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

SIB
Swiss Institute of
Bioinformatics

Universität
Zürich[UZH]

Beaconize this: Federated Data Discovery in Biomedical Genomics

Michael Baudis
Professor of Bioinformatics
University of Zürich
Swiss Institute of Bioinformatics **SIB**
GA4GH Workstream Co-lead *DISCOVERY*
Co-lead ELIXIR Beacon API Development

Republic of Korea
Switzerland 60th anniversary
1963–2023

200+ Genomic Data Initiatives Globally

Clinical/Genomic Medicine

Research

National

Cohorts

How Many Genomes?

RESEARCH

HEALTHCARE

60M individuals
132.5M sequences

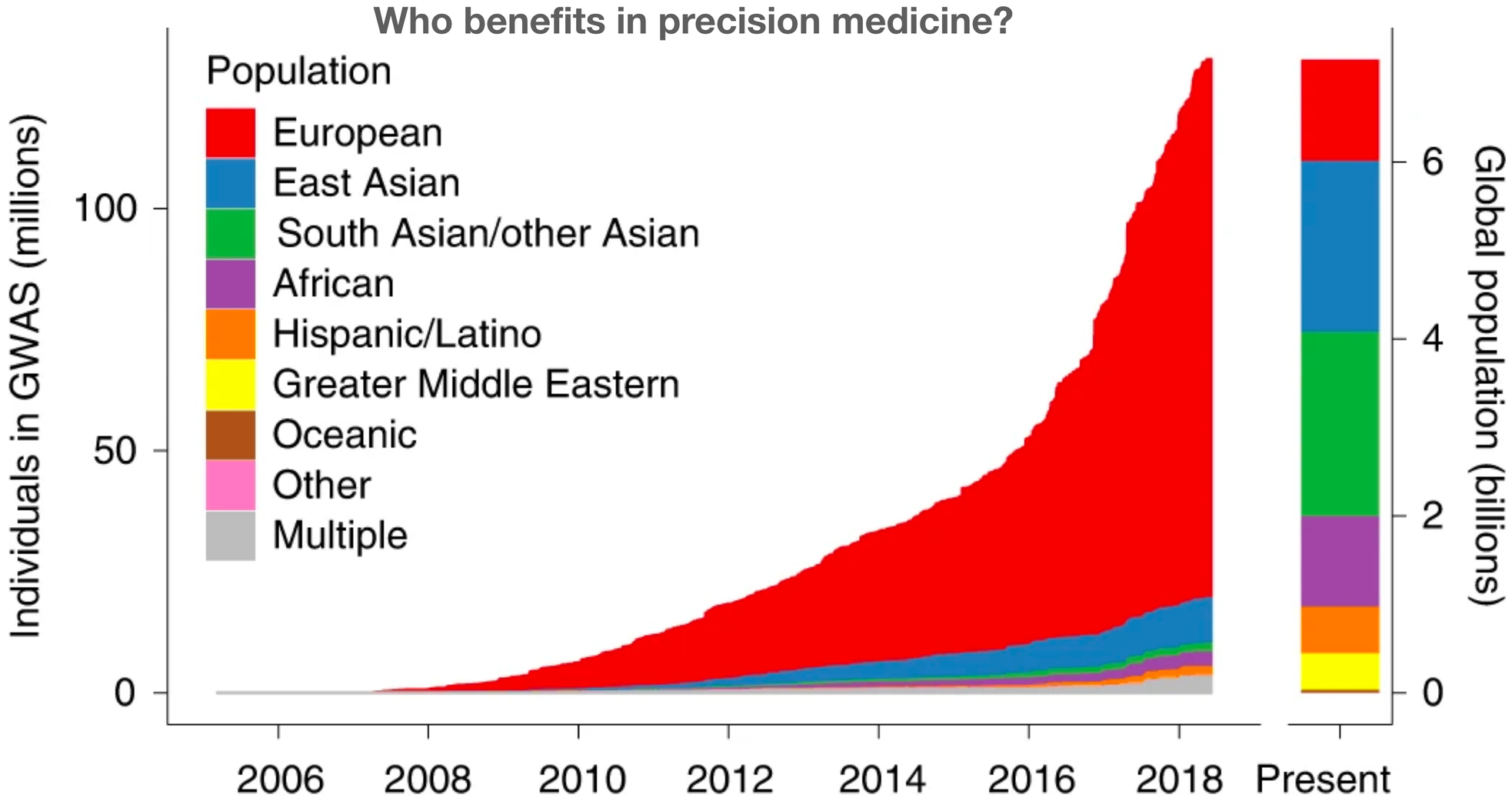
CLINICAL TRIALS

2.7-3M individuals

COHORTS

140M individuals

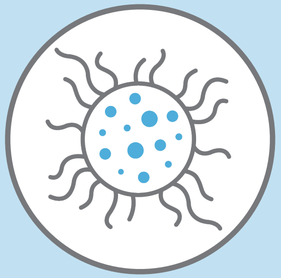
Genomic research has long-standing problems with diversity



Slide: Karen-Miga - The Human Pangenome Project @ GA4GH 2023

Martin et al 2019

Global Genomic Data Sharing Can…



Demonstrate patterns in health & disease

Increase statistical significance of analyses

Lead to “stronger” variant interpretations

Increase accurate diagnosis

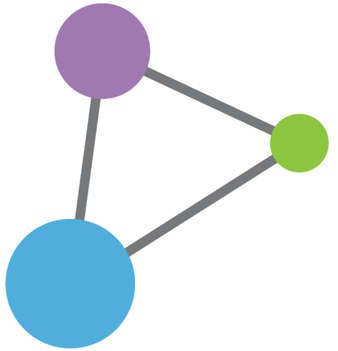
Advance precision medicine

Different Approaches to Data Sharing

Centralized Genomic Knowledge Bases

Data Commons
Trusted, controlled repository of multiple datasets

Hub and Spoke
Common data elements, access, and usage rules



Linkage of distributed and disparate datasets

Different Approaches to Data Sharing

progenetix

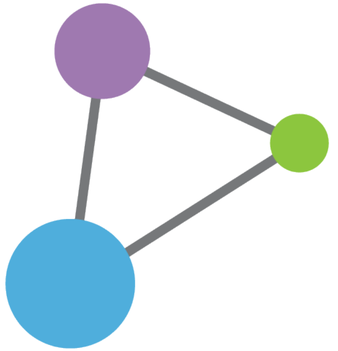
European Genome-Phenome Archive

SPHN Swiss Personalized Health Network

Centralized Genomic Knowledge Bases

Data Commons
Trusted, controlled repository of multiple datasets

Hub and Spoke
Common data elements, access, and usage rules



Linkage of distributed and disparate datasets

Different Approaches to Data Sharing

progenetix

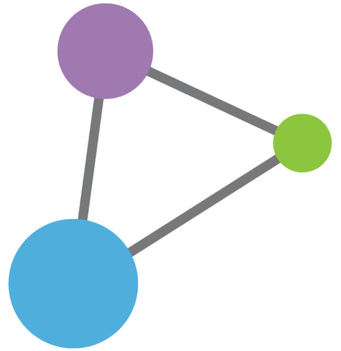
Centralized Genomic Knowledge Bases

European Genome-Phenome Archive

Data Commons
Trusted, controlled repository of multiple datasets

SPHN Swiss Personalized Health Network

Hub and Spoke
Common data elements, access, and usage rules



Linkage of distributed and disparate datasets

progenetix.org

Cancer Genomics Reference Resource

- ***open*** resource for oncogenomic profiles
- over **116'000 cancer CNV profiles**
- more than **800 diagnostic types**
- inclusion of reference datasets (e.g. TCGA)
- standardized encodings (e.g. NCIt, ICD-O 3)
- identifier mapping for PMID, GEO, Cellosaurus, TCGA, cBioPortal where appropriate
- core clinical data (TNM, sex, survival ...)
- data mapping services
- recent addition of SNV data for some series

Universität Zürich[UZH]

progenetix

SIB Swiss Institute of Bioinformatics

Different Approaches to Data Sharing

progenetix

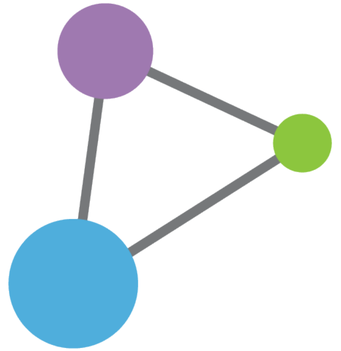
Centralized Genomic Knowledge Bases

European Genome-Phenome Archive

Data Commons
Trusted, controlled repository of multiple datasets

SPHN Swiss Personalized Health Network

Hub and Spoke
Common data elements, access, and usage rules



Linkage of distributed and disparate datasets

The EGA

European Genome-Phenome Archive

Long term secure archive for human biomedical research sensitive data, with focus on reuse of the data for further research (or "*broad and responsible use of genomic data*")

EUROPEAN GENOME-PHENOME ARCHIVE

CRG Centre for Genomic Regulation

EMBL-EBI

Global Alliance for Genomics & Health

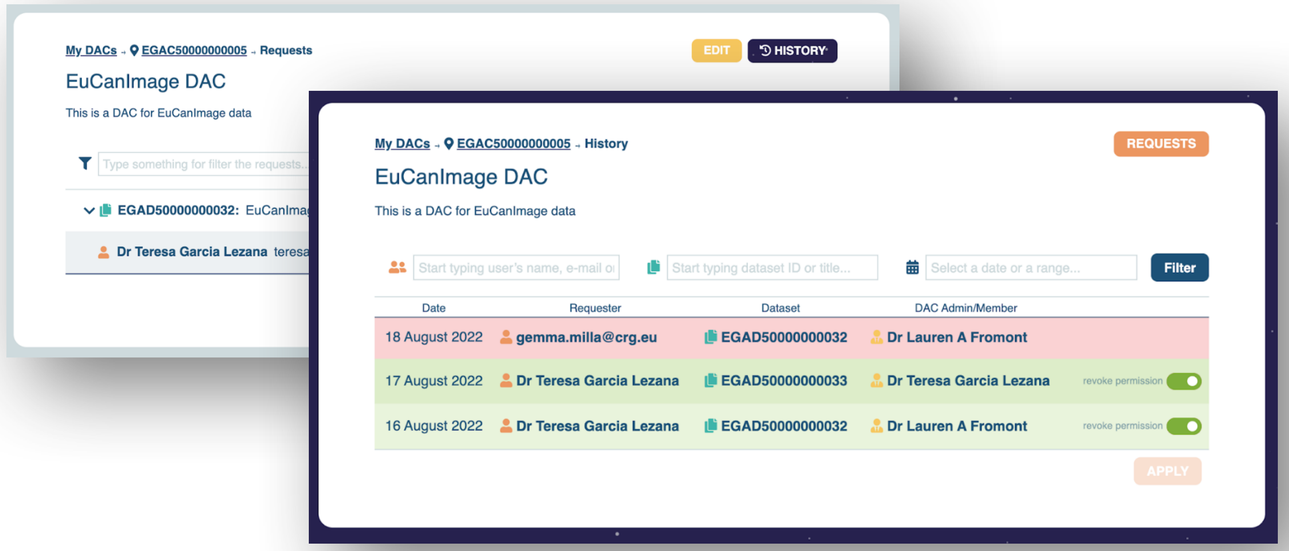
DRIVER PROJECT

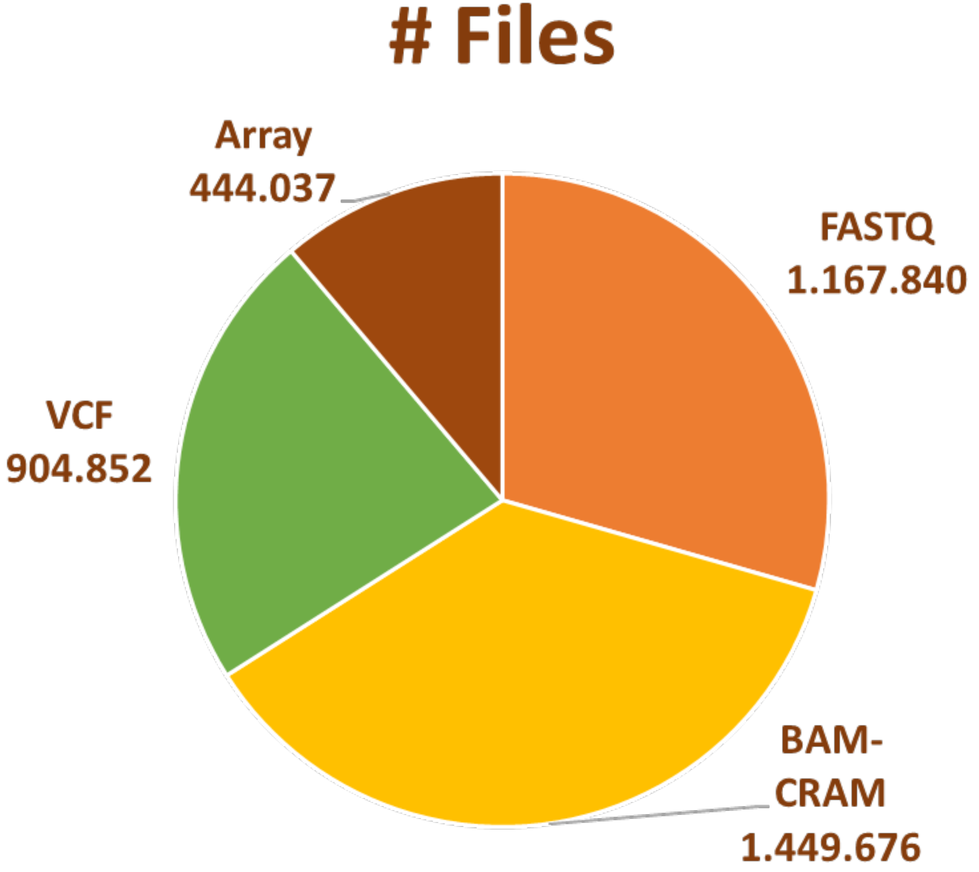
Slide: adapted from Jordi Rambla@ GA4GH 2023

The EGA

European Genome-Phenome Archive

- EGA “owns” nothing; data controllers tell who is authorized to access ***their*** datasets
- EGA admins provide smooth “all or nothing" data sharing process





4,328 Studies released
10,470 Datasets
2,309 Data Access Committees

Slide: adapted from Jordi Rambla@ GA4GH 2023

Different Approaches to Data Sharing

progenetix

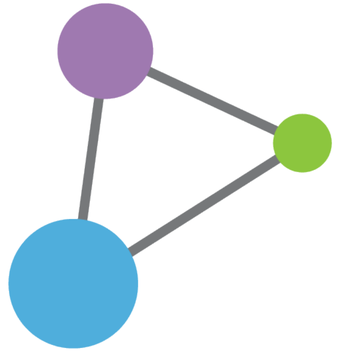
Centralized Genomic Knowledge Bases

European Genome-Phenome Archive

Data Commons
Trusted, controlled repository of multiple datasets

SPHN Swiss Personalized Health Network

Hub and Spoke
Common data elements, access, and usage rules



Linkage of distributed and disparate datasets

SPHN: a federal mandate

MANDATES

SAMS
Swiss Academy of Medical Sciences

SIB
Swiss Institute of Bioinformatics

GOVERNANCE

NSB
National Steering Board

BioMedIT Board

IMPLEMENTATION

Management Office
- Management of the overall initiative
- Daily operations

Personalized Health Informatics Group
- SPHN Data Coordination Centre (DCC)
- BioMedIT project

Expert Advisory Groups
- International Advisory Board (IAB)
- Scientific Expert Board (SEB)
- Ethic Legal Societal Implications Advisory Group (ELSIag)

DCC Working Groups
- Clinical Data Semantic Interoperability
- Bioinformatics and Data Analytics
- BioMedIT Interoperability
- Hospital IT
- IT Security

BioMedIT nodes
1. Core-IT/Vital-IT
2. sciCORE
3. SIS

PARTNERS

Universities, ETHZ, EPFL
University Hospitals
Personalized Health and Related Technologies (PHRT)

COLLABORATIONS

- eHealth Suisse
- Health 2030
- Personalized Health Alliance Zurich Basel (PH ZH-BS)
- swissuniversities
- Swiss Biobanking Platform (SBP)
- Swiss Clinical Trial Organisation (SCTO)
- Swiss Data Science Centre (SDSC)
- Swiss National Science Foundation (SNSF)
- unimedsuisse

SPHN FUNDED PROJECTS (29)

- Collaboration Agreements with University Hospitals
- Infrastructure Development Projects (2 jointly funded with PHRT)
- Driver Projects (4 jointly funded with PHRT)

PHRT FUNDED PROJECTS (56)

- Research Projects (4 jointly funded with SPHN)
- Translational Technology (2 projects jointly funded with SPHN)

Different Approaches to Data Sharing

progenetix

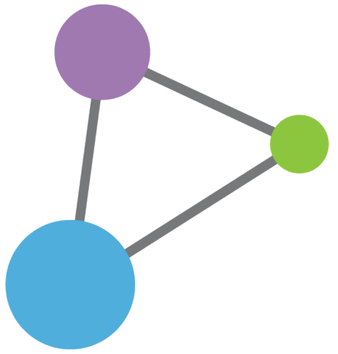
European Genome-Phenome Archive

SPHN Swiss Personalized Health Network

Centralized Genomic Knowledge Bases

Data Commons
Trusted, controlled repository of multiple datasets

Hub and Spoke
Common data elements, access, and usage rules



Linkage of distributed and disparate datasets

Federation

A New Paradigm for Data Sharing

FROM

TO

Data Copying

Data Visiting

Federation

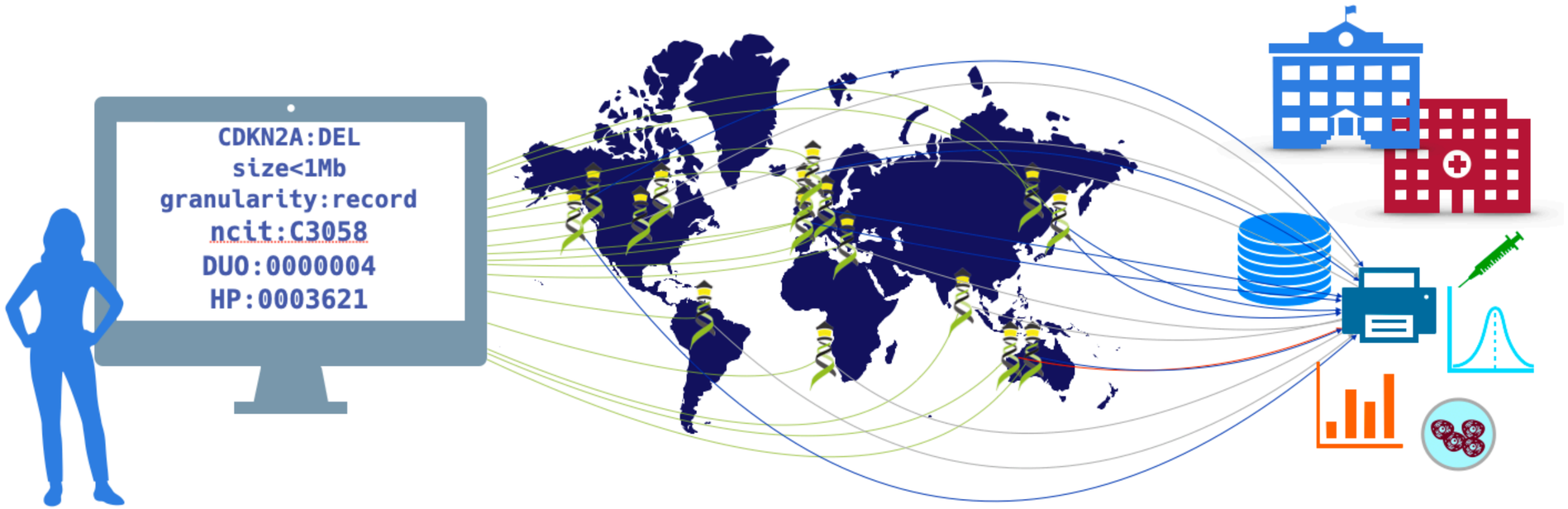
“A grouping of autonomous organizations and datasets with a centralised control”

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

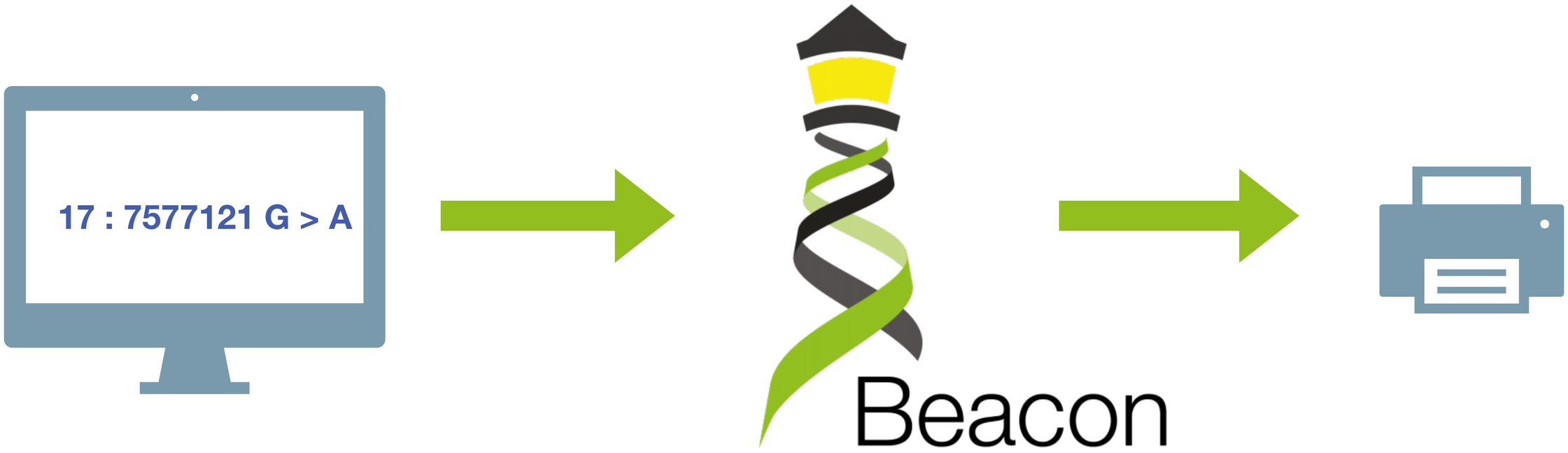
Global Alliance
for Genomics & Health

Collaborate. Innovate. Accelerate.



The GA4GH Beacon Protocol

Federating Genomic Discoveries



A ***Beacon*** answers a query for a specific genome variant against individual or aggregate genome collections

YES | **NO** | **\0**

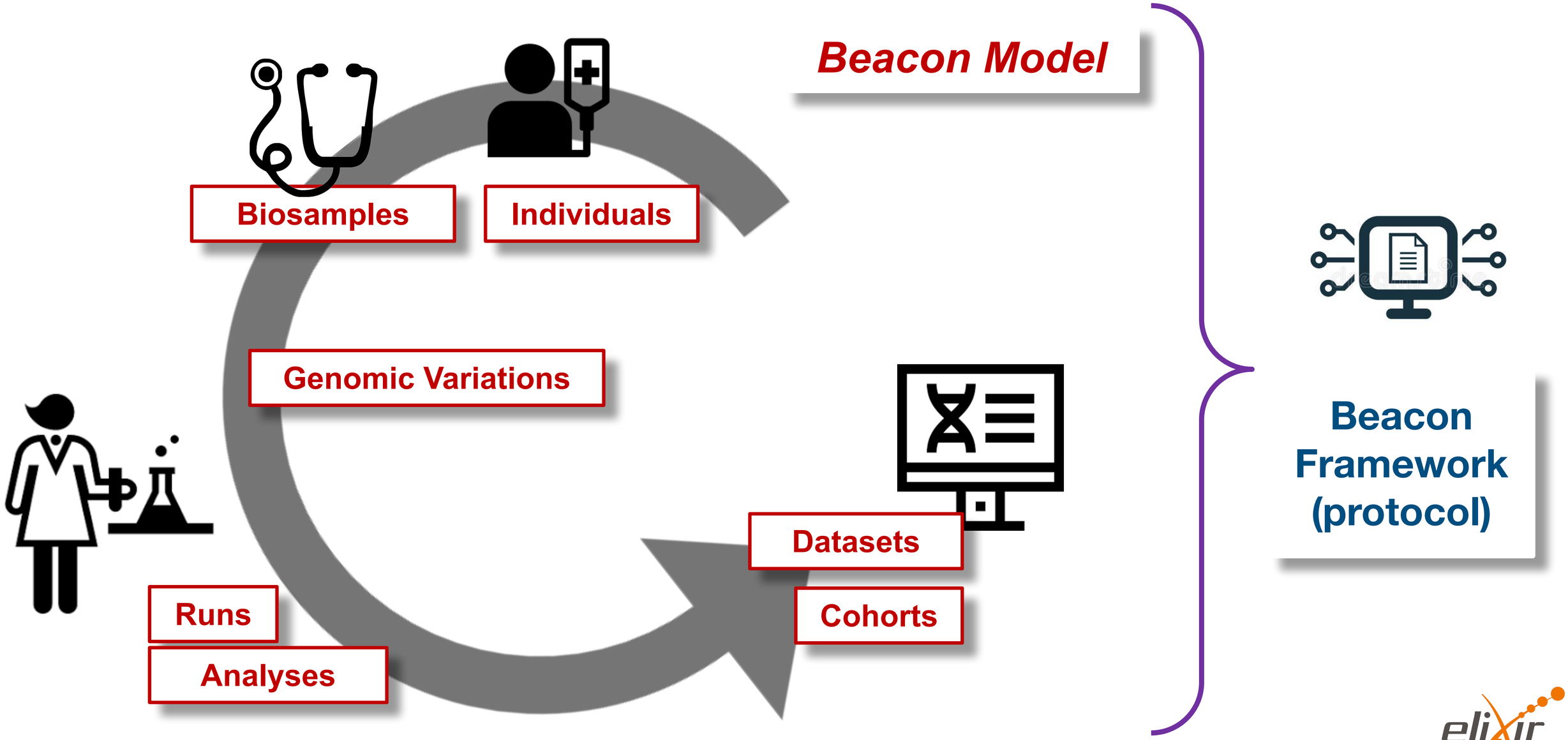
Have you seen this variant? It came up in my patient and we don't know if this is a common SNP or worth following up.

A Beacon network federates *genome variant queries* across databases that support the ***Beacon API***

Here: The variant has been found in **few** resources, and those are from **disease** specific **collections**.

Beacon v2

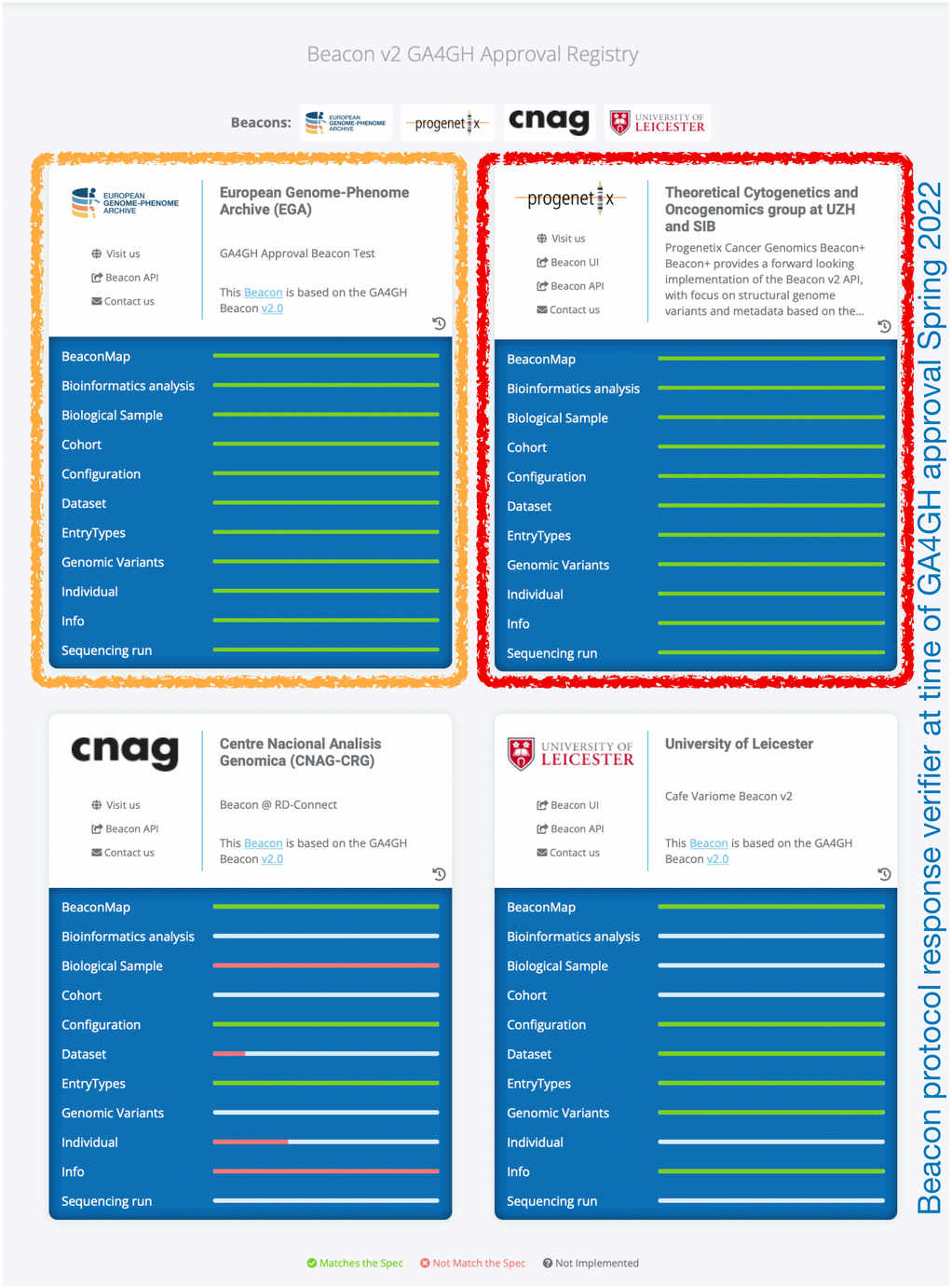
docs.genomebeacons.org



Progenetix & Beacon

Implementation driven standards development

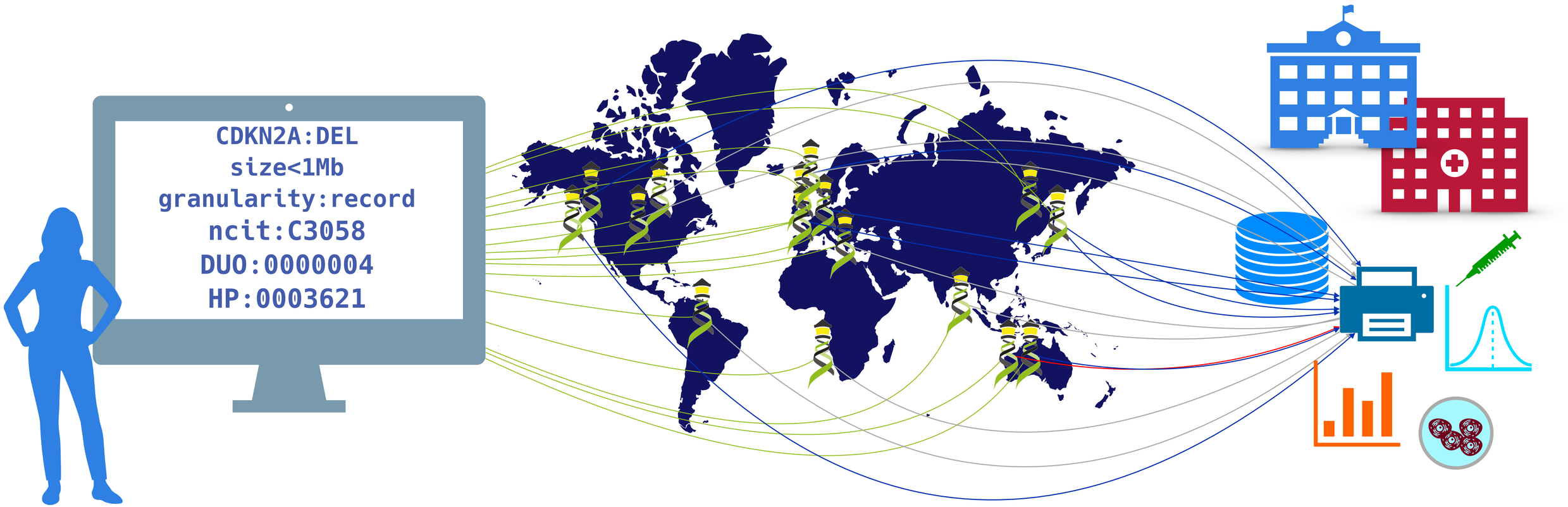
- Progenetix Beacon+ has served as implementation driver since 2016
- prototyping of advanced Beacon features such as
  - ➡ structural variant queries
  - ➡ data handovers
  - ➡ Phenopackets integration



Global Alliance for Genomics & Health

progenetix

elixir



Can you provide data about focal deletions in CDKN2A in Glioblastomas from juvenile patients with unrestricted access?

Beacon *v2* API

The Beacon API v2 represents a simple but powerful **genomics API** for ***federated*** data discovery and retrieval

The Beacon team through the ages

EUROPEAN GENOME-PHENOME ARCHIVE

CRG Centre for Genomic Regulation

Jordi Rambla
Arcadi Navarro
Roberto Ariosa
Manuel Rueda
Lauren Fromont
Mauricio Moldes
Claudia Vasallo
Babita Singh
Sabela de la Torre
Marta Ferri
Fred Haziza

CSC

Juha Törnroos
Teemu Kataja
Ilkka Lappalainen
Dylan Spalding

UNIVERSITY OF LEICESTER

Cafe Variome Central

Tony Brookes
Tim Beck
Colin Veal
Tom Shorter

Swiss Personalized Health Network (SPHN)

University of Zurich[UZH]

Michael Baudis
Rahel Paloots
Hangjia Zhao
Ziying Yang
Bo Gao

Genomics england

Augusto Rendon
Ignacio Medina
Javier López
Jacobo Coll
Antonio Rueda

cnag
centre nacional d'anàlisi genòmica
centro nacional de análisis genómico

Sergi Beltran
Carles Hernandez

Inserm
Institut national de la santé et de la recherche médicale

David Salgado

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Salvador Capella
Dmitry Repchevski
JM Fernández

DisGeNET

Laura Furlong
Janet Piñero

elixir

B1MG

Serena Scollen
Gary Saunders
Giselle Kerry
David Lloyd

H3Africa
Human Heredity & Health in Africa

Nicola Mulder
Mamana Mbiyavanga
Ziyaad Parker

EU Can CAN.

David Torrents

AUTISM SPEAKS

Dean Hartley

JUNTA DE ANDALUCIA
Fundación Progreso y Salud
CONSEJERÍA DE SALUD

Joaquin Dopazo
Javier Pérez
J.L. Fernández
Gema Roldan

CINECA

Thomas Keane
Melanie Courtot
Jonathan Dursi

Heidi Rehm
Ben Hutton

Toshiaki Katayama

GEM Japan

Stephane Dyke

DNASTACK

Marc Fiume
Miro Cupak

BRCA EXCHANGE

Melissa Cline

ENA

EMBL-EBI

Diana Lemos

EUROPEAN JOINT PROGRAMME RARE DISEASES

VICC Variant Interpretation for Cancer Consortium

GA4GH Phenopackets
Peter Robinson
Jules Jacobsen

GA4GH VRS
Alex Wagner
Reece Hart

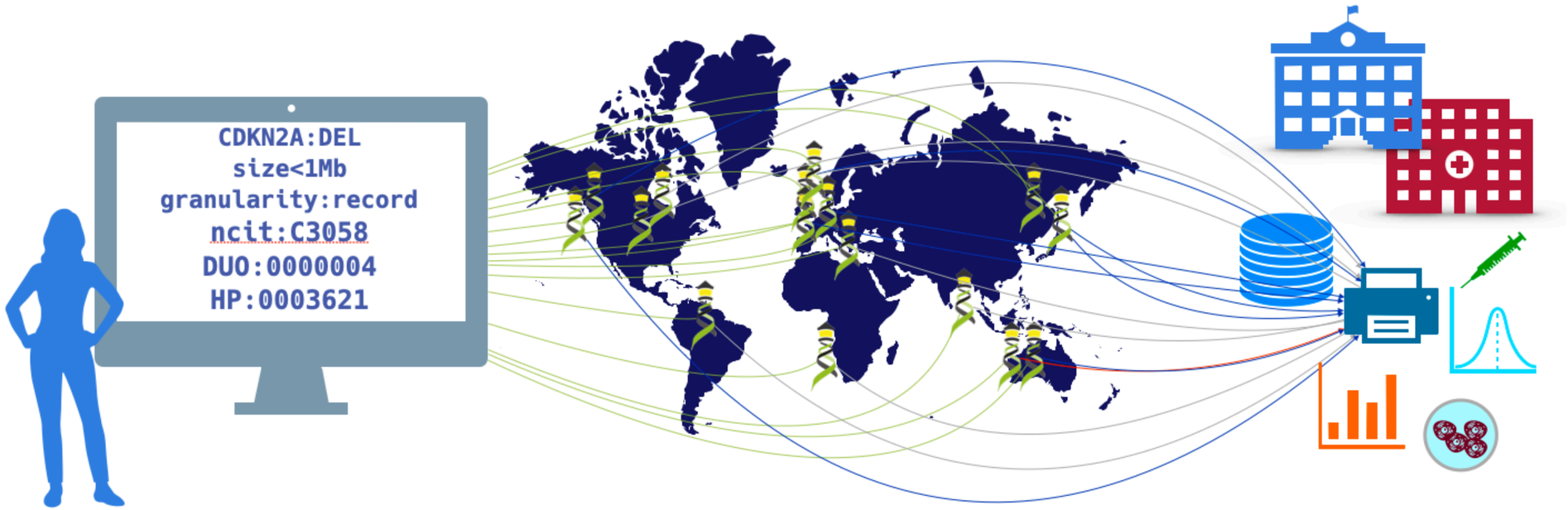
Beacon PRC
Alex Wagner
Jonathan Dursi
Mamana Mbiyavanga

Alice Mann
Neerjah Skantharajah

elixir

Global Alliance
for Genomics & Health

Collaborate. Innovate. Accelerate.

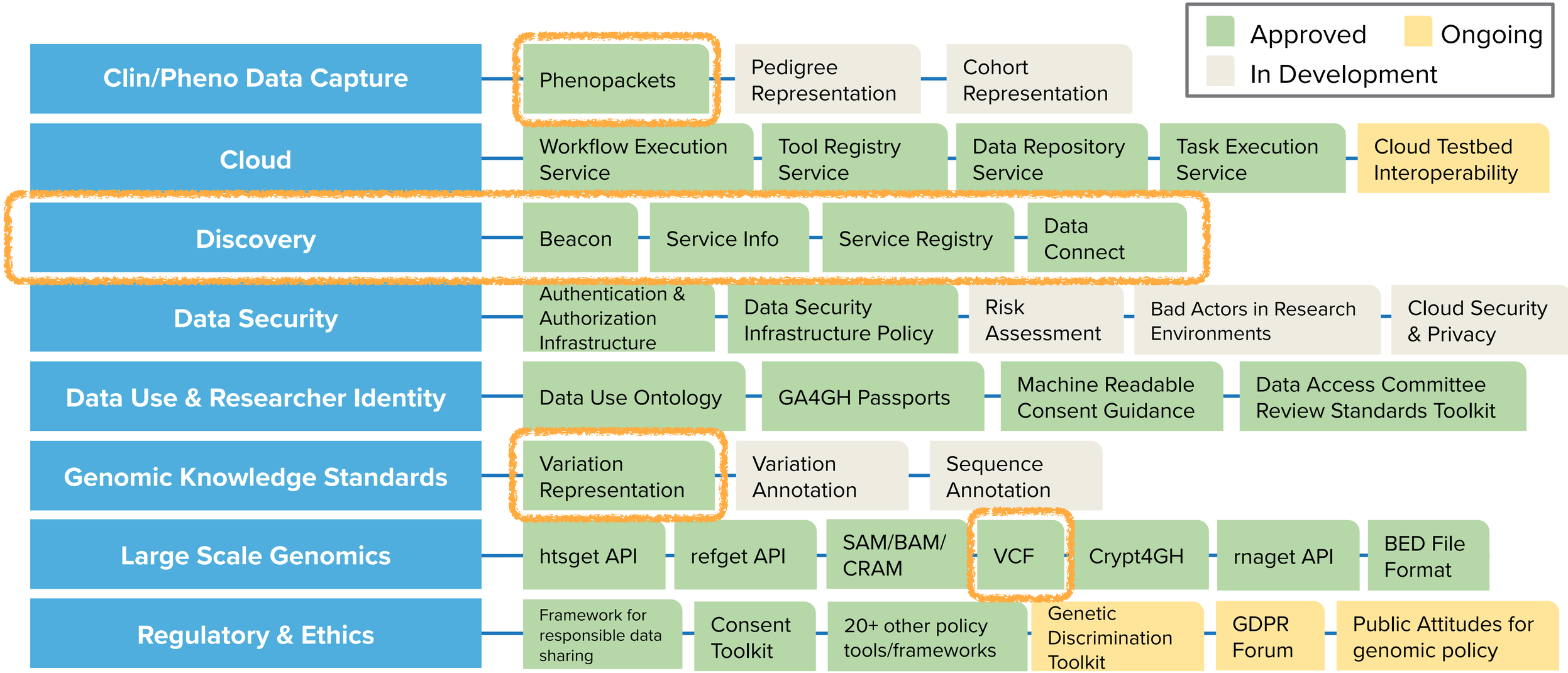


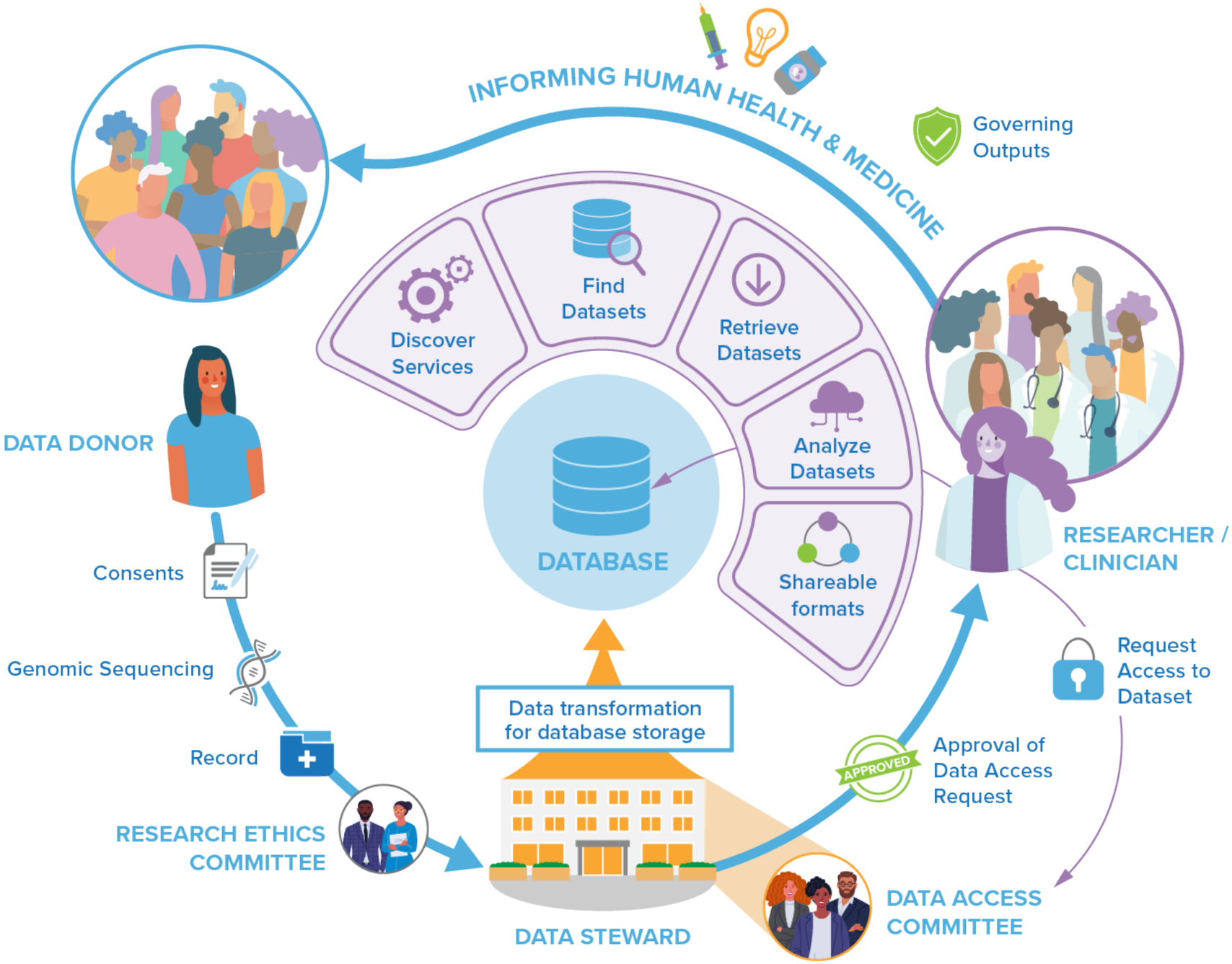
The GA4GH Beacon Protocol

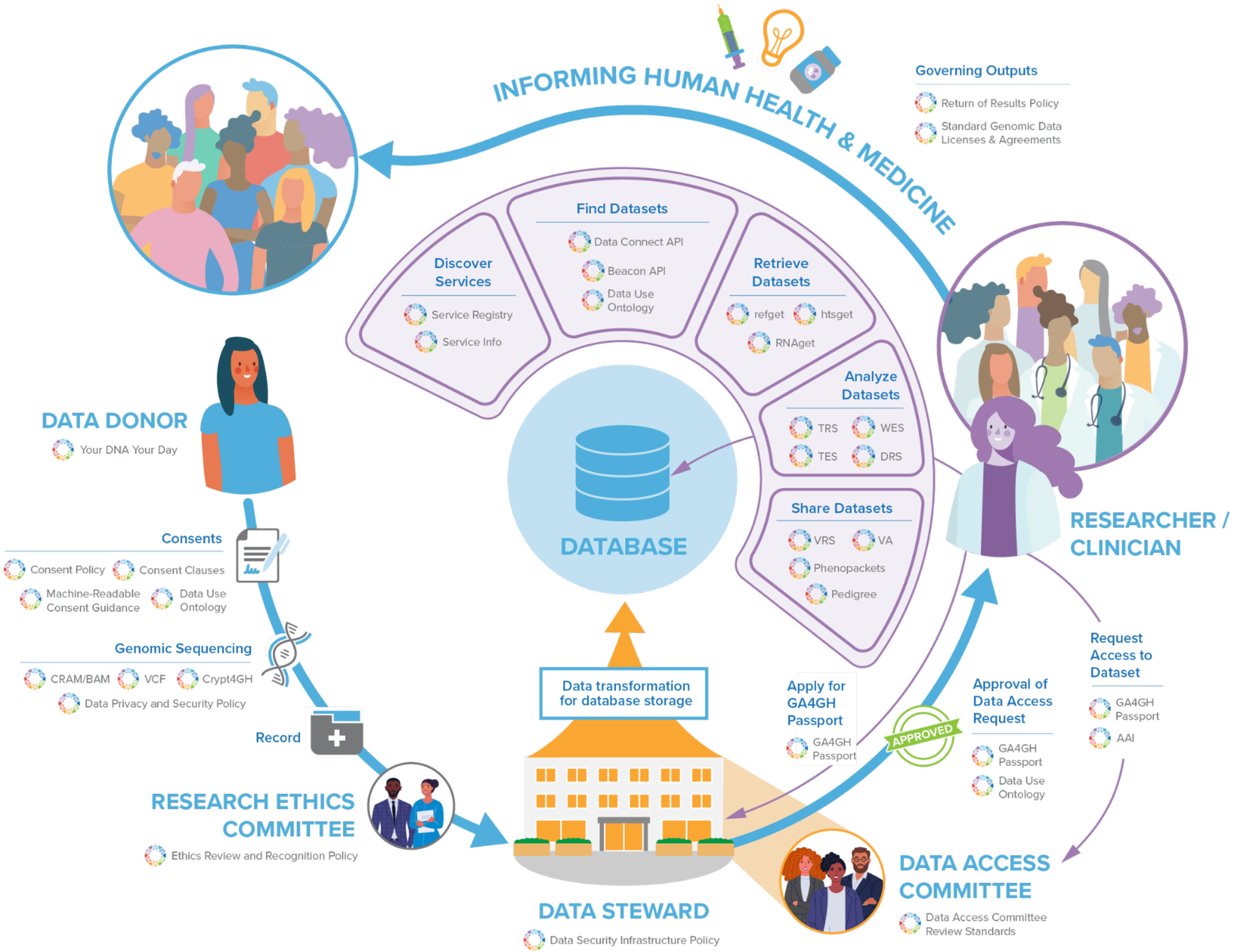
... is only one part of the GA4GH ecosystem

GA4GH ecosystem and outputs

Overview of GA4GH standards and frameworks




INFORMING HUMAN HEALTH & MEDICINE
Governing Outputs
Discover Services
Find Datasets
Retrieve Datasets
Analyze Datasets
Shareable formats
DATABASE
DATA DONOR
Consents
Genomic Sequencing
Record
RESEARCH ETHICS COMMITTEE
Data transformation for database storage
DATA STEWARD
APPROVED
Approval of Data Access Request
Request Access to Dataset
RESEARCHER / CLINICIAN
DATA ACCESS COMMITTEE


INFORMING HUMAN HEALTH & MEDICINE
Governing Outputs
Return of Results Policy
Standard Genomic Data Licenses & Agreements
Find Datasets
Data Connect API
Beacon API
Data Use Ontology
Discover Services
Service Registry
Service Info
Retrieve Datasets
refget
htsget
RNAget
Analyze Datasets
TRS
WES
TES
DRS
Share Datasets
VRS
VA
Phenopackets
Pedigree
DATABASE
DATA DONOR
Your DNA Your Day
RESEARCHER / CLINICIAN
Consents
Consent Policy
Consent Clauses
Machine-Readable Consent Guidance
Data Use Ontology
Genomic Sequencing
CRAM/BAM
VCF
Crypt4GH
Data Privacy and Security Policy
Record
Data transformation for database storage
Apply for GA4GH Passport
GA4GH Passport
APPROVED
Approval of Data Access Request
GA4GH Passport
Data Use Ontology
Request Access to Dataset
GA4GH Passport
AAI
RESEARCH ETHICS COMMITTEE
Ethics Review and Recognition Policy
DATA STEWARD
Data Security Infrastructure Policy
DATA ACCESS COMMITTEE
Data Access Committee Review Standards

Global Collaborations

Global Alliance for Genomics & Health

International HundredK+ Cohorts Consortium

Global Genomic Medicine Consortium

Beyond One Million Genomes Project

WORLD ECONOMIC FORUM

THE MEDICAL GENOME INITIATIVE

World Health Organization Science Council

Human Pangenome Project (HPP) / Human Pangenome Reference Consortium (HPRC)

INTERNATIONAL COMMON DISEASE ALLIANCE | Maps to Mechanisms to Medicine

Public Health Alliance for Genomic Epidemiology

CRDSA Clinical Research Data Sharing Alliance

ICODA International COVID-19 Data Alliance

elixir

HUMAN GENOME ORGANISATION HUGO

THE HUMAN VARIOME PROJECT

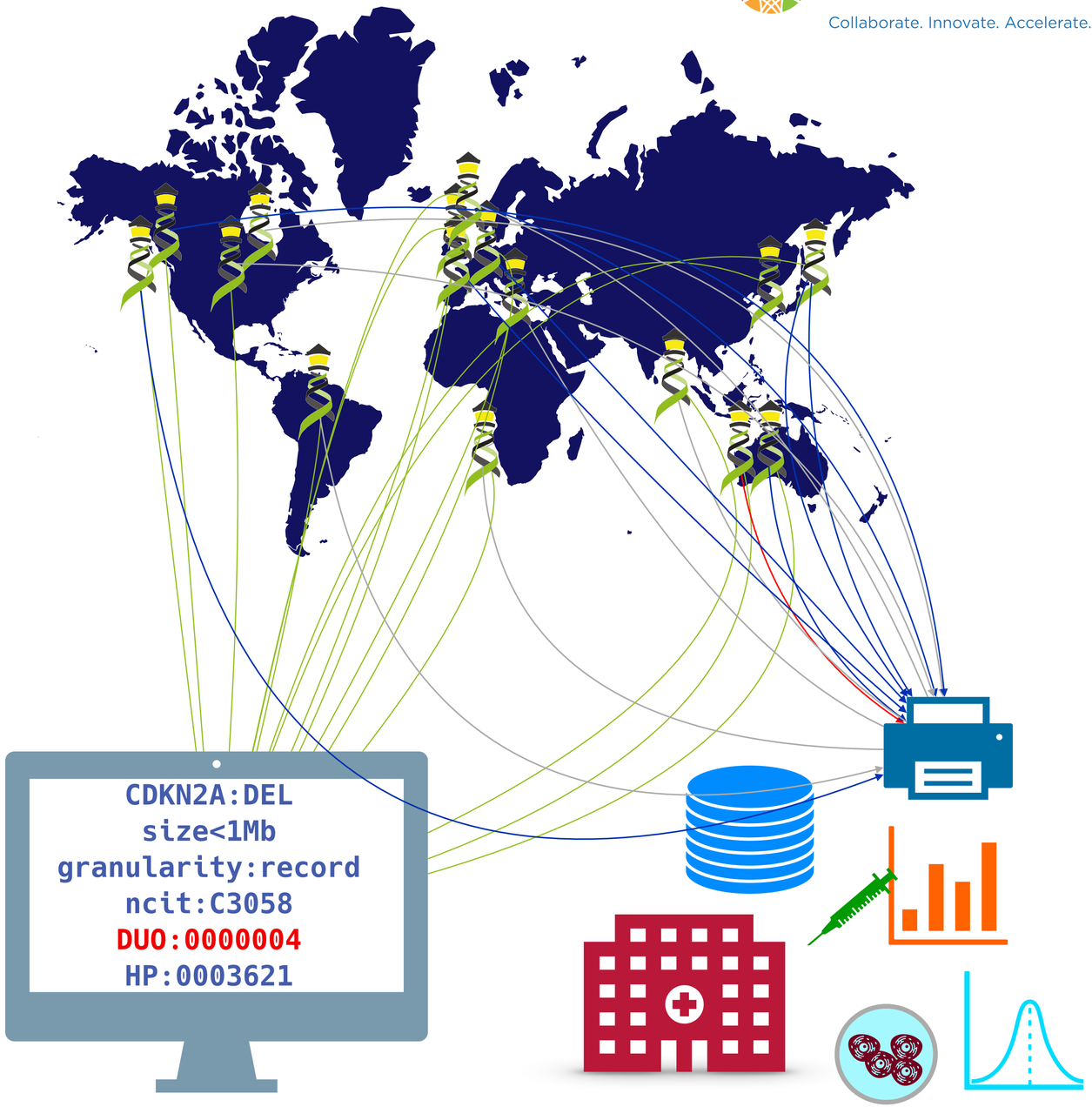
HGVS HUMAN GENOME VARIATION SOCIETY

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

What Can You Do?

- implement procedures and standards supporting **data discovery** (FAIR principles) and federation approaches
- forward looking consent and data protection models adhering to **ORD** principles ("as secure as necessary, as open as possible")
- get involved with international **data standards** efforts and projects

➡ **Collaborate!**

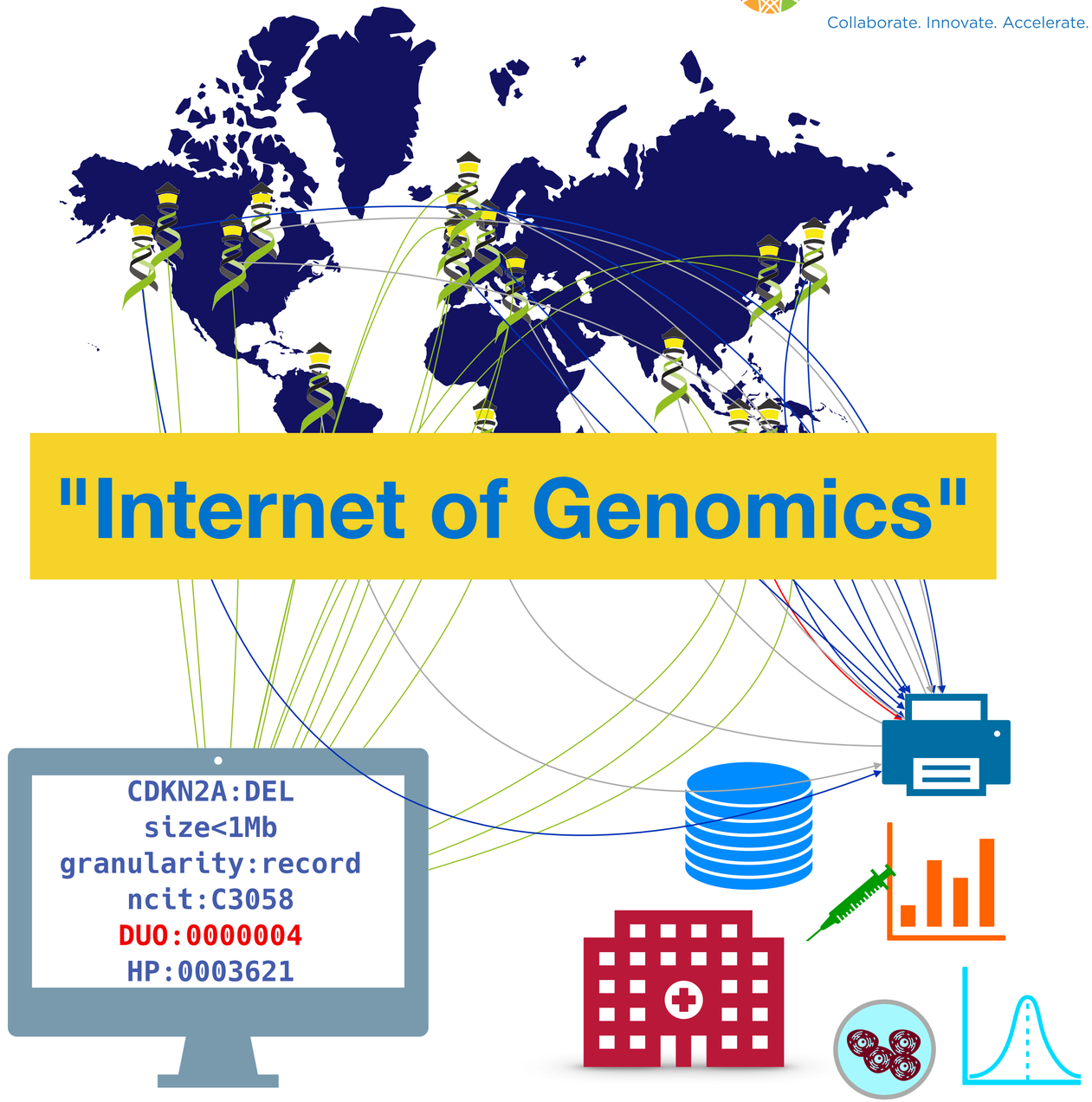


Global Alliance
for Genomics & Health
Collaborate. Innovate. Accelerate.

What Can You Do?

- implement procedures and standards supporting **data discovery** (FAIR principles) and federation approaches
- forward looking consent and data protection models adhering to **ORD** principles ("as secure as necessary, as open as possible")
- get involved with international **data standards** efforts and projects

➡ **Collaborate!**




Universität Zürich[UZH]
SIB
Swiss Institute of Bioinformatics
Global Alliance
for Genomics & Health
Collaborate. Innovate. Accelerate.
elixir
SPHN