# The Digital Library: Myths and Challenges

Terry Kuny
Global Village Research
Hull, Quebec
Email: terry.kuny@xist.com

Gary Cleveland
Information Technology Services
National Library of Canada
Email: gary.cleveland@nlc-bnc.ca

## Introduction

The purpose of this paper is simple: it is a provocation. It aims to rebut the technolust and rhetorical excess that has characterized much of the reporting on "digital libraries." The polemic here aspires to provide a corrective to the reporting that is whipped up by the technology companies, politicians, and *Wired* magazine—the same forces that gave us the "paperless office" and predict the "demise of the book." But it does not suggest that "digital libraries" are a waste of time and money. There is no call for a neo-Luddite uprising against "digital libraries" developments and technology. But the polemic tone is intended, for a call-to-arms is needed in the library community to meet the challenges of "digital libraries"—and we must attempt to recognize these challenges clearly. To further this process, a suggestion for a new focus in the pursuit of "digital libraries" is suggested which present a renewed role for librarians in the digital future.

Technological progress has changed *how* libraries do their work, not *why*. But the most profound technological development—a connection of computer to computer in an unbroken chain around the world—may alter the fundamental concept of the library in the twenty-first century. But we would suggest that technology will not substantially alter the business of librarians—connecting people with information.

If librarians and information professionals are going to progress into the 21st century then a clear and effective "digital library" model for library services and development will be increasingly important. An increasingly complex technological, social, legal, and economic environment defines many boundaries within which "digital library" services will evolve. Librarians may discover that "libraries-without-walls" are actually only libraries with new walls—technologically bounded, legally restricted, and administratively hamstrung. The "digital library" may be equally impenetrable and as profoundly limiting to their patrons as

*The authors would like to thank theirs colleagues at the National Library of Canada for their support and insights as we rant, rave, explore and develop, digital collections and services. Their tolerance of eccentricity and provision of a creative work environment has allowed us to become better digital librarians.*

the physical library which techno-pundits would suggest digital collections are intended to replace.

Exposing some myths that permeate the popular press reporting about digital libraries sets the stage for a closer examination of the significant challenges to "digital library" development.

## Digital Library: Myths

**Myth 1: The Internet is the digital library.**

The word “library” has been appropriated by many different groups to signify simply a collection of digital objects that people can access from their desktops—digital objects like electronic documents, and digitized pictures, sound, and video.

A global information network, of which the Internet is the seed, has the illusion of promising fingertip access to the world’s information. A fairly spectacular example of what many people consider to be a digital library today is the World Wide Web. The Web is a gathering of thousands and thousands of documents. Many would call this huge collection of documents a “digital library” because they can read and use whatever they wish by accessing the Web, just as one can use technology to do banking in a “digital bank” or buy compact discs in a “digital record store”.

But is this a "digital library"? In reality, the Internet and the World Wide Web are to libraries what a fleamarket is to the Library of Congress. For many common library requests, locating information on the Internet remains highly inefficient compared to traditional library sources, especially for unfamiliar users. Finding information is difficult, the quality of the information is quite variable, and reliable, professional assistance for the confused and lost is lacking.

What challenges emerge? The development of an infrastructure for the networked resource discovery and retrieval of highly distributed, autonomously created, and diverse electronic information is required. Above all, this infrastructure will need to be managed by professionals who understand information needs and uses. An often repeated quote among library lists is that the Internet is the place to find an answer in 3 days for a query that would take 3 hours in a library. Evaluation and authority are required in order to ensure the "slow horse of meaning" is not overtaken by the "fast horse of mere information." [Klapp] There remains much work to be done before the Internet will have the coherence and user-friendliness of a library.

**Myth 2: The myth of a single digital library or one-window view of digital library collections.**

Nicholas Negroponte, a guru of the digital cognoscenti, has called for the U.S. Congress to pass a "digital deposit" act to change the Library of Congress from a "depository" to a

"retrievatory". His vision of the library is one where a "Library of Progress could be in the pockets of tomorrow's kids" and where citizens can get electronic access to a library-without-walls where information is accessible anywhere and anytime. [Negro 95]

The challenges to this vision? Despite the utopianism of Negroponte's view, even modest moves towards increasing digital collections and services will be strongly affected by future copyright and licensing regimes, as well as prohibitive costs for digitization and support of technical infrastructure.

But more importantly, the digital future will be an unruly one composed of multiplicity of competing information providers. Libraries will be only one source of information. "Prime" information resources will probably be locked into proprietary collections—essentially "private digital libraries" which are accessible on a subscription or pay-per use basis. Developing interoperability standards for locating and retrieving information in this highly distributed and heterogeneous environment will be a considerable challenge in their own right.

**Myth 3: Digital libraries will provide more equitable access, anywhere, any time.**

A great deal of work must be done to turn this myth into reality. We can assume that a global computer network—the Internet or some descendant—will be the primary delivery mechanism for digital information. Equitable access is currently compromised by the fact that the Internet is not as ubiquitous as the computing press would have us believe. There are relatively few connections outside the more populated centers, the costs of access can remain high, and for the vast majority of the world's population in developing countries, having widespread Internet access may be the equivalent of walking on the moon. Furthermore, the connections that do exist for most people are slow. For a digital library to provide equitable access to information, it is imperative that the same universal availability that is a characteristic of the telephone system is also a characteristic of the network.

In the future, complex multimedia resources and services may have specialized hardware and software requirements such that only a limited number of workstations can actually access the information. Limits of network bandwidth and slow transmission speeds may make the effective access to information problematic for many users.

Given the immense technical and legal hurdles involved, the prospects for equitable access to digital collections and services seems increasingly problematic. Copyright reform will be a slow process and has the potential to derail the very idea of "digital libraries". The technologies on the desktop, between computers, and for storing and processing information are dynamic variables. What is certain is that the management of technology for digital libraries are becoming more complex as is the administration of licenses and user access. The impact upon equitable access could be considerable.

**Myth 4: Digital libraries will be cheaper than print libraries.**

A common assumption among technology reporters about the costs of "digital libraries" is that digital is cheaper than paper. This contention is far from established in fact or in practice. Although many libraries project savings, especially when substitution strategies are used which replace selected serials titles with document delivery services, the cost/benefit analysis of making this switch remains unclear. In some cases, the switch to electronic serials may save the library money by offsetting the cost to users who must pick up the charge for document delivery.

Furthermore, the costs of "being digital" are substantive ones. Many libraries now devote significant resources for hardware and software infrastructure. These expenses will increase—new hardware will be required, more licenses to software, increased infrastructure administration and training. And these costs are borne by libraries who only be acquiring digital materials and have limited electronic services. Those institutions that aspire to the development of digital collections and services can expect all of the above plus extensive design, digitization, and implementation costs. Are digital library budgets evolving at the expense of decreasing acquisitions budgets? At the end of the day, how many libraries can afford the effort? And at what cost to the valuable existing services they perform?

## The Digital Library: Setting out the Challenges

Creating “effective” digital libraries poses serious challenges for existing and future technologies. The integration of digital media into traditional collections will not be straightforward, like previous new media (e.g., video audio tapes), because of the unique nature of digital information—it is less fixed, easily copied, and remotely accessible by multiple users simultaneously. Traditional library processes such as collection development and reference, though forming a potential basis for "digital library" work, will have to be revised and enhanced to accommodate these differences. Taking what we know about libraries as a starting point, we can begin to examine in more detail what the specific challenges might be.

### Resource Discovery

Digital information on the Internet is characterized by the fact that digital documents can exist in several formats, possibly in several versions, in locations that are not yet fixed. A document or resource may exist at one network location one day, and disappear the next. Services such as AltaVista, YAHOO, and other WWW services are increasingly popular. These indexing services provide an essential service in assisting users to find information. But users are already noting that these services are becoming overburdened and that obtaining meaningful results can be frustratingly elusive.

The indexing service developers are still trying to establish a cost recovery model that will allow them to match capabilities with the growth of information on the Web. Currently this is through advertising. But there is no guarantee that these services will provide high-quality service to the general public for indefinite return on investment. A reasonable expectation is that there will be levels of service established. At a base level, anyone can access the database without charge but the results will be of limited usefulness and good information will be swamped by a flood of the bad. Searching will not be efficient. But a "higher" level of users who can pay for better service may get benefits such as increased search functionality, better algorithms for ranking information, or higher speed access. Presumably the techno-peasants, to paraphrase the famous saying of Marie Antoinette, are expected to eat "digital cake."

There are also important questions as to whether the existing technologies can scale sufficiently to accommodate growth in the number of users and the volume of information on the Internet. Users are discovering the quality of information retrieved from large heterogeneous databases may be lost in a flood of irrelevant results. Large centralized databases represent single points of failure and bottlenecks.

Librarians organize knowledge through the processes of subject analysis and cataloguing—creating information about information, or what is known as “metadata”. A major challenge exists to develop methods of consistently and uniquely identifying and retrieving networked information, no matter what format they are or where they reside. Metadata standards are still in their infancy. Initiatives and research such as the Dublin Core/Warwick Framework for metadata, Government Information Locator Services (GILS), and Uniform Resource Characteristics (URC) are promising ways forward. But discussions by many of the participants in these efforts suggest that given the complexity of metadata issues, a solution to the global resource discovery problem remains distant.

Librarians provide another role in supporting resource discovery, one in which digital technologies play only a small role. This is in providing a "trusted" service. This trust and authority is based upon librarians' making choices, evaluating information as a part of collection development and with a thorough understanding of what users need. Librarians are do not only create pathfinders and guides for their users, they *are* the pathfinders and guides.

### Digital Collection Development

Librarians have considerable experience in digitization, although the profession has tended to call it something else. The "retrospective conversion" of printed library cards into machine-readable catalog records represents one of the earliest widespread, digitization efforts. What was learned in the process? One lesson is surely that conversion of paper into digital is expensive and time-consuming. The cost/benefit analysis to librarians and users for enhanced bibliographic access established the benefit of the expense, but it was acknowledged that the costs involved were higher than anticipated.

Some types of media reveal themselves to be more suited to digitization than others. Photographic collections, bibliographic resources, statistical collection, and even some kinds of journal literature are amenable to digitization. Other materials such as maps and books may be less amenable to digitization. Considerable study of what users need, how they use information, and whether digital formats serve their needs effectively is still required. Undertaking large digitization initiatives without a fundamental understanding of user needs is putting the cart before donkey. Being digital is not necessarily commensurate with being useable.

In an era of difficult to obtain resources, questioning the efficacy of undertaking the expensive process of digitizing specialized collections of materials that may support a handful of scholars is not only legitimate but essential. It is possible that some digitization efforts will create collections of what is essentially, dead digital information—information with low market value, of limited interest and utility, and whose circulation is no greater than its paper or microfilm equivalent. It may be that most of the important information that the poor in any country need may not be in databases.

Librarians collect published information in a variety of formats—books, journals, CD-ROMs, audio and video tapes and discs. To this growing set of media, libraries are adding repositories of digital information, on-line databases of documents and images in various formats. It will not generally be the case that libraries will *replace* older media with digital media, but that they will collect them *in addition to* established media. The reasons this substitution will not easily occur are many: user resistance, limitations on use, poor digital product design, or the medium may not be effective to satisfy the user requirements. The challenge here will be to “span both print and digital materials... [and to] ...provide a coherent view of a very large collection of information.” [Lynch 95]

**Preservation**

If we assume that libraries are able to build and/or acquire some types of digital collections, there remains a significant challenge inherent in preserving these collections. Pre-digital libraries have had to worry about climate control and the de-acidification of books, but the preservation of digital information will make these time-consuming and costly problems look easy. For example, digital storage media are "fragile", with a limited shelf life. Worse yet, the digital information on those storage media, even if they do survive will be rendered unreadable by obsolescence of technology—the fact that as information technology evolves, older systems disappear taking with them the ability to read the information they managed.

To preserve digital information, digital libraries will continually have to “migrate” information from one digital hardware and software configuration to another. The Report of the Task Force on Archiving of Digital Information suggests that "rapid changes in the means of recording information, in the formats for storage, and in the technologies for use threaten to render the life of information in the digital age as, to borrow a phrase from

Hobbes, 'nasty, brutish and short.'" [TFADI 96] The cost of such migration is unknown and there is no guarantee that future generations will have the funds to do this. The digital preservation function must be attended to in all digital collections. Even libraries which do not normally have a significant preservation concern will find that digital collections will require "refreshing" and migration to new systems to maintain their accessibility. Technological obsolescence, migration of digital information, legal and organizational issues all test the "limits of digital technology." There are no preservation standards for digital information.

In the pre-digital world, libraries have had a long-standing tradition of resource sharing. This resource sharing takes the form of reciprocal borrowing privileges, coordinated collection development, preservation programs, shared cataloguing, and union lists of serials. It has been carried out through many associations, consortia, cooperative projects, and other formal and informal resource sharing agreements. No single library can take upon the responsibility of "doing it all".

The same resource-sharing will be necessary in the coming digital era. For example, in the United States, libraries have begun to create strategies for sharing digital information. The U.S. National Digital Library Federation, made up of the largest American research libraries and archives and the Commission on Preservation and Access are working to develop a coordinated funding strategy and formulate selection guidelines to collect electronic information in the US. Without such arrangements, there will be no one to ensure that the terabytes of digital information that will be scattered about the network will be collected, ordered and preserved. In Canada, the Data Liberation Initiative is an co-operative effort by Canadian universities to increase access to Canadian statistical databases through common licensing and access arrangements.

Librarians argue that if we do not emphasize the *library* in the phrase "digital library" and build collections that can be preserved, then future generations will look back at this time as a digital Dark Ages—a time when, somehow, the record of human knowledge went missing. In the final report of the Task Force on Archiving of Digital Information, the first of the general conclusions was that "the first line of defense against the loss of valuable information rests with the creators, providers and owners of digital information." [TFADI 96] Are libraries the second line of defense? Can and should we entrust our electronic legacy to creators, providers and owners of digital information whose interest in preserving information may only be as long as it has market value—materials which may disappear when that market value approaches zero?

### Digital Library Administration

Peter Graham of the Rutgers University Libraries, suggests that for implementation of a Digital Research Library, long-term organizational, fiscal, and institutional commitments will be necessary. The technical tasks are "the easiest to solve; they will only cost money"—it is the institutional commitments that "will be much more difficult to achieve."

[Graham 95] The TULIP final report provides supporting evidence of the importance of this commitment:

> *Politics, lack of priority and lack of responsibilities can cause long delays and have all but killed the [TULIP] project in a few of the TULIP universities.*

Management of the technical infrastructure for "digital library" services will be a significant hurdle for most libraries, especially as budgets continue to shrink and the costs of developing and maintaining collections increases. The recently released final report of TULIP (The University Licensing Program), a major project between Elsevier Science and 9 leading American universities to test systems for the networked delivery of electronic publications, concluded that "managing large digital collections locally, is harder and more expensive than managing a comparable print collection." [TULIP 95]

#### Copyright and Licensing

If libraries do begin to systematically collect digital information on a larger scale, the provision of effective access could be questionable. In fact, copyright could end up preventing libraries from providing open access to the digital information they collect. Questions of copyright must be managed so that digital information can be created and distributed throughout "digital libraries" in a manner that is equitable for both information producers and information consumers. Copyright could become an insurmountable barrier to the development of digital collections.

There are indications that content providers unhappy with the protections afforded them under copyright law, will turn to contract law and licensing for protection. Libraries are already experiencing the administrative burden of managing site licenses for electronic information such as CD-ROMs and data files. Licensing provides content providers with a stronger mechanism to control the transmission and use of information. This has the effect of moving information from a realm where ideas are allowed to flow in the public domain, to one where this flow is controlled by the provider.

There is an increasing unease among members of the library community that copyright changes will adversely affect the ability of libraries to provide digital collections and services. The discomfort librarians feel is justified. One has only to consider the statement of the International Publishers Copyright Council. on digital library collections to sense the challenge that librarians face:

> *Many national and regional libraries contemplate digitizing their print collections to facilitate a virtual library that can provide service to patrons at remote locations and facilitate resource-sharing. Such a concept will destroy not only the incentive to create new copyrighted works, but the revenue from existing works that provides the investment in new works by authors and publishers. [I.P.C.C. 96]*

Information providers such as publishers increasingly see libraries and themselves as sharing the same customer base. Publishers view libraries as threats to their market. What is being established is a sense among publishers that they are in the same business as libraries:

> *No longer will libraries be the sole repository of published matter. No longer will libraries be the only means of obtaining archival information. In some areas, libraries will be able to fulfill their function by merely pointing to other electronic repositories and in others they will seek out more active roles. [I.P.C.C. 96]*

But remains important to remember the "public" is not the same as a "customer" and access to "publicly available information" is not a product. Herein lies a fundamental difference between libraries and commercial information providers.

Under restrictive conditions of use, whether imposed by contract or some revised copyright legislation, "digital libraries" will not be able to satisfy many of the imperatives of information anywhere, anytime. Libraries will be required to provide reasonable assurances to content providers that the terms of their licenses can be maintained, and that distribution of copyrighted materials are restricted to particular users or locations. It is even likely that users will have to visit the physical building of the library since the digital collections may only be available on particular workstations or require special equipment to access the materials.

### Cost

Information has never really been free. There is always a cost in its creation, its production, and its dissemination. Freely-accessible public libraries, subsidized through taxation, largely hide the real cost of information from library patrons and this is appropriate where libraries are considered as a public good. "Digital libraries" introduce new and uncertain economic realities and relationships into libraries. Where the costs of accessing information were once hidden to “patrons”, the digital era is likely to require “customers” who will be required to pay fees for access to digital services and collections.

A major assumption of the information age, is that information will be available to all—for a fee. This assumption runs counter to the ethos that underlie libraries. It will be a cruel irony that the very technology that holds so much promise of providing access to digital information *en masse* will end up restricting it to only the very few that can afford it. What is affordable for some users, isn't for many others. "Digital libraries" may be privately-owned corporate services and collections to which subscription, pay-per-use, or licensing fees may apply. Libraries are already having a taste of this future as they wrestle with restrictive licenses for the use of data tapes and CD-ROMs. Users are often required to

use digital materials on-site in order to satisfy the contract requirements of the information providers.

The TULIP final report suggested that "building digital libraries will be a costly and lengthy process" and that making additional funds available for this content "will not be a trivial issue." The "harsh economic realities" are that digital collection development entails heavy costs for implementation, licensing, training, promotion, and the development and support of a technical infrastructure. Furthermore, the report suggested the one critical issue which was not resolved was "how to make the transition to digital libraries work economically." [TULIP 96] Economic models for making the "digital library" work, in terms of real costs and benefits, have neither been clearly articulated nor established.

## Conclusion: Rethinking Digital Libraries, Reinventing Librarians

Digital information is, and will be, treated differently than paper-based information. It is likely that in the near future, the terms of accessibility and the conditions for management and collection of electronic information will not be determined by the library profession within the context of traditional library services, but rather by information professionals working to maximize return on a corporate information resource. Making the distinction between "public digital libraries" and "private digital libraries" will become an increasingly important consideration.

In the view of some librarians, a "digital library" should do all the things that traditional libraries have done for hundreds of years, and play the same essential role in society that libraries have always played. Accordingly, a true "digital library" will build on the central library ethic: it would exist as a sustainable information commons that supports open access to a wide variety of material expressing diverse viewpoints. The only difference is that a "digital library" operates in an electronic medium.

In this view, a "digital library" should include a collection of digital objects, but it would encompass much more than such a collection. A "digital library" would also include all the processes and services—collection development and management, subject analysis, index creation, reference work, and preservation—that are the backbone and nervous system of contemporary libraries. These are the processes and services that are invisible in a well-run library, and thus are taken for granted.

Although this perspective of the "digital library" is predictable because of existing library models, there remains an anachronistic quality to it. As suggested above, the costs, technologies, legal issues and administration of "digital libraries" militate against achieving this old paradigm vision.

One important consequence of the information revolution is that the costs of organizing information are beginning to match the costs of producing the information. A number of

technology experts have suggested that the future electronic information environment should be based on an "underlying ethos of abundance rather than scarcity" of information. [ARL 94] In this view, it is the *context* not the content that will be locus for value. "The future belongs to neither the conduit or content players," posits Paul Saffo of the Institute for the Future, "but to those who control filtering, searching, and sense-making tools we will rely on to navigate through the expanses of cyberspace." [Saffo 94] Esther Dyson, a well-respected commentator on technological developments concurs. The "value shifts from the transformation of bits rather than bits themselves, to services, to the selection of content, to the presence of other people, and to the assurance of authenticity - reliable information about sources of bits and their future flows." [Dyson, 94]

Librarians should be heartened by this future. Computers only manipulate numbers—it is people that connect them to meaning. Librarians provide context to users. Even as the stuff of library collections begins to change and become collocated in the private digital libraries of publishers and content owners, the value of librarians who can effectively turn mere data into knowledge will be paramount.

A "library" has always been more than a building containing books, or a computer on a network full of documents. In some respects, "digital libraries" are not new: libraries have been using technology to facilitate access to information for years and telephone reference can easily be considered a type of "digital library" service. Following the direction suggested by Esther Dyson, librarians might willing accept the "depreciation of intellectual assets and property", i.e. digital collections, while finding a greater appreciation in the "intellectual processes and services" that a "digital librarian" might provide.

A different view of the future might be one where a "digital library" is more like a "knowledge center", where a complex system of professionals whose expertise supports access to information and acts as an intermediary to a variety of digital and other sources. These digital librarians/knowledge workers, who imbued with an ethic of equitable access, would function as well-trained intermediaries in an heterogeneous information environment—an environment that if not actively hostile to users is certainly confusing—to find and make sense of the masses of data for their users. The knowledge that "digital librarians" bring to this information environment would make sense of a multiplicity of digital and paper-based collections and resources, provide access to a network of key contacts, identify cost effective strategies for information retrieval, and assist users in the publication and creation of new information.

Open access to information—it is this principle that lies at the heart of the modern library, digital or otherwise. It is this principle which must be upheld against the many forces which might diminish its enlightening force. But perhaps open access to information in the future does not mean open stacks and digital collections. Is it possible that the principle might be changed slightly to "open access to knowledge", a principle which suggests a right to publicly accessible professional services that can guide users through information

flows and mediate information overload? Is it not possible that the value of libraries is not in the collections, but in the librarians? In a turbulent technological environment, perhaps a change of scenery is required. Redirecting the focus of librarians' attention and resources from the development of "digital libraries" to the development of "digital librarians" will be vital to the future of the profession. The time has come to invest in people and not in technology. Central to the vision of the new digital library is a digital librarian/knowledge worker who cares about people.

## Footnotes


[Bush 45]. Bush, V. "As we may think." *Atlantic Monthly.* July, 1945.

[Dyson 94]. Dyson. Esther. **"**Intellectual Value." *Release 1.0* December, 1994.

[Graham 95]. Graham, Peter. *The Digital Research Library: Tasks and Commitments.* Proceedings of Digital Libraries '95, Austin, Texas, June 11-13, 1995. <URL: http://www.csdl.tamu.edu/DL95/>

[I.P.C.C. 96]. International Publishers Copyright Council. *IPCC Statement: Libraries, Copyright and the Electronic Environment.* April 22, 1996.

[Klapp 86]. Klapp, Orrin. *Overload and Boredom: Essays on the Quality of Life in the Information Society.* New York: Greenword Press, 1986.

[Lynch 95]. Lynch, C., and Garcia-Molina, H. (1995). *Interoperability, Scaling, and the Digital Libraries Research Agenda: A Report on the May 18-19, 1995 IITA Digital Libraries Workshop, August 22, 1995.* <URL:http://www-diglib.stanford.edu/diglib/pub/reports/iita-dlw/main.html>

[Negro 95]. Nicholas Negroponte. "A Bill of Writes." *Wired.* 1995.

[Saffo 94]. Saffo, Paul. "It's the Context, Stupid." *Wired.* 1994.

[TFADI 96]. Task Force on Archiving of Digital Information. *Preserving Digital Information: Final Report and Recommendations.* 1996. <URL:http://lyra.rlg.org/ArchTF/>

[TULIP 95]. Elsevier Science. *TULIP Final Report.* July, 1996. <URL:http://www.elsevier.nl:80/homepage/about/resproj/trmenu.htm>