<u>**An unpublished study**</u>

Digital Library Research and Digital Library Practice: How Do they Inform Each Other?

Tefko Saracevic and **Marija Dalbello**

School of Communication, Information and Library Studies, Rutgers University,

4 Huntington Street, New Brunswick, NJ 08901. Email: {tefko,dalbello}@scils.rutgers.edu

The study surveys two large sets of activities concentrating on digital libraries to examine the following questions: ***Does digital library research inform digital library practice? And vice versa? To what extent are they connected, now that nearly a decade has passed since they began?*** **Examined were research projects supported by the first and second Digital Library Initiative (DLI), digital library projects listed by the Association for Research Libraries (ARL) and Digital Library Federation (DFL), and selected literature, focusing on the last five years. Methods concentrate only on examination of visible or "surface" sources or records, i.e. information that can be gathered from web sites, open literature, and published data. Limitations of the method are acknowledged; accordingly, caveats are made about conclusions. From this data we conclude that the two activities are not as yet demonstratively connected. A set of differing interpretations and conclusions are included.**

1. Introduction

In many fields, research and practice have a complex relationship or connection. In an ideal paradigm, (some) research, particularly toward the applied end, informs and even transforms practice and (some) practice informs research, especially in the selection of problems. Research and practice converge. However, in reality it rarely works exactly that way. The links between research and practice are neither always linear nor are they often easy to discern. Their connections may be serendipitous. Time and social context play a significant role as well. Transfer of ideas is complex, as the classic Rogers' (1995) study of diffusion of innovation, and Bijker's (1994) study of

sociotechnical change have amply demonstrated. There are further considerations. Research often raises expectations, and, by definition, it neither promises nor produces predictable outcomes. Practice may advance without direct input of research.

In this study, we are trying to examine the complex relations and connections between research and practice in the area of digital libraries solely through records that digital library projects in both research and practice generated on their web sites, and from the literature reporting on digital libraries. In other words, we concentrate solely on visible or "surface" evidence. The strengths and limitations of the method are elaborated in the methodology section and again revisited in conclusions at the end.

We asked the following questions related to numerous activities in digital libraries:

- *Does digital library research inform digital library practice? And vice versa?*

- *To what extents are they connected now, nearly a decade after they began?*

"Digital library research" refers to projects in Digital Library Initiatives (DLI) 1 and 2 (described below) and research reports in the literature. We interpret "digital library practice" to include any working digital library (as categorized below), and/or demos or testbeds reflecting any practical, operational library-oriented achievements. "Inform" refers here to a visible connection based on evidence (1) in the sites of research projects and in the research literature that points to any consideration of or link to an operational digital library project, or to demos, and testbeds, or (2) in digital library practice any consideration of or link to research projects in DLI, or any other research**. Research and practice we covered are mostly US based and oriented; we did not cover similar and sizable activities elsewhere.

2. Framework

Big science, as characterized a generation ago by Derek de Solla Price (1963), is heavily institutionalized, subsidized, and driven by pre-set agendas. In the U.S., research agendas and subsidies are generally set by national agencies chartered to support research, such as the National Science Foundation (NSF), National Institutes for Health (NIH), and others, often in consultation with different constituencies, including researchers. For some time,

research supported by NSF is to a large extent directed toward pragmatic problems, with aims to push the envelopes of applications and extend innovation. The reasons are political, economic, and social; a payoff is to be expected.

In the U.S., the agenda for digital library research is under the same umbrella. It is set and conducted through multiagency Digital Library Initiatives (DLI) lead by NSF. While the agenda is set by participating agencies, constituencies have been consulted in various ways, e.g. through NSF organized workshops. DLI 1 (1994-1998) involved six projects and some $24 million; DLI 2 (1999-2006) involves 77 projects in various programs and some $60 million (but it is hard to find the overall sum). While the agendas for both DLIs were relatively broad, their base rested firmly in technology (Lesk, 1999; panels in Schatz & Chen, 1999). These agendas are the primary (if not the only) driving force for digital library research in the U.S. since its beginnings in the early 1990s. In his keynote address to the Association for Computing Machinery (ACM) Digital Libraries '99 conference, David Levy (2000) concluded that "the current digital library agenda has largely been set by the computer science community, and clearly bears the imprint of this community's interests and vision. But there are other constituencies whose voices need to be heard."

Starting in 2001, NSF also funds a newer, related and larger program, National Science Digital Library (NSDL), subtitled as "The comprehensive source for science, technology, engineering and mathematics education." The NSDL mission, as stated on its web site, is: " … to both deepen and extend science literacy through access to materials and methods that reveal the nature of the physical universe and the intellectual means by which we discover and understand it." We did not explore NSDL because it just started when we begun our analysis and furthermore, because their primary emphasis is on education. It includes components of digital libraries, but also many other and different aspects and projects. For instance, while it includes projects such as "*A Digital Library of Ceramic Microstructures"* and "*Bridging the Gap Between Libraries and Data Archives*," it also has projects such as "*Thematic Real-time Environmental Data Distributed Services (THREDDS)*," and "*Virtual Telescopes in Education (TIE)*" (Zia, 2001). However, as they mature, a number of NSDL projects should be explored as to a connection to digital libraries in general..

Digital library practice is institutionally/organizationally based and oriented toward a given community, pragmatic development, and practical operations. As expected, the aims are toward pragmatic problems at hand. Among others, this involves:

- Digitizing and providing access to specialized materials in possession of many institutions, such as the [American Memory Project of the Library of Congress](#).

- Incorporating digital dimensions and providing access to electronic collections and resources, with a variety of associated services (i.e. creating and managing so-called hybrid libraries) by hundreds of academic, research, public, and special libraries, such as the [U of California at Berkeley's Sunsite Digital Library](#).

- Building digital libraries by professional and other organizations, such as the subscription-based ACM (Association for Computing Machinery) Portal, incorporating the [ACM Digital Library](#).

- Developing collections in specific domains, such as the [Perseus Digital Library](#), covering materials from antiquity to the Renaissance.

These activities are hardly a decade old, but their explosive growth resulted in hundreds of projects and practical digital libraries.

Practical efforts in digital libraries share a common characteristic. Agendas were set at grassroots, by individual libraries, academic departments, professional organizations, museums, publishers ... often driven by enthusiastic individuals. Pioneering projects from the early 1990s, such as those at the Library of Congress mentioned above, served as examples for a great many institutions to follow. Electronic publishing, the development of digital collections, preservation, and management of digital resources with myriad issues and challenges above and beyond technology are also part of these pragmatic efforts.

In sum, the efforts and expenditures in both digital library research and digital library practice are substantial and the question of their connections is warranted and important to raise. But, the answers are not easy to discern and interpretations may differ. Our study aims to open a dialogue on the nature of these connections at present.

3. Methodology

Our study is qualitative and impressionistic, with all the well-known strengths, weaknesses, and limitations of such studies. Basically, the strengths lie in the power to analyze and interpret evidence that is qualitative in nature, and the weaknesses are connected with the lack of formal testing of hypotheses and resulting interpretations that may be more subjective. To some extent, our approach is also related to bibliometrics and webmetrics, in that we also derived some statistics from the data.

We culled data from publicly available web sites, articles, citations, and databases. We examined in detail web sites of many projects and digital libraries, as described below. We simply took them "as is," using the public statements they offered as of **January and February 2002**, about their goals, activities, results, and publications. We gathered data that was publicly available through these sources; we use the term "evidence" in that limited sense. We did *not* evaluate anything - any program, project, or results.

We used a classification of research projects, practical projects, and literature to characterize and sort the findings, as described below.

The limitations of the study are as follows. Examination of "surface" or visible data, while powerful evidence, is limited. We did *not* explore relations and connections between research and practice that are based on transfer and translation of ideas, results, and practices through a variety of indirect means and "invisible" contacts, which often happen and which may provide a fuller and possibly even different picture. For instance, we did *not* examine contacts through conferences, tutorials, and similar gatherings where much transfer may take place. We did *not* conduct interviews with participants in digital library research or practice, which may reveal much more. We did *not* examine any context, role of organizations, or any connection to predecessors or related activities. We did *not* investigate where do people in research or practice get their ideas. We stuck only to that that is visible in public record. This means that we have ignored the tacit knowledge that may be underlying information transfer in this field of activity.

4. Digital library research

In order to answer: *To what extent can we find evidence(in the sense as described above) that projects in Digital Library Initiatives are connected in some way to digital library practice*?, we visited all of the available Web sites of projects in DLI 1 and 2.

As to the literature, the papers in Harum & Twidale (2000) described and, to some extent, evaluated DLI 1 projects; some of the discussions in the compendium have relevance to the question raised here. Otherwise, we could not find in the literature any other assessment or evaluation of DLI 1 or 2 projects or of DLI as a research program, for possible use in relation to questions raised in this study, aside the paper by Levy (2000) already mentioned.

4.1 Digital Library Initiative 1

DLI 1 included six institutions, funded from 1994-1998, as listed by the National Science Foundation. It would be more advantageous to have the benefit of detachment provided by time and distance from the projects. Instead, looking at current projects through the lens of their sites provides immediacy yet makes it hard to discern what was actually accomplished. The results can be only surmised. Four DLI 1 projects are continuing into DLI 2 projects (UC Santa Barbara, Berkeley, Carnegie Mellon, and Stanford) and their sites incorporate both projects with minimal, if any, differentiation. The results of site visits show the following connections of research and practice:

1. University of California at Santa Barbara's "Alexandria Digital Library (ADL) Project" concentrated on developing tools for and a collection of geographic data and map browsers. The project does have a visible practical connection; the University's Davidson Library hosts the ADL map browser and catalog with a link to the California Digital Library (CDL), encompassing the nine campuses of the University of California system. The project bibliography lists close to 140 entries. With very few exceptions, the publications are oriented toward computer maps and spatial information, but many reflect work beyond the

project. The project has been continued in DLI 2 under the title "Alexandria Digital Earth Prototype (ADEPT)," with a practical component as one of the goals.

2. University of California at Berkeley's "Environmental Planning and Geographic Information Systems". However, the site refers only to the current project in DLI 2 under the title "Re-inventing Scholarly Information Dissemination and Use." It is hard to find results from the DLI 1 project. Most of the materials on the site refer to images; it is not clear how that content is connected to the current title. The site leads to "Digital Library Collections" consisting of image files, and botanical, zoological, and geographic data, including about 30,000 photographs of California plants, documents on California environment, and links to maps and databases such as "Museum of Vertebrate Zoology Data Access". It also provides access to Blobworld, a Corel collection of 35,000 images and a search engine for images by keyword or shape (blob). These are practical demonstrations. About 40 publications are listed in two Progress Reports (1996 and 1998). Some are about digital libraries in general; some about user studies, and others are related mostly to computer images and vision.

3. Carnegie Mellon University's "Informedia Digital Video Library." The description for both DLI 1 and DLI 2 projects is rolled into one. It deals with "how multimedia digital libraries can be established and used." It does have a separate page for Informedia 1, done under DLI 1, and offers a description of an approach to integrating multimedia objects into a collection. A demo under Informedia 2 is "under construction." It lists some 60 publications, mostly on computer vision and multimedia.

4. University of Illinois at Urbana-Champaign's "Federating repositories of scientific literature." A practical result is the "UIUC Digital Library Testbed," described as "providing access to the full-text of articles from over 50 journals in civil engineering, computer science, electrical engineering, and physics" through DeLIver, an experimental search system, also available through the engineering library. For the DLI 1 project, some 100 publications are listed; they treat a wide range of topics even above and beyond the topic of the project, and include a number of user studies.

5. University of Michigan's "Intelligent agents for information location." While demo sites are mentioned, no connection to a prototype, testbed, or practical library can be discerned. About 60 publications are listed. The topic most discussed is intelligent agents, but many publications are above and beyond the project. No other results are identified from the site. Based on what is on the web, it seems that this DLI 1 project has the least results and connections.

6. Stanford University's "Building the InfoBus: Interoperation mechanisms among heterogeneous services." The site merges the DLI 1 project with the current project in DLI 2 under the title, "Stanford Digital Library Technologies Project." DLI 1 is reflected through a review of technical accomplishments. The review lists 12 publications, while the list of "Working papers" on the site lists some 140 publications on a wide variety of topics, many above and beyond the project. A testbed is provided. There is a link from the project site to the University Library although we could not discern any connection from the Stanford U Library site to the project or testbed.

Literature, of course, is an important vehicle for communicating and informing, thus we took a closer look at the literature or bibliographies on DLI sites. A large proportion of the items listed in all of the projects belongs to gray literature – technical reports, notes, annual reports and the like that are difficult, if not impossible, to retrieve by subject access, thus for all practical purposes they are invisible. Of the open literature, the largest proportions by far are papers in conference proceedings by various ACM Special Interest Groups (SIGs). Small proportion is journal articles. Overwhelmingly, the literature is oriented toward computer science and scientists, rather than other fields or practice. This is not surprising, for a large majority of investigators listed in the projects were associated with a computer science department; five out of six (83%) Principal Investigators (PIs) were from computer science, one from geography. While there were many other investigators and project participants, it was not possible to investigate fully their composition on the basis of available data. But most of them listed a computer science department as their affiliation.

We classified the projects into *domain-oriented* (concentrating more on techniques of use in specific domains, topics or subjects) and *general technology* (concentrating more on techniques that are domain independent, even

though examples may involve given domains). Two projects (33%) were domain-oriented (UC Santa Barbara and UC Berkeley), while the rest were more oriented toward general technology.

As shown here, two of the projects (UC Santa Barbara and Illinois) established a visible connection with a practical digital library, i.e. a library at their universities. The Corporation for National Research Initiatives (CNRI) is sponsoring the D-Lib Test Suite — "a group of digital library testbeds that are made available over the Internet for research in digital libraries, information management, collaboration, visualization, and related disciplines". Included are testbeds from Carnegie Mellon, Cornell, UC Berkeley, UC Santa Barbara, Illinois and Tennessee-Knoxville. Out of these six testbeds, four are from DLI 1 projects and their continuations in DLI 2. These could be considered as practical demo-outcomes. However, from the information provided, we cannot discern if they are actually being used, and if so, how and by whom. There is no literature on the use of these testbeds that we could find. This is in contrast to the testbeds provided by the Text Retrieval Conference (TREC); the results from use of TREC testbeds in testing various approaches to information retrieval (IR) are widely reported in open literature and technical reports.

Thus, either through testbeds or through a library link, four out of six DLI 1 projects (66%) have visible links to practice.

4.2 Digital Library Initiative 2

Under "DLI 2 Funded Projects," the NSF site lists 77 projects comprising 28 main projects, eight projects with undergraduate emphasis, 11 international projects, 14 in the Special Projects Program, and 16 in the Special Projects in Information Technology Research Program. These are funded for the period 1999 to 2006, however, some are targeted for shorter periods or different start years. The amounts for all projects range from $33,000 to $7.5 million. For this analysis, we concentrated on the 28 main projects only. We did not include study of other than the 28 major projects, basically because their emphasis is less on digital libraries, and more on some other aspect, such as education.

Of the 28, 18 (64%) can be classified as domain-oriented, and 10 as general technology. This is a significant shift from DLI 1 projects, where only 33% were domain-oriented. Of the PIs, 15 (53%) were from computer science departments, and the rest from a range of other departments — languages, classics, philosophy, sociology, geography, geology, history, and biomedicine. This is also a significant difference from DLI 1, where 83% of PIs were from computer science departments. Still, from the list of all the investigators in addition to PIs, a large majority is from computer science. In general, DLI 2 is much more domain-oriented than DLI 1, and the spread of disciplines involved is wider. The reason may be that the spread of agencies involved in DLI 2 is also wider.

Of the 28 projects, two have no direct link from the NSF site; one of these (Illinois) has a missing link and one (South Carolina) has an invitation for students to participate but no other information. For these two, we made no further effort to find project information (if it exists at all). Four projects from DLI 1 were also continued in DLI 2, (UC Berkeley, UC Santa Barbara, Carnegie Melon, Stanford) and they were discussed above. That means we further investigated 22 DLI 2 projects.

The amount and quality of information that can be gleaned from these 22 sites is highly uneven. Five include a demonstration of actual practical libraries in their domains, but no DLI project information beyond that (UC Davis, Eckerd, Johns Hopkins, one of Stanford's three, and Tufts). Some of these projects existed prior to (and independently of) DLI 2. However, it is not clear whether what is shown are the developments before or after DLI 2, but it is clear that these represent practical digital libraries. Nine sites show demos of their work (Arizona, UCLA, Columbia, Harvard, one of Indiana's two, Hawaii, Massachusetts, Oregon, Texas). The rest have project descriptions of various depths.

Counting all 28 DLI 2 projects as to practical results (including those that have been continued from DLI 1), 17 (61%) have so far produced a practical digital library or are showing demos of their results on their sites. Not surprisingly, the majority or 13 of the 17 (76%) are domain-oriented; the other four are technology-oriented.

Thirteen projects also provide a list of publications ranging from 3 to 70. Included are technical reports and other gray literature; some proceedings papers; and a few journal articles. Many publications are general, above and beyond the project; some are dated even long before the project.

We conclude that, although a number of DLI 2 projects have practical results or dimensions, it is too early to discern the overall results.

5. Digital library practice

We considered several information sources to tap into the large and diverse universe of digital library practice:

- Digital library projects as identified in databases of the Association of Research Libraries (ARL) and Digital Library Federation (DLF). ARL has 125 members in North America. DFL "is a consortium of libraries and related agencies that are pioneering in the use of electronic-information technologies to extend their collections and services" which has 28 partners.

- Operational digital libraries as identified in Libweb, a directory of library servers on the web at UC Berkeley. "Libweb currently lists over 6100 pages from libraries in over 100 countries."

- The "Featured Collection" appearing in each issue of D-Lib Magazine.

- Digital libraries in professional societies: ACM Digital Library and the IEEE/IEE Electronic Library.

This section is divided into "projects" and "operations." The "projects" refer to a variety of developmental and operational undertakings by a variety of organizations, as described below. "Operations" refers to operational digital libraries. Many of these also include digital library projects. (We took the established terminology of NSF and ARL — both refer to "projects," but very different types of projects).

A 2001 survey of digital projects in 21 large libraries in DLF provides, among other things, some economic data (Greenstein, Thorin, & Mckinney, 2001) indicating that the average expenditures for these libraries for year 2000 related to "digital activities" was $4.2 million; the range of annual investments in "content creation" was from $5,000 to $3 million. ARL report states that 105 large research libraries in the U.S. spent close to $100 million in 1999-2000 just for the purchase of electronic resources, an increase of $23 million over the previous year (ARL, 2001). While these surveys cover only a portion of libraries, the figures illustrate the magnitude of expenditures for electronic resources in digital libraries and digital library projects.

5.1 Projects

ARL maintains the ARL Digital Initiatives Database, "a Web registry for description of digital initiatives in or involving libraries", while DLF maintains Public Access Collections, a "searchable database of members' public domain, online digital collections." The ARL database lists 427 digital library projects in 13 countries. Of these, 374 are in the U.S.. DFL lists 288 collections from 28 member institutions, all in the U.S. There is some overlap: 78 projects/collections (all U.S.) are listed in both databases. Thus between the two, we have 584 (U.S.) listings. ARL database is broader than the DLF database and not restricted to members. Typically, the projects involve digital conversion and development of digital collections, including encoding, organizing, and providing access to texts, images, video, film, and sounds, or technological developments.

Table 1. of institutional or organizational affiliation of 374 US projects follows. It shows that a whole range of institutions, along with libraries, is involved in digital library projects. Included in the list is the DLI testbed from Illinois, as the only connection between DLI and ARL projects.

Table 1. Distribution of projects in the ARL database according to institutions or organizations under which the projects are listed (N=374 U.S. projects)

Institutional/organizational affiliation	No.	%
Universities	191	51%
Agencies of the federal government (such as Library of Congress or Smithsonian);	91	24
Historical societies, museums, or archives	32	9
Trans-institutional (collaborative, regional)	23	6
Public libraries	18	5
Professional societies	6	2
Publishers	2	>1
Other*	11	3

*one each is in a gallery, school or bibliographic utility; one is a DLI testbed; and seven fall into
other categories.

The domains of these projects vary widely including *souvenir photographs of battlefields of Virginia; Philadelphia's Chinatown; architecture in fine prints; guide to African-American resources; 14th century physician's belt books; Al Capone; talking history*. Of the 374 projects, the majority, or 315 (84%) were domain oriented, and explicitly retrospective in nature, having a strong historical component. The rest, or 56 projects

(15%), deal with a variety of technological and operational aspects (electronic reserve, digital photo duplication, support services for digital library developers, and the like). Distribution of more specific domains in the 315 domain projects is presented in Table 2.

Table 2. Distribution of projects in the ARL database according to domain or (N=315 U.S. projects)

Domain or topic of projects	No.	%
Historical treatments of various topics and cultures, such as the history of the written word and print culture	65	21%
Topics related to Europe	46	15
U.S. regional	43	14
Organizations, firms, societies, universities and their histories	32	10
Biographical	30	9
Ethnic and immigrant communities	30	9
Periods and events in U.S. history	22	7
Archival and library collections, their retrospectives and histories	18	6
Social and political movements	12	4
Other countries or continents, exclusive of Europe	9	3
Geographic	8	3

Several agendas are evident among the projects:

- Providing access to specialized materials and collections from an institution (or several institutions) that are otherwise not accessible to the broader public,

- Covering in an integral way a topic with a range of sources, or

- Providing technological support for specific functions in digital libraries.

The host institutions funded the majority of the projects. External funding sources supporting the projects include grants from federal agencies (National Endowment for the Humanities, National Endowment for the Arts, and Department of Education; local or state governments; a variety of foundations (David and Lucille Packard Foundation, Andrew W. Mellon Foundation, Reuters Foundation, Arthur R. Marshall Foundation, and the Ahmanson Foundation); and corporations (such as Ameritech, and BellSouth). This demonstrates broad support for the ARL-listed projects, and thus a wide interest in digital libraries.

Of the 28 institutions in DLI 2, 17 (60%) had also ARL-listed projects. All together there were 64 ARL projects listed in these 17 institutions. One of the 64 was listed as a DLI project (Illinois). The other 63 had no

connection with DLI or vice versa. Thus, the institutions that host a DLI 2 project also host a number of other, practical digital library projects, but beyond that single one, we could not find any other connection. While in the same institution, the vast majority of practical projects are independent of DLI projects and vice versa.

In sum, as far as we can determine, the ARL-listed projects, by and large, have been developed and are maintained independently of DLI research. These were developmental projects with direct operations as a goal, with little or no evidence of any involvement of research beyond that of development.

5.2 Operations

In addition to the projects reviewed, there are numerous operational digital libraries offering access to electronic resources and a range of services. It is next to impossible to enumerate all the libraries and other organizations/institutions that have an operational, practical digital library actively serving their communities. Many existing libraries have a Web presence as listed in Libweb. As to the U.S., Libweb lists close to 1,000 libraries, with a high number having a digital library incorporating digital resources, collections, and services of one sort or another. By now, all larger academic and research libraries have an associated digital library, but the depths of their digital collections and services vary significantly. In addition, library consortia now reach small libraries. For instance, OhioLINK, a statewide consortium of 66 academic libraries in Ohio, enables all of them, small and large, to have access to extensive digital collections and services. While technology issues and developments predominate, as "hybrid libraries" they are facing many other challenges as well (Schwartz, 2000).

A sample of 58 U.S. libraries listed in Libweb was examined for links with research. First, we went to the 28 libraries of universities that have DLI 2 projects to observe whether there are any links to DLI projects, demos or testbeds, or if there are any mentions of DLI projects. All 28 libraries at these institutions have well developed digital libraries incorporating a large array of resources and services; they are among the leading practical digital libraries. Three of these 28 library sites (11%) have links to DLI projects (UC Santa Barbara, Johns Hopkins, Tufts). Four libraries (14%) (Berkeley, Harvard, Cornell, and Texas) have a list of digital library projects, but none of them are DLI projects — these are projects found in the ARL database or are other projects. Secondly, we

examined another sample of 30 libraries at large institutions, such as Rutgers, Princeton, Yale, New York Public Library, with strong digital libraries of their own, but could not find any connections or mention of DLI projects.

In sum, the connections between DLI research and practical libraries range from very minimal to none. But here is a note of caution: While we found only three connections and nothing more, this does not mean that there are no other connections; it means that we just could not find them.

5.3 Collections

In every issue, D-Lib Magazine presents a "Featured Collection" showcasing digital collections and libraries in different domains. It is an eclectic feature, demonstrating a wide variety of applications and great imagination. Collectively, they demonstrate a startling diversity. Some of these are imaginative Web sites, others are true collections, and some are full-fledged digital libraries. The domains are diverse: *nuclear physics, ragtime, geology, molecule of the month, plant kingdom, aquarium, Chaucer, mammalian brain, T. rex Sue, bugscope, castles, classical music, vaudeville, math, entomology* – you name it. But the feature also includes digital libraries of long standing, as Internet Public Library, the American Memory Historical Collection, and MEDLINEplus. We examined all 33 Featured Collections for years 1999 (since the inception of this feature), 2000, and 2001. None of them has any connection with DLI projects, being more or less independently conceived, developed and constructed. This is another sample of operational digital libraries or collections that are not connected to digital library research, or at least having no visible connection that we could discern.

5.4 Digital libraries in professional societies

Numerous organizations outside of hybrid libraries have operating digital libraries related to collections in their own domain, represented by ACM Portal, incorporating the ACM Digital Library, and IEEE Xplore, incorporating the IEEE/IEE Electronic Library.

Both provide access to their publications, including conference proceedings. We chose these two societal digital libraries because, among other materials, they include many articles about digital libraries that appeared over the years, and because the publications of these societies were the prime outlets for reporting from DLIs. A

search in IEEE Xplore for "digital libraries" retrieved 403 documents, and in the ACM Digital Library 1,057 documents. Rous (2001) described the background and design principles for the ACM Digital Library, while Durmiak (2000) described IEEE Xplore. Combined with the information in these articles and extensive examination of and searching in the respective sites, we concluded that their design and operations mirror a number of other operational digital libraries. As far as we can see, they have no visible connection to work, demos or testbeds in DLI projects. It is surprising that although members of these societies dominate DLI research, the efforts of these societies regarding their own digital libraries are independent of DLI advances.

5.5 Commercial products

We did not intend to review this area, but could not help noticing while exploring the issues raised here that there is another world very much involved with digital libraries. Many commercial vendors of library systems have moved toward developing and offering a variety of packages that deal with supporting development and maintenance of digital libraries. So, too, have non-profit service organizations such as OCLC and CrossRef. Web-oriented companies are also entering the market. Their customers are libraries making an evolutionary transition from library automation to digital libraries. The customers are not only traditional libraries, but also other organizations, such as societies..

These vendors offer, among others, digital management systems for libraries; integrated library systems that now include digital library components; access, search and delivery systems; digital content conversion services; license and rights management systems; security systems; navigation, discovery and interoperability systems; interfaces; and digital reference services in a variety of packages directly related to digital libraries. For instance, the company Ex Libris "a worldwide supplier of software solutions and related services for libraries and information centers," is offering MetaLib, a standardized interface and portal, incorporating SFX, an interoperability system, for hybrid libraries and information systems.

Many libraries and other organizations buy or license these packages for building and managing their own versions of digital libraries. Vendor evaluation and selection (rather than research evaluation) has become a major

activity in libraries. For example, Condit & Calloway (2001) provide an evaluation of reference technology offered by commercial vendors.

These products and services are discussed in magazines such as Information Today, Computers in Libraries, Information Technology and Libraries, and others. They are presented in exhibits at professional-society meetings, including gatherings of the American Library Association and Special Library Association; and at trade shows such as InfoToday, which has a separate section on E-Libraries. Web sites such as Vanderbilt's Library Technology Guides provide information on a variety of aspects related to technology; this particular site lists 46 active "library automation companies" in the U.S.; not all into digital libraries, but most heading there.

Are digital libraries heading toward commercialization? As they say in the social sciences: "This is beyond the scope of the paper." But there clearly is a strong connection between operational digital libraries and commercial developments and offerings.

6. Literature

We classified the topics of papers in the literature into four classes:

1. *Issues*: Papers that are mostly devoted to broad issues, commentary, analysis, context, and meta-discussions. For instance, these include papers on what are digital libraries, the role or value of digital libraries, underlying assumptions, digital libraries for developing countries, and the like.

2. *Technology*: Papers that primarily concentrate on techniques and supporting technology that are domain independent, even though examples may involve given domains.

3. *Projects*: Papers that are primarily devoted either to reports on developments related to a specific, single project in a domain, or to general descriptions of various initiatives, activities, or plans that involve numerous projects. Accordingly, we subdivide project articles into *specific* and *general*.

4. *Research*: Papers that report on scholarly activities in the traditional sense. Includes papers on research results related to a problem or topic, rather than reports on development related to a specific project or technology. Such articles contain theory, models, hypotheses, experiments, observations, and/or data.

The literature on digital libraries is extensive. The recent comprehensive review of digital libraries by Fox & Urs (2002) has an extensive bibliography of nearly 500 items. Under the term "digital library or libraries," the Library and Information Science Abstracts database lists over 1,600 documents and INSPEC (a science and technology database) lists over 2,400 documents. We already noted the number of digital library documents in ACM and IEEE. Of this literature, we concentrated on some specific publications, and not on the literature as a whole.

6.1 Special issues

We looked at four journals: Communications of the ACM (CACM), IEEE Computer, Journal of the American Society for Information Science (JASIS), and Information Processing & Management (IP&M). They differ as to their orientation. The first two are general review and advances-reporting journals; while they do not publish research articles per se, they often, among others, report on research undertaken and on project accomplishments. The last two are research-oriented journals.

Communications of the ACM (CACM) has three special issues devoted to digital libraries (vol. 38 (4) 1995; vol. 41 (4) 1998, and vol. 44 (5) 2001). The 1995 issue has 23 articles — four are longer and issue-oriented, six are devoted to technology, 12 are short descriptions of projects, of which 11 report on specific projects, including the six DLI 1 projects, and one is a research article. The 1998 issue has 19 articles - four are on issues, six on technology and nine on projects, of which eight (88%) are on specific projects. The 2001 issue has 21 articles - six on issues, six on technology, and 10 on projects , of which seven (70 %) are on specific projects. The orientation in the issues is toward description, rather than research. The progression is evident. The first special issue was mostly of future projects, while subsequent ones report on a number of more mature projects. The 1995 issue prominently dealt with DLI 1. Interestingly, subsequent issues do not mention DLI 1 at all. One possible exception is an article in the 1998 issue from a team in the DLI 1 project at Stanford, but even that one did not mention DLI but had to be surmised from authorship and topic. One DLI 2 project (Tufts) is represented in the 2001 issue, with acknowledgment to DLI 2. Otherwise, DLI projects are not incorporated. Of the 26 specific projects in all issues,

only one of the projects (Library of Congress) is listed in the ARL database, thus the projects in <u>CACM</u> and the projects in ARL represent different universes of projects. But they also include a number of projects from other countries. This demonstrates an international spread of digital libraries beyond the U.S., an issue not covered in this study.

<u>IEEE Computer</u> has two special issues in digital libraries (vol. 29 (5) 1996 and vol. 32 (2) 1999). The 1996 issue has seven articles, one is a general introduction, and the rest describe the six DLI 1 projects as presented in proposals and releases. The 1999 issue has six articles - one on issues, and five on technology. Of the five, two are from DLI 1 projects (Illinois, Stanford), one is related to DLI 2 (Carnegie Mellon), and two report on other projects (JSTOR, New Zealand); none are related to ARL projects. The majority of articles in these special issues are devoted to DLI projects and descriptions.

<u>Journal of the American Society for Information Science</u> (JASIS) has two special issues on digital libraries (vol. 44 (8), 1993 and vol. 51 (3 & 4) 2000). The 1993 issue contains six articles: three are on issues, and three on projects - since this was pre-DLI and pre-ARL, none of the projects is connected with DLI or library projects. The two-part 2000 issue contains 16 articles: two are on issues (introductory statements), one on technology, three on projects (both related to projects in the ARL database), and 10 on research. Of the research articles, three are from DLI 1 projects (Santa Barbara, Berkeley, Illinois).

<u>Information Processing & Management</u> has one special issue on digital libraries (vol. 35 (3), 1999). Of the 11 articles, two are on issues, two on technology, and seven on research. One of the research articles reports the results of a DLI 1 project (Illinois).

In sum, we can see two patterns in these special issues. <u>CACM</u> and <u>IEEE Computer</u> are oriented toward reporting of projects and technology in a more general way, while <u>JASIS</u> and <u>IP&M</u> are more research-oriented. These describe two different orientations of reported works on digital libraries. The authors in the first two journals are mostly from computer science departments, while in the last two in addition to authors from computer science departments that are in the majority, authors from a few other departments and agencies are represented as well.

Three DLI 1 projects contributed to the research literature. In this set of articles, the presence of projects in the ARL database and other operational projects is minimal.

6.2 Conference proceedings

We concentrated here on the <u>Proceedings of ACM Conference on Digital Libraries</u> for three years: 1999, 2001, and 2002 when it became a <u>Joint Conference on Digital Libraries</u> (with IEEE Computer Society). Starting in 1996, this annual event developed into the premier conference on digital libraries in the U.S., particularly from the computer science perspective.

The 1999 <u>Proceedings</u> have 23 main papers (we did not consider in depth the panels and poster papers). Of the 23, 11 (48%) are on research (although including for the most part evaluation of a project), nine (39%)are on projects, and three (13%) on technology. In the acknowledgments, six (23%) explicitly acknowledge DLI support. Of the 57 authors, as best as we can determine only seven (12%) came out of institutions associated with other than computer science.

The 2000 <u>Proceedings</u> have 23 papers. As to topic, 10 (43%) of papers are on research, nine (39%) on projects, and four (17%) on technology. Of the 69 authors, as best as we can determine 10 (14%) come from outside of computer science institutions. Only one paper had a direct acknowledgement to DLI.

The 2001 proceedings have a different format, both long and short (2-page papers) are integrated. We considered only the 39 long papers: 24 (62%) are on projects and 15 (38%) on research. 13 (33%) have acknowledgement of support to various DLI programs. Of the 120 authors, as best as we can determine, 24 (20%) come from outside of computer science.

None of the 85 papers in these three <u>Proceedings</u> is about any practical digital library project, as listed by ARL, or any operational digital library, as listed in Libweb. We also undertook a cursory examination of panels, posters and short papers in these <u>Proceedings</u> and similarly, could not find any connection to practical projects.

In sum, papers at these conferences represent an impressive diversity of efforts in digital libraries, with the proportion of project descriptions increasing and research decreasing in 2001. Also, a growing proportion of papers

is coming out of DLI projects, about one-fifth of papers has a DLI acknowledgement. While the proportion of authors outside computer science is rising, only 16% of all authors over these years comes from outside. These conferences mainly represent efforts coming out of the computer science community, and provide a minimal connection to efforts involving broader communities. But as the projects and research in digital libraries began involving specific domains, where subject expertise is a critical component, we see a broadening of participation, as in the 2001 Proceedings.

Papers related to practical projects and operational digital libraries are presented at (and integrated within) a variety of disciplinary and trade conferences, thus, comparisons cannot be made easily, and we did not attempt them.

6.3 D-Lib Magazine

From its start in 1995, D-Lib Magazine evolved into a primary vehicle for reporting on many facets of digital libraries. We analyzed 153 papers that appeared in the main sections variously titled "Articles," "Stories," and "Project Briefings" from January 1999 to January 2002. We did not consider other materials in the Magazine - and there are plenty of these. As to the topics of the 153 papers, 42 (27%) are on issues, 37 (24%) on technology, 65 (42%) on projects (of these, 16 are on projects on a general level and 49 on specific projects), and nine (6%) on research.

Of the 49 papers that reported on specific projects, 36 (73%) are from the U.S.; the rest (27%) are projects in the European Union, the UK, Germany, Netherlands, Australia, and Canada. Some of the projects are on digital collections, others report on services, processes (such as reference), or technology. Of the 36 U.S. specific projects, 18 (50%) are on operating digital libraries either in a domain, or describing a digital library or service in an institution. A few university libraries are represented. Nine of 36 (25%) are related to DLI. We could not find descriptions of the ARL listed projects, or operational digital libraries that we culled from Libweb.

Of the nine research articles, none mentions or deals directly with the DLI projects. The articles belong to several categories, but most of them (six out of nine) are in the category of "assessment and evaluation" of the

economics of digital libraries, usage statistics and evaluation of use, tools, projects and services. The three remaining articles include: Study of education for digital libraries, survey paper on social informatics, and study of end-user search patterns. The authors of three of the papers are LIS faculty; two are authored by researchers in government institutions; the rest of the papers are by authors from a computer science department, a consultant, and other organizations.

The 153 papers contain a total of 59 acknowledgment statements (38%). We looked at acknowledged support by granting agencies. In 11 acknowledgments, DLI support is mentioned, however, five additional projects acknowledge NSF support, which in all probability is through the DLI program. Support from other agencies, from government, to industry, to foundations and private individuals or groups, is also acknowledged, again showing the widespread interest in digital library work.

Among others, we looked at authors. All together, 393 authors were associated with these 153 papers; 68 papers (44%) had single authors, 33 (21%) had two authors and the rest or 52 papers (33%) had three or more authors. Put another way, out of 393 authors, 325 (83%) were involved in collaboration to produce papers, and presumably underlying work. Thus, digital libraries present a highly collaborative activity, which comes as no surprise.

As to the country of origin, 304 (77%) are affiliated with various agencies in the U.S., 36 (9%) come from the UK, and the rest from13 other countries. The Magazine is international, but with a decidedly U.S. flavor. A further analysis of e-mail domains of the 304 U.S. authors reveals that 208 (68%) have an .edu domain, and thus have affiliation with educational institutions, 34 (11%) have a .com affiliation, 27 (9%) are with government (.gov), 20 (7%) are with an organization (.org), and the rest are with state, military and network domains. While educational affiliation of authors predominates, there are still many authors involved with other organizations and commerce, showing a spread of activities in digital libraries outside academe.

Of the 85 papers with multiple authors, 49 (58%) are by authors who work in the same institution (not necessarily the same department), and the rest (36 or 42%) involved authors from different institutions. While production of most of the collaborative papers (and presumably underlying work) is bound by a single institution,

there is a surprisingly large number of papers (presumably work as well) based on cross-institutional cooperation. While a number of these institutions involve operating libraries, most still involve computer centers or computer science departments, and research institutes. It is hard to do such fine grain analysis, because the data is not readily available.

D-Lib Magazine papers provide a look at the rich panoply that represents people, institutions, and agencies involved in digital libraries. But, it is not representative of operational digital libraries, and practical projects (as listed in ARL).

7. Conclusions

The study examined projects in digital library research and digital library practice in the U.S., with the aim of determining whether they inform each other, and whether there is a connection. We consulted information provided on the Web sites of a large number of digital library projects reporting research or practice, and a representative set of literature on the topic. In other words, we looked at what is visible and on the surface. The approach has obvious limitations - we took the information provided "as is;" and we did not pursue any deeper analysis of connections, if any, below the surface. We acknowledge, as enumerated in the section Methodology, significant limitations to the method. Thus, we also acknowledge that conclusions should be taken with that caveat in mind. In all of this, we do not criticize or evaluate either research or practice in general, or any undertaking or project in particular. We did not look at accomplishments, but only at possible visible connections.

A brief answer is this: We believe that presently, digital library research and digital library practice are conducted by and large mostly independent of each other, minimally informing each other, and having slight, or no connection. But, since they are still in progress and the diffusion process is a function of time, we may expect changes. The agenda for digital library research, as reflected by Digital Library Initiatives, is set from the top down, although with some consultation with some, mostly computer science, constituencies. In that respect, we concur with David Levy's conclusion, quoted in the introduction, that the research agenda largely bears the imprint of the computer science community's interests and vision. The agenda for digital library practice is set from the

bottom up, by the institutions and organizations involved, and bears the imprint of institutional interests, priorities, visions, and missions. Considerable resources and efforts are spent in each. In many instances, digital library research projects are conducted at the same institutions that have sizable digital library practical projects, but they have no visible connection.

However, the dynamics of the situation are more complex than this brief answer suggests. In DLI 2, as opposed to DLI 1, a majority of projects is domain-oriented. A number of these have produced or are in the process of establishing demos, testbeds, or practical digital libraries in their domains. While most of the PIs and project staff in DLI 2 are still associated with computer science departments, the proportion of PIs and project staff from other departments and fields has risen. The rise in research oriented toward specific domains corresponds with a rise in the potential realization of a visible connection between digital library research and digital library practice.

Referring to DLI 1, the report on digital libraries by the President's Information Technology Advisory Committee (2001) states, "Many of today's digital library accomplishments can be directly traced to early Digital Libraries Initiative (DLI) funding." Our aim was not to investigate the issue of "accomplishment", but a question can be raised about the inferred "direct tracing." If we consider practical digital libraries, we could not find such a trace or connection. It seems to us, that the development of the vast majority of practical digital libraries proceeded independently of any connections to DLI 1.

We did not discuss the $$$$ factor, but it cannot be ignored. Millions of dollars are involved in both digital library research and digital library practice; the economic aspects are critical to both. Digital library research was driven by availability of massive funding. Digital library practice is flourishing because of massive direction of funds to development and operations. More often than not, contemporary choices in research topics and in technology transfer hinge on economics. In other words, economic factors and interests may be the deciding factor in possible connection, or lack thereof, between research and practice.

Why this divide between research and practice? A panel at the 2001 Digital Library Conference (Levy et al., 2001) discussed, among others, topics related to the necessity and role of traditional versus digital libraries, the role of paper and related issues. They noted a polarization of viewpoints on many issues. We provide a preliminary

interpretation about these issues and associated polarization, with a suggestion that they should be examined further. As all activities, digital library research and digital library practice proceed from a number of assumptions and premises. Among others, these deal with the use and role of technology in digital libraries, the items and content to be handled, the role of the human element, and the overall context. Here are some of the questions for the premises:

- **Technology:** What can/cannot and should/should not be automated? What should be emphasized?

- **Objects:** What objects and collections are to be treated? Created? How should they be handled? What about their persistence? What is the role of paper? Of its continuing existence? Connection to digital libraries?

- **People:** What is the significance of human element? The role and extent of human intervention, human intelligence and interpretation? The place of people in relation to technology?

- **Context:** What institutional context may be appropriate? What roles do the social and cultural contexts play? Politics? Economics? Legal structures? What role exists for traditional libraries?

The premises, resulting from answers to each of these questions, can be placed on a continuum. It seems to us that the premises for digital library research are on one end, and for digital library practice are on the other end of the continuum – they are polarized. A technological perspective heavily influences the research end of the spectrum, while the practice end is influenced by an institutional and user/use perspective. Consequently, they formulate the premises about these issues quite differently. We suggest that the polarization on these issues may explain the divide and the relatively low extent of contact; further, we suggest that as long as this state of divide in premises continues, there is little likelihood on real and productive contact and interaction. Again, this conclusion should be taken with caveats expressed.

In an idealized paradigm, research is supposed to inform practice by suggesting innovation. But diffusion of technology and innovation is not a straight line. It is not predictable. Transfer can come from a number of directions, some wholly unexpected, and from interactions that are not easily observable, as argued in the next

section. It can take a short or long time. It can be filtered indirectly through numerous, often-invisible channels. In general, sociotechnical change depends on a number of factors such as infrastructure and technology, social and economic aspects, and contagion effects. As Bijker (1994) has shown the acceptance of a technology (closure and stabilization) are determined by the acceptance by a relevant social group of a working artifact – invention and social relevance converge. In the case of digital library technology, we see a complexity in which the institutional environment, the marketplace, the knowledge-producing communities (represented by the research communities involved in DLI 1 & 2 research), and the practice communities in traditional library settings, are involved in a somewhat chaotic manner. It seems that each community is addressing a different aspect of digital libraries – one more technical and other more institutional. It is also possible, that because of different premises, they are building quite different digital libraries. However, as mentioned, we observed only the situation on the surface, as it exists at present, and we are not predicting anything about possible connection in the future. It is a challenge for all stakeholders.

History of information retrieval (IR) and of information science in general provides an example of the convoluted path between research and innovation (Salton, 1987, Hahn & Buckland, 1998). Research on advanced IR was begun in a laboratory setting by Gerard Salton and colleagues in the early 1960s. Funded by NSF and other agencies, IR research flourished in the following decades. But large commercial search vendors and services, such as Dialog and LexisNexis, ignored research results and proceeded with development of their own IR technology. Only in the 1990s did they and other vendors incorporate some of the advanced IR research results into their own search technology, basically because of market/user pressures and dissatisfaction, and growing competition from alternate search engines. Today, many, but not all, commercial web search engines use IR research from that era as the basis for further development; in the process they have developed advanced but proprietary IR. However, we believe that the parallel for digital libraries may not hold and may not be valid to start with. Then, (up to late 1990's) only a very few large vendors (less than 10 in the world) were the only practical, large, and universally used IR applications; they had the market regardless of research-based innovation, and accordingly did not care. Now, a great number and variety of practical digital libraries are on their way. They may and do provide,

independently of research, for their own developmental advances. The very quantities of IR systems then and digital libraries now are not comparable.

In sum, as it stands now, we believe that digital library research on the one hand, and digital library practice on the other, reside in parallel universes with little visible contact and intersection, as demonstrated by the diffusion channels examined here. We think that, while they are both about digital libraries, there is a digital divide between them. At present, the two communities disseminate ideas in detached formal networks of communication that are more or less self-referential. But things and connections may change. The few connections that have been established are now an exception rather than a rule. Perhaps they are a sign of things to come.

8. Differing interpretations

Our research addresses a number of sensitive issues. We fully acknowledge limitations of the sources and methods used, and therefore we issued caveats. Accordingly, we prefaced all our conclusions with "we believe that…" rather than "we show that… ." On the one hand, we know that the issues and questions raised here are important for the future of both, digital library research and digital library practice. But on the other hand, we only believe in and stand by our derivations and surmises. We accept that objections could and should be raised. There are differing "I believe that …" interpretations. Also, different questions, methods and conclusions are possible. To stimulate thinking, investigation, and discussion of the questions on relations, here are some issues and interpretations raised by the reviewers:

"[T]he argument for this [conclusions] as presented seems not convincing. It seems quite plausible to argue instead other conclusions from the data presented, e.g. that researchers in the DL field have not adequately publicized their findings, that practitioners have not been sufficiently scholarly in looking for solutions, that commercial DL products do not adequately give credit to the underlying research etc."

"It is well known that technology transfer may take place through informal as well as formal channels and records. This indeed has been common in the DL field, a fact that should not be ignored in this paper. For example, many people developing commercial DL products have attended DL conferences and learned of research work.

Likewise a good percentage of those attending DL conferences, including particularly conferences outside USA …

are practitioners, who bring back to their libraries and projects what they have learned from research presentations.

Further, invited talks, panel discussions, short papers, posters, and workshops are key parts of conferences where

technology transfer takes place in both directions, and these have been ignored in the analysis."

"There are notable examples of DL research going into practice. For example, Google was launched by

students working on the Stanford DLI-1 project even as that project was still underway, and clearly has had

tremendous influence on practice! Further, Google today has very strong ties to the DL research community."

"Using info about ARL is not necessary predictive of the connection between research and practice. One

alternate explanation is that the DL field as a whole attempts to bring together a number of communities to work

together on integrative problems. Thus, there are computer scientists, librarians, policy makers, content publishers,

etc. It is well known that people in the DL field often have allegiances to their "home" field and secondary

allegiances to DL. Then, measures of cohesion between fields of primary allegiance are not likely to change if they

are based on counts of projects and publications in the home field. In other words, ARL topics are likely to continue

as ARL topics, and this fact does not necessarily show that people connected with ARL don't also apply DL

research when doing DL project activities. … is it not feasible that since D-Lib Magazine covers matters so well,

that ARL feels it does not need to replicate that in its own list?"

"Citations is used as a key indication that there is not much connection between research and practice. One

problem with this measure is that practitioners rarely cite, and often read handbooks or digested version of research,

making it unlikely that they would cite the original research. Further, there are many other measures of connection

between research and practice that in a young field often are more useful. For example, how many researchers

consult with people engaged in practice? How many startups have a connection with the research effort (recall

Google)? How many practitioners attend research overview tutorials? What about perhaps 2/3 of the 600+

attendees of ICADL'2001 who were practitioners and heard about research and practical efforts? What about the

hundred of teams that wrote proposals for DLI, DLI-2, NSDL, ITR, and other initiatives related to DL that were

not funded but spent a great deal of time studying DL research."

"The use of principal investigators for projects is problematic in the analysis of projects and publications. … The fact that DLI projects were strongly dominated by CS departments is not disputable, but the interplay between the CS departments and the various partners (in other academic units as well as government, business, and industry) must be addressed as this is where some of the technology transfer takes place. In fact, the national policies of technology transfer and practical applications in DLs are not addressed at all in this paper and this seems an important facet given the inherent tensions among basic research funded by NSF/NIH/DARPA and the special nature of the DL initiatives. What are the analogs to PIs for practical libraries…? Where do they get their ideas? What roles do CNRI, CNI, Educause, and other NGOs play? Along these lines, it would be useful to look at companies spun off from the DLI-1 projects. How did the early DL research influence IBM, Sun, and others to create DL 'solutions' (now called content management)?"

"Looking at software that practical DLs use might also provide a trace of connection between research and practice. Looking at Cheshire over a decade evolving from an OPAC to a DL platform might be one example. Practical DLs might not directly adopt Informedia video algorithms, but the video skim concept is finding its way into practical DLs. What about open-source solutions?"

"Another omission is in the area of federated projects. … Two examples are NCSTRL (www.ncstrl.org) and NDLTD (www.ndltd.org), each of which has over 100 sites using research systems that are periodically enhanced with new technology, and where the community of practitioners learns about new methods. Also through these efforts and others like them, there are clearly over 200 universities or documentation centers that have been exposed to DL research even though they might not realize it."

"Another omission is the area of standards. Much research has led to standards that are widely used in practice. In the DL field clear examples are OAI and DC. Should not these connections be measured in the analysis?"

"The multiple decade technology transfer process for IR research (requiring a major shift in underlying technology infrastructure from centralized hosts to client-server WWW architectures) would tell us that it is much too soon to expect much connection between DL research and practice yet. The conclusion that such transfer may

not take place in DLs due to the complexity of the application seems both highly premature and in need of much more explanation."

While we did not make a conclusion that such transfer may not take place, we believe that so far, after a decade or so of digital library research and digital library practice there is little evidence that they are indeed informing each other and that significant transfer has taken place. We believe that, as yet, they happily live separately – each does its own thing. We also believe that this is not good for progress in either.

Acknowledgements

<div align="center">

References

</div>

Association for Research Libraries. (2001). ARL libraries spend nearly $100 million on electronic resources. <u>ARL Bimonthly Report</u> 219.Retrieved 28 Jan 2002 from <u>http://www.arl.org/newsltr/219/eresources.html</u>

Bijker, W. A. (1994) <u>Of bicycles, bakelites, and bulbs: Toward a theory of sociotechnical change.</u> Cambridge, MA: MIT Press.

Condit, J. F. & Calloway, M. (2001). Creating an instant messaging reference system. <u>Information Technology and Libraries, 20</u>, (4), 202-212.

Durmiak, A (2000). Welcome to IEEE Xplore. <u>IEEE Power Engineering Review, 20</u> (11), 12.

Fox, E.A. & Urs, S.R. (2002). Digital Libraries. In Annual Review of Information Science and Technology, 36, 503-589.

Greenstein, D., Thorin, S., & Mckinney, D. (2001). Draft report of a meeting held on 10 April in Washington DC to discuss preliminary results of a survey issued by the DLF to its members. Digital Library Federation. Retrieved 30 Jan. 2002 from http://www.diglib.org/roles/prelim.htm#results

Hahn, T. B. (1996). Pioneers of the online age. Information Processing & Management, 32, (1), 33-48.

Hahn, T. B. & Buckland, M. (Eds.) (1998). Historical studies in information science. Medford, NJ: Information Today.

Harum-S; & Twidale-M. (Eds.) (2000) Successes and failures of digital libraries. 1998 Annual-Clinic-on-Library-Applications-of-Data-Processing. Graduate School of Library and Information Science, Illinois University at Urbana-Champaign.

Lesk, M. (1999). Perspectives on DLI-2 - Growing the field. D-Lib Magazine, 5 (7/8). Retrieved 30 Jan. 2002 from http://www.dlib.org/dlib/july99/07lesk.html

Levy, D.A. (2000). Digital libraries and the problem of purpose. D-Lib Magazine, 6 (1). Retrieved 15 Jan. 2002 from http://www.dlib.org/dlib/january00/01levy.html. Also appears in: Bulletin of the American Society for Information Science, Aug./Sept 2000, 22-26.

Levy, D. et al. (2001). Panel: High tech or high touch: Automation and human mediation in libraries. Proceedings of the First Joint ACM/IEEE-CS Joint Conference on Digital Libraries. p. 345.

President's Information Technology Advisory Committee (PITAC) (2001). Digital libraries: Universal access to human knowledge. Retrieved 10 Feb. 2002 from http://www.itrd.gov/pubs/pitac/pitac-dl-9feb01.pdf

Price, D. J. deS. (1963). Little science, big science. New York: Columbia University Press.

Rogers, E. (1995). Diffusion of innovation. (4th ed.). Free Press.

Rous, B. (2001). The ACM Digital Library. Communications of the ACM, 44, (5), 90-91.

Salton, G. (1987). A historical note: The past 30 years in information retrieval. Journal of the American Society for Information Science, 38, (5), 375-380.

Schatz, B. & Chen, H. (1999). Digital libraries: Technological advances and social impact. IEEE Computer, 32, (2) 45-50.

Schwartz, C. (2000). Digital libraries: an overview The Journal of Academic Librarianship, 26 (6), 385-393.

Zia, L. L. (2001). The NSF National Science, Technology, Engineering, and Mathematics Education Digital Library (NSDL) Program: New Projects and a Progress Report. D-Lib Magazine, 7 (11).

Uniform Resources Locators (URLs) for sites mentioned in the paper

American Library Association http://www.ala.org/

Association for Computing Machinery. ACM Digital Library http://portal.acm.org/

Association for Computing Machinery. ACM Special Interest Groups (SIGs)

http://www.acm.org/sigs/guide98.html

Association of Research Libraries (ARL) http://www.arl.org/

Association of Research Libraries. Digital Initiatives Database http://www.arl.org/did/

California Digital Library (CDL) http://www.cdlib.org/

Carnegie Mellon University. Informedia Digital Video Library http://www.informedia.cs.cmu.edu/

Corporation for National Research Initiatives (CNRI) http://www.cnri.reston.va.us/

Computers in Libraries http://www.infotoday.com/cilmag/ciltop.htm

CrossRef http://www.crossref.org/

Dialog http://www.dialog.com/

Digital Library Federation (DLF) http://www.diglib.org/

Digital Library Federation. Public Access Collections http://www.hti.umich.edu/cgi/b/bib/bib-idx?c=dlfcoll

Digital Library Initiatives http://www.dli2.nsf.gov/

Digital Library Initiative 1 (DLI 1). http://www.dli2.nsf.gov/dlione/

Digital Library Initiative 2 (DLI 2). http://www.dli2.nsf.gov/projects.html

D-Lib Magazine http://www.dlib.org/

D-Lib Test Suite http://www.dlib.org/test-suite/index.html

IEEE/IEE Electronic Library http://ieeexplore.ieee.org/lpdocs/epic03/

Ex Libris http://www.exlibris-usa.com/

Information Technology and Libraries http://www.lita.org/ital/index.htm

Information Today http://www.infotoday.com/it/itnew.htm

InfoToday http://www.infotoday.com/

LexisNexis http://www.lexisnexis.com/

Library of Congress. American Memory Project http://www.loc.gov/mem

National Science Digital Library http://nsdl.org/

National Institutes for Health (NIH) http://www.nih.gov/

National Science Foundation (NSF) http://www.nsf.gov/

OCLC http://www.oclc.org/home/

OhioLINK http://www.ohiolink.edu/

Perseus Digital Library http://www.perseus.tufts.edu/

Special Library Association http://www.sla.org/

Stanford U. Building the InfoBus: Interoperation mechanisms among heterogeneous services http://www-diglib.stanford.edu/diglib/index.html/

Stanford U. Technical accomplishments http://dbpubs.stanford.edu:8090/pub/2000-50

Text REtrieval Conference (TREC). http://trec.nist.gov/

U of California at Berkeley. Sunsite Digital Library http://sunsite.berkeley.edu/

U of California at Berkeley. Sunsite Digital Library. Libweb http://sunsite.berkeley.edu/Libweb/

U of California at Berkeley. Environmental Planning and Geographic Information Systems

http://elib.cs.berkeley.edu/

U of California at Santa Barbara. Alexandria Digital Library (ADL) Project http://www.alexandria.ucsb.edu/

U of Illinois at Urbana-Champaign. Federating repositories of scientific literature

http://dli.grainger.uiuc.edu/default_old.htm

U of Illinois at Urbana-Champaign. UIUC Digital Library Testbed http://dli.grainger.uiuc.edu/idli/idli.htm

U of Michigan. Intelligent agents for information location http://www.si.umich.edu/UMDL/

Vanderbilt's Library Technology Guides http://staffweb.library.vanderbilt.edu/breeding/ltg.html