

Interoperability in the life sciences

Converging on practices for sharing research information and tools

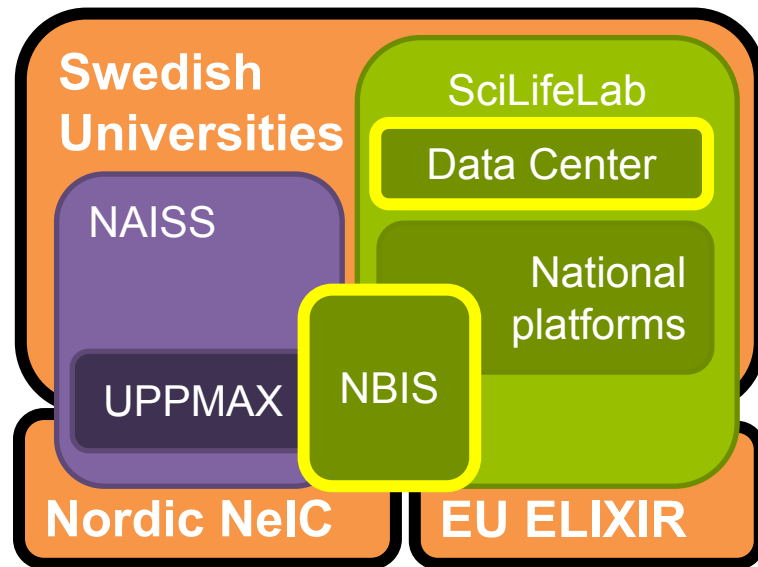
Wolmar Nyberg Åkerström, Data steward / bioinformatician
NBIS – National Bioinformatics Infrastructure Sweden

ENLIGHT Workshop#2 on Connecting Digital Research Infrastructures
with European Infrastructures - Experiences and Lessons Learned
15 March 2023, Online

Connecting infrastructures



- **NBIS - National Bioinformatics Infrastructure Sweden** has staff at Uppsala University and in 5 other cities
- **ELIXIR Sweden** is a node in a European infrastructure for life science resources
- **Enabling open science** through national, Nordic and European collaborations
 - Support
 - Training
 - Infrastructure



“Swedish Life Science researchers **apply good Data Management practices** so that the **research outputs** produced are **available** to the global research community, and to society at large, according to the principles of *Open Science, Reproducible Research, and FAIR*”



Niclas Jareborg
Data Manager



Markus Englund
Data Steward



Yvonne Kallberg
Data Steward



Elin Kronander
Data Steward



Stephan Nylinder
Data Steward



Maja Pelve
Data Steward



Mattias Strömberg
Data Steward



Wolmar N. Åkerström
Data Steward



Erik Hedman
Data Steward



Support

Support services ranging from short consultation, consultancy to long-term embedded bioinformaticians.

- ❑ Data management planning and good practice
- ❑ Publishing data, code and other outputs
- ❑ Considerations for working with sensitive data
- ❑ Preparing data and research outputs for sharing and reuse



Training

Training events in advanced
and applied bioinformatics.

- ❑ Courses and reusable training materials on data management practices and reproducibility
- ❑ Events and seminars designed to cater to life science researchers at all career stages
- ❑ Embedded courses and presentations on request



Infrastructure

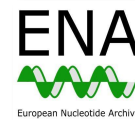
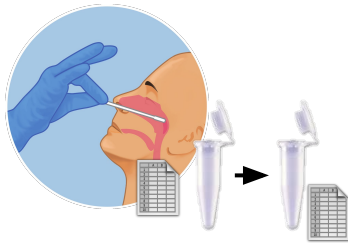
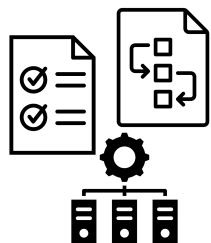
Providing infrastructure in the form of services, computational resources, tools and guidelines to the life science community.

- ❑ Data management guidelines for the Swedish life sciences
- ❑ Federated EGA for transnational discovery and access to sensitive human research data
- ❑ AIDA Data Hub for machine learning in medical imaging diagnostics
- ❑ Genomic Data Infrastructure for access to genomic, phenotypic and clinical data across Europe

Promoting FAIR by design



“Protocol” & “project plan” icons by Justin Blake, and “infrastructure” icon by Eko Purnomo, from thenounproject.com



Study & data
design

Sampling
& specimen
collection

Sample
preparation

Sample analysis
& data generation

Data processing
to prepare inputs
for analysis

Data
analysis

Communicating
results

Procedures

data protection,
ethics permit,
infrastructure,
standards,
protocols,
data dictionaries,
data access, ...

Biosamples and instruments

populations (statistical) and inclusion criteria,
physical processing steps,
working storage conditions,
long-term storage location,
sample quality assessment,
sample annotations,
reagents, instruments, kits, ...

Data and computational workflows

digital processing steps,
working storage conditions,
long-term storage location,
data quality assessment,
sample/data annotations,
reference data,
analysis method...

Outputs

publications,
data,
tools,
workflows,
reports,
dashboards, ...



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

Guidance for data stewards, project managers and researchers



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



Detailed recipes for making FAIR data

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



Registry of FAIR standards & resources

<https://fairsharing.org>



FAIR Data Stewardship Guidance

<https://ds-wizard.org>



A collection of **recipes by the bioscience community** that cover the specific operation steps of FAIR data management

Search the cookbook recipes

Filters:

Search recipes names

Search recipes by type

Search recipes by audience

Reading Time

From 0 min To 60 min
Executable code

Yes Both No

Maturity Level

○ ○ ○ ○ ○

RESET

Identifier	Recipe Name	Recipe Type	Reading Time	Executable Code	Audience	Maturity Level
FCB015	Downloading data with Aspera	Hands-on	15 min	✓	Principal Investigator, Data Manager, Data Scientist	○ ○ ○ ○ ○
FCB014	Transferring data with SFTP	Hands-on	15 min	✓	Principal Investigator, Data Manager, Data Scientist	○ ○ ○ ○ ○
FCB054	EHDEN OHDSI discovery with Schema.org	Experience Report / Applied Example	15 min	✗	Terminology Manager, Data Manager, Data Scientist, Ontologist	● ● ○ ○ ○
FCB042	eTox - omics datasets	Experience Report / Applied Example				○ ○ ○ ○ ○
FCB067	FAIR High-Content Screening data deposition	Experience Report / Applied Example				○ ○ ○ ○ ○
FCB037	Making omics data matrix FAIR	Hands-on				○ ○ ○ ○ ○
FCB038	FAIR Data Matrices	Hands-on				○ ○ ○ ○ ○
FCB039	Structuring data matrices	Hands-on				○ ○ ○ ○ ○
FCB040	Exploring data with SPARQL	Hands-on				○ ○ ○ ○ ○
FCB041	Integrating data	Hands-on				○ ○ ○ ○ ○
FCB062	FAIR Computational Workflows	Hands-on				○ ○ ○ ○ ○
FCB043	ND4BB - chemical activities datasets	Experience Report / Applied Example				○ ○ ○ ○ ○



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?

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.



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>

The screenshot shows the WorkflowHub homepage. At the top is a navigation bar with the WorkflowHub logo, a 'Browse' dropdown, a search bar, and a settings icon. Below the navigation bar is a main content area with a welcome message: 'WorkflowHub is a registry for describing, sharing and publishing scientific computational workflows. The registry supports any workflow in its native repository.' A link is provided to 'Click here to see COVID-19 related workflows'. Below this is a section with three large icons: 'Discover' (magnifying glass), 'Contribute' (upload arrow), and 'About' (information icon). The page is divided into two main columns. The left column, 'Latest additions', lists recent workflow uploads with icons and brief descriptions, including 'X-omics ACTIONdemonstrator analysis workflow', 'IndexReferenceFasta-nt', 'CroMaSt: A workflow for domain family curation through cross-mapping of structural instances between protein domain databases', and 'DRC_biochemical_toECBD'. The right column, 'Find content', features two buttons: 'Browse Spaces' and 'Browse Teams'. Below these buttons is a list of workflow categories: 'Alignment Assembly Bioinformatics covid-19', 'CWL Galaxy GATK4 Genomics INDELs Machine Learning Metagenomics Nextflow Python ma-seq', 'RNAseq SNPs Transcriptomics workflow Workflows', and a '(Show all)' link. At the bottom is an 'Integrations' section with six categories: 'Authentication' (LS LOGIN), 'Ontologies' (OLS), 'Assign DOIs' (DataCite, DOI), 'Publish' (zenodo), 'Programmatic access' (json:api), and 'Markup' (Bioschemas).

287
workflows

11
system types

112
teams







105
organisations

364
people

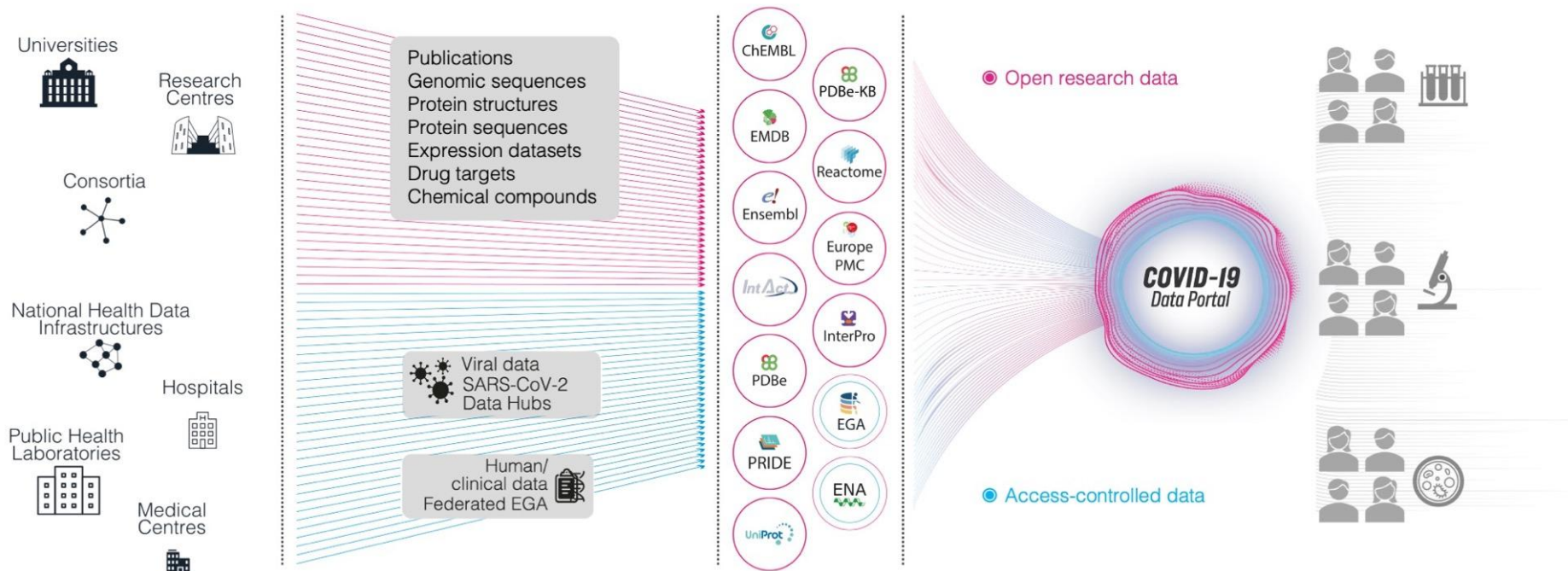


Converging on shared resources



		Core Data Resources (CDRs)	Data resources of fundamental importance for the long-term preservation of life science data
		ELIXIR Deposition Databases (EDDs)	Recommended resources for the deposition of experimental data
		Recommended Interoperability Resources (RIRs)	Tools and resources to support FAIR research

Converging on shared infrastructure





- ELIXIR is a network of 23 national Nodes & the ELIXIR Hub connecting 245 research institutes
 - 5 Platforms
 - 15 Communities
 - 10 Focus groups
 - EU projects & internal projects
- The Nodes collectively run hundreds of bioinformatics services



● ELIXIR institutes
★ ELIXIR Lead Institutes

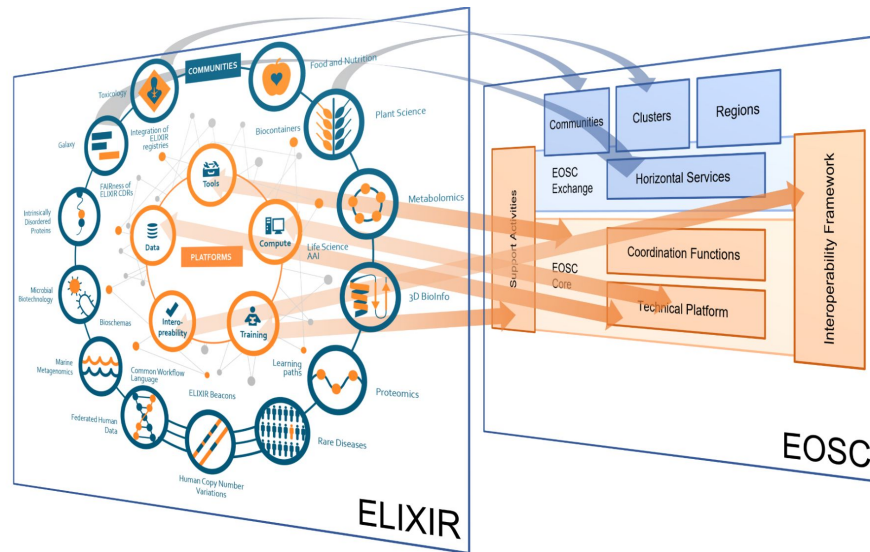


- Throughout the EOSC Core and Interoperability frameworks
- Bidirectional relationships, with best of breed solutions being adopted by both sides
- ELIXIR Communities will over time adopt EOSC Services and become driving user communities for both organisations





- ELIXIR is represented in all 5 advisory groups and most of the Task Forces
- ELIXIR-EOSC Focus Group convenes monthly to report and discuss EOSC initiatives
- 3 NBIS staff, 2 co-chairs



Semantic Interoperability Task Force

Highlight = Start in 2023

Task Force Charter

Focus ≈ Good practices for encoding research information for a web of FAIR data and services—ensuring that the precise meaning of the information is preserved and can be understood by tools, workflows and data infrastructures.

Thematic explorations (Objective 1)

Depart from the task force charter, EOSC projects and interoperability initiatives around the Task Force members

Strategic knowledge exchange (Objective 2)

Organise workshops and collaborative problem-solving activities, accumulating literature, examples of solutions and feedback.

Converging on recommendations (Objective 3)

Produce a set of recommendations on Semantic Interoperability to support the wider EOSC community

Outputs

Scientific articles and/or conference presentations, events organised with stakeholders, and reports and/or guidelines on semantic interoperability, etc.

Work streams update

(Meta)data standards (Theme 1)

Survey the landscape of semantic interoperability around EOSC

Propose a minimum set of (meta)data, indicators and crosswalks

Semantic artefact catalogues (Theme 2)

Converge on a definition and the dimensions of a maturity model

Analyse catalogues using the model and suggest a path forward

Use cases (Theme 3)

Capturing interoperability case studies from EOSC stakeholders

Indexing and presenting use cases to adopters and builders

Road to completion

EOSC Association endorsed deliverables

Consolidate outputs across themes and align with EOSC projects

Converge on recommendations for EOSC going forward

Sort outputs into core deliverables and supporting outputs



Semantic Interoperability
Task Force



Case studies and use cases

User journeys as means of effectively demonstrating value and engaging stakeholders

Engagement of research communities and service providers in EOSC
EOSC Symposium, Prague, Czech Republic
15 November 2022

Presented by Peter Maccallum (ELIXIR)
on behalf of Wolmar Nyberg Åkerström (ELIXIR Sweden)
and the EOSC Semantic Interoperability TF



This work is licensed under a Creative Commons
Attribution-NonCommercial 4.0 International
License

<https://doi.org/10.17044/scilifelab.21542313>

Journeys come in many shapes

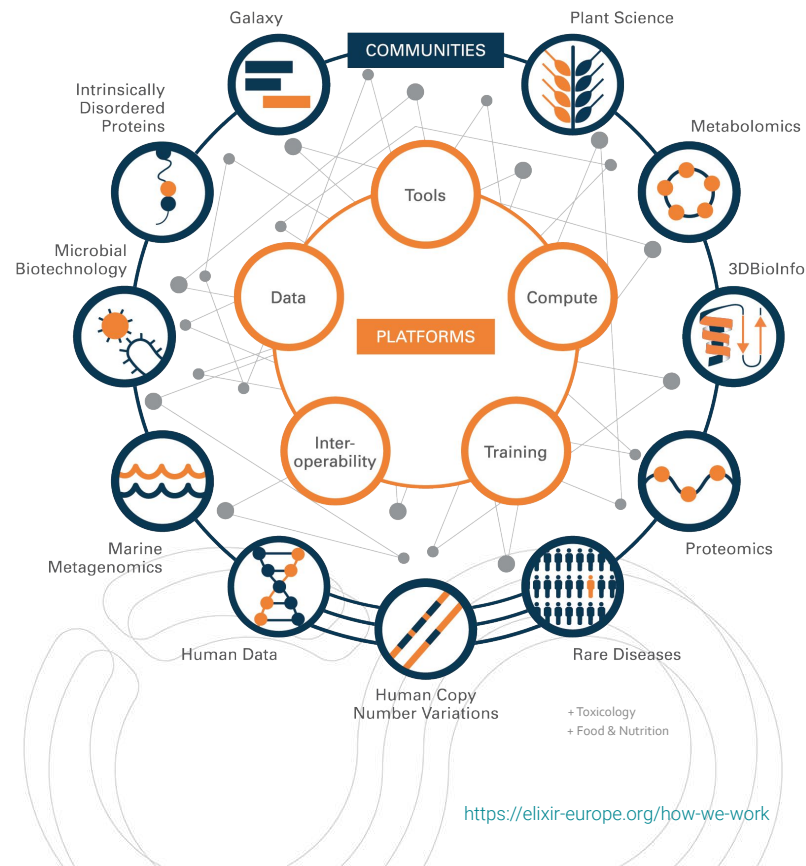
Use cases and case studies can demonstrate value and engage stakeholders in requirements gathering

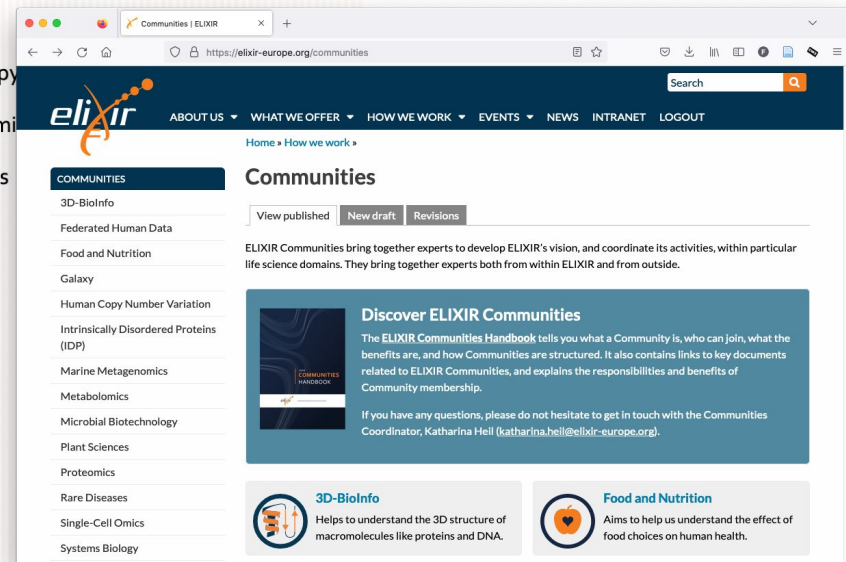
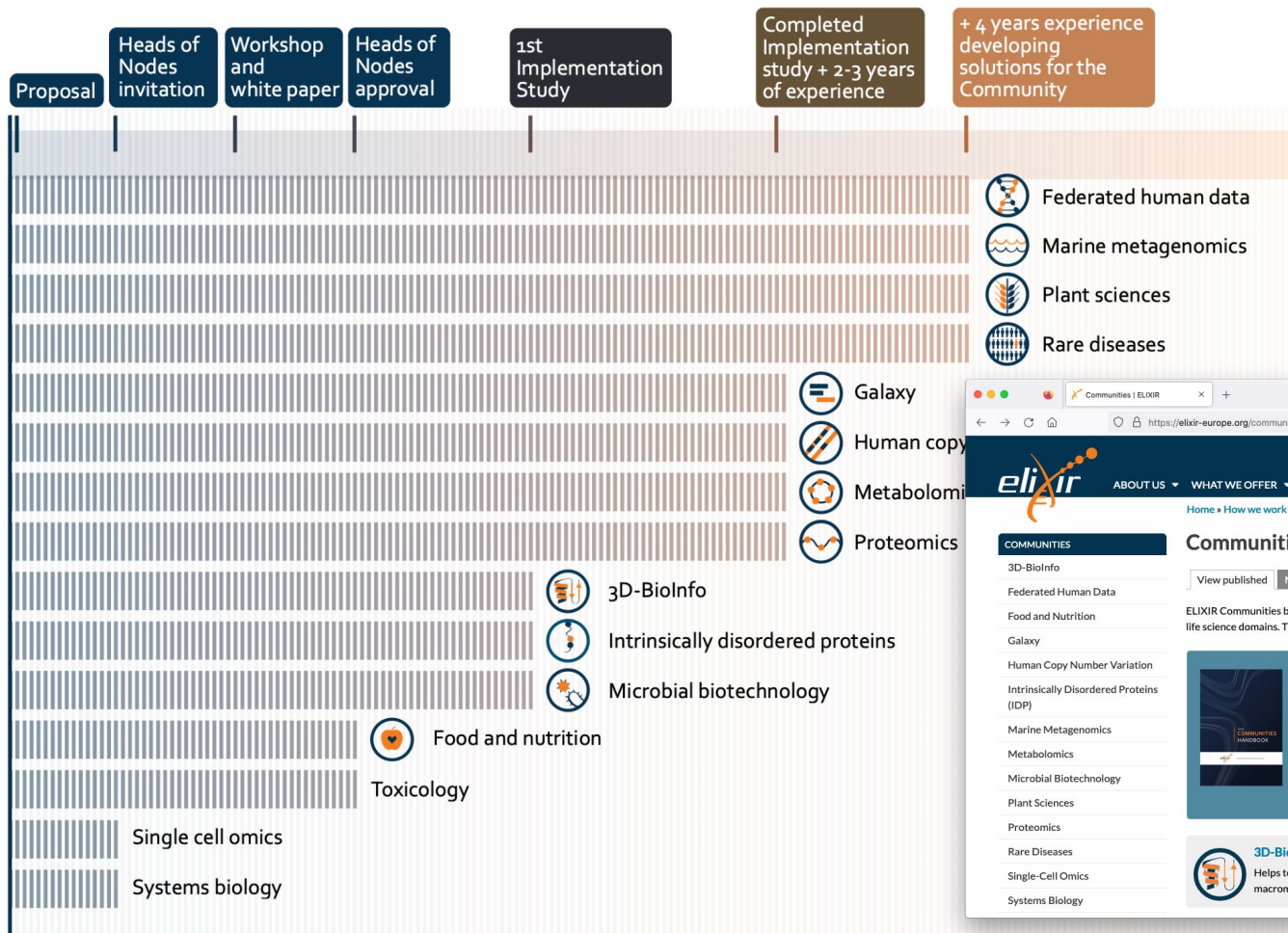
Use cases and communities

ELIXIR draws on communities to share success stories and define use cases for its services

Semantic interoperability use cases

Provide 1) input to the initiatives that are shaping EOSC and 2) examples and lessons learned to stakeholders





Use cases demonstrate value

Prevalent in contexts ranging from product sales, to EOSC projects, and healthcare governance

Adopters recognise themselves

Used in requirements gathering and communication

Builders can see the full picture

Used in development and validation of services

Policy can ensure representation

Mapping the landscape of stakeholders and how they are served by different use cases



Community Use Cases | EOSC


https://eosc-portal.eu/eosc-in-practice/use-cases

Home » Use Cases » Community Use Cases

Community Use Cases

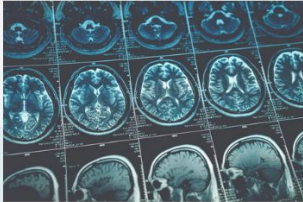
This page provides examples of "EOSC in practice" use cases or success stories that highlight how EOSC services and resources can support the daily work of researchers and innovators. Please scroll down to make the use cases appear.

If you wish to share your EOSC use case, please [fill out the webform](#).




Run4science.org - Measuring environmental and biodiversity data... while running!

We want citizens to measure their environment by using smartphones. Most of the citizen science initiatives are focused on one topic: it can be air quality, biodiversity.





Kampal Artificial Intelligence for rare disease diagnosis

In the context of the EOSC-hub project, Kampal Data Solutions is benefitting from storing the healthy and ill patients' registries to a database on EOSC infrastructure.




Guardomic- bot mitigation engine

Web services owners struggle daily to protect their websites from bot traffic and their users from fraudulent digital ads or cryptocurrency web mining.

BBC Research & Development





ELIXIR and the life sciences community

https://eosc-portal.eu/elixir-and-life-sciences


ELIXIR and the life sciences community

[Go to Website](#)





Engineering & Technology



Natural Sciences

Societal challenges

ELIXIR is an organisation aiming to coordinate, integrate and sustain bioinformatics resources – such as databases, computational services, applications – across its member states and enables users in academia and industry to access what is vital for their research.

The challenge is to unite Europe's leading life science organisations in managing the increasing volume of data generated by publicly-funded research.

Technical challenges

ELIXIR wants to establish a federation of cloud sites, each providing storage and compute capacity for researchers. An established Competence Centre in EOSC-hub is supporting this activity. The CC team has been supporting ELIXIR to set up a compute platform that allows ELIXIR cloud and data providers to share cloud compute and storage capacity to replicate and share reference datasets with each other and with their users.

How EOSC can help and add value

The ELIXIR Compute Platform aims to enable researchers to combine technical components of the platform services into a seamless ecosystem, creating a science-ready interface to the key resources and technological capabilities that are available for life sciences.

The ELIXIR Compute Platform has been leveraging the EOSC Service Catalogue, especially in the area of aligning ELIXIR with the EOSC AAL.

Share this: [f](#) [g+](#) [t](#)

Case studies are real-world scenarios

Actual people and organisations adopting a solution in a specific context and outcomes of their efforts

Use cases can be business focused

How a stakeholder group can approach adopting a solution to achieve broader goals and outcomes

Use cases can be task focused

How a user interacts with different services and resources to complete a task

Conceptual models tie them together

Actors, components and resources involved and how they relate to more general categories



Capturing case studies and use cases

Encourage the wider EOSC stakeholder community to contribute interesting and representative examples

Indexing for adopters and builders

Compare and consolidate across stakeholders, tasks, goals and component of the EOSC IF

Representing the EOSC community

Identify and fill gaps in the types of organisations, domains and goals captured by the task force



Task Force Charter

Semantic Interoperability

Introduction

EOSC and semantic interoperability
Challenges to semantic interoperability

Main aims

Objectives

- O1: Explorations into Semantic Interoperability
- O2: Knowledge exchange around Semantic Interoperability
- O3: Recommendations

Core activities of the Task Force

Planned duration of the Task Force

Working methodology

Explorations/Case Studies

Methods
Output

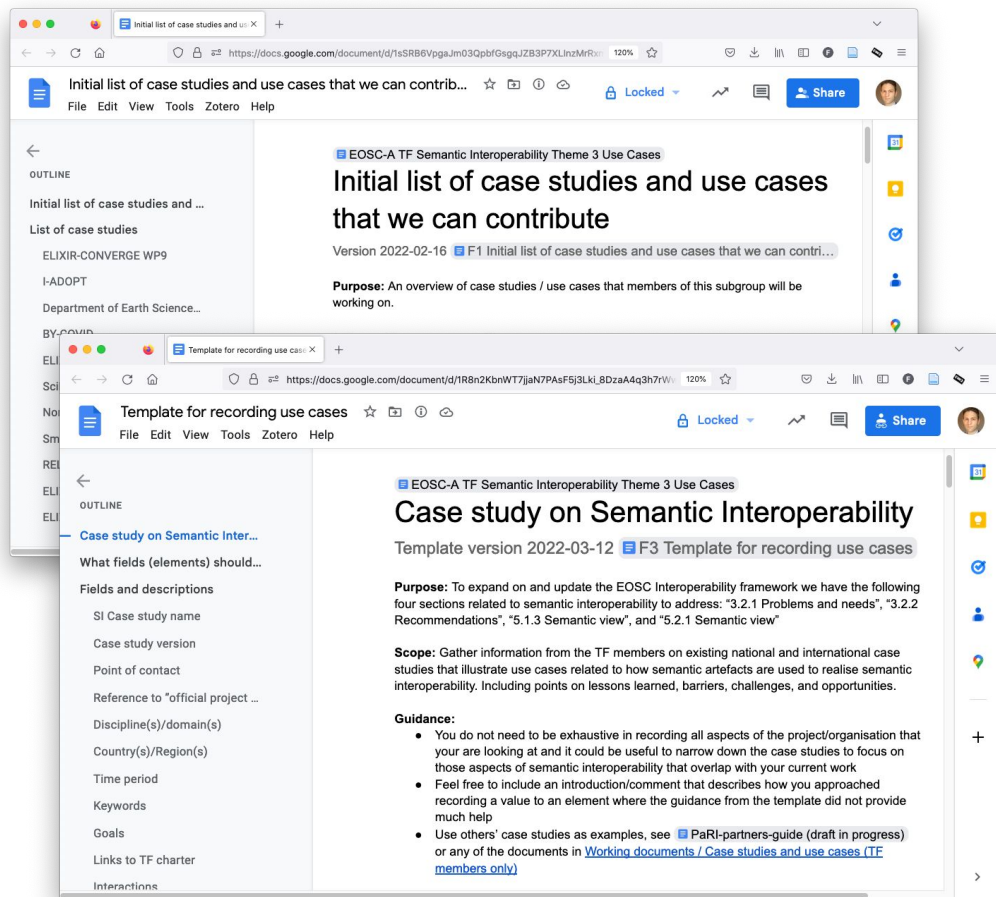
Dependencies

Membership

References

APPENDIX

Case studies (examples)



Guide initiatives through journeys

Coordination and steering through the TFs could be better with a shared vision and comprehensive picture

Host a shared task force catalogue

Keeping accumulated case studies in one place could increase their visibility and avoid redundant efforts

Promote best practices and resources

Pooling experiences, procedures and skills will help TF members focus on content over process

Liaise between EOSC projects and TFs

Creating opportunities to communicate can foster further collaboration and synergies



Thank you



“Swedish Life Science researchers
apply good Data Management practices
so that the **research outputs** produced
are **available** to the global research
community, and to society at large,
according to the principles of *Open
Science, Reproducible Research, and FAIR*”



Niclas Jareborg
Data Manager



Markus Englund
Data Steward



Yvonne Kallberg
Data Steward



Elin Kronander
Data Steward



Stephan Nylinder
Data Steward



Maja Pelve
Data Steward



Mattias Strömberg
Data Steward



Wolmar N. Åkerström
Data Steward



Erik Hedman
Data Steward