PSDI

PHYSICAL SCIENCES DATA INFRASTRUCTURE

Connecting Digital Research Infrastructures for the Physical Sciences

PHASE 1 PILOT REPORT

November 2021—March 2022

Contents

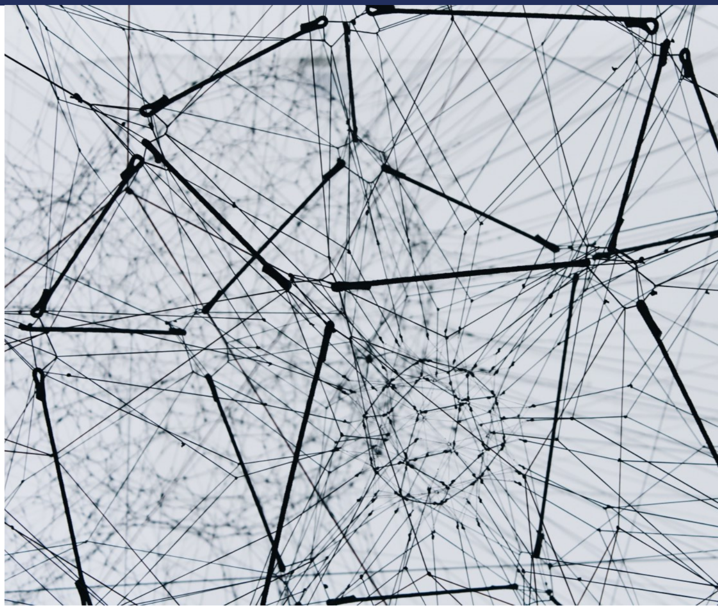
Through broad engagement with researchers and infrastructure providers, the Physical Sciences Data Infrastructure pilot elaborated a vision for how the physical sciences can transition to become a fully-connected, digitally-enabled discipline. This report summarises the PSDI pilot phase project and outlines its key findings.

Introduction to PSDI	2
Physical Sciences Stakeholders	3
Pilot Activities	
Co-ordination	4
Engagement	5
Architecture	6
Case Studies	7
Future Vision	9
Recommendations	10
Connecting Existing Infrastructures	11
Best Use of Data	12
Best Use of People	13
Best Use of Technology	14
PSDI Development	15
Pathfinder Candidates	16
PSDI in the Future	17
Acknowledgements	18

Introduction to PSDI

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 an output of research, but a driver of further discovery in and across the physical sciences. But mostly every research infrastructure, from large facility to laboratory, has its own data infrastructure with limited ability to share, integrate and reuse data across systems.



PSDI aims to enable researchers to:

- Find and Access reference quality data from commercial and open sources
- Combine data from different sources
- Share data, software and models including experimental and simulation data
- Deploy close-to-data computation and containerisation of data and software
- Use AI to explore data
- Learn how to make the results of their research open and FAIR

Connecting the research landscape

The PSDI aims to serve the data needs of the physical sciences community principally by connecting existing digital tools and data sources with each other, and with support provided for and by other disciplines. Here, 'Physical Sciences' primarily means the areas of physical science research funded by EPSRC, but reaching out to related themes in Energy, Engineering and Manufacturing and also to other domains. Accordingly, PSDI will connect with related data infrastructures across research council boundaries where useful, e.g., chemistry related aspects of natural environment, life science and medical research.

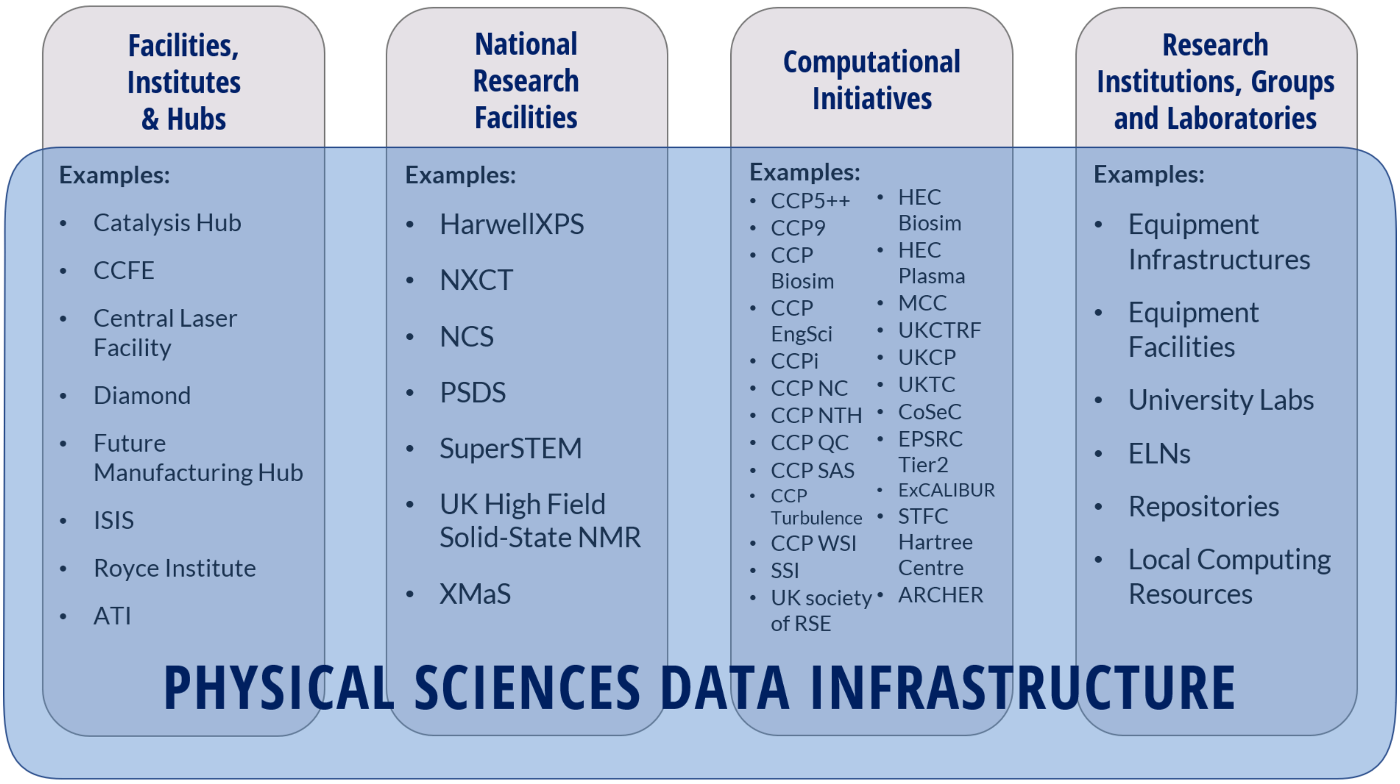
The need for a connected digital research infrastructure serving the physical sciences was identified in a *Statement of Need* presented to EPSRC in early 2021. At EPSRC's request a proposal for a pilot project was submitted in October 2021. The PSDI-pilot ran from November 2021 to March 2022, supported by EPSRC DRI (digital research infrastructure) funding.

The PSDI Statement of Need can be downloaded from https://www.psdi.ac.uk/background/

Physical Sciences Stakeholders

The Statement of Need identified four "pillars" of stakeholders for physical sciences research, each pillar representing a significant group of data producers and consumers.

- **Pillar 1. Facilities, Institutes and Hubs** – significant centralised national facilities and activities that serve many researchers based on a common need.
- **Pillar 2. National Research Facilities** – medium-scale centralised facilities operating at a level that cannot be addressed in a standard laboratory.
- **Pillar 3. Computational Initiatives** – digital research infrastructure and national research communities primarily focused on computational "physical sciences" (as defined above).
- **Pillar 4. Research Institutions, Groups and Laboratories** – the community of research teams in institutions with strong scientific profiles and some data management and computing capabilities.



In the PSDI Pilot the majority of the engagement centred around pillars 1 to 3 due to the short timescales required for delivery. Although the longer tail of pillar 4 was engaged with through networks, presentations and some case study work.

Further expansion of consultation with pillar 4 will be a key area of future work on PSDI.

Pilot Coordination

The PSDI pilot project worked with researchers and infrastructure providers across all of stakeholder pillars. The aims of the pilot were:

- Engage with the PSDI stakeholder community and build support for its creation
- Undertake some case studies to demonstrate the potential scientific benefits
- Trial some relevant technologies and investigate their interoperability
- Gather requirements arising from the case studies and trials and wider consultation
- Analyse these requirements, and propose technology recommendations for PSDI
- Produce a detailed plan for future phases of PSDI

The work was structured into 4 work packages, with one of the PSDI leadership team taking lead on each of the work packages.

WP 1 - Coordination, Governance and Strategy oversaw the pilot project organisation, considered the governance framework that will be required for later phases, and outlined a plan for PSDI going forward.

WP 2 - Stakeholder Engagement undertook a broad programme of workshops, focus groups, discussions and interviews, engaging primarily with a wide range of stakeholders across the four pillars to elicit requirements.

WP 3 - Architecture & Technology investigated and evaluated technology options and potential architecture designs for PSDI, including small-scale technology trials investigating interoperability, security and data semantics.

WP 4 - Case Studies undertook case studies which prototyped systems that demonstrated how proposed elements of PSDI could influence cutting-edge research.

The work was led by a team from the School of Chemistry at the University of Southampton and the Scientific Computing Department at STFC, with additional contributions from the Universities of Sheffield, Liverpool and Cardiff, and the Cambridge Crystallographic Data Centre.

In total, more than 120 Staff Months of effort were delivered during the five-month period with over 50 individuals contributing to the work.

Our website can be found at

www.psdi.ac.uk

We are also on Twitter at

@PSDI_UK

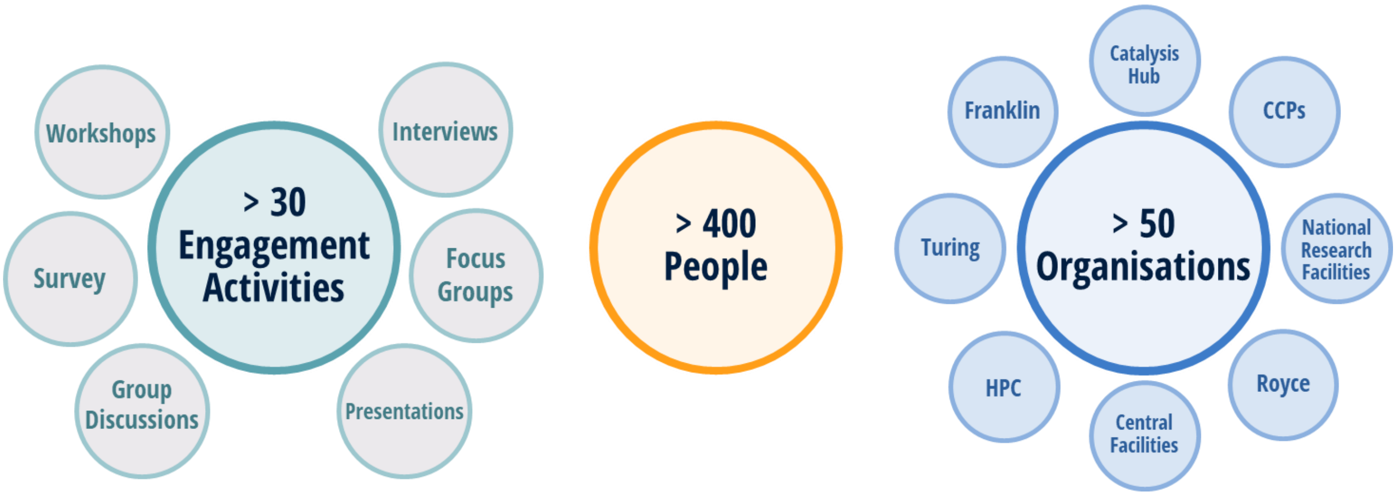
Engagement

The pilot project team engaged extensively with a widely range of stakeholders from across the four pillars and more broadly with other domains and internationally. There were more than **30 engagement activities**, involving in excess of **400 people** from over **50 organisations**.

Engagement was undertaken through a variety of modalities, including:

- workshops
- bilateral meetings
- presentations at events
- group discussions
- focus groups
- interviews
- survey

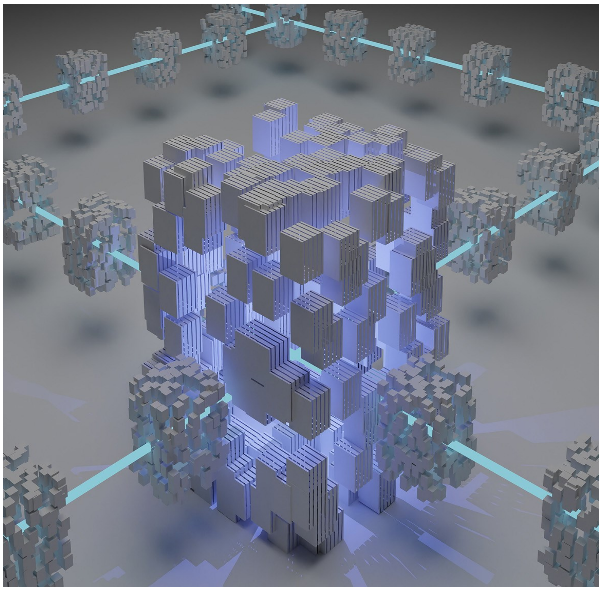
Interactive on-line boards were used to facilitate input/feedback. Engagement activities were structured around a common set of questions of interest, covering data management, data analysis, standards for data exchange, workflows, and collaborations. They were aimed to investigate data requirements and elaborate on the benefits that could be achieved from a more integrated data infrastructure. The Statement of Need provided a very useful baseline vision upon which to base discussions.



Delivering such a large-scale evaluation exercise in such a short timescale was possible due to the significant expertise and contacts the Southampton-STFC team has built up over decades of pushing forwards the digital infrastructure agenda, and developing and operating data infrastructures for the physical sciences and other domains. While the level of engagement was very significant during the very limited duration of this pilot, continuation and expansion of the consultation will be required to further steer the development of PSDI.

Outcomes from each of the engagement activities fed in to the recommendations from the pilot phase, in particular looking at the synergies between engagement in different areas. These are incorporated in reports on the PSDI website : www.psdi.ac.uk/the-pilot

Architecture



The pilot explored some candidate technological approaches to provisioning a data infrastructure for the Physical Sciences and more widely.

The PSDI proposes to provide a state-of-the-art, enterprise-scale digital infrastructure for research that is distributed across a number of centres and is responsive to the needs of the user community. It should be a scalable, adaptable and accessible infrastructure that is secure and reliable, while also being compatible with related data infrastructure initiatives supporting research.

Architectural options were explored through group exercises to capture and survey the technical requirements for a data infrastructure for physical sciences. These identified a set of capabilities the infrastructure should provide to satisfy the requirements and pointed to a service architecture model that would deliver those capabilities.

Three detailed studies were also undertaken focussed on particular technology areas:

- **Workflow Orchestration:** a study to evaluate and demonstrate the use of the AiiDA workflow orchestration tool for the capture of data provenance for molecular simulation. The study trialled AiiDA within the bio-molecular simulation field using the UK's national supercomputer ARCHER2 and other HPC platforms.
- **Materials Ontologies:** considered the best practise and state of the art in the definition and use of ontologies to support information management within materials science.
- **User identity and security infrastructure:** considered requirements for a federated authentication system for PSDI to give a high level of personal privacy and data integrity. This would need to be deployed in collaboration with the wider UK research infrastructure.

WP3 identified the following principles for designing PSDI architecture:

- Don't reinvent
- Openly accessible
- Secure as needed
- Support FAIR
- Integrate with existing services and capabilities
- Interoperate internationally
- Use Open Standards

Outcomes from the architecture design investigations fed in to the recommendations from the pilot phase, and are incorporated in reports on the PSDI website : www.psdi.ac.uk/the-pilot

Case Studies

The pilot undertook 8 case studies looking in more depth at how aspects of functionality of a future PSDI could be developed. They were designed both as demonstrators and to provide a practical test of how to digitally underpin key aspects of physical sciences research. They covered a wide range of research areas, techniques and infrastructure requirements with some covering science driven investigations and some exploring cross-domain underpinning functions. They predominantly spanned pillars 1 to 3 as these were the most representative and tractable to engage with in the given timescale.

SCIENCE DRIVEN

CS1 Data and Simulation driven understanding of catalytic activity

Demonstrate the practice and value of linking and combining data from across experimental data facilities. The initial target area was X-Ray Absorption Spectroscopy (XAS) data processing. Publications of the UKCH were analysed to explore reproduction of results and the barriers to finding and reusing data.

CS2 Exploring CSD-Theory as a tool for assisting materials discovery

Exploration and testing of the CCDC's "CSD-Theory" software for storing, sharing, and analysing results of computational crystal structure prediction on porous organic materials.
It investigated CSD-Theory's ease-of-use, available tools, and extendability for the purposes of sharing computational results with experimental colleagues, and for expediting analytical calculations on a published example dataset.

CS3 Combining data sources in Materials Physics

Evaluation of the requirements for storing experimental and Natural Language Processing (NLP) mined data. The main areas of focus for this case study were:

- the application of Natural Language Processing (NLP) to extract data from the functional magnetic material literature
- the architecture required to merge this extracted data with laboratory data to form a resource description framework (RDF) database.

Summary and detail reports can be found for each case study on the PSDI website : www.psdi.ac.uk/the-pilot

UNDERPINNING FUNCTIONS

CS4 Spectroscopic data infrastructure

Evaluate technology and data requirements to underpin spectroscopy characterisation techniques across all disciplines using the infrastructure. This included technical requirements for a repository system from data management and domain perspectives along with integration with workflows and 3rd-party tools. Exemplar data and workflows from Supramolecular Chemistry were used as the initial proof of concept.

CS5 Data curation and availability at instrument-based facilities

Exploration of data management necessary to publish standalone datasets, for formal publishing routes or machine learning, in the NRF for X-ray CT (NXCT) at Southampton. The O(PB) archive of computed tomography data includes captured metadata, with the work assessing if this is “good enough” to publish, and prototyping tools to “bridge the gap”. It also examines if this can translate to other instrument facilities.

CS6 Process Recording and Digital Research Notebooks

Assessment of the digital landscape around process recording and scientific notetaking, particularly Digital Research Notebooks (DRNs). This included; a survey on working practices, examination of the diversity of tools available and technologies for improving process recording, and exploring user requirements and interactions through personas, storyboards and user/data flows.

CS7 Data trust, sharing & preservation

Exploration of data trust and sharing framework and the applicability to PSDI. This also explored broader options for data exchange and sharing frameworks and existing systems set up as data exchanges or to facilitate data sharing. The case study also undertook some exploratory work into potential requirements of data exchange infrastructures and the scope of what such a mechanism could offer.

CS8 The Role of Structure in Physical Sciences Data Management

A review of the importance and challenge of reliable structure representation, the need for enabling workflows and infrastructure, the importance of having curated structure-based data resources, and the opportunities for global collaboration with industry and academia.

Each case study produced a detail report of their work and a summary which focused on the key outcomes and recommendations made by the case study. Consultation with the researchers, along with their recommendations and regular all-project discussions were a key component that fed into the overall recommendations from the PSDI pilot.

Future Vision

An integrated DRI for PS

PSDI aims to accelerate research in the physical sciences by providing a digital research infrastructure (DRI) that brings together and builds upon the systems researchers currently use. PSDI will bring together existing support infrastructures, widening their applicability, and adding value through aggregation.

We envision that PSDI will enable researchers to:

- Find and Access to reference quality data from commercial and open sources
- Combine data from different sources
- Make the results of their research open and FAIR
- Use AI to explore data
- Leverage simulation data to drive experimental science and vice versa
- Share data, software and models including experimental and simulation data
- Curate legacy resources beyond individual projects

These will be achieved by:

- Surfacing data from many sources
- Providing reference-quality data
- Standardising, normalising and aggregating data and metadata
- Supporting workflows that automate data processing
- Supporting software curation and publication
- Supporting multiscale modelling and multimodal research
- Providing a common platform to run models and codes from different sources
- Enabling access to performance compute for scaling up

PSDI Recommendations

To enable a series of recommendations to be produced across the whole pilot project, the findings from all the work packages in the exploratory pilot were brought together and examined. This included findings and recommendations from the case study activities, the technical investigations, the community engagement activities and expert input from the project team.

From the discussion surrounding these findings we produced a series of top level recommendations for future development of data infrastructures in the Physical Sciences.

These recommendations were grouped into 4 areas:

Connecting existing infrastructure

Best use of data

Best use of people

Best use of technology

These areas of top level recommendations are expanded on, with details of the individual recommendations, over the next few pages.

Recommendations at a more detailed level were also generated in many of the work packages and case studies. These findings can be explored in the additional reports produced through the pilot work. A listing of the pilot reports can be found on the website at www.psdi.ac.uk/the-pilot

Connecting Existing Infrastructures

The primary finding of the pilot confirmed the view set out in the Statement of Need, that the current data landscape is fragmented, making data analysis and reuse unduly complex, and that physical sciences research would be greatly accelerated by more integration of the systems that handle data. This would enable researchers not only to undertake their own analyses more effectively, but also to make their data products available as inputs for further research.

There is strong view in the community that the major need is for a data infrastructure that connects existing systems, widening their applicability and adding value through aggregation, rather than for the development of new functions. Such an integration would support data workflows enabling researchers to concentrate on their science rather than spend time on data management activities.

Stakeholders were understandably concerned that a data infrastructure must be trustworthy and enduring as without assurance of its longevity, researchers would be reticent to invest the time required to engage with a system which may be temporary or fail to gain traction as key infrastructure.

R1	Construct an integrated digital infrastructure for physical sciences that connects existing research data services, both with each other, with those from other domains, and internationally.
R2	The infrastructure must facilitate curation, and long-term support for data, software and services beyond the lifespan of individual projects.
R3	Encourage and implement governance and communication mechanisms whereby co-operation and co-creation between all stakeholder organisations is enabled and fostered across domains and internationally.

There is great opportunity for alignment with data management activities, across the UKRI councils, with international projects, and with industry. From our analysis and discussions with key overseas players, it is clear that other countries, most notably Germany, the US, Australia and Japan, are forging ahead with building data infrastructures. With the rise of AI and digitally-driven working practices, we are on the verge of a digital revolution in research and industry and, without coherent support for UK physical sciences to exploit this opportunity, the UK risks falling behind in an area where it has hitherto been a leader.

Best Use of Data

There is a need to open up data for reuse and aggregation into collections that add value, and to link up with data sources from other domains for cross-disciplinary, multiscale modelling and multimodal research. It should be possible to readily access provenanced data, including reference quality data, and secondary data underpinning publications. Availability of data should support reproducibility and validation of research, in addition to application in further research including machine learning and AI.

There is also a crucial need to establish better data-level connectivity across the pillars, particularly bridging between experimental and computational activities.

For an integrated, distributed physical sciences data landscape to be realised, some new connecting functionality will need to be developed. The new infrastructure should support the overarching principles of data being as open and FAIR as possible, and drive international collaborations and interdisciplinary research through the use of open standards.

R4

Develop a toolkit for publishing data collections for physical science researchers, covering data management, metadata, ontologies, curation, provenance and tools supporting research teams to implement data polices.

R5

Establish mechanisms to provide access to provenanced data from many sources, notably:

- reference quality data from commercial and open sources
- original data generated from experiments and simulations
- secondary data underpinning articles, theses and reports
- derived or analysed intermediate data
- collections of results data representing aggregations of properties or features of analysed data

R6

Provide tooling to support reproduceable data processing workflows, including providing a common platform to run models and code; access, transfer and integrate data from different sources; and more easily access performance compute services for scaling up.

R7

Develop support for transforming data to knowledge, including visualisation, discovery, data mining, aggregating data for integrative science, and AI techniques.

Best Use of People

It was clearly recognised that an effective research ecosystem requires not only investment in technology, but also needs support professionals to make it usable, and appropriately trained people to fully exploit it. We observed a wide variation in levels of data skills in different groups. This highlighted an opportunity for sharing knowledge and best practice between projects, disciplines and research domains.

Much of a physical sciences researcher time is spent finding, cleaning, transforming and importing/exporting data. There is a need for dedicated professionals who can either fully support researchers' data workflows enabling them to concentrate on research without being impeded by cumbersome data management, or provide streamlined tools supporting data intensive research in the physical sciences enabling researchers to more easily support themselves. The role of these professionals must be fully established, recognised and sustained.

The physical sciences research community is extensive and varied and therefore needs broad community participation in its governance, planning and development.

R8	Provide co-ordination for community engagement activities, discussions and training, supporting community adoption, and enabling community input to ensure gaps are identified and plans developed to address them in a coordinated way.

R9	Nucleate expertise for the support and service of research groups in data science for physical sciences, providing community training and support.

R10	Establish recognition and professionalisation for roles related to provision of data tools and infrastructure such as RSEs, data scientists, engineers and curators.

R11	Establish a governance structure for PSDI involving wide-ranging expertise to provide steering such as prioritising activities, advising on allocation of resources and overseeing developments, as well as to work with other projects to deliver complementary functionality across the UK's Digital Research Infrastructure.

Best Use of Technology

Information Technology for data management and data analysis is rapidly changing and often diverging, with important new functionality emerging continuously. Physical science researchers currently have to navigate a wide diversity of provision in a highly heterogeneous technological environment. Physical Science research workflows should "Ride the wave" of technology evolution and integrate the latest technological developments.

Providing an integrated infrastructure where researchers can adopt diverse tools, yet continue to work together will require agreement on and maintenance of the vocabulary, interfaces and tools that enable interoperability. An essential feature of a data infrastructure for physical sciences should thus be to develop and maintain interoperability standards and the associated supporting tools that enable sharing and discovery of metadata and data.

R12 Early-stage deployment of the PSDI should include services enabling the connection of existing provision, extensible to under-supported areas of physical science, together with the standards to underpin the interoperability of those services. In particular, the PSDI should deploy tools and services

R12.1 *Connecting data:* aggregated catalogues of data and services; creation of standards for data formats, metadata, vocabularies and ontologies.

R12.2 *Connecting Services:* protocols and APIs; security and access control mechanisms; data storage and transfer systems; cloud services; monitoring and reporting systems; computation workflow orchestration systems.

R13 An underlying principle of the PSDI should be to adopt existing technology where it has already been successfully deployed elsewhere, adapting it if necessary, working in partnership with those developing it for other domains.

Developing a technological environment that is integrated and consistent, will require components that bridge between existing technology. Examples of such connected infrastructures can be found in initiatives in related disciplines such as the computing centres within the IRIS programme for fundamental physics, the NERC data archives and JASMIN system, the Ada Lovelace programme for national facilities, and the EBI for life sciences. We propose a similar approach within Physical Sciences.

PSDI Development

The focus of PSDI is not to build a centralised new system, but rather to bring together existing support infrastructures; widening their applicability, and adding value through aggregation, making the whole greater than the sum of its parts. To make this aggregation succeed new functionality is required, with some components best provided by a central hub, and others being distributed. Existing components will also need adaptation or extension to be available through PSDI.

The proposed programme for the next stage of PSDI focuses on the following activity:

The PSDI-Hub develop core functionality to make PSDI work as an integrated whole

Pathfinder projects small scale projects will create prototype tools and services. These may focus on domain integration, cross-domain application or expertise delivery.

PSDI Hub

The PSDI-Hub will cover core services required to support the platform and integration with other infrastructures. Examples of Hub activities include the following.

Governance & coordination : PSDI will serve a wide community so needs broad participation in governance, planning and development. A governance structure with broad representation will be coordinated to oversee the project. The Hub will also liaise with other DRI initiatives and related activities.

PSDI Management : The PSDI management office will provide project management, report generation, finance support, admin support, communications, events and training logistics.

Development of core data infrastructure components : These include provision of catalogues of data and services, and support for community creation of standards, ontologies, protocols and APIs. The Hub will enable access to scalable data storage and transfer systems, cloud services, etc. and will adopt access control mechanisms with sufficient flexibility to enable effective content maintenance.

Hub Operations : Activities related to technical operation of the PSDI Hub services. They include platform provision and maintenance, system administration, security, user support, licensing, and system monitoring, evaluation and reporting.

Communication & Engagement : Expanding the broad programme of engagement from the pilot. Engaging with UKRI infrastructure providers will be key to effective data integration across sources and research disciplines. International engagement will ensure interoperability with infrastructures overseas. PSDI stakeholders will be kept informed via channels including the website, social media, webinars, publications, conferences etc.

People & Skills : Working with other initiatives across the UKRI DRI people and skills theme, PSDI will coordinate development activities for data-focussed DRI professionals. It will advocate for recognition of the key research contribution of DRI professionals. It will co-design and co-deliver programmes of education and training for researchers in order to support adoption and appropriately skilled use of PSDI and UKRI data infrastructures.

Pathfinder Candidates

Pathfinders are early adoption activities where the pilot project demonstrated readiness and potential added value. These activities would be developed in parallel with the Hub and focus on delivering first content through PSDI, with additional features added incrementally. They are designed to act as exemplars for subsequent data integration work. Areas identified for potential pathfinders are organised into three categories.

Domain Integration Pathfinders

Domain Integration pathfinders are selected with a focus on integrating tools, services or data from a specific research domain and will be constructed with key stakeholders. Although these may be initially focused on a single domain, any development will also consider wider adaption and implementation. Candidate areas include:

Catalysis research involves combining data from simulation, characterisation and activity/properties of a huge array of candidate compounds.

Materials design is predicated on understanding and controlling the interplay between processing, microstructure and properties data.

Biomolecular science links physical science with biological and medical science., with a focus on biomolecular simulation trajectories

Cross-domain Pathfinders

Cross-domain pathfinders do not focus on a specific domain but instead will focus on areas where services and tools can be developed and integrated which have an application across many domains. Candidate areas include:

Reference data typically comes in many forms (proprietary, licensed, Open) and is the basis for context for novel research, as well as scientific discovery and enabling derivation of assertions, trends, rules and increasingly AI.

Spectroscopy is a key underpinning technique in environmental, life and physical sciences providing not only characterisation of samples but also in-situ analysis of dynamic systems.

Process recording is fundamental to driving validation, trust, reuse of data across the sciences and is a crucial aspect of data management and integration.

Expertise-based Pathfinders

Expertise-based pathfinders will be formed through partnerships with key established organisations, augmenting activities specifically for physical sciences and related disciplines. Candidates include:

How to publish data collections : expertise in data management, curation, and the metadata, ontology and provenance standards for FAIR data that support researchers to implement funders' data policy.

Data Science Skills : including professionalisation and recognition of expertise combining Research Software Engineering for the physical sciences, Data science, Data engineering and carpentry, development of pedagogical data examples, training through CDTs and Departments.

Data to Knowledge : including expertise in workflows connecting various data sources (e.g., from computational and experimental facilities to ELNs) for data ingestion, discovery, integration, analysis, and visualisation; knowledge discovery will be underpinned by expertise in computational science, data mining and AI.

PSDI in the Future

The development of PSDI will necessarily be incremental; where the challenges and benefits are already known, some development can begin immediately, other areas require more community discussion and scoping before commencing. This incremental process will allow additional pathfinders to be identified and incorporated into PSDI.

As PSDI continues the **PSDI Hub** will remain as the integrating and coordinating entity. It will run as a production level service, but will also be developed to extend its functionality and, if necessary, adapt to enable integration of additional initiatives and pathfinders. The PSDI would be part of the federated platform which brings together these services and offers them, free at point of access, to the UK research community.

Pathfinders in PSDI will have two routes for further development.

- Where services are core to PSDI, they may be incorporated into Hub operations.
- Others may evolve into **Competency Centres** providing services to the community within their specific fields of expertise.

PSDI will also enable co-creation of tools, services and resources with new research projects which interface with PSDI, or extend it in novel ways.

As PSDI progresses beyond its initial phases, data federation with **other UKRI data initiatives** will be a focus, as will embedding UKRI federated data infrastructure into research culture, and supporting the **skills** and **DRI-focused careers** agenda.

Physical sciences data repositories and tools will be linked with those in other disciplines to facilitate the use of physical and chemical data in research in environmental, biological, medical and other sciences. We envisage programmes of co-design of data workflows that underpin **cross-disciplinary research**, including combining information sources currently siloed in different disciplines. Developing **common standards** will be a key activity to enable cross-disciplinary data sharing.

These different types of activity, imply a **programme** comprising of a mixture of operational, development and research activities, with funding from different sources appropriate to the type of activity.

Key to the future success of the PSDI is recognition of the need for sustained investment in DRI. This allows for **long-term reliability** of services and allows the community to **trust** and invest time into those resources.

Acknowledgements

PSDI Leadership Team

Dr Juan Bicarregui, STFC (Principal Investigator)
Dr Barbara Montanari, STFC
Professor Simon Coles, University of Southampton
Professor Jeremy Frey, University of Southampton
Dr Brian Matthews, STFC
Dr Vasily Bunakov, STFC
Dr Nicola Knight, University of Southampton

We thank EPSRC for their funding for this project through grant EP/W032252/1 and their staff, advisory members and steering panels.

PSDI Pilot Project Team

Alexander Dibbo, STFC
Alin Elena, STFC
Cheney Ketley, STFC
Damian Jones, STFC
Dave Emerson, STFC
David Crooks, STFC
Derek Ross, STFC
Dominik Jochym, STFC
Edoardo Pasca, STFC
Esme Williams, STFC
Georgia Lomas, STFC
Gilberto Teobaldi, STFC
Illian Todorov, STFC
James Gebbie, STFC
Jens Jensen, STFC
Leon Petit, STFC
Marion Samler, STFC
Martin Horsch, STFC
Martin Plummer, STFC
Martyn Winn, STFC
Michael Seaton, STFC
Nick Hill, STFC
Noel Vizcaino, STFC
Rocio Garavito Ramirez, STFC
Silvia Chiaccheria, STFC
Abraham Nieva de la Hidalga, UK Catalysis Hub
Richard Catlow, UK Catalysis Hub
Steve Brewer, University of Lincoln
Andrew Cooper, University of Liverpool
Joseph Thacker, University of Liverpool
Nicola Morley, University of Sheffield
Richard Rowan Robinson, University of Sheffield
Zhaoyuan Leong, University of Sheffield
Ian Bruno, CCDC
Suzanna Ward, CCDC
Cerys Willoughby, University of Southampton
Charlie Holdship, University of Southampton
Christopher Taylor, University of Southampton
Colin Bird, University of Southampton
Don Cruickshank, University of Southampton
Graeme Day, University of Southampton
Ian Sinclair, University of Southampton
Richard Boardman, University of Southampton
Robert Bannister, University of Southampton
Samantha Kanza, University of Southampton
Thomas Allam, University of Southampton
Victoria Hooper, University of Southampton
Ronald Swart – Centre for Process Innovation

Organisations

We extend our thanks to all of the organisations that engaged with us during this pilot project. We have used their experiences and knowledge to guide this report, but the content does not necessarily reflect the views of the organisations.

Alan Turing Institute (ATI)
UK Catalysis Hub
Central Laser Facility
Diamond Light Source
ISIS Neutron and Muon Source
Rosalind Franklin Institute (RFI)
Henry Royce Institute
EPR National Research Facility (NRF)
HarwellXPS NRF
Ion Beam Facility NRF
NCS NRF
NXCT NRF
PSDS NRF
UK High-Field Solid-State NMR NRF
XMaS NRF
CCP Steering Panel (I.e., Chairs of CCPs, representatives of Research Councils, intl representatives from CECAM etc.)
CCP-Biosim
CCP-NC
CCP-NTH
CCPWSI+
CCP4
CCP5++
CCP9
CCP Turbulence
CCP-EM
European Psi-k Network
AiiDA
European Materials Modelling Council
MMMHub
High-End Computing Consortium (HEC) MCC
HEC UKCP
HEC UKCT
HEC UKCTRF
HEC UK-AMOR
HEC Plasma
HEC Biosim
HEC UKCOMES
RSE
SSI
IRIS
DAFNI
JASMIN
Tier 2 HPC Providers
N8-CIR
Wider PS Community
CCDC
EMBL-EBI
NERC
EPSRC
STFC
NOMAD
CECAM

Logo by Weareherd
Report design & preparation by Nicola Knight

Contact Info:

www.psds.ac.uk

Twitter @PSDI_UK

PSDI
PHYSICAL SCIENCES
DATA INFRASTRUCTURE

UKRI Engineering and Physical Sciences Research Council

Funded by EPSRC grant EP/W032252/1