

Ten Years of TRAIL

Daureen Nesdill, Laura Sare, Alice Trussell, Marilyn Von Seggern

The year 2016 marks the tenth year the members of the Technical Report Archive and Image Library (TRAIL) have been providing open access to US federal technical reports (figure 1). Because TRAIL has created a substantive open access resource over the last ten years, it seems appropriate to look back and reflect on the work of TRAIL.

History

Historically, many library patrons and librarians have viewed government documents as a collection of unknown and difficult-to-find materials. Their indexing and call numbers often vary from that of other library resources, and finding the key to unlocking those mysteries fell to a select few librarians.

The Superintendent of Documents Classification system (SuDoc), fundamentally a classification based on agency names, presents challenges to patrons because call numbers change as agencies are created, dissolved, or merged with other agencies. An assigned SuDoc call number may not be used by a library depending on the organization and arrangement of government documents within that library.

Technical reports, a subset of government documents, can be even more challenging to discover and access. These reports, the communications of government research progress in technology and science, contain important information serving specialized audiences of researchers. Even older reports, dating back to the 1920s, may be of interest to engineers, scientists, and also researchers in other fields. Most institutions have librarians who actually know the location of the reports and how to find them. Library users seeking information often need to talk to these specialized librarians, who facilitate discovery of the gems hidden on shelves and in drawers. Over the years, paper reports were replaced by microcards, microfilm, and microfiche to save paper and storage space. Unfortunately, these formats required bulky reading equipment, further impeding access. The bulky technology and variation in cataloging coupled with the transition to the use of electronic resources rendered legacy reports virtually invisible. Library users now depend on easily accessible



Figure 1. A logo was created to celebrate ten years of TRAIL.

electronic resources and have left the printed reports on library shelves.

Concerned about the inaccessibility of the wealth of science and engineering information contained in government technical reports, in 2004 librarians at the University of Arizona (UA) began conversations with other engineering librarians across the US. A vision of digitized, freely available federal technical reports created a spark of enthusiasm.

UA librarian, Maliaca Oxnam held informal sessions with members of the Engineering Libraries Division of the American Society for Engineering Education (ASEE) to gauge interest and enthusiasm. As a result of that interest, UA submitted a proposal to the Greater Western Library Alliance (GWLA) for a project that would identify, digitize, and provide open access to technical reports published prior to 1976. The project was selected by GWLA to move forward, and in concert with the Center for Research Libraries (CRL) a formal agreement was announced in the spring of 2006.

Table 1. Information from a survey of librarians resulted in a listing of the top ten federal agencies whose publications should be considered for digitization first.

1	Department of Energy, DOE
2	Environmental Protection Agency, EPA
3	National Aeronautics and Space Administration, NASA
4	US Department of Agriculture, USDA
5	National Bureau of Standards, NBS
6	US Geological Survey, USGS
7	Atomic Energy Commission, AEC
8	National Advisory Committee for Aeronautics
9	US Army Research
10	US Bureau of Mines

Soon after the announcement a fortuitous opportunity emerged. The Linda Hall Library, in Kansas City, MO (LHL), was interested in digitization and agreed to work on the project and conduct a cost analysis by digitizing 200-500 reports. A task force of representatives from six academic libraries, in addition to representatives from LHL, GWLA, and CRL met at the Center in August 2006. During that meeting, an ambitious plan was developed to gather, digitize, and establish a prototype website to host digitized technical reports. The goal was to have a website up and running within eight months using a budget of \$65,000. The National Bureau of Standards (NBS) Monograph Series was chosen as the first series to be digitized due to its unique depth and breadth of information. A small subset of Atomic Energy Commission (AEC) reports from the Division of Biology & Medicine were also chosen for digitization.

Since the six institutions involved in the project spanned four time zones, GWLA hosted a web-based workspace to enable the task force to communicate and store project documentation. Weekly conference calls moved the project forward. A 'call for interest' survey was sent to science, technology, engineering, math (STEM) and government documents librarians through numerous listservs, asking respondents which agencies' content should be the project's highest priority. The enthusiastic responses to the November 2006 survey were gathered from eighty-four individuals representing sixty-one university libraries and sixteen government agencies and special libraries. From the results the top ten agencies whose collections were priorities for digitization were identified (table 1).

It was Robert Schwarzwald, while a librarian at the University of Hawaii at Manoa (UH Manoa), who volunteered his institution's resources to create and host the file storage and website page for the pilot project using Streetprint. The digitized reports from LHL and commercial scanning vendors were sent

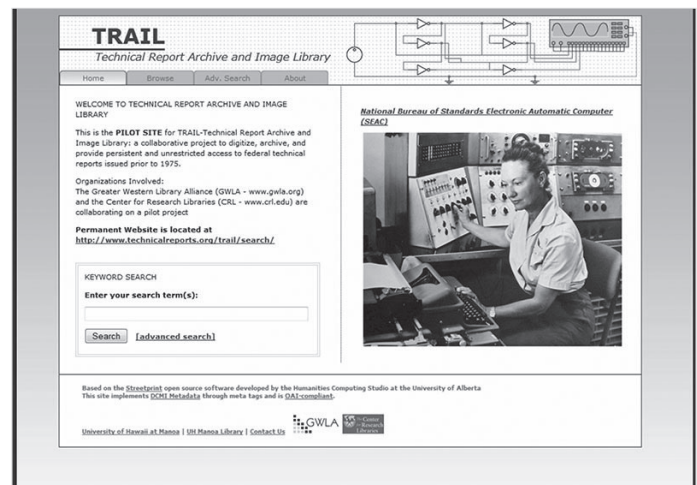


Figure 2. The pilot TRAIL website from 2009 was created at the University of Hawaii-Manoa. The site no longer exists.

to UH Manoa to be loaded onto servers and made accessible through the pilot website. Before the website could go live, a name for the project had to be determined. Various names and acronyms were proposed and TRAIL, an acronym for Technical Report Archive and Image Library, was selected. The website for accessing reports with the new name and logo went live in March, 2007 (figure 2). Shortly afterward the University of Michigan (UM) contacted GWLA and offered to host the TRAIL website. The discussion surrounding that offer fell through, but it led to an offer to join the UM Digitization Project and have the technical reports digitized through the Google Books Program. Because the UM Digitization Project is able to scan materials at no cost to TRAIL, the overall cost of scanning materials was considerably reduced.

TRAIL started attracting attention. An invitation was received to present to the Commerce, Energy, NASA, and Defense Information Managers Group (CENDI) at the May 2007 meeting in Washington, DC. The CENDI meeting was an important step in facilitating communication with government agencies.

The original members of TRAIL were engineering librarians plus one government documents librarian. Five additional government document librarians were recruited in 2008 to increase the project's expertise in government documents. Now the project had an even split between engineering and government document librarians. Working groups were formed to address the various types of tasks, such as collecting reports for processing, processing the documents to be digitized, developing and maintaining a website about the project, and addressing the technology issues with scanning.

The Workflow

The basic workflow has not changed substantially since the beginning of the project. Two workflow streams for digitization were developed. While all content was sent from partner institutions to UA for processing (reviewed, inventoried, cataloged and shipped), in the first stream, material of uniform page size was sent to the UM Digitization Project for scanning and providing online access. The preference was for donated items that could be deconstructed for the scanning process and not returned to the owning library.

The second workflow stream consists of the reports requiring special handling. Publications with maps, foldouts or anything warranting nondestructive scanning were initially digitized by vendors and stored on servers at UH Manoa. In 2008 the UM Digitization Project began depositing its digitized collection in the HathiTrust Digital Library, often referred to as just HathiTrust. As a result, all TRAIL documents scanned through the UM Project are now deposited in HathiTrust.

By 2009 the project was outgrowing the capacity of the pilot website. Mel DeSart, librarian at the University of Washington (UW) offered the expertise of the library's IT department in developing a search engine and interface (figure 3). At the same time the University of North Texas (UNT) agreed to provide servers for the non-Google scans. The new interface being developed by UW would therefore have a search engine able to search both HathiTrust and UNT content for TRAIL reports. UNT also volunteered to scan the reports containing maps, foldouts or anything warranting nondestructive scanning. Thus UNT took over both third party digitization vendors and UH Manoa's positions in the workflow.

Two major events occurred in 2010. TRAIL won the LexisNexis/GODORT/ALA "Documents to the People" award (figure 4). This increased the project's visibility in the government documents community. The second event was the administrative move of TRAIL from GWLA to CRL and becoming the Center's newest Global Resources Network. In addition to administrative expertise, CRL provides online workspace and a web presence, www.crl.edu/programs/trail. As a result of the move, bylaws were adopted and TRAIL became more formally organized. The elected positions of chair, chair-elect and secretary were established along with a steering committee as the governing body. The four working groups, Collections, Processing, Technology and Communications were formalized. The elected position of treasurer was added in 2012.

While TRAIL was a GWLA initiative, the institutions belonging to GWLA provided financial support for the project. Once TRAIL moved administratively to CRL, a new support structure had to be developed. In 2011 a new member recruitment task force was established, which later morphed into a

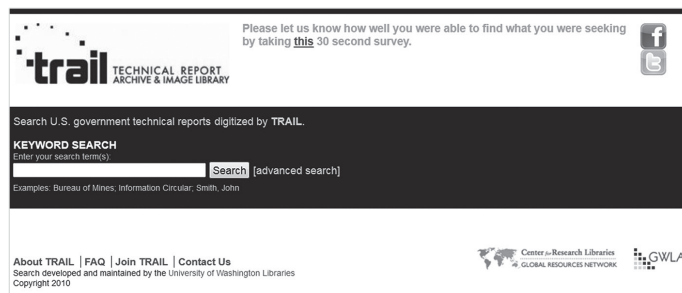


Figure 3. The present TRAIL website for searching technical reports was created at the University of Washington. www.technicalreports.org/trail/search/

working group. Membership is important to TRAIL not only for financial support, but because members demonstrate support for the project through content and volunteer assistance. As of December 2015 TRAIL had thirty-nine member institutions.

Four core working groups, <http://www.crl.edu/grn/trail/working-groups>, move TRAIL forward by focusing on specific aspects of the overall process. Each working group and the Steering Committee have separate scheduled time for conference calls and workspace on CRL's Confluence. The Collections Working Group is responsible for all aspects of the TRAIL project related to the identification, selection, and acquisition of report series for scanning. All aspects related to the cataloging, scanning, and deposit of the resulting electronic TRAIL materials into the appropriate archive are done by the Processing Working Group. The Communications Working Group is responsible for all communication and promotion regarding TRAIL, such as providing descriptive project content on the TRAIL web pages, assisting other working groups with communication development, and oversight and general help with reference questions sent to TRAIL. The responsibility assigned to the Membership Working Group is for recruiting additional organizations and individuals to become members of TRAIL, for creating and conducting orientation sessions for new TRAIL members, and for all member-specific communication within TRAIL.

Changes in Philosophy Years of Coverage

Initially, selection of the reports to be digitized was based on the results of the initial survey, print format, and publication date prior to 1976. The date limiter became problematic early on, since some series that began decades earlier continued years past 1975. It was decided to complete these series rather than adhere to the 1975 cutoff date. In 2014 the reference to "pre-1976" in the Goals and Bylaws was eliminated.

Non-print Formats Member institutions were requesting microforms be digitized since this format was difficult for library patrons to use and microform collections took up valuable space in libraries. In 2011 TRAIL initiated a pilot project to digitize

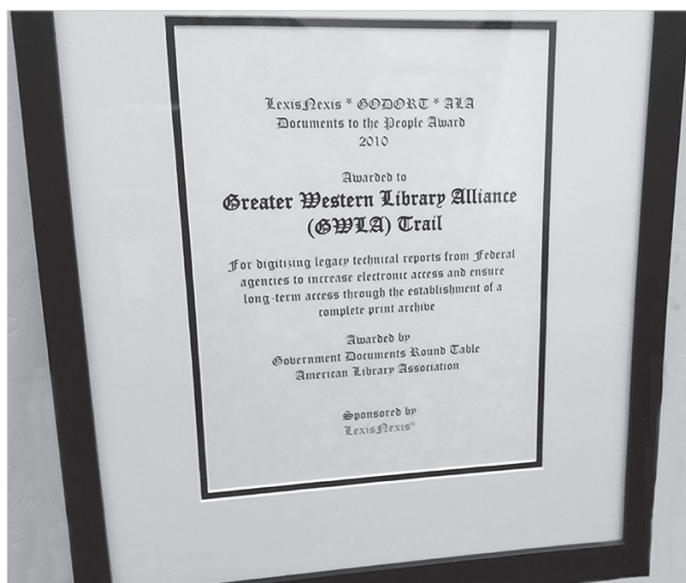


Figure 4. TRAIL won the LexisNexis/GODORT/ALA “Documents to the People” award in 2010.

technical reports on microform in response to these requests. Workflow and standards were investigated and a decision to have UNT be responsible for microforms was agreed upon.

Harvesting

The National Aeronautics and Space Administration (NASA) officials contacted TRAIL and requested assistance in locating some of their missing National Advisory Committee for Aeronautics (NACA) reports on microfiche. Library personnel at University of California, San Diego, (UCSD) a TRAIL member, were just completing an assessment of UCSD NACA microfiche. They contacted TRAIL about reports they owned that were not listed on NASA servers. The two groups were put in contact and worked together to complete the collection and as a result TRAIL was able to harvest the entire digitized NACA collection. A decision was then made by the Steering Committee to work with other federal agencies in harvesting useful collections.

Personal Members

In 2014, TRAIL was advised that some of the CRL Global Resources Network groups have provisions in their bylaws to include “personal members.” These are personnel at institutions or organizations that are not members of TRAIL who are interested in participating in TRAIL’s Working Groups and activities. Discussions about how personal members could assist TRAIL with its goals for the future ensued and in early 2015 the TRAIL bylaws were amended to include personal membership. By December 2015, TRAIL had nine personal members from two federal agencies, six universities and one public library.

Series Selection

Some of the more well-known and readily available sci-tech government series were among the first to be digitized. These included technical report series from the NBS, the Bureau of Mines, and the Fish and Wildlife Service. Libraries engaged in weeding projects or downsizing for moves or remodeling donated many items. Other donated volumes were second copies and no longer needed.

When most of a series is digitized the missing pieces are identified on the TRAIL Needs List, www.crl.edu/grn/trail/current-activities/needs-list. At that point TRAIL accepts the pieces in any way they can be sent—as returnable loans or as microfiche, for example.

As widely held series neared completion the TRAIL Collections Working Group moved on to other agencies and series. Examples include United States Earthquakes 1928–62 and 1963–68, produced by the Coast and Geodetic Survey (now the National Geodetic Survey) of the Department of Commerce, the Federal Energy Administration (later merged with the Department of Energy) Conservation Papers and other series, and the Biological Services Program (Department of the Interior) report series.

Another major set of government technical report series were those from the AEC, established in 1945 and abolished in 1974. Research done in the 1940’s leading up to the development of the atomic bomb and continuing after the war with peacetime uses of atomic energy resulted in numerous laboratory technical report series. Some were generated directly by government labs and others were produced under contract with research and commercial entities such as Armour Research Foundation, General Electric, Babcock and Wilcox Company, Battelle Memorial Institute, DuPont de Nemours and Company, and General Dynamics Corporation.

Copyright becomes an issue when private entities are involved, even when the reports are issued in government series from the AEC. Some technical report series from government laboratories such as Argonne, Brookhaven, Livermore, Sandia, and Hanford were classified, but many had no restrictions and were distributed to academic libraries. Even items in these series could run into copyright restrictions, however, if a photograph or any other part of a report had private ownership or authorship. If TRAIL is aware of a copyright restriction, the report will not be freely available. As the number of TRAIL-generated technical reports grew in HathiTrust it was noticed that many were inaccessible due to copyright restrictions. The problem was investigated and found to be due to the absence of a cataloging record field indicator for government documents. Though

some are still not available due to valid copyright restrictions, the addition of this field code has opened many to full view.

A need was recognized for brief descriptions of series to provide some background on the sponsoring agency to users of TRAIL. Summaries including the history of the agency and any other pertinent information are included for most series on the Collections Working Group site at www.crl.edu/grn/trail/current-activities/series-list.

Progress by the Numbers

TRAIL has now digitized and made available more than 50,000 US agency technical reports, all accessible from the TRAIL interface at www.technicalreports.org (fig. 3). Thirty-seven percent of the collection resides in the UNT Digital Library and the rest in HathiTrust. Detailed holdings and use statistics are not available from HathiTrust at this time, but UNT Digital Library statistics provides an estimate of TRAIL's digitizing effort and the use it is receiving. In November 2015, the TRAIL holdings at UNT resulted in a page count of over one million, a document count of over 18,500, and usage totals of over two million. Based on UNT Digital Library statistical reports, the TRAIL content receives the highest use of any collection in the Library. The number of uses has grown from 4,705 in 2010 to 554,743 in 2013 and 623,383 in December 2015. TRAIL's objective to provide the public with full-text access to federal technical reports has been making steady, substantial progress during the past ten years.

Future Plans

As TRAIL moves into its tenth year, the future looks bright. TRAIL, no longer a start-up organization, is working to improve upon the foundation developed over the last ten years. TRAIL will expand on the institutional cooperation and partnerships that provided a solid foundation. A future goal of TRAIL is using technology to improve discovery and usability of the information in the collection. An area of focus will be the website interface. A technology task force has been established to address issues such as the metadata required to increase discovery of material. Discussion among the members of TRAIL have occurred about how to apply text mining to increase the usefulness of the reports, which have broad content both in subject and type of information. The reports may include pages of data, charts, standards, and series of reports on a single research topic (e.g., Strontium 90).

TRAIL will be creating additional outreach opportunities through Internet-based services. An announcement will be made in 2016 when the TRAIL-related Libguides go live. TRAIL also does outreach through Social Media.



Figure 5. The locations of institutional and personal members of TRAIL as of December 2015 including both the US and Canada. The red pins are institutional members and the green pins are personal members.

TRAIL's Facebook page is Technical-Report-Archive-Image-Library-TRAIL, the Twitter account is @TRAILTechReport and the Wikipedia page can be found at en.wikipedia.org/wiki/Technical_Report_Archive_%26_Image_Library. The Communications Working Group is responsible for maintaining these services.

Members of TRAIL are also reviewing processes and procedures to determine what present procedures are working, and what can be improved. TRAIL is documenting activities to provide more transparency to current members. This will also provide potential members with information about TRAIL. Another initiative is to establish metrics so that statistics can be gathered from both HathiTrust and UNT regularly to demonstrate the importance of TRAIL.

At the ten-year mark TRAIL has thirty-nine member institutions and nine personal members (figure 5). The most recent addition to member institutions is the Government Publishing Office. They will be bringing their expertise in cataloging technical reports and join the group in the conversation about the future of TRAIL. If you are interested in joining the conversation, TRAIL invites you to learn more about TRAIL and become a member at www.crl.edu/grn/trail/about-trail.

Daureen Nesdill (daureen.nesdill@utah.edu) is Research Data Management Librarian, University of Utah. **Laura Sare** (lsare@tamu.edu) is Government Information Librarian, Texas A&M University. **Alice Trussell** (alitrus@ksu.edu) is Engineering Librarian, Kansas State University. **Marilyn Von Seggern** (m_vonseggern@wsu.edu) is Government Information Librarian, Washington State University Libraries.