

European Genomic Data Infrastructure

GDI and genome data whether or not under the EHDS

Frederik Coppens

Head of VIB Data Core

Head of Node ELIXIR Belgium



May 8th, 2023

Impact of the EHDS regulation on secondary use of health data

VIB Data Core



To offer an answer to **data-related infrastructural challenges** associated with **data-driven life sciences**

Open & FAIR data
Implementation

Quality
Standardisation
Reproducibility
Efficiency

Across the data
lifecycle



Infrastructure
Applications
& Services

non-sensitive &
sensitive data

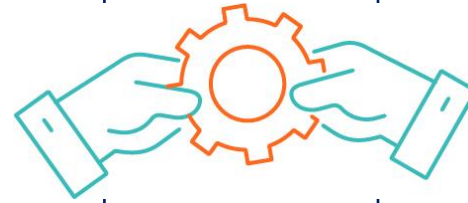
VIB Technologies



Tech Watch



Innovation Lab



Core Facilities

ELIXIR Europe



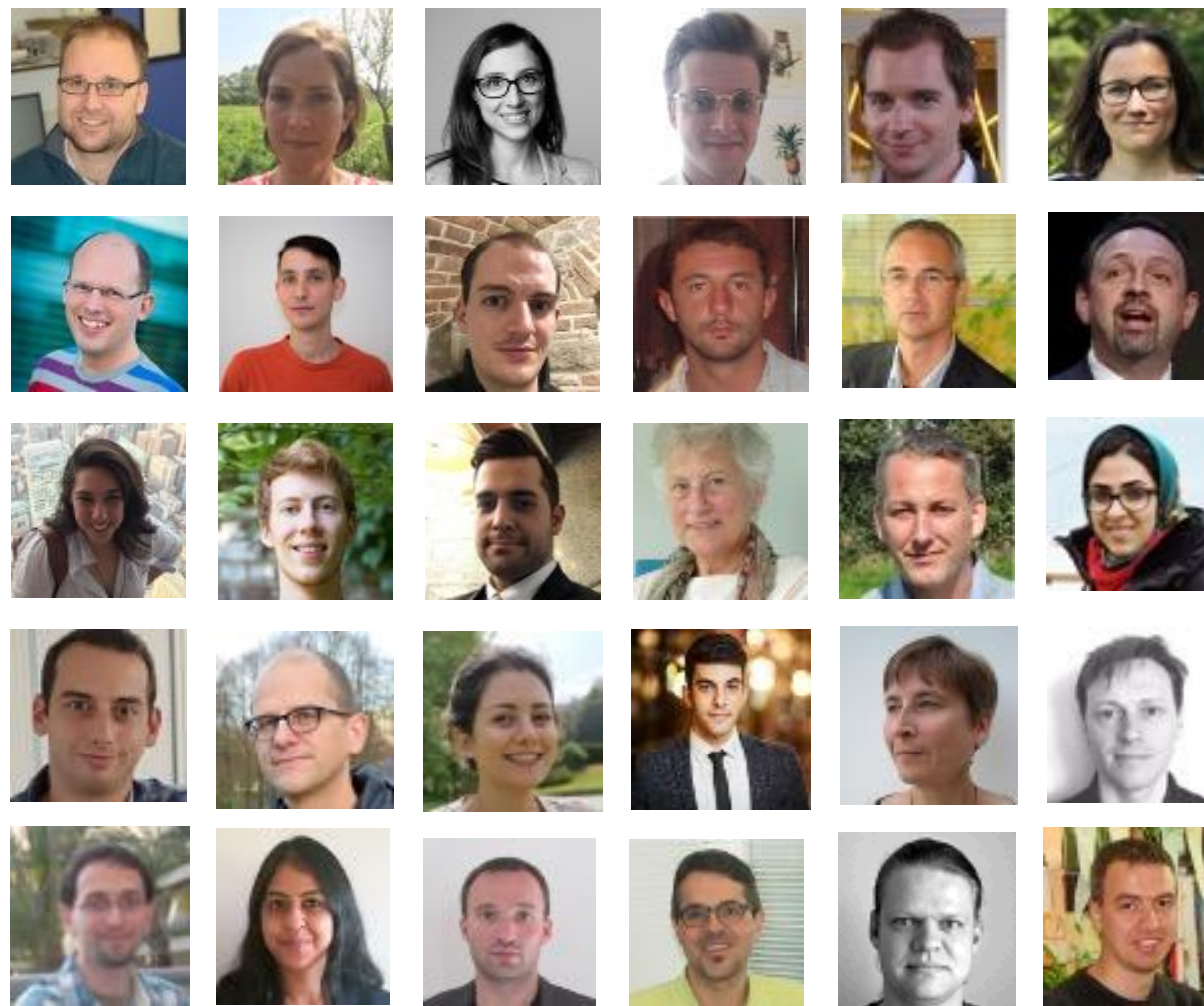
intergovernmental organisation that brings together life science resources such as

- databases
- software tools
- training materials
- interoperability resources
- compute resources
- data management support



The goal of ELIXIR is to **coordinate bioinformatics resources from across Europe so they form a single infrastructure.**

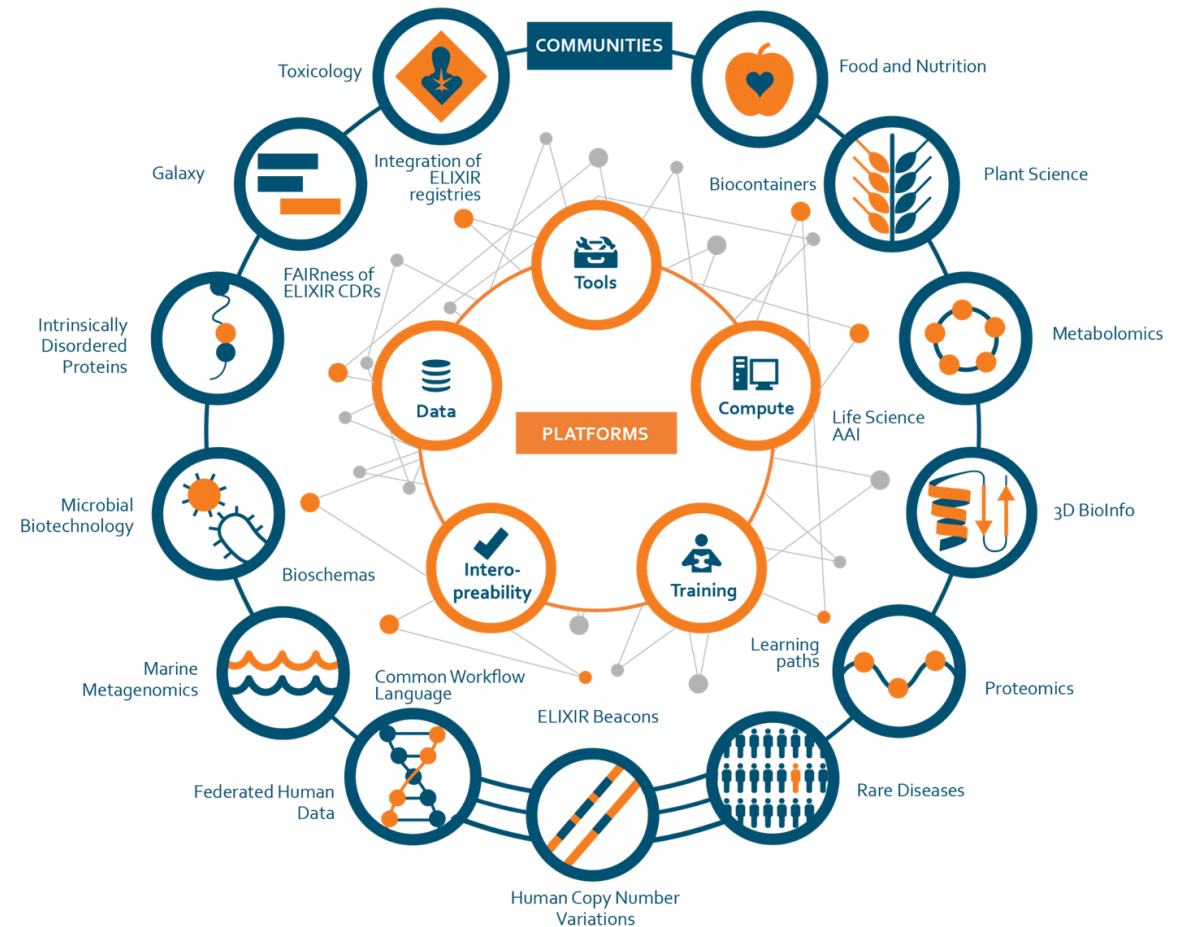
ELIXIR Belgium



SCIENCE MEETS LIFE

ELIXIR Communities connect infrastructure with life-science research experts

- Formed around domain experts in ELIXIR Nodes (including non-ELIXIR partners)
- Provide a mechanism for long-term collaborations with other ESFRIs and large-scale initiatives
- Drive service developments in the ELIXIR Platforms
- Provide framework to develop and maintain community standards

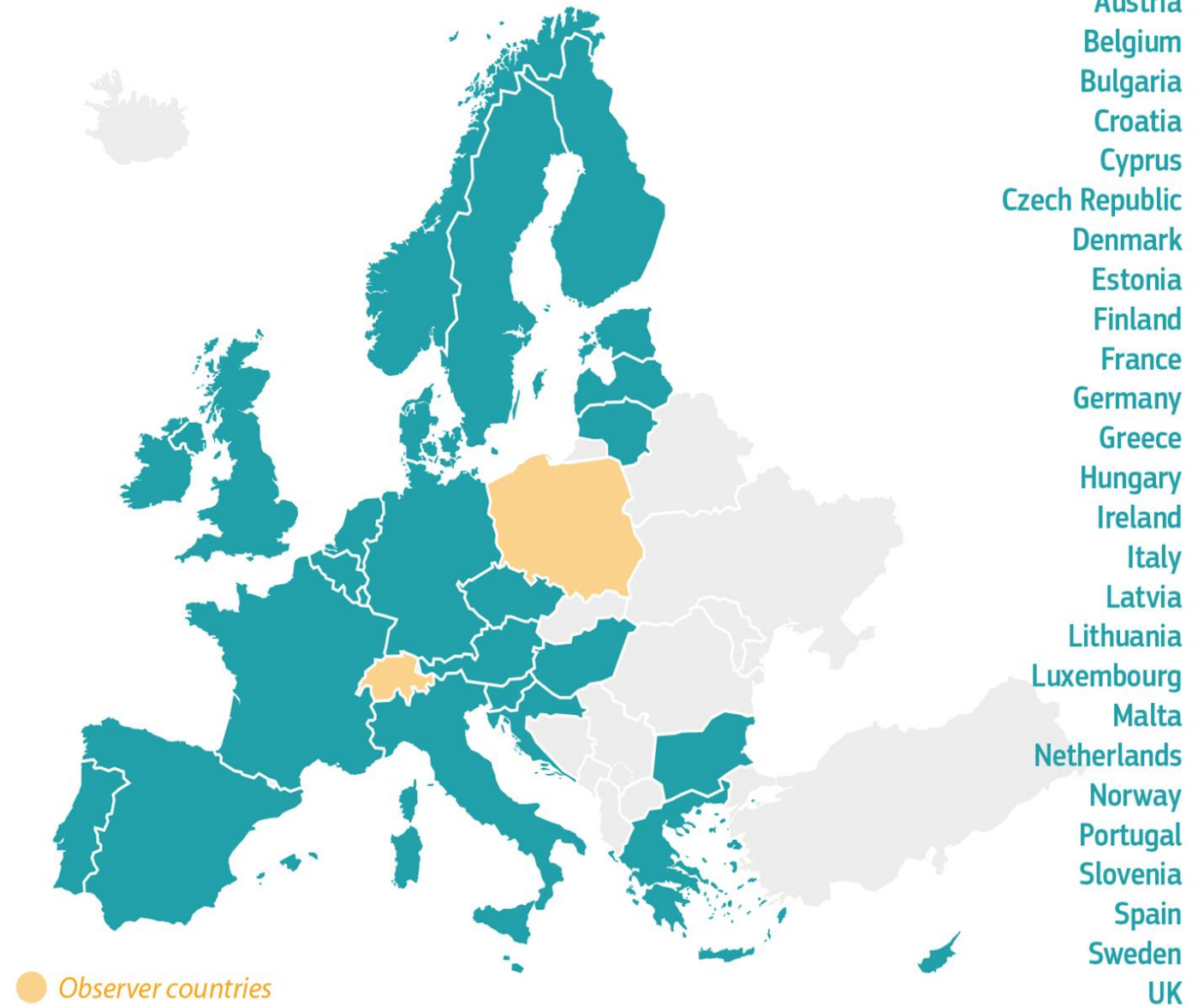


Team of infrastructure experts, guided by research

In the context of a European vision

1+MG initiative

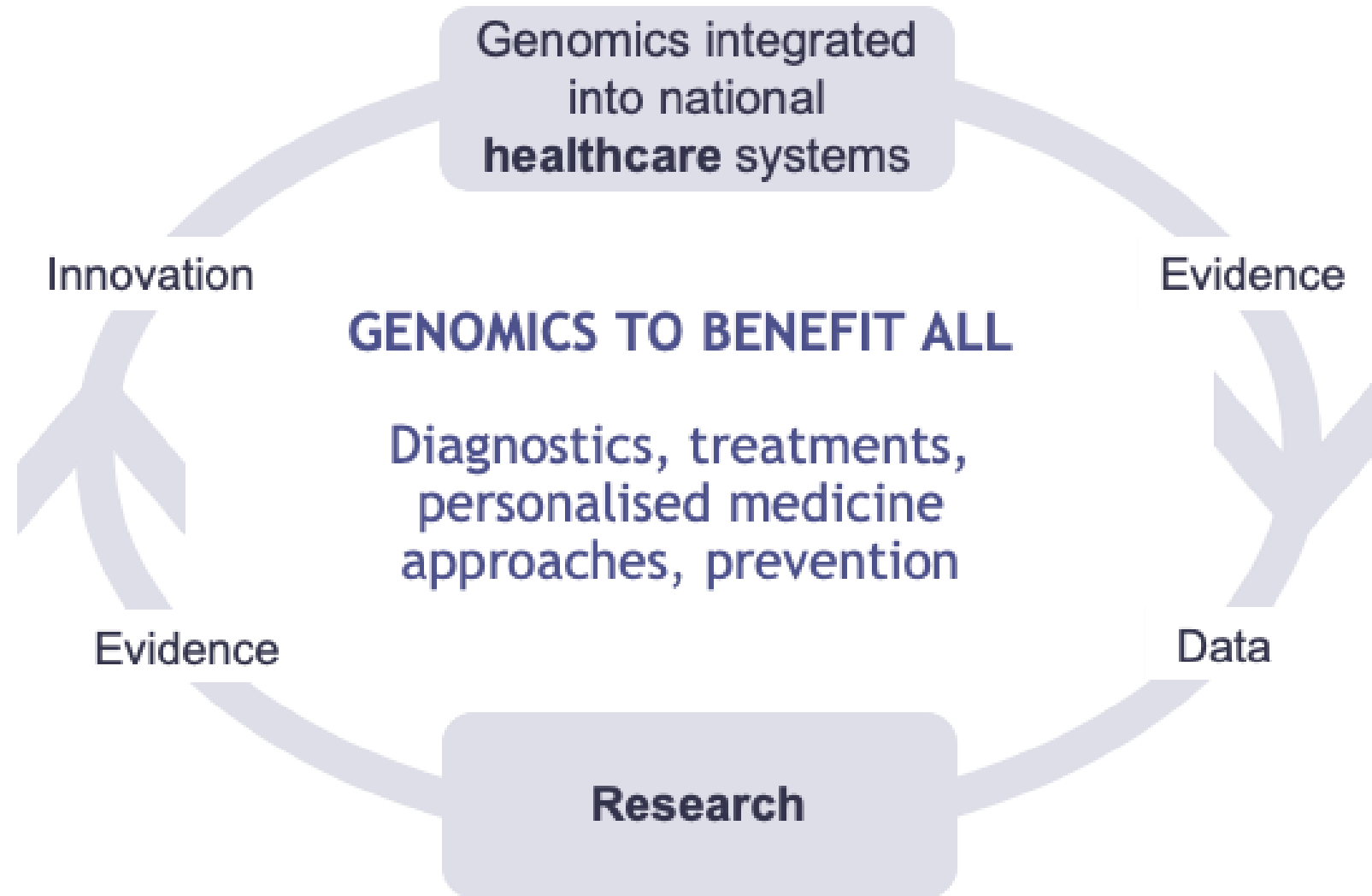
realise a practice of personalised medicine and health, based upon a shared 'trust framework' and the infrastructure to safely access and integrate high quality genomic data and other health data across borders



updated on 16 November 2022

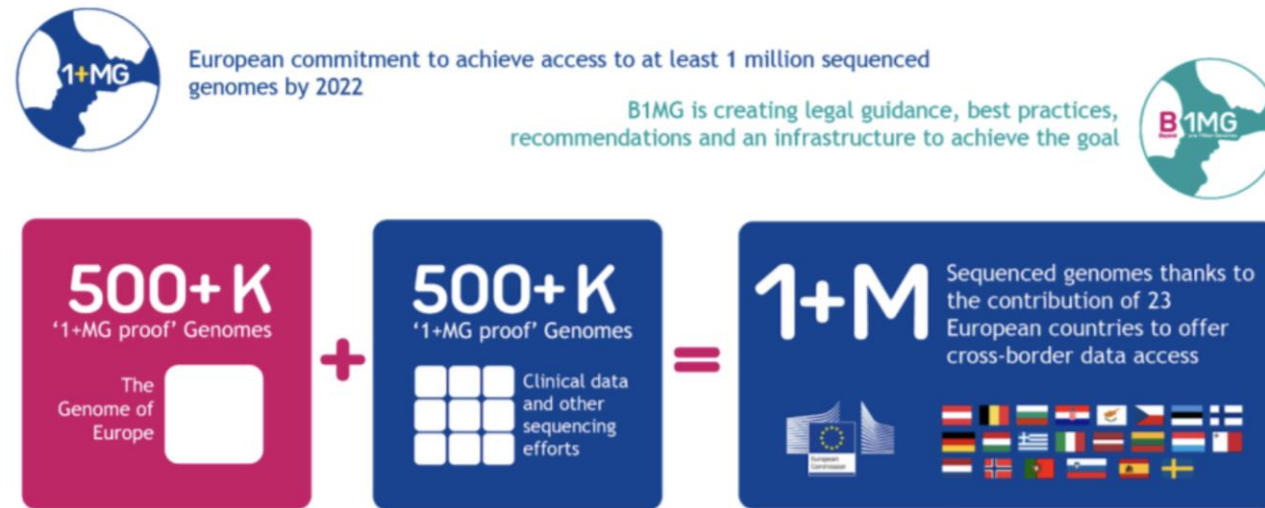
Aims of 1+MG

+ Healthcare
+ Policy
+ Research



Access to & linking of data as driver for innovation

Genome of Europe



build a high-quality European network of national genomic reference cohorts, representative of the European population

In Belgium started by the Belgian Genome Biobank project (FWO)

A European vision for the future of healthcare

Building on technical solutions, coordinated across borders

Federated European Genome-Phenome Archive (FEGA)

FEGA as primary resource for the discovery of and access to human data for research in Europe

- A federated network of human data repositories, regional hubs, and cloud analysis environments
- Platform for discovery, access, distribution of human data
- Coordination and consistent standards for discovery and access
- Network of experts to build capacity across Europe



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



SCIENCE MEETS LIFE

FEGA Maturity Model

- Path for nodes to join Federated EGA
- Aligned with other maturity models, e.g. 1+MG

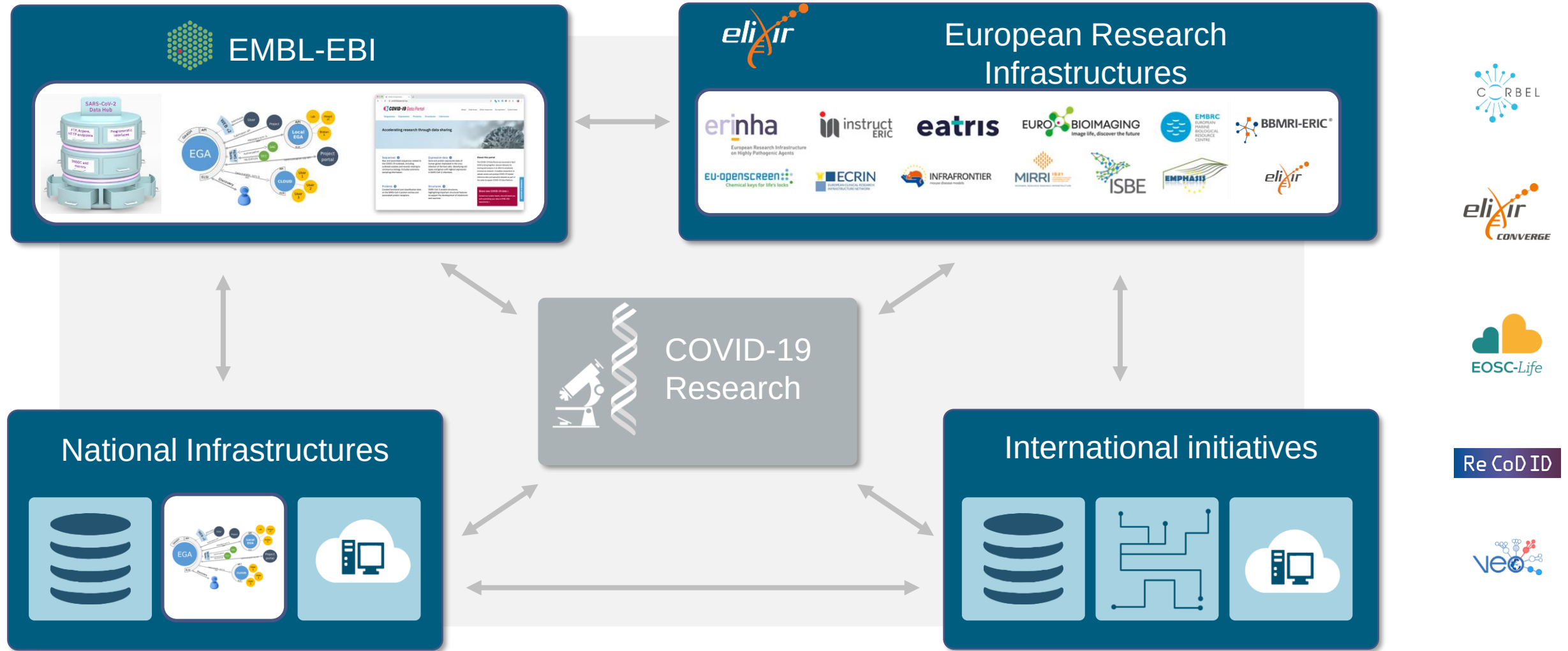
FEGA MM	Level definitions
[1] Governance, Strategy & Sustainability	+
[2] Legal	+
[3] Data & metadata management	+
[4] Technical Infrastructure	+
[5] Operations Support	+
[6] Communications, Community Building, and Engagement	+



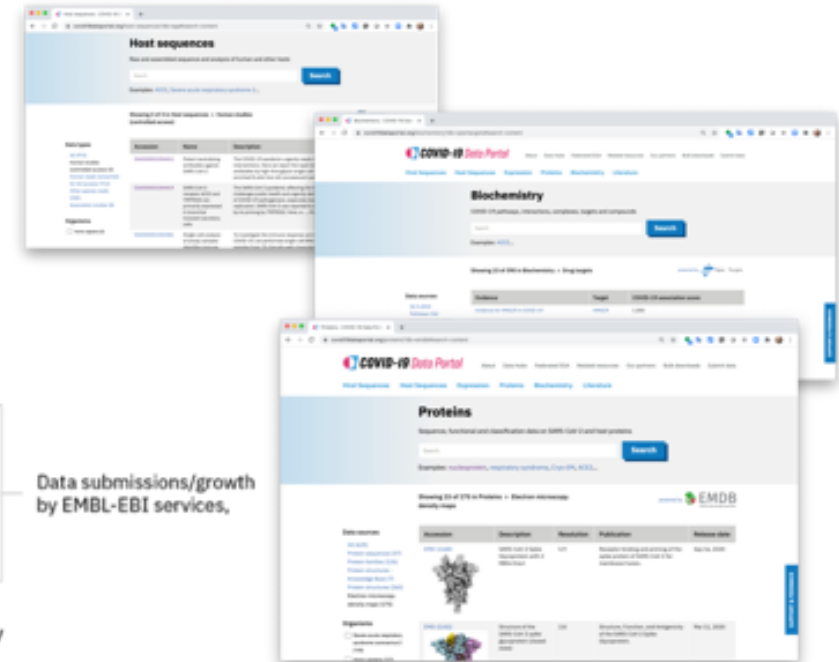
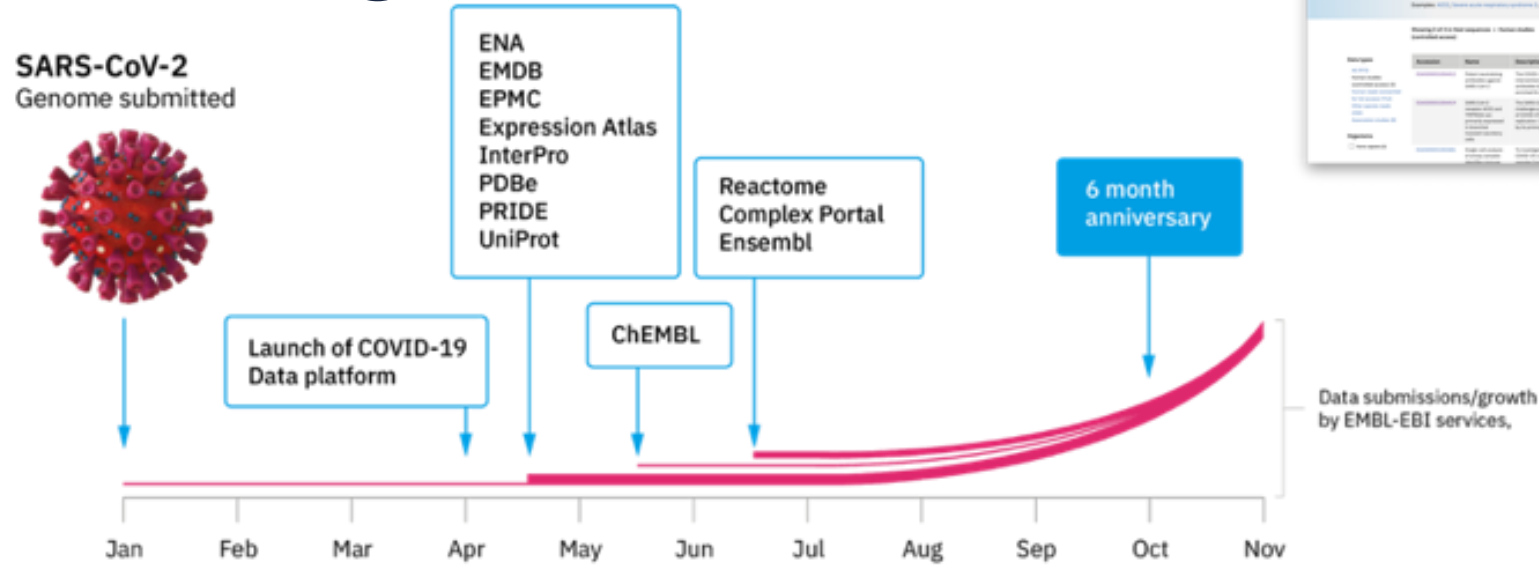
<https://inab.github.io/fega-mm/>

SCIENCE MEETS LIFE

European COVID-19 Data Platform



COVID-19 Data Portal



Viral Sequences

All (369,429)
 Sequences (46,327)
 Reference sequences (1)
 Raw reads (188,088)
 Sequenced samples (122,092)
 Studies (207)
 Genes (22)
 Browser (1)
 Variants (12,691)

Host Sequences

All (5,413)
 Human studies (controlled access) (6)
 Human reads (consented for full access) (3,693)
 Other species reads (1,692)
 Association studies (22)

Expression

All (87)
 Gene expression (4)
 Gene expression experiments (23)
 Single cell expression (4)
 Single cell expression experiments (19)
 Protein expression experiments (37)

Proteins

All (1,010)
 Protein sequences (79)
 Protein families (315)
 Protein structures - Knowledge Base (7)
 Protein structures (260)
 Electron microscopy density maps (334)
 Electron microscopy public image archive (15)

Biochemistry

All (2,709)
 Pathways (16)
 Interactions (2,202)
 Complexes (29)
 Compound document (8)
 Drug targets (453)
 Metabolomics experiments (1)

Imaging

All (17)
 Images (2)
 Electron microscopy public image archive (15)

Literature

All (266,965)
 Coronaviruses (134,918)
 Diseases (123,805)
 Related viruses and diseases (2,231)
 Genes, receptors and antibodies (6,006)
 Supplementary material (5)

<https://www.covid19dataportal.org/>

Slide by Guy Cochrane

SCIENCE MEETS LIFE

**Technical solutions, coordinated
by ELIXIR, grounded in Open
Science & FAIR data principles**

European Genomic Data Infrastructure

Support better healthcare & research



**European
Genomic Data
Infrastructure**

- Deploy the infrastructure for 1+MG in European countries
- Develop innovative solutions to tackle disease scenarios
- Models & frameworks for long-term sustainability & governance of the infrastructure



+ Healthcare
+ Policy
+ Research



SCIENCE MEETS LIFE

GDI participation in numbers

44 project beneficiaries

4 Affiliated Entities

6 Associated Partners

54 project partners

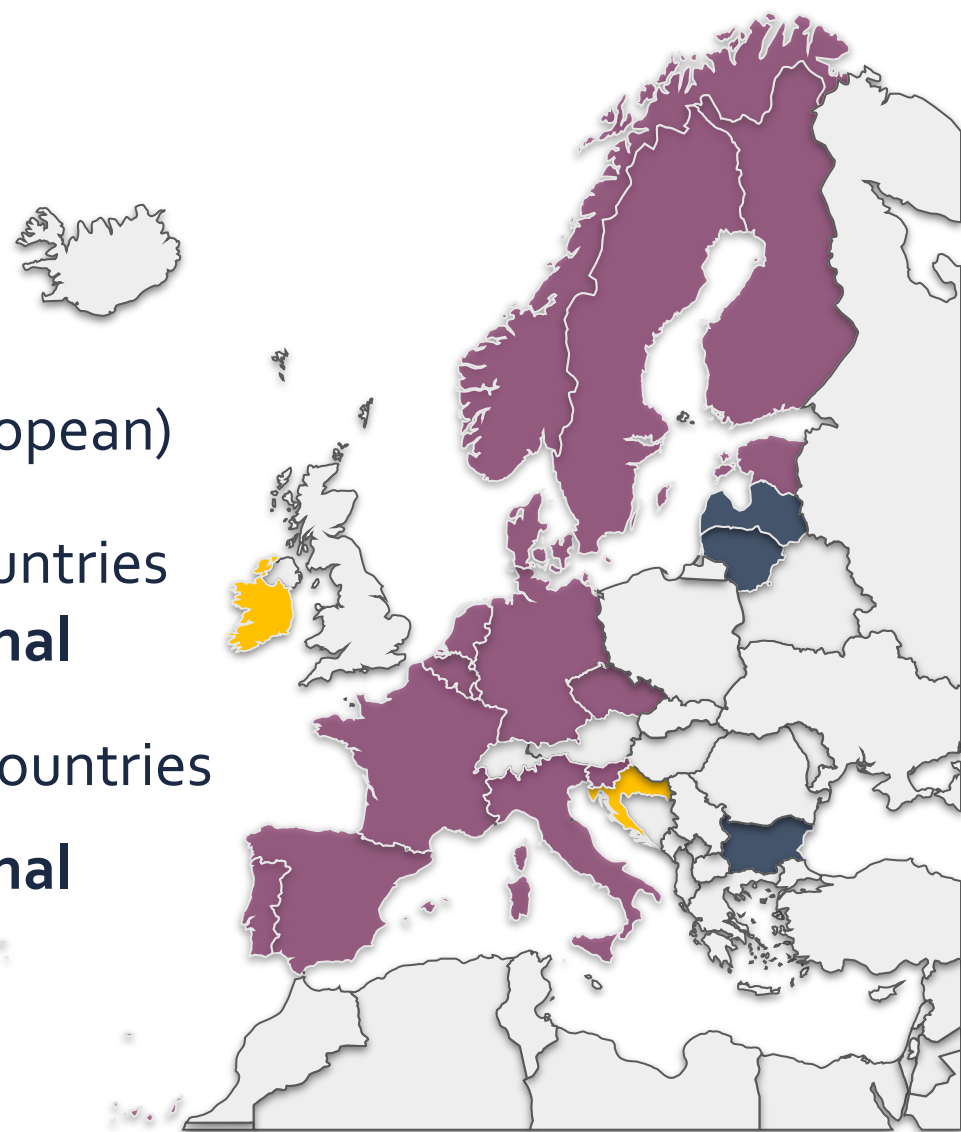
2 Research Infrastructures
(ELIXIR & BBMRI)

€40m co-funded **50%**

20 countries (all European)

2024 at least **6** countries
technically operational

2026 at least **15** countries
technically operational



GDI outcomes

- Build infrastructure to store, process & share sensitive data
 - Federated
 - Sustainable
 - Secure
 - Based on Open community standards
- Interoperability with health & care data
- Enable globally competitive research, personalised healthcare and health policy making

Infrastructure of strategic importance for healthcare & research

Open & FAIR data in research

Facilitate

- cross-border exchange of data through common interfaces
- harmonisation by further developing & applying common standards

To enable

- Re-use of data
- Integration of diverse datasets, across domains e.g. health & care
- Access to a large compendium of high quality data for ML/AI approaches

EHDS as a partner

Driver for a similar revolution in health & care

- Make health & care data discoverable & accessible for re-use
- Harmonisation of health & care data

Find the right approach to strengthen the existing ecosystem

Genomic data paves the way

- Linking genomic & health data crucial for e.g. personalised medicine
- Differences between genomic data & health and care data e.g. EHR
- Dedicated technical solutions needed to handle
 - ▶ Large volume of data (Terabyte to Petabyte)
 - ▶ Access to large compute facilities for standard processing
 - ▶ Existing ecosystem of repositories & standards
- Changing approaches
 - ▶ Bring analysis to the data
 - ▶ Federated learning

Building an ecosystem of initiatives

- Build on existing & emerging fit-for-purpose solutions
- Reuse solutions & infrastructure where possible
- Focus on interoperability : connecting & integrating
- Without additional requirements & procedures where solutions exist



SCIENCE MEETS LIFE

Acknowledgements



DEPARTEMENT
ECONOMIE
WETENSCHAP &
INNOVATIE



SCIENCE MEETS LIFE

