Global Alliance
for Genomics & Health

Collaborate. Innovate. Accelerate.

Genomic and Health Related Data Sharing

October 18th, 2018

About Me

Global Alliance for Genomics & Health

Not a genomic or data scientist, not a clinician or an informatician

Med Tech Executive

Patient Advocate and Navigator

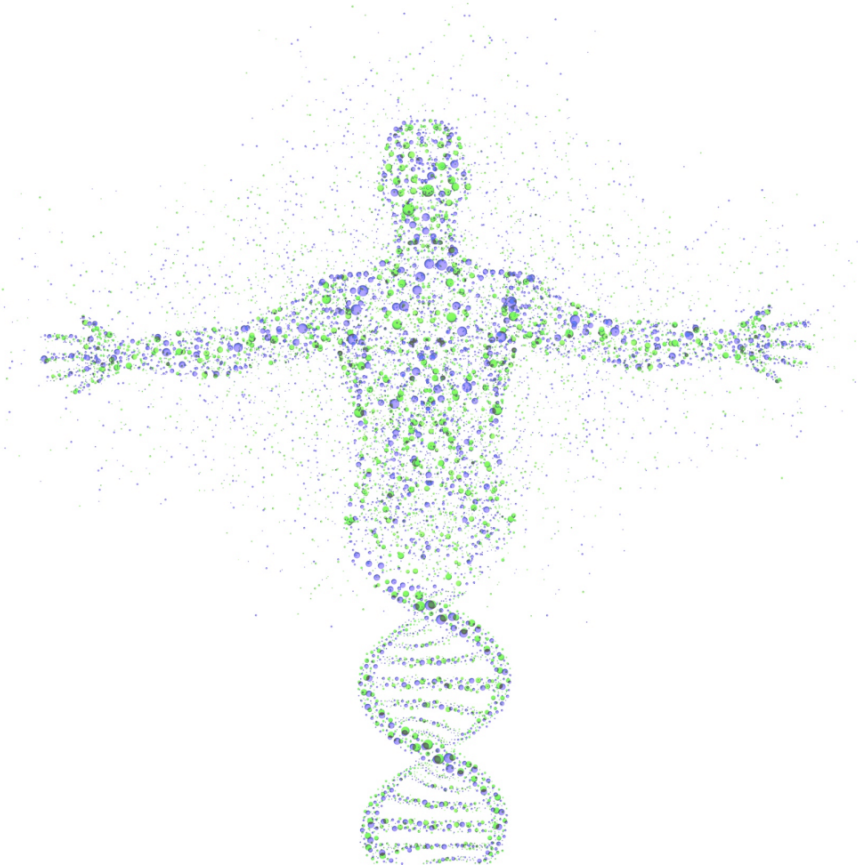
Cancer Charitable Sector Executive

Moving Health Research into Products, Policy, Practise

- GA4GH and OICR

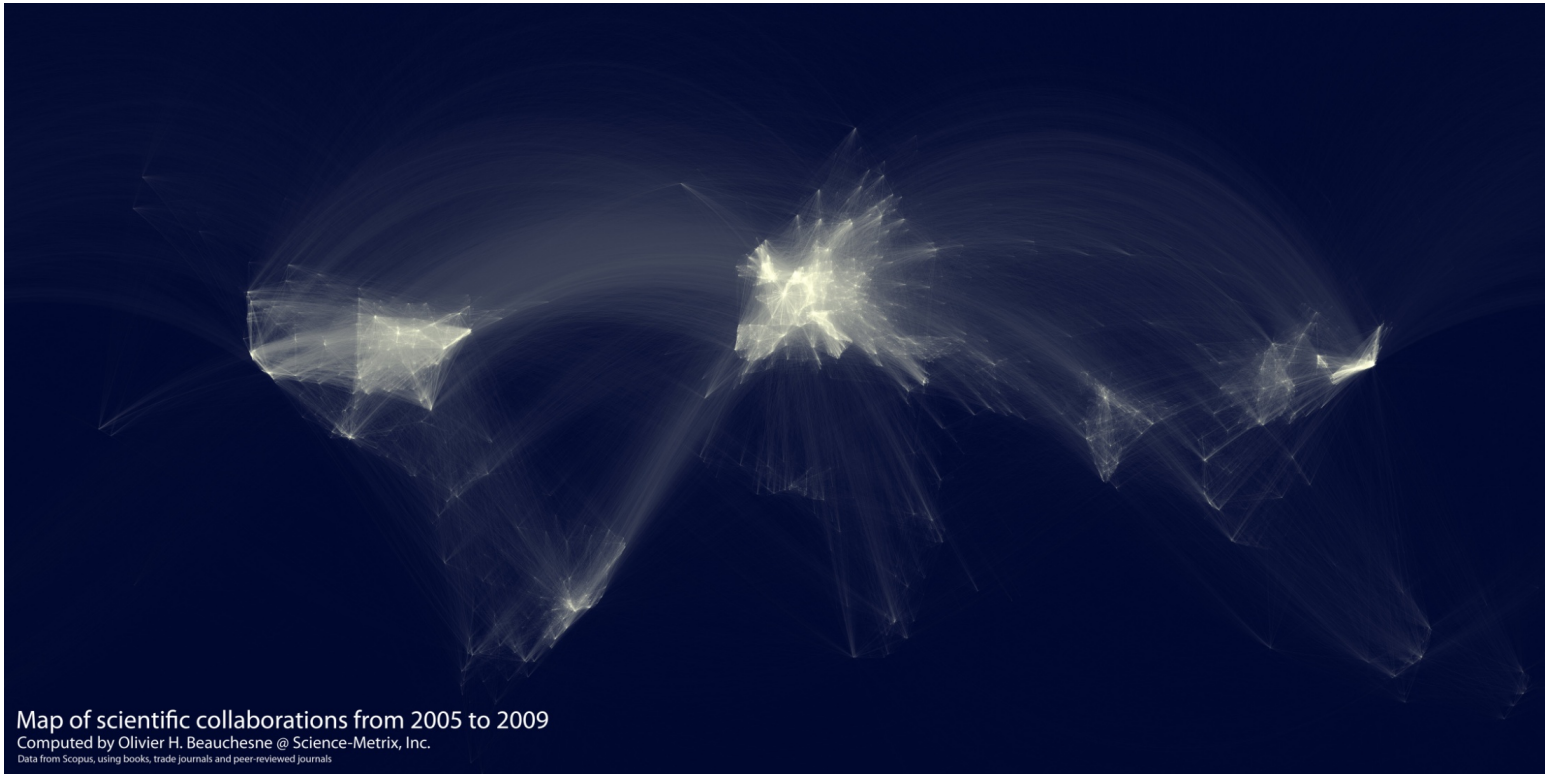
NO Conflicts

We have been living through a revolution – or 3



International Scientific Collaboration 2005-9

Global Alliance for Genomics & Health

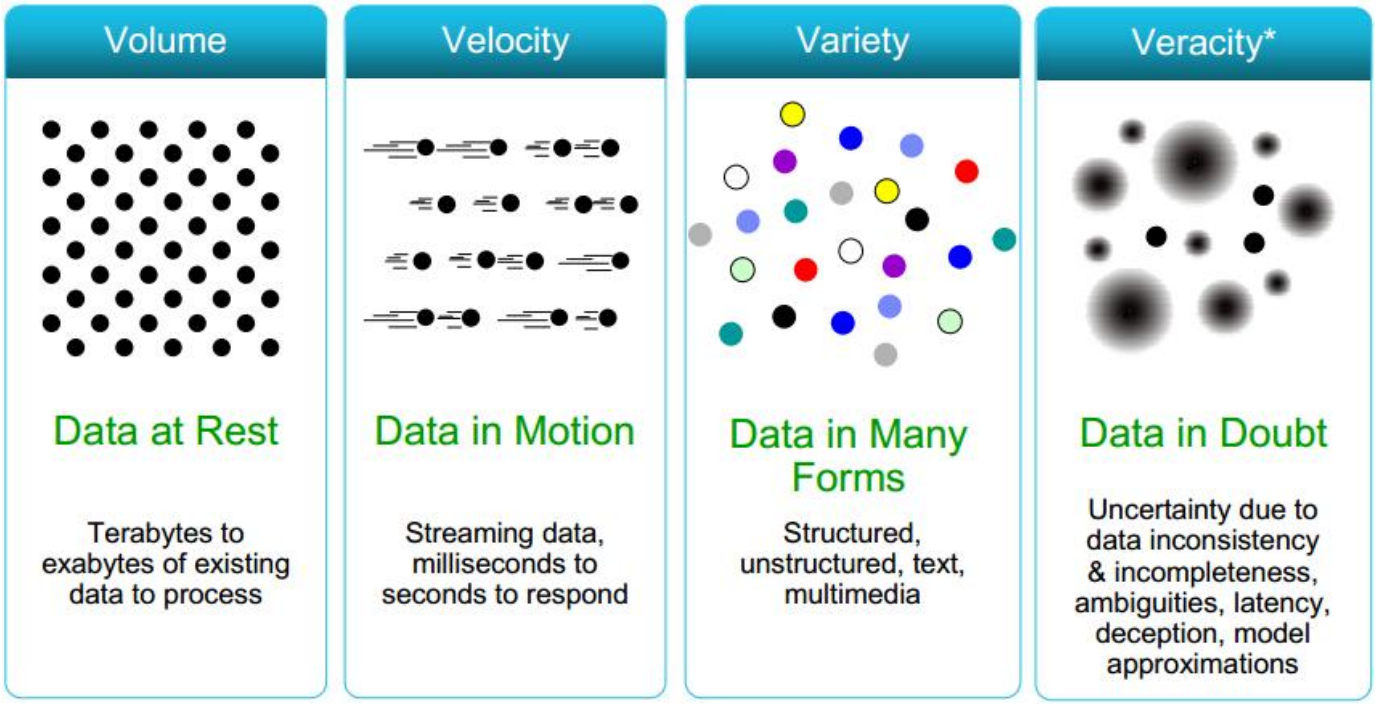


And today

Global Alliance for Genomics & Health

The 3 V's of Big Data … and the 4th

Global Alliance for Genomics & Health



*Truthfulness, accuracy or precision, correctness

Source: http://www-05.ibm.com/fr/events/netezzaDM_2012/Solutions_Big_Data.pdf

BIG DATA in Health

Astronomical – Genomical - not really!

- BUT Put together
  - OMICS + phenotype + clinical + longitudinal + wearables + exposome + data exhaust
- AI – ML- DL – NN
  - Pattern, clusters, data shaping the hypothesis

Cost and time for a whole genome

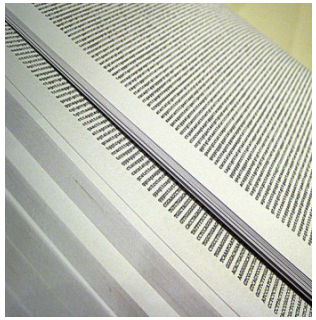
Global Alliance for Genomics & Health

2003	2018
Cost - $3,000,000,000	Cost $1000
Time – 15 years	Time – 1 day

The Challenge

Global Alliance for Genomics & Health

Unparalleled generation of human genetic data



In a way that allows data to be shared on a global level

How do we unlock its potential?

Thus empowering new knowledge, new diagnostics and new therapeutics for patients and populations

Courtesy of Mark Lawler

The GA4GH Mission

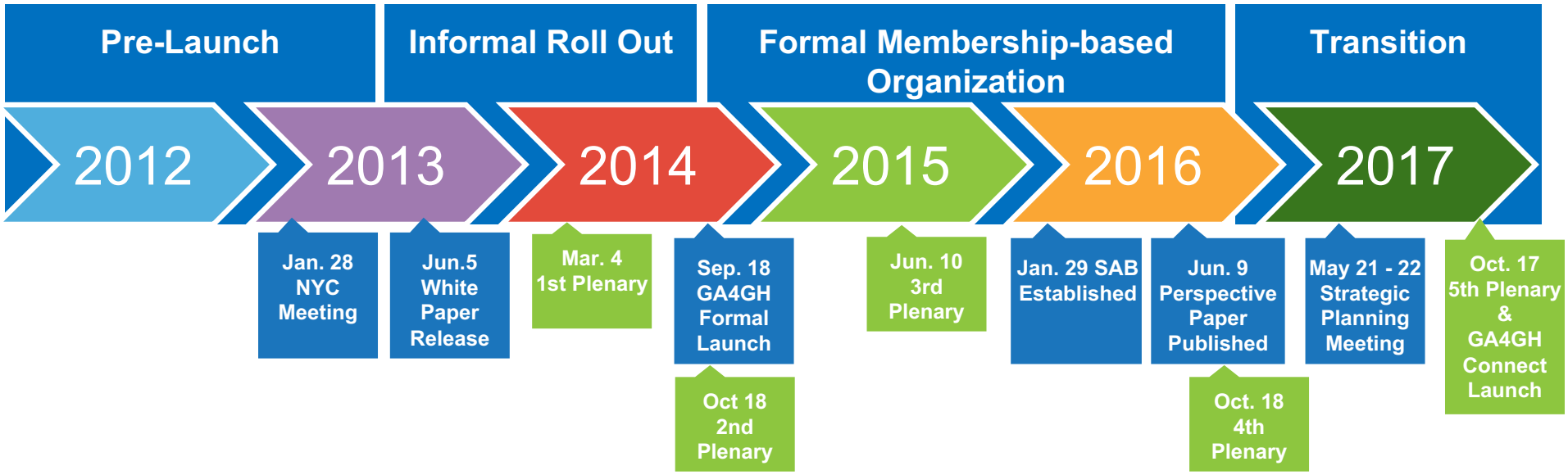
Global Alliance for Genomics & Health

LAUNCH of GA4GH in 2013

The Global Alliance for Genomics and Health aims to accelerate progress in genomic science and human health by developing standards and framing policy for responsible genomic and health-related data sharing.

GA4GH: 2012-17

Global Alliance for Genomics & Health



Core Principles of Data Sharing

- Enable **international** data sharing
- Follow **FAIR** principles
- Promote sharing across the **translational continuum** (discovery research, clinical trials, clinics, diagnostic labs, industry)
- Encourages technology-enabled **federated approaches** (bring analysis to the data)
- Promote **interoperability**: scientific, technical, and ethical
  - Adopt standards, have transparent documentation
  - Standardized file formats, bioinformatics & variant calling protocols, variant annotation
  - Consent policies to ensure data can be shared internationally

Universal Declaration of Human Rights, (1948)

"The Right to Science"

27(1)

"**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.**"

"The Right to Recognition"

27(2)

"Everyone has the right to the protection of the moral and material interests resulting from any scientific, literary or artistic production of which he is the author."

Accelerated Shift from Research to Healthcare

Percentage of whole genomes and exomes
that are funded by **healthcare** systems

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.

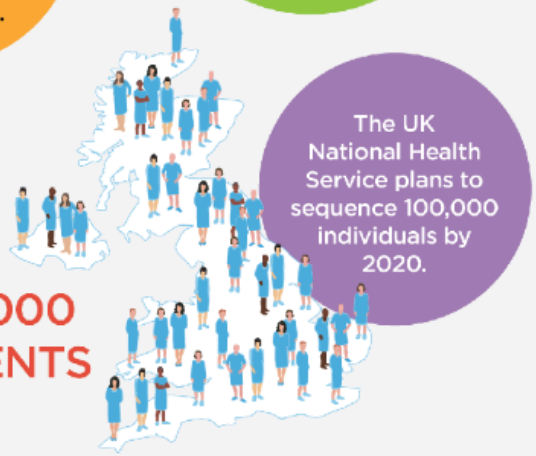
1 MILLION PEOPLE

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

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 is expected to have some genomic data in healthcare by 2020.

10%



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

100,000 PATIENTS

National Medical Genome Projects and Cohorts

Global Alliance for Genomics & Health

Medical Genomes

- Countries with active national medical genome projects
- Countries with some activity of medical genomics
- Countries planning medical genome projects

Cohorts

- National cohorts > 100k genotyped or sequenced at least 25k
- National cohorts > 100k people active collection now
- Planning national cohorts > 100k

Medical Genomes

Global Alliance for Genomics & Health

Countries with active national medical genome projects	Countries with some activity of medical genomes	Countries planning medical genome projects
Australia Denmark Estonia Finland France Iceland Saudi Arabia UK	Canada Japan Qatar Spain USA	Brazil China India Ireland Jordan Kuwait Netherlands Scotland Singapore South Korea Switzerland Turkey United Arab Emirates

National Cohorts

Global Alliance for Genomics & Health

National cohorts >100K genotyped or sequenced at least 25K	National Cohorts >100K people active collection now	Planning national cohorts >100K
Canada China Denmark Iceland Japan Norway South Korea Sweden UK USA	Africa (H3Africa) Australia Finland France Germany Iran Israel Malaysia Mexico Netherlands Saudi Arabia South Africa Taiwan	Austria India Singapore Switzerland

Opportunity

- If we can enable secondary use of clinical genomic data for research we will have a **>60 million virtual cohort by 2025**
- Data from **millions of samples** may be needed to show patterns in health/disease
- Humans will be the **best studied organisms** on the planet due to healthcare

Research to Medical

Research

- English as language
- Lightweight legal
- Identical/similar systems
- Open data
- Publications
- Grant-funding

Practicing Medicine

- National language
- Heavy legal framework
- Very different systems
- Closed data
- Not published
- Contract-funding

Solution - Federation

Global Alliance for Genomics & Health

Move Analysis to the data, not aggregate data close to each researcher.

Have broad, reciprocal data access methods, as open as possible but respecting national processes and thus implicitly patient consent

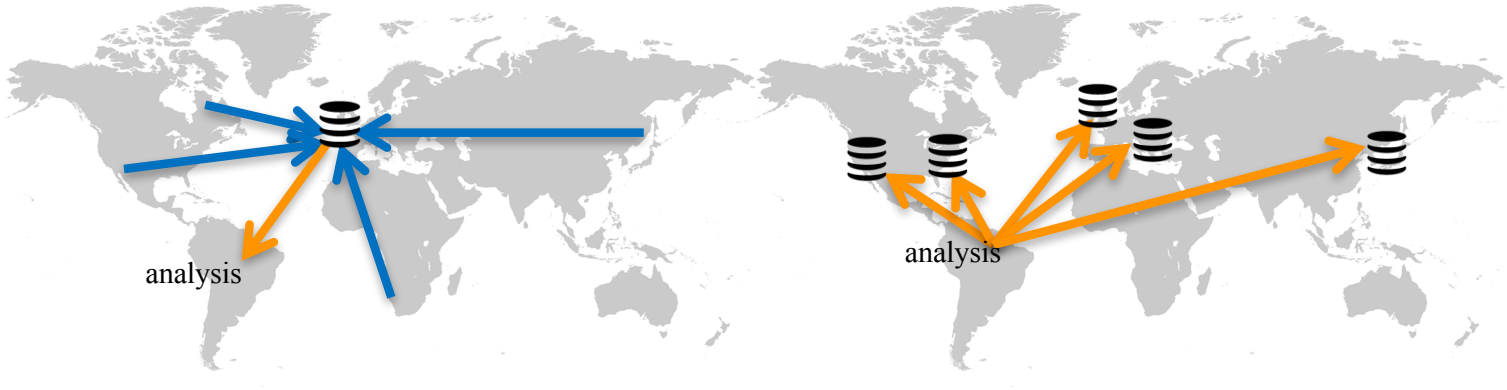
Transfer methods and skills into the healthcare sector

Leverage healthcare data to make more discoveries on humans

Federation

Open research data

Healthcare data
with research use



Aggregate data globally

Download, analyse locally

Continues for basic research

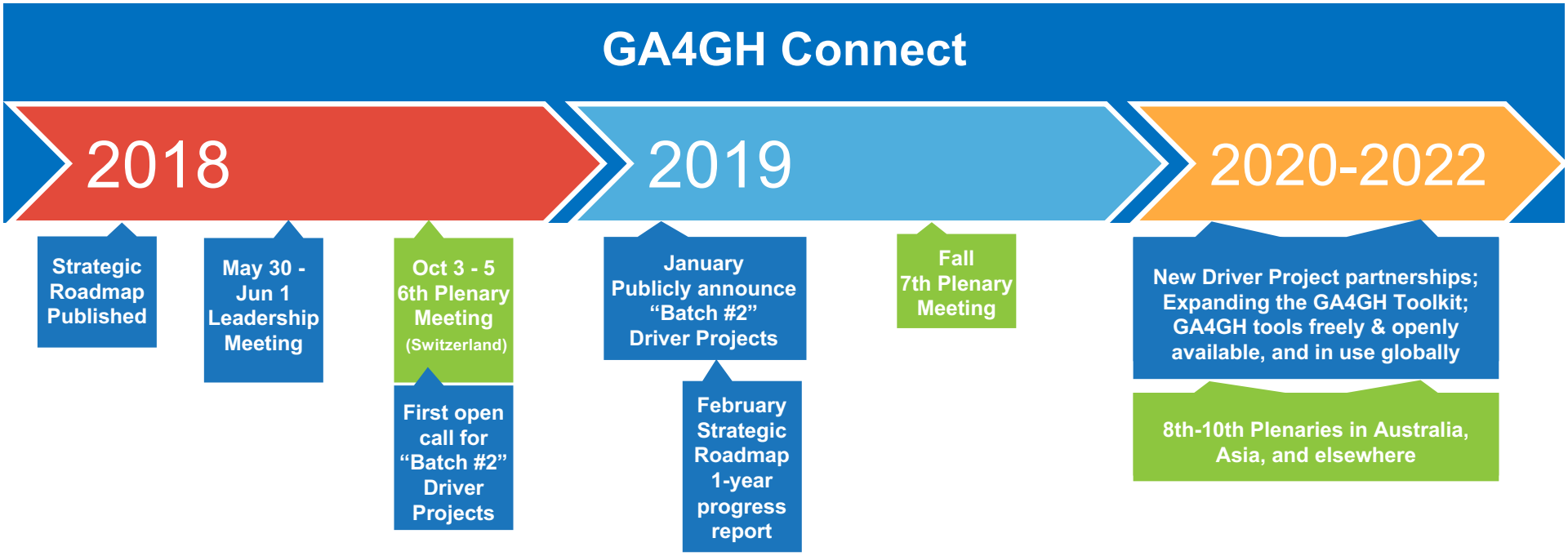
Analyse data locally (via VMs)

Collate analyses

New approach for healthcare

GA4GH: Present, and Future

Global Alliance for Genomics & Health



The GA4GH Ecosystem

Global Alliance for Genomics & Health

Global Alliance members include:

Universities and research institutes (32%)

Academic medical centers and health systems (10%)

Disease advocacy organizations and patient groups (5%)

Consortia and professional societies (5%)

Funders and agencies (5%)

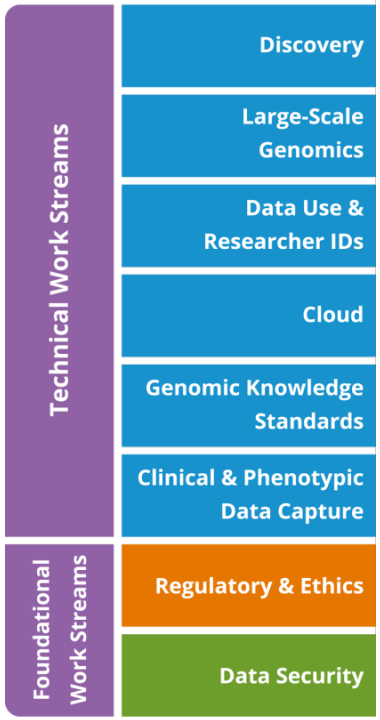
Life science and technology companies (43%)

Operations and Funding

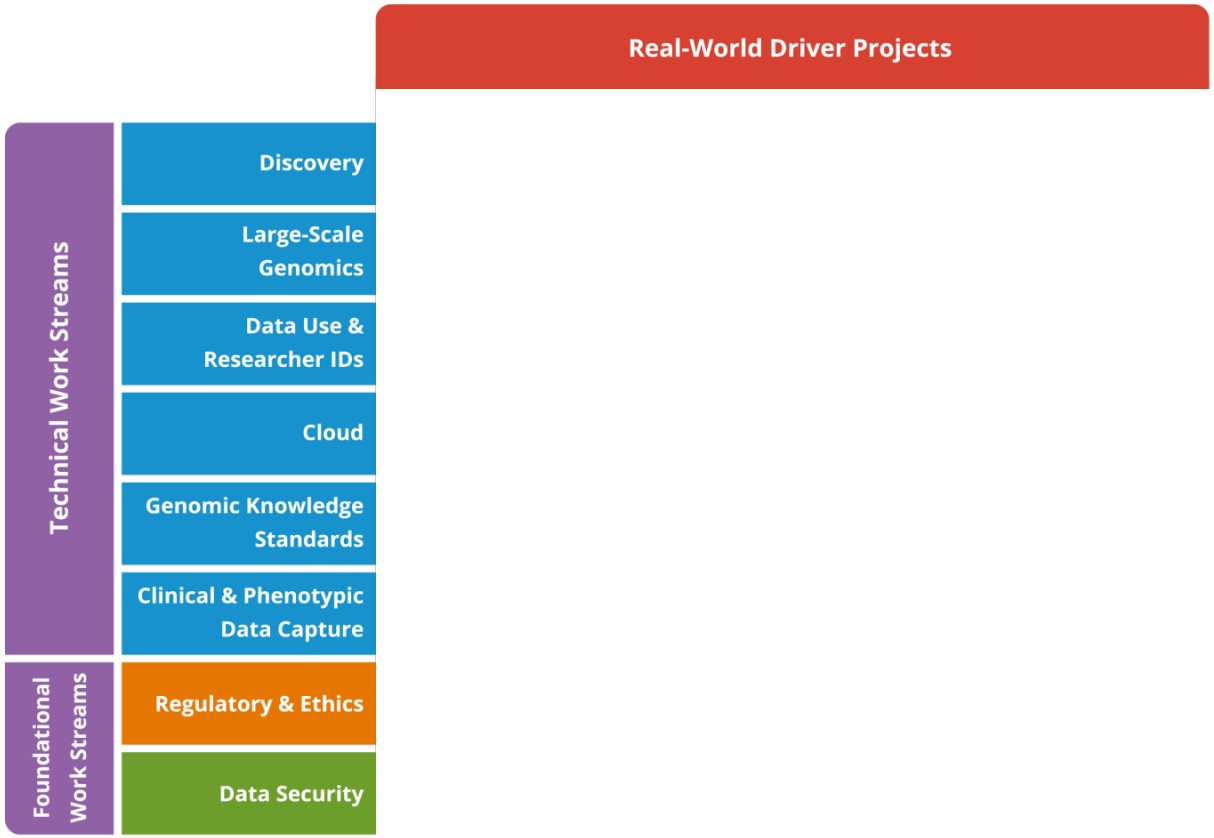
- **Host institutions**
  - Ontario Institute for Cancer Research
  - Wellcome Sanger Institute
  - Broad Institute of MIT and Harvard

- **Funding**
  - Working with major international funders (NIH, Wellcome, Genome Canada/CIHR) to support and expand core funding
  - Outreach to other public and philanthropic funders to support reference implementations, pilots, and global reach

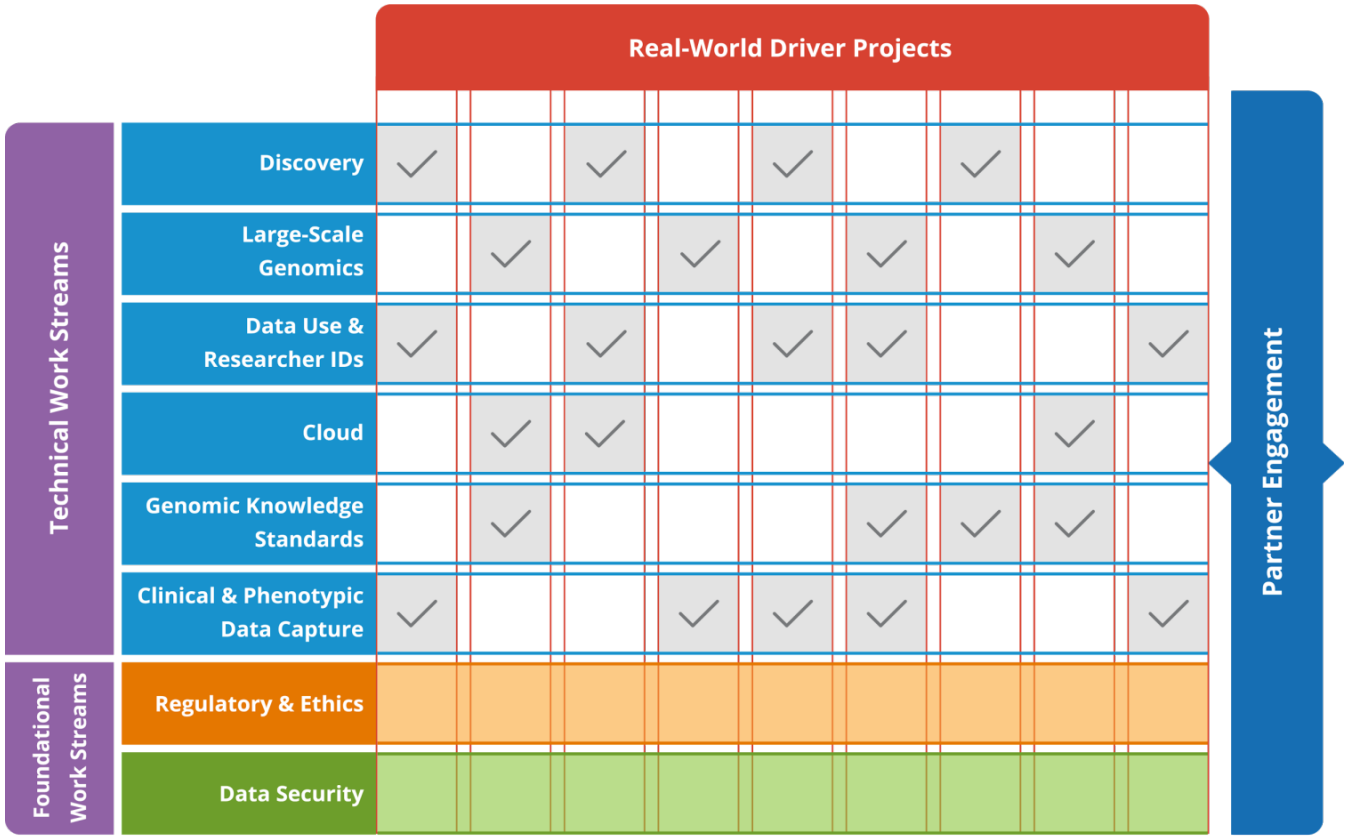
How We Work


Technical Work Streams
Discovery
Large-Scale Genomics
Data Use & Researcher IDs
Cloud
Genomic Knowledge Standards
Clinical & Phenotypic Data Capture
Foundational Work Streams
Regulatory & Ethics
Data Security

How We Work


Real-World Driver Projects
Technical Work Streams
Discovery
Large-Scale Genomics
Data Use & Researcher IDs
Cloud
Genomic Knowledge Standards
Clinical & Phenotypic Data Capture
Foundational Work Streams
Regulatory & Ethics
Data Security

How We Work


Real-World Driver Projects
Technical Work Streams
Discovery
Large-Scale Genomics
Data Use & Researcher IDs
Cloud
Genomic Knowledge Standards
Clinical & Phenotypic Data Capture
Foundational Work Streams
Regulatory & Ethics
Data Security
Partner Engagement

GA4GH Connect: 2018 Driver Projects

All of Us Research Program
United States

Australian Genomics
Australia

BRCA Challenge
International

CanDIG
Canada

ClinGen
International

ELIXIR Beacon
Europe

ENA / EVA / EGA
Europe

Genomics England
United Kingdom

Human Cell Atlas
International

ICGC-ARGO
International

Matchmaker Exchange
International

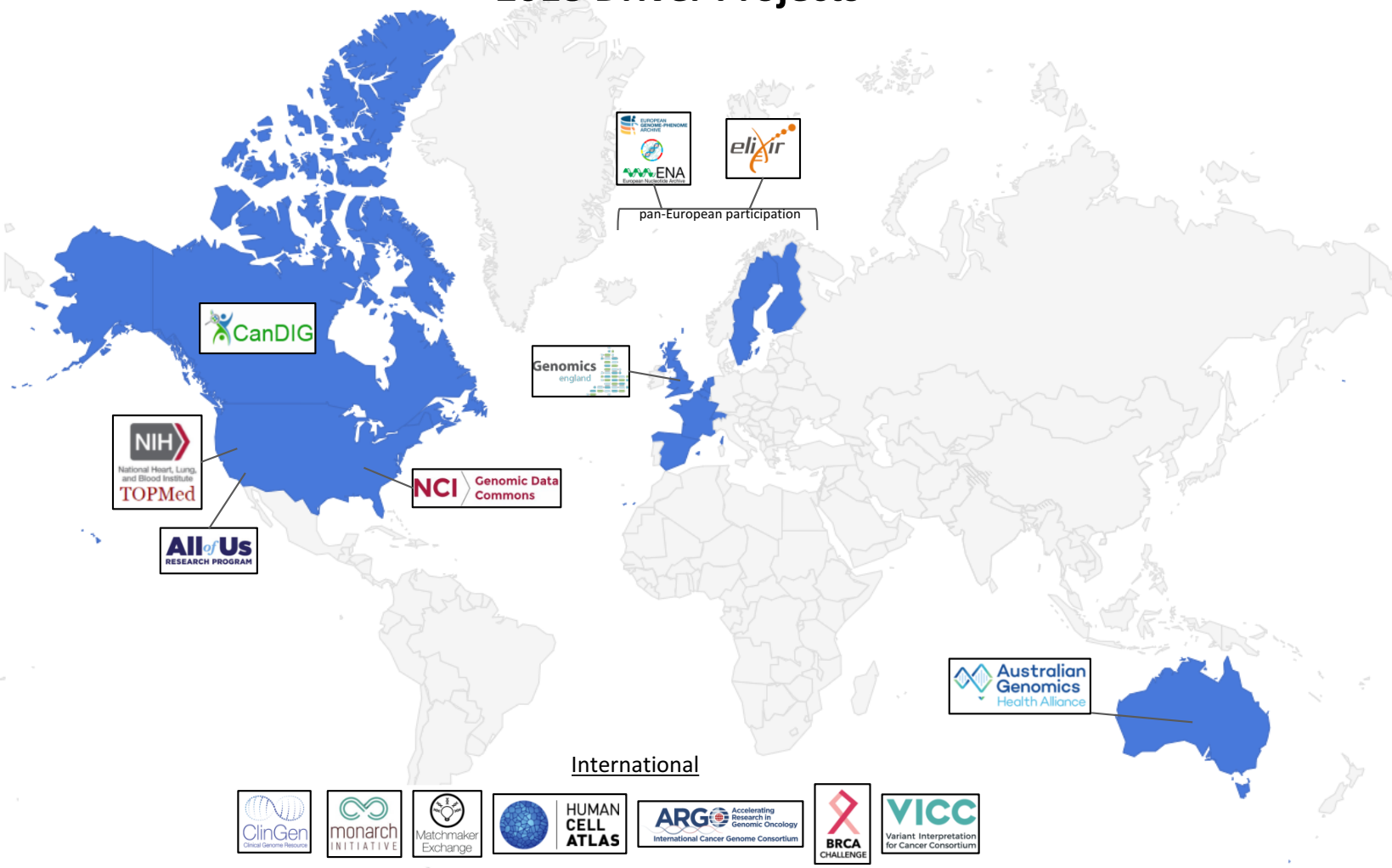
Monarch Initiative
International

National Cancer Institute Genomic Data Commons (NCI GDC)
National Cancer Institute Data Commons Framework (NCI DCF)
United States

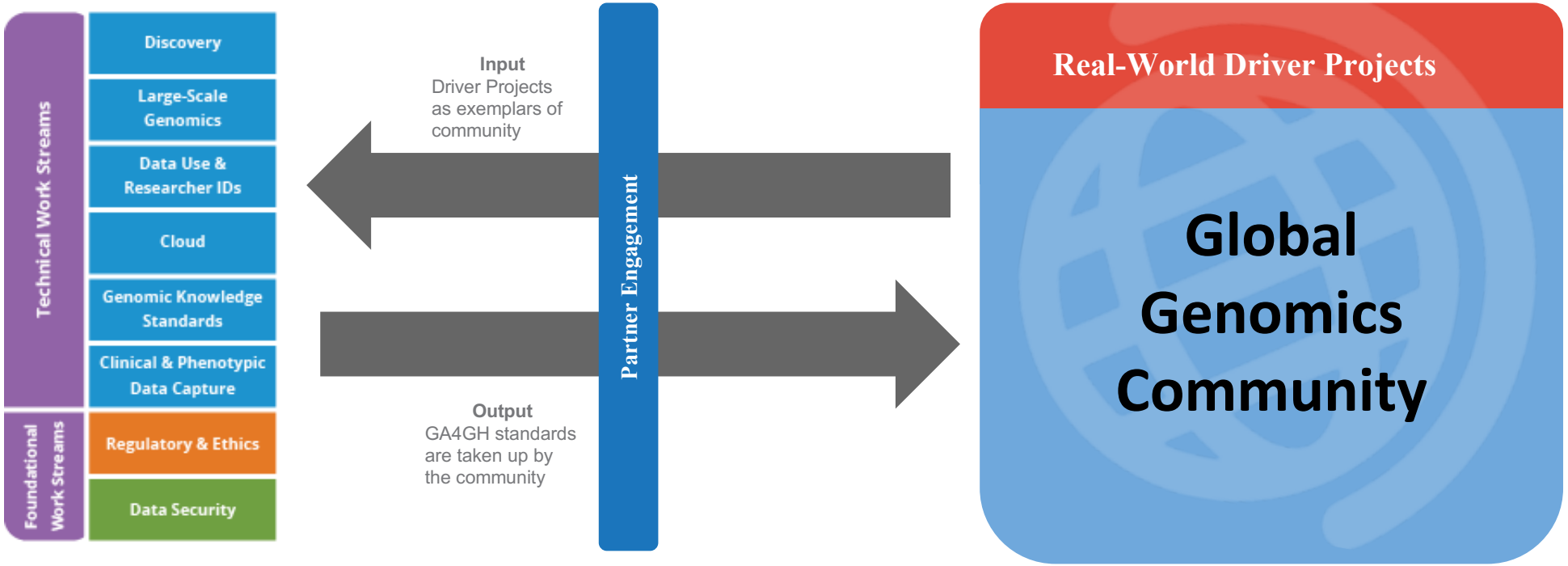
TOPMed
United States

Variant Interpretation for Cancer Consortium
International

2018 Driver Projects



GA4GH and the Community



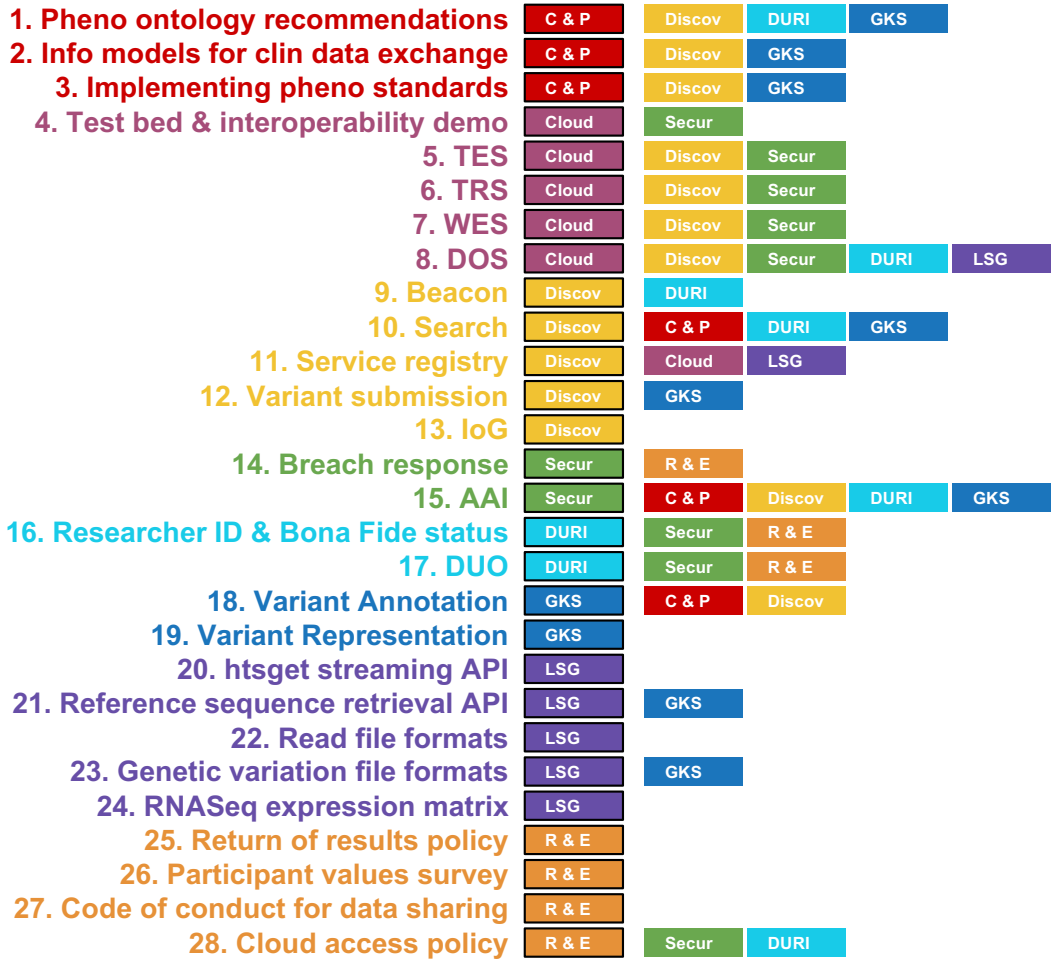
Community Engagement

Global Alliance for Genomics & Health

- **Driver Projects -** Capacity to Contribute to the development of standards
- **Collaborative Projects** - Projects with high interest to adopt or adapt standards and frameworks
- **Strategic relationships -**
  - with NIH and ELIXIR across several projects and work streams
  - XSDO – SNOMED, HL7, ISO, CDISC

GA4GH 2018 Strategic Roadmap

Deliverables



Work Streams

- Clinical & Phenotypic
- Cloud
- Discovery
- Data Security
- Data Use & Researcher IDs
- Genomic Knowledge Standards
- Large-Scale Genomics
- Regulatory & Ethics

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

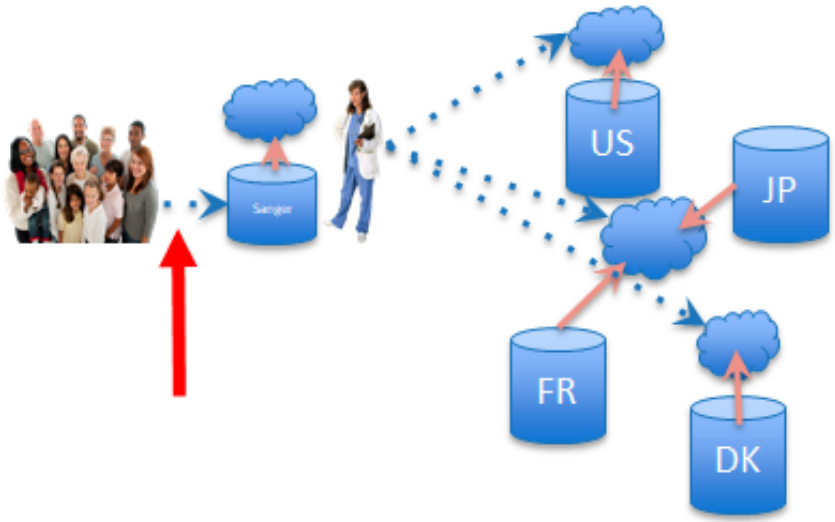
- htsget Streaming API
- Variant annotation

Global Alliance
for Genomics & Health

Collaborate. Innovate. Accelerate.

Clinical & Phenotypic Data Capture Work Stream

Establishing standard ontologies and information models to describe the clinical phenotype.



Clinical & Phenotypic Data Capture

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

Work with other SDO's to 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.
- **Phenotype and disease ontology**: ontologies and terminologies for capturing the clinical phenotype as well as harmonisation policies to enable machine-readablity.

International Collaboration and Data Sharing over 30 years

Consortia Data Release, Data Use, and Publication Policies (2009)

NIH Genomic Data Sharing Policy (2014)

Human Genome Project (1990)

Bermuda Principles (1996)

1000 Genomes Project (2008)

Policy for Release and Database Deposition of Sequence Data (2000)

International Cancer Genome Consortium (2008)

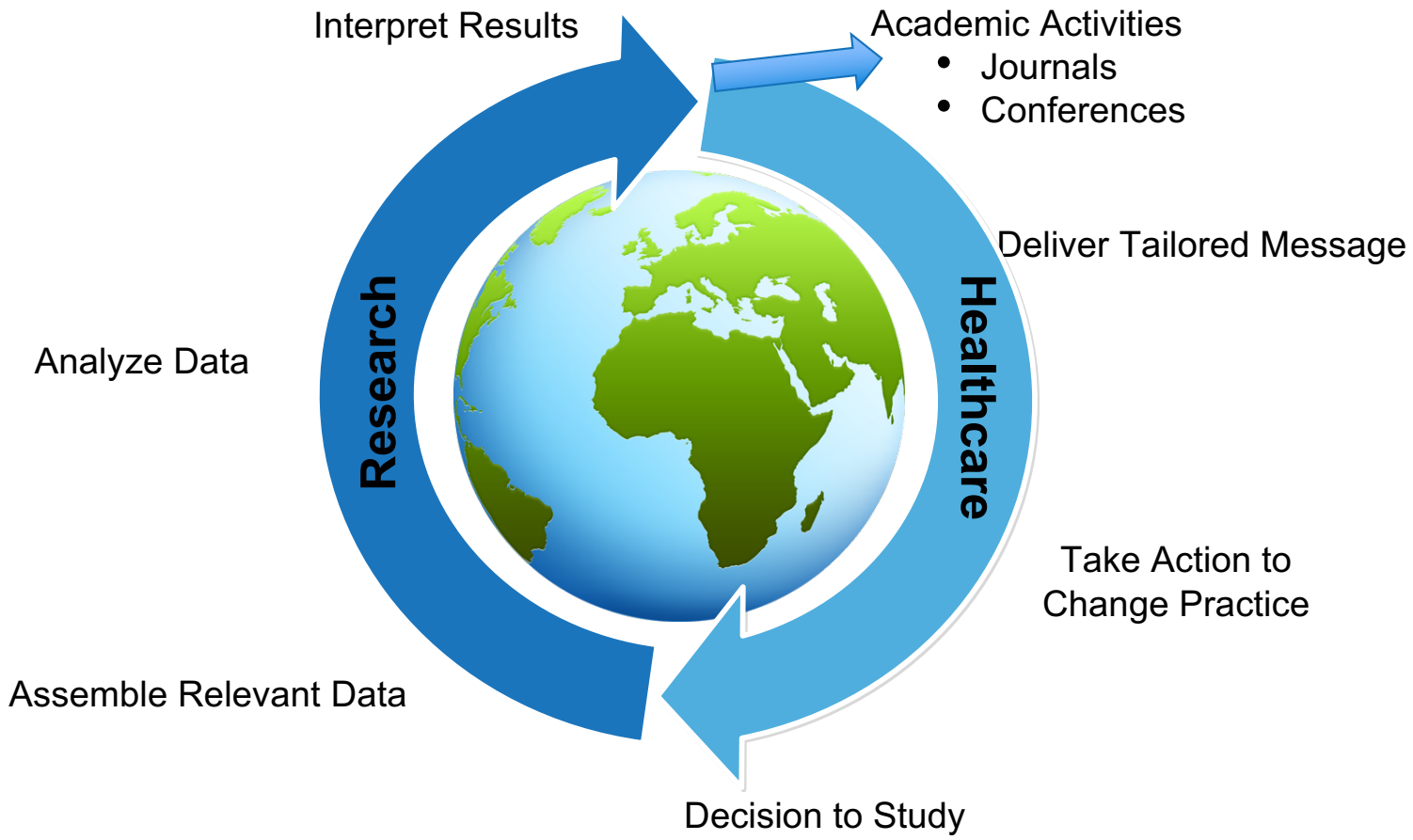
Policy for Sharing of Data Obtained in NIH Supported or Conducted Genome-Wide Association Studies (2007)

Personal Genome Project (2005)

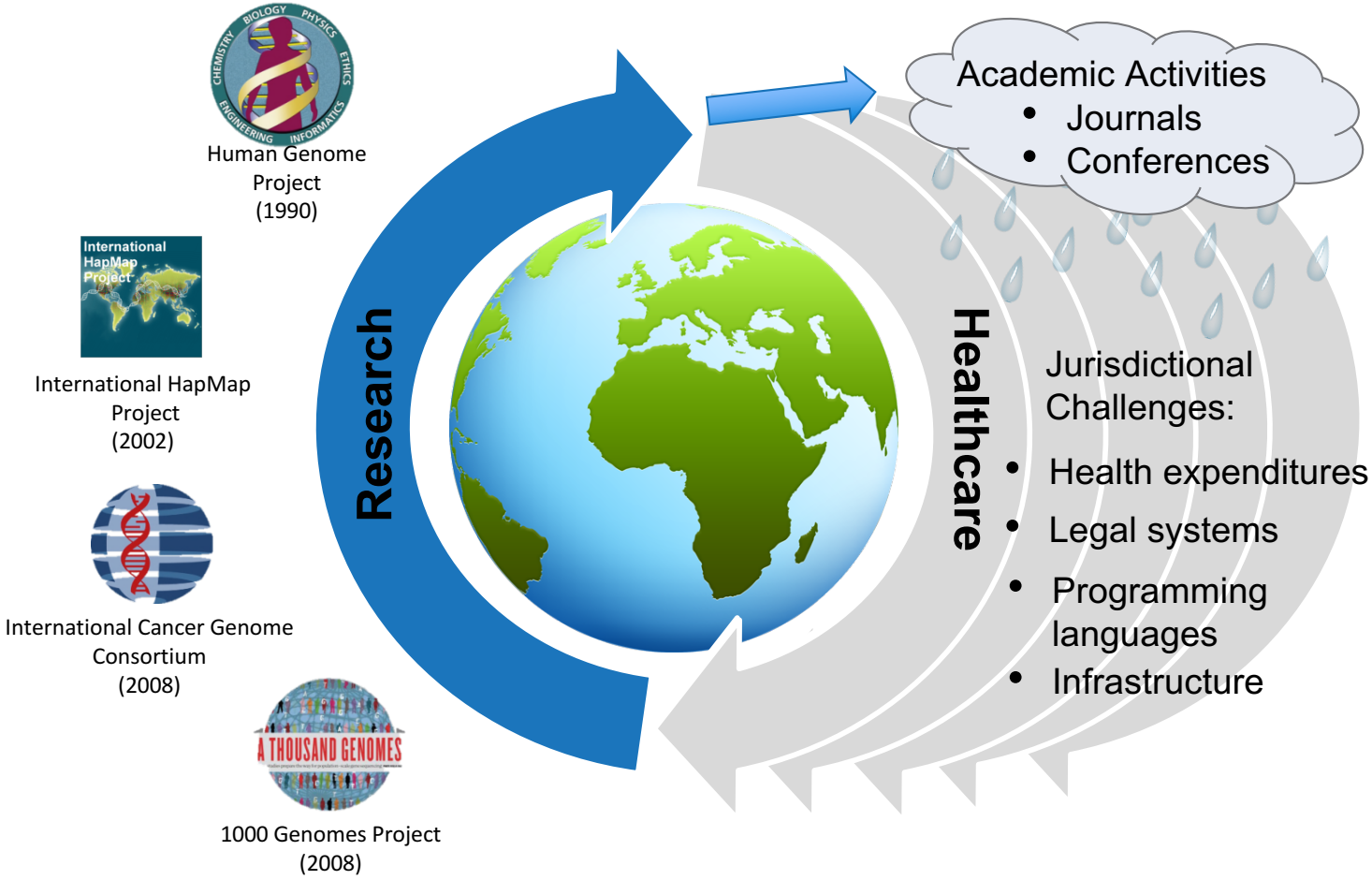
Uniform Principle for Sharing Integral Data and MAterials Expeditiously (UPSIDE) (2004)

International HapMap Project (2002)

Global Learning for Health … or not



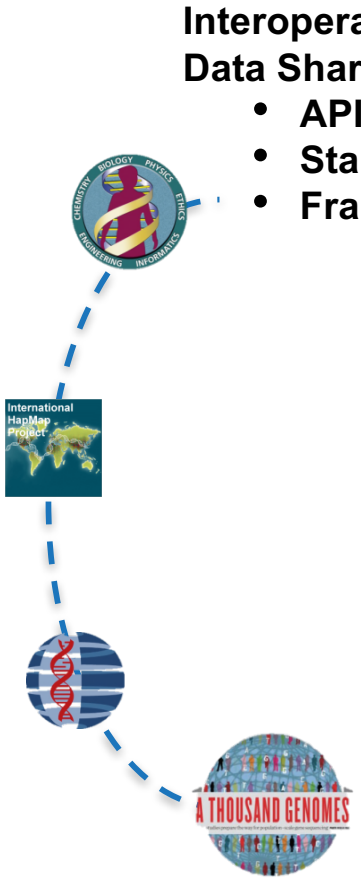
What is holding us back?

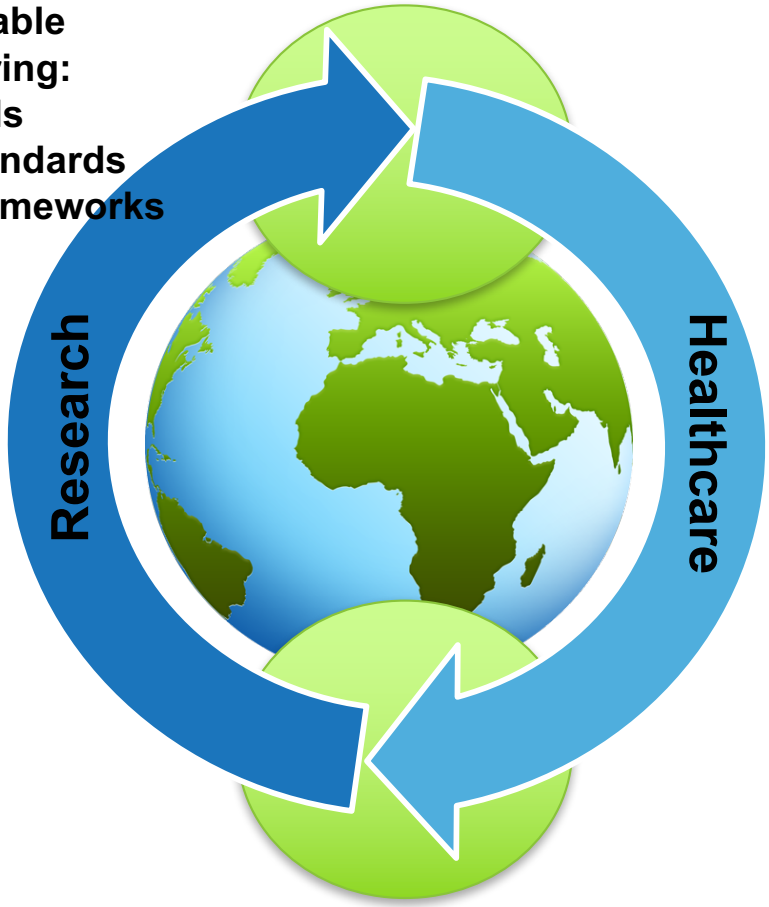


Global Learning for Health

Interoperable Data Sharing:

- **APIs**
- **Standards**
- **Frameworks**





Genomic Knowledge Exchanges

BRCA CHALLENGE

Matchmaker Exchange

Bridges need at least two anchors – and rules!

Shared Responsibility

- Much of the technical knowhow around genomics is in the research community
- The technical knowhow around clinical features and diagnosis is in the clinical community
- Clinical and phenotypic data much more diverse and complex
- We have a shared responsibility to make this work for patients

Thank You

New website	ga4gh.org
5-year Strategic Plan	ga4gh.org/aboutus
2018 Strategic Roadmap	ga4gh.org/howwework/strategic-roadmap
Op-Ed on BioRxiv	doi.org/10.1101/203554
Perspective paper in *Science*	PMID: 27284183
Thoughts on data sharing?	YourDNAYourSay.com
Plenary meeting slides	bit.ly/5thPlenary
Plenary videos on YouTube	GA4GH YouTube Channel