

## 167. Developing for KBase

Thomas Brettin\*<sup>1</sup> ([brettin@cels.anl.gov](mailto:brettin@cels.anl.gov)), **Rick Stevens**<sup>1</sup>, Jason Baumohl<sup>2</sup>, Aaron Best<sup>3</sup>, Jared Bischof<sup>1</sup>, Ben Bowen<sup>2</sup>, Tom Brown<sup>1</sup>, Shane Canon<sup>1</sup>, Stephen Chan<sup>2</sup>, John-Marc Chandonia<sup>2</sup>, Dylan Chivian<sup>2</sup>, Ric Colasanti<sup>1</sup>, Neal Conrad<sup>1</sup>, Brian Davison<sup>4</sup>, Matt DeJongh<sup>3</sup>, Paramvir Dehal<sup>2</sup>, Narayan Desai<sup>1</sup>, Scott Devoid<sup>1</sup>, Terry Disz<sup>1</sup>, Meghan Drake<sup>4</sup>, Janaka Edirisinghe<sup>1</sup>, Gang Fang<sup>7</sup>, José Pedro Lopes Faria<sup>1</sup>, Mark Gerstein<sup>7</sup>, Elizabeth M. Glass<sup>1</sup>, Annette Greiner<sup>2</sup>, Dan Gunter<sup>2</sup>, James Gurtowski<sup>6</sup>, Nomi Harris<sup>2</sup>, Travis Harrison<sup>1</sup>, Fei He<sup>5</sup>, Matt Henderson<sup>2</sup>, Chris Henry<sup>1</sup>, Adina Howe<sup>1</sup>, Marcin Joachimiak<sup>2</sup>, Kevin Keegan<sup>1</sup>, Keith Keller<sup>2</sup>, Guruprasad Kora<sup>4</sup>, Sunita Kumari<sup>6</sup>, Miriam Land<sup>4</sup>, Folker Meyer<sup>1</sup>, Steve Moulton<sup>4</sup>, Pavel Novichkov<sup>2</sup>, Taeyun Oh<sup>8</sup>, Gary Olsen<sup>9</sup>, Bob Olson<sup>1</sup>, Dan Olson<sup>1</sup>, Ross Overbeek<sup>1</sup>, Tobias Paczian<sup>1</sup>, Bruce Parrello<sup>1</sup>, Shiran Pasternak<sup>6</sup>, Sarah Poon<sup>2</sup>, Gavin Price<sup>2</sup>, Srivdya Ramakrishnan<sup>6</sup>, Priya Ranjan<sup>4</sup>, Bill Riehl<sup>2</sup>, Pamela Ronald<sup>8</sup>, Michael Schatz<sup>6</sup>, Lynn Schriml<sup>10</sup>, Sam Seaver<sup>1</sup>, Michael W. Sneddon<sup>2</sup>, Roman Sutormin<sup>2</sup>, Mustafa Syed<sup>4</sup>, James Thomason<sup>6</sup>, Nathan Tintle<sup>3</sup>, Will Trimble<sup>1</sup>, Daifeng Wang<sup>7</sup>, Doreen Ware<sup>5,6</sup>, David Weston<sup>4</sup>, Andreas Wilke<sup>1</sup>, Fangfang Xia<sup>1</sup>, Shinjae Yoo<sup>5</sup>, Dantong Yu<sup>5</sup>, **Robert Cottingham**<sup>4</sup>, **Sergei Maslov**<sup>5</sup>, **Adam P. Arkin**<sup>2</sup>

<sup>1</sup>Argonne National Laboratory, Argonne, IL, <sup>2</sup>Lawrence Berkeley National Laboratory, Berkeley, CA, <sup>3</sup>Hope College, Holland, MI, <sup>4</sup>Oak Ridge National Laboratory, Oak Ridge, TN, <sup>5</sup>Brookhaven National Laboratory, Upton, NY, <sup>6</sup>Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, <sup>7</sup>Yale University, New Haven, CT, <sup>8</sup>University of California, Davis, CA, <sup>9</sup>University of Illinois at Champaign-Urbana, Champaign, IL, <sup>10</sup>University of Maryland, College Park, MD

<http://kbase.us>

**Project Goals: The KBase project aims to provide the capabilities needed to address the grand challenge of systems biology: to predict and ultimately design biological function. KBase enables users to collaboratively integrate the array of heterogeneous datasets, analysis tools and workflows needed to achieve a predictive understanding of biological systems. It incorporates functional genomic and metagenomic data for thousands of organisms, and diverse tools for (meta)genomic assembly, annotation, network inference and modeling, allowing researchers to combine diverse lines of evidence to create increasingly accurate models of the physiology and community dynamics of microbes and plants. KBase will soon allow models to be compared to observations and dynamically revised. A new prototype Narrative interface lets users create a reproducible record of the data, computational steps and thought process leading from hypothesis to result in the form of interactive publications.**

One of the key operating principles of KBase is to allow the scientific community to incorporate their own new algorithms into the system to make them available to others easily; to avail themselves of the KBase computational architecture; and to make use of the KBase data sources. The KBase team aims to make this process simple and to provide an easy route for dissemination of new tools and comparison to existing tools in a common framework.

To achieve these goals KBase system design is firmly rooted in our service-oriented architecture and in our commitment to open source development and distribution models. These aspects of the project have been with the system since its inception, and will continue to be a guiding force for future work.

The KBase system design is based on several sound best practice principles including consistent code use, code re-use, and the decoupling of modular system components. We have established standard software engineering processes for version control, software and data builds, testing, QA/QC, deployments and releases. These enable the deployment of a large number of services by a relatively small release engineering team.

To prepare for potential future services contributed by the community, we provide developer training materials on our website as well in hands-on developer training sessions called bootcamps. In the future, we plan to offer a wider range of bootcamps and webinars to target different types of developers and different scientific focus areas. In the meantime, prospective developers and computational biologists can find information about KBase service design at <http://kbase.us/developer-zone/>.

*KBase is funded by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research.*