Appendix A. PBTF Data Sharing & Repository Guide

Purpose

PBTF is committed to accelerating pediatric brain tumor research through the responsible sharing, integration, and reuse of high-quality data. This appendix provides practical guidance to help investigators prepare competitive applications, develop data sharing plans, and generate research-ready datasets and unique material resources that maximize scientific value beyond the original study.

This guide complements the PBTF Research Program Guide and should be used by investigators, bioinformaticians, data managers, and research administrators when planning PBTF-funded projects.

Key Definitions

Research-ready data: Data that are complete, interpretable, quality-controlled, and sufficiently documented so they can be understood and reused by another investigator without contacting the originating laboratory.

Research Resources: All unique research outputs including model organisms, cell lines, plasmids, protocols, software, and data**.**

Foundational data: Assay-generated data and the minimum processed data, quality control information, and metadata required to enable broad secondary use.

Primary processed data: Standardized analysis-ready datasets derived from raw data that support downstream research and reproducibility.

Open-access data: Data that may be accessed without controlled-access authorization, consistent with participant privacy and consent.

Controlled-access data: Data requiring formal authorization before access, typically including individual-level genomic data and sensitive linked clinical information.

General Research Use (GRU): Participant consent that permits broad secondary research use of data without unnecessary restrictions, allowing data to be integrated and reused across studies.

Data Use Limitations (DUL): Restrictions associated with participant consent that define how data may be shared and reused.

Institutional Certification: Documentation completed through an institution confirming that participant consent, privacy protections, and data governance support the proposed data sharing in accordance with NIH Genomic Data Sharing Policy.

Planning Your Data Strategy

Every PBTF application must include a Data Use Strategy, a Data Contribution & Deposit Plan, and a Technology (AI/ML) Integration Approach. Together, these components describe how the proposed research will leverage existing pediatric brain tumor research resources, how resulting data and other research products will be prepared for sharing and reuse, and how AI/ML methods will be applied and validated.

The level of detail expected differs between the Letter of Intent and Full Proposal stages.

LETTER OF INTENT:

At the Letter of Intent stage, applicants should provide a concise overview of their proposed data strategy using the PBTF Data LOI form. The purpose is to help PBTF evaluate project feasibility, identify potential data-sharing or resource considerations, and determine whether additional consultation may be beneficial before invitation to submit a Full Proposal.

Applicants should briefly address the following:

Research Resource & Data Use Strategy

- What existing datasets, biospecimens, models, repositories, or other research resources will support the proposed project?
- Have these resources been identified and, where applicable, has access been confirmed or initiated?
- If new data will be generated, why are existing resources alone insufficient to answer the proposed research question?

Data Contribution & Deposit Plan

- What new data or other reusable research products will the project generate?
- How do you anticipate sharing these resources (e.g., repository, network, or other access mechanism), if known?
- Are there any anticipated consent, data-use, repository, or feasibility considerations that could affect data sharing or integration?

Technology (AI/ML) Integration Approach

- What AI/ML methods or computational approaches do you anticipate using, if known at this stage?
- How will this approach leverage or scale across existing federated, multi-modal datasets?
- Are there any anticipated validation strategies or data-access considerations relevant to your technology approach?

FULL PROPOSAL:

Applicants invited to submit a Full Proposal should provide a comprehensive data strategy that addresses the following areas.

Research Resource & Data Use Strategy

Describe how existing research resources will support the proposed work.
Include, as applicable:

- datasets, repositories, biospecimens, cohorts, imaging collections, preclinical models, digital pathology or other research resources to be used;
- why these resources are appropriate for the proposed research;
- whether access has been secured or is pending; and
- the rationale for generating new data, where applicable.

Applicants are encouraged to leverage existing pediatric brain tumor resources whenever appropriate to maximize scientific value and reduce unnecessary duplication of data generation.

Data Contribution & Deposit Plan

Describe the research products expected to result from the project and how they will be prepared for sharing and reuse.
Include, as applicable:

- anticipated datasets and other reusable research products; (e.g., cell lines, reagents, and software/code)
- repository strategy;
- Material Transfer Agreement (MTA) status for any materials already in hand;
- documentation, metadata, and quality-control activities;
- anticipated timeline for data sharing;
- applicable consent, Institutional Certification, or Data Use Limitations; and
- any requested budget for data stewardship activities, including harmonization, documentation, quality assurance, and repository submission.

Technology (AI/ML) Integration Approach

Describe the AI/ML methods or computational approaches to be applied and how they will be validated.
Include, as applicable:

- the specific AI/ML methods, models, or computational approaches to be used;
- how the approach will leverage or scale across existing federated, multi-modal datasets;
- the validation strategy, including performance metrics and comparison approaches appropriate to the method; and
- any data access, computational resources, or governance considerations relevant to the proposed approach.

Applicants should ensure that the proposed data strategy is appropriate to the project's scope and align with the research-ready data expectations described in this guide. Projects that generate patient-derived data

should also ensure that participant consent and institutional approvals support the proposed data-sharing strategy as shared in the Program Guide document.

What Makes Data Research Ready?

A research-ready dataset contains the information necessary for another qualified investigator to understand, reproduce, and reuse the data without contacting the originating laboratory. At a minimum, research-ready datasets should include:

Core Data

- Appropriate raw and primary processed data
- Data generated by the funded project

Metadata

- Participant and sample identifiers
- Assay information
- Clinical annotations
- File relationships

Documentation

- Study README
- Data dictionary
- Provenance documentation
- Version history
- Protocol sharing link (e.g., protocols.io) for core experimental protocols

Quality Control

- Quality control summaries
- Processing documentation
- Validation metrics

Reproducibility

- Workflow or pipeline description
- Software and reference versions
- Processing parameters
- Versioned data releases
- Publicly shared protocols for key experimental methods (e.g., via protocols.io)

Rule of thumb: If another qualified investigator could understand and reuse the dataset without contacting the originating laboratory, it is likely ready for integration.

Repository Strategy

Repository Consideration	Guidance
Controlled-access genomic data	Typically deposited into NIH-approved repositories such as dbGaP
Harmonized research assets	May also be integrated within the CBTN ecosystem, as appropriate
Multiple repositories	Some projects may require deposition into more than one repository
Repository planning	Applicants should describe their proposed repository strategy within the Data Contribution & Deposit Plan
Data from CBTN-sourced biospecimens or resources	Any resulting data files are expected to be returned to CBTN, which manages further dissemination, including deposition to dbGaP for controlled-access genomic data.
Data from PNOC-sourced biospecimens or resources	PNOC is treated as its own repository, on par with dbGaP. Resulting data files are expected to be deposited with PNOC, which manages further dissemination; controlled-access genomic data are additionally deposited to dbGaP as applicable.
Data from PCDC-sourced biospecimens or resources	PCDC is treated as its own repository, on par with dbGaP. Resulting data files are expected to be deposited with PCDC, which manages further dissemination; controlled-access genomic data are additionally deposited to dbGaP as applicable.
Plasmids/DNA/Viruses	Deposited into Addgene.
Antibodies	Deposited into Addgene, the Antibody Registry, or the ABCD Database.
Model Organisms	Deposited into Jackson Laboratories (JAX) or the Mutant Mouse Resource & Research Center (MMRRC).
Cell Lines/Organoids	Deposited into ATCC, CBTN.

Data Sharing Expectations

Projects generating patient-derived data should:

- Obtain participant consent supporting broad secondary research use whenever possible.
- Ensure consent aligns with General Research Use (GRU) principles.
- Understand applicable Data Use Limitations (DUL).
- Obtain Institutional Certification when required.
- Develop a plan for timely data sharing and integration.

Budgeting for Data Sharing

Preparing research-ready datasets requires dedicated expertise and resources. Applicants should budget for data-sharing activities when necessary to prepare, harmonize, document, quality-control, or deposit research-ready datasets. Appropriate budget items may include:

- Data preparation
- Data harmonization
- Metadata creation
- Quality control
- Documentation
- Repository submission
- Data governance
- Bioinformatics support

Where investigators do not have the resources or infrastructure needed to prepare and deposit research-ready data, PBTF may consult CBTN, where appropriate, to help define the scope of work and associated budget required to support these activities.

Investigator Support & Technical Resources for CBTN

PBTF encourages investigators to take advantage of the technical resources, training opportunities, and consultation services available through CBTN when planning and conducting PBTF-funded research. The examples below represent current resources available to support investigators; additional support may be available based on project needs.

Getting Started with CBTN

Investigators new to CBTN can participate in:

- **Getting Started** sessions (first Tuesday of each month, 10:00 a.m. ET)
- **Advanced Q&A** sessions (third Thursday of each month, 2:00 p.m. ET) for active projects
- Monthly drop-in office hours

Additional self-guided resources are available through the **Kids First Data Resource Portal** and **CBTN Quick Start Videos**.

Application Development

To support investigators preparing PBTF applications, PBTF and CBTN may also offer pre-application webinars covering:

- the four project data scenarios;
- data sharing and deposit requirements;
- strategies for accessing existing CBTN resources; and
- developing a strong Data Contribution & Deposit Plan.

Applicants are encouraged to take advantage of these resources before preparing an LOI or full proposal.

Material Resource Sharing

In addition to data, PBTF-funded projects that generate physical research materials, including cell lines, plasmids, antibodies, organoids, and model organisms are expected to deposit these materials into established public repositories to maximize their reuse across the field.

Materials must be made available no later than the date of publication of findings that rely upon them, or within 12 months of the award's end date, whichever comes first.

PBTF discourages a "sharing upon request" model for these materials. Requiring other investigators to contact the originating lab creates unnecessary burden and reduces discoverability; depositing directly into a public repository ensures materials remain accessible regardless of future changes in lab personnel or contact information.

Computational Resources

Eligible investigators may also take advantage of the **NIH Cloud Credit Program** to help offset cloud computing costs associated with data analysis.

Questions

Questions regarding CBTN data resources, investigator support, or technical assistance may be directed to **research@cbtn.org** and **research@curethekids.org**