Herding Cats
Reducing Metadata Noise for Undergraduate Honors Theses

G.W. Swicord, Theses and Dissertations Manager, George A. Smathers Libraries at the University of Florida
USETDA2019 in Charleston, South Carolina
Overview

• Brief history of the IR@UF’s role in electronic undergrad Honors projects at UF
  ◦ Prominent aspects of the organizational environment

• Processes during our first years managing the Honors projects

• Current state of efforts to significantly automate metadata creation and submission processing
Undergraduate Honors Projects in the IR@UF

THE SEMI-FERAL DAYS
How Honors projects are produced at UF

- **UF is highly decentralized**
  - Unlike graduate ETDs, Honors projects are required/accepted by individual academic units
  - Communication with many units can be slow and complicated
  - Responses to centralizing efforts can be very different across units

- **Deliverable standards, deadlines, and formal acceptance processes vary**
  - Some terminal projects are non-traditional, e.g. web portfolios, while others are laid out like a journal galley
  - Some projects are term papers for a specific Honors-track course, others are handled like master's theses and require defense before a committee
How we ended up hosting Honors Theses

• Electronic Honors theses first managed by Honors Program in 2009
  ◦ They used a vended content system to accept full theses and abstract-only submissions

• 2013 –Director of Honors Program and Library Dean met
  ◦ Decision taken to transfer responsibility to the Libraries
    ▪ Theses are required by departments/programs, not directly by Honors Program
  ◦ Made sense to move to IR@UF
    ▪ The IR@UF already hosted graduate ETDs
How we ended up hosting Honors theses

• 2014 – Students began submitting directly to the Libraries
  ◦ Submissions came in email with project file(s) and paperwork files attached

• 2016 – Honors Program website redesigned
  ◦ Theses database no longer supported
  ◦ Transferred legacy files and records to library for inclusion in IR@UF
2014-2017: Learning the lay of the land

THE FREE RANGE YEARS
Volume of email interactions was very labor intensive

- Questions before, during, and after submitting
- Problems with submission metadata (departmental forms)
  - Widely varied spellings of colleges, departments, and majors
  - Mixed success with guidance to submit typed data or forms completed in Acrobat
- Problems with thesis/project files themselves
  - Scanned printouts instead of PDFs saved from Word
  - Revisions and comments present
Undergraduate Honors ETD metadata

- In early semesters, Honors metadata was limited
  - Title, type, author, publication date, poor keywords, and abstract
- In 2016, we began enriching the metadata with info drawn from student records
  - We used Access and Excel to make the info resemble graduate metadata
2018-2019: Getting the Ranch in Shape

PLENTY OF MANUAL WORK LEFT TO DO
Trial online submission form using Qualtrics

- Proof of concept for students adopting online submission form
  - Intended to reduce manual effort of copy-pasting from email submissions
  - Design process helped us realize we should ignore college/department form structure

- Metadata data creation still very laborious due to data cleanup time for Qualtrics output
  - All information still manually entered by students, spelling still required extensive manual cleanup
Custom web form backed by SQL database

• In fall 2018, we finally rolled out the initial version of our online submission form

• Benefits of the form include:
  ◦ Connection to student information system
    ▪ Colleges, departments, majors spelled consistently
    ▪ Final Honors level available without need to ingest spreadsheet from Records Office
  ◦ Interactive guidance/control
    ▪ Branch points provide paths for students with less common circumstances, e.g. co-authors, multiple project advisors
    ▪ Students cannot complete the form without entering their parts of the metadata and uploading at least one file (or indicating that they need help with a very large file)
Work still in progress

Next development steps include:

- Self-service portal to eliminate many requests for IT services
- Connection to manually-maintained table of project advisors
- Refinement of Microsoft Access expressions used to generate metadata
  - Automate (at least partially) processing multi-major students
  - Distinguish single-advisor from multi-advisor students
    - To avoid leader tags such as “Advisor(s):” in metadata content
Thank you for attending!

• Our undergraduate Honors Projects are available at [http://ufdc.ufl.edu/ufhonors](http://ufdc.ufl.edu/ufhonors)

• PowerPoint and PDF versions of this presentation will be available at [https://ufdc.ufl.edu/l/AA00068921/](https://ufdc.ufl.edu/l/AA00068921/)
  ◦ I hope to add a set of notes as a download on that item

• My contact info:
  G.W. Swicord
  gwswicord@ufl.edu
  352-273-2763