MEETING REPORT

GA4GH April Connect 2022

The 2022 GA4GH April Connect meeting was a hybrid event, with in person attendees joining in Montreal, Canada. The meeting brought together members of the GA4GH Work Streams, Driver Projects, implementers, and the broader community to collaborate and advance standards development work, with a focus on cross-Work Stream initiatives. This document includes an overview of all sessions, with links to recordings, slides, and meeting agendas and minutes.

Session Table of Contents

Opening Remarks & GA4GH Updates

Fist Full of Categorical Variants

Implementing Data Connect - Levels of Harmonisation

FASP Collaboration on Passports, DRS, and Discovery

Strategic Refresh Brainstorm

REWS General Meeting

Cloud Work Stream Roadmapping for 2022

Bridging Data & Workflows from Research into Healthcare

Building Globally Impactful Standards through Inclusion

GA4GH Product Development & Approval

Hang 'em High Standards

Report Backs & Looking Forward

The Stable, The Draft, and the Trial Use

Update on Technical Direction of Passports v1.2

For a Few Standards More

Testbeds and Validation of GA4GH Standards

Unpacking the Many Meanings of Genetic Discrimination

FASP Use Case Session

GA4GH Cloud API Implementers Workshop

Computable Cohort Representation Hackathon Kick-Off

Standardising Approaches to Sharing Data Models via APIs

20 APRIL 2022

Opening remarks and GA4GH updates

Recording • **Slides** • **Agenda**

GA4GH Chief Standards Officer Susan Fairley highlighted developments the GA4GH community has achieved since the first Work Stream Firehose session in January 2018, such as progress reported in a special issue of the journal *Cell Genomics* in November 2021. In order to move toward a future of seamless federated genomic analysis, the next phase of work involves engagement, integration and filling gaps, and building community. The session continued with updates from the GA4GH Equity, Diversity, and Inclusion (EDI) Advisory Group, Federated Analysis Systems Project (FASP), and Work Streams.

Strategic Refresh Brainstorm

Recording • **Slides**

GA4GH Steering Committee Chair Ewan Birney chaired the session, which was one in a series of ten open town hall brainstorms to gather input from the community on the focus and directionof GA4GH. Through a gap analysis in 2020 chaired by Heidi Rehm and Andrew Morris, the GA4GH community previously identified three strategic priorities: interoperability and technical alignment, implementation support, and engagement with healthcare and clinical standards. The 2022 strategic refresh sessions provide an opportunity for members to share feedback on GA4GH's success with these priorities so far, and provide suggestions for improvement. This session focused heavily on promoting implementation around the globe.

Building globally impactful standards through inclusion

Recording • **Agenda**

Partnering with the EDI Advisory Group and Regulatory and Ethics Work Stream (REWS) Diversity Group, a representative from Aleria led an interactive workshop on how increasing inclusion can boost both diversity and performance within an organisation. The speaker emphasised that inclusion is invisible, with some groups experiencing exclusion much more commonly and acutely than other groups. A review of case studies and a sample measurement of exclusion experiences among attendees highlighted challenges and opportunities for incorporating equity, diversity, and inclusion into GA4GH standards creation.

GA4GH Cloud API implementers workshop

Recording • **Slides** • **Minutes / Agenda**

Implementers of GA4GH Cloud deliverables such as the Workflow Execution Service (WES) and Tool Registry Service (TRS) APIs presented their experiences with real-world use cases. Representatives from Amazon Web Services (AWS), DNAStack, Broad/Cromwell, and Dockstore shared their target users, goals, challenges, and suggestions for GA4GH Cloud APIs. The group then discussed pain points and opportunities for improvement in greater detail. Implementers will next focus on submitting pull requests and contributing to the Cloud Work Stream Roadmap, especially for the next release of WES.

Testbeds and validation of GA4GH standards

Recording • **Slides**

The session began with several brief presentations from current GA4GH testing initiatives, including Acidbio/BED, VCF validator, Refget, RNAget, and the recently approved Beacon v2.0. A demonstration of a testbed infrastructure provided a proof of concept for the unification of GA4GH testing initiatives. The group discussed the next steps for testbed infrastructure, which included sharing building blocks, interacting with journal editors, defining file format accompaniments, and harmonising efforts to reduce workload.

The stable, the draft and the trial use (ClinGen and VICC implementations of GKS standards)

Recording • **Slides**

The session chairs walked through plans by two Driver Projects — Variant Interpretation for Cancer Consortium (VICC) and Clinical Genome Resource (ClinGen) — to adopt Genomic Knowledge Standards (GKS) within the framework of a maturity model. The structure and terms of the maturity model are aligned to those seen in the HL7 FHIR specifications, which include three maturity level categories: draft, trial use, or stable. Next, members of the GKS Work Stream aim to draft classes used in the Variation Representation Standard (VRS) (and associated VRSATILE tooling) and Variant Annotation (VA) standard with the maturity model in mind.

GA4GH product development and approval

Recording • **Agenda**

GA4GH is currently reviewing and revising the process used to propose, develop, and approve standards. Chief Standards Officer (CSO) Susan Fairly discussed the detail of the updated GA4GH product development and approval document. The modifications include who votes on product

approvals, the timelines of approvals, how to make changes to approval decisions, and product documentation. The discussion notes and feedback were recorded using a miro board. The first draft of the new product approval is now open for comments here.

Fist full of categorical variants (VRS)

Recording • **Slides** • **Minutes / Agenda**

Members of the Genomic Knowledge Standards (GKS) Work Stream discussed the genomic knowledge search problem and proposed a mechanism for representing variation as a categorical concept, or definition. This approach would provide a framework for computable descriptions of variation. Moving forward, the team aims to develop a working community around categorical variation and to draft and vet several categorical variation models.

REWS general meeting

Recording • **Driver Project Slides** • **GHIF and ELSI Slides** • **Minutes / Agenda**

The Regulatory and Ethics Work Stream (REWS) met to share updates on current activities, hear from Driver Projects, and give a preview of new groups. The Data Access Committee Review Standards (DACReS) team requested feedback via survey, and the Genetic Discrimination group invited members to a working meeting to draft the detailed plan of a short commentary that would define the notion of genetic discrimination. Presentations from the GDPR Forum, Consent Toolkit team, and REWS Diversity group covered recent progress on projects. Several new groups are being launched: in 2022, Diversity in Datasets, Ethical Provenance Toolkit, Public Attitudes for Genomic Policy, Clinical Data Sharing and Consent; and in 2023, Newborn Whole Genome Sequencing. Finally, representatives from Australian Genomics, EUCANCan, the Human Cell Atlas (HCA), and EpiShare gave overviews of their efforts.

21 APRIL 2022

Update on the technical direction of Passports v1.2

Recording • **Slides**

The session began with a brief overview of OpenID Connect, OAuth 2.0, JSON Web Tokens, JSON Web Signature, and Claims. The revisions introduced in Passports v1.2 will be minor and will not need approval from the GA4GH Steering Committee (SC). The revision allows Passports to be issued as signed tokens using RFC 8693 “OAuth 2.0 Token Exchange“ mechanism, which will allow mutually trusting systems to pass signed tokens among them. Passport v2.0 will include

major edits requiring the SC's approval. This includes the introduction of a new type of token, a work-order token, which will be limited to specific datasets, used by specified agents only, and small enough to fit into HTTP headers. This new token will also be issued using the OAuth 2.0 Token Exchange mechanism that will be introduced in v1.2.

Unpacking the many meanings of genetic discrimination (Genetic Discrimination Group)

Recording • **Minutes / Agenda**

The Genetic Discrimination (GD) group reached a consensus on a person-centric approach to focus on discrimination from the perspective of those who experience it. There was also a discussion on the need to define genetic discrimination broadly and the need to acknowledge issues of public concerns surrounding GD. The majority of the session focused on preparing an outline for an EDI bioethics publication.

Computable cohort representation hackathon kickoff

Recording • **Minutes / Agenda**

The recently formed Computable Cohort Representation group kicked off a hackathon series to tackle a number of questions, including: how does one represent a set of patient attributes to assemble a cohort from different systems and datasets? How does one formulate the queries? Pain points, such as gaining access, to data which might block the primary goal were also identified. Preparatory work was described. Over the course of several weeks, the team will host five hackathons on key topical areas in order to show an example phenotype across different definitions, specify the phenotype in an executable format, and test across implementations in a federated manner. To learn more about hackathon plans or to get involved, you can read the team's planning document.

Implementing Data Connect — levels of harmonization

Recording • **Slides** • **Minutes / Agenda**

The Data Connect team walked through several realistic datasets to illustrate the ease of sharing data using the Data Connect reference implementation. The ability of data providers to use Data Connect whether or not they have full semantic annotation or harmonized data was shown. This illustrated that Data Connect allows for inclusivity, and a pathway to more advanced levels of harmonization as and when providers are able to take that resource intensive step. Participants discussed issues related to identifiers, and prototype exploration of schema representation on top of Data Connect's use of json-schema.

Hang ‘em high standards (Variation Annotation)

Recording • **Slides** • **Minutes / Agenda**

Contributors to the Variation Annotation (VA) subgroup of the Genomic Knowledge Standards (GKS) Work Stream presented the current state of and future plans for the VA specification. For the initial release, the group plans to focus on creating a new modelling framework with base profiles (such as pathogenicity or therapeutic response profiles) that basic users can adopt, and that power users can adapt into their own specialised models to share back with the community. A lively discussion addressed the connection between VA and file formats such as VCF. Before GA4GH 10th Plenary, the team plans to draft the framework and base profiles, incorporating VRS Added Tools for Interoperable Loquacious Exchange (VRSATILE) value object descriptors.

Cloud Work Stream roadmapping for 2022

Recording • **Slides** • **Minutes / Agenda**

Attendees discussed updates to the completed suite of Cloud APIs, which have all been approved and are in maintenance mode. This year, the Data Repository Service (DRS) team aims to release v1.3 and refine integration with the Passports specification. The Workflow Execution Service (WES) API team aims to look into workflow and engine parameters. The Tool Registry Service (TRS) API team is looking at a potential breaking change, to account for more possible representations of a workflow resource. The Task Execution Service (TES) API group is looking at a number of updates to roll into the next release, including tag filtering.

For a few standards more (GKS)

Recording • **Minutes / Agenda**

Members of the Genomic Knowledge Standards (GKS) Work Stream and other attendees reviewed outcomes from their previous Connect sessions and developed a roadmap for 2022, with an aim to ratify and adopt a maturity and data versioning model to help current and potential adopters implement the evolving GKS standards. By GA4GH 10th Plenary, the team also hopes to draft a formal VRSATILE specification, which supports organisations who wish to adopt VRS.

FASP use case session

Recording • **Slides** • **Minutes / Agenda** • **Miro Board**

Representatives from the Beacon v2 standard, Genomics England, the Autism Sharing Initiative, the NIH Cloud Platform Interoperability Effort (NCPI), and Human Heredity & Health in Africa

(H3Africa) presented use cases to the Federated Analysis Systems Project (FASP) group. Participants asked questions, identified patterns, and generated ideas for future demonstrations that build on user interest in improved data access and federated analysis, including in the areas of end-to-end gene panels and federated imputation.

22 APRIL 2022

GA4GH standardizing approaches to sharing data models via APIs

Recording • **Slides** • **Minutes / Agenda**

At this session, contributors discussed ways to achieve common approaches to share data and the models used to describe data, to support interoperability. The idea of a schema registry service (similar to the tool registry service developed by the GA4GH Cloud Work Stream, but for identifying and accessing schema instead of tools) was discussed. As a first step, they aim to establish a group to explore this idea, with a view to supporting the Data Connect and Beacon use cases.

FASP collaboration on Passports, DRS, and Discovery

Recording • **Slides** • **Minutes / Agenda**

The session began with contributors reviewing use cases for DRS + Passports and Discovery + Passports integrations, focusing on how to authenticate API access for various biomedical platforms. The group discussed that in order to solve this, a standardised authorisation is needed. The session was divided into three main categories: current problems that need to be solved, potential solutions, and remaining problems without solutions. The ideas and discussions were tracked using a Miro board.

Bridging data and workflows from research into healthcare — a gap analysis

Recording • **Slides** • **Mentimeter Agenda**

More genomic datasets are being created in various international healthcare settings, but often in different and fragmented environments — making it difficult to share interoperably between research and healthcare. In this workshop, participants looked at a rare disease proof of concept from the Beyond One Million Genomes Project (B1MG) that used the ELIXIR network. Then they discussed challenges, opportunities, and successes of integrating healthcare and research through technology and operations within their own organisations. The session chairs plan to compile the information into a gap analysis.

Report backs and looking forward

Recording • **Slides**

During the course of the meeting, 408 individauls (355 virtual and 53 in-person from 35 countries gathered together to advance standards development work and cross-work stream collaborations and projects. To conclude the meeting, GA4GH Chief Standards Officer (CSO) Susan Fairley reported on the progress made during all 19 working sessions and next steps to further the goals of each group.