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
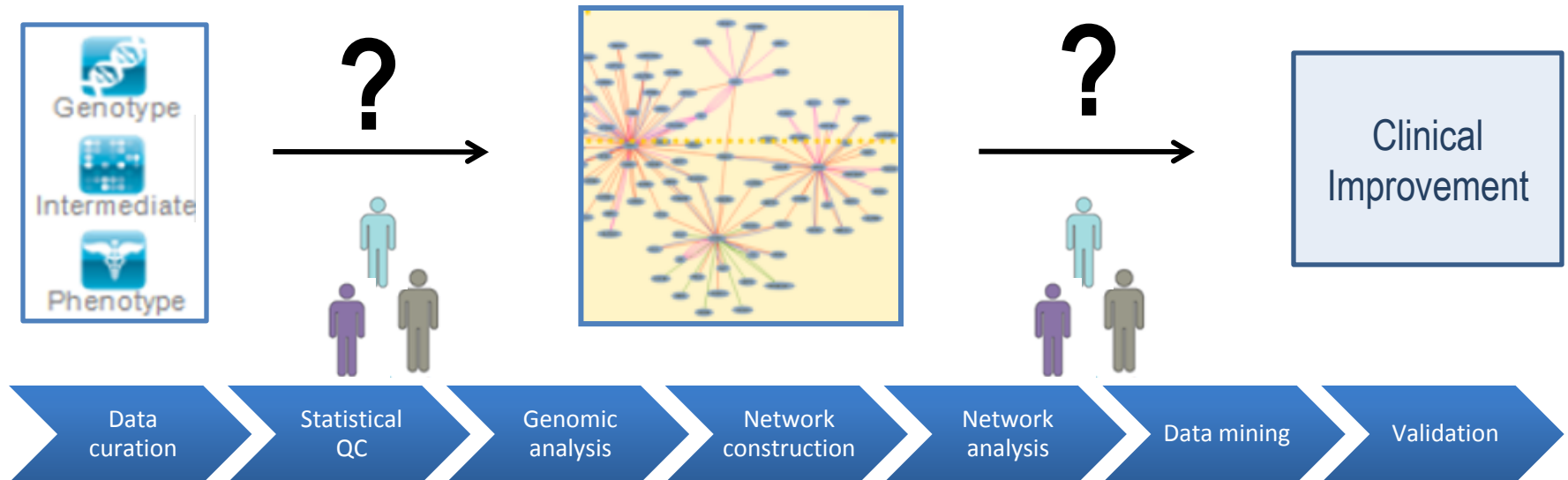
Building Disease Maps

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

Real-time tools to track data analysis projects

Reduce, Reuse, Recycle

A Repository of Data, Tools, and Models

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

An online community to connect people and projects

My Other Computer is Amazon

Cloud-based data stores and compute resources 3

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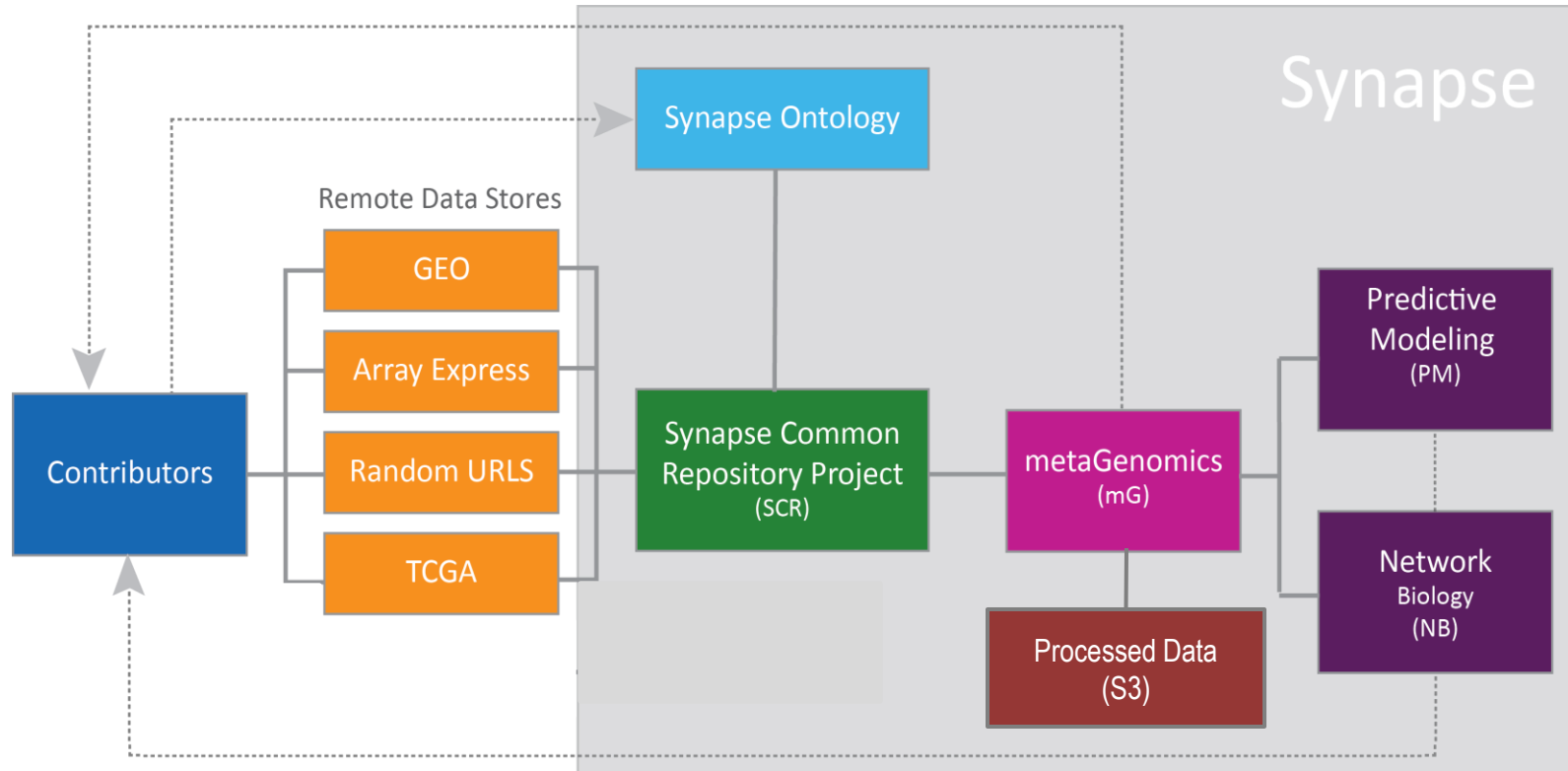
- > 12K genomic and expression standardized datasets indexed in SCR
- Error detection, normalization in mG
- Access raw or processed data via download or API in downstream analysis
- Building towards open, continuous community curation

Sage Metagenomics using Amazon Simple Workflow
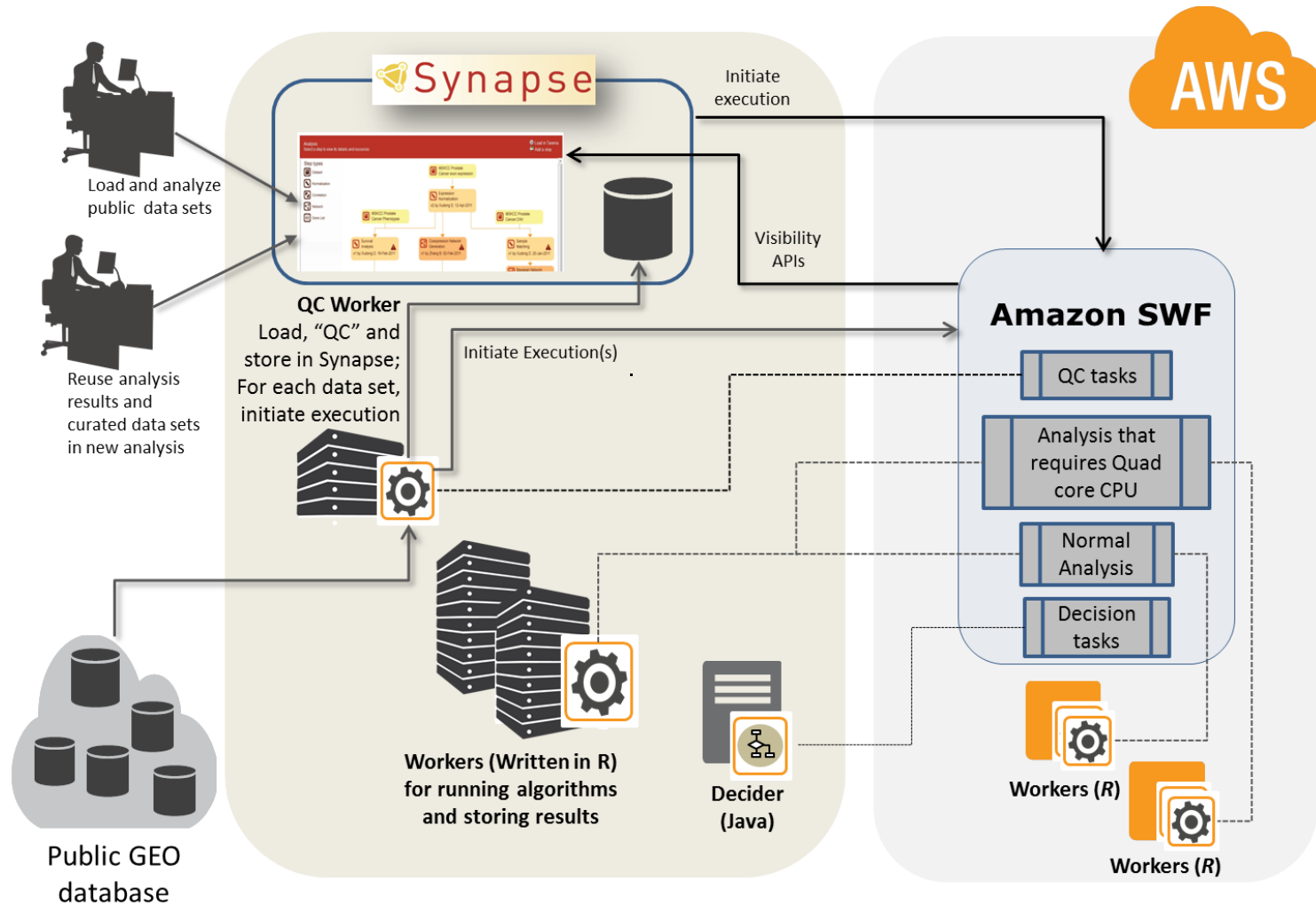
Use of Ontologies in Synapse

- 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

$\Rightarrow$ Expect iterative refinement of ontology and system using it (we are still early in process)

$\Rightarrow$ Expose users to SKOS semantics; use more complicated constructs only on back end if necessary.

$\Rightarrow$ Service-oriented architectures allow separation of code and ontology updates.

$\Rightarrow$ Need to establish feedback mechanisms from end users to ontology designers directly from point of use.