

# Galaxy Workflows to BioCompute Objects

Charles Hadley King, Christopher A. Armstrong, Jonathon Keeney, Janisha Patel, Stephanie S. Singleton, Stephen D. Shank, Raja Mazumder

At GCC 2020 we first presented “Implementation of the IEEE-2791-2020 standard (BioCompute Objects) in Galaxy workflow invocations” (<https://vimeo.com/437704225>) in which a Galaxy invocation and workflow object were used to construct a rough BioCompute Object for download.

Workflow Invocations

Workflow: Galaxy BioComput Object D... BioComput... about 1 month ago about 1 month ago scheduled

Invocation: ebf8f50c6abde6d

BioCompute Object

A BioCompute Object (BCO) is the unofficial name for a JSON object that adheres to the IEEE-2791-2020 standard. A BCO is designed to communicate High-throughput Sequencing (HTS) analysis results, data set creation, data curation, and bioinformatics verification protocols.

Learn more about BioCompute Objects.

Download

At GCC 2021, we presented “Publication of BioCompute Objects (IEEE-2791-2020) created from Galaxy workflow invocations” (<https://vimeo.com/567195992>). Here, we demonstrated how a Galaxy BCO could be deposited to the BioCompute database (BCODB) via API. The BioCompute Object DB APIs are all available via at <https://biocomputeobject.org/api/docs/>.

## BioCompute participants



BCOs were conceived by researchers at the US FDA and The George Washington University. The goal was to ease the burden associated with regulatory review of large high-throughput sequencing projects by offering a standardized reporting template. This standard is supported by the FDA Centers, CDER, CBER, and CFSAN, for regulatory submissions on a voluntary submitter basis. Contributions to the BioCompute community have been made by individuals from NIH, Harvard, several biotech and pharma companies, EMBL-EBI, FDA-H.I.V.E, DANAexus, Seven Bridges, and the Galaxy community.

A BioCompute Object (BCO) is the unofficial name for a JSON object that adheres to the IEEE-2791-2020 standard. A BCO is designed to communicate High-throughput Sequencing (HTS) analysis results, data set creation, data curation, and bioinformatics verification protocols.

When all the pieces are in place for a platform (like Galaxy) users will be able to auto generate and submit a BCO to the BCODB. This will be common place for all these different workflows with a common format.

Resources	Builder	Account	BioCompute Objects	Log Out
BCO Accession	Name	State ↑	Prefix	Last Updated
ARGOS_000018DRAFT	"ngs ID List and Selection Criteria"	DRAFT	ARGOS	2022-06-22T19:43:29.759813Z
BCO_000094DRAFT	"Hadley King"	DRAFT	BCO	2022-03-16T19:19:04.589295Z
GALXY_000001DRAFT	"Example for Intro2Bioinformatics"	DRAFT	GALXY	2022-04-12T17:08:15.799243Z
BCO_000113DRAFT	"seekCRIT: Detecting and characterizing differentially expressed circular RNAs using high-throughput sequencing data"	DRAFT	BCO	2022-04-13T13:55:14.460025Z
ARGOS_000002DRAFT	"Influenza A reference proteome sequences"	DRAFT	ARGOS	2022-06-15T13:40:23.731488Z
ARGOS_000003DRAFT	"Influenza A reference proteins list"	DRAFT	ARGOS	2022-06-15T14:48:24.921924Z
ARGOS_000006DRAFT	"SARS-CoV-2 reference proteins list"	DRAFT	ARGOS	2022-06-21T20:09:52.295794Z
ARGOS_000008DRAFT	"Original FDA-ARGOS BioProject Assembly Metadata from NCBI"	DRAFT	ARGOS	2022-06-22T14:40:56.687427Z
ARGOS_000009DRAFT	"SRA Metadata and QC metrics from the Original FDA BioProject and the National Collection of Pathogenic Viruses"	DRAFT	ARGOS	2022-06-22T15:40:49.82210Z

BioCompute Portal 22.07 | BCODB 22.07 | Portal UserDB 22.07 | Contact Us | Report bug or request feature

## For our GCC 2022 demo, we aim to present:

1) An (published) update to Galaxy generated BCOs, and upcoming direct submission PR. With the addition of a direct submission function to BCODB from Galaxy updates that is no longer an issue.

POST /api/objects/drafts/create/ Create BCO Draft

Creates a new BCO draft object.

Parameters

Name Description

data \*required object (body) Edit Value Model

Create BCO Draft Schema

```

POST /api/objects/drafts/create/ {
  "description": "Parameters that are supported when trying to create a draft BCO.",
  "parameters": {
    "BCO Drafts to create": {
      "prefix": "string", "description": "BCO Prefix to use",
      "owner_group": "string", "description": "Group which owns the BCO draft.",
      "object_id": "string", "description": "BCO Object ID.",
      "schema": "string", "description": "Which schema the BCO satisfies.",
      "contexts": "> (...)"
    }
  }
}

```

We plan to submit a pull request for a direct form submission via API in the next Galaxy release.

2) In the BCODB, a BCO identifier ([https://biocomputeobject.org/GALXY\\_000001/1.0](https://biocomputeobject.org/GALXY_000001/1.0)) is composed of the following: a protocol [<https://>], a hostname [[biocomputeobject.org/](https://biocomputeobject.org/)], prefix [GALXY], number [\_000001/], and version [1.0]. We have created a reserved prefix as the default for BCOs submitted from a Galaxy instance. This is available to all users on our public instance.

The "object\_id" consists of four different parts:

<protocol>	<hostname>	<BCO accession>	<BCO version>
https://	biocomputeobject.org/	/BCO_000001	/1.5

- Protocol: It is recommended that this be `https` or at least `http` see Scheme in RFC 3986 for more detail on protocols.
- Hostname: see Host in RFC\_3986
- BCO Accession: composed of 2 parts separated by an underscore( \_ ):
  - the BCO Prefix: A 3-5 alphanumeric charaters. Preferably one registered via the BioCompute Registry
  - BCO identifier: a numeric identifier appended to the prefix
- BCO version: This is the version of the BCO, as recorded on the Provenance Domain

"object\_id" `https://biocomputeobject.org/BCO_000001/1.5"`

3) The Database for Reference Grade Microbial Sequences (FDA-ARGOS) project has been utilizing Galaxy BCOs to document pipelines. These regulatory-grade computational workflows are used for curation of a knowledgebase. Galaxy generated BCO are used (along with BCOs from other platforms) for data curation and deposited into knowledgebases [GlyGen ([www.glygen.org](http://www.glygen.org)) and FDA\_ARGOS (<https://data.argosdb.org/>)].

ArgosDB Home FAQ Release History About

Search by: BCID, dataset file name, title, description or categories

3 results found, after filters: 'assembly': 'Influenza A'

reviewed dataset ARGOS\_000012 in TSV format. [multiple]

assemblyQC-HIVE Lab

organism_name...	lineage...
Influenza_A_...	strain_APuer...
Salmonella_t...	str. LT2 (en...

QC of short read sequences assembled into genomes.

reviewed dataset ARGOS\_000022 in TSV format. [multiple]

assemblyQC - Pond Lab

organism_name...	lineage...
SARS-CoV-2...	B.1.617.2...
SARS-CoV-2...	B.1.1.52...

QC of short read sequences re-assembled into genomes.

reviewed nucleotide dataset ARGOS\_000001 in FASTA format. [Influenza A]

Influenza A reference gene sequences

```

>>> [ORF8]4-
1[CHM2A_<ORF>
9[ORF3]4-
1[CHM2A_<ORF>
Vacuolar protein sorting-
associated protein 2
DS=dsipuluc base=
GN=Ccdc15 PE=1 SV=122
MDLLFGRKTPF

```

Influenza A (A/Puerto Rico/8/1934 H1N1)