

Improving Discovery of Hidden Technical Report Collections

Ms. Linda R. Musser, Pennsylvania State University

Linda Musser is Head of the Fletcher L. Byrom Earth and Mineral Sciences Library at the Pennsylvania State University. She received her B.S. in Civil Engineering and worked in industry prior to receiving her M.S. in Library and Information Science from the University of Illinois.

Improving Discovery of Hidden Technical Report Collections

This paper describes how software tools were utilized to aid in cataloging works issued as part of the U.S. Atomic Energy Commission (AEC) depository program.

From the mid-1940s to mid-1970s the U.S. Atomic Energy Commission (AEC) operated a depository program with selected libraries. The program distributed technical information about atomic energy in various formats – paper, microfiche, microopaques. - and receiving libraries were tasked with making these materials available to the public [1]. While there was some overlap with AEC material distributed via the Federal Depository Library Program (FDLP), the majority of items – hundreds of thousands of titles – were distributed only via the AEC depository program. These library collections remain largely invisible primarily due to the lack of cataloging. Historically, many libraries chose not to add technical reports to their catalogs [2] and most of those collections remain outside of these libraries' online catalogs to this day [3]. This lack of visibility makes these technical reports difficult to locate, which reduces their use and concomitantly their value to the institution.

While it is tempting to save space by discarding these low-use collections, the availability of online repositories such as those by the U.S. Department of Energy Office of Scientific and Technical Information (OSTI) and the National Technical Reports Library (NTRL) are not always acceptable surrogates for having physical copies. In the first place, everything in these repositories is not fully digitized. Secondly, digitized versions are not always of sufficient quality nor are they always complete and access to original or alternate versions is required. Finally, online repositories sometimes go down (or are taken down) so, using the LOCKSS principle (*Lots Of Copies Keep Stuff Safe*), having access to other copies is valuable. In recent years, some progress has been made in addressing the cataloging challenge of technical reports generally and specifically the AEC materials. The Technical Report Archive and Image Library (TRAIL) program scanned and cataloged portions of the AEC materials issued on paper and, in 2019, began to scan those reports issued on microopaques, around 30,000 titles [4]. Much of the AEC collection was issued on microfiche, however, and most of those have yet to be cataloged in any way.

Penn State University participated in the AEC depository program for most of its existence and amassed a very large collection of reports, over 500 linear feet, most of which had been transferred to storage decades ago. The collection contents were recorded on check-in cards, which reflected the format of the individual report received, i.e., paper, microfiche, small microfiche, or microopaque. Over the years, the print reports had been cataloged and added to the online catalog but the other formats remained in storage and invisible to users. In 2018, we began to investigate how to improve access to these AEC materials. Initial efforts focused on the bulk of the stored collection, which had been issued on 4-inch x 6-inch microfiche, however, in support of the work of the TRAIL project, we switched our focus to the smaller formats on microfiche and microopaque.

At the start of the project, we cataloged each individual title in the standard fashion – checking OCLC for existing records and creating descriptive metadata manually if not found. It

quickly became apparent that OCLC had few records for these items and the majority would require individual, i.e., original cataloging. Even moving quickly, it would take decades to complete the project in this manner so we explored ways to automate parts of the process. Several combinations of programs were employed, each adding a level of efficiency to the process of cataloging these reports. The programs used included Microsoft's Power Automate and Excel, Java program language, and MarcEdit. The steps in the project are described below.

Step 1: Have or create an electronic inventory of the items in the collection. In our case, items were stored in small boxes, which were numbered sequentially, but lacked an online inventory. We briefly considered using the check-in cards to create an inventory but found that they were not reliably accurate as a source for report numbers. Boxes were pulled from storage and personnel input report numbers into an Excel spreadsheet.

Step 2: Use a software program to retrieve metadata from an online source. Bibliographic information about these AEC reports is available in several free online databases including NTRL [5], OSTI [6], and the International Nuclear Information System (INIS) Repository of the International Atomic Energy Agency [7]. We chose to use each of these databases to some extent but primarily relied upon OSTI as a metadata source, based in part on its robust API. We experimented with two different methods to extract the metadata from the OSTI database: Power Automate [8] to query the OSTI website interface and the OSTI API [9]. Both methods were successful in retrieving metadata.

Power Automate is part of the Microsoft suite of software and is designed to automate repetitive tasks. A program, called a flow, was created to query the OSTI website interface (Figure 1) using the report number from the Excel spreadsheet. If a result was found, the program copied the OSTI ID number and title to an Excel spreadsheet (Figure 2). (The title information allowed verification that the correct record was retrieved.) The OSTI ID number could then be used to extract metadata from the OSTI database by using MarcEdit (with the appropriate plug-in). While this method essentially queried the OSTI database twice - first to get the OSTI ID and then again with MarcEdit - the program could be run from any PC with access to the Power Automate program. While relatively slow (around 1 record per second), the process was effective and reliable.

Use of the OSTI API offered the potential to obtain OSTI IDs much more quickly since it did not require use of the public OSTI website interface. Indeed, this turned out to be the case however identifying the correct matching OSTI record proved to be a bit more challenging. A Java program (actually several) was written to query the OSTI database using the report numbers from the Excel spreadsheet. Results returned all the bibliographic information about the item (Figure 3). Technically, since the API pulled all the bibliographic metadata, we could have utilized MarcEdit with the retrieved data directly rather than re-querying the OSTI database. However, to maintain consistency in the records we were creating, both methods were designed to retrieve the OSTI ID, followed by use of the MarcEdit program to pull selected metadata elements from the OSTI database.

Step 3: Create MARC records. The MarcEdit program [10] was used to retrieve metadata from the OSTI database using the OSTI ID identified in step 2. Once records were retrieved (Figure 4), various checks were run. For example, superscripts and subscripts were usually encoded using special character strings, e.g., XeF₄ was coded as XeF/sub 4/ or XeF\$sub 4\$, so these extra characters were removed for clarity. Where practical, non-English language information was added as was information about errata and translations. Format data (microfiche, microopaque) was added. Unique collection information was added as was information from which to generate a local call number. In cases where metadata was not available from OSTI, records were created using data manually input into an Excel file, sometimes using metadata retrieved from NTRL or the INIS Repository. Using the MarcEdit Delimited Text Translator tool, MARC records were created from this spreadsheet data.

Various shortcuts were taken in the process of creating these records, largely in order to save time and keep the project moving forward. For example, the MarcEdit plug-in for OSTI does not retrieve all the bibliographic information available, a limitation that we chose to accept. The OSTI metadata does not include information on the format in hand or on the number of pieces of fiche or cards. While we were able to code the format appropriately, we had not asked the inventory personnel to capture the number of pieces so we created records with an unspecified number of pieces. At the start of the project, we had limited our work to the creation of descriptive records only but subject keyword data was captured as were abstracts, if available. When capturing metadata manually, we were even more restrictive in how much data were transcribed, using only the fiche or card header. For example, even if visible, we limited to three the number of personal authors that were added.

Step 4: Upload MARC records into the local catalog and OCLC. This step was handled by personnel in the Cataloging and Metadata Department of the University Libraries. In some instances, duplicate records were identified and required resolution.

Step 5: Perform a quality check on the loaded records. Sometimes fiche or cards were missed in the inventory, so an item by item check was performed after records were loaded into the catalog. This check proved to be useful as it was rare to find a box with no issues that required attention. In some cases, a particular title was missed in the inventory. In other cases, a degree symbol was interpreted as a 'zero'. Some coding of super- and subscripts was missed. As always, the occasional typo appeared – usually due to messy OSTI metadata.

Discussion

Both of the methods tested for retrieving metadata from the OSTI database yielded good results for creating bibliographic records. The OSTI API was the fastest in terms of processing time and completeness of the resulting metadata available, and is the method we continue to use. To date, we have created catalog records for over 149,000 works and estimate we have about 58,000 works remaining to complete the cataloging of this collection. Use of the cataloged materials has improved and continues to grow exponentially. Given this success, we have begun to use this method to pull metadata on the Penn State University collection of U.S. Department of Energy (DE) reports on microfiche in preparation for creating MARC records for the nearly

200,000 reports that are in storage. Beyond the AEC and DE reports, we have had some small successes in utilizing the OSTI database for other collections, most notably a small collection of Nuclear Regulatory Commission (NUREG) reports. Unfortunately, the OSTI database only covers selected subjects and so large quantities of reports, e.g., NASA and the Department of Defense, still require records. APIs for these resources do not exist or are limited, however, the method of using Power Automate to query a web resource to obtain metadata remains a viable option. Looking beyond Penn State University to the approximately fifty other libraries across the U.S. that participated in the AEC depository program, records for over 207,000 titles from the AEC collection will be shared in OCLC and available to be reused. Other mechanisms to share the resulting records, e.g., extraction and loading of bibliographic data, may be possible on a case by case basis.

In summary, there are hundreds of thousands of technical report collections in academic and special libraries that are not cataloged. Although subject specialists may not have expertise in traditional cataloging, we do have knowledge of specialized tools that can help facilitate the creation of catalog records. This project is an example of how subject knowledge can be leveraged to help create access to previously hidden resources by using a combination of software tools and free online resources. Once resources become visible, usage of these collections will increase, a result for which we have ample evidence at our institution.

References

- [1] I. A. Warheit, "The Library Program of the U.S. Atomic Energy Commission Technical Information Service," *Bulletin of the Medical Library Association*, vol. 40, no. 1, pp. 1–5, Jan. 1952.
- [2] H. F. Redman, "Technical Reports - Problems and Predictions," *Arizona Libraries*, vol. 23, no. 1, pp. 11–17, Winter 1965/66.
- [3] L. Musser. "Atomic Energy Commission Depository Collection: Shining a Light on a Hidden Resource," Presented at the Special Libraries Association Annual Meeting, Baltimore, MD, 2018. Poster. DOI:10.18113/S14D18. [Online]. Available: <https://scholarsphere.psu.edu/resources/d740783e-76f2-48b1-bee6-08c3cd21916d>
- [4] J. Kirk, S. Wood, and L. Sare, "Filling in the Gaps, Doing What We Have Always Done in TRAIL," *Documents to the People*, vol. 51, no. 4, 2023, [Online]. Available: <https://journals.ala.org/index.php/dtp/article/view/8151/11353>
- [5] "National Technical Reports Library". [Online]. Available: <https://ntrl.ntis.gov/NTRL/>
- [6] "OSTI.GOV". [Online]. Available: <https://www.osti.gov/>
- [7] "International Nuclear Information System (INIS)". [Online]. Available: <https://www.iaea.org/resources/databases/inis>
- [8] *Power Automate, version 2.34.187.23206*. (2023). Microsoft.

[9] U.S. Department of Energy, Office of Scientific and Technical Information. *API Documentation*. (2024). [Online]. Available: <https://www.osti.gov/xml-services>

[10] Reese, Terry. *MarcEdit, version 7.5.98*. (2024). [Online]. Available: <https://marcedit.reeset.net/>

Figure 1 – Example of metadata on OSTI.GOV website

REPORT ON TEST OF EXTINGUISHING AGENTS FOR LITHIUM METAL FIRES

TECHNICAL REPORT

18 June 1951

OSTI ID: 4357448

Wilson, E S; Nadler, P O

❗

OSTI does not have a digital full text copy available. For more information, please see [document availability](#), search [WorldCat](#), or search [Google Scholar](#).

Cite

Export

Share

Save

Print

Details

Similar Records / Subjects

Research Organization:	Oak Ridge National Lab., Y-12 Area, Tenn.
DOE Contract Number:	W-7405-ENG-26
NSA Number:	NSA-11-003855
OSTI ID:	4357448
Report Number(s):	Y-F30-1
Resource Relation:	Other Information: Changed from OFFICIAL USE ONLY Dec. 7, 1956. Orig.

Figure 2 – Results using Power Automate to extract OSTI IDs

1	Series # Box 115	OSTI record	Title	Author 1	Author 2	Author 3	Author 4	Date	Note		
338	Y-F15-3	4394515	TRANSIENT TEMPERATURE AND THERMAL STRESS IN AN INFINITELY LONG SOLID CYLINDRICAL ROD WITH							1 Search Results	
339	Y-F17-9	4373232	PERFORMANCE CHARACTERISTICS FOR A GENERAL ELECTRIC G-3 ELECTROMAGNETIC PUMP							1 Search Results	
340	Y-F20-12		On the (N,2N) reaction in beryllium with n Houtermans, F. G.					1951		0 Search Results	
341	Y-F30-1	4357448	REPORT ON TEST OF EXTINGUISHING AGENTS FOR LITHIUM METAL FIRES							1 Search Results	
342	Y-F31-4	4402398	TESTING AND EXAMINATION OF THERMAL CONVECTION LOOPS OPERATED WITH LITHIUM AND LEAD							1 Search Results	
343	Y-F33-3	4317682	THE SOLID PHASES OF ALKALI AND URANIUM FLUORIDE SYSTEMS							1 Search Results	
344	Y-KB-1	4728195	EXCRETION OF URANIUM FROM MIXED EXPOSURES UNDER INDUSTRIAL CONDITIONS							1 Search Results	
345	Y-KB-7	4731513	USE OF STATISTICS IN AN APPLIED HEALTH PHYSICS PROGRAM							1 Search Results	
346	Y-KB-22	4756709	TESTS OF A PROPOSED URANIUM CONTAINER							1 Search Results	
347	YAEC-1	4319258								3 Search Results	

Figure 3 – Sample output using OSTI API

```
[{"rel":"citation","href":"https://www.osti.gov/biblio/6035946"}],
{"osti_id":"10108720","title":"Preliminary comparison with 40 CFR Part 191,
Subpart B for the Waste Isolation Pilot Plant. Volume 2: Probability and
consequence modeling","report_number":"SAND-91-
0893/2","product_type":"Technical
Report","language":"English","country_publication":"United
States","description":"This second volume documents the probability and
consequence modeling done by the Performance Assessment Division of
Sandia National Laboratories for the 1991 preliminary performance
assessment (PA) of the Waste Isolation Pilot Plant (WIPP). The volume
provides an overview of the PA calculations; discusses the mechanics of
the probability modeling and construction of the complementary
cumulative distribution functions (CCDFs); discusses the generic
computational models and the applied (or site-specific) models used in
consequence analysis and the results that these models predict for both
undisturbed conditions (base case) and disturbed conditions (in which one
or more hypothetical boreholes intrude the repository during the 10,000-
year regulatory period); and tabulates the calculational results used to
construct the CCDFs reported in Volume 1. 87 figs., 18
tabs.","availability":"OSTI; NTIS; INIS; GPO Dep.","publication_date":"1991-
12-01T00:00:00Z","entry_date":"2010-02-18T00:00:00Z","format":"Medium:
X; Size: 388 p.","coverage":"Topical","subjects":["12 MANAGEMENT OF
RADIOACTIVE AND NON-RADIOACTIVE WASTES FROM NUCLEAR
FACILITIES","WIPP","PERFORMANCE","MATHEMATICAL MODELS","ALPHA-
BEARING WASTES","UNDERGROUND DISPOSAL","GEOLOGY","SHAFT
EXCAVATIONS","BOUNDARY CONDITIONS","RADIONUCLIDE
MIGRATION","052002","WASTE DISPOSAL AND
STORAGE"],"doe_contract_number":"AC04-76DP00789","authors":
[],"sponsor_orgs":["USDOE, Washington, DC (United
States)"],"research_orgs":["Sandia National Labs., Albuquerque, NM (United
States)"],"other_identifiers":["ON: DE92004914"],"links":
[{"rel":"citation","href":"https://www.osti.gov/biblio/10108720"}]}]
C:\Users\jaf6\Desktop\osti\DERedos\DE110001-120000>curl
https://www.osti.gov/api/v1/records?identifier={"DE92004926"}
```

Figure 4 – Example of bibliographic record as output from MarcEdit program.

```
=LDR 00878nmm a22002657i 4500
=001 os4357448
=007 he\\m|||||
=008 250221s1951\\|||xx\\|||b\\|||u\\|||eng\\d
=040 \\$aTOE$beng$erda$cPSt
=088 \\$aY-F30-1
=100 1\\$aWilson, E S,$eauthor.
=245 10$aREPORT ON TEST OF EXTINGUISHING AGENTS FOR LITHIUM METAL FIRES.
=264 \\1$aUnited States :$b[publisher not identified].$c1951
=300 \\$a1 microfiche ;$c4 x 6 in.
=336 \\$atext$btxt$2rdacontent
=337 \\$amicroform$bh$2rdamedia
=338 \\$amicrofiche$bhe$2rdacarrier
=500 \\$aNSA number: NSA-11-003855
=500 \\$aDOE contract number: W-7405-ENG-26
=500 \\$aResearch organization: Oak Ridge National Lab., Y-12 Area, Tenn..
=653 \\0$aMINERALOGY, METALLURGY, AND CERAMICS.
=653 \\0$aCOMBUSTION.
=653 \\0$aLITHIUM.
=653 \\0$aTESTING.
=700 1\\$aNadler, P O,$eauthor.
```