

MAGLAB AMRIS FACILITY USER DATA MANAGEMENT PLAN

Contact: Joanna Long (jrlong@mbi.ufl.edu)

Facility Location: University of Florida, Gainesville, FL

Funder: National Science Foundation (NSF)

Last modified: 2021-06-11

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 (Findability, Accessibility, Interoperability, and Reuse) 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 Advance Magnetic Resonance Imaging and Spectroscopy (AMRIS) User Facility be digitally accessible upon publication, or within 3 years. This data management plan (DMP) details resources available to AMRIS 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

Raw data are single or a collection of one-, two-, or three-dimensional (1D, 2D, 3D) NMR spectra or MRI images, depending on the research conducted. Metadata and 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 preparation, and relevant quantitative and qualitative information about the samples (e.g. experimental conditions, sample quantity and concentration, solvents/buffers, etc.). AMRIS facility personnel will assist users with the capture, storage, and organization of all relevant sample processes, AMRIS experimental conditions, MR sequences and relevant parameters, and data and metadata resulting from work taking place at the MagLab. This includes details of any further sample processing prior to the performance of experiments, as well as any files resulting from data analysis via AMRIS facility data processing workflows.

DATA FORMAT

Raw data is electronically recorded and stored in a variety of formats depending on the instrumentation and software used. Data, metadata, and the products of research that are stored in unusual or not generally accessible formats will be converted to more accessible formats, or necessary software will be made publicly available to support users, and independent investigators with data processing/visualization, as described below.

The facility uses a variety of software and hardware vendors which use their own custom file formats. Bruker NMR TopSpin software (Bruker Corporation) uses a nested file structure, which stores both experimental parameters, raw data and processed data together. A 'title' file is also stored with the data sets, allowing for additional information regarding sample composition, etc., to be stored. Being the predominate vendor in NMR, the Bruker file format is widely used and the TopSpin software is available for free for academic use from Bruker. TopSpin can also convert data to the open standard JCAMP-DX format [2] that is used for uploading to online repositories.

The Varian and Agilent NMR/MRI instruments use VNMR and VNMRJ software, which combine raw and processed data with experimental parameters. This software allows for data export to JCAMP-DX format [2]. VNMRJ is now free and open source [3]. For preclinical MRI, Bruker hardware uses its Paravision software (Bruker Corporation) that uses nested folder data sets that include experimental details, raw data, and processed data. Varian/Agilent VNMRJ software



produces its own proprietary data format, which contains similar data. For data sharing purposes, both imaging software packages allow data export to the open standard medical imaging format DICOM files [4,5].

The majority of detailed data analysis is handled by the users themselves and is highly specific to the data collected and research undertaken. Both Varian and Bruker software packages allow for direct analysis of the files, and a wide range of proprietary and open source software are available that can read these data formats [6].

[2] <http://www.jcamp-dx.org/>

[3] <https://openvnmrj.org/>

[4] <https://www.dicomstandard.org/>

[5] Bidgood Jr, W. Dean, et al. "Understanding and using DICOM, the data interchange standard for biomedical imaging." *Journal of the American Medical Informatics Association* 4.3 (1997): 199-212.

[6] <https://amris.mbi.ufl.edu/user-resources/software/>

DATA ACCESS AND SHARING

RESPONSIBILITIES OF THE PRINCIPAL INVESTIGATOR

For each project in the user's portal, the designated Principal Investigator (PI) is the steward of the research data, will select the vehicle(s) for publication or presentation products of research, and will have ultimate authority in their initial use.

Research activities detailed in AMRIS 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. Journal articles should include digital object identifiers (DOIs) and other appropriate persistent identifiers to indicate where relevant data and metadata can be accessed. Users are welcomed and encouraged to work in collaboration with AMRIS facility personnel to verify data or results before use in publications, meeting presentations and 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, including deidentification requirements for human subject-associated data. The PI is encouraged to disclose such requirements to AMRIS facility staff listed as collaborators 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 AMRIS facility data management policies.

For research involving animals or human subjects, the PI's host institutional animal care and use committee (IACUC) or institutional review board (IRB) approval must be secured before a project is granted magnet time. Additionally, the PI must comply with all public access requirements that are laid out by other funding agencies sponsoring the research, in addition to AMRIS Facility data management policies.

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 considered on a case-by-case basis, and are subject to approval of the PI. The PI initial use authority does not control sharing data with AMRIS facility staff to gauge instrument performance, or meet reporting requirements for the facility.

Users can obtain data in person using an external storage device such as a USB drive or via remote access to the machines while on the University of Florida (UF) campus or when connected to the UF network via VPN. 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 parcel post.

The MagLab is exploring the Open Science Framework (OSF) [7] to serve as a project management and data sharing platform between AMRIS 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.



[7] <https://osf.io/>

DATA ACCESS POLICY

This policy applies only to data and metadata collected at the AMRIS User Facility under the MagLab user program. 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 are to be made available to the public, as outlined below:

All data and metadata associated with a user project will become publicly available when an associated manuscript is published, or within 3 years of the date the project was last assigned magnet time. Extension of the embargo period can be requested by the PI on a case-by-case basis. Exceptions will generally be granted if a publication or patent application is under review. Repository, entry, DOI, and other relevant accession information must be reported to the MagLab at the time of publication or conclusion of the data embargo period [8].

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 AMRIS 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 AMRIS Facility staff regarding the type of data collected and its suitability for public consumption.

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

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, or choose from a recommended repository listed in the table, below.

Data-type, Field, or Funding Agency	Repository	Link to homepage
NMR metabolic spectra	Metabolomics Workbench	https://www.metabolomicsworkbench.org/
Biologically relevant NMR spectra and structures	Biological Magnetic Resonance Data Bank	https://bmrbl.io/
NMR spectra and structures	NMRShiftDB	https://nmrshiftdb.nmr.uni-koeln.de/
Biomolecular structures	Protein Data Bank	https://www.rcsb.org/
Neuroimaging data	NeuroImaging Tools and Resources Collaboratory	https://www.nitrc.org/projects/nitrc/
Traumatic brain injury	Federal Interagency Traumatic Brain Injury Research (FITBIR) informatics system	https://fitbir.nih.gov/
Alzheimer's disease	Alzheimer's Disease Neuroimaging Initiative	http://adni.loni.usc.edu/
Generalist	Zenodo	https://zenodo.org/
Generalist	Mendeley Data	https://data.mendeley.com/
Generalist	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 [9]. 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 [7].



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.

[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 AMRIS 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 AMRIS Facility, and facility support (including NSF grant number) as outlined below:

"A portion of this work was performed in the McKnight Brain Institute at the National High Magnetic Field Laboratory's Advanced Magnetic Resonance Imaging and Spectroscopy (AMRIS) Facility, which is supported by National Science Foundation Cooperative Agreement No. DMR-1644779 and the State of Florida."

For data collected from 2018-2022, the grant number is DMR-1644779. Data collected from 2012-2017, the grant number is DMR-1157490. Please include both grant numbers if your data was collected during both time periods.

Additionally, some AMRIS instrumentation was funded by NIH grants and require the following acknowledgements:

Publications using data collected on the 11T instrument should include

"This work was supported in part by an NIH award, S10 RR025671, for MRI/S instrumentation"

Publications using data collected on the 600 MHz Avance III should include:

"This work was supported in part by an NIH award, S10 RR031637, for magnetic resonance instrumentation."

Publications using data collected with the 1.5-mm HTS Cold Probe should include:

"A portion of this work was carried out using a 1.5 mm High Temperature Superconducting Cryogenic Probe developed with support from NIH award R01 EB009772." and cite in materials/methods: Ramaswamy, Vijaykumar, et al. "Development of a ¹³C-optimized 1.5-mm high temperature superconducting NMR probe" *Journal of Magnetic Resonance* 235 (2013): 58-65.

Publications using data collected on the 600 MHz WB Neo should include:

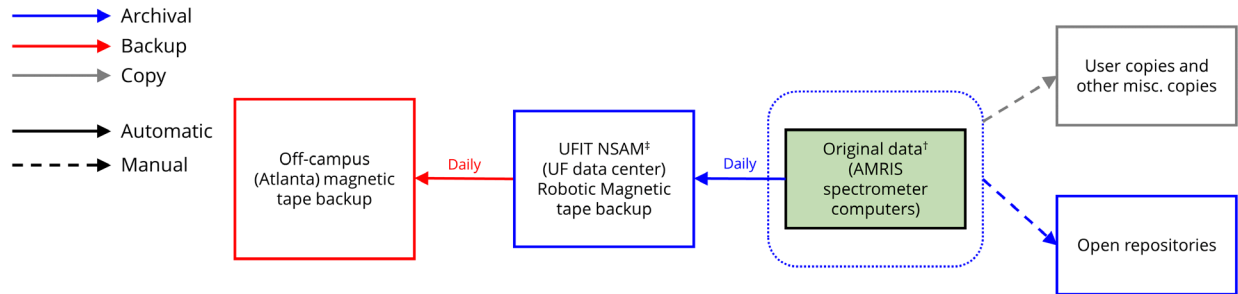
"This work was supported in part by an NIH award, S10 OD028753, for magnetic resonance instrumentation."

ARCHIVING OF DATA

All raw data collection is performed on AMRIS Facility computers and hardware. Data is stored on the instruments for a minimum of 6 weeks post collection. Data is automatically archived daily to University of Florida's enterprise-scale network robotic magnetic tape storage facility based on IBM's Tivoli Storage Manager. This facility is secure, climate-controlled, 24-hour x 365-day staffed data-center maintained by University of Florida's Infrastructure and Communication Technology (ICT) group. Off-site backups are stored in Atlanta, GA. Two most recent generations of data backups are retained.



DATA MANAGEMENT MAP



† : Data stations run Centos 7. Acquired data is stored in /opt/data.

‡ : Contents are automatically backed up daily using a service based on IBM's Tivoli Storage Manager (<https://hosting.it.ufl.edu/services/backup-archive/>) and AMRIS staff have 24-hour access.