LIS 505 Reading Notes
2010-10-24

Encoded Archival Description: The Development of an Encoding Standard for Archival Finding Aids
By Daniel V. Pitti
2010-10-24

- Encoded Archival Description (EAD)'s underlying technology is Standard Generalized Markup Language (SGML) and Extensible Markup Language (XML).
- EAD was created to try and solve the problem of collections being around the world
  - Physical distance makes finding and accessing materials difficult
  - Since the mid-19th century, archivists have been trying to solve this problem
  - This is really about providing universal **intellectual access**
    - Since you can't ship the physical items
  - EAD ties in with trying to get finding aids standardized
    - Standardization is a community issue
      - Must reflect a community's interests
      - Community must be involved in developing the standard

Universal Access via Printed Catalogs
- Explores the history of the card catalog
  - LC was responsible for creating its own card catalog and then asked select elite libraries to send in their cards as well.
  - This became known as the National Union Catalog (NUC)
    - It only covered published materials, not primary resources
    - Was published for awhile
- *A Guide to Archives and Manuscripts in the United States* (1961)
  - Listed collection-level information about the holdings of 1300 archives
  - Later they tried to update it but found out that the collections had grown so much that they couldn't do it, so it became a repository listing instead.
- *National Union Catalog of Manuscript Collections* (NUCMC)
  - By the LC
  - Did the same thing for manuscript collections that the NUC did
    - Created catalog cards
  - Lasted from 1962-1994 in print
    - The decision was to use networked computers instead

Universal Access via Online Catalogs
- MARC came into being for catalog cards
- OCLC and Research Library Group (RLG) databases became the de facto union catalogs
- MARC Archival and Manuscripts Control (MARC AMC)
  - "The AMC format made it feasible for archives and manuscript repositories to provide brief, synoptic surrogates for collections in their care in bibliographic catalogs. The AMC format by itself, however, only specified content encoding standards; it did not provide standards or the actual content of the records themselves, and without such standards, he format was simply an empty vessel."

- - **Basically, they summarized what was in a collection.**
- AACR2 didn't work for archives because "its chapter on manuscript cataloging abandoned longstanding archival descriptive principles."
- Archives, Personal Papers, and Manuscripts (APPM) + AMC was a better fit for archives
- Finding aids are the next step after MARC AMC records in trying to determine individual items/series
    - Order: MARC AMC > Finding Aids > Specific items

The Value of Standards
- Standards are awesome
- They make resources available
- **Archivists like to whine and don't want to be standardized because they're all super special twinkly princesses. (my words)**

The Lessons of MARC
- By encoding MARC fields with specific names, the information can be extracted in the future for various devices.
- MARC is open source
- MARC is 30 years old.
- Most MARC systems only import and export MARC records
- Selection criteria for EAD:
    - Open source standards
    - Could handle complex finding aid information and make it searchable
- Why MARC wouldn't work:
    - Limited to 100k characters
    - Doesn't support hierarchically structured documents
    - The MARC-community doesn't have the money to fund research for a more complex version
        - Only used by libraries, museums, and archives instead of a larger audience

SGML, HTML, XML, and EAD
- Why Standard Generalized Markup Language (SGML) is awesome:
    - Open source standards
    - Supports hierarchically structured documents
    - No restrictions on file size
    - A much wider audience of users
- It is a standard for constructing markup languages (metastandard!)
- It is not perfect off the shelf for archives; needs modifications
- **A community of users can decide on naming conventions, syntaxes and language for the structure of documents and thus been able to create a standard unique for their community.**
    - **Document Definition Types (DTDs)**: when a markup language is written in compliance with SGML requirements
    - Each community has developed their own DTDs which becomes their standard
- SGML supports descriptive markup of text
- Descriptive markup
    - Structural
        - Identifies document components and their logical relationships
        - Structural elements generally are components that warrant distinct visual presentation:

- - - Ex: chapter titles, paragraphs, lists, and block quotes
  - ○ Nominal
    - Identifies named entities, both concrete and abstract
      - ▪ Ex: corporate names, personal names, topical subjects, genres, and geographic names.
    - EAD can understand Scope, Content, Biographies, etc. as named entities and thus can search for them
- SGML also supports "referential markup" aka links to external data.
  - ○ **You can link repositories together this way!**
- HTML is the most wildly popular usage of SGML
  - ○ Can't look for very specific information though
  - ○ I.e. it just says "list" and not what this is a list *of*
- SGML information can't be used on the web. It's too raw. Thus came XML!
  - ○ XML is more complex than HTML (learning curve!) but easier than XML
    - Thus more powerful than HTML

Overview of EAD Development
- EAD sprang from the Berkley Finding Aid Project (BFAP)
- 1995 EAD started coming together
- The job of the EADWG was:
  1. assisting
     in developing and reviewing a data model for archival finding aids;
  2. reviewing the EAD DTD;
  3. testing and evaluating the EAD DTD;
  4. reviewing application guidelines; and
  5. initiating review of EAD by the SAA Standards Board and SAA Council
- September 1996: EAD was made public
- Many high level universities worked on
- RLG developed the Finding Aid SGML Training (FAST) Workshop to get libraries and archives on board with accepting EAD