

MAGLAB PULSED FIELD FACILITY USER DATA MANAGEMENT PLAN

Contact: Laurel Elaine Winter Stritzinger (lwinter@lanl.gov)

Facility: Pulsed Field

Facility Location: Los Alamos National Laboratory

Last modified: 2021-07-12

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 (Findable, Accessible, Interoperable, and Reusable) 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 Pulsed Field Facility (PFF) at Los Alamos National Laboratory (LANL) be digitally accessible upon publication, or within 3 years. This data management plan (DMP) details resources available to PFF 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 are numerical time-dependent information on physical processes. The data is accompanied by metadata, which describes various sample and experimental parameters. Users are responsible for capturing and organizing descriptions of samples, protocols for their preparation, and relevant quantitative and qualitative information about the samples. Users are also responsible for capturing and organizing information as it pertains to the PFF experimental set-up and related measurement details.

DATA FORMAT

Raw data are primarily recorded on the facility's custom data acquisition systems. The high data sampling rate, methods/equipment, and complexity of PFF experiments necessitate use of a variety of industry-standard data formats, including National Instruments Technical Data Management Streaming (NI TDMS) and Hierarchical Data Formats (HDF5). The data can be imported into many products, including Excel, OpenOffice, Origin, Igor, NumPy, and MATLAB as well as converted to more accessible formats such as ASCII or UTF-8 encoded plaintext files as needed. Additionally, a custom software suite "High Magnetic Field Science Toolset" is continuously developed and supported by the PFF. The open-source software is designed to acquire, analyze, annotate, convert, and interact with PFF experimental data [2].

Information provided by users including sample description, preparation protocols, and experimental details can be recorded in various electronic file formats (e.g. *.xlsx, *.docx, *.pptx, *.html, *.txt, or *.pdf) and stored with the associated data set and other products of research. However, metadata describing various PFF experimental attributes are often recorded in users' laboratory notebooks. To implement FAIR data practices, users can elect to have experiment data, metadata, and other materials automatically uploaded to an associated Open Science Framework (OSF) [3] private project with software developed by PFF staff [4].

[2] DOI: 10.5281/zenodo.5076133, <https://github.com/ffb-LANL/High-Magnetic-Field-Science-Toolset>

[3] <http://osf.io>



[4] DOI: 10.5281/zenodo.5076633, <https://github.com/ffb-LANL/OSF-Tools>

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. 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 PFF staff 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, in addition to the PFF data access policies outlined below.

Research activities detailed in PFF user proposals and approved for magnet time are expected to result in presentations, publications, or other vehicles for dissemination of data and results. Whenever possible, published manuscripts should include digital object identifiers (DOIs) and other appropriate persistent identifiers to indicate where relevant data and metadata can be accessed. The NSF Division of Materials Research encourages users to include metadata (e.g., sample description and preparation methods, PFF experimental protocols, software and data processing workflows etc.) with published data to provide adequate information about the data to enable reproduction [5].

The NSF Public Access Policy 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 [6]. 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 [7].

[5] <https://www.nsf.gov/bfa/dias/policy/dmpdocs/dmr.pdf>

[6] https://www.nsf.gov/news/special_reports/public_access/index.jsp

[7] 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 PFF staff to gauge instrument performance, meet reporting requirements for the facility, or for preservation and archival purposes.

PFF users that collect data on-site can receive copies on physical media. 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 physical media delivered via parcel post. LANL and the MagLab are exploring the OSF to serve as a project management, data sharing and access platform [3]. PFF users are encouraged to use the OSF to organize and share data, but it is not required.

DATA ACCESS POLICIES

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 and must be reported to the MagLab [8] at the time of publication or conclusion of the data embargo period.

Some data are not expected to be made publicly available. These are data that will not form the basis of publishable research findings such as data from experiments known to be faulty in some regard, and data generated to verify successful operation of the instrument. Users may consult PFF staff regarding the type of data collected and its suitability for public consumption.



[8] <https://reporting.magnet.fsu.edu/>

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 responsible for selecting an appropriate repository. Recommended repositories are listed in the table, below.

Data-type, Field, or Funding Agency	Repository	Link to homepage
Generalist repository	Open Science Framework	Osf.io
Generalist repository	figshare	figshare.com
Generalist repository	Harvard Dataverse	dataverse.harvard.edu
Generalist repository	Zenodo	zenodo.org
Generalist repository	Dryad	datadryad.org

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 [9]. Among them, the MagLab recommends the OSF (Center for Open Science), a free and open platform for research project management, and a reliable data repository [3].

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 in 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.

[9] <https://www.nature.com/sdata/policies/repositories#general>

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

Authors of any publications or presentations that utilize PFF 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 MagLab PFF and facility support as outlined below:

“A portion of this work was performed at the Pulsed Field Facility at the National High Magnetic Field Laboratory, which is supported by the U.S. Department of Energy and the National Science Foundation Division of Materials Research through DMR-1644779, and the State of Florida.”

The current grant number, DMR-1644779, applies to data and products of research generated in the five-year period from 2018 to 2022. For data collected from 2012-2017, the appropriate grant number is DMR-1157490. Both grant numbers should be included if they apply.

ARCHIVING OF DATA

Data collected and stored on PFF data acquisition systems are archived to LANL institutional servers. Archived user data are retained at the PFF indefinitely. This retention policy is reviewed annually and may be revised at the request



of our user community or in response to the continually evolving capabilities and costs of data storage. Users are encouraged to utilize any institutional data archival services available to them in addition to PFF resources.

Data stored on LANL Institutional servers is currently backed up two different ways at separate physical locations. This ensures availability of the data in case of a disaster where one physical location might be inaccessible due to events such as fire or flooding.

In addition to back-up LANL provides Snapshots. A Snapshot is a read-only copy of the file system taken at specified intervals. Snapshots are scheduled to occur automatically on a recurring basis. Scheduled intervals are weekly, nightly, and hourly.

Files in a snapshot of the file system have the same read access privileges as in the active file system. A user with permission to read a file in the active file system can read it in a Snapshot. Users without read permission still cannot read the file. Write access privileges do not apply because files within a Snapshot cannot be changed; Snapshots are read-only. Files must be updated on the file's image in the current, active file system.

DATA MANAGEMENT MAP

