Chapter XV
Digital Libraries

Micah Altman
Harvard University, USA

INTRODUCTION

Digital libraries are collections of digital content and services selected by a curator for use by a particular user community. Digital libraries offer direct access to the content of a wide variety of intellectual works, including text, audio, video, and data; and may offer a variety of services supporting search, access, and collaboration. In the last decade digital libraries have rapidly become ubiquitous because they offer convenience, expanded access, and search capabilities not present in traditional libraries. This has greatly altered how library users find and access information, and has put pressure on traditional libraries to take on new roles. However, information professionals have raised compelling concerns regarding the sizeable gaps in the holdings of digital libraries, about the preservation of existing holdings, and about sustainable economic models.

This chapter presents an overview of the history, advantages, disadvantages, and design principles relating to digital libraries, and highlights important controversies and trends. For an excellent comprehensive discussion of the use, cost and benefits of digital libraries see Lesk (2005), for further discussion of architectural and design issues see Arms (2000), and see Witten and Bainbridge (2002) for a detailed example of the mechanics of implementing a digital library.

BACKGROUND

In 1939, before the first digital computer system was designed, Vannevar-Bush, a professor of electrical engineering at MIT, proposed a system that in many ways foreshadowed modern digital libraries. (Bush, 1939, 1945) (Bush would become head of the Office of Scientific Research and Development during World War II and then one of the chief advocates for the creation of the National Science Foundation.) This system, the "Memex," was designed to microfilm entire libraries of books and journals, combine these with individuals' private notes and indexes, and make them available on the desktop. Bush envisioned that the Memex would enable users and information professionals to create new organizations of knowledge through 'associative trails,' links among parts of different

documents. Although this system was never built, Bush's ideas inspired generations of future computer scientists, including J.C.R. Licklider, who made fundamental contributions to the development of personal computer interfaces, artificial intelligence, the internet, and digital libraries. Licklider envisioned much of the design of modern digital libraries, including the integration of indexing, search, retrieval, and storage services. (Licklider, 1965)

Although lacking the characteristic search and direct access capabilities of modern digital libraries, social science data archives were, in a sense, the first digital libraries, since they maintained large collections of digital material and provided access to it to outside users. Many of these collections were started in the 1950s when social scientists realized that it was crucial that their research surveys, and so forth, were recorded in digital form, in order to be preserved for future research. (Bisco, 1970) In the 1970s through the late 80s, digital technology was adopted in most libraries, primarily in the form of OPAC's (online public access catalog), which replaced card catalogs. It was not until the early 1990s, when the burgeoning World Wide Web, made dramatically more useful by indexing services such as Lycos (one of the early and dramatic successes of Internet search), greatly accelerated the growth of digital libraries and brought the combination of access and content that is their modern hallmark.

Government funding was crucial to early developments in digital library technology, and continues to remain important. For example, the Lycos search engine emerged from work done by the Informedia project at Carnegie-Mellon, and the immensely popular Google search service emerged from Stanford's Interlib project. Both of these projects were initially funded under the Digital Library Initiative, a joint project of NSF, NASA, and DARPA. The two phases of this initiative sponsored some of the most innovative efforts in digital libraries across a decade. (see Griffin, 1998) Other U.S. government programs such as the National Digital Information Infrastructure Preservation program (NDIIPP), funded by the Library of Congress, and the NSF's National Science Digital Library, continue to support innovative research in this area. Other countries have also contributed funding, mostly focused on the digitization of content, although some organizations such as the U.K.'s JISC (Joint Information Systems Committee) have funded a mix of content and innovative research.

Search and information retrieval have long been significant components of digital libraries, and commercial search engines such as Google, Yahoo, and MSN are now extremely popular. Search engines do not, however, constitute digital libraries, which integrate collection management, access, and other services. Some notable examples of modern digital libraries include the arXiv preprint server (McKiernan, 2000); and the many online electronic journals collections made available through the JSTOR project (Guthrie, 2001) and by many of the major commercial and open publishers. (Also see D-lib magazine, which routinely highlights notable digital collections.) Moreover, within the last five years, software systems that provide complete digital library services have become available, including: Greenstone (Witten & Bainbridge, 2002), VDC (Altman, Andreev, Diggory, et al. 1999), Fedora (see Lagoze, Krafft, Payette, et al., 2005), and DSPACE (Tandley, Bass, Stuve, et al., 2003).

DIGITAL LIBRARIES: ADVANTAGES AND DESIGN

Digital collections have a number of distinct advantages over traditional physical collections. One of the largest is that they typically allow convenient access, at all hours, to a wide variety of materials, from any location that provides Internet access. By offering different forms of access, digital libraries may also expand the potential user community, by offering, for example, more convenient access to disabled users, digitizing fragile materials for popular use, or offering access to remote populations. As a case in point, by digitizing collections of older journal articles, JSTOR increased the usage of these articles by the same user communities by a factor of 10 (Guthrie, 2001). Offering open access to journals has been shown to further increase usage (See Wilinsky, 2005).