

Intended Users:	Information Professionals	Researchers
-----------------	---------------------------	-------------



Dataset Catalog

Description: The Dataset Catalog is a freely available catalog of biomedical datasets from various repositories for users to search, discover, retrieve, and connect with datasets to accelerate scientific research. Currently, the Dataset Catalog includes metadata for more than a million datasets across 15 major data repositories, including dbGap (genotypes and phenotypes), ImmPort (clinical study and trial data), Figshare, Dryad, and Harvard Dataverse. Datasets coming from different repositories may be insufficiently described, lack key metadata terms, or use different metadata standards. The Dataset Catalog standardizes the metadata, making it easier for people to find relevant datasets all in one place.

Popular uses of this product:

Information Professionals	Researchers
- Assist researchers in finding and using datasets. - Inform researcher about open licensing terms for sharing and reusing datasets - Explore DATMM, a semantic metadata model to provide information about biomedical datasets. - Use the API to integrate dataset information directly into your applications.	- Find biomedical datasets, their licensing terms, and access options. - Find publications associated with datasets. - Explore datasets related by subject or contributor.

Key Points:

1. Use the Dataset Catalog to find biomedical datasets across several different repositories. Filter by repository, time frame, and MeSH terms to find the most relevant datasets.
2. Find key information about a dataset such as access and rights, contributors, associated publications, and links to the dataset.

Considerations:

1. This resource is a catalog only and does not directly house or archive the datasets. Use the Dataset Catalog to search multiple repositories at once and find permalinks to the desired datasets.

Teaching Examples:

1. For Open Access Week, include the Dataset Catalog when highlighting OA resources. Offer an instructional session on open data and demonstrate finding relevant datasets using the Dataset Catalog.
2. Have participants create a list of keywords and MeSH terms relevant to their research question. Use those terms to search the catalog and find potentially relevant datasets. Identify other important pieces of metadata such as other keywords, repository, contributor, access details, and licensing information.

Real Life Examples:

1. A researcher receives an NIH grant and must create a data management and sharing plan (DMSP). Both Dryad and Harvard Dataverse are indexed in the Dataset Catalog, so the researcher plans to use one of those repositories to share their research data.
2. A researcher is looking for potential datasets to use on a project. They use the Dataset Catalog to search several repositories at once. They find a dataset they are interested in, but according to the Access Information, the dataset is only available to authorized users. They follow the listed URL to find more information about the dataset and how to request access.

More Information:

[User Guide](#)

[FAIR Data Principles](#)

[NIH DMSP- Selecting a Data Repository](#)

[Data Curation Primers](#)- peer-reviewed, living documents that detail a specific subject area or curation task and that can be used as a reference to curate research data (See Clinical Trials, Genetic Data, Human Participants, etc)