Global Alliance for Genomics & Health

Meeting Summary

GA4GH Leadership Meeting

May 2018, Toronto, Canada

Introduction

The GA4GH 2018 Leadership Meeting brought together more than 120 active GA4GH Contributors to discuss the current state of Roadmap Deliverables. The meeting included full-group sessions, working meetings of each of the eight Work Streams, and three cross-cutting breakout sessions. *Access the full agenda, with links to individual meeting agendas*.

Work Streams

Each Work Stream gave a ten-minute presentation on Day 2 of the meeting, highlighting current work and goals for the meeting. They had the opportunity to meet as a small group three times and gave a report-back during to the full group during the closing session on Day 3. Key highlights, as well as links to meeting minutes (where available) are captured below.

- *Access the Work Stream Update Slides and detailed minutes from Day 2*
- *Access the Work Stream Report Back Slides and detailed minutes from Day 3*

Clinical & Phenotypic Data Capture

Several members of the Clinical & Phenotypic Data Capture (Clin/Pheno) Work Stream met on May 29 and 30 as part of a PhenoPackets hackathon, organized by the Monarch Initiative. Much of the discussion at this meeting was focused on aligning PhenoPackets with the HL7 FHIR standard. At the GA4GH Meeting, the Clin/Pheno group continued hacking on this topic, and also reviewed terminologies, ontologies, and data types of interest to Driver Projects. Key actions for this group are to define the use cases for PhenoPackets, extend the GKS Variant Annotation Modelling requirements to include best practices for the use of ontologies and terminologies, and to develop an engagement strategy with the Partner Engagement Group. *Read the Clin/Pheno Meeting minutes*.

Cloud

During the meeting, the Cloud Work Stream met with its “pressing need” Driver Projects: Human Cell Atlas, Genomics England, and TOPMed/NIH Genomic Data Commons. They discussed a 1x1 demo of the Cloud testbed, and a path to a more complete demonstration, which was identified as a key action from the meeting. *Read the Cloud Meeting minutes*.

Data Security

The Data Security Work Stream focused its discussions on progress within its two subgroups: Breach Response and AAI. The group proposed adding two more deliverables from the Breach Response subgroup: in addition to developing a Data Breach Reporting Protocol, they will also work on Vulnerability Management and a Data Breach Management Best Practice Guide for use within GA4GH Driver Projects. They also aim to review a draft AAI profile by 6th Plenary. *Read the Breach Response Meeting minutes*.

Data Use & Researcher Identities

The Data Use & Researcher Identities (DURI) Work Stream used the meeting to clarify definitions of data use categories based on driver project input, to review the next draft of the Data Use Ontology, and to define the most critical researcher identity claim types. Key actions from the meeting include working with Driver Projects to develop a DUO discovery demonstration and a Researcher Identity interoperability demonstration. *Read the DURI Meeting minutes.*

Discovery

The Discovery Work Stream discussed collaborating across Work Streams to promote sharing between large-scale initiatives and the development of a global network for genomics and clinical data discovery. They also discussed re-identification risk inherent in many GA4GH deliverables proposed a GA4GH-wide position statement on re-identification to reduce perceived novelty. In addition to pushing forward the Search and Beacon API's, the team's next steps include following up with ClinVar on alternate strategies for its "VariantEx" deliverable and piloting an international case exchange between Australian Genomics and Genomics England.

Genomic Knowledge Standards

The Genomic Knowledge Standards (GKS) Work Stream discussed progress on its roadmap deliverables, including a focus on coordinating the two GKS subgroups, reviewing types of variants that are in scope for the GKS standards, and a presentation of a use case from ClinGen. The group also discussed overlap with other Work Streams that are addressing different aspect of variant annotation and representation. *Read the GKS Meeting minutes.*

Large Scale Genomics

The Large Scale Genomics (LSG) Work Stream discussed progress on streaming sequence data/htsget, read file formats, variation file formats, RNA sequencing data, and reference retrieval. The team's next steps include convening new subgroups to discuss variation scaling and nanopore sequencing and to push the Reference Retrieval API and the RNASeq expression matrix toward product approval. *Read the LSG Meeting minutes*.

Regulatory & Ethics

The Regulatory and Ethics Work Stream (REWS) discussed progress on their primary roadmap deliverables: Return of Results, Participant Values Survey, and the International Code of Conduct for Health Research. Key actions for the REWS include (1) updating the GA4GH website with a statement on responsible genomic data sharing, and (2) developing a regulatory & ethics review process for all GA4GH deliverables.

Summary: Driver Project Technical Panel

A panel discussion of six Driver Project technical representatives addressed how projects plan to share data and the types of data they plan to share; what GA4GH tools they are using or plan to use; and issues of privacy and security around sharing genomic and health related data. *Read the detailed panel discussion minutes*.

Cross-Cutting Breakouts

Variants and Structuring Associations

Representatives from the Discovery, Clin/Pheno, Large Scale Genomics, Cloud, and GKS Work Streams discussed questions of variant annotation and representation across a variety of Roadmap activities, including aligning the VCF reads file format maintained by the LSG Work Stream and the Variant Modeling Consortium standard being developed by the GKS Work Stream. During the larger group meeting, the topic was raised again and it was noted that outreach to the Hail group may be warranted. *Read the Variant Meeting minutes*.

Access

Representatives from the REWS, Data Security, Cloud and DURI Work Streams focused primarily on establishing a universally accepted researcher identity mechanism in order to streamline responsible access to data. They discussed the concept of "registered access", current mechanisms for accessing All of Us and Australian Genomics data, and the need for a flexible data access “passport” or “library card” that accommodate varying levels of stringency. They also discussed whether implementation of this type of mechanism would mean data stewards would drop existing data access agreement requirements when sharing data with other organizations. They recommended the library card model be piloted within ELIXIR and an NIH project, like All of Us.

Data Retrieval

Representatives from the Cloud, Discovery, Genomic Knowledge Standards, DURI, and Large Scale Genomics Work Streams discussed Roadmap deliverables with multiple Work Stream stakeholders

including the Data Object Service-supported object formats, Discovery search, and htsget. *Read the Data Retrieval meeting minutes*.

Priority Deliverables

During the meeting, Several Driver Projects named deliverables of particular interest and Work Streams stated deliverables to be prioritized for October 2018 release:

Driver Project Priorities

- **All of Us**: Researcher IDs. AoU will allow researcher access by end of 2018/beginning of 2019.
- **Australian Genomics**: ClinVar API ("VariantEx")
- **CanDIG**: all of them
- **ClinGen**: First step toward VariantEx: alternative that uses with ClinGen as an intermediary
- **ELIXIR-Beacons**: end-to-end implementation of Cloud standards
- **ENA/EGA/EVA**: DUO, Researcher IDs, and nanopore standardization (among others)
- **Genomics England**: Cloud containers and workflows, DUOS/data usage; variant representation; Search API.
- **Human Cell Atlas**: Open data coordination platform will launch in 6-12 months, requiring expression matrices for single cell sequencing. Also, WES and Researcher IDs.
- **ICGC-ARGO**: WES, DOS, registered access, ethics and security work.
- **Matchmaker Exchange**: one-sided matching
- **Monarch Initiative**: PhenoPackets and FHIR coordination; variant representation and annotation.
- **NIH GDC/DCF**: Outputs from Cloud, DURI, and LSG Work Streams
- **TOPMed**: Variant representation, ability to go from VCF to other formats.
- **VICC**: Variant representation specifications

Work Stream Priorities

- **Cloud**: Workflow Execution Service (WES) API; cross-cutting demo with htsget
- **Genomic Knowledge Standards**: VMC V1.0
- **Discovery**: Beacon API and Search API
- **Large Scale Genomics**: Reference Sequence Retrieval API
- **DURI**: Data Use Ontology
- **Clinical & Phenotypic Data Capture**: Investigating PhenoPackets v1.0 mapping/alignment with FHIR
- **Data Security**: Breach Response Strategy
- **REWS**: Interpretation of Participant Values Survey Results; focus on responsible data sharing in GA4GH messaging