LA-UR-24-21362

Approved for public release; distribution is unlimited.

Title:	Data Management Across DOE Labs: LANL Research Library Overview
Author(s):	Cain, Brian J. Hamalainen, Hannah Winkler Klein, Martin
Intended for:	Presentation for "Data Management Across DOE Labs" event
Issued:	2024-02-14

Los Alamos
NATIONAL LABORATORY

Los Alamos National Laboratory, an affirmative action/equal opportunity employer, is operated by Triad National Security, LLC for the National Nuclear Security Administration of U.S. Department of Energy under contract 89233218CNA000001. By approving this article, the publisher recognizes that the U.S. Government retains nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or to allow others to do so, for U.S. Government purposes. Los Alamos National Laboratory requests that the publisher identify this article as work performed under the auspices of the U.S. Department of Energy. Los Alamos National Laboratory strongly supports academic freedom and a researcher's right to publish; as an institution, however, the Laboratory does not endorse the viewpoint of a publication or guarantee its technical correctness.

Data Management Across DOE Labs:
LANL Research Library Overview

Brian Cain, Librarian
cain_b@lanl.gov
https://orcid.org/0000-0001-7356-5860

Hannah Hamalainen, Data Management Librarian
hannahh@lanl.gov
https://orcid.org/0000-0002-7783-7736

Martin Klein, Scientist
mklein@lanl.gov
https://orcid.org/0000-0003-0130-2097

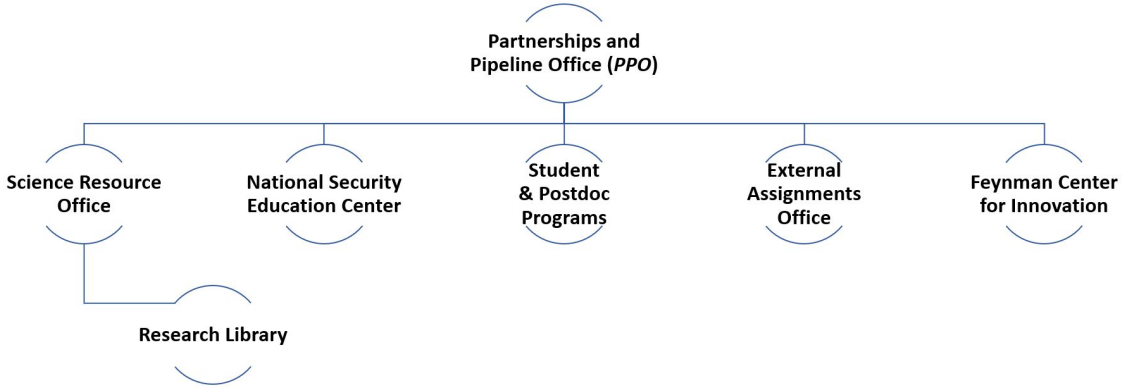
Love Data Week
Feb 15 2024

LANL Data Services Landscape

- Research Library
  - Research Services & Resource Discovery
  - Technology Strategy and Services

- Information Science and Technology Institute (ISTI)

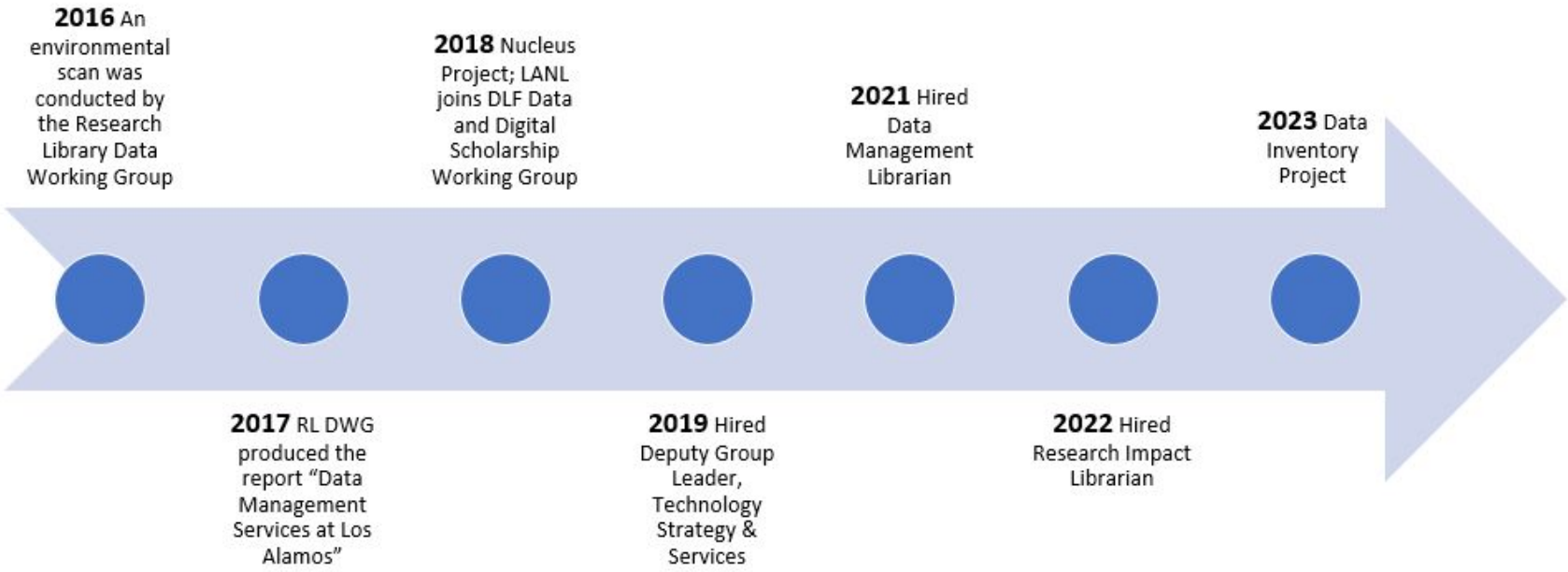
- High Performance Computing



Previous Data Stakeholder Outreach

- Needs assessment through surveys and user interviews
- Respondent areas of greatest need in 2016
  - Data storage and archiving
    - Two thirds of respondents were primarily storing research data on their personal computers, unsure how to preserve for the long term
  - Data Review and Release/Sharing
    - Clarifying and streamlining review and release process for data
    - Best methods for sharing internally and externally
  - Data Management Plans
    - How to create a DMP and what language to include as it relates to LANL/funder policy

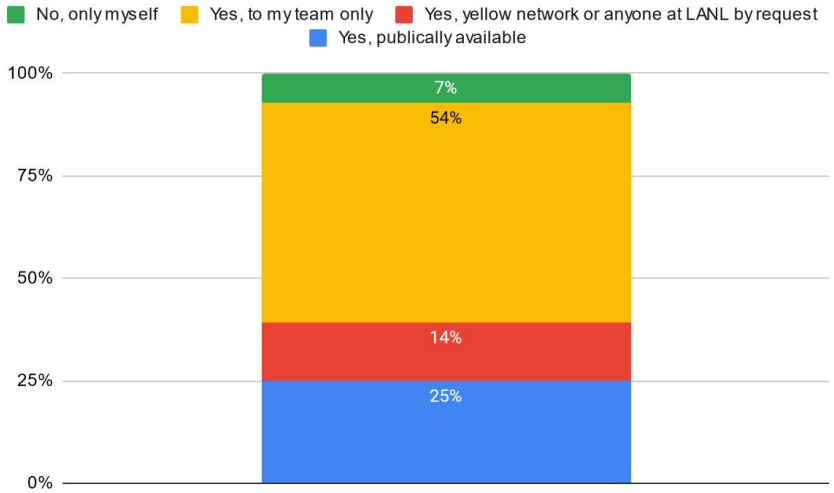
Moving Forward and Outreach



Data Inventory

- The majority of respondents shared data within their group or team only – 75% of the data is internal to LANL

- Only three people of twenty-seven responded that they had a data management plan in place.



- Future wants/needs:
  - Repository feature needs: a data “phonebook” or directory of data owners with interoperability and a data preview search functionality
  - Data policy guidance for offboarding data
  - Further outreach

Available resources and services for LANL researchers

RESEARCH LIBRARY / SERVICES / DATAMANAGEMENT

About Us | Research Tools | Find Materials | Services | RASSTI & ROSY | Library FAQs | LANL Databases

DATA MANAGEMENT

- Library Services
- Circulation
- Data Management
- Document Delivery
- Interlibrary Loan (ILL)
- Remote Access
- Research Assistance
- Research Impact
- Retiree Borrowing Program
- Scanning
- Training
- Data Dose

Data Management Consulting Services

request

We offer help with:

- Persistent Identifiers (ORCID, DOIs, etc.)
- Data Management Plans (DMPTool)
- Reference Management (EndNote, Zotero, etc.)
- Finding and Using Data Repositories
- Organizing Data and Data Documentation (Good practices for data documentation, file organization, file naming, metadata)
- Scholarly Communication (Publons, open access, etc.)
- RASSTI (Review and Approval System for Scientific and Technical Information) & ROSY (Review Only SYstem)
- Data Vizualization (Data Cleaning, GIS, Excel, etc.)
- Researcher Impact (Bibliometrics, altmetrics, etc.)
- Reproducible Research (Electronic Lab Notebooks, etc.)
- Data After Your Project (Archival storage, sharing, and preservation)

Services

Discovering and Accessing Data

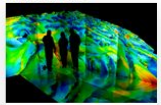
The SRO Research Library offers access to a wealth of data collections from LANL and other government sources, proprietary databases, not-for-profit agencies, and other institutions. Research Data Management Services can help you navigate these and other publicly available sources.

Data Management Plans (DMP)

DMPTool
Build your Data Management Plan

Data management plans should address how data will be managed, disseminated and preserved over the course of a research project. We recommend using the DMPTool, a free, web-based editor and template, to help prospective PIs efficiently write their data plans. Email dataquestions@lanl.gov for help.

Data Visualization



Data visualization support to transform data into visual representations can happen at any stage of your research. We can suggest ways to clean your data for visualization, brainstorm on visualization techniques, or to learn about helpful data visualization tools.

Documentation & Organization

Organizing your data involves developing a directory structure for your files, standard file naming conventions, and documention of the data planning. Creating a plan for how to organize your data from the beginning of a research project can streamline access to and management of your data.

Data Citation and Metadata

Metadata, or "data about data" is documentation that helps others discover and cite your dataset. Our services help you determine what documentation and metadata to create, how to create it, how to store it, and how to use citable persistent identifiers such as a DOI or ORCID.

Storage and Backup

Our consulting services help you prepare your project data for longer-term access by you and those in your research community, and to come up with a storage plan for your data that addresses your and your team's needs for access and data security.

LANL Citation Classification

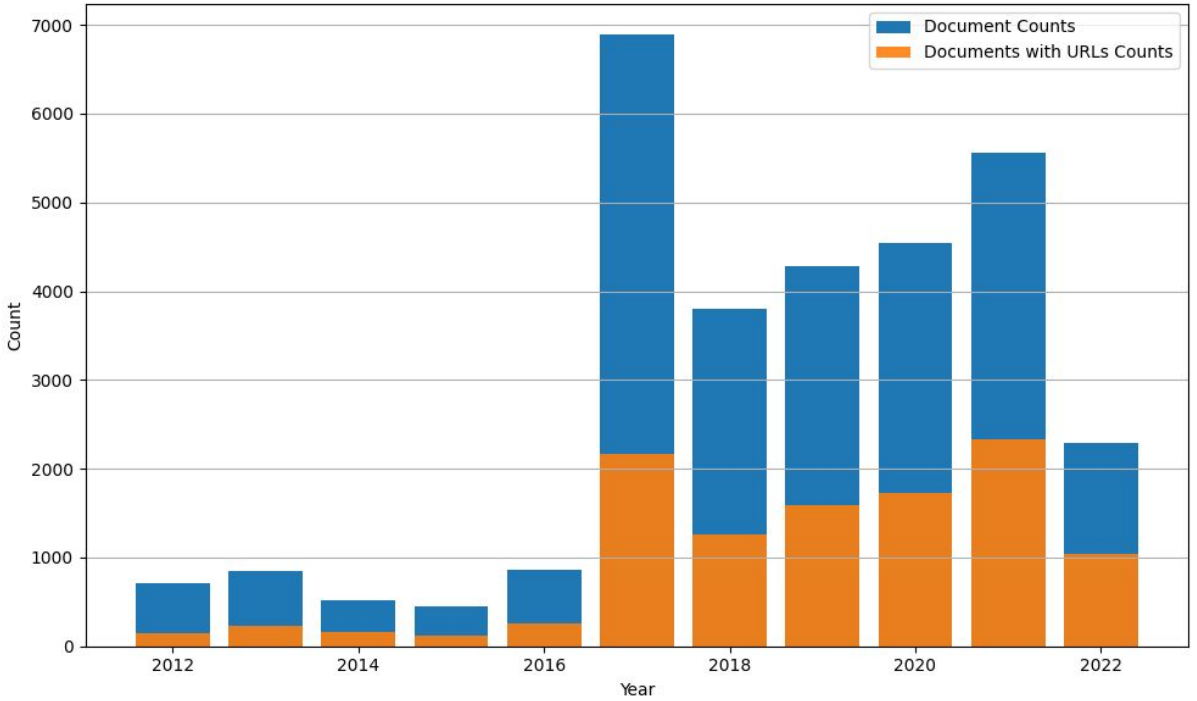
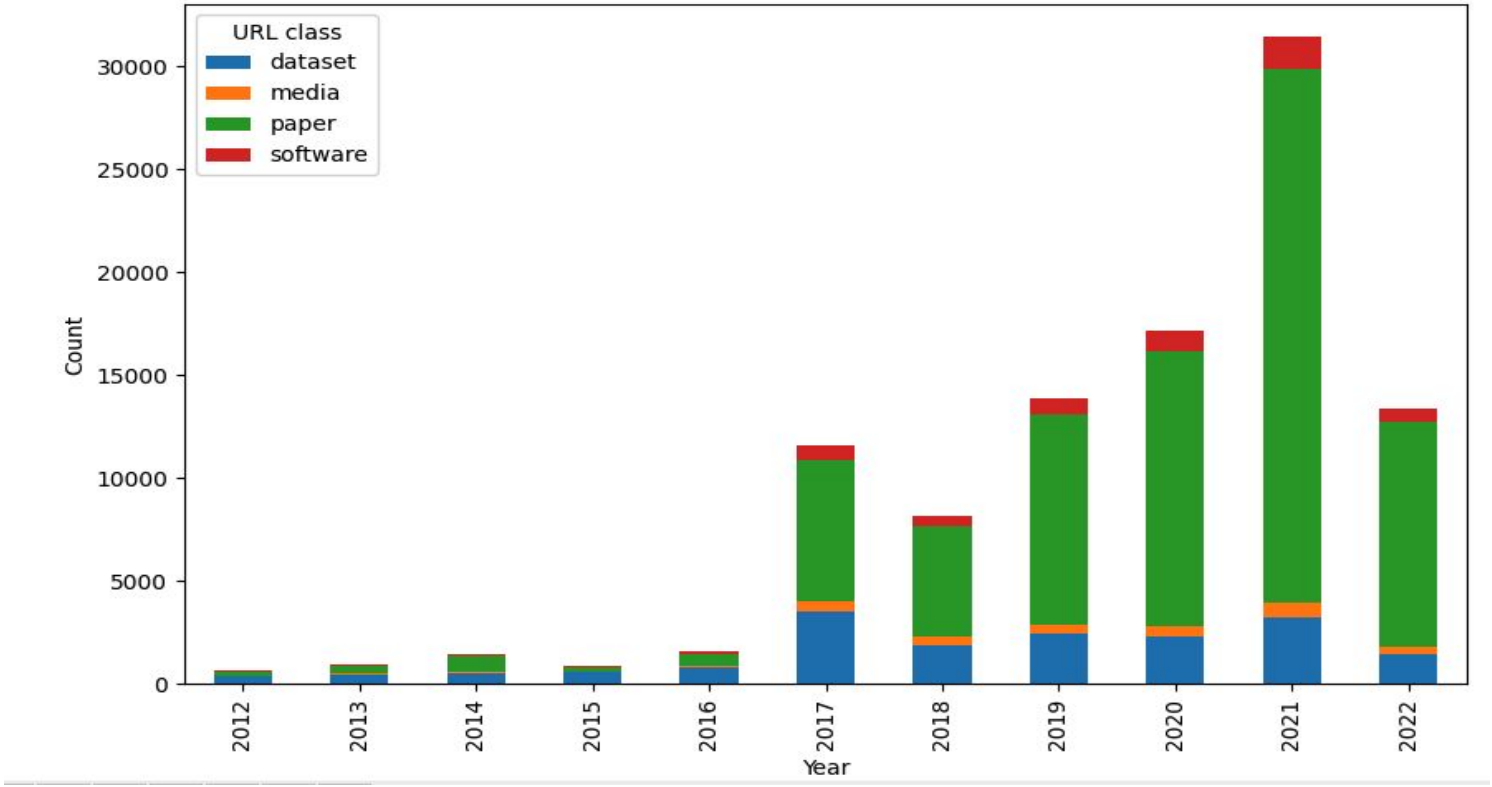
We are creating and (re-)using software, datasets, and other scholarly artifacts!

- Do we know of them?
- Do we know where they are?
- Do we know who reuses them?

Experiment:

- RASSTI corpus of LANL-authored publications
- PDF -> XML, extraction of URLs and their citation context
- Train ML models to predict citation class: code, data, media, paper
- Train ML models to assess citation intent (created, used, mentioned)

LANL RASSTI Approved Submissions



LANL Citation Classification