Federal Interagency for Traumatic Brain Injury Research (FITBIR) -

Data Preservation and Access Practices

The Federal Interagency Traumatic Brain Injury Research (FITBIR) Informatics System is a central repository and resource for sharing data that was developed by the Department of Defense (DOD) and the National Institutes of Health (NIH) to promote collaboration, accelerate research, and advance knowledge on the characterization, prevention, diagnosis and treatment of traumatic brain injury (TBI). FITBIR provides a common platform and standardized format for data collection, retrieval, and archiving, while allowing for flexibility in data entry and analysis.

This document describes the FITBIR Data Preservation and Access Practices for submitting, accessing, and archival of data in the FITBIR Data Repository.

I. Submission Requirements

Investigators submitting study research data must comply with the specific procedures described below.

1. Data Redaction

Investigators submitting FITBIR data are expected to submit a Data Submission Request, providing assurance that all data are submitted in accord with applicable laws and regulations, and that the identities of research participants will not be disclosed to the FITBIR Informatics System. Data and/or images submitted to the FITBIR Informatics System will be de-identified such that the identities of data subjects cannot be readily ascertained or otherwise associated with the data by the FITBIR staff or secondary data users. In addition, de-identified data will be coded using a unique code known as a Global Unique Identifier (GUID). Use of the GUID minimizes risks to study participants because it keeps one individual's information separate from that of another person without using names, addresses, or other identifying information. The unique code also allows FITBIR to link together all submitted information on a single participant, giving researchers access to information that may have been collected elsewhere. The GUID is a computer-generated alphanumeric code [example: 1A462BS] that is unique to each research participant (i.e., each person's information in FITBIR—or each subject's record—has a different GUID). FITBIR will assist investigators in how to create the GUID, which is an essential requirement for uploading data to FITBIR. Data submitters agree to notify FITBIR as soon as possible if, upon review of FITBIR data, the Submitter discovers identifying information in that data.

2. Material Transfer Agreement

The NINDS Human Biospecimen and Data Repository is called BioSEND and all FITBIR-related biospecimens can be queried through the FITBIR portal site under the FITBIR Biorepository Biosample Catalog form. The purpose of the BioSEND repository is to receive, process, store, and distribute biospecimens resources that can be shared by the neuroscience research community, and to encourage the research efforts of established scientists, junior investigators, and scientists with novel approaches. BioSEND will work closely with FITBIR to assure standardization of biological sample collection, protection of subject privacy and rights, and transparency regarding availability of samples for Traumatic Brain Injury disease biomarker discovery research.

For access to FITBIR-related samples from the NINDS Biorepository, Investigators must first apply for a FITBIR Data Access account to query for sample availability. Once an account is obtained, biosample information can be found in the FITBIR Biorepository Biosample Catalog form within the Query module. All applications are submitted through an online webform. Within the webform, applicants are required to upload the following files (in pdf, gif, jpg, jpeg, or png file formats):

- Biosketch
- Research Strategy (4-page limit)
- Table Summary of Samples of Interest (This should include information on sample availability.)

The TBI Biospecimen Review Access Committee (TBI BRAC) committee will assess the applications based on experimental rationale, feasibility/reproducibility of the assays, expertise of the investigator, availability of institutional resources to support the study, and the statistical analysis of the number of samples required for the hypothesis testing. Investigators will be notified by email of the outcome of the review (i.e. approve, approve upon revisions, or deny) within a week after the review meeting. Summary statements are released within 2-3 weeks after review. For applications that were 'approved upon revision', investigators will be notified about concerns that would need to be addressed before the application could move forward. Please note that the TBI BRAC does not provide funding. If sample access is approved by the TBI BRAC, the two possible outcomes are:

1. If the study has funding, the samples are distributed to the investigator following a virtual meeting to introduce the sample distribution process ("onboarding") and after fulfillment by the investigator of the additional requirements specified by the selected repository (i.e. MTA, Data Use Agreement, fees, data analysis and sharing plan). This post-TBI BRAC process for biosample distribution is outlined in these documents (FITBIR).
2. If the investigator has yet to obtain funding for the study, the TBI BRAC will issue a letter to the applicant documenting provisional access to the samples requested. This letter can be used to support an application for funding opportunities from the NIH or other organizations. Conditional approvals will be valid for a period of up to 12 months.

Please read BioSEND's Material Transfer Agreement (MTA) for details regarding transferring of biospecimen resources.

3. Data Quality and Preservation

The FITBIR system was designed with long-term data preservation in the forefront. For a more in-depth description of how FITBIR supports long-term data preservation, please see the publication, "Development of an informatics system for accelerating biomedical research," which elaborates on the system design and architecture of the underlying BRICS platform. For long-term preservation of data, the Open Archival Information System (OAIS) model highlights the importance of six functions—ingest, access, data management, archival storage, administration, and preservation planning. Specifically, the model provides a framework for preserving information for a designated community (group of potential consumers and multiple stakeholders). The model is unique because it is content and technology agnostic. FITBIR has applied the OAIS model for long term preservation of biomedical data collected for traumatic brain injury (TBI) research, implementing the concept of Submission, Archival and Dissemination Information Packages (SIP, AIP, DIP) for processing data for the TBI research community. Imaging data SIPs are produced by the Image Submission tool.

FITBIR's required usage of Common Data Elements (CDEs) and a well annotated and versioned data dictionary supports the development of Archival Information Packages (AIPs), which are preserved in the FITBIR Data Repository. Data Submitters are expected to use the CORE TBI Common Data Elements (CDEs) at a minimum.

FITBIR mandates collection, validation, and submission of data using Common Data Elements (CDEs). These standards are transparently documented on FITBIR's public site. The required use of CDEs, availability of documentation combined with detailed version history, and metadata associated with study-related data make it possible for long term preservation and use of data.

FITBIR Operations team works with researchers to map their study variables to specific CDEs. The FITBIR Operations team consults with researchers to ensure the formats of the CDEs collected are compatible with the FITBIR Informatics System. In addition to CDE formatted data, FITBIR will accept raw data from imaging, biomarker, or physiologic studies.

Data are submitted in Comma Separated Values (CSV) format, a universal standard ensuring ease of mapping and interoperability. CSV's simplicity, platform independence, compatibility, and transparency make obsolescence unlikely and make it a viable option for long-term data preservation. Utilization of the CSV format significantly reduces the need for future format migrations or emulations. There are no plans to change the CSV file format or metadata schemas.

Submitted data must conform to the CSV form structure templates and CDEs as defined in the Data Dictionary. Before data can be uploaded, the data must first pass validation procedures and checks which ensure the data schema and permissible values are consistent with the Data Dictionary and that the GUID assigned to participant data is valid i.e., has been created in FITBIR.

All research data submitted to FITBIR must be accompanied by proper documentation to ensure meaningful use of the data. Study documentation should be comprehensive and sufficiently clear to enable investigators to understand the study and data. Study profiles created in the Data Repository support built in metadata field annotations (e.g., Aims; Study Type; Site information) as well as uploading of supporting documentation (e.g., study protocols; manual of operations; variables measured; case report forms; other relevant documents)

All data submitted into FITBIR is initially in a Private status, meaning only users assigned permissions to the study can access them. Once a study is completed or ready to share some portion of their data, the study team will notify FITBIR Operations. FITBIR Operations performs a study closeout analysis which entails running a python script that checks for the following:

1. Demographic data have been submitted for all participants
2. Duplicate entries
3. Partial data (i.e., a subset of participants are missing values for a given data point)
4. Extra-validation scoring algorithms errors
5. Correspondence between GUIDs and reported subject IDs (if submitted) throughout all submissions.

Overall, these checks ensure that data for each participant are consistent, properly reported and as complete as possible. These detailed findings are reported to the study team and will be corrected by the study team (if necessary) before sharing the data. If study data involves imaging submissions, a separate study closeout analysis is run by FITBIR Imaging Operations which runs similar consistency checks and, additionally, checks for any hidden PII in metadata header tags.

Once the study team approves the data quality and accuracy, they will request the status of their datasets be changed to Shared. Once in a Shared status, these data are accessible by FITBIR users with Query Tool privileges. FITBIR Operations will also notify program (DoD and NINDS) of the study's fulfillment of their data submission agreement.

The DOIs generated by FITBIR through an NIH contract with the Interagency Data ID Service (IAD), which is operated by the U.S. Department of Energy Office of Scientific and Technical Information (OSTI). The IAD service acts as a bridge to DataCite, which is one of the major registries of DOIs. The DOIs are assigned to individual research studies and are findable within the established repositories, available also from open sites with core metadata supported via Data Tag Suite (DATS v2.2).

II. Archiving Procedures

The following section describes the processes and procedures for archiving resources once submitted to FITBIR:

1. Storage

All stored data are hosted at the CIT's secure facility, with strict adherence to security measures defined by the Federal Information Security Management Act (FISMA). Furthermore, the system undergoes an annual review by the NIH Security team. FITBIR utilizes backup storage in an AWS cloud environment, which provides on-site and off-site storage layers, thereby further enhancing data redundancy and resilience. Regular contingency and restore tests, performed yearly and monthly respectively, ensure the backup systems' functionality and data integrity.

2. Infrastructure Operations

The operation of FITBIR is 24x7 with redundancies and backups done on a nightly schedule. The FITBIR database is backed up in accordance with NINDS and CIT Security Policies and Guidelines and provides a restore capability to revert to in the event of a database corruption or system failure.

CIT policies and procedures are available on an internal content management system. We have our roadmap items in Jira and track any future plans through this mechanism. All Infrastructure planning and roadmap items are documented internally within the Jira system. The FITBIR development team under the direction of the government tracks the progress for all roadmap items and plans for enhancements and improvements to the FITBIR infrastructure. The plan is reviewed by the system owners (the government funders NIH and DoD, and Matthew McAuliffe, co-Director of FITBIR) and prioritized based on requirements.

3. Data Integrity

FITBIR supports data integrity in the following ways:

1. *Curation*:
    - Ph.D. level of curation for Metadata and data is provided by the FITBIR and BRICS lead curator and the supporting operations team.
2. *Validation*:
    - The FITBIR Data Dictionary and Data Repository modules enable validation of study data and metadata. Any modifications to data structures and CDEs are made only

by the authorized FITBIR CDE curator. Any changes to datasets in the Data Repository are recorded and viewable to users through a version history log and accompanying metadata. The FITBIR system is 21 CFR Part 11 certified.

3. *Quality Checks*:

    - Each primary research study in the FITBIR Data Repository undergoes a quality control check (referred to as a "study closeout analysis") before the data are shared with FITBIR Data Access users. FITBIR Operations runs a series of python scripts that analyze the data for consistency and potential issues including: (1) missing subject demographic information (2) duplicate records (3) partial data (4) incorrect instrument scoring or data entry and (5) subject identification inconsistencies across records. A separate quality check is performed for imaging data (if applicable) that analyzes images for: (1) duplicates (2) unnecessary images (3) subject identification inconsistencies across records and (4) potential personal identifiable information (PII). These findings are compiled in a report sent to the study team and -- with further discussion and clarification -- any remaining issues are resolved. This may take the form of resubmissions that are given a new dataset ID and whose relationship to other datasets can be specified with a link. Once the data has passed these quality checks and the finalized datasets have been verified by the study team, the data are shared within FITBIR.

4. *Archival/Modifications*:

    - If a study team notifies FITBIR operations that their shared datasets need to be revised, these datasets are archived, corrected datasets are resubmitted (and given a new dataset ID), and users who had previously accessed the now archived datasets are notified of the change. Flags, descriptions, and dataset links can be utilized to further annotate datasets and specify contextual information e.g., reason for archiving.

4. Security

The FITBIR system resides in a badged, monitored, and audited secure data center within CIT on the restricted NIH Campus. The backup infrastructure is currently supported in an AWS cloud environment. The hardware, software, networking and applications are all maintained by the FITBIR system administration team, in accordance with NIH policies and procedures. The FITBIR admin team maintains all servers and storage according to strict and well-defined laws and regulations (e.g. FISMA moderate). As a FISMA Moderate system, FITBIR adheres to the NIST 800-53 security standards and guidelines. All documentation is held in either NSAT system at NIH or Confluence, the content management system. Both NSAT and Confluence are internal only.

The FITBIR design incorporates several security and integrity controls to ensure the system and its associated systems are continually protected. This is done through a multi-tiered approach to ensuring data integrity is achieved through only authorized user functions and assignments.
The first design consideration is user authorization and permissions. These users will be unable to perform any transactions outside of their assigned areas. System administrators will grant proper roles and operating boundaries for each of their users. The next design consideration is to establish control points. Firewalls will be placed to partition the functions each instance is able to perform. In addition to NIH firewalls and intrusion detection, the FITBIR servers all have firewalls implemented to only allow required ports. The purpose is to reinforce work areas, permissions, and access to prevent any duplication, unintentional changes, or malicious changes of data.

The system design also incorporates the important audit trail capability which will allow administrators to track the history of all users in order to provide history, error identification, and accountability for system users. The NIH Incident Response Team (IRT) team regularly scans the FITBIR system for security and privacy vulnerabilities. Any issues are addressed within established timeframes as set by NIH security policy.

III. Access Requirements

Investigators requesting access to FITBIR must comply with the specific procedures described below. FITBIR account requests are reviewed and assessed by the Data Access and Quality Committee (DAC) which is composed of stakeholders from NINDS and DoD. Biospecimen applications are reviewed and assessed by the TBI BRAC which is composed of NINDS TBI researchers.

1. Data Access Requirements

Data Access requires submission of the following forms:

a. Data Access Request
b. Biographical Sketch

2. Data Submission Requirements

Data Submission requires submission of the following forms:

a. Data Submission Request

3. Biosample Access Requirements

Data Access users that wish to access FITBIR-related biosamples from the NINDS Biorepository (BioSEND) must fill out an online webform application and attach the following forms:

a. Biosketch
b. Research Strategy
c. Table Summary of Samples of Interest

IV. Inquiries

For inquiries on the FITBIR Preservation and Access Practices, please send an email to FITBIR-OPS@mail.nih.gov.