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





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

UK Data Research Infrastructures Biosciences and related areas



CLIMB BIG DATA

Cloud Infrastructure for Big Data Microbial Bioinformatics













data commons

UK Data Research Infrastructures Biosciences and related areas

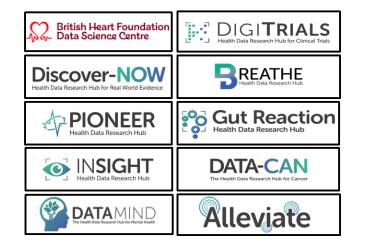


National institute for health data science. Large-scale data and advanced analytics

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



UK Health Data Research Alliance

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/Al
- 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-classresearch-and-innovation-infrastructure/digital-researchinfrastructure/

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

Landscaping exercise Next meeting 4 July 2023 (tomorrow!).



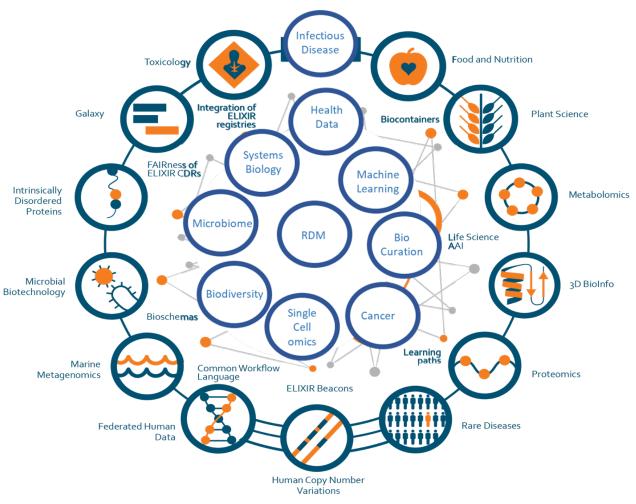
ELIXIR-UK, a node of ELIXIR the European Research Infrastructure for Life Science Data

Intergovernmental 23 nodes, 250+ organisations

A federated European digital infrastructure for Life Science Data,

Coordinate national and European capabilities. Sister to Euro-Biolmaging & 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



Collect Plan 🚹 Feide Tryggve Conceptualization Grant Application DSW training nor seq Direct upload from Core Facilities 000 </> Data Storage Process & Analyse Metadata Feide tracking Feide 🚹 Feide Data integration Sample & Data p Bioinformatician University of York **NeLS** Norwegian e-Infrastructure for Life Sciences DisGeNET . ISA Format 8 Flexible Storage Sample/Data provenance Long term storage JWS online for SBML models Analysis Pipelines = Galaxy 🏀 HMINT SILVA K STRING 😾 CHEBI Snapshots for publication Dr Elvina Gountouna Research Data Manager C A T H 🕞 BioSample 🥢 InterPro Oxford Brooke Preserve Knowledge & Share elixir European Nucleotide Archive Data A MetaboLight Rhea Feide Deposition Brokering THE HUMAN PROTEIN ATLAS* 88PDBe-KB orphadata Databases reactome eh Cellosaurus **Bio**Studies. European Variation Archive 👹 BRENDA **NORWAY** training **Bioimage metadata** - REMBI

National project

Trusted data resources

RDM Stewardship network, fellowships,







Dr Rob Ewin

Iniversity of



Image Data 11 videos

ELIXIR-UK, Research Data Knowledge

RDMkit

Epitranscriptome data

Toxicology data

Structural bioinfo

Rare disease data

our tasks

All tools and

All training resources





Introduction

Bioimaging specialists are acquiring an ever growing amount of data: images, associated metadata, etc. Ho management often does not receive the attention in requirise or is avoided altogether since its consistered a budgetmestine task. At the same time, storing images on personal computers or USE keys is no longet an option, assuming it ever was taba volume is exponentially increasing, and not just the acquired images need storing bud potentially processed images will be generated and will need to be keyst alongist the term optimal images. The storing bud potentially processed images will be generated and will need to be keyst alongist the term optimal images. The storing bud potentially processed images will be stored. For how long, who will cover the cost of the hardware, and who will cover the cost of managing the infrastructure. All the stakeholders

What constitutes bioimage data

requirements are understood and met.

An image is much more than a collection of zeros and ones. The image will contain the binary representing the pixels on scree but it is usually packed with useful metadata. You will find the obvious keys indicating how to interpret the zeros and ones, you can also find a lot of acquisition metadata a gh andware/insument used, settings used, etc. Managing images immediately becomes a larger problem, not only the binary files need to be handled, but also the associated metadata. Serveral efforts have been made and still ongoing to capture those metadata. Understanding and capturing the metadata are critical for many reasons, just to mention a few analysis, detection of possible faults in acquisition systems. It is important to detide how much details will be recorded since this could dramatically increase the metadata volume and therefore the effort required to capture

el

Te55

Events

d to be involved in the preliminary discussions; biologists, facility managers, data analysis, IT support, etc. to ensure that the

The Biolm

licoue

How can TeSS help you?

Browse the catalogue

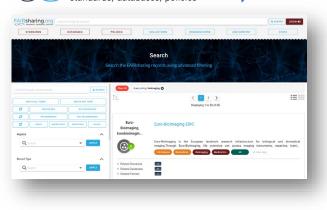
Providers

A Workflows

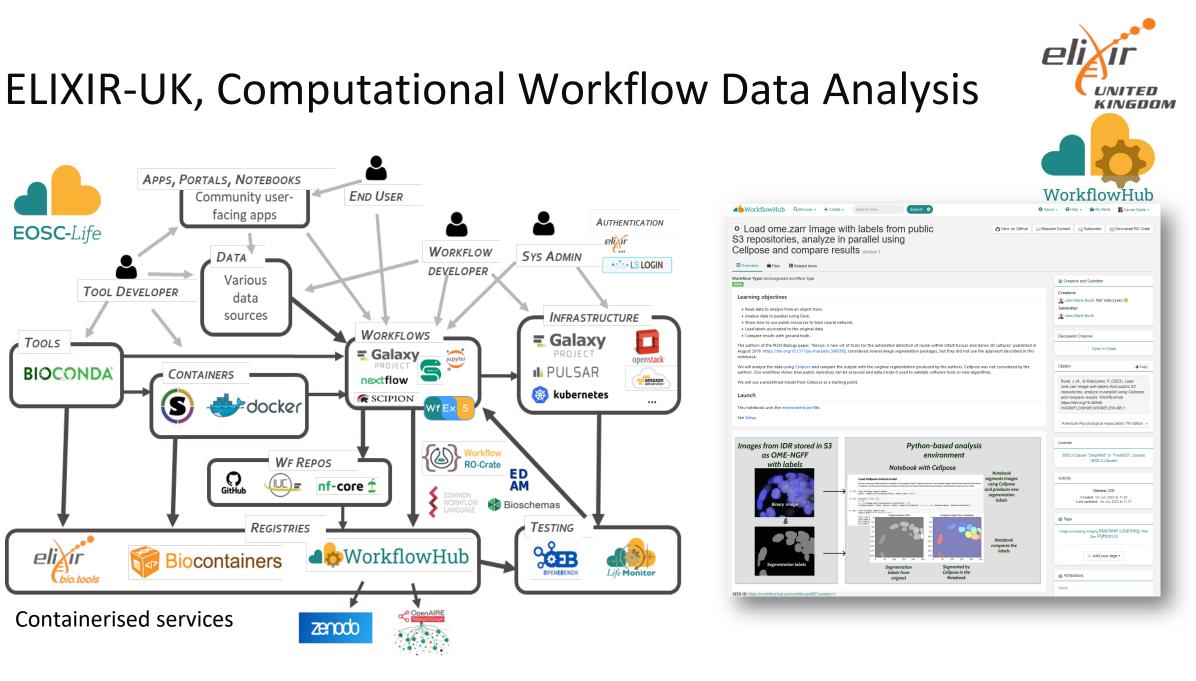
00000

Naterials

FAIR sharing.org



DSW Product v Solutions v Learn v About **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 🔂 FAIR elevant tools and resources cookbook Registry 🛛 Stendards/Detabases Date stonege 1 Tool info Standards/Databases 1 Teel info .1 Data 8.1 Data 1 Identifi apping 7.1 Data hosting access licensing Standards/Databases 1.2 Data 7.2 Data versioning 8.2 Data anonymisation Standards/Databases 6.3 Data mode 7.3 Data transfer 8.3 Data release .4 Managing i Tool info 1 Tool info Plant genomic and genetic variation 64 data submission to EMBL-EBI Reading Tim 30 minutes Guidance i Teelinto or managing, visualizing and analyzing Data Steward: research Difficulty Maturity Level & Indicator [F+MM-2.3R] [F+MM-3.3C] Cite me with FCB061 28231 tools OMERO diseastor ID Verified COPP Imaging > Data visualisation Desktop application Java Python 👌 🖬 🖷 Image analysis -KINGDOM

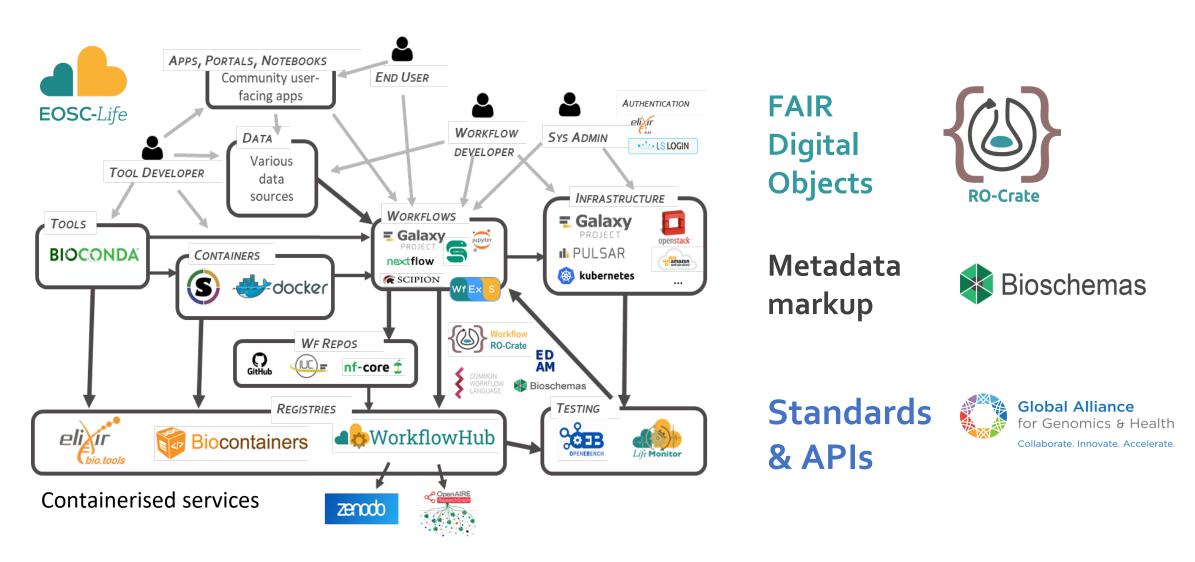


Tools

elixir



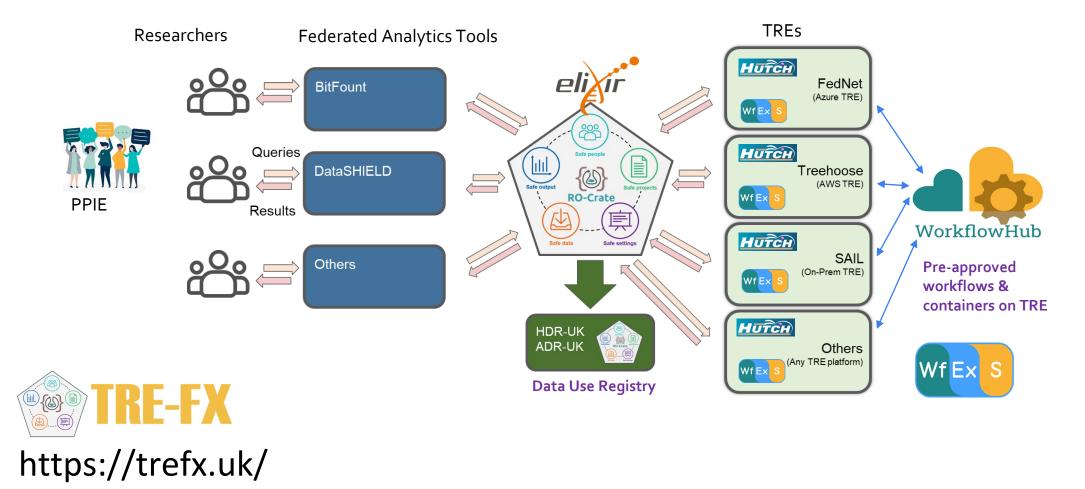
ELIXIR-UK, Computational Workflow Data Analysis



Data Research Infrastructure **Collaboration**



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





KINGDO



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

19 June 2023



News Blog Events Voices Horizon Europe association delay Investing across the UK Climate change 101 jobs that change the world COVID-19

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.

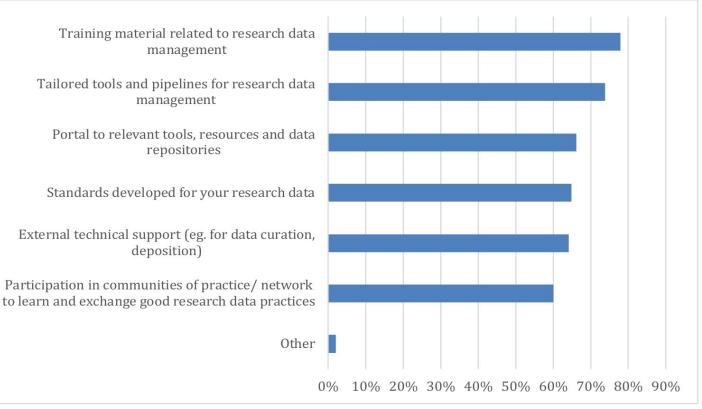
Funding of £34 million over five years is subject to business case approvals.





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)

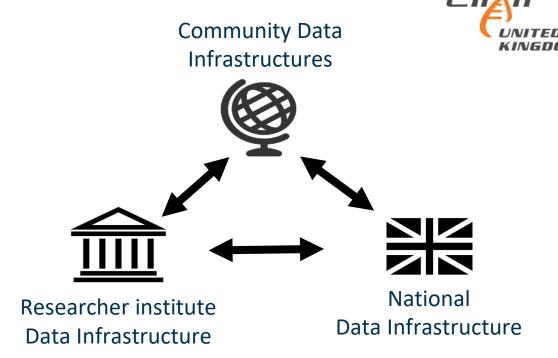


BioFAIR Feasibility Study (2020-2021)



End-to-end project data management and analysis Coherent ecosystem of tools and services

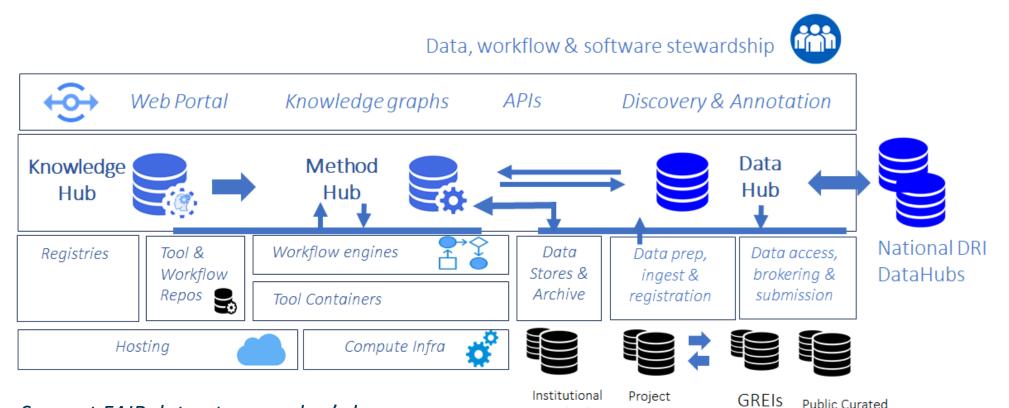
- Method Hub National data analysis facility using workflows and notebooks over dedicated & national compute.
- Data Hub Data standards, cataloguing, support over federated data lake, linked to other data hubs.
- Data brokering deposition archives, data store & retention for projects, QA, AI readiness.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



repositories

repositories

Data Resources

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/