Synapse Commons Data Use Procedure

Last Update: Oct 2022

This document describes the standard operating procedure for using data within Synapse Commons ("**Commons Data Use Procedure**"). Synapse is a software platform designed to enable combined analyses of biological data with the goal of speeding improvements in healthcare through collaborative research. A central feature is a repository of datasets that contain genomic, molecular, and phenotypic data from human and non-human sources. These datasets are available in standardized formats for analysis.

Data hosting and sharing through Synapse must comply with all data protection or privacy laws and regulations applicable to processing personal data in the jurisdiction where the data was collected. **Data contributors must de-identify human data before making it available in the Synapse Commons to protect subject privacy.** Data may be de-identified according to the United States Health Insurance Portability and Accountability Act of 1996 ("HIPAA") standards for de-identification outlined in U.S. 45 CFR 164.514(b). Limited identifiers may be included with the Data when unambiguously approved by the Data subjects, an Institutional Review Board (IRB), or a similar ethics committee.

Collaborations within Synapse are facilitated by creating project workspaces where files such as analytical workflows, datasets, codes, figures, and documentation may be stored and shared. Synapse users may make their projects accessible to all registered Synapse users ("**Public**") or choose to make their projects accessible only to a subset of those users selected by the project owner or owner-designated project administrator(s) ("**Private**"). By default, all content residing within a Project inherits the Project sharing settings. More information on adjusting the sharing settings of specific folders, files, or tables, is available in the [Sharing Settings](#) documentation.

Although data contributors may broadly share data with the public, data use conditions may be required to protect the individuals from whom the data is derived.

Data use in the Synapse Commons is governed by a tiered system designed to ensure that each dataset is shared using the appropriate level of protection against the risk of harm to human participants. This system is enacted through data contribution and use procedures (Figure 1). Only users who demonstrate comprehension of the tiered system, risks of harm to human participants, ethical issues related to using data in the cloud, and key Synapse functionalities may become independent data contributors in Synapse. Users will demonstrate comprehension by completing an examination administered by Sage Bionetworks. Data contributors must identify the risks and appropriate protections required for sharing each dataset. Synapse informs users of these conditions and records users' agreement to adhere to all dataset-specific data use conditions before making the data available.

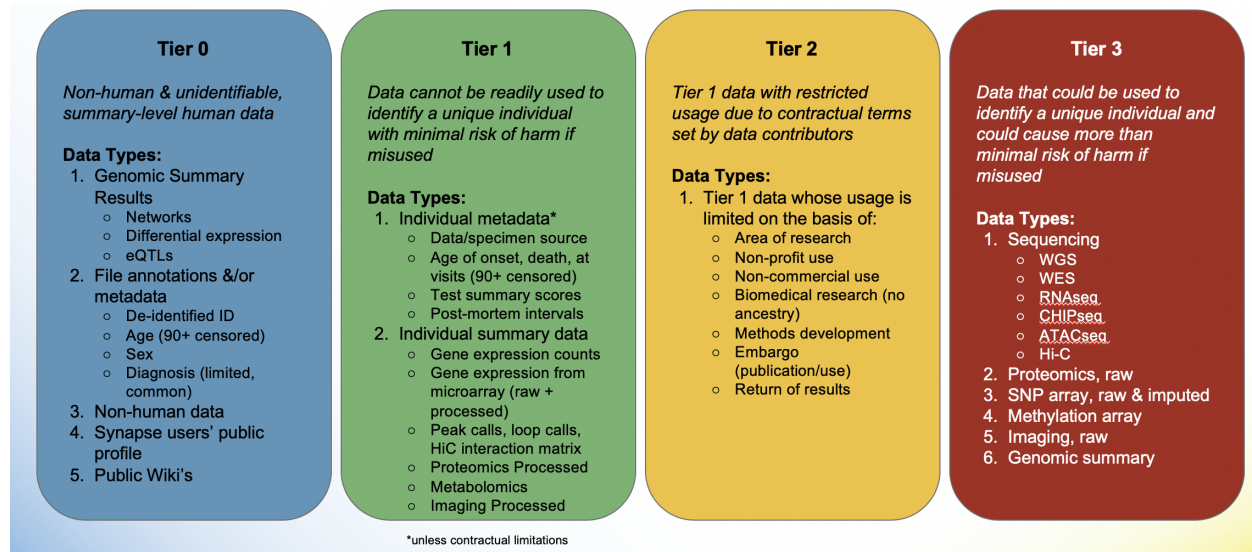**Figure 1. Process for protected data access.**
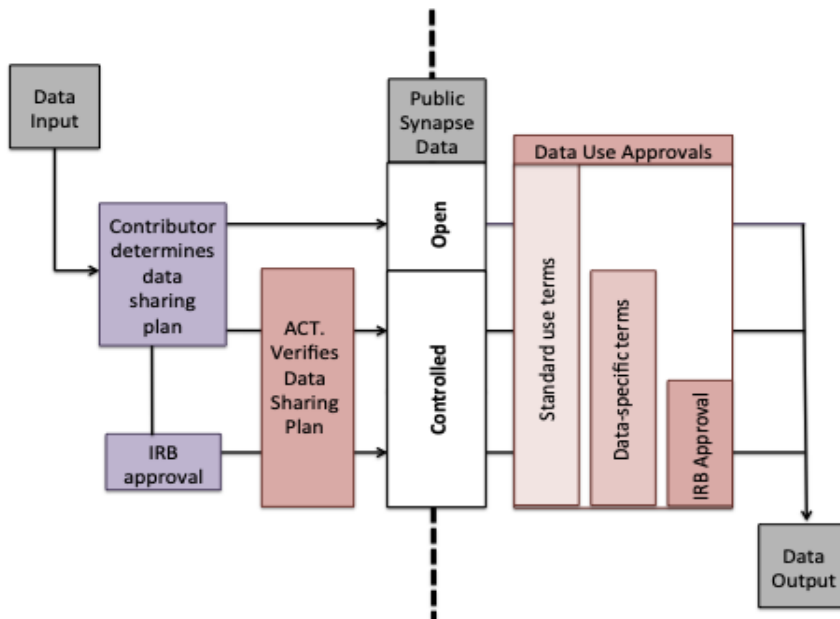
1. Data sharing plans

While Sage Bionetworks' policy promotes sharing data with as few restrictions as possible, the data access tiered system also accommodates sharing of data under conditions. For example, conditions may be required by informed consent, legal contract, or regulations disclosed by the data contributor. This process allows data sharing within the legal and ethical framework for the responsible use of human data.

- **Tier 0:** Data available to anyone on the web for any purpose. Access to Tier 0 Data is not tracked.
- **Tier 1:** Data available to all registered Synapse users without use limitations (Synapse "**Open Data**").
- **Tier 2:** Data available to all registered Synapse users whose use is limited by terms typically derived from the informed consent documents signed by participants contributing data for that study or from IRB-approved proposals (Synapse "**Controlled Data**").
- **Tier 3:** Individual-level **Controlled Data** requiring additional protections for human subjects, like sensitive information or data from "vulnerable populations" defined using OHRP guidelines or equivalent and its implementation in applicable local law. Sharing Tier 3 Data is considered to confer some risk of re-identification or harm to human participants. Contribution or use of these data requires either agreement to additional data-specific terms of use or independent review/monitoring by an ethics committee or Institutional Review Board ("**IRB**").

Here are examples of the types of data typically found in each tier. In cases where an individual has consented to share data unrestrictedly, data will be provided under the Open Data guidelines

2

regardless of what type of data it contains. In cases where data was collected in a jurisdiction under which local laws or regulations restrict data sharing, data sharing must comply with those terms and restrictions.

Tier 0	Tier 1	Tier 2	Tier 3
Non-human & unidentifiable, summary-level human data	*Data cannot be readily used to identify a unique individual with minimal risk of harm if misused*	*Tier 1 data with restricted usage due to contractual terms set by data contributors*	*Data that could be used to identify a unique individual and could cause more than minimal risk of harm if misused*
Data Types: 1. Genomic Summary Results ○ Networks ○ Differential expression ○ eQTLs 2. File annotations &/or metadata ○ De-identified ID ○ Age (90+ censored) ○ Sex ○ Diagnosis (limited, common) 3. Non-human data 4. Synapse users' public profile 5. Public Wiki's	**Data Types:** 1. Individual metadata* ○ Data/specimen source ○ Age of onset, death, at visits (90+ censored) ○ Test summary scores ○ Post-mortem intervals 2. Individual summary data ○ Gene expression counts ○ Gene expression from microarray (raw + processed) ○ Peak calls, loop calls, HiC interaction matrix ○ Proteomics Processed ○ Metabolomics ○ Imaging Processed *unless contractual limitations*	**Data Types:** 1. Tier 1 data whose usage is limited on the basis of: ○ Area of research ○ Non-profit use ○ Non-commercial use ○ Biomedical research (no ancestry) ○ Methods development ○ Embargo (publication/use) ○ Return of results	**Data Types:** 1. Sequencing ○ WGS ○ WES ○ RNAseq ○ CHIPseq ○ ATACseq ○ Hi-C 2. Proteomics, raw 3. SNP array, raw & imputed 4. Methylation array 5. Imaging, raw 6. Genomic summary

Note that access restrictions apply to the dataset as well as the analysis results derived from it unless the analyses have been summarized to a point where the original Data is no longer identifiable (e.g., figures of results, CNV summaries of SNP data, or gene expression summaries from RNAseq data).

2. Browsing/Searching for a dataset of interest.

Anyone can search or browse the list of Public datasets available in the Synapse Commons using key search terms or tags. The displayed information may include the name of the dataset and a brief description from the data contributor, the username of the data contributor, data use conditions, and the Synapse dataset identifier. Registered Synapse users can also search/ browse Private datasets for which they have been authorized viewing privileges.

3. Registered Users

All Synapse Commons users must register and create a Synapse user account to contribute or use data within Synapse. Users register by providing a name and valid e-mail address and agreeing to the Synapse Terms and Conditions of Use. As part of the registration process, users agree to each of the terms of the Synapse Pledge, recapitulating the themes of the Synapse Terms and Conditions of Use, Privacy Policy, and Code of Conduct. Registered users create a username and a unique password for their account, allowing them to operate within Synapse.

SYNAPSE Pledge

Please read the following statements, check the circles and sign your name.

I ___First Name, Last Name___ ,
reaffirm my commitment to all Synapse Governance policies for responsible research and
data handling (linked below), including:

I will adhere to the Synapse Community Standards of inclusion and respect.

A participant in the Synapse Community:
- Welcomes others
- Uses inclusive language
- Shares experiences and knowledge
- Respects other viewpoints and ideas
- Shows empathy and kindness when interacting with others

Please refer to our full **Synapse Community Standards**.

I agree

I will adhere to all conditions and use limitations, including privacy laws and regulations.

For more information about how Sage Bionetworks regulates data access and privacy, please refer to the **Synapse Commons Data Use Procedure**.

I agree

I will act ethically and responsibly.

You agree to the following Synapse Operating Ethics Principles, outlined in our **Terms and Conditions of Use**.

Consequences of misconduct can include loss of both data use privileges and future use of Synapse.

I agree

I will use appropriate physical, technical and administrative measures to keep data secure and protect participants' privacy.

In your use of Synapse, you agree to:
- Keep your login information secure and not share with others.
- Keep data safe from breach or misuse through appropriate security measures.
- Not attempt to re-identify or contact participants.
- Refrain from unauthorized data redistribution. Please refer to the **Synapse Commons Data Use Procedure**.

For more information, please refer to the full **Synapse Terms and Conditions of Use**.

I agree

SYNAPSE Pledge

I will support open access best practices for public facing data.

The Synapse open access platform promotes data accessibility and collaboration. We encourage you to contribute your research findings to open access journals when applicable.

You agree to support open access best practices when possible (e.g., sharing code, metadata, annotating files for discovery).

[ ] **I agree**

I will credit research participants and all data sources.

You agree to acknowledge data participants and to cite contributors and data sources using the language provided with the dataset.

It is our policy to terminate the access privileges of those who infringe the intellectual property rights of others. For more information, please refer to our full **Terms and Conditions of Use**.

[ ] **I agree**

I will not use data for marketing and/or advertising.

Data may not be used for marketing or advertising purposes. You agree not to upload or otherwise transmit any unsolicited or unauthorized advertising, promotional materials, junk mail, spam, or any other form of solicitation (commercial or otherwise).

Please review the **Synapse Privacy Policy** for Sage's commitment to privacy protections.

[ ] **I agree**

I will report any suspected data breach or misuse to privacyofficer@sagebase.org within 2 business days of initial incident knowledge.

Data misuse includes violating dataset access requirements, unauthorized use or redistribution, lack of data accreditation, and unethical data handling.

For more information about what classifies as a violation of our data governance policies, please refer to our **Data Use Procedure** and **Terms and Conditions of Use**.

[ ] **I agree**

Sign Your Name _____ **Date** _____

2020 SAGE BIONETWORKS

4. Certified Users

Once a user is registered, they can complete a certification process with training and an examination to demonstrate their understanding of the ethical, legal, and technical issues associated with using genomic and health information in the Synapse environment. Only users who complete this certification process will have access to the full Synapse functionality.

5. **Validated Users**

Users can apply to have their profile validated. A Synapse user with a Validated Profile is eligible to request access to Bridge and mHealth data (data collected in research studies conducted via self-guided mobile applications). To validate their profile, users will need to recommit to the Synapse Pledge and establish their identity by providing a combination of profile information in their Synapse profile and creating an ORCID profile (if they do not have one already). Acceptable identity attestation documentation includes a letter from a signing official on letterhead attesting to their identity, a notarized letter attesting to their identity, or a copy of their professional license. Custom identity attestation may be offered for users unable/unwilling to submit any of the documents mentioned above.

We distinguish users who are Anonymous, Registered, Certified, and Validated with distinct Synapse access privileges.

	Anonymous	Registered	Certified	Validated
View Public Wiki Pages	X	X	X	X
Browse Public Project Catalog	X	X	X	X
Browse Public File Catalog	X	X	X	X
Create a Project		X	X	X
Add Wiki Content		X	X	X
Download Files/Tables*		X	X	X
Upload Files/Tables			X	X
Create Folders			X	X
Add Provenance			X	X
Upload Docker Containers			X	X
Request Access to Specific Controlled Access Data**			X	X
Request access to mHealth data				X

*A user's ability to download files/tables depends on the local sharing settings of the content, the user

credentials, and on fulfilling the 'Conditions for Use', if any, of the file/table they wish to access. For example, some files/tables may require a Data Use Agreement before downloading.
** A user may need to be certified and/or validated for data access.

6. Procedure for data contribution
a. Submitting a new dataset

Individuals interested in independently submitting a dataset to Synapse ("data contributors") must be registered, have completed the Synapse examination, and be a 'Certified user'. Data contributors are responsible for adhering to all applicable data protection laws and regulations. They must de-identify data (to HIPAA standards in the US and/or other privacy laws and regulations applicable to processing personal information) before inclusion in Synapse Commons. Where required by applicable law, data contributors must verify that specific, informed, and freely given consents have been obtained for the use and sharing of the data for research purposes.

The dataset contribution process on Synapse is intended to be semi-automated. Data contributors are responsible for indicating the suitable dataset-specific conditions required for data use through Synapse. The flow chart in Figure 2 illustrates the questions to address to assign appropriate data use conditions. In cases where an individual (study participant) has consented to share data unrestrictedly, that data will be provided as Open Data regardless of what type of data it contains. For individual human data consented to be shared as "Open Data," data contributors are encouraged to upload the template consent form(s) associated with data collection alongside the data set. If questions remain about tier assignment, particularly for human subjects data, users are encouraged to seek advice from an ethics committee or independent IRB. For the contribution of Controlled Data, contributors must provide documentation verifying the appropriateness of the proposed data-sharing plan.

Sage Bionetworks may but is under no obligation to seek the opinion of an independent ethics team or accredited IRB to confirm the data contributor's selection of the data-sharing plan. In these cases, Sage Bionetworks can request, and the data contributor must provide, the (English-translated) templates of the informed consent document or the protocol governing the dataset in replacement of an IRB decision.

Figure 2. Flow chart of considerations for data contributions.
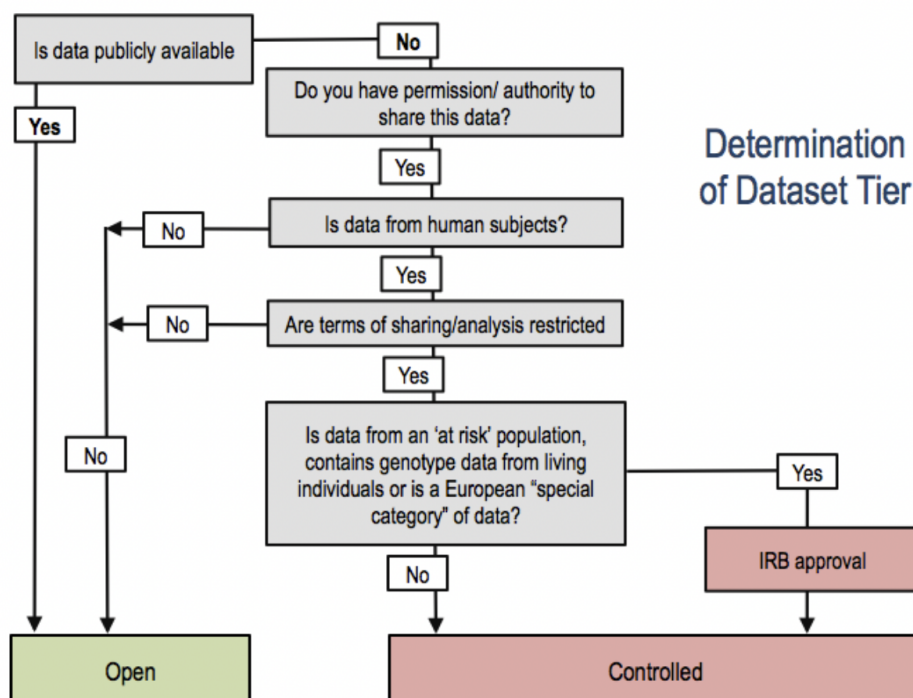
b. Review of data contributions

The Sage Bionetworks Access Compliance Team (ACT) will assist data contributors upon their request to determine the data use conditions for each dataset. To ensure that all required information has been provided, the ACT will inspect submissions of Controlled Data to the Synapse Commons. Exemplary reasons for rejecting data contribution include improper data sharing plan, lack of authority of the contributor to share a given dataset or the need for an IRB opinion on data sharing. Users who repeatedly misapply Controlled Data labels to data sets that should be Open Data may have their data contribution rights revoked to prevent overburdening the ACT. Submission of Open Data is automated and is not reviewed by the ACT, although we do provide a reporting function to notify us of mislabeled data. Submitted datasets are given a data accession number and are tagged for easy sorting or searching on Synapse.

c. Sharing of contributed data

Data use conditions are distinct from data sharing privileges. Data contributors determine with whom they choose to share each dataset. They may share data with: (1) all registered Synapse users, (2) a limited set of investigators, (3) or they may retain access solely for themselves. Data use and sharing conditions (tiers) are determined by legal, regulatory, and privacy rules and must be set for all data shared through Synapse.

To use a dataset, a registered Synapse user must (1) be granted access to the data from the data contributor and (2) agree to all data use conditions that apply to that dataset. Groups of investigators using Synapse to share data Privately are not required to set data use conditions for

sharing within their group. Still, data use conditions must be set to allow data access to the Synapse Commons.

7. Data Use Terms and Conditions
a. Requesting access to datasets

Registered Synapse users have agreed to standard terms of use for all data and may request the use of any dataset within the Synapse Commons and those shared with them by collaborators. They are automatically permitted to use all Open Data in Synapse.

Use of Controlled Data requires agreement to additional conditions to maintain privacy protections, including respect for all regulatory or legal conditions delineated in the informed consent or legal contracts as disclosed by the data contributor.

The Synapse ACT monitors requests to use Controlled Data. Requests may include a data analysis plan and documented proof that the user is authorized to use these data within the confines of that plan. Authorization of a data analysis plan can be in the form of approval or exemption from an ethical governing body (e.g., IRB or ethics committee). **Users may not redistribute Controlled Data. Users who access Controlled Data may be required to provide periodic reports on their data use. Likewise, Users may not share individual-level human data or results outside their Synapse Commons projects space.** Individual-level data or results refers to any file with values for an individual, as opposed to aggregate data, which are data combined from several individuals. Note this is true for files obtained directly from Synapse Commons or generated through the user's own analysis.

Conditions for data use		
	Open	***Controlled***
Registration into Synapse	X	X
Standard Terms of Use and Privacy Policy	X	X
Dataset-specific Terms of Use		X
User agrees not to redistribute		X
User provides project description/Statement of Intended Use	Optional	X
Certified User		Optional
Validated User profile		Optional

Use of Controlled Data is granted for 1 year or longer as determined by the data contributor. Authorization for data use of Controlled Data will be renewable.

8. Terms and Conditions of use of Synapse

The Synapse Commons is designed to host project workspaces where researchers can share tools, analytical code, and results in an open-source manner. Although most of these algorithms may be shared under permissive licenses, Sage Bionetworks cannot preclude the possibility that algorithms may be licensed, patented, or shared by individuals that do not have the appropriate permissions to share such information. As such, we will require all Synapse users to take full responsibility for using code obtained through Synapse.

9. Working within Synapse
a. Creating or joining a Project

Synapse Projects are online workspaces where researchers can collaborate and organize their work. Synapse supports all kinds of working groups: individuals, small teams, and large consortia. Any Synapse user can create a Project, populate it with content, including data, invite other users to the Project, and assign administrative privileges over it to other users. Sage Bionetworks is not responsible for the content, ideas, or opinions posted in Projects.
There are different levels of Project sharing settings: "View," "Download," "Edit," "Edit and Delete," and "Administrator." The user creating a new project is automatically given "Administrator" rights and can assign access levels, including "Administrator," to other users on the Project.
By default, projects marked "Public" confer "View" privilege to any web user. Additional privileges are only granted to registered Synapse users.

Public Project Access Privileges

	View	Download	Edit	Edit & Delete	Administrator
Search for and see the existence of a Project, File, or Folder	X	X	X	X	X
View content annotations	X	X	X	X	X
View contents of a file		X	X	X	X
Download content		X	X	X	X
Edit content			X	X	X
Delete an entity				X	X
Change a folder or file's metadata					X
Assign/change sharing settings for project users, including removing themselves					X
Create Project					X

b. Adding dataset(s) to a Public Project

To upload data in a given project, one must be a registered Synapse user with edit permissions to upload data to a Project. In addition, users also must pass the Certification Quiz to become 'Certified User' (see above).

c. Working as a team

Users may organize into Teams to work jointly on various Projects. In such a case, the rules for data contribution and use remain in effect. Only team members with data use approval can download datasets, and only team members who completed the Synapse Certification Quiz satisfactorily are permitted to upload data.

10. Working outside of Synapse

Data use is not restricted to individuals seeking to work within Synapse. Unless specified otherwise, users may choose to work on other platforms or within their local systems, contingent on an agreement to abide by the same terms and conditions delineated within Synapse.

11. Procedure for monitoring data use

Adherence to Controlled Data use conditions is the responsibility of the users and their institutions. Transparency is critical in promoting the proper use of the data and ensuring that unintended consequences of misusing the data are identified and addressed. For this reason, users wishing to access Controlled Data may be required to submit an Intended Data Use Statement; these statements may be made publically accessible within Synapse. Use of Controlled Data is monitored through random audits, and by using any Controlled Data, users agree to permit Sage Bionetworks, or third parties acting on Sage Bionetworks' behalf, to conduct such audits and to cooperate reasonably with it. To promote collaboration, users are encouraged to perform all analyses within publically accessible project workspaces on the Synapse Commons or in a similar open system. Users may be asked to submit a short report about their activities annually. Synapse activities are monitored regularly, and monitoring reports are shared as needed with external ethics advisors. Periodically we will consider mechanisms to improve the data governance process.

12. Consequences of misconduct

Consequences of misconduct include loss of current data use privileges and future use of Synapse. Sage Bionetworks will communicate any intentional or accidental violation of terms of use to the data user, institutional officials, and the data contributor and revoke data access if data misuse is identified or upon institutional request. Users and institutions will be allowed to petition for the right to return to Synapse, and the ACT will evaluate such requests. Users may file an appeal, which could result in a reversal or modification of the ACT findings of research misconduct.