KBase: The Systems Biology Knowledgebase for Predictive Biological and Environmental Research in an Integrated Data Platform

Adam P. Arkin[1], **Robert Cottingham[3]**, **Chris Henry[2]**, Benjamin Allen*[3] (allenbh@ornl.gov), Jason Baumohl[1], Jay Bolton[1], Shane Canon[1], Stephen Chan[1], John-Marc Chandonia[1], Dylan Chivian[1], Mikayla Clark[3], Paramvir Dehal[1], Meghan Drake[3], Janaka N. Edirisinghe[2], José P. Faria[2],, Annette Greiner[1], Tianhao Gu[2], James Jeffryes[2], Marcin P. Joachimiak[1], Roy Kamimura[1], Keith Keller[1], Vivek Kumar[5], Sunita Kumari[5], Miriam Land[3], Sebastian Le Bras[1], Zhenyuan Lu[5], Akiyo Marukawa[1], Sean McCorkle[4], Cheyenne Nelson[1], Dan Murphy-Olson[2],, Erik Pearson[1], Gavin Price[1], Priya Ranjan[3], William Riehl[1], Boris Sadkhin[2], Samuel Seaver[2], Alan Seleman[2], Gwyenth Terry[1], James Thomason[5], Doreen Ware[5], Pamela Weisenhorn[2], Elisha Wood-Charlson[1], Shinjae Yoo[4], Qizhi Zhang[2], Diane Zheng[1]

[1]Lawrence Berkeley National Laboratory, Berkeley, CA; [2]Argonne National Laboratory, Argonne, IL; [3]Oak Ridge National Laboratory, Oak Ridge, TN; [4] Brookhaven National Laboratory, Upton, NY; [5] Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

http://kbase.us

Project Goals: The Department of Energy Systems Biology Knowledgebase (KBase) is a knowledge creation and discovery environment designed for both biologists and bioinformaticians. KBase integrates a large variety of data and analysis tools, from DOE and other public services, into an easy-to-use platform that leverages scalable computing infrastructure to perform sophisticated systems biology analyses. KBase is a freely available and developer extensible platform that enables scientists to analyze their own data within the context of public data and share their findings across the system.

The U.S. Department of Energy (DOE) supports biological and environmental research to investigate the complex interactions within biological systems and the processes that shape soil, water, and ecological dynamics of our biosphere. KBase is a an open-source software and data platform funded by DOE to enable sharing, integration, and analysis of many types of data associated with microbes, plants, and their communities using scalable computing infrastructure. This extensive community resource facilitates large-scale analyses of biological systems and is designed to accelerate scientific discovery, improve reproducibility, and foster open collaboration.

KBase offers a suite of scientific applications to enable users to build sophisticated analytical workflows and share their findings. The nearly 200 apps in KBase offer diverse scientific functionality across the realms of comparative genomics, communities analysis, metabolic modeling, and transcriptomics. Additionally, a number of tools and services within KBase have been co-developed with developers at the DOE Joint Genome Institute. Users can build and share sophisticated workflows through a combination of chaining together multiple analysis tools, writing scripts for automation, and using batch processing, all within notebook-style *Narrative*s that contain the employed data and tools. Developers can build, test, register, and

deploy new or existing software as KBase apps using the Software Development Kit, thereby extending the platform's scientific capabilities.

Recently developed features allow for greater organization of collaborative projects and increased depth of discovery within massive datasets. Projects, laboratories, and even whole institutions can organize their users and associated Narratives into a shared *Organization* with multiple permission levels and management features. Additionally, an early version of social feeds informs users of changes happening within their Organizations. Newly added services enable the platform to find and suggest data sets or Narratives that may be of interest to a particular user, based on searching interconnections between the data in KBase. These services will ultimately evolve into knowledge-discovery features, enabling KBase to propose new hypotheses by making connections across the system. KBase is unique in offering these diverse and integrated capabilities to a growing user community that is actively pioneering the use of the platform in their publications.
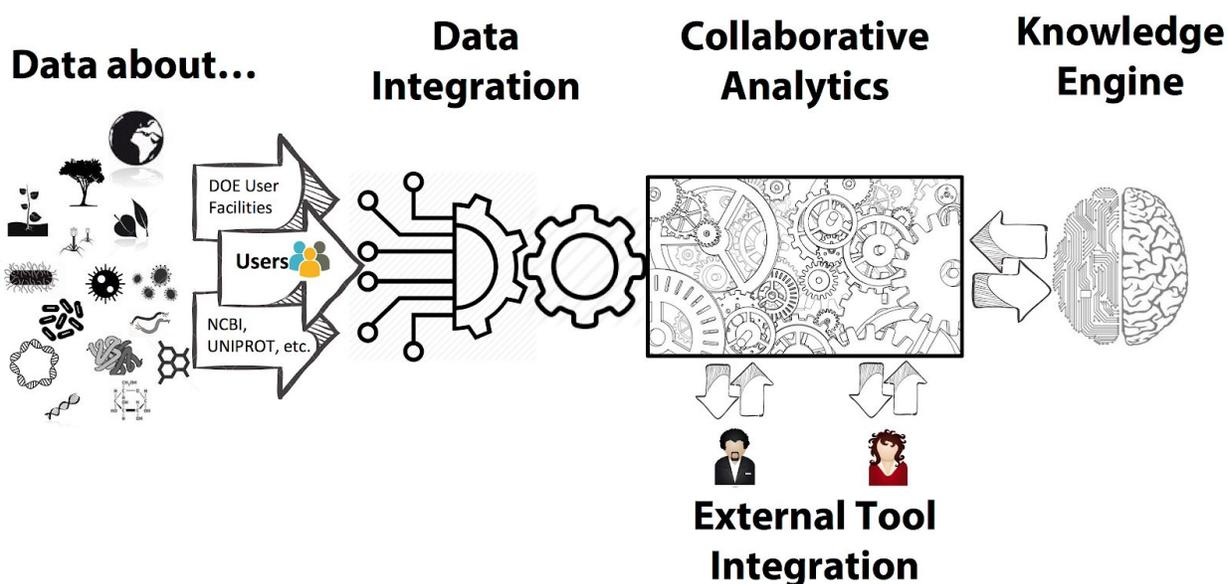

Figure 1. Integration of data and tools into KBase.