Surveying the Digital Library Landscape

Trends and Observations

Michael J. Giarlo
An overview of the digital library landscape
Fundamental Concepts

CCSDS Open Archival Information System Reference Model (2002)
Growing recognition that the digital library is an institutional endeavor
Digital Curation

"IRENE," courtesy of kqedquest@Flickr (CC)
Enabling digital curators to **mitigate risks** to their data

Abandoning the notion of **preservation as place**

Acknowledging that the **curation process** begins early
Emerging principles

First and foremost, preserve the bits

"Lots of copies keep stuff safe"
Divisions between access and curation continue to blur
Challenges

1. Cost and **scalability** of, e.g., storage
2. New and unfamiliar **formats**, maybe
3. Scholarly communication **and collaboration**
Repositories

Courtesy of davidflanders@Flickr (CC)
The evolution of repositories

- **collection-building tools** (Greenstone, DLXS)
- **display applications** with harvestable metadata (DSpace, CONTENTdm)
- **OAIS-inspired monolithic frameworks** (Fedora, aDORe)
- **decoupled micro-services** + web architecture (California Digital Library, SWORD, OAI-ORE)
Building trust in repositories
How? By ensuring:

- content will **not change** except, e.g., for planned format migration
- derivatives and migrations of content are **faithful** to the original masters
- content is **authentic** and has documented provenance
- content is **accessible** via known identifiers
Technologies for trust

- digital signatures
- persistent identifiers
- audit trails / provenance metadata
- content versioning
- authorization / authentication
- rights / access control
Growing the repository
Broad growth

electronic theses & dissertations
research data
learning objects
pre-prints
electronic records
public domain documents
Deep growth

Incorporating the repository into institutional programs and workflows
Acknowledging the tension between preservation and growth
Serving as a **conduit** between library metadata and the web's best practices and "2.0"-style uncontrolled metadata
Outward

Contributing authority metadata and schemas to efforts to build a more descriptive web
Inward

Adding value to services by augmenting them with "crowd-sourced" metadata
Emergence of the Linked Data pattern

● "Use URIs as names of things.
● Use HTTP URIs so that people can look up those names.
● When someone looks up a URI, provide useful information.
● Include links to other URIs, so they can discover more things."
Linked Data sites in library-land

- id.loc.gov
- dewey.info
- metadataregistry.org
- viaf.org
- chroniclingamerica.loc.gov
- wdl.org (soon)
Challenges

- Expensive human description efforts do not scale with all the data
- Will provenance and preservation metadata suffice in the trust framework?
- To what extent should embedded metadata be used?
Interoperability

Courtesy of atomatic@Flickr (CC)
Much of the recent convergence between library technology and the web has occurred in the digital library context.
Libraries and cyberinfrastructure
Integrating library systems and processes into campus cyberinfrastructure
Standards and best practices help, as does alignment with technologies from outside libraries
Sustainability

Courtesy of the Toronto Society of Architects (CC)
Sustain the digital library through a framework of partnerships, institutional commitments, and technology strategies.
Bibliography


Thanks!

Blog:  
http://lackoftalent.org/michael/blog/  

Slides:  