

MEETING REPORT

GA4GH October Connect 2021

The GA4GH October Connect meeting occurred from October 12–14, 2021 to provide opportunities for collaboration across the GA4GH Work Streams, Driver Projects, and broader community and to support contributors in advancing work on the GA4GH Roadmap. Below, you will find an overview of all sessions along with links to recordings, slides, and meeting minutes.

Overview of Sessions Held

Opening Remarks	Work Stream Updates	Presentations from External Initiatives	Cloud Work Stream	Pedigree Standard
Federated Analysis Systems Project	Passports Implementation	Cohort Representation	Specification Development Processes	Genetic Discrimination
Future of VCF	Regulatory & Ethics Work Stream	Data Access Committee Review Standards	Variation Standards	Data Connect & Cloud
EDI Workshop	Consent Workshop	Refget & Sequence Collections	Data Model Library Feedback	Security Workshop
AI/ML Bias	Clin/Pheno Driver Project Roundtable	Sequence Annotation	Closing Remarks	

OCTOBER 12, 2021

Opening Remarks: Building Momentum on Implementation

Recording • Slides

GA4GH Chief Standards Officer Susan Fairley shared an overview of GA4GH and next steps to accelerate our work and implementation of GA4GH standards, emphasizing the importance of applying genomics into healthcare, enabling global interoperability, and putting GA4GH standards into practice across a broad range of settings. Several GA4GH initiatives are supporting these efforts, including the Federated Analysis Systems Project (FASP), the Technical Alignment Sub Committee (TASC), and the GA4GH Starter Kit. The expansion of the GA4GH Technical Team will also help to accelerate these efforts, providing additional capacity and

support to the contributor community. The session then continued with updates from all GA4GH Work Streams and Initiatives.

Work Stream Updates

[Recording](#) • [Slides](#)

All the GA4GH Work Streams, the Federated Analysis Systems Project (FASP), and the Equity, Diversity, and Inclusion (EDI) Advisory Group gave updates on their initiatives and discussed the sessions they will host at GA4GH Connect.

Opportunities for Collaboration: Presentations from External Initiatives

[Recording](#) • [Slides](#)

The Human Pangenome Reference Consortium (HPRC) aims to improve representation of sequence diversity in the human population by sequencing 350 diverse human genomes and creating a comprehensive map of genome variation. While currently using GA4GH-maintained standards such as VCF, BAM, and CRAM, HPRC hopes to explore variation representation with the GA4GH Genomic Knowledge Standards (GKS) Work Stream and seek global partnerships to increase sample diversity, share data, and coordinate on standards.

PHA4GE aims to improve mechanisms for protected pathogen data sharing and standardize approaches to compute in this space. Through the work of several working groups, PHA4GE has adapted GA4GH's approach but for the pathogen genomics community, with a focus on bridging research with public health and government.

CINECA aims to deploy GA4GH standards to enable human cohort interoperability through a paradigm of federated analysis. The group aims to tackle a series of challenges, including: federated data discovery, interoperable authentication and authorization, harmonized cohort level metadata, federated analysis interoperability for research and healthcare applications, and trans-national harmonised ELSI framework.

Cloud Work Stream

[Recording](#) • [Meeting Minutes / Slides](#)

The Cloud Work Stream identified major roadmap items and specification updates across all their APIs. The Workflow Execution Service (WES) API team aims to examine feasibility of generic WES input and output formats and support for GA4GH Passports; the Task Execution Service (TES) API team will continue non-breaking improvements to current AI capabilities, along with supporting the GA4GH Passports standard; the Tool Registry Service (TRS) API team aims to explore TRS URIs; and the Data Repository Service (DRS) API team will continue iterating on proposals for batch requests for DRS objects and pagination of results.

Pedigree Standard Implementation

[Recording](#) • [Meeting Minutes / Slides](#)

The GA4GH Pedigree team demonstrated use of the draft Pedigree standard in data collection, management, and analysis tools. In terms of next steps, the team aims to share the minimum core dataset for feedback and approval; demonstrate use within the GA4GH Phenopackets standard;

submit Pedigree Standard for GA4GH approval; and extend converter tooling to support additional formats.

Federated Analysis Systems Project (FASP)

[Recording](#) • [Meeting Minutes / Slides](#)

The FASP team aims to collect use cases for 2022 and beyond, with a goal of collaborating with groups such as NCPI and asking permission to use and expand use cases. In particular, the team hopes to expand geography and participation of their vertical and horizontal demos shown previously, to continue promoting the theme of same compute in different places. To support these efforts, FASP hopes to collaborate more closely with the Technical Alignment Sub Committee (TASC) and the other GA4GH Work Streams to develop and harmonize testbeds and API testing across GA4GH.

OCTOBER 13, 2021

Cohort Representation

[Recording](#) • [Meeting Minutes / Slides](#)

The newly-formed Cohort Representation subgroup presented a landscape review and path forward. The group has decided to split into two “pizza teams” to focus on key topics: 1) minimal information for a computable cohort, which will aim to describe sets of attributes required to describe a cohort consistently; and 2) Phenotypes, which will look at existing phenotypic standards and combining phenotypic information into cohort representation.

Genetic Discrimination

[Recording](#) • [Meeting Minutes](#)

The Genetic Discrimination Information Document—a GA4GH deliverable in development—was reviewed at the meeting. The group discussed strengthening the document by extending the consent clauses section to encompass all tools and resources relevant to anti-discrimination on the basis of genetics. The team also discussed the ongoing Delphi Study, in which analysis of round 1 results are taking place (and round 2 beginning in November). The team aims to begin brainstorming what to do with the results of the data, as well as overall next steps for the Genetic Discrimination Observatory initiative within GA4GH.

Passports Implementation

[Recording](#) • [Meeting Minutes / Slides](#)

The Passports team discussed successes, challenges, and next steps for the standard. The standard has been used successfully in single broker systems with some level of trust; and the next update, Passports v1.2, will aim to tighten up some ambiguity in the specification, with the overall goal of allowing loosely-coupled institutes in federated systems to share data. The next challenge is figuring out what is required to allow a system of multiple brokers across boundaries, which will require discussion and alignment of policy, governance, and tech. The team is [collecting use cases](#) to collect requirements for further technical design on this question.

LSG infrastructure & specification development processes

[Recording](#) • [Meeting Minutes / Slides](#)

Members of the Large Scale Genomics Work Stream explored the technical and social aspects of specification development. The team began with an overview of technical procedures that are currently used to maintain both PDF and HTML specifications, and discussed implementing Architecture Design Records to help document and crystallize important decisions made in specification development. The team aims to continue these process discussions, with the goal of sharing these techniques that other Work Streams and groups can put into action.

Future of VCF

[Recording](#) • [Meeting Minutes](#)

The VCF group from the Large Scale Genomics Work Stream reviewed and determined gaps within their landscape analysis, and narrowed down use cases that were high priority. The team aims to set things into motion to further establish the group and will focus on advancing work on simulated data and the requirements that it must fulfill.

Regulatory & Ethics Work Stream

[Recording](#) • [Meeting Minutes](#)

The Regulatory & Ethics Work Stream (REWS) met to move forward current roadmap deliverables. The Data Access Committee Review Standards (DACReS) Policy was unanimously approved by the Work Stream, and will be brought to the next Steering Committee for GA4GH approval; and Consent Clauses for Large Scale Sequencing was unanimously approved to move forward to public comment. REWS also began brainstorming new areas to explore for their 2020-2021 roadmap, including diversity of datasets, benefit sharing, remote participation in research, approaches to consent for recontact, and communication around public/private collaborations.

Data Access Committee Review Standards (DACReS)

[Recording](#) • [Meeting Minutes](#)

With the support of NIH funding, the DACReS team aims to focus on two areas of work: 1) Capacity building for procedural standardization, including testing and validating the policy through consultation workshops and pilot exercises; and 2) Explore the emergence of automated governance solutions in the data access management space. In terms of next steps, the team aims to use surveys and interviews to explore quality indicators of Data Access Committees and check interest in or concerns for adopting automated approaches. The group is open to other ideas and suggestions for further exploration.

Variation Standards

[Recording](#) • [Meeting Minutes / Slides](#)

The Genomic Knowledge Standards (GKS) Work Stream presented updates on their latest standards: Variation Representation Specification and Variant Annotation. For next steps, the Work Stream aims to focus on demonstrable applications and tools and provide solutions to the application of congruent variation.

FASP (Data Connect + Cloud)

[Recording](#) • [Meeting Minutes / Slides](#)

The Data Connect and Cloud teams ran through a series of demos involving searching for COVID-19 and cancer datasets. The teams aim to continue aggregating use cases—particularly how Data Connect could be used for workflow-based use cases, such as monitoring the status of workflow runs. The team also aims to advance integration of Passports in Data Connect and explore federated use cases involving sharing data models across multiple sites.

OCTOBER 14, 2021

Consent Workshop

[Recording](#) • [Meeting Minutes / Slides](#)

Several Driver Projects, institutions, and organizations were invited to their current process for genomic testing in “ordinary” medical care, with the goal of gathering new insights in order to draft a model template for genomic research consent clauses for data sharing. A number of key challenges and barriers were identified, along with key considerations for implementing a dynamic consent system. Next, the team aims to collect consent clauses from the community and discuss whether they typify what is currently happening in the world of clinical genomics consent.

EDI Workshop

[Recording](#) • [Meeting Minutes / Slides](#)

The EDI Advisory Group announced the launch of the Onboarding initiative, after a successful pilot program that paired four newcomers with four Work Stream Guides. Next, the team shared results from an EDI survey this year, which revealed that more than 50% of respondents were not as involved in GA4GH as they wanted to be. These individuals felt that a lack of time and support impeded their ability to engage with the community. With this in mind, the group held an interactive brainstorming session to determine the EDI Advisory Group’s next project. The group decided that their efforts would be best geared towards developing a Work Stream Best Practices document.

Data Model Library Feedback

[Recording](#) • [Meeting Minutes / Slides](#)

The goal of the data model library is to design a platform that will encourage interoperability across specifications from different GA4GH Work Streams, encourage schema reuse for new standards, and provide stable endpoints for GA4GH schemas. The group discussed cross Work Stream feedback on requirements for a GA4GH-wide schema library, with the aim of reaching broad agreement on key requirements and features. Next, the group aims to reach a common understanding of the scope of the project, including what the data platform will provide for internal contributors as well as downstream consumers of data models, continue discussion of unresolved items, and discuss ideas for a new GA4GH API specification, the Data Model Registry Service.

Refget & Sequence Collections Updates

[Recording](#) • [Meeting Minute / Slides](#)

The refget and sequence collections groups presented brief summaries of five major topic areas with the goal of increasing participation. These include: the adoption of an array based format for defining sequence collections; sequence order within collections; compatibility functions and comparing collections; reverse lookup; and integration into other GA4GH products. The teams aim to carry on these conversations at regular meetings and develop and distribute better use cases for the reverse lookup.

Security Workshop

[Recording](#) • [Meeting Minutes / Slides](#)

The Security Workshop included a presentation on privacy-preserving federated analytics for personalized medicine, a solution that uses multi-party homomorphic encryption. MEdCo uses such a system to provide a distributed software platform for federated cohort exploration and analytics of clinical and genomic data. There are ongoing discussions with the ELIXIR Cloud & AAI Driver Project for the integration of MEdCo with the GA4GH Cloud APIs.

Clin/Pheno Driver Project Roundtable

[Recording](#) • [Meeting Minutes / Slides](#)

The Clinical and Phenotypic Data Capture (Clin/Pheno) Work Stream held a roundtable to build a collaborative community across the GA4GH network, ensure that future Clin/Pheno deliverables can meet community needs, and receive input and feedback on the next Clin/Pheno roadmap. The Work Stream heard from many Driver Projects and groups, including: EUCANCan, EJP-RD, SPHN, SNOMED, HPO & Clinical Terminologies, CanDIG & EpiShare, Australian Genomics, ISO/TC215/SC1, and H3Africa.

AI/ML Bias

[Recording](#) • [Meeting Minutes](#)

This session provided the current landscape of AI/ML bias, beginning with examples of AI/ML bias, biases that can occur throughout the AI lifecycle, types of biases in the data, model evaluations, and the importance of building trust through transparency. Afterwards, the group conducted a discussion on opportunities for AI/ML bias standards within GA4GH. A new cross Work Stream initiative will be created to continue discussions and work in this space, and the team will aim to conduct a survey of AI/ML use across GA4GH.

Sequence Annotation

[Recording](#) • [Meeting Minutes](#)

The Sequence Annotation group reviewed the current draft of the Sequence Annotation model. The team discussed the current draft structure and relationships in the model, as well as implementing a controlled vocabulary for transcript types. The team is open to feedback!

Closing Remarks

[Recording](#) • [Slides](#)

Chief Standards Officer Susan Fairley recapped all twenty sessions from Connect and shared next steps and upcoming announcements from GA4GH.