GA4GH 9th Plenary Meeting Report

The Global Alliance for Genomics and Health (GA4GH) held its annual plenary from September 28 to 29, 2021. Held virtually this year due to the ongoing COVID-19 Pandemic, the 9th Plenary meeting brought together 511 participants across 57 countries, featuring 25 sessions on topics such as real world implementations of GA4GH standards, engaging with underrepresented communities, responsible data, and COVID data sharing. The meeting offered English closed captioning and live interpretation in Spanish, French, German, and Japanese.

Overview

Track 1: The GA4GH Story So Far

- Opening Remarks
- Keynote: "A new appraisal of genomics at the World Health Organization" *(Harold Varmus, Weill Cornell Medicine)*
- GA4GH Overview & Updates
- EDI Advisory Group
- Standards Spotlight
- Federated Analysis Systems Project (FASP)
- Introduction to the GA4GH Starter Kit

Track 2: Collaborate. Innovate. Accelerate.

- Genomics in Health Implementation Forum (GHIF)
- Real World Implementers
  - Australian Genomics
  - PRECISE
  - GEM Japan
  - IRUD
- Panel: Engaging with Underrepresented Communities
  - CIEDAR
  - Human Pangenome Project
  - All of Us Research Program
  - NCIG

Track 3: Responsible Data

- Panel: Responsible Data
  - Homomorphic Encryption in MedCO
  - CanDIG
  - Ethical, legal, and governance considerations in the processing, use, and sharing of real-world data

Track 4: Expanding the Scope & Impact of GA4GH

- Keynote: The Role of PH4AGE in response to barriers in global (pathogen) data sharing *(Alan Christoffels, PHA4GE)*
- Panel: COVID-19 Data Sharing
  - COVID-19 Host Genetics Initiative

    - Return of Results
- Real World Implementers
    - B1MG
    - NIH Cloud Interoperability
    - H3Africa
    - Enabling clinical data discovery with Beacon
    - National COVID Cohort Collaborative
    - CanCOGeN
- Panel: Building Bridges to Industry and Clinical Care
    - Clinical Research Data Sharing Alliance
    - Geisinger Health System
    - Invitae

Track 1: The GA4GH Story So Far

Opening Remarks

Ewan Birney (EMBL-EBI)

Presentation Video | **Slides**

Birney opened GA4GH 9th Plenary with an overview of the meeting, sharing it's our most geographically diverse meeting to date with 57 countries represented. The meeting would follow four key themes: The GA4GH Story So Far; Collaborate. Innovate. Accelerate.; Responsible Data; and Expanding the Scope & Impact of GA4GH.

Keynote: A new appraisal of genomics at the World Health Organization

Harold Varmus (Weill Cornell Medicine)

Presentation Video | **Slides**

In his keynote talk, Varmus discussed approaches to increasing equity in the use of genomics through two examples: 1) advising the WHO on expanding applications of genomics in global health; and 2) Building a research consortium in New York City to study the role of genetic ancestry in cancer risk, progression, outcomes, and more.

Overview & Updates from GA4GH

Peter Goodhand (GA4GH) & Susan Fairley (GA4GH)

Presentation Video | **Slides**

Goodhand shared key updates to GA4GH, including the formation of a non-profit to ensure the stability of GA4GH, expansion of staff to accelerate GA4GH's core work and mission, and launch of a new COVID-19 Interoperability Initiative, and directive grant to help with pandemic preparedness in the future. Next, Chief Standards Officer Susan Fairley provided an update on the technical side of GA4GH and the variety of activities that are occurring to support standards development.

EDI Advisory Group: Building Intentional Communities

Laura Paglione (Spherical Cow Group)

Presentation Video | **Slides**

Paglione discussed the work of the GA4GH EDI Advisory Group to date, and how the team has taken an intentional community approach. Work has culminated into three overarching categories—Onboarding, Participation Levels, and Equity by Design—and encouraged the community to get involved and contribute to EDI efforts to effect change.

Standards Spotlight

GA4GH Standards Integration: Phenopackets + VRS

Monica Munoz-Torres (University of Colorado), Alex Wagner (Nationwide Children's Hospital), Peter Robinson (Jackson Laboratory)

Presentation Video | **Slides**

Munoz-Torres provided an overview of the GA4GH Phenopackets standard, which is now on its v2.0 release. The new version brings significant changes to enable better representation of cancer, common disease, and COVID-19. Munoz-Torres then described efforts to integrate the Variation Represerntation Specification (VRS) and VRSATILE—standards developed by the GA4GH Genomic Knowledge Standards Work Stream—into Phenopackets v2.0,

Beacon API

Jordi Rambla (CRG)

Presentation Video | **Slides**

Beacon is a discovery tool aiming to be a "lingua franca." Rambla shared an overview of Beacon's development, and how the current version—Beacon v2—can provide more information on sequences and better support clinical genomics research and its application to healthcare.

Passports

Sarion Bowers (Wellcome Sanger Institute)

Presentation Video

Bowers presented three ways to currently access data—open access, registered access, and controlled access. For sensitive data that is under controlled access, the Passports standard uses digital identities to help Data Access Committees (DACs) review data access requests. In the future, the team aims to combine Passports with other GA4GH standards and to allow for further automation to alleviate burden on DACs.

Passports & DRS Integration

Max Barkley (DNAstack)

Presentation Video

Next, Barkley discussed efforts to integrate Passports standard with the Data Repository Service (DRS) API, which is used for accessing files in the cloud or on remote servers. The team discussed challenges, progress, and next steps to allow for the integration of the two standards.

DUOS

Pamela Bretscher (Broad Institute)

Presentation Video

Bretscher shared an example of DUO in action, in the Broad Institute's DUOS system. A matching algorithm can instantly determine whether a data access request is within the parameters of its original consents, allowing Data Access Committees to more easily make a final decision on whether to grant access. In the future, DUOS hopes to make use of the Passports Standard and to continually update processes based on the GA4GH DACReS group.

Data Connect

Miro Cupak (DNAstack)

Presentation Video | **Slides**

A newly-approved standard, the Data Connect API provides a flexible mechanism for data discovery and search. Cupak shared the key features of the Data API, which allows data custodians to easily describe their data as it sits without being overly prescriptive or going through an intensive transformation process, lowering the barrier to sharing and searching data.

CRAM Update

James Bonfield (Wellcome Sanger Institute)

Presentation Video | **Slides**

Bonfield presented an update on CRAM, a file format for genomic data compression. Updates were made to v3.1, which reports significant boosts in terms of the speed/size tradeoff of the data file. Challenges and next steps were presented to further improve the standard.

Future of VCF

Albert Smith (University of Michigan)

Presentation Video | **Slides**

Smith provided an overview of the history of VCF, a tab-delimited format for storing variant calls and individual genotypes. Addressing VCF scalability is next on the radar, with benchmarking being planned to evaluate possible solutions and proposals.

GA4GH Federated Analysis Systems Project (FASP)

Max Barkley (DNAstack), Ian Fore (NIH), Brian O'Connor (Broad Institute)
Presentation Video | **Slides**
Now two years in the works, FASP has shown how GA4GH APIs can be used in concert to enable real-world scientific use cases. The team has brought people together through regular meetings and hackathons to zero in on specific use cases. Barkley provided an update on two key demos—the Vertical and Horizontal demos—and progress made to date.

Introduction to the GA4GH Starter Kit

Jeremy Adams (GA4GH)
Presentation Video | **Slides**
Adams presented a new GA4GH initiative, the GA4GH Starter Kit, which serves as a simple, comprehensive suite of open source applications to help new organizations implement or adopt GA4GH standards. Two GA4GH standards—DRS and WES—have starter kits, with more standards coming soon.

Track 2: Collaborate. Innovate. Accelerate.

Implementing GA4GH Standards: Genomics in Health Implementation Forum (GHIF)

Kathryn North (Australian Genomics)
Presentation Video | **Slides**
North opened track 2 by sharing more on the Genomics in Health Implementation Forum, a community aimed at putting GA4GH standards into practice and enabling collective learning in an effort to solve the ability to share data across borders. North presented a brief history of GHIF, benefits of participating in GHIF, and how to join.

Real World Implementers

Accelerating rare disease diagnosis

Zornitza Stark (Australian Genomics)
Presentation Video | **Slides**
Stark presented work done through the Acute Care Genomics project, which aims to pilot a national approach for rapid genomic testing for infants admitted into the intensive care unit. The project is using interoperable formats to share the data for research use, establishing electronic workflows, and streamlining clinical data capture. The project aims to accelerate the implementation of genomics within the Australian healthcare system and foster a virtuous cycle between clinical and research genomics.

Precision Health Research Singapore

Mar Gonzalez-Porta (Genome Institute of Singapore)

Presentation Video | **Slides**

Gonzalez-Porta discussed Singapore's Precision Medicine Strategy. Phase 1 is complete, which aimed to sequence 10,000 health Singaporeans and enable clinical adoption through return of results. The team is transitioning now to Phase 2, which aims to build a database of 100K genomes from the local population and link it to corresponding phenotypic data. GA4GH standards such as Beacon and CRAM have aided in their efforts thus far.

Activities for Developing and Implementing Framework for Responsible Sharing of Clinical Data & Genomic Data

Soichi Ogishima (Tohoku University)

Presentation Video | **Slides**

Ogishima provided an overview of the GA4GH standards implemented in Japan's biobank network, including the Data Use Ontology (DUO) and the Machine Readable Consent Guidance (MRCG) for research purpose matching and Phenopackets v2 for clinical data capture and exchange. The teams aims to expand the biobank network and cohort network to enable a responsible data sharing ecosystem between clinical genomics and research.

Initiative on Rare and Undiagnosed Diseases (IRUD) and real-world use of Phenopackets

Kenjiro Kosaki (Keio University)

Presentation Video | **Slides**

Kosaki discussed use of Phenopackets in IRUD, where the accumulation of phenotype-genotype data within Phenopackets has proven valuable. Through the IRUD-Exchange data sharing platform, new disorders—such as a new overgrowth syndrome—were able to be identified, along with re-analysis of unresolved cases.

Panel: Engaging with Underrepresented Communities

CIEDAR: CoVaRR-Net Indigenous Engagement, Development, and Research

Michelle Johnson-Jennings (University of Washington), Katie Collins (University of Saskatchewan)

Presentation Video | **Slides**

Johnson-Jennings and Collins discussed the work of CIEDAR, one of the eight pillars of Canada's Coronavirus Variants Rapid Response Network (CoVaRR Net). The inclusion of CIEDAR was key to integrating indigenous peoples in the network itself—to build true partnerships after a history of distrust and data mismanagement and to enable effective community health policies during the COVID-19 pandemic.

The Human Pangenome Project

Karen Miga (UCSC Genomics Institute)

Presentation Video | **Slides**

The reference genome was largely derived from the genomic information of a single individual and has historically consisted of data from individuals of European descent. The Human Pangenome Consortium aims to improve the representation of sequence diversity in the human population to build a new reference structure, in order to improve variant detection and clinical outcomes for all people. Miga shared current work to date from the project and the current goal of launching global partnerships to expand representation.

Ethical and Policy Considerations for Meaningful Inclusion in the *All of Us* Research Program

Katherine D. Blizinsky (All of Us Research Program)

Presentation Video | **Slides**

Blizinsky discussed the goal of the *All of Us* research program: to build a longitudinal cohort of deep, rich data on people living in the United States. A key part of this project is diversity and meaningful engagement—balancing participant interests and autonomy with maximizing the use of donated data. Blizinsky shared an update on the reach and engagement of the program, but also emphasized key factors that must continue to be fostered or built, including trust and mechanisms for engagement across barriers.

NCIG: National Centre for Indigenous Genomics

Hardip Patel (Australian National University)

Presentation Video | **Slides**

Patel discussed the inequitable health outcomes that indigenous communities in Australia face today, and emphasized the need to increase diversity in genomics studies to improve clinical outcomes. Underpinning these efforts are the need to respect data sovereignty of indigenous people, including collective benefit and authority to control the data. Building upon these principles, Patel shared current efforts, including community engagement and new studies underway.

Panel Q&A

Presentation Video

Track 3: Responsible Data

Panel: Responsible Data

Truly Privacy-Preserving Federated Analytics for Precision Medicine with Multiparty Homomorphic Encryption

Jean-Pierre Hubaux (EPFL)

Presentation Video | **Slides**

Hubaux discussed homomorphic encryption as a viable and secure way to carry out federated analysis for health data and precision medicine and streamline the process for conducting large-scale collaborative GWAS studies. Hubaux then shared the MedCO project, a developing system deployed in Swiss hospitals that uses homomorphic encryption as a building block.

CanDIG and Responsible Data

Jonathan Dursi (CanDIG)

Presentation Video | **Slides**

Dursi shared CanDIG's responsible data sharing principles for federation design and when building technical solutions, emphasizing that building trust among collaborators is key. Dursi then discussed implementation of these principles and current progress to date across CanDIG's distributed infrastructure.

Ethical, legal, and governance considerations in the processing, use, and sharing of real-world data for healthcare and health research

Rachele Hendricks-Sturrup (Duke-Margolis Center for Health Policy)

Presentation Video | **Slides**

Hendricks-Sturrup discussed the implication of real-world data cross the patient-consumer spectrum, a concept in which healthcare is transitioning into an around-the-clock activity involving the generation, use, and integration of all types of data. This leads to the intersection of regulated and unregulated health data, which must be considered from ethical, legal, and social perspectives so that in the face of health innovation, no one gets harmed.

Responsible Use: Return of Results

Anna Lewis (Harvard University)

Presentation Video | **Slides**

Lewis discussed the emerging ethical and legal aspects of supporting the return of clinically actionable results back to patients. While return of results is hard, resource-intensive, and in tension with the notion of meeting larger sample sizes, Lewis emphasized the duty to return results and the cultural shift taking place of viewing patients as partners and collaborators in the

research endeavor. Lewis discussed the GA4GH approved Return of Results policy as a helpful resource and reference point for managing the return of results.

Panel Q&A

Presentation Video

Panel: Real World Implementers

End-to-End Implementation of Standards: B1MG Proof of Concept

Tommi Nyronen (CSC-IT Center for Science)

Presentation Video | **Slides**

Nyronen presented a background of the EU's "1 + Million Genomes" project, which aims to bring together 22 EU countries, the UK, and Norway with the goal to have at least 1 million sequenced genomes accessible in the EU by 2022. Next, Nyronen shared an infrastructure proof of concept with a rare diseases use case utilizing GA4GH standard, before discussing next steps on the roadmap.

Opportunities for NIH Cloud Interoperability approaches to improve outcomes of pediatric diseases

Alisa Manning (Broad Institute)

Presentation Video | **Slides**

Manning discussed the potential for NIH cloud interoperability approaches to improve outcomes of pediatric diseases. The Pediatric Cardiac Genetics Consortium aims to bring together researchers to study data for the first time in the cloud, using health controls from two NHLBI cohorts.

Leveraging the H3Africa Infrastructure for effective and responsible genomic data sharing to accelerate genomic research in Africa

Mamana Mbiyavanga (University of Capetown)

Presentation Video | **Slides**

The H3Africa Consortium aims to facilitate African-based research with the goal of improving the health of African populations by building capacity in human and research infrastructure. Mbiyavanga shared the GA4GH standards used to date within the Consortium that are key to building infrastructure on the continent.

Enabling Clinical Data Discovery with Beacon

Lauren Fromont (Centre for Genomic Regulation)

Presentation Video | **Slides**

Fromont discussed the evolution of the Beacon API, from its initial "yes/no" response for data discovery to making the standard more relevant for clinical use. Fromont then shared an example

of implementing Beacon v2 in hospitals in Catalunya. The Beacon v2 marks an important step in enabling a robust and global Beacon network.

Track 4: Expanding the scope and Impact of GA4GH

Keynote: The role of PHA4GE in response to barriers in global (pathogen) data sharing

Alan Christoffels (SANBI)

Presentation Video | **Q&A Video**

There are many barriers to data sharing in the public health space, from lack of adequate technical solutions to legal and ethical implications. In this keynote talk, Christoffels discussed the role of the Public Health Alliance for Genomic Epidemiology (PHA4GE) in addressing these barriers. Christoffels shared the work of three groups within PHA4GE—the Data Structures Working Group, the Ethics and Data Sharing Working Group, and the Infrastructure Working Group—in facilitating and supporting responsible data sharing during an outbreak.

Panel: COVID-19 Data Sharing

COVID-19 Host Genomics Initiative

Ben Neale (Broad Institute)

Presentation Video | **Slides**

Neale discussed the aim of the COVID-19 Host Genetic Initiative, which aims to map the human genetic architecture of COVID-19 to understand why a large proportion of individuals are asymptomatic while others experience progressive, even life-threatening consequences. Neale shared findings from several studies on what is known so far, and emphasized the importance of an open community to allow for robust, real-time replication of findings and exchange of results.

National COVID Cohort Collaborative (N3C)

Melissa Haendel (University of Colorado Anschutz Medical School)

Presentation Video | **Slides**

Data sharing is critical to addressing unanswered questions in COVID-19, from predicting acute COVID-19 severity early in its course to diagnosing and treating long COVID. With the largest public HIPAA-limited dataset in US history, Haendel discussed the role of the National COVID Cohort Collaborative (N3C) to help tackle these topics.

Genomics to help us understand and control the COVID-19 Pandemic

Catalina Lopez-Correa (Genome Canada)

Presentation Video | **Slides**

Lopez-Correa discussed the role of CanCOGeN, a Genome Canada-led consortium, in using genomics to better understand and control the COVID-19 pandemic. Lopez-Correa dove into the work of the HostSeq and ViralSeq implementation committees, which aim to sequence and analyze host genomes and viral genomes, respectively. To unlock the power of this data, Lopez-Correa emphasized the importance of data sharing and interoperability to improve clinical outcomes for patients.

Panel Q&A

Presentation Video

Panel: Building Bridges to Industry and Clinical Care

Clinical Research Data Sharing Alliance: A multi-stakeholder alliance serving the clinical data sharing ecosystem

Bron Kisler, Clinical Research Data Sharing Alliance

Presentation Video | **Slides**

The Clinical Research Data Sharing Alliance (CRDSA) seeks to maximize the value of patient-donated data by ensuring broad secondary use and understanding of high-quality clinical research data. Kisler shared that key to accomplishing this goal is balancing the conflicting priorities of open science and patient privacy. To address this tension, CRDSA is developing collaborative solutions to address these challenges and share best practices in order to bridge the worlds of genomic data sharing and clinical research data sharing.

The Importance of data standards in the study of genomic medicine implementation

Marc Williams (Geisinger Health System)

Presentation Video | **Slides**

Williams shared the work of the MyCode Community Health Initiative, examining how lack of standards impacts testing, interpretation, and reporting—ultimately impacting the ability to scale health systems. Williams discussed challenges and potential solutions, sharing that the learning health system model is a viable framework for implementing precision health.

Data challenges in clinical lab testing

Rebecca Truty (Invitae)

Presentation Video | **Slides**

Invitae aims to bring comprehensive genetic information into mainstream medicine to improve healthcare. Truty described Invitae's two core tenets—that patients own their health data and health data is more valuable when shared—and the ways Invitae is supporting these goals. Truty discussed current challenges and potential solutions, emphasizing the importance of standards throughout the process.

Panel Q&A

Presentation Video

Closing Remarks

Presentation Video