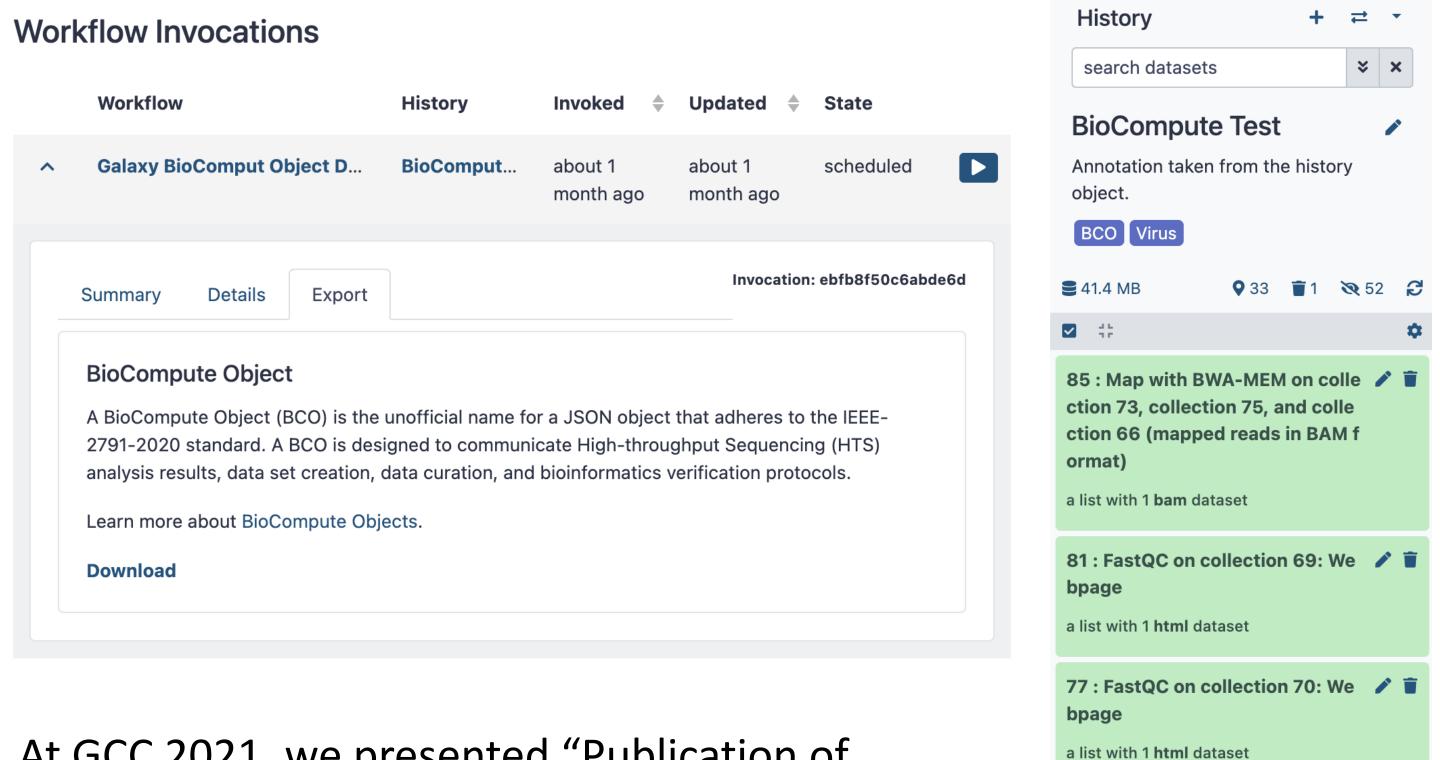


Galaxy Workflows to BioCompute Objects

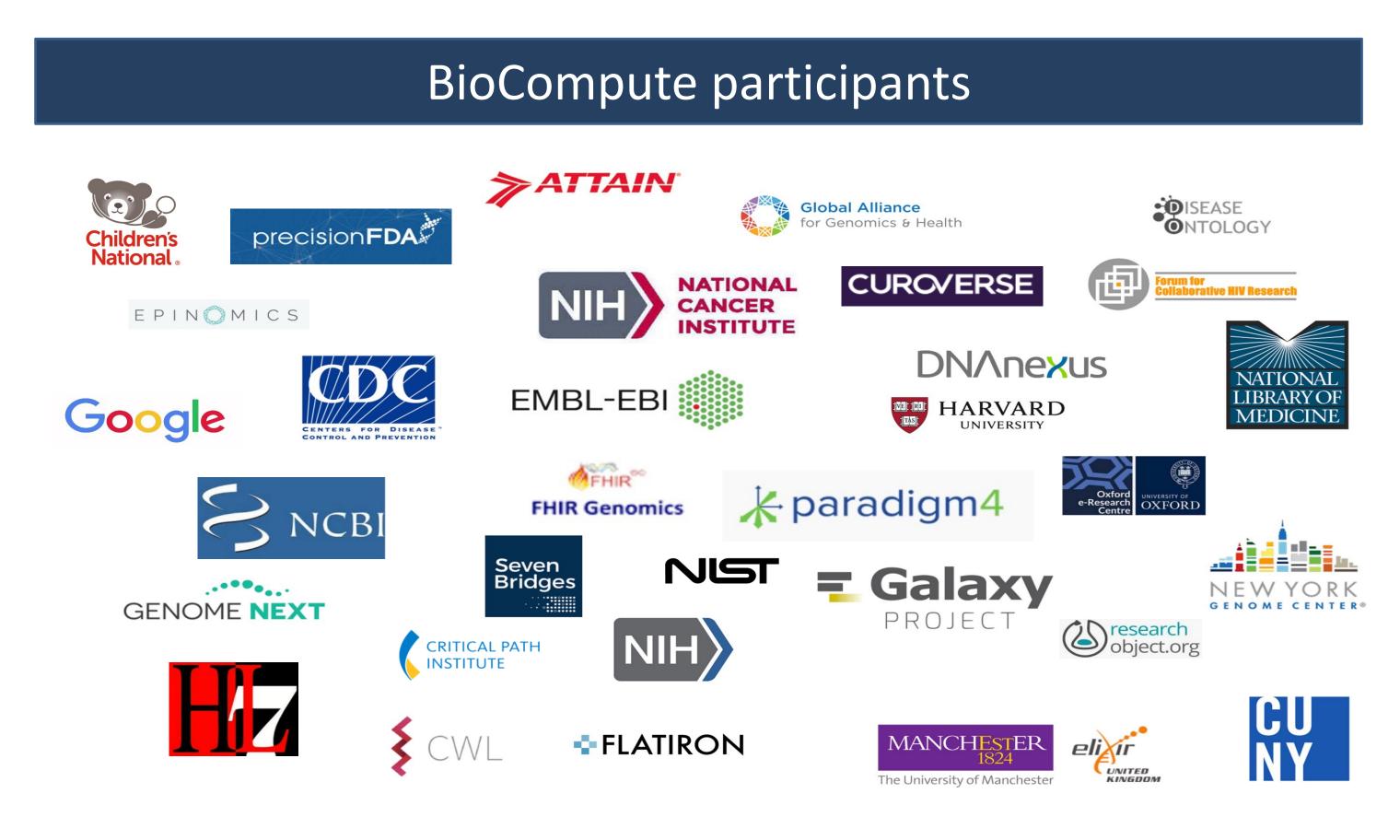
Charles Hadley King, Christopher A. Armstrong, Jonathon Keeney, Janisha Patel, Stephanie S. Singleton, Stephen D. Shank, Raja Mazumder

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.



At GCC 2021, we presented "Publication of BioCompute Objects (IEEE-2791-2020) created

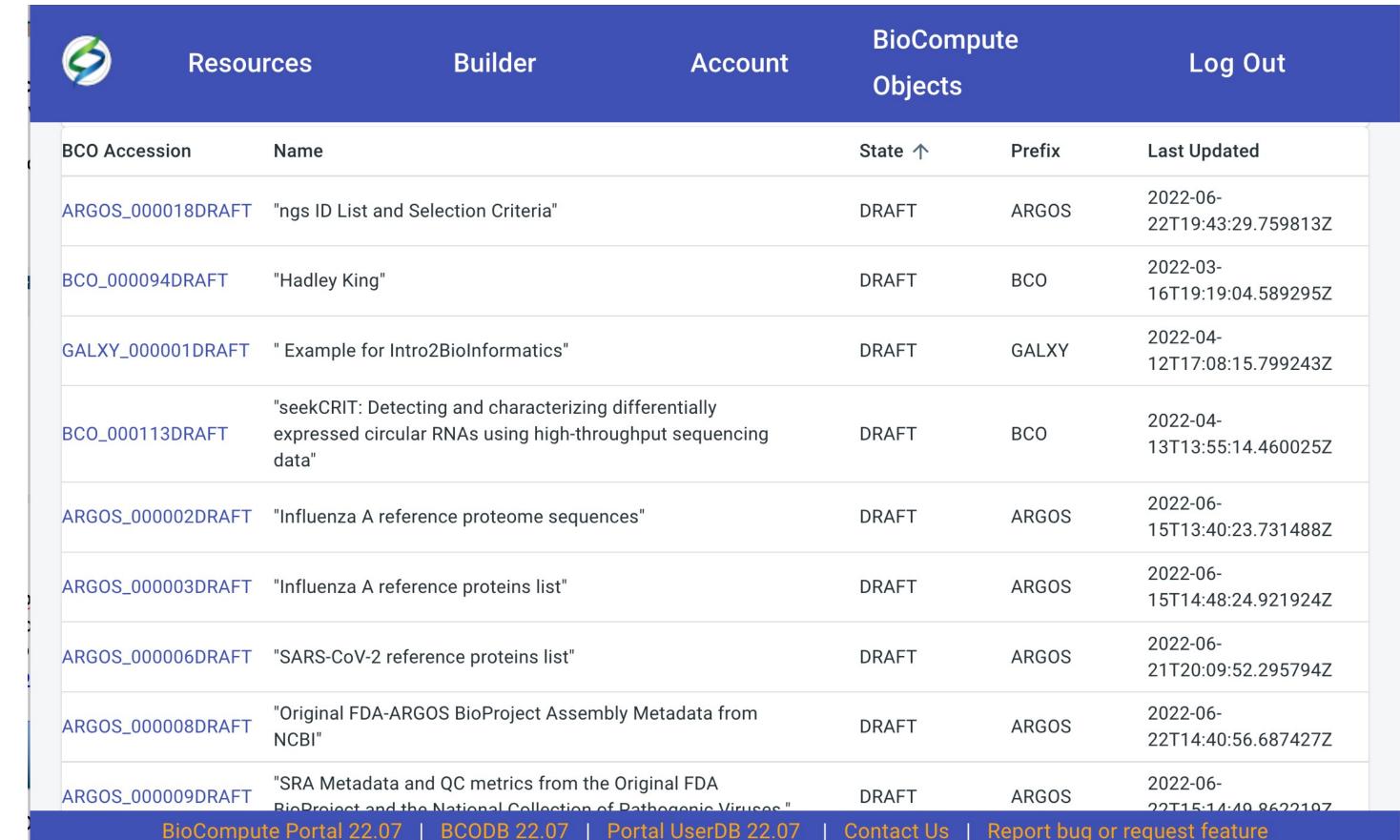
from Galaxy workflow invocations" (https://vimeo.com/567195992). Here, we demonstrated how a Galaxy BCO could be deposited to the BioCompute database (BCODB) via API. The BioCompute Object DB APIs are all available via at https://biocomputeobject.org/api/docs/.



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, DNAnexus, Seven Bridges, and the Galaxy community.

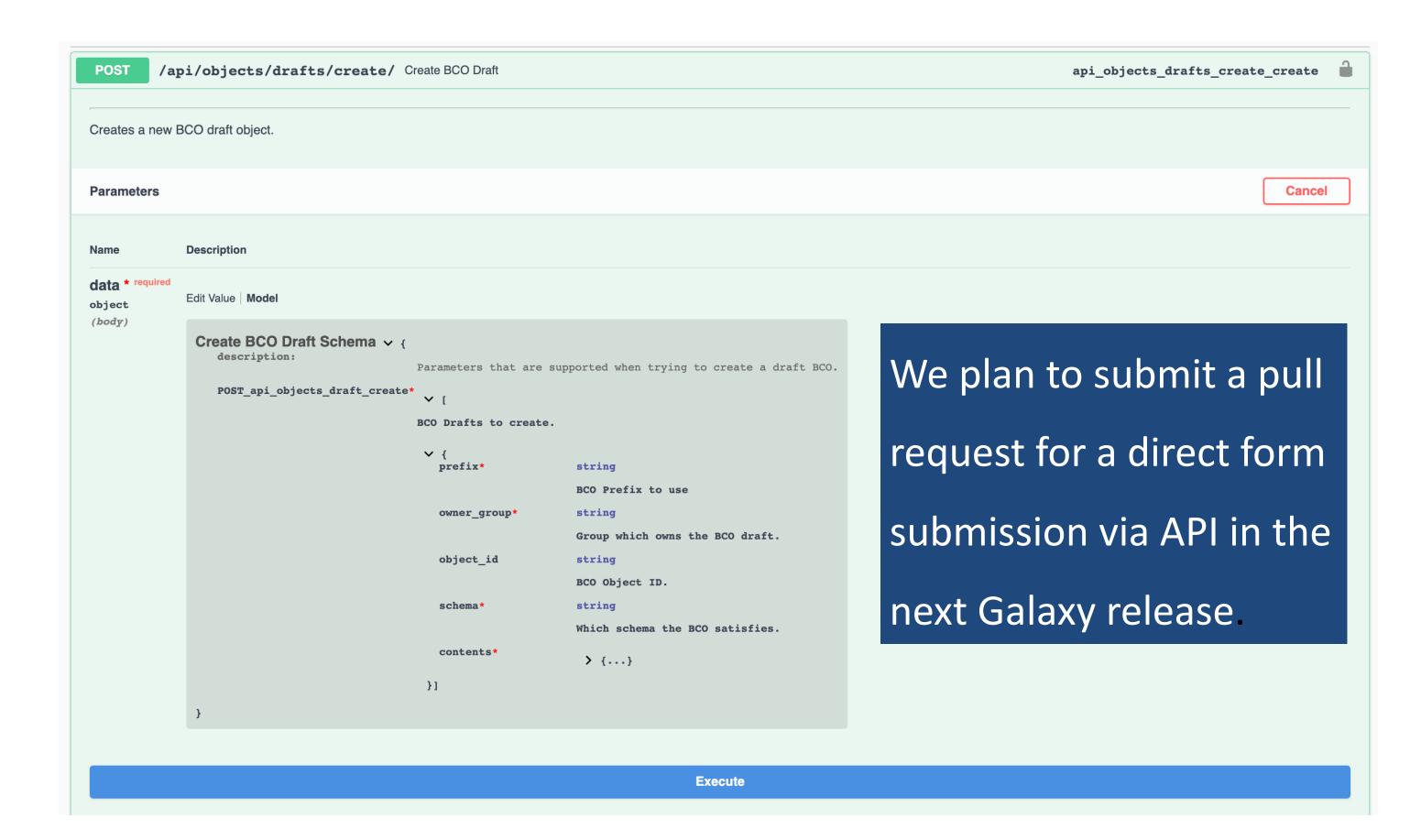
A BioCompute Object (BCO) is the unofficial name for a JSON object that adheres to the IEEE-2791-2020 standard. A BCO is designed to communicate High-throughput Sequencing (HTS) analysis results, data set creation, data curation, and bioinformatics verification protocols.

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.



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.



2) In the BCODB, a BCO identifier

(https://biocomputeobject.org/GALXY_000001/1.0) is composed of the following: a protocol [https://], a hostname [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.

<pre><pre><pre><pre><pre><pre><pre>https://</pre></pre></pre></pre></pre></pre></pre>	<pre> <hostname> biocomputeobject.org/</hostname></pre>	<bco accession=""> /BCO_000001</bco>	<bco version=""> - /1.5 </bco>
1. Protocol:	t is recommended tha	at this be https	or at least http see Scheme in RFC 3986 for more detail on protocols
2. Hostname	e: see Host in RFC_39	86	
3. BCO Acce	ssion: composed of 2	parts separated	by an underscore(_):
∘ th	e BCO Prefix: A 3-5 a	lphanumeric cha	araters. Preferably one registed via the BioCompute Registry
	CO identifier: a numer	ic identifier appe	ended to the prefix
。 B0			

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) and FDA_ARGOS (https://data.argosdb.org/)].

