

# ePTRI

EUROPEAN PAEDIATRIC TRANSLATIONAL RESEARCH INFRASTRUCTURE

## EPTRI-ELIXIR common service "Paediatric Data Interoperability"

EPTRI Stakeholder Round Table– July 9, 2020



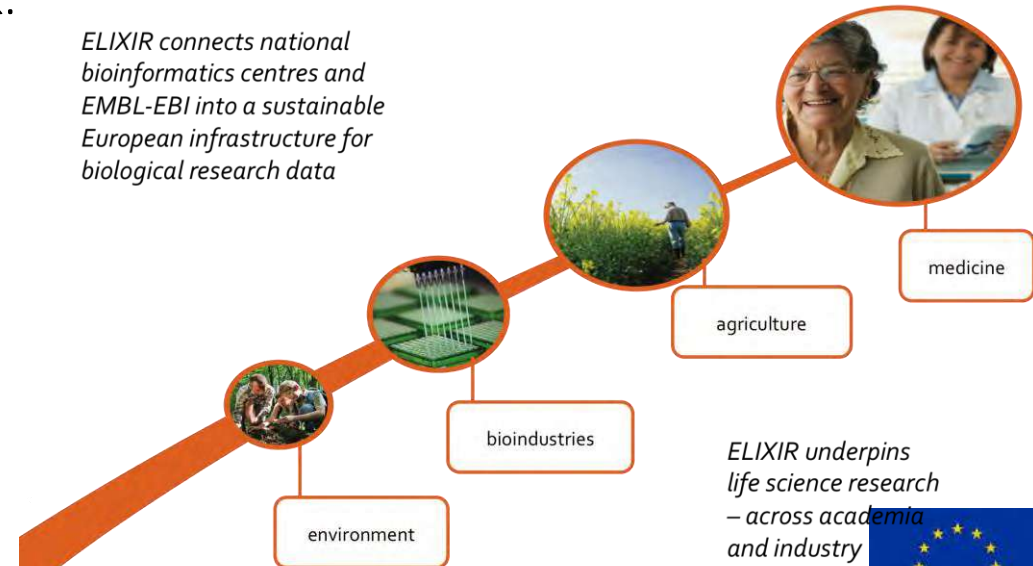
This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No 777554

# ELIXIR: a Research Infrastructure to face the Big Data challenge in Biology in Biology

ELIXIR is an **intergovernmental organisation**, formally established in 2016 as a Landmark European Research Infrastructure, that brings together “**bioinformatic resources**” for **life sciences** from across Europe. These resources include **databases, software tools, training materials, best practices, cloud storage and supercomputers**.

The goal of ELIXIR is to coordinate these resources so that they form a **single infrastructure**. This infrastructure makes it easier for scientists to **find and share data, exchange expertise, and agree on best practices**. Ultimately, it will help them gain **new insights** into how living organisms work.

*ELIXIR connects national bioinformatics centres and EMBL-EBI into a sustainable European infrastructure for biological research data*

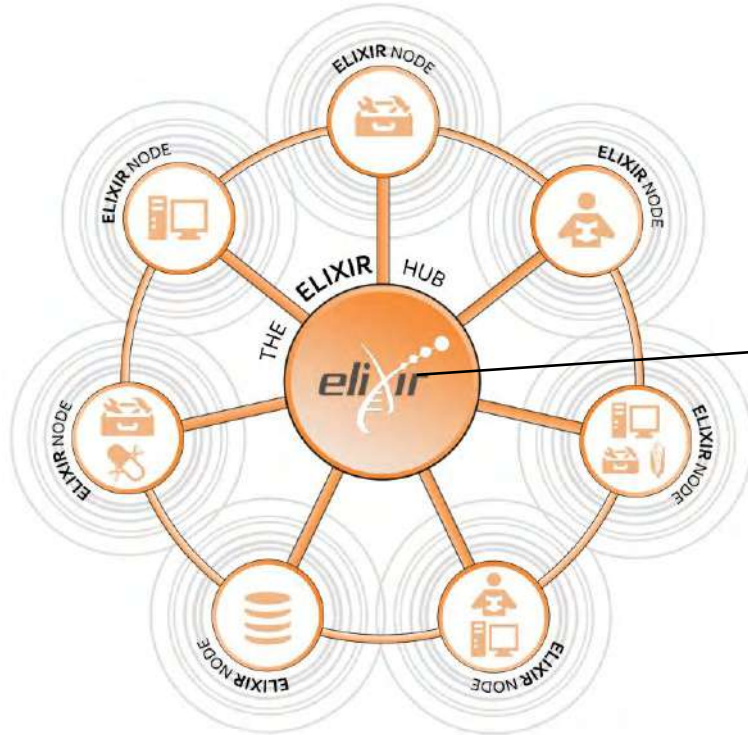


*ELIXIR underpins life science research – across academia and industry*



# ELIXIR: A pan-european distributed Infrastructure for Bioinformatics

**ELIXIR** is structured as a central hub, located in the Wellcome Genome Campus (Hinxton, UK) and 23 national nodes including over 160 Research Organizations.



# ELIXIR Organization



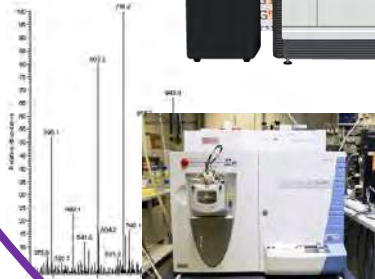
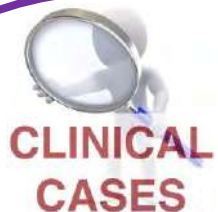
Five technical **platforms** for **Compute, Data, Tools, Interoperability** and **Training**

Complemented by several **user communities**

In the **2019-23 Scientific Programme** use cases evolved in “User Communities” enlarging the ELIXIR portfolio such as Proteomics, Metabolomics, Galaxy, ..

# Big Data-driven innovation requires complex eco-systems

**DATA**



**Heterogeneous and fragmented data**

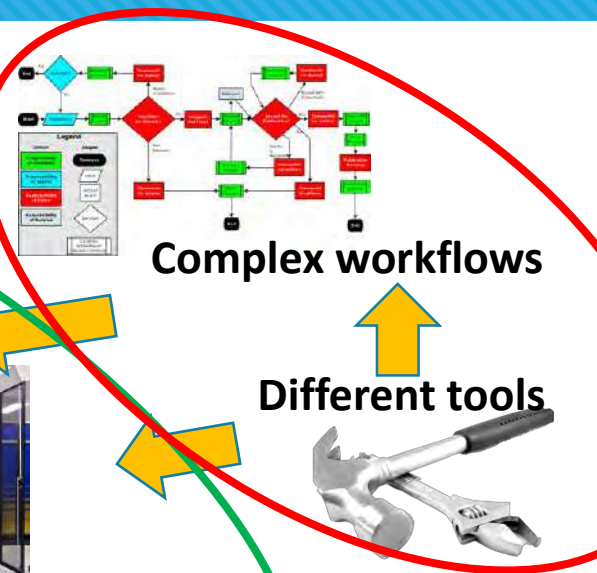


**Heterogeneous and fragmented databases**

**COMPUTE**



**TOOLS**



**Complex workflows**

**Different tools**



**Heterogeneous computational systems**



**ePTRI**

EUROPEAN PAEDIATRIC TRANSLATIONAL RESEARCH INFRASTRUCTURE

EPTRI Stakeholder Round Table– July 9, 2020

# What is needed to efficiently connect the ecosystem?

## ***INTEROPERABILITY***

- Standard formats
- Standard description of concepts (Ontologies)
- Standard and stable identifiers
- Rich, standard and machine-readable description of resources(data and tools) with metadata
- Clear access/privacy policies
- Technologies to deploy tools on different computational architectures
- Languages to easily connect different tools/data in workflows



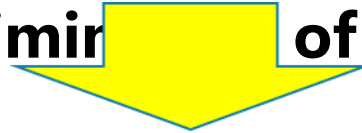
# Make the data FAIR!

Findable, Accessible, Interoperable, Reusable

**Process automation** through machine readability (of data and metadata)

Effective **streamlined acquisition, integration and analysis of data**

**Minimization/elimination** of data wrangling



**Scientific queries are answered more rapidly in a flexible way.**

**time-to-value are significantly reduced**

**R&D can be accelerated.**

**developing more-segmented or -personalized medicines**



# Common service for data interoperability

<https://elixir-europe.org/platforms/interoperability>

**Standards:** formats, reporting guidelines, ontologies

**Metadata services:** ontology, annotation, validation, harvesting, Indexing

**Register** services and datasets

**Search engine** for datasets.

**Identifier** resolution & management

**Identifier** mapping services

Describing and sharing **workflows** between different systems

Harmonisation of **tools and pipelines**

**Common Programmable Interfaces**

Best practice.

**FAIRsharing.org**  
standards, databases, policies



**Bioschemas**



**Identifiers.org**  
Resolution service



BioContainers



COMMON  
WORKFLOW  
LANGUAGE



**research  
object.org**



Beacon

**ePTRI**

EUROPEAN PAEDIATRIC TRANSLATIONAL RESEARCH INFRASTRUCTURE

EPTRI Stakeholder Round Table– July 9, 2020





# e-resources for EPTRI data

## Electronic documents and data e-library

Resources to *store, cure and preserve all the digital documents and data* produced as a result of research activities during EPTRI, offering a central location where authorised users can upload and download files, in several different forms and formats.

## “Data interoperability” Common Service

Tools for *discovering, accessing, integrating and analysing biological data to facilitate sharing and re-use* of data according to the FAIR principles.

## Text mining and Natural Language Processing tools

Tools *for semantic search and classification to index all documents and tag them with the appropriate metadata*. The final goal is to extract quality and coherent knowledge from digital documents and data. Elixir can offer some solutions

*Elixir-IT, Elixir-LU*



# Common service for data interoperability

- Exploratory phase to collect needs and requirements of the EPTRI community
- Prioritization of the possible interventions
- Definition of standard formats, metadata, ontologies (FAIRsharing) and implementation of new standards when needed
- Adoption of best practices to make existing resources findable and interoperable (Bioschemas, Identifiers.org)
- Inclusion in federated resources for sharing data on human variations (Beacon)
- Training on interoperability best practices to people involved in data management
- Training on best practices to release new software and analysis workflows ensuring reusability (Biocontainers, CWL)



