GA4GH April Connect 2026 Meeting Report



The GA4GH April Connect meeting in Montreal, Canada brought together global leaders in genomics and health data to progress our technical standards into real-world clinical impact. Over four days, the community focused on interoperability, federated learning, ethical governance, integrating GA4GH standards into national health initiatives, and more.

Bridging the gap: from technical readiness to clinical impact

A primary thread throughout the conference was the transition of GA4GH products from technical standards into practical, real-world applications. In Canada, funders such as the Canadian Institutes of Health Research (CIHR) and Genome Canada are requiring compliance with GA4GH standards to receive funding, highlighting a commitment to building an interoperable genomic and health data ecosystem. Sessions also emphasised moving towards “user-driven science,” focusing on what clinicians and researchers need from this space in order to advance patient care. There was also a focus on integrating GA4GH standards with clinically-based standards, such as HL7 FHIR and OMOP, signaling a growing effort to link the research lab to healthcare.

Equitable governance and stewardship

The Connect meeting underscored that technical standards are only as effective as the ethical and legal frameworks supporting them. The Regulatory & Ethics Work Stream is developing a governance checklist tailored to low- and middle-income countries (LMICs) that aims to address power imbalances in global partnerships. The community also explored opportunities to develop tools for computable governance, which would turn complex legal policies into machine-readable formats to facilitate transparent and efficient data access. A “stewardship maturity matrix” was also proposed, aiming to help measure how repositories are responsibly managing their data.

The launch of the GA4GH AI Work Stream

GA4GH has launched a new Work Stream focused on developing governance recommendations and data standards to foster responsible integration of artificial intelligence (AI) in genomics and health. The Connect meeting served as a kick-off for the Work Stream, with a dynamic, interactive plenary session that asked all community members to imagine the ways in which GA4GH could support relevant AI projects. Explore a list of AI projects generated by the community here, or add your own. To get involved in the Work Stream, register your interest here.

Federated ecosystems and “data visiting”

The Connect meeting demonstrated a global shift in how genomic data is accessed, moving away from large-scale data transfer to federated models. The session “From capacity to collaboration: genomic data sharing in Asia” showed an emphasis on data visiting rather than data transfer to navigate fragmented regulatory frameworks and geopolitical concerns. The GA4GH community is also actively prototyping privacy-preserving methods and other advanced privacy technologies that keep raw data in its home jurisdiction. The expansion of initiatives like the Federated Variant-Level Matching network shows a growing commitment to connecting diverse clinical and research nodes globally.

Harmonisation and schema interoperability

As the suite of GA4GH products grows, a central theme emerged regarding internal alignment of these standards. Several sessions focused on the risk of fragmentation when Work Streams use different schemas or data formats. There was also a concerted push to align how complex biological data is represented across all GA4GH tools.

Table of contents

Connect Day 1

- GA4GH Connect Welcome
- Work Streams and Communities of Interest Showcase
- The TASC-athon: Le Monstre
- GA4GH au Canada, eh?
- Scaling the Federated Variant-Level Matching Network
- Passports for... data and people and agents, oh my!: computable governance for data sharing

---

Connect Day 2

- From capacity to collaboration: genomic data sharing in Asia
- Next steps for federated learning and GA4GH: discussion and hackathon
- GKS and Clin/Pheno schema interoperability
- Future-proofing clinical data in Beacon: collaboratively reviewing next-gen phenotypic schemas
- Operationalising genomic data stewardship: refining assessment criteria and maturity measures for cloud-based environments
- Harmonising Categorical Variant Representation with the Beacon Protocol
- Guidance on human genomic data sharing in collaborative research: ethical and governance considerations for global partnerships, with a focus on low- and middle-income country contexts
- Experiments Metadata: expanding coverage and Beacon integration
- The Dating Game: Coordinating Large-Scale Implementation of GA4GH Standards
- What difference does REWS make? Identifying and measuring our impact and that of ELSI research, more generally
- GA4GH Analytics Dashboard (web edition): what's new and roadmap working session

Connect Day 3

- Building trustworthy AI in genomics and health: launching the GA4GH AI Work Stream
- Building a translational bridge: a cross-community workshop on Beacon v2 and HL7 FHIR integration
- Data passports and eIDAS: extending data access beyond research
- Verifiable Credentials infrastructure for Passports' visas use
- From guidance to practice: pilot-testing responsible generative AI guidance points for genomic data sharing
- Describing rare disease trajectories: Bring Your Own Disease and collaborate
- Towards Passport / AAI 2.0 - Data Access Pathways for All Policies
- Provenance metadata for genomic datasets

Connect Day 4

- Register-a-thon
- Structural, categorical, and long-read–derived variation in Beacon: challenges and requirements for Beacon v3.0
- The Status of Neuroscience Data Sharing in 2026: Data Governance, Tooling, and Education
- Building a Community Roadmap for Beacon v3.0
- Integrating GA4GH APIs
- GKS road map
- GIF Project showcase: progress, activities, and connections

Connect Day 1

GA4GH Connect Welcome

Recording | Agenda | Slides

The opening session set the stage for the GA4GH Connect meeting with a forward-looking conversation on the state of genomic and health data sharing and what it will take to move from technical readiness into real-world impact. Naveed Aziz (Vice President of Research and Innovation) and Fei-Fei Liu (Scientific Director) of the Canadian Institutes of Health Research highlighted how national funding and governance frameworks can help catalyse genomic and health data sharing at scale for precision medicine. In particular, the Canadian Precision Health Initiative and Pan-Canadian Genome Library are required to comply with GA4GH standards, along with all genomics-related projects receiving funding from CIHR.

Heidi Rehm, Chair of GA4GH, shared progress and highlights from GA4GH in 2025 and 2026, including new leadership, newly approved GA4GH products, an update on funding, and impact stories highlighting GA4GH products in action. Then, members of the GA4GH community provided deep dives on specific projects, including updates from the Variant-Level Matching network, an allele frequency Beacon demonstration from the European Genomic Data Infrastructure Project, the development of the ClinVar This! tool to standardise genomic knowledge dissemination, and the development of RefgetStore, which makes the retrieval of reference sequences and sequence collections more efficient. The session concluded with updates on the shifting landscape of U.S. data sharing and data security policy, followed by recent progress from the GA4GH Technical Alignment Subcommittee (TASC).

Work Streams and Communities of Interest Showcase

Recording | Agenda | Slides

The Work Streams and Communities of Interest Showcase gave attendees the opportunity to engage directly with each group, learn about current projects and upcoming priorities, and identify opportunities for collaboration. Each Work Stream and Community of Interest delivered a three-minute presentation, after which attendees moved between booths to connect with the groups. The presentations emphasised the importance of real-world adoption, usability and interoperability of standards, and community engagement.

The Genomic Knowledge Standards (GKS) Work Stream reported a pivot from standards development to adoption following the approval of several specifications (Variation Representation Specification [VRS] v2.0, Categorical Variation Representation Specification [Cat-VRS] v1.0, and Variant Annotation [VA-Spec] v1.0) last year. The GKS Work Stream also introduced regular “Think Tank” hackathons and Asia-Pacific–friendly sessions. The Clinical & Phenotypic Data Capture (Clin/Pheno) Work Stream

presented its multi-layered model of harmonising clinical context around genomic data, anchored by the ISO-approved GA4GH standard, Phenopackets. The Discovery Work Stream outlined progress on the Experiment Metadata Checklist, Beacon v2.0, and plans for Beacon v3.0. The Cloud Work Stream described its ecosystem of Cloud standards, highlighting that the Data Repository Service (DRS) now serves around 72 petabytes of accessible data to the community. The Large-Scale Genomics (LSG) Work Stream covered work on quantum-resistant encrypted formats, updates on the RNAget standard and the “Future of VCF” initiative, and plans to have the Whole Genome Sequencing (WGS) Quality Control (QC) standards achieve ISO approval. The Data Security Work Stream discussed the launch of the Data Passports Study Group and alignment with the Verifiable Credentials model, while the Data Use & Researcher Identities (DURI) Work Stream discussed refinement of the Data Use Ontology (DUO) and how access requests should function in AI-driven, agent-based environments. The Regulatory & Ethics Work Stream (REWS) highlighted two recent outputs: the Genomic Sequencing for Newborn Screening Decision Tool and the Position Statement on Genetic Discrimination (GD), declaring that GD is a violation of human rights.

The Rare Disease Community described a shift from outreach toward implementation, particularly improving data collection and standardisation in clinical contexts. The Infectious Disease Community stressed cross-border data sharing during outbreaks and ongoing alignment with Public Health Alliance for Genomic Epidemiology (PHA4GE) standards. The Cancer Community outlined priorities in structural variant representation, OncoBeacon, and clinical trial matching. The Neuroscience Community presented a joint effort with the International Neuroinformatics Coordinating Facility (INCF) and the Brain Research International Data Governance Exchange (BRIDGE) Project to integrate genomic, clinical, and imaging data. The Clinical Genomics Laboratory Community (CGLC) reported rapid growth, with a focus on variant classification standards and bridging development with real-world clinical implementation.

The TASC-athon: Le Monstre

Recording | Agenda | Slides

This session evaluated the Verified Implementations Framework for GA4GH products. Participants broke into groups to assess different implementations against specified tracks. The framework included five central tenets: transparency, interoperability, sustainability, proportionality, and recognition (not endorsement), with three tracks covering open-source software, closed-source platforms, and non-software products. During the session, participants engaged in a "Dragon's Den" style presentation where each group pitched their fictional product implementation to the Dragons. The exercise helped identify gaps and strengths in the Verified Implementations Framework across different product types, with particular discussion around documentation requirements, compliance testing, and support structures for various implementation scenarios.

Key takeaways

The session demonstrated that the proposed Verified Implementations Framework:

- is conceptually strong and applicable across diverse implementation scenarios;
- requires refinement, particularly in:
    - clearly defining what constitutes compliance;
    - supporting non-software implementations;
    - providing practical tooling and validation mechanisms.

Priority action items

- Identify example products to be pushed through tracks one, two, and three.
- Run the framework for an example product in track one.
- Complete the implementation of the framework for track one based on feedback and approve the process needed for tracks two and three (issue #74).
- Execute the same process for tracks two and three to validate the overall ethos and plan of implementation for the framework.

GA4GH au Canada, etc?

Recording | Agenda | Slides

The session, opened by Peter Goodhand and Anne-Cecile Desfaits, positioned Canada at a key inflection point in advancing its genomics strategy. While the country benefits from strong scientific leadership and aligned national funders such as Genome Canada and the Canadian Institutes of Health Research, participants emphasised persistent fragmentation across data platforms and infrastructure. The central opportunity is to leverage GA4GH standards to enable a more coordinated, interoperable ecosystem.

Across presentations and discussions, a consistent message emerged: technical solutions for interoperability largely exist, but progress is constrained by governance, policy, and coordination challenges. Key priorities include improving data harmonisation, aligning consent and access frameworks, investing in shared infrastructure such as trusted research environments (TREs), and strengthening workforce capacity in federated data approaches. Participants also highlighted the need for clearer roles across platforms and more coordinated funding strategies to avoid duplication and support scalable solutions.

The session concluded with a call for measurable, system-level progress. Proposed next steps include developing a national metrics framework, advancing adoption of key standards, and establishing sustainable models for infrastructure and tools. Canada's size and collaborative culture were seen as advantages in building a connected, learning health system that can integrate local, national, and global data efforts.

Key takeaways

- Interoperability is the priority; fragmentation is expected but manageable.
- Barriers are mainly policy, governance, and coordination — not technical.
- Sustainable infrastructure and workforce capacity are critical gaps.
- Stronger funder alignment is needed to scale national impact.

Priority action items

- Define a national metrics framework for adoption and impact.
- Coordinate funding and asset mapping to reduce duplication.
- Accelerate GA4GH standards implementation across platforms.
- Advance Trusted Research Environment infrastructure.
- Harmonise consent and data access frameworks.

Scaling the Federated Variant-Level Matching Network

Recording | Agenda | Slides

This session focused on expanding the Federated Variant Level Matching (VLM) network to integrate diverse data sources (clinical, research, patient-led) and onboard new nodes, including DNAstack (long-read data for indels/CNVs/SVs), Seqr in Genomics (Canada), and Breakthrough Genomics. Key technical developments include extending beyond single nucleotide variants to handle indels and structural variants, improving cross-node querying (especially gene-level searches), and enabling better access to phenotype information via new API endpoints integrated with Phenopackets and Beacon V2.

There was discussion about a privacy-preserving method to identify duplicate individuals across datasets using hashed genomic "fingerprints" derived from a fixed panel of common SNPs. Property-preserving hashing (PPH) was highlighted as essential to tolerate sequencing errors, and zero-knowledge proofs were suggested as a potential enhancement.

For indel matching, the Variation Representation Specification (VRS) was presented as a standardised model using inter-residue coordinates and normalisation to resolve representation ambiguities (e.g. VCF vs. HGVS). The AnyVLM software stack (built on AnyVar) translates between formats, normalises variants, handles GRCh37/38 liftover, and exposes a VLM-compatible API — allowing nodes to adopt robust standards without full re-engineering.

The session concluded with agreement that indel support and gene-level queries are near-term priorities, and that community consensus on implementation, standards, and governance remains essential.

Key takeaways

- Genomic hashing with property-preserving hashing can deduplicate individuals across nodes, solving a critical scalability and accuracy problem.
- Indel matching requires normalisation — VRS + AnyVLM provides a practical path forward without forcing nodes to rebuild their systems.
- Gene-level and phenotype queries are the next major features after the phenotype endpoint goes live.

Priority action items

- Prototype hashing for deduplication: select SNP panel and test property-preserving hashing.
- Pilot AnyVLM for indels: test across nodes, document integration steps.
- Finalise gene query roadmap: define API, annotation standards, and timeline.

Passports for... data and people and agents, oh my!: computable governance for data sharing

Recording | Agenda | Slides

This session explored Data Passports, and the product's benefits for offering a computable way to represent data governance policies. Discussions highlighted the strong potential for improved interoperability and reduced fragmentation, along with key challenges such as overlapping Visa categories and the need to handle complex, dynamic data policies.

Key takeaways

- There is interest in Data Passports as a standardised way of providing machine-readable policy metadata with benefits of improving interoperability and transparency in data governance. Key benefits include increased interoperability, reproducibility, and auditability.

- Data Passports can act in parallel to complement the existing Passports.
- Avoiding oversimplification when grouping Visa types is key to prevent incorrect access decisions.

Priority action items

- The proposed Visa types offer a useful starting point; though, there should be continued work to address the potential overlap between categories.
- A clearer conceptual framework may be needed, given that data governance is context-dependent.
- Continue exploring discussions around Data Passports' granularity. Details on these additional discussions are available here.

Connect Day Two

From capacity to collaboration: genomic data sharing in Asia

Recording | Agenda | Slides

This session focused on examining the legal and technical landscapes of genomic data sharing across Singapore, China, Japan, South Korea, Hong Kong, and Taiwan. While these regions have developed robust national precision medicine initiatives and biobanks, participants identified significant barriers to international collaboration, including fragmented regulatory frameworks, lengthy approval, and geopolitical security concerns. The discussions looked at more secure models, such as trusted research environments (TREs) and federated analysis, highlighting the need for GA4GH protocols to balance privacy with scientific progress.

Key takeaways

- High-capacity jurisdictions in Asia have established massive genomic datasets, but remain siloed by local governance.
- There is an increased emphasis on shifting away from data transfer towards data visiting approaches.
- Public engagement shows general support for research aimed at public benefit, but reluctance remains regarding access by the insurance and commercial sectors.

Priority action items

- Conduct a survey across Asian jurisdictions to map data sharing practices and barriers.
- Develop a regional white paper on genomic data sharing, focusing on risk-based classification and governance mechanisms.

Next steps for federated learning and GA4GH: discussion and hackathon

Recording | Agenda | Slides

This session focused on Federated Learning (FL). FL enables collaborative model training across platforms without sharing raw data, improving privacy and leveraging GA4GH building blocks such as WES, TES, DRS, and Passports/AAI. The session will include a concise overview, a collaborative discussion to prioritise a backlog of follow-up work, and a hackathon to turn a small set of high-priority items into prototypes or draft pull requests. Candidate hackathon tasks include an authorisation handoff demo between WES and TES, a compute-and-billing metadata exchange prototype, and a model write-back workflow to DRS with provenance and access control. Expected outcomes are a prioritised action list, potential external collaborators, non-FLWG collaborators, and at least one reproducible artifact or PR or prototype from the hackathon.

Key takeaways

- Five core challenge areas identified: data and compute access; billing and cost accounting; telemetry and service information; policy and governance; and real-world implementation
- Two use case models: formal consortia (2–10 institutions with legal agreements) versus ephemeral, edge-based federated learning at scale
- GA4GH's focus: to develop interoperable governance frameworks and enable different federated learning technologies to work together, rather than building new solutions from scratch
- Privacy considerations: differential privacy, with appropriate budget management, can help protect data; alignment with existing NIST frameworks is recommended
- Practical deliverables: governance frameworks, cost profiles, telemetry standards, provenance queries, GWAS demonstrations, and Passport integration
- Collaboration approach: GitHub-based workflows, with pull requests as the primary mechanism for tracking contributions; continued collaboration encouraged beyond Connect
- Success factors: real-world driver projects, cloud-agnostic solutions, data standardisation, and clear consortium sizing to support privacy benefits

Priority action items

- Coordinate with leadership from the new GA4GH AI Work Stream on next steps.
- Determine a set of features for existing GA4GH products needed to address FL-related use cases, and develop a release plan for these features.
- For FL-related challenges that may result in new GA4GH products, continue to formulate product concepts in collaboration with the AI Work Stream.
- Hackathon leadership: drive a structured GitHub workflow, including pull request management, group-based labeling, and weekly summaries.
- Policy and governance: develop a minimum viable governance framework covering data use agreements, cost sharing, authentication, authorisation, compliance, and readiness checklists.
- Billing and cost accounting: create node-specific unit cost profiles, assess open costing standards, and design job cost monitoring and budgeting capabilities.
- Telemetry and service information: define minimal telemetry fields and service descriptors for federated infrastructures, with extensibility aligned to GA4GH standards.
- Data access and compute: develop provenance query mechanisms and cohort selection tools, integrating Beacon and Phenopackets standards.
- Real-world implementation: demonstrate federated GWAS across distributed sites with privacy-preserving features, and explore expanding participation.
- Passport and workflow integration: advance GA4GH Passport integration with WES/TES systems to support controlled data access.

GKS and Clin/Pheno schema interoperability

Recording | Agenda | Slides

The session centered on launching a pilot project to achieve technical interoperability between two GA4GH Work Streams: the Genomic Knowledge Standards (GKS) and Clinical & Phenotypic Data Capture Work Streams.

A framework was laid out to align products in the Work Streams that were developed independently and now create fragmentation for implementers. Key interoperability gaps were illustrated and included divergent representations of external concepts, variant interpretation models, and technical mismatches, yielding incompatible JSON schemas.

The proposed resolution is a “meet-in-the-middle” process under GA4GH’s Technical Alignment Subcommittee (TASC), featuring bi-weekly meetings, pre-circulated draft models, voting, and milestone-driven road maps that prioritise translation layers over breaking changes to existing specifications.

Breakout group discussions surfaced several cross-cutting themes: the critical need to engage actual Phenopackets users (including clinical and research implementers) to understand their needs and constraints; a desire for a minimal shared core of entities (e.g. variants, genes, phenotypes) while preserving richer data in each specification; the importance of validation tools, example conversions, and clear documentation of decisions; and a call for top-down enforcement of aligned models for new products, accompanied by gentle migration paths for existing users.

Key takeaways

- Significant interoperability gaps exist between GA4GH Work Streams. Products from GKS and Clin/Pheno were not designed to be interoperable. They use different schemas, data formats (JSON schema vs. Google Protocol Buffers), and representations of shared concepts like genes and variant interpretations, forcing implementers to do extra work.
- A "meet-in-the-middle" pilot project was proposed to drive alignment, with the goal of bringing together developers and implementers from relevant Work Streams to align on core data models and create a road map for future interoperability.
- A key point of discussion was whether to prioritise pure technical interoperability or focus on scientific impact. Some participants emphasised that resources should be directed toward user-requested features and demonstrating clear added value, rather than pursuing interoperability as the primary goal.
- Multiple breakout groups stressed the importance of understanding real-world use cases.

Priority action items

- Present the proposed "meet-in-the-middle" framework and pilot project plan to all relevant Work Stream Leads to identify all impacted products, gather broad input, and engage additional participants.
- Engage end-users and implementers early in the process. Actions include creating a registry of Phenopackets users, conducting user surveys to gather requirements, and involving clinicians and implementers to understand their pain points and transition challenges before making changes.
- To support the alignment process, the group agreed to produce practical tools, including: validation sets for new schemas, example data conversions between existing formats, clear

documentation of all design decisions (e.g. why a minimal representation was chosen), and a formal versioning road map.

Future-proofing clinical data in Beacon: collaboratively reviewing next-gen phenotypic schemas

Recording | Agenda | Slides

The session opened with a framing of the "fragmentation risk" between Beacon and Phenopackets, highlighting missing elements such as medical actions and SDOH. The distinction between Beacon's query model (limited filters) and representation model (derived from Phenopackets) were clarified with notes that temporal and relational queries are particularly constrained.

Breakout groups mapped synthetic clinical cases (rare disease and cancer) onto Beacon and Phenopackets schemas. Key gaps identified included:

- inability to determine patient age from sample data alone;
- lack of phenotype information in some case records;
- no standardised place for disease stage, metastasis timing, or treatment-outcome linkages;
- difficulty representing longitudinal data (e.g. cancer progression, multiple treatments over time).

Several themes emerged: the complexity of sex and gender representation, the intentional omission of date of birth for privacy, and the need for better documentation of what Beacon can and cannot do. One group proposed the "handshake model" (Beacon as a discovery layer pointing to external rich data sources like OMOP or Phenopackets), which resonated with many attendees.

The AI segment demonstrated a prompt-engineered LLM tool that converts unstructured clinical notes into Beacon queries. While promising for scaling semantic mapping, the group noted that hallucinations and inconsistent ontology mapping remain barriers. The session closed with a commitment to synthesise feedback into a prioritised roadmap for Beacon and Phenopackets releases.

Key takeaways

- Beacon cannot link related events (e.g. treatment to specific disease episode).
- Sex and gender representation is incomplete for clinical needs.
- There is a preference for the "handshake" model: Beacon for lightweight discovery and pointer to richer data sources.

Priority action items

- Synthesise breakout group feedback to identify priority gaps.

- Review Beacon model for potential additions (e.g. medical actions, temporal linkages).
- Align Beacon and Phenopackets versioning to avoid fragmentation.

Operationalising genomic data stewardship: refining assessment criteria and maturity measures for cloud-based environments

Recording | Agenda | Slides

This session focused on the development of a stewardship maturity matrix for datasets for maximising the stewardship impact in cloud environments. The group reviewed progress from a multi-year Delphi study which established core outcomes for responsible stewardship. Key objectives of the session included gathering community input on the proposed criteria to support the eventual draft of a stewardship maturity matrix for repositories.

Key takeaways

- Ambiguity between data repository and data contributor responsibilities around de-identification highlights the need for clearer delineation.
- While robust enforcement is critical, criteria must also remain practical and not overly burdensome, especially given resource variability across repositories.

Priority action items

- Refine the proposed stewardship criteria by incorporating stakeholder feedback gathered during the workshop discussions.
- Develop a draft of the stewardship maturity matrix that translates criteria into measurable items for repositories.
- Pilot the framework with selected repositories and data stewards to assess usability, feasibility, and clarity of the maturity indicators.

Harmonising Categorical Variant Representation with the Beacon protocol

Recording | Agenda | Slides

The session opened with an overview of the agenda and goals: exploring how Beacon and GKS standards (particularly Cat-VRS built on VRS) can align to improve variant querying and representation. There was an introduction on Cat-VRS capabilities, followed by context on Beacon's current query vs. representation models and its philosophy of referencing external standards rather than owning variant representation.

Participants discussed current blockers to VRS/Cat-VRS adoption in Beacon, aspirational variant types for Beacon v3.0, and potential entry points via new VQS request parameters. Key topics included the fuzzy nature of CNV/SV matching, the definitional vs. query-oriented role of Cat-VRS, challenges with normalisation from VCFs, and the desire for a collaborative "variant query standard."

A lively discussion highlighted the need for better tooling and normalisation guidance while preserving Beacon's flexibility. The group explored ways to handle complex representations (adjacencies, fusions, pharmacogenomic effects) without creating forks. The session closed with strong interest in continued collaboration and concrete next steps for Beacon v3.0.

Key takeaways

- Cat-VRS and Beacon can strongly support each other: Cat-VRS provides a powerful constraint-based model for defining and matching categorical variants, addressing limitations in Beacon v2.0.
- Tension exists between Beacon's permissive implementation model and the need for more consistent, standardised query parameters aligned with VRS/Cat-VRS.
- Beacon v3.0 should prioritise support for structural variants, gene fusions, adjacencies, pharmacogenomics (including star alleles), co-occurrence queries, and improved location-based matching for fuzzy variants such as CNVs.

Priority action items

- Launch a joint pilot or working group (building on Beacon Scouts, potentially linked to GKS Think Tank) with interested participants for ongoing collaboration on variant queries.
- Co-develop new VQS query parameters and request types for Cat-VRS objects, fusions, adjacencies, and other complex variants in Beacon v3.0, ensuring backwards compatibility.
- Promote GKS tooling and explore recommendations for normalisation to help Beacon nodes move toward standardised VRS/Cat-VRS representations.

Guidance on human genomic data sharing in collaborative research: ethical and governance considerations for global partnerships, with a focus on low- and middle-income country contexts

Recording | Agenda | Slides

This session introduced a draft data governance checklist aimed at promoting equitable sharing of genomic data globally, particularly for LMICs. Discussions highlighted a persistent gap between the principles of current frameworks and their real-world implementation in resource-constrained settings. The checklist aims to bridge these divides by offering practical guidance components, such as

informed consent, capacity building, empowerment, accountability and enforcement.Attendees discussed how to operationalise and pilot the tool in practice as well as potential challenges related to genomic illiteracy, data ownership, and benefit-sharing mechanisms. The session also reinforced the importance of governance tools supporting context-specific, co-developed approaches that reflect diverse implementation settings.

Key takeaways

- Genomic data sharing is shaped by power dynamics, governance structures, and capacity constraints.
- Strong global frameworks exist, but they lack practical implementation guidance in LMIC contexts. Common implementation challenges include limited legal and institutional capacity, weak enforcement mechanisms, and administrative burden.
- The checklist is intended for piloting in additional African country contexts to assess its applicability across diverse regulatory and research environments. While grounded in African and broader LMIC experiences, the tool is designed to inform equitable genomic data sharing in global collaborative research.
- To ensure maximum practicality and utility of the checklist, capacity building, informed consent, benefit sharing, and community engagement must be adapted to local contexts and viewed as continuous processes without imposing unrealistic operational burdens.

Priority action items

- Continue refining the checklist using feedback gathered from the session, along with iterative testing.
- Strong emphasis on co-developing with stakeholders and adapting to local contexts to ensure more meaningful participation and oversight.

Experiments Metadata: expanding coverage and Beacon integration

Recording | Agenda | Slides

The session opened with an introduction to the Experiments Metadata Checklist, approved in November 2025 by the GA4GH Product Steering Committee. Genomic files often lack experimental context, hindering data discovery. The Checklist describes the most important variables within a sequencing experiment, focusing on properties that are standards across all -omics disciplines. The product has been implemented by organisations and initiatives such as the Pan-Canadian Genome Library and the National Cancer Institute. The team also shared their product road map, focusing on

expanding core concepts, developing a computable schema (LinkML), aligning ontologies, and collaborating with ELIXIR and the GA4GH Infectious Disease Community.

Breakout groups then explored specific topics to refine metadata properties for long-read sequencing, single-cell sequencing, and targeted sequencing, resulting in actionable feedback.

Jordi Rambla also presented on opportunities for integration with Beacon, sharing that Beacon v2.0 introduced two endpoints — analysis (dry lab) and run (wet lab) — mainly to explain why a variant wasn't detected. The endpoints are lightly used, with changes welcome for Beacon v3.0.

Key takeaways

- Breakout groups delivered concrete, actionable feedback in three core areas.
    - The long-read sequencing group recommended consolidating fields, removing raw signal files, and using conditional fields for PacBio vs. Nanopore.
    - The single-cell sequencing group recommended distinguishing cells loaded (raw) from actual cells (processed). Many properties also apply to bulk sequencing.
    - The targeted sequencing group recommended splitting the "library selection" vocabulary and adding pointers to library BED files.
- Beacon integration is feasible — the v2.0 "run/analysis" endpoints are under-used, so changes are welcome for Beacon v3.0.

Priority action items

- Finalise the three category extensions with conditional fields and resolve overlaps.
- Push core metadata into the Genome Standards Consortium (GSC) and the International Nucleotide Sequence Database Collaboration (INSDC).
- Develop a computable schema using LinkML and generate a JSON schema for Beacon integration.

The dating game: coordinating large-scale implementation of GA4GH standards

Recording | Agenda | Slides

This session focused on identifying concrete, near-term data sharing opportunities between GA4GH National Initiatives Forum (NIF) members, with an emphasis on implementing GA4GH standards within real-world technical and policy constraints.

Participant “dat(a shar)ing profiles” and short presentations provided a snapshot of each initiative’s infrastructure, data types, and sharing goals. These inputs helped surface areas of alignment and informed the breakout discussions, where smaller groups worked through what would be required to enable data sharing between specific systems. Profiles were presented for: NCI Cancer Research Data Commons, China Digital Lung, Ontario Brain Institute - Brain-CODE, Pan-Canadian Genome Library, Genomics England, Hartwig Medical Foundation, NHLBI BioData Catalyst, European Genome-phenome Archive, GUARDIANS, Genomics Thailand, BGSI, GSA Human and CNCB Cloud, AnVIL, Synapse/Sage Bionetworks, Tohoku Medical Megabank project, and NIH All of Us

Discussions highlighted that while many initiatives have elements of the necessary technical infrastructure in place, moving toward cross-border data sharing is often shaped by governance frameworks, consent models, and institutional policies. In several cases, participants were able to outline potential pilot collaborations; in others, the conversations clarified where constraints currently limit implementation.

Overall, the session helped define both emerging pilot opportunities and a more concrete understanding of the steps and conditions required to move toward real-world data sharing.

Key takeaways

- Many initiatives have foundational technical components in place (e.g. APIs, controlled access systems), but implementation approaches vary.
- Policy, governance, and consent frameworks are key factors shaping what is feasible.
- Cross-border data sharing introduces additional complexity, particularly around legal and institutional alignment.
- There is a need for clearer, practical guidance on how to implement GA4GH standards in specific contexts.
- Small groups with aligned interests were able to explore more concrete implementation steps.
- In some cases, potential pilot collaborations began to take shape; in others, key constraints were more clearly defined.

Priority action items

- Identify and follow up on potential pilot collaborations discussed during the session.
- Compile and analyse dat(a shar)ing profiles into the NIF Data Sharing Readiness Report.
- Synthesise common technical and policy considerations raised across groups.
- Develop practical, implementation-focused guidance based on recurring needs.
- Explore mechanisms to continue discussions and track progress on emerging pilots.

What difference does REWS make? Identifying and measuring our impact and that of ELSI research, more generally

Recording | Agenda | Slides

This session explored how to capture and clearly communicate the impact of REWS work, ELSI- related research within GA4GH and supporting the broader community as a whole. Discussions highlighted key challenges with ELSI contributions often being less visible in citation counts, as well as issues related to traceability, attribution, and feasibility. Through these discussions, the group identified a broad set of potential impact indicators, and identified the need for a mixed-methods approach that combines quantitative metrics and qualitative insights. Renewed emphasis on aligning with GA4GH-wide metrics initiatives, such as the new dashboard, were also deemed critical for capturing REWS' impact.

Key takeaways

- The impact of REWS is real but largely invisible under the current metrics.
- Traditional indicators (e.g. citations) fail to capture the policy and governance influence.
- There is a strong need for a mixed approach combining quantitative tracking and qualitative evidence.

Priority action items

- Incorporate session feedback into the framework of impact indicators.
- Align with the efforts of the GA4GH metrics and analytics dashboard.
- Pilot targeted qualitative approaches (e.g. Driver Project engagement, case studies).
- Explore funding opportunities to support further development and implementation.

GA4GH Analytics Dashboard (web edition): what's new and road map working session

Recording | Agenda | Slides

The session started with Sasha Siegel (GA4GH Chief Product Officer) giving a brief overview of the purpose of creating the analytics dashboard as well as the changes from the v0.1 version to the new v1.0 version. She then presented the data sources that were used to create the dashboard.

The GA4GH Technical Team presented the new Implementation Registry that is in development. Data from the Implementation Registry would then be used for the implementation map on the dashboard.

Next there was a demo of the new dashboard, with an introduction of the glossary and key terms section as well as the core KPIs. This was followed by an explanation of the data being displayed in the plots and a demo of the data tables with explanations on how to use them to filter and export data.

Session participants tried out the dashboard and discussed how they would use it from the perspective of researchers using the tool or the product team shaping the tool. The tech team received feedback from each discussion group and answered remaining questions. The session closed with future plans.

Key takeaways

- Given the use case for the analytics dashboard, the default plots do not fit all personas using the dashboard.
- This would be a good platform to include REWS data in the future.

Priority action items

- Toggle of view based on persona: implementers, funders, analysts.
- Beacon to be included: GitHub and data from Beacon repositories.
- GA4GH product-based artefacts classification data.
- Work with REWS on policy framework and toolkit data to be included in the dashboard.

Connect Day Three

Building trustworthy AI in genomics and health: launching the GA4GH AI Work Stream

Recording | Agenda | Slides

Artificial intelligence (AI) is rapidly reshaping human genomics — from discovery and biobank-scale analysis to clinical interpretation support. The opportunity is immense, but so is the risk of fragmentation, irreproducibility, inequitable performance, and the rapid development of tools that outpace governance.

GA4GH has long provided the technical and policy foundation that enables genomic data to move responsibly across institutions and borders. As AI becomes embedded in genomics workflows, the community now needs shared standards that ensure transparency, replicability, equity, and real-world effectiveness.

This interactive workshop began with an introduction to the GA4GH AI Work Stream, its proposed charter, the scope of the group, and alignment across existing Work Streams. Participants then collaboratively identified opportunities for potential projects and initiatives.

Key takeaways

- GA4GH is launching an AI Work Stream to standardise responsible AI in genomics globally. This foundational Work Stream will aim to coordinate related activities across the other Work Streams and collaborate with the Regulatory & Ethics Work Stream on responsible AI guidance.
- Five cross-cutting principles will guide the work of the AI Work Stream: equity, attribution, governance, reproducibility, and transparency
- More than 40 project ideas were generated by the community brainstorming session, and organised into key thematic areas: agentic AI, data harmonisation, benchmarking, compliance, privacy and security, and documentation.

Priority action items

- Finalise and ratify the AI Work Stream charter with community amendments and confirm liaisons with other GA4GH Work Streams.
- Publish the AI Works Stream road map, including mapping the 40+ community project ideas into thematic working groups, organising GitHub repositories, and setting up mailing lists.
- Begin regular meetings to advance project work. Coordinate with the Regulatory & Ethics Work Stream on prioritising responsible AI guidance points.

Building a translational bridge: a cross-community workshop on Beacon v2.0 and HL7 FHIR integration

Recording | Agenda | Slides

The session opened with an overview of Beacon v2.0 as a lightweight API designed for genomic data discovery, emphasising its role as a unifying query layer across heterogeneous systems. While Beacon provides a flexible default model and standardised interface, it relies on implementers to perform the necessary data harmonisation and mapping.

FHIR was then introduced as a widely adopted clinical standard built around modular resources such as patients, observations, and diagnostic reports. Its design prioritises clinical workflows and patient-centric data exchange, making it fundamentally different from Beacon's discovery-oriented approach. This divergence in design philosophy emerged as a central theme throughout the discussion.

While it is relatively straightforward to wrap APIs or create translation layers, true interoperability depends on aligning the underlying data models. Examples were shared demonstrating successful mappings between GA4GH variant representations and FHIR genomic structures, including large-scale transformations of millions of variants, confirming that integration is feasible but resource-intensive.

A detailed case study on the Variant Level Matching (VLM) network demonstrated that query translation between Beacon and FHIR is largely straightforward, involving parameter mapping rather than conceptual redesign. However, response transformation is more complex, requiring de-identification of patient data, conversion into the Phenopackets format, and alignment of phenotype and metadata representations.

Finally, a cancer-focused use case highlighted deeper challenges in expressing complex queries that link diseases, treatments, and variants within a single clinical context. These limitations were attributed not only to data models but also to how data is captured in practice, where relationships between clinical elements are not always explicitly recorded.

Key takeaways

- While API-level integration between Beacon and FHIR is achievable, aligning the underlying data models requires significant effort and remains the primary bottleneck.
- FHIR is designed around patient-level clinical workflows, whereas Beacon is designed for population-level discovery queries. This mismatch affects query direction, response structure, and system design.
- Beyond technical work, privacy constraints, institutional policies, and consent frameworks significantly impact feasibility and adoption.
- Existing FHIR operations provide partial alignment. Operations like population-level variant queries can approximate Beacon functionality, but limitations remain.
- Differences in phenotype representation (e.g. HPO vs ICD/SNOMED), missing metadata (e.g. provenance, handover info), and lack of direct equivalents for some Beacon concepts complicate interoperability.

Priority action items

- Continue developing and refining mappings between Beacon, FHIR, and related standards (e.g. Variation Representation Specification [VRS] and Phenopackets).
- Explore and compare API-level vs. database-level translation strategies, including implementation of adapters that handle query conversion, response transformation, and de-identification.
- Engage stakeholders to define consent frameworks, privacy safeguards, and institutional participation approaches for integrating clinical systems into Beacon networks.
- Document best practices for implementing Beacon-FHIR integrations, including query patterns, data transformation approaches, and known limitations.

Data Passports and eIDAS: extending data access beyond research

Recording | Agenda | Slides

This session focused on looking at the intersection of the GA4GH Data Passport with the Electronic Identification, Authentication and Trust Services (eIDAS) regulation, and whether GA4GH standards can be used in this context. Discussions explored the requirements of the European Genomic Data Infrastructure (GDI) project, as well as issues related to trust, scalability, and user control within the current Passport context. The eID levels of identity assurance and schemes under the eIDAS were also examined, as well as the limitations in connectivity across countries. The presentations also explored the decentralised Web 3.0 model and its use of distributed ledger technology, digital wallets, and Verifiable Credentials for bolstering interoperability and privacy, with a practical demo included.

Key takeaways

- Distinctions between the eID and digital wallet demonstrates the comparative benefits of the latter in improving privacy.
- Moving to the EU Digital Identity Wallet model shifts control back to the user via Verifiable Credentials.
- The Verifiable Credentials model within the context of Web 3.0 offers better capabilities for cryptographically verifying information through decentralised access methods, improving access, reliability, and selective disclosure.

Priority action items

- Continue exploring GDI's requirements to act as an authorised participant in the European Health Data Space to focus on the linkage between clinical health and genomic datasets.
- Address how to handle access to compute resources and correct billing profiles using Verifiable Credentials.
- Explore how the EU digital identity wallet can bolster interoperability within the GA4GH context.
- Review the "Verifiable Credentials infrastructure for Passports' visas use" Connect session for additional technical examples of Verifiable Credentials.

Verifiable credentials infrastructure for Passports' visas use

Recording | Agenda | Slides

This session focused on exploring the Verifiable Credentials (VC) infrastructure to determine if the existing GA4GH Passport information can be represented using the VC format. The discussions explored the VC formats that extend beyond W3C, current protocols, and OpenID Connect profiles.

Key highlights included examining how the Passport ecosystem fits into the VC infrastructure, and the importance of selecting a profile that ensures global interoperability with emerging regulatory requirements.

Key takeaways

- VC provides a robust infrastructure for mapping the existing GA4GH Passports, offering highly interoperable formats including SD-JWT and ISO/IEC 18013-5 Mdoc.
- There's an emerging focus on key profiles, such as OpenID4VC-HAIP, DIIP, and ETSI, with HAIP identified as a robust option for high-security environments due to its strong holder authentication.
- Exploratory discussions on the role of AI agents, the required formats and protocols, and security levels reiterated the need for strong guardrails.

Priority action items

- Update the current Passport ecosystem architecture diagrams to include the verification of Passport and Visa signatures.
- Continue exploring the interaction with the Verifiable Credentials model and Passports, in parallel to ongoing improvements to the existing Researcher Passport specification.
- Produce a first draft of a new profile, and continue discussions during the biweekly Data Security Work Stream (DSQS) meetings. Further details on the DSWS meetings can be found here.

From guidance to practice: pilot-testing responsible generative AI guidance points for genomic data sharing

Recording | Agenda | Slides

This session focused on defining the approaches for pilot-testing the guidance points on the responsible use of generative AI for genomic data sharing. Participants reviewed the proposed guidance points and engaged in a prioritisation exercise, identifying three primary pillars including "Transparency and Explainability ," Technical Robustness and Accuracy," and "Privacy, Data Security, and Responsible Data Governance". Groups additionally considered various implementation pathways across key categories, like system design, data governance, human oversight, and operational processes, to shape the next phase of work. These insights will also help inform the continued partnership with the new AI Work Stream to ensure that ethical oversight remains iterative and embedded in the development of generative AI tools.

Key takeaways

- The ten REWS guidance points, synthesised from 90+ regulatory frameworks across 60+ countries, represent a near-complete first phase of the working group's mandate.
- Participants prioritised three guidance points for pilot-testing: transparency and explainability; technical robustness and accuracy; and privacy and data governance.
- Implementation pathways discussed include system design choices, data governance mechanisms, human oversight structures, and operational processes.
- Defining KPIs, end users, and evidentiary thresholds for harm reporting were identified as key challenges requiring further elaboration.
- Ongoing collaboration with the GA4GH AI Work Stream is essential to translating ethical guidance into technical practice.

Priority action items

- Circulate the next draft of the guidance points paper to interested participants for pre-submission review.
- Share prioritisation and implementation outputs with the GA4GH AI Work Stream for integration into their technical road map.
- Continue iterative collaboration between REWS and the AI Work Stream to refine and pilot-test the selected guidance points in real-world genomic data contexts.

Describing rare disease trajectories: bring your own disease and collaborate

Recording | Agenda | Slides

This session focused on how to tell the complete “story” of real-world cases using a structured narrative that can be mapped to ontologies, standards, and registries. Experts provided an overview of current ontologies and models (e.g. HPO, MONDO, Phenopackets, PhenoDB), outlining their development, usability and limitations. Danielle Boyce from ALS-TDI provided an overview of the Natural History Study, explaining how electronic health record (EHR) integration has led to large, complex datasets. She also demonstrated why capturing exposure history data in a structured, computable way so it can be linked to biological samples remains difficult. Zohreh Talebizadeh from Global Genes discussed the RARE-X PRO surveys, and emphasised the meaningful insights patients and caregivers provide; however, personalised data and patient language is difficult to map to structured ontologies. The final case study was presented by Janine Lewis on IAMRARE, a

patient-powered registry. Key challenges to creating trial-ready datasets include semantic mismatches, varying levels of granularity, and missing terms for newly described features.

The session concluded with a discussion on the minimal information required for describing patient phenotypes and how to make data more interpretable and useful for researchers. The conversation started with ClinGen's head-to-toe surveys, and how they can be paired with disease-specific extensions in order to capture a shared, core dataset without compromising customisation. Participants also discussed capturing objective details (e.g. treatment, hospitalisation) alongside generic modifiers (mild/moderate/severe) to avoid misinterpretation. From the patient and caregiver perspective, tools that reflect real experiences and allow flexible approaches to data collection can enhance engagement. Participants agreed to develop toolkits, standardised templates, and lay-friendly terminology to improve usability of patient-recorded data.

Key takeaways

- Problems with current ontologies and phenotyping models stem from how data are collected, metadata requirements, a lack of documented negatives and exclusions, and heterogeneity across EHR systems.
- Interoperable, community-built ontologies are necessary to create a more consistent and computable representation of disease and exposure data.
- HPO provides standardised terminology while Phenopackets enables richer, more clinically meaningful representations of patient data. Both structured data and narrative context are needed to enhance the understanding of rare disease cases.
- Different imaging modalities (e.g. facial photos, X-rays, and retinal scans) can be integrated into standardised frameworks (e.g. Phenopackets) to enhance clinical decision-making and AI-driven diagnostics.
- The most critical step in PhenoDB's workflow is accurately identifying a list of possible diagnoses to guide gene and variant interpretation; experts are exploring the use of AI for diagnosis prediction from phenotypic data.
- Variability in clinical systems and structural gaps in HPO decrease diagnostic yield (i.e. accurate phenotyping).
- The "minimum computable profile" for ALS requires better standardised exposure and contextual data.
- Personalised data capture, integration of genotype-phenotype relationships, data integration across registries, and gaps in mapping patient language to ontologies limit phenotyping on the RARE-X platform
- IAMRARE wants to produce standardised, trial-ready datasets to reduce the burden on patients and accelerate therapy development; however, semantic mismatches, varying levels of

granularity, and missing terms for newly described features make it difficult to map data to ontologies

Priority action items

- Schedule follow-up meetings and strategise how to move forward with data standardisation and phenotyping initiatives.
- Organise disease-specific working sessions to build data collection standards, with ALS as a pilot, and improve the forms used for data collection based on initial feedback from workshop participants.
- Encourage the community to bring use cases and real world experiences to the GA4GH Human Exposome Data Standards group and the Rare Disease Community meetings.
- Explore creating a series of resources, including: a "beginner's guide" catalogue of existing phenotyping tools and resources;,best practices for patient-led phenotyping, a list of synonyms of HPO terms geared towards the layperson, and a toolkit for patient advocacy groups to implement HPO and standardised phenotyping in their registries.

Towards Passport /AAI v2.0: data access pathways for all policies

Recording | Agenda | Slides

This session focused on defining the road map and a set of features for the upcoming major release of Passports and the Authorisation and Authentication Infrastructure (AAI) v2.0. Current gaps and challenges of the existing specification were discussed, along with a set of proposed potential features. These include: service-to-service AuthN; task-specific tokens; and Passport Lite. AI agent delegation and Passport Lite were voted on for prioritisation during the group breakout discussions. Next steps include plans to produce the pull requests by GA4GH Plenary 2026, and a full v2.0 release within the next year.

Key takeaways

- The current Passports specification faces several challenges, such as JSON payloads becoming too large, resulting in “fat Passports,” lack of trust, and identity and authentication issues.
- Attendee voting for prioritised features included support for Passport Lite, as well as formalising how AI agents are identified, minted, and traced within the ecosystem.

Priority action items

- Continue to refine the Passports and AAI release timeline.

- Map priority features into the release plan, alongside the development of mini-milestones and targets for each feature.
- Revise plans for the development of Passports and AAI relative to the broader Data Security road map, incorporating session input.
- Create pull requests for several major features by 14th Plenary (September 2026), with a Passports and AAI v2.0 release by the next Connect meeting (April 2027).
- Begin planning a Data Security Work Stream Hackathon prior to Plenary 2026 (tentatively July 2026).

Provenance metadata for genomic datasets

Recording | Agenda | Slides

The session opened with an introduction to dataset provenance metadata, which refers to the contextual information about a genomic dataset's origin and production. This metadata is crucial for federated data discovery across datasets, building trustworthy data catalogues, and connecting datasets to their scientific context. Real-world examples were shared when missing provenance caused harm.

Next, Driver Projects and initiatives shared their approaches, challenges, or needs around provenance metadata. Beacon, for example, had lacked standardised dataset-level representation of provenance metadata; the new aggregated response proposal offers a way to provide summaries and provenance-related properties. The European Health Data Space will mandate the use of the Health Data Catalog Vocabulary–Application Profile (DCAT-AP) for dataset discovery, with the 1+Million Genomes Initiatives already aligning its minimal data model to reference Health (DCAT-AP). On the policy side, Data Passports was shared as a way to bundle governance metadata directly with datasets to help researchers and data access committees (DACs) make informed data access decisions. In infectious disease, PHA4GE emphasised the importance of capturing agents and processes, maintaining organisational continuity, and collecting the purpose of sampling or sequencing in the metadata. In the U.S., it was mentioned that legal compliance with the U.S. bulk data rule is needed when sharing datasets overseas. Afterwards, attendees discussed next steps towards building a minimum viable set of provenance properties that can be adopted.

Key takeaways

- Dataset provenance is a critical gap: Researchers struggle to answer basic questions about dataset origin, funding, geographic collection, publications, and contacts. This impacts discovery, trust, attribution, and equitable benefit sharing.

- No one-size-fits-all, but a minimal core is needed: multiple standards exist (e.g. W3C PROV-O, DCAT, Health DCAT-AP, schema.org/BioSchema, RO-Crate), but no single standard meets all genomics use cases. Participants strongly supported defining a minimal viable set of provenance properties that could be widely adopted, with extensions for domain-specific needs.
- The European Health Data Space (EHDS) will mandate Health DCAT-AP for describing health datasets. GA4GH should ensure interoperability with this standard rather than reinventing the wheel.
- Participants cautioned against over-engineering a standard. Starting with one or two concrete use cases (e.g. federated discovery, legal compliance with the US bulk data rule, patient transparency) will help define a practical MVP.
- Beacon currently lacks dedicated dataset-level provenance fields. The Metadata Scout initiative is a natural home for this work, but requires clear requirements.

Priority action items

- Conduct a landscape analysis — collect and compare provenance fields from existing models (e.g. Beacon, Phage, Bento, PCGL, HCA, DCAT, Health DCAT-AP, RO-Crate, ISO standards) to identify common core fields and gaps.
- Define an MVP provenance core, starting with one to two concrete use cases (e.g. federated discovery, legal compliance) to define a minimal viable set of dataset-level provenance properties that can be widely adopted.
- Given broad interest and the cross-cutting nature of this work, form a small group (building on the Beacon Metadata Scout and interested participants) to carry forward the landscape analysis and MVP definition.

Connect Day Four

Register-a-thon

Recording | Agenda | Slides

The aim of the meeting was to present the latest GA4GH Registry tooling to the community and explain the process for registering and maintaining service entries. The session included a practical walkthrough of the registration workflow, with hands-on support to help onboard as many Driver Project implementations into the registry as possible. The session's goal was to produce an updated GA4GH map reflecting the newly registered services and overall progress achieved during the session.

Key takeaways

- GA4GH Registry is a central hub for identifying and connecting global service deployments worldwide.
- Proven success: previously identified 23 DRS servers across 4 countries holding 72+ petabytes of biomedical data.
- Sustainable tools: moving from ad-hoc collections to permanent, user-friendly self-service registration tools.
- Quality control: data flows from staging to production through hand-curation to maintain integrity.
- Global expansion: registry now includes deployments across Asia, Europe, and multiple server types.

Priority action items

- Register GA4GH service instances using the self-service registration tool during and after the event.
- Hand-curate data from staging registry to production registry to ensure security and quality control.
- Maintain the interactive registry map displaying all registered instances.
- Provide remote support for online participants needing help with registry registration.
- Continue developing and improving registry tooling and public website integration.
- Promote the permanent registry URL as the ongoing resource for GA4GH instance registration.

Structural, categorical, and long-read-derived variation in Beacon: challenges and requirements for Beacon v3.0

Recording | Agenda | Slides

The session opened with an introduction to Cat-VRS and its constraint-based modeling approach, followed by an introduction to Beacon's current query landscape and the need for tighter alignment with VRS/Cat-VRS. The discussion highlighted Beacon’s permissive node model, slow VRS adoption, and the aspiration to prioritise SVs, fusions, and adjacencies for Beacon v3.0.

Participants then walked through concrete query examples — from simple indels and CNVs to complex fusions, cytoband-based queries, and repeat expansions. Key discussion points included:

- the ambiguity of current query parameters and the value of request profiles to constrain and standardise queries;
- how Cat-VRS can serve as a translator between fuzzy front-end concepts and precise back-end representations, especially for cytobands, gene-level queries, and high-level copy number events;
- challenges with repeat expansions: VRS reference length expressions provide one approach, but participants noted the need to handle non-canonical repeat sequences, VNTR-internal STR variation, and multi-level fidelity;
- the multivariant query gap: Beacon cannot return samples with multiple specified variants — while workarounds exist via record-level responses and multiple requests, a standardised approach is needed;
- GET vs. POST for complex queries: Participants discussed the trade-off between human-readable GET requests and the flexibility of POST for nested/structured query objects.

Key takeaways

- Adjacencies and structural variants are top priorities for Beacon v3.0: fusions, translocations, and other SVs remain difficult to query in Beacon V2. The Cat-VRS adjacency class supports both precise VRS locations and broader mappable concepts (gene names, cytobands, unspecified elements), capturing real-world data fuzziness.
- Normalisation remains a major barrier: Beacon v2.0 is permissive by design, but this leads to inconsistent data across nodes. VRS v1.3 adoption has been slow. Collaborative development of data translators and clearer normalisation guidance is essential.
- Complex queries need structured solutions: multivariant requests, repeat expansions, and systemic events (chromothripsis) are not yet well-supported and require joint design work.
- Long-read data amplifies representation challenges: different technologies produce systematically different breakpoint calls and representations. Normalising across these remains an open problem.

Priority action items

- Launch a joint pilot/working group to develop and test new VQS request profiles for adjacencies, fusions, structural variants, and repeat expansions.
- Co-develop standardised query parameters for Cat-VRS concepts in Beacon v3.0, including naming conventions for adjacency sides, order semantics, and repeat representation.
- Promote GKS tooling and normalisation guidance to help Beacon nodes adopt consistent VRS/Cat-VRS representations and reduce implementation barriers.

- Solicit use cases and requirements for multivariant/haplotype queries and systemic events (e.g. chromothripsis) to inform future Beacon and Cat-VRS roadmap decisions.

The status of neuroscience data sharing in 2026: data governance, tooling, and education

Recording | Agenda | Slides

In this session, the community leads highlighted the work-to-date and next steps for each of the three theme teams. JB Poline discussed the Data Harmonisation team's living, community-driven resource that documents data models, ontologies, and terminologies, helping to create a "common language" across datasets. The purpose of this resource is to help researchers choose and apply standards, while also enabling interoperability between siloed systems. Poline emphasised greater participation and engagement and hopes the community can serve as a forum, run AI seminars, and create pilot use-cases. Kim Ray, of the data governance team, shared updates on their landscape analysis of over 30 global documents (e.g. data use agreements and management plans). The gaps, inconsistencies, and common elements that the team identifies will help support work in data harmonisation. The team is planning sessions with representatives from data repositories to see how different groups handle governance. Francis Jeanson represented the Scientific Collaboration and Education team and highlighted the *Neuroscience Educational Readiness Checklist (NERC),* a tool to help institutions assess and prepare effective training programs. The team's manuscript for scalable, adaptable, training resources to support large-scale brain data science is near publication.

Overall, the session demonstrated how data governance, harmonisation, tooling, and education are interdependent. Better data standardisation enables more consistent governance, which in turn enables scalable data access and analysis. The community will look at creating a living resource to collate all the standards and resources across all three teams, with separate catalogues for different resource types (harmonisation, governance, tooling, training etc.). While a spreadsheet can serve as a starting point, the community must also address maintenance responsibilities and funding.

Key takeaways

- Standards need to be paired with tools for annotation, search, and data exchange and the outputs should be accessible for real research workflows (e.g. querying cohorts, running analyses).
- It may not be necessary to develop entirely new frameworks for data sharing; standards for clinical and genomic data can apply to neurodata.
- Emphasise automated data annotation (over panel automation) to reduce human error, inconsistencies, and researcher burden.

- While the community should agree on standards, institutions must help support their implementation.
- Future projects to explore how AI can be used to standardise more complex, inconsistently labelled data can be achieved via use-cases, common data elements, and knowledge graphs.
- Institutional risk aversion, an increase in sovereignty laws, privacy and re-identification risks, and legal complexities in international sharing are the main challenges driving data governance work.
- Integrate governance earlier on in the data lifecycle to improve efficiency and avoid issues with data release downstream.
- There is a need for a common global data cataloguing approach (standardising naming, clear definitions, how different data types relate) to allow government frameworks to interoperate.
- To enhance adoption for non-technical researchers, map training resources to objectives and skills and provide more visibility to who is producing data, where it resides, and what policies are being applied. Map the ecosystem before building new solutions.
- Future developments may require stronger safeguards for neurological data (i.e. mental privacy, personal identity).

Priority action items

- Kim Ray: reach out to representatives from databases and data collection groups to present their data governance practices at future meetings.
- Community: contribute to the development of a living resource (e.g. spreadsheet) for collating and cataloguing tools, resources, and standards across the three Theme Teams (i.e. governance, harmonisation, education).
- Community: determine which approach is more beneficial for data standards development — a top down approach (data governance leads) or a bottom up approach (data harmonisation leads, governance feeds into it)
- Leadership: Consider how the Rare Disease Community can collaborate with the Neuroscience Community to create a beginner's toolkit of available resources.
- Leadership and GA4GH support: explore how the GA4GH technical and leadership teams can help with hosting and maintaining a catalogue of neuroscience resources

Building a community road map for Beacon v3.0

Recording | Agenda | Slides

There was an introduction to the purpose of a third version of Beacon to add additional aspects, including from community feedback and relevant requests.

The “OR” operator was presented with a concrete syntax: nested arrays inside the filters array represent OR, while the top level is AND. Beacon’s response makes the interpretation explicit, disambiguating queries that would otherwise be ambiguous for human readers. The group noted that the current design follows Disjunctive Normal Form (DNF), which is clearer for clinical users.

Rounding counts were introduced as a response to rare-disease community requests. Instead of a single integer, a beacon can return a string-based bucket or a range. This allows privacy-preserving counts while still providing useful signals.

A large part of the session was devoted to integrating RNAget into Beacon. Emilio Righi has been bridging the two communities. Beacon gains the ability to query expression matrices, genomic features, and expression values; RNAget gains Beacon’s rich filtering and context (individuals, datasets, cohorts). Participants suggested renaming “expression values” to a more generic term (e.g. “signal values”) to cover other -omics technologies.

Michael Baudis demoed the aggregated granularity feature, which is already running in a production beacon (Progenetics). It supports one-dimensional counts (e.g. disease distribution) and two-dimensional cross-tabs (e.g. disease vs. sex). The implementation uses a flexible “splits” mechanism for numerical values (age buckets). The demo showed live queries with filters (e.g. CDKN2A deletion + glioblastoma) returning aggregated counts rather than individual records.

Later topics included multivariant requests (no solution yet, but recognised as necessary), the deprecation of OpenAPI (to be replaced either by generation tooling or removed entirely), and the status of the Global Beacon Atlas. The session ended with a call for community implementation of the aggregation features while they are still in draft, so that real-world feedback can shape the final specification.

Key takeaways

- “OR” operator will be added using nested arrays (DNF style), resolving long-standing ambiguity.
- Rounding counts enable privacy-preserving responses (e.g. “1-5”, “45k-55k”), answering rare-disease requirements.
- RNAget integration brings expression matrices, genomic features, and expression values into Beacon (e.g. APOE expression in Parkinson’s brain).

Priority action items

- Finalise aggregation PR — resolve discrepancies (use of filtering terms, binning) and merge.

- Implement OR operator in reference implementations to validate syntax and response disambiguation.
- Draft multivariant query proposal — explore accession ID mechanism for anonymous result set requerying.

Integrating GA4GH APIs

Recording | Agenda | Slides

This session explored the integration of Cloud and other GA4GH APIs. The group looked at pain points and rough edges to identify work in 2026 to improve integration within and across Work Streams.

Key takeaways

- An artifact documenting some key integration points and fixes that need to be done, who needs them, and when they are needed should be shared with the Work Streams and Driver Projects.
- Key accomplishments from the Federated Learning hackathon were noted, especially as it relates to facilitating the broader topic of API integration.

Priority action items

- API Integration and WES Development
    - Lead WES-TES-DRS integration efforts with passport delegation
    - Implement GA4GH In Focus project for WES-TES-DRS integration testing
    - Develop Google Summer of Code project on WES/TES log parsing and self-healing workflows
- Logging and standardisation
    - Investigate OpenTelemetry integration for standardised WES and TES logging
    - Synchronise WES and TES logging schemas
- Service Info and Metadata
    - Expand Service Info with structured capability descriptions (GPU/CPU specs, geolocation)
    - Migrate Service Info repository under GA4GH organisation
- Federated learning and policy
    - Establish monthly Policy Working Group meetings for federated learning governance
    - Document integration pain points and API improvement ideas in shared notes
- Collect and prioritise API integration feedback from participants for Cloud Work Stream

GKS road map

Recording | Agenda | Slides

The session opened with Alex Wagner (GKS Work Stream Co-Lead) framing the shift from innovator-focused work to early-majority adoption, highlighting the need to break down engagement barriers. Kori Kuzma reported on the GKS Think Tank series, Melissa Cline (GKS Work Stream Co-Lead) presented hackathon outcomes, and Larry Babb then outlined the implementer-driven roadmap process, emphasising that producers are now feeding consumers, and that the Work Stream's priority is unblocking active implementers.

Product-specific updates:

- Variant Representation Standard (VRS) survey showed top requests: intronic variants, phased genotype models (2nd), and pharmacogenomic variation. Implementers want example data and how-to guides more than a maturity dashboard.
- Categorical Variant Representation Specification (Cat-VRS) survey highlighted oncogenicity/pathogenicity and co-occurring categorical variants as top new features. Notebook-based examples and recipe-oriented tutorials were strongly requested.
- Variant Annotation Specification (VA-Spec) hackathon discussion identified five areas, with unanimous support for simplified, non-redundant data exchange patterns ("normalised" documents), followed by reference tooling and concept diagrams. This led to the Genomic Knowledge Starter Kit proposal.

Breakout groups identified major barriers (lack of real-world examples, complex dependencies, need for evangelism tools) and confirmed the Starter Kit as the highest-priority road map item.

Key takeaways

- Documentation and onboarding are the top barriers to entry. New users consistently highlighted the lack of real-world examples, jargon-heavy specifications, and the overwhelming mental load.
- Community priorities are shifting from "building the spec" to "making it usable". Survey results and breakout feedback strongly favour example datasets, how-to guides, simplified (non-redundant) data structures, and reference tooling — moving the Work Stream from early adopters toward the early majority.
- A "Genomic Knowledge Starter Kit" is the new flagship objective. The Work Stream will create a centralised collection of profiled, idealised example data documents with accompanying tooling and concept diagrams to lower the adoption barrier.

Priority action items

- Develop the Genomic Knowledge Starter Kit: product teams to build profiled, idealised example data documents (VA-Spec, Cat-VRS, VRS) with reference tooling, to be showcased at the October Connect.
- Launch an Asia-Pacific-friendly Think Tank: coordinate a new monthly slot to enable global participation; spread the word to colleagues in that region.
- Improve onboarding documentation: create a glossary of terms, plain-language concept descriptions, and use-case-driven "recipe" tutorials (e.g. "how to represent a canonical allele").

GIF Project showcase: progress, activities, and connections

Recording | Agenda | Slides

The GA4GH Implementation Forum (GIF) focuses on bridging the gap between developers of GA4GH standards and real-world users. The group is guided by three main pillars, implementation, interoperability, and inclusivity. Over the last year, GIF welcomed two new co-leads, Brian O'Connor and Wafaa M. Rashed. GIF also launched the *Spotlight* programme and hosted three Ask Me Anything events (Phenopackets, Crypt4GH, and Passports). There are currently four GIF projects underway (Federated VLM, Crypt4GH, TRE-Open Suite, and eLwazi).

Heidi Rehm and Nara Sobreira presented the Variant Level Matching (VLM) platform, which aims to connect rare disease databases through a federated network to allow identification of cases with specified variant types and associated phenotype. Muhammad Usama Sardar highlighted the progress the TRE-Open Suite project has made via several collaborations – including Internet Engineering Task Force (IETF), Internet Research Task Force (IRTF), and Confidential Computing Consortium (CCC) – on attested TLS and remote attestation primitives. Mamana Mbiyavanga discussed the eLwazi platform, highlighting how it connects and scales genomic data analysis across Africa using GA4GH standards as a foundation. The conversation ended with an interactive discussion where participants shared ideas for future GIF projects, spotlights, and AMAs, with particular interest in collaboration between GA4GH and NIF efforts.

Key takeaways

- *Seqr* (Gene43, Genomics4RD) and *HIFI Solves* (DNAstack, PacBio) databases were recently approved to integrate into the VLM platform. *Breakthrough Genomics* is pending review.
- Standardisation of the main match endpoint enables more nodes to join and share data via VLM; integration of Phenopackets allows for a detailed endpoint (patient symptoms and clinical features as opposed to just genomic data).

- VLM is focused on improving tooling for data harmonisation and privacy-preserving genomic hashing to identify duplicates.
- The TRE Open suite team is working to ensure data security between data sources and cloud providers through attested TLS; a previous design was found to be vulnerable with a high-severity CVE (CVE-2026-33697) of CVSS score 7.5.
- The TRE Open suite team requested support from the community, including a liaison statement and call for adoption for the SEAT working group proposal (draft-fossati-seat-expat), to move the TRE Open Suite work forward with IETF.
- The eLwazi team rebuilt their imputation service using GA4GH standards and enabled a federated architecture. The system integrates standardised APIs for workflows, federated identity and access control (passports/visas), metadata querying, workflow execution services, and provenance tracking.
- Next steps for the eLwazi project include: a second federation node (e.g. Mali, Uganda), production launch, full deployment of H3Africa v7, service information extensions, workflow expansion via TRS, and RO-Crate.

Priority action items

- TRE Open Suites: follow up with Cloud Work Stream regarding whether the proposed attested TLS solution will require changes to GA4GH APIs, and provide details on implementation plans in relation to GA4GH projects and APIs.
- eLwazi: deploy the federated imputation platform to additional nodes.
- eLwazi: consider and decide on the architecture for Data Connect instances (single vs. per-node) for the federated network and implement accordingly.
- eLwazi: explore using RO crates with DRS/TRS for automated provenance management in the imputation workflow.
- All interested parties: submit ideas for future GIF projects, spotlights, or AMAs or follow the process to apply for new GIF projects or spotlights.
- GIF leadership: review and process incoming ideas for GIF projects, spotlights, and AMAs from the Slido and follow up with submitters as needed.
- GIF leadership: receive and process incoming spotlight submissions and coordinate with the GA4GH Comms Team for publication.

This meeting report was developed with the aid of AI.