MAGLAB EMR FACILITY USER DATA MANAGEMENT PLAN

Contact: Stephen Hill (shill@magnet.fsu.edu)

Facility: Electron Magnetic Resonance (EMR)

Facility Location: 1800 E. Paul Dirac Dr., Tallahassee, FL 32310

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 Electron Magnetic Resonance (EMR) User Facility be digitally accessible upon publication, or within 3 years. This data management plan (DMP) details resources available to EMR facility 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

The EMR facility's data contains mainly electronic records of measurements taken during an experiment. Raw data are a single, or a collection of EMR spectra. The raw data are collected on local computers controlling the homebuilt and commercial spectrometers located at the EMR facility. The products generated by research vary depending on sample and application type. For samples prepared externally, users are responsible for capturing and organizing descriptions of samples, protocols for their preparations, and relevant quantitative and qualitative information about the samples. EMR facility personnel will assist users with the capture, storage, and organization of the data.

DATA FORMAT

EMR data formats are defined by the equipment used. For homebuilt spectrometers, the raw data are recorded and stored in txt or dat files in columnar formats. For the commercial Bruker spectrometer, the raw data are saved as a binary format (DSC and DTA files) specified by the vendor. The Bruker spectrometer also provides an option to store the raw data in plaintext files. All the raw data can be readily processed and analyzed using software provided by the EPR community, such as EasySpin [2], or commercial vendors. Homemade analysis software is also available and freely accessible through the NHMFL website, such as Spin [3]. Metadata can be recorded with the raw data files at the option of the researchers. Management of the meta-data associated with standard data files is exclusively the purview of the principal investigator.

[2] https://www.easyspin.org
[3] https://nationalmaglab.org/user-facilities/emr/software/

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 EMR user proposals and approved for magnet time are expected to result in presentations, publications, or other vehicles for dissemination of 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. Users are encouraged to work in collaboration with EMR facility personnel to verify data or results before use in forums such as publications, meeting presentations, grant or patent applications.

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 EMR facility staff listed as collaborators to the extent necessary to facilitate compliance. For research involving human subjects, the PI must include the relevant institutional review board (IRB) number(s) in the submitted user proposal, or magnet time cannot be granted. 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 EMR facility data access policies, below.

The NSF Public Access Policy [4] 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 [4]. 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 [5].

[4] https://www.nsf.gov/news/special_reports/public_access/index.jsp
[5] 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 EMR facility personnel to gauge instrument performance, meet reporting requirements for the facility, or 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 EMR facility 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 on physical media in person, through parcel post, or through their virtual delivery mechanisms of choice in consultation with EMR facility staff.

DATA ACCESS POLICY

This policy applies only to data and metadata collected at the EMR User Facility 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, data from preliminary experiments that are not intended to be delivered to EMR facility users, standards/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. Users should consult EMR facility staff regarding the type of data collected and its suitability for public consumption.

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 should select an appropriate repository [6].

If no suitable, community-recognized resource is available, data and associated metadata should be submitted to a recognized generalist repository, such as those listed in the table, below. The journal, *Scientific Data* (Springer Nature), recommends several generalist repositories [7]. Among them, the MagLab recommends the Open Science Framework (Center for Open Science), a free and open platform for research project management and a reliable data repository.

Data-type, Field, or Funding Agency	Repository	Link to homepage
Generalist	Dryad Digital Repository	https://datadryad.org/stash
Generalist	figshare	https://figshare.com/
Generalist	Open Science Framework	http://osf.io/

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.

[1] http://www.go-fair.org/fair-principles
[6] https://www.nature.com/sdata/policies/repositories
[7] 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 EMR facility 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 EMR Facility, and facility support (including NSF grant number) as outlined below:

"A portion of this work was performed at the Electron Magnetic Resonance User Facility at the National High Magnetic Field Laboratory, which is supported by the National Science Foundation Division of Materials Research and Division of Chemistry through DMR-2128556, and the State of Florida."
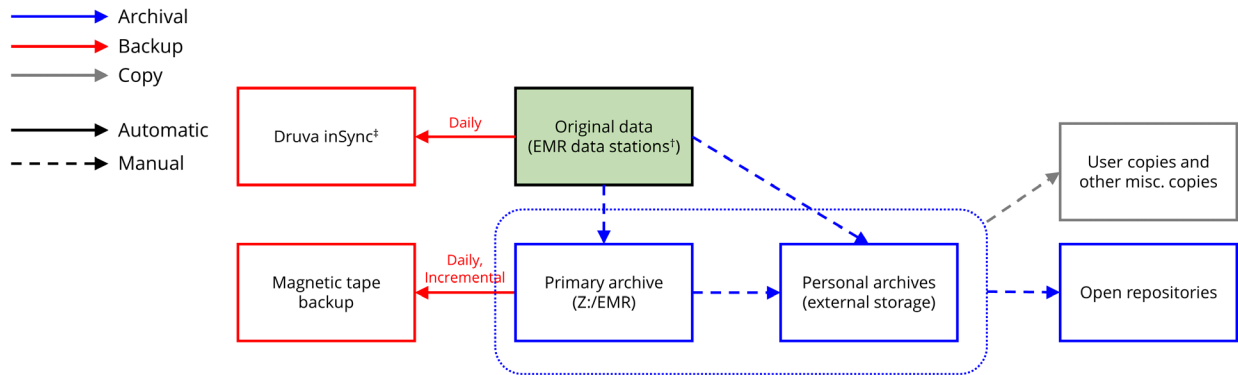
For data collected from 2012-2017, the appropriate grant number is DMR-1157490. For data collected from 2018-2022, the grant number is DMR-1644779. For data collected from 2023-2027, the grant number is DMR-2128556. Please include all grant numbers corresponding to the periods during which data were collected.

ARCHIVING OF DATA

The EMR facility provides data archival and backup services for all user facility data. All raw data collection is performed on EMR facility hardware and instruments, and all raw data are archived on MagLab backup server (the Z drive). Data archived on the Z drive are subjected to weekly backups to ensure data durability. The backup server is managed by the MagLab Computer Support Group. Additionally, data collected using Windows 10 computers are subjected to backups using Druva inSync on a daily basis. All data that has ever been archived has been retained and this will continue indefinitely. PIs are encouraged to utilize any institutional data archival services available to them in addition to MagLab resources.

DATA MANAGEMENT MAP

Legend:
- → Archival (blue)
- → Backup (red)
- → Copy (gray)
- → Automatic (solid black)
- --→ Manual (dashed)

Diagram:

Druva inSync‡ ← Daily ← Original data (EMR data stations†)

Original data → Primary archive (Z:/EMR)

Original data --→ Personal archives (external storage)

Magnetic tape backup ← Daily, Incremental ← Primary archive (Z:/EMR)

Primary archive (Z:/EMR) --→ Personal archives (external storage)

Personal archives (external storage) --→ Open repositories

Personal archives / Open repositories --→ User copies and other misc. copies

† : Most data stations run Windows 10, two run Linux.
‡: Only Windows 10 data stations are backed up with Druva inSync.