

ELIXIR - EOSC strategy plan

ELIXIR - a pan-European ESFRI within the life science domain



Jessica Lindvall on behalf of ELIXIR team

September 27, 2023, Swiss EOSC Coffee

www.elixir-europe.org

25 minutes talk at the Swiss EOSC Coffee 27.9.2023 (by Jessica Lindvall,
jessica.lindvall@scilifelab.se)

Jessica Lindvall - Dep. Head of Node ELIXIR-SE

- Associate Professor in Bioinformatics (Stockholm University)
- ELIXIR Training Platform ExCo co-lead
- EOSCTask Force co-chair “Upskilling countries to engage in EOSC”
- Head of Training (NBIS Training and SciLifeLab Training Hub)
- ... and other things...



Overview of Jessica Lindvall

Introduction to ELIXIR – how we work



Short introduction to ELIXIR and the work

ELIXIR - data for life

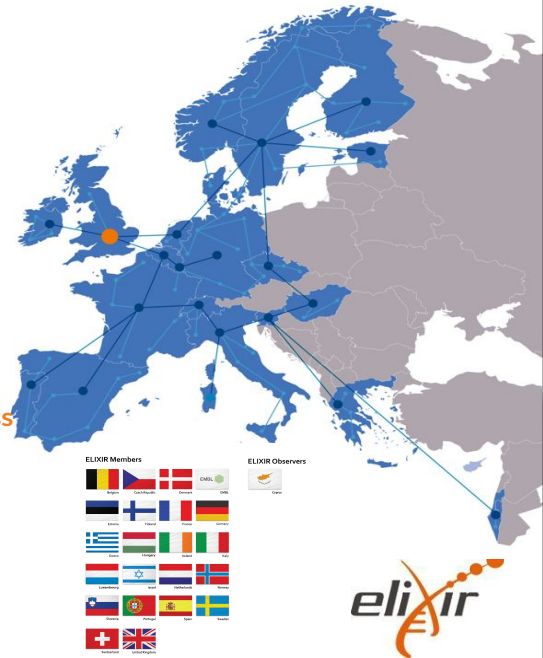
ELIXIR is a pan-European research infrastructure (ESFRI) that brings together life science resources such as

- databases
- software tools
- training materials
- data standards
- compute resources
- people infrastructure

The goal of ELIXIR is to **coordinate life science resources from across Europe, so they form a single infrastructure**

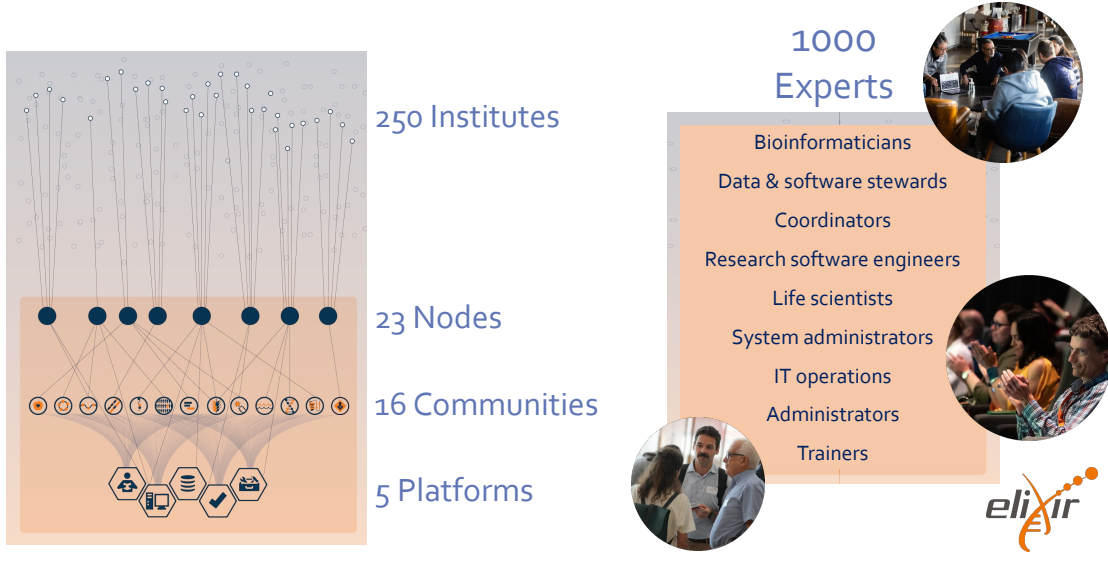
ELIXIR in numbers

- 23 Countries
- ~ 250 research institutes involved
- 1000+ staff



Introduction to ELIXIR as an ESFRI and pan-European Research Infrastructure.

ELIXIR's core structure



ELIXIR core structure. ELIXIR is built around the core (orange) 5 platforms (core and these always exist), Communities (scientific drivers and evolving. Communities have a life cycle) and the Nodes (currently 23 but more are onboarded). The Core (orange) consists of many experts with a broad range of expertise, skills and knowledge

ELIXIR stakeholders - organisations

ELIXIR Hub
1 organisation

Nodes
>250 organisations

Collaborators
25-50 organisations

Users
>10 000
organisations

To show that ELIXIR is reaching many, from the central Hub via the Nodes and our collaborators to the users

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

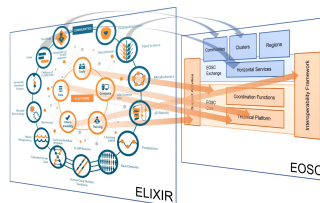


<https://doi.org/10.5281/zenodo.7120997>

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

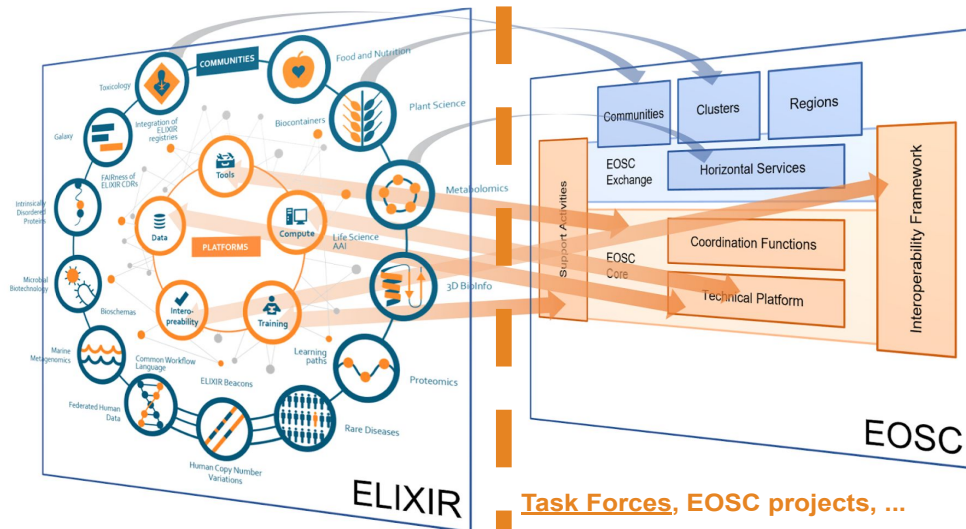


EOSC-Future



One of the collaborators are the EOSC ecosystem. The picture shows some use cases and examples here ELIXIR is involved.

ELIXIR connection to EOSC



The connection between ELIXIR ecosystem and EOSC ecosystem where we envision the EOSC-Association Task Forces and various EOSC projects and others to be the 'filter' between the two ecosystem, bi-directional feeding into each other. Across the ELIXIR ecosystem we have many members involved in various EOSC initiatives, taking the strength from one another.

New EOSC projects (involvement of ELIXIR)

ELIXIR Hub involvement

- OSCARS
- EVERSE
- ENTRUST

ELIXIR Node involvement

- OSTrails



An overview of new EOSC (from 2023) projects where ELIXIR is involved.

ELIXIR expert groups

ELIXIR brings groups of people together as **Communities, Focus Groups and Platforms**. There are **over 30 individual groups** which interact in a range of different ways.

3D-BioInfo	Metabolomics	Systems Biology	Biocuration	Health Data	Data Platform
Food and Nutrition	Microbial Biotechnology	Toxicology	Biodiversity	Impact	Tools Platform
Galaxy	Plant Sciences	Federated Human Data	Cancer Data	Innovation and Industry	Interoperability Platform
Intrinsically Disordered Proteins	Proteomics	Human Copy Number Variation	EOSC	Machine Learning (AI)	Compute Platform
Marine Metagenomics	Single-Cell Omics	Rare Diseases	FAIR Training	Research Data Alliance (RDA) Activities	Training Platform



ELIXIR have a specific Focus Group on EOSC that gathers stakeholders across the ELIXIR landscape to come together and discuss and drive the ELIXIR-EOSC initiatives. E.g the Focus Group helped draft the EOSC-ELIXIR Strategy. A group that meets monthly to share EOSC specific things

Current (fall 2023) ELIXIR-EOSC status

EOSC Nodes concept

- ELIXIR will monitor closely how this develops - [paper just released by GEANT on this](#)

2024 - with the ELIXIR-EOSC Focus Group

- ELIXIR-EOSC Strategy to be updates (led by Jonathan Tedds)
- Consider to develop guidelines/best practices for ELIXIR services to be onboarded to EOSC

EOSC Future - extended 6 months (monitored closely by ELIXIR)

- Focusing on EOSC Core components e.g. Marketplace/Helpdesk/Monitoring etc

EOSC Beyond project - start 2024

- Develop *new* EOSC Core components and developing EOSC Node concept (CESNET (CZ) pilot Node)

European Data Spaces coming together under [DSSC](#) e.g. [EHDS](#) (in alignment with [GDI](#) and other)



Current status on the work for the ELIXIR-EOSC Focus Group

**ELIXIR services and
resources – benefits to
the EOSC ecosystem**




ELIXIR Nodes run 430 services for life scientists & bioinformaticians

- 21 Compute services – cloud, compute, storage and access services
- 128 Data resources – sustainable data resources within a scalable, connected ecosystem
- 47 Interoperability services – standardisation, metadata and vocabularies
- 282 Tools – accessible and benchmarked software tools conforming to information standards
- 27 Training services – developing scientific and technical experts and users




Overview in numbers what thematic services ELIXIR provides

Resources to access knowledge and curated digital objects




bio.tools

bio.tools helps you find and select bioinformatics 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, Dmitry Fishman, Dario Garcia-Gasulla, Tina Tirma, 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, high volumes of biological data are being generated and made accessible to researchers. Machine learning (ML) has come into



EDITORIAL

Ten simple rules for making a software tool workflow-ready

Paul Branch^{1,2}, Peter Crook^{3,4}, Brian Ballard-Rapley^{1,2,4}, Stuart Owen^{1,2}, Douglas Lowe⁵, Aljos R. Williams⁶, Quentin Groom⁶, Matthew O'Brien⁷, Frederik Coppens^{8,9}, Robin Drilling¹⁰, Ignace Eggenmeier¹⁰, Philip Ewels¹¹, Carole Godley¹²

¹ Department of Computer Science, The University of Manchester, Manchester, United Kingdom; ² MRC Centre for Drug Use and Dependence Research, University of Manchester, Manchester, United Kingdom; ³ The Wellcome Centre for Human Genetics, University of Oxford, Oxford, United Kingdom; ⁴ MRC Centre for Drug Use and Dependence Research, University of Manchester, Manchester, United Kingdom; ⁵ MRC Centre for Drug Use and Dependence Research, University of Manchester, Manchester, United Kingdom; ⁶ MRC Centre for Drug Use and Dependence Research, University of Manchester, Manchester, United Kingdom; ⁷ MRC Centre for Drug Use and Dependence Research, University of Manchester, Manchester, United Kingdom; ⁸ MRC Centre for Drug Use and Dependence Research, University of Manchester, Manchester, United Kingdom; ⁹ MRC Centre for Drug Use and Dependence Research, University of Manchester, Manchester, United Kingdom; ¹⁰ MRC Centre for Drug Use and Dependence Research, University of Manchester, Manchester, United Kingdom; ¹¹ MRC Centre for Drug Use and Dependence Research, University of Manchester, Manchester, United Kingdom; ¹² MRC Centre for Drug Use and Dependence Research, University of Manchester, Manchester, United Kingdom

^{*} andrew.reynolds@man.ac.uk

Introduction

In recent years, the volume of data to be analysed, as well as the complexity of that analysis, across many scientific fields (from genomics through to complex systems) have increased

SCIENTIFIC DATA

Comment: The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson et al.^{*}

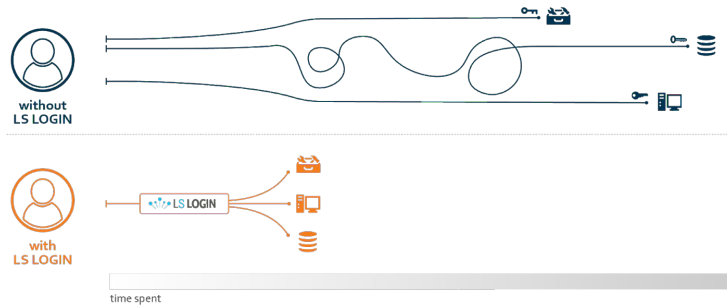
There is an urgent need to improve the infrastructure supporting the reuse of scholarly data. A diverse set of stakeholders—representing academia, industry, funding agencies, and scholarly publishers—have come together to design and jointly endorse a concise and measurable set of principles that we refer to as the FAIR Data Principles. The intent is that these may act as a guideline for those seeking to enhance the reusability of their data holdings. Distinct from past initiatives that focus on the human aspect, the FAIR Principles are specifically designed to enhance the ability of machines to automatically find and use the data, in addition to supporting its reuse by individuals. The document is the first formal publication of the FAIR Principles, and includes the rationale behind them, and some exemplar implementations in this 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



Examples on resources and the impact ELIXIR made across various thamtics.

Life Science Login as a single-entry point

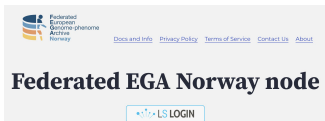


Originally ELIXIR AAI

Co-developed through EOSC-Life Project (WP5) into current Life Science Login

Applicability for federated authentication and access management for LS services

Use case example: Federated EGA - Norway













15

Single entry point to all services – LS LOGIN

1. Remove the burden of maintaining and securing different digital identities from users
2. Simplify situation when accessing several services/data sources for a single workflow
3. Remove the burden of maintaining and securing user's access credentials from service providers
4. Simplify situation for service providers, allowing them to focus on their service

Examples of ELIXIR high impact services

	Life Science Login	Authentication service	 13K Users
	FAIRsharing	Data and metadata resource including standards, databases and policies	 15K Standards
	ELIXIR TeSS	ELIXIR's training portal	 16K Training materials
	BioContainers	Software standardisation resource	 70K Containers & packages
	BioTools	Registry of tools, databases and services	 35K Monthly visits

Metrics from 2021



Numbers quoted from 2021 annual report

ELIXIR's support for FAIR data and software



FAIR data services & resources

Open registries, ontologies, identifiers, data management platforms, stewardship tools, data FAIRification methodology, standards



Trusted data resources

Open deposition databases and portals, scalable curation, sustainability



Data analytics & platforms

Workflows, reproducible and portable processing, software and AI best practice, FAIR assessment, federated analytics



Open & FAIR policy/advocacy

FAIR principles, FAIR leadership & partnering at the global, European and national level



Specific communities

Human Data, Structural Bioinformatics, Rare Diseases, Plant Sciences, Microbial Biotechnology, Proteomics, Metagenomics, Systems Biology...



Stewardship and training

Capability frameworks, skills, data managers network, training portal



ELIXIR supports FAIR data through several threads.

Services and resources, building blocks that can be used various RDM solutions.

Repositories; ELIXIR deposition databases for the sustainability of FAIR lifescience research data

Techniques for data to be born FAIR such as workflows, study reporting guidelines

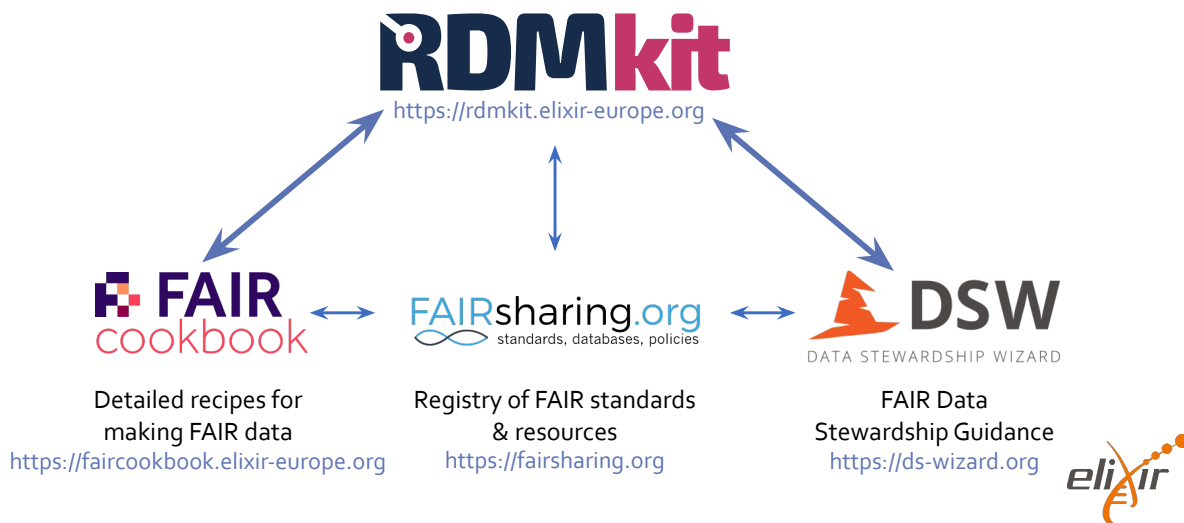
And techniques for existing datasets to be assessed for FAIRness and further FAIRified.

Communities around lifescience disciplines, data types or infrastructures.

Training for the building and expanding FAIR expertise.

Research Data Management Toolkit

Guidance for data stewards, project managers and researchers



<https://faircookbook.elixir-europe.org/>



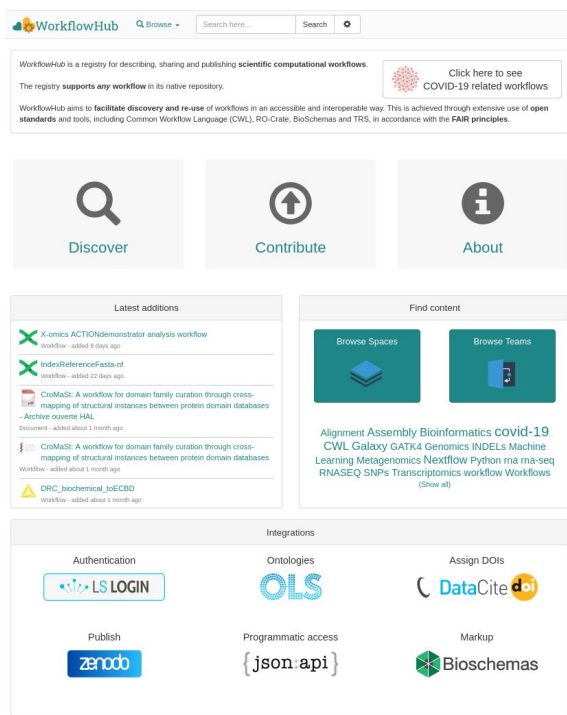
A FAIR workflow registry

Launched 2020

An EOSC service provided by ELIXIR, the University of Manchester and the EOSC-Life project

Open development

<https://workflowhub.eu>



287
workflows

11
system types

112
teams

105
organisations

364
people



The first F of FAIR is Findability, so for that we have made WorkflowHub, a registry of now more than 280 computational workflows. It was launched in 2020, initially to host COVID workflows, then in EOSC-Life and ELIXIR, and now it's an EOSC service supported by multiple projects and institutions through Open Development.

ELIXIR's Training Portal- TeSS (Training eSupport System)

Single entry point for trainers and trainees to discover online information and content

Includes training materials, events, collections and learning paths

Training material collected from 59 content providers

Event information collected from 71 content providers

How can TeSS help you?

Search the portal for courses, events, videos, presentations, learning pathways, handbooks...
All types of resources at all levels for leveraging computational resources in the life sciences.

Search TeSS...



Browse the catalogue



Events

Discover the latest training events and news from ELIXIR nodes and 3rd-party providers.



Materials

Browse the catalogue of training materials offered by ELIXIR nodes and 3rd-party providers.



Workflows

Create training workflows to visualise learning steps and link to resources specific to your training needs.

Niall Beard, Finn Bacall, Aleksandra Nenadic, Milo Thurston, Carole A Goble, Susanna-Assunta Sansone, Teresa K Attwood

TeSS: a platform for discovering life-science training opportunities, *Bioinformatics*, Volume 36, Issue 10, 15 May 2020, Pp 3290-3291

<https://doi.org/10.1093/bioinformatics/btaa047>



ELIXIR Flag-ship TeSS the training registry which gather training resources, tools and support including events and material that is a product from ELIXIR.

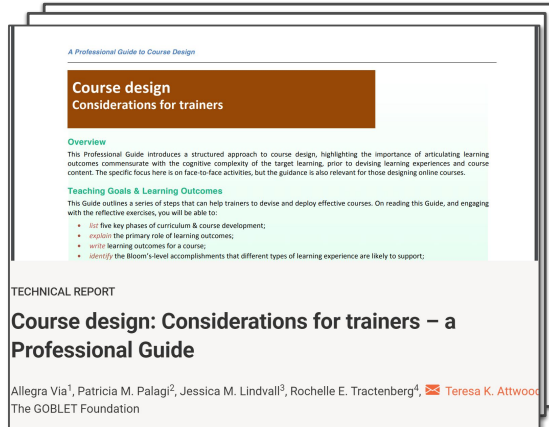
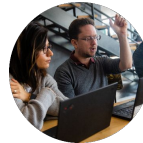
Train-the-trainer (TtT)

398 trainers trained since ELIXIR started in over 32 TtT courses

ELIXIR TtT instructor community (~20 active members)

Training materials developed to deliver courses face-to-face, hybrid and online (synchronously)

Professional guidelines available for trainers



Train-the_Trainer programme is an impactful way to increase capability in researchers and staff-scientists across ELIXIR to share knowledge and expertise.

Opportunities and Challenges



Opportunities ELIXIR and EOSC

Data Integration: ELIXIR provides a range of data resources and tools that can be integrated into the EOSC ecosystem. This would enable EOSC users to access and utilize ELIXIR's vast collection of biological data and resources, including genomics, proteomics, metabolomics, and more.

Interoperability: ELIXIR's expertise in developing standards for data and tool interoperability can be leveraged to ensure that EOSC components are interoperable and can communicate effectively with each other. This would facilitate the sharing and reuse of data and tools across the EOSC.

Training: ELIXIR has a strong focus on training, with a range of resources available for researchers, developers, and other stakeholders. This expertise can be harnessed to develop training programs for EOSC users, helping to build capacity and ensure that researchers can take full advantage of the EOSC.

FAIR Data: Collaborating between ELIXIR and the EOSC would help ensure that data stored in EOSC repositories is compliant with FAIR principles, facilitating its use and reuse by researchers.

Cross-disciplinary Collaboration: ELIXIR's expertise in biological data and tools can be applied to a wide range of scientific fields, from bioinformatics to biomedicine, agronomy, environmental science, and more. Collaborating with EOSC would enable ELIXIR to expand its reach and work with researchers from other disciplines to find solutions to complex scientific challenges.



Jessicas thoughts on some challenges that exists between ELIXIR and EOSC

Collaborative challenges ELIXIR and EOSC

User Engagement: ELIXIR and EOSC will need to engage with a diverse range of users, from researchers and developers to policy makers and funders. Ensuring that these users are aware of the benefits of collaboration and engaged in the development of policies and services will be crucial for success.

Cultural Differences: ELIXIR and EOSC are both large, complex organizations with their own cultures, values, and priorities. Finding ways to bridge these cultural differences and work together effectively will require strong leadership, communication, and collaboration skills.

Data Governance: One of the main challenges in data sharing is ensuring that data is managed in accordance with ethical, legal, and regulatory requirements. ELIXIR and EOSC will need to work together to establish data governance policies that address issues such as privacy, confidentiality, consent, and data protection.

Technical Compatibility: While ELIXIR and EOSC share a commitment to open science and data sharing, they may use different technical standards, protocols, and infrastructure. Harmonizing these systems and ensuring technical compatibility will be an important challenge for both organizations.



Jessica's thoughts on the collaborative challenges that exist, bottom-up and top-down

Ideas on way forward, for an efficient collaboration

User Engagement: ELIXIR and EOSC *collaboratively develop user-centric policies and services*

Establish user advisory boards to provide feedback and suggestions on how to improve services and policies
Joint outreach activities, such as workshops, training sessions etc.

Foster a Culture of Collaboration: KEY for successful collaboration and work and require a strong commitment to collaboration and communication

Leaders from both organizations should work to foster a culture of collaboration and openness, actively seeking opportunities for synergies (joint projects and initiatives)

Regular meetings and forums for dialogue can also help facilitate communication and build trust between organisations and EOSC

Develop Common Data Governance Policies: *collaborate to develop common data governance policies* that address ethical, legal, and regulatory issues related to data sharing

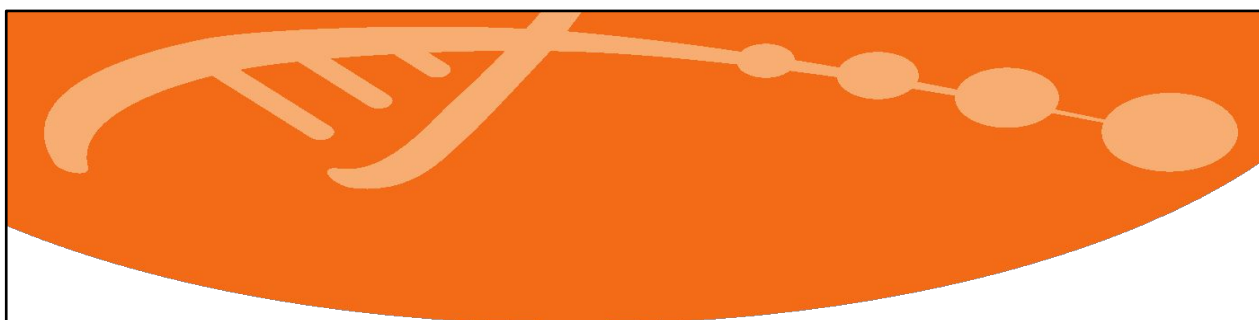
Developed with input from a wide range of stakeholders, including researchers, policy makers, and data protection experts.

Establish Technical Working Groups: *Establish joint technical working groups to explore and develop common standards, protocols, and infrastructure.*

E.g. Task Force(s) to focus on areas of overlap and should be composed of technical experts across organisations



Ideas on how to overcome these challenges, bottom.up and top-down



Thank you!

jessica.lindvall@scilifelab.se



www.elixir-europe.org