

UK Data Research Infrastructures

Professor Carole Goble

The University of Manchester, UK

carole.goble@manchester.ac.uk

Joint Head of Node ELIXIR-UK, Joint lead Federated Analytics HDR UK

Investigator Software Sustainability Institute UK



BiolmagingUK 2023, 3 July 2023



<https://elixiruknode.org>
www.elixir-europe.org

UK Data Research Infrastructures

Biosciences and related areas

data commons



Cloud Infrastructure for Big
Data Microbial Bioinformatics



NERC
Environmental
Data Service



UK Data Research Infrastructures

Biosciences and related areas



National institute for health data science.
Large-scale data and advanced analytics

DARE UK

Programme - coordinated and trustworthy national data research infrastructure, cross-domain research.

Centres of excellence with expertise and tools developing data to provide insights



Ethical use of UK health data for research

HDR-UK Technology Ecosystem

Trusted Research Environments, Gateways, Federated Analytics, **Five Safes**



Common Themes- Collaborative Commons



- **Aggregating data** generated from different sources
- Making data and software **FAIR**
- **Integrating** researcher data with reference data
- Exploiting the accumulation of data to enable data analytics and **Machine Learning/AI**
- **Data Integration & ETL** – ingestion, validation, transformation & loading into public data stores
- Maintaining and **curating data** for the long term
- **Secure Data Environment** for scenarios involving sensitive data
- Supporting **analysis workflows** from different data sources at scale, federated analytics
- One stop support for accessing and executing tools
- **Key services** for AAI, storage, cataloguing, compute, metadata management etc
- Data Stewardship and Research Software Engineering, **best practices & capacity building**
- Setting and supporting **Standards**
- Data and software **publishing**
- **International and national partnerships**

Wayfair/ Van Guard / Driver
Projects & Communities

UKRI DRI Congress

<https://www.ukri.org/what-we-offer/creating-world-class-research-and-innovation-infrastructure/digital-research-infrastructure/>

UKRI DRI Club

<https://tinyurl.com/UKRI-DRI-Club>

Landscaping exercise

Next meeting 4 July 2023 (tomorrow!).

ELIXIR-UK, a node of ELIXIR the European Research Infrastructure for Life Science Data

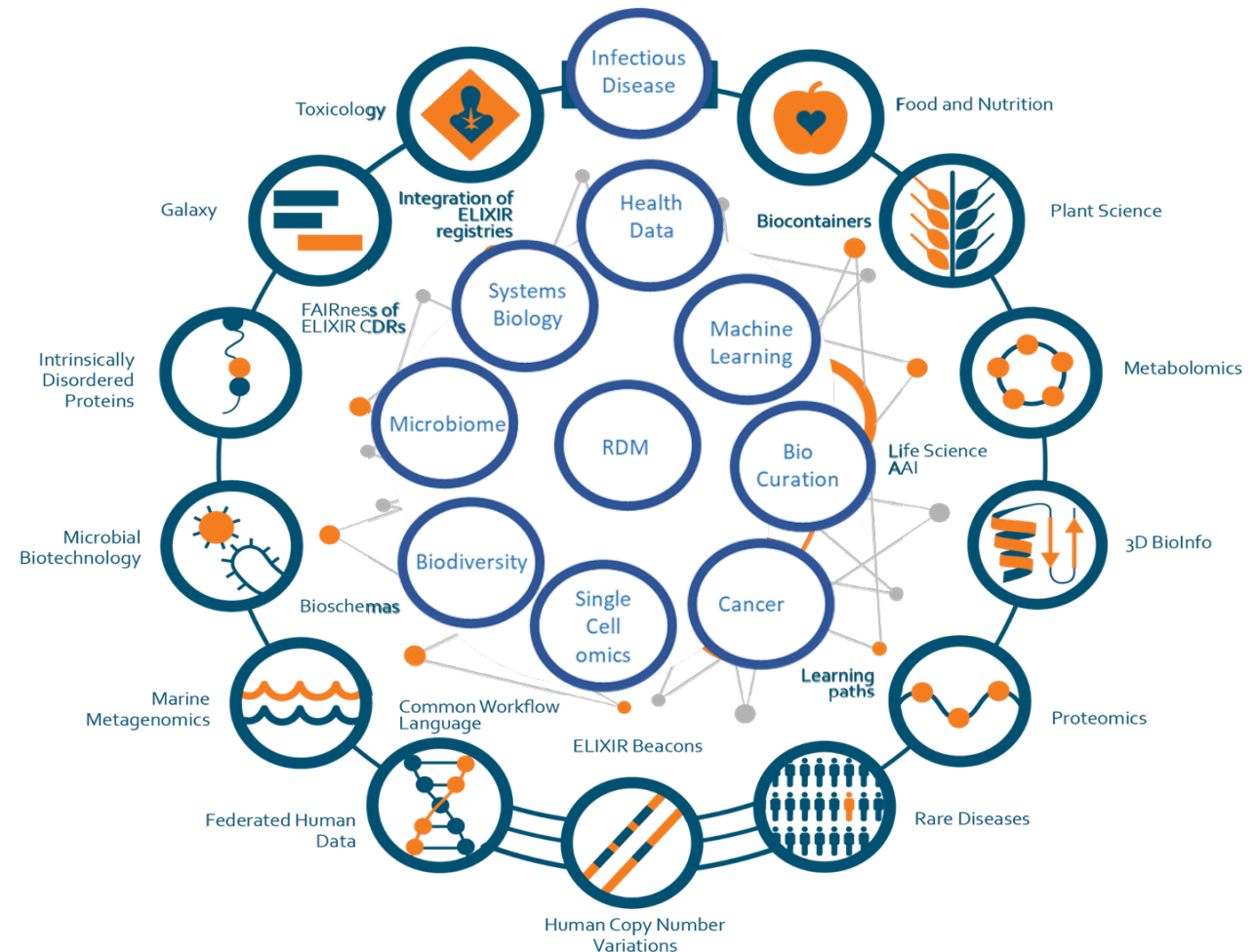
Intergovernmental 23 nodes, 250+ organisations

A federated European digital infrastructure for Life Science Data,

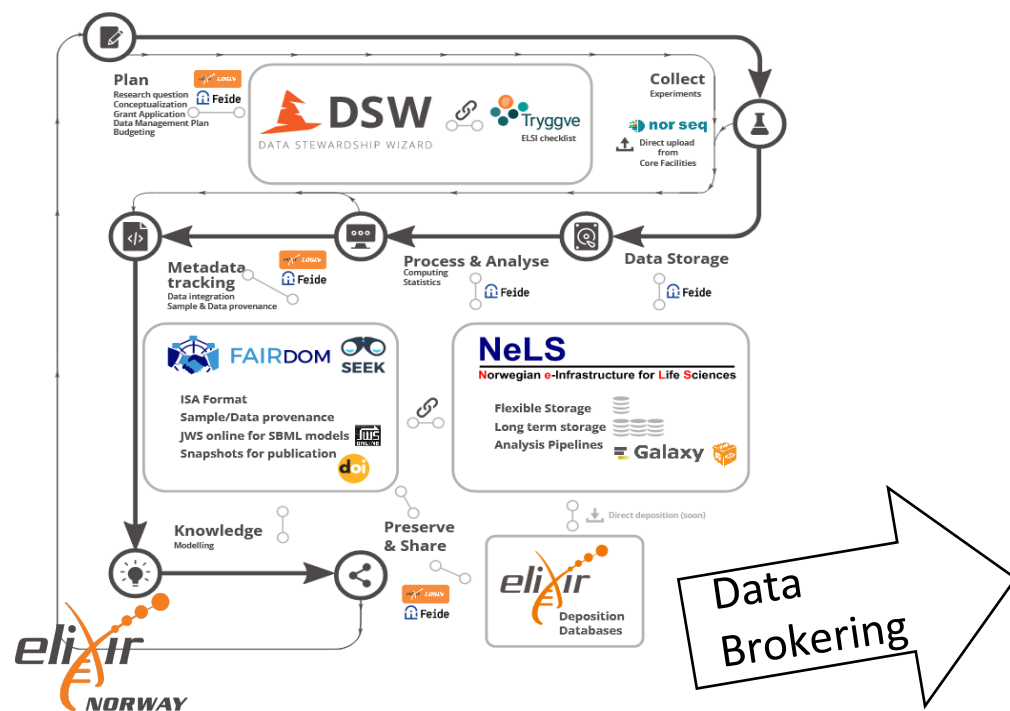
Coordinate national and European capabilities.
Sister to Euro-BioImaging & partner in EOSC-Life and other projects

Data & software FAIR and open as possible
Data analytics

Research Data Management knowledge and capability.



ELIXIR-UK, Research Data Management



National project research data support

Trusted data resources



RDM Stewardship network, fellowships, training

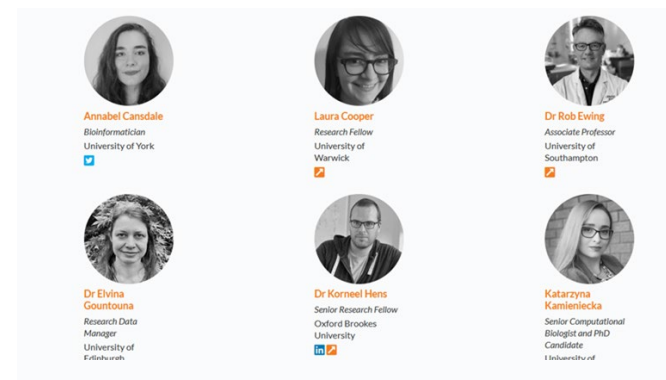


Image Data
11 videos

ELIXIR-UK, Research Data Knowledge

RDMkit

The RDMkit website features a grid of logos for various research data management tools and a central circular diagram with icons for different data types. The logos include PIPPA, CorkOakDB, DisProt, UniProt, ENA, ChEMBL, PRIDE Archive, IntAct, MetaboLights, Rhea, European Variation Archive, BioStudies, and BRENDA. The central diagram has icons for Data, Metadata, Images, Audio, Video, Text, and Software.

FAIRsharing.org
standards, databases, policies

The FAIRsharing.org website shows a search bar and a list of search results. The search results include a table with columns for Name, Description, and Date. The table lists various standards and databases, including the European Bioimaging ERIC, the European Variation Archive, and the BioStudies database.

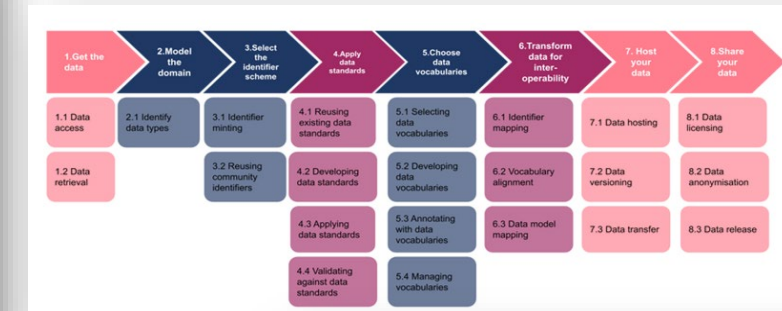
elixir
TeSS

The elixir TeSS website shows a search bar and a list of search results. The search results include a table with columns for Name, Description, and Date. The table lists various standards and databases, including the European Bioimaging ERIC, the European Variation Archive, and the BioStudies database.

elixir
bio.tools

The elixir bio.tools website shows a search bar and a list of search results. The search results include a table with columns for Name, Description, and Date. The table lists various standards and databases, including the European Bioimaging ERIC, the European Variation Archive, and the BioStudies database.

FAIR
cookbook

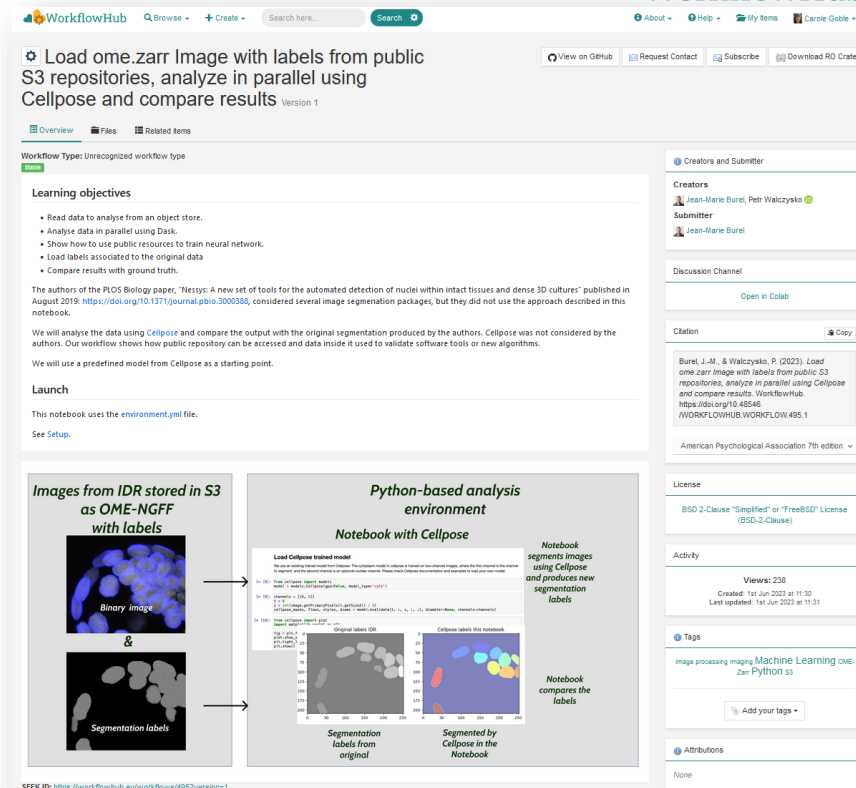
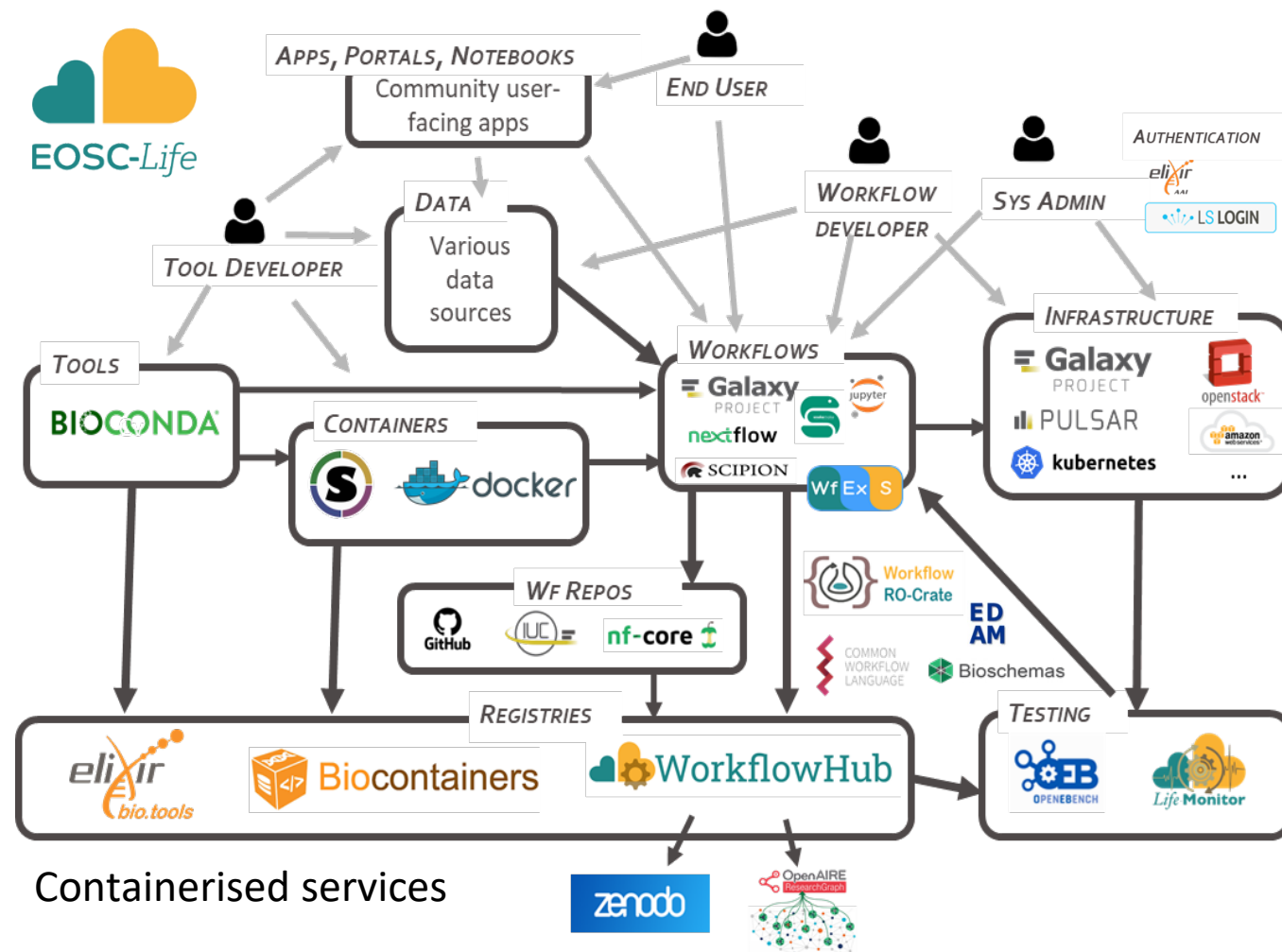


The Plant genomic and genetic variation data submission to EMBL-EBI website shows a form for submitting data. The form includes fields for Name, Description, and Date. The website also features a search bar and a list of search results.

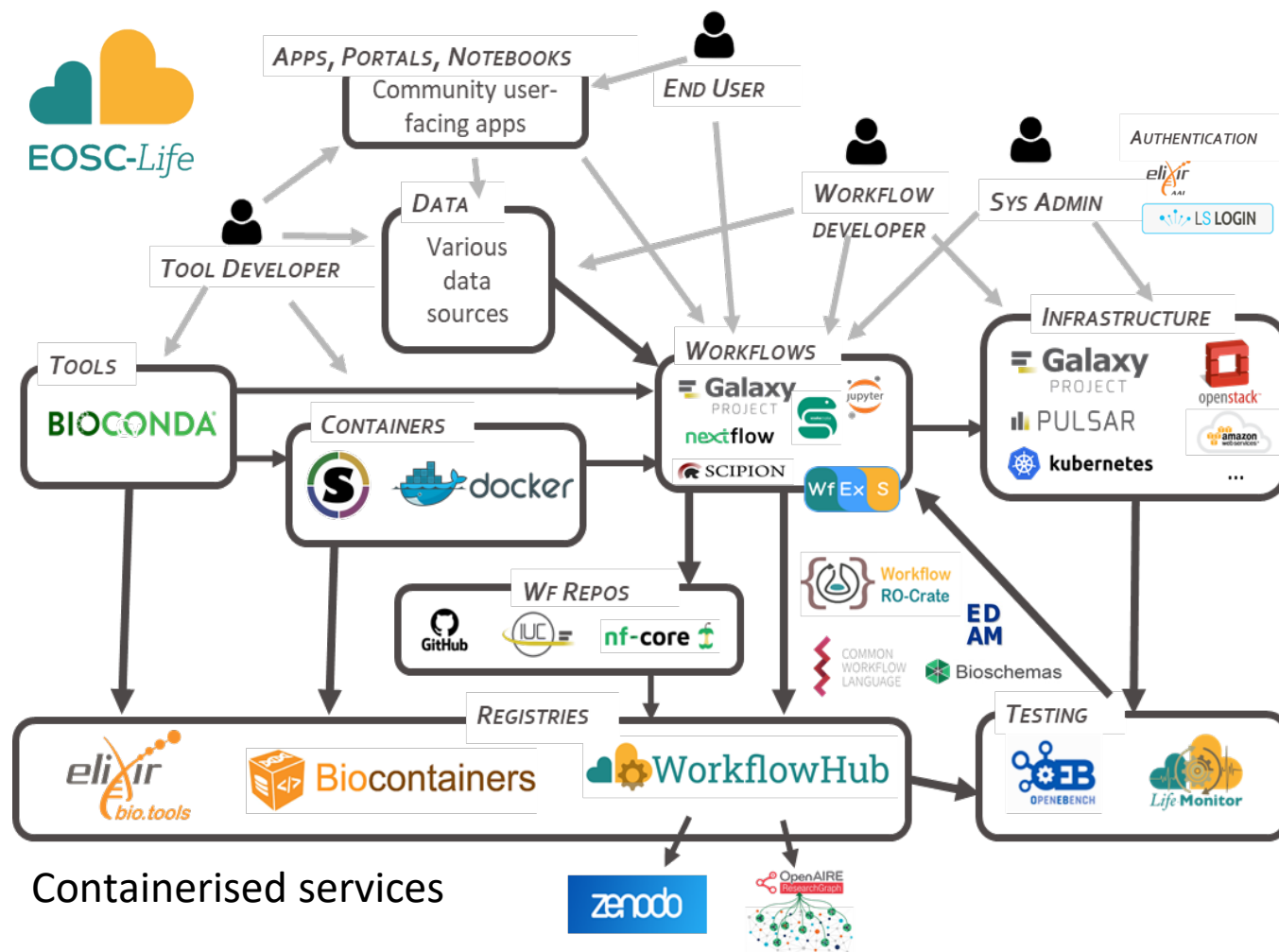
elixir
UNITED KINGDOM

The Data Stewardship Wizard website features a large illustration of a person standing next to a large screen displaying a play button. The text on the page reads: "Data Stewardship Wizard. Create, plan, collaborate, and bring your data management plans to life with a tool trusted by thousands of people worldwide — from data management pioneers, to international research institutes."

ELIXIR-UK, Computational Workflow Data Analysis



ELIXIR-UK, Computational Workflow Data Analysis



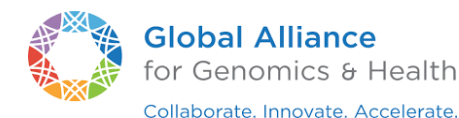
FAIR
Digital
Objects



Metadata
markup

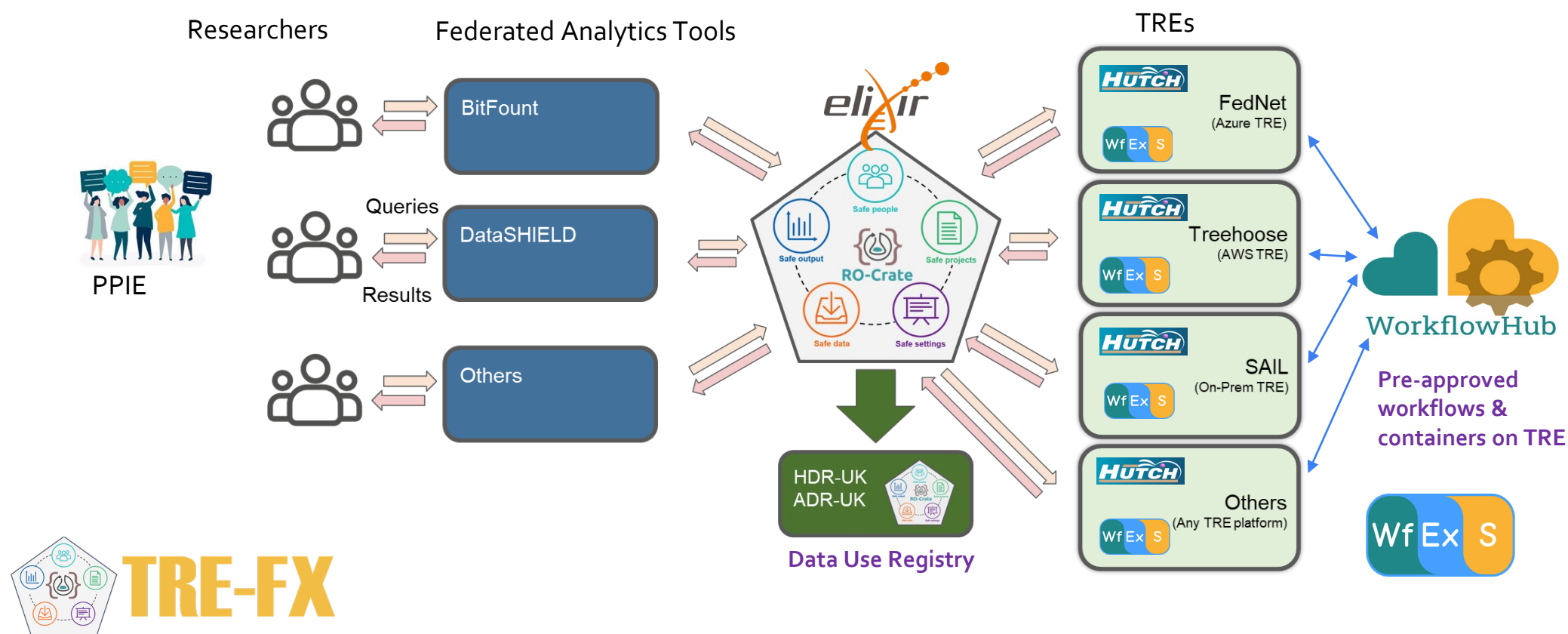


Standards
& APIs



Data Research Infrastructure **Collaboration**

Federated analytics across HDR-UK TREs using ELIXIR's RO-Crate Digital Objects and Computational Workflow Infrastructure



<https://trefx.uk/>



<https://biofair.uk/>

A Data and Computational
Method Commons
infrastructure for biological
and biomedical sciences

19 June 2023



[Home](#) > [News](#) > UKRI invests £72 million upgrading UK research infrastructure

UKRI invests £72 million upgrading UK research infrastructure

BioFAIR

BioFAIR, a new digital infrastructure, will be established to support research in biological and biomedical sciences. It will widen access to existing data processing, analysis and repository infrastructures, from organisations like [ELIXIR](#), to maximise the findability, accessibility, interoperability, reusability (FAIR) and reproducibility of researchers' data.

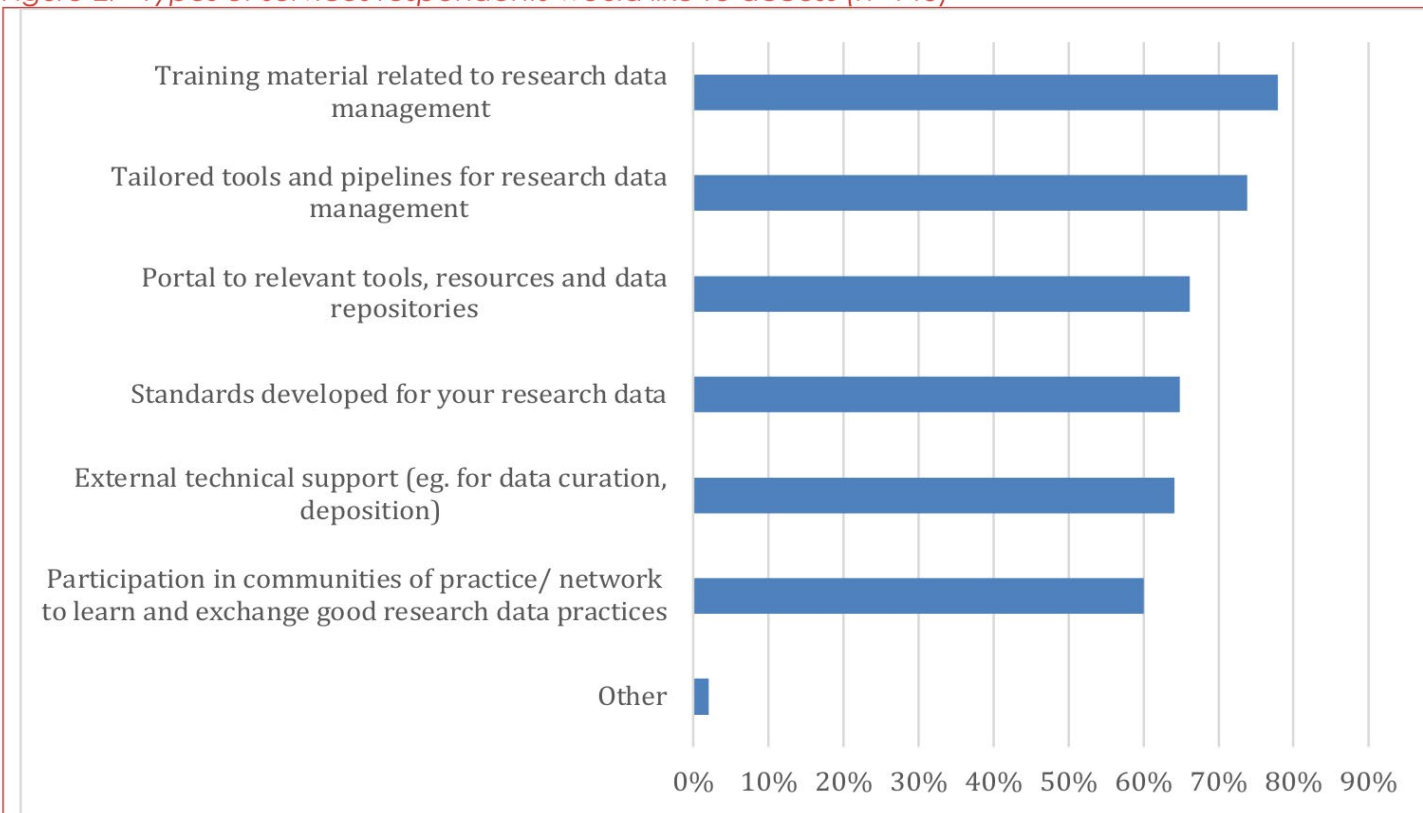
Bridging the gap between researchers, data sources at individual institutions, and existing data infrastructures, BioFAIR will accelerate discovery and aid research advances in fields as diverse as medicine and agriculture.

Funding of £34 million over five years is subject to business case approvals.

Address fragmented,
patchy, DIY data
management and data
analysis

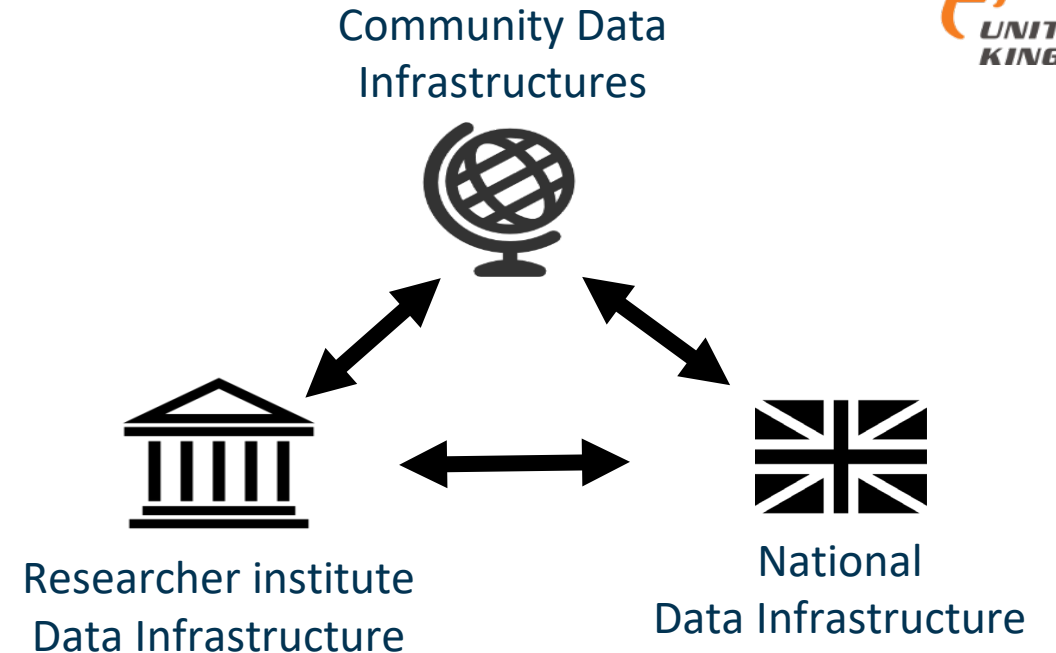
Support UK's researchers,
data stewards and DRI
providers with a national
capability

Figure 27 Types of services respondents would like to access (n=145)



BioFAIR Feasibility Study (2020-2021)

- **Method Hub** - National data analysis facility using workflows and notebooks over dedicated & national compute.
- **Data Hub** - Data standards, cataloguing, support over federated data lake, linked to other data hubs.
- **Data brokering** - deposition archives, data store & retention for projects, QA, AI readiness.
- **Knowledge Hub** - RDM and analysis **know-how**, stewardship and training.
- National **Portal**



Own data, own analysis, own tools
Share data, shared analysis, shared
tools

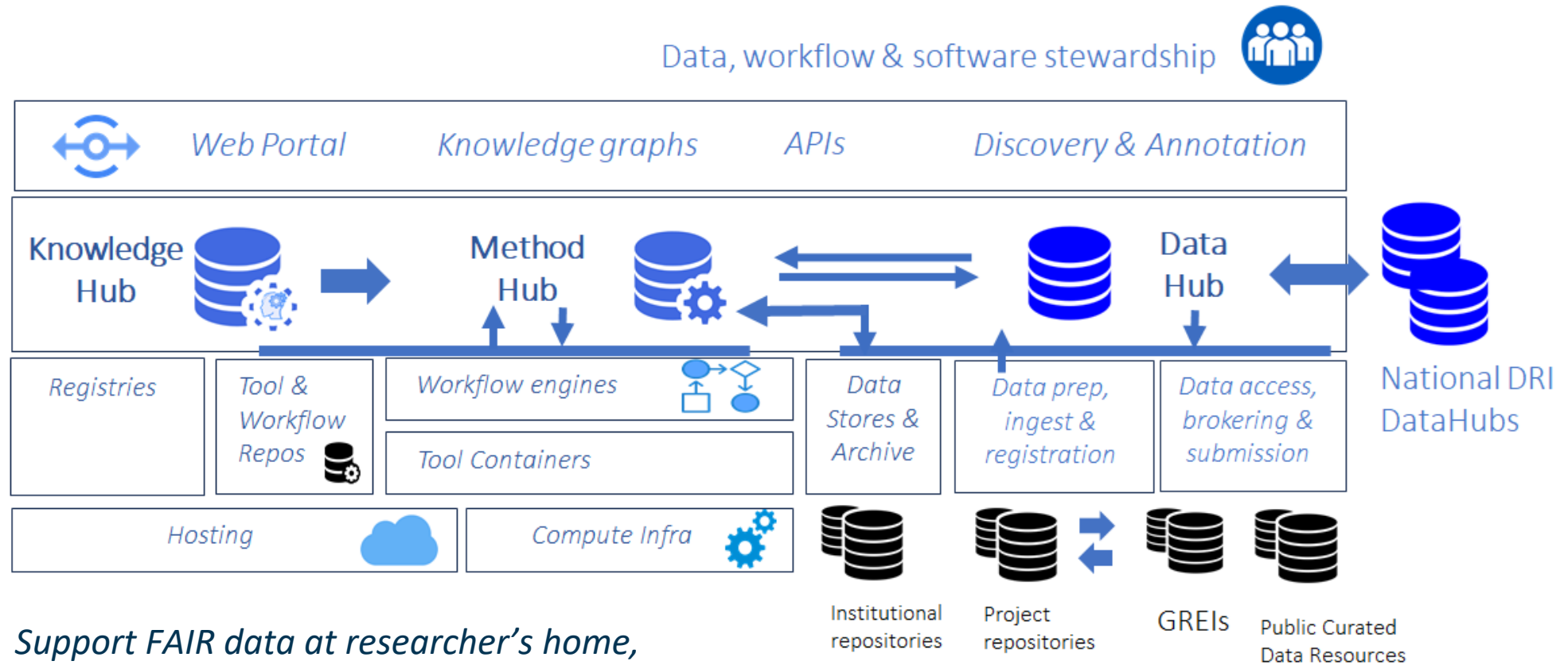
Digital Objects & Standards



End-to-end project data management and analysis

Coherent ecosystem of tools and services

FAIR Digital Objects based



*Support FAIR data at researcher's home,
FAIR analysis using their data safely and
confidentially*

Shameless advert



1st Conference on Research Data Infrastructure

12-14 September 2023
Europe/Berlin timezone

<https://indico.scc.kit.edu/event/3453/>