Metadata Assessment Workshop (31 Mar 17)

Thank you everyone for signing up for today's Metadata Assessment Workshop.

We will be meeting in **Uris Lib B05 Classroom from 1 to 2:30 PM today**. This classroom has computers available, and we will be walking people through setting up the hosted options for OpenRefine and Python/Bash at the beginning (so you don't have to bring your own computer, though please do if you can).

If you were unable to register but want to attend, please email user-9f226 or Marcie Suzanne Farwell to check first (we'll try to respond ASAP). We ask that for this particular session, people do not just drop in.

Minimal Setup

If you have 10 minutes this morning and will be using the hosted options for the tools, please sign up for (free or 1 month test) accounts at Python Anywhere (https://www.pythonanywhere.com/pricing/, choose the Free beginner account) and RefinePro (which offers each new account a free month trial period): https://app.refinepro.com/register/. We will alert you to these options at the beginning of the workshop as well if you don't have time. Do not worry about any other setup, we will do those at the event.

If you are bringing your own laptop, you need Python 2.7 Pip (usually included with 2.7), and OpenRefine 2.7rc1 (rc1 recommended, rc2 or 2.6 *should* both also work) installed. If you have any trouble installing these, just use the hosted versions mentioned above for now, and we can chat at the end of the workshop about how to get your laptop set up for this work later on.

Agenda

We have a short time to cover a meaty topic, so this should be treated as an introduction to 2 methods for doing this work for jumping off in your own daily practice.

- Introduction / Setup Help : 10 minutes
- Metrics for Metadata Assessment : 10 minutes
- OpenRefine for Metadata Assessment : 30 minutes
  - Overview of OpenRefine
  - Loading a File
  - Facets
  - GREL or Google Refine Expression Language
  - Using Regular Expressions
  - Completeness Rankings
  - Export Reports
- Python for Metadata Assessment : 30 minutes
  - Overview of Python MetadataBreaker Scripts
  - Harvesting Metadata
  - General Report
  - Looking at a Specific Field
  - Using SORT, UNIQ, GREP, Regular Expressions
  - Export Reports
- Wrap up / Next Steps : 10 minutes

Sample Data

I've gotten requests to work with the following data sources:

- eCommons
- Fedora
- FGDC
- MARC
- Solr
- Shared shelf