Data sharing for pre-clinical projects

Background

The Focused Ultrasound Foundation (FUSF) will require that the source data underlying the results obtained through projects funded by FUSF be made publicly available at the completion of the study after publication of the results. At minima, recipients of FUSF funding will continue to be required to make public their final report.

Our objective is to create an ecosystem of FUS data comprising data obtained in all projects funded by FUSF, and to provide researchers the opportunity to access the source data in order to be able to:

- replicate the study,
- re-analyze and/or reprocess of the data,
- reuse the data in some circumstances,
- perform cross-comparison,
- support the development of their own projects or protocols,
- aggregate data from multiple sites/projects.

These data will include

- The source data underlying the results presented in the final report,
- Reporting on the methodology, including experimental systems and their characterization,
- Details of any software used to process the results.

These guidelines provide information about which data need to be included, where the data should be stored, and how data should be presented. A data sharing clause will be included in funding contracts when appropriate.

Open data policy

The FUS Foundation advocates for open data practices and has created and will maintain a public data repository. The creation and maintenance of projects in the collection will come at no cost for the PIs.

Public data should be published under a Creative Commons license Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0), that allows copying and redistribution of the data in their original form (no remix, transformation or built-upon) for non-commercial purposes under requirement that appropriate credit be given.
Data that cannot be made available to a general public, but can be made available to FUS research community, will be shared under the same license with access granted by the PI upon request.

Data repository

For data storage and dissemination, FUSF has created a data repository collection, hosted by the Center for Open Science (OSF).

Upon signature of a research funding agreement, the PI will create a project within the OSF system and grant an administrator access to FUSF. FUSF will verify that the project has been created, and will assist with troubleshooting or technical issues.

Each project will then be managed by the PI. Projects can be sub-divided into private and public parts. Private parts of the projects can be used for project participants to share data and/or project management, and to store data only accessible upon request.

Data publication

Data publication will be a two-step process:

1. At the start of a study, the project's PI will publicly release a summary of the study, including context, objectives, methodology, expected results, and anticipated deadline.
2. At completion of the study or after publication, the project's data will be made publicly available. At that time, FUSF will create a registration of the project to generate a time-stamped immutable version of the project and all contents/files, and create a DOI on this registration.

Data format and Detailed description

1. Data Description: The type of data to be stored will be contingent on the study, but should include description of experimental systems and treatment parameters, data (such as PCD recording, temperature measurements…), images such as pre and post-treatment images, US, MRI, CT; thermal MRI or histology, biological and immunological date, as appropriate.
2. Formats: Data should be stored in a format accessible to the community at large, by privileging open source formats. We will ask the PIs to endorse the FAIR Data Principles (See Annex 1) as a framework to promote the broadest reuse of research data.
3. Metadata and Documentation: There is currently no intention to create uniform metadata formats or standards, as the collection will not be a database. Precise documentation will be required from the PI to describe the data as they relate to specific experiments, animals, or sets of treatment parameters. They will also include details of any software and methods used to process the results, in order for others to be able to replicate the analysis.
4. Platform for data storage: It is possible to store data directly on the OSF server in the collection. A list of supported data formats is available here, mainly including images, video, spreadsheets, plain text, and word processing. For other data types, several storage providers are supported by OSF, and can be linked to the projects. FUSF will also provide access to a storage option for the data, linked to the projects, and will require that a copy of all final data for public access be stored there. This will ensure the data will remain available, even if changes are made on the project's OSF page.
5. Access and sharing: Projects in the collection will be created and administered by the PI. PI will grant administrator access to FUSF. Data will be published in project components created by the PI.
6. Budget: The creation and maintenance of projects in the collection will come at no cost for the PIs. The maintenance of the collection will be provided by FUSF.
7. Privacy, Intellectual Property: Public data will be published under a Creative Commons license Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0), that allows copying and redistribution of the data in their original form (no remix, transformation or built-upon) for non-commercial purposes under requirement that appropriate credit be given. Data that cannot be made available to a

general public will be shared under the same license, but will be accessible only after the PI grants authorization.

8. Archiving, Preservation, Long-term Access: The intention is to create open-ended access to the data after their publication.
9. Biological data: Biological data obtained during FUSF funded projects should also be made public in a community-endorsed, public repository. See Annex 2 for suggestions.

Annex 1 – General Guidance for Data formatting

We will ask the PIs to endorse the FAIR Data Principles as a framework to promote the broadest reuse of research data.

The basics of the FAIR Data Principles are that the data be

- **Findable**: The objective of the FUSF Data Repository is to make FUS related data easily discoverable. Our collection is hosted in an open repository, where each project is assigned its own unique, persistent URL. It is also possible to create DOIs for your public research (although at the time of publication of these guidelines DOIs point to the current version of the project or registration, as OSF does not support DOI versioning).
- **Accessible**: as defined by the presence of a user license
- **Interoperable**: to ensure interoperability, comparisons and combinations with data from different sources by both humans and machines, data should be stored in a non-proprietary open file format and described using a standard vocabulary (where available).
- **Reusable**: ensure that the data are findable, accessible, and interoperable. Metadata and documentation should allow someone not familiar with the project to understand what the project and the data are about.

The FAIR guiding principles as described in this publication:

Box 2 | The FAIR Guiding Principles

To be Findable:
F1. (meta)data are assigned a globally unique and persistent identifier
F2. data are described with rich metadata (defined by R1 below)
F3. metadata clearly and explicitly include the identifier of the data it describes
F4. (meta)data are registered or indexed in a searchable resource

To be Accessible:
A1. (meta)data are retrievable by their identifier using a standardized communications protocol
A1.1 the protocol is open, free, and universally implementable
A1.2 the protocol allows for an authentication and authorization procedure, where necessary
A2. metadata are accessible, even when the data are no longer available

To be Interoperable:
I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
I2. (meta)data use vocabularies that follow FAIR principles
I3. (meta)data include qualified references to other (meta)data

To be Reusable:
R1. meta(data) are richly described with a plurality of accurate and relevant attributes
R1.1. (meta)data are released with a clear and accessible data usage license
R1.2. (meta)data are associated with detailed provenance
R1.3. (meta)data meet domain-relevant community standards

Annex 2 - Examples of public repositories for specific datasets

Following recommendations from Nature research journals[1], below is a list of community-endorsed, public repositories where the following types of data should be published in addition to publication within the FUSF data repository.

Data Type	**Suitable repositories**
Protein sequences	Uniprot
DNA and RNA sequences	Genbank
	DNA DataBank of Japan (DDBJ)
	EMBL Nucleotide Sequence Database (ENA)
DNA and RNA sequencing data	NCBI Trace Archive
	NCBI Sequence Read Archive (SRA)
Genetic polymorphisms	dbSNP
	dbVar
	European Variation Archive (EVA)
Linked genotype and phenotype data	dbGAP
	The European Genome-phenome Archive (EGA)
Macromolecular structure	Worldwide Protein Data Bank (wwPDB)
	Biological Magnetic Resonance Data Bank (BMRB)
	Electron Microscopy Data Bank (EMDB)
Gene expression data (must be MIAME compliant)	Gene Expression Omnibus (GEO)
	ArrayExpress
Crystallographic data for small molecules	Cambridge Structural Database
Proteomics data	PRIDE
Cytometry and Immunology	FlowRepository
	ImmPort

[1] https://www.nature.com/nature-research/editorial-policies/reporting-standards#availability-of-materials

Annex 3 – General Resources and Information on data sharing

FairSharing.org A curated, informative and educational resource on data and metadata *standards*, inter-related to *databases* and data *policies*.

Gates Open Research Data Guidelines A platform for rapid author-led publication and open peer review of research funded by the Bill & Melinda Gates Foundation

Nature recommended Data Repositories A list of repositories that can be useful to publish data (such as biological sciences data) in addition to the FUSF repository.