

Tips for Successful Data Contribution

This guide provides a step-by-step roadmap for contributing and releasing data through Sage Bionetworks portals. Our goal is to help Data Contributors (Principal Investigators (PIs) and their teams) prepare data that is formatted for reuse to be shared securely and compliantly.

This document is organized into four lifecycle phases to proactively address governance, privacy, and technical standards. We look forward to partnering with you early in the planning phase to ensure a smooth, efficient data contribution and release process.

The Four Phases of the Data Sharing Lifecycle:

- ☐ [Phase 1: Project Initiation and Data Landscape](#)
- ☐ [Phase 2: Governance and Contract Execution](#)
- ☐ [Phase 3: Data Upload, Curation, and Quality Assurance](#)
- ☐ [Phase 4: Discovery, Access & Reuse](#)

[Glossary of Terms](#)

Responsibility Key	Role
Data Contributor	Principal Investigator, Data Liaison, and Institutional Staff
Sage Bionetworks	Privacy & Compliance, Data Management, Program and Scientific Teams, and Portal Teams

Version Date: 07/17/2026

Disclosure: This document was drafted with the assistance from AI tools.

Phase 1: Project Initiation & Data Landscape

This phase occurs early in the grant or project lifecycle, ensuring the intent to share data aligns with funders' expectations, institutional policies, and Sage Bionetworks governance requirements.

Key Step	Responsibility	Description of Activities
1. Confirm Informed Consent, Ethical and Contractual Compliance	Data Contributor	Verify the authority to share data (e.g., Institutional Review Board [IRB] determination of approval, waivers, or non-human subject research). Determine the scope of informed consent for data sharing and potential reuse. Identify any contractual obligation or institutional stipulation impacting data sharing.
2. Review Sage Bionetworks Terms of Service (ToS) and Data Coordinating Center (DCC) Specific Data Use Terms	Data Contributor	PIs review Sage Bionetworks ToS and, as applicable, the Data Coordination Center (DCC)-specific data use terms via a DUC or similar documentation to understand Sage Bionetwork's baseline terms and conditions for data sharing. This information should be included within a relevant NIH Data Management and Sharing Plan (DMSP).
3. Initiate Contact with SageTeam	Data Contributor	Inform the relevant Sage Bionetworks Program or Data Coordinating Center (DCC) of the intent to share data. Review program-specific workflows.
4. Create a Synapse Account and Complete Certification Quiz	Data Contributor	All members of the Data Contributor team must have their own synapse accounts . Uploading and curating (applying metadata and annotation) require completion of the brief Synapse certification quiz.
5. Establish Data Ingress Requirements & Data Model Scope	Sage Team	Define the technical and curation requirements for data contribution, including: the data model schemas that apply to the contributor's data types, the Synapse folder structure required for metadata submission (including Curator compatibility where applicable), and the metadata templates the contributor will use.

Version Date: 07/17/2026

Disclosure: This document was drafted with the assistance from AI tools.

		Inform the Data Contributor of any gaps - such as data types that require new schema development - so that model work can begin before upload, rather than blocking it. Early scoping at this step is the single most effective way to reduce delays in Phase 3.
6. Familiarize yourself with your Synapse project	Data Contributor	In most cases, Data contributors are added to Synapse projects by the Sage team and can start exploring the features of the system, like the staging folder structure (e.g., will further sub-folders be needed for certain data types), data upload options (e.g., programmatically or via Synapse UI), and Synapse search functionality.
7. Complete the Data Landscape	Data Contributor	Complete a Data Landscape Survey or equivalent document detailing the expected data types, assay modalities, file formats, sample counts, release timeline, and any known data use limitations. Contributions that include this level of detail at the landscape stage consistently move through Phases 3 and 4 more quickly than those where data types are clarified late. This information is used by the Sage team to: confirm which existing data model schemas and metadata templates apply to the contribution; identify whether new schema or template development is needed before upload can begin; scope capacity and assign the appropriate named relationship lead; and determine Curator readiness for the contribution.
8. Identify Institutional Signatory/Liaison	Data Contributor	Provide agreement/contract contact, designate a primary Data Uploader for technical coordination, and a primary/enduring contact for future questions about the data.
9. Conduct Preliminary Privacy Review	Sage Team	Conduct a high-level data privacy risk assessment or Data Protection Impact Assessment (DPIA) based on the data source, type, and origin, and identify any data access/sharing constraints to determine potential access tiers (Open vs. Controlled Access).

Version Date: 07/17/2026

Disclosure: This document was drafted with the assistance from AI tools.

Phase 2: Governance & Contract Execution

This phase focuses on formalizing the legal and ethical framework for data sharing and reuse.

Key Step	Responsibility	Description of Activities
1. Execute Applicable Data Sharing Agreement (DSA)	Data Contributor / Sage Team	The PI's institution and Sage Bionetworks execute a Data Sharing Agreement (DSA) (sometimes referred to as Data Contribution Agreement [DCA], or Data Sharing and Use Agreement [DSUA], etc.), or append study-specific exhibits to an existing agreement. If data-sharing conditions have been unambiguously confirmed (for example, as a condition of funding), a stand-alone Exhibit will be executed. This agreement or Exhibit defines the parties' responsibilities and the expected data to be shared. The Sage team will provide an Agreement Template.
2. Set Up Data Access Conditions and Use Limitations	Data Contributor / Sage Team	Based on the DSA and Data Privacy Risk Assessment/DPIA, the Privacy & Compliance Team and PI will agree on the Access conditions and Use Limitations for each data type (e.g., define who can access the data and for what purpose).
3. Set up Data Contributor's project space on Synapse	Sage Team	Set up the designated Synapse location and folder structure with necessary access controls for data ingress

Phase 3: Data Upload, Curation & Quality Assurance

The focus here shifts to the technical preparation of the data, ensuring it is ready for public access and meets FAIR principles.

Key Step	Responsibility	Description of Activities
1. Prepare Data for Submission: Complete Data De-identification	Data Contributor	The Data Contributor is responsible for completing all necessary de-identification or pseudonymization steps in accordance with the DSA and Data Use Limitations (DULs) before data upload.
2. Data Upload	Data Contributor	The Data Contributor submits all raw data, processed files, data dictionaries, study protocols, and analysis code to Sage-approved locations (e.g., Synapse, protocols.io , Github). Following the Sage Team's instructions for folder structure, file organization, and access permissions.
3. Metadata Documentation Upload	Data Contributor	Provide standardized metadata (e.g., study, participant, biospecimen, assay, file-level) using Curator or approved templates. Accurate annotation is required to support data validation, quality review, and downstream portal discoverability and interoperability.
4. Validate and Curate Submissions and prepare for release	Sage Team	The Sage Team Data Managers review submissions for completeness, consistency, schema conformance, and compliance with applicable quality standards. This includes validating metadata against the portal's data model, resolving any issues with the Data Contributor, ensuring all necessary annotations are applied, and preparing the dataset for release.
5. Mint Digital Object Identifiers (DOIs)	Sage Team / Data Contributor	Confirm a permanent identifier (DOI) is minted for the dataset/project, which can be used to reference the data availability in associated publication(s).

Phase 4: Discovery, Access & Reuse

Version Date: 07/17/2026

Disclosure: This document was drafted with the assistance from AI tools.

This phase makes the released data discoverable and usable through Sage-supported portals and access workflows. It covers ensuring accurate data access controls, enabling public discovery, supporting responsible data use, and monitoring acknowledgments, publications, and other evidence of reuse.

Key Step	Responsibility	Description of Activities
1. Confirm Access Controls and Release Conditions	Sage Team	Confirm that the appropriate data access controls, data use limitations, and release conditions are applied. This includes verifying the data is set to be released according to its level of sensitivity (e.g., open, registered, controlled) and consistent with approved Data Use Limitations, required governance, privacy, and compliance conditions before data is released.
2. Prepare Materials For Data Discoverability and Reuse	Data Contributor	Work with the Sage Team to prepare the materials needed to support data discovery and reuse. This may include a Data Availability Statement, portal summary, dataset description, acknowledgment language, related publication links, and any contributor-approved messaging for launch or downstream reuse.
3. Enable Data Discovery and Access	Sage Team	Upon final approval, the Sage Team makes the dataset discoverable and accessible through the appropriate Synapse-backed portal or repository, including relevant study pages, FileViews, search interfaces, DOI-linked records, or other approved discovery surfaces.
4. Support Responsible Data Access	Sage Team or Designated Data Access Committee.	Serve as, or facilitate operations of, the designated Data Access Committee (DAC) responsible for granting researchers access to controlled-access data. The DAC ensures data access requests are ethically sound and adhere to established governance policies, legal requirements, and terms of agreements.

Version Date: 07/17/2026

Disclosure: This document was drafted with the assistance from AI tools.

5. Monitor Access Compliance and Renewals	Sage Team	Monitors data access activity, user attestations, renewal requirements, and revocation needs to ensure continued compliance with the approved Data Use Limitations, Data Use Certificate terms, governance policies, and applicable agreements
6.Track Data Reuse, Acknowledgement, and Impact	Data Contributor and Sage Team	The Data Contributor and Sage Team may monitor publications, citations, acknowledgments, and other evidence of downstream data reuse. Publications using the data should acknowledge Sage Bionetworks, the relevant portal or DCC, the original funding source, and the contributing investigators or consortium as required by the applicable Data Use Certificate, Data Sharing Agreement, or portal policy. Data Contributor reports issues to Sage Team as necessary.

Glossary:

Access Conditions

Rules that define who can access data and under what circumstances.

Access Controls

Technical and administrative restrictions applied to data to enforce approved access levels, such as open or controlled access.

Controlled Access

A data access level in which only approved users can view or use sensitive data under specific conditions.

Data Access Committee (DAC)

The group responsible for reviewing and approving requests for access to controlled-access data.

Data Contribution Agreement (DCA)

An alternate term for an agreement governing data contribution to Sage.

Data Contributor

The principal investigator and/or team responsible for preparing and sharing data with Sage Bionetworks.

Data Coordinating Center (DCC)

A coordinating body or program that manages data workflows, governance, and sharing requirements for a specific initiative.

Data Availability Statement

A statement in a publication or project record that explains where the data can be found and how it can be accessed.

Data Curation

The process of reviewing, organizing, standardizing, and improving data and metadata to make them suitable for sharing and reuse.

Data De-identification

The removal or modification of personal identifiers from data to reduce privacy risk.

Data Dictionary

Documentation that defines the structure, fields, values, and meaning of a dataset.

Data Ingress Requirements

The technical and curation requirements for submitting data, such as metadata expectations and file format standards.

Data Landscape

A summary of the data being contributed, including its type, volume, and expected release timeline.

Version Date: 07/17/2026

Disclosure: This document was drafted with the assistance from AI tools.

Data Landscape Survey

A form or document used to collect structured information about a dataset before contribution and release.

Data Liaison

A designated person who coordinates data submission and related communications between the contributor and Sage.

Data Management Sharing Plan

A document describing the research data collected, access policies and limitations, data standards and metadata, provenance, public archiving, and timeline for public accessibility.

Data Management Team

The Sage team responsible for managing uploaded datasets and making them available through the portal.

Data Model

A structured framework that organizes elements of data and standardizes how they relate to one another.

Data Privacy Risk Assessment

An evaluation of privacy risks associated with sharing a dataset.

Data Protection Impact Assessment (DPIA)

A formal assessment used to identify and reduce privacy and data protection risks.

Data Sharing Agreement (DSA)

A legal agreement between the contributor's institution and Sage that defines the terms for sharing and managing data.

Data Sharing and Use Agreement (DSUA)

Another name for an agreement covering both data sharing and downstream use conditions.

Data Use Certificate (DUC)

The terms and conditions governing access to and use of data in a specific Data Coordinating Center or program.

Data Use Limitations (DULs)

Restrictions on how shared data may be used, often based on consent, agreements, or privacy requirements.

Demographics

Descriptive participant information, such as age, sex, race, or ethnicity, often included as metadata.

Digital Object Identifier (DOI)

A permanent identifier assigned to a dataset or project so it can be persistently referenced and cited.

Version Date: 07/17/2026

Disclosure: This document was drafted with the assistance from AI tools.

Discoverability

The extent to which data can be found by users through metadata, search, and portal interfaces.

Ethical Compliance

Adherence to ethics requirements, including informed consent and research oversight.

FAIR Principles

Guidelines stating that data should be **Findable, Accessible, Interoperable, and Reusable**.

Findability

A FAIR principle referring to how easily data can be located using metadata and identifiers.

GitHub

A code hosting platform referenced in the document as a possible location for analysis code or related materials.

Governance

The legal, ethical, and policy framework that governs how data are shared, accessed, and reused.

Informed Consent

Permission given by research participants that may include conditions for data sharing and reuse.

Institutional Review Board (IRB)

A committee that reviews research involving humans to ensure ethical compliance.

Institutional Signatory

The authorized representative at an institution who can approve or sign agreements.

Interoperability

A FAIR principle describing how well data can work with other datasets, systems, or standards.

Metadata

Structured descriptive information about data, such as study design, file contents, demographics, and protocols.

Minting a DOI

The act of registering and assigning a DOI to a dataset or project.

National Institutes of Health (NIH)

The U.S. biomedical research funding agency referenced in connection with data sharing planning requirements.

Open Access / Open Data

Data that are made broadly available without restrictive access controls.

Version Date: 07/17/2026

Disclosure: This document was drafted with the assistance from AI tools.

Portal

A Synapse-backed platform through which data are organized, discovered, and accessed.

Principal Investigator (PI)

The person with primary responsibility for the research project and related data sharing decisions.

Privacy & Compliance Team

The Sage team responsible for privacy review, compliance monitoring, and implementation of data access conditions.

Protocols / Study Protocols

Documents describing how a study was conducted, often shared alongside data.

protocol.io

A platform referenced as a possible place to store or share protocols.

Pseudonymization

A privacy-protective process in which identifiers are replaced with codes, while some link to identity may still exist separately.

QC (Quality Control)

Checks performed to ensure data and metadata are complete, accurate, and compliant with standards.

Release Timeline

The planned schedule for making data publicly available.

Responsibility Key

A section that identifies which party is responsible for each task in the lifecycle.

Reuse

The ability for shared data to be used again in future research or analysis.

Sage Bionetworks Terms of Service (ToS)

The baseline legal terms and conditions governing use of Sage systems and services.

Schema

A defined structure or model used to validate and standardize data and metadata.

Secure Synapse Workspace

A protected Synapse location used for uploading and managing data before release.

Synapse

Sage Bionetworks' platform for storing, managing, governing, and sharing research data.

Technical Coordination

Operational work related to data upload, formatting, metadata, and system requirements.

Version Date: 07/17/2026

Disclosure: This document was drafted with the assistance from AI tools.