

Digital library

From Wikipedia, the free encyclopedia

Jump to: [navigation](#), [search](#)



This article **needs additional citations for verification**. Relevant discussion may be found on the [talk page](#). Please help [improve this article](#) by adding citations to [reliable sources](#). Unsourced material may be [challenged](#) and [removed](#). *(March 2012)*

A **digital library** is a [library](#) in which collections are stored in digital formats (as opposed to print, [microform](#), or other media) and accessible by computers.^[1]^[not in citation given] The digital content may be stored locally, or accessed remotely via computer networks. A digital library is a type of [information retrieval](#) system.

In the context of the [DELOS](#), a Network of Excellence on Digital Libraries, and [DL.org](#), a Coordination Action on *Digital Library Interoperability, Best Practices and Modelling Foundations*, Digital Library researchers and practitioners produced a **Digital Library Reference Model**^{[2][3]} which defines a digital library as: "A potentially virtual organisation, that comprehensively collects, manages and preserves for the long depth of time rich digital content, and offers to its targeted user communities specialised functionality on that content, of defined quality and according to comprehensive codified policies."^[4]

The first use of the term *digital library* in print may have been in a 1988 report to the [Corporation for National Research Initiatives](#)^[5]^[not in citation given] The term *digital libraries* was first popularized by the [NSF/DARPA/NASA](#) Digital Libraries Initiative in 1994.^[6] These draw heavily on [As We May Think](#) by [Vannevar Bush](#) in 1945, which set out a vision not in terms of technology, but user experience.^[7] The term *virtual library* was initially used interchangeably with *digital library*, but is now primarily used for libraries that are virtual in other senses (such as libraries which aggregate distributed content).

A distinction is often made between content that was created in a digital format, known as [born-digital](#), and information that has been converted from a physical medium, e.g. paper, by [digitizing](#). The term [hybrid library](#) is sometimes used for libraries that have both physical collections and digital collections. For example, [American Memory](#) is a digital library within the [Library of Congress](#). Some important digital libraries also serve as long term archives, such as [arXiv](#) and the [Internet Archive](#).

Contents

- [1 Academic repositories](#)
- [2 Digital archives](#)
- [3 The future](#)
- [4 Searching](#)

- [5 Frameworks](#)
- [6 Construction and organization](#)
 - [6.1 Software](#)
 - [6.2 Digitization](#)
- [7 Advantages](#)
- [8 Challenges](#)
 - [8.1 Digital preservation](#)
 - [8.2 Copyright and licensing](#)
 - [8.3 Metadata creation](#)
- [9 See also](#)
- [10 References](#)
- [11 External links](#)
 - [11.1 Conferences](#)

Academic repositories

Many academic libraries are actively involved in building [institutional repositories](#) of the institution's books, papers, theses, and other works which can be digitized or were 'born digital'. Many of these repositories are made available to the general public with few restrictions, in accordance with the goals of [open access](#), in contrast to the publication of research in commercial journals, where the publishers often limit access rights. Institutional, truly free, and corporate repositories are sometimes referred to as digital libraries.

Digital archives

Physical [archives](#) differ from physical libraries in several ways. Traditionally, archives are defined as:

1. Containing [primary sources](#) of information (typically letters and papers directly produced by an individual or organization) rather than the secondary sources found in a library (books, periodicals, etc.).
2. Having their contents organized in groups rather than individual items.
3. Having unique contents.

The technology used to create digital libraries is even more revolutionary for archives since it breaks down the second and third of these general rules. In other words, "digital archives" or "online archives" will still generally contain primary sources, but they are likely to be described individually rather than (or in addition to) in groups or collections. Further, because they are digital their contents are easily reproducible and may indeed have been reproduced from elsewhere. The [Oxford Text Archive](#) is generally considered to be the oldest digital archive of academic physical primary source materials.

The future

Large scale digitization projects are underway at [Google](#), the [Million Book Project](#), and [Internet Archive](#). With continued improvements in book handling and presentation technologies such as [optical character recognition](#) and [ebooks](#), and development of alternative depositories and business models, digital libraries are rapidly growing in popularity. Just as libraries have ventured into audio and video collections, so have digital libraries such as the [Internet Archive](#).

According to Larry Lannom, Director of Information Management Technology at the nonprofit Corporation should be for National Research Initiatives, "all the problems associated with digital libraries are wrapped up in archiving." He goes on to state, "If in 100 years people can still read your article, we'll have solved the problem." Daniel Akst, author of *The Webster Chronicle*, proposes that "the future of libraries — and of information — is digital." Peter Lyman and Hal Varian, information scientists at the [University of California, Berkeley](#), estimate that "the world's total yearly production of print, film, optical, and magnetic content would require roughly 1.5 billion gigabytes of storage." Therefore, they believe that "soon it will be technologically possible for an average person to access virtually all recorded information."^[8]

Searching

Most digital libraries provide a search interface which allows resources to be found. These resources are typically [deep web](#) (or invisible web) resources since they frequently cannot be located by [search engine crawlers](#). Some digital libraries create special pages or [sitemaps](#) to allow search engines to find all their resources. Digital libraries frequently use the [Open Archives Initiative Protocol for Metadata Harvesting](#) (OAI-PMH) to expose their metadata to other digital libraries, and search engines like [Google Scholar](#), [Yahoo!](#) and [Scirus](#) can also use OAI-PMH to find these deep web resources.^[9]

There are two general strategies for searching a *federation* of digital libraries:

1. distributed searching, and
2. searching previously harvested [metadata](#).

Distributed searching typically involves a client sending multiple search requests in parallel to a number of servers in the federation. The results are gathered, duplicates are eliminated or clustered, and the remaining items are sorted and presented back to the client. Protocols like [Z39.50](#) are frequently used in distributed searching. A benefit to this approach is that the resource-intensive tasks of indexing and storage are left to the respective servers in the federation. A drawback to this approach is that the search mechanism is limited by the different indexing and ranking capabilities of each database, making it difficult to assemble a combined result consisting of the most relevant found items.

Searching over previously harvested metadata involves searching a locally stored [index](#) of information that has previously been collected from the libraries in the federation. When a search is performed, the search mechanism does not need to make connections

with the digital libraries it is searching - it already has a local representation of the information. This approach requires the creation of an indexing and harvesting mechanism which operates regularly, connecting to all the digital libraries and querying the whole collection in order to discover new and updated resources. [OAI-PMH](#) is frequently used by digital libraries for allowing metadata to be harvested. A benefit to this approach is that the search mechanism has full control over indexing and ranking algorithms, possibly allowing more consistent results. A drawback is that harvesting and indexing systems are more resource-intensive and therefore expensive.

Frameworks

The formal reference models include the DELOS Digital Library Reference Model (Agosti, et al., 2006)^[10] and the Streams, Structures, Spaces, Scenarios, Societies (5S) formal framework.^[11] The Reference Model for an Open Archival Information System ([OAIS](#)) provides a framework to address [digital preservation](#).^[12]

Construction and organization

See also [Digital Collections Selection Criteria](#).

Software

There are a number of software packages for use in general digital libraries, for notable ones see [Digital library software](#). Institutional repository software, which focuses primarily on ingest, preservation and access of locally produced documents, particularly locally produced academic outputs, can be found in [Institutional repository software](#).

Digitization

In the past few years, procedures for [digitizing](#) books at high speed and comparatively low cost have improved considerably with the result that it is now possible to digitize millions of books per year.^[13]

Advantages

The advantages of digital libraries as a means of easily and rapidly accessing books, archives and images of various types are now widely recognized by commercial interests and public bodies alike.^[14]

Traditional libraries are limited by storage space; digital libraries have the potential to store much more information, simply because digital information requires very little physical space to contain it. As such, the cost of maintaining a digital library can be much lower than that of a traditional library. A physical library must spend large sums of money paying for staff, book maintenance, rent, and additional books. Digital libraries may reduce or, in some instances, do away with these fees. Both types of library require

cataloguing input to allow users to locate and retrieve material. Digital libraries may be more willing to adopt innovations in technology providing users with improvements in electronic and audio book technology as well as presenting new forms of communication such as wikis and blogs; conventional libraries may consider that providing online access to their OPAC catalogue is sufficient. An important advantage to digital conversion is increased accessibility to users. They also increase availability to individuals who may not be traditional patrons of a library, due to geographic location or organizational affiliation.

- **No physical boundary.** The user of a digital library need not to go to the library physically; people from all over the world can gain access to the same information, as long as an Internet connection is available.
- **Round the clock availability** A major advantage of digital libraries is that people can gain access 24/7 to the information.
- **Multiple access.** The same resources can be used simultaneously by a number of institutions and patrons. This may not be the case for copyrighted material: a library may have a license for "lending out" only one copy at a time; this is achieved with a system of [digital rights management](#) where a resource can become inaccessible after expiration of the lending period or after the lender chooses to make it inaccessible (equivalent to returning the resource).
- **Information retrieval.** The user is able to use any search term (word, phrase, title, name, subject) to search the entire collection. Digital libraries can provide very user-friendly interfaces, giving clickable access to its resources.
- **Preservation and conservation.** Digitization is not a long-term preservation solution for physical collections, but does succeed in providing access copies for materials that would otherwise fall to degradation from repeated use. Digitized collections and born-digital objects pose many preservation and conservation concerns that analog materials do not. Please see the following "Problems" section of this page for examples.
- **Space.** Whereas traditional libraries are limited by storage space, digital libraries have the potential to store much more information, simply because digital information requires very little physical space to contain them and media storage technologies are more affordable than ever before.
- **Added value.** Certain characteristics of objects, primarily the quality of images, may be improved. Digitization can enhance legibility and remove visible flaws such as stains and discoloration. ^[15]
- **Easily accessible.**

Challenges



This article's [tone](#) or style may not reflect the encyclopedic tone used on **Wikipedia**. See Wikipedia's [guide to writing better articles](#) for suggestions. (May 2008)

Digital preservation

Main article: [Digital preservation](#)

Digital preservation aims to ensure that digital media and information systems are still interpretable into the indefinite future. Each necessary component of this must be migrated, preserved or [emulated](#).^[16] Typically lower levels of systems ([floppy disks](#) for example) are emulated, bit-streams (the actual files stored in the disks) are preserved and operating systems are emulated as a [virtual machine](#). Only where the meaning and content of digital media and information systems are well understood is migration possible, as is the case for office documents.^{[16][17][18]} However, at least one organization, the [WiderNet Project](#), has created an offline digital library, the [eGranary](#), by reproducing materials on a 4 [TB hard drive](#). Instead of a bit-stream environment, the digital library contains a built-in [proxy server](#) and [search engine](#) so the digital materials can be accessed using an [Internet browser](#).^[19] Also, the materials are not preserved for the future. The eGranary is intended for use in places or situations where Internet connectivity is very slow, non-existent, unreliable, unsuitable or too expensive.

Copyright and licensing

Digital libraries are hampered by [copyright](#) law because, unlike with traditional libraries, digital libraries do not have access to works from every time period. The republication of material on the web by libraries may require permission from rights holders, and there is a conflict of interest between libraries and the publishers who may wish to create online versions of their acquired content for commercial purposes. In the year 2010 it was estimated that twenty-three percent of books in existence were created before 1923 and thus out of copyright. Of those printed after this date, only five percent were still in print as of 2010. Thus, approximately seventy-two percent of books were not available to the public.^[20]

There is a dilution of responsibility that occurs as a result of the distributed nature of digital resources. Complex intellectual property matters may become involved since digital material is not always owned by a library.^[21] The content is, in many cases, [public domain](#) or self-generated content only. Some digital libraries, such as [Project Gutenberg](#), work to digitize out-of-copyright works and make them freely available to the public. An estimate of the number of distinct books still existent in library catalogues from 2000 BC to 1960, has been made.^{[22][23]}

The [Fair Use Provisions \(17 USC § 107\)](#) under the [Copyright Act of 1976](#) provide specific guidelines under which circumstances libraries are allowed to copy digital resources. Four factors that constitute fair use are "Purpose of the use, Nature of the work, Amount or substantiality used and Market impact."^[24]

Some digital libraries acquire a license to lend their resources. This may involve the restriction of lending out only one copy at a time for each license, and applying a system of [digital rights management](#) for this purpose (see also above).

The [Digital Millennium Copyright Act](#) of 1998 was an act created in the United States to attempt to deal with the introduction of digital works. This Act incorporates two treaties from the year 1996. It criminalizes the attempt to circumvent measures which limit access to copyrighted materials. It also criminalizes the act of attempting to circumvent access control.^[25] This act provides an exemption for nonprofit libraries and archives which allows up to three copies to be made, one of which may be digital. This may not be made public or distributed on the web, however. Further, it allows libraries and archives to copy a work if its format becomes obsolete.^[26]

Copyright issues persist. As such, proposals have been put forward suggesting that digital libraries be exempt from copyright law. Although this would be very beneficial to the public, it may have a negative economic effect and authors may be less inclined to create new works.^[27]

Metadata creation

In traditional libraries, the ability to find works of interest was directly related to how well they were catalogued. While cataloguing electronic works digitized from a library's existing holding may be as simple as copying or moving a record from the print to the electronic form, complex and born-digital works require substantially more effort. To handle the growing volume of electronic publications, new tools and technologies have to be designed to allow effective automated semantic classification and searching. While [full text search](#) can be used for some searches, there are many common catalog searches which cannot be performed using full text, including:

- finding texts which are translations of other texts
- linking texts published under pseudonyms to the real authors ([Samuel Clemens](#) and [Mark Twain](#), for example)
- differentiating non-fiction from parody (*[The Onion](#)* from *[The New York Times](#)*, for example)

See also



Wikiversity has learning materials about [Curriculum on Digital Libraries](#)

- [List of digital library projects](#)
- [List of online encyclopedias](#)
- [Category:Digital library software](#)
- [Category:Institutional repository software](#)
- [Digital library](#) article at [LISWiki](#), a [Library science wiki](#)
- [Libraries in Second Life](#)

References

1. [^](#) Greenstein, Daniel I., Thorin, Suzanne Elizabeth. [The Digital Library: A Biography](#). Digital Library Federation (2002) [ISBN 1-933645-18-0](#). Accessed June 25, 2007.
2. [^](#) L. Candela, G. Athanasopoulos, D. Castelli, K. El Raheb, P. Innocenti, Y. Ioannidis, A. Katifori, A. Nika, G. Vullo, S. Ross: *The Digital Library Reference Model*. April 2011 ([PDF](#))
3. [^](#) L. Candela et al.: *The DELOS Digital Library Reference Model - Foundations for Digital Libraries*. Version 0.98, February 2008 ([PDF](#))
4. [^](#) L. Candela, G. Athanasopoulos, D. Castelli, K. El Raheb, P. Innocenti, Y. Ioannidis, A. Katifori, A. Nika, G. Vullo, S. Ross: *The Digital Library Reference Model*. April 2011, 17. ([PDF](#))
5. [^](#) Kahn, R. E., & Cerf, V. G. (1988). [The Digital Library Project Volume I: The World of Knowbots, \(DRAFT\): An Open Architecture For a Digital Library System and a Plan For Its Development](#). Reston, VA: Corporation for National Research Initiatives.
6. [^](#) Edward A. Fox. [The Digital Libraries Initiative - Update and Discussion](#), Bulletin of the America Society of Information Science, Vol. 26, No 1, October/November 1999.
7. [^](#) Candela, L.; Castelli, D. & Pagano, [History, Evolution and Impact of Digital Libraries](#). In P. Iglezakis, I.; Synodinou, T. & Kapidakis, S. (ed.) *E-Publishing and Digital Libraries: Legal and Organizational Issues*, IGI Global, 2011, 1- 30
8. [^](#) Akst, D. (2003). The Digital Library: Its Future Has Arrived. *Carnegie Reporter*, 2(3), 4-8.
9. [^](#) Koehler, AEC. *Some Thoughts on the Meaning of Open Access for University Library Technical Services* *Serials Review* Vol. 32, 1, 2006, p. 17
10. [^](#) Agosti, M., Candela, L., Castelli, D., Ferro, N., Ioannidis, Y., Koutrika, G., Meghini, C., Pagano, P., Ross, S., Schek, H.-J., & Schuldt, H. (2006). A Reference Model for DLMSs Interim Report. In L. Candela, & D. Castelli (Eds.), Deliverable D1.4.2 - Reference Model for Digital Library Management Systems [Draft 1]. DELOS, A Network of Excellence on Digital Libraries -- IST-2002-2.3.1.12, Technology-enhanced Learning and Access to Cultural Heritage. Online at: http://146.48.87.122:8003/OLP/Repository/1.0/Disseminate/delos/2006_WP1_D1_42/content/pdf?version=1
11. [^](#) Gonçalves, M. A., Fox, E. A., Watson, L. T., & Kipp, N. A. (2004). Streams, Structures, Spaces, Scenarios, Societies (5S): A Formal Model for Digital Libraries. *ACM Transactions on Information Systems (TOIS)*, 22 (2), 270-312.
12. [^](#) "The DSpace team recognized the value of the OAIS framework and recast the repository's architecture to accommodate this archival framework" Baudoin, P.; M. Branschovsky (2004), [MIT's DSpace experience: a case study](#)
13. [^](#) Committee on Institutional Cooperation: [Partnership announced between CIC and Google](#), 6 June 2007, Retrieved 7.
14. [^](#) [European Commission steps up efforts to put Europe's memory](#) on the Web via a "European Digital Library" Europa press release, 2 March 2006

15. [^](#) Gertz, Janet. "Selection for Preservation in the Digital Age." *Library Resources & Technical Services*. 44(2) (2000):97-104.
16. [^](#) [a](#) [b](#) Cain, Mark. "Managing Technology: Being a Library of Record in a Digital Age", *Journal of Academic Librarianship* 29:6 (2003).
17. [^](#) Breeding, Marshall. "Preserving Digital Information.". *Information Today* 19:5 (2002).
18. [^](#) Teper, Thomas H. "Where Next? Long-Term Considerations for Digital Initiatives." *Kentucky Libraries* 65(2)(2001):12-18.
19. [^](#) ["WiderNet: About the eGranary"](#).
20. [^](#) Van Le, Christopher, "Opening the Doors to Digital Libraries: A Proposal to Exempt Digital Libraries From the Copyright Act," *Case Western Reserve Journal of Law, Technology & The Internet*, 1.2 (Spring 2010),135.
21. [^](#) Pymm, Bob. "Building Collections for All Time: The Issue of Significance." *Australian Academic & Research Libraries*. 37(1) (2006):61-73.
22. [^](#) [Antique Books](#)
23. [^](#) Kelly, Kevin (2006-05-14). ["Scan This Book!"](#). [New York Times Magazine](#). Retrieved 2008-03-07. "When Google announced in December 2004 that it would digitally scan the books of five major research libraries to make their contents searchable, the promise of a [universal library](#) was resurrected. ... From the days of Sumerian clay tablets till now, humans have "published" at least 32 million books, 750 million articles and essays, 25 million songs, 500 million images, 500,000 movies, 3 million videos, TV shows and short films and 100 billion public Web pages."
24. [^](#) Hirtle, Peter B., ["Digital Preservation and Copyright,"](#) *Stanford University Libraries*. Retrieved October 24, 2011.
25. [^](#) United States Copyright Office, ["The Digital Millennium Copyright Act of 1998 - U.S. Copyright Office Summary"](#) 1998, 2.
26. [^](#) United States Copyright Office, ["The Digital Millennium Copyright Act of 1998 - U.S. Copyright Office Summary"](#) 1998, 15.
27. [^](#) Van Le, Christopher, "Opening the Doors to Digital Libraries: A Proposal to Exempt Digital Libraries From the Copyright Act," *Case Western Reserve Journal of Law, Technology & The Internet*, 1.2 (Spring 2010),145.

External links

- [CNRI-DARPA: D-Lib Magazine](#) Electronic publication that primarily focuses on digital library research and development

Conferences

- [TPDL](#) - International Conference on Theory and Practice of Digital Libraries
- [ECDL](#) - European Conference on Digital Libraries
- [ICADL](#) - International Conference on Asian Digital Libraries
- [JCDL](#) - ACM and IEEE Joint Conference on Digital Libraries
- [ICSD](#) - International Conference for Digital Libraries and the Semantic Web

