Unlocking Biomedical Research Results: Resources from the National Library of Medicine

Peter Seibert, MLIS
Librarian

Brittany Chao, DPhil
Senior Health Science Policy Analyst

Teresa Zayas Cabán, PhD, FACMI
Assistant Director

American Medical Informatics Association
January 29, 2026

Topics for Today

- Introduction
- Dataset Catalog 2.0
- NIH Sharing Policies: NIH Public Access Policy
- Discussion

Topics for Today

- Introduction
- Dataset Catalog 2.0
- NIH Sharing Policies: NIH Public Access Policy
- Discussion

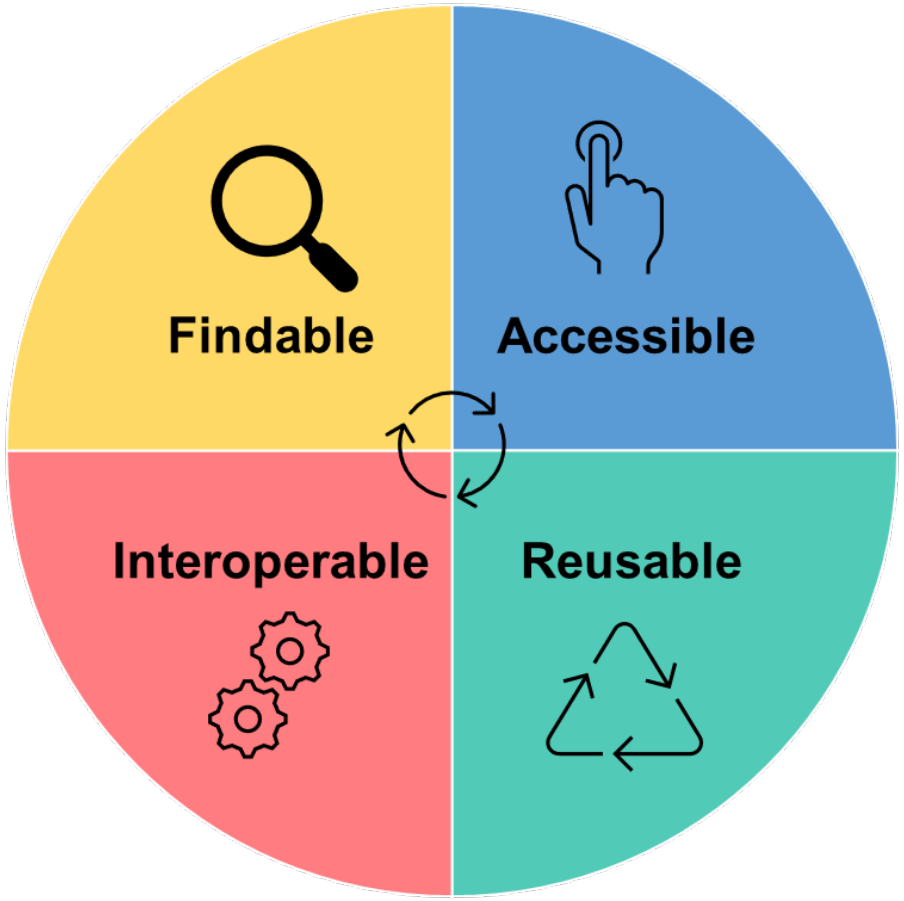
Importance of Sharing Research Results

- Advance rigorous and reproducible research

- Promote public trust in research

Enabling Findability of Research Datasets

Access to data key to scientific discovery



Adapted from National Institute of Environmental Health Sciences

Expediting Access to Results of NIH-Funded Research

- Efforts to ensure:
  - Scientific data underlying publications are made accessible at the time of publication
  - Publications are made freely available and publicly accessible in repositories without embargo
  - Digital persistent identifiers are included in published research outputs

Critical Role of Informatics

Topics for Today

- Introduction
- Dataset Catalog 2.0
- NIH Sharing Policies: NIH Public Access Policy
- Discussion

NLM Dataset Catalog: Expanding Biomedical Data Discovery at Scale

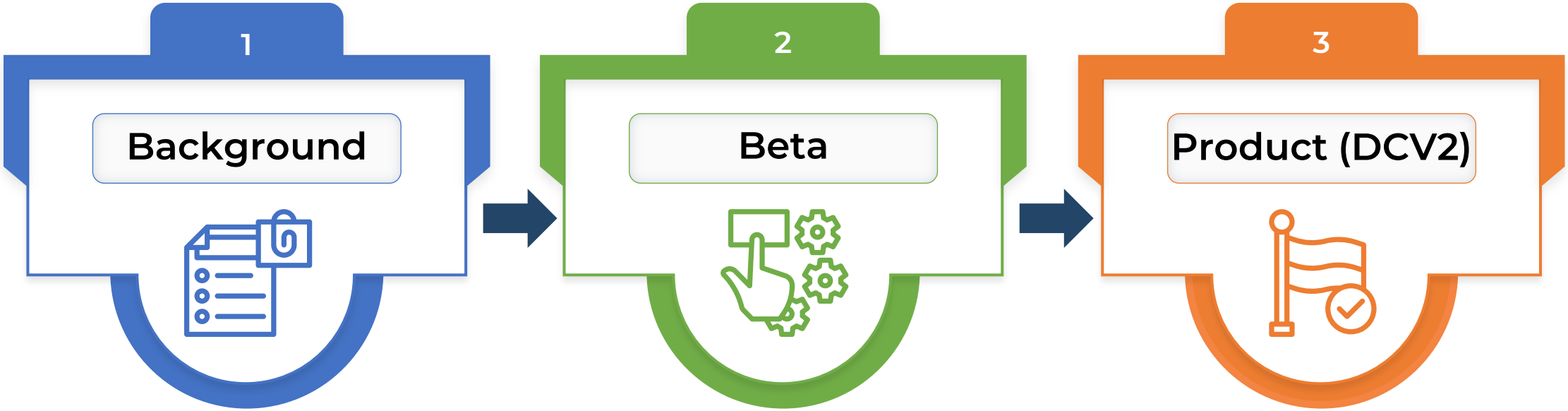
National Library of Medicine

January 2026

Peter Seibert

NLM Dataset Catalog

Content Agenda



NLM Dataset Catalog

PubMed of Datasets!

WHAT

- A catalog of biomedical datasets from selected publicly available repositories for researchers

WHY

- Accelerate biomedical research and discovery
- Leverage AI to scale dataset discoverability
- Facilitate open access to NIH-funded research

VISION

- A single, “all-in-one” tool helps users to search, find, & retrieve biomedical datasets

STRATEGIC GOALS

- Provide discovery systems for users
- Allow researchers to discover biomedical datasets from many repositories
- Drive adoption and acceptance of DATaset Metadata Model (DATMM) as a standard

DATMM Overview

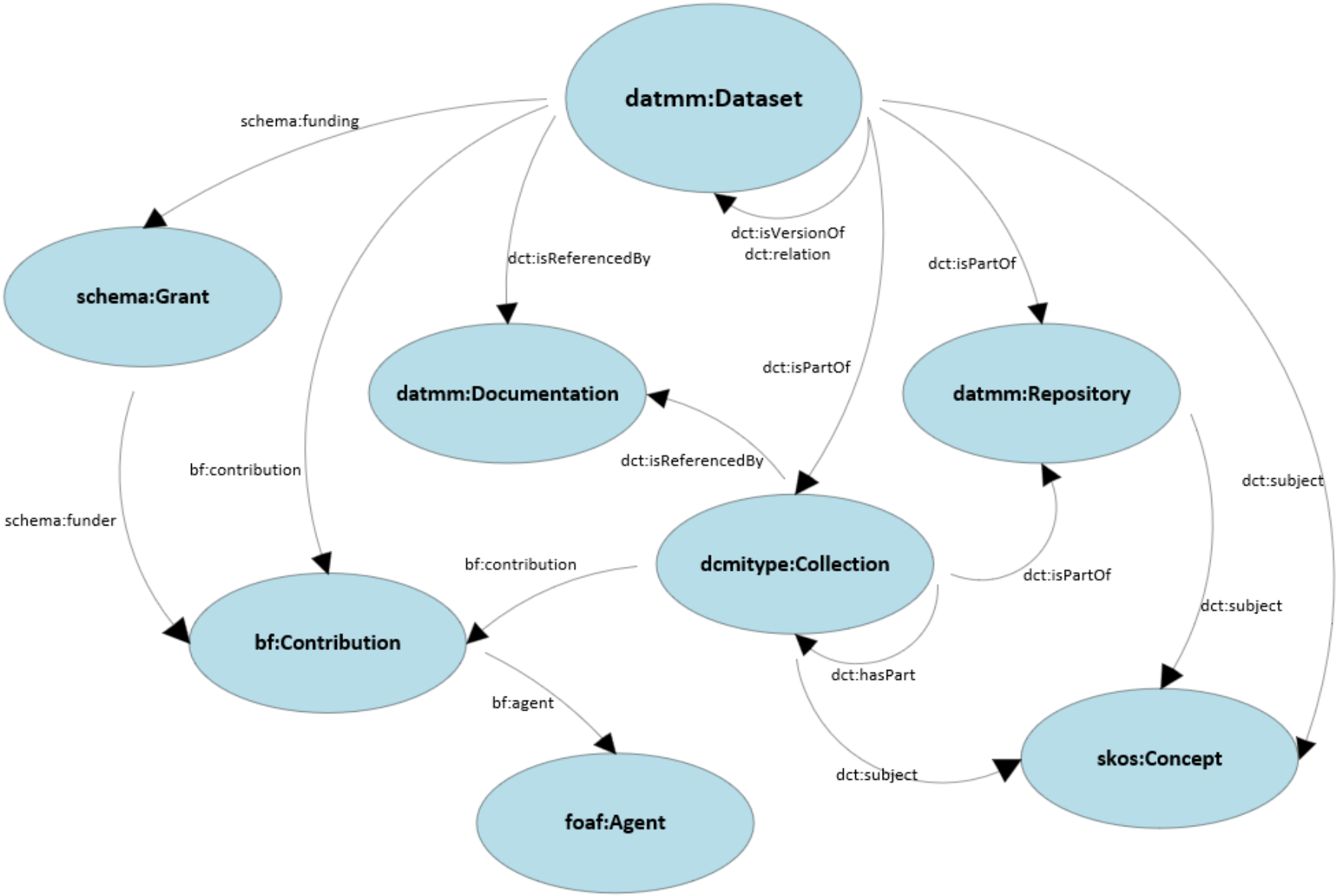
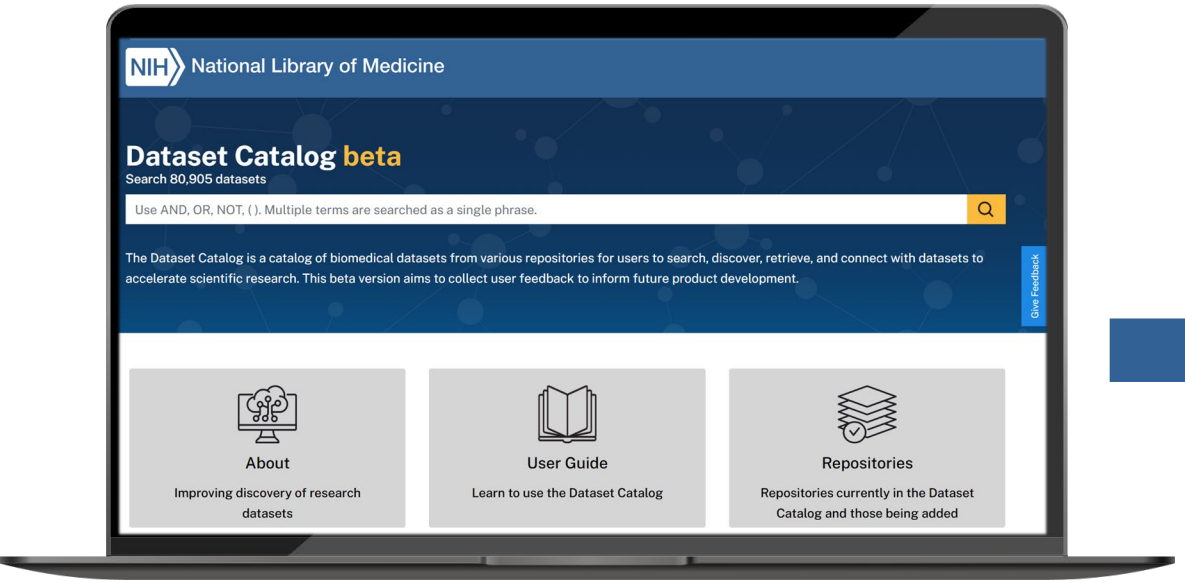


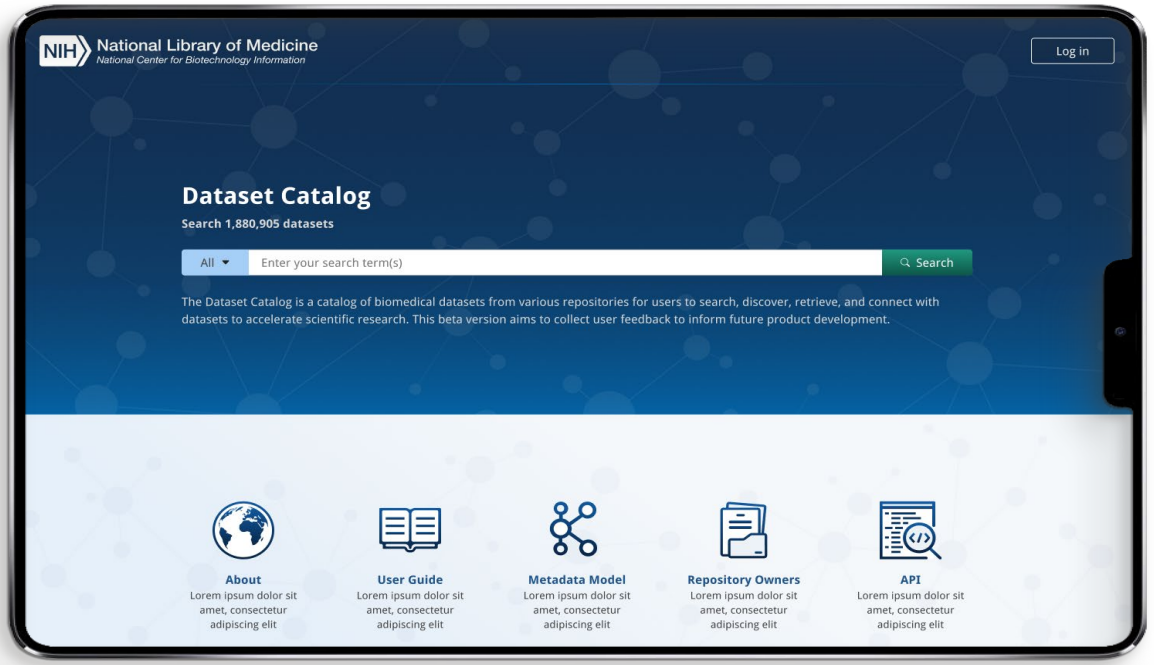
Figure 1. DATMM top-level Classes and their relationships to each other

From Beta to Product (DCv2)



Beta Version (2023)

Current Interface (2026)



Beta Ingested Repositories

7
REPOSITORIES
80,000+
DATASETS
and counting!

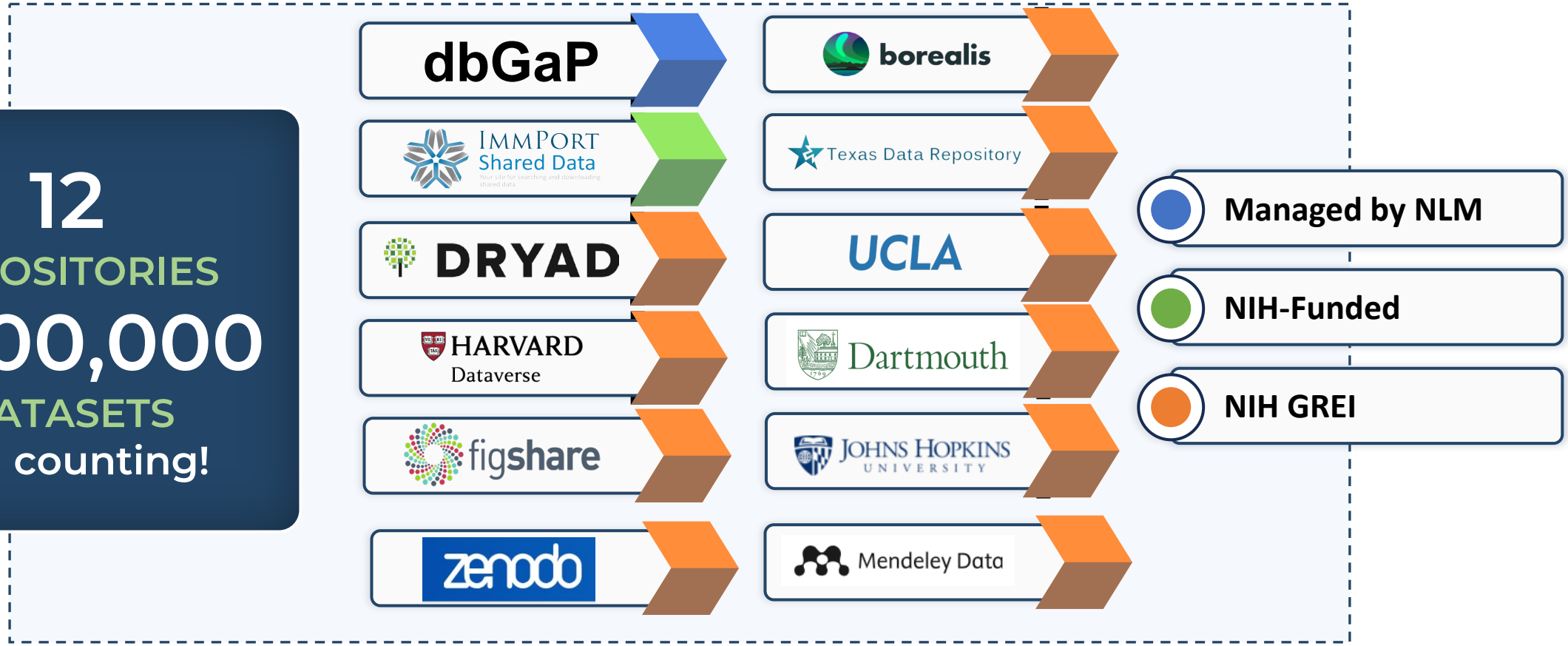
- Managed by NLM
- NIH-Funded
- NIH GREI

Feedback Implemented

1	Search not precise with too broad of results	Targeted Searching of Key Facets ☑
2	No personalization	Login to save searches ☑
3	Limited Number of Datasets	More than 2M Datasets ☑
4	Complex Tool Needs Better Instruction	Improved User Guide ☑
5	Limited Number of Repositories	Repository Application Portal ☑
6	No programmatic access to NLM metadata	Introduction of API ☑
7	Limited MeSH indexes	Automatic Indexing with MeSH ☑
8	No Way to Find Content on Web	Published Content to Semantic Web ☑
9	Cost Prohibits Maintenance & Scaling	Ability to Scale ☑
10	No way to cite sources	Ability to Create Data Citations ☑

DCv2 Ingested Repositories

12
REPOSITORIES
2,300,000
DATASETS
and counting!



NIH National Library of Medicine

Log in

A New Look!

NLM Dataset Catalog

Search 2,322,888 datasets

All ▾ | Enter your search term(s) | Search

The NLM Dataset Catalog is a catalog of biomedical datasets from various repositories for users to search, discover, retrieve, and connect with datasets to accelerate scientific research.

Give Feedback

About
Review general information and policies

User Guide
Learn tips and tricks for searching and using

Metadata Model
Explore the DATaset Metadata Model (DATMM) Schema

Repository Owners
Evaluate requirements and apply for inclusion

SPARQL Query Builder
Access the SPARQL query editor

Searchable Repositories
Find repositories currently available for searching

NIH National Library of Medicine

NIH National Library of Medicine

Log in

New Search Options

NLM Dataset Catalog

Search 2,322,888 datasets

All ▼ | Enter your search term(s) | Search

- All
- Title
- Funding ID
- Contributor

t Catalog is a catalog of biomedical datasets from various repositories for users to search, discover, retrieve, and connect accelerate scientific research.

Give Feedback

About
Review general information and policies

User Guide
Learn tips and tricks for searching and using

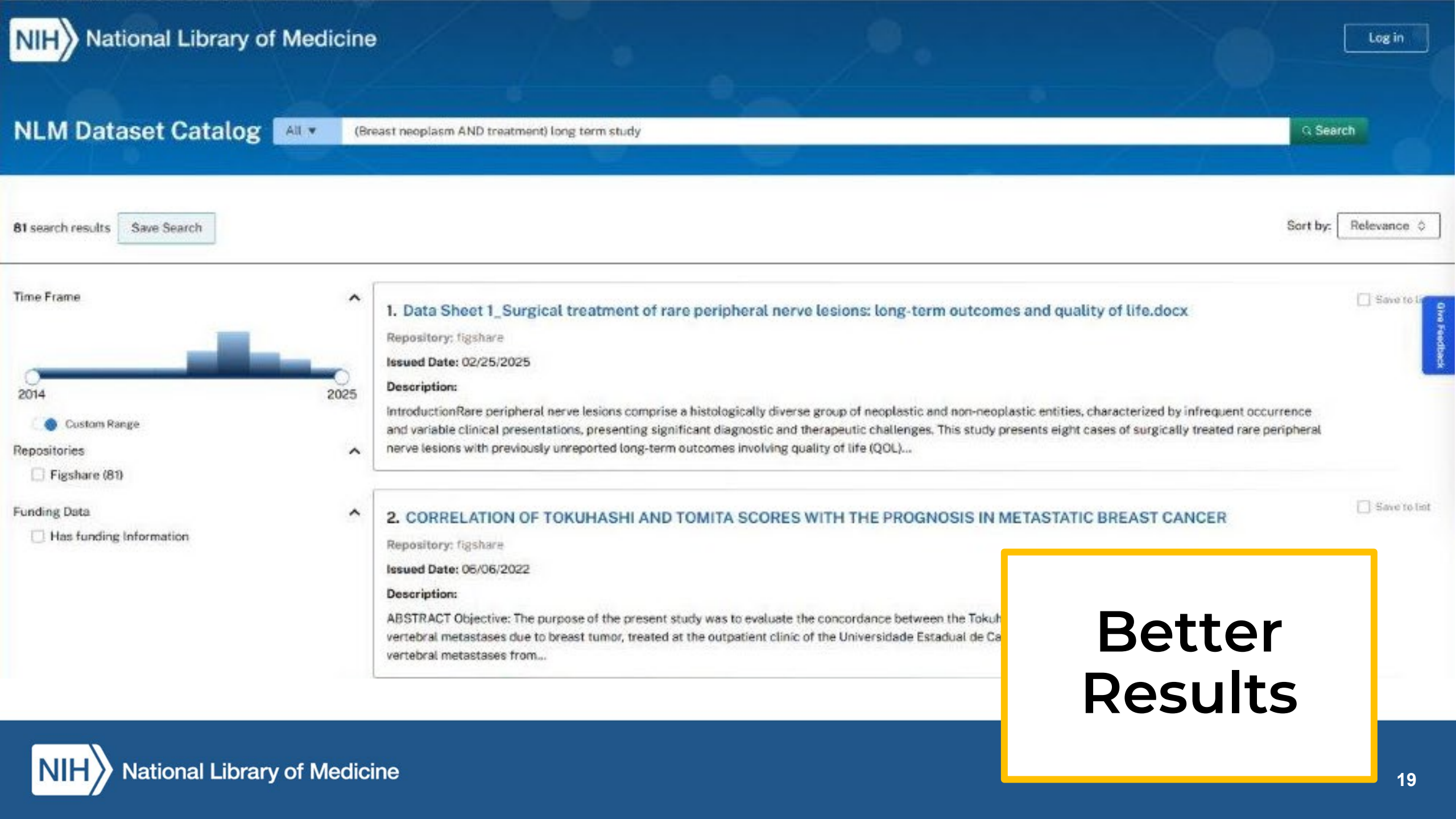
Metadata Model
Explore the DATaset Metadata Model (DATMM) Schema

Repository Owners
Evaluate requirements and apply for inclusion

SPARQL Query Builder
Access the SPARQL query editor

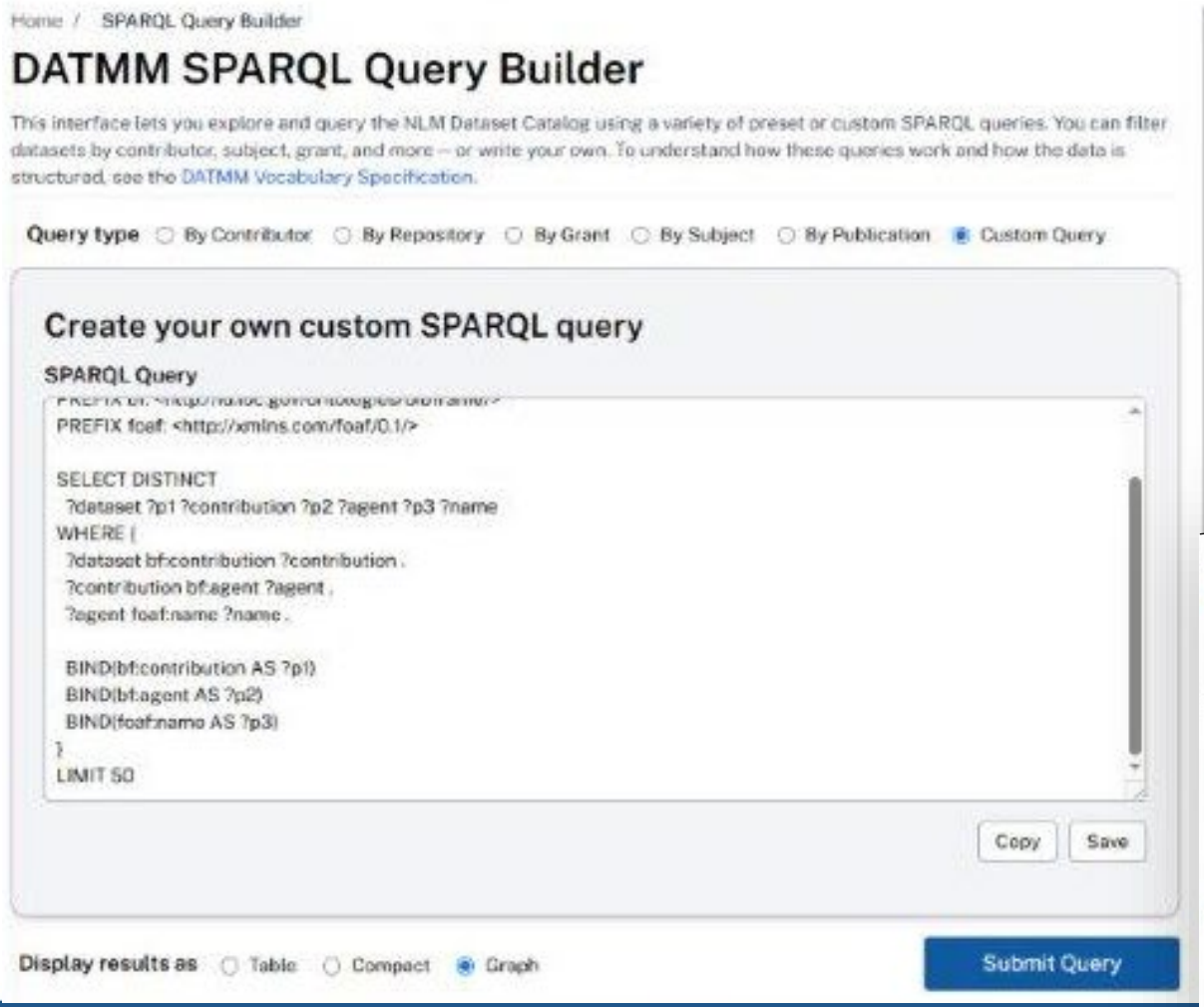
Searchable Repositories
Find repositories currently available for searching

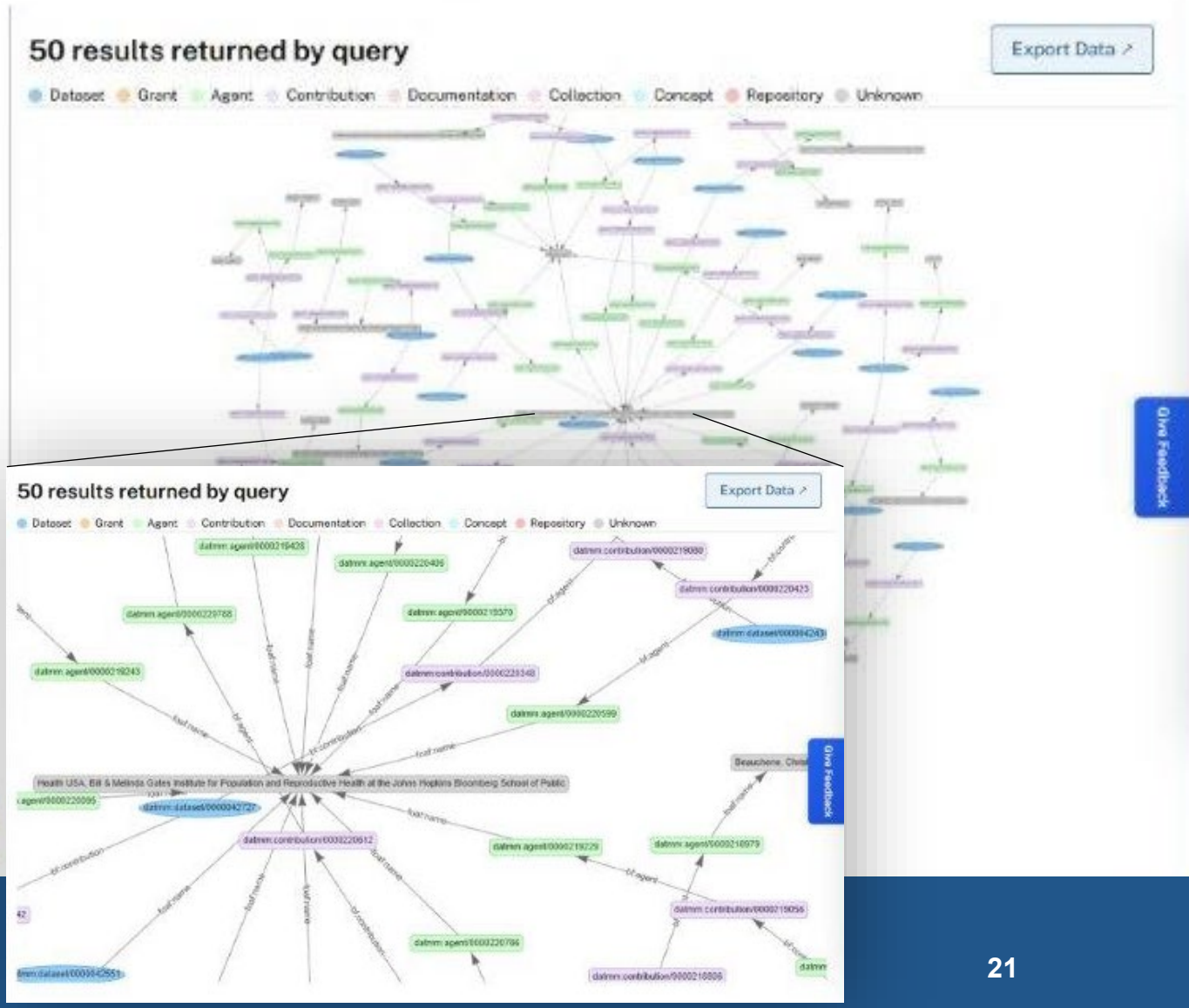
NIH National Library of Medicine


National Library of Medicine
Log in
NLM Dataset Catalog
All
(Breast neoplasm AND treatment) long term study
Search
81 search results
Save Search
Sort by:
Relevance
Time Frame
2014
2025
Custom Range
Repositories
Figshare (81)
Funding Data
Has funding information
1. Data Sheet 1_Surgical treatment of rare peripheral nerve lesions: long-term outcomes and quality of life.docx
Repository: figshare
Issued Date: 02/25/2025
Description:
IntroductionRare peripheral nerve lesions comprise a histologically diverse group of neoplastic and non-neoplastic entities, characterized by infrequent occurrence and variable clinical presentations, presenting significant diagnostic and therapeutic challenges. This study presents eight cases of surgically treated rare peripheral nerve lesions with previously unreported long-term outcomes involving quality of life (QOL)...
2. CORRELATION OF TOKUHASHI AND TOMITA SCORES WITH THE PROGNOSIS IN METASTATIC BREAST CANCER
Repository: figshare
Issued Date: 06/06/2022
Description:
ABSTRACT Objective: The purpose of the present study was to evaluate the concordance between the Tokuh
vertebral metastases due to breast tumor, treated at the outpatient clinic of the Universidade Estadual de Ca
vertebral metastases from...
Better Results
National Library of Medicine
19

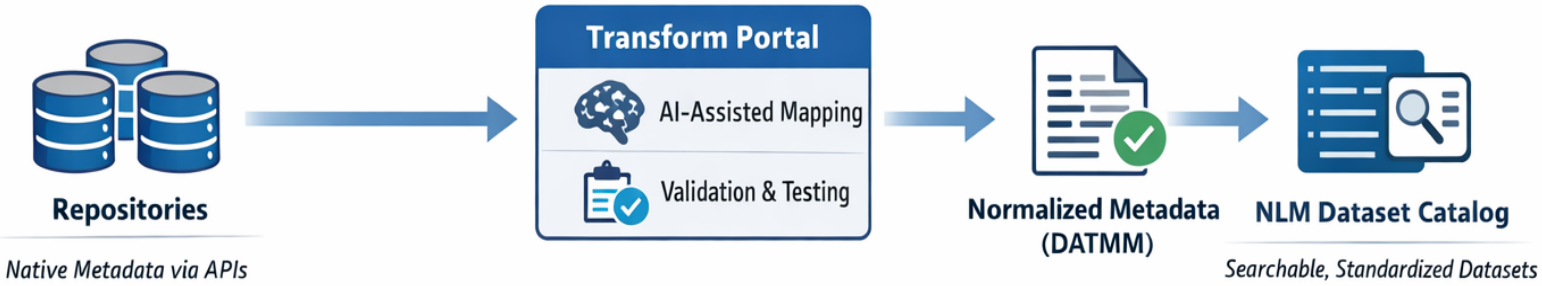
Saving Citations

Access Programmatically with SPARQL



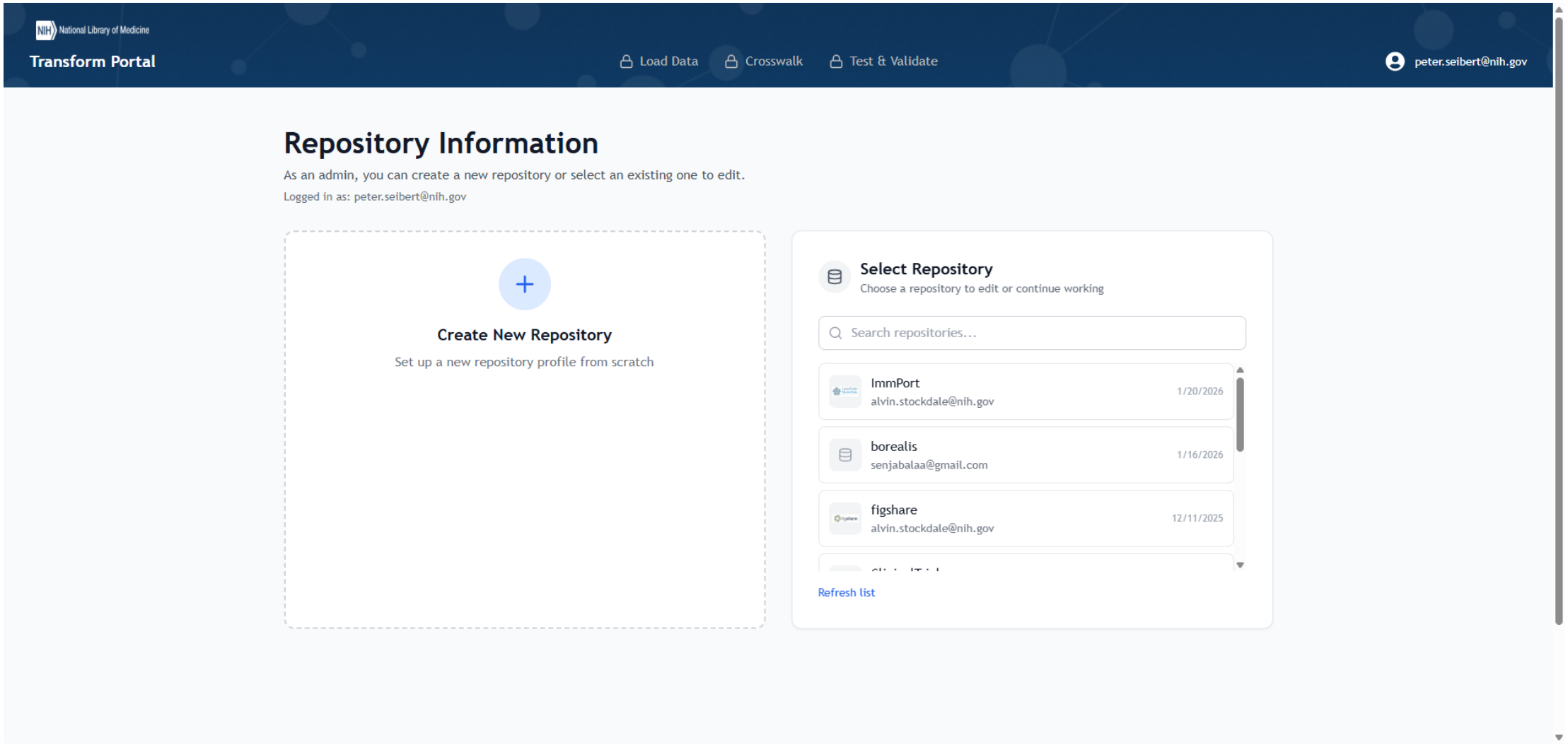


Transform Portal: Scaling Biomedical Data Access

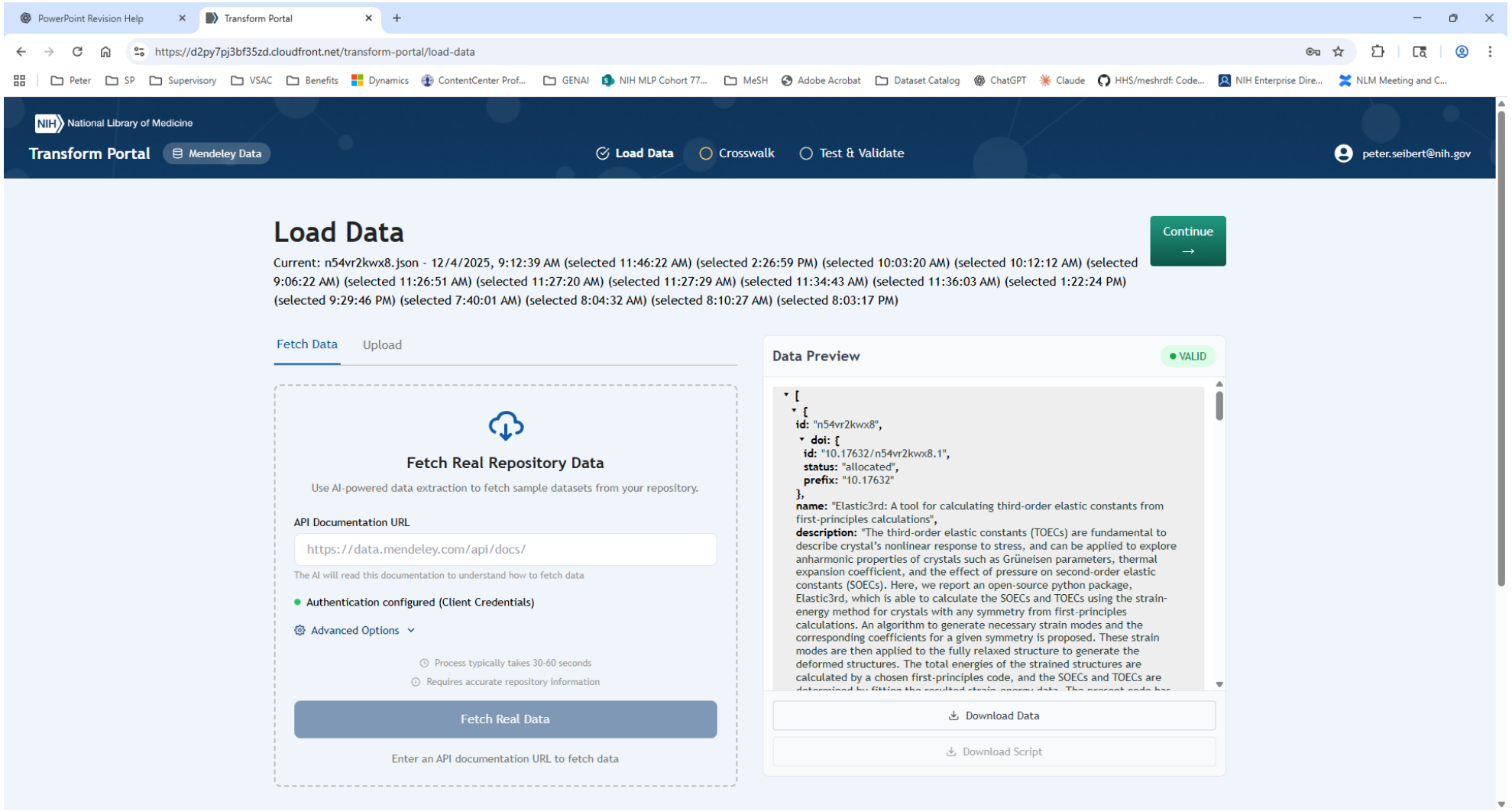


- Enables normalization of repository metadata to the NLM Dataset Catalog metadata model (DATMM)
- Uses AI-assisted methods to analyze sample metadata retrieved via repository APIs
- Recommends mappings (“crosswalks”) between repository metadata and DATMM elements
- Generates and validates transformation configurations and scripts
- Supports testing and loading of normalized metadata into the Dataset Catalog
- Automates ongoing updates to keep repository metadata current over time

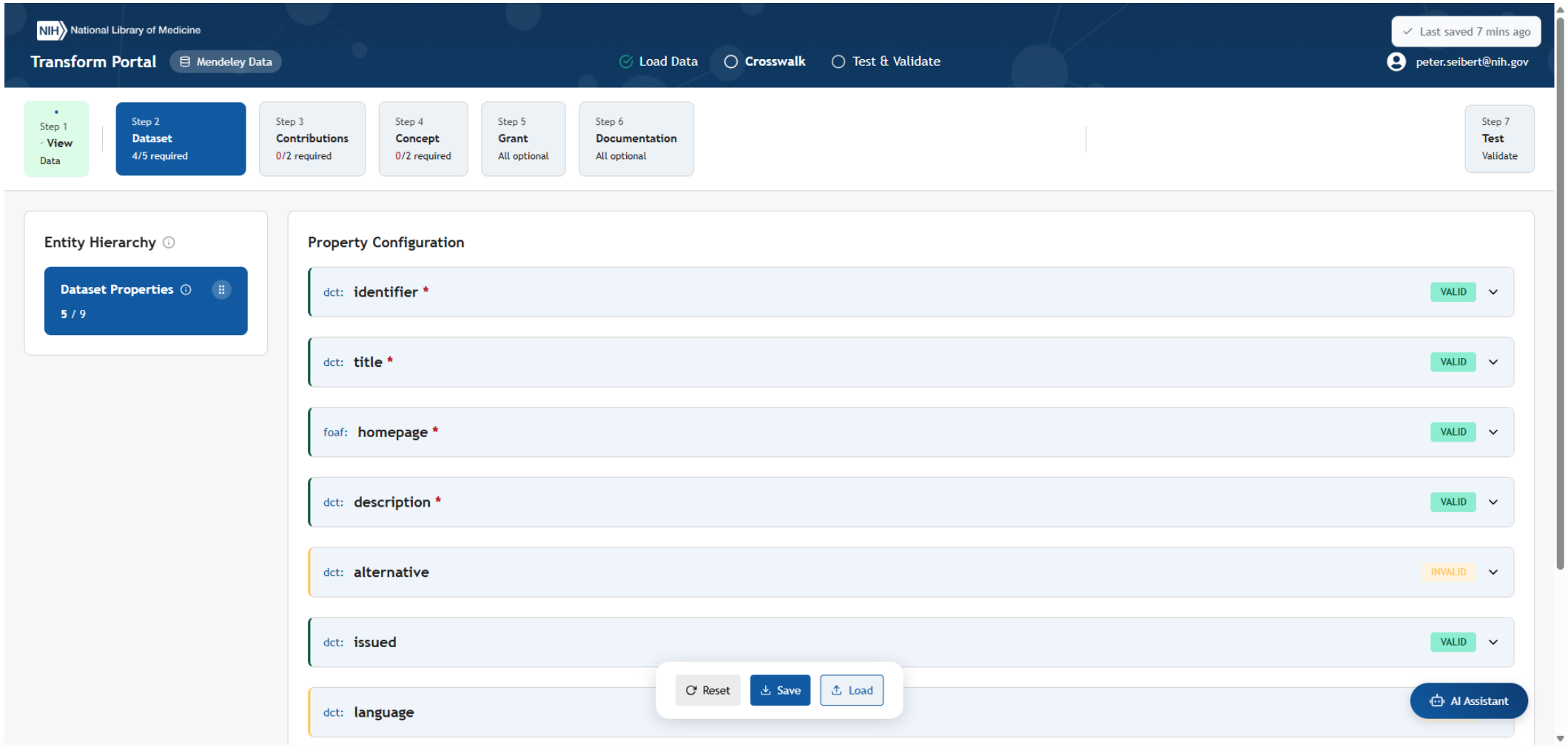
Transform Portal: Scaling Biomedical Data Access



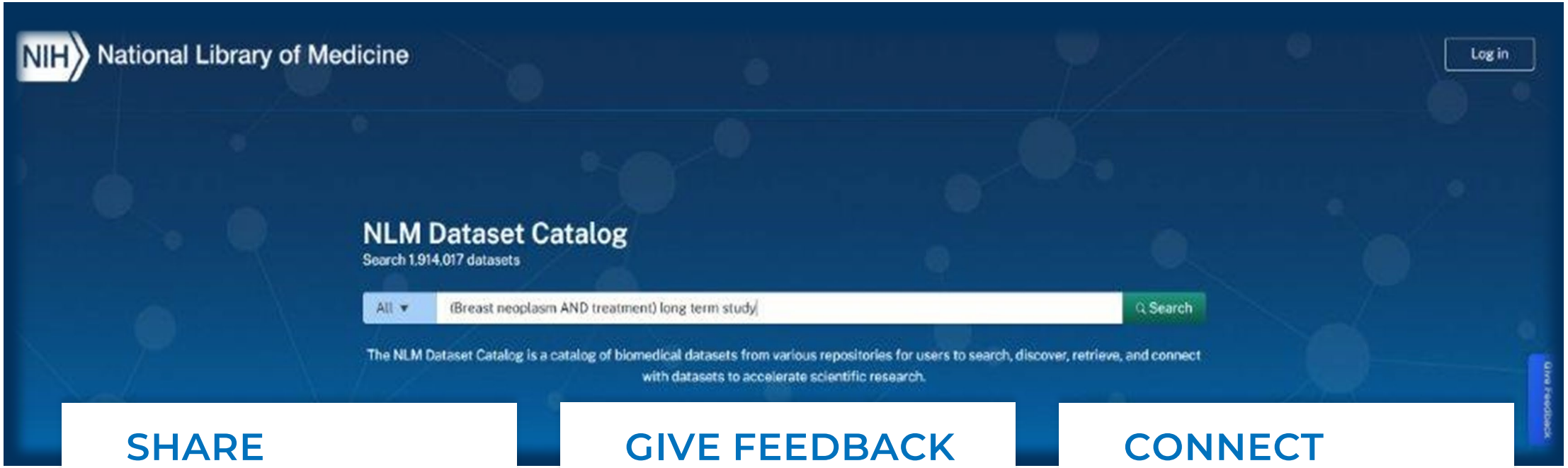
Transform Portal: Scaling Biomedical Data Access



Transform Portal: Scaling Biomedical Data Access



How You Can Help



SHARE

- Share with your colleagues and community.

GIVE FEEDBACK

- Provide feedback using the blue feedback button.

CONNECT

- Contact Peter Seibert peter.seibert@nih.gov.

Topics for Today

- Introduction
- Dataset Catalogue 2.0
- NIH Sharing Policies: NIH Public Access Policy
- Discussion

NIH Sharing Policies

2023 NIH Data Management and Sharing Policy	• Submission of Data Management and Sharing Plan to promote sharing of scientific data
NIH Genomic Data Sharing Policy	• Sharing of large-scale human or non-human genomic data
Clinical Trial Requirements for Grants and Contracts	• Registration and submission of results information to ClinicalTrials.gov
2024 NIH Public Access Policy	• Access to publications resulting from NIH funding

https://grants.nih.gov/policy-and-compliance/policy-topics

What's New in the NIH Public Access Policy?

2008 Public Access Policy	New 2024 Policy
• Final, peer-reviewed manuscripts to be **made publicly available no later than 12-months after the official date of publication**	• Author Accepted Manuscripts accepted for publication in a journal, on or after July 1, 2025, to be submitted to PubMed Central upon acceptance for publication, **for public availability without embargo** upon the Official Date of Publication.

https://grants.nih.gov/policy-and-compliance/policy-topics/public-access

A Few Definitions

- **Author Accepted Manuscript (AAM):** The author's final version that has been accepted for journal publication and includes all revisions resulting from the peer review process, including all associated tables, graphics, and supplemental material.

- **Final Published Article:** The journal's authoritative copy, including journal or publisher copyediting and stylistic edits, and formatting changes, even prior to the compilation of a volume or issue or the assignment of associated metadata.

Scope

Any Author Accepted Manuscript accepted for publication in a journal, on or after July 1, 2025, that is the result of funding by NIH in whole or in part through:

- Grant or cooperative agreement, including training grants
- Contract
- Other Transaction
- NIH intramural research
- The official work of an NIH employee

Applies regardless of whether the NIH-funded principal investigator or project director is an author and regardless of whether non-NIH funds contributed to developing or writing the Author Accepted Manuscript.

Scope (Cont.)

- Policy applies based on acceptance date of AAM
- 2008 Policy applies for awards closed prior to July 1, 2025
- 2024 Policy applies for awards active on or after July 1, 2025

Requirements

Submission	• of AAM to PMC upon acceptance for publication, for public availability without embargo upon the Official Date of Publication
Acknowledgement	• of federal funding
Agreement	• to submission statement granting NIH the right to make AAM available without embargo

Compliance: 2 Pathways

1

- Submission of the electronic version of the Author Accepted Manuscript to PubMed Central **upon its acceptance** for publication, for public availability without embargo upon the Official Date of Publication, or

2

- Submission of the Final Published Article to PubMed Central from journals or publishers with formal agreements with NLM, upon the Official Date of Publication, for public availability without embargo.

Must be in PMC. Posting on journal website alone doesn't satisfy Policy.

Government Use License and Rights

- NIH given rights to make AAM available in PMC upon the Official Date of Publication, at acceptance of funding
- Authors agree to a submission statement when submitting the AAM in the NIH Manuscript Submission System
- Authors encouraged to include statement in manuscript of Policy applicability

https://grants.nih.gov/grants/guide/notice-files/NOT-OD-25-049.html

Publication Costs

- The submission of the AAM is a free pathway to compliance
- Reasonable costs that are allowable may be paid for publication
- Fees for submission of the AAM or for a right or a license to submit an AAM to PubMed Central are not allowable

https://grants.nih.gov/grants/guide/notice-files/NOT-OD-25-048.html

Resources

- 2024 NIH Public Access Policy: https://grants.nih.gov/grants/guide/notice-files/NOT-OD-25-047.html
- Supplemental Guidance: Government Use License and Rights: https://grants.nih.gov/grants/guide/notice-files/NOT-OD-25-049.html
- Supplemental Guidance: Publication Costs: https://grants.nih.gov/grants/guide/notice-files/NOT-OD-25-048.html
- Notice of Updated Effective Date (July 1, 2025): https://grants.nih.gov/grants/guide/notice-files/NOT-OD-25-101.html
- Director's Statement: https://www.nih.gov/about-nih/nih-director/statements/accelerating-access-research-results-new-implementation-date-2024-nih-public-access-policy
- Website: https://grants.nih.gov/policy-and-compliance/policy-topics/public-access
- Frequently Asked Questions: https://grants.nih.gov/faqs#/public-access-policy
- Reputable Journals (https://grants.nih.gov/grants/guide/notice-files/NOT-OD-18-011.html) and Guidance: Requirement for Instruction in the Responsible Conduct of Research (https://grants.nih.gov/grants/guide/notice-files/NOT-OD-22-055.html)

Questions about the Policy? Email SciencePolicy@od.nih.gov

Other NIH Sharing Policy Updates

Request for Information (RFI) on Maximizing Research Funds by Limiting Allowable Publishing Costs

- Proposal to establish new policy limiting allowable publication costs for grants, contracts, and Other Transactions
- Public comments available from: https://osp.od.nih.gov/wp-content/uploads/2025/12/Compiled-Public-Comments-on-the-RFI-on-Maximizing-Research-Funds-by-Limiting-Allowable-Publishing-Costs.pdf
- RFI available from: https://grants.nih.gov/grants/guide/notice-files/NOT-OD-25-138.html

RFI on Draft NIH Controlled-Access Data Policy and Proposed Revisions to NIH Genomic Data Sharing

- NIH proposes:
  - New NIH Controlled-Access Data Policy
  - Revision of the NIH Genomic Data Sharing Policy
- Public comments due March 18
- RFI available from: https://grants.nih.gov/grants/guide/notice-files/NOT-OD-26-023.html

Questions?