

Implementing an Open Access Policy:

Building the Workflow to Identify and Gather Content

Mathew Willmott (willmott@mit.edu) and Ellen Duranceau (efinnie@mit.edu)

Background

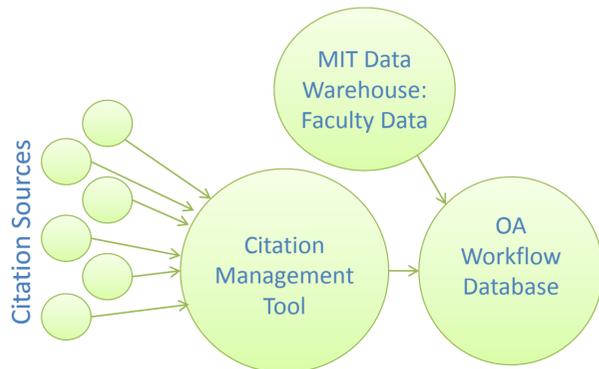
We describe a workflow process that we built to support the implementation of the MIT Faculty Open Access Policy (approved in March 2009). This process is applicable to any campus that aims to locate and gather content for populating a repository whether or not the campus has an open access policy in place.

The Policy grants MIT nonexclusive permission to exercise all rights under copyright, provided articles are not sold for a profit; authors may opt-out of the Policy on a per-paper basis. The MIT Libraries administer the Policy under the guidance of the Faculty Committee on the Library System. DSpace@MIT, MIT's repository, was chosen as the vehicle for the open dissemination of papers under the Policy.

What did the Policy mean for the Libraries? We needed to:

- ...find out what papers MIT faculty are publishing
- ...find or create a system that could keep track of bibliographic data about these papers and our progress on obtaining them
- ...have the capability to track and assess our progress
- ...get started right away
- ...make the process as convenient for faculty as possible

Gathering bibliographic data



The process:

- Gather citations to MIT-authored papers from several different licensed databases. (MIT does not maintain a central database of faculty publications.)
- Import all citations to a citation management tool using existing filters; normalize and check for duplicates.
- Export citations from the citation management tool and import them into a custom-built relational database (known as the OA Workflow Database); this database is a prototype used to get the process started and identify system requirements.
- Import names and affiliations of faculty members into this database.
- Run a script to match each faculty member with papers they authored.

The benefits:

- Resource savings
 - Used many of our existing tools
- Flexibility
 - Workflow customization
 - Outputs of bibliographic data
- Speed
 - Up and running quicker than expected
 - Rapid cycles of iterative development
- Accountability
 - Full control of data collection for assessment

Data flow management

The flexibility of a custom system allowed us to create our own displays and data structures as needed:

- Workflow controls were added to mark our progress on obtaining each paper
- Notes were added to track our interactions with particular faculty members
- An automatically generated email was created which could be used to contact a faculty member about outstanding papers
- Tables of publishers and journals were created to track various policies with respect to author rights, harvesting, version requirements, etc.
- Queries were written to generate statistics on various facets of the project, allowing for a thorough assessment of the process.

Obtaining and depositing papers

Acquiring from Publishers

Several publishers have given us permission to download papers directly from their platforms for deposit into the OA collection.

We use three methods for this:

- Exporting bibliographic data from OA Workflow Database to citation management tool, where we use a "Find Full Text" service to download papers en masse from many publishers.
- Using the SWORD protocol to automatically deposit papers from BioMed Central.
- Manually visiting publisher sites to download any papers not otherwise obtained.

Contacting Faculty

For papers we cannot not obtain from publishers, we need to request the final author's manuscript from the authors.

We use the OA Workflow Database to support this process:

- We send messages generated through the database to each author, once per year. Each message contains a list of papers needed from that author.
- The OA Workflow database automatically removes articles from the author's list that we can obtain from publisher, or have already received, or which have been opted out; database also offers capability to label papers authors can supply in final published form.
- Subject librarians send centrally-built outreach messages.

Using Other Sources

While building each message generated by the OA Workflow Database, we check other sources for the author's papers. Through analysis, we determined that the MIT web domain and certain subject repositories housed many relevant papers, which in some cases we have permission to copy.

We seek appropriate and legal copies from:

- The MIT web domain
- arXiv
- Social Science Research Network

Adding Metadata

Adding metadata is a significant value-added service that allows readers to distinguish versions of papers, cite and link to the final published version, and captures controlled forms of MIT names, among other benefits.

We add this data in two stages:

- The acquisitions team assigns license, version, and peer-review related metadata as well as the DOI (or URL) for the published version.
- The metadata team adds bibliographic metadata, using authority control for MIT names and departments; builds full citation; and adds grant numbers and research sponsor names.



Next steps

- **Review commercial or open source options** in order to make an informed decision on whether to develop a more robust custom solution or adopt an existing product.
- **Develop dual repository deposit mechanisms** so that authors who post their work in other repositories only need to submit their paper in one location.
- **Expand use of SWORD protocol** to support automated deposit of papers from more publishers.
- **Expand reuse of automatically harvested metadata** to increase efficiency in cataloging and depositing papers.