

CHAPTER 1

Digital libraries and information access: introduction

Gobinda Chowdhury and Schubert Foo

Introduction

Systematic research and development activities in digital libraries began just over 20 years ago, and during this short period of time the field of digital libraries has progressed significantly. Over the past two decades a large number of digital libraries have appeared in different countries that cover different subjects, and disciplines from health to science, engineering to arts and culture, and others. Likewise, digital libraries have been designed, developed and used by a wide range of user communities that include school children, academics, scholars, scientists and the general public. Different types of content have been created and stored in these digital library repositories, ranging from basic digital objects like photographs, music and film to more research-oriented scholarly, scientific and research data.

Alongside this, a significant amount of resources and efforts have been invested in research into digital libraries that have given rise to over 8000 journal and conference papers and a large number of books, theses, research reports and other kinds of scholarly publications. Experts from a number of disciplines, like library and information science, computer science, engineering, psychology, business management, law, economics and others, have joined hands to address and resolve a variety of research issues and challenges associated with digital libraries.

The field of digital libraries has evolved significantly over the past two decades, both in terms of the nature and characteristics of research and in terms of the objectives and functionalities of digital libraries. While the first phase of digital library research in the early 1990s put the focus on building technologies for management of large volumes of digital information for remote access, this focus subsequently shifted to users, usability and impact studies, open access and so on. Subsequently, with the rapid progress in web and social networking technologies, the focus of digital library research has been extended to new and upcoming challenges such as semantic access, social information retrieval and social network analysis (see for details, Theng et al., 2009; Borgman, 2007; Goh and Foo, 2007). Furthermore, digital library research that once focused primarily on content – books, journals, music, video etc. – is now being

2 DIGITAL LIBRARIES AND INFORMATION ACCESS

extended to managing research data as well as research output, i.e. scholarly publications, in order to take a holistic view of research and scholarly activities and to develop information management systems that support all such activities (Borgman, 2011).

What is a digital library?

The term ‘digital library’ has been defined differently within the research communities, and over the years its definition has changed, reflecting the shifting focus of digital library research. Several researchers have discussed and analysed different definitions of digital libraries (see, for example, Arms, 2001; Borgman, 1999, 2000; Chowdhury and Chowdhury, 1999, 2003; Lesk, 2005). An earlier definition that came out of a workshop in 1994 (Gladney et al., 1994) considers a digital library as:

an assemblage of digital computing, storage, and communications machinery together with the content and software needed to reproduce, emulate, and extend the services provided by conventional libraries based on paper and other material means of collecting, cataloguing, finding, and disseminating information.

It is clear from this definition that earlier digital libraries aimed to emulate conventional analogue libraries. This is more obvious from the second part of the definition, which emphasizes that:

A full service digital library must accomplish all essential services of traditional libraries and also exploit the well-known advantages of digital storage, searching, and communication.

(Gladney et al., 1994)

Of course there are other definitions that point out other functions of a digital library; for example, the following one provided by the Digital Library Federation emphasizes the service as well as preservation activities:

Digital libraries are organizations that provide the 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.

(Digital Library Federation, 2004)

These early definitions emphasize that a digital library:

- is an organization, rather than a service
- provides access to digital works rather than information and data

- promises to provide future access, i.e. it has the responsibility of preservation and
- should provide access to digital works easily and economically.

However, the perception and connotation of a digital library have changed with the progress of research in digital libraries and, consequently, the emphasis has shifted from managing documents to managing information and data. The complexity of the notion of a digital library grew as more and more researchers joined the field from different disciplines. According to the digital library manifesto produced by the DELOS network of excellence (Candela et al., 2007):

Digital Libraries represent the meeting point of many disciplines and fields, including data management, information retrieval, library sciences, document management, information systems, the web, image processing, artificial intelligence, human–computer interaction, and digital curation.

The DELOS digital library manifesto (Candela et al., 2007) defines a digital library as a combination of:

- 1 An organization: A possibly virtual organization that comprehensively collects, manages, and preserves for the long term rich digital content, and offers to its user communities specialized functionality of that content, of measurable quality and according to codified policies;
- 2 A system: A software system that is based on a defined (possibly distributed) architecture and provides all functionality required by a particular Digital Library. Users interact with a Digital Library through the corresponding Digital Library System; and
- 3 A Digital Library Management System: A generic software system that provides the appropriate software infrastructure both (i) to produce and administer a Digital Library System incorporating the suite of functionality considered foundational for Digital Libraries and (ii) to integrate additional software offering more refined, specialized, or advanced functionality.

Furthermore, the DELOS manifesto points out that,

For every Digital Library [DL], there is a unique Digital Library System [DLS] in operation (possibly consisting of many interconnected smaller Digital Library Systems), whereas all Digital Library Systems are based on a handful of Digital Library Management Systems. For instance, through DILIGENT it is possible to build and run a number of DLSs, each realising a DL serving a target community.

(Candela et al., 2007)

The above definition indicates that a digital library uses a variety of software,

networking technologies and standards to facilitate access to digital content and data to a designated user community. However, a digital library is much more than an online search and retrieval system comprised of hardware, software, standards and networking technologies. Lagoze et al. (2005) comment that while search and access are its basic functions, a digital library 'facilitates the creation of collaborative and contextual knowledge environments'. The same notion of the digital library as a place for activities and collaboration has been put forward by other researchers as well. For example, Pomerantz and Marchionini (2007) argue that libraries have always been places that are dependent on the functionality, community and personal experiences of users, and digital libraries support new kinds of functionality for much broader communities and thus bring emerging senses of place.

Research in digital libraries

Many researchers have reviewed research and development in digital libraries and identified major areas of research interests and trends. In the first comprehensive review of digital library research during the first ten years (1989–99), Chowdhury and Chowdhury (1999) grouped digital library research papers into the following areas: collection development; development methodology and design issues; user interfaces; information organization: classification and indexing; resource discovery: metadata; access and file management; user studies; information retrieval; legal issues; social issues; evaluation of digital information; evaluation of digital libraries; standards; preservation; and management. They commented that while certain areas such as metadata, indexing and information retrieval, were well researched during the first decade, certain other areas, such as social issues of digital libraries, and the evaluation and impact of digital libraries were not studied widely. In a more recent study, Liew (2009) analysed research papers published during the decade 1997–2007 that focused primarily on social, cultural, legal, ethical, organizational and use dimensions of digital library research. She observed that while there is a significant body of research in the use and usability of digital libraries, as well as on the organizational, economic and legal issues, there is a lack of research on the ethical, social and cultural aspects of digital libraries.

It should be noted that while the first decade of digital libraries focused more on the organization and retrieval aspects, research in the second decade moved on to user and usability issues and the corresponding organization, economic and legal issues. However, until a few years ago, the social and cultural issues were not studied extensively. Of late there is an increasing amount of interest in the social and cultural aspects of digital libraries, as can be gleaned from the themes of the major digital library conferences; for example, cultural heritage is the main theme of two major international conferences on digital libraries to be held in 2012, viz. TPDL (Theory and Practice in Digital Libraries, formerly ECDL, European Digital Library Conference) and ICADL (International Conference on Asia Pacific Digital Libraries).

In another piece of research, focusing primarily on curriculum development for

digital library education, Fox and his collaborators at Virginia Tech University in the USA have developed the 5S model of digital libraries, where each of the five 'S' stands for (Pomerantz et al., 2006):

- Streams that describe all types of content – text, image, audio, video – and communication channels
- Structures that describe various schemes of organization of information such as catalogue codes, metadata schema, taxonomy, etc.
- Spaces that describe tools and techniques used in indexing, browsing, searching and interfaces
- Scenarios specified as system states, and events including machine processes, and
- Societies that include human users, system managers and machines.

The digital library curriculum based on the 5S model proposes 18 modules that cover different aspects of digital libraries, ranging from digitization and digital objects to metadata, thesaurus and ontology, architecture, searching and information seeking, filtering, summarization, visualization, archiving and preservation, intellectual property, etc. More recently Chowdhury and Son have attempted to create a knowledge map of digital library concepts, which contains 21 main topics and 1015 subtopics (for details, see Nguyen and Chowdhury, 2011).

Information access in digital libraries

Many large digital libraries exist today and the volume of their content and their growth rate are quite amazing. For example, the European Library (2011) provides access to 200 million records held in Europe's national libraries; the Europeana (2011) digital library, established in 2008, now has a collection of about 20 million digital objects; Trove (2011), a digital library service of the National Library of Australia, has over 250 million records; and so on.

Digital libraries are created to facilitate access to digital information by local or remote users. To be more precise, a digital library is created for a designated local or remote user community so that users can easily search, retrieve and use digital information that is relevant to their specific information needs. Thus information search and access is at the heart of digital library research and development activities. However, the concept of a designated user community is also very important here. All the activities related to the search, retrieval and use of information in a digital library should be designed to meet the information tasks and needs of a designated user community. Thus, information access has a broader scope than information retrieval. Agosti (2008, 2) comments:

The approach to the modeling of the information retrieval process has dramatically changed over the years, mostly in a positive and evolutionary way, with the final aim of

passing from an information retrieval approach towards an information access one where the real user is the focus of interest.

Thus, we may say that information access is the *raison d'être* of any digital library, and as an area of study it includes all the typical information retrieval processes and activities, ranging from content and data selection, and processing and indexing by a digital library, to search and retrieval, and use of information and data by a designated community of users in order to meet their information requirements. This book aims to cover research and development in different areas related to information access in digital libraries.

About this book

As discussed so far in this chapter, the field of digital libraries continues to be a very important area of research and development. The pervasive nature of the internet, the multi-disciplinary nature of digital libraries, and the myriad of information available through different forms of devices and channels provide a rich avenue for further studies, learning, innovation, research and development (R&D) and use. Research challenges and issues continue to evolve as researchers and developers are expected to meet the increasing demands of users, and therefore need to develop more user-friendly interfaces, intelligent search and retrieval capabilities, well endowed metadata description and content organization in scalable system architectures. The number of digital libraries around the world, and the amount of money invested in digital libraries R&D, have continued to grow over the past two decades. The field of research attracts a large and diverse audience and many new books on the subject have appeared since the early 1990s. In addition, there are several academic conferences focusing on digital libraries – notably the three established annual conferences taking place in different parts of the world: the ACM/IEE Joint Conference on Digital Libraries (JCDL), TPDF (formerly ECDL) and Asia's ICADL.

Over the years several texts and references on different aspects of digital libraries have appeared, a number of which have been published by Facet Publishing. However, the emergence of new research findings and the future importance of digital libraries provide a constant scope for new books on the subject. Furthermore, given the diversity of the field and the extensiveness of research in it, it is not always easy to provide a comprehensive view of the entire spectrum of digital library research in one book. Consequently, many recent books on digital libraries focus on one or more specific aspects of digital libraries. In a similar vein, this book focuses on information access and interactions in digital libraries.

However, given the diversity of digital libraries and the extensiveness of research, it is not always easy to provide a comprehensive view of the entire field of information access in digital libraries by just one or two experts. Keeping this in view, and based on

the editors' experience and an understanding of the level of research by various experts in relevant areas, an outline was prepared for this book that addressed the key areas of research in information access and interactions in digital libraries. A list of renowned experts in the field was then drawn from around the world to produce an up-to-date and authoritative resource on digital libraries that provides truly global coverage in terms of both content and expertise.

Altogether there are 15 chapters in this book. This chapter provides the background and an introduction to the book, beginning with an introduction to the concept of the digital library and then providing a brief overview of research in digital libraries, and finishing with a brief introduction to the content of the book.

Access to information depends quite significantly on the architecture of a digital library. In Chapter 2, Suleman discusses various issues related to the architecture of digital libraries. Core digital library design considerations are presented first, followed by a discussion of how these principles are used to build large digital libraries. The chapter then discusses how various digital libraries are interconnected into larger, networked digital library systems, exemplified by international projects such as the Networked Digital Library of Theses and Dissertations (NDLTD). Various challenges related to integration and scalability are then discussed. This is followed by a case study of an architecture designed specifically for digital libraries in developing countries that as yet do not have reliable and high penetration of internet access.

Metadata forms the foundation of a digital library in that it allows us to design a structure for better organization of and access to information. Metadata standards are created by expert communities of practice that specify the data elements and the corresponding features to be created in order to design a digital library database of one or more specific types of content and data. In Chapter 3, Shiri and Rathi discuss the importance of metadata, with special reference to user interface and interaction features of digital libraries. They review four digital libraries in order to discuss new developments in the use of metadata and to explore the emerging trends and new features and functionalities, such as social tags, recommendations, reviews and ratings in digital library user interfaces.

In Chapter 4, Chowdhury and Foo discuss various issues related to information access in digital libraries. Using examples from several large digital libraries, they discuss how access to information and its use depend on different factors such as the nature of the content and data, users, search interface features and the overall objectives and business of the digital library and the organization responsible for the digital library service concerned. The chapter also provides a brief review of literature and research projects in order to show the trends in R&D activities in relation to information access in digital libraries. Thus, this chapter provides the background for discussions of various issues related to information access that come later, in subsequent chapters of the book.

Recent developments in the web and social networking technologies have had significant implications for the access and use of digital libraries, especially in relation

to collaborative search and retrieval. In Chapter 5, Goh discusses collaborative search and retrieval in digital libraries. Drawing on the relevant research in information seeking and retrieval, he argues that collaboration with other users has always remained a central feature of information seeking. He then discusses research on collaborative querying, collaborative filtering and collaborative tagging or social tagging, which facilitate access to digital information.

In Chapter 6, Pang discusses how the penetration of the internet and social media, accompanied by the adoption of computing devices by the masses, has led to an increase in the amount of information accumulating online. She examines how social media technologies have implications for information access in digital libraries and points out a number of further developments of digital libraries in the online world.

Access to digital libraries requires extensive use of information and communication technologies (ICTs). However, there are significant gaps in our society in terms of access to technologies, as well as in users' information and digital literacy skills. Thus, various social issues such as the digital divide, social inclusion, information literacy and web accessibility are associated with access to information in digital libraries. In Chapter 7, Liew discusses these issues in the context of access to and use of digital libraries. Focusing specifically on the issue of social inclusion, she argues that the mere digitization of cultural heritage of a community does not guarantee social inclusion. She explores and seeks some answers to two major questions: what are the current barriers to a socially inclusive use of digital libraries, and what initiatives towards overcoming exclusion can be incorporated and implemented in digital library projects?

A digital library is created for users. Sometimes these users are well defined and their needs are known, while in most other cases the target user communities of a digital library could be anyone anywhere, and consequently their information needs and information behaviour are not well defined. Coupled with this, the rapid changes in ICT, especially in the area of web, mobile and social media technologies, have continuously reshaped and redefined users' needs and their interaction behaviour with digital libraries. In Chapter 8, Wilson and Macevičiūtė provide an overview of the current trends in research on the interactions between users and digital libraries. More specifically, they discuss the main results of user research in digital libraries and explore the applicability of existing models of information behaviour to the digital library sphere.

Many new developments have taken place that have direct or indirect implications for digital information access and use. Alongside the institutional and commercial digital libraries and information services, many new players have appeared in the digital information field such as Google Books and Google Scholar and Microsoft Academic Search, which have brought many new opportunities and challenges for digital library designers as well as for users. However, one technology that has permeated almost every sphere of today's society is the mobile technology. In Chapter 9, Kim, Durr and Hawamdeh discuss mobile technology for providing digital information services in libraries.

Parallel to the developments in commercial digital libraries and information services, a number of open access digital libraries have appeared and a variety of open scholarship activities have taken place over the past few years that have significantly influenced the digital library landscape, from both research and user perspectives. In Chapter 10, Chowdhury and Foo discuss the issues of open access and institutional repositories in the context of open access digital libraries. They briefly introduce the concept of open archives initiative and institutional repositories and then discuss how open access initiatives, coupled with recent developments in cloud computing technologies and associated developments, can help us to build digital libraries that are free at the point of use, and thus facilitate better access to and wider dissemination of knowledge.

A challenge is often faced in integrating several subject open access repositories and collections, like the National STEM Digital Library (NSDL), arXiv, Engineering Pathway and Internet Public Library, so as to enable the building of larger, interconnected and distributed digital libraries. In Chapter 11, Yang and Park discuss ongoing research on integrating and expanding taxonomy and subject categories derived from multiple repositories in science, technology, engineering and mathematics. They discuss the challenges involved in determining the semantic relationships between subject categories from different repositories through several text classification models and developing operations and processes for integration based on the identified subject category relationships from different repositories.

Usability is an important measure of the success of a digital library. Usability and user-centred design issues, their importance and the associated challenges and research issues, with particular reference to information access and interactions in digital libraries, are discussed by Chowdhury in Chapter 12.

Digital libraries provide access to a variety of information and data, but while some digital libraries can be accessed for free, such as the National Science Digital Library (NSDL) and Europeana, others require registration and payment of fees, such as the ACM Digital Library. In this respect, intellectual property and copyright are very important issues related to the design and management of digital libraries. In Chapter 13, Fraser discusses the economic challenges facing library and information services, open access initiatives and the provision of current copyright laws with regard to information services. After addressing the current debates on the economic viability of digital information services and the inadequacies of current intellectual property laws, he proposes a new framework for a digital copyright registry and discusses its features.

As discussed earlier in this chapter, one of the objectives of digital libraries is to ensure future access to information and data. This calls for preservation of digital information and data. In Chapter 14, Dobрева and Ruusalepp discuss current research and development activities in digital preservation and point out that ensuring interoperability among various systems and services remains a major challenge for future access to digital information and data.

In the last chapter of the book, Chowdhury and Foo summarize and discuss the trends in research in digital libraries, with special reference to information access and interactions. The digital library as a field of study is significantly influenced not only by research and developments within the field, but also by a variety of other factors, including the developments and challenges associated with the web, social networking, green ICT and cloud computing technologies, and various other socio-economic and political challenges, including the digital divide and digital literacy. In Chapter 15 Chowdhury and Foo address some of these issues and discuss how they will influence access and interactions in digital libraries.

Summary

As discussed in this chapter, research and development activities in digital library work have progressed quite significantly over the past two decades, giving rise to numerous large and small digital libraries, over 8000 journal and conference papers, and several books covering different aspects of digital libraries. Nevertheless, given the fast progress of research in the field and rapid developments in digital library tools, technology and standards and others, the need for more authoritative books is now felt more than ever. However, given the depth and diversity of research in the field, it is becoming increasingly difficult for any one author or any one book to cover the whole spectrum of research and development in digital library work. This book specifically covers the important aspect of information access and interactions in digital libraries. Written by a team of digital library experts from around the world, it aims to be an authoritative source for students, practitioners, researchers and developers in the field of digital libraries. The book is particularly well suited as a text and reference book, owing to the carefully selected scope of its coverage and its balanced focus on what has been achieved and what we can expect in the future, together with its identification of future challenges and trends in information access and interaction.

References

- Agosti, M. (2008) Information Access Using the Guide of User Requirements. In Agosti, M. (ed.), *Access through Search Engines and Digital Libraries*, Springer, 1–12.
- Arms, W. (2001) *Digital Libraries*, MIT Press.
- Borgman, C. L. (1999) What are Digital Libraries? Competing visions, *Information Processing and Management*, **35** (3), 227–43.
- Borgman, C. L. (2000) *From Gutenberg to the Global Information Infrastructure: access to information in the networked world*, ACM Press.
- Borgman, C. L. (2007) *Scholarship in the Digital Age: information, infrastructure, and the internet*, MIT Press.
- Borgman, C. L. (2011) The Conundrum of Sharing Research Data, *Journal of the American Society for Information Science and Technology*,

- http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1869155.
- Candela, L., Castelli, D., Ioannidis, Y., Koutrika, G., Pagano, P., Ross, S., Schek, H., Schuldt, H. and Thanos, C. (2007) Setting the Foundations of Digital Libraries: the DELOS manifesto, *D-Lib Magazine*, **13** (3/4), www.dlib.org/dlib/march07/castelli/03castelli.html.
- Chowdhury, G. G. and Chowdhury, S. (1999) Digital Library Research: major issues and trends, *Journal of Documentation*, **55** (4), 409–48.
- Chowdhury, G. G. and Chowdhury, S. (2003) *Introduction to Digital Libraries*, Facet Publishing.
- Digital Library Federation (2004) *A Working Definition of Digital Library* [1998], <http://old.diglib.org/about/dldefinition.htm>.
- European Library, The (2011) <http://search.theeuropeanlibrary.org/portal/en/index.html>.
- Europeana (2011) www.europeana.eu/portal/.
- Gladney, H. H., Fox, E. A., Ahmed, Z., Asany, R., Belkin, N. J. and Zemankova, M. (1994) *Digital Library: gross structure and requirements: report from a March 1994 workshop*, www.csdlib.tamu.edu/DL94/paper/fox.html.
- Goh, H. L. D. and Foo, S. (2007) *Social Information Retrieval Systems: emerging technologies and applications for searching the web effectively*, Information Science Reference, an imprint of IGI Global.
- Lagoze, C., Krafft, D. B., Oayette, S. and Jesuroga, S. (2005) What is a Digital Library Anymore, Anyway? Beyond search and access in the NSDL, *D-Lib Magazine*, **11** (11), www.dlib.org/dlib/november05/lagoze/11lagoze.html.
- Lesk, M. (2005) *Understanding Digital Libraries*, 2nd edn, Morgan Kaufman
- Liew, C. L. (2009) Digital Library Research 1997–2007: organisational and people issues, *Journal of Documentation*, **65** (2), 245–66
- Nguyen, H. S. and Chowdhury, G. (2011) Digital Library Research (1999–2010): a knowledge map of core topics and subtopics. In Xing, C., Crestani, F. and Rauber, A. (eds), *Digital Libraries: for cultural heritage, knowledge dissemination and future creation, 13th International Conference on Asia-Pacific Digital Libraries, ICADL2011, Beijing, October 24–27, 2011*, 367–71.
- Pomerantz, J. and Marchionini, G. (2007) The Digital Library as Place, *Journal of Documentation*, **63** (4), 505–33.
- Pomerantz, J., Wildemuth, B. M., Yang, S. and Fox, E. A. (2006) Curriculum Development for Digital Libraries, *Proceedings of the 6th ACM/IEEE-CS Joint Conference on Digital Libraries, JCDL 2006, June 11–15, 2006, Chapel Hill, North Carolina*, 175–84.
- Theng, Y. L., Foo, S., Goh, H. L. D. and Na, J. C. (2009) *Handbook of Research on Digital Libraries: design, development and impact*, IGI Global.
- Trove (2011) *One Search: a wealth of information*, National Library of Australia, <http://trove.nla.gov.au>.