

Proceedings of the 9th Annual Federal Depository Library Conference

October 22 - 25, 2000

The Digital Libraries Initiative: A USA Federal Program of Research and Applications

Stephen M. Griffin, National Science Foundation
Arlington, VA

Digital Libraries Initiative (DLI)

<http://www.dli2.nsf.gov>

Phase 1: Program Profile

- Sponsored by NSF, DARPA, NASA
- 1994 - 1998
- Six university-led projects; similar project model for each
- \$24M total over five years, ending fall 1998.
- A program of fundamental digital libraries research, testbed building and partnerships

Project/Research Focus

Carnegie Mellon University: Digital Video Libraries

- speech, image and natural language technologies integration

University of Michigan: Intelligent Agent Architectures

- software agents; resource federation; artificial service market economies; educational impact

Stanford University: Uniform Access

- interoperability; protocols & standards; distributed object architectures; interface design for distributed information retrieval

University of California, Santa Barbara: Geographic Information Systems

- spatially-indexed data; content-based retrieval; image-compression; metadata

University of Illinois: Intelligent Search and the Net

- large-scale information retrieval across knowledge domains; semantic search; SGML; user/usage studies

University of California, Berkeley: Media Integration and Access

- new models of "documents"; natural language processing; content-based image retrieval; innovative interface design

DLI Phase 1 Collaboration and Partnering

DLI Lead Institutions:

Carnegie Mellon
University of California, Berkeley
University of Illinois
Stanford University
University of Michigan
University of California, Santa Barbara

Flow of Resources, Technologies, Knowledge, Intellectual Products

Computer & Communications Companies

Digital Equipment Corp.

Xerox Corp.
Xerox PARC
Intel Corp.
Apple Corporation
Bellcore
Eastman Kodak Co.
IBM
Lockheed
Interconnect Tech Corp.
Enterprise Integration (EIT)
Bellcore
Interval
Microsoft Corp.
Bell Atlantic Network Services
AT&T
Hewlett Packard
United Technologies

Softquad
BRS/Dataware
Spyglass
Hitachi

Publishers/Content Providers

Elsevier Science Group
Encyclopedia Britannica
McGraw-Hill Publishers
Dialog Information Services
O'Reilly
WAIS, Inc.
QED Communications
John Wiley & Sons
U.S. News & World Report
M&T Publishing
Tribune Company
UMI

Professional Societies

American Math Society (AMA)
ACM
IEEE
American Institute of Aeronautics and Astronautics (AIAA)
American Physical Society
American Institute of Physics
NCGIA
Association of Research Libraries

Other Universities

SUNY Buffalo
University of Maine
University of Arizona
Open University, U.K.
University of Wisconsin
University of Colorado
MIT
Cornell University

Libraries

Project Site University Libraries
USGS Library
Library of Congress
California State Library
Sonoma County Library
St. Louis Public Library
New York Public Libraries

International Organizations

ERCIM

Primary & Secondary Schools

Project-local community schools
Fairfax County Public Schools
Winchester-Thurston School
Ann Arbor Public Schools
Stuyvesant High School, NYC
Shasta County Office of Education

Government Agencies and Labs

DMA/CIO
US Navy
USGS
NASA/ARC
Research Agency of California
San Diego Association of Governments

Other/Non-Profits

CNRI
Environmental Systems Research Institute
Mellon Foundation
Kellogg Foundation
Getty Foundation

Digital Libraries Initiative - Phase 2

- Core Sponsors: NSF, DARPA, NLM, LoC, NASA, NEH
~\$8-10 million/yr for 5 years (beginning FY98)
- sponsor a full-spectrum of activities: fundamental research, content & collections development, domain applications, testbeds, operational environments, new resources for education and preserving America's cultural heritage
- address topics over entire DL lifecycle: information creation, dissemination, access, use, preservation, impact, contexts
- implement a modular, open program structure: add new sponsors, performers, projects at any time

Program Goals:

New DL research, technologies and applications to advance the use of distributed, networked information of all types around the nation and the world

DLI Phase 2 Collaboration and Partnering

DLI2 Academic Institutions

Flow of Resources, Technologies, Knowledge, Intellectual Products

Computer & Communications Companies

Digital Equipment Corp.

Xerox Corp.

Xerox PARC

Intel Corp.

Apple Corporation

IBM

SRI International

Oracle

GE

Interval

Microsoft Corp.

Bell Atlantic Network Services

AT&T

Lucent Technologies

Hewlett Packard

Informix

Sharp

NEC

Hitachi

Sun Microsystems

Healthwise

Welch Allyn

Government Agencies and Labs

Smithsonian Institution

US Navy

Los Alamos National Laboratory

National Park Service

California Academy of Sciences

CA Env. Res. Eval. Sys. (CERES)

CA Dept. of Water Resources

San Diego Supercomputer Center

USGS

NASA/ARC

Resources Agency of California

S. California Earthquake Center

Consortium of Research Libraries-UK

UK Office for Library & Information Networking

Libraries/Museums

Library of Congress

California Digital Library
New York Public Library
NASA Ames Library
USGS Library
Museum Fine Arts, Boston

Professional Societies

Modern Language Association
ACM
Oral History Association
NCGIA
Association of Research Libraries
Chicago Historical Society

Other/Non-Profits

Mellon Foundation
Parkard Humanities Institute
Getty Foundation
Columbia Presbyterian Medical Center

International Organizations

EU/ERCIM
JISC
DFG

Content Providers

CNN
The News Hour with Jim Lehrer
Dialog Information Services

Academic Projects Partners

University Arizona
University of Bath
University of Bristol
University of California at Berkeley
University of California at Davis
University of California at Los Angeles
University of California at Santa Barbara
Carnegie Mellon
Columbia University
Cornell University
Eckerd College
Georgia State University
Harvard University
University of Illinois at Chicago
Indiana University
John Hopkins University
University of Kentucky
King's College, London
University of Leeds

University of Liverpool
University of Maryland
University of Massachusetts
University of Michigan
Michigan State University
University of North Carolina
Old Dominion University
Oregon Health Sciences University
Oregon Graduate Institute of Science and Technology
University of Pennsylvania
University of Texas at Austin
University of South Carolina
Southampton University
Stanford University
Swarthmore College
Tufts University
University of Washington
University of Wisconsin at Madison

Comparison of DLI with DLI - Phase 2 DLI - Phase 1 (1994-1998)

Research: broad, technology-centered
Testbeds: for technology research
Content/collections: donated to projects
Infrastructure: limited testbed development
Context: primarily user evaluation

DLI - Phase 2 (1998-2002)

Research: refined technical scope; extend to new areas and dimensions in the DL information lifecycle

Testbeds: for DL research with added emphasis on interoperability & technology integration

Content/collections: increased emphasis on content, collections development and management

Infrastructure: operational DLs with collections of value to domain and other "communities" of users

Context: understanding DLs in domain, economic, social, international contexts

The Federal High Performance Computing and Communications Program, 1992-1996

- Early focus on speed and bandwidth
- Two dimensional thinking of early 1990s

- Three dimensional thinking of mid-1990s

Next: Advanced functional capabilities, wide use

- Digital libraries must present vastly different content at the use level yet maintain striking similarities at the digital level. To do this requires interdisciplinary research at all stages of the content lifecycle and layers of networking infrastructure.

Add context and structure to digital content in early stages of preparation

- adding metadata to digital content early makes a digital library much more useful and inexpensive than trying to create more intelligent software to compensate for it later

Challenges for Digital Libraries

- use the Internet to enhance creation, access, and usability of globally distributed content-of-value
- build information technologies to acquire new knowledge and understanding from the world's stores of information
- maintain the substance, form, and function of information objects from source through network to user (skeuomorph)

A Major Issue for Sponsors

What proportion of resources should go to:

1. Efforts to make software intelligent?
2. Efforts to make content intelligible?

Worldwide Production of Original Content (Estimates, 1999)

Storage Medium	TB/Year	TB/Year	Growth rate, Percent
	Upper Estimate	Lower Estimate	
Paper	240	23	2
Film	427,216	58,216	4
Optical	83	31	70
Magnetic	1,693,000	635,660	55

TOTAL	2,120,539	693,930	50
--------------	------------------	----------------	-----------

Source: School of Information Management and Systems, University of California, Berkeley

<http://www.sims.berkeley.edu/how-much-info/>

- Information technology is pushed by research and applications in other disciplines.
- Computer Science is stressed by and enlivened by engagement in new topical problem areas.
- Interdisciplinarity beyond the sciences has much to offer.

Changing Scales and Contexts of Interaction and Collaboration

- International collaborative efforts are essential to achieving a content-rich, balanced Global Information Infrastructure. Issues must be addressed through collaborations at many levels.

Making Global Digital Libraries Infrastructure Means:

- Merging intellectual perspectives
- Dealing with heterogeneity at many levels
- Achieving interoperability at many levels
- Integrating information technologies
- Building large collections of great diversity
- Supporting functions beyond search and query
- New conceptualizations of the future (imagination)
- Global participation
- Economic and IP models for new information use

Building Large Scale Operational Systems

- Our understanding of the impacts of digital libraries on social institutions and practices is limited because we do not yet have large-scale systems being heavily used to observe and analyze.
- The reflexive behaviors of systems, supporting infrastructures and user populations become apparent when millions of people use digital libraries, not thousands.

Making Digital Libraries Infrastructure Requires Dealing with Heterogeneity at Many Levels:

- Objects, collections, services, platforms

Making Digital Libraries Infrastructure Requires Merging Intellectual Perspectives

Traditional Libraries Stress:

- Service
- Selection, Organization, Structure for Access
- Centralization, Standards
- Physical objects & standard genres

Contemporary Technological Capabilities (e.g., WWW) Stress:

- Flexibility, Openness
- Rapid Evolution
- Decentralization (geographic, administrative)
- Digital objects, old and new genres

Making Digital Libraries Infrastructure Requires

- Application of Integrated Technologies
- Building Large Collections of Diverse Information
- Supporting More than Query
- New Conceptualizations of the Future (imagination)

Digital Libraries Initiative Project Highlights

- Basic Representations of Music & Audio
- Blobworld Update
- Open Archives Metadata Set
- Alexandria Digital Library
- Informedia-II: Integrated Video Information Extraction and Synthesis
- Example of a Large Data Object: Michelangelo's David
- The Digital Atheneum
- Cervantes Project

Goals for the Future

- Gather information and build collections (to better use what we have and discover what is missing...)
- Create new global communities (to communicate and collaborate)
- Make technology disappear (from our awareness and experience)

The definition of "digital library" continues to evolve

Internet accessible digital objects (representing text, data, documents, images, sound, video, agents, databases, middleware...) with sufficient identity, structure and contextual information to allow creating coherent collections on demand to service the needs of diverse user communities (query, analysis, communication, collaboration, ...)

For More Information:

- Digital Libraries Initiative Home Page:
<http://www.dli2.nsf.gov/>