

ELIXIR - CORBEL and contributions of the EU health related infrastructures to Personalised Medicine in Europe

Jerry Lanfear, ELIXIR CTO

INFRAFRONTIER Stakeholder Meeting, November 2017



www.elixir-europe.org

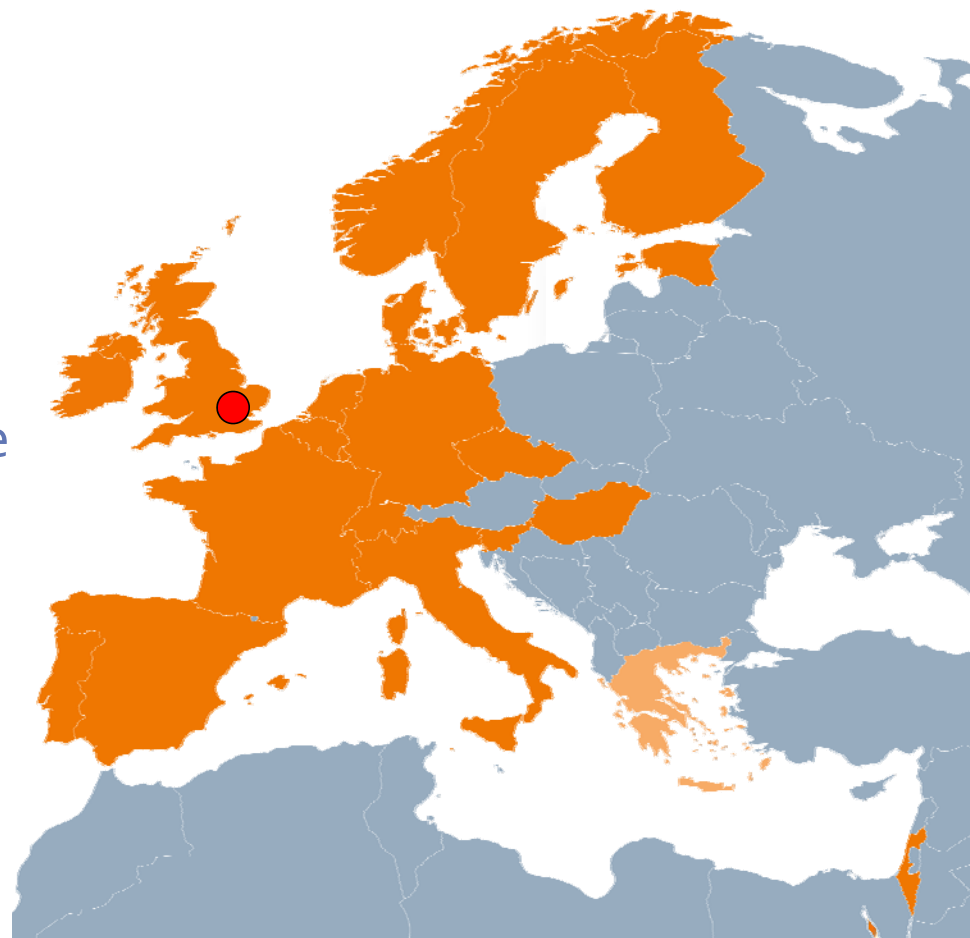
ELIXIR Members



ELIXIR Observers



ELIXIR, a
distributed
infrastructure
for biological
data



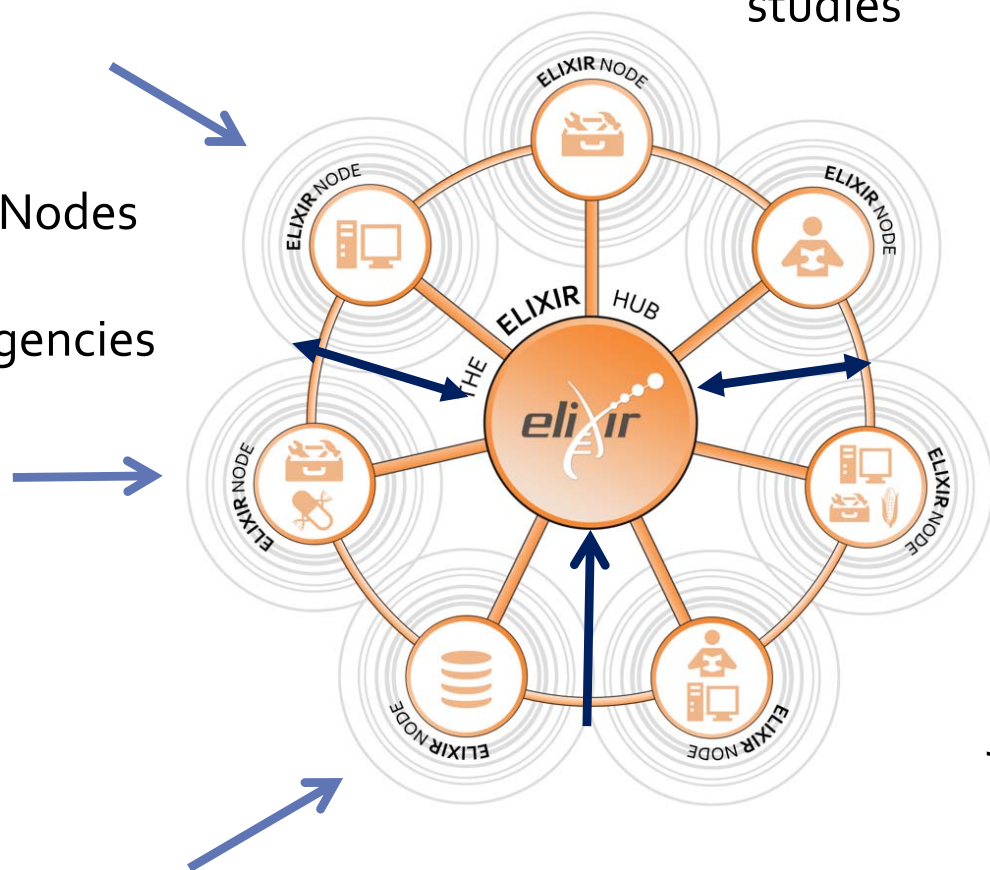
Note: ELIXIR Hub uses EMBL's legal model => ELIXIR categorised as international treaty organisation



ELIXIR funding

Implementation studies

Members/Nodes funded by national agencies



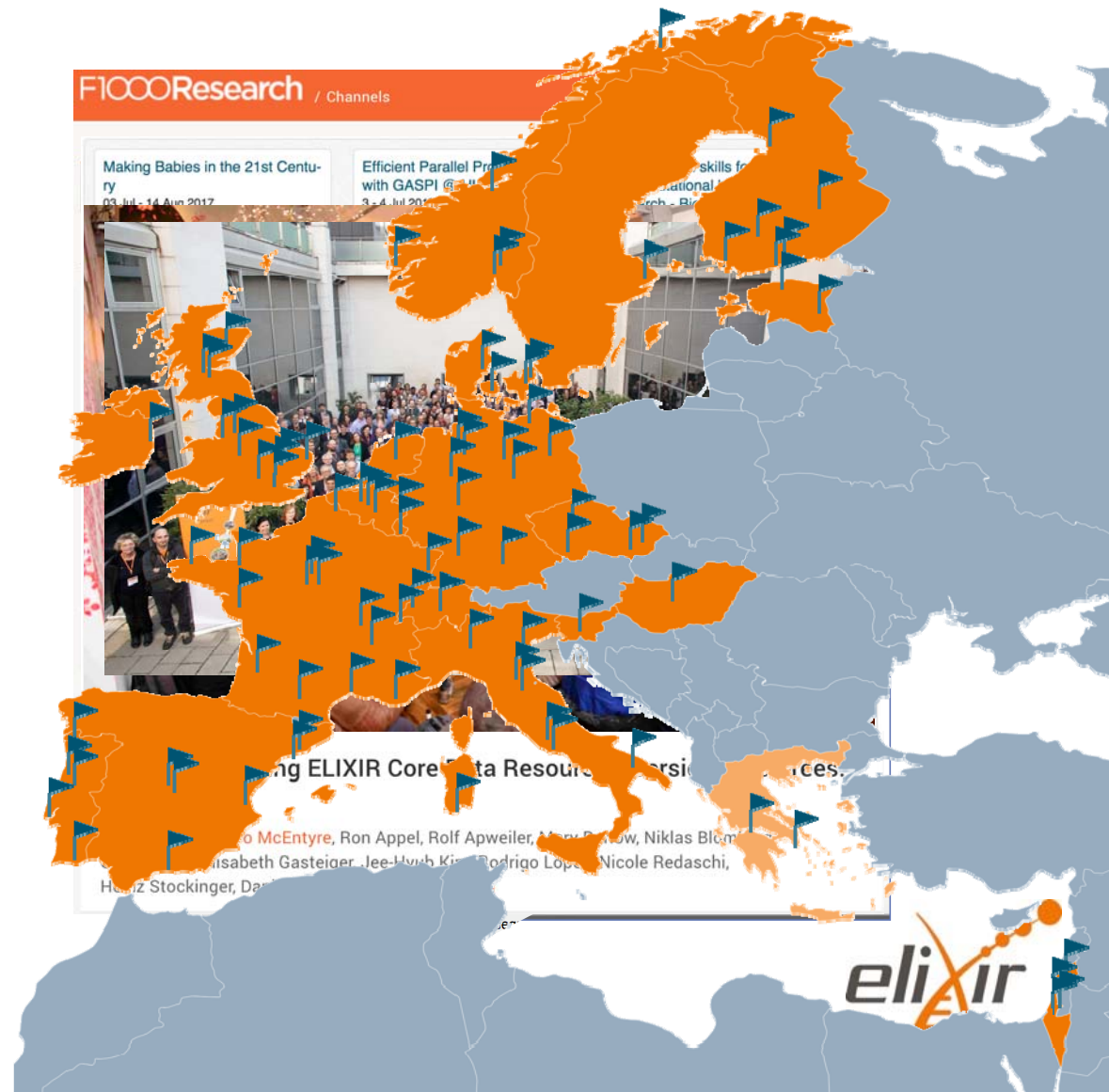
Hub funded jointly by ELIXIR member states

Hub & Nodes - Joint applications to EU funding



ELIXIR in numbers

- 21 Members and 1 Observer
- ~ 180 institutes involved
- 600+ staff
- 16 Core Data Resources
- 23 Implementation Studies ongoing or soon to start
- 17 papers in ELIXIR F1000 channel
- 264 live events in TeSS
- 350 companies attended Innovation and SME programme



ELIXIR: European infrastructure for biological information

Data infrastructure for Europe's life-science research:

Platforms



Data



Interoperability



Tools



Compute



Training

Use cases



Marine metagenomics



Crop and forest plants



Human data



Rare diseases



 [@ELIXIREurope](https://twitter.com/ELIXIREurope)

www.elixir-europe.org



Vision

To facilitate discoverability, access, sharing and analysis of genomics data, including rare disease, linked to other data types, at scale

(4-5M participants)

To demonstrate how use of infrastructure can impact translation of genomics research into medicine

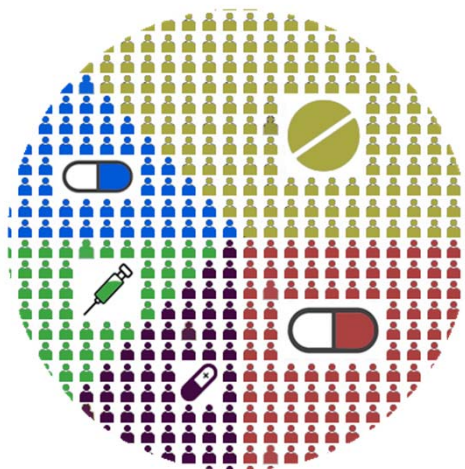


Infrastructure for Human Genomics and Translational data

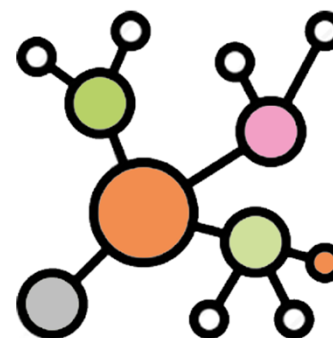
Understand disease



Patient Stratification and PM



Develop and test novel
pharmacological hypotheses



Unleash the possibilities for genomics and health



Mission

- To construct and operate a sustainable infrastructure for Human Genomics and Translational data in Europe to support life science research and its translation to medicine



"Whenever possible, biological research data should be submitted to the recommended community deposition databases"

ELIXIR Deposition Database list

Deposition Database	Data type	International collaboration framework ¹
ArrayExpress	Functional genomics data. Stores data from high-throughput functional genomics experiments.	
BioModels	Computational models of biological processes.	
EGA	Personally identifiable genetic and phenotypic data resulting from biomedical research projects.	European Bioinformatics Institute and the Centre for Genomic Regulation
ENA	Nucleotide sequence information, covering raw sequencing data, contextual data, sequence assembly information and functional and taxonomic annotation.	International Nucleotide Sequence Database Collaboration
IntAct	IntAct provides a freely available, open source database system and analysis tools for molecular interaction data.	The International Molecular Exchange Consortium
MetaboLights	Metabolite structures and their reference spectra as well as their biological roles, locations and concentrations, and experimental data from metabolic experiments.	
PDBe	Biological macromolecular structures.	
PRIDE	Mass spectrometry-based proteomics data (protein expression information (intensity values) and the support	

- The ELIXIR Deposition Databases meet the technical quality and governance criteria expected of ELIXIR [Core Data Resources](#)
- Agreed collectively by 21 Heads of Nodes
- International collaborative effort

<https://elixir-europe.org/platforms/data/elixir-deposition-databases>



ELIXIR Core Data Resources

ELIXIR Core Data Resource list

Core Data Resource	Data type
ArrayExpress	Functional Genomics Data from high-throughput functional genomics experiments.
CATH	A hierarchical domain classification of protein structures in the Protein Data Bank.
ChEBI	Dictionary of molecular entities focused on 'small' chemical compounds.
ChEMBL	Database of bioactive drug-like small molecules, it contains 2-D structures, calculated properties and abstracted bioactivities.
EGA	Personally identifiable genetic and phenotypic data resulting from biomedical research projects.
ENA	Nucleotide sequencing information, covering raw sequencing data, sequence assembly information and functional annotation.
Ensembl	Genome browser for vertebrate genomes that supports research in comparative genomics, evolution, sequence variation and transcriptional regulation.
Ensembl Genomes	Comparative analysis, data mining and visualisation for the genomes of non-vertebrate species.
Europe PMC	Europe PMC is a repository, providing access to worldwide life sciences articles, books, patents and clinical guidelines.
Human Protein Atlas	The Human Protein Atlas contains information for a large majority of all human protein-coding genes regarding the expression and localization of the corresponding proteins based on both RNA and protein data.
The IMEx Consortium: represented by IntAct and MINT	IntAct provides a freely available, open source database system and analysis of molecular interaction data. MINT focuses on experimental interactions mined from the scientific literature.
InterPro	Functional analysis of protein sequences, predicting the presence of

- Set of European data resources of fundamental importance to the wider life-science community and the long-term preservation of biological data.
- Agreed collectively by 21 Heads of Nodes
- International collaborative effort

<https://elixir-europe.org/platforms/data/core-data-resources>



European Genome-phenome Archive (EGA)

Controlled access for human genomic data



Data submitters: >600
(Universities, research institutes,
consortia, pharma, ...)



International
Cancer Genome
Consortium



IRDIRC

INTERNATIONAL
RARE DISEASES RESEARCH
CONSORTIUM



wellcome trust
sanger
institute



UK
10K



Data requesters:
> 7,100 (worldwide)
Controlled access:
Data Access Committees



www.ega-archive.org



EUROPEAN
GENOME-PHENOME
ARCHIVE

EMBL-EBI



CRG
Centre
for Genomic
Regulation

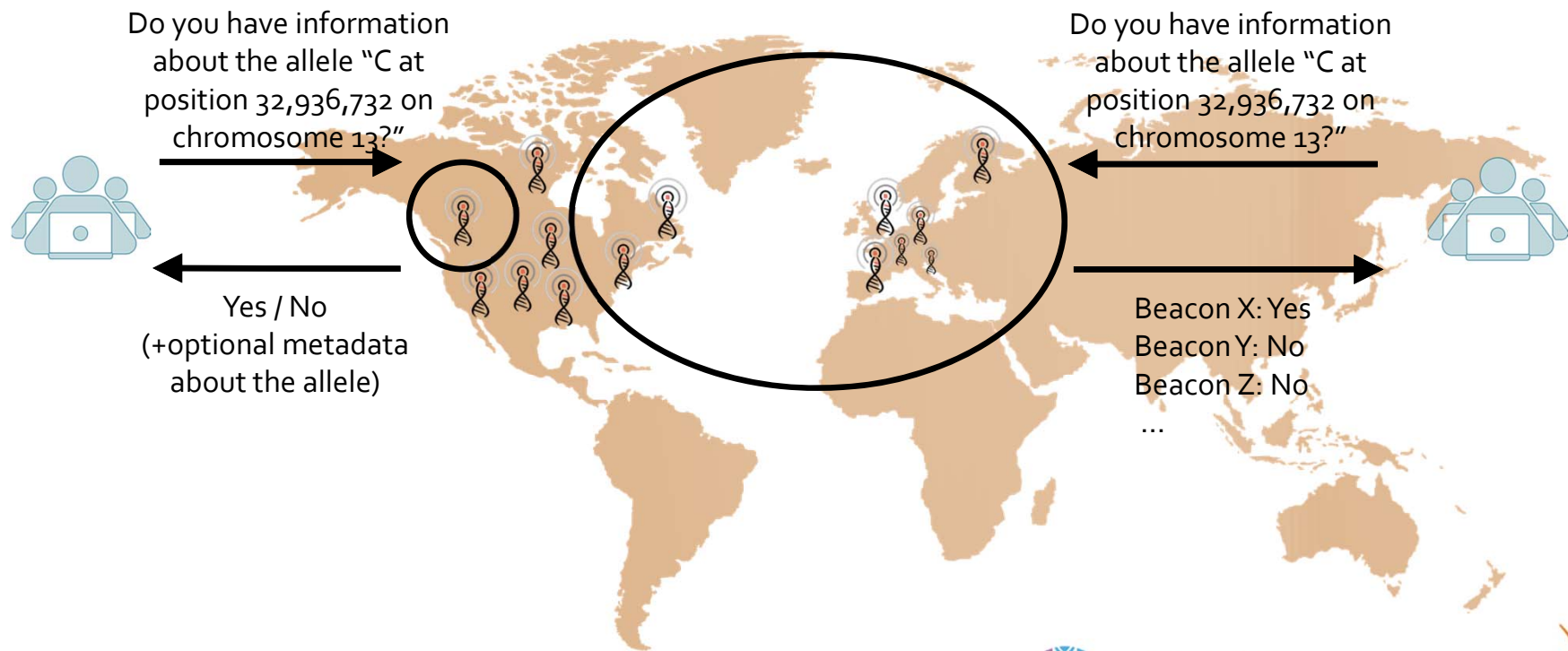


Local EGA

- Sensitive data are stored locally
- EGA provides the software platform
- First prototype in 2017 with national partners



Beacons: Public data discovery web-service



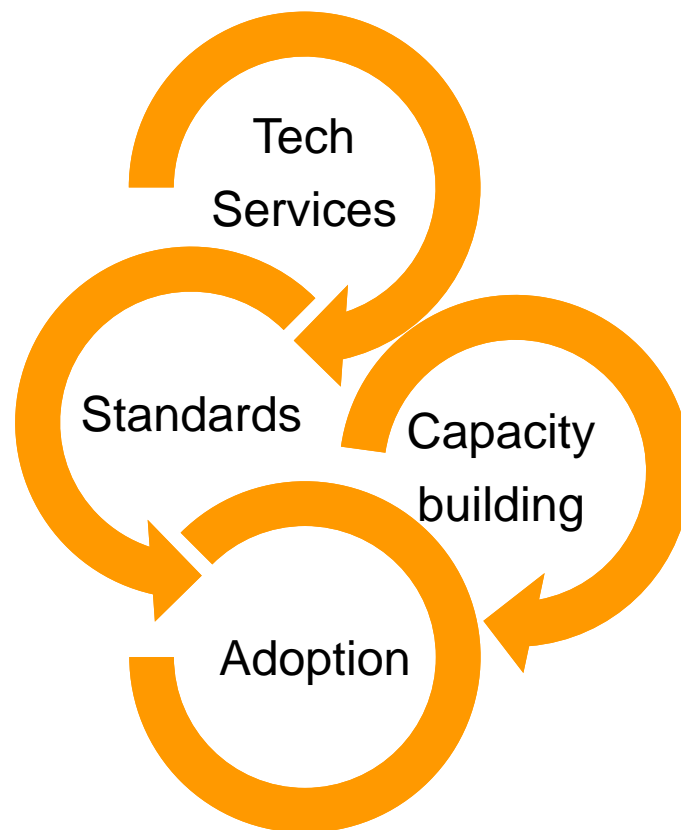
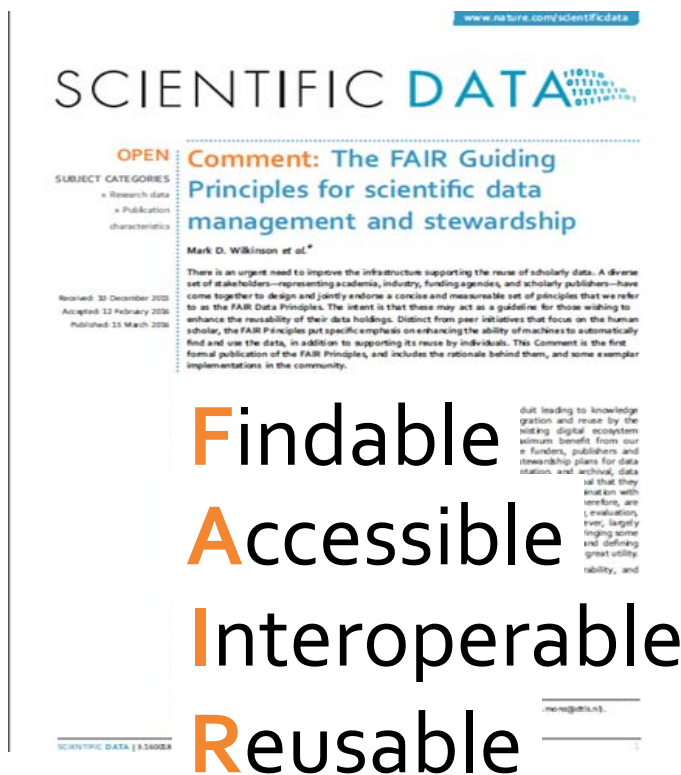
beacon-network.org/



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



ELIXIR Interoperability



Interoperability services and practices to support FAIR data and interoperability activities

Use case focused
Data providers
Data integrators

With international initiatives, from community grassroots to government programmes.





Bioschemas.org

ELIXIR Implementation study led by ELIXIR UK



<http://bioschemas.org>

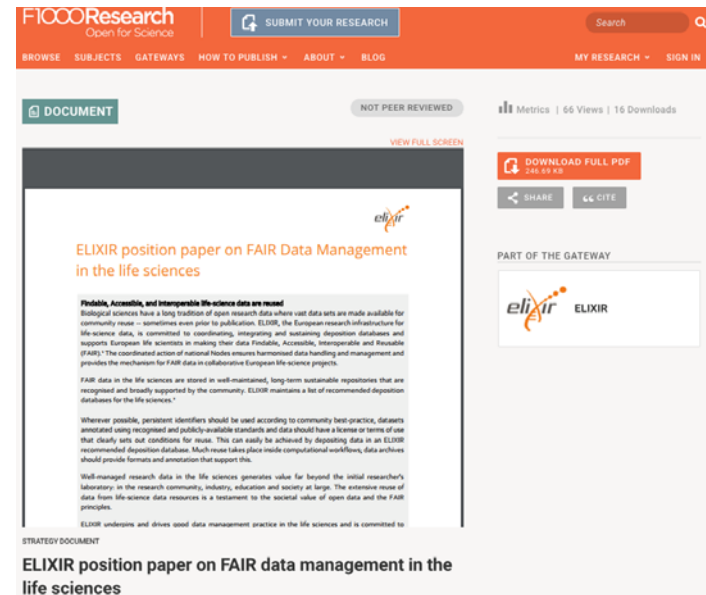
- Schema.org = **Structured data markup for web pages**
- Bioschema.org = schema.org for biological information (dataset, sample, standard, protein, training material ...)
 - Specification on top of schema.org
 - Layer of constraints + documentation + extensions
- Minimum properties for finding data

Specification	schema.org
Data model	
Minimum information	
Controlled vocabularies	
Cardinality	
New (properties types)	
Documentation	
Examples	



ELIXIR Position Paper on FAIR data management in the life sciences

1. Open sharing of research data is a core principle
2. Data Management is crucial to science
3. Data should be submitted to deposition databases
4. All data submitted to Open Data archives should align with community-defined standards
5. ELIXIR Nodes implement FAIR for their respective nations
6. Professional skills, adequate resources and appropriate funding are needed for Data Management and infrastructure

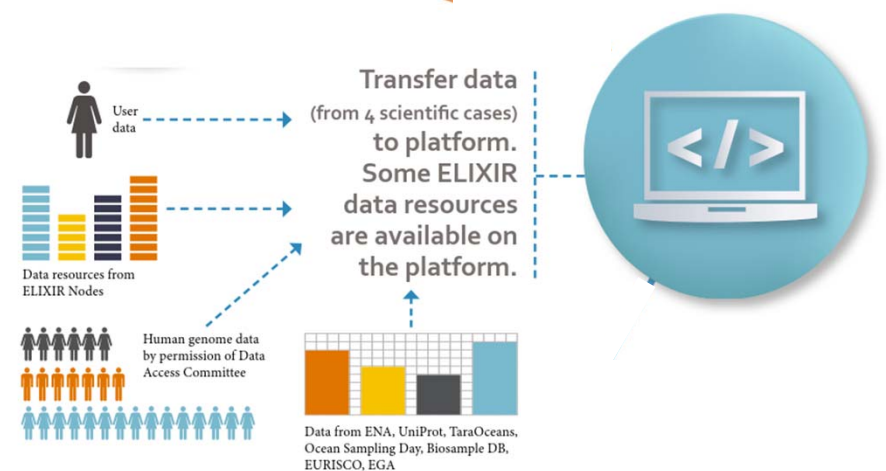


Blomberg N and ELIXIR Consortium. ELIXIR position paper on FAIR data management in the life sciences. **F1000Research** 2017, **6**(ELIXIR):1857 (document) (doi: [10.7490/f1000research.1114985.1](https://doi.org/10.7490/f1000research.1114985.1))



Practical demonstration: Data storage and transfer, coupled to security

- **ELIXIR Webinar:** Transfer of large volumes of confidential human data, while maintaining appropriate access rights
- 13 November 2017
- Slides and recording available at: <https://www.elixir-europe.org/events/elixir-webinar-transfer-large-volume-data>



ELIXIR Authorisation and Authentication Infrastructure

Enables life science researchers to use their institutional IDs to access services and data:

- **Reduced bureaucracy and costs**
- **Improved vetting:** federated identities provide greater confidence to the service and data providers
- **Regular updates:** as researchers join leave institutions, their affiliation information is maintained regularly
- **Improved access to usage metrics:** consistent use of accounts allows service providers to better analyse the use of their services
- **Applicable to other research infrastructures (CORBEL)**





CORBEL: SHARED SERVICES FOR LIFE SCIENCES

Thirteen ESFRI Research Infrastructures in Bio-Medical Sciences to **harmonise access and services** for complex research projects involving more than one research infrastructure.



- *biobanking & biomolecular resources*



- *curated databases*



- *highly pathogenic microorganisms*



- *functional genomics*



- *microorganisms*



- *translational research*



- *marine model organisms*



- *screening & medicinal chemistry*



- *structural biology*



- *clinical trials*



- *plant phenotyping*



- *biological/medical imaging*



- *systems biology*



BACKGROUND

- 4 year project: 2015-2019
- 37 partners in 13 BMS RIs
- budget: €14.8 million
- builds on BioMedBridges (2012-2015)
- co-coordinated by ELIXIR and
BBMRI-ERIC





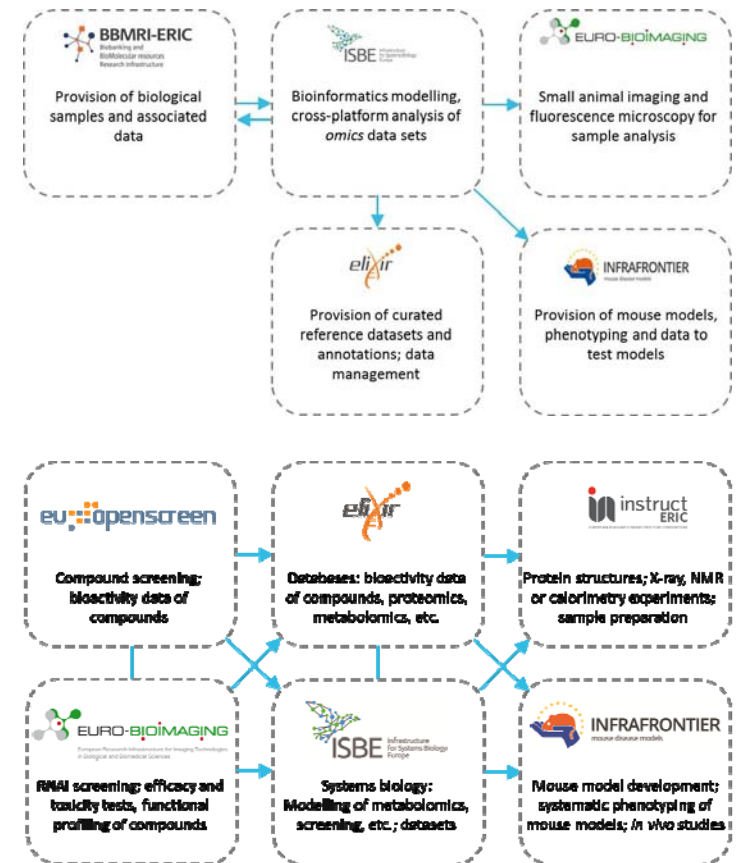
EXAMPLE OF CORBEL INNOVATION PIPELINE

Use Case 1

Genotype-to-Phenotype analysis
based on models and experimental
data

Use Case 2

Predictive systems pharmacology for
safer drugs and chemical products



ELIXIR: European infrastructure for biological information

Data infrastructure for Europe's life-science research:

Platforms



Data



Interoperability



Tools



Compute



Training

Use cases



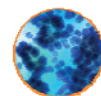
Marine metagenomics



Crop and forest plants



Human data



Rare diseases



 [@ELIXIREurope](https://twitter.com/ELIXIREurope)

www.elixir-europe.org

