

# Exploring Digital Libraries



# Exploring Digital Libraries

Foundations, Practice, Prospects

KAREN CALHOUN

**Neal-Schuman**

An imprint of the American Library Association

Chicago • 2014

[www.alastore.ala.org](http://www.alastore.ala.org)

© 2014 Karen Calhoun

First published in the United Kingdom by Facet Publishing, 2014.  
This simultaneous U.S. Edition published by Neal-Schuman,  
an imprint of the American Library Association, 2014.

18 17 16 15 14 5 4 3 2 1

ISBN: 978-1-55570-985-3 (paper)

[www.alastore.ala.org](http://www.alastore.ala.org)

.....

# Contents

<b>List of figures and tables.....</b>	<b>ix</b>
<b>Preface .....</b>	<b>xi</b>
<b>Acknowledgements .....</b>	<b>xv</b>
<b>Glossary .....</b>	<b>xix</b>
<b>1 Emergence and definitions of digital libraries .....</b>	<b>1</b>
Overview .....	1
The emergence of digital libraries (1991–2001) .....	1
Early digital-library projects.....	15
Definitions of digital libraries .....	18
Conclusion .....	26
<b>2 Outcomes of digital libraries’ first decade .....</b>	<b>27</b>
Overview.....	27
A new field of research and practice .....	28
The transformation of scholarly communication processes.....	32
Technical innovations .....	36
Digitization and digital preservation.....	42
Metadata and standards .....	48
Working digital libraries .....	51
Conclusion .....	52
References to websites in Table 2.1 .....	52

<b>3</b>	<b>Key themes and challenges in digital libraries.....</b>	<b>59</b>
	Overview.....	59
	The key themes of digital library work.....	59
	Key challenges.....	65
	Conclusion.....	84
<b>4</b>	<b>Digital library collections: repositories.....</b>	<b>85</b>
	Overview.....	85
	The traditional library worldview.....	85
	Repositories, libraries and the web.....	87
	The evolution of digital library repositories.....	89
	Conclusion.....	109
<b>5</b>	<b>Hybrid libraries.....</b>	<b>111</b>
	Overview.....	111
	Changing information-seeking behaviors.....	111
	Libraries’ response: changing hybrid library collections.....	116
	Changing technologies for hybrid libraries.....	123
	Conclusion.....	135
<b>6</b>	<b>Social roles of digital libraries.....</b>	<b>139</b>
	Overview.....	139
	Introduction.....	139
	Foundations of digital libraries’ social roles.....	142
	A possible framework of social roles.....	144
	Conclusion.....	157
<b>7</b>	<b>Digital libraries and their communities.....</b>	<b>159</b>
	Overview.....	159
	Approach.....	159
	Successful, sustainable digital libraries.....	159
	Inception: purpose and focus.....	162
	Inception: branding and awareness.....	163
	Creation: community and needs orientation.....	165
	Creation: user-centered design, ease of use and reliability.....	169
	Growth: quality content.....	171
	Maturity: funding and sustainability.....	171
	Conclusion.....	176

<b>8</b>	<b>The prospects of open access repositories .....</b>	<b>179</b>
	Overview .....	179
	Successful subject-based repositories .....	179
	The value of institutional repositories .....	180
	The policy and legal frameworks .....	186
	Deposit mandates.....	190
	Other issues with self-archiving.....	192
	'Google has won' .....	194
	Making institutional repositories more valuable.....	196
	The future of repositories.....	198
	Conclusion .....	207
<b>9</b>	<b>Digital libraries and the social web: scholarship.....</b>	<b>209</b>
	Overview .....	209
	Introduction .....	209
	Background: Web 2.0 and Library 2.0 .....	210
	What is the social web? .....	211
	Digital libraries and the social web.....	212
	Digital libraries' social evolution: a visual framework .....	214
	Scholarship .....	215
	From personalization to collaboration.....	219
	Scholarly collaboration on the social web .....	222
	Conclusion .....	234
<b>10</b>	<b>Digital libraries and the social web: collections and platforms ....</b>	<b>237</b>
	Overview .....	237
	Visualizing the shift from collections to platforms .....	237
	The dilemma of the national or local collections focus.....	237
	Optimizing the reach and visibility of digital libraries.....	241
	Mass digitization and digital libraries.....	250
	Large-scale digital libraries, portals and platforms.....	251
	Crowdsourcing and citizen science .....	252
	Conclusion .....	254
	<b>References .....</b>	<b>257</b>
	<b>Index.....</b>	<b>311</b>

.....

## List of figures and tables

### Figures

Figure 1.1	Validation of the predictions of runaway growth: more information and more consumption .....	8
Figure 2.1	Key outcomes of the first decade of digital libraries .....	27
Figure 2.2	Digital library documents indexed in Google Scholar and Scopus, 1990–2012 .....	30
Figure 3.1	Concept map of major themes of the digital library literature (2002–2012) .....	64
Figure 3.2	Aspects of digital library sustainability .....	83
Figure 4.1	Timelines for repositories, the internet and web, and hybrid libraries .....	90
Figure 4.2	Contents of OpenDOAR repositories, mid-2013 .....	95
Figure 5.1	Trends in the use of public and academic research library traditional collections .....	114
Figure 5.2	Rising e-serials expenditures, dropping monograph expenditures .....	117
Figure 5.3	Technologies and tools supporting hybrid library business processes.....	124
Figure 6.1	A framework of social roles of digital libraries.....	145
Figure 7.1	Life-cycle model of success factors for digital libraries in social environments .....	161
Figure 8.1	Force field analysis of the emergence of a sustainable, global ecosystem of repositories .....	206
Figure 9.1	Transitions associated with the shift to social digital libraries .....	214



Figure 9.2 The evolution of digital libraries toward new roles on the social web.....215

Figure 9.3 Social web impacts on researchers and scholarship .....217

Figure 10.1 The evolution of digital library collections on the social web .....238

**Tables**

Table 1.1 A progression of digital library definitions.....21

Table 2.1 A sample of digital libraries, their histories and funding .....53

Table 3.1 Summary of frequently occurring keywords or keyword phrases .....62

Table 3.2 Issues associated with digital libraries and copyright.....81

Table 4.1 Summary information: a sample of subject-based repositories worldwide .....100

Table 4.2 Summary information: a sample of institutional repositories worldwide .....101

Table 6.1 Potential social contributions of digital libraries.....146

Table 6.2 Some brief notes on digital preservation of selected types of content .....154

Table 7.1 Aligning digital libraries with community needs and practices: some examples.....166

Table 7.2 Key technological success factors .....170

Table 7.3 Strategies for sustainability .....173

Table 8.1 The value of institutional repositories to stakeholders and target audiences.....183

Table 8.2 Barriers and possible service responses for institutional repositories.....197

Table 9.1 Leading social networks specifically oriented to researchers, 2013.....226

Table 9.2 Selected researcher identification services .....231

.....

## Preface

The emergence of digital libraries in the early 1990s was a turning point and a critical component of the worldwide, world-changing shift to networked information. The turning point crystallized from myriad disparate elements that had influenced and combined with one another, in the process creating something entirely new. The high energy and important outcomes of that first, formative digital library decade created further winds of change that in turn produced more vigorous activity. More than 20 years on, we can observe a young, multidisciplinary field of digital library research and practice that has now reached a new intersection of dynamic forces – one dominated by the social web. This is the context in which I wrote this book.

The intention of the book is to help advance learning, professional discourse, research and practice at this particular moment in the evolution of digital libraries. I have done my best to take a broad, international perspective. At one level, I have approached the subject generally: the foundations, current practice and prospects of digital libraries. A second structural theme runs through the book as well: digital libraries in the context of the web-based communities they serve. In particular I explore the possibilities of a shift toward a community-centered perspective based on what digital libraries contribute or might contribute in a time of large-scale, web- and network-driven societal change.

The book attempts to place insights gleaned from the record of digital libraries' first decade, the second decade's literature of research and practice, conversations with several digital library experts, and my own research and experiences into an overarching framework that can make sense of an enormous amount of activity. Over the course of twenty-some years, thousands of people from multiple disciplines and domains have contributed

to this young field of endeavor. The book's structure, topics and analyses reflect my own perspective, interpretations and knowledge of all this work; it may be an understatement to say there are other perspectives. I have cited many sources in this book, not only to place the chapters and sections in the context of related work, but also to guide those who wish to explore specific topics in more depth or from other points of view.

The challenges facing libraries, including digital libraries, are momentous. In this book, I have tried to lay out the issues and opportunities in a way that invites progress toward addressing the issues and seizing the opportunities. While it explores the past and present, the book looks forward. Thus, I have written mainly for the individuals and groups who will define and carry out the next stages of the digital shift in libraries, cultural memory institutions and information infrastructures: practicing professionals, graduate students and their professors. In addition, library and information science leaders and researchers may find some of the book's analytical frameworks of interest. Finally, the second half of the book may contain some relevant material for those interested in the broad question of the social roles of libraries in the digital age.

Chapter 1 of the book traces the emergence of the field and offers a definition of 'digital libraries' in an environment of multiple perspectives and continuous change. Chapter 2 describes the outcomes of the first decade of digital libraries (1991–2001) that continue to shape digital library research and practice today: among them the transformation of scholarly communication processes, the open access movement, large-scale digitization, and working digital libraries. Chapter 3 introduces a concept map that identifies the key themes and challenges of digital libraries' second decade (2002–12). Chapters 4 and 5 discuss digital libraries and the web through the lens of collections and collection building, with Chapter 4 devoted to repositories and Chapter 5 examining hybrid libraries (collections containing both digital and non-digital resources).

The second half of the book turns to the social roles of digital libraries and their value in their communities. Chapters 6 to 8 explore digital libraries in the context of a framework of actual and potential social roles, with Chapter 6 laying out the framework, Chapter 7 examining the nature of online communities, and Chapter 8 looking into the prospects of open access repositories. Chapters 9 and 10 offer a visual framework for considering the possibilities of digital libraries as social platforms on the web, with Chapter 9 exploring the influence of the social web on scholarship, and Chapter 10 examining a number of opportunities afforded by the social web to go beyond

what digital libraries have achieved in their first two decades.

A great deal has been written and said about specific aspects of digital libraries, but not that much that examines their progress or prospects in terms of their overarching societal or community impact. In addition to this book's aim to summarize a large body of research and practice into a manageable resource for teaching and learning, it has the goal of encouraging new discourse and reflection about digital libraries' contributions to democratic societies. Recognizing that we find ourselves in a chaotic social and technical environment, facing significant uncertainty about the future, it is more important than ever to collectively take up this conversation.

.....

## Acknowledgements

The process of creating this book has involved examining the work of hundreds of researchers and practitioners who took the time to collect their thoughts and write. I want to start by thanking them, as this book builds on what they studied, learned and shared. The digital library field has produced a large body of literature, impossible for any one person to absorb, so I apologize in advance for any key contributions I have missed, and I welcome hearing about them.

Many individuals have personally supported me through the lengthy process of forming and articulating the ideas in this book. First and foremost, I would like to thank my husband for his unflagging patience and support during the 16 months or so when I was devoting any free time that I had to writing. Second, without the encouragement, generosity and flexibility of Rush Miller, university librarian at the University of Pittsburgh, I would not have been able to write this book.

Many have helped to guide the writing. Among them, I want to thank Howard D. White, professor emeritus of Drexel University's College of Information Science and Technology, whose teaching and advice resulted in my first steps, over 30 years ago, toward professional research, writing and publication. More recently, my friend and former colleague Lorcan Dempsey, OCLC's chief strategist and vice president of research, has helped to shape my thinking over the past decade; he was the first to make me believe that I could write this book, and he has been generous with looking at drafts, helping me sort out ideas and referring me to useful sources. Sheila Corrall, professor and chair of the University of Pittsburgh's library and information science program, took time from an incredibly busy year to read drafts and provide very specific help and guidance; I cannot thank her

enough. In addition, the encouragement, enthusiasm and comments of Helen Carley and Sarah Busby, my editor and former editor at Facet Publishing, have been most welcome.

Early conversations with Ron Larsen, dean and professor of Pitt's School of Information Sciences (SIS), and Stephen Griffin, visiting professor at SIS and a former program director at the National Science Foundation (NSF), helped me to understand the impact and outcomes of NSF's Digital Library Initiatives. Aaron Brenner, head of Pitt's digital library program, helped from the writing's beginning to the final drafts with contributions of time, critiques and suggestions. Jordan Hatcher, an expert in intellectual property and internet law, kindly found time to talk with me about legal language. Elvia Arroyo-Ramirez, at the time a new graduate of the University of Pittsburgh, brought the perspective of a new professional to reviewing early chapters and helped me to find this book's voice.

The later chapters of this book incorporate the results of interviews with nine well known digital library experts. I am grateful for their willingness to share their experience and insights, which have underpinned and enriched these chapters. They are:

- William Arms, professor emeritus, Computing & Information Science, Cornell University
- Jill Cousins, executive director, Europeana
- Ricky Erway, senior program officer, OCLC Research and former associate co-ordinator of the American Memory program at the Library of Congress
- Rose Holley, digital library specialist, National Archives of Australia and former manager of the Australian Newspaper Digitization Program at the National Library of Australia
- Carl Lagoze, associate professor, School of Information, University of Michigan
- Clifford Lynch, executive director, Coalition for Networked Information
- Deanna Marcum, managing director, ITHAKA S+R
- Oya Rieger, associate university librarian for digital scholarship and preservation services, Cornell University Library
- Roy Tennant, senior program officer, OCLC Research.

I also want to express my gratitude to Dawn and Jim Metz, owners of an inn in the small town of Ligonier, where I spent three highly productive writing retreats. Their warm hospitality and wonderful house gave me space and

peace to fill these pages. I also benefited greatly from the space and kind support of two public libraries; my thanks are due to Diana Falk and her staff of the Norwin Public Library and to Janet Hudson and her staff of the Ligonier Valley Public Library.

I did not realize what the journey of writing this book was going to entail when I first began. Perhaps if I had fully realized the scope of what I was setting out to do, I would not have started. I want to express my heartfelt appreciation to those I have acknowledged here; they made the journey possible. Finally, I am grateful to the many other colleagues, friends and family who contributed to completing this book, perhaps without knowing they were doing so.

**Karen Calhoun**

.....

## Glossary

This glossary provides brief, informal definitions of terms in the context of digital libraries; some of the terms have other meanings in different contexts.

**ACM** Association for Computing Machinery

**altmetrics** A subfield of bibliometrics that includes quantitative measures based on engagement and impact in the context of the social web

**API** Application Programming Interface; a machine interface that allows and responds to requests for services from other computer programs and/or allows data to be exchanged

**ARL** Association of Research Libraries (US)

**artifacts** Physical items in library, museum or archival collections

**arXiv** A subject-based repository of pre-prints in the fields of mathematics, physics, astronomy, computer science, quantitative biology, statistics, and quantitative finance

**ASEO** Academic search engine optimization; the creation, publication, and modification of scholarly content in a way that makes it easier for academic search engines to both crawl and index it

**associative indexing** A form of information organization and retrieval based on the relationships (links) between things; a concept often credited to Vannevar Bush and his idea for the 'memex'

**authentication** The automated process of validating the identity of a person or computer

**authorization** The automated process of providing the appropriate rights to access information or carry out particular actions

**BASE** Bielefeld Academic Search Engine; BASE collects and indexes digital academic content from repositories and other sources worldwide



**BIBFRAME** An undertaking by the Library of Congress and the library community to determine a transition path for MARC21 to linked data standards

**bibliometrics** A research method in library and information science that uses quantitative assessment and statistics to describe publication patterns. One bibliometric method is citation analysis

**BnF** Bibliothèque nationale de France

**branding** In the context of this book, the process of successfully communicating the identity, intent and nature of a digital library to the target audiences it aims to serve

**business model** How a product, service or organization will generate income and/or cover costs

**business processes** Activities that produce services or projects. Some library business processes are selecting, acquiring, describing and managing, supporting discovery and access, circulating or linking and preserving collections

**CERN** European Organization for Nuclear Research

**citation management tools** Tools that support scholarly research and writing; they capture, organize and maintain citations (references) and enable using the captured data in multiple ways, most commonly to create lists of references or bibliographies

**citizen science** A form of research collaboration involving members of the public in scientific research projects that address real-world problems

**cloud services** Web-based applications that are accessed via common web browsers and whose infrastructure is hosted on the network (i.e., 'in the cloud')

**CNI** Coalition for Networked Information

**Creative Commons licenses** A set of template copyright licenses available to the public that allow for certain types of free, worldwide distribution of copyrighted works. Creative Commons ([creativecommons.org](http://creativecommons.org)) is the best known of several new models for licensing rights to content

**CrossRef** A reference linking service operated by a consortium of publishers; the official DOI registration agency for scholarly and professional publications

**crowdsourcing** A massive collaboration technique that enables individuals, working as a virtual group, to collectively accomplish a shared, large and significant goal

**cyberinfrastructure** An infrastructure of high-speed networks, computing and data systems, repositories, and instruments that supports data-

- intensive research environments and enables advanced capabilities
- DARPA** Defense Advanced Research Projects Agency (US)
- data curation** Activities including selection, organization and access, preservation, maintenance and archiving of scientific data
- deep web** The part of the web that is not indexed by common search engines; also called the invisible or hidden web
- deposit mandates** Policies requiring authors to make their work available via open access in a particular repository or repositories
- destination site** On the web, a site that people visit by going directly to its URL through typing or bookmarks
- digital curation** A series of activities that support creating, acquiring, appraising, repurposing, accessing, and preserving digital collections
- digital divide** The gap between those with and without access to quality information resources and to information and communications (ICT) technologies
- digital libraries** 1. A field of endeavor with participants chiefly from the computer, library and information sciences, publishing, the cultural heritage sector, and education; 2. Systems and services, often openly available, that support the advancement of knowledge and culture, contain managed collections of digital resources, and often use an architecture centered on a repository
- digital preservation** The active management and appraisal of digital information over its entire life cycle, from creation and active use to selection, transfer and preservation, access and re-use
- digitization** The process of converting a physical item into a digital format, sometimes known as digital reformatting
- disclosure** In the library context, the results of activities to make library content discoverable and visible in or from external systems and services, especially search engines and on high-traffic sites
- discovery service** In the library context, a type of web-based user interface that provides for unified, integrated search and retrieval based on a pre-harvested, centralized index to heterogeneous resources
- DLF** Digital Library Federation
- DLI** Digital Libraries Initiative, a US-based digital library research program that began in the early 1990s
- DOAJ** Directory of Open Access Journals
- DOI** Digital Object Identifier, a type of identifier used to identify scholarly publications
- DRIVER** Digital Repository Infrastructure Vision European Research;

- DRIVER** harvests and provides access to content from repositories in 38 countries
- DSpace** A well known open source software package for building and managing repositories
- Dublin Core** A small set of metadata elements used to describe digital resources; maintained by the Dublin Core Metadata Initiative
- EAD** Encoded Archival Description, used to create electronic finding aids for archival materials
- EDM** Europeana Data Model
- eLib** The Electronic Libraries Programme, a UK-based digital library development program that ran from 1994 to 2001
- EPrints** A well known open source software package for building and managing repositories
- e-research** Data- and computationally-intensive, distributed, collaborative research activities that use advanced information and communication technologies; see also **cyberinfrastructure**
- ERM** E-resource management system
- Europeana** A portal to a multi-lingual online collection of millions of digitized items from European museums, libraries, archives and multi-media collections
- fair dealing, fair use** A legal provision that provides exceptions to copyright law when the use of copyrighted material is for the purposes of non-commercial research or study, criticism, review or reporting. It is related to the US concept of 'fair use'
- Fedora** Flexible Extensible Digital Object Repository Architecture; an open source software system for providing a core repository service
- FRBR** *Functional Requirements for Bibliographic Records*, a conceptual model for metadata developed by IFLA that provides a framework for expressing a work in terms of its attributes and relationships to other objects/resources
- Gold open access** The provision of open access by publishing in an open access journal
- Google 5** The five libraries that entered into a 2004 agreement with Google to mass-digitize their collections; this developed into the **Google Library Project**, with more research library participants
- Green open access** The provision of open access by publishing in journals and self-archiving a copy on the web or in an open access repository
- HathiTrust** A membership organization for creating a shared repository of digital collections; most of the content is from libraries that participated

in the Google Library Project

**HTML** Hypertext Markup Language, a simple language used on the web for marking up and formatting text and linking to other objects

**HTTP** Hypertext Transfer Protocol, the basic protocol of the web for communication between browsers and websites

**hybrid library** Combinations of traditional collections, licensed e-resources and openly available digital collections produced in-house or elsewhere

**i2010 Digital Libraries** A European Commission plan, announced in 2005, to build a European digital library by 2010; Europeana developed from this initiative

**identifiers** In the digital library domain, references to digital objects that are intended to be long-lasting or persistent; an essential characteristic of a digital library resource. Some examples of identifiers discussed in this book are DOIs, PURLs, URIs, and ORCIDs

**IEEE** A professional association for advancing technology. Originally the acronym stood for Institute of Electrical and Electronics Engineers, but the scope of the association is now much broader; therefore the organization is now known simply as IEEE

**IFLA** International Federation of Library Associations

**IMLS** Institute of Museum and Library Services (US)

**institutional repository** A repository for collecting, preserving and disseminating, in digital form, the intellectual output of an institution

**interoperability** The provision of uniform, coherent access for users to diverse information from different, independently managed systems

**IR** Information retrieval

**ISO** International Organization for Standardization

**JISC** Joint Information Systems Committee (UK)

**JSTOR** A digital library subscription service for scholarly journals, primary sources and books

**Kahn-Wilensky architecture** Principles for the design of a digital library that is open in its architecture, with four main components: a repository, mechanisms to support search, identifier systems, and user interfaces. The architecture could contain additional components for supporting other services

**knowledgebases** Digital registries (machine- or human-readable, usually both) that collect and organize metadata and content needed for specific functions, like managing access rights information

**landing page** The web page that appears when a user clicks on a link from a search engine result list or other referring site

**learning management systems** Integrated software applications for supporting network-based learning environments

**Library 2.0** A conceptualization of the 'next generation' library, characterized by user interaction, participation and user-centered design

**library management systems** Integrated software applications, generally based on relational databases, that support the business processes of libraries

**linked data** The framework for publishing and consuming semantic web content, which consists of interlinked, structured data. Linked data can also refer to the collection of interrelated, linked datasets in the web of data

**LIS** Library and information science

**LOCKSS** Lots of Copies Keep Stuff Safe, a community effort and open source system allowing libraries to collect, preserve and provide access to digital content

**long tail** A concept made popular by Chris Anderson of *Wired* magazine that refers to obscure, low-use content as the 'long tail' of a demand curve. His hypothesis is that network-level aggregation and disclosure of these obscure items will also generate demand and open new markets for their use

**MARC** Machine-Readable Cataloging

**mashups** Interactive web applications that gather content from diverse sources to create new services. An example is the combination of Google Maps content with other information displayed to a user

**mass digitization** The digitization of very large, whole collections of content, with no or minimal selection

**mediated deposit** Self-archiving that is accomplished through a mediator, e.g., a manager of an institutional repository

**metadata** Structured information that describes, explains, locates and otherwise makes it easier to retrieve and use an information resource (NISO definition)

**MIT** Massachusetts Institute of Technology

**MOOC** Massive open online course

**NCBI** National Center for Biotechnology Information (US)

**NDIIPP** National Digital Information Infrastructure and Preservation Program (US)

**NDLTD** Networked Digital Library of Theses and Dissertations

**needs assessment** A process that studies user needs, wants and tasks in order to use this information in designing or selecting systems or services

- NIH** National Institutes of Health (US)
- NISO** National Information Standards Organization (US)
- NLA** National Library of Australia
- NSF** National Science Foundation (US)
- OAI** Open Archives Initiative
- OAI-ORE** Open Archives Initiative Object Reuse and Exchange; a set of standards for the description and exchange of aggregations of web resources
- OAI-PMH** Open Archives Initiative Protocol for Metadata Harvesting
- OAIS** Open Archival Information System
- OASIS** Organization for the Advancement of Structured Information Standards
- OCLC** Online Computer Library Center
- OCR** Optical character recognition, an automated process for creating a digital representation of text
- OECD** Organization for Economic Cooperation and Development
- online communities** Network-based spaces in which participants communicate and interact, share and distribute content, and build or maintain relationships. Examples of types of online communities are support, interest-based, gaming and knowledge communities
- open access** The provision of online, free-of-charge access to content that is also free of most copyright and licensing restrictions
- open source software** Software for which the source code is freely available. To be open source software, it should be licensed under an Open Source Initiative (OSI)-compliant open source software license (<http://opensource.org> has details)
- OpenDOAR** Directory of Open Access Repositories
- OpenURL** A standard for enabling linking from one information resource to another; OpenURLs work by capturing and transferring metadata about an information resource to another service, then retrieving and displaying the information resource to the user
- ORCID** Open Researcher and Contributor ID
- orphan works** Copyrighted works whose owner is unknown or cannot be located. Orphan works are a problem because potential users of the orphan works have no one to seek permission from when doing clearances
- PageRank** The link analysis algorithm used by the Google search engine, based on the principle of associative indexing and named after Larry Page, a co-founder of Google

**PDA** Patron Driven Acquisitions, a model for licensing and purchasing e-books ‘just in time,’ based on library user selections

**peer review** The professional evaluation of a work by a colleague in the same field

**personas** In user-centered design, a representation of the goals and behavior of a hypothesized type of user. Personas, typically developed from user interviews, are used to guide service design, development or enhancement

**portal** A term used in the library community to refer to a service for simplifying searching across and linking from and to diverse collections and making it easier to authenticate and authorize access to licensed resources. Portal implementations in the first half of the 2000s were generally associated with metasearch or federated searching

**post-prints** Digital representations of scholarly articles after they have been peer-reviewed

**pre-prints** Digital pre-publications (drafts) of scholarly articles

**PubMed Central** A subject-based repository of biomedical and life sciences journal literature operated by the National Center for Biotechnology Information (US). PubMed Central also has centers in Europe and Canada

**PURL** Persistent Uniform Resource Locator

**RDA** *Resource Description and Access*, an international standard for cataloging and the successor to AACR2 (the *Anglo-American Cataloging Rules* second edition revised)

**RDF** Resource Description Framework, a method supporting interoperability that allows structured and semi-structured data to be mixed, exposed and shared across different applications

**referee** See **peer review**

**reference linking** Linking from one information resource to another

**reference management tools** See **citation management tools**

**registries** Databases that store specific types of information needed for enabling the exchange of data between systems

**RePEC** Research Papers in Economics; a subject-based repository

**repositories** Open access digital libraries of research-quality resources. They can be subject- or institutionally based, or they can be aggregations of repositories

**ROAR** Registry of Open Access Repositories

**RSS** Really Simple Syndication; supports subscription and automatic, immediate distribution of content so that users receive updates

**scholarly communication** The communication activity of scholars, for example how they communicate as writers, cite each others' work, choose channels for disseminating their work and collaborate with other scholars

**Scopus** A database of citations and abstracts for scholarly journal articles

**self-archiving** Depositing a digital copy of a work on the web (e.g., on a personal or group web page or in a repository) to provide open access to it

**semantic web** A common framework that enables large-scale integration of, and reasoning on, data on the web and allows data to be shared and re-used across boundaries. Also known as 'a web of data' that can be processed by machines

**SEO** Search engine optimization

**serials crisis** Continual, above-inflation increases in the costs of subscriptions to scholarly content, while the budgets of subscribing libraries remain static or decline

**SGML** Standard Generalized Markup Language, a system for creating markup languages and a precursor to XML

**SHERPA RoMEO** A searchable database of publisher's policies regarding self-archiving of journal articles

**SIP** Simple Publishing Interface, a publishing protocol facilitating the exchange and re-use of data in multiple applications and systems

**social web** A phrase used to refer collectively to the websites, tools and services that facilitate interactions, collaboration, content creation and sharing, contribution and participation on the web. Examples include e-mail and discussion forums, bookmarking, wikis and blogs, microblogs, media sharing services and social networks

**socio-technical systems** In digital libraries, systems that support not only information seeking and discovery but also community interaction and collaboration. Socio-technical system design is based on the interplay between technologies, information, and people

**SRU/SRW** Protocols for internet search queries maintained by the Library of Congress. SRU (an OASIS standard) is the Search/Retrieve URL Service and SRW is the Search/Retrieve Web Service. SRU/SRW draw on the model for Z39.50 but are not as complex

**SSRN** Social Science Research Network; a subject-based repository

**standards** Established technical norms, requirements, specifications, processes or practices that are ratified, proposed or in draft form, accepted, or recommended by international, regional or national organizations



- subject-based repository** A repository for collecting and providing access to digital objects related to a subject or group of subjects; also called a discipline-based repository
- success factor** A requirement or necessary condition for achieving an organization's or project's mission
- SWORD** Simple Web-Service Offering Repository Deposit, an interoperability standard facilitating deposit in repositories
- syndication** The automatic distribution of web content to multiple other sites
- target audience** A specific group (or groups) of people for whom a product or service is intended
- TEI** Text Encoding Initiative
- TEL** The European Library
- TRAC** Trustworthy Repositories Audit and Certification
- UKOLN** UK Office for Library and Information Networking
- URI** Uniform Resource Identifier
- usability** A quality attribute that assesses how easy user interfaces are to use (Jakob Nielsen)
- user-centered design** A philosophy and process for designing interactive systems in which the needs and practices of end-users receive extensive attention; sometimes called human-centered design
- value proposition** A brief, clear statement of the benefits of using a particular service or product, e.g., what problems it solves for the user
- VIAF** Virtual International Authority File
- VLEs** Virtual learning environments
- VREs** Virtual research environments
- W3C** World Wide Web Consortium, an international standards body that works on the creation and maintenance of web standards. W3C develops specifications (called recommendations) that describe the building blocks of the web. Tim Berners-Lee is the director of W3C
- Web 2.0** A conceptualization of the web as a platform for participation, emphasizing user interaction, sharing of content and social networking
- web referrer** An online site or service that drives visits and visitors to another site
- web services, web APIs** Enabling technologies that support the machine-to-machine exchange and re-use of content between sites
- WoS** Web of Science; a citation index to scholarly literature
- XML** eXtensible Markup Language, a subset of SGML that provides a simple, flexible text format. XML has played an important role in the

publication and exchange of data on the web

**Z39.50** A standard protocol for searching collections on remote systems, gathering and displaying a unified set of results; used in early digital library efforts to achieve interoperability

## CHAPTER 1

.....

# **Emergence and definitions of digital libraries**

### **Overview**

This chapter traces the first decade of progress in digital libraries (1991–2001), with emphasis on the foundational innovations, vision, motivations, new technology, funding and early programs that prompted their emergence and rapid development. It next turns to the question of how to define the concept of ‘digital libraries’ in an environment of multiple perspectives and continuous technological and societal change. The chapter’s intent is to orient the reader to the field as well as to ground the rest of the book in the context of the aspirations and efforts of many diverse communities and individuals.

### **The emergence of digital libraries (1991–2001)**

This book places the beginning of digital libraries in 1991, the year in which the National Science Foundation (NSF) in the US sponsored a series of workshops on how to make digital libraries a reality, not just a dream. At the same time, digital libraries are an outcome of the revolution in computing, telecommunications and information systems that began almost 40 years ago, around 1965. This section frames the emergence of digital libraries as a recognized field of endeavor in terms of four requirements for viability and growth: a compelling vision, strong motivating factors, technology and funding.

### ***A compelling vision***

Many authors (Arms, 2000; Fox, 1993a; Lesk, 2004; Tedd and Large, 2005) trace the vision of digital libraries to a post-World War 2 paper by Vannevar Bush

called *As We May Think* (1945) and a book called *Libraries of the Future* by J. C. R. Licklider (1965). Licklider's research for the book was sponsored by the US Council on Library Resources (Clapp, 1965, ix). Bush, at that time director of the US Office of Scientific Research and Development, called for a new approach to information organization and discovery based on the visionary concept of a 'memex' – a fast, flexible and efficient desktop device enabling associative indexing and instant access to both a vast library and a scientist's personal files.

The ideas and writings of Licklider, a professor of computer science at MIT, vice president of a high-technology company and imminent researcher for the Defense Advanced Research Projects Agency (DARPA), eventually led to ARPANET, a system of networked computers that preceded the internet. At the outset Licklider's *Libraries of the Future* focuses less on technology and more on solving the basic limitations of printed materials and the bricks-and-mortar libraries of the time:

If books are intrinsically less than satisfactory for the storage, organization, retrieval, and display of information, then libraries of books are bound to be less than satisfactory also. We may seek out inefficiencies in the organization of libraries, but the fundamental problem is not to be solved solely by improving library organization at the system level. Indeed, if human interaction with the body of knowledge is conceived of as a dynamic process involving repeated examinations and intercomparisons of very many small and scattered parts, then any concept of a library that begins with books on shelves is sure to encounter trouble

Licklider, 1965, 5

Noting that 'the "libraries" of the phrase, "libraries of the future", may not be very much like present-day libraries,' and 'in the present century, we may be technically capable of processing the entire body of knowledge in almost any way we can describe,' Licklider went on to create a prescient list of criteria for the future library that reflects both the progress and aspirations of 21st-century libraries (Licklider, 1965, 1, 20, 36–9).

### Key developments from 1965 to the early 1990s

Licklider laid out his challenging 'libraries of the future' vision in 1965. Over the next 25 years, the technologies needed to build digital libraries became not only available but affordable – for example, digital storage, processors,

connectivity, natural language processing, text formatting and scanning, optical character recognition (OCR), indexing and more (as discussed by Lesk, 2004, 16–89). Perhaps most importantly, the promise of the internet (dedicated in its earliest years to research-oriented use) for public and commercial use had captured the public imagination as well as the interest of the private sector and research professionals (Weingarten, 1993; Stoker, 1994; Ginsparg, 2011).

#### *The computer and information sciences*

Computer and information scientists made enormous progress in information retrieval theory and systems between 1965 and 1990. Computer scientists advanced the knowledge and understanding of architecture and systems, and information scientists complemented their work (Arms, 2012, 581). Howard D. White and Kate McCain's renowned analysis of the structure of the information science discipline between 1972 and 1995 indicates that the discipline was principally focused on information retrieval and user-system relationships; bibliometrics; automated library systems and online catalogs; science communication; and user theory (White and McCain, 1998). All of these created a solid foundation for the emergence of new research on digital libraries.

#### *Online information industry*

The online information industry predates the internet and the web. It had its start in the 1970s and by the early 1990s, it was a US\$12 billion industry (1992 dollars), serving mainly the business sector (Calhoun, 1994, 2). There was, however, a segment of the industry called 'scientific, technical and diversified online services' that served primarily research and education; the market leaders in the early 1990s were Mead Data Central (NEXIS/MEDIS), Dialog and InfoPro Technologies (BRS/ORBIT) (Calhoun, 1994, 4–5). Dialog dates to 1972 (O'Leary, 1993).

In the early 1990s the online information industry took the form of online host services that mounted databases and software from which subscribers could retrieve information using first, dedicated terminals and later, personal computers. The firms that offered these services relied on content providers (database producers, publishers, abstracting and indexing services) and reliable, commercially available telecommunications networks (providing dial-up services). The supply of online content was already relatively large

by the early 1990s; online databases grew from around 300 in 1979 to nearly 5200 in 1993 (Calhoun, 1994). CD-ROM database vendors had also entered the market for digital information by that time.

The growing adoption of personal computers not just by businesses and other organizations but also in homes, together with the advent of internet access (which was faster and cheaper than the existing commercial networks) led to both amazing opportunities and large challenges for the most successful online services and content providers of the time. The internet has long roots, and many had been aware of its potential for years. For example, in a 20-year retrospective piece he wrote in 2011, the well known physicist Paul Ginsparg notes that he first used e-mail on the original ARPANET, which preceded the internet, while a freshman at Harvard in 1973.

Online information industry services and content providers (e.g., publishers and professional societies) were faced with managing the disruptive risks and opportunities of the 'information superhighway' and full-text digital content in order to maintain (or improve) their positions – or else risk extinction. This same set of new conditions encouraged the entry of many new players providing online information and services.

#### *Libraries, standards and automation*

Libraries were early adopters of online information systems, and highly trained reference librarians served as intermediaries conducting searches of the very expensive online services, which had expert, non-intuitive interfaces *not* designed for end-user searching. In addition, for library information technology and technical services operations, the first distribution of MARC (Machine-Readable Cataloging) records from the Library of Congress in 1968 (Avram, 1969) was a great leap forward. Over the ensuing years MARC had a transformative influence on libraries, as did the founding in 1967 of the first shared computerized cataloging system based on MARC, the Ohio College Library Center (now OCLC Online Computer Library Center; Kilgour, 1969).

The MARC record and these new systems quickly created a new plane for library technological advances. MARC made it possible to aggregate large structured datasets to underpin the conversion from printed to online catalogs of library holdings; the first generation of robust automated systems for libraries; and many new services in libraries (for example, interlibrary lending became much easier, faster and less costly). All of these developments together put libraries in a position to be early adopters of many new information technologies, the internet and the web. Thanks to this long

foreground, libraries were also ready for digital library collections, systems and services (Calhoun, 2003, 282).

#### *The Follett report*

In the UK, a great deal of experience and knowledge of the latest information technologies and networks, predating the internet and the web, led up to the Follett report (1993). The UKOLN (UK Office for Library and Information Networking) had been established in 1990 (Stoker, 1994, 119). Just two examples that reflect the current topics of the 1980s are from Brindley, who was writing about strategy for the 'electronic campus' and the shift from print and CD-ROM to online dissemination of scholarly content (Brindley, 1988; 1989); and from Law, an expert on library automation since the 1970s, who among other topics was writing about projects to get the nation's academic library catalogs online (Law, 1988). The logical extension of all this work was the Follett report, which placed academic libraries high on the UK national agenda for higher education and quickly generated large-scale national funding for the development of 'electronic' or 'virtual' libraries (the eLib Programme, discussed later in this chapter).

#### *Archives and other professional communities*

A foundational development that came out of the archives, humanities computing, linguistics and other professional communities was the Text Encoding Initiative (TEI), which produced a standard for encoding scholarly texts in machine-readable form. TEI, intended to support data interchange in humanities research, can be traced to a conference of the Association for Computers and the Humanities in 1987. The then newly available Standard Generalized Markup Language (SGML) was the spark needed to kick off the development of TEI and a new way of supporting textual research on the network (Ide and Sperberg-McQueen, 1995).

Daniel Pitti (1997) describes how the advent of the internet inspired the archival community to renew its efforts to bring geographically distributed primary resources together in a way that would enable universal intellectual access. Foundational (pre-internet) work was accomplished from 1981 to 1984 when a US National Information Systems Task Force of the Society for American Archivists paved the way to a MARC standard for the encoding of records describing archives and manuscripts. MARC records provide for the online discovery of archives and manuscript collections at the collection level,

and in a library context; but machine-encoded finding aids were needed to actually lead to the materials in the collection. Archivists' next step was to develop a standard, computer-based encoding structure for finding aids. This work began in 1993 and produced the Encoded Archival Description or EAD standard. SGML is the technology underlying EADs. The development of EAD and experience with SGML were momentous developments that aligned the archival community's work with the web and the networked digital environment that was emerging.

### *Other developments*

Given the limits of this space and my time to conduct the necessary research, this section's mini-analysis of the conditions leading to digital libraries from 1965 to the early 1990s is far from complete. I have merely touched on the work of some disciplines, organizations and communities of practice and not discussed others' contributions at all. In addition to the roles played by early research on the internet (which goes back to the 1970s) and by computer and information scientists, the online information industry, archives and libraries, the efforts of countless researchers and implementers intersected with, ran parallel with, or contributed directly to the origins of digital libraries. These include the individuals and groups who developed the internet and web standards, open systems and other core aspects of networking; those who pioneered new ways of marking up and encoding text; the geospatial or informatics communities; teaching and learning communities; and more. While recognizing these many contributions, I have focused this and the next chapter chiefly on the roles of computer and information scientists; libraries and the cultural heritage sector; and scholarly communities, content providers and online services.

### **An ambitious agenda**

Christine Borgman (2007, 21), writing of the political aspects of new, large-scale research programs, noted 'visions must be grand to attract attention and the promised outcomes must be ambitious to attract money.' By the start of the last decade of the 20th century, computer and information scientists, scholarly content providers and libraries were ready to embrace an ambitious agenda. They were ready for the next steps toward the systems that Bush and Licklider had envisioned in 1965. Building the first digital libraries was not just feasible: it was the logical next step for researchers and professionals in



many fields. Elements of the vision of digital libraries that fueled scholarly and public interest in the first decade of digital library research and development, starting around 1991 included:

- easy, fast, and convenient access to the world's information (regardless of where that information is stored) at any time, from anywhere in the world
- effective storage and organization of massive amounts of text, multimedia and data beyond the bounds of what even the largest single library could provide
- organization and access to materials in many languages
- greatly improved searching and browsing capabilities
- interoperability enabling the cross-searching of many diverse collections at once
- direct, instant delivery of information and data to multiple users at the same time
- transformative improvements in support for research and education globally; better support for interdisciplinary work and scholarly collaboration across institutions and around the world
- significant cost savings over traditional (duplicative) methods for cataloging, storing and preserving analog materials.

### ***Strong motivating factors***

As if the grand opportunities were not enough, two more powerful motivating factors converged in the early 1990s to make the time right for digital libraries. One was a sense of urgency to solve the pressing issue of an explosion of scholarly information; the other, already mentioned, was a sense of opportunity that arose in firms and communities of practice supporting scholarship. First, publishers, professional societies and indexing services seized on technological advances to improve the information storage and retrieval systems they used. Second, libraries and cultural organizations saw an opportunity to preserve and extend access to valuable collections through digitization.

### **A sense of urgency**

#### *Runaway growth*

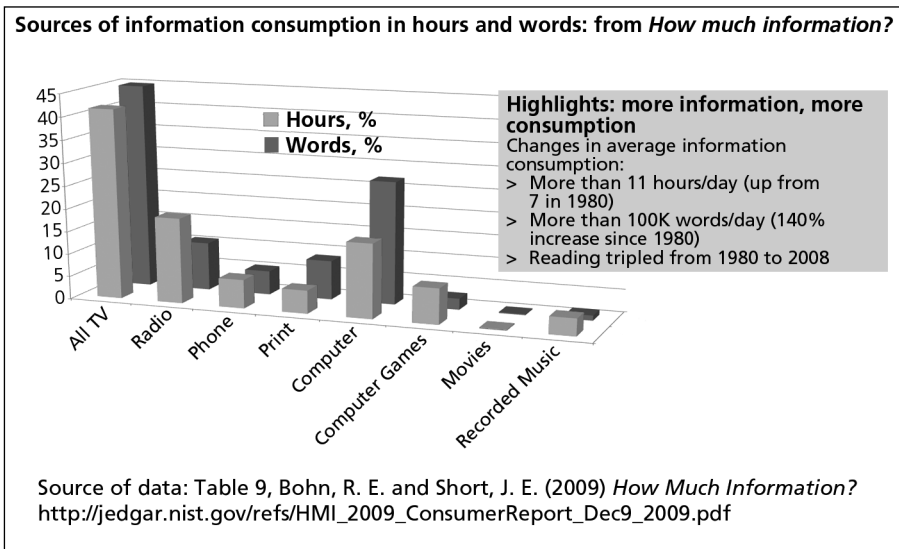
Both scholars and librarians have considered digital libraries to solve large-scale, long-standing challenges. Chief among them is the need to make an increasingly overwhelming volume of material accessible and available. Writing at the

conclusion of World War 2, Bush (1945) noted the ‘growing mountain of research’ and that the difficulty posed by an explosion of scientific publications ‘extended far beyond our present ability to make real use of the record.’ In the UK the sense of urgency was similar, in that it was centered on the perception of runaway growth, but of a different nature. UK national attention was focused on specific problems facing UK higher education and academic libraries – increasing costs for materials coupled with a huge expansion in student populations – and the opportunity to solve them by effectively harnessing the technologies of the global information revolution (Carr, 2002).

The notions of runaway growth were fueled by other early predictions as well. Although his methodology was later called into question (Molyneux, 1994), librarian Fremont Rider’s conclusion in *The Scholar and the Future of the Research Library* (1944) – that research libraries would double in size every 16 years – firmly established a sense of urgency around solving the problem of runaway library growth.

*How much information?*

Rider’s methods may have been flawed, but he was not wrong about runaway growth in the world’s information, including scholarly information. Figure 1.1 pulls estimates from a report of how much information was



**Figure 1.1** Validation of the predictions of runaway growth: more information and more consumption

consumed by Americans, from what sources, in 2008 compared to 1980 (Bohn and Short, 2009). The bars in Figure 1.1 compare estimated 2008 information consumption in hours and words.

As it turns out, there was and continues to be an information explosion, although not perhaps in the ways that Rider and digital library pioneers anticipated. Much more information does exist, and people spend more of their time consuming more of it. Bohn and Short estimate that the number of hours of information consumption per person grew 2.6% a year from 1980 to 2008 (2009, 7). The size of research library collections has not doubled every 16 years, but the amount of information available and of interest to an academic community has exceeded that growth rate; there is more to read than ever, and reading has increased since 1980: this is because there are now so many more ways to consume words (text). The report estimates that a third of the words that people consume come through interactions with computers, and the overwhelmingly preferred way to receive words is via the internet.

## A new world

### *Scholarly communication*

‘Scholarly communication’ alludes to the communicative activity of scholars (people engaged in creating original scholarly works), in particular how they communicate as writers, linkers (e.g., citing others’ work), submitters/disseminators (the choice of formal and informal communication channels, e.g., journals, conferences, wikis, blogs), and collaborators (Borgman and Furner, 2002, 6). Quite a few groups, classes of individuals and tools contribute to the *process* of scholarly communication, but stand outside scholarly communication itself: a few of these are peer reviewers, tenure review boards, evaluation tools or systems (e.g., citation counts), editors, publishers, professional societies, online information services, libraries, and of course readers/information seekers and those who annotate or comment on scholarly work informally.

Innovation in the process of scholarly communication was already well under way by the early 1990s, and the early achievements of online information services, publishers, professional societies and indexing services are impressive. At the time when digital library work got under way, major scholarly societies and publishers saw new opportunities and were keenly interested in developing better systems for publishing full-text journals and articles.

*Mercury and CORE*

Two early projects, Mercury and CORE, influenced the rapid development of new kinds of networked, online retrieval systems that made papers from many independent sources appear as one integrated service. Mercury began in 1991; it was a pre-web solution for bringing together computer science articles from three different scholarly publishers. It validated new concepts for converting, storing and delivering page images from distributed sources over the Carnegie Mellon University's campus network (Arms, 2012, 581). CORE was an early project and key influencer of methods for scanning document collections. It ran from 1991 to 1995, digitizing about 400,000 pages from 20 chemistry journals and demonstrating successful ways to build a full-text index for retrieval and display of digitized documents (Arms, 2012, 588). Chapter 2 discusses other early projects (TULIP, Red Sage and others).

The scholarly content providers and services that supported these projects brought a sense of opportunity and substantial resources to early digital library research and development, and these projects had a powerful impact. Indeed, together with many subsequent investments by scholarly societies, publishers and indexing and online service providers, the early projects eventually transformed scholarly publishing and the expectations of faculty and researchers that scholarly content will be not only online, but also interlinked.

*Digitization*

Librarians and other professional communities also drove the development of early digital libraries and the technologies underpinning them. Early efforts to preserve the treasures held in library collections, archives and museums through digital reformatting (digitization) took off in the early 1990s; these are also discussed in Chapter 2. By 1995, there were baseline standards, working principles, and a small but growing community of digitization specialists available for digital imaging projects for texts, pictorial images and more. This field of specialization eventually grew beyond the library and cultural heritage community and spawned mass digitization projects and the public's growing demand for books in digital form (e-books).

**Technology****The last barriers removed**

As mentioned already, personal computers, the internet and the web were also catalysts enabling research and development of digital libraries; these

technologies were firmly in place before the web's first iteration in 1989 (attributed to Tim Berners-Lee, then at CERN) and the enthusiastic take-up of the Mosaic browser starting in 1993 (the US National Center for Supercomputer Applications at the University of Illinois had developed Mosaic) (Ginsparg, 2011, 5). These new innovations scaled up the size of prior opportunities to build services for collections stored in digital forms, retrievable over networks.

Arms (2000, 10) reported a series of technical developments in the early 1990s that 'removed the last fundamental barriers to building digital libraries':

- Storing information on computers became significantly less costly.
- Major advances had been made in the quality of personal computer displays.
- Receiving information over the internet became fast, affordable and reliable.
- Portable personal computers had become affordable and powerful.

### The Kahn Wilensky architecture

The general principles for the design of a digital library that is 'open in its architecture and which supports a large and extensible class of distributed digital information services' were laid out by Kahn and Wilensky (Kahn and Wilensky, 1995; Arms, 1995). Kahn and Wilensky strongly influenced how early digital libraries were built by technologists. Micah Altman (2008) characterizes the 'Kahn-Wilensky architecture' as having four main types of components:

- *repositories*, ranging from file systems to distributed storage systems for content
- mechanisms to support *search* (indexing or metadata)
- *identifier systems* for identifying and locating digital objects
- *user interfaces* to perform user services (for example searching, browsing, visualization, delivery).

This is not an exhaustive list. Other components of digital library architecture include, for example, security systems for authenticating users, services to aggregate search results from multiple sources, and tools for supporting collaboration and other types of interaction.

### Interoperable, web-based digital libraries

By 1991, computer scientists already had extensive experience with the development of information retrieval (IR) systems. Fox and Sornil (2003) wrote: 'DLs can be regarded as extended IR systems with multiple media and federation.' As web search and retrieval tools improved and gained acceptance over the course of the decade, digital library researchers and professionals sought new approaches to integrate their methods with web technology and the network. Lorcan Dempsey (1994), then at UKOLN, offered a prescient analysis of how the internet and web would generate entirely new kinds of systems and move information creation, publication and discovery to the network. This is indeed what has happened. He also foresaw the immense challenges that libraries would face aligning and integrating their traditional knowledge organization practices, metadata silos and fragmented information systems with the new networked environment – things he has continued to write and speak about today.

Computer scientists and librarians began to respond to the challenges. For example, early in the new millennium, Cornell and the University of Virginia began work on Fedora (Flexible Extensible Digital Object and Repository Architecture), a new system for digital library architecture. The intent was to provide a new framework for interoperable, web-based digital libraries (Payette and Staples, 2002). As will be discussed in Chapter 3, interoperability (the provision of uniform access to diverse information stored on different computer systems in different locations) has proved to be an ongoing challenge facing digital library developers.

### Hybrid libraries

Interoperability became increasingly important as digital library projects, publishers, professional societies, indexing and online services brought content online and demand grew for unified access to content locked up in separate systems with separate interfaces. Furthermore, libraries began looking for ways to integrate the digital content with their predominantly non-digital collections (printed books and journals, prints, slides, maps, analog sound recordings and films, government documents, etc.) Rusbridge (1998) usefully described library collections in four categories: *legacy* (non-digital), *transitional* (legacy resources that have been or will be digitized), *new* digital resources (those expressly created as digital) and *future* digital resources. This book refers to the third and fourth categories as *born digital* resources. Rusbridge called for the development of technologies, systems and

services for the 'hybrid library,' which would integrate all four categories of resources. As discussed in Chapter 5, from early days to the present, the necessity of accommodating the requirements of hybrid libraries has been a key driver in the field of digital libraries and the profession of librarianship.

## **Funding**

This section provides a high-level summary of key national and international funding sources and programs in the first decade of digital libraries. A subsequent section, which contains a review of early large-scale digital library programs, also contains information about funding. The next chapter also incorporates information about funding from foundations, membership organizations, individuals, commercial or non-profit entities, universities and national libraries.

### **Funding streams**

Federal and international agencies, national libraries, higher education institutions, public and private sector organizations, even individuals – all provided streams of funding for the early development of digital libraries. First-decade digital library funding tended to gravitate to national or local institutional levels, or it was invested as a result of the strategic capital budgeting decisions of commercial firms. The variety of streams has resulted in many technical advances, diverse digital libraries, and a complex landscape.

### **Large-scale efforts**

Large-scale efforts tended to be funded by international bodies, government agencies, foundations and non-profit organizations. Some libraries invested heavily in digital library programs in keeping with their missions to support historical and cultural studies, provide a national research information infrastructure and preserve their nations' creative output (examples from Australia, France, the Netherlands, New Zealand, the UK, the US and elsewhere are represented in the next chapter). As noted previously, the investments of scholarly societies, publishers and indexing and online services also considerably advanced the early efforts to put scholarly content online; the amounts invested are unknown but collectively it must have been substantial.

## Universities and institutions

It should also be mentioned that the funding from many universities and institutions supporting individual library projects, when taken as a whole, probably surpasses the financing provided by the centrally organized programs. Daniel Greenstein and Suzanne Thorin's report (2002) of a survey conducted by the Digital Library Foundation indicated that in 2000 responding libraries spent an average of over US\$1 million each on digital conversion and digital library personnel (see their Table 3.1). University library projects at that time focused predominantly on digitization of cultural heritage materials.

## Other sources

In a few cases the vision, commitment and financial resources of single individuals produced lasting and influential digital libraries. Brewster Kahle, for example, founded the Internet Archive in 1995, providing the capital himself. In 2003 one journalist reported that 'the ten million-dollar annual budget [of the Internet Archive] continues to come primarily out of Kahle's pocket' (Womack, 2003). Chapter 2 continues the discussion of how a variety of types of organizations supported the emergence of digital libraries as a new field of endeavor.

## The one universal digital library

National agendas have contributed to the sense of urgency that spurred the eventual development of digital libraries in many countries. The dream of one universal, global digital library has been relevant everywhere to some degree, and it still is (see Arms, 2005, for a discussion of the user's viewpoint). However, while digital libraries are relevant globally, with some notable exceptions they have been funded at the national or local level. Writing for the 2007 issue of the *Annual Review of Information Science and Technology*, David Bearman (2007, 223–4) stated that 'although the vision of a singular "Digital Library" was what captured the popular and political imagination, and was promoted especially by Vice President Al Gore in the 1992 election campaign, through the 1990s the United States government supported "digital libraries" in the plural.' Bearman's perspective is supported by a review of the *Source Book on Digital Libraries* (Fox, 1993b), which reports on a series of NSF invitational workshops that preceded the NSF's call for the Digital Library Initiative (DLI-1) proposals. That work in the foreground of funded projects had two long-lasting outcomes: a preference



for the term 'digital library' over 'electronic library' and a shift from the goal to 'develop a prototype national digital library' (singular) toward funding opportunities for the development of digital libraries (plural). The last chapter of this book returns to a consideration of the dilemma created for digital library implementers as a result of the disunion between who funds digital libraries and who benefits from them.

### **Early digital library projects**

UK, US and multinational programs had considerable influence on digital library development and they produced significant outcomes that defined the way forward as digital libraries continued to evolve. The key projects included:

#### ***UK eLib Programme (eLib)***

The driving force for the commissioning of eLib was the Follett report (1993), which reviewed the system of UK academic libraries in light of the problems of huge expansion of undergraduate populations, rising costs for library materials and the opportunities of new forms of information storage, access and retrieval over networks. Recommending that the problems be addressed through the use of information technology, the Follett report was highly influential and released the funding for eLib (Dempsey, 2006b). Managed by the Joint Information Systems Committee (JISC), eLib ran for seven years (1995–2001) and involved 70 projects. For more information see the first feature article in the first issue of *Ariadne*, which itself grew out of eLib (Kirriemuir, 1996). Pinfield (2004) offers a detailed review of eLib's influential outcomes.

#### ***DLI-1***

The first large-scale funding for digital libraries in the US began in 1994 with an initial four-year Digital Library Initiative (DLI-1) sponsored by NSF, the National Aeronautical and Space Agency (NASA) and DARPA (Defense Advanced Research Projects Agency) (Arms, 2000, 62–3; National Science Foundation, 1993). The projects emphasized mainly technical aspects of digital libraries (Mischo, 2004, 6) and were led for the most part by computer scientists. Behavioral, social and economic issues got little attention during the first round of NSF funding.

**DLI-2**

In 1998 NSF issued a second call for proposals (National Science Foundation, 1998a; Griffin, 1999; Mischo, 2004). DLI-2 began with more concern for the social, behavioral and economic aspects of digital libraries and attracted funding from multiple agencies, including national libraries and the Institute of Museum and Library Services (IMLS).

**Other US national programs**

Arms (2012) reports on American Memory, a digital library that started in 1995 as a result of the Librarian of Congress' establishment of a project to digitize five million items and make them available on the web within five years (see Table 2.1 in this book). Arms, Blanchi and Overly (1997) discussed the technical building blocks, which came from the National Digital Library Project (NDLP) at the Library of Congress. The US National Institutes of Health (NIH) engaged early with digital library efforts. In February 2000 they launched the digital library PubMed Central, which as of this writing contains 2.7 million articles. The US PubMed Central was developed and is managed by the US National Center for Biotechnology Information (NCBI) (Humphreys, 2000).

**Joint NSF/JISC international projects**

In 1998 NSF called for proposals for multi-country, multi-team projects. In the UK, JISC issued a matching call. Six projects were funded jointly by NSF and JISC to explore cross-domain resource discovery, digital archiving, search and retrieval for musical information, reference linking, subject gateways, and metadata for multimedia digital objects (Chowdhury and Chowdhury, 2003, 56–7).

**European Commission (EC)**

Even before the first decade of digital library research and practice, the European Commission devoted substantial attention and funding to library-related programs. As Dempsey (2006b) notes, 'the first EU call for proposals in the libraries area was as far back as July 1991. The motivating framework for this and later calls was established in the Libraries Action Plan, a document first circulated in 1988.' Digital library programs were funded under the European Union's Framework Programmes, beginning with the Third. Funding for digital library research has continued at generous levels

(<http://cordis.europa.eu>; see also Collier, Ramsden and Zhao, 1995; Dempsey, 1995; Dempsey, 2006b).

### **Projects in China and India**

A considerable body of digital library research and development has occurred in China and India. Zhou (2005) and Shen et al. (2008) describe a number of large-scale digital library projects in China, starting with the introduction of CALIS (Chinese Academic Library Information System) in 1998, followed by CADLIS (Chinese Academic Digital Library), completed in 2005. Kumar (2010) and Das, Sen and Dutta (2010) describe digital library research and development in India, which began early in the new millennium and now includes open repositories, a number of cultural heritage digital libraries and the Digital Library of India.

### **Other projects**

A number of large-scale, ambitious projects were inspired by democratic ideals and attracted multiple sources of funding and voluntary support:

#### **Project Gutenberg**

Project Gutenberg ([www.gutenberg.org](http://www.gutenberg.org)) is the first and oldest digital library. It began in 1971 as an idea from Michael Hart, who, given free computer time at the University of Illinois, decided to type in the US *Declaration of Independence* and then tried (unsuccessfully) to send it to everyone on the campus network (Hart, 1992). Gutenberg's goal has been to provide public domain e-texts a short time after they enter the public domain, for free, using only volunteers and donations to get the work done.

#### **Internet Archive**

Brewster Kahle started the Internet Archive in 1995. The Internet Archive (IA) has numerous components, but the Wayback Machine, which provides access to archived versions of an estimated 220+ million websites, may be the best known. The IA is an advocate for universal and free access to knowledge and it founded a co-operative project called the Open Content Alliance to build and preserve a massive digital library of multilingual digitized text and multimedia content (Dougan, 2010).

## The Million Books project

The Million Books project ([www.ulib.org](http://www.ulib.org); the first project of the Universal Digital Library) began with some preliminary test projects that led to an initial grant from NSF in 2000 (Linke, 2003; St Clair, 2008). Raj Reddy, an award-winning computer science professor at Carnegie Mellon University, continues to inspire and direct it. The Universal Digital Library's mission is to foster creativity and free access to all human knowledge; its purpose is to make digital texts freely available to anyone who can read and has access to the network ([www.ulib.org/ULIBAboutUs.htm](http://www.ulib.org/ULIBAboutUs.htm)). Partners came from China, Egypt, India and the US. It reached and exceeded its goal of a million scanned books in 2007. Collections are represented on mirror sites in China and India.

## Definitions of digital libraries

### *The definition used in this book*

The definition of 'digital libraries' that underpins this book has two parts. Digital libraries are:

- 1 A field of research and practice with participants from many disciplines and professions, chiefly the computer, information and library sciences; publishing; the cultural heritage sector; and education.
- 2 Systems and services, often openly available, that (a) support the advancement of knowledge and culture; (b) contain managed collections of digital content (objects or links to objects, annotations and metadata) intended to serve the needs of defined communities; (c) often use an architecture that first emerged in the computer and information science/library domain and that typically features a repository, mechanisms supporting search and other services, resource identifiers, and user interfaces (human and machine).

My intention is to provide a practical definition that reflects the current situation, but can evolve as digital libraries evolve in the context of the web. Lagoze (2010, 25–31) has persuasively discussed the trend of digital libraries toward the resource-centered architecture of the web (mentioned again in subsequent chapters of this book). The definition used in this book refers for the most part to the traditional repository-centered architecture, because this model remains characteristic of most digital libraries today. Through the chapters of this book, I attempt to make the case that the important characteristics of digital libraries are (in this order) the social roles they play;

the communities they serve; the collections they gather for those communities; and the enabling technologies that support them. Social roles and communities are more likely to abide over time; collections and enabling technologies are more likely to shift.

### ***Other definitions of digital libraries***

#### Different perspectives

At the start of digital libraries' first decade, what came to be called a digital library had a number of names – electronic library, virtual library, library without walls. The first decade's explosion of activity and funding for digital library research and practice engendered many diffuse definitions of the phrases *digital library* or *digital libraries*. Some of the principal authors during this first decade paid little heed to definitions; others' discussions of definitions are lengthy. Considered as a whole, the digital library literature contains an enormous amount about how to define digital libraries. Fox et al. (1995, 24) suggest an explanation: 'the phrase "digital library" evokes a different impression in each reader.'

The public on the one hand, and those involved in building digital libraries, on the other, naturally had a variety of perspectives on the nature of digital libraries, when they were first conceived. The following list represents a few of these initial perspectives:

- a computerization of traditional libraries (people in general)
- a framework for carrying out the functions of libraries in a new way with new types of information resources (librarians)
- a new set of methods to innovate and improve fee- or membership-based indexing, full-text repositories and hyperlinking systems (publishers, online information services, professional societies, indexing services)
- a distributed text-based information system (computer and information scientists)
- a collection of distributed information services (computer and information scientists)
- a distributed space of interlinked information (computer and information scientists)
- a networked multimedia information system (computer and information scientists)
- a space in which people can collaborate to share and produce new knowledge (those working on collaboration technologies)
- support for formal and informal teaching and learning (educators).

Arguably the most comprehensive and thoughtful discussion of first-decade digital library definitions is by Borgman (1999 and 2000, 35–52), who notes that the many definitions arise because ‘research and practice in digital libraries are being conducted concurrently’ and by individuals and teams from different fields. Borgman made sense of the definitions that had emerged by 2000 by grouping and discussing them in a variety of ways, including:

- orientation (research-oriented versus practice-oriented definitions)
- concept of a library (narrow – library as a collection of content supporting information retrieval – versus broad – library as a continuous and trusted social entity)
- emphasis (definitions emphasizing collections, a particular type of content or communities versus those with an emphasis on institutions or services).

### A sample of definitions

Table 1.1 builds out from the core of definitions considered by Borgman in 1999 and 2000. It offers a sample of definitions, considers their principal facets and cites their sources. The sample is far from comprehensive but attempts to show the progression of definitions from those emphasizing the enabling technologies of digital libraries (text analysis, distributed retrieval systems, metadata, indexing and knowledge representation, data communication networks, intelligent agents, interface design, multimedia storage, etc.) toward a new generation of definitions that place more emphasis on the communities and social roles of digital libraries. A number of authors have made the point that early research engendered definitions that focused more on technical issues and less on the broader social context of digital libraries (for example, Lagoze, 2010, 6).

### Discussion

The DELOS definition offers a framework for understanding, planning and evaluating digital libraries. Another model is the ‘5S’ framework (Streams, Structures, Spaces, Scenarios, and Societies) introduced in the dissertation of Marcos André Gonçalves (2004), which has been used to inform the development of a curriculum for digital library education and for other purposes. Another influential definitional model – one that pushes beyond read-only digital library repositories – is the one proposed by Lagoze and

others (2005). This paper introduced a more flexible, richer information model for digital libraries based on an ‘information network overlay’ for modeling resources, their descriptions and relationships. It represented breakthrough thinking that led to new possibilities for digital libraries that facilitate ‘the creation of collaborative and contextual knowledge environments.’

<b>Table 1.1</b> A progression of digital library definitions		
<b>Definition</b>	<b>Facets</b>	<b>Source and comments</b>
‘The library of the future will be based on electronic data ... contain both text and graphics and be widely available via electronic networks. It is likely to be decentralized ...’	<ul style="list-style-type: none"> <li>• Digital data (collections)</li> <li>• Multimedia</li> <li>• Services (widely accessible)</li> <li>• Networked</li> <li>• Distributed</li> <li>• Enabling technologies</li> </ul>	Lesk, Fox and McGill, 1993, 12, 19–24  This was a white paper for NSF created in 1991. It led to the series of NSF workshops and the first NSF call for proposals.  The focus of the definition and white paper was on enabling technologies and maintaining US national competitiveness.
‘A service; an architecture ... a set of information resources, databases of text, numbers, graphics, sound, video, etc.; a set of tools and capabilities ... [with] users ... [and] contributors ...’  Another key assumption: For use on the network	<ul style="list-style-type: none"> <li>• Services (networked; with tools/capabilities)</li> <li>• Architecture (enabling technologies)</li> <li>• Digital data (collections)</li> <li>• Multimedia</li> <li>• Community-based (users/contributors)</li> </ul>	Borgman, 1993, 122
‘Systems providing a community of users with coherent access to a large, organized repository of information and knowledge ... enriched by the capabilities of digital technology ... span[ning] both print and digital materials ... provid[ing] a coherent view of a very large collection of information ... integrat[ing] materials in digital formats ... such as multimedia, geospatial data, or numerical datasets ... [characterized by] continuity [with] traditional library roles and missions ... [and with] many digital repositories ... appear[ing] to be a single digital library system ...’	<ul style="list-style-type: none"> <li>• Systems</li> <li>• Community-based</li> <li>• Services (coherence; collected and organized)</li> <li>• Enabling technologies</li> <li>• Distributed, interoperable</li> <li>• Digital and non-digital data (hybrid)</li> <li>• Multimedia</li> <li>• Extension of existing libraries</li> </ul>	Lynch and Garcia-Molina, 1995
‘A large collection of the full contents of high use materials including books, journals, course materials, and multimedia learning packages, which can be directly accessed by students and staff’ [with personal computers]	<ul style="list-style-type: none"> <li>• Multimedia</li> <li>• Terms and conditions (licensed content)</li> <li>• Collection</li> <li>• Digital data (digitized)</li> </ul>	Zhao and Ramsden, 1995  ELINOR project; concerned with digital library development for teaching and learning. Led to insights and progress on copyright and publisher content licensing issues (see Collier, Ramsden and Zhao, 1995).  <i>(Continued on next page.)</i>

<b>Definition</b>	<b>Facets</b>	<b>Source and comments</b>
'Organized collections of digital information. They combine the structuring and gathering of information, which libraries have always done, with the digital representation of information that computers have made possible.'	<ul style="list-style-type: none"> <li>• Services (organized, structured and gathered)</li> <li>• Digital data (collections)</li> <li>• Extension of existing libraries</li> <li>• Computers (enabling technologies)</li> </ul>	<p>Lesk, 1997, xx, xxii</p> <p>Lesk also stressed the importance of the economics of digital libraries: 'We know how to build a digital library ... we do not know how to make it economically supportable.'</p>
<p>'The definition of the digital library will require an understanding of the role and nature of public institutions in a postindustrial society.'</p> <p>'A realm of free speech and association as well as an information market place.'</p>	<ul style="list-style-type: none"> <li>• Extensions of existing libraries (but not as collections; rather in their societal roles)</li> <li>• Social (emphasis on social aspects)</li> </ul>	<p>Lyman, 1996</p> <p>Emphasizes the social role of libraries offering free and equal access to knowledge and ponders the question of how digital libraries might support the traditional role of the library as a 'marketplace of ideas' and the public interest in education and democratic participation.</p>
'Organizations [i.e., institutions] 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.'	<ul style="list-style-type: none"> <li>• Organizations (institutions)</li> <li>• Digital data (collections)</li> <li>• Community-based</li> <li>• Services (selecting, collecting, organizing, providing access, delivering, preserving)</li> </ul>	<p>Waters, 1998</p> <p>The definition developed by the Digital Library Federation.</p> <p>Services encompass a curatorial role.</p> <p>See also Deegan and Tanner (2002, 22)</p>
<p>1 Digital libraries are a set of electronic resources and associated technical capabilities for creating, searching, and using information.</p> <p>2 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.</p>	<ul style="list-style-type: none"> <li>• Digital data (collections)</li> <li>• Enabling technologies</li> <li>• Services (collecting, organizing, searching, using information)</li> <li>• Community-based</li> <li>• Use- and user-centered</li> <li>• Emphasis on social aspects (life cycle of information)</li> </ul>	<p>Shortened version of Borgman, 2000, 42.</p> <p>This definition has been very influential in the digital library field.</p> <p>From the beginning, Borgman has stressed the importance of the social aspects of digital libraries.</p>
'Sociotechnical systems – networks of technology, information, documents, people, and practices.'	<ul style="list-style-type: none"> <li>• Systems</li> <li>• Networked</li> <li>• Community-based</li> <li>• Use- and user-centered (work practices and people)</li> <li>• Emphasis on social aspects</li> <li>• Systems</li> <li>• Enabling technologies</li> <li>• Collections</li> </ul>	<p>Bishop, Van House and Buttenfield, 2003</p> <p>Emphasis on balancing the needs of people with the requirements for collections and enabling technologies.</p>



**Table 1.1** (continued)

Definition	Facets	Source and comments
'A tool at the center of intellectual activity having no logical, conceptual, physical, temporal, or personal borders or barriers to information. Generally accepted conceptions have shifted from a content-centric system that merely supports the organization and provision of access to particular collections of data and information, to a person-centric system that delivers innovative, evolving, and personalized services to users. Conceptions of the role of Digital Libraries have shifted from static storage and retrieval of information to facilitation of communication, collaboration, and other forms of dynamic interaction . . . [and] the capabilities of Digital Libraries have evolved from handling mostly centrally located text to synthesizing distributed multimedia document collections, sensor data, mobile information, and pervasive computing services.'	<ul style="list-style-type: none"> <li>• Service (Tool)</li> <li>• Systems</li> <li>• Use- and user-centered</li> <li>• Community-based</li> <li>• Social (communication, collaboration, dynamic interaction)</li> <li>• Multimedia</li> <li>• Mobile</li> <li>• Terms and conditions (policies)</li> </ul>	<p>Candela et al., 2007</p> <p>A conceptual definition from the DELOS Digital Library Manifesto (Candela et al., 2006). Defines six core components of digital libraries: content, users (both humans and machines), functionality, quality, policy (e.g., rights) and architecture. The Manifesto contains a useful discussion of digital library definitions.</p>

Other authors have also contributed insightful commentary on how to define digital libraries, rather than specific or formal definitions or frameworks (two examples are Chowdhury and Chowdhury, 2003, 4–9; Chowdhury and Foo, 2012, 2–4). Bill Arms offers an informal definition ('a managed collection of information, with associated services, where the information is stored in digital formats and accessible over a network,' but at the same time Arms has consistently emphasized that digital libraries must be understood as an 'interplay of people, organizations and technology' (2000, ix, 2). The already cited article by Peter Lyman offers another, quite different perspective; I recommend it to anyone with an interest in libraries' (and digital libraries') roles supporting the public good. The IFLA/UNESCO manifesto on digital libraries ([www.ifla.org/digital-libraries/manifesto](http://www.ifla.org/digital-libraries/manifesto)), which contains a definition of a digital library, also emphasizes the role of digital libraries in bridging the 'digital divide' (discussed in Chapter 6).

Levy and Marshall's article (1995, 78, 80, 82–3) is particularly important because it applies a work-oriented (ethnographic) perspective, noting that the emergence of digital libraries challenges the assumptions but not the basic character of libraries as an interplay of collections, enabling technologies, and services supporting the work that communities of users want to do. Noting

'an infrastructure by itself does not constitute a library' and 'the highest priority of a library, digital or otherwise, is to service the research needs of its constituents,' their article presaged the ensuing shift away from enabling technologies and digital collections as ends in themselves and toward user-centered design and networked services supporting collaborative work.

### ***A few definitional issues***

There are many challenges associated with attempting to define digital libraries. Some of the issues discussed by various authors include the following:

#### **Distributed digital libraries**

Some digital libraries are central archives that provide digital content storage and deliver services from a single system; others' content and services are distributed in multiple locations on the network. Still others aggregate the content of many digital libraries (repositories of repositories). Suleman (2012; 17–21) discusses centralized and distributed digital libraries. It should be noted here that digital libraries that are crawled and indexed by common or academically oriented search engines are discoverable in search engine results as if they were aggregated. The definition of digital libraries in this book covers some of these but excludes the virtual aggregations offered by common search engines like Google or Bing. The academic search engine Google Scholar, however, has characteristics of a digital library (it has a social role and it is intended for scholars' use). Beel, Gipp and Wilde (2010, under section 2.1) further discuss academic search engines, including Google Scholar, PubMed and IEEE Explore.

#### **Hybrid libraries**

As already noted, Rusbridge coined the term 'hybrid library' to refer to combinations of traditional collections, licensed e-resources and openly available digital collections produced in-house or elsewhere. Some digital content is directly accessible; other content can be linked to; still other content is represented only by citations (metadata). Schwartz (2000) writes 'the hybrid library is the context within which most academic digital libraries are found – the ecosystem of the digital library, as it were.' Chowdhury and Chowdhury (2003, 6–7) confirm this view; in their book they use the phrase 'digital

libraries' to denote both digital-only and digital-plus-analog (hybrid) libraries. As Bearman (2007, 223) remarks, an assumption that a digital library contains only digital works is overly limiting; it is necessary to include within the scope of digital libraries those that 'service some physical items in addition to digital content.' The definition of digital libraries in this book includes hybrid libraries provided that the amount of digital content directly available or accessible through links exceeds the content represented by metadata only. Databases of metadata only fall outside the definition.

### 'Library' or 'digital library'?

As library collections are increasingly dominated by online content, the concepts of 'digital libraries' and 'libraries' are less distinguishable than they were in the 1990s, when digital libraries began to emerge. Chapter 5 discusses the possible convergence of strategic agendas for digital libraries and traditional libraries. However, the definition provided in this book does not conflate digital libraries and libraries.

### Preservation

Deegan and Tanner's definition (2002, 22) is a set of principles emphasizing the curatorial role of digital libraries as managed collections, requiring that digital objects be selected, made accessible, and preserved as *long-term, stable resources*. The definition of digital libraries used in this book does not explicitly require a preservation mission.

### Open or restricted content?

As discussed in Chapter 4 of this book, digital library innovations have led to rapid growth in the availability of open, freely available digital content and a culture of open data interchange. The definition used in this book notes that digital libraries are often open. However, the definition does not exclude fee-based or restricted-access digital libraries such as those produced by publishers and other e-resource providers, provided they are intended to serve defined communities. Borgman (2000, 46–7) and Chowdhury and Chowdhury (2003, 8–9) also discuss this issue. The definition in this book includes, for example, open or fee-based digital libraries from scholarly publishers, professional societies, aggregators like JSTOR or the Directory of Open Access Journals. It also includes library or consortially provided, cloud-

based library discovery layers that provide access to a substantial amount of open, licensed and/or fee-based digital content.

### Global digital library

Borgman proposed a working definition of a ‘global digital library’ as ‘a useful construct that encompasses all the digital libraries that are connected to and accessible through a global information infrastructure’ (2000, 48). Such a construct does not exist as of this writing. The world wide web, in and of itself, or its representation in a search engine like Google, falls outside the definition of a digital library that is used in this book.

### Conclusion

This chapter has traced the antecedents of digital libraries to 1965 and J. C. R. Licklider’s challenging vision for ‘libraries of the future,’ which, he noted, ‘may not be very much like present-day libraries.’ Key developments from 1965 to 1990 in computer and information science, telecommunications and networks, online publishing, personal computer ownership, libraries, archives and other professional communities – not to mention the internet and web – prepared the ground for an ambitious digital library research and development agenda. The vision for digital libraries was grand, and it attracted top research and professional talent and generous funding.

Early projects in the US and UK, programs funded by the European Commission, scholarly publishing projects, a number of projects inspired by democratic ideals, and many other initiatives led to groundbreaking innovations and the emergence of a new field of endeavor. Multifaceted and surrounded by dynamic technological and societal conditions, digital libraries are challenging to define, because they evoke diffuse impressions and continually evolve. The chapter concludes with a practical definition that underpins the use of the phrase ‘digital libraries’ in this book.

The next chapter examines the outcomes of digital libraries’ exhilarating first decade: a new field of endeavor; transformative change in the processes of scholarly communication and in how (and where) people look for information; new ways of organizing, interlinking, and aggregating digital content; large-scale digitization; digital preservation; the open access movement; and working digital libraries.

.....

# Index

Note: The following abbreviations have been used: *f* = figure;  
*t* = table

- Academia.edu 225, 226*t*  
academic libraries 8, 9, 15, 36, 130, 243, 250  
    circulation trends 114*f*, 115  
    expenditures of materials budgets 116, 117*f*  
academic search engine optimization (ASEO) 195–6  
'Academic Working Environment' (AWE) 204  
accessibility 77, 86, 93–4, 95*f*, 96–7  
ACDP *see* Australian Cooperative Digitization Project  
ACM *see* Association for Computing Machinery  
ACM Digital Library 54*t*  
Adamick, J. 90, 99, 180  
Adewumi, A.O. 243  
aDORe 102–3, 105  
ADS *see* Astrophysics Data System  
Afzal, W. 201, 202  
AgEcon Search 100*t*  
Alexa 52  
Alicante, University of 247  
Altman, M. 11, 213  
altmetrics 218–19  
Amazon 109, 115  
American Astronomical Society 106  
American Memory 16, 53*t*, 149–50, 162, 251  
Anaraki, L.N. 144  
Anderson, S. 203  
annotation tools 241  
*Annual Review of Information Science and Technology* 14  
Application Programming Interfaces (APIs) 103, 104, 211  
AquaBrowser 129  
ARCHER 200  
architecture *see* Kahn Wilensky architecture  
Archival Sound Recordings project (British Library) 45  
archives 5–6, 34, 48, 118–19  
Archives of American Art (Smithsonian Institute) 165  
*Ariadne* (magazine) 15, 29–30  
ARL *see* Association of Research Libraries  
Arlitsch, K. 93, 195, 196, 244  
Armbruster, C. 97  
Arms, W. Y. 16, 23, 29, 40, 70, 88  
    funding 171, 172  
ArnetMiner 224  
ARPANET 2, 4  
ARROW 134, 200  
arXiv.org 53*t*, 163  
    funding 172, 239  
    open access 34–5, 38  
    repositories 89, 97, 100*t*, 105, 107  
    research identifiers 228, 231*t*, 233  
*As We May Think* (Bush) 1–2  
ASEO *see* academic search engine optimization  
ASIS&T *see* Association for Information Science and Technology  
Association for Computers and the Humanities 5  
Association for Computing Machinery (ACM) 29, 54*t*  
Association for Information Science and Technology (ASIS&T) 49  
Association of Research Libraries (ARL) 32, 170, 201  
    hybrid libraries 114*f*, 116, 117*f*; 122, 130  
associations: community building 28–9  
associative indexing 37, 65  
Astronomical Society, American 106  
astronomy 106, 254  
Astrophysics Data System (ADS) 100*t*  
Athens (authentication/authorization service) 132  
AtomPub 107  
audience-directed digital libraries 55–6*t*  
audiences: defining 170, 196, 198  
audio technology 45  
audiovisual materials 115  
Australia 13, 43, 97, 191, 200, 201  
    national libraries 46, 53*t*, 148, 253  
Australian Cooperative Digitization Project (ACDP) 43  
Australian Research Data Commons 200  
authentication 127, 132, 213, 214  
AuthorClaim 231*t*, 233  
authorization 127, 132  
Authors Guild 122  
automated deposit 194  
automation 4–5  
awareness 163–5, 181  
Awre, C. 108, 109  
  
Bailey Jr, C. 30, 201  
Baker, T. 74  
Bankier, J.-G. 192  
BASE *see* Bielefeld Academic Search Engine  
BBC *see* British Broadcasting Corporation  
Beagrie, N. 149, 219–20  
Bearman, D. 14, 25, 51, 120  
Beel, J. 24, 195, 244  
Bell Labs 33  
Bell, S. 184

- Bergman, M.K. 66  
 Berners-Lee, Tim 11, 72–3  
 Berry, L.L. 163  
 Besek, J. M. 157  
 best practice: digitization 47–8  
 Beunen, A. 174  
 Bhargava, R. 244–5  
 Bibliographic Framework Initiative (BIBFRAME) 133–4  
 Biblioteca Virtual Miguel de Cervantes 247  
 Bibliothèque Nationale de France (BnF) 45–6, 53*t*, 247, 251  
 BibUsages project (France) 251  
 Bielefeld Academic Search Engine (BASE) 92, 102, 128–9  
 Bing 249  
 BioMed Central 107  
 Biotechnology Information, National Center for (NCBI) (US) 16  
 Bizer, C. 72, 74, 75  
 Björk, B.-C. 95–6  
 Black, K. 99  
 Blanchi, C. 16  
 Blixrud, J. 131  
 Block, D. 117  
 blogs 30  
 Blue Ribbon Task Force 155  
 Blyberg, John 210  
 BNB *see* British National Bibliography  
 BnF *see* Bibliothèque Nationale de France  
 Bohn, R.E. 8*t*, 9  
 Bollen, J. 61  
 bookmarking 220  
 Borgman, C. 6, 20, 25, 26, 31, 67–8, 143  
   community partnerships 170  
   cyberinfrastructure 199  
   data curation 201  
   online communities 160, 161  
 born digital resources 12, 111, 156  
 Brabham, D. 253–4  
 branding 127, 163–5  
 Branin, J. 87  
 Brar, N. 170  
 Brase, J. 200  
 Bray, P. 242  
 Brazil 232  
 Breeding, M. 129, 133  
 Brin, S. 36–7, 65, 88, 195  
 Brindley, L.J. 5  
 British Broadcasting Corporation (BBC) 75–6  
 British Library 27, 164, 253  
   Online Newspaper Archive 47  
   researcher identifiers 232  
   Sounds.bl.uk 45, 55*t*  
 British National Bibliography (BNB) 249  
 Brody, T. 94, 95, 96, 97  
 Brown, S. 182, 189, 191  
 Brusilovsky, P. 212  
 Buchanan, S. 169  
 Budd, J.M. 202  
 Burns, C.S. 92, 184–5, 202  
 Bush, V. 1–2, 6, 7–8, 37  
 business planning 174–5  
 business processes: hybrid libraries 123, 124*f*
- CADLIS *see* Chinese Academic Digital Library  
 California Digital Library 38  
 California-San Francisco, University of 33, 121  
 CALIS *see* Chinese Academic Library Information System  
 Candela, L. 212  
 Caplan, P. 40
- Carlton College (Science Education Resource Center) 167–8  
 Carnegie Mellon University 10  
 Carr, L. 96, 97, 191  
 Carr, R. 163  
 Casey, M. 210  
 Castells, M. 151  
 cataloging and classification 49, 113, 128  
 Cathro, W. 148  
 Cave, R. 219  
 CENL *see* Conference of European National Librarians  
 Center for History and New Media 221  
 Center for Research Libraries (US) 156  
 CERN *see* European Organization for Nuclear Research  
 Chakraborty, A.K. 212  
 Chapman, S. 47  
 China 17, 18, 43–4, 101*t*, 102  
 Chinese Academic Digital Library (CADLIS) 17  
 Chinese Academic Library Information System (CALIS) 17  
 Chinese Academy of Sciences 101*t*  
 Chowdhury, G.G. 24–5, 60  
 Chowdhury, S.C. 24–5  
 Chronicling America 247  
 Chute, R. 102  
 CIA World Fact Book 76  
 CIBER Research 201  
 CIC *see* Committee on Institutional Cooperation  
 circulation trends 114*f*, 115  
 citation advantage: scholarly communication 94, 95  
 citation management 220–2, 225  
 CiteSeerX 100*t*, 220  
 CiteULike 212, 220  
 citizen science 254  
 Clegg, W. 29  
 CLIR *see* Council on Library and Information Resources  
 cloud-based services 25–6, 109, 125, 129–30 134, 221  
 CNRI *see* Corporation for National Research Initiatives  
 ‘co-curation’ 166  
 co-operatives 28–9, 47  
 COAR 102  
 CogPrints 38  
 Cohen, Dan 168–9, 213, 221, 252  
 collections: traditional libraries 85–6, 140  
 ‘collective collecting’ 123, 136  
 Collier, M. 175  
 Collins, L.M. 99  
 Colorado, University of 185  
 Committee on Institutional Cooperation (CIC) 121  
 commons-based digital libraries 149, 241–2  
 community building 28–9  
 community centers 166*t*  
 community engagement 77–8, 226–7  
 community orientation 163, 209  
 ‘Comparison of Research Networking Tools and Research Profiling Systems’ (Wikipedia) 224  
 compete.com 52  
 computer and information sciences 3, 6, 10, 48, 72, 87, 89  
 Computers and the Humanities, Association for 5  
 Computing Machinery, Association for (ACM) 29, 54*t*  
 Conference of European National Librarians (CENL) 120  
 conferences/conference proceedings 9, 28–9, 31, 65  
 conglomerates 204–5  
 Connaway, L.S. 119  
 Connotea 220  
 content 3–4, 6, 82*t*, 94, 113, 122, 143–4  
   licensing 130, 131, 132  
   recruiting 181–2, 183*t*, 184  
   types of 154*t*, 155

- useability of 167–8, 169, 170*t*, 171  
*see also e-content*  
 Conway, P. 154  
 Cool, C. 29  
 copyright 34, 121–2, 130, 147  
   key issues 79–80, 81–2*t*  
 CORE 10  
 Cornell University 12, 41, 99, 172, 202, 222  
 Cornish, G.P. 79  
 Corporation for National Research Initiatives (CNRI) 61  
 Corral, S. 29, 201  
 costs: repositories 92  
 Council on Library and Information Resources (CLIR) (US) 47  
 Cousins, J. 252  
 Covenant University (Nigeria) 243  
 Covey, T. 182, 188, 189  
 Cox, J. and L. 235  
 Cramer, T. 109  
 Crane, G. 253  
 Craven, J. 144  
 crawling: web 24, 86, 88, 113, 125, 149, 204  
 Creaser, C. 144, 192  
 Creative Commons CC0 1.0 Universal Public Domain Dedication 248  
 Crosas, M. 200  
 CrossRef 39, 233–4  
 CrossRef DOI Resolver 247  
 Crow, R. 152  
 crowdsourcing 251, 252–4  
 Crowston, K. 254  
 cultural and heritage collections/organizations 7, 10, 14, 17, 77, 99, 133  
   early defining digitization projects 43–7  
   Google and 195  
   hybrid libraries 116, 118  
   linked data 247–8  
   public-private partnerships 174  
   social role 145, 148, 149–50, 152, 153, 156  
*Current Cites* 30  
 cyberinfrastructure 199  
 Cybermetrics Lab 91, 92, 100*t*  
  
 ‘d-facts’ 153  
*D-Lib Magazine* 29–30, 59–60, 61, 62*t*, 63, 64*f*, 65  
 D4Science-II project 203  
 DAI *see* Dutch Author Identifier  
 Daniel, R. 50  
 DAREnet 102, 203  
 DARPA *see* Defense Advanced Research Projects Agency  
 DART 200  
 Das, A. K. 17  
 data curation 200–1, 202  
 ‘data deluge’ 199  
 Data Digital Libraries (France) 247  
 Data Hub 75, 246  
 DataCite 200  
 DataOne 200  
 DataVerse Network (DVN) 200  
 Davis, P. M. 148–9  
 DBPedia 72  
 De Angeli, A. 225, 226*t*  
 De Castro, P. 107  
 De Laurentis, C. 152  
 De Rosa, C. 164  
 De Roure, D. 108  
 Deegan, M. 25, 144, 150, 151, 152, 255  
 deep web 66  
 Defense Advanced Research Projects Agency (DARPA) 2, 15, 61  
 delicious.com 220  
 DeLlver *see* Desktop link to Virtual Engineering Resources  
 DELOS  
   definition of digital libraries 20  
   Network of Excellence on Digital Libraries 212  
   Reference Model/Manifesto 203  
 Dempsey, L. 12, 16, 119, 144, 153, 239  
   hybrid libraries 135, 136  
 Denicola, R. C. 186  
 depositing 96, 97, 108, 205  
   mandates 190–2  
   open access repositories 181, 182, 187  
 Desktop link to Virtual Engineering Resources (DeLlver) 34  
 destination sites 196  
 Dewey Decimal Classification System 49  
 Dialog 3  
 Dickinson, D. 180, 204  
 DIEPER 43  
 Digital Commons 97  
 Digital Curation Centre (UK) 154  
 ‘digital divide’ 23, 44, 144–5, 205  
*Digital Futures* (Deegan and Tanner) 255  
 digital libraries  
   definitions 18–20, 21–3*t*, 24–6  
   early developments 1–7  
   funding 13–15  
   key outcomes 27*f*  
   motivating factors for 7, 8*f*, 9–10  
   sample of working libraries 51, 52, 53–6*t*  
   as ‘socio-technical systems’ 60  
   University of Virginia 12, 99 *see also* Fedora  
   Virginia Tech 55*t see also* NDLTD  
   vision for 6–7  
 Digital Library for Earth Science Education (DLESE) 167  
 Digital Library Foundation 14  
 Digital Library of India 17  
 Digital Library Initiative (DLI-1/-2) 15, 16, 36, 37, 142  
 Digital Library Manifesto (DELOS) 212  
 Digital Museum Project (China) 102  
 Digital Object Identifier (DOI) 39, 200, 231  
 Digital Public Library of America (DPLA) 121, 168–9, 248, 252  
 digital reformatting *see* digitization  
 Digital Repositories Programme (Joint Information Systems Committee (JISC)) 107  
 Digital Shift, The (blog) 30  
 digital value chain 152–3  
 Digital Vita 224  
 digitization 7, 10, 43–8, 122–3, 250  
   research library collections 119–23  
 Directory of Open Access Journals 25  
 Directory of Open Access Repositories (OpenDOAR) 91, 92, 93–4, 95*f*, 97–8  
 ‘disaggregated’ model: scholarly content 152  
 discipline-specific reward systems 192–3  
 disclosure 132–3  
 discoverability 86, 119, online communities 165  
   disclosure/visibility 132–5  
   hybrid libraries 124, 125  
   institutional repositories 197*t*  
   researchers 228–9  
   scholarly communication 77, 93–4, 95*f*, 96–7, 125–9  
   services 126–9, 135  
 distributed digital libraries 24  
 Distributed Proofreaders 253  
 DLESE *see* Digital Library for Earth Science Education  
 DLI-1/-2 *see* Digital Library Initiative  
 DOI *see* Digital Object Identifier

- Dougherty, D. 210  
 DPLA *see* Digital Public Library of America  
 DRIVER 96, 102, 203  
 Drupal (web content manager) 108  
 Dspace 172, 243–4  
   repositories and 92, 97, 98–9, 101*t*, 105, 106, 107  
 Dublin Core Metadata Initiative/Element Set 30, 48–9, 73  
 Dunhuang Caves 43–4  
 Dunleavy, P. 221–2  
 Duranceau, E.F. 107, 194  
 Duraspace 98–9, 109  
 Dutch Author Identifier (DAI) 231*t*, 232  
 Dutta, C. 17  
 DVN *see* DataVerse Network
- EAD *see* Encoded Archival Description  
 Earth Science Education, Digital Library for (DLESE) 167  
 e-books 115–16, 118, 126, 131, 135, 163, 253  
 e-commerce 249  
 economic sustainability 83  
 economic value 151–3  
 e-content 33, 115–16, 118, 126, 127, 129  
   *see also* content  
 ecosystem 203–4, 206*f*  
 e-Depot (Netherlands) 34  
 e-discussion lists 30  
 EDM *see* Europeana Data Model  
 educational digital libraries 150, 167, 219  
   *see also* teaching and learning communities  
 Eduserv 109  
 Egypt 18, 173*t*  
 e-health 214  
 ‘e-inclusion’ 144  
 Eisenberg, M.B. 112  
 e-journals/serials 33, 115, 117*f*, 131, 134, 154*t*, 235  
   *see also* journals/serials  
 ‘electronic campus’ 5  
 ‘electronic library’ 15, 19  
 eLib Programme (UK) 5, 15, 29, 33  
 E-LIS 100*t*  
 Elsevier 31, 33, 54*t*, 222, 224, 233  
 empowerment 148–9  
 enabling technologies 82*t*, 84, 124*f*  
   hybrid libraries 131, 132, 133, 135, 136  
 Encoded Archival Description (EAD) 6  
 Endeca 128–9  
 EndNote 221  
 end-users: academic institutions 185  
 EPrints *see* GNU EPrints  
 EPUB (e-book format) 115  
 e-publishing (UK) 33  
 e-research 199–202  
 e-resources 112  
   discovery services 126, 133  
   licensing 131, 132  
   management systems (ERMs) 131, 135  
 Erway, R. 172, 174, 179  
 e-science 199, 201  
 ethical sustainability 83  
 European Commission (EC) 16–17, 26, 102, 120, 144, 216  
   open access 186, 187  
 European Committee for Standardization 107  
 The European Library (TEL) 68, 120, 127  
 European Organization for Nuclear Research (CERN) 11, 101*t*  
 European Research Council 190  
 Europeana 106, 121, 122, 134, 135, 153, 203  
   knowledgebases/registries 164, 166  
   linked data 247–8, 251–2, 253  
 Europeana Data Model (EDM) 248  
 Ex-Libris 41  
 eXtensible Markup Language (XML) 73, 103–4  
 Eysenbach, G. 94, 95, 214
- Facebook 225, 231, 242  
 FamilySearch Indexing 164, 253  
 Fazel-Zarandi, M. 224  
 federated searching 127–8  
 Fedora *see* Flexible Extensible Digital Object and Repository Architecture  
 ‘Fedorazon’ 109  
 fee-based digital libraries 25, 26, 149  
 Fenner, M. 221, 230  
 Ferreira, M. 192  
 Finch Report (2012) (UK) 187, 216  
 Find in a Library service 134  
 Finland 117  
 ‘5S’ framework (Streams, Structures, Spaces, Scenarios, and Societies) 20, 59  
 Fleet, C. 253  
 Flexible Extensible Digital Object and Repository Architecture (Fedora) 12, 92, 99, 105, 107, 108  
 Flickr 149, 153, 165, 241–2  
 Flickr Commons 242  
 Follett report 5, 15  
 force field analysis 205, 206*f*  
 Foresite 106  
 format-based digital libraries 54–5*t*, 150  
 foundations: community building 28–9  
 Fox, E. A. 12, 14, 19, 142, 143  
*Framework of Guidance for Building Good Digital Collections*, A (National Information Standards Organization) 47–8  
 Framework Programmes (European Commission) 16–17  
 France 13, 45–6, 53*t*, 140, 173*t*, 247, 251  
 Franssen, J. 125  
 Freebase 76  
 French Publishers Association 46  
 French Society of Literary Authors 46  
 Fressard, O. 140  
 ‘frozen’ digital libraries 212  
 full-text repositories 19, 34  
 funding 13–15, 28, 99, 105, 120, 239–40  
   academic libraries 5  
   deposit mandates 190  
   early development and 51, 53–6*t*  
   impact of shifts in 31–2  
   online communities 171–2, 173*t*, 174–6, 177  
   open access and 186–7  
 Funding Councils (UK) 186, 187  
 future digital resources 12
- GABRIEL *see* Gateway and Bridge to Europe’s National Libraries  
 Galaxy Zoo 254  
 Gallica 45–6, 53*t*, 150, 162, 164, 247, 251  
 Garcia-Molina, Hector 37, 69  
 Gateway and Bridge to Europe’s National Libraries (GABRIEL) 127  
 Gazan, R. 241  
 Geleijnse, H. 32  
 genre-based digital libraries 54–5*t*, 150  
 Georeferencer 253, 254  
 geospatial communities 6  
 Germany 173*t*  
 Gerolimos, M. 240–1  
 Ginsparg, P. 4, 35, 36, 163  
 Gipp, B. 24, 244  
 global digital library 26



- GNU EPrints (software) 42, 106, 107, 232  
 mobile devices 243  
 repositories and 91, 92, 97, 98
- Goble, C. 108
- Goertzen, M. 227
- Gold, A. 200
- Gold open access 187
- Gonçalves, M.A. 20
- Google 5, 34, 37, 65, 93–4, 119–20, 231  
 educational digital libraries 167  
 information-seeking behaviors and 113  
 metadata 249  
 subject-based repositories and 180
- Google Books 115, 134
- Google Knowledge Graph 75, 76
- Google Library Project 121, 250
- Google and the Myth of Universal Knowledge: a view from Europe* (Jeanneney) 120
- Google Scholar 24, 30, 31, 67, 93–4, 132, 134  
 citation management 222  
 open access 195, 196, 204, 205
- Gore, Al 14
- Göttingen University 43
- government agencies 183*f*, 186
- Green open access 187
- Green, R. 108
- Greenstein, D. 14
- Greenstone (software) 44, 46, 243
- Griffin, S. 31
- Gruzd, A. 227
- Guédon, J.-C. 187
- Gutenberg Bibles 44  
*see also* Project Gutenberg
- Guthrie, K. 174–5
- Hagedorn, S. 71, 101
- HAL Institut National de Recherche en Informatique et en Automatique Archive Ouverte (HAL-INRIA) 101*t*, 164
- HAL Sciences de l'Homme et de la Société (HAL-SHS) 101*t*
- Hammond, T. 220
- Handle System 39
- Hanlon, A. 189, 194
- Hanson, C. 113, 125, 132
- Hargreaves Review of IP and Growth (UK) 80
- Harle, J. 144
- Harley, D. 151
- Harmony Project 222–3
- Harnad, Stevan 34, 94, 95, 187, 190–1
- Harriman, J. H. P. 175
- Harris Interactive 112
- Hart, Micheal 17, 163
- Harvard Depository 117
- Harvard Law School 190
- Harvard University (Faculty of Arts and Science) 190, 200
- harvesting: metadata 45, 86, 100, 101, 233
- Haslhofer, B. 106
- HathiTrust Digital Library 121–2, 128
- Head, A.J. 112
- Health, National Institutes of (NIH) (US) 16, 95
- healthcare *see* medicine and healthcare
- Heath, T. 72, 74, 75
- Heery, R. 203
- Heidari, A. 144
- Hendler, J. 73
- Henry, G. 136, 144
- Henty, M. 97
- Hessel, H. 113, 132
- heterogeneity 67–8
- Hewlett Packard (HP) 98, 172
- Hey, T. 199
- Hill, A. 232
- Hirtle, P.B. 79
- historical pictorial collections 47, 241–2
- History and New Media, Center for 221
- Hitchcock, S. 41, 154
- Hochstenbach, P. 102
- Holley, R. 253
- Horava, T. 131
- hosts 183*t*, 184
- How Much Information?* (Bohn and Short) 8*t*, 9
- Howe, Jeff 252
- HTTP *see* Hypertext Transfer Protocol
- Hudson, E. 79
- Hull, D. 212
- Hull, University of 108
- human-centered design *see* user-centred design
- Humanities Media Interface (HUMI) 44
- humanities research 5, 164, 182, 253
- HUMI *see* Humanities Media Interface
- Huntington, P. 188
- hybrid libraries 12–13, 24–5, 66, 111–13, 135–7, 200  
 changing collections 116, 117*f*, 118–19  
 changing technologies 123, 124*f*, 125  
 changing use and engagement with 114*f*, 115–16  
 digitizing research library collections 119–23  
 E-resource management 131–5  
 interoperability/discoverability 40, 68, 87, 125–9  
 licensing terms and conditions 130–1  
 next-generation systems (cloud-based) 129–30  
 timeline 90*f*
- Hydra project 108, 213
- hyperlinking systems 19, 34, 88
- Hypertext Transfer Protocol (HTTP) 73, 104
- i2010 Digital libraries 120, 121, 144
- IDEALS repository (University of Illinois) 154
- identifiers 11, 39, 73, 74, 246–7
- IEEE Explore 24, 195
- IEEE (formerly Institute of Electrical and Electronics Engineers) 29
- IFLA *see* International Federation of Library Associations
- Illinois, University of 17, 154
- IMLS *see* Institute of Museum and Library Services
- INASP *see* International Network for the Availability of Scientific Publications
- indexing 7, 9, 11, 19, 24, 88, 128  
 associative 37, 65  
 by Google 113, 119  
 deep web 66  
 family history 253  
 search engines 218
- India 17, 18
- InfoPro Technologies 3
- informatics communities 6
- Information Access Platform (Endeca) 128–9
- information consumption 8*f*, 9
- 'information network overlay' 21
- information retrieval (IR) theory and systems 3, 12
- 'information superhighway' 4
- information-seeking behaviors 111–13
- initial funding 172
- Institut Cartografic de Catalunya 253–4
- Institute of Electrical and Electronics Engineers *see* IEEE
- Institute of Museum and Library Services (IMLS) 16, 115
- Institute for Statistics (UNESCO) 226
- institutional repositories 89, 91, 92, 96–7, 109, 179, 196, 197*t*

- content 94, 95*f*
- mediated deposit 194
- social role 154
- traditional scholarly reward system 216
- value propositions 182, 183*t*, 184
- institutions 14, 28–9 47
- ‘intellectual infrastructure’ 142, 156
- Intellectual Property Office (UK) 78–9
- intellectual property rights 78–80, 120, 123, 130, 148
- interfaces *see* software-to-software interfaces; user interfaces
- interlibrary lending 4
- International Conference on Open Repositories (2009) 228
- International Conference on the World Wide Web 72, 73
- International DOI Foundation 39
- International Dunhuang Project (IDP) 43–4
- International Federation of Library Associations (IFLA) 23, 187
- International Network for the Availability of Scientific Publications (INASP) 89
- International Publishers Association 187
- international repositories 100
- International Standard Name Identifiers (ISNIs) 230, 231*t*, 234
- International Standards Organization (ISO) 68–9, 230
- internet 3, 4, 5, 6, 9, 11, 90*f*
- Internet Archive (IA) 14, 17, 55*t*, 120, 121, 163, 164
- Internet Public Library 150
- interoperability 7, 12, 37–40
  - hybrid libraries 123, 124, 125–9
  - as key challenge 65–71
  - repositories 101, 102, 107, 108, 204
  - see also* reference linking
- Ireland 201
- Iriberrri, A. 160, 167, 169, 171, 180, 181
- Islandora 108
- ISNIs *see* International Standard Name Identifiers
- ISO 23950 (Z39.50) (International Standards Organization (ISO)) 38, 68–9, 127–8
- ISO *see* International Standards Organization
- ITHAKA 112–13, 195, 216, 218, 225
  - online communities 170, 173, 174, 175
- Jaeger, P.T. 140, 141
- Japan 43, 44, 232
- Jeanneney, J.-N. 120
- Jeevan, V.K.J. 115
- Jefferson, Thomas 143
- Jilovsky, C. 117
- JISC *see* Joint Information Systems Committee (JISC) (UK)
- Johns Hopkins University 106, 202
- Johnson, T.R. 224
- Joint Information Systems Committee (JISC) (UK) 15, 16, 29–30, 45, 98, 203
  - repositories 106, 107, 108, 109
  - researcher identifiers 232
- Jones, R. 107
- journals/serials 184–5, 187, 188, 195, 207
  - see also* e-journals
- Jouve Group 46
- JSON 104
- JSTOR 25, 42–3, 54*t*, 106, 164
- Kahle, B. 14, 17, 163
- Kahn Wilensky architecture 11, 39, 87, 213, 246
- Keio University (Japan) 44
- Kell, D.B. 212
- Keller, M.A. 72
- Keller, K. L. 163
- Kennan, M.A. 191, 201, 202
- Kenney, A.R. 47
- Kenyon, A.T. 79
- Kieft, R. 123–4
- Kim, J. 182, 188, 192–3
- knowledge economy 151
- Knowledge Ecosystem model 203
- ‘knowledge map’: digital library field 60
- knowledge work 150–1
- knowledgebases 131, 134–5
- Konsta, R. 241
- Kowal, K.C. 253
- Kress, N. 123–4
- Krichel, T. 227
- Krownne, A. 149
- Kumar, B. 17
- Kushmerick, A. 228–9, 233
- Lage, K. 170
- Lagoze, C. 18, 20–1, 50, 66–7, 69, 86, 105
- Lampert, C. 166
- Lana, A. 202
- Lane, J. 228
- Lankes, R.D. 140, 212, 255–6
- LANL *see* Los Alamos National Laboratory
- Large, J.A. 69, 70
- large-scale funding 13, 15, 17–18, 31
- Lassila, O. 73
- Lattes Platform 231*t*, 232
- Lavoie, B. 119, 136, 144, 153
- Law, D. 5, 240, 255, 256
- legacy collections 12, 118, 250
- legal deposit 34, 46, 95, 148, 157
- legal frameworks 156–7
- Leggott, M. 108
- Lercher, A. 182
- Leroy, G. 160, 167, 169, 171, 180, 181
- Lesk, M. 176
- Leuven University 175
- Levy, D. M. 23
- Lewis, S. 107
- Li, Y. 249
- librarianship education 29, 49
- libraries *see* digital libraries; traditional libraries
- Libraries Action Plan (European Commission) 16–17
- Libraries of the Future* (Licklider) 2, 26
- Library 2.0 concept 210
- Library Associations, International Federation of (IFLA) 23, 187
- Library of Congress (US) 4, 16, 53*t*, 109, 120, 133–4, 140, 241–2
- Library and Information Networking, UK Office for (UKOLN) 5
- Library Journal* 116
- library management systems 123
- Library Materials Survey (2012) (US) 116
- Library Resources, US Council on 2
- ‘library without walls’ 19
- LibraryCrunch 210
- LibrarySearch 200
- LIBRIS 247
- licensing 115–16 118, 155, 189
  - hybrid libraries 130–1, 132, 135
- Licklider, J.C.R. 2, 26, 245, 256
- Liew, Chern Li 60, 144, 170
- life-cycle model: online communities 160, 161*f*, 169, 176–7, 180, 219
- linked data 104, 105, 133, 245–50
  - applications for 71, 72, 74, 75–7
- LinkedIn 225
- Linking Open Data Project 74–5

- Lippincott, J. K. 243  
 literature: key themes 19, 29, 30*f*, 32, 160  
   concept map 59–61, 62*t*, 63, 64*f*, 65, 139  
 Liu, Y.Q. 29, 230  
 Live Search Books 120, 121  
 Los Alamos National Laboratory (LANL) 41, 53*t*, 102–3  
 Los Alamos Physics Archive 34–5  
 Lossau, N. 102  
 Lowry, C. 131  
 Loy, M. 173  
 Lund University (Sweden) 129  
 Luther, J. 218–19  
 Lyman, P. 23, 66  
 Lynch, C.A. 28, 31–2, 35, 50, 69, 96, 239
- Ma, Y. 29  
 Machine-Readable Cataloging (MARC) 4, 5–6, 48, 49, 134  
 machine-to-machine data exchange 72  
 Madison, James 140  
 Making Institutional Repositories A Collaborative Learning Environment (MIRACLE) 184, 185  
 Malpas, C. 122  
 Manchester, University of 232  
 Manduca, C.A. 167, 168  
 Maness, J.M. 185  
 Manghi, P. 102  
 MARC *see* Machine-Readable Cataloging  
 Markey, K. 96, 181  
 Maron, N. 173  
 Marshak, D. W. 224  
 Marshall, C.C. 23  
 ‘mashups’ 104, 211  
 Maslov, A. 106  
 mass digitization 10, 81*t*, 250  
   hybrid libraries 117, 119, 120, 122–3, 135, 136  
   preservation and 154*t*  
 Massachusetts Institute of Technology (MIT) 2, 98, 101*t*, 172  
 massive open online courses (MOOCs) 168  
 materials budgets 116  
 McCain, K. 3  
 McClure, C.R. 140, 141, 143, 209  
 McDonald, R.H. 97, 191  
 McGovern, N.Y. 156, 191–2  
 Mead Data Central 3  
 media sites 113  
 mediated deposit 194  
 ‘Medicine 2.0’ (Eysenbach) 214  
 medicine and healthcare 214, 218, 247  
 Medline 247  
 Mellon Foundation 101, 105, 122, 172, 183  
 ‘memex’ 2, 37, 86  
 Mendeley 164, 212, 221–2, 225, 226*t*  
 Menendez, M. 225, 226*t*  
 Menestrina, Z. 225, 226*t*  
 Mercer, H. 97  
 Mercury 10  
 MetaArchive Cooperative 156  
 metadata 11, 25, 101, 103, 126, 156  
   early development 48–51  
   linked data 248, 249  
 metasearch 127–8  
 Meyer, L. 155–6  
 Miaskiewicz, T. 185  
 Michigan, University of 33, 101  
 microdata 249  
 Microsoft 120, 121, 231  
 Microsystem and Information Technology, Shanghai Institute of 101*t*  
 Miller, R. 131
- Million books project 18  
 MIMAS Names Project 231*t*, 232  
 Minnesota, University of 113, 124, 125, 132  
 MIRACLE *see* Making Institutional Repositories A Collaborative Learning Environment  
 mission-directed digital libraries 55–6*t*  
 MIT *see* Massachusetts Institute of Technology  
 Mitchell, C. 243  
 Mitchell, E. 245  
 mobile devices 144, 211, 243–4  
 Moen, W. E. 68  
 Monash University 200  
 MOOCs *see* massive open online courses  
 Moore, G. 181, 186, 188, 192, 193  
 Moravian Library 253  
 Morris, S. 188  
 Mostafa, J. 61  
 multimedia resources 249  
 Museum and Library Services, Institute of (IMLS) 16, 115  
 MusicBrainz 76
- NACSIS-ELS 43  
 name authority control systems 229  
 Name Identifier Summit (2009) 233  
 Names Project (UK) 232  
 NASA *see* National Aeronautical and Space Agency  
 Nationaal Archief (The Hague) 253  
 National Aeronautical and Space Agency (NASA) 15, 100*t*, 101*t*  
 National Center for Biotechnology Information (NCBI) (US) 16  
 National Center for Supercomputing Applications (NCSA) (US) 48  
 National Digital Information Infrastructure and Preservation Program (NDIIPP) (US) 153, 157  
 National Digital Library Project (NDLP) 16, 251  
 National Information Standards Organization (NISO) (US) 41, 47–8  
 National Institutes of Health (NIH) (US) 16, 95, 186, 190, 191, 223  
 national libraries  
   digitization programs 13, 45, 53*t*, 116, 135, 149–50, 156–7  
   success factors 162  
 National Library of Australia (NLA) 46, 53*t*, 148, 253  
 National Library of New Zealand 44, 46  
 National Library of Scotland 253  
 national programs (US) 16  
 National Science Digital Library (NSDL) (US) 102, 167, 168  
 National Science Foundation (NSF) (US) 1, 29, 36, 41, 61, 105, 143  
   e-research 199, 201  
   funding from 15, 16  
 National Taiwan University 101*t*  
 National Virtual Observatory (US) 106  
 national-level research evaluation 232  
 Nature Publishing Group 220, 233  
 NCBI *see* National Center for Biotechnology Information  
 NCSA *see* National Center for Supercomputing Applications  
 NCSTRL 38  
 NDIIPP *see* National Digital Information Infrastructure and Preservation Program  
 NDLP *see* National Digital Library Project  
 NDLTD *see* Networked Digital Library of Theses and Dissertations  
 NEDEC *see* Northeast Document Conservation Center  
 needs 181, 182

- assessment 45, 162–3, 170
- orientation 165, 166*f*, 167–9
- negotiation: licensing 130, 131, 135
- Netherlands 13, 96, 129, 141, 148, 203
  - researcher identifiers 231*f*, 232
  - scholarly communication 32, 34
- Networked Digital Library of Theses and Dissertations (NDLTD) 55*t*, 68
- new digital resources 12
- New York University 122
- New Zealand 13, 44, 46, 201
- newspapers 46, 47, 148
- Newton, M.P. 183, 186
- next-generation
  - hybrid library systems 129–30
  - repositories 108–9
- Nguyen, Son Hoang 28, 60
- Nicholas, D. 188
- Nielsen, J. 169
- NIH *see* National Institutes of Health
- NISO *see* National Information Standards Organization
- Niu, J. 229
- NLA *see* National Library of Australia
- non-digital collections 12, 66, 86, 87, 111, 127
- Norris, M. 93–4
- North Carolina State University 128–9, 249
- Northeast Document Conservation Center (NEDEC) 47
- Norway 232
- Nottingham University 189
- NSDL *see* National Science Digital Library
- NSF *see* National Science Foundation (US)
  
- OAI *see* Open Archives Initiative
- OAI2LOD Server 106
- OAIS *see* Open Archival Information System
- OAIster 101, 93–4, 203
- OASIS *see* Organization for the Advancement of Structured Information Standards
- Object Reuse and Exchange (ORE) 104–6
- O'Brien, A. 29, 196
- O'Brien, P.S. 93, 195, 244
- Ockham Initiative 213
- OCLC 134, 156, 249–50
- OCLC Online Computer Library Center (formerly Ohio College Library Center) 4, 48, 49
- OCLC Research 40, 122
- O'Connor, S. 117
- OECD *see* Organization for Economic Cooperation and Development
- OECD Principles and Guidelines for Access to Research Data from Public Funding* (Organization for Economic Cooperation and Development) 199
- Ohio College Library Center (later OCLC Online Computer Library Center) 4
- Oliver, K. B. 91
- Onaifo, D. 66, 195, 244
- one universal digital library 14–15
- ongoing funding 172
- online communities 52, 176–7
  - community and needs orientation 165, 166*t*, 167–9
  - funding and sustainability 171–2, 173*t*, 174–6
  - growth and quality content 171
  - inception of 162–5
  - sustainability 159–60, 161*f*, 162
  - user-centered design/ease of use/reliability 167, 169, 170*t*, 171
- online information industry 3–4, 6, 9
- Online Newspaper Archive (British Library) 47
- OpCit *see* Open Citation Project
- open access 17, 18, 87, 135, 147–8
  - copyright 82*t*, 188–9
  - cultural and heritage content 148
  - deposit mandates 190–2
  - digital divide 144
  - early development 34–5, 41–2
  - future trends 198–205, 206*f*
  - Google 194–6
  - intellectual property rights 78
  - licensing 130
  - policy and legal frameworks 186–7
  - repositories 204–5, 207–8
  - scholarly information 93–4, 95*f*, 96–7, 148, 151, 152
  - self-archiving 192–4
  - subject-based repositories 179–80
  - support and advocacy for 98
  - sustainability 174
  - value of institutional repositories 180–2, 183*t*, 184–6, 196, 197, 198, 218
- Open Access Journals, Directory of 25
- Open Access Repositories, Directory of (OpenDOAR) 91, 92, 93–4, 95*f*, 97–8
- Open Archival Information System (OAIS) 156
- Open Archives Initiative (OAI) 69, 91
  - Object Reuse and Exchange (ORE) 104–6, 245, 248, 249
  - Protocol for Metadata Harvesting (OAI-PMH) 38–9, 91, 100, 101
- Open Citation Project (OpCit) 41–2, 91
- Open Content Alliance 17, 120
- Open Contributor and Researcher ID (ORCID) 231*t*, 233–4
- open data interchange 25–6
- Open Journal demonstration project 40
- open repositories 219
- Open Repositories Conference (2008) 106
- Open Repositories, International Conference on (2009) 228
- OpenDOAR *see* Directory of Open Access Repositories
- OpenID 231*t*
- OpenURLs 40–1
- Oppenheim, C. 93–4
- ORCID *see* Open Contributor and Researcher ID
- ORE *see* Object Reuse and Exchange
- O'Reilly, Tim 210, 239
- Organ, M. 94
- Organic ePrints 100*t*
- Organization for the Advancement of Structured Information Standards (OASIS) 92
- Organization for Economic Cooperation and Development (OECD) 199
- orphan works 81*t*
- Ostrow, S.E. 47
- outcomes-based assessment 140, 141, 157–8
- Overly, E.A. 16
  
- Paepcke, A. 69
- Page, L. 36–7, 65, 88, 195
- PageRank 36–7, 65, 88
- Palmer, C. L. 183, 186, 192
- Papers Past 44, 46, 54*t*, 162
- parent institutions 183*t*, 185
- Park, T.K. 62
- 'participatory librarianship' 256
- partnerships 28–9, 118, 121, 170, 202, 250, 252
  - knowledgebases/registries 134, 135
  - public-private 174
- Pathways project (National Science Foundation) 105
- Patron-Driven Acquisitions (PDA) 118
- Payne, L. 123–4, 194
- PayPal 231
- peer-reviewed publications 60, 61, 65
- Perseus Digital library 43, 53*t*, 100*t*, 253

- persistent identifiers 39–40, 41, 69, 200, 228  
 Persistent Uniform Resource Locator (PURL) 40  
 personal computers 4, 11  
 personal digital libraries 149, 219–20, 221  
 Peters, D. 102  
 Pettifer, S.R. 212  
 Pew Internet Project 211, 243  
 photographic collections 46, 241–2, 253–4  
 Picture Australia 46  
 Pinfield, S. 15, 191  
 Pinterest 242  
 Pitti, D. 5  
 PlanetMath 149  
 Pridal, P. 253  
 Pomerantz, J. 59–60  
 portals 126–7  
 Powell, A.L. 61, 203  
 Powell, J.E. 99  
 Poynder, R. 191  
 pre-prints 34, 151, 192  
 Preserv 2 project 106  
 preservation 25, 34, 42, 47, 87, 108  
   copyright 81*t*  
     hybrid libraries 117, 119  
     institutional libraries 197*t*  
     intellectual/cultural assets 120, 121, 153, 154*t*, 155–7  
     social role 147*t*  
 Priem, J. 218  
 Prince Edward Island, University of (UPEI) 108  
 Pringle, J. 233  
 print collection management 117  
 professional societies 7, 9  
 Proffitt, M. 165, 242  
 ‘Project Digitise’ (British Library) 45  
 Project Gutenberg 17, 55*t*, 163, 164, 251, 253  
   *see also* Gutenberg Bibles  
 Proquest 34, 233  
 Protocol for Metadata Harvesting (OAI-PMH) 100  
 proxy servers 132  
 public domain 81*t*  
 public libraries 114*f*, 115, 140–1, 196, 209, 243, 244  
 Public Library Data Services (US) 114*f*  
 Public Library Surveys (2009/2010) (Institute of  
   Museum and Library Services) 115  
 public policy environment 156–7  
 publication patterns: digital library topics 30*f*, 31  
 public–private partnerships 174  
 publishers 7, 9, 233  
 Publishers Association, International 187  
 PubMed Central 16, 24, 89, 95, 114, 246  
   open access 190, 195  
   subject-based repository 100*t*  
 Puckett, J. 221  
 PURL *see* Persistent Uniform Resource Locator  
 ‘push’ initiatives 165  
  
 ‘Quand Google défie l’Europe’ (‘When Google  
   Challenges Europe’) (Jeanneney) 120  
 Queensland University of Technology 190  
  
 Ramirez, M. 189, 194  
 Ranking Web of World Repositories 91  
 Rasmussen, D. 66, 196, 244  
 RDA *see* Research Data Australia  
 RDF *see* Resource Description Framework  
 reach: scholarly communication 94, 95  
 read-only sites 21, 159, 212, 213, 251  
 Really Simple Syndication (RSS) feeds 220, 241  
 Rebollo, R. 200  
 ReCAP *see* Research Collections Access & Preservation  
 Recker, M. 167  
  
 recognition: branding 181  
 Red Sage 33  
 Reddy, Raj 18  
 reference linking 40, 69, 113  
   *see also* interoperability  
 referrers 113, 165, 180, 195, 196, 242  
 reformatting *see* digitization  
 RefWorks 221  
 registries 131, 134–5  
 Registry of Open Access Repositories (ROAR) 42, 91,  
   92, 97–8, 190  
 relevance 137  
 reliability: content 167, 169, 170*t*, 171  
 REMAP 108  
 remote access: e-resources 132  
 RePEC *see* Research Papers in Economics  
 RepoMMan 108  
 repositories 11, 18  
   common search services 100, 101*t*, 102–3  
   current position and roles 93–4, 95*f*, 96–7  
   data re-use/disclosure/dissemination 104–7  
   evolution of 89, 90*f*, 91–3  
   international 100  
   multiverse of research and practice 88–9  
   next-generation 108–9  
   open access 41–2, 219  
   repository architecture 87  
   researcher identification systems 233  
   subject-based 99, 100*t*  
   systems and software 97–9  
   traditional libraries 85–7  
   web crawling and indexing 88  
   web services 103–4  
 ‘Repositories and the Cloud’ (JISC/Eduserv) 109  
 Research Assessment Exercise (2008) (UK) 232  
 Research Collections Access & Preservation (ReCAP)  
   122  
 Research Councils (UK) 186, 187  
 research data 154*t*  
 Research Data Australia (RDA) 200  
 Research Excellence Framework (UK) 187  
 Research Information Network 201  
 Research Libraries, Association of *see* Association of  
   Research Libraries (ARL)  
 Research Libraries, Center for (US) 156  
 Research Papers in Economics (RePEc) 38, 89, 100*t*, 227,  
   231*t*, 233  
 ResearcherID 224, 231*t*, 233  
 researchers  
   identifiers 227–30, 231*t*, 232–4  
   profiling systems 222–4  
 ResearchGate 225, 226*t*  
 Resource Description Framework (RDF) 49, 73, 74, 76,  
   249  
 restricted-access 25–6  
 reward systems: scholarly 192–3, 216, 218  
 Reznik-Zellen, R. 90, 99, 180  
 Richardson, J. 200  
 Rider, F. 8, 9  
 Rieger, O. 181, 193, 216  
 rights: authors 188–9, 248  
 RLG Cultural Materials Alliance 172  
 ROAR *see* Registry of Open Access Repositories  
 Rochester, University of 182, 184  
 Rodgers, R. 107  
 Romary, L. 97  
 Ronallo, J. 249  
 Rosa, I.B. da 243  
 Ross, S. 153  
 Rotenberg, E. 228–9, 233  
 Rowland, F. 93–4

- Rowlands, I. 188  
 RSS *see* Really Simple Syndication (RSS)  
 runaway growth 7, 8f, 9  
 Rusbridge, C. 12–13, 24  
 Rutner, J. 195
- Safire, W. 252  
 Salako, A. 169  
 Sale, A. 191  
 Sanderson, R. 77  
 Santa Fe Convention 38  
 Sarr, N. 184  
 Sattler, K.-U. 71  
 Savenije, B. 141, 144, 148, 174  
 Schandl, B. 106  
 Schema.org 76, 249–50  
*Scholar and the Future of the Research Library, The* (Rider) 8, 9  
 Scholar Universe 233  
 scholarly communication 6, 7, 9, 32–4, 35, 36, 42–3, 111  
   citation management 221–2  
   collaboration/process of knowledge 150–1  
   copyright 82t, 188–9  
   data re-use/disclosure/dissemination 104–6  
   linked data 246  
   open access 91, 93–4, 95f, 96–7, 135, 152, 207  
   self-archiving 192–3  
   social networks 225, 226t, 227  
   social role 145, 148, 207  
   social web 215f, 216, 217f, 218  
 scholarly name disambiguation 228–9, 230  
 Scholarly Publishing and Academic Resources Coalition (SPARC) 152  
 Scholar's Workbench 108  
 Schonfeld, R.C. 195  
 'School for Scanning' (Northeast Document Conservation Center) 47  
 Schrier, R.A. 241  
 Schwartz, C. 24  
 SciELO *see* Scientific Electronic Library Online  
 Science Education Resource Center (Carleton College) 167–8  
 ScienceDirect 54t  
 Scientific American 73  
 Scientific Electronic Library Online (SciELO) 56t  
 scientific publications 8  
 Scientific Publications, International Network for the Availability of (INASP) 89  
 SciPlore 195  
 SciVal 224  
 Scopus 28, 31, 60, 231t, 233  
 Scotland, National Library of 253  
 search engine optimization (SEO) 93, 195–6, 244–5, 249  
 search engines 67, 128, 149, 204, 205, 218, 249  
   academic 195, 244  
   information-seeking behaviors 112, 113  
   visibility of repositories 93, 94, 125, 133  
 search services: repositories 100, 101t, 102–3  
 Search/Retrieve URL Service/Web Service (SRU/SRW) 68  
 Sefton, P. 180, 204  
 self-archiving 34, 35  
   open access 182, 187, 188–9, 191, 192–3, 194  
 semantic interoperability 69, 70–1, 107  
 semantic web 49, 71–7, 133, 223, 245–50  
   repositories 104, 105, 107, 108  
*semantic web, The* (Berners-Lee, Hendler and Lassila) 73  
 Sen, B.K. 17  
 SEO *see* search engine optimization  
 serials *see* journals and serials
- Seventh International World Wide Web Conference 36–7  
 SFX 41  
 SGML *see* Standard Generalized Markup Language  
 Shanghai Institute of Microsystem and Information Technology 101t  
 shared print repositories 123  
 Shen, X. 17  
 Sherpa RoMEO 98, 189  
 Shibboleth (authentication/authorization service) 132  
 Shirky, C. 255  
 Short, J.E. 8f, 9  
 Shreeves, S. L. 154, 194  
 SIDLP *see* Stanford Integrated Digital Library Project  
 Simister, Shawn 76  
 Simons, N. 200  
 Simple Publishing Interface (SPI) 106, 107, 194  
 Simple Web-service Offering Repository Deposit (SWORD) 106, 107, 194, 206  
 Skinner, K. 156  
 Smith, A. 47  
 Smith, C. 192  
 Smith, K.K. 173  
 Smithsonian Institute 165, 242  
 SMO *see* social media optimization  
 Snyder, S. 165, 242  
 Social Aspects of Digital Libraries Workshop (1996) 143  
 'social audits' 140  
 social environments 160, 161f  
 social inclusion 144  
 social media optimization (SMO) 244–5  
 social networks: scholarly communication 225, 226t, 227  
 Social Science Research Network (SSRN) 89, 97, 100t  
 social sustainability 83  
 social value 18–19, 20, 139–42, 157–8  
   community engagement 77  
   economic value 151–3  
   empowering/informing individuals 148–9  
   foundations of social role 142–4  
   intellectual property rights 78  
   knowledge work 150–1  
   open access 147–8  
   possible framework of social roles 144, 145f, 146–7t  
   preservation of intellectual and cultural assets 153, 154t, 155–7  
   supporting teaching and learning 149–50  
 social web 136, 209, 223, 234–5, 254–6  
   citizen science 254  
   crowdsourcing 252–4  
   definition 211  
   digital libraries and 212–2, 214f  
   digital libraries' social evolution 214, 215f  
   dilemma of national/local collections focus 237, 238f, 239–41  
   large-scale libraries/portals/platforms 251–2  
   mass digitization 250  
   personalization/collaboration 219–24  
   reach/visibility of digital libraries 241–50  
   scholarly collaboration 222–5, 226t, 227–30, 231t, 232–4  
   scholarly communication 215f, 216, 217f, 218  
   Web 2.0/Library 2.0 concepts 210  
 social websites 113  
 Society for American Archivists 5  
 software-to-software interfaces 103–4  
   *see also* user interfaces  
 Somerville, M. M. 170  
 Sornil, O. 12  
 Sound Archive 162  
 Sounds.bl.uk. 45, 55t  
*Source Book on Digital Libraries* (Fox) 14, 142, 143

- Southampton, University of 41, 98  
 Spain 247  
 SPARC *see* Scholarly Publishing and Academic Resources Coalition  
 special collections 116, 118–19, 170, 250  
 SPI *see* Simple Publishing Interface  
 Spink, A. 29  
 Springer-Verlag 33  
 SRU/SRW *see* Search/Retrieve URL Service/Web Service  
 SSRN *see* Social Science Research Network  
 St Jean, B. 185  
 Standard Generalized Markup Language (SGML) 5, 6, 33  
 Standardization, European Committee for 107  
 standards 4–5, 48–51, 68–70  
     researcher identification 230, 231  
 Stanford Integrated Digital Library Project (SIDLP) 36–7, 72  
 Stanford Prizes for Innovation in Research Libraries 247  
 Starting Point Digital Library 168  
 Stevens, R. 108  
 Streams, Structures, Spaces, Scenarios, and Societies ('55' framework) 20  
 Suber, P. 190  
 subject-based repositories 53–4*f*, 99, 100*t*, 109, 174, 194, 219  
     content 94, 95*f*, 96  
     evolution of 89, 91, 92  
     social role 150, 154  
 Suchy, D. 243  
 Suleman, H. 24  
 Sumner, T. 185  
 Supercomputing Applications, National Center for (NCSA) (US) 48  
 superficial uniformity 69  
 SURF Foundation 96  
 sustainability 52, 80, 82–4, 83*f*, 155, 204  
     hybrid libraries 136  
     institutional libraries 197*t*  
     online communities 159–60, 161*f*, 162, 171–2, 173*t*, 174–6 177  
 Sustainable Digital Preservation and Access (Blue Ribbon Task Force) 155  
 Swain, R. 91  
 Swan, A. 173, 174, 182, 184, 186, 189, 191  
 Sweden 247  
 SWORD *see* Simple Web-service Offering Repository Deposit  
 syntactic interoperability 69, 70  
 Syracuse University (School of Information Studies) 254  
 systems and software 97–9, 171  
  
 tagging 242  
 Tammaro, A.M. 29  
 Tang, J. 223–4  
 Tanner, S. 25, 144, 148, 150, 151, 152, 255  
 Tarrant, D. 105, 106  
 Tarrant, J. 144  
 teaching and learning community 6, 19, 149–50, 180, 207, 219, 251  
     context of use 167, 168  
     *see also* educational digital libraries  
 technological developments 2–6, 10–13, 36–41, 171, 239  
     hybrid libraries 123, 124*f*, 125  
     success factors 170*t*  
     user-centred design 169, 170*t*  
 Tedd, L.A. 69, 70  
 Teffeau, L.C. 183, 186  
 TEI *see* Text Encoding Initiative  
 TEL *see* The European Library  
 Tennant, Roy 30, 101  
  
 Ternier, S. 107  
 Texas Digital Library 106  
 Text Encoding Initiative (TEI) 5  
 Theses and Dissertations, Networked Digital Library of (NDLTD) 55*t*, 68  
 Thomas, C. 97, 191  
 Thomas, L.C. 243  
 Thomson-Reuters 31, 224, 233  
 Thorin, S. 14  
 Tilburg University 32, 33  
 Tonta, Y. 47  
 Toronto, University of 181  
 TRAC *see* Trustworthy Repositories Audit & Certification  
 traditional libraries 2, 4, 6, 7, 10, 19, 25, 114  
     collections 85–7, 89  
 training  
     digitization 47–8  
     librarianship 29  
 transitional collections 12  
 Trefethen, A. 199  
 Tripathy, H. 61, 115  
 Trove 46, 53*t*, 162, 251, 253  
 Trustworthy Repositories Audit & Certification (TRAC) 156  
 Tsinghua University 223  
 Tufts University 43, 53*t*  
 TULIP (The University Licensing Program) 33  
 Twitter 225  
  
 UK Office for Library and Information Networking (UKOLN) 5  
 undergraduates: information-seeking behaviors 36, 111, 112  
 UNESCO 23, 226  
 Uniform Resource Identifiers (URIs) 40–1, 246–7  
 Uniform Resource Locator (URL) 39–41, 73, 239  
 United Kingdom (UK) 13, 201  
     e-publishing projects 33  
     information-seeking behaviors 112  
     public libraries 140, 141  
     runaway growth 8  
     sustainability 173*t*  
     technological developments 5  
 United Nations 44  
 United States (US) 13, 14, 18  
     digitization projects 119–20  
     information-seeking behaviors 112  
     public libraries 115, 140, 243  
     sustainability 173*t*  
 Universal Digital Library 18  
 Universidade de São Paulo Biblioteca Digital de Teses e Dissertações 101*t*  
 Universitat Autònoma de Barcelona Dipòsit Digital de Documents 101*t*  
 Universiteit Utrecht Igitur Archive 101*t*  
 universities  
     copyright 188–9  
     deposit mandates 191, 192, 193  
     e-research 200  
     faculty engagement 10, 111, 166, 261  
     funding from 14, 28  
     information-seeking behaviors 36, 112–13  
     licensing 132  
     open access 181–2, 183*t*, 186, 188,  
     researcher profiling systems 224  
     size of repositories 96–7  
     undergraduates 36, 111, 112  
     universities: individual  
         Alicante 247  
         California-San Francisco 33, 121

- Carnegie Mellon 10  
 Colorado 185  
 Cornell 12, 41, 99, 172, 202, 222  
 Covenant (Nigeria) 243  
 Göttingen 43  
 Harvard 190  
 Hull 108  
 Illinois 17, 154  
 Keio(Japan) 44  
 Leuven 175  
 Lund (Sweden) 129  
 Manchester 232  
 Michigan 33, 101  
 Minnesota 113, 124, 125, 132  
 Monash 200  
 National Taiwan 101*t*  
 New York 122  
 North Carolina State 128–9, 249  
 Nottingham 189  
 Prince Edward Island (UPEI) 108  
 Rochester River Campus Libraries 182, 184  
 Southampton 98  
 Syracuse 254  
 Technology (Queensland) 190  
 Toronto 181, 183  
 Tsinghua 223  
 Tufts 43, 53*t*  
 Utah State 167  
 Virginia 12, 99 *see also* Fedora  
 Waikato 44  
 Washington 242  
 Wollongong (Australia) 94  
 University Licensing Program (TULIP) 33  
 UPEI *see* University of Prince Edward Island  
 URIs *see* Uniform Resource Identifiers  
 URL *see* Uniform Resource Locator  
 US Council on Library Resources 2  
 US National Information Systems Task Force (Society for American Archivists) 5  
 useability: content 167–8, 169, 170*t*, 171  
 usefulness 169  
 user interfaces 180  
   hybrid libraries 12, 123, 126, 127, 128  
   institutional libraries 198  
   usability 169  
   *see also* software-to-software interfaces  
 user-centered design 167, 169, 170*t*, 171  
 Usherwood, B. 112, 140, 141, 157  
 Utah State University 167  
  
 value propositions 137, 182, 183*t*, 184–6  
 Van de Sompel, H. 102, 104, 105, 202, 203  
 Van House, N. A. 144, 150, 161  
 Van Westrienen, G. 96  
 Varian, H.R. 66  
 Vattulainen, P. 117  
 Vaughan, J. 166  
 VIAF *see* Virtual International Authority File  
 Vickery, G. 211  
 Virginia, University of 12, 99 *see also* Fedora  
 Virginia Tech 55*t*  
   *see also* NDLTD  
 Virtual International Authority File (VIAF) 229, 231*t*, 247  
 ‘virtual library’ 19  
 virtual private networks (VPNs) 132  
 ‘Virtual Research Environments’ (VREs) 108, 203  
 visibility: scholarly communication 94, 95  
 VIVO 108, 222–3, 224  
 VPNs *see* virtual private networks  
 VREs *see* ‘Virtual Research Environments’  
  
 W3C *see* World Wide Web Consortium  
 Wacha, M. 182  
 Waikato, University of 44  
 Wald, M. 249  
 Walters, T.O. 156  
 Warner, S. 227, 228  
 Warwick Framework 48  
 Washington, University of (UW) 242  
 Waters, D. 155  
 Wayback Machine 17  
 Web 2.0 concept 210, 239, 246  
 web 3, 10, 11, 18, 26, 211  
   architecture 70–1, 87, 105, 109  
   archiving 154*t*  
   crawling 24, 88  
   of documents/data 71  
   information-seeking behaviors 112  
   referral traffic and destinations 113  
   repository services 103–4  
   timeline 90*t*  
   traffic analytics 196  
   visibility of hybrid libraries 132–5  
 Web of Science (WoS) 31  
 web-based bibliographic description 134  
 web-based digital libraries 12, 86  
 Wellcome Trust 190, 191  
 Welsh Repository Network 194  
 White, H.D. 3  
 Wiggins, A. 254  
 ‘wiki’ 211  
 Wikipedia 76, 112, 149, 165,  
   social web 211, 224, 242, 253  
 Wilde, E. 24, 244  
 Wilensky, R. 32  
 Wills, G. 249  
 Wilson, B. 61  
 Wilson, R. 166  
 Wired magazine 252  
 Wisner, J. 123–4  
 Wisner, M. 182  
 Witt, Michael 106  
 Wollongong, University of (Australia) 94  
 Wolski, M. 200  
 work-oriented (ethnographic) perspective 23  
 World Digital Library 120  
 World Wide Web Conference, Seventh International 36–7  
 World Wide Web Consortium (W3C) 49, 72–3, 74, 75  
 World Wide Web, International Conference on the 72, 73  
 WorldCat 134, 250  
 WoS *see* Web of Science  
 Wunsch-Vincent, S. 211  
  
 XML *see* eXtensible Markup Language  
  
 Yahoo 249  
 Yakel, E. 154  
 YouTube 242, 249  
  
 Z39.50 *see* (ISO 23950)  
 Zaugg, H. 221  
 Zepheira 134  
 Zhang, J. 61  
 Zhou, Q. 17  
 Zillman, M.P. 66  
 Zimmermann, C. 227  
 Zotero 212, 221, 222