



# ELIXIR

EDITH meeting , May 2023

[www.elixir-europe.org](http://www.elixir-europe.org)



# ELIXIR – what do we do

We build life science informatics **capacity and infrastructure** in Europe, connect and develop a **network of experts** and provide hundreds of high quality **services and resources** available to all



Databases



Training



Software tools



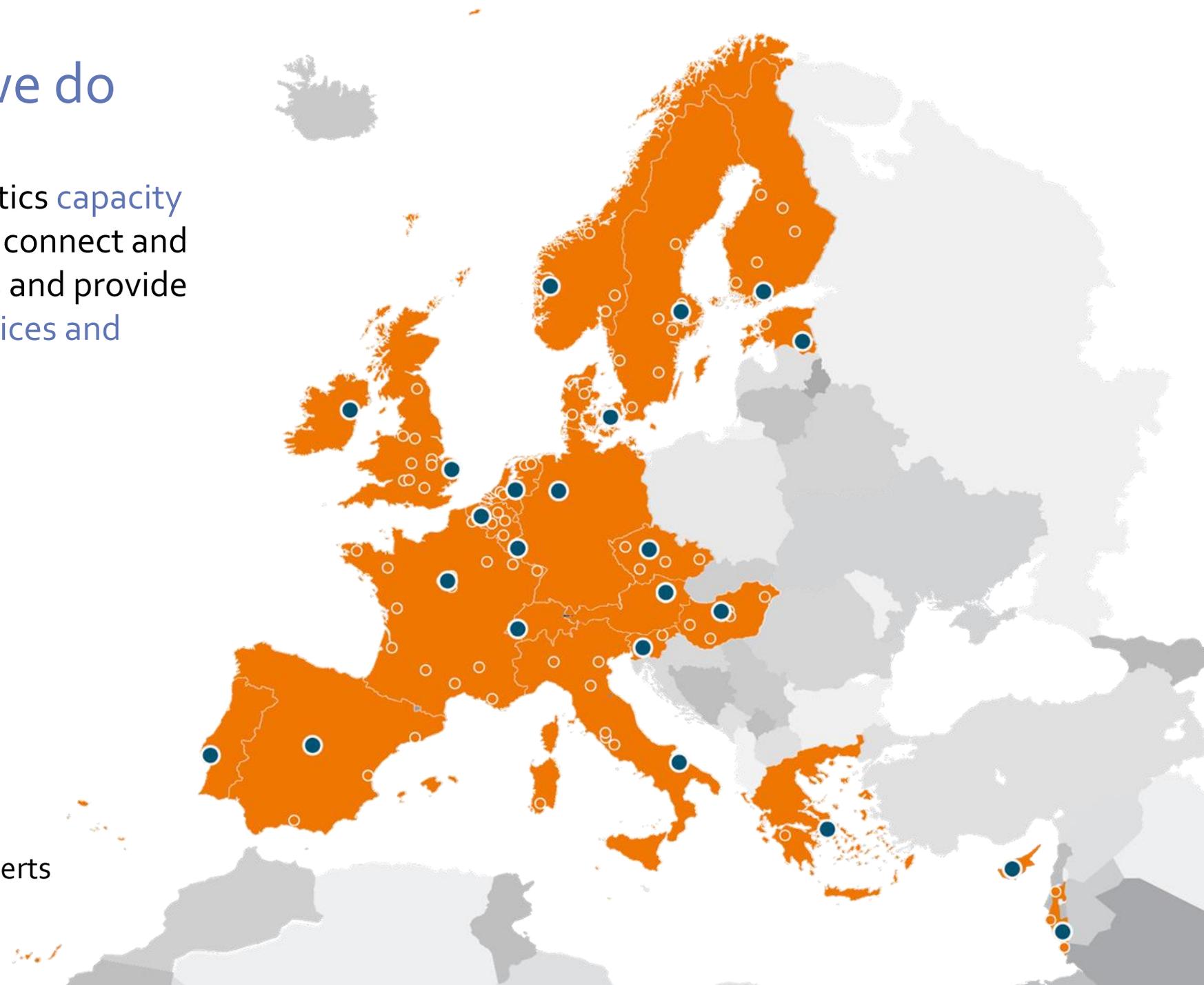
Data standards



Compute resources



Scientific & technical experts



# Resources to access knowledge and curated digital objects

bio.tools



bio.tools helps you find and select bioinformatics software and connect it in workflows.

BioContainers



Search a repository of containerised software that you can build into workflows.

WorkflowHub



A registry for sharing and publishing scientific computational workflows.

FAIRsharing.org



FAIRsharing.org allows you to search for databases and data policies by aspects such as domain, species and country.

TeSS



Search for training courses, webinars, training materials and workflows in TeSS, ELIXIR's training portal.

COMMENT | FOCUS

## DOME: recommendations for supervised machine learning validation in biology

DOME is a set of community-wide recommendations for reporting supervised machine learning-based analyses applied to biological studies. Broad adoption of these recommendations will help improve machine learning assessment and reproducibility.

Ian Walsh, Dmytro Fishman, Dario Garcia-Gasulla, Tiina Titma, Gianluca Pollastri, ELIXIR Machine Learning Focus Group, Jennifer Harrow, Fotis E. Psomopoulos and Silvio C. E. Tosatto

With the steep decline in the cost of many high-throughput technologies, large amounts of biological data are being generated and made accessible to researchers. Machine learning (ML) has come into



Check for updates

OPEN ACCESS

Citation: Brack P, Crowther P, Soiland-Reyes S, Owen S, Lowe D, Williams AR, et al. (2022) Ten

NATIONAL BIOLOGY

EDITORIAL

## Ten simple rules for making a software tool workflow-ready

Paul Brack<sup>1</sup>, Peter Crowther<sup>2</sup>, Stian Soiland-Reyes<sup>1,3\*</sup>, Stuart Owen<sup>1</sup>, Douglas Lowe<sup>4</sup>, Alan R. Williams<sup>1</sup>, Quentin Groom<sup>5</sup>, Mathias Dillen<sup>6</sup>, Frederik Coppens<sup>6,7</sup>, Björn Grüning<sup>8</sup>, Ignacio Eguinoa<sup>6,7</sup>, Philip Ewels<sup>9</sup>, Carole Goble<sup>1</sup>

1 Department of Computer Science, The University of Manchester, Manchester, United Kingdom, 2 Melandra Limited, Stockport, United Kingdom, 3 Informatics Institute, University of Amsterdam, Amsterdam, The Netherlands, 4 Research IT, IT Services, University of Manchester, Manchester, United Kingdom, 5 Meise Botanic Garden, Meise, Belgium, 6 Department of Plant Biotechnology and Bioinformatics, Ghent University, Ghent, Belgium, 7 VIB Center for Plant Systems Biology, Ghent, Belgium, 8 Bioinformatics Group, Department of Computer Science, Albert-Ludwigs-University Freiburg, Freiburg, Germany, 9 Science for Life Laboratory (SciLifeLab), Department of Biochemistry and Biophysics, Stockholm University, Stockholm, Sweden

\* soiland-reyes@manchester.ac.uk

### Introduction

In recent years, the volumes of data to be analyzed, as well as the complexity of that analysis, across many scientific fields (from genomics through to exoplanet exploration) have increased

## SCIENTIFIC DATA

Amended: Addendum

OPEN

CATEGORIES  
» Research data  
» Publication characteristics

## Comment: The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson *et al.*\*

10 December 2015  
12 February 2016  
15 March 2016

There is an urgent need to improve the infrastructure supporting the reuse of scholarly data. A diverse set of stakeholders—representing academia, industry, funding agencies, and scholarly publishers—have come together to design and jointly endorse a concise and measurable set of principles that we refer to as the FAIR Data Principles. The intent is that these may act as a guideline for those wishing to enhance the reusability of their data holdings. Distinct from peer initiatives that focus on the human scholar, the FAIR Principles put specific emphasis on enhancing the ability of machines to automatically find and use the data, in addition to supporting its reuse by individuals. This Comment is the first formal publication of the FAIR Principles, and includes the rationale behind them, and some exemplar implementations in the community.

### Supporting discovery through good data management

Good data management is not a goal in itself, but rather is the key conduit leading to knowledge discovery and innovation, and to subsequent data and knowledge integration and reuse by the



RDM support throughout the entire **life cycle of projects** as outlined in **DMPs**

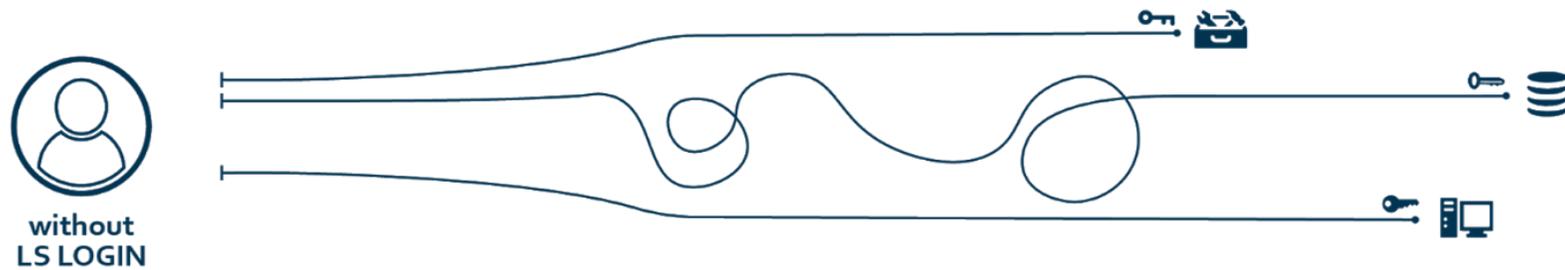


**Online focal point** for guidance, information, best practice, examples

Context and **signpost** for FAIR data resources as a Hub for a RDM Knowledge Commons



# Life Science Login as a single-entry point



time spent

Originally ELIXIR AAI

Co-developed through EOSC-Life Project (WP5) into current Life Science Login

Applicability for federated authentication and access management for LS services

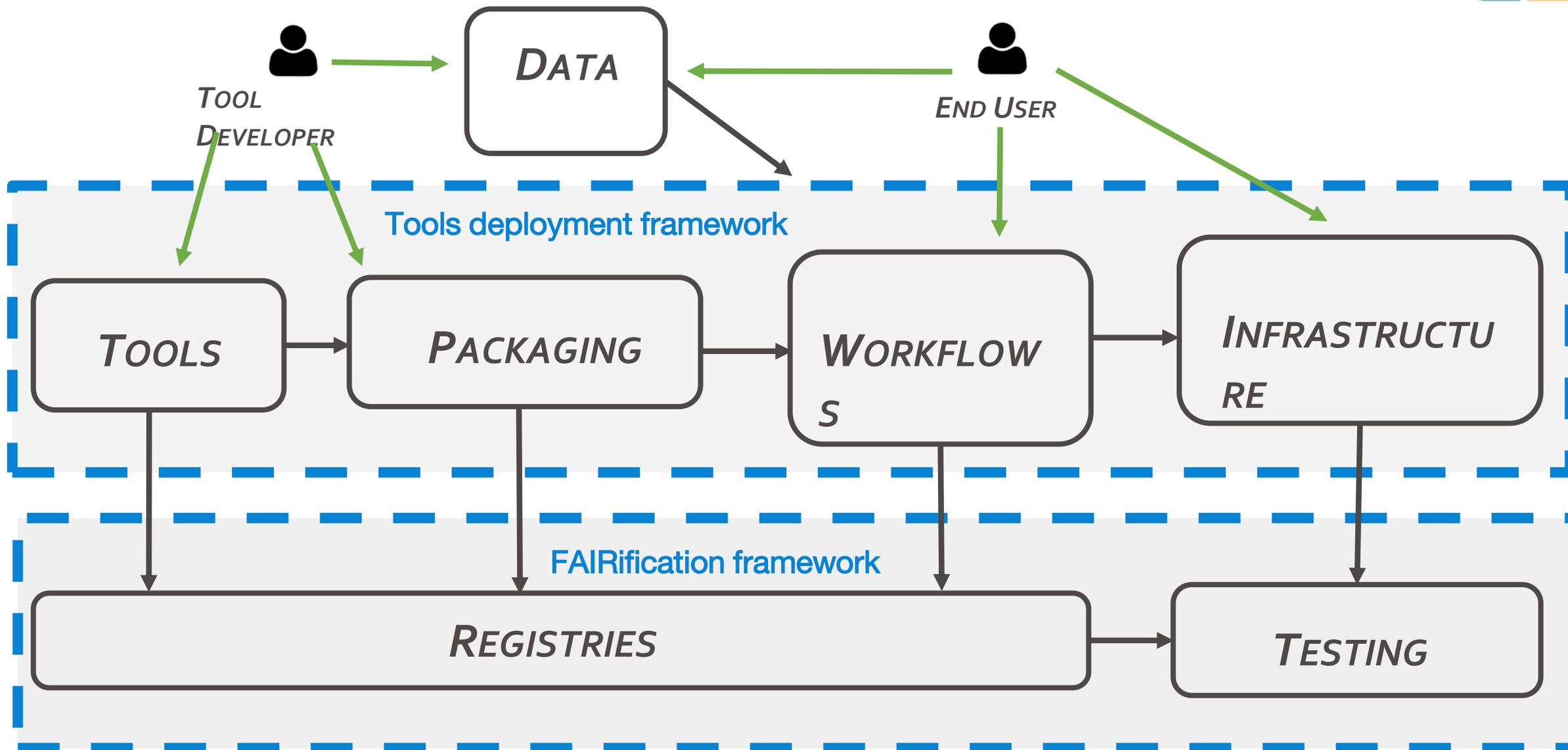
Use case example: Federated EGA - Norway



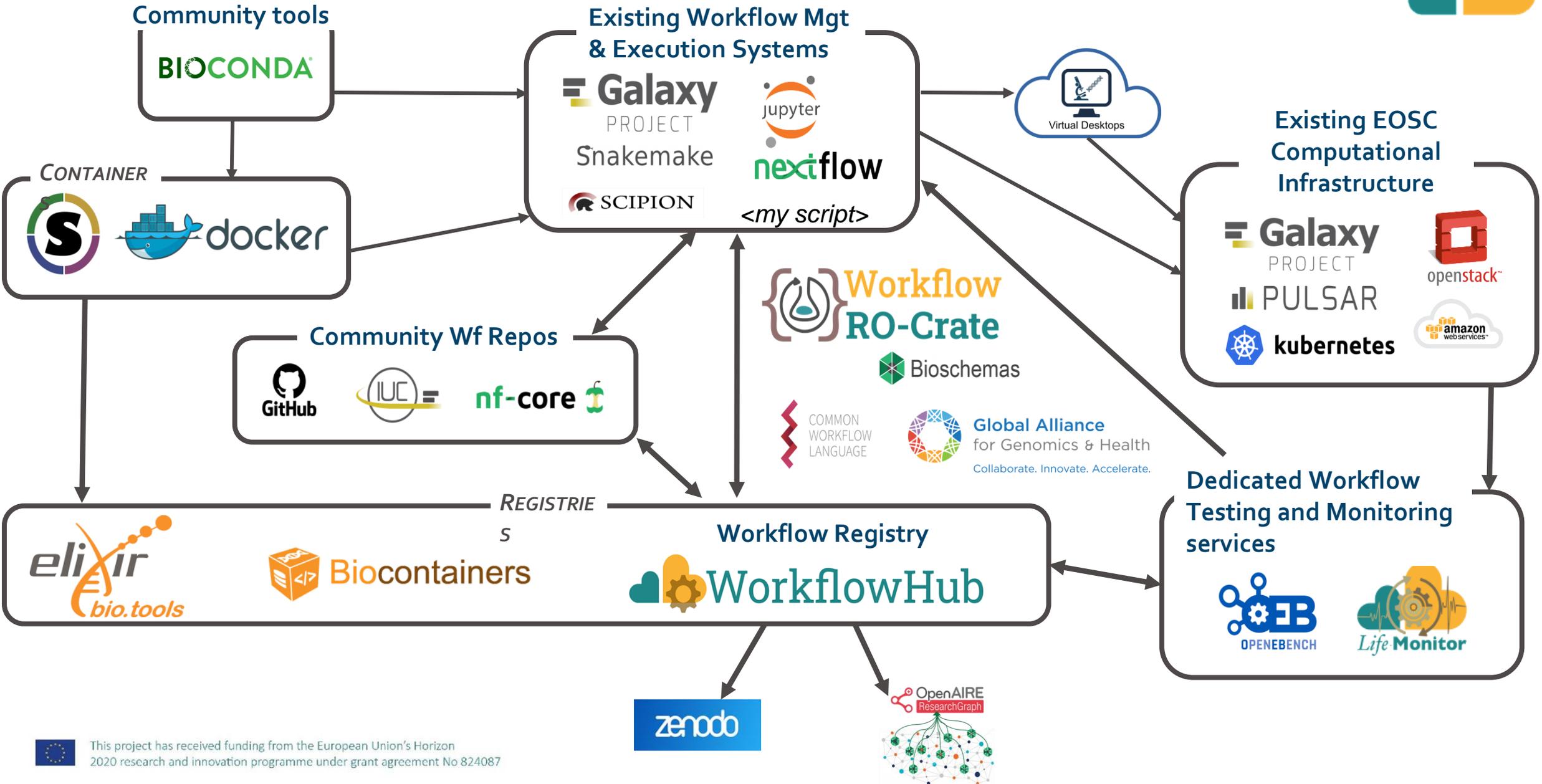
[Docs and Info](#) [Privacy Policy](#) [Terms of Service](#) [Contact Us](#) [About](#)

## Federated EGA Norway node

LS LOGIN



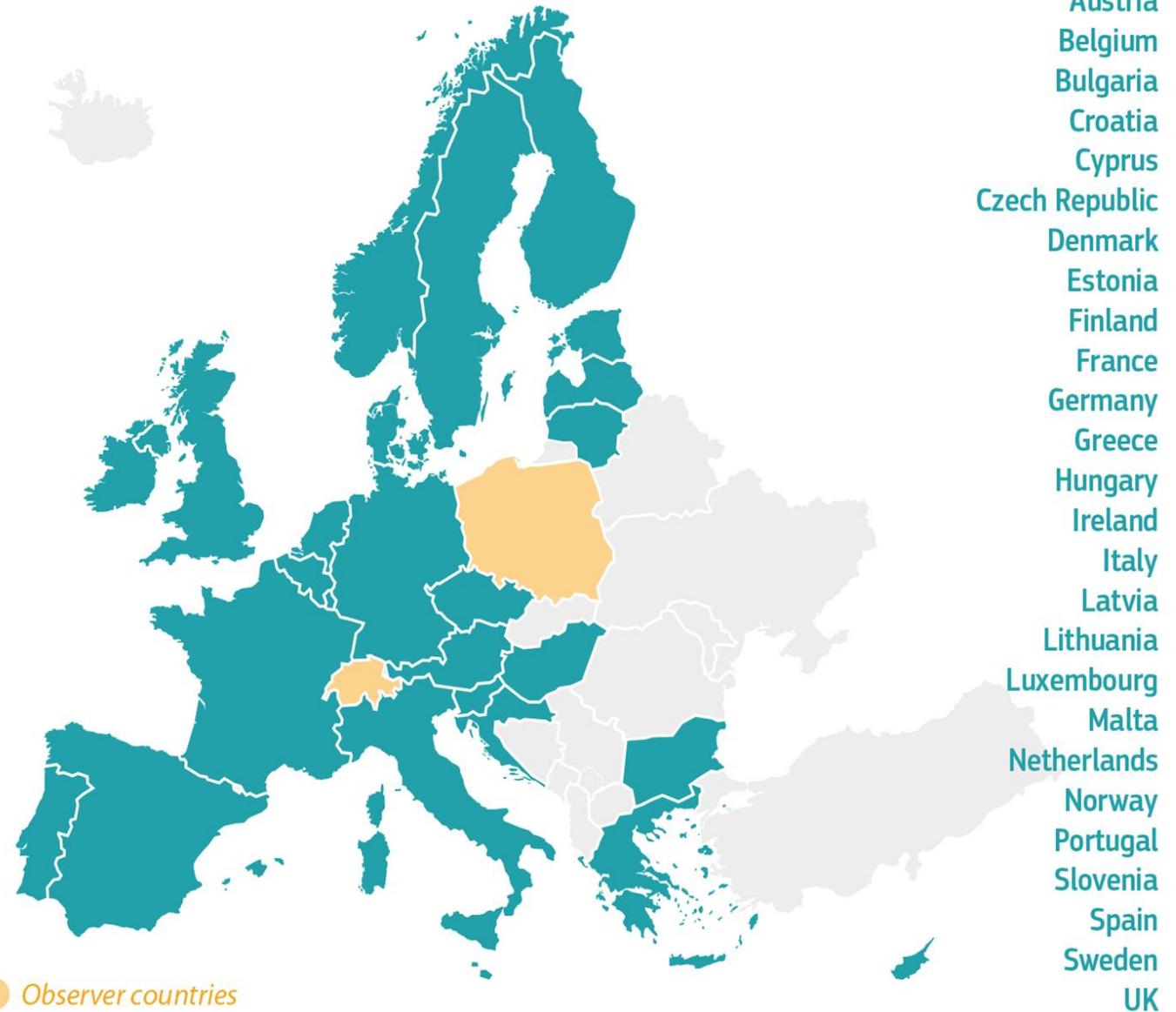
# The EOSC virtual research environment for the life sciences





## 1+MG initiative

“1+MG facilitates countries to 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”



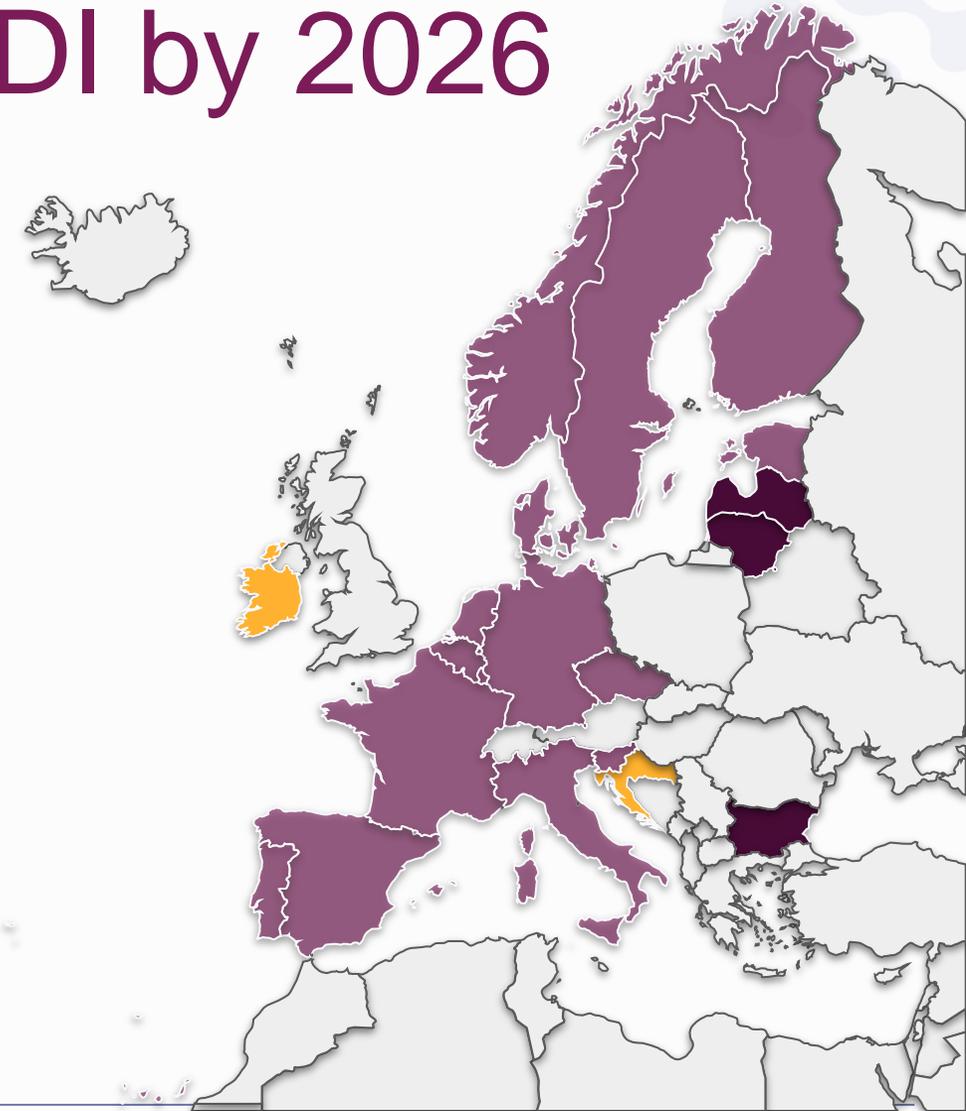
updated on 16 November 2022

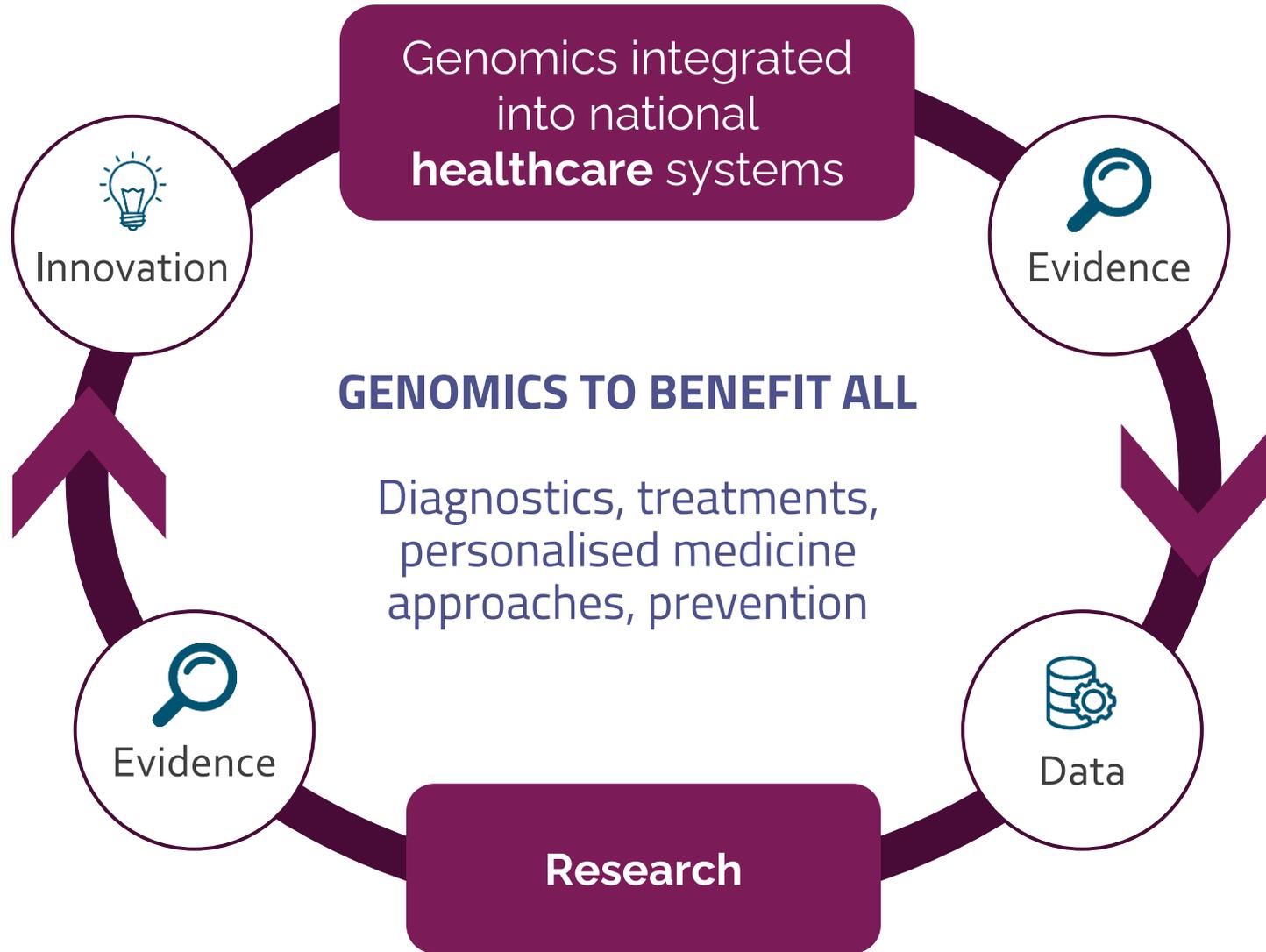


# Countries commitment to GDI by 2026

- Fully operational and integrated into 1+MG infrastructure: **Belgium, Czech Republic, Denmark, Estonia, Finland, France, Germany, Italy, Luxembourg, Portugal, Slovenia, Spain, Sweden, The Netherlands, Norway**
- Fully operational national node but not yet integrated in the 1+MG infrastructure: **Bulgaria, Latvia, Lithuania**
- Onboarding: **Croatia, Ireland** *Cyprus, Hungary, Malta, Romania,*

Infrastructure will exist in 2024 with at least 6 countries technically operational







# A user toolkit for federated research data



Deposit

Archive & Publish

Find Data

Negotiate  
Access

Remote  
analysis



Data  
reception



Storage and  
interfaces



Data  
discovery



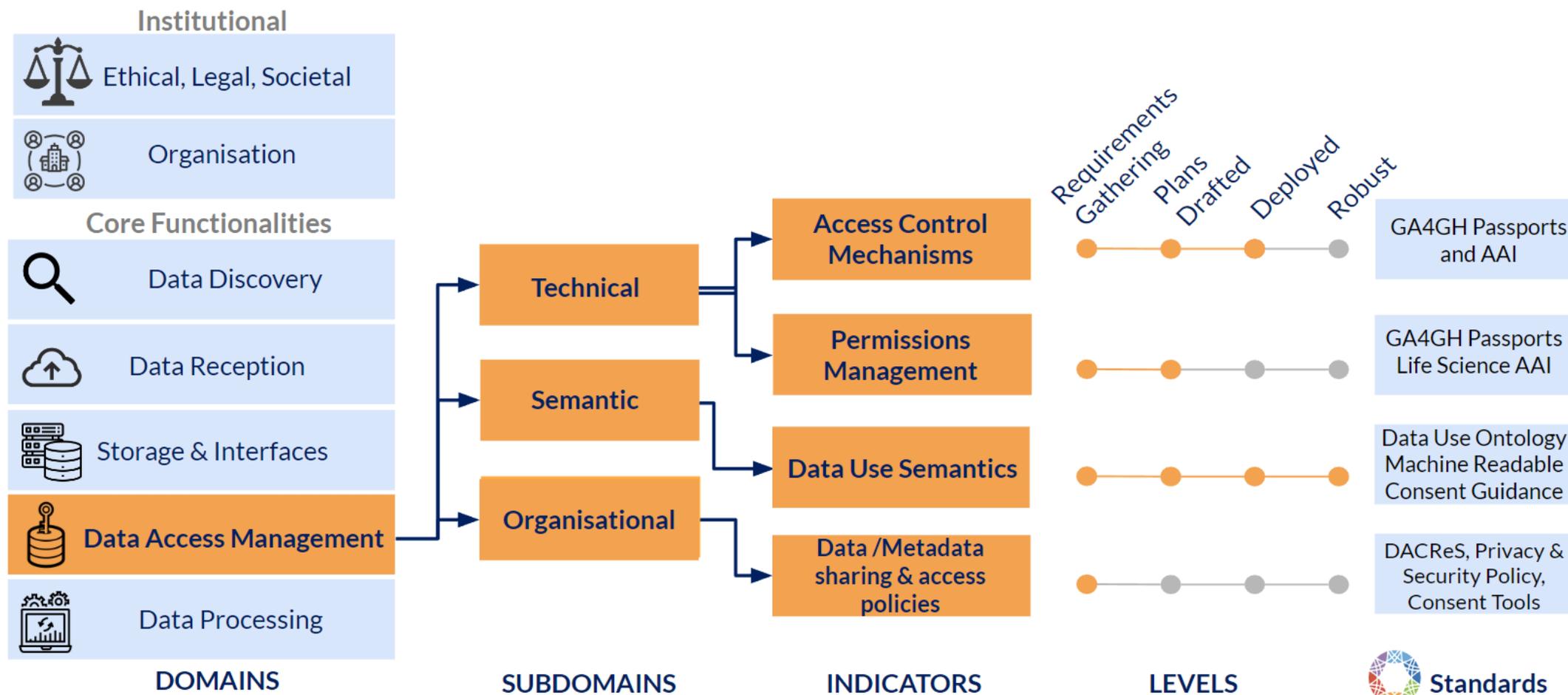
Access management  
tools



Data  
processing



# ELIXIR Human Data Infrastructure Maturation Model





# GDI Starter Kit - launch in June 2023

Product	Outline	Functionality
Sensitive Data Archive	Securely stores data (FEGA technologies)	
LifeScience AAI	Provides a federated Identity	
REMS	Tool to allow data access applications and decisions	
Beacon	Genetic and phenotypic data discovery	
Beacon Network	Federated network of Beacons	
Synthetic Data	Realistic synthetic anonymous data	
htsget	Secure genetic data distribution standard	
Containerised Computation	Computation via containers, e.g docker or singularity	
Federated Computation	Federated workflows, e.g. Nextflow	
Packaging and Deployment	Packaging and deployment of the starter kit	
User Portal – Data Catalogue	European level catalogue of data within deployed nodes	
User Portal – Access management	European level data application and access management tool	

External to the starter kit

Node Connection

# What we offer



## Guidelines

Guidelines and best practices to help you manage life science data, run training courses, develop software and more.



## Web portals

Find the right software, training courses, standards and more in our interlinked portals to life science resources.



## Services

Find compute services, databases, and the full list of resources ELIXIR coordinates.



## Partnerships with Industry and SMEs

Join events and projects that bring the private and public sectors together.



## Opportunities to work together

Join a scientific group in ELIXIR or partner with us to apply for EC funding.



## For ELIXIR members

If you work at an institute that is part of ELIXIR, then remember to take advantage of the benefits ELIXIR offers.

ELIXIR coordinates the provision of life science services and resources, developed and managed by ELIXIR Nodes

The services and resources are freely available to researchers around the world

The services and resources support efficient manipulation, analysis, storage and exchange of life science data