

Surya Saha¹, Michele Mattioni¹, Milos Trboljevac¹, Eric Wenger², Allison Heath², Yuankun Zhu², Robert Carter³, Michelle Giglio³, Suvarna Nadendla³, C. Titus Brown⁴, Bailey K Farrow², Daniel J. B. Clarke⁵, Adam Kraya², Kristin Ardlie⁶, Jared Nedzel⁶, Lan Nguyen⁵, Avi Ma'ayan⁴, Owen White³, Jack DiGiovanna¹, Adam Lesnick² ¹ Seven Bridges Inc., Charlestown, MA 02129, USA; ² Center for Data-Driven Discovery in Biomedicine, Children's Hospital of Philadelphia, Philadelphia, PA 19146, USA; ³ Institute for Genome Sciences, University of Maryland, Baltimore, MD 21201, USA; ⁴ School of Veterinary Medicine, University of California, Davis, CA 95616, USA; ⁵ The Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA; ⁶ Department of Pharmacological Sciences, Mount Sinai Center for Bioinformatics, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

Platform-based discovery powered by DRS

The diagram illustrates a workflow for creating a cohort on the CFDE portal and exporting a manifest file, which is then used by the CAVATICA DRS manifest importer to create a cohort on the Global Alliance for Genomics & Health Data Repository Service (DRS) API.

Step 1: Create cohort on CFDE portal and export manifest file

The CFDE portal interface shows a table of samples with columns for ID, Name, Sex, Age, Race, and various genomic data fields. A green arrow points from the CFDE portal to the CAVATICA DRS manifest importer.

Step 2: CAVATICA DRS manifest importer

The CAVATICA DRS manifest importer interface shows a form for creating a new cohort. It includes fields for Cohort Name, Cohort Description, and Cohort Type. A green arrow points from the CAVATICA DRS manifest importer to the DRS API.

Step 3: Global Alliance for Genomics & Health Data Repository Service (DRS) API

The DRS API interface shows a form for creating a new cohort. It includes fields for Cohort Name, Cohort Description, and Cohort Type. A green arrow points from the DRS API to the Discovery analysis and visualization section.

Step 4: Discovery analysis and visualization

The Discovery analysis and visualization section shows a table of samples with columns for ID, Name, Sex, Age, Race, and various genomic data fields. It also includes a section for Discovery analysis and visualization, showing a table of samples with columns for ID, Name, Sex, Age, Race, and various genomic data fields.

The diagram also includes logos for the Global Alliance for Genomics & Health, RStudio, Jupyter, Scala, and Appyters, indicating the tools and services used in the workflow.

Manifest-based DRS import



Attribute	Definition	Example
drs_uri (required)	DRS URI as defined by GA4GH DRS spec for pointers to file objects.	drs://dg.ANV0:dg.ANV0/97e349fa-88f5-4fef-8008-87a3aad3560a
study_registration	External source from which the identifier included in study_id originates (answer can be dbGaP for example)	gtex_id
study_id	Unique identifier that can be used to retrieve more information for a study	GTE _x
participant_id	Unique identifier that can be used to retrieve more information for a participant	GTE _x -N7MS
specimen_id	Unique identifier that can be used to retrieve more information for a specimen	GTE _x -N7MS-2526-SM-2D7W3
experimental_strategy	The experimental strategy used to generate the data file referred to by the ga4gh_drs_uri. (Based on GDC definition)	RNA sequencing assay
file_format	The format of the data, see possible values from the data_format fields in GDC. Can use whatever values make sense for the particular implementation.	BAM
fhir_document_reference	Optional fhir url pointing to the FHIR Document Reference, if metadata available on a FHIR Server	https://include-api-fhir-service.includedcc.org/DocumentReference?identifier=HTP.2885a2e1-001b-4751-b622-fdfa2795ea6.rsem.isoforms.result.s.gz

- Establish dedicated Git repo on official NCPI Github organization
- Socialize the standards to the community to get feedback for further improvements and broader adoption

