Metadata Assessment Tools Hackathon

Kate Flynn
Amelia Mowry

Agenda

Settling time (15 minutes)

Slides (15 minutes)

Work (45 minutes)

Report (15 minutes)

Background

DLF Assessment Interest Group Metadata Working Group Environmental Scan in 2016

Updates in 2017

Metadata Assessment Tools Repository

How it works

Think up a task (or use one of ours)

Find some data (or use our samples)

Test!

Let us know how it went!

Data Sets

- DPLA Bulk Download
  - https://dp.la/info/developers/download/ (can download sets from individual contributors)
  - https://digital.library.unt.edu/ark:/67531/metadc502991/ (2015)
- MARC Distribution Services (data set)
  - https://www.loc.gov/cds/products/marcDist.php
- UNT Metadata Edit Dataset
  - https://digital.library.unt.edu/ark:/67531/metadc304852/
- Find more at https://github.com/kateefly/ToolsDataSet

Potential Tasks

- Checking that the metadata in a specific field matches a specific controlled vocabulary
  - OpenRefine reconciliation service
- Checking that there are no spelling errors and punctuation follows the guidelines you've set
  - 5" x 8" and NOT 5 in. x 8 in.
-

Potential Tasks

- Checking a field for data type conformance
  - Dates should be in YYYY-MM-DD. Values like 'n.d.' and 'ca. 1915' are not allowed
- Checking that required elements are present
  - You have a new set of required elements and now need to check to see if they are present in your legacy metadata (hint: you can try using the requirements found the documents found at the [Metadata Application Profile Clearinghouse Project](#) as guide for this task)

Potential Tasks

- Checking that the meaning of the field remains consistent
  - **Physical Description**
    - Image **[DCMI Type]**
    - black-and-white prints (photographs) **[AAT]**
    - 5 x 8
- Edit content

Failure is ok!

You do **not** have to succeed at the task you set for yourself

If you fail, that lets us know what skills a given tool might require or how much time you may need to put in to make it useful

Example Test - Catmandu

User wanted to use Catmandu for assessment, but couldn't get it installed even though she had command-line experience. Further research suggested it may be better suited for Linux.

Takeaway: Failures provide useful information. Don't be embarrassed to share what went wrong.

Example Test - OpenRefine

User wanted to use OpenRefine to assess quality of ebook records. He found it helpful for "semi-consistent" fields, but not as helpful for large, textual blocks. He was frustrated by the limitation in number or rows, which limited the usefulness of the tool for large sets.

Takeaway: Context is important. Provide detail on what you're attempting and what you're experiencing.

Example Test - MarcEdit

Two users tested MarcEdit. One was attempting to batch edit MARC records and was generally successful, though she was confused by some of the tools. The other was trying to assess new records. She experienced problems using any of the MarcEdit capabilities on records that weren't already valid.

Takeaway: Even common tools benefit from tests by multiple users to uncover useful capabilities and limitations.

Let's Get Testing!

Testing form: http://bit.ly/MetadataTest

Link to spreadsheet: http://bit.ly/MetadataTools

Sample data & tests: https://github.com/kateefly/ToolsDataSet

Reports

- Did you experience anything during your test that you didn't know how to include in the form?
- Did something work in a way you didn't expect?
- Did you have trouble with installation of a tool?
- Did you need outside resources to use a tool?

Stay involved

Keep testing.

If you're interested in working with us on this, join the DLF AIG Metadata Working Group (http://dlfmetadataassessment.github.io) .

Feel free to suggest new tools.

Metadata Assessment Tools Repository.