Challenges of a Sustainable Software Platform for Predictive Biology: Lessons Learned on the KBase Project

Dan Gunter
LBNL
What is KBase?

Open software and data platform for addressing the grand challenge of systems biology: Predicting and designing biological function

Unified system that integrates data and analytical tools for comparative functional genomics of microbes, plants, and their communities

Collaborative environment for sharing methods and results and placing those results in the context of knowledge in the field

9/27/15
Dan Gunter, Berkeley Lab
What do we need to build?

"Reference" data

Users work on the system

Knowledge base built from users' public results

Doe Systems Biology Knowledgebase

Lead PI: Adam Arkin (LBNL)
Multi-institutional project

Lawrence Berkeley National Laboratory

Argonne National Laboratory

Brookhaven National Laboratory

Oak Ridge National Laboratory

Cold Spring Harbor Laboratory

University of California, Davis

Hope College

University of Illinois at Urbana-Champaign

Yale University

University of Tennessee

9/27/15

Dan Gunter, Berkeley Lab
Challenge: pull together to make a single, maintainable system
What doesn't work

Working in isolation

Hacking what's there vs. Design

No standards for software engineering
Umm.. why think that *would* work?

- It (mostly) works for C.S. research
  - avoiding long design cycles, not standardizing
    `
    save time
    ` for the `highly rewarded`
    activities like publishing, doing performance studies, graduating
    students. Isolation can help "protect" ideas before they are published.

- Going from research groups to a cohesive
  software team is a `deep`
  shift
What we are trying:

(1) "Agile" for large, distributed science teams

Pick "new" sprints and teams

Two-week "Sprint" x 4-5 Teams

1-week Evaluation

Executive Committee

Github x many repos

JIRA

9/27/15

Dan Gunter, Berkeley Lab
1. **Openness of process**

   https://github.com/kbase/project_guides

   Index of documents:
   - Process guide
   - Acknowledgment guidelines
   - External data sources
   - How to add a method
   - How to add a new data type
   - Developer deployment
   - Setting up Travis CI and Coveralls
   - ThirdPartyDevelopmentProtocol
   - Towards a Social Platform
   - Feature requests
   - Help Desk Process

9/27/15

Dan Gunter, Berkeley Lab