About Synapse

The past two decades have seen an amazing exponential growth in the availability of genetic and other important biomolecular data fueled by incredible advances in measurement technologies. These breakthroughs have resulted in an increasing amount of money spent on research by industry and research papers written by academia. However, despite a few examples of new block-buster drugs, we have not witnessed a general rapid improvement in the treatment of many significant human diseases. Indeed, the numbers of new drugs approved by the FDA has actually declined over this same period.

A fundamental reason improvement in clinical outcomes does not match the pace of biological data production is that the analysis and interpretation of this data remain largely an individual activity limited by the bandwidth of an individual scientist. Drug development pipelines require a series of handoffs among many individual scientists, each with different areas of expertise. In our experience, the ability to access, understand, reformat, and reuse data, analysis methods, or models of disease at each step in the pipeline is a significant rate-limiting step, even within the confines of a single company or research institution. Additionally, much of the relevant data to answer a particular research question is spread among multiple public and private repositories. Because each pharmaceutical company and academic group protects their own data, the end result is enormous duplication of effort and missed opportunities across the industry as a whole.

Sage Bionetworks' mission is to catalyze a cultural transition from the traditional single lab, single-company, and single-therapy R&D paradigm to a model based on broad precompetitive collaboration on the analysis of large scale data in medical sciences. If this were to happen it would benefit future patients through improved treatments to diseases. It would benefit society through reduced costs of health care. It could even benefit the pharmaceutical industry, which is struggling to replace the revenue lost as old medicines go off patent production, by seeding increased innovation in the sector. Sage Bionetworks is actively engaged with academic, industrial, government, and philanthropic collaborators in developing solutions to these issues.

The technology component of Sage's solution strategy is Synapse, an informatics platform for open data-driven science. Synapse will provide support for Sage Bionetwork's own research initiatives, but more importantly serve as a public resource for the broader scientific community.

Catalyzing a transformation to truly collaborative research requires that the platform help scientists solve a series of problems that impede truly collaborative work today:

1. **Finding and using relevant data -** Frequently today scientists have difficulty just tracking down and gaining access to resources created by others, even within the confines of the same organization. Even when data is available, ~75% of the work of an analysis project is spent understanding data structures and preprocessing data for analysis. Synapse provides ways to not only make data discoverable, but push data to leverage common formats, controlled vocabularies, and annotation standards that allow data to be described and exposed to analysis unambiguously.
2. **Understanding analysis workflows -** Synapse accounts for the fact that much research is experimental and ad hoc in nature, and that hardened analysis methods only emerge over time. Tracking who has run what version of code on what version of the data in a project is the key to reproducible, and therefore improvable, research.

3. **Supporting genome-scale analysis** - Analyzing datasets with information on whole genomes is currently limited to those with access to large computational resources and significant IT support. Synapse makes cloud computing technologies accessible to scientists.
4. **Forming and maintaining productive collaborations** – Scientists tend to start from scratch on a project rather than take work in an unknown state. The platform must help scientists track what work has already been done in a particular area and help create and sustain collaborations.

To support the vision described above, a variety of types of users will need to interact with Synapse for different purposes. Initially the Synapse system will be focused on supporting the statistical or computational biologist power users in manipulating clinical and genomics data sets; over time support could grow for more biological or clinically focused researchers. Synapse leverages a web service-based architecture in which a common set of services is accessed via different sets of client applications to support a growing set of use cases over time.

One client application is the Synapse web portal. This is a "Web 2.0" environment for scientists to discover and share data, models, and analysis methods. The portal is organized around projects, which any scientist can create and invite collaborators to join. These online workspaces then serve as the glue to help teams of researchers collaborate to solve complex scientific analysis problems. Additional capabilities to visualize data and interact with disease models through the web portal will be developed over time.

We also expect our target users to be already proficient using data analysis tools, and to want to continue using those tools to work with data. Synapse's first analysis client will be an integration with the R / Bioconductor statistical package allowing users to track analyses performed with published packages or custom scripts to complete their work. Additional analysis platforms (e.g. Gene Pattern or Matlab) will get supported over time. All analysis platforms would interact with the Synapse system by calling the Synapse web service layer. This is a set of REST-based services providing support for annotating, querying, and updating data, analysis code, and models, and controlling access to these resources. These services will also allow tracking of the provenance of a multi-step analysis procedure, and executing analysis steps on cloud-based resources.

For more information, please see the Synapse Vision Document or contact synapseInfo@sagebase.org.