Global Alliance
for Genomics & Health

Collaborate. Innovate. Accelerate.

GA4GH:

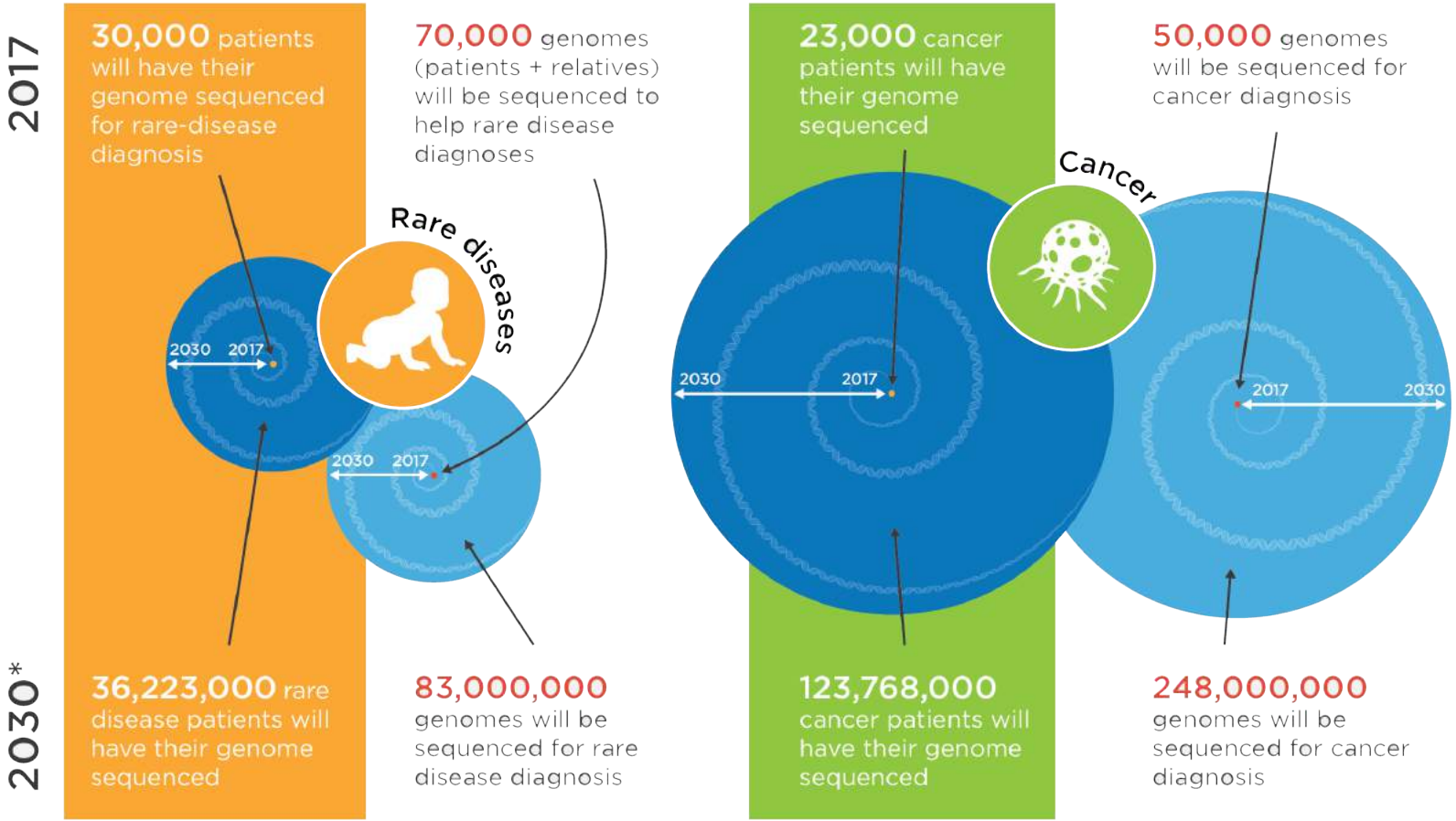
Empowering personalized health through data sharing

Melissa Konopko, PhD

Genomics = Science?

The world is changing

Percentage of whole genome and exome sequencing funded solely by healthcare



* Projected figures, based on current data and known status of genomics initiatives worldwide.

Genomics in Cancer

- When cancer patients have their genomes sequenced it informs clinical decisions ~10% of the time

- The clinical research community is confident this will increase steadily over the coming years

Genomics enters healthcare

In 2017 active genomic medicine programmes are already underway in many countries. Finland, the UK, the US, and Australia are a few examples.

1 MILLION PEOPLE

The National Institutes of Health plans to recruit over 1 million people into its All of Us research programme.

18,000 people by 2021

To improve patient outcomes and support research, the Australian healthcare system is building a Federation of clinical and genomic data.

10% of Finland's population expected to have some genomic data in healthcare by 2020.

10%

100,000 PATIENTS

The UK National Health System plans to sequence 100,000 individuals by 2020.

Global Alliance
for Genomics & Health

Collaborate. Innovate. Accelerate.

Responsibilities
Opportunities
Challenges

HUMAN RIGHTS FOUNDATION

Universal Declaration of Human Rights, (1948)

27(1) **"The Right to Science"**

"**Everyone has the right** freely to participate in the cultural life of the community, to enjoy the arts and **to share in scientific advancement and its benefits.**"

Responsibility

- Technical knowhow around genomics is in the research community
- Technical knowhow around clinical features and diagnosis is in the clinical community

We have a joint responsibility to make this work for patients

Opportunity

- If we can enable secondary use of clinical genomic data for research we will have a >60 million virtual cohort by 2025
- Humans will be the best studied organisms on the planet due to healthcare

Opportunities from cohorts of 10M+

1. Diagnose rare disease associated with low frequency alleles
2. Treat heterogeneous diseases like cancer
3. Discover and characterise epistasis
4. Understand selection pressures
5. Understand mutational penetrance in observational studies
6. Understand variation in allele frequency in different environments

Surmountable Challenges

- Healthcare is not used to this type, amount of data; we must draw on skills, learnings of research.
- As genomic datasets grow from terabyte to petabyte to exabyte scale, the community must re-tool.
- Clinical data are not interoperable: differences between individual and international healthcare technical systems. Portable analysis routines must be developed.
- Different nations have unique regulatory requirements that must be met while maximizing data access.

Global Alliance for Genomics & Health

The GA4GH mission

Global Alliance for Genomics & Health

We are an international non-profit that aims to:

> *"…accelerate progress in genomic research and human health by cultivating a common framework of standards and harmonized approaches for effective and responsible genomic and health-related data sharing."*

The solution is federation

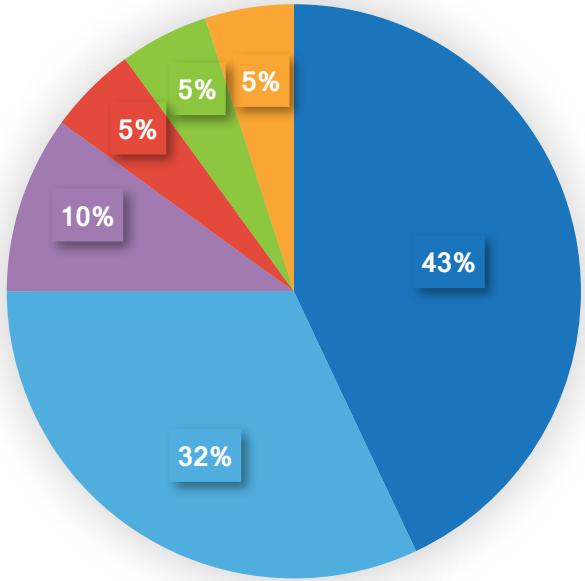
“a grouping of autonomous organisations and datasets with a centralised control”

Federation allows us to….

1. move analysis to data, not aggregate data close to each researcher
2. have broad, reciprocal data access methods which respect national processes and patient consent
3. transfer methods and skills into the healthcare sector
4. leverage healthcare data to make more discoveries on humans

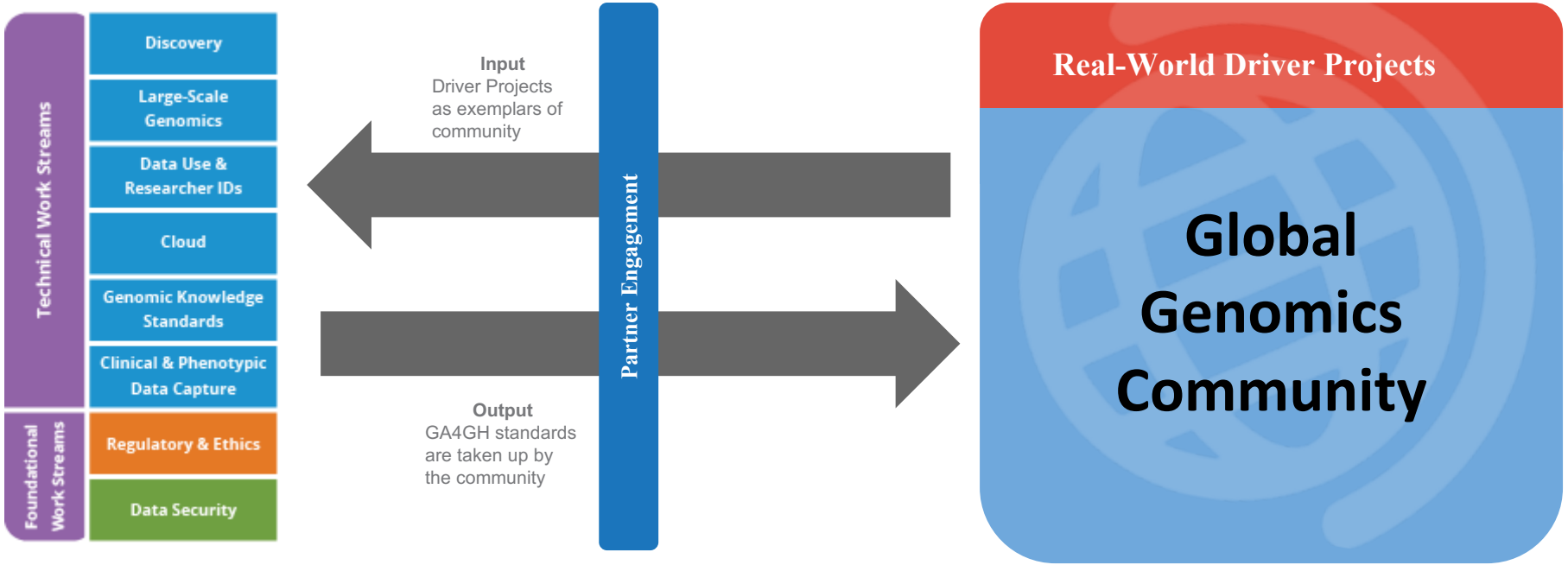
What is GA4GH?

The GA4GH ecosystem



- Life science and information technology companies
- Universities and research institutes
- Academic medical centers and health systems
- Disease advocacy organizations and patient groups
- Consortia and professional societies
- Funders and agencies

GA4GH and the Community



GA4GH Connect: 2017 Driver Projects

All of Us Research Program
United States

Australian Genomics
Australia

BRCA Challenge
International

CanDIG
Canada

ClinGen
United States

ELIXIR Beacon
Europe

ENA / EVA / EGA
Europe

Genomics England
United Kingdom

Human Cell Atlas
International

ICGC-ARGO
International

Matchmaker Exchange
International

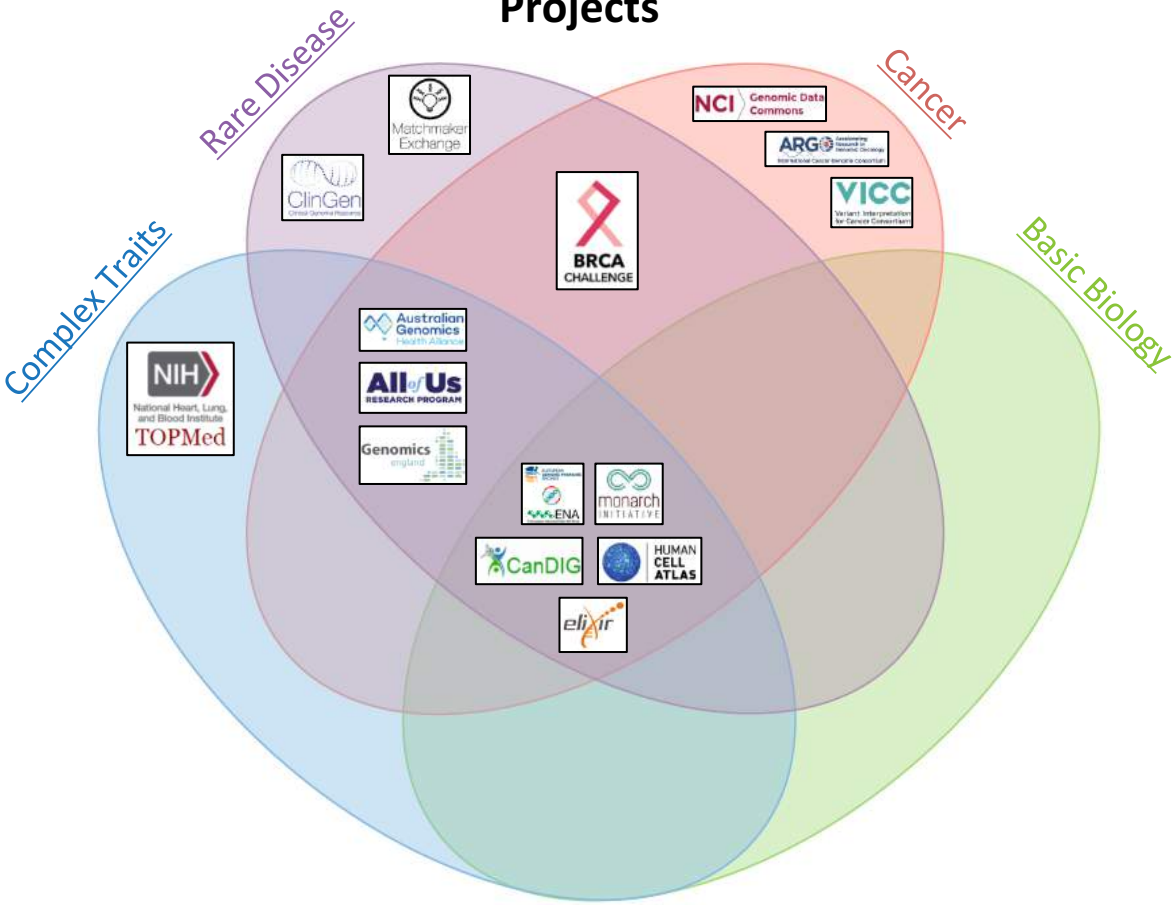
Monarch Initiative
International

National Cancer Institute Genomic Data Commons
United States

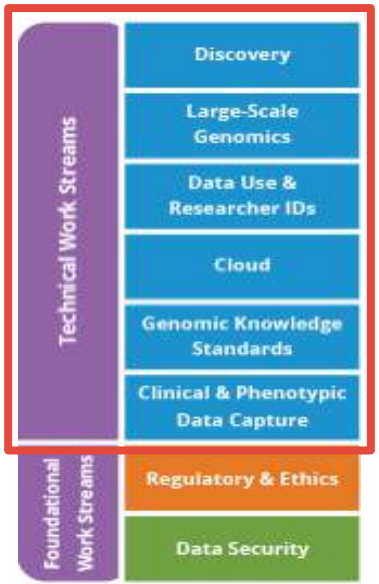
TOPMed
United States

Variant Interpretation for Cancer Consortium
International

2017 Driver Projects

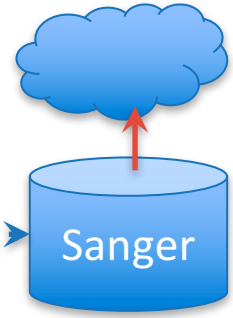


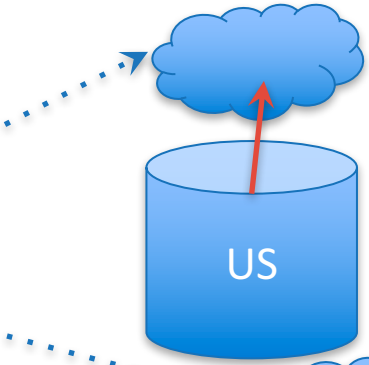
Our “Work Streams”

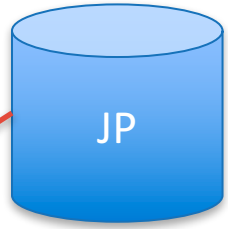


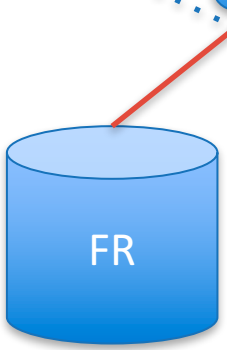
Creating a suite of standards and APIs for genomic data sharing:

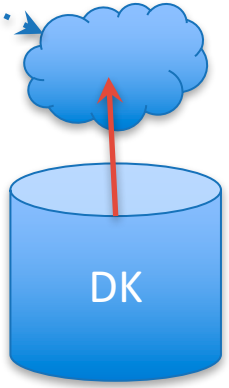
- **Cloud**: *Bringing algorithms to the data by creating standards for portable workflows.*
- **Large Scale Genomics**: *Creating standardized methods for accessing large-scale genomic data.*
- **Data Use & Researcher Identities**: *Harmonizing researcher identities and generating a widely applicable data use ontology.*
- **Clinical and Phenotypic Data Capture**: *Establishing standard ontologies and information models to describe clinical the phenotype.*
- **Discovery**: *Creating a unified data discovery platform to accelerate genomics and clinical data access.*
- **Genomic Knowledge Standards**: *Developing standards to represent genomic variants and their annotations.*











- Researcher wants to run new analysis method on >10M genomes; Sanger only has 1 million on site.
- More data available behind technical, logistical, regulatory and security barriers.
- With GA4GH Solutions: They register and can be electronically authenticated to run their workflow on interoperable data in the cloud on data from all over the world.

The Current GA4GH Toolkit

Genomic Data

- File Formats Standards – VCF/BCF, BAM/SAM/CRAM
- GA4GH Streaming API – htsget

Regulatory & Ethics

- Framework for Responsible Sharing of Genomic and Health Related Data
- Accountability Policy
- Data Sharing Lexicon
- Consent Policy

Data Security

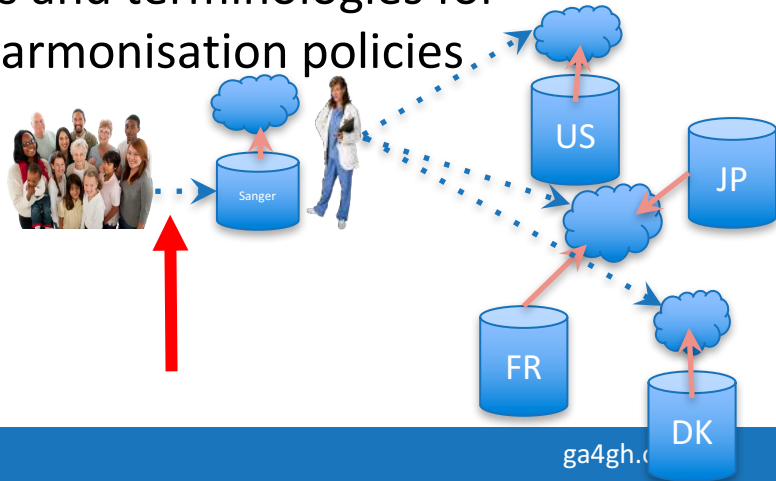
- Security Infrastructure
- Privacy and Security Policy

Clinical and Phenotypic Data Capture

Global Alliance for Genomics & Health

Establish standard ontologies and data models to describe clinical phenotypes

- **Information models for clinical/phenotypic data exchange**: standard to enable the exchange of clinical phenotype information. Alignment with HL7/FHIR
- **Phenotype and disease ontology**: ontologies and terminologies for capturing the clinical phenotype as well as harmonisation policies to enable machine-readablity

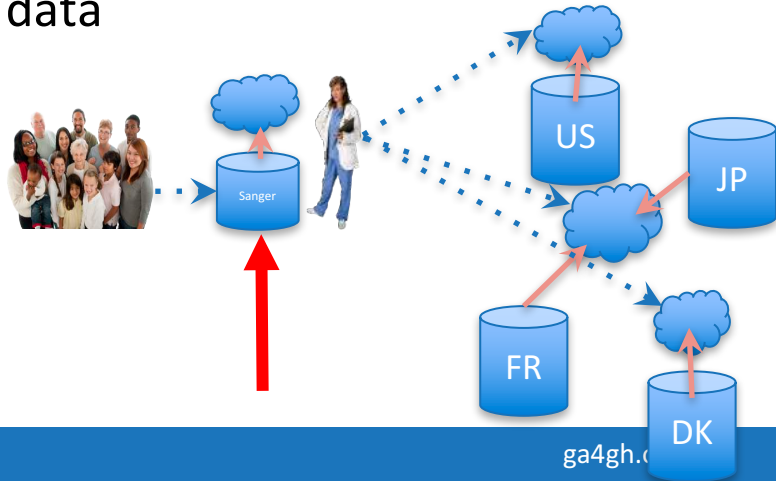


Genomic Knowledge Standards

Global Alliance for Genomics & Health

Standards-based components for exchange of genomic information

- **Variant representation: data model/specification** extensible data model and message schema specification for the representation of variants
- **Variant annotation: data model** to guide the linkage of annotations and structured clinical interpretations to variant data

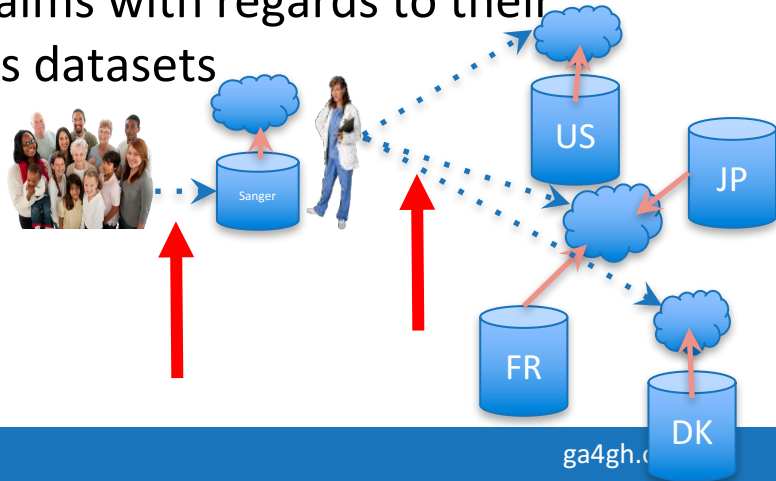


Data Use and Research Identity (DURI)

Global Alliance for Genomics & Health

Harmonization of researcher identities

- **Data use ontology**: tag datasets with restrictions about their usage, making them automatically discoverable based on the intended usage
- **Researcher identity and bona fide status**: allow researchers and other users to establish identity and credentials claims with regards to their professional identity to acquire access across datasets



Foundational Work Streams

- **Data Security**: *Developing and using technology standards to protect genomic and related data.*
- **Regulatory and Ethics**: *Providing guidance on the ethical, legal, and social implications of international data sharing.*

GA4GH 2018 Strategic Roadmap...Alignment to F.A.I.R. Principles

Tools for making data...

Findable

- Beacon
- Data Use Ontology
- Internet of genomics
- Reference sequence retrieval
- Search API
- Service Registry
- Tool Registry Service

Accessible

- Authentication and Authorization Infrastructure
- Cloud access policy
- Data Object Service
- Data Use Ontology
- Researcher ID & bona fide status

Interoperable

- Clin/pheno data exchange info models
- Data Object Service
- Genetic variant file formats
- Ontology best practices
- Read file formats
- RNAseq expression matrix
- Variant annotation
- Variant representation
- Variant submission API
- Task Execution Service
- Testbed & interoperability demos
- Tool Registry Service
- Workflow Execution Service

Reusable

- htsget Streaming API
- Variant annotation

With standards genomic medicine will:

- enter clinical practise at a much faster rate
- be less expensive and have reduced risk, decreasing costs to healthcare systems and individuals
- harmonise work around the globe
- reach many more patients worldwide
- be faster to deliver further advances for basic research and human health

Global Alliance
for Genomics & Health

Collaborate. Innovate. Accelerate.

Have ideas?
Want to get involved?

melissa.konopko@ga4gh.org

www.ga4gh.org