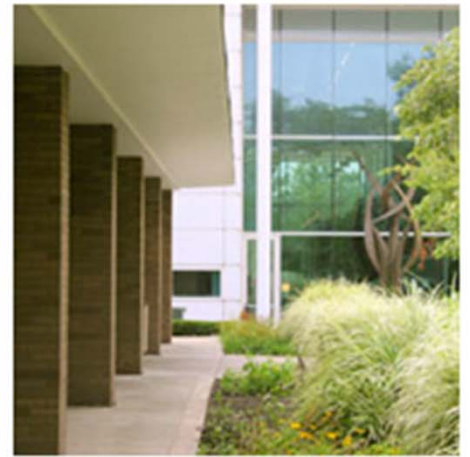


(Semi) Automated ETD Ingest into Digital Commons

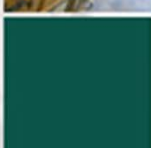
WAYNE STATE
UNIVERSITY
LIBRARY SYSTEM

Joshua Neds-Fox, Coordinator for Digital Publishing | August 3, 2012



ETD Ingest

The problem: get ETDs from ProQuest and archive them in DigitalCommons@Wayne



ETD Ingest

Step 1: Masters and PhD candidates deliver ETDs to ProQuest, choose IR permissions online

ETD Ingest



WAYNE STATE UNIVERSITY

UNIVERSITY LIBRARIES NEF LAW SHIFFMAN MEDICAL REUTHER LIBRARY

LOGIN ▾

UNIVERSITY LIBRARIES

- RESOURCES
- SERVICES
- LIBRARY INFO
- QUICKLINKS

ELECTRONIC THESIS AND DISSERTATION PERMISSIONS FORM

Wayne State University Libraries would like to permanently archive and disseminate an electronic version of your Thesis or Dissertation to the Wayne State University community and general public using [Digital Commons@WSU](#). This form controls the level of access you would like the WSU community and/or general public to have to your Dissertation or Thesis in DigitalCommons@WSU. **Please read the agreements below carefully and consult the members of your advisory committee before submitting this form.**

[...for answers to FAQs.](#)

[...for more information about Digital Commons@WSU.](#)

Please direct questions or comments about this form and Digital Commons@WSU to digitalcommons@wayne.edu.

First Name: *

Last Name: *



ETD Ingest

Step 2: ProQuest delivers ETDs – PDF and XML – to FTP server, emails WSU

ETD Ingest

Step 3: Given a reasonable # of ETDs, WSU runs shell script.



ETD Ingest

- Unzip ETDs
- Concatenate XML
- Replace char. Encodings
- Determine DC article IDs
- XSLT ProQuest > Bepress XML
- Split into Diss / Thes XML upload files
- XSLT ProQuest > Bepress > Marc XML > Marc Records
- Copy PDFs to web-accessible folder
- Archive ZIP files

ETD Ingest

**Step 4: Batch XML
upload to DC@WSU**



ETD Ingest

Step 5: Batch revise uploads to encode permissions choices (doctype, embargo date); upload MARC to catalog



ETD Ingest

Step 6: Record new ETDs, permissions in tracking spreadsheet; add tickler to embargo end date on calendar.



ETD Ingest

**Step 7: Batch
revise series when
embargos end to
change doctype.**

ETD Ingest

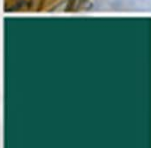
Where it can fail:

- Shell script: XSLT fails on non-standard chars.
- Batch upload fails on non-standard “descriptions” of disciplines.

ETD Ingest

What's next:

- Port tracking spreadsheet to MySQL table
- Automate updating of revision spreadsheet by querying against permissions dB.



Questions?

Joshua Neds-Fox

Coordinator for Digital Publishing
Wayne State University Libraries

jnf@wayne.edu

313 577 4460

