Where Do We Go From Here? The Next Decade for Digital Libraries
By Clifford Lynch
Notes
2010-08-31

- Digital libraries' roots can be traced back to 1965 when Libraries of the Future by J. C. R. Licklider was beginning to imagine a world of broad access to knowledge.
- Automated machines that run behind the scenes at digital libraries came into existence in the 1960s.
- You could call Lexis-Nexis a digital library which emerged in the mid-1980s.
- However, what most would call a "digital library" didn't spring into popular thought till the mid 1990s when the Internet really became available for many people.
- The 1994-2004 period is characterized by major programmatic funding from the U.S. government (and around the world too!).
    - Major sources of funding:
        - Digital Libraries Initiative
        - DLI-2
    - This led to academic research in the field of digital libraries which legitimized them and got the attention of the public.
    - Several organizations got together to push for an inclusive, international effort to pool resources, knowledge, and experience to create digital libraries.
    - Funding has pretty much discontinued from the U.S. government (the economy, the research had mostly been completed).
    - A digital library community however remains firmly in place.
        - They appeal to other sources for funding.
- The goal now is to find other areas to work in such as the production systems where digital libraries can handle massive amounts of data that needs to be shared and used between various parties.
- Academic community also relies heavily on digital library technology (digital asset management, digital collection creation and management, institutional repositories).
- The use of digital libraries is that they offer many tools which can be utilized in many areas.
- We need to do more research in how to preserve data and for how long.
    - Various agencies and stakeholders need to be involved in this!
- Upcoming areas of interest:
    - Personal information management as more of our activities are captured.
    - "Long term relationships between humans and information collections and systems. This is related to personal information management, but also considers evolutionary characteristics of behavior, systems that learn, personalization, system to system migration across generations of technologies, and similar questions. This is connected to human-computer interface studies and also to studies of how individuals and groups seek, discover, use and share information, but goes beyond the typical concerns of both to take a very long time horizon perspective."
    - How can information services and resources be used in all aspects of human development over the course of our lives?
    - The creation of "colaboratories" where libraries and laboratories/collaboration work together.

Digital Libraries: An Overview
By Candy Schwartz
Notes
2010-08-31

- There is no one definition of digital libraries but generally share these traits:
  - " Serve a defined community or set of communities;
  - May not be a single entity;
  - Are underpinned by a unified and logical organizational structure;
  - Incorporate learning as well as access;
  - Make the most of human ("librarian") as well as technological resources;
  - Provide fast and efficient accessing, with multiple access modes;
  - Provide free access (perhaps just to the specified community);
  - Own and control their resources (some of which may be purchased); and
  - Have collections that:
    - Are large, and persist over time,
    - Are well organized and managed,
    - Contain many formats,
    - Contain objects, not just representations,
    - Contain objects that may be otherwise unobtainable, and
    - Contain some objects that are digital ab origine." (digitally born)
- "Hybrid libraries" are those seamlessly integrate digital and analog materials together in a way that is easy to navigate for all users.
- The creation of digital libraries is a big endeavor concerning both financial resources and the skills of staff, along with questions about how to support the creation of a digital library, the long-term ability to maintain it, etc.

Setting the Wheels in Motion

- Libraries either create their own digital works or purchase access/use of pre-existing records.
- Digital libraries got their start with government and other big funding.
- Knowledge has been shared about what has gone right, wrong, the best technology.
- The types of DLs has also grown from institutional to corporation to private, etc.

Why Digital Libraries?

- "The focus of the Digital Library Initiative is "to dramatically advance the means to collect, store, and organize information in digital forms, and make it available for searching, retrieval, and processing via communication networks—all in user-friendly ways."
- The costs have gone down, and cultural changes has demanded more user-friendly ways to access information (i.e. from home and on mobile devices).
- Collaboration and "crowdsourcing" is possible in a digital environment.
- Long-term preservation of knowledge and the audience can be greater.

The Resources

- Questions to ask:
  - What to acquire?
    - Balance between user needs and costs.
  - From what sources?
  - In what formats?
    - Text, image, video, audio, games, etc.

        - Text can be broken into multiple types such as plain text, rtf, pdf, etc.
    - How to preserve materials?
- What makes an object? The whole digital book or each chapter, illustration, index, etc.?
- Intellectual property rights and copyright come into play.
- Digitalizing is a lengthy process.
- Media deterioration, stability of technology and access.

A Place for Everything, And…

- Digital collections must be treated like physical ones, identified, described, stored, and disseminated.
- Unique identification numbers and names
- The Corporation for National Research Initiatives (CNRI) "handle" system is one of the best for naming and object management systems.
- Lots of metadata: descriptive (end user), administrative, and structural (how it looks on the page).
- Then got to worry about software and hardware to support the system.

Services

- What does a digilib offer the user?
    - Information seeking and retrieval
    - Reference query fulfillment
    - User training
        - Guides, manuals, pathfinders
    - Current awareness
        - I.e. what's new
    - Learning facilitation and Collaborative Activities
        - Guides/information on how to do activities in groups

Users and DL Interaction

- Users may be a restricted community (i.e. authenticated) or all Internet users.
- Must be able to provide bandwidth and servers that can withstand the traffic.
- Users have different abilities that may or may not be able to use common or all input devices.
- Where is the website being viewed? Desktop, laptop, netbook, cell phone, tablet computer?
- Different levels of web savvy
- Search methods:
    - Keyword, author, etc.
    - Advanced with Boolean, etc.
    - Popular results algorithms like what Google uses.
    - Digilibs usually use Z39.50
- Browsing and Navigation
    - Not only by text, but by theme, subject, geospatial data, and even image pattern matching **what's this?**

Funding

- Usually from grants
- Consortium of institutions sharing the costs
- Advertising
- Sponsorships from major companies

Maintenance
- Hardware and the collection itself

Rights Management
- Nearly everything you can put in a digilib has a copyright.
- Lots of restrictions may arise (ex: only so many can access it at one time)

Evaluation
- By knowing the usage, time spent on the site, etc. One can:
  - Argue for increased funding
  - Knowing when goals have been met
  - Test the system
  - Compare several solutions to a problem
  - Look for problems

The Open Right (aka DigiLib management)
- Have to know the rights of content creators
- Where to acquire content from
- Select the best infrastructure
- Encourage ongoing Digilib research
- Maintain funding
- Continue to find better usability
- Compete against other information services for attention

Understanding Digital Libraries
Chapter 1
Notes
2010-09-01

Why Digital Libraries?
- **Digital Library:** a collection of information which is both digitized and organized.
  - Can be searched for keywords
  - Accessed anywhere in the world
  - Copied without error
  - Is service-orientated meeting the time and location needs of users
  - Takes up very little physical space (hardware only)
- What it takes to build a digilib:
  - Ability to put stuff into it
    - Add content to it.
  - Ability to get stuff out of it
    - Access, search, and retrieve the content.
    - Information must be protected/preserved so it remains online
  - Ability to pay for it
- Digilibs change culture because of the way it collects, stores, and shares information.

History of Libraries
- **Technological determinism:** the idea that it is hopeless to fight against the change brought about by new technologies.
- The move is from "who has the content" to "ability to find the content."

- Digilibs require more than hardware, they require people and machines, a culture to be produced that can help people find the information they're looking for.
- Information transportation (from person to person) is very important.
- Preservation of information over time is also important.
- Access to information must be widespread.
- Who has the most books (in general):
    - National collections
    - Universities
    - Public libraries (i.e. New York Public Library)

Vannevar Bush

- "As we may think" written by Vannevar Bush and published in *Atlantic Monthly* in July 1945.
- He wrote that great scientific progress had been made during WWII because people (scientists) worked together, not alone.
- He talked about:
    - Organize scientific effort
    - Factory process turned ideas into reality
    - Mass production and sharing of knowledge
- His system was called, Memex.
    - Barcoded microfilm
    - Precursor to hypertext links
    - Ease of posting content
- He didn't know/think about keyword searching all text.
    - He envisioned a community of *humans* who worked together that labeled information.
- Warren Weaver in 1947
    - Thought computers could simply translate languages with ease.
    - He envisioned searches being done by statistical processing: **what is this?**

Early Language Processing

- It took a long time to figure out how to do even basic database queries.
- You have to be specially trained to do it.
- Speech recognition software will have to be handled statistically and it is still a very in-exact science.