

# Data Infrastructure Issues

## Questions:

- What scientific questions are we working to answer?
- How are we answering similar questions today?
- How is/will data be collected and analyzed?



## Issues:

- **Scientific Workflow** – Working in collaboration with scientists to understand and define workflow, current tools, and how scientist approaches problems
- **Organization & Structure of Data** – Leverage from existing GTL projects, and ongoing biology programs in the community to start the definition phase now and work towards solutions that are capable of evolving
- **Quality/Pedigree** – Define a structured set of automatic quality assessments and metadata that will provide confidence and support -> moving towards validation
- **Interoperability** – Transparency is an evolutionary process. Publish ontologies/schemas to support interoperability
- **Data Overload** – Real/time processing/summary/mining tools. Interfaces that provide easy access to data, computing, and analytical tools
- **Multi-Scale, Multi-Disciplinary Knowledge Transformation** – Integrating data, computational results, information, and analytical tools across multiple scale and domains
- **Collaborative Technologies** – Interconnecting people, facilities and information into a virtual presence