



Datasets

Updated: December 8, 2023.

This category includes structured collections of related digital files suitable for communication, interpretation, or processing. ([MeSH](#))

The Library produces or acquires, makes available, and serves as a permanent repository of datasets including those found in various sequence and literature databases. The Library is particularly interested in supporting the discovery and downloading of genomic data. [NCBI Datasets](#) provides web-based and command line access to these data and their associated metadata sourced from across different databases.

NLM is committed to preserving and sharing scientific data that are consistent with FAIR (Findable, Accessible, Interoperable, and Re-usable) data principles.

Data in NLM resources may have unique policies regarding curation, access, and retention. Documentation for using these resources should be consulted first.

NLM also may acquire selected datasets that are germane to research initiatives at the Library.

The [NLM Strategic Plan 2017-2027](#) noted that "[c]reating efficient ways to link the literature with associated datasets enables knowledge generation and discovery." To this end, PubMed Central (PMC) and Bookshelf contain datasets that are provided as supplemental materials to full-text articles. Additionally, PMC, Bookshelf, and PubMed records may include links to datasets held in other repositories. These data links may be to other NLM resources (e.g., GenBank) or external data repositories (e.g., figshare, Dryad).

Since many datasets of potential use to biomedical researchers and health professionals are archived by institutional repositories, such as government agencies and major research institutions, NLM may provide access to this data, as noted above.