

Synapse Commons Data Use Procedure

Last Update: March 18, 2013

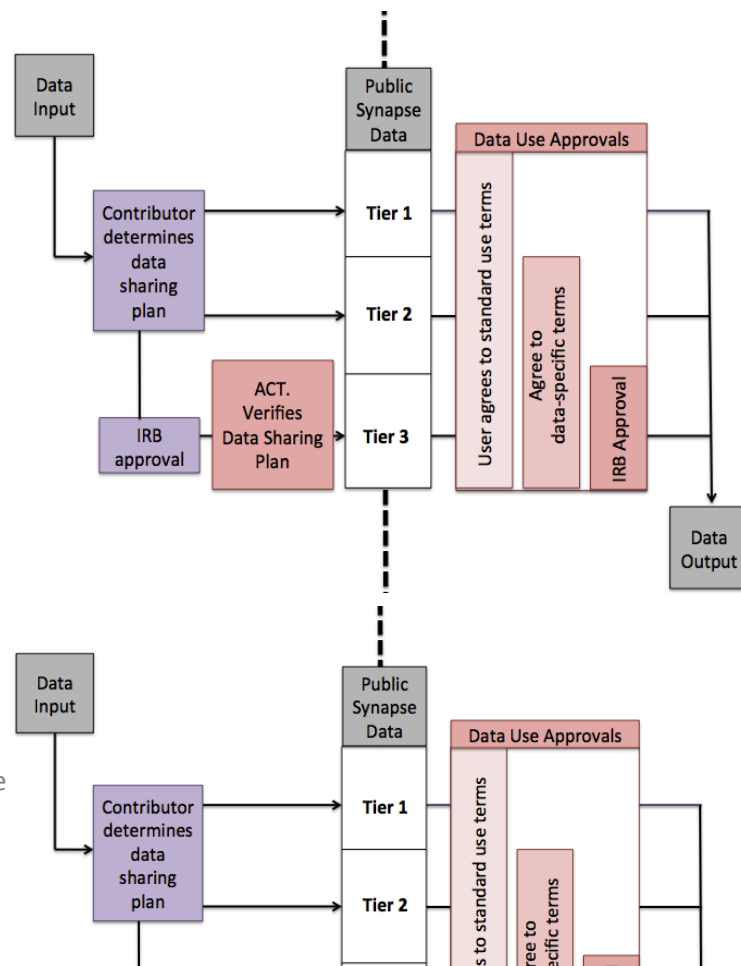
This document describes the standard operating procedure for use of data within the Synapse Commons. Synapse is a software platform designed to enable combined analysis of biological data with the goal of speeding improvements in healthcare through collaborative research. A central feature is a repository of datasets that contain genomic, molecular and/or phenotypic data from human and non-human sources. These datasets are available in standardized formats for analysis. **Human data must be de-identified according to HIPAA standards prior to being stored in the Synapse Commons in order to protect subject privacy.** For information about de-identification in accordance with the HIPAA Privacy Rule, please visit hhs.org.

Collaborations within Synapse are facilitated through the creation of 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 and/or owner-designated project administrator(s) (“**Private**”).

This policy describes governance for data use in the public content of Synapse, termed the Synapse Commons. Although data contributors may choose to broadly share data with the public, data use conditions may be required to protect the individuals from whom the data is derived.

Figure 1. Process for protected data access.

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 protections against risk for harm to human participants. This system is enacted through procedures for data contribution and data use (Figure 1). Data contributors must identify the specific risks and appropriate protections required for sharing of each dataset. Synapse informs users of these conditions and records users’ agreement to adhere to all dataset-specific data use conditions prior to making the data available.



1.Data sharing plans

Data is made available through the Synapse Commons using a tiered system. While it is Sage Bionetworks' policy to promote data sharing with as few restrictions as possible ("Open Tier"), this system accommodates data sharing under restricted conditions when required by informed consent, legal contract or other privacy requirements as disclosed by the data contributor ("Restricted and Controlled Tiers"). This process is designed to allow data sharing within the legal and ethical framework for responsible use of human data.

- **Open Data** includes data that may be shared with every Registered Synapse user in an unrestricted manner.
- **Restricted Data** includes data that may be shared only following user agreement to comply with additional data-specific terms. These terms are typically derived from the informed consent documents signed by participants contributing data for that study and/or from IRB-approved proposals. Sharing restricted access data confers minimal additional risk of harm to human participants and is managed through an automated agreement and approval process.
- **Controlled Data** includes data requiring additional protections for human subjects, like sensitive information or data from "vulnerable populations" as defined using [OHRP guidelines](#) or by the [European Data Protection Directive](#) and its implementation in applicable local law. Sharing of controlled data is considered to confer more than minimal additional risk of re-identification and/or harm to human participants and contribution or use of these data requires independent review/monitoring.

Here are examples of the types of data typically found in each Tier: In cases where an individual has consented to share data in an unrestricted manner, that data will be provided under the Open Access guidelines 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.

Open Data	Restricted Data	Controlled Data
-Nonhuman data -Human data: <ul style="list-style-type: none">- Publically available elsewhere- De-identified data with no known sharing or use restrictions- Self-contributed data unambiguously consented for Open Access data sharing- Copy number variation (CNV) data	Data that can only be shared under conditions outlined in informed consents or contracts: <ul style="list-style-type: none">-Research field of use-Type of analysis-Obligation to partner with the data generator-Obligation to share results with the data contributor-User affiliation limitation (e.g., nonprofit only)	Unless otherwise directed by research participants, controlled access data include consented: <ul style="list-style-type: none">- Sequencing and/or genotype data from living individuals.- Sensitive data from "vulnerable populations" as defined by OHRP- Health & genomics data from European individuals

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 data is no longer identifiable (e.g., figures of results, CNV summaries of SNP data, and/or gene expression summaries from RNAseq data).

2. Browsing/Searching for 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 dataset, a brief description from the data contributor, species, disease area, data type, number of individuals in the dataset, name/contact of 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. Registering in Synapse

All Synapse Commons users must register and create a Synapse user account in order to either contribute or use data within Synapse. Users register by providing a full name and valid e-mail address as well as agreeing to the Synapse Terms and Conditions of Use and the Synapse Privacy Policy. These terms include the agreement to:

- Properly attribute data generators/contributors
- Adhere to ethical and responsible use of data
- Protect and respect the privacy and confidentiality of individuals from whom data is derived
- Understand and respect the rules for use and redistribution of data, software and other content
- Comply with all regulatory mandates and all applicable laws and regulations

Registered users receive a Synapse account, a user name and a unique password that permit them to operate within Synapse.

4. Procedure for data contribution

a. Submitting a new dataset

Individuals interested in submitting a dataset to the Synapse Commons (“data contributors”) must be a registered Synapse user. Data contributors are responsible for adhering to all applicable data protection laws and regulations. In addition, data contributor must be de-identified data to HIPAA standards prior to inclusion in Synapse and where required by applicable law, 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. A [tutorial](#) helps Contributors assess the suitability of the contribution and determine the conditions to implement for sharing each dataset. The flow chart in Figure 2 illustrates the questions to address in order to assign the appropriate data use conditions. In cases where an individual (study participant) has consented to share data in an unrestricted manner, that data will be provided as Open Data regardless of what type of data it contains. If questions remain about Tier assignment, particularly for data involving human subjects, users are encouraged to seek advice from an ethics committee or independent Institutional Review Board (IRB). For contribution of Controlled Data, contributors must provide documentation verifying that an IRB or an ethics board has approved the proposed data-sharing plan.

Data contributors are responsible for indicating the dataset-specific conditions required for data use through Synapse. Sage Bionetworks may, but is under no obligation to, assist data contributors with their IRB/Ethics review submissions. Sage Bionetworks also 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 data sharing plan. In these cases, Sage Bionetworks can request, and the data contributor must provide, the (English translated) templates of the informed consent documentation and/or protocol governing the dataset in replacement of an IRB decision.

b. Review of data contributions

The Sage Bionetworks Access Compliance Team (ACT), a team of Sage Bionetworks employees trained in human data protection, will assist data contributors upon their request to help determine the data use conditions for each dataset. In order to ensure that all required information has been provided, the ACT will review all submissions of Controlled Data to the Synapse Commons. Anticipated reasons for rejecting data contribution include improper data sharing plan, lack of authority of the contributor to share a given dataset, and/or the need for an IRB opinion on data sharing. Submission of Open and Restricted Data is automated and is not reviewed by the ACT. 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 (Tier) are determined by legal, regulatory, and privacy rules, and must be set for all data shared through Synapse.

In order 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 in a private setting are not required to set data use conditions for sharing within their group but data use conditions must be set in order to open data access up to the Synapse community.

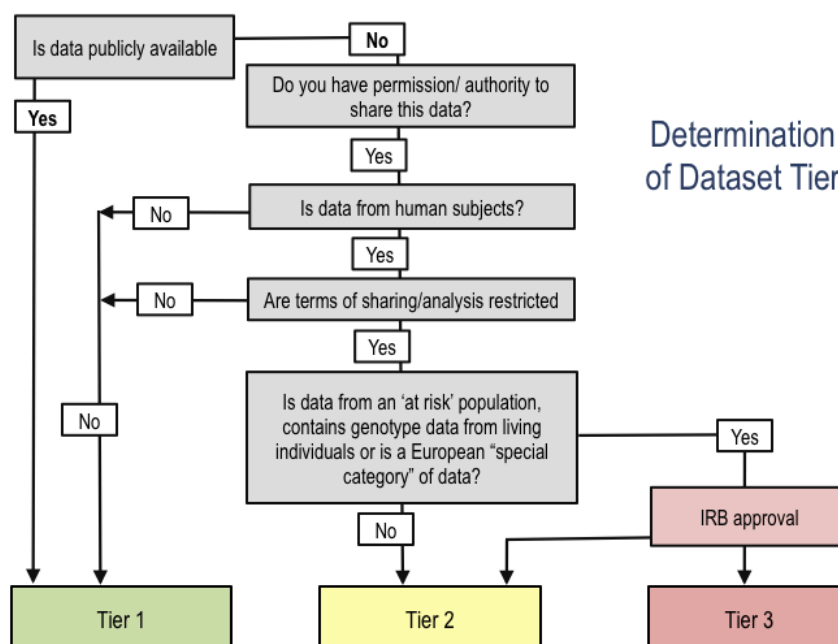
5. Data Use Terms and Conditions

a. Requesting access to datasets

Registered

Synapse users may request use of any dataset that has been shared with them, including those within the Synapse Commons and those shared with them by collaborators. By registering in Synapse, users agree to the standard Synapse Terms and Conditions of Use and Synapse Privacy Policy as stated above and, as such, are automatically permitted to use all Open Data. Use of Restricted or Controlled Data requires agreement to additional data use conditions to maintain privacy protections, including respect of all regulatory or

Figure 2. Flow chart of considerations for data contributions.



legal conditions delineated in informed consent or legal contracts as disclosed by the data contributor. **Open and Restricted Data may be redistributed, but only if the redistributing user assures that all subsequent recipients comply with all applicable terms and conditions and privacy protections.**

<i>Standard Terms of Use for all datasets</i>
Attribution of Data Generator/Contributor
Legal Compliance with HIPAA rules (no re-identification) and other data protection laws.
Ethical Behavior (prevent misuse of data, discrimination, harm to population)

<i>Examples of Dataset Specific Terms</i>
Limit on research field of use (e.g. disease-specific)
Limit on type of analysis
Requires return of analytical results to data contributor
Limit to biological/medical research use (e.g. health-care focus)
Limit on user affiliation (e.g. commercial, not-for-profit)
Requires annual report to data generator

Requests to use Controlled Data are monitored by the Synapse ACT. They must 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 data analysis plan can be in the form of an approval or exemption from an ethical governing body (e.g., IRB or ethics committee). **Controlled Data may not be redistributed and users will be required to provide periodic reports on data use.**

Application Process:			
	<i>Open</i>	<i>Restricted</i>	<i>Controlled</i>
Registration into Synapse	X	X	X
Standard Terms of Use and Privacy Policy	X	X	X
Dataset-specific Terms of Use (as applicable)		X	X
User agrees not to redistribute			X
User provides project description	Optional	Optional	X
Proof of IRB authorization (Approval or Exemption)			X

b. Review of data use requests

Users must read and electronically agree to the standard Terms and Conditions of Use and Privacy Policy upon registration. Agreement to these enables access and use of all Open Data. In order to access Restricted Data, users are required to agree to additional terms on a data-specific basis. Because distribution of Restricted Access data is considered to confer minimal additional risk of harm to participants, these agreements will be automated.

For access to Controlled Data, users must submit proof of authorization for data analysis plan to the Sage Bionetworks Access and Compliance Team (ACT) for review. In some cases the ACT may request additional information in order to properly assess a request. Incomplete submissions will not be reviewed or considered. Use of Controlled Data is granted for 1 year or longer as determined by the data contributor. Continued authorization for data use is renewable with proof of continued IRB/ethical review or exemption.

6. Terms and Conditions of use of Synapse

The Synapse Commons is designed to host project workspaces where researchers may share tools, analytical code, and results in an open source manner. Although the majority of these algorithms are expected to be shared under permissive licenses, Sage Bionetworks cannot preclude the possibility that algorithms may be licensed or patented and/or shared by individuals that do not have the appropriate permissions to share such information. As such, we will require all Synapse users to adhere to Terms and Conditions of Use and Privacy Policy indicating, among other things, their awareness of these possibilities and their agreement to take full responsibility for their use of code obtained through Synapse.

7. Working within Synapse

a. Creating or joining a Workspace

Synapse users can collaborate with other users within project workspaces. Any Synapse user can create a workspace, populate it with content including data, invite other users into the workspace, and assign administrative privileges over the project to other users. Sage Bionetworks is not responsible for the content, ideas or opinions posted on workspaces.

There are 3 levels of workspace access available within Synapse: “Manage”, “View” or “Edit”. Workspace owners have “managing” rights and may select those users that may “Manage”, ‘View’ and/or ‘Edit’ the project. Use of Restricted or Controlled data within a workspace requires:

- 1) Permission to view the workspace (at the discretion of workspace managers)
- 2) Agreement to the data use conditions (as described in this document)..

Public Workspace Access Privileges				
	Read	Edit	Manage	Owner
Create Workspace				X
Assign/change access right of project users			X	X
View data and related analysis results		X	X	X
Edit content		X	X	X
Review project description	X	X	X	X
Review list of datasets	X	X	X	X
Review list of workspace contributors	X	X	X	X

b. Adding dataset(s) to a public workspace

Users may not make new datasets publicly accessible without completing the Synapse data contribution process (Section 4). Only workspace members with editing privileges can add datasets to an existing workspace. When new data is imported into a workspace (either from the Synapse Commons or an external source), access is restricted to those workspace members who have agreed to the appropriate data use conditions. Users will be allowed to upload new datasets into private projects without

undergoing this process. However, they will be required to complete the data contribution process if they decide to make the project or data public.

8. Working outside of Synapse

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

9. Procedure for monitoring data use

Adherence to Controlled Data use conditions is the responsibility of the users and their institutions. Transparency is a key element in promoting proper use of the data and in ensuring that unintended consequences to use or misuse of the data are identified and addressed. Use of Controlled Data will be monitored through random audits, and by using any Controlled Data, you agree to permit us or third parties acting on our behalf to conduct such audits and to cooperate reasonably therewith. 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 working outside of the Synapse system will be asked to submit a short report about their activities annually. It is our intention to link these reports to the dataset(s) and to users profile in Synapse.

Monitoring of Synapse activities will be discussed on an annual basis with an external Ethics Advisory Team, (SB-EAT). The SB-EAT is composed of independent international experts in bioethics and will provide guidance on the implementation of the Sage Bionetworks Data Access Policy and on the ethical, social and policy implications of the Synapse Commons. The goal of these efforts is to identify risks and unintended consequences of data use and to confirm that appropriate human data use protections are in place. Periodically we will consider mechanisms to improve the data governance process. It is also our intention to make documents associated with SB-EAT/ACT activities accessible through the Synapse Commons to ensure transparency and encourage comments/feedback from the entire community.

10. Consequences of misconduct

Consequences of misconduct include loss of both 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 will 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 such requests will be evaluated by the external SB-EAT.