American Brain Tumor Association®
Providing and pursuing answers®

ABTA Approved Repositories for Data and Protocols

Because of the wide range of projects funded by the ABTA, grantees may select the repository most compatible with their data. When available, subject-focused repositories are preferred over general repositories because increased discoverability of the data. If a repository emerges as the standard resource in a field (e.g., GenBank for DNA sequences), the awardee is encouraged to use that repository to better disseminate the research results to like-minded investigators interested in building upon the research. Any repository approved by the ABTA must meet the following criteria:

- *Re-Use*: The repository must guarantee to any interested party free access to the data without restriction on research reuse.
- *Security*: The repository must describe how datasets are stored, and confidential information is protected.
- *Stability*: The repository must assure that the data will be available for the indefinite future, regardless of whether the repository is dismantled.
- *Fee Structure*: The repository must define a fee structure to the investigator depositing the data (not to a third-party user accessing the data).
- *Metadata*: The repository must require the awardee to provide sufficient metadata to explain the data to others. These metadata must be searchable so that repository visitors can easily locate desired datasets.
- *File Formats*: The repository should accommodate all file types generated by the awardee.
- *Machine Extraction*: Preferably, the repository will feature machine-readable and machine-interpretable functionality to enable third-party users to more easily locate the data.
- *Reception to ABTA Data*: The repository must be willing to accept data submitted by ABTA-funded researchers.

If a desired repository is not currently approved by the ABTA, applicants may request in the data sharing plan that the ABTA consider the repository for approval. If the repository is approved, it will be added to the list of ABTA Acceptable Repositories.

Approved Data Repositories

ArrayExpress: ArrayExpress Archive of Functional Genomics Data stores data from high-throughput functional genomics experiments, and provides these data for reuse to the research community.

BLOODPAC Data Commons: The Blood Profiling Atlas in Cancer contains scientific and clinical liquid biopsy data from blood, saliva, urine, and cerebral spinal fluid, across many types of cancer.

BRAIN Commons: BRAIN (Brain Research and Innovation Network) Commons aims to create a place in a cloud that is accessible to everyone interested in contributing to our understanding of brain disorders and diseases.

cBioPortal: The cBioPortal for Cancer Genomics provides visualization, analysis and download of large-scale cancer genomics data sets.

ClinicalTrials.gov: ClinicalTrials.gov is a Web-based resource that provides patients, their family members, health care professionals, researchers, and the public with easy access to information on publicly and privately supported clinical studies on a wide range of diseases and conditions.

COSMIC: The Catalogue Of Somatic Mutations In Cancer, is the world's largest and most comprehensive resource for exploring the impact of somatic mutations in human cancer.

dbGAP: The database of Genotypes and Phenotypes (dbGaP) was developed to archive and distribute the data and results from studies that have investigated the interaction of genotype and phenotype in Humans.

European Genome-Phenome Archive: EGA is a service for permanent archiving and sharing of personally identifiable genetic, phenotypic, and clinical data generated for the purposes of biomedical research projects of in the context of research-focused healthcare systems.

Figshare: Figshare allows users to upload any file format to be previewed in the browser so that any research output, from posters and presentations to datasets and code, can be disseminated in a way that the current scholarly publishing model does not allow.

Flow Repository: FlowRepository is a database of flow cytometry experiment, primarily for experimental findings published in peer-reviewed journals in the flow cytometry field.

GDC: The Genomic Data Commons (GDC) is a research program of the National Cancer Institute (NCI). The mission of the GDC is to provide the cancer research community with a unified data repository that enables data sharing across cancer genomic studies in support of precision medicine.

GenBank: GenBank is an annotated collection of publicly available DNA sequences available through the National Center for Biotechnology Information databases. GenBank contains over 135,000,000 sequence records and is updated every two months. GenBank is part of the International Nucleotide Sequence Database Collaboration along with the DNA DataBank of Japan and the European Molecular Biology Laboratory.

GENIE: AACR Project Genomics Evidence Neoplasia Information Exchange is a multi-phase, multi-year, national and international project that aggregates and links clinical-grade cancer genomic data with clinical outcomes from cancer patients. The data within GENIE are shared with the global research community after defined periods of time through cBioPortal and the Synapse Platform.

GitHub: Repository for open-source code.

Metabolomics Workbench: The Metabolomics Workbench serves as a national and international repository for metabolomics data and metadata and provides analysis tools and access to metabolite standards, protocols, tutorials, training, and more.

Nanomaterial Registry: The Nanomaterial Registry is an authoritative, fully curated resource that archives research data on nanomaterials and their biological and environmental implications.

NCBI BioProject: A BioProject is a collection of biological data related to a single initiative, originating from a single organization or from a consortium. A BioProject record provides users a single place to find links to the diverse data types generated for that project.

NCBI BioSample: The BioSample database contains descriptions of biological source materials used in experimental assays.

NCBI GEO Datasets: This database stores curated gene expression DataSets, as well as original Series and Platform records in the Gene Expression Omnibus (GEO) repository. Enter search terms to locate experiments of interest. DataSet records contain additional resources including cluster tools and differential expression queries.

NCBI Protein: The Protein database is a collection of sequences from several sources, including translations from annotated coding regions in GenBank, RefSeq and TPA, as well as records from

SwissProt, PIR, PRF, and PDB. Protein sequences are the fundamental determinants of biological structure and function.

NCBI Reference Sequence: A comprehensive, integrated, non-redundant, well-annotated set of reference sequences including genomic, transcript, and protein.

Open Science Data Cloud: The Open Science Data Cloud provides the scientific community with resources for storing, sharing, and analyzing terabyte and petabyte-scale scientific datasets.

Open Science Framework: Open Science Framework is a repository hosted by Center for Open Science, which is a non-profit technology company providing free and open services to increase inclusivity and transparency of research.

PRIDE: The PRIDE PRoteomics IDEntifications (PRIDE) database is a centralized, standards compliant, public data repository for proteomics data, including protein and peptide identifications, post-translational modifications and supporting spectral evidence.

ProteomeXChange: ProteomeXchange Consortium was established to provide globally coordinated standard data submission and dissemination pipelines involving the main proteomics repositories, and to encourage open data policies in the field.

PubChem: PubChem is an open chemistry database at the National Institutes of Health (NIH). PubChem collects information on chemical structures, identifiers, chemical and physical properties, biological activities, patents, health, safety, toxicity data, and many others.

Sequence Read Archive: Sequence Read Archive (SRA) makes biological sequence data available to the research community to enhance reproducibility and allow for new discoveries by comparing data sets. The SRA stores raw sequencing data and alignment information from high-throughput sequencing platforms, including Roche 454 GS System®, Illumina Genome Analyzer®, Applied Biosystems SOLiD System®, Helicos Heliscope®, Complete Genomics®, and Pacific Biosciences SMRT®.

Sage Bionetworks: A data sharing platform for accessing data, analyses, and research toolkits to accelerate discoveries and their translation to clinical outcomes.

Single Cell Portal: From the Broad Institute, this platform is open-access for sharing and exploring single-cell genomics data.

The Cancer Imaging Archive: TCIA hosts a large archive of medical images of cancer accessible for public download. Supporting data related to the images such as patient outcomes, treatment details, genomics and expert analyses are also provided when available.

Zenodo: Zenodo is a general repository that collects all research outputs from across all fields of research, and accepts any file format as well as both positive and negative results.

Approved Protocol Repositories

Bio-protocol: Bio-protocol is a peer-reviewed open-access protocol journal with no publication fees. Protocols are validated and updatable.

Protocol Exchange: The Protocol Exchange is an Open Repository for the deposition and sharing of protocols for scientific research. These protocols are posted directly on the Protocol Exchange by their authors and hence have not been further styled, peer reviewed or copy edited. Rather they are made freely available to the scientific community for use and comment.

Protocols.io: Protocols.io is a free, central, up-to-date, crowd-sourced protocol repository.

Scientific Protocols: Scientific Protocols is a free and easy way to share scientific protocols, with the intent to encourage transparency in research and promote reproducibility of published results. Every protocol created on the website is also a GitHub Gist.