Electronic Finding Aids

Some Issues and Notes from
Emory University

Martin Halbert, Director for Library Systems, Emory University
DLF Forum Presentation   7/17/99

Big Issues We All Face

- Emerging data standards (EAD, XML).
- Technology in rapid flux.
- Integrating technology work units into our traditional organizational structures.
- Developing sustainable initiatives.

Emory's Interest

- Engaged in many digital projects, most relevant is the SAGE Project, a $1.5 M digital archives grant-funded endeavor.

- SAGE is particularly interested in designing for sustainability and interoperability.

- Time for initial digitization experiments is past. Now the challenge is to begin to make digital archives stable and viable in long term, like print counterparts.

Digital Finding Aids & Archives

- Immediate confusion encountered: Are digital finding aids different from digital archives? Technology blurs the difference in a way that does not happen with print artifacts.

- Immediate challenge encountered: It is easy to replicate a static hierarchical finding aid digitally. But is that all we want to do?

SAGE Project Work

- ## Technology and Formats:
  - Exploration of the EAD format for both finding aids and digital object reference. How far can we push the standard?
  - Adaptable search engine (CNIDR Isite) as an integrated complement to the traditional finding aid. Dynamic finding aids?
  - Automated migration technologies for ongoing data conversion: EAD/SGML to HTML now, EAD/SGML to XML later.

- ## Sustainable practices:
  - Dirty OCR text for search engine / Clean PDFs for display.
  - Streamline and automate the workflow to make use of students and limited number of staff.
  - Scalable, inexpensive SW & HW tools.

Mindset

We must actively work to change mindsets from static print approaches to dynamic digital practices:

- – enormous differences between print finding aids and digital archive systems.
- – hierarchical print finding aids versus recombinant (ad hoc) electronic finding aids.
- – time scale for transcription/migration of finding aids: print = decades, digital = months.

Differences in Online versus In-Person Expectations

- Need to manage expectations of online users. (In person, we'd give them a bunch of boxes of stuff to browse. Online, do we really want to give them the equivalent: 2,000 hits?)

- We have no well-accepted studies of the value of item level descriptions.

- Online users expect to be able to search down to the word level. Can we afford this?

What Contributes to Viability?

- ## Interoperability / Open Source Technology
  - Demonstrated success of Linux/Open Source movement as a potential model. Give away what you develop. The more people that use it, <u>the more they will help develop it.</u>
  - Buy into community synergy and stone soup, rather than building beautiful, expensive, isolated ivory towers. Soup will sustain you, towers will drain you.
  - The more phones there are that can interoperate, the better off all their users are. We have to face the Z39.50 issue or its equivalent.

- ## Modularity / Immediate Utility
  - Build it in chunks. Make something that is useful right away. That way everyone is interested in continuing to develop it. Institute version control and migration tools. Be ready, willing, and able to switch systems.

What Would Help

- Expanded development communities (if we all go our own way, we all go the same way). The more alliances and forums there are for sharing tools and encouraging joint development, the better (call us if you want to talk, we're happy to talk to anybody!).

- More tests and exploration of what distributed searching means in digital archives environments as opposed to bib catalogs.

Contacts / Webcites

- Martin Halbert:
  - Email: mhalbert@emory.edu
  - Phone: 404-727-2204
  - URL: `http://elan.library.emory.edu/Staff/Mhalbert/`
- SAGE Project:
  - URL: `http://sage.library.emory.edu/`
- CNIDR Isite Isearch (Open Source) Search Engine:
  - URL: `http://www.cnidr.org/ir/isite.html`