Outline, February 20, 2008

1. Appreciation of a multi-disciplinary project
   
a. Benefits – very interesting work!

Working on a team with a blend of professional skills, with everyone working together towards the accomplishment of team goals.


Computer science - Archives, as in the Open Archives Initiative, is understood as a repository of scholarly papers

Archival profession, the term Archives is marked by characteristics such as an emphasis upon preservation and the existence of an institutional policy defining its mission.

c. Important need – mutual engagement. With archivists and technologists communicating their professional knowledge and expertise to one another; to be part of a learning process.

2. Building and extending a search system

I worked with my project colleagues in the evaluation and selection of EAD search solutions in mid-2003.

a. Needed a defined functionality – search and retrieval of XML documents;

b. Lack of market options; tasked with investigating search and retrieval options for Encoded Archival Description finding aids in XML format, the options were quite limited.

Online Archive of California, Adviser organization to the Northwest Digital Archives: During the 2001 to 2004 time period, migrated from Dynatext, a commercial SGML publishing solution, to DLXS XPAT to XTF, the latter a locally-developed open source application, to support online EAD search and retrieval.

c. Limits on open source solutions – one driver is the typical use of Relational Database Management System databases to host EAD content; true of two EAD indexing tools currently available, Archon, developed at UIUC; Archivists Toolkit, which is more of a processing than web search tool. Both require the mapping of the finding aid’s XML content to the table and columnar structure of a relational database management system.

i. Content can be lost, if it’s not accounted for in the mapping.
ii. Lose relationships that exist in the XML after the mapping to relational is performed.

d. Needed the ability to develop applications now (2003, 2004) and in the future; in other words, a closed architecture system or a system enabling local development and extension...

commercial versus OSS: Open source, such as Xindice; need to embrace a developmental over a proven solution, at least at that time, because of the lack of maturity in open source native XML search solutions.

Open API (or ruleset enabling local development) versus a solution which would have required communication with a mediator, a technology company, to enhance the site.

Cut out the mediator: focus communications on the issues that normally arise in a multidisciplinary project, that I touched on earlier.

IXIASOFT TEXTML database that’s been used to support the NWDA union catalog since 2003

1. XML documents are the basic unit of storage; versioning in supported in the database

2. Documents can be queried using an XML-based query language...

3. Use the .NET API...

Some of the services that we’ve built in the last five years include:
i. Built a web-based user interface, the NWDA researcher site. This site now has all durable URLs that enables searches to be passed from external systems to the NWDA catalog.

ii. An RSS feed application showing recent resources added to the NWDA system; again, including the browsing terms as categories, so that they can be used as the basis for filtering a feed.

iii. An Open Archives Initiative – Protocol for Metadata Harvesting data provider instance that enables repository harvesting by criteria such as NWDA member institution or by NWDA browsing term (which is a local subject term assigned by an archivist).

iv. An application that feeds NWDA finding aid information to Google, consistent with Google’s sitemap specification.

Able to perform all of this work locally, using the vendor’s API information and standards such as the Encoded Archival Description and the NWDA program’s best practices guidelines, using a single data source, the NWDA union catalog.

While there are defects with this approach, paying an ongoing support fee of around $2,500 per year, API documentation that’s incomplete, it’s still proven to be a viable approach; because of our ability to extend the application locally in a way that’s consistent with the archival needs of the program.

3. **Digital program – integration of finding aid content with digital objects (for example, digitized photos...)**

a. Even in 2008, the lack of market options for archival finding aid searching and digital object integration is still a problem (CONTENTdm, EAD and digital objects...; TEXTML).

CONTENTdm: A tool that we’ve used extensively for digital collections development, and one that’s used at a number of Alliance institutions and by NWDA member institutions.

EAD indexing in the product is limited to the first 128,000 characters, for full text searching; seven EAD tags, such as <titleproper> and <unittitle>, are extracted into CONTENTdm for indexing.

Archon, which I mentioned earlier, is an open source tool that supports the integration of finding aid content with digital objects. One problem, like with CONTENTdm, is that parts of the finding aid may be lost as it’s imported into Archon’s MySQL relational database.

Vision: Provide access to digital objects while keeping the full archival record, the finding aid, accessible and intact.
If you're interested in access to online archival materials, I recommend that you read an article authored by Elizabeth Yakel in the May/June 2007 *D-Lib Magazine* on next generation archival finding aids.

She describes a prototype online site that supports end-user tagging of archival documents. *This is important, because site users may know things about individuals and collections that the local archivist doesn't.* But, as Yakel notes, one requirement of this tagging software, being created at the University of Michigan, is keeping the archival record intact.

b. Systematic approach through a planning grant; the planning is being done with this technical backdrop…

4. My summary

*Donald Waters of ARJ noted, that given the increasing development of book digitization projects by companies such as Google and Microsoft, [research libraries] will increasingly be distinguished by the special collections of rare and unique materials which they hold and by the scholarly services they provide for these materials, especially in conjunction with similar collections at museums and archives locally and around the world.*

Underscores the importance of this work, and that it be approach with an eye towards sustainability.

One goal of a digital program

Integrate EAD and digital services

Logical next step