New approaches to document analysis: MVD (Phelps and Wilensky, 1996) • Data compression and communication for remote browsing (House, 1995) 	<ul style="list-style-type: none"> • HTTP server • CGI • Illustra DBMS • Dieast doe server • File system • Metadata • Raw data 	Environmental and GIS related Data <ul style="list-style-type: none"> • Photographs • Videos • Aerial Photos • Geographic Data • Botanical Datasets 	<ul style="list-style-type: none"> • California Agencies and Organizations • U.S. Geological Survey • Industrial Partners 	<ul style="list-style-type: none"> • Berkeley Digital Library Sun-SITE • CSTR Project • NASA EOS-DIS • BADGER • Sequoia 2000 • OASIS • Access USGS
UMDL (U of Mich.): http://www.si.umich.edu/UMDL/ : Intelligent Agents for Information Location				
<ul style="list-style-type: none"> • Technical & socioeconomic issues (Crum, 1995) • Evaluation by real users (Crum, 1995) • Market based resource allocation (Mullen and Wellman, 1995; Atkins et al., 1996) • Inquiry-based education through interacting acting software agents (Birmingham, 1995; Wallace et al., 1996) 	<ul style="list-style-type: none"> • User Interface agents • Mediation agents • Collection interface agents 	Earth and space science Journal <ul style="list-style-type: none"> • Encyclopedia of science and technology • Core and popular journals (UMI) • Encyclopedia Americana • Scientific journals • Encyclopedia Britannica 	<ul style="list-style-type: none"> • Expert researchers • Students • General public 	<ul style="list-style-type: none"> • JSTOR • Elsevier PEAK • UM Gateway • Visual Image Initiative • Making of America • Humanities Text Initiative
Building the Interspace (UIUC): http://dli.grainger.uiuc.edu : Digital Library Infrastructure for a University Engineering Community				
<ul style="list-style-type: none"> • Building a large-scale information retrieval across knowledge domains (Schatz, 1995; McGrath, 1996) • Building concept space indexing for scaleable semantic retrieval (Chen et al., 1996) • Federating distributed repositories (Schatz et al., 1996) • Semantic federation across repositories (Harum et al., 1996) • Manually structured text document (Schatz and Chen, 1996) 	<ul style="list-style-type: none"> • User Interface • Collaborative Kernel • Domain manager • MCE • Concept assigner • Concept space generator • Category map generator • Validation services • Object model 	Engineering and Science Literature <ul style="list-style-type: none"> • Academic Press • AAAS/AAS • ACS/AIAA • AIP/APS • ASAE/ASCE • ASME • IEE/IEEE • IEEE Computer Society • John Wiley & Sons 	<ul style="list-style-type: none"> • Faculty and students at Univ. of Illinois • CIC consortium 	<ul style="list-style-type: none"> • Global Cultural Memory Project • MESL Project • Horizon Project • The Daily Planet™ • Astronomy Digital Image Lab (ADIL)

Table 1. (Continued)

Project foci	Architecture	Contents	Users	Related projects
Alexandria DL (UC SB): http://alexandria.sdc.ucsb.edu/ : Towards a Distributed Digital Library with Comprehensive Services for Images and Spatially Referenced Information				
<ul style="list-style-type: none"> • Collections of spatially-indexed and graphical information (Smith and Frew, 1995; Bittenfield and Goodchild, 1996) • Compatibility with standards (Smith, 1996b) • Image processing and parallel processing (Manjunath, 1998) • Automatic image processing and knowledge about region metadata (Smith, 1996b; Larsgaard, 1997) 	<ul style="list-style-type: none"> • User interface • Catalog system • Ingest system • Storage system 	<ul style="list-style-type: none"> • Digitized maps • Satellite images • Digitized aerial photos • Specialized textual material • Other graphical information 	<ul style="list-style-type: none"> • School children • Academic researchers • General public 	<ul style="list-style-type: none"> • Sierra Nevada Ecosystem • Mojave Desert Ecosystem
Stanford Integrated DL (Stanford Univ.): http://diglib.stanford.edu/ : Interoperation Mechanisms Among Heterogeneous Services				
<ul style="list-style-type: none"> • Interoperability across protocol domains, standards, and payment (Paepcke et al., 1996) • Distributed object architectures (Wang, 1998) • Information sharing and communication models (Kamiya et al., 1996) 	<ul style="list-style-type: none"> • DLITE • InterOp Protocol • Z Server • Proxy Generator • Infobus Socket Interface • Query Translator • JYLU • SenseMaker • SCAM • RManage/FIRM 	<ul style="list-style-type: none"> • Computing literature • Knight-Ridder's Dialog service • MIT Press • ACM • Stanford libraries 	<ul style="list-style-type: none"> • Stanford faculty, staff, and students • Scientists at the NASA Ames Research Center • Customers at partners • Students and researchers at other universities 	<ul style="list-style-type: none"> • CSTR/dienst Project • GIOSS Project • Boolean Query Translation • STARTS • SONIA • ComMentor

CORDS) to accept information in digitized form for copyright registration (Arms, 1996).

American Heritage Virtual Archive Project (<http://sunsite.berkeley.edu/amher/>): The American Heritage Virtual Archive Project, funded partly by the National Endowment for the Humanities (NEH), is a joint proposition between academic institutions such as University of California, Berkeley, Stanford University, Duke University, and the University of Virginia. Together, these universities are building a database of encoded searching aids called Encoded Archival Description (EAD). These EAD aids provide access to American historical and cultural collections. This project aims to develop a structure that can evaluate the prototype's "virtual archive" capability to make distributed digital library resources available, and the potential of the project's decentralized production operations. The U.S. Department of Education, Berkeley Finding Aid Project (<http://sunsite.berkeley.edu/FindingAids/EAD/bfap.html>), and the California Heritage Digital Image Access Project (<http://sunsite.berkeley.edu/calheritage/>) are involved with the design of the proposed virtual archive, and will benefit from the combination of finding aids from numerous institutions into one

database that can provide easy access to information detailing physically distributed archives.

Making of America (MOA) Project (http://moa.cit.cornell.edu/MOA/moa-main_page.html): The MOA Project aims to make a digital library of important United States historical information available over the Internet. The initial phase of this project was funded by the Andrew W. Mellon Foundation and the Charles E. Culpeper Foundation from 1994 to 1996, and was framed around the cooperation between the University of Michigan library and Cornell University library. By making scanned digital images of the complementary journals and monographs from both libraries' 19th century United States history collection, MOA plans to fully capture all signification information and make that information easily accessible on both campuses.

Computer Science Technical Reports (CS-TR) Project (<http://www.cnri.reston.va.us/home/cstr.html>): The CS-TR Project is a joint project started in 1992 involving five universities, the Library of Congress, and the Corporation for National Research Initiatives (CNRI). With help from DARPA, CS-TR is developing network access to computer science technical information held by

Cornell University, Carnegie Mellon University, MIT, Stanford University, and the University of California, Berkeley. The project has not only built an archive of information, but has also created an architecture and experimental tools for bringing a vision of seamless access to a network of distributed information more clearly into focus. Initially, CNRI and the universities were to develop digital libraries that could utilize bulk information and improve existing knowledge of information search, storage, and retrieval. The results of this project have been made available for use in numerous environments such as the Networked Computer Science Technical Reports Library (NCSTRL) project (<http://www.ncstrl.org>). The NCSTRL project hosts an international collection of computer science reports and articles provided by participating universities.

Project Open Book (<http://www.library.yale.edu/preservation/pobweb.htm>): Project Open Book is a multi-stage program at Yale University, exploring the feasibility of converting microfilm to digital imagery (about 10,000 volume digital images) through laboratory modeling processes (Conway, 1996). Access to the materials is enhanced by developing structured indexes and distributed access over Yale's campus network. The project also evaluates the issues of selection, quality, and cost. Its principal partner is Xerox Corporation.

Perseus Project (<http://www.perseus.tufts.edu/>): The Perseus Project, since 1987, has been developing a continually expanding digital library in the area of the humanities, specifically ancient world studies such as ancient Greek culture, Roman civilization, and the Renaissance (Crane, 1998). This project is supported by the Annenberg/Corporation for Public Broadcasting (CPB) Project, NSF, Apple Computer, the National Endowment for the Arts, the National Endowment for the Humanities, the Packard Humanities Institute, the Getty Grant program, Xerox Corporation, and the Fund to Improve Post-Secondary Education. Several academic institutions including Boston University and Harvard University have banded together to gather materials for the Perseus Project. These materials include maps, art catalogs, philological tools, secondary essays, ancient texts and translations which are now being shared by over 70 museums. Perseus allows viewers to examine objects in particular detail illustrating these objects in far greater detail than would be possible in a print publication.

Saguaro Digital Library Project (<http://krishna.bpa.arizona.edu/usmstar>): The Saguaro Digital Library (SDL) is being developed by a consortium of the University of Arizona research groups, Federal and State agencies in conjunction with industrial partners. The SDL directly addresses the goals of the National Biological Information Infrastructure (NBII). This project is intended to develop a comprehensive digital library system in support of a full range of services to enhance our understanding of the interdependence between the economy and the environment. The ultimate goal of the project is to allow components of the digital library to evolve independently and yet be able to call on one another efficiently and conveniently. The SDL supports heterogeneous and federated collections of digital content, including data, metadata, models, tools, and algorithms. This project also addresses the semantic gap between the vast digital repositories of data and the end user. It directly serves education, research, policy development, and the public interest by providing decision support tools for improved monitoring of ecosystem status and optimization of sustainable productivity.

2.3. DL projects in Europe

eLib Programme (<http://www.ukoln.ac.uk/services/elib/>): In response to pressures from growing student numbers and a growing base of available academic information, the Joint Information Systems Committee (JISC) of the Higher Education Funding Councils for England, Scotland and Wales and the Department of Education for Northern Ireland, started the Electronic Libraries Programme (eLib) in 1995. The eLib program comprises approximately 60 projects focusing on diverse areas of interest such as network access, document delivery, on-demand publishing, digitization, images, training and awareness, pre-prints and gray literature, and quality assurance. In addition, many projects have overlapping themes, and several other projects are involved in other digital library efforts internationally. The eLib program is more of a developmental enterprise than a research program, and its purpose is to provide electronic resources for the United Kingdom's higher education system as well as to initiate a cultural shift toward the acceptance of digital libraries and electronic resources. Unlike the fundamental research characteristics of the NSF and similar agencies, JISC's

projects are concentrated at the near-market, practical application end of the spectrum (Rusbridge, 1998)

VD17 Project (http://www.forwiss.tu-muenchen.de/public/forwiss/projekte/vd17/index_eng.html): The VD17 project is attempting to digitize 17th century printed material collected from German-speaking countries in a distributed digital database system (OMNIS/Myriad). Sponsored by Deutsche Forschungsgemeinschaft (DFG), VD17 is registering and scanning approximately 250,000 titles with over one million pixel images to jukebox-like storage media. Users will be able to search the catalog over the Internet and see the information using a traditional WWW browser. In the 10 years of the main phase of the project, 1.75 million prints are expected to be gathered.

British Library Research and Innovation Centre Project (BLRIC) (<http://www.ukoln.ac.uk/services/bl/>): The BLRIC project is studying the ways that networking technology can improve library and information services, including difficulties faced by library staff and patrons in this era of rapid technological change. The project is bringing together universities, libraries, and outside experts through conferences and seminars to further the study and exchange of information. To gain insight from international researchers, BLRIC keeps close ties with the Nordic council for Scientific Information (NORDINFO) and the British Library's European counterpart. Their digital library hosts a core of digitized information, sounds, and images that are available on demand anywhere in the world, and the information can be hosted and dispatched by more than one institution.

Mediterranean Digital Library (MEDLIB) Project (<http://www.unesco.org/webworld/mediter/medlib.htm>): The MEDLIB Project is an effort by United Nations Educational, Scientific, and Cultural Organization (UNESCO) to build a digital library for the Mediterranean region through cooperation with scholars and information reserves around the Mediterranean Sea. MEDLIB digitization efforts are focused on area's national and regional heritage (e.g., manuscripts, famous literary, old stocks, philosophical and religious works, etc.) and contemporary issues important to the region (e.g., environment, population migration, hydrological or mining resources, etc). To establish open network, this project aims to set up policies and mechanisms for the

coordinated development of information sources to be implemented by the participating institutions.

2.4. DL projects in pacific region and other countries

New Zealand Digital Library (NZDL) (<http://www.nzdl.org/>): The NZDL project is a research program in the Computer Science Department at Waikato University. Its purpose is to develop digital library technology, and enable other organizations to create their own digital collections (Witten et al., 1996). NZDL has built several demonstration digital libraries that are available over the WWW for computer science technical reports, literature, and frequently asked questions about the Internet. The ultimate goal of the project is to develop systems that can automatically build a framework out of uncatalogued, unstructured distributed repositories of information.

Cooperative Research Centre for Distributed Systems Technology (DSTC) Project (<http://www.dstc.edu.au/>): Started in 1992 and scheduled to finish in 1999, the DSTC project is a collaborative effort funded by the Australian Government's Cooperative Research Centres Program and more than thirty other organizations for the development of a technological foundation for future global distributed systems. Key areas of focus for DSTC are Java, XML, Internet searching, knowledge management, collaborative computing, and security.

Japanese Electronic Library Pilot (JELP) Project (http://www.cii.go.jp/el/index_e.html): As one of the three main projects under the Center for Information Infrastructure (CII) in Japan, the JELP project, jointly sponsored by the Information-technology Promotion Agency (IPA) and the National Diet Library (NDL), offers experimental electronic access to a variety of books and other resources existing in libraries and tests the latest advances in data-processing and network technologies to prepare for the establishment of future electronic libraries. This project is conducting two experiments: (1) National Union Catalog Network Experiment that is verifying methods for locating target documents via networks and procedures for efficient system operation, and, (2) Electronic Library Practical Experiment that focuses on digitizing information recorded on prints and microfilm materials and accumulating the digitized data into a database.

NACSIS Electronic Library System (NACSIS-

ELS) (<http://www.nacsis.ac.jp/els/else.html>): NACSIS-ELS is a digital library funded by National Center for Science Information Systems (NACSIS) of the Ministry of Education, Science, Sports, and Culture in Japan since 1995. The library provides primary information such as catalogue and bibliographic information from Japanese academic journals, especially scientific papers. The primary users are: (1) researchers and postgraduate students in universities, junior and technical colleges in Japan, (2) researchers in inter-university research institutes in Japan, and (3) researchers at academic societies cooperating with NACSIS-ELS in Japan.

Korean Digital Library Pilot (KDLP) Project (<http://www.dlibrary.or.kr/>): The KDLP project is one of the information superhighway projects sponsored by the Ministry of Information and Communication in Korea. The project is a five-library consortium consisting of The National Library of Korea, The National Assembly Library, The Science Library in KAIST (Korea Advanced Institute of Science and Technology), The Korea Research and Development Information Center in KAIST, and Korea Research Foundation. This project was developed for systematic servicing of full-text information for users by providing useful information for academic and research activities through the Internet. The primary objectives of the project are (1) to develop a common software system among participating digital libraries, (2) to implement full-text databases within the limitations of copyright law, and (3) to develop a digital library system by which user can access non-stop information services.

Timely information for All, Relevant and Affordable (TiARA) (<http://www.digilib.org.sg/>): TiARA is a multiple agency effort of the National Computer Board, the National Library Board, the National Science and Technology Board and participating libraries in Singapore. TiARA provides access to library catalogues, online databases, Internet resources and a calendar of events. TiARA also alerts users to new articles, books and magazines in their interest area. It consists of online data from content providers (e.g., Dialog Corporation, Faulkner, Engineering Information and Elsevier Science) and library catalogues from various Singapore libraries.

Canadian Initiative on Digital Libraries (CIDL) (<http://www.nlc-bnc.ca/cidl/cidle.htm>): The CIDL project, a coalition of Canadian libraries proposed in 1997, strives to coordinate the digital collections in

Canada in order to improve long-term access to digital library services. CIDL's objectives are to educate the public in digital library issues, promote digital library standards, coordinate activities to avoid duplicating digital information, build close ties with informational organizations such as museums, publishers, creators, and government agencies, and advance awareness of Canada's digital library efforts.

Although many other interesting digital library development efforts are underway in the areas of biological and medical digital libraries, these projects are not discussed in this paper. Readers are, however, encouraged to visit them and to explore this exciting area of development. Some representative digital libraries include GALEN II (<http://www.library.ucsf.edu/>), the National Library of Medicine (NLM)'s MEDLINE plus (<http://medlineplus.nlm.nih.gov/medlineplus/>); Organising Medical Networked Information (OMNI) (<http://omni.ac.uk/>), healthfinder^R (<http://www.healthfinder.org/aboutus/default.htm>), BioMedNet's free MEDLINE (<http://www.biomednet.com/>) service, Princeton University Biology Library (<http://www.princeton.edu/biolib/>), and Marine Biological Laboratory (MBL) (<http://www.mbl.edu/>).

3. Research Issues in Digital Library Development and Deployment

Research in digital libraries has resulted in a wide range of technological development, which is just a prelude for reshaping the way science and education are practiced. In the development of digital libraries, much of the research has focused on technical issues. Libraries are social institutions that have supported rich interactions among information providers, intermediaries including librarians, information, and information seekers for many centuries. We believe that digital libraries are logical extensions of traditional libraries. Therefore, the development of digital libraries should be grounded in a firm understanding of all the components and interactions among these components as depicted in Fig. 1. Fig. 1 illustrates major components that have been drawn from the results of an analysis of the on-going digital library development efforts described in the previous section.

As explained earlier, a digital library is a

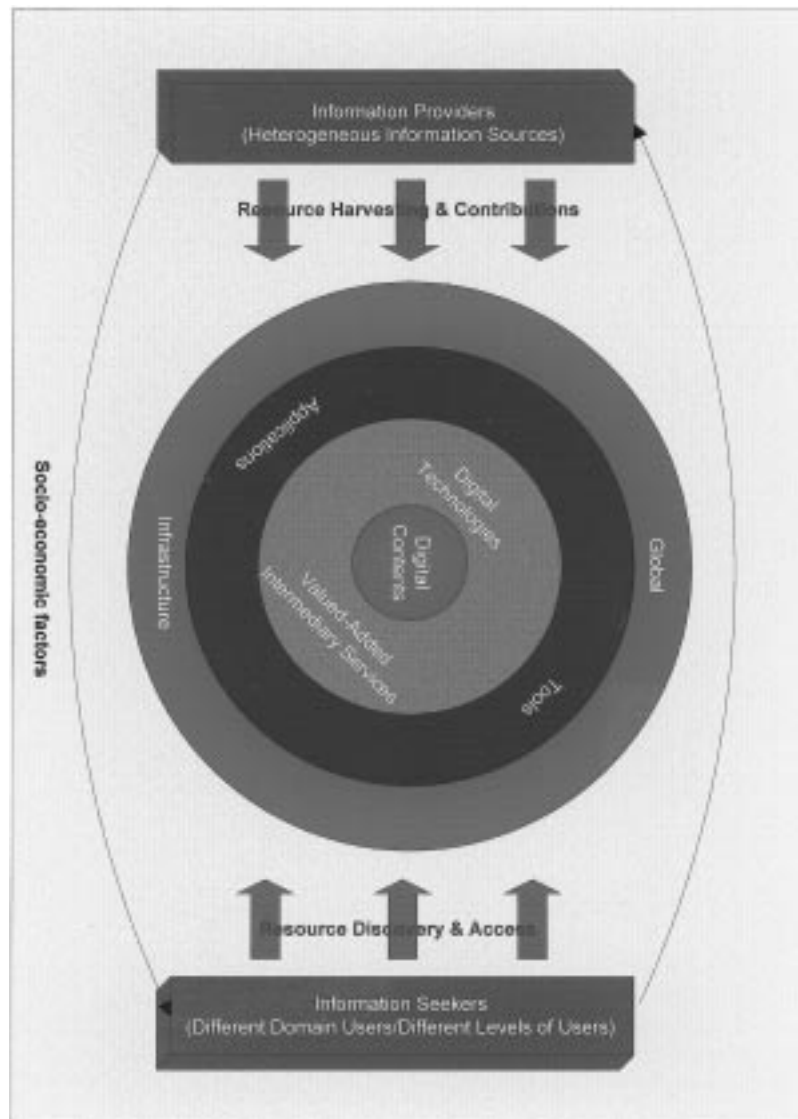


Fig. 1. Key components in digital library environment.

networked system environment where individual components interact with each other to allow users to submit and access digital contents. Digital contents refer to digitized text, image, audio and video information stored in the digital library. Users can be information seekers, information providers, or both. Information seekers are those who access and acquire information from the digital library, while information providers are those who provide data sources to the digital library. Such data sources

contributed by information providers are either raw data or processed information. If the data source has not been processed and needs to be processed, value can be added by intermediary services and digital technologies. Such value-added intermediary services include search, classification, filtering, translation, and publishing (Brewer et al., 1996). Detailed discussion of intermediary services can be found in Brewer et al. (1996) and Wiederhold (1995). These services and digital technologies are aided by various

applications and tools to ensure quality and reliability of the digital contents and to assist users (both information providers and seekers) who have vastly different backgrounds, information needs and levels of expertise. We believe that the components illustrated in Fig. 1 are the key to opening the library and its resources to more users.

The solutions to fundamental problems in the design of digital libraries can benefit not only from the study and development of each component, but also from the facilitation of interactions among such components. It is therefore important to articulate potential obstacles we should overcome during the development and deployment of digital libraries.

3.1. Interoperability

Paepcke et al. (1998) view "interoperability" as the most important issue when building digital libraries since collections of independently developed components rely on each other to accomplish larger tasks. There are two different types of interoperability: syntactic interoperability and semantic interoperability. *Syntactic interoperability* is the ability of multiple software components to cooperate although their implementation languages, interfaces, and execution platforms are different. We believe, however, that emerging standards (i.e., CORBA, DCOM, Z39.50, HTTP, KIF, etc.) can resolve most of the syntactic level interoperability. *Semantic interoperability* is the ability of participating system domains to understand the meaning and use of terminology from different domains and their axiomatic mapping ability between agreed concepts to make a semantically compatible information environment. Establishing semantic interoperability among heterogeneous and disparate information sources has been a critical issue in the past two decades largely in the database community and will be even more important for digital library community.

Much of the existing data in any digital library has problems with semantic interoperability. One of the important issues in semantic interoperability is how to identify and resolve various data and schematic-level conflicts among disparate information sources. Conflicts arise when semantically similar information is represented by different names and meanings, different unit measurements, different levels of precision, and/or different data structures among information sources. Good understanding of semantics of the information components, such as attributes,

object classes, and relationships between object classes is a key to resolving such conflicts. The result provides interoperability among different information sources. A well-defined classification scheme for semantic conflicts can provide clear guidelines in understanding and resolving different types of conflicts (Sheth, 1991). Although classifications of semantic inconsistencies appear in traditional multidatabase literature (Kim and Seo, 1991, Ventrone and Heiler, 1991, Kashyap and Sheth, 1996), no classification scheme has been proposed for the wide variety of resources available through digital libraries. Ram and Park (1997) have proposed a comprehensive framework for classifying semantic inconsistencies in heterogeneous geographic databases at both the data and schematic level. These techniques can be easily extended to support digital libraries.

As previously stated, Stanford's digital library architecture, called InfoBus, is based on the premise of a hardware bus that provides services for information storehouses and clients to interoperate through functions built into the testbed (Paepcke, 1996). This approach tries to solve syntactic interoperability. The architecture is implemented using the CORBA distributed-object technology through Xerox PARC's InterLanguage Unification facility (ILU) and Inprise's Visbroker, in the form of an information retrieval and payment protocol called InterPay (Paepcke et al., 1996). Digital Library Interoperation Protocol (DLIOP) is an asynchronous protocol that increases robustness of information retrieval during network or server outages. This protocol has been adopted by other Digital Library Initiative participants such as the University of Michigan and the University of California, Santa Barbara.

The Saguaro Digital Library is developing a set of techniques to provide both syntactic and semantic interoperability (Ram and Park, 1998b). This project uses standard architectures and protocols (e.g., CORBA, Z39.50, Java RMI, HTTP, etc.) to allow communication between library components and collections. To provide semantic interoperability, semantic mediators are being developed. The best way to provide very high levels of autonomy for components of the digital library is to locate the interoperability machinery outside the participating local systems to mediate between components. Wiederhold (1992) defines a mediator as a software module that exploits encoded knowledge about a

particular dataset to bring the source information into a common form for a higher layer of applications. Mediators not only provide methods to access and integrate data from multiple databases, but also facilitate interoperability among disparate information sources, because each mediator contains specific domain knowledge for the interoperability, such as mapping between source database schemas and the integrated schema. In the Saguaro Digital Library, semantic mediators automatically detect and resolve data level and schema level conflicts from multiple sources. These mediators use an ontology called SCROL (Semantic Conflict Resolution OntoLogY) to resolve conflicts (Ram and Park, 1998a).

3.2. Knowledge representation and ontologies for mediation

The use of standard ontologies is critical in digital libraries to provide semantic interoperability because knowledge of the domain specific context is important in understanding the meanings of terms. Descriptions, terms, and the conditions for using information in goods or services make up one of the crucial issues in the University of Michigan Digital Library. This library has structured procedures for describing information objects and services that are the basis for a metadata framework called *Conspectus* (Atkins, 1996). Through the use of a modified intellectual work hierarchy originally developed by the International Federation of Library Associations (IFLA), the project has been able to create a high level ontological description of intellectual work.

In the Saguaro Digital Library, SCROL is used to facilitate knowledge sharing and reuse among mediators as well as to assist semantic mediators automatically detect and resolve semantic conflicts. The concepts/sub-concepts and instances from the shared ontology are mapped onto an individual schema, and then manipulated by domain specific semantic mediators. For example, when the temperature of a certain area is requested and the results are returned, the semantic mediator will automatically detect (by traversing the ontology) that the units of measurements are in conflict (Celsius in one case and Fahrenheit in the other). Using a conversion algorithm, it converts to the unit required by the user and offers a comprehensive set of results to the user. In addition, the semantic mediator can automatically detect schema level conflicts because SCROL stores mapping information for each schema. It is important

to note that it may not always be possible to resolve every conflict. In cases where resolution is impossible, the semantic mediator simply passes on details of the type of conflict to the user.

3.3. Metadata

Metadata refers to summary information about the content, organization, or purpose of data. The primary role of metadata is to enable users to access information that is relevant. However, because the organization and use of metadata rely heavily on human judgement, domain knowledge, intended use, purpose and classification policy, different information sources may use different metadata formats and standards. For instance, metadata may be the structural description of the content of the data (e.g., creator, creation date, classification code, abstract, and keywords), or the description of the physical aspects of the data (e.g., data type, data format, data size, and location). Due to differences in the contents of metadata, it is problematic to combine and use metadata from different information sources.

Since the contents of metadata significantly affect the browsing, searching, indexing, and cataloging capabilities, it is necessary to establish a well-defined metadata standard. There are several metadata standards in existence. Some examples of these relevant to digital libraries are: the U.S. Machine Readable Cataloging (US MARC) Standard (<http://www.tlcdelivers.com/tlcrs/gen0001.htm>) that is used for representing and communicating bibliographic and related information in machine-readable form, the Federal Geographic Data Committee (FGDC) Standard (<http://www.fgdc.gov/index.html>) that describes different types of geospatial standards, and Dublin Core (<http://purl.org/dc/>) which is a content description model for textual resources.

The Alexandria Digital Library (ADL) uses a catalog with components from the US MARC and FGDC standards which are both useful for representing spatial metadata. FGDC is a federally commissioned standard used since 1995 for sharing geospatial metadata. Its main benefits are that it informs users of what data is available, whether or not the data meets their criteria, where the data is located, and how to retrieve the data. The US MARC standard, on the other hand, is a group of codes and functions for encoding machine-readable records—each record being comprised of a record structure, content designation, and data content. The metadata for a

US MARC holding is kept in one record containing the following four elements: a leader, a record directory, control fields, and variable fields.

The Beethoven Project, a part of UMDL, utilizes metadata generated from US MARC records on works by or about Beethoven. The generated metadata is then mapped to an ontology. This knowledge-based metadata is used for reasoning about the data to establish ontogenic relationships. The metadata is also used for online textual, image, and multimedia materials, which is updated and modified in conjunction with changes in the ontology. The Stanford Digital Library has developed a metadata architecture because InfoBus facilities require a generalized approach for interoperability. For example, metadata is required to determine what collections are relevant for a particular task and to understand their surrounding context. The metadata architecture provides “Attribute Models” which are generalizations of the attribute sets, such as Dublin Core, and “Attribute Model Translators” which performs translation between Dublin Core and US MARC.

In the Saguaro Digital Library (SDL), a semantic data model is used to describe the various objects contained in each dataset or collection, and their interrelationships. This semantic data model is a constitutive part of the metadata for the SDL. It describes concepts in terms that users understand, such as forests, watersheds, and spatial relationships. Therefore, the metadata, along with the ontology, plays a key role in supporting interoperability. Since the proficiency of metadata contributors directly impacts the usefulness of the data itself, the SDL is exploring factors influencing user participation and quality of input. Ultimately, this project aims to establish a system for metadata review.

3.4. User interface

One of the most important components in digital libraries is an easy-to-use user interface. The interface for digital libraries must provide a uniform interface and intelligence to support rich interaction between various levels of users and the resources that reside in the digital library. Users may range from research scientists, domain experts, college students, and K-12 students, to the general public. Thus, the user interface should be easy to use and robust enough to effectively display various data types (e.g., text, images, audio, and videos) in a single interface if necessary. The user interface must be flexible and highly customizable so

that users can save not only their entire interaction session but also the presentation output format. Displaying text and images in the same user interface may delay the presentation of some portion (e.g., heavy graphical images vs. simple text display), hence the user interface (or user interface intelligent agent) should be aware of this and be able to adjust the delivery of output.

The Saguaro Digital Library supports users who have different levels of expertise in the use of the library, as well as in the domain of resources provided by the library. For example, K-12 students may use the library to learn about fire risk and stewardship of natural resources. Policy makers may use the data, models and tools in the library to develop integrated landscape maps for strategic fire vulnerability assessment. Senior researchers may use the library to study the socioeconomic risks associated with fire. At the same time, the library should have the ability to capture information about the user (to create a user profile) and allow him to expand his use of the library as he becomes more familiar with the materials and the search procedures. The primary objective of the SDL is to make resources available to end users who do not have the expertise in remote sensing or geographic information systems. We believe this is the key to opening the library and its resources to more users.

To provide this type of “contextual” assistance to users, the SDL uses a set of collaborating intelligent software agents (Ram et al., 1997). This type of contextual assistance to users distinguishes the SDL from previous approaches such as the Alexandria Digital Library (Smith, 1996a), Interspace (Schatz et al., 1996), Informedia (Wactlar et al., 1996), the Stanford Integrated Digital Library (Paepcke et al., 1996), and the Environmental Electronic Library (Wilensky, 1996). The following table (Table 2) illustrates various user interface components, their characteristics, and technologies used in the major projects funded through the DLI—Phase 1 Initiative.

3.5. Efficient query processing

In movies, searches on the Internet always return what you are looking for no matter how vague your keywords are. For instance, in the movie, “Mission Impossible,” Tom Cruise searches with keywords like “file” and “computer” and three results are returned. In reality, this is possible only if there are efficient query mechanisms to retrieve the most relevant

Table 2. The analysis of user interface in DLI—Phase 1 projects

Digital Library	Components	Characteristics	Technologies
CMU Informedia (IDVLS)	Query (Video Metadata, Multimedia Abstractions), Information Visualizers (Survey of Past Work, VIBE. Information Filtering), etc.	<ul style="list-style-type: none"> • Parallel presentation • Context sizing slide switching • Skimming dial • Synthetic interviews • Reuse • Select and view multiple layers of geographic data 	Dynamic HTML, JavaScript
Berkeley (GIS Viewer)	Panning, Zooming, GIS Layer Selection, Query, Annotations, Editing and Removing Layers, Saving Configurations, etc.	<ul style="list-style-type: none"> • Pan and zoom to select a region and level of resolution • Browse and annotate network-ready photographs as well as geographic data 	CGJ, Java
Michigan (ARTEMIS)	Past Searches, Keywords Hierarchy, Recommender, Web Collections, Abstracts, Driving Question Folders, etc.	<ul style="list-style-type: none"> • Interface for K-12 students to the UMDL collections • Provide access to a persistent, individual workspace 	Java, HTML, CORBA, ILU, KQML
Illinois Interspace (DeLiver)	Search Screen, Quick Tips Page, In-Depth Help Pages, Results Page, DLI Journal Titles, Table of Contents of IEE Electronic Letters, etc.	<ul style="list-style-type: none"> • Full text searching • Controlled vocabulary searching • Thin, browser-neutral web client; n-tier server approach 	HTML, DDE, OLE, CCI, SGML, XML, PDF
Alexandria (JiGi)	Welcome & Login, Map Browser, Search Options, Workspace Component Metadata Browser, etc.	<ul style="list-style-type: none"> • The user can select items and build personal folders of information. • An easy process for downloading the client programs • Direct manipulation metaphor of iconic representations 	HTML, Java (JFC/Swing, JTable, JTree, aelfred parser), XML, HTTP, VRML
Stanford (DLITE)	Documents, Collections, Queries, Constructors, Services, Representations of People, etc.	<ul style="list-style-type: none"> • Distributed, client/server application • Support sharing, re-use, and persistence • Deploy drag-and-drop style interface 	Python/Tk, Java/AWT, CORBA, ILU.

information from multiple collections in the digital library. Many efforts are underway to develop efficient query schemes. Rao et al. (1995) describe the concept of iterative query refinement, in which the process of query formulation is repetitive. The user submits a query, browses the results, and then reformulates the query until the results contain the most relevant information. This is an interactive learning process if the user is not sure how to formulate the exact questions or express them effectively. Other approaches use agent technology. While the previous approach requires intensive user involvement since iterative query refinement is a user-guided activity, the agent-based approach can significantly reduce user involvement because the agent can automatically perform most of the content filtering and evaluation tasks.

Every data object in the Alexandria Digital

Library (ADL) has individual catalog entries, and is made up of bibliographic records at the item level. Each of these records may also have information common to all the data objects belonging to a single publication project, also known as parent catalog entries. For instance, the ADL gazetteer, a place-name index that allows users to locate earth features by entering the name associated with that feature, can be used as a spatial search tool or as a stand-alone reference.

The Saguaro Digital Library uses “contextual search” mechanisms based on agents. As described previously, since each user visits the library with a different purpose, the agent extracts the needs of the user through a dialog that assists the user in formulating queries, and uses it to establish the context of the search. This is necessary because some users are very clear about what they need, while others

do not know exactly what to look for. The context of the query is established by asking the user a number of questions relating to his topic of interest, the intended use of the information (i.e., do a prescribed burn or write a term paper on wildland fires), the level of expertise of the user, the time constraints for the search, the level of quality required and the extent of fee that he/she is willing to pay for the resources. The user is also allowed to explore the "information space" of the digital library showing the types of resources available, and determine what he needs (Ram and Ramesh, 1998, Ram and Shankaranarayan, 1998).

3.6. Resource harvesting and dynamic evolution

Most digital libraries are mausoleums, i.e., static information sets that require periodic updates. Updates are often ephemeral and costly. In reality, a digital library should support dynamic resources that can be updated by users or an information provider at any time. Using appropriate data management techniques to insure quality, security, searchability and a means to generate metadata information, it is possible to create dynamic digital libraries. To collect the information, it is essential to provide "digital forms" for the appropriate information set and develop update rules. In addition, the evolution of the library needs to be supported by tools that allow new resources to be added and old/obsolete resources to be removed periodically.

Providing mechanisms for resource harvesting and supporting dynamic evolution of digital libraries would have enormous practical benefits to the general public at large. Some of the benefits are: (1) reduced cost of obtaining updated information, (2) direct communication with the end user to discover what information is used and why, (3) creation of a cooperative atmosphere to encourage information generators and users to exchange ideas and information, (4) increase in the cost effectiveness of the research effort by expediting and increasing the flow of information, and (5) providing a mechanism for direct public participation in the information gathering process. It further creates an environment where users and the public have a significant role and responsibility to add information that they deem necessary to effect the outcome maps or decision surfaces that may be created through information exchange.

The Berkeley digital library supports the idea, in

part, of resource harvesting by allowing untrained users to furnish environmental information (House, 1996). People interested in becoming data providers can supply relevant information to the library through online access once they are approved by a current data provider who is an expert in the field. The Saguaro Digital Library is attempting to develop a "harvesting agent (HA)" in support of resource harvesting. Several types of resources may be added to the library as it grows and evolves over time. Resource providers interact with the HA which conducts a dialog with the user through the WWW GUI using its "harvesting rule base." The latter contains a set of rules to direct questions, determines what type of resource the provider wishes to add, and invokes the appropriate harvesting tool and quality control tool to collect information about the new resource or update to an existing resource.

3.7. Socioeconomic issues

Research in digital libraries should deal with many other complex social and economic issues, as well as legal issues. These include, education, information and computer literacy, human resources, cultural and multilingual issues, conservation and preservation of digital materials, copyright, and privacy. As mentioned earlier, libraries are social institutions in which people interact with each other, using information resources and facilities that libraries provide. Digital libraries should reflect the social context in which they exist.

Marchiononi and Maurer (1995) discuss the social and intellectual roles of a library, and mention that a library is a physical place for people with different perspectives to meet and interact in a knowledge place. Digital libraries should extend such a social function by providing a "virtual place" where individuals can interact with each other and share information resources. For example, digital libraries can provide collaboration technologies or computer-mediated communication support that can act as a forum for facilitating information and knowledge exchange and retention for a particular user group or the entire user community around the globe.

As the use of digital materials increases, a major issue that has surfaced is how to pay for accessing and obtaining digital information. As digital libraries become involved in priced services, many interesting questions can be asked, such as: how much should the digital library charge its users? Should the digital

library allow free access to all collections, or associate materials with different price tags based on access time or quantity? It is clear that digital technologies have changed the cost structure of the publishing industry. While the costs of publishing and distribution of digital materials are significantly less than printed materials, overhead costs for value-added services (e.g., multimedia content services) and maintaining digital libraries are much higher than printed materials. The economic model of digital libraries is unclear. This important topic requires more attention in digital library research.

There are several different pricing mechanisms and cost models in existence. For example, some digital libraries use subscription and time based models so that users can have unlimited access to the service within the subscription period. The ACM Digital Library (<http://www.acm.org/dll>) and the IEEE Computer Society Digital Library (<http://www.computer.org/epub/>) are popular examples of this type of model. Some commercial publishers such as Link (<http://link.springer.de/>) provide services to libraries or other defined groups based on institutional or IP addresses, while other publishers such as ACM SIGMOD (<http://www.acm.org/sigmod/>) provide access to its electronic journals based on membership. The question is whether information should be freely disseminated or charged. Traditionally, public libraries do not charge their users, thus it is difficult for them to find funding sources to provide new services. Should digital libraries charge their patrons? Should they cooperate with each other or compete with other electronic publishers? The bottom line is that we should deliver integrated and unified services and provide access to various resources available for our community users.

3.8. Legal issues

As discussed in Samuelson (1995), establishing new copyright laws is an arduous task of balancing between an author's and his publisher's interests (i.e., receiving formal attribution of their work and compensation), and a user's and the library's interests (i.e., having public access to information products at reasonable terms). Copyright law prohibits copying and distribution of copyrighted work without permission from its owner. New technologies have now made it easier to access and reproduce digital materials compared to photocopying printed materials. These replicated digital materials can be easily

redistributed electronically worldwide in a short period of time. Abuse of digital materials (e.g., copyright infringement) can stunt the development of new intellectual activities.

Several projects are investigating copyright and intellectual property issues in digital libraries. An example of a technical contribution to this subject is the RightsManager System (<http://dmlwww.cwru.edu/doc/rmgim.htm>) in the Case Western Reserve University (CWRU) digital library project. This pilot implementation is a joint effort undertaken by CWRU and IBM. The main goal of this system is to provide intellectual property management that is intended to protect intellectual property owners' rights and users' privacy. To achieve this goal, the system has adopted several safeguard mechanisms, such as envelopes for safe document delivery and authenticity control, encryption keys, and document markings to identify the document source ("watermark") and its destination ("fingerprint") (Gladney and Lotspiech, 1997). The system server is designed to act as both a content repository and as a clearinghouse for all client requests for access to copyrighted digital materials (Alrashid et al., 1998).

Another copyright issue is related to plagiarism on the Internet (Denning, 1995). Since research work can be easily accessed and downloaded via the Internet, anyone can copy an author's work without proper authorization. This is particularly true of work-in-progress papers and technical reports because such items can be easily copied and wrongfully attributed to another person. The Stanford digital library project developed the SCAM (Stanford Copy Analysis Mechanism) copy detection engine to detect copyright violations in digital libraries (<http://www.db.stanford.edu/~shiva/SCAM/scamInfo.html>). This tool has been proven to be effective in finding several instances of plagiarism.

Privacy is another important issue in digital libraries. Traditionally, libraries are responsible for protecting user privacy and are not legally allowed to reveal personal records of user activities in the library (e.g., checked-out items, recalls, billing information, and other personal information). However, computer systems and software (e.g., agent) can easily track and monitor user activity and store such information in user profile databases or user log files for later use by the system. This type of information can typically be used to fine-tune system performance and to provide

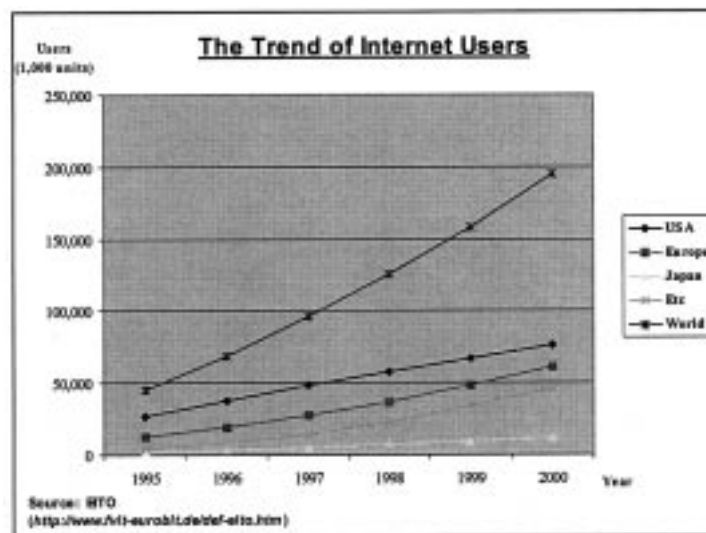


Fig. 2. The trend of internet users.

better user services. However, it is unclear that the use of such information is in favor of the user. Moreover, misuse of personal information should be prevented. Appropriate legislation and policy for this privacy issue should be established to dictate the use and dissemination of user information.

4. Conclusion

The number of Internet users around the world is growing exponentially. The European Information Technology Observatory (EITO) (<http://www.fvit-eurobit.de/def-eito.htm>) has reported that, by the year 2000, nearly 200 million people around the world will have Internet access. Fig. 2 shows that the number of users is expected to quadruple from 1995 to January 2000. The increasing trend of Internet users provides digital libraries with both opportunities and risks. As the number of Internet users increases and more information sources are made available on the Internet, we may experience "information overloading." It is already getting very difficult to find information that is relevant to the user. In this sense, effective and intelligent search mechanisms are very important. Another problem associated with information overloading is information quality. It is apparent that, without value-added services, Internet users may have higher risks in terms of information quality. In

such an environment, well-developed digital libraries can play a key role in providing high quality digital information to users.

Digital libraries require research in many other areas including dynamic interoperability, support for library evolution, contextual search mechanisms, and social issues. Such research requires interdisciplinary efforts from information systems, computer science, library science, management science, social science, and other fields. In this paper, we presented several research issues and on-going efforts that span these disciplines. We believe that digital libraries are ripe with research opportunities, offer many challenges, and will continue to grow in the next several years.

References

- Alrashid TM, Barker JA, Christian BS, Cox SC, Rabne MW, Slotta EA, Upthegrove LR. Safeguarding Copyrighted Contents: Digital Libraries and Intellectual Property Management. *D-Lib Magazine* April. (<http://www.dlib.org/dlib/april98/04barker.html>) 1998.
- Arms CR. Historical Collections for the National Digital Library: Lessons and Challenges at the Library of Congress. *D-Lib Magazine* April/May. (<http://www.dlib.org/dlib/april96/loc/04c-arms.html>, <http://www.dlib.org/dlib/may96/loc/05c-arms.html>) 1996.
- Atkins DE. The University of Michigan Digital Library Project. *D-Lib Magazine* July/August. (<http://www.dlib.org/dlib/july96/07atkins.html>) 1996.
- Atkins DE. Report of the Santa Fe Planning Workshop on

- Distributed Knowledge Work Environments: Digital Libraries. (<http://www.si.umich.edu/SantaFel/>) 1997.
- Atkins DE, Birmingham WP, Durfee EH, Glover EJ, Mullen T, Rundensteiner EA, Soloway E, Vidal JV, Wallace R, Wellman MP. Toward Inquiry-Based Education Through Interacting Software Agents. *IEEE Computer* 1996;29(5):69–76.
- Birmingham WP, Drabenstott KM, Frost CO, Warner AJ, Willis K. The University of Michigan Digital Library: This is not your father's library. *Proceedings of the 1st Annual Conference on the Theory and Practice of Digital Libraries*. (<http://www.csdl.tamu.edu/DL94/paper/umdl.html>) 1994.
- Birmingham WP. An Agent-Based Architecture for Digital Libraries. *D-Lib Magazine* July. (<http://www.cnri.reston.va.us/home/dlib/July95/07birmingham.html>) 1995.
- Brewer A, Ding W, Hahn K, Komlodi A. The Role of Intermediary Services in Emerging Digital Libraries. *Proceedings of the 1st ACM international Conference on Digital Libraries* Bethesda, MD, March 20–23, 1996, pp. 29–35.
- Brown C. AI on the WWW: Supply and Demand Agents. *IEEE Expert* 1995;10(4):50–55.
- Buttenfield BP, Goodchild MF. The Alexandria Digital Library Project: Distributed library services for spatially referenced data. *Proceedings of GIS/LIS '96* Denver, Colorado 76–84, 1996.
- Chen H, Schatz BR, Ng TD, Martinez JP, Kirchhoff AJ, Lin C. A Parallel Computing Approach to Creating Engineering Concept Spaces for Semantic Retrieval: The Illinois Digital Library Initiative Project. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 1996;18(8):771–782.
- Christel M, Kanade T, Mauldin M, Reddy R, Sirbu M, Stevens S, Wactlar H. Informedia Digital Video Library. *Communications of the ACM* 1995;38(4):57–58.
- Cole T, Harum S. Digital Library Projects. *LITA Newsletter* 1995;16(2): (<http://www.lita.org/newslett/v16n2/diglibs.html>)
- Conway P. Yale University Library's Project Open Book. *D-Lib Magazine* February. (<http://www.dlib.org/dlib/february96/yale/02conway.html>) 1996.
- Cox B, Tygar JD, Sirbu M. NetBill Security and Transaction Protocol. *Proceedings of the USENIX Workshop on Electronic Commerce*. (<http://informedia.cs.cmu.edu/pubs/netbill.pdf>) 1995.
- Crane G. The Perseus Project and Beyond. *D-Lib Magazine* January. (<http://www.dlib.org/dlib/january98/01crane.html>) 1998.
- Crum L. University of Michigan Digital Library Project. *Communications of the ACM* 1995;38(4):63–64.
- Denning PJ. Editorial: Plagiarism in the Web. *Communications of the ACM* 1995;38 12:29.
- Etzinoni O, Weld DS. 1995. Intelligent Agents on the Internet: Fact, Fiction, and Forecast. *IEEE Expert* 1995;10(4):44–49.
- Fox EA, Akscgn RM, Furuta RK, Leggett JJ. Digital Libraries. *Communications of the ACM* 1995;38(4):23–28.
- Frew J, Freeston M, Kemp RB, Simpson J, Smith T, Wells A, Zheng Q. The Alexandria Digital Library Testbed. *D-Lib Magazine* July/August. (<http://www.dlib.org/dlib/july96/alexandrial/07frew.html>) 1996.
- Gladney HM, Lotspiech JB. Safeguarding Digital Library Contents and Users: Assuring Convenient Security and data Quality. *D-Lib Magazine* May. (<http://www.dlib.org/dlib/may97/ibml/05gladney.html>) 1997.
- Griffin SM. NSF/DARPA/NASA Digital Libraries Initiative Projects. Division of Information and Intelligent Systems, National Science Foundation (NSF). (http://www.cise.nsf.gov/iis/dli/_home.html) 1998.
- Gruber TR. A Translation Approach to Portable Ontologies. *Knowledge Acquisition* 1993;5(2):199–220.
- Harum S, Mischo W, Schatz BR. Federating Repositories of Scientific Literature: An Update on the Digital Library Initiative at the University of Illinois at Urbana-Champaign. *D-Lib Magazine* July/August. (<http://www.dlib.org/dlib/july96/07harum.html>) 1996.
- Hauptmann AG. Speech Recognition in the Informedia Digital Video Library: Uses and Limitations. *Proceedings of the 7th IEEE International Conference on Tools with AI(ICTAI-95)* 1995.
- House NA. UC Berkeley's NSF/ARPA/NASA Digital Libraries Project. *Proceedings of Allerton Conference*. (<http://edfu.lis.uiuc.edu/allerton/95/s3/vanhouse.html>) 1995.
- House NA, NA, Butler MH, Ogle V, Schiff L. User-Centered Iterative Design for Digital Libraries. *D-Lib Magazine* February. (<http://www.dlib.org/dlib/february96/02vanhouse.html>) 1996.
- Kamiya K, Röscheisen M, Winograd T. Grassroots: A System Providing a Uniform Framework for Communicating, Structuring, Sharing Information, and Organizing People. Computer Science Department, Stanford University. (<http://walrus.stanford.edu/Grassroots>) 1996.
- Kashyap V, Sheth A. Semantic and Schematic Similarities Between Database Objects: A Context-Based Approach. *The VLDB Journal* 1996;5(4):279–304.
- Kim W, Seo J. Classifying Schematic and Data Heterogeneity in Multidatabase Systems. *IEEE Computer* 1991;2412:12–18.
- Larsgaard ML. Providing access to Earth-sciences spatial data: metadata and the Alexandria Digital Library. In: DeFelicce BJ (Ed.). 1996. *Proceedings of Annual Meeting of the Geoscience Information Society in conjunction with Annual Meeting of the Geological Society of America* Denver Alexandria, VA: Geoscience Inf. Soc. 27–30, 1997.
- Manjunath BS. Image processing in the Alexandria Digital Library project. *Proceedings of the IEEE International Forum on Research and Technology Advances in Digital Libraries (ADL'98)* Santa Barbara, CA. 180–187, 1998.
- Marchionini G, Maurer H. The Roles of Digital Libraries in Teaching and Learning. *Communications of the ACM* 1995;38(4):67–75.
- McGrath RE. Caching for Large Scale Systems: Lessons from the WWW. *D-Lib Magazine* January (<http://www.dlib.org/dlib/january96/ncsa/01mcgrath.html>) 1996.
- Mullen T, Wellman MP. A Simple Computational Market for Network Information Services. *Proceedings of First International Conference on Multiagent Systems (ICMAS-95)*. (<ftp://ftp.eecs.umich.edu/people/wellman/icmas95.ps.Z>) 1995.
- Ogle V, Stonebraker M. 1995. Chabot: Retrieval from a Relational Database of Images. *IEEE Computer* 1995;28(9):40–48.
- Ogle V, Wilensky R. Testbed Development for the Berkeley Digital Library Project. *D-Lib Magazine* July/August. (<http://www.dlib.org/dlib/july96/berkeley/07ogle.html>) 1996.
- Paepcke A, Cousins SB, Garcia-Molina H, Hassan SW, Ketchpel SP, Röscheisen M, Winograd T. Using Distributed Objects for Digital Library Interoperability. *IEEE Computer* 1996;29(5):61–68.
- Paepcke A. Information Needs in Technical Work Settings and their Implications for the Design of Computer Tools. *Computer*

- Supported Cooperative Work: The Journal of Collaborative Computing* 1996;5:63–92.
- Paepcke A, Chang CK, Garcia-Molina H, Winograd T. Interoperability for Digital Libraries Worldwide. *Communications of the ACM* 1998;41(4):33–43.
- Phelps TA, Wilensky R. Multivalent Documents: Inducing Structure and Behavior in Online Digital Documents. *Proceedings of the 29th Hawaii International Conference on System Science* 1996.
- Ram S, Venkatsubramanyan S, Marsh S, Ball G. Resource Discovery and Intelligent Image Retrieval in a Distributed Environment. *Proceedings of the 30th Annual Hawaii International Conference on System Sciences* 7–10, 1997.
- Ram S, Park J. A Comprehensive Framework for Classifying Data- and Schematic-Level Semantic Conflicts in Geographic Databases. Department of Management Information Systems, The University of Arizona, Working Paper, 1997.
- Ram S, Park J. Semantic Conflict Resolution Ontology (SCROL): An Efficient Hierarchical Ontology Scheme for Detecting and Resolving Data- and Schema-Level Semantic Conflicts. Department of Management Information Systems, The University of Arizona, Working Paper, 1998a.
- Ram S, Park J. Semantic Mediators to Facilitate Interoperability Among Distributed, Heterogeneous Spatial Information Sources: An Ontology and Metadata Based Approach. Department of Management Information Systems, The University of Arizona, Working Paper, 1998b.
- Ram S, Ramesh V. Collaborative Conceptual Schema Design: A Process Model and a Prototype Systems. forthcoming, *ACM Transactions on Information Systems* 1998.
- Ram S, Shankaranarayan G. Modeling and Navigating of Large Information Spaces: A Semantic Based Approach. Under Review at *HICCS '98*. 1998.
- Ram S, Venkatsubramanyan S, Yool S, Marsh S, Park J. SAGE: An Intelligent Agent Based Toolkit for Discovering and Accessing Distributed Earth Observing Data. *Proceedings of Earth Observation & Geo-Spatial Web and Internet Workshop '98* 1998;17–19.
- Rao R, Pedersen JO, Hearst MA, Mackinlay JD, Card SK, Masinter L, Halvorsen PK, Robertson GG. Rich Interaction in the Digital Library. *Communications of the ACM* 1995;38(4):29–39.
- Rowley HA, Baluja S, Kanade T. Human Face Detection in Visual Scenes. *Computer Science Technical Report*. CMU-CS-95-158. (<http://informedia.cs.cmu.edu/pubs/faces.pdf>) 1995.
- Rusbridge C. Towards the Hybrid Library. *D-Lib Magazine* July/August. (<http://www.dlib.org/dlib/july98/rusbridgel07rusbridge.html>) 1998.
- Samuelson P. Copyright and Digital Libraries. *Communications of the ACM* 1995;38(4):15–21, 110.
- Sato S, Kanade T. NAME-IT: Association of Face and Name in Video. *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition* San Juan, Puerto Rico, 1997.
- Schatz B. Building the Interspace: The Illinois Digital Library Project. *Communications of the ACM* 1995;38(4):62–63.
- Schatz B, Chen H. Building Large-Scale Digital Libraries, *IEEE Computer* 1996;29(5):22–26.
- Schatz B, Mischo WH, Cole TW, Hardin JB, Bishop AP, Chen H. Federating Diverse Collections of Scientific Literature. *IEEE Computer* 1996;29(5):28–36.
- Sheth AP. Semantic Issues in Multidatabase Systems. *SIGMOD Record* 1991;40(4):5–9.
- Sirbu M, Tygar JD. NetBill—An Internet Commerce System Optimized for Network-Delivered Services. *IEEE Personal Communications* 1995;2(4):34–39.
- Smith MA, Kanade T. Video Skimming and Characterization through the Combination of Image and Language Understanding. *Proceedings of IEEE International Workshop on Content-Based Access of Image and Video Databases (ICCV98)*. (<http://informedia.cs.cmu.edu/documents/ICCV98.pdf>) 1998.
- Smith TR, Frew. Alexandria Digital Library. *Communications of the ACM* 1995;38(4):61–62.
- Smith TR, A Digital Library for Geographically referenced Materials. *IEEE Computer* 1996a;29(5):54–60.
- Smith TR, The Meta-Information Environment of Digital Libraries. *D-Lib Magazine* July/August. (<http://www.dlib.org/dlib/july96/new/07smith.html>) 1996b.
- The Stanford Digital Libraries Group. The Standard Digital Library Project. *Communications of the ACM* 1995;38(4):59–60.
- Ventrone V, Heiler S. Semantic Heterogeneity as a Result of Domain Evolution. *SIGMOD Record* 1990;20(4):16–20.
- Wactlar HD, Kanade T, Smith MA, Stevens SM. Intelligent Access to Digital Video: Informedia Project. *IEEE Computer* 1996;29(5):46–52.
- Wallace R, Krajcik J, Soloway E. Digital Libraries in the Science Classroom: An Opportunity for Inquiry. *D-Lib Magazine* September. (<http://www.dlib.org/dlib/september96/umdl/09wallace.html>) 1996.
- Wang MQ. An Interactive, Structure-Mediated Approach to Exploring Information in a Heterogeneous, Distributed Environment. Computer Science Department, Stanford University, Ph.D. Dissertation. (<http://www-pcd.stanford.edu/michelle/diss.pdf>) 1998.
- Wiederhold G. Mediators in the Architecture of Future Information Systems. *IEEE Computer* 1992;25(3):38–49.
- Wiederhold G. Digital Libraries, Value, and Productivity. *Communications of the ACM* 1995;38(4):85–96.
- Wilensky R. UC Berkeley Digital Library Project. *Communications of the ACM* 1995;38(4):60.
- Wilensky R. Toward Work-Centered Digital Information Services. *IEEE Computer* 1996;29(5):37–45.
- Williams M. Digital Library Technology (DLT). *Research on Digital Libraries*. (<http://dlt.gsfc.nasa.gov/Projects/Abstracts/projectRDL.shtml>) 1997.
- Witten IH, Cunningham SJ, Apperley MD. The New Zealand Digital Library Project. *D-Lib Magazine* November. (<http://www.dlib.org/dlib/november96/newzealand/11witten.html>) 1996.

Sudha Ram is Professor, Management Information Systems in the College of Business and Public Administration at the University of Arizona. She received a B.S. degree in mathematics, physics and chemistry from the University of Madras in 1979, PGDM from the Indian Institute of Management, Calcutta in 1981 and a Ph.D. from the University of Illinois at Urbana Champaign in 1985. Dr. Ram has published articles in such journals as *Communications*

of the ACM, *IEEE Expert*, *IEEE Transactions on Knowledge and Data Engineering*, *Information Systems*, *Information Science*, and *Management Science*. She has also presented her research at several conferences such as International Conference on Information Systems, International Conference on Data Engineering and other IEEE and ACM conferences. She was the guest-editor for a special issue of *IEEE Computer* on "Heterogeneous Distributed Database Systems". She was the General chair for the 1997 ACM SIGMOD International Conference on Management of Data. She was also the Program Co-Chair for the 17th International Conference on Conceptual Modeling (ER98). Dr. Ram's research deals with modeling and analysis of database and knowledge based systems for manufacturing, scientific and business applications. Her research has been funded by IBM, NCR, US ARMY, NIST, NSF, NASA, and ORD(CIA). Specifically, the research deals with Interoperability among Heterogeneous Database Systems, Semantic Modeling, Data Allocation, Schema and View Integration, Intelligent Agents for Data Management, and Tools for database design. Dr. Ram serves on editorial board for the *Journal of Database Management*, *Information Systems Frontiers*, *Journal of Information Technology and Management*, and as associate editor for the *Journal of Systems and Software*, *INFORMS Journal on Computing* and editor for the *IEEE Computer Society Press—Advances in Computer Science and Engineering*. She has chaired several workshops and conferences supported by ACM and IEEE. She is a member of ACM, IEEE Computer

Society and INFORMS. She is also the leader of the Advanced Database Research Group based at the University of Arizona.

Jinsoo Park is a doctoral candidate, Management Information Systems at the University of Arizona. He received a B.A. degree in American Studies from Keimyung University, Korea in 1991 and MBA and M.S. degrees in MIS from the University of Pittsburgh in 1994. His research interests are in the areas of interoperability among heterogeneous databases, metadata management in interorganizational information systems, knowledge management, semantic modeling, geographic database systems, agent-based electronic business transactions. His current research focus is on developing a methodology and technology for facilitating interoperability among distributed and heterogeneous geographic databases. He is a member of ACM, IEEE, AIS, INFORMS and USENIX. He is also a member of the Advanced Database Research Group (ADRG) at the University of Arizona.

Dongwon Lee is a Ph.D. student of Management Information Systems (MIS) at the University of Arizona. His research interests are in the fields of semantic modeling, object-oriented and distributed database systems, agent-based e-commerce, and agent-based digital library. He received a BBA degree in 1996 and an M.S./MBA degree in 1998 from Seoul National University, Korea. He is a member of ACM and AIS. He is also a member of the Advanced Database Research Group (ADRG) at the University of Arizona.