

NSF/DARPA/NASA DIGITAL LIBRARIES INITIATIVE

A PROGRAM MANAGER'S PERSPECTIVE*

Stephen M. Griffin

INTRODUCTION

The Digital Libraries Initiative (DLI) was the result of a community-based process which began in the late 1980s with informal discussions between researchers and agency program managers. These discussions progressed to planning workshops designed to develop research values and agendas and culminated in the National Science Foundation (NSF)/Defense Advanced Research Projects Agency (DARPA)/National Aeronautics and Space Administration (NASA) Research in Digital Libraries Initiative announced in late 1993. With the selection and funding of the six DLI projects [http://www.cise.nsf.gov/iis/dli_home.html], interest and activities related to digital libraries accelerated rapidly. The six DLI projects became highly visible and influential efforts and grew in scope, participation, and influence. NSF and DARPA funded additional workshops as part of the DLI to develop consensus on specific digital libraries topical areas and boundaries, to bring together researchers to stimulate cross-disciplinary interaction, and to ponder together how best to adapt to a rapidly changing global information environment. By now, researchers and practitioners from many disciplines have been drawn into digital libraries research and related activities, from subject domains reaching far beyond the sciences into the arts and humanities.

Based on the recognized achievements of DLI and the promise of additional Federal investment in digital libraries, a follow-up program was announced in the spring of 1998. In the new program, "Digital Libraries Initiative—Phase 2," NSF, DARPA, and NASA are joined by the National

*This paper first appeared in *D-Lib Magazine* (July/August 1998), ISSN 1082-9873, <<http://www.dlib.org>>.

Library of Medicine, the Library of Congress, and the National Endowment for the Humanities as primary sponsors [<http://www.nsf.gov/pubs/1998/nsf9863/nsf9863.htm>]. First round awards are expected to be made beginning in September 1998.

Digital libraries are meant to provide intellectual access to distributed stores of information by creating information environments which advance access beyond electronic access to raw data—i.e., the bits—to the fuller knowledge and meaning contained in digital collections. Electronic access is increasing at a rapid pace through global efforts to increase network connectivity and bandwidth, new information management tools, and importantly, interoperability across systems and information content. The quantity of online digital information is increasing tenfold each year in a relatively uncontrolled open environment. This pace of information creation far surpasses that of the development of technologies to use the material effectively. The number of people accessing digital collections through the WWW also shows explosive rates of growth. Finally, internationalization is making a “global information environment” a reality. The World Wide Web (WWW) offers a bounty of certain kinds of information to those willing to struggle through the repetitive searching and sifting often required. Digital libraries research is essential to enabling more people to better create and use vast amounts of distributed information and to contribute to the quality and quantity available via the Web and future access frameworks. But it is often not sufficiently appreciated that it is the content that motivates most people to use the Internet and digital libraries. Many Americans and others around the globe are increasingly turning to Internet-based repositories as the primary source of information about many subjects. People of all ages and backgrounds, it turns out, love to browse, explore, and accumulate new knowledge—in short, to learn.

Ultimately, it is the demand for high quality content and ease of access and use that will drive the funding and development of digital libraries. And expectations are high. Users routinely issue 10^1 size queries into 10^{14+} size data spaces. They look to get the information requested—all of it—but only that which is relevant—and it should take just a few seconds.

Efficient information retrieval—identifying all relevant sources quickly—is one aspect of digital libraries research. Another, and potentially even more valuable, aspect of digital libraries, is their ability to preserve and extend discourse—to provide richer contexts for people to interact with information. The real value of digital libraries may prove to be in their ability to alter the way individuals, groups, organizations, etc., behave, communicate, and conduct their affairs. New forms of collaboration in scholarly and other endeavors are appearing regularly. In this role, digital libraries are powerful instruments of change in social and work practices.

Programmatically, digital libraries remain closely linked to advances in high performance computing and networking and both contribute to and validate these technologies. The merging of advanced computing and communications technologies with massive volumes of digital content will dramatically alter knowledge generation and contexts of use.

DLI AND THE FEDERAL CONTEXT: HPCC

The Digital Libraries Initiative emerged programmatically within the structure of the Federal High Performance Computing and Communications Program (HPCC) [<http://www.hpcc.gov/>]. HPCC was introduced in a supplementary report to the President's FY1992 Budget and consisted of coordinated efforts in four general focus areas that were executed, for the most part, within established programs in the eight participating agencies. The HPCC Program was the product of several years of planning and discussion within the Federal Coordinating Council for Science, Engineering and Technology (FCCSET). The 1992 report was entitled "Grand Challenges: High Performance Computing and Communications." Grand challenges were driving applications for developing teraflop computing systems and a national high bandwidth network for research and education. The bulk of the program involved funding of high performance computing systems, advanced software technologies and algorithms, and networking infrastructure.

In 1994, the HPCC Program was expanded to include a fifth program component, Information Infrastructure Technology and Applications (IITA). IITA was intended to provide for the research and development needed to develop an underlying technology base for the National Information Infrastructure (NII) and to address National Challenges. National Challenges were seen to be those applications, benefited by HPCC technologies and resources, that could have broad and direct impact on the Nation's competitiveness and the well-being of its citizens. Included in the list of National Challenges were "digital and electronic libraries."

HPCC has evolved into the Federal Computing, Information, and Communications (CIC) programs. The Executive Summary to the FY 1998 Supplement acknowledges the breadth of influence of the Internet and technologies culture of change: "There is little historical precedent for the swift and dramatic growth of the Internet, which, just a few short years ago, was a limited scientific communications network developed by the Government to facilitate cooperation among Federal researchers and the university research community."

The Digital Libraries Initiative is featured prominently in the FY 1998 document as a CIC R&D Highlight, testimony to both its achievements and to the mounting importance of the area generally. This is compelling

given that the DLI represents only about 0.6 percent or \$6M of the CIC Program's \$1100M budget.

A distinctive feature of the continuing Federal computing, information, and communications planning dialogue is a relatively narrow focus on developing technologies for scientific applications and education. Strong reliance is placed on traditional institutional forms, mature disciplinary research communities, and quantitative methodological approaches to problem solving.

DLI PROGRAMMATIC CONTEXT

In the early 1990s, NSF, DARPA and NASA were individually supporting basic research in computing and communications and viewed digital libraries as a broad, newly-emerging topical area of great potential. Informal working groups of agency managers were formed and met regularly over a period of time to define programmatic goals and discuss alternative research agendas. These were, then, the topics of technical workshops funded by the agencies to reconcile with community values and expectations. The reports emanating from the workshops provided the intellectual content of the first program announcement—Research in Digital Libraries—which was released in the fall of 1993. The Digital Libraries Initiative was designed as a basic research initiative to advance the means to collect, store, organize, and access information in digital form via communication networks. Projects were expected to perform high-risk research and to test and demonstrate new technologies.

The program was broadly cast. It was quickly realized, once the proposals were received and reviewed, that DLI needed additional direction and coherence. The 1995 IITA Digital Libraries Workshop entitled "Interoperability, Scaling, and the Digital Library Research Agenda" refined the scope and added coherence to the DLI research agenda. The report coming out of the workshop defined digital libraries as: "An organized collection of multimedia data with information management methods that represent the data as information and knowledge" [<http://www.ccic.gov/pubs/iita-dlw/>].

The discussions continued as the program evolved. The directions for digital libraries research and benefits of deployment were actively debated within and across technical, library, and other communities. Tensions, some still unresolved, have inhibited interaction and exchange between various communities. An important point of issue is that many see advances in digital libraries research dependent on efforts in domains other than computer and information sciences. The phrase "digital libraries" has been adopted widely over electronic libraries, virtual libraries, and others with the understanding that "electronic" refers primarily to

the nature of the technologies which operate on information and “virtual” implies a synthetic environment which resembles the original physical environment. This is as it should be. “Digital” refers to a representation of information on electronic (and other) media. Digital representation adds great potential for enhanced functionality and utility of information corpora. Once information has been digitally encoded, tools and systems can be invented to create altogether new ways to extract meaning from the collection.

Definitions continue to change as researchers and users stretch our thinking about them. The NSF sponsored Santa Fe planning workshop on distributed knowledge work environments [<http://www.si.umich.edu/SantaFe/>] held in March 1997, broadened the definition of a digital library as follows: “[T]he concept of a ‘digital library’ is not merely equivalent to a digitized collection with information management tools. It is rather an environment to bring together collections, services, and people in support of the full life cycle of creation, dissemination, use, and preservation of data, information, and knowledge.”

The Santa Fe workshop set the intellectual directions and content for Digital Libraries Initiative—Phase 2 (DLI-2). DLI-2 will address a narrower technology research agenda than DLI. Progress to date has suggested areas of greater importance and will support research across the information lifecycle including content creation, access, use and usability, and preservation and archiving. DLI-2 will place emphasis on interoperability and technology integration, content and collection development and management, applications and operational infrastructure, and understanding digital libraries in domain-specific, economic, social, and international contexts—i.e., digital libraries as human-centered systems. The program will go beyond computing and communications specialty communities and proposes to engage scholars, practitioners, and learners with many ambitions including not only science and engineering but also the arts and humanities. By doing so, DLI-2 acknowledges that significant advances in technology will result from the perspectives, methods, and applications of non-science domains—that important new research questions for computer and information sciences will be raised and perhaps answered in venues other than academic computer science research laboratories.

While DLI-2 is part of the Human Centered Systems (HuCS) component of the Federal CIC Programs, its projects are expected to involve content in subject areas across the continuum of human interest. The topical boundaries of DLI-2 activities will be set according to the availability and the character of the sources of program and project investment. DLI-2 also recognizes that collection building and knowledge access are inherently international and will actively promote activities and processes that bridge political and language boundaries. It is hoped that these work-

ing groups will provide valuable advice for stimulating international efforts [http://www.si.umich.edu/UMDL/EU_Grant/home.htm].

The new NSF Knowledge and Distributed Intelligence Initiative (KDI) coincides with, and is closely related to DLI-2 [<http://www.ehr.nsf.gov/kdi/default.htm>]. KDI acknowledges the commonality of approaches in research and development emerging across scientific and engineering disciplines as a result of the deployment of new information technologies and infrastructure. A symposium on KDI was held at the National Academy of Sciences in September 1997 and attended by policy makers and executives from public and private foundations. In calling for the broadest possible dialogue and input, Neal Lane, then director of NSF, stated: "The access we have gained to widely distributed sources of information marks a major accomplishment for human civilization. . . . It is, nevertheless, only a first step. Access to information is one thing. But intelligently absorbing, refining, and analyzing this information to glean useful knowledge is another altogether."

DLI-2 differs from KDI programmatically in that DLI-2 is focused on users and collections—DLI-2 projects are expected to point to future use and usability. Information and processes for delivering information are emphasized across the entire digital libraries life cycle. KDI, particularly the "knowledge networking component," is targeted at fundamental interdisciplinary research about knowledge and knowledge access. While KDI is an NSF-only interdisciplinary program, executed within existing program structures, DLI-2 includes multiple agencies, some which go far beyond interest in science and engineering such as the Library of Congress and the National Endowment for the Humanities. DLI-2 is neutral with respect to subject matter.

DLI PROGRAM CONSTRAINTS

Digital libraries research has not yet gained Federal support commensurate with evident levels of community interest and activity. The supply of research funding falls far short of the demand. Individual agencies continue to be constrained by mission and one year budget cycles. While longer term activities and programs of support have been established, many of these undergo scrutiny on a regular basis—at least every four years with the elections. Digital libraries projects can extend in scope well beyond agency missions and demand support beyond a single agency's means. Larger-scale projects require several years to complete and require a stable and predictable funding stream to retain essential staff and resources.

Research agencies tend to be limited in supporting those research activities and infrastructure building that stay within their defined missions. Maintaining existing disciplinary programs often is favored over beginning true interdisciplinary ones. While encouraging collaboration on

the part of performers, it is difficult for sponsors to do the same—to build multiple sponsorship arrangements for single large projects.

Digital libraries projects are typically multi-modal—a mix of research, application, and development of operational systems. It is a fact that some of the most important research issues are bound into the process of building operational systems and analyzing the use and performance of these systems. It is also important that the “libraries” (which may be testbeds or experimental systems) contain collections of value. It is these final two stages of digital libraries research (i.e., what might be considered prototypical operational systems containing content of significant value but still subject to research based on broad-based use) for which it is difficult to find funding from agencies like NSF and DARPA. To achieve a limited expansion of scope, each of the six original DLI projects formed partnerships with various organizations—public (other federal, state, and local governments) and private (major technology vendors, publishers, libraries, schools, etc.). Taken together, over eighty major partnerships were formed which provided the projects with substantially more resources, testing environments, and, importantly, fresh perspectives on their activities. Cost-sharing of more than 100 percent on average was generated. In DLI-2, the challenges are greater still, and agency managers hope that certain selected projects will have sufficient appeal to attract additional funding from other Federal and non-Federal sources.

CONCLUSION

The 1990s are a critical decade when information technology intersects with, and becomes drawn into, endeavors cutting across domain-specific research, education, and commercial and social practices. As a result, many people and organizations are crossing into unfamiliar territory with unpredictable consequences. These are times of enormous opportunity in which decisions made in the present augur prominently in shaping the future. Digital libraries as global multilingual repositories of data, knowledge, sound, and images invite people everywhere to become users and learners.

Digital libraries are inherently international. Knowledge is recorded and stored in many forms, often using different languages and symbol systems. That which exists in one language or located in one country may be only a small part of a corpus of interest. Fuller access to information across language, location, and cultures means fuller understanding of a particular topic and the relationships among topics. Researchers and users must have opportunities to work together if they are to see globally distributed, interoperable, content rich systems. Yet, while scientists and information professionals around the world are engaged in digital libraries research and development, as of now there is little coordination or col-

laboration because of lack of implementation mechanisms. As part of the Digital Libraries Initiative, a modest step was taken to establish five international working groups to help build DL research agendas for technical, content, social, and economic issues. This effort is jointly funded by the National Science Foundation and the European Union. It is hoped that these working groups will provide valuable advice for international efforts.

There remains an unnatural separation between the producers and consumers of digital libraries resources. A proper balance of attention (and support) among research, applications, content, and collections has yet to be achieved. Localizing research efforts in computer and information sciences venues is limiting, and many believe that efforts in libraries, museums, art departments, schools of music, archeology, history, and other humanities departments are necessary to advance digital libraries research. Yet the science agencies, like NSF, DARPA, and NASA, can only make awards to performers in non-science venues with difficulty—and frequently accompanied by protest from the disciplinary communities normally receiving support from specific agency programs.

DLI benefitted from “bottom-up” program development. It was conceived and planned by program managers at the agencies relying heavily upon community input—not as part of a grander programmatic scheme influenced by transient political value. As such, the monies invested were from the base budgets of the programs involved. (About fifteen separate programs from NSF, DARPA, and NASA contributed funds.) Program managers believed strongly in the values and goals of the initiative—and acted with considerable independence in implementing and executing the program.

By adopting a participatory consensus-based management approach, one that was open, adaptive, and responsive to a larger community, the program was able to be particularly effective in exploiting aspects of the global information infrastructure revolution that was underway. In many ways, the management culture reflected the positive aspects of the open culture of the Internet which the program was attempting to enrich.

The interagency management group for DLI-2, presently composed of managers from the sponsoring agencies, meets regularly to discuss current developments and consider the future. The group adheres to the spirit of DLI management and hopes for broad representation and community involvement and consideration of a large analytical framework to help shape future directions.