#472

Quality Assurance and Evaluation of Change for Patent Metadata

Dr. Oksana L. Zavalina (Oksana.Zavalina@unt.edu), Mark E. Phillips, & Hannah S. Tarver

Background & Problem Statement

- Metadata quality greatly impacts access to information resources.
- Metadata quality assurance is especially important for unique valuable materials that are not widely available outside of the specialized digital collection.
  - e.g., digitized historical patents.
- Metadata is edited in the process of quality assurance
  - Lack of studies evaluating metadata change
    - Often due to lack of data: no metadata versioning

Texas Patents collection

- Part of UNT Portal to Texas History
  - Metadata versioning enabled since 2009
- Collection of historical patents (19th – early 20th centuries)
- Testbed collection for metadata training, metadata quality and metadata change evaluation

Case study

- Metadata management in *Texas Patents* Collection
- Evaluation of metadata change intended to support the metadata quality assurance for this collection.

Data

All (31,068) versions of 13,025 unique metadata records (as of May 2017)

Research Questions

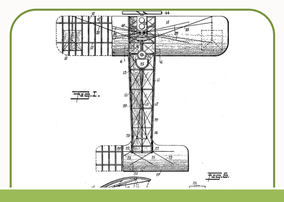
- Level of change in metadata records and fields:
  - frequency of metadata change events
- Distribution of metadata change among editors
- Characteristics of change between versions: fields, etc.

Research Method

Content analysis of metadata records

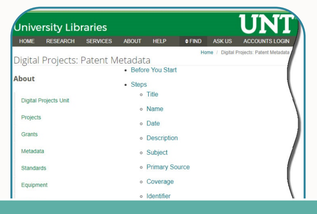
Metadata in Texas Patents Collection

- UNTL metadata scheme based on Qualified Dublin Core
  - Hierarchical
  - 20 descriptive metadata elements (data about information object)
  - 1 administrative metadata element (automatically captured data about metadata record)
- Available for harvesting via OAI-PMH in UNTL, DC, and METS formats



•1. Digitized patent is loaded into repository with the hidden minimal UNTL metadata record

- Several automatically prepopulated descriptive fields: *format*, *collection*, etc.



•2. The rest of the UNTL record is completed by metadata editor

- All applicable descriptive metadata fields (use online entry form)
- Metadata editors rely on detailed collection-specific metadata guidelines
- Over 400 metadata editors, including specialists, students, volunteers

3. Record is made visible

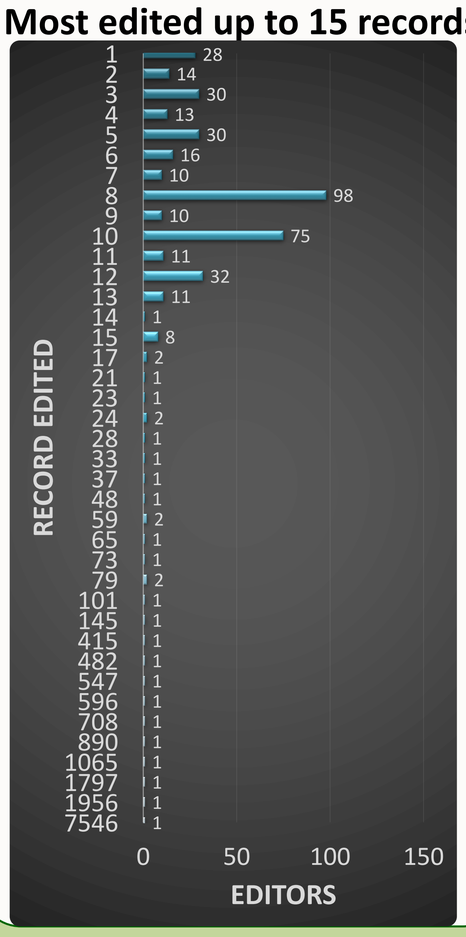
- After that, it may or may not be edited again

Frequency of metadata change events

- Only 69 records were not changed
- Most changed between 1 and 3 times
- 40% had 3 editing events

Level of metadata change by editors

Most edited up to 15 records

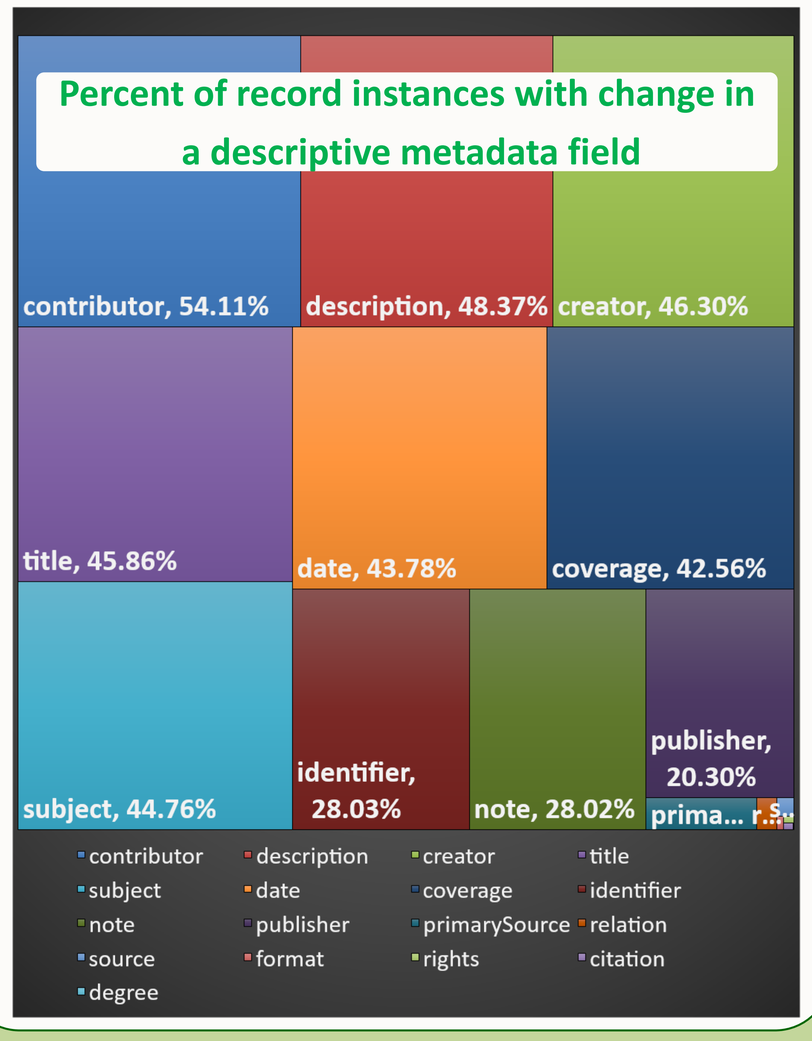


Level of metadata change by fields per editing event

- Largest number of editing events included changes made to 2 metadata fields in a record
- BUT over 2,000 edits included change to 10 fields in a record

Overall Change by Metadata Field

- Most (17) descriptive fields changed at least once
- 10 fields were edited in 20% or more of all record instances



Concurrent research

Comparative analysis of each pair of consecutive versions of the same metadata record to identify categories and subcategories of metadata change (e.g., additions, deletions, modifications, etc.

Conclusions

More studies in different repositories are needed to contribute to understanding of metadata management and metadata quality assurance and the role of metadata change in these processes.