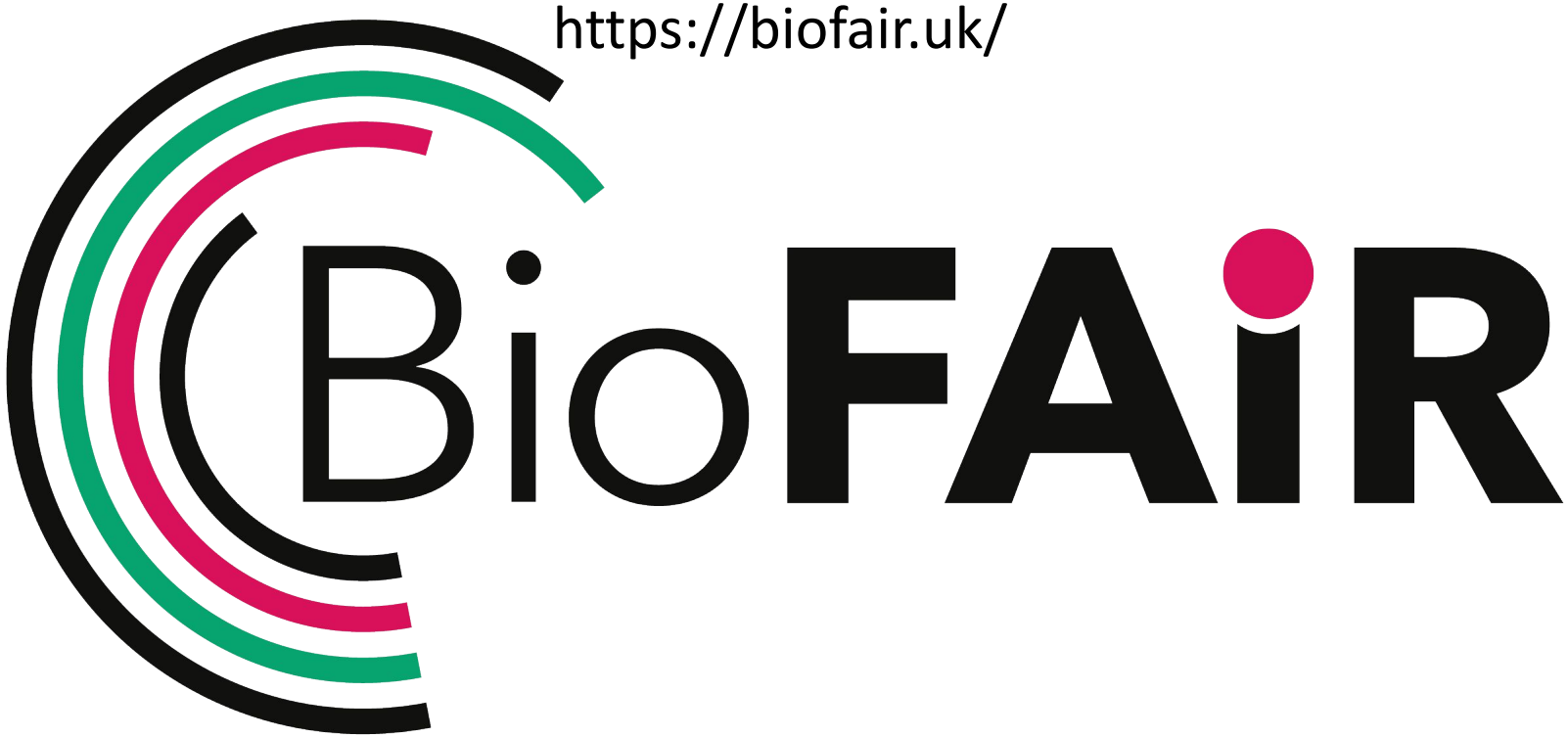
https://biofair.uk/



UK Research and Innovation

A Digital Research Infrastructure for UK Life Sciences

Gerry Reilly
Interim Director, BioFAIR UK
gerry.reilly@earlham.ac.uk

Eva Wan
Project Manager, BioFAIR UK
eva.wan@earlham.ac.uk

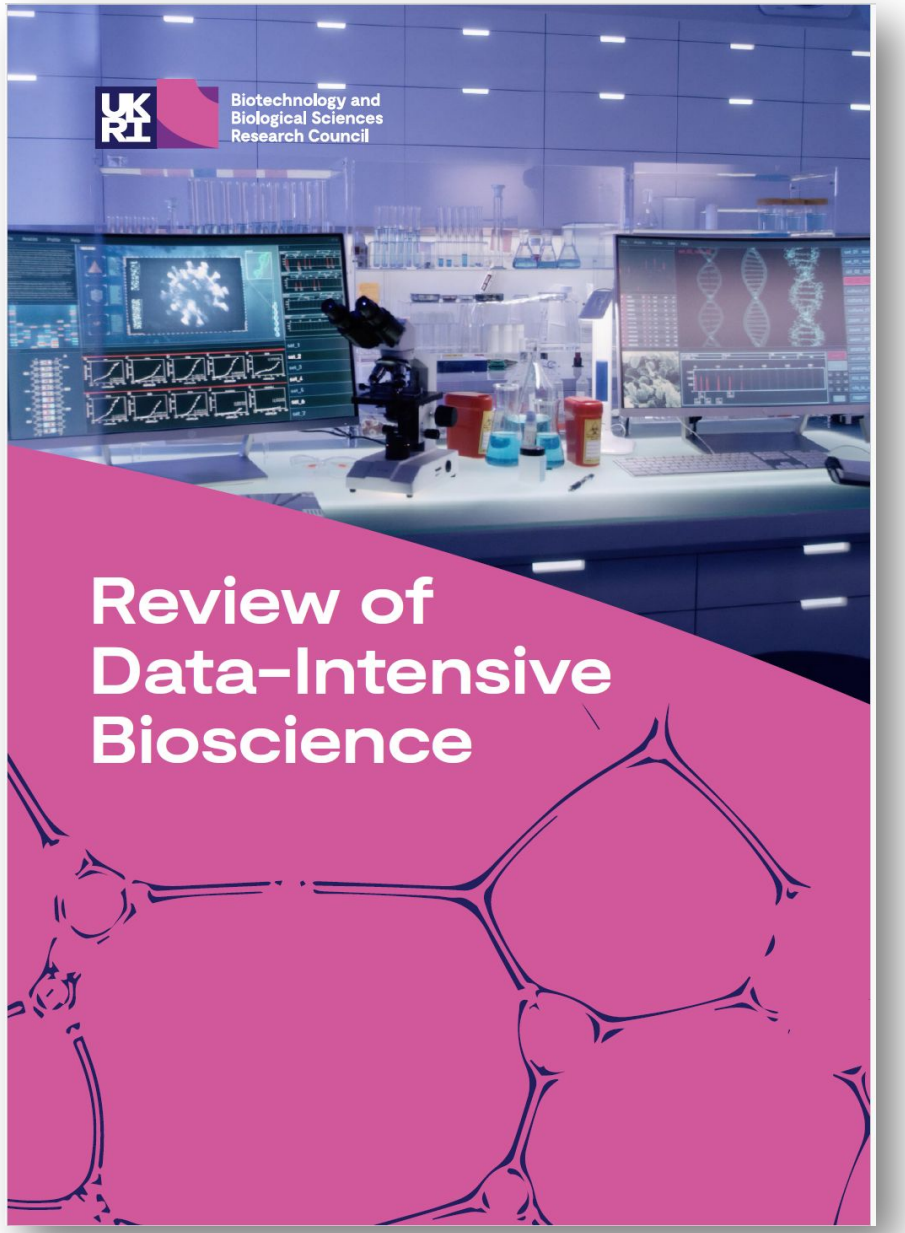
The life sciences are data rich disciplines

…

as reliant on sophisticated analysis of multiple complex datasets as laboratory-based experimentation

with vast increase in pace of experimentation, computational analysis and volumes of data

a team sport using different data types

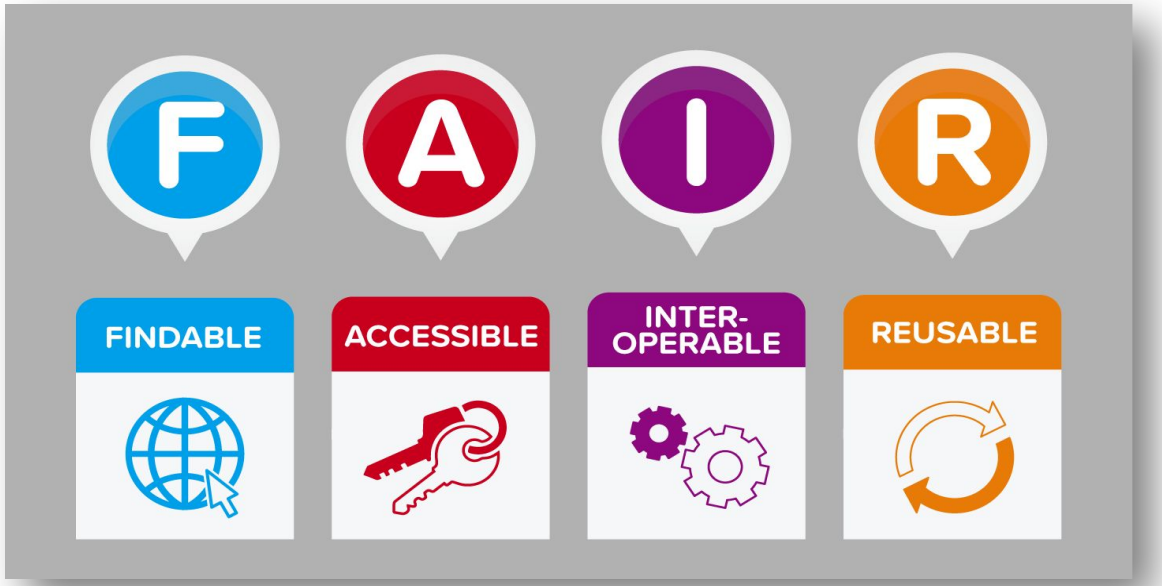


100+ UK Research Performing Organisations (RPOs)

100,000+ UK Researchers

100+ Project Consortia

…that should operate using FAIR Principles



FAIR Data drives AI

What does FAIR include?

Data

Metadata

Data Methods, Workflows, SOPs

Software

Training materials

Other data services and tools

What doesn't FAIR include?

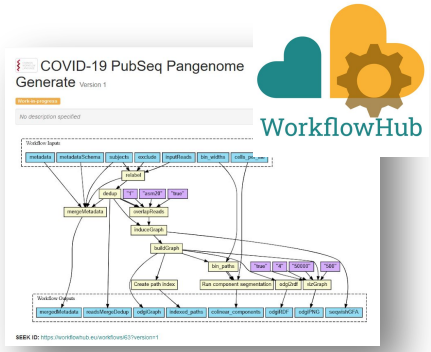
Open data

Data quality

Wilkinson *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 3 (2016)
Goble *et al* FAIR computational workflows. Data Intelligence 2(2020)
Barker *et al*, Introducing the FAIR Principles for research software. Sci Data 9, 622 (2022)

Life sciences researchers need to have access to rich data services

Data methods – to manipulate and analyse their own and other data reproducibly

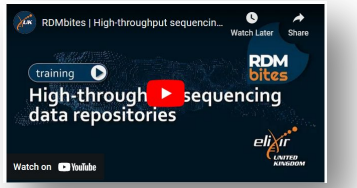


Research Software Engineers & Stewards

Data resources – to manage, store and curate their data, or search for and access others' data

Data Stewards

Training and Knowledge – to learn how to use and apply these data methods and resources and share knowledge



RDMkit
FAIR cookbook
FAIRsharing.org standards, databases, policies
elixir TeSS

Trainers

Communities – to discuss issues, share and promote good practice, support stewards and engineers, make pathways for FAIR data in their fields.

Champions

enabling and applying FAIR principles

https://elixir-europe.org/
https://elixiruknode.org/

European wide FAIR data services with UK leadership

25 Countries

28 Organisations

Navigating a rich and complex landscape

BioFAiR

Data methods and services

RStudio
Docker
Life Monitor
elixir bio.tools
jupyter
kubernetes
elixir bio.tools
COMMON WORKFLOW LANGUAGE
Galaxy EUROPE
WorkflowHub
snakemake
OPENEBENCH
nextflow

Training & Knowledge resources

FAIR cookbook
DSW DATA STEWARDSHIP WIZARD
RDM bites
RDMkit
elixir TeSS
FAIRsharing.org standards, databases, policies

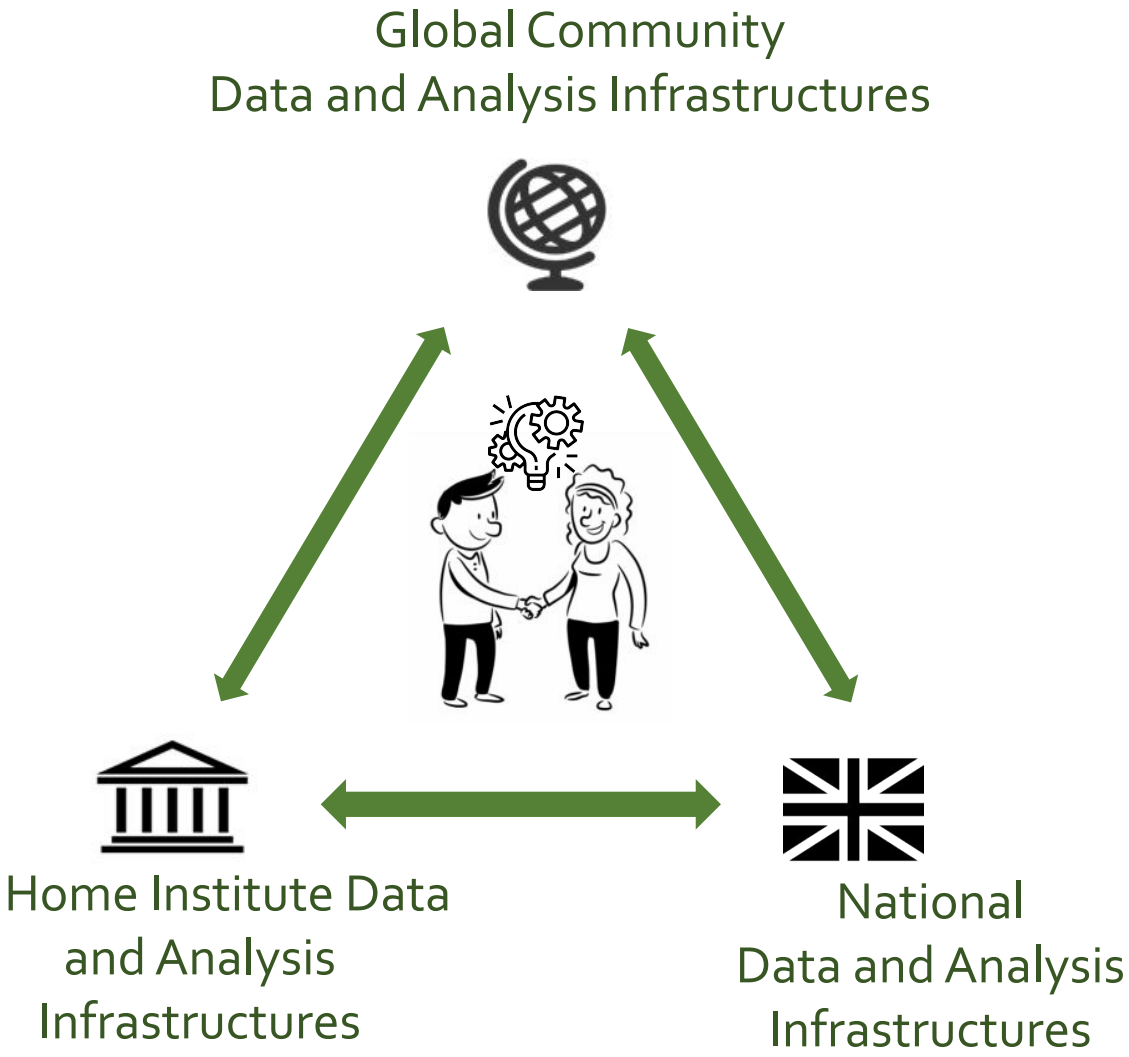
Data resources

Bioschemas.org
Global Alliance for Genomics & Health Collaborate. Innovate. Accelerate.
ROCrate
PRIDE PRoteomics IDEntifications Database
isatools
FAIRDOM SEEK
COPO
EMBL-EBI

Community groups

elixir DaSH UNITED KINGDOM
Galaxy Training Network
Ed-DaSH
AIBIO UK
THE CARPENTRIES
RSE

Navigating a rich and complex landscape



A coherent ecosystem of services for the UK

Own data & analysis

Shared data & analysis

Common data, analysis, tools, standards, services, support

Technopolis Study of the UK (2021)

Figure 19 Main reasons why research data and other digital assets are not made available for secondary use or reuse (n=140)

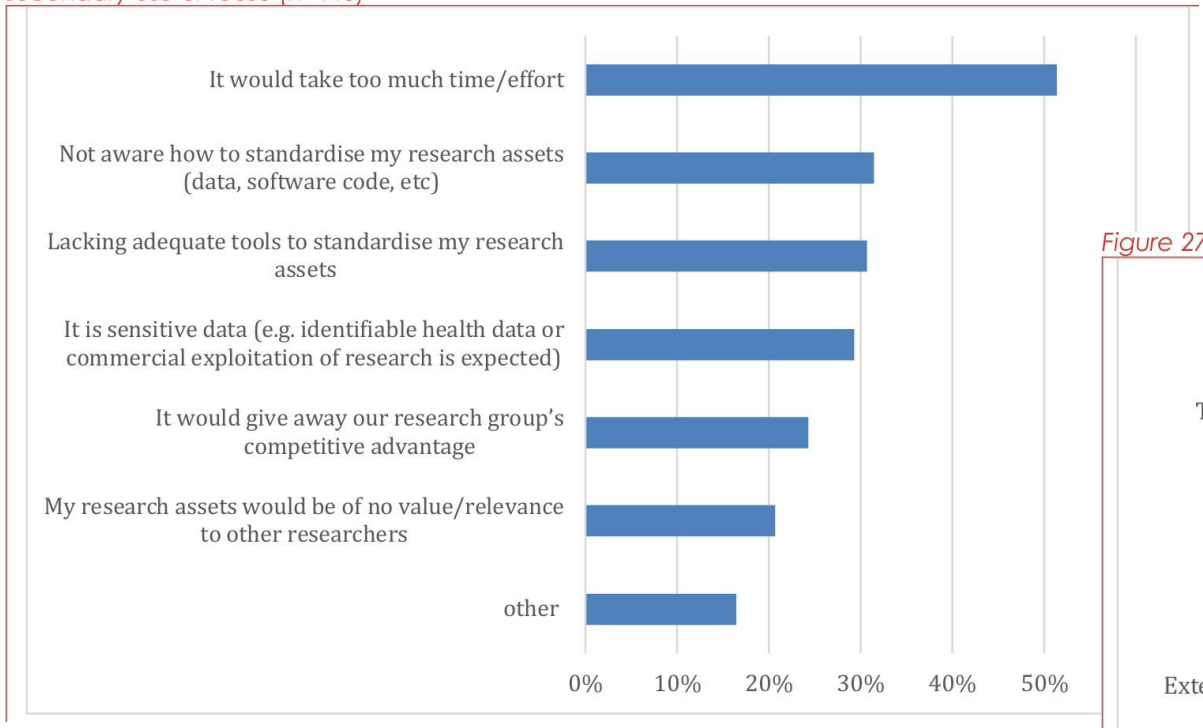
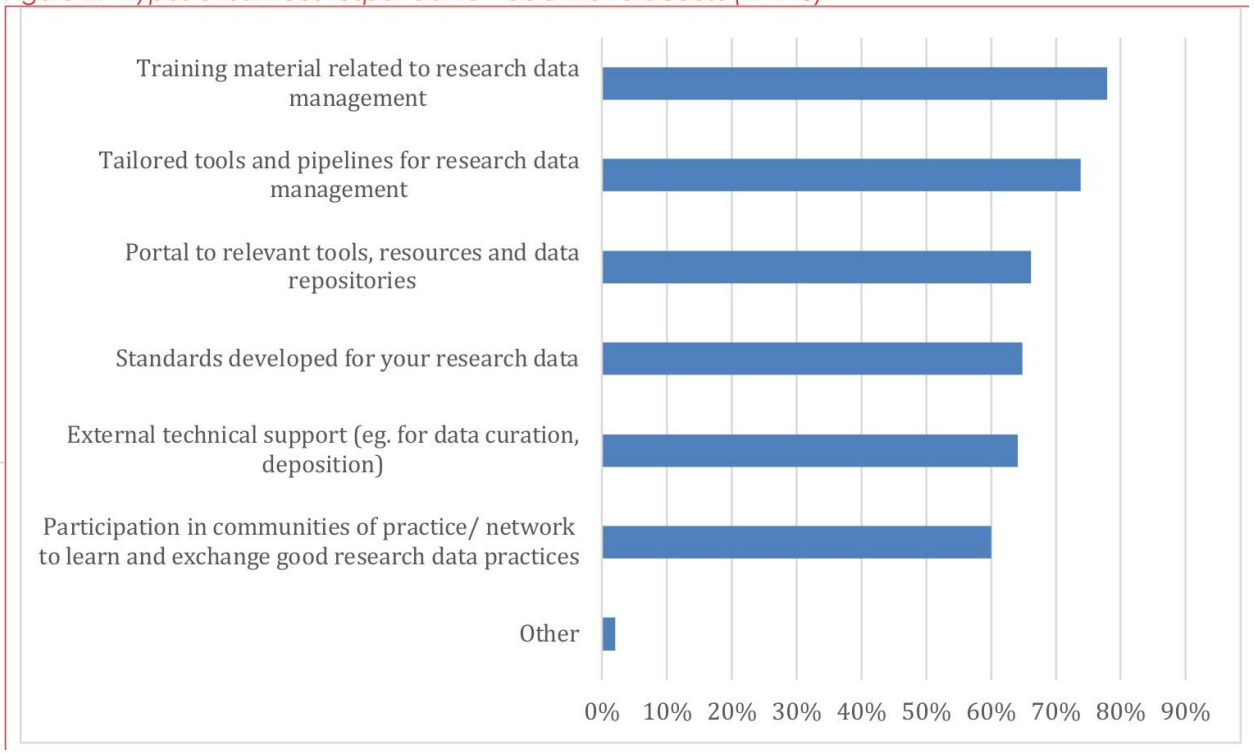
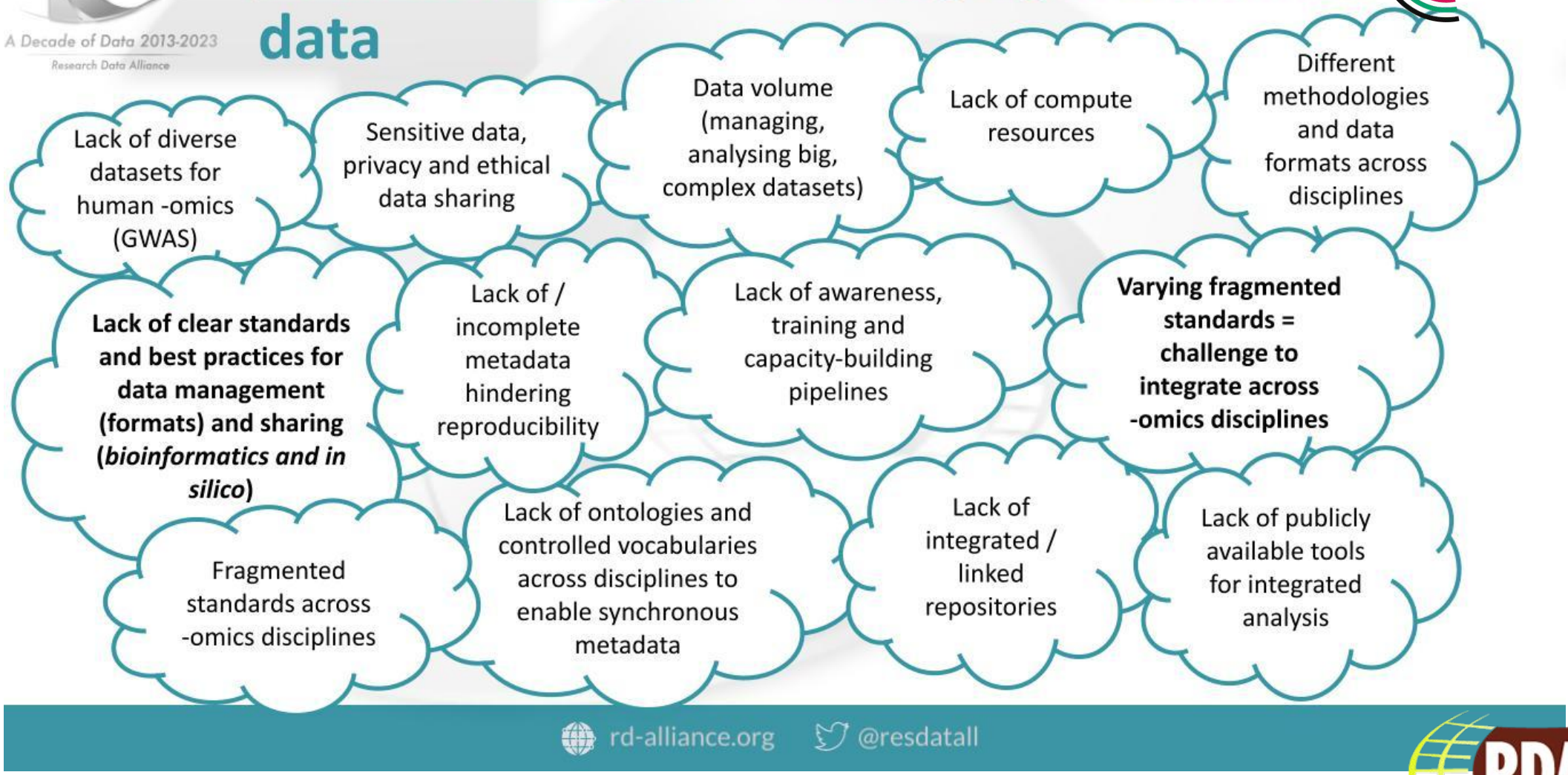


Figure 27 Types of services respondents would like to access (n=145)



Respondents reported 56 different repositories used

Your challenges of managing multiomics data



Squandering the value of data-intensive research

Fragmented and confusing landscape

Data methods and services lack connectivity
Lack of awareness of methods and services researchers could/should use
End to end data lifecycle support is hard

Lack of research data management (RDM) skills

Majority of researchers lack the basic skills to use existing tools and services
Shortage of highly skilled experts

Inconsistent usage

Uneven access data services and data methods, with supporting services
Mostly researcher DIY

Data often not reusable

Lack of skills, shared data poor quality non-reusable by other researchers

UK lacks a FAIR culture

Limited adoption of FAIR principles across UK life sciences researchers

This will only increase as the life sciences becomes increasingly data-intensive

Other countries are establishing national bioscience data capabilities

Strategic priority for UK

BioFAIR Objectives

1. Culture change: To drive adoption of FAIR principles and open data across the UK life sciences to provide high quality FAIR datasets ready for uptake and reuse.
2. Defragmentation: To increase the coordination, collaboration and cohesiveness of the UK's life science related data landscape, enhancing its effectiveness and efficiency.
3. Access: To enable democratised access to data and data methods within UK life sciences via a national capability.
4. Skills: To attract, develop and retain excellent research data management skills and capacity for UK life sciences.
5. Data reuse: To improve the efficiency of research and enabling new research by increased re-use of data in the life sciences.

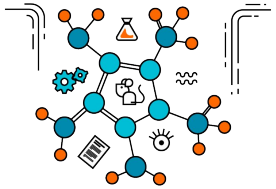
Scope: Users and Generators

The long tail

UK researchers and project teams

Collaborations between institutions and partners

Communities



Data Coordination Centres

Support their Data Hubs, archives and processing activities

Researchers

- Limited skills in research data management & analysis
- High level of skill in research data management & analysis

Research Supporters

- Institutional data managers
- Community data stewards
- Workflow developers
- Trainers

Tools and Service Providers

- Tools, data resources, analysis
- Improvements for the UK
- Data AI-readiness

Scope: Services and Data

Data tools & services

- Datasets
- Software code
- Analysis tools
- Workflows
- Standards for data and metadata

UK Project Data

- Humans,
- Animals, Plants, Fungi
- Microbes

Out of scope

- Sensitive data, e.g. patient data: Partner with Health Data Research UK
- Data beyond the life sciences: Partner with NERC Environmental Data Commons
- Industry R&D data infrastructure

Support the Data Journey

The first & last mile

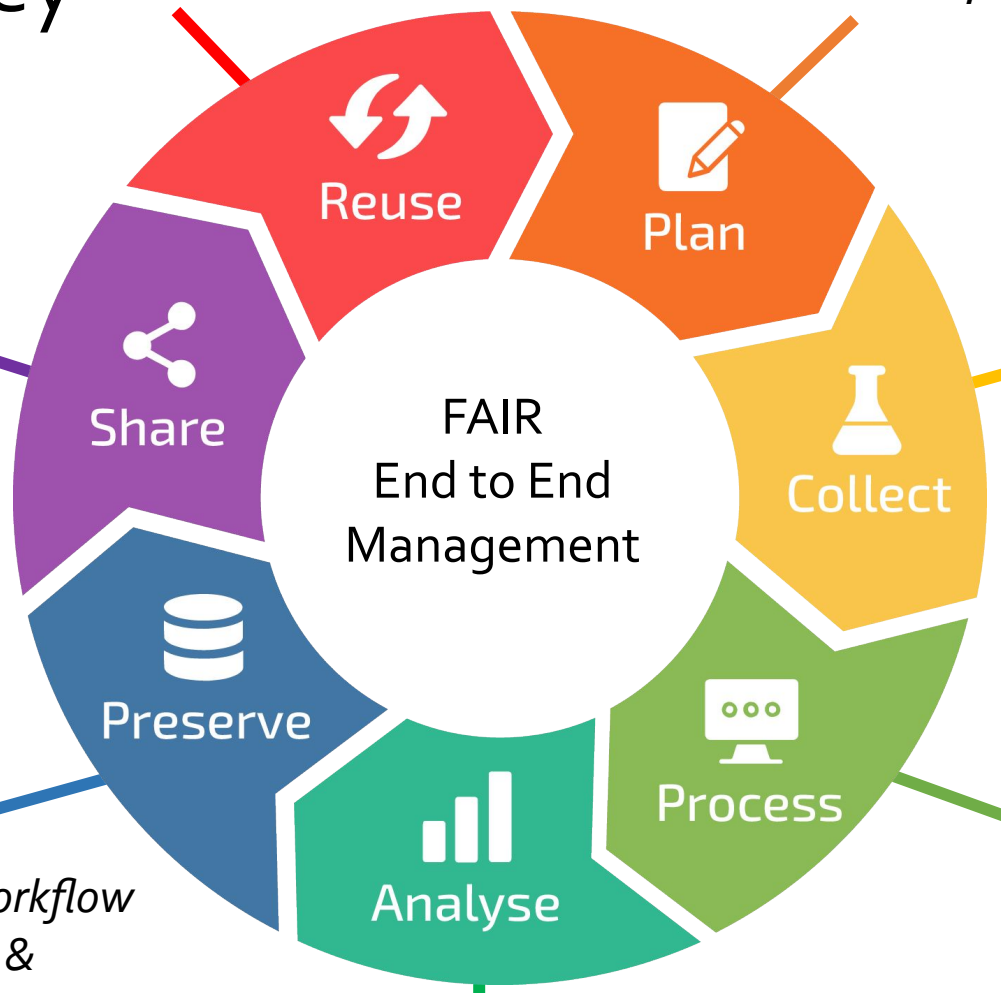
Data & Metadata Collection

Data Processing and Analysis

Data Deposition Archives

Data Processing and Analysis

FAIR Data Journey



Data and workflow publishing, credit

Project planning
DMP, Software planning

FAIR data collection
Data Standards
AI readiness

Data preparation pipelines
FAIR Data curation
Standardisation and harmonisation

Simple access to tools
Reproducible QA data analysis
ML best practice
Link with other National Hubs

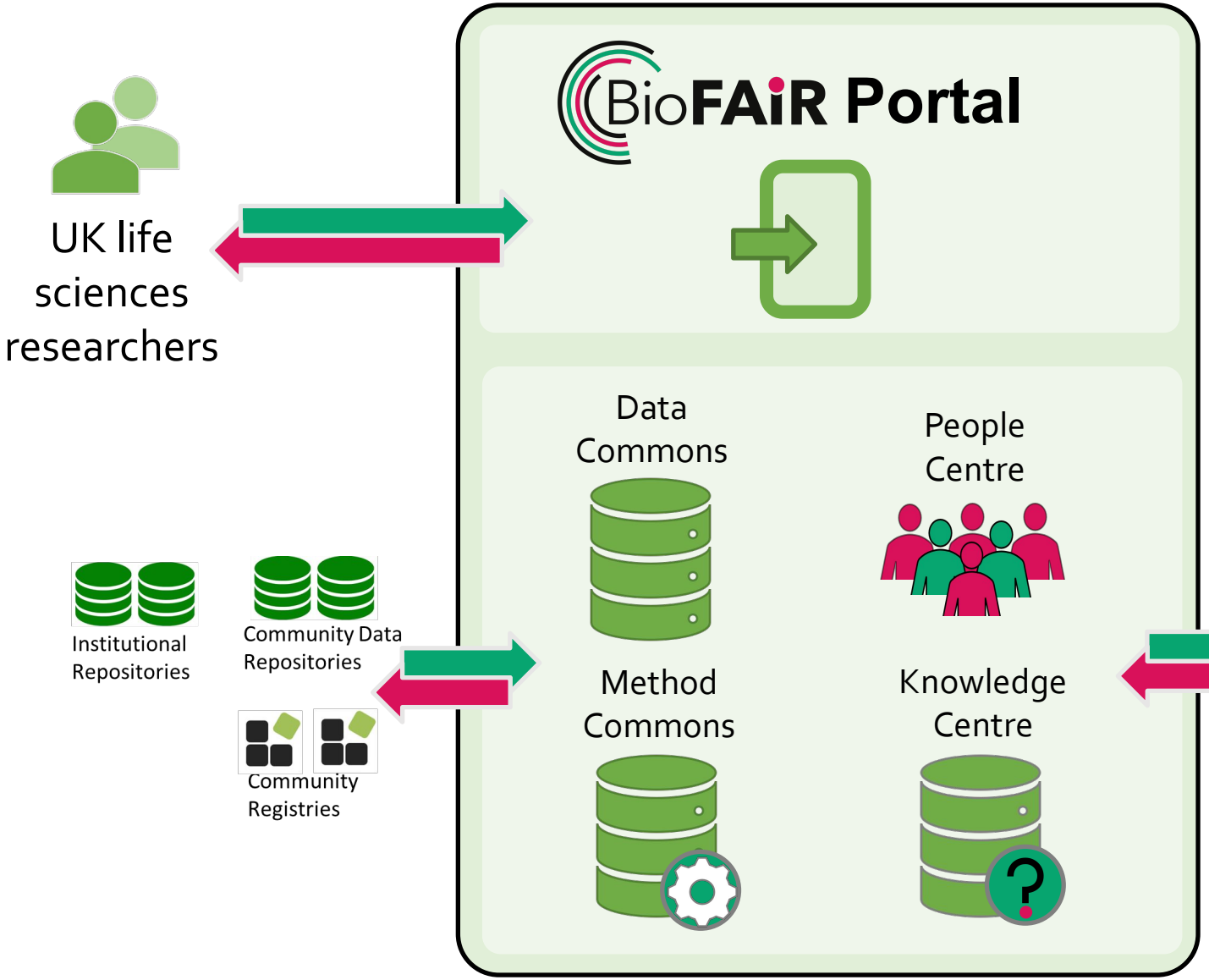
Analysis & workflow maintenance & reproducibility

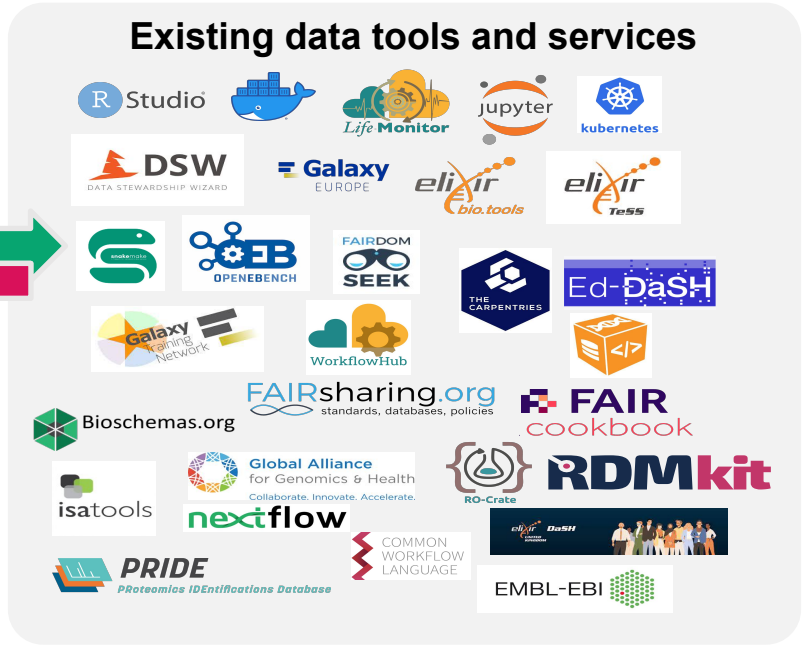
Post-project data retention

Workflow, SOP and data registration,
Collaboration spaces
Streamline data deposition into public archives
Links with other National Hubs

Joining up the Dots

BioFAIR will join the dots between DRI investments, act as a catalyst for integration between DRI data hubs, build a cooperative collective of analysis and FAIR capability and provide cohesive capability using cross cutting technologies and techniques.

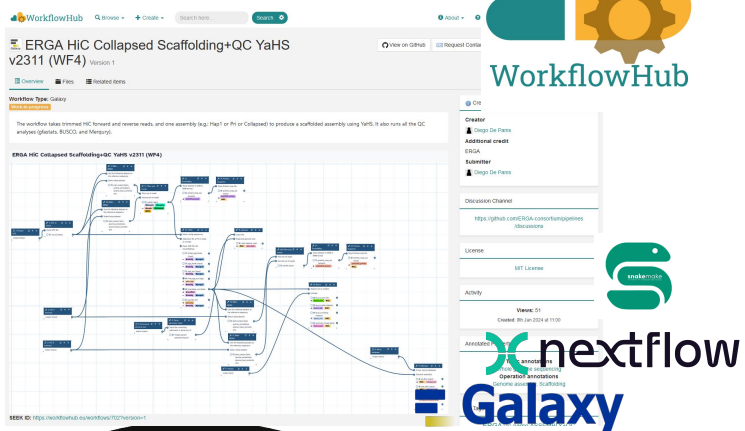




Using existing services and national delivery partners

Method Commons

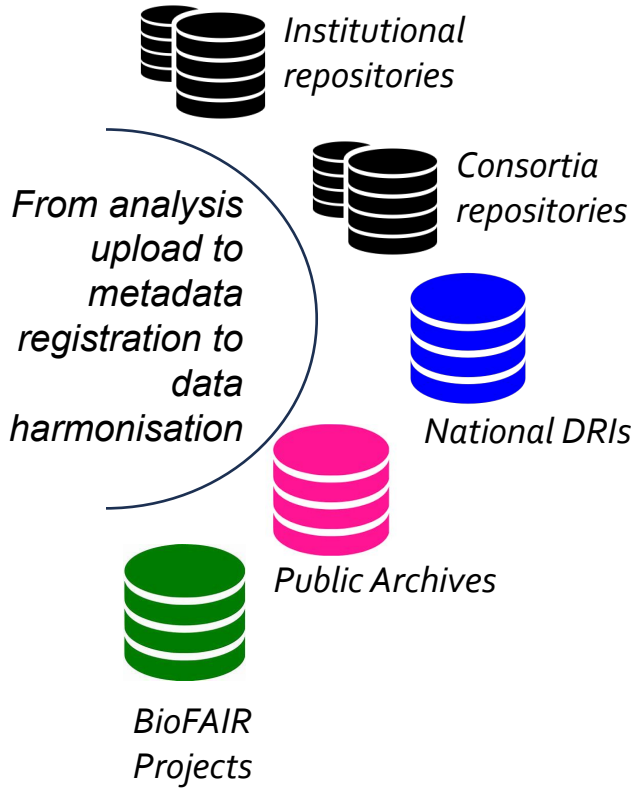
National workflow capability
Workflow development
National compute
Electronic Lab Notebooks
Workflow sharing – public & project enclaves
No-code analysis



Data Commons

Standards
Data Catalogue
Federated Data Lake
National data storage
Project data spaces
Analysis data ingress & egress
FAIR data services
FAIR Digital Objects

Data Management Planning
Islands of harmonised data
Submission pipelines to archives
Data brokering for projects
Post-project data retention
Data sharing – public & project enclaves



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

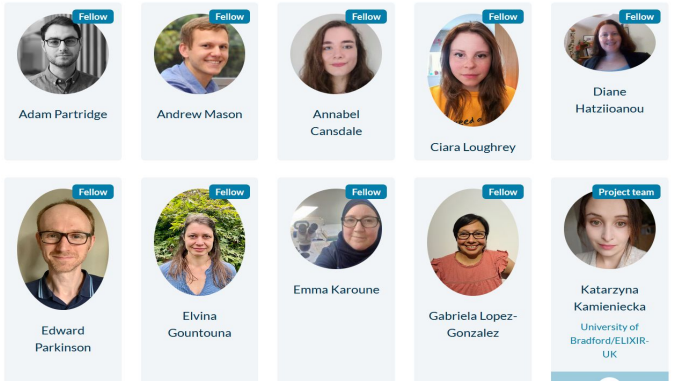
Using existing services and national delivery partners

Knowledge Centre

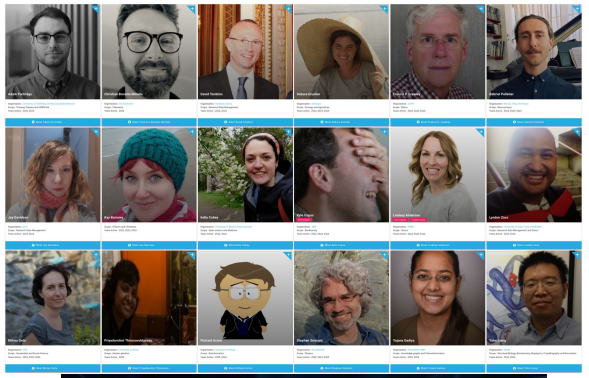
RDM Toolkits & Cookbooks, DM Plans
Standards Registries
Training Portals

People Centre

Train the Trainers
Dedicated Data Stewards
Dedicated Software and workflow developers
Community workers



elixir UNITED KINGDOM **Fellowship**



FAIRsharing Community Champions Programme

A data and analysis concierge service

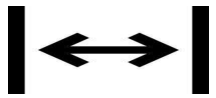
Aim to Accelerate Discovery

Maximise FAIR and Reproducible Data and Methods

Improve culture
Mobilise researchers
Reduce waste time

AI Readiness

Widen use of existing data methods and services

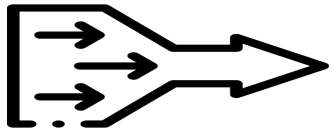


Optimise best practice
Exploit UK and global investments and skills

Support UK service providers

Benefit from international services

Ease and streamline collaboration



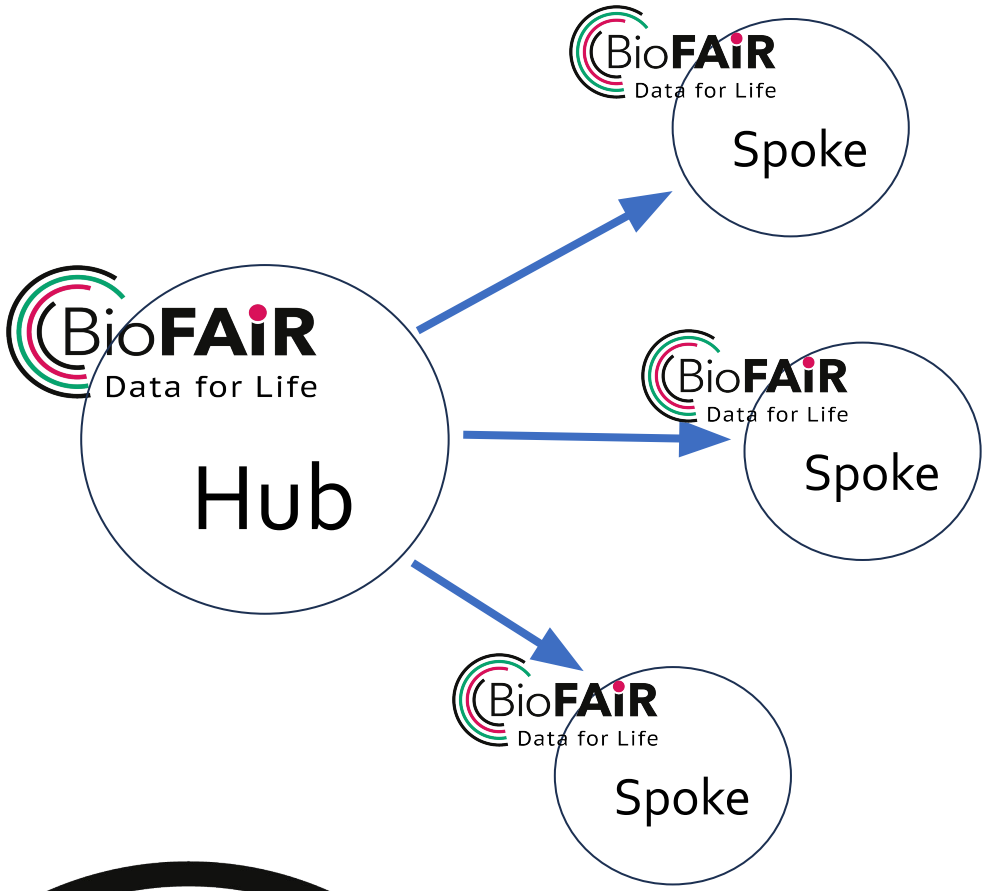
Provide an outside organisation data and methods brokering and data retention service

Bridge the gap

Between researchers, data sources at individual institutions, and existing data infrastructures

Increase efficiency of research

BioFAIR Delivery model



Delivery spokes operate the tools and services available through BioFAIR to form a coherent federated infrastructure

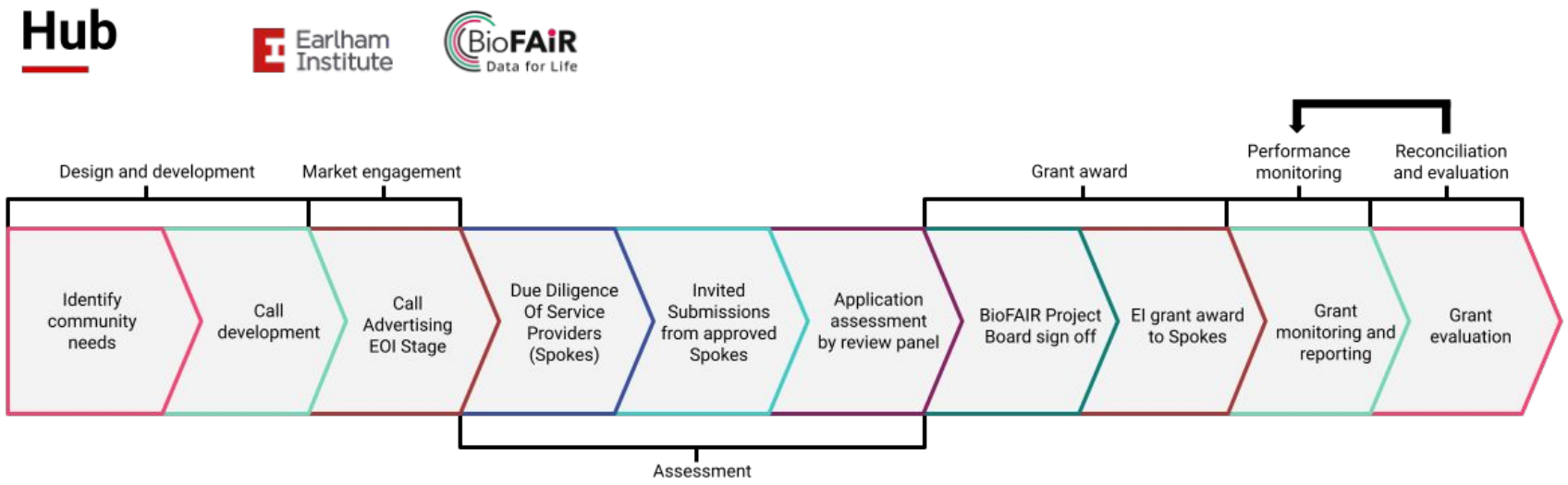
Complements existing UK Digital Research Infrastructures

Partnerships

Pathfinder communities

Spoke selection

- Delivery team and leadership will be distributed across Hub and Spokes
- Spokes will be selected as long term partners with responsibility for key components and services
- Eligibility based on track record of delivery and operational excellence
- Open and transparent selection process



BioFAIR Delivery Phases

Phase 0 Pre-launch	Phase 1 Launch *Month 1 – 8*	Phase 2 Set up *Month 9 – 24*	Phase 3 Establish *Month 25 – 48*	Phase 4 Expand *Month 49 – 60*
• Interim Director appointed • Business Case approval • Initial stakeholder engagement	• Recruit core staff • Establish governance • Finalise M&E plan • Continue stakeholder engagement	• Develop and deploy MVP infrastructure • First pathfinder projects • First Community of Practice • Establish outreach • Initial evaluation report	• Establish operational infrastructure • Delivery training programme • Expand CoPs • More pathfinder projects • Full Evaluation • Future investment strategy developed and agreed	• Maturation of operational infrastructure • Review coverage • Expand training and outreach • Future strategy published

Target for Launch - Summer 2024

Help us start to build BioFAIR

Workshop 1 Topics

- At the heart of BioFAIR is a commitment to the FAIR principles. We want to facilitate their wide spread adoption to enable better access and reuse of data, however we also understand that researchers want to be able to demonstrate their active adoption of FAIR to funders. What services should BioFAIR provide to enable both adoption and demonstration?

- We are early in defining the priorities for BioFAIR and its architecture. What services and capabilities would you like us to focus on delivering in the first year?

Workshop 2 Topics

- The BioFAIR project was conceived from the ELIXIR community and we wish to continue its deep commitment to be open, transparent and community-led. How would you like to engage with BioFAIR and support the community led design and delivery? What communications approaches would you like us to adopt?

- BioFAIR aims to serve researchers and their facility support staff, RSEs, bioinformaticians and data stewards in RPOs. How should BioFAIR and local staff and and data policies/platforms best interact?

https://biofair.uk/

UK Research and Innovation