Updates to the BioCompute Tools and Guidelines for Use

Jonathon Keeney[1], Charles Hadley King[1], Tianyi Wang[1], Raja Mazumder[1], Dennis Dean[2], Phil Webster[2], Saul Acevedo[2], Omar Serang[3], Sam Westreich[3] Luis Santana-Quintero[4] Mark Walderhaug[4]

1: Department of Biochemistry and Molecular Medicine, The George Washington School of Medicine and Health Sciences, Washington, DC, 20052, USA

2: DNAnexus, Mountain View, CA, USA

3: Seven Bridges Genomics, Charlestown, MA, 02129, USA

4: Center for Biologics Evaluation and Research, FDA, Silver Spring, MD 20993

Introduction

Difficulties in documenting computational workflows have long troubled researchers and reviewers. Without guidelines for describing a workflow, definitions and descriptions are usually ad hoc and rarely adequate for comprehension or reproducibility. BioCompute is shorthand for an IEEE standard (2791-2020) meant to provide a framework for describing pipelines in a rigorous but flexible way, and a BioCompute Object (BCO) is an instance of a workflow that validates against that standard. The standard abstracts a pipeline into its conceptual components, and represents those components according to existing data standards and ontologies. BCO has been adopted by three FDA centers to help smooth major delays in processing regulatory submissions.

Throughout the last year, our central repository for BCOs (called the Portal) has undergone major updates to improve functionality and o achieve a more user-friendly experience. On the newly redesigned portal, Google and ORCID authentication are integrated in the form of JSON Web Tokens (JWT). Users can edit, validate and publish the workflow to a public BCO DataBase (BCODB) in JSON format and directly interact with API endpoints. The next step for BioCompute is a pilot project which will involve the BioCompute team, FDA, and industry sponsors to ensure that BCO as a vehicle can increase communication efficiency and fluidity during the submission and review process.

Web Portal

https://biocomputeobject.org

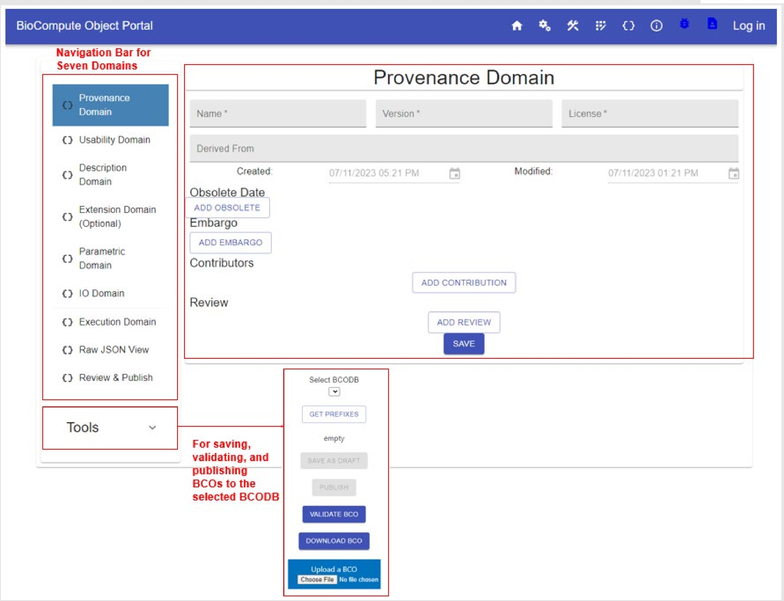


Figure 1: The BCO Builder interface. The web Portal hosts tools for working with BCOs, including a Builder tool for creating BCOs, shown here. Components of a workflow are organized into eight different "domains" in the BCO system, five of which are required, and three are optional. They are Provenance, Usability, Extension, Description, Execution, I/O, Parametric and Error domains. Each domain records relative information of the analysis workflow.

BCOnexus

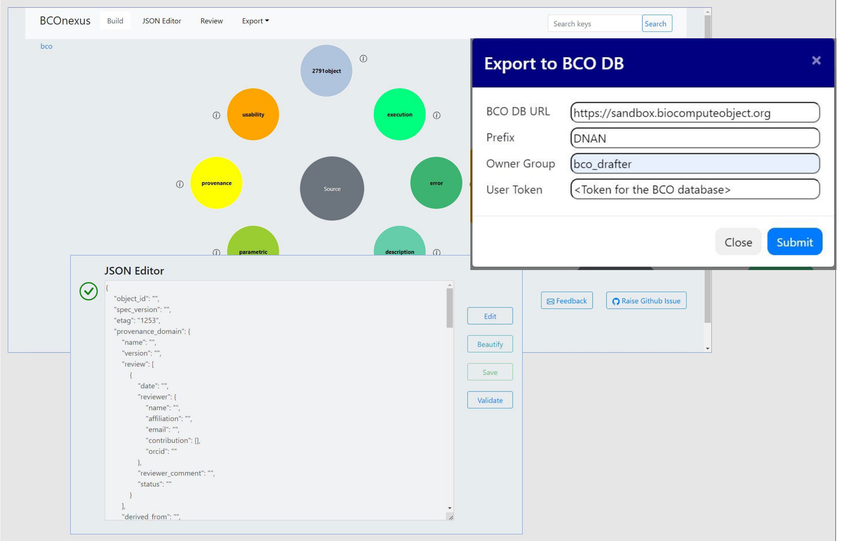


Figure 2: BCOnexus released a platform-free Docker-based schema-driven BCO editor. BCOnexus enables visual creation, exploration, and editing of BCOs. Created to provide schema-driven navigation and manipulation of BCOs, the application now provides a JSON text interface for validation and editing. BCOnexus can export to multiple destinations, including BCO repositories and DNAnexus workflows.

Seven Bridges APP

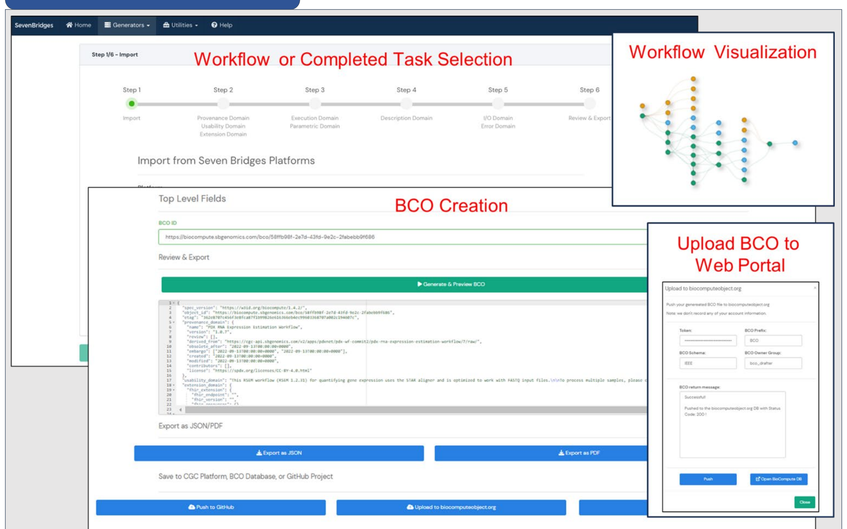
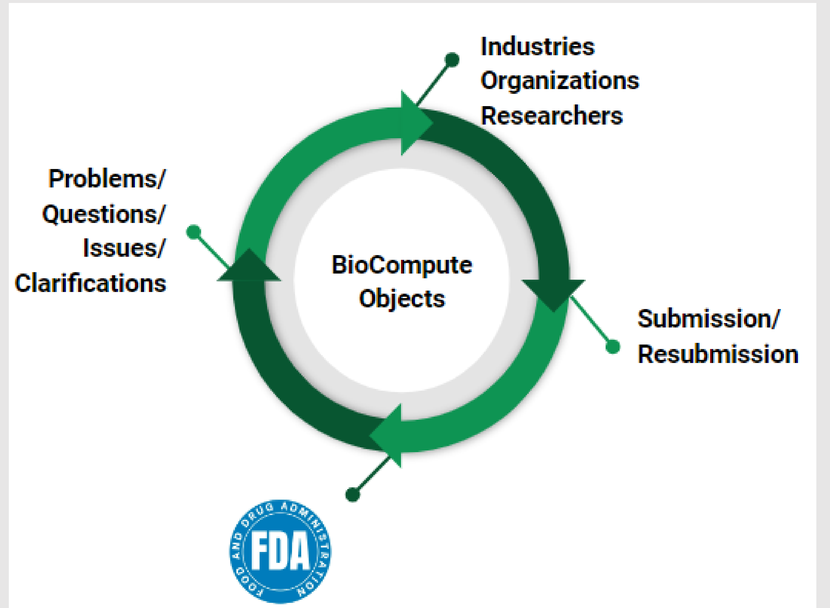


Figure 3: The BCO App supports BCO generation from free text, Common Workflow Laguage, and a completed task. BCO information can be entered from interactive input, uploaded from a local computer, or any Seven Bridges Platforms. The figure shows screen captures of Workflow or Completed Task Selection, Workflow Visualization, BCO Creation, and BCO upload to the Web Portal.

Pilot Project

The pilot project involves the BioCompute team, FDA, and the industry to enhance the adoption of BioCompute and raise community awareness. Ultimately, a best practices document will be generated for creating, submitting, and reviewing BCOs in a regulatory context. This project will use BCO as a vehicle to ensure communication efficiency and fluidity during the submission and review process. The project is currently in the design and recruiting stage.



Discussion

BioCompute plays a important role in harmonizing workflow data, managing complexity across multiple versioned resources, and boosting reproducibility. Relying on ad hoc descriptions often results in a loose description that is assembled as an afterthought, and which may not address important questions. As the field matures, it will be important to be able to address known concerns in regulatory submissions, and for all parties to have a common language to do so. The more than 400 individuals who have participated in the BioCompute community sought to put thought into the framework in a way that provides guidance for addressing these questions, without creating excessive burden on the sponsor or the reviewer.

With Google and ORCID integration, improved user interface, a new Portal instance at the FDA, four tools for directly authoring BCOs from a platform (including the two presented here, and the HIVE and Galaxy platforms), and expanded documentation, there are now many tools online for creating, sharing, editing, and publishing, BCOs. Overall, the project is working towards a smoother and more secure submission process. In support of this, the FDA can pull BCOs from the public BCODB as requested by the sponsor. With APIs and authentication tokens, users can also directly submit to the FDA portal API endpoints, and can easily interact with BCO and BCO implementation.

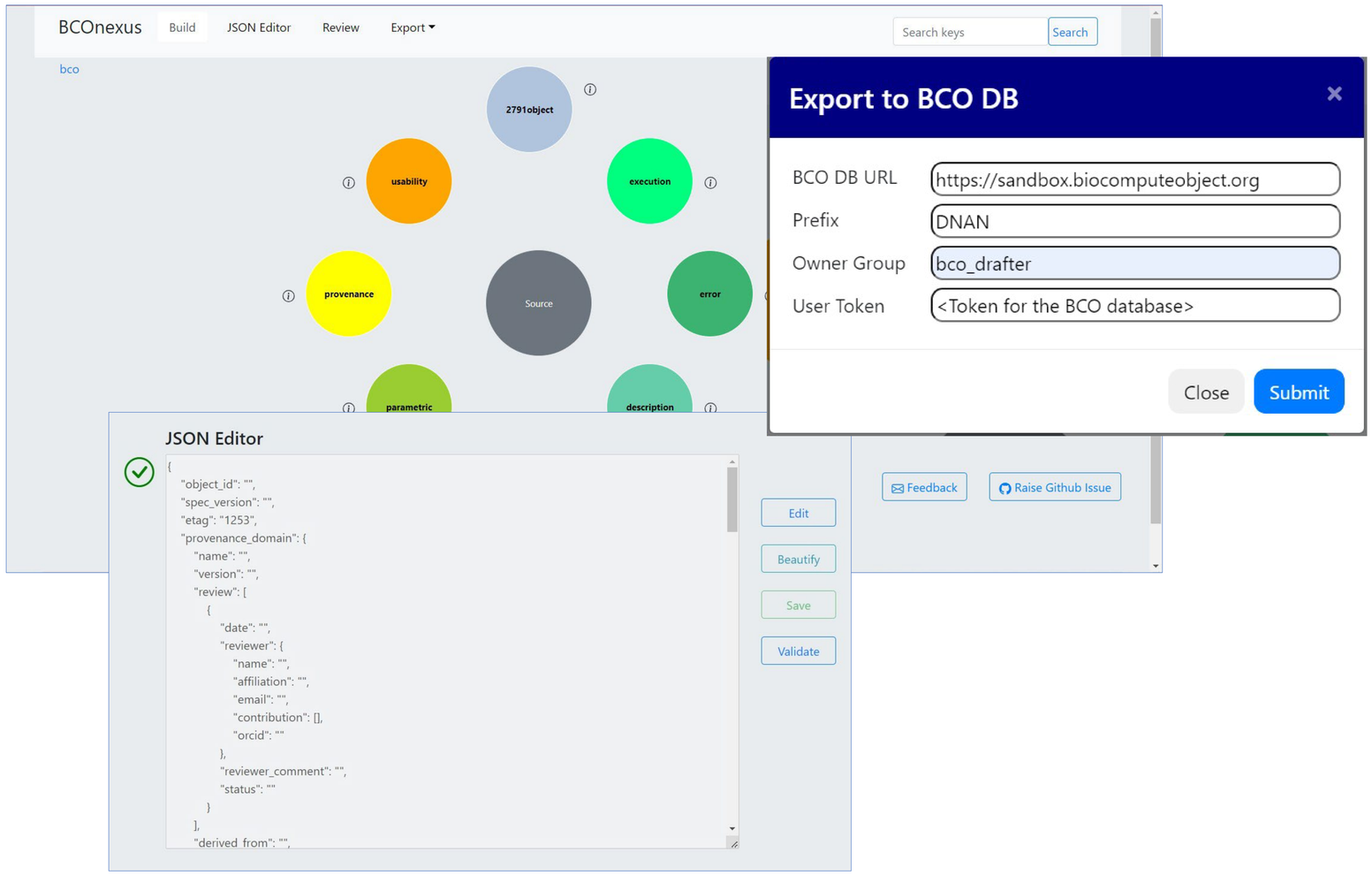


Figure 2: BCOnexus released. A platform-free Docker-based schema-driven BCO editor, BCOnexus enables visual creation, exploration, and editing of BCOs. Created to provide schema-driven navigation and manipulation of BCOs, the application now provides a JSON text interface for validation and editing. BCOnexus can export to multiple destinations including BCO repositories and DNAnexus workflows.