FDA U.S. FOOD & DRUG ADMINISTRATION

Coordination of BioCompute Objects between Seven Bridges, DNAnexus, and the FDA


Jonathon Keeney

Presenter, George Washington University
Contractor at CBER


Omar Serang


Sam Westreich

DNAnexus


Dennis Dean


Phil Webster

Seven Bridges


Konstantinos Karaggiannis


Mark Walderhaug

FDA


Raja Mazumder


Hadley King

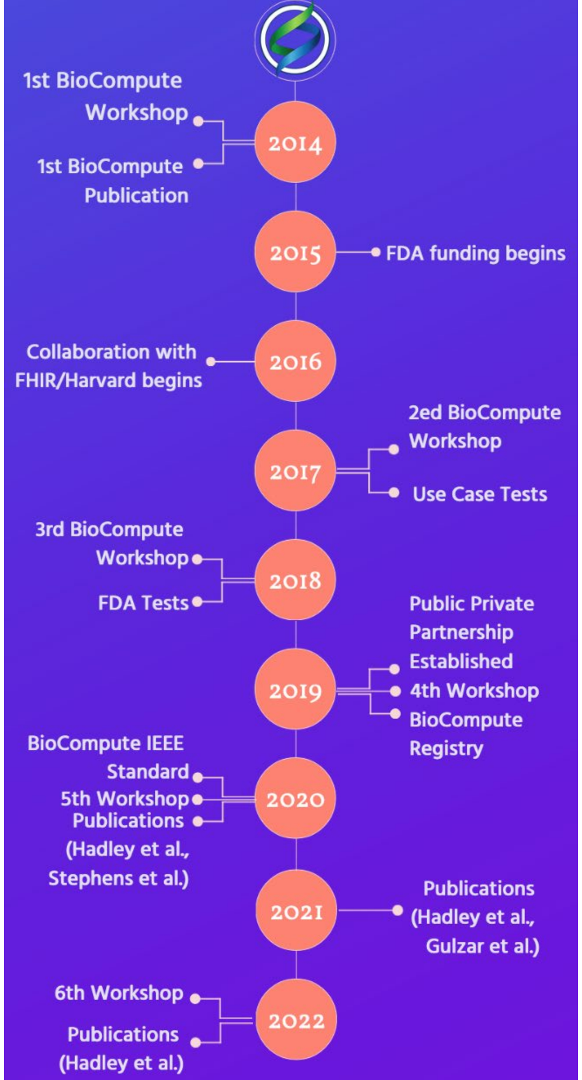

Tianyi Wang

THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC

Others: Anton Golikov

Introduction/Hypothesis - BioCompute

- Working groups initiated in 2014, standardized in 2020
    - Official name: IEEE 2791-2020
- A reporting and documentation protocol for computational pipelines
- A BioCompute Object (BCO) can be combined with workflow languages
    - Existing examples include CWL, WDL, and Nextflow
- Need a secure mechanism to communicate BCOs to FDA from popular bioinformatics platforms



2014
1st BioCompute Workshop
1st BioCompute Publication

2015
FDA funding begins

2016
Collaboration with FHIR/Harvard begins

2017
2ed BioCompute Workshop
Use Case Tests

2018
3rd BioCompute Workshop
FDA Tests

2019
Public Private Partnership Established
4th Workshop
BioCompute Registry

2020
BioCompute IEEE Standard
5th Workshop
Publications (Hadley et al., Stephens et al.)

2021
Publications (Hadley et al., Gulzar et al.)

2022
6th Workshop
Publications (Hadley et al.)

Findings/Results - GWU

- “Platform-free” tools, with support on the Seven Bridges and DNAnexus platforms
- Easily record full pipeline as a BCO
- Interactive with centralized Portal at the FDA
  - Directly upload a BCO

Findings/Results – Seven Bridges

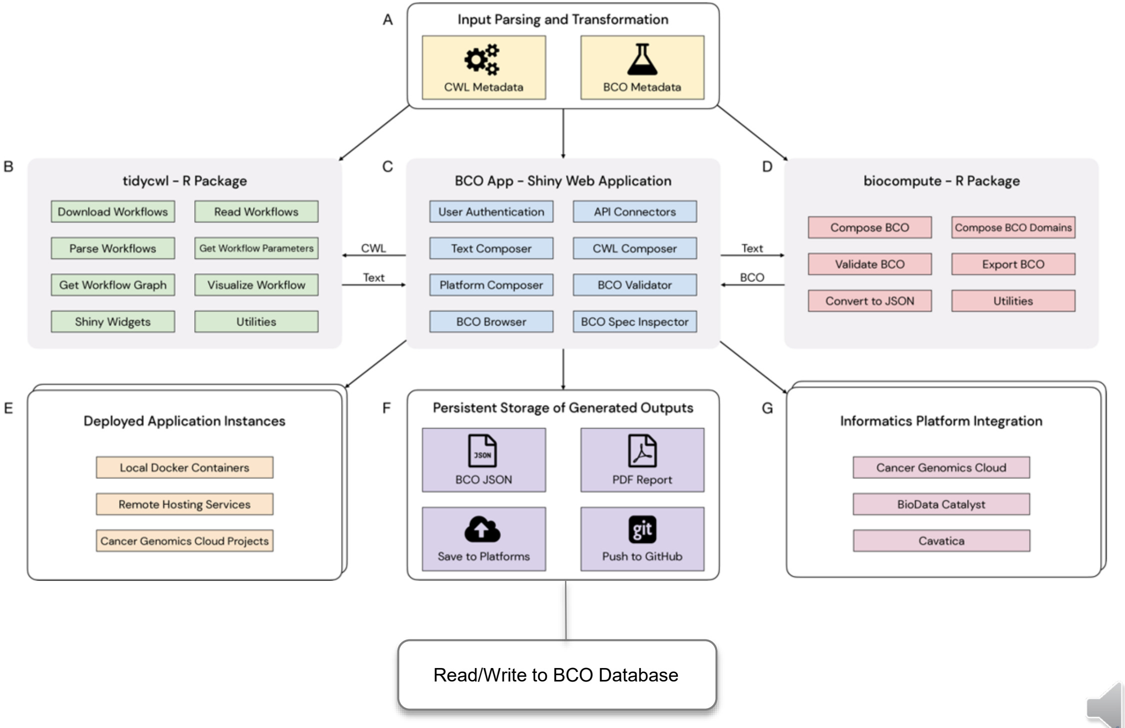
Overview

- Flexible architecture support deployment in multiple environments
- Open source CRAN libraries support collaboration
- Text, CWL, and Task input support automated BCO generation

New Functionality

- BCO Database functionality supports BCO sharing
- BCO Best Practices Documents

Seven Bridges' BCO App Architecture



A
Input Parsing and Transformation
CWL Metadata
BCO Metadata

B
tidycwl – R Package
Download Workflows
Read Workflows
Parse Workflows
Get Workflow Parameters
Get Workflow Graph
Visualize Workflow
Shiny Widgets
Utilities

CWL
Text

C
BCO App – Shiny Web Application
User Authentication
API Connectors
Text Composer
CWL Composer
Platform Composer
BCO Validator
BCO Browser
BCO Spec Inspector

Text
BCO

D
biocompute – R Package
Compose BCO
Compose BCO Domains
Validate BCO
Export BCO
Convert to JSON
Utilities

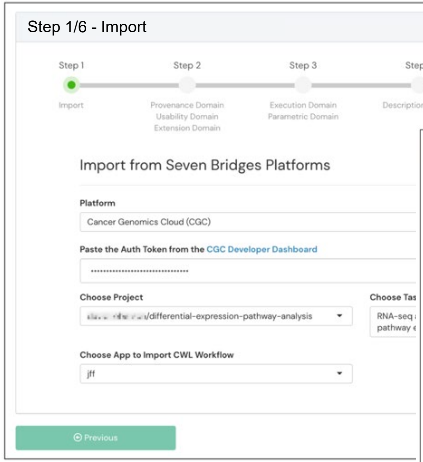
E
Deployed Application Instances
Local Docker Containers
Remote Hosting Services
Cancer Genomics Cloud Projects

F
Persistent Storage of Generated Outputs
BCO JSON
PDF Report
Save to Platforms
Push to GitHub

G
Informatics Platform Integration
Cancer Genomics Cloud
BioData Catalyst
Cavatica

Read/Write to BCO Database

Seven Bridges BCO App Screen Shots



Step 1/6 - Import

Import from Seven Bridges Platforms

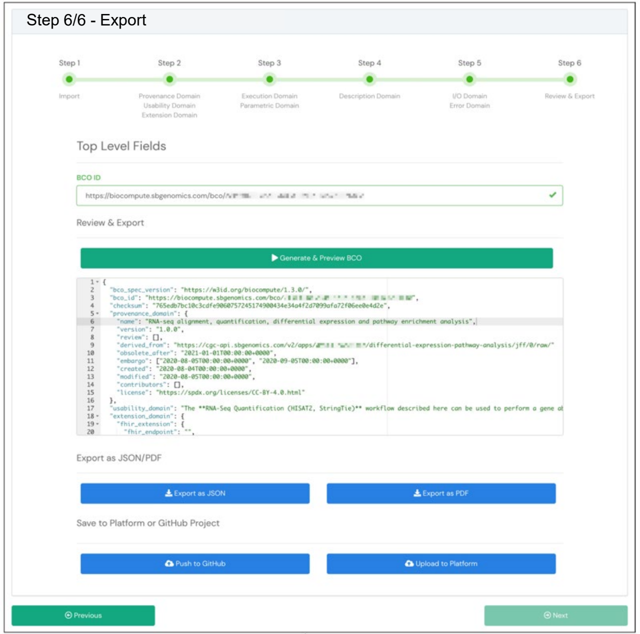
Platform
Cancer Genomics Cloud (CGC)

Paste the Auth Token from the CGC Developer Dashboard
••••••••••••••••••••••••••••••••

Choose Project
[redacted]/differential-expression-pathway-analysis

Choose Tas
RNA-seq
pathway

Choose App to Import CWL Workflow
jff



Step 6/6 - Export

Top Level Fields

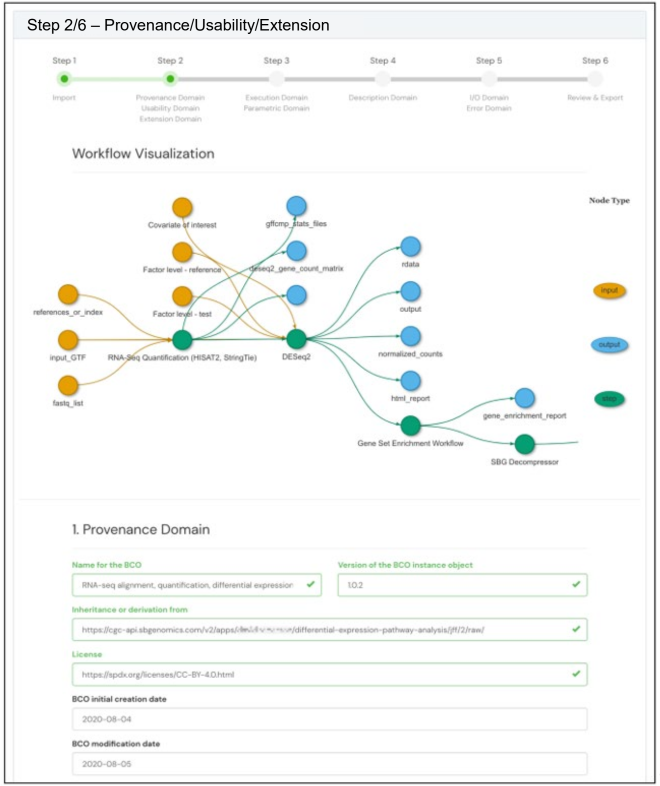
BCO ID
https://biocompute.sbgenomics.com/bco/[redacted] ✓

Review & Export

```
{
  "bco_spec_version": "https://w3id.org/biocompute/1.3.0/",
  "bco_id": "https://biocompute.sbgenomics.com/bco/[redacted]",
  "checksum": "765edb7bc10c3cdfe9060757245174900434e34a4f2d7099afa72f06ee0e4d2e",
  "provenance_domain": {
    "name": "RNA-seq alignment, quantification, differential expression and pathway enrichment analysis",
    "version": "1.0.0",
    "review": [],
    "derived_from": "https://cgc-api.sbgenomics.com/v2/apps/[redacted]/differential-expression-pathway-analysis/jff/0/raw/"
    "obsolete_after": "2021-01-01T00:00:00+0000",
    "embargo": ["2020-08-05T00:00:00+0000", "2020-09-05T00:00:00+0000"],
    "created": "2020-08-04T00:00:00+0000",
    "modified": "2020-08-05T00:00:00+0000",
    "contributors": [],
    "license": "https://spdx.org/licenses/CC-BY-4.0.html"
  },
  "usability_domain": "The **RNA-Seq Quantification (HISAT2, StringTie)** workflow described here can be used to perform a gene at
  "extension_domain": {
    "fhir_extension": {
      "fhir_endpoint": "",
```

Export as JSON/PDF

Save to Platform or GitHub Project



Step 2/6 – Provenance/Usability/Extension

Workflow Visualization

1. Provenance Domain

Name for the BCO
RNA-seq alignment, quantification, differential expression ✓

Version of the BCO instance object
1.0.2 ✓

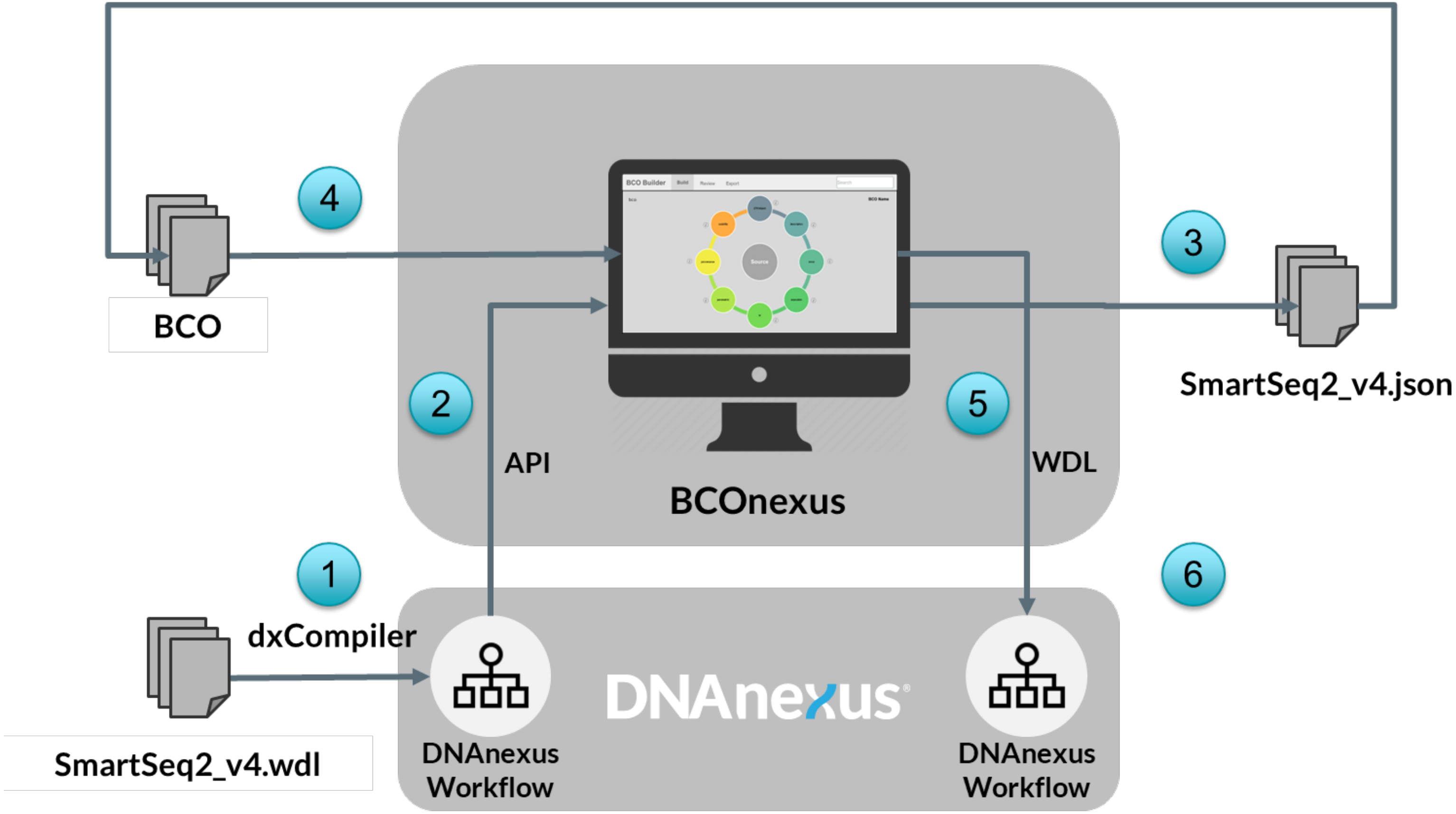
Inheritance or derivation from
https://cgc-api.sbgenomics.com/v2/apps/[redacted]/differential-expression-pathway-analysis/jff/2/raw/ ✓

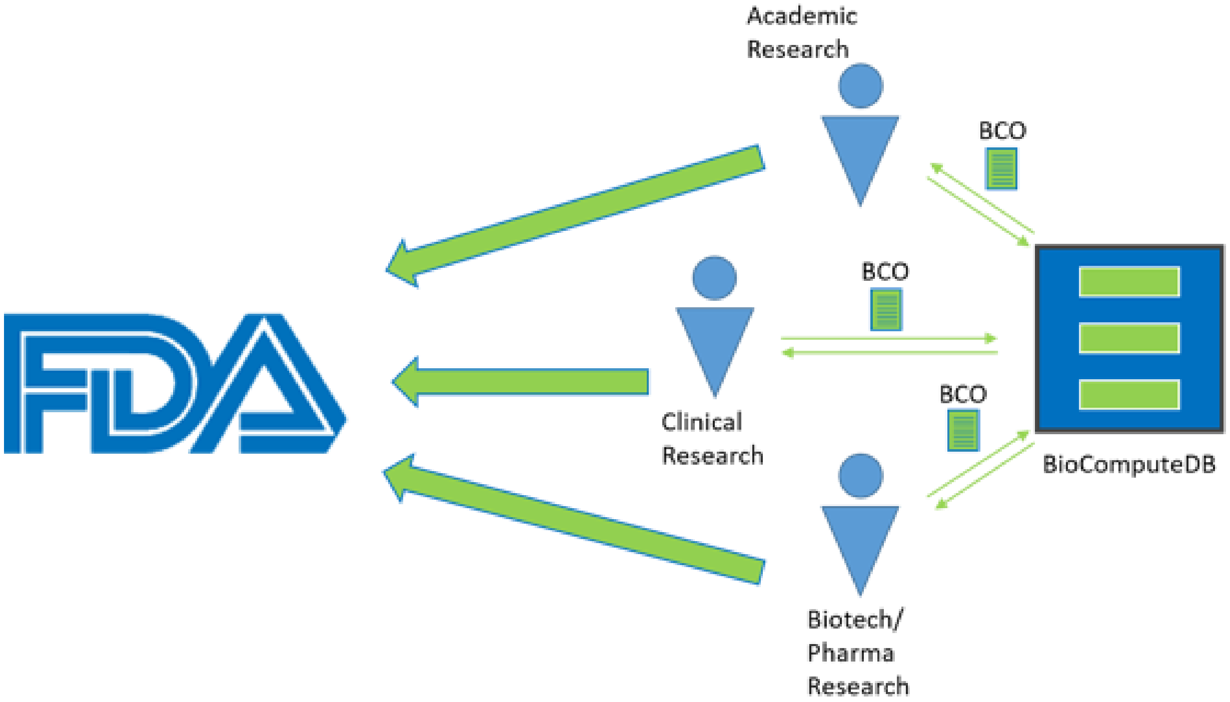
License
https://spdx.org/licenses/CC-BY-4.0.html ✓

BCO initial creation date
2020-08-04

BCO modification date
2020-08-05

BCOnexus SmartSeq2 Workflow




Academic
Research
BCO
Clinical
Research
BCO
BCO
BioComputeDB
Biotech/
Pharma
Research

Conclusion

- Produce and transmit BCOs to the FDA;
- Improve communication of computational analysis pipelines and ease burdens on both reviewers and regulated industry;
- Additional features including aggregation and statistical reporting of BCOs, comparison of two BCOs for discrepancies, and “template” and “run” BCOs that can be versioned and individually certified.
- **Hands on workshop for both platforms:**
  - **Tuesday, September 13th at 10AM**
  - **Virtual event, listed on FDA calendar**

Thank you!

FDA U.S. FOOD & DRUG
ADMINISTRATION