e-INFRA CZ

SensitiveCloud

Conference e-INFRA 2024, 2024-04-30

Michal Růžička <ruzicka@ics.muni.cz>
Jaroslav Juráček <juracek@ics.muni.cz>

cesnet

MUNI
CERIT-SC

VŠB TECHNICKÁ UNIVERZITA OSTRAVA

IT4INNOVATIONS
NÁRODNÍ SUPERPOČÍTAČOVÉ CENTRUM

Motivation

Infrastructure for Sensitive Data

- Security is an increasingly emphasized area.
  - Users are gradually placing more and more emphasis on the security of their data.
  - Various parts of the university (Med, CEITEC, RECETOX, CERIT-SC, ...) work with medical and other sensitive data.
- Two “hot topics” in data management and processing.
  - Open and FAIR data – The principles of Open and FAIR data (Findable, Accessible, Interoperable, Reusable, ...) are needed as the simple availability of data does not guarantee its usability.
  - Sensitive Data – Primarily in Health and Life Science, which creates a clear demand for processing sensitive data.
- A strategic goal of CERIT-SC / e-INFRA CZ – Supporting life-science and high-value users.

SensitiveCloud

Infrastructures for Sensitive Data

Our Team

- Selected specialists from CERIT-SC team.
    - Technical implementation (storage and computation software and hardware, data networks, datacenters personnel, …).
    - Compliance, risk management, ...
    - Service design.
    - Legal consultation.
    - Security research (forensic readiness).

Environment Design and UX

- Higher security == worse UX.
- Higher security == requirement for higher service maturity (ITIL, ISO, ...).
- Environment design based on:
  - “General security best-practice”.
  - User interviews and testing.
- Developed semi-formal framework:
  - On-boarding process for users: first contact, interview(s), contract signing, access and training, deployment.
  - Description of the environment.
  - User rights and responsibilities.
  - Contract template.
  - General conditions of use of the service.

Implementation

- Two components:
  - SensitiveCloud Compute (including GPUs).
  - SensitiveCloud Storage.
- SensitiveCloud Compute is provided as a PaaS built on Kubernetes with Rancher:
  - The user only manages the application, not the entire VM.
  - Ideal SaaS – R-Studio, Jupyter Notebook, ...
- SensitiveCloud Storage.
  - Integration within computing via NFS-CSI.
- Separate network, WireGuard VPN access, Perun AAI with multi-factor, iron behind the lock with camera surveillance, ...
- BUT! ... Some users would still like VMs...

General Security Best Practice

- Rational search for possible vulnerabilities, attacks and countermeasures.
  - From formal risk analysis to wise men over coffee meetings...
- ISO 27k:
  - Not a prerequisite.
  - Not a sufficient condition.
  - By itself, addressing process maturity and higher security de facto is a by-product.
  - SensitiveCloud certified in summer 2023.
- BUT:
  - The correct solution implicitly must necessarily be ISO 27k compliant.
- Forensic readiness.
  - The security methods research applied to SensitiveCloud infrastructure.

Forensic Readiness

- (Not only) a research area dealing with environmental preparedness for incident impact analysis.
  - “Trouble comes; after the trouble comes the cop, and then what?”
  - Risk description, analysis (access token theft, VPN connection breach, ...).
- Scenario analysis.
  - Data tracing.
- Simulation of scenarios.
  - Controllably populate risk on running infrastructure (simulated incident).
  - Observe how teamwork works when responding to an incident.
  - Verify that the collected data (logs, ...) is sufficient for a proper investigation of the incident.
    - ... and that it matches our expectations...

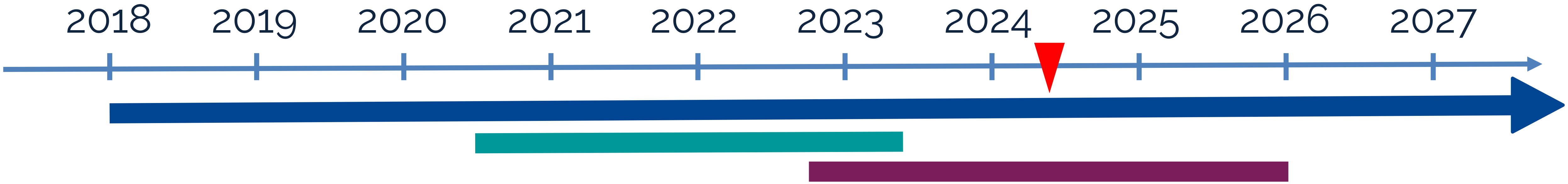
Continuous Process

- ISO 27k is about continuous planning, monitoring and improvements.
- In the process of applying for projects planning to use SensitiveCloud.
  - Possible extension of hardware resources.
  - Possible new integrations with infrastructures.
  - Possible new use-cases.
- Continuous improvements of processes, technical security measures, onboarding process, and SensitiveCloud public presentation, ...
- Thinking of SensitiveCloud 2.0, 3.0, …
  - Full virtual machines, virtual desktops in the secure environment.
  - Dedicated Kubernetes/OpenStack platform deployed on-demand.

GDI/FEGA

Infrastructures and Tools to Share Genomic Data across Europe

Genomic Data Infrastructures



1+MillionGenomes

Creation of **infrastructure** objectives and rules that will enable **secure access to genomic data** across European countries.

Beyond 1 Million Genomes

H2020 project.

Concept and testing phases.

Fulfilling the goals of the 1+MG initiative.

Legal and technical **guidance**, data **standards**, best practices.

European Genomic Data Infrastructure

Digital Europe project co-funded by Member States.

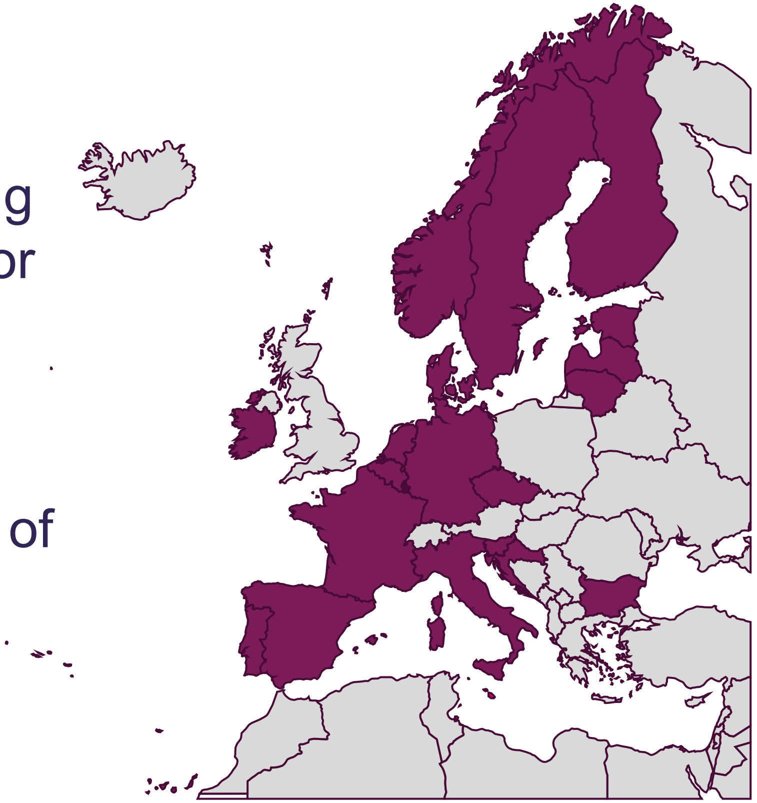
Growth and sustainability phases.

It builds on the results of 1+MG, B1MG and related projects of European countries.

Creates a federated data infrastructure.

Goals and Structure of the Genomic Data Infrastructure (GDI)

- It will enable access to genomic (and related phenotypic and clinical data) across Europe. Creating a federated, sustainable, and secure infrastructure for data access.
- Fulfilling the 1+MG vision – ensuring the readiness and sustainability of the infrastructure of individual member countries that will enable federated sharing of genomic data.
- 1+MG, B1MG and GDI are independent entities (structure, organization, financing).
- A total of 54 partners from 20 European countries.
- Primary use for clinicians.



Genomic Data Infrastructure – Member Countries

- A Member State undertakes to:
  - Provides a node or data center within the network.
  - Each country manages its own data (e.g. national/regional nodes).
  - The data nodes will make cross-border data analytics available using a common standards framework and APIs.
- The overall data infrastructure provides 5 main functions:

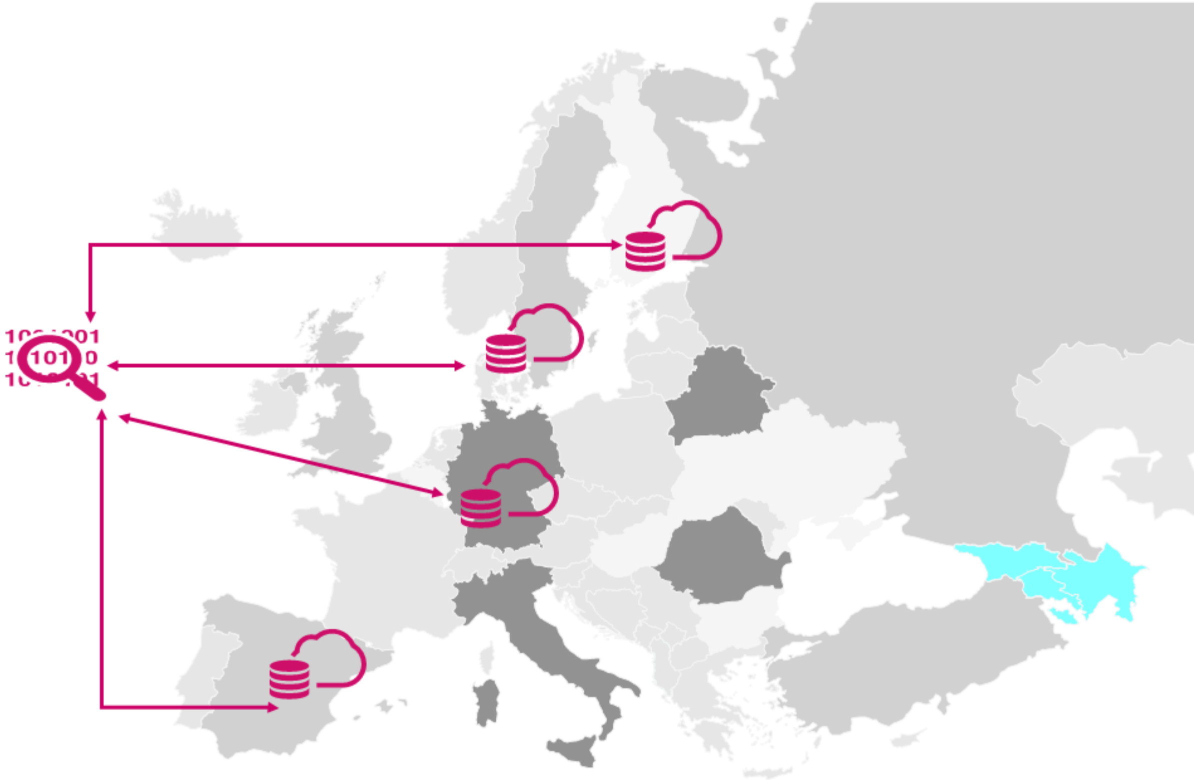
Data discovery

Access management tools

Data processing

Data reception

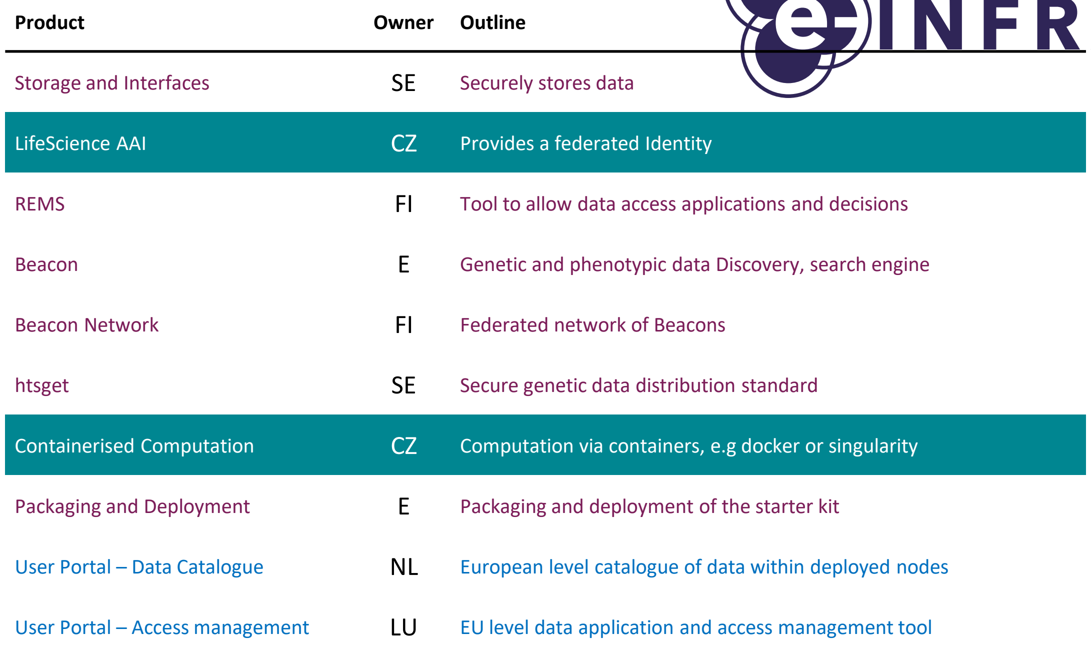
Storage and interfaces



GDI Technical Solution

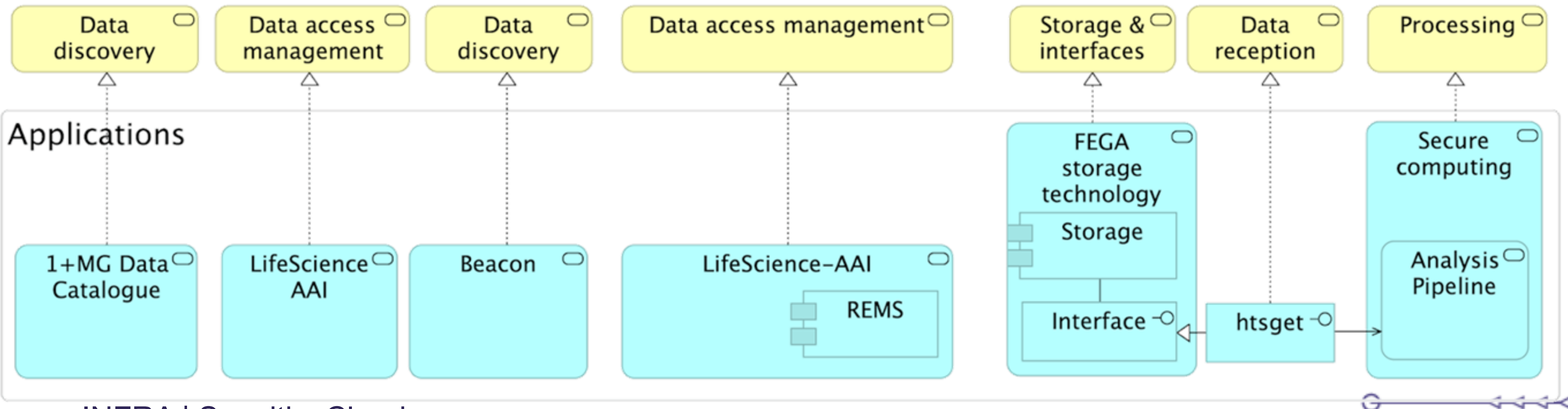
- A set of tools and components shared between member states.
- First version (starter kit) – some tools and components.
  - June 2023 (9th ELIXIR All Hands meeting, Dublin, Ireland).
  - Demonstration of access to genomic and phenotypic data between member states / nodes.
  - Set designed with the use of existing components, based mainly on the functionalities of the B1MG.

e-INFRA CZ



Product	Owner	Outline
Storage and Interfaces	SE	Securely stores data
LifeScience AAI	CZ	Provides a federated Identity
REMS	FI	Tool to allow data access applications and decisions
Beacon	E	Genetic and phenotypic data Discovery, search engine
Beacon Network	FI	Federated network of Beacons
htsget	SE	Secure genetic data distribution standard
Containerised Computation	CZ	Computation via containers, e.g docker or singularity
Packaging and Deployment	E	Packaging and deployment of the starter kit
User Portal – Data Catalogue	NL	European level catalogue of data within deployed nodes
User Portal – Access management	LU	EU level data application and access management tool

https://github.com/GenomicDataInfrastructure



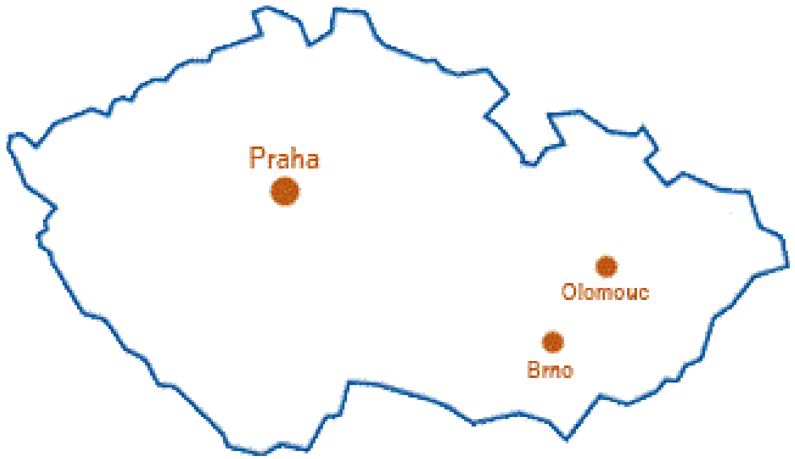
Galaxy PROJECT

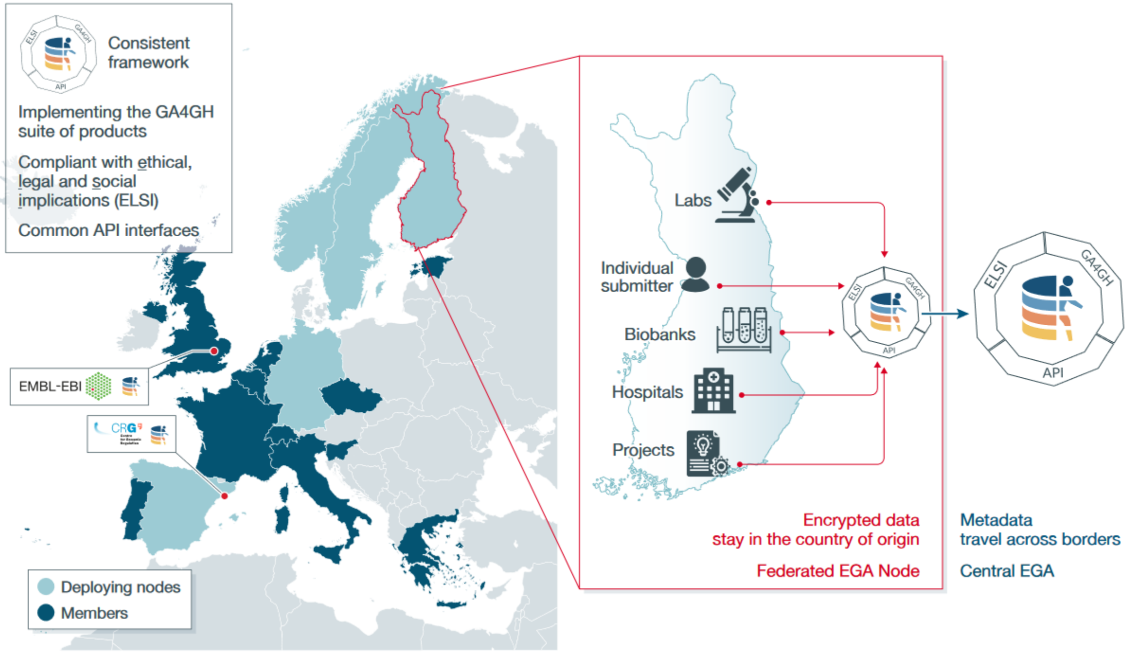
BIOCONDA

Flower

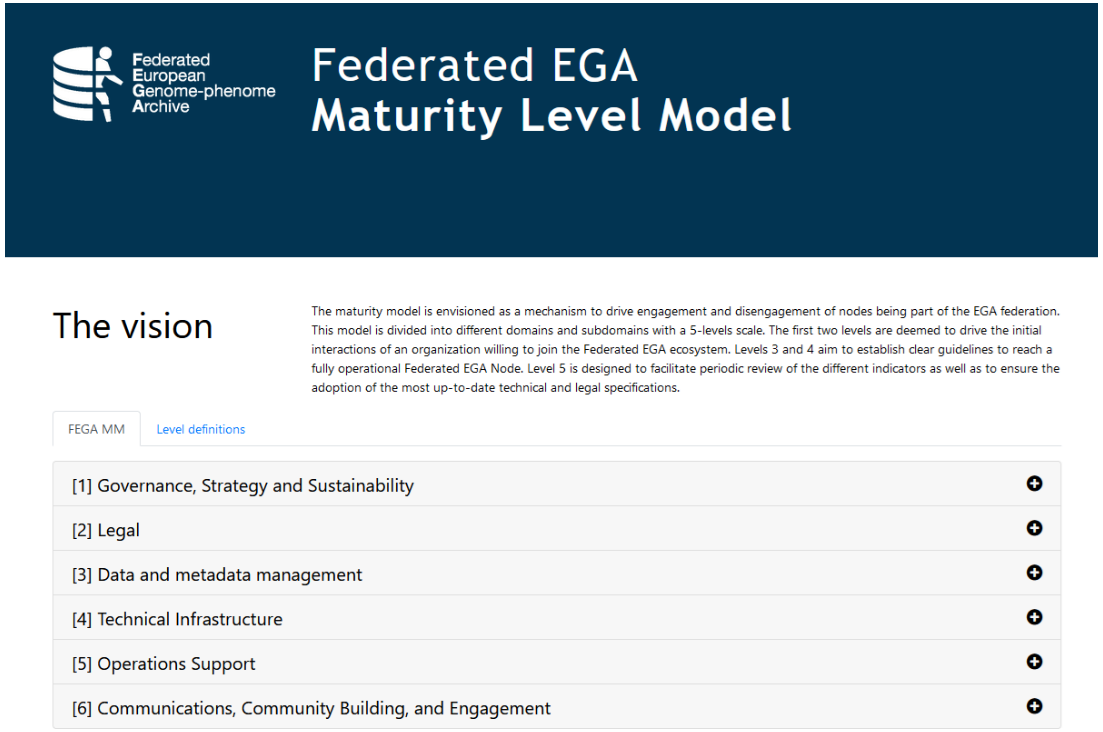
FEGA Node / Access Point

- The creation of a node is associated with the necessary technical and process competence.
- It is not only about creating a node, but also about ensuring its sustainability.
- Organizational and legal aspects are directly covered by FEGA/Elixir:
  - https://ega-archive.github.io/FEGA-onboarding/
- Each node has the following tasks:
  - implement the EGA ETL pipeline (Extract, transform, and load),
  - provide helpdesk to users and DACs,
  - provide user experience according to FEGA SOPs, data access services,
  - export metadata to CEGA,
  - contribute to the development of common APIs, tools, and resources, participate in strategic committee meetings, provide CEGA with summary operational reports.
- SensitiveCloud to be used for FEGA node hosting.




Consistent framework
Implementing the GA4GH suite of products
Compliant with ethical, legal and social implications (ELSI)
Common API interfaces
EMBL-EBI
CRG
Labs
Individual submitter
Biobanks
Hospitals
Projects
ELSI
GA4GH
API
Encrypted data stay in the country of origin
Federated EGA Node
Metadata travel across borders
Central EGA
Deploying nodes
Members

FEGA Node / Access Point



Federated European Genome-phenome Archive

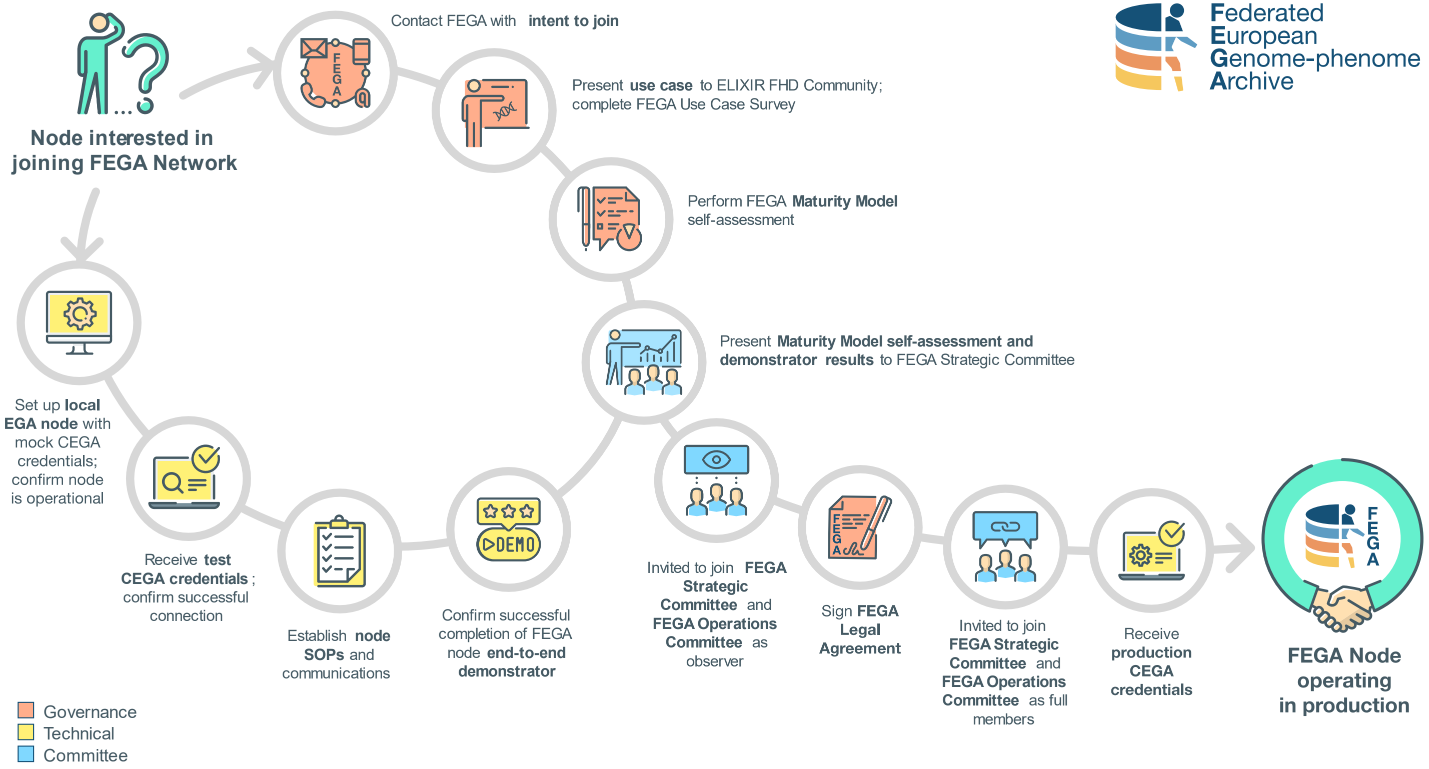
Federated EGA Maturity Level Model

The vision

The maturity model is envisioned as a mechanism to drive engagement and disengagement of nodes being part of the EGA federation. This model is divided into different domains and subdomains with a 5-levels scale. The first two levels are deemed to drive the initial interactions of an organization willing to join the Federated EGA ecosystem. Levels 3 and 4 aim to establish clear guidelines to reach a fully operational Federated EGA Node. Level 5 is designed to facilitate periodic review of the different indicators as well as to ensure the adoption of the most up-to-date technical and legal specifications.

FEGA MM | Level definitions

- [1] Governance, Strategy and Sustainability
- [2] Legal
- [3] Data and metadata management
- [4] Technical Infrastructure
- [5] Operations Support
- [6] Communications, Community Building, and Engagement



Federated European Genome-phenome Archive

Node interested in joining FEGA Network

Contact FEGA with **intent to join**

Present **use case** to ELIXIR FHD Community; complete FEGA Use Case Survey

Perform FEGA **Maturity Model** self-assessment

Present **Maturity Model self-assessment and demonstrator results** to FEGA Strategic Committee

Set up **local EGA node** with mock CEGA credentials; confirm node is operational

Receive **test CEGA credentials**; confirm successful connection

Establish **node SOPs** and communications

Confirm successful completion of FEGA node **end-to-end demonstrator**

Invited to join **FEGA Strategic Committee** and **FEGA Operations Committee** as observer

Sign **FEGA Legal Agreement**

Invited to join **FEGA Strategic Committee** and **FEGA Operations Committee** as full members

Receive **production CEGA credentials**

FEGA Node operating in production

🟧 Governance
🟨 Technical
🟦 Committee

- https://ega-archive.github.io/FEGA-onboarding/
- https://ega-archive.github.io/FEGA-onboarding/topics/maturity-model/
- https://github.com/EGA-archive/FEGA-onboarding

Summary

SensitiveCloud Infrastructure

- SensitiveCloud up and running.
  - ISO 27k certified.
  - Onboarding process, processes, technical measures, service agreement template, …
- Continuous process of improvements.
  - Projects in progress, FEGA.
  - Thinking of SensitiveCloud 2.0, 3.0, …

1.

We find out whether SensitiveCloud fulfils your requirements.

2.

We explore all the specifics of your use case and, if necessary, make appropriate changes.

3.

We sign an agreement confirming our duties and expectations.

4.

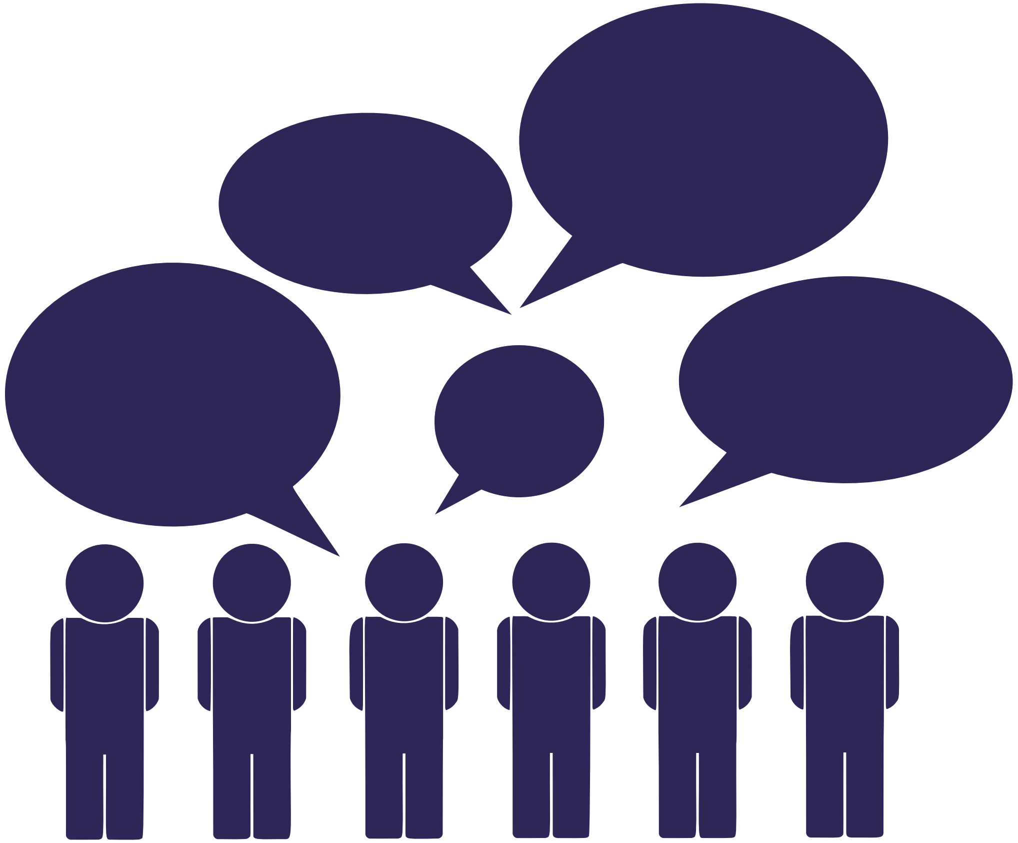
We train you to harness the most of the environment.

5.

You work on your research, and we continuously monitor your needs and satisfaction.

https://www.cerit-sc.cz/infrastructure-services/sensitivecloud

Discussion



Source: Communicate_communication_conference_2028004 od OpenClipart-Vectors z Pixabay