



ELIXIR : the distributed infrastructure for life-science information

Anne-Françoise Adam-Blondon

► To cite this version:

Anne-Françoise Adam-Blondon. ELIXIR : the distributed infrastructure for life-science information. LERU Open Science Ambassador meeting, Université Paris-Saclay, Nov 2023, Gif-sur-Yvette, France. hal-04503092

HAL Id: hal-04503092

<https://hal.inrae.fr/hal-04503092>

Submitted on 13 Mar 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



 @ELIXIREurope

 /company/elixir-europe

ELIXIR: the distributed infrastructure for life-science information

A-F Adam-Blondon, INRAE, ELIXIR-FR Head of node

LERU Open Science Ambassadors meeting,
Université Paris-Saclay, France, 2023 Nov 9th

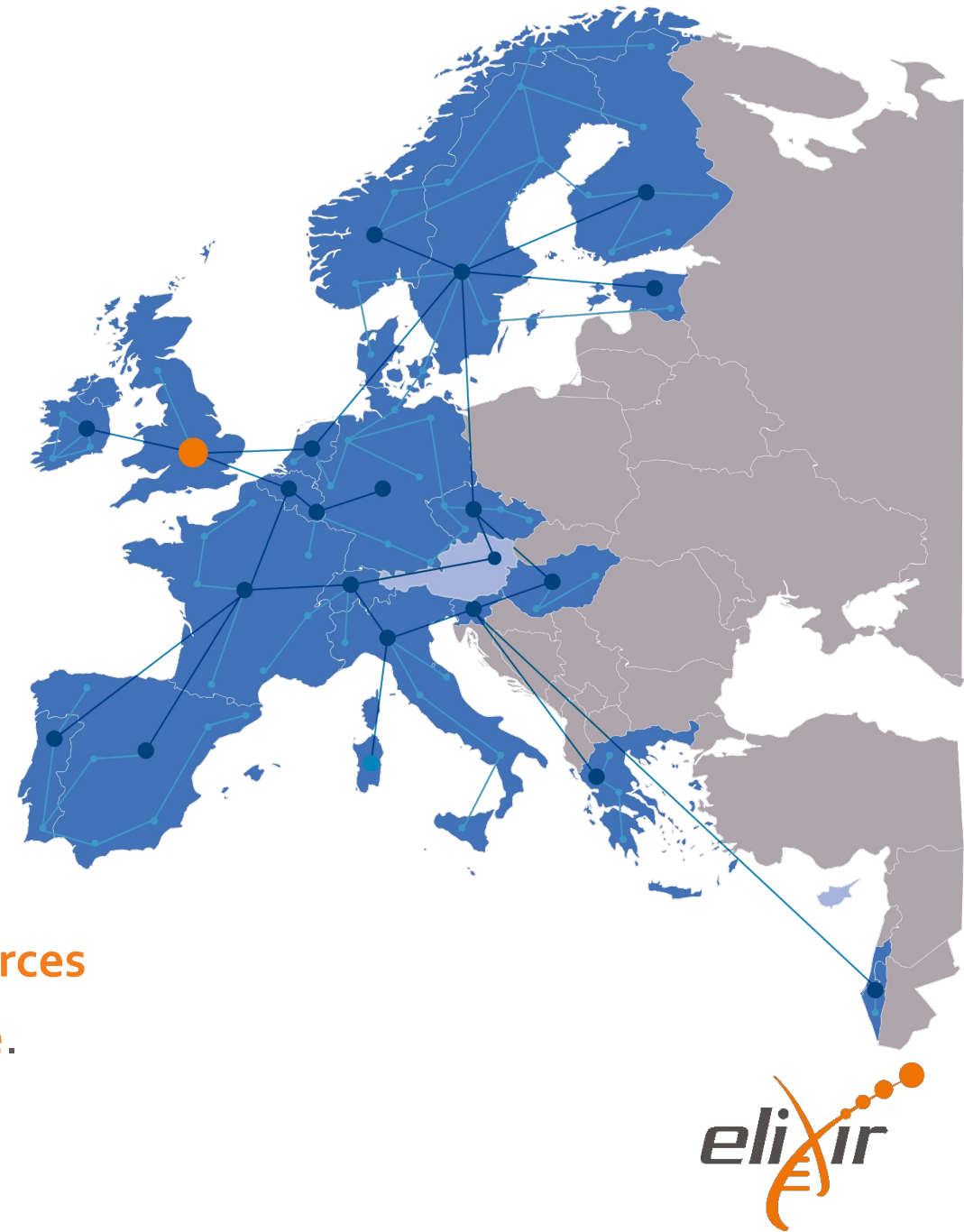
www.elixir-europe.org

ELIXIR Europe

ELIXIR is an intergovernmental organisation that brings together **life science** resources such as

- databases
- software tools
- training resources
- interoperability resources
- compute resources
- data management support

The goal of ELIXIR is to **coordinate bioinformatics resources from across Europe so they form a single infrastructure.**



ELIXIR – who we are

24 Nodes

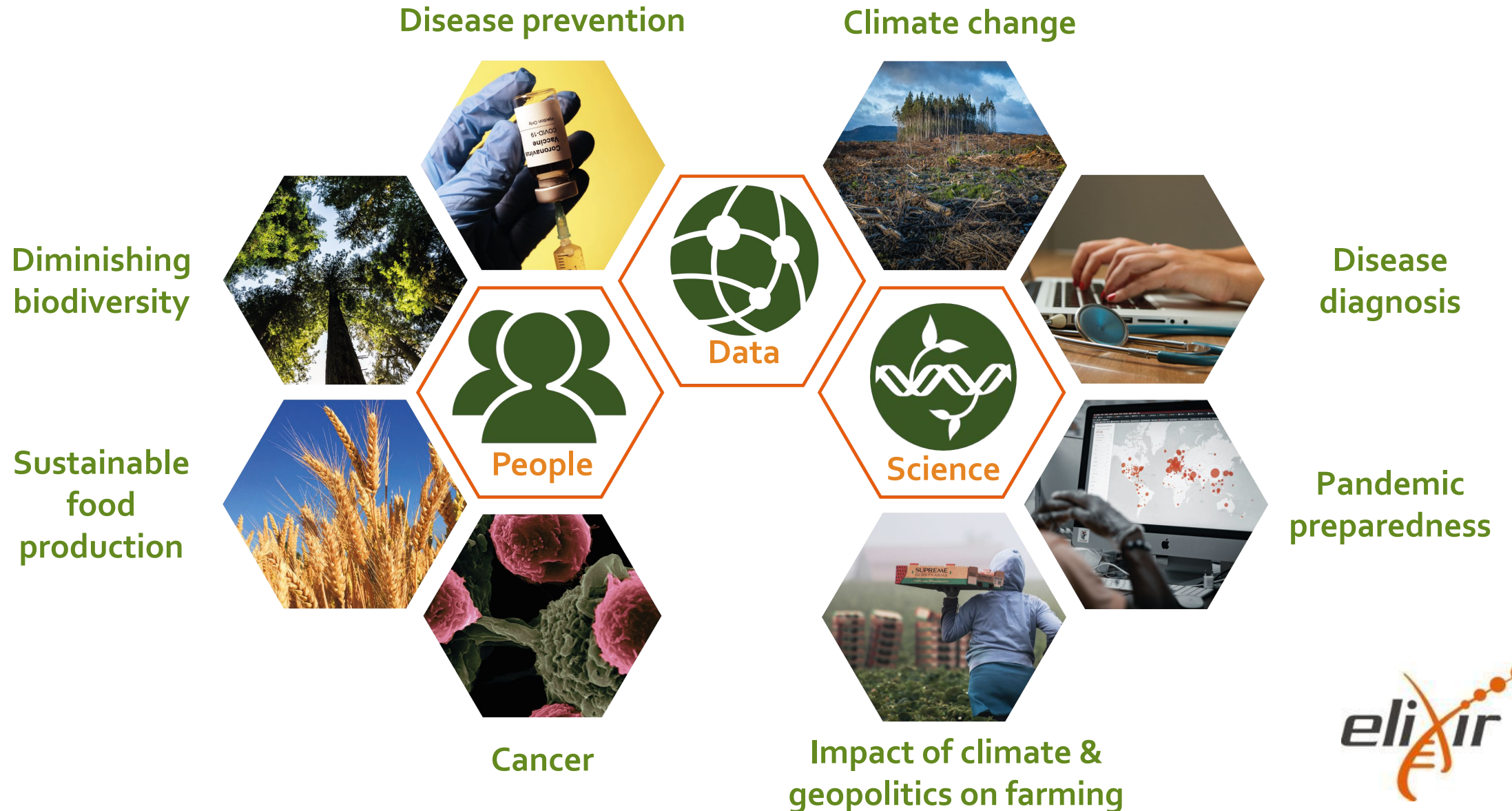
- ELIXIR Nodes form national networks of excellence, containing a total of **245** institutes/universities
- **EMBL-EBI** is one of the nodes

1 Hub

The Hub, located at EMBL-EBI, Cambridge, UK, provides the secretariat and coordinating activities for ELIXIR



Enable data-driven solutions to societal challenges



IFB is a distributed French National Research Infrastructure

- 36 platforms and research teams spread all over France

Coordination: IFB-core

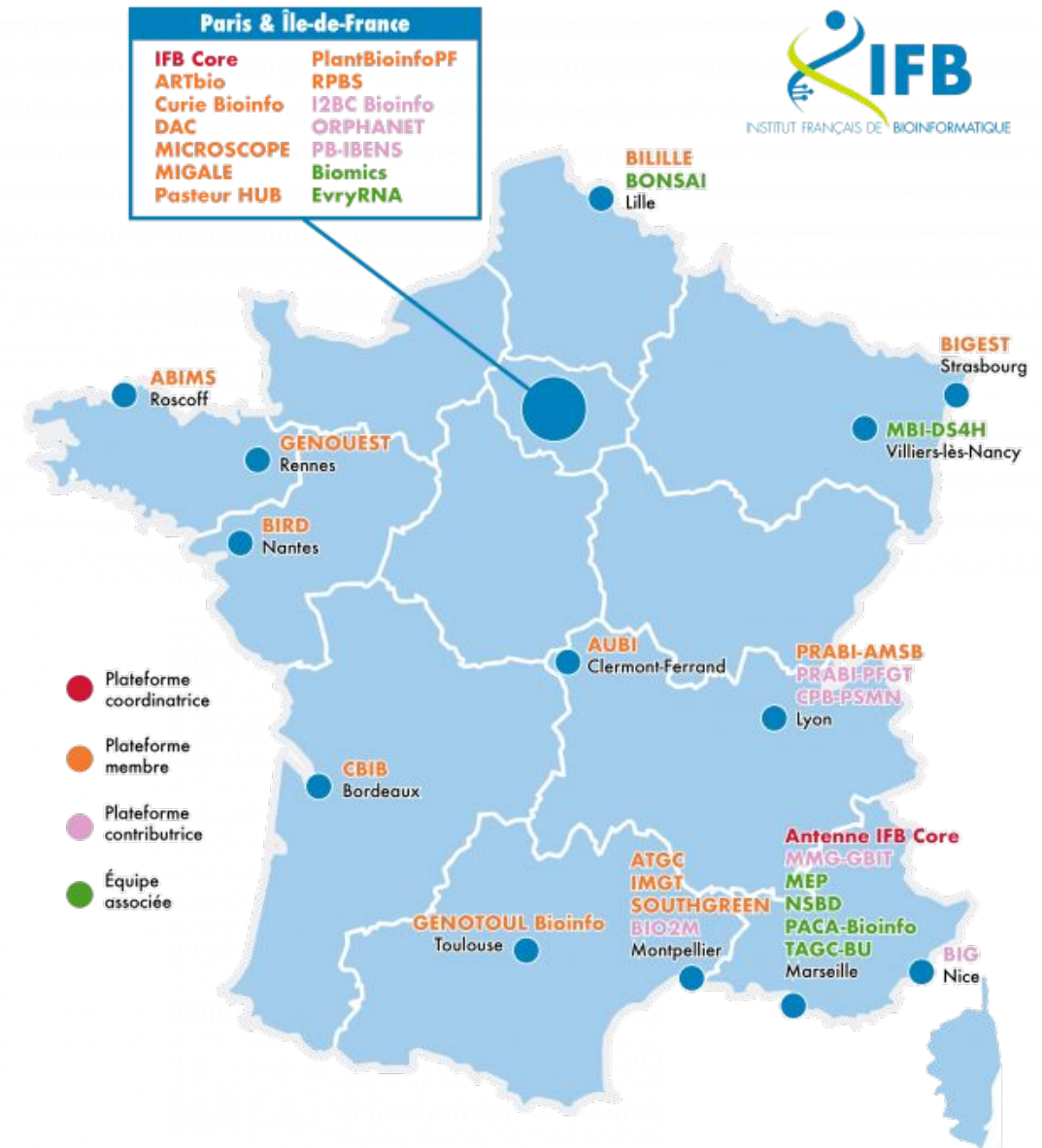
- Multi-research organizations : CNRS, Inserm, INRAE, CEA

Common resources

- Supported by the French Ministry of Higher Education and Research since 2012

A large set of expertise in bioinformatics

- >400 experts (~200 FTE)





IFB's domains of application encompass the three scientific priorities of the 2024-2028 programme of ELIXIR



Enabling and responding to
advances in cellular and molecular
research will drive life science

Connect the latest developments and
established data resources to realise
the potential of cellular and molecular
biology



Biodiversity, food security and
pathogens are interrelated
societal challenges

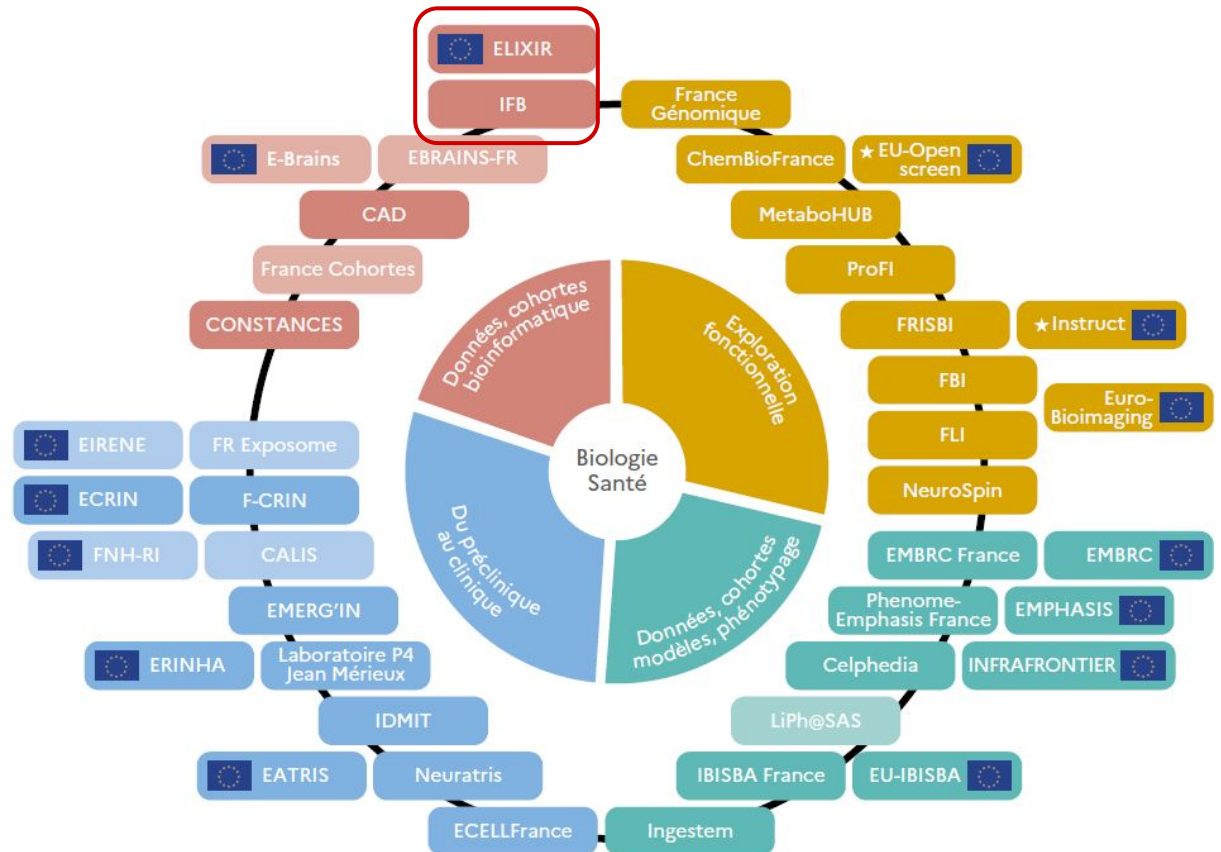
Mobilise and integrate molecular data
to support transnational research
programmes in biodiversity, food
security and pathogens



The promise of genomics
research to improve health
and disease outcomes

Provide the infrastructure to support
the discovery, access, sharing and
analysis of human genomics data and
linked data on a massive scale

- Sharing expertise and working groups on subjects of common interest
- Partnership in large national projects
- Enhance partnerships in ELIXIR activities and in European projects



How we work - ELIXIR Platforms

Platforms bring together experts from Nodes to develop ELIXIR's technical vision and coordinate activities in defined technical areas. There are five Platforms:

Compute

Builds and integrates **cloud, compute, storage and access services** for the life-science research community

Data

Drives the use, re-use and value of life science data by providing **curated, robust, long-term sustainable data resources** within a coordinated, scalable and connected data ecosystem

Tools

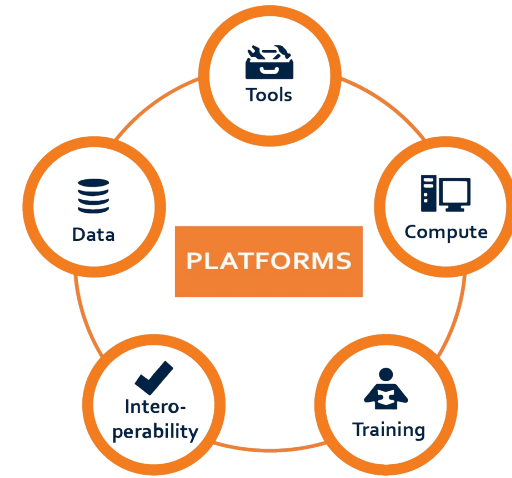
Helps communities **find, register and benchmark** software tools; maintains information standards and produces, adopts and promotes **best practices for tool development**

Interoperability

Helps people and machines to **discover, access, integrate and analyse biological data**; encourages the life science community to adopt **standardised file formats, metadata & vocabularies**

Training

Strengthens **national training** programmes; grows bioinformatics and research data management **training capacity and competence across Europe**; empowers researchers to use ELIXIR's resources



Galaxy

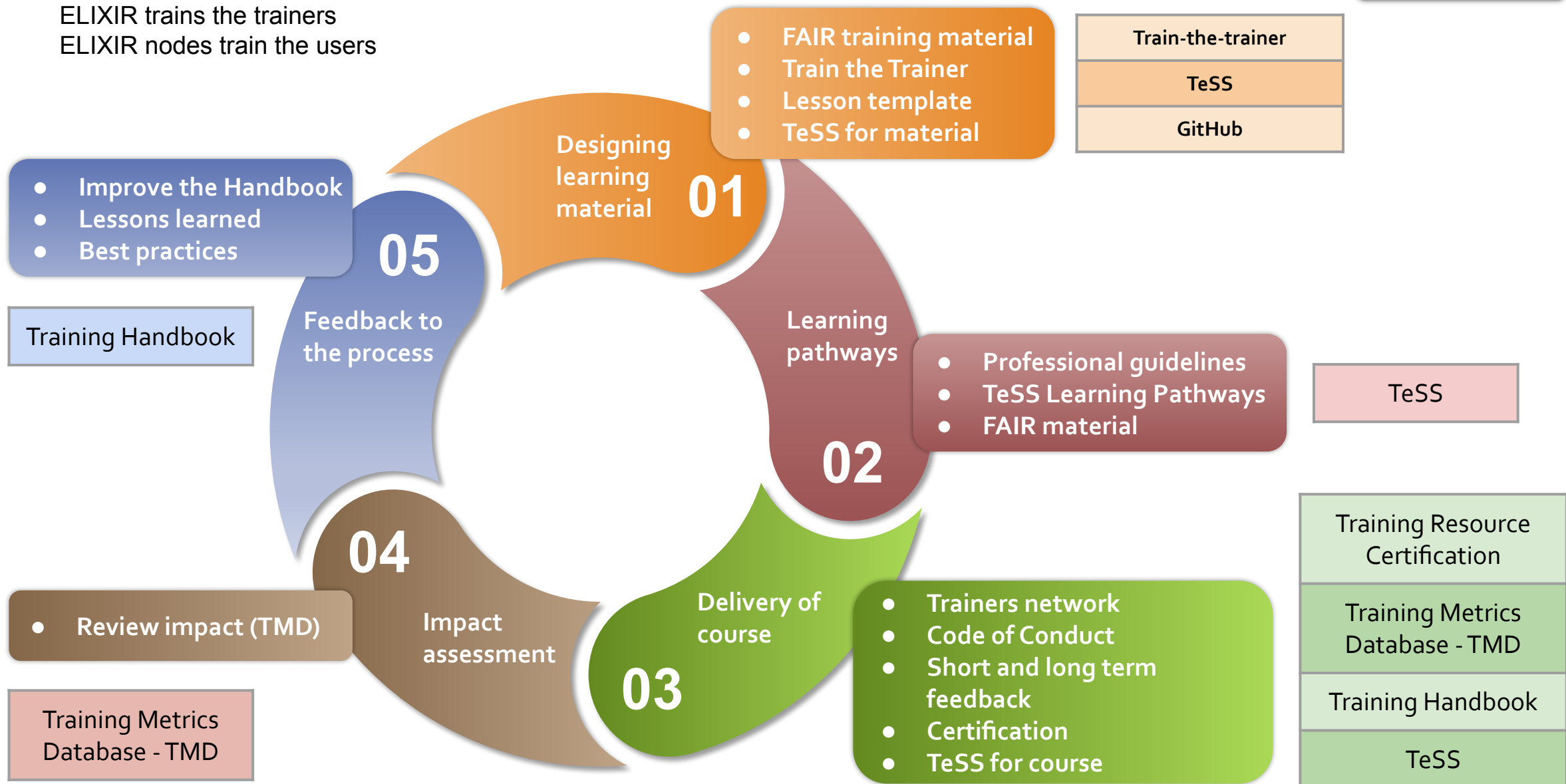


ELIXIR Training Platform

Service/Product

• Activity

ELIXIR trains the trainers
ELIXIR nodes train the users



How we work - ELIXIR Communities: connecting infrastructure & life science experts



Formed around domain experts in ELIXIR Nodes (including non-ELIXIR partners)



Provide a mechanism for long-term collaborations with other ESFRIs and large-scale initiatives



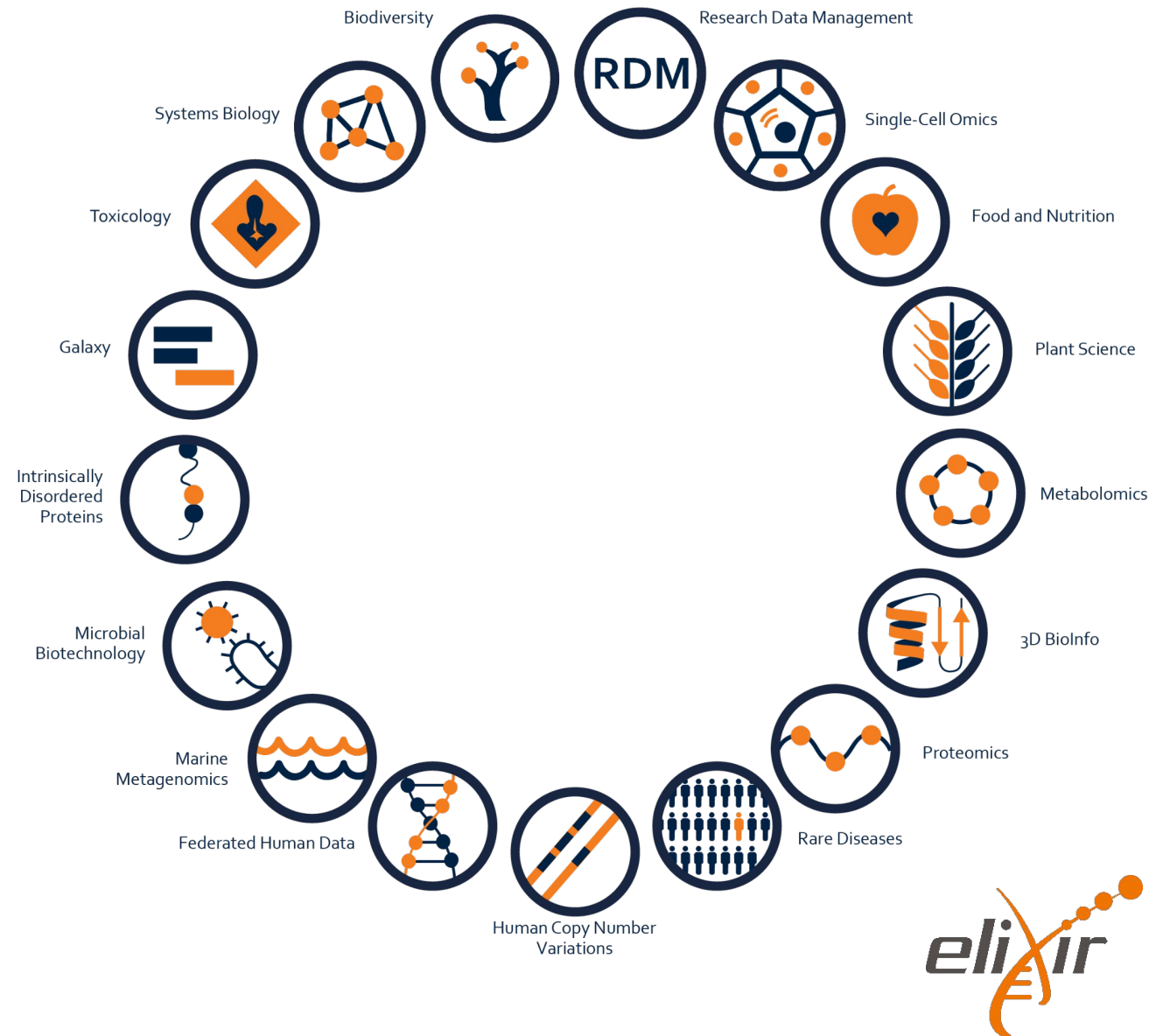
Drive service developments in the ELIXIR Platforms



Provide a framework to develop and maintain community standards



The [ELIXIR Communities Handbook](#) tells you what a Community is, who can join, what the benefits are, and how Communities are structured.



- **Liaising and networking** with all platforms and most of the communities and contributing to their activities
- **Co-coordinating and contributing to the strategic planning of some of them =>** 
- Contributing to **European projects**: 9 since 2020




Technology

Platforms

ELIXIR Platforms bring together experts from Nodes to develop ELIXIR's technical vision and coordinate activities in defined technical areas. There are five Platforms: Data, Tools, Interoperability, Compute and Training.



Tools Platform

Helps researchers find the best software tools to analyse their data. 



Data Platform

Aims to identify key data resources across Europe and support the linkages between data and literature.



Compute Platform

Develops services to make it easier to store, share and analyse large datasets.



Interoperability Platform

Develops and encourages the adoption of standards to describe life science data.



Training Platform

Helps scientists and developers find the training they need, and also provides that training.



Science

Communities



3D-Bioinfo



Biodiversity



Food and nutrition



Galaxy



Intrinsically Disordered Proteins



Marine Metagenomics 



Metabolomics



Microbial Biotechnology



Plant Sciences

litate crop and



Proteomics



Single-Cell Omics



Systems Biology



Toxicology



R

Human Data Communities



Federated Human Data




Human Copy Number Variation 



Rare diseases



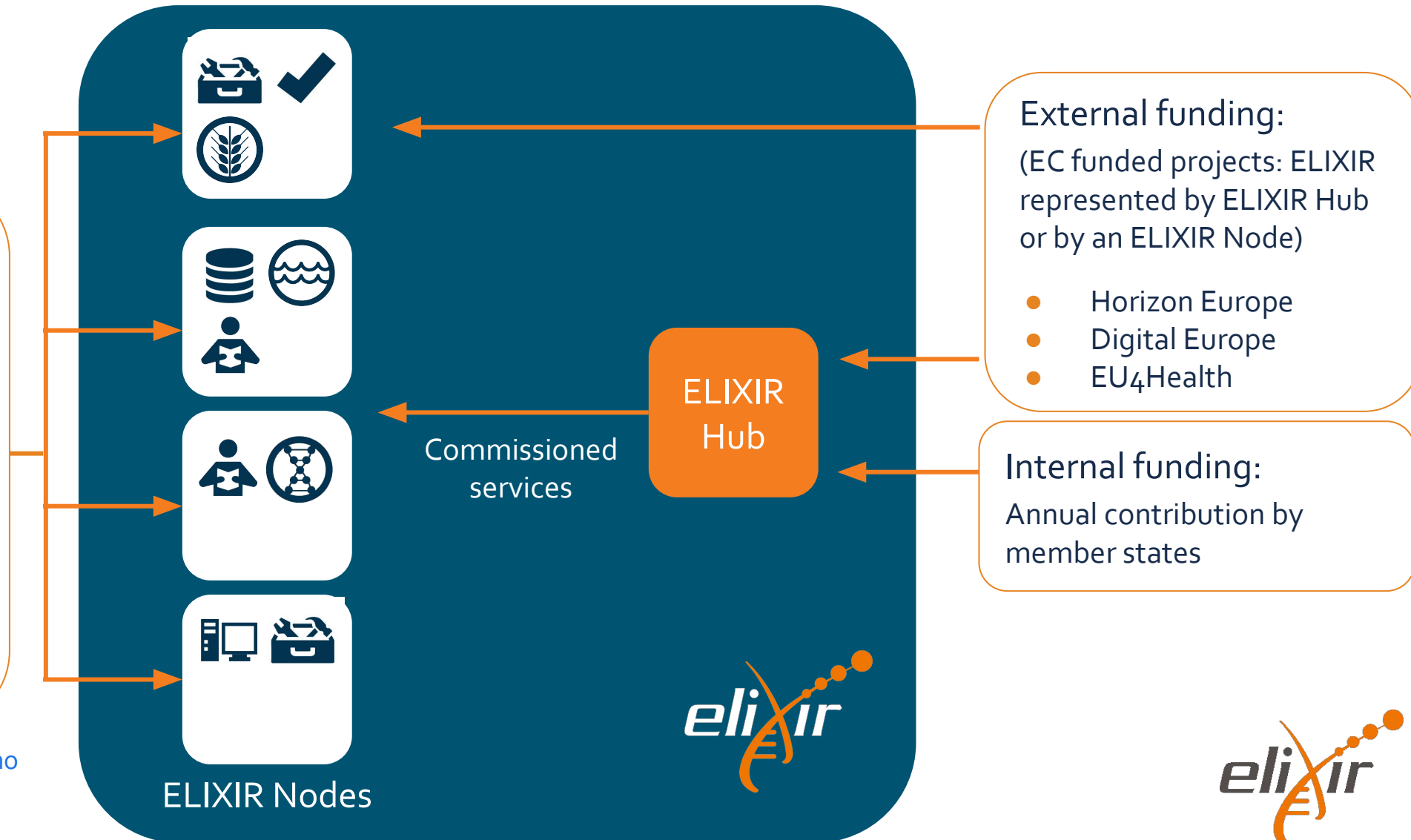
Cancer Data (focus group) 

ELIXIR's funding model

Node funding:

- National roadmap funding
- Competitive research grants
- EU Structural Funds or Recovery and Resilience Facility (RRF)
- Trusts and foundations
- Industry collaboration

<https://elixir-europe.org/about-us/how-funded/sustainability-plan>



ELIXIR's key stakeholders



ELIXIR Nodes

24 ELIXIR Nodes
245+ institutes
850+ scientists



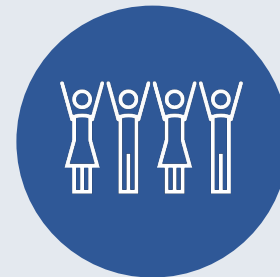
Users

Bioinformaticians
500,000+ life science researchers
Users in industry



Funders & decision makers

European Union
National funding agencies
ESFRI delegates
ELIXIR Board members



Collaborators

ESFRI RIs
Global initiatives (GA4GH, Galaxy)
National initiatives (Australian
BioCommons, NIH)

Collaborator example – the ELIXIR EOSC strategic vision

ELIXIR maps to EOSC at all levels

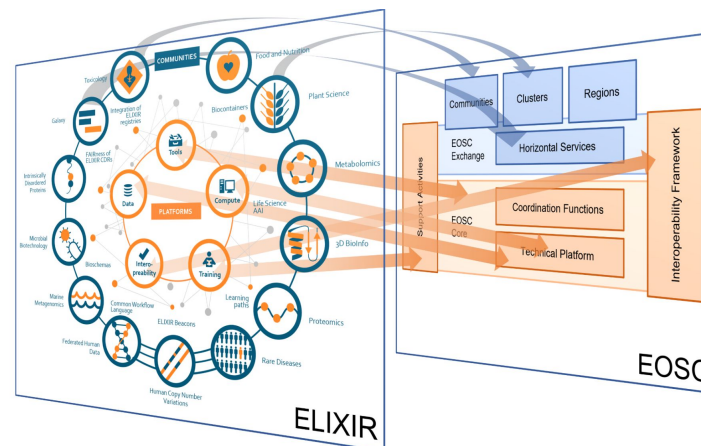
- Throughout the EOSC Core and Interoperability frameworks
- Bidirectional relationships, with best of breed solutions being adopted by both sides
- ELIXIR Communities will over time adopt EOSC Services and become driving user communities for both organisations



ELIXIR and the EOSC Association Task Forces

ELIXIR on Advisory Groups

- ✓ The Implementation of EOSC
- ✓ Metadata and Data Quality
- ✓ Research Careers and Curricula
- ✓ Sustaining EOSC
- ✓ Technical Challenges on EOS



ELIXIR involvement in EOSC projects



EOSC-Pilot



EOSC-Enhance



EOSC-Life



EOSC-Future



ELIXIR flagship events

BioHackathon Europe

- Bringing together bioinformaticians for **five days of hacking**
- **Five successful annual events** since 2018
- Projects to advance **open source infrastructure** for data integration to accelerate scientific innovation
- Supporting operations across ELIXIR Platforms, Communities and Focus Groups through technology implementations (e.g. **FAIR, identifiers, metadata standards, ontologies and metadata catalogues**)



Innovation and SME Forums

- Approximately **two one-day** events per year
- Providing **Small to Medium Sized Enterprises** (SMEs) the opportunity to present their innovative ideas
- Enabling companies to learn more about **current and emerging ELIXIR services**
- Forging strong links with the **local ELIXIR Node representatives** running ELIXIR services



ELIXIR Bioinformatics Industry Forum (EBIF)

- **One-day** annual event
- Discussions with industry experts around **visionary ideas, bottlenecks and solutions** to major challenges in the data-driven life science sector
- A forum for **knowledge exchange and collaboration** in the pre-competitive space
- **Networking** opportunities with bioinformatics opinion leaders, academic experts in ELIXIR and the commercial sector



ELIXIR's Code of Conduct – the principles



The Code of Conduct pertains to ELIXIR organised or funded events



We value
each other's perspectives

We adopt
a zero-tolerance approach to harassment
and discrimination

We maintain
high ethical standards

We'll apply
honesty and integrity in the dealing of any
transgressions against the Code.

We're committed
to making ELIXIR events collaborative,
supportive and enjoyable

We'll ensure
a respectful and inclusive environment



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

Examples of ELIXIR high impact services

	Life Science Login	Authentication service		Users
	FAIRsharing	Data and metadata resource including standards, databases and policies		Standards
	ELIXIR TeSS	ELIXIR's training portal		Training materials
	BioContainers	Software standardisation resource		Containers & packages
	BioTools	Registry of tools, databases and services		Monthly visits

ELIXIR research data management resources

Guidance for data stewards, project managers and researchers

Overview of good data management practices



The Research Data Management Kit (RDMkit) guides you through the whole data management life cycle and includes advice specific to your domain, your role and your country.

Step-by-step instructions



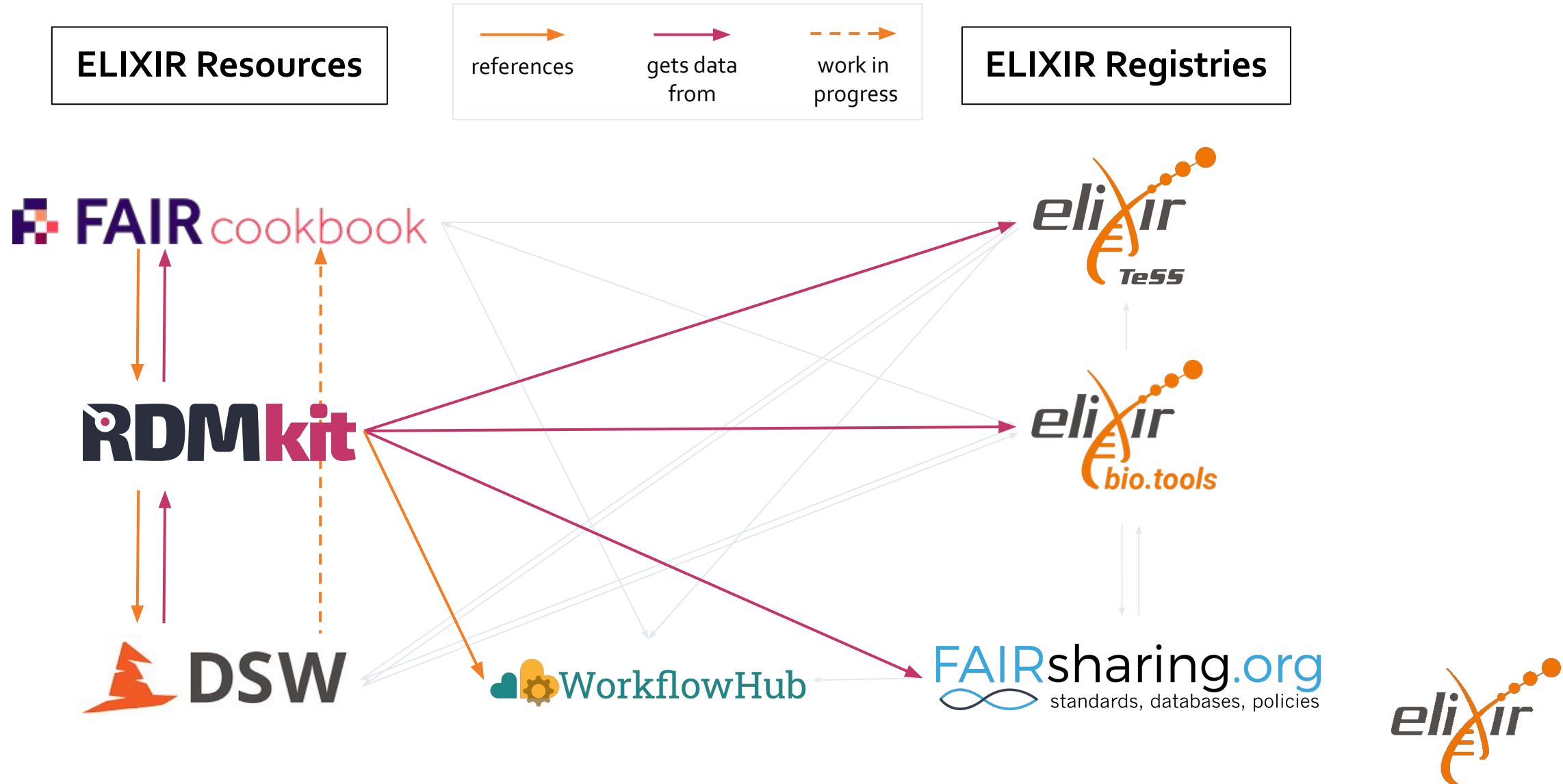
The FAIR Cookbook contains step-by-step recipes to accomplish specific data management tasks and to make your data FAIR (Findable, Accessible, Interoperable, Reusable).

Data management plan wizard



The Data Stewardship Wizard (DSW) is an online tool that guides researchers and data stewards through their data management planning.

Research Data Management Toolkit



A web-based toolkit for the
bioscience community written by
the bioscience community

RDMkit in numbers

151

Contributors
The force behind RDMkit



336

Tools & resources
Explained in the context of
real world problems



101









Pages
Helping you with data
management



What can we help you find?

Search RDMkit

Browse all topics by

- 
Data life cycle
Start here to get an overview of research data management based on stages in the data life cycle.
- 
Your role
Identify your role in research data management, find data management resources relevant for you, and information to help you progress in your career path.
- 
Your domain
Learn about data management tasks that affect your domain or research community, and the solutions adopted to address them.
- 
Your tasks
Find guidelines and solutions for tackling common data management tasks.
- 
Tool assembly
Find concrete combinations of tools and resources assembled into an ecosystem for research data management.
- 
National resources
Find pointers to country specific information resources and national research data management practices.
- 
All tools and resources
Browse the RDMkit's catalogue of tools and resources for research data management.
- 
All training resources
Browse all training resources mentioned in RDMkit pages.

Resources for reproducible environment of analysis

bio.tools



bio.tools helps you find and select bionformatics 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

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 Brack¹, Peter Crowther², Stian Soiland-Reyes^{1,3*}, Stuart Owen¹, Douglas Lowe⁴, Alan R. Williams¹, Quentin Groom⁵, Mathias Dillen⁶, Frederik Coppens^{6,7}, Björn Grüning⁸, Ignacio Eguinoa^{6,7}, Philip Ewels⁹, Carole Goble¹

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

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

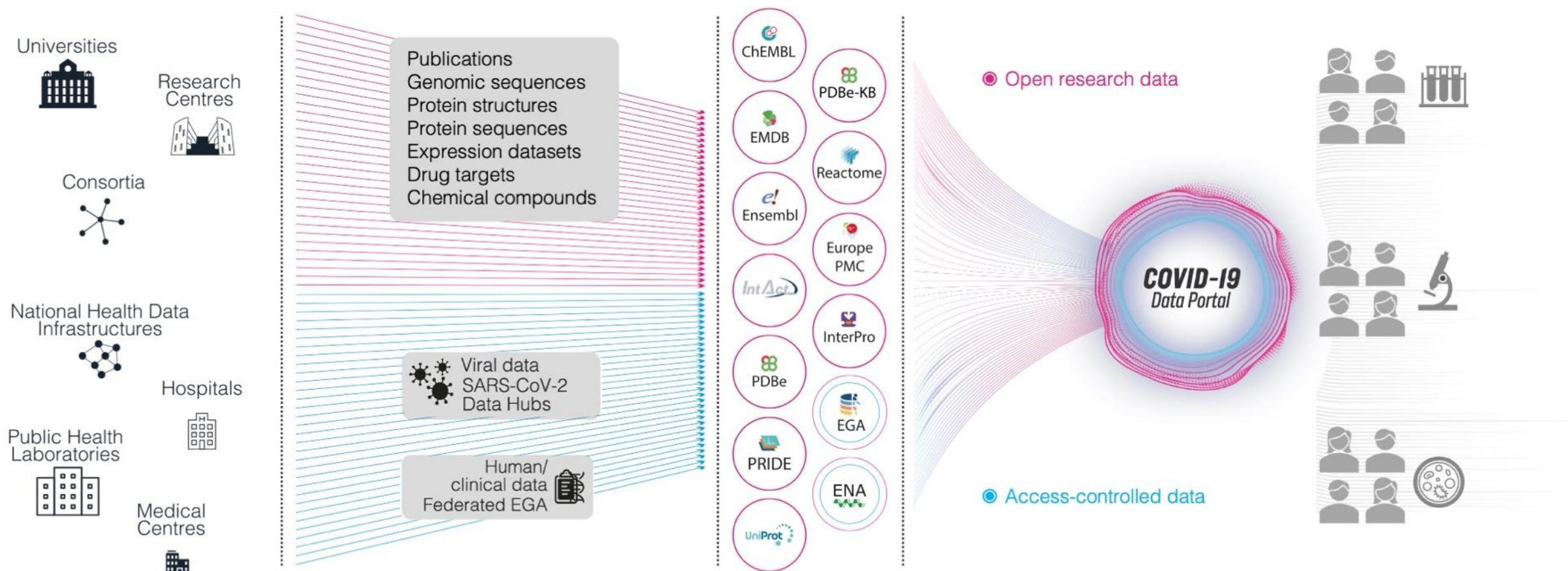
There is an urgent need to improve the infrastructure supporting the reuse of scholarly data. A diverse set of stakeholders—representing academia, industry, funding agencies, and scholarly publishers—have come together to design and jointly endorse a concise and measurable set of principles that we refer 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



The European COVID-19 Data Platform



<https://www.covid19dataportal.org>



Guy Perrière
ELIXIR-FR Head of Node



Anne-Françoise Adam-Blondon
ELIXIR-FR Head of Node



Angela Sáenz
ELIXIR-FR Node coordinator



Many thanks to the ELIXIR hub
team for their slide deck!

