Getting to know the GA4GH Workflow Execution Service (WES) API

NCI Containers & Workflows Webinar

James Eddy @ Sage Bionetworks | Ian Fore @ NCI CBIIT

April 8, 2022

SageBionetworks

These slides: *bit.ly/nci-cwig-wes*

Summary

- Today we'll share an introduction to and overview of the **Workflow Execution Service (WES) API**, a standard developed by members of the **Cloud Work Stream** and associated Driver Projects in the **Global Alliance for Genomics and Health (GA4GH)**
- This is *not* a technical deep dive — to learn more about WES (and to get involved with the Cloud WS), check out ***ga4gh-cloud.github.io***

Agenda

- **Chapter 1:** Sage Bionetworks & DREAM Challenges
- **Chapter 2:** A Crowd-Sourced Workflow Execution Challenge
- **Chapter 3:** GA4GH Cloud Work Stream & the WES API
- **Chapter 4:** WES Use Cases & Implementations
- **Chapter 5:** Workflow Interop Testbed & Federated Analysis Systems Pilot
- **Chapter 6:** Ongoing Development with WES

Chapter 1:

Sage Bionetworks & DREAM Challenges

SageBionetworks

Who is Sage Bionetworks?

Non-profit research institute based in Seattle

Mission to accelerate biomedical discoveries by improving methods for scientific *collaboration* and *communication*

Better Science Together

Core Values @ Sage

- **Be intentional** – consider solutions from more than one perspective.
- Promote an **ecosystem of sharing** – with proper **attribution**.
- Solve specific problems with general solutions – make these available for **reuse and adaptation**.
- Do **trustworthy, impactful work** – prioritize **outcomes over ego**.
- **Be bold** and willing to experiment.

What do we do?

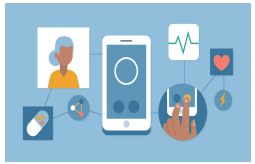
At Sage we believe that by harnessing the power of open science, we help research communities develop reliable outcomes to advance our understanding of human health.

Responsible Data Sharing

Sage supports research collaborations by overseeing data coordination, visualization, and analytics across distributed teams. We manage grant- or project-based research consortia to share, evaluate and distribute data, methods, and insights.

Benchmarking Reliable Methods

Because we are all susceptible to the self-assessment bias, Sage has developed tools that help researchers to objectively benchmark the performance of computational methods, and to disseminate community-verified methods.

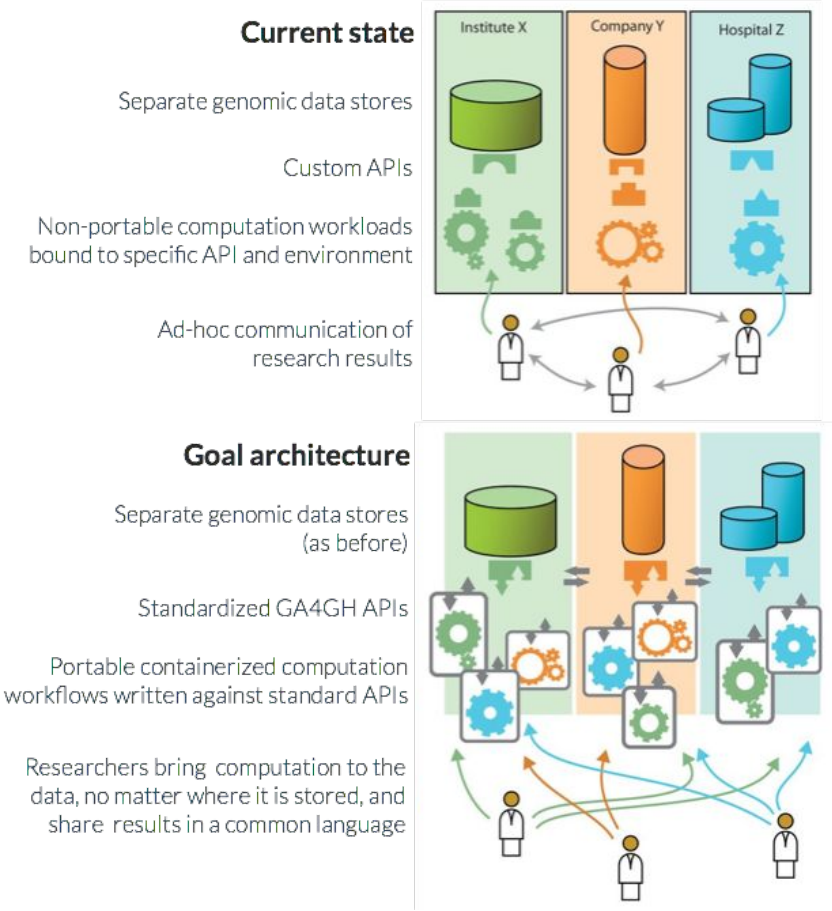


Understand Real-World Evidence

By applying our approach to digital health, Sage works with participants and researchers to understand how real-world environments impact our individual experience of health and disease.

Synapse: platform for open science

- Enabling FAIR ecosystems requires scalable infrastructure, community standards/APIs, and tools for researchers of all types
- Sage has pioneered open science frameworks and approaches, with Synapse as the primary platform we use to support these efforts



SageBionetworks

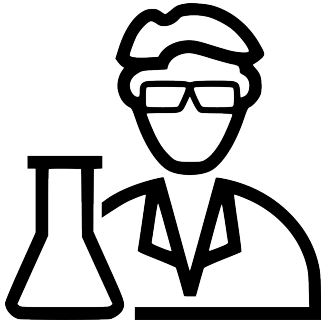
Paten, B. et al. The NIH BD2K center for big data in translational genomics. JAMIA, 2015 . doi:10.1093/jamia/ocv047

Data & Tooling @ Sage

Dr. Brian O'Connor
Chief Data Officer

Dr. James Eddy
Director of Informatics & Biocomputing

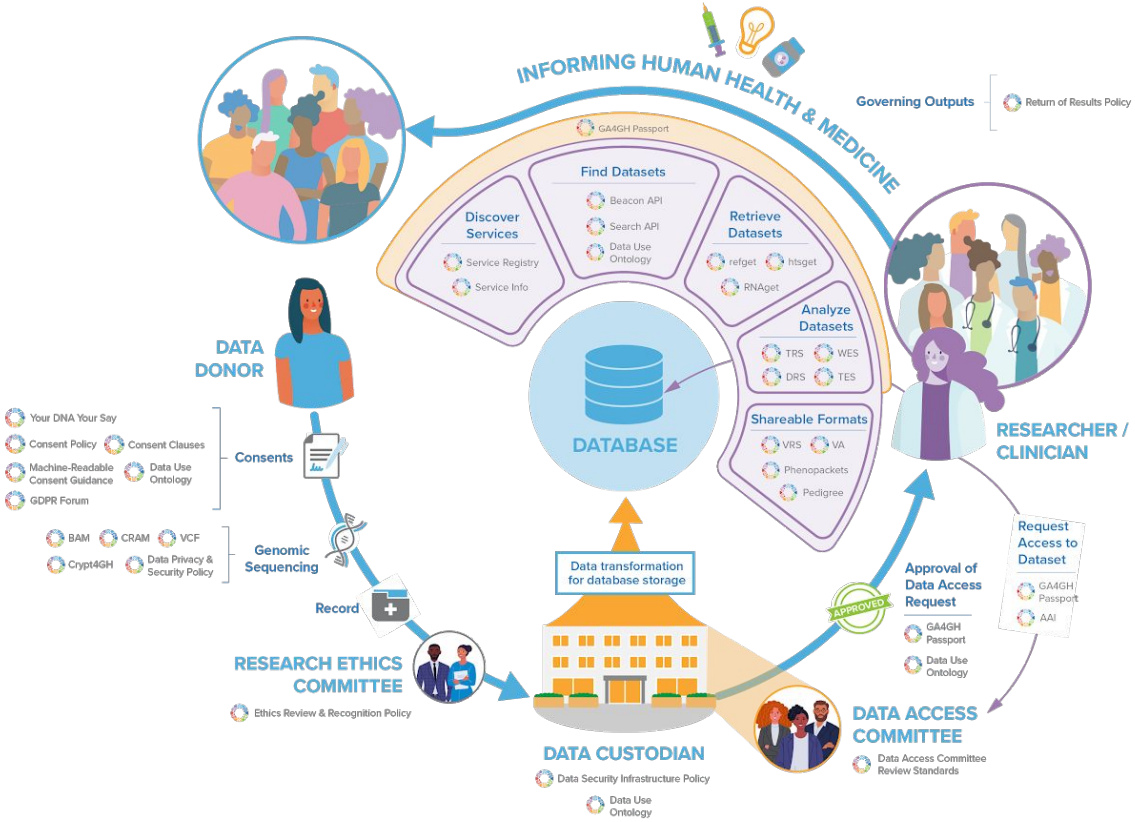
We utilize scalable/robust technologies and adopt "industry standard" best practices in our daily work to build systems that enable researchers to access data and generate impactful and measurable outcomes.



SageBionetworks

Global Alliance for Genomics and Health (GA4GH)

- Promotes sharing across the translational continuum (discovery research, clinical trials, clinics, diagnostic labs, industry)
- Standards for interoperability: scientific, technical, and ethical
  - File formats, variant annotation schemas, phenotype data serialization, etc.
  - Consent policies to ensure that data can be shared internationally
  - ***Common APIs for data access, sharing, analysis web services***



SageBionetworks

DREAM CHALLENGES

Our mission is …

- to contribute to the solution of important **biomedical** problems
- to foster **collaboration** between research groups
- to **democratize access** to data
- to **accelerate research**
- to **objectively assess** algorithms and their performance

ITCR U24 for advancing method benchmarking

U24	2020	• James Eddy • Paul Christopher Boutros	• Sage Bionetworks • University of California Los Angeles	Advancing Method Benchmarking and Data Sharing Through Crowd-Sourced Competitions in Cancer Research	Active

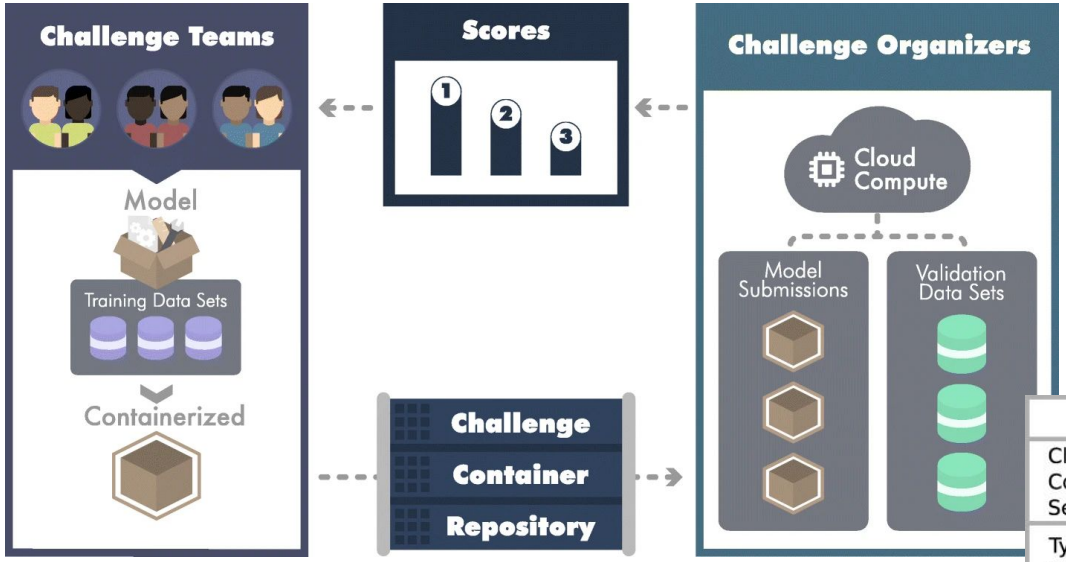
- **AIM 1:** Develop a community hub and ***benchmarking toolkit*** for biomedical challenges.
- **AIM 2:** Develop ***portable software and services for distributed benchmarking*** on sensitive and protected data.
- **AIM 3:** Expand the biomedical challenge community through improvements in education, outreach, and empowering the organization of independent challenges and benchmarking projects.

Dr. Jake Albrecht
Director of Challenges & Benchmarking

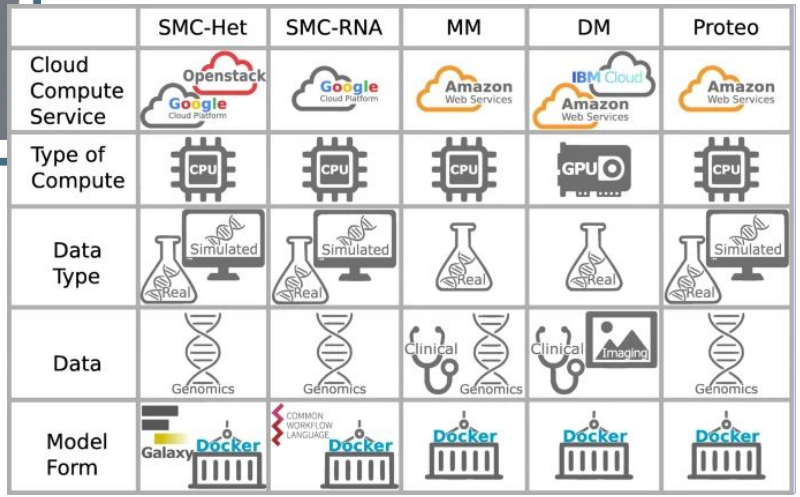
Dr. Paul Boutros
Director of Cancer Data Science at UCLA

SageBionetworks

Model-to-data for benchmarking



- Permits use of sensitive, proprietary, or otherwise hard-to-move data
- Preserves integrity of gold-standard validation data
- Algorithm reproducibility and re-usability
- Prospective assessment



Images from Ellrott, et al., Genome Biol (2019):
Reproducible biomedical benchmarking in the cloud: lessons from crowd-sourced data challenges

SageBionetworks

Chapter 2:

A Crowd-Sourced Workflow Execution Challenge

SageBionetworks

Problem: containers are a small slice of the picture

Demonstrating utility of GA4GH Cloud standards

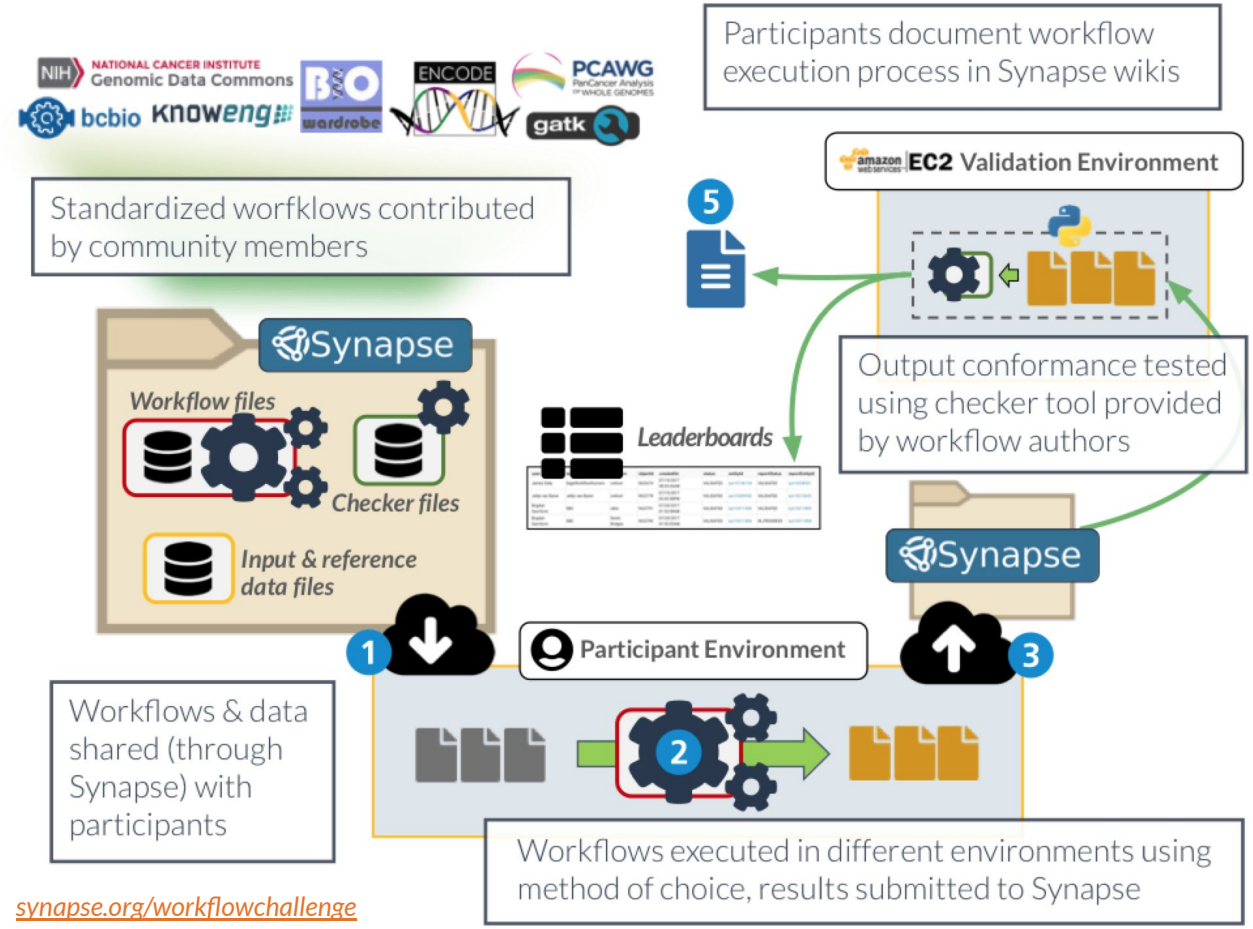
- The **GA4GH/DREAM Infrastructure Challenges**

DREAM CHALLENGES powered by Sage Bionetworks + Global Alliance for Genomics & Health

- Not your standard DREAM competition: ***community effort*** to evaluate, benchmark, and develop tools/platforms for reproducible, portable, and scalable pipelines
- Series of challenges focused on individual aspects of GA4GH vision — and ultimately, integration of all components
- Will also contribute to enhanced infrastructure for other DREAM challenges

GA4GH/DREAM Workflow Execution Challenge

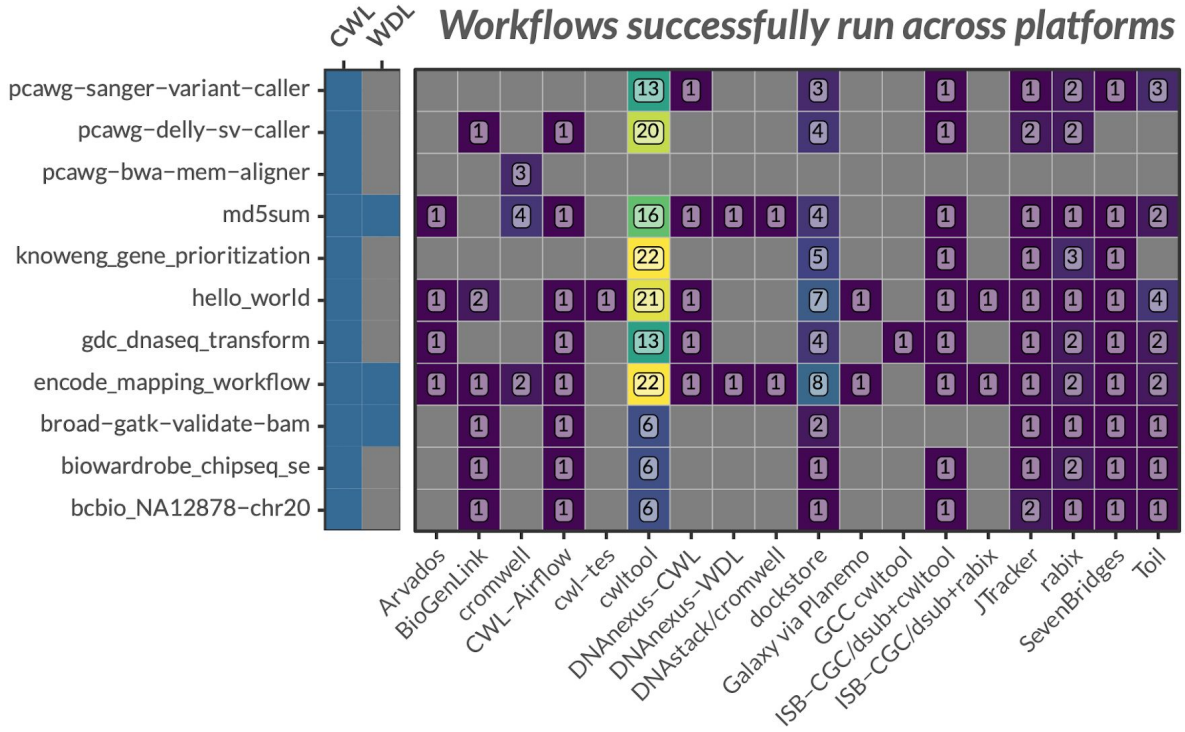
- Collaborative effort between GA4GH, multiple working groups focused on containers/workflows, and DREAM (Sage)
- Community, crowd-sourced testing and evaluation of workflow portability, workflow engine/platform compliance

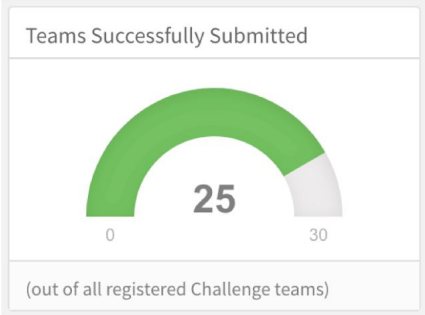


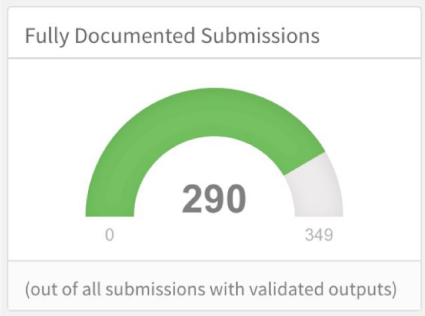
synapse.org/workflowchallenge

SageBionetworks

Teamwork makes the DREAM work(flow)





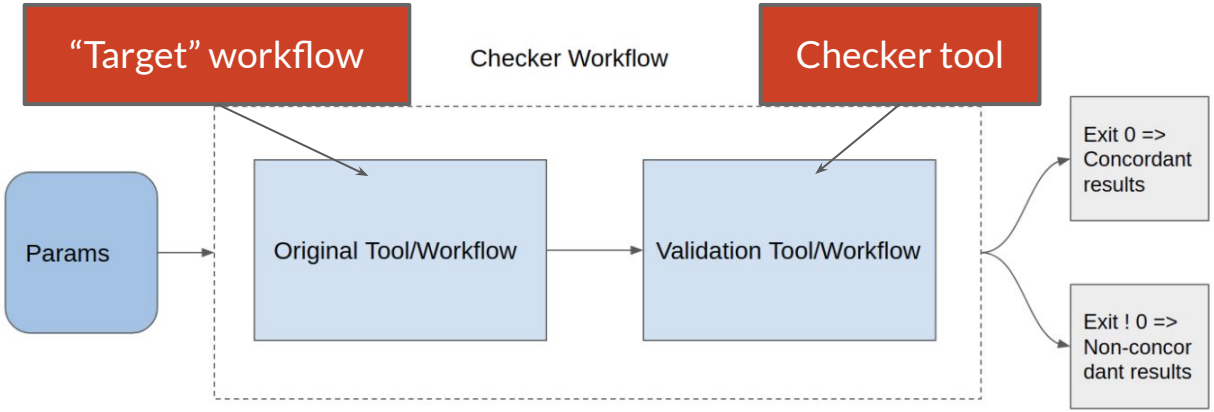


synapse.org/workflowchallenge

SageBionetworks

Standardizing workflow validation

Checker workflows are additional workflows you can associate with a tool or workflow. The purpose of them is to ensure that a tool or workflow, given some inputs, produces the expected outputs. Below is a visual overview of how a checker workflow looks.



See github.com/ga4gh/dockstore/wiki/WIP-Checker-Workflow-Support-Tutorial

Chapter 3:

GA4GH Cloud Work Stream & the Workflow Execution Service (WES) API

SageBionetworks

GA4GH Cloud Work Stream

The ***Cloud Work Stream*** is focused on creating specific standards for defining, sharing, and executing portable workflows and accessing data across clouds.

AMED

Australian Genomics Health Alliance

EUROPEAN GENOME-PHENOME ARCHIVE

ENA European Nucleotide Archive

HUMAN CELL ATLAS

Genomics england

elixir

NIH National Heart, Lung, and Blood Institute

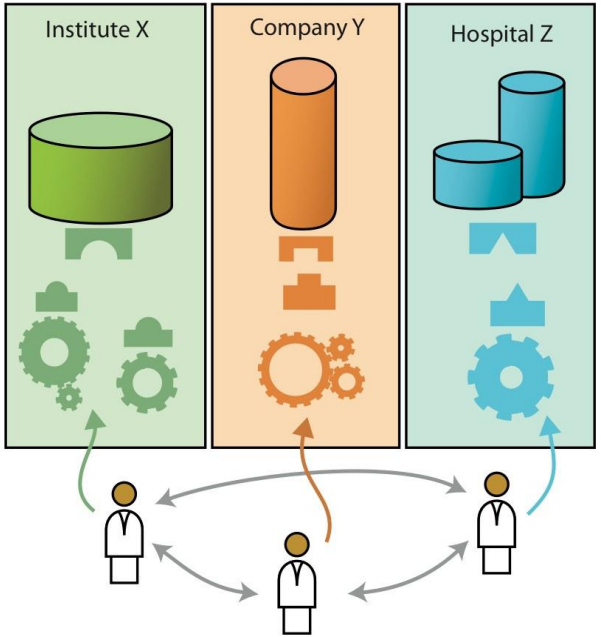
ARGO

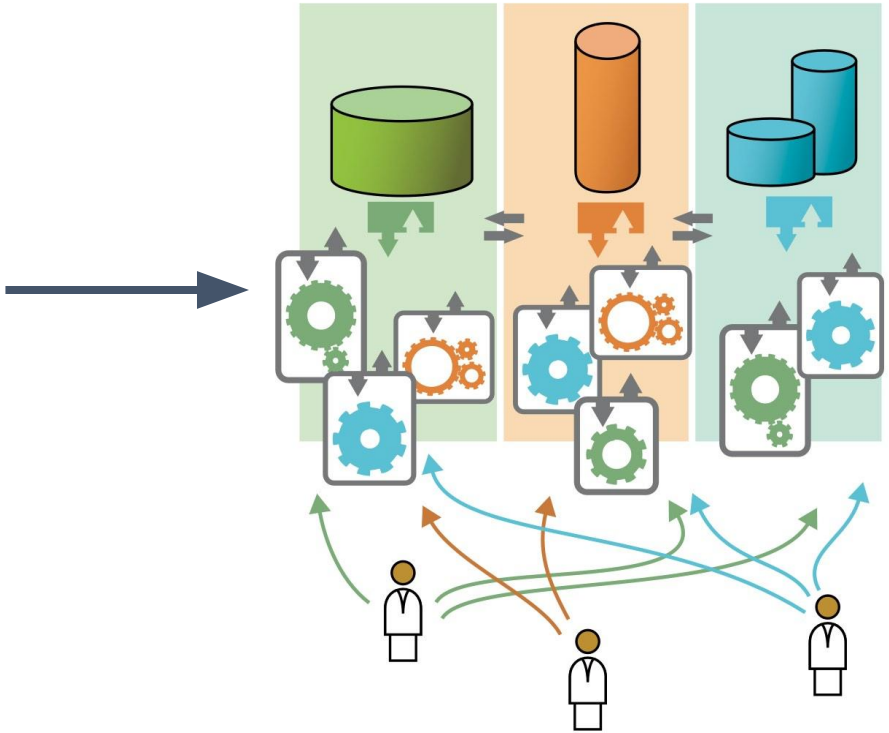
And more!

We work with many different **Driver Projects** to develop, enhance, test, and use the **Cloud WS** APIs.

SageBionetworks

Cloud Work Stream vision





SageBionetworks

Cloud Work Stream APIs

Cloud Work Stream APIs

TRS: Tool Registry Service API

Provides workflows and container images

DRS: Data Repository Service API

Provides access to data sets

WES: Workflow Execution Service API

Interprets and executes workflows
***or** schedules execution via TES*

TES: Task Execution Service API

Executes individual tasks and stages data in/out

SageBionetworks

elixir

Global Alliance
for Genomics & Health
Collaborate. Innovate. Accelerate.

Search...

Workflow Execution Service (1.0.1)

Download OpenAPI specification: Download

Executive Summary

The Workflow Execution Service (WES) API provides a standard way for users to submit workflow requests to workflow execution systems, and to monitor their execution. This API lets users run a single workflow (currently **CWL** or **WDL** formatted workflows, other types may be supported in the future) on multiple different platforms, clouds, and environments. Key features of the API:

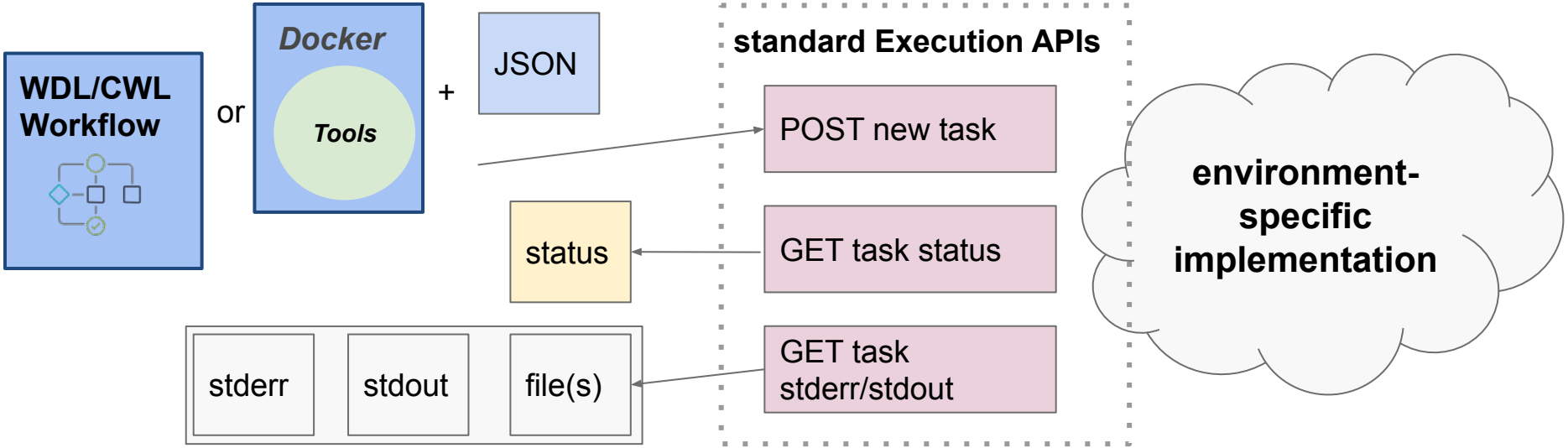
- can request that a workflow be run
- can pass parameters to that workflow (e.g. input files, cmdline arguments)
- can get information about running workflows (e.g. status, errors, output file locations)
- can cancel a running workflow

SageBionetworks

API Docs: *https://ga4gh.github.io/workflow-execution-service-schemas/docs/*

Workflow Execution (as a) Service

- **WES API** provides a way to send a request to run a CWL or WDL-described workflow in a remote environment, monitor progress, and retrieve the result



GitHub: https://github.com/ga4gh/workflow-execution-service-schemas

Latest release: 1.0.0

SageBionetworks

WES as a *meta*-API

- Common instructions for submitting jobs to a workflow engine/platform
- WES server handles execution logic

```
RunRequest ⌄ {
  description:
                To execute a workflow, send a run request including all the details needed to begin
                downloading
                and executing a given workflow.

  workflow_params
                ⌄ {
                  description:
                                REQUIRED
                                The workflow run parameterizations (JSON encoded), including input and
                                output file locations

                }
  workflow_type   string

                REQUIRED
                The workflow descriptor type, must be "CWL" or "WDL" currently (or another alternative
                supported by this WES instance)

  workflow_type_version   string

                REQUIRED
                The workflow descriptor type version, must be one supported by this WES instance

  tags
                > {...}
  workflow_engine_parameters
                > {...}
  workflow_url    string

                REQUIRED
                The workflow CWL or WDL document. When workflow_attachments is used to attach files, the
                workflow_url may be a relative path to one of the attachments.

}
```

WorkflowExecutionService

GET `/service-info` Get information about Workflow Execution Service.

GET `/runs` List the workflow runs.

POST `/runs` Run a workflow.

This endpoint creates a new workflow run and returns a `RunId` to monitor its progress.

The `workflow_attachment` array may be used to upload files that are required to execute the workflow, including the primary workflow, tools imported by the workflow, other files referenced by the workflow, or files which are part of the input. The implementation should stage these files to a temporary directory and execute the workflow from there. These parts must have a Content-Disposition header with a "filename" provided for each part. Filenames may include subdirectories, but must not include references to parent directories with '..' -- implementations should guard against maliciously constructed filenames.

The `workflow_url` is either an absolute URL to a workflow file that is accessible by the WES endpoint, or a relative URL corresponding to one of the files attached using `workflow_attachment`.

The `workflow_params` JSON object specifies input parameters, such as input files. The exact format of the JSON object depends on the conventions of the workflow language being used. Input files should either be absolute URLs, or relative URLs corresponding to files uploaded using `workflow_attachment`. The WES endpoint must understand and be able to access URLs supplied in the input. This is implementation specific.

The `workflow_type` is the type of workflow language and must be "CWL" or "WDL" currently (or another alternative supported by this WES instance).

The `workflow_type_version` is the version of the workflow language submitted and must be one supported by this WES instance.

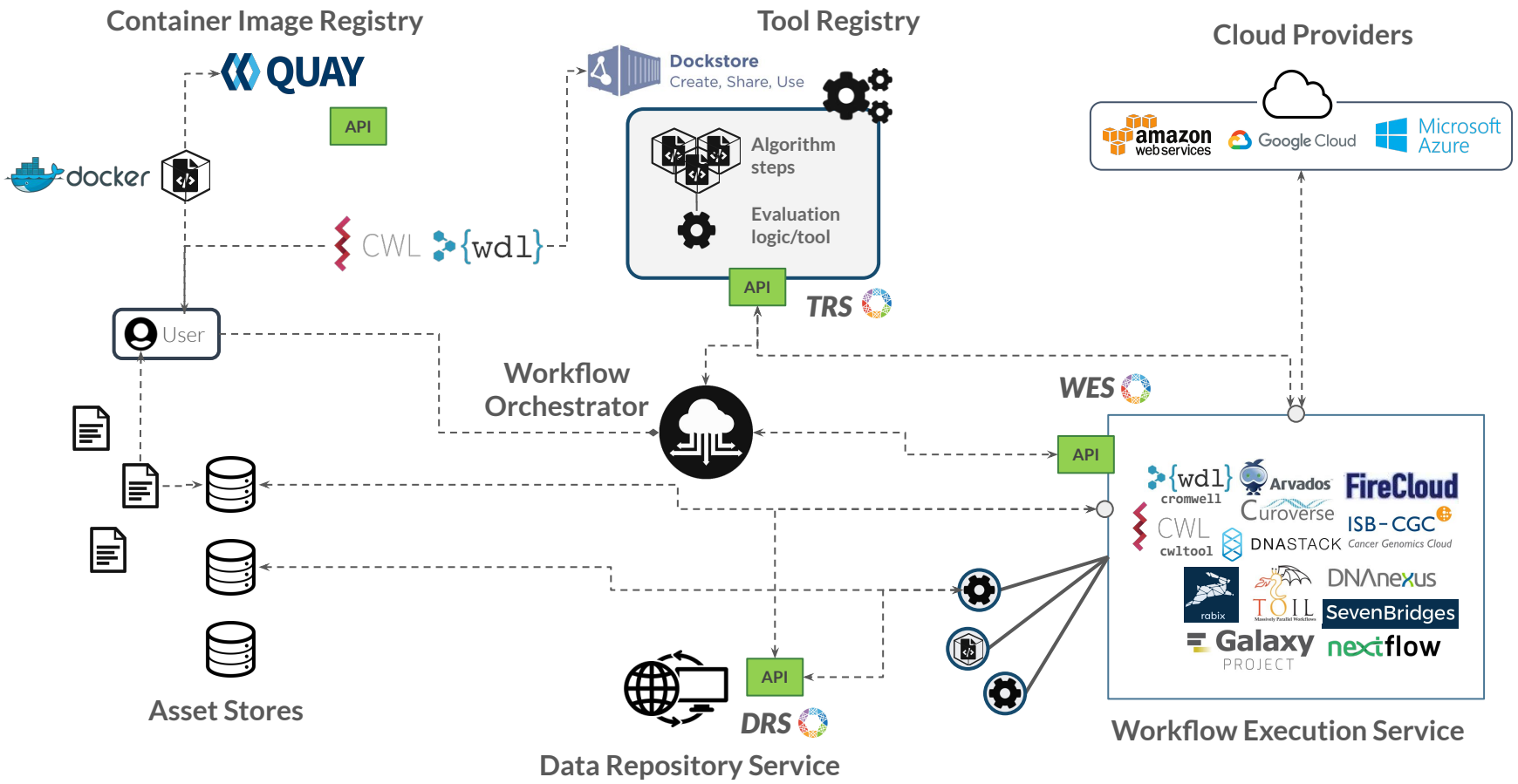
See the `RunRequest` documentation for details about other fields.

Parameters Try it out

Name	Description
workflow_params string($application/json) (formData)	
workflow_type string (formData)	
workflow_type_version string (formData)	
tags string($application/json) (formData)	
workflow_engine_parameters string($application/json) (formData)	
workflow_url string (formData)	
workflow_attachment array[string] (formData)	

SageBionetworks

WES in the ecosystem of GA4GH APIs



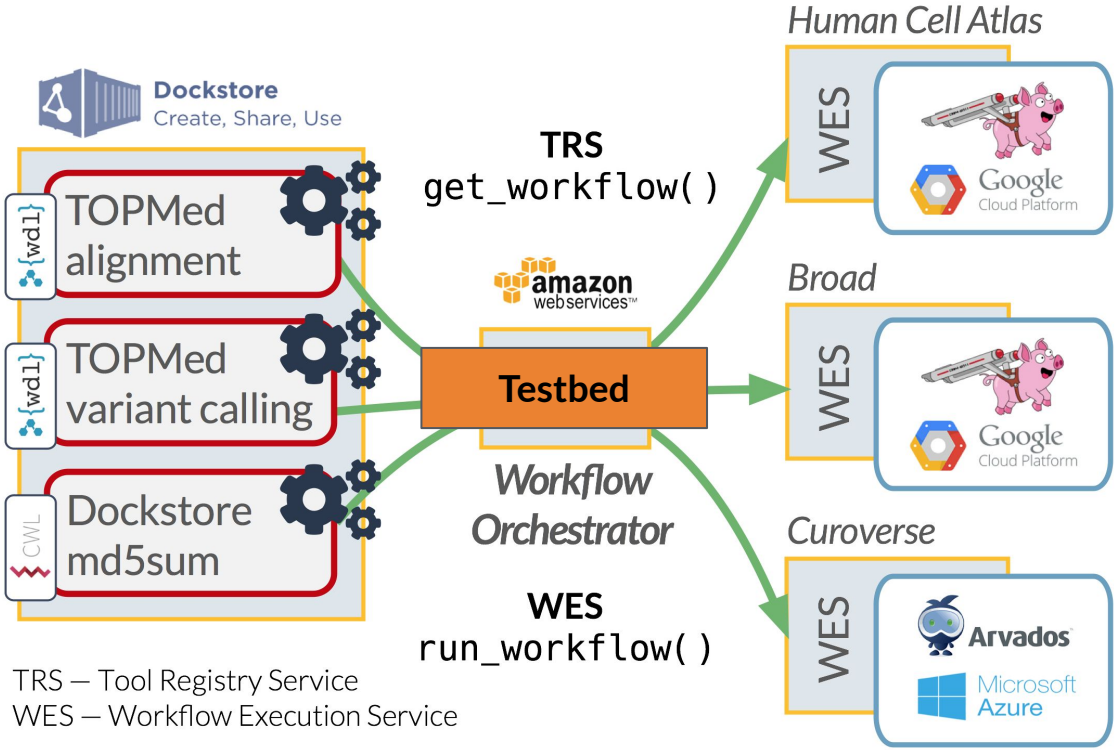
Chapter 4:

WES Use Cases & Implementations

SageBionetworks

Workflow Interoperability Testbed

- **WES/TRS client modules:** communicate with registered API server endpoints
- **Testbed:** iterate through registered workflows, identify corresponding checker workflows, create jobs for each WES endpoint, add to queue for execution
- **Orchestrator:** use transform module(s) and WES client to submit, manage, and monitor workflow jobs



TRS — Tool Registry Service
WES — Workflow Execution Service

Demonstrated at GA4GH F2F in Toronto, May 2018

SageBionetworks

Testbed results

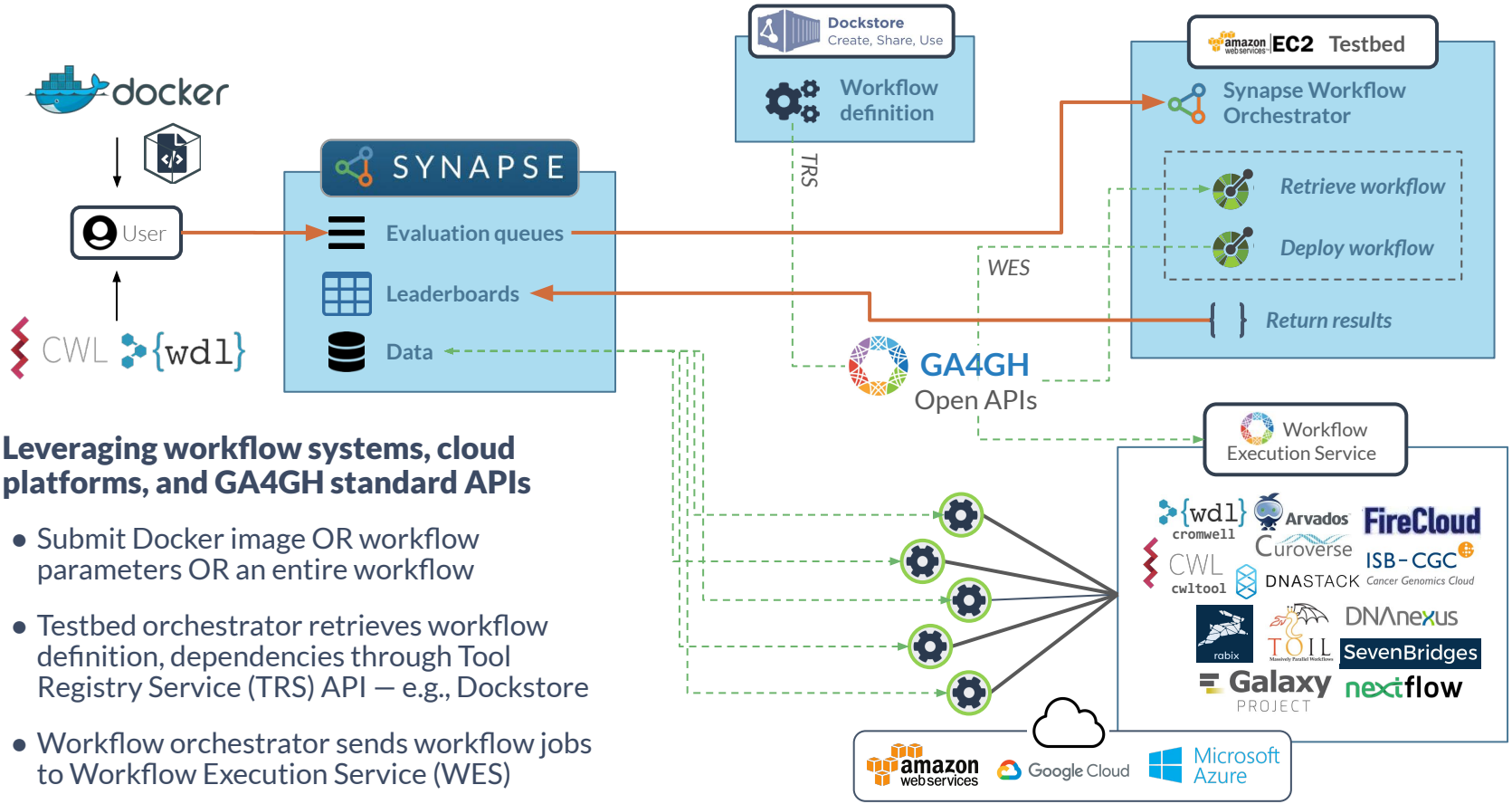
- CWL/WDL (through WES) are generally portable, reproducible
- Data access/storage protocols create mismatches
- Auth is hard...

					Workflows											
				Driver	n/a		**HCA**	**TOPMed**		**TOPMed**		**TOPMed**		**PCAWG**	**AGHA**	
				Workflow type	CWL	WDL	WDL	CWL		WDL		WDL		CWL	CWL	
				Input source	GitHub	GItHub	GS	GS	GS	GS	GS	GS	GS	various	GS	Keep
				Input protocol	relative	relative	https	gs	https	gs	https	gs	https	https	gs	keep
				Dockstore version	develop	develop	dockstore	1.29.0	1.31.0	1.29.0	1.31.0	1.29.0	1.31.0	checker	master	master
WES endpoints																
Driver	*Engine*	*WES version*	*FS supported*	*Types*	md5sum-checker	md5sum-checker/wdl	HCA_Smartseq2	TOPMed_alignment_pipeline_CWL		u_of_Michigan_alignment_pipeline		TopMed_Variant_Caller		pcawg-bwa-mem-workflow	wes-agha-test-gcp	wes-agha-test-arvados
HCA via CZI	Cromwell	v1.0	gs, http(s)	WDL		☐	✓			✓	✓					
TOPMed, HCA via Broad	Cromwell	v1.0	gs, http(s)*	CWL**, WDL		✓				✓		✓				
AGHA via Veritas	Arvados	v1.0	http(s), keep	CWL	✓				✓					✓		✓
GEL, AGHA via Illumina	Cromwell	v1.0	gs, http(s)*	CWL, WDL	☐	✓				✓		✓			✓	

SageBionetworks

Basel testbed

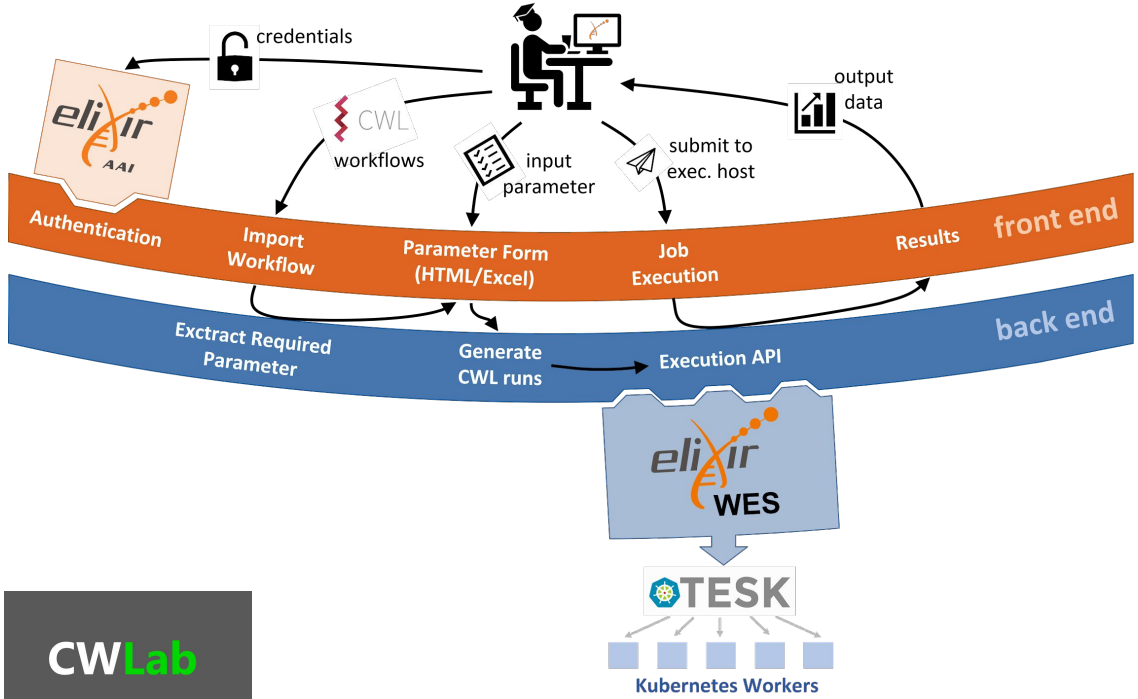
Model-to-data: why not workflows?



Leveraging workflow systems, cloud platforms, and GA4GH standard APIs

- Submit Docker image OR workflow parameters OR an entire workflow
- Testbed orchestrator retrieves workflow definition, dependencies through Tool Registry Service (TRS) API — e.g., Dockstore
- Workflow orchestrator sends workflow jobs to Workflow Execution Service (WES)

Elixir: CWLab



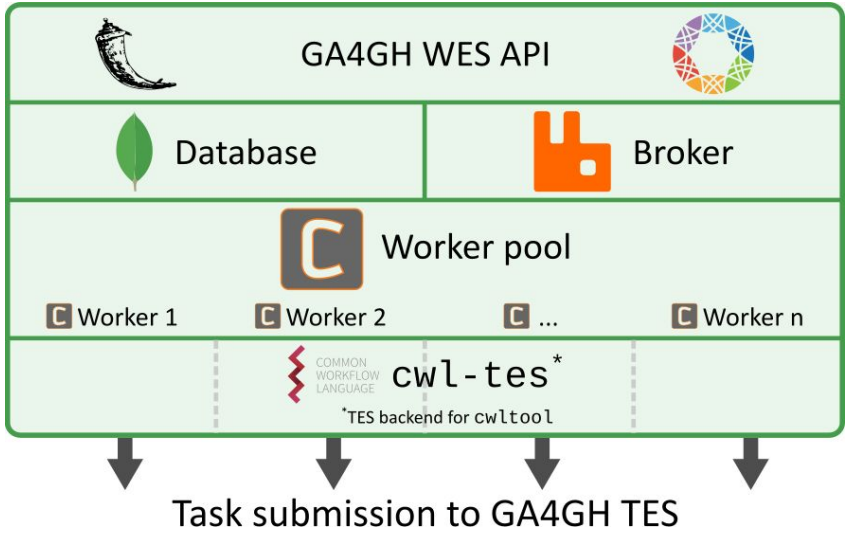
- User-friendly web interface
- GA4GH WES-compliant
- Import any CWL workflow
- Batch processing
- Abstraction model for workflow params

CWLab

https://github.com/CompEpigen/CWLab

Elixir: *cwl*-WES

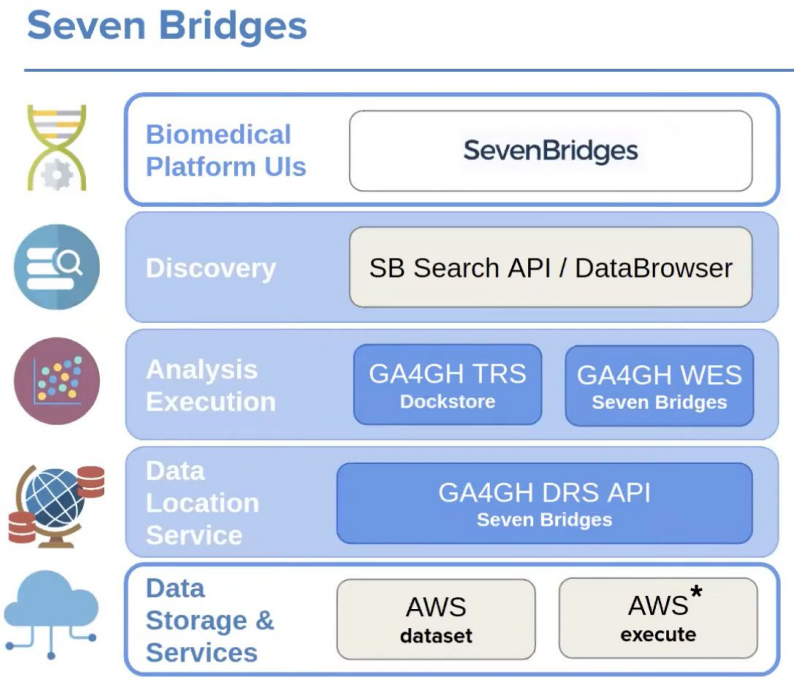
- Interpretation and scheduling of CWL workflows
- Designed to delegate execution to GA4GH TES instance



https://github.com/elixir-cloud-aai/cwl-WES

Other implementations

- **Seven Bridges**
- **Sapporo**
  https://github.com/sapporo-wes
- **ICGC ARGO**
  https://github.com/icgc-argo/workflow-api
- **DNAstack**
- **Nextflow Tower**
  https://tower.nf/openapi/index.html
- **GA4GH Starter Kit**
  https://github.com/ga4gh/ga4gh-starter-kit-wes



nextflow tower

ga4gh

GET	/ga4gh/wes/v1/runs	GA4GH list runs
GET	/ga4gh/wes/v1/runs/{run_id}	GA4GH describe run
GET	/ga4gh/wes/v1/runs/{run_id}/status	GA4GH retrieve run status
GET	/ga4gh/wes/v1/service-info	GA4GH service info
POST	/ga4gh/wes/v1/runs	GA4GH create a new run
POST	/ga4gh/wes/v1/runs/{run_id}/cancel	GA4GH cancel a run

Chapter 5:

WES in context and Federating analyses

Ian Fore
NIH, CBIIT

SageBionetworks

FASP Federation Demos

Global Alliance for Genomics & Health

Vertical demos: not covered here…

Horizontal demos:

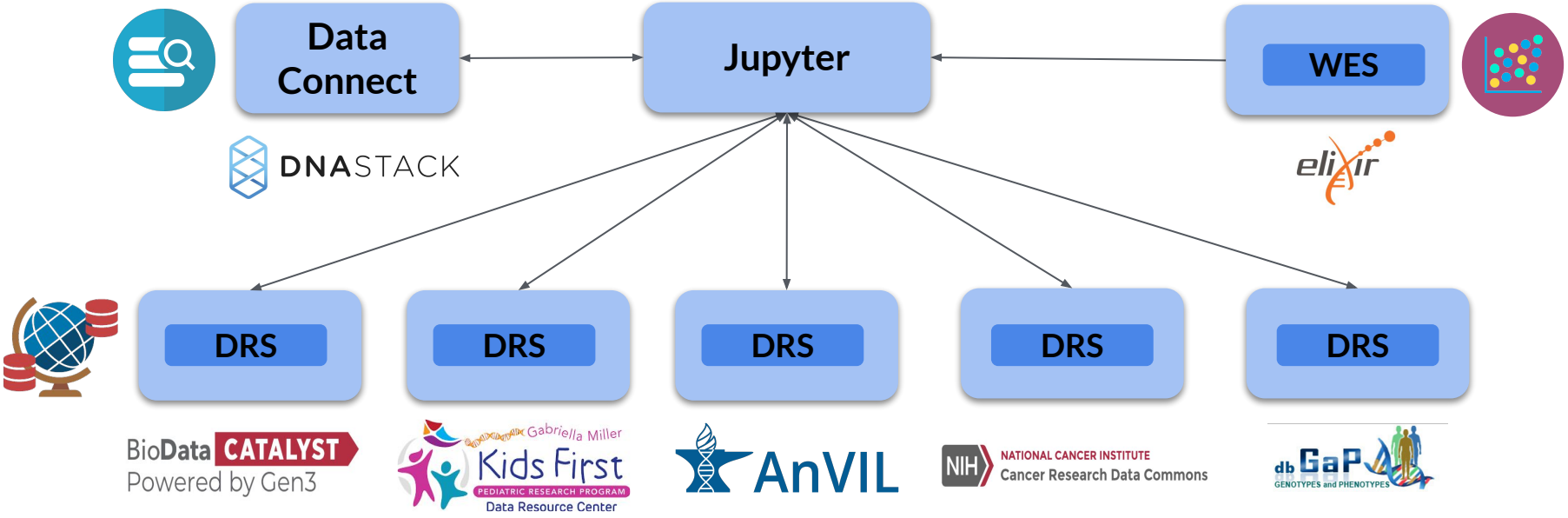
- Focused on a particular task
- Multiple vendors/systems
- Shows GA4GH standards used across systems with production and near-production implementations

Example - Horizontal Demo from FASP-Scripts

Global Alliance for Genomics & Health

Orchestrating GA4GH Services via Jupyter Notebooks

Multiple Data Repositories | Accessing Petabytes of Data

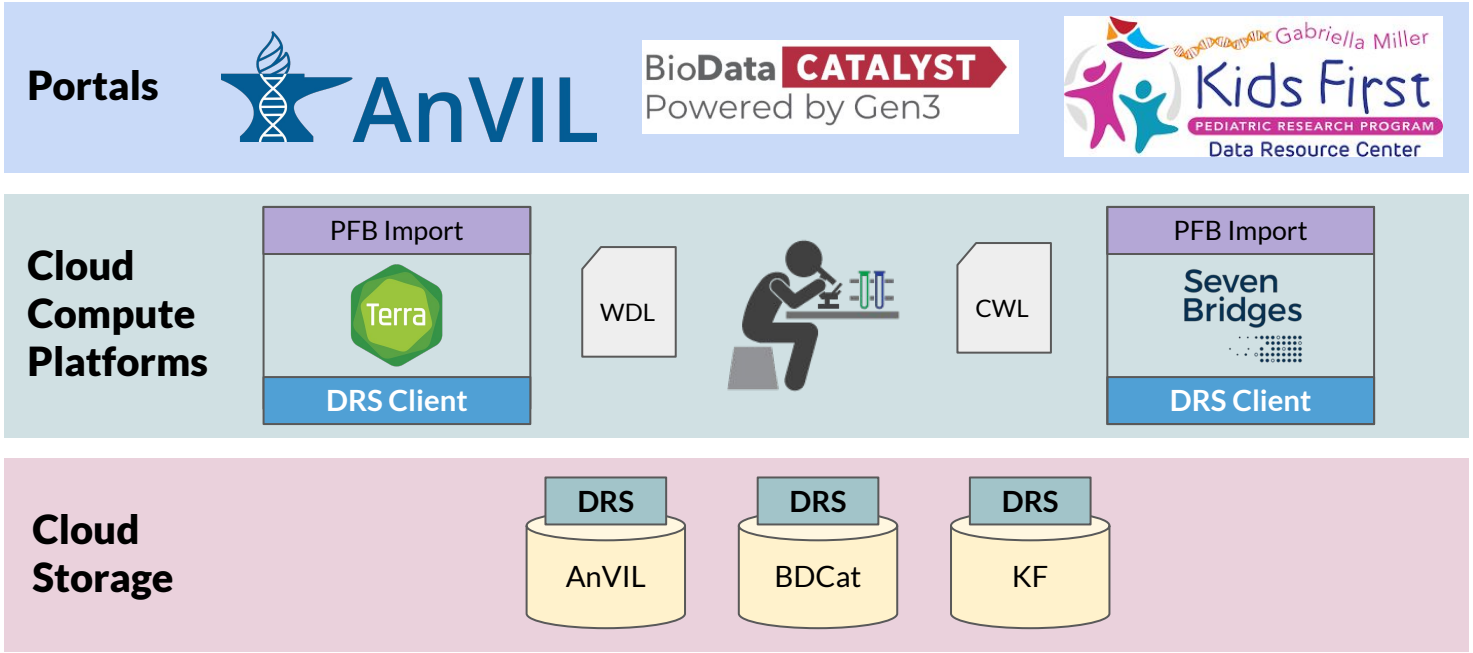


Example - Horizontal from NCPI

Global Alliance for Genomics & Health

Use Case #7: Tim Majarian's cross dataset analysis for Congenital Heart Disease

"We performed an association analysis, interrogating the effect of rare exonic variation on CHD risk at a fraction of the cost that would have otherwise been incurred without these interoperability tools."



① *Find Data*

② *Compute on Cloud Workspaces*

③ *Access Data on Cloud Storage*

Federated Analysis Systems Project (FASP)

Brian O'Connor
Sage Bionetworks

Max Barkley
DNAstack

Ian Fore
NCI, CBIIT

WES in Context - Demo from FASP-Scripts

Global Alliance for Genomics & Health

Available implementations of GA4GH components can be strung together to do something useful

- **Data Connect** - select cases and files
- **DRS** - workflow access to files
- **WES** - execute the workflow
- **DRS** - result retrieval

Complete notebook

https://github.com/ga4gh/fasp-scripts/blob/master/notebooks/GECCO_Gen3_on_SB.ipynb

Example - Horizontal Demo from FASP-Scripts

Global Alliance for Genomics & Health

Available implementations of GA4GH components can be strung together to do something useful

- Data Connect - select cases and files
- DRS - workflow access to files
- WES - execute the workflow
- DRS - result retrieval

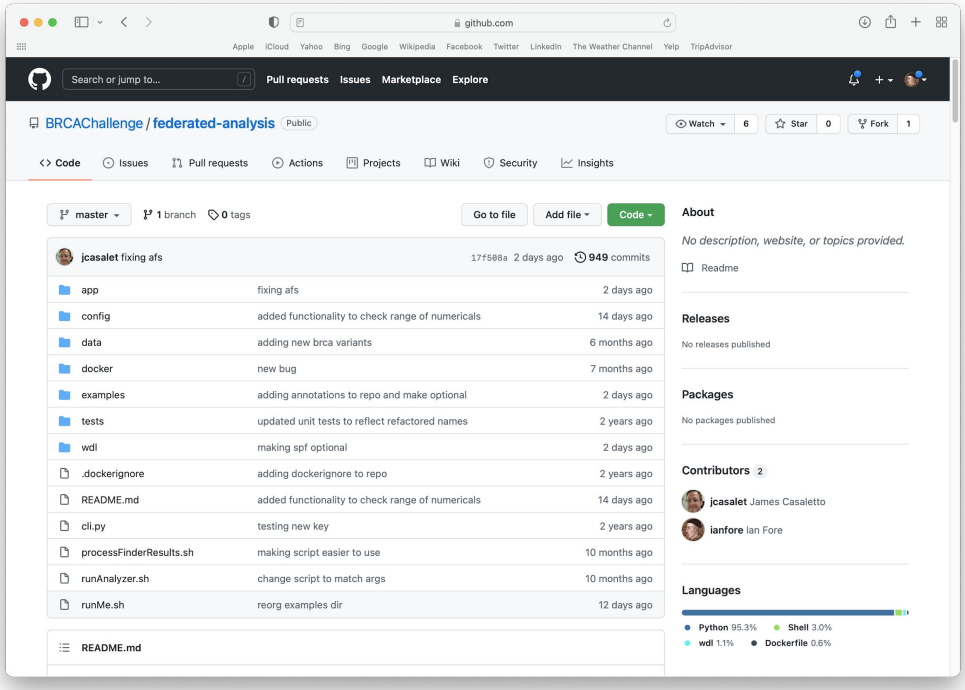
	acc	sample_id	sex	age	insert size average	insert size standard deviation		ge
0	SRR7271762	117454	Male	54	353.2	90.4	TE	4
1	SRR7271780	117477	Female	55	346.4	84.8	TE	5
2	SRR7271789	117486	Male	55	334.9	92.1	TE	5

Complete notebook

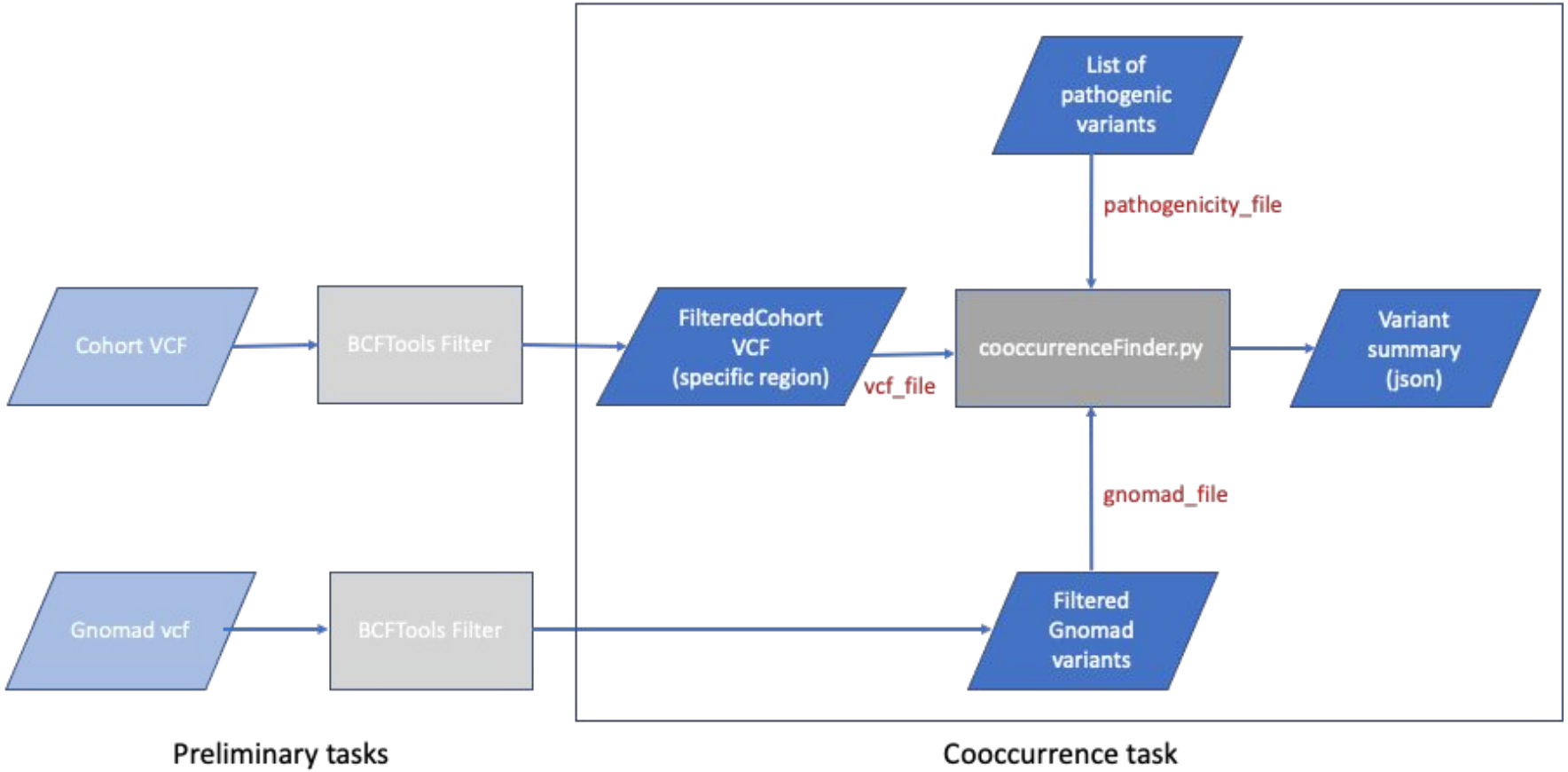
https://github.com/ga4gh/fasp-scripts/blob/master/notebooks/GECCO_Gen3_on_SB.ipynb

VUS container

BRCA Challenge – Federated Analysis - Melissa Cline, James Casaletto - UC Santa Cruz



https://github.com/BRCAChallenge/federated-analysis


List of pathogenic variants
pathogenicity_file
Cohort VCF
BCFTools Filter
FilteredCohort VCF (specific region)
vcf_file
cooccurrenceFinder.py
Variant summary (json)
gnomad_file
Gnomad vcf
BCFTools Filter
Filtered Gnomad variants
Preliminary tasks
Cooccurrence task

Pathogenic Germline Variants in 10,389 Adult Cancers

TCGA PanCanAtlas

Cell. Volume 173 Issue 2: p355–370.e14, 5 April 2018 10.1016/j.cell.2018.03.039

We conducted the largest investigation of predisposition variants in cancer to date, discovering 853 pathogenic or likely pathogenic variants in 8% of 10,389 cases from 33 cancer types. Twenty-one genes showed single or cross-cancer associations, including novel associations of SDHA in melanoma and PALB2 in stomach adenocarcinoma. The 659 predisposition variants and 18 additional large deletions in tumor suppressors, including ATM, BRCA1, and NF1, showed low gene expression and frequent (43%) loss of heterozygosity or biallelic two-hit events. We also discovered 33 such variants in oncogenes, including missenses in MET, RET, and PTPN11 associated with high gene expression. We nominated 47 additional predisposition variants from prioritized VUSs supported by multiple evidences involving case-control frequency, loss of heterozygosity, expression effect, and co-localization with mutations and modified residues. Our integrative approach links rare predisposition variants to functional consequences, informing future guidelines of variant classification and germline genetic testing in cancer.

About the Data

- Data Types and File Formats
- High Level Data Generation
- Data Dictionary
- GDC Data Processing
- Data Standards
- Data Sources
- Publications

See our Data Model

Supplemental Data

- **Data Files**
  - Compressed VCF file of the combined, filtered variant calls using GATK, VarScan2, and Pindel on WES data of the 10,389 final passed-QC samples. - PCA.r1.TCGAbarcode.merge.tnSwapCorrected.10389.vcf.gz
  - Tabix file of the compressed VCF file of the combined, filtered variant calls using GATK, VarScan2, and Pindel on WES data of the 10,389 final passed-QC samples. - PCA.r1.TCGAbarcode.merge.tnSwapCorrected.10389.vcf.gz.tbi
  - Prioritized, cancer related variants discovered in 10,389 cases. Please use "Overall_Classification" column to distinguish between Pathogenic, Likely Pathogenic and Priortizied VUSs. - PCA_pathVar_integrated_filtered_adjusted.tsv

COPDGene HMB
VCF files
In
BioDataCATALYST

Browse Data | Documentation | forei | Logout

NIH National Heart, Lung, and Blood Institute | BioData CATALYST Powered by Gen3

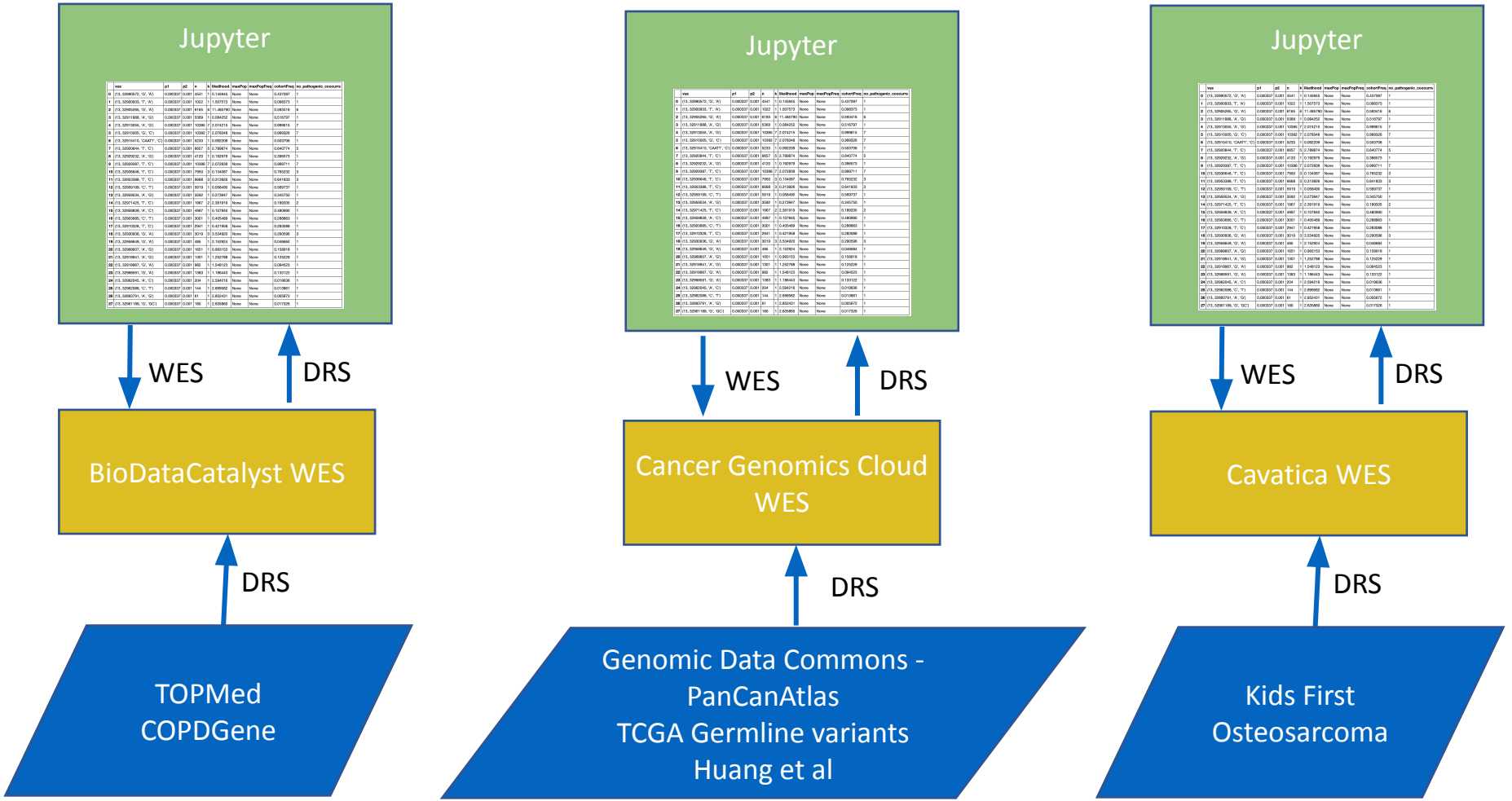
Dictionary | Exploration | Discovery | Workspace | Profile

Data | File

Explorer Filters | Data Tools | Summary Statistics | Table of Records

Project Id	File Name	File Size	GUID
topmed-COPDGene_HMB	COPDGene_phs000951_TOPMed_WGS_freeze.8.chr16.hg38.c1.vc	18.17 GB	dg.4503/2e9e9381-8ad0-4e08-a610-c5feff688788
topmed-COPDGene_HMB	phg001124.v1.TOPMed_WGS_COPDGene_v3.genotype-calls-vcf.WGS_markerset_grc37.c1.HMB-MDS.tar.gz	45.36 GB	dg.4503/6830a2b5-9789-41e2-befb-c28085e127db
topmed-COPDGene_HMB	COPDGene_phs000951_TOPMed_WGS_freeze.8.chr11.hg38.c1.vc	26.32 GB	dg.4503/49b64e41-9777-4516-aff3-47fd89604c1d
topmed-COPDGene_HMB	phg000794.v1.TOPMed_WGS_COPDGene.genotype-calls-vcf.WGS_markerset_grc37.c1.HMB-MDS.tar.gz	30.91 GB	dg.4503/8bf1f59b-c98a-44e0-9179-c690cb4eb1d7
topmed-COPDGene_HMB	COPDGene_phs000951_TOPMed_WGS_freeze.8.chr10.hg38.c1.vc	24.86 GB	dg.4503/8746aa6d-6164-46ef-b4da-ccd47afec05b
topmed-COPDGene_HMB	COPDGene_phs000951_TOPMed_WGS_freeze.8.chr5.hg38.c1.vcf.	32.92 GB	dg.4503/6aad51c2-2ea9-4248-8a38-7d52c10cfc89
topmed-COPDGene_HMB	COPDGene_phs000951_TOPMed_WGS_freeze.8.chr9.hg38.c1.vcf.	23.43 GB	dg.4503/59a1e9b4-d793-4d35-a7e8-3d6d5adfd221
topmed-COPDGene_HMB	COPDGene_phs000951_TOPMed_WGS_freeze.8.chr7.hg38.c1.vcf.	30.21 GB	dg.4503/09bda6d1-56cd-48cd-9592-06853cc4d347
topmed-COPDGene_HMB	COPDGene_phs000951_TOPMed_WGS_freeze.8.chr4.hg38.c1.vcf.	34.93 GB	dg.4503/ce11d3a6-177f-4d01-9c88-0344c9fbc5a2
topmed-COPDGene_HMB	COPDGene_phs000951_TOPMed_WGS_freeze.8.chr19.hg38.c1.vc	12.75 GB	dg.4503/5ab3b171-b7d5-447d-88a0-3616fe317969
topmed-COPDGene_HMB	COPDGene_phs000951_TOPMed_WGS_freeze.8.chr3.hg38.c1.vcf.	37.45 GB	dg.4503/feceeeca-a2b9-4e85-bbfe-53bd4fd4f747
topmed-COPDGene_HMB	COPDGene_phs000951_TOPMed_WGS_freeze.8.chr18.hg38.c1.vc	14.68 GB	dg.4503/5e488099-9c46-4440-8d24-8c82ca09f4a6
topmed-COPDGene_HMB	COPDGene_phs000951_TOPMed_WGS_freeze.8.chr14.hg38.c1.vc	16.97 GB	dg.4503/9a909fe9-68a7-4f39-aff6-acab7ea7ec6a
topmed-COPDGene_HMB	COPDGene_phs000951_TOPMed_WGS_freeze.8.chr22.hg38.c1.vc	7.93 GB	dg.4503/b192348c-47e3-41ff-bc9a-f6b92c7f4ef7
topmed-COPDGene_HMB	COPDGene_phs000951_TOPMed_WGS_freeze.8.chr12.hg38.c1.vc	25.46 GB	dg.4503/c05e6d6b-064c-4862-ad32-38cf72811a49
topmed-COPDGene_HMB	COPDGene_phs000951_TOPMed_WGS_freeze.8.chrX.hg38.c1.vcf.	20.37 GB	dg.4503/f2c0e270-3649-4be1-8a4c-8e066b7324b2
topmed-COPDGene_HMB	COPDGene_phs000951_TOPMed_WGS_freeze.8.chr20.hg38.c1.vc	11.94 GB	dg.4503/20e52355-af9f-4da0-9670-8b6f78859c3e
topmed-COPDGene_HMB	COPDGene_phs000951_TOPMed_WGS_freeze.8.chr15.hg38.c1.vc	16.23 GB	dg.4503/9be2bc5c-6da2-4a76-a096-436cfee157b9
topmed-COPDGene_HMB	COPDGene_phs000951_TOPMed_WGS_freeze.8.chr21.hg38.c1.vc	7.16 GB	dg.4503/1b3391fb-02d3-418a-b5d9-8d4274061c1a
topmed-COPDGene_HMB	COPDGene_phs000951_TOPMed_WGS_freeze.8.chr13.hg38.c1.vc	17.92 GB	dg.4503/fe831cc7-4c40-4a87-b19c-2aad305f6e82

Subjects
10,333



	ves	p1	p2	n	k	likelihood	maxPop	maxPopFreq	cohortFreq	no_pathogenic_cooccurrs
0	(13, 32969072, 'G', 'A')	0.000037	0.001	4541	1	0.169843	None	None	0.437097	1
1	(13, 32900933, 'T', 'A')	0.000037	0.001	1022	1	1.537573	None	None	0.098373	1
2	(13, 32992916, 'G', 'A')	0.000037	0.001	8166	6	11.695790	None	None	0.983416	8
3	(13, 32511888, 'A', 'G')	0.000037	0.001	5369	1	0.084252	None	None	0.516797	1
4	(13, 32913856, 'A', 'G')	0.000037	0.001	10386	7	2.074215	None	None	0.998615	7
5	(13, 32910505, 'G', 'C')	0.000037	0.001	10382	7	2.076348	None	None	0.999326	7
6	(13, 32910410, 'CAATT', 'C')	0.000037	0.001	8253	5	0.092208	None	None	0.953706	5
7	(13, 32920844, 'T', 'C')	0.000037	0.001	8657	5	0.786874	None	None	0.840774	5
8	(13, 32929232, 'A', 'G')	0.000037	0.001	4133	1	0.193979	None	None	0.399973	1
9	(13, 32920937, 'T', 'C')	0.000037	0.001	10386	7	2.075839	None	None	0.999711	7
10	(13, 32909546, 'T', 'C')	0.000037	0.001	7993	3	0.134397	None	None	0.769232	3
11	(13, 32953398, 'T', 'C')	0.000037	0.001	8668	3	0.313926	None	None	0.841833	3
12	(13, 32959195, 'C', 'T')	0.000037	0.001	9919	1	0.056495	None	None	0.958737	1
13	(13, 32959534, 'A', 'G')	0.000037	0.001	3582	1	0.273947	None	None	0.345750	1
14	(13, 32971425, 'T', 'C')	0.000037	0.001	1967	2	2.391919	None	None	0.189335	2
15	(13, 32958838, 'A', 'C')	0.000037	0.001	4967	1	0.107843	None	None	0.480990	1
16	(13, 32935585, 'C', 'T')	0.000037	0.001	3001	1	0.425489	None	None	0.289803	1
17	(13, 32910326, 'T', 'C')	0.000037	0.001	2941	1	0.431968	None	None	0.283086	1
18	(13, 32930936, 'G', 'A')	0.000037	0.001	3019	3	3.534825	None	None	0.290596	3
19	(13, 32959618, 'G', 'A')	0.000037	0.001	486	1	2.132924	None	None	0.046984	1
20	(13, 32958607, 'A', 'G')	0.000037	0.001	1051	1	0.950153	None	None	0.158916	1
21	(13, 32919841, 'A', 'G')	0.000037	0.001	1361	1	1.252708	None	None	0.129229	1
22	(13, 32910967, 'G', 'A')	0.000037	0.001	862	1	1.548123	None	None	0.084523	1
23	(13, 32968991, 'G', 'A')	0.000037	0.001	1363	1	1.166443	None	None	0.133122	1
24	(13, 32982545, 'A', 'C')	0.000037	0.001	204	2	2.594216	None	None	0.019836	1
25	(13, 32962596, 'C', 'T')	0.000037	0.001	144	2	2.699562	None	None	0.013861	1
26	(13, 32983791, 'A', 'G')	0.000037	0.001	61	1	2.652421	None	None	0.005872	1
27	(13, 32981188, 'G', 'GC')	0.000037	0.001	166	2	2.630968	None	None	0.017326	1

Merged analyses

	n			cohortFreq			no_pathogenic_coocurrs		
source	**copdgene**	**osteosarcoma**	**tcga**	**copdgene**	**osteosarcoma**	**tcga**	**copdgene**	**osteosarcoma**	**tcga**
vus									
('13', 32315831, 'G', 'A')	9345.0	28.0	NaN	0.914742	0.848485	NaN	1.0	1.0	NaN
('13', 32318080, 'C', 'T')	10195.0	33.0	NaN	0.997944	1.000000	NaN	1.0	1.0	NaN
('13', 32318598, 'T', 'C')	9047.0	32.0	NaN	0.885572	0.969697	NaN	1.0	1.0	NaN
('13', 32319654, 'A', 'G')	8119.0	NaN	NaN	0.794734	NaN	NaN	1.0	NaN	NaN
('13', 32321240, 'G', 'C')	10196.0	33.0	NaN	0.998042	1.000000	NaN	1.0	1.0	NaN
('13', 32323151, 'ATT', 'A')	4106.0	NaN	NaN	0.401919	NaN	NaN	1.0	NaN	NaN
('13', 32325741, 'C', 'T')	10194.0	33.0	NaN	0.997847	1.000000	NaN	1.0	1.0	NaN
('13', 32331128, 'G', 'A')	10194.0	33.0	NaN	0.997847	1.000000	NaN	1.0	1.0	NaN
('13', 32333969, 'A', 'G')	10194.0	33.0	NaN	0.997847	1.000000	NaN	1.0	1.0	NaN
('13', 32338918, 'A', 'G')	10194.0	33.0	NaN	0.997847	1.000000	NaN	1.0	1.0	NaN
('13', 32340868, 'G', 'C')	10193.0	33.0	NaN	0.997749	1.000000	NaN	1.0	1.0	NaN
('13', 32342270, 'CAAA', 'CA')	NaN	24.0	NaN	NaN	0.727273	NaN	NaN	1.0	NaN
('13', 32343709, 'G', 'GA')	635.0	NaN	NaN	0.062157	NaN	NaN	1.0	NaN	NaN
('13', 32343709, 'GA', 'G')	3580.0	NaN	NaN	0.350431	NaN	NaN	1.0	NaN	NaN
('13', 32344166, 'GA', 'G')	7460.0	NaN	NaN	0.730227	NaN	NaN	1.0	NaN	NaN
('13', 32345879, 'G', 'A')	10194.0	33.0	NaN	0.997847	1.000000	NaN	1.0	1.0	NaN
('13', 32346707, 'T', 'C')	10194.0	33.0	NaN	0.997847	1.000000	NaN	1.0	1.0	NaN
('13', 32353757, 'C', 'T')	5163.0	NaN	NaN	0.505384	NaN	NaN	1.0	NaN	NaN
('13', 32354180, 'T', 'C')	10202.0	33.0	NaN	0.998630	1.000000	NaN	1.0	1.0	NaN

FASP-scripts

Global Alliance for Genomics & Health

For more examples see...
https://github.com/ga4gh/fasp-scripts

Chapter 6:

Ongoing Development with WES

SageBionetworks

2022 Strategic Roadmap priorities — GA4GH Cloud

- Continue to **expand the registry** of known implementations and instances of GA4GH Cloud APIs
- **Continue to refine and release versions of APIs** based on FASP and Driver Project feedback
- **Encourage new Driver Projects to leverage Cloud APIs** and providers to **create new implementations**
  - *Support **"Starter Kit"** implementations*
  - *Support **FASP demos** that use Cloud APIs*
- **Continue to publish new API releases ~2x year**, posts on wiki, and white paper describing APIs and best practices

Current keepers of WES

Patrick Magee
DNAstack

Alex Kanitz
Elixir

44 Open ✓ 64 Closed — Author · Label · Projects · Milestones · Assignee · Sort

- Migrate build docs over to github actions [Type: Documentation]
  #178 opened 16 days ago by patmagee
- Input and Output Format Specification [EPIC] — 9
  #176 opened on Feb 23 by patmagee
- Support passing TRS URIs to workflow_url [Type: Documentation] — 4
  #175 opened on Dec 8, 2021 by uniqueg
- Add API call for workflow and inputs validation — 2
  #174 opened on Oct 14, 2021 by markjschreiber
- Support and standard keyword for WES — 15
  #173 opened on Oct 6, 2021 by jb-adams
- Pass TES endpoints to WES request — 2
  #170 opened on May 6, 2021 by uniqueg
- Cross-Cloud Workstream [EPIC]
  #168 opened on Jan 26, 2021 by ruchim
- TES x WES [EPIC] — 1

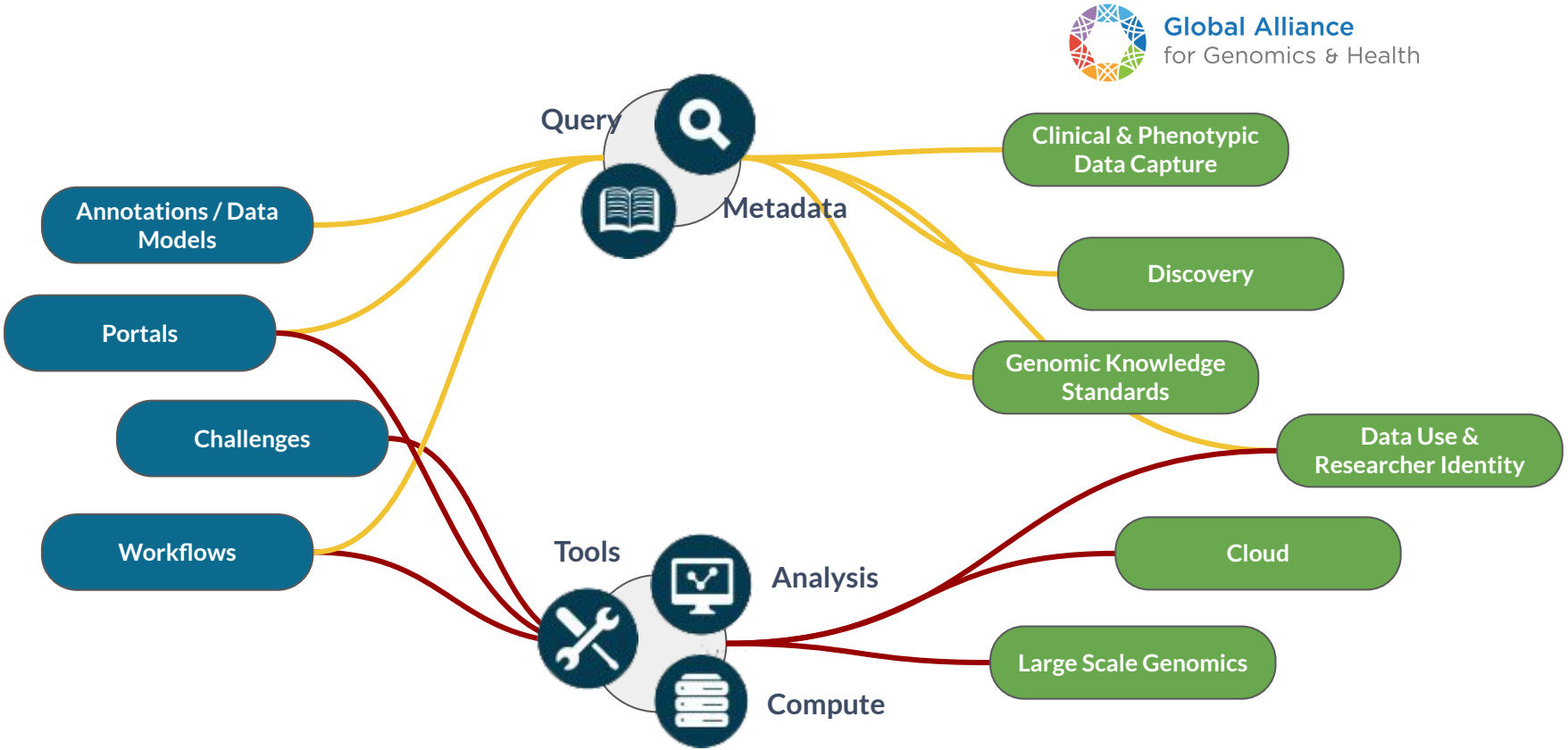
Active development/discussion items

- Standardizing workflow parameters format [***WES#161***]
- Improved input/output format specification (and validation) [***WES#176***]
- Better integration with TRS [***WES#175***]
- Improved paging for workflow run logs [***WES#177***]
- Integration with DRS and GA4GH Passports [***WES#156***, ***DRS#294***, ***DRSxWES***]
- Many more open tickets…

2022 Strategic Roadmap Priorities — WES

- **Key roadmap priorities:**
  1. (Q1) Improve usability of Run Listing API (Driver: Elixir)
  2. Passport integration (Driver: Elixir)
  3. (Q1) Support pagination for tasks within a workflow
  4. (Q1) Start discussion on API scalability (Driver: Elixir, implementers: Seven Bridges, AWS, Terra)
  5. (Q1) Start discussion on unified input/output format and how proscriptive WES should be
- **Year-end goal:** solution for providing implementers with a scalable API
- **Next steps:** create forums to discuss implications of (4) and (5) best practices

Get involved!



Work Stream meeting minutes openly available on ga4gh.org — check some out and join a call!

Thanks to …

Acknowledgements

Sage Bionetworks
Brian O'Connor
Jake Albrecht
Thomas Yu
Bruno Grande

UCLA
Paul Boutros

GA4GH
Ian Fore (NCI)
Patrick Magee (DNAstack)
Alex Kanitz (Elixir)
Ruchi Munshi (Broad)
Walt Shands (Dockstore)
Max Barkley (DNAstack)
David Glazer (Verily)
Jeff Gentry (Foundation Medicine)
Peter Amstutz (Curii)
Denis Yuen (Dockstore)
Pratik Soares (Illumina)
Many others!

Questions?