Work Stream Updates
March 2022

Global Alliance for Genomics & Health

Dear colleagues,

Welcome to the Global Alliance for Genomics and Health (GA4GH) quarterly Work Stream update newsletter. In these pages, you can learn about progress from each Work Stream and cross-Work-Stream initiatives over the last quarter, including actions for the community, key achievements, next steps, and opportunities for collaboration.

Before diving in, we want to share the following announcements:

GA4GH Events

- **GA4GH April Connect 2022**: GA4GH's upcoming working meeting will be a hybrid event, taking place both in Montreal, Canada, and online. Members of the GA4GH community are invited to propose sessions for the agenda by 18 March 2022.
- **International Society for Computational Biology 2022 (ISMB)**: This year, members of the GA4GH community will hold a hackathon to collaboratively shape and develop useful features supporting for the adoption of the Variation Representation Specification (VRS) and a tutorial on the Federated Analysis Systems Project (FASP). Stay tuned for more details soon!
- **GHIF Newborn Sequencing Virtual Workshop**: Join the Genomics in Health Implementation Forum (GHIF) for a virtual workshop on international efforts to establish genomic newborn screening programs on 5 April 2022 from 20:00 to 22:00 UTC.

GA4GH Strategic Refresh Town Hall Meetings

GA4GH is undergoing a strategic refresh to evaluate and discuss progress on the organization's strategic imperatives. We will be inviting the community to attend a series of town hall meetings to provide their perspective and share any feedback on what is working and what needs improvement within GA4GH strategies and structures. Share your thoughts with us by completing a brief survey or by joining us at one of our upcoming town hall meetings.

Google Summer of Code Call for Submissions

For the past four years, the GA4GH community has participated in the Google Summer of Code program (GSoC) — a global program that brings student developers into open source software development. Interested in submitting a potential project and being a mentor? Please submit any GA4GH project ideas to jeremy.adams@ga4gh.org.

GA4GH Special Issue in *Cell Genomics*

Looking for some good bedtime reading? Last November, the journal *Cell Genomics* ran a special issue on GA4GH, featuring a compendium of twelve papers and commentaries written by the GA4GH community. Grab a cup of herbal tea and settle in to learn more about GA4GH, our goals, our outputs, and how to use them.

New to GA4GH or Interested in Getting More Involved?

Join the Onboarding Program presented by the GA4GH Equity, Diversity and Inclusion (EDI) Advisory Group! The program provides new members with the knowledge needed to start participating fully in GA4GH Work Streams. You'll be paired with an active member of a Work Stream (also known as a "Work Stream guide") to help you orient to the GA4GH community and Work-Stream-specific projects . If you are interested in contributing to any of the GA4GH Work Streams, start by reviewing this informational packet and reaching out to secretariat@ga4gh.org.

Sincerely,
The GA4GH Secretariat

Table of Contents

Work Stream Updates

Clinical & Phenotypic Data Capture (Clin/Pheno)

Clin/Pheno Meeting Minutes | Clin/Pheno Roadmap

The Clinical & Phenotypic Data Capture Work Stream aims to support clinical adoption of genomics through information models and standards for describing and exchanging clinical phenotypes. The team is developing the Phenopackets and Pedigree standards.

Take Action

- **Interested in learning more about the ISO draft International Standard ballot process?** Contact Lindsay Smith.
- **Requesting feedback**: the Pedigree team welcomes the community to test the Pedigree Standard Model. Reach out to Lindsay Smith to learn more and provide feedback.

Key Achievements

- Phenopackets v2.0 was approved and is now available for adoption. Hear from leaders about the changes and enhancements of version 2.0 and the benefits for clinicians and researchers.
- A manuscript describing the Phenopackets ecosystem was submitted to *Nature Biotechnology*.
- Phenopackets has passed the ISO draft International Standard ballot and will move forward to publication in April 2022.
- The HL7 FHIR Phenopackets Implementation Guide was presented to the National Library of Medicine (NLM) on 18 February 2022.
- The HAPI Terminology server has been developed that allows genomic resources based on Phenopackets to be distributed to the electronic health record (EHR) community.

Cross-Work Stream Coordination

- Computable Cohort Representation, a new joint subgroup with the GA4GH Discovery Work Stream, aims to focus on minimal information about the cohort high level descriptors and build a computable phenotype standard.

Next Steps

- Complete Pedigree standard documentation and other materials needed for approval, validation, and conversion.
- The Computable Cohort Representation aims to take asthma cohort definitions from various libraries and harmonize them into a computable standardized phenotype.

Cloud

Cloud Meeting Minutes | Cloud Roadmap

The Cloud Work Stream aims to help the genomics and health communities take full advantage of modern cloud environments by bringing algorithms to data. The team develops the Workflow Execution Service (WES) API, Tool Registry Service (TRS) API, Data Repository Service (DRS) API, and Task Execution Service (TES) API.

> **Take Action**
>
> - **Interested in joining the Cloud Work Stream?** Now is the time to expand the team and update its APIs**.** Reach out to Reggan Thomas to get involved.

Key Achievements

- All of the Cloud services have been approved and are readily available.

Cross-Work-Stream Coordination Goals

- In collaboration with the Federated Analysis System Project (FASP), the Cloud Work Stream team hopes to continue to demonstrate how its services can work together with APIs from other Work Streams, such as the Data Use & Researcher Identities (DURI) and Discovery Work Streams.
- FASP also wishes to continue to provide feedback on real-world use of Cloud APIs.

Next Steps

- Refine and release versions of the APIs based on feedback, continuing to publish new API releases roughly twice yearly.
- Develop resources such as Wikipedia posts and papers describing APIs and best practices.

Data Use & Researcher Identities (DURI)

DURI Meeting Minutes | DURI Roadmap

The DURI Work Stream aims to help the community facilitate and streamline data access processes by standardizing researcher identity and data use. The team develops the Passport Standard and the Data Use Ontology (DUO).

Take Action

- **Interested in joining DURI?** The team seeks contributors to help reach its 2022 goals. Contact the Secretariat to get involved.
- Calling for community input on Passport v1.2: share your comments on GitHub.

Key Achievements

- Machine-Readable Consent Guidance was approved in June 2021.
- The Data Access Committee Guiding Principles and Procedural Standards Policy was approved in August 2021.
- Papers on the Passport standard, Data Use Ontology (DUO), and Data Use Ontology System (DUOS) were published in the GA4GH special issue of *Cell Genomics*.
- The DUO standard was updated this month.

Cross-Work Stream Coordination Goals

- DURI and the Regulatory & Ethics Work Stream (REWS) are discussing ways to improve and standardize data use agreements in their future roadmaps.
- The team plans to continue developing Machine-Readable Consent Guidance, with potential tooling to support DUO terms in consent forms and protocols to assist the community with consent form translation.
- A DRS + Passports Hackathon led by the Federated Analysis Systems Project (FASP) aims to address multiple subproblems for the token handoff and to test interoperability.

Next Steps

- DURI 2022 Roadmap plans include working towards the release of Passport v1.2, as well as merging the v1.2 specification with the Authentication & Authorization Infrastructure (AAI) specification into a v2.0 release.

Discovery

Discovery Meeting Minutes | Discovery Roadmap

The Discovery Work Stream aims to facilitate the discovery and utilization of data sources and services. The team develops the Beacon API, Data Connect (formerly Search) API, and Service Info/Service Registry APIs.

Take Action

- **Interested in joining the Discovery Work Stream?** The Data Connect team is looking for groups to test and implement the Data Connect API. Reach out to the Secretariat to get involved.
- **Sharing data models and describing data:** The Discovery team is interested in working with others on ways of describing shared data, with uses relating to both Data Connect and Beacon. If you are interested in exploring this, please get in touch with the Secretariat.

Key Achievements

- Data Connect 1.0 standard has been released, and work has started on the next minor release (v1.1). Discover the value and impact of this come-as-you-are data sharing approach in a news story, and dig deeper in our podcast episode.
- The Federated Analysis Systems Project (FASP) held a hackathon at the end of January, including a Data Connect session on getting started with the standard and a "how to add custom schemas" challenge.
- For its data delivery protocol, Beacon v2.0 has adopted Phenopackets and the Variation Representation Specification (VRS).

Cross-Work-Stream Coordination Goals

- The team aims to expand into integrated applications—for example, query Beacon v2.0 with Phenopackets and VRS in matchmaking applications.

Next Steps

- Currently, the team is working to increase Data Connect adoption with groups that have been previously contacted.
- The Work Stream members will continue to push for wider distributions of registry standards for network services.

- Submission of the Beacon v2.0 specification to the GA4GH Steering Committee is planned for April 2022.

Genomic Knowledge Standards (GKS)

GKS Meeting Minutes | GKS Roadmap

The GKS Work Stream aims to enable downstream analysis of genomic data by developing common APIs for exchanging reference genomic information. The team develops the Variation Representation Specification (VRS) and Variant Annotation (VA) Framework.

Take Action

- **Interested in joining GKS?** Virtual meetings now occurg as follows:
  - **Variation Representation**
    - 2nd and 4th Monday: 11 am - 12 pm ET
    - 3rd Monday: 5 - 6 pm ET / Tuesday: 9 - 10 am AEDT
  - **Variation Annotation**
    - Every Wednesday: 11 am - 12 pm ET
  - **VRS-VCF**
    - 2nd Wednesday: 3 - 4 pm ET
  - **Sequence Annotation**
    - 1st Thursday: 10 - 11 am ET

Reach out to Reggan Thomas to attend.

Key Achievements

- A paper on VRS 1.2 was published in the GA4GH special issue of *Cell Genomics* in November 2021.
- Welcome Andreas Prilić, who is stepping in for Reece Hart as VRS co-lead.
- The Variant Annotation team reviewed the foundational model after receiving feedback from ClinGen/VICC, and started a reference document containing MolecularConsequence, VariantPathogenicity, and TherapeuticResponse statements.

Cross-Work-Stream Coordination

- VRS and VCF leads have been working towards understanding common goals, so they can improve the focus and priorities across their overlapping efforts.

Next Steps

- Attending ISMB 2022? Join us for the VRS Hackathon 11 to 12 July 2022. Register for the

event and propose projects to support the adoption of VRS in bioinformatics software, genomics databases, and analysis pipelines.

- VRS 1.3 is under development, focusing on relative copy number and genotypes.
- Members will promote VRSATILE as a GA4GH-recognized framework for applying VRS.
- The team continues work on sequence annotation, and the first models are expected to be passed back to the VRS and VA teams in the coming months.
- Ongoing proposal development for restructuring the GKS Work Stream aims to reduce siloisation of projects and foster shared ownership of products and standards across the Work Stream.
- The Variant Annotation team plans to formalise and document the foundational, domain-agnostic conceptual model for representing knowledge statements and their provenance. They also plan to create a minimal modelling framework that supports community-driven development of models for specific types of knowledge statements.

Large Scale Genomics (LSG)

LSG Meeting Minutes | LSG Roadmap

The LSG Work Stream aims to develop efficient formats to store, access, and analyze sequencing reads, genetic variation, and gene expression information. The team develops htsget, refget, CRAM, VCF, and Crypt4GH.

> **Take Action**
>
> - **Interested in getting involved?** The team is recruiting a lead for the Large Scale Genomics Work Stream as a whole and a co-lead for the Future of VCF project. Please reach out to Reggan Thomas to learn more.

Key Achievements

- The formalized BED standard was approved by the GA4GH Standards Steering Committee in October 2021.
- A paper on CRAM 3.1 has been published in *Bioinformatics*, describing updates made to the format. A second implementation is being developed by Yasasvini (Yash) Puligundla at the Broad Institute.
- The Crypt4gh team has begun initial discussions with cloud vendors regarding native support and key management for the Crypt4GH standard.
- After collecting user stories, the future of VCF team has identified key fields for the format

and defined initial aims.

Cross-Work-Stream Coordination

- The RNAget team is working with the Data Connect team from the Discovery Work Stream to support search queries in the RNAget API.

Next Steps

- The htsget team is focused on expanding adoption, supporting non-standard file formats, and making changes to HTSlib to delegate index support to htsget.
- The refget team is exploring support of refget in other LSG file formats.
- As a whole, the Work Stream team is focused on increasing participation and recruiting new contributors for their projects.

Data Security

Data Security Meeting Minutes | Data Security 2020 Roadmap

The Data Security Work Stream aims to help the community keep data protected and to ensure that GA4GH standards are developed within a sound risk-management framework. The team develops the Data Security and Infrastructure Policy (DSIP) and the Authentication & Authorization (AAI) API. The team is also looking towards developing Risk Assessment Methodology for Software Stacks and a Cloud Security and Privacy Guide.

Take Action

- **Interested in getting involved?** The Work Stream team seeks volunteers to lead and contribute to the Breach Response Policy and Malfeasance Rules. Reach out to the Secretariat.
- **Calling all Driver Projects or Initiatives:** Contact the Secretariat if interested in contributing to or providing use cases for Malfeasance Rules.

Key Achievements

- NIH's RAS project is in production implementing the GA4GH Passport standard, which uses the Authorization and Authenticaiton Infrastructure Specification.
- Institutions such as the University of Chicago, the Broad Institute, Seven Bridges Genomics, and Harvard University are using GA4GH Passports to test interoperability between their projects, leading to changes in the Authentication & Authorization Infrastructure (AAI) API and Passport Standard to make things work more smoothly.

- After the October 2021 Connect Meeting, a call for volunteers went out for interest to work on Malfeasance Rules. The NHLBI TopMED Driver Project expressed interest in the project.

Cross-Work-Stream Coordination

- The team formed a technical working group between the Data Use & Researcher Identities (DURI) Work Stream and the Data Security Work Stream to collaborate on GA4GH Passport v1.2.
- Members continued collaboration with the Cloud, Discovery, and DURI Work Streams through the Federated Analysis Systems Project (FASP).

Next Steps

- The plan is to continue to advance all standards, with the aim to release Breach Response Policy at the 2022 Plenary.
- There is a continued focus on increasing participation and involvement in Data Security Work Stream deliverables.

Regulatory & Ethics (REWS)

Regulatory & Ethics Meeting Minutes | Regulatory & Ethics Roadmap

The Regulatory & Ethics Work Stream aims to create and harmonize forward-looking consent and privacy policies and anticipatory data governance models — building upon the human right to benefit from science. View the list of approved and ongoing REWS deliverables here.

> **Take Action**
>
> - **Interested in getting involved?** This is a great time to join as REWS develops a new roadmap for future projects (listed below). Reach out to Lindsay Smith to get involved.

Key Achievements

- Genetic Discrimination: Implications for Data Sharing Projects (GeDI) — a collaboration with the Genetic Discrimination Observatory (GDO) to develop an informational resource on genetic discrimination — was approved by the GA4GH Standards Steering Committee in January 2022!
- The following REWS policies were approved last year by the GA4GH Steering Committee:
    - Data Access Committee Guiding Principles and Procedural Standards Policy

produced by the Data Access Review Standards (DACReS) working group.

    - GA4GH Policy on Clinically Actionable Genomic Research Results. Learn about the first international policy on how to ethically return results that could lead to medical benefits for research participants.
    - Engagement Framework. Read a news story on how the framework helps build inclusive engagement practices in genomics.

- The Consent Task Force published Pediatric Consent to Genetic Research: Clauses, which provide generic consent clauses particular to the pediatric research context for mature minors and for parents consenting on behalf of children. An accompanying generic Assent Form and Information Sheet (for ages approximately 7 to 11) is also provided.
- The Genetic Discrimination team aims to start a new project on addressing the challenges raised by stigmatization and discrimination in genomics at the population level. Reach out to Lindsay Smith to get involved.
- The *Your DNA, Your Say* team published a paper, "Return of genomic results doe snot motivate intent to participate in research for all: Perspectives across 22 countries," in *Genetics in Medicine*.

Cross-Work-Stream Coordination

- REWS is collaborating with the Data Use & Researcher Identities (DURI) Work Stream on the Data Access Committee Guiding Principles and Procedural Standards Policy, which aims to develop procedural standards for Data Access Committees (DACs).

Next Steps

- Large Scale Initiatives Consent Clauses is undergoing the GA4GH product review process and will go to the April 2022 Steering Committee meeting for approval.
- The DACReS working group aims to interview and develop a survey for DAC members to inform the DACReS policy and assess potential for automated decision making to support access decisions.
- REWS members are developing their 2022/2023 roadmap and identifying new potential projects to explore, including: Artificial Intelligence/Machine Learning Bias, Diversity in Datasets, Newborn Sequencing, Global Genomics, "Global Public, Benefit Sharing," Blockchain and Ethical Provenance, and Consent and Clinical Data Sharing.

Project Updates

Federated Analysis Systems Project (FASP)

FASP Meeting Minutes

FASP was established to show that GA4GH APIs, when used in concert, can facilitate real-world, scientific use cases by conducting genomic analysis in the cloud. FASP aims to simulate how a researcher would search, access, and analyze genomic data within the GA4GH ecosystem via end-to-end test scenarios involving multiple Driver Projects. Learn more about FASP here.

> **Take Action**
>
> - **Do you have a use case or project to share with the FASP team?** Fill out the FASP Use Case / Project Survey.

Key Achievements

- The team welcomed Ian Fore as a new co-lead for FASP in December 2021.
- The team developed testbed coverage for the Data Repository Service (DRS) API and GA4GH Passports integration.
- FASP members held a hackathon at the end of January 2022, including a Data Connect session on getting started with the standard and a “how to add custom schemas” challenge.

Next Steps

- User feedback from the January hackathons have helped inform API changes and specification improvements. Projects include:
    - Validating DRS in the context of other APIs and continuing work on DRS batch operations and pagination.
    - Refining DRS and GA4GH Passports test scenarios.
    - Supporting GA4GH Passports/AAI subgroup work towards v1.2 release.
- The FASP team aims to support the GA4GH technical team on Starter Kit activities:
    - Distribute recording of Data Connect deployment.
    - Prepare tutorial at ISMB on the GA4GH Starter Kit and FASP Scripts.
- The plan is to continue to work with Driver Projects to understand use cases for FASP, ultimately using use cases to drive real world implementations of GA4GH APIs.

Equity, Diversity, and Inclusion (EDI) Advisory Group

EDI Meeting Minutes

The EDI Advisory Group aims to examine and address equity, diversity, and inclusion within the GA4GH standards development community.

Take Action

- **New to GA4GH?** Reach out to Neerjah Skantharajah to be placed in our onboarding program.
- **Become a GA4GH Work Stream guide!** Help orient newcomers to the inner workings of your GA4GH Work Stream/subgroup. Please reach out to Neerjah Skantharajah to volunteer to be a Work Stream guide!
- **Join our EDI work on bringing diversity and inclusion to the GA4GH Community**. Reach out to Neerjah Skantharajah to be added to the mailing list.

Key Achievements

- The onboarding initiative has now successfully onboarded 18 newcomers, with overall positive feedback from newcomers and high retention rate.
- A new Work Stream Information Packet was created to highlight opportunities for people to learn about the breadth of work across GA4GH and get involved.
- The group held a successful Miro Board brainstorming session at the end of last year to begin outlining best practices for building inclusive teams within GA4GH, leading to the creation of a Work Stream Best Practices Document.

Next Steps

- The group will add a check-in process with newcomers to continue assessing and improving the GA4GH Onboarding program.
- The team will continue work on EDI's second deliverable, a Work Stream Best Practices document. A Google form will be circulated to the Work Streams and Sub Groups to identify current best practices in use by Work Streams that have fostered inclusivity, and areas where best practices still require development.
- Plans are in place to develop an EDI webpage on ga4gh.org to increase visibility of EDI work across GA4GH.

Technical Alignment Sub-Committee (TASC)

TASC Meeting Minutes

TASC aims to provide mechanisms and recommendations to create internal consistency and technical alignment across GA4GH Work Streams and product deliverables.

Take Action

- **Do you have a technical alignment issue for GA4GH?** Please add it to the TASC issues list on GitHub.

Key Achievements

- The sub-committee discussed key topics such as a GA4GH data model library and data model reuse, in collaboration with Schemablocks.
- The team discussed the Variation Representation Specification (VRS) namespace topic.

Next Steps

- The TASC team will take forward work on the GA4GH data model library with GA4GH staff, in line with plans discussed at October 2021 Connect.
- The sub-committee will continue the namespace discussion, with a deep dive on type prefix between TASC and the VRS teams.