

Connect Agenda

Getting Started with GA4GH APIs

Date & Time: November 15th, 2022 at 2:30pm - 4pm ET ([click here for your local time](#))

Chairs Ian Fore, Brian O'Connor

**Join from PC, Mac,
Linux, iOS or Android:**

Dial in by phone:

Meeting ID:

Password:

Slides:  Getting Started with GA4GH APIs Slide Deck

Meeting Recording:

Meeting Transcript

Attendees: Kathy Reinold, Gerrit Botha (UCT/eLwazi), Angela Page (GA4GH/Broad), Daniel Park (Broad/eLwazi), Miro Cupak (DNASTACK), Patrick Cheung (Philips), Raja Kathirvel (Philips), Ray Krasinski (Philips), Tim Slade (RTI), Tony Handstad, Yash Puligundla (GA4GH/Broad), Jeremy Adams (DNASTACK), Justina Richardson, Teresa Abok, Evan Christensen (University of Utah), Mamana Mbiyavanga (UCT/H3Africa), Nicky Mulder (UCT/H3Africa/eLwazi)

Aim of the Meeting: To understand needs of:

- Projects that want to make data available by deploying implementations of GA4GH APIs
- Users that want to perform tasks that access multiple sources via GA4GH APIs

Do this via

- Understand the role and purpose of GA4GH starter kit
- Gather feedback on starter kit

- Identify areas where users may need support in implementing standards within their environments
 - What use cases?
 - Who?
 - What APIs are needed?
 - What APIs/features are missing?

Meeting Takeaways

- Today's discussion contributed towards the identification of "blue bars"
- Potential "blue bars" include annotation and sequence variants, synthetic cohort handoff, use and availability of APIs in a health care context at Philips, eLwazi

Actions:

- ☐ Plan for April Connect
 - ☐ TBD as the role of FASP is evolving
 - ☐ x
- ☐ Plan follow-up meetings
 - ☐ X
 - ☐ x
- ☐ Any other business arising
 - ☐ X
 - ☐ x

Session description

In this session, Yash Puligundla, a member of the GA4GH technical team, will provide a detailed overview of starter kit and gather feedback from members on any other types of support required by contributors. Following this, the group will be encouraged to discuss use cases that may benefit from the implementation of GA4HG standards. This discussion will be used to identify use cases and areas where further FASP follow up, or other GA4GH follow up, could be beneficial.

Workshop Agenda

	Agenda Item	Presented by	Item Type	Time
1.0	Introduction	Chairs		
2.0	Resources Review			30 mins
2.1	• Detailed overview of starter kit	Yash Puligundla	Presentation	

2.2	Starter kit feedback gathering	Yash Puligundla/all	Guided Discussion	
3.0	Group discussion prompt - What is the use case you are interested in? - Who is interested in working on this? - What APIs are needed? - What APIs/features are missing?	All	Guided Discussion	1 hour
3.1	Intro - Structure Overview	Angela Page	Presentation	10 minutes
3.2	Possible use cases to discuss Other groups are welcome to discuss these questions!! Bold were discussed			
	H3Africa - General API adoption and Federated imputation – confirmed - Nicky Mulder and/or Gerrit Botha from H3Africa/eLwazi			
	- ELIXIR - Access control - consider the use cases different orgs may be facing over the next couple of years - Better Cloud API support for workflow and/or task federation - Alex Kanitz			
	- Variants of Uncertain Significance (VUS) Classification – confirmed topic - Melissa Cline			
	- Undiagnosed Disease Network (UDN) Kids First / Common Fund – possible topic - Valerie Cotton, Chris Kinsinger			
	Synthetic cohort handoff between systems. Kathy brought up as			

	possible use case... maybe cross over with NCPI? – Kathy Reinold			
	• Annotation and sequence variants. Major challenge GFF3 col 9 (specification: https://github.com/The-Sequence-ontology/Specifications/blob/master/gff3.md)!! Just a list of values but high value. Looking to develop more structured format there. Collaboration in the future, could use something like a GIF use case around exchange between systems. See agenda for the GKS session during connect: https://docs.google.com/document/d/1TPmV3-5MN5iMitBtVsvj_RwVrcG4n8le/edit – Evan Christensen			
	•			
	• <i>Any other groups have proposed use cases to discuss?</i>			
4.0	Wrapup	Chairs		

Background

The following diagram illustrates an evolving strategy for applying GA4GH standards to specific projects

Rather than requiring coding of new tools for these projects they would be addressed by combinations of the following, as appropriate to the project.

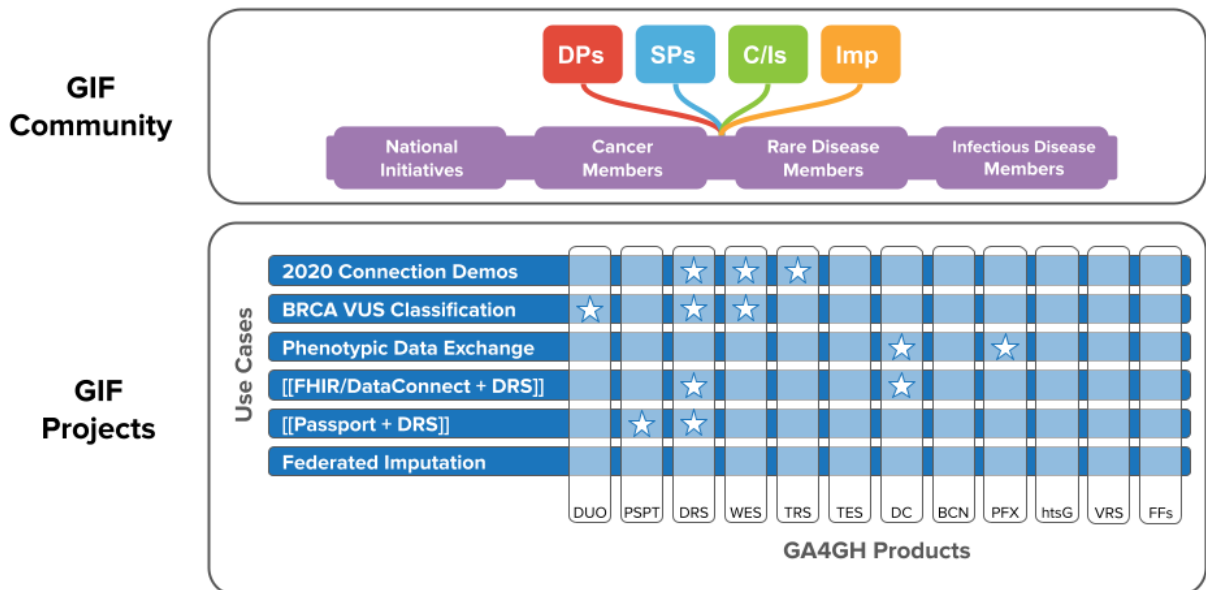
- Deployment of existing implementations
- Access to data in existing implementations
- Developing capability within project teams
 - Training
 - Client code examples

The session will explore

- Questions for Users/Champions/Project Leads

- Tech team questions re:starter kit and feedback (Yash)
- Would tutorials be helpful?
- What other kind of support do you need?
 - Role of assigned experts?

GA4GH Implementation Forum (GIF)



A review of the following NCPI use cases would be suitable for the approach described above. Increased awareness of the resources available would allow the originators of these projects to move forward with them under their own momentum.

Controlled Access FHIR

https://github.com/NIH-NCPI/NCPI_use_case_tracker/issues/29

Search studies by variables

https://github.com/NIH-NCPI/NCPI_use_case_tracker/issues/25

Determine if a study has enough of the type of data I need to motivate applying for access

https://github.com/NIH-NCPI/NCPI_use_case_tracker/issues/24

Genome-wide Sequencing Analysis to Identify the Genes Responsible for Enchondromatoses and Related Malignant Tumors

https://github.com/NIH-NCPI/NCPI_use_case_tracker/issues/21

Uniform search of public sample and sequence read information across NCBI and Kids First repositories

https://github.com/NIH-NCPI/NCPI_use_case_tracker/issues/19

SRA & Kids First DRC for Kids First & UDN co-analysis

https://github.com/NIH-NCPI/NCPI_use_case_tracker/issues/10

This is well advanced in many areas and a notebook (fasp-scripts) exists to perform analysis via WES.

Whole slide images

https://github.com/NIH-NCPI/NCPI_use_case_tracker/issues/9

Notes

Let's keep notes from the session here:

- Yash presenting the resources
 - FASP scripts
 - Starter kit

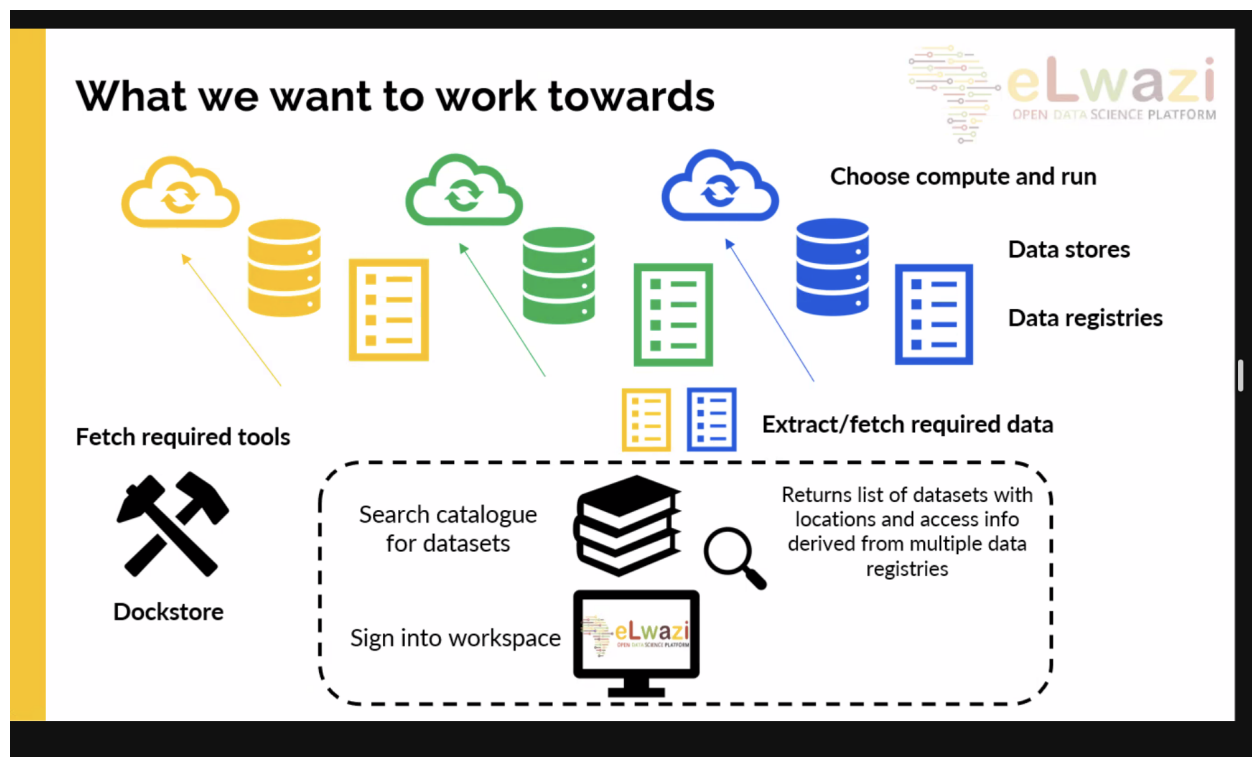
GA4GH standards used in this tutorial			
 Global Alliance for Genomics & Health			
Standard	Standard Version implemented in the starter kit	Starter Kit (GitHub and Docker Hub)	Starter Kit Version
Data Repository Service (DRS)	1.3experimental	ga4gh/ ga4gh-starter-kit-drs	0.3.2
Workflow Execution Service (WES)	1.0.0	ga4gh/ ga4gh-starter-kit-wes	0.3.2
Data Connect	1.0.0	ga4gh/ ga4gh-starter-kit-data-connect	0.1.1
Passports	1.2	ga4gh/ ga4gh-starter-kit-passport-broker	0.0.2
		ga4gh/ ga4gh-starter-kit-passport-ui	0.0.2

ga4gh.org

- Fasp-clients are client wrappers around the APIs
 - Fasp-scripts show analysis examples
 - Yash walked through the available implementations of APIs in the starter kit
- Discussion with Ian
 - Discussion questions
 - Data providers or starter kit users... can we get your feedback?

- Gerrit Botha (eLwazi) has been using WES and DRS starter kits... gives good indication of what's needed for implementation. Interested in Passports next.
 - Group discussion about what missing APIs... interest in Data Connect from Kathy for example
- Structure overview with Angela
 - More activities than workstreams... more contributors than driver projects
 - Want to recognize contributions
 - Structure
 - Study groups
 - Work streams
 - Now (new) Implementation Forum
 - Brings together FASP and GHIF
 - Replacement for both
 - Blue Bars: GA4GH Implementation Forum (GIF)
 - Organizing into GIF use cases, FASP was an example of that.
 - GIF use case... mission driven use case, committed partners, and ga4gh products (+product gaps)
 - GA4GH looking for well-defined Blue Bars beyond just FASP...
 - Discussion today looking for group to identify examples
- Use Case discussion
 - Annotation and sequence variants – Evan
 - Synthetic cohort handoff – Kathy
 - Use and availability of APIs in a health care context at Philips – Ray Krasinski
 - Wants to see more 1) use cases and 2) example docs/implementations like starter kit
 - Federated analysis... DNASTack interested in this. Autism sharing driver project –Jeremy Adams
 - eLwazi – Nicky Mulder and Gerrit Botha
 - Terra workspace
 - Provides access to data on the cloud
 - Use of DRS, TRS, WES ... Terra tied to GCP...
 - Second workspace... gen3... provides AWS
 - Third type... local HPC
 - Lack workspace
 - Lack common APIs above
 - Desire
 - One workspace to find all the datasets across eLwazi
 - Want to be able to compute on those data using workflows from Dockstore
 - There isn't a single frontend to do this right now
 - Want to have GA4GH APIs to figure out what data and where to compute
 - What APIs

- Data Connect ... for the central catalog implementation
- DRS for data access
- WES for compute
- TRS for workflow sharing
- Passports for who you are, authN/Z
- Need to decide which implementations of the above... starter kit, Terra, Gen3?
 - They have HPC centers to partner with on this
 - Gen3 exists and can be used
 - Are there other systems that can be used? Lifebit CloudOS? Starter kit for production?
- Need data locality in this project, that is important
- In 3 years... want to have a real ecosystem where each environment provides APIs and Terra or another client can compute on them
- What's missing
 - Frontend experience
 - Logic layer to direct work to the “right” place? This is what Nicky showed below:



- This may be a new set of standards or just extensions of DRS, WES, etc to have more geographic information, cost information, etc. Could be linked to discovery.
- Federated Imputation use case – Nicky Mulder and Gerrit Botha
 - Situation: certain panels, only allowed at certain compute facilities...
 - TOPMed imputation panel

- African imputation panel
 - Need to be able to federate the imputation
- Can we use the default APIs?
 - DRS, AAI/Passports to share data
 - Merged data... something needs to provide the merged data... only the person that requested imputation to be done should be able to access this
- Looking at a few options, tools
- Need to
 - Find a better way to merge
 - homomorphic encryption
 - Preference now is to find a federated way of imputation
 - APIs are not the problem... need to figure out the merge process.
- Observation, using both panels at the same time gave better results
- Could submissions be encrypted so the imputation servers can encrypt data from users
- So there's two ways to approach
 - 1) compute on both panels at the same time in some way that's secure
 - Or
 - 2) impute at each location but figure out how to merge after
- APIs from GA4GH
 - Maybe encryption would be the key thing to add