10 Digital Libraries and the Social Web: Collections and Platforms

Karen Calhoun
Cornell University Library (retired)
ksc10@cornell.edu

Note: This is a preprint of a chapter whose final and definitive form was co-published in *Exploring Digital Libraries: Foundations, Practice, Prospects* by Facet Publishing (2014) and ALA Neal-Schuman (2014).

Overview

This chapter continues the consideration of digital libraries' responses to the social web. It builds on the visual framework introduced in chapter 9 to consider the transition of digital collections to platforms that align well with how people find information, work and play on the social web; are highly visible and invite interaction; and re-mix and re-use data from other sources. The chapter closes with some thoughts about future digital libraries and libraries' digital future.

Visualizing the shift from collections to platforms

The starting point for this chapter is figure 9.2, specifically the branches that pertain to the shift of digital libraries from collections to platforms. Figure 10.1 offers a closer look at these branches and individual sections of this chapter discuss each branch. The left side of figure 10.1 visualizes the current situation; the right side visualizes some innovations, experiments or possibilities. The discussion proceeds from the top toward the bottom of the tree.

Keywords: Digital libraries—Citizen participation; Online social networks; Libraries and community; Cultural heritage collections; Linked data; Discoverability
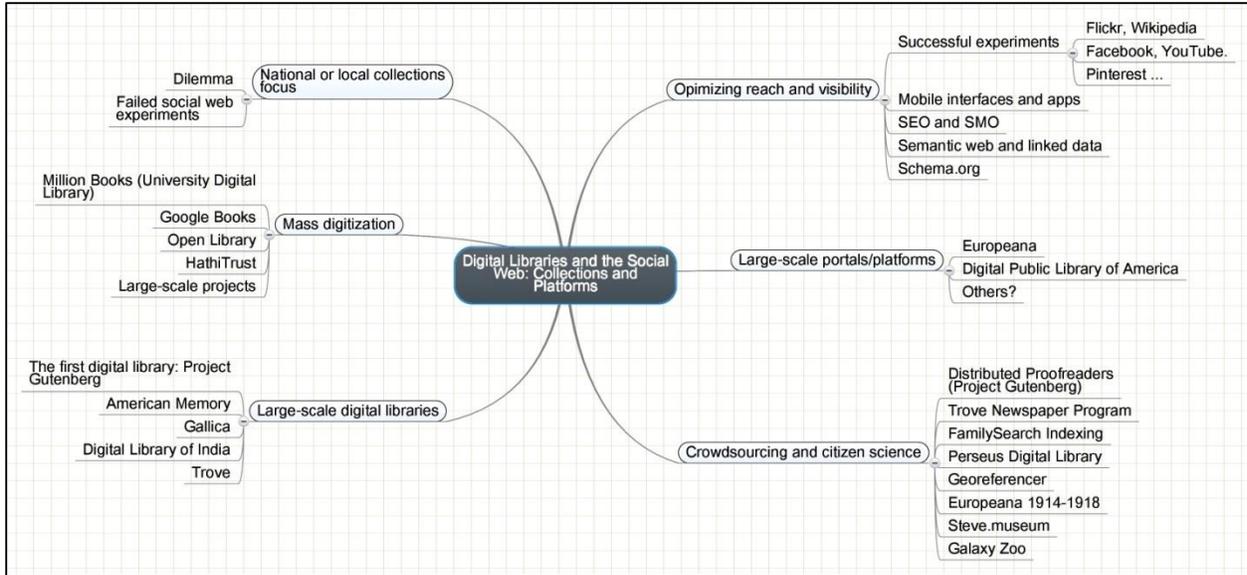
Figure 10.1 The evolution of digital library collections on the social web

The dilemma of the national or local collections focus

The dilemma

The builders and maintainers of real-world digital libraries face a dilemma that comes from two

sources:

- Core assumptions about digital libraries as destination sites, complete in themselves

- The tension between who uses digital libraries and who pays for them

Digital library builders constructed the first wave of digital libraries at national or local

institutional levels, or scholarly publishers built them to move their content online. As the web

grew up around them, second wave initiatives and technologies to give digital libraries web

interfaces and to make them more interoperable with each other developed (like metasearch or

OAI-PMH). Nevertheless a core assumption was that the communities for whom the digital

library was built would visit the digital library at its own URL (that is, the digital library would be a

destination site).

The third wave (where the field is now) is integrating digital libraries more fully with the web and web technologies. Much of that work is described in this book. A key realization of digital library builders has been that most digital libraries are not destination sites, and even ones that are need to be discoverable on the larger web, not only by people but also by machines (i.e., via web services and APIs). Lorcan Dempsey has been making these points for years on his blog; his post after reading Tim O'Reilly's now famous essay on Web 2.0 is particularly interesting in this regard (O'Reilly 2005; Dempsey 2005; see also Dempsey 2006b).

The fourth wave began as the social web grew up around third wave initiatives. As the social web's impact is felt, the goals of making digital libraries more compatible with web technologies, more interoperable with each other, and more visible in search engines are shifting again. The fourth wave is about fully responding to how different communities of people work and play on the web—how they learn and get things done, how they look for other people and information (using search engines and social sites), how they share what they create or find, how information shows up in diverse contexts on the social web, and more. This fourth wave requires an even higher commitment to disclosing digital library content and services in external contexts and for global audiences that the builders never see. This new wave is a disruptive and destabilizing force, but transition is necessary if digital libraries are to continue to thrive.

At the same time, library collections are usually funded at local, regional or national levels, and their funding is intended to support communities at local, regional or national levels. Even before digital libraries began to evolve toward the social web, traditional library funding models (from local or national parent bodies, for local or national communities) did not work well for open digital libraries, which from the beginning transcended the boundaries of place and attracted global audiences. But the global audience for an open digital library does not fund it; the funding

body does. As Cliff Lynch pointed out ten years ago in his insightful piece on political and economic aspects of digital library development (2003), tension around the design, development and continuous improvement of open digital libraries arises when there is a mismatch between who benefits from the digital library and who pays for it. The tension created by this mismatch can be considerable, but it can be reduced through careful work with a digital library's political and economic stakeholders. New business models for open digital libraries can also help; the story of moving to a community-based model to fund arXiv.org is a case in point (Rieger 2011).

Failed social web experiments

Digital libraries have tended to offer simple information access; along the continuums depicted in figure 9.1, many if not most digital libraries today belong at the left (focus on collections and expert communications; read-only or authorized contributions only). As such, they do not offer a favorable "habitat" for successful implementations of social web tools, which arose from quite different conditions (active online communities). This has led to a number of failed experiments introducing social web tools to digital libraries. Derek Law, in an essay about future digital developments for libraries, characterizes some library attempts to apply social web tools as well-intentioned attempts to use new information spaces to deliver old information. He writes "librarians have engaged in almost every fad … without perhaps considering how service philosophy should change … the [key definers of social networks] need to underpin any decision to use the tools … or else we run the risk of further littering the web with inactive library blogs, lifeless virtual library communities and out-of-date Facebook pages" (2011, 367).

The social web is not simply a new fashion; it represents a new way of thinking and doing things. Applied superficially, social web tools will lead to results like those reported by Gerolimos (2011), whose review of the literature about libraries on Facebook and his own study led him to conclude that library Facebook pages are unlikely to stimulate significant interaction between

libraries and their communities. Another study by Gerolimos and Konsta (2011) suggested that except for RSS, academic library implementations of social media tools generated low levels of use and participation.

A source of this failure is that some implementers have used social web tools to merely promote what they are doing, instead of using them to establish and maintain relationships or help their communities do what *they* want to do. Schrier (2011) emphasizes that a successful social media strategy for digital libraries involves becoming "intertwined with the knowledge creation processes relevant to their collections." This approach differs from learning how to use a few social media tools and then adding them to a web interface. Schrier offers a set of five general principles (listening, participation, transparency, policy, and strategy) for planning digital library strategic options and becoming "facilitators of conversations" on the social web. Digital library experiments with social media tools can also benefit from studying what has made social web tools successful in other settings. Such an approach is likely to uncover not only a solid strategy and receptive audience(s) for digital library experiments, but also any technical or organizational barriers to overcome. Gazan, for example, constructed a decision model for examining the prospects of introducing digital library annotation tools (2008).

Being successful on the social web will also require digital library builders to understand, innovate and continue responding to how their communities look for people and information using search engines and social sites, how they share what they create or find, and how information shows up in diverse contexts on the social web, Among other things this involves optimizing the reach and visibility of digital library content.

Optimizing the reach and visibility of digital libraries

Some successful experiments

Digital library managers have successfully experimented with social web techniques to make their collections easier to find and use in an environment where even well-known digital libraries must compete for the attention of online communities. The following list provides two of the numerous examples of early experiments.

- **Flickr.** In 2007 a small team at the Library of Congress began a low-cost pilot project to test ways to increase awareness of historic photographs from the collection. Other project objectives were to gain experience and understanding of the social web, tagging and community interaction. The team chose Flickr (flickr.com; a popular photosharing site) for the experiment (Springer et al. 2008). Their sharing of two collections of 4,615 digitized photographs beginning in January 2008 met with overwhelming positive response. A few months later, the photos had been viewed over ten million times; 79% had been added to Flickr members' personal collections; over 67,000 tags had been added; thousands of comments had been left; and average monthly visits to the LC's own photographs site had risen 20%. Subsequent evaluations demonstrated sustained high community interaction (Bray et al. 2011). The Library learned it could reach new audiences and demonstrate its value to the public using social web approaches.

  Conversations between the Library and Flickr also led to the launch of the Flickr Commons (flickr.com/commons), where many cultural heritage institutions now share public photo collections, including the Smithsonian (Kalfatovic et al. 2008). Bray and others (2011) provide information about the highly positive outcomes achieved by several others participants in the Flickr Commons.

- **Wikipedia.** Motivated by a well-known survey (De Rosa 2005) indicating low use of library websites for starting research, the University of Washington (UW) Libraries initiated a project in early 2006 to have students insert links about the libraries' digital collections into Wikipedia (Lally and Dunford 2007). The objectives were to test this new way to reach out to users and determine if it would drive new traffic to UW web sites. The team chose Wikipedia because it was already a top referrer to the collections. Analysis of their server statistics over the ensuing months indicated that the added links drove a sustained upward climb of traffic from Wikipedia to UW digital libraries. They concluded that Wikipedia is an essential, low-cost tool for making digital library content highly visible outside the library's web pages. Proffitt and Snyder's more recent findings (2012) confirmed that links and images added to Wikipedia can help to build a digital library's reach and visibility.

Libraries and museums have also reported successful experiments with tagging, Facebook, Twitter, YouTube and Pinterest (just a few of the available analyses are Trant 2009a; Thornton 2012; King 2012). These types of experiments by individual libraries have continued in parallel with the emergence of new, large-scale methods and technologies for reaching new audiences and making digital library content more visible in search engines and in many contexts on the social web.

Mobile apps and interfaces

The use of mobile devices is becoming ubiquitous around the world and increasingly there is an expectation that services will provide mobile applications (apps) and interfaces (Purcell 2010). Trends reported in Pew Internet reports suggest that 56% of American adults owned a smartphone, 34% owned a tablet, half reported having apps on their phones and 82% had them on their tablets (Anderson and Rainie 2012; pewinternet.org trend data, device ownership, May 2013). Seeking for online news and information increasingly relies on portable devices (Purcell

2010, slide 66). Mobile is linked to the social web too: a global web research firm reported that mobile is the main factor driving the use of social platforms across all markets (GlobalWebIndex 2013).

Lippincott (2010) examined the implications of increasing ownership of mobile devices for academic libraries and concluded there are both opportunities and challenges for academic libraries that develop mobile applications. Since then, libraries have begun to go mobile: Thomas (2012) reported continued growth in the implementation of mobile services by American public and academic libraries from the baseline captured in 2010, when 34% of American public libraries and 44% of academic libraries reported providing some type of mobile service.

Mobile access to digital libraries

Mitchell and Suchy (2012) examine mobile access to *digital libraries* based on four case studies and found that developing mobile access to digital collections remains in early stages at the time of this writing. Using the EPrints software, Adewumi (2013) built a repository for Covenant University in Nigeria and tested its usability on various mobile devices. Noting that only the Greenstone platform provided a mobile interface to digital repositories at the time, Rosa and others (2012) presented their design process for a mechanism for making DSpace repositories accessible on mobile phones. The intent was to meet a need in regions of the world where widespread access to the internet is not common but use of cell phones is. The paper is well worth consulting as the design process featured the use of surveys and the development of personas to learn the characteristics, needs and expectations of intended users of the mobile interface. A mobile interface was added to the DSpace 3.0 code this year (Tzoc 2013).

A quick literature review of work on mobile interfaces and apps for digital libraries suggests only nascent development in this area. However, given that social media use and information seeking preferences are shifting strongly in the direction of mobile devices, it seems inevitable that digital libraries will eventually want to be able to reach their communities through mobile devices.

SEO and SMO

SEO is important on the social web because the more frequently a site is included the first few pages of search result lists, the more visitors it will receive. The more visitors a site receives, the greater the likelihood that selected content will be tweeted, tagged, bookmarked, shared on a social network, and linked to. This cycle of activity also works in reverse: findings of a number of the experiments discussed in an earlier section of this chapter demonstrated a positive correlation between sharing selected digital library content on social media sites and increased visits to the digital libraries' web sites. Onaifo and Rasmussen's analysis (2013) found that the amount of traffic that a site receives is a factor driving its ranking by search engines, which in turn is a factor driving where information indexed from that site appears in search engine results.

Most digital library managers do not have an SEO strategy for improving the reach and visibility of their digital libraries. The literature about SEO in any library context is quite small. As discussed in chapter 8, Beel, Gipp and Wilde (2010) and Arltisch and O'Brien (2013) provide information on how to utilize SEO techniques to increase the visibility of digital library content in academic search engines. Onaifo and Rasmussen offer the most thorough recent analysis of SEO and the findability of library web pages. Through an evaluation of the findings of a study of Ontario public library web sites, most of which were poorly ranked in search engine results, they

identified which web site characteristics are positively correlated with increased findability and ranking by search engines.

The Onaifo and Rasmussen article also contains a reference to a new set of techniques—social media optimization or SMO, a phrase first introduced in 2006 by Bhargava (2010)—that libraries can use to increase their linkability from social media sites. Onaifo and Rasmussen remark "it is insufficient to use social media simply as a store front, as many libraries do, if the goal is to attract users to the library's website. It is also a good SMO strategy to make it easier for others to tag and bookmark library content, as well as engage with library content through such means as comments and content sharing (within copyright limits)." An updated list of Bhargava's five rules for SMO follow:

1. Create shareable content

2. Make sharing easy

3. Reward engagement

4. Proactively share content

5. Encourage the mashup

It will be an important step in the evolution of digital libraries on the social web for digital library managers to begin to apply SEO and SMO best practices to digital library sites and content.

The semantic web and linked data

Chapter 1 examines the compelling vision that led to the emergence of digital libraries. One section quotes J.C.R. Licklider, an internet pioneer who foresaw the power of human interaction with the body of knowledge "conceived of as a dynamic process involving repeated examinations and intercomparisons of very many small and scattered parts" (1965. 5). The semantic web and linked data have the potential to realize this aspect of Licklider's vision for

libraries of the future. Chapter 3 introduces the semantic web and linked data and explains why these "very many small and scattered parts" could be important to the advancement of knowledge and culture themselves. Chapter 4 discusses the history and current situation for OAI-ORE, a relatively new standard that uses a semantic web approach for describing and exchanging aggregations of web resources (usually scholarly resources). Chapter 5 provides an introduction to library linked data that describes traditional library collections (rather than digital libraries). How to deploy these new technologies and standards in digital libraries is a new grand challenge to the field.

Erik Mitchell (2012) explains why the semantic web and linked data are important to the digital libraries built and maintained by libraries, archives and museums. The semantic web and linked data are important to the social web because they produce open, reusable bits of data that facilitate machine-to-machine interactions, in turn enabling better integration and interoperability of digital library information in other contexts.

Scholarly research and linked data

The semantic web and linked data offer the possibility of "a more data-centric, semantically-linked, and social network-embedded scholarly communication model that resembles the profound changes in social, political, and economic discourse characteristic of Web 2.0" (Van de Sompel et al. 2009). Put another way, the semantic web and linked data offer new opportunities for scholars to share the results of their work in more dynamic, interactive ways.

An examination of the Data Hub (datahub.io) for linked data sets representing scholarly content suggests that scholarly linked data has begun to be available to remix and re-use, especially in the following areas:

- Computer and information science and engineering – examples include linked data sets representing DBLP and ACM publications and IEEE papers and e-research data. The contents of the ECS (Electronics and Computer Science) Southhampton repository are also available as linked data.

- Life sciences – the prominent example is PubMed

- Repositories – examples include the linked data sets representing the contents of EPrints Southampton and several institutional repositories at the Open University in the UK

In addition, Data Hub searches reveal quite a number of e-research data sets, particularly in the areas of the life sciences, chemistry and environmental science, although only a small subset are structured as linked data.

Identifiers, interlinking and linked data

Identifiers are an essential component of the Kahn-Wilensky architecture of digital libraries. Key outcomes of the first decade of digital libraries were a keen understanding of the role of identifiers and their importance for reliably linking between and across web resources and sites (think of DOIs).

The new vision of the web as a semantic, global web of data has renewed attention to URIs (Uniform Resource Identifiers) for both persistently identifying a resource and providing the means to express relationships and link to other resources. The digital library field has continued to contribute substantially to the work to build linked data sets and to further develop the utility of a range of identifiers in linked data. Some examples of this work include:

- CrossRef DOI Resolver – a linked data set that contains URIs for every DOI that CrossRef manages. This supports the use of DOIs in linked data applications.

- Medline – a linked data implementation of 19 million Medline articles, linked to their DOI URIs and journal identifiers.

- VIAF – a representation as linked data that includes URIs for VIAF IDs in the data set. These URIs link to the clusters of multilingual forms of names in multiple cultural heritage institutions' name authority files (e.g., consider the many ways in which different nations express the name of the playwright Anton Chekhov).

- The British National Bibliography (BNB), LIBRIS (Swedish academic library union catalog), and several other national library catalogs – the linked data representations of these link to URIs in VIAF

Digital library linked data

Linked data sets have been seen as a way to make library metadata—not just resource descriptions, but vocabularies and metadata schema—more relevant and interoperable on the web. Digital library linked data is important on the social web because it has the potential to surface digital library content—and its relationships to other content—much more easily on social sites, as well as to support new social web services. So far this potential has not been realized, but the first step is to make digital library linked data available to be used in applications.

There are some linked data sets that represent library-managed digital libraries; one example is Chronicling America (chroniclingamerica.loc.gov). At the time of this writing the most prominent of the digital libraries represented by linked data is Europeana (europeana.eu). Other leaders in the deployment of semantic web and linked data approaches in digital libraries include the Bibliothèque nationale de France (for its Gallica and Data Digital Libraries; see Edelstein et al. 2013) and the University of Alicante in Spain (for its Biblioteca Virtual Miguel de Cervantes).

These digital library initiatives were honored with 2013 Stanford Prizes for Innovation in Research Libraries (Zaino 2013).

Europeana and linked data

Europeana (europeana.eu), the European Union's flagship digital library project, provides a portal to discover cultural heritage materials held in libraries, museums, archives and audiovisual collections across Europe. In early 2013 Europeana was providing discovery services for 26 million objects in cultural heritage institutions in 28 countries (Cousins 2013; Europeana Foundation 2013, 8-9). The Europeana Data Model (EDM) is a framework for ingesting, managing and publishing metadata from its contributing organizations (Isaac 2012). Its basis is in OAI-ORE and the principles of linked data. The EDM is designed so that it can be used by other organizations to structure their metadata. The Digital Public Library of America's model builds on the experience of EDM (DPLA 2012).

Following their work on the EDM specification, the Europeana team soon mounted a project to represent Europeana metadata using the EDM and make the results available as linked data (Haslhofer and Isaac 2011). In 2012 Europeana released the restructured metadata as open linked data, first in a pilot. Later the same year, Europeana released all Europeana metadata as open linked data; at the time of the initial release the data set contained metadata on 20 million objects (Europeana Foundation 2013). This means there is now a substantial and significant body of linked data about European digital library collections to be re-used and re-mixed in other contexts, including social web sites.

Rights issues and digital library linked data

The semantic web has been conceived as a public data commons, open to anyone. Linked data sets function most effectively if they are open and available for re-use with no or minimal

restrictions. "Europeana Terms for User Contributions" (europeana.eu/portal/rights/terms-for-user-contributions.html) specifies that Europeana will make organizations' contributed metadata available under the terms of the Creative Commons CC0 1.0 Universal Public Domain Dedication (CC0; creativecommons.org/publicdomain/zero/1.0/). This means that the contributor is dedicating the metadata to the public domain and waiving all rights to it. The legal issues related to aggregating, exchanging and re-using data from cultural heritage institutions can be complex, especially in an international context. Rights to metadata may be restricted, terms for re-use and exchange may be unclear, or the metadata might be an organization's key business asset (Baker et al. 2011, under 3.3).

Digital library contributions to the semantic web

There is ample reason for digital library researchers and builders to be inspired by the prospects of the semantic web and linked data, and good progress is being made. As noted in chapter 4, next generation repositories and projects related to OAI-ORE are using semantic web approaches. Tools for publishing metadata as RDF are readily available (Byrne and Goddard 2010, Appendix). At this time however, digital library specialists are contributing more to the development of the semantic web by publishing linked data sets than by building applications that use linked data. It will take time for developers inside and outside the digital library space to build applications that consume linked data.

Schema.org

Schema.org (schema.org) is an organization formed in June 2011 through a joint effort of search engine leaders Bing, Google and Yahoo. It offers a collection of metadata element sets that enable webmasters to take advantage of semantic web approaches to structuring data and then use them for SEO and other purposes. A simple explanation of schema.org is that it provides a vocabulary of types of things (movies, books, events, etc.) and uses *microdata* (a

mechanism for embedding structured machine-readable data in HTML documents). For example, the schema.org vocabulary, combined with properties defined by microdata, can alert a search engine that the marked-up section of a web page communicates the name of a person, and what that name is. In that sense schema.org is a *semantic* web approach (that is, it is based on the machine encoding of *meaning*). The major search engines (Google, Bing and Yahoo) "understand" and deploy this structured data to assess relevance and augment the display of search engine results.

Semantic web approaches based on schema.org and microdata are increasingly being used in e-commerce (Hepp 2012). Li, Wald and Wills (2012) describe how these techniques are being used in multimedia applications to expose the inside content of multimedia resources ("media fragments") for indexing by search engines. Examples are part of a YouTube video or part of a music recording.

In the digital library domain, Ronallo (2012) offers an overview, tutorial and examples of schema.org and microdata used to mark up web pages for digital objects in the North Carolina State University Libraries Digital Collections. Ronallo also explains some current limitations of these techniques for the cultural heritage sector. In the library space in general, OCLC is experimenting with embedding structured metadata in each WorldCat.org record based on schema.org (Miller et al. 2012, 34; Breeding 2013b) to promote better discovery of library materials in schema.org participating search engines. The prospects for more digital library applications of schema.org seem good, but at the time of this writing, it is early days for these techniques.

Mass digitization and digital libraries

Chapter 5 discusses mass digitization and a number of other large-scale digitization projects for book collections. There are now millions and millions of digitized books around the world. What does it mean for digital libraries? No one knows. At the time of this writing, except for public domain books, the legal challenges and publisher resistance are preventing the entire impact of the Google Library Project and other initiatives from rolling out. But mass digitization of widely-held materials in libraries has already had a good deal of influence on the digital library landscape for individual libraries.

For many reasons, the role and primacy of huge legacy book collections held by academic research libraries are changing. Special collections and archives are now widely perceived as key assets of research collections, because they are what make research library collections distinctive. Unfortunately, small special collections budgets and other problems constrain the possibilities to pursue high-cost strategies around them (see for example Education Advisory Board 2011, xiii, 78; Maron and Pickle 2013, 2-3). If fresh new approaches to mounting and sustaining projects could be found, given how many of these collections remain hidden (see for example Dooley and Luce 2010), there is the potential for many new digital library initiatives.

Not just new investment, but social web, participatory approaches could help to add new and enhance existing digital libraries, make these collections visible on the web, preserve more unique materials for future generations, and spawn outreach programs to make this digital content accessible in new contexts and to many communities. Whether this actually happens or not remains unpredictable. Continued pressure on library budgets is a hindering factor, but the possibilities of partnerships and increasing public and research interest in digital representations of cultural heritage content, driven by large initiatives such as those in the Netherlands and France and by Europeana and DPLA, are driving factors.

Large-scale digital libraries, portals and platforms

Chapter 2 discusses a sample of 15 working digital libraries that have endured since the start of digital libraries. The oldest is Project Gutenberg (gutenberg.org), which is not only the first digital library, dating from 1971, but also the project that has exemplified strong connections to a community of participants from the beginning (Lebert 2008).

Another one of the sample projects from the first decade of digital libraries is Trove (trove.nla.gov.au; Holley 2010b), which developed a strong community around its digitized newspaper content, whose OCR text accuracy was a concern for the NLA (National Library of Australia). The decision was made to expose the raw OCR text to the public for correction, starting in December 2007. The public's response was immediate and positive beyond any expectation (Holley 2009; 2010a). The project is now famous for crowdsourcing work on a large-scale digitization project. Crowdsourcing, discussed in the next section, is a massive collaboration technique that enables individuals, working as a virtual group, to collectively accomplish a shared, large and significant goal.

The implementers of two other early projects, Gallica, from the Bibliothèque nationale de France (gallica.bnf.fr) and American Memory, the flagship service of the US National Digital Library Program (memory.loc.gov), launched traditional read-only digital libraries but made unanticipated discoveries about the communities that engaged deeply with their content and services. The BnF's 2002 BibUsages project studied the usage and users of Gallica and concluded that "digital libraries, far from being simple digital versions of library holdings, are now attracting a new type of public, bringing about new, unique and original ways for reading and understanding texts" (Assadi et al. 2003). In the course of the interviews I conducted for this book, I learned that American Memory was initially targeted for professors and others in

university settings; a pre-test revealed the true primary audience to be grade school and high school teachers. An interviewee for this book recounted her insight that thanks to the end-user test, "teachers developed curricular ideas and shared them. As a result, many years later, many digital resources continue to be used, discussed and remembered."

Europeana (europeana.eu) has begun experimenting with projects to engage ordinary citizens, scholars, teachers and children with its massive, cross-cultural and multilingual digital library. In March 2011 it launched the highly interactive Europeana 1914-1918 (europeana1914-1918.eu/en), which allows people to connect their stories and memorabilia to Europeana. The online social features are supplemented with road shows; the ones in Germany alone resulted in 25,000 artifacts being scanned (Charlton 2012) and added to the collection. Another community engagement project, in partnership with the Digital Public Library of America (DPLA), focuses on Europeans immigrating to America (Berkman Center 2011).

Europeana's leaders have conceived the prorgram's public mission in the context of the social web. In a 2012 interview, Jill Cousins, executive director of Europeana, said "The whole Europeana concept is not about creating a destination site in Europeana.eu but about distributing the aggregated data into other systems, mobile applications and so forth so that the content can be used in many different ways and sustain different ways of looking at the material – e.g. in higher education and schools" (Cranfield 2012).

The Digital Public Library of America, or DPLA (dp.la) is intended to be "a digital library in service of the American public" (Cranfield 2012) bringing digital content from many sources together in one platform. DPLA launched in April 2013, having recently hired Dan Cohen, an historian and well-known leader of social web-inspired initiatives at the Center for History and New Media, as DPLA's founding Executive Director. In an interview shortly after the hiring was

announced, Cohen remarked that "successful digital projects mainly involve getting diverse people working together towards an ideal." Cohen envisons DPLA as a portal that "will bring entirely new audiences to formerly scattered collections" and "a large open storehouse for classroom use and scholarly investigation" (Enis 2013).

Crowdsourcing and citizen science

Crowdsourcing background

This section is a brief treatment of a large and significant subject; it offers some background, some sources for further study, and a few examples. William Safire (2009), the well known columnist, explored the origins of the word "crowdsourcing" in a piece for the *New York Times Magazine.* He reported that Jeff Howe of *Wired* magazine had coined the term in 2005 and wrote about it in 2006 (Howe 2006). An examination of these sources and several others reveals that crowdsourcing arose in the business sector as an innovative way to outsource work to ordinary people instead of employees or contractors.

Daren Brabham (2008), a researcher who wrote his dissertation on crowdsourcing and the collective intelligence of online communities, explains why crowdsourcing has had so much success in the business community. He concludes that crowdsourcing is "a model [enabled by the web] capable of aggregating talent, leveraging ingenuity while reducing the costs and time formerly needed to solve problems" (87). Brabham's dissertation (2010) treats the topic of crowdsourcing much more broadly and thoroughly. Many other types of organizations outside the business world rapidly embraced crowdsourcing. Of course, Wikipedia is the epitome of crowdsourcing and the power of the social web (and it predates the coining of both terms).

Digital library crowdsourcing

In the digital library space, the promising results of Europeana 1914-1918 have already been mentioned as well as the digital program for historic newspapers of the National Library of Australia's Trove service. Another noted experiment is the steve.museum project, which found crowdsourcing to be successful for enhancing access (through tagging) and creating alternative vocabulary to museum documentation (Trant 2009b).

Holley's article on crowdsourcing (2010a) discusses several other organizations' successful experiments with the technique. Two of the six projects she discusses are FamilySearch Indexing and Distributed Proofreaders. FamilySearch Indexing (familysearch.org/volunteer/indexing) crowdsources the indexing of family history records and make the results freely available. Volunteers had indexed over a billion family history records as of the time of this writing. Project Gutenberg, the first digital library, fittingly inspired one of the earliest crowdsourcing projects, which dates from 2000 and is called Distributed Proofreaders (pgdp.net/c/). As of this writing, Distributed Proofreader volunteers had helped convert nearly 26,000 public domain titles into freely available e-books.

Using amateurs to address large scale professional or technical challenges was once considered to have no chance of succeeding. In the digital library arena, Gregory Crane of the Perseus Digital Library has documented the challenges of extracting geospatial data from a very large number of unstructured historical textual sources (2004). The Perseus Digital Library recently began experimenting with crowdsourcing (Davis 2012). Fleet, Kowal and Přidal (2012) describe other crowdsourcing efforts related to digitized historical maps—an online initiative to crowdsource the georeferencing of historical map images using the Georeferencer application. Five institutions had implemented projects as of late 2012: the Moravian Library (Brno), the Nationaal Archief (The Hague), the National Library of Scotland (Edinburgh), the British Library

(London), and the Institut Cartografic de Catalunya (Barcelona). At each location, the public's online participation significantly increased the number of historical maps that were georeferenced. At the Institut Cartografic de Catalunya, for example, 1,000 early printed and manuscript maps and aerial photographs were all georeferenced in 24 days by 88 volunteers.

Citizen science

Experiences with crowdsourcing have contributed to the rapid development of citizen science initiatives. Andrea Wiggins and Kevin Crowston have been reporting on the impressive work being at the Syracuse University School of Information Studies to study and understand citizen science, which they define as a form of research collaboration involving members of the public in scientific research projects that address real-world problems.

Wiggins and Crowston (2011, 2012) present their analyses and categorization of citizen science initiatives, concluding "under the right circumstances, citizen science can work on a massive scale, generating high quality data that lead to reliable, valid scientific outcomes as well as unexpected insights and innovation" (2012, 3426). They particularly note the efficacy of virtual (web-based) citizen science, such as the well-known Galazy Zoo (galaxyzoo.org) project, which crowdsources the classification of images of galaxies. Over three years in the project's life, over 250,000 volunteers participated in the classification of over 56 million galaxies (2011, 7).

Crowdsourcing and citizen science may be the most important social web phenomena for digital library specialists to watch, as there are many potential ideas and applications that would be worth considering as digital libraries move into the future.

Conclusion

The social web is an emergent, highly chaotic space. Social web initiatives related to the digital library field are equally chaotic, representing the convergence of many overlapping, parallel or directly competing efforts. More contenders for the attention of a digital library researcher or professional seem to enter this confusing new space every week. One of the digital library leaders interviewed for this book said "this space is a mess at the moment. Many projects will fail. It is hard to predict which ones will succeed."

In the evolution of digital libraries toward new roles on the social web, existing conventions and supporting systems are already undergoing significant disruption, and while the disruption is painful, it also presents new opportunities for digital libraries to go beyond what they have achieved in their first two decades. While these achievements are impressive, they will not guarantee future success, especially in the fast-moving world of the social web. The best way to honor digital libraries' past is to participate in creating their future.

Chapter 5 of this book makes a case for merging hybrid library and digital library strategic agendas; it simply makes no sense to continue separate, parallel lines of development. Chapter 6 makes a case for evolving digital library research and practice by shifting the primary focus toward digital libraries' social roles, in particular how they might empower individuals and support the progress of knowledge, learning, the free flow of ideas and an informed citizenry. Marilyn Deegan and Simon Tanner (2002, 216-217), in an eloquent passage in their book *Digital Futures*, write "librarians should redefine the profession, not in terms of the collections we hold, but in terms of the skills, abilities and value we bring to our communities … the time has come for us not to be defined by infrastructure." Infrastructure can be buildings, collections, or enabling technologies. Libraries' and librarians' social roles and communities are more likely to abide over time; collections and enabling technologies are more likely to shift.

Derek Law, writing of digital developments in the library landscape, speaks of the urgent need for an "overarching philosophical redefinition of what libraries should be" (2011, 374). To emphasize his point about the trouble that libraries are having staying relevant and viable in a world dominated by the web, Law rewords quotes from Clay Shirky's famous blog post on the demise of the newspaper industry (Shirky 2009). Law's version replaces some of Shirky's words with words related to libraries (362-363). I have followed Law's lead but supplied some of my own rewording:

> Society doesn't need ~~newspapers~~ **libraries.** What we need is ~~journalism~~ **knowledge and understanding** … When we shift our attention from "save ~~newspapers~~ **libraries**" to "save society," the imperative changes from "preserve the current institutions" to "do whatever works." And what works today isn't the same as what used to work … No one experiment is going to replace what we are now losing with the demise of ~~news on paper~~ **libraries as they have been**, but over time, the collection of new experiments that do work might give us the ~~journalism~~ **libraries and librarians** we need.

R. David Lankes has written and spoken widely on the importance of a new worldview that is free of assumptions carried forward from libraries' past successes and librarians' traditional roles. He is well known for his conceptualization of "participatory librarianship" and for his emphasis of librarians' societal roles:

> The mission of librarians is to improve society through facilitating knowledge creation in their communities (Lankes 2011).

Lankes has provided new conceptual models not only for librarianship, but also for libraries as community platforms. While the models are intended for redefining libraries generally, they are equally useful for defining the next steps for digital libraries' participation in the social web.

Most digital libraries continue to operate from a traditional, collections-centered service model. Change will be difficult, especially in the realm of scholarly practices and norms, where the roots of tradition are deep. Lankes and Law offer clear-eyed appraisals of current prospects and new models for rethinking what libraries and digital libraries should and can do for a society that is now so dominated by the social web. Their work and that of the hundreds of other people cited in this book provide ample reason for optimism.

Internet pioneer J.C.R. Licklider predicted long ago that the "libraries of the future" may not resemble libraries as they have been. Digital libraries are moving to the mainstream, but for them too, the future may not look much like the past. The social web opens the door to new possibilities, but experimenting superficially with social media will not increase libraries' or digital libraries' value in a globally networked world. Community-centered strategies, aligned with the large changes shaping the web and society, are required.

The first grand vision of digital libraries inspired two decades of digital library research and practice that have been, and continue to be a powerful force for advancing the pursuit of knowledge and culture. Emulating the creativity and pioneering spirit of digital libraries' first twenty years is the starting point for creating the next grand vision for libraries' digital future.