

UCLA

Publications

Title

Social Aspects of Digital Libraries. Final Report to the National Science Foundation

Permalink

<https://escholarship.org/uc/item/7tw0x377>

Authors

Borgman, Christine L.
Bates, Marcia J.
Cloonan, Michele V.
[et al.](#)

Publication Date

1996-09-28

Copyright Information

This work is made available under the terms of a Creative Commons Attribution-NonCommercial-NoDerivatives License, available at <https://creativecommons.org/licenses/by-nc-nd/4.0/>

Peer reviewed

Social Aspects of Digital Libraries

**A Report on the February 15-17, 1996
UCLA-NSF Social Aspects of Digital Libraries Workshop
<http://www.gslis.ucla.edu/DL/>**

**Third in the Digital Libraries Initiative
Workshop Publication Series**

2
699
.A1
2545
1998
v. 3

Table of Contents

Letter from NSF Program Official	p. i
NSF/DARPA/NASA Digital Libraries Initiative Projects - Project titles and URLs	p. ii
Final Report to the National Science Foundation	
Acknowledgements	p. 1
Abstract	p. 2
I. Introduction	p. 3
II. Research Framework for Social Aspects of Digital Libraries	p. 4
II.A. Information Life Cycle Model	p. 6
II.B. Scenarios	p. 6
II.B.1. Human-Centered Scenario: The Information Life Cycle in the Arts	p. 8
II.B.2. Artifacts Scenario: Business Records in the Information Life Cycle	p. 8
II.B.3. Systems Scenario: Health Information Systems and the Information Life Cycle	p. 8
III. Research Agenda	p. 8
III. A. Human-Centered Research Issues in Digital Libraries	p. 10
III.A.1. State of the Art	p. 10
III.A.2. Research Issues	p. 10
III.B. Artifact-Centered Research Issues in Digital Libraries	p. 10
III.B.1. State of the Art	p. 12
III.B.2. Research Issues	p. 12
III.C. Systems-centered Research Issues in Digital Libraries	p. 13
III.C.1. State of the Art	p. 14
III.C.2. Research Issues	p. 14
III.D. Methods To Evaluate The Social Aspects Of Digital Libraries	p. 15
III.D.1. State of the Art	p. 16
III.D.2. Research Issues	p. 16
IV. Conclusions and Recommendations	p. 17
V. Appendices	p. 17
Appendix 1: Agenda	p. 19
Appendix 2: Workshop Investigators, Staff, and Participants	p. 21
Appendix 3: Guidelines for Research Agenda Papers and Presentations	p. 23
Appendix 4: Background Paper	p. 25
Appendix 5: General Conference Information	p. 31

**National Synchronization for the Digital Libraries Initiative
NSF/DARPA/NASA Digital Libraries Initiative Projects
<http://dli.grainger.uiuc.edu/national.htm>**

This is one of a series of reports on workshops sponsored in conjunction with the Digital Libraries Initiative. The Digital Libraries Initiative (DLI) is an interagency program sponsored by the National Science Foundation (NSF), the Department of Defense Advanced Research Projects Agency (DARPA), and the National Aeronautics and Space Administration (NASA).

The broad goal of the Digital Libraries Initiative is to dramatically advance the means to collect, store, organize and use widely distributed knowledge resources containing diverse types of information and content stored in a variety of electronic forms. An essential component in achieving this goal are those activities designed to engage the broader community. Toward this end, a number of workshops have been sponsored to focus on topical areas of importance to the development and use of digital libraries.

More complete information on each of these events and ongoing research and testbed activities of the six Digital Libraries Initiative projects can be obtained through the internet at the addresses given on the next page. A limited number of paper reports have been printed and are being made available to increase the dissemination of workshop findings and for the added convenience of interested parties.

Stephen M. Griffin, NSF
DLI Interagency Coordinating Committee Chair

National Synchronization for the Digital Libraries Initiative NSF/DARPA/NASA Digital Libraries Initiative Projects

University of California, Berkeley and Partners, "The Environmental Electronic Library: A Prototype of a Scalable, Intelligent, Distributed Electronic Library"
(<http://elib.cs.berkeley.edu/>)

University of California, Santa Barbara and Partners, "The Alexandria Project: Towards a Distributed Digital Library with Comprehensive Services for Images and Spatially Referenced Information"
(<http://alexandria.sdc.ucsb.edu/>)

University of Illinois, Urbana-Champaign and Partners, "Building the Interspace: Digital Library Infrastructure for a University Engineering Community"
(<http://dli.granger.uiuc.edu/>)

University of Michigan, Ann Arbor and Partners, "The University of Michigan Digital Libraries Research Proposal"
(<http://www.sils.umich.edu/UMDL/HomePage.html>)

Stanford University and partners, "The Stanford Integrated Digital Library Project"
(<http://www.diglib.stanford.edu/diglib/pub/>)

Carnegie Mellon University and partners, "Informedia: Integrated Speech, Image and Language Understanding for Creation and Exploration of Digital Video Libraries"
(<http://informedia.cs.cmu.edu/>)

UCLA-NSF Social Aspects of Digital Libraries Workshop

Invitational workshop held at UCLA, February 15-17, 1996

Final Report to the National Science Foundation (NSF)

FINAL REPORT TO THE NATIONAL SCIENCE FOUNDATION

**Computer, Information Science, and Engineering Directorate
Division of Information, Robotics, and Intelligent Systems
Information Technology and Organizations Program**

Award number 9528808

Principal Investigator:

Christine L. Borgman, Department of Library and Information Science

Co-principal investigators:

Marcia J. Bates, Department of Library and Information Science

Michele V. Cloonan, Department of Library and Information Science

Efthimis N. Efthimiadis, Department of Library and Information Science

Anne J. Gilliland-Swetland, Department of Library and Information Science

Yasmin B. Kafai, Department of Education

Gregory H. Leazer, Department of Library and Information Science

Anthony B. Maddox, Department of Education

**Graduate School of Education & Information Studies
University of California, Los Angeles**

November, 1996

ACKNOWLEDGEMENTS

Many people besides the investigators were involved in the development, management, and report writing for this workshop. The report was drafted by the investigator and co-investigators at UCLA, with review and additional contributions from the workshop participants. Leah Lievrouw, who joined the UCLA faculty after the proposal was funded, quickly became a full member of the workshop team and made substantial contributions to the report as well. Our external advisory board also guided the selection of participants and the design of the program: Dan Atkins, Edward Fox, Michael Lesk, David Levy, Clifford Lynch, and Gary Marchionini.

Special thanks for the intellectual oversight of the project at the National Science Foundation are due to Su-Shing Chen, Director of the Information Technology and Organizations Program who guided and funded the proposal; his successor as Program Director, Les Gasser, who served as coordinator for the workshop and provided continuing guidance; Stephen M. Griffin, Program Manager for the Digital Libraries Initiative, who gave us invaluable assistance with the workshop and coordinated our work with that of other digital library projects; and Y.T. Chien, Division Director, whose long-term commitment to extending the scope of information science research in general and digital library research in particular led to the digital library initiative and to many related interdisciplinary projects such as this workshop.

At UCLA, we received the strong support of the Graduate School of Education & Information Studies, and from our dean, Ted Mitchell. This was the first major joint project of the two departments of the newly-formed school, established in 1994. Anthony Maddox served as project manager, ably coordinating the myriad administrative aspects of the workshop, before, during, and after the event. Mary King, Events Manager, and her staff turned a classroom building into a conference center, housed and fed participants, and provided technical support to the workshop with efficiency and grand style, all on a National Science Foundation budget. They established a comfortable and effective working environment that contributed substantially to the success of the workshop.

Our team of graduate students, drawn from both departments, not only enlivened our sessions, but took and transcribed notes throughout all the sessions. We are grateful to them for their intellectual contributions and for the many long hours they contributed to the process: Nadia Caidi, Venkatachallam Maithili, Marlene Martin, John Schacter, Susan Schreiner, and Claude Zachary.

Most of all, we thank the workshop participants, who came from around the country to spend a warm February weekend in Los Angeles, for their many contributions, before, during, and after the workshop -- discussion papers, presentations, working groups, editorial review, and contributions to the final report. Philip Agre, Raya Fidel, Rob Kling, and Susan Leigh Star were especially helpful in contributing detailed comments and suggestions for the final draft. The report summarizes the discussions from the workshop and attempts to frame the issues for a much larger group of prospective researchers, designers, and users.

All materials from the workshop, including background paper, participants' discussion papers, and information about the organizers and participants, are available at <http://www.gslis.ucla.edu/DL/>

ABSTRACT

This workshop brought together scholars, researchers, and practitioners from the emerging community of scholars concerned with social aspects of digital libraries. Our goals were to assess existing knowledge that might inform research and to propose a research agenda that would pose new questions.

We propose a definition of digital libraries that encompasses two complementary ideas, one emphasizing that they extend and enhance existing information storage and retrieval systems, incorporating digital data and metadata in any form; the other emphasizing that design, policy, and practice should reflect the social context in which they exist. We propose an information life cycle model to illustrate the flow of human activities in creating, searching, and using information and the stages through which information artifacts may pass: activity, inactivity, and disposal.

Research issues raised in the workshop were organized into three foci: human-centered, artifact-centered, and systems-centered. We recommend that research be conducted on these themes, that scholars from multiple disciplines be encouraged to develop joint projects, that scholars and practitioners work together, and that digital libraries be developed and evaluated in operational, as well as experimental, work environments. Only in this way can we build digital libraries to support diverse communities of users in their professional, educational, and recreational activities.

I. Introduction

This workshop was a result of a series of informal conversations that took place over the last several years with increasing frequency, between members of multiple disciplinary and professional communities, regarding the need for more research on the social aspects of digital libraries. Many scholars are recognizing that a new intellectual community of interest is forming around these issues. Although we came from very different disciplines, our paths had crossed or paralleled for years. The emergence of this community reflects a joint sensibility that we are experiencing a major social transformation, and that digital libraries are a crucible for this transformation. Some of us knew each other from concerns with ethics and privacy; some came from science and technology studies; some knew of each other through methodological conversations; some knew each other's work through seeking abstract connections in the literature. No individual at the workshop knew all the other participants; rather, the group was selected to represent a diverse but complementary set of interests, drawing from networks of people known to the organizers and the advisory board¹.

The workshop served as a place to strengthen the bonds among the emerging community, identify new members, and identify issues that would draw the interest of a much larger research community. Conversations were lively and rich; we all left with a sense of excitement about this rapidly growing community with so many common interests and deeply intersecting roots.

It is not by accident that a term for this community, "social informatics," originated at the UCLA workshop. In the few months since the workshop that term already is in use at the National Science Foundation, in the title of a new research center at Indiana University, the title of a 1996 chapter in the *Annual Review of Information Science and Technology*, and the title of a forthcoming special issue of the *Journal of the American Society for Information Science*.

The core premise of the workshop was that digital libraries represent a set of significant social problems that require human and technological resources to solve. Workshop participants were charged with appraising the scope of social aspects of digital libraries, assessing what is known about these problems, and identifying the research and development issues that need to be addressed to solve them. Our first task was to define "digital libraries." We determined that digital libraries encompass two complementary ideas:

1. Digital libraries are a set of electronic resources and associated technical capabilities for creating, searching, and using information. In this sense they are an extension and enhancement of information storage and retrieval systems that manipulate digital data in any medium (text, images, sounds; static or dynamic images) and exist in distributed networks. The content of digital libraries includes data, metadata that describe various aspects of the data (e.g., representation, creator, owner, reproduction rights), and metadata that consist of links or relationships to other data or metadata, whether internal or external to the digital library.
2. Digital libraries are constructed -- collected and organized -- by a community of users, and their functional capabilities support the information needs and uses of that community. They are a component of communities in which individuals and groups interact with each other, using data, information, and knowledge resources and systems. In this sense they are an extension, enhancement, and integration of a variety of information institutions as physical places where resources are selected, collected, organized, preserved, and accessed

¹ Among the disciplines represented, whether by current affiliation or graduate degrees, were archives, art history, communication, computer science, education, library and information science, management, philosophy, psychology, and sociology. Participants included university faculty, researchers in private industry and in non-profit institutions, and library and archive practitioners involved in digital library research.

in support of a user community. These information institutions include, among others, libraries, museums, archives, and schools, but digital libraries also extend and serve other community settings, including classrooms, offices, laboratories, homes, and public spaces.

The first idea emphasizes the fact that digital libraries are computer-based systems constructed for people to use and that they are extensions of information storage and retrieval systems. The second emphasizes the belief that digital libraries should be constructed in a way that accommodates the actual tasks and activities that people engage in when they create, seek, and use information resources; in this sense they are an extension of physical environments. Both assert that digital libraries are sets of information resources collected and organized on behalf of a community.

Embedded in this definition are complex concepts with meanings that vary by context and by field of study. The terms "information," "community," and "library" are the most problematic. Definitions of "information" abound: signal processing; sensory perception; data generated by individuals and groups; objects that can be managed in retrieval systems; intellectual commodities that can be exchanged in the marketplace; etc. "Community" implies a group of people with something in common, but those common features may be permanent or temporary, static or dynamic, innate or selected; biological or cultural, etc. -- and any one individual can be a member of many communities at once. A "library" is often narrowly defined in technical contexts as a database application, while in other contexts a "library" is a social institution that selects, collects, organizes, preserves, conserves, and provides access to information on behalf of a community. Even the term "digital" is problematic, for it reflects both "digital objects" -- those created in digital form, and "digitized objects" -- those that are representations (e.g., scanned images, keyed text) of objects in other forms.

We cannot resolve these definitions here, nor is it fruitful to do so. Rather, we recognize that many perspectives exist and that research on digital libraries will benefit by study from the largest possible number of perspectives. We do find it helpful for the purposes of this report to distinguish between information entities as the objects that can be collected and organized into digital libraries and information in the sense of communication processes involved in the creation and use of those information entities. Entities in digital libraries are representations of human communication and are thus artifacts of that communication. Those artifacts can be described and represented in many ways, depending on the social context, motivation for using digital libraries, and other aspects of the application. As we illustrate below, the same artifact might be collected for multiple purposes and organized in multiple ways, depending on the community and application served.

While it is possible to build systems independent of human activities that will satisfy technical specifications, systems that work for people must be based on analyses of learning and other life activities. Empirical research on users should be influencing design in three ways: (1) by discovering which functionalities user communities regard as priorities; (2) by developing basic analytical categories that influence the design of system architecture; and (3) by generating integrated design processes that include empirical research and user community participation throughout the design cycle. Important decisions frequently are made at the very beginning of the design process, often without the designers realizing it, because they are using concepts that do not align accurately with user communities' concepts or with empirical reality. It would be unfortunate if this happened with digital libraries. Furthermore, given that such decisions are being made today, we are at a crucial turning point in the history of the infrastructure of collective human cognition.

In considering a research agenda, we acknowledge that digital libraries will continue to be constructed by the research and development community on behalf of users, but that users also will construct digital libraries on their own behalf. Thus we should create functional capabilities and

tools that enable people to construct and tailor digital libraries to their own circumstances. The phrase "social aspects" in this report refers to the perspective that human considerations -- the individual, group, and community -- should be the starting point for digital library design.

Our purpose in this report is to identify research issues arising from the many different disciplines concerned with the theory and practice of digital library development. This disparate research community needs a framework within which to identify complementary interests and areas of collaboration. Claiming a single set of definitions or perspectives would be contradictory to that goal. Our objectives in this report are to outline existing knowledge that might inform research and to propose a research agenda that builds upon that knowledge to pose new questions about the social aspects of digital libraries.

II. Research Framework for Social Aspects of Digital Libraries

We based the selection of workshop participants and the workshop discussion around two social aspects of digital libraries: information needs and end-user searching and filtering. These aspects, their component topics, and discussion questions are presented in the background papers in the Appendix. Discussion papers by the workshop participants responded to the UCLA background paper and identified many other issues. While the UCLA background paper provided a fruitful starting point for the workshop, we quickly expanded the boundaries of our concerns in several directions. Rather than focusing solely on the individual user who interacts with a digital library, we considered also the group, organization, and community activities and concern which give rise to information-related behavior. We expanded our interest in information storage and retrieval to include preceding and succeeding phases, incorporating the processes of creating, using, and disposing of information.

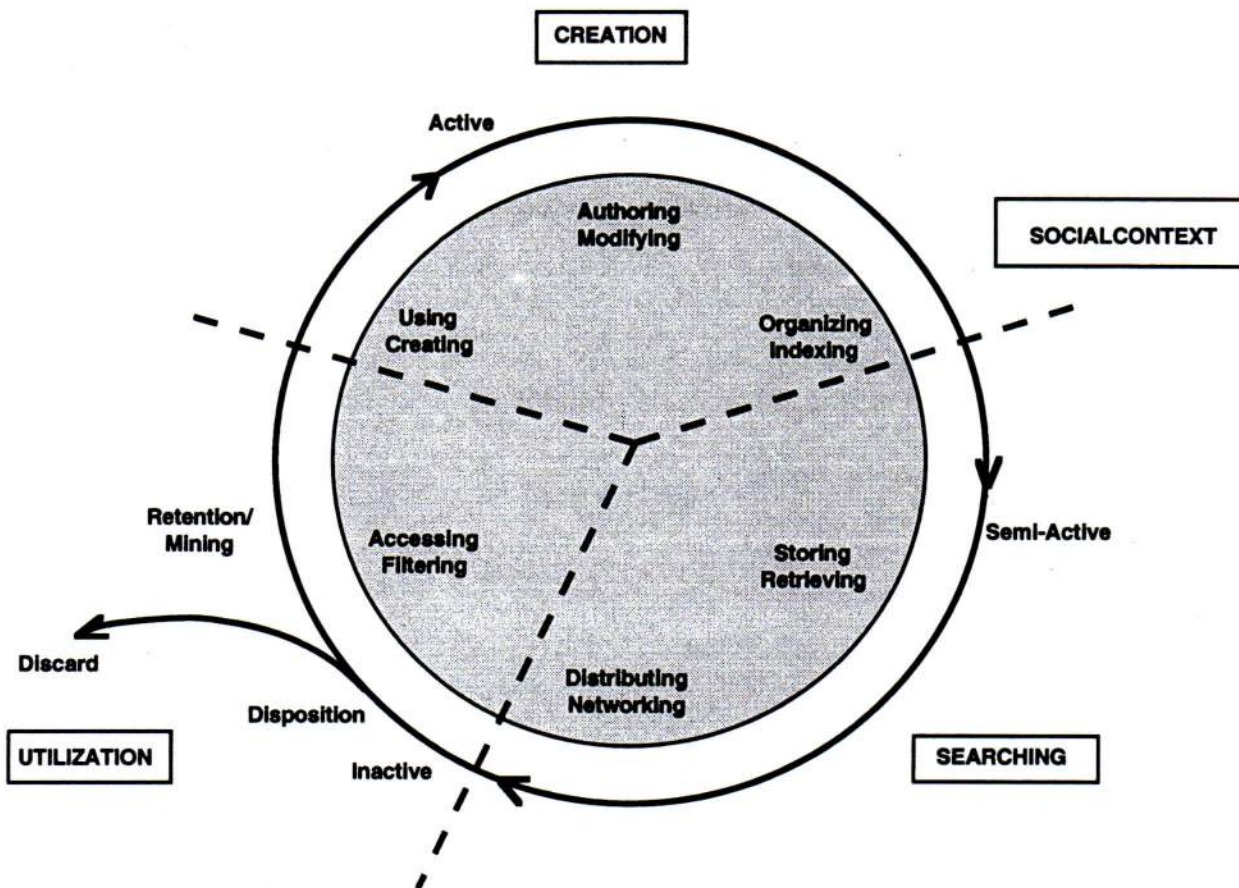
Our discussions resulted in the two-part definition of digital libraries stated above, in several common themes, and in a general model of the life cycle of information and information processes. We present the model, illustrate it with scenarios, and then organize the research issues around these three themes:

- Human-centered research issues: a focus on people, both as individual users and as members of groups and communities, communicators, creators, users, learners, or managers of information. We are concerned with groups and communities as units of analysis as well as with individual users.
- Artifact-centered research issues: a focus on creating, organizing, representing, storing, and retrieving the artifacts of human communication.
- Systems-centered research issues: a focus on digital libraries as systems that enable interaction with these artifacts and that support related communication processes.

II.A. Information Life Cycle Model

The Information Life Cycle depicted here is one schematic attempt to represent the flow of information, both as artifact and as social process, in a given social system (Figure 1). The outer ring indicates the life cycle stages (active, semi-active, and inactive) for a given type of information artifact (such as business records, artworks, documents, or scientific data). The stages are superimposed on six types of information uses or processes (shaded circle). The cycle furthermore has three major phases: information creation, searching, and utilization. The alignment of the cycle stages with the steps of information handling and process phases may vary according to the particular social or institutional context.

Information Life Cycle



NOTE: The outer ring indicates the life cycle stages (active, semi-active, and inactive) for a given type of information artifact (such as business records, artworks, documents, or scientific data). The stages are superimposed on six types of information uses or processes (shaded circle). The cycle has three major phases: information creation, searching, and utilization. The alignment of the cycle stages with the steps of information handling and process phases may vary according to the particular social or institutional context.

Though this figure shows only a single round of the cycle, it is important to note that cycles may intersect, overlap or "stack" as information moves across social settings. Information may be removed from active use at one or more points in the cycle. Disposal does not necessarily imply that information is destroyed; rather, it may be stored for later use by others in different circumstances, set aside, or may otherwise continue to exist. While social context is not explicitly represented in the figure, it is environmental and pervasive throughout the cycle. Creating, seeking, and using information are socially-situated human activities.

Some activities may evolve in the predicted directions; others may iterate between phases, skip phases, or end before the cycle is complete. People's encounters with digital libraries -- or any type of information system -- are reflexive; that is, each encounter influences the next. The user's situation and knowledge change continually and some systems are able to respond to these changing states.

II.B. Scenarios

The UCLA report team also developed several scenarios to illustrate both the model and the three themes. The art world scenario demonstrates the human-centered focus; the business records scenario illustrates the artifact focus; and the health information scenario exemplifies the technology focus.

II.B.1. Human-Centered Scenario: The Information Life Cycle in the Arts

Artists, curators, dealers, students, lay people, and audiences create, search, or use art content or processes in a virtual community that is sometimes called the "art world." In the creation phase of the cycle, artists' production of new works often depends on their ability to use or "mine" information in innovative ways. They may draw on others' ideas or works as influences, to contradict or react against, or to incorporate elements into new works. "Authoring" in this sense is a creative response, as the artist incorporates themes, ideas, or images from diverse sources into his or her own insights and representations.

Other arts professionals, such as art historians, musicologists, music librarians, literary critics, or other gatekeepers sort, organize and evaluate cultural works.

The convergence and conflicts among these groups' views of the same information is seen in the searching and utilization phases of the cycle. Are musical pieces organized and searchable by date, style, composer, melodic theme, performance, performers, length, genre, storage medium, or all of these? Are visual art works retrievable by their formal characteristics, mythological references, concepts, "schools," places of origin, figures depicted in them, artist's name, owner's name, provenance, medium, or all of these? As each community organizes and represents the content for its own use, unfamiliar language, representations, and functional capabilities may present barriers to use by other communities.

Distribution and access to cultural works involves yet other organizations and people -- galleries, magazines and journals, museums, and libraries all play a role. In the performing arts, producers, critics, theater companies, and publishers of plays perform the same function. Judgment is key at this point in the cycle; the art dealer decides which artists to represent and show, the museum curator decides which works to acquire and exhibit, and the theater company director selects which plays to produce. Some art works will necessarily be discarded (physically destroyed or not recorded in a useful form).

Finally, works available in a given place and time provide the basis for artists to make new works in a renewed cycle of creative borrowing, influence, use, and originality.

II.B.2. Artifacts Scenario: Business Records in the Information Life Cycle

Businesses continuously produce, search, use, and discard records, and develop record-keeping and information systems to do so. Business records include operational data (e.g., asset management, market profiles, scheduling projections), related transactional metadata (e.g., audit trails, use statistics), and strategic information (e.g., annual reports, product designs, patents, executive correspondence). Information itself, in the form of digital materials such as graphic design or software, may be the business's product. In most cases, business information systems and the records they contain are considered either as assets or as by-products of business operations. Organizations increasingly view the information artifacts they generate as their "institutional memory," and are seeking ways to capture and exploit "intellectual capital" (e.g., as profiles of employee expertise) for new purposes.

Traditionally, business records (artifacts) move through the information life cycle from a period of intense use shortly after they are created, through a period of occasional use, to a period of inactivity. Records that are no longer used are discarded according to a systematic records retention schedule, or transferred to an archive for preservation. Preservation decisions are based on whether materials have enduring legal, fiscal, or administrative value for their creators or subsequent historical or research value to other users.

The life cycle of digital business records is now often seen as asset management; fewer corporate records (especially operational data and transactional metadata) are being retired systematically. Digital artifacts are stored for unforeseen uses (i.e., data warehousing), are used by different workers for new reasons (e.g., training, work practice analyses), are analyzed and cross-compiled to serve new management objectives (e.g., data mining), and are combined into new products. Artifacts must be reorganized, re-indexed, and searchable in new ways to be useful for new purposes.

II.B.3. Systems Scenario: Health Information Systems and the Information Life Cycle

In the context of digital libraries, the creation and use of health-related information requires a wide array of technological capabilities so that health care providers, researchers, policy makers, the general public, and others can use the information according to their needs. Many sources generate health information, including patient care units, clinical laboratories, insurance companies, government, health clubs, research and educational institutions, and individuals themselves. Data are stored in financial, telemedicine, and public and private health information systems, and are used for patient care, financial management, legal compliance, clinical and public health research and teaching, and so on.

While the artifacts needed for all of these applications may be the same or similar, the communities and purposes for use are different. At present, the applications are served by multiple digital libraries and multiple systems, each with different methods of organization, representations of artifacts, and functional capabilities. They might be served better by a single digital library if it could support multiple representations, methods of organization, and multiple functional capabilities tailored to different audiences. Alternatively, they could be served by multiple digital libraries with links among the representations, enabling them to function as a single system. From a systems perspective, digital libraries for health care applications should be interoperable, support platform portability, verification and authorization of data from many sources, and reduce redundancy. Records may be active, inactive, or eligible for disposal according to different applications. At the same time, the network of systems should provide interfaces tailored to each group of users that would allow them to create, search, and use information in their own ways. The design of such digital libraries must be based on an understanding of work practices and other information related behavior in the health care context.

III. Research Agenda

We organize the research agenda around the three themes introduced earlier: human-centered, artifact-centered, and systems-centered aspects of digital library research. Within each, we present a brief summary of the state of the art and a list of issues. No rank order is implied, nor should be inferred. While we make no claim that the research issues identified are either mutually-exclusive or exhaustive, this list represents issues that workshop participants identified as urgent and solvable, since sufficient knowledge exists to frame them and to establish their significance. We conclude with a section on methods to evaluate the social aspects of digital libraries.

III. A. Human-Centered Research Issues in Digital Libraries

III.A.1 State of the Art

Research on individuals usually falls in different disciplines than does research on groups, communities, and social context and culture. Individual users of information technology are studied in communication, library and information science, education, psychology, human factors, and linguistics, among others. Most of the research in these disciplines views the individual as an actor who employs the technology for instrumental purposes. We understand basic characteristics of individual information use within groups such as professionals (engineers, art scholars, social workers, etc.), the general public, members of age groups (children, seniors, etc.), and members of other special groups (disabled, prisoners, etc.). Adult users are far better studied than are children, and goal-directed information seeking is far better studied than browsing and serendipitous behavior. Characteristics of information usage vary widely among these groups, raising questions of when systems can be generalized and when they should be tailored to specific groups, or even to individuals. While we have a basic understanding of human communication processes, both oral and written, we have only rudimentary knowledge of how these processes change when conducted via new media.

The social context and culture of information technologies, including digital libraries, has been the subject of a substantial body of social research. Much of this research has been conducted by scholars who anchor their analyses in social studies of science and technology, institutional analysis/political science, symbolic interactionism, ethnomethodology, organizational and group communication research, cultural and linguistic anthropology, political economy, and activity theory, among others. They all share similar social approaches to technology; i.e., they focus on technologies as they are situated in and arise from social relationships, communities, power, and the creation and sharing of meaning. These traditions tend to examine visible behavior rather than cognition, and relationships rather than individuals; and reject simple, technologically-deterministic frameworks in favor of more socialconstructivist views of technological development and diffusion in society. They recognize that the acceptance and use of information technologies reflects ongoing negotiations among social groups with divergent economic, political and cultural interests.

Among the better understood topics at this level are the relationship between work practices and the design of systems and user interfaces; evolution, implementation, and evaluation of information technologies, especially in organizations; and user perceptions of and participation in development. A substantial body of work extending over several decades has demonstrated enduring inequities in the distribution of and access to information and related technologies across social groups.

III.A.2. Research Issues

We identified the following topics as significant human-centered research issues in digital libraries. We do not claim that this is a complete list; rather, it reflects the themes most commonly identified by the workshop participants. No rank order is implied.

Heterogeneous populations and applications: When should digital libraries be tailored to individual users, groups, and communities? When should they be generalized? What social, demographic, or other variables should be considered in digital library design? How do we accommodate the varying understanding of the same content by different communities? For example, current legal information systems are predicated on a thorough understanding of the law, yet non-lawyers have great needs for legal materials as well. Similarly, how do we make the same scientific materials useful for scientists and school children? Whereas professionals know the domain, are motivated, and are a homogeneous population with the goal to increase the organization's success, students do not know the domain, often are not motivated, and encompass very diverse populations. How do we incorporate this disparate range of behaviors into digital library design?

Institutions/cultural objects of study: Can cross-institutional frameworks be developed for describing digital library development and impact? What are the cultural responses to technology (e.g., social differentiation versus integration)? Can integrated systems be built that reflect a complete sense of community, incorporating publishing, support for conversation, and computer-supported cooperative work, as well as information retrieval?

Information literacy skills: What kinds of information literacy skills are required for digital libraries? What do we need to teach and how do we teach it? To what extent can digital libraries be self-instructional? What old behaviors and expectations about information and information systems will users carry into digital libraries?

Designing for richness: How can digital libraries both embody and support new ways of doing things; e.g., changing literacies? What is the relationship between digital libraries and emerging practices like knowledge brokering? Will they support or threaten national traditions (e.g., languages and cultural practices)? How will digital libraries be built and situated in information environments characterized by browsing, varying levels of social intelligence, changing demands for information, and subjective experience? How may digital libraries complement or disrupt the rhythms, routines, and interruptions of work life?

Studies of situated use: How do people actually use or otherwise engage with information now — e.g., what comprises reading in a multimedia environment? What can be learned by studying new or novice users, on one hand, versus those who resist or abandon new technologies, on the other? What can be learned from historical studies of the development and politics of technological standardization?

Design world/Content world interface: What is the social role or social life of different types of content? Does that role change from system to system, across social groups, or across geographic areas? How can design priorities better support the meanings and relationships of people who create and share content? How can we employ what people know about their subject domain and work practices in the design of interfaces and functional capabilities?

Tools for content creators: Digital libraries will enable everyone, including children, to be authors, producers, and creators of information—whether as simple as a home page or as sophisticated as a novel or the resources to support an electronic community. What kinds of help do people need, and what kinds of information do they need to achieve their objectives as producers of information?

III.B. Artifact-Centered Research Issues in Digital Libraries

III.B.1. State of the Art

Digital libraries contain information entities collected and organized on behalf of communities. These entities are artifacts of human communication or are digital representations of artifacts. Artifacts may be text, images, numeric data, sounds, or other information created in digital form; they may be representations of other online or offline artifacts. Information entities are data and usually carry associated metadata that is necessary to identify, manage, and use the data. Metadata may be descriptions of content (author/creator, title, subject, summaries, classification codes, etc.), descriptions of an artifact (format, software that created it, granularity of image, etc.), ownership, reproduction rights, security (cryptographic technique, etc.), relational metadata that provide links to other versions, source codes, viewers, related materials, etc. Some artifacts will be static objects (e.g., published documents), others will be dynamic (e.g., intermediate versions of documents), or continuous (e.g., conversations, transaction data streams). And some artifacts will consist of metadata describing non-digital objects (e.g., catalog records for printed books; descriptions of people, museum objects, geological sites, public buildings, etc.). The line between data and metadata is a fuzzy one in digital libraries.

The study of artifacts in digital libraries builds on the knowledge of artifact creation discussed in the prior section and incorporates research and practice in the description, organization, and representation of information objects. Theoretical constructions of how people naturally describe and organize objects are studied in philosophy, psychology, education, and linguistics, among other fields, and extended into theoretical models and practice in archival studies and library and information science (description, cataloging, classification, indexing, abstracting) and computer science (knowledge representation).

Most of the research and development on organization of resources within collections has taken place in separate professional contexts such as librarianship, archives, museum curation, and expert systems. Significant cross-professional cooperation between these communities is a relatively recent phenomenon, although each community established professional practices for the organization of digital resources as they were introduced. The library community established international standards for the communication of digital resources in the 1960s, resulting in the hundreds of millions of cataloging records (metadata) now extant in digital form. Research efforts in information organization and retrieval in these applied settings continue to result in improvements in the design of specific information systems. Research and development in other communities has resulted in standards such as SGML (Standard Generalized Markup Language) and HTML (HyperText Markup Language). A variety of public domain and proprietary representation structures for images, text, and other objects are appearing, such as TIFF, JPEG, MPEG, TEI, etc. While many of these formats are incompatible, some progress is being made in exchange mechanisms.

Digital library design will likely draw from a number of organizational and representational techniques; no one approach fulfills all kinds of information needs. A number of models exist for the organization of materials in a single collection, but no similar model exists for organizing resources across multiple collections. Rapid changes in the industries and institutions that produce and manage artifacts, such as publishing, film studios, software developers, and telecommunications law, are shaping the ways that new kinds of materials serving new purposes are generated and distributed.

The description and organization of artifacts relies heavily on human judgement, applying knowledge of the subject domain, of the intended user communities, and of principles of indexing, abstracting, classification, and categorization. While formal characteristics such as size, color, and format can be assigned automatically, description of content usually requires assigning

characteristics of meaning to the artifact, a distinctly human task. Searching by text contained in artifacts is notoriously difficult, due to the variation in uses of a given term in different contexts (Paris, the city; Paris, the god; plaster of Paris), variation in terms for a given concept by different communities (e.g., botanists vs. gardeners; scientists vs. schoolchildren; physicians or lawyers vs. lay persons) and in different contexts; and the variety of terms by which any concept is labeled. Promising avenues of exploration include "vocabulary switching" databases to translate among the terminology of communities, and computational techniques to identify latent concepts. Computational linguistics, including automatic language translation, will be important to creating, searching, and utilizing artifacts in digital libraries. We need to extend these techniques to content other than text, and find new ways to describe and organize images and sounds.

III.B.2. Research Issues

We identified the following topics as significant artifact-centered research issues in digital libraries. We do not claim that this is a complete list; rather, it reflects the themes most commonly identified by the workshop participants. No rank order is implied.

Making artifacts useful within a community: Studies of information-seeking behavior and of work practices yield insights into organizing for a given community. How can we generalize these assessment methods to determine optimal organizational methods for a given community? The attempt to tailor organizational representations of digital libraries for specific communities reaches its logical conclusion when digital libraries are organized for a single individual user, or a single particular use. How can we make it possible for users to personalize existing organizational schemes, or to create their own?

Making artifacts useful to multiple communities: Information organization strategies facilitate sharing across multiple communities of users. For example, how can legal or medical materials be useful both to experts and to the average citizen? What do we need to do to make digital libraries useful for other communities? How can collections of historical records or of scientific images be arranged in order to promote use by scholars? Can these same collections be organized for use by school children?

Dynamic artifacts: How do we organize and represent rapidly changing material or multiple manifestations of substantially similar materials? What sorts of schemes must be developed to keep surrogates and other descriptions of rapidly changing digital materials up-to-date; to represent and describe multiple manifestations of the same work?

Hybrid digital libraries: Digital artifacts will supplement, not supplant hard-copy artifacts. Non-digital materials (paper, film, microfiche, etc.) must be integrated with digital materials for combined access. How can we agglomerate and reconcile earlier non-digital control technologies, such as library catalogs, museum registrarial systems, and archival finding aids into digital libraries?

Professional practices and principles: What are the appropriate contributions of cataloging, indexing, archives, museum informatics, and information system design to the organization of resources in a digital library? Can specific organizing techniques developed for non-digital materials be applied in the new digital environment? What about the applicability of principles developed for an earlier time? Have others with a useful professional contribution to make been excluded in digital library design? What principles from these areas are relevant to digital libraries? Are all general principles relevant? How do relevant principles apply to digital libraries and what form do they take? What modifications in the practice of applying these principles are required?

Human vs. automated indexing: Digital libraries will be far too large to rely entirely on manual description and organization, thus more research effort is needed in automated description and

organization. While digital artifacts will be easier to describe automatically than non-digital artifacts, description of meaning will continue to be a problem. Most importantly, we need to achieve a workable balance between automation and human intervention. Only the most superficial indexing of works can be done automatically, and human indexing of content is expensive. What is indexed best by humans and what by machines? How do the two complement one another?

Legacy data: Massive amounts of data and metadata about artifacts already exist in digital form, some to current standards and much in non-standard formats. What are the principles and the selection criteria for migrating these data and metadata to new forms for digital libraries?

Hierarchies of description: We need description and organization not only within digital libraries, but among them. Searchers must be able to identify the existence of a digital library before being able to locate an artifact it contains. We need to identify relationships among digital libraries. The arrangement and organization of entire collections—the interoperability of a digital library's organizational component—might be achieved through the use of standards, but these standards and the systems that exploit them need to be developed. How can we develop compatible representations at the level of individual digital libraries and at the level of collections of libraries?

Portability: The range of content, formats, and users of digital libraries will result in a comparable range of standards and mechanisms for description and organization, yet each community may wish to interact with artifacts originating in another. How can we move data and metadata between different representations and encoding schemes?

Artifactual relationships: Can we develop schemes to represent the relationships among digital materials? One way to deal with highly similar manifestations of the same resource and rapidly changing digital material may be to develop automated means to represent relationships among digital items such as whole/part, same origin of content in different medium (e.g., book, script, film, play), multiple instances of an artifact, original and translation, etc.

Level of representation: Preferences for level of description vary by collection and by community. For example, how fine should the resolution be in a collection of stored images of American cities or farmland? That may depend on what kinds of data that scientists—or teachers and their students—will subsequently want to extract from the images. Shall a literary manuscript be stored as natural-language-searchable text or as a digital image? Some scholars may want to search for key words or phrases, and prefer the former, while others may want to see every mark on the digital image of the original manuscript page. How shall we determine the level of representation for a collection or a community?

III.C. Systems-Centered Research Issues in Digital Libraries

III.C.1. State of the Art

From a systems-centered perspective on the social aspects of digital libraries, our goal is to construct digital libraries as systems that enable interaction with these artifacts and that support related communication processes. The systems-centered perspective integrates the human and artifact perspectives. While a wide range of technologies and functional capabilities are required for the design and development of digital libraries, most are beyond the scope of this report. We restrict our discussion to systems-centered research issues that follow directly from the human-centered and artifact-centered issues presented above.

Individuals, groups, and communities require a variety of technologies in their interaction with digital libraries, whether as communicators, creators, users, or managers of information. Technologies are needed to support the creation, description, organization, representation, and

utilization of the artifacts of human communication. The choice of capabilities and degree of use will vary throughout the information life cycle.

The social aspects of digital libraries meet technology at the user interface because the interface reflects deeply-embedded design decisions and implicit assumptions about peoples' goals, communication, cognition, and behavior related to the system. All too often, interface design focuses on the surface characteristics of the system, attempting to "patch" inelegant or cumbersome systems.

Computer-based technologies exist in support of all steps in the information life cycle, but usually were developed for specific purposes at that step and are not capable of transferring content among steps. Although technologies exist to cross platforms with ease for those with good technical infrastructure, the real world of digital libraries must cope with the realities of severe budget limits and hereditary systems. Especially as digital libraries cross borders into schools, commerce and the home, the pragmatics of maintenance and support for the following issues need to be understood and taken into account.

For example, we have technologies for creating and authoring text, images, and music, but few technologies for organizing, indexing, storing, or retrieving the products of those technologies directly. Word processing files usually require manual markup for typesetting; word processing and typesetting files rarely enter digital libraries without further manual markup for indexing and retrieval. The manual intervention often is so cumbersome that it is easier to recreate the data (e.g., through scanning or keying) than to reuse it. Despite the great strides in word processing technology in the last decade, it remains difficult for authors using different software and computing platforms to share files, especially if they need to exchange them intact over the Internet. Exchanging digital data in other media (images, sounds) remains yet more problematic, despite progress in technical standards.

We have more advanced tools for creating digital objects, especially for text, and progress is being made in tools to create still and moving images. Research on computer-supported cooperative work is increasing our understanding of group processes related to information technologies.

Research in retrieval of text is the most advanced area of digital libraries technology, with a history dating from the 1950s. To the extent that any information entities can be managed with textual metadata, text retrieval techniques are generalizable. Searching for objects by non-textual characteristics is most easily done by formal features such as shapes or colors, but even these techniques are in early stages of development. Little work has been done in tools to support other steps in the information life cycle, such as tools for communication (e.g., how to share data), tools for interpretations (e.g., how to process data), tools for creation (e.g., how to contribute to information), tools for documentation (e.g., search history), and tools for protection (e.g., privacy). These tools need to be adaptable in two ways: how the system adapts to the user and how users customize the system to their needs.

III.C.2. Research Issues

We identified the following topics as significant systems-centered research issues in digital libraries. We do not claim that this is a complete list; rather, it reflects the themes most commonly identified by the workshop participants. No rank order is implied.

Community-based development tools: Digital libraries need to be tailored to the context of their target audience, providing effective search methods suitable for diverse communities, varying from the untrained user to specialists, from occasional to expert users, from the general population to narrowly defined groups. Individual communities may be multi-cultural and multi-lingual, and digital libraries supporting different cultural and linguistic groups need to be able to interact with

each other. How can we promote customized development of large numbers of digital libraries that are interpretable and can be tailored to individuals and communities?

Multiple interfaces: Each digital library may have multiple user communities. Is it more appropriate or effective to develop multiple interfaces representing different learning stages or categories of information needs, or to develop a single generic interface coupled with diverse navigation and data manipulation tools?

Social interfaces: How can "social interfaces" facilitate the creation, retrieval, and filtering of information, while facilitating the communication essential to building online communities? How can the interface facilitate, but not impose, community views and values?

Mediating interaction: How can interfaces be both generic and infinitely flexible, taking into account how people do things in the world, and what they want to do? How can interfaces provide tools for mediated creation and retrieval, but not themselves mediate?

Intelligent agents, user models: what kinds of access are desired by users? What role can and should human intermediaries have? Computational agents? Can we identify patterns in information seeking styles that might translate into user models for digital library design? What design features and search capabilities in existing related systems best meet user needs and capabilities? What kinds of filtering can be taught users, and what kinds of automatic filters can be designed to do for users what they would do for themselves?

Information presentation: The manner in which information is presented or delivered will influence the way that it is received and interpreted. How can tools for presentation design support the creation, searching, and utilization stages of the information life cycle?

Open architecture: The balance of generalizing and tailoring digital libraries to communities will require that multiple digital libraries be interoperable. How can we create the open architectures necessary for data exchange, portability, and interoperability?

Development methods: Incorporating human-centered approaches to digital library design requires an iterative cycle of design-test-redesign. How should current methods be adapted to support general purpose digital libraries and digital libraries tailored to well-defined user communities?

Tools for accessing and filtering information: At the core of the information retrieval problem is the need to locate the relevant information while filtering out the abundant irrelevant information. How can digital libraries incorporate native abilities in accessing, filtering, navigating, browsing, and searching for information?

III.D. Methods To Evaluate The Social Aspects Of Digital Libraries

III.D.1. State of the Art

Designing real systems for real people requires that we have a means to evaluate them, not just against a set of technical specifications but within the social context of their use. While reliable and valid methods exist, they have not been widely applied in digital library design, and new methods are needed as we extend the scope of digital libraries and their communities of users.

Studies of the individual and of the social contexts and culture of information technologies have employed a wide range of data-gathering and analysis techniques, including controlled experiments with operational or prototype systems, unobtrusive online collection of behavioral data (e.g., logging keystrokes), ethnographic techniques like participant observation or interviewing, content analysis, and network analysis. Some types of data, such as network or logging data, may be

subjected to quantitative, multivariate analysis; qualitative data may be analyzed thematically or using techniques from criticism such as literary or genre analysis, dramatic or rhetorical analysis. Research in human-computer interaction indicates that even briefest evaluation efforts significantly increase the quality of design.

III.D.2. Methods Issues

We identified the following topics as significant methods issues in digital libraries. We do not claim that this is a complete list; rather, it reflects the themes most commonly identified by the workshop participants. No rank order is implied.

Participatory design: How can we involve digital library users in the design and evaluation processes?

Studying new activities: What new techniques are needed to study virtual institutionalization? How can new types of discursive practices (e.g., chat rooms, online help or advice networks) be observed and analyzed both validly and reliably? What can be learned methodologically from the study of existing systems? Can system designers be encouraged to employ social analysis methods in the design process? How can studies of users and practices be designed to be more longitudinal, to take advantage of multi-disciplinary research teams, to cross-train methodological specialists, or to triangulate among multiple methodologies?

Levels of evaluation: We need to evaluate components of digital libraries as well as relate multiple perspectives on how the social context influences the design of artifacts. What kind of comprehensive measures do we need to design that evaluate the whole information and learning experience? What kind of evaluation processes (and supporting tools) will provide timely and valid predictions about individual steps, features, and capabilities?

Iterative methods: How can we extend methods of iterative design to include evaluation during and after system use through which we gather information while people are using the system? How can we study groups engaged in rapid development and formative and summative evaluation of digital libraries?

Tailoring methods: We need methods and measures to evaluate digital library designs in relation to potential users and contexts. For example, what works well in professional and academic settings may not be appropriate for the average user.

IV. Conclusions and Recommendations

We brought together scholars, researchers, and practitioners from the many disciplines that study the ways people create and use information, and those who study methods and techniques for creating, representing, and organizing information. Our discussions addressed a wide range of social aspects of digital libraries, considering information creation and use among individuals, groups, organizations, and society, and the technology required to support them. Our goals were to assess existing knowledge that might inform research and to identify a research agenda that would pose new questions.

As a result of our discussions, we propose a definition of digital libraries that encompasses two complementary ideas, one emphasizing that they extend and enhance existing information storage and retrieval systems, incorporating digital data and metadata in any form; the other emphasizing that design, policy, and practice should reflect the social context in which they exist. The first idea emphasizes the systems perspective, that digital libraries extend and enhance existing information storage and retrieval systems, incorporating digital data and metadata in any form. The second

emphasizes that digital libraries exist in a social context and that design, policy, and practice must reflect that context.

We propose an information life cycle model to illustrate the flow of human activities in creating, searching, and using information and the stages through which information artifacts may pass: activity, inactivity, and disposal.

The two-part definition of digital libraries and the information life cycle model reflects the complementary perspectives of many disciplines and professions with an interest in information creation, use, and management and the convergence of information and communication technologies in the networked world of the National Information Infrastructure and the Global Information Infrastructure. Scholars, researchers, and practitioners from a variety of perspectives must address a large number of complementary research issues, which we organized into three foci: human-centered, artifact-centered, and systems-centered. Some of these research issues can be addressed within individual disciplines but most will require multi-disciplinary teams.

We conclude this report by recommending that research be conducted on these themes, that scholars from multiple disciplines be encouraged to develop joint projects, that scholars and practitioners work together, and that digital libraries be developed and evaluated in operational, as well as experimental, work environments. Only in this way can we build digital libraries to support diverse communities of users in their professional, educational, and recreational activities.

UCLA-NSF Social Aspects of Digital Libraries Workshop

Invitational workshop held at UCLA, February 15-17, 1996

Appendix 1: Agenda

Thursday, February 15

12:00 p.m. - 8:30 p.m. Participant arrivals and registration
7:00 p.m. - 8:30 p.m. Reception at the Summit Hotel Bel-Air (Refreshments, Hors d'oeuvres)

Friday, February 16

7:30 a.m. - 8:00 a.m. Shuttle bus to UCLA
8:00 a.m. - 9:00 a.m. Continental Breakfast at GSE&IS Building

9:00 a.m. - 9:05 a.m. Introduction
Christine Borgman, Chair, UCLA Department of Library and Information Science

9:05 a.m. - 9:15 a.m. Comments
Stephen Griffin, National Science Foundation

9:15 a.m. - 9:30 a.m. Workshop Goals
Christine Borgman

9:30 a.m. - 10:15 a.m. Session 1: Social Context and Culture
Facilitator: Leah Lievrouw Discussants: Philip Agre and Rob Kling

10:15 a.m. - 10:30 a.m. Refreshment Break

10:30 a.m. - 11:15 a.m. Session 2: Information Needs and Information Seeking
Facilitator: Marcia Bates Discussants: Raya Fidel and Gary Marchionini

11:15 a.m. - 12:00 p.m. Session 3: Linking User-Learner Needs and Behavior to Digital Library Design
Facilitator: Yasmin Kafai Discussants: Su-Shing Chen and Nancy Van House

12:00 p.m. - 1:00 p.m. Sandwich Buffet Lunch at GSE&IS Building

1:15 p.m. - 2:00 p.m. Session 4: Organization, Description and Representation of Information
Facilitator: Gregory Leazer Discussants: Karen Drabenstott and David Levy

2:00 p.m. - 2:45 p.m. Session 5: Search Capabilities for Users
Facilitator: Efthimis Efthimiadis Discussants: Edward Fox and Clifford Lynch

2:45 p.m. - 3:00 p.m. Break

3:00 p.m. - 3:45 p.m. Session 6: Interface Design for Information Retrieval
Facilitator: Anne Gilliland-Swetland Discussants: Joseph Busch and Susan Dumais

3:45 p.m. - 5:00 p.m. Campus Free Time

5:00 p.m. - 7:00 p.m. Keynote Address and Reception, Moore Hall 100 and Patio
Keynote Speaker: Clifford Lynch

7:00 p.m. - 9:00 p.m. Dinner in Moore Hall Reading Room, Moore Hall 3340
9:00 p.m. - 9:30 p.m. Shuttle bus to Hotel

Saturday, February 17

7:30 a.m. - 8:00 a.m. Shuttle bus to UCLA
8:00 a.m. - 9:00 a.m. Buffet Breakfast at GSE&IS Building
9:00 a.m. - 10:30 a.m. Topic Breakout Sessions

Session 1: Social Context and Culture, Room 111

Facilitators: Leah Lievrouw and Nadia Caidi.

Participants: Philip Agre, Tora Bikson, Ann Bishop, Rob Kling,
Ronald Rice, Velimir Srica, Susan Leigh Star.

Session 2: Information Needs and Information Seeking, Room 121

Facilitators: Marcia Bates and Susan Schreiner.

Participants: Donald Case, Brenda Dervin, Raya Fidel,
Carol Kuhlthau, Gary Marchionini.

Session 3: Linking User-Learner Needs and Behavior to Digital Library Design, Room 202

Facilitators: Yasmin Kafai and John Schacter.

Participants: Elfreda Chatman, Su-Shing Chen, Paul Conway,
Aimee Dorr, Joseph Krajcik, Nancy Van House.

Session 4: Organization, Description and Representation of Information, Room 208

Facilitators: Gregory Leazer and Marlene Martin.

Participants: Karen Drabenstott, Michele Cloonan, Raymond D'Amore,
David Levy, Daniel Pitti, Cecelia Preston.

Session 5: Search Capabilities for Users, Room 245

Facilitators: Efthimis Efthimiadis and Venkatachallam Maithili.

Participants: Edward Fox, Thomas Landauer, Ray Larson,
Clifford Lynch, Philip Smith.

Session 6: Interface Design for Information Retrieval, DS Lounge

Facilitators: Anne Gilliland-Swetland and Claude Zachary.

Participants: Joseph Busch, Andrew Dillon, Susan Dumais,
Edie Rasmussen, Vicky Reich.

10:30 a.m. - 11:00 a.m. Refreshment Break
11:00 a.m. - 12:30 p.m. Topic Breakout Sessions
12:30 p.m. - 1:30 p.m. Working Lunch on Campus

1:30 p.m. - 3:30 p.m. Breakout reports and discussion
Christine Borgman

3:30 p.m. - 4:00 p.m. Refreshment Break

4:00 p.m. - 5:30 p.m. Final report planning, structure, responsibilities and wrap-up
Christine Borgman

5:30 p.m. - 6:00 p.m. Shuttle bus to Hotel
6:30 p.m. - 7:00 p.m. Shuttle bus to Beverly Hills
7:00 p.m. - 10:00 p.m. Reception and Dinner
10:00 p.m. - 10:30 p.m. Shuttle bus to Hotel

Sunday, February 18

7:00 a.m. - 12:00 p.m. Hotel check-out and participant departures

UCLA-NSF Social Aspects of Digital Libraries Workshop

Invitational workshop held at UCLA, February 15-17, 1996

Appendix 2: List of Participants

Investigators

Marcia Bates, University of California, Los Angeles; mjbates@ucla.edu
Christine Borgman, University of California, Los Angeles; cborgman@ucla.edu
Michele Cloonan, UCLA and Smith College, mcloonan@library.smith.edu
Efthimis Efthimiadis, University of California, Los Angeles; ene@argo.gslis.ucla.edu
Anne Gilliland-Swetland, University of California, Los Angeles; swetland@ucla.edu
Yasmin Kafai, University of California, Los Angeles; kafai@gse.ucla.edu
Gregory Leazer, University of California, Los Angeles; leazer@ucla.edu
Anthony Maddox, University of California, Los Angeles; amaddox@ucla.edu

Staff

Keri Botello, Dept. of Library and Information Science, UCLA; kbotello@ucla.edu
Nadia Caidi, Dept. of Library and Information Science, UCLA; ncaidi@ucla.edu
Jann Cripp, Graduate School of Education and Information Studies, UCLA, cripp@gse.ucla.edu
Lydia Doplemore, Dept. of Library and Information Sci., UCLA; doplemore@gseis.ucla.edu
John Houser, Dept. of Library and Information Science, UCLA; jhouser@ucla.edu
Mary King, Graduate School of Education and Information Studies, UCLA, king@gse.ucla.edu
Renée Kneer, Dept. of Library and Information Science, UCLA; rkneer@ucla.edu
Venkatachallam Maithili, Dept. of Education, UCLA; maithili@gseis.ucla.edu
Marlene Martin, Dept. of Education, UCLA; marl@ucla.edu
John Schacter, Dept. of Education, UCLA; schacter@mailmac.cse.ucla.edu
Susan Schreiner, Dept. of Library and Information Science, UCLA; sschrein@ucla.edu
Claude Zachary, Dept. of Library and Information Science, UCLA; czachary@ucla.edu

Participants

Philip Agre, University of California, San Diego; pagre@weber.ucsd.edu
Tora Bikson, Rand Corporation; tora@monty.rand.org
Ann Bishop, University of Illinois at Urbana-Champaign; bishop@alexia.lis.uiuc.edu
Joseph Busch, Getty Art History Information Program; jbusch@getty.edu
Donald Case, University of Kentucky; dcase@ukcc.uky.edu
Elfreda Chatman, University of North Carolina, Chapel Hill; chatman@ils.unc.edu
Su-Shing Chen, University of North Carolina, Charlotte; schen@uncc.edu
Paul Conway, Yale University; pconway@yalevm.ycc.yale.edu
Raymond D'Amore, Mitre Corporation; rdamore@mitre.org
Brenda Dervin, Ohio State University; bdervin@magnus.acs.ohio-state.edu
Andrew Dillon, Indiana University; adillon@indiana.edu
Aimée Dorr, University of California, Los Angeles; dorr@gse.ucla.edu
Karen Drabenstott, University of Michigan, Ann Arbor; karen.drabenstott@umich.edu
Susan Dumais, Bell Communications Research; std@bellcore.com
Raya Fidel, University of Washington; fidelr@u.washington.edu
Edward Fox, Virginia Polytechnic Institute and State University; fox@vt.edu
Rob Kling, University of California, Irvine; kling@ics.uci.edu
Joseph Krajcik, University of Michigan, Ann Arbor; krajcik@umich.edu
Carol Kuhlthau, Rutgers University; kuhlthau@zodiac.rutgers.edu
Thomas Landauer, University of Colorado; landauer@psych.colorado.edu
Ray Larson, University of California, Berkeley; ray@sherlock.berkeley.edu
David Levy, Xerox Palo Alto Research Center; dlevy@parc.xerox.com
Leah Lievrouw, University of California, Los Angeles; llievrou@ucla.edu
Clifford Lynch, University of California-DLA; Clifford.Lynch@ucop.edu
Gary Marchionini, University of Maryland, College Park; march@oriole.umd.edu
Daniel Pitti, University of California, Berkeley; dpitti@library.berkeley.edu
Cecelia Preston, University of California, Berkeley; cpreston@info.sims.berkeley.edu
Edie Rasmussen, University of Pittsburgh; erasmus@lis.pitt.edu
Vicky Reich, Stanford University; vicky.reich@forsythe.stanford.edu
Ronald Rice, Rutgers University; rrice@scils.rutgers.edu
Philip Smith, Ohio State University; psmith@magnus.acs.ohio-state.edu
Velimir Srica, University of California, Los Angeles; vsrica@ucla.edu
Susan Leigh Star, University of Illinois at Urbana-Champaign; star@alexia.lis.uiuc.edu
Nancy Van House, University of California, Berkeley; vanhouse@sims.berkeley.edu

UCLA-NSF Social Aspects of Digital Libraries Workshop

Invitational workshop held at UCLA, February 15-17, 1996

Appendix 3: Guidelines for Research Agenda Papers and Presentations

We are inviting two people in each of 6 topic areas to make a short presentation that will provoke discussion of a research agenda, based on a very short paper to be distributed in advance. We have scheduled each topic for 45 minutes (see program schedule) in the Friday plenary sessions, consisting of 15 minutes of presentation and 30 minutes of discussion. Discussion will continue in the Saturday breakout sessions, with the goal of drafting a research agenda document by the end of the workshop.

Our aspiration in drafting a research agenda on social impacts of digital libraries is to engage a broader design community that consists of

- traditional systems design communities
- new design communities in the content areas of the digital libraries
- audiences for digital libraries

Your paper is to consist of

- 3 challenges
- 3 opportunities

in your topic area that will lead to the engagement of these digital library design communities. It is to be no more than 2 pages in length; a half-page listing the challenges and opportunities in a succinct, provocative form would be optimal! Electronic mail submissions of the paper, in ASCII format, are preferred. Please include your name, title, address, telephone and facsimile numbers, Internet address, and a 350-word biography.

Topic areas and invited speakers:

Information needs

Session 1: Social context and culture

- Philip Agre, University of California, San Diego
Internet: pagre@weber.ucsd.edu; voice: (619-538-6328)
- Rob Kling, University of California, Irvine
Internet: kling@ics.uci.edu; voice: (714-824-5955, 5160)

Session 2: Information needs and information seeking

- Raya Fidel, University of Washington
Internet: fidelr@u.washington.edu; voice: (206-543-1888)
- Gary Marchionini, University of Maryland at College Park
Internet: march@oriole.umd.edu; voice: (301-405-2053)

Session 3: Linking user-learner needs and behavior to digital library design.

- Su-Shing Chen, University of North Carolina, Charlotte
Internet: schen@uncc.edu; voice: (704-547-4885)
- Nancy Van House, University of California, Berkeley
Internet: nav-lis@cmsa.berkeley.edu; voice: (510-642-0855)

End user searching and filtering

Session 4: Organization, description and representation of information

- David Levy, Xerox Palo Alto Research Center
Internet: dlevy@parc.xerox.com; voice: (415-812-4376)
- Karen Markey Drabenstott, University of Michigan
Internet: karen.drabenstott@umich.edu; voice: (313 763-3581)

Session 5: Search capabilities for users

- Edward Fox, Virginia Polytechnic Institute & State University
Internet: fox@vt.edu; voice: (540-231-5113)
- Clifford Lynch, University of California-DLA
Internet: Clifford.Lynch@ucop.edu ; voice: (510 987-0522)

Session 6: Interface design for information retrieval

- Joseph Busch, Getty Art History Information Program
Internet: jbusch@getty.edu; voice: (310-395-1025)
- Micheline Beaulieu, City University, London
Internet: mmb@is.city.ac.uk;

UCLA-NSF Social Aspects of Digital Libraries Workshop

Invitational workshop held at UCLA, February 15-17, 1996

Appendix 4: Background Paper

Christine L. Borgman
Marcia J. Bates
Michele V. Cloonan
Efthimis N. Efthimiadis
Anne Gilliland-Swetland
Yasmin Kafai
Gregory H. Leazer
Anthony Maddox

Graduate School of Education & Information Studies
University of California, Los Angeles

June, 1995

Overview Of Research and Application Issues

Digital Libraries is a National Challenge Application designated by the Information Infrastructure Technology and Applications Task Group under the High Performance Computing and Communications Initiative. The Digital Libraries application has brought together researchers from computer science, communications, library and information science, psychology, linguistics, and from the disciplines in which digital libraries are being created, including the sciences, social sciences, arts, and humanities. National Challenge projects are intended to focus on large societal problems and bring human and technological resources to bear on their solution. Digital Libraries are a prime example of such problems, for they cross all disciplines and all sectors of society.

Many social aspects of digital libraries need to be addressed, as we come to understand the full range of issues they encompass. The research workshop will focus on two social problems that are urgent in developing the National and Global Information Infrastructures:

- Information Needs: Identifying real information needs and developing digital libraries to meet those needs.
- End User Searching And Filtering: Designing digital libraries in which it is possible to find the right information in a glut of information.

We have chosen these two problems because they are urgent, enough research exists to frame them but not enough to solve them, and the work on these problems is scattered across multiple disciplines that need to be brought together to form a research community.

Other social aspects of digital libraries include use and usability by a range of user populations; ethical concerns; data/information validation, authentication, and peer review issues; cognitive authority (how can we trust what we are seeing/reading?); privacy vs. accessibility; short-term development vs. long-term preservation (cutting edge vs. standards); user costs and the impact of commercial components of the library on users; and the power and biases of digital libraries for the process of transmitting and shaping culture and cultural heritage across geographic and temporal boundaries. The real potential for digital libraries revolves around being able to think outside the

scope of the system -- imagining new possibilities and paradigms for the collaborative development, maintenance, and use of knowledge as derived from information content, context, and structure. Although we use the term "library" we are actually building entities that blend not only information types, media, and uses, but also professional and disciplinary approaches to their construction. For digital libraries to achieve their full potential, technologically and socially, we should be able to capitalize on any disciplinary or professional paradigm for arrangement and description that might add richness and utility, whether that of libraries, archives, museums, or other perspectives.

While we will focus on the two primary themes, we will set them in the context of the other issues above. The goal of the research workshop is to identify specific research questions that need to be addressed to further research in digital libraries. We expand on these themes:

Information Needs

Historically, much of information retrieval research has taken the information query as a given. That is, the user comes to the system with a query, while the source of the query, and the ultimate usefulness of the information retrieved to meet that query are not examined. But, in fact, users tend to ask questions of information systems that they think, rightly or wrongly, the system can answer. There may be other types of queries, other types of information resources, and other social and institutional ways of making the information available that are needed and are not revealed when only the information retrieval system design itself is studied.

Several linked areas of research need to be examined and modeled in order to produce the desired end result of satisfied users meeting real needs.

Social Context and Culture: Information needs must arise from somewhere. Researchers, professionals, and schoolchildren are seeking information in a dense and complex social context. Information seeking often arises out of a matrix of social pressures, expectations, and mores, as well as from an individual's thought processes. Research in scholarly communication and the sociology of science has described much of this social context. Research is in its infancy, however, on the link between that context and the particular information needs and information seeking behaviors that arise out of that context.

Much of the research on digital libraries may assume implicitly that basic components such as document representation, interfaces, and retrieval algorithms can be generalized across document types, user groups, and application domains. This assumption has not been tested explicitly -- and research on the social context of information needs suggests that such generalization may not be possible. We may need to tailor many aspects of digital libraries to their environment. As the NII becomes the GII and we build multi-lingual, multi-media, multi-level digital libraries, the generalizability issues will be critical.

Information Needs and Information Seeking: The large body of research on information needs of various groups consists mostly of cross-sectional studies in which average percentages of types of need or of types of resource used are discovered. With this body of research as a basis, what is needed now are more organic studies of behavior, in which particular users are followed through time in solving their information problems, and types of need are seen to be in relation to particular types of conditions encountered by users. We need to move from the study of the objective facts of the various types of use to a study of the meaning, motivation, and logic that drive the user from one action to the next. With such information, we can then design information systems that facilitate the user in following a natural-feeling path to the desired end result in an information search.

Most of the research in this area has focused on the information needs and uses of professionals or experts in a subject domain. Building digital libraries to exist on the NII/GII means creating information spaces that can serve the needs of novices in a subject domain, especially students of all ages. The increasing use of computational media to support learning activities in school settings introduces a different kind of user with some distinctive features: whereas professionals know the domain, are motivated, and are an homogeneous population with the goal to increase their success, students do not know the domain, often are not motivated, and encompass very diverse populations.

While this distinction between users and learners could simply define learners as one subgroup of users, we need to recognize that learning is not just for students in the classroom but professionals are (or should be) constantly learning too. Moreover, when the professional is acting as a learner, that person is susceptible to all the challenges faced by students. Information seeking and learning appear to be closely related cognitive activities, but this relationship has not been studied explicitly, as the research tends to be conducted in different disciplines.

Linking User Needs and Behavior to System Design: Many of the research studies on users and many of those on information retrieval system design and improvement have been conducted independently of each other. We need to start with the results of research on users, draw implications for information system design from those results, and then design and test systems that better meet real user needs.

In the last ten years, human computer interaction (HCI) research has been dominated by the view that the user should be at the center of software environment design to make computers easier-to-use (propagated by such seminal publications as Card, Moran, and Newell's "Human Computer Interaction" (1983) and Norman and Draper's "User-Centered Design Systems" (1985)). Most software design places the user at the center of three essential issues: the tasks that need to be undertaken by the software, the tools that are provided by the software to cope with the task and the interfaces to those tools. Placing the learner at the center recognizes the special needs such as understanding the goal, the motivation, the diversity and the potential growth of the learner-user of digital libraries. While research exists specifically at the intersection of HCI and information retrieval, the HCI perspective has not been a strong influence on IR system design overall.

End User Searching And Filtering

Information retrieval research generally has focused on a model of retrieval in which the user presents a query to the system, the system searches, sometimes with user relevance feedback, and then comes up with the best answer possible within the design of the system. The emphasis has been placed on finding all the relevant records in the system, with as few irrelevant ones being retrieved as possible.

As information systems and computer capabilities become more sophisticated, users are able to conduct much more interactive searches, in which they use a variety of search techniques in a variety of sources over time for a given search. Users often want to do the searching themselves. The process of searching and seeking preliminary results enables them to clarify their information needs in their own minds as they go along--without having to articulate the query for a search intermediary or an automatic information system. Currently, users may not want every generally relevant record in the system, but rather they need a way to filter out the few records that are sufficient and of good quality for their purposes. Filtering is the process of sifting and winnowing through a retrieval set, finding potentially interesting records. To facilitate this process, descriptive records must describe the information resources accurately enough, relative to the user's perception of the question, to discriminate between relevant and irrelevant records. With the right kind of

support through sophisticated system design, the user can interactively filter and refine search results until a satisfactory retrieval set is achieved.

In this context, digital library design needs to refocus (or add to current research streams) in two ways: looking more at ways to help the user in doing the searching, rather than aiming for the system to do it all for the user, and providing tools to the user to aid in filtering.

Both of these objectives can be simultaneously met through research in three areas:

Organization, Description, and Representation of Information: A mix of automatic and human intellectual organization and indexing has proven quite robust in information retrieval research. Much research is needed on optimal methods to organize information to aid the ultimate end user in searching and filtering in interactive searching.

To be able to facilitate the information seeking process, we also need to be able to understand how and why people create the information in the first place (assuming that the scope of some of the digital libraries encompasses such objects as raw data, full text of papers, remotely sensed data, clinical imaging, and user annotations). Trying to facilitate such an understanding leads to issues of the primary and secondary functionality of information objects, the structure of those objects, and documentation and exploitation of their context. For example, an object's relationship to similar materials, or materials that are part of the same transaction, or materials that are generated by the same process or function. The successful development of various searching agents and an investigation of how they might work together is a requirement for the development of successful large-scale digital library projects.

Search Capabilities for Users. If users are to take a more active role in their own information searching, then the digital library should provide them with an array of search capabilities that match their needs and preferences as they proceed in a search. For example, the user might have available a number of different types of intelligent agents, each of which searches in a different way in the files -- one looking for text words or phrases in titles, another searching for shapes in image files, still another looking for broadly-coded classificatory categories, etc.

Interface Design for Information Retrieval. We need to study both general interface design issues and those specific to the information retrieval situation. For instance, different types of indexing of the digital library may require different types of on-screen arrangements and search capabilities for the user.

As large-scale digital libraries become widely available on the NII and GII to a broad user community, the information process cycle will be extended to include users-learners' incorporation of the information retrieved into their own information environments. Information seeking, retrieving, and use is an iterative process. We should consider how learners can store the information found in a way that is beneficial to their learning experience. In this environment, we can study the kind of information structures and links that learners build to record their search processes, which will assist in designing digital libraries that support the entire information cycle. The construction of any database or information structure can be considered a learning experience, which is an aspect of digital libraries that has received little attention, if any, from the research community. As we seek to expand our understanding of information seeking and use in a social context, we also expand the scope and nature of interface design for information retrieval.

Summary

The research workshop on the social aspects of digital libraries will address two problems that are urgent in developing the National and Global Information Infrastructures: (1) Information Needs: Identifying real information needs and developing digital libraries to meet those needs; and (2) End User Searching And Filtering: Designing digital libraries in which it is possible to find the right information in a glut of information.

Each of these problems requires research on multiple issues that cross multiple disciplines, primarily library and information science, education, computer science, communication, and some of the problem domain areas. Many of the researchers working on these problems would not identify themselves as addressing digital libraries problems. If these problems are to be addressed adequately, however, we need to bring together key people from these various disciplines, both those who identify themselves as digital libraries researchers and those who do not. Our goal is to form a research community that can focus on the social aspects of digital libraries. The product of the workshop will be a research agenda that will be widely distributed to the various constituent communities in hopes of stimulating research that converges on these problems.

Workshop Topics

The workshop will identify the research questions to be addressed in the social aspects of digital libraries related to these topics. We propose the following research questions to provide starting points for discussion:

Information needs

- Social context and culture

To what extent can digital library interfaces, information retrieval algorithms, intelligent agents, and other system components be generalized across application domains and to what extent must they be tailored to each environment?

- Information needs and information seeking

To what extent are information needs and uses generalizable across user and learner groups and to what extent do they need to be tailored?

What is the relationship between information seeking and learning in digital libraries?

- Linking user-learner needs and behavior to digital library design.

What systems design techniques are appropriate in applying user needs research to digital library design?

End user searching and filtering

- Organization, description and representation of information

Which methods of organization can be generalized for digital libraries applications? Which cannot? How can methods developed for single database, single system applications be adapted to multiple database distributed applications?

How well do current standards and structures work, such as the Anglo-American Cataloging Rules (AACR), Machine Readable Cataloging (MARC), SGML, TEI, UNICODE, etc.? How do these standards interact and conflict? What new standards are needed? How useful will these and other standards be in facilitating multi-lingual, multi-media, multi-level information retrieval in the Global Information Infrastructure?

- Search capabilities for users

What search capabilities are specific to individual problem domains and which are generic? How should problem domain areas be divided? By subject area (e.g., science, medicine, arts), by age group (children, adults), by problem goal (e.g., fundamental research, business application), by form of content (text, numeric, graphics, moving images, sound), etc.

- Interface design for information retrieval

What human-computer interaction principles can be applied to the information retrieval environment and which are unique to IR? How can we extend interface design to encompass a broader definition of the information process cycle?

How can we facilitate interaction among the various digital libraries communities, and the related communities providing the technical computing and communications infrastructure on which digital libraries rely?

UCLA-NSF Social Aspects of Digital Libraries Workshop

Invitational workshop held at UCLA, February 15-17, 1996

Appendix 5: General Conference Information

Hosted By:

The Department of Library and Information Science
Graduate School of Education & Information Studies (GSE&IS)
University of California, Los Angeles

Sponsored By:

Information Technology and Organizations Program
Information, Robotics, and Intelligent Systems Division
Computer and Information Science and Engineering Directorate
The National Science Foundation

Co-Investigators:

Marcia J. Bates
Christine L. Borgman, Principal Investigator
Michele V. Cloonan
Efthimis N. Efthimiadis
Anne Gilliland-Swetland
Yasmin Kafai
Gregory H. Leazer

Workshop Coordinator:

Anthony B. Maddox
1-310-206-5865 (voice)
1-310-206-4460 (facsimile)
digital-libraries@gslis.ucla.edu (Internet)

Additional members of the steering committee:

Dan Atkins, University of Michigan at Ann Arbor
Edward Fox, Virginia Polytechnic Institute and State University
Michael Lesk, Bell Communications Research
David Levy, Xerox Palo Alto Research Center
Clifford Lynch, University of California-DLA
Gary Marchionini, University of Maryland, College Park

GSE&IS Staff Coordinators:

Keri Botello
Lydia Doplemore
John Houser
Renée Kneer