GA4GH in Real World

Work Stream / Subgroup: Cloud

SELECT A GA4GH GROUP

Date (Day/Month)**:** Thursday, 9 October at 9:00 - 10:30 CEST
Time (CEST)**: 9 - 10:30 am**
Session chair(s):
- Name **In-person :** Brian D OConnor
- Name **Virtual : Venkat Malladi**

Link to slides: P GA4GH in Real World - Connect slide.pptx

Session description:

This session aims to showcase real-world implementations of GA4GH standards across research, clinical, and industry settings, highlighting their impact and adoption. Attendees will also identify challenges and opportunities in applying GA4GH standards, fostering discussions on improving usability, interoperability, and global adoption and identifying gaps.

Meeting aims:
- **Showcase real-world implementations of GA4GH standards across research, clinical, and industry settings, highlighting their impact and adoption.**
- **Identify challenges and opportunities in applying GA4GH standards, fostering discussions on improving usability, interoperability, and global adoption.**
- **Develop a list of requirements and use cases for existing API**
- **Determine if a new API standards needs to be developed or adopted**

Resources and links:

- 目 GA4GH in the real world
- 目 GA4GH Cloud AI/ML Working Group (Federated Machine Learning)

Agenda:

	Agenda item	Presented by	Item type	Time
1.0	Introduction	Brian & Venkat		5 min
1.1	Discuss minimal Starter Kit	Venkat	Presentation	5 minutes
2.0	eLWAZI Use case and hackathon results	Mamana Mbiyavanga	Presentation	20 minutes
2.1			Discussion	10 Minutes
3.0	Federated Machine Learning	Alex, Venkat and Kyle	Presentation	20 minutes
3.1			Discussion	10 minutes
4.0	From Prototypes to Production: Pacific Analytics' Kit for Modular GA4GH-Compliant Ecosystems	Javed and Anurag	Presentation	10 minutes
4.1	Expansion of minimal Starter Kit		Discussion	10 minutes
5.0	Wrap up			

Zoom recording:

hhttps://us02web.zoom.us/rec/share/mIWPQMJTmWI6YtOBQ5Su6cONEFNB6ZyK2LrII7UETBf7
jMsbmLMTWMVoDDUyoHss.yw_3j33cFr6tjO57
Passcode: .P#Rgz0E

Summary:

The session showcased real-world implementations of GA4GH standards, highlighting interoperable, federated, and privacy-preserving approaches to data sharing and analysis. Case studies such as IPCHiP demonstrated cross-border data sharing, while projects like the Federated Imputation Platform and Pacific Analytics Toolkit illustrated scalable, standards-based interoperability. Discussions emphasized the development of a minimal starter kit for practical deployments, the formation of a new Cloud, ML, and AI Working Group to advance federated machine learning, and the importance of regional engagement, particularly in Africa. Next steps include improving integration across GA4GH APIs, piloting the starter kit, expanding federated compute initiatives, and documenting community experiences to guide broader adoption.

High level key takeaways (5 overarching points from the key takeaways)

· **Advancing Real-World Interoperability Through GA4GH Standards**
Initiatives such as the International Precision Child Health Partnership (IPCHiP) and the Pacific Analytics Toolkit demonstrate the practical application of GA4GH principles in enabling cross-border, real-world data sharing. These efforts showcase how interoperable systems can bridge on-premise and cloud environments while maintaining data security and transparency in both successes and challenges.

· **Building Federated and Scalable Data Ecosystems**
Developments such as the Federated Variant-Level Matching and Imputation Platform and ongoing work in Federated Machine Learning (FML) reflect a broader move toward distributed, privacy-preserving data analysis. These approaches enhance efficiency and accuracy while minimizing the need for centralized data aggregation and supporting interoperability across regions and platforms.

· **Developing Foundational Infrastructure for Implementation**
The proposed "Minimal Starter Kit for Interoperability" highlights the importance of establishing a practical foundation for deploying interoperable systems beyond demonstration environments. Core components include authentication, authorization, data models, APIs, storage, logging, and auditing—all aligned with GA4GH standards and informed by defined scientific use cases.

· **Strengthening Regional and Global Collaboration**
Several initiatives underscore the value of inclusive engagement—such as regional collaborations in Africa and the Pacific—and cross-group coordination to ensure that GA4GH standards and tools address diverse implementation contexts. Transparent communication around both achievements and setbacks remains key to collective progress.

· **Integrating Ethical, Legal, and Security Considerations**
Across all projects, there is a shared emphasis on embedding strong ethical, legal, and security practices within technical frameworks. This ensures that innovations in federated and interoperable data systems are developed responsibly, maintaining trust and compliance in multi-jurisdictional environments.

Key takeaways:

1. Real-World Data Sharing: IPCHiP Case Study

- The International Precision Child Health Partnership (IPCHiP) exemplifies borderless data sharing using GA4GH principles.
- Key challenges include managing on-prem vs. cloud environments, ensuring data security, and providing metadata access.
- Emphasis on transparency in both successes and setbacks when adopting GA4GH standards.

2. Federated Variant-Level Matching and Imputation Platform

- A federated imputation platform, presented by Mamana Mbiyavanga, connects multiple imputation services through one interface, improving accuracy and efficiency.
- Built using GA4GH standards, currently supporting Michigan imputation software and multiple reference panels.
- Highlights the need for better integration between DRS and WES.
- Discussions emphasized regional engagement (e.g., in Uganda and Kenya) and starting with small-scale implementations.

3. Pacific Analytics Toolkit and Cloud Components

- Pacific Analytics demonstrated an open-source, framework-agnostic toolkit for building interoperable platforms using GA4GH standards.
- The system supports data discovery, workflow execution, and repository integration.
- Encourages collaboration across groups and integration of GA4GH standards with existing infrastructure.

4. Minimal Starter Kit for Interoperability

- Proposal for a "minimal starter kit" to deploy real-world interoperable systems beyond demo environments.
- Focuses on essential components: authentication, authorization, data models, APIs, storage, logging, and auditing.
- The approach involves defining scientific use cases, aligning on relevant standards, and deploying public data demos with stakeholder input.

5. Federated Machine Learning (FML) and Community Collaboration

- Federated ML discussed as a privacy-preserving alternative to centralized models.
- Challenges include data preparation, access permissions, and authentication/authorization.
- A new Cloud, ML, and AI Working Group formed to develop frameworks supporting federated ML across GA4GH standards.

- Emphasis on community engagement and addressing legal, ethical, and security aspects in federated environments**.**

Next steps:

1) Enhance Standards Integration
- Improve interoperability between DRS, WES, and other GA4GH APIs to support federated analysis platforms.

2) Advance the Minimal Starter Kit
- Define scope, core components, and priority standards.
- Develop and test a prototype deployment using public data.

3) Expand Federated Compute and ML Initiatives
- Continue work within the Cloud, ML, and AI group to define use cases and build a shared framework.
- Ensure strong links with security, ethics, and governance workstreams.

4) Promote Community Adoption
- Organize regional workshops (e.g., in Africa) to raise awareness and support small-scale pilot implementations.
- Foster collaboration among groups like Pacific Analytics and federated data initiatives.

5) Document and Share Learnings
- Publish case studies (e.g., IPCHiP, Federated Imputation) demonstrating real-world adoption of GA4GH standards.
- Encourage sharing of implementation experiences — both challenges and successes — to build community confidence.

Meeting minutes:

Case Study: [IPCHiP Guest Blog Post](#)
- Transparent sharing of successes and failures help the community learn
- Adaptability is key
- Build trust, strong communication, and governance frameworks early
- Lead pilot projects to assess approaches

"Minimal Starter Kit"
- Address authentication and authorization
- API interface
- Data model/schema
- Storage and indexing

Summary
- GA4GH Starter Kit enables practical, interoperable deployments
- Encourages incremental, standards-based adoption
- Demonstrates real-world federation, compute interoperability, and scalable workflows
- Builds a foundation for future production-ready implementations

eLwazi use case
- Implementation of GA4GH Standards
  - DRS for data access
  - WES for compute
  - TRS for workflow sharing
  - Using Terra workspaces
  - Cloud and HPC Environments: Azure, GCP, AWS, and Local
- Ability for researchers to access the data without moving the data
- eLwazi pilot platform using GA4GH starterkits (Refer to slides)
- eLwazi GA4GH Hackathon 2025
  - Understanding how GA4GH standards (DRS, Passports, WES, Data Connect) and TREs can be implemented and integrated
  - Exploring federated access

—----------

1. Case Study: International Precision Child Health Partnership (IPCHiP)

- Venkat presents a case study on **borderless data sharing** under the International Precision Child Health Partnership.
- The initiative aims to share **genomic and phenotypic data across institutions** while respecting local constraints.
- Challenges discussed include **on-premises vs. cloud environments**, **data security**, and **metadata access**.
- The group emphasizes the importance of **transparency** in both successes and challenges when adopting GA4GH standards.

2. Federated Variant-Level Matching and Imputation Platform
Use Case Overview

- Mamana Mbiyavanga presents the **federated variant-level matching platform** built using GA4GH standards.
- The platform connects multiple imputation services through a **single interface**, allowing researchers to impute from different reference panels.
- It saves time and improves accuracy by merging results from multiple panels.
- Current implementation uses **Michigan imputation software** and supports **multiple reference panels**.

Demo

- A live demo shows how users can **select services, upload data, and submit jobs**.
- Users can monitor, download, and merge results from different imputation services.
- Additional use cases include **organoid research** and **federated phenotype imputation**.

Discussion

- When asked about missing standards, Mamana highlights the need to better integrate **DRS** and **WES**.
- Deborah (Macquarie University) raises questions about expanding to **Uganda and Kenya**; Mamana stresses **workshops and awareness** as key enablers.
- Participants agree that **small-scale implementations** are an effective way to begin.

3. Pacific Analytics Toolkit and Cloud Components
Overview

- The **Pacific Analytics** team showcases their **open-source toolkit** for building platforms using GA4GH standards.
- Their **cloud components library** offers a **framework-agnostic and protocol-agnostic UI** for integrating GA4GH repositories.
- It supports **data discovery, workflow execution, and integration** with existing repositories and tools.

- Demonstrated capabilities include **running workflows in federated environments**, along with lessons learned.

Discussion

- **Monica Muñoz Torres** asks about the toolkit's primary use case and real-world impact.
- **Alex (Pacific Analytics)** explains that it enables **modular, enterprise-grade platforms** for clients.
- The conversation highlights the **open-source nature** of the components and potential for **cross-group collaboration**.
- **David (Camber)** suggests working together to integrate GA4GH standards with **existing infrastructure and tools**.
-

4. Proposal: Minimal Starter Kit for Interoperability

- Venkat introduces a **"minimal starter kit"** concept focusing on **core interoperable components**.
- The kit addresses **authentication, authorization, API interfaces, data models, storage, indexing, logging, and auditing**.
- The goal is to enable **real-world data sharing and computation**, moving beyond isolated demonstrations.
- The plan includes:
  - Understanding key **scientific scenarios**.
  - Defining **relevant standards**.
  - **Deploying public data demos**.
  - **Engaging stakeholders** to tailor implementations.

5. Federated Machine Learning: Concepts and Community Engagement
Introduction

- Venkat introduces **federated machine learning (FML)**, comparing it to centralized models and outlining challenges in **data discovery, access permissions, and privacy**.
- Issues such as **data formatting, preparation, and authentication/authorization** are highlighted.
- Emphasis is placed on ensuring **data protection and privacy** while enabling collaborative analysis.

Community and Next Steps

- Venkat announces a new **Cloud, ML, and AI working group** to tackle FML challenges.
- The group meets **monthly** to share use cases, discuss implementations, and develop solutions.
- Priorities include creating a **common framework** to support FML across GA4GH standards.
- The importance of **community engagement**, and addressing **legal, ethical, and security concerns**, is strongly emphasized.