CHAPTER

INTRODUCTION TO DIGITAL LIBRARIES

1

INTRODUCTION

Digital libraries emerged in the early 1990s but were preceded by inspiring visions of innovative thinkers and several decades of intensive development within information technologies. The innovative ideas for using information technology to organize and disseminate knowledge go back to futuristic essays by H.G. Wells and the work of early information scientists Paul Otlet, Vannevar Bush, and J.C.R. Licklider (Grudin, 2011; Lynch, 2005; Rayward, 1994, 1997, 2005). Unveiling his vision of a "world brain," a universal encyclopedia, H.G. Wells wrote in 1938: "The time is close at hand when any student, in any part of the world, will be able to sit with his projector in his own study at his or her own convenience to examine *any* book, *any* document, in an exact replica" (p. 77). The ideas of H.G. Wells and Vannevar Bush captured popular imagination and inspired future information scientists and inventors, but the technology that led to the development of digital libraries turned out to be quite different from that which they envisioned.

It was progress in digital computing (rather than the analog machines proposed by Bush) and the growth of computer networks that have enabled the construction of digital libraries and remote access to digital representations of scholarly and cultural resources held in libraries, archives, and museums. Nonetheless, the futuristic visions, as reflected in H.G. Wells' quotation in the preceding paragraph became a reality to a certain extent. Nowadays, students can easily access books and other scholarly resources, including Wells' *World Brain,* from their laptops or mobile devices. Digital libraries transcend physical and technical barriers to give access to information resources and enable novel ways of examining and linking these resources together. The digitized version of Wells' book, its "exact replica," is technically available through the Google Book Project and HathiTrust Digital Library. Due to copyright restrictions, however, access to its content is limited to students and faculty whose universities are members of HathiTrust. Universal access to "*any* book, *any* document" as envisioned by Wells (1938, p. 77), is technically feasible but is currently constrained by social and legal barriers. As Michael Lesk (2012) points out, the technological obstacles that were dominant in the first phase of digital library development have been generally overcome by progress in computers and networking, but the legal and social challenges remain.

Digital libraries encompass a wide range of materials, from books to representations of three-dimensional artifacts. The content is either created digitally or converted from a variety of analog sources through digitization. Extensive digitization efforts have accompanied the construction of digital libraries to transform the wealth of traditional scholarly and cultural materials held in libraries, archives, and museums into a digital format. The conversion process is far from complete. Nonetheless, for the first time in the history of recorded knowledge, information resources can be free from physical carriers and are available in a uniform digital form, regardless of their original sources and types of

Discover Digital Libraries. http://dx.doi.org/10.1016/B978-0-12-417112-1.00001-6

presentations. David M. Levy (2003) observes, "a single medium or representational format (ones and zeros) is now capable of representing all the forms of talk we have so far managed to create: text and graphics, voice, and moving images. And a single device is capable of making all these forms manifest" (p. 36). Levy (2003) also notes that all artifacts and documents are fundamentally social, created and used in the context of human activities.

On one hand, this shift from analog to digital methods in recording, transmission, and storage offers tremendous benefits for access and new forms of interaction with text and image. On the other hand, it poses unique challenges for organizing, presenting, and preserving digital resources and serving user communities in virtual information spaces. The streams of ones and zeros don't have much value unless they can be transformed into useful and usable scholarly, educational, and personal resources. The organizational aspect became the center of attention in the early phase of digital library development as the library cataloging standards developed in the print environment did not translate well into the digital realm. The initial construction of digital libraries was accompanied by the explosion of new metadata schemas, a so-called metadata renaissance as described by Calhoun (2014). Digital libraries have emerged as complex systems that serve not only as repositories of digital objects with associated metadata but also as information systems in a networked environment providing search and retrieval mechanisms and supporting user interaction. The contributions from computer science and the advances in information search and retrieval have enhanced the functionality and technical capabilities of DL systems.

Multiple, and often competing, definitions of what is a DL have emerged in the library and computer science communities. The concept was extensively debated during the formative period in the 1990s. Researchers were trying to reconcile the mission and principles of traditional libraries with the digital format of information resources, distributed network access, and new interaction capabilities and at the same time address the uniqueness of digital libraries, especially in contrast to resources available through the open web. The debate on what constitutes a DL is very important, as it not only advances research and practice but also has broader implications for the evaluation of digital libraries and their educational and social use (Bawden and Rowlands, 1999; Borgman, 2000). Digital libraries are viewed as multidimensional phenomena consisting of multiple layers and building blocks, available in the distributed network environment, with resources and associated services developed, organized, and managed to support users' scholarly and educational activities as well as personal research. The social–technical perspective is applied to consider the complexity of digital libraries as systems of technology, documents, users, and practices existing in social contexts (Bishop et al., 2003).

The evolution of digital libraries is marked by several phases, comprising the early transitional projects and the formative decade in the 1990s, mass digitization in the 2000s, and the large-scale aggregations undertaken in the last few years. Digital libraries have evolved into complex, multilayered, distributed systems since the first digital collections were made available over the Internet in the mid-1990s. For the first two decades, digital libraries were constructed primarily as standalone entities with strong institutional ties to libraries, archives, and museums. This landscape of multiple, discrete, and dispersed collections proves to be challenging for resource discovery as it requires locating and searching individual digital libraries. The difficulties in resource discovery, however, have begun to be addressed in recent years. Large-scale digital libraries, such as the Digital Public Library of America, Europeana, or HathiTrust, aggregate content from smaller individual digital libraries and provide portals for global searching and retrieval.

The initial development of digital libraries was focused on building the technical infrastructure. At the time, they tended to be system centered and rarely incorporated research on user needs or evaluation from the user perspective. However, as digital libraries expanded, it became clear that they were

difficult to use and posed usability problems (Blandford et al., 2004; Borgman, 2003). Usability became an important area of research in the digital library field, as well as becoming a form of evaluation (Buttenfield, 1999; Chowdhury et al., 2006; Jeng, 2005). Buttenfield (1999) recognized early on that digital libraries are information systems that people use to satisfy information needs not easily met in a traditional library and called for the adaptation of a wide range of usability methods. Usability has been identified by users as the most important criterion in evaluating digital libraries (Xie, 2006). The perceptions of usability and usefulness also play an important role in user adoption of digital libraries for educational use (Liu and Luo, 2011; Matusiak, 2012).

The development of digital libraries occurred concurrently with the emergence of the web. Digital libraries have adopted some Internet technical standards and have become part of the global networking infrastructure. But in this new and dramatically changing environment, libraries lost not only their primary role as information providers but also a visible and unique identity (De Rosa et al., 2005; Law, 2009, 2011; Lagoze, 2010). Digital libraries are now part of a broader information landscape, often competing for users' attention with a multitude of other information resources. Lagoze (2010) argues persuasively that digital libraries, with their institutional affiliation and traditional information models, have been less responsive to user expectations and changing information behaviors. Digital scholarly publications have become a mainstream resource in academic research, but the adoption of digital libraries in personal research and for educational use has been limited (Bearman, 2007; Liu and Luo, 2011; Matusiak, 2012; McMartin et al., 2008). Digital libraries are relatively new phenomena, and, like many new and emergent information systems, they face challenges of discovery, acceptance, and utilization. The social aspects of digital libraries and support for users' scholarly and educational activities represent areas that require further attention in research and practice.

DEFINITIONS AND FRAMEWORKS OF DIGITAL LIBRARIES

The concept of a digital library, as an entity separating from a traditional library, emerged in the mid-1990s. Prior to that, many institutions maintained collections of purchased electronic resources or even digitized materials, but did not refer to them as separate collections or libraries (Schwartz, 2000). Several different terms, including "electronic library," "virtual library," "network-accessible libraries," or "libraries without walls" were used to describe the new phenomenon in the early phase of digital library development. The term "digital library" came from the National Information Infrastructure Initiative (Bearman, 2007; Lagoze, 2010). It quickly gained acceptance despite some concerns that the combination of "digital" and "libraries" was somewhat misleading and blurred the distinction between collections of network-accessible electronic resources and libraries as institutions (Lynch, 1993). *Communications of the ACM* devoted an entire issue to digital libraries in May 1995. Fox (1995), in the introductory article, noticed a shift in terminology from "electronic library" to "digital library."

The first definitions that originated in the research community focused on the digital content and enabling technologies. The library community joined the debate in the late 1990s, emphasizing traditional library functions and services. The subsequent definitions attempted to reconcile the views of researchers and librarians and provide multifaceted perspectives emphasizing heterogeneous content, technical capabilities, and new functionality supporting diverse communities of users. The concept of digital libraries as a sociotechnical construct captures the complex nature of user interaction with these systems in the dynamic context of social practices. Digital library frameworks have emerged in response to the difficulty of describing the complexity of these systems by a single definition.

This review is limited to selected definitions illustrating the evolution of understanding of digital libraries. The proliferation of digital library definitions makes a comprehensive review very difficult. Schwartz (2000) identifies 64 formal and informal definitions. The purpose of these intensive intellectual activities has been to understand the nature of the emerging phenomenon and to provide a theoretical foundation for research and future development of digital libraries. The evolution of the concepts of digital libraries demonstrates a shift from the early focus on the traditional library model and the system-centered approach to emphasizing the complexity of these systems, their multiple dimensions, and the social context of use.

EARLY DEFINITIONS

A significant part of early digital library research efforts concentrated on defining the new phenomenon in an attempt to articulate its purpose and find answers to the fundamental question: what is a digital library? (Levy, 2000; Lyman, 1996; Marcum, 1997). Researchers not only differ in their answers and offer competing visions, but they don't even agree on whether an explicit definition is possible or necessary (Greenstein, 2000). Many researchers emphasize that the term "digital library" evokes different meanings to different people (Borgman, 1999; Fox, 1995; Schwartz, 2000). The diversity of opinions stems from the fact that the early concepts of digital libraries combined the missions, techniques, and cultures of traditional libraries with the capabilities and cultures of computing and telecommunications (Marchionini, 2000).

The early definitions of digital libraries illustrate the tension between researchers and practitioners and further focus on the networking technologies, digital format of collections, organization, and promise of universal access. The Association of Research Libraries (ARL) provided one of the early definitions of a digital library. The ARL definition is based on a book by Karen Drabenstott, *Analytical Review of the Library of the Future*, published in 1994 (Drabenstott, 1994). The definition identifies common elements of a digital library:

- The digital library is not a single entity.
- The digital library requires technology to link the resources of many.
- The linkages between the many digital libraries and information services are transparent to the end users.
- Universal access to digital libraries and information services is a goal.
- Digital library collections are not limited to document surrogates; they extend to digital artifacts that cannot be represented or distributed in printed formats (Association of Research Libraries, 1995).

The aforementioned definition focuses on the digital nature of collections, enabling network technology, and issues of access. In this respect, it reflects the early stage of thinking about digital libraries. It was, however, widely adapted in research projects (Koohang and Ondracek, 2005; Schwartz, 2000; Xie, 2006) and incorporated into other definitions.

Fox (1995) emphasizes the great potential of digital libraries in fulfilling "the age-old dream of every human being: gaining ready access to humanity's store of information" (p. 23). He embraces the computer science approach along with the traditional library perspective and sees digital libraries as networked information systems carrying out the functions of libraries in a new way, and they offer new possibilities to organize and access information resources. Fox notices that the metaphor of a traditional library is "both empowering and constraining" and points out that much of the power of digital libraries is in access to actual objects (p. 25). Lesk (1997) also highlights their new capabilities, describing

them as the "powers we never had with traditional libraries" (p. 1). In his opinion, the great advantage of digital libraries lies not only in access but also in the organization of digital content enhanced by indexing and full-text retrieval. He emphasizes organization as a key element and simply defines a digital library as "a collection of information that is both digitized and organized" (p. 1).

Definitions offered by the library community shift the focus from the word "digital" to "library" and elaborate on the role and functions of libraries in the new digital environment. Digital libraries are not seen as new or unique phenomena but rather as extensions of traditional libraries delivering new types of information resources and offering new user services. The Digital Library Federation (DLF) presents a definition that de-emphasizes the digital nature of collections but stresses the functions and services offered by libraries as organizations. The DLF describes digital libraries as "organizations that provide resources, including the specialized staff, to select, structure, offer intellectual access to, interpret, distribute, preserve the integrity of, and ensure the persistence over time of collections of digital works so that they are readily and economically available for use by a defined community or set of communities" (Waters, 1998, para. 3). This definition reflects the conviction that, with time, there will be less emphasis on the digital nature of material and more on traditional roles of libraries. The DLF definition includes distinct traditional library roles—such as selection, intellectual control, providing access, and preservation—applied to digital works.

MULTIFACETED DEFINITIONS AND PERSPECTIVES

The definitions that emerged in the second decade of digital library development emphasize multiple facets of these systems and combine the technical components with services and social aspects. Borgman (1999), in her comprehensive overview of early definitions, points to the diverging views of the research and library communities. Librarians envision digital libraries as extensions or augmentations of traditional libraries with resources in digital format and new types of services, while computer scientists view digital libraries as enabling technologies and networks.

Research-oriented definitions tend to give a narrower view as they primarily concentrate on the technical aspects of digital format, information architecture, and information retrieval. Practice-oriented definitions see digital libraries in social and institutional contexts and emphasize services. Borgman proposes a definition that would bridge the two conflicting approaches. Her definition has two elements:

- Digital libraries are a set of electronic resources and associated technical capabilities for creating, searching and using information.
- Digital libraries are constructed, collected, and organized by (and for) a community of users, and their functional capabilities support the information needs and uses of that community (Borgman, 1999).

Arms (2000) also attempts to reconcile the computer science approach with the librarians' perspective. He offers a succinct definition of a digital library as a managed collection of information in a digital format, with associated services, and accessible over networks. Arms places importance on the quality of managed, curatorial collections, where digital objects are described and organized systematically and made available to the public through a searchable interface.

The concept of digital libraries as unique or complex phenomena has emerged in studies examining their actual use and in the context of constructing such systems. Marchionini (2000) stresses that

digital libraries have a combination of traditional library roles as well as aspects of computing. However, in his reflections on the multiyear evaluation of the Perseus Project, he describes digital libraries as "emergent complex systems" (p. 326). Witten and Bainbridge (2003) echo earlier definitions by underscoring the principles of selection and organization. They define a digital library as "a focused collection of digital objects, including text, video, and audio, along with methods for access and retrieval, and for selection, organization, and maintenance of the collection" (p. 6). The concepts of selection, organization, and management are central to the authors' understanding of digital libraries. Although these concepts reflect traditional values of librarianship, Witten and Bainbridge do not perceive digital libraries as extensions of existing institutions, simply computerized or digitized libraries, but rather as a unique phenomenon that offers new ways of creating knowledge.

The sociotechnical perspective shifts the focus of the debate from technical aspects to the social context of digital library use and evaluation (Bishop et al., 2003). The editors of the book *Digital Library Use: Social Practice in Design and Evaluation* highlight the complexity of digital library systems, especially if viewed as part of interactions with the larger world of work, study, and collaborative activities of users and developers. Digital libraries are broadly defined as "sociotechnical systems – networks of technology, information, documents, people, and practices" (Bishop et al., 2003, p. 1). The emphasis of this perspective is on the design of digital libraries, based on an understanding of user needs and activities and on their role in the processes of knowledge construction viewed in a broader social context. The authors comment on the relationship between digital and traditional libraries, noting that some digital libraries are an outgrowth of traditional libraries, while other digital libraries only relate to traditional libraries metaphorically. The view of digital libraries as a social–technical phenomenon is espoused by several authors of chapters in the *Digital Library Use: Social Practice in Design and Evaluation* book, with Levy (2003) examining the social nature of documents and Agre (2003) commenting on digital libraries embedded in the social world.

Lagoze et al. (2005) reflect on the state of digital library development in the age of Google and argue that digital libraries should move away from the legacy of the traditional library information model built around metadata repositories. The new information model should move beyond search and access functionality and enable creating collaborative and contextual environments where information resources are "shared, aggregated, manipulated, and refined" (Lagoze et al., 2005, para. 10). The authors don't propose yet another definition but describe digital libraries in terms of desired characteristics, including

- Selection of resources according to the criteria relevant to the digital library mission
- Services to facilitate the use of resources by the target community
- Collaborative features, allowing users to contribute knowledge and reuse resources
- Contextual features enabling the relationships between the resources

Furthermore, Lagoze et al. (2005) describe their work of extending the functionality of the National Science Digital Library (NSDL), where they propose a new, resource-centric information model for managing, manipulating, and processing content and metadata.

In a recent publication, Calhoun (2014) expands the understanding of digital libraries by incorporating the architecture of digital library systems and the concepts of open access. She proposes a practical definition that combines multiple components, including systems and services, managed collections of digital content "intended to serve the needs of defined communities" (p. 18), a system architecture centered on a repository, search features, and user interfaces. This definition focuses on the technical infrastructure built with repository systems, reflecting the current state of digital library development.

Table 1.1 Selected Concepts of Digital Libraries

Author(s)	Digital Library Components	Emphasis
Single-field perspective		
Association of Research Libraries (1995)	Not a single entity = bibliographic control + digital objects + enabling network technologies	Digital nature of collections; access to full-text documents; universal access
Fox (1995)	Networked information systems = information resources + new ways to organize + new ways to access and retrieve	Distributed networks; information retrieval; extended access to information resources
Digital Library Federation (Waters, 1998)	Organizations = distributed digital resources + staff + library services	Extension of traditional libraries as organizations; traditional library roles and services
Multifaceted perspectives		
Borgman (1999)	Digital libraries = digital resources + associated technical capabilities + network distribution User-centered approach = services supporting user needs	Digital format and enabling technologies; community of users; user support
Arms (2000)	Managed collection of digital information = resources in a digital format + associated services + network access	Curatorial responsibility: selection, organization, and preservation; user services
Witten and Bainbridge (2003)	Heterogeneous systems = digital resources in multiple modes of representation + metadata + methods for access and retrieval	Selection, organization, and maintenance; new ways of creating knowledge
Bishop et al. (2003)	Sociotechnical systems = networks of technology + information + documents + people + practices	Digital technology; knowledge work; social practices; user-centered approach
Lagoze et al. (2005)	Multilayered resource-centric model = network of selected resources + structural and semantic relationships	New information model going beyond search and access; collaborative and contextual environment
Calhoun (2014)	Digital libraries = systems and services + managed collections of digital content + repository-centered architecture	Open access; support for the advancement of knowledge and culture

Unlike the earlier perspectives, Calhoun's definition calls attention to digital libraries' purpose in supporting the "advancement of knowledge and culture" (p. 18) and the availability of their content in open access. Calhoun recognizes the importance of social roles of digital libraries, noting, "social roles and communities are more likely to abide over time; collections and enabling technologies are more likely to shift" (p. 19). Table 1.1 provides a summary of selected definitions and perspectives.

Table 1.2 Digital Library Frameworks

Name/Authors	Digital Library Components	Emphasis
The DELOS Manifesto (Candela et al., 2007b)	A three-tier framework = Digital libraries as virtual organizations + DLS that users interact + DLMS providing software infrastructure	Six fundamental concepts: content, user, functionality, quality, policy, and architecture
The 5S (Fox et al., 2012)	Complex systems defined in terms of Streams Structures Spaces Scenarios Societies	Theoretical constructs capturing the essence of an information lifecycle

DIGITAL LIBRARY FRAMEWORKS

Digital library frameworks have emerged in recent years as a sign of the digital library research field becoming more mature and in response to a growing realization that the complex nature of these systems and their multiple facets are impossible to capture by a single definition or perspective (Candela et al., 2007b). The authors of the frameworks build upon the previous conceptual work and advance the understanding of digital libraries as unique, multidimensional phenomena by identifying the core concepts and outlining the relationships between them. The goals of the theoretical models are to provide a foundation for digital library research, a common vocabulary, and to further the development of information models and such systems. The two frameworks reviewed here were developed by interdisciplinary teams of researchers. Table 1.2 provides a summary of their key features.

The DELOS Manifesto is a conceptual framework developed by the members of the DELOS Network of Excellence in Digital Libraries, a research group funded by the European Union (Candela et al., 2007b). *The Manifesto* identifies key concepts, sets the foundation to facilitate the integration of research, and supports the development of improved and more flexible digital library systems. Digital libraries are understood broadly as the center of intellectual activity that enables collaboration, communication, and other forms of dynamic interaction and research activities. As demonstrated in Fig. 1.1, the digital library framework consists of three interrelated tiers:

- Digital library—represents an organization that collects, manages, and preserves the rich digital content on behalf of users.
- Digital library system (DLS)—a software system that is based on a defined architecture and provides all functionality.
- Digital library management system (DLMS)—a generic software system that provides the infrastructure to produce and administer a digital library.

Six core concepts are associated with the proposed digital library framework: content, user, functionality, quality, policy, and architecture. The three-tier framework provides a systematic approach to all levels of digital library development and use. It distinguishes between the technical infrastructure and software needed to develop and administer instances of digital libraries, the digital library systems

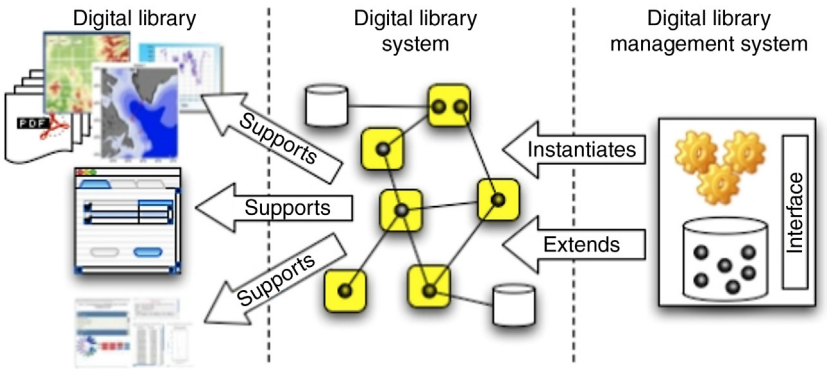

FIGURE 1.1 Digital Library, DLS, and DLMS: A Three-Tier Framework (Candela et al., 2007b)

that users interact with, and organization(s) responsible for collecting and managing digital content. The framework serves as a foundation for the Digital Library System Reference Architecture (Candela et al., 2007a). The Reference Model and DLMS are reviewed in Chapter 6. Candela et al. (2007b) also discuss the roles of actors in digital library systems: digital library end users, digital library designers, digital library system administrators, and digital library application developers. Digital library end users use the content and services via different digital library functions. Digital library designers define, customize, and maintain digital libraries based on their own expertise and knowledge. Digital library system administrators are responsible for the selection of software components, in particular the identification of the architectural configuration to construct the digital library system. Digital library application developers develop the software components of DLMS and DLS for different types of functionality.

The 5S—Societies, Scenarios, Spaces, Structures, Streams model provides a theoretical foundation for defining key constructs of digital libraries, which are viewed as complex systems of digital content, people, and technology. The authors propose a simple definition of digital libraries and define a set of abstractions representing the fundamental entities involved in the process of digital library development and use (Fox et al., 2012). This model builds on the authors' previous work in which the concepts of streams, structures, spaces, scenarios, and societies were defined (Gonçalves et al., 2004). Digital libraries are defined as complex systems that

- Help satisfy info needs of users (societies)
- Provide info services (scenarios)
- Organize info in usable ways (structures)
- Present info in usable ways (spaces)
- Communicate info with users (streams) (Fox et al., 2012, p. 6)

The focus of this definition is on the information lifecycle in which users perform tasks and interact with organized information sources in order to accomplish their goals. The theoretical elements of the framework are described as follows:

- Streams represent information flow and can be used to model content, which can be static (text, image) or dynamic (video).
- Structures support the organization of information in a usable and meaningful way.

- Spaces are collections of documents and are used in the context of access and presentation.
- Scenarios are used to describe user tasks and activities, which, in the context of digital libraries, can convey services.
- Societies refer to different communities of users; members have different roles and can undertake a range of activities.

The essential constructs are highly abstract and do not correspond directly to digital library concepts of digital objects, collections, services, etc., but can be used to define them. For example, digital objects can be described in terms of streams and structures. The *5S* framework provides a foundation for taxonomy of digital library terms and was used in developing an library and information science curriculum for educating future digital library professionals.

The review of research literature demonstrates an evolution of concepts since the emergence of digital libraries in the mid-1990s. The understanding of digital libraries has evolved from a one-dimensional perspective, seeing them as extensions of traditional libraries or network systems, to views of them as unique, multifaceted, and social phenomena that play an important role in knowledge construction. This review is based on previous work that was revised and expanded for the purpose of this chapter (Matusiak, 2010).

EVOLUTION OF DIGITAL LIBRARIES

"Digital libraries have a short yet turbulent and explosive history," remarked Tefko Saracevic in 2000 (p. 350). Although more than a decade passed, his statement is still true, especially in light of the long history of recording and organizing information resources. Despite an intensive period of development, digital libraries are a relatively new and emergent phenomenon. The history of digital libraries is divided into two decades, with some researchers placing the beginning in 1991 (Arms, 2012; Calhoun, 2014), while others point to 1994, as it was the year when systematic funding began to be available for digital library research and development (Dempsey, 2006; Lynch, 2005). The mid-1990s also coincided with the emergence of the web and the release of the first browsers. Lynch (2005), however, states that "very substantial digital library systems were developed prior to the World Wide Web" (para. 4). This section provides a brief overview of the evolution of digital libraries. It begins with highlighting the innovative ideas of early visionaries and key contributions of the pre-1990s era. Next, it traces the major phases in the development of digital libraries and reviews exemplary projects in the scientific and cultural heritage communities.

PRELUDE: EARLY VISIONARIES AND THE PRE-1990S ERA

Vannevar Bush and J.C.R. Licklider are widely recognized as digital library pioneers (Arms, 2000; Calhoun, 2014; Lesk, 2005, 2012). Lynch (2005) also points to H.G. Wells and Paul Otlet as early thinkers in the "prehistory" of digital libraries. H.G. Wells, a famous writer of science fiction, was also a utopian socialist who believed that some form of permanent world encyclopedia would bring together scattered intellectual resources and advance universal education. His idea of universal access to knowledge is formidable and some of the passages predicting the use of imaging as a duplication method to create "microscopic libraries of record" are indeed striking (Wells, 1938, p. 76). Rayward (1999) argues in his critical reassessment that Wells' vision of a "world brain," with its central control,

is rather troubling. Other researchers note a collaborative nature to Wells' universal encyclopedia and see it as a precursor to Wikipedia (Reagle, 2010).

Paul Otlet was a contemporary of H.G. Wells and a strong proponent of a universal encyclopedia as well. Otlet was a Belgian bibliographer, a founder of documentation, and an innovative thinker whose contributions to knowledge organization and information science, largely forgotten, have been rediscovered thanks to the historical studies of W. Boyd Rayward (1994, 1997). As Rayward (1997) points out, Otlet was concerned with the growth of publications and believed that technological innovations would provide a solution to storage, organization, and retrieval of records of knowledge. The technologies that he had at his disposal were index cards and microphotography. He constructed systems of interlinked bibliographic records, images, and excerpts from publications. These analog databases, or "repertories" as Otlet referred to them, were standardized and used Universal Decimal Classification as a common organizational schema. Rayward (1994) describes repertories as prototypes of hypertext systems consisting of "nodes or chunks organized by a system of links and navigational devices that allowed the movement of the user from bibliographic reference to full text to image and object" (p. 240). Otlet's writings also include innovative ideas about the power of multimedia in knowledge diffusion and concepts of mechanized information retrieval. His contributions are now acknowledged and afforded him recognition as a "forefather of the Internet" (Manfroid et al., 2013, p. 312).

Vannevar Bush, an American scientist and engineer, published his essay "As We May Think" in 1945 when he was the head of the US Office of Scientific Research and Development. In his position at the Office of Scientific Research and Development, Bush oversaw the Manhattan Project and witnessed firsthand the expansion of scientific research and its unprecedented role in warfare. In his essay, he commented on the exponential growth of scientific publications and increasing specialization and argued that traditional indexing methods were inadequate to meet the demands of modern scientists. He proposed organizing resources by association, which, in his opinion, reflected the way a human mind works. Furthermore, in reviewing advancements in photography and microfilming, he predicted the use of imaging for reproduction and efficient storage, which he called the "compression" of library collections in his words. Bush envisioned automating the process of storage and retrieval and using microfilm for copies of books and other documents stored in a new kind of device that he called a memex. He described it as a "sort of mechanized private file and library" (Bush, 1945, para. 55). The memex was meant to extend researchers' memory through a trail of associations that would link documents. The physical device that Bush described resembled a desk and included screens on which material was supposed to be projected. The design of the memex was analog using the technology available at the time. As his biographer states, Bush was an expert in analog computing and code-breaking machines but never became comfortable with digital computing (Zachary, 1997). Drawings of the memex appeared in *Life* magazine, but the actual machine was never built. However, the ideas expressed in "As We May Think" inspired the new generation of information scientists and inventors.

J.C.R. Licklider was the first researcher who envisioned the use of digital technology not only to make the body of recorded knowledge available to users in a more efficient way but also to enable new forms of interactions. In 1965, when he published his book *Libraries of the Future,* digital computing was primarily conducted on large mainframe computers in research laboratories. His idea of extending the notion of libraries to computers was radical at the time. Like Bush, Licklider was concerned with the proliferation of scientific publications. He dismissed the notion of physical libraries

and books as unsatisfactory forms of information storage, organization, and retrieval. His vision moves away from books and documents to transformable information, representing ideas and facts in classes of information and domains of knowledge. Licklider (1965) states: "we need to substitute for the book a device that will make it easy to transmit information without transporting material, and that will not only present the information to people but also process it for them" (p. 6). Libraries of the future, "procognitive systems" as he calls them (p. 6), are described as a meld of a structured body of knowledge and intelligent computer processing. Licklider imagined highly interactive, "question–answer" systems capable of analyzing information on behalf of users, including reading, comprehending information, and compiling abstracts. As a computer scientist and psychologist, Licklider was interested in human–computer interaction and envisioned new systems as a way of augmenting human processing.

Libraries of the Future is also known for its quite accurate prediction of the emergence of digital libraries in 1994 (Arms, 2000). Licklider (1965) indeed points to 1994 as a possible date but cautiously notes, "we expect that computers will be capable of making quite 'intelligent' contributions by 1994, […] but we prefer not to count on it" (p. 58). Present-day digital library systems include many features predicted by Licklider, but they also fall short of his vision. Despite the digital format and more efficient storage and retrieval, digital libraries remain primarily collections of books and documents, and the level of interaction and processing is not near that which Licklider envisioned. In addition to his prescient and innovative ideas for the design of future libraries, Licklider also contributed to the development of computer networking, operating systems, and artificial intelligence. As a director of one of the agencies within the US Department of Defense's Advanced Research Projects Agency (ARPA), he was involved in developing the ARPANET, the direct predecessor to the Internet. The publication of *Libraries of the Future* in 1965 marks a transition from the visionary designs to the developments of technical infrastructure and standards that provide a foundation for constructing digital libraries (Lynch, 2005).

The period between 1965 and 1990 saw the expansion of computing beyond research labs, with a transition from mainframe systems to personal computers and the development of a globally distributed information environment. Computer networks enabled the sharing of information between interconnected sets of computers, but it was the expansion of ARPANET and the introduction of standard networking protocols that led to the emergence of the worldwide network of networks, the Internet. The advancements in networking and improvements in computer processing and storage capabilities were accompanied by research in indexing, natural language processing, and information retrieval (Calhoun, 2014; Lesk, 2005). The first commercial retrieval systems were built in the 1970s with LEXIS providing access to legal information and DIALOG serving the scientific community. As Calhoun (2014) notes "libraries were early adopters of online information systems" (p. 4) with librarians serving as intermediaries and expert searchers.

Computing was introduced in libraries in the 1960s and the library community developed their own online services and standards, such as the machine readable cataloging (MARC) format and the Z39.50 protocol. MARC was developed in the 1960s as a standardized format for exchanging cataloging records. The Library of Congress began distributing MARC records in 1969 (Arms, 2012; Calhoun, 2014; Lesk, 2012). Online library catalogs were developed in the late 1970s and became part of automated library systems. Z39.50 was one of the first protocols for distributed computing and enabled searching collections on remote systems (Arms, 2012). MARC and Z39.50 were used in some early digital projects, and although later replaced by newer standards, nonetheless, they provided a

foundation in the initial phase of digital library development. Improvements in scanning technologies in the late 1980s encouraged libraries to experiment with the digitization of selected cultural heritage materials. The first digitized collections were available on CD-ROMs or through local library networks. The invention of the web by Tim Berners-Lee in 1990 changed the landscape dramatically. The early experimental projects moved to the web, and the development of digital libraries began in earnest.

THE FORMATIVE YEARS: 1991–2001

Several research findings point to 1990 as a turning point when technological advancements in computing made it possible to move from a vision and experimental project into digital library practice (Arms, 2012; Calhoun, 2014). Arms notes, "about 1990, computing reached a level where it became economically possible to mount large collections online and to access them over networks… [The] libraries of today were formed by the energy and creativity of these efforts" (2012, pp. 579–580). The first decade was a period of intensive interdisciplinary research on concepts, architectural models, metadata standards, and digitization best practices and guidelines. Digital libraries emerged as a field of scientific inquiry with research agenda focused on digital libraries as networked information systems. Research efforts were accompanied by prototype building and developing digital library technical infrastructure.

Digital Library Initiatives

Sponsored research initiatives gave considerable impetus for digital library development during the first decade. In the United States, the federal government provided systematic funding for digital library research that was formulated in a series of planning workshops sponsored by the National Science Foundation (NSF) in 1993–94 and then established as the Digital Library Initiative (Griffin, 2005; Mischo, 2005). The Digital Library Initiative consisted of two phases:

- Phase I (1994–98) involved three US federal agencies: the NSF, the National Aeronautics and Space Administration, and the Defense Advanced Research Projects Agency. The funding was awarded to six university-based projects with the focus on information technology and testbeds. Support was not provided for the creation or conversion of digital content (Griffin, 2005).
- Phase II (1998–2002) had support from NSF, the National Aeronautics and Space Administration, and the Defense Advanced Research Projects Agency as well as additional agencies, including the National Library of Medicine, the Library of Congress, the Federal Bureau of Investigation, and the National Endowment for the Humanities. The second phase of the Digital Library Initiative was envisioned as a broader program extending beyond research on information systems into content development, use, and usability. Fifty projects were awarded funding during the second phase. As Griffin (2005) notes, "the projects addressed topics spanning the entire information lifecycle—creation, access, dissemination, use, and preservation—and placed additional emphasis on measures of impact" (pp. 22).

In addition to funding through the Digital Library Initiative, support for digital projects in the United States has also been available through the Library of Congress, the Institute of Museum and Library Services, and private foundations. In the United Kingdom, government funding was provided through the Joint Information Systems Committee for the eLib program that began in 1995 (Calhoun, 2014;

Carr, 2009). The European Commission funded digital library programs through the European Union's Framework Programme.

These funded digital library projects represent the top research on different technical and social aspects of digital libraries. Among them, the Alexandria Digital Library is a typical example. It sets out to build a digital library service for spatial and geographic data. It consists of maps, remote-sensing maps, pictures, and text materials. The testbed system entails four components: a graphical user interface, a catalog component, a component for adding new items, and a storage component. The main approaches for the project are: (1) many classes of collection items, (2) user interface digital library architecture with catalog components, (3) Internet access by a variety of users, (4) interoperability with other digital library activities, (5) an iterative design by incorporating new technologies, and (6) support for traditional library functionality (Larsgaard and Carver, 1995; Smith, 1996; Smith and Frew, 1995). Fig. 1.2 presents the legacy search page of Alexandria Digital Library. This project has evolved into the current Alexandria Digital Research Library (http://alexandria.ucsb.edu/).

One unique contribution of the Alexandria Digital Library project is its iterative design process based on user feedback. A series of studies were conducted to test user interfaces of the Alexandria Digital Library. Generating data from users' interactions with the interfaces of Alexandria Digital Library, Hill et al. (2000) identify problems with the interfaces, the requirements of system functionality, and the collection of the digital library. The following implications for the design of digital library interfaces are suggested: unified and simplified search, being able to manage sessions, more options for results display, offering user workspace, offering more help functions, allowing easy data distribution, and informing users of the process status. Analyzing a 12-month time series of transaction logs from the Alexandria Digital Library, Buttenfield and Reitsma (2002) developed a model of

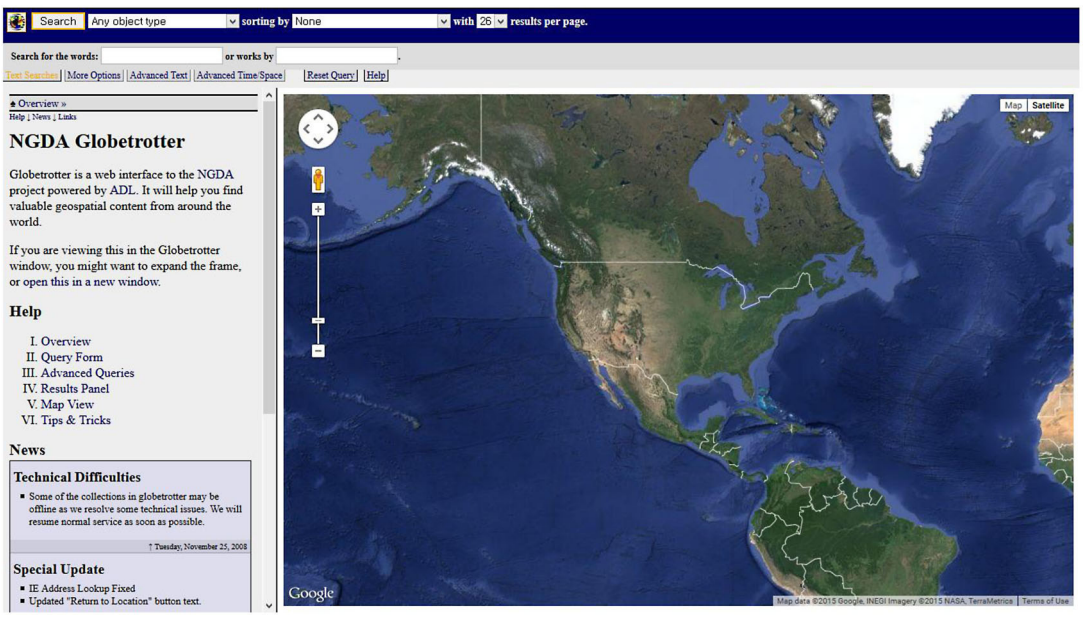

FIGURE 1.2 Alexandria Digital Library Legacy Search Page

transactions in relation to time, origin, and destination components to detect patterns of user navigation within the digital library. The findings show that user training, instead of changes to the user interface, had a significant effect on transaction patterns. Through interviews with faculty members, Borgman et al. (2005) reveal that faculty members preferred searching for locations, place names, concept, or themes and that their personal research resources were used extensively in their teaching. The findings yield implications for the creation of personal digital libraries as well as the capability to import data from different formats and standards. In addition, digital library design also needs to facilitate the sharing of resources among faculty members.

Early Digital Library Projects

The early digital library projects arose out of the traditional library environment. The Internet gave new possibilities for global information sharing. Early digital library projects gradually moved to the web environment. Examples of these transitional digital library projects include: Project Mercury sponsored by Carnegie Mellon University, the Perseus Digital Library, the Chemistry Online Retrieval Experiment (CORE), Elsevier's TULIP project, and the Envision Project. Following is a discussion of two digital library projects that represent a focus on humanities and science in this period of time: The Perseus Digital Library and the CORE project.

The Perseus Digital Library is one of the key early digital library projects. Perseus version 1.0 started as a CD-ROM of mainly Greek texts and English translations. The Perseus Digital Library went live on the web in 1995 and further expanded its collections to Greco-Roman materials. The turning point in this project was receiving funding from the Digital Library Initiative Phase 2, which enabled the Perseus Digital Library to include more digital collections in the humanities. Perseus 3.0 and the Java-based 4.0 version were released on the Internet. It moved from a teaching tool to a research tool (Crane, 2015; Preece and Zepeda, 2009). The Perseus digital library is considered as the most important resource in the study of Greece and Rome, and it is also regarded as a role model in the adoption of technology in the humanities (Dubis, 2003; Wilson, 2000). Fig. 1.3 presents the homepage of Perseus Digital Library.

Perseus is considered by researchers as a typical digital library and has been examined to generate recommendations for digital libraries in general. Several cases studies have closely analyzed the use of the Perseus project. Three years of investigation of the Perseus project in different learning environments shows that it offers information and resources for users and also requires users to apply new strategies to interact with the system (Marchionini and Crane, 1994). The Perseus Digital Library illustrates the importance of longitudinal and multifaceted evaluation of digital libraries. The subsequent recommendations for the evaluation of digital libraries are proposed: evaluation needs to consider system testing along the way; evaluation needs to adapt to changes; evaluation needs to apply both quantitative and qualitative data (Marchionini, 2000). Another case study identifies what needs to be improved in the Perseus Digital Library including the problem of implicit hyperlinks, path tools, authoring tools, etc. (Yang, 2001).

In contrast to the Perseus Digital Library, the CORE project is a digital library that mainly consists of journal articles in chemistry. The CORE project sets up a model for converting large collections of texts and graphics in a distributed network environment. The unique characteristics of this project are (1) scanning and converting a large amount of information, (2) including both text and graphical data, and (3) focusing on a specific subject area. Unlike many digital library projects, CORE consists of both a scanned image and a marked-up ASCII version for each page of the publisher's database. Each

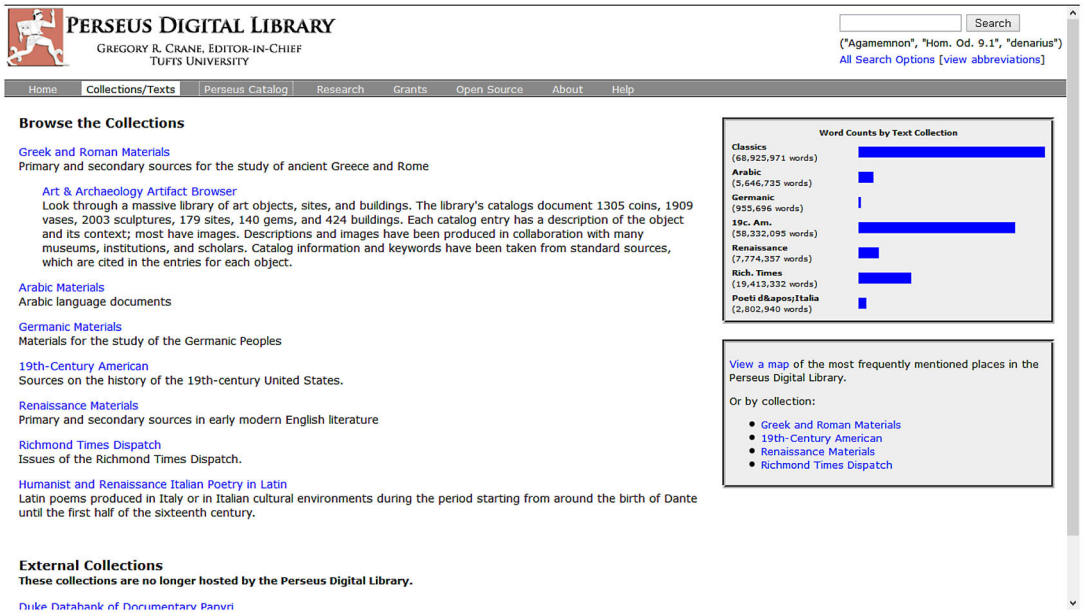

FIGURE 1.3 Perseus Digital Library Homepage

page was scanned and segmented, and graphics were separated and linked to figure references in the articles. Lack of standards for the conversion of special characters and equations was the key challenge for building CORE.

Evaluation studies were conducted to assess the CORE project. Five tasks (citation, query, browse, essay, and transform) were performed by 36 chemistry students at Cornell University, with 12 using the paper-based version, 12 using the Pixbook (a full-text search/image display system), and 12 using the Superbook (a full-text search/ASCII display system). On the query search tasks, the digital systems were better than the paper version in both performance and speed. Electronic systems performed better than the paper version on the essay task even though there was no difference in the time the task took. On the browsing task, the performance results were competitive, but the print and ASCII display were faster than the image display. For the citation task, the performance results were again the same, but the ASCII display was slower than the image display and paper version (Lesk, 1991). Based on the data collected from transaction logs, online questionnaires, online comments, interviews, and anecdotes, Entlich et al. (1996) report that the top 35% of users accounted for 80% of the usage. The analysis of users' searching, viewing, reading, and printing habits reveals that users appreciated the full-text searching capability. Among different types of searches, the author search was the most popular search, representing 32.1% of all searches. Users were still in the process of getting familiar with the digital library, and syntax and format errors accounted for 17.6% of searches. Usage data provide not only information about how users interact with digital libraries but also what affect users' interaction with digital libraries.

This period saw the first decade of digital library projects for the scientific and cultural heritage community being developed. Exemplary projects of the first decade of digital libraries in the scientific

and cultural heritage community are: arXiv (pronounced AR-KIVE), a pioneering repository project; American Memory, a Library of Congress project of digitized cultural heritage collections; and JSTOR, a digital library of primarily academic journals

Launched in 1991, arXiv was created by Paul Ginsparg as an electronic bulletin board for colleagues to exchange unpublished manuscripts in the field of theoretical high-energy physics. The arXiv is also known as the "Los Alamos e-print archive." Twenty years later, the system contains about 7,000,000 full texts. It receives 75,000 new texts each year and serves close to 1 million full-text downloads to 400,000 unique users every week. ArXiv mainly covers research on physics but has extended to other related fields: mathematics, nonlinear science, computer science, statistics, etc. (Ginsparg, 2011; Luce, 2001). Further research has been conducted to investigate the effect of position in announcements of newly received articles on citation and readership in arXiv. The findings discover multiple effects in play in the digital library, ranging from the "visibility" effect for positions near the beginnings of the announcements, to the "self-promotion effect," to the "reverse-visibility" effect for positions near the ends of announcements, and a "procrastination" effect associated with submission made at the end of the day (Haque and Ginsparg, 2009, 2010). The main contribution of arXiv is the implementation of the idea of a central repository for researchers to share articles all over the world. In addition, it prompted the creation of PubMed Central by the US National Institutes of Health. Moreover, it blazed a trail for a new direction for scholarly communication. Fig. 1.4 presents the homepage of arXiv.

American Memory plays a key role in the development of digital libraries as one of the earliest large-scale projects. The Library of Congress started the American Memory as a pilot project in 1989, which became the National Digital Library Program in 1995. The American Memory project focuses on American history and culture. It was funded by congressional appropriations and private sources. The unique strength

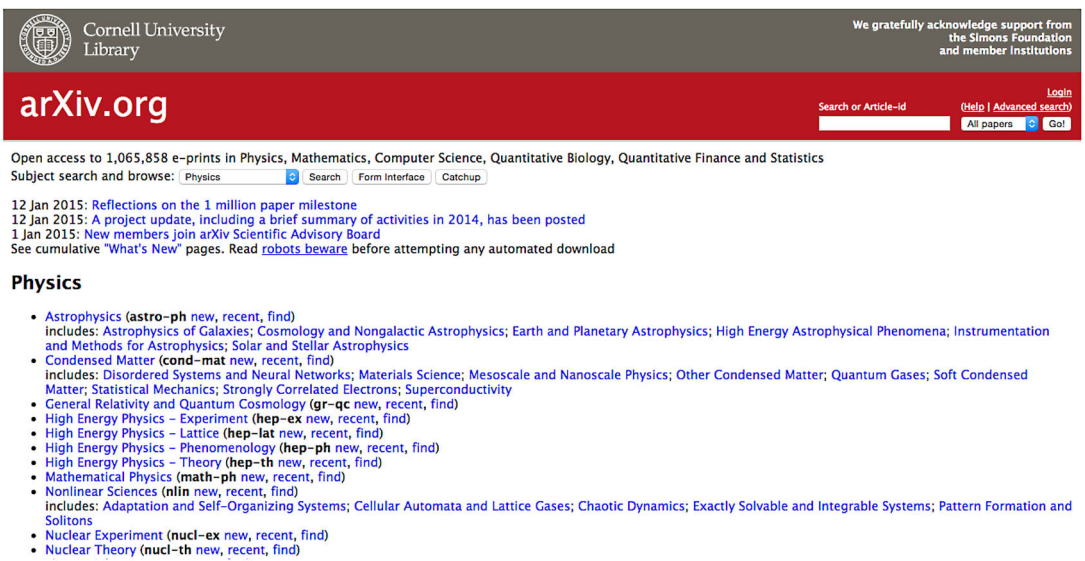

FIGURE 1.4 arXiv Homepage

of American Memory is its digital reproductions of original documents, which include texts, photographs, maps, motion pictures, and sound recordings. The selection criteria for the collections take into account the following: the uniqueness of the collection, usefulness in fulfilling other tasks including preservation, suitability for digitalization technology, and the educational value. Another strength is that it offers a standardized interface for users to navigate multiple collections. In addition to searching and browsing functions, there is an education program "Learning Page" associated with the American Memory collections. This page offers information for both teachers and students to teach and learn about not only American Memory collections but also American history (Fesenko, 2007; Lucas, 2000). Fig. 1.5 presents the homepage of American Memory.

As one of the early digital library projects, the development of American Memory provides invaluable lessons for the creation of digital libraries. Madden (2007, 2008) summarizes the contributions of the American Memory project:

- Data, in particular usage data, can help the enhancement of the digital library.
- Automation can reduce human error and stimulate standards development.
- Exceptions reduce the sustainability of data and the scalability of production.
- Interoperability is required for collaborative digitization projects.
- Flexible data model is needed to satisfy diverse needs.
- Data can be reused to accommodate lifecycle of different digital productions.

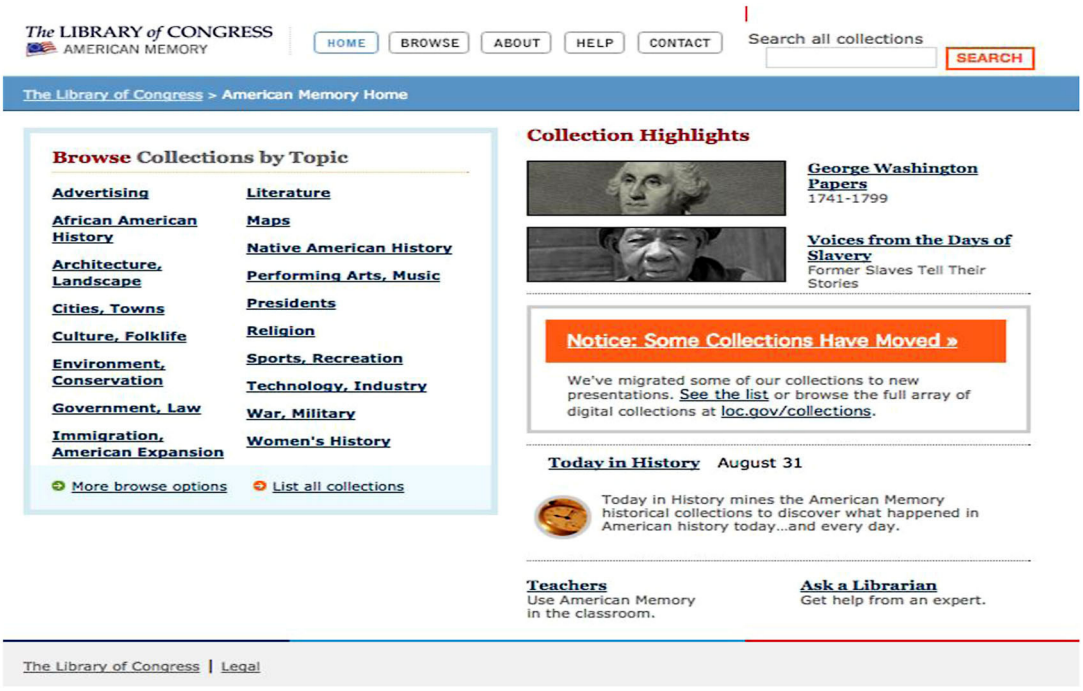

FIGURE 1.5 American Memory Homepage

JSTOR is primarily a digital library of academic journals, providing access to back issues of journal publications in social sciences, arts and humanities, and sciences. JSTOR started in 1995 as a digitization project funded by a grant from the Andrew W. Mellon Foundation. The first JSTOR collection, Arts & Sciences, became available in 1997. JSTOR contains complete back files of over 1600 core scholarly journals. Digital versions of print journals were created by scanning each page at a high resolution and searchable text was created by using optical character recognition (OCR) software plus intensive proofreading. More recently, JSTOR began adding books and primary sources. Preservation and access are the two main objectives of the JSTOR project. On the one hand, JSTOR was created to preserve digital scholarly content; on the other hand, it offers more efficient and effective access for scholarly communication. Another goal is to save a substantial amount of investment in physical library space. Fig. 1.6 presents the homepage of JSTOR.

Spinella (2007, 2008) highlights four aspects of JSTOR's success.

- First, the number of community participants gradually increased in all categories including libraries, publishers, higher education institutions, linking partners, and other types of collaborating organizations. There were more than 4300 institutions supporting JSTOR in 2007.
- Second, there is a large amount of archived journal publications. In 2007, JSTOR contained 750 journals with 25 million pages dating back to 1665. The collection spans 50 disciplines covering the humanities, social sciences, economics, education, law, and the life sciences. Moreover, JSTOR implemented standards for digitization, indexing, and metadata.

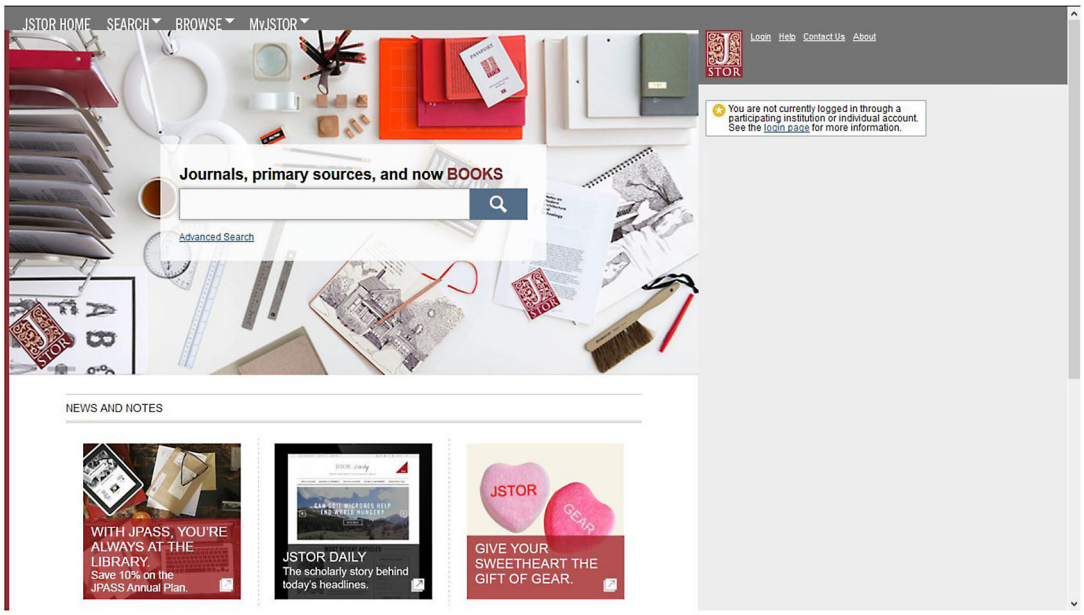

FIGURE 1.6 JSTOR Homepage

- Third, usage has increased at an average of 50% every year since 2001 and reached 500 million accesses in 2007. Large institutions are the main users of JSTOR. Disciplinary usage is affected by the number of researchers/teachers and students, as well as the number of the titles in the system. Users viewed 2–2.5 pages per article from 1997 to 2007.
- Fourth, the average cost to maintain the content and access per institution may decrease along the way. The costs for larger institutions and medium/small institutions are different. While large institutions experience declining cost per use, medium/small institutions' costs increase.

JSTOR played a significant role in facilitating the transition from print to electronic publishing in serial publications. However, creating the JSTOR repository was not an easy task. Walker et al. (2010) demonstrate the need for more resources, cohesive leadership, and open communication among all the parties in building a JSTOR repository.

It is worth noting that the development of digital libraries in this period also led to the creation and advancement of digital library architecture. In order to solve the problems of linked heterogeneous digital library services, Kahn and Wilensky (1995) created an important digital library architectural framework that defines the key entities in a digital library. It consists of four components: a digital object, a repository, a repository access protocol, and dissemination. A digital object is an instance of an abstract data type consisting of *data* and *key-metadata* in which "a handle," an identifier unique to the digital object, is a critical part. A *repository* is a storage system in which digital objects are stored for access and retrieval. The Repository Access Protocol (RAP) allows access to a stored digital object or its metadata by specifying its handle. A *dissemination* is the result of an access service request. The main contribution of this framework is the naming, identifying, and/or invoking of digital objects. Arms (1995) further enhances the key concepts in digital library architecture and identifies principles to highlight the key issues in the transition from the network services to a digital library:

- The technical framework is part of a larger legal and social context.
- The limitations of terminology hinder the understanding of digital library concepts.
- The structure and content of a digital library should be separate from each other.
- Names and identifiers are critical for a digital library.
- Digital library objects are beyond merely just data.
- Users may use a digital library object differently from the stored object.
- Repositories should be the custodian for the stored object.
- Users need useful intellectual works instead of just digital objects.

Digital library research and projects in their formative years have contributed to the development of digital library architecture, standards, iterative design, and usability studies, as well as best practices for their development. Digital library projects have contributed to the effective retrieval of multimedia objects across multiple repositories on the Internet and semantic interoperability (Schatz and Chen, 1996). After comprehensively reviewing digital library literature in terms of its content, organization, services, access, evaluation, as well as social, economic, and legal issues, Fox and Urs (2002) identify two types of challenges for digital libraries: research and practice. From the research perspective, a theory is needed that specifically accounts for different aspects of research in digital libraries; at the same time, a methodology for the design, development, and enhancement of digital libraries is also required. From the practice perspective, there is a lack of guidelines for the management of digital libraries.

BUILDING THE CONTENT AND OPENING ACCESS: 2000s

In the 2000s, digital library development rose to a new level. "By the end of the 1990s, digital libraries were no longer a novelty" (Arms, 2012, p. 583). Meanwhile, government funding ended. "As of 2005, it seems a virtual certainty that substantial programmatic US government funding of digital libraries research in terms of the construction of prototype systems is at an end" (Lynch, 2005, para. 9). In this period of time, two characteristics exemplify the development of digital libraries: Mass Digitization Projects and Open Access Repositories.

Mass Digitization Projects

Mass digitization refers to converting materials on an industrial scale without curating specific materials for digitization. OCR is used to make the full text of digitized documents searchable (Coyle, 2006). Coyle (2006) recognizes several issues in mass digitization:

- Include selection of items for digitizing, scanning, quality control, the OCR process, the creation of metadata, the creation of technical metadata, and storage
- Create book structure in a digital format
- Develop user interface for the large collection with different formats
- Develop new digitization standards for mass digitization
- Control quality of mass digitization for preservation purpose
- Define the scope of the mass digitization project

The Google Book Project represents a major mass digitization initiative undertaken by a commercial company in cooperation with the cultural heritage sector. Google initiated a massive project in 2004 to digitize millions of books and make their text searchable. The Google project is an expansion of the Early Google Print. At the beginning, five of the world's largest libraries joined the project: Harvard University, Stanford University, the University of Michigan at Ann Arbor, the University of Oxford, and the New York Public Library. Taking into consideration copyright laws, the project started digitizing works in the public domain first. To deal with copyrighted books, search engines only offer a few lines of text associated with a search term (Carlson and Young, 2005; Hanratty, 2005). The Google Book Project promotes the discussion of multiple issues and concerns. Bearman (2006) cites the five principles enumerated by the national librarians of the La Francophonie Meeting held in 2006: free access to publicly owned resources, nonexclusive agreements with content providers, preservation of standard images for long-term accessibility, protection of the integrity of the original materials, and provision of multilingual and multicultural access. However, the Google Book Project is also criticized for its image and metadata quality (Leetaru, 2008).

The Google Book Project is not the only mass digitization project although the other ones are not comparable in scale. The Internet Archive, a nonprofit group, worked with major libraries in six countries—Canada, China, Egypt, India, The Netherlands, and the United States—to incorporate their digitized books into the group's collection. They had about 77,000 books online in early 2005. These mass digitization projects are interrelated. For example, Brewster Kahle, the creator of the Open Content Alliance (OCA), is also the driving force behind the Internet Archive. Multilingual digitized text and multimedia documents are the foci of OCA. Over thirty library and cultural organizations shared their digital collections, and the content was available through the OCA web site and Yahoo (Carlson and Young, 2005; St. Clair, 2008).

Funded by the NSF, the Million Book Project is a collaborative project with international partners in China, India, and the Biblioteca Alexandrina. The main objective of the project is to advance research and development in the following areas: machine translation, massive distributed databases, storage formats, the use of digital libraries, distribution and sustainability, security, search engines, image processing, OCR, language processing, and copyright laws (St. Clair, 2008).

These mass digitization projects share similarities and differences. For example, Google Books and the Open Content Alliance represent two different approaches for mass digitized projects: one represents a proprietary model, and the other one represents an open model. After comparing the two projects, Leetaru (2008) concludes that they share many similarities even though they seem different. Google creates more powerful and intuitive interfaces for users to access the digital materials. The OCA effort ensures long-term preservation if the funding source is stable. Crane (2006) highlights the unique dimensions of a mass digitized project:

- Large scale in quantities
- Heterogeneity of content
- Granularity of objects from single catalogue entry to tagged objects
- More noise
- More audience as the result of open access
- Reduction in collections and entry points

These projects contribute to the enhancement of technology as the result of automatic processes, support for machine translation, and the improvement of information extraction. Copyright is the critical barrier for mass digitized projects, although copyright laws intend to protect new innovations. St. Clair (2008) proposes several approaches to overcome the barriers:

- Identify nonrenewed books
- Create a list of US books published between 1923 and 1963
- Obtain permission from publishers
- Digitize government publications
- Create synthetic documents, in particular in science and technology fields

The favorable rulings for HathiTrust and Google Books in *Authors Guild, Inc. v. HathiTrust* and *Authors Guild v. Google Books* have significant impact on the growth of digital libraries. Chapter 2 offers detailed discussion of the two cases, in particular how the four factors of fair use were weighted.

Open Access Repositories

An open access repository is defined as a collection of full-text documents available in online databases on the Internet that can be accessed freely and instantly. Institutional repositories are managed by research institutions to house their own authors' works (Pinfield, 2005). The creation of open access institutional repositories significantly promotes scholarly communication. Open access repositories achieved remarkable successes in this period of time. The origin of the open access repositories is associated with cost of serials. Now, a measure of high quality in repositories is whether it can also help to make research visible (Joint, 2008). Suber (2009) created and updated a timeline of the open access movement, including the development of open access repositories, such as 1991's arXiv, 1994's Perseus project, etc. Many open access repositories have been established. The Open Directory of Open Access Repositories offers a quality listing of international open access repositories. According to its

About page (OpenDOAR, 2014, para. 7), "*Open*DOAR is primarily a service to enhance and support the academic and research activities of the global community. *Open*DOAR maintains a comprehensive and authoritative list of institutional and subject-based repositories. It also encompasses archives supported by funding agencies like the National Institutes for Health in the USA or the Wellcome Trust in the UK and Europe." It had more than 2600 listings in August 2015.

On the one hand, the establishment of institutional repositories and the development of digital libraries play a critical role in the success of open access (Bailey, 2007). On the other hand, open access publishing increases the complexity and diversity of the development of digital libraries. Open access repositories can be classified as subject repository or institutional. Creaser et al. (2010) conducted an international survey and focus groups regarding awareness and attitudes toward open access repositories. The results show that in general authors support open access repositories. However, there are disciplinary differences in their preference on repositories. While authors from physical science and mathematics prefer subject-based repositories, authors from the social sciences and the humanities and arts like institutional repositories. ArXiv is a good example of a subject-based repository.

Tsakonas and Papatheodorou (2007) performed a usefulness and usability study to evaluate an open access digital library, a subject-based repository: e-prints in Library and Information Science. The findings of the study indicate that user satisfaction is influenced by usefulness and usability. Usefulness is measured by how relevant the information is, as well as the breadth of coverage. Ease of use, aesthetics, terminology, and learnability determine whether the digital library is usable. An increase in usability of a digital library is mainly owed to its open access nature and personalized features. Just as Krishnamurthy (2008) points out, the key term for open access, open access repository, and digital libraries is open.

LARGE-SCALE DIGITAL LIBRARIES: 2010–

Connecting discrete, standalone digital libraries has been a major issue since the beginning of their development. "The plurality of digital libraries is an ongoing technical challenge and major source of user frustration" (Bearman, 2007, p. 224). Major efforts have been undertaken to aggregate content from smaller individual digital libraries and to provide portals for global searching and retrieval. The large-scale digital libraries represent the next generation of discovery systems, with centralized access to a wide variety of scientific and cultural heritage materials. Characteristics of large-scale digital libraries can be summarized as follows:

- Large collection size: most of them contain more than one million items.
- Diverse formats: collection items may contain text, images, audio, video, etc.
- General and specific collection development policy: it not only contains general collection scope, quality guidelines, and selection responsibility but also depends on each participating member's policy for its specific collections.
- Copyright concern: some of them may not have copyright clearance.
- Level of access: single interface for all collection items; depending on copyright, some items have full-text access while others might have access limited to just the citation or abstract.
- Interoperability: metadata mapping is used to ensure the exchange of metadata between collections, or a single metadata schema is applied across all collections.

This section highlights the developments of four large-scale digital libraries, representing the work in building digital libraries with multiple subjects and with different ways of collaboration:

the National Science Digital Library (NSDL), HathiTrust, the Digital Public Library of America (DPLA), and Europeana.

The National Science Digital Library

The NSDL is one of the largest digital libraries constructed to advance science and math education in the so-called STEM disciplines (Science, Technology, Engineering, and Mathematics). It was funded by the US NSF initially. Most of the 64 projects focus on the development of collections in a variety of subject areas including STEM. The objective of the NSDL is to provide students and teachers a variety of interactive learning and teaching resources that are comprehensive, high quality, and authoritative. The initial effort focuses on collection building, technical infrastructure, and metadata creation (Lagoze et al., 2002; Lagoze et al., 2006; McCown et al., 2005; Zia et al., 2001). Fig. 1.7 presents the homepage of the NSDL.

The contribution of the NSDL is not just as a repository for digital collections. Lagoze et al. (2002) illustrate the core components of the NSDL architecture, which is based on previous digital library research:

- Metadata repository contains metadata.
- Search and discovery offer capabilities for users to locate and find resources in the digital library.
- Access management sets the requirements for users of the digital library and providers of intellectual property.
- User interface and portals facilitate interactions among users, collections, and services of the digital library.

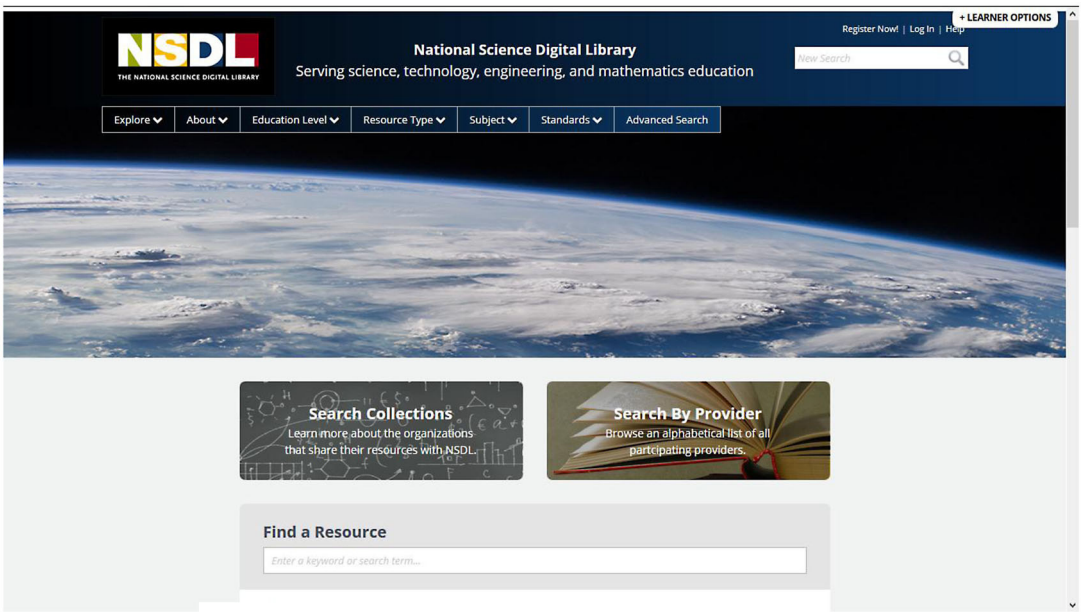

FIGURE 1.7 The National Science Digital Library Homepage

The NSDL added new standards on metadata and universal object identifiers that assist users to identify high quality resources. In addition, interoperability and reusability standards enable resources from different disciplines and different times to integrate together (Fortenberry and Powlik, 1999). Arms et al. (2003) discuss the NSDL's Metadata Repository based on the Open Archives Initiatives Protocol. It selected Dublin Core, qualified Dublin Core, IEEE Learning Technology Standards committee, Advanced Distributed Learning, MARC 21, Content Standards for Digital Geospatial Metadata, Global Information Locator Service, and Encoded Archival Description as metadata schemas. Its metadata was created at the collection level instead of item level. The Open Archives Initiative Protocol for Metadata Harvesting was considered to be used for harvesting. Lagoze et al. (2006) introduce the NSDL Data Repository, which offers a platform for new NSDL applications that promote user participation in the digital library. Its Data Repository consists of four applications: Expert Voices, a collaborative blogging system that supports Q/A discussions, recommendations and annotations, structured metadata, and the creation of relationships among resources in the NSDL; On Ramp, a distributed system that allows multiple users/groups to create, edit, and disseminate information in different formats; Instructional Architect, a system that enables users to discover, select, and design instruction materials; and the Content Alignment Tool, a system that aligns NSDL collections to state and national educational standards.

McCown et al. (2005) performed an experiment using educators to evaluate the relevance of the search results in the NSDL and Google. The results show that Google outperformed in the following areas: (1) useful educational materials, (2) retrieval precision, with Google's 38.2% compared to NSDL's 17.1%. At the same time, NSDL was found to contain higher quality resources and unique collections. One in four NSDL resources were not indexed by Google, and little overlap was identified between the two collections. NSDL has become a popular teaching and learning resources (Chen et al., 2015). A large-scale research project has been conducted nationally by a group researchers on instructors' and students' use of digital learning materials, in particular NSDL (Morgan et al., 2013, 2014; Wolf et al., 2008). The findings show that 80% of the faculty participants ranked peer review of materials as their first or second priority in relation to important digital library features. In addition, 30% of them indicated sources of pedagogical supplements as first or second priority. Students had diverse learning needs, styles, and different preferences in terms of technology and course mode. The information about students helped enhance the design of NSDL.

HathiTrust

HathiTrust represents a successful example of collaborative work on a large-scale repository/digital library. The Hathi name represents the value of the organization; Hathi in Hindi means elephant, which is well known for its memory, wisdom, and strength (Christenson, 2011). The HathiTrust started in 2006 when the University of Michigan proposed to the libraries associated with the Committee on Institutional Cooperation to build a shared digital repository to store the large files that Google digitized from the Committee on Institutional Cooperation libraries' book collections. HathiTrust, named in 2008, includes both digitized books and journal articles. This digital library contains materials in both the public domain and copyrighted works. The main issues are quality control, public search interfaces, ingestion of non-Google and nonbook content, access issues for people with disabilities, collection grouping, data mining, and academic research tools. HathiTrust is not dependent on the Google Book Project, and it has more resources from the public domain. The repository enables academic libraries to preserve collections from the last 150 years and may lead to a new direction for the future. Moreover,

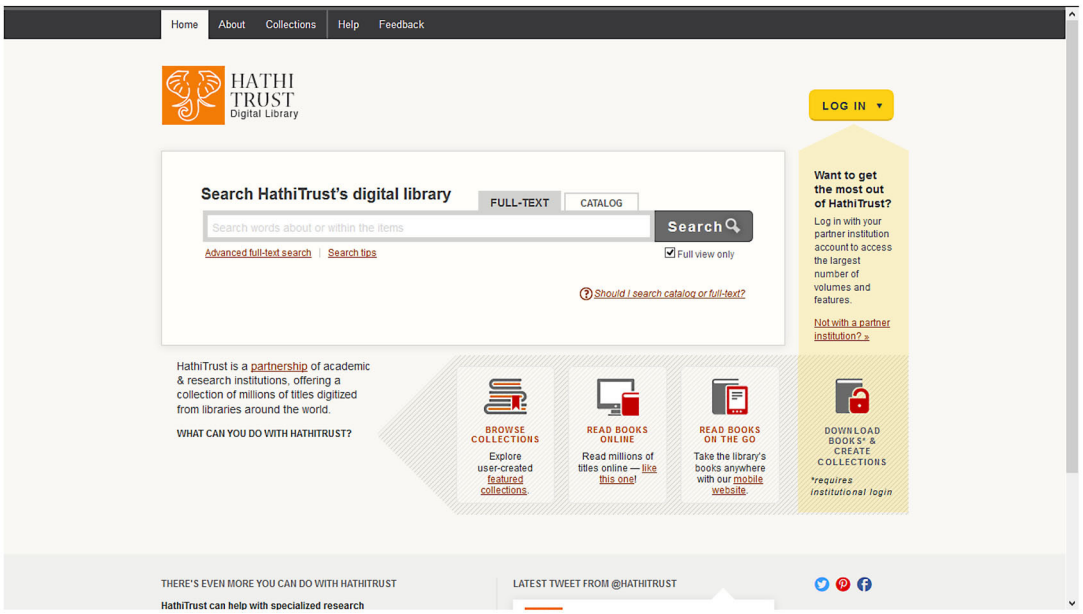

FIGURE 1.8 HathiTrust Homepage

OCLC (a global library cooperative) records the digital titles in HathiTrust in addition to printed copies in academic libraries (Pritchard, 2012). As of August 2015, the HathiTrust had more than 100 partners, and it is open to institutions all over the world. It contained more than 6 million book titles and 350,000 serial titles (HathiTrust, n.d.). Fig. 1.8 shows the homepage of HathiTrust.

The main objective of HathiTrust is to create a comprehensive digital collection of library materials owned by the participating research institutions. Simultaneously, it improves broad access to these materials, preserve important research records, coordinate shared collection management strategies to save costs, create and sustain the "public good," and develop a technical framework that enables members to build and share functionality (HathiTrust, n.d.). It is not only a digital library but also a collaborative group that works on key issues in creating and preserving a large collection of digital volumes. The wide collaboration, aggregated expertise, and integrated digital collections benefit both the participating libraries and users (Christenson, 2011). HathiTrust's metadata management system "Zephir" is a model of a metadata management system that offers the "best practical solution to organize and automate the multiple processes of metadata conversion, quality control and ingest, including making inventories and error reports" (Mallery, 2015, p. 354). Moreover, a web site was created (https://www.hathitrust.org/zephir) to provide comprehensive documentation to illustrate this multifaceted system.

The main challenge facing HathiTrust is copyright. Researchers can search for copyrighted documents but are unable to access them if their institutions are not members. Although multimedia information is included in the digital library, there is a lack of audio functionality which makes it difficult to become a digital library for special user groups, such as musicians (Downie et al., 2014). While 68% of HathiTrust's collection items are "in copyright," the other 32% are in the public domain.

The compositions of the 32% in the public domain consist of 21% in public domain worldwide including about 4% US federal government documents and 11% in US public domain (Eichenlaub, 2013).

User feedback is a key for the creation of a successful digital library. Fenlon et al. (2014) investigated user requirements for collection building in the HathiTrust Digital Library. The analysis from focus groups and interviews indicates that scholars consider collection building as a key scholarly activity and highly heterogeneous. They expect better metadata offering rich data about the documents and are willing to participate in the metadata creation and sharing process. After comparing the functionality of Google Books and HathiTrust on federal government publication use, Sare (2012) concludes that Google Books and HathiTrust each has its own strength and limitation. While Google Books has more government documents in general, HathiTrust is best for locating full-text government documents published after 1923. Google Books has an advantage in providing the added functionality of data visualization.

The Digital Public Library of America

The DPLA, a national public digital library, launched in 2013. The DPLA intends to not only serve as a center for the public to access all cultural materials free of charge but also as a platform for libraries and other cultural heritage institutions to thrive in the digital age. Its digital metadata come from large institutions, such as the National Archives and Records Administration and HathiTrust. In addition, regional service hubs contribute to metadata harvested by the DPLA. The DPLA is an aggregator of existing digital materials available in a distributed environment rather than a central repository (Cottrell, 2013; Gregory and Williams, 2014). As of 2015, the DPLA offered more than 10 million items from 1600 institutions to the public (DPLA, 2015). Fig. 1.9 presents the homepage of the DPLA.

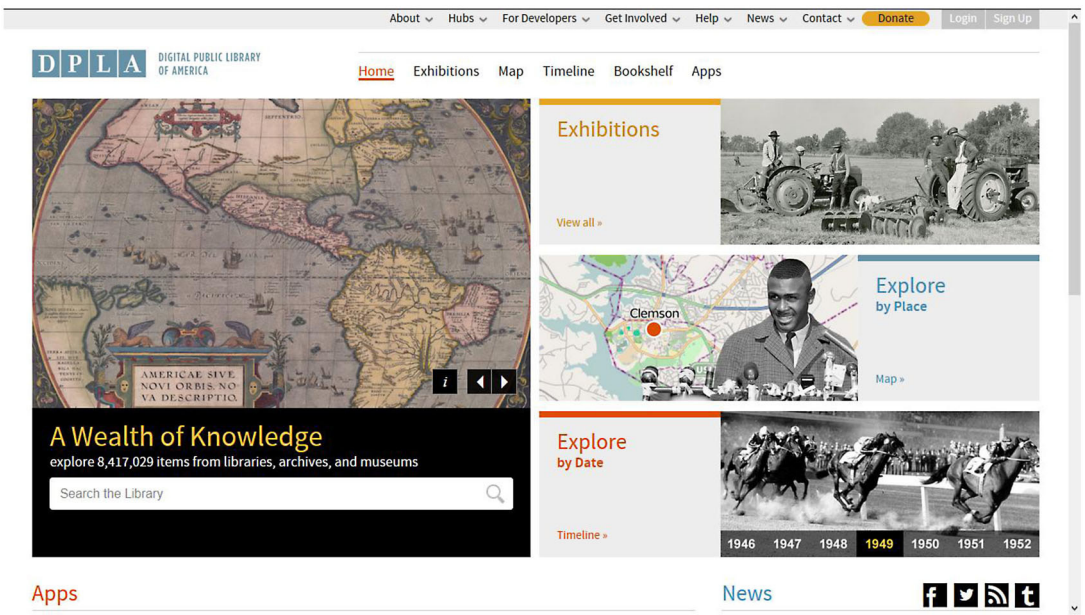

FIGURE 1.9 The Digital Public Library of America Homepage

The DPLA consists of three layers: a portal to a public web site with search functions; a platform containing the technical infrastructure, which is built using open source software; and a partnership with libraries, museums, archives, funders, universities, schools, and other institutions to promote broad access to information (Vandegrift, 2013). In addition, it is a repository for metadata. Content remains with the institutions, and only the metadata about the digital objects are harvested. Another benefit of the DPLA is that it provides APIs and open data for software developers, researchers, and others to create different types of tools for discovery, access, and communication (Eichenlaub, 2013; Ma, 2014).

The unique characteristic of the DPLA is acting as service and content hubs. According to Paulmeno (2015), DPLA partnered with different organizations and created 13 services hubs as of April 2015. These hubs collect metadata from different organizations and send them to the DPLA. At the same time, 16 content hubs had been established as of 2015. Content hubs are large institutions such as Harvard University, The New York Public Library, and The Smithsonian. Instead of collecting metadata, they each supply more than 200,000 records with their own metadata. Expanding their service and content hubs is the primary goal of the DPLA. Simultaneously, the goal of promoting DPLA usage is also vital. DPLA sponsors its annual DPLAfest in different parts of the country to encourage the involvement and use of the DPLA. With more funding, the DPLA is in the process of creating more service hubs to eventually connect online collections and serve all 50 states of United States in 2017 (DPLA, 2015). The DPLA infrastructure was also designed to be interoperable with the Europeana, which makes it possible for the creation of a worldwide network (Eichenlaub, 2013).

Just as Vandegrift (2013, p. 3) puts it, "DPLA is not a public library, a content repository, or a threat to traditional library services." The DPLA becomes the partner of, rather than a replacement for, public and academic libraries. It will greatly enhance their collections and services. The ultimate goal of DPLA is to serve as a center for the general public to explore libraries, archives, and museums' scattered collections in the United States. The DPLA also explores alternative licensing models with authors and publishers (Palfrey, 2013).

Europeana

Europeana, Europe's digital library, museum, and archive, is an across-domain and cross-cultural heritage site. Europeana means "thinks European" in Latin, and it is self-explanatory for users to understand the coverage of the digital library. Europeana was launched in 2008 and funded by the European Commission. Europeana covers Europe's cultural and scientific heritage materials and consists of pictures, files, books, photos, sounds, newspapers, manuscripts, and archival records digitized in museums, libraries, archives, and audio–visual collections (Purday, 2009). In 2014, the digital library contained 10 million digital objects from every member of the European Commission, including approximately 130 institutions (Weiss, 2014). The main objective of Europeana is to develop an open platform so that users and cultural institutions can access and manage a large collection of surrogate objects representing digital objects through an application program interface (API) (Concordia et al., 2010). Fig. 1.10 shows the homepage of Europeana.

As a large-scale and aggregate digital library, the key issues for building the digital library are split into work packages that focus on human, political, and intercommunity interoperability; interoperability of metadata standards; semantic and multilingual interoperability; technical architecture; and users and usability. The European Semantic Elements (ESE) consists of a subset of the Dublin Core and a set of 12 elements to meet its own needs (Purday, 2009). The Europeana Data Model (EDM) is

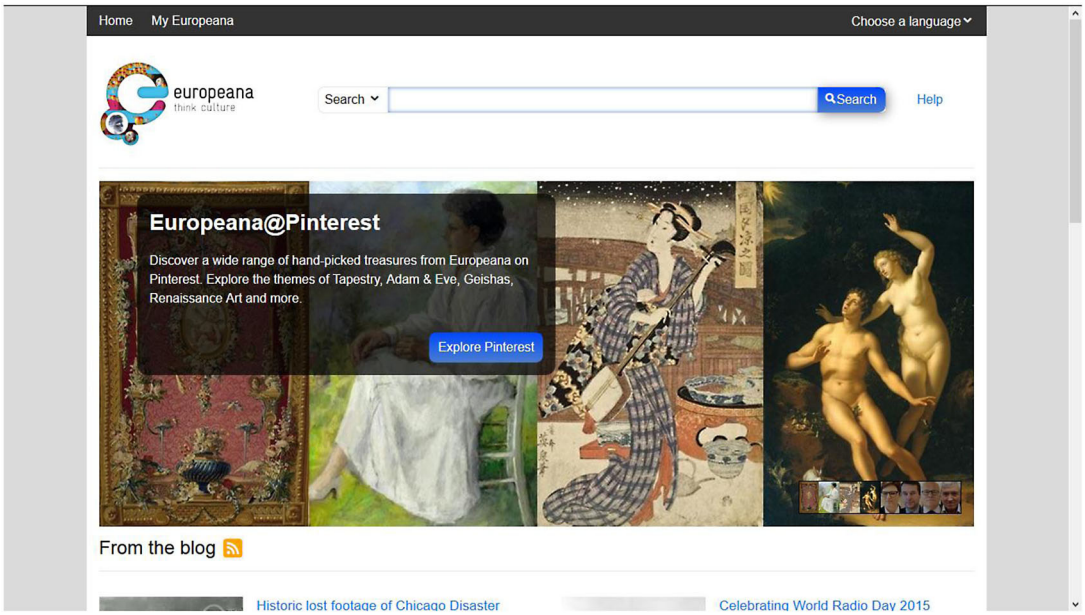

FIGURE 1.10 Europeana Homepage

replacing the European Semantic Elements to better represent millions of objects semantically across different types of Europeana's cultural heritage communities. The EDM integrates existing standards like RDF(s), Open Archive Object Reuse & Exchange, SKOS, and Dublin Core. It offers the possibility for cultural heritage institutions to move from closed information architectures to open and linked environments (Doerr et al., 2010; Haslhofer and Isaac, 2011). Peroni et al. (2013) identify some issues of EDM and suggest that content providers need to comply with the correct EDM Mapping Guideline and improve different levels of description of the objects.

Users' opinions have been solicited to enhance the design of Europeana, as has been the case with some previously discussed digital libraries. Dobreva and Chowdhury (2010) conducted a usability study of Europeana. They found that the majority of users rated the following functions highly: navigation, search functions, search results presentation, and ease of access. Users would like to have results ranked better and have higher precision results, to be able to refine their searches, to be able to overcome the language barriers, to have more help functions, to be able to use multiple ways of browsing including visualization tools, to be able to customize the interface, to have more links between items, and to be able to go back to their original searches. Nicholas and Clark (2014) analyzed log data to generate information-seeking behaviors of Europeana users. Log data analysis shows: (1) regarding site loyalty, less loyal users were found in the Europeana site than in other scholarly sites; and cultural institutions and their members were loyal users. (2) Regarding social media referrals, the overall visitors from social media referrals growth was lower (34%) than the overall visitor growth (90%) of Europeana. (3) Virtual exhibitions attracted high levels of engagement. More interestingly, from the beginning, Europeana took into consideration the design of a "lite" interface for mobile users. After analyzing log

files of user interactions, in particular mobile interactions with Europeana, Nicholas et al. (2013) stress the following unique findings:

- It is extremely popular for users to use personal devices to interact with Europeana. Use of personal devices is growing four times faster than office devices and will become a critical part of future traffic.
- Unique information-seeking and use behaviors of mobile users focus on consuming content instead of creating content. They are short one-time visitors with less intensive engagement, and more visits occur during evenings and weekends.

These findings of user studies have great implications for the enhancement of Europeana.

The development of large-scale digital libraries is a necessary step in the evolution of digital libraries. On the one hand, it brings many benefits to developers to share architectures, metadata, other technical standards and guidelines, and costs. It also offers the opportunity for users to access widely available resources organized in a unified format and interface. On the other hand, it brings challenges and problems for the development of digital libraries. Copyrights, metadata harvesting, interoperability, digitizing quality, and multilanguage access are several typical challenges that researchers and practitioners have to face in the creation of large-scale digital libraries. Please read Chapter 11 for a detailed discussion of the challenges of digital library development.

REFERENCES

Agre, P.E., 2003. Information and institutional change. The case of digital libraries. In: Bishop, A.P., Van House, N.A., Buttenfield, B.P. (Eds.), Digital Library Use: Social Practice in Design and Evaluation. MIT Press, Cambridge, MA, pp. 219–240.

Arms, W.Y., 1995. Key concepts in the architecture of the digital library. D-lib Magazine 1 (1). Available from: http://www.dlib.org/dlib/July95/07arms.html.

Arms, W.Y., 2000. Digital Libraries. MIT Press, Cambridge, MA.

Arms, W.Y., 2012. The 1990s: the formative years of digital libraries. Lib. Hi Tech 30 (4), 579–591.

Arms, W.Y., Dushay, N., Fulker, D., Lagoze, C., 2003. A case study in metadata harvesting: the NSDL. Lib. Hi Tech 21 (2), 228–237.

Association of Research Libraries (ARL), 1995. Definition and purposes of a digital library. Available from: http://old.arl.org/resources/pubs/mmproceedings/126mmappen2

Bailey, C.W., 2007. Open access and libraries. Collect. Manage. 32 (3–4), 351–383.

Bawden, D., Rowlands, I., 1999. Digital libraries: assumptions and concepts. Libri 49 (4), 181–191.

Bearman, D., 2006. Jean-Noël Jeanneney's critique of Google: private sector book digitization and digital library policy. D-Lib Mag. 12 (12), 1–7.

Bearman, D., 2007. Digital libraries. Ann. Rev. Inform. Sci. Technol. 41 (1), 223–272.

Bishop, A.P., Van House, N.A., Buttenfield, B.P., 2003. Digital Library Use: Social Practice in Design and Evaluation. MIT Press, Cambridge, MA.

Blandford, A., Keith, S., Connell, I., Edwards, H., 2004. Analytical usability evaluation for digital libraries: a case study. In: Digital Libraries, Proceedings of the 2004 Joint ACM/IEEE Conference. IEEE. pp. 27–36.

Borgman, C.L., 1999. What are digital libraries? Competing visions. Inform. Process. Manage. 35 (3), 227–243.

Borgman, C.L., 2000. From Gutenberg to the Global Information Infrastructure: Access to Information in the Networked World. MIT, Cambridge, MA.

Borgman, C.L., 2003. Designing digital libraries for usability. In: Bishop, A.P., Van House, N.A., Buttenfield, B.P. (Eds.), Digital Library Use: Social Practice in Design and Evaluation. MIT Press, Cambridge, MA, pp. 85–118.

Borgman, C.L., Smart, L.J., Millwood, K.A., Finley, J.R., Champeny, L., Gilliland-Swetland, A.J., Leazer, G.L., 2005. Comparing faculty information seeking in teaching and research: implications for the design of digital libraries. J. Assoc. Inf. Sci. Technol. 56 (6), 636–657.

Bush, V., 1945a. As we may think. Atlantic Mon. 176 (1), 101–108, Available from: http://www.theatlantic.com/magazine/archive/1945/07/as-we-may-think/303881/.

Bush, V., 1945b. As we may think. Life, 19(11), 112-114, 116, 121, 123-24. [Reprinted from Bush, V. (1945). As we may think. Atlantic, 176(1), 101–108].

Buttenfield, B., 1999. Usability evaluation of digital libraries. Sci. Technol. Lib. 17 (3–4), 39–59.

Buttenfield, B.P., Reitsma, R.F., 2002. Loglinear and multidimensional scaling models of digital library navigation. Int. J. Hum-Comput. Stud. 57 (2), 101–119.

Calhoun, K., 2014. Exploring Digital Libraries: Foundations, Practice, Prospects. Neal-Schuman, Chicago.

Candela, L., Castelli, D., Pagano, P., 2007a. A reference architecture for digital library systems: principles and applications. In: Digital Libraries: Research and Development. Springer, Berlin, pp. 22–35.

Candela, L., Castelli, D., Pagano, P., Thano, C., Ioannidis, Y., Koutrika, G., Schuldt, H., 2007b. Setting the foundations of digital libraries: the DELOS manifesto. D-Lib Mag. 13 (3–4). Available from: http://www.dlib.org/dlib/march07/castelli/03castelli.html.

Carlson, S., Young, J., 2005. Google will digitize and search millions of books from 5 top research libraries. Chron. Higher Edu. 51 (18), A37–A40.

Carr, R., 2009. A history of digital library economics. In: Baker, D., Evans, W. (Eds.), Digital Library Economics: An Academic Perspective. Chandos Publishing, Oxford, UK, pp. 57–70.

Chen, H., Moore, J.L., Chen, W., 2015. Understand and analyzing learning objects: a foundation for long-term substantiality and use for elearning. Know. Manage. E-Learn. 7 (2), 280–296.

Chowdhury, S., Landoni, M., Gibb, F., 2006. Usability and impact of digital libraries: a review. Online Inform. Rev. 30 (6), 656–680.

Christenson, H., 2011. HathiTrust: a research library at web scale. Lib. Res. Tech. Serv. 55 (2), 93–102.

Concordia, C., Gradmann, S., Siebinga, S., 2010. Not just another portal, not just another digital library: a portrait of Europeana as an application program interface. IFLA J. 36 (1), 61–69.

Cottrell, M., 2013. A digital library for everybody. Am. Lib. 44 (3/4), 44–47.

Coyle, K., 2006. Mass digitization of books. J. Acad. Librar. 32 (6), 641–645.

Crane, G., 2006. What do you do with a million books? D-Lib Mag. 12 (3). Available from: http://www.dlib.org/dlib/march06/crane/03crane.html.

Crane, G. (2015). Perseus digital library. Available from: http://www.perseus.tufts.edu/hopper/

Creaser, C., Fry, J., Greenwood, H., Oppenheim, C., Probets, S., Spezi, V., White, S., 2010. Authors' awareness and attitudes toward open access repositories. New Rev. Acad. Librar. 16 (S1), 145–161.

De Rosa, C., Cantrell, J., Carlson, M., Gallagher, P., Hawk, J., Sturtz, C., 2005. Perceptions of libraries and information resources. A Report to the OCLC Membership. OCLC, Dublin, Ohio.

Dempsey, L., 2006. The (digital) library environment: ten years after. Ariadne 46. Available from: http://www.ariadne.ac.uk/issue46/dempsey.

Dobreva, M., Chowdhury, S., 2010. A user-centric evaluation of the Europeana digital library. In: Chowdhury, G., Khoo, C., Hunter, J. (Eds.), The Role of Digital Libraries in a Time of Global Change. Springer, Berlin, pp. 148–157.

Doerr, M., Gradmann, S., Hennicke, S., Isaac, A., Meghini, C., van de Sompel, H., 2010. The Europeana data model (EDM). In: World Library and Information Congress: 76th IFLA general conference and assembly pp. 10–15.

Downie, J.S., Dougan, K., Bhattacharyya, S., Fallaw, C., 2014. The HathiTrust Corpus: A Digital Library for Musicology Research? In: Proceeding of First International Digital Libraries for Musicology Workshop (DLfM 2014), ACM, New York, NY.

DPLA, 2015. Digital Public Library of America makes push to serve all 50 states by 2017 with $3.4 million from the Sloan and Knight foundations. Available from: http://dp.la/info/2015/06/26/digital-public-library-of-america-makes-push-to-serve-all-50-states-by-2017-with-3-4-million-from-the-sloan-and-knight-foundations/

Drabenstott, K., 1994. Analytical Review of the Library of the Future. Council on Library Resources, Washington, DC.

Dubis, M., 2003. Web resources for the study of New Testament backgrounds. JRTI 6 (1), 3–9.

Eichenlaub, N., 2013. Checking in with Google Books, HathiTrust, and other open access ebook collections. Comp. Lib. 33 (9), 4–9.

Entlich, R., Garson, L., Lesk, M., Normore, L., Olsen, J., Weibel, S., 1996. Testing a digital library: user response to the CORE project. Lib. Hi Tech 14 (4), 99–118.

Fenlon, K., Senseney, M., Green, H., Bhattacharyya, S., Willis, C., Downie, J.S., 2014. Scholar-built collections: a study of user requirements for research in large-scale digital libraries. Proc. Assoc. Inform. Sci. Technol. 51 (1), 1–10.

Fesenko, K., 2007. Models of digital cooperation. Slavic East Eur. Inform. Res. 8 (4), 87–97.

Fortenberry, N.L., Powlik, J.J., 1999. A national digital library for science, mathematics, engineering, and technology education. Comp. Appl. Eng. Edu. 7 (1), 45–49.

Fox, E., 1995. Digital libraries. Comm. ACM 38 (4), 23–31.

Fox, E.A., Gonçalves, M.A., Shen, R., 2012. Theoretical foundations for digital libraries: the 5S (societies, scenarios, spaces, structures, streams) approach. Synth. Lectures. Inform. Concepts Retrieval Serv. 4 (2), 1–180.

Fox, E.A., Urs, S.R., 2002. Digital libraries. Ann. Rev. Inform. Sci. Technol. 36, 503–589.

Ginsparg, P., 2011. ArXiv at 20. Nature 476 (7359), 145–147. Available from: http://www.nature.com/nature/journal/v476/n7359/full/476145a.html.

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 Trans. Inform. Syst. (TOIS) 22 (2), 270–312.

Greenstein, D., 2000. Digital libraries and their challenges. Lib. Trend. 49 (2), 290–303.

Gregory, L., Williams, S., 2014. On being a hub: some details behind providing metadata for the Digital Public Library of America. D-Lib Mag. 20 (7/8). Available from: http://dx.doi.org/10.1045/july2014-gregory.

Griffin, S.M., 2005. Funding for digital libraries: research past and present. D-Lib Mag. 11 (7/8). Available from: http://www.dlib.org/dlib/july05/griffin/07griffin.html.

Grudin, J., 2011. Human-computer interaction. Ann. Rev. Inform. Sci. Technol. 45 (1), 367–430.

Hanratty, E., 2005. Google Library: beyond fair use? Duke L. Tech. Rev. 2005, 10–26.

Haque, A.U., Ginsparg, P., 2009. Positional effects on citation and readership in arXiv. J. Am. Soc. Inform. Sci. Technol. 60 (11), 2203–2218.

Haque, A.U., Ginsparg, P., 2010. Last but not least: additional positional effects on citation and readership in arXiv. J. Am. Soc. Inform. Sci. Technol. 61 (12), 2381–2388.

Haslhofer, B., Isaac, A., 2011. Data. Europeana.eu: the Europeana linked open data pilot. In: International Conference on Dublin Core and Metadata Applications, pp. 94–104.

HathiTrust Digital Library, n.d. Millions of books online. Available from: http://www.hathitrust.org/.

Hill, L.L., Carver, L., Larsgaard, M., Dolin, R., Smith, T.R., Frew, J., et al., 2000. Alexandria Digital Library: user evaluation studies and system design. J. Am. Soc. Inform. Sci. 51 (3), 246–259.

Jeng, J., 2005. What is usability in the context of the digital library and how can it be measured? Inform. Technol. Lib. 24 (2), 47–56.

Joint, N., 2008. Current research information systems, open access repositories and libraries: ANTAEUS. Lib. Rev. 57 (8), 570–575.

Kahn, R., Wilensky, R., 1995. A framework for distributed digital object services. Corporation for National Research Initiatives. Available from: http://www.cnri.reston.va.us/kw.html.

Koohang, A., Ondracek, J., 2005. Users' views about the usability of digital libraries. British J. Edu. Technol. 36 (3), 407–423.

Krishnamurthy, M., 2008. Open access, open source and digital libraries: a current trend in university libraries around the world. Prog. Electron. Lib. Inform. Syst. 42 (1), 48–55.

Lagoze, C.J., 2010. Lost identity: the assimilation of digital libraries into the web. PhD dissertation, Cornell University.

Lagoze, C., Arms, W., Gan, S., Hillmann, D., Ingram, C., Krafft, D., et al., 2002. Core services in the architecture of the national science digital library (NSDL). In: Proceedings of the 2nd ACM/IEEE-CS joint conference on Digital libraries.ACM, New York, NY. pp. 201–209.

Lagoze, C., Krafft, D., Cornwell, T., Dushay, N., Eckstrom, D., & Saylor, J. (2006). Metadata aggregation and automated digital libraries: a retrospective on the NSDL experience. In: Proceedings of the 6th ACM/IEEE-CS joint conference on Digital libraries. ACM, New York, NY, pp. 230–239.

Lagoze, C., Krafft, D., Payette, S., Jesuroga, S., 2005. What is a digital library anyway, anymore? Beyond search and access in the NSDL. D-Lib Mag. 11 (11), Available from: http://dlib.org/dlib/november05/lagoze/11lagoze.html.

Larsgaard, M.L., Carver, L., 1995. Accessing spatial data online: project Alexandria. Information technology and libraries 14 (2), 93.

Law, D., 2009. Academic digital libraries of the future: an environment scan. New Rev. Acad. Librar. 15 (1), 53–67.

Law, D., 2011. Library landscapes: digital developments. In: Baker, D., Evans, W. (Eds.), Libraries and Society: Role, Responsibility and Future in an Age of Change. Chandos Publishing, Oxford, UK, pp. 361–377.

Leetaru, K., 2008. Mass book digitization: the deeper story of Google Books and the Open Content Alliance. First Monday 13 (10). Available from: http://firstmonday.org/ojs/index.php/fm/article/view/2101/2037.

Lesk, M. (1991, September). The CORE electronic chemistry library. In: Proceedings of the 14th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM, New York, NY, pp. 93–112.

Lesk, M., 1997. Practical Digital Libraries: Books, Bytes & Bucks. Morgan Kaufmann Publishers, San Franciso, CA.

Lesk, M., 2005. Understanding Digital Libraries. Morgan Kaufmann, San Francisco, CA.

Lesk, M., 2012. A personal history of digital libraries. Lib. Hi Tech 30 (4), 592–603.

Levy, D.M., 2000. Digital libraries and the problem of purpose. D-Lib Mag. 6 (1). Available from: http://www.dlib.org/dlib/january00/01levy.html.

Levy, D.M., 2003. Documents and libraries: a sociotechnical perspective. In: Bishop, A.P., Van House, N.A., Buttenfield, B.P. (Eds.), Digital Library Use: Social Practice in Design and Evaluation. MIT Press, Cambridge, MA, pp. 25–42.

Licklider, J.C.R., 1965. Libraries of the Future. MIT Press, Cambridge, MA.

Liu, Z., Luo, L., 2011. A comparative study of digital library use: factors, perceived influences, and satisfaction. J. Acad. Librar. 37 (3), 230–236.

Lucas, B.E., 2000. Media review essays: learning American memories on the whole World Wide Web. Hist. Rev. 27 (1), 152–156.

Luce, R.E., 2001. E-prints intersect the digital library: inside the Los Alamos arXiv. Issues Sci. Technol. Librar. 29.

Lyman, P., 1996. What is a digital library? Technology, intellectual property and the public interest. Daedalus 125 (4), 1–33.

Lynch, C.A., 1993. Accessibility and integrity of networked information collections. Background paper no. OTA-BP-TCT-109. Washington, DC: Office of Technology Assessmnet. Available from: http://files.eric.ed.gov/fulltext/ED368360.pdf.

Lynch, C., 2005. Where do we go from here?: the next decade for digital libraries. D-Lib Mag. 11 (7/8). Available from: http://www.dlib.org/dlib/july05/lynch/07lynch.html.

Ma, H., 2014. Tech services on the Web: DPLA: Digital Public Library of America http://dp.la. Tech. Serv. Quart. 31 (1), 83–84.

Madden, L., 2007. Digital Curation at the Library of Congress: Lessons Learned From American Memory and the Archive Ingest and Handling Test. Library of Congress, Washington, DC.

Madden, L., 2008. Applying the digital curation lessons learned from American Memory. Int. J. Digital Curation 3 (2), 121–129.

Mallery, M., 2015. Zephir, the HathiTrust metadata management system; http://www. hathitrust. org/zephir. Tech. Serv. Quart. 32 (3), 354–356.
Manfroid, S., Gillen, J., Phillips-Batoma, P.M., 2013. The Archives of Paul Otlet: between appreciation and rediscovery, 1944–2013. Lib. Trend. 62 (2), 311–328.
Marchionini, G., 2000. Evaluating digital libraries: a longitudinal and multifaceted view. Lib. Trend. 49 (2), 304–333.
Marchionini, G., Crane, G., 1994. Evaluating hypermedia and learning: methods and results from the Perseus Project. ACM Trans. Inform. Sys. (TOIS) 12 (1), 5–34.
Marcum, D.B., 1997. Digital libraries: for whom? For what? J.Acad. Librar. 23. (2), 81–84.
Matusiak, K.K., 2010. Use of digital resources in an academic environment: a qualitative study of students' perceptions, experiences, and digital literacy skills. PhD dissertation, University of Wisconsin, Milwaukee.
Matusiak, K.K., 2012. Perceptions of usability and usefulness of digital libraries. Int. J. Humanities Arts Comp. 6 (1–2), 133–147.
McCown, F., Bollen, J., Nelson, M.L., 2005. Evaluation of the NSDL and Google for obtaining pedagogical resources. In: Research and Advanced Technology for Digital Libraries. Springer, Berlin, pp. 344–355.
McMartin, F., Iverson, E., Wolf, A., Morrill, J., Morgan, G., Manduca, C., 2008. The use of online digital resources and educational digital libraries in higher education. Int. J. Digital Lib. 9 (1), 65–79.
Mischo, W.H., 2005. Digital libraries: challenges and influential work. D-Lib Mag. 11 (7/8), 636–651. Available from: http://www.dlib.org/dlib/july05/mischo/07mischo.html.
Morgan, G., Dziuban, C., McMartin, F., Morrill, J., Moskal, P., Wolf, A., 2013. Evaluating Faculty and Student Use of Digital Resources for Teaching and Learning. Available from: http://hdl.handle.net/2142/55341.
Morgan, G., Dziuban, C., McMartin, F., Morrill, J., Moskal, P., Wolf, A., 2014. Technical Report: Results From the Study: Student Use of Digital Learning Materials: implications for the NSDL. Available from: https://www.ideals.illinois.edu/handle/2142/55346.
Nicholas, D., Clark, D., 2014. Information seeking behaviour and usage on a multi-media platform: case study Europeana. In: Library and Information Sciences: Trends and Research, Springer, Berlin, pp. 57–78.
Nicholas, D., Clark, D., Rowlands, I., Jamali, H.R., 2013. Information on the go: a case study of Europeana mobile users. J. Am. Soc. Inform. Sci. Technol. 64 (7), 1311–1322.
OpenDOAR, 2014. About OpenDOAR. Available from: http://www.opendoar.org/about.html.
Palfrey, J., 2013. What is the DPLA? Lib. J. 138 (7), 38–41.
Paulmeno, M., 2015. The Digital Public Library of America: building a national digital library. Mississippi Lib. 78 (2). Available from: http://epubs.library.msstate.edu/index.php/MSLib/article/view/262/303.
Peroni, S., Tomasi, F., Vitali, F., 2013. Reflecting on the Europeana data model. In: Digital Libraries and Archives. Springer, Berlin, pp. 228–240.
Pinfield, S., 2005. A mandate to self archive? The role of open access institutional repositories. Serials: J. Serials Commun. 18 (1), 30–34.
Preece, E., Zepeda, C., 2009. The Perseus Digital Library: A case study. Available from: https://repositories2.lib.utexas.edu/bitstream/handle/2152/6836/perseus_case_study.pdf?sequence=5.
Pritchard, S.M., 2012. HathiTrust libraries map a shared path: a turning point in information access. Portal Lib. Acad. 12 (1), 1–3.
Purday, J., 2009. Think culture: Europeana.eu from concept to construction. Electron. Lib. 27 (6), 919–937.
Rayward, W.B., 1994. Visions of Xanadu: Paul Otlet (1868–1944) and hypertext. J. Am. Soc. Inform. Sci. 45 (4), 235.
Rayward, W.B., 1997. The origins of information science and the International Institute of Bibliography/International Federation for Information and Documentation (FID). J. Am. Soc. Inform. Sci. 48 (4), 299–300.
Rayward, W.B., 1999. H. G. Wells's idea of a world brain: a critical reassessment. J. Am. Soc. Inform. Sci. 50 (7), 270–312.
Rayward, W.B., 2005. The historical development of information infrastructures and the dissemination of knowledge: a personal reflection. Bull. Am. Soc. Inform. Sci. Technol. 31 (4), 19–22.

Reagle, J.M., 2010. Good Faith Collaboration: The Culture of Wikipedia. MIT Press, Cambridge, MA.
Saracevic, T., 2000. Digital library evaluation: toward an evolution of concepts. Lib. Trend. 49, 350–369.
Sare, L., 2012. A Comparison of HathiTrust and Google Books using federal publications. Pract. Acad. Librar. Int. J. SLA Acad. Div. 2 (1), 1–25.
Schatz, B., Chen, H., 1996. Building large-scale digital libraries. Computer 29 (5), 22–26.
Schwartz, C., 2000. Digital libraries: an overview. J. Acad. Librar. 26 (6), 385–393.
Smith, T.R., 1996. A digital library for geographically referenced materials. Computer 29 (5), 54–60.
Smith, T.R., Frew, J., 1995. Alexandria Digital Library. Commun. ACM 38 (4), 61–62.
Spinella, M.P., 2007. JSTOR: past, present, and future. J. Lib. Admin. 46 (2), 55–78.
Spinella, M., 2008. JSTOR and the changing digital landscape. Interlending Doc. Supply 36 (2), 79–85.
St. Clair, G., 2008. The Million Book project in relation to Google. J. Lib. Admin. 47 (1–2), 151–163.
Suber, P., 2009. Open access in 2008. J. Electron. Pub. 12 (1). Available from: http://quod.lib.umich.edu/cgi/t/text/idx/j/jep/3336451.0012.104/--open-access-in-2008?rgn=main;view=fulltext.
Tsakonas, G., Papatheodorou, C., 2007. Critical constructs of digital library interaction. In: Proceedings of the 11th Panhellenic Conference in Informatics (PCI 2007) pp. 57–66. Available from: http://eprints.rclis.org/10500/1/PCI2007_Tsakonas_Papatheodorou_eprints_.pdf.
Vandegrift, M., 2013. The Digital Public Library of America: details, the librarian response and the future. In: The Library with the Lead Pipe. Available from: http://www.inthelibrarywiththeleadpipe.org/2013/dpla/.
Walker, B., Schoonover, D., Margjoni, R., 2010. Creating a statewide JSTOR repository: initial steps taken by the Florida State University System. J. Interlib. Loan Doc. Del. Electron. Res. 20 (3), 159–172.
Waters, D.J., 1998. What are digital libraries? CLIR Issue 4, 1–6.
Weiss, A., 2014. Using Massive Digital Libraries: A LITA Guide. ALA Tech Source, Chicago.
Wells, H.G., 1938. The World Brain. Doubleday, Doran & Co, Garden City, NY.
Wilson, S., 2000. Navigating ancient worlds. Humanities 21 (5), 18–20.
Witten, I.H., Bainbridge, D., 2003. How to Build a Digital Library. Morgan Kaufmann Publishers, San Franciso, CA.
Wolf, A., Morgan, G., McMartin, F., Manduca, C., 2008. Faculty use of digital resources and the implications for digital libraries. Poster presented at 2008 Educause Annual Conference. Available from: http://d32ogoqmya1dw8.cloudfront.net/files/facultypart/faculty_use_digital_resources.pdf.
Xie, H.I., 2006. Evaluation of digital libraries: criteria and problems from users' perspectives. Lib. Inform. Sci. Res. 28 (3), 433–452.
Yang, S.C., 2001. An interpretive and situated approach to an evaluation of Perseus digital libraries. J. Am. Soc. Inform. Sci. Tech. 52 (14), 1210–1223.
Zachary, G.P., 1997. Endless Frontier: Vannevar Bush, Engineer of the American Century. Free Press, New York.
Zia, L. L., Pearce, L., Prinsen, J. G. B., Ticer, B. V., Peters, D., Pickover, M., Matylonek, J. C., et al., 2001. The NSF National Science, Technology, Engineering, and Mathematics Education Digital Library (NSDL) Program: New Projects and a Progress Report. D-lib Mag. 7(11). Available from: http://dlib.org/dlib/november01/zia/11zia.html.