On the EGA* Ecosystem

(*European Genome-phenome Archive)

Arcadi Navarro

November the 13th, 2023

EUROPEAN GENOME-PHENOME ARCHIVE

elixir CONVERGE

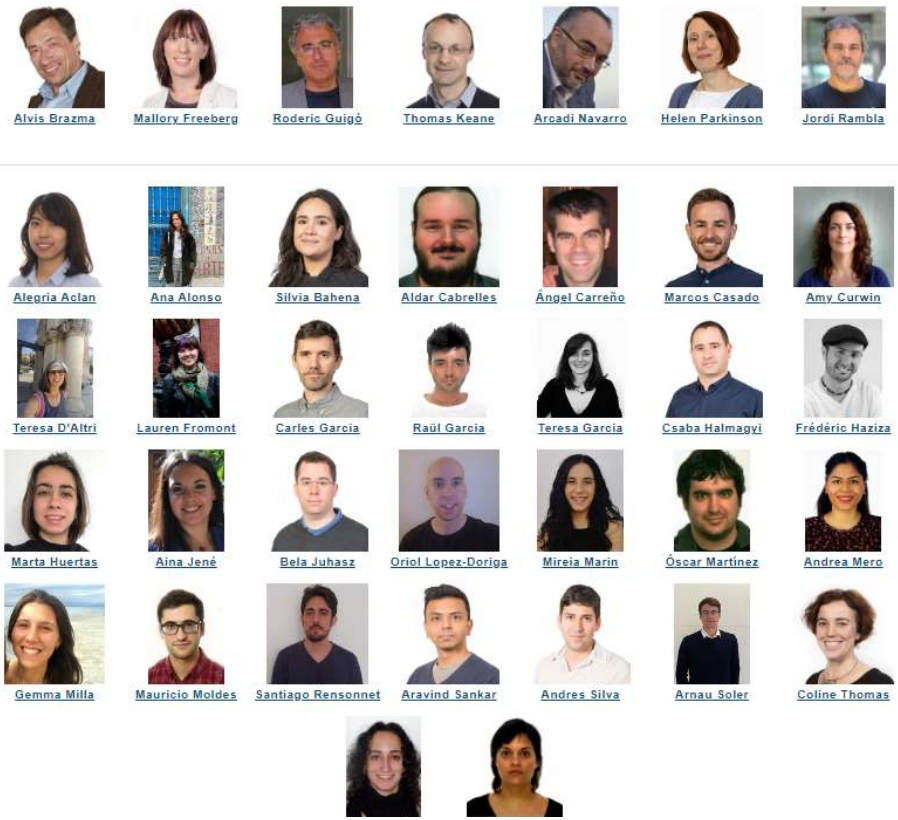
These projects have received funding from the European Union's Horizon 2020 research and innovation programme

CRG Centre for Genomic Regulation

EMBL-EBI

Thanks!

The Team



Core organizations:

CRG Centre for Genomic Regulation

EMBL

Additional sources:

Gobierno de España · Ministerio de Economía, Industria y Competitividad · Instituto de Salud Carlos III

UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional
"Una manera de hacer Europa"

Generalitat de Catalunya

"la Caixa"

elixir

HORIZON 2020

SEVENTH FRAMEWORK PROGRAMME

EXCELENCIA SEVERO OCHOA

FNIH
Foundation for the National Institutes of Health

Accelerating Medicines Partnership

RD-Connect

And critical support from the following sources:

BSC Barcelona Supercomputing Center
Centro Nacional de Supercomputación

upf. Universitat Pompeu Fabra Barcelona

INB Spanish National Bioinformatics Institute

https://ega-archive.org/support

Outline

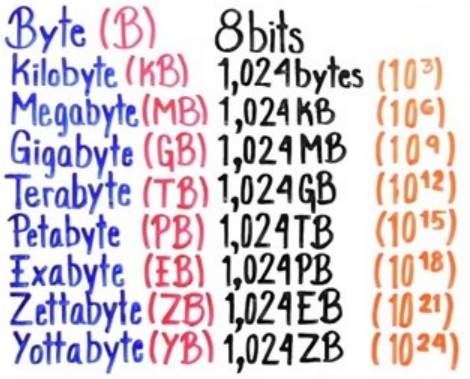
1. Big Data & Genomic Big Data
2. Genomic Data
3. The sharing dilemma
4. What is the EGA for?
5. Beacon
6. The EGA Federation

Outline

1. Big Data & Genomic Big Data
2. Genomic Data
3. The sharing dilemma
4. What is the EGA for?
5. Beacon
6. The EGA Federation

what do all of these have in common?

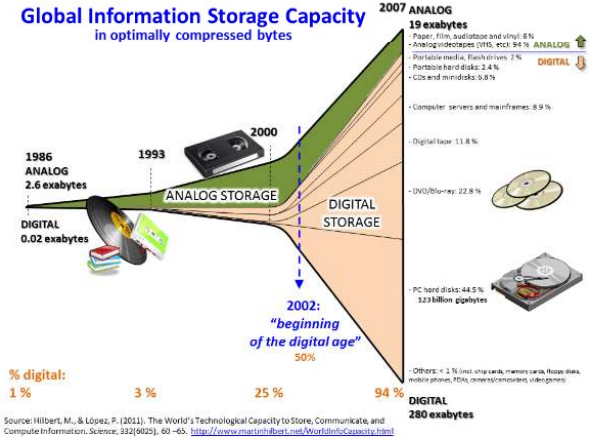
what is “BigData”?



Think for a sec… what does “Big Data” mean?

There are huge volumes of data in the world:

- From the beginning of recorded time until 2003, humans created 5 billion gigabytes of data.
- In 2011, the same amount was created every two days
- In 2013, every 10 minutes.
- The best estimation for 2020 was that 1.7MB of data was created **every second by every person** on earth*. This is 13PB per second or >1 exabyte per minute.
- By 2025, there will be 175 zettabytes of data in the global datasphere.



*https://www.domo.com/solution/data-never-sleeps-6

what is “BigData”?

Think for a sec… what does “Big Data” mean?

This is growing so much that **90% of the data in the world today has been created in less than the last year**. This data comes from everywhere: sensors used to gather climate information, posts to social media sites, digital pictures and videos, medical records, purchase transaction records or cell phone GPS signals to name a few.

This data is Big Data

what is “BigData”?

…and a tiny fraction of these data belongs to the fields of Medical Genomics

…but this tiny fraction is still a huge amount of data!

where does genomic data come from?

The promise in the 90s: “In 10 years we will unravel the genetic bases of complex diseases!!” … **It was the reason under the Human Genome Project**

20TH Anniversary of The e

2021)

Science aspects: High impact on science and healthcare

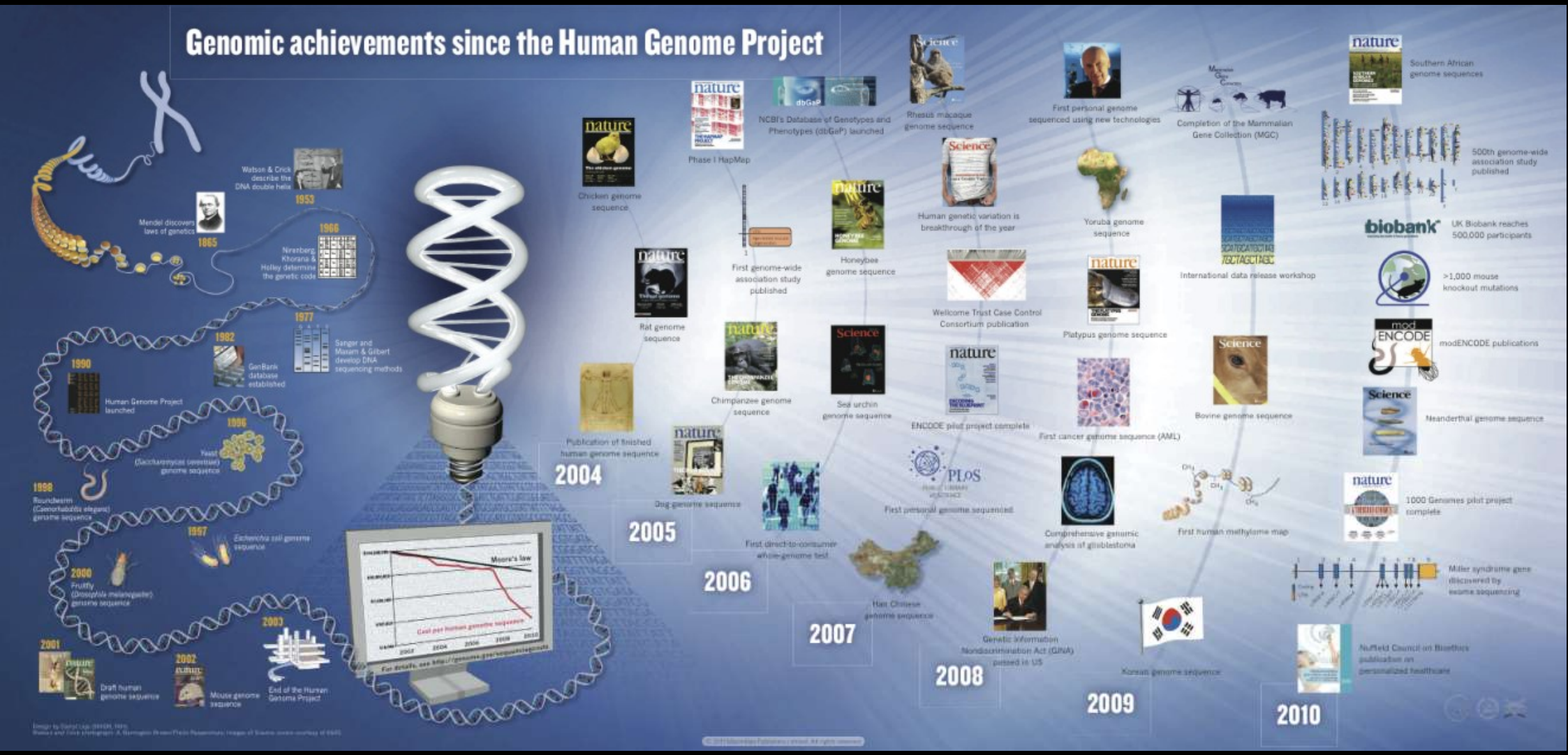
Data sharing aspects: Bermuda & Fort Lauderdale Principles > or "*Good vs Evil*"

Technological advances that these projects had generated:

- Genomic Browsers (and tracks)
- Haplotype viewers
- VCF format
- …

Issues:

- Lack of diversity
- Going backwards on data sharing?


Genomic achievements since the Human Genome Project
1865
Mendel discovers laws of genetics
1953
Watson & Crick describe the DNA double helix
1966
Nirenberg, Khorana & Holley determine the genetic code
1977
Sanger and Maxam & Gilbert develop DNA sequencing methods
1982
GenBank database established
1990
Human Genome Project launched
1996
Yeast (Saccharomyces cerevisiae) genome sequence
1998
Roundworm (Caenorhabditis elegans) genome sequence
1997
Escherichia coli genome sequence
2000
Fruitfly (Drosophila melanogaster) genome sequence
2001
Draft human genome sequence
2002
Mouse genome sequence
2003
End of the Human Genome Project
Moore's law
Cost per human genome sequence
2004
Publication of finished human genome sequence
Chicken genome sequence
Rat genome sequence
Dog genome sequence
2005
Phase I HapMap
First genome-wide association study published
Chimpanzee genome sequence
First direct-to-consumer whole-genome test
2006
NCBI's Database of Genotypes and Phenotypes (dbGaP) launched
Honeybee genome sequence
Sea urchin genome sequence
Han Chinese genome sequence
2007
Rhesus macaque genome sequence
Human genetic variation is breakthrough of the year
Wellcome Trust Case Control Consortium publication
ENCODE pilot project complete
First personal genome sequenced
Genetic Information Nondiscrimination Act (GINA) passed in US
2008
First personal genome sequenced using new technologies
Yoruba genome sequence
Platypus genome sequence
First cancer genome sequence (AML)
Comprehensive genomic analysis of glioblastoma
Korean genome sequence
2009
Completion of the Mammalian Gene Collection (MGC)
International data release workshop
Bovine genome sequence
First human methylome map
2010
Southern African genome sequences
500th genome-wide association study published
UK Biobank reaches 500,000 participants
>1,000 mouse knockout mutations
modENCODE publications
Neanderthal genome sequence
1000 Genomes pilot project complete
Miller syndrome gene discovered by exome sequencing
Nuffield Council on Bioethics publication on personalized healthcare

Outline

1. Big Data & Genomic Big Data
2. Genomic Data
3. The sharing dilemma
4. What is the EGA for?
5. Beacon
6. The EGA Federation

The enablers: The HapMap project

The International HapMap Consortium
A haplotype map of the human genome
Nature **437,** 1299–1320 (2005)
https://doi.org/10.1038/nature04226

The HapMap is a map of haplotype blocks and the specific SNPs that identify the haplotypes (tag SNPs).

The HapMap is valuable by reducing the number of SNPs required to examine the entire genome for association with a phenotype from the 10 million SNPs that exist to roughly 500,000 tag SNPs.

269 DNA samples from four “populations”

The project follows the data release principles of sharing information rapidly and without restriction on its use.

- Officially started October 27 to 29, 2002
- The complete data Phase I published 27 October 2005.
- Phase II dataset published October 2007.
- Phase III dataset released spring 2009 published in September 2010.

Data available at Index of /hapmap (nih.gov)

The enablers: 1000 Genomes Project

https://www.internationalgenome.org

Launched in January 2008. Since 2015 managed by the IGSR: *The International Genome Sample Resource*

Has had several phases: Pilot (2010),Phase I (2012): 1,092 samples, Phase II (technical), Phase III (2015)

Uses a combination of low-coverage whole-genome sequencing, deep exome sequencing, and dense microarray genotyping

In early 2020, released high-coverage (30x) data for an additional 698 samples.

The samples are anonymous and have **no associated medical or phenotype data**. The project holds self-reported ethnicity and gender. All participants declared themselves to be healthy at the time the samples were collected.

Data was quickly made available to the scientific community through freely accessible public databases

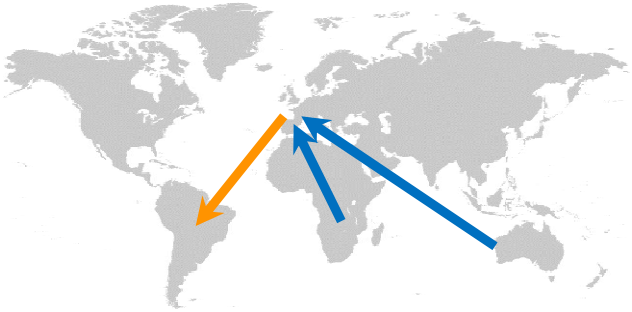
Files accessible from FTP sites hosted by EMBL-EBI and NCBI.

Now also hosted on Amazon Web Services (AWS) and AnVIL.

Thousands of papers have been published that use these datasets

Data access models

Fully Open research data



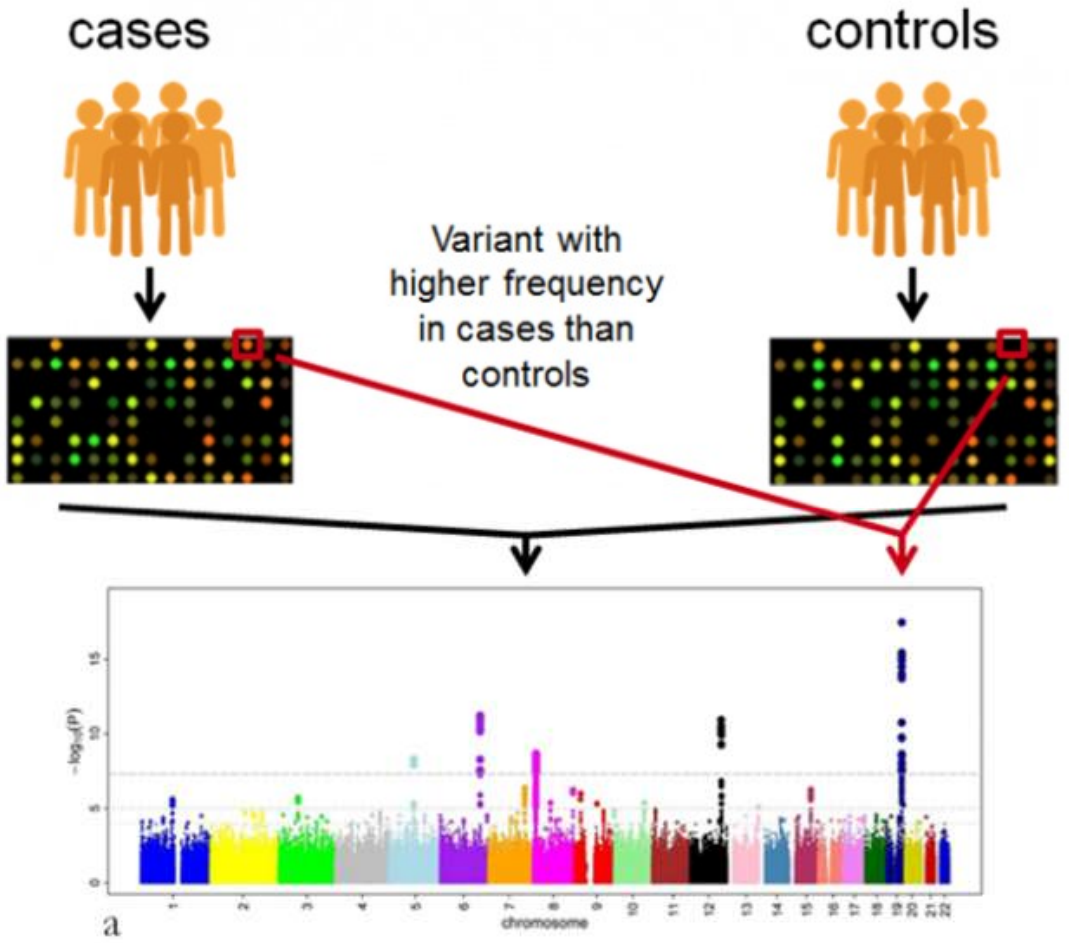
Aggregate data globally

Download, analyse locally

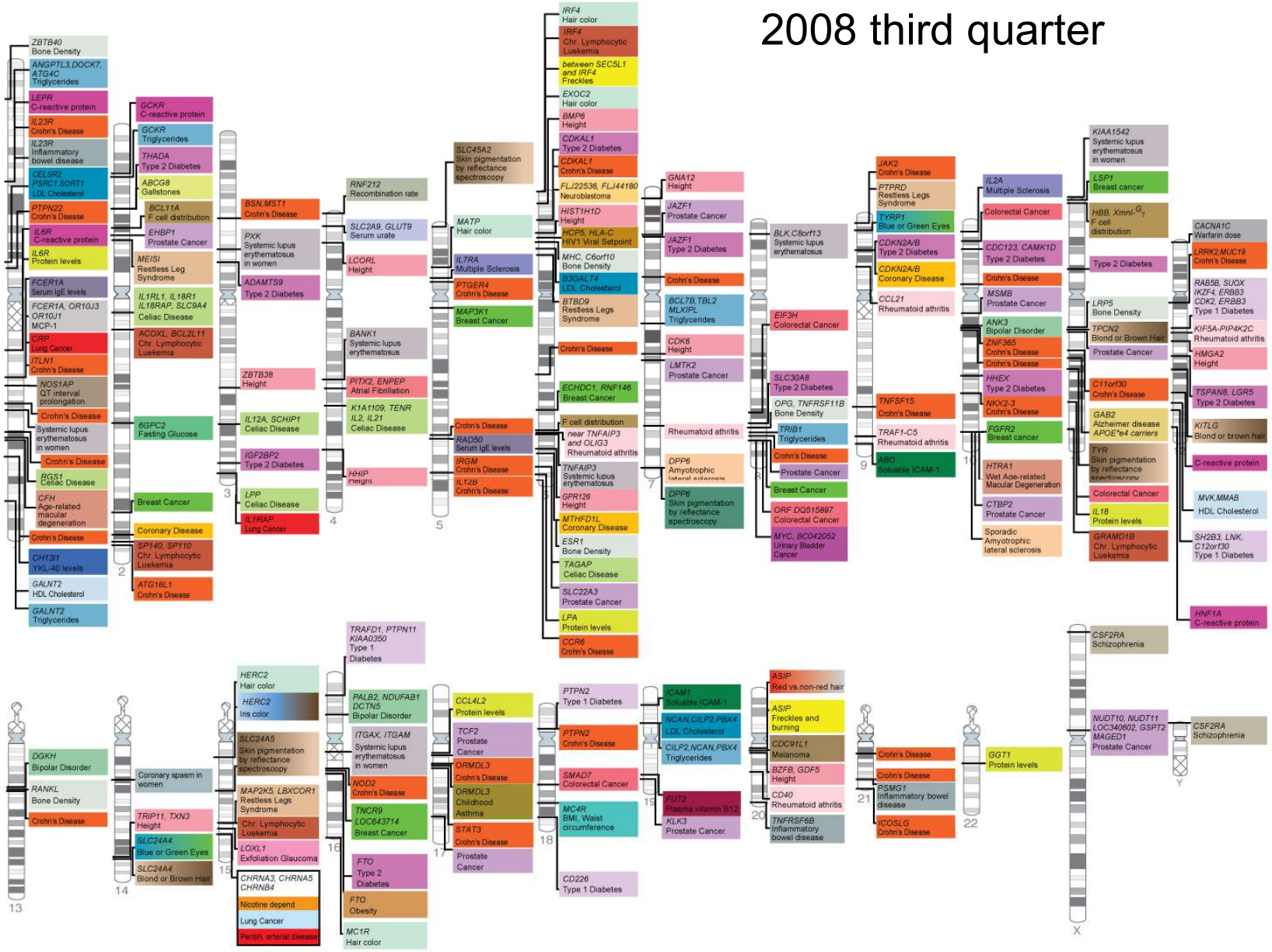
e.g., Plant and animal genomes, Population-based results, Some GWAS summary statistics

Accumulation of human data

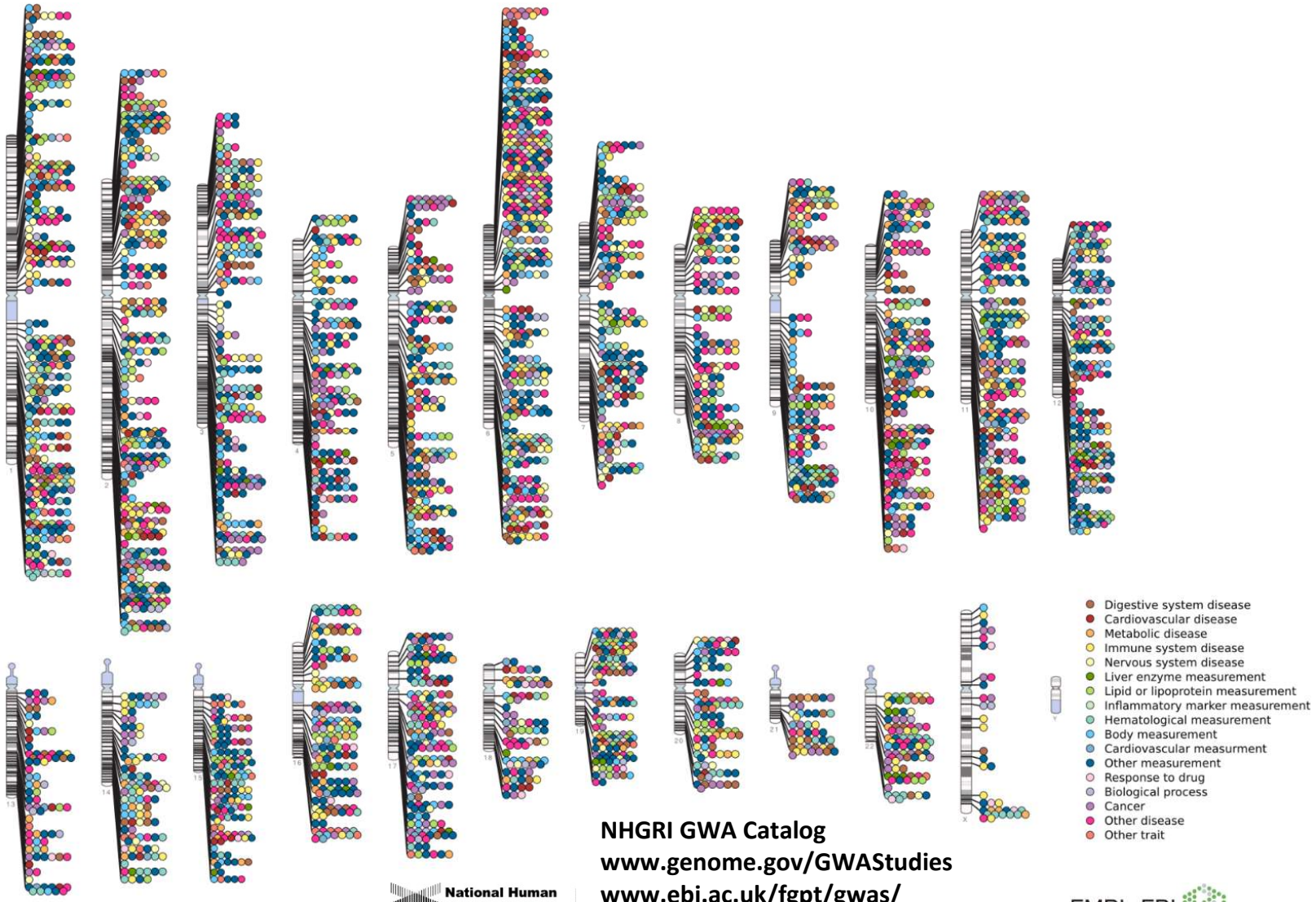
Using all these advances, many data relevant for human health has accumulated. For instance, take Genome Wide Association Studis (GWAS), which started in 2005-2007



2008 third quarter



Published Genome-Wide Associations through 12/2013
Published GWA at $p \leq 5X10^{-8}$ for 17 trait categories



NHGRI GWA Catalog
www.genome.gov/GWAStudies
www.ebi.ac.uk/fgpt/gwas/

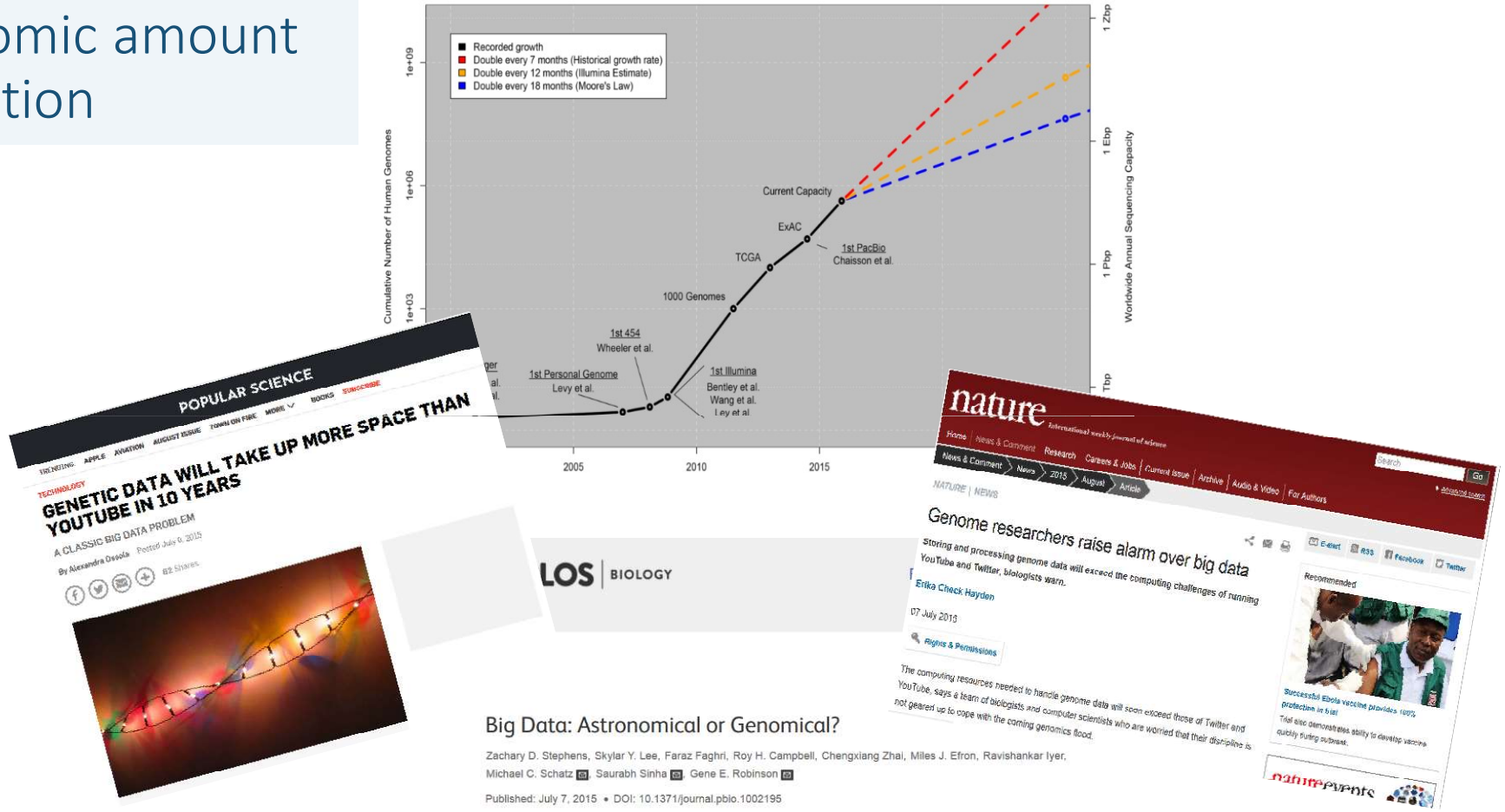
National Human Genome Research Institute

EMBL-EBI

Yesterday's version


Abdominal aortic aneurysm
Acute lymphoblastic leukemia
Adhesion molecules
Adverse response to carbamapezine
Adiponectin levels
Age-related macular degeneration
AIDS progression
Alcohol dependence
Alopecia areata
Alzheimer disease
Amyloid A levels
Amyotrophic lateral sclerosis
Angiotensin-converting enzyme activity
Ankylosing spondylitis
Arterial stiffness
Asparagus anosmia
Asthma
Atherosclerosis in HIV
Atrial fibrillation
Attention deficit hyperactivity disorder
Autism
Basal cell cancer
Behcet's disease
Bipolar disorder
Biliary atresia
Bilirubin
Bitter taste response
Birth weight
Bladder cancer
Bleomycin sensitivity
Blond or brown hair
Blood pressure
Blue or green eyes
BMI, waist circumference
Bone density
Breast cancer
C-reactive protein
Calcium levels
Cardiac structure/function
Carnitine levels
Carotenoid/tocopherol levels
Celiac disease
Cerebral atrophy measures
Chronic lymphocytic leukemia
Cleft lip/palate
Cognitive function
Conduct disorder
Colorectal cancer
Corneal thickness
Coronary disease
Creutzfeldt-Jakob disease
Crohn's disease
Cutaneous nevi
Dermatitis
Drug-induced liver injury
Endometriosis
Eosinophil count
Eosinophilic esophagitis
Erectile dysfunction and prostate cancer treatment
Erythrocyte parameters
Esophageal cancer
Essential tremor
Exfoliation glaucoma
Eye color traits
F cell distribution
Fibrinogen levels
Folate pathway vitamins
Follicular lymphoma
Fuch's corneal dystrophy
Freckles and burning
Gallstones
Gastric cancer
Glioma
Glycemic traits
Hair color
Hair morphology
Handedness in dyslexia
HDL cholesterol
Heart failure
Heart rate
Height
Hemostasis parameters
Hepatic steatosis
Hepatitis
Hepatocellular carcinoma
Hirschsprung's disease
HIV-1 control
Hodgkin's lymphoma
Homocysteine levels
Hypospadias
Idiopathic pulmonary fibrosis
IgA levels
IgE levels
Inflammatory bowel disease
Intracranial aneurysm
Iris color
Iron status markers
Ischemic stroke
Juvenile idiopathic arthritis
Keloid
Kidney stones
LDL cholesterol
Leprosy
Leptin receptor levels
Liver enzymes
Longevity
LP (a) levels
LpPLA(2) activity and mass
Lung cancer
Magnesium levels
Major mood disorders
Malaria
Male pattern baldness
Matrix metalloproteinase levels
MCP-1
Melanoma
Menarche & menopause
Meningococcal disease
Metabolic syndrome
Migraine
Moyamoya disease
Multiple sclerosis
Myeloproliferative neoplasms
N-glycan levels
Narcolepsy
Nasopharyngeal cancer
Neuroblastoma
Nicotine dependence
Obesity
Open angle glaucoma
Open personality
Optic disc parameters
Osteoarthritis
Osteoporosis
Otosclerosis
Other metabolic traits
Ovarian cancer
Pancreatic cancer
Pain
Paget's disease
Panic disorder
Parkinson's disease
Periodontitis
Peripheral arterial disease
Phosphatidylcholine levels
Phosphorus levels
Photic sneeze
Phytosterol levels
Platelet count
Polycystic ovary syndrome
Primary biliary cirrhosis
Primary sclerosing cholangitis
PR interval
Progranulin levels
Prostate cancer
Protein levels
PSA levels
Psoriasis
Psoriatic arthritis
Pulmonary funct. COPD
QRS interval
QT interval
Quantitative traits
Recombination rate
Red vs.non-red hair
Refractive error
Renal cell carcinoma
Renal function
Response to antidepressants
Response to antipsychotic therapy
Response to hepatitis C treat
Response to metaformin
Response to statin therapy
Restless legs syndrome
Retinal vascular caliber
Rheumatoid arthritis
Ribavirin-induced anemia
Schizophrenia
Serum metabolites
Skin pigmentation
Smoking behavior
Speech perception
Sphingolipid levels
Statin-induced myopathy
Stroke
Systemic lupus erythematosus
Systemic sclerosis
T-tau levels
Tau AB1-42 levels
Telomere length
Testicular germ cell tumor
Thyroid cancer
Tooth development
Total cholesterol
Triglycerides
Tuberculosis
Type 1 diabetes
Type 2 diabetes
Ulcerative colitis
Urate
Venous thromboembolism
Ventricular conduction
Vertical cup-disc ratio
Vitamin B12 levels
Vitamin D insuffiency
Vitiligo
Warfarin dose
Weight
White cell count
YKL-40 levels

An astronomic amount of information



Amazing results!

nature reviews drug discovery

Explore content · About the journal · Publish with us · Subscribe

nature › nature reviews drug discovery › biobusiness briefs › article

BIOBUSINESS BRIEFS | 08 July 2022

Human genetics evidence supports two-thirds of the 2021 FDA-approved drugs

David Ochoa, Mohd Karim, Maya Ghoussaini, David G. Hulcoop, Ellen M. McDonagh & Ian Dunham

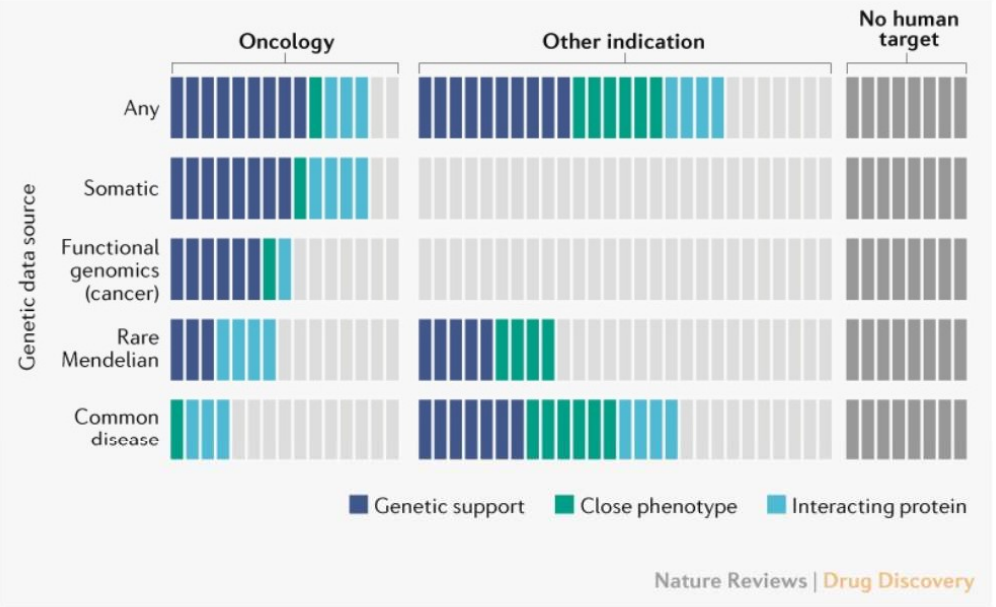


Fig. 1 | **Supporting human genetic evidence for new drugs approved by the FDA in 2021.** Availability of genetic evidence implicating each of the 50 drug targets (columns) as likely causal genes for the disease for which the drug is indicated. Genetic data sources are stratified (rows) based on the predominant nature of the genetic information. Evidence is classified based on whether it directly supports the gene–indication pair or any of the closely related phenotypes or proteins. See Supplementary information for details and an expanded figure. Data source: Open Targets Platform (November 2021).

Outline

1. Big Data & Genomic Big Data
2. Genomic Data
3. The sharing dilemma
4. What is the EGA for?
5. Beacon
6. The EGA Federation

But… what did actually happen in medical genomics?
(without many people noticing)

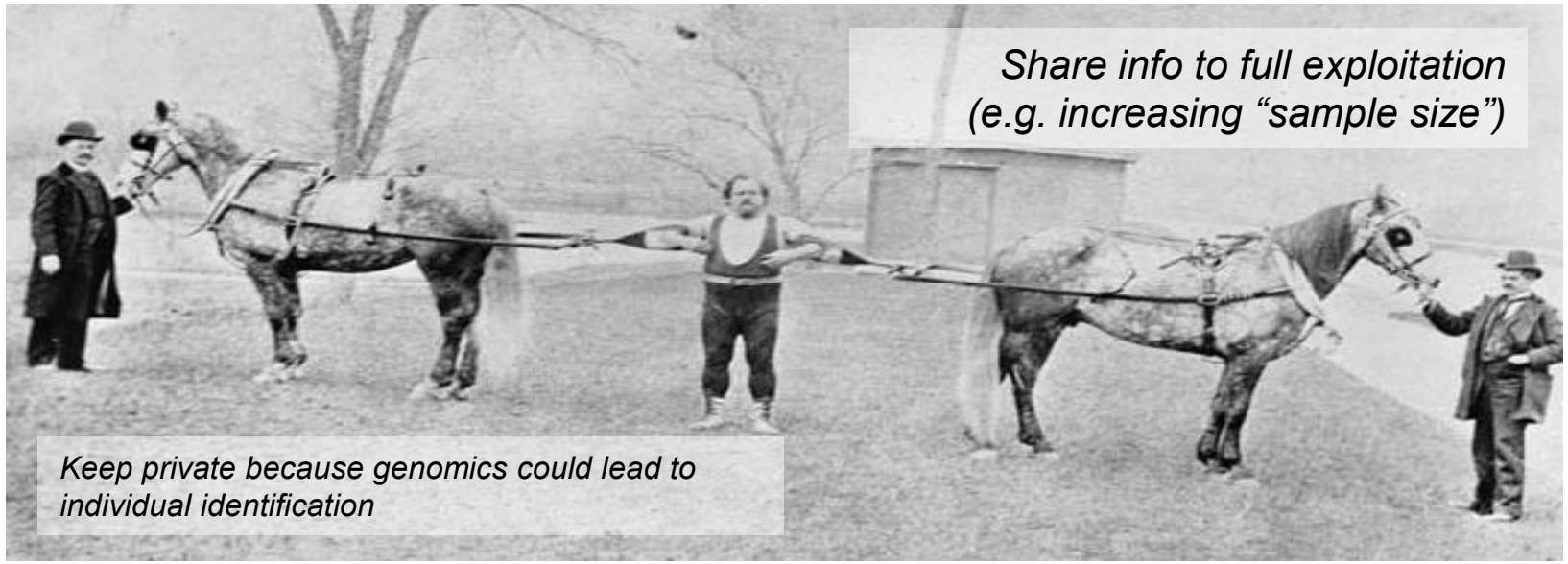
"We will only share your data for research purposes."

The power of a simple sentence… in a consent form

What is needed for you to guarantee that?

the “Sharing dilemma”

A human rights problem* with genomic information

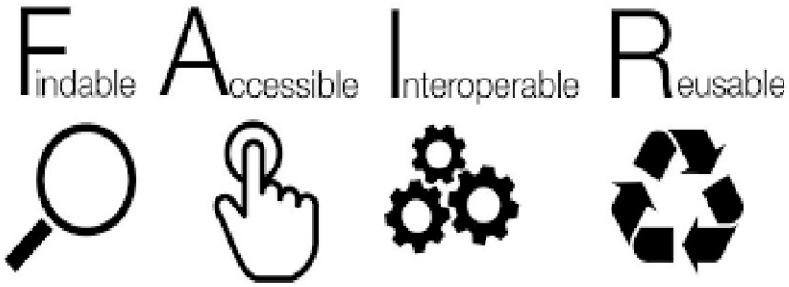


- Private Identifying Information – need to be kept confidential
- Generated with taxpayer’s money – need to be made public

*Arts. 12 and 27 of the Universal Declaration of Human rights

my own position in the dilemma:

Scientific data must be FAIR



EGA European Genome-Phenom

ega-archive.org

EUROPEAN GENOME-PHENOME ARCHIVE

Login | Register | Need Help?

ABOUT | DISCOVERY | SUBMISSION | ACCESS

Search...

The European Genome-phenome Archive (EGA) is a service for permanent archiving and sharing of personally identifiable genetic, phenotypic, and clinical data generated for the purposes of biomedical research projects or in the context of research-focused healthcare systems.

I want to...

DOWNLOAD DATA

DEPOSIT DATA

MANAGE DATA

Events

16-17 NOVEMBER 2023 | ELIXIR COMMUNITIES | BARCELONA, SPAIN / VIRTUAL

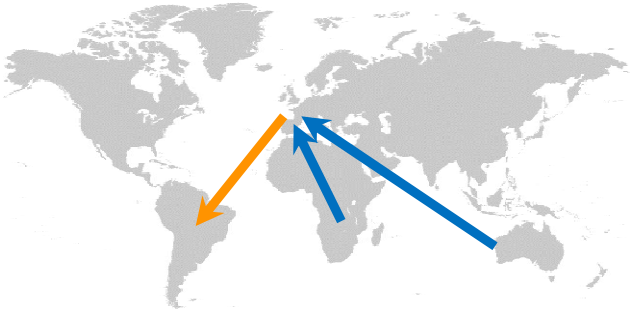
ELIXIR Federated Human Data Community Day

Latest Blog Posts

ENG 17:16 11/11/2023

Data access models

Fully Open research data

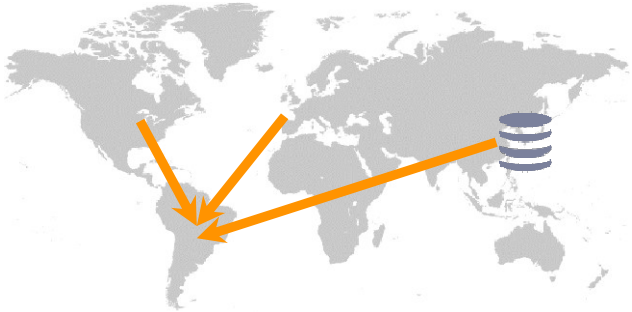


Aggregate data globally

Download, analyse locally

e.g., Plant and animal genomes, Population-based results, Some GWAS summary statistics

Managed access data



Different **controlled access** silos

Download, analyse locally

e.g., UK BioBank, dbGaP/EGA, GWAS consortia

Outline

1. Big Data & Genomic Big Data
2. Genomic Data
3. The sharing dilemma
4. What is the EGA for?
5. Beacon
6. The EGA Federation

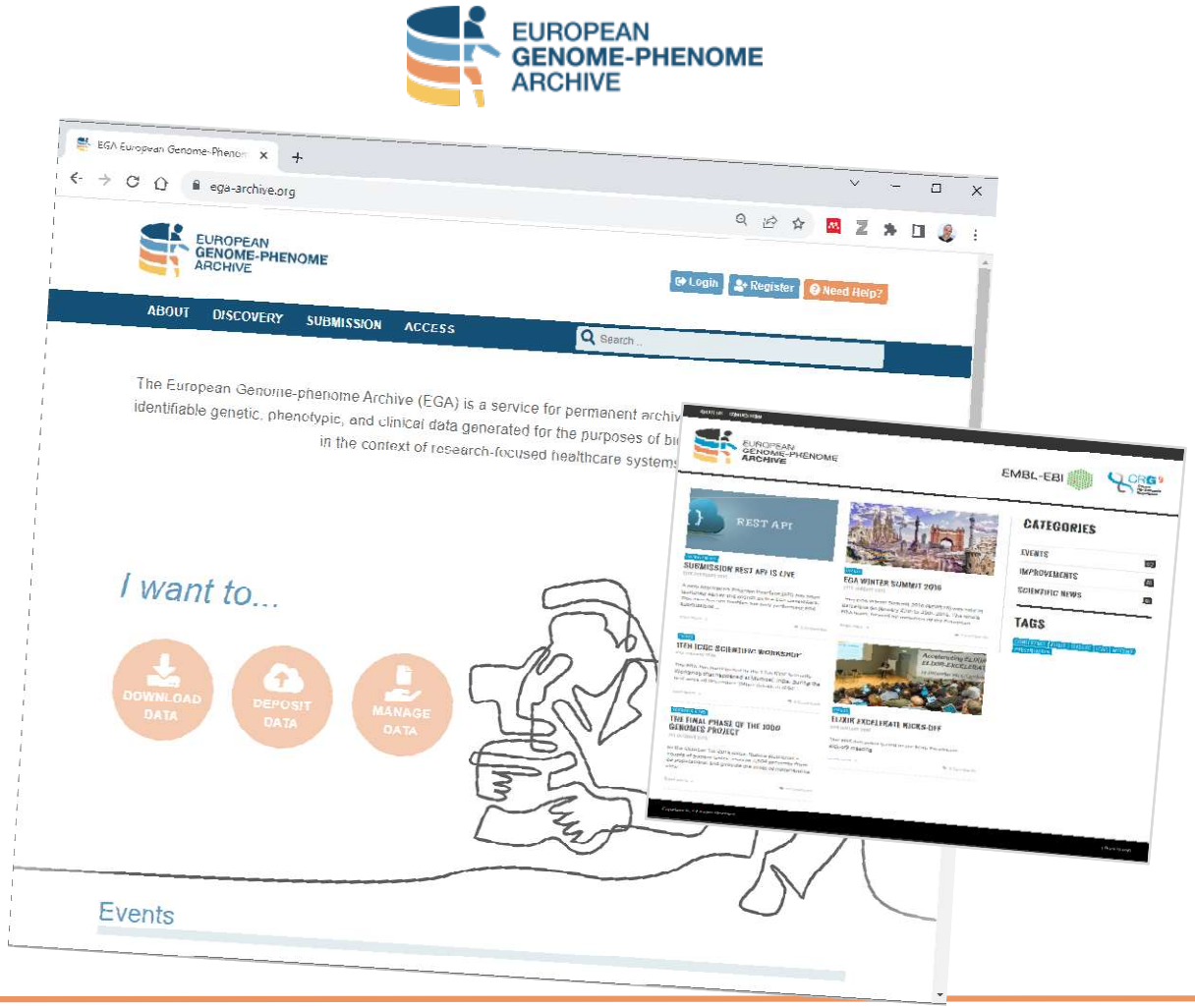
The EU Solution: the European Genome-phenome Archive (EGA)

The EGA is an Archive for human genomic data that should have controlled access because it allows identifying individuals

- ✓ Data is provided by research centers and health care institutions
- ✓ Access is controlled by Data Access Committees
- ✓ Data requesters are researchers from other research or health care institutions

Our goal is to foster responsible data reuse for the good of science and all humankind

https://ega-archive.org



What is the EGA for?

- Science needs data:
  - Data you collect,
  - … that journals & funders will ask you to share for reproducibility,
  - …and other scientist will leverage in re-analysis and meta-analyses
- EGA supports Open Science for human sensitive data, in the frontier between pure and health-bound biomedical research

Who manages the EGA?

The EGA was create
2008, to host the W

CRG joined in **2013**

Project goal:

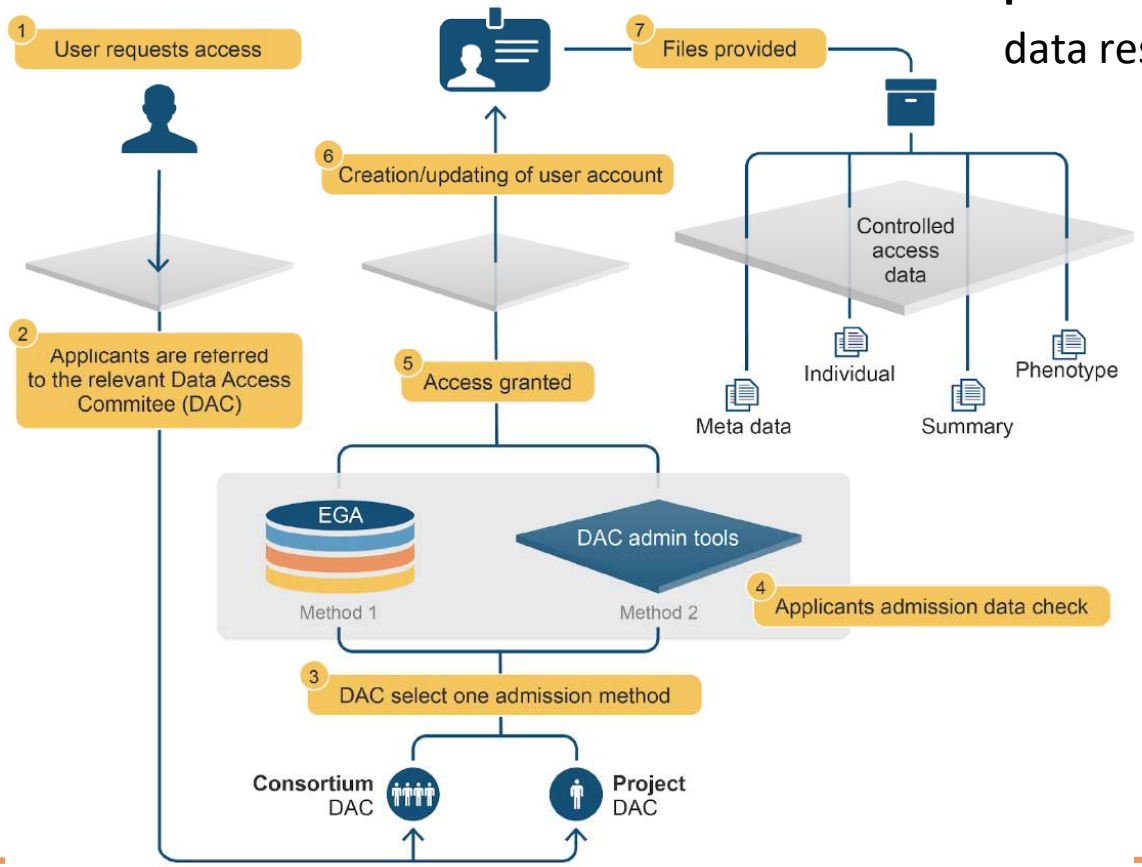
To transform the EGA to a joint

BL-EBI

CRG Centre for Genomic Regulation

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

EGA's differential assets



The EGA is **free worldwide** resource for **permanent secure archiving** and sharing of all types of **potentially identifiable** biomolecular and *phenoclinic* data resulting from biomedical research projects.

Data is provided by research centers and health care institutions.

Control of access is kept by every study Data Access Committee (*data controller*).

Data requesters are researchers from other research or health care institutions.

The EGA contains a huge variety of data

Studies ~7,600

Datasets ~15,000

DACs ~3,500

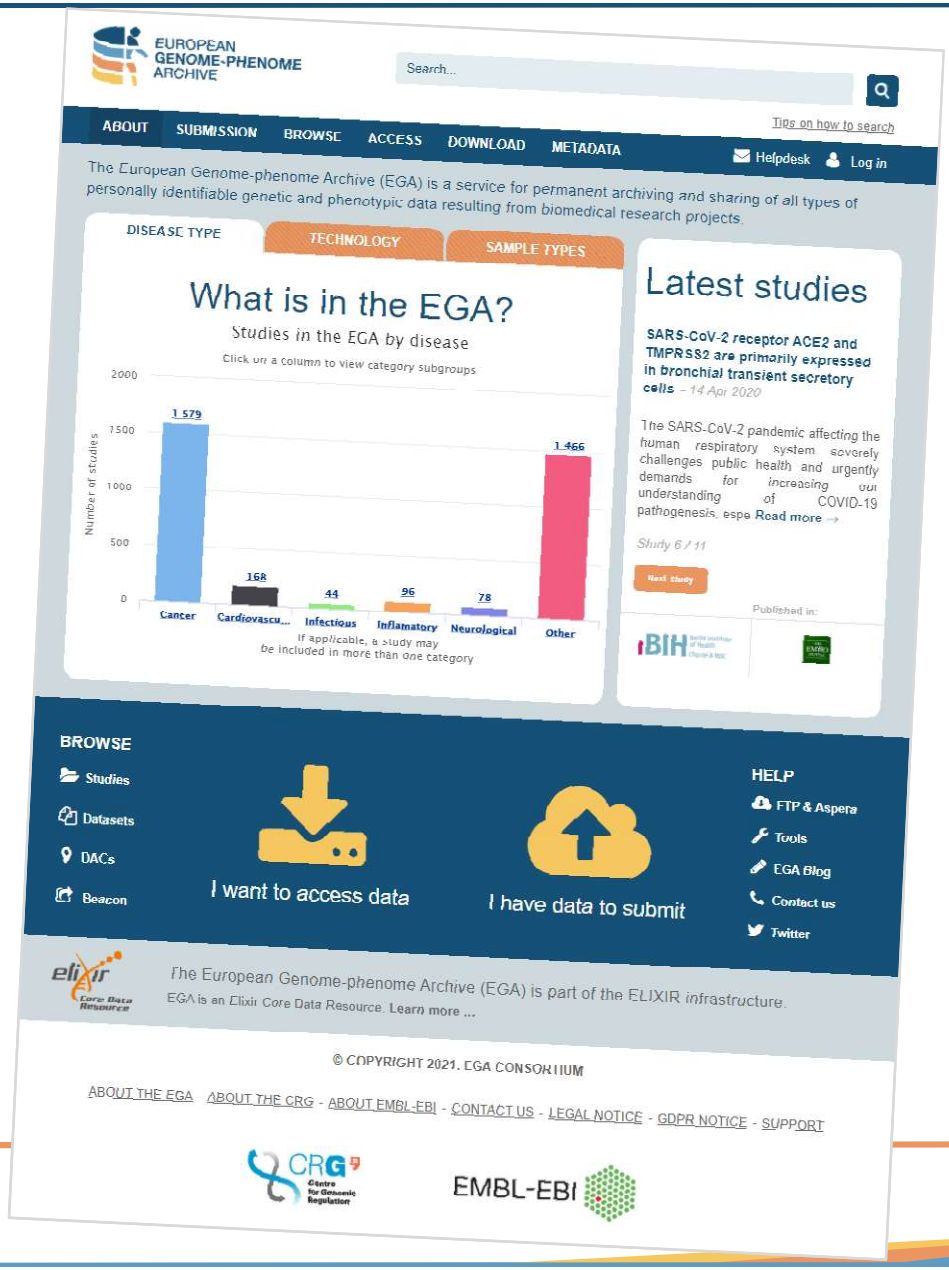
Submitters ~2,250

Requesters ~32,500

Files ~3.5 Million

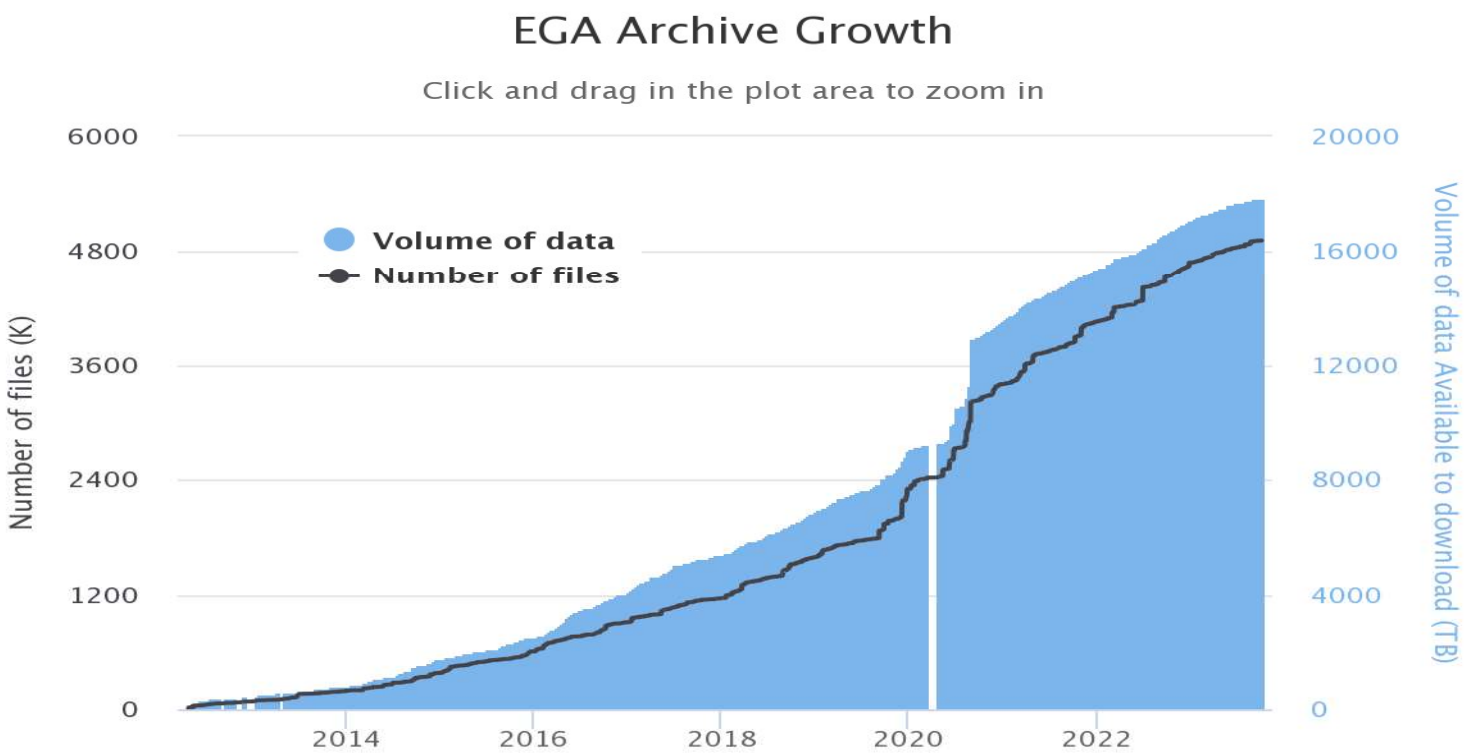
Archive size ~17,4 Petabytes

Updated Nov 11th



https://ega-archive.org

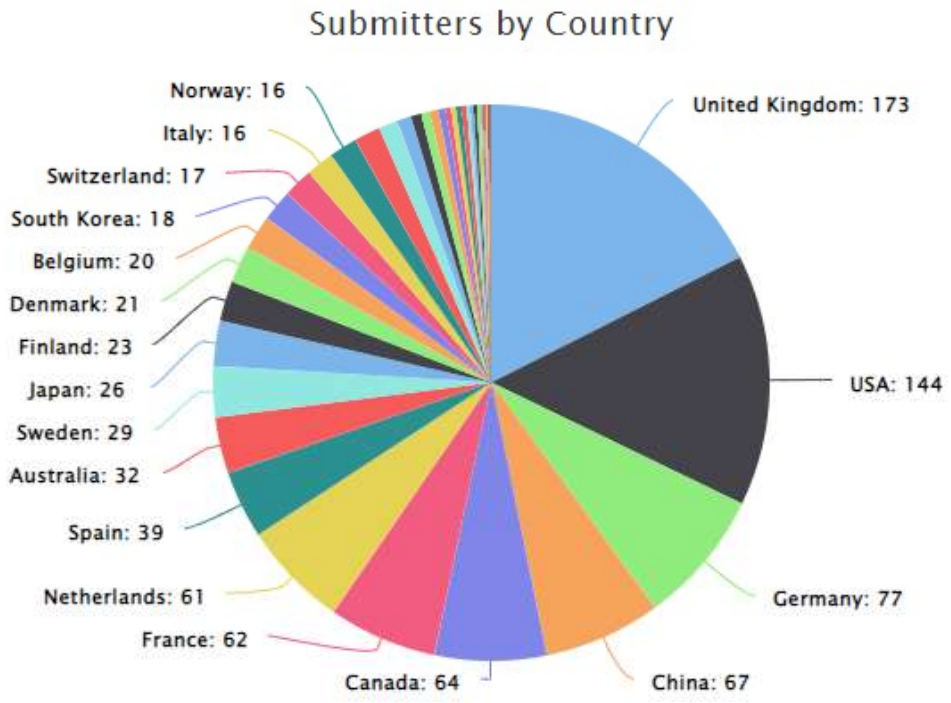
EGA's Growth



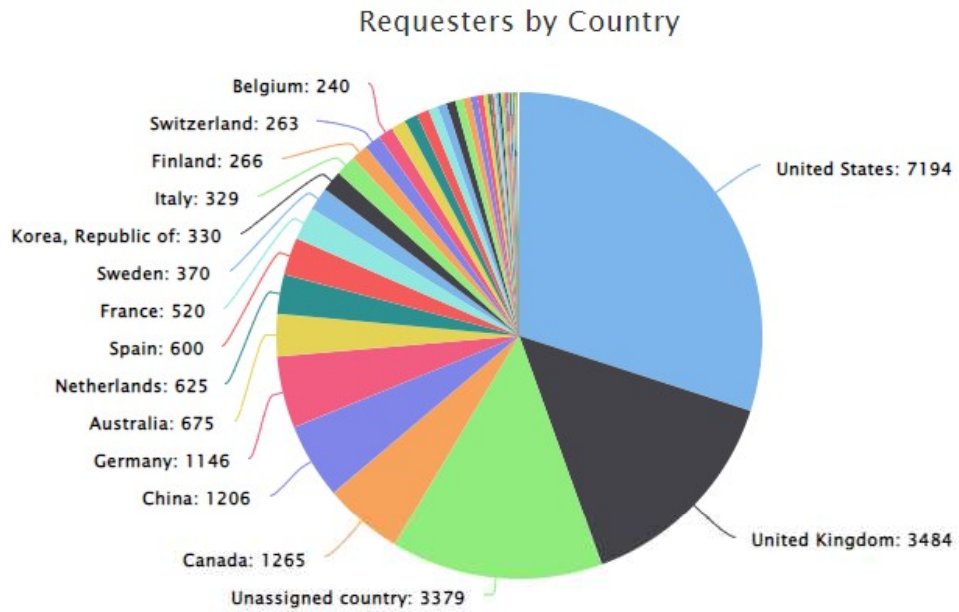
EUROPEAN GENOME-PHENOME ARCHIVE

Who is using the EGA?

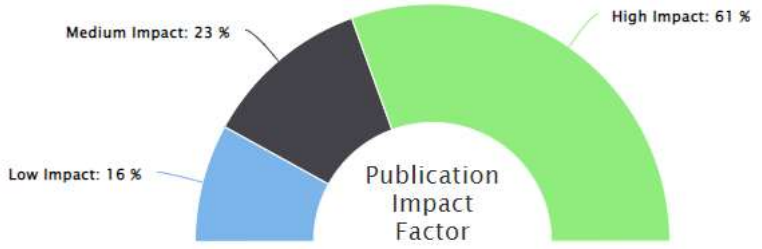
Submitter Accounts Created by Country



Requester Accounts Created by Country



Journal Impact Factor of Publications Citing Data Stored in the EGA



Outline

1. Big Data & Genomic Big Data
2. Genomic Data
3. The sharing dilemma
4. What is the EGA for?
5. Beacon
6. The EGA Federation

is all this enough?

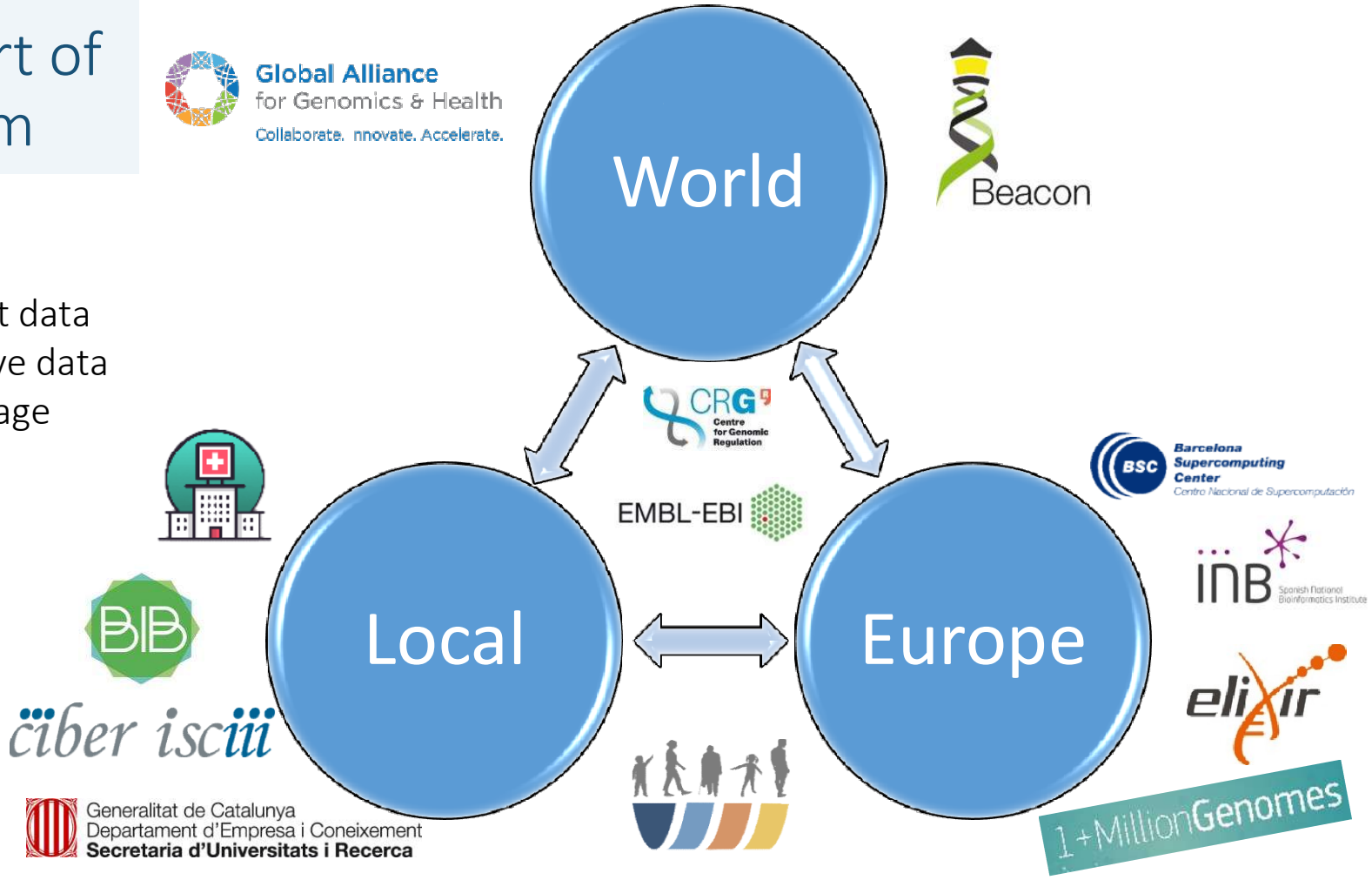
GDPR and its multifaceted interpretations, plus different jurisdictions and interests, plus practical and economic issues make the things difficult.

Fortunately…

*GDPR is the EC's General Data Protection Regulation

...the EGA is part of a rich ecosystem

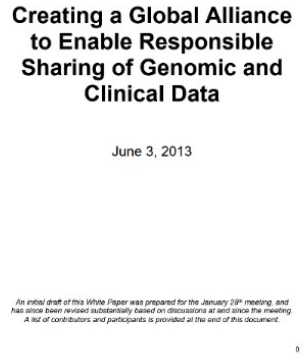
Pioneering GDPR-compliant data stewardship (comprehensive data discovery, secure data storage and legally-compliant data sharing)



The GA4GH the Global Alliance for Genomics and Health

https://ga4gh.org

Global Alliance for Genomics & Health



Creating a Global Alliance to Enable Responsible Sharing of Genomic and Clinical Data

June 3, 2013

An initial draft of this White Paper was prepared for the January 28th meeting, and has since been revised substantially based on discussions at and since the meeting. A list of contributors and participants is provided at the end of this document.

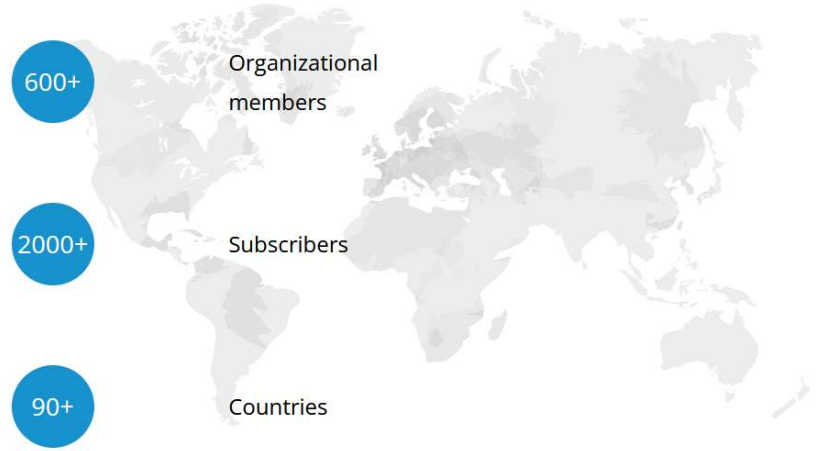
[GA4GH: The Global Standards Organization for Genomics - YouTube]

Mission

The Global Alliance for Genomics and Health 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.

Members

Experts in healthcare, research, patient advocacy, life science, and information technology.



GA4GH **Driver Projects** are real-world genomic data initiatives that provide input on the standards most needed for the international genomics community to share data. Driver Projects put finished standards and frameworks to immediate use.

GA4GH Driver Projects

National Initiatives

NIH National Heart, Lung, and Blood Institute
TOPMED
NIH National Cancer Institute
Genomics england
Australian Genomics Health Alliance
AMED
SPHN
All of Us Research Program

Pan-African

H3Africa

Pan-European

European Genome-Phenome Archive
ENA European Nucleotide Archive
elixir

International

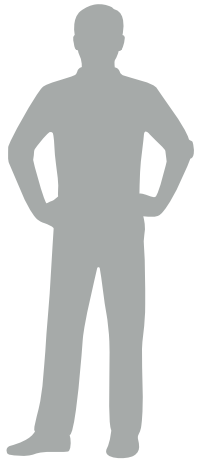
BRCA Exchange
Human Cell Atlas
Matchmaker Exchange
ClinGen Clinical Genome Resource
monarch initiative
ARGO
Autism Speaks
EU Can Can.
epi share
VICC Variant Interpretation for Cancer Consortium

lighting beacons

we have made an enormous effort to distinguish discovery from sharing

Beacon

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



"Do you have a 'C' at chromosome 13 at position 32,936,7 ?"

"Yes" (or "no")

chr13 32,936,732

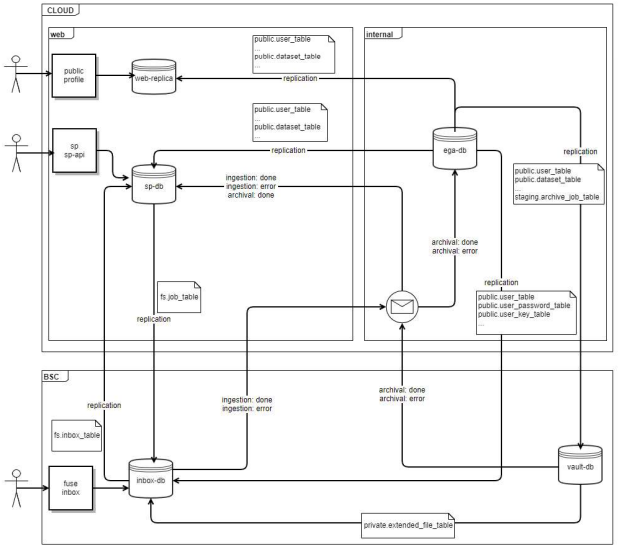
C

Geneticist

Data Holder

lighting beacons

Beacon allows for increasingly sophisticated queries



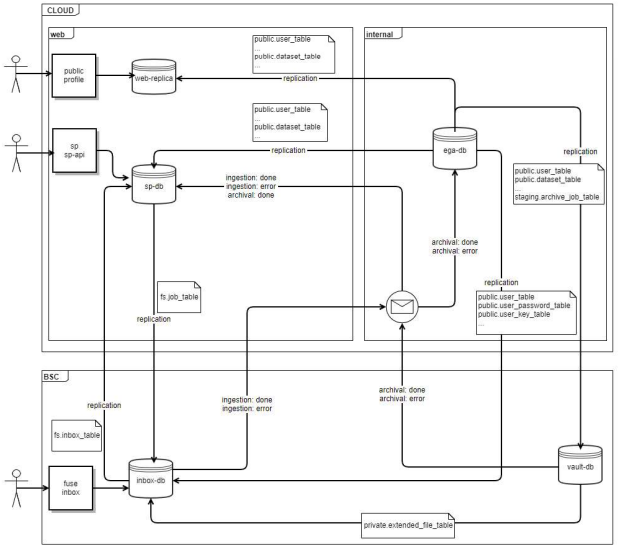
Have you seen deletions in this region on chromosome 9 in Glioblastomas from a juvenile patient, in a dataset with unrestricted access?

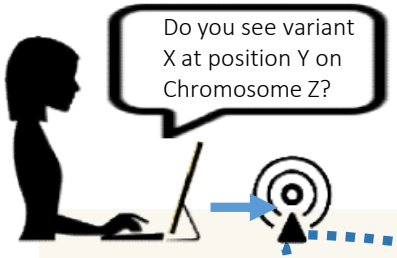
Beacon *v2* API

The Beacon API v2 proposal opens the way for the design of a simple but powerful "**genomics API**".

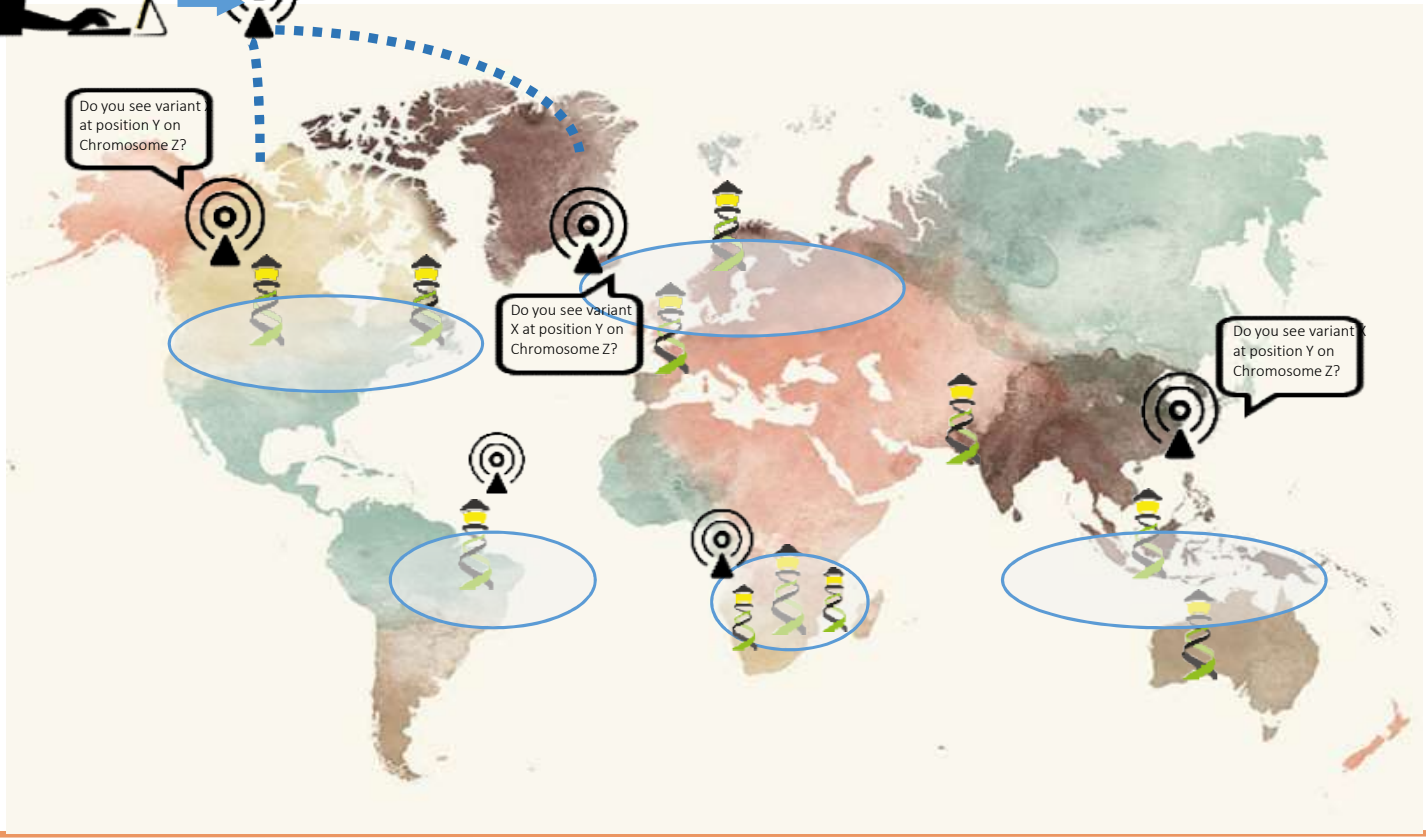
lighting beacons

we have made an enormous effort to distinguish discovery from sharing



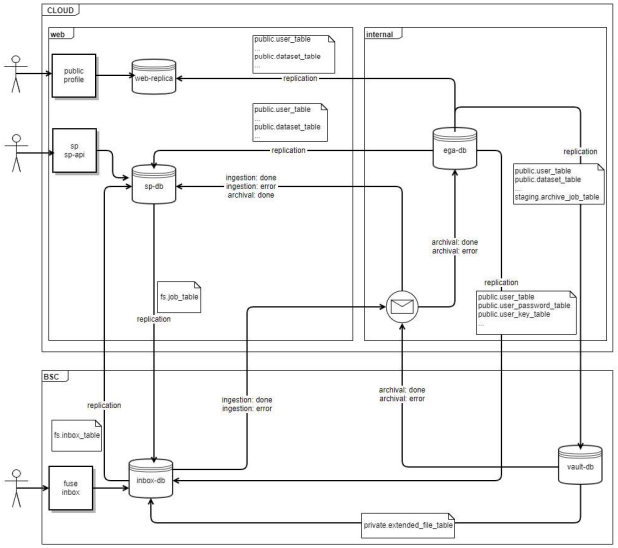


A growing number of Beacons

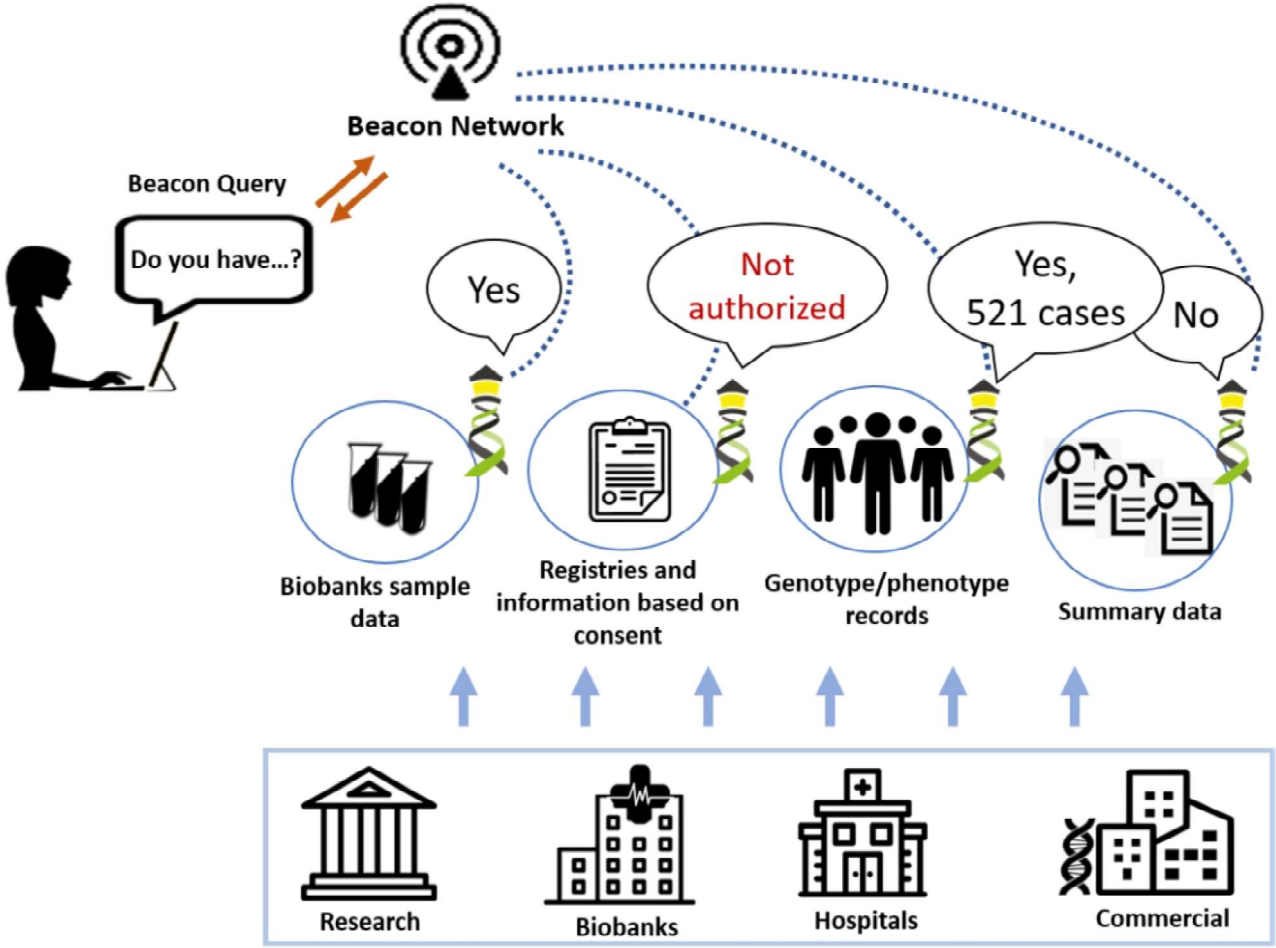


lighting beacons

And the Beacon Network allows for simple entry points

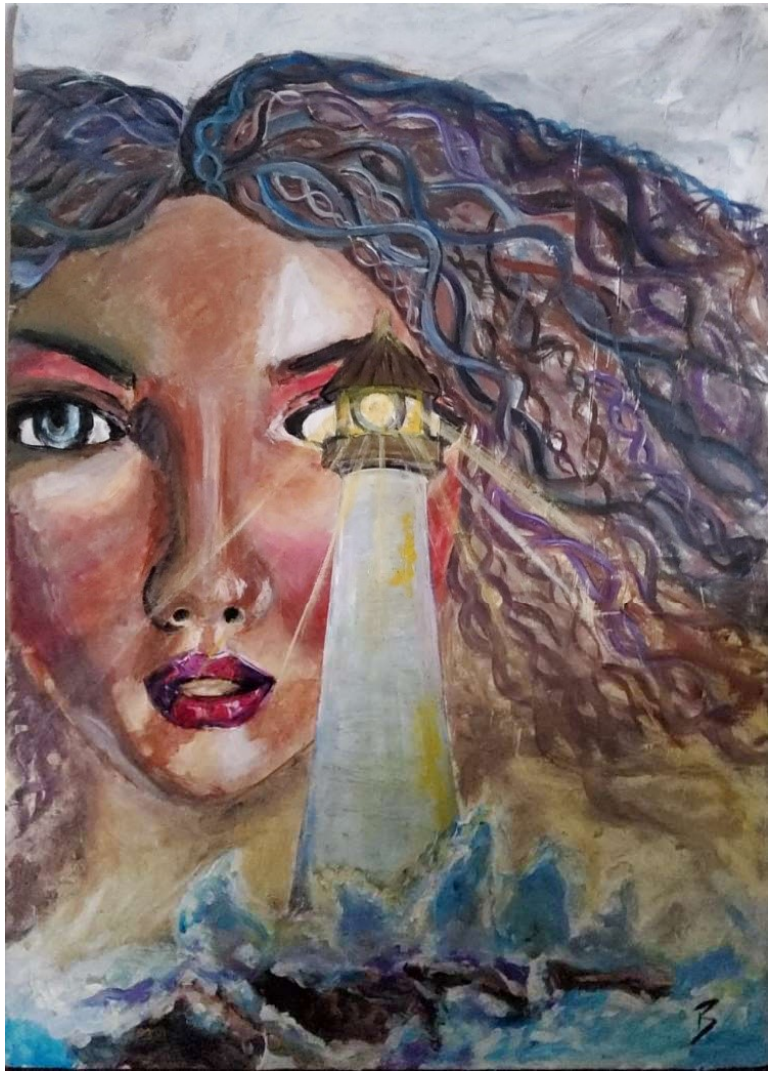


Beacon Network: Aggregator of Beacons



lighting beacons

And so, Beacon is becoming more are more useful, even beyond genomics



Towards the Internet of Genomics

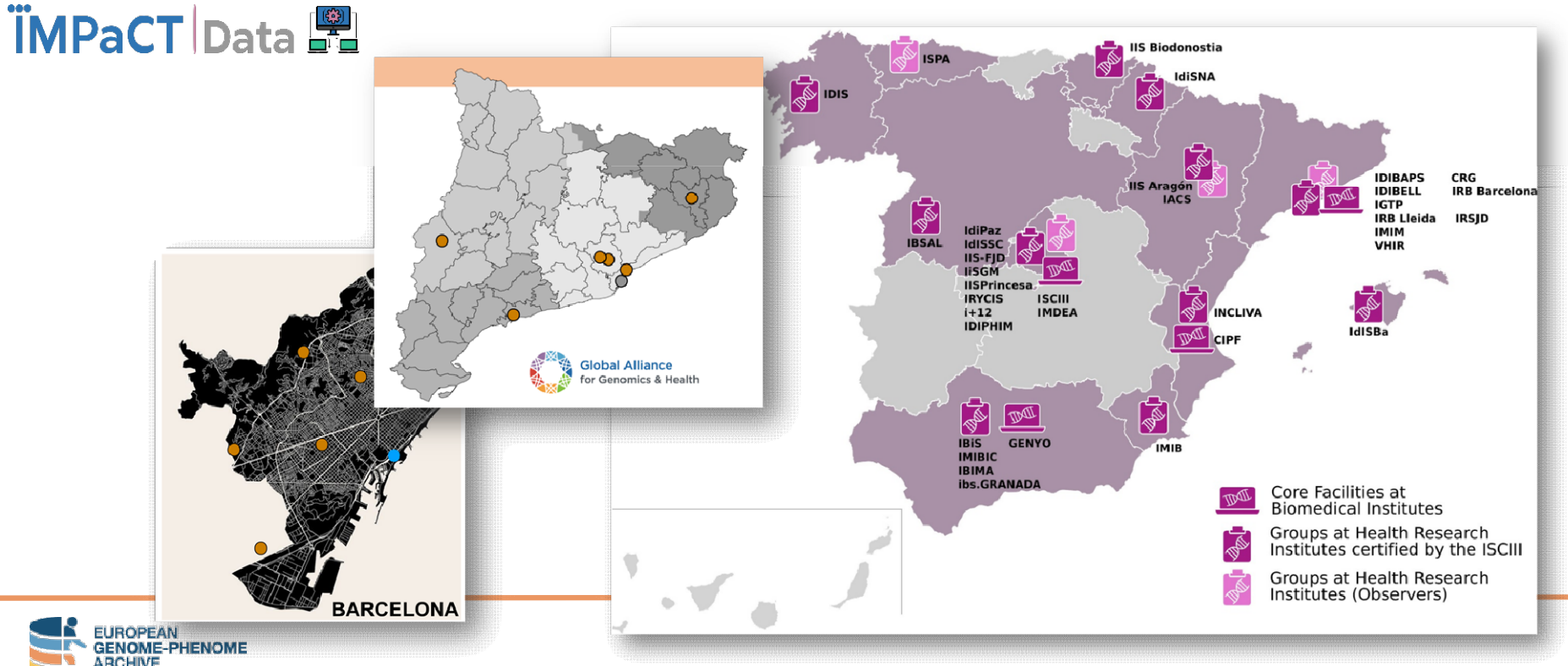
It starts with a local Beacon network…

Towards the Internet of Genomics

It starts with a local Beacon network…

Towards the Internet of Genomics

That may be part of a state-wide network…



is all this enough?

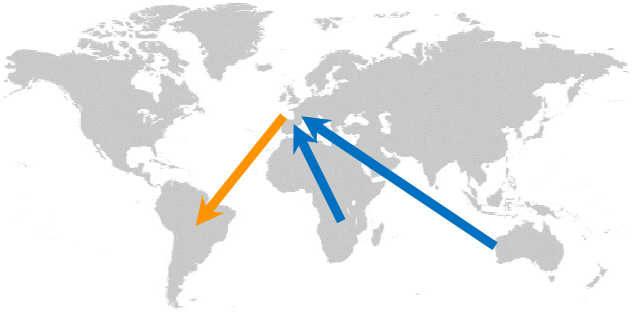
GDPR and its multifaceted* interpretations, plus different jurisdictions and interests, plus practical and economic issues make the paperwork cumbersome, plus **the Centralized Stewardship model difficult to sustain**

Oh, gosh!!

*and nefarious

Data access models

Fully Open research data

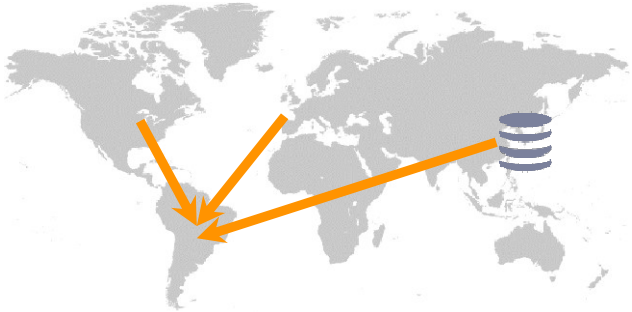


Aggregate data globally

Download, analyse locally

e.g., Plant and animal genomes, Population-based results, Some GWAS summary statistics

Managed access data



Different **controlled access** silos

Download, analyse locally

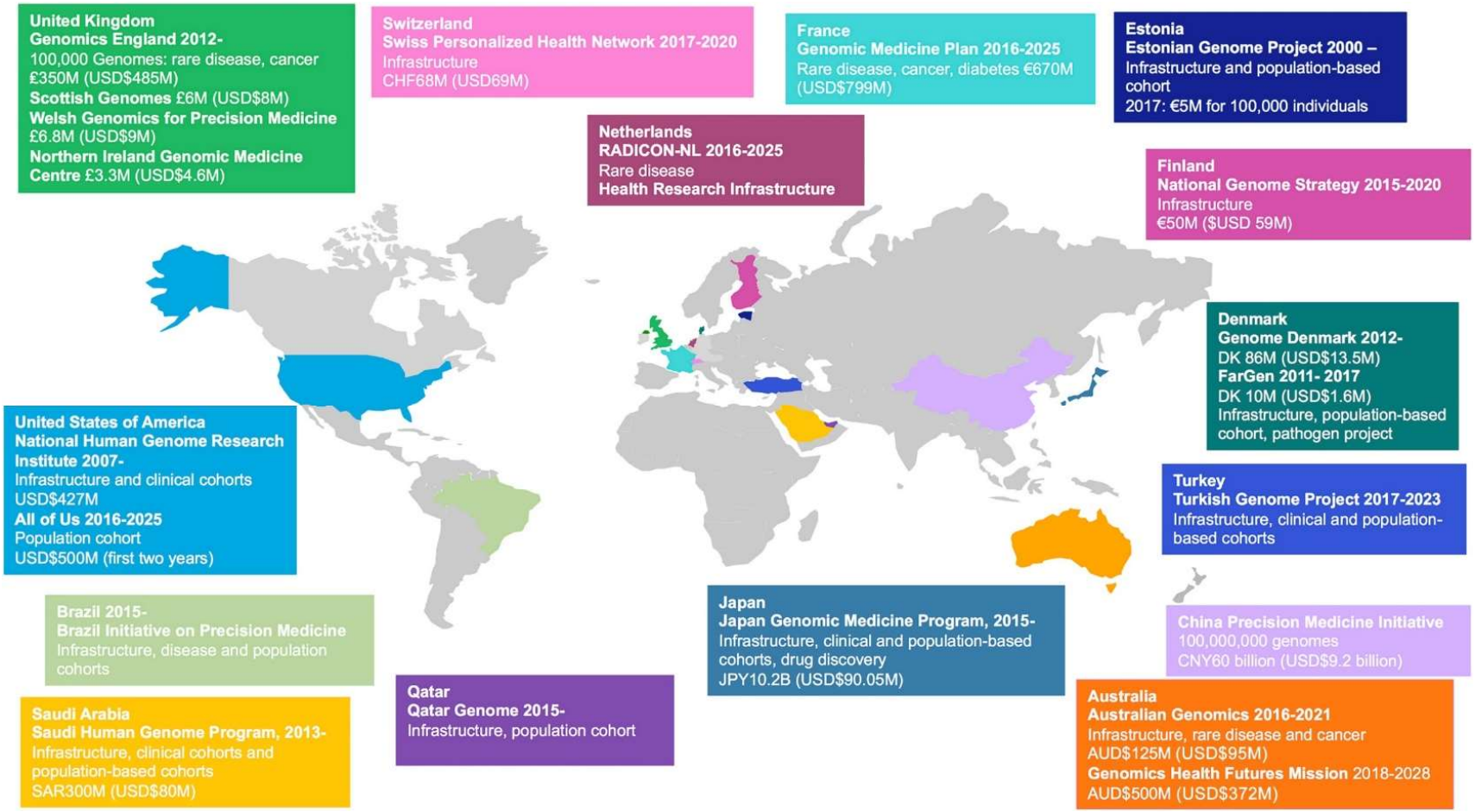
e.g., UK BioBank, dbGaP/EGA, GWAS consortia

Integrating Genomics into Healthcare: A Global Responsibility

Zornitza Stark, Lena Dolman, Teri A. Manolio, Brad Ozenberger, Sue L. Hill, Mark J. Caulfied, Yves Levy, David Glazer, Julia Wilson, Mark Lawler, Tiffany Boughtwood, Jeffrey Braithwaite, Peter Goodhand, Ewan Birney, Kathryn N. North

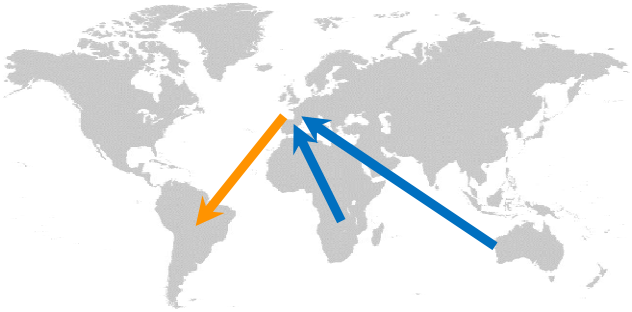
Global Alliance for Genomics & Health

The American Journal of Human Genetics 2019 10413-20DOI: (10.1016/j.ajhg.2018.11.014)



Data access models

Fully Open research data

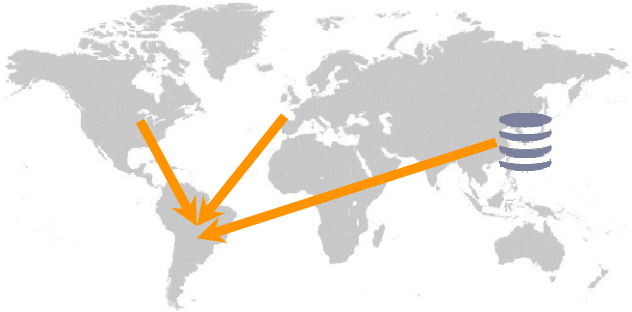


Aggregate data globally

Download, analyse locally

e.g., Plant and animal genomes, GWAS results

Managed access data

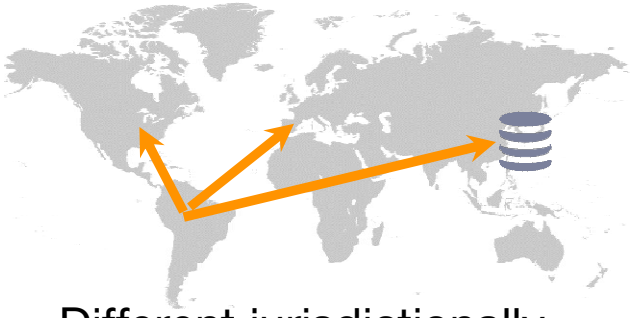


Different **controlled access** silos

Download, analyse locally

e.g., UK BioBank, dbGaP/EGA, GWAS consortia

Secondary Research use of healthcare data



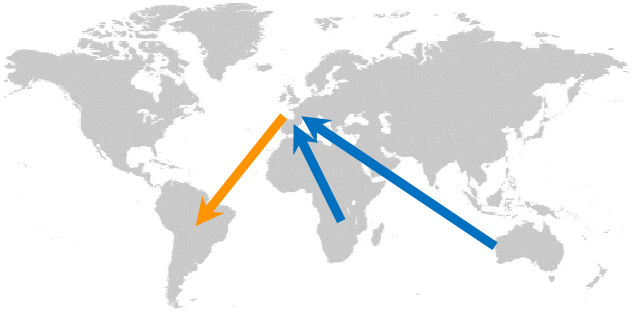
Different jurisdictionally restricted datasets

Send analysis containers into trusted research environments

e.g., Genomics England data, Danish Genome Cohort

Data access models

Fully Open research data

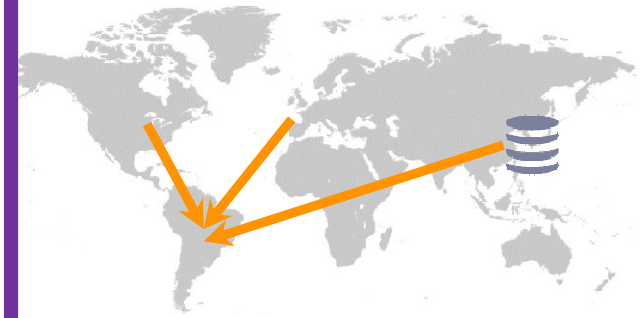


Aggregate data globally

Download, analyse locally

e.g., Plant and animal genomes, GWAS results

Managed access data

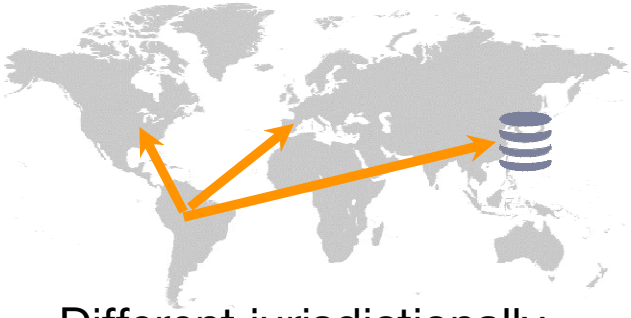


Different **controlled access** silos

Download, analyse locally

e.g., UK BioBank, dbGaP/EGA, GWAS consortia

Secondary Research use of healthcare data



Different jurisdictionally restricted datasets

Send analysis containers into trusted research environments

e.g., Genomics England data, Danish Genome Cohort

Outline

1. Big Data & Genomic Big Data
2. Genomic Data
3. The sharing dilemma
4. What is the EGA for?
5. Beacon
6. The EGA Federation

A potential solution

Federated EGA

Setting the path for actual federated sharing of sensitive human data

1+ Million Genomes

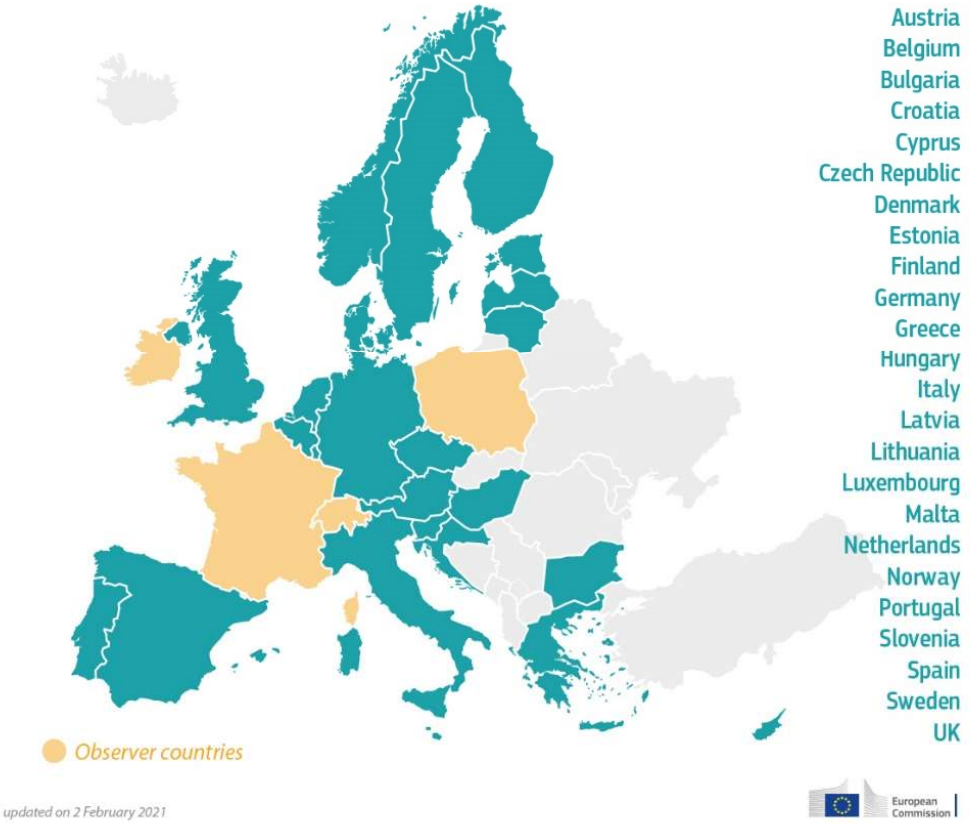
DECLARATION OF COOPERATION

Towards access to at least 1 million sequenced genomes in the European Union by 2022

Mariya Gabriel, Commissioner for Digital Economy and Society welcomed this new step forward:

> *" "EU cooperation on linking genomic health data becomes ever more important to advance personalised health care, better prevention and excellence in research. We are happy to see Latvia joining the other 18 signatory Member States and look forward to enhance this European collaboration. " "*

Countries that have signed the 1+MG Declaration since 2018



Austria
Belgium
Bulgaria
Croatia
Cyprus
Czech Republic
Denmark
Estonia
Finland
Germany
Greece
Hungary
Italy
Latvia
Lithuania
Luxembourg
Malta
Netherlands
Norway
Portugal
Slovenia
Spain
Sweden
UK

Link to initial Declaration

a complex European context

Endeavors such as the 1+Million Genomes Initiative & the Beyond 1MGenomes Project are made possible by the **Federated EGA or the Federated EGA Technology**

Mariya Gabriel, Commissioner for Digital Economy and Society welcomed this new step forward:

> "EU cooperation on linking gen becomes ever more important to personalised health care, better excellence in research. We are l Latvia joining the other 18 signa States and look forward to enha collaboration." "

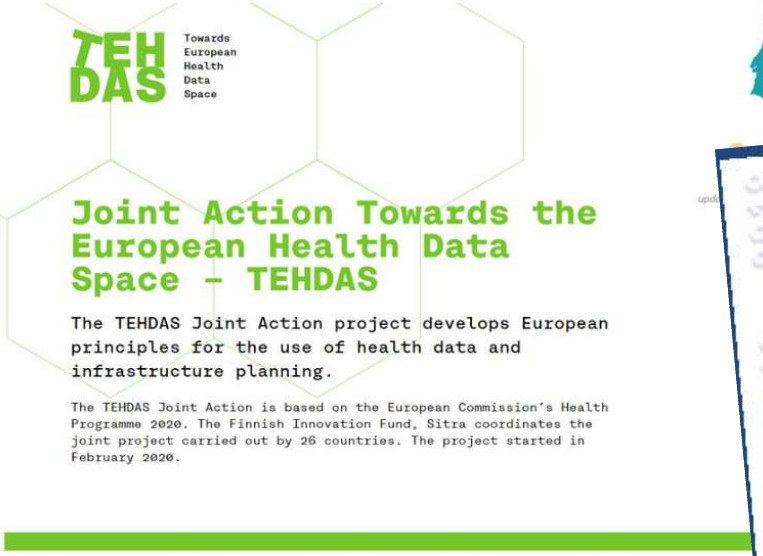

NATURE REVIEWS | GENETICS

Corrected: Author Correction

VOLUME 20 | NOVEMBER 2

ROADMAP

Leveraging European infrastructures to access 1 million human genomes by 2022


TEHDAS Towards European Health Data Space

Joint Action Towards the European Health Data Space – TEHDAS

The TEHDAS Joint Action project develops European principles for the use of health data and infrastructure planning.

The TEHDAS Joint Action is based on the European Commission's Health Programme 2020. The Finnish Innovation Fund, Sitra coordinates the joint project carried out by 26 countries. The project started in February 2020.


DECLARATION OF COOPERATION

Towards access to at least 1 million sequenced genomes in the European Union by 2022

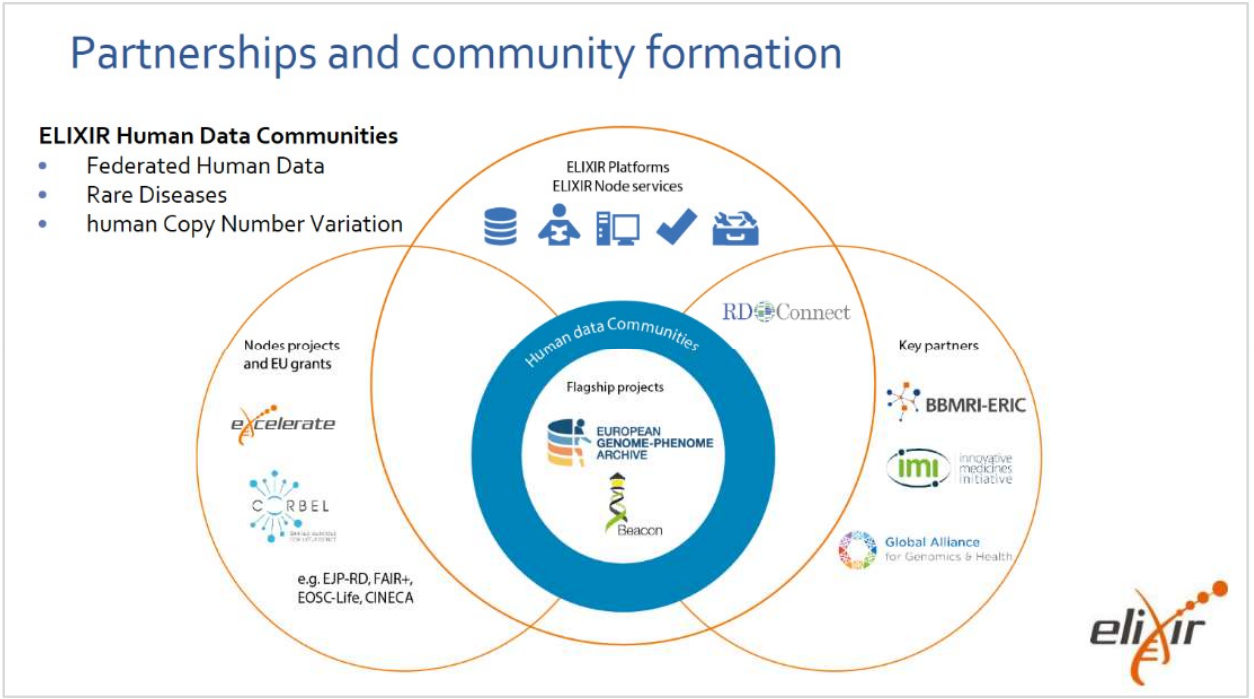
Česká republika

and

Eesti Vabariik

the European context

Endeavors such as the 1+Million Genomes Initiative & the Beyond 1MGenomes Project are made possible by the **Federated EGA or the Federated EGA Technology**



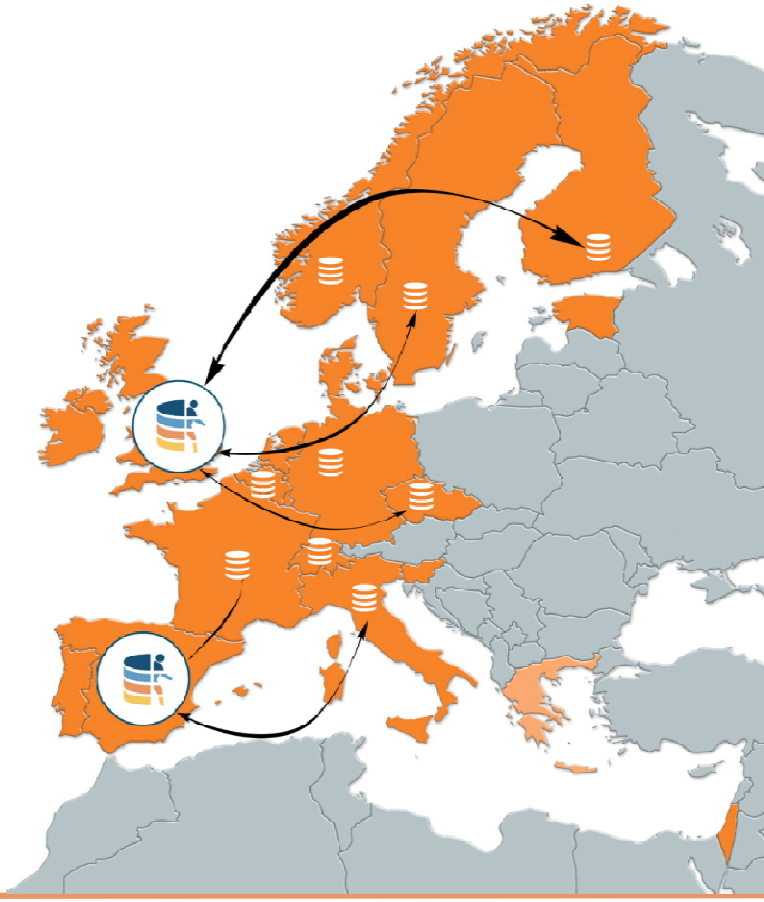
Our role in the GDI* project

*European Genome Data Infrastructure

- CEGA and FEGA nodes are **key** partners of the European Genomic Data Infrastructure initiative
- FEGA software is one possible solution to become **operational** within GDI
- Joining FEGA is a possible way to **comply** with GDI mandate

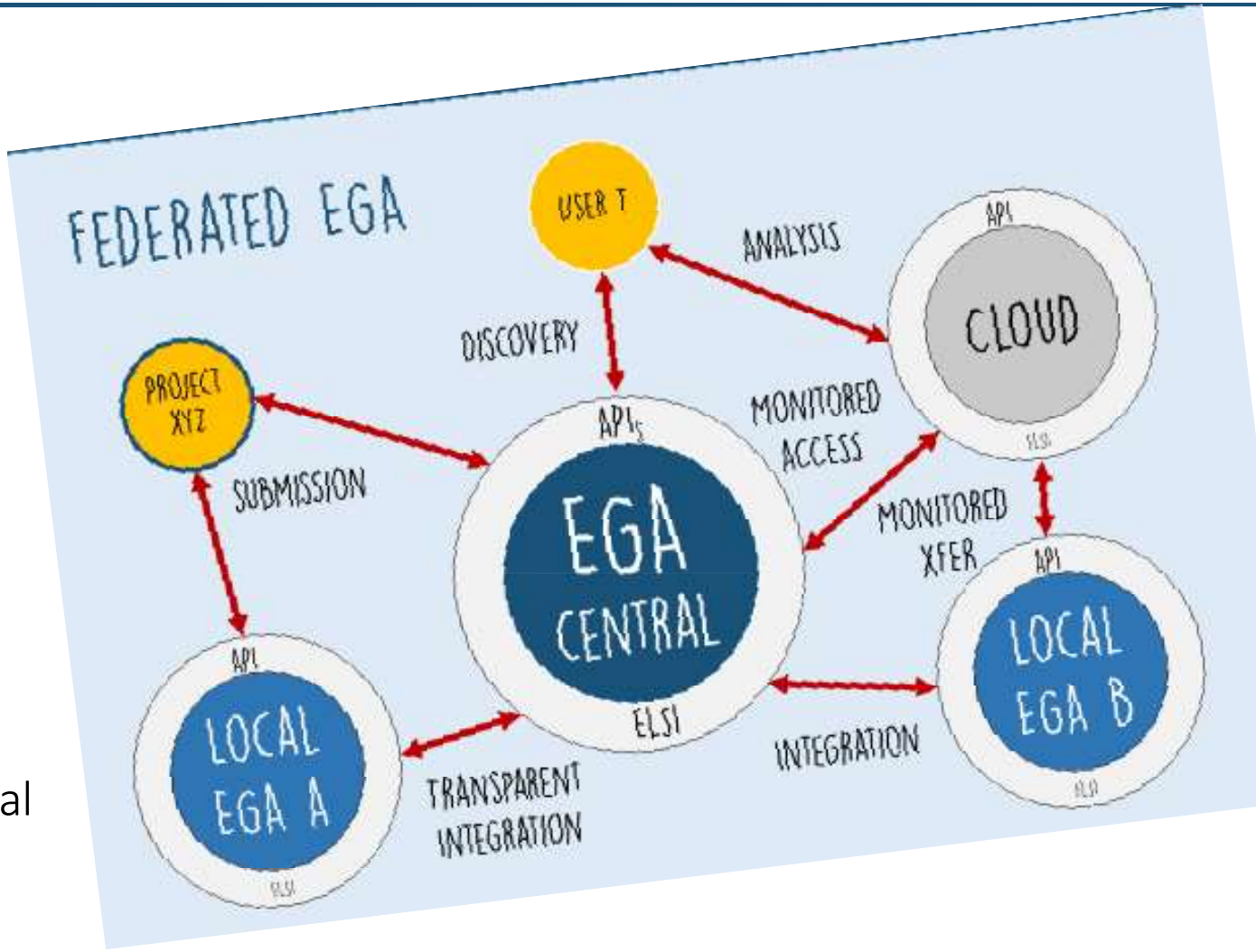
Federated EGA: The Vision

Federated EGA strives to support the discovery of and secure access to human data globally, while respecting national data protection regulations, with the goal of accelerating disease research and understanding and improving human health.

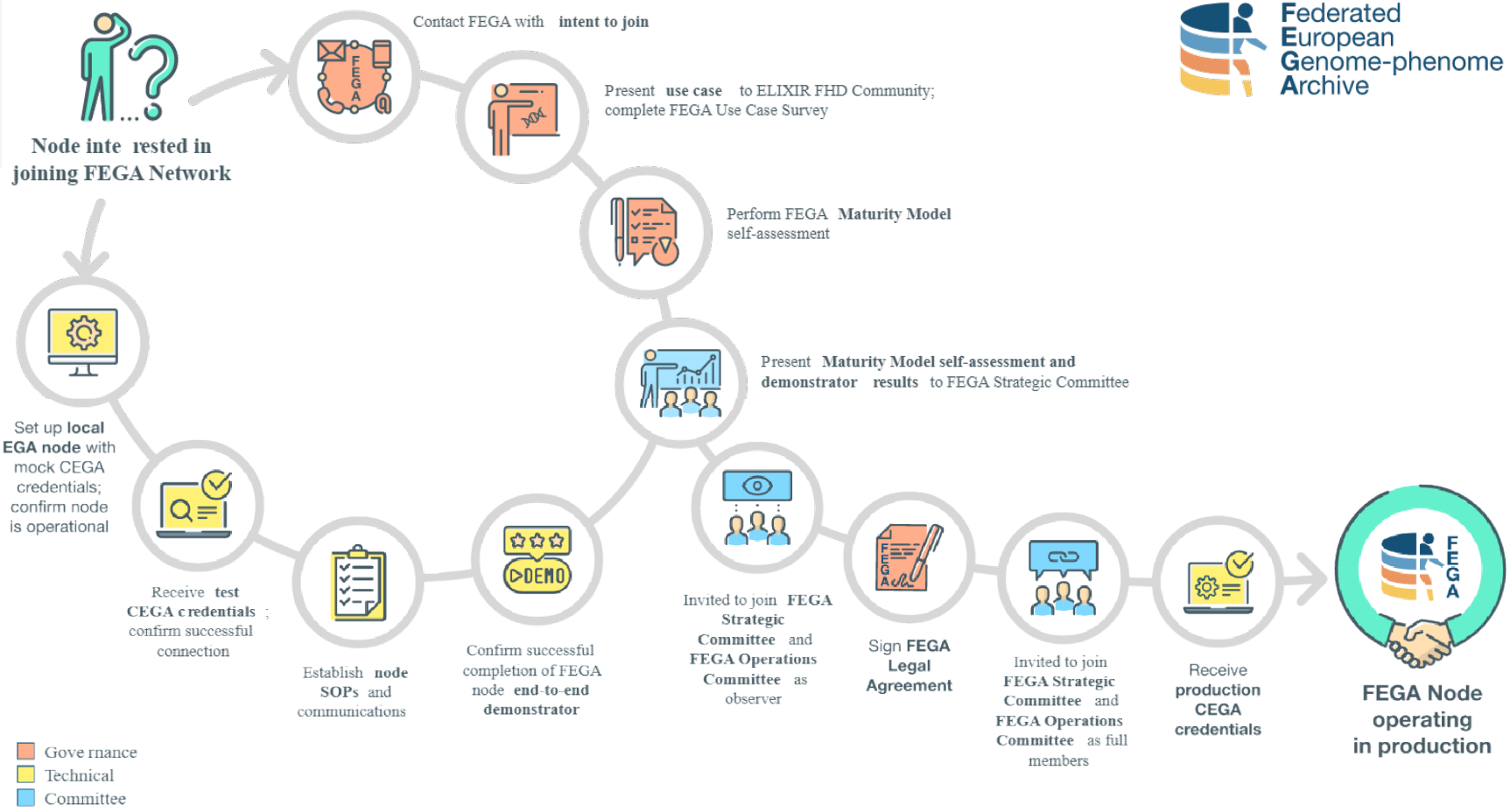


the European context

The EGA Federation is a way to achieve secure (and legal!) data discovery and sharing, including cloud features to allow for **federated analysis** (sending the algorithm to bits of data, rather than sending the data to a central computer with the algorithm)

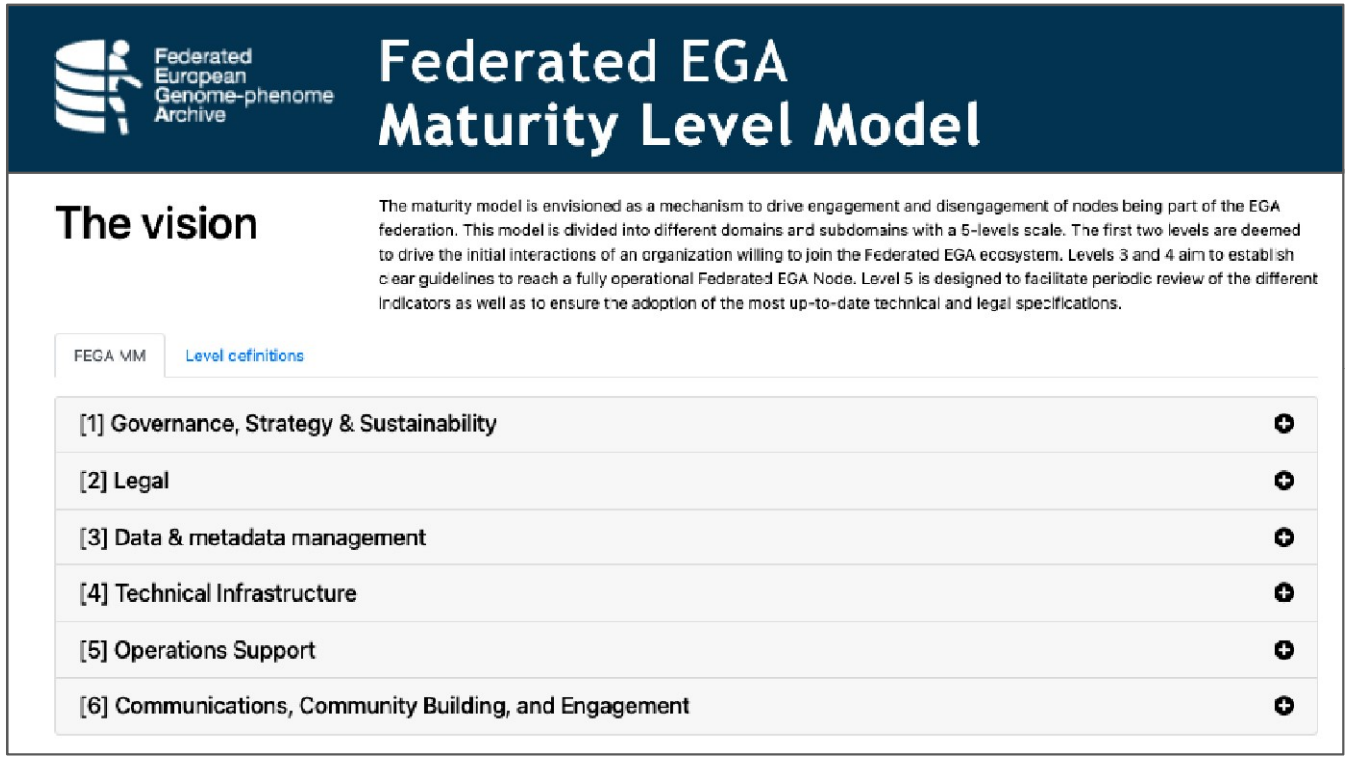


How does the journey look like?



Federated European Genome-phenome Archive

Started self-assessments of FEGA Maturity Model



Align with other Maturity Models

B1MG

HEALTHYCLOUD
Health Research & Innovation Cloud

FAIRplus

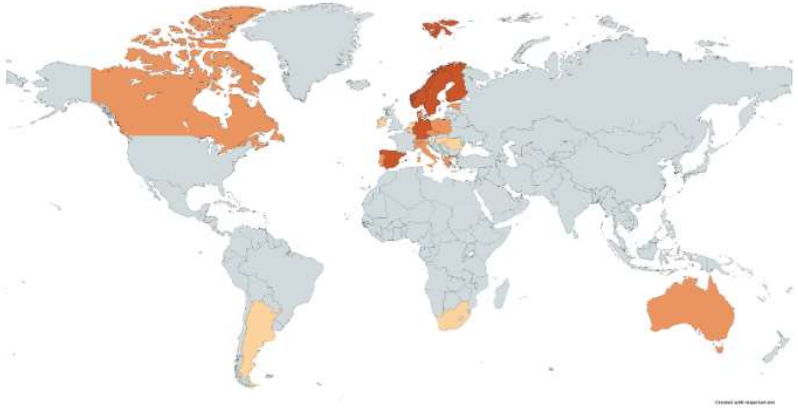
Global Alliance for Genomics & Health
Collaborate. Innovate. Accelerate.

elixir

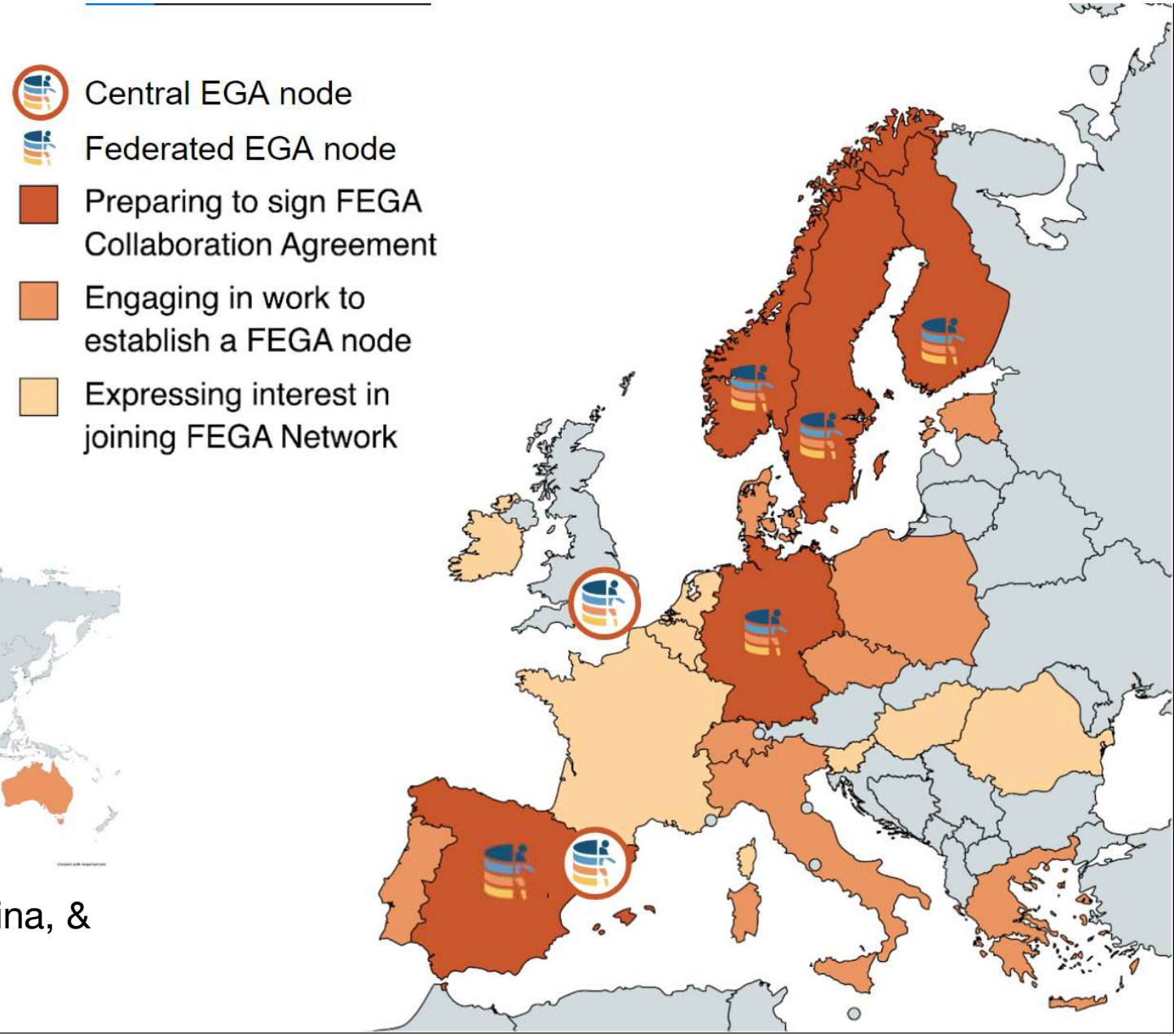
Albert Hornos, Amy Curwin, Kjell Petersen, Laura Portell-Silva, Mallory Freeberg, Melissa Konopko, Teresa D'Altri, Giselle Kerry, Salvador Capella-Gutierrez

Status as to September 2022

>20 countries / national initiatives and counting!

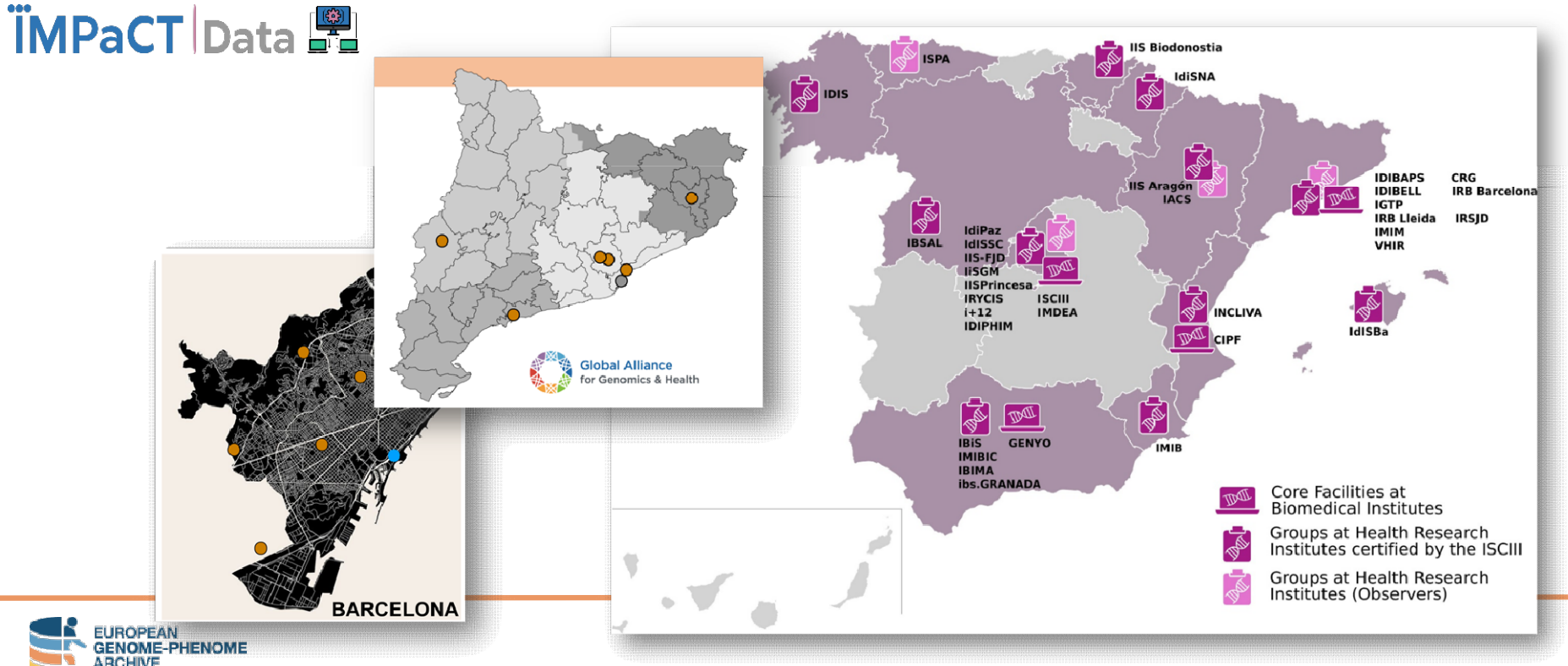


Also Canada, Australia, Argentina, & H3Africa / South Africa



Towards the Internet of Genomics

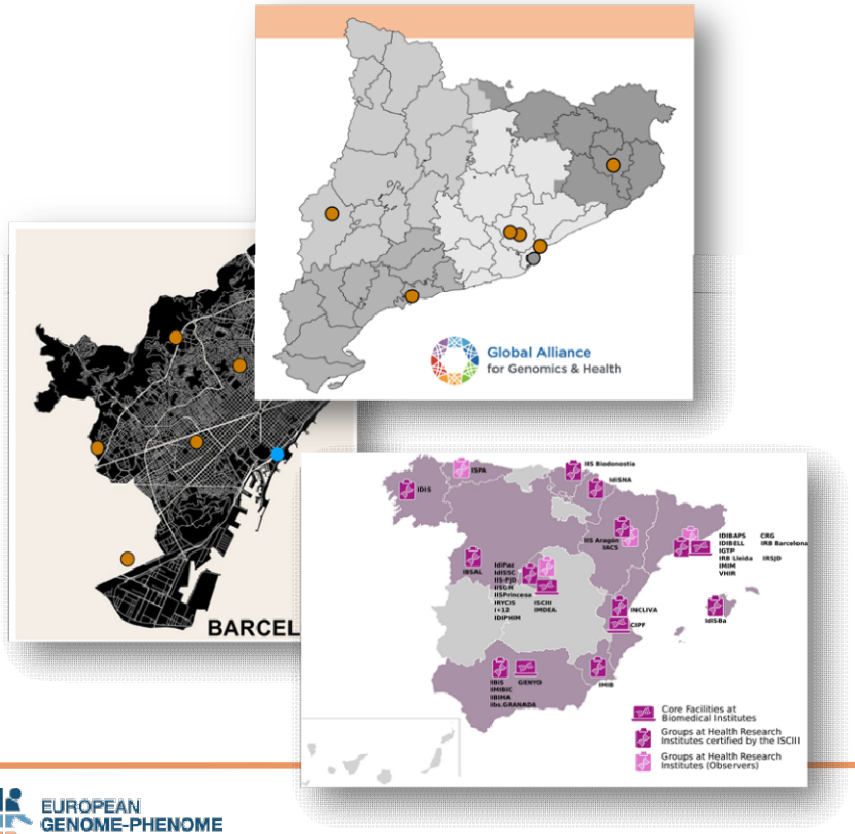
That may be part of a state-wide network…



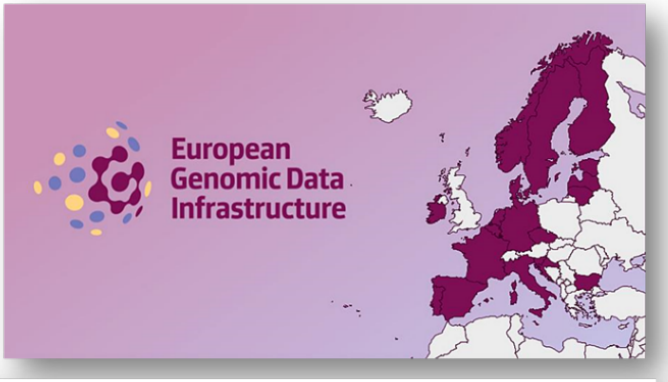
EUROPEAN GENOME-PHENOME ARCHIVE

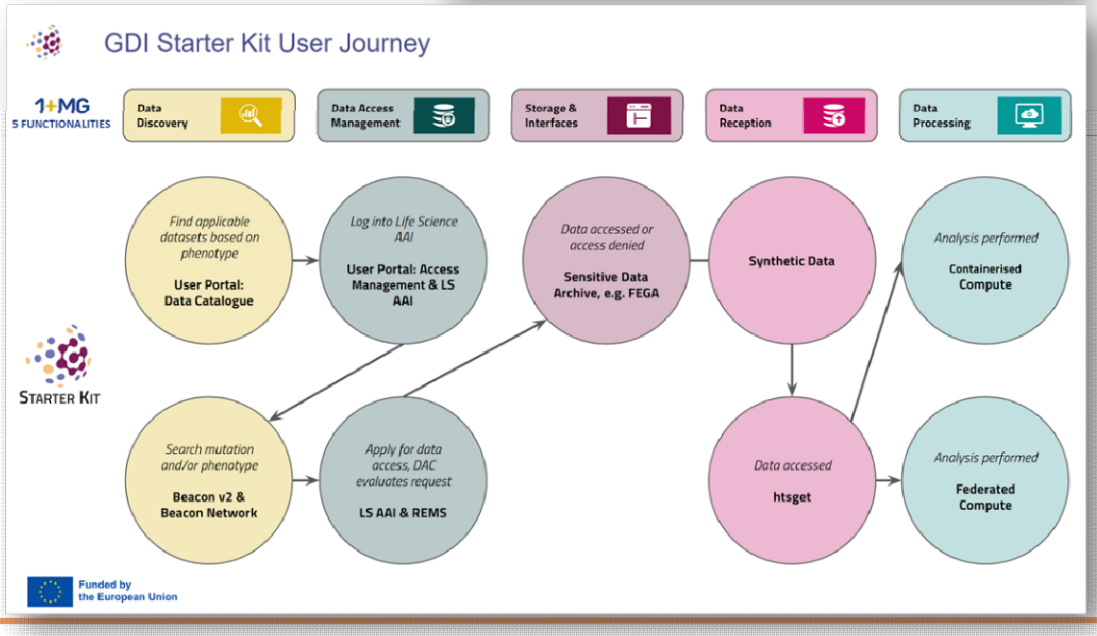
Towards the Internet of Genomics

Which is part of a European Network



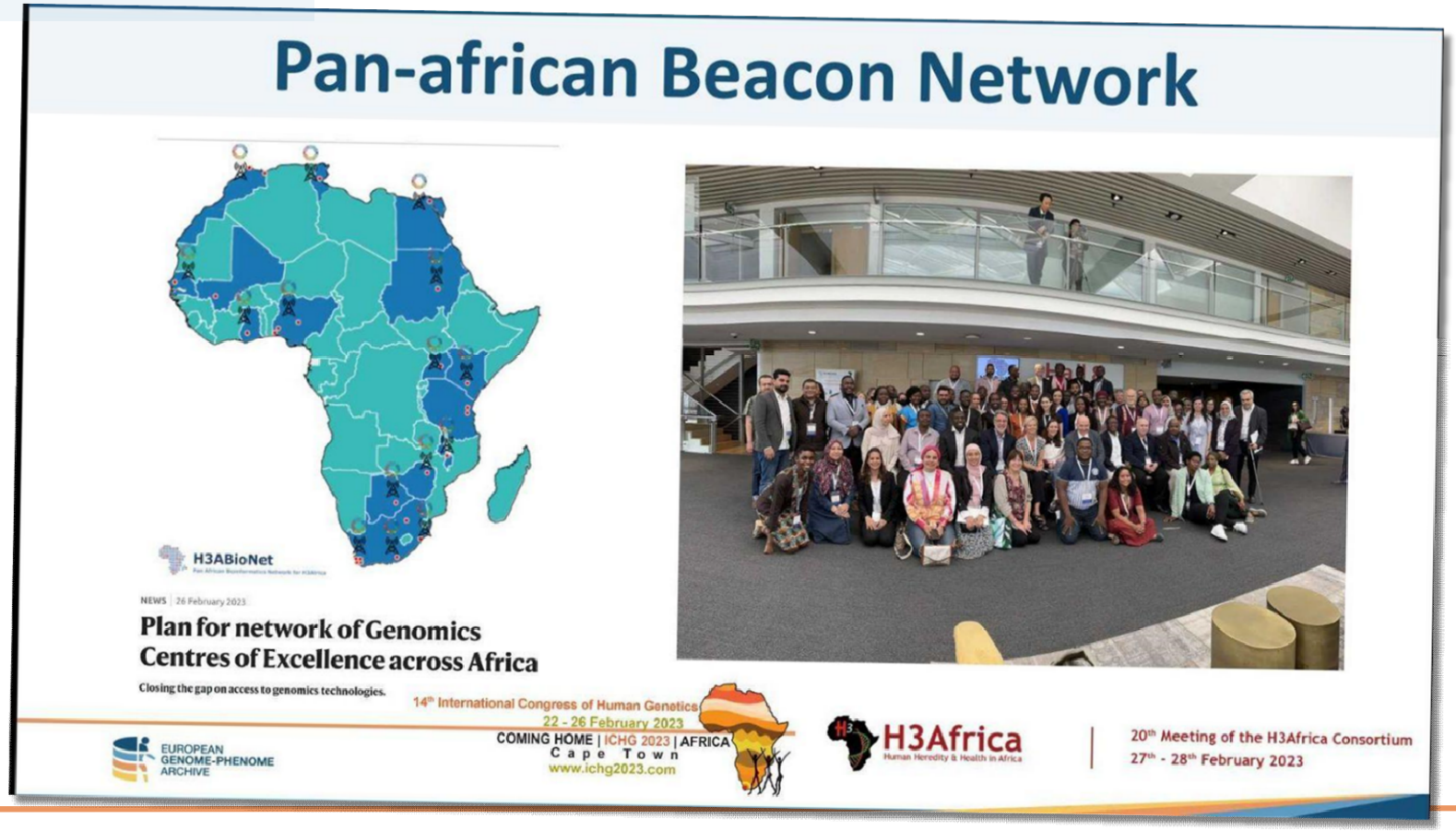
Federated European Genome-phenome Archive





Towards the Internet of Genomics

And perhaps not just European



Will we have a global network?