

ACM Multimedia '96

Course MP 2

TUTORIAL TITLE: Building and Applying Digital Libraries II

INSTRUCTORS: Edward A. Fox, Robert M. Akscyn

SCHEDULE: Monday, Nov. 18, 1996, 2 p.m. - 5:30 p.m.

PARTICIPANTS: Level: Intermediate

Participants are encouraged from those working in fields like computing, library science, information science, multimedia, and publishing. Those with some background on digital libraries are invited (e.g., having taken the morning session, or having read the literature).

COURSE DESCRIPTION:

Objectives:

Attendees should be able to help in design, development, evaluation, and standardization efforts related to digital libraries. They should understand research and development issues, including principles and guidelines for design of scalable, sustainable DLs.

Planned Activities:

There will be coverage of the following SOURCES:

- Digital Library conferences;
- IITA meetings (e.g., May 1995 workshop);
- Allerton Institutes (from U. Illinois, NSF);
- D-Lib (research, magazine, working groups);
- D-Lib research articles (architecture, metadata, URNs, use);
- Virginia Tech information (DL page, Sourcebook);
- Virginia Tech projects (Envision, ILDLCS, WWW traffic analysis/mod/sim);
- Z39.50 (overview, OCLC, CNIDR);
- Library of Congress;
- CNRI (architecture, handles);
- UMBC agents, preservation, TEI, ...

The last hour will involve group activities for specifying requirements and developing designs for the National Digital Library of Theses and Dissertations, or other effort(s).

OUTLINE:

SELECTED PAGES FROM WWW ---

SOURCES:

* Conferences

- ACM SIGIR DL Page http://info.sigir.acm.org/sigir/digital_lib.html
- DL'97: Philadelphia, July 23-26 (before ACM SIGIR'97)
- DL'96: Bethesda, March (1st ACM ...)
- DL'95: Austin, June [link] <http://csdl.tamu.edu/DL95/>
- earlier in DL'9x and ADL'9x series from 1994-1996
- IITA Digital Libraries Workshop, 1995 [link] <http://www-diglib.stanford.edu/diglib/pub/reports/iita-dlw/main.html>
- Allerton, 1995 [link] <http://edfu.lis.uiuc.edu/allerton/95/>

* Journals

- J. of Visual Communication and Image Representation, 1996
- IEEE Computer, May 1996 - Communications of the ACM, Apr. 1995
- J. American Society for Information Science, Sept. 1993
- (new) British Computer Society J. of Digital Information

* WWW

- D-Lib <http://www.dlib.org/>
 - + Research Projects (incl. DLI) ([link]) <http://WWW.dlib.org/projects.html>
- D-Lib Articles ([link]) <http://www.dlib.org/title-index.html> ([link])
- D-Lib Working Groups [link] <http://www.dlib.org/groups.html>
 - + Metadata [link] <http://www.dlib.org/metadata/overview.html>
 - + Naming [link] <http://www.dlib.org/naming/overview.html>
 - + Repository Interfaces [link] <http://www.dlib.org/repository/overview.html>
 - + Social Aspects [link] <http://www.dlib.org/social/overview.html>
- D-Lib Magazine
 - + Agents [link] <http://www.dlib.org/dlib/July95/07birmingham.html>
 - + Architecture (incl. handles) [link] <http://www.cnri.reston.va.us/home/dlib/July95/07arms.html>
 - + Metadata [link] <http://www.dlib.org/dlib/July95/07weibel.html>
 - + Uniform Resource Names (URNs) [link] <http://www.dlib.org/dlib/february96/02arms.html>
 - + Use [link] <http://www.dlib.org/dlib/october95/10bishop.html>
- Virginia Tech ptrs [link] <http://scholar.lib.vt.edu/digilib/>
- Digital Library Source Book, 1993 [link] <http://fox.cs.vt.edu/DLSB.html>

RESEARCH:

* Virginia Tech

- Envision (CS literature) [link] <http://ei.cs.vt.edu/~cs5604/Adv/Adv-Envision.html>
- Envision report [link] <http://ei.cs.vt.edu/papers/ENVreport/final.html>
- Interactive Learning with a Digital Library in CS [link]
<http://ei.cs.vt.edu/~cs5604/Adv/Adv-ILDLCs.html>
- WWW Traffic Analysis, Modeling and Simulation [link]
<http://www.cs.vt.edu/~chitra/www.html>
 - + Sources of variation: daily, semester, local/remote
 - + Excel graphs (18 months) (see <http://ei.cs.vt.edu/~fox/EDMEDIA96>)
 - Accesses [link]
 - Bytes [link]

* OCLC

- SiteSearch [link] <http://www.oclc.org/oclc/menu/site.htm>

* architectures, components, protocols

- Z39.50
 - + ptrs [link] <http://ds.internic.net/z3950/z3950.html>
 - + Isite [link] <http://vinca.cnidr.org/software/Isite/Isite.html>
 - + architecture [link] http://vinca.cnidr.org/resources/isite_overall.gif

* National

- Library of Congress
- American Memory [link] <http://lcweb2.loc.gov/ammem/ammemhome.html>
- Call [link] <http://lcweb2.loc.gov/ammem/award/>

* CNRI

- key issues [link] <http://WWW.CNRI.Reston.VA.US/home/cstr/arch/slides.html>
- repositories of digital objects [link] <http://www.cnri.reston.va.us/home/doa.html>

* naming

- PURLs [link] <http://purl.oclc.org/>
- handles [link] <http://www.cnri.reston.va.us/home/cstr/handle-intro.html>

* agents

- UMBC [link] <http://www.cs.umbc.edu/agents/>
- Michigan [see prior link]

* library and information science

- intellectual property rights
 - + CMU NetBill [see prior link]
 - + preservation [link] <http://www-rlg.stanford.edu/ArchTF/>
- representations
 - + Text Encoding Initiative [link] <http://etext.virginia.edu/TEI.html>
- principles [link] <http://ei.cs.vt.edu/~cs5604/DL/DL7.html>

PERSPECTIVES:

- * user and social needs
 - Allerton [see prior link]
 - D-Lib WG [see prior link]
- * interfaces and user interaction
 - monitoring, analyzing, visualizing and modeling traffic (see above)
- * architectures, components, protocols
 - Z39.50 (see above)
 - CNRI (see above)
 - agents (see above)
 - bus: Stanford [see previous links]
 - systems, engines, and operations
- * naming (see above)

RECOMMENDATIONS:

- * Join projects: electronic theses and dissertations, CS technical reports
- * Use our courseware and add to it
- * Work toward “Open Digital Library”

ACKNOWLEDGEMENTS

Sponsors and Partners: Elsevier, IBM, NSF, OCLC, SURA

Co-PIs, Students, Staff

Projects

- * NSF IRI-9116991, CDA 9312611, EID-9109853, CDA-9308259
- * NCSTRL (ARPA, NSF)



Digital Libraries '95

The Second Annual Conference on the Theory and Practice of Digital Libraries

June 11-13, 1995 - Austin, Texas, USA

Getting a physical copy of the DL 95 Proceedings

Sponsors and cooperating institutions

From the Conference Chair, David M. Levy

From the Program Chair, Richard Furuta

Conference Committee

Attendee List

Full Papers

Delivering Technology for Digital Libraries: Experiences as Vendors,
William T. Crocca and William L. Anderson

InterPay: Managing Multiple Payment Mechanisms in Digital Libraries,
Steve B. Cousins, Steven P. Ketchpel, Andreas Paepcke, Hector Garcia-Molina, Scott W. Hassan, and Martin Roescheisen

Providing Government Information on the Internet: Experiences with THOMAS,
W. Bruce Croft, Robert Cook, and Dean Wilder

A New Zealand Digital Library for Computer Science Research,
Ian H. Witten, Sally Jo Cunningham, Mahendra Vallabh, and Timothy C. Bell

Cataloging in the Digital Order,
David M. Levy

Collection Maintenance in the Digital Library,
Mark S. Ackerman and Roy T. Fielding

Interoperability, Scaling, and the Digital Libraries Research Agenda:

A Report on the May 18-19, 1995

IITA Digital Libraries Workshop

August 22, 1995

Clifford Lynch (clifford.lynch@ucop.edu)

Hector Garcia-Molina (hector@db.stanford.edu)

Converted to HTML using GradStudentWare 2.2

Contact [Christian Mogensen](#) with bug reports.

Introduction

Definitions and Roles of Digital Libraries

Defining Interoperability in the Digital Library Environment

Infrastructure Requirements for Digital Library Research

Research Issues and Priorities

1. Interoperability

2. Description of Objects and Repositories

3. Collection Management and Organization

4. User Interfaces and Human-Computer Interaction

Conclusions

Executive Summary

Appendix 1 - List of Participants

Appendix 2 - Strawman Report

Appendix 3 - Report of the working groups

3-1 - The Publishing Perspective

3-2 - The Commercial Perspective

3-3 - The Library Perspective

3-4 - The Internet Perspective

3-5 - The Multimedia Perspective

Introduction

This report summarizes the results of a workshop on Digital Libraries held under the auspices of the U.S. Government's Information Infrastructure Technology and Applications (IITA) Working Group in Reston, Virginia on May 18-19, 1995. The objective of the workshop was to refine the research agenda for digital libraries with specific emphasis on issues of scaling and interoperability, and to identify the infrastructure developments needed to make progress on these issues.

While there have been a number of workshops and other meetings examining the broader questions of support for applications in the National Information Infrastructure (NII), we believe this was the first workshop that focused specifically on Digital Libraries in this context. In the past year, Digital Libraries have emerged as one of the central and most compelling applications enabled by the NII; numerous digital library research projects are underway, including six large-scale pilot projects that have been funded jointly by ARPA, NASA, and NSF. While Digital Libraries are now a vibrant research area, and also a field in which considerable commercial development is taking place (presaging the future economic importance of Digital Library technology to the United States), many new questions are emerging as a result of this flowering of research

37th Allerton Institute 1995

Graduate School of Library and Information Science
University of Illinois at Urbana-Champaign

How We Do User-Centered Design and Evaluation of Digital Libraries: A Methodological Forum

This conference was sponsored by the [National Science Foundation](#)



Introduction, Ann P. Bishop

Session 1 - Migrating Foundational Study Approaches to the Virtual Environment

Special Presentation: Findings from Digital Library Studies

 Annelise Mark Petjersen, *Designing for Retrieval in Library Collections: Lessons from Book House*

 Michael Twidale, *How to Study and Design for Collaborative Browsing in the Digital Library*

Session 2 - Co-Design in Digital Libraries

Session 3 - Work Practice and Institutional Change

Session 4 - Electronic Information Seeking and Use

Special Presentation: Social and Organization Issues in Classification (notes only) - S. Leigh Star and Geof Bowker

Session 5 - Users, Diversity, and Change

Session 6 - Wrap-up

List of Participants

The 1995 Allerton site is available via the [EDFU Electronic Library](#)

The Publications Office

Graduate School of Library and Information Science

University of Illinois, Urbana-Champaign

Last updated: 17 January 1996

Digital Library Research

From here, you can follow pointers to some of the major cooperative projects, funding and coordinating agencies, and associated activities in digital library research. Many of the research centers maintain lists of projects and project descriptions as well as small collections of technical papers, typically authored by staff.

- [Coordinating and funding bodies](#)
- [U.S. federally funded cooperative projects](#)
- [Centers for research on digital libraries in the U.S.A.](#)
- [Programs and projects outside the U.S.A.](#)

For collections of technical papers and a selection of individual items, see:

- [**Technical Reports and Papers**](#)

Coordinating and funding bodies

[NASA's Digital Library Technology Project](#). A project that supports the development of new technologies to facilitate public access to NASA data via computer networks.

[The Coalition for Networked Information](#). A joint project of the Association of Research Libraries, CAUSE, and EDUCOM to promote information resources in networked environments.

[The Internet Engineering Task Force](#). The protocol engineering and development arm of the Internet.

[The World Wide Web Consortium](#). The W3 Consortium exists to develop common standards for the evolution of the World Wide Web.

[European Research Consortium for Informatics and Mathematics](#). A consortium of leading research establishments in Europe that encourages collaborative work among researchers and with industry.

European Commission

- [Science, Research, and Development: The 4th Framework Programme](#)
- [I*M-EUROPE](#)
- [General Information](#)

[Research Libraries Group \(RLG\): Strategy for 2000, Supporting Projects](#). A not-for-profit membership organization for improving access to information to support research and education. Among other services and activities is a research program that facilitates collaboration among institutions in relevant areas of research (e.g., library services, information access and delivery).

Federally funded cooperative projects

[The NSF/DARPA/NASA Digital Library Initiative \(DLI\)](#). Six federally funded projects in digital library research, with partnerships led by universities. The individual projects are listed below.

University of California, Berkeley: An Electronic Environmental Library Project. (A DLI project.)

University of California, Santa Barbara: The Alexandria Project: Towards a Distributed Digital Library with Comprehensive Services for Images and Spatially Referenced Information. (A DLI project.)

Carnegie Mellon University: Informedia: Integrated Speech, Image and Language Understanding for Creation and Exploration of Digital Video Libraries. (A DLI project.)

University of Illinois at Urbana- Champaign: Building the Interspace: Digital Library Infrastructure for a University Engineering Community. (A DLI project.)

University of Michigan: The University of Michigan Digital Library Project. (A DLI project.)

Stanford University: Stanford University Digital Libraries Project. (A DLI project.)

The Computer Science Technical Reports Project (CSTR). A collaboration involving CNRI, five universities, and the Library of Congress.

D-Lib. A forum for researchers and developers of advanced digital libraries.

Centers for research on digital libraries in the U.S.A.

- Center for Intelligent Information Retrieval
- Center for the Study of Digital Libraries
- Center for Electronic Texts in the Humanities
- Center for Research on Information Access
- The Electronic Text Center at the University of Virginia
- The Information Infrastructure Project
- OCLC Online Computer Library Center, Inc.: Office of Research and Special Projects
- Rutgers Center for Information Management, Connectivity, and Integration

Programs and projects outside the U.S.A.

eLib: Electronic Libraries Program (UK): A broad program of projects addressing a wide range of digital library issues.

European Research Consortium for Informatics and Mathematics (ERCIM): A consortium of research organizations from thirteen European countries, which provides a framework for collaboration.

Nordinfo: Coordinating framework for collaboration among three centers devoted to digital libraries and electronic publishing: National Digital Library Centre (NDLC), Nordic Net Centre (NNC), and NorDEP.

Distributed Systems Technology Centre (DSTC): A joint venture by the Australian Government's Cooperative Research Centres (CRC) Program and participating organizations to develop the technological infrastructure for global distributed systems.

New Zealand Digital Library: An interactive system with collections of computer science technical reports and literary materials. The site includes pointers to the relevant project descriptions, technology, and related information.

Digital Library Network (DLNet): This page describes digital library work at the University of Library and Information Science, Tsukuba Science City, Japan, in English. From here, follow links to discussions of

D-lib magazine

The Magazine of Digital Library Research

ISSN 1082-9873

Access to Stories, Briefings, and Editorials by Title

[A](#) [B](#) [C](#) [D](#) [E](#) [F](#) [G](#) [H](#) [I](#) [J](#) [K](#) [L](#) [M](#) [N](#) [O](#) [P](#) [Q](#) [R](#) [S](#) [T](#) [U](#) [V](#) [W](#) [X](#) [Y](#) [Z](#)

You may also browse by *author*.

1990 Census LOOKUP: Mining a Mountain of Data, Deane W. Merrill, Nathan G. Parker, Harvard H. Holmes, Chris Stuber, Valerie J. Gregg

A

Access and Discovery: Issues and Choices in Designing DIFWICS, Jeremy Hylton

Access to Digital Objects: A Communications Law Strategy, Patrice A. Lyons

Accessing the Visible Human Project, Michael J. Ackerman

Advanced Web Presentation through Data Modeling: An Open Architecture for the Personalized Webs of the Future, Leon Shklar

Agent-based Architecture for Digital Libraries, William P. Birmingham

Alexandria Digital Library Testbed, James Frew, Michael Freeston, Randy Kemp, Jason Simpson, Terence Smith, Alex Wells, and Qi Zheng

B

Berkeley Digital Library SunSITE, Roy Tennant

Briefings:

- ARTFL, Andrea Doane
- Brief Update on the Alexandria Digital Library Project: Constructing a Digital Library for Geographically-Referenced Materials, Terence R. Smith
- CyberStacks(sm): A 'Library-Organized' Virtual Science and Technology Reference Collection, Gerry McKiernan
- Digital Libraries in the Classroom, Elliott Soloway
- EduPort, Miriam Masullo
- Highlights related to the Government Information Locator Service (GILS): Toward a Global Information Locator, Eliot Christian
- Integrated Document Access (IDA) Project, Margaret Colmer
- Making a Digital Library: The Chemistry Online Retrieval Experiment -- A Summary of the CORE Project (1991-1995), Richard Entlich, Lorrin Garson, Michael Lesk, Lorraine Normore, Jan Olsen, Stuart Weibel
- OCLC Internet Cataloging Project, Erik Jul
- Oregon State University's Government Information Sharing Project, Jacquelyn Miller
- Project Muse: 43 Humanities and Social Sciences Journals to Come on the Network

- SuperJournal, David Pullinger, Christine Baldwin
- TURNIP: The URN Interoperability Project, Renato Iannella
- UK Electronic Libraries Programme, Chris Rusbridge
- Z39.50 and the World Wide Web, Sebastian Hammer, John Favaro

C

Caching for Large Scale Systems, Lessons from the WWW, Robert E. McGrath
Coming Soon to Your Favorite Library: Decision Support on Demand, Hemant Bhargava, Bob Norris
Content Ratings and Other Third-Party Value- Added Information: Defining an Enabling Platform, Martin Röscheisen, Terry Winograd, Andreas Paepcke
Creating a Networked Computer Science Technical Report Library, James R. Davis

D

Digital Libraries and Corporate Technology Reuse, Jonathan T. Hujsak
Digital Libraries: Searching Is Not Enough; What We Learned On-Site, Andreas Paepcke

E

Economic Framework for Pricing and Charging in Digital Libraries, J. Sairamesh, C. Nikolaou, D. Ferguson, Y. Yemini

Editorials:

- All Things in Good Time
- Double-Edged Sword of Access
- Future is a Complex Place
- Getting Used to Technology and Revolutions in the Making
- In This Issue [July/August 1996]
- Levels of Abstraction
- Taking the Measure of the Net
- Telling Time
- Text Is More Than Just Words on a Page, Susan Hockey
- Tragedy of the Commons, Revisited (Again)
- When is Honesty the Best Policy?
- Where Are We in Space?
- Word (or two) of welcome

F

Federating Repositories of Scientific Literature: An Update on the Digital Library Initiative at the University of Illinois at Urbana-Champaign, Susan L. Harum, William H. Mischo, and Bruce R. Schatz
French Minitel: Is There Digital Life Outside of the "US ASCII" Internet? A Challenge or Convergence? Jack Kessler

G

Global Change Data and Information System-Assisted Search for Knowledge (GC-ASK) Project, Roberta Y. Rand

H

Historical Collections for the National Digital Library: Lessons and Challenges at the Library of Congress, (Part I), Caroline R. Arms
Historical Collections for the National Digital Library: Lessons and Challenges at the Library of Congress (Part

II), Caroline R. Arms

I

Image Browsing in the Alexandria Digital Library (ADL) Project, B.S. Manjunath
Informedia Digital Video Library: Technology Outreach, Howard D. Wactlar

K

Key concepts in the architecture of the digital library, William Y. Arms

M

Meta-Information Environment of Digital Libraries, Terence R. Smith
Metadata: The Foundations of Resource Description, Stuart Weibel
Model Editions Partnership: Historical Editions in the Digital Age, David Chesnutt

N

Need for a Common Infrastructure: Digital Libraries and Electronic Commerce, Daniel Schutzer
Netlib Mathematical Software Repository, Shirley Browne, Eric Grosse, Tom Rowan
New Center at Columbia University for Digital Library Research: Fostering Interdisciplinary Research and Bridging Cultural Clashes, Judith Klavans
News-on-Demand: An Application of Informedia® Technology, Alexander G. Hauptmann, Michael J. Witbrock and Michael G. Christel

O

Options for the Future, Joshua Lederberg

P

Pricing Electronic Journals, Hal R. Varian

R

Recent Developments in GALEN II: Evolution of a Digital Library for the Health Sciences, John A. Kunze, Brian N. Warling
The Red Sage Project: An Experimental Digital Journal Library for the Health Sciences, A Descriptive Overview, Richard E. Lucier and Peter Brantley
Research in Support of Digital Libraries at Xerox PARC; Part I: The Changing Social Roles of Documents, Marti A. Hearst
Research in Support of Digital Libraries at Xerox PARC: Part II: Paper and Digital Documents, Marti Hearst, Gary Kopec, Dan Brotsky
ROADS to Desire: Some UK and Other European Metadata and Resource Discovery Projects, Lorcan Dempsey

S

SCAM Approach to Copy Detection in Digital Libraries, Narayanan Shivakumar, Hector Garcia-Molina
Secure Repository Design for Digital Libraries, Carl Lagoze
Summary of Stanford's Digital Library Testbed Design and Status, Andreas Paepcke
SunSITE: Serving Your Internet Needs Since 1992, Judson Knott, Paul Jones

T

Task Force on Archiving of Digital Information, John R. Garrett
Testbed Development for the Berkeley Digital Library Project, Virginia Ogle and Robert Wilensky
Text Is More Than Just Words on a Page, Susan Hockey (*guest editorial*)

U

Uniform Resource Names: A Progress Report, The URN Implementors
University of Michigan Digital Library Project: The Testbed, Daniel E. Atkins
User-Centered Iterative Design for Digital Libraries: The Cypress Experience, Nancy A. Van House, Mark H. Butler, Virginia Ogle, Lisa Schiff

V

The VARIATIONS Project at Indiana University's Music Library, David E. Fenske, Jon W. Dunn

W

Warwick Framework: A Container Architecture for Diverse Sets of Metadata, Carl Lagoze
Warwick Metadata Workshop: A Framework for the Deployment of Resource Description, Lorcan Dempsey, Stuart L. Weibel
What Do People Want from Information Retrieval? (The Top 10 Research Issues for Companies that Use and Sell IR Systems), W. Bruce Croft
Working Towards an Understanding of Digital Library Use: A Report on the User Research Efforts of the NSF/ARPA/NASA DLI Projects, Ann Peterson Bishop

Y

Yale University Library's Project Open Book: Preliminary Research Findings, Paul Conway

Copyright © 1996 Corporation for National Research Initiatives



hdl://cnri.dlib/september96-title.index

Working Groups

One of D-Lib's principal activities is stimulating and supporting working groups that address aspects of Digital Library research. Some of these groups are created by D-Lib; some are affiliated with the Digital Library Initiative, or other federally funded projects; and some are independent groups.

The following working groups in Digital Library research are currently associated with D-Lib:

- [Metadata to Describe Information in Digital Libraries](#)
- [User Needs Assessment and Evaluation](#)
- [Social Aspects of Digital Libraries](#)
- [Repository Interfaces](#)
- [Digitization and Conversion](#)
- [Naming Objects in the Digital Library](#)
- [Networked Computer Science Technical Report Library \(NCSTRL\)](#)
- [Task Force on Archiving Digital Information](#)

D-Lib is also a sponsor of:

- [The First ACM International Conference on Digital Libraries: Program and Proceedings](#)

D-Lib is coordinated by CNRI and is sponsored by the Defense Advanced Research Projects Agency (DARPA) on behalf of the Information Infrastructure Technology and Applications (IITA) Working Group of the High Performance Computing and Communications (HPCC) program.



wya/af/reb-a
Last revised: June 14, 1996

WORKING GROUPS

D-Lib Working Group on Metadata to Describe Information in Digital Libraries

Joint Chairs: Michael F. Goodchild, Terence R. Smith, University of California, Santa Barbara

If sense is to be made of the flood of information that will be available through digital libraries, it must be described effectively, so that it can be found, its value assessed, and its acquisition handled efficiently. Metadata is the term most often used to refer to the description of information objects to support these three functions of digital libraries. Digital library technology is capable of both supporting major augmentations to traditional metadata activities and providing a basis for catalog interoperability.

D-Lib is associated with two activities in this field. Both focus on the process by which creators of digital information can add metadata to their work at the time of creation. This metadata is then available for computer programs to use in building indexes and other access tools. It is also available as a basis for subsequent cataloguing or the creation of secondary information services.

The first of these activities comes out of the [Alexandria Digital Library](#) project at the University of California, Santa Barbara. This project concentrates on geospatial information, such as maps, but its studies of metadata are broad based and applicable to all types of on-line data. Alexandria is one of the projects in the ARPA/NSF/NASA Digital Library Initiative (DLI) and its metadata studies involve members of several of the other DLI projects.

The second activity is the [Metadata I](#) and [Metadata II](#) invited workshop series. The first of these was sponsored by OCLC and NCSA in March 1995, chaired by Stuart Weibel of OCLC. Its major contribution was the "Dublin Core" metadata elements. D-Lib has agreed to be a sponsor of subsequent workshops.

These two activities are inter-related. In particular, Alexandria is using the Dublin Core as a building block for its own developments.



wya/reb-a

Last revised: March 17, 1996

WORKING GROUPS

D-Lib Group on Naming in Digital Libraries

The D-Lib Group on Naming in Digital Libraries covers all aspects of naming of digital resources. This topic, which appears simple on the surface, proves to be remarkably subtle when applied to the complex world of digital libraries.

For several years, the Internet Engineering Task Force (IETF) was a focus for efforts to develop Uniform Resource Names (URNs). These are globally- unique, persistent, location-independent names that can be applied to any network resource. This work is being continued by an informal group of URN implementors. The focus of the D-Lib group is on the next stage, how to use names in large scale libraries.

User groups that wish to assign names to objects in a digital library are faced with a variety of issues. One type of question is the relationship of names to semantic concepts such as uniqueness, mutability, etc. Are these managed by the naming system or by an external system? In a large library, rules and conventions for assigning names can be very complicated. If users are to see the names, it is helpful if they have some structure to help them be remembered or recognized, but there are real dangers in attempting to embed semantic information into names.

There will be many naming schemes. Some, already exist and must be merged into the digital library. The integration of naming schemes is a technical challenge and an organizational one, requiring decisions about the registration of naming schemes, and the allocation of top-level names.

Few digital objects exist by themselves. They are parts of larger groups or made up of many components. Naming such complex and compound items proves to be intimately connected to questions of what metadata to keep for each component and how to represent the relationships between them. Proposed solutions include composite objects (which contain several separate objects) and meta-objects (which provide links to other digital objects).

Finally, all questions of naming must consider scale. Processes for naming and organizing small numbers of objects may be totally inadequate for large collections.



wya

Last revised: February 5, 1996

WORKING GROUPS

D-Lib Working Group on Repository Interfaces

Chair: William L. Scherlis, Carnegie Mellon University

This working group focuses on technical issues associated with repository interoperation. As digital libraries proliferate, many approaches to managing digital assets and associated meta-data are emerging. There are important differences among these approaches, and these differences have technical, legal, social, economic, and political dimensions. How can multiple repositories coexist and interact effectively?

The working group is motivated by several important trends: The complexity and semantic richness of objects and meta-data managed by repositories is increasing. Information objects of greater value are now being managed more routinely, raising issues of security, access control, and support for commerce. Performance demands are increasing, as is the quantity and size of information objects, particularly in multimedia applications. Digital libraries are interacting more often with personal, group, and wide area information services. Finally, the distinction is blurring between digital libraries and other institutional information resources such as databases and corporate webs.

The starting points for the working group are technologies that support management of information objects, their names, and associated meta-data-databases, distributed file systems, object bases, and the Web. Several digital library research groups have started to develop concepts that could provide a basis for repository interoperation, including the CS-TR architectural work of Kahn and Wilensky, the Stanford Infobus project of Garcia-Molina and Winograd, and the agent architecture of the Michigan DLI project. In addition to the need to reconcile these various approaches, there is a broader need to put them in the context of standards efforts in the wider community, including Web-associated standards, CORBA, OLE, z39.50, and SQL and its successors. All of these deal with resolving names to objects, and all deal in some measure with meta-data.

The initial effort of the working group is (1) to identify the dimensions of the space of repository interaction and interoperability, and the issues associated with achieving some transparency for users of the digital libraries, and (2) to assess current research and development efforts to understand the differences among them.



wya/reb-a

Last revised: February 5, 1996

WORKING GROUPS

D-Lib Group on Social Aspects of Digital Libraries

I. UCLA-NSF Workshop on Social Aspects of Digital Libraries

An invitational workshop was held at UCLA, February 15-17, 1996; 32 researchers, developers, and practitioners, 9 UCLA faculty facilitators, and 6 UCLA graduate research assistants participated. All materials from the workshop, including schedule and agenda, list of participants, participants' discussion papers and biographical statements, and summary reports presented at the meeting are available on the web site (<http://www.gslis.ucla.edu/DL/>).

We selected two research areas, each with three sub-topics, as focal points for a two-day workshop:

Information Needs: Identifying real information needs and developing digital libraries to meet those needs.

- Social context and culture
- Information needs and information seeking
- Linking user-learner needs and behavior to digital library design

End user searching and filtering: Designing digital libraries in which it is possible to find the right information in a glut of information.

- Organization, description and representation of information
- Search capabilities for users
- Interface design for information retrieval

II. Results of the workshop

While we bounded the scope of the workshop to provide a starting point for discussion and a set of criteria for selecting participants, our participants quickly expanded those boundaries.

The boundaries expanded in several directions:

- Level of analysis: Our scope, as stated in the background paper (see web site), focused on the needs and activities of the individual user. While important, we must recognize that individuals do not work with information resources in isolation from their communities. They perform individual tasks in the context of their work teams, classroom, and other social organizations. Many tasks are performed in group contexts; we must consider CSCW and collaboratory environments as well. Multiple levels of analysis are required.
- Scope of analysis: Our scope addressed information searching and retrieval processes. While important, we must set searching in the context of the cycle of information creation and utilization. People will create information in digitized form that becomes part of digital libraries and need tools and functional capabilities for doing so. They will search for information created by other people, and for purposes other than those intended by the creators, requiring a variety of searching functions. Once located, they will incorporate new information into other products and processes that become part of the life-cycle. We need consistent means to organize, describe, represent, and dispose of information throughout these activities and processes.
- Content vs. process: Our scope addressed digital libraries as a set of digitized resources and associated technical capabilities for searching for information, which is roughly the scope defined in the digital libraries initiative. This scope statement addresses the digitized content of digital libraries but does not recognize the social processes around digital libraries -- the "library" in digital libraries. We need to

address both, hence the distinction made in the second definition stated in the beginning of this report.

III. Research agenda for Social Aspects Of Digital Libraries

We will present the research agenda with respect to the two definitions of digital libraries outlined above. These two definitions converge in a model of the life cycle of information and information processes.

The model covers the sequence from the creation of information (author, artist, memo-writer, data-generation scientist, publisher, etc.), through the searching for information, and the utilization of it, often for very different purposes than it was originally created. An exit from the loop is given to indicate that we do not need to save everything created in digital form -- indeed, we need criteria and mechanisms to decide what to keep and what to destroy. The model addresses the social context for all aspects of the cycle -- people create information for one purpose, search for it for another, and utilize for another. We need to organize, describe, and represent for multiple uses but we must design based on an understanding of what those uses might be. Similarly, we need searching and utilization interfaces that support many perspectives and purposes, with a variety of functional capabilities -- but all must be based on some understanding of the underlying tasks/roles that the information will play in a social context.



clb/wya

Last revised: March 18, 1996

An Agent-Based Architecture for Digital Libraries

William P. Birmingham
The University of Michigan
Electrical Engineering and Computer Science Department
School of Information Science and Library Studies
Ann Arbor, MI 48109
wpb@eecs.umich.edu

D-Lib Magazine, July 1995

-
- [Introduction](#)
 - [Agents](#)
 - [What the architecture provides](#)
 - [The Conspectus and the conspectus language](#)
 - [Status and summary](#)
 - [Acknowledgements](#)
 - [References](#)
-

d-lib forum

d-lib magazine

Introduction

One of the most exciting promises of digital libraries is access to a great variety of information and *services* that transcend what is available today through on-line services, such as the World-Wide Web (WWW). A library is more than just stacks of materials on shelves; it is also highly trained people that provide valuable services. These services include such things as *organization and cataloging*, research, notification of new publications, and so forth. Indeed, one of the greatest assets of libraries are these high-valued services. The WWW, while it probably contains more information than any single traditional library, is arguably not as useful as a traditional library because it lacks these services (particularly organization and sophisticated search support). No one is dismantling their libraries because of the WWW yet. The University of Michigan Digital Library Project (UMDL) [1,2] believes that a successful digital library needs to provide both access to a wide variety of valuable content and services.

Because the range of both content and services that are possible for a digital library are potentially large (we cannot even imagine what will be available or needed in the future), there will be no single, complete digital-library solution. Rather, we expect that as editing tools become better and access to networks becomes easier and cheaper, there will be millions of content suppliers; "everyman" can become a vanity press on the information superhighway. We believe that the days of centralized suppliers of information (e.g., large publishing houses and traditional libraries) are numbered, and that the traditional notion of a "collection" will

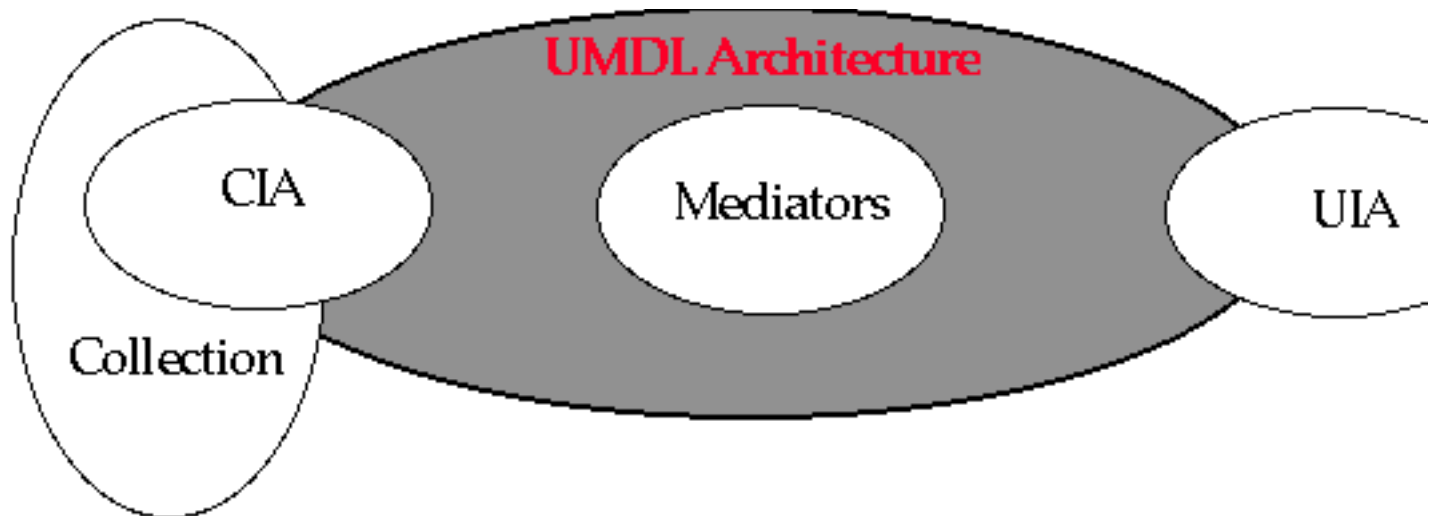


Figure 1: UMDL agent types

As the architecture is developed, the broad classes of agents depicted in Figure 1 will be continually refined; specialized agents will be added to the system as needed (the modularity property). For example, we can create user interfaces that are customized to a particular class of users, rather than to a particular collection or access mechanism (e.g., Boolean search over controlled vocabulary). In addition, the ability to *team* agents (as described in the next section, "What the architecture provides") dynamically creates new services with new agents, which is especially important since we anticipate the agent population will be constantly changing.

What the architecture provides

From a user's perspective, the types of high-level support that make a digital library worth using, such as searching, will be performed by a team of agents. For example, consider Figure 2, where a user (through the UIA) is searching for all articles by "Joan Q. Publique". Assuming that all agents have registered with the registry agent, the UIA contacts a query planner by first requesting the registry for a query planner that knows about author searching. The query planner then goes to the registry to get the addresses of a name authority (meta data that gives variations of Joan Q. Publique) and a name index (a partial listing of collections that contain works sorted by author). The planner then interrogates the authority, and then the index, finally determining the address of a particular collection. The collection is then accessed by the UIA using a protocol specific to the CIA.

It is easy to image how this process can be extended for different types of search by adding new types of agents (e.g., subject indexes and new kinds of query planners). The teaming methods gives the architecture a dynamic planning ability[5] that is critical for finding the best way to perform some service, as well as easily incorporate new types of search methods. There is, however, a cost.

Key Concepts in the Architecture of the Digital Library

William Y. Arms
Corporation for National Research Initiatives
Reston, Virginia
warms@cnri.reston.va.us

D-Lib Magazine, July 1995

Introduction

For the past two years, the Computer Science Technical Reports project (CS-TR) has been developing an architecture for a digital library with funding from the Department of Defense's Advanced Research Projects Agency (ARPA). This is a general purpose framework for a digital library in which very large numbers of objects, comprising all types of material, are accessible over national computer networks. It is described in a paper by Robert Kahn and Robert Wilensky (cnri.dlib/tn95-01).

This introduction describes the author's view of eight general concepts that emerged from the discussions. These concepts are key issues in the transition to a true digital library from the network services that we have today. The Kahn/Wilensky paper contains a comprehensive framework for resolving the issues.

General Principles

- 1. The technical framework exists within a legal and social framework
 - 2. Understanding of digital library concepts is hampered by terminology
 - 3. The underlying architecture should be separate from the content stored in the library
 - 4. Names and identifiers are the basic building block for the digital library
 - 5. Digital library objects are more than collections of bits
 - 6. The digital library object that is used is different from the stored object
 - 7. Repositories must look after the information they hold
 - 8. Users want intellectual works, not digital objects
 - Reference
-

d-lib forum

d-lib magazine

General Principles

1. The technical framework exists within a legal and social framework



Metadata: The Foundations of Resource Description

Stuart Weibel
Office of Research, OCLC Online Computer Library Center, Inc.
weibel@oclc.org

D-Lib Magazine, July 1995

This paper is an abbreviated version of the Summary Report of the OCLC/NCSA Metadata Workshop. It sets forth a proposal for the content of a simple resource description record (the Dublin Core Metadata Element Set) and outlines a series of further steps to advance the standards for the description of networked information resources.

- Introduction
- Underlying Assumptions
- Implementations
- Next Steps
- References

d-lib forum

d-lib magazine

Introduction

The explosive growth of interest in the Internet in recent years has created a digital extension of the academic research library for certain kinds of materials. Valuable collections of texts, images and sounds from many scholarly communities -- collections that may even be the subject of state-of-the-art discussions in these communities--now exist only in electronic form and may be accessible from the Internet. Knowledge regarding the whereabouts and status of this material is often passed on by word of mouth among members of a given community. For outsiders, however, much of this material is so difficult to locate that it is effectively unavailable.

Why is it so difficult to find items of interest on the Internet or the World Wide Web? A number of well-designed locator services, such as Lycos [<http://lycos.cs.cmu.edu/>], are now available that automatically index many of the resources available on the Web and maintain up-to-date databases of locations. But indexes are most useful in small collections within a given domain. As the scope of their coverage expands, indexes succumb to problems of large retrieval sets and problems of cross disciplinary semantic drift. Richer records, created by content experts, are necessary to improve search and retrieval. Formal standards such as the TEI Header and MARC cataloging) will provide the necessary richness, but such records are time consuming to create and maintain, and hence may be created for only the most important resources.

An alternative solution that promises to mediate these extremes involves the creation of a record that is more informative than an index entry but is less complete than a formal cataloging record. If only a small amount of human effort were required to create such records, more objects could be described, especially if the author of the resource could be encouraged to create the description. And if the description followed an established standard, only the creation of the record would require human intervention; automated tools could discover these descriptions and collect them.

Can a simple metadata record be defined that sufficiently describes a wide range of electronic objects? The Online Computer Library Center (OCLC) and the National Center for Supercomputing Applications (NCSA) convened the invitational Metadata Workshop on March 1-3, 1995, in Dublin, Ohio to address this issue. Fifty-two librarians, archivists, humanities scholars and geographers, as well as standards makers in the Internet, Z39.50 and Standard Generalized Markup Language (SGML) communities, met to identify the scope of the problem, to achieve consensus on a list of metadata elements that would yield simple descriptions of data in a wide range of subject areas, and to lay the groundwork for achieving further progress in the definition of metadata elements that describe electronic information.

Goals

Goals of the workshop included fostering a common understanding of the problems and potential solutions among the stakeholders and promoting a consensus on a core set of metadata elements to describe networked resources.

Scope

Since the Internet contains more information than professional abstractors, indexers and catalogers can manage using existing methods and systems, it was agreed that a reasonable alternative way to obtain usable metadata for electronic resources is to give authors and information providers a means to describe the resources themselves. The major task of the Metadata Workshop was to identify and define a simple set of elements for describing networked electronic resources. To make this task manageable, it was limited in two ways. First, only those elements necessary for the discovery of the resource were considered. It was believed that resource discovery is the most pressing need that metadata can satisfy, and one that would have to be satisfied regardless of the subject matter or internal complexity of the object.

Secondly, the discussion was further restricted to the metadata elements required for the discovery of what were called **document-like objects**, or **DLOs** by the workshop participants. It was believed that DLOs are still the most common type of resource sought in the Internet and that whatever solution could be proposed for DLOs could be extended to other kinds of resources. More importantly, the likelihood of making progress on this challenging problem would be increased if attention could initially be restricted to something familiar.

DLOs were not rigorously defined, but were understood by example. For example, an electronic version of a newspaper article or a dictionary is a DLO, while an unannotated collection of slides is not. Of course, the crux of the problem is that in a networked environment, DLOs can be arbitrarily complex because they can consist of text with callouts to images, audio or video clips, or to other hypertext documents. The Metadata Workshop participants made no attempt to limit the complexity of DLOs, except to say that the intellectual content of a DLO is primarily text, and that the metadata required for describing DLOs will bear a strong resemblance to the metadata that describes traditional printed texts.

As a result of the restricted focus of the workshop, certain issues required for a complete description of DLOs, such as cost, archival status and copyright information, were eliminated from the scope of the discussion. Elements required for the description of objects other than DLOs, such as the elements required for the description of complex geological strata in a geospatial resource, were also beyond the scope of the discussion. The goal was to define a core set of metadata elements that would allow authors and information providers to describe their work and to facilitate interoperability among resource discovery tools. But because the core elements do not yield a complete description of objects in a networked environment, careful consideration was also given to mechanisms for extending the element set.

The primary deliverable from the workshop was a set of thirteen metadata elements, named the **Dublin Core Metadata Element Set** (or Dublin Core, for short). The Dublin Core was proposed as the minimum number of metadata elements required to facilitate the discovery of document-like objects in a networked environment such as the Internet. The syntax was deliberately left unspecified as an implementation detail. The semantics of these elements was intended to be clear enough to be understood by a wide range of users.

Below is a brief description of the elements in the Dublin Core **Dublin Core Element Description**

- **Subject:** The topic addressed by the work
- **Title:** The name of the object
- **Author:** The person(s) primarily responsible for the intellectual content of the object
- **Publisher:** The agent or agency responsible for making the object available
- **OtherAgent:** The person(s), such as editors and transcribers, who have made other significant intellectual contributions to the work
- **Date:** The date of publication
- **ObjectType:** The genre of the object, such as novel, poem, or dictionary
- **Form:** The physical manifestation of the object, such as Postscript file or Windows executable file
- **Identifier:** String or number used to uniquely identify the object
- **Relation:** Relationship to other objects
- **Source:** Objects, either print or electronic, from which this object is derived, if applicable
- **Language:** Language of the intellectual content
- **Coverage:** The spatial locations and temporal durations characteristic of the object

To make this discussion concrete, consider an electronic a record created with the relevant portions of the Dublin Core, and a sample syntax, that describes an electronic version of Maya Angelou's poem "On the Pulse of Morning". This description is based on a record created by the University of Virginia Library's Electronic Text Center. (For a description of that project, see Gaynor [[Gaynor](#)].)

- **Subject:** Poetry
- **Title:** On the Pulse of Morning
- **Author:** Maya Angelou
- **Publisher:** University of Virginia Library Electronic Text Center
- **OtherAgent:** Transcribed by the University of Virginia Electronic Text Center
- **Date:** 1993
- **Object:** Poem
- **Form:** 1 ASCII file
- **Identifier:** AngPuls1
- **Source:** Newspaper stories and oral performance of text at the presidential inauguration of Bill Clinton
- **Language:** English

Underlying Assumptions

The discussions at the Metadata Workshop revealed several principles that should guide the further development of the element set. Adherence to these principles increases the likelihood that the core element set will be kept as small as possible, that the meanings of the elements will be understood by most users, and that the element set will be flexible enough for the description of resources in a wide range of subject areas. These principles are intrinsicity, extensibility, syntax independence, optionality, repeatability, and modifiability.

Intrinsicity

The Dublin Core concentrates on describing intrinsic properties of the object. Intrinsic data refer to the properties of the work that could be discovered by having the work in hand, such as its intellectual content and physical form. This is distinguished from extrinsic data, which describe the context in which the work is used.

MAGAZINE

Uniform Resource Names

A Progress Report

The URN Implementors

D-Lib Magazine, February 1996

ISSN 1082-9873

Introduction

The development of networked information requires reliable ways to name resources on networks. The Internet community has adopted the term, "Uniform Resource Name (URN)", for a name that identifies a resource or unit of information independent of its location. URNs are globally unique, persistent, and accessible over the network.

The concept of universal names has been warmly embraced by the networking and library communities, but convergence on the details proved difficult until recently. During fall 1995, however, members of the principal groups that are actively working in the field reached outline agreement on most of the major topics. The main characteristics of this agreement are described in this paper.

The catalyst for the recent progress was a meeting in October 1995 hosted by Keith Moore at the University of Tennessee. Invitations were sent to every group that had a current Internet draft on this subject. The URN groups represented are listed at the end of this report. This meeting was followed by a series of discussions including informal sessions at the December meeting in Dallas, Texas, of the Internet Engineering Task Force (IETF).

Convergence is important because many people who manage large collections of on-line information have been reluctant to commit to using any form of URN during a period of flux. The present consensus has two major results:

- Users who wish to give permanent names to on-line resources can now plan to incorporate URNs from existing naming schemes in documents, indexes, and on-line systems. They can be reasonably confident that future developments of the URN framework will not force them to reformat or otherwise modify existing URNs.
- The implementation of this framework will remove the concern that using a particular name scheme might affect longevity or the future usefulness of assigned URNs. The framework allows continued support for existing URNs, through other resolution systems, if one name scheme ceases to be supported in its original form. Thus users who assign names within any of the agreed-upon schemes are assured against obsolescence.

This report summarizes the emerging consensus. A strength of the framework is that it allows different approaches to be pursued, and the framework has the ability to evolve over the long term. Naming is a complex issue and the groups are interested in URNs for a variety of different reasons. They bring different philosophies and different technical approaches. Their implementations range in scope and complexity. It is therefore encouraging for the community that they have reached general agreement and are working together to find technical solutions to the outstanding questions.

Working Towards an Understanding of Digital Library Use

A Report on the User Research Efforts of the NSF/ARPA/NASA DLI Projects

Ann Peterson Bishop
Graduate School of Library and Information Science
University of Illinois at Urbana-Champaign
218 LIS Building, 501 East Daniel Street
Champaign, IL 61820
abishop@uiuc.edu

D-Lib Magazine, October 1995

Introduction

The Digital Library Initiative (DLI) projects, funded jointly by the National Science Foundation (NSF), the Advanced Research Projects Agency (ARPA), and the National Aeronautics and Space Administration (NASA) began about a year ago. Their user study teams have already produced some valuable findings and described some provocative theoretical and methodological challenges. From my vantage point as coordinator of the University of Illinois DLI Social Science team, I will highlight the efforts of the six projects to communicate with each other about user research. Links to the DLI project home pages and to some of the papers published by project members have been included for more in-depth coverage of some of the issues summarized. In this article, I will also discuss the upcoming Allerton Institute at the University of Illinois, a methodological forum on digital library use that will provide another means for researchers in a variety of disciplines and settings to share their ideas and concerns about the conduct of social science research related to digital library use.

The Growth of Digital Libraries and the Challenge of Understanding Their Use

Improvements in information technologies and increased support directed towards our national information infrastructure have led to the development of a wide range of digital library collections and services. Academic, special, and public libraries are implementing on-line systems that provide their patrons with electronic access to library catalogs and a variety of other information resources. NASA is developing on-line collections of images and data for scientists and engineers. Museums are digitizing their collections and making them available on the Internet. Members of scientific communities are building collaboratories to support their work and communication. Publishers are experimenting with the creation of digital archives of their journals and books. And individuals and groups from all walks of life are using community-based networks to provide local and global access to information resources they have created. In addition to this array of existing networked information tools and resources--all of which can be thought of as variations on theme of the digital library or as pieces and layers of the digital information infrastructure--research and development projects related to building the next generation of digital library systems are also flourishing.

Digital libraries pose fascinating socio-technical challenges for understanding their use. Those supporting the construction of digital libraries are naturally concerned that their investments pay off in terms of attracting users and making information services more effective and efficient. The design and evaluation of digital libraries,



- [Electronic Document Collections at Virginia Tech](#)
 - [Imagebase \(Uses PURLs and Dublin Core Metadata standard\)](#)
 - [Meeting Schedule](#)
 - [Pointers to information about Digital Libraries](#)
 - [Project Reports](#)
 - [References: Research Department Virginia Tech Computing Center](#)
-

References

- **General**
 - [Digital Libraries Research and Development Forum \(D-Lib\)](#)
 - [Florida Center for Library Automation's Digital Library Project](#)
 - [IBM Digital Library](#)
- **Specific Topics**
 - **Agents**
 - [UMBC Intelligent Software Agent Resources](#)
 - [Survey of Intelligent Software Agents](#)
 - [More agent links](#)
 - **Metadata**
 - [Metadata: the Foundations of Resource Description](#)
 - [OCLC/NCSA Metadata Workshop Report](#)
 - [RFC-1807](#)
 - [TEI](#)
 - **Naming**
 - [Handles](#)
 - [PURL](#)
 - **Z39.50**
 - [Isite Software](#)
 - [Library of Congress WWW/Z39.50 Gateway/Info](#)
 - [Prise 1.0 Software](#)
 - [Willow](#)

Some of these documents are in Adobe's Portable Document Format (PDF). In order to view them, you will need a [PDF viewer](#)

University Libraries, Virginia Tech
Send Suggestions or Comments to webmaster@scholar.lib.vt.edu
Last updated: October 1, 1996

URL: <http://scholar.lib.vt.edu/digilib/>

Digital Library Source Book, 1993, ed. E. Fox

To order a paper copy, or find out background information please look at the [README](#) file. To use an Adobe Acrobat Reader or Exchange to work with the book, look at the [PDF version](#). Otherwise, use the PostScript version that appears below in sections.

- [Title Page](#)
- [Table of Contents](#)
- [Chapters 1-7 all together \(1.26 M\)](#)
- [Chapter 1: Future Directions in Text Analysis, Retrieval and Understanding \(esp. white paper on A National Electronic Science, Engineering, and Technology Library\)](#)
- [Chapter 2: July 1992 Workshop](#)
- [Chapter 3: December 1992 Workshop](#)
- [Chapter 4: Notable Events](#)
- [Chapter 5: Directory of Interested Parties](#)
- [Chapter 6: Summary and Recommendations](#)
- [Chapter 7: Glossary](#)
- [Index](#)

See also more information of interest:

- [April 1995 Communications of the ACM](#)
- [Gladney et al. report on DL requirements and architecture \(PostScript\)](#)
- [PowerPoint presentation by Fox for 1994 Digital Libraries Workshop at Rutgers \(to be decoded by binhex\)](#)
- [PowerPoint presentation by Fox for 1994 Digital Libraries Workshop at Texas A&M \(printable, in black and white, to be decoded by binhex\)](#)
- [PowerPoint presentation by Fox for DL Keynote at EG-MM'94 in Graz \(to be decoded by binhex\)](#)
- [PowerPoint presentation by Fox for DL Keynote at ISMIS'94 in Charlotte \(to be decoded by binhex\)](#)
- [WWW Pages for CS2984 Course Notes on Digital Libraries](#)

Envision

The Envision Project was funded as **A User Centered Database from the Computer Science Literature** by NSF for 1991-95. ACM has provided free access to their publications.

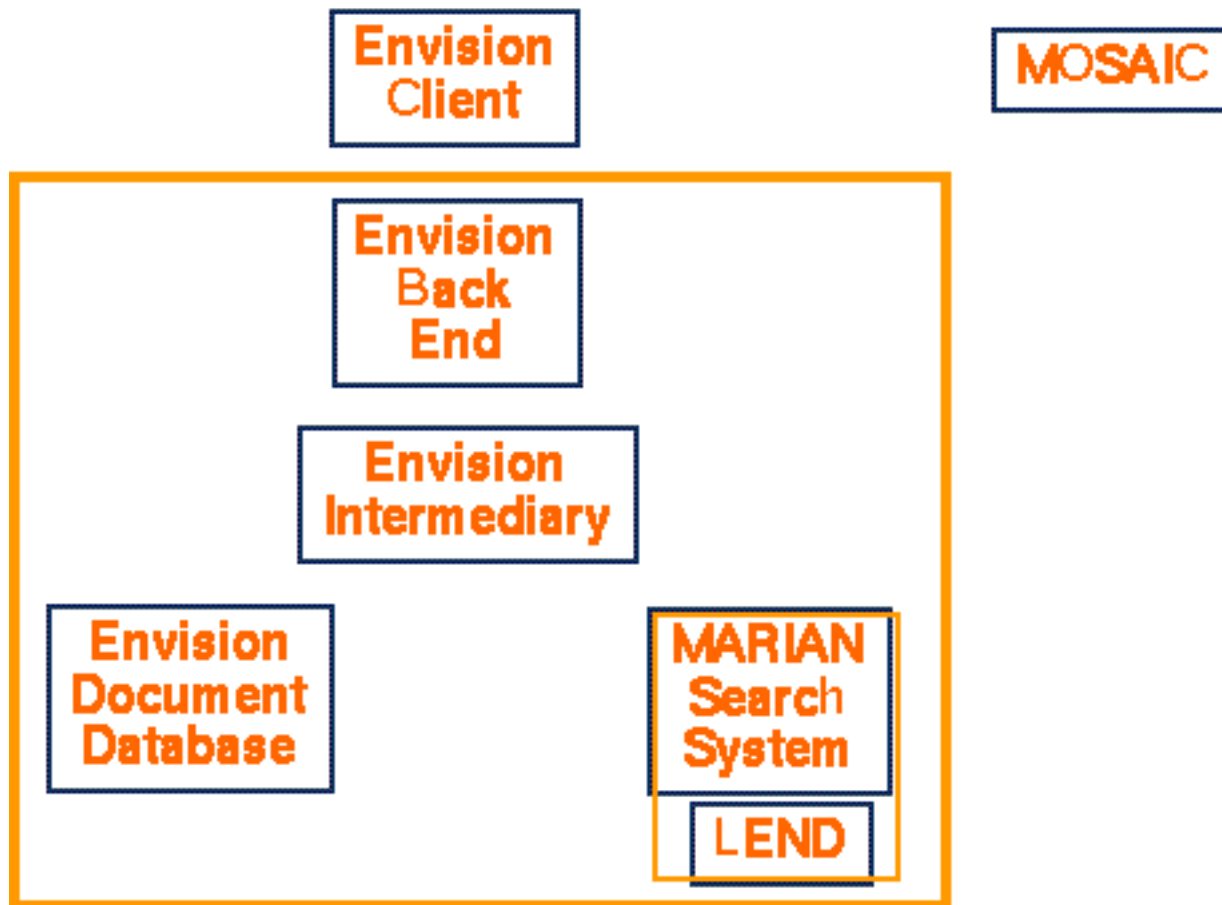
Efforts have concentrated on building an archive based upon SGML, developing an object-oriented database, applying the MARIAN retrieval system and WWW, and constructing a special search interface based upon user wishes.

The interface includes:

- a query screen
- a results list screen
- a results visualization screen
- Mosaic display of retrieved documents

The system architecture is a combination of various elements:

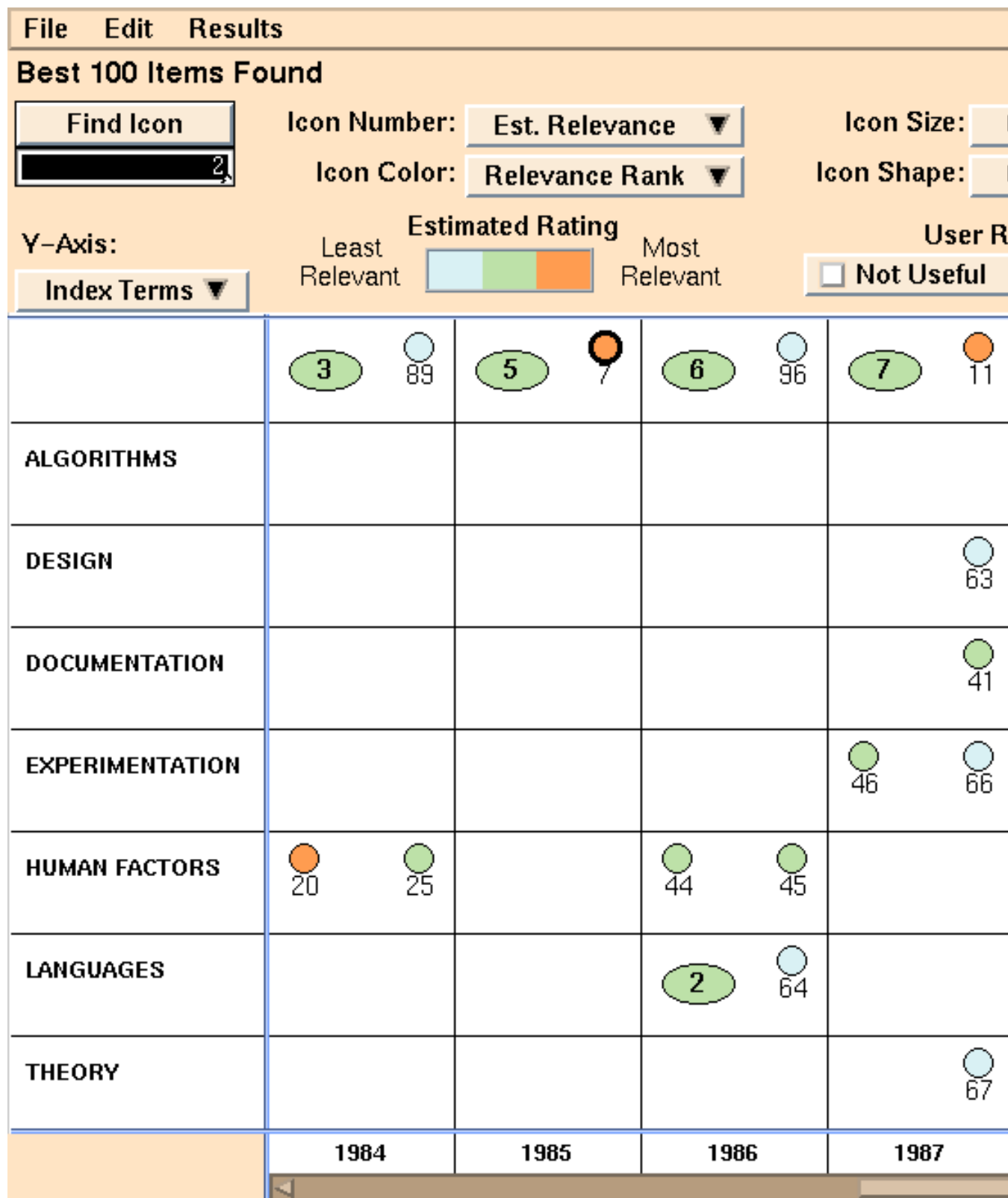
Envision



Envision - Results Screens

The interface includes:

- Graphic View 1:



PROJECT ENVISION FINAL REPORT

A User-Centered Database from the Computer Science Literature NSF Grant IRI-9116991

Edward A. Fox, Lenwood S. Heath, Deborah Hix
Department of Computer Science
Virginia Polytechnic Institute and State University
Blacksburg, VA 24061-0106

Converted to HTML Wed Jul 5 17:41:14 EDT 1995

Summary of Completed Project

With the support of the National Science Foundation and the Association for Computing Machinery (ACM), the Envision project has developed a prototype digital library of computer science literature that is highly usable (from user-centered design), highly structured (from SGML and an object database), and highly integrated (from hypertext links among objects). The result is a representation of part of the computer science literature as a cohesive body of knowledge that can be searched and viewed in innovative ways. The user interface was designed with careful attention to user needs and desires (through interviews with potential users), to graphic detail (through involvement of an artist and attention to the research literature on graphical perception and psychophysics), and to usability (through an iterative process of usability evaluation). Recognizing the need to translate enormous quantities of documents in an unlimited variety of input formats into a single standard format, the project developed a flexible system for analyzing the structures (e.g., titles, authors, paragraphs, and references) within a document and translating that structure into any standard markup scheme. The Envision distributed server supports simultaneous access to the library by a number of users and in a variety of ways. The Envision software is soon to be installed at ACM headquarters and made available to ACM members. The Envision system will continue in use at Virginia Tech and Norfolk State University to support the work of a related NSF Educational Infrastructure grant.

Technical Information

The list of publications resulting from Envision research appears in the References section. The data collected during this project include electronic versions of computer science literature (Section [2.1](#)). A great deal of software was created or adapted during this project (Section [2.2](#)). A number of people have contributed to the success of the Envision project. These are listed in Appendix [A](#). We are particularly proud of the number of undergraduate students who were able to obtain research experience on the Envision project.

Computer Science Literature

The library contains bibliographic records, full-text articles, and scanned page images. The bulk of the approximately 100,000 bibliographic records are from ACM's *Computing Archive*. We have also incorporated publicly available bibliographies from Ohio State University, the University of Arizona, and the University of Melbourne. We have approximately 700 full-text articles from *Communications of the ACM* and several of the

ACM *Transactions*. Finally, we have about 13,000 scanned page images, from various ACM publications and the technical report series of the Virginia Tech Department of Computer Science.

Envision Software

The major software components of the Envision system are the following.

1. **The Envision Client.** This component interacts with a user to accomplish the tasks of querying the Envision library and visualizing result sets in the Envision graphical display. This client interface is a major innovation of the Envision project and required the greatest amount of effort in interaction design and evaluation, in software design, and in software development.
2. **A WWW Viewer.** Envision employs a WWW browser as its presentation front end. Currently we use Mosaic running on a UNIX workstation.
3. **The Envision Intermediary.** This component communicates with the Envision client over the network to maintain session information, packages queries for the MARIAN search system, and packages result sets to pass back to the Envision client.
4. **The MARIAN Search System.** This component, developed in a separate research effort to access a library catalog, searches the Envision library for documents relevant to the user's query. The search can be based on a combination of title, author, and content words. Result sets are ranked by estimated relevance.
5. **Enhanced WWW Server.** Envision documents are viewed via a WWW interface that accesses a WWW server enhanced by CGI scripts that retrieve Envision objects from the object database and package them into HTML for presentation.
6. **The Object Database.** The Envision object database maintains our view of the structure of the library in terms of classes such as document, person (author), institution, publication, and keywords. Objects in this database refer to related objects, providing a rich hypermedia structure.
7. **The DELTO System.** The DELTO (Document Analysis and Translation) system addresses the need to convert documents in many ill-defined input formats that are received for inclusion in the Envision library into the standard SGML structural representation needed by the Envision object database and MARIAN searchers. This system emphasizes flexibility and automation. DELTO is a major innovation of the Envision project.

Components 1 and 2 run under the X Window System; these have been tested on Sun, DECstation, and DEC Alpha workstations. Components 3 and 4 run on a NextStation. Components 5, 6, and 7 run on a DEC Alpha and should port easily to other UNIX systems.

A public release of the Envision software is due during the summer of 1995. The Envision client will be freely available over the Internet by anonymous ftp from Virginia Tech. Initially, the server components (3, 4, 5, 6, and 7) and the actual library of electronic documents will be released to the ACM, as well as used in a related NSF Educational Infrastructure project at Virginia Tech and Norfolk State University.

References

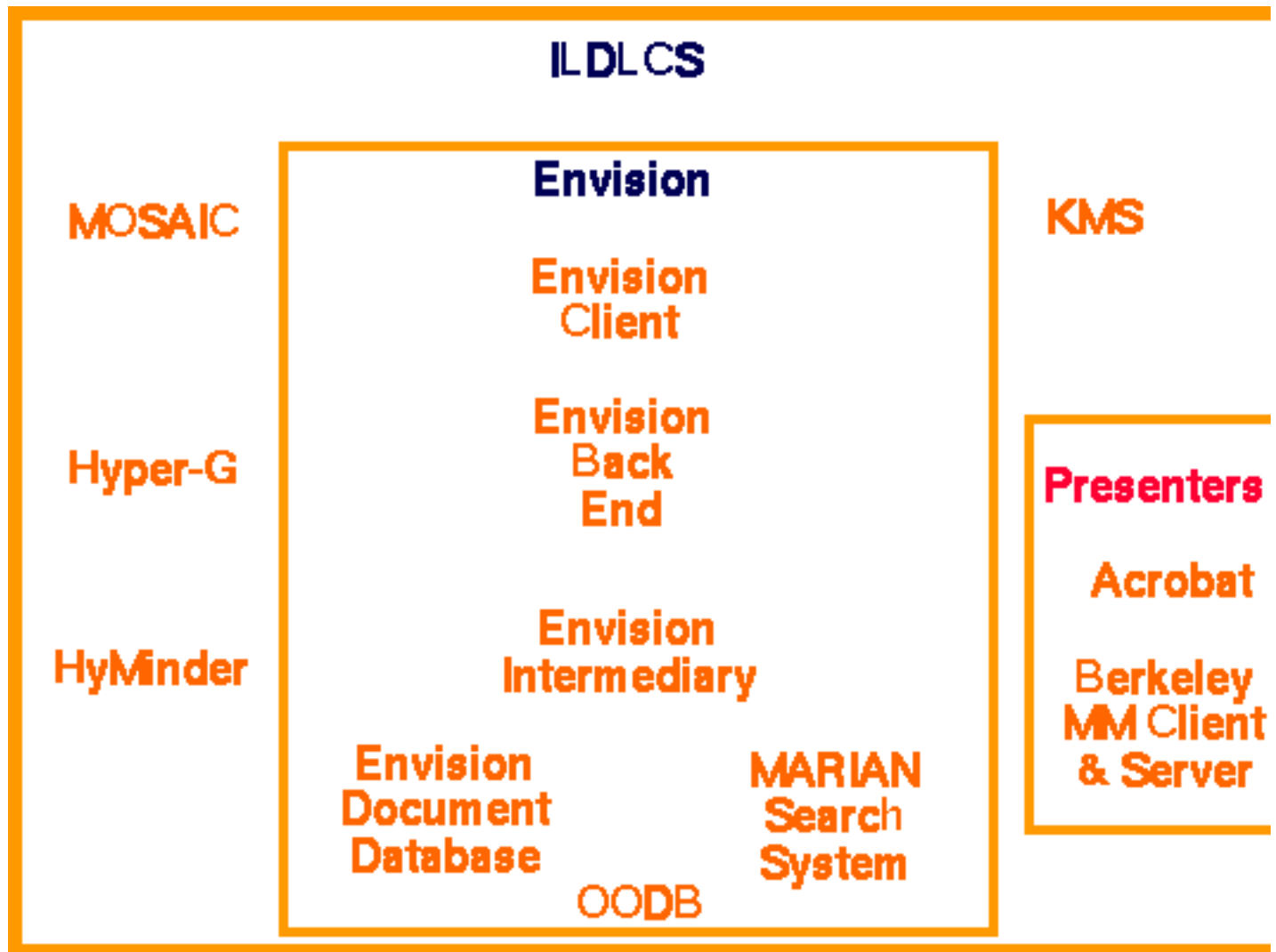
- 1 G. A. Averbach. A system for document analysis, translation, and automatic hypertext linking. Master's thesis, Department of Computer Science, Virginia Polytechnic Institute and State University, Blacksburg, Virginia, 1995.
- 2 S. Betrabet, E. A. Fox, and Q. Chen. A query language for information graphs. Technical Report TR 93-03, Department of Computer Science, Virginia Polytechnic Institute and State University, 1993.
- 3 D. J. Brueni, B. Cross, E. A. Fox, L. S. Heath, D. Hix, L. T. Nowell, and W. C. Wake. What if there were desktop access to the computer science literature? In *Proceedings of the 21st Annual ACM Computer Science Conference*, pages 15-22, 1993. Also available as Tech. Report TR 92-42,

ILDLCS

The ILDLCS Project was funded as **Interactive Learning with a Digital Library in Computer Science** by NSF for 1993-96. ACM has provided free access to their publications, as have several other publishers. Norfolk State University is a partner in this effort, which building upon the Envision Project. More details are given online.

Efforts have concentrated on developing courseware for 4 courses that have been redone in paperless manner, constructing tools to help with algorithm visualization, and extending the Envision efforts to help with as many CS courses as possible.

The system architecture is a combination of various elements:





World Wide Web (WWW) Traffic Analysis Research

Computer Science Department
Virginia Polytechnic and State University
Blacksburg, VA 24061-0106

What's New

11 September 1996

Slide presentation [gzip'd slides] from our SIGCOMM96 paper "Removal Policies in Network Caches for World-Wide Web Documents" is now available.

The mission of our research group is to

- collect and make available to other researchers a collection of Web traffic traces from a variety of networks,
 - work to make the use of proxy caches more effective through performance evaluation of different proposed cache designs, and
 - production of tools to assist in the collection and analysis of Web traffic and in the evaluation of cache designs.
-

Resources currently available:



Marc Abrams, Stephen Williams, "Complementing Surveying and Demographics with Automated Network Monitoring," *World Wide Web Journal*, No. 3, Vol. 1., June 1996.



Stephen Williams, Marc Abrams, Charles R. Standridge, Ghaleb Abdulla, Edward A. Fox, "Removal Policies in Network Caches for World-Wide Web Documents," *Proceedings, ACM SIGCOMM*, Stanford, CA, August 1996, pp. 293-305, and slides [gzip'd slides] from presentation.



Marc Abrams, Stephen Williams, Ghaleb Abdulla, Shashin Patel, Randy Ribler, Edward A. Fox, "Multimedia Traffic Analysis Using Chitra95," *Proceedings: ACM Multimedia '95*, San Francisco CA, November 1995, pp 267-276.



Marc Abrams, Charles R. Standridge, Ghaleb Abdulla, Stephen Williams, Edward A. Fox, "Caching Proxies: Limitations and Potentials," *Proceedings: 4th Inter. World-Wide Web Conference*, Boston, MA, Dec. 1995, pp 119-133.



Ghaleb Abdulla, Marc Abrams, Edward A. Fox, *Scaling the WWW*, working paper, March 1996.



Experience in Network Delivery of Computer Science Courseware, transparencies from presentation at the 2nd Annual SUCCEED Conference, N.C. State, March 1995. Describes using Chitra94 to analyze World-Wide Web traffic from Computer Science courses at Virginia Tech.



WWW data collection tools used in certain papers listed above, and slides from a presentation on some of the tools [Adobe pdf or postscript versions].

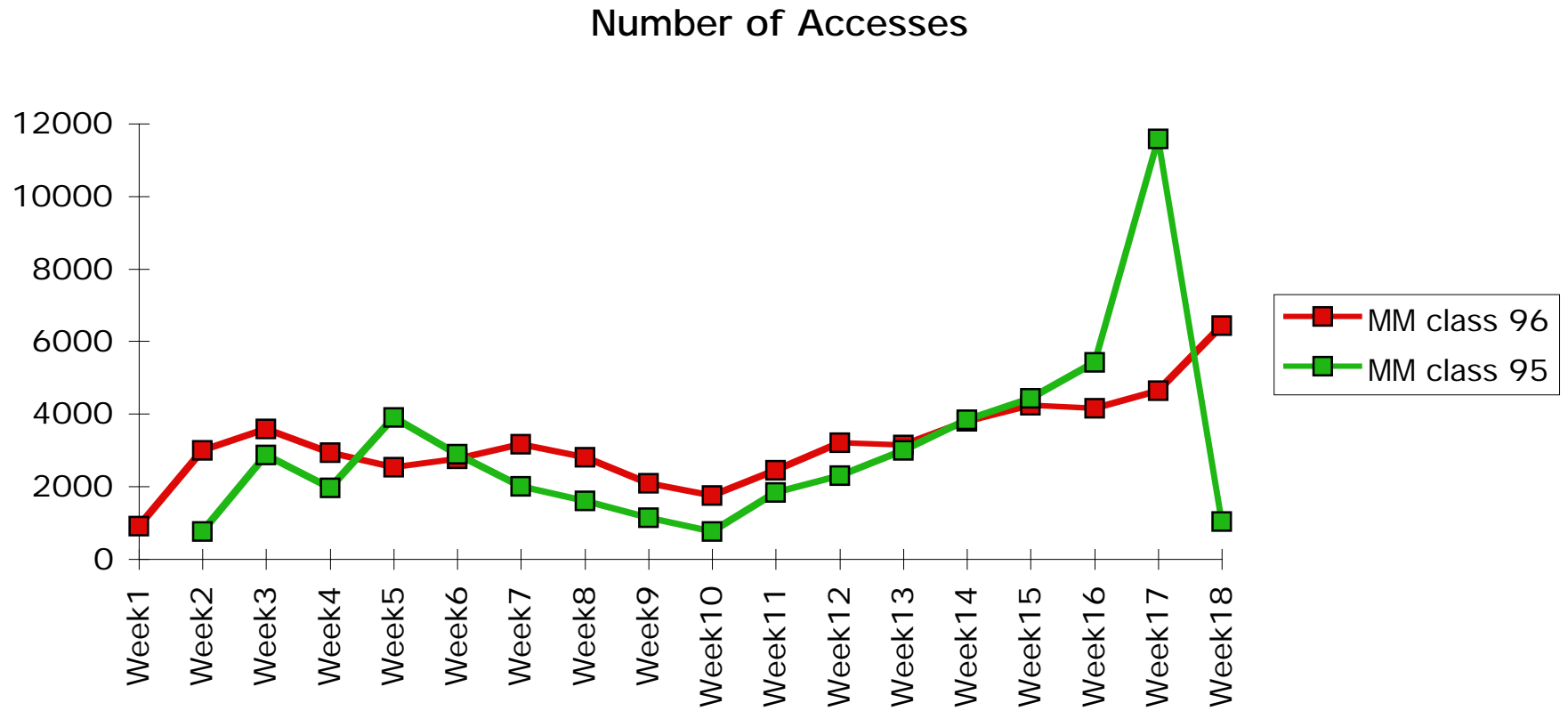


Chitra, a tool that can be used to analyze trace data, including traces of traffic from the WWW. The trace formats currently supported are common log format" and the TCPdump tool. (The next release of the tool is expected in Spring 1996.)

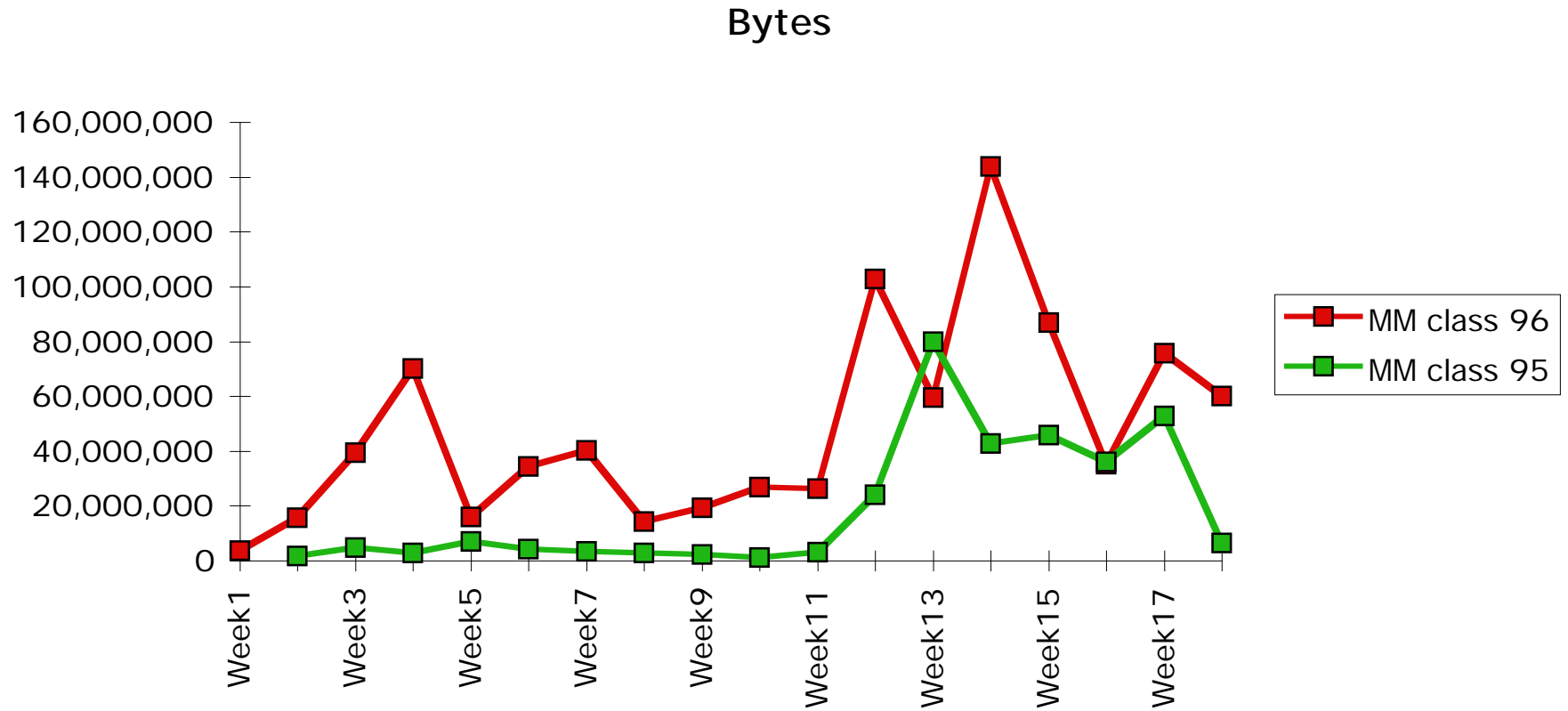
Research team members:

- Marc Abrams [URL: <http://www.cs.vt.edu/vitae/Abrams.html>]
Computer Science Department
Virginia Tech
Blacksburg, VA 24061-0106 USA
E-Mail: abrams@vt.edu
 - Ghaleb Abdulla [URL: <http://csgrad.cs.vt.edu/~abdulla/>]
Computer Science Department
Virginia Tech
Blacksburg, VA 24061-0106 USA
E-Mail: abdulla@csgrad.cs.vt.edu
 - Edward A. Fox [URL: <http://fox.cs.vt.edu/>]
Computer Science Department
Virginia Tech
Blacksburg, VA 24061-0106 USA
E-Mail: fox@vt.edu
 - Charles R. Standridge
FAMU-FSU College of Engineering
Tallahassee, FL 32310, USA
E-Mail: stand@evax11.eng.fsu.edu
 - Stephen Williams [URL: <http://csgrad.cs.vt.edu/~williams/>]
Computer Science Department
Virginia Tech
Blacksburg, VA 24061-0106 USA
E-Mail: williams@csgrad.cs.vt.edu
-

mm-96 Chart 6



mm-96 Chart 5



SiteSearch

[Search for specific topics](#)

[Go to OCLC Home Page](#)

- [Demonstration](#)
 - [Descriptions of Products and Services](#)
 - [News](#)
 - [Publications](#)
 - [User Documentation](#)
-

Demonstration

- The WebZ Demo Page lets you search OPAC data and several other reference databases (<http://tikal.dev.oclc.org:2000>)

Descriptions of Products and Services

- [Introducing OCLC SiteSearch: To the Next Stage of the Electronic Library](#)
- [Elsevier Science/OCLC Electronic Publishing Pilot Program](#)
- [WebZ Server Questions & Answers](#)
- [Z39.50 Server System Questions and Answers](#)

News

OCLC Newsletter Features

- [Georgia's GALILEO Project](#)
- [Interview: Merryll Penson, Ralph E. Russell, and William Gray Potter](#). The directors of three libraries involved in the GALILEO project discuss the creation of the statewide project, its current status and future plans

News Releases

- ['GALILEO' to Use OCLC SiteSearch Software to Deliver Information, FirstSearch to Georgia Libraries--December 1, 1995](#)

See the [complete news release list](#) for earlier news releases.

Publications

Reference News

- [Winter 1996, No. 29](#)
- [Fall 1995, No. 28](#)
- [Summer 1995, No. 27](#)
- [Spring 1995, No. 26](#)
- [December 1994, No. 25](#)

Z39.50 resources - a pointer page

The Library of Congress is the official maintenance agency for Z39.50. As such they are the place to go to get the most official current legal information related to Z39.50. This page you are reading may phase out (though not soon) as they develop their page (started July 1995).

This page is meant as a reference point for resources related to the Information Retrieval Service and Protocol standard, ANSI / NISO Z39.50. This standard was first successfully balloted in 1988; several companies implemented this standard or variants of this; but it did not develop large scale acceptance. A noteworthy implementation based on this standard is WAIS (Wide Area Information Services). Also see the **Profiles** section for more info on present development of WAIS within Z39.50.

The standard was significantly rewritten for its next version. This is ANSI/NISO Z39.50-1992 (Version 2). One important step in this version of the standard was alignment with ISO 10162/10163, the Search and Retrieval (SR) Service Definition and Protocol Definition. Also beginning with this version, the protocol data units are described in ASN.1 (A "Layman's Guide" to ASN.1 is available from RSA) -- The Version 3 ASN.1 is available as flat ascii as well as in a wonderfully useful HTML format. from Library of Congress's various servers.

The next version (Version 3) of the standard was balloted in December 1994, and officially accepted by ANSI in July 1995. The official version of the standard is available electronically, at the Library of Congress's ftp server (ftp.loc.gov). Note this is a copyrighted document - many thanks to whoever achieved this electronic availability. The official text is available in postScript and wordPerfect, in four parts:

postscript: Part1, Part2, Part3, and Part4.

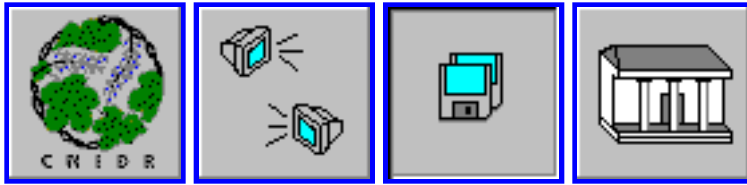
WordPerfect: Part1, Part2, Part3, and Part4.

The Z39.50 ImplementorsGroup (ZIG) works closely with the standard's maintenance agency, the Library of Congress. This group meets 2 - 3 times a year and has discussions on its listserv Z3950IW@NERVM.NERDC.UFL.EDU. For meeting minutes, more about the LISTSERV, scheduled future meetings, and other related information check out the relevant sub-section at Library of Congress

Freely available implementations of Z39.50 and related code are starting to become available. Those I know of (let me know of others) are:

- CNIDR's Isite, Isearch, FreeWAIS, etc
- Index Data, a software development enterprise operating out of Copenhagen, Denmark has developed a Version 3 API toolkit to aid in the implementation of the ISO SR and Z39.50-1995 protocols. They say: "software is available free of charge, on a liberal license: Commercial re-use is explicitly permitted."
- National Library of Canada has made its client and server code available;
- NIST is making available a Z39.50 client/server package based on the PRISE search engines.
- OCLC has made its Z39.50 Client API available to the public
- University of California - Berkeley demonstration client/server protocol engine
- USGS is making available a freeware implementation of Z39.50 as an OLE add-on to WWW browsers. You can fetch the executable software, README.TXT, and source files by anonymous FTP to host www.usgs.gov, in the directory /gils/ciir/dtic_a02.
- Willow -- the Washington Information Looker-upper Layered Over Windows.
- John Lamp is doing a good job tracking sites with Z39.50 tools and resources.

Electronic documents of interest (let me know of more) are:



CNIDR Isite

CNIDR Isite is an integrated Internet publishing software package including a text indexer, a search engine and Z39.50 communication tools to access databases. Isite includes the CNIDR ZDist, Isearch and Search API distributions.

See what Isite can do for you

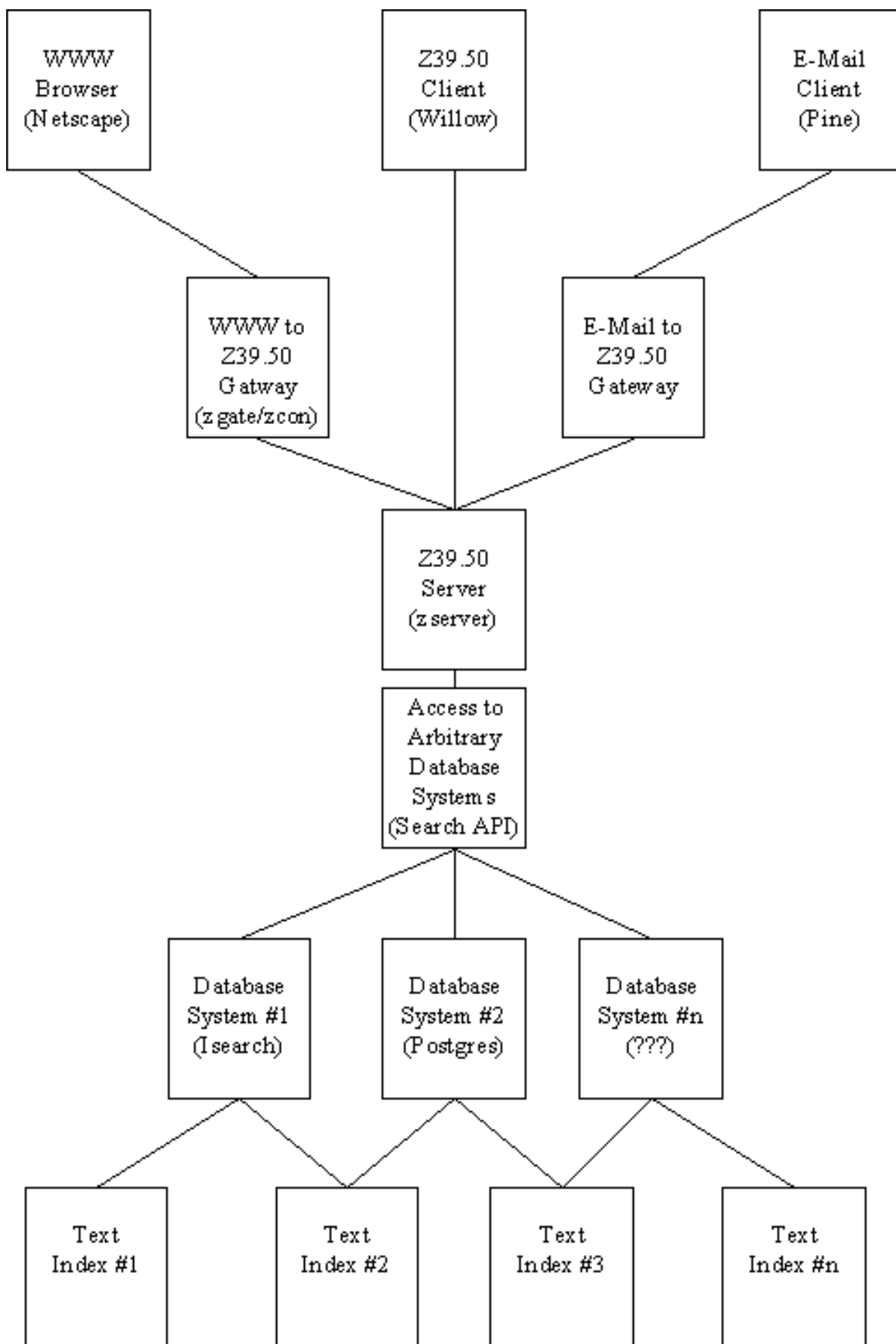
- **Help us to better serve you!**
- [Diagram of Overall Architecture](#) - Details available via Administrator's Guide below
- [Demo of Stateful http to Z39.50 Gateway](#) - Demonstrates access to various database systems
- Other systems using Isite
 - [NASA Global Change Master Directory](#)
 - [Z39.50 Ranked Search](#)
 - [Z39.50 Boolean Search](#)
 - [Distributed Document Search](#)
 - [American Astronomical Society: Electronic Astrophysical Journal Letters](#)
 - [United Nations International Drug Control Programme](#)
 - [University of Tennessee Office of Research Services: Friends and Partners Cookbook](#)
 - [Microlytics, Inc.](#)
 - [Library of Congress Z39.50 Gateway](#)
 - [U.S. Department of Housing and Urban Development GILS Service](#)
 - **YOUR LINK GOES HERE** - *Please send me pointers to your Isite-based systems!!*

Download a copy

- [Stable Version](#) - includes precompiled binaries
- [Untested Versions](#) - require a C++ compiler

Read the documentation

- [Isite Administrator's Guide](#) - Refers to stable versions
- [Untested Isite Administrator's Guide](#) - Refers to untested versions
- [Isearch Tutorial](#) - Step-by-step guide on building databases with Isearch
- [Z39.50 Maintenance Agency](#) - Everything you always wanted to know about Z39.50 and more!
 - Includes electronic copies of the ANSI/NISO Z39.50 standard
 - Includes implementor agreements
 - Includes various papers written by experts in the field
 - Includes lots of other stuff you will need to get the most out of Isite
- [BSn Doctypes](#) - Many of the input files supported by the Isearch indexer are documented here



[text-only]

The LIBRARY *of* CONGRESS

AMERICAN MEMORY

Historical Collections for the National Digital Library

SEARCH

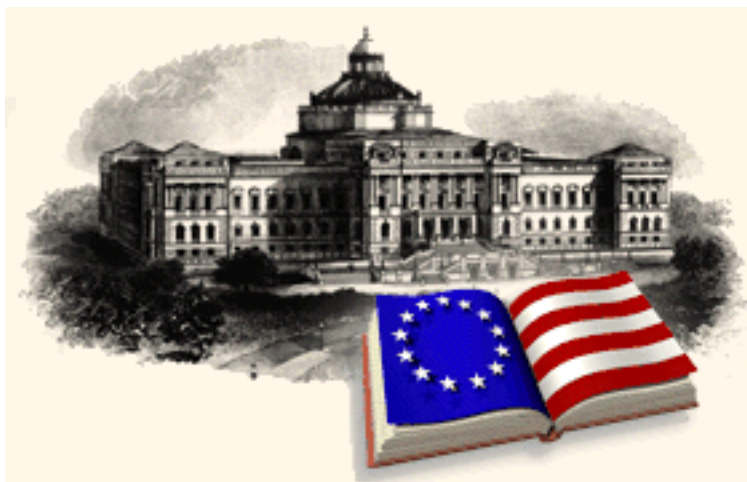
American Memory Collections

BROWSE

List of all American Memory Collections

LEARN

Organized help for using the collections



American Memory consists of primary source and archival materials relating to American culture and history. These *historical collections* are the key contribution of the Library of Congress to the National Digital Library. Most of these offerings are from the Library's unparalleled special collections.

Access Collections by Type



Prints & Photos



Documents



Motion Pictures



Sound Recordings

S h o w c a s e

Three new collections:

Evolution of the Conservation Movement, 1850-1920 (manuscripts, legal documents, photographs)

Gottschow-Schleisner (photographs)

Horydczak (photographs)



Introduction

Announcing the National Digital Library Competition

Summarized Project Guidelines

Awards

Application Process

Evaluation of Proposals

For More Information

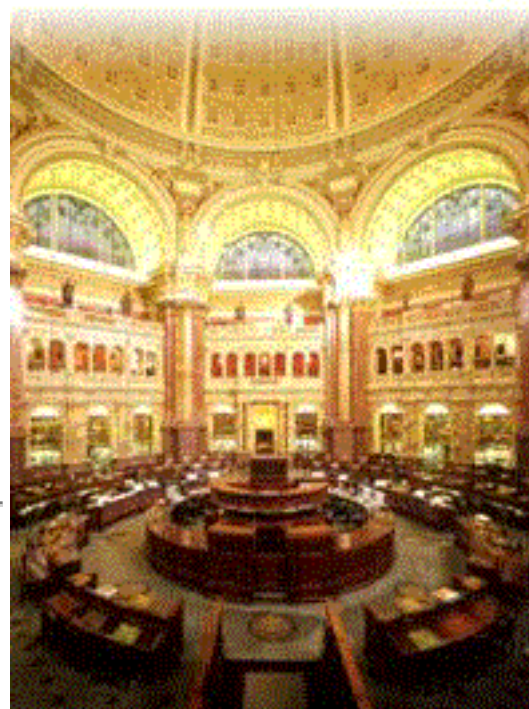
Includes a recommended reading list

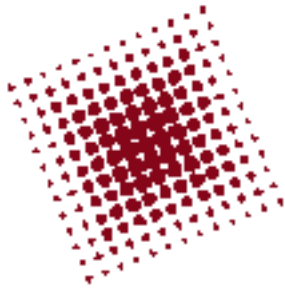


[The Library of Congress Home Page](#)

Library of Congress

Comments: lcweb@loc.gov (07/03/96)





Corporation for National Research Initiatives

Key Architectural Issues in The Digital Library

William Y. Arms

Acknowledgments

- This is work in progress.
 - This is a personal interpretation of ideas developed by the CSTR Project.
 - CSTR is a joint project of CNRI with Carnegie Mellon, Cornell, MIT, Stanford and UC Berkeley, funded by ARPA.
 - For background information, see the [CSTR home page](#).
 - The architecture is more fully described in a [paper by Robert Kahn and Robert Wilensky](#).
-

Key Issues and CSTR Terminology

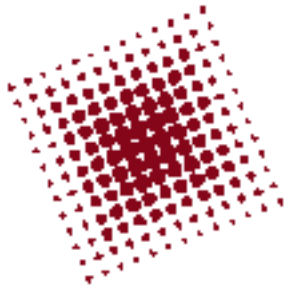
This set of WWW pages looks at the following six key issues in the architecture of the digital library.

- Items in the library - [digital object](#).
- Identifiers - [handle](#).
- Storage - [repository](#).
- Sets of objects - [composite and meta-object](#).
- Information about objects - [properties](#).
- Semantic layering (schema) - [data model](#).

The architecture under development is an open architecture. In general, it allows these topics to be considered separately.

The CSTR Architecture and the World Wide Web

Many of the concepts in the CSTR architecture can be partially implemented within the framework of the World Wide Web and fit with recent IETF discussions.



Corporation for National
Research Initiatives

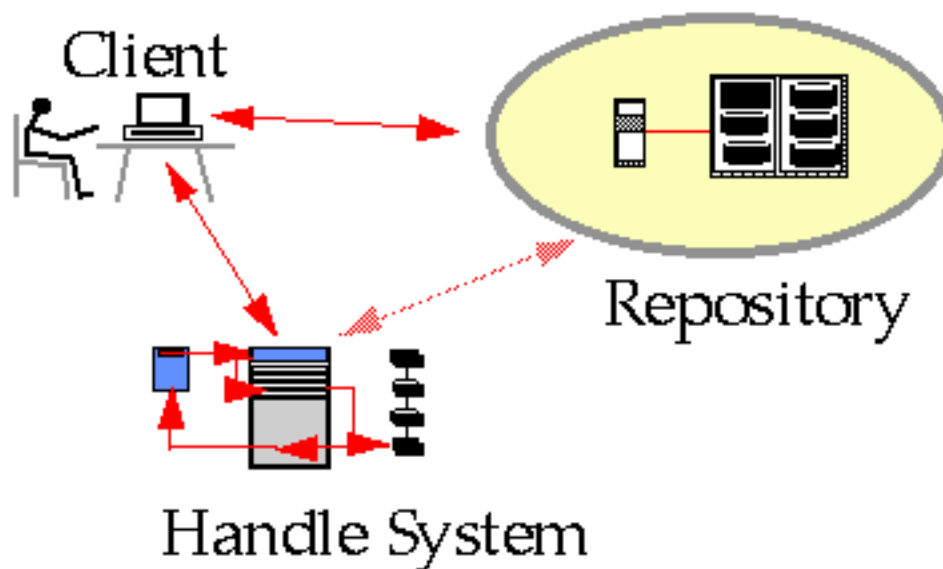
Digital Object Architecture Project

Principal investigators

Robert E. Kahn
William Y. Arms

Summary of the project

This project continues the architectural work of the DARPA-funded Computer Science Technical Reports (CS-TR) project. That project developed a Framework for Distributed Digital Object Services and implemented some key components. This project continues research and development of this framework and two extensive testbeds at the Library of Congress.



The basic entity in the system architecture is the "digital object", which contains copyright material or other material in which other rights and interests are manifest. There may also be rights and interests associated with digital objects themselves. The major components of the system are: (a) repositories of digital objects that allow network based deposit and access, (b) handle servers that record the location of digital objects over long periods of time, (c) registration and recordation mechanisms to keep track of rights and interests associated with digital objects, and (d) client software to enable use of these components over the network. Digital object

fingerprints are used in the registration system to permit validation of the objects at a later time.

The first testbed is with the Copyright Office at the Library of Congress. This is a system to register electronic materials for copyright and recordation of changes in copyright ownership. The second testbed is with the National Digital Library Program at the Library of Congress. This is a very large scale project to convert historic materials from the library's collections to digital form and make them available to the world.

Background papers

- A Framework for Distributed Digital Object Services by Robert Kahn and Robert Wilensky, May 1995
 - Key Concepts in the Architecture of the Digital Library by William Y. Arms, D-Lib Magazine, July 1995
 - "Implementation Issues in an Open Architecture Framework for Digital Object Services" by Carl Lagoze and David Ely. Cornell Computer Science Technical Report TR95-1540
 - "A Design for Inter-Operable Secure Object Stores (ISOS)" by Carl Lagoze, Robert McGrath, Ed Overly, Nancy Yeager. Cornell Computer Science Technical Report TR95-1558
 - Uniform Resource Names: A Progress Report The URN Implementors, D-Lib Magazine, February 1996
 - Historical Collections for the National Digital Library: Lessons and Challenges at the Library of Congress Caroline R. Arms, D-Lib Magazine, April 1996. Part 2
-

Funding

Funding for this work is provided by the Defense Advanced Research Projects Agency (DARPA) and the Library of Congress.



wya
6/30/96



A PURL is a **P**ersistent **U**niform **R**esource **L**ocator. Functionally, a PURL is a URL. However, instead of pointing directly to the location of an Internet resource, a PURL points to an intermediate resolution service. The PURL resolution service associates the PURL with the actual URL and returns that URL to the client. The client can then complete the URL transaction in the normal fashion. In Web parlance, this is a standard HTTP *redirect*.

The OCLC PURL Service has been strongly influenced by the active participation of OCLC's Office of Research in the IETF Uniform Resource Identifier working groups. There is nothing incompatible between PURLs and the ongoing URN work. PURLs satisfy many of the requirements of URNs using currently deployed technologies and can be transitioned smoothly into a URN architecture once it is deployed.

Further Information and Resources

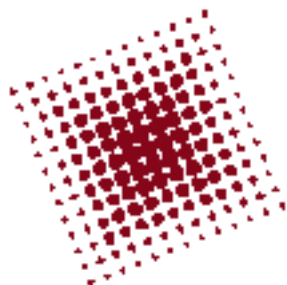
- A brief introduction to PURLs
- A longer introduction to PURLs
- Frequently Asked Questions
- Download the PURL software **NEW**
- PURL-L mailing list
- More info

Interacting with This Resolver

- Create your first PURL
- Register as a user
- Create PURLs, domains, groups
- Modify PURLs, domains, groups, users
- Search this resolver
- Power user's page (all features)

As of *Sat Jul 13 13:26:28 PDT 1996* : PURLs Created = **6768** , PURLs Resolved = **473905** and Unique Client Systems = **13121** (See the complete Database Stats for more details.)

The PURL Team
purl@oclc.org



Corporation for National
Research Initiatives

Handles and the Handle System

Forms for handle administration

Forms to add and edit handles, to create naming authorities, and to set up groups of administrators are available through the [Handle Administration Page](#).

Information about the Handle System

Technical information

- [An overview of the system.](#)
- [Implementation of the Handle Management System.](#)
- [FTP server](#) to download documentation and code.
- [Browsers](#) that support handles.

Architectural considerations

- The use of handles within a [framework for distributed digital object services](#).
- Handles as a [key concept in the digital library](#).
- The IETF's work on [Universal Reference Names](#).

Presentations and demonstrations

- [An architectural overview.](#)
 - [The handle system.](#)
 - [D-Lib Magazine](#) with handles.
-

A brief introduction to Handles

A **handle** is a unique identifier for a digital object. This object can be stored in a digital library repository, in an ftp archive, in a World Wide Web server, or any other digital store. Handles can also be used for other forms of identification, such as electronic mail addresses. A high performance Handle Management System is publicly available on the Internet. The useful properties of handles include the following.

- Handles are guaranteed to be unique.
- Handles are permanent. Therefore, they can be used to identify objects for purposes of copyright or archiving.
- Handles are location independent. The object may be moved to a different location without changing its handle. This enables handles to be used to refer to an object, for example, in a bibliographic citation.

A handle has the syntax:

naming authority / string

or: *hdl://naming authority / string*

The **naming authority** is a globally unique name. The **string** is unique for that naming authority.

[Return to CNRI home page](#)

hdl://cnri/handle-intro

wya

Last revised: November 11, 95

UMBC

AgentWeb

UMBC

An Honors University in Maryland

Laboratory for Advanced Information Technology



UMBC AgentWeb

Intelligent Software Agents



[UMBC LAIT](#) | [AgentWeb](#) | [NEW!](#) | [AgentNews](#) | [KQML](#) | [Search](#) | [Help](#)

Information and resources about intelligent information agents, intentional agents, software agents, softbots, knowbots, infobots, etc. Send comments and suggestions to [Tim Finin \(finin@umbc.edu\)](mailto:finin@umbc.edu).

- **About the AgentWeb...**

- [What's new...](#) **NEW**
- [Current AgentNews webletter](#) **NEW**
- [About the AgentNews webletter and mailing lists](#)
- [AgentWeb help...](#)
- [AgentWeb salon ...](#) **NEW**
- [About the UMBC Laboratory for Advanced Information Technology](#)

- **Agent basics ...**

- [Introductory material](#)
- [Agent FAQ](#)
- [Agent theory - philosophy, formalisms, ...](#)
- [Agent technology - systems, tools, languages, standards, ...](#)
- [Mobile agents ...](#) **NEW**

- **Agent resources ...**

- [Agent papers](#)
- [Agent events, conferences, workshops, ...](#)
- [Agent mailing lists and newsgroups](#)
- [Agent courses and seminars](#)
- [Other agent related web resources](#)

- **Who is doing what ...**

- [Agent-related R&D groups and companies](#)
- [Agent-related projects](#)
- [Example Agents](#)

[Home](#)

[Search](#)

[Contents](#)

[News](#)

[Contacts](#)

Preserving Digital Information: Final Report and Recommendations

May 20, 1996

At the end of 1994 the Commission on Preservation and Access (CPA) and RLG created a Task Force on Archiving of Digital Information charged with investigating and recommending means to ensure "continued access indefinitely into the future of records stored in digital electronic form." The 21-member task force, co-chaired with distinction by Donald Waters, Associate University Librarian, Yale University, and John Garrett, Chief Executive Officer of CyberVillages Corporation, recently completed their final report. RLG and CPA are making this widely available online and in print.

Electronic versions are available from RLG's FTP server ([ftp.rlg.org](ftp://ftp.rlg.org)) and this Web site:

[HTML version](#)

[Adobe Acrobat version: /pub/archtf/final-report.pdf](#)

[Microsoft Word for Windows 6.0 version: /pub/archtf/final-report.doc](#)

[ASCII Rich Text Format version: /pub/archtf/final-report.rtf](#)

Notes:

To download an Adobe® Acrobat® viewer to use as a helper application with your web browser, connect to the [Adobe web site](#).

Copies of the printed, bound report are available for \$15.00 (prepayment required) from the Commission on Preservation and Access, 1400 16th Street, N.W., Suite 740, Washington, DC 20036-2217.

RLG will be mailing the printed report to the member representative at each of our [member institutions](#) in North America and Europe as well as to each member liaison in our collaborative [SHARES](#) (Shared Resources) and [PRESERV](#) (Preservation) programs.

The task force's final report benefits from their action last September to make a draft version available online and to open a listserv for comments by the community. Many thanks to all of you who responded. That [draft report](#) can still be found on RLG's server and Web site:

[Adobe Acrobat version: /pub/ArchTF/Draft-Report.pdf](#)

[Microsoft Word for Windows 6.0 version: /pub/ArchTF/Draft-Report.doc](#)

[ASCII version: /pub/ArchTF/Draft-Report.txt](#)

RLG has already built into its agenda work on several of the task force's nine recommendations. (Our [archival server](#) and [digital collections](#) projects are directly related.) We will be following up on other recommendations with other stakeholders.

Please share your comments and advice with us regarding this report and the specific recommendations; you can send them by e-mail to [Nancy Elkington](#), RLG member services officer and member of the task force.

Sincerely,

James Michalko
President

TEI Guidelines for Electronic Text Encoding and Interchange (P3)

Made available from the Electronic Text Center at the
University of Virginia.

Search the *TEI Guidelines*.

Word or phrase (omit all quotes):

Search

1-50

Other types of searches:

You may also combine words or phrases within a specified proximity, or locate segments such as sections where two words or phrases both occur.

Browse the *TEI Guidelines*.

- [Bibliographic header of the TEI Guidelines](#)
 - [Preface](#)
 - [Acknowledgments](#)
 - TEI Working Committees (1990-1993)
 - Advisory Board
 - Steering Committee Membership
 - [Changes from TEI P1 to TEI P3](#)
 - [Part 1: Introduction](#)
 - [Part 2: Core Tags and General Rules](#)
 - [Part 3: Base Tag Sets](#)
 - [Part 4: Additional Tag Sets](#)
 - [Part 5: Auxiliary Document Types](#)
 - [Part 6: Technical Topics](#)
 - [Part 7: Alphabetical Reference List of Tags and Attributes](#)
 - [Part 8: Reference Material](#)
-

Resources of Related Interest

- [The Text Encoding Initiative Home Page](#)
- [Other Electronic Versions of the TEI Guidelines](#)
- [TEI P3 now available on CD-ROM](#)
- [The Electronic Text Center Introduction to TEI and Guide to Document Preparation.](#)
- [TEI DTD Browser](#), courtesy of CETH

Digital Libraries - Implementation Principles

As we build digital libraries, it is important to consider key principles so that these libraries will be easily usable, and have long-term archival value.

1. Declarative representations of documents should be used.
2. Document components should be represented using natural forms, namely objects that can be manipulated by users familiar with those objects.
3. Links should be recorded, preserved, organized and generalized.
4. There should be a separation between the digital library and user interfaces to it.
5. Searching should make use of advanced retrieval methods.
6. Open systems that include the user, and where (some of) the functions of librarians are carried out by the computer, must be developed.
7. Task-oriented access to electronic archives must be supported.
8. A user-centered development approach should be adopted.
9. Users should work with objects at the right level of generality.

Electronic Thesis and Dissertation Project

[http://etd.vt.edu/etd/
etd@vt.edu](http://etd.vt.edu/etd/etd@vt.edu)

For a good overview, see the Sept. *D-Lib Magazine* coverage:

National Digital Library of Theses and Dissertations (NDLTD)
A Scalable and Sustainable Approach to Unlock University Resources
<http://www.dlib.org/dlib/september96/theses/09fox.html>

Purpose:

Our project is primarily an effort to improve graduate education. We will work so that graduate students become information literate, learning how to become electronic publishers and knowing how to use digital libraries in their research. True success in these projects will potentially mean a permanent change in graduate education and scholarly publishing, with digital libraries playing a more dominant role in supporting and disseminating research.

History:

- * 1987: Discussion about SGML for dissertations in Ann Arbor, Michigan
- * 1992: Project discovery meeting led by Coalition for Networked Information, UMI, Council of Graduate Schools(CGS), and Virginia Tech
- * 1994: Monticello Electronic Library meeting at Virginia Tech decided to use PDF and SGML (with templates and conversion tools)
- * 1996: Southeastern Universities Research Association (SURA) funds pilot project in South, led by team at Virginia Tech
- * 1996-99: U.S. Department of Education's Fund for the Improvement of Post Secondary Education supports NDLTD along with:

ACM, Adobe, ArborText, Council of Graduate Schools, Coalition for Networked Information, Cornell Digital Library Research Group, Council of Southern Graduate Schools, IBM, Microsoft, OCLC, State Council of Higher Education for Virginia, SOLINET, SURA, UMI, and University of Utah Graduate School

Invitation:

Please investigate given the pointers above, or contact Ed Fox (fox@vt.edu) for more information. We welcome every institution of higher education to participate, so ultimately every thesis, dissertation (and related reports) will become a part of this comprehensive digital library of research results.