Foreword

> The time is close at hand when any student, in any part of the world, will be able to sit with his projector in his own study at his or her own convenience to examine any book, any document, in an exact replica
>
> **H.G. Wells, 1938. (p. 77).**

The above quote from H. G. Wells, which also appears in the first chapter of the book, is appropriate because it foresees digital libraries, even without mentioning them by name. In addition, in 2016, when this book was published, we celebrate 150 years since Wells' birth and mourn 70 years since his death.

This book identifies the challenges, current trends, and future directions of digital library development, use, and evaluation. The coverage of the book is comprehensive, as can be easily determined by each chapter and the accompanying bibliographies.

This Foreword has a hazier aim—I am trying to examine the complex relations and connections between research and practice in the area of digital libraries:

- *Does digital library research inform digital library practice? And vice versa?*
- *To what extents are they connected now, nearly two decades after they began?*

"Digital library research" refers to various research projects funded historically by the National Science Foundation (NSF) in the United States and by European Union programs in Europe, as well as research reports presented at various digital library conferences to this day. "Digital library practice" includes working digital libraries found on the Web, reflecting any practical, operational library-oriented achievements. "Inform" refers here to a visible connection based on evidence either (1) in the sites of research projects and/or in the research literature that points to any consideration of or link to an operational digital library project, or (2) in digital library practice showing any consideration of or link to research projects found in the literature. In other words, concentration here is solely on visible or "surface" evidence.

In many fields, research and practice have a complex relationship or connection. In an ideal paradigm, (some) research, particularly toward the applied end, informs and even transforms practice and (some) practice informs research, especially in the selection of problems. In an ideal world, research and practice converge. However, in reality, it rarely works exactly that way. The links between research and practice are neither always linear nor are they often easy to discern. Their connections may be serendipitous or vague, even weak or nonexistent. Time and social context play a significant role as well. Transfer of ideas is complex. There are further considerations. Research often raises expectations, and, by definition, it neither promises nor produces predictable outcomes. Practice may advance, and in many areas has done so, without direct input of research. The area of digital libraries is not an exception to any of these points.

Historically, in the United States, the original agenda for digital library research was set and conducted through the multiagency Digital Library Initiatives (DLIs) lead by the National Science Foundation (NSF). DLI 1 (1994–98) involved six awards and $25 million; DLI 2 (1999–2003) involved 34 awards and $48 million (Griffin, 2005). It may be of interest to note that the origins of Google were acknowledged as being supported by a DLI 1 grant at Stanford University (National Science Foundation, 2004).

While the agendas for both DLIs were relatively broad, their base rested firmly in technology. Levy (2000) describes this early conclusion regarding DLI projects as reported in the research literature or at various library conferences:

> "The current digital library agenda has largely been set by the computer science community, and clearly bears the imprint of this community's interests and vision. But there are other constituencies whose voices need to be heard."

That indeed there are different viewpoints about digital libraries was also recognized half a decade later by Arms (2005). Both conclusions still hold today.

Since the end of DLI 2, NSF no longer supports research in digital libraries. All federal US support is now channelled through the Institute of Museum and Library Services (IMLS), established by law in 1996 combining the Institute of Museum Services and the Library Program Office from Department of Education. The very name (including museums) shows a shift and broadening in the coverage of the area of support. The majority of grants are for pragmatic operational projects—some involving digital libraries—as is clearly visible in the enumeration of their past grants (https://www.imls.gov/grants/awarded-grants).

Again historically, in Europe, digital library research was supported by the European Union in two programs: DELOS: Network of Excellence on Digital Libraries (2004–07, at a cost of 950,000 EUR) and DL.org: Digital Library Interoperability, Best Practices and Modelling Foundations (2008–11, at a cost of 1,200,000 EUR). Both projects involved a large number of participants and workshops across Europe. Crowning achievements for both projects were two landmark publications: The DELOS Manifesto (Candela et al., 2007) and The Digital Library Reference Model (DL.org, 2010). The Manifesto lays out basic definitions and structure of digital libraries. The Reference Model provides detailed descriptions of concepts involved, together with models of various applications and domains of digital libraries. Both documents consider digital libraries as "the meeting point of many disciplines and fields, including data management, information retrieval, library sciences, document management, information systems, the web, image processing, artificial intelligence, human-computer interaction, and digital curation." (Candela et al., 2007).

The US DLI projects were primarily oriented toward development of applications—testbed/prototype building. The European digital libraries projects were primarily oriented toward conceptual definitions and Europe-wide cooperation among researchers. However, government support for digital library research has ended in both the US and European Union.

Digital library practice is institutionally/organizationally based and oriented toward a given community, pragmatic development, and practical operations. As expected, the aims are toward the pragmatic problems at hand. Typical examples—rather than exhaustive enumerations—involve the following:

- Digitizing and providing access to specialized and historic materials by the Library of Congress through the *American Memory Project of the Library of Congress* (launched in 1990, first on CD-ROM and then in 1996 on the Web) (https://memory.loc.gov/ammem/index.html).
- Incorporating digital dimensions and providing access to electronic collections and resources, with a variety of associated services (ie, creating and managing hybrid libraries—incorporating both traditional and digital materials and services) by thousands of academic, research, public,

and special libraries, such as the *University of California Berkeley Library* (http://www.lib.berkeley.edu/node).

- Building digital libraries by professional and other organizations, such as the subscription-based Association for Computing Machinery (ACM) *ACM Digital Library* (launched in 1998) (http://dl.acm.org/).
- Developing digital collections in specific domains, such as the *Perseus Digital Library*, covering digitized materials from antiquity to the Renaissance (launched on CD-ROM in 1987—four years before the advent of the Web; it was later established on the Web in 1995) (http://www.perseus.tufts.edu/hopper/).
- Developing and building huge continent- or nation-wide portals that act as an interface to many cultural institutions, providing millions of books, paintings, films, museum objects, archival records, and music and sounds that have been digitized. Examples include: (1) *Europeana* (launched in 2008), under a subtitle "think culture" calls to "explore 48,738,306 artworks, artifacts, books, videos and sounds from across Europe." (2) *Digital Public Library of America* (DPLA) (launched in 2013), under a subtitle "a Wealth of Knowledge" calls to "explore 11,425,950 items from libraries, archives, and museums." (Quotes with given numbers from both sites are as of Jan. 2016) (http://www.europeana.eu/portal/) (http://dp.la/).

These are just a very few examples of digital and hybrid libraries. They represent only a sliver of an explosive growth that resulted in a multitude of practical digital libraries worldwide.

Practical efforts in digital libraries share a common characteristic. Agendas were set at grassroots— by individual libraries, academic departments, professional organizations, museums, publishers— often driven by enthusiastic individuals. Pioneering projects from the early 1990s, such as those at the Library of Congress mentioned above, served as examples for a great many institutions to follow. Development of digital collections, management and preservation of digital resources, user-oriented services, electronic publishing—with myriad issues and challenges beyond technology—are also a part of these pragmatic efforts.

As mentioned at the beginning of this Foreword, evidence analyzed here is based either in (1) sites of research projects that show a direct connection to practice or in (2) practical sites of digital libraries that show a direct connection to research; in other words, it is based solely on evidence that is directly visible or on "surface" evidence.

Here is a sample of the literature on digital libraries. *Communication of the ACM* (CACM) is the flagship journal of the Association of Computing Machinery (ACM). CACM had three special issues devoted to digital libraries [CACM vol. 38 (4) 1995; vol. 41 (4) 1998; and vol. 44 (5) 2001]. In the most recent special issue on digital libraries, the article "The ACM Digital Library" stated:

> "ACM distinguished itself in its advanced planning for its digital library by paying close attention to three such issues: changing patterns in scholarly behavior; the functioning of copyright law in a networked environment; and the development of a business model" (Rous, 2001, p. 90).

No research project was mentioned or cited in the article, even though this and other two CACM special issues on digital libraries contained other articles about DLI projects that involved specific digital libraries.

An article by Candela et al. (2007) about the Digital Library Manifesto has 17 references; none of them cite an operational digital library project. In turn, (as of Jan. 2016) the article is cited in Scopus (the largest bibliographic database covering over 22,000 titles) 31 times. A cursory examination of titles and references in those citations did not find a connection to any operational digital library project.

In contrast, the article by Bearman (2007) (among the best, if not even the best review of digital library literature up to that time) has 311 references, many of which refer to practical digital libraries. In a section devoted to practical systems, Bearman classifies them as to "Discipline- and Subject-Based Digital Libraries; Genre- and Format-Based Digital Libraries; Institutional Repositories; and Mission- and Audience-Directed Digital Libraries." The last one included children's digital libraries. This is the only article found that had descriptions of and references to practical digital libraries.

There are numerous international conferences devoted to digital libraries:

- Joint Conference on Digital Libraries (JCDL); since 2001 cosponsored by ACM and Institute of Electrical and Electronics Engineers, Computer Society (IEEE-CS), thus "joint" in the title— before 2001 they were separate.
- International Conference on Theory and Practice of Digital Libraries (TPDL); started in 1997 as the European Conference on Research and Advanced Technology on Digital Libraries (ECDL).
- International Conference on Asian Digital Libraries (ICADL).
- In addition, many other—including regional—conferences are listed in each issue of D-Lib Magazine (http://www.dlib.org/groups.html).

In other words, conferencing on digital libraries is a rich, international tradition and venue. All the major conferences publish conference proceedings, of which only two Proceedings of the Joint Conference on Digital Libraries (JCDL) are examined here in some detail: the first one that was jointly held in 2001, and the most recent one held in 2015.

Papers and presentations in JCDL 2001 are simply listed without a particular categorization. The first paper was entitled "Integrating automatic genre analysis into digital libraries," and the last was "The virtual naval hospital: the digital library as knowledge management tool for nomadic patrons." Over 80 contributions are listed; a good number are only one page long, consisting only of an abstract and references with no actual paper included.

Papers and presentations in JCDL 2015 are divided into categories labeled by sessions: "People and Their Books; Information Extraction; Big Data, Big Resources; Working the Crowd; User Issues; Ontologies and Semantics; Non-text Collections; Temporality; and Archiving, Repositories, and Content." There were 18 full and 30 short research reports. Numerous examples of applications were given, such as a papers with the title: "No More 404s: Predicting Referenced Link Rot in Scholarly Articles for Pro-Active Archiving;" or "iCrawl: Improving the Freshness of Web Collections by Integrating Social Web and Focused Web Crawling."

A random examination of papers in both JCDL conferences, particularly as to their citations, yielded no reference to a practical digital library. However, a large proportion made specific and practical suggestion how to handle digitally given objects (eg, sounds) or did research examining a process (eg, digital reading).

In sum, papers at these conferences represent an impressive diversity of efforts in digital libraries. As for authors, these conferences mainly represent efforts coming out of the computer science community and provide a minimal connection to efforts involving broader communities. While the proportion of authors outside computer science is rising, less than 20% of all authors during these years comes from outside the discipline.

Time for conclusion. A brief answer posed by questions at the outset is this:

> As it stands now, I believe that digital library research on the one hand, and digital library practice on the other, reside in parallel universes with little visible contact and intersection, as demonstrated by the diffusion channels examined here. I think that, while they are both about digital libraries, there is a digital divide between them. At present, the two communities disseminate ideas in detached formal networks of communication that are more or less self-referential. However, things and connections may change.
>
> In other words, I believe that presently, digital library research and digital library practice are conducted mostly independent of each other, minimally informing each other, and having slight or no connection.
>
> Furthermore, I also concur with David Levy's conclusion, quoted above, that the research agenda largely bears the imprint of the computer science community's interests and vision.

However, since both research and practice are in progress and the diffusion process is a function of time, we may expect changes. The approach and method adopted has obvious limitations—I took the information provided "as is" and did not pursue any deeper analysis of connections, if any, below the surface. Here are a few more limitations. It is well known that technology transfer may take place through informal as well as formal channels and records. This indeed has been common in the digital library field, a fact that cannot be ignored. Many people that are designing and developing digital libraries have attended digital library conferences and learned of research work. Likewise, a good percentage of those attending digital library conferences are practitioners who bring back to their libraries and projects what they have learned from research presentations. Further, invited talks, panel discussions, short papers, posters, and workshops are key parts of conferences where technology transfer takes place in both directions, and these have been ignored in the present analysis. Thus, conclusions here can easily be questions. Still, the subject in the leading questions should be raised.

This book covers large issues facing digital libraries. It incorporates current research perspectives on the development and evaluation of digital libraries, as well as an overview of best practices and standards for high performance. This combination of research and up-to-date practical guidelines is a unique strength of this book.

T. Saracevic

School of Communication and Information, Rutgers University, New Brunswick, New Jersey, United States

REFERENCES

Arms, W.Y., 2005. A viewpoint analysis of the digital library. D-Lib Mag. 11(7–8). Available from: http://dlib.org/dlib/july05/arms/07arms.html.

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

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

DL.org., 2010. The Digital Library Reference Model. (Also contains the DELOS Manifesto). Available from: http://www.dlorg.eu/uploads/DL%20Reference%20Models/The%20Digital%20Library%20Reference%20Model_v1.0.pdf.

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

Levy, D.A., 2000. Digital libraries and the problem of purpose. D-Lib Mag. 6(1). Available from: National Science Foundation. (2004). On the origins of Google. Available from: http://www.nsf.gov/discoveries/disc_summ.jsp?cntn_id=100660.

National Science Foundation (2004). On the origins of Google. Available from: http://www.nsf.gov/discoveries/disc_summ.jsp?cntn_id=100660&org=IIS.

Rous, B., 2001. The ACM Digital Library. Commun. ACM 44 (5), 90–91.

Wells, H.G., 1938. The World Brain. Doubleday, Doran & Co, Garden City, NY.