Toward Participatory Digital Libraries

Gary Marchionini
University of North Carolina at Chapel Hill
2010 International Symposium on the Transformation of Library and Information Science
November 16-17, 2010

Taipei, Taiwan
Outline

• DLs reflect what we value and what we are. Physical libraries have been cathedrals of knowledge, learning, and thus power. DLs give us a lens on what we are becoming in the digital age.

• Electronic Technologies change information work: Hybrid Libraries and Offices
• More active participation: Sharium Model
• Personal DLs join with Institutional DLs
• R&D Challenges
Information Resources Trends

• Content Features (queries too)
  – Not only text
    • Statistics, images, music, code, streams, biochemical
  – Multimedia, multilingual
  – Dynamic
    • Temporal (e.g., blogs, wikis, sensor streams)
    • Conditional (e.g., computed links, recommendations)

• Content Relationships
  – Hyperlinks, new metadata, aggregations
  – Digital Libraries, personal collections

• Content acquires history

Gary Marchionini, UNC-CH
Digital Objects

Gary Marchionini, UNC-CH
The Social Side

– Libraries are social organisms.
– They have foundational missions and policies that reflect their institutional ‘genetics’.
– They are influenced by environmental conditions that affect their collections and services (phenotypes).
– Library collections and services reflect the social organisms that support them.
– Digital Libraries offer broader kinds of collections and services (different phenotypes).
– Digital Libraries respond to digital environmental conditions and thus provide reflective lenses for understanding evolving digital societies.
Context

• In addition to the continual evolution of the DL objects themselves, many layers of context also evolve

• Context is manifested through USE that is made harvestable by Cyberinfrastructure
  – Click streams/logs/
  – Explicit hyperlinks in and out [e.g., Citeseer, DBLB]
  – Implicit relationships [e.g., recommendations]
  – Other relationships [e.g., temporal, spatial, conceptual]
Content, Metadata, & Context: Boundaries?

Gary Marchionini, UNC-CH
Argument: Institutional and Personal Time

- Libraries are memory institutions
  - Access implies persistence
  - There is substantial responsibility in distinguishing between what is ephemeral and what is worth keeping
  - Digital libraries emphasize local content
- Digital age now spans 3 generations
- Personal memories increase in value with age
- Digital assets require attention: no reliable attics
- Digital libraries must participate in collection, management, and perpetuation of personal memories

Gary Marchionini, UNC-CH
Participatory Trend

Institutional DL(1)

Institutional DL(n)

Personal DL(1)

Personal DL(2)

Personal DL(n)

Gary Marchionini, UNC-CH
The Institutional DL Milieu

- 20 years of DL research and development
- Hybrid libraries
  - Academic, research
  - Publishing (e.g., ACM DL)
  - Secondary (e.g., OCLC, Research Index, ISI, institutional repositories)
- Born Digital
  - Special library projects (e.g., Perseus, Open Video)
  - E-science databases (e.g., Genbank, EOS, sensor nets) Petabytes per day, exabytes per year
  - Indexes (e.g., Google)
  - Contributor Run (e.g., ibiblio)
  - Social Networks (e.g., Twitter, FB)
- Continued trend toward integration and linkages of physical and digital information resources
The Personal DL Milieu

- Multiple capture streams of personal memories
  - Files: photos, music, videos, texts, e-books
  - Communiqués: emails, blog posts, wall posts, tweets
  - Streams: SMS feeds, PHRs, sensor streams (lifelogs, smart devices)
  - Secondary: annotations, hyperlinks, friend networks
  - Profiles, passwords, access and activity logs

- Multiple generations
  - Annotations, edits, versions of objects
  - Formats and applications

- Multiple devices and formats: Cloud solution?
  - How many devices do you use? Have you lost data?

Gary Marchionini, UNC-CH
DL Collections and Services: The Sharium

- Digital Libraries are active workspaces in which many stakeholders participate
  - Multimedia streams rather than files/objects
  - Systems exhibit behavior (dynamic and interactive; computational; memorial)
  - End user interactions: contributions; annotations/tags; crowd sourcing—these become part of the collection and must also be managed

Gary Marchionini, UNC-CH
Sharium Workspace

The Sharium Work Space

- Messaging
- Search/Discovery
- Problem Solving/Construction
- Contribution
- Presentation

Digital Library
- Channels
- Files
- Tools

Gary Marchionini, UNC-CH
Spectrum of Examples

• Valley of the Shadow [http://valley.lib.virginia.edu/](http://valley.lib.virginia.edu/)
  – Contributions solicited physically
• Baltimore Learning Community [now defunct]
  – Content centralized, Teacher lesson plans and use notes contributed
  – Registered users, contributor responsibility
• Worm Community System (and other scientific collaboratories)
  – Data sharing, some with embargo (e.g., dbGaP)
• Ibiblio [http://ibiblio.org/index.html](http://ibiblio.org/index.html)
  – Contributor run DL with 1500+ collections
• Europeana [http://europeana.eu/portal/](http://europeana.eu/portal/)
  – Member libraries, end user feedback
• Wikipedia
  – Evolving policies
BitTorrent

- Virtual communities/co-ops
- **What.cd** for music sharing
  - Participation is payment
  - 100,000 participants; ~400,000 music albums
  - Invitation only; heavily regulated (one must earn rights to invite); the actions of invitees propagate to inviter (invite losers, you lose and vice versa)
  - Ratio system to modulate participation (including incentives and punishments)
  - Allow sharing as well as seeding new torrents
  - File integrity standards (e.g., format, bit rate) as well as metadata/authority standards
    - These quality standards are a hallmark of carefully controlled private sites like **what.cd**
  - Requests can be made and voted on/discussed by community—a kind of collection development mechanism

Gary Marchionini, UNC-CH
What are the roles of institutional DLs in Personal DLs?

- Storage? [economies of scale; digital estates; trusted imprimatur]
- Technical and informational consulting? [reference and training]
- Customized services? [indexes, apps]
- Local history/cultural memory? [personal digital heritage blends into local heritage]
- Vocabulary and open source standards? [connect my digital life to other digital lives]

Gary Marchionini, UNC-CH
Personal and Cultural Identity

• PIM bleeds into GIM into DL
  – Implications for privacy
  – Implications for identity

• Sensor streams, click streams, and personal histories.

• Projections+Reflections=Proflections

• Institutional networks (e.g., DLs) recapitulate our personal networks

• DLs can become trusted personal repositories

Gary Marchionini, UNC-CH
DL Model Clash

• Inside out: core is curated by expert stakeholders, content added with deliberation
  • Most national and institutional DLs
  • Pre-coordinate finding aids within collection
  • Interoperation becomes a challenge

• Self-organizing systems
  • Contributor run DLs (Wikipedia, ibiblio)
  • Post coordinate linkages become a challenge
  • Sustainability also a challenge

• These models will surely meet
Managing the Clash

- Parallel Services
- Distinct services with referral
- Integrated services with Levels of ‘Blessedness’
  - Expert curated
  - Community curated
  - Non-curated

Gary Marchionini, UNC-CH
Authority and Trust

• Whether for selection and collection building, preservation, or services, the days are past when scholarly authority alone determines what is saved, learned, and therefore used.
  – Data generation takes advantage of computation, simulation, and mass scale human and sensor contributions
  – Scientific discovery takes advantage of data mining and analysis
  – Indexing and access benefit from social tagging: Expertise plus the long tail
  – Preservation benefits from collective use rather than provenance and authority alone

• Digital librarians must share control while instilling trust: Balance expertise and the wisdom in the long tail

Gary Marchionini, UNC-CH
R&D Challenges

• Interoperation
  – Technical (e.g., hardware, software)
  – Data and metadata (e.g., formats, protocols)
  – People (e.g., language, culture)
  – Institutions (e.g., consortia)

• Discovery and Use
  – Indexing and representation
  – Retrieval algorithms (e.g., multiple sources of evidence)
  – Interactive interfaces (e.g., agile views, visualizations)
R&D Challenges (cont’)

• Collection Development and Contributions
  – Degree of control
  – Version control

• Help/Reference
  – Automatic/human mix (e.g., from FAQ to chat)
  – Need analysis/ (‘reference interview’)

• Maintenance and Preservation
  – Assuring persistence and stability/authority
  – Harvesting context
R&D Challenges (cont’)

• Intellectual Property
  – Own/license(rent), free/fee
  – Securing, tracking
  – Confidentiality/privacy

• Hybrid Libraries
  – Parallel systems (costs, redundancies)
  – Informing users

Gary Marchionini, UNC-CH
Library Augmentations

• New types of reuse and sharing
• Patron Contributions
• Virtual communities and collaboratories
• Direct support for creation and use (entire information life spiral)
• Collaborative filtering, cataloging, question answering
• Open-source libraries
Preservation

• What is worth preserving?
  – Genes (genotypes) vs expressions (phenotypes)
• What context to include?
• Who decides?
• Who pays? How much?
• Storage model (replication, migration, emulation)
• Storage policies (e.g., authority, cost)

Gary Marchionini, UNC-CH
Preservation Challenges

- Physics: Petabytes per day
- HD film: 2-10PB; at least one per day created

Storage costs: $500/TB/year [$500K/PB/yr]
- disk, tape, verify, system admin, upgrades

Archival desiderata: 100 years

Verification methods and costs (impossible to read and verify exabytes, new stochastic techniques and accompanying risks)

What about ephemera? Interactions? WOW?

Gary Marchionini, UNC-CH
Challenge: Managing Blurred Boundaries

Applications, Interfaces, Security

Open WWW

Curated Web

Private Web

Gary Marchionini, UNC-CH
Pointers

• UNC-CH School of Information and Library Science:  www.ils.unc.edu

• Gary Marchionini:  www.ils.unc.edu/~march
  gary@ils.unc.edu