# Strategies for Sustaining Digital Libraries

Edited by Katherine Skinner and Martin Halbert

Emory University
Digital Library Publications
Atlanta, Georgia

# Publication and Cataloging Information:

ISBN: 0-9772994-1-4

Editors: Katherine Skinner, Martin Halbert
Copy Editors: Mary Battle

Publisher: Emory University
Digital Library Publications
Atlanta, GA 30322

# ONCE IN A HUNDRED GENERATIONS

Paul Arthur Berkman (University of California, Santa Barbara)

**Abstract:** Once in a hundred generations – every 2,000 years – an information technology threshold is reached that changes human capacity to manage and discover knowledge. Invention of the digital medium created such a paradigm shift and we are now faced with the challenge of sustaining the information products generated with this transformational technology. For the last several thousand years, libraries and archives have provided the architectures to manage information based on their content and context, respectively. With digital technologies, however, the inherent structure of information (i.e., boundaries between granules of content) also can be applied to information management. Lessons learned from the National Science Digital Library (http://www.nsdl.org) reveal that technological as well as organizational and economic strategies are necessary to sustain digital libraries as "public goods." Implementation of a national task force on digital library sustainability is recommended to elaborate visionary solutions for knowledge management and discovery in our evolving digital era.

## A BRIEF HISTORY OF HUMAN COMMUNICATION

Understanding where we have been is a key to the future. The opportunity to transform human communication on a global scale happens once in a hundred generations – every 2,000 years – and we are living during such a period (Fig. 1.1).

**Question 1: What are the distinctions between the digital medium and all of its hardcopy predecessors?**

For thousands of years Neolithic humans shared their life stories on cave walls (with smoke handprints and colored animal drawings) or on rocks (with stick figures and symbols) etched for future generations. Immovable, these images on stone have weathered the test of time.

Then, nearly 5,500 years ago, clay tablets awakened a new capacity for humans to share experiences and insights. Rolling devices – the ancestor of all typesetting – enabled humans to imprint and reproduce symbols in clay. Clay also had the advantage of being much easier to transport than stone, but it was more fragile.

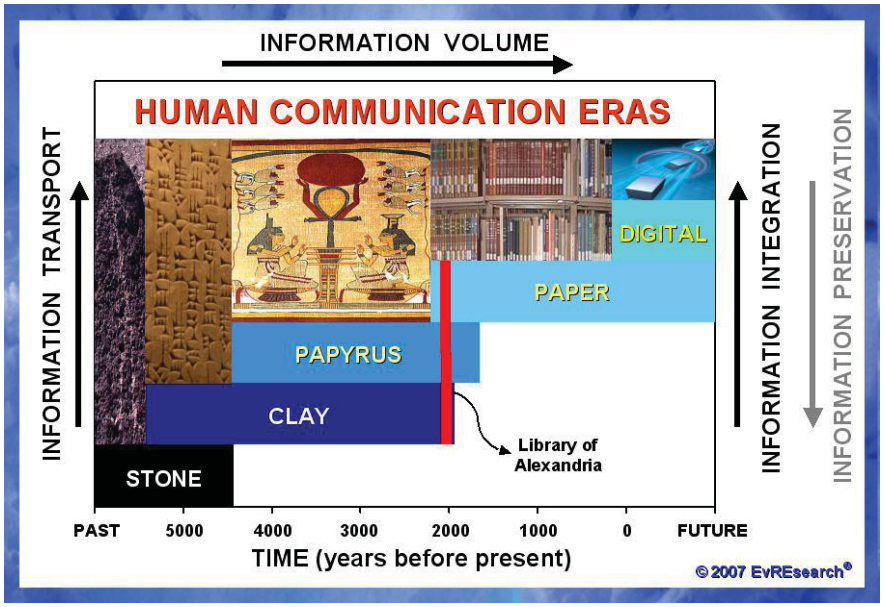


FIGURE 1.1: Eras in our civilization based on the media that humans have used to communicate beyond face-to-face. Each new communication medium has increased human capacity to: (a) transport information across time and space; (b) produce more information faster; and (c) integrate information into relational schema. Conversely, information has become more ethereal and difficult to preserve from stone to digital. Modified from Berkman et al. (2006a,b).

A thousand years later, humans invented papyrus to exchange information with much greater detail and color than ever before. Papyrus was lighter and more pliable than clay, which made it easier to distribute. Pieces of papyrus also could be combined to create complex information sources.

After another two millennia, we saw the advent of paper, which certainly must rank as one of the most significant inventions in our civilization. During this period with the Great Library of Alexandria, clay, papyrus, and paper coexisted as media to share data and other information beyond face-to-face communication. On a global scale, paper then took off as the principal medium for communicating across space and time.

Until the invention (or rather harnessing) of electricity, paper was unrivaled in the role of sharing knowledge in our world. Then came digital devices to collect, store, transmit, and display information. It has only been in the past fifty years that digital devices have become the communication backbone in our world information society.

Each era of global communication, from stone to digital (Fig. 1.1), has been accompanied by a threshold increase in human capacity to transport information. Similarly, each new communication medium has significantly increased our capacity to produce information, as indicated by the relative volumes of information that emerged. Moreover, the ability to integrate information has increased over time with tablets, folios, books, and now websites.

In contrast, the most resilient medium was stone with petroglyphs and pictographs that have stood the test of time through rain, snow, wind, and even fire. Subsequent media have been much more fragile. In fact, the digital medium has been like a black hole where most of the information produced has been lost because of limited preservation strategies and rapid obsolescence of storage devices.

Over the past 6,000 years, there has been global transformation in the information management medium every couple millennia. Paper was most recent with its invention in China around 2,000 years ago, curiously near the start of the Common Era that has since marked time across our civilization. If the past is any indication of the future, the digital medium will be with us for millennia to come. The challenge is to manage our digital information and to facilitate knowledge discovery for the benefit of future generations around the world.

## KNOWLEDGE MANAGEMENT AND DISCOVERY

Looking backward through time, we recognize that information in our civilization has been managed largely through libraries and archives. While similar in their needs to facilitate information access and preservation, these two architectures possess fundamental differences. Archives manage information based on the *context* of records linked to specific activities and transactions, like the Bureau of Motor Vehicle records of your car title. Libraries largely manage information based on the *content* of the information resources, as with the subject categories in the Dewey Decimal System. Beyond content and context there is a third element of information to establish meaning and that is its *structure* (Fig. 1.2).

**Question 2: Are there unique aspects of the digital medium that will enhance knowledge management and discovery?**

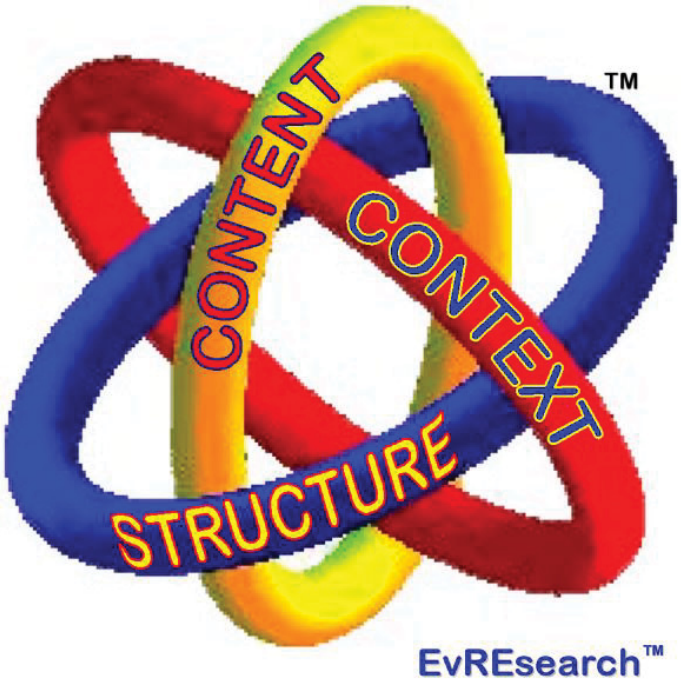


FIGURE 1.2: "Borromean Rings of Meaning" illustrate the three inseparable elements of information (content, context, and structure) that provide the basis for understanding and synthesizing knowledge. From Berkman et al. (2006a,b).

For example, when a message is encrypted (i.e., the *structure* is altered) it still has *content* and *context*, but no meaning absent the key to unlock the encryption. Alternatively, if the names or dates and places are removed from an information resource, it still has *context* and *structure*, but limited meaning without the salient facts. Similarly, meaning will be compromised by removing the context that can be used to authenticate an information resource or establish its provenance.

The paradigm shift created by digital technologies is the opportunity to dynamically utilize the *structure* of information as well as its *content* and *context* for the purposes of knowledge management and discovery. A hardcopy book can be managed based on its *content* (as in libraries) or its *context* (as in archives). However, it is not possible to automatically break a printed book into smaller granules of information (chapters, pages, paragraphs, etc.) that can be managed or discovered independently.

With the digital medium it has become possible to utilize the *content* and *context* as well as *structural* patterns (such as the white space formed by an indent or carriage return) to manage sets, subsets, and supersets of information resources. It is this ability to

dynamically manage the granularity of information that distinguishes the digital medium from all of its hardcopy predecessors in our civilization (Fig. 1.1).

*Content*, *context*, and *structure* of information create meaning that can be interpreted across a spectrum of understanding (Liebowitz 1999). The value of information is that it provides the foundation to synthesize knowledge that enables individuals to determine the course of their actions. Knowledge, which can be simply defined in terms of information relationships, is the epitome of learning (Bloom 1956) and the aspiration of all educated people.

## DIGITAL INFORMATION SUSTAINABILITY

Digital libraries and archives, which are emerging around the world (Arms 2000; Thibodeau 2001; NDIIPP 2002; Greenstein and Thorin 2002; Hodges et al. 2003; Lesk 2004; Duranti 2005), reflect the issues of sustainability. The following lessons are from the *National Science, Technology, Engineering and Mathematics Education Digital Library*, or NSDL, (http://www.nsdl.org) that originated in 2000 as a "community based endeavor" supported by the National Science Foundation (http://www.nsf.gov).

The NSDL established a "working structure" with a *Core-Integration Team*, *Policy Committee*, five *Standing Committees*, a *National Visiting Committee* and other entities as approved by an *Assembly* of the projects (http://sustain.comm.nsdl.org/). Supported projects contribute to the NSDL program by producing collections and services that have value to user, producer, and sponsor communities. Technical innovations are woven throughout so that the digital library can be effectively operated and applied. Generalizing, the NSDL "working structure" reveals underlying sustainability elements of any digital information organization (Table 1.1).

| TABLE 1.1: Sustainability Elements of Digital Information Organizations[a] | |
|---|---|
| **ELEMENT** | **SCOPE OF ACTIVITIES** |
| Program | Long-term administrative strategies for collaboration among developers, users, sponsors, and other stakeholders to "anchor" the digital information organization |
| Projects | Public-private-university-government strategies to support the creation, maintenance, funding and evolution of needed collections and services |
| Communities | Engagement, networking, and evaluation strategies to meet the demands of users, developers, and sponsors |
| Technical | Application strategies to achieve long-term preservation, access, and knowledge discovery with digital information |
| [a] See the Sustainability Standing Committee homepage (http://sustain.comm.nsdl.org/). Adapted from Berkman (2004). | |

Organizational strategies to implement the NSDL are further reflected by the projects that have been funded, effectively in two phases before and after 2003 (Fig. 1.3). Between 2000 and 2003, NSDL funded 88 collection, 45 service, 29 *Core Integration*, and 19 research projects. In 2004, characteristics of the NSDL conceptually changed with elimination of the track for collection projects and the emergence of pathways projects "to provide stewardship for the content and services needed by major communities of learners" (http://www.nsdl.org). From 2004 to 2006, there have been an additional 31 *Core Integration*, 21 pathways, 22 service, and eight research project awards. Together, these NSDL awards have been distributed across 35 states (NSDL 2007).

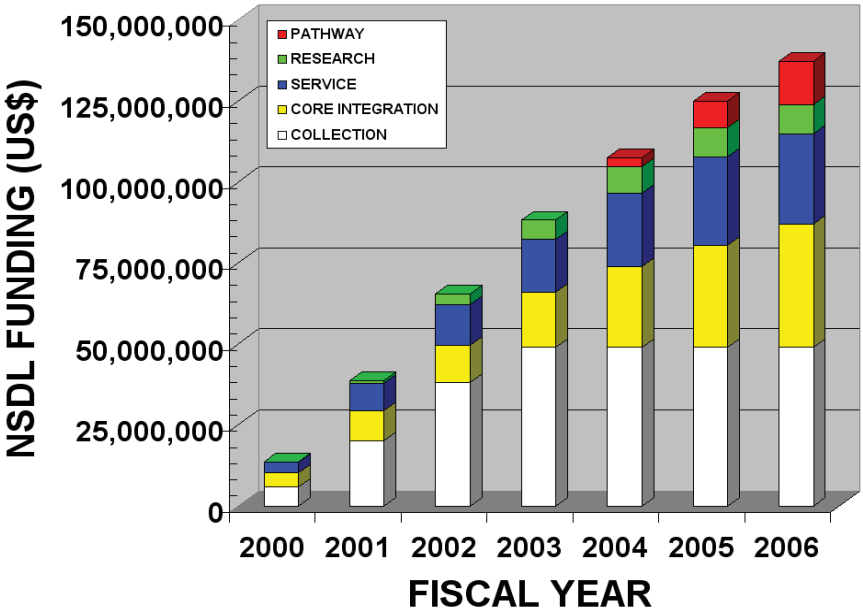


FIGURE 1.3: Cumulative funding by the National Science Foundation for different types of projects (legend) in the National Science, Technology, Engineering, and Mathematics Education Digital Library (http://www.nsdl.org). Data are from NSDL (2007).

In addition to conceptual changes, the shift in organizational emphasis before and after 2003 is represented by the relative support for *Core Integration*, which is responsible for integrating the NSDL projects. During the 2000-2003 period, *Core Integration* accounted for 16% of the projects and 19% of the NSDL funding. Afterward, these percentages increased to 34% and 43%, respectively. These adjustments in the NSDL reflect the distributed-centralized continuum of architectures that can be implemented for digital information organizations in general.

**Question 3: What is the optimal allocation of resources to balance the elements (Table 1.1) that are needed to sustain a digital information organization?**

## FUNDING PUBLIC GOODS

To better understand the economics of digital libraries, stories from NSDL projects that were considered to be sustainable were captured in a series of written vignettes (Table 1.2). These projects all existed prior to 2000 and provide potential anchors for long-term development of the NSDL organization, which is why many of them have received pathways funding.

**TABLE 1.2: Matrix of "Sustainability Vignettes" Written for the NSDL[b]**

| NSDL PROJECT | USERS | FUNDING | STRUCTURE |
|---|---|---|---|
| **Earth Science Information Partnership Federation:** http://esipfed.org Formed 1998 | 370 on list server; 83 partners include national data centers | government, meeting registration | not-for-profit corporation (federated partnership) |
| **Electronic Environmental Resources Library:** http://eerl.org Formed 1994 | Educators, librarians (about 3,000 visitors/day) | government, university gifts, corporations | not-for-profit corporation |
| **Journal of Chemical Education:** http://jce.divched.org JCE founded 1924. NSDL pathways funding 2006 | Chemical science teachers (about 12,000) | government, corporation, subscription, advertising | not-for-profit corporation (division within professional society) |
| **The Macaulay Library:** http://birds.cornell.edu/macaulaylibrary/ Audio collection initiated 1930s with Cornell Laboratory of Ornithology | museums, science centers, educators, researchers, corporations | government, university, gifts, sales | not-for-profit corporation (membership organization within university) |
| **Mathematical Association of America Digital Library:** http://mathdl.ma.org & **Math Gateway:** http://mathgateway.maa.org MAA incorporated 1920. NSDL pathways funding 2004 | about 15,000 visitors daily | government | not-for-profit corporation |
| **WGBH – Teachers' Domain:** http://teachersdomain.org WGBH radio began 1951. NSDL pathways funding 2004. | K-12 teachers and students (about 60,000 registered) | government, corporations, gifts, licensing | not-for-profit corporation (department within local media network) |

[b]See the Sustainability Standing Committee homepage – http://sustain.comm.nsdl.org/

All of the vignette projects involve not-for-profit corporations, suggesting that a corporate framework is necessary for large or small digital information organizations to manage their fiscal and

legal responsibilities in a sustainable manner. Moreover, all of the vignette projects involve government funding to produce results that can be openly disseminated, which effectively makes them *"public goods"* (Varian 1998, Stiglitz 1999). As such, these projects produce non-rival resources that can be consumed by anyone without diminishing the availability for others.

A significant hurdle for the NSDL, as with many digital information organizations, is to leverage current support into future revenue streams that will promote its long-term stability. Government agencies, universities, and other institutions with public mandates, resilient infrastructures, and access to long-term support may provide societal anchors to sustain networks of digital information resources. Philanthropic contributions, as with the Carnegie libraries (Bobinski 1969, Slyck 1995), also may be part of the solution. Moreover, sustainability likely will involve strategies to sell valued information goods and services (Stein 2007), such as providing access to scholarly journals through online databases (http://www.jstor.org/).

**Question 4: How is value established with digital information organizations that user, sponsor, and developer communities (Table 1.1) will financially support?**

## CONCLUSION

From stone to digital (Fig. 1.1), each era of global communication has been accompanied by a threshold increase in human capacity to transport, produce, and integrate information. As a civilization, our legacy is wrapped into this information that historically has been safeguarded in libraries and archives.

However, we have yet to build the information management architectures that will effectively preserve digital information (Boeke 2006). Technical difficulties with long-term preservation underscore the challenges to sustain digital information over decades, let alone centuries and millennia. The above types of questions underlie the technical, organizational, and economic issues that must be considered to sustain digital information organizations.

Practical strategies to sustain digital information in the public good will come from targeted discussions that engage stakeholder experts throughout society to think out-of-the-box into the distant future. Along these lines, in January 2005, a national task force on digital library sustainability was proposed to twelve federal

agencies through the *Federal Science and Technology Information Managers Group* (http://www.cendi.gov/minutes/pa_0105.html). The closing panel of the NSDL annual meeting in October 2006 and a subsequent discussion at the Library of Congress (http://www.digitalpreservation.gov/) in November 2006 further revealed actionable interest in implementing such a task force (minutes of meetings can be accessed through the NSDL Sustainability Standing Committee homepage: http://sustain.comm.nsdl.org).

We are living during a rare transition between global communication eras – which happens once in a hundred generations (Fig. 1.1) – and there is no roadmap. It is clear, however, that digital information sustainability is essential to the knowledge management and discovery opportunities that will empower an enlightened society.

Our generation has serious responsibilities to manage digital information into the future for, as observed by the convener of the *United Nations World Summit on the Information Society* (http://www.itu.int/wsis/), Adama Samassekou (personal communication 2004):

***"Knowledge is the common wealth of humanity."***

## ACKNOWLEDGEMENTS

I thank the diverse members of the Sustainability Standing Committee (SSC) for sharing insights, passion and creativity to sustain the *National Science, Technology, Engineering and Mathematics Education Digital Library* (NSDL). The opportunity to craft this paper, as chair of the SSC, was generously supported by the NSDL (Grant No. NSF / DUE 0329044).

## REFERENCES

Arms, William Y. 2000. *Digital libraries.* Cambridge: MIT Press.

Berkman, Paul A. 2004. Sustaining the National Science Digital Library. *Project Kaleidoscope Newsletter.* August 20, 2004 http://www.pkal.org/template2.cfm?c_id=1383.

Berkman, Paul A., George J. Morgan, Reagan Moore, Babak Hamidzadeh. 2006a. Automated granularity to integrate digital records: The "Antarctic treaty searchable database" case study. *Data Science Journal* 5:84-99. http://www.jstage.jst.go.jp/article/dsj/5/0/84/_pdf.

--. 2006b. Automated granularity to integrate digital records: The "Antarctic treaty searchable database" case study. *Data Science*

*Journal* 5:84-99. Translation – Archivo Municipal, Cartagena, Spain. http://archivo.cartagena.es/recursos/texto0_antarctica_dos.pdf.

Bobinski, George. 1969. *Carnegie Libraries: Their History and Impact on American Public Library Development.* Chicago: American Library Association.

Boeke, Cindy. 2006. IPRES 2006 conference report: Digital preservation takes off in the e-environment. *D-Lib Magazine* 12, no. 12. http://www.dlib.org/dlib/december06/boeke/12boeke.html.

Bloom Benjamin S. 1956. *Taxonomy of educational objectives, handbook I: The cognitive domain.* New York: David McKay.

Duranti. Luciana (ed.). 2005. *The long-term preservation of authentic electronic records: Findings of the InterPARES Project.* San Miniato: Archilab. http://www.interpares.org/book/index.htm.

Greenstein, David I. and Thorin, Suzanne E. 2002. *The Digital Library: A Biography. 2nd Edition.* Washington, D.C.: Digital Library Federation.

Hodges, Patricia, Maria Bonn, Mark Sandler, and John P. Wilkin, (eds.). 2003. *Digital libraries: A vision for the 21st century.* Ann Arbor: University of Michigan Scholarly Publishing Office.

Lesk, Michael. 2004. *Understanding digital libraries. 2nd Edition.* New York: Morgan Kaufmann.

Liebowitz, Jay (ed.). 1999. *Knowledge management handbook.* Boca Raton: CRC Press.

NSDL. 2007. *2006 annual report: Leveraging collaborative networks.* Boulder: National Science Digital Library. http://nsdl.org/about/download/misc/NSDL_ANNUAL_REPORT_2006.pdf.

NDIIPP. 2002. *Preserving our digital heritage: Plan for the national digital information infrastructure and preservation program.* Washington, D.C.: Library of Congress.

Stein, Seth. 2007. Education, Outreach and Marketing. *EOS Transactions of the American Geophysical Union.* 88, no. 4:39-40. http://www.agu.org/pubs/crossref/2007/2007EO040007.shtml.

Stiglitz, Joseph E. 1999. Knowledge as a public good. In *Global public goods*, eds. Kaul, Inge, Isabelle Grunberg, and Marc Stern, 308-326. Oxford: Oxford Scholarship Online Monographs.

Slyck, Abigail A.V. 1995. *Free to all: Carnegie libraries and American culture, 1890-1920.* Chicago: University of Chicago Press.

Thibodeau, Kenneth. 2001. Building the archives of the future: Advances in preserving electronic records at the National Archives and Records Administration. *D-Lib Magazine* 7, no. 2. http://www.dlib.org/dlib/february01/thibodeau/02thibodeau.html.

Varian, Hal R. 1998. *Markets for information goods.* Berkeley: University of California.