

Digital Libraries and Their Technologies: A Look Back and a Peek Forward

Nicholas J Belkin

School of Communication & Information

Rutgers University, Emeritus

What is/could be a Digital Library

Early Concepts of “Libraries of the Future”

- H.G. Wells’s World Brain (1938)
 - World encyclopedia
- Vannevar Bush’s MEMEX (1945)
 - Constructing and sharing associative “trails” from document to document to mimic “As we may think”
- J.C.R. Licklider +’s Libraries of the future (1965)
 - Procognitive systems, not libraries
- Manfred Kochen’s The growth of knowledge (1967)
 - Extending the goals of information retrieval to impart understanding

H.G. Wells's World Brain (1938)

- A World Encyclopaedia no longer presents itself to a modern imagination as a row of volumes printed and published once for all, but as a sort of mental clearing house for the mind, a depot where knowledge and ideas are received, sorted, summarised, digested, clarified and compared.
- It would develop a directorate and a staff of men of its own type, specialised editors and summarists.
- This Encyclopaedic organisation need not be concentrated now in one place; it might have the form of a network
- It would constitute the material beginning of a real World Brain.
- By means of microfilm, the creation of an efficient index to all human knowledge, ideas and achievements, to the creation, that is, of a complete planetary memory for all mankind.

Vannevar Bush's MEMEX (1945)

- Microfilm as the means for production and distribution of the corpus
- Speech to text as the method for producing new documents
- Unspecified means for meaningful addition of new documents to the corpus
- Manually produced associative indexing of the corpus
 - By individuals, of their own corpus
 - Pooled, for indexing the entire corpus

Licklider's Procognitive Systems

- Limited to “transformable” information
- Criteria that such systems must meet (in part)
 - Be available when and where needed;
 - Handle both documents and “facts” (i.e. items of information derived from one or more documents without regard to the form of the source passage);
 - Permit various categories of input to the body of knowledge;
 - Make available a body of knowledge organized broadly and deeply;
 - Facilitate its own development;
 - Converse or negotiate with the user while formulating requests and responding to them;
 - Permit user interaction with metadata and source information;
 - Facilitate joint contribution of users to, and use of the body of knowledge.

Kochen's Growth of Knowledge (1967)

- Alternative goal for information retrieval: supporting understanding rather than merely dispensing information
- Newly generated knowledge welded into a coherent whole
- A move from supporting access, to providing evaluation and synthesis
- Book brings together an excellent collection of readings supporting this position, including an excerpt from Wells (1938) and the article by Bush (1945).

Common elements of these proposals

- Universal, remote access to the store of knowledge
- Direct access to the contents of the documents in the store of knowledge
- Providing at least two levels of representation of knowledge, the original documents, and some representation constructed from the contents of the original documents
- Meaningful linking of documents to one another, and to the constructed representations
- Support for addition and integration of new information into the store of knowledge
- Support for use of interaction with the store of knowledge in resolution of the reason for interaction with it

Traditional functions of the library

- Preservation and maintenance of the cultural heritage
 - *Cultural heritage* defined broadly to mean the declarative knowledge, art, crafts and traditions of any social community
- Supporting access to, and understanding and use of, information in the cultural heritage
- Supporting addition to the cultural heritage
- Supporting communities of users of the cultural heritage

Early DL organizational history

NSF/NASA/DARPA Digital Library Initiative

- Informal series of meetings/workshops in late 1980s
 - In response to research/practice in the field
- 1993, formal announcement of Phase 1 of the Joint DLI
 - Six five-year projects funded
- 1998, Phase 2 announced
 - Broader funding opportunities

Early DL meetings history

- Late 1980s – early 1990s, invitational workshops funded by NSF, DARPA and NASA
- March, 1994 IEEE CAIA'94 Workshop on Intelligent Access to On-Line Digital Libraries was held as part of the Tenth IEEE Conference on Artificial Intelligence for Applications (CAIA'94)

More early DL meeting history

- 1994 – 1995 First and Second Annual Conference on the Theory and Practice of Digital Libraries
 - Independently funded
 - Morphed into
- 1996 – 2000 ACM SIGIR & SIG LINK(WEB) International Conference on Digital Libraries (DL '96 to DL 2000)
- 1994 – 2000 IEEE-CS Forum on Research and Technology Advances in Digital Libraries (ADL '94 to ADL '2000)
- 2001 DL and ADL merged to form the ACM IEEE Joint Conference on Digital Libraries

Still more early DL meeting history

- 1997 – 2010 European Conference on Digital Libraries, initially organized as a part of the first DELOS Working Group, subsequently partially funded by European Union, through establishment of the European Research Community on Digital Libraries
 - 2011 renamed the International Conference on Theory Practice of Digital Libraries
- 2000 – Libraries in the Digital Age (LIDA) established at the InterUniversity Centre in Dubrovnik
- 1998, the International Conference on Asia-Pacific Digital Libraries (ICADL) began its annual continuing series

Research and practice responses to
technologies enabling DLs

Late 1980s

Saw the beginnings of a shift from print to digitization

- OPACs
 - Remote access
 - Search
- MARC standard
 - Sharing of data among libraries
- Storage technologies
 - Digital collections on CD-ROM and other devices
 - Sharing and search

Early 1990s

- High speed scanners & OCR
 - Supported the development of local, stand-alone DLs
- Software distribution techniques
 - Supported the development of tools/packages for building DLs
- Maturation of hypertext & of image and video technology
 - Supported the development of multimedia, highly intralinked DLs
- Major focus was development of increasingly sophisticated local DLs

Mid to late 1990s

- DLI Phase 1 Projects initiated in 1994
- Six five-year projects
- Collaboration amongst recipients
- Explored the use of a variety of technologies for a variety of advanced DL functions
- Emphasis on building complete systems, with defined community of users

DLI Phase 1 projects

- **University of California, Berkeley** – *Environmental Electronic Library*
A prototype scalable, intelligent, distributed electronic library focused on environmental information. Collected and organized diverse environmental data for analysis, reporting, and decision-making, with eventual deployment in California's CERES systems.
- **University of California, Santa Barbara** – *Alexandria*
Developed a scalable, distributed digital library platform for heterogeneous resources, emphasizing interoperability, resource federation, and user-friendly access.
- **Stanford University** – *Uniform Access*
Researched interoperability, protocols, standards, and distributed object architectures to enable uniform access to heterogeneous distributed information resources.
- **Carnegie Mellon University** – *Digital Video Libraries*
Focused on full-content search and retrieval of video segments, integrating speech, image, and natural language technologies.
- **University of Illinois at Urbana-Champaign** – *GRAINGER*
Explored new models for distributed documents, semantic interoperability, and strategies for rights management in digital collections.
- **University of Michigan** – *Intelligent Agent Architectures*
Investigated software agents, resource federation, artificial service market economies, and their educational impact.

Mid to late 1990s, cntd

- **The World Wide Web (WWW) and HTTP**
 - Initially difficult to incorporate into DL functions, but with increased capabilities, incorporated to support wide-spread access to many varieties of content, including user-generated
 - Linking to used to promote retrieval by association, as well as to rank retrieval results
- **High-Speed Scanners & OCR**
 - Maturation of these technologies were used to shift digitization of images to readable, and therefore searchable texts
 - This led to the ability to apply sophisticated retrieval technology
- **TREC program**
 - Results of this IR program began to be incorporated in DLs

DLI Phase 2 projects

- **High-Performance Digital Library Classification Systems: From Information Retrieval to Knowledge Management** University of Arizona, <<http://ai.bpa.arizona.edu/go/dl/>>.
- **Re-inventing Scholarly Information Dissemination and Use** University of California, Berkeley, <<http://elib.cs.berkeley.edu/>>
- **A Multimedia Digital Library of Folk Literature** University of California, Davis, <<http://philo.ucdavis.edu/SEFARAD/>>
- **Alexandria Digital Earth Prototype** University of California, Santa Barbara, <<http://www.alexandria.ucsb.edu/adept/adept.html>>
- **Informedia II: Auto-Summarization and Visualization Over Multiple Video Documents and Libraries** Carnegie Mellon University, <<http://www.informedia.cs.cmu.edu/dli2/>>
- **Simplifying Interactive Layout and Video Editing and Reuse** Carnegie Mellon University, <<http://www.cs.cmu.edu/%7Esilver/#About%20SILVER>>
- **Digital Library for Human Movement**
- **A Patient Care Digital Library: Personalized Search and Summarization over Multimedia Information** Columbia University, <<http://www.cs.columbia.edu/diglib/PERSIVAL/>>

DL I Phase Two projects

- **Project Prism at Cornell University: Information Integrity in Digital Libraries** Cornell University, <<http://www.prism.cornell.edu/>>
- **Digital Analysis and Recognition of Whale Images on a Network (DARWIN)** Eckard College, <<http://pythagoras.eckerd.edu/darwin>>
- **An Operational Social Science Digital Data Library** Harvard University, <<http://www.thedata.org/>>
- **A Distributed Information Filtering System for Digital Libraries** Indiana University Indianapolis/Bloomington, <<http://sifter.indiana.edu/>>
- **Digital Workflow Management: The Lester S. Levy Digitized Collection of Sheet Music, Phase Two** Johns Hopkins University, <<http://levysheetmusic.mse.jhu.edu/>>
- **The Digital Atheneum: New Techniques for Restoring, Searching, and Editing Humanities Collections** University of Kentucky, <<http://www.digitalatheneum.org/>>
- **Founding a National Gallery of the Spoken Word** Michigan State University, <<http://www.ngsw.org/>>
- **Tracking Footprints through an Information Space: Leveraging the Document Selections of Expert Problem Solvers** Oregon Health Sciences University, <<http://www.cse.ogi.edu/dot/research/footprints/>>

DLI Phase 2 Projects

- **Data Provenance** University of Pennsylvania, <<http://db.cis.upenn.edu/Research/provenance.html>>
- **A Software and Data Library for Experiments, Simulations, and Archiving** University of South Carolina, <<http://econ.badm.sc.edu/beam/>>
- **Stanford Interlib Technologies** Stanford University, <<http://www-diglib.stanford.edu/>>
- **Image Filtering for Secure Distribution of Medical Information** Stanford University, <<http://www-db.stanford.edu/pub/gio/TIHI/TID.html>>
- **A Digital Library of Vertebrate Morphology, Using High-Resolution X-ray CT** University of Texas at Austin, <<http://www.ctlab.geo.utexas.edu/dmg/index.html>>
- **A Digital Library for the Humanities** Tufts University, <<http://www.perseus.tufts.edu/>>
- **Automatic Reference Librarians for the World Wide Web** University of Washington, <<http://www.cs.washington.edu/research/diglib/>>

DLI Phase 2 projects

- **Undergraduate Emphasis**
- **Using the National Engineering Education Delivery System as the Foundation for Building a Test-Bed Digital Library for Science, Mathematics, Engineering and Technology Education** University of California, Berkeley,
<<http://www.needs.org/>>
- **Columbia Earthscape: A Model for a Sustainable Online Educational Resource in Earth Sciences** Columbia University,
<<https://wwwc.cc.columbia.edu/sec/dlc/earthscape/>>
- **Research on a Digital Library for Graphics and Visualization Education** Georgia State University,
<<http://asec.cs.gsu.edu/asecdl-nsf-dli2/index.html>>
- **Digital Libraries for Children: Computational Tools that Support Children as Researchers** University of Maryland,
<<http://www.cs.umd.edu/hcil/kiddiglib/>>
- **A Digital Library of Reusable Science and Math Resources for Undergraduate Education** University of North Carolina, Wilmington.
- **Planning Grant for the Use of Digital Libraries in Undergraduate Learning in Science** Old Dominion University,
<<http://dlib.cs.odu.edu/>>
- **The JOMA Applet Project: Applet Support for the Undergraduate Mathematics Curriculum** Swarthmore College,
<http://forum.swarthmore.edu/joma_applet/>
- **Virtual Skeletons in Three Dimensions: The Digital Library as a Platform for Studying Anatomical Form and Function** University of Texas at Austin, <<http://www.eskeletons.org/>>

2000s

- Standards for metadata, such as Dublin Core and OAI-PMH
 - Support for communication and integration among and across DLs
- Even better scanning technologies, used by, e.g. Hathi Trust and Google Books
 - Greatly increased size and breadth of available corpus
- Human-computer interaction and interactive information retrieval technologies and techniques
 - Better interface design
 - User-centered evaluation

2010s

- **Cloud Infrastructure technologies**
 - Libraries migrated their infrastructure to cloud environments not only to save costs, but also to increase access
- **Linked Open Data**
 - Frameworks like RDF allowed metadata to become interconnected, web-searchable data points
- **Smart Phones and Mobile Apps**
 - DL interfaces were redesigned to enable searching on mobile devices
- **Internet technologies**
 - Web archiving increased the corpus
 - Social tagging and crowdsourcing augmented or created metadata
- **Bibliometric methods**
 - These were adopted to study and enhance the use and usability of the corpus

2020s (so far)

- **Vector embedding**
 - Retrieval greatly enhanced
- **Knowledge graphs**
 - New representations of the corpus
 - New methods of access to the corpus
- **LLMs, especially RAG**
 - Used to support automatic classification, retrieval, and a variety of other functions of the DL
 - You know better than I about this

Relationship of DL research/practice to traditional functions of libraries, and to early concepts of libraries of the future

DL research/practice and library functions

- Preservation and maintenance of the cultural heritage
 - DLs have successfully applied technologies to this function, primarily through construction of individual libraries, and through web archiving
- Supporting access to, and understanding and use of, information in the cultural heritage
 - DLs have successfully applied technologies to increase access to information, but have been less successful in supporting understanding and use
- Supporting addition to the cultural heritage
 - DLs have applied technologies that enable direct addition to the cultural heritage to some, but not a great extent.
 - DLs have more successfully applied technologies to increase the scope of the cultural heritage
- Supporting communities of users of the cultural heritage
 - DLs have successfully applied technologies to enhance support for specific communities

Libraries of the future and DL research

- Universal, remote access to the store of knowledge – To some extent
- Direct access to the contents of the documents in the store of knowledge - Yes
- Providing at least two levels of representation of knowledge, the original documents, and some representation constructed from the contents of the original documents – To some extent
- Meaningful linking of documents to one another, and to the constructed representations – Yes, and To some extent
- Support for addition and integration of new information into the store of knowledge - Yes
- Support for use of interaction with the store of knowledge in resolution of the reason for interaction with it – Not really

Some Concluding Remarks

Have DL responses to technologies enabled truly new library functions?

- No truly new functions for libraries
- But enhancements such that they may be understood as having changed the quality of how the functions are supported

LLMs and Libraries of the Future

- Are LLMs the final solution to early concepts of the library of the future?
- Many of the criteria set by Licklider et al. seem to be met
- The crucial criterion of *meaningful* construction of new representations may not be met
- But that may not make any difference

How will DLs maintain their support of library functions in the face of LLM applications?

- LLMs provide the answers and responses that individuals find sufficient to at least some of their needs
- Will people still want to engage with DLs if this is the case?
- What can DLs do that LLMs cannot?
 - Libraries support communities, even DLs
 - Context matters
 - Value of human curation

Thank you for your kind attention.

References

- Arms, W. Y. (2012). The 1990s: the formative years of digital libraries. *Library hi tech*, 30(4), 579-591.
- Bush, V. (1945) As we may think. *Atlantic Monthly*, 176(1): 101-108.
- Fox, E.A. (1999) Digital Library Initiative projects 1994-1999. *Bulletin of the American Society for Information Science*, October-November 1999: 7-11.
- Gladney, H.M., E. A. Fox, Z. Ahmed, R. Ashany, N.J. Belkin & M. Zemankova. (1994) Digital libraries: Gross structure and requirements. In: *Digital Libraries '94, Proceedings of the Symposium on Theory and Practice of Digital Libraries*, College Station, Texas, 1994, 10 pp.
<http://csdl.tamu.edu/DL94/paper/fox.html>
- *The growth of knowledge: Readings on organization and retrieval of information* (1967). M. Kochen, ed. New York: John Wiley & Sons.
- Licklider, J.C.R. (1965). *Libraries of the future*. Cambridge, MA: MIT Press.
- Saracevic, T. (2004, September). Evaluation of digital libraries: An overview. In *Notes of the DELOS WP7 workshop on the evaluation of Digital Libraries*, Padua, Italy
- Wells, H.G. (1938). *World Brain*. Garden City, NY: Doubleday, Doran & Co.
- Xie, I., & Matusiak, K. (2016). *Discover digital libraries: Theory and practice*. Amsterdam: Elsevier.