MAGLAB DC FIELD FACILITY USER DATA MANAGEMENT PLAN

Contact: Scott Hannahs (sth@magnet.fsu.edu)

Funder: National Science Foundation Division of Materials Research

Last modified: 2023-01-10

ABSTRACT:

Ensuring that publicly-funded research data is preserved and freely available safeguards efficient use of government resources and facilitates efficient delivery of scientific discoveries to maximize impact. The National Science Foundation (NSF) supports FAIR (**F**indability, **A**ccessibility, **I**nteroperability, and **R**euse) data guiding principles [1], and considers data management planning as integral to any NSF-funded research. Therefore, products of research generated at MagLab user facilities should be made available to the scientific community and general public. Specifically, this policy requires that all research products generated at the MagLab DC Field Facility (DCFF) be digitally accessible upon publication, or within 3 years. This data management plan (DMP) details resources available to DCFF users, and outlines procedures for managing data and the products of research in alignment with FAIR principles.

[1] http://www.go-fair.org/fair-principles

PRODUCTS OF THE RESEARCH

Raw data vary widely by experiment but generally consist of multiple columns of numerical values representing applied magnetic field, temperature, voltages, frequency, etc. Metadata and the products generated by research vary depending on the material and experiment being conducted. Users are responsible for capturing and organizing descriptions of samples, protocols for their preparation, and relevant quantitative and qualitative information about the samples (e.g. experimental conditions, crystal structure, doping levels, etc.). DCFF personnel will assist users with the experimental setup, instrumentation, cryostat and magnet operation, etc. and may also assist with analysis of experimental data and metadata resulting from work taking place at the MagLab. This includes details of any further on-site sample and/or experimental stage preparation prior to placing the sample in the high field environment, as well as any products of research resulting from data analysis.

DATA FORMAT

Users that employ the DCFF data acquisition computers record and store raw data in the form of ASCII or UTF-8 encoded plaintext files in a columnar format, e.g. tab-separated values (.tsv) files. The file formats of research products (including data and metadata) generated by experiments vary depending on raw data and application type. Data may also be acquired on a user-owned computer employing a format and organizational structure as dictated by the needs of the user and the experiment.

All MagLab-developed formats are open. Specifications and software to read and analyze data in these formats is available to the scientific community for free or at nominal reproduction costs. These software tools are provided on laboratory web sites and software storage areas. Metadata can be recorded with the raw data files at the option of the researchers. Other metadata is recorded in the users written notebooks, computer files, or other media at the option of the principal investigator. Management of the metadata associated with standard data files is exclusively the purview of the principal investigator.

DATA SHARING AND ACCESS

RESPONSIBILITIES OF THE PRINCIPAL INVESTIGATOR

The Principal Investigator (PI) is the steward of the research data, will select the vehicle(s) for publication or presentation of research products, and will have ultimate authority in their initial use.

Research activities detailed in DCFF user proposals and approved for magnet time are expected to result in presentations, publications, or other vehicles for dissemination of the data and results. Details of experimental work and metadata (e.g., description of samples, experimental protocols, algorithm specifications, database schemas, etc.) should be included with published data. Published manuscripts should include digital object identifiers (DOIs) and other appropriate persistent identifiers to indicate where relevant data and metadata can be accessed.

It is the responsibility of the PI to ensure protection of privacy, confidentiality, intellectual property, national security, or other rights or requirements. The PI is encouraged to disclose such requirements to DCFF staff listed as collaborators and participants to the extent necessary to facilitate compliance. Additionally, the PI must comply with all public access requirements that are laid out by other funding agencies sponsoring the research.

The NSF Public Access Policy [2] requires PIs who publish peer-reviewed journal articles or juried conference papers to make copies of such items (either the final accepted version, or the version of record) available to the public free of charge within one year of publication [2]. The NSF Public Access Repository (NSF-PAR), provides mechanisms that enable NSF-supported investigators to meet this requirement, and provides search mechanisms to enable the public to find and use these materials [3].

[2] https://www.nsf.gov/news/special_reports/public_access/index.jsp
[3] https://www.nsf.gov/news/special_reports/public_access/about_repository.jsp

DATA SHARING PRACTICES

Prior to publication, project data and metadata will be shared with registered MagLab users listed as project collaborators. Requests from other interested parties will be directed to the PI. The PI initial use authority does not control sharing data with DCFF personnel for preservation and archival purposes.

The MagLab is exploring the Open Science Framework (OSF) to serve as a project management and data sharing platform between DCFF personnel and external users. Users and staff are encouraged to use the OSF for data transfer, access, and storage, but it is not required. The PI can send/receive and share materials and data through their virtual delivery mechanisms of choice in consultation with DCFF staff.

DATA ACCESS POLICY

This policy applies only to data and metadata collected at the DCFF under the user program. Products of proprietary research, not funded by the NSF, are exempt from these data access requirements. To balance the need to make data openly available to the community with user expectations that they will be able to publish results of their scientific efforts without fear of preemption, data and metadata associated with a user project are expected to be made publicly available when an associated manuscript is published, or within 3 years of the date the project was last assigned magnet time, unless a related publication or patent application is actively under review. Repository entry, DOI, and other relevant accession information should be included in publications.

As exceptions to the requirements in the previous section, some data are not required to be made publicly available. These are data that will not form the basis of publishable research findings nor are associated with a user project. These include data from experiments known to be faulty in some regard, e.g. through mishap or due to a flawed experimental design, calibration runs for which results are not needed to interpret legitimate project data, and data generated to verify successful operation of the instrument or demonstrate capability.

PUBLIC DATA REPOSITORIES

FAIR guidelines [1] stipulate that data and associated metadata should be submitted to a discipline-specific, community-recognized, public repository. The project PI is expected to utilize an appropriate repository. Recommended repositories are listed in the table, below.

Data-type, Field, or Funding Agency	Repository	Link to homepage
General Materials Research and Condensed Matter Physics	Open Science Framework	http://osf.io/

If no suitable, community-recognized resource is available, data and associated metadata should be submitted to a recognized generalist repository. The journal, *Scientific Data* (Springer Nature), recommends several generalist repositories [4]. Among them, the DCFF recommends the Open Science Framework (OSF) [5], a free and open platform developed by the Center for Open Science for research project management and as a reliable data repository.

The OSF supports the ability to embargo data and metadata in accordance with the policies outlined above. While embargoed, all submitted materials or datasets are given their own unique, persistent URLs. DOIs can be generated when projects or selected components are made public. These may be cited and accessed by the public, and are indexed in Google Scholar. The OSF is a flexible alternative to some field-specific repositories to efficiently, and wholly disseminate all data and metadata related to complex, large-scale projects spanning multiple disciplines.

Any materials deposited and accessed from public data repositories should include the "Policies for Re-use, Re-distribution, and Production of Derivatives" section, below. Data that is submitted to repositories is made available per the terms, conditions, and licenses adopted by the repository.

[4] https://www.nature.com/sdata/policies/repositories#general
[5] https://osf.io/

POLICIES FOR RE-USE, RE-DISTRIBUTION, AND PRODUCTION OF DERIVATIVES

Authors of any publications or presentations that utilize DCFF data, results, software, or other resources are encouraged to cite relevant literature, include relevant DOIs, or otherwise acknowledge the researchers who generated the samples, data, results, software, or other materials.

In addition, all published manuscripts, datasets, and presentations must acknowledge the DCFF and facility support (including NSF grant number) as outlined below:

"A portion of this work was performed at the DC Field Facility at the National High Magnetic Field Laboratory, which is supported by the National Science Foundation Division of Materials Research through DMR-2128556, and the State of Florida."

For data collected from 2012-2017, the appropriate grant number is DMR-1157490. For data collected from 2018-2022, the grant number is DMR-1644779. For data collected from 2023-2027, the grant number is DMR-2128556. Please include all grant numbers corresponding to the periods during which data were collected.

ARCHIVING OF DATA

The DCFF provides data archival and backup services for all user facility data that is collected using DCFF data acquisition computers. Data are subjected to automated archival from each data acquisition computer to the DCFF's OpsServer, which is subjected to regular backups to a custom RAID array. All data that has ever been archived/backed up to DCFF servers has been retained and this will continue indefinitely.

DCFF users may duplicate raw data for transport in a number of ways either to physical media (portable external storage or internal storage of user-owned computers) or file transfer to remote computers. Remote users are provided data through various virtual delivery mechanisms chosen by the PI, private repository submission (to be made public at the appropriate time), or through physical media delivered via physical delivery. PIs are encouraged to utilize any institutional data archival services available to them, in addition to MagLab resources.

DATA MANAGEMENT MAP
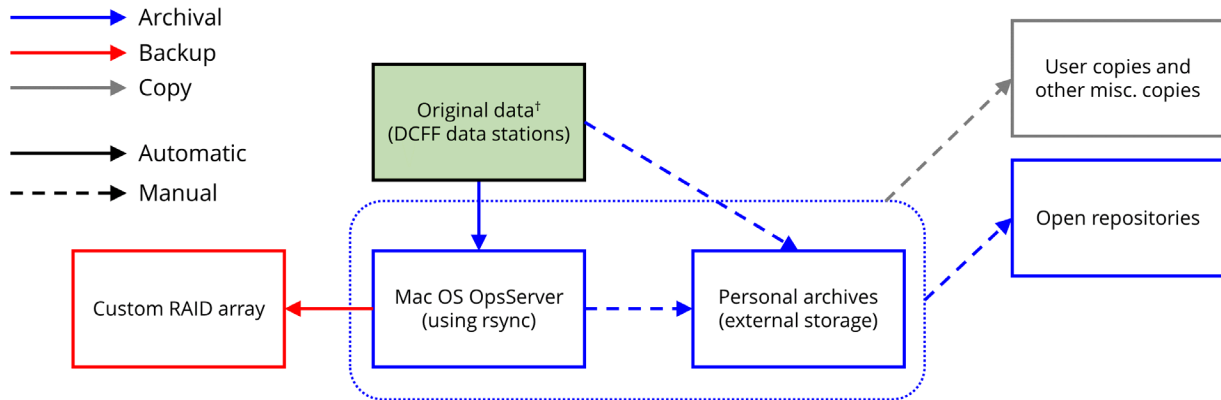
† : User data is not always collected on DCFF data stations. This map only applies to data for which a copy exists at the MagLab.