

# Sage Bionetworks Mission

Sage Bionetworks is a non-profit organization with a vision to create a “commons” where integrative bionetworks are evolved by contributor scientists with a shared vision to accelerate the elimination of human disease

**Building Disease Maps**

**Data Repository**

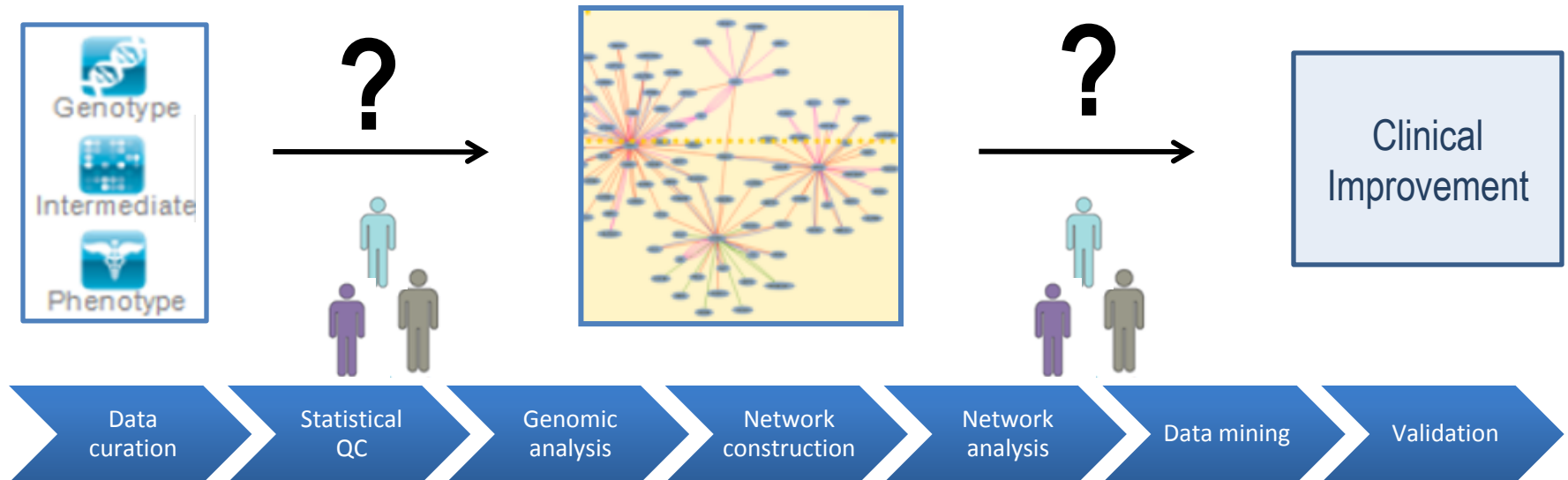


**Commons Pilots**

**Discovery Platform**

Michael Kellen  
Director of Technology

# Need: Community-based analysis is required to build accurate models of disease



## Within the analytical community:

- Quality assessment of data, tools and models
- Reproducibility of models
- Reusability of data
- Validation of models across multiple disease models and patient subtypes

## Across the field of biology:

- Validation of model predictions in experimental/clinical setting
- Accelerated pace towards improved therapeutic development

# Solution: Sage Bionetwork Synapse Project

Watch What I Do, Not What I Say

Reduce, Reuse, Recycle

The screenshot shows the Synapse 'Network Generation Analysis' page. It features a navigation bar with 'Datasets', 'Tools', 'Networks', 'People', and 'Projects'. The main content area is divided into several sections: 'Overview' with a description of the analysis, 'Notes for this analysis' with a list of actions, 'Analysis' with a step-by-step flowchart, and 'Step types' with a list of analysis steps. The flowchart shows a sequence of steps: 'MSKCC Prostate Cancer each expression', 'Expression Normalization v2 by Xuting D. 12-Apr-2011', 'MSKCC Prostate Cancer Phenotypes', 'Survival Analysis v1 by Xuting D. 16-Feb-2011', 'Compression Network Generation v1 by Zhang B. 02-Feb-2011', and 'Sample Matching v1 by Xuting D. 20-Jan-2011'.

Real-time tools to track data analysis projects

Most of the People You Need to Work with Don't Work with You

The screenshot shows a Synapse profile page for Charles Sawyers, Howard Hughes Medical Institute. The profile includes a photo, name, affiliation, and a detailed biography. The biography states: 'Dr. Sawyers is investigating the signaling pathways that drive the growth of cancer cells, with an eye toward designing new treatment options for patients with chronic myeloid leukemia, prostate cancer, and glioblastoma. Our laboratory is focused on characterizing signal transduction pathway abnormalities in various cancers with an eye toward translational implications. One example is chronic myeloid leukemia (CML), which is caused by the BCR-ABL fusion gene, expressed as a consequence of the Philadelphia chromosome translocation.' The page also lists 'Followed Projects' and 'Datasets'.

An online community to connect people and projects

The screenshot shows the Synapse 'All Datasets' page. It features a navigation bar with 'Datasets', 'Tools', 'Networks', 'People', and 'Projects'. The main content area includes a 'Filter List' with columns for 'Dataset Name', 'Layers', 'Number of Samples', 'Status', 'Species', 'Tumor / Tissue Type', 'Disease', 'Investigator', and 'Created On'. The table lists various datasets such as 'MSKCC Prostate Cancer', 'Mouse Model of Diet-Induced Adiposclerosis', 'Human Liver Cohort', 'Hippocampus, Lung, Liver', 'Cancer Cell Line Panel', 'Human Brain Tissue Resource Center', 'Mouse Model of Blood Pressure', and 'LIFEspan PHASE1'.

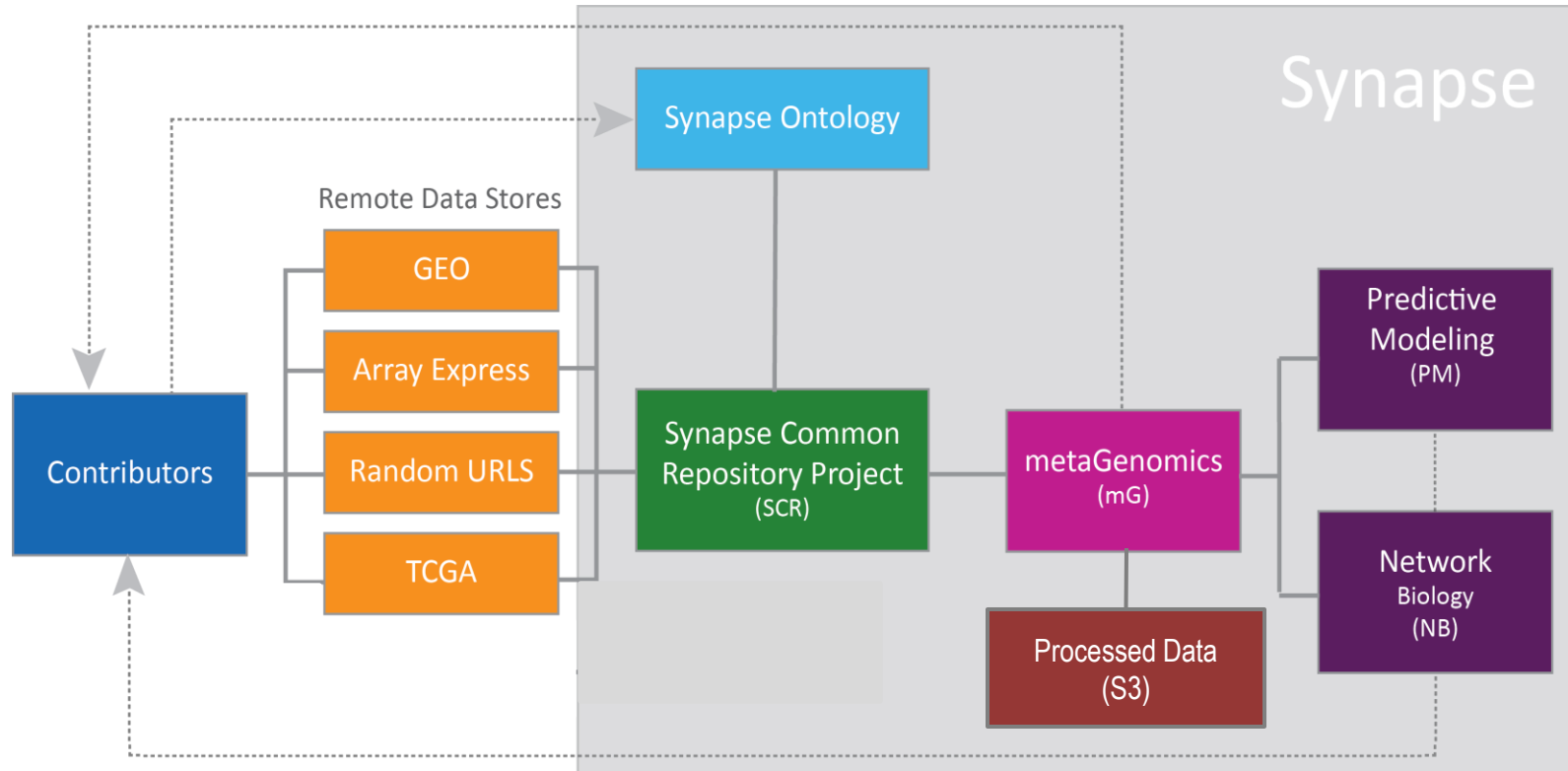
A Repository of Data, Tools, and Models

My Other Computer is Amazon

The screenshot shows the Synapse 'Welcome to SYNAPSE' page. It features a navigation bar with 'Datasets', 'Tools', 'Networks', 'People', and 'Projects'. The main content area includes a 'Welcome to SYNAPSE' section, a 'Featured PROJECT' section, a 'Top Synapse CONTRIBUTORS' section, and a 'Synapse NEWS' section. The 'Featured PROJECT' section highlights the 'The Federation Aging project'. The 'Top Synapse CONTRIBUTORS' section lists Charles Sawyers, Amanda Myers, and Health CareWorks. The 'Synapse NEWS' section includes a link to a new analysis method for remodeling DNA methylation.

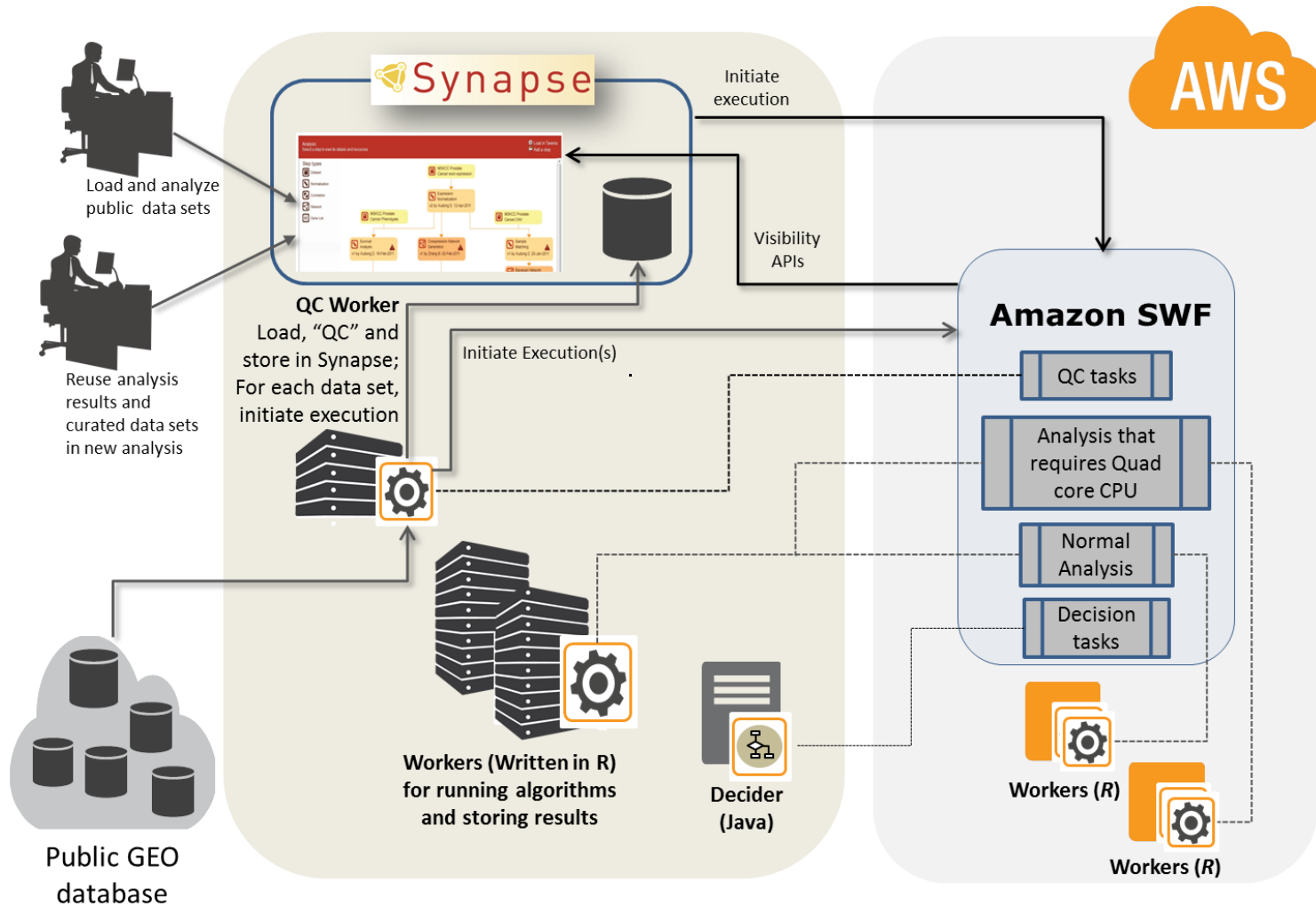
Cloud-based data stores and compute resources 3

# Sage Metagenomics Project



- > 12K genomic and expression standardized datasets indexed in SCR
- Error detection, normalization in mG
- Access raw or processed data via download or API in downstream analysis
- Building towards open, continuous community curation

# Sage Metagenomics using Amazon Simple Workflow



# Use of Ontologies in Synapse

- Synapse UI helps users annotate data / other objects in system with right concepts
- Automated data ingestion pipelines autogenerate appropriate annotations
- Ontologies drive faceted search
- Future development: consistency of clinical trail variables across studies
- Future development: integration with external indices

# Ontologies in Synapse

## Ontology Benefits

- + Consistency of term use across many data sets facilitates large-scale meta analysis (e.g. finding gene expression signatures across all GEO prostate cancer studies)
- + Users can find public data to integrate with their own private studies.
- + Increase ability to integrate across systems (future)

## Ontology Difficulties

- End users are the domain experts, but can't articulate everything they need up front
- End users can not understand complex ontological constructs e.g. OWL class restrictions / upper ontology schemes
- Many ontologies of dubious quality: incomplete term coverage; definition of concepts not mapping to terms as they are used in practice by people.

# Recommendations

- ⇒ Expect iterative refinement of ontology and system using it (we are still early in process)
- ⇒ Expose users to SKOS semantics; use more complicated constructs only on back end if necessary.
- ⇒ Service-oriented architectures allow separation of code and ontology updates.
- ⇒ Need to establish feedback mechanisms from end users to ontology designers directly from point of use.