Digitization Cost Calculator


- Provides time and cost estimates for various tasks related to digitization
  - Includes prepping material, rights review, reshelving material – not just scanning and metadata

- Still in development

- Uses averages from places that have submitted data

- Looking for more data, especially from smaller places, to make it more accurate
  - To find out more about contributing data, visit: [https://wiki.diglib.org/Data_Collection_and_Contribution, Digitization_Cost_Calculator](https://wiki.diglib.org/Data_Collection_and_Contribution, Digitization_Cost_Calculator)
Metadata Quality

Info on these slides taken from the session speakers:

• Gretchen Gueguen – Data Services Coordinator, DPLA
• Chris Stanton – Metadata Specialist for Empire State Digital Network
• Cory Harper – Metadata Services Librarian, New York University
Concerns for Metadata in Aggregations

• Many institutions still creating metadata focused on internal use
  • E.g. “bridge,” “campus,” without indicating which bridge on which campus. Makes sense when on your own page, but not when pulled up in DPLA search

• DPLA is looking for:
  • Content
  • Context
  • Consistency
  • Coherence
  • Conformance to standards
  • Which standards is still a question
ESDN Concerns

• Lack of context, especially in titles (e.g. “letter”)

• Missing fields – especially minimal or no subject headings
  • Headings not using a controlled vocabulary

• Identifying data that is “pointless” for DPLA so it can be masked
  • E.g. “no publisher,” “unknown,” “9999,” “no date”

• Rights statements need to be more useful and accurate
Corey Harper’s Metadata Research

• Looking at how metadata quantity and quality correlates to use of digital material
  • Very detailed – looking at things like field word counts and number of hits
  • Analyzing common searches in DPLA
  • Which fields are most useful to populate?

• Huge variety between hubs and institutions in terms of amount of metadata and fields

• Hubs have great range of percentage of items with hits in DPLA – from 1% of ESDN (don’t panic, we are pretty new to DPLA!) to 21% of Harvard University. How is this related to their metadata?

• Recording of a similar presentation, made at Digital Library Federation Forum in 2015:
  https://open.library.ubc.ca/clRcle/collections/55474/items/1.0220818
More than ¼ of words are rights statements!
Rightsstatements.org

• Supported by DPLA and Europeana

• Provides 12 standardized rights statements to describe copyright and re-use status of digital objects. Similar to Creative Commons. Statements fall in 3 categories:
  • In copyright
  • Not in copyright
  • Copyright status unclear

• Will help searchers who need items with a particular copyright status

• May force Rights fields to be more accurate, rather than blanket statements

• At some point, our material will need to comply with these standardized statements
  • We will provide training
DPLA Metadata Working Group

• Looking at ways to describe things in aggregate and make connections between items and collections
  • What is a collection?
    ◦ Collection information is useful when evaluating individual digital objects
    ◦ Confusing for users when items and collections are being searched simultaneously
    ◦ Finding or mapping the data can be difficult

• They have developed a number of user scenarios to help guide the discussion:

• After group discussions, will issue white paper for comment

• Both conceptual and technical considerations