

Synapse Commons Data Access and Use Procedure

Last updated: November 22, 2011

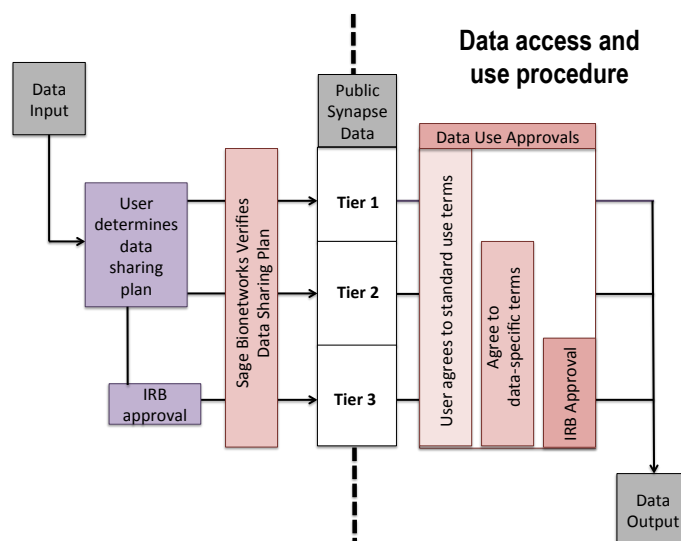
This document describes the standard operating procedure for access to and use of data within Synapse Commons. Synapse is a software platform designed to enable combined analysis of genomic data with the goal of speeding improvements in healthcare through collaborative research. A central feature is 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 and integration into computational disease models. **All human data contributed to Synapse must be de-identified prior to being provided for use on the Synapse Commons and must be coded to maintain strict compliance with HIPAA guidelines for subject privacy.**

Collaborations within Synapse are facilitated through the creation of project workspaces where analytical workflows and datasets may be shared. Each project is organized with a home page, a workflow page, selected datasets and custom analysis content. The project homepage includes general information about the project including a project summary, a list and brief description of relevant datasets used in the analysis, the name of the workspace creator and the list of active collaborators with full access to the workspace content.

Synapse users may create projects with either public or private access rights. By default, Synapse projects are public, meaning that the content is readable and/or editable by the entire Synapse user community. Users wishing to work on these projects or access the project content may join the workspace as collaborators. Users can also elect to create a Private workspace. This may happen, for example, under cases where accessibility to work is embargoed by a legal contract. The existence of Private projects is confidential and access to content is restricted to designated investigators selected by the project owner. This policy describes governance for data access and use in the public component within Synapse, termed the Synapse Commons.

Data access and use through the Synapse Commons will be governed using a tiered access system designed to ensure that each dataset is shared using the appropriate level of protections against risk for harm to human participants. This system will be enacted through procedures that govern data contribution and data use (Figure 1). The data contribution procedure will identify the appropriate risks and protections for each dataset. The data use procedure will inform data users of these risks and require their agreement to adhere to the identified protections.

Figure 1. Process for protected data access.



1. Data sharing plans: Tiered Access

Data will be made available through the Synapse Commons using a three-tier data access system. While it is Sage Bionetworks' policy to promote data sharing with as few restrictions as possible (Tier 1), 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 (Tiers 2 and 3). This process is designed to allow data sharing within the legal and ethical framework for responsible use of genomic data.

-**Tier 1** includes data that may be shared with every Synapse user in an unrestricted manner.

-**Tier 2** includes data that may be shared only following agreement to additional data-specific terms of use usually outlined in the informed consent documents for that study.

- **Tier 3** is reserved for data requiring additional protections due to inclusion of potentially sensitive information such as genotype data from living individuals, rare information or inclusion of vulnerable populations as defined using OHRP guidelines (<http://www.hhs.gov/ohrp/policy/index.html>) .

Sharing of Tier 1 and Tier 2 data confers minimal additional risk of harm to human participants. Although all human data in Synapse Commons is de-identified according to HIPAA regulations, sharing of Tier 3 data is considered to confer more than minimal additional risk of re-identification and potential harm to study participants.

Here are examples of the data types in each tier:

Tier 1	Tier 2	Tier 3
<ul style="list-style-type: none">-Nonhuman data-Human non-genotype data without additional limitations-Human genotype data obtained from:<ul style="list-style-type: none">Deceased individualsData already public elsewhere without restrictionsIndividuals consented for data sharing-Human copy number variation (CNV) data	<p>Human non-genotype data with conditions outlined in informed consent such as limitations on:</p> <ul style="list-style-type: none">-Research field of use-Type of analysis-Users must be a research partner with data generator-Requirement for return of results-User affiliations (e.g., nonprofit only)	<p>Human sequencing and/or genotype data from living individuals</p> <p>Human non-genotype data from sensitive populations</p>

Note that access restrictions apply to the dataset as well as the analysis results derived from it.

2. Browsing/Searching for dataset of interest.

Anyone can search or browse the list of public datasets available on Synapse using key search terms or tags. The displayed information includes:

- Name of dataset
- Brief description from data contributor
- Species
- Disease area
- Data type
- Number of individuals
- Name/Contact of data contributor
- Tier assignment or access conditions
- Synapse dataset identifier

3. Procedure for data contribution

a. Submitting a new dataset

Individuals interested in submitting a dataset to the Synapse Commons must work with Sage Bionetworks to ensure that data is shared using the appropriate terms of use and restrictions. Data contributors must be a registered Synapse user and is responsible for ensuring that data adheres to all applicable laws and regulations regarding the protection of participant privacy. All human data must be de-identified to HIPAA standards.

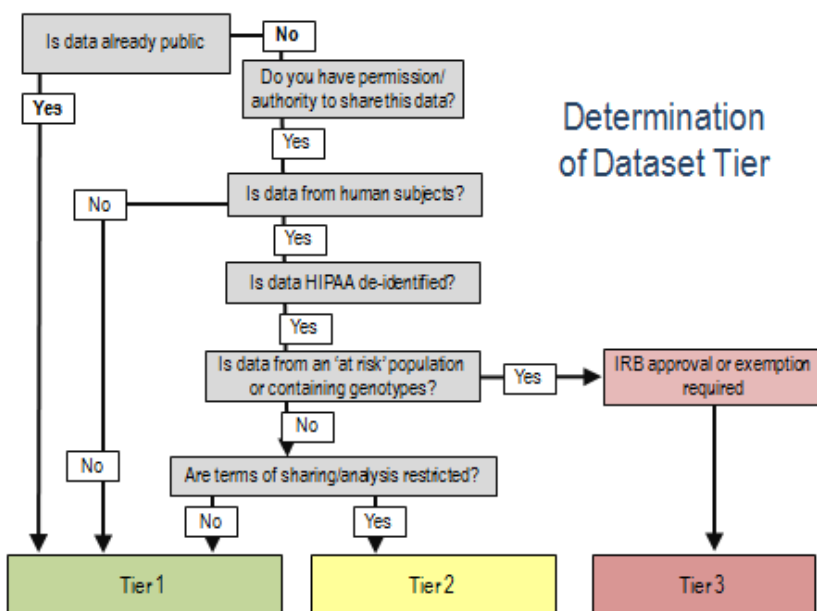
The dataset contribution process on Synapse is semi-automated. Contributors are guided through a Web-based questionnaire to help them determine the eligibility and appropriate protections for sharing of each dataset (Figure 2). If questions remain about Tier assignment, particularly for data involving human subjects, users are encouraged to seek advice from an ethics committee or independent IRB. For contribution of human non-genotype data, submission of informed consent documentation and/or IRB opinion is recommended but not required as these datasets are not considered to confer more than minimal risk of harm to participants (Tier 1 and Tier 2). For submission of Tier 3 data, contributors must also provide documentation that an IRB or ethics board had approved the data-sharing plan.

In addition to contribution of data, data contributors are also required to provide the following information for the purpose of describing the data source and characteristics for use by other interested Synapse users:

- Data contributor name/institution
- Dataset annotations (e.g., species, disease, tissue, number of samples)
- Proposed data sharing plan (Tier assignment and/or dataset-specific conditions for sharing)
- Proof of Ethics or IRB review (for Tier3) Brief Dataset description (optional)

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 Institutional Review Board (IRB) to confirm the data contributor's selection of data sharing plan. In these cases, Sage Bionetworks will request, and the data contributor will provide, the templates of the Informed Consent documentation 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 work closely with data contributors upon their request to ensure appropriate Tier assignment of each dataset. Submission of Tier 1 and Tier 2 data is automated but the ACT will manually review submissions of Tier 3 data prior to posting the Data on the Synapse Commons to confirm that Contributors have certified that data distribution and use conform to HIPAA and other applicable regulations and approvals. If a data sharing application is incomplete the data contributor will be notified and will be invited to resubmit. Anticipated reasons for rejection 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.

Data submissions will be given a data accession number and will be tagged for easy sorting or searching on Synapse. Data access will be provided through Synapse according to the sharing conditions described in this document.

c. Sharing of contributed data

Data contributed to Synapse may be shared in one of three ways: (1) shared with all registered Synapse users, (2) shared with a limited set of designated investigators, (3) or access solely by the data contributor. Sharing data through either (1) or (2) requires completion of the data contribution process in order to define the dataset Tier assignment and must include proof of IRB approval for data use by all registered Users or the designated investigators for Tier 3 datasets before data may be shared.

4. Accessing data: Terms and Conditions

a. Registering in Synapse

All Synapse Commons users must be registered in order to access data within Synapse. User registration requires submission of full name and a valid e-mail address and agreement to the standard Synapse Terms and Conditions of Use including:

- Properly attributing data generators/contributors
- Adherence to ethical use of data
- Compliance with all regulatory mandates and local laws.

User information is verified through an 'Open Access' registered identifier system. Registered users receive a Synapse account, user name and unique password that permits them to operate within Synapse.

b. Requesting access to datasets

Registered Synapse users may request access to any dataset within the Synapse Commons. Prior to being granted data access, users must agree to the standard Terms and Conditions of Use. These include the agreement to properly attribute data generators/contributors, to adhere to ethical use of data, and to comply with all legal and regulatory mandates. In addition to respecting the conditions set forth in the Synapse data access process, users must also abide by all local laws. Users will be encouraged to provide a short description of intended data use upon initial request to access each dataset. This description will be recorded and made publicly available through Synapse on the user's Synapse page and through the dataset page. All data access requests will be made public through Synapse. Users will also be encouraged to report any publications or other products or data use so that such knowledge can be shared among all Synapse users.

Following agreement to the Synapse standard Terms and Conditions of Use, users may access all Tier 1 datasets. Additional dataset-specific terms of use are also required for access to Tier 2 datasets. These terms are dataset-specific and are imposed only when required to maintain protections, regulatory or legal conditions described as part of informed consent or legal contracts enacted during the process of data generation and disclosed by the data contributor. **Tier 1 and Tier 2 data may be redistributed, but only if the redistributing user assures that all**

subsequent recipients will comply with the standard Terms and Conditions and any dataset-specific terms of use.

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

<i>Examples of Dataset Specific Terms (only applied as required in Tier 2 or 3)</i>
Limit on research field of use (e.g. disease-specific)
Limit on type of analysis
Requires return of analytical results
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

Tier 3 data use requires IRB authorization of intended data use, in the form of an approval or exemption. Users may seek IRB authorization to access an individual dataset or a group of datasets. Sage Bionetworks encourages researchers to access only those Tier 3 datasets that are specifically related to their research interests. For example, a diabetes researcher may seek approval to access all data related to metabolic disease. Synapse contains guidance to facilitate users through the IRB submission process. Users are required to provide documented proof of IRB authorization including an IRB reference number and an IRB contact for verification. Data within Tier 3 will be categorized into groups of datasets with similar research topics (typically by disease). **Tier 3 data may not be redistributed.**

Application Process:			
	<i>Tier 1</i>	<i>Tier 2</i>	<i>Tier 3</i>
Registration into Synapse	X	X	X
Standard Terms of Use	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

c. Review of data access requests

Because distribution of Tier 1 and Tier 2 data is considered to confer minimal additional risk of harm to participants, user agreement to adhere to data use terms will be considered sufficient to authorize data access without any additional authorizations. Tier 1 and Tier 2 data access application and approval processes are automated. Users will be directed to read and submit electronic agreement to the standard Terms and Conditions of Use. Sage Bionetworks will also explore the utility of a short instructional video that describes data use terms and their implications. Terms of Use agreements are recorded electronically and made public through Synapse. The standard Terms and Conditions of Use will be displayed every time a user requests access to an additional dataset.

For Tier 3 data, a complete application including proof of IRB authorization will be submitted to the Sage Bionetworks Access and Compliance Team (ACT). This team consists of 3 or more Sage Bionetworks employees with experience and training in human data protection. A minimum of 2 ACT members will review each application. In some cases the ACT may request additional information in order to properly assess a data access request. Incomplete submissions will not be reviewed or considered. Access to Tier 3 data will be granted for 1 year, renewable with proof of annual IRB review/exemption and approval confirmation. Sage Bionetworks tracks all data access requests and data uses.

5. Terms and Conditions of use of Synapse

The Synapse Commons is designed to host public 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 indicating, among other things, their awareness of these possibilities and their agreement to take full responsibility for their use of code obtained through Synapse.

6. 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 datasets according to his/her specific dataset permissions and invite collaborators or guests to the workspace. The workspace creator is responsible for managing and approving access rights to the workspace. 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 ‘View’ and/or ‘Edit’ the project. By default, Synapse workspaces are Public, meaning that content is viewable by the entire Synapse user community. Access to a workspace does not automatically provide access to datasets that are used within that workspace. When data is used within a Synapse project, the data access permissions are set independently from the project permissions. Users wishing to access the content of a Public workspace and collaborate on the project need to:

- 1) Join the workspace (at the discretion of workspace owner)
- 2) Obtain permission to all the datasets included in the workspace (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 Sage Bionetworks data contribution process (Section 4). If users attempt to upload new datasets into a public project, they will be required to

complete this process before data will be accepted. 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), data access is restricted to those workspace members with the appropriate data access permissions. 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 public.

7. Working outside of Synapse

Data access is not restricted to individuals seeking to work within Synapse. Users may choose to work on other platforms or within their own local systems. In cases where users download data for use outside of Synapse, they will be periodically requested to provide a short description of the resultant work. This information will be publicly accessible through the user's Synapse profile as well as through the dataset page.

8. Procedure for monitoring data use

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. All public data access requests descriptions and annual reports of public projects will be broadly accessible through Synapse Commons. In addition, to promote open collaborations, users are encouraged to perform all analyses within publically accessible project workspaces on the Synapse Commons or in a similar open system.

Tier 3 data use will be monitored through annual evaluations of data use. Tier 3 data users will be asked to submit proof of IRB authorization, a description of data use, a list of collaborators, and description of publications or other research products stemming from these projects. When analyses are conducted within Synapse Commons, the yearly project update will be generated automatically from information within that project and users will be asked to verify that the content is up to date. Tier 3 data users that work outside of the Synapse system will be required to submit a short report to be linked to the dataset(s). Adherence to Tier 3 terms of use is the responsibility of the users and their institutions.

Monitoring of Synapse activities will be discussed on a bi-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 Sage Bionetworks Commons. The goal of these efforts will be 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 access process. Documents associated with SB-EAT/ACT activities will also be posted within the Synapse Commons to ensure transparency and encourage comments/feedback from the entire community.

9. Consequences of misconduct

Consequences of misconduct include loss of both data access privileges and future use of Synapse. Sage Bionetworks will communicate any intentional or accidental violation of terms of use to the data user as well as institutional officials 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.