Sensitive data archiving in Nordics and Federated EGA

2019-05-15
NeIC 2019 - Open Science with Sensitive Data workshop

Niclas Jareborg, NBIS / ELIXIR-SE

Slides adapted from other presentations by
Jordi Rambla, CRG / Thomas Keane, EMBL-EBI / Gary Saunders, ELIXIR-Hub / Tryggve2 team members
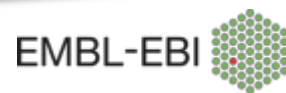
Tryggve2 major deliverables

1. **Sensitive data archiving**
2. Production quality processing services
3. Homogenized user experience
   - User mobility
   - Workflow mobility
   - Data mobility
4. Nordic use cases
   - Research
   - Infrastructure development
5. ELIXIR AAI
6. IT Security
7. ELSI Topics

EGA

The EGA was created by the EBI, in 2008, as an extension of the ENA…

EGA Mission Statement

"The European Genome-phenome Archive (EGA) seeks to accelerate data sharing for human genetic and phenotype data, consistent with local ethical agreements and international standards, to enable molecular medicine."

What is the EGA?

The EGA is a resource for permanent secure archiving and sharing of all types of potentially identifiable biomolecular and phenotypic data resulting from biomedical research projects.
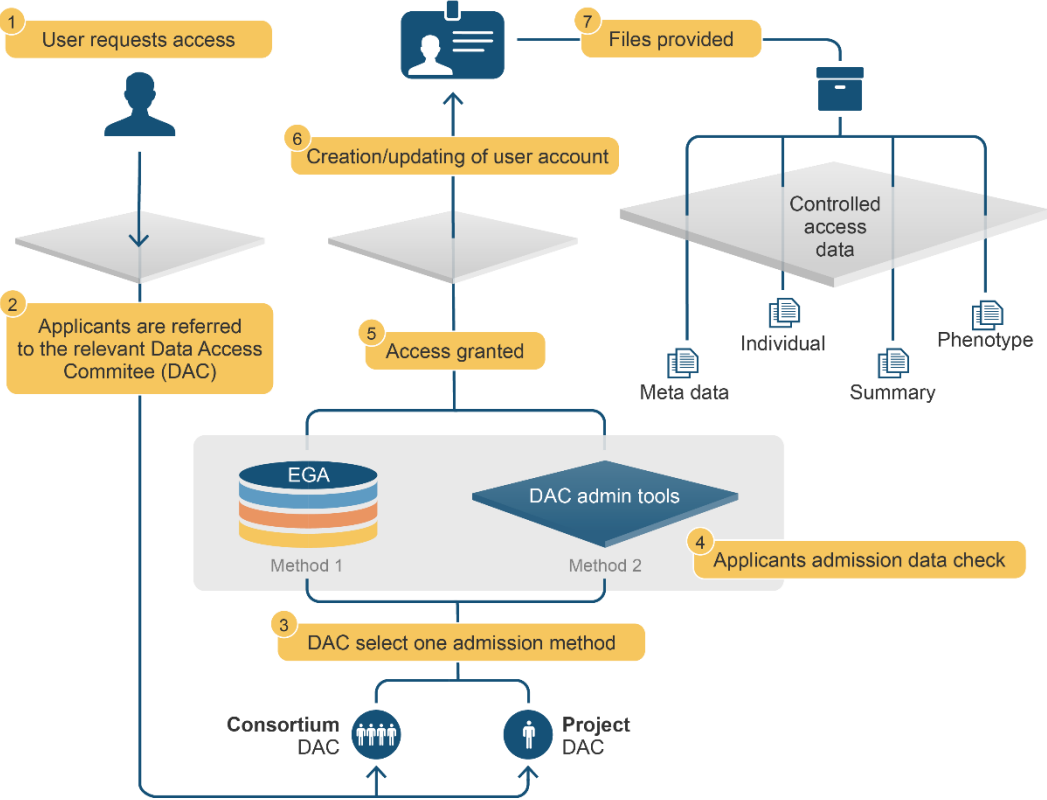
Data is provided by research centers and health care institutions.
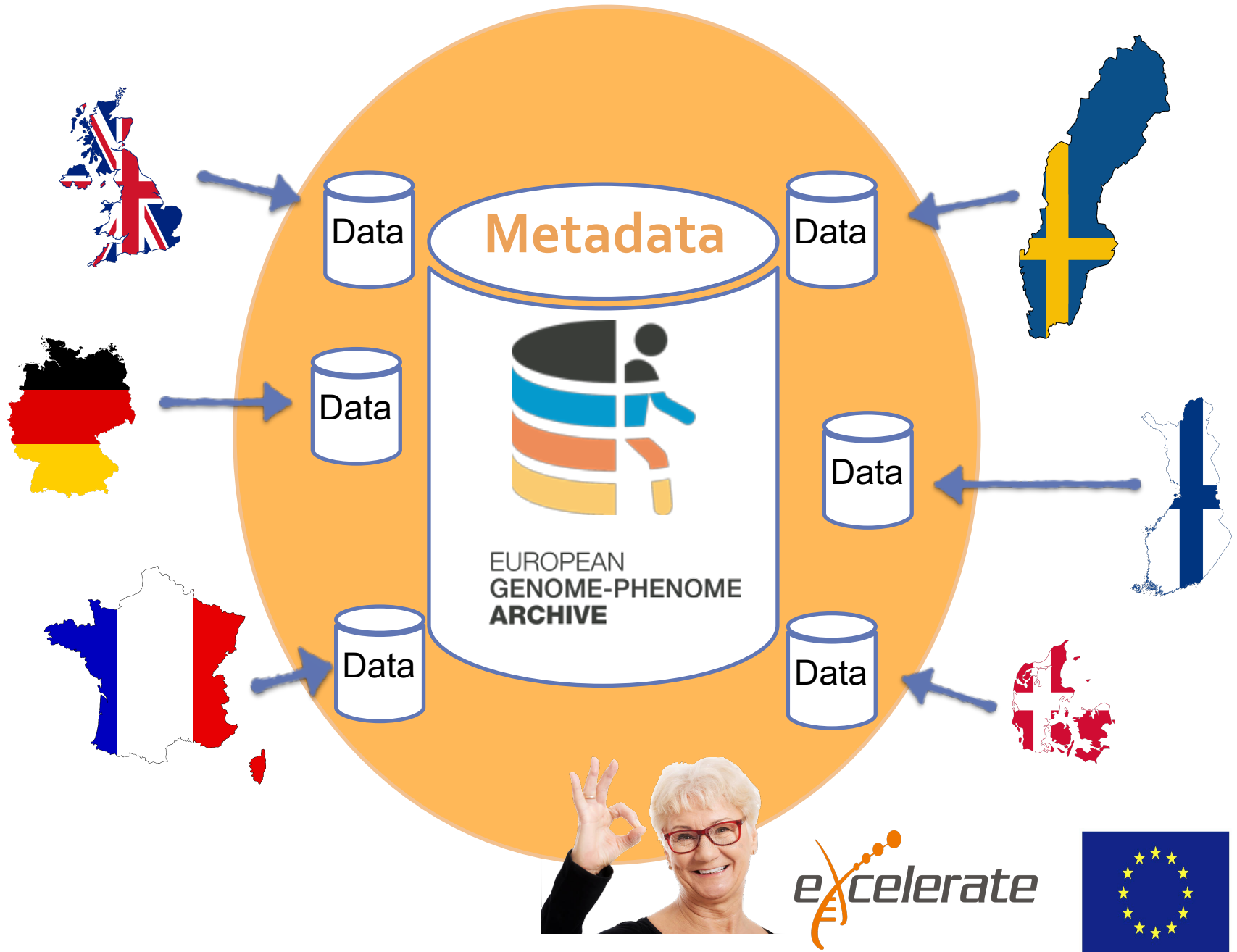
Access is controlled by Data Access Committees.

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

https://ega-archive.org

EUROPEAN GENOME-PHENOME ARCHIVE

5

EUROPEAN
GENOME-PHENOME
ARCHIVE

excelerate

D-1 Sensitive data archiving - Key Targets

- Develop a **code base** for archiving of sensitive data
  - *Compatible with the needs of ELIXIR-Excelerate WP 9.1.2*
  - Packaged for convenient deployment
  - **Support core processes & other relevant processes**
    - "Data in" & "Data out" (through several mechanisms)
    - Federated querying

  - *Also for data that will **not** be exposed through EGA*

- Establishment of **service instances** in member nodes

Cross-border team

- Tryggve partners
    - ELIXIR Nodes: Finland, Norway, Sweden
- Central EGA
    - CRG (ELIXIR-ES) & EMBL-EBI
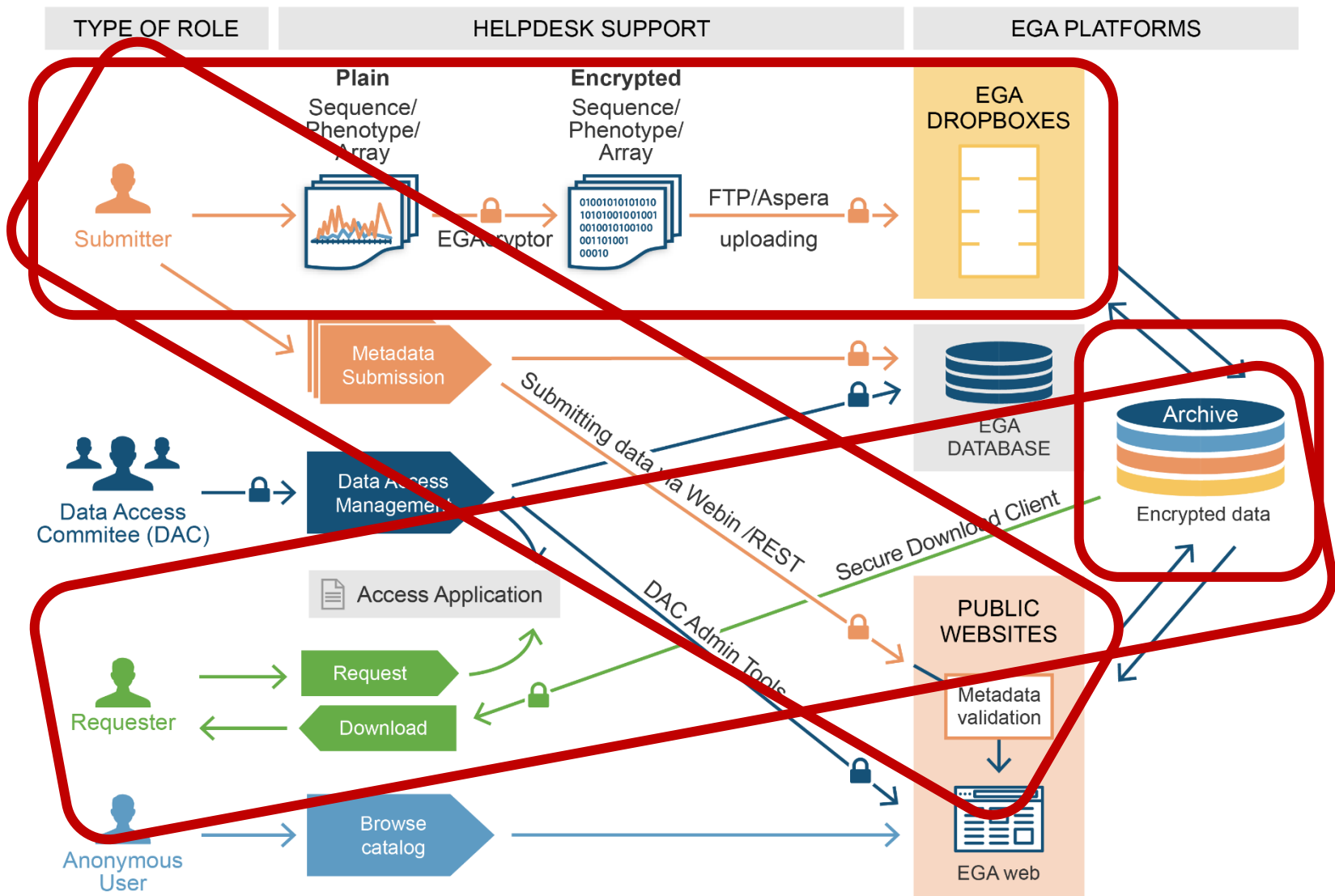
D-1 Sensitive data archiving - milestones

- **2018-12** – Deployment of core functionality code base in a production-like environment demonstrated -   *(2019-02)*

- **2019-05** – At least one sensitive data archiving instance technically ready to join the Federated EGA

Tryggve

Nordic Vision for Data Management

Developing competence and capacity to manage sensitive data –

Federation of EGA service to the Nordic Countries

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

elixir

Tryggve
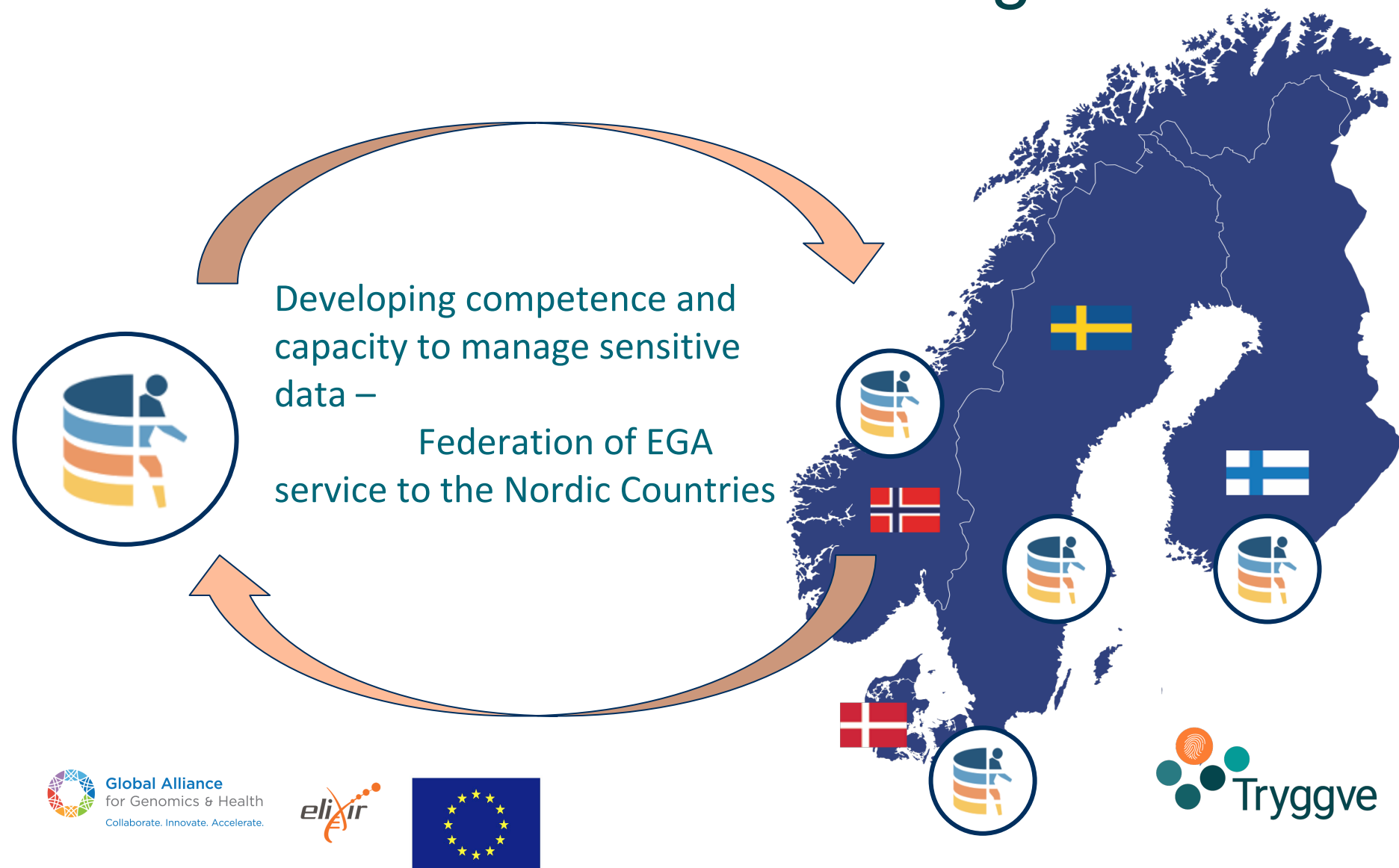
Future ways of accessing data
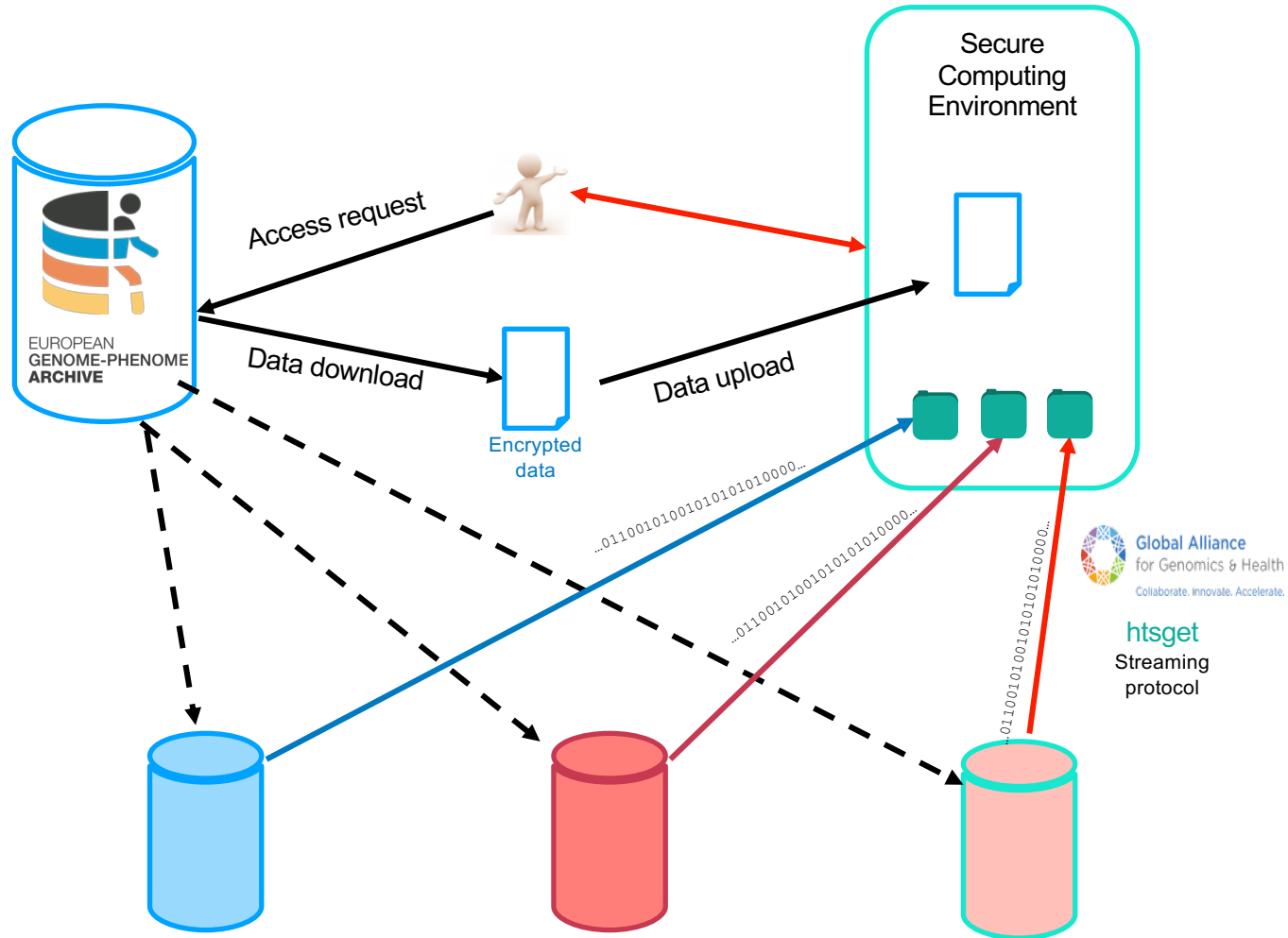
NorTwinCan use case
Possible technical solution

The GA4GH Mission

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**.
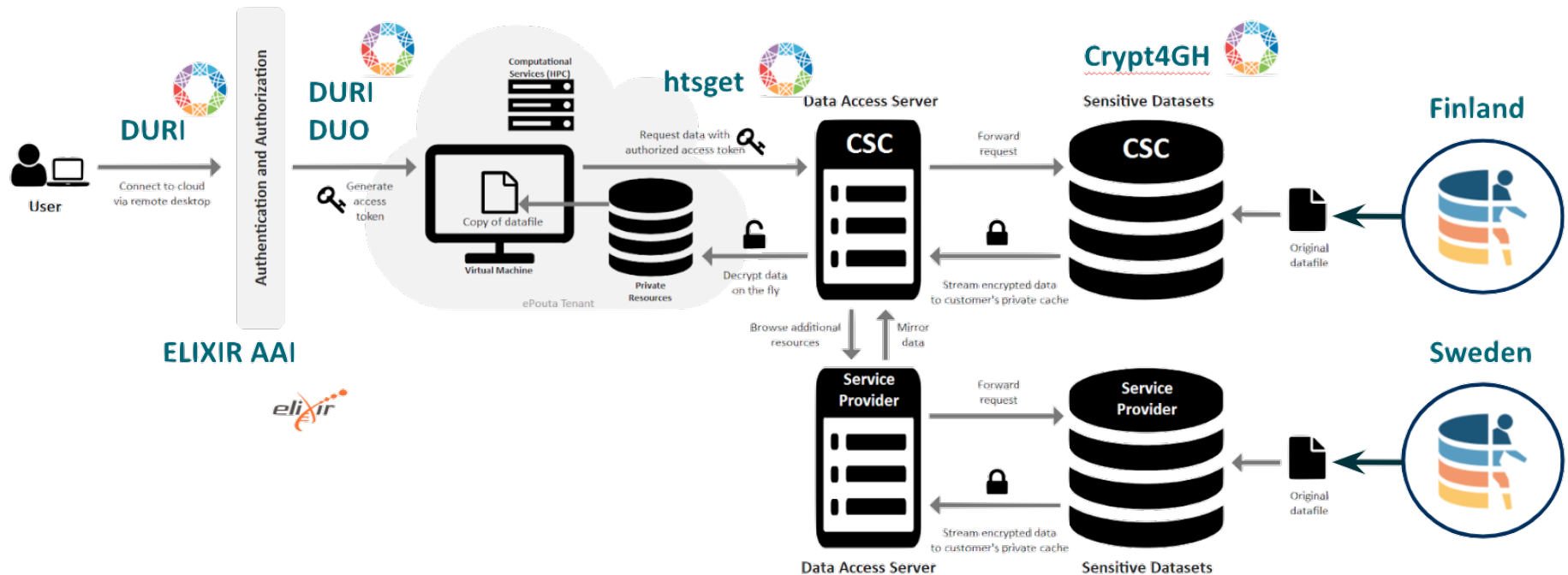
Core Principles of Data Sharing

- Enable **international** data sharing

- 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

GA4GH and ELIXIR have formed strategic partnership

GA4GH

Developing standards for international data sharing

ELIXIR

Adopting and implementing standards focused on European countries
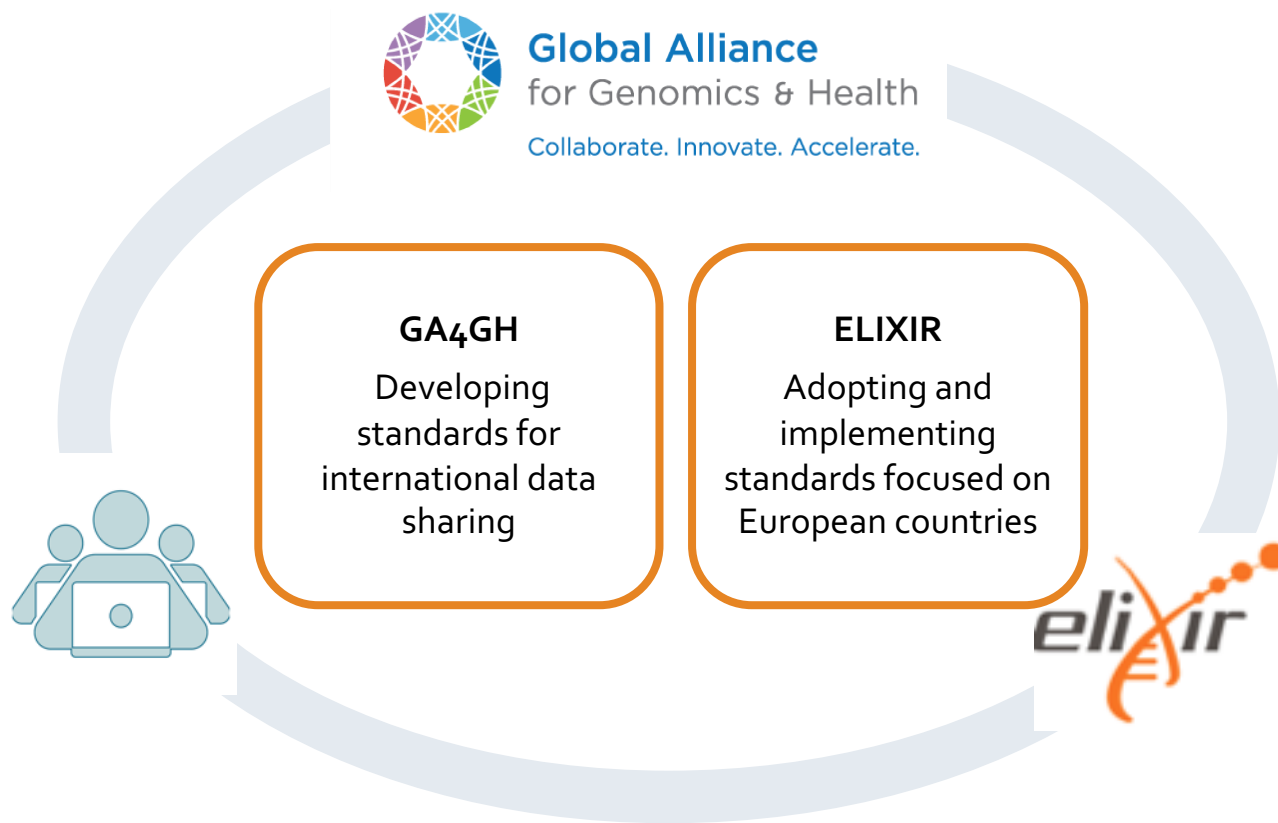
Working towards GA4GH standards, APIs and toolkits to be used throughout ELIXIR Nodes for human data discovery and access

ELIXIR Human Genomics & Translational Data - technically

Data

Training

Compute

Interoperability

Tools

Data Discoverability
Federating lightweight discoverability of data, and datasets across ELIXIR

ELIXIR Beacon - GA4GH Driver Project

Data Archival
Utilising the ELIXIR Deposition Databases to ensure secure, long-term, efficient archival of data

EUROPEAN GENOME-PHENOME ARCHIVE

Federated Data Access
Coordinating a collection of interoperable EGA-like resources to ensure secure management of sensitive data across the ELIXIR Nodes

ELIXIR Federated Human Data Community - htsget/refget

Data Analysis
Bringing 'analysis to data' via common workflow languages, workflows, containers, and tools

trait
a Lygature platform
bio.tools
RD-Connect

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

ELIXIR Federated Human Data (2019-21)

- Many human data initiatives across Europe

- ELIXIR human data community to coordinate approaches
  - Metadata, phenotypes, AAI, ELSI, discovery

DECLARATION OF COOPERATION

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

Česká republika
and
Eesti Vabariik
and
Reino de España

OpenRiskNet

iASiS
Big Data for Precision Medicine

NeIC Tryggve
Collaboration on sensitive data

CORBEL

RD Connect

BBMRI-ERIC®

EOSCpilot
The European Open Science Cloud for Research Pilot Project

EUROPEAN GENOME-PHENOME ARCHIVE

ELIX

- M

- EL
  co
  -

DECLARATION OF COOPERATION

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

Česká republika

and

Eesti Vabariik

and

Reino de España