ELIXIR Resources

Enabling life science researchers to access and analyse data

Peter Maccallum, ELIXIR Chief Technical Officer

European Biomedical Research Infrastructures' resources for Researchers in Neurosciences, 1st Feb 2022

www.elixir-europe.org

ELIXIR Life Science Resources

ELIXIR is Europe's Infrastructure for Life Science Data.

- The concept was developed in close collaboration with the European Bioinformatics Institute and consultation with the EU
- Development has been aligned with the European Strategy Forum on Research Infrastructures (ESFRI) and ELIXIR has been an ESFRI 'landmark' since 2016

Life Science Data seems quite broad?

- Our core is the 'omics databases and tools
- In some topics - human data, imaging - other ESFRI partners have strengths and we collaborate
- Our early focus on Open Science and FAIR have made us significant contributors to FAIR toolkits and the European Open Science Cloud (EOSC)

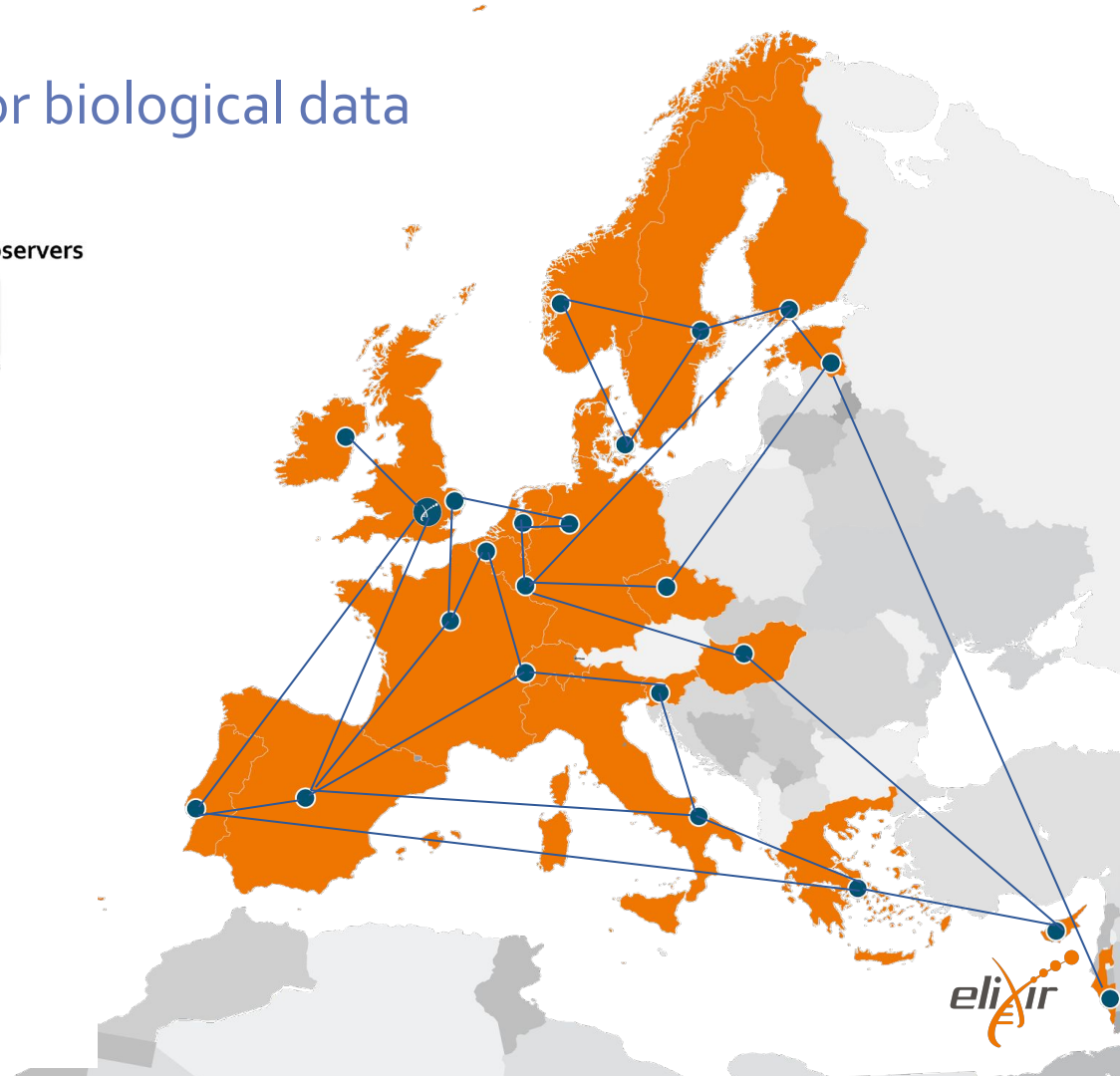
A sustainable infrastructure for biological data

ELIXIR Members

Belgium
Czech Republic
Denmark
EMBL
Estonia
Finland
France
Germany
Greece
Hungary
Ireland
Italy
Luxembourg
Israel
Netherlands
Norway
Slovenia
Portugal
Spain
Sweden
Switzerland
United Kingdom

ELIXIR Observers

Cyprus



ELIXIR's Technical Framework - some terminology

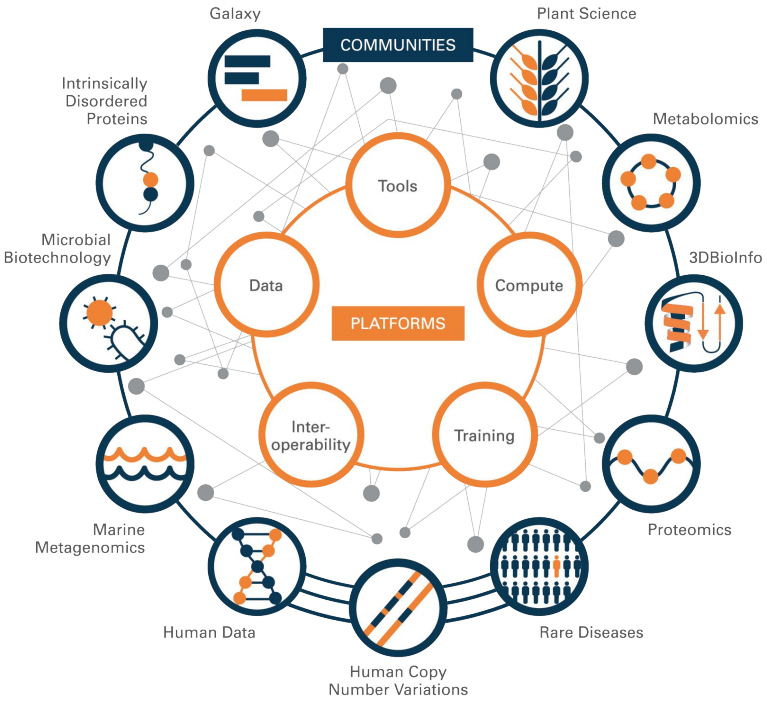
ELIXIR offers over 300 distinct services, form specialist tools to full research environments.

We group together technical aspects of these services through our ELIXIR Platforms

- Compute, Data, Interoperability, Tools and Training

We collect the driving use cases and specialist topics through the ELIXIR Communities

https://elixir-europe.org/services/list



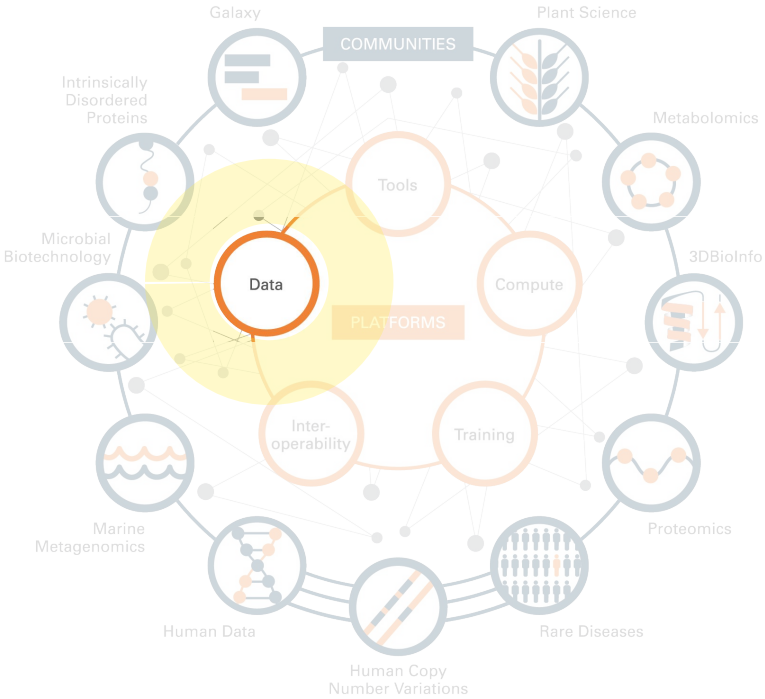
ELIXIR Data Platform - sustainable data resources

The goal of the ELIXIR Data Platform is to drive the use, re-use and value of life science data.

- for bioinformaticians and life science researchers
- users in academic and industrial settings
- sound governance, life cycle management, and long-term sustainability

Researchers need open access to technically and scientifically excellent data resources for effective data discovery, deposition, and re-use.

- Open Access as a core principle for publicly funded research.
- terms of use or a licence that enable the reuse and remixing of data (see Open Definition)



Data for Life: Sustaining Europe's life science data infrastructure

Mission: drive long-term sustainability of life science data resources

ELIXIR Core Data Resources and Deposition Databases - Resources of fundamental importance for the research community and the long-term preservation of life-science data.

- Basis for policy actions
- Basis for technical actions
- Capacity building
- Life cycle management

Core Data Resource	Data type
ArrayExpress	Functional Genomics Data from high-throughput functional genomics experiments.
BRENDA	Database of enzyme and enzyme-ligand information, across all taxonomic groups, manually extracted from primary literature and extended by text mining procedures, integration of external data and prediction algorithms.
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 vertebrate s
Europe PMC	Europe PMC books, pater
Human Protein Atlas	The Human protein-codi proteins bas
The IMEx Consortium: represented by IntAct and MINT	IntAct provi molecular in interactions
	Functional a

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 da protein expression information values) and the suppo	

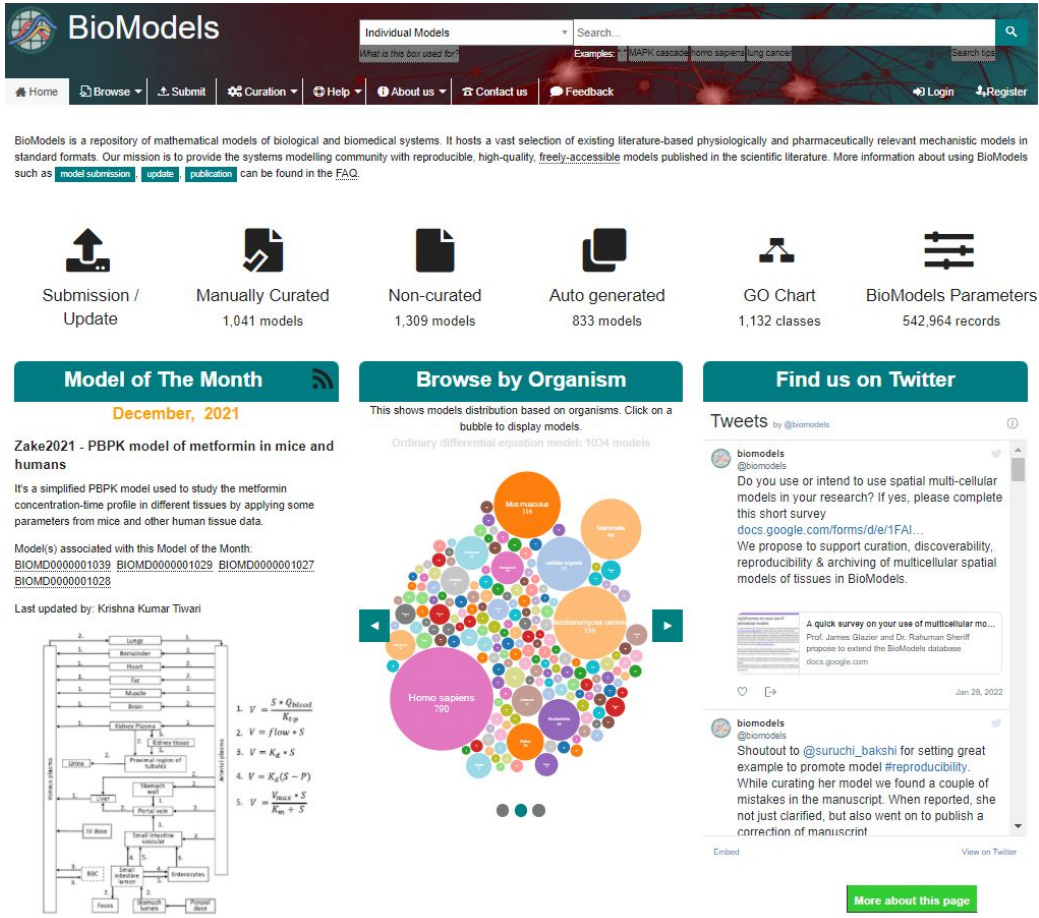
https://elixir-europe.org/core-data-resources
https://elixir-europe.org/recommended-deposition-databases

ELIXIR Deposition Database example - BioModels

The BioModels ELIXIR Deposition Database allows users to deposit and re-use physiologically and pharmaceutically relevant mechanistic models in standard formats

'BioModels is a repository of mathematical models of biological and biomedical systems… Our mission is to provide the systems modelling community with reproducible, high-quality, freely-accessible models published in the scientific literature.'

https://www.ebi.ac.uk/biomodels/



ELIXIR Interoperability Platform - FAIR and Open Science

The goal of the ELIXIR Interoperability Platform is to help people and machines to discover, access, integrate and analyse biological data.

It encourages the life science community to adopt

- standardised file formats,
- metadata,
- vocabularies and
- Identifiers

...and works internationally to achieve its goals.

Identifies and promotes interoperability best practices for data providers and data integrators.

Delivers the interoperability services that underpin ELIXIR Communities and Platforms.

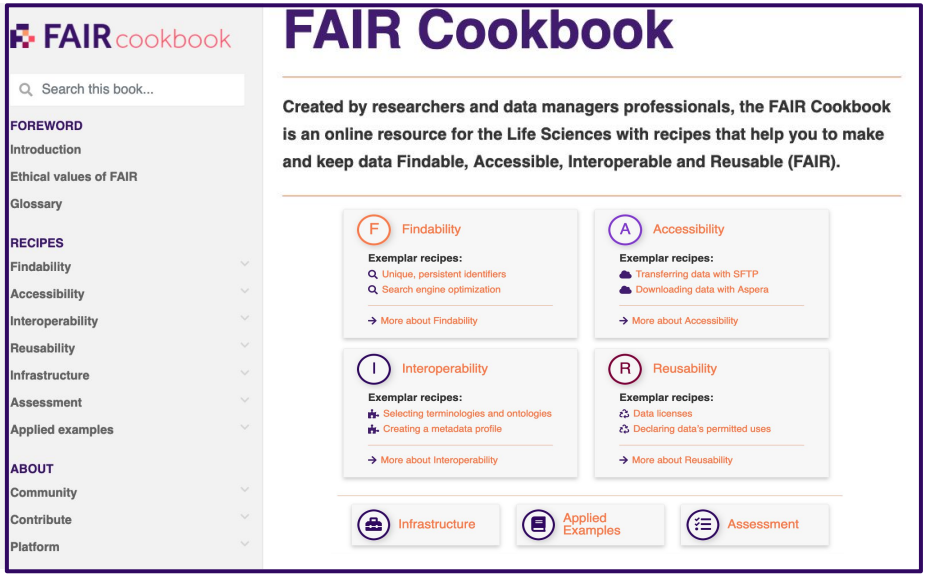
Interoperability - Recommended Interoperability Resources

- ELIXIR's Recommended Interoperability Resources.
- These are resources that:
  - Establish connections between resources
  - Acquire and expose metadata of different resources
  - Create the infrastructure needed to build integratable data collections
- They will help you translate your data across databases or resources or find the best standards for your data.

Resource	Description
FAIRsharing	A registry of the curated metadata of databases, policies and standards.
g:Profiler	A gene-centric data integrator with web UI and API services.
Identifiers.org	An identification and resolution system for life science, and a provider of compact identifiers and URIs.
InterMine	Framework to integrate life sciences data based on an extensible data model, providing web interface and RESTful web services.
ISA Framework	ISA (Investigation > Study > Assay) helps researchers to provide rich description of experimental metadata so that the resulting data and discoveries are reproducible and reusable.
Ontology Lookup Service	Repository for biomedical ontologies that aims to provide a single point of access to the latest ontology versions through web UI or RESTFUL API
3DBIONOTES API	A reusable platform-independent API call component for protein metadata alignment, annotation, and integration across major protein data resources.
BridgeDb	A combination of a software framework and an API for mapping identifiers for related objects in life sciences.
DisGeNET API	An API SPARQL Endpoint for genetic variants (hu
MOLGENIS	A software package to help res supports data queries

elixir
Recommended Interoperability Resources

Interoperability example: FAIR Cookbook



https://fairplus.github.io/the-fair-cookbook

FAIRplus

IMI FAIRplus, private-public partnership:

- Addresses common FAIRification problems
- Uses data from other IMI projects: from clinical (observational & event based) to molecular data

FAIR Cookbook *flagship output*:

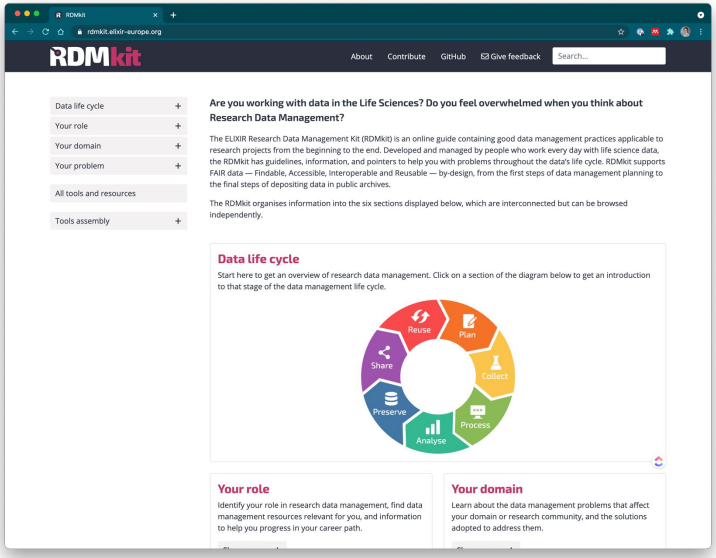
- 50+ recipes a- guidance, technical, hands-on & reviews
- 50+ contributors, incl many ELIXIR Nodes
- Part of ELIXIR FAIR services ecosystem: RDMkit, FAIRsharing, biotools
- Complementary to Pistoia Alliance FAIR Toolkit

FAIR Cookbook already adopted by IMI in their RDM guidance

Interoperability example: RDMkit — supporting FAIR-by-design

Designed to enable data lifecycle management

Website focal point for guidance, information, best practice, examples



RDMkit

https://rdmkit.elixir-europe.org

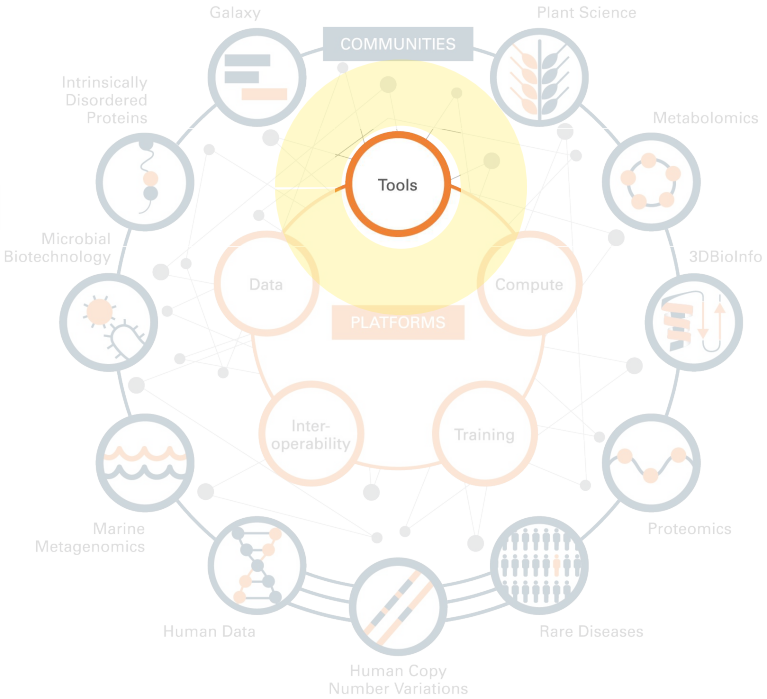
elixir CONVERGE

ELIXIR Tools Platform - an ecosystem for analysis

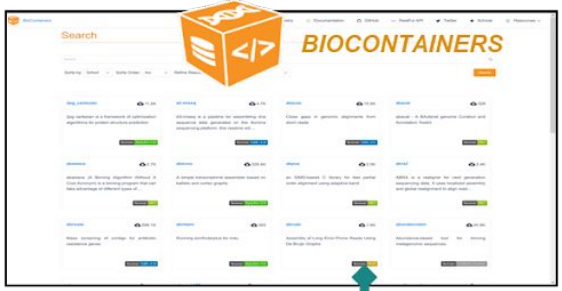
The ELIXIR Tools Platform helps communities find, register and benchmark software tools. These tools help researchers access, analyse and integrate biological data, and so drive scientific discovery across the life sciences.

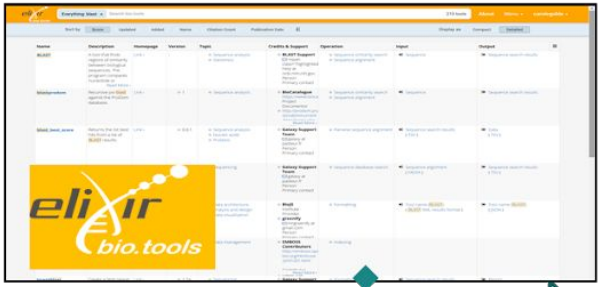
We maintain information standards for these tools, and produce, adopt and promote best practices for their development. We also:

- connect tools and data services
- link tools to training materials
- make it easy to download, deploy, and benchmark tools
- put tools in the context of workflows, including container technologies and Galaxy.



Tool and Workflow Registries





COMMON WORKFLOW LANGUAGE

Workflow RO-Crate

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

Bioschemas.org

EDAM

Metadata & API Framework

Testing

OPENEBENCH

Life Monitor

Workflow Systems

nextflow

Galaxy

SCIPION

Containers

docker

Workflow Repos

GitHub

Galaxy EUROPE

nf-core

Tool Packages

Bioconductor

debian med

IUC

BIOCONDA

elixir

Tools example: bio.tools

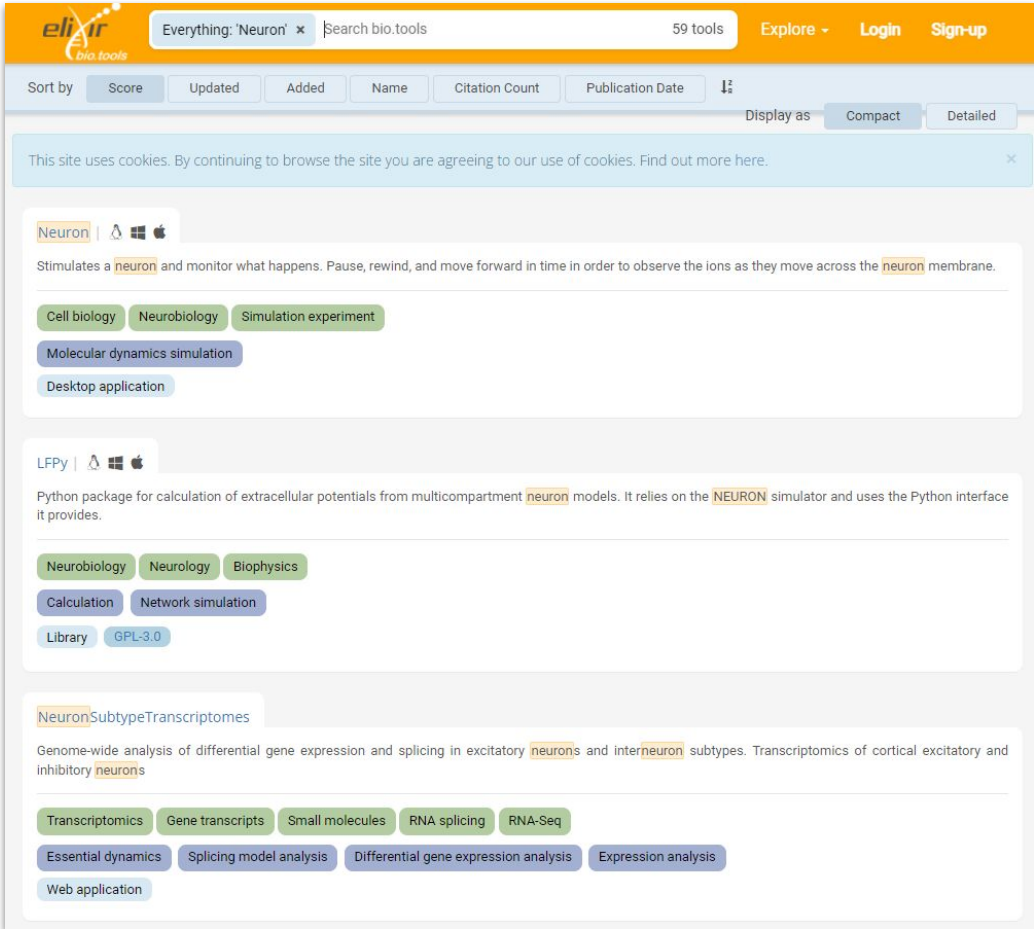
Bio.tools is a discovery portal for bioinformatics software information, providing a curated, annotated description of tools and data services.

'a comprehensive registry of software and databases,

- *…simple command-line tools and online services,*
- *…databases and complex, multi-functional analysis workflows*
- *…described in a rigorous semantics and syntax,*
- *…concise, consistent and therefore comparable information.'*

The EDAM Ontology of bioscientific data analysis and data management enables the search and comparison.

https://bio.tools/
https://edamontology.org/page

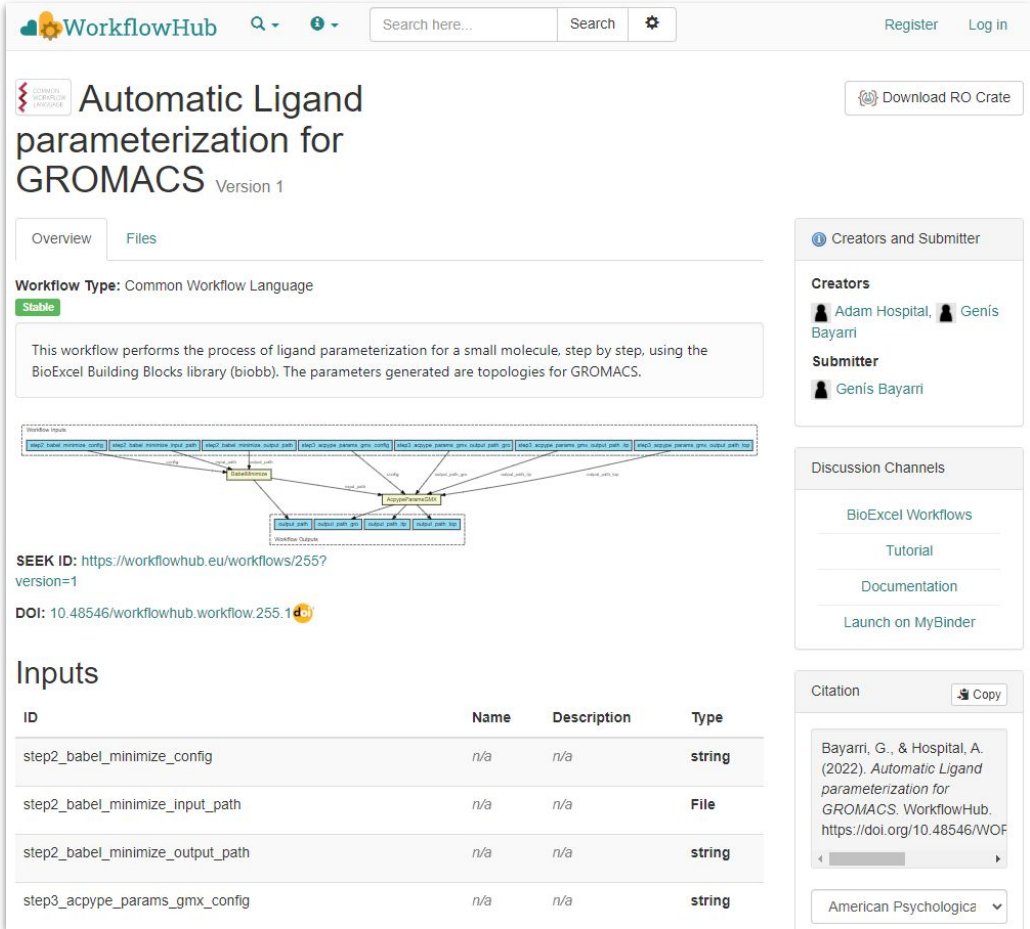


Tools example: WorkflowHub

Developed though the EOSC-Life project, WorkflowHub adds a registry for workflows to the ELIXIR Tools Ecosystem.

- *'describing, sharing and publishing scientific computational workflows.*
- *supports any workflow in its native repository.*
- *use of open standards and tools, including Common Workflow Language (CWL), RO-Crate, BioSchemas and (GA4GH) Tools Registry Service'*

https://workflowhub.eu/



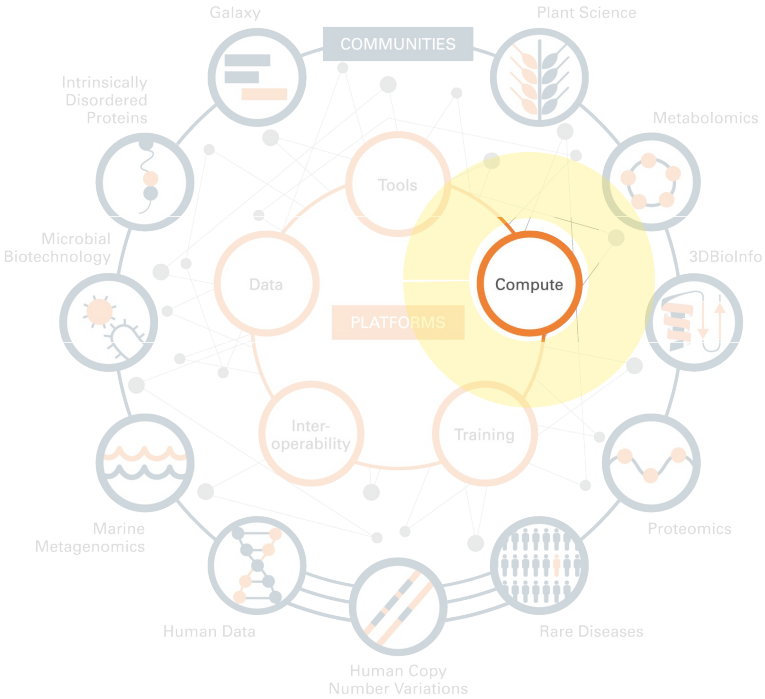
ELIXIR Compute Platform - data-centric computing solutions

The ELIXIR Compute Platform was established to build and integrate cloud, compute, storage and access services for the life-science research community.

Downloading and analysing data locally is no longer viable due to both the data size and scope of the analysis activities.

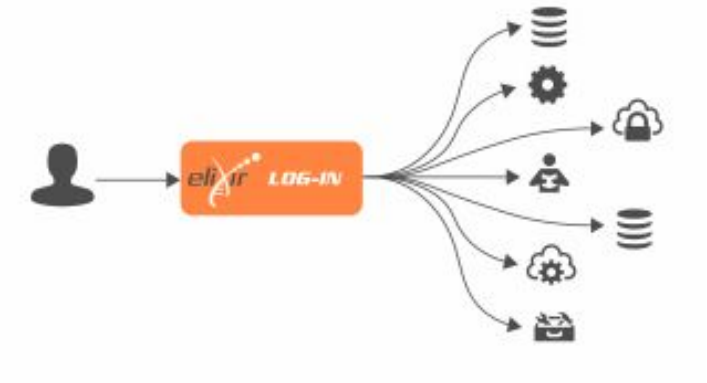
- data needs to be managed as a federation
- data providers work as a single infrastructure
- researchers can bring their analysis to where the data is located

The objective is to combine all components of the ELIXIR Compute services into a seamless workflow.



Compute example: Authentication & Authorisation

- Majority of life science services are openly accessible but may require researchers to sign in using a username and password
- Sensitive data and licensed resources of course require strong security for access
- Researchers quickly become overloaded with having to remember numerous login credentials



ELIXIR Authorisation and Authentication Infrastructure (AAI)

- Enables researchers to use their home organisation credentials - or community or commercial identities (e.g. ORCID, LinkedIn) - to sign in and access data and services they need
- allows service providers to control and manage the access rights of their users and create different access levels for research groups or international projects

Look for the 'elixir LOG-IN' logo on services.

Compute example : EBI Embassy Cloud

Compute Platform services include Data-centric Cloud environments in Nodes.

'Embassy Cloud users (tenants) have direct access to the EMBL-EBI data, services and compute. This is a practical and cost-effective alternative to replicating services and downloading vast, public datasets locally. Tenants can access their workspace from anywhere in the world, reducing the need for capital investments in hardware and related operational costs.'

https://www.embassycloud.org/

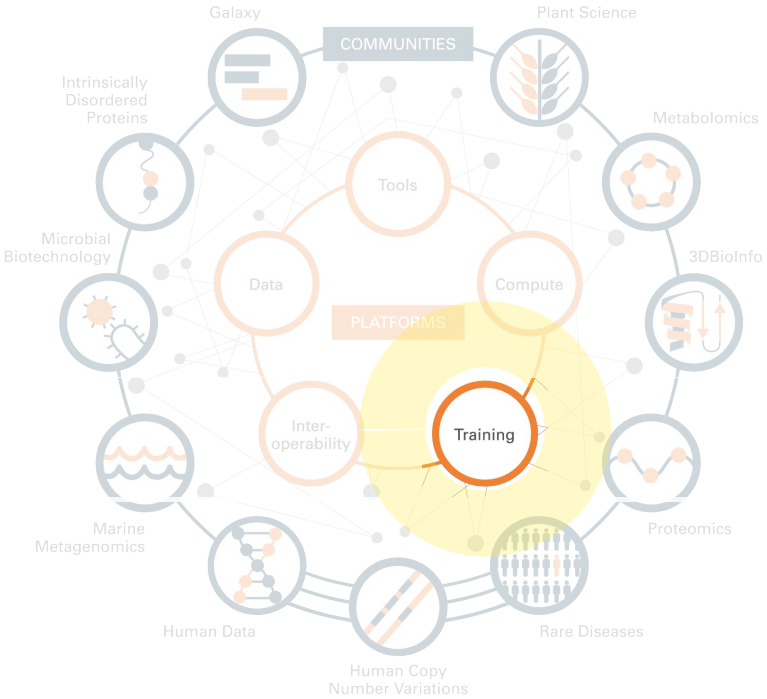
ELIXIR Training Platform

The ELIXIR Training Platform was established to develop a training community that spans all ELIXIR member states (see the list of Training Coordinators). It aims to:

- strengthen national training programmes,
- grow bioinformatics training capacity and competence across Europe, and
- empower researchers to use ELIXIR's services and tools.

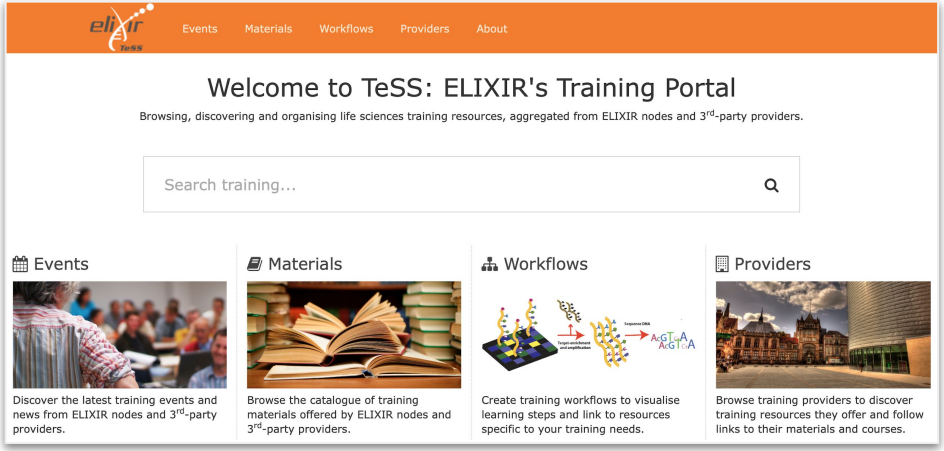
The Platform:

- establishes and implements best practices in bioinformatics training through a training Toolkit
- supports training providers across Europe in creating and delivering training for developers, researchers and trainers
- builds a sustainable training infrastructure



Training example: TeSS

- Need to find training courses or materials?
- ELIXIR's Training e-Support System (TeSS) Platform enables you to search, share & package training resources, training materials and events
- Any content providers can upload and include training courses or content

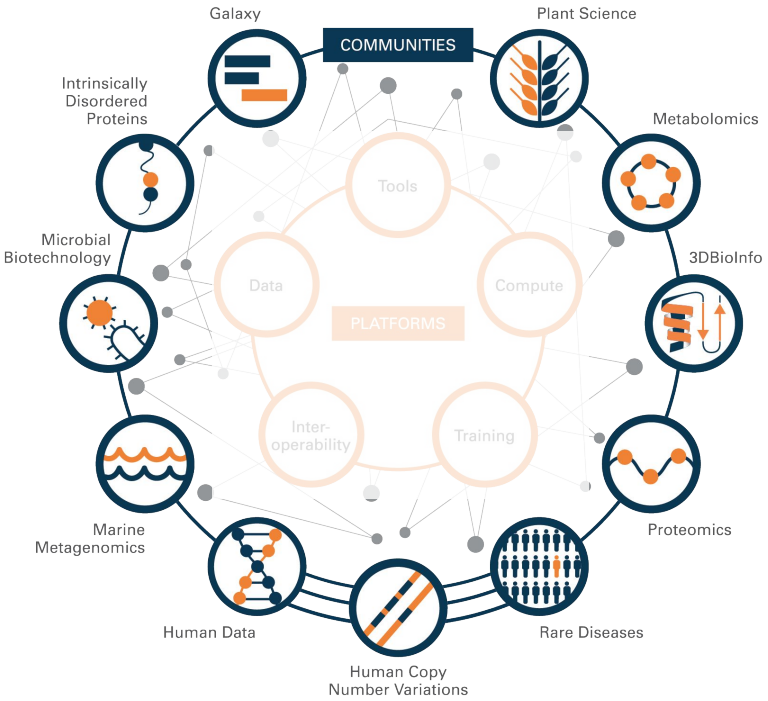


Search for events held near you...

https://tess.elixir-europe.org

ELIXIR Communities

- 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 framework to develop and maintain community standards



ELIXIR Proteomics, Metabolomics Communities

The goal of the ELIXIR Proteomics Community is to develop and maintain sustainable proteomics tools and data resources.

- An essential part of the development will also be the 'FAIRification' of the resources.

The goal of the ELIXIR Metabolomics Community is to work with experimental scientists and developers to provide the resources, analysis tools and infrastructure that will help metabolite identification.

- The community will establish an infrastructure of services, standards and datasets.

F1000Research F1000Research 2017, 6:875 Last updated: 21 JAN 2022

Check for updates

OPINION ARTICLE

A community proposal to integrate proteomics activities in ELIXIR [version 1; peer review: 2 approved]

Juan Antonio Vizca
Wout Bittremieux
Myriam Ferro[4], Al
Lydie Lane[10,11], Kr
Petr Novak[15], Mag
Veit Schwämmle
Jiri Vondrasek[9], M
Oliver Kohlbacher

[1]European Molecular Biology
[2]ELIXIR Hub, Cambridge, CB
[3]Department of Mathematic
[4]French Proteomics Infrastru
de Toulouse, CNRS, UPS), Str
[5]ProteoRed, Proteomics Unit
[6]Biomolecular Mass Spectro
Pharmaceutical Sciences, Un
[7]Netherlands Proteomics Ce
[8]Analytical Biochemistry, Dep
[9]Institute of Organic Chemis
[10]CALIPHO Group, SIB Swiss
[11]Department of Human Pro
[12]National Bioinformatics Inf
62, Sweden
[13]Proteome Informatics Gro
[14]Computer Science Departm
[15]Institute of Microbiology,
[16]Center for Proteomics and
[17]Department of Biomedical
[18]Center for Biotechnology,
[19]Department of Biochemist
[20]Interuniversity Institute fo
[21]Center for Proteomics, Uni
[22]Applied Bio & Molecular Sy
[23]Netherlands Metabolomics
[24]Dutch Techcentre for Life S
[25]Medical Bioinformatics, Me
[26]VIB-UGent Center for Medi

F1000Research F1000Research 2017, 6(ELIXIR):1649 Last updated: 21 JAN 2

Check for updat

OPINION ARTICLE

REVISED **The future of metabolomics in ELIXIR [version 2; peer review: 3 approved]**

Merlijn van Rijswijk[1,2], Charlie Beirnaert[3], Christophe Caron[4], Marta Cascante[5], Victoria Dominguez[4], Warwick B. Dunn[6], Timothy M. D. Ebbels[7], Franck Giacomoni[8], Alejandra Gonzalez-Beltran[9], Thomas Hankemeier[2,10], Kenneth Haug[11], Jose L. Izquierdo-Garcia[12,13], Rafael C. Jimenez[14], Fabien Jourdan[15], Namrata Kale[11], Maria I. Klapa[16], Oliver Kohlbacher[17-19], Kairi Koort[20,21], Kim Kultima[22], Gildas Le Corguillé[4,23], Pablo Moreno[11], Nicholas K. Moschonas[16,24], Steffen Neumann[25], Claire O'Donovan[11], Martin Reczko[26], Philippe Rocca-Serra[9], Antonio Rosato[27], Reza M. Salek[11], Susanna-Assunta Sansone[9], Venkata Satagopam[28], Daniel Schober[25], Ruth Shimmo[20,21], Rachel A. Spicer[11], Ola Spjuth[29], Etienne A. Thévenot[30], Mark R. Viant[6], Ralf J. M. Weber[6], Egon L. Willighagen[31], Gianluigi Zanetti[32], Christoph Steinbeck[33]

[1]ELIXIR-NL, Dutch Techcentre for Life Sciences, Utrecht, 3503 RM, The Netherlands
[2]Netherlands Metabolomics Center, Leiden, 2333 CC, The Netherlands
[3]ADReM, Department of Mathematics and Computer Science, University of Antwerp, Antwerp, 2020, Belgium
[4]ELIXIR-FR, French Institute of Bioinformatics, Gif-sur-Yvette, F-91198, France
[5]Department of Biochemistry and Molecular Biomedicine, Faculty of Biology, Universitat de Barcelona, Barcelona, 08028, Spain
[6]School of Biosciences, Phenome Centre Birmingham and Birmingham Metabolomics Training Centre, University of Birmingham, Birmingham, B15 2TT, UK
[7]Computational and Systems Medicine, Department of Surgery and Cancer, Imperial College London, London, SW7 2AZ, UK
[8]INRA, UNH, Human Nutrition Unit, PFEM, Metabolism Exploration Platform, MetaboHUB-Clermont, Clermont Auvergne Universit
Clermont-Ferrand, F-63000, France

Galaxy Community (web)

Mission/Vision

Fosters a Galaxy community in Europe, together with Galaxy resources and training.
Monitors and fosters the use of Galaxy in ELIXIR.

Goals

- Build a European network of Galaxy communities
- Extend Galaxy training provision
- Create a Galaxy cloud infrastructure across Europe
- Make it easier to access and transfer data
- Improve tools and data integration
- Promote FAIR principles in Galaxy

Galaxy Europe

Some ELIXIR nodes already offer a centralised instance of Galaxy

- IFB, France
- de.NBI, Germany
- usegalaxy.be, Belgium

→ collaboration with the Galaxy Core Team: launched usegalaxy.eu (2018), hosted in Freiburg (de.NBI)
→ Aim to expand this into a network of Galaxy instances worldwide, guaranteeing a base level of compatibility and enabling all training materials of the Galaxy Training Network.

Acknowledgements & contacts

- Sirarat Sarntivijai on behalf of the ELIXIR Data Platform and ELIXIR Interoperability Platform
- Jennifer Harrow on behalf of the ELIXIR Tools Platform
- Jonathan Tedds on behalf of the ELIXIR Compute Platform
- Katharina Heil on behalf of the ELIXIR Training Platform and the Proteomics, Metabolomics, and Galaxy ELIXIR Communities
- Andy Smith & the ELIXIR Hub

Follow ELIXIR:

- Post job vacancies here: https://elixir-europe.org/about-us/vacancies
- Follow us on social media - Twitter: @ELIXIREurope
- BioHackathon Europe: https://biohackathon-europe.org
- ELIXIR Webinars: https://elixir-europe.org/events
- Any post-meeting questions - peter.maccallum@elixir-europe.org