Acknowledgements

Michelle Dalmau, Digital Projects and Usability Librarian, Digital Library Program, Indiana University

Jenn Riley, Metadata Librarian, Digital Library Program, Indiana University

Digital Library Program website:
  ◦ www.dlib.indiana.edu
Focus on Users

- Cannot digitize everything in the collections, so usually consider heavily-used collections for digitization.
- Once a collection or specific items are selected for digitization, the next question is usually, “What about the metadata?”
- In considering how to enhance or create metadata, many options.
- Look to use of collections/items and potential use for guidance in metadata plans.
Indiana University Experience

- Digital Library Program established in November 1997
- Built on early digital library development in three areas:
  - Digital music library -- Variations
  - Electronic text – Victorian Women Writers
  - Digital art images – DIDO
- Began seeking grants from the beginning
- Many grant funders required use and usability work:
  - Does project serve an identified need?
  - Does it function as promised?
- Created position for a use and usability specialist
Case study: Sheet Music

- Indiana University’s Lilly Library had two collections of sheet music, Sam DeVincent and Starr.
- Sam DeVincent Collection of American Sheet Music
  - 24,000 pieces of sheet music, songbooks, and folios, acquired in 1998.
  - In 1988 DeVincent donated a large portion of his collection to the Smithsonian Institution's National Museum of American History.
  - Lilly collection contains duplicates of some of the pieces in the Smithsonian collection additional materials DeVincent subsequently acquired.
  - Arranged by collector: personal names of musicians or performers or on subjects he defined that were as diverse as the American Red Cross and Halloween.
  - Created an in-house database.
Starr Sheet Music Collection

- More than 100,000 pieces of sheet music arranged in three categories.
- Composers, lyricists, performers, and literary figures
- Subjects and types of music
- Chronological miscellany

Primarily a collection of American popular music, which extends from the late eighteenth century through the 1950's.

- No database listing items in this collection.
- No written record of all titles in any given category.
Descriptive Metadata

- Lilly cataloger had done early work to develop a MARC record for sheet music, but it was never officially adopted.
  - About 500 pieces from the Starr Collection were cataloged experimentally and added to the library’s online catalog.
- DeVincent database was useful in-house but did not adhere to standards; no authority control.
- Union catalog of historical sheet music seemed like a good idea, but what about cataloging and metadata?
Sheet Music Consortium

- Four universities with substantial sheet music collections met in 2001 to begin discussing a union catalog for sheet music.
- Meeting for all interested parties at IU in 2002.
- Production site launched in 2003.
- Not just for online sheet music.
- Provides an aggregated search, then sends the user back to the home institution for access.
Same question: Metadata?

- Intention from the beginning to respond to user needs.
- Faculty users of Lilly sheet music collections were invited to 2002 meeting.
- Uses are varied.
  - Study of musical content, lyrics, cover art, and advertisements
  - Performance
- Asked how they would like to search.
  - Scholars had different needs from performers.
Consortium Metadata

- OAI–PMH for harvesting metadata
- Harvested data in the 15 fields of unqualified Dublin Core: Title, Creator, Subject, Description, Publisher, Contributor, Date, Type, Format, Identifier, Source, Language, Relation, Coverage, Rights.
- Some participants worried about usefulness of these fields and added labels to aid discovery:
  - Creators: Keates, Henri A. [lyricist]
    - Robinson, Harry I. [lyricist]
    - Robinson, Louis [lyricist]
    - Keates, Henri A. [composer]
    - Robinson, Harry I. [composer]
    - Robinson, Louis [composer]
    - Fred Waring [performer]
- These labels came from discussions with users regarding the importance of differentiating roles
Project Website

- Knew that the key to providing users with a useful resource would be what data providers did with their native metadata.
- Significant variety in quality and quantity of metadata.
- Website* offers guidelines for mapping metadata to Dublin Core, including answers to questions and examples.

*http://digital.library.ucla.edu/sheetmusic/OAIProject.html
Lessons Learned

- We realized that we needed more systematic ways to gather input from users regarding their search and discovery needs.
- The Sheet Music Consortium helped us realize that early online sheet music collections use customized metadata models.
  - Many similarities
  - Many differences
- Next sheet music project would build upon lessons learned from Sheet Music Consortium (SMC).
IN Harmony: Sheet Music from Indiana

- In 2004 IMLS-funded project to digitize and create metadata for 10,000 pieces of sheet music from four diverse partners:
  - Indiana University Lilly Library
  - Indiana State Library
  - Indiana State Museum
  - Indiana Historical Society

- Goals included creating
  - standards-based images that could be used to print and perform; and
  - rich, interoperable metadata, responsive to user needs.
Users are limited by the search options provided by the specific institution or resource; how would they like to search?

Experts are one source of information but users themselves are key to answering this question.

SMC experience and research revealed:
- Users of sheet music have unique discovery needs, especially with regard to subject access.
- Uses do not approach subject access in a uniform or predictable way.
- Subject searches typically include topic, form, genre, style, and geographic terms.
User Studies

- Built into the project from the earliest stages
- Four major user studies
  - Query logs analysis
  - Card sort
  - Task scenario
  - Email content analysis
- Purpose was to inform the design of the metadata model, the sheet music cataloging tool and the collection web site.
Conducted to assess actual user queries and discovery patterns.

Derived from server logs generated by two online sheet music websites, Sheet Music Consortium and the Indiana University Sheet Music Collection.

IU Sheet Music Collection comprised of homogeneous metadata described by custom fields.

Designed to learn what type of searches users perform, how often they conduct known-item vs. unknown-item searches, what kinds of searches users conduct (topical, genre, etc.).

Learned that known-item searches predominate, but among subject searches most common types are genre/form/style, topic, and instrumentation.
Card Sort

- Designed to understand how representative users approach subject-related discovery.
- Built upon query logs analysis by using actual queries harvested for card sort concepts.
- Also tested their own categorical constructs of subject against the users’ construct.
- Nine users of online sheet music grouped 55 pre-defined concepts and assigned category names to their groupings.
- Revealed that structured metadata is essential to aiding discovery.
- Cross-relationships were judged important by participants for access.
- Confirmed research finding that it is difficult to distinguish among genre, form, and style.
- Influenced modular metadata structure that could support flexible discovery mechanisms in the end-user interface.
Task Scenario Study

- Designed to further examine subject–related access.
- Compared user–provided terms with user–selected descriptors from pre–identified controlled vocabularies.
- Tasks designed to determine which, if any, controlled vocabularies best match participants’ natural language and search strategies.
- None of the vocabularies emerged clearly as most appropriate for describing sheet music.
- Results reinforced design of modular metadata model to accommodate multi–faceted access to sheet music.
Email Content Analysis Study

- Content analysis of ~50 emails about sheet music sent to the Lilly Library.
- This method calls for a systematic deconstruction of email content based on a predefined set of objectives:
  - Content
  - Search and retrieval strategy
  - User profile
- Resulted in twelve, pre-defined specific coding categories for each of the above objectives; assigned to each email query (as appropriate) based on a set of coding guidelines.
- Reinforced findings from other studies: prevalence of known-item searching, multi-faceted queries comprised of two categories (e.g., war marches), and users’ conception of title extending to first line of song or chorus.
Conclusions

- The two sheet music projects offer insights into the need to work closely with users in designing metadata for digitization projects.
- Specific findings are less important than use of multiple methodologies to assess user needs.
- Digitization is too costly and complex to proceed without user input; those users may be few in number but must be representative.
- It’s expensive work but we already knew that...