ELIXIR

a pan-European ESFRI within the life science domain

Jessica Lindvall on behalf of ELIXIR team

March 13, 2023, TF Upskilling monthly presentations

www.elixir-europe.org

Jessica Lindvall - Associate Professor (Stockholm University)

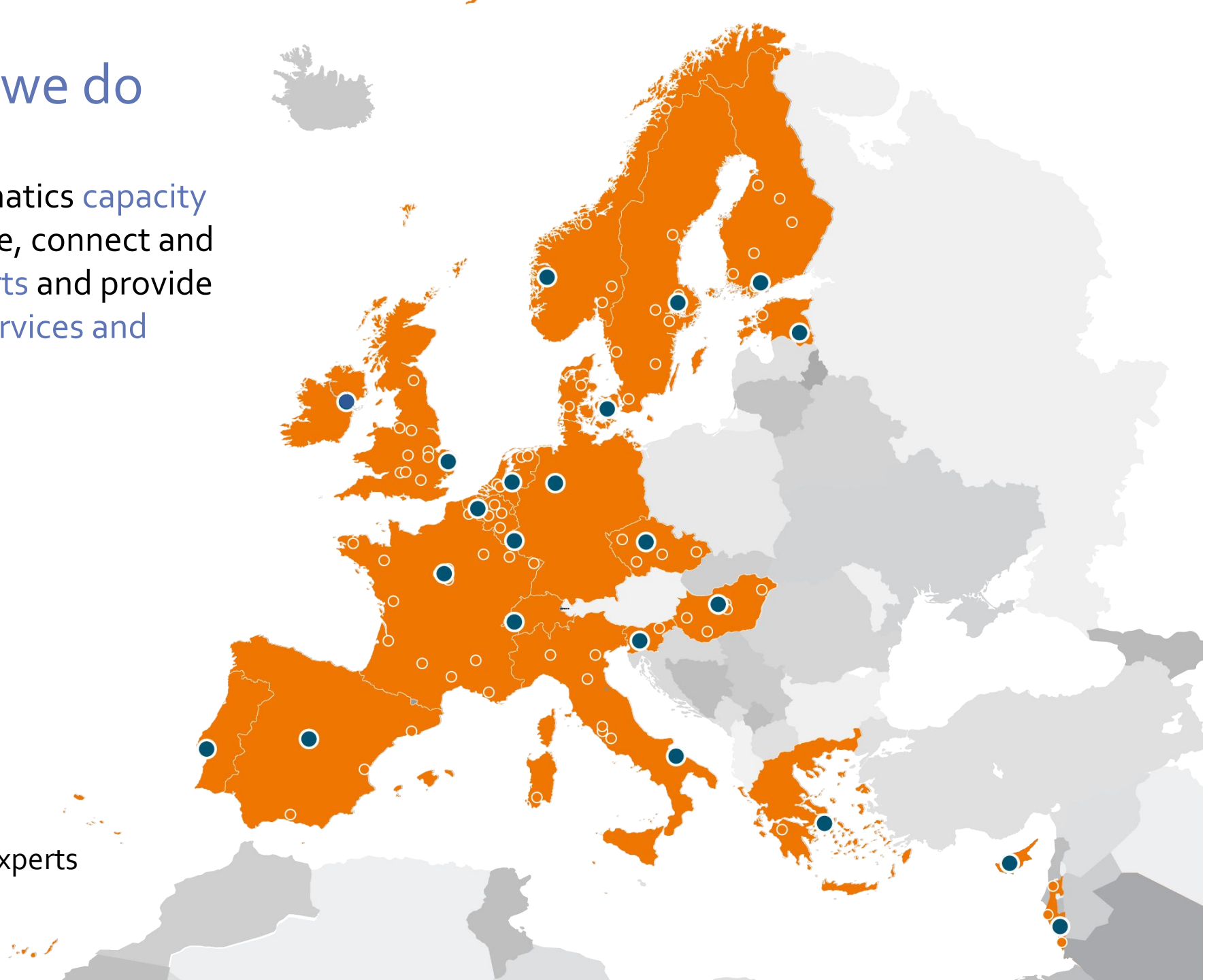
- EOSC Task Force co-chair "Upskilling countries to engage in EOSC"
- ELIXIR Training Platform ExCo co-lead
- ELIXIR Sweden (NBIS, National RI), Deputy Head of Node
- Head of Training (NBIS Training and SciLifeLab Training Hub)
- … and other things…

Introduction to ELIXIR – how we work

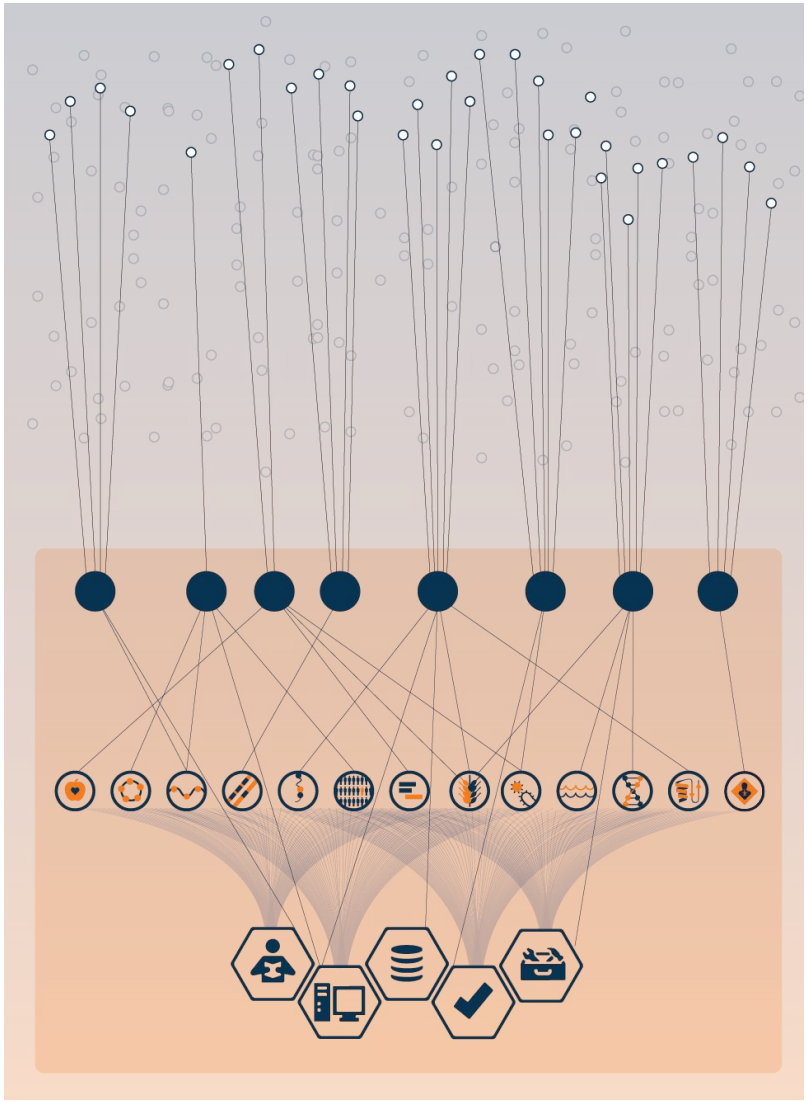
ELIXIR – what do we do

We build life science informatics capacity and infrastructure in Europe, connect and develop a network of experts and provide hundreds of high quality services and resources available to all

- Databases
- Training
- Software tools
- Data standards
- Compute resources
- Scientific & technical experts



ELIXIR's core structure



250 Institutes

23 Nodes

15 Communities

5 Platforms

800 Experts

- Bioinformaticians
- Data & software stewards
- Coordinators
- Research software engineers
- Life scientists
- System administrators
- IT operations
- Administrators
- Trainers

ELIXIR expert groups

ELIXIR brings groups of people together as **Communities, Focus Groups and Platforms**. There are **over 30 individual group**s 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 stakeholders - organisations

ELIXIR Hub
1 organisation

Nodes
>245 organisations

Collaborators
25-50 organisations

Users
>10 000s
organisations

Collaborator example – the ELIXIR EOSC strategic vision

ELIXIR maps to EOSC at all levels

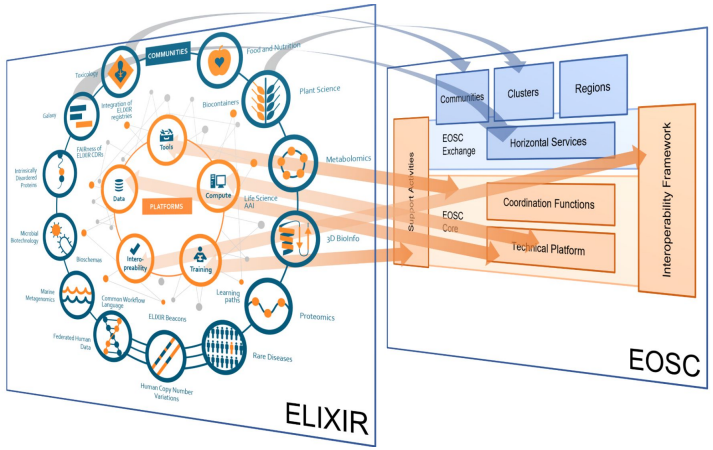
- Throughout the EOSC Core and Interoperability frameworks
- Bidirectional relationships, with best of breed solutions being adopted by both sides
- ELIXIR Communities will over time adopt EOSC Services and become driving user communities for both organisations

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

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**

Resources to access knowledge and curated digital objects

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

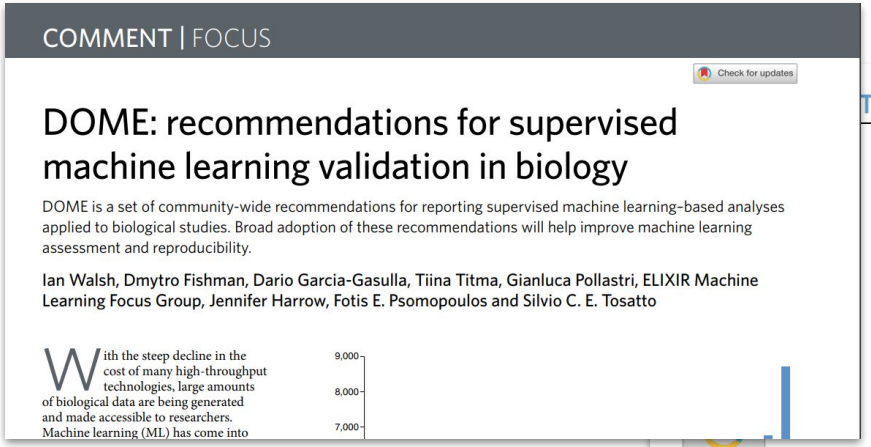
A registry for sharing and publishing scientific computational workflows.

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

Check for updates

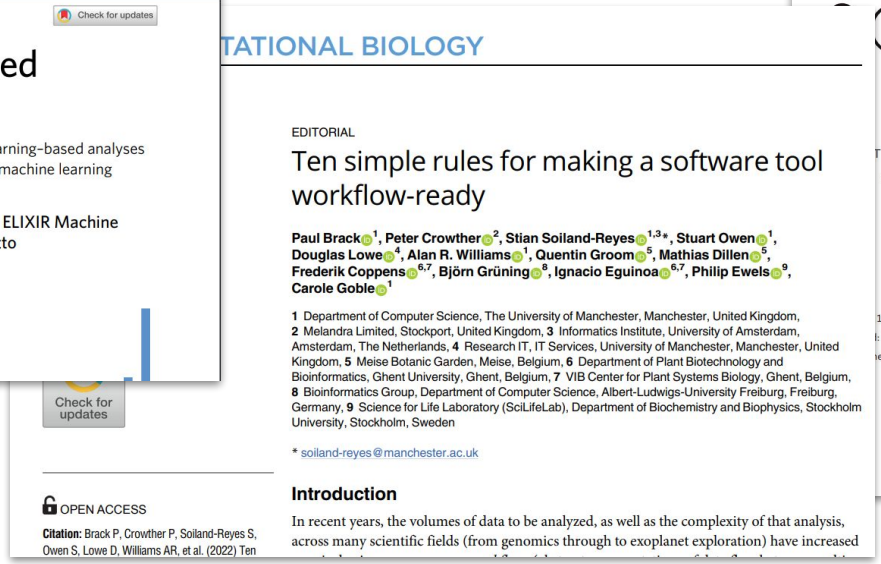
DOME: recommendations for supervised machine learning validation in biology

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

9,000
8,000
7,000


TATIONAL BIOLOGY

EDITORIAL

Ten simple rules for making a software tool workflow-ready

Paul Brack[1], Peter Crowther[2], Stian Soiland-Reyes[1,3]*, Stuart Owen[1], Douglas Lowe[4], Alan R. Williams[1], Quentin Groom[5], Mathias Dillen[5], Frederik Coppens[6,7], Björn Grüning[8], Ignacio Eguinoa[6,7], Philip Ewels[9], Carole Goble[1]

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

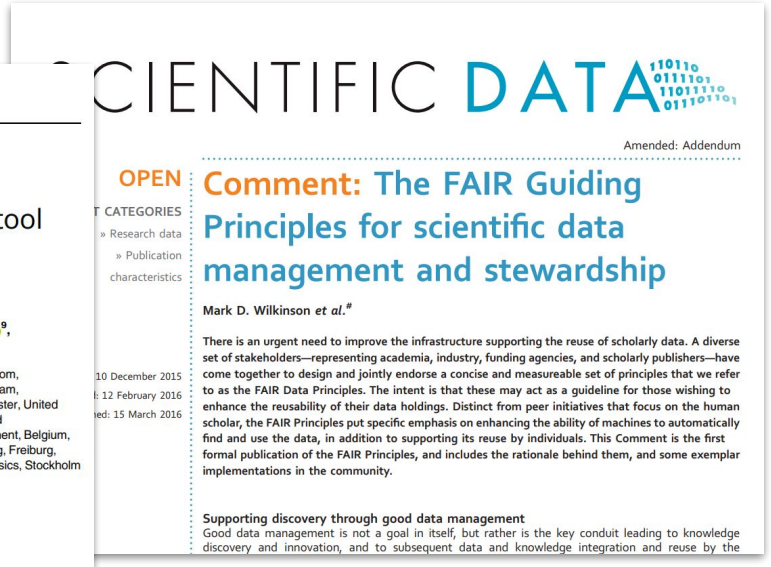
Check for updates

OPEN ACCESS

Citation: Brack P, Crowther P, Soiland-Reyes S, Owen S, Lowe D, Williams AR, et al. (2022) Ten

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


CIENTIFIC DATA

Amended: Addendum

OPEN

T CATEGORIES

» Research data

» Publication characteristics

10 December 2015
: 12 February 2016
ed: 15 March 2016

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

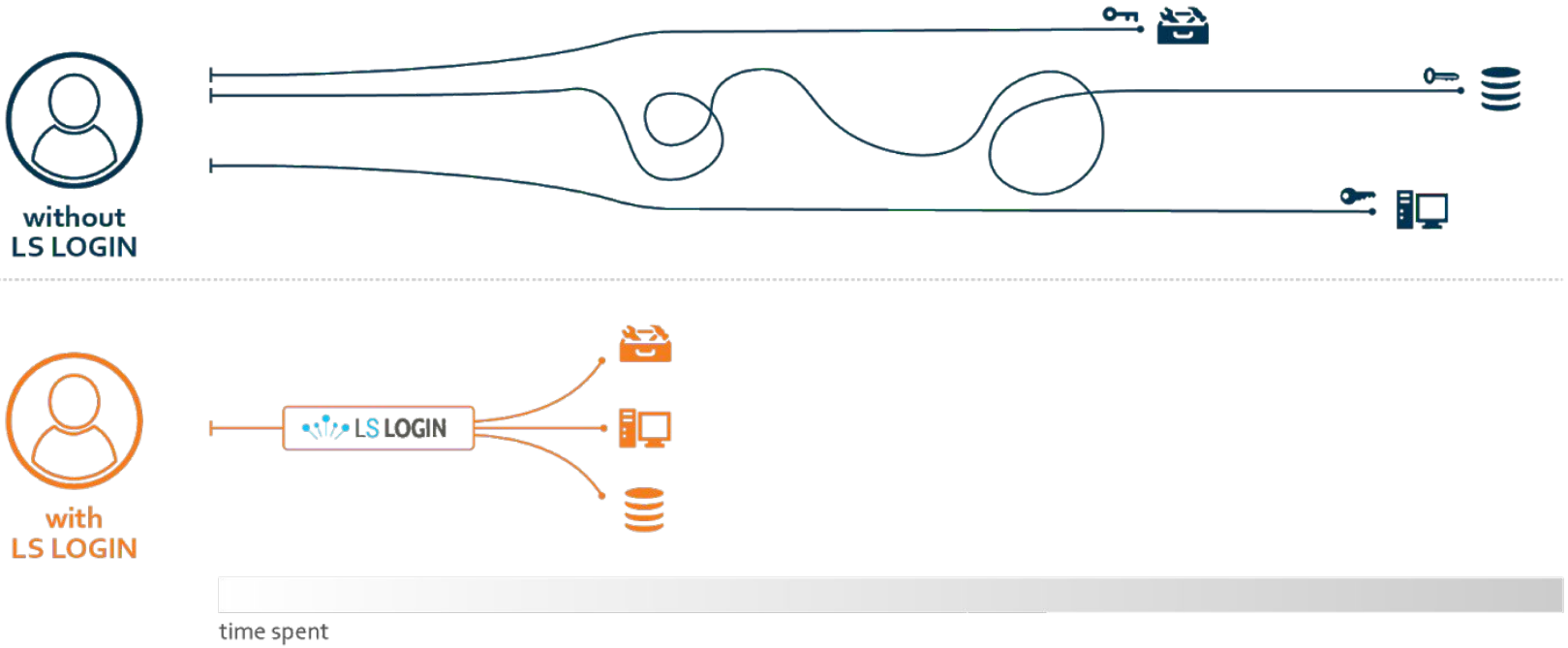
Mark D. Wilkinson *et al.*#

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 measureable 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

Life Science Login as a single-entry point

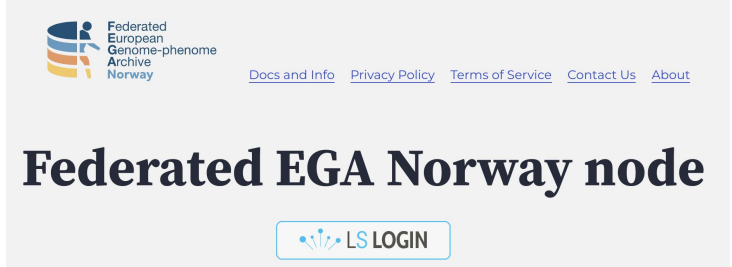


Originally ELIXIR AAI

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



Examples of ELIXIR high impact services

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

Metrics from 2021

elixir

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

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

elixir

WorkflowHub

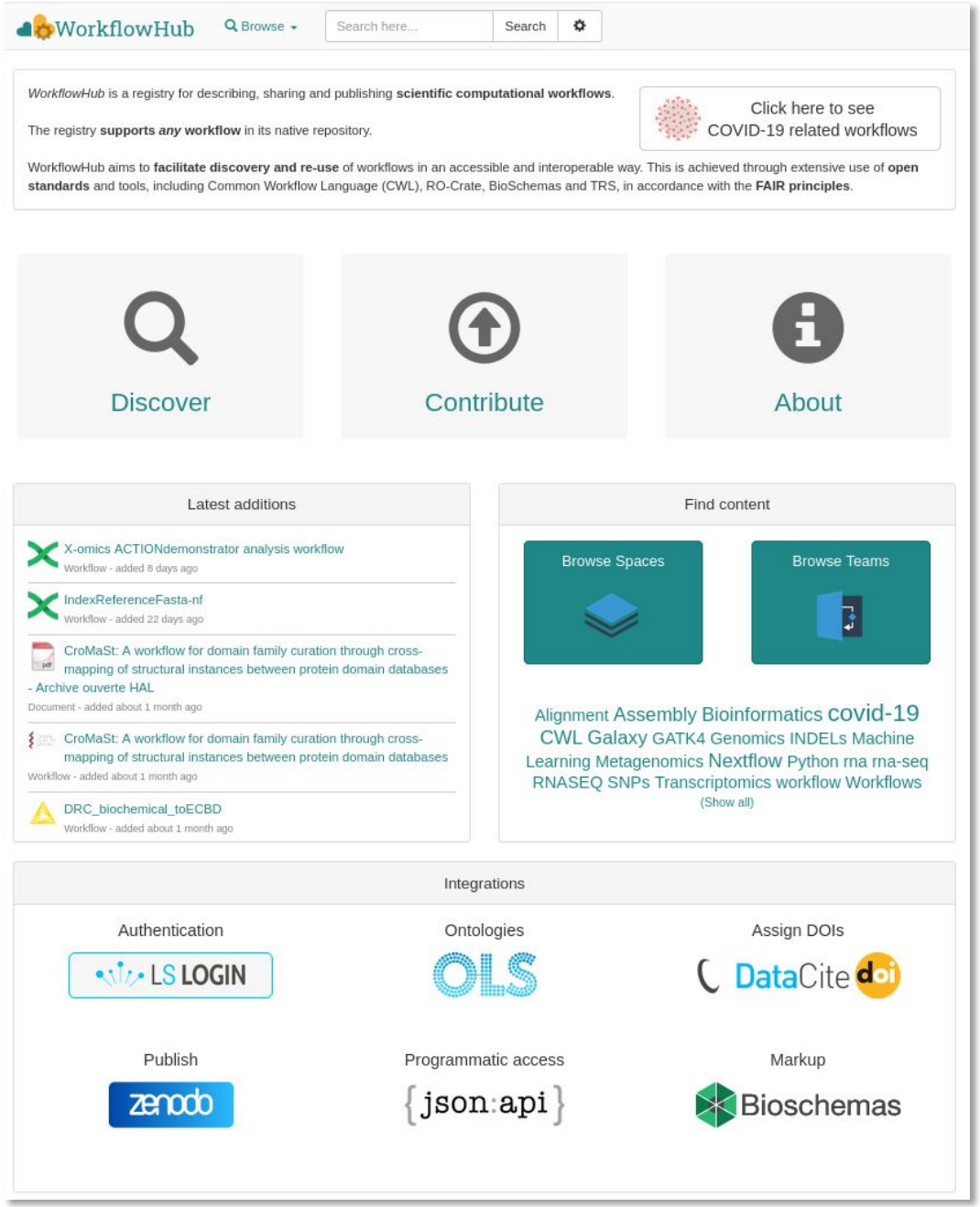
A FAIR workflow registry

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

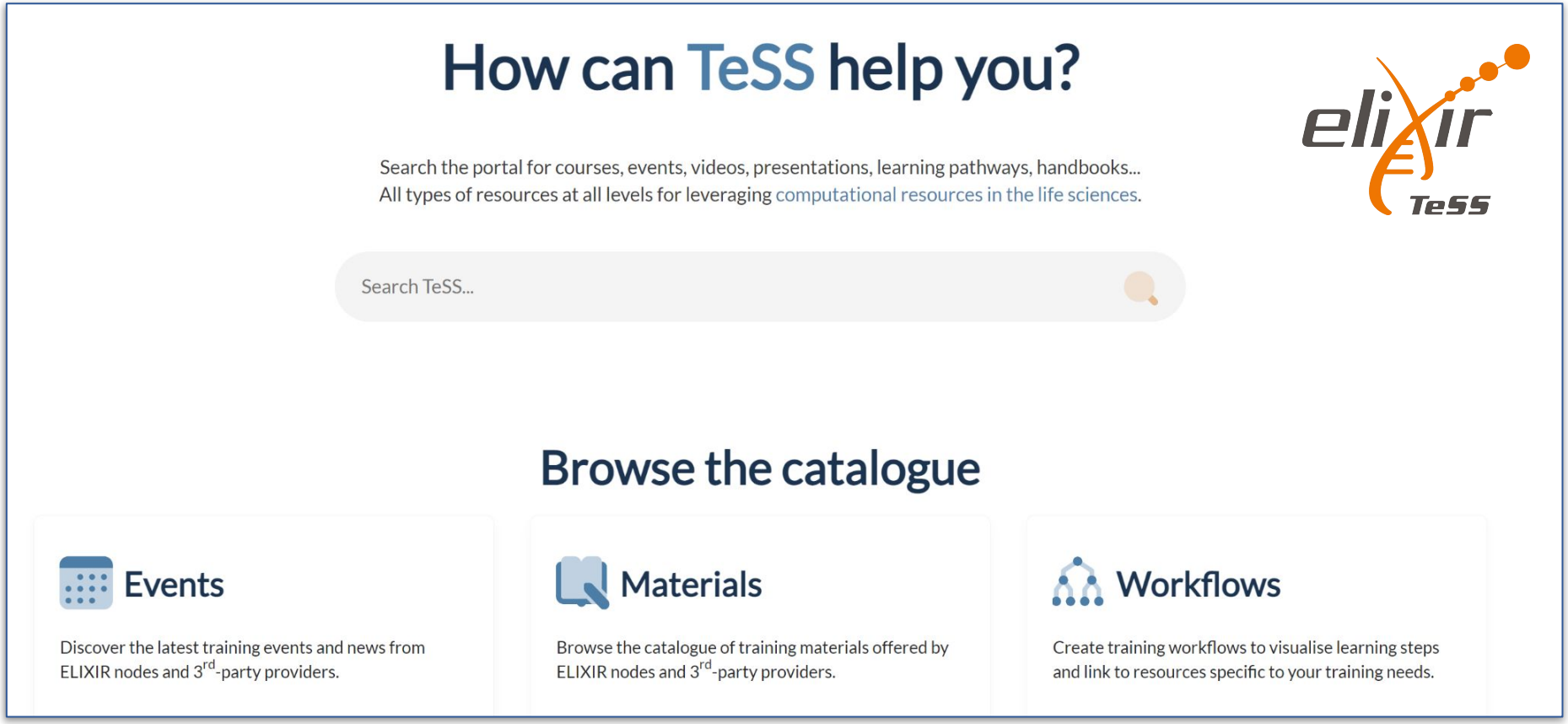
ELIXIR's Training Portal- TeSS (Training eSupport System)

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**



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

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

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.

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.

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

Thank you!

jessica.lindvall@scilifelab.se

www.elixir-europe.org

Overview of the presentation outline

Introduction: Start with a brief introduction to ELIXIR and EOSC. Explain what these acronyms stand for and why they are important. Give an overview of their objectives and the scope of their activities.

ELIXIR: ELIXIR is a research infrastructure for life science data. Explain what ELIXIR is and how it is organised. Provide an overview of its services, tools, and resources. Mention any relevant partnerships or collaborations that ELIXIR is involved in.

EOSC: EOSC is the European Open Science Cloud. Explain what EOSC is and how it relates to ELIXIR. Provide an overview of its objectives and the scope of its activities. Mention any relevant policies or initiatives that EOSC is involved in.

Benefits: Explain the benefits of ELIXIR and EOSC. How do they contribute to the advancement of science and technology? What impact do they have on research and innovation? Provide examples of successful projects or use cases that have benefited from ELIXIR or EOSC.

Challenges: Discuss the challenges faced by ELIXIR and EOSC. What are the obstacles to achieving their objectives? How are they addressing these challenges? What opportunities exist for further collaboration and innovation?

Future directions: Discuss the future directions of ELIXIR and EOSC. What are their priorities for the coming years? What are the emerging trends in life science data management and open science? How are ELIXIR and EOSC positioning themselves to address these trends?

Conclusion: Summarize your presentation and reiterate the key points you want your audience to remember. Encourage questions and discussion to engage your audience and ensure that they leave with a clear understanding of ELIXIR and EOSC.