

1. Best practices & Open terminologies

- Delivering an environment for open terminology work - terminology mapping, FHIR server, tooling support - e.g., Redcap

2. Patient-level phenotype exchange

- Phenopackets exchange standard - moving phenotypic info around like genomic info, in partnership with HL7 and ISO/TC215 (Health Informatics) to get data into and out of EHR

3. Family health history capture and interoperability

- Better management of pedigree data and how we use it (long-term goal of propagating info between family members)

Family health history capture and interoperability

- Better management of pedigree data and how we use it (long-term goal of propagating info between family members)
- What's in and out of scope?
 - **family history** and/or **pedigree** (family structure)? **scope**: common schema, APIs, best practices? **priorities**: EHR, patient, CDS, genomic tools?
- New Product Proposal being prepared
 - **pedigree**, common schema, APIs, best practices
- Driver Projects: Survey on use of Pedigree Information is coming your way soon!

Patient-level phenotype exchange: goal is freely exchangeable “FASTA for phenotypes” <https://phenopackets-schema.readthedocs.io>

- Phenopackets exchange standard - moving phenotypic info around like genomic info, in partnership with HL7 and ISO/TC215 (Health Informatics) to get data into and out of EHR
- Phenopackets v1.0
 - Great work from Jules, Peter etc to advance towards a v1.0
 - Discussed what’s in and out of scope
 - Interoperability with FHIR was seen as a key feature
 - Proposal for standard in June
- All: Drivers and other interacting workstreams please review and make requests

CONTENTS

A short introduction to protobuf

Phenopacket basics

Requirement Levels

Phenopacket building blocks

▢ Phenopacket

Family

Cohort

Biosample

Working with Phenopackets in Java

Examples



Move fast and fix things!

Resolve production errors quickly, and
deploy code with confidence. Give
Rollbar a try.

Sponsored · Ads served ethically

📖 Read the Docs

v: latest ▾

Docs » Phenopacket

[Edit on GitHub](#)

Phenopacket

A Phenopacket is an anonymous phenotypic description of an individual or biosample with potential genes of interest and/or diagnoses. It can be used for multiple use cases. For instance, it can be used to describe the phenotypic findings observed in an individual with a rare disease. The phenopacket can contain information about genetic findings that are causative of the disease, or alternatively it can contain a reference to a VCF file if exome sequencing is being performed as a part of the differential diagnostic process. A Phenopacket can also be used to describe phenotypic abnormalities observed in a genetically modified animal model, or to describe the constitutional phenotypic findings of an individual with cancer (a [Biosample](#) should be used to describe the phenotypic abnormalities directly associated with an extirpated or biopsied tumor).

```
message Phenopacket {
  string id = 1;
  org.phenopackets.schema.v1.core.Individual subject = 2;
  repeated org.phenopackets.schema.v1.core.Phentype phenotypes = 3;
  repeated org.phenopackets.schema.v1.core.Biosample biosamples = 5;
  repeated org.phenopackets.schema.v1.core.Gene genes = 6;
  repeated org.phenopackets.schema.v1.core.Variant variants = 7;
  repeated org.phenopackets.schema.v1.core.Disease diseases = 8;
  repeated org.phenopackets.schema.v1.core.HttpFile hts_files = 9;
  org.phenopackets.schema.v1.core.Metadata meta_data = 10;
}
```

id

Phenopacket structure



patient:

id: "PROBAND#1"

sex:

id: "PATO:0000384"

label: "male"

V1 finalized at this meeting!!!



who?



Global Alliance
for Genomics & Health

phenotypes:

- type:

id: "HP:0000520"

label: "Proptosis"

severity:

id: "HP:0012828"

label: "Severe"

classOfOnset:

id: "HP:0003577"

label: "Congenital onset"

what?

how?

when?

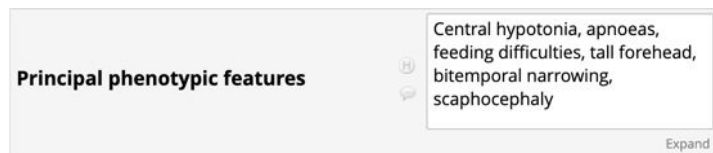
Best practices & Open terminologies

- Delivering an environment for open terminology work - terminology mapping, FHIR server, tooling support - e.g., Redcap
- Discussed Genomics England and Australian Genomics exemplars
 - great interactions with new and existing Driver Project

Clinical & Phenotypic Data Capture

Are you using REDCap?

- REDCap now supports capturing standard terminologies via an external FHIR Terminology server - support for HPO, SNOMED CT, OMIM other



Principal phenotypic features	Central hypotonia HP:0011398 HPO terms
Principal phenotypic features	Apnea HP:0002104 HPO terms
Principal phenotypic features	Feeding difficulties ... HP:0008872 HPO terms
Principal phenotypic features	Bitemporal hollow... HP:0025386 HPO terms
Principal phenotypic features	Scaphocephaly HP:0030799 HPO terms

Get the REDCap plugin here:
https://github.com/aehec/redcap_fhir_ontology_provider

Try Ontoserver here: <http://genomics.ontoserver.csiro.au>

Other potential products useful for Driver Projects

- Data standards
 - **Acute Care data set specification** - Australian Genomics and Genomics England
 - FHIR Implementation Guide
 - » FHIR Profile, extension, code systems and value sets (HPO, SNOMED CT)
 - Supports clinical implementation and research
 - Phenopackets / FHIR Implementation Guide

Cloud Overview: before (Monday AM)

	Champions	Purpose	Status
Data Repository Service (DRS)	Brian O'Connor David Glazer	map a logical ID to a physical means of retrieval	v0.1 → v1.0
Workflow Execution Service (WES)	James Eddy	send the algorithm to the data	v1.0 → v1.1
Tool Repository Service (TRS)	Denis Yuen Susheel Varma	share workflows and tools	2.0.0-beta.2 → v2.0
Task Execution Service (TES)	Angel Pizarro Susheel Varma	run a single job	v0.3 → v1.0

Cloud Overview: after (Tuesday PM)

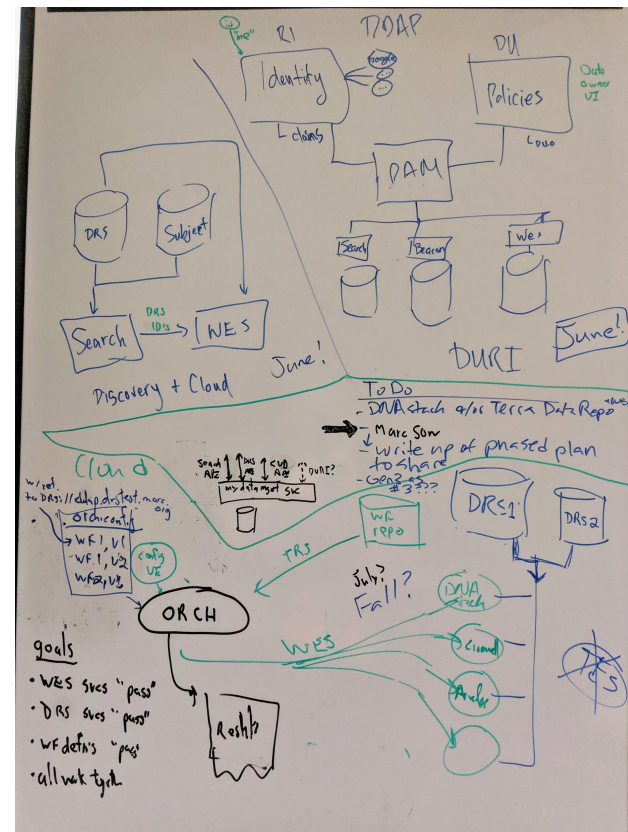
API	Work to be done
DRS → 1.0	URI schema (PR #261)
	Merging object/bundle into single endpoint/response (PR #258)
	File pairs/secondary file support and effects on URI schema
	Retry-after support
TRS → 2.0	Nothing critical for 2.0, just official docs and submission process
WES → 1.1	Specify auth headers for DRS inputs in the workflow run request
	Light changes based on driver feedback (e.g. GEM Japan)
TES	Will continue to evolve, <i>not going for 1.0 this year.</i>
WES + TRS + DRS	1) libraries/local testing ability 2) centralized testbed
General	Driver use cases on ga4gh.cloud, list of implementations of each API and links out. Helping new drivers with available solutions and use cases for our APIs.

Cloud Flow: Search + Cloud WES/DRS + DURi



Global Alliance
for Genomics & Health

1. Test user starts the flow
2. Testbed calls search endpoint, passing the OIDC auth token for the user
3. Search returns a list of DRS URIs
4. Testbed passes workflow, DRS URI, and user creds to WES endpoint
5. WES fetches the input data, having first obtained appropriate credentials
 - multiple WES/DRS/DURi interaction options
6. WES runs the workflow



Cloud Drivers: multiple configurations

- private environment; federated nodes of data; each node runs WES to compute on its local data (e.g. SPHN, ICGC Argo)
- private environment; one or more WES endpoints that run jobs across data at one or more locations (e.g. Elixir)
- wide-audience data provider; central WES endpoint that runs jobs across its own data, possibly in many locations (e.g. DataStage)
- multi-tenant environment; central WES endpoint that runs jobs across data in many locations, much of which is controlled by other entities (e.g. Terra, Illumina, Elixir)

Meeting goals

1. [Researcher Identity specification v1 DRAFT RFC presented](#)
2. Gathering feedback from DPs, Cloud, Discovery, Security
3. RI and DUO integration discussion
4. Technical demonstrations Google in Researcher Identity stream. Claims lab to capture new claims
5. DURi Roadmap until 4/2020

Meeting outcomes and next steps

RI spec -- One issue identified and fixed: persistent identifiers for organisation (s) -> **experts recommend to freeze the spec for technical implementations** by Google and ELIXIR. Not all cohorts will be able to utilise RI spec v1 claims

[Driver Projects and the RI and DUO interactions](#) on Researcher Identification, shared data resources and their granularity and data access committee structures. List of DPs interested in DURi: **CanDIG, ASI, EpiShare, GEL, NCI, VICC, EJP-RD, Australian genomics, EGA, Beacon, ELIXIR, EGA, H3Africa, AMED/GEM, ICGC/ARGO** (in addition to existing interactions)

Technical demonstrations Google in Researcher Identity stream. [Claims lab](#) deep dived to technology to identify additional claims (for v2).

Pilot ideas emerged: e.g. NCI relying on ELIXIR AAI, RI spec technical implementations by Google and ELIXIR, [DUO – RI integration example with the real DAC](#) (with the EGA), RI product approval process in sync with Security WS AAI

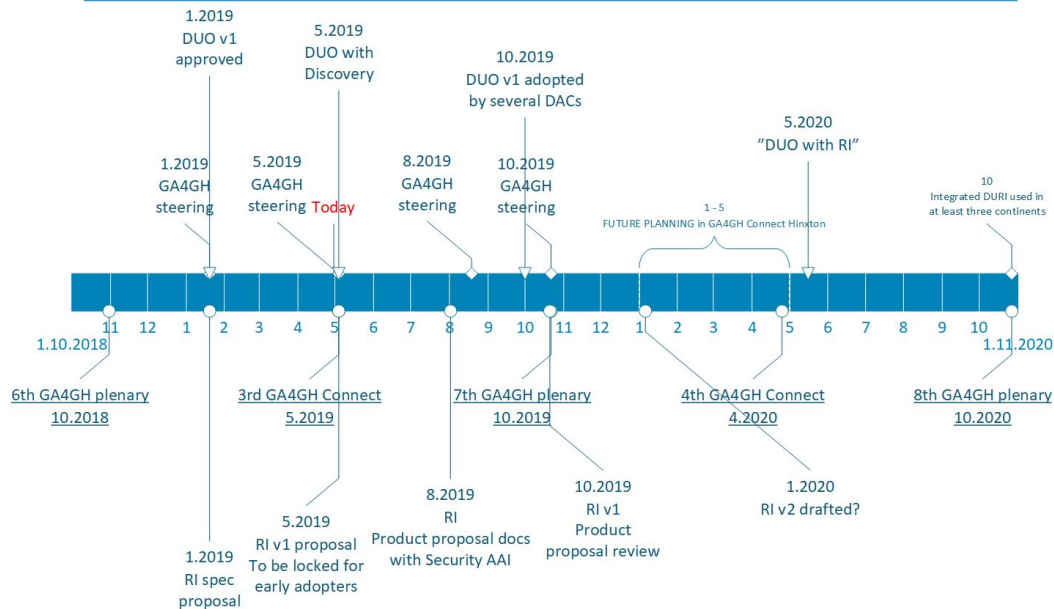
DURI roadmap goal suggestions

- Refactor the ga4gh namespace so that (signed) claims can come from multiple sources without collision
- Explore workflows for addressing contractual/legal/liability needs
- Agree on a worked out use case incorporating RI + DU (eg EGA THL dataset, Broad DUOS)
- Integration demo with WES/DRS
- Implementation showing DUO and RI brought into DAC evaluation
- Have a draft v1.1 RI spec WIP to share in May 2020 and frozen in October 2020
- Integration of new needs from CanDIG and NIH to the next RI claims version
- Modularisation of DU and RI components so that they can be integrated quickly into existing systems engagement with clinical driver projects to develop patient accessible DUO codes
- Plain language DUO
- Consider a shallow evaluation of driver project use cases and legal frameworks to inform long-term strategy. “Shallow” is intended to reflect an Minimum Viable Product approach.

-> next DURl steering meeting will update roadmap to cover 4.2019-4.2020

DU

Data Use -- Melanie Courtot, Jonathan Lawson coordinate in Monthly Telco
Key contributors:
Products: Data Use Ontology DUO v1



RI

Research Identity - Ilia Tulchinsky and Stephanie Dyke coordinate in monthly telco
Key contributors: Craig Voisin, Mikael Linden, Alexandra
Products: Researched Identity RI v1 (DRAFT)

Dugnaðr!



HEALTH



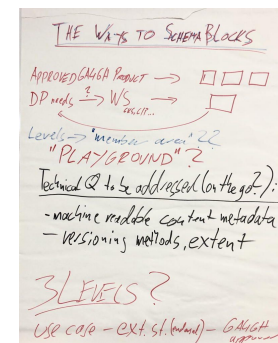
Beacon {↑}

- Beacon API
 - presentation of Beacon v2 prototype
 - filters for phenotypes, sample specs.
 - access levels
 - focused queries & responses
 - handover => data access
 - keeping a “simple” concept compatible to the original Beacon concept
 - “clinical” targeted, with demonstrated use cases => broaden engagement, uptake
- Beacon in Discovery
 - engagement of driver projects
 - evaluation of new Beacon concepts
 - monthly open meetings + DP invites, engagement
 - first target: *EvidenceBeacon* concept



SchemaBlocks {S}[B]

- code, standards for cross-project uptake
- primary development in WS & DP products
- broad consent about usefulness of concept
 - several soft commitments for contributions :-)
- converging on technical representation (demo of JSON Schema implementation example)
- emerging new use cases from DP + WS
 - technical blocks (sugg. by EpiShare...)
 - query elements
 - definition of ontology scoping
 - Beacon v2 query types
- levels of recommendation
- present update at Sep SC meeting

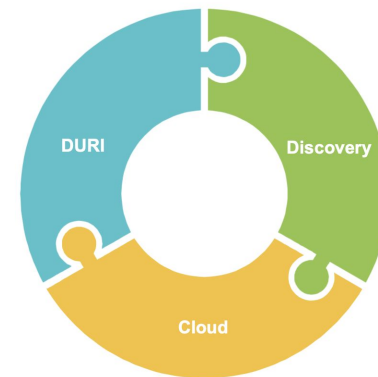


Networks: registry of GA4GH services

- Identified outstanding features and issues
- Developed a plan to drive towards product approval with support by ELIXIR as the primary driver (submission TBD, some time in 2019)

Search: framework for multi-modal search

- Aligned on a common goal, good driver engagement
- Clarified scope:
 - We should not define **data models** themselves, defer those activities to rest of GA4GH
 - We should develop **meta-models**, i.e. a way to list what is searchable;
 - We have two proposals for **how to search**, with mostly syntax differences, plans to drive convergence through structured analysis
- We have an architecture diagram for the integration of Discovery (**Search, Beacon**) + DUR (RI) + Cloud (**DRS, TRS, WES**); these interactions make Discovery actionable!



Variant Representation (VR) and Annotation (VA)

- Product review submission: Allele model
 - v1 will support small variations only (SNVs and indels)
 - Minimal spec, subset of current work
 - Schema-only submission (no API at this time), working through process requirements
 - June submission to target September SC meeting (tentative) [lag?]

Andy Yates, Bob Freimuth

Matt Brush, Javi Lopez, Alex Wagner, Larry Babb, Reece Hart

Variant Representation (VR) and Annotation (VA)

- Product review submission: Allele model
 - v1 will support small variations only (SNVs and indels)
 - Minimal spec, subset of current work
 - Schema-only submission (no API at this time), working through process requirements
 - June submission to target September SC meeting (tentative) [lag?]
- Joint meeting with Clin/Pheno
 - Coordination: scope of work, use of VR/VA models within Phenopackets
- Drop-in requirements from LSG
 - Applying VA work to support VCF annotation
 - Scope of VA: genes, sequences?
- Future
 - Expansion and maturation of VR and VA models (DP speed dating)
 - Ongoing implementations of GKS products by several DPs

Variant Representation (VR) and Annotation (VA)

- Product review submission: Allele model
 - v1 will support small variations only (SNVs and indels)
 - Minimal spec, subset of current work
 - Schema-only submission (no API at this time), working through process requirements
 - June submission to target September SC meeting (tentative) [lag?]
- Joint meeting with Clin/Pheno
 - Coordination: scope of work, use of VR/VA models within Phenopackets
- Drop-in requirements from LSG
 - Applying VA work to support VCF annotation
 - Scope of VA: genes, sequences?
- Future
 - Expansion and maturation of VR and VA models (DP speed dating)
 - Ongoing implementations of GKS products by several DPs

Enterprise-Level Questions

- github: need policy about repository usage (working, pre-approval space vs. main repository)
- Terminologies and value sets
 - Registering, sharing, management and governance



Global Alliance
for Genomics & Health
Collaborate. Innovate. Accelerate.

Large Scale Genomics



File Formats

James Bonfield, Louis Bergelson

- Driver project: how to protect against accidental DDOS on refget server?
- Storing base modifications in SAM/BAM/CRAM; useful links to key people.
- Storage of vendor specific data (eg ONT signal), ongoing work.
- CRAM 3.1 presentations.
 - Spec as text or ref implementation?
 - Standard library of codecs to avoid *requirement* to reimplement



Future of VCF

Yossi Farjoun, Cristina Gonzalez

- From the driver projects, most pressing issues are:
 - **Scaling** of callsets of large numbers of samples
 - Arbitrary **subsetting** of large callsets
- Discussed a proposed solution:
 - Use “**look up**” notation for reference blocks (+ checkpointing)
 - Use “**local alleles**” for preventing the PL’s from exploding (with non-breaking and breaking options)
 - Optionally store with **Tachyon/Pil** for columnar compression + random access.
- Arising need: **base modification** in VCF
- Discussed proposed notations for **sample filtering** and high-ploidy **phasing**.



Encrypted Container Format

Alexander Senf, Rob Davies

Workshop discussions:

- Presented Crypt4GH Design Goals and current Specifications
 - Enable tools (htsjdk, htslib, ..) to read/use encrypted data natively both from files and via htsget.
- Potential new attack vectors discussed
 - Possibility to reorder/delete parts of a file undetected.
- New Driver Project interest: H3Africa
 - Addresses a real need. Interest to test any products released in this context.



RNASeq

Roderic Guigo, Barbara Wold, Sean Upchurch

- Add continuous endpoint to deliver signal data; bigwig matrix
- Driver metadata survey
 - goal: determine intersect and common naming
- Discussion of search filters - trim to “highly relevant” set
- Starting up EBI/Expression Atlas implementation



htsget

Mike Lin, Jerome Kelleher

Supporting driver project deployments of mature spec; represented here:

In use- EGA/Elixir

Eng- AGHA, GeL

Scoping- CanDIG, H3Africa

Applications: genome browsers, analytics pipelines & notebooks, data distribution

Minor features & improvements to spec in process



Refget

Andy Yates

v1 targets: AWS public data set load (early July), AWS serverless, application note

v2 targets: Alias to checksum and sequence collections on the roadmap for late 2018 first draft

Concurrent: Discovery network integration. How to integrate dependent schemas/endpoints (/service-info). How to use htsgget to get around issues of the spec and HTTP redirects

<https://docs.google.com/document/d/1uBP8SBrE1CZxfdhaSL6SQRrLGsEhWPsjqJzF3c5bIFk>

Data Security

- **Update of the security review process**
 - **Revised security questionnaires for proposal and approval**
- **Update on the security infrastructure document**
- **Progress on Breach Notification and Response**
- **Progress on Authentication and Authorization Infrastructure**
- **GenoPri'19**



Product-Proposal and Approval Security Questionnaires

- **Positive reception of the more product-tailored security review process**
- **Consensus on the new structure, more guided questions and OWASP-based risk analysis**
- **Next steps**
 - **Consolidate the questionnaires and share them with the product development teams**
 - **Identify security liaisons with technical work streams and driver projects, as the main communication channel with the DSWS**
 - **Set up periodic meetings between DSWS and technical work streams to follow security considerations on products under development**

Security Technology Infrastructure

- Updated references to the *GA4GH Authentication and Authorization Infrastructure (AAI)* and the *Breach Response* standards (both yet to be approved).
- Added “Open Web Application Security Project (OWASP) top 10 security risks”
- Forward-looking approach to technological privacy-protection: Homomorphic Encryption, Secure Multi-party Computation (SMC), Post-Quantum cryptography
- Broadened the scope and increased the generality of the recommendations and guidelines
- Updated references and added recent attacks to -omics systems.
- Next steps
 - Add explicit recommendations to Cloud security, coordinated with REWS and Cloud
 - Revise the allocation of stakeholder responsibilities according to the Breach Response policy
 - Start a 1-month review process before the document is consolidated
 - Align with updates from the privacy and security policy

Data Breach Process Update (AGHA team)

- **GA4GH Data Breach Reporting Protocol**
 - Decision tree identifying conditions under which reporting to GA4GH is needed
 - Definition of Accountable and Responsible actors within GA4GH
 - Definition of the information to be provided in the breach report
- **Draft GA4GH Breach Response Plan**
 - High level decision process for the identification of the to-be-informed agencies, definition of agreements on external communications and task group actions and mitigations
- **Ongoing discussions and next steps**
 - Vulnerability management (NIST-based)
 - Data Breach Drill Kit and Breach Management Best Practice Guide
 - Survey to identify requirements (we welcome responses from all driver projects)

<https://is.gd/GA4GHdatabreachsurvey>

Authentication and Authorization Infrastructure

- Collaboration with REWS, Cloud, DUR, LSG
- Draft spec available and almost finalized for AAI and Identity Claims: [AAI/RI](#)
- Ongoing discussions on
 - Frameworks and identification of responsibilities in federated identity/claims systems
 - Need of persistent identifiers for claim sources, to enable audit trails and provenance tracking (responsibility of the implementers in v1)
- Next steps
 - Finalize the specification with the feedback received from driver projects
 - Open source example implementation, released for trials in summer, conditioned on the evolution of the standard
 - Follow-up with the interested driver projects for additional implementations of the AAI
 - Identify frameworks to deal with identity breaches in federated identity systems
 - Follow-up synchronization meeting scheduled in May

6th Workshop on Genome Privacy and Security, GenoPri'2019

- <https://genopri.org>
- Main venue on genome privacy and security
- Co-located with GA4GH 7th Plenary Meeting
- Brings together:
 - Computer Scientists
 - Bioinformaticians
 - Ethicists, Legal Scholars, Social Scientists
- **Contributions from GA4GH WS and DP are particularly welcome and encouraged**
- **Deadline for talk proposals: July 19, 2019**



[Home](#) [Call for Talks](#) [Organizers](#) [Program](#) [Local Info](#) [Past Events](#)

6th International Workshop on Genome Privacy and Security (GenoPri'19)

October 21-22, 2019, Boston, USA

Co-located with [GA4GH 7th Plenary Meeting](#), The Hynes Convention Center, Boston, USA

October 21-23, 2019

#genopri19

What we discussed

- Adapting data sharing to the **EU *General Data Protection Regulation***
 - The Human Cell Atlas: A Legal Basis for Data Sharing - public interest? consent? already made public?
 - Updates to GA4GH [Privacy and Security, Consent Policies and Templates](#) (by October)
- **Goals of the GA4GH IP Sub-group, under an iterative process**
 1. Ensure GA4GH products are shared under open licenses.
 2. Ensure GA4GH has clear right to member contributions under copyright law.
 3. Address the use of GA4GH Trademarks.
 4. Ensure standard-essential patents from GA4GH members are openly licensed, so they do not impede adoption.
 5. Outline a consensus process for developing final IP policy.

What we discussed with Security workstream

- Approval of the draft GA4GH [Policy Framework](#) Approval Process in joint REWS/Security meeting

REWS next goals

- “Sense checking” Framework and Policies (Indigenous, socio-cultural)
- Data Steward/Trust concepts
- Lexicon & the GDPR
- Cloud Policy

Meeting: 2019 October meeting + GenoPri (present updated Privacy & Security Policy)

- Organize our potential contributions to the main meeting



Global Alliance
for Genomics & Health
Collaborate. Innovate. Accelerate.

Looking Ahead

Ewan Birney



Goal: Approve at least 3 new deliverables on Sep 3



7th Plenary Registration



Global Alliance
for Genomics & Health

October 21 - 23, 2019
Hynes Convention Center
Boston, USA

Register Early

<http://bit.ly/7thplenary>



7th Plenary

October 21

Morning ***invitation only**

- Breakouts with focus on advancing Strategic Roadmap deliverables

Afternoon ***invitation only**

- Breakouts continued
- GenoPri

Evening

- Welcome Remarks
- Networking Dinner

October 22

Full Day

- Keynote
- GA4GH Updates
- Workshops
- Poster Sessions

October 23

Morning & Early Afternoon

- Keynotes
- Workshops

Late Afternoon / Dinner

- Steering Committee

Can't register for October 21? Contact Justina.Chung@ga4gh.org

THANK YOU!

2019 GA4GH Connect Implementation Meeting Programme Planning Committee

- Laura Clarke (EMBL-EBI)
- Cristina Gonzalez (EMBL-EBI)
- Brian O'Connor (UCSC)
- Andy Yates (EMBL-EBI)
- Christina Yung (OICR)