



WHEN METRIC TOOLS AND FORMS SEEM TO BE BORING, VIDEO GAMES COLLECT COGNITIVE DATA

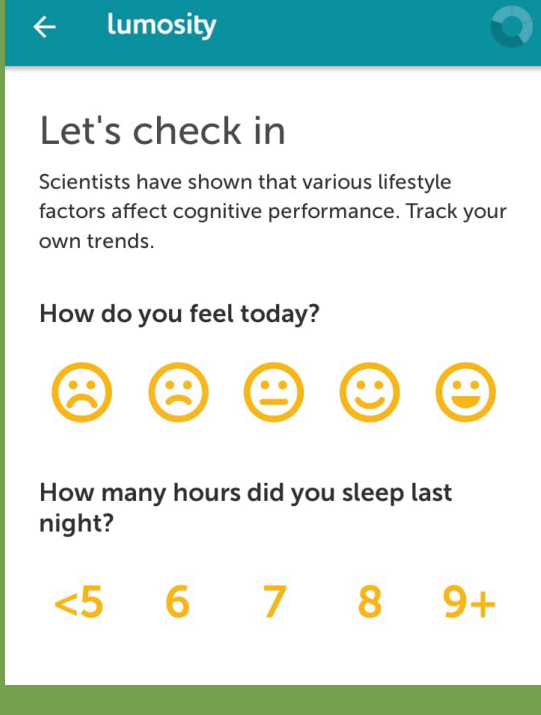
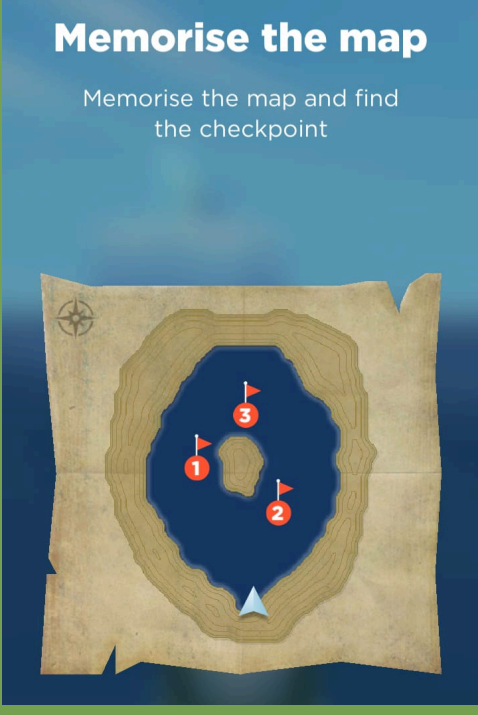
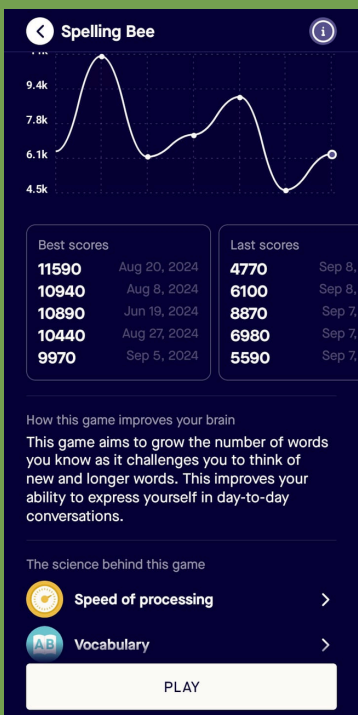
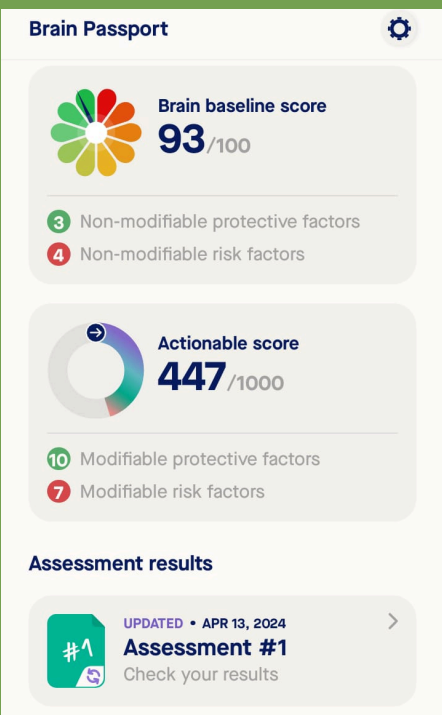
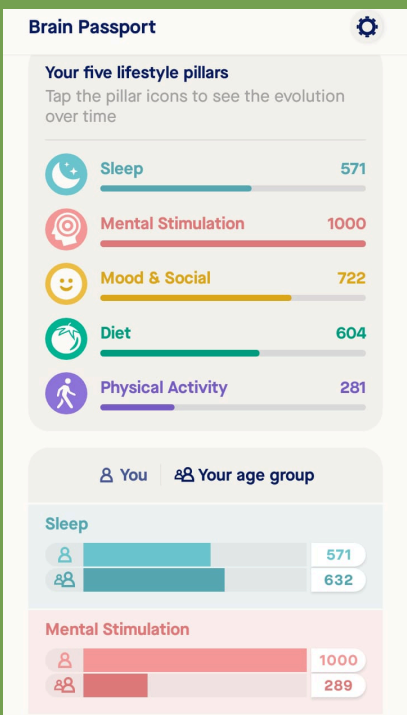
THAREQ BARASABHA^{1,2} & IVAN KOYCHEV¹

¹ PSYCHIATRY DEPARTMENT, UNIVERSITY OF OXFORD
² FACULTY OF MEDICINE, UNIVERSITAS BRAWIJAYA

INTRODUCTION

- Video games
 - Powerful tools for entertainment also for cognitive and physical rehabilitation (1)
 - Immersive experiences that align with scientific standards while maintaining high levels of player engagement (2)
 - Enable continuous and passive collection of cognitive data (2)
- Challenges:
 - ensuring sustained user engagement
 - validating the accuracy and reliability of the data collected
- Conventional metric tools and forms:
 - low participant engagement
 - potential bias due to their repetitive and unengaging nature

METHOD



- Three mobile-app games compared:
 - Five Lives
 - Sea Hero Quest
 - Lumosity
- The data collected classified into:
 - demographic data
 - behavioural data
 - cognitive data

Five Lives
(>80 K users)

Sea Hero Quest
(>4.5 M users)

Lumosity
(>100 M users)

RESULT

Game	Data Type	Data Collected
Five Lives	Demographic	Age, gender, education level, work, home environment, height, weight, family with dementia
	Behavioural	Sleep, mental stimulation, mood & social, diet, physical activity, etc.
	Cognitive	Language, attention, memory, agility
Sea Hero Quest	Demographic	Age, education, gender, home environment, nation
	Behavioural	Handedness, navigating skills, sleep duration, traveling time
	Cognitive	Spatial (wayfinding, path integration, radial navigation)
Lumosity	Demographic	Gender, age, education level
	Behavioural	Daily mood and sleep duration
	Cognitive	Speed, memory, attention, flexibility, problem-solving, math, language

REFERENCES

1. de Souza Miguel, G.F., de Sá, A.A.R., Tannús, J. et al. Proposal of a game streaming based framework for a telerehabilitation system. Multimed Tools Appl 83, 33333-33350 (2024). <https://doi.org/10.1007/s11042-023-16741-8>
2. Laura, Levy., Amy, J., Lambeth., Rob, Solomon., Maribeth, Gandy. Method in the madness: the design of games as valid and reliable scientific tools. (2018).9-. doi: 10.1145/3235765.3235793
3. <https://seaheroquest.alzheimersresearchuk.org/wiki/#game-play-data>

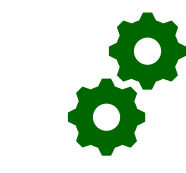
CONCLUSION: Video games can be used to collect demography, cognitive components, and protective / maladaptive daily habits of the users, as an option to the conventional metric tools and forms.

Supporting FAIR: Experiences in Improving the Scores on the Doors

Catherine Jones, Peter Holt, and Oliver Brough (Technology Department) Karen VanHaltren and Teagan Zoldoske (Scientific Computing Department) UKRI/STFC

The UK Energy Research Centre (www.ukerc.ac.uk) undertakes “Independent whole systems research for a sustainable energy future” and was formed in 2004. The Energy Data Centre (EDC) is a capability of UKERC and provides a discovery portal and data management expertise for researchers funded through UKERC.

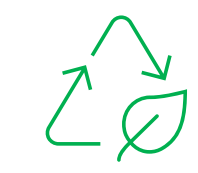
Current key priorities for the EDC are:



Continuing service development and maintenance



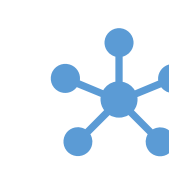
Enabling FAIR data



Focusing on Environmental sustainability



Preserving energy knowledge through building a grey literature collection



Contributing to UKRI Digital Research Infrastructure

FAIR Data: Data that is Findable, Accessible, Interoperable and Reuseable

EDC Policies Review

We have reviewed our existing policies over the years to continuously improve the EDC’s functionality as a FAIR facing repository. This has come under 2 approaches, both resulting in action plans:

Rapid Assessment Model (RAM) Approach: Identify current status and aspirations across a set of topics needed for a service to effectively **preserve content**

FAIR Approach: Starting from the principles, assess whether the EDC meets them and what is needed to improve

By using RAM in 2020/21 to convert implicit policy to explicit policy, and revising our existing policy in 2024 to include clearer links to FAIR, we have been able to explore the similarities and differences between FAIR and Forever policies and highlighted the key areas for the EDC to focus on to meet these goals.

Find out more about RAM at: <https://www.dpconline.org/digipres/implement-digipres/dpc-ram>

FAIR Assessment Tools Review

FAIR Assessment Tools generate machine-oriented tests based on FAIR Metrics. These tests result in a quantifiable “FAIRness” score for a piece of data.

This score can not only improve trust in the use of data we hold by showing how Reusable and Interoperable it is but also shows the EDC’s due-diligence in supporting FAIR data.

Using these tools on samples of data within the EDC showed that the FAIRsFAIR tool F-UJI was the most relevant due to its domain-agnostic nature and easily retrievable results.

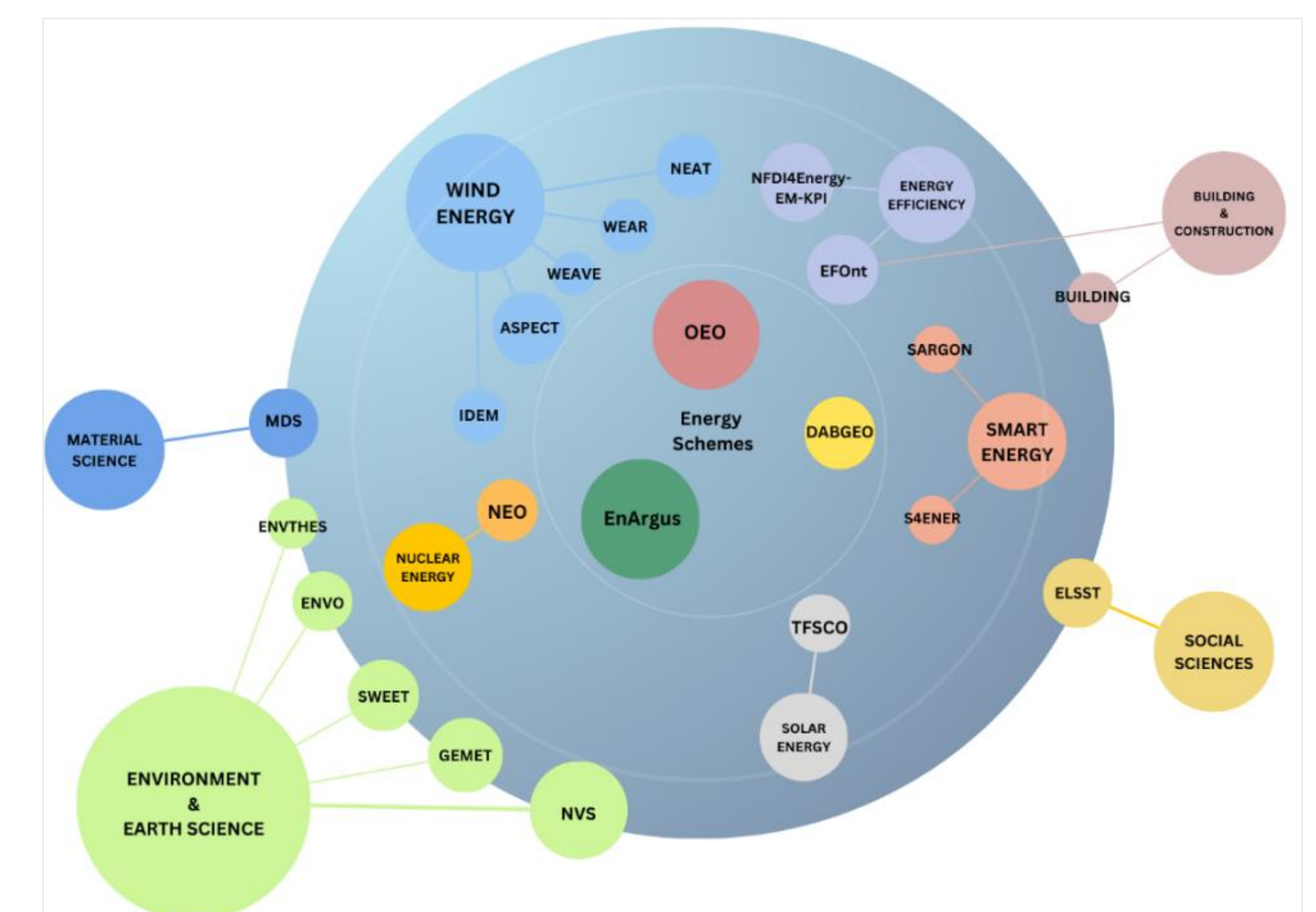
Discussions on how a data’s FAIRness score could be displayed to the user have involved showing averages across each letter of FAIR, as well as visualisations of these scores.

Energy Ontology Landscape Review

The FAIR principles recommend that for interoperability FAIR vocabularies should be used, with an emphasis on domain standards. This work reviewed the landscape of energy semantic artefacts.

The **Open Energy Ontology (OEO)** was identified as the largest energy focused ontology

The EDC uses energy categories to add subjects which based on the **IEA** scheme from 2004.



Use of energy categories means that the EDC has consistently classified records but issues with terminology changes and no ability to traverse relationships. These will need to be addressed to further refine FAIR practice.

FAIR-Enabled API

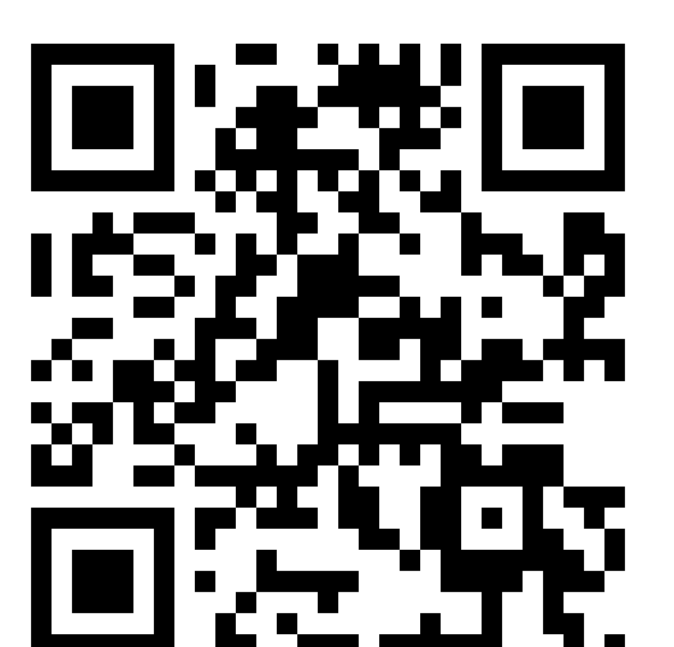
The Digital Infrastructure for National Infrastructure (DINI) Project gave us an opportunity to specify an API to enable machine access which has FAIR considerations at the heart of the design.

To demonstrate this API is FAIR, we will:

- Use the Open API standard
- Include clear Terms of Service and licenses
- Have good documentation for the API

To demonstrate the repository is Trustworthy, we will:

- Be certified – On the EDC Service development roadmap
- Enable metadata about the repository itself to be interrogated so that the service using the API based on RDA recommendations



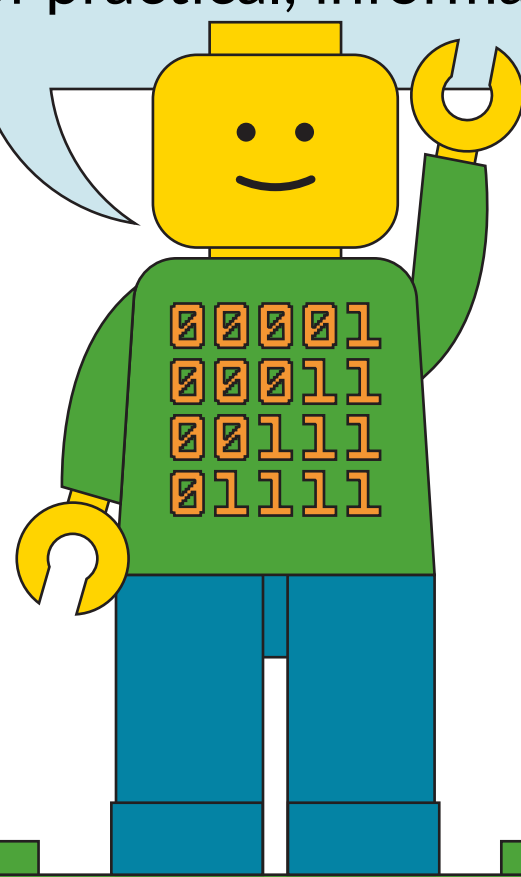
The UK Centre for Ecology & Hydrology (UKCEH) is a world-leading independent environmental research institute, producing data from a wide range of scientific disciplines.

Over 20 years ago, we recognised the need to improve data management practices and created the Informatics Liaison Network: staff connected through discipline-specific data expertise, distributed across all our sites.

The network developed into the Data Stewardship Team by improving ways of working, introducing tools, and expanding services.

However, the core of our work is still very much liaison with researchers at all stages of the data lifecycle, with an overarching aim to make research data and other digital objects increasingly FAIR.

And LEGO® bricks? Well, it turns out researchers love a bit of practical, informative play in their training!



1994

Four research institutions representing different scientific domains merged to form the Centre for Ecology and Hydrology (CEH):

- Institute of Hydrology,
- Institute of Terrestrial Ecology,
- Institute of Freshwater Ecology,
- Institute of Virology.

It was recognised that the research and data from these different domains could be brought together to answer bigger, more complex questions. This led to the establishment of the Environmental Informatics programme.

2006

A network of Informatics Liaison Officers (ILOs) was formed, primarily focussed on the management, sharing and reuse of data. The ILOs were based within different science areas, reflecting their expertise. This was the foundation of the Data Stewardship Team.

2007

The first Data Management Plan (DMP) template was designed to provide a framework for data management. The template went through various amendments to improve user engagement.

2010

The CEH Information Gateway was launched. For the first time, researchers could find and download datasets held by CEH. The ILOs were instrumental in helping to populate the catalogue and advertise it within the organisation.

2011

The ILO competency framework was launched as a way to improve career development and progression. This was based on a set of themes and skills of which all ILOs had a core knowledge, but with the opportunity to develop advanced skills and become theme champions.

The ILO wiki was launched as an internal resource on all things data management.

2014

An InfoPath template was developed, transforming DMPs from isolated project documents to online, machine-readable versions.

2016

The Environmental Information Data Centre (EIDC) catalogue was launched, making data discoverable, accessible and reusable for all NERC-funded research. ILOs have a joint role with the EIDC, liaising with internal and external researchers to get their data published.

2019

UKCEH was formed when CEH became an independent research organisation.

2022

The Informatics Liaison Network was rebranded as the Data Stewardship Team. The Research Data Management (RDM) hub was launched as an extension and modernisation of our ILO wiki, and is available to all at UKCEH.

A customised version of the Data Stewardship Wizard (DSW) was developed: a smart questionnaire with machine-readable DMP output and targeted guidance linked to our RDM hub. It has an intuitive interface, instant feedback and metrics, and easy-to-follow progress. The emphasis is on data management activities during the project and working towards reusable datasets, with guidance on what is required to make them FAIR.

2024

Our training programme for researchers was developed and launched, focussing on short, interactive training courses: Research Data Management principles, FAIR data, how to use the DSW, and data publishing. To maintain engagement, our training courses are interactive and use a variety of tools, including LEGO® bricks, to inform and entertain.*

2025

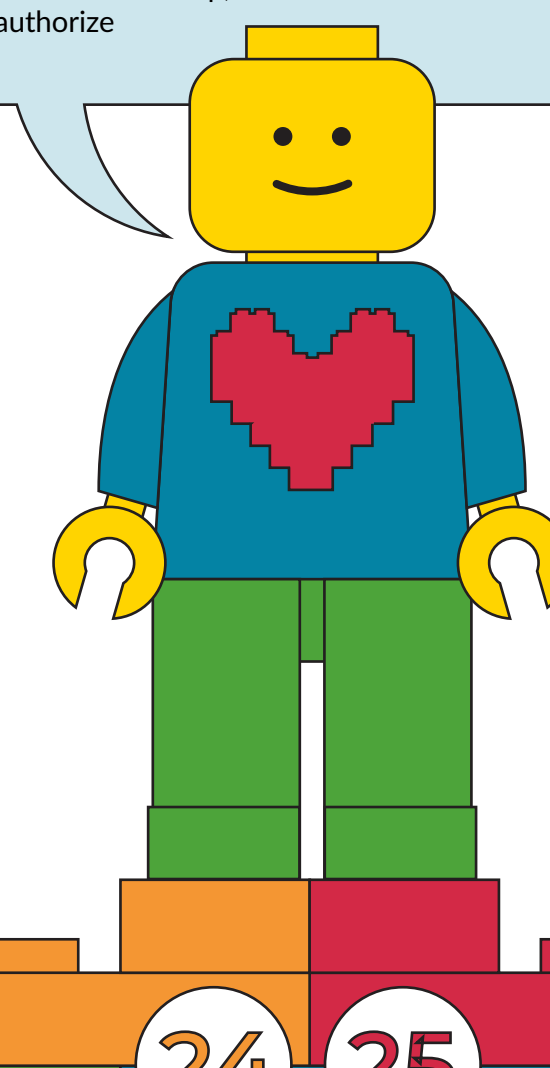
Our recent developments include a reporting tool that aggregates DMP responses across the organisation, providing high-level insights that can be used to highlight data trends, inform resource planning, and reveal opportunities that may otherwise have been overlooked.

*The idea of LEGO® bricks as a practical teaching aid was inspired by the quote:

“Data without metadata is like a LEGO set without the instructions”

Acevedo, 2023, Nature, <https://doi.org/10.1038/d41586-023-01929-7>

LEGO® is a trademark of the LEGO® Group, which does not sponsor, authorize or endorse this poster.



Methods as data: Investigating how reporting in methods sections describing mouse models of breast or prostate cancer affects perceived replicability of models

Anna Korzeniowska^a, Louise Saul^{b, c}

a. eLife Publishing Ltd, Cambridge CB2 1AW ; b. University of Southampton, Southampton, SO17 1BJ; c. Careers and Skills for Data-driven Research (CaSDaR)

Abstract

The potential for replicability and reproducibility of an experiment is vital for trustworthy research and to verify claims made when publishing the research. However, there are increasing reports of cancer study designs not being reproducible¹. Detailed contents of the methods sections of manuscripts are essential for replicability. This is a moral imperative when considering experimental animal research undertaken in accordance with ASPA and for compliance with the 3Rs initiative (Reduce, Refine, Replace), as researchers should show that they are undertaking animal experimentation with an appropriate level of care and consideration towards the experimental animals.

However, there is no established method for assessing the replicability of animal models during peer review², and whilst there exist papers that describe how to embed practices to support making models replicable at the point of experimental design³, we propose that promotion of extensive reporting in methods sections may be a way of promoting perceived replicability of animal models to ensure trust in research , with a focus on mouse models of breast and prostate cancer. To support method sharing, we plan to generate a checklist by searching available literature and consulting with researchers to determine the factors that should be included in methods sections to ensure replicability, then determine the representation of those factors in mouse models of breast and prostate cancer described in manuscripts published to determine the current level of perceived replicability in these manuscripts.

Scoping Review

To identify which areas to include in the checklist to promote perceived replicability, the following was used:

- 1. Resources from institutions known for support of animal welfare in experimentation (NC3Rs⁴, CAMARADES³)
- 2. A review of manuscripts published in the last 5 years that published immunocompetent mouse models of breast or prostate models.

Manuscripts with the following breast or prostate models were excluded from the scoping review: experimental models published more than 5 years ago, models with transgenic mouse models (athymic mouse models were included), and only the methods sections were examined. Information was not collected from text elsewhere in the manuscript.

The above was used to generate a list of 43 factors that we propose are relevant to the replicability of mouse models used in breast and/or prostate cancer models that were divided into 3 different areas: Husbandry, experimental, and procedural.

As well as recording the factors that we proposed were important for perceived replicability, we collected information on the geographical location of the lead author to ensure that we were representing a diverse sample.

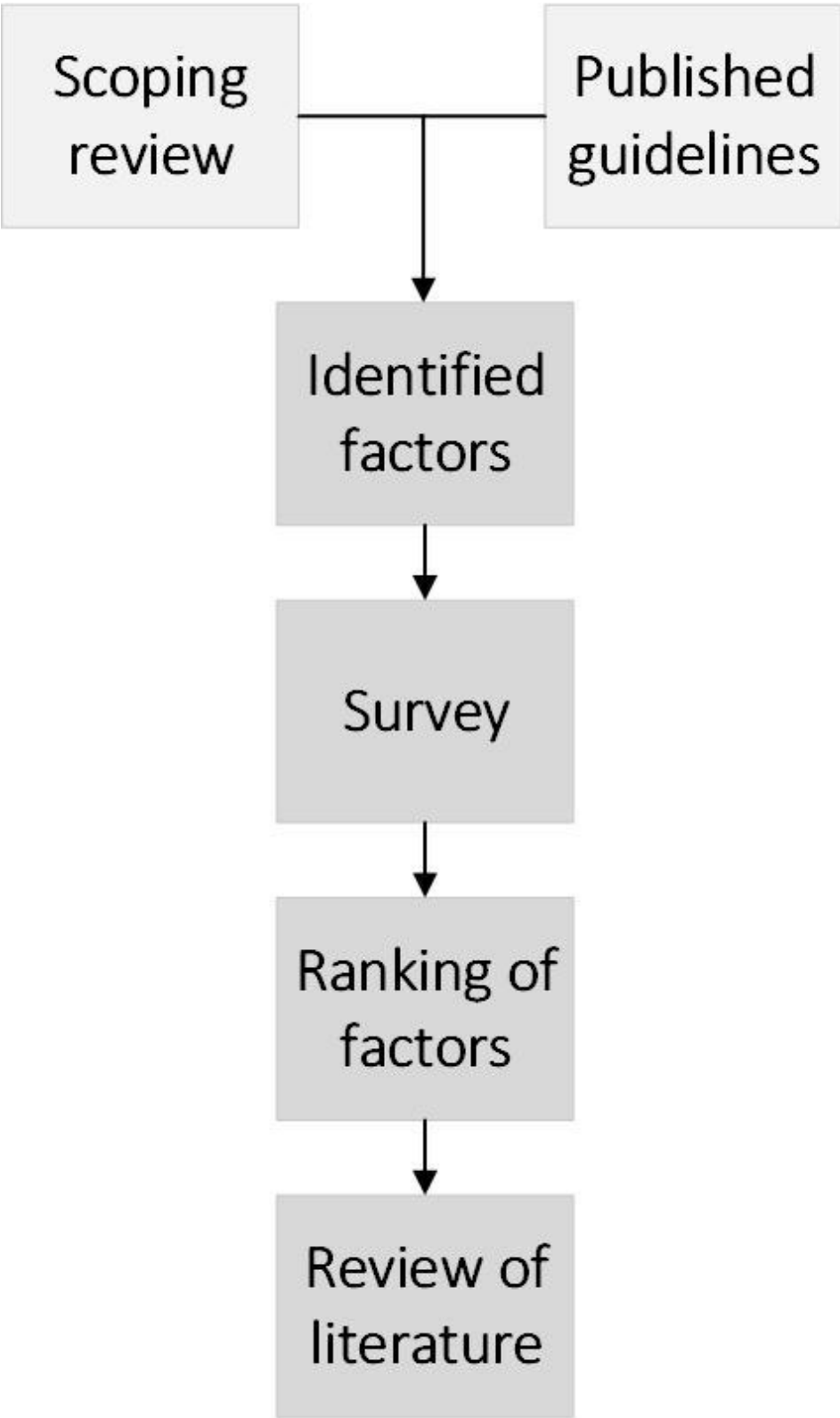


Figure 1. Image showing workflow for the project

Survey Design

We used the factors identified to create a survey for researchers currently using experimental animal models for oncology research, where we asked them to rate on a scale of 1-5 (1 meaning not important and 5 meaning essential) the importance of various factors that we have identified from a literature review as being relevant to the replicability of breast/prostate cancer solid tumour models when included in the methods sections of manuscripts.

We recruited the participants as researchers who are either involved with the NC3Rs working group with a focus on oncology research or know someone who is attached to this group who has forwarded on the survey to them. The participants will be asked for additional information about their career stage, involvement in replicability studies, their geographic location, and their institution type.

The ratings of the factors will support us to create a list of factors important for replicability; based upon the ratings from the survey, the factors will be ranked, and a weighting will be given to each factor. We are currently awaiting the results of the survey.

Future Work

Having identified the factors relevant to perceived replicability of mouse models, ranked them, and applied a weighting factor to each of them, we plan to query the PubMed database to identify the whether manuscripts describing models of breast or prostate cancer published in the last 5 years could be perceived to be replicable and therefore trustworthy by researchers, based upon the survey results.

There are two options to do this work:

Method	Description	Barriers
E-utilities	Entrez programming utilities ⁵ , (E-utilities) which allow search and retrieval of information from the PubMed database with the use of a public API to retrieve numbers of manuscripts from the database which use specific keywords identified from the survey.	The data in the PubMed database is unstructured, meaning that it is difficult to determine if the entries returned have the keywords present in the methods or elsewhere in the text. This presents an issue as we are focusing solely on the methods sections.
Manual search	A systematic review of the literature by creating an exclusion criteria of specific literature, searching the PubMed database, and manually collecting data from the methods sections of the returned manuscripts.	This will need to be completed by a named researcher; therefore position statements will need to be completed, and additional testing will need to be implemented to ensure impartiality in scoring.

The selected method will largely depend on the survey results and the rankings allocated.

1. Errington, T. M., Mathur, M., Soderberg, C. K., Denis, A., Perfito, N., Iorns, E., & Nosek, B. A. (2021). Investigating the replicability of preclinical cancer biology. *eLife*, 10, e71601. <https://doi.org/10.7554/eLife.71601>

2. Landi, M., Everitt, J., & Berridge, B. (2021). Bioethical, Reproducibility, and Translational Challenges of Animal Models. *ILAR Journal*, 62(1–2), 60–65. <https://doi.org/10.1093/ilar/ila027>

3. Wilson, E., Ramage, F. J., Wever, K. E., Sena, E. S., Macleod, M. R., & Currie, G. L. (2023). Designing, conducting, and reporting reproducible animal experiments. *Journal of Endocrinology*, 258(1), e220330. <https://doi.org/10.1530/JOE-22-0330>

4. Percie Du Sert, N., Ahluwalia, A., Alam, S., Avey, M. T., Baker, M., Browne, W. J., Clark, A., Cuthill, I. C., Dirnagl, U., Emerson, M., Garner, P., Holgate, S. T., Howells, D. W., Hurst, V., Karp, N. A., Lazic, S. E., Lidster, K., MacCallum, C. J., Macleod, M., ... Würbel, H. (2020). Reporting animal research: Explanation and elaboration for the ARRIVE guidelines 2.0. *PLOS Biology*, 18(7), e3000411. <https://doi.org/10.1371/journal.pbio.3000411>

5. Sayers E. E-utilities Quick Start. 2008 Dec 12 [Updated 2018 Oct 24].

Data management planning at UKCEH using the Data Stewardship Wizard

Nichols, M., Ferguson, S., Stuart, R., Mobbs, D., Leaver, D., Dean, H., and Zwagerman, T.
UK Centre for Ecology & Hydrology

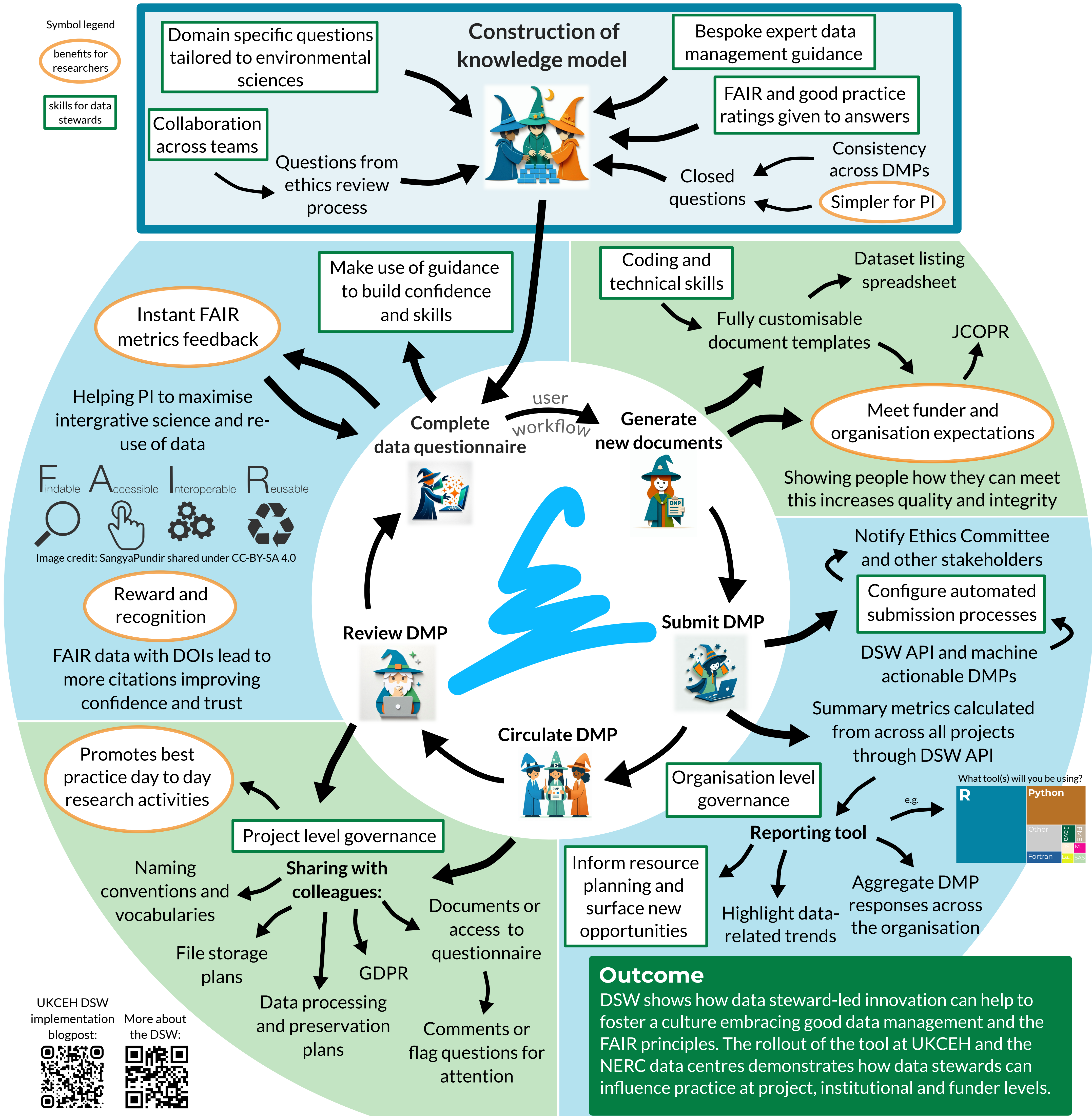
Contact: datastewards@ceh.ac.uk

Introduction

Data Stewardship Wizard (DSW) is a tool with a fundamentally different approach to data management, designed to help deliver FAIR data in practice, and demonstrates how data stewards can **embed good practice directly into the research process**. The tool guides researchers through a series of relevant, multi-choice questions in a smart questionnaire, providing expert guidance and instant FAIR metrics feedback based on the answers. It automatically generates a range of bespoke data management plan documents, **reducing the burden** on researchers.

Context

The role of data stewards is becoming more vital as environmental scientists tackle increasingly complex and interdisciplinary challenges. Researchers are facing **growing expectations** around transparency, efficiency and compliance in the age of open science. **Data stewards play a key role** in meeting these needs, through bridging technical expertise, data management knowledge and policy requirements. DSW enables data stewards to disseminate these skills on a wider scale.



Physical Chemistry Properties Data Collection

PChProp

Matthew Partridge, Samantha Pearman-Kanza, and Jeremy Frey
PSDI, School of Chemistry and Chemical Engineering, University of Southampton, SO17 1BJ

Context & Challenge

- There are multiple large **disparate datasets** in physical chemistry
- Many contain errors and require careful **data auditing**
- Disparate datasets make it hard to use for multi property modelling
- Data is being lost or made **inaccessible**

Impact

- The data ingest auditing ensures all the data included is of high quality
- Opens up the data for modelling and electronic accessibility
- Ensures important datasets are easily available to all across the Physical Chemistry domain
- AI and ML ready data collections



Physical Chemistry Properties Data Collection Version 1

110,804

Unique compounds

each described with **23** metadata fields

from **9** data sources

describing **5** different properties

8,811
melting points

6,544
boiling points

102,927
aqueous
solubility values

1,224
HL constants

236
miscibility values

Future

Physical Chemistry Properties Data Collection Version 2 (in development)

400,000
Unique compounds

from **17** data sources

describing **6** different properties

200,000
melting points

16,000
boiling points

240,000
aqueous
solubility values

1,224
HL constants

10,000
miscibility values

450
CMC Values
236
SHG values

- New datasets have been provided to us
- This will require careful integration and ongoing auditing
- Will need data stewardship to ensure continued quality

The Physical Sciences Data Infrastructure is funded through EPSRC Digital Research Infrastructure Funding – Grants EP/X032701/1, EP/X032663/1 and EP/W032252/1.

Acknowledgements: We would like to acknowledge and thank all the people involved in the PSDI Statement of Need, Pilot, Phase 1 and Phase 2.

Find us on social media – just search “PSDI”



Explore more on our website:
www.psd.ac.uk

ELIXIR-UK

Discover the building blocks behind a decade in data management and stewardship for the life sciences and beyond

ELIXIR-UK is the national Node of ELIXIR – the European life science research infrastructure. Since ELIXIR-UK was formed, it has grown its work and expertise in research data management, coordinating a distributed network of 30 UK organisations and communities across the UK.



Biotechnology and
Biological Sciences
Research Council

ELIXIR-UK is supported by the Biotechnology and Biological Sciences Research Council – Current grant: BB/C001384/1.



Interested in what we do?

Come talk to us!



Robert Andrews
RDM Lead
Cardiff University



Xenia Perez Sitja
Deputy Node Coordinator
Earlham Institute

2025: Expanding Data Stewardship guidance

Co-leading the development of the Data Stewardship Handbook, with the vision of extending the handbook beyond life sciences to other disciplines.

2024: Co-Leading ELIXIR-RDM Community

Co-leading the official RDM Community of ELIXIR at European level



2020: Leadership in Resource Development

Co-led the development of the RDMkit and FAIR Cookbook – 2 Horizon Europe recommended resources supporting FAIR best practices for life sciences and health data.



2025: Published UK specific DMP templates and examples



ELIXIR-UK is developing and promoting UK-specific templates in the Data Stewardship Wizard (DSW) – an ELIXIR-recommended interoperability resource and Horizon-Europe-recommended resource for DMP creation.

The work includes 3 case studies and a total of 17 example DMPs for different data types and research projects.

2018: First ELIXIR-UK data stewardship project

Members of the UK Node participate in the first data stewardship projects within ELIXIR, collaborating with other Nodes in Europe.

2023: Launch of the RDM Club

Created a national forum for data stewards and researchers to connect, learn, and share best practices.

From just 10 attendees in early 2023, the club now reaches up to 70 attendees per month, attracting people at all career stages and UK funders.



2021-2024: Data Stewardship Fellowship

Won MRC/BBSRC-funded project to coordinate a nation-wide fellowship programme to upskill UK life sciences teams in data stewardship.



24
Fellows

17
Organisations

60
RDMbites
Short informative videos (3-5 minutes) to describe basic and specialised aspects of RDM.

27
Contributions to resources
Contribute to existing resources with knowledge and expertise from Fellows.

12
Training courses
Developed courses and workshops on basic and specialised topics in an open and reusable format

17
Training delivered
Bringing the knowledge and expertise to professionals within Fellows' organisations and virtual communities

9
Outreach activities
Embedding resources locally and amplifying the work of ELIXIR and this Fellowship

10
Sponsorships
Sponsored Fellows to attend conferences and meetings with a focus on FAIR data or capacity building in bioinformatics

2016

ELIXIR-UK members co-author the FAIR guiding principles.



Wilkinson, M., Dumontier, M., Aalberts, L. et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 5, 160018 (2016)

2013-2015 – ELIXIR-UK joins the ELIXIR Europe Consortium

The UK Node of ELIXIR is established to bring together researchers across the UK working on life sciences data, bioinformatics and computational biology, share expertise and reap the benefits of knowledge exchange with our European counterparts.

mTeSS-X:

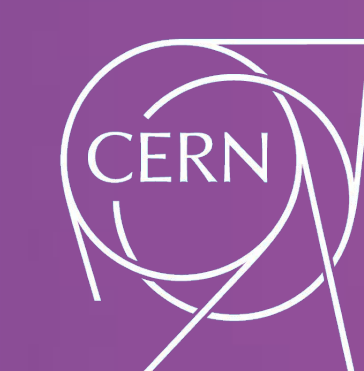
A Federated, FAIR-Aligned Platform for Distributed Management and Exchange of Training Resources

Oliver Knodel, Martin Voigt, Munazah Andrabi, Finn Bacall, Phil Reed, Kenneth Rioja, Guido Juckeland and Carole Goble

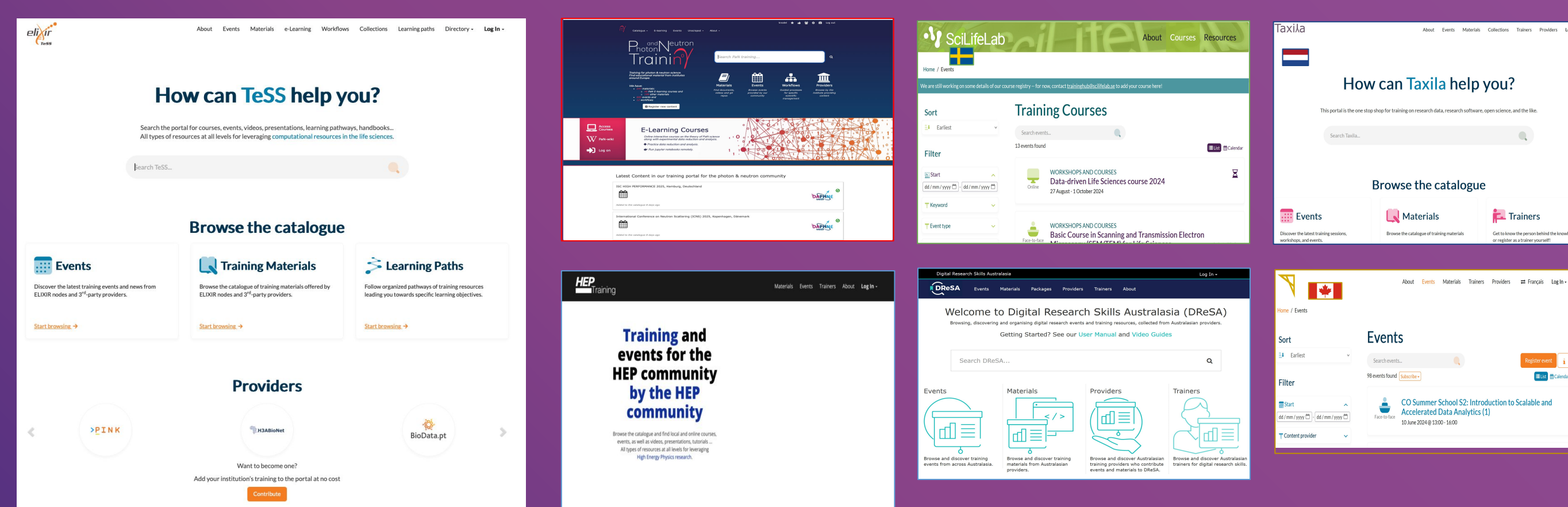
Description

ELIXIR's open-source Training e-Support System (TeSS) provides a central platform for accessing training materials and events in life science disciplines. While widely adopted, TeSS currently lacks support for content sharing between independent instances.

The **OSCARS project mTeSS-X** enhances **TeSS** by enabling **metadata exchange** and **shared content** across instances through “**spaces**”—self-managed community catalogues within a central “**hub**”. This fosters collaboration and integration across research infrastructures, creating a more connected training ecosystem.



Training catalogues based on TeSS

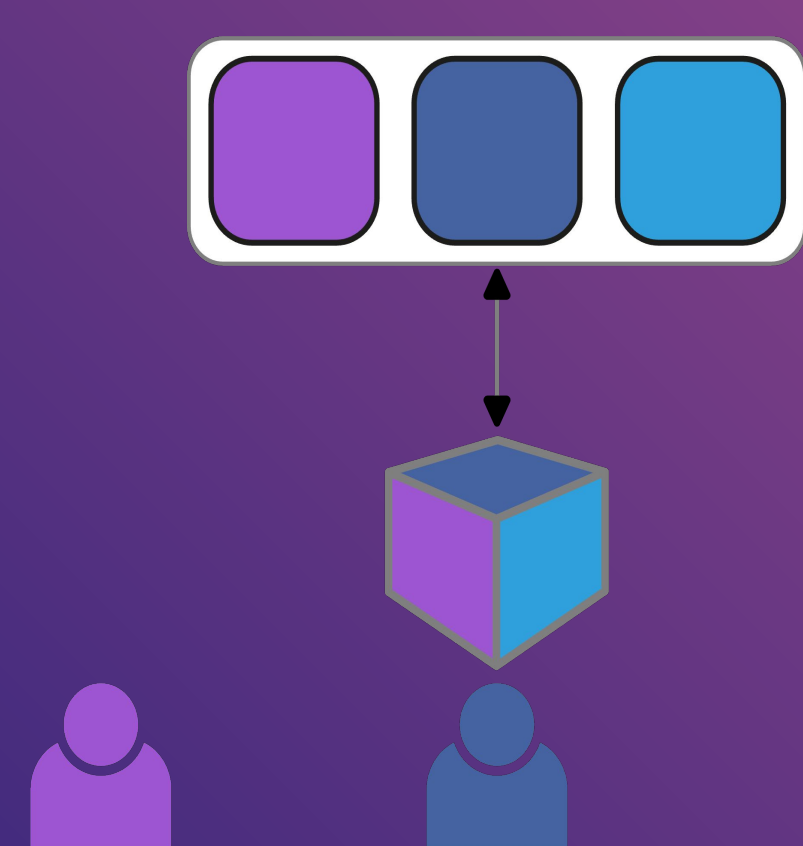


- ELIXIR TeSS Life Science training – tess.elixir-europe.org
- CERN's High Energy Physics training – training.cern.ch
- Photon and Neutron (PaN) training – pan-training.eu
- Digital Research Skills Australasia – dresa.org.au
- Taxila Open science training in the Netherlands – taxila.nl
- SciLifeLab – training.scilifelab.se

Challenge: Independent Fragmented Training Portals

- While widely adopted in a diversity of science clusters TeSS currently lacks support for content sharing between independent TeSSinstances.
- Catalogues supported by TeSS are currently:
 - Completely independant
 - Operate as siloed systems
 - Do not support content exchange
 - Lack shared metadata
- No support for cross-discipline training resource exchange
- Duplication of operational effort across platforms

mTeSS: Multi-space catalogue



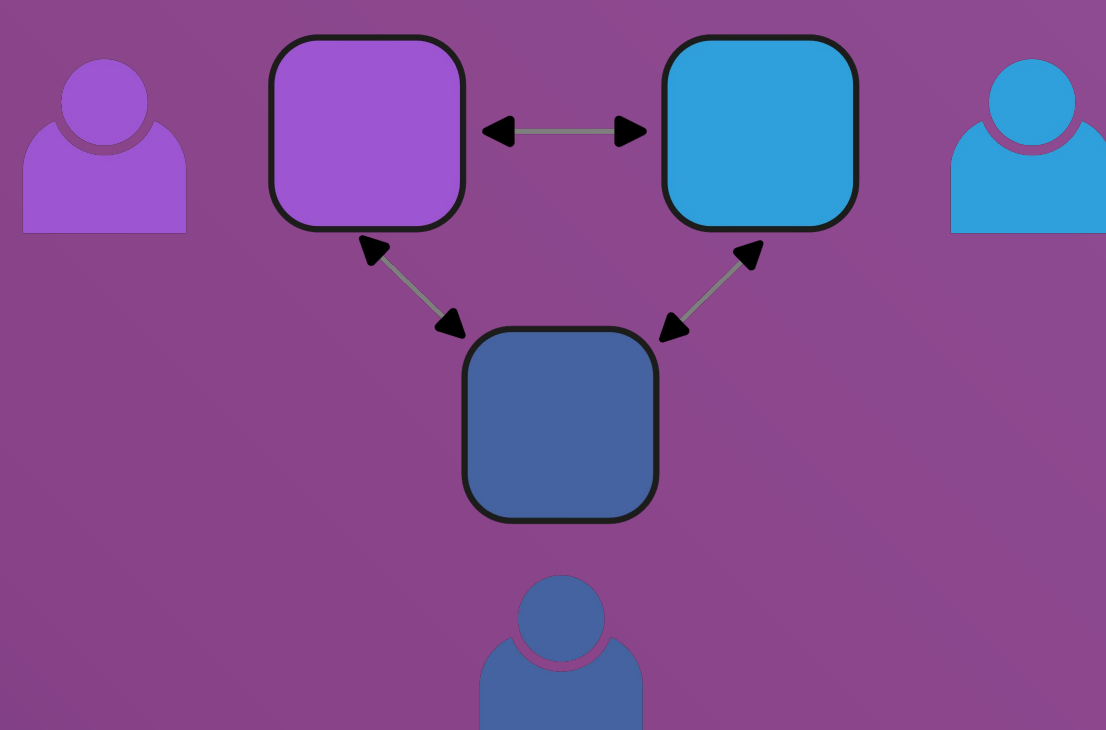
- Catalogues pooled into a **shared TeSS** instance supporting distinct community-specific “**views**” or “**spaces**” with:
 - Unique **branding** and identity
 - **Targeted** content selection
- Shared catalogue management

Scientific Impact

- **Improved Findability & Accessibility:** Federated catalogue system enables unified discovery across disciplines
- **Increased Reusability:** Content can be reused across domains and portals, promoting FAIR training principles
- **Enhanced Sustainability:** Reduced duplication and operational overhead through shared infrastructure
- **Scalability:** Framework supports future inclusion of more science clusters, communities, and RIs

TeSS-X: Exchange between catalogues

- Automatic **exchange** of nominated **content** between the spaces in a **TeSS hub and TeSS instances**
- Done via a dedicated interface for **metadata harvesting (OAI-PMH)**



Join the mTeSS-X Community

The mTeSS-X Club is an open community to help its users and gather feedback for continuous development of the mTeSS-X.

- Bi-weekly Zoom meetings
- Quarterly focus group meetings
- Slack channel: [#tess-club \(elixir-europe.slack.com\)](https://tess-club@elixir-europe.slack.com)



Learn more about
OSCARS funded projects!

www.oscars-project.eu [/company/oscars-project](https://company.oscars-project) @OSCARS-project-EU oscars-project.eu/#newsletter



Funded by
the European Union

UK Human Functional Genomics Initiative: Data Coordination Centre



Dorothea Seiler Vellame^{2,3,5}, Craig Willis^{2,3,5}, Paul Kainth^{1,5}, Starr Young^{1,5}, Jonathan Mill^{1,4,5}.

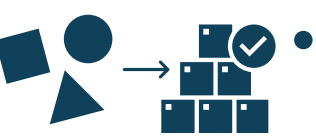
(1) FGx Coordination Hub. (2) Data Coordination Centre. (3) Research Software and Analytics Group. (4) Complex Disease Epigenomics Group. (5) University of Exeter.

FGx Initiative

The FGx is a network of UK functional genomic research to facilitate **innovation** and **collaboration**. Funded until 2029, it contains 5 research clusters, grouped by biological theme.

The DCC

Aims to:



Standardise meta-data and workflows



- Build a platform for **data linkage**



- Upskill** the community in FAIR and data stewardship

Initial steps:

Understand cluster **needs**



- Meet and **survey** data generators and bioinformaticians



- What are the data **priorities**?

Understand the **state of the field** and what is being developed by others



- Meet with partners (BioFAIR, EMBL-EBI, NMGN)



- Foster collaborations and avoid duplicating efforts

Interested in working with us? Contact here:

D.Seiler-Vellame@exeter.ac.uk

DCC-FGx@exeter.ac.uk



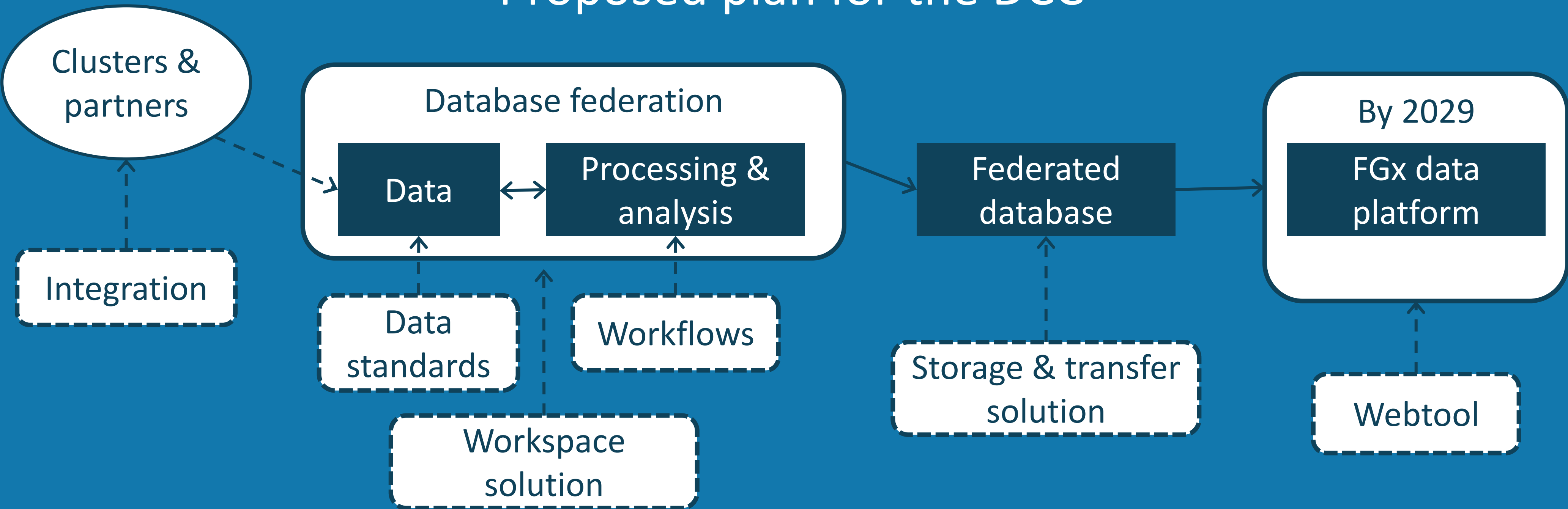
Data being generated

	Cluster				
	FGSL	Brain development	Protein PTM	Musculoskeletal	Tissue QTLs
Clinical data					
Imaging	Histochemistry				
	Multiplex immunofluorescence				
	Microscopy				
	Cellular imaging				
	Super high-resolution imaging				
	<i>Estimated total size</i>	>100s TB	5 TB	Multiple TBs	?
Genomic	SNP arrays				
	Whole genome sequencing (SR)				
	Whole genome sequencing (LR)				
	Gene expression (SR)				
	Gene expression (LR)				
	DNA modifications				
	ATAC-seq				
	scRNA-seq				
	scATAC-seq				
	Spatial transcriptomics				
	<i>Estimated total size</i>	?	>10 TB	Multiple TB	?
Proteomic	Mass-spectrometry				
	Olink/SomaScan				
	Spatial proteomics				
	<i>Estimated total size</i>	0	0 TB	Multiple TB	0

*FGSL: Functional Genomics Screening Lab LR: Long read sequencing, SR: Short read sequencing.

Our data challenge: Even common data types, when generated by different groups will **differ in meta data collected**. The data is subject to **batch effects** so is not easily merged. **Complex workflows** will be required.

Proposed plan for the DCC



Funders



Medical Research Council



Biotechnology and Biological Sciences Research Council

Research clusters



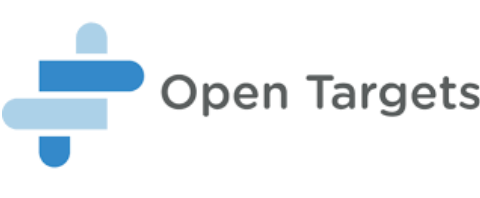
IMPERIAL



Additional partners



MRC National Mouse Genetics Network



EMBL-EBI



NIHR | National Institute for Health and Care Research



PSDI
PHYSICAL SCIENCES
DATA INFRASTRUCTURE

Designing support strategies for enabling cultural change in data practices for the physical sciences

Nicola Knight, Samantha Pearman-Kanza, Louise Saul and Cerys Willoughby

Foundations from Our Community Statement of Need

PSDI is a vision for how the physical sciences can transition to become a fully connected and digital discipline. To gather initial requirements from the community we created a Statement of Need: a rapid scoping exercise that engaged broadly with the physical sciences and related communities involving over 30 Engagement Activities with over 50 organisations.

These organisations formed part of the four pillars we identified as essential to providing resources to meet the needs of the community:

- Pillar 1. Facilities, Institutes and Hubs
- Pillar 2. National Research Facilities
- Pillar 3. Computational Initiatives
- Pillar 4. Research Institutions, research groups and laboratories

Community Events for acquiring feedback

In collaboration with our Partners and other groups, we regularly run events for the community. These events provide opportunities to learn about PSDI, exchange knowledge, share experiences, and connect with peers.

We also use these events as an opportunity for the community to share their requirements for PSDI and provide feedback on our resources and ideas.

Machine Learning Schools

Machine Learning Schools run by PSDI in collaboration with Alchemy, PSDS, AHSO, STFC-SCD and CCPs are examples of in-person education we provide. These schools help raise awareness of state-of-the-art methods and give participants hands on learning.

Technical Workshops

In addition to our in-person training, we also run workshops focused on a particular topic, e.g. Data in Polymer Science, and Units, Symbols, and Terminology. These events bring together participants with similar challenges to work on solutions that benefit the whole community.

Self-paced Learning

PSDI provides many self-paced learning resources including interactive Moodle courses and tutorials. These are designed to help researchers develop essential skills in areas such as data, computing, research data management, and other transferable research skills.

Resource Development by our Pathfinders

Pathfinders are a means to focus resource development in key strategic areas and establish exemplar approaches and systems into the PSDI infrastructure that can be expanded into further domains, data types, techniques and user communities over time.

Our initial Pathfinders focused on Catalysis, Process Recording, Machine Learning, Data Collection, Biomolecular Simulation, Reproducible Computational Workflows, and more.

PSDI

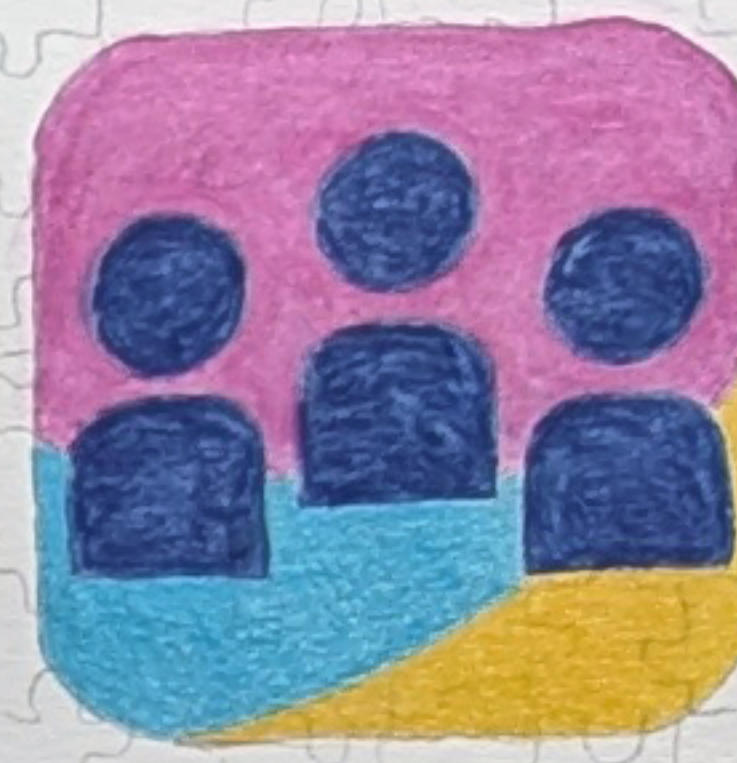
PHYSICAL SCIENCES
DATA INFRASTRUCTURE



Services



Tools



Community



Data



Guidance &
Training



Collaboration

Knowledge Base

The PSDI Knowledge Base provides access to guidance, support and training materials. A Getting Started section helps new users to find their way around PSDI resources and a Guidance section provides an introduction to data-related topics and using PSDI resources.

Webinars

PSDI hosts regular webinars, which are promoted on our Events page. The Webinars cover a wide variety of topics that are of interest to the physical sciences community and beyond. Recordings of past Webinars are available on our YouTube channel.

Case Studies

Our community value seeing examples of how technology or best practices can be used to overcome Shared Challenges. For that reason, we look to share our experiences through Case Studies. Examples include implementing ELNs and resurrecting legacy data sets.

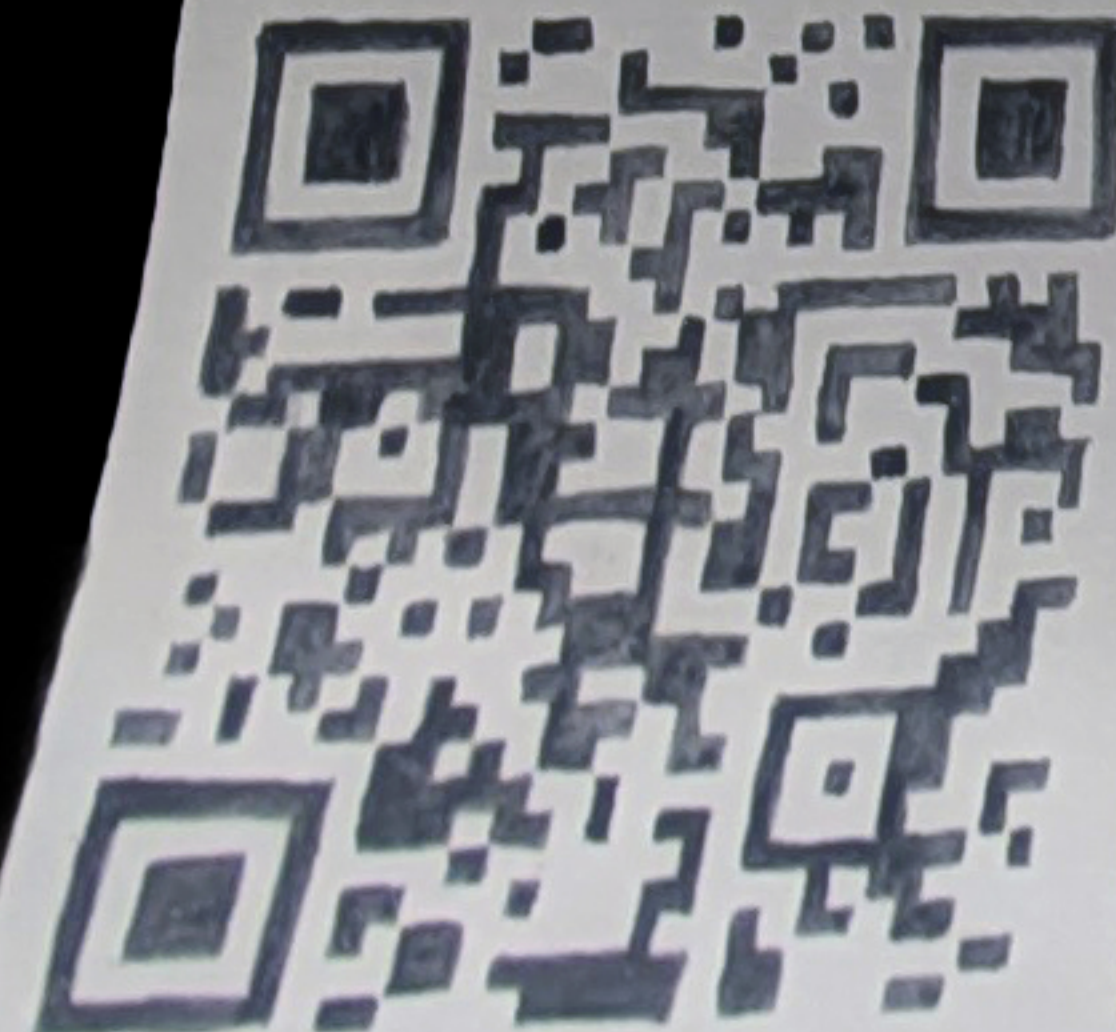
Local to International Community Networks

PSDI isn't just a single institution or organisation but a group of individuals, organisations and institutions. The main partners in PSDI are The University of Southampton and the Science Technology Facilities Council, but we are joined by a growing number of partners.

In addition to our collaborators, we have connections with related networks and organisations, big and small. These are not just those in the physical sciences but also more broadly those with related interests such as Electronic Research Notebooks, Data Stewards, and FAIR.

Visit our website at www.psdia.ac.uk

Sign up to our newsletter!



UK Research
and Innovation



University of
Southampton



Science and
Technology
Facilities Council



Engineering and
Physical Sciences
Research Council