



# ELIXIR Europe

*Data as working capital in the science enterprise*



Niklas Blomberg, ELIXIR Director

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

# Open access and sharable data - Why should data be shared?

- Reproducible science: data shared for validation and independent review?
  - Institutional repositories, limited metadata encoding, small group of users
- Secondary research use: “meta analysis” & reuse in new research
  - large user base, metadata standards, provenance & citation: requires different scale of investment.

# Open access and sharable data - Why should data be shared?

- Reproducible science: data shared for validation and independent review?
  - Institutional repositories, limited metadata encoding, small group of users
- Secondary research use: “meta analysis” & reuse in new research
  - large user base, metadata standards, provenance & citation: requires different scale of investment.

*If something is unsustainable it will stop*  
(Herbert Stein, chairman of the Council of Economic Advisers 72-74)



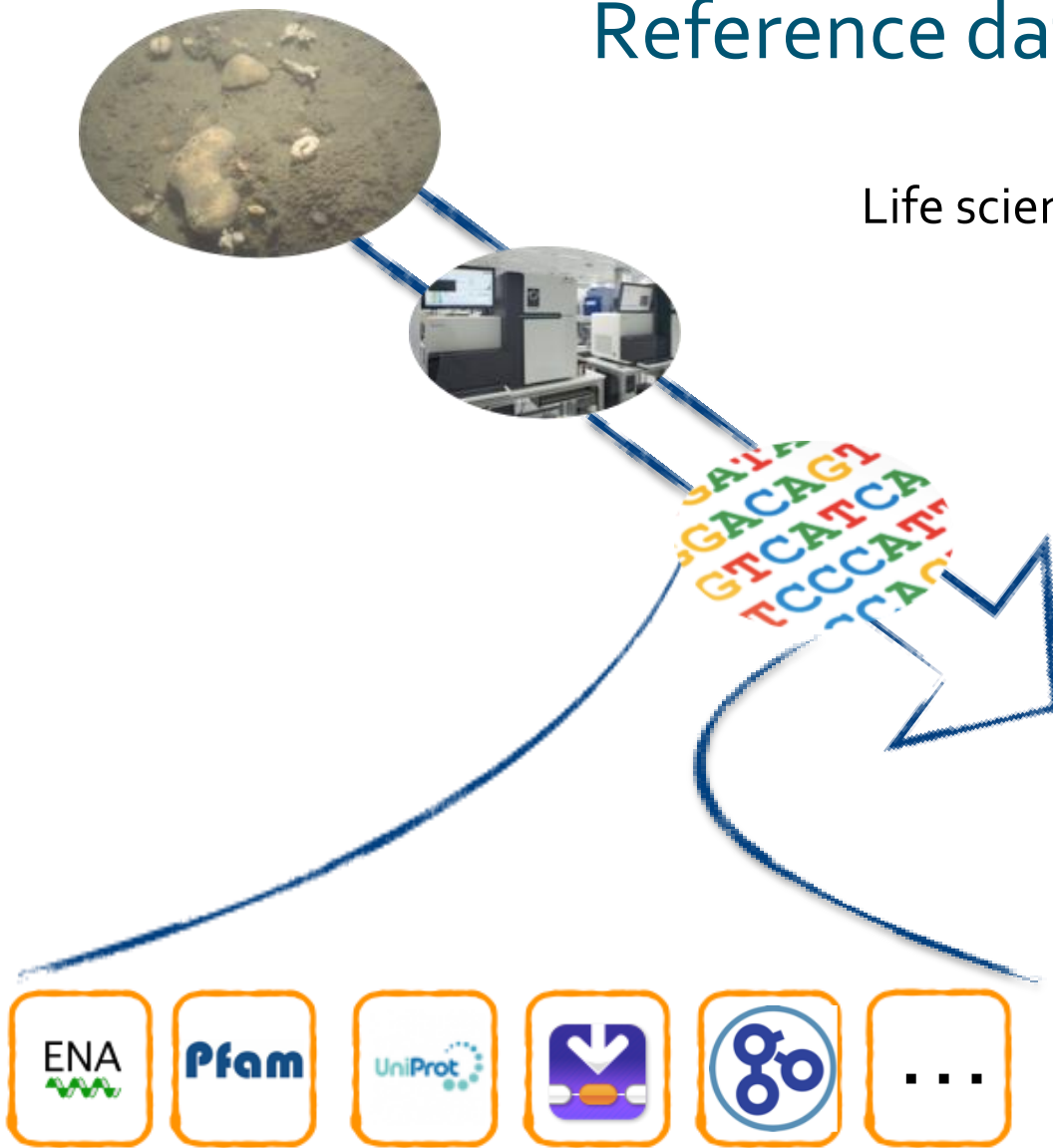
# TARA oceans: Global ocean microbiome



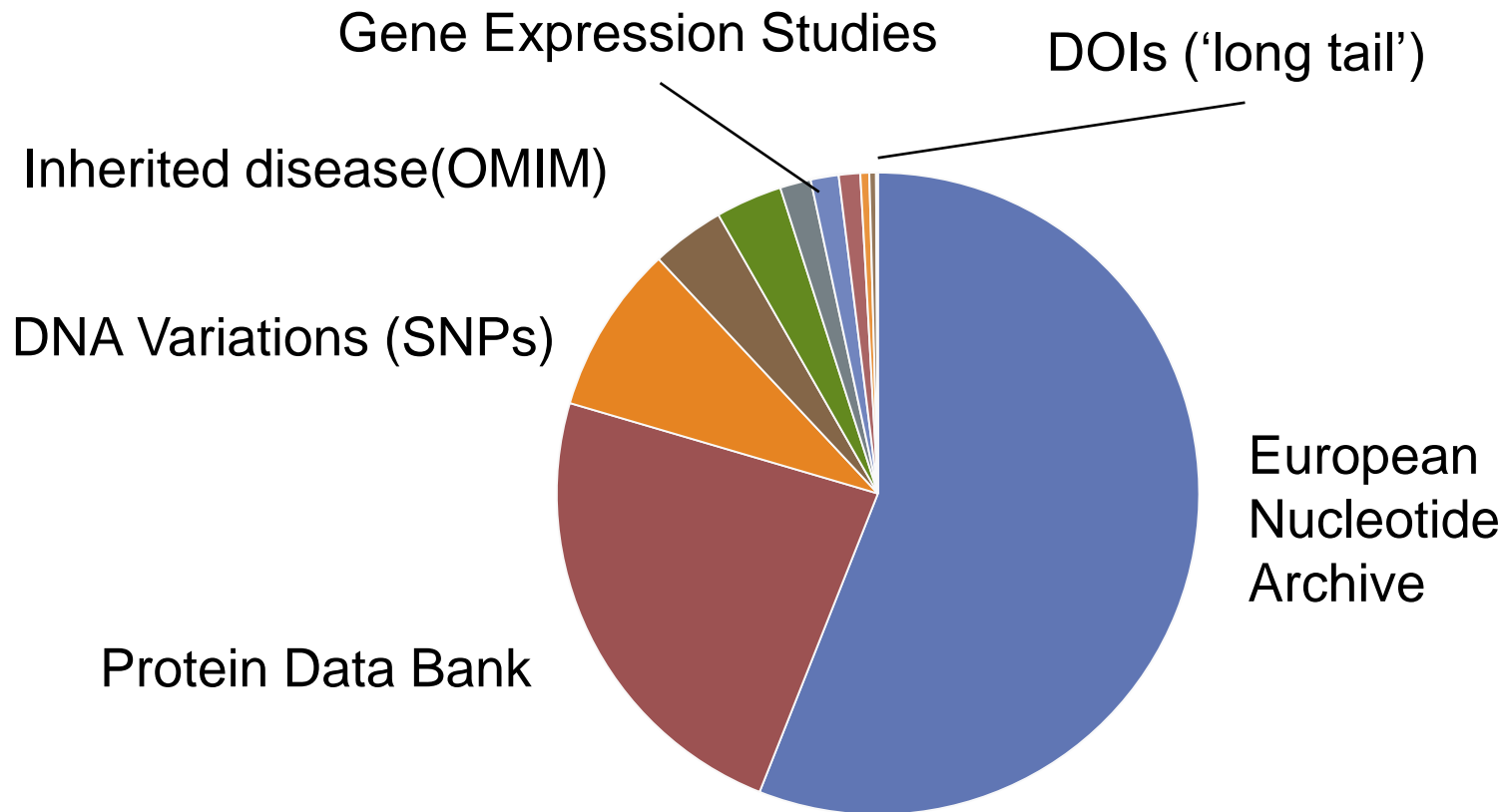
# Reference data is critical to biology

Life science archives and knowledge bases:

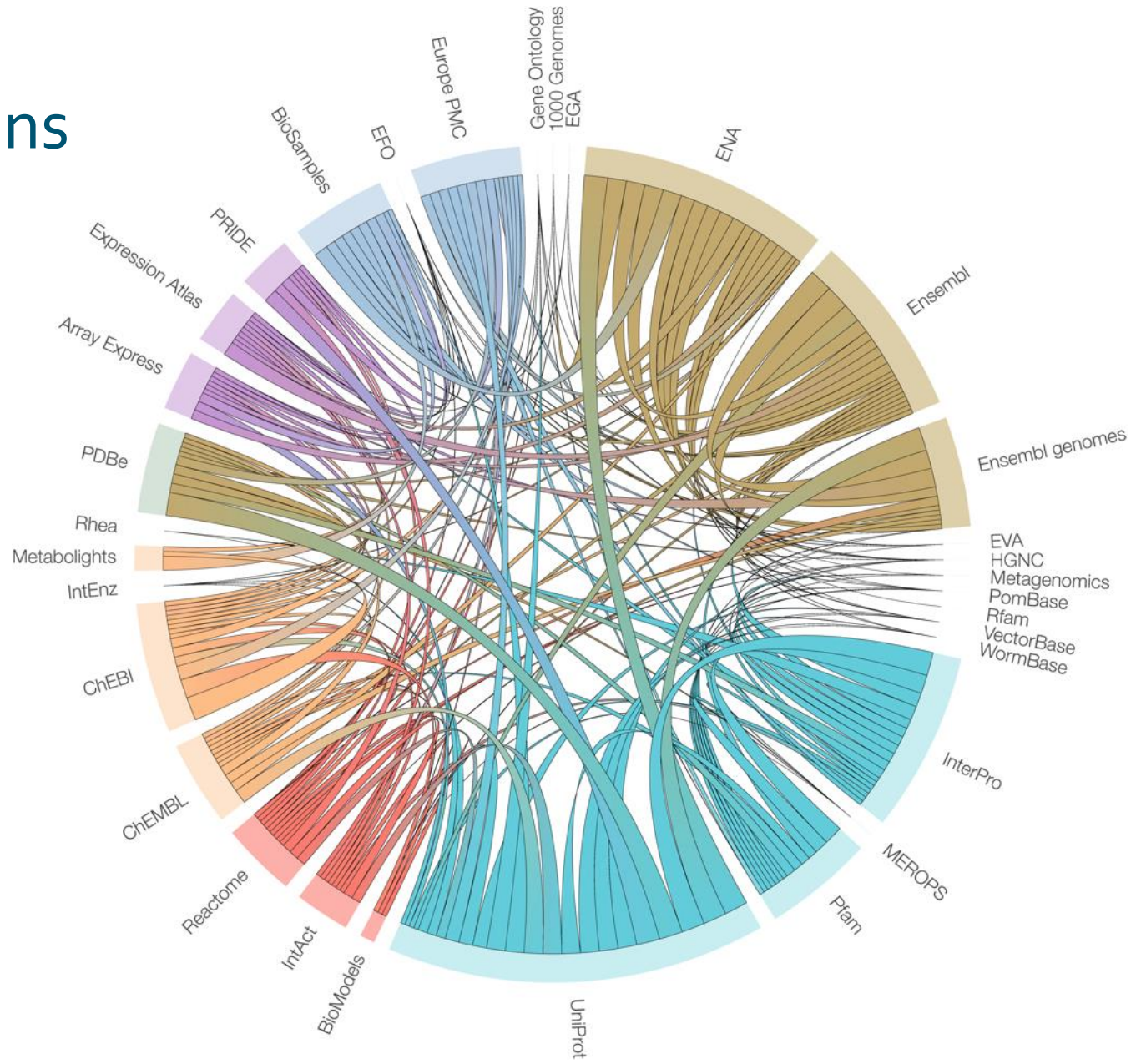
- European Nucleotide Archive
- Pfam
- UniProt
- Gene ontology
- ...



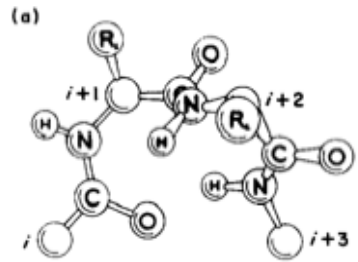
# Findable? Accessible? Interoperable? Reusable?



# Database Interactions

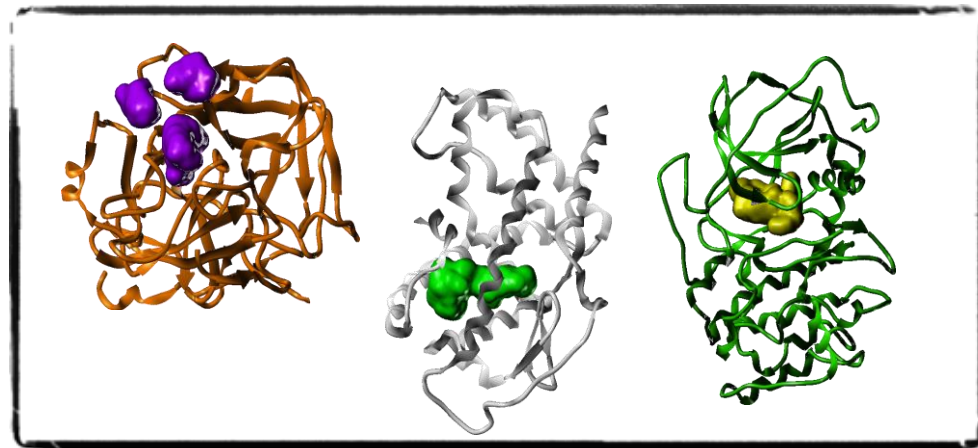
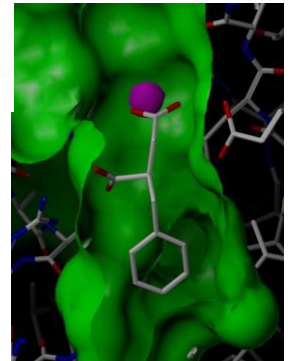
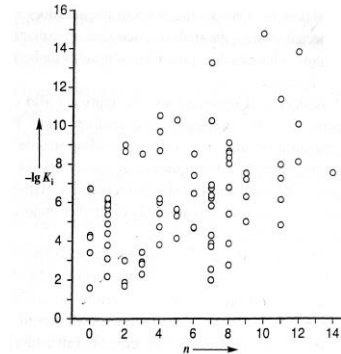


# The value of open data?



(a) Data extraction

$\beta$ -Turns were extracted from 59 non-identical proteins (resolution  $\leq 2$  Å) using data derived from the Brookhaven Protein Data Bank (Bernstein *et al.*, 1977).



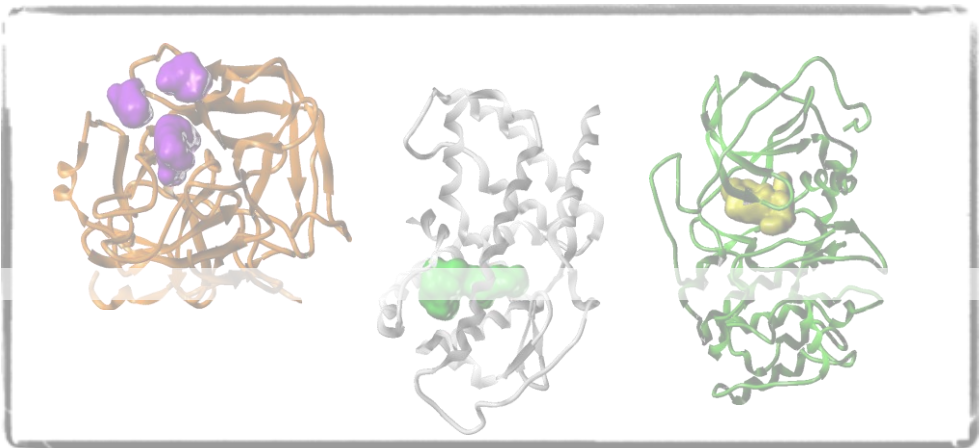


# The value of open data?

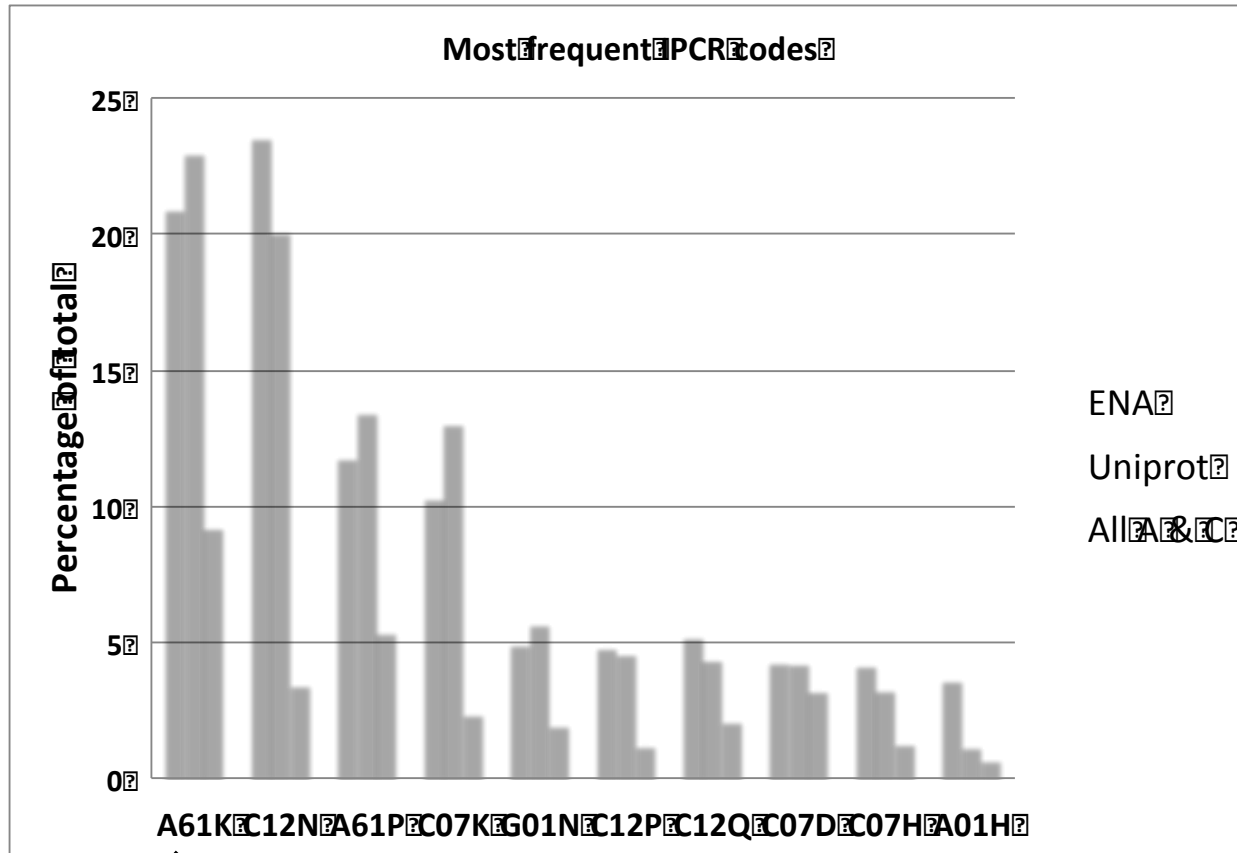


(a)     
Implementing Translational Science  
*Shaping the Industry*

 **SCHRÖDINGER.**  **CHEMICAL COMPUTING GROUP** 



# In 2014: 117k database citations in patents



Drugs, Beauty-care, ...  
Microorganisms



# The data challenges

- Data production sites increasing across Europe
- Data growth in life sciences
- Secure access and governance of human data
- Open data mandates of National and European funders

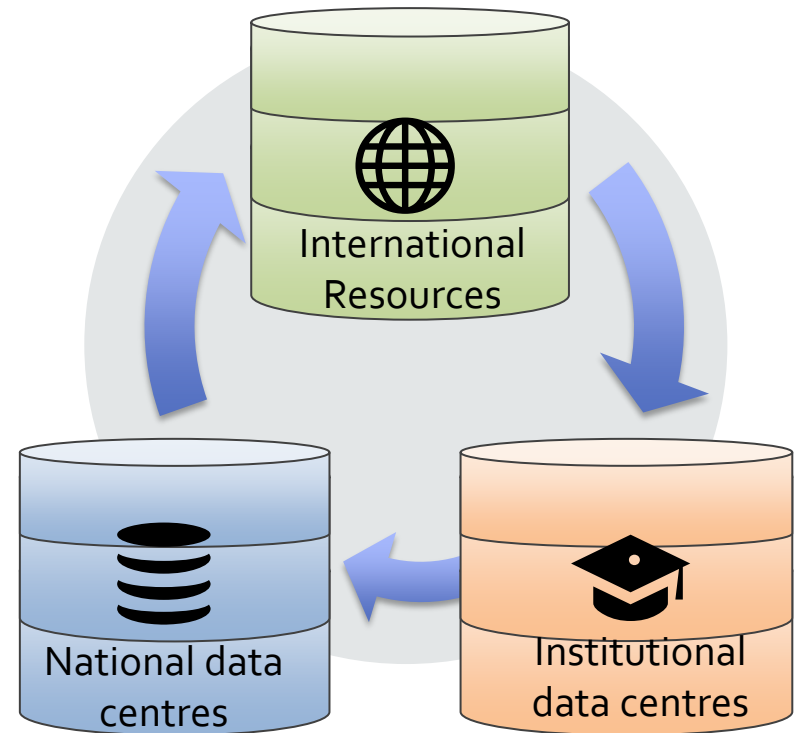


wellcome trust



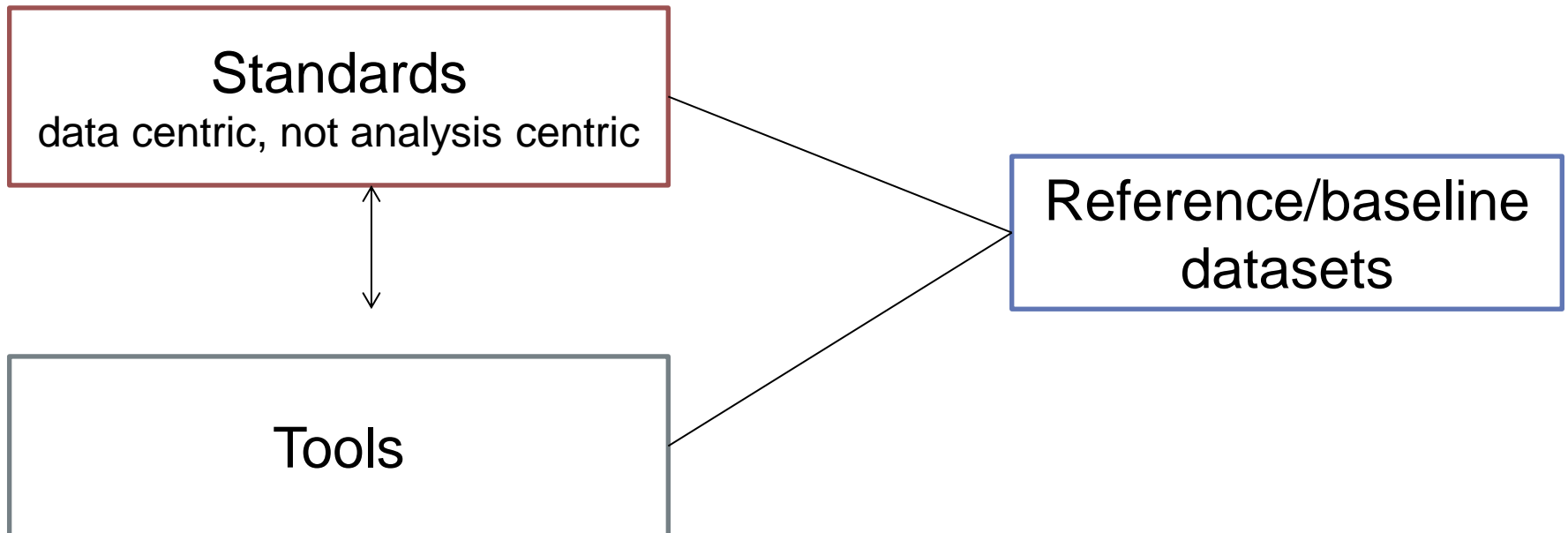
# Changing landscape with many actors

- Highly distributed data-generating & monitoring
- Distributed analysis requires reference datasets (organized centrally, locally or in distributed networks)



# The age of phenotyping?

- Phenotyping will become more important
  - RNA, Protein, Metabolites, Images, continuous sensing, ...
- Analysis is the bottleneck – will require:



***Will require rethinking of computational backbone & access models***



# Membership and collaborations

Five new ELIXIR  
Members in 2015-2016:  
France, Spain,  
Belgium, Italy and  
Slovenia

Ireland joined as  
observer

Collaboration Strategy  
signed with de.NBI  
(Germany)

Advanced discussions  
with Luxembourg

## Members



Belgium



Czech Republic



Denmark



EMBL



Estonia



Finland



France



Israel



Italy



Netherlands



Norway



Portugal



Slovenia



Spain



Sweden



Switzerland



United Kingdom

## Observers



Greece




Ireland

# A distributed infrastructure to scale with the challenges

- **ELIXIR** deliver services through national ELIXIR Nodes
- **ELIXIR** Nodes build local bioinformatics capacity throughout Europe
- **ELIXIR** Nodes build on national strengths and priorities



## ELIXIR: The Norway Node



ELIXIR's Norway Node will provide competence and infrastructure building on key areas for Norway, in particular marine resources and medical research. Challenges related to processing and analysis of data from next-generation sequencing and other high-throughput methods are important both to basic research in these areas, and to the development of new enterprises. The node will also provide training and support toward researchers.

### Collaborating organisations


**University of Bergen**  
The coordinating partner of the Elixir Norway node. The main focus is of marine genomics and e-infrastructure. An early deliverable is Licbase developed in tight collaboration with the Sea Lice Research Centre.

**University of Oslo**  
Emphasizes biomedical resource provision and analysis, leveraging public resources in integrative statistical genomics, with secure management of person sensitive data.


**Norwegian University of Life Sciences**  
Main focus on providing genomic resources for species-oriented and comparative fish genomics. Provision of web-based solutions for services, toolboxes, and computational access to these data.

**Norwegian University of Science and Technology**  
Tools and resources for analysing genome data, with focus on gene regulation, non-coding RNAs and epigenetics, but also bacterial genomics. Handling and analysis of data from human biobanks.

**University of Tromsø**  
Tools and pipelines for analyzing metagenomic (and genomic) data, with a particular focus on taxonomic classification and bioprospecting (functional and metabolic potential).




**Marine research**  
The Norwegian ELIXIR Node will provide services and resources toward marine genomics including researchers, government, and industry. The Norwegian Node will offer several integrated packages geared towards large-scale analysis of marine genomic and metagenomic data (e.g. fish genomics and marine bioprospecting). This also includes provision of web-based solutions for services, toolboxes, and computational access to reference data provided by the ELIXIR infrastructure.



**Health and biobanks**  
The Norway Node supports infrastructure for handling and analysis of data for medical research, including human biobanks. Such data may be sensitive, and must be stored with secure access. The node is developing infrastructure for sensitive data. Tools for data analysis are integrated into Hec3, the Norwegian e-infrastructure for life sciences. This provides user-friendly solutions for example for human re-sequencing data and other genome-scale analyses.

### Contact

Fullan.egg@unibergen.no  
ELIXIR Norway Node  
Computational Biology Unit  
Department of Informatics  
University of Bergen, Norway  
http://www.unibergen.no/elixir  
www.norwaynode.org



The Research Council of Norway

# ELIXIR Services



**Data deposition:**  
ENA, EGA, PDBe, EuropePMC, ...



**Compute:**  
Secure data transfer, cloud computing, AAI



**Added value data:**  
UniProt, Ensembl, OrphaNet, ...



**Bioinformatics tools:**  
Bio.tools



**Data management:**  
Genome annotation  
Data management plans



**Industry:**  
Innovation and SME programme  
Bespoke collaborations



**Data Interoperability:**  
Standards, Identifiers,  
Ontologies



**Training:**  
TeSS, Data Carpentry,  
eLearning





# Building a sustainable infrastructure for biological information across Europe

## ELIXIR Mission

- Building a sustainable infrastructure for biological information across Europe

## ELIXIR Services

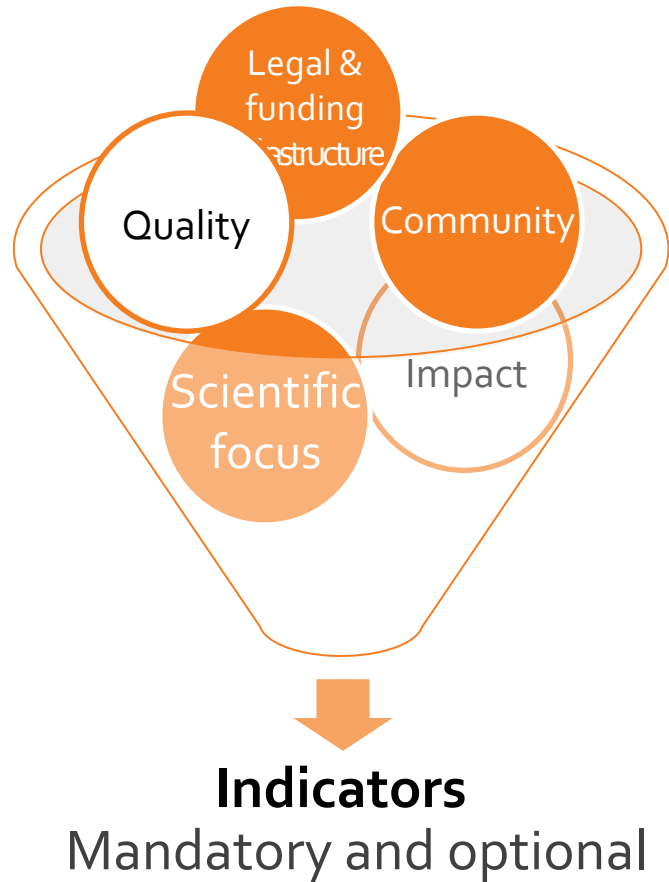
- Backbone of the ELIXIR life science data infrastructure providing stable access to biological data

## ELIXIR Core Data Resources

- Key reference datasets; Authority on identifiers

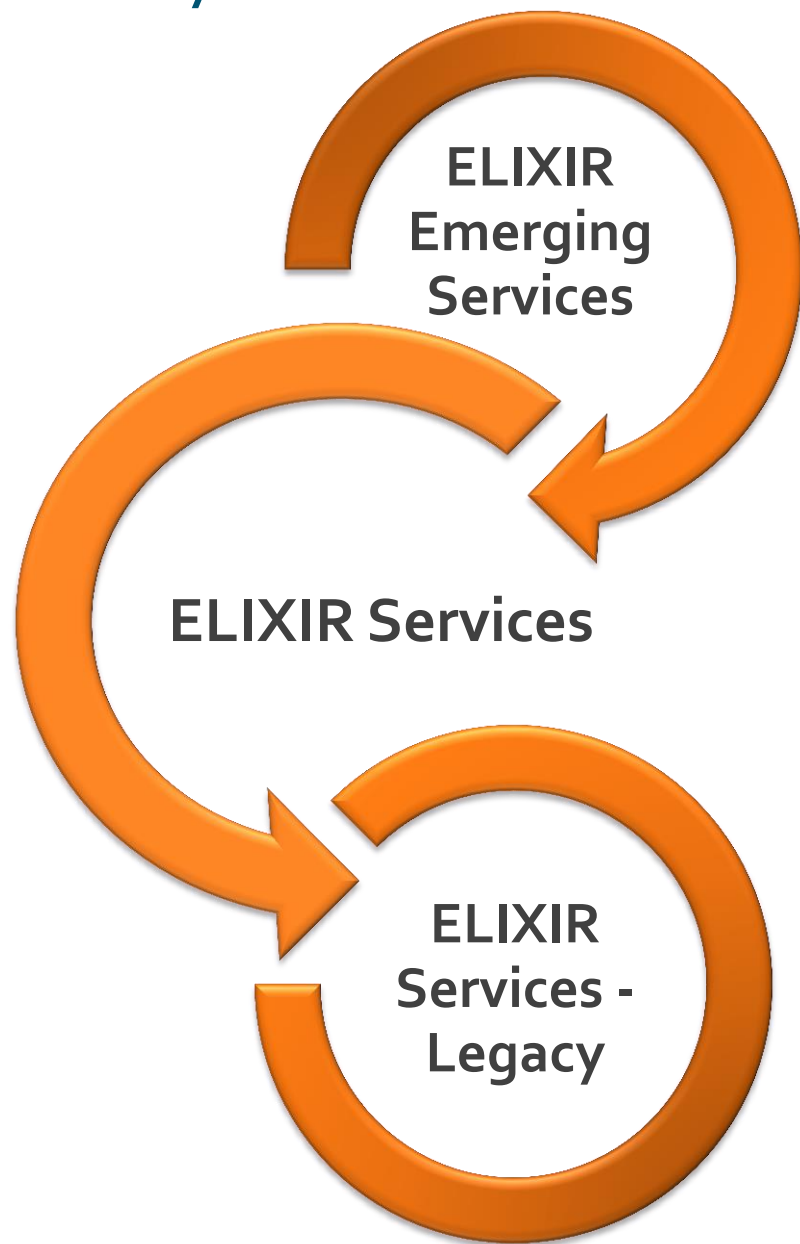


# Develop and sustain our core infrastructure of data resources



- 1) **Scientific focus** and quality of science  
e.g. curational effort, benchmarking
- 2) **Community** served by the resource  
e.g. web statistics
- 3) **Quality of service**  
e.g. uptime, user support and training
- 4) **Legal and funding infrastructure**  
e.g. institutional support, use policy
- 5) **Impact** and translational stories  
e.g. foundational role

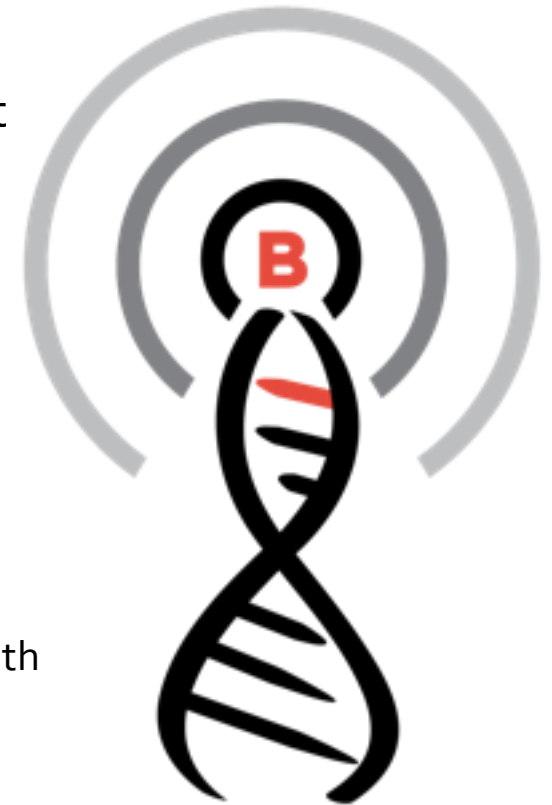
# Life cycle – ELIXIR Services



- Indicators guide and inform the managers of Emerging Services
- Capacity building support from ELIXIR community
- Monitoring of usage trends and the scientific impact provides information to support life cycle management

# Findable data: ELIXIR GA<sub>4</sub>GH Beacon Pilot

- **GA<sub>4</sub>GH Beacons are a discovery service:**
  - which datasets include genomes with allele of interest
- ELIXIR pilot project with partners from the Netherlands, Sweden, Finland, France and Spain
- Complies to the GA<sub>4</sub>GH Beacon project standard and security working group policies
- **Three objectives:**
  - Provide ELIXIR reference implementation on GA<sub>4</sub>GH Beacon with 3 authorization levels
  - Provide ELIXIR standards for data types
  - Provide an example on capacity and expertise build across ELIXIR Nodes to integrate national resources as part of a joint service interface

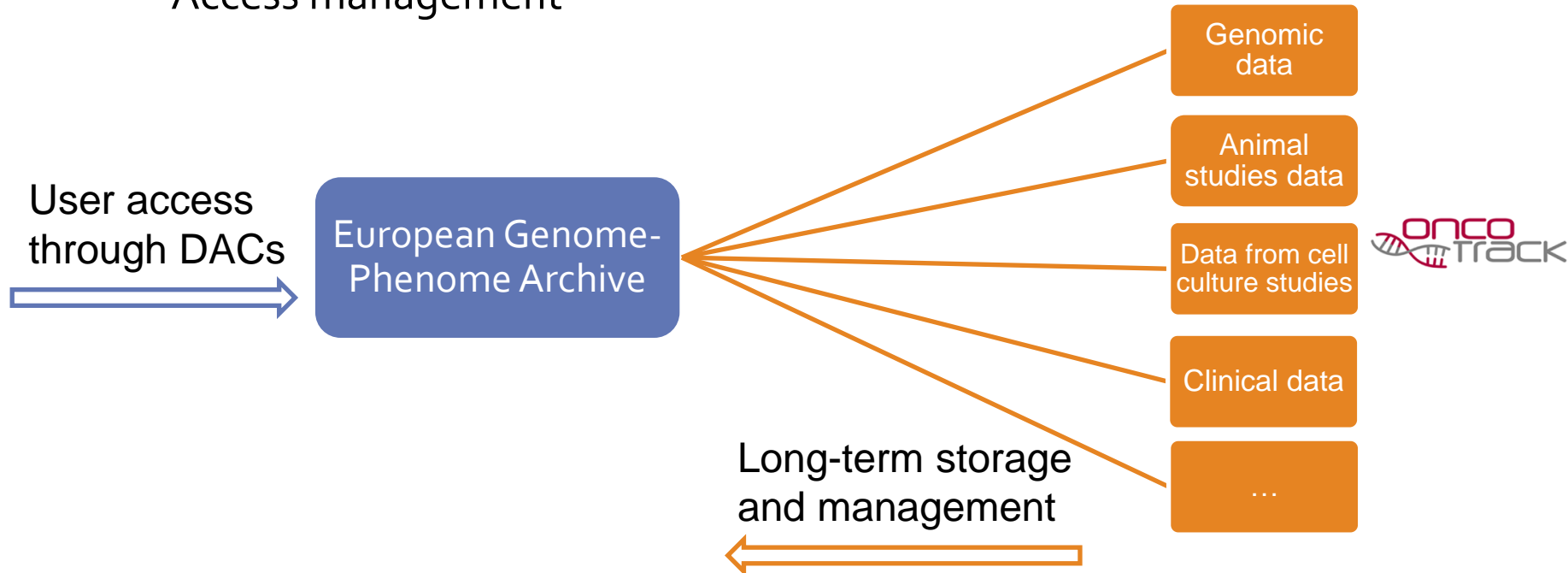


**Global Alliance**  
for Genomics & Health  
Collaborate. Innovate. Accelerate.



# Accessible data: ELIXIR – IMI OncoTrack

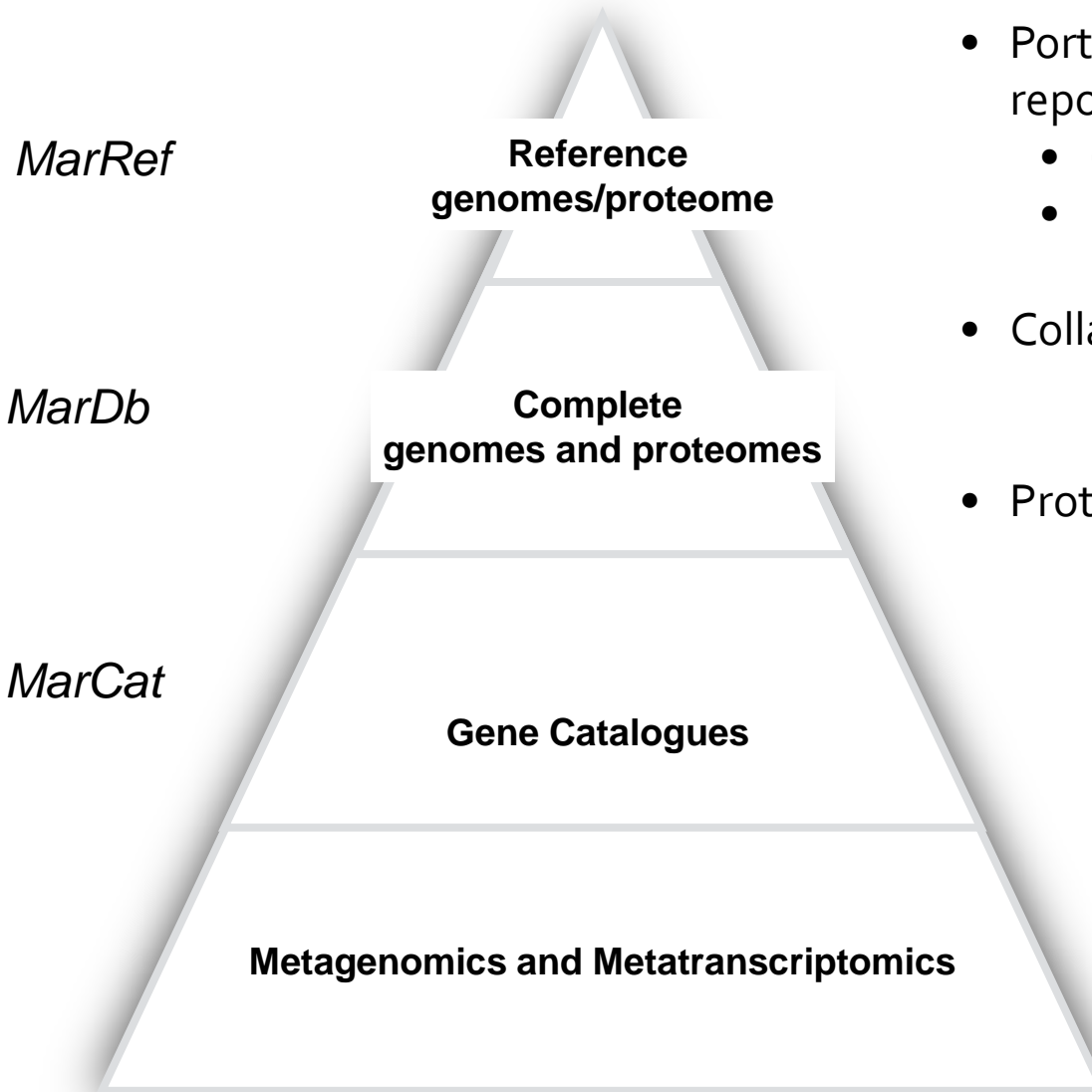
- Scoping study to understand long-term knowledge management requirements
  - Data storage, Meta data mappings, Data governance - Consent and Access management



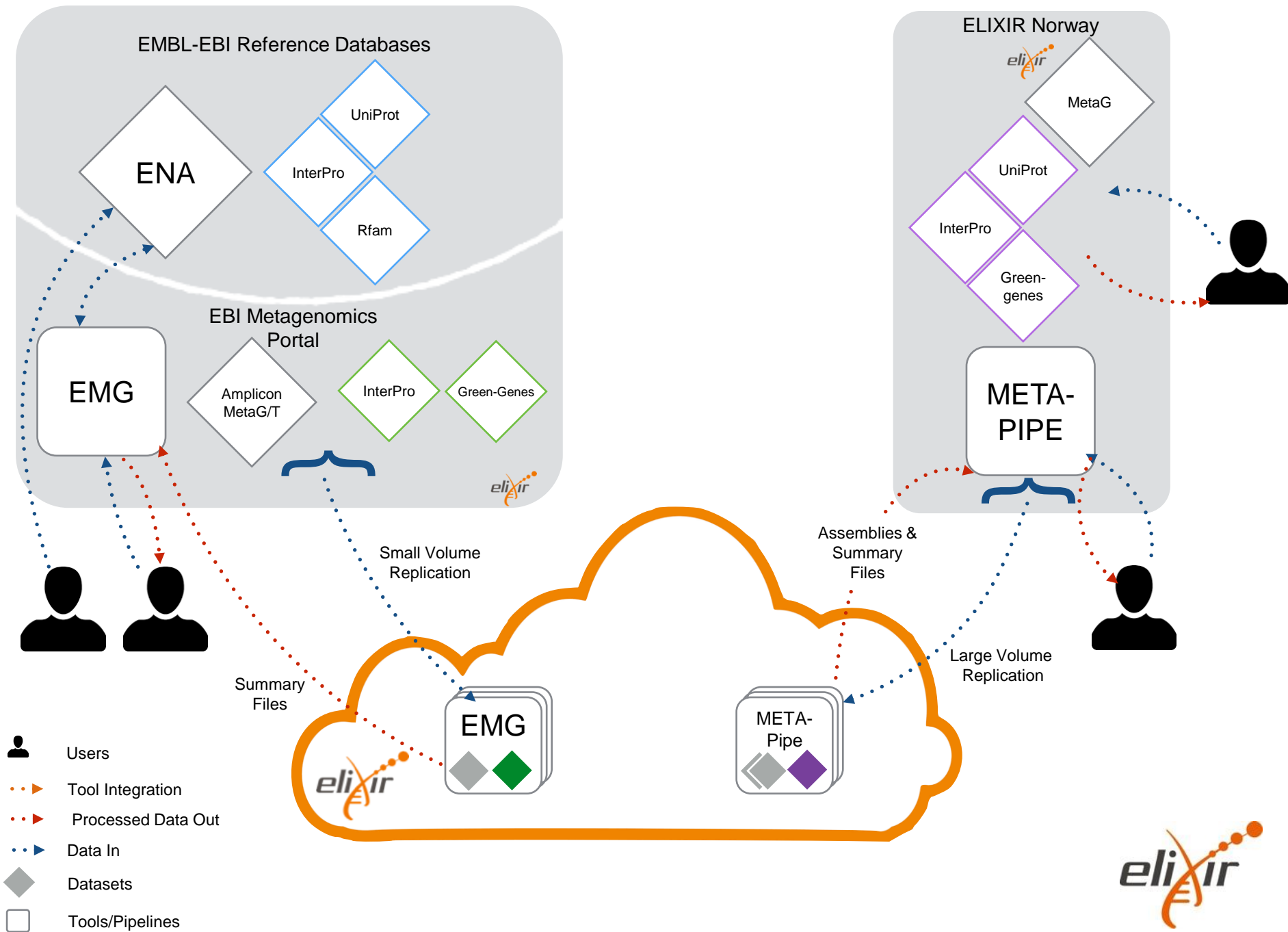
Innovative Medicines Initiative



# Marine reference annotations



- Portal that points to datasets existing in repositories
  - Curated
  - Filtered(?)
- Collate from Nodes and resources
- Prototype by end of December 2016



# Data-Driven Innovation in the aquaculture and marine industries

- 12-13 May 2016, Oslo, Norway
- **Programme:**
  - Introduction to marine science resources and services available through ELIXIR Norway and partners
  - Examples of companies making use of big data in marine sciences:  
**ArcticZymes, XELECT, Marine Harvest, AquaGenome**
  - Three Tech Track sessions: **Bioinformatics, Marine Genomics** and **Metagenomics**
- More information and registration:  
<http://innovation-marine.eventbrite.co.uk>







**Thank you**



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

