CREATING A REPOSITORY OF OLD EXAMS WITH DSPACE

Mariya Maistrovskaya
Rachel Wang
University of Toronto Libraries

DSpace North American User Group Meeting
September 2019
Old Exams Repository

Search or browse the 3 most recent years of exams supplied by the Faculty of Arts and Science (St. George), UTM, and UTSC.

Search by course code, campus, department

Browse Exam Collections

- 5,500 exams available from 3 U of T campuses
- 3 most recent years of final exam papers (no answers)
- Restricted to campus access
- Batch generated metadata and batch loaded
## DSpace for IR vs exams

<table>
<thead>
<tr>
<th>Institutional Repository</th>
<th>Exams Repository</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Open access</strong></td>
<td>Campus-restricted access</td>
</tr>
<tr>
<td><strong>Long-term storage</strong></td>
<td>Retention schedule &amp; removals</td>
</tr>
<tr>
<td><strong>Permanent URL (handle / DOI)</strong></td>
<td>No permanent URL</td>
</tr>
<tr>
<td><strong>Varied metadata &amp; object types</strong></td>
<td>Uniform metadata &amp; single object type</td>
</tr>
</tbody>
</table>
Where do exams come from?

Establishing partnerships with:

- Registrar’s office
- University archives
- Campus libraries
- Campus IT
- Individual departments and faculty members
Restriction and retention

At U of T the balance between student and faculty interests is achieved by:

- Only sharing the **last 3 years of exams** (when a new batch is received, the oldest one is removed)
- **Opt-out model** where faculty can request to restrict their exams from being posted (FAS, UTM)
- **Opt-in model** at the UTSC campus where faculty send in exams ad hoc
Exam processing pipeline at U of T
1. Scanning and filenaming

- Print exams are received as a batch once a year (FAS) and 3 times a year (UTM)
- Exams are scanned into PDF with file names:

```plaintext
St. George / UTM Exam Filename Convention
CSC490H1-d17-L5101.pdf

Course code
Department
Exam date
Section (optional)

UTSC Exam Filename Convention
CSCD84H3-d17-LEC01.pdf

Course code
Department
Exam date
Section (optional)
```
2. Metadata generation

- Dublin Core metadata is generated from the file names using a Python script with Beautiful Soup library
- The script also uses a CSV spreadsheet of departmental codes per campus for mapping

```python
def get_depts_mapping(csv_filename):
    mapping = {}
    with open(csv_filename) as raw_csv:
        for row in csv.DictReader(raw_csv):
            mapping[row['Dept_code']] = row['Department']
    return mapping
```
3. DSpace simple archive

- A different Python script is used to package the PDFs and metadata into DSpace simple archives for ingest.
4. Batch import into DSpace

- DSpace simple archives are imported into their respective collections via batch import
- Collections older than 3 years old are removed
The University of Toronto Mississauga and you, as a student, share a commitment to academic integrity. You are reminded that you may be charged with an academic offence for possessing any unauthorized aids during the writing of an exam. Clear, sealable, plastic bags have been provided for all electronic devices with storage, including but not limited to: cell phones, SMART devices, tablets, laptops, calculators, and MP3 players. Please turn off all devices, seal them in the bag provided, and place the bag under your desk for the duration of the examination. You will not be able to touch the bag or its contents until the exam is over.
Scripts and slides

UTL’s scripts for the Old Exams Repository:
https://github.com/utlib/dspace-exams-ingest-scripts/

These slides:
http://hdl.handle.net/1807/96626
QUESTIONS

Mariya Maistrovskaya
Digital Publishing Librarian
University of Toronto Libraries
mariya.maistrovskaya@utoronto.ca