GA4GH 13th Plenary Meeting Report



The GA4GH 13th Plenary Meeting, co-hosted by SciLifeLab, brought together members of the global genomics and health community for a week of collaboration, technical advancement, and strategic reflection. Nearly 300 attendees gathered in Uppsala, Sweden, and more than 300 participants joined virtually from around the world. The GA4GH Plenary meetings are crucial forums to convene people from across a wide variety of sectors — including research, healthcare, government, and academia — to discuss the path forward for responsible global genomic and health data sharing. The meeting centred on core themes of the value of genomic medicine, the evolving role of artificial intelligence (AI) in genomics, the importance of collaboration, and the vital cultivation of trust amongst all participants in the healthcare ecosystem.

The meeting was also highlighted in *Technology Networks* and several blog posts from meeting attendees (Ewan Birney, Deborah Ekusai Sebatta, Rukiya Mohamed Haji). You may also learn more about the meeting in our press release and in a video produced by SciLifeLab.

This Plenary, we had more than 50 posters as part of the GA4GH poster session. To view all posters presented during the meeting, visit the Poster Guide.

Thank you to our meeting sponsors, funders, and Plenary Programme Committee Members for dedicating time, energy, and resources to shaping this year's meeting and making 13th Plenary possible.

Demonstrating the value of genomic medicine

A recurring theme of the Plenary meeting was the need to demonstrate tangible clinical and societal benefits of genomic medicine, with a lens on how genomic medicine can improve patient outcomes, while also being cost-effective for healthcare systems. Presentations from both regional and national genomic initiatives highlighted a focus on forging collaborations between academia, government, and industry, sometimes through the development of nationally funded genomic platforms. Speakers emphasised the need to clearly communicate the value and long-term benefit of genomic medicine to educate the public and build trust.

The intersection of AI and genomics

The meeting highlighted how artificial intelligence is poised to transform genomics from a descriptive science into a predictive, AI-native engineering discipline. It also emphasised that this progress is fundamentally constrained by biomedical data that remains siloed, heterogeneous, and unready for AI. To enable this convergent future, GA4GH must evolve from building the foundational "highway" for data sharing to defining the "rules of the road" for AI, championing global standards for AI-readiness and ethical implementation.

The role of standards and international collaboration

Global challenges cannot be solved by any one country alone. Plenary speakers recognised the vital role organisations such as GA4GH play in developing interoperable standards for the secure sharing of data across borders. International convening bodies also play a crucial role in fostering dialogue and collaboration, across a wide array of geographic contexts. Speakers also supported the idea of creating GA4GH "regional hubs" to engage more deeply with local scientific communities and support adoption of GA4GH standards.

The importance of trust in data sharing

Building trust is fundamental to the success of global data sharing initiatives. Trust is built on more than just technical security; it requires robust governance, clear legal frameworks, and transparency. Our Plenary speakers emphasised the importance of acknowledging diverse cultural values and engaging meaningfully with patient and participant communities.

Table of contents

Plenary Day 1

- Opening remarks
- Local welcome
- Keynote Presentation: Patrick Tan
- Panel: Global and regional perspectives on data sharing
    - Binay Panda
    - Hyun-Young Park
- Keynote presentation: Annemieke Ålenius
- Panel: European health data sharing
    - Sandrine Richard
    - Kjetil Taskén
    - Samuli Ripatti
    - Raili Sillart
- Keynote presentation: Nara Sobreira
- Panel: Real-world implementation of GA4GH standards
    - Frederik Otzen-Bagger
    - Wendi Bacon
    - Marko Malenic
- Evening talkshow: Future landscape of genomic data sharing

Plenary Day 2

- Opening remarks
- Keynote presentation: Carolina Vitabäck
- Panel: Making the case for data sharing to advance genomics
    - Lauren Maxwell
    - Wee Hwee Lin
    - Christian Martshall
    - James Buchanan
- Keynote presentation: Richard Rosenquist Brandell
- Panel: GA4GH supporting genomics in clinical practice at the national level
    - Frédérique Nowak
    - Jakris Eu-ahsunthornwattana
    - Tomi Mäkelä
    - Sergi Beltran
- Keynote presentation: Ewan Birney
- Panel: Applications of AI and ML for genomics and global health
    - Mathias Ekman
    - Joep de Ligt
    - Nevenka Dimitrova
    - Ma'n Zawati

Connect Day 1

- Connect welcome and GA4GH updates
- The Future is FAIR beyond genomics: building bridges to the Research Data Alliance (RDA) and other FAIR research data initiatives
- DUO Diligence: spring cleaning and roadmapping
- From global standards to local implementation: applying REWS' Model Consent Clauses in African research settings
- GIF TRE Open Suite: towards GA4GH-compliant secure and privacy-preserving computations
- Discovery Work Stream in 2026
- Bridging location and phenotype
- Agentic Science: Enabling AI Agents to Reason with Biomedical Data via GA4GH Standards

Connect Day 2

- Literature derived variant evidence: toward GA4GH compatible representation
- Influencing EPR and LIMS suppliers to adopt FHIR standards and native interoperability (part 1)
- GA4GH in the real world
- Capturing cancer genomics data attributes in Beacon v2.0

(Connect day 2 cont.)

- FAIR in action using GA4GH products: hands-on FAIRification from a fresh start to the FAIRest of them all
- Variant Representation alignment: bridging VRS, VCF, HGVS, and SPDI
- Influencing EPR and LIMS suppliers to adopt FHIR standards and native interoperability (part 2)
- Passports that scale: lightweight tokens and federated trust
- Where is global? Geolocation in genomics
- Leveraging GA4GH's Tools & Standards to Balance Pathogen Data Sovereignty with Open Access
- Next steps on the use of Verifiable Credentials (models and protocols) for improving Passports
- Scaling the Federated Variant-Level Matching Network
- An analytics dashboard for everything GA4GH

Connect Day 3

- The next step toward clinical application of GA4GH products: a sequel! (Part 1)
- Securing Beacon Networks with Passports and trust federation
- The Cloud Work Stream road map 2026 and beyond
- The next step toward clinical application of GA4GH products: a sequel! (Part 2)
- Solving Beacon queries
- Shaping the future of the GA4GH Cancer Community of Interest
- Generating guidelines for community-driven portals for infectious disease research
- In the (GIF) Spotlight: highlighting GA4GH implementations
- Path to normative: implementing a GA4GH maturity model

National Initiatives Forum

- Introduction to NIF
- NHS genomic updates
- Updates from National Initiatives
- Updates from community tools
- Discussion: strategy for NIF going forward

Plenary Day 1

Opening remarks

Recording | Slides

GA4GH Chair Heidi Rehm kicked off the 13th Plenary meeting with opening remarks, providing an overview of the meeting's key themes. This year's plenary meeting focused on global and regional perspectives on data sharing, real-world implementation of GA4GH standards, the future landscape of genomic data sharing, and applications of artificial intelligence (AI) and machine learning (ML) for genomics and global health.

Local welcome

Recording | Slides

Co-Director of SciLifeLab Mia Phillipson welcomed attendees to Uppsala, Sweden, highlighting the city's long academic history and prominent researchers such as Carl Linnaeus, who pioneered life science standardisation. Phillipson shared more about SciLifeLab, Sweden's largest national research infrastructure, which was founded 15 years ago and aims to enable groundbreaking life science discoveries.

Keynote presentation by Patrick Tan: “The Singapore National Precision Medicine Programme — from local impact to global dialogue”

Recording | Slides

Patrick Tan, Executive Director of Precision Health Research, Singapore (PRECISE) discussed Singapore’s approach to precision health, emphasising its use of genomics to address rising healthcare costs and an ageing population. The PRECISE-SG100K project, which has sequenced 100,000 individuals, has created a multi-ancestry Asian dataset to identify genetic risks and develop cost-effective interventions. The long-term goal is to build a genomics-informed learning healthcare system. Tan highlighted the importance of GA4GH standards in fostering international collaboration and announced that the GA4GH 14th Plenary meeting will be held in Singapore.

Panel: Global and regional perspectives on data sharing

Binay Panda: “Data sharing in the world’s most populous country”

Recording | Slides

Binay Panda, Professor at Jawaharlal Nehru University, provided an overview of the challenges and early stages of data sharing in India. The country’s immense diversity, with over 122 major languages and multiple ethnolinguistic groups, presents both a unique opportunity for genetic research and significant legal complexities. The right to privacy is a fundamental right established under article 21 of the Indian Constitution, and a new 2023 law on Digital Personal Data Protection provides the first-ever legal framework for digital data, including genomics. While this law is a crucial first step, it lacks specific guidance for sectors such as healthcare. Panda emphasised the importance of establishing accurate, specific consent processes and learning from international collaborations such as those from GA4GH.

Hyun-Young Park: “KNIH’s experiences on data sharing”

Recording | Slides

The Korea National Institute of Health manages more than 30 large-scale cohorts and registries. Hyun-Young Park, Director of the National Institute of Health in South Korea, highlighted key initiatives, such as the Korean Genome and Epidemiology Study and the National Bio Big Data Project, that are generating whole-genome sequencing data. To address the unique genetic makeup of the Korean population, the country developed a specific biobank array and is now working on a telomere-to-telomere reference panel. All data sharing is managed through a central repository, and is governed by strict laws, such as the Personal Information Protection Act and the Bioethics and Safety Act. While international data sharing is legally challenging, Park discussed the use of federated analysis models to promote global collaboration without transferring raw data abroad.

Keynote presentation by Anniemieke Ålenius: “Harmonising health data sharing across Europe: opportunities and challenges.

Recording | Slides

Annemieke Ålenius, Director of the Swedish eHealth Network, presented on the new European Health Data Space (EHDS) regulation, which aims to improve data interoperability and use across Europe. The EHDS was developed in response to the COVID-19 pandemic, which highlighted a lack of cross-border data sharing. The regulation establishes four core pillars: 1) citizens’ rights to their health data; 2) rules for primary use of data; 3) guidelines for secondary use of data; and 4) standards for electronic health record systems.

Panel: European health data sharing

Sandrine Richard: “Europe’s health data in a global genomic era: trust, technology, and transformation”

Recording | Slides

Sandrine Richard, a board member of Ethik-IA, continued discussion of the European Health Data Space (EHDS), stating that this new regulation is part of Europe’s response to the paradox of having vast amounts of health data that are fragmented and not interoperable. Richard emphasised that the key challenges to genomic and health data sharing are not technical, but rather related to trust, governance, and ethics. The EHDS, along with the Artificial Intelligence Act, aim to provide the legal framework for a unified data strategy based on standards and transparency, with the goal of creating a secure, interoperable system for both primary and secondary use of data.

Kjetil Taskén: “European data sharing between precision medicine cancer trials”

Recording | Slides

Kjetil Taskén, Professor and Head of Institute at Oslo University Hospital, presented on the Precision Cancer Medicine Repurposing System Using Pragmatic Clinical Trials (PRIME-ROSE), a multi-stakeholder project that aggregates genomic data for cancer clinical trials across Europe. With 28 partners across 19 countries, PRIME-ROSE is designed to overcome the rarity of specific genomic profiles in precision medicine. By standardising clinical data to the OMOP standard and aligning biomarker definitions, PRIME-ROSE enables data aggregation, accelerating trial recruitment and filling cohorts more efficiently. Based on the project’s success, the Joint Action for Personalised Cancer Medicine, a new government-based initiative, will use PRIME-ROSE as a blueprint to expand data aggregation across Europe.

Samuli Ripatti: “Translational research at the time of large-scale biobank studies”

Recording | Slides

Samuli Ripatti, Director of the Institute for Molecular Medicine Finland and Professor at the University of Helsinki, discussed Finland’s robust health data strategy. This includes the use of nationwide registries and the FinnGen project, a public-private partnership with genetic data from more than 500,000 individuals. This comprehensive data is used for diverse research, from large-scale epidemiological studies to training generative AI models for precision health. He showcased the practical application of polygenic risk scores to predict conditions and diseases, such as breast cancer, and highlighted the effort to integrate these tools into clinical settings. Ripatti emphasised Finland’s commitment to building trusted research environments and developing advanced analytical tools to ensure the data’s utility for research and healthcare.

Raili Sillart: “Health wealth: unlocking Europe’s health and genomic data for innovation and better lives”

Recording | Slides

Raili Sillart, Vice Chair of ICPerMed and Innovation Adviser at the Estonian Ministry of Social Affairs, highlighted Europe’s unique position in health data, due to its rich longitudinal health records and a strong foundation of public trust. Estonia, with its small size and innovative spirit, serves as a model for a digitally connected society. The country’s biobank and e-health system demonstrate how to use genomic data for personalised health and prevention, empowering citizens. The vision is to avoid fragmentation by connecting research, clinical practice, and industry, and ensure that health data becomes a source of innovation and resilience across Europe and beyond — while keeping the patient health outcome as the main focus of all efforts.

Keynote presentation by Nara Sobreira: “Phenotypic and genomic data sharing to facilitate variant classification and gene discovery”

Recording | Slides

Nara Sobreira, Associate Professor at Johns Hopkins University, discussed her work on sharing phenotypic and genomic data to identify novel disease genes. She introduced several tools, Gene Matcher and Variant Matcher, which are part of the Matchmaker Exchange — a network that connects databases globally. These tools have helped in the discovery of more than 720 novel disease genes and have reclassified a significant number of uncertain variants. Nara stressed the importance of continued data sharing, collaboration, and the need for new partners, especially clinical labs, to join this global effort.

Panel: Real-world implementation of GA4GH standards

Frederik Otzen Bagger: “Real-world implementation of GA4GH standards”

Recording | Slides

Frederik Otzen Bagger, Head of Bioinformatics, Genomic Medicine at Rigshospitalet, discussed the importance of data standards in clinical genomics. Low-level standards such as VCF help simplify internal pipeline work, while high-level standards such as Beacon v2.0 and OMOP have enabled Rigshospitalet to securely share and analyse data, accelerating the discovery of novel disease genes. Looking ahead, the hospital plans to adopt other GA4GH standards such as Phenopackets and the Data Use Ontology. Bagger concluded by calling for greater standardisation of data production, emphasising that it is essential for the efficiency and success of future federated data networks.

Wendi Bacon: “HDR UK technology and GA4GH standards”

Recording | Slides

Wendi Bacon, Product Owner at Health Data Research UK (HDR UK), provided an overview of HDR UK, a UK-based charity focused on streamlining health data research to prevent duplicated efforts. The organisation operates on a non-profit, open-source model and works with government, research councils, and other charities. Key initiatives include the Health Data Research Gateway for discovering data and Cohort Discovery for generating aggregated information from federated datasets. They also developed a Safe People Registry to securely verify and monitor researcher access to sensitive data, aligning with the “five safes” principles. HDR UK aims to integrate its tools with GA4GH standards, such as Beacon, Passports, the Task Execution Service (TES), and Data Use Ontology (DUO), to improve interoperability and foster global collaboration.

Marco Malenic: “Real-world htsget for data sharing”

Recording | Slides

Marco Malenic, a software engineer at the University of Melbourne, discussed htsget, a GA4GH standard that addresses the challenge of sharing only the necessary and relevant genomic data. Malenic presented htsget-rs, an implementation that dynamically sends required reads and variant data to a user based on their authorisation. This method eliminates the need to statically strip data, saving storage and allowing for flexible access control. Malenic highlighted the difficulty of implementing authorisation and the importance of separating authorisation from the data itself. He suggested that future work could involve integrating with other standards, such as the GA4GH Data Repository Service (DRS) and Passports, and called for community discussion to clarify the htsget specification.

Evening talk show: Future landscape of genomic data sharing

Recording

Paul Franks (Lund University), Mercury Shitindo (Africa Bioethics Network), Tazro Ohta (Chiba University), and Tiffany Boughtwood (Genomics Australia) discussed the challenges and opportunities of genomic data sharing. They highlighted the need for culturally and legally sensitive approaches, due to fragmented systems, diverse values and needs, the history of colonialism, and a priority to ensure Indigenous data sovereignty concerns. The speakers emphasised the importance of building trust by ensuring shared benefits and direct engagement with communities across the healthcare ecosystem. They also addressed the complexities of clinical translation, stressing the need to align research with local health burdens and for standards to facilitate the translation of genomics into clinical care. The discussion concluded by underscoring the value of communication and openness in a world of shifting political landscapes and data security risks.

Plenary Day 2

Opening remarks

Recording | Slides

At the start of day two, Peter Goodhand, CEO of GA4GH, and Heidi Rehm, Chair of GA4GH, provided a summary of GA4GH's journey and progress. GA4GH has grown to include 5 Host Institutions, 28 Driver Projects, and several hundred contributors worldwide. The organisation's ecosystem now includes eight Work Streams, five Communities of Interest, and collaborations with other standards organisations. Over the years, GA4GH has delivered more than 40 technical standards and policy frameworks for the responsible sharing of genomic and health data. A major focus moving forward is ensuring that these products are implemented effectively across a variety of contexts to improve health outcomes.

Keynote presentation by Carolina Vitabäck: “When data means hope: a personal perspective on sharing responsibly”

Recording | Slides

Carolina Vitabäck, a consultant in information governance and information security and Board Member of Barncancerfonden South, shared her daughter’s two-and-a-half-year battle with leukemia. She recounted the decision to consent to data sharing, a process she believes could be improved with more straightforward information. Carolina shared that the "All Together" protocol personalised her daughter’s treatment and acknowledged that her survival was a result of the progress made through

data and research. She stressed the importance of global collaboration and safe data management, believing that "shared data means shared hope." Vitabäck's story was a poignant reminder to all about the importance of the work we do, underscoring the critical role of data in advancing precision medicine and saving lives.

Panel: Making the case for data sharing to advance genomics

Lauren Maxwell: "How do we measure the costs, benefits, and harms of biomedical data and sample reuse?"

Recording | Slides

Lauren Maxwell, Junior Group Leader at the Heidelberg Institute of Global Health, delivered a presentation on the importance of demonstrating the impact of both sharing and not sharing biomedical data. She argued that while the benefits are obvious to researchers, they are not always clear to policymakers, leading to an overemphasis on hypothetical harms. Citing Hans Rosling, she stressed the importance of a data-driven approach to prove that the benefits outweigh the harms. Her research, a scoping review, revealed that while many studies focus on benefits, there is a lack of rigorous, comparative analysis on the costs of both reusing and not reusing data. The presentation concluded that transparent metadata and clear governance, as seen in Estonia, are essential for building trust and achieving the ultimate goal: positive impact on health.

Wee Hwee Lin: "Economy of scale: population-level implementation of pharmacogenetic testing in Singapore"

Recording | Slides

Wee Hwee Lin, Associate Professor at the National University of Singapore, discussed the Precision Health Research, Singapore (PRECISE) programme, which aims to integrate genomics into healthcare. Having completed phase two (SG100K) with data from 100,000 individuals, they are now scaling up to implement precision medicine at a population level. This requires rigorous evaluation, particularly through a cost-effectiveness analysis for pharmacogenomic testing. The challenge lies in the lack of local, granular data representative of the Asian population. To address this, PRECISE is linking SG100K data with electronic medical records in a secure platform called "TRUST," guided by responsible data sharing principles. The initiative is now looking to explore the most cost-effective approach to implementing pharmacogenetics panel testing at a national level in Singapore.

Christian Marshall: “Genomic data sharing in a diagnostic lab network”

Recording | Slides

Christian Marshall, Clinical Laboratory Director at the Hospital for Sick Children (SickKids), provided an overview of the AllforOne Clinical Genomics Network in Canada. The project focuses on sharing genomic data from clinical labs to improve rare disease diagnosis. In the province of Ontario, more than 14,000 individuals have been sequenced for suspected rare genetic diseases, with a 28% diagnostic rate. The network, initially a partnership between SickKids and the Children’s Hospital of Eastern Ontario (CHEO), aims to expand nationally in Canada. Data is shared across three levels, from basic variant information to detailed phenotypic data, with over 90% of patients consenting. This sharing has reduced analysis time, increased diagnostic accuracy, and allowed for the evaluation of phenotypic expansion. The ultimate goal is to standardise and ensure equitable diagnostics across Canada.

James Buchanan: “The value of data sharing to advance genomics: a health economics perspective”

Recording | Slides

James Buchanan, Senior Lecturer at Queen Mary University of London, discussed a health economics perspective to determine the value of data sharing, and practical steps to generate more evidence in the future. He highlighted the need to move beyond traditional economic evaluations and quantify the broader value of data sharing, such as accelerating discoveries and improving diagnostic yields. Buchanan cited a study from Australia that showed how re-analysing exome data in infants with rare diseases was cost-saving and improved diagnostic yield, and highlighted how the discovery of ReNU syndrome was enabled by international data sharing. He concluded that while direct empirical evidence on the economic value of data sharing is scarce, a signal of its value is emerging. Buchanan recommended that recent case studies can be examined to generate more evidence.

Keynote presentation by Richard Rosenquist Brandell: “Genomic Medicine Sweden: advancing precision medicine at the national level”

Recording | Slides

Richard Rosenquist Brandell, Director of Genomic Medicine Sweden (GMS), shared the evolution of GMS — a national initiative with the goal of implementing genomic technologies in healthcare. Starting in 2017 as a bottom-up effort, GMS has grown into a government-funded infrastructure. The initiative is a collaborative network of seven genomic medicine centres that perform precision diagnostics for rare diseases, cancer, and infectious diseases. GMS has developed a National Genomics Platform to facilitate data sharing and is actively incorporating technologies such as whole-genome sequencing and long-read sequencing. Rosenquist Brandell highlighted successful projects, such as the Testbed Sweden for Precision Health in Cancer, and ambitious plans such as the Precision Omics Initiative Sweden (PROMISE) initiative, which aims to create large population cohorts for multi-omics research.

Panel: GA4GH supporting genomics in clinical practice at the national level

Frédérique Nowak: “PFMG2025: Integrating genomic medicine into the national healthcare system in France”

Recording | Slides

Frédérique Nowak, Coordinator at Inserm, discussed the French Genomic Medicine Initiative, a project launched in 2016 to integrate genomic medicine into France's healthcare system. The program is built on a strong, national framework with two clinical laboratories providing whole genome sequencing for rare diseases and cancer, ensuring equal access across the country. To date, more than 50,000 requests have been received, with a 31.2% diagnostic rate for rare diseases. Nowak discussed a continuum between care and research, where data reuse and reanalysis can accelerate research and improve diagnosis and care. Challenges remain, such as clinical biological interpretation and turnaround time, the complexity of the genomic medical pathway, and data access and sharing for research purposes, which will be addressed in a future phase.

Jakris Eu-ahsunthornwattana: “Genomics Thailand: five years, 50,000 genomes, and beyond”

Recording | Slides

Jakris Eu-ahsunthornwattana, Project Lead at Genomics Thailand, discussed Thailand’s national genome programme that began in 2019 to sequence 50,000 individuals over five years. The initiative, funded by the cabinet, initially focused on research to build a reference genome but has since expanded to include ethics, data management, and workforce development components. The

programme's second phase aims to expand sequencing efforts, including incorporating long-read technology for rare diseases and sequencing of 200,000 newborns, and integrate findings into the healthcare system. To ensure compliance to laws, regulations, and ethical standards, a data access committee will govern sharing within a trusted research environment.

Tomi Mäkelä: “Cancer genomics supporting clinical practice”

Recording | Slides

Tomi Mäkelä, Director of the Finnish Cancer Center, presented on Finland's iCAN project, an initiative that combines precision cancer medicine with digital health to advance research in cancer. The project aims to understand cancer-host crosstalk, identify actionable targets, develop digital diagnostics and therapies. Tumour samples, liquid biopsies, and dynamic health data from national registries and clinical systems are integrated into a secure research environment for data analysis and discovery. With more than 7,400 recruited patients, over 1,000 have been identified with actionable alterations, enabling a mechanism to return research findings to patients. This effort is shaping national plans for further integrating research into clinical care, and actively promoting national and international data sharing.

Sergi Beltran: “Genomic data sharing in Spain: examples from clinical research and healthcare”

Recording | Slides

Sergi Beltran, Head of Bioinformatics Unit at the Centre Nacional d'Anàlisi Genòmica (CNAG) in Barcelona, Spain, discussed the IMPaCT-GENóMICA initiative. The goal is to establish a coordinated framework with the National Health System, the Spanish Autonomous Regions, and research institutes to equitably integrate precision medicine across Spain. The project brings together more than 100 hospitals and three sequencing centres to provide whole genome sequencing for rare diseases and cancer. Data is shared through a central platform that collects phenotypic information and leverages tools such as Beacon v2.0 for federated data discovery. While still in its early stages, IMPaCT-GENóMICA has already shown a significant increase in diagnostic yield for rare diseases, demonstrating the value of collaboration and coordination.

Keynote presentation by Ewan Birney: “Genomics, Imaging, and AI — three technologies spanning research to healthcare delivery”

Recording | Slides

Ewan Birney, Deputy Director General at the European Molecular Biology Laboratory (EMBL), discussed the convergence of genomics, imaging, and AI. Birney highlighted the success of AI models, such as the protein-folding tool AlphaFold, which can tackle grand biological challenges. He then delved into a new AI model, "Delphi," developed by his research group. This model applies large language model principles to healthcare data, treating health events as a sequence of "tokens" with variable time. Trained on UK Biobank data and validated on Danish health records, Delphi accurately predicts a wide range of future health events like heart attacks and dementia, demonstrating the potential for AI in epidemiology and healthcare planning.

Panel: Applications of artificial intelligence (AI) and machine learning (ML) for genomics and global health

Mathias Ekman: “Multimodal biomedical GenAI for precision medicine”

Recording | Slides

Mathias Ekman, Director of Digital Health Platforms and Partner Ecosystems at Microsoft, discussed how generative AI and large language models (LLMs) are accelerating the slow cycles of drug discovery and clinical implementation. Ekman highlighted the emergence of multimodal biomedical AI, which combines different data types such as images, clinical records, and even proteins, which are “languages” too. A key example is the BrainChild project in Sweden, which integrates diverse data from childhood cancer patients to improve survival rates. The core idea is to leverage AI technologies to translate cutting-edge research findings into tangible healthcare benefits faster and more effectively.

Joep de Ligt: “Improving outcomes for cancer patients, the need for federated learning health systems”

Recording | Slides

Joep de Ligt, Data Lead at the Hartwig Medical Foundation, presented on the Netherlands' goal to integrate genomics into clinical practice for adult cancer patients. His team established a whole-genome sequencing service for routine care, while simultaneously collecting consent for data reuse. This data has been shared with over 300 research groups globally. A key success story involved using machine learning to predict tumour origin. This led to improved treatment options for 67% of patients and demonstrated a significant survival benefit. The talk also highlighted a challenge: validating biomarkers for non-response, which requires a much larger patient cohort than is available in the Netherlands, underscoring the need for international data sharing.

Nevenka Dimitrova: “Towards federated learning with genomics and spatial technologies”

Recording | Slides

Nevenka Dimitrova, Computational Oncology Advisor at the Memorial Sloan Kettering Cancer Center, presented on the application of AI and spatial technologies to cancer research, highlighting their approach to looking at cancer ecosystems. She explained how sophisticated imaging techniques allow us to zoom in on cancer biology, revealing the physical interactions between tumour cells and the immune system. Her team uses AI foundation models and vision transformers to detect, classify, and quantify these markers, with the goal of predicting treatment responses. Dimitrova also discussed a successful hackathon using microbiome data and federated learning, which demonstrated high accuracy in classifying sequencing data. She concluded by advocating for the new GA4GH AI Work Stream to aggregate spatial data, breaking down silos and fostering global collaboration for shared discovery.

Ma’n H. Zawati: “Responsible use of AI: key normative updates”

Recording | Slides

Ma’n H. Zawati, Research Director and Associate Professor at the Centre of Genomics and Policy at McGill University presented on the regulatory landscape of AI and genomics, highlighting its convergence with law and ethics. He noted that 77 countries have launched national AI strategies, and there is a global push for frameworks to address issues of equity and bias. He outlined three key normative approaches: risk-based, principle-based, and a self-regulation approach. Developing guidance for the responsible use of generative AI are key next steps of the GA4GH Regulatory & Ethics Work Stream (REWS). This includes principles such as ensuring human oversight, promoting transparency, and mitigating biases. The goal is to move beyond identifying challenges to actually implementing practical solutions for safety and accuracy in AI.

Connect Day One

Connect welcome and GA4GH updates

Recording | Agenda | Slides

GA4GH Chief Product Officer Sasha Siegel shared recent progress within GA4GH, noting a transition towards focusing on the implementation of GA4GH standards. GA4GH's forward-looking strategy focuses on product uptake, implementation, and engagement by developing tools such as a product maturity model, an analytics dashboard, and an implementation registry. With the growing use of artificial intelligence (AI) in the field, GA4GH is launching its first new Work Stream in eight years, dedicated to AI in genomics.

Following Siegel's welcome, Co-Leads from GA4GH's eight Work Streams provided updates to the community.

Genomic Knowledge Standards (GKS) Work Stream update

- In 2025, the GKS Work Stream released three specifications: Variation Representation Specification (VRS) v2.0, Categorical Variation Representation Specification (Cat-VRS) v1.0, and the Variation Annotation Specification (VA-Spec) v1.0.
- To drive engagement, the GKS Work Stream holds quarterly "Open House" meetings on key progress and updates and more technically focused "Think Tanks" for developers.

Clinical & Phenotypic Data Capture (Clin/Pheno) Work Stream update

- The Phenopackets standard is undergoing technical revisions and enhancements based on community feedback to inspire wider adoption.
- Family Health History standards are under development, including Pedigree and the Kinship Ontology.
- A newly formed Exposome Study Group is assessing the need to integrate environmental exposure and genomics data to better understand their interplay in disease, and considering the role GA4GH could play in standardising exposure data.

Discovery Work Stream update

- The Data Connect standard is seeking collaborators for the development of v2.0, which includes enhanced support for GA4GH Passports, linking across tables, and adding privacy-preserving queries.
- The Beacon standard for federated querying of genomic and phenotypical data is experiencing increasing adoption, with v3.0 actively in development.
- The Experiments Metadata Standard, which provides a minimal checklist and dictionary of properties to characterise -omics experiments, has undergone the product development and

approval process and plans to expand its scope to include deeper characterisation for specific categories such as transcriptomics and targeted sequencing.

Cloud Work Stream update

- The Cloud Work Stream mapped 22 registered Data Repository Service (DRS) servers, revealing an ecosystem with over 72 petabytes of data. The Work Stream aims to push for wider mapping of its other standards.
- The Cloud Work Stream aims to evolve new complementary standards, such as the Portable Format for Biomedical (PFB) data and RO-Crates.

Large Scale Genomics (LSG) Work Stream update

- The Whole Genome Sequencing (WGS) Quality Control (QC) standards, which defines a minimal set of QC metrics, has undergone the product development and approval process. The group has future plans to expand to long-read sequencing and somatic variant approaches.
- The team behind the refget Sequence Collections standard, approved earlier this year, shared future plans to focus on integration into file formats and extend the concept to pangenomes.

Data Security Work Stream update

- The Data Security Work Stream's primary focus is on the Passports standard. The team is tackling implementation challenges, such as scalability and performance due to large token sizes. The team is developing task-specific tokens to address this issue.
- The Work Stream is also exploring Verifiable Credentials as a new way to represent Passport Visas.
- The launch of the newly formed Data Passports Study Group is imminent. The standard aims to enable computable data access policy governance, tied to the data itself.

Data Use & Researcher Identities (DURI) Work Stream update

- DURI is planning a v3.0 release of the Data Use Ontology (DUO).
- The Work Stream plans to collaborate with the Data Security Work Stream on the new Data Passports initiative.

Regulatory & Ethics Work Stream (REWS) update

- REWS highlighted key updates and initiatives, one of which is the ongoing Public Attitudes for Genomic Policy blog post series. The posts aim to make findings on public trust in genomics and data sharing more accessible to policymakers and the public.
- REWS is moving forward with a new Position Statement on Genetic Discrimination, taking a human rights and multi-stakeholder approach to develop a flexible policy framework that can evolve with changes in the genomics field. Future plans aim to address potential discrimination issues in the context of AI.

The future is FAIR beyond genomics: building bridges to the Research Data Alliance and other FAIR research data initiatives

Recording | Agenda | Slides

The meeting focused on fostering collaboration between GA4GH and the Research Data Alliance (RDA) to extend Findable, Accessible, Interoperable, and Reusable (FAIR) data principles across domains. The session highlighted RDA's community-driven approach to research data sharing, its interest and working groups, and a strategic partnership with GA4GH established in 2023.

Lightning talks covered RDA initiatives such as the Verification of Genomic Annotations Working Group (which is adopting GA4GH standards such as refget and the Data Repository Service [DRS]), RO-Crate for packaging research objects with JSON-LD, FAIR Mappings for bridging metadata models, and HealthDCAT AP for harmonising health data catalogues in the European Health Data Space (EHDS).

Discussions explored aligning GA4GH products (e.g. Passports/Visas, Phenopackets, Beacon, and the Data Use Ontology [DUO]) with non-genomics standards, leveraging technologies such as linked data, JSON-LD, and signposting for better interoperability, and addressing obstacles such as silos and the need for extensible APIs.

The session ended with a Q&A and calls for ongoing engagement, including joint sessions and hackathons, to avoid reinventing standards and ensure genomics fits into broader ecosystems such as healthcare and biodiversity.

Key takeaways

- Bridging domain-specific (GA4GH) and domain-agnostic (RDA) communities can prevent silos, promote reusable standards, and extend FAIR principles to non-genomics areas such as biodiversity and healthcare, but requires active alignment to avoid "not-invented-here" issues.
- Technologies such as JSON-LD, RO-Crate, and FAIR mappings offer practical ways to enhance interoperability, enabling GA4GH standards to integrate with broader ecosystems without major reinvention.
- Ongoing collaboration is essential, with opportunities for genomics to influence and adopt from RDA while pushing research into healthcare applications.

Priority action items

- Organise joint activities, such as talks, hackathons, or co-located events between GA4GH Work Streams and RDA groups (e.g. Trusted Research Environments, FAIR Mappings).

- Set up a mechanism for regular check-ins (e.g. biannual meetings) to identify high-priority alignments between GA4GH products and RDA outputs, involving individuals familiar with both organisations.
- Explore integration of linked data technologies (e.g. JSON-LD) into GA4GH APIs, potentially starting with variant annotations, and mapping them to broader standards such as schema.org or HealthDCAT.
- Join relevant RDA groups (e.g. Health Data Interest Group) and attend plenaries (e.g. Brisbane event) to facilitate collaboration.

DUO diligence: spring cleaning and roadmapping

Recording | Agenda | Slides

This session provided a productive opportunity to assess the status of the Data Use Ontology (DUO) through interactive, hands-on tracker reviews. Participants explored the extensibility of DUO and encouraged the idea of a DUO core with bespoke modules and extensions. The group also highlighted the applicability of Data Passports and other ontologies to handle questions around dataset access request fees, as well as the need for specific use cases and requirements.

Key takeaways

- DUO should be kept small and focused; additional needs could be accommodated through separate modules and extensions for specific communities.
- There is support for extending the standard, but requires following the hierarchy and classes of terms as they are represented.
- Data access fees may be better handled in Data Passports or other ontologies instead of DUO, given the latter's emphasis on control and consent conditions

Priority action items

- Future Data Use & Researcher Identities (DURI) Work Stream meetings will continue discussions around statistics of usage, potential deprecation, and development of an ontology for terms related to representing data access requests.
- Assess needs from specific communities to extend DUO and assess how to best accommodate DUO codes — balancing their stability with a desire to not reidentify terms at a later stage if they get promoted to "core."
- Propose a DUO schema as part of Data Passports.

- Review current DUO content and assess which terms are derived from individuals' consent and which are not. The latter could be a candidate for a DUO extension into a Data Access Request ontology (which would also cover terms such as "access fee required").

Get involved!

Join conversations around DUO and other DURI activities on the DURI Slack channel! Any interested individuals are also invited to attend the upcoming DURI meetings. For more information, please contact Jessica Seegobin.

From global standards to local implementations: applying REWS' Model Consent Clauses in African research settings

Recording | Agenda | Slides

The session explored the challenges of adapting global genomics standards, such as the GA4GH Regulatory & Ethics Work Stream (REWS) Consent Clauses, to ensure applicability in African research settings. Key input from the African Bioethics Network (ABN) revealed the importance of representing local relational ethics and consent frameworks and prioritising community and collective decision making. The discussions reiterated the significance of integrating diverse ethical frameworks from the start to ensure standards are globally usable, while balancing trust and autonomy.

Key takeaways

- Genomic standards and consent tools must account for cultural, ethical, and infrastructural diversity, particularly in low to middle income countries (LMICs), to ensure global applicability. Current frameworks often reflect high-income contexts, creating adaptation burdens and potential inequities.
- African research emphasises collective decision-making, traditional authority, and relational frameworks such as Ubuntu and Utu, which must be integrated alongside individual consent to ensure ethical and culturally appropriate practices.
- Language, comprehension, and practicality are critical. Technical terminology, Western metaphors, and assumptions about literacy or infrastructure can hinder trust.
- Training, translation, iterative feedback, and support for local GA4GH nodes are necessary to make standards usable and sustainable across diverse research environments.
- REWS should be embedded across all standards and technical Work Streams, ensuring that ethical, legal, and social considerations inform every aspect of genomic data sharing.

Priority action items

- The group will expand its consultations with more countries in the continent to account for regional differences, and eventually work towards developing draft adapted clauses.

GIF TRE Open Suite: towards GA4GH-compliant secure and privacy-preserving computations

Recording | Agenda | Slides

This session focused on how to automate Trusted Research Environment (TRE) deployment and trust management within the GA4GH ecosystem, moving the discussion towards solutions for enabling federated analytics. Three discussion tables addressed the following topics: 1) policies and governance; 2) trust and privacy enhancing technologies (PETs); and 3) GA4GH product integration and architecture.

Discussions highlighted the complexity of compliance, as multiple legal and geographic frameworks apply, making extensive automation difficult. Researchers typically trust a system based on legal contracts and service reliability rather than direct technical evaluation. The group plans to study existing TRE implementations to find areas for GA4GH standardisation and consider lessons from other communities dealing with sensitive data. Technical solutions discussed included the need for robust federated authentication (e.g. via OIDC, Passports), the role of technologies such as Confidential Computing (e.g. Trusted Execution Environments [TEEs]), and the potential for GA4GH services to present trust assertions (e.g. legal entity identity, geolocation, workload identity, ISO 27001 claim) and respective evidences (e.g. X.509, TEE Report, ISO certificate) to aid trust decisions.

Key takeaways

- Extensive automation of TRE compliance is difficult because multiple, often conflicting, legal and geographic compliance frameworks must be satisfied simultaneously.
- Researchers primarily decide to trust a TRE based on legal contracts and the demonstrated reliability/availability of the service, rather than directly evaluating the underlying security technology.
- Technical efforts must focus on providing robust federated authentication, using standards such as OIDC and Passports, and leveraging specific Privacy-Enhancing Technologies (PETs) such as differential privacy, homomorphic encryption, and confidential computing to secure the environment.

Priority action items

- Research and document how current Trusted Research Environments (TREs) are implemented and operated to identify common functional and security features that are suitable for GA4GH standardisation and automation efforts.
- Prioritise the technical development and promotion of robust federated authentication and authorisation mechanisms, leveraging existing GA4GH standards such as Passports and industry standards such as OIDC, to securely manage user access across multiple TREs.
- Focus technical efforts on exploring and integrating advanced security technologies, particularly Confidential Computing, as a foundational element for securing the processing environment (TRE).
- Create a standardised mechanism for GA4GH services to generate and present formal, machine-readable "trust assertions" (e.g. verifiable evidence of compliance such as ISO 27001) to aid in the digital verification of a service's trustworthiness.
- Proactively collaborate with and learn from other communities (e.g. in finance or other sensitive data domains) and standards organisations to adopt best practices and ensure that GA4GH solutions are technically sound and broadly applicable.

Discovery Work Stream in 2026

Recording | Agenda | Slides

The session discussed current and future projects of the GA4GH Discovery Work Stream. The Federated Variant Level Matching Project provided an update, sharing its move towards using the Beacon standard for connecting VCF and phenotype databases, adopting Auth0 for security over tokens/passports, and addressing ongoing challenges in variant representation, such as indels and structural variants, and phenotype exchange. The project is planning to deploy new features such as direct contact, phenotypic features, and affected status, which are supported by Beacon's current API.

Updates on the Discovery Work Stream products were also shared:

- Beacon v3.0 plans were detailed, including adding aggregated granularity, supporting some logical operators, new enumerations in filters, and better support for genomic representation standards (such as the Variation Representation Specification [VRS]).
- The Experiments Metadata checklist is nearing approval and aims to standardise metadata for omics data discovery and reuse, with an initial focus on core sequencing properties before expanding to assays such as transcriptomics and targeted sequencing.

- Data Connect is seeking feedback on a potential v2.0 release to enhance features such as GA4GH Passports support and improve schema consistency, with a call for community members to help with its development.
- Service Info and Service Registry will move into maintenance mode and transition to the purview of the Technical Alignment Sub-committee (TASC).

Key takeaways

- For the Variant Level Matching Project, the use of simple tokens was deemed insufficient for exchanging sensitive VCF and phenotype data. The project has adopted Auth0 for authentication, finding it a more practical solution than full GA4GH Passports for their current needs.
- Plans for Beacon v3.0 and Data Connect v2.0 releases are under discussion, including enhancements and integration with other GA4GH standards.
- The Experiment Metadata checklist is in its final stages, having passed the Product Review Committee and pending a final vote.

Priority action items

- Proceed with the deployment of the next set of features for the Federated Variant Level Matching Project, leveraging Beacon's API capabilities and the newly adopted Auth0 security model.
- Initiate a focused effort within the Variant Level Matching Project and relevant Work Streams (such as the Genomic Knowledge Standards [GKS] Work Stream) to define the standardised nomenclature for exchanging indels and structural variants across the network, which is a required step before deployment.
- Push the Experiment Metadata checklist through its final step — the Product Steering Committee vote at the end of October — to officially establish the v1.0 minimum standard for omics data discovery.

Get involved!

Interested in enabling broad and more effective genomic and health data discovery? Consider contributing to the development of Beacon v3.0 and Data Connect v2.0! For more information, please contact Beatrice Amos.

Bridging location and phenotype

Recording | Agenda | Slides

The session focused on bridging geolocation data with health phenotypes to advance exposomics research, emphasising how environmental exposures (the exposome) influence biology and account for 70% of premature mortality risk compared to under 2% from genetics. Key discussions included challenges in recording precise, longitudinal geolocation data for linkage with environmental variables, ensuring shareable geolocation IDs that balance privacy and scientific utility, and enabling federated data analysis to compute exposures securely.

Presentations highlighted the Secure Anonymised Information Linkage (SAIL) Databank's anonymised population-level data linkage in Wales, using fields such as the Anonymous Linking Field (ALF) and Residential Anonymous Linking Field (RALF) for dynamic exposure modeling, and the newly funded Cohort and Environmental Data Integration for Epidemiological Innovation (CONNECT) project's efforts to create hyper-resolution environmental maps (e.g. air pollution) linked to large UK cohorts such as the UK Biobank and NHS Digital.

The session also covered harmonisation of geolocation standards, geocoding methods, privacy strategies such as fuzzing and synthetic data, and breakout discussions on generating standards for secure environmental-genomic linkages.

Key takeaways

- Precise, longitudinal geolocation data is essential for accurate exposome research, but requires balancing granularity with privacy to prevent re-identification.
- Projects such as SAIL Databank and CONNECT demonstrate scalable approaches to linking environmental exposures with health data, highlighting the need for federated systems and hyper-resolution mapping.
- Harmonisation across regions and cohorts is challenging due to varying data collection methods, but synthetic data and portable linkage frameworks offer promising solutions for global applicability.

Priority action items

- Define initial conceptual models or frameworks for linking geolocation identifiers with environmental exposure data.
- Outline a road map of next steps based on discussions to develop and implement standards for geolocation identifiers and environmental data linkage across cohorts.
- Develop standards for harmonising geolocation data and geocoding across cohorts to enable scalable exposome integration.

- Create standards for secure, federated linking of environmental data with genomics and clinical cohorts, building on efforts such as those of GA4GH.
- Solicit suggestions from interested cohorts and projects for implementing these standards.
- Disseminate environmental data from projects such as CONNECT via public repositories, with tutorials and pipelines for user access.

Agentic science: enabling AI agents to reason with biomedical data via GA4GH standards

Recording | Agenda | Slides

The session centered on establishing a new foundational Artificial Intelligence (AI) Work Stream to address the rapid integration of AI in genomics and health data. Its potential for accelerating research, data analysis, and discovery must also be balanced with responsible, ethical, and equitable use.

Presenters gave lightning talks, including: 1) practical AI applications, such as variant annotation using generative pre-trained transformer (GPT) models with retrieval-augmented generation (RAG) to eliminate hallucinations; 2) the Narrative Project for privacy-preserving data descriptions; 3) agentic scientists leveraging GA4GH standards for hypothesis generation and workflow execution; and 4) BRCA variant analysis.

A collaborative Figma board brainstorm explored AI's possible impacts over the next three years, including autonomous agents supporting scientific discovery as well as increased accessibility for tools and software, along with risks such as data silos and biases. The group then explored GA4GH's role in developing standards, governance, and cross-Work-Stream partnerships to enable responsible AI use.

Discussions highlighted challenges such as data quality control, hallucinations, synthetic data generation, metadata standards for provenance, human oversight, and the need for industry-academia collaboration to make genomics data AI-ready.

Key takeaways

- AI is transforming genomics from descriptive to predictive science, requiring urgent GA4GH standards for data readiness, ethical use, and governance to handle its rapid pace and risks.
- Success demands industry-academia partnerships, cross-Work-Stream alignment, and emphasis on ethics and equity to democratise access and ensure responsible AI deployment.
- While embracing agentic science and hallucinations for discovery, prioritising quality control via synthetic data and RAG, human oversight, and metadata provenance will be key to maintain trust and accuracy.

Priority action items

- Launch the GA4GH AI Work Stream by establishing a charter, leadership, principles, and road map.
- Organise an open strategy session to categorise and prioritise the topics discussed on the Figma board, and determine next steps in collaboration with other GA4GH Work Streams.
- Conduct a landscaping session with other GA4GH Work Streams to define territories, enable AI integration, and address painful interoperability issues such as shared data models.
- Explore partnerships with external groups (e.g. ML Commons) for metadata standards and cross-pollination on consent and provenance.

Get involved!

Interested in co-leading the GA4GH AI Work Stream? Learn more and apply to the opportunity.

Connect Day Two

Literature derived variant evidence: toward GA4GH compatible representation

Recording | Agenda | Slides

The meeting explored how literature-derived variant evidence can be better integrated with GA4GH standards to improve discoverability, representation, provenance, and supporting evidence. This session focused on advancing variant representation and extraction from scientific literature to bridge observational and interpretative data in genomics.

The SIB Text Mining Group presented tools such as LitVar and Variomes for searching variants in PubMed and PMC, emphasising the need for database-independent normalisation using HGVS and GA4GH standards such as the Variation Representation Specification (VRS) and Categorical Variation Representation Specification (Cat-VRS) to handle underspecified or categorical variants, improve interoperability, and potentially integrate with Beacon for unified access.

Discussions highlighted challenges such as non-single nucleotide polymorphisms (SNP) variants, multilingual support (e.g. Japanese case reports), confidence scoring, and evidence extraction using language models.

Somatic copy number variants (CNVs) in cancer were also discussed. A key point emphasised was the "pathogenicity annotation gap" due to under-annotation, and that data-driven approaches with Cat-VRS could help integrate functional evidence and observational data from resources such as Progenetix.

Attendees, including experts from the GA4GH Genomic Knowledge Standards (GKS) Work Stream, explored collaborations for stepwise implementation, precise vs. categorical representations, and hybrid evidence aggregation.

Key takeaways

- Literature-based variant search tools like Variomes provide higher coverage and recall for new or understudied variants compared to curated databases, serving as a vital complement for curation and evidence tracking.
- GA4GH standards such as VRS, Cat-VRS, and VRSATILE enable precise representation of imprecise and underspecified variants from literature, facilitating interoperability, discovery, and integration with knowledge-bases and tools such as CIViC or Beacon.
- Addressing gaps in CNV annotations for cancer requires hybridising observational data with literature evidence, using data-driven frameworks to identify drivers and improve pathogenicity assessments.

Priority action items

- Develop and test a prototype for representing literature-extracted variants in JSON files using Cat-VRS, VRS, and the Variant Annotation Specification, focusing on a small literature subset for benchmarking and validation.
- Explore API development for variant annotation in non-English languages, potentially incorporating machine translation and multilingual models.
- Coordinate with teams (e.g. Microsoft/Evidence Aggregator collaboration) to incorporate Variomes into broader literature search assessments and manual validation efforts.
- Define a minimal Beacon dataset schema for literature variants.

Get involved!

Join the GKS Think Tank channel to collaborate on use cases for variant extraction, normalisation, and representation (e.g. starting with SNVs/small indels and abstracts/full-text papers). Reach out to Work Stream Manager Beatrice Amos for more information.

Influencing Electronic Patient Record (EPR) and Laboratory Information Management Systems (LIMS) suppliers to adopt FHIR standards and native interoperability (part 1)

Recording | Agenda | Slides

The NHS Genomic Medicine Service (GMS) is advancing a digital strategy for interoperability, aiming for genomics to influence 50% of healthcare by 2035. The plan centers on developing a Unified Genomic Record (UGR), a core component of the national Single Patient Record.

A centralised order management broker is being built to connect over 240 NHS trusts and their disparate EPR (Epic, Cerner) and LIMS systems. The main challenge is compelling global suppliers to adopt shared standards, specifically from HL7 FHIR and GA4GH, to simplify system integration, enable responsible cross-border data sharing, and accelerate biomedical research. The effort seeks collective action to influence suppliers rather than rely on fragmented institutional negotiations.

Key takeaways

- The NHS faces a highly complex and fragmented system, with 189 secondary care trusts, over 6,400 GP practices, and more than 120 unique EPR/LIMS suppliers often running bespoke, disconnected instances. The lack of standardised data exchange across these systems is the main barrier to delivering integrated genomic care.
- The entire digital strategy is fundamentally dependent on standardising data exchange using HL7 FHIR (Fast Healthcare Interoperability Resources) and GA4GH standards. These standards are necessary for ensuring responsible data sharing, machine readability, and enabling advanced features such as pharmacogenomic guidance and clinical trials matching.

Priority action items

- Work internationally and nationally to create a single, unified voice to persuade global EPR and LIMS suppliers to proactively adopt HL7 FHIR and GA4GH interoperability standards. This is necessary to overcome the current landscape of bespoke, non-communicating systems without relying on multi-year legislative mandates.
- Build the central, events-based broker system by the target date (anticipated April 2026). This system acts as the "plumbing" to facilitate requests and returns between the many fragmented local hospital systems and the genomic laboratory hubs, which is essential for placing and tracking genomic test orders.
- Finalise the development of the Digitised Test Directory (DGTS) to provide a machine-readable, structured, searchable, and version-controlled reference dataset. This foundational common data model will underpin the entire Order Management system and ensure accurate communication of available genomic tests across the NHS.

Influencing EPR and LIMS suppliers to adopt FHIR standards and native interoperability (part 2)

Recording | Agenda | Slides

The NHS is advancing its digital strategy by developing a Single Patient Record (SPR), with the Unified Genomic Record (UGR) serving as a pathfinder. This system, supported by FHIR and other common standards, spans direct care, research, and population health across England's 57 million people. Four major projects underpin this: the Digitised Test Directory (an interactive catalogue), Digitised Order Management (a central broker), the unified genomic record (UGR), and a move toward common standards, enforced eventually by the Data Use and Access Bill. Architecturally, the system leverages a federated object store (using the Data Repository Service [DRS] and htsget) and a single FHIR store for metadata, all accessible via the NHS API Management platform.

Key takeaways

- To overcome vendor fragmentation, the Data Use and Access Bill will eventually force vendors to adopt NHS standards. However, the immediate strategy is to encourage vendor adoption now, before the legislation is operationalised, to accelerate the integration of disparate EPR and LIMS systems across the country.
- The NHS is committed to developing a SPR for all patient interactions. Crucially, the UGR — which integrates patient, clinical, and diagnostic data — is about two years ahead in development and is positioned to act as one of the key pathfinders for the entire national Single Patient Record architecture.
- To manage large genomic files such as VCF and BAM files, a federated object store has been implemented using GA4GH standards such as DRS and htsget. This approach allows external systems to find and securely stream genomic data through a central API platform, using a FHIR DocumentReference to link rich metadata to the large DRS-managed files.

Priority action items

- Continue the development and deployment of the Digitised Order Management and Unified Genomic Record systems using FHIR and GA4GH standards
- Operationalise the UGR as the SPR Pathfinder, ensuring it supports a population-level system covering direct care, research, management, and public health.
- Establish fine-grained, purpose-based data access control to enable safe research.

GA4GH in the real world

Recording | Agenda | Slides

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 the International Precision Child Health Partnership (IPCHiP) demonstrated cross-border data sharing, while projects such as the Federated Imputation Platform and Pacific Analytics Toolkit illustrated scalable, standards-based interoperability. Discussions emphasised the development of a minimal starter kit for practical deployments, advancing federated machine learning through the AI Work Stream, and the importance of regional engagement. Next steps include improving integration across GA4GH APIs, piloting the starter kit, expanding federated compute initiatives, and documenting community experiences to guide broader adoption.

Key takeaways:

- 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.
- 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.
- 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, authorisation, data models, APIs, storage, logging, and auditing — all aligned with GA4GH standards and informed by defined scientific use cases.
- Several initiatives underscored 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.

- 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.

Priority action items

- Improve interoperability between the Data Repository Service (DRS), Workflow Execution Service (WES), and other GA4GH APIs to support federated analysis platforms.
- Advance the minimal starter kit by defining its scope and core components, then testing a prototype deployment using public data.
- Expand federated compute and machine learning initiatives, ensuring strong links with the Data Security and Regulatory & Ethics Work Streams
- Promote community adoption and collaboration by organising regional workshops to raise awareness and support small-scale pilot implementations.
- Document and share learnings by publishing case studies and sharing implementation experiences.

Capturing cancer genomics data attributes in Beacon v2.0

Recording | Agenda | Slides

The meeting explored how Beacon v2.0 can be extended to better support cancer genomics use cases, including rich clinical metadata queries, matched samples, complex variants, and clinical intervention timelines. It focused on the development and implementation of OncoBeacons, a cancer-specific adaptation of the GA4GH Beacon standard for genomic and clinical data discovery and sharing.

Participants from Dutch cancer centers (Hartwig Medical Foundation and Princess Máxima Center) shared their journey in building Beacons, including data modelling challenges, ontology usage, privacy considerations for somatic vs. germline data, integration with models such as OMOP, and the balance between flexibility and strictness in standards to ensure utility.

Discussions also covered historical context of Beacon evolution, implementation differences across platforms, query complexities, and the need for community-driven harmonisation to enable federated queries and biomarker discovery.

Emphasis was placed on starting small, addressing privacy risks, and differentiating between data discovery and deeper analytics.

Key takeaways

- Beacon implementations for cancer must prioritise utility by harmonising data placement, ontologies, and query mechanisms — balancing flexibility for local needs with strictness for interoperability.
- Privacy and consent are critical, with aggregate data being safest for public sharing, but somatic variants offering lower re-identification risks; start with conservative access levels and build confidence gradually.
- Community involvement is essential — start small with implementations, report issues to developers, and focus on use cases to drive improvements, as perfect standards evolve through practical testing and feedback.

Priority action items

- Provide feedback from the session to the GA4GH Beacon and Cancer Communities to inform ongoing developments.
- Define minimal standards for cancer data in Beacons, including tiers for discovery and analytics, through community collaboration (e.g. via GitHub issues and upcoming sessions such as the GA4GH Cancer Community meetings).
- Launch operational OncoBeacons by the end of the year at the Dutch centers, with transparent mappings and testing of federated queries.
- Develop sanity checks and simulated datasets for validating Beacon filters and implementations.
- Engage big players (e.g. large US initiatives) and advocate for funding to support tool development and standardisation.
- Develop a community sourced list of use cases
- Identify cancer centres working on implementation and develop basic guidelines for how they can get started with Beacon.

FAIR in action using GA4GH products: hands-on FAIRification from a fresh start to the FAIRest of them all

Recording | Agenda | Slides

The session focused on applying FAIR (Findable, Accessible, Interoperable, Reusable) principles to genomic data, digital assets, and pipelines. A core theme was moving beyond simple data sharing to making data machine-actionable, which requires using widely recognised standards from catalogues such as the Fairsharing registry.

The discussion introduced the G4GH Task Execution Service (TES) as an example for applying FAIR principles across specifications, software implementations, and service instances, including consideration for AI-ready data. A significant takeaway from the group discussion was that many GA4GH resources currently face a high barrier to entry and need improved onboarding and accessibility to fully achieve the "Findable" (F) principle.

Key takeaways

- FAIR implementation should be incremental and context-specific; starting with widely adopted repositories and standards can significantly advance FAIR compliance without excessive cost.
- FAIR principles apply not only to data but also to software, workflows, and infrastructure, each requiring tailored approaches to metadata, interoperability, and reusability.
- Enhancing the findability and usability of GA4GH products is critical for broader adoption, requiring concerted efforts in communication, documentation, and community support.

Priority action items

- Continue discussions and follow up in early November to review progress and insights from the breakout groups.
- Improve onboarding materials and guidance to lower barriers for new users of GA4GH standards and resources.
- Explore creating implementation profiles or clearer FAIR adoption pathways tailored to GA4GH products.

Variant representation alignment: bridging VRS, VCF, HGVS, and SPDI

Recording | Agenda | Slides

This session focused on the challenges in variant representation and interoperability. Different identifier systems were discussed, and widespread issues with ambiguity, inconsistency, and misalignment across formats were highlighted. Specific challenges include:

- VCF: inconsistent implementations, non-standard separators, and ambiguous info fields.
- HGVS and SPDI: transcript ambiguities, outdated documentation, and coordinate confusion.
- RSIDs: location-based rather than variant-specific, limiting precision.
- Structural variants: fuzzy breakpoints and aggregation difficulties.
- Cross-cutting issues: liftover ambiguities, one-to-many mappings, and scaling limitations of text-based formats.

Breakout groups emphasised inertia in adopting updates, day-to-day pain points for researchers and bioinformaticians, and the need for standardisation, improved computational identifiers, and scalable formats to support complex variants.

Key takeaways

- A unified, precise, and computable variant representation standard is critically needed to ensure interoperability and accurate cross-platform data exchange.
- Current systems and legacy formats introduce ambiguities and inconsistencies that hinder accurate variant identification and integration.
- Modern use cases, such as structural, complex, and pangenomic variants, require scalable, structured formats that support ambiguity handling, multi-assay data, and flexible aggregation.

Priority action items

- Consolidate feedback from breakout sessions to inform future GA4GH standardisation efforts.
- Encourage community alignment on a standardised, unambiguous variant representation format for display and search.
- Improve documentation and tooling support for existing specifications (e.g. VCF, HGVS) to reduce implementation variability.

Passports that scale: lightweight tokens and federated trust

Recording | Agenda | Slides

This session reviewed the Passports structure and standard Visa types, including the specific National Institutes of Health (NIH) Researcher Auth Service (RAS) implementation. It also explored the challenges related to large Passports, as well as the potential for improved computable policy representation through the proposed Data Passports concept.

Key takeaways

- Passports can support more than role-based access control, despite the challenges with its large size and the cumbersome process of validating permissions.
- Task-specific tokens may be a useful approach for improving management and handling metadata, possibly integrating a refresh mechanism.
- Data Passports provides a computable approach to represent data access policies and manage governance, in light of fragmentation challenges, but requires further exploration within the Data Passports Study Group.

Priority action items

- Continue discussions on Data Passports and task-specific tokens at upcoming Data Security Work Stream meetings.

Get involved!

- Participate in the Data Passports Study Group to help shape these conversations! For more information or to join the mailing list, please contact Jessica Seegobin.

Where is global? Geolocation in genomics

Recording | Agenda | Slides

The session focused on integrating geolocation data into GA4GH standards, emphasising its role in service discovery, data interpretation, and global collaboration.

Historical context was provided, highlighting long-standing discussions on geolocation formats, privacy concerns such as obfuscation and granularity, and a proposed GeoJSON-based schema for reusable objects in standards such as Service Info.

A practical use case from the Cloud Work Stream was presented, showcasing a prototype map of 23 Data Repository Service (DRS) servers, and stressed the need for better visibility of GA4GH implementations through an updated registry.

Discussions covered extending geolocation to non-human/environmental data, analytical applications such as querying samples by proximity in Beacon, and feedback on schema elements such as precision, ISO codes, and time-series tracking. Compatibility with other standards such as Darwin Core, automation for registry updates, and broader use cases for mapping growth in GA4GH adoption were also explored.

Key takeaways

- Geolocation enhances GA4GH's global visibility and utility, enabling service discovery, regulatory compliance, and data analysis (e.g. population background inference or environmental sampling), but requires careful handling of privacy through obfuscation and granularity.
- A standardised GeoJSON subset in Service Info offers a flexible, reusable format for implementations, with potential extensions for non-human data and time-linked trajectories.
- The registry should evolve into a dynamic, automated tool for mapping and tracking GA4GH growth, starting with known services to demonstrate impact and identify gaps in global adoption.

Priority action items

- Update the GA4GH registry to incorporate geolocation data using a standardised GeoJSON-based schema, with input from the GA4GH Tech Team.
- Develop a user-friendly form on the GA4GH website for registering implementations, and transition from temporary tools such as Google Forms to automated updates via Service Info endpoints.
- Bootstrap the registry by compiling known implementations from Work Streams and the website, and evolve it into a time-series database to track changes over time (e.g. growth in servers and data volume).

- Review and refine the geolocation schema for precision/obfuscation handling, ISO code integration, and compatibility with external standards such as Darwin Core, with feedback from implementers.
- Reach out to known projects (e.g. Beacon and Phenopackets) to register services and aim for a comprehensive map by the 2026 Plenary meeting in Singapore.

Leveraging GA4GH's tools and standards to balance pathogen data sovereignty with open access

Recording | Agenda | Slides

This session explored the challenges and opportunities for equitable pathogen data sharing, while addressing challenges in data sovereignty and open access. Discussions highlighted the importance of supporting local governance, protecting data contributors, and ensuring transparency in tracking data use in light of barriers to data sharing, particularly in low to middle income country (LMIC) contexts.

Key takeaways

- There is a pressing need for a harmonised, equitable pathogen data sharing framework that balances sovereignty, access, and trust.
- Track-and-trace and data provenance tools can serve as enabling technologies for equitable and transparent data use.
- Political, infrastructural, and publication-related barriers remain significant obstacles for LMIC participation.
- Avoiding data contributors and depositors from experiencing reputational or political harm is essential for sustainable collaboration.
- Capacity-building and regional oversight mechanisms are vital to empower countries with limited legal frameworks.

Priority action items

- Comprehensive stakeholder analysis is needed to identify community contributors, with extensive consultation efforts needed before work can proceed further.
- Policy needs to reflect the variation in scientific and policy use cases, the unique epidemiology of pathogens, and the political context that influences data sharing.
- The team will host another meeting in three to four months to review progress on the draft policy recommendations, track-and-trace prototype framework, and outcomes from stakeholder consultations.

Next steps on the use of verified credentials (models and protocols) for improving passports

Recording | Agenda | Slides

The meeting discussed the Verifiable Credentials model, and potentially developing a new specification. The session reiterated the potential to leverage existing technologies, including mapping GA4GH Passports and Visas into established infrastructures using the OID4VC model. There was also emphasis on further liaising with standards groups, such as ISO, while exploring the genomic-related needs for secure data access.

Key takeaways

- Verifiable Credentials can allow for richer semantics and the use of automatic reasoning with the reference of ontologies.
- Alignment with OID4VC's specifications, including OID4VCI, OID4VP, SIOPv2, will be helpful for avoiding redundancy.
- Gaps in trusted research environments (TRE)-related specifications may be addressed by forming new Visa types or using the AARC Blueprint Architecture.

Priority action items

- Consider forming a GA4GH Study Group to explore Verifiable Credentials further.
- Differentiate work with ISO by formats and protocols, allowing ISO to focus on the mapping of APIs.
- Engage with other groups facilitating authentication and authorisation to ensure representation of specific genomics needs are incorporated.
- Liaise with the OpenID Digital Credentials Protocols (DCP) and the W3C Verifiable Credentials working groups.
- Share developments at the Ad Hoc Group 2 Verifiable Credentials for Genomics Data segment during the ISO/TC 215/SC1 meeting on 21 October.

Scaling the Federated Variant-Level Matching Network

Recording | Agenda | Slides

The meeting focused on advancing Variant Level Matching (VLM) to address challenges in rare disease genetics, such as high rates of variants of uncertain significance (VUS) and the rarity of variants and genes, emphasising the need for global data sharing to enable gene discovery and variant interpretation.

An overview of the VLM system was provided, highlighting implementations in databases such as Variant Matcher, Franklin, Seqr, and Geno2MP, connected via the Beacon v2.0 API and Auth0 for authentication, with examples of successful gene discoveries and variant reclassifications.

The Variation Representation Specification (VRS) and the AnyVar tool were discussed to standardise variant representation, facilitate liftover between genome builds, and create off-the-shelf solutions for ad hoc VLM networks, including loading GREGoR Consortium data.

A pilot with clinical labs in Brazil was shared, demonstrating how VLM aids in reclassifying VUS, while addressing barriers such as proprietary data concerns, regulatory issues, and the need for flexible sharing levels.

Discussions covered technical challenges such as indel representation, phenotype granularity, duplicate data handling, and creating Beacon profiles, with a focus on scaling the network inclusively across research, clinical, and patient-led groups.

Key takeaways

- Data sharing via VLM significantly reduces VUS rates by enabling variant reclassification, but requires balancing openness with authentication to prevent misuse, such as by AI companies.
- Standardisation (e.g. via VRS and Beacon profiles) and flexible implementations are crucial for scaling VLM networks inclusively, especially for under-resourced or regulated entities such as clinical labs in low-income countries.
- While research databases are connecting successfully, engaging commercial/clinical labs demands proving tangible benefits (e.g. time savings and better reports) and addressing burdens such as email overload and proprietary concerns through phased, reciprocal sharing.

Priority action items

- Develop and finalise technical documentation for joining the VLM network, including a guide on endpoints, expectations, and a Beacon v2.0 profile to lower barriers for new nodes.
- Implement next-phase features such as returning direct contact emails, detailed phenotypes (e.g. HPO terms and ICD/OMIM codes), affected status, and variant type queries per gene and region, while addressing indel and structural variant representation using tools such as the Variation Representation Specification (VRS).

- Fix identified bugs (e.g. liftover issues in Seqr) and integrate more nodes to expand the network.
- Explore genomic tags or decentralised persistent identifiers to handle duplicate datasets and identify matching individuals without excessive emails.
- Continue pilots with clinical and patient-led groups to demonstrate value, refine consent and regulatory approaches, and ensure flexibility for varying sharing capabilities.

An analytics dashboard for everything GA4GH

Recording | Agenda | Slides

The meeting focused on introducing a new analytics dashboard prototype developed by the GA4GH tech team to track and analyse the impact of GA4GH standards and products across various platforms. The team demonstrated features for tracking GitHub engagement, PubMed article analysis, and PyPi data, while discussing plans to expand the dashboard to include more organisations and harmonise data from different sources. The team also presented the potential to measure implementation of the policies and frameworks produced and shared by the Regulatory & Ethics Work Stream (REWS) through the use of natural language processing and supervised learning. The presentations highlighted the importance of measuring GA4GH's impact through comprehensive data collection and visualisation tools, with discussions around potential improvements and integration of these tools into the main website.

Key takeaways

- The community was engaged and interested in knowing more about the use cases for the dashboard.
- For any data used in drawing the analytics, validation is necessary and would help build trust in using the raw data retrieved from the data sources.
- Although notebooks are easy to use, a web-based dashboard would serve the community better. Functionality to customise the graphs on the data was also a requested feature.
- The GitHub repository of this project should also be opened up for the community to use, review, and raise issues.
- Any use of artificial intelligence and machine learning in analysing the data should have oversight to avoid hallucinations or inconsistencies.

Priority action items

- Ingest and validate the data from these three data sources — GitHub, PubMed, and PyPI.
- Enhance the user experience and accessibility of the notebooks, making it easier to understand the data and resulting graphs.

- Draft and publish the metadata used in the data validation and visualisation.
- Develop and publish a web-based dashboard.
- Open the GitHub repository to the community and public soon.

Connect Day Three

The next step toward clinical application of GA4GH products: a sequel (part one)

Recording | Agenda | Slides

This first part of the clinical applications session focused on advancing clinical adoption of GA4GH standards by addressing barriers to implementation, building on insights from the April 2025 Connect meeting in Boston. After that session, we synthesised 30 pages of community notes into four categories: 1) outreach and engagement; 2) implementation and adoption support; 3) technical development and interoperability; and 4) governance, regulatory, and policy. During this session, attendees reviewed fifteen derived action items, emphasising the need to demonstrate value through use cases, foster partnerships, and integrate with existing clinical systems. The goal of both parts of this Connect experience was to take time for the community to prioritise and choose the first set of items to help bring GA4GH products into clinical use.

The next step toward clinical application of GA4GH products: a sequel (part two)

Recording | Agenda | Slides

Following a prioritisation exercise and a community vote, participants extracted three priority topics, then broke out into breakout groups to generate tangible recommendations for GA4GH leadership to accelerate the adoption of standards into real-world healthcare applications. The first discussion focused on developing and showcasing tangible clinical use cases and vignettes; the second one centered on encouraging and supporting implementation programmes and incremental adoption; and the third discussion zeroed in on increasing outreach and presence at clinical conferences and in health IT communities. Report-outs from the groups emphasised developing regional hubs to customise outreach, engaging patient support organisations, creating vignettes, and establishing a clinically-oriented use case catalogue. The groups were also asked to define how GA4GH could implement their suggestions.

Key takeaways from both sessions

- GA4GH aims to bridge research and clinical application via use cases and partnerships.
- There was consensus across all discussions on the need to demonstrate the value of GA4GH standards through real, tangible clinical success stories and use cases.
- There was a repeated emphasis on moving beyond a centralised approach to reach diverse global stakeholders by establishing and empowering regional and local support structures.
- Community input drives prioritisation — targeting vendors and engaging local governance emerged as critical points of connection and engagement.
- Actionable outcomes should focus on supporting implementations, soliciting feedback, and working towards interoperability.
- Finding tactics to lower the barriers to adoption and integration is key to facilitating adoption.

Priority action items

- Develop and showcase tangible, narrative-style clinical use cases and vignettes to provide clinical audiences with examples of how GA4GH standards can facilitate diagnostics and data reuse. Stories should demonstrate how adopting standards makes clinical work faster, cheaper, or easier. This was highlighted as the most fundamental first step to getting clinical buy-in.
- Cultivate regional hubs and targeted communications that are audience-specific and locally relevant.
- Lower the barriers to adoption by providing resources, tools, and mentorship; improving the GA4GH website to have clear entry points for different audiences; and providing guidance on where and how to ask technical questions.
- Develop targeted communication materials for institutional leadership that specifically address the value of GA4GH products and their return on investment.
- Engage stakeholders throughout the GA4GH product development process to ensure standards are inherently useful and accessible to clinical practitioners.
- To lower the barrier to adoption, encourage and support small-scale pilot programmes and incremental adoption.
- Drive efforts towards developing APIs and translators for clinical standards compatibility, ensuring seamless integration with existing clinical infrastructure and industry standards such as HL7, FHIR.

Securing Beacon networks with Passports and trust federation

Recording | Agenda | Slides

This session explored the model of the Beacon Network, as well as the role of Passports and Visas. The meeting also looked at using Verifiable Credentials and the Decentralised Discovery Network with Authorisation as potential solutions, while acknowledging the challenges around interoperability, storage, and management.

Key takeaways

- There is a further need to explore a system where the Beacon Network can aggregate and validate Visas to manage user access while preserving trust across different sources.
- Collective community engagement may be useful to ensure the Beacon Network supports flexibility and can evolve without limiting scalability.
- Verifiable Credentials may offer a decentralised model for Beacon Network authentication, though challenges of interoperability, storage, and trust anchor distribution persist.
- The Decentralised Discovery Network with Authorisation may provide an alternative approach to federated authentication, where each node manages its own trust and permissions.

Priority action items

- Explore the establishment of federated trust registries and profiles for biomedical research.
- Continue more community focused discussions on design and deployment.

The Cloud Work Stream road map 2026 and beyond

Recording | Agenda | Slides

The Cloud Work Stream has evolved from developing individual cloud abstraction APIs (e.g. Tool Registry Service [TRS], Workflow Execution Service [WES], Task Execution Service [TES], and Data Repository Service [DRS]) into advancing an integrated ecosystem supporting findable, accessible, interoperable, and reusable (FAIR) data discovery and democratised compute. In 2025, key achievements include the expansion of the DRS network to 23 servers hosting 72 petabytes of data, the launch of the Federated Compute Initiative, and updates to the Cloud Work Stream's standards to enhance interoperability and cloud integration. Emerging efforts such as Mesh cards and the Portable Format for Biomedical (PFB) data aim to strengthen data and metadata federation across systems. Next steps include publishing DRS and WES papers, defining and developing new specifications, hosting open house and hackathon events, and fostering cross-Work-Stream collaboration on access control, ontology, and cost-sharing models.

Key takeaways

- Continue enhancements across all Cloud Work Stream APIs while also working towards an integrated cloud ecosystem
- New tools such as Mesh cards and PFB are enabling more flexible and interoperable data exchanges across federated systems.
- Collaboration, community engagement, and cross-Work-Stream integration will be key to further strengthening integration across the GA4GH ecosystem.

Priority action items

- Advance new initiatives, including:
    - developing and circulating PFB specifications for community review;
    - defining the structure and goals for federated compute;
    - investigating Mesh cards.
- Launch open house sessions and “Ask Me Anything” (AMA) events in collaboration with the GA4GH Implementation Forum.
- Further develop current Cloud Work Stream products, including:
    - organising a WES “write-a-thon” focused on federated compute and interoperability;
    - extending TRS to support AI models and multiple workflow languages;
    - developing frameworks for managing compute and storage usage across APIs.
- For long-term term exploration, the Work Stream aims to:
    - evaluate cost-sharing and incentivised data contribution models;
    - refine self-assessment tools for cloud repositories;
    - explore multi-use ontologies for fine-grained data access control.

Solving Beacon queries

Recording | Agenda | Slides

This session explained Beacon v2.0's framework — which includes a query and response structure — separate from its clinical genomics model, using REST APIs for entity queries such as range or individual-based genomic variations. Discussions focused on filters (e.g. ontology, alphanumeric, custom) combined via `AND`, cross-queries intersecting entities, real-world cancer examples highlighting limitations (e.g. unspecific results, false positives), and proposals for specificity via model paths, grouping (`OR` expressions), excluded flags, and request profiles for standardised variation queries.

Future alignments with VRS v2.0 were outlined, alongside challenges in backend resolution, heterogeneity, and shifting toward domain-specific standards over universal ones, with audience input on GraphQL, record access, and metadata catalogues.

Key takeaways

- Beacon emphasises discovery simplicity but needs enhancements such as paths and grouping to handle complex queries and reduce ambiguities without custom parsers.
- Genomic queries evolve via request profiles (non-breaking) then the Variation Representation Specification (VRS) v2.0 for typed, declarative support, enabling Beacons to specify capabilities.
- Flexibility supports diverse uses but requires harmonisation, bridging IT/researcher gaps, and community-driven restrictions for networks vs. aggregators.

Priority action items

- Submit feedback on filter grouping options and query enhancements via GitHub pull requests or Beacon calls.
- Join Variant Scouts meetings to influence parameter definitions and VRS integration.

Shaping the future of the GA4GH Cancer Community of Interest

Recording | Agenda | Slides

The session explored the future direction of the GA4GH Cancer Community. Discussions highlighted barriers to the uptake of GA4GH standards and challenges in applying GA4GH standards to real world use cases. A federated analysis demo underscored the need for harmonisation, as results were dataset-specific. The community suggested defining a minimum set of standards, potentially through a dedicated working group, and identified clinical trial matching as a priority.

To channel feedback on GA4GH standards back to the Work Streams, the group emphasised the importance of structured and low-burden feedback mechanisms. Ideas included leveraging GitHub or Think Tanks for technical input, and potentially hosting GA4GH Implementation Forum (GIF) “Ask Me Anything” (AMA) events.

Overall, the session emphasised the importance of structured processes, deliberate meeting design, and furthering collaborations with Driver Projects, consortia, and vendors.

Key takeaways

- The Cancer Community should evolve from information sharing to actionable collaboration, aligning with Driver Project needs. Structured one-on-one engagement with Driver Projects, monthly office hours, and lightweight study groups were suggested as ways to vet emerging ideas.
- The group emphasised the need to demonstrate the value of GA4GH standards and ensure community-driven framing.
- The group recognised the need to broaden representation across cancer subtypes and build stronger connections with large consortia (e.g., AACR Genie, NCI, VIC, ClinVar Somatic WG), along with the notion of “meeting communities where they are.”

Priority action items

- Engage with GA4GH Driver Projects by setting up dedicated meetings with cancer-focused Driver Projects.
- Identify and appoint a project lead for a Clinical Trial Matching initiative to explore structured approaches to convert written eligibility criteria into machine-readable formats
- Continue development of OncoBeacon, engaging related groups such as VICC and the ClinVar Somatic working group.
- Further engage the broader community, including consortia, vendors, and external groups (e.g. AACR Genie, NCI). This could include exploring the GIF AMA events and other accessible engagement formats.

Generating guidelines for community-driven portals for infectious disease research

Recording | Agenda | Slides

This session introduced the concept of Local Pathogen Portals — web-based nodes that promote country or region-specific infectious disease research and connect to a central portal at EMBL-EBI for pandemic preparedness. The core challenge is defining a common set of expectations for what content should be in these portals, while maintaining flexibility to serve local community needs. The aim of the session was to launch a working group to develop these content guidelines. The agenda featured a presentation on the overarching Pathogen Data Network (PDN), followed by detailed case studies from the Swedish, Spanish, and Dutch national portals on their content prioritisation and outreach strategies.

Key takeaways

- The core challenge identified was the lack of guidance on what should be in an infectious disease portal (e.g., expected content, prioritisation, and production methods).
- Local Pathogen Portals are a critical part of the Pathogen Data Network (PDN), which connects them bi-directionally with a central portal to ensure efficient data flow and future pandemic preparedness.
- Case studies from the Swedish, Spanish, and Dutch portals showed that content prioritisation is driven by local mandates and community needs. The Dutch portal noted that providing a local, visible platform makes researchers more likely to deposit data (including experimental data), while the Spanish and Swedish portals rely on direct feedback and collaboration with national reference labs and local research consortia to shape their resources.

Priority action items

- Start a working group to define content guidelines and discuss what should be in an infectious disease portal.
    - This group must establish a balance, ensuring that portals have a recognisable, predictable structure and content while retaining the flexibility for each local node to tailor resources to its specific community, regional regulations, and interests.
- Foster global data discovery and interoperability by connecting local portals with the central EMBL-EBI Pathogens Portal. This means continuing efforts to integrate, reference, and annotate scattered local data and catalogs, making sure that data can be found and properly accessed by researchers globally.
    - Future work is needed to explore using GA4GH standards such as Beacon to facilitate node-to-node communication.

- Assess results from the capacity framework and self-assessment tool to help local data hubs evaluate their readiness across five categories: services, data management, best practices, infrastructure, and governance.

In the (GIF) Spotlight: highlighting GA4GH implementations

Recording | Agenda | Slides

The GA4GH Implementation Forum (GIF) session opened with brief updates and a focus on community feedback. Leaders highlighted the Q&A-driven "Ask Me Anything" (AMA) series, with events held about Beacon, Phenopackets, and Passports. The Co-Leads also shared the current portfolio of small, time-bounded GIF Projects, which includes the Federated Variant-Level Matching project, eLwazi, Cloud-based BRCA Exchange environment, and TRE Open Suite. The GIF Co-Leads also introduced a new initiative, GIF Spotlights — a lightweight, template-supported way to share real-world implementation stories, code snippets, lessons learned, and adoption tips that will build a searchable library.

Breakout discussions surfaced recurring implementation challenges, including: making the case for adopting GA4GH standards over mature in-house solutions; unclear product boundaries and expectations; gaps in practical documentation; weak feedback loops from implementers back to standards teams; and irregular community cadence. The group explored ways to tighten collaboration within the GA4GH ecosystem, improve communication and mentorship, curate implementation guidelines and a registry of live deployments, and proactively bridge regional efforts.

Key takeaways

- The community wants practical, searchable, and concrete implementation help. These could include decision guides, validation tools, examples, and a Q&A repository.
- GA4GH Spotlights can accelerate adoption by lowering barriers to sharing implementation experiences, celebrating diverse regions, and capturing lessons quickly.
- Participants called for greater clarity on the ecosystem of GA4GH groups and how they work together, along with product guidance to determine when to use which standards.

Priority action items

- Finalise and launch GA4GH Spotlights. Early Spotlight candidates include the Data Repository Service (DRS) servers map, VRS clinical integration, eLwazi "what we learned" story, BRCA Exchange, and progress from the TRE Open Suite initiative.
- Clarify ecosystem roles by developing a one-pager diagram explaining Driver Projects, GIF Projects, Spotlights, and Use Cases, and their links to GA4GH Work Streams, TASC, and the Tech Team.

- Strengthen support channels by identifying a searchable forum (e.g. Discourse) for implementation Q&A's beyond Slack.
- Create a minimum implementation guide template and a registry of live implementations in collaboration with the Work Streams and TASC.
- Launch a mentor roster and regional matchmaking.

Path to normative: implementing a GA4GH maturity model

Recording | Agenda | Slides

The meeting focused on introducing and discussing the GA4GH Product Feature Maturity Model, a framework designed to communicate the readiness and stability of GA4GH products, such as APIs, data models, and policies, to facilitate interoperability and adoption.

The session included an in-depth exploration of applying the maturity model across various product types, with breakout discussions on APIs, data models, policies, and protocols.

Key points included defining product features, distinguishing between draft, trial use, and normative stages, and addressing challenges in applying the model to non-technical products such as policies.

The meeting emphasised community feedback to refine the model and ensure its broad applicability.

Key takeaways

- The maturity model is comprehensive and adaptable, suitable for technical products such as APIs and data models, but requires customisation for policies due to their unique nature.
- Broad community feedback is essential to stress-test and refine the model, ensuring it aligns with diverse product needs and supports interoperability.
- The model enables clear communication of feature readiness to adopters, from early innovators to those requiring stable, normative standards, enhancing trust and usability.

Priority action items

- Product leads should maintain regular communication with implementers to track feedback and progress, ensuring features meet maturity criteria (e.g. moving from draft to trial use or normative).
- Develop tailored guidelines or appendices for different product types, particularly policies, to address nuances in applying the maturity model.
- Gradually introduce the maturity model by auditing existing GA4GH specifications and discussing implementation with Work Stream Co-Leads to prepare for broader adoption, potentially by the next Connect meeting.

National Initiatives Forum (NIF)

View presentation recording | View strategy discussion recording

The National Initiatives Forum (NIF) meeting focused on the advancements and challenges in genomics and healthcare data sharing across various national initiatives. The day opened with a presentation on the NHS Genomic Medicine Service, followed by updates from eight national initiatives (Canada, Singapore, Portugal, France, Germany, Finland, Australia, and Japan). The afternoon session included presentations on the New Zealand's use of the B1MG Maturity Model to benchmark its progress, Federated gnomAD and complementary projects such as the Gene Curation Coalition (GenCC) and Variant Level-Matching (VLM), the International Precision Child Health Partnership (IPCHiP), newborn screening initiatives such as the UK Generation Study, and the Australian Hospital Toolkit. The NIF meeting concluded with a discussion session to gather reflections, priorities, and ideas.

Introduction to NIF

Richard Scott (Genomics England) and Kathryn North (Murdoch Childrens Research Institute)

The meeting opened with an overview of the National Initiatives Forum (NIF), established in 2016 to advance the integration of genomics into healthcare. NIF Co-chairs Richard Scott and Kathryn North emphasised the value of in-person collaboration, the importance of respectful dialogue, and the shared responsibility of developing practical tools and resources for large-scale data sharing. The forum's role in aligning with GA4GH was highlighted, particularly in ensuring standards can be applied across diverse healthcare systems. Plans were discussed to increase workshops, broaden membership, and strengthen global engagement channels.

NHS genomic updates

Dame Sue Hill and Deborah Porter (NHS England)

Representatives from the NHS, including Dame Sue Hill and Deborah Porter, outlined its ambitious 10-year plan, which aims to incorporate genomics into 50% of healthcare episodes by 2035. Key developments included the expansion of the Genomic Medicine Service, comprehensive cancer analysis, and population health initiatives. Digital transformation is at the core of this strategy, with projects such as a unified genomic record and the digitisation of the test directory. These efforts require legislative updates to enable NHS England to serve as the data controller and to ensure strong interoperability across systems, underpinned by common data models and HL7 FHIR standards.

Updates from National Initiatives

Canada

Rob Annan (Genome Canada)

Rob Annan presented Canada's Precision Health Initiative, The program aims to bring together all provinces, strengthen industry partnerships, and address governance issues, particularly regarding Indigenous data sovereignty. While challenges remain around jurisdictional complexities and data sharing, the initiative is framed as not only a healthcare advancement but also a driver of economic development and innovation in precision medicine.

Singapore

Shih Wee Seow (Precision Health Research, Singapore [PRECISE])

Shih Wee Seow shared progress from Singapore's National Precision Medicine Programme, launched in 2017 and structured in three phases. Phase one successfully sequenced 10,000 healthy individuals, and phase two expanded this to 100,000. With strong public trust and significant industry partnerships, the initiative is now entering phase three, which will extend sequencing to 10% of the population. The Ministry of Health views genomics as a key tool for mitigating rising healthcare costs, with the rollout of national genetic testing programmes such as familial hypercholesterolemia (FH) screening. The program represents a model of rapid scale-up and strong government alignment.

Portugal

Astrid Vicente (National Institute of Health Doutor Ricardo Jorge)

Astrid Vicente provided an update on Portugal's Genomic Medicine Initiative, which is focused on creating a coherent strategy to integrate genomics across its three healthcare systems. The country is working closely with the European Commission's Directorate-General Reform to design a legal and ethical framework for genomic and health data. Using a maturity level model to assess progress across multiple domains, Portugal has identified seven priorities, including governance, data management, and clinical service delivery. These steps are intended to ensure high-quality patient care and prepare for long-term integration of genomics into national health policy.

Implementation of the Maturity Model in New Zealand

Cristin Print (University of Auckland)

Cristin Print provided an update on the integration of genomics within New Zealand's unique healthcare landscape, particularly emphasising Indigenous leadership and equity. Current efforts include a Ministry of Health precision health briefing and ongoing projects despite financial constraints and significant inequities. A key tool is the Beyond One Million Genomes (B1MG) Maturity Level Model, which New Zealand plans to run both nationally and within a remote Māori-led health organisation by year's end. This will generate actionable insights for government investment, strengthen advocacy, and

provide a framework for international collaboration, with support already received from Australia, the UK, and other NIF partners.

France

Frédérique Nowak (Inserm)

Frédérique Nowak presented on the launch of France's Genomic Medicine Initiative with the goal of embedding genomics directly into clinical practice for rare diseases and cancer. Building on established national rare disease and cancer plans, the programme created structured networks of diagnostic laboratories, cancer genetics services, and clinical trial centers. Two national whole-genome sequencing laboratories were established in Paris and Lyon, ensuring equitable access across all regions, including overseas territories. Having recently completed a maturity level model assessment, France is now preparing the next phase of its strategy, with working groups drafting updates to expand whole-genome sequencing and further integrate genomic medicine nationwide.

Germany

Jan Eufinger (German Human Genome Phenome Archive [GHGA])

Jan Eufinger from Germany presented two major initiatives. The German Human Genome Phenome Archive (GHGA), launched in 2020, serves as the national node of the Federated European Genome-Phenome Archive, providing controlled-access storage and standards for omics data, with outreach, legal frameworks, and community engagement in cancer and rare disease. In parallel, a new national genome sequencing programme has begun, enabled by law and health insurance coverage, aiming to sequence around 100,000 genomes/exomes over five years for rare diseases and cancer. Data will feed into the new genomeDE platform, overseen by the Federal Institute for Drugs and Medical Devices, with GHGA supporting infrastructure and plans to enable research use through federated analysis and a secure processing environment.

Finland

Helen Cooper (FinnGen)

Helen Cooper highlighted the FinnGen project, a large public–private partnership leveraging the country's unique population history, permissive biobank law, and rich health registries. With around 500,000 participants collected mainly from hospital biobanks, FinnGen combines genotype, clinical, and registry data to accelerate research and drug development. The project has returned data to biobanks for broader use, released large-scale genome-wide association study results publicly, and built secure cloud-based research environments for partners. Current efforts focus on integrating omics and clinical datasets, harmonising health records via the Observational Medical Outcomes Partnership Common Data Model, and ensuring sustainability beyond the project's research phase, potentially through legislative change to establish FinnGen as a permanent national resource.

Japan

Soichi Ogishima (Tohoku University)

Soichi Ogishima outlined two major national genomic medicine initiatives: a whole genome implementation programme led by the Ministry of Health, and a large biobank project supported by the Ministry of Education. The biobank includes more than 150,000 participants, featuring both a population cohort and a unique three-generation cohort, with extensive genomic, exposome, and clinical data integrated into a single database. To date, 100,000 genomes have been sequenced in collaboration with government and industry. Japan is also advancing standards for pedigree and exposome data, connecting with Biobank Network Japan and international partners, and piloting the return of results to participants. Current challenges include restrictive policies on cross-border data sharing and the need for common governance frameworks, as well as building federated research environments to support international collaboration.

Australia

Tiffany Boughtwood (Genomics Australia)

Tiffany Boughtwood outlined the transition from Australian Genomics to Genomics Australia within the federal Department of Health, building on strong momentum. Given Australia's federated health system, the new body is coordinating policy and exploring future governance with states/territories via the Health Technology & Genomics Collaboration. Early priorities include improving access to precision oncology beyond research projects, tackling research-to-clinical translation barriers, advancing information management, and embedding First Nations-led engagement. Genomics Australia is forming an Advisory Council, pursuing international partnerships, and supporting legislation to limit life insurers' use of genetic information — all aligned to a 10-15 year vision for routine, equitable genomic care.

Updates from community tools

Federated gnomAD and elements of GenCC and VLM

Heidi Rehm (Broad Institute)

Heidi Rehm provided an update on expanding gnomAD to better capture global diversity via a federated model. Countries can run the gnomAD pipeline locally and share aggregated variant frequencies, avoiding cross-border raw data transfers. Complementary efforts include GenCC (which provides crowdsourced, harmonised gene-disease validity, including candidate genes) and variant-level matching (VLM) for "one-sided" queries that find variant-phenotype matches across datasets to reduce variants of unknown significance rates. GA4GH-aligned APIs and standards underpin these efforts to speed gene discovery and make rare-disease diagnosis more equitable.

International Precision Child Health Partnership (IPCHiP)

Lindsay Smith (The Hospital for Sick Children)

Lindsay Smith presented on IPCHiP, which links paediatric centres in Toronto, Boston, London, and Melbourne to pool data and expertise. Its flagship study, Gene-STEPS, which focuses on infants with unexplained epilepsy, shows ~43% diagnostic yield across over 500 cases and has informed local policy (e.g. funding for clinical genome sequencing). A new study, GEMSTone, will focus on hypotonia. The collaboration is tackling legal and technical barriers to international data sharing and is exploring federated approaches and shared analysis tools to connect sites without forcing clinicians to learn new platforms.

Newborn screening

Richard Scott (Genomics England)

Richard Scott presented on newborn genome sequencing pilots, which have expanded rapidly to more than 50 programs worldwide. Common elements include consent, sequencing, returning actionable results, programme evaluation, and strategies for data reuse — implemented differently across health systems. In the UK, the Generation Study aims for 100,000 newborns, with more than 23,000 enrolled, building evidence for potential national adoption within 10 years. Ongoing priorities include harmonising gene lists, ensuring equitable recruitment, aligning with health-economic methods, and planning for future reuse (e.g. pharmacogenomics and common-disease risk).

Hospital toolkit

Clara Gaff (Melbourne Genomics Health Alliance)

Clara Gaff presented on Australia's new Hospital Genomics Toolkit, which distills organisational actions needed to embed genomics in routine care. It outlines seven key actions — starting with creating a genomic leadership group spanning executives, digital/IT, pathology, quality, and clinical services — to align electronic medical records and laboratory information management systems, governance, and workforce training. Piloted in several large hospitals and transitioning to health department stewardship, the toolkit is designed for varied contexts and applies across rare disease, cancer, pathogen genomics, and pharmacogenomics. It helps bridge the gap between genomic experts and hospital leadership to deliver safer, higher-quality, and sustainable genomic services.

Discussion: the strategy of NIF going forward

Collaboration and cross-learning

Participants highlighted that collaboration across initiatives is "ripe." Key opportunities include benchmarking across countries, comparing governance models (clinical vs. research vs. hybrid), and advancing implementation science to bridge technical capability with real-world adoption. Gathering together fosters trust, cross-talk, and mentorship — which is particularly valuable for newer initiatives.

A "coalition of the willing" was proposed, allowing members to self-nominate for working groups and indicate time commitments. Given the diversity of national approaches (research-focused vs. healthcare implementation), there may be value in distinct streams for each.

Expanding the Federated gnomAD project

The federated gnomAD initiative drew strong interest, with over 12 countries engaged and the first integration expected soon. The goal is to expand steadily, demonstrate scalability, and seek international funding under GA4GH leadership. Equity and inclusion were emphasised, particularly ensuring representation of under-studied populations.

Governance and signaling

Participants stressed the importance of clearly signaling the governance context of data, whether research, clinical, or hybrid. Misunderstandings can be avoided by developing simple indicators that specify the operating framework.

Challenges and considerations

Data sharing

Challenges to data sharing persist, including:

- International sharing faces legal and forensic risks, requiring a staged approach that begins with lower-risk strategies, such as:
    - One-sided matching (e.g., Matchmaker exchange, where a query is sent to another dataset, but the data itself is not fully shared).
    - Federated approaches, where data can be analyzed without transferring.
    - Controlled access and trusted research environments (TREs), which mitigate risks but limit utility.
- A common language is needed to explain risks and benefits to governments, funders, and the public. This can also clarify differences between data access and data sharing

Resource Limitations

Financial and human resource constraints remain a recurring barrier to sustained progress.

- Many initiatives operate with limited dedicated staff and rely heavily on volunteer contributions. This constrains their ability to maintain shared resources, such as the NIF Toolkit

Equity

Persistent disparities continue to affect equitable participation in genomics.

- Indigenous and low- and middle-income country (LMIC) communities face unique governance and trust challenges, often lacking representation in data-sharing discussions. Participants emphasized the importance of Indigenous leadership and the co-development of culturally respectful frameworks for data governance.
- While technical standards can be global, the ethical, legal, and cultural contexts vary widely. Inclusive approaches must therefore extend beyond technical alignment to include shared frameworks for trust, reciprocity, and local benefit.

Tools, models, and resources

- **Beyond 1 Million Genomes (B1MG) Maturity level model (MLM):** Used by countries such as Portugal, New Zealand, and Denmark to identify gaps and guide priorities. Coordination nodes across Europe and Asia-Pacific could support benchmarking.
- **Toolkits and resources:** The existing toolkit should be updated based on members' needs. Consider moving from a static repository toward a catalogue of expertise.
- **Industry partnerships:** Different national models offer lessons on consent and data sharing. Government support remains a key enabler.

Community engagement

Workshops should first reduce duplication with GA4GH Work Streams, then target priority areas defined by NIF members. They should also have a clear purpose: either launching collaborative projects or providing feedback on GA4GH standards.

A mentorship programme is being developed, emphasising multi-level learning (from peers and senior leaders).

Strategic actions

- Advocate for international genomic data sharing, supported by a forthcoming “call to arms” publication.
- Develop a risk-benefit matrix for data sharing, comparing data types (raw vs. aggregate; cancer vs. germline), and share it with the Regulatory & Ethics Work Stream
- Explore how GA4GH standards capture data-reuse benefits and communicate value to governments.
- Engage the Beyond One Million Genomes (B1MG) project and the World Health Organization (WHO) on ethics, legal issues, and coordination.
- Strengthen links with GA4GH Communities of Interest, starting with the Cancer and Rare Disease communities.
- Promote intra-country data sharing as a stepping stone toward broader adoption.
- Hold a follow-up planning workshop in early 2026.
- Establish indicators to clarify research vs. clinical contexts.
- Circulate advocacy materials, such as the “The value of large-scale programmes in human genomics” paper (Horn et al.)and Toolkit documents.
- Compile self nominations for potential working groups
- Explore Toolkit updates via contractors, junior staff, or expertise catalogue

---

This meeting report was developed with the aid of Google Gemini.