Gaining Intellectual Control over Technical Reports and Grey Literature Collections

Abstract

This paper will describe the solution that the author employed to facilitate uniform, online, web-based access to detailed information about the library’s various collections of technical reports and grey literature. By working closely with colleagues from the University Archives, the Engineering Library adopted the Encoded Archival Description (EAD) standard (typically used to provide standardized, digital description of archival and manuscript collections) and applied it to create machine readable finding aids for all departmental research and technical reports collections of the School of Engineering. The Archivists’ Toolkit (AT), an open source application, has been used to create bibliographic descriptions, to establish name and subject entries, to manage locations and to export EAD finding aids and MARCXML records for technical reports collections. The project has been completed in house, using existing resources; given the increased number of requests the library received for these materials since the completion of the project, it is considered a success. By creating descriptive bibliographies (finding aids) that incorporate metadata standards, materials that were once hidden and lost to researchers, are now easily discovered and used.

Introduction

Even without financial pressures on library budgets, collections of technical reports typically don’t rank too high on priority lists for processing and inclusion in libraries’ holdings. Technical reports however, always found their way on library shelves or cabinets where they would be filed for easy physical browsing and discovery by library users. With the ongoing pressures on library spaces resulting in libraries being consolidated and closed, these collections have often ended up in remote storage locations waiting for more favorable circumstances when technical services operations could afford to invest the time and human resources to catalog and process them. While initiatives such as the Technical Reports Archive and Image Library (TRAIL) provide viable solutions to preserve and make available these very important resources, libraries still have to contend with materials that do not fit the scope of TRAIL or other collaborative digital initiatives. This paper will focus on the solution used to gain bibliographic and intellectual control over the collection of departmental technical reports (a small section of the larger grey literature collection) in the context where on-site browsing of shelves to find the necessary material is no longer feasible or deemed an acceptable way of responding to users’ needs to locate and use these materials.

Technical Reports and Grey Literature Collections in Academic Libraries

In the past, one would walk into any technical library, be it academic or special and ask the person at the desk about the reports or grey literature section. Once pointed to the specific floor or area, the researcher would browse the cabinets where reports (issued in print or on microfiche or on microfilm) would be filed by issuing agency, department or corporate entity and then by what is commonly known as the Technical Report (TR) number. Occasionally, the library would
maintain inventory lists for these collections, or even better, maintain local catalogs. However, years of technological progress and innovations in library systems have often missed the local databases used to catalog and maintain technical reports holdings in libraries, and we find today that libraries have little bibliographic control over these collections. In addition to this, with increased transition to electronic collections and space configurations in libraries, these materials, typically of odd size and format (oversized, with insets and fold-outs), have slowly been moved to storage areas, consolidated with other materials of same format (i.e. microforms), or incorporated in the library’s Government Documents collections, making it even more challenging for users to locate and retrieve these materials.

These collections used to represent a significant fraction of a technical or science library’s collection and that was the case until 2001 for the Engineering Library at Princeton University. The Engineering Library’s collection of technical reports consists of materials received over years on the Federal Depository Library Program (FDLP), non-depository materials (materials received through exchange programs with other institutions) and departmental technical reports. Until 2001, when the library moved to a new building, all these materials were available in one physical location with on-site reference assistance to users. In 2001, when the library moved to its new building, only a small part of the reports collection was transferred to the new facility. The FDLP print reports have been moved to an off-site remote storage with retrieval capabilities for on campus use; the FDLP microform reports have been partly incorporated with the Government Documents collection and partly moved to an on-campus storage facility. The only unit of the reports collection retained at the library was the departmental technical reports.

The library has collected departmental reports issued by departments, laboratories and centers affiliated with the School of Engineering and Applied Science (SEAS) from the late 1940s to the late 1990s. Today, SEAS at Princeton University consists of 6 departments and 6 interdisciplinary centers but the school evolved over the years and that evolution is reflected in the departmental affiliation noted for each publication in the collection. Many of the centers, departments or laboratories that issued reports during that period of time are no longer in existence, either due to consolidations or due to completion of projects. SEAS reports were produced mainly as a record of publicly funded research undertaken at the University. While results of the research were often published in peer-reviewed literature, the reports frequently contain results of experiments, computations and primary data that are not included in the published literature. Some reports, especially those in high demand, have been cataloged individually and holdings information has been accurately maintained for retrieval purposes. Yet, these are very few cases, as less than 5% of the SEAS departmental reports have full bibliographic information available in the library online catalog. Even without availability information, the Interlibrary Loan Office and the Engineering Library, consistently receive inquiries about reports by Princeton authors that have been cited in the published literature. Researchers in need of these materials would trace the author’s affiliation and send inquiries either directly to the Engineering Library or to the Library’s Interlibrary Services, despite the lack of any type of information in the library catalog about the availability of the material. With limited in house resources to undertake full cataloging work for technical reports, a feasible and sustainable solution was explored to allow the Engineering Library staff to create a window that would allow any user to take a look and determine if what is needed can be found on the library’s shelves.
Grey Literature Collections as Special Collections

A Special Collection typically implies a collection that contains rare, unique and valuable materials that require special handling and a secure location; it may also imply a collection built around a single idea, topic or even a personality. Archives, Manuscripts and Rare Books typically form the core of academic special collections. Such collections are processed based on standards universally accepted by the profession. Levels of processing however differ, and archivists have developed various criteria to describe a collection as “fully processed”, “minimally processed” or “unprocessed”, but these criteria differ from institution to institution. Slotkin and Lynch’s principle that a collection is “processed” whenever it can be productively used for research seems to be still current and practiced by the majority of Association of Research Libraries (ARL) archives.

Applying principles of special collections and archives processing to grey literature collections is not a common procedure, and a search of library literature revealed no examples to demonstrate that the practice exists. However, in building grey literature collections and caring for them, academic librarians show a level of special attention and management analogue to the curatorial management that forms the backbone of Special Collections operations. Moreover, as Gelfand pointed out, “grey literature ages well and, for some users, becomes more valuable over time” which is why research libraries have been committed to collecting and preserving these materials, even though resources for proper bibliographic and intellectual control are limited or nonexistent. Since grey literature collections are built and managed in similar ways to special collections, the project described in this paper aims to demonstrate that well established methods and standards in use for bibliographic control of special collections can be adapted and used for grey literature collections.

Project Specifications

Having a precedent set by the creation and launch of the Project Matterhorn Digital Collection in 2005, where the early publications issued by the Princeton Plasma Physics Laboratory between 1951 and 1958 under the code name Project Matterhorn have been digitized and processed as a special collection, it was decided that this solution will be adapted and applied to the print collections of SEAS departmental technical reports located at the Engineering Library. The goal of the project was to process the technical reports using archival principles and methods, so that they would be available for research use as soon as possible, while digitization of these works would be approached as a different project, at a different time.

The multi-year project coordinated by the Engineering Librarian, started in summer 2007. At that time, discussions were initiated with colleagues from the University Archives to assess the feasibility of archival processing for the SEAS technical reports. These discussions have proved to be extremely informative and educational, especially for Engineering Library staff that typically has little if any interaction with the library’s Special Collections, Manuscripts and Archives units. From the discussions held, it was learned that archivists use the term processing to refer to activities of arranging, describing and preserving archival materials. Processing allows researchers access to the collection through a variety of means. After processing is
Describing the contents of a collection is the most important aspect of processing, as description provides information on the context in which materials were created, as well as relating the physical characteristics and the intellectual contents. Accurate description allows researchers to determine whether the materials are relevant to their research.

As a result of the consultations and discussions that took place, the following Technical Reports Processing Plan has emerged:

- A separate Engineering Library Finding Aids repository, named Technical and Scientific Reports will be created on the Library’s Finding Aids site: http://diglib.princeton.edu/ead/search?pi=eng&sortkey=creator where finding aids for all Engineering Technical Reports collections will be stored.
- The processing undertaken by the Engineering Library will consist of only arrangement and description of materials. No preservation activities have been deemed necessary.
- Archival arrangement is based on two core principles, provenance and original order. In accordance with these principles, distinct collections will be formed around the issuing body (department, project or program) responsible for the work recorded by the reports.
- A finding aid based on the Encoded Archival Description (EAD) standard will be created for each collection of reports by using Archivists’ Toolkit (AT), an open source application widely used by the archival community.
- The online finding aid will describe the collection and list individual reports in each collection, recorded in the order of issuance.
- A collection record will be created for the online public catalog directing researchers to the collection finding aid (online), which gives more explicit information about the collection. MARC records (in XML encoding) for the public online catalog will be generated using AT.
- The collection record will contain the following metadata:
  - Call Number, Repository, Title and Dates, Extent (size), Creator (issuing body), Scope Content (collection description), Subject(s), Access, Language of Materials
- The metadata used to describe each item (report) consists of Author, Report Title, Date and Report Number information.

Adapting Archivists’ Tools to Engineers’ Needs

The Archivists’ Toolkit is an open source relational database for archival data management developed collaboratively by Five Colleges, Inc., New York University Libraries, and the UCSD Libraries with funding from Andrew W. Mellon Foundation. Its main function is to support the description of archival resources. The system is highly customizable and extremely user friendly, making it very easy to learn and use without any prior knowledge of archival processing principles and terminology. The client application of AT was installed on computers located at the Engineering Library and linked to the newly created repository of finding aids designated as “Engineering Technical Reports and Publications”. The repository was created on an existing MYSQL backend database on a Windows server (Fig. 1).
The client software facilitates the data entry workflow for creating collection records and item records. For the purpose of the work described in this paper, only the AT client software was used by the Engineering Library staff. All database related maintenance and installations have been performed by the Library Systems (IT) staff.

AT supports the creation of “resource records” and “component records” (this is standard archival terminology). The resource records created are the collection records (Fig. 2) and the component records are the item records describing the reports that form each collection of reports (Fig. 3).
For each collection record, the fields under Basic Description, Names & Subjects, Notes, Etc. & Deaccessions and Finding Aid Data (Fig. 4) have been filled following the guidelines described in Appendix 1.
Once the collection record has been created, component (item) records have been added for each report by entering data under Basic Description and Names & Subjects following guidelines described in Appendix 1 (Fig. 5).

Fig. 5 AT Item Record Components

After all data has been entered in the AT database, MARC XML records have been exported for each collection, which than have been loaded in the library online catalog. Each record in the online catalog links to the corresponding finding aid for each collection. The finding aids available online have been created using the EAD export feature of AT (Fig. 6).
Project Resources and Timeline

Over the summer months of 2007 (July and August) one full-time student research assistant was hired to work on the project. The main responsibilities assigned to the student assistant were:

- Identify the distinct issuing bodies (labs, centers, programs) for each technical reports series.
- Build collections around the issuing bodies (Creator). Once the collections have been identified, the Engineering Librarian would write a collection description and create a Collection Record using Archivist Toolkit.
- Arrange the contents of each collection (file the reports in a specific collection in order of publication).
- Create Item Records for each report associated with a specific collection using Archivist Toolkit.

In 2008 the process and resources (one student assistant) used in the previous year were employed again for two months (July and August) and 19 more collections were added to the AT database.
To put things into a fiscal context, according to the 2009-2010 ARL Annual Salary Survey, the average annual salary of a cataloger with 1 to 5 years experience in the Middle Atlantic Region was $46,801. Should we have employed the services of a cataloger to process these collections with item level cataloging, it would have required two months of a cataloger’s time or roughly $7800. The students employed for the project over a period of 4 months worked a total of 480 hours. The hourly rate for the students employed was $12/hour, bringing the cost of the salaries paid for the project to $5760. While the salary savings may not be significant, and the project did not produce item level catalog records, the most cost-effective aspect of the project was the ability to tap in and use existing institutional resources. We drew on the expertise of our archivist colleagues to conceptualize the intellectual arrangement of the collections and to set up the processing workflow; we used existing infrastructure resources such as the AT tool and the finding aids repository; and we benefited from existent institutional systems support and IT support.

Project Outcome

At the time of this writing, 11 out of 30 technical reports collections have been fully processed. These finding aids form the core of the Engineering Repository and are publicly available, and open to Internet search. For the remaining 19 collections basic collection records and item records have been created and complete finding aids will be available online by June 2011. The project was successful as a proof of concept demonstration for applying archival processing principles, standards and tools to grey literature collections, therefore, the methodology described here will be applied to process the remaining technical collections at the Engineering Library.

The access logs for the finding aids repository show that since the collections have been opened to Google and other search engines, awareness of the existence of these materials within the research community has increased significantly. Groundbreaking research conducted at the University which lead to technological advancement, such as research done in 1970s by Thomas Sweeney on Sailwing design, is now easily located in the technical reports collections at the Engineering Library. While the reports are not digitally available, they can be easily obtained through regular interlibrary loan services or by direct inquiry with the Engineering Library.

One of the more interesting collections that have received attention from the public since the availability of the finding aid is the Princeton Engineering Anomalies Research (PEAR) Technical Notes. Research conducted at the PEAR laboratory from 1979 to 2007 has been surrounded by controversy and is refuted by the larger scientific community. Nevertheless, ESP and telekinesis research has always been a hot topic and since the closure of the lab in 2007, shortly before the collection finding aid went online, the library has received a constant flow of requests for copies of the reports. Peer reviewed journals have systematically refused to publish papers from PEAR researchers, which is why the technical notes collection at the Engineering Library is even more valuable. Table 1 summarizes the yearly access data for the most used finding aids from the Engineering Technical Reports and Publications Repository (as of January 2011).
Table 1. Top Viewed Finding Aids

<table>
<thead>
<tr>
<th>Finding Aid</th>
<th>2008</th>
<th>2009</th>
<th>2010</th>
<th>Total views</th>
</tr>
</thead>
<tbody>
<tr>
<td>Department of Mechanical and Aerospace Engineering</td>
<td>0</td>
<td>346</td>
<td>436</td>
<td>782</td>
</tr>
<tr>
<td>Technical Reports</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Department of Civil Engineering and Operations</td>
<td>18</td>
<td>205</td>
<td>174</td>
<td>397</td>
</tr>
<tr>
<td>Research Reports</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Department of Civil and Geologic Engineering Research</td>
<td>30</td>
<td>186</td>
<td>116</td>
<td>332</td>
</tr>
<tr>
<td>Research Reports</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Department of Civil Engineering. Engineering-Economics</td>
<td>0</td>
<td>105</td>
<td>116</td>
<td>221</td>
</tr>
<tr>
<td>Series Reports</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Princeton Engineering Anomalies Research</td>
<td>0</td>
<td>11</td>
<td>92</td>
<td>103</td>
</tr>
<tr>
<td>Technical Notes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Conclusions

Most academic libraries hold treasures such as the PEAR collection, but they may get little if any visibility in the outside world. When grey literature collections are treated as special collections, basic archival principles for defining and describing these collections can be followed and tools used by archivist colleagues can be employed. By using this methodology, successful, cost-effective solutions can be found for bringing these treasures to light. The Archivists’ Toolkit system has proved to be a robust system with strong support from the professional archival community, which continuously contributes to its enhancement. Many archives in academic libraries have adopted it, which makes it readily available to academic technical libraries interested in working collaboratively with archivist colleagues to employ this tool for access and control of technical reports and grey literature collections. Thinking outside the box and exploring and borrowing tools from related professional fields can reveal solutions to some of today’s collection management challenges faced by libraries.
Appendix 1 – Archivists’ Toolkit Instructions for data entry to create a Resource Description Record (Collection Record) and Component Description Record (Item Record)

Data Entry Guidelines

Select Resources.

To add a new series, click on New Record.

Under Basic Description tab, select Level: collection and fill out:
Title: […Department or Creator Name], Technical Reports
Date Begin: Enter the date range of the series.
Date End:
Language Code:
Resource Identifier:
Extent: Enter 1 and select Linear feet in Extent Number, and then enter the number of volumes in the series in Container Summary, e.g. 13 volumes.

Under Names & Subjects tab, click on Add Name Link > Create Name.
Fill out the Corporate Primary Name and Subordinate Names, e.g. Princeton University is the primary name, and Communications Laboratory is the subordinate name.
Select the Name Source and Name Rule. This should follow the LOC Name Authority File and AACR2.

Under Notes, etc. & Deaccessions tab, fill out:
Abstract: A very brief summary of the materials being described (2-3 sentences); usually a bit of biographical or historical information about the department or creator and abridged statements about the scope, content of the collection being described.
Arrangement note: A brief statement about how the content is arranged.
Conditions Governing Access note: Enter “Collection is open for research use.”, unless there are restrictions on the materials.
Conditions Governing Use note: Enter the copyright statement.
Scope and Content note: A brief statement summarizing the range and topical coverage of the described materials. The purpose is to assist readers in evaluating the potential relevance of the materials to their research. It may highlight particular strengths of, or gaps in, the described materials.

Under Finding Aid Data tab, fill out:
Finding Aid Title:
 […] Technical Reports: Finding Aid
Finding Aid Date: Date of Entry
Author: Name of person who entered the record.
Description Rules: Select DACS.
Language of Finding Aid: Finding aid written in English.

To add a new technical report, open the series from the resource list and click on Add Child.

Under Basic Description tab, select Level: item.
Fill out the Title and Date.

Under Names tab, click on Add Name Link.
Click on Create Name to add a new entry. Select Person. Fill out:
Personal Primary Name: Enter the author’s last name.
Personal Rest of Name: Enter the author’s first name, etc.
Name Source: select Local Sources.
Name Rule: select Local.
Or select the author from the existing list, Function: creator, and hit Link.


