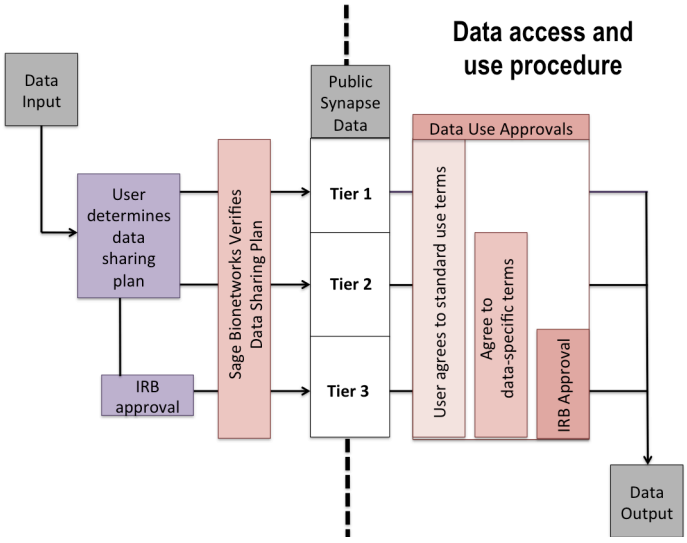
Governance of Human Data Use within the Synapse Commons

Synapse is an open compute space designed to speed the pace by which genomic science can impact human health. We aim to collectively enhance biological knowledge by providing a platform for researchers to share genomic datasets, methodological tools and analytical results in a dynamic fashion prior to formal publication. Information shared within Synapse may contain genomic, molecular and/or phenotypic data derived from human or non-human samples and may originate from either public (GEO, Array Express, TCGA) or private sources. Users may share data to encourage community participation in analytical development within the scope of their own research. Alternatively, users may utilize Synapse to contribute data to the community for inclusion in projects initiated by other researchers.

Synapse data governance policies have been developed to allow data sharing within a legal and ethical framework for responsible use of genomic data (**Figure 1**). These guidelines balance appropriate protection of human participants with the facilitation of collaborative data sharing for advancement of community-based science. Data governance within Synapse is managed through using both policy and technological implementation. The underlying principles of this system have been distilled into a set of terms and conditions that must be signed by all Synapse users. These include requirements to: use and contribute only data that have been de-identified to HIPPA standards; not to discriminate, identify or recontact individuals or groups represented by the data; attribute data contributors when discussing results from data; and redistribute data including these terms of use but only as allowed for each dataset.

Figure 1. Synapse process for contribution of and access to data.



Recognizing that the risks conferred on human participants from data sharing differ across studies, Synapse tailors data use restrictions across datasets using a tiered data access system. These tiers differ in use restrictions required to appropriately protect the individuals from whom the data is derived. **Tier 1** contains data for which there is minimal risk of re-identification or harm to individuals and these data are shared in an unrestricted manner with all Synapse users. **Tier 2** contains data that confers minimal risk to individuals but for which specific data use limitations are described. These include limitations on research field of use (e.g., data may only be used for cardiovascular disease research) or on designated researchers (e.g., data may only be used by researchers at non-profit institutions) as outlined within the informed consent process. **Tier 3** includes data with potentially sensitive information that may confer more than minimal additional risk of re-identification or harm to study participants. This includes genotypes or sequencing data derived from living individuals and all data derived from vulnerable or easily identified populations such as rare disease groups. Sharing of Tier 3 data requires the approval of an external ethics committee designated by the data contributor and may only be shared with Synapse users that provide annual proof that an ethics committee has approved their data analysis plan. Examples of data categorized within each tier are

Lara Mangravite, PhD
Sage Bionetworks

April 6 2012

described in **Table 1**. Synapse provides tools to assist data contributors in categorizing data but does not actively participate in the tier-assignment process. Permissions for data contribution and access are managed by an automated system within Synapse. Tier 3 data use is monitored.

Synapse is designed to enable transparency in data use with the dual goals of improving collaborative efforts within research and ensuring that unintended consequences of data use or misuse are addressed. Users are encouraged to perform all analyses within publically accessible project workspaces such as within the Synapse Commons. Data access requests, project descriptions, and data use reports are made public within Synapse through an automated, auditable tracking system. The public nature of Synapse ensures that data use can also be directly monitored by data contributors, researchers, funding agencies, human participants, and by the general public. As such, this is expected to develop into a system in which data use is monitored across the community. Misconduct or misuse of Synapse or Synapse content will result in loss of data access privileges and future use of Synapse.

Table 1* Data types by tier.

Tier 1	Tier 2	Tier 3
-**Nonhuman data** -**Human non-genotype** data without additional limitations -Human genotype data obtained from: **Deceased** individuals Data **already public** elsewhere without restrictions Individuals **consented for data sharing** -Human **copy number variation** (CNV) data	Human non-genotype data with **conditions outlined in informed consent** such as limitations on: -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)	Human **sequencing** and/or **genotype** data from living individuals Human non-genotype data from **sensitive populations**

***Access restrictions apply to the dataset and all resultant analyses.**

Sage Bionetworks has worked closely with an independent Ethics Advisory Team (SB-EAT) to develop these governance processes. The SB-EAT expert ethicists will continue to provide guidance on the evaluation of the ethical, social and policy implications of the Synapse collaborative research environment. Synapse activities will be regularly audited and procedures may be adapted based on observations and feedback from users and the community. The goal of these efforts is to develop an environment that enables the community to work together to advance research in a transparent, collaborative manner.