

Synapse Vision Document

Michael Kellen

Purpose

The purpose of this document is to describe at a high level Sage's vision of where we want to take the Sage Platform called "Synapse" over 3-5 years¹. It is intended to be the first written summary of the platform that we could hand to any new employee, collaborator, or funding agency. It should be considered a "living document" that changes as the vision matures. Please contact the author with feedback of any kind; it is only through iterations with team members that the document will become useful.

Summary

"Significantly... [Sage's] mission will be pursued in an open access "scientific commons" that will provide a resource for the general scientific community and a forum for the evolution of better models through a distributed network of contributor investigators." – Description of Sage Platform in LSDF 2009 grant.

Despite an increasing amount of money spent and research papers written, the global biomedical research enterprise as a whole is failing to generate new products that can replace the revenues lost as existing medications go off patent². In the pharmaceutical industry as a whole, an exponential growth of both biological "omics" data and research papers is occurring while the numbers of NMEs approved by the FDA is flat or declining. Sage believes that a fundamental reason improvements in clinical outcomes do not scale with 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. Pharmaceutical R&D pipelines, and even many pre-commercial research programs, consist of a series of handoffs among individual scientists with different areas of expertise. In our experience, the ability to access, understand, reformat, and reuse data, analysis methods, or models at each step is a significant rate-limiting step in research progress, 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 locations. 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.

¹ In this document "Synapse" is intended to mean "the complete set of software infrastructure that supports the sharing of data, tools, and models among Sage employees, collaborators, and the general scientific community".

² <http://www.nature.com/nrd/journal/v9/n3/full/nrd3078.html>

Is there a better way to do research? Compare the situation in pharmaceuticals to the technology industry. In that industry Moore's Law, an exponential increase in basic compute power, is successfully translated into a wealth of new product launches every year. Here, some of the most widely-used software projects in the world are open source, including the Android operating system, the Apache web server, and the MySQL and PostgreSQL databases. Furthermore, the successes of these open-source projects have not killed off large corporations. Rather, by making some commonly-needed basic infrastructure available at little to no cost, open source software has lowered the barrier to entry to the industry as a whole, and seeded innovation as entire new businesses (Facebook, Twitter, etc.) spring up by innovating on top of this open base.

It is worth considering how development occurs in the tech industry. Over the last several decades, the software industry has evolved from its roots as individual hobbyist "hackers", each maintaining his own idiosyncratic development environments to a models where large, distributed development teams increasingly managed their development efforts using standardized software development infrastructure. Currently, the trend is for software teams to outsource the hosting of code and supporting resources to organizations dedicated for this purpose (e.g. SourceForge, GitHub, Google Code, Atlassian). The tools provided by these sites (e.g. version control, bug tracking, wikis, automated build systems) let developers anywhere create a software development project and instantly access many supporting tools, and are used by open-source and professional software engineering teams alike. The success of many open-source development projects demonstrates that even highly distributed and decentralized teams can effectively collaborate on complex and large-scale projects given an appropriate collaboration framework.

Obviously, there are some fundamental reasons why the pharmaceutical industry cannot simply copy the tech industry: the high cost and inaccessibility of experimental biology and the complexity of the regulatory environment make for a substantially different environment. However, when it comes to the analysis of large-scale biological data sets, particularly in areas of basic research not directly related to the performance of a particular proprietary compound, big Pharma could stand to act a bit more like Silicon Valley. If Netflix can put data on which people like which movies into the public domain to improve the methodology of their recommendation engine, why can we not do the same for (suitably deidentified and protected) human health data for a far more important purpose?

Some of the basic pieces to create this open access space for biological research already exist. A large number of repositories provide access to a wealth of biological data from a variety of studies (dbGaP, GEO, etc). A large number of tools are freely available that provide mechanisms to analyze common data formats (Bioconductor, GenePattern, etc). The leading computational biology groups can pull together multiple datasets and tools into sophisticated analysis pipelines that allow them to construct models that predict biological behavior, although this methodology is still at a relatively immature state. However, the only way you can typically learn what these scientists has done is through a journal article that describes the work in a manner most often wholly inadequate to reproduce the analysis or apply it to a new project. What is missing is a place where biological data scientists can collaborate on these analysis projects in real-time and across organizational boundaries, a place where young investigators can actively learn methods from the leading scientists in the field by seeing exactly what was done, and

a place where relevant data, tools, and models are brought into a single shared space with sufficient compute resources to support modern biological data analysis. Sage Bionetworks is leading a community-based effort to build an open platform called Synapse which will address precisely these issues.

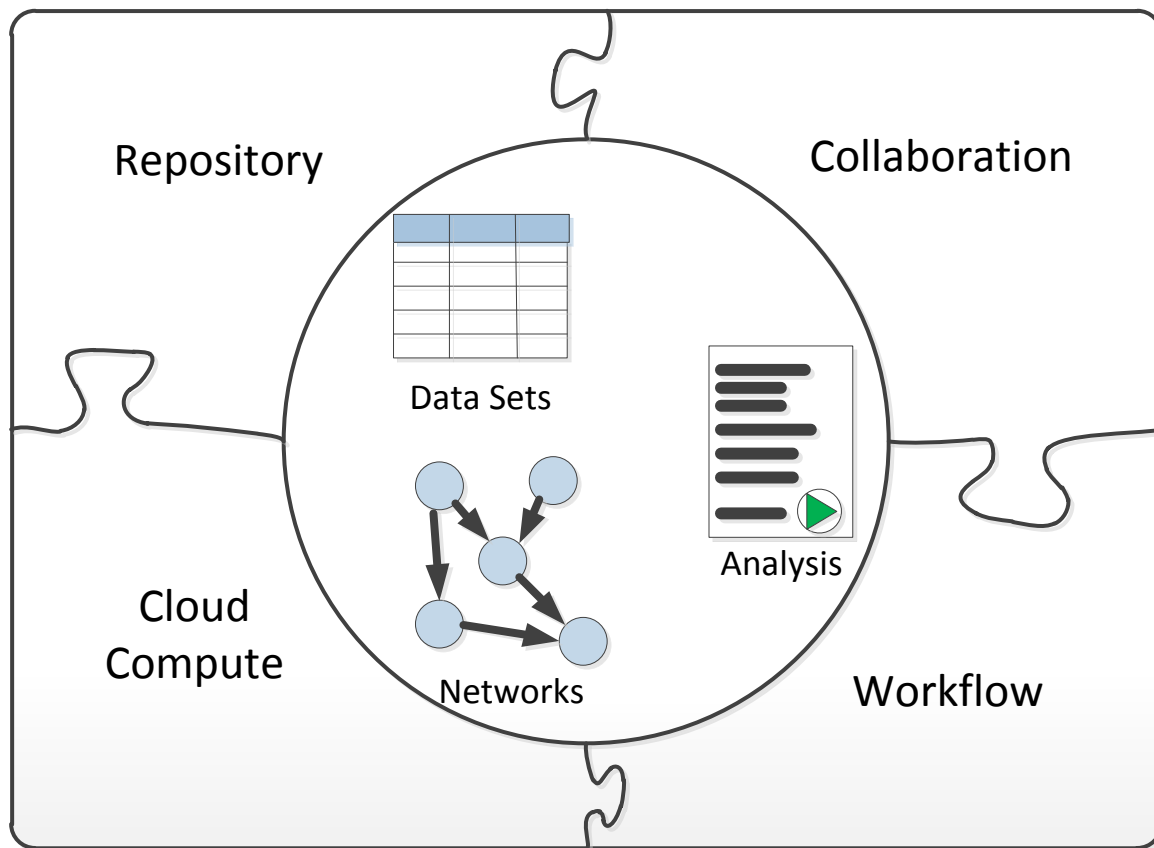
Although Sage does have some ability to create top-down pressure to drive people towards a platform (e.g. by influencing publishers or funders requirements), it is illustrating that open source code repositories do not grow by mandate. They grow because individual development teams are more effective when they let somebody else create, host, and maintain both this infrastructure and their own project code, and when development team members can easily build on each other's work in a shared environment. Sage platform developers must aim to create a product useful enough that early-career scientists would actually choose to use it on their own volition, regardless of any top-down push from senior researchers.

Synapse will make research more effective by bringing scientists together with the data, tools, and models needed to solve complex scientific problems as a team. The platform has immediate benefit for research conducted by Sage and its collaborators, but ultimately will be a vehicle that can be used by any research team anywhere. 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. **Accessing relevant data, tools, and models** - Frequently today scientists have difficulty just tracking down and gaining access to resources created by others, even within the confines of the same company. The platform must provide ways for scientists to easily discover what has already been done and make it available.
2. **Understanding data** - Sage finds that ~75% of the work of an analysis project is spent understanding data structures and formats, and reformatting, integrating and normalizing data. Access is not enough. The platform must push data to use common formats, controlled vocabularies, and ontologies that allow data to be described and exposed to analysis unambiguously.
3. **Understanding analysis methods** - Scientists commonly write disposable "research code" to answer particular questions. This leads to considerable effort expended as scientists "reinvent the wheel" to make their own particular task run. The platform must help scientists focus on only writing novel algorithms, and in understanding what other existing analysis methods already do. At the same time, Synapse must account for the fact that much research is experimental and ad hoc in nature, and that hardened analysis methods 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.
4. **Supporting analysis on 'omics scale data** - Analyzing large datasets currently is the domain of experts with access to large computational resources. The platform will provide computational flexibility and increased scientific focus by making cloud computing technologies accessible to scientists; we are "All-in"³ for the cloud⁴.

³ Steve Balmer in [Network World](#)

5. **Understanding the state of a project** – When in doubt, 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 on a particular research project.
6. **Finding the biological context for results** – Scientists spend much time accessing information on particular biological entities in the literature and public databases. The platform should provide easy integration with information from a variety of sources to help scientists understand a particular result.



At the core of the Sage platform is the management of three major types of scientific resources: data sets that contain experimental results or intermediate steps of a computation pipeline, the analysis tools and scripts used to process the data, and network models that represent the underlying biology. The platform will address the problems described above by providing functionality across 4 major functional areas: A Repository that provides distributed access, discovery, and query across various storage mechanisms for these objects, Collaboration tools that allow distributed teams of people to work together more effectively, Cloud compute and storage to deal with the scale of modern scientific data and processing, and workflow that tracks what has been done for a project.

⁴ See also the [Cloud Computing Requirements](#) document on the intranet.

Use Cases

The following constitute the highest-level use cases around the Sage Platform, and are intended to describe how scientists collaborate to solve research problems. It is expected that more granular use cases will be developed to drive particular platform feature areas.

Sage Predictive Genetics Project

This use case describes an idealized version of a typical predictive genetics project as currently run by Sage Bionetworks. In this sort of project, a combination of phenotypic, genetic, and expression data (aka a “globally coherent data set”) are processed by Sage scientists to determine a list of genes that are interesting for some reason. A typical Sage project will run through the following phases to produce a list of “interesting” genes:

- **Data acquisition** – Generation of a dataset, performed outside of Sage
- **Basic data curation** – Basic data integrity and completeness checks of the raw data
- **Statistical QC** – Statistical quality control and corrections / normalizations of data
- **Genomic Analysis** – Statistical approaches to find genes that correlate with phenotypes and expression levels
- **Network construction** – Construction of a correlation or causation network
- **Network analysis** – Identification of a subnetwork or genes of interest in a network
- **Data Mining** – Searching public literature / databases for information on genes of interest
- **Validation** – Experimental validation of a genetic prediction, performed outside of Sage

The pain points for these projects center on the difficulty scientists have tracking data, analysis methodology, and network versions and the sequences in which they are executed over the course of a project. Communication with internal and external collaborators and efficient re-use of prior work are also big issues. A software system that tracks data, analysis, and network versions and the relationships among resources in a project would alleviate much of this pain. For more details, see the [Sage Predictive Genetics Project Use Case](#).

Federation Project

One of the focus areas of the federation is on ensemble methods. This is really just a form of network analysis / construction as defined above: a number of network models are inputs to some statistical methodology that produces an output network. Sage scientists may employ the “Sage Pipeline” described above as one way to analyze the data; other Federation members each have their own existing methodology and computational environment for getting through basic data validation, QC, and prediction of relevant networks or gene lists. Ideally each lab would start to gain exposure to other labs methodology via sharing data and code on Synapse. As volunteers for a scientific collaboration around data and research sharing, federation members should have some willingness to act as test sites for initial platform prototyping. As a collaboration among scientists at 4 different research institutions, the Federation projects are a good initial target for initial Synapse beta users; see the [Federation Project Use Cases](#) for details.

MetaGEO

MetaGEO is an effort lead by Brig Mecham to perform large scale analysis of multiple gene expression datasets hosted by GEO or other public databases. This sort of functionality is provided commercially by companies like NextBio and OncoPrint, but these tools are not necessarily available to researchers outside of large pharma, and in any case the statistical processing of these organizations is not open and extensible by the community. The goals of the project are to:

1. Improve the annotation of acquired data to facilitate finding relevant groups of datasets to analyze for particular purposes (e.g. “all the datasets related to Prostate Cancer”)
2. Improve the automated statistical processing of raw data to get cleaner signals of gene expression under various conditions
3. Identify and remove / correct poor quality data, e.g. bad probes, bad arrays, batch effects via manual inspection and validation of automatically flagged problems
4. Do all the above in a open source, open access environment where MetaGEO can be a resource for the community to use and extend.

See <http://sagebionetworks.jira.com/wiki/display/METAGEO/Home> for more details.

Sage Commons

The ultimate goal of the platform is to support the Sage Commons: a distributed community of academic, non-profit, and corporate researchers working in an open-access environment to advance scientific knowledge and develop treatments for diseases. In this type of environment, the individual types of activities outlined in the Sage pipeline are a representative, but incomplete, subset of the activities that occur. As an example, one could envision extending the activities supported forward into data validation, or backwards into the planning and capture of a globally coherent dataset. If you look at biological research as a whole, compared to research at Sage, some of the main differences are:

- Globally, there are many more biologists than computational specialists. Most of these biologists will lack even the familiarity with computational techniques that the biologists at Sage have. Simple data mining, visualization, and analysis techniques become increasingly important, ease of use becomes more important, and supporting wide flexibility in analysis methodology becomes less important. Commercial products like NextBio and Ingenuity provide examples of interfaces designed to allow biologists to ask relatively straightforward questions about biology and answer them using data integrated from a variety of sources using predefined methodologies. More advanced questions could be asked using specialist bioinformatics tools like Cytoscape or a Genome Browser which could be integrated into the platform.
- The scale of the Commons in terms of users and data is many orders of magnitude beyond what Sage Bionetworks would generate based on its own research needs. For example, Sage will be submitting a TR01 grant proposal to the NIH in which Sage would act as a neutral 3rd party that would host data captured from the comparator arms of clinical trials. The initial grant proposal would need to identify 3-20 clinical studies which contain clinical data and at least genetic or expression data (both is not yet common) on hundreds to thousands of patients who received either a placebo or standard of care in a stage 2 or 3 clinical study. The end goal of this project

would be to create an environment in which all new clinical studies would have such data released to the public domain (20+ studies / year).

Software Architecture

Overview

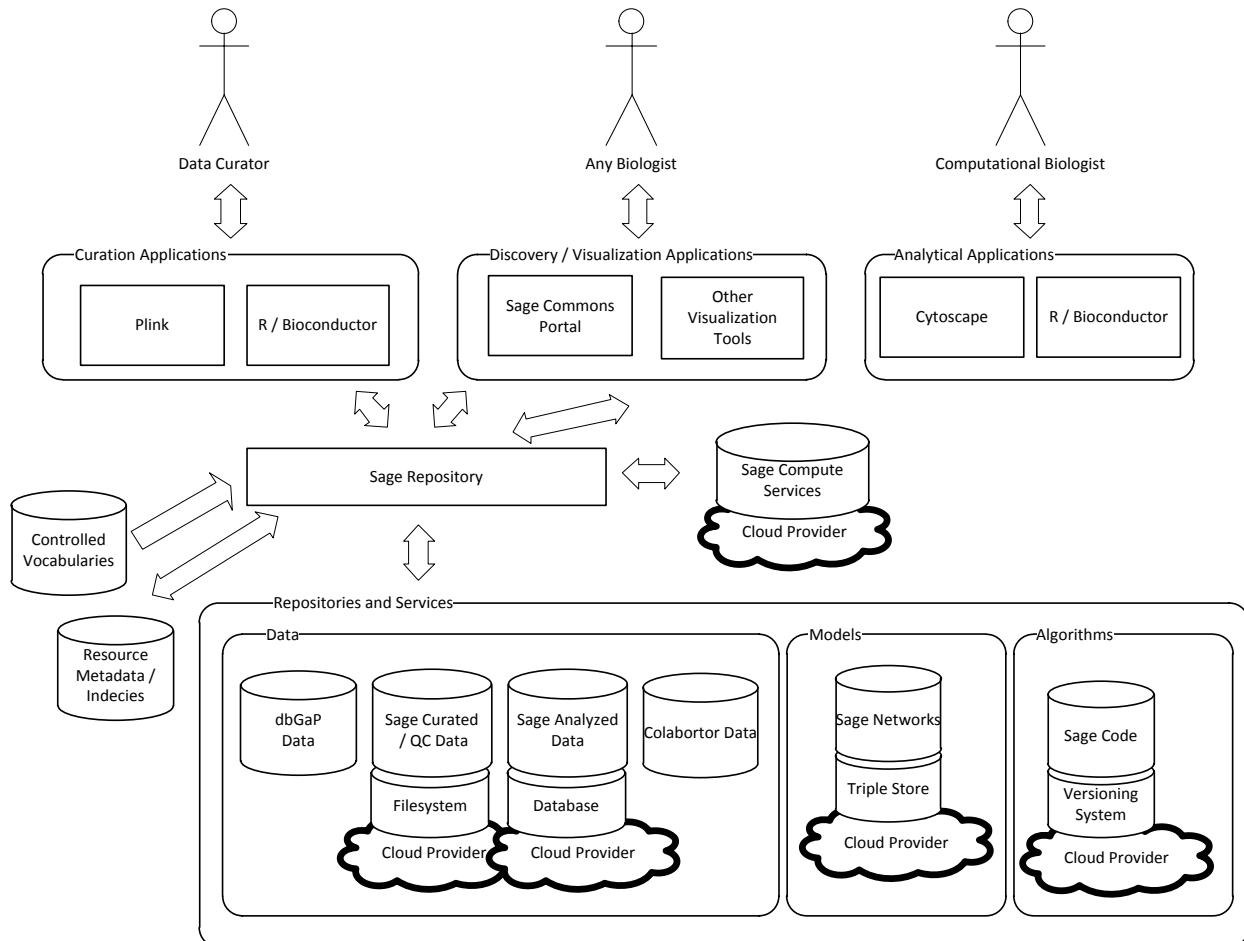


Figure 1: Platform Architecture Overview

A variety of users need to interact with the Sage Platform for different purposes via different sets of applications depending upon their interests and expertise. Data curators and computational biologists are used to working with a combination of dedicated tools (e.g. Plink or Cytoscape) or custom scripts (e.g. R / Bioconductor) to complete their work. These tools will need to be extended to work with platform data in as transparent a manner as possible. A Sage Commons portal will be a “Web 2.0” environment for end user scientists to discover and share data, models, and analysis methods. This new web application would be the point of introduction to the system for new users. Additional capabilities to visualize data and models, or run simple analyses are also of interest to the general biological user, provided they can be delivered through a greatly simplified UI.

This extensible set of applications will need to access a variety of information including experimental data, network models, and algorithms. In some cases, this data could be geographically dispersed for a number of reasons (e.g. human genetic data can only be stored in dbGap). Even if data can be centralized, different types of data will likely need to be stored in different structures. Consequently, a repository services layer is necessary to insulate each application from the specifics of how each resource type is accessed and structured. This service provides a single location to handle annotation, indexing, auditing, security, and versioning.

Recently there has been an accelerating trend in the IT industry towards “cloud computing” environments in which large service providers (e.g. Amazon, Microsoft, and Google) provide on-demand access over the internet to shared pools of computing resources that can be provisioned, used, and released as needed. In addition to becoming an increasingly cost-effective strategy to provide compute and storage resources, cloud computing puts many basic IT management tasks (e.g. maintaining hardware, patching software, backing up data) into the hands of the cloud provider, thus keeping Sage focused on tasks that require scientific expertise. As a new development effort with little existing legacy, the Sage Platform will look to aggressively leverage and optimize its architecture to take full advantage of these existing, and coming services to provide large scale computation on demand to scientific applications.

At the same time, it will be important to avoid locking in users to particular cloud vendors, or to offer users the ability to conduct analysis locally when dictated by circumstance (e.g. requirements for particular hardware or to extract maximal value from servers already paid for). By taking advantage of abstraction layers like JClouds, Eucalyptus or Nimbus that provide a single interface to local or cloud-based resources, the platform will allow collaborators to move computation and/or data among environments as needed. Note that the volumes of data preclude moving data on demand at analysis time. Rather, each site where computation may be located will need a local copy of the data, and the repository services will need to insure that requests for data resolve locally when appropriate and synchronize data sets in the background; see also the [Cloud Computing Requirements](#) document on the intranet.

Sage Web Portal

The Sage Web Portal is envisioned as one of likely many client applications that use Sage Service. The Sage Portal will be a “Web 2.0”⁵ environment for end user scientists to interact and share data, models, and analysis methods, both in the context of specific research projects, and broadly across otherwise disparate projects. General web-based collaboration / project management functionality is available from many large software vendors, and we expect to integrate with existing products in this area, see [Project Collaboration Requirements](#). The Sage Web Portal should provide:

⁵ I know, one of the most overloaded and useless marketing terms ever coined. For me, Web 1.0 means one set of people publishing content on the web using some publishing tool, and other people reading the content using a browser. Web 2.0 means end users, using the browser itself, contribute to the content seen by other users.

- Ways to easily search and navigate through content relevant to their research interests. This will include a combination of browsing, free text search, and structured queries to find relevant resources
- Probably some facilities for connecting people with overlapping research interests. The web portal is not intended to be a social networking site, however integrations with social networking sites like Linked-in or Facebook could potentially be interesting.
- A framework for integrating and organizing the UI around a variety of tools added into the framework, conceptually similar to the way Facebook allows external developers to embed “apps” within Facebook’s pages that access Facebook data.

Users will require an easy way to visualize networks stored in Sage through a client application that accesses Sage Platform network data. Export to / integration of Sage services with Cytoscape or other off-line tools is one option, but requires users to install and learn to use that particular tool. It could be useful to integrate a web-based network viewer that could provide basic network visualization to a broad group of users. This would not be a replacement for an expert-user tool, but a way to get very basic functionality to the masses (and good eye candy for fund-raising). There has recently been a release of a [Cytoscape web version](#) that would be a first choice for this purpose.

A common follow-up activity to data analysis is mining the existing literature and a variety of public databases for more information on genes of interest uncovered by the analysis. In this area, there are a large number of potential commercial and open source solutions available to researchers including NextBio, Ingenuity, tranSMART, Gene Atlas etc. Replicating this functionality is outside the scope of Synapse. Synapse’s core focus is supporting data analysis, and we will look to integrate with one or more of these sorts of platforms for follow up knowledge management and data mining.

Sage Repository Services

The variety of resource types managed by the Sage platform will require their own dedicated storage: analysis code will need to be in a software versioning system, raw experimental data will be stored as flat files, analyzed data could be in a relational database, and network data could be in a triple store. Data may be additionally fragmented geographically due to a variety of non-technical requirements (e.g. dbGaP requires that human genetic data be shared via their own servers). Due to the pace of change of scientific data and requirements, there is a need to insulate calling applications from the specific details of how resources are discovered, identified and accessed. The Sage repository will provide a generic API for working with resources which may be biological data, networks, or algorithms via a URL following linked data principles⁶. The benefit of this approach is the reuse of these services across a number of potentially growing classes of resources. This layer provides a single place for a variety of general-purpose platform features:

- **Annotation** – Facilities for managing resource metadata associated with Sage resources that describe their structure and context. This includes leveraging controlled vocabularies or

⁶ <http://linkeddata.org/>

ontologies to ensure consistency across different resources. Where appropriate, emerging standards like CDISC for clinical data will be leveraged.

- **Search / Indexing** - Both structured and unstructured federated query mechanisms to find resources via indexes created by this layer⁷.
- **Auditing** – A recording of the history of who did what to produce a particular resource, resulting in a high level history / work flow for projects run on the platform.
- **Security** – Resource level control on user level access and guest level access that can evolve over time to reflect the changing nature of resource availability with project life cycle. This layer will leverage emerging standards (e.g. Open ID) to manage access to platform resources.
- **Versioning** – Object level version history, with relationships between resources tracked at the version level. There will also need to be a distinction between published resources ready to be reused and work in progress.

In collaboration with the Institute for Systems Biology, the platform team is evaluating if the Addama software package⁸ can provide some of the functionally needed by this component.

Sage Data Services

These services provide means for working with “globally coherent datasets” as described above that contain a mix of clinical / phenotypic, genetic, and expression data on a population of individuals. Practically speaking, some datasets may be missing some of this information for a variety of reasons.

Sage data services should provide the following:

- **Data Upload** – Some sort of curation procedures will need to be enforced to bring disparate datasets into some sort of consistent structure with well-structured metadata, and ensure general integrity of the dataset.
- **Data Download** – Will need to provide means of accessing the entire dataset. However, downloading the entire dataset for analysis should be discouraged as much as possible in favor of uploading analysis programs to the Sage platform to operate on Sage or other linked datasets in the cloud.
- **Query** – Ability to run structured (SQL-like) queries on a datasets and return (paged) results to a client program (visualization or analysis package). Note that it is important that the service not be constrained to query against a particular dataset format, there will likely be some sort of query rewriting mechanism needed to insulate the UI from the raw data schemas.
- **Data Edit** – Curation / QC of datasets will require the integration of dedicated tools for working with structured data from various sources. Representative examples include Freebase

⁷ Although a counter-argument could be made that just “Google searching” free text is good enough, talk to Atul Butte on this point.

⁸ <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2670281/pdf/1471-2105-10-79.pdf>

Gridworks for curation, Tableau for data visualization or Spotfire for guided biological workflows.

Sage may not be the ultimate storage location for many datasets. Instead, it is desirable that certain datasets or parts of datasets live online in other public repositories where appropriate (e.g. NCBI GEO for gene expression data and dbGaP for clinical / genetic data), and that Sage use linked data principles to make data sets from these repositories available when needed⁹.

Sage Network Data Services

These services deal explicitly with managing biological network data. Here the nodes in a network represent genes, proteins, or other biological entities. Nodes may be annotated with a variety of data from various sources. Edges between nodes represent relationships between pairs of nodes. Edges may be typed (e.g. interacts with, binds to, inhibits), directional, and potentially annotated as well (e.g. with a number denoting the strength of a particular interaction).

Key network data services are (in rough order of priority):

- **Import a network** from another tool; many in-house tools are likely sources of networks. May want to import networks from online sources (like?...)
- **Export a network** – likely need to export to Cytoscape, some web-based network viewer. Also need to support passing networks as inputs to further analysis methods.
- **Query** – search a network to find patterns / subnetworks of interest. Could be used to provide basic network statistics as well as more advanced analysis.
- **Integration** – Integration of data from a variety of public sources that provide additional information on the entities in a network
- **Diff** - Compare related networks to find the overlap / differences. Or, compare a set of networks generated by different methods to some “gold standard” to rank different construction methodologies.
- **Merge** - combine two or more networks into a composite summary
- **Edit** – support highly granular individual changes to nodes / edges in a network

Semantic web technologies provide a natural framework for integrating network-centric data that may originate from a variety of sources¹⁰.

Sage Analysis Services

An emerging vision of cloud-computing is that data and the software that operates on it will mostly live on remote servers, accessed by people through a variety of mechanisms¹¹. Instead of scientists moving

⁹ This statement should not be taken to mean that Sage can not cache a *copy* of a remote dataset for performance reasons if needed, just that the platform must track who is the single point of truth for the data.

¹⁰ <http://www.biomedcentral.com/1471-2105/10/S10/S10>

¹¹ In one example, [Tim O'Reilly discusses his cloud computing vision](#). And yes, this is a bit consumer-focused and intentionally forward looking and it might not be reasonable to expect a high-throughput biologics platform to work the same way as your smart phone. But then again, we must heed O'Reilly's advice to build software for the world as it will exist in 5 years, not the world as it exists today.

large amounts of data around to use with local software applications running on local compute resources, the Sage Platform should encourage scientists to upload analysis methods / tools, and run analysis on the Sage Platform. There are a couple reasons for this: one is that the economics of the cloud model are becoming increasingly attractive and it will become increasingly rare for research teams to manage their own custom Linux cluster just so their scientists can do high performance computing¹². An even more important reason is that the fundamental goal of the Sage Platform is to make sharing scientific data really happen. This won't happen by magic, the software must be designed with a "social interface"¹³ that strongly encourages people to reuse other people's work **and then open up the results of their work for others to reuse**. Scientists in general like to gain access to other people's work. Getting a scientist to go to a Sage Platform that contains a dataset (or analysis tool, or network model) that he already wants to work with will be about as hard as knocking a drunk off a tightrope with a baseball bat¹⁴. However, if that scientist is allowed to happily download that dataset, run local analysis, and write his Nature paper without recontributing anything to the Sage platform, there will be a very strong chance that nobody will go to Sage because nothing is in Sage.

Of course running an analysis on Sage must be as easy and flexible as running analysis on a local machine, or it won't happen. The Sage analysis services must provide:

- **Data import / export** into commonly used analysis environments. A scientist should be able to work in R (or Perl, Python, Matlab, other specialty tool etc.) and load Sage data via a single call passing in the URI of the dataset.
- **Easily scale analysis** – mechanisms like increasing the size / numbers of VMs are an option, as well as providing analysis written for parallelization frameworks like MapReduce.
- **Capabilities for managing publication of results** – because there is a lot of experimenting and people probably don't want their colleagues looking at intermediate steps of debugging a new analysis routine. However, it's also pretty important that scientists don't just keep their own work private indefinitely while they go off to write their papers and grants.
- **Ways to distribute cost** – If Sage is a cloud service, it must be paid for somehow. Maybe funding can be found to provide free access to services to the scientific community at large, but there will probably need to be some mechanism for tracking individual's usage of the system, and charging at least heavy users for their loads. This applies to compute cycles, but may actually be even more critical to consider in the context of the storage / IO required to maintain large datasets or network models. Amazon, for example, has [Flexible Payment](#) and [DevPay](#) services that are designed to allow other organizations to build applications on top of Amazon web services, and bill users for their usage (the organization is free to charge any price, or none at all) for their "value-add" services. Use of cloud services opens a number of ways in which the revenue needed to maintain the platform would grow with increasing use.

¹² [Translational bioinformatics in the cloud: an affordable alternative](#) Dudley et al, Genome Medicine 2010, 2:51

¹³ <http://www.joelonsoftware.com/articles/NotJustUsability.html>

¹⁴ I know. But now you're paying enough attention to even read the footnotes.

Release Plan

The team has gone through some rapid prototyping sprints leading up to a successful public demo at the Sage Congress in April 2011. We are now preparing to enter internal alpha testing and further public development after that. See the platform team [Release Plan](#) for details.