BioFAIR: a new BioCommons infrastructure for UK life science

ELIXIR Europe

Intergovernmental organisation (EMBL special project) founded in 2013.

Coordinates life science resources towards a federated European digital infrastructure:
- Data & software FAIR and open as possible
- Transnational access and analysis

National Nodes + EMBL-EBI Node provide **Services** to the ELIXIR Portfolio (300+).

Gateway
- to international initiatives & standards bodies
- for Communities of Practice

Advocacy and policy making.

23 Nodes

220+ orgs

Hub @Wellcome Genome Campus

https://elixir-europe.org

Trusted data resources
Public deposition databases, data resources and portals, scalable curation, sustainability.

RDM platforms & support
Data Management Planning, data FAIRification, data brokering to public deposition databases.

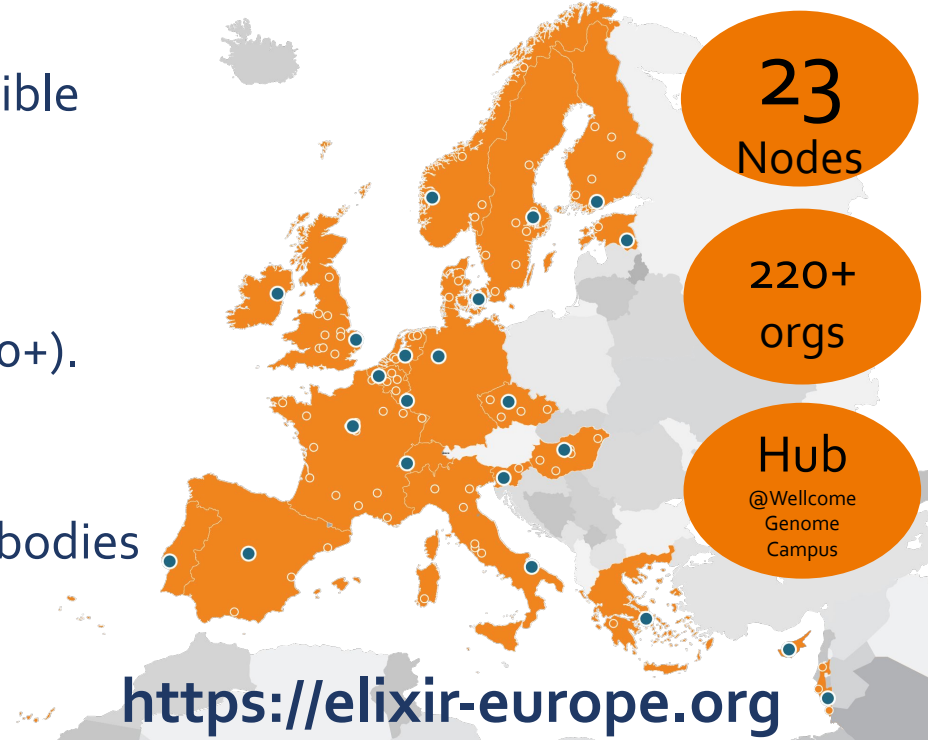### FAIR data services
AAI, data transfer, workflow, tool and container registries, ontologies, identifiers, stewardship tools, standards.

Data analytics and platforms
Computational workflow platforms, reproducible and portable processing, software and AI best practices, federated analytics.

Knowledge management
Best practice, guidelines, shared experience and know-how, data FAIRification methodologies.

Stewardship and training
Capability frameworks, skills, data managers network, training portal, train the trainers, data stewardship fellowships.

The ELIXIR-UK Node

The national Node of ELIXIR in the United Kingdom, providing coordination and capacity building.

A consortium of 23 UK research organisations.

Supplies many of the ELIXIR Services to European and international partners.

Participates in a raft of ELIXIR Communities, activities and projects, leading many of them.

Works with UK funders on policy & infrastructure provision & skills.

23 orgs

https://elixiruknode.org

The ELIXIR-UK DaSH Fellowship

Project to embed Research Data Management (RDM) know-how into UK universities and institutes:
- Building capacity and professionalising **data stewardship** while leveraging the local expertise
- Creating a data management **culture change** from within the institutions

24 fellows (2 cohorts, 17 organisations) producing and delivering **training** in FAIR data stewardship using ELIXIR-UK knowledge and resources, e.g. RDMkit and the FAIR Cookbook.

Funded as part of the "UKRI Innovation Scholars:Data Science Training in Health and Bioscience (DaSH)" call.

RDMbites. a series of short videos on all things research data management

The need for a UK BioCommons

In 2021 ELIXIR-UK (together with Technopolis Ltd) undertook a 9-month **study** to:
- identify the **Research Data Management (RDM) needs** of UK bioscience researchers
- evaluate the **feasibility and impact** of the creation of a new UK Institute delivering a FAIR Data and Method BioCommons

This was funded by BBSRC, MRC and Wellcome, and included desk research, a community survey, dedicated interviews, and a technical analysis and scope.

Figure 19 Main reasons why research data and other digital assets are not made available for secondary use or reuse (n=140)
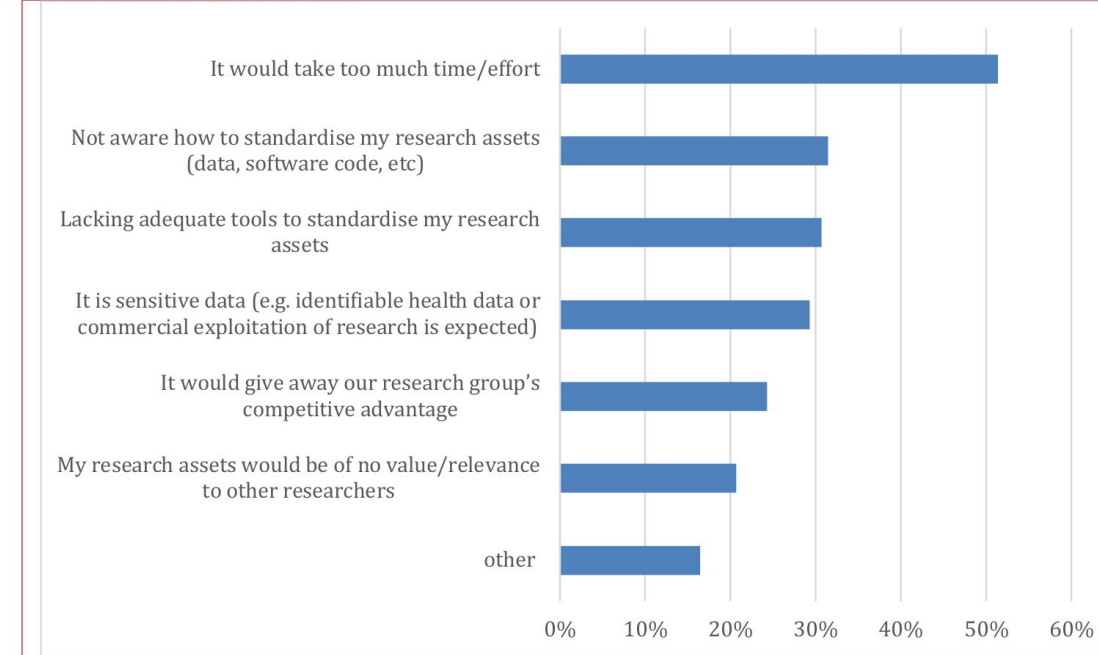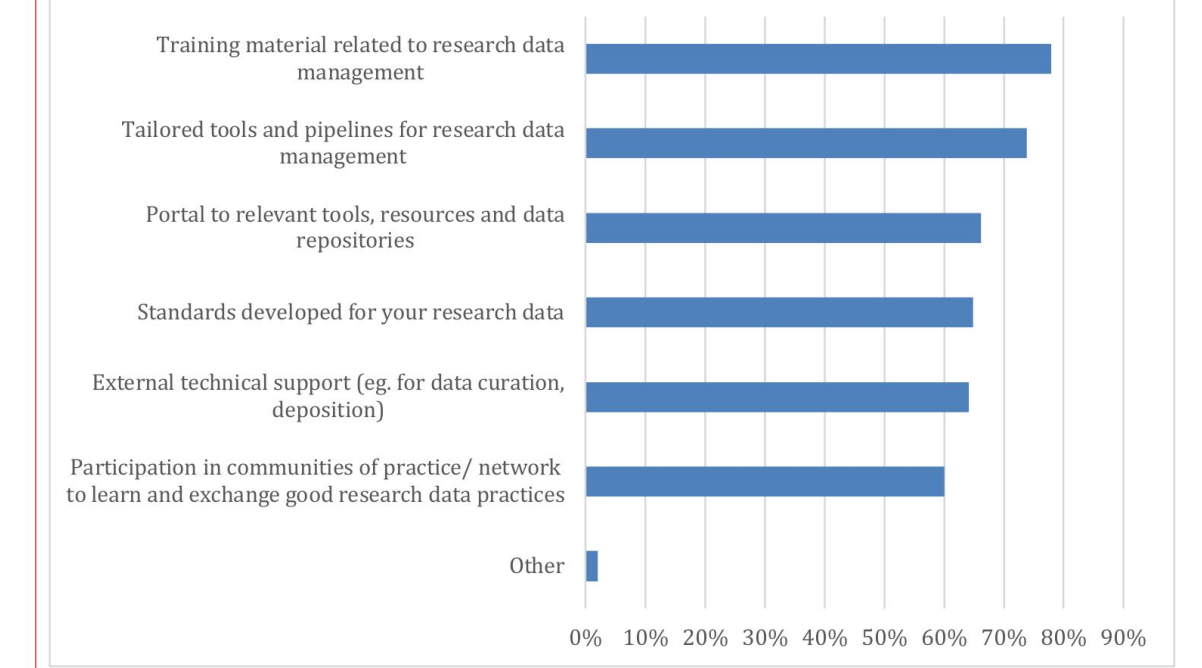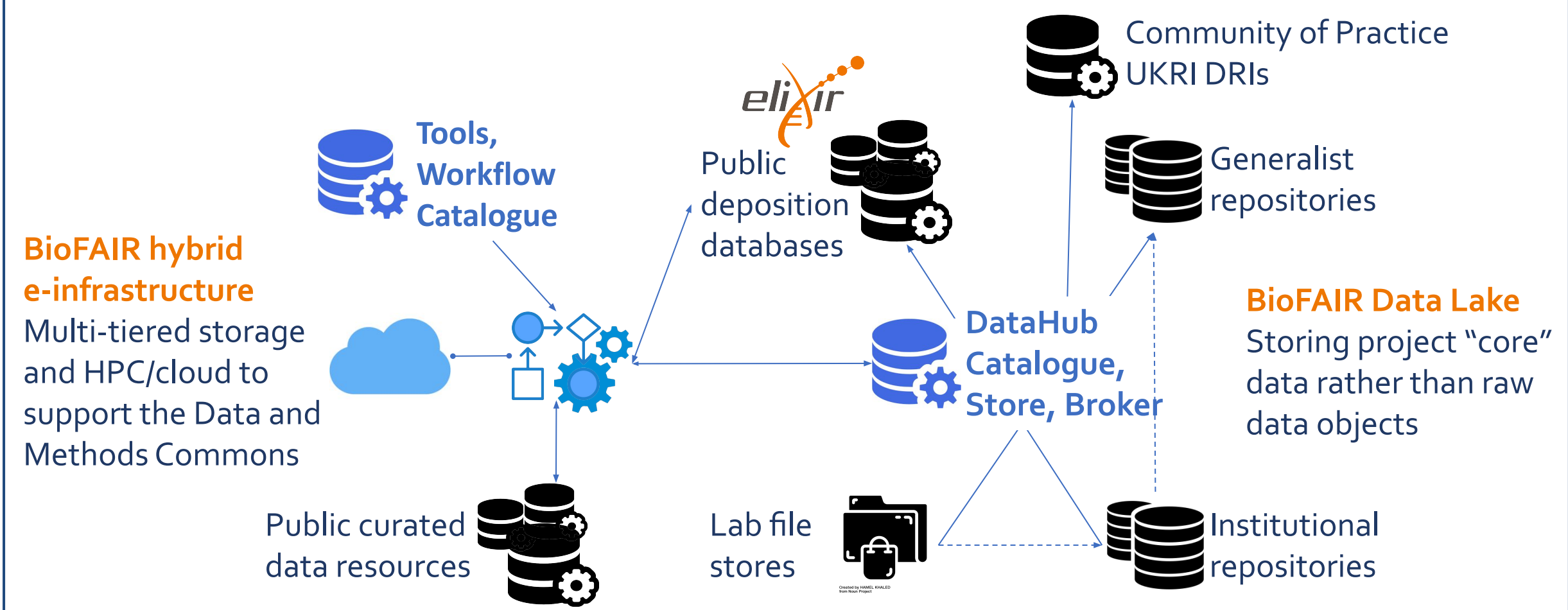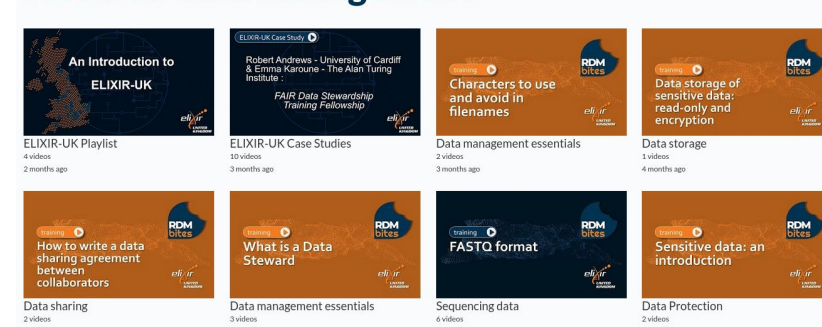
Figure 27 Types of services respondents would like to access (n=145)

- Only **30%** of the UK life sciences community are compliant with the FAIR principles
- The study recommended the creation of a BioFAIR virtual institute to:
  - provide a data commons and a new digital infrastructure for the whole of the UK biosciences research community (20,000+ researchers)
  - support the adoption of FAIR RDM principles in order to produce a step-change in the sharing and re-use of research data and meaningful improvements in academic productivity, scientific quality and reproducibility and social benefits through greater knowledge spillovers.

BioFAIR: a new Data and Methods Commons

Tools, Workflow Catalogue

Public deposition databases

Community of Practice UKRI DRIs

Generalist repositories

DataHub Catalogue, Store, Broker

BioFAIR hybrid e-infrastructure
Multi-tiered storage and HPC/cloud to support the Data and Methods Commons

BioFAIR Data Lake
Storing project "core" data rather than raw data objects

Public curated data resources

Lab file stores

Institutional repositories

BioFAIR Data Commons
- National **data catalogue** over a federated data lake
- Support **FAIR data at Research Performing Orgs** and project sites
- **Data brokering** to public archives
- Leverage ELIXIR's FAIR services, catalogues and data archives.
- FAIR Research Data Management advocacy

BioFAIR Computational Methods Commons
- National **analysis capability** using computational workflows and notebooks
- Curate and share **best practice tools & workflows**
- Integration with data catalogue & stores
- **AI readiness** and QA
- Leverage ELIXIR's computational tools and workflow capability

BioFAIR Integrated Data Services
- **Metadata and Standards** framework based on **FAIR Digital Objects and APIs**
  - Bioschemas, RO-Crate, ISA; leveraging EU infrastructure developments.
- Life Science AAI Infrastructure (LS-Login)
- Data and method **stewardship** & User support
- **Communities of Practice**
  - working with other DRIs
  - C-MASS, EuroBioimaging-UK, HDR-UK, NERC Data Commons, etc.
- **Training** about the Commons and data stewardship; train-the-trainers
- One-stop shop web portal and APIs

UKRI's Infrastructure Fund will invest initial seed funding of £6.3 million over the next three years; case for further funding proposed to IF/DRI.

Contact

Prof. Carole Goble
carole.goble@manchester.ac.uk
The University of Manchester
Manchester, M13 9PL, UK

Dr. Nicola Soranzo
nicola.soranzo@earlham.ac.uk
Earlham Institute
Norwich, NR4 7UZ, UK

ELIXIR-UK is coordinated from the Earlham Institute in Norwich and led by its Joint Heads of Nodes: Prof. Carole Goble (The University of Manchester) and Prof. Neil Hall (Earlham Institute).

contact@elixiruknode.org
@ElixirNodeUK