


TeraGrid™
[About](#)
[News](#)
[Outreach](#)
[Science Gateways](#)
[User Support](#)
[User Support](#)
[Resources](#)

Data Collections

[Access](#)
[Compute](#)
[Data & Visualization](#)
[Home > User Support](#)

What Is a Data Collection?

[Portal](#)

Data collections represent permanent data storage that is organized, searchable, and available to a wide audience, either a collaborative group or the scientific public in general. Data collections usually have a Web interface or portal for displaying and retrieving data.

The term "collections" is also used to refer to the libraries or groupings of data in [Storage Resource Broker](#) (SRB), which is a client/server-based suite of data storage and movement tools.

Hosting Your Collection on the TeraGrid

To make your collection available as a TeraGrid resource, you will need a data allocation. See the [TeraGrid Resource Catalog](#) to see allocable data resources, and see [Allocations and Accounts](#) for information on requesting an allocation. We recommend that first-time TeraGrid users utilize the [TeraGrid Getting Started Guide](#).

What Collections Are Available on the TeraGrid?

A number of data collection resources are available at the TeraGrid sites. The data collections table below lists collections that are available to or created by TeraGrid users. The list contains a brief description or abstract of each collection that is currently in production at TeraGrid sites. URL links will connect directly to the collection interface or to more detailed information. Collections without URLs are used by a community of researchers but are not available to the public.

[TACC](#)

Related links

- [Data and Visualization Overview](#)
- [Accounts and Allocations](#)
- [TeraGrid Resource Catalog](#)
- [Science Gateways](#)

TeraGrid Getting Started Guide

Need Help?

- [Phone Toll-free 1.866.907.2383](#)
- [Submit a Ticket \(online form -- fastest\)](#)
- [Submit a Ticket via email](#)
- [TeraGrid Knowledge Base](#)

Table of Collections

Collection	Description	How to Access	Size
IU			
BioMirror	Several genome databases, combined distribution	http://www.bio-mirror.net	3 TB
BIOSCI/Bionet	A set of electronic communication forums - the bionet USENET newsgroups and email lists.	http://www.bio.net	0.01
Central Life Sciences Data (CLSD)	A collection of Life Sciences data from a variety of public sources that may be queried as if it were a unified database. BLAST is rendered as a relational operation within this database.	http://racinfo.indiana.edu/clsd/	7 GB
ChEMBL	Database of chemical structures based on compounds published in PubChem, including Pub3D and PubDock	http://www.ebi.ac.uk/chembl/	36.4 TB
CLIOH	A collection of images, video, sound, narratives, VR and graphical recreations of endangered cultural heritage sites worldwide.	http://clioh.informatics.iupui.edu/	
Daphnia (water flea) genome	Annotated Genome of Daphnia (water flea)	http://daphnia.cgb.indiana.edu	
DisProt	Database of proteins that contain regions of intrinsic disorder or that are entirely disordered as determined by one or more of 30 different experimental methods	http://www.disprot.org	
DroSpeGe	Comparative Drosophila Species Genomes database	http://eugenics.org/drospege	0.1
EthoBank	A public repository for animal behavior data. It is part of EthoSource, a global initiative to store, share, and combine behavioral information.	http://www.indiana.edu/~ethobank	
euGenes	euGenes provides a common summary of gene and genomic information from various eukaryotic organism databases.	http://iubio.bio.indiana.edu:8089/	
FlyBase	FlyBase is a repository of data on fruit fly (Drosophila spp.) anatomy, physiology, genetics, genomes, gene expression, gene products, publications and researchers. FlyBase's section of 'Bulk Data Downloads' is available on the TeraGrid by grid ftp.	http://flybase.bio.indiana.edu/	2.3 GB
HAPPI	A comprehensive high-quality human protein interaction data	http://bio.informatics.iupui.edu/HAPPI	
HIP2	A repository of health human plasma protein database	http://bio.informatics.iupui.edu/HIP2	
Indiana GIS data	A comprehensive collection of Indiana geospatial data, including aerial photos, topographic maps, and digital elevation data.	http://www.indiana.edu/~gisdata/	7.5 TB
IUBio Archive	Archive of Biology software and data	http://iubio.bio.indiana.edu	
MutDB	Annotated human variation data with protein structural information and other functionally relevant information	http://mutdb.org	
Network Workbench	A Large-Scale Network Analysis, Modeling and Visualization Toolkit for Biomedical, Social Science and Physics Research.	http://nwb.slis.indiana.edu	
SBLEST	Structure-Based Local Environment Search Tool uses vectors of amino acid structural environments to perform K Nearest Neighbor queries against a database of protein structures.	http://sblest.org	

Scholarly Database	Approximately 18 million publications, patents and grants, ten percent of which contain full-text abstracts	https://sdb.slis.indiana.edu	
wFleaBase	Daphnia Genome Database	http://wfleabase.org	0.05
Purdue			
Climate Modeling Data	This dataset is the output from the Community Climate System Model (CCSM) to simulate global climate changes. It consists of four dynamic geophysical models simulating the atmosphere, ocean, land surface and sea-ice, and one central coupler component. It facilitates fundamental research on the earth's past, present, and future climate states. This set of data is in NetCDF file format.	http://www.purdue.teragrid.org/portal	400 GB
LARS Data	Provided by Laboratory for Applications of Remote Sensing (LARS) at Purdue University, the LARS image data collection includes multi-spectral (3 to 15 or so wavelength bands) and hyper-spectral (dozens to hundreds of wavelength bands) image data that are being used for education and research purposes. Most of the image data are for locations in the State of Indiana, dated from 1972 to 2004. The data have been collected by satellite-borne and aircraft-borne sensors. Each band of data in a data set represents the energy received by the sensor that is within the wavelength range of that band. Different sensors cover different portions of the optical spectrum. The primary data formats are ERDAS LAN, Leica Geosystems Imagine, GeoTIFF, and HDF. In addition, some are in LARSYS MIST format.	http://www.purdue.teragrid.org/portal http://www.indianaview.org/	150 GB
National Weather Service (NWS) Doppler Radar Data	Purdue University is one of the country's four top-level distributors of high-resolution radar data from the national network of Next Generation Radar (NEXRAD). The NEXRAD Radar system comprises 159 Weather Surveillance Radar-1988 Doppler (WSR-88D) sites across the United States and in selected overseas locations. NWS data are collected and redistributed in real time. Level II data are in three meteorological base data quantities: reflectivity, mean radial velocity, and spectrum width.	http://www.purdue.teragrid.org/portal	100 GB
Purdue Terrestrial Observatory (PTO) satellite data	Provided by Purdue Terrestrial Observatory (PTO), the PTO image data sets currently include real-time remote sensing data from the GIVSSR sensor on the GOES-12 (also called GOES-East) satellite. These data are collected every 15 to 30 minutes covering different portions or sectors of the earth disk. A full disk scan sector is obtained every 3 hours. The current data products published online include the most recent images for the eastern United States and images for the continental United States in JPEG and HDF formats. PTO image data sets also include data from a 4.5 meter tracking antenna including that from the AVHRR sensor on several NOAA satellites, the MVISR sensor on the Chinese Fengyun 1D satellite and the MODIS sensors on the NASA Terra and Aqua satellites. The current online data products include JPEG images of the most recent pass coverages.	http://www.purdue.teragrid.org/portal http://www.itap.purdue.edu/pto/	1.4 TB (tape)
SDSC			
AfCS Molecule Pages	The Alliance for Cellular Signaling (AfCS) performs comprehensive experimental analyses of selected signaling systems and archives these data for the research community in the Data Center of the Signaling Gateway. These include: database of cell signaling protein information, interaction and function of signaling proteins, AfCS experimental data and Molecule Pages curated/automated data.	http://www.signaling-gateway.org/ http://www.signaling-gateway.org/data/Data.html#	462 GB
Alexandria Digital Library	Photographs		
APOPTOSIS DB	Proteins related to cell death data	http://www.apoptosis-db.org/	
Backbone Packet Header Traces	Anonymized OC48 data is available for use by academic researchers and US government agencies. The dataset is also available for corporate entities (including corporate researchers) who participate in CAIDA's membership program.	http://www.caida.org/funding/trends/data/ http://www.caida.org/analysis/measurement/oc48_data_request.xml	
Backscatter Data	From UCSD network telescope	http://www.caida.org/research/security/telescope/	
Biocyc (SRI)	Collection of Pathway/Genome Databases		
Bionome	Biology network of modeling efforts		
BIRN	Biomedical Informatics - neuroscience data		5416 GB
Braindata	Rutgers neuroscience collection		
CHRONOS	Interactive chronostratigraphy and stratigraphic database - Federated Geog. DBs	http://www.chronos.org/	
CKAAPS	Protein evolutionary information	http://ckaaps.sdsc.edu/	
CUAHSI	Community hydrological collection		
DigEmbryo	Visible Embryo-human embryology Support for the Visible Embryo digital library and image archive. Accessed through a portal at Armed Forces Institute.		720 GB
Digital Earth Data Library	Earth sciences related datasets		300 GB
Digsky	2Mass (2 Micron All Sky Survey), DPOSS (Digital Palomar Observatory Sky Survey Collection), NVO (National Virtual Observatory)	http://astro.ncsa.uiuc.edu/catalogs/dposs/	51380 GB
EarthRef Digital Archive	Earth Science information		
EcoGrid			
Encyclopedia of Life	Genomic data	http://www.eolproject.info/	
GAPP	Great Ape Phenome Project - Primate information		
GEON	The Geosciences Network	http://www.geongrid.org/	
GEOROC	Petrological and geochemical data for igneous rocks	http://georoc.mpch-mainz.gwdg.de/	

GERM	Earth reservoir information- the Geochemical Earth Reference Model	http://earthref.org/GERM/index.html	
Hayden Planetarium Collection	Hayden Planetarium Collection		7210 GB
HPWREN	High Performance Wireless Research and Education Network Sensor Network Data & Wireless Network Analysis Data	http://hpwren.ucsd.edu/	1-2 TB
HyperLter	HyperSpectral Images		233 GB
Interpro Mirror	Protein data		
IPBIR	Primate information- Integrated primate biomaterial and information resource		
JCSG	Structural genomics data		
Kd's DB	Rocks and minerals		
KNB	Knowledge networks for biocomplexity		
LDAS	Land data assimilation system- forecast simulations by numerical weather prediction models		
LDAS/SALK	land/neuro		4562 GB
NARA/Collection	Archival Documents		63 GB
National Archives	Persistent archive		
NAVDAT	Geochemistry data - Western North American Volcanic and Intrusive Rock Database	http://navdat.geongrid.org/	
Network Topology Data	Skitter macroscopic network topology project	http://navdat.geongrid.org/	
Nobel Foundation Mirror	Nobel Foundation Information		
NPACI	NPACI Users - scientific simulation output		17578 GB
NSDL/CI	K-12 Curriculum Web-sites-education collection		2785 GB
NSDL-NSDL	National Science, Mathematics, Engineering, and Technology Education Digital Library. Implementation of a persistent archive of education material registered into the NSDL repository.	http://crs.nsdl.org/collection http://nsdl.org/	
NSDL/SIO Exp	SIO Explorer Documents-oceanographic voyages		
PETDB	Petrological and chemical data- Petrological Database of the Ocean Floor	http://www.petdb.org/petdb/	
PlantsP	Plant kinase/phosphatase information	http://plantsp.sdsc.edu/	
PlantsT	Plant transporter information	http://plantst.sdsc.edu/	
PlantsUBQ	Plant protein information	http://plantsubq.sdsc.edu/	
PMAG	Paleomagnetic information		
Portal	Grid Portal	http://www.afcs.org/	1745 GB
Protein Data Bank	Protein data	http://www.rcsb.org/	
Protein Kinase Resource	Protein Kinase Resource	http://pkrr.sdsc.edu/	
Protein Mutation Resource	Protein mutations/ effect on structure	http://pdbnode8.sdsc.edu:8080/pdb/static.do?p=pmr/	
ROADNET	Real time observatories, applications and Data Management Network - Sensor data		
SALK	SALK biology data archive		
San Diego and Tijuana Watersheds	Water resources mapping		0.1 GB
San Diego Conservation Resources Network	Sensitive species map server		3 GB
SCEC	The Southern California Earthquake Center (SCEC), headquartered at the University of Southern California, hopes to gather new information about earthquakes in Southern California, integrate this information into a comprehensive and predictive understanding of earthquake phenomena, and to communicate this understanding to the general public in order to increase earthquake awareness.	http://visservices.sdsc.edu/projects/scec/	15246 GB
Scripps	Oceanographic research data		
SeamountsOnline	Seamounts species distributions	http://seamounts.sdsc.edu/	1 GB
Security logs and archives	Security information		
SEEK	Ecology data - ecological niche modelling applications	http://biodi.sdsc.edu/ww_home.html	
SIO Exp	SIO Explorer Documents-oceanographic voyages	http://nsdl.sdsc.edu/tools/index.html	
SLAC	Protein Crystallography		
SLAC/JCSG	Protein Crystallography-structural genomics data	http://www.jcsg.org/	4137 GB
Small Molecule Database	Quantum calculations on small molecules		
Structural Genomics Targets DB	Models of structural genomics targets	http://spam.sdsc.edu/perl/browser_beta.pl	
TeraBridge	Robust infrastructure for data management, querying, and mining of sensor stream data		
TeraGrid	TeraGrid science and engineering collections	http://portal.teragrid.org/ http://www.teragrid.org/	80354 GB
Transana	Classroom Videos		92 GB
Transport Classification Database	Protein information - membrane transport protein information	http://www.tcdb.org/	
TreeBase	Phylogeny and ontology information	http://www.treebase.org/	
UCSDLib - Libraries Image Collection	Archival Image Files - ArtStore		
WebBase	Web crawls		127 GB
Whole gene ontology data			
WhyWhere	WhyWhere Niche Modeling Image files Offered in a desktop version. Includes access to a remote SRB datastore of almost 1000 global coverage datasets	http://www.landshape.org/enm/	19 GB
Yeast regulatory network	Yeast gene/protein interaction information	http://imhotep.ucsd.edu:7873/knownme/ygrn.html	0.1 GB
TACC			

LiDAR	This is a LiDAR-based digital terrain collection from the Bureau of Economic Geology (BEG) containing high-resolution data on the Texas and California coastlines, Brownsville and Wyoming. This data is used by the geological science community to study shoreline migration, geomorphic change, conduct hydro-modeling and wetlands classification.		0.15 TB
MODIS Satellite Imagery of the Earth	This collection is from the Center for Space Research and contains remotely sensed imagery of the earth, including derived products such as NDVI, cloud classification, surface temperature, surface type, and aerosol optical thickness. It is used by a wide range of Earth scientists in areas such as detecting the level and evolution of air pollution sources or tracking algae blooms in bodies of waters. There are a total of 14 feeds from a handful of satellites that result into numerous derivative products. This data collection grows by approximately 2.4 GB per day.		6 TB
NEXRAD	The Center for Research in Water Resources will provide data from the National Weather Service's Next Generation Weather Radar (NEXRAD) Program. This data will be streamed directly from the NOAA port in Boulder and will be used to study various atmospheric conditions, such as cold fronts, dry lines, and thunderstorm gust fronts, never before visible within storms. (Note: the Global Hydrology Atlas which was specified in TACC's original TeraGrid proposal is insufficiently large to justify TG hosting, for this reason, the collection was altered to be more in line with the TG mission.) National coverage for this collection at the ten minute resolution should produce on the order of approximately 1.5 TB per year.		1.2 TB
UTCT	The TACC UTCU data collection makes allows browsing, downloading, and analysis of high-resolution X-ray computed tomography (HRXCT) data of biological and geological specimens.	http://utct.tacc.utexas.edu/about/utct.php	0.25 TB

Results generated from TeraGrid Resources Database.
If you are not able to find the information you're looking for, email the [TeraGrid Help Desk](#).



The TeraGrid project is funded by the [National Science Foundation](#) and includes 11 partners:
[Indiana](#), [LONI](#), [NCAR](#), [NCSA](#), [NICS](#), [ORNL](#), [PSC](#), [Purdue](#), [SDSC](#), [TACC](#) and [UC/ANL](#).

Please email help@teragrid.org with questions or comments.

The TeraGrid 1.0 Transitional CSS compliant.

