



Enabling access to large-scale European life science data resources for AI development

Tommi Nyrönen CSC
For Nordunet conference Luleå
8.9.2026

elixir-europe.org

-
- A map of Europe with the 15 member states of the European Union in 1993 highlighted in orange. The map includes the United Kingdom, Ireland, France, Germany, Italy, Spain, Portugal, Greece, and the Benelux countries. Each country has a small orange dot indicating its location. The map also shows the surrounding landmasses of North Africa, the Middle East, and Iceland, which is not a member state.



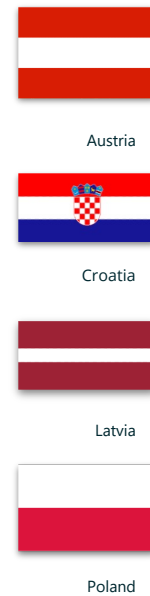
ELIXIR is a sustainable European infrastructure for biological data, est. 2014



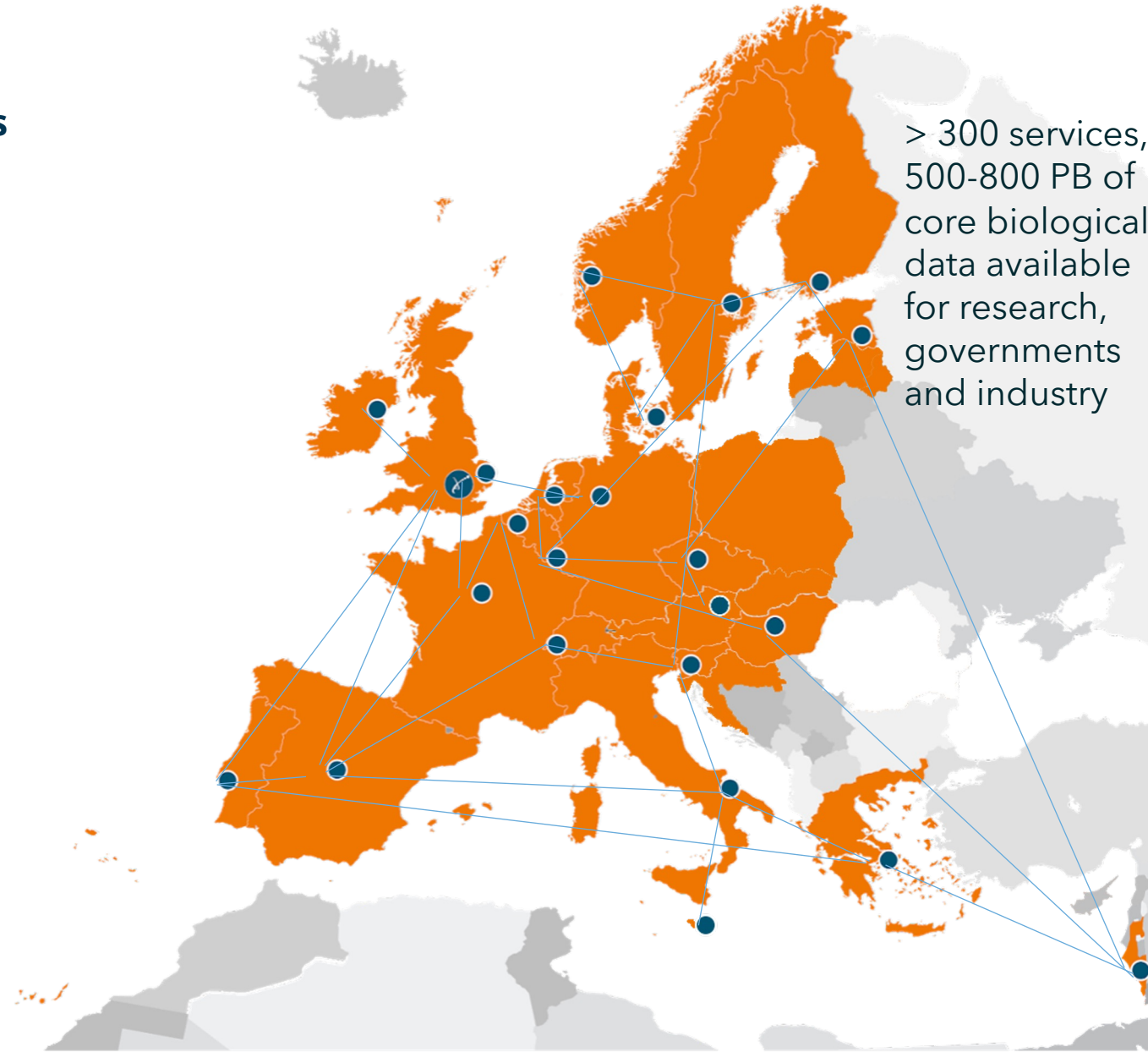
Members



Observers



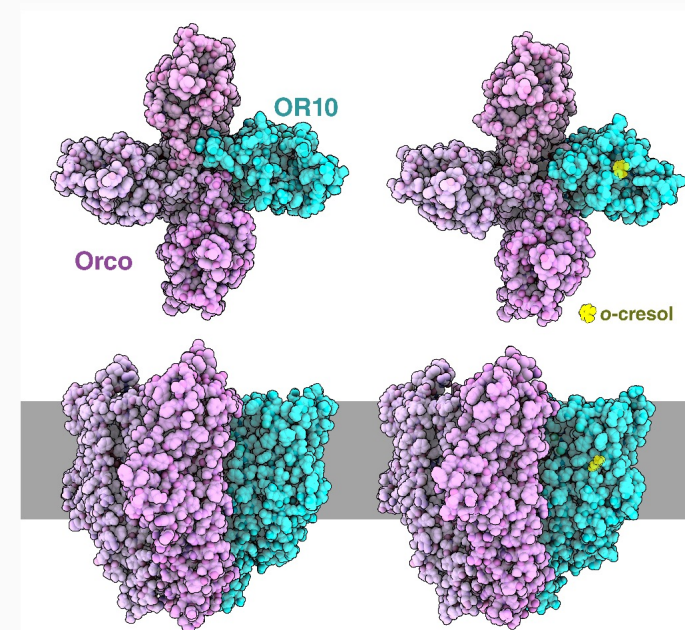
> 300 services,
500-800 PB of
core biological
data available
for research,
governments
and industry



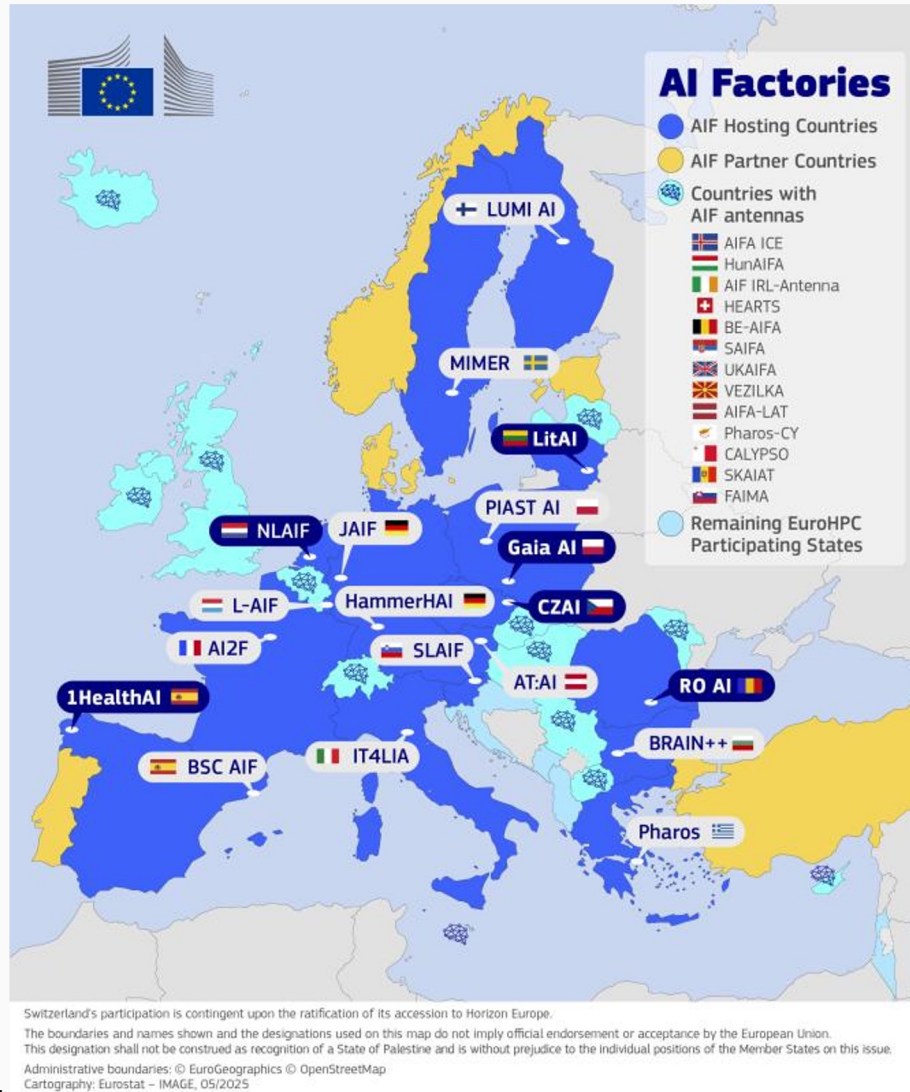
Nobel 2024 and ELIXIR data

- **Protein Data Bank** (PDB, ELIXIR core data resource): Experimental 3D structures used as the primary training targets and validation benchmarks
- **UniProt** (SIB, ELIXIR core data resource): Curated protein sequences and functional annotations mapped evolutionary relationships
- **MGnify** (EMBL-EBI, ELIXIR core data resource): Microbial sequence data expanding the evolutionary dataset to uncultured and unknown organisms
- *AlphaFold DB* (EMBL-EBI & Deepmind) New infrastructure built to distribute DeepMind's AI-predicted structures

“Demis Hassabis and John Jumper have developed an AI model to solve a 50-year-old problem: predicting proteins’ complex structures.”



Leveraging European e-Infrastructure in Life Science?



Demand for service: *large-scale computational data access* lacks service experts and investment

Life sciences tend to stretch EuroHPC/AI Factory data handling capability and usage policies

- **Sensitive data (e.g. from human subjects - EHDS)** needs to be integrated into European supercomputer ecosystem to optimise broader EU-level long-term investment
- **Life Science datasets** need to exist for research project duration in logical/physical proximity of AI Factory computing resources

Talent integrating across HPC-Quantum & AI & DATA (biomolecular)

- Teams to empower SMEs and academia to collaborate in supercomputing and life science data.



LUMI AI Factory



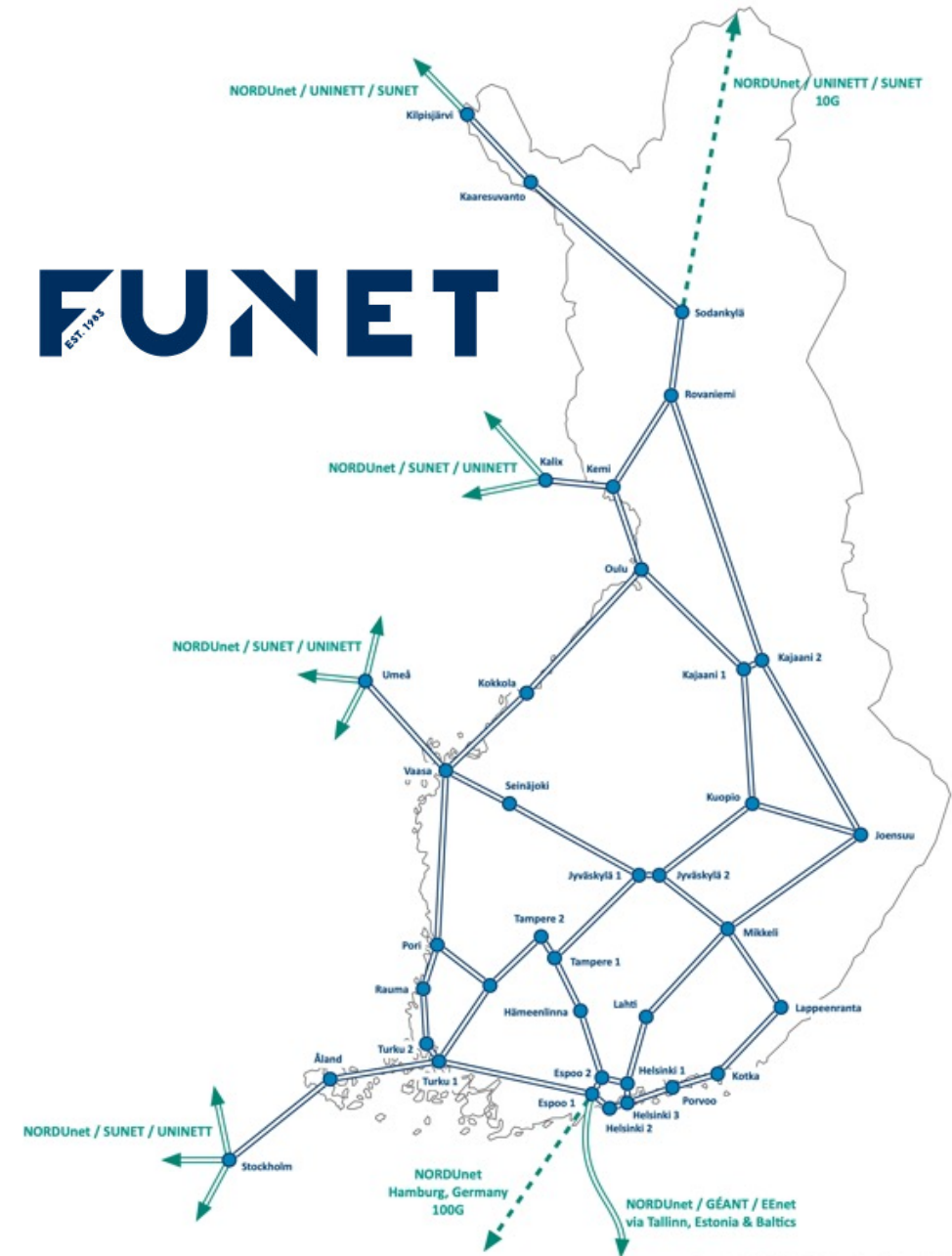


The three pillars of LUMI AI Factory

1. AI-optimised supercomputer **LUMI-AI**
2. Expert support by **LUMI AI Factory Service Center**
3. Experimental quantum-computing platform **LUMI-IQ**

Reliable and secure high-speed data communications

- LUMI ecosystem in CSC Kajaani datacentre is directly connected to the Nordic network backbone
- Scalability: multi-terabit transmission capability
- Hyper Connectivity Service Area (HCSA) node
 - e.g. **secure data processing zones** for EHDS/GDPR compliance
- The next-generation NORDUnet connects site to GÉANT, ensuring European-wide service availability
- **Key for enabling data-intensive Life Science**



Key industrial sectors and focus areas

Manufacturing
industries

Health and life
sciences

Communication
technologies
and networks

Data
ecosystems



Sensitive
data

Computing
continuum,
dynamic data

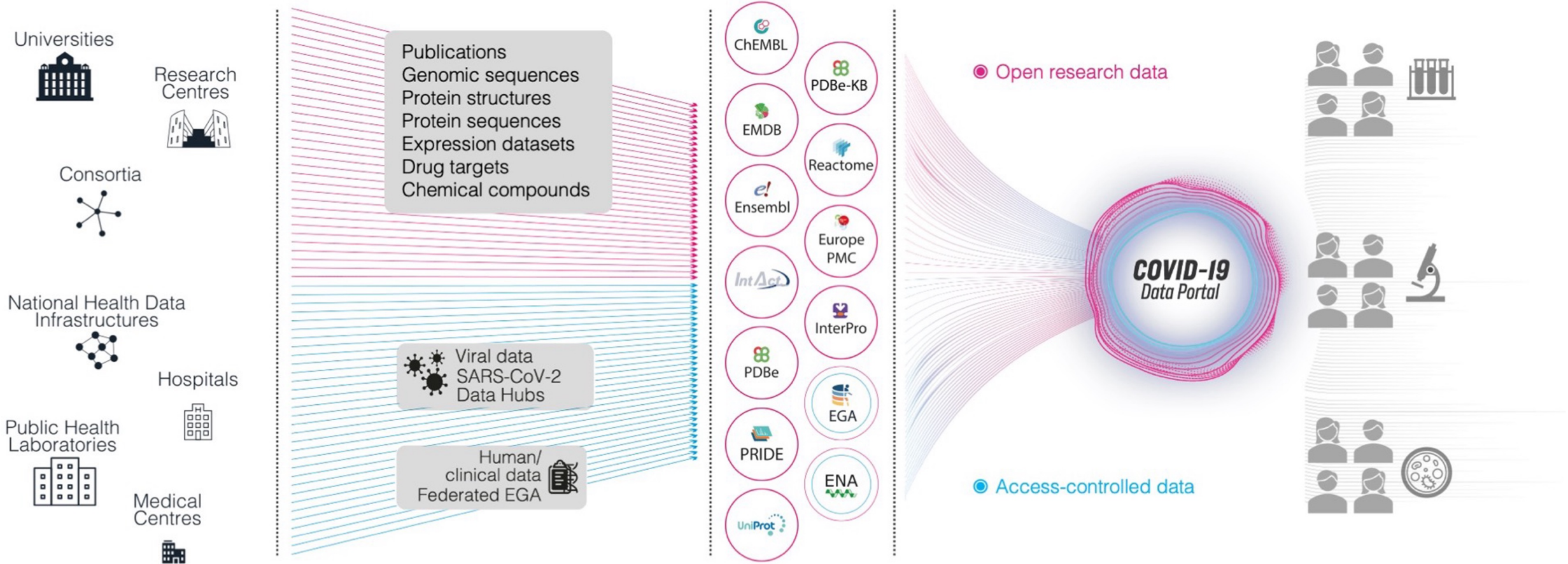
Simulation & AI

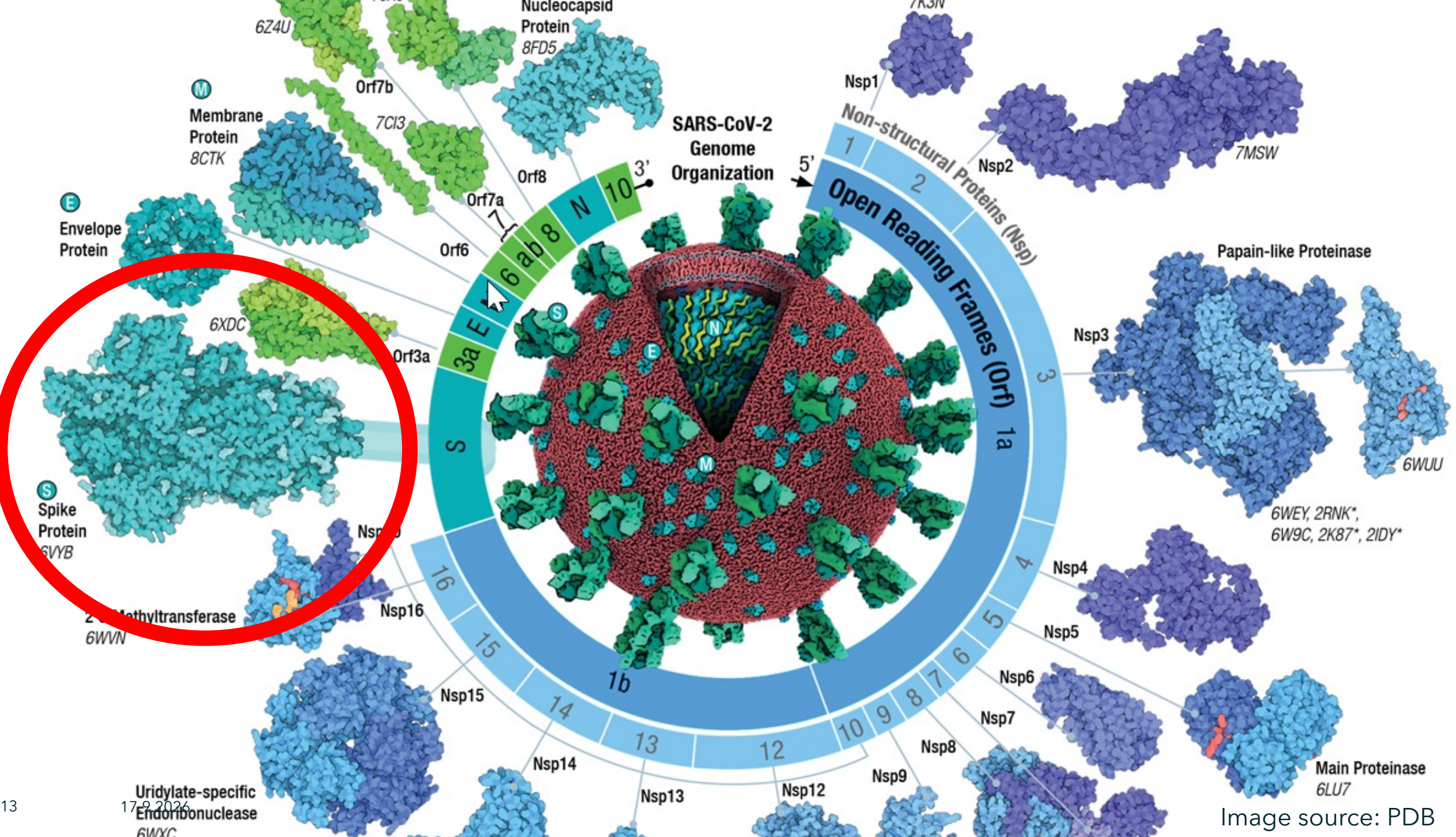
AI startups, agile projects

Digital twins and "AI
for science"

Ease of use and
streamlined access

The European COVID-19 Data Management





Bringing life science data resources to AI Factories

- Data availability is key – *no datasets, no computational analysis*

- Back in 2020
took about 1
time to anal
A dataset co
occupy 100
tens of millio

- Data manag
a large exte

- Network performance is not the bottleneck!

- Case study: <https://www.lumi-supercomputer.eu/empowering-pandemic-preparedness-with-european-covid-19->

**EU27 prevention spending rose from
~€51 bn (2020) to ~€90.4 bn (2022)
during COVID-19 according to
OECD/Eurostat.**

LUMI

EMBL-EBI 

elixir 





From EU strategy to secure high-performance data access

EU data union and life science strategy

Data Union (enabler) priorities

- **Scale up access to data for AI**
- Streamlining data rules
- Strengthen global position on data flows

Life Science (demand) strategy

- Leading global hub for R&D by 2030
- End-to-End value chain approach
- **Turn breakthroughs into products**

Data Labs: Service providers linking data spaces to AI ecosystems.

Action: ELIXIR in the EU AI ecosystem, focusing on Health & Life Science data know-how in the ELIXIR nodes

Possible high-level scientific use cases to help drive technology (ELIXIR)

EU RAISE: AI-Driven Climate Adaptation for the European Agriculture and Food Cluster

- Can EU AI resources support federated AI computing workflows on **non-sensitive** but technically complex agricultural, environmental, and Earth observation datasets?
- Expected gaps include metadata immaturity, interoperability frictions, and insufficiently productised data services

Diagnosis Market: Federated Discovery and Validation of Blood-Based Epigenetic Biomarkers

- Can EU AI resources support secure and scientifically robust management and analysis of **sensitive** biomolecular health data?
- Expected gaps: heterogeneous data governance, incomplete federation, limited safe outputs, and mismatch between compliance and practical usability

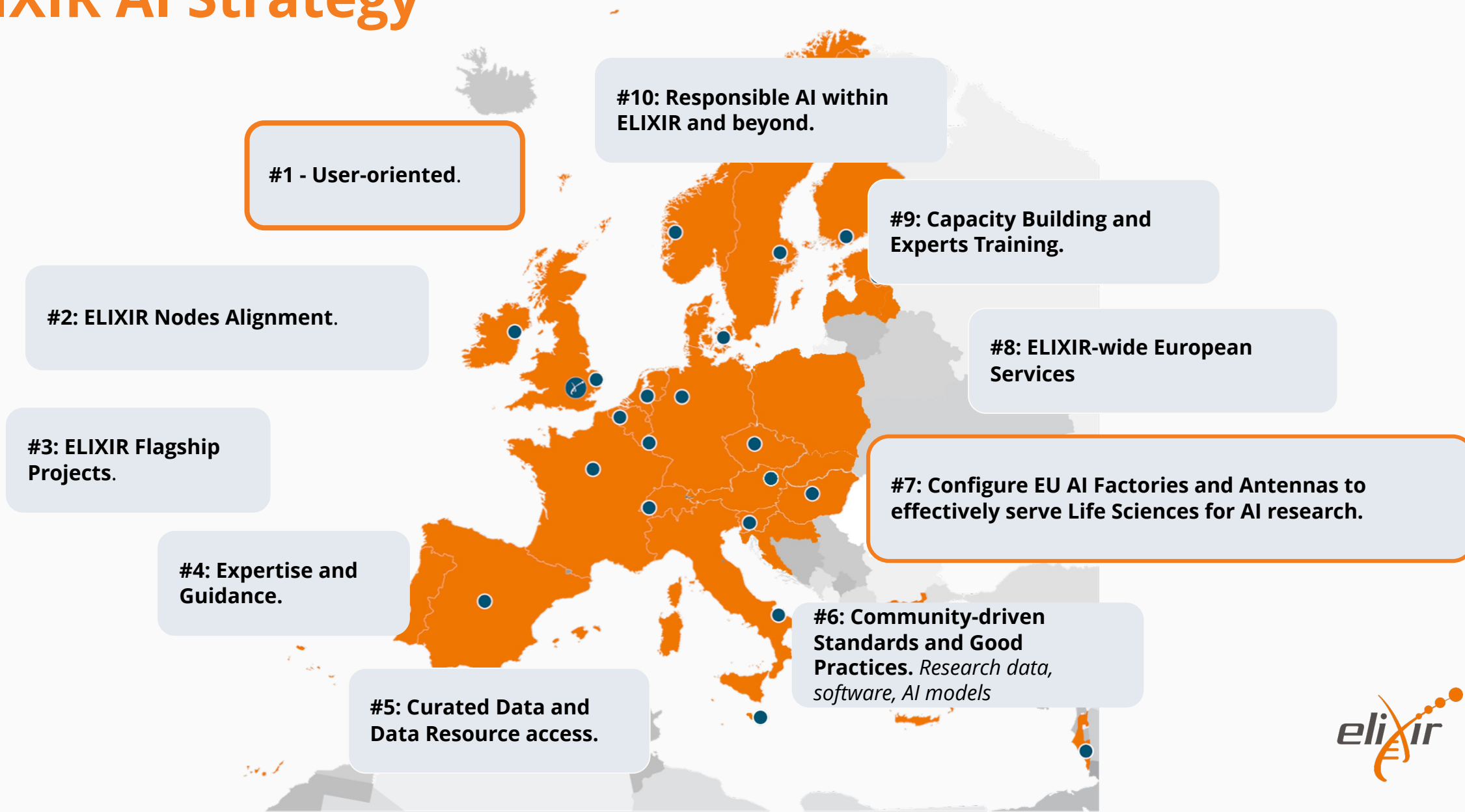
EU missions – e.g. Cancer data federated learning on distributed datasets

- Combination of federated AI computing workflows and sensitive data management to understand molecular mechanisms and improve diagnosis accuracy and speed.

Security and confidentiality by design; Trustworthy AI by design

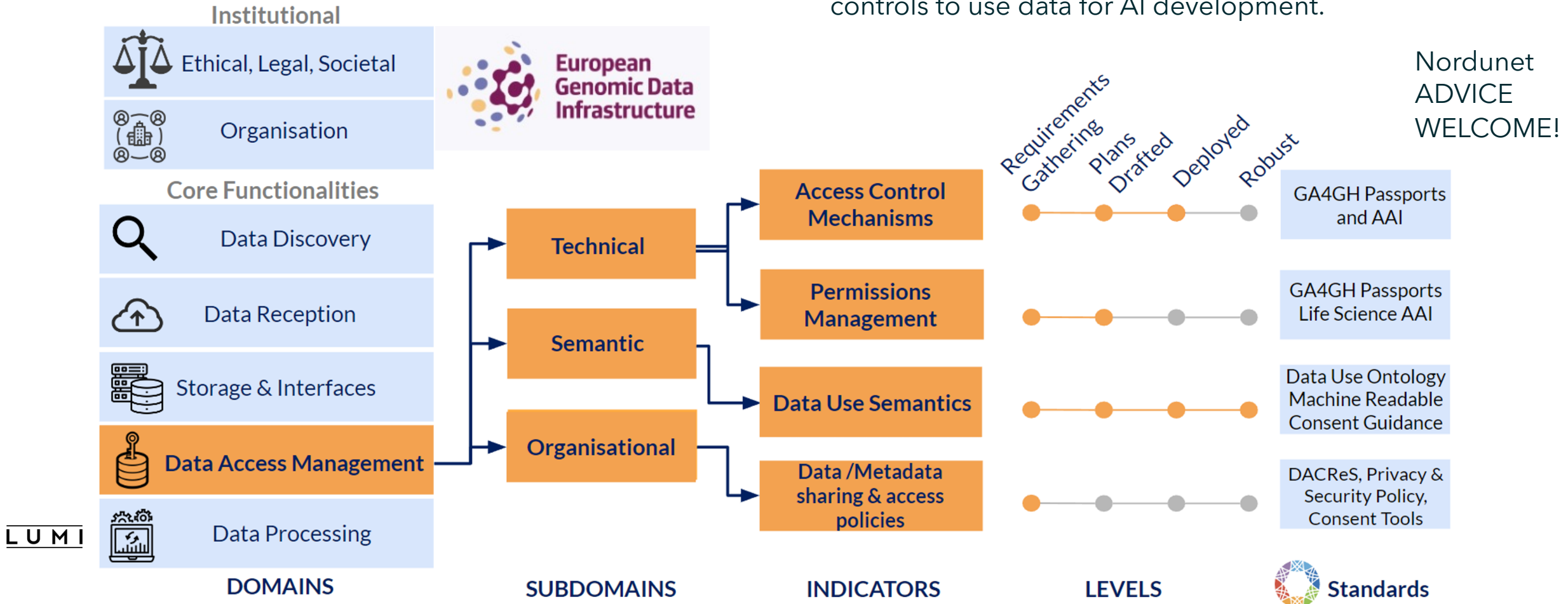
Regulatory: AI Act, GDPR, EHDS

ELIXIR AI Strategy



Investments ongoing in Sensitive data management

Health Data generated from human subjects is typically subject to GDPR, EHDS, AI Act. We need a legally compliant secure cross-border data infrastructure and standards-compliant data management - including storage, transfer, and access controls to use data for AI development.



Common AAI needed for Life Science data services

- Global Alliance for Genomics and Health **GA4GH Passport** consist of Visas that are signed Java Web Token retrievable from an Open ID Connect endpoint. This standard format implemented across ecosystem enables implementation of European Genomic Data Infrastructure data governance workflows.



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



In addition to reliable Identity (e.g. MyAccessId as an authentication source in LS Login operated by ELIXIR), Life Science data access also requires a common AAI with **versatile authorisation attributes** that encode data access authority decisions.

- AffiliationAndRole** - role within the user's affiliated institution
- AcceptedTermsAndPolicies** - user accepted specific terms, policies, conditions or meets particular criteria (e.g. ethics compliance)
- ResearcherStatus** - the user has been acknowledged to be a researcher of a particular type or standard
- LinkedIdentities** - assertions of differing user identifiers to identify the same person
- ControlledAccessGrants** - grants of controlled access to data sets

DATA ACCESS METADATA ATTRIBUTES FOR ORCHESTRATING DATA ACCESS

Conclusion

Without collaborative improvement to European health and life science data access capability, AI will not work as intended.

Strategic Opportunities and Structural gaps

- **European health and life-science data generation is substantially larger** than current European data and compute infrastructure can absorb
 - Fragmented data stewardship
- Most **health data are produced in hospitals** rather than in research
 - Limited sensitive-data readiness
- AI Factories strengthen European **AI capability but...life-science data currently stretch their data access handling capability, policy, and service models**
 - Uneven science data-to-compute proximity
- Coordinated improvement of **high-quality datasets as a service near AI/HPC** capacity.
 - Talent gaps across data, AI, and HPC to enable science breakthroughs
- ELIXIR is mapping AI actions to its 5-year programme timeframes: **2027-28** and **2029-33** (2 + 5 years)



Thanks!



Tommi Nyrönen CSC

nyronen csc.fi

[https://www.linkedin.com
/company/elixir-finland](https://www.linkedin.com/company/elixir-finland)

Follow us

[LinkedIn](#)

[Instagram](#)

[Facebook](#)

[YouTube](#)

[csc.fi](#)