# Creating and Managing a Repository of Past Exam Papers

*Mariya Maistrovskaya and Rachel Wang*

**ABSTRACT**

*Exam period can be a stressful time for students, and having examples of past papers to help prepare for the tests can be extremely helpful. It is possible that past exams are already shared on your campus—by professors in their specific courses, via student unions or groups, or between individual students. In this article, we will go over the workflows and infrastructure to support the systematic collection, provision of access to, and repository management of past exam papers. We will discuss platform-agnostic considerations of opt-in versus opt-out submission, access restriction, discovery, retention schedules, and more. Finally, we will share the University of Toronto setup, including a dedicated instance of DSpace, batch metadata creation and ingest scripts, and our submission and retention workflows that take into account the varying needs of stakeholders across our three campuses.*

**BACKGROUND**

The University of Toronto (U of T) is the largest academic institution in Canada. It spans across three campuses and serves more than 90,000 students through its 700 undergraduate and 200 graduate programs.[1]

The University of Toronto structure is the product of its rich history and is thus largely decentralized. As a result, the management of undergraduate exams is carried out individually by each major faculty at the Downtown (St. George) Campus, and centrally at the University of Toronto Mississauga (UTM) Campus and the University of Toronto Scarborough (UTSC) Campus.

The Faculty of Arts and Science (FAS) at the St. George Campus has traditionally made exams from its departments available to students. In the pre-internet era, students were able to consult print and bound exams in departmental and college libraries' reference collections. With the rise of online technologies, the FAS Registrar's Office seized the opportunity to make access to past exams more equitable for students and worked with the University of Toronto Libraries (UTL) Information Technology Services (ITS) to digitize and make exams available online. They were initially shared electronically via the Gopher protocol and later via Docutek ERes, one of the first available course e-reserves systems.

After the UTL became an early adopter of the DSpace ([https://duraspace.org/dspace/](https://duraspace.org/dspace/)) open source platform for its institutional repository in 2003, the UTL ITS created a separate instance of DSpace to serve as a repository of old exams. The repository makes the last three years of exams from the FAS, UTM, and UTSC available online in PDF. About 5,500 exam papers are available to students with U of T login at any given time.

Discussed below are some of the considerations in establishing and maintaining a repository of old exams on campus, along with practical recommendations and shared workflows from the UTL.

**Mariya Maistrovskaya** ([mariya.maistrovskaya@utoronto.ca](mailto:mariya.maistrovskaya@utoronto.ca)) is Digital Publishing Librarian, University of Toronto. **Rachel Wang** ([rachel.wang@utoronto.ca](mailto:rachel.wang@utoronto.ca)) is Application Programmer Analyst, University of Toronto.

**CONSIDERATIONS IN ESTABLISHING A REPOSITORY OF OLD EXAMS**

If you are looking to establish a repository of old exams, these are some of the considerations to take into account when planning a new service or evaluating an existing one.

*The Source of Old Exams*
Depending on the level of centralization on your campus, exams may be administered by individual academic departments or submitted by instructors/admins into a single location and managed centrally. The stakeholders involved in this process may include the office of the registrar, campus IT, departmental admins or libraries, etc. Establishing a relationship with such stakeholders is key in getting access to the files. When arranging to receive electronic files, consider whether they could be accompanied with existing metadata.

Alternatively, if the university archives or records management already receive copies of campus exams, you may be able to obtain them there. Print versions will need to be digitized for online access—later in this article we will share metadata creation strategies in this scenario. It is also possible that exams may be collected in less formal ways, for example, via exam drives by student unions and groups.

The UTL works closely with the FAS Registrar's Office to receive a batch of exams annually. The UTL receives a copy of print FAS exams that get digitized by the ITS staff. The UTL also receives exams from two U of T campuses, UTM and UTSC, that arrive in electronic format via the campus libraries. The U of T Engineering Society and the Faculty of Law each maintain their individual exam repositories, and the Arts and Science Student Union maintains a bank of term tests donated by students.

*Content Hosting and Management*
One of the key questions to answer is which campus department or unit will be responsible for hosting the exams, managing content collection, processing and uploads, and providing technical and user support. These responsibilities may be within the purview of a single unit or may be shared between stakeholders. Here are some examples of the tasks to consider:

1. Collecting exams from faculty or receiving them from a central location
2. Managing restrictions (exams that will not be made available online)
3. Digitizing exams received in print
4. Creating metadata or converting metadata received with the files
5. Uploading exams to the online repository
6. Removing exams from the online repository
7. Providing technical support and maintenance (e.g., platform upgrades, troubleshooting)
8. Providing user support (e.g., assistance with locating exams)

At U of T, tasks 1–2 are taken care of by Registrar Offices at FAS and UTM and by the Library at UTSC. Tasks 3–8 are performed centrally by the UTL ITS, with the exception of digitization services for exams received from the UTM and UTSC campuses.

Further details and considerations related to the content management system and processing pipelines are outlined in the "Infrastructure and Workflows" section below.

*Collection Scope*

Depending on the sources of your exams, you may need to establish the scope rules for what gets included in the collection. For example:

- Will you only include final exams? Will term tests also be included?
- Will solutions be posted with the exams?
- Will additional materials, such as course syllabi, also be included?

At the UTL, only final exams are included in the repository, and no answers are supplied.

*Exam Retention*

Making old exams available online is always a balancing act between the interests of students who want to have access to past test questions and the interests of instructors who may have a limited pool of questions to draw from or who may teach different course content over time and want to ensure that the questions continue to be relevant.

At the UTL, in consultation with campus partners, the balance was achieved by only posting the three most recent years of exams in the repository. As soon as a new batch is received, the UTL removes a batch of exams more than three years old.

*Opt-In versus Opt-Out Approach*

Where exam collection is driven centrally by a registrar's office, for example, that office may require that all past exams be made available to students. Similarly to the retention considerations, the needs of instructors who draw questions from a limited pool can be accommodated via opt-outs, individual exam restrictions, and ad hoc take-down requests.

An alternative approach to exam collection would be an opt-in model where faculty choose to submit exam questions on their own schedule.

At the UTL, the FAS and the UTM campus both operate under the opt-out model. The UTL receives all exam questions in regular batches unless they have been restricted by instructors' requests. Occasional withdrawal requests from instructors require an approval from the Registrar's Office.

Conversely, the UTSC campus operates under the opt-in model where individual departments submit their exams to the library. While this model provides the most flexibility, the volume of exams received from this campus is subsequently relatively small.

*Repository Access*

When making old exams available online, one of the things to consider is who will have access to them. Will the exams only be available to students of the respective academic department, or to all students, or to the general public? Will access be possible on campus as well as off campus? If the decision is made to restrict access, is there an existing authorization infrastructure in place that the repository could take advantage of, such as an institutional single sign-on or library's proxy access?

At the UTL, access to the Old Exams Repository is provided through EZProxy in the same fashion as subscription resources made available via the library.

*Discoverability and Promotion*
How will students find out about the exams available in the repository? Will the repository be advertised via the library's website, promoted by course instructors, or linked with the other course materials? Considering the challenge of promoting a resource like this along with a variety of other library resources, it will be preferable to make it known to students via the same channels through which they receive other course information. For many institutions this would be via their learning management system or their course information system.

At U of T, the Old Exams Repository is linked from the Library website. Previously, the link was embedded in the university's learning management system course template. With a recent transition to a new learning management engine, such exposure is yet to be reestablished.

**INFRASTRUCTURE AND WORKFLOWS**

*Minimum CMS Requirements*
A repository of old exams does not require a specific content management system (CMS) or an off-the-shelf platform. Your institution may already have all the components in place to make it happen. Here are the minimum requirements you will want to see in such a system:

- File upload by staff (preferably in batch)
- File download by end users
- Basic descriptive metadata
- Search / browse interface
- Access control / authentication (if you choose to restrict access)

The UTL uses a stand-alone instance of DSpace for its Old Exams Repository. DSpace is an open-source software for digital repositories used across the globe primarily in academic institutions.

The UTL chose this platform since it was already running an instance of DSpace for its institutional repository (IR) and had the infrastructure and expertise on site. However, this is not a solution we would recommend to an institution with no existing DSpace experience. While DSpace is an open-source platform, maintaining it locally requires significant staff expertise that may not be warranted considering that a collection of exams would only use a fraction of its robust functionality. If you do consider using DSpace, a hosted solution may be preferable in a situation when local IT resources and expertise are limited.

*Distributing Past Exams via an Existing Digital Repository*
An institution that already maintains a digital repository may consider adding exams as a collection to the existing infrastructure. When choosing to do so it is important to consider whether the exams use case may be different from your IR use case, and whether the new collection will fit in the existing mission and policies. Differences may include the following:

- Access level. IR missions tend to revolve around providing openly accessible materials, whereas exams may need to be restricted. Will your repository allow selective access restrictions to the exams collection?
- Longevity. IR materials are usually intended to be kept long-term, whereas exams may be on a retention schedule. For that reason, it also does not make sense to assign permanent identifiers to exams as many repositories do for their other materials.

- File types and metadata. Unlike a variety of research outputs and metadata usually captured in an IR, exams would have uniform metadata and object type. This makes them suitable for batch transformations and uploads.

### Batch Metadata Creation Options

Because of the uniform object type, exams are well suited to batch processing, transformations, and uploads. At UTL, metadata is created from the filenames of scanned PDF files by a Python script.[2] The script breaks up the filename into Dublin Core metadata fields based on the pattern shown in figure 1. See figure 2 for a snippet of the script populating Dublin Core metadata fields.
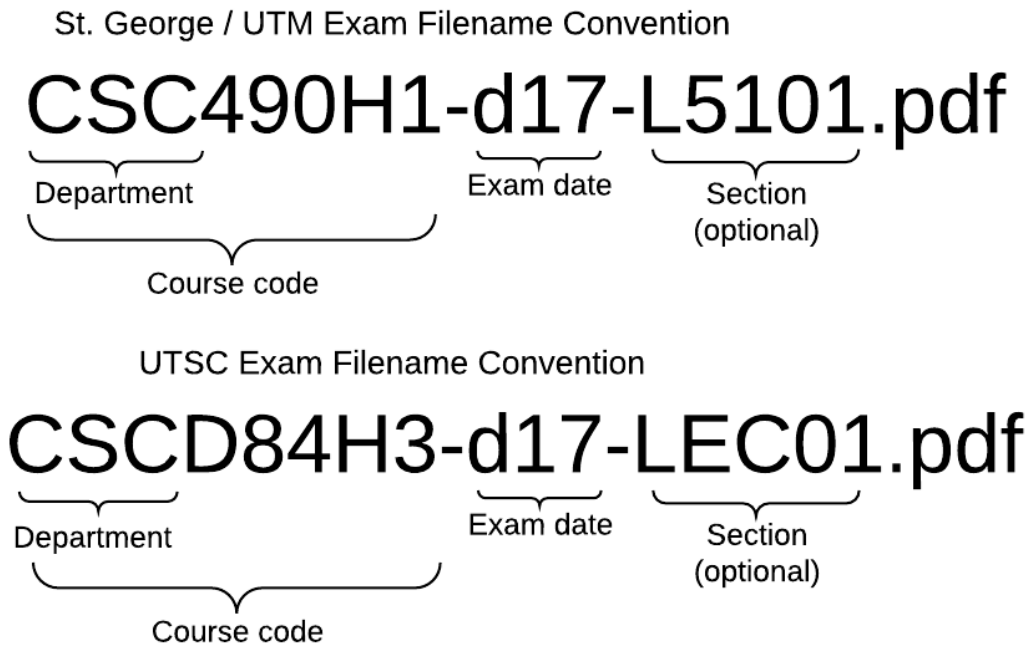


**Figure 1.** File-naming pattern for metadata creation at UTL.



**Figure 2.** A screenshot of the UTL script generating Dublin Core metadata from filenames.

Once metadata is generated, the second Python script (figure 3) packages the PDF and metadata file into a DSpace Simple Archive (DSA) which is the format that DSpace accepts for batch ingests.
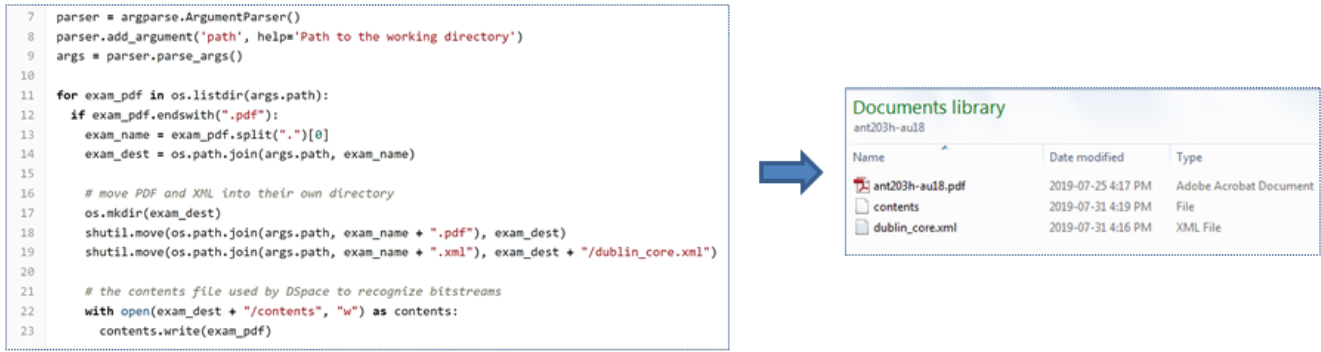
```
7   parser = argparse.ArgumentParser()
8   parser.add_argument('path', help='Path to the working directory')
9   args = parser.parse_args()
10
11  for exam_pdf in os.listdir(args.path):
12      if exam_pdf.endswith(".pdf"):
13          exam_name = exam_pdf.split(".")[0]
14          exam_dest = os.path.join(args.path, exam_name)
15
16          # move PDF and XML into their own directory
17          os.mkdir(exam_dest)
18          shutil.move(os.path.join(args.path, exam_name + ".pdf"), exam_dest)
19          shutil.move(os.path.join(args.path, exam_name + ".xml"), exam_dest + "/dublin_core.xml")
20
21          # the contents file used by DSpace to recognize bitstreams
22          with open(exam_dest + "/contents", "w") as contents:
23              contents.write(exam_pdf)
```

**Figure 3.** A screenshot of the UTL script packaging a PDF and metadata into a DSpace Simple Archive.

The DSpace Simple Archive (DSA) then gets batch uploaded into the respective campus and exam-period collections (figure 4) using the DSpace native batch import functionality. Figure 5 shows what an individual exam record looks like in the repository.

After a new batch is uploaded, collections older than three years are removed from the repository. The UTL's exams processing scripts are openly available in Github under an Apache License 2.0 (https://github.com/utlib/dspace-exams-ingest-scripts/).
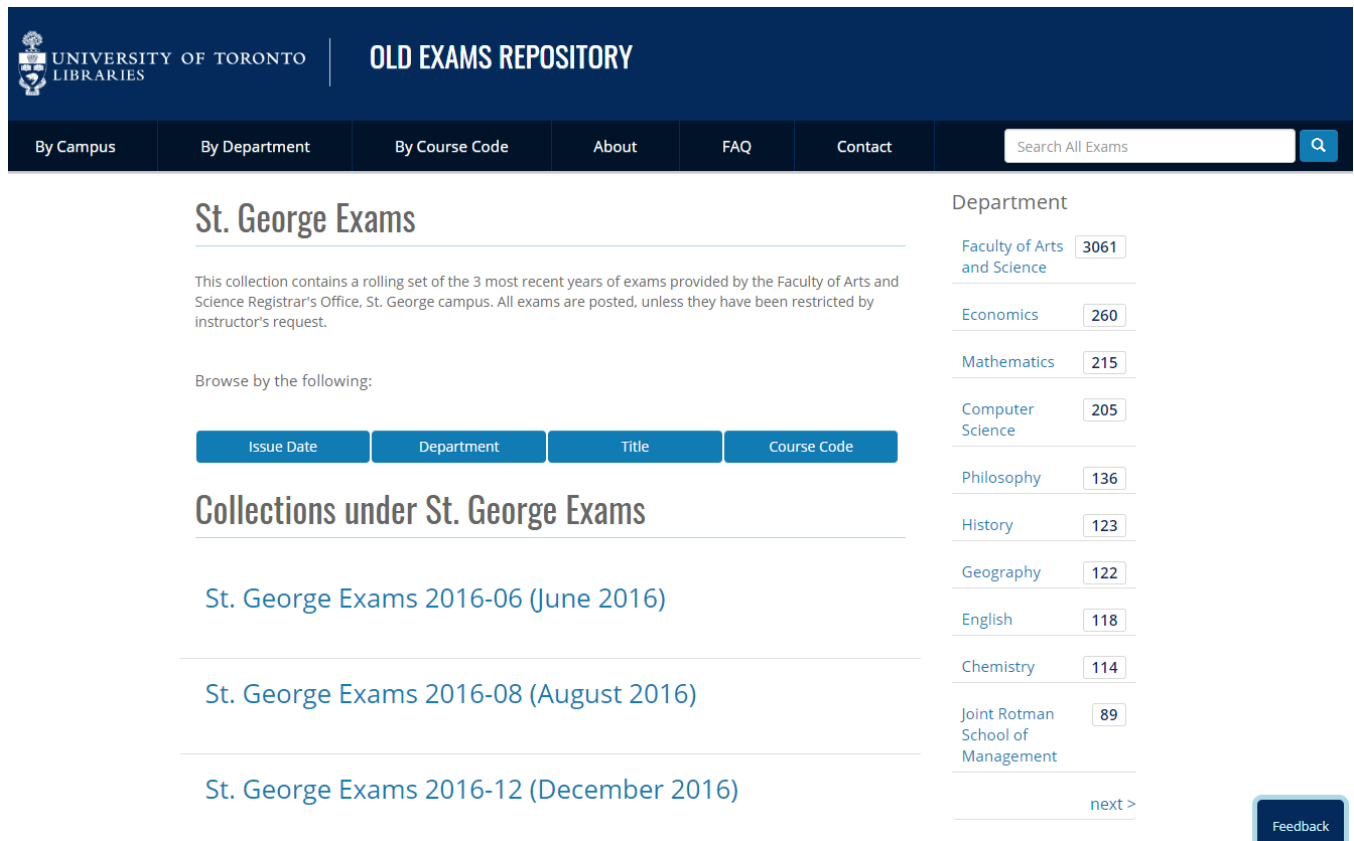
**Figure 4.** A screenshot of collections in the UTL's Old Exams Repository.

**Figure 5.** A screenshot of a record in the UTL's Old Exams Repository.

**CONCLUSION**

Having access to examples of past exam questions can be extremely helpful to students in preparing for upcoming tests. It is possible that old exams are already being shared on your campus in official or unofficial ways, in print or electronically. Facilitating online sharing of electronic copies means that all students, on and off campus, will have equitable access to these valuable resources. We hope that the considerations and workflows outlined in this article will help institutions establish such services locally.

**ACKNOWLEDGEMENTS**

The authors would like to acknowledge the UTL librarians and staff who contributed to the setup and maintenance of the Old Exams Repository over the years: Marlene Van Ballegooie, metadata technologies manager, who operated the filename-to-Dublin Core metadata crosswalk; Sean Xiao Zhao, former applications programmer analyst, who converted it into Python; and Sian Meikle, associate chief librarian for digital strategies and technology, who was at the inception of the original exam-sharing service and provided valuable historical context and feedback on this article.

**ENDNOTES**

[1] University of Toronto, "Quick Facts," accessed November 4, 2019, https://www.utoronto.ca/about-u-of-t/quick-facts.

[2] University of Toronto Libraries, "Exam Metadata Generation and Ingest for DSpace," GitHub Repository, last modified September 20, 2019, https://github.com/utlib/dspace-exams-ingest-scripts/.