GA4GH April Connect 2025 Meeting Report



Executive summary

The GA4GH April Connect 2025 Meeting brought together members of the GA4GH community for four days of collaboration, technical advancement, and strategic reflection. 150 attendees convened at the Broad Institute of MIT and Harvard, and more than 400 participants from around the world joined virtually. The meeting focused on driving actionable outcomes and advancing the GA4GH mission to ensure standards are effectively implemented, widely adopted, and make a meaningful impact in real-world settings. The tone throughout the meeting was pragmatic and forward-looking, and focused on closing the gap between aspiration and action.

Sessions explored challenges across the GA4GH product lifecycle — from onboarding and documentation to federated infrastructure and policy harmonisation. Multiple hackathons and working groups convened to refine standards like the Variation Representation Specification (VRS), refget, and htsget, while others focused on ways to leverage them, such as schemas, registries, and tooling to support implementation at scale. Policy-focused sessions tackled consent, data access, governance frameworks, and regulatory gaps, underscoring the need to align technical and ethical infrastructure. Communities using GA4GH standards — ranging from rare disease to

prenatal genomics, pathogens, and neuroscience — shared both successes and barriers, reinforcing the need for tailored implementation support, responsive governance, and cross-Work-Stream collaboration.

Strategic sessions also tackled big-picture questions. What does it take to build trust in federated data sharing systems? How can GA4GH better support low- and middle-income regions? What role should GA4GH play in the responsible adoption of generative AI for health? Answers to these questions centered on adaptability, inclusivity, and shared understanding — highlighting the importance of real-world partnerships, strong infrastructure, and policy frameworks that account for both ethical nuance and technical complexity. At the same time, participants emphasised that many of these challenges span the entire healthcare ecosystem — not just genomics — and called on GA4GH to focus its contributions where its expertise and standards can make a unique and complementary impact.

Product-focused sessions unveiled exciting progress: Beacon v2.5 is expanding support for metadata and clinical integration; VRS and Beacon are aligning for more expressive variant discovery; and the Cloud Work Stream proposed new standards for trust and system orchestration including those supporting federated AI. Sessions on compliance, data access modelling, and consent continued to emphasise the importance of harmonised governance — particularly in light of international data movement and emerging research consortia like the Human Pangenome Project.

Education and engagement remained recurring themes. GA4GH's evolving global engagement strategy highlighted work to expand educational materials, regional representation, and multilingual resources. The relaunch of the GA4GH Implementation Forum (GIF) further reinforced the move from standards development to real-world problem solving. Communities called for clearer documentation, more accessible tooling, and co-designed implementation pilots that reflect their needs, priorities, and constraints.

In sum, the week's discussions underscored a shared commitment to GA4GH's mission — not only to develop consensus-based standards, but to foster a global community that is advancing the responsible, interoperable, and impactful sharing of genomic data. Across sessions, participants returned to a common goal: transforming technical progress into real-world benefit for science, medicine, and society.

Key themes and takeaways

From specification to real-world implementation

- **Starter kits evolve:** GA4GH aims to create complete, use-case-driven starter kits to demonstrate real-world value.
- **Compliance suites and testbeds:** broader adoption of test suites (e.g. htsget) is revealing friction points and helping track readiness and interoperability across standards.
- **Tooling and documentation are essential:** the need for improved onboarding, documentation, and centralised reference implementations were top priorities in nearly every product-focused session.

Global equity and inclusion

- **Regional liaisons and local workshops:** new roles and outreach programs should foster meaningful global participation while avoiding fragmentation.
- **Education series:** a modular learning series, tailored to both technical and non-technical users, should be developed to broaden understanding of GA4GH products.
- **Multilingual and localised engagement:** participants called for non-English materials, culturally adapted outreach, and mentorship programs to support broader accessibility.

Policy, consent, and governance

- **Data access models must be computable:** access policies need machine-readable formats for automation and transparency — balanced with ethical safeguards and human review.
- **The Regulatory & Ethics Work Stream (REWS) goes practical:** REWS is focused on implementation-ready guidance (e.g. model policies, checklists, and templates) that are adaptable across jurisdictions.
- **Consent compatibility matters:** the session “identifying core consent elements for pangenomic studies” highlighted that trust-building requires transparency around data sharing, communicating the risks of re-identification, and community-driven consent models.

Federation and infrastructure

- **Beacon v2.5 and Variation Representation Specification (VRS) Integration:** GA4GH is co-developing tools for more expressive and precise federated queries, starting with pilot implementations.
- **Cloud ecosystem alignment:** tools like the Data Repository Service (DRS), the Workflow Execution Service (WES), and Passports are aligning on the need for trust metadata, cold storage flags, and cross-system compatibility.
- **Flexible federation models:** lightweight “relay + peer” architectures are enabling more groups to join federated networks without heavy engineering overhead.
- **Data discovery convergence:** Data Connect, Experiments Metadata Standard, Whole Genome Sequencing (WGS) Quality Control (QC) standards, and the Data Model and Schema Consensus (DaMaSC) working group are aligning to improve discoverability and schema interoperability across GA4GH ecosystems.

Artificial intelligence (AI), large language models (LLMs), and innovation

- **Phenopackets meets LLMs:** GA4GH is investigating how structured formats like Phenopackets could support the safe and effective benchmarking of large language models, with early community interest in defining shared evaluation frameworks and datasets.
- **Federated machine learning:** a new GA4GH Implementation Forum (GIF) project is exploring federated learning (FL) pipelines using existing GA4GH standards — anchored in real-world use cases.
- **Verifiable credentials for visas:** early conversations are exploring how cryptographic credentials could support more expressive and secure authorisation models.

Emerging standards and use cases

- **Refget Sequence Collections:** now GA4GH-approved, this new standard helps to unambiguously define and identify collections of reference genomes — including for pangenome use cases.
- **Experiments Metadata Standard:** a draft metadata schema is ready for comment and slated for product approval by the October Plenary.
- **Prenatal genomics prototypes:** Beacon and Phenopackets are powering real-world prenatal repositories — tailored for trio genomes and ultrasound-derived phenotypes.

Organisational evolution

- **Chief Product Officer (CPO) onboarded:** Sasha Siegal, the new CPO, will drive community-informed product development and streamline coordination across Work Streams.
- **GA4GH Implementation Forum (GIF) relaunch gains traction:** GIF now supports product “Ask Me Anything” (AMA) events and multi-institutional collaborations through GIF Projects.
- **Cross-Work-Stream governance in focus:** alignment efforts between the Technical Alignment Subcommittee (TASC) and the Data Model and Schema Consensus (DaMaSC) working group are helping to clarify product life cycles, schema ownership, and deprecation policies.

Session summaries

Day 1

- Opening remarks
- From standards to practice: overcoming barriers in genomic data sharing
- Workshopping best practices for community engagement in product development
- Expanding and evolving the Federated Variant Level Matching platform for rare disease
- Leveraging use cases for standardising data structures, variant stores, and file formats working session
- gNBSeq: progress and planning
- GA4GH's global engagement strategy updates: progress on key deliverables
- Genomic Knowledge Standards (GKS) Connect hackathon report-out: outcomes, challenges, and next steps
- Service Info and Service Registry: present and future
- Data visiting
- Genomic health data ecosystem (HDE) development and maturity level

Day 2

- Beacon v2.0: updates and insights
- Towards clinical application of GA4GH products
- (hts)Get compliant with it: hackathon edition
- Beacon for metadata discovery: expanding the scope
- Data access modelling for GA4GH ecosystems
- GA4GH data model best practices lunch session
- The path to implementation of Regulatory & Ethics Work Stream (REWS) products
- Cloud Work Stream: new standards
- The "TASC-athon"
- Future of the Data Connect standard
- Experiments Metadata Standard: now and beyond
- Refget Sequence Collections standard
- Genomic Knowledge Standards (GKS) Work Stream road map

Day 3

- Brainstorming technical next steps for the Pathoplexus pathogen genomic data sharing ecosystem
- Passports, visas, and task-specific tokens
- Standards for enabling federated machine learning
- Visa encoding in verifiable credentials data model v2.0
- Exploring AI-driven innovation: benchmarking large language models for rare disease diagnostics
- Data Repository Service (DRS) working session
- Finding synergy: Technical Alignment Sub-committee (TASC) and Data Model and Schema Consensus (DaMaSC)
- Collaborative brainstorming for the next phase of GA4GH Implementation Forum (GIF) activities
- VLM + VRS = NextGen Beacon
- Regulatory & Ethics Work Stream (REWS) in focus: bridging regulatory and ethical gaps across GA4GH
- Adapting GA4GH standards to meet the needs of the Neuroscience Community

Day 4

- Informing clinical data sharing standards
- Querying Phenopackets with Beacon v2.0: a prototype implementation in the prenatal space
- GA4GH Schema Registry specification
- Automating patient matching to clinical trials
- GA4GH Whole Genome Sequencing (WGS) Quality Control (QC) standards meeting
- Pangenome structural variant reporting and representation
- GA4GH in the real world
- Identifying core consent elements for pangenomic studies

Day 1

Opening remarks

View Recording | Slides

Session Chair: Peter Goodhand, Heidi Rehm

Meeting summary

This opening session focused on reaffirming the mission and global impact of GA4GH while emphasising the importance of sustainable funding, effective implementation of standards, and community engagement. The conversation highlighted both organisational stability and ongoing challenges in global collaboration. Speakers encouraged the community to think proactively about how GA4GH can better support the implementation of its products, especially in the context of cross-border projects and underfunded initiatives. The launch of a new Chief Product Officer role and reflections on GA4GH's funding, history, and future direction anchored a broader call for innovation, inclusivity, and collective responsibility.

Key takeaways

- **More support is needed for standards implementation**: while GA4GH has historically focused on developing foundational standards, there is increasing recognition that additional support is needed to help implement those standards — especially in global, federated, or under-resourced settings.
- **Sustainable, multi-national funding is crucial**: GA4GH has typically relied on a core group of funders. Despite the critical new five-year funding recently secured from Canada, ongoing funding from diverse global sources — through both grants and creative partnerships — remains essential.
- **Leadership and community engagement are evolving**: the appointment of Sasha Siegal as Chief Product Officer (CPO) signals a push for deeper community-informed processes, better product development infrastructure, and more streamlined engagement across GA4GH.

Priority action items

- **Support implementation projects**: actively identify and pursue funding opportunities to help community members implement GA4GH products — especially where centralised or cross-border coordination is needed. GA4GH can serve as a grant applicant or fund distributor when appropriate.
- **Enhance accessibility and infrastructure**: improve documentation, GitHub organisation, and community tools to make standards easier to adopt. This includes overhauling the "starter kit" and embedding analytics to track product use and impact.
- **Support inclusive and global participation**: leverage assigned experts, travel grants, and sponsorships to ensure inclusive participation. Encourage institutions and funders to support both financial and in-kind contributions to sustain and diversify engagement.

Day 1

Plenary session: from standards to practice: overcoming barriers in genomic data sharing

Recording (start at 18:00) | Agenda | Slides

Session Chairs: Lindsay Smith, Christian Marshall, Augusto Rendon, Graham Ritchie

Meeting summary

This session centered on real-world implementation challenges of GA4GH standards, with presentations from Genomics England and the International Precision Child Health Partnership (IPCHiP). Genomics England showcased its evolving national infrastructure for clinical and research genomics, emphasising how GA4GH standards such as the Data Repository Service (DRS), htsget, Data Connect, and the Data Use Ontology (DUO) are critical to achieving scalable, federated, and platform-agnostic data sharing across the UK's NHS. IPCHiP shared a contrasting experience of international federated data sharing across four paediatric hospitals, revealing significant technical and governance barriers that have stalled progress despite strong shared goals. A key theme was the tension between aspiration and implementation — standards are essential, but not sufficient. Resources, institutional alignment, data harmonisation, governance constraints, and the sheer complexity of connecting heterogeneous environments emerged as pressing barriers.

Key takeaways

- **Standards enable scalability, but not without investment:** Genomics England's success illustrates how standards like DRS and Data Connect can simplify complex infrastructures — but only when paired with sustained funding, policy alignment, and central support.
- **Federation is harder than it looks:** IPCHiP's experience shows that even small networks struggle with federated data sharing due to disparate IT policies, resource limitations, and lack of shared governance or infrastructure.
- **Governance barriers often manifest as technical ones:** many technical challenges arise from security and data protection requirements set by governance structures — highlighting the inseparability of the two.

Priority action items

- **Develop practical, flexible federation models:** promote modular, cost-aware federation strategies that work across diverse environments — including clinical settings with restrictive governance.
- **Support the community with tools and templates:** provide plug-and-play infrastructure options, policy templates, and practical implementation guidance — especially for small or resource-limited projects.
- **Elevate the value proposition for funders:** advocate for more evidence-based assessments of the impact of data sharing, including cost-benefit analyses, to help secure long-term investment in federated systems.

Day 1

Plenary session: workshopping best practices for community engagement in product development

Recording | Agenda | Slides

Session Chairs: Mónica Muñoz Torres, Melissa Cline, Venkat Malladi

Meeting summary

This interactive workshop focused on improving global engagement and inclusivity in GA4GH product development and implementation. Rooted in real-world survey data, participants identified key barriers such as time zone challenges, email fatigue, technical complexity, and lack of clarity around how to engage. The highly interactive session used live Sli.do polling and open discussion to explore practical solutions, including better documentation, up-to-date reference implementations, localised workshops and hackathons, improved feedback mechanisms, and mentorship programs. Emphasis was placed on reaching underrepresented regions and stakeholder groups — including patient advocates, regulators, and industry — and building a more navigable and welcoming entry point into the GA4GH ecosystem.

Key takeaways

- **Engagement gaps persist despite interest:** many individuals are eager to participate in GA4GH work but face barriers such as unclear entry points, limited time, and misaligned communication channels.
- **Reference implementations and real-world use cases are critical:** developers and implementers consistently emphasised the need for working examples, testing tools, and compliance suites to support adoption.
- **Broader and more inclusive outreach is needed:** the GA4GH community must better engage stakeholders such as patient advocates, regulators, and newcomers from underrepresented regions, possibly through local language content, targeted outreach, and mentorship.

Priority action items

- **Create and maintain a centralised repository on the GA4GH website** (e.g. repurposing ga4gh.org/our-products/implementations) of up-to-date reference implementations, testing tools, and example use cases for GA4GH standards.
- **Develop lightweight, accessible onboarding materials** (e.g. short explainer videos and “getting started” guides), and mentorship or buddy programs.
- **Pilot more inclusive engagement formats**, such as hosting implementation-focused fora, translating documentation into multiple languages, conducting regional hackathons, and establishing local-language outreach chapters, in collaboration with the Communities of Interest.

Expanding and evolving the Federated Variant Level Matching platform for rare disease

Recording | Agenda | Slides

Session Chairs: Heidi Rehm, Nara Sobreira

Meeting summary

This session focused on the progress, challenges, and next steps for building a federated Variant Level Matching (VLM) network using the GA4GH Beacon v2.0 API, specifically for rare disease genomics. Presenters outlined current node implementations, with three already connected (VariantMatcher, Franklin, and Geno2MP/MyGene2), and discussed the technical and governance aspects of adding new nodes — including centralised authentication using Auth0, data sharing models, and clinical versus research data regulations. The session also highlighted a proof-of-concept for a scalable and lightweight federation architecture, enabling small projects to participate more easily. Finally, there was an open discussion on expanding support for complex queries, integrating clinical data, and ensuring privacy, transparency, and scalability as the network grows.

Key takeaways

- **Federated VLM is live and expanding:** several research-based nodes are already connected using Beacon v2.0, with additional nodes (e.g., seqr on AnVIL, RD-Connect) close to joining.
- **Scalable authentication via Auth0:** the adoption of centralised Auth0-based authentication overcomes the scalability limitations of point-to-point token sharing and simplifies node onboarding.
- **Flexible architecture supports broader participation:** a proof-of-concept "relay + peer" model using lightweight Postgres replication and Dockerised deployment allows smaller groups to contribute variant-level data without extensive engineering overhead.

Priority action items

- **Refine and standardise phenotype-rich, one-sided and gene-level queries** across all VLM nodes to maximise the clinical utility of matches beyond simple variant lookup.
- **Engage clinical labs and institutions to define regulatory, consent, and technical requirements** for including clinical (non-research) datasets in the VLM network.
- **Establish and publish clear onboarding guides for new nodes**, including Auth0 setup, beaconisation, and minimal data models — especially for groups planning to federate small datasets or use the "relay" model.

Day 1

Leveraging use cases for standardising data structures, variant stores, and file formats working session

Recording | Agenda | Slides

Session Chair: Heidi Rehm

Meeting summary

This session explored scalable variant file formats and community needs around large-scale genomic datasets. Rob Davies presented an overview of the evolution of VCF up to version 4.5, detailing critical updates like local allele indexing and structural variant support. Katherine Chao followed with a description of the Hail Variant Dataset (VDS) used in the Genome Aggregation Database (gnomAD) to enable efficient, scalable, and sparse storage of variant data. Lee Lichtenstein presented the *All of Us* Research Program Genomic Variant Selection (GVS) system, a scalable genomic variant store using BigQuery, and emphasised supporting diverse user needs across data sets and formats. Jonathan LeFaive introduced the sparse allele vectors (SAV) format, a sparse extension of BCF with potential for more efficient analyses. Oliver Hofmann spoke about HTSJDK and he and other speakers raised concerns about ecosystem support, especially Java tools stuck on older VCF versions, and encouraged consideration of APIs, tooling compatibility, and developer resourcing. The session concluded with rich discussion on potential solutions, including parquet/iceberg adoption, standardised APIs, and defining use cases for a “future of VCF” approach.

Key takeaways

- **VCF v4.5 introduces valuable enhancements (e.g. local alleles and better phasing) but is only a partial solution to scalability and ecosystem compatibility challenges** — particularly around tool support beyond C-based implementations.
- **Alternative representations like Hail VDS, GVS, and SAV** offer improvements in storage efficiency, data sparsity, and flexibility for large-scale analysis, but lack standardisation and universal support.
- **The community is grappling with diverging needs and fragmented tooling** — leading to a consensus that future progress may require standardised APIs, use-case-driven approaches, and thoughtful coordination across stakeholders.

Priority action items

- **Facilitate an inclusive, ongoing community dialogue** to assess core use cases for variant data representation — such as data exchange, population-scale analysis, and visualisation — with the goal of identifying where standardisation or convergence is needed across existing and emerging formats (e.g. VCF, VDS, SAV, and GVS), rather than assuming further evolution of VCF alone.
- **Acknowledge and address resource constraints in the tooling ecosystem**, especially for widely used but under-resourced libraries such as HTSJDK and the Integrative Genomics Viewer (IGV). Rather than focusing solely on early engagement, support should be mobilised to modernise these implementations — ideally aligning new formats and APIs with the Variation Representation Specification (VRS) information model to promote long-term interoperability and sustainability.

- **Explore and potentially pilot an interoperable API layer** (e.g. htsget-style or Beacon-style queries) that abstracts backend implementations (e.g. VDS, GVS, parquet) while supporting scalable access, federation, and reuse.

gNBSeq: progress and planning

Recording | Agenda | Slides

Session Chairs: Anna Lewis, Yvonne Bombard

Meeting summary

This GA4GH REWS' newborn screening working group is developing a tool to help implementers navigate ethical, legal, and social issues (ELSI) in genomic sequencing programs for healthy newborns (gNBSeq). Anna Lewis introduced a literature review powered by AI (Anthropic's Claude) that extracted and synthesised over 2,000 ELSI-related issues across 30 key decision points in gNBSeq programs. The aim is to create a practical, non-prescriptive tool for implementers — such as research projects and public health bodies — that can support thoughtful, context-specific program design. Participants discussed gaps in the literature, emphasised the importance of considering public health framing and not just clinical ethics, and stressed the need for the tool to accommodate diverse legal, regulatory, and cultural environments globally. Key topics included the limits of consent models, data interoperability, translation to clinical practice, and capacity in low-resource settings.

Key takeaways

- **A strong consensus emerged that the tool should be practical, flexible, and internationally adaptable,** recognising variation in consent models, public trust, regulatory frameworks, and healthcare infrastructure across jurisdictions.
- **Key decision points must account for both current research contexts and potential public health implementation,** including clear articulation of when and how genomic sequencing might transition from pilot projects into routine newborn screening.
- **Stakeholders emphasised the need to foreground children's rights, public health ethics (distinct from individual autonomy frameworks), and long-term data stewardship,** ensuring that issues like data re-use, system capacity, and equity are not sidelined.

Priority action items

- **Refactor the tool's schema** to more clearly reflect high-priority decision points (e.g. consent models, clinical validation, access to treatment, and data reuse), organised by distinct use cases: research project setup, transitioning to public health implementation, and transparency and comparability across projects.
- **Human-review and refinement of AI-generated summaries** of each decision point's literature base, drawing on expertise from the working group to validate, supplement, and interpret the evidence — particularly where key perspectives (e.g. gray literature, policy, lived experience) may be missing.
- **Iterate and pilot the tool with real-world projects** like the Genomics England Generation Study and others in the International Consortium on Newborn Sequencing (ICoNS), to ensure usability and relevance, and consider developing multiple dissemination formats (e.g. guidance documentation, slide decks, and interactive website) to suit different user groups.

Day 1

GA4GH's global engagement strategy updates: progress on key deliverables

Recording | Agenda | Slides

Session Chair: Neerjah Skantharajah

Meeting summary

This interactive session explored GA4GH's evolving Global Engagement Strategy (formerly EDI Strategic Plan), which aims to embed global participation more deeply across the organisation. It began with an overview of completed and ongoing deliverables — such as the updated Data Sharing Lexicon, the Genomic Data Initiatives Survey, and the Travel Fellowship Program — before diving into three specific deliverables: (i) a strategy to ensure regional representation in use cases through guidelines, liaisons, and regional workshops; (ii) updating GA4GH's implementation registries to support better discoverability and consistency across standards; and (iii) plans for an educational series to increase product uptake, particularly across diverse global communities. In-person and virtual attendees provided detailed, collaborative feedback via sticky notes digital whiteboard.

Key takeaways

- **Regional liaisons are central to meaningful global engagement:** participants overwhelmingly supported the creation of regional liaisons, noting their potential to improve outreach, collect geographically relevant use cases, and help overcome time zone and resource barriers. However, care must be taken not to over-fragment already limited contributor pools.
- **The implementation registry must evolve for clarity and utility:** community members highlighted the need to differentiate between "implementations" and "deployments," include richer metadata on implementation (e.g. contact information, maturity, and proof of interoperability), and introduce a streamlined, community-updatable process to ensure accuracy and utility.
- **Product education should be diverse, targeted, and accessible:** attendees emphasised that educational materials must cater to both technical and non-technical audiences. Ideas included modular learning paths, short explainer videos, use-case-driven storytelling, and improved "how-to" documentation — prioritised by product impact and user type.

Priority action items

- **Pilot and fund regional liaison roles:** define clear expectations and support mechanisms (not volunteer-only) for liaisons. Explore co-hosted satellite events and local engagement models such as viewing parties or use-case feedback workshops.
- **Consolidate and revamp the implementation registry:** align the website and Service Registry into a unified, crowdsourced system. Define clear terminology for implementations versus deployments and ensure discoverability, regular updates, and endorsement pathways.
- **Launch a product education strategy with multi-format content:** develop a short introductory series ("what is GA4GH?" and "how does this product work and why is it important?"), prioritise product areas with broad applicability or interoperability needs, and ensure consistency and accessibility across all materials.

Day 1

GKS Connect Hackathon report-out: outcomes, challenges, and next steps

Recording | Agenda | Slides

Session Chairs: Larry Babb, Kyle Ferriter, Kori Kuzma

Meeting summary

The Genomic Knowledge Standards (GKS) Connect Hackathon report-out session highlighted significant advancements and collaborations across the GA4GH GKS community. Participants shared outcomes from their one-and-a-half-day intensive work, involving around 25 contributors focused on refining and implementing the Variant Annotation Specification (VA-Spec), the Categorical Variation Representation Specification (Cat-VRS), and the Variation Representation Specification (VRS) standards. Major projects included schema improvements for MaveDB, development of new study result profiles, enhancements to documentation and examples, creation of a unified schema browsing tool, FHIR translation to and from VRS, and Docker-based “VRS in a Box” utilities. Community-driven innovation and real-world implementation needs strongly influenced the hackathon’s direction, demonstrating how hands-on collaboration rapidly accelerated development and identified actionable gaps, especially in tooling, documentation, and onboarding.

Key takeaways

- **Implementer-driven development works:** shifting focus to real-time engagement with implementers through hackathons has greatly accelerated GKS product adoption and refinement, notably for the VA-Spec and Cat-VRS.
- **Documentation and accessibility remain bottlenecks:** participants highlighted the ongoing need for easier navigation of schemas, more intuitive documentation, and better tooling to reduce barriers for new adopters.
- **Prototype successes point to scalable tools:** innovations like the new tumor variant frequency profile, the FHIR-to-VRS translator, and “VRS in a Box” are paving the way for broader adoption and integration of GKS tools in diverse systems, including clinical applications.

Priority action items

- **Improve and consolidate documentation:** launch a working group or focused sprints to enhance schema documentation, add intuitive and educational examples, and simplify navigation across GKS specifications (e.g. through Tristan’s unified browser).
- **Standardise and scale tooling:** finalise and distribute the “VRS in a Box” containers, consider splitting VRS-python into core/reference and utility libraries, and evaluate moving to more modern/refget-compatible sequence repositories.
- **Build implementation support structures:** establish recurring co-working or “doc jam” sessions, support outreach to implementers, and consolidate use cases and tooling (such as the Fast Healthcare Interoperability Resource [FHIR] translator and builder classes) into a centralised onboarding guide or toolkit.

Day 1

Service Info and Service Registry: present and future

Recording | Agenda | Slides

Session Chairs: Miro Cupak, Jonathan Fuerth

Meeting summary

This session revisited the GA4GH-approved Service Info and Service Registry standards — two lightweight APIs that support service discovery and metadata sharing within federated networks. Though both have seen widespread adoption (especially Service Info), limited connection to adopters prompted this community-driven session to crowdsource requests for enhancements of the standards. Using a “lean coffee” format, attendees proposed, voted on, and discussed top priorities such as discovery of access mechanisms supported by services, compliance testing and how to better support discovery of optional service capabilities and API versions.

Key takeaways

- **Service Info and Registry are widely adopted but under-evolved:** despite being stable and well adopted, there are enhancements the community would like to see implemented — highlighting a need for evolving the standards.
- **Strong interest in discoverability and interoperability features:** attendees emphasised the need for features like authentication metadata (e.g. OAuth/Passport support) and capability broadcasting to enhance machine-readable service discovery.
- **Community wants clarity and coordination:** there’s a clear demand for improved documentation, examples, and coordination around standard extensions, particularly to handle partial implementations and federated authentication scenarios.

Priority action items

- **Develop guidance for extensions and capabilities:** create clear documentation for how to extend Service Info, broadcast capabilities, and advertise partial API implementations in a machine-readable way.
- **Integrate with compliance verification tools and testbed:** leverage the GA4GH testbed infrastructure (currently under development by the GA4GH tech team) to provide lightweight validation tools for implementers and potentially publish public compliance results.
- **Evaluate inclusion of optional metadata fields:** assess the feasibility and usefulness of incorporating fields for organisational logo URLs, geographic information and access setup (e.g. authentication endpoints, flows, and client IDs) within Service Info and Service Registry.

Data visiting

Recording | Agenda | Slides

Session Chairs: Donrich Thaldar, Yann Joly

Meeting summary

This session provided a comprehensive exploration of the concept and implementation of data visiting — a model for secure data sharing where analysis occurs within the data provider's computing environment. The session began with a landscape overview of five organisational models (UK Biobank, the *All of Us* Research Program, Australian Genomics, ELIXIR, and Apheris), focusing on where data resides, how it is accessed, and what can be taken away. Apheris presented their computational governance framework, emphasising technical enforcement of policy and auditability. Discussions highlighted the need for standardised definitions, airlock protocols, policy frameworks, and developer experience enhancements. Participants also called for greater inclusion of research participants and funders in governance, and the importance of aligning legal, technical, and ethical standards globally. A qualitative study to deepen understanding of governance challenges and opportunities was also proposed.

Key takeaways

- **Data visiting models vary widely:** implementation models differ significantly across institutions, but all share a core principle — data remains in place, and researchers or their code visit it. A common tension exists between data custodians' need for control and researchers' need for usability.
- **Standardisation opportunities identified:** participants flagged key areas for GA4GH-led standardisation, including access and egress controls (airlocks), auditability, computational governance frameworks, cross border data transfer, and certification of trusted research environments (TREs).
- **Stakeholder inclusivity is critical:** there was strong consensus that research participants, funders, developers, and private sector actors must all be engaged in governance design to ensure solutions are equitable, scalable, and ethically robust.
- **Sustainable models for funding compute resources are essential**: when compute is hosted by the data provider, there must be mechanisms to allocate and fund those compute resources for external researchers — such as usage-based billing, shared infrastructure agreements, or dedicated research support programs.

Priority action items

- **Launch a multistakeholder landscape study:** conduct a structured qualitative study with researchers, data custodians, developers, participants, and funders to identify experiences with data visiting in different countries and the legal and ethical challenges faced. It was suggested to carry out this study using a deliberative workshop format instead of interviews to avoid over-solicitation of the GA4GH community and to promote community debate.
- **Establish a model data visiting agreement:** initiate working groups to define common technical, legal, and ethical requirements for data visiting — including egress protocols, classification of output data (summary versus identifiable), and cross-border compliance — building toward a model data visiting agreement and certification process for TREs.

Genomic health data ecosystem (HDE) development and maturity level

Recording | Agenda | Slides

Session Chairs: Jordan Lerner-Ellis, Adrian Thorogood

Meeting summary

This session focused on benchmarking national health data ecosystems (HDEs) and the tools available to assess their progress toward integrating genomics into healthcare. Jordan Lerner-Ellis opened with an overview of HDEs and the importance of benchmarking as these systems mature and become critical to national health strategies. Caitlin Howley presented findings from a comparative review of seven national genomics programs, identifying six shared priority areas such as participant engagement, data infrastructure, and ethical integration. Astrid Vicente described the development and application of a comprehensive maturity level model (MLM) used to benchmark national genomics implementation across eight European countries. She highlighted variations across domains such as public awareness, investment, and policy. Anya Kovanda shared Slovenia's experience in developing and implementing genomic medicine into routine care, emphasising a bottom-up approach guided by clinical genetics professionals and contextualised data sharing. The session concluded with a discussion on GA4GH's potential role in promoting uptake of benchmarking tools, adapting models for diverse jurisdictions, and expanding assessments to include societal acceptability and equity.

Key takeaways

- **There is no universal model for a successful HDE, but common priorities** — such as ethical considerations, data infrastructure, and public engagement — can serve as shared benchmarks.
- **The MLM is a practical and scalable tool** for assessing national readiness and progress in implementing genomics in healthcare, with growing adoption across Europe and interest globally.
- **Benchmarking requires both bottom-up expertise and top-down support,** ideally coordinated through multidisciplinary teams and repeat assessments to monitor progress, adapt strategies, and share best practices.

Priority action items

- **Package and disseminate maturity assessment tools** (such as the MLM and Caitlin's review framework) via the GA4GH website and National Initiatives Forum (NIF), along with guidance on how to get started.
- **Host a webinar** explaining how to implement the MLM, showcasing existing country examples and highlighting use cases for longitudinal tracking and policy development.
- **Expand benchmarking domains** to include equity, public acceptability, and better representation in high-resource settings — supported by shared tools like public perception surveys and genomic diversity metrics.

Day 2

Beacon v2.0: updates and insights

Recording | Agenda | Slides

Session Chairs: Jordi Rambla, Michael Baudis

Meeting summary

This session provided a comprehensive update on the Beacon protocol and its various extensions and developments. The meeting began with an overview of Beacon's progress, emphasising its flexibility and its role in both data discovery and delivery. Jordi Rambla highlighted the strength of community-driven implementation across various disease areas, but also acknowledged ongoing challenges such as the lengthy deployment process and blurred lines between specification and implementation. A series of presentations from Scout groups followed, each focusing on advancing specific Beacon features: variant-level query precision (Variant Scout), Boolean and complex logic support in filters (Filter Scout), aggregate queries for dashboarding and faceted search (Aggregation Scout), core specification improvements (Dev Scout), integration with RNA expression data (RNAget), and finally, a proposal from the Genomic Knowledge Standards (GKS) Work Stream to model categorical variation using the Categorical Variation Representation Specification (Cat-VRS). The session concluded with discussion on cross-Beacon interoperability, links to Fast Healthcare Interoperability Resources (FHIR) and the Observational Medical Outcomes Partnership (OMOP) data model, and future road map planning for Beacon v2.5.

Key takeaways

- **Beacon has evolved from simple Boolean queries to a rich, flexible protocol supporting discovery, delivery, and soon aggregated and categorical queries,** thanks to Scout-driven extensions and alignment with community needs (e.g. variant types, filters, and data models).
- **Interoperability with existing health informatics standards such as FHIR, OMOP, and new GA4GH efforts such as Cat-VRS and RNAget is a growing focus,** with interest in enabling richer clinical and expression-based queries while maintaining simplicity and usability.
- **Deployment is technically straightforward but operationally slow due to legal, ethical, and institutional challenges,** which must be factored into any plans for widespread Beacon-based federation or data delivery.

Priority action items

- **Finalise and release Beacon v2.5 in autumn/winter 2025,** incorporating key updates from all active Scouts, especially around filters, aggregation, and variant precision.
- **Initiate concrete efforts to model and integrate clinical data formats (e.g. FHIR) and categorical variation (via Cat-VRS),** with clear documentation and prototype implementations to guide adopters.
- **Clarify specification versus implementation boundaries and improve supporting materials** (e.g. documentation, client/server libraries, and example Beacons) to reduce ambiguity, speed up adoption, and foster more consistent implementations.

Day 2

Towards clinical application of GA4GH products

Recording | Agenda | Slides | Shared notes document

Session Chairs: Mónica Muñoz Torres, Melanie Courtot, Bob Freimuth, Benjamin Berk, Nephi Walton

Meeting summary

This interactive session brought together community members to explore how GA4GH standards can be better integrated into clinical workflows. Through presentations and nine facilitated breakout groups (in person and online), participants discussed clinical use cases, critical gaps, integration challenges, and opportunities to improve adoption. Across diverse geographies and roles — such as clinicians, software vendors, data scientists, and policy experts — key themes emerged around the need for better alignment with existing clinical infrastructure, enhanced outreach to stakeholders, and the development of concrete implementation support tools. The group emphasised actionable steps for bridging the translational divide between genomics research and patient care.

Key takeaways

- **Clinicians and vendors need mature, plug-and-play tools:** there is strong demand for products that are invisible to the end user but deeply embedded into clinical workflows — tools that 'just work." Many clinicians are unaware of GA4GH, and adoption hinges on user-friendly, value-driven implementations with clear return on investment.
- **Interoperability is critical but must be pragmatic:** true interoperability means aligning with dominant clinical systems such as Epic, Health Level Seven (HL7) Fast Healthcare Interoperability Resources (FHIR), and Observational Medical Outcomes Partnership (OMOP). GA4GH should focus on filling genomics-specific gaps and providing APIs, converters, and bridges — not replacing established standards.
- **Success relies on engagement, governance, and incentives:** adoption in clinical contexts requires engaging the right stakeholders — including lab vendors, hospital Chief Information Officers (CIOs), clinical societies, and funders. Without incentives, mandates, or policy levers, integration remains unlikely regardless of technical readiness.

Priority action items

- **Develop and showcase end-to-end clinical use cases:** produce real-world vignettes with accompanying technical workflows, showing how GA4GH products improve patient care, save costs, or enable discovery. These should be easy to share and targeted at diverse audiences (e.g. CIOs, clinicians, and funders).
- **Launch an "implementation starter kit" and playbook:** package key standards (e.g. Variation Representation Specification [VRS], Data Use Ontology [DUO], Passports, and Phenopackets) with example code, sandboxes, decision trees, and one-pagers explaining use, benefits, and what's "good enough" for interoperability.
- **Establish strategic partnerships with clinical and standards bodies:** organise joint meetings or working groups with HL7, American College of Medical Genetics and Genomics (ACMG), European Society of Human Genetics (ESHG), clinical genetics societies, and key electronic health record (EHR) vendors.

Include voices from clinician-researchers and clinical practitioners alike, and advocate for funding requirements that include GA4GH standards.

(hts)Get compliant with it: hackathon edition

Recording | Agenda | Slides

Session Chairs: Deeptha Srirangam, Chen Chen, Jimmy Payyappilly

Meeting summary

This hands-on hackathon session brought together a small but engaged group to collaboratively explore and expand GA4GH's htsget compliance suite. Participants were introduced to the technical infrastructure, including the compliance test suite and testbed report submission process. The team worked through the installation and debugging of the compliance suite, wrote and executed new test cases (e.g. validating the correct data type response in service-info endpoints), and submitted a successful compliance report to the GA4GH testbed. The discussion also touched on the broader adoption of the compliance suite framework, ideas for future engagement (e.g. badges for implementers), and GA4GH's long-term goals around implementation support and standards alignment.

Key takeaways

- **Real-world demonstration of compliance tools:** the session offered a step-by-step example of writing, executing, and submitting a compliance test for htsget, helping attendees understand the full workflow from code to visualisation in the testbed.
- **Barriers to adoption identified:** participants highlighted common installation and dependency challenges that may limit broader uptake, underscoring the need for streamlined tooling and clearer documentation.
- **Desire for broader engagement:** attendees, including newcomers, expressed interest in contributing to compliance suite development. There was strong support for more visible value-adds, such as badges and integration into the GA4GH product approval process.

Priority action items

- **Streamline installation and setup:** address key pain points by simplifying dependencies, improving documentation, and aligning compliance suites (e.g. htsget) with shared libraries such as the GA4GH testbed-lib.
- **Promote broader adoption of compliance suites:** encourage other GA4GH Work Streams to develop or refine compliance suites and integrate testbed usage into the product development lifecycle.
- **Develop contributor onboarding pathways:** create clear, accessible entry points for newcomers — including guides, recommended first GitHub issues, and contact touchpoints — starting with follow-up communications to participants from this session.

Day 2

Beacon for metadata discovery: expanding the scope

Recording | Agenda | Slides

Session Chairs: Jordi Rambla, Michael Baudis

Meeting summary

This session explored whether GA4GH Beacon should expand its scope to better support metadata, particularly in light of real-world demands for discovery and interoperability across sensitive and federated data. Jordi Rambla and Michael Baudis presented the current Beacon structure — its model (data content) and framework (querying system) — and highlighted its blurry boundary between data and metadata. Real-world examples from initiatives like the European Genomic Data Infrastructure (GDI) project, the European Genome-phenome Archive (EGA), and Federated EGA illustrated the challenges of harmonising Beacon with metadata-centric systems like DCAT. Marcos Casado introduced the evolving Federated EGA metadata model, emphasising submission metadata and validation strategies using JSON Schema and BioValidator. David Bujold then presented the GA4GH Experiments Metadata Standard, designed to aid discoverability of genomic sequencing experiments and potentially integrate into Beacon searches. A lively discussion followed around the necessity, scope, and complexity of metadata integration, highlighting tensions between standardisation and flexibility, the importance of provenance, and community readiness to contribute to pilot implementations.

Key takeaways

- **There is strong community interest in expanding Beacon's metadata capabilities,** especially for catalogue-aware and aggregated data, but consensus on scope, terminology, and implementation strategy remains a challenge.
- **Submission and experimental metadata are already being modeled (e.g. by Federated EGA and GDI projects),** offering an opportunity for Beacon to align and avoid duplication — provided interoperability is thoughtfully managed.
- **Community input stressed the importance of use-case-driven development, modularity, and optionality,** especially given diverse needs and limited existing support for metadata completeness and consistency.

Priority action items

- **Define concrete pilot use cases** with Driver Projects to test metadata queries in Beacon, particularly using the Federated EGA and experimental metadata models as reference implementations.
- **Engage metadata stakeholders (e.g. the Data Model and Schema Consensus [DaMaSC] working group, GDI, and ELIXIR) to co-develop minimal, harmonised metadata profiles,** ensuring alignment on high-level discovery fields while preserving deeper semantic flexibility.
- **Refine the Beacon specification to support metadata without compromising existing simplicity,** possibly by introducing optional metadata modules and ensuring clear provenance, ontology mapping, and ethical use indicators.

Day 2

Data access modelling for GA4GH ecosystems

Recording | Agenda | Slides

Session Chairs: Tom Conner, Jaime Delgado

Meeting summary

This session explored the complex interplay between data access policies, privacy, security controls, and real-world use cases within the GA4GH ecosystem. Speakers Tom Connor and Jaime Delgado, alongside participants from NIH, Broad Institute, and others, examined how regulations such as the General Data Protection Regulation (GDPR) and Health Insurance Portability and Accountability Act (HIPAA), along with evolving standards and user expectations, impact technical implementations of consent and access. Key topics included dynamic consent, privacy-preserving API design, provenance metadata, and defining computable access policies. Lively discussion emphasised the balance between automation and human oversight, the challenge of policy heterogeneity across jurisdictions, and the need for both transparency and standard vocabularies to build scalable, trustworthy infrastructures.

Key takeaways

- **Dynamic consent requires technical and process innovation:** implementing dynamic, revocable consent at scale depends as much on streamlining institutional workflows as on technical automation and standardisation.
- **Infrastructure versus dataset governance needs clear separation:** policies often apply at the infrastructure level, not the dataset level, which complicates enforcement and calls for structured vocabularies to describe both.
- **Automation must be transparent, trustworthy, and auditable:** while automation can improve efficiency, it must be paired with clear documentation, human-readable decision logs, and buy-in from policy owners to be acceptable in sensitive settings.

Priority action items

- **Develop a minimal, common vocabulary for dataset access policies:** initiate a collaborative effort (possibly across the Data Use & Researcher Identities [DURI] Work Stream, Regulatory & Ethics Work stream [REWS], and Data Security Work Stream) to define a simple, widely adoptable policy vocabulary for describing data access controls at the dataset level.
- **Prototype tools for ethical and legal provenance tracking:** specify mechanisms to encode legal terms and governance clauses into computable, auditable metadata objects that follow data across platforms.
- **Explore a standardised risk report tool for data access committees (DACs) and institutions:** design a human-readable, standardised security and compliance summary (e.g. SecOps risk digest) to support data access committees (DACs), Chief Information Security Officers (CISOs), and legal teams in evaluating and authorising access infrastructure.

GA4GH data model best practices lunch session

Recording | Agenda | Slides

Session Chairs: Kathy Reinold, Heather Creasy, Karen Cranston

Meeting summary

This session, led by the GA4GH Data Model and Schema Consensus (DaMaSC) working group, focused on advancing best practices for data modelling and ontology use across GA4GH. The team presented a draft set of recommended ontologies, outlined strategies for representing and publishing data models, and sought community input on balancing prescriptiveness and flexibility. Participants engaged in breakout discussions to share real-world experiences, highlighting both successes and challenges with specific data models and schemas, ranging from VCF and Phenopackets to Observational Medical Outcomes Partnership (OMOP) data models, Fast Healthcare Interoperability Resources (FHIR), and Beacon. A key tension emerged between the desire for interoperability via standardisation and the need to accommodate domain-specific flexibility.

Key takeaways

- **Documentation and real-world examples are critical:** across all discussions, clear documentation — including assumptions, real-world data examples, and schema versioning — was seen as essential for adoption and reuse.
- **Interoperability requires balance between prescriptiveness and flexibility:** many advocated for standardising on a small set of formats (e.g. JSON Schema, LinkML); others emphasised maintaining the flexibility to support diverse data types and communities by extension.
- **Ontology use should be guided, not mandated:** participants supported recommending ontologies such as EDAM, OMOP, Systematized Nomenclature of Medicine (SNOMED), especially for newcomers, but stressed the importance of being ontology-agnostic when necessary to support broader data integration.

Priority action items

- **Develop and publish a minimal, prescriptive representation standard:** define a minimal set of required information for GA4GH product compliance, while allowing optional alternatives for broader interoperability.
- **Create a GA4GH data model packaging template:** establish a recommended template for publishing data models that includes purpose, scope, diagrams, versioning, digital object identifiers (DOIs), example datasets, and links to reference implementations.
- **Integrate and expand the ontology recommendations list:** finalise and publish the DaMaSC Ontology Recommendation list with domain context, example use cases, and clear disclaimers that promote reuse while allowing justified deviation.

Day 2

The path to implementation of Regulatory & Ethics Work Stream (REWS) products

Recording | Agenda | Slides

Session Chairs: Vasilki Rahimzadeh, Dianne Nicol, Ma'n Zawati

Meeting summary

The meeting focused on the challenges of implementing GA4GH Regulatory & Ethics Work Stream (REWS) products, highlighting a need for clearer, more practical guidance that can be adapted across legal, cultural, and institutional contexts. Participants emphasised that while the products are valuable, they are often too abstract and not widely known or accessible. Suggestions included developing implementation toolkits, creating more targeted communication strategies (such as podcasts and summary guides), and embedding implementation planning earlier in the product development process. There was strong support for ongoing metrics and feedback mechanisms to better understand and improve real-world uptake.

Key takeaways

- **Need for tailored implementation support:** REWS products are often high-level and abstract, which limits their immediate usability in diverse local or institutional contexts. There's a clear need for practical, context-sensitive guidance and examples for implementation.
- **Challenges with legal interpretation and institutional readiness:** stakeholders often struggle with inconsistent or unclear interpretations of legal requirements (e.g. the General Data Protection Regulation [GDPR]). There is a desire for REWS to be more prescriptive, helping institutions navigate regulatory ambiguity and support confidence in use.
- **Awareness and accessibility barriers:** many stakeholders are unaware of existing REWS products. Improving discoverability, translation (both linguistic and cultural), and using alternate formats (e.g. podcasts, summary guides) could broaden reach and impact.

Priority action items

- **Develop implementation starter guides and toolkits:** create clear, step-by-step documentation that includes use-case-specific implementation strategies, guidance for institutional review boards (IRBS) and ethics boards, and practical templates — particularly for complex products like model consent clauses and data access agreements.
- **Launch metrics and feedback mechanisms:** accelerate the ongoing "metrics group" work to evaluate adoption, usage, and impact of REWS products. This includes the identification of indicators and a set of metrics that can be used to guide uptake of Ethical, Legal, and Social Implications (ELSI) related research.
- **Expand translation and communication channels:** identify translational approaches to adapt products for global use (e.g. language, cultural context, and legal systems). Explore alternative outreach methods such as podcasts, webinars, or region-specific engagement strategies to improve awareness and usability.

Day 2

Cloud Work Stream: new standards

Recording | Agenda | Slides | Transcript

Session Chairs: Michael Lukowski, Michael Baumann, Brian O'Connor

Meeting summary

This session focused on presenting and discussing new ideas for standards or extensions to GA4GH Cloud Work Stream APIs, including well-supported efforts like Portable Format for Bioinformatics (PFB), node and mesh cards for federated data systems, improved support for workflow input/output specification, mechanisms to establish trust between systems sharing sensitive data, and cross-standard awareness. The session included six short presentations (or "pitches") followed by questions, emphasising real-world implementations, championing new work, and connecting related efforts across GA4GH. Presenters included Michael Bauman (PFB), Michael Lukowski (mesh cards), Alex Kanitz (API descriptors), Brian O'Connor (trust networks), Ian Fore (cross-standard alignment), and Pavel Nykamov (attested TLS). Participants discussed where existing standards might already address some needs and stressed the importance of prototyping and collaboration to move ideas forward.

Key takeaways

- **Practical adoption is driving standards forward:** Proven tools such as PFB have demonstrated success across major NIH and GA4GH Driver Projects but lack formal standardisation. Adoption has outpaced governance — now is the time to bring those tools into GA4GH's official scope.
- **Federated ecosystems need discoverability and trust:** both mesh/node cards and the discussion around trust networks highlight a maturing ecosystem where systems and organisations need to discover, evaluate, and safely interoperate with each other. Trust isn't just about users — it's about systems trusting systems.
- **We need more interoperability between standards:** many pitches (e.g. PFB and Schema Registry; mesh cards and Service Registry; test/API descriptors and workflow languages) intersect with existing GA4GH work. Clear communication between Work Streams, and experimentation across them, is crucial to avoid reinventing the wheel.

Priority action items

- **Standardise PFB with GA4GH and align with Schema Registry:** form a working group to draft a GA4GH standard for PFB, evaluating and integrating with Schema Registry, Data Connect, and Research Object Crate (RO-Crate) where appropriate.
- **Prototype mesh/node cards and define trust metadata:** create example mesh cards, simulate mesh formation, and explore how trust metadata (e.g. security protocols, data types, and authZ) can be embedded or verified in practice. Consider aligning with the GA4GH Service Registry.
- **Launch a Cross-Work-Stream hackathon or demo project:** choose a realistic federated use case (e.g. federated machine learning, multi-cloud analysis), and have teams prototype it using cloud standards, mesh cards, API descriptors, Passports, and Schema Registry. Document the process to identify where new standards are needed — or where existing ones can be extended. The suggestion was to start with a hackathon but this may evolve into a GIF project.

The "TASC-athon"

Recording | Agenda | Slides

Session Chairs: Andy Yates, Mamana Mbiyavanga

Meeting summary

The TASC-athon session brought together GA4GH developers, staff, and contributors to collaboratively address the backlog of open issues in the Technical Alignment Subcommittee (TASC) GitHub repository, aimed at improving technical alignment across Work Streams. After a brief overview of the purpose and process of TASC, participants formed breakout groups (both in-person and virtual) to assess, refine, or close specific issues. Key themes included how to manage GA4GH namespaces, documentation tooling, GitHub repository organisation, cross-Work-Stream coordination, and open-source licensing. The session resulted in the closure of outdated issues, clearer paths forward for persistent blockers, and momentum for restructuring GitHub governance under a more formalised set of standards and best practices.

Key takeaways

- **Need for a centralised and clear GitHub governance framework:** discussions revealed a strong demand for standardised GitHub workflows, contribution guidelines, API specifications, and licensing recommendations across GA4GH projects to ease onboarding and maintain consistency.
- **Issue triage and maintenance must be more active:** several issues dating back to 2020 were deemed irrelevant or solved through other initiatives, highlighting the importance of regular backlog grooming and clearer ownership models.
- **Lightweight, actionable sub-issues can drive progress:** large or blocked issues (e.g. namespace resolution) were broken into smaller technical tasks — demonstrating that even long-stalled efforts can be revived with structured planning and new ownership.

Priority action items

- **Close or consolidate tale and redundant issues:** officially close low-priority or outdated issues (e.g. Issue #12 on OpenAPI HTML generator and Issue #40 on repository splitting) and consolidate others into active broader initiatives like the GitHub harmonisation project.
- **Formalise GitHub standards and create a maintained Style Guide:** develop and publish a GA4GH GitHub Style Guide encompassing branching, tagging, change logs, API specification locations, documentation best practices, and license recommendations. Consider creating a working group to own this effort.
- **Implement GA4GH namespace resolution service:** advance Issue #16 by implementing a simple resolver for GA4GH-computed identifiers (e.g. for Variation Representation Specification [VRS] objects), linking to documentation, and registering patterns with identifiers.org. Begin with a lightweight HTML-based v1.0 and expand over time.
- **Prioritise development of terms of reference for TASC:** ensure broader representation and clarify the role and remit of TASC to help better prioritise and potentially resolve issues.

Day 2

Future of the Data Connect standard

Recording | Agenda | Slides

Session Chairs: Miro Cupak, Jonathan Fuerth

Meeting summary

The GA4GH Data Connect standard enables data custodians to expose structured and semantically described data tables without prescribing a particular data model, and data consumers to search this data using a flexible query language. The session emphasised its flexible adoption model, allowing for gradual semantic enrichment of data. Participants identified priorities for future versions of the standard, including improved interoperability with other GA4GH APIs, more sophisticated access control (e.g. integration with Passports), and foreign key–like table linkages. Use cases and implementation challenges were shared by a diverse group of stakeholders, including those using Data Connect for variant-level matching, cohort discovery, and downstream tool interoperability. The group collectively voted on high-priority areas for development.

Key takeaways

- **Flexible semantics + practical querying = broad adoption potential:** Data Connect provides value at all stages of data sharing maturity — by allowing data custodians to semantically describe existing structures using JSON Schema and enabling searching viaSQL queries.
- **Growing need for cross-standard integration:** there is strong interest in integrating Data Connect with other GA4GH APIs, particularly the Data Repository Service (DRS), the Workflow Execution Service (WES), and Passports, to enable cohesive end-to-end workflows. Participants expressed a desire for clearer guidance to support such interoperability.
- **Developer support and tooling are just as important as specification development:** community feedback underscored the importance of supporting implementations through documentation, reference implementations, and utilities like JDBC drivers or Python clients. This support is as crucial as evolving the specification itself.

Priority action items

- **Advance Integration with other GA4GH standards (DRS, WES, and Passports):** align Data Connect with DRS (e.g. standardised inclusion of drs_ids in tables), define access using Passports, and explore practical workflows that link Data Connect queries to WES jobs. Clarify access token handling and consider future use of smaller, task-specific tokens.
- **Enable table linking and fine-grained access control:** develop native support for linking across tables (e.g. foreign-key–like constructs) and explore standardised patterns for fine-grained access control (e.g. column- or row-level restrictions). Capture and document implementation best practices from existing deployments.
- **Expand documentation, tooling, and adoption resources:** build out example integrations (e.g. with Phenopackets) and develop utility libraries (e.g. Python or JDBC drivers). Support developers in connecting Data Connect to real-world data environments and workflows.

Day 2

Experiments Metadata Standard: now and beyond

Recording | Agenda | Slides

Session Chairs: David Bujold, Peter Woollard

Meeting summary

This session provided a comprehensive overview of the Experiments Metadata Standard group's work on defining a core checklist of metadata fields to describe genomic experiments, with the aim of enhancing data discovery and interoperability. David Bujold introduced the session, highlighting the project's focus on experimental methods and equipment rather than full protocol reproducibility. Peter Woollard walked through recent improvements, including ontology-based refinements, new identifier fields, and revised property documentation. Attendees provided feedback on batch effect capture, sample quality metrics, and ontology term availability, with Charlie Barclay outlining how the community can request missing terms. Future plans include extending the checklist to specific experiment types (e.g. transcriptomics, single-cell), implementing schemas for integration, encouraging broader adoption, and launching the GA4GH product approval process by July. The session closed with a summary of next steps.

Key takeaways

- **Core metadata checklist finalised:** the group has developed a draft metadata checklist focused on experiment-level properties, emphasising ontology-based entries (e.g. for instrument models and experiment types) to improve discoverability and interoperability.
- **Community feedback is essential:** input from GA4GH Driver Projects and domain experts (e.g. in cancer and transcriptomics) is critical for refining checklist fields, ensuring ontology coverage, and expanding into more specific experimental categories.
- **GA4GH product approval process underway:** the group aims to launch the open for comment period by mid-July to seek formal GA4GH approval, with a target to achieve product status by the October 2025 Plenary.

Priority action items

- **Solicit broader community review:** distribute the draft checklist and feedback forms beyond the Discovery Work Stream (e.g. via mailing lists and individual outreach) to gather input from additional Driver Projects and domain experts.
- **Prepare for open for comment period:** finalise supporting materials (e.g. implementation notes, use cases, and decision records) required to begin the GA4GH product approval process by the July 14 deadline.
- **Expand domain-specific coverage:** recruit contributors with expertise in high-priority experiment types (e.g. single-cell, arrays, targeted sequencing) to begin populating relevant metadata fields and ontology mappings for those categories.

Day 2

refget Sequence Collections standard

Recording | Agenda | Slides

Session Chairs: Nathan Sheffield, Timothee Cezard, Andy Yates

Meeting summary

This session provided a comprehensive overview of the newly approved refget Sequence Collections standard, its motivations, current and planned implementations, and future directions including support for pangenomes. The presenters detailed the structure and purpose of the standard — creating verifiable, unique identifiers for sets of sequences such as reference genomes — and walked through its encoding process, API endpoints, and use cases. Key implementers from Ensembl, European Variation Archive (EVA), ELIXIR Norway, and others shared real-world applications including validating VCFs, managing annotations, and future integration with tools and file formats. The group discussed challenges such as long-term server sustainability, integration into existing formats (e.g. BAM and VCF), and potential for hierarchical or nested sequence collections. The session concluded with an open discussion focused on adoption pathways and collaborative next steps.

Key takeaways

- **Refget Sequence Collections enable unambiguous identification and comparison of reference genome builds** — including their names, sequences, and order — through a standardised digest and API interface.
- **Widespread implementation is underway,** including by EVA, Ensembl, and ELIXIR, with support for validation of user-submitted data and potential use in annotations and pangenome structures.
- **Sustainability and interoperability are key concerns,** especially ensuring multiple mirrored servers, integration with existing bioinformatics formats, and adoption by major toolkit providers.

Priority action items

- **Establish additional public servers and mirrors for sequence collections,** ideally hosted by multiple institutions or through a cloud or content delivery network (CDN) infrastructure, to ensure redundancy and reliability.
- **Pilot inclusion of sequence collection digests in common genomic file formats** (e.g. BAM and VCF) and coordinate with toolkit maintainers (e.g. HTSlib and Picard) to enable interpretation and retrieval.
- **Advance the development of pangenome support** by formalising how collections of collections (hierarchical sequence collections) are digested, compared, and retrieved, especially for alternate loci use cases.

Genomic Knowledge Standards (GKS) Work Stream road map

Recording | Agenda | Slides

Session Chairs: Alex Wagner, Melissa Cline, Larry Babb

Meeting summary

This session provided an in-depth look at the GA4GH Genomic Knowledge Standards (GKS) Work Stream, with updates from each of its product development teams — Variation Representation Specification (VRS), Categorical Variation Representation Specification (Cat-VRS), and Variant Annotation Specification (VA-Spec) — as well as the Sequence Annotation Study Group. The discussion highlighted recent achievements, including the release of VRS v2.0, the Sequence Annotation Study Group's recommendations for future work, and the current community review phase of Cat-VRS and VA-Spec. A consistent theme throughout the session was the importance of user-centered development, cross-specification alignment, and implementation support. The latter portion of the session was dedicated to an open discussion where community members emphasised the need for better documentation, implementer guidance, and shared infrastructure across GA4GH products. Future hackathons, tooling development, and a focus on education and documentation were also discussed.

Key takeaways

- **User-driven implementation is a top priority:** all GKS product teams emphasised the need to focus on growing adoption and usability for real-world implementers. This includes ensuring the availability of sample datasets (e.g. ClinVar and MaveDB), developing reference implementations, and improving ease-of-use tooling like "VRS in a Box."
- **Documentation and onboarding need major improvements:** there was broad consensus that GKS documentation is too complex for new implementers and calls for clearer "getting started" materials, harmonised schemas, and unified entry points to understand how the specifications are connected.
- **Cross-specification alignment and collaborative governance are emerging needs:** As GKS products become more interdependent (e.g. VA-Spec referencing VRS and Cat-VRS), participants urged for structured processes to manage versioning, compatibility, and possibly organisational-level conformance testing and documentation tooling.

Priority action items

- **Develop centralised, user-Friendly documentation and examples:** create "starter kits" and dynamic Jupyter Notebooks across VRS, Cat-VRS, and VA-Spec; establish a GKS-wide landing page explaining products and how they interact; improve schema-level documentation and tooling for implementer-specific documentation reuse.
- **Launch targeted implementer-focused hackathons and support:** organise a dedicated "implementer hackathon" (e.g. at an upcoming Epic or GA4GH meeting) to help teams model and validate their data against GKS specifications; provide 1:1 mentoring, "office hours," and peer-to-peer onboarding sessions.
- **Establish processes for product interoperability and validation:** coordinate with other GA4GH Work Streams (e.g. Phenopackets and Beacon) to harmonise dependencies and encourage schema reuse; explore a lightweight conformance framework or "verified implementation" badge system to help users ensure and demonstrate compliance with GKS specifications.

Day 3

Brainstorming technical next steps for the Pathoplexus pathogen genomic data sharing ecosystem

Recording | Agenda | Slides

Session Chairs: Danny Park, Anderson Brito, Theo Sanderson, Anna Parker

Meeting summary

This GA4GH Connect session focused on showcasing Pathoplexus, a pathogen data platform built to address unmet needs in pathogen genomics, with special emphasis on usability, ethical data sharing, and open governance. Pathoplexus and its underlying open-source software, Loculus, were introduced via presentations. The session also explored integration opportunities with GA4GH standards, including the Data Repository Service (DRS) for data access, Data Use Ontology (DUO) for data use restrictions, and Beacon for federated discovery. Discussion covered real-world use cases, such as managing restricted-use data with publication restrictions, metadata challenges (especially geolocation), and future needs for federated pathogen genomics infrastructure. Participants from GA4GH, Public Health Alliance for Genomic Epidemiology (PHA4GE), and the European Genome-phenome Archive (EGA) highlighted areas for standardisation, technical alignment, and community-driven next steps.

Key takeaways

- **Pathoplexus fills critical gaps in pathogen data sharing** by enabling easier, more ethical, and rapid submission of viral genomes with structured metadata, including temporary data use restrictions that encourage earlier data release.
- **There are potential opportunities to deploy GA4GH standards in Pathoplexus,** with possible products including DRS (for persistent identifiers and data access), DUO (for modeling usage restrictions, though current terms are likely not directly suitable), and Beacon (for federated queries across systems).
- **Future federation and scale may make GA4GH standards relevant** — both technically (e.g. persistent IDs, scalable storage, and data curation tools) and strategically (e.g. aligning metadata models and coordinating among multiple Loculus instances and external resources such as the International Nucleotide Sequence Database Collaboration [INSDC]).

Priority action items

- **Prototype DRS integration for Pathoplexus,** providing a compliant interface for viral genome files, especially as raw read data is added, which could be directly used by platforms like Terra.
- **Consider whether Pathoplexus data restrictions could be modelled with DUO,** identifying whether existing DUO terms could model some aspects of the restricted-use terms, and the extent to which terms could be added to better model this.
- **Consider whether Beacon could support federated querying across Loculus instances,** acting as a standards-compliant API to support scalable pathogen data discovery, which would be especially impactful if other international resources adopted Beacon.

Day 3

Passports, visas, and task-specific tokens

Recording | Agenda | Slides

Session Chairs: Tom Conner, Surya Saha

Meeting summary

This session of the GA4GH Data Security Work Stream focused on the current state and implementation of GA4GH Passports, particularly around authentication and authorisation workflows for controlled access datasets. Presentations explored different architectural models using Researcher Auth Service (RAS) Passports, including a working prototype from Lister Hill that interfaces with the database of Genotypes and Phenotypes (dbGaP) Fast Healthcare Interoperability Resources (FHIR) server, and implementations using GA4GH Data Connect. Demonstrations highlighted how task-specific tokens derived from Passports can be used in real-world environments, while discussions focused on when and how to transmit Passports, performance tradeoffs (especially with large datasets and pagination), and evolving the model to better support attribute-based access control and potential work order tokens. The breakout discussions surfaced issues around interoperability, the limits of the current Passports specification, and ideas for a potential 2.5 or 3.0 version of the standard.

Key takeaways

- **The current GA4GH Passports specification is widely adapted but often stretched:** implementations are common but typically involve workarounds or patches, suggesting it is time to revisit the standard for improved clarity, usability, and fit across use cases.
- **Performance can be significantly impacted by how Passports is used:** for example, repeated validation via a clearinghouse on every pagination request in FHIR can add substantial overhead; thoughtful decisions on caching or token lifecycle are critical.
- **Passports alone is not sufficient for full access control decisions:** real-world authorisation requires additional contextual attributes (action, resource availability, and environment).

Priority action items

- **Kick off a structured assessment of current Passports deployments** to identify pain points and drive design of a v2.5 or v2.0 specification, including input from groups such as the Data Repository Service (DRS) team and the Cloud Work Stream as a whole, the Data Connect team, and Driver Projects.
- **Develop guidance or profiles for standardising Passports use across various implementations,** including recommendations for managing large payloads, token passing, and clearinghouse interactions across pagination.
- **Explore formalising work order tokens** as a complementary model that incorporates user attributes from Passports along with action, resource availability, and environment parameters to enable fine-grained, scalable, and efficient access control of data as well as compute.

Day 3

Standards for enabling federated machine learning

Recording | Agenda | Slides

Session Chairs: Kyle Ellrot, Alexander Kanitz, Venkat Malladi and Ravi Madduri

Meeting summary

This session was an exploratory and discussion-based workshop focused on enabling federated machine learning (FL) in the context of genomic and health-related data. There is no existing GA4GH product for FL, and the session aimed to gather input, identify stakeholders, and assess whether GA4GH can and should contribute standards or coordination. Presentations outlined the promise and pitfalls of FL, showcased various FL frameworks (e.g. NVIDIA FLARE, OpenFL, APFL), and emphasised the technical, ethical, and logistical challenges of data heterogeneity, privacy risks, model governance, and computational disparity across sites. The discussion featured a rich set of use cases, ranging from Parkinson's progression (ENIGMA) and diagnostic odyssey prediction (Epic) to microbiome modelling and clinical trial matching. The group concluded that GA4GH is well-positioned to support FL implementations by leveraging existing standards and coordinating cross-framework harmonisation, especially through a new GA4GH Implementation Forum (GIF) Project.

Key takeaways

- **Federated learning is a promising but complex opportunity for GA4GH:** FL offers a path to privacy-preserving, generalisable AI models using decentralised health and genomic data — but requires coordinated effort on data harmonisation, workflow orchestration, and model governance.
- **Existing GA4GH standards are a strong foundation:** technologies like the Task Execution Service (TES), Data Repository Service (DRS), Passports, Data Use Ontology (DUO), and Workflow Execution Service (WES) are well-suited to support FL workflows, especially for model execution, security, access control, and artifact sharing.
- **A real-world use case–driven approach is essential:** rather than standardising in the abstract, the group agreed to anchor the effort in two to three concrete use cases, such as microbiome-based disease modelling, Parkinson's progression from imaging-genetics, and diagnostic decision support, and build a minimal viable ecosystem around those.

Priority action items

- **Launch a federated machine learning GA4GH Implementation Forum (GIF) Project:** coordinate technical contributors, Driver Projects, and Work Streams to iteratively develop and test FL-enabled pipelines using existing GA4GH standards and real data.
- **Define minimum viable tooling and infrastructure requirements:** create a working document or white paper that outlines the minimal GA4GH stack needed (e.g. TES, WES, DRS, Passports, and metadata schemas) to support FL use cases across heterogeneous environments.
- **Engage stakeholders and build community:** grow the conversation through the newly opened #federated-ml Slack channel, collect contact information from session participants, and set up regular coordination calls to maintain momentum and prepare deliverables for the next GA4GH meeting.

Day 3

Visa encoding in verifiable credentials data model v2.0

Recording | Agenda | Slides

Session Chairs: Martin Kuba, Jaime Delgado

Meeting summary

This session explored the potential for extending the GA4GH Passports visa model to align with the World Wide Web Consortium (W3C) Verifiable Credentials (VC) data model. Martin Kuba presented the current structure and limitations of GA4GH visas — namely their inability to express complex conditions or bind securely to the presenting entity — and proposed using VCs to overcome these. He provided a technical deep dive into VCs, digital wallets, RDF/JSON-LD, and ontologies (e.g. Data Use Ontology [DUO] and Mondo Disease Ontology [MONDO]), demonstrating how these could enhance expressiveness and trust in access control mechanisms. The group discussions centered on the need for more specific role designations in visas, challenges around interpreting researcher context, and whether shifting to VCs would yield real-world benefit. Participants emphasised the importance of identifying concrete use cases before committing to development, with several expressing interest but uncertainty over the immediate applicability.

Key takeaways

- **Current GA4GH visas lack expressiveness:** they cannot easily represent complex conditions such as time-bound access or research context (e.g. disease-specific or non-profit use), nor are they cryptographically bound to a user or device.
- **VCs offer potential improvements:** VCs can support selective disclosure, richer semantics via ontologies, and cryptographic binding to user-controlled wallets — enabling more precise, interoperable, and secure representations of user attributes.
- **Adoption requires clear demand and use cases:** there is interest in exploring this further, but consensus was that development should not proceed without commitment from real-world implementers (e.g. Driver Projects or NIH/ELIXIR-backed services) to test and validate the model.

Priority action items

- **Identify concrete use cases** from Driver Projects or federated access scenarios (e.g. cross-continental data sharing) that are blocked or limited by current Passport/visa models.
- **Develop a proof of concept** that maps existing GA4GH visas to VCs using DUO or MONDO ontologies, and demonstrate how they could express complex policies.
- **Host focused follow-up sessions** within the Data Security Work Stream to assess interest, refine technical scope, and align on next steps for potential pilot implementations.

Day 3

Exploring AI-driven innovation: benchmarking large language models for rare disease diagnostics

Recording | Agenda | Slides

Session Chairs: Benjamin Berk, Moni Muñoz Torres, Nicole Vasilevsky, Donrich Thaldar

Meeting summary

This highly engaging session, co-led by Monica Muñoz-Torres, Nicole Vasilevsky, Benjamin Berk, and Donrich Thaldar, explored the role of large language models (LLMs) in aiding rare disease diagnostics using GA4GH standards like Phenopackets. The team shared a demo of converting clinical notes to structured Phenopackets and using LLMs to generate differential diagnoses. The session addressed the challenges of benchmarking LLM performance, including dataset representativeness, gold standard answer creation, evaluation metrics, and legal and ethical implications. Participants split into breakout groups to explore these aspects in depth, emphasising the need for expert oversight, data governance, and realistic benchmarks. There was a strong sense of momentum toward building a shared strategy for trustworthy and effective use of LLMs in clinical genomics.

Key takeaways

- **Benchmarking LLMs for rare disease diagnosis is both a technical and ethical imperative** — requiring diverse datasets, expert-reviewed answer keys, and carefully chosen evaluation metrics to ensure clinical safety and utility.
- **Structured data formats like Phenopackets remain crucial,** even in an LLM-enabled future, providing consistent, computable inputs for model evaluation and clinical integration.
- **Legal, ethical, and human factors are central:** considerations include liability, model explainability, user accountability, and patient trust in LLM-assisted care, especially in low-resource settings.

Priority action items

- **Create a benchmark dataset** for LLM evaluation in rare disease diagnosis, including diverse and representative Phenopackets, expert-reviewed gold standards, and known challenging cases.
- **Define a shared evaluation framework** with clearly documented metrics, thresholds, and methodologies to assess LLM outputs, accounting for model uncertainty and clinical applicability.
- **Develop GA4GH guidance on responsible LLM use,** including ethical and legal frameworks, data governance protocols, and community-driven recommendations for transparency and trust.

Day 3

Data Repository Service (DRS) working session

Recording | Agenda | Slides

Session Chairs: Michael Lukoski, Brian O'Connor

Meeting summary

This interactive session on DRS (Data Repository Service) focused on discussing practical enhancements to the specification, rooted in real-world implementation feedback and emerging use cases. The discussion covered three core topics: adding cloud and geographic constraints to data access methods, establishing trust relationships between compute environments and DRS servers, and better communicating data availability via cold storage indicators. The first segment explored constraints on data movement, especially for jurisdictions with strict regional or cloud-provider restrictions. The second part centered on trust models and how to securely and scalably share access credentials between platforms. The final discussion tackled how DRS can better signal data accessibility when items are in cold or delayed-access storage. Throughout, attendees emphasised cross-Work-Stream collaboration, keeping the specification backwards-compatible, and the importance of balancing policy signalling with technical feasibility.

Key takeaways

- **Cloud and regional constraints need representation — but not enforcement — in DRS:** implementers need a way to indicate location- or provider-based data access constraints, such as "AWS-only" or "must remain in countries or cloud regions," but these should be non-binding metadata, not enforced rules, due to legal/implementation variability and complexity.
- **Scalable trust between systems is essential:** As cloud ecosystems grow, there's a pressing need for scalable, flexible mechanisms to establish trust between DRS servers and compute platforms. Ideas ranged from mutual TLS and static lists to verifiable credentials and attestation-based models.
- **Cold storage status should be explicit and actionable:** there is strong support for DRS to communicate data availability beyond a simple Boolean — potentially including delay durations, last-known status, and rehydration status — to help platforms optimise job scheduling and user feedback.

Priority action items

- **Advance and finalise cloud constraint proposal in DRS:** revise and socialize the draft pull request (PR) by adding usageConstraint fields to DRS access methods, ensuring optional backwards compatibility, and evaluating integration with existing vocabularies (e.g. Data Use Ontology or related ontologies).
- **Convene a Trust working group:** form a team within the Cloud Work Stream to draft an RFC (request for comment) outlining possible trust mechanisms (e.g. verified lists, mutual TLS, and verifiable credentials). Ensure participation from DRS implementers, compute platforms, and security experts.
- **Define and prototype cold storage availability extensions:** propose a new metadata model for availability in DRS v1.6+, possibly including enum values (e.g. "immediate," "delayed," and "offline") and estimated rehydration times. Coordinate testing through hackathons to assess real-world feasibility.

Day 3

Finding synergy: TASC and DaMaSC

Recording | Agenda | Slides

Session Chairs: Andy Yates, Kathy Reinold and Mamana Mbiyavanga

Meeting summary

This joint session brought together members of the GA4GH Technical Alignment Subcommittee (TASC) and the Data Modeling and Schema Consensus group (DaMaSC) to clarify their respective roles, explore synergies, and align on shared objectives. The DaMaSC team outlined its focus on best practices for data representation and the development of technical specifications (e.g. schema registry), while TASC described its role as a clearinghouse for technical alignment, with a remit for recommendations but not products. A robust discussion followed on governance, lifecycle management, product versus recommendation definitions, registry ownership, and the importance of harmonising GA4GH standards in service of real-world utility and efficiency. Participants also emphasised the need for clarity, community engagement, and formal mechanisms for prioritisation, deprecation, and feedback. The session ended with a shared commitment to closer integration, with DaMaSC functioning as a group within TASC, similar to product teams in other GA4GH Work Streams.

Key takeaways

- **Formal integration and clear roles:** DaMaSC will function as a technical working group within TASC, following the pattern of product teams within other Work Streams. It will offer a focused, agile team to advance interoperability recommendations and products around data representation and sharing.
- **Schema Registry Specification ≠ Implementation:** DaMaSC is developing a specification for a schema registry — not its implementation or governance. TASC (and possibly GA4GH leadership) will determine ownership, policy, and operational governance of any official GA4GH registry.
- **DaMaSC will continue to develop recommendations and products around data representation**, adding GitHub issues to TASC's repository and presenting initiatives to TASC for input and collaboration.
- **Need for broader harmonisation and strategic alignment:** participants stressed the importance of aligning technical efforts (e.g. schema development, deprecation, and lifecycle) with broader GA4GH goals, such as usability, interoperability, and measurable impact — while also calling for analytic tools to track product usage and better connect with initiatives such as the GA4GH Implementation Forum (GIF).

Priority action items

- **Create and maintain GitHub issues for DaMaSC work in the TASC repository:** DaMaSC will create GitHub issues in the TASC repository to document active initiatives, support transparency, and facilitate coordination and feedback across GA4GH.
- **Establish regular representation and reporting at TASC meetings:** DaMaSC will identify primary and secondary representatives to attend TASC meetings regularly and report on progress. This ensures bi-directional communication and alignment on governance, process, and integration points.
- **Clarify governance and ownership of Schema Registry implementation:** TASC, in consultation with the GA4GH tech team and GA4GH leadership, should initiate a formal decision-making process to determine if and how GA4GH will host a Schema Registry, who will own and maintain it, and how submission, review, and deprecation processes will be governed.

Day 3

Collaborative brainstorming for the next phase of GA4GH Implementation Forum (GIF) activities

Recording | Agenda | Slides

Session Chairs: Mallory Freeberg, Neerjah Skantharajah

Meeting summary

This session, led by Mallory Freeberg, provided a comprehensive update on the rebooted GA4GH Implementation Forum (GIF), including a new charter focused on implementation, interoperability, and inclusivity. It featured an overview of recent and planned activities — such as the GIF "Ask Me Anything" (AMA) program and the onboarding of new GIF Projects — followed by four flash talks from current and newly accepted GIF Projects (Federated Variant-Level Matching [VLM], eLwazi, BRCA Exchange, and the Trusted Research Environment [TRE] Open Suite). The session wrapped up with discussion around improving clarity on the distinction between GIF Projects and AMA events, ideas for integrating multiple standards into single implementation discussions, and proposals for including hackathons and office hours as part of GIF-supported activities. Feedback was gathered through a live poll to prioritise upcoming GIF activities.

Key takeaways

- **GIF has shifted focus toward community interaction and implementation support,** especially through AMA sessions, which flip the usual presentation-heavy model of GA4GH-related meetings to an interactive Q&A format that connects standard developers with implementers.
- **Clearer distinction is needed between GIF Projects and other GIF initiatives** (e.g. AMA events), as the terminology and structure can be confusing for attendees unfamiliar with the broader context.
- **There is strong community interest in more hands-on formats like hackathons and office hours,** particularly for multi-standard use cases, federated implementations, and collaborative problem solving during GA4GH meetings.

Priority action items

- **Clarify GIF branding and communications,** especially the difference between GIF Projects and AMA sessions, possibly reconsidering terminology like “AMA” to better reflect a global and inclusive audience.
- **Pilot hackathons or “office hours” at upcoming GA4GH meetings,** possibly under GIF coordination, to support real-time implementation challenges, peer learning, and mentoring.
- **Scale up GIF AMA sessions,** including multi-standard themes and deeper integrations with implementation communities, while expanding resources (e.g. best practices, recordings, and Q&A documentation) for broader asynchronous access.
- **Prioritised activities from the audience poll include**:
    - raise awareness of current GA4GH implementation successes and impact (62.5%);
    - raise awareness of existing GIF Projects and how to get involved (58.3%);
    - host a session at every GA4GH Connect meeting (54.2%);
    - provide GIF updates at every GA4GH Plenary (54.2%).

VLM + VRS = NextGen Beacon

Recording | Agenda | Slides

Session Chairs: Alex Wagner, Jordi Rambla, Michael Baudis, Larry Babb

Meeting summary

This working session explored the integration of GA4GH VRS (Variation Representation Specification) into the Beacon framework to create a more robust, interoperable query system — referred to as “NextGen Beacon.” Co-led by Alex Wagner, Larry Babb, Michael Baudis, and Jordi Rambla, the session focused on designing a scalable, federated network for variant-level matching (VLM) using standardised query formats and data models. Presentations and live discussion covered advances in VRS v2.0, the proposed Variant Query Schema (VQS), real-world implementation challenges, and interest from data providers in low-lift pilot participation. Several participants shared that they already store VRS-formatted data and expressed willingness to pilot a lightweight Beacon setup if toolkits or middleware were made available. Others emphasised the need to balance standardisation with usability, calling for flexible front ends that can accept familiar formats while translating to VRS-compliant queries under the hood. The group largely aligned on a phased approach: start with exact variant matching using VRS IDs, pilot with real datasets, and incrementally scale to support fuzzier and more complex queries.

Key takeaways

- **Consensus to anchor Beacon queries on VRS:** participants generally agreed that aligning Beacon query syntax and variant representation with VRS v2.0 is the ideal long-term goal, provided sufficient tooling and backwards compatibility are in place.
- **Immediate need for lightweight pilots:** there was strong interest in piloting a minimal version of a VRS-enabled Beacon, especially among groups who already have data in the VRS format but lack the resources to build custom endpoints.
- **Separation of concerns is critical:** the group emphasised that human-readable input formats and front-end usability must be decoupled from the technical representation of variants passed over the wire (e.g. via JSON/VRS), to avoid losing engagement from end users.

Priority action items

- **Launch a pilot project using “AnyVar” Toolkit:** identify one to three data providers (e.g. MaveDB and BRCA Exchange) to participate in a pilot that loads VRS-formatted data into the AnyVar tool and exposes it via a minimal Beacon-compliant API for exact variant matching.
- **Develop and share reusable middleware:** build middleware that can take common input formats (e.g. HGVS strings and protein-level descriptions) and convert them into VRS IDs or objects suitable for Beacon query submission.
- **Draft a simple phonebook/metadata profile for nodes:** create a shared schema for Beacon node metadata ("phonebook") that documents query capabilities, supported query types, and data content — enabling network-level discovery and federation.

Day 3

Regulatory & Ethics Work Stream (REWS) in focus: bridging regulatory and ethical gaps across GA4GH

Recording | Agenda | Slides

Session Chairs: Vasilki Rahimzadeh, Dianne Nicol, Ma'n Zawati

Meeting summary

This session focused on the GA4GH Regulatory and Ethics Work Stream (REWS) product development road map. The discussion began with presentations showcasing different types of REWS products — policy frameworks, tools, and resources — developed in response to community-identified needs. Case studies included the Diversity in Datasets policy, the Model Data Access Agreement clauses, a terminology resource for data visiting, and early efforts on responsible generative AI tools. The interactive portion engaged participants in exploring how needs for REWS products are identified, communicated, and resourced, and how to improve coordination with other GA4GH Work Streams. Key themes included the challenge of funding empirical research, ensuring contextual adaptability (especially for low-resource settings), clarifying REWS's role in legal and ethical advising, and embedding REWS earlier and more consistently in technical product development.

Key takeaways

- **REWS products fill essential ethical and legal gaps — but must be more adaptable:** participants highlighted the need for localised versions or contextualisation, especially for under-resourced regions, where assumptions about infrastructure or legal frameworks may not hold.
- **Cross-Work-Stream collaboration and early involvement are crucial:** REWS input is often needed at the ideation stage of technical products — not just at the review phase — but current pathways to collaboration are inconsistent and often dependent on informal relationships or personal initiative.
- **Funding and recognition of effort remain structural barriers:** empirical research necessary for REWS product development is often unfunded. Participants called for clearer guidance on how to propose REWS products requiring new research and emphasised the need for compensated roles for policy contributors.

Priority action items

- **Improve visibility and accessibility of REWS guidance and process documentation:** update the website to include a clear, detailed road map of REWS product development and explicitly list contact points and engagement pathways.
- **Create mechanisms for proactive and structured cross-Work-Stream engagement:** consider formally embedding REWS representatives in other Work Streams and developing lightweight consultation models to surface ethical, legal, and social implications (ELSI) earlier.
- **Develop regional adaptation resources and clearer guidance on legal limitations:** provide model product adaptations or use cases from diverse contexts, and clarify the distinction between academic ethical advice and legal counsel in documentation and communications with technical groups to manage expectations.

Adapting GA4GH standards to meet the needs of the Neuroscience Community

Recording | Agenda | Slides

Session Chairs: Francis Jeanson, Randy McIntosh

Meeting summary

This session marked the one-year anniversary of the GA4GH Neuroscience Community of Interest, highlighting progress made by each of the three theme teams — Scientific Collaboration and Education, Data Governance, and Data Harmonisation — toward promoting global collaboration and responsible data sharing in neuroscience. Presentations and breakout discussions focused on aligning neuroscience data practices with GA4GH standards, emphasising the importance of interoperability, ethical governance, community engagement, and security. Central to the conversation was the challenge of balancing generic standards with the specific needs of neuroscience and mental health research, and how to foster adoption through clear incentives, use-case driven development, and integration with existing efforts.

Key takeaways

- **Shared challenges and synergies:** while many technical and governance needs are not neuroscience-specific, developing focused use cases within the domain (e.g. Parkinson's and mental health) is critical to driving adoption and aligning with existing standards such as Passports, Phenopackets, and Health Level Seven (HL7) Fast Healthcare Interoperability Resources (FHIR).
- **Governance and consent are top priorities:** breakout discussions reiterated ongoing challenges around consent language, cross-jurisdictional legal frameworks, and the need to address ambiguous distinctions between anonymisation and de-identification, especially for sensitive data like mental health.
- **Collaboration is critical to avoid reinvention:** both theme leads and participants emphasised the value of building on existing consortia (e.g. Enigma, Brain-CODE, and Accelerating Medicines Partnership [AMP] initiatives) and working closely with GA4GH Work Streams (especially the Data Security and Data Use & Researcher Identities [DURI] Work Streams) to evolve and apply existing tools and standards rather than creating new ones from scratch.

Priority action items

- **Launch a Slack channel for ongoing coordination:** as suggested by participants, a dedicated Slack space will help maintain momentum, allow easy updates, and foster cross-theme communication between meetings.
- **Develop a harmonisation landscape and road map:** the Data Harmonisation Theme Team will continue cataloguing existing ontologies and standards in use across neuroscience projects, identifying overlaps and gaps, and proposing a path toward interoperability.
- **Engage with the GA4GH Data Security, Data Use & Researcher Identities (DURI), and Regulatory & Ethics (REWS) Work Streams:** begin formal discussions to explore how Passports, Authorisation and Authentication Infrastructure (AAI), and REWS policies and frameworks can be adapted or extended for neuroscience use cases, including potential contributions to new or updated GA4GH standards based on neuroscience community needs.

Day 4

Informing clinical data sharing standards

Recording | Agenda | Slides

Session Chairs: Yann Joly, Juliann Savatt

Meeting summary

This session focused on advancing guidance for responsible sharing of individual-level clinical genomic data in the absence of explicit patient consent. The group presented findings from two key activities: (1) a 2023-2024 survey of clinical laboratory and institutional perspectives, which found general support for pseudonymised data sharing without explicit consent, and (2) a newly completed policy analysis of 33 institutions across 17 jurisdictions, highlighting significant variability in how clinical genetic data can be shared, what data types are involved, and what (if any) data security measures are documented in institutional policies. While many institutions allow sharing under the assumption of implicit consent for direct patient care, the study revealed ambiguity in language, inconsistent transparency, and frequent omission of technical safeguards in written policies. The session concluded with discussion on identifying gaps and establishing more harmonised and transparent policy frameworks.

Key takeaways

- **There is widespread but varied acceptance of sharing without explicit consent:** most institutions reviewed in the survey permit the sharing of clinical genetic data for patient care purposes without explicit consent, but the rationale and conditions for such sharing are inconsistently described and rarely detail the types of data involved.
- **Policies lack specificity and transparency:** the majority of institutional policies do not clearly define data types, recipients, or safety measures associated with clinical genetic data sharing. Many omit explicit references to technical safeguards like pseudonymisation or encryption.
- **Consent terminology is ambiguous and potentially misleading:** terms such as "without consent" or "implicit consent" generated confusion during the discussion, highlighting the need to carefully distinguish between lack of explicit consent, consent for care, and formal research consent.

Priority action items

- **Refine and harmonise terminology:** in collaboration with members of the more technical Work Streams, develop a shared vocabulary to clarify distinctions between different types of consent (e.g. explicit, implicit, assumed, or waived) in clinical settings, particularly when policies reference data sharing for patient care.
- **Engage broader constituents for deeper insight:** conduct a small number of follow-up case studies with clinicians, genetic counselors, legal advisors, and IT security leads at a smaller subset of institutions to explore policy intent and practical implementation in more depth.
- **Create a model policy template:** use findings from the institutional analysis to draft a model policy for clinical data sharing that clearly outlines allowable sharing practices, recipient types, data categories, and required security measures. Members of the Clinical Genomics Laboratory Community will be involved in this activity.

Day 4

Querying Phenopackets with Beacon v2.0: a prototype implementation in the prenatal space

Recording | Agenda | Slides

Session Chairs: Michael Duyzend, Ron Wapner, Peter Robinson, Jess Giordano, Amy Brower

Meeting summary

This session introduced the development of a prenatal genomic data repository integrated with GA4GH Beacon v2.0 and Phenopackets, with the aim of bridging clinical and computational expertise to improve diagnosis, gene discovery, and research in fetal genomics. The session outlined challenges unique to the prenatal setting — such as limited phenotypes, complex trio genomes (fetus/placenta, mother, and father), the evolving nature of fetal development, and the need for standardised terminology, secure data sharing, and longitudinal data integration. The team presented a prototype implementation of Beacon v2.0 on the Terra platform, enabling federated phenotype- and variant– driven querying of genomic data and demonstrating the feasibility of scalable, standards-based infrastructure for clinical use. The discussion emphasised the need for a collaborative, cross-disciplinary approach to ensure adoption, enhance interoperability, and enable patient matching and discovery in real-time clinical and research settings.

Key takeaways

- **Prenatal genomics requires unique infrastructure:** standard clinical and research repositories lack fetal-specific phenotypic annotations and temporal detail. Building a dedicated prenatal repository is essential for accurate diagnosis, gene discovery, and longitudinal care.
- **Beacon v2.0 and Phenopackets integration is feasible and valuable:** the prototype deployed on Terra demonstrates that Phenopackets can be effectively queried via Beacon, enabling federated access to fetal genomic and phenotypic data, though ease of implementation and query customisation remain key challenges.
- **Clinical and technical collaboration is critical:** close coordination between clinicians and developers was essential to expand Human Phenotype Onotology (HPO) terms, standardise phenotyping from ultrasound data, and implement usable systems. The prenatal space offers a unique opportunity to co-develop tools grounded in real-world clinical need.

Priority action items

- **Develop a plug-and-play Beacon implementation:** create a streamlined, template-based deployment package tailored to prenatal genomics and beyond, to support broader adoption by sites with limited technical expertise.
- **Establish a technical working group:** convene a dedicated group to address technical issues such as variant representation, fuzzy phenotype matching, and integration with electronic medical record (EMR) systems such as Epic, with regular meetings to iterate on specifications and needs. This can be done in collaboration with the Clinical & Phenotypic Data Capture Work Stream and Beacon working group.
- **Enable scalable phenotype standardisation:** collaborate with EMR vendors and sonography software providers to export standardised Phenopackets at the point of care, and expand tools for automatic extraction of HPO terms from clinical notes and ultrasound reports.

Day 4

GA4GH Schema Registry specification

Recording| Agenda | Slides

Session Chairs: Nathan Sheffield, Kathy Reinold, Jonathan Fuerth

Meeting summary

This session focused on the Data Model and Schema Consensus (DaMaSC) group's efforts to build schema consensus and foster interoperability through two main initiatives: a lightweight Schema Registry API and a best practices guide for ontology alignment. The discussion centered on the Schema Registry API — intended to be a minimal, discoverable, and open way to find and share JSON Schemas within and beyond GA4GH. Several implementation demos showcased the potential to enable version control, governance models, and schema federation across communities. The lively discussion raised critical points about scope, format flexibility (e.g. JSON Schema versus LinkML and other model formats), discoverability, governance, and how this registry could integrate into GA4GH's product lifecycle. Participants urged a focus on user needs, clarity on use cases, and pragmatism in evolving the effort iteratively while ensuring alignment with existing global initiatives.

Key takeaways

- **Simplicity with extensibility is key:** the proposed Schema Registry API prioritises minimalism and ease of implementation, but should remain flexible to support multiple governance models, schema formats, and federated deployments. At the same time, it is critical to narrow the focus of the effort (by restricting the format or other characteristics in the short run) in the interests of providing a workable prototype. A working prototype will enable the team to explore, flesh out, and refine the specification.
- **Community value lies in discoverability and lifecycle support:** there is strong demand for a central place to discover GA4GH schemas, including both approved and in-development versions, with clear maturity indicators and update pathways.
- **Governance and maintenance matter as much as technology:** the success of any registry depends on consistent policy enforcement, version tracking, and community uptake — not just API design. Clear policies on maturity levels, update responsibilities, and approved versus sandbox schemas are critical.

Priority action items

- **Define a clear minimum viable product (MVP) scope:** focus initial implementation on approved GA4GH Work Stream schemas and establish clear policies for versioning, maturity levels, and submission governance.
- **Validate with real-world users and use cases:** engage groups such as the Genomic Knowledge Standards (GKS) Work Stream, Experiments Metadata Standard, and refget to populate the registry and validate its utility through actual adoption and integration into their workflows.
- **Coordinate with external schema initiatives:** explore alignment and interoperability with global schema efforts (e.g. Research Data Alliance [RDA], European Open Science Cloud [EOSC], and LinkML) to avoid duplication and ensure long-term relevance and compatibility.

Day 4

Automating patient matching to clinical trials

Recording | Agenda | Slides

Session Chairs: Zinaida Perova, Bernie Pope, Benjamin Haibe-Kains, Salvador Capella-Gutierrez, Romina Royo

Meeting summary

This session, which served as part of the GA4GH Cancer Community of Interest meeting, focused on automating patient matching to clinical trials. Presenters representing academic medical centers, non-profits, tech startups, and patient advocacy groups shared real-world efforts to improve trial accessibility using structured data, AI tools, and patient-centric platforms. Speakers discussed the technical and social challenges of converting free-text eligibility criteria into computable formats, the need for structured and federated data sharing, and the importance of engaging patients directly in the process. The session ended with a panel discussion calling for a GA4GH-led Study Group to advance data and interoperability standards for trial matching.

Key takeaways

- **Structured eligibility criteria are essential but still lacking:** trial eligibility criteria are often vague, unstructured, and proprietary, posing major challenges to automation. Formats and approaches such as clinical trial markup language (CTML) and large language models (LLMs) offer promise but need refinement and broader adoption.
- **Federated approaches and open standards are the future:** scaling trial matching will require shared standards for encoding clinical and trial data, real-time trial availability information, and interoperability across systems and institutions.
- **Patients must be central to the process:** patients are highly motivated to find treatment options and often drive trial discovery themselves. Tools must be accessible, empathetic, and transparent, and should support — not replace — shared decision-making with clinicians.

Priority action items

- **Form a GA4GH Study Group on clinical trial standards:** launch a dedicated group to advance computable trial eligibility formats (e.g. CTML), engage with global stakeholders (e.g. National Cancer Institute [NCI] and industry), and develop guidance for prospective trial design using standardised, structured formats.
- **Promote integration of AI/LLM tools to bridge unstructured data gaps:** support projects that apply AI to extract structured data from electronic health Records (EHRs) and trial protocols, including tooling to map free text to computable criteria for eligibility assessment.
- **Develop patient-centric trial matching interfaces and governance models:** prioritise inclusive, multilingual, and accessible interfaces for patients; establish policies that support data portability, transparency in site/trial availability, and patient navigation assistance.

GA4GH Whole Genome Sequencing Quality Control Standards meeting

Recording | Agenda | Slides

Session Chairs: Justin Jeyakani, Maxime Hebrard

Meeting summary

This session featured an in-depth presentation by Justin Jeyakani and Maxime Hebrard, who introduced the GA4GH Whole Genome Sequencing (WGS) Quality Control (QC) Standards and its reference implementations. They walked through the rationale for standardising QC metrics, details of the current deliverables (e.g. metrics definitions, reference implementation, and benchmarking resources), and plans for scope extension to long-read and somatic pipelines in the 2025 road map. Miro Cupak then presented an overview and demo of the GA4GH Data Connect standard, emphasising its flexibility, support for external data models, and integration patterns with other GA4GH products. The session concluded with a robust discussion on integrating WGS QC with the broader GA4GH ecosystem, including Data Connect, Experiments Metadata Standard, Data Use & Researcher Identities (DURI) Work Stream standards, Data Model and Scheme Consensus (DaMaSC) working group, and potential collaborations with the Genome Aggregation Database (gnomAD), European Genome-phenome Archive (EGA), and International Standards Organization (ISO).

Key takeaways

- **WGS QC is now a mature GA4GH product:** it defines formal metrics and reference pipelines for germline short-read data, with upcoming support for long-read and somatic data. It addresses a key harmonisation challenge in global genomics.
- **Data Connect enables discoverability and integration:** through flexible JSON Schema descriptions and SQL-based querying, Data Connect can federate diverse QC and metadata tables, improving data transparency and access.
- **Opportunities for broader ecosystem alignment:** multiple potential touchpoints with other GA4GH standards (e.g. DaMaSC Schema Registry, DURI Work Stream standards, and Experiments Metadata Standard) were identified, along with interest in aligning with external efforts from ISO, gnomAD, and EGA.

Priority action items

- **Submit WGS QC Metric Schema to the DaMaSC Registry:** finalise the JSON Schema for WGS QC output and submit it to the emerging DaMaSC Schema Registry to support wider reuse and harmonisation across GA4GH products. Align with Experiments Metadata Standard on the matter.
- **Develop a Data Connect example implementation for QC outputs:** package a minimal Data Connect service exposing WGS QC metrics to demonstrate feasibility for data producers and support adoption across Driver Projects and national sequencing initiatives.
- **Engage with gnomAD and EGA:** initiate targeted outreach to the gnomAD and EGA teams to explore adoption of WGS QC metrics and Experiments Metadata Standards in their submission and discovery pipelines, enhancing visibility and consistency.

Pangenome structural variant reporting and representation

Recording | Agenda | Slides

Session Chairs: Benedict Paten, Karen Miga, Heng Li, Glenn Hickey, Melissa Cline, Alex Wagner and Nathan Sheffield

Meeting summary

This session explored emerging strategies for representing structural variation within the human pangenome, highlighting recent advances and ongoing challenges in graph-based genome modeling. Benedict Paten introduced the goals of the Human Pangenome Reference Consortium (HPRC) to replace the linear reference with a diverse, phased collection of telomere-to-telomere assemblies. Glenn Hickey detailed how variation is captured in pangenome graphs and the need for nested variant representation to avoid losing complex signals. Heng Li proposed an alternative approach using a reference tree to separate variation from coordinate systems, offering improved representation of complex and overlapping variants. The session concluded with presentations on how current GA4GH standards (e.g. refget and VRS [Variation Representation Specification]) could evolve to support pangenome structures, followed by a broad discussion on the future of graph-centric references, the need for format standardisation (e.g., Graphical Fragment Assembly [GFA]), and the importance of fostering collaboration between HPRC and GA4GH communities.

Key takeaways

- **Graph-based representations are essential** for accurately capturing structural variation and common polymorphisms in human genomes, but they challenge existing variant formats such as VCF, particularly in complex or deeply nested regions.
- **There is no one-size-fits-all model yet:** multiple strategies (e.g. nested VCFs, tree-based references, and computed identifiers) are being explored to support compatibility, representation accuracy, and future scalability.
- **Closer collaboration between GA4GH and HPRC is needed** to align emerging graph-based formats with GA4GH standards, ensure toolchain interoperability, and co-develop stable, scalable infrastructure for the future of genomics.

Priority action items

- **Launch pilot projects** between GA4GH and HPRC to test the application of refget, VRS, and new variant representation strategies (e.g. nested VCF or tree-based models) on the current pangenome draft data.
- **Convene a focused working group** to evaluate and potentially standardise graph genome formats (e.g. GFA) under the GA4GH umbrella, aligning with needs from HPRC and broader genomics communities.
- **Create cross-project participation pathways** to integrate pangenome experts into GA4GH Work Streams (e.g. Genomic Knowledge Standards [GKS] Work Stream) and vice versa, fostering ongoing collaboration around variant normalisation, coordinate systems, and file format development.

Day 4

GA4GH in the real world

Recording | Agenda | Slides

Session Chairs: Venkat Malladi, Brian O'Connor

Meeting summary

The final session of the day focused on moving beyond theoretical standards to real-world implementation of GA4GH technologies. It opened with a presentation from Jilong Liu on the BIOS platform in China, which integrates several GA4GH Cloud APIs for federated data analysis, demonstrating practical adoption across institutions. This was followed by technical insights into the current state and limitations of the GA4GH starter kits, especially regarding real-world production use. The discussion evolved into a collaborative brainstorming session on creating a minimal, reusable, open-source data-sharing suite to demonstrate the impact of GA4GH standards through real-world scientific use cases. The group emphasised starting with a specific biological question to stress-test interoperability, inspire adoption, and potentially identify gaps in the existing standards.

Key takeaways

- **Starter kits are useful, but not enough:** starter kits are great for understanding and prototyping GA4GH standards, but fall short for enterprise-level deployment. There is a need for clearer purpose — whether they are for education, prototyping, or production.
- **Need for an end-to-end starter kit:** a bare-bones, open-source, end-to-end data sharing suite could help stakeholders see the real-world value of GA4GH standards. It should focus on solving one tractable scientific problem, like breast cancer or malaria variant analysis examples.
- **Use case-driven development is key:** to showcase the value of GA4GH, it's critical to identify communities already organised around specific scientific questions and support them with tools that solve their real problems — then communicate that success widely.

Priority action items

- **Draft a one-pager:** summarise the vision for the "GA4GH starter suite" with a concrete example, purpose, scope, and potential user story. Circulate it to align stakeholders and refine the concept.
- **Engage community groups:** Reach out to GA4GH Community of Interest groups (e.g. Neuroscience Community and Infectious Disease Community) to identify biological questions they are trying to solve and co-develop tailored demo implementations.
- **Establish modular, template-based starter kits:** use templated infrastructure (e.g. Helm charts and Backstage templates) to create multiple lightweight, scenario-specific starter kits, allowing customisation without starting from scratch.

Day 4

Identifying core consent elements for pangenomic studies

Recording | Agenda | Slides

Session Chairs: Heather Lawson, Eimear Kenny, Ma'n Zawati, Vasiliki Rahimzadeh, Karen Miga

Meeting summary

This session focused on defining core consent elements needed to support global pangenome research and building a federated Human Pangenome Project (HPP). The Human Pangenome Reference Consortium (HPRC) team presented a preliminary set of consent elements based on an international inventory of consent forms. Participants, both in person and online, engaged in an interactive Sli.do exercise to prioritise these elements and identify what might be missing. The group explored challenges around harmonising diverse consent models, especially balancing international data sharing, risks of re-identification, and respecting local governance and legal frameworks. The session concluded with a proposal to develop a red/yellow/green classification framework to evaluate consent compatibility, with future integration into GA4GH policy resources through the Regulatory and Ethics Work Stream (REWS).

Key takeaways

- **Risks of re-identification and international data sharing are top concerns:** participants ranked transparency around re-identification risks and whether a study participates in international data sharing as the most critical elements of consent — particularly important for building trust in underrepresented communities.
- **Plurality of consent models must be embraced, not eliminated:** the goal is not to create a one-size-fits-all consent but rather to recognise a range of local policies and participant expectations — requiring thoughtful harmonisation rather than forced standardisation.
- **Terminology and context matter deeply:** seemingly interchangeable terms like "controlled access," "open access," or even "sample" can be interpreted differently by participants, institutions, and researchers. Consent tools must reflect and respect these differences, especially in global or multilingual contexts.

Priority action items

- **Develop a consent compatibility matrix (red/yellow/green framework):** create a structured evaluation tool to classify consents based on how easily they can be integrated into a global pangenome resource — highlighting areas of compatibility, uncertainty, and incompatibility.
- **Extend and contextualise existing GA4GH consent resources:** build on the existing REWS Consent Toolkit to create a tailored set of model clauses and guidance documents specific to pangenome research, especially addressing open access, re-identification, and international data use.
- **Map stakeholder roles in consent development and harmonisation:** define and document which stakeholders (e.g. local institutional review boards [IRBs], community advisory boards, legal experts, and technologists) need to be engaged at each stage of the consent lifecycle to support ethical and scalable pangenome collaboration.