

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

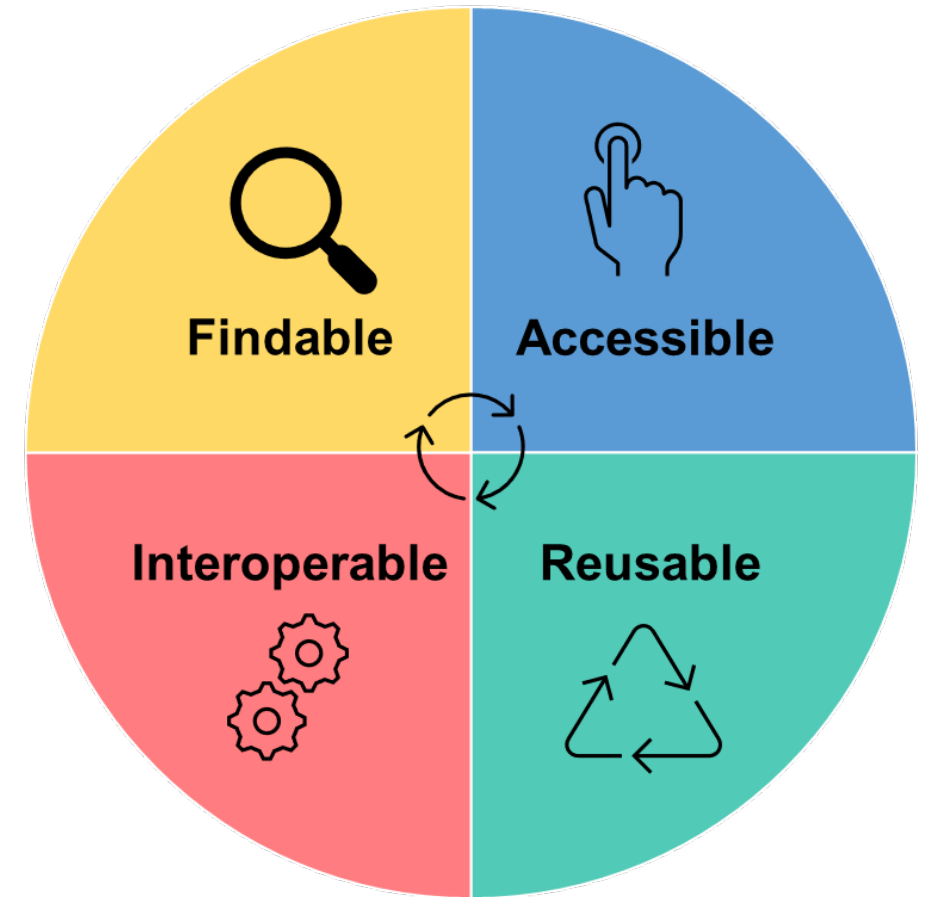
- Advance rigorous and reproducible research



- 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

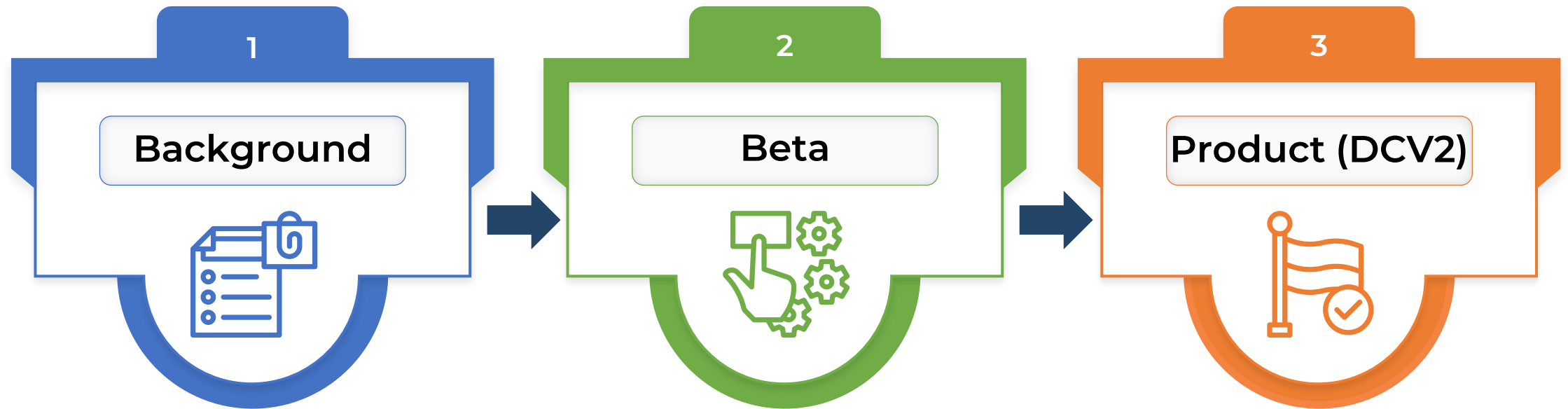
Peter Seibert



National Library of Medicine

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

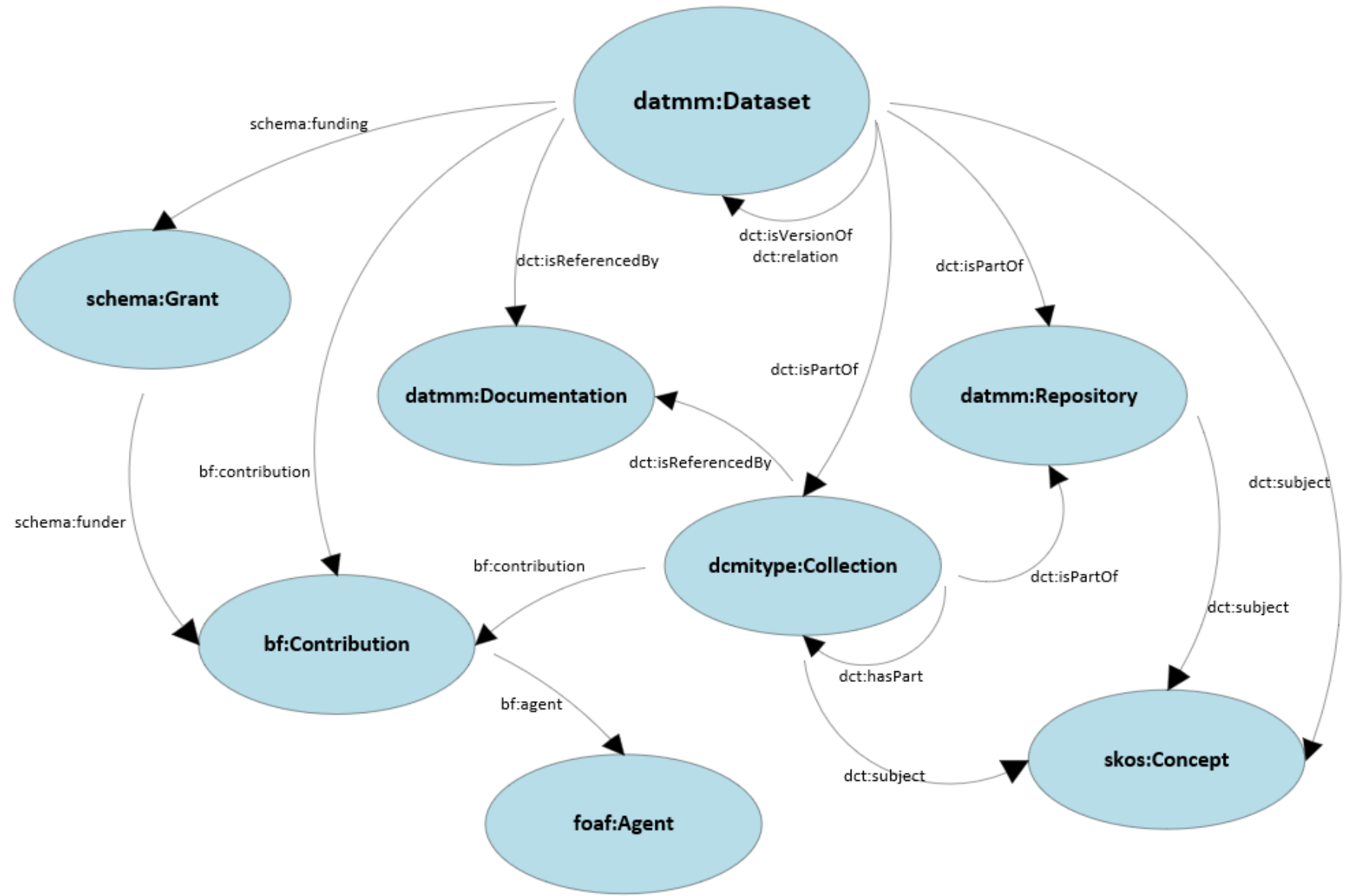
VISION

- 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



From Beta to Product (DCv2)

Current Interface (2026)

NIH National Library of Medicine

Dataset Catalog beta

Search 80,905 datasets

Use AND, OR, NOT, (). Multiple terms are searched as a single phrase.

The 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. This beta version aims to collect user feedback to inform future product development.



About

Improving discovery of research datasets



User Guide

Learn to use the Dataset Catalog



Repositories

Repositories currently in the Dataset Catalog and those being added

Beta Version (2023)

NIH National Library of Medicine
National Center for Biotechnology Information

Log in

Dataset Catalog

Search 1,880,905 datasets

All Enter your search term(s)

Search

The 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. This beta version aims to collect user feedback to inform future product development.



About

Lorem ipsum dolor sit amet, consectetur adipiscing elit



User Guide

Lorem ipsum dolor sit amet, consectetur adipiscing elit



Metadata Model

Lorem ipsum dolor sit amet, consectetur adipiscing elit



Repository Owners

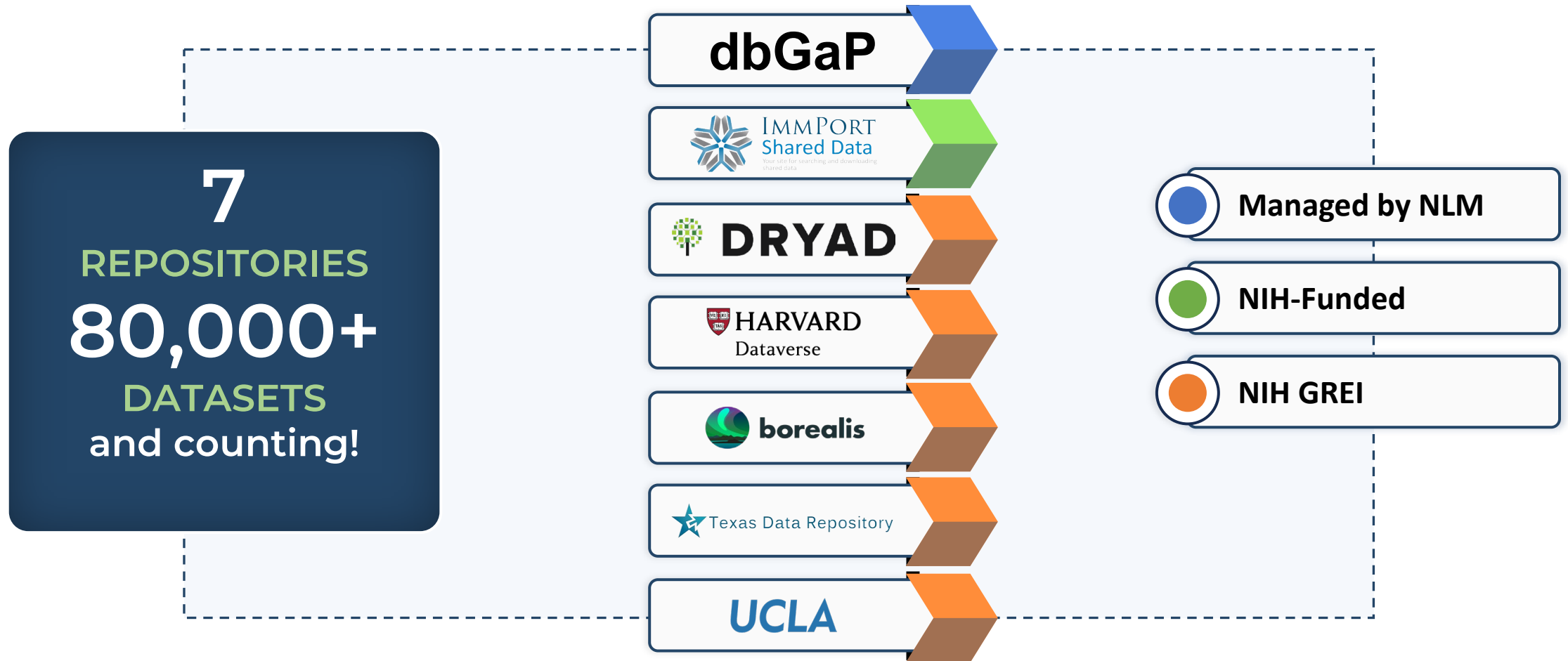
Lorem ipsum dolor sit amet, consectetur adipiscing elit



API

Lorem ipsum dolor sit amet, consectetur adipiscing elit

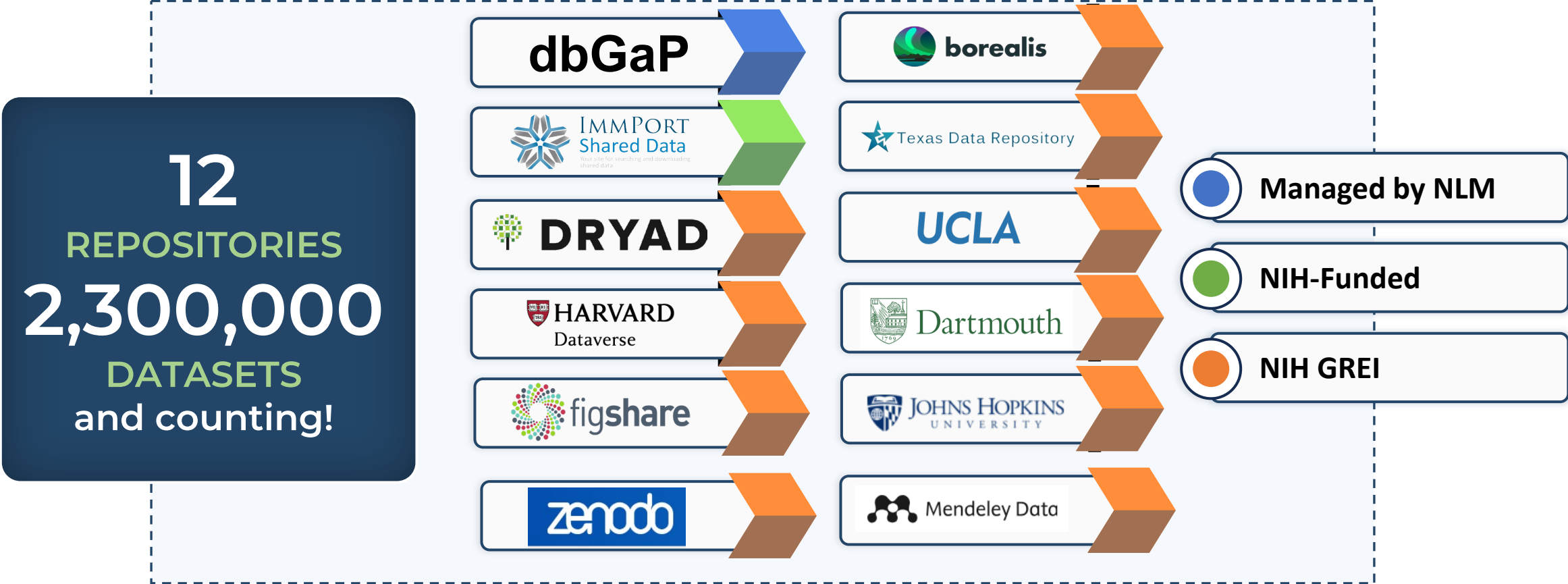
Beta Ingested Repositories



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





NLM Dataset Catalog

Search 2,322,888 datasets

All ▾

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

New Search Options

NLM Dataset Catalog

Search 2,322,888 datasets

All ▾

- All
- Title
- Funding ID
- Contributor

The Dataset Catalog is a catalog of biomedical datasets from various repositories for users to search, discover, retrieve, and connect 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

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 Save to list

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 Save to list

Repository: figshare

Issued Date: 06/06/2022

Description:

ABSTRACT Objective: The purpose of the present study was to evaluate the concordance between the Tokuhashi and Tomita scores with the prognosis in metastatic breast cancer. The study included 100 patients with vertebral metastases due to breast tumor, treated at the outpatient clinic of the Universidade Estadual de Campinas. The results showed a strong correlation between the scores and the prognosis of vertebral metastases from...

Better Results

Saving Citations

← Back to Search

Copy Link

Download Citation

figshare · 2025 Feb

Data Sheet 1_Surgical treatment of rare peripheral nerve lesions: long-term outcomes and quality of life.docx

Go to the dataset:



Contributors:

- [Micić, Aleksa](#)
- [Vujić, Lazar](#)
- [Terzić, Andrej](#)
- [Putra, Gunna Hutomo](#)
- [Savić, Andrija](#)
- [Rasulić, Lukas](#)
- [Grujić, Jovan](#)
- [Lepić, Milan](#)
- [Stojiljković, Aleksandra](#)

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)

PAGE NAVIGATION

- Contributors
- Dataset URI
- Description
- Keyword(s)
- Rights
- Grants and Funding

Access Programmatically with SPARQL

Home / SPARQL Query Builder

DATMM SPARQL Query Builder

This interface lets you explore and query the NLM Dataset Catalog using a variety of preset or custom SPARQL queries. You can filter datasets by contributor, subject, grant, and more – or write your own. To understand how these queries work and how the data is structured, see the [DATMM Vocabulary Specification](#).

Query type By Contributor By Repository By Grant By Subject By Publication Custom Query

Create your own custom SPARQL query

SPARQL Query

```
PREFIX rd: <http://rdfs.org/ns/foaf/0.1/>  
PREFIX foaf: <http://xmlns.com/foaf/0.1/>  
  
SELECT DISTINCT  
  ?dataset ?p1 ?contribution ?p2 ?agent ?p3 ?name  
WHERE {  
  ?dataset bf:contribution ?contribution .  
  ?contribution bf:agent ?agent .  
  ?agent foaf:name ?name .  
  
  BIND(bf:contribution AS ?p1)  
  BIND(bf:agent AS ?p2)  
  BIND(foaf:name AS ?p3)  
}
```

LIMIT 50

Copy Save

Submit Query

Display results as Table Compact Graph

50 results returned by query

Export Data >

Dataset Grant Agent Contribution Documentation Collection Concept Repository Unknown

50 results returned by query

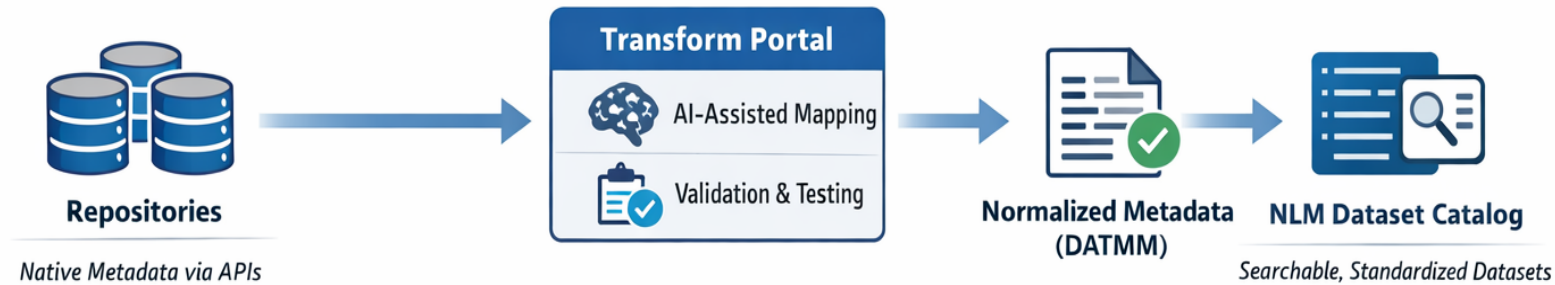
Export Data >

Dataset Grant Agent Contribution Documentation Collection Concept Repository Unknown

NIH National Library of Medicine

21

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


NIH National Library of Medicine
Transform Portal

Load Data Crosswalk Test & Validate

peter.seibert@nih.gov

Repository Information

As an admin, you can create a new repository or select an existing one to edit.
Logged in as: peter.seibert@nih.gov






Create New Repository

Set up a new repository profile from scratch

Select Repository

Choose a repository to edit or continue working

 ImmPort alvin.stockdale@nih.gov	1/20/2026
 borealis senjabalaa@gmail.com	1/16/2026
 figshare alvin.stockdale@nih.gov	12/11/2025

[Refresh list](#)

Transform Portal: Scaling Biomedical Data Access

PowerPoint Revision Help x Transform Portal x +

https://d2py7pj3bf35zd.cloudfront.net/transform-portal/load-data

NIH National Library of Medicine

Transform Portal Mendeley Data

Load Data Crosswalk Test & Validate

peter.seibert@nih.gov

Load Data

Continue

Current: n54vr2kwx8.json - 12/4/2025, 9:12:39 AM (selected 11:46:22 AM) (selected 2:26:59 PM) (selected 10:03:20 AM) (selected 10:12:12 AM) (selected 9:06:22 AM) (selected 11:26:51 AM) (selected 11:27:20 AM) (selected 11:27:29 AM) (selected 11:34:43 AM) (selected 11:36:03 AM) (selected 1:22:24 PM) (selected 9:29:46 PM) (selected 7:40:01 AM) (selected 8:04:32 AM) (selected 8:10:27 AM) (selected 8:03:17 PM)

Fetch Data Upload

Fetch Real Repository Data

Use AI-powered data extraction to fetch sample datasets from your repository.

API Documentation URL

The AI will read this documentation to understand how to fetch data

- Authentication configured (Client Credentials)

Advanced Options

- Process typically takes 30-60 seconds
- Requires accurate repository information

Fetch Real Data

Enter an API documentation URL to fetch data

Data Preview

VALID

```
{
  "id": "n54vr2kwx8",
  "doi": {
    "id": "10.17632/n54vr2kwx8.1",
    "status": "allocated",
    "prefix": "10.17632"
  },
  "name": "Elastic3rd: A tool for calculating third-order elastic constants from first-principles calculations",
  "description": "The third-order elastic constants (TOECs) are fundamental to describe crystal's nonlinear response to stress, and can be applied to explore anharmonic properties of crystals such as Grüneisen parameters, thermal expansion coefficient, and the effect of pressure on second-order elastic constants (SOECs). Here, we report an open-source python package, Elastic3rd, which is able to calculate the SOECs and TOECs using the strain-energy method for crystals with any symmetry from first-principles calculations. An algorithm to generate necessary strain modes and the corresponding coefficients for a given symmetry is proposed. These strain modes are then applied to the fully relaxed structure to generate the deformed structures. The total energies of the strained structures are calculated by a chosen first-principles code, and the SOECs and TOECs are determined by fitting the resulted static energy data. The present code has
```

Download Data

Download Script

Transform Portal: Scaling Biomedical Data Access

The screenshot displays the Transform Portal interface. At the top, the NIH National Library of Medicine logo is on the left, and a user profile for peter.seibert@nih.gov is on the right. The main navigation bar includes 'Transform Portal', 'Mendeley Data', and three active tabs: 'Load Data', 'Crosswalk', and 'Test & Validate'. A progress indicator shows seven steps: Step 1 (View Data), Step 2 (Dataset, 4/5 required), Step 3 (Contributions, 0/2 required), Step 4 (Concept, 0/2 required), Step 5 (Grant, All optional), Step 6 (Documentation, All optional), and Step 7 (Test & Validate). The main content area is divided into 'Entity Hierarchy' (showing 'Dataset Properties' 5/9) and 'Property Configuration'. The configuration table lists properties with their validation status:

Property	Validation Status
dct: identifier *	VALID
dct: title *	VALID
foaf: homepage *	VALID
dct: description *	VALID
dct: alternative	INVALID
dct: issued	VALID
dct: language	INVALID

At the bottom of the configuration area, there are 'Reset', 'Save', and 'Load' buttons, and an 'AI Assistant' button.

How You Can Help

The screenshot shows the NLM Dataset Catalog homepage. At the top left is the NIH logo and 'National Library of Medicine'. At the top right is a 'Log in' button. The main heading is 'NLM Dataset Catalog' with the subtext 'Search 1,914,017 datasets'. Below this is a search bar containing the text '(Breast neoplasm AND treatment) long term study' and a 'Search' button. A descriptive paragraph below the search bar reads: '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.' On the right side of the screenshot, there is a vertical blue button labeled 'Give feedback'.

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

What's New in the NIH Public Access Policy?

2008 Public Access Policy

- Final, peer-reviewed manuscripts to be **made publicly available no later than 12-months after the official date of publication**

New 2024 Policy

- 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

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?

