Access in a Networked World: Scholars Portal in Context

JERRY D. CAMPBELL

ABSTRACT

SINCE THE 1960S, LIBRARIANS HAVE projected a vision of a digital library that offers seamless access to a vast world of scholarly information of all types. Until the 1990s, however, digital technologies lacked the power and capacity to deliver on the vision. During the 1990s, technology platforms, networking technologies, electronic resources, and the evolution of standards matured sufficiently to lay the foundation for the vision to be fulfilled. As this maturation was taking place, the rapid growth of electronic resources was based on numerous proprietary systems making access across such systems impossible. The scholars portal project is an effort to create a search and retrieval tool that will provide an interim solution to this problem until such time as those systems are built on a unified set of standards and data formats.

Since the publication of the Association of Research Libraries (ARL) white paper on the need for a research library portal (Campbell, 2000), the concept of a *scholars portal* (SP) has generated much interest. Illustrative of this interest are what might be described as several independent demonstration projects sponsored by a number of entities including individual libraries, ARL,[1] the Ontario Council of University Libraries ("What's New," 2002), and the Council of Australian University Librarians.[2]

Interest in an SP also generically characterizes a number of efforts to create specialized subject portals for researchers.[3] Initially described as "the place to start for anyone seeking academically sound information" (Campbell, 2000, p. 211), the SP concept has been widely explored, developed, and refined.[4]

Jerry D. Campbell, Dean, University Librarian, and Chief Information Officer, University of Southern California, Doheny Memorial Library, Los Angeles, CA 90089-0182
LIBRARY TRENDS, Vol. 52, No. 2, Fall 2003, pp. 247–255

The purpose of this essay, therefore, is not to reiterate the definition of the now well-published concept of an SP but rather to outline the larger context within which SP exists. It is often our tendency to place exaggerated expectations on new technologies and thereby diminish the value of their eventual impact. Alone, SP constitutes only one small but vital step in the much larger jigsaw puzzle of the evolving digital library. Thus, it is important to place SP in this larger context and to understand the nature of the contribution it will make to the advancement of digital libraries.

Grand Vision

Since the earliest days of digital technology, we have been sharing grand visions about how it would transform the things we do. These dreams, of course, have included speculations about how libraries might be changed. By their nature, such visions often do not deal with details but rather focus on the larger dream. After the first decade of serious speculation about technology and libraries,[5] one study suggested that technology would transform the structure and function of libraries of the 1980s by storing materials in new formats, by making "obsolete the concept of the catalog and the book stack" as they were then known, and by linking them by means of a nationwide network (Conference Board, 1972, pp. 116–117). One of the best recent statements of the vision is equally elegant in its simplicity and challenging in its scope: "The dream to which we need to aspire is that all scholarly and research publications (including university, governmental, research, and museum sites) be universally available on the Internet in perpetuity" (Hawkins, 2000). Taken together, such recurring projections of the technology empowered library of the future have kept before us the vision of a slowly emerging digital library.

At almost any point over the past four decades, however, the challenge of these visions has outstripped the actual capacity of digital technology to deliver the dream. With hindsight, we can see that over the years virtually every aspect of the technology, from power and capacity to programming language, has been unequal to the challenge. Absent also have been the required infrastructures of connectivity and data standards necessary for the dreamed digital library to function. We now know that an operating digital library requires a vast number of elements functioning flawlessly together. Glimpsing the vision, it turns out, has been far easier than bringing it to reality.

All Things Necessary

Only recently have we begun to have in prospect all things necessary to implement the library envisioned for a generation. Among the vast number of improvements in computer technology, the following categories are key for the development of digital libraries.

Perfecting the Platforms

Sometime during the 1990s the persistence of the Moore's Law phenomenon produced computing hardware platforms of sufficient speed and memory to begin to implement the long-articulated vision. Amidst constantly improving machine specifications, it is not possible to isolate the moment it happened. Indeed, it is not possible to articulate exactly what speed and memory capacity were necessary. It is only possible to look back and recognize that during that decade we began to have access to hardware platforms of sufficient capacities to develop functional, if still rudimentary, digital libraries.

Similarly, these capable computers were accompanied by the development of other necessary components. Among these were

- the evolution of online storage systems with the capacity to hold and make available quickly massive amounts of information pertinent to digital libraries;
- the availability of increasingly sophisticated programming languages and software platforms;
- the introduction of improved systems for authentication and authorization.

This is not to argue that the computer platforms reached an end point or even slowed in their evolution in the 1990s. To the contrary, Moore's Law continues unabated (Kurzweil, 1999, pp. 20–25). It is only that they achieved sufficient capacities to allow implementation of digital libraries to begin in earnest.

Achieving the Connectivity

Even with the early Internet at our disposal, the possibility of highly networked libraries could not be realized as the 1990s arrived. Unlike with platforms, however, it is clear when a world-changing advance in connectivity occurred. In 1993 Tim Berners-Lee introduced the World Wide Web (the Web), which provided the infrastructure for flexible use of the Internet, thereby transforming connectivity (Berners-Lee & Fischetti, 1999). The Web rapidly became the mechanism by which libraries (and everything else) sought to achieve giant strides in the networking of resources.

Soon thereafter, the Internet itself was challenged to meet the growing bandwidth requirements stimulated by the Web as the so-called "commodity" Internet was born. As a consequence, Internet2 (I2) was launched to provide much greater bandwidths in a separate network environment for the research community. I2 technology made possible the rapid exchange of large files and empowered libraries to move from text to larger files such as those containing graphic materials.

As with platforms, the technology of connectivity continues to improve. Efforts are already underway to develop even higher speed optical networks. In addition, wireless data networks have recently added much needed nomadic flexibility for network users. Though wireless networks are comparatively slow shared environments, they, too, are rapidly increasing in bandwidth.

Evolving a Critical Mass of E-Resources

The 1990s also saw an amazing growth in the amount of information available on the Web (Lyman & Varian, 2000, p. 5). The surge in online information included journal literature from both for-profit and not-for-profit publishers as well as a rich sampling of archival resources principally from universities and national libraries. In addition, the decade brought a phenomenal surge in the production of raw data from the growth of computational science. In the panoply of published information, only monographic literature lagged behind, primarily because monographic publishers were slow to embrace the technology and because copyright restrictions served as an impediment (Lynch, 2002). All in all, therefore, the 1990s witnessed an extraordinary growth in the availability of digital information.

Creating Containers and Hooks

Unfortunately, the rapid growth of information available in digital format was marked by a significant problem from the standpoint of library users: it was characterized by a multitude of formats that did not offer a uniform means by which it might be found. In other words, the rapid growth of digital information began before we had developed common standards for data and metadata. As a result, users have not had an easy way to identify and retrieve information with thoroughness and precision. In this situation, users, especially neophytes, have had a difficult time identifying pertinent information, and thorough research has often been difficult even for experienced users. In addition, as the wealth of digital information continued to grow, the problem only worsened because the variety of formats and lack of metadata persisted.

Thus, the 1990s saw significant efforts to address the need for common formats for data and metadata. Centering on the Web environment, a number of data formats were introduced, chief among them being SGML (on which HTML is based) and more recently XML (Berners-Lee, 2002). Similarly, a number of metadata formats were developed, including Dublin Core, EAD, METS, and MODS (Tennent, 2002). These efforts, however, are ongoing and cannot be said to be fully mature. Thus, while they have had a positive impact by slowing the proliferation of data and metadata variations, they have not fully resolved the problems associated with what might be described as islands of disconnected data and information.

The Rise of Search Engines

With the early rise of information contained in data sources (databases and data repositories) in the 1960s came the need to provide tools for its retrieval. Such tools were first developed in connection with specific data sources. To function, information or objects in a data source first had to be described or indexed in a manner that could be interpreted by machine. These are the descriptions we now generally refer to as metadata. A "search engine" was then developed and customized to the indexing scheme, making it possible to retrieve data. Depending upon the quality of the indexing and search engine, such tools allowed users to find and retrieve specific data objects from data sources with speed and precision. Most significant data sources were accompanied by customized, unique search engines.

Later, as database architecture became better developed, a number of major providers offered the capacity to search across large amounts of information as long as it was stored in their proprietary syntax for expressing structure in data. This improved the general situation, but still left it difficult to work with information spread among different proprietary solutions. The economics of information and the mechanism of copyright law have provided significant disincentives for solving this problem of standalone data sources.

With the introduction of the Web and the vast amount of information located there (or, perhaps, lost there) came the urgent need to develop tools for retrieval in the Web environment. Thus, in the mid-1990s, a number of agencies developed search engines designed specifically for the Web. These engines, commonly referred to as portals, primarily searched for keywords and phrases and applied relevancy ranking schemes to determine validity. They also developed their own indexes from keywords and phrases in order to carry out searches quickly. As they have improved, they have become sophisticated, powerful, and amazingly adept at locating information on the Internet.

The sophistication and power of evolving search engines notwithstanding, however, certain limitations still confronted library users. The data sources served by specific search engines were still a disconnected sea of information islands whose contents could not be discovered and retrieved by a single search tool. These unique search engines characterized most digital information and data licensed by libraries. Furthermore, the extraordinary Web search engines could not see the actual data beneath such search engine–driven data sources even when those data sources were Web enabled because the Web only provided access to the unique search engines themselves. Indeed, the Web itself became another information island (albeit from the standpoint of size, continent might be a more accurate metaphor) and compounded the problem for those seeking information.

Scholars Portal

It was with respect to the context outlined above that the concept of an SP was developed. The growing plethora of high-quality E-resources created a clear need for a tool that could, with a single search interface with a multitude of unique search engines (including Web engines), discover and retrieve relevant information from each, and present a single, merged set of results to the researcher. While other desirable features have been included in the Association of Research Libraries SP Project, this is its fundamental purpose and the need that drove its conceptualization there and elsewhere. The combination of capable platforms, high bandwidth connectivity, maturing data and metadata formats, and sophisticated but target-limited search engines was sufficient to indicate that such a portal was possible and, moreover, suggested where it should fit in the panoply of information technology.

Specifically, an SP would serve as an aggregator of search engine–driven data sources. Initially, it would necessarily be limited in scope because it would require that interfaces be created (programmed) and kept up-to-date for each data source. In their first implementations, therefore, SPs will be institution- or agency-specific with a defined set of data sources identified as targets to be aggregated. They might be thought of as offering second-level search engines in that an SP engine would sit above other search engines. In many cases, the nature of license restrictions on data sources suggests this approach. Even when local SPs become linked via the Web, certain limitations on the retrieval of information may be required at each location, based on the number and nature of institution- or agency-specific licenses. Nonetheless, by searching across even a limited set of data sources, SPs will vastly improve the prospect that digital library users will be able to discover and retrieve high-quality information.

Glimpsing the Future

As much as they are needed, SPs constitute a poor solution to a complex problem. They are poor solutions because they represent yet another layer of technology necessary to solve problems created by earlier layers of technology and because they must be adjusted each time underlying technology changes. Thus, SPs may best be thought of as an interim but necessary step in the evolution of tomorrow's digital library, a step that will be made obsolete upon the eventual emergence and utilization of accepted standards for data and metadata along with a new generation of tools for searching. Such standards and new tools could both reduce the number of technology layers and increase the ease of information discovery and retrieval.

It should be noted that the perfection of standards and new discovery and retrieval tools will not alone alter the impact of economic incentives for providers to continue to maintain separate data sources. Only if the habits of scholars come to express a preference for standards-based sources and tools to the exclusion of separate data sources will the economic incentives be reversed.

Such standards and tools are only now being developed and will be some time in development. Until the Web is supplanted, they will also be Web-based technologies. In addition to the continued evolution of the standards for data and metadata noted above, an example of a new approach is the OpenURL effort (Stern, 2001). This solution would function by running data mining processes (so-called "smart agents") against a generic, public syntax (OpenURL) resource identification system, by utilizing an identified local "resolver" machine to validate and give the location of items, and by employing extensible metadata syntax to standardize and store query data from a variety of metadata formats. While such a system is feasible, as the previous sentence indicates, it is complex and requires not only agreement on the OpenURL syntax but the development of viable smart agents and "resolver" machines loaded with appropriate data. It will also require time.

Indeed, given the current situation and in spite of our considerable technological prowess, no ultimate solution to the fragmentation of data sources is likely to be simple or quick in its development. And compared to any ultimate solution, scholars portals are much simpler and already available in first-generation versions. This indicates that efforts to test scholars portals, even if only as interim solutions to the problem, are necessary and justifiable. Not only will they move us a little closer to the dream of a universal, networked digital library, they will also give our users something they urgently need today.

As for the grand vision of the digital library of the future, it will eventually come to pass. In time, "all scholarly and research publications (including university, governmental, research, and museum sites)" will indeed "be universally available on the Internet in perpetuity" (Hawkins, 2000). It may be hard today to believe that such an outcome will be achieved, but a scant decade ago it would have been equally hard to believe that something called "the Web" would transform not only the distribution of knowledge but the habits of the workplace as well. It is important, therefore, that we continue to believe in the vision and that we continue to articulate it. It is also important that we work to make it a reality.

Notes

1. See "Seven ARL Libraries" (2002) and Quint (2002).
2. See http://library.queensu.ca/libguides/databases/scholarsportal.htm; http://anu.edu.au/caul/caul-doc/caul20022aarlin.doc.
3. See Halbert (2002). See also Technical Issues ad hoc Committee (2002).
4. See the substantive article by the ARL Portal Project manager M. E. Jackson (2002). See also Thomas (2000a).
5. See the summary in Fussler (1973, pp. 1–11).

References

Berners-Lee, T. (2002). *W3C data formats.* Retrieved December 2, 2002, from http://www.w3.org/TR/NOTE-rdfarch.html.

Berners-Lee, T., & Fischetti, M. (1999). *Weaving the Web: The original design and ultimate destiny of the World Wide Web by its inventor,* 1st ed. New York: HarperCollins.

Brown, J. S. (2002). Learning in the digital age. In *Futures Forum 2002: Exploring the future of higher education* (pp. 20–23). Cambridge, MA: Forum for the Future of Higher Education.

Campbell, J. D. (2000). The case for creating a scholars portal to the Web: A white paper. *ARL Newsletter 211.*

Conference Board. (1972). *Information technology: Some critical implications for decision makers.* New York, NY: Conference Board, Inc.

Fussler, H. H. (1973). *Research libraries and technology.* Chicago: University of Chicago Press.

Fyffe, R. (2002). Technological change and the scholarly Communications Reform Movement. *Library Resources & Technical Services, 45*(2), 50–61.

Gillespie, R. G., & Dicaro, D. A. (1981). *Computing and higher education: An accidental revolution.* Seattle: University of Washington Press.

Greenstein, D., & Thorin, S. E. (2002). *The digital library: A biography.* Washington, DC: Digital Library Federation and Council on Library and Information Resources.

Halbert, M. (2002). The Metaarchive.org Project: A joint project of Emory University and ASERL. Retrieved August 26, 2002, from http://www.metascholar.org/modules.php?op=modload&name=News&file=article&sid=4.

Hawkins, B. (2000). Libraries, knowledge management, and higher education in an electronic environment. Retrieved October 7, 2002, from http://www.alia.org.au/conferences/alia2000/proceedings/brian.hawkins.html.

Herring, S. D. (2002). Use of electronic resources in scholarly electronic journals: A citation analysis. *College & Research Libraries, 63*(4), 334–340. Retrieved June 23, 2003, from http://library.queensu.ca/libguides/databases/scholarsportal.htm.

Jackson, M. E. (2002). The advent of portals. *Library Journal.* Retrieved October 28, 2002, from http://libraryjournal.reviewsnews.com/index.asp?layout=articlePrint&articleIDCA242296.

Kurzweil, R. (1999). *The age of spiritual machines: When computers exceed human intelligence.* New York: Penguin Books.

Lougee, W. P. (2002). *Diffuse libraries: Emergent roles for the research library in the digital age.* Washington, DC: Council on Library and Information Resources.

Lyman, P., & Varian, H. (2000). How much information? Retrieved August 8, 2002, from http://www.sims.berkeley.edu/research/projects/how-much-info/summary.html, p. 5.

Lynch, C. (2002). What's become of the digital library? Retrieved December 4, 2002, from http://www.educause.edu/asp/doclib/abstract.asp?ID=EDU0207.

NetLibrary and eBooks an excellent fit for OCLC. (2002). *OCLC Newsletter, 256,* 30–31.

Project MARC Reports. (1968). *Library Resources and Technical Services, 12*(3), 245–319.

Quint, B. (2002). Academic libraries develop integrated portal software package. In *Newsbreaks.* Retrieved December 4, 2002, from http://www.infotoday.com/newsbreaks/nb020513-2.htm.

Seven ARL libraries launch Scholars Portal Project. (2002). *Library Journal.* Retrieved December 4, 2002, from http://libraryjournal.reviewsnews.com/index.asp?layout=articlePrint&articleID=CA215696.

Stern, D. (2001). Automating enhanced discovery and delivery: The OpenURL possibilities. *Online.* Retrieved August 29, 2002, from http://www.infotoday.com/online/OL2001/stern3_01.html.

Technical Issues ad hoc Committee on Content Linking and Management for Electronic Serials and other Digital Materials. (2002). Working bibliography. Retrieved August 28, 2002, from http://www.viva.lib.va.us/viva/tech/cat/link/working_bibliography.html.

Tennant, R. (2002). Digital libraries—Metadata as if libraries depended on it. Retrieved August 26, 2002, from http://libraryjournal.reviewsnews.com/index.asp?layout=articleID=CA206408&.

Thomas, S. E. (2000). Abundance, attention, and access: Of portals and catalogs. Retrieved December 6, 2002, from http://www.arl.org/newsltr/212/portal.html.

Thomas, S. E. (2000). The catalog as portal to the Internet. Retrieved August 26, 2002, from http://lcweb.loc.gov/catdir/bibcontrol/thomas_paper.html.

Travis, T. A., & Norlin, E. (2002). Testing the competition: Usability of commercial information sites compared to academic library Web sites. *College & Research Libraries, 63*(5) 433–448.

What's new: Ontario Scholars Portal. (2002). Retrieved December 4, 2002, from http://www.uottawa.ca/library/actnew-e.html.