
Anderson, Greg, Rebecca Lasher, and Vicky Reich. "The Computer Science Technical Report (CS-TR) Project: A Pioneering Digital Library Project Viewed from a Library Perspective." The Public-Access Computer Systems Review 7, no. 2 (1996): 6-26. (Refereed Article)

1.0 Overview

Be favorable to bold beginnings.
--Virgil

In 1992, the Advanced Research Projects Agency (ARPA) awarded a three-year grant to the Corporation for National Research Initiatives (CNRI) and five research universities to build a large-scale, distributed digital library of computer science technical reports produced by project participants. The participating universities were Carnegie Mellon University, Cornell University, the Massachusetts Institute of Technology, Stanford University, and the University of California at Berkeley. CNRI served as a collaborator and agent for the project.

The Computer Science Technical Reports (CS-TR) project was one of the earliest sustained investigations into the system engineering of digital libraries, and it pioneered multi-institutional collaborative research in this increasingly important area. The CS-TR project investigated a broad spectrum of technical, social, and legal issues related to the development and implementation a very large, heterogeneous, distributed digital library.

The project's main accomplishments can be summarized as follows:

1. The most enduring accomplishments were the mutual respect and the research partnership that developed between the computer scientists and the librarians who worked together to investigate digital library issues.
2. The project created a prototype digital library service that included a large collection of technical reports; an exchange format for bibliographic data (RFC 1357, which was superseded by RFC 1807); a distributed delivery protocol (Dienst) for information on the World-Wide Web; an information awareness service (Sift); an approach to interoperability (the Kahn and Wilensky paper); and a Web catalog tool (Lycos).

3. The critical issues associated with the evolving concept of digital libraries were articulated through practice and deeper research into key issues.

2.0 Project Planning

CS-TR project planning began in 1990 with discussions among staff

from the participating institutions. Computer science technical reports are an important body of knowledge; however, they are often difficult to locate because they are normally published by academic/research departments. The original question posed for the project was straightforward: how can we make computer science technical reports more accessible to researchers? Project participants initially believed that the intellectual property issues associated with distributing the technical reports were not terribly complex.

As a result of these early discussions, a variety of broader issues were identified, such as:

- o How do we build technologies that make scholarship more effective?
- o What do we really mean by a digital library?

This more comprehensive view was presented to potential funding agencies. ARPA funding was secured in 1992, and CNRI assumed the role of contract administrator. At this stage, it was apparent that the project had the potential to set the pace for several important aspects of the digital library: distributed, virtual collections spread across the network; sophisticated linking mechanisms that would enable the location and retrieval of information no matter where it was located; tools to handle intellectual property issues; and identification and resolution of service and scholarly productivity issues.

The consortial arrangement of the project enabled each participating institution to pursue separate, but linked, approaches to these issues. Each of the five participants placed its own technical reports online at its site. Through network-based searching and retrieval mechanisms, the project explored the issues involved in sharing, rather than duplicating, online information.

+ Page 8 +

The research goals of the project varied with each participant. In "A Proposal for MIT Participation in an Electronic Library Plan" most of the key points involving technical, organizational, service, and data questions were enumerated:

1. To obtain early experience with a core function of the distributed electronic library of the future.
2. To work with a database that is readily available, has a critical time-sensitive value, and is already well-known and valued by its target audience.
3. To explore the architecture, design, and workflow issues associated with making information available in digital form.
4. To work within the research/prototype domain with a volume of information large enough to be useful, interesting, and scalable to an operational system.
5. To provide an important service to an audience of researchers, faculty, and students who are motivated

and likely to have access to appropriately powerful workstations to use the library from their offices. [1]

Each campus pursued its own research questions within the framework of these common goals. CNRI led the coordination, discussion, and facilitation of the individual efforts and contributed its own research on linking mechanisms and electronic copyright management.

3.0 Design and Development Issues

The project's core design was based upon the construction of a bibliographic records database that described the technical reports and provided links to the page-image representations of the reports. In addition to images, the project obtained the full text of the technical reports from either the reports' source files or OCR conversions. Using this full-text information, the project evaluated different retrieval mechanisms; explored data integrity issues for huge stores of data; and developed citation linking strategies for references across documents (e.g., a link from a footnote or citation in one document to the cited document itself). [2]

+ Page 9 +

3.1 Bibliographic Record Format

Many computer science R&D organizations routinely announce new technical reports by mailing (via the postal service) the bibliographic records for these reports. These bibliographic records are usually produced by secretaries or publications coordinators. This paper alert service has some obvious drawbacks: mailing costs, postal delays, and an inflexible format that is not amenable to convenient filing for later retrieval. The CS-TR project participants wanted to shift to electronic bibliographic records distribution; however, in order to do so, they needed to use the same bibliographic record exchange format.

The project participants wanted a format that was simple (for people and for machines), easy to read, and easy to create. It was recognized that this was likely to be an interim format, because automatic and full-text indexing methods could supersede bibliographic records.

Early in the project, use of the USMARC format was considered and discarded. USMARC is very complex, not easily taught, and not accepted by non-catalogers. Project staff were concerned that the complexity and the high level of training necessary to catalog in USMARC could cause significant time delays between report publication and bibliographic record creation. For the CS-TR project, the possibility of a delay was unacceptable.

The BibTeX and Refer formats were also considered and rejected. Neither had the required computer science technical report fields (e.g., Computing Reviews category, monitoring, funding, contract organizations, and grant number).

The project participants created their own bibliographic format: "RFC 1357, A Format for Mailing Bibliographic Records" (this format was subsequently superseded by RFC 1807, "A Format for Bibliographic Records"). The basic design principles of the RFC

1357/RFC 1807 formats were the:

1. Identification and creation of data elements needed for citation creation, management, and retrieval.
2. Creation of bibliographic records that coincided as closely as possible with the publication of the technical report.
3. Creation of RFC 1357/RFC 1807 records via machine parsing of the report's title page data and/or by staff in the participating computer science departments (library catalogers would not be needed).

+ Page 10 +

4. Provision of core information for more formal library bibliographic records. Berkeley, Stanford, and Cornell built translators that map the Project's record formats into other formats, including USMARC.

Once the project participants decided to create a new bibliographic format, the development and implementation of the format proceeded quickly. The project was not constrained by older formats and could add fields as desired.

Project participants came to agreement on name authority conventions for institutions; however, use of AACR2 was never discussed as a tool for bibliographic description.

3.2 Centralized Versus Distributed Indexes

Once the bibliographic record format was created, the project considered the issue of centralized versus distributed indexes. Project participants had long discussions where they argued the virtues, value, and scalability of centralized and/or decentralized indexes for very large distributed collections.

One of the early goals of the project was to develop an interoperable, distributed collection that would allow each site to develop its own testbed architecture, create consistent content based on the TIFF-B standard, experiment with interoperable systems, and share digitized technical reports across different systems. In the end, no conclusions were reached, and the above goal was not met.

The project participants recognized that neither centralized nor decentralized servers would scale-up well. Eventually a more complicated, yet to be determined, architecture could emerge that would involve replication of an institution's indexes on several servers around the country.

In order to get started, Cornell developed Dienst--a protocol and an operational system that provided Internet access to the project's distributed collections. Indexes were produced and kept at each institution. Each institution was required to run the Dienst server protocol. Dienst did permit a "single distributed collection model," but it was not an interoperable model running on different software and server platforms. [3] Some institutions implemented a full-text searching capability limited to that site's reports.

There were four classes of Dienst services:

- o A Repository Service stored digital documents, each of which had a unique name and could exist in several different formats.
- o An Index Service server searched a collection and returned a list of documents that matched the search.
- o A single, centralized Meta Service (also called a Contact Service) provided a directory of locations of all other services.
- o A User Interface Service mediated human access to the digital library.

A group of sites sharing the Dienst protocol formed a single distributed collection. Each site typically ran Repository, Index, and User Interface Services for documents issued by that site. One site ran a Meta Service, which defined the set of sites that make up the collection.

Davis et al. describe Dienst as follows:

From the standpoint of a Dienst user, a document collection consists of a unified space of uniquely identified documents, each of which may be available in a variety of formats. Using publicly available World Wide Web clients, users may search the collection, browse and read individual documents in any of their available formats, and download or print a document. [4]

With the Dienst system, users could query all or selected institutions using combinations of keywords in fields (e.g., author and title). The search was performed in parallel at user-selected sites. If a server was unavailable, the search would time-out and display a message to the user that the server was down.

Davis et al. indicate that "further work needs to be done in two areas: begin replicating index servers to increase availability and response time; add persistent search which continues to attempt to contact non-responsive sites." [5]

3.3 Technical Report File Format

The pros and cons of a standardized technical report file format (e.g., images, SGML, PostScript, and ASCII) was vigorously debated. The TIFF-B image format (also called Group IV fax compression in TIFF format) was selected as the project standard. This decision was supported by the following factors: (1) in 1992, image formats were standard and many commercial image software packages were available on multiple platforms; (2) retrospective paper reports could be easily converted to the image format; (3) project participants were eager to populate servers with both retrospective and prospective reports; and (4)

researchers did not want to engage in document markup, convert documents, or develop new standards.

Some project members believed (and continue to believe) that image files were the ultimate version of record, because they provided the simplest exact representation of the document and could be exported to new software and platforms over time.

Many of the participating institutions made multiple file formats available on their servers. All formats were available through the Dienst protocol. Use of the TIFF-B format was a requirement for the project, but most institutions also offered PostScript and ASCII files (particularly for the newer reports).

3.4 Scanning and OCR

Project participants conducted an in-depth investigation of scanning and OCR hardware and software. Although there was no dpi requirement, the project participants agreed to scan pages at 300 dpi or greater because use of a lower resolution might require rescanning as more sophisticated systems were developed. Each institution purchased different equipment and software. As long as TIFF-B image files were produced, project participants did not need to use the same equipment. In fact, the project encouraged different scanning and OCR implementations.

MIT conducted the most in-depth research on the high-volume production, archival, and record keeping aspects of the scanning process. The MIT Library 2000 testbed effort focused significant attention on production scanning.

+ Page 13 +

This emphasis was based upon the hypotheses that scanned images of documents will be an important component of any future electronic environment. At its core, the digital library must contain high-quality content, and, for the foreseeable future, much of that content will come from the conversion of paper-format information to scanned images. The creation of a large corpus of quality information provides the testbed content for investigations into system architecture, electronic information management, retrieval, and long-term storage issues.

Basic principles of the MIT scanning effort included:

1. Materials should only be handled once. The design of the scanning environment should strive to achieve the greatest advantage in terms of price, performance, and quality. Libraries and publishers cannot afford to rescan materials as technological capability increases. To limit potential damage to the original paper artifact, scanning once is preferable. To adhere to these principles, good paper workflow, management, and content selection are important.
2. As much information as possible should be captured in a single scan. Although current technology cannot exploit all of the bits captured, future technologies will be able to do so. The MIT scanners were capable of a resolution of 400 pixels per inch, with eight bits of gray-scale per pixel. This created very large files

(about 16 MB per scanned page), which were rendered down to the agreed-upon interchange format for the project: 300 dpi, one bit per pixel, in TIFF-B format.

3. Quality control is critical. In order to achieve the first two principles, quality control methods must assure a high degree of integrity and confidence in the production environment. The MIT Libraries' Document Services department adapted procedures from its microreproduction heritage for this new production scanning effort. Document Services was using test targets from the Association for Information and Image Management (AIIM) and the Institute of Electrical and Electronic Engineers (IEEE) to test calibrations on the scanner. Quality control was checked via file checksums and visual review of selected images.

+ Page 14 +

4. Context of the images is important now and in the future. Because the underlying technologies will change and improve in the future, the scanned images must provide enough context for humans and machines to understand both their content and structure in order to use them effectively. The MIT scanning effort created a metadata record to provide information about the scanned document and the environment in which it was created. This record specified both the form and content of the information that must be captured when a document is scanned, and it became a component of the scanned form of the document. The record assisted in viewing, displaying, or printing the image correctly; in understanding how to interpret the image; and in meeting contractual or legal requirements.

3.5 Distributed Digital Object Services

The most important design issue for the CS-TR project was to determine an appropriate infrastructure and architecture for a large distributed digital library. The outcome of the lengthy discussions of this issue is captured in a paper by Kahn and Wilensky:

This document describes fundamental aspects of an infrastructure that is open in its architecture and which supports a large and extensible class of distributed digital information services. Digital libraries are one example of such services; numerous other examples of such services may be found in emerging electronic commerce applications. Here we define basic entities to be found in such a system, in which information in the form of "digital objects" is stored, accessed, disseminated and managed. We provide naming conventions for identifying and locating digital objects, describe a service for using object names to locate and disseminate objects, and provide elements of an access protocol. [6]

The most important concept in the Kahn and Wilensky paper is the creation of the "handle" concept, which seeks to separate document naming issues from network address issues. Handles are not URLs; handles are an approach to a large-scale problem of

naming objects that may change location over time. A handle is a unique, permanent identifier for a document, and it is used to name the document on a server. A mechanism called a "handle server" maps the handle to the document's real network address. A working prototype of the handle server is available at CNRI, and handle functionality is being integrated into World-Wide Web browsers, such as Netscape.

+ Page 15 +

In the future, a Web browser will send a message to a handle server that gives the handle for the desired document. The handle server will send the Web browser the actual network address of the target document, which the browser will then retrieve. Handles and handle servers will be very powerful tools for digital libraries. No longer will Web servers contain false links, because handle servers can update documents' network addresses on a nightly basis.

For libraries to move beyond their physical walls (and campus boundaries) and to leverage the power of the distributed information base of the network to enrich services for their local community of users, a basic architecture for naming, locating, and accessing network information must be well-understood and adopted. The handle concept accomplishes this important goal.

3.6 Copyright

Copyright is a key issue in building digital libraries. At the beginning of this project, participants assumed that there would be few (or no) copyright issues associated with distributing computer science technical reports. They assumed that the reports published at their schools were either in the public domain or that the rights were held by the publishing university. Later, as copyright questions arose, the project participants assumed that a single strategy would work for every institution. These assumptions proved to be naive. Upon investigation with legal counsel, researchers discovered that each school had different intellectual property policies, and, consequently, five different approaches to the copyright issue evolved.

At Stanford, librarians took on the role of ensuring that these copyright issues did not pose a risk to the university or to the faculty. Librarians identified scenarios that needed attention, and they began to meet with legal counsel to determine appropriate responses. These efforts helped them to articulate a set of copyright guidelines now used by the CS-TR projects at Stanford and Cornell.

+ Page 16 +

The major findings and recommendations of the Stanford guidelines are presented below. Other institutions may find this information helpful; however, they should not view it as legal advice. The worldwide legal environment is undergoing rapid change, and the project's approach may become obsolete in the face of new laws and treaties.

- o At most U.S. academic institutions, the author owns the copyright to any books, articles, or technical reports.

Works published prior to March 1989 without a copyright notice are in the public domain (unless steps were taken within 5 years to establish copyright ownership). Works, with or without a copyright notice, published after March 1989 are copyrighted.

- o In most cases, reports that are produced (or report on work sponsored by) the government are not in the public domain. The government can make copies, but, at most institutions, the author owns the rights.
- o Most CS-TR project institutions ask authors to sign a form granting the institution nonexclusive, revocable, royalty-free license to publish, perform, display, and distribute the works. One author's signature is binding on multiple-author works.
- o If an author has signed or plans to sign an exclusive agreement with a publisher for a particular work (or for substantially the same work) in a particular format, that author cannot then sign a nonexclusive agreement with the institution for the same work in the same format.
- o If an author signs a nonexclusive agreement with an institution for a technical report and then decides to publish the same work elsewhere, the author should inform the publisher of this previous agreement. The author should then grant the institution written permission for the nonexclusive rights to publish, perform, and display the works before any works are loaded on that institution's servers. If the author indicates he or she has already signed an exclusive agreement with a publisher, the technical report should probably not be mounted on the server without permission of that publisher.

+ Page 17 +

- o At some institutions, the authors do not own the rights to their works. Each institution should be clear about copyright ownership before mounting technical reports on servers. The CS-TR project did not address the issue of third-party rights in technical reports. When authors sign agreements it is assumed that the entire work is original or that the author has the rights to include non-original tables, charts, and figures. This is one area that could be pursued by asking authors specifically about the originality of their works.

There are several ways to manage technical reports that are submitted to publishers as articles.

- o Ask the authors not to sign exclusive agreements with publishers. Ask them to modify the publisher's standard agreement to allow the institution to keep the work on a server.
- o Make special arrangements with the publishers so the technical reports can stay on the servers even if an article is published.

- o If the author requests, remove the technical report from the server and point to the printed article.
- o Include a notice with the technical reports to inform viewers of their rights (e.g., transmitting documents over the network and viewing them, which may be legally considered a "performance" of the work; making printed copies; distributing copies to others; and selling copies to others). This relieves the users of guessing what restrictions might apply. Most likely, the user will properly assume fair-use restrictions apply, view the work, and perhaps make a personal copy. But can the user legally send a copy to a colleague? Cornell has chosen to clarify these issues by explicitly sublicensing rights to the user. [7] This clarifies the user's rights to quote, redistribute, or make copies of the technical report. However, the sublicense must preserve the author's right to withdraw the technical report from further distribution.

+ Page 18 +

3.7 Tension Between Research and Operational Objectives

After a certain point in the CS-TR project's development, the project's prototype systems were used as both experimental and production services. The prototype systems that were available for public use changed constantly. This created a tension between providing reliable operational services while developing new experimental capabilities.

In the CS-TR project, librarians continuously examined the long-term viability of the effort. At each stage of the project, it was important to remember that the project was primarily conducting research and that digital libraries are in a nascent state. Whatever we built would be superseded by more powerful knowledge and services in the future.

Several public systems were implemented with support from the CS-TR project:

- o Dienst, a distributed search system for technical reports (Cornell). [8]
- o Mercury, a centralized search system for technical reports (Carnegie Mellon).
- o GLOSS, a system to help find relevant data sources (Stanford). [9]
- o SIFT, a system for performing wide-area information dissemination on USENET newsgroups and computer science technical reports (Stanford). [10]
- o Lycos, a searchable catalog of Internet resources (Carnegie Mellon). [11]
- o A handle server to maintain unique identifiers to objects in the digital library (CNRI). [12]

During the CS-TR project, these prototype systems were quite successful. The Lycos system had thousands of accesses every day. The Sift system had over 10,000 subscribers. Fourteen institutions used Dienst as a production system to disseminate their technical reports.

However, using prototype systems as production systems was challenging. Enhancements and changes to the Dienst system were problematic because the institutions using the system all had to implement the upgrades. In a similar fashion, changes to Lycos or Shift system affected the Internet users of these systems.

Today, many of the project's prototype systems have evolved into true production systems; however, they will continue to be used as testbeds for digital library experimentation and research. They offer an opportunity to examine a variety of new issues, such as the linkage of large-scale, distributed digital object collections; the cognitive efforts needed to identify and present coherent collections to users; and the effective integration and evaluation of services for all media, examining both content and user issues.

4.0 Collaboration in the CS-TR Project

The CS-TR project involved significant collaboration between the participating institutions. It also required extensive collaboration between librarians and computer scientists.

4.1 Collaboration Between Participating Institutions

As a result of many long discussions and compromises, the CS-TR project created systems that are more logical than they would have been without this collaborative effort. However, collaborations of this kind create tensions. Each institution was primarily funded to study specific areas of the overall digital library research domain. All of the participating institutions wanted to make their technical reports available on their servers as soon as possible so that their research could commence, and they wanted their prototype systems to reach the broadest possible audiences. While project participants had a common overall objective, the above considerations sometimes made multi-institutional collaboration a challenging endeavor.

4.2 Collaboration Between Librarians and Computer Scientists

If we accept that we are living in an information age and that a central challenge for this age is to give people tools with which they can successfully use networked information, then librarians and computer scientists are natural collaborators to address this challenge. Computer scientists and librarians each bring to the discussion complementary technical skills and perspectives. Computer scientists have a broad view of the network, new approaches to information retrieval, and an openness to change. Librarians have content expertise, responsibility for significant collections of scholarly material, a strong service orientation, and a historical commitment to the preservation of our intellectual heritage. Both communities share the academic

values of the open sharing of information and the desire to foster the creation of new knowledge.

From the inception of the CS-TR project, librarians worked closely with computer scientists. Both groups brought strengths to the project, and the cooperative results were superior to those that would have occurred if either group had conducted the project alone. Through ongoing discussions and consideration of common problems, such as the proposed handle mechanism, an atmosphere of trust and respect was created. The librarians benefitted from the computer scientists' cultural values of exploration and learning by doing. The computer scientists benefitted from the librarians' broad perspective and integrative skills. The mutual respect of these two groups for each other's professional knowledge and abilities created a productive, dynamic atmosphere.

For example, early in the design stage of the project, the development of bibliographic records for the technical reports was a key discussion topic. The computer scientists wanted a variety of departmental staff to be able to quickly and easily create bibliographic records. The librarians wanted consistent record content and the ability to make multiple uses of the record. The resultant record structure (RFC 1807) accommodated both sets of requirements in a sustainable, scalable manner. The records can be immediately created upon acceptance of the technical report by publishing assistants. The records have a consistent definition, and the use of record fields is well-understood. There are conversion routines to facilitate MARC record creation (or use of the record in other formats).

+ Page 21 +

Another example is the collaboration of staff in the MIT Libraries' Document Services department with researchers in the MIT Laboratory for Computer Science's Library 2000 project to create an operational scanning service. This collaboration resulted in other opportunities for joint work on scanning issues.

The collaborative efforts of librarians and computer scientists created mutual respect that will continue to bear fruit long after the CS-TR project's termination.

5.0 Expanding the CS-TR Project

At the June 1995 CS-TR meeting, the project participants agreed to ask the Computing Research Association (CRA) to endorse and to encourage the dissemination of this technology. A new consortium effort called Networked Computer Science Technical Report Library (NCSTRL) was created to merge the CS-TR project (sponsored by ARPA) and the WATERS (Wide Area Technical Report Service) project (sponsored by the National Science Foundation). [13]

Institutions interested in participating in NCSTRL should consider the following qualifying criteria:

1. Participating sites are required to adopt, implement, and use RFC 1807 and Dienst. Adoption of these tools allows the site to automate the collection, management, and network availability of its own repository of

computer science technical reports. The institution's report collection will become part of an expanding distributed library of technical reports through interoperation with other cooperating sites.

2. Doctoral granting U.S. institutions in computer science are invited to participate. Other institutions of higher education (or commercial or government research laboratories) who wish to participate should contact Rebecca Lasher (rlasher@forsythe.stanford.edu) to inquire about their possible involvement.
3. Institutions should only join if they feel they have a long-term commitment to disseminating computer science technical reports using NCSTRL tools.

+ Page 22 +

6.0 Lessons Learned

Over the three years of the project, every participant gained a better understanding of the intellectual, organizational, social, and legal complexities embodied in library services. Building sophisticated digital library services while preserving the enduring values of a traditional library is a difficult endeavor.

Among the lessons learned are:

- o Providing digital library services raises very difficult issues related to intellectual property and system scale, content, and use.
- o The underlying foundation of the digital library is content, structure, and organization. This foundation must be durable, but flexible enough to be useful in future environments.
- o From the beginning, create good content. Libraries cannot afford to redo the digital library with each new iteration of system design or access method.
- o A focus on system openness and interoperability is critical. The digital library is integrally involved with the nature of public and scholarly communication, information formats, and the economic and political environments within which information is created and sought. [14]

7.0 Conclusion

Libraries are operational, production-oriented service organizations. A librarian's evaluation of a research project tends to focus on how successfully the products of this project are integrated with (or replace) existing services and how well they can be supported and renewed in a production environment. The CS-TR project built several new prototypes, which became true production systems. During the course of the project, it addressed many key aspects of designing a digital library:

1. Discovery: matching the technology with the service vision.

2. **Delivery:** nurturing and developing this match in a prototype atmosphere to examine its feasibility and readiness for implementation.

+ Page 23 +

3. **Service:** the ongoing operations of the service and the continuous improvement of the service.
4. **Support:** provision of assistance, documentation, and training.
5. **Integration:** fit of the new service with the organization's overall architecture and services.

The CS-TR project made the most progress in the areas of discovery and delivery. More precise questions for each of the above processes were articulated. The project's discussions about integration issues related large-scale, distributed digital libraries will have a lasting impact on the field.

The CS-TR project provides a model of a working distributed digital library that will be useful to participants in the NSF Joint Initiative Digital Library Projects and as the conceptual framework for further research by other digital library developers. The NCSTRL system that evolved from the CS-TR and WATERS projects will contribute significantly to the broader digital library community. [15]

From a librarian's perspective, the CS-TR project offered the opportunity to work with and contribute to a world-class effort to transform scholarly communication. The learning experience was intense and gratifying. More questions have been formulated than were answered, but the new questions are better articulated and understood. One key question is whether a "digital library" is a real library as we understand it today or just a metaphor for something entirely different.

Notes

1. Jerome H. Saltzer, "A Proposal for M.I.T. Participation in an Electronic Library Plan" (Cambridge: Massachusetts Institute for Technology, 1992).

2. A great deal of research was done by the participating institutions that is not mentioned in this article. Detailed descriptions of these activities can be found on each project participant's Web page. See
<URL:<http://www.cnri.reston.va.us/home/cstr.html#participant>>.

+ Page 24 +

3. See
<URL:<http://www.ncstrl.org/Dienst/htdocs/Info/protocol4.html>>.

4. James R. Davis, Carl Lagoze, and Dean B. Kraft, "Dienst: Building a Production Technical Report Server" (Paper delivered at ADL '95: A Forum for Research and Technology Advances in Digital Libraries, Tysons Center, VA, 17 May 1995).

5. Ibid.

6. Robert Kahn and Robert Wilensky, A Framework for Distributed Digital Object Services (Reston, VA: Corporation for National Research Initiatives, 13 May 1995). See <URL:<http://www.cnri.reston.va.us/home/cstr/arch/k-w.html>>.

7. See <URL:<ftp://elib.stanford.edu/pub/reports/rebecca/copyright.html>>.

8. See <URL:<http://www.ncstrl.org/Dienst/htdocs/Info/protocol4.html>>.

9. Gloss is a research system, and the server may be unavailable at times. See <URL:<http://gloss.stanford.edu>>.

10. Sift is a research system, and the server may be unavailable at times. See <URL:<http://sift.stanford.edu>>.

11. See <URL:<http://lycos.cs.cmu.edu>>.

12. See <URL:<http://www.cnri.reston.va.us/home/cstr/handle-intro.html>>.

13. See <URL:<http://www.ncstrl.org>>.

14. Adapted from: Sarah M. Pritchard, "Librarians: Real Expertise for a Virtual World," Library Issues: Briefings for Faculty and Administrators 15, no. 5 (1995).

15. Clifford Lynch and Hector Garcia-Molina, Interoperability, Scaling, and the Digital Libraries Research Agenda: A Report on the May 18-19, 1995 IITA Digital Libraries Workshop. See <URL:<http://www-diglib.stanford.edu/diglib/pub/reports/iita-dlw/main.html>>.

+ Page 25 +

Acknowledgments

The research report upon which this article is based was sponsored in part by the Corporation for National Research Initiatives, using funds from the Advanced Research Projects Agency of the United States Department of Defense under CNRI's grant no. MDA-972-92-J-1029. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies or endorsement, whether expressed or implied, of ARPA, the U.S. Government, or CNRI.

Appendix A. RFC 1807 Fields

Bibliographic record fields should follow the format described below. "<M>" means the field is mandatory; records must include all mandatory fields. "<O>" means the field is optional.

The tags (a.k.a. the Field IDs) are shown in upper case.

<M> BIB-VERSION of this bibliographic records format
<M> ID
<M> ENTRY date

<O> ORGANIZATION
<O> TITLE
<O> TYPE
<O> REVISION
<O> WITHDRAW
<O> AUTHOR
<O> CORP-AUTHOR
<O> CONTACT for the author(s)
<O> DATE of publication
<O> PAGES count
<O> COPYRIGHT, permissions and disclaimers
<O> HANDLE
<O> OTHER_ACCESS
<O> RETRIEVAL
<O> KEYWORD
<O> CR-CATEGORY
<O> PERIOD
<O> SERIES
<O> MONITORING organization(s)
<O> FUNDING organization(s)
<O> CONTRACT number(s)
<O> GRANT number(s)
<O> LANGUAGE name
<O> NOTES
<O> ABSTRACT
<M> END

+ Page 26 +

For the text of the entire RFC 1807 standard, see
<URL:<http://ds.internic.net/rfc/rfc1807.txt>>.

About the Authors

Greg Anderson, Director, IT Discovery Process, MIT Information Systems, 77 Massachusetts Ave., Room E19-324, Cambridge, MA 02139. Internet: ganderso@mit.edu. (During the CS-TR project, Mr. Anderson was the Associate Director for Systems and Planning at the MIT Libraries.)

Rebecca Lasher, Head Librarian, Mathematical and Computer Sciences Library, Stanford University, Stanford, CA 94305-2125. Internet: rlasher@forsythe.stanford.edu.

Vicky Reich, Assistant Director Highwire Press and Information Access Analyst, Green Library, Stanford University, Stanford, CA 94305-6004. Internet: vicky.reich@forsythe.stanford.edu.

About the Journal

The World-Wide Web home page for The Public-Access Computer Systems Review provides detailed information about the journal and access to all article files:

<URL:<http://info.lib.uh.edu/pacsrev.html>>

Copyright

This article is Copyright (C) 1996 by Greg Anderson, Rebecca Lasher, and Vicky Reich. All Rights Reserved.

The Public-Access Computer Systems Review is Copyright (C) 1996 by the University Libraries, University of Houston. All Rights Reserved.

Copying is permitted for noncommercial, educational use by academic computer centers, individual scholars, and libraries. This message must appear on all copied material. All commercial use requires permission.