Zotero: Low-budget, high-power middleware for electronic theses and dissertation (ETD) repository workflows

Esther Jackson, Kathryn Pope, Jeremiah Mercurio
Columbia University Libraries
Today’s presentation

- The problem
- The solutions we explored
  - Why Zotero?
- The Zotero process
  - Pain points
- Next steps
  - Caveats
The problem

- Columbia (CU) students currently submit dissertations into two repositories (ProQuest Dissertations & Theses and Academic Commons) through one interface (ProQuest ETD Dashboard).
- ProQuest’s publishing agreement does not make an accommodation for fair use claims.
- Clarivate’s acquisition of ProQuest accelerates the potential for dissertation content to be repacked and sold for large-data and analytics products.
- CU students are preparing two versions of their dissertations: one for defense and one for deposit.
Possible solutions

- New software solutions
  - Vireo
  - OJS
  - thing - th(esis).ing(est) from MIT

- Existing tools
  - Google Sheets
  - Zotero
Enter Zotero

- Open Source and community maintained
- Robust, open data structures to connect metadata with asset files and to easily export libraries
- Allows for collaborative collection management
- Cloud-hosted, with no need for Libraries (or other campus partners) to host and/or maintain a separate software product
- Low-cost
Process

● Evolving!
● Currently
  ○ Graduate School of Arts & Sciences (GSAS) staff collect author information and accept dissertation deposit through Google Form (along with receiving a signed author agreement)
  ○ GSAS sends Libraries staff metadata in two spreadsheets (information they already have about authors, and information they collect from authors through Google form)
  ○ Libraries staff merge csv files, convert to ris, and import into a Zotero group library.
  ○ GSAS staff associate final version of the dissertation with the appropriate Zotero record.
  ○ Libraries staff export Zotero library metadata and files, transform with OpenRefine to format needed for repository, and import them into the repository
  ○ Subject headings are added and the dissertations are published on a rolling basis
● Overall, smooth implementation in pilot phase
Example Zotero record

**Item Type**: Thesis

**Title**: Urban Politics of Death: Memorabilia and Mutual Aid amid COVID–19 Pandemic in New York City

**Contributor**: Devlin, Ryan

**Author**: Arora, Shreyas

**Abstract**: The thesis focuses on inquiring about, what were the challenges faced by low...

**Type**: Mutual Aid; Covid–19 Pandemic; Community Solidarity; Social Networks; Self–help Organizing; Grassroot Movements; Civic Engagement; Urban Informality; New York City

**University**: Columbia University

**Place**: Urban Planning

**Date**: 2022

**# of Pages**: 2024–09–30

**Language**

**Short Title**

**URL**

**Accessed**

**Archive**: sa3862

**Loc. in Archive**: rtd2101

**Library Catalog**

**Call Number**: M.S.

**Rights**

**Extra**
Pain points

- Finding the balance where we:
  - use the metadata the partner has
  - ask students for the information they are willing to provide
  - are spending minimal time on individual items
  - are processing bulk deposits on a reasonable schedule

- Tweaking the workflow to meet partners where they are

- Having the workflow be understandable and usable by different libraries staff
Next steps

- Work with GSAS (Graduate School of Arts & Sciences) to ingest a batch of dissertations
- Continue to document and refine workflow for other doctoral degree programs, for our Master’s and undergraduate degree program partners, and for Libraries staff
- Assess the amount of time used for this new workflow
- Reconsider pilot workflow in late 2023
Caveat

While this solution will meet the immediate needs of GSAS and accommodate their desire to stop using the ProQuest platform, Zotero was not created to manage electronic theses and dissertations, and is not a long-term solution for the deposit of theses and dissertations.

Libraries staff are currently reviewing electronic theses and dissertation deposit management across Columbia University and will make recommendations about appropriate software that GSAS may choose to use as a long-term solution.