



Datasets

Updated: September 10, 2021.

This category includes structured collections of related digital files.

The Library produces, makes available, and serves as a permanent repository of datasets such as [GenBank](#), the [Visible Human](#), the [Unified Medical Language System \(UMLS\)](#), and many other resources created or maintained by the NLM National Center for Biotechnology Information, Lister Hill National Center for Biomedical Communications, and the National Information Center on Health Services Research and Health Care Technology.

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 published as supplemental materials or otherwise linked to full-text articles. Additionally, PMC, Bookshelf, and PubMed records may include data links to datasets held in other repositories. These data links may be to records in other NLM databases (e.g., GenBank) or external data repositories (e.g., figshare, Dryad). NIH-funded authors may also deposit datasets to PMC via the NIH Manuscript Submission (NIHMS) system with a peer-reviewed author manuscript.

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.