

# PSDI

## PHYSICAL SCIENCES DATA INFRASTRUCTURE



Science and  
Technology  
Facilities Council



Engineering and  
Physical Sciences  
Research Council

Physical Sciences Data Infrastructure project 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, and Pilot and Phase 1b projects.

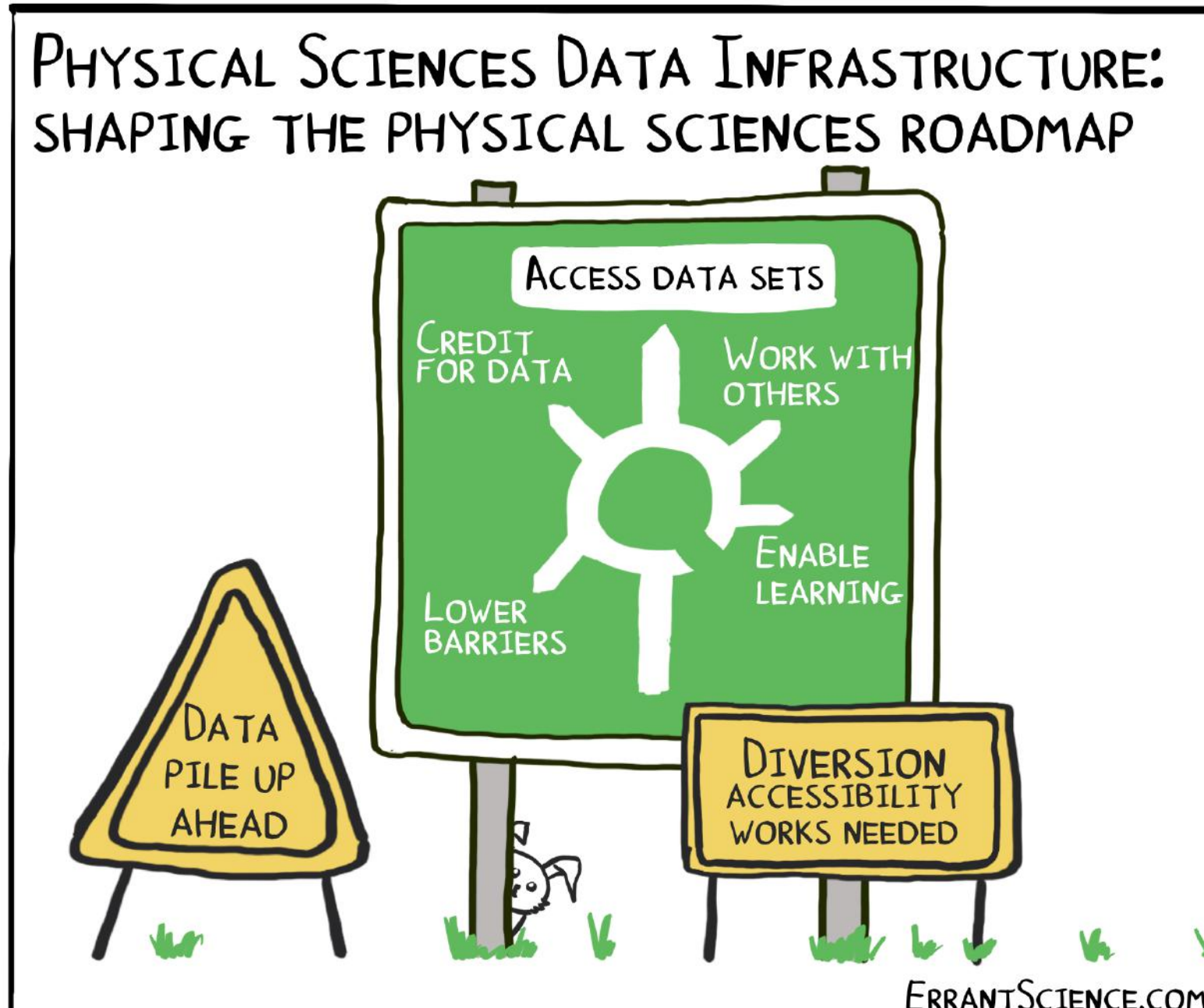
Juan Bicarregui<sup>1</sup>, Simon Coles<sup>2</sup>, Brian Matthews<sup>1</sup>, Jeremy Frey<sup>2</sup>, Barbara Montanari<sup>1</sup>  
Vasily Bunakov<sup>1</sup>, Samantha Kanza<sup>2</sup>, Nicola Knight<sup>2</sup>, Ilian Todorov<sup>1</sup>

<sup>1</sup>Scientific Computing Department, Science and Technologies Facilities Council, Rutherford Appleton Laboratory, Harwell Campus, Didcot, OX11 0QX

<sup>2</sup>School of Chemistry, Faculty of Engineering and Physical Sciences, University of Southampton, University Road, Southampton, SO17 1BJ

### Integrated Data Infrastructure for the Physical Sciences

PSDI is building a data infrastructure that brings together and extends the data systems researchers currently use. PSDI will bring together existing support infrastructures, widening their applicability, and adding value through aggregation.



### The Driver

Data needs in research are growing at previously unimaginable rates and the need for collaboration around data has never been clearer. Data is not simply a research output, but a driver of further discovery across all research. However, in the Physical Sciences, many research bodies, from large facility to laboratory, have their own data infrastructure with limited ability to share, integrate and reuse data across systems.

**Pilot Phase** - The PSDI exploratory pilot ran from Nov 2021 – Mar 2022. The pilot phase undertook a wide range of community engagement, scoping and design work. This involved partners from STFC, University of Southampton, Cardiff University, The University of Sheffield, University of Liverpool and CCDC.

### Our Recommendations

From our exploratory pilot, we produced a series of recommendations for future development of data infrastructures in the Physical Sciences. These focused around 4 areas:

- Connecting existing infrastructure
- Best use of data
- Best use of technology
- Best use of people

See [www.psd.ac.uk/the-pilot/recommendations](http://www.psd.ac.uk/the-pilot/recommendations)



## PSDI Initiation

The current phase of PSDI runs from Oct 2022 to Sept 2023 and will continue our stakeholder engagement and initiate development of components to connect existing infrastructures

The activity is organised into 5 work packages in two areas:

- The PSDI-Hub** – will provide core functionality and organisational structures to underpin PSDI operations
- Pathfinder projects** – are early deployment activities to seed the infrastructure with exemplar tools, data and services

#### WP1: Management and Governance

Project co-ordination and management. Establishing governance and oversight mechanisms.

#### WP2: Communications and Training

Communication of project activities and development to the wider community. Includes a programme of engagement activities with PSDI stakeholders.

#### WP3: Platform Development & Operation

Evaluation of technology options and development of the technical platform. Works closely with Pathfinders.

#### WP4: First round Pathfinders

Activities to develop initial content available through PSDI with development focused around key strategic areas.

#### WP5: Future Pathfinders

Development of candidates for future pathfinders. Builds on case studies and engagement from the pilot phase as well as wider community engagement.

## Current Pathfinders

**Pathfinder 1: Experimental data capture** - co-create an integrated environment to support, capture, analyse reuse of XAS data.

**Pathfinder 2: Process Recording** - implement process recording into the Hub and develop of metadata and ontology layers to support data analysis and search.

**Pathfinder 3: Building Data Collections** - explore and develop methods to build, store, manage and access data collections across the physical sciences.

**Pathfinder 4: Process Orchestration** - create a repository for bimolecular simulation trajectories that includes data storage, and automated capture of metadata and provenance.

**Pathfinder 5: Data to Knowledge** - construct workflows to connect data sources for data ingestion, discovery, integration, analysis and visualisation.



Find us on Twitter:  
[@PSDI\\_UK](https://twitter.com/PSDI_UK)

Reports and more  
information on our website:  
[www.psd.ac.uk](http://www.psd.ac.uk)