

PSDI is an integrated data infrastructure for empowering your physical sciences data and streamlining your processes.

Get started with PSDI



# **Explore our Resources**



# **Demos and Exemplars**

# Biomolecular simulation



A case study of technologies to enable and simplify the capture and sharing of data from within biomolecular simulation workflows.

# Data to knowledge infrastructure



Creating infrastructure and training in the field of machine learning interatomic potentials.

# File format converter service



A catalogue of scientific file formats and a tool to convert between them.

#### Your case study here, click to find out more!



## **TOWNHALL** 20th June, IOM3 London psdi.ac.uk/event/psdi-townhall/

This area is intended to showcase

**Contact us** 

This area is intended to sho work contributed by our pathfinders and other collaborators.

# Get involved







Engineering and Physical Sciences Research Council Quick Links

Social	
SOCIAI	PSDI acknowledges the funding support by the
Join the email list   Twitter	EPSRC grants EP/X032701/1, EP/X032663/1 and EP/W/032252/1
<u>LinkedIn   YouTube</u>	

Terms of Use	Data	About us
Privacy	Services	Partners
Data protection	Communities	Support
Policy	Guidance and Training	Contact us
FAQ	News & Events	Sitemap



# Get started with PSDI

Home » Get started with PSDI

This page will contain details about the data, tools and services that PSDI provides. It will also contain information about the PSDI project, links to the guidance materials and training that we provide. There will also be content about our communities and various events and activities that PSDI runs.







Engin Physic Resea

Social

<u>Join the email list | Twitter |</u> <u>LinkedIn | YouTube</u> PSDI acknowledges the funding EPSRC grants EP/X032701/1, E and EP/W032252/1

cal Sciences support Support by the PrivacyTerms of UseDataAbout usPrivacyServicesPartnersData protectionCommunitiesSupportPolicyGuidance and TrainingContact usFAQNews & EventsSitemap	eering and		Quick Links	
Terms of UseDataAbout ussupport by the P/X032663/1PrivacyServicesPartnersData protectionCommunitiesSupportPolicyGuidance and TrainingContact usFAQNews & EventsSitemap	cal Sciences Irch Council			
Support by the P/X032663/1PrivacyServicesPartnersData protectionCommunitiesSupportPolicyGuidance and TrainingContact usFAQNews & EventsSitemap		Terms of Use	Data	About us
Data protection     Communities     Support       Policy     Guidance and Training     Contact us       FAQ     News & Events     Sitemap	P/X032663/1	Privacy	Services	Partners
PolicyGuidance and TrainingContact usFAQNews & EventsSitemap		Data protection	Communities	Support
FAQ News & Events Sitemap		Policy	Guidance and Training	Contact us
		FAQ	News & Events	Sitemap



# **All Resources**

Home » All Resources

••

..

..

••

..

••

. .

. .

. .

License3

It is intended that in the future it will be possible to search across PSDI, including advanced searching capabilities. Exactly what this will look like is in development, but this is a mock-up of how it might look with various options for filtering the search results.

New	vs Events Community	Guidance Training	Data Services	Tools Support
(a sec	arch			
□ Sho	w advanced search			
Resource type	114 resources found f	or "crystallography"		Sort by: Relevance -
Repo1 11				
Repo2 123	Resource Name	Owner	Property \$	Description
Reno3	Crystallography1	Owner1	Property1	short description
nepoo	Crystallography2	Owner2	Property2	short description
-	Crystallography3	Owner14	Property1	short description
Institutions	Crystallography4	Owner7	Property4	short description
Tag1 11	Crystallography5	Owner2	Property1	short description
Tag2	Crystallography6	Owner3	Property7	short description
Tog2	Crystallography7	Owner1	Property1	short description
lags	Crystallography8	Owner4	Property4	short description
Tag4 11	Crystallography9	Owner17	Property1	short description
Tag5 (123)	Crystallography10	Owner5	Property5	short description
····	Crystallography11	Owner9	Property16	short description
People	Crystallography12	Owner11	Property1	short description
Format1	Crystallography13	Owner8	Property12	short description
Earmat2	Crystallography14	Owner3	Property1	short description
Formatz 23	Crystallography15	Owner14	Property61	short description
Format3 11	Crystallography16	Owner12	Property1	short description
	Crystallography17	Owner1	Property1	short description
Research area	Crystallography18	Owner1	Property18	short description
License1	Crystallography19	Owner4	Property1	short description
License2	Crystallography20	Owner6	Property21	short description





Quick Links

Social

11

.....

Join the email list | Twitter | LinkedIn | YouTube

PSDI acknowledges the funding support by the EPSRC grants EP/X032701/1, EP/X032663/1 and EP/W032252/1

Terms of Use	Data	About us
Privacy	Services	Partners
Data protection	Communities	Support
Policy	Guidance and Training	Contact us



#### © PSDI | Cookies & Privacy | Website by Herd



# Your resource here

Home » Your resource here

This is a placeholder page for a 'Highlighted resource' linked to from the PSDI homepage. This resource may be a commonly used or new resource. We provide a quick link to the resource from the PSDI homepage to help the community locate it more easily.

If you have a data collection, tool, service, learning materials or training that you want to share with the physical sciences community, then we want to hear from you! Click the contact us button below.

This page will likely contain a description of the highlighted resource and information about the benefits of using it. The page will also detail who the resource is provided by and information about how it can be used or reused. A link will be provided to the resource, along with links to related resources such as guidance and training.







Social

Join the email list | Twitter | LinkedIn | YouTube

PSDI acknowledges the funding support by the EPSRC grants EP/X032701/1, EP/X032663/1 and EP/W032252/1

#### Engineering and **Physical Sciences Research Council**

#### Quick Links

Terms of Use
Privacy
Data protectior
Policy
FAQ

Data
Services
Communities
Guidance and Trainin
News & Events

About us
Partners
Support
Contact us

Sitemap



# **ELN Finder**

Home » ELN Finder

This a placeholder page for the ELN Finder tool. A description of the tool will be provided here with a link. It is also likely that we will include some links to our guidance materials providing an introduction to ELNs, considerations for choosing an ELN, transitioning from paper-based note-taking to digital, and tips for implementing ELNs.







Engin Physi Resea

Social

<u>Join the email list | Twitter |</u> <u>LinkedIn | YouTube</u> PSDI acknowledges the funding EPSRC grants EP/X032701/1, EI and EP/W032252/1

Peering and cal Sciences arch CouncilQuick LinksTerms of UseDataAbout usSupport by the P/X032663/1PrivacyServicesPartnersData protectionCommunitiesSupportPolicyGuidance and TrainingContact usFAQNews & EventsSitemap
Terms of UseDataAbout ussupport by the PrivacyPrivacyServicesPartnersData protectionCommunitiesSupportPolicyGuidance and TrainingContact usFAQNews & EventsSitemap
g support by the P/X032663/1PrivacyServicesPartnersData protectionCommunitiesSupportPolicyGuidance and TrainingContact usFAQNews & EventsSitemap
EP/X032003/1       Data protection       Communities       Support         Policy       Guidance and Training       Contact us         FAQ       News & Events       Sitemap
PolicyGuidance and TrainingContact usFAQNews & EventsSitemap
FAQ News & Events Sitemap



# Highlight placeholder page

Home » Highlight placeholder page

This is a placeholder page for a 'Highlighted resource' linked to from the PSDI homepage. This resource may be a commonly used or new resource. We provide a quick link to the resource from the PSDI homepage to help the community locate it more easily.

This page will likely contain a description of the highlighted resource and information about the benefits of using it. The page will also detail who the resource is provided by and information about how it can be used or reused. A link will be provided to the resource, along with links to related resources such as guidance and training.





Social

<u>Join the email list | Twitter |</u> <u>LinkedIn | YouTube</u> KK

Engin Physi Resea

PSDI acknowledges the funding EPSRC grants EP/X032701/1, E and EP/W032252/1

neering and ical Sciences arch Council		Quick Links	
g support by the	Terms of Use	Data	About us
EP/X032663/1	Privacy	Services	Partners
	Data protection	Communities	Support
	Policy	Guidance and Training	Contact us
	FAQ	News & Events	Sitemap



# Data to Knowledge

Home » Data to Knowledge

This is a placeholder page for a 'Demo or Exemplar' linked to from the PSDI homepage. The demo or exemplar may be a resource or case study produced by one of our pathfinder teams to show how the technologies and infrastructure provided by PSDI can help to solve a problem relevant to our community.

This page will likely contain a description of the highlighted resource or case study including information about the problem it is solving and how the solution can be used by the community. A link will be provided to the resource, along with links to related resources such as guidance and training.

The Data to Knowledge pathfinder aims to create production data infrastructure services that will enable researchers to maximise the value and quality within their data and to enable the development of future novel methods and technologies, that are currently not possible due to an absence of centralised approaches to data. It will do this by focusing on providing a technological solution to enable and simplify the capture and sharing of data from within biomolecular simulation workflows, and in a way that does not require huge cultural shifts in ways of working thus lowering the barriers to adoption. The intention is to capture the full pipeline of data from experimental input through to analytical outputs to preserve the full data provenance in how scientific studies are performed. The work here will be in partnership with and informed by the biomolecular simulation community via existing relationships under the STFC CoSeC programme in the form of CCPBioSim (scientific) and HECBioSim (HPC) consortia along with external partner institutions such as the EBI.



DATA INFRASTRUCTURE

Engir Phys Rese

Social

Join the email list | Twitter | LinkedIn | YouTube

PSDI acknowledges the fundin EPSRC grants EP/X032701/1, and EP/W032252/1

neering and ical Sciences arch Council		Quick Links	
g support by the EP/X032663/1	Terms of Use Privacy Data protection Policy FAQ	Data Services Communities Guidance and Training News & Events	About us Partners Support Contact us Sitemap



# Data Infrastructure and Tooling for Biomolecular Simulation

Home » Data Infrastructure and Tooling for Biomolecular Simulation

This is a placeholder page for a 'Demo or Exemplar' linked to from the PSDI homepage. The demo or exemplar may be a resource or case study produced by one of our pathfinder teams to show how the technologies and infrastructure provided by PSDI can help to solve a problem relevant to our community.

This page will likely contain a description of the highlighted resource or case study including information about the problem it is solving and how the solution can be used by the community. A link will be provided to the resource, along with links to related resources such as guidance and training.

The biomolecular simulation pathfinder aims to create production data infrastructure services that will enable researchers to maximise the value and quality within their data and to enable the development of future novel methods and technologies, that are currently not possible due to an absence of centralised approaches to data. It will do this by focusing on providing a technological solution to enable and simplify the capture and sharing of data from within biomolecular simulation workflows, and in a way that does not require huge cultural shifts in ways of working thus lowering the barriers to adoption. The intention is to capture the full pipeline of data from experimental input through to analytical outputs to preserve the full data provenance in how scientific studies are performed. The work here will be in partnership with and informed by the biomolecular simulation community via existing relationships under the STFC CoSeC programme in the form of CCPBioSim (scientific) and HECBioSim (HPC) consortia along with external partner institutions such as the EBI.





КК

Engin Physi Resea

Social

<u>Join the email list | Twitter |</u> <u>LinkedIn | YouTube</u> PSDI acknowledges the funding EPSRC grants EP/X032701/1, E and EP/W032252/1

neering and ical Sciences arch Council		Quick Links	
	Terms of Use	Data	About us
g support by the	Privacy	Services	Partners
EP/X032663/1	Data protection	Communities	Support
	Policy	Guidance and Training	Contact us
	FAQ	News & Events	Sitemap



# File format converter

Home » File format converter

This is a placeholder page for a 'Demo or Exemplar' linked to from the PSDI homepage. The demo or exemplar may be a resource or case study produced by one of our pathfinder teams to show how the technologies and infrastructure provided by PSDI can help to solve a problem relevant to our community.

This page will likely contain a description of the highlighted resource or case study including information about the problem it is solving and how the solution can be used by the community. A link will be provided to the resource, along with links to related resources such as guidance and training.

The file format conversion tool is one part of the Process recording pathfinder enabling users to convert files for sharing or use in across different software providers. Process recording is fundamental to driving validation, trust, reuse of data across the sciences and is a crucial aspect of data management and integration. It provides crucial support for the researcher in the laboratory whilst simultaneously structuring data for subsequent downstream management, publication, and reuse. It encompasses workflow support and data management systems, but more importantly tools for ingest that seamlessly fit into existing laboratory researcher practices and account for a range of processes from capturing handwritten observations to interfacing with instruments and recording data analysis. There are also socio-cultural aspects around disciplinary working practices that must be considered, and the support must seamlessly interface with other aspects of scholarly practice such as linking to published literature, database searching and writing reports/papers/theses. This pathfinder will investigate generic and domain specific tools for process recording and assess the landscape of semantic web technologies and metadata schemas for the physical sciences to aid with this process. This pathfinder will produce services to empower researchers to choose the best tools for them and provide exemplars and guidelines on best practices for publishing and sharing data.







Social

Join the email list | Twitter | LinkedIn | YouTube

PSDI acknowledges the funding support by the EPSRC grants EP/X032701/1, EP/X032663/1 and EP/W032252/1

#### Quick Links Engineering and **Physical Sciences Research Council** Terms of Use Data About us Services Privacy Partners Data protection Communities Support **Guidance and Training** Policy Contact us FAQ News & Events Sitemap



# Your pathfinder project

Home » Your pathfinder project

This is a placeholder page for a 'Demo or Exemplar' linked to from the PSDI homepage. The demo or exemplar may be a resource or case study produced by one of our pathfinder teams to show how the technologies and infrastructure provided by PSDI can help to solve a problem relevant to our community.

Interested in collaborating on a pathfinder project for the benefit of the physical sciences community? We want to hear from you! Click the contact us button below.

This page will likely contain a description of the highlighted resource or case study including information about the problem it is solving and how the solution can be used by the community. A link will be provided to the resource, along with links to related resources such as guidance and training.

#### Contact us





Social

Join the email list | Twitter | LinkedIn | YouTube



PSDI acknowledges the funding support by the EPSRC grants EP/X032701/1, EP/X032663/1 and EP/W032252/1

## Engineering and Physical Sciences **Research Council**

#### Quick Links

lerms of Use
Privacy
Data protection
Policy
FAQ

Data
Services
Communities
Guidance and Trai
News & Events

About us Partners Support Contact us Sitemap



# Sign up for newsletter

Home » Sign up for newsletter

This is a placeholder for 'Subscribe to our newsletter"

Contact us



Social

<u>Join the email list | Twitter |</u> <u>LinkedIn | YouTube</u> КК

Engin Physi Resea

PSDI acknowledges the funding EPSRC grants EP/X032701/1, E and EP/W032252/1

neering and ical Sciences arch Council	Quick Links		
g support by the EP/X032663/1	Terms of Use Privacy	Data Services	About us Partners
	Data protection Policy	Communities	Support Contact us
	FAQ	News & Events	Sitemap



# Contact

Home » Contact

ïrst	Las	t	
Email *			
Subject *			
bubject			
lessage *			



Social

<u>Join the email list | Twitter |</u> <u>LinkedIn | YouTube</u>



Engin Physi Resea

PSDI acknowledges the funding EPSRC grants EP/X032701/1, E and EP/W032252/1

## Get in touch

you have a question about the pilot work being undertaken for PSDI, or you ish to contribute some comments and feedback to the project, you can get contact via this contact form.

member of the team will get back to you as soon as possible.

FAQ

neering and ical Sciences		Quick Links	
arch Council			
g support by the	Terms of Use Privacy	Data Services	About us Partners
EP/X032663/1	Data protection	Communities	Support
	Policy	Guidance and Training	Contact us

News & Events

Sitemap





Home » Data

÷.

..

..

••

It is intended that in the future it will be possible to search for data across PSDI, including advanced searching capabilities. Exactly what this will look like is in development, but this is a mock-up of how it might look with various options for filtering and displaying the search results.

	Datasets Molecules Models Workflows
	<pre>asearch</pre>
	Show advanced search
Resource	510 datasets found for "benzene" 🗧 🗉 Sort by: Relevance -
Repo1	
Repo2	Benzene Dataset
Repo3	This <b>Repo1</b> record with <b>Identifier</b> contains a MS spectrum of Benzene with an
Tags	InChIkey of XXXXX. Further information from the short description provided by the record and bibliographic metadata.
Tog1	Repo Tag Tag Format License Access control

ugi	
ag2 123	
ag3 (1)	Benzene Dataset
ag4 11 ag5 123	This <b>Repo2</b> record with <b>Identifier</b> contains a MS spectrum of Benzene with an InChIkey of XXXXX. Further information from the short description provided by the record and bibliographic metadata.
Formats	Repo Tag Tag Format License Access control
Format1 (11)	
ormat2 123	Benzene Dataset
ormat3 11	This <b>Repo7</b> record with <b>Identifier</b> contains a MS spectrum of Benzene with an InChIkey of XXXXX. Further information from the short description provided by
Licenses	the record and bibliographic metadata.
icense1 (11	Repo Tag Tag Format License Access control
icense2 (123)	
icense3 1	<< <

Quick Links

 Social
 Terms of Use

 PSDI acknowledges the funding support by the
 Privacy

 Join the email list | Twitter |
 EPSRC grants EP/X032701/1, EP/X032663/1
 Data protection

 LinkedIn | YouTube
 Policy

Terms of Use	Data	About us
Privacy	Services	Partners
Data protection	Communities	Support
Policy	Guidance and Training	Contact us
FAO	News & Events	Siteman







# Services and Tools

Home » Services and Tools

It is intended that in the future it will be possible to search for tools and services across PSDI, including advanced searching capabilities. Exactly what this will look like is in development, but this is a mock-up of how it might look with various options for filtering the search results.



© PSDI | Cookies & Privacy | Website by Herd



# Guidance

Home » Guidance

search

# **Knowledge Base**

Learning by role Learning by discipline **Research data management** Research data lifecycle FAIR data Metadata & provence **Digital notebooks and ELNs** <u>Smart Labs</u> <u>Data analysis</u> **Workflows Programming** ??? ???

# Support

**Helpdesk** <u>FAQs</u> System Status Known Issues **Policies** 

Contact us

**Training** 

Skills4Scientists **Training Events** In person training **Online Training** ??? <u>???</u>

# **PSDI** Publications & Articles

#### **Digital note taking**



Kanza, Samantha. (May, 2024). The role of digital note taking for the 21st Century Scientist.

#### **Data Revival**



Frey, Jeremy, Munday, Samuel. (April, 2024). Reviving & Extracting (Old) Data.

### **Transparent research**



Pearman-Kanza, Samantha, Knight, Nicola. (April, 2024). Promoting Open & Transparent Research Practices through PSDI.

#### **Can AI replace us?**



Frey, Jeremy. (March, 2024). Will an AI win the next Chemistry Nobel Prize and replace us?





Engineering and Physical Sciences Research Council

#### Quick Links

erms of Use	Data	About us
Privacy	Services	Partners

Data protectic

Social

#### Join the email list | Twitter |

PSDI acknowledges the funding support by the EPSRC grants EP/X032701/1, EP/X032663/1 and EP/W032252/1





Support







# Knowledge Base

Home » Knowledge Base

The knowledge base will contain a range of guidance materials and documentation related to working with and sharing data. This material may include learning paths, tutorials and "how do I?" content.

Examples of topics that we might provide guidance on include:

- Learning paths by user role / or by discipline
- Introduction to data science
- Research data managment
- Data sharing
- Data citation
- Data standards
- Data analysis and data visualisation
- Jupyter notebooks
- Coding
- Semantic web and metadata

Let us know what guidance you would like to see PSDI provide by clicking the button below.

#### Contact us





Social

Join the email list | Twitter | LinkedIn | YouTube

PSDI acknowledges the funding support by the EPSRC grants EP/X032701/1, EP/X032663/1 and EP/W032252/1

Engineering and Physical Sciences Research Council

Terms of Use Privacy Data protection Policy

FAQ

Data

Services Communities **Guidance and Training** News & Events

Quick Links

About us Partners

Support

Contact us

Sitemap



# Training

Home » Training

PSDI will provide training content to enable the community to improve their skills in relation to research data management, working with data and data sharing practices. Initially materials from the Skills4Scientists courses will be available online and training associated with our Pathfinder communities and their case studies. In the future we may also provide in person training and workshops.

Examples of topics that we might provide training on include:

- Learning paths by user role / or by discipline
- Introduction to data science
- Research data managment
- Data sharing
- Data citation
- Data standards
- Data analysis and data visualisation
- Jupyter notebooks
- Coding
- Semantic web and metadata

Let us know what training you would like to see PSDI provide by clicking the button below.

#### Contact us





Social

Join the email list | Twitter | LinkedIn | YouTube

PSDI acknowledges the funding support by the EPSRC grants EP/X032701/1, EP/X032663/1 and EP/W032252/1

Engineering and Physical Sciences Research Council

Terms of Use Privacy Data protection Policy

FAQ

Services Communities **Guidance and Training** News & Events

Data

Quick Links

About us Partners

Support Contact us Sitemap



# The role of digital note taking for the 21st Century Scientist

Home » The role of digital note taking for the 21st Century Scientist

May 9, 2024

On 2 May 2024, Dr Samantha Pearman-Kanza presented at "The Chemistry Laboratory: Evaluation, Assessment and Research Symposium 2024" (<u>CLEAR 24</u>) Global Virtual Symposium. The talk was about the role of digital note taking for the 21st Century Scientist and included topics about Scientific Record Keeping, Electronic Lab Notebooks, Successful Digital Transformations and Skills4Scientists.





"The Chemistry Laboratory: Evaluation, Assessment and Research Symposium 2024" (CLEAR 24) Global Virtual Symposium 2<sup>nd</sup> May 2024 di.ac.uk/ Dr Samantha Pearman-Kanza

https://www.psdi.ac.uk/

University of Southampton



For further information the presentation can be download <u>here</u>

Click <u>**#CLEAR25</u>** for details about next year's symposium.</u>



Quick Links

Social

<u>Join the email list | Twitter |</u> <u>LinkedIn | YouTube</u> PSDI acknowledges the funding support by the EPSRC grants EP/X032701/1, EP/X032663/1 and EP/W032252/1

Terms of Use	Data	About us
Privacy	Services	Partners
Data protection	Communities	Support
Policy	Guidance and Training	Contact us
FAQ	News & Events	Sitemap





Search records...



The role of digita	l noto takina for	the 21st Co	ntury Scientist
The fole of digita	i note taking for		muly oblemist

Q



Published May 2, 2024   Version v1			Presentation C Open	72	36
The role of digital note Kanza, Samantha <sup>1</sup> 🝺	e taking for the 21st Century Sc	ientist	Show affiliations	<ul> <li>VIEWS</li> <li>Show more</li> </ul>	<b>DOWNLOADS</b>
Presentation by Samantha Pearman-Kanza Virtual Symposium. The talk was about the r	for the PSDI Project at the Chemistry Laboratory: Evaluation, A ole of digital note taking for the 21st Century Scientist and inc	Assessment and Research Sympo luded topics about Scientific Rec	sium 2024" (CLEAR 24) Global ord Keeping, Electronic Lab	Versions	
Files				Version v1 10.5281/zenodo.11163093	May 2, 2024
20240502-PSDI_CLEAR.pdf □	of 22 — + Automatic Zoom÷			<b>Cite all versions?</b> You can cite all vers 10.5281/zenodo.11163092. This DOI re always resolve to the latest one. Read r	ions by using the DOI presents all versions, and will nore.
	PHYSICAL SCIENC DATA INFRASTRU	CES		External resources Indexed in <b>OpenAIRE</b>	
Th	a role of digital note taking for the 2	Olst Contury		Communities Physical Sciences Data	Infrastructure
" https://www.psdi	The Chemistry Laboratory: Evaluation, Assessment Symposium 2024" (CLEAR 24) Global Virtual Sy 2 <sup>nd</sup> May 2024 ac.uk/ Dr Samantha Pearman-Kanza University of Southampton	t and Research		Details DOI DOI 10.5281/zenodo.11163093 Resource type Presentation Publisher	
Files (9.0 MB)			►	Zenodo Conference	
Name		Size	Download all	"The Chemistry Laboratory: Evaluation, Symposium 2024" (CLEAR 24) Global V 24), Virtual, 02/05/2024	Assessment and Research 'irtual Symposium (CLEAR
20240502-PSDI_CLEAR.pdf md5:e2c96dcdd965da847f34a1cac09495ed ?		9.0 MB	Preview Download		
Additional details				Rights	on 4.0 International
Funding PSDI Pha UK Resea Physical	se 1b EP/X032663/1 rch and Innovation Sciences Data Infrastructure Phase 1b EP/X032701/1			Citation	
UK Resea	rch and Innovation			Kanza, S. (2024, May 2). The role of dig Century Scientist. "The Chemistry Labo Assessment and Research Symposium Virtual Symposium (CLEAB 24). Virtual	ital note taking for the 21st ratory: Evaluation, 2024" (CLEAR 24) Global Zenodo
Citations ?			~	https://doi.org/10.5281/zenodo.1116309	93
Show only:   Literature (0)     Citations To This	Dataset (0) Software (0) Unknown (0) Version	Search for citation	Search	Style APA -	
No citations found				Export	
				JSON	- Export

#### Technical metadata

Created May 9, 2024

Modified May 9, 2024



About	Blog	Help	Developers	Contribute	Funded by
About Policies Infrastructure Principles Projects Roadmap Contact	Blog	FAQ Docs Guides Support	REST API OAI-PMH	<ul> <li>☑ GitHub</li> <li>☑ Donate</li> </ul>	CERN       Image: Cernical control of the



Search records...



Q



Published March 27, 2024   Vers	sion v1		Presentation Gen	86 51
Will an Al win th at FlavourTalk C	e next Chemistry Nobel Prize and r onference 2024	replace us? - Keynote	e presentation	<ul> <li>VIEWS &amp; DOWNLOADS</li> <li>Show more details</li> </ul>
Frey, Jeremy <sup>1</sup> 🕩			Show affiliations	
A Kouncto talk from Professor J	aremy Frey at the FlavourTalk 2024 Conference (The Future for Flavou	rs:AL Biotechnology and Precision Breed	ding) In this presentation	Versions
Professor Frey presents some o development in the near future.	f the technological advances that have impacted science over the las	st century and discusses how these might	t change research and	Version v1         Mar 27, 2024           10.5281/zenodo.10886524
Files				<b>Cite all versions?</b> You can cite all versions by using the DOI 10.5281/zenodo.10886523. This DOI represents all versions, and will
20240319-FreyFlavourTalk.pd	f		~	always resolve to the latest one. Read more.
🔲 🔎 🌪 Page	e: 1 of 46 - + Automatic Zoom=		23 🔒 🖸 🛛 »	
				External resources
		I AM THE FIRST AI TO RECEIVE THE		
	"Will an Al win the	PRESTIGIOUS NOBEL PRIZE	HAVE	OpenAIRE
	nevt Chemistry	BEEN POSSIBLE WITHOUT ERROR	OUND	Communities
		FTZZZZ		Physical Sciences Data Infrastructure
	Nobel Prize and			
	replace us?	TURN IT OFF AND		Details
	Jeremy G. Frey	CHECK STACKOVERFLOW	ILICK UPDATE TELDS!	DOI
	Head of Computational Systems Chemistry		EnaurySerroute con	DOI 10.5281/zenodo.10886524
	School of Chemistry, University of Southampton		CRRANTSCIENCECON	Resource type Presentation
08/03/202	24 Jeremy Frey Flavour Talk		1	<b>Publisher</b> Zenodo
Files (13.5 MB)			~	FLAVOURTALK 2024 CONFERENCE - The Future for Flavours: AI, Biotechnology, and Precision Breeding (FlavourTalk 2024), London, 19 March 2024
Name		Size	Download all	
20240319-FreyFlavourTalk.pd md5:ffffb86e041b319198a835122aec0d7	f 3 😧	13.5 MB	Preview Download	Rights
				© • Creative Commons Attribution 4.0 International
Additional details				
Funding	<b>PSDI Phase 1b</b> EP/X032663/1 UK Research and Innovation			Citation
	Physical Sciences Data Infrastructure Phase 1b EP/X032701/1 UK Research and Innovation			Frey, J. (2024, March 27). Will an AI win the next Chemistry Nobel Prize and replace us? - Keynote presentation at FlavourTalk
	Artificial and Augmented Intelligence for Automated Scientific I UK Research and Innovation	Discovery EP/S000356/1		Conference 2024. FLAVOURTALK 2024 CONFERENCE - The Future for Flavours: AI, Biotechnology, and Precision Breeding (FlavourTalk 2024), London. Zenodo. https://doi.org/10.5281/zenodo.10886524
Dates	Created			Style APA -
	2024-03-19			
				Export
Citations ?			~	JSON - Export
Show only:				

Citations To This Version			Search for citation	Search			
No citations found						Technical metadata Created March 27, 2024 Modified March 27, 2024	
		$\textcircled{\textbf{(f)}}$	Jump up				
About	Blog	Help	Developers	Contribute			Funded by
About Policies Infrastructure Principles Projects Roadmap Contact	Blog	FAQ Docs Guides Support	REST API OAI-PMH	GitHub ☑ Donate			CERN       OpenAIRE



Files

Search records...



Published April 10, 2024 | Version v1

Promoting Open & Transparent Research Practices in the Physical Sciences through PSDI

Q

Pearman-Kanza, Samantha<sup>1</sup> (D; Knight, Nicola<sup>1</sup> (D

Presentation given by Dr Samantha Pearman-Kanza and Dr Nicola Knight on behalf of PSDI for the Keele Open Research Network. Openness and transparency are among two of the most important factors to consider with respect to conducting scientific research. However, the data, tools and education available to enable these processes are often lacking. The Physical Sciences Data Infrastructure (PSDI) Initiative is conducting research and developing services to mitigate this and to provide the physical sciences community with the knowledge, exemplars and tools to promote FAIR (Findable, Accessible, Interoperable, Re-useable) research and data. This presentation covers a number of the different areas PSDI is working on. It will explain the barriers and challenges to open digital research, discuss the vital aspect of process recording (with respect to understanding what researchers actually want and how to choose the right tools for you to record your processes in the first place), and outline the considerations needed for producing fully FAIR data (and metadata), research and code. It also discusses the relevant initiatives within PSDI (training, workshops, services) that have been designed to pave the way for the scientific community to be more open and transparent with their research practices.

71 ⊛ views	26 Lownloads		
► Show	more details		
Versions			
Version v1 10.5281/zenodo.10974152	Apr 10, 2024		
<b>Cite all versions?</b> You can cite all versions by using the DOI 10.5281/zenodo.10974151. This DOI represents all versions, and will always resolve to the latest one. Bead more			

🔓 Open

Show affiliations

Presentation



	Promoting t https://www.psdi.ac.uk/	DATA INFI Open & Transparent he Physical Sciences to Keele University Open Resea 10 <sup>th</sup> April 2024 r Samantha Pearman-Kanza & University of Southar	RASTRUCTURE Research Practices brough PSDI arch Network Dr Nicola Knight mpton	in		Keywords and subjects   electronic lab notebooks   ELNs   metadata   FAIR   data   process recording   openness   transparency      Details Dol Dol 10.5281/zenodo.10974152 Resentation Publisher Zenodo
Name			Size			Conference
20240410-Keele md5:d4981efe5dfde9	le.pdf 97fd9e90c97cb7b301f <b>3</b>		8.0 MB	Pr	review Download	Keele Open Research Network , Online, 10 April 2024 Languages English
Additional d	details					Rights
Funding	<b>PSDI Phase 1b</b> EP/ UK Research and Inr	x032663/1 novation				Creative Commons Attribution 4.0 International
Dates	<b>Available</b> 2024-04-10 Presentation given or	n this date				<b>Citation</b> Pearman-Kanza, S., & Knight, N. (2024, April 10). Promoting Open & Transparent Research Practices in the Physical Sciences through PSDI. Keele Open Research Network, Online. Zenodo.
Citations 😮					~	https://doi.org/10.5281/zenodo.10974152
Show only:	<ul><li>Literature (0)</li><li>Dataset</li><li>Citations To This Version</li></ul>	(0) Software (0) Unknown	(0) Search fo	or citation	Search	Style APA •
No citations for	und					Export
						JSON - Export
						Technical metadata Created April 15, 2024 Modified April 15, 2024
		Jump	up			
About About Policies Infrastructure Principles Projects Roadmap Contact	Blog Blog	Help FAQ Docs Guides Support	Developers REST API OAI-PMH	Contribute ☑ GitHub ☑ Donate		Funded by         OpenAIRE



Search records...

Q



Published April 17, 2024   Version v1	Prese	entation Gpen	50	Δ
Reviving & Extracting (Old) Data			<ul><li>O∠</li><li>O∠</li><li>VIEWS</li><li>▶ Show model</li></ul>	DOWNLOADS
Frey, Jeremy ' (D; Munday, Samuel ' (D)		Show affiliations		
Presentation from Professor Jeremy Frey at the Institute of Process Research and Development Industrial Day scientific research; why data is important and some of the things that new technologies enable us to do with a work including PSDI and start company Data Revival.	he need for data in ne past and current	Versions		
Files			Version v1 10.5281/zenodo.11065555	Apr 17, 2024
Files (58.2 MB)	~	<b>Cite all versions?</b> You can cite all versions by using the DOI 10.5281/zenodo.11065554. This DOI represents all versions, and will always resolve to the latest one. Read more.		
Name	Size	Download all		
Leeds frey 2.pptx md5:cfe271e3ea3868d51677a4b19a4840f5 3	58.2 MB	La Download	External resources	
Citations 😮		~	<b>OpenAIRE</b>	
Show only:       Literature (0)       Dataset (0)       Software (0)       Unknown (0)         Citations To This Version	Search for citation	Search	Communities	
No citations found			Physical Sciences Da	ata Infrastructure

#### Details

#### DOI

DOI 10.5281/zenodo.11065555

#### **Resource type**

Presentation

#### Publisher

Zenodo

#### Conference

Institute of Process Research and Development, Industrial Day, University of Leeds, 17 April 2024

#### Rights

© © Creative Commons Attribution 4.0 International

#### Citation

Frey, J., & Munday, S. (2024, April 17). Reviving & Extracting (Old) Data. Institute of Process Research and Development, Industrial Day, University of Leeds. Zenodo. https://doi.org/10.5281/zenodo.11065555

Style	APA	-	٤

Export		
JSON	-	Export

Technical metadata

Created April 25, 2024 Modified April 25, 2024

#### Jump up

About	Blog	Help	Developers	Contribute	Funded by
About Policies Infrastructure Principles Projects Roadmap Contact	Blog	FAQ Docs Guides Support	REST API OAI-PMH	GitHub C Donate	CERN       Image: Cernical control of the

×



# Community

Home » Community

# Latest News



#### MADICES 2 Workshop 2024

by Victoria Hooper | May 29, 2024 | Updates

Dr Samantha Pearman-Kanza attended the 2nd Edition of the MADICES Workshop in April 2024



#### Lab Horizons Magazine Issue 3 – May 2024

by Victoria Hooper | May 24, 2024 | Updates

Dr Samantha Pearman-Kanza writes another article for the new Errant Science Lab Horizons Magazine.



#### The role of digital note taking for the 21st Century Scientist

by Victoria Hooper | May 9, 2024 | Updates

On 2 May 2024 Dr Samantha Pearman-Kanza gave a presentation on behalf of PSDI at #CLEAR24

« Older Entries

# **Events**

### **Featured event**

#### 20

June

# **PSDI Townhall**

📩 **Date** 20th June 2024

**• Time** 10:00 am - 4:30 pm

**Venue** IOM3

**Organiser** PSDI

Price Free

The Physical Sciences Data Infrastructure (PSDI) is running a Town hall event on 20th June and would like to invite representatives from across the Physical Sciences and related domains to attend. At this Townhall event we will be showcasing some of the development acti...

<u>More Info</u>

#### **Upcoming events**

June 2024

#### **PSDI Townhall**

20th June 2024 10:00 am Free

<u>More Info</u>

#### **Previous events**

June 2023

#### Webinar: Introduction to PSDI

29th June 2023 2:00 pm Free

<u>More Info</u>

#### July 2023

#### Webinar: Pathfinders - Process Recording

27th July 2023 2:00 pm Free



September 2023

#### Machine Learning for Atomistic Modelling Autumn School 2023

Sep 18 12:00 pm £100

<u>More Info</u>

# **Explore our communities**



# **Our Collaborators**

DCC



**Catalysis Hub** 



CCDC





#### Henry Royce Institute

The Catalysis Hub aims to establish a world-leading, comprehensive and coordinated programme of catalytic science in the UK. The Cambridge Crystallographic Data Centre (CCDC) is the home of the Cambridge Structural Database (CSD). The Digital Curation Centre (DCC) is a world-leading centre of expertise in digital information curation with a focus on building capacity, capability and skills for research data management. The Henry Royce Institute is the UK's national institute for advanced materials research and innovation working with the UK materials community to develop solutions to national and global challenges.

# **Get Involved**





# **News and Updates**

#### Home » News and Updates

# **Updates**



### MADICES 2 Workshop 2024

May 29, 2024

Dr Samantha Pearman-Kanza attended the 2nd Edition of the MADICES Workshop in April 2024



#### Lab Horizons Magazine Issue 3 – May 2024

May 24, 2024

Dr Samantha Pearman-Kanza writes another article for the new Errant Science Lab Horizons Magazine.



# The role of digital note taking for the 21st Century Scientist

#### May 9, 2024

On 2 May 2024 Dr Samantha Pearman-Kanza gave a presentation on behalf of PSDI at #CLEAR24







#### **Promoting Open & Transparent Research Practices in the Physical Sciences through PSDI**

#### Apr 22, 2024

On 10 April 2024 Dr Samantha Pearman-Kanza and Dr Nicola Knight gave a presentation on behalf of PSDI for the Keele Open Research Network.



#### Towards data sharing service for Physical Sciences Data Infrastructure

Mar 27, 2024

Members of the PSDI technology team (Jonathan Bathe and Vasily Bunakov) presented recently in CS3 2024.



#### The Development of the Chemists Notebook

Mar 20, 2024

Dr Samantha Pearman-Kanza gives talk at The Royal Society of Chemistry – Historical Group event.

# Data Revival: webinar recording available

Mar 12, 2024

For the fifth PSDI webinar, Samuel Munday presented about extraction of knowledge from lab books.

#### Two Articles by Samantha for Lab Horizons Magazine in February 2024

Mar 6, 2024

Dr Samantha Pearman-Kanza writes her second article for the new Errant Science Lab Horizons Magazine.

#### Love Data Week 2024

Feb 16, 2024 International Love Data Week 2024 #LoveData24

**Next Entries** 

# Newsletters

#### **PSDI Updates: March 2024**

Mar 28, 2024

Check out our March 2024 newsletter for updates on upcoming PSDI talks and links to past webinar recordings.

#### **Upcoming webinar 28th March**

Mar 14, 2024

In this newsletter you can find information about our upcoming webinar on Ampletracks and links to the recordings of our previous webinars

#### **PSDI Updates: November 2023**

Nov 14, 2023

Check out our November 2023 newsletter for updates on upcoming PSDI talks and links to past webinar recordings.

#### **PSDI Updates: September 2023**

Sep 25, 2023 Check out our September 2023 newsletter for

#### **PSDI Updates: August 2023**

Aug 11, 2023 Check out our August 2023 newsletter for

# **PSDI Updates: Webinars July 2023**

Jul 19, 2023

updates on upcoming PSDI talks and links to past webinar recordings.

updates on upcoming PSDI talks and links to past webinar recordings.

Welcome to our July 2023 PSDI update! Discover upcoming events and last month's webinar recordings in this newsletter.

#### **PSDI Updates: June 2023**

Jun 1, 2023 Welcome to our June 2023 PSDI update! Explore upcoming events and our outputs from the pilot phase in this newsletter.

Quick Links

Social

<u>Join the email list | Twitter |</u> <u>LinkedIn | YouTube</u> PSDI acknowledges the funding support by the EPSRC grants EP/X032701/1, EP/X032663/1 and EP/W032252/1

Terms of Use	Data	About us
Privacy	Services	Partners
Data protection	Communities	Support
Policy	Guidance and Training	Contact us
FAQ	News & Events	Sitemap



# MADICES 2 Workshop 2024

Home » MADICES 2 Workshop 2024

May 29, 2024

At the end of April 2024, Dr Samantha Pearman-Kanza attended the 2nd Edition of the MADICES Workshop (Machine-actionable data interoperability for the chemical sciences) in snowy Berlin, on behalf of PSDI.



This was a really interesting workshop that explored many of the research areas that PSDI is interested in, including metadata, interoperability, data sharing and processing and the use of semantic web technologies in producing FAIR data. The workshop was a mix of discussion and hackathon style activities, and a number of datasets were made available in advance for use in the hackathon.

As a result of several pre-planning calls, and a pre setup GitHub Repository, workshop members had been able to suggest discussion topics and areas of interest, and these were formalised into three main working groups on day 1:

1. Semantic Annotation Standards & Tools – This looked at identifying best practices for semantically annotating and structuring data and the range of tools available to automate this process. GitHub: https://github.com/MADICES/MADICES-2024/discussions/9

 Platform Communication – This looked at how different platforms should communicate requests, what these should look like and what data formats should be used. GitHub: https://github.com/MADICES/MADICES-2024/discussions/10
 Proprietary Dataset Processing – This looked at the different methods for processing proprietary datasets. GitHub: https://github.com/MADICES/MADICES-2024/discussions/11

Despite all three areas being of interest, naturally Samantha took part in the Semantic Web Annotation Standards & Tools Group. The group started by conducting a knowledge sharing exercise, which was really interesting as it exposed the range of tools and techniques that different group members were and weren't aware of, and these were documented on the GitHub: <u>https://github.com/MADICES/MADICES-2024/discussions/9</u>.

The main interests of the group centred around a few key topics, namely generating semantic metadata, how to create JSON-LD files to do this, and the different ways of representing some of the scientific information semantically, e.g. Units. These were formalised into some specific tasks for the hackathon to gather information about unit implementations and create a package on PyPi, to create an exemplar of a semantic metadata record in JSON-LD from a pre-existing metadata record, and to write a simple automation script to identify which areas were simple/complex to automate. The results of these activities can be found in this repository: <u>https://github.com/MADICES/semantic-annotation</u>.

Undertaking these activities raised some really important points. With respect to units, which at first glance feels like it should be a simple area to work in, there are actually a range of different ontologies and methods to represent units and measurements semantically, and they can prove quite complex to get your head around. Further, it can also be quite tricky to ensure that you have selected the appropriate unit to use with your data.

With respect to semantic annotation, this exercise highlighted just how important it was to have the combination of semantic knowledge and domain knowledge. Firstly, to ensure that people creating the semantic annotations can actually understand what the data is about, and to correctly identify the required ontology terms, as many of the same terms are used across chemistry and biology to mean slightly different things, and it's important to ensure that the selected classes actually have the same meaning as the data represented in the datasets. It's all about semantics, literally! An example of the JSON-LD that was created can be found in this JSON-LD Playground Link, and the test data and conversion scripts can be found on the GitHub repository.

This was a highly pertinent workshop for PSDI as it aligns with many of our research interests and goals for the future. We are actively looking into automated metadata generation, best practices for metadata, and how to usefully implement ontologies and linked data within the physical sciences to improve interoperability.

The full conference program and participants list can be found here: <u>https://www.cecam.org/workshop-details/machine-</u> actionable-data-interoperability-for-the-chemical-sciences-madices-2-1321

#### See all news

#### Quick Links

So	cia	

Join the email list | Twitter | LinkedIn | YouTube PSDI acknowledges the funding support by the EPSRC grants EP/X032701/1, EP/X032663/1 and EP/W032252/1

Terms of Use	Data	About us
Privacy	Services	Partners
Data protection	Communities	Support
Policy	Guidance and Training	Contact us
FAQ	News & Events	Sitemap



# Lab Horizons Magazine Issue 3 – May 2024

Home » Lab Horizons Magazine Issue 3 – May 2024

May 24, 2024

Issue 3 of the Lab Horizons magazine was published in May 2024 and Dr Samantha Pearman-Kanza has written another article for them titled "The Evolution of the Scientists Notebook. One step forward or two steps back?" and Samantha looks at where the labnote book came from. To have a read check out page 10: Lab Horizons Issue 3 – Digital Edition.



Keep an eye out for what's in store later this year!

If you have any ideas on important areas you would like to see featured feel free to get in touch.

We need to think about how to capture these more personal aspects of scientists' thinking

#### See all news

Quick Links

Social

<u>Join the email list | Twitter |</u> <u>LinkedIn | YouTube</u> PSDI acknowledges the funding support by the EPSRC grants EP/X032701/1, EP/X032663/1 and EP/W032252/1

Terms of Use	Data	About us
Privacy	Services	Partners
Data protection	Communities	Support
Policy	Guidance and Training	Contact us
FAQ	News & Events	Sitemap



Q Search for e	events
🗙 🔪 Today	Upcoming ~
June 2024	
THU <b>20</b>	Featured 20 June @ 10:00 am - 4:30 pm BST <b>DSDI TOWNALS Automatical Structure of Materials</b> , Minerals and Mining, 297 Euston Road, London The Physical Sciences Data Infrastructure (PSDI) is running a Town hall event on 20th like to invite representatives from across the Physical Sciences and related domains to Townhall event we will be showcasing some of the development activities that are cure PSDI, gathering feedback from the community and providing information on how to g with PSDI activities. Free

**<** Previous Events



<u>Join the email list | Twitter |</u> <u>LinkedIn | YouTube</u> PSDI acknowledges the funding EPSRC grants EP/X032701/1, E and EP/W032252/1



th June and would to attend. At this urrently going on in get more involved

Next Events >

Subscribe to calendar 💉

Detering and ical Sciences arch CouncilQuick LinksImage support by the EP/X032663/1Terms of UseDataAbout usPrivacyDataAbout usData protectionCommunitiesPartnersData protectionCommunitiesSupportPolicyGuidance and TrainingContact usFAQNews & EventsSitemap				
Terms of UseDataAbout usBesupport by the EP/X032663/1PrivacyServicesPartnersData protectionCommunitiesSupportPolicyGuidance and TrainingContact usFAQNews & EventsSitemap	neering and ical Sciences arch Council		Quick Links	
PrivacyServicesPartnersEP/X032663/1Data protectionCommunitiesSupportPolicyGuidance and TrainingContact usFAQNews & EventsSitemap	a cupport by the	Terms of Use	Data	About us
Data protection       Communities       Support         Policy       Guidance and Training       Contact us         FAQ       News & Events       Sitemap	g support by the FP/X032663/1	Privacy	Services	Partners
PolicyGuidance and TrainingContact usFAQNews & EventsSitemap		Data protection	Communities	Support
FAQ News & Events Sitemap		Policy	Guidance and Training	Contact us
		FAQ	News & Events	Sitemap



« All Events

# **PSDI Townhall**

## 20 June @ 10:00 am - 4:30 pm BST FREE

The Physical Sciences Data Infrastructure (PSDI) is running a Townhall event on 20th June and would like to invite representatives from across the Physical Sciences and related domains to attend. At this Townhall event we will be showcasing some of the development activities that are currently going on in PSDI, gathering feedback from the community and providing information on how to get more involved with PSDI activities.

#### Book your place at the Townhall (Tickettailor link)

- The provisional agenda for the day is the following: 10:00 – Start 10:00 – 10:25 Registration & Coffee 10:25 – 10:30 Housekeeping 10:30 – 10:40 Introduction 10:40 – 11:15 PSDI Overview 11:15 – 11:55 Demonstrators 11:55 – 12:15 Invited Talk 12:15 – 12:25 Logistics about Exhibition 12:25 – 13:55 Lunch + Exhibition
- 13:55 14:35 Contributed Lightning Talks

14:35 - 15:25 Interactive Feedback & Discussion session
15:25 - 15:40 Closing - next steps
15:40 - 16:30 Coffee + informal networking

# **Register for the event**

Please book your in person place by "purchasing" a ticket **through ticket tailor**. There is no charge, this enables us to manage our attendee list and cater to any dietary requirements.

#### Register now

This event will be run as an in person event only, however, there will be the opportunity to see the event materials and provide feedback alongside the Townhall meeting. If you are unable to attend the meeting but want to hear more about the event please make sure you are signed up to our PSDI mailing list (https://www.jiscmail.ac.uk/cgi-bin/wa-jisc.exe?SUBED1=PSDI&A=1)

 $\blacksquare$  Add to calendar  $\checkmark$ 

## **DETAILS ORGANISER VENUE**

Date:	PSDI	IOM3
20 June	View Organizar Mahaita	Institute of Materials,
Time:	view Organiser website	Minerals and Mining, 297
10.00 am 4.20 nm PST		Euston Road
10.00 am - 4.30 pm 631		London, NW1 3AD +
Cost:		Google Map
Free		View Venue Website

Website: https://buytickets.at/physic alsciencesdatainfrastructur e/1246138

**<** Webinar: Using the Galaxy Platform in Large Scale Experiments

Quick Links

us

Social		Terms of Use	Data	About u
	PSDI acknowledges the funding support by the	Privacy	Services	Partner
	EPSRC grants EP/X032701/1, EP/X032663/1	Data protection	Communities	Support
<u>LinkedIn   YouTube</u>		Policy	Guidance and Training	Contact
		FAQ	News & Events	Sitemap



Back to all Events





# Webinar: Pathfinders – Process Recording

In the <u>first webinar of the series</u> (a recording of this now available on the <u>PSDI YouTube page</u>) we took a look at PSDI as a whole. In this second webinar we focus down into work package 4 of the current phase, which contains our 5 pathfinders, in this webinar we looked at the work in Pathfinder 2: Process Recording, which is led by Dr Samantha Kanza.

The recording of this webinar is now available **on Youtube**.

Abstract: Process recording is fundamental to ensuring the preservation of the scientific record and enabling reproducibility. The Electronic Lab Notebook (ELN) was originally created to serve as a direct replacement for the paper lab notebook, to ensure the digital capture and retention of the scientific record. What seemed like an obvious simple software-based solution has been anything but, and there are now over 80 active ELNs on the market and a wide range of different approaches have been taken to digitizing scientific research across academia and industry.

This talk will discuss the shift in software offerings and attitudes to process recording software and report on the results of a recent survey on ELN and Notebook Usage in our physical sciences community. We will also discuss the current work that is being undertaken in this pathfinder to produce a comprehensive resource on process recording tools to aid researchers in identifying the best solution(s) for them. It will also cover research being undertaken around the different aspects of process recording, and community recommendations for making

improvements in these areas, such as metadata, semantics, data standards and the use of hybrid/voice technologies.



Bio: Dr Samantha Kanza is a Senior Enterprise Fellow at the University of Southampton. She is a Pathfinder Lead and researcher in the Physical Sciences Data Infrastructure initiative. She also coordinates the Future Blood Testing Network (www.futurebloodtesting.org) run out of Reading as well as previously co-ordinating the AI 4 Scientific Discovery Network (AI4SD – www.ai4science.network). Samantha works in the interdisciplinary research area of applying computer science techniques to the scientific domain, specifically through the use of semantic web technologies and artificial intelligence. Her research includes looking at electronic lab notebooks and smart laboratories, to improve the digitization and knowledge management of the scientific record using semantic web technologies; and using IoT devices in the laboratory. She has also worked on a number of interdisciplinary Semantic Web projects in different domains, including agriculture, chemistry and the social sciences.

## Watch the recording

You can watch the recording of this webinar via our youtube channel.

The pathfinders aim to focus development in strategic areas with the aim to explore and establish exemplar systems that can be brought under the PSDI umbrella and act as templates for future phases of the project. An overview of the pathfinder will be presented as well as their goals over the course of the project. For those that wish to get some more information about the pathfinders before the webinar please go to: <u>https://psdi.ac.uk/current-work/wp4-pathfinders/</u>

The PSDI team looks forward to seeing you at the webinar, if you have any questions you can always get in <u>contact</u> <u>with us</u>.

20 June

#### **PSDI Townhall**

Date 20th June 2024 Time 10:00 am - 4:30 pm Venue IOM3 Price Free



Quick Links

Social	PSDI acknowledges the funding support by the	Terms of Use Privacy	Data Services	About us Partners
Join the email list   Twitter	EPSRC grants EP/X032701/1, EP/X032663/1 and	Data protection	Communities	Support
<u>LinkedIn   YouTube</u>		Policy	Guidance and Training	Contact us
		FAQ	News & Events	Sitemap





Back to all Events





**Registration Link** 

# Webinar: Introduction to PSDI

Welcome to the Physical Sciences Data Infrastructure (PSDI) webinar series. This webinar series is designed to communicate the PSDI work to a wider audience!

In the first webinar of the series we will be presenting a broad introduction to PSDI. If you are wondering what PSDI is working on, then this webinar will be an excellent introduction for you. In this webinar we will cover the vision of the project, an overview of the work we have undertaken in the PSDI pilot, the work we are currently undertaking and where we are focusing in the shorter term. Future webinars will focus in more depth on specific elements of the project. Our **second webinar** running in July will focus more on work undertaken in work package 4 through our pathfinders.

This webinar will be a presentation from the PSDI team with a chance for those attending to ask questions at the end, the webinar will be recorded and a <u>copy of the recording</u> is now available on the <u>PSDI Youtube</u> <u>Channel</u> – you can subscribe to the channel to find out when future videos go live.

The PSDI team looks forward to seeing you at the webinar, if you have any questions you can always get in **contact with us**.

## Watch the recording

You can watch the recording of this webinar via our Youtube Channel.

#### 20 June

#### **PSDI Townhall**

Date 20th June 2024 Time 10:00 am - 4:30 pm Venue IOM3 Price Free



			Quick Links	
Social	PSDI acknowledges the funding support by the	Terms of Use	Data	About us
Join the email list   Twitter	EPSRC grants EP/X032701/1, EP/X032663/1 and EP/W032252/1	Data protection	Communities	Support
<u>Linkedin</u>   <u>You lube</u>		FAQ	News & Events	Sitemap



Back to all Events





Date 18th September 2023 - 20th September 2023

> Time 12:00 pm - 2:00 pm

Venue	

# **Machine Learning for Atomistic Modelling Autumn School 2023**

All places for the school have now been finalised, for those attending please ensure your payment is made by 1st September.

- Date of Event: 18th 12:00 20th September 14:00
- Location: Daresbury Laboratory, in person event
- Fee: £100 (covers 2 nights accommodation and catering)
- Pre-requisites: Students will be expected to bring their own laptop, to have a decent level of coding experience (see pre-requisites below) and provide a letter of support from their supervisor

Daresbury Laboratory Location Keckwick Lane, Daresbury, WA4 4AD

> Price £100

## Description

This machine learning for materials training course is being run by the Physical Sciences Data Infrastructure (PSDI) initiative in collaboration with <u>PSDS</u>, <u>AI4SD</u>, <u>STFC-SCD</u> and <u>CCP5</u>. This training is targeted towards PhD students, in particular those in the Materials and Molecular Simulations field. The aim of this training is to introduce attendees to the latest methods of machine learning applied to atomistic simulation of materials.

This training will encompass a number of talks and practical sessions, focusing on the basics of machine learning, machine learning interatomic potentials and graph neural networks. There will also be the opportunity for attendees to present a poster on their work.

## Learning outcomes

- Awareness of the state-of-the-art methods for machine learning for atomic and molecular simulations
- Hands on experience of using machine learning for atomic and molecular simulations

## **Outline Agenda – Draft**

## Day 1 – 18th

- 12:00 13:00: Registration & Lunch
- 13:00 13:30: Introduction
- 13:30 15:00: Lecture Session: Basic introduction to ML topics – Reinhard Maurer
- 15:00 15:30: Coffee Break
- 15:30 17:15: Practical session: Basic ML worked example – Reinhard Maurer
- 17:30 18:30: Research Seminar – Aron Walsh
- 19:00: BBQ

## Day 2 – 19th

#### • 09:00 – 10:30: Lectures (1h30) Machine Learning Interatomic Potentials – Ioan Magdau

- 10:30 11:00: Coffee
- 11:00 12:30: Practical Session MLIP (1h30) – Ioan Magdau + Alin-Marin Elena
- 12:30 14:00: Lunch
- 14:00 15:30: Practical Session MLIP (1h30) – Ioan Magdau + Alin-Marin Elena
- 15:30 16:00: Coffee
- 16:00 18:00: Lectures: GNN talks - Keith Butler + Alex Ganose
- Poster session evening +

## **Day 3 – 20th**

- 09:00 10:30: Practical Session: Building and training GNN – Keith Butler & Alex Ganose
- 10:30 11:00: Coffee
- 11:00 12:30: Practical Session: Using pre trained networks - Keith Butler & Alex Ganose
- 12:30 14:00: Lunch & Departure



## **Pre-requisites**

Students attending this course must already have a foundational level of Python experience and hands on experience of using Python in their research. You will be expected to provide your own laptop for the training course, although software installation will not be required. A letter of support will be required from your supervisor alongside your application. This letter of support is to show the backing of your supervisor to attend the training and must be completed on headed paper, but does not need to be detailed. A template of the minimum required content is available in this word document.

## Timelines

The application deadline is **30th June 2023** (including letter of support from your supervisor). Applications are now closed. You will be informed of the outcome of your application on 1st August, you will have to accept your place within 1 week and payment is required by 1st September. These timelines have been amended due to the exceptionally high number of applications we received.

Food and 2 nights accommodation is included in the £100 fee paid for this event, travel to Daresbury is not included and will need to be covered by the attendee. Please note: places on this course are limited and in the event of oversubscription to the training course we will favour a diverse group of attendees.

All places for the school have now been finalised, for those attending please ensure your payment is made by 1st September.

## **Organising Committee**

- Alin-Marin Elena, Scientific Computing Department STFC
- Keith Butler, Queen Mary University London
- Reinhard Maurer, University of Warwick
- Kim Jelfs, Imperial College London
- Alex Ganose, Imperial College London
- Simon Coles, University of Southampton
- Samantha Kanza, University of Southampton
- Nicola Knight, University of Southampton

#### 20 June

#### **PSDI Townhall**

**Date** 20th June 2024 **Time** 10:00 am - 4:30 pm Venue IOM3 Price Free

#### <u>More Info</u>

Quick Links

Social

Join the email list | Twitter | LinkedIn | YouTube

PSDI acknowledges the funding support by the EPSRC grants EP/X032701/1, EP/X032663/1 and EP/W032252/1

Terms of Use	Data	About us
Privacy	Services	Partners
Data protection	Communities	Support
Policy	Guidance and Training	Contact us
FAQ	News & Events	Sitemap



# **Collaborator Profile: CatalysisHub**

Home » Collaborator Profile: CatalysisHub

# **Catalysis Hub**

The Catalysis Hub was created with EPSRC funding with the aim to first establish a world-leading, comprehensive and coordinated programme of catalytic science in the UK; secondly to develop new knowledge and promote innovation in and translation of catalytic science and technology; and thirdly to enable the UK to regain and retain its world leading position in catalysis. The Hub was launched in April 2013 at its physical centre in the Research Complex at Harwell with a program of initial scientific projects starting from July 2013.





**PSD** PHYSICAL SCIENCES DATA INFRASTRUCTURE

Social

Join the email list | Twitter | LinkedIn | YouTube



Engineering and Physical Sciences Research Council

PSDI acknowledges the funding support by the EPSRC grants EP/X032701/1, EP/X032663/1 and EP/W032252/1



Quick Links

Terms of Use Privacy Data protection Policy FAQ Data
Services
Communities
Guidance and Training
News & Events

About us Partners Support Contact us

Sitemap



# **Collaborator Profile: CCDC**

Home » Collaborator Profile: CCDC

# **The Cambridge Crystallographic Data Centre**

The Cambridge Crystallographic Data Centre (CCDC) is the home of the Cambridge Structural Database (CSD). Established almost 60 years ago, the CSD contains data for over 1.25 million experimental crystal structure determinations, the majority of which today are deposited directly with the CCDC by researchers worldwide. The CCDC develops scientific software and services that enable discovery and reuse of data and knowledge derived from the CSD in support of both industrial and academic research. The Centre also engages directly in research and undertakes a range of education and outreach activities.



## **CCDC and PSDI**

Critical to the success of data infrastructure in the physical sciences is availability of trustworthy FAIR data collections that can be readily accessed for use in AI and other data-driven research activities. The CCDC will be working with other partners to establish the policies and best practice necessary to ensure that data resources made available through PSDI can be trusted and are AI-ready. We will also be contributing to consideration of licensing models that can be adopted to enable access to highly curated physical sciences data by the UK academic community. Building on this we will support the specification and implementation of federated data services that allow data from diverse data collections to be combined and aggregated in pursuit of research goals. Finally, we will together support activities aimed at establishing data standards critical to data interoperability and reuse, in particular in the area of Crystal Structure Prediction.





Engir Physi Resea

Social

Join the email list | Twitter | LinkedIn | YouTube

PSDI acknowledges the fundin EPSRC grants EP/X032701/1, and EP/W032252/1

# advancing structural science

neering and Ical Sciences arch Council		Quick Links	
g support by the	Terms of Use Privacy	Data Services	About us Partners
EP/X032663/1	Data protection Policy	Communities Guidance and Training	Support Contact us
	FAQ	News & Events	Sitemap



# **Collaborator Profile: DCC**

Home » Collaborator Profile: DCC

## The Digital Curation Centre

The Digital Curation Centre (DCC) is a world-leading centre of expertise in digital information curation with a focus on building capacity, capability and skills for research data management.

The DCC provides expert advice and practical help on how to store, manage, protect and share digital research data. We provide a broad range of resources including online tools, guidance and training. We also provides consultancy services on issues such as policy development and data management planning.

Services are targeted primarily at the higher education community, both in the UK and internationally, but our resources are of benefit to the commercial sector too.



DATA INFRASTRUCTURE

Social

Join the email list | Twitter | LinkedIn | YouTube



PSDI acknowledges the funding support by the EPSRC grants EP/X032701/1, EP/X032663/1 and EP/W032252/1

# DCC

## Engineering and Physical Sciences **Research Council**

#### Quick Links

Terms of Use Privacy Data protection Policy FAQ

Data Services Communities **Guidance and Training** News & Events

About us Partners Support Contact us

Sitemap



# **Collaborator Profile: The Henry Royce Institute**

Home » Collaborator Profile: The Henry Royce Institute

# The Henry Royce Institute

The Henry Royce Institute is the UK's national institute for advanced materials research and innovation, supporting world leading excellence in UK materials research. Having invested over £330 million in world-class equipment, infrastructure, training and outreach across 11 institutions, Royce works with the UK materials community to develop solutions to national and global challenges.



## The Royce Discovery Centre

The Royce Discovery Centre, based at the University of Sheffield City Centre Campus, focuses on early-stage research on materials discovery and processing. The facility takes materials and processing concepts, and develops these from basic principles through analytical and experimental processes, with the aim of proving the concept in terms of feasibility and applicability for industrial use. Housing state-of-the-art specialist laboratories, workshops, and office spaces, the Royce Discovery Centre draws on expertise present within the Royce team and the Department of Materials Science and Engineering, to provide real solutions in the discovery and creation of innovative material systems.

## **The Royce Pathfinder Project** Title: Data-capture for advanced metals processing

The Henry Royce Institute will lead a Pathfinder project on developing data infrastructure for exploring Processing-Structure-Property relationships (P-S-P) in metallurgy. The project will have three objectives: (i) develop ontology / metadata schema for describing metals processing; (ii) develop microstructural descriptors and software tools to digitally fingerprint microscopy images; (iii) develop and install infrastructure (physical and digital) for capturing and storing P-S-P information (digital thread). It will build on work in previous PSDI pathfinders and adapt some of these learnings to the particularly challenging metallurgical space, where the processing history strongly affects microstructure and performance and where considerable redundancy and inefficiency may exist, and current data capture is often unstructured.

The pathfinder will be led through the Royce Discovery Centre in Sheffield and will make use of their extensive additive manufacturing facilities, which provide a data-rich platform in which to develop the capabilities that form part of this Pathfinder with a view to generalising them for deployment in more traditional metallurgical applications, such as casting and forging.



Social

Join the email list | Twitter | LinkedIn | YouTube



Engin Physi Resea

PSDI acknowledges the funding EPSRC grants EP/X032701/1, E and EP/W032252/1

# University of Sheffield



DescriptionQuick LinksCouncilCuick LinksPrivacyDataPrivacyDataData protectionServicesData protectionCommunitiesDiloyGuidance and TrainingFAQNews & Events				
Support by the g support by the P/X032663/1Terms of UseDataAbout usPrivacyServicesPartnersData protectionCommunitiesSupportPolicyGuidance and TrainingContact usFAQNews & EventsSitemap	neering and ical Sciences arch Council		Quick Links	
Privacy     Services     Partners       EP/X032663/1     Data protection     Communities     Support       Policy     Guidance and Training     Contact us       FAQ     News & Events     Sitemap	g support by the	Terms of Use	Data	About us
Data protectionCommunitiesSupportPolicyGuidance and TrainingContact usFAQNews & EventsSitemap	EP/X032663/1	Privacy	Services	Partners
PolicyGuidance and TrainingContact usFAQNews & EventsSitemap		Data protection	Communities	Support
FAQ News & Events Sitemap		Policy	Guidance and Training	Contact us
		FAQ	News & Events	Sitemap



# **Community placeholder**

Home » Community placeholder

This is a placeholder for 'Join a community' and individual 'Communities'

Our community engagement activities are currently under development, but we anticipate that we will form communities of interest around different disciplines, domains and technologies. These communities can share their expertise and work together to develop standards, tools, case studies, and guidance that can be shared both within the specialist and wider PSDI communities.

Let us know what communities you would like to see represented and what activities those communities could be engaged with by clicking the button below.

#### Contact us



DATA INFRASTRUCTURE

Social

Join the email list | Twitter | LinkedIn | YouTube

PSDI acknowledges the funding support by the EPSRC grants EP/X032701/1, EP/X032663/1 and EP/W032252/1

# Engineering and Physical Sciences Research Council

#### Quick Links

Terms of Use
Privacy
Data protection
Policy
FAQ

Data
Services
Communities
Guidance and Training
News & Events

About us
Partners
Support
Contact us

Sitemap



# **Contribute to the community placeholder**

Home » Contribute to the community placeholder

Community engagement activities are currently under development but we hope to encourage users to get involved in the PSDI community by submitting data collections, developing services or tools, and sharing case studies using our infrastructure.

Let us know what activities you think our communities should be doing by clicking the button below.

Contact us



Social

Join the email list | Twitter | LinkedIn | YouTube



PSDI acknowledges the funding support by the EPSRC grants EP/X032701/1, EP/X032663/1 and EP/W032252/1

## Engineering and Physical Sciences **Research Council**

Quick Links

Terms of Use Data Services Privacy Data protection Communities Policy **Guidance and Training** FAQ News & Events

About us Partners Support Contact us Sitemap



# About PSDI

Home » About PSDI

# The vision of PSDI

The data needs for research are growing at previously unimaginable rates and the need for collaboration around data has never been clearer. Data cannot be considered as simply an output of research, as it is itself a driver of further discovery. Experiments, observations, computations and simulations all generate data and data flows form the very fabric of research in and across the physical sciences. But, for the most part, each physical science research infrastructure, from laboratory to large facility, has essentially its own data infrastructure.

Whilst centralised, data-centric infrastructures for collecting and reusing data can act as community hubs and drive new methods and discoveries, the current diversity of data infrastructures enables each platform to be tailored to the specific needs of its field. PSDI, therefore aims to provide an additional layer of infrastructure that enables sharing of existing resources whilst ensuring that each can remain dedicated to its specific application.



Data Services and Tools Guidance Community About PSDI



# **Our objectives**

The aim of PSDI is to enable researchers in the physical sciences to handle data more easily by connecting the different data infrastructures they use. PSDI will connect and enhance existing infrastructure in Physical Sciences.

Through PSDI researchers will be able to:

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

# **Our team**



PSDI is composed of collaborators across multiple institutions, departments and geographic sites in the United Kingdom. The core of the team are based at the Rutherford Appleton & Daresbury Laboratory sites of STFC (Science and Technology Facilities Council) and the University of Southampton with the assistance and support of a number of partner organisations. Find out more about the **PSDI team and our collaborators**.

# Our work

The current phase of PSDI, Phase 2, which started in January 2024 continues to build upon the work carried out in The Pilot and Phase 1 of the project.

This phase of PSDI begins the second half of the initial 30-month development (started in Phase 1). This will lead to the launch of version 1 of PSDI in March 2025. PSDI will then continue in a dual mode, maintaining the operational services of version 1 and continued development to provide further services. Our work is broken down into a number of different work packages with additional contributions in the form of pathfinders. Pathfinders are a means to focus some development activity in key strategic areas and explored and established exemplar approaches and systems that can be folded into the PSDI infrastructure and act as a template or starting point to bring in further domains, data types, techniques and user communities in future phases. You can read more about the structure of our **work packages** and our current **pathfinders**.

The PSDI initiative is funded through EPSRC DRI funding (EP/X032701/1 and EP/X032663/1).

# Get involved



Quick Links

About us

Partners

Support

Contact us

Sitemap

#### Social Terms of Use Data PSDI acknowledges the funding support by the Privacy Services EPSRC grants EP/X032701/1, EP/X032663/1 Join the email list | Twitter | Data protection Communities and EP/W032252/1 LinkedIn | YouTube **Guidance and Training** Policy FAQ News & Events





Home » People

# **PSDI Executive team**



Juan Bicarregui Project PI

Juan Bicarregui is Head of the Data Division in the Scientific Computing Department at STFC. Juan heads up the investigator team ...



# Barbara Montanari

Work Package 5 Lead Barbara Montanari is Head of Computational Science and Engineering Division at STFC and Director of the Computational Science Centre ...



# **Simon Coles**

Work Package 4 Lead Simon Coles is Professor of Structural Chemistry and Director of both the UK National Crystallography Service and the UK Physical ...



## **Nicola Knight** Project Coordinator / Work

Package 2 Lead Nicola Knight is a research fellow at the University of Southampton and will be coordinating activities across the pilot project and ...







# Jeremy Frey

Work Package 1 Lead / Pathfinder 3 Lead Jeremy Frey is Professor of Physical Chemistry and Head of Computational Systems Chemistry at University of Southampton. Jeremy takes ...

# Vasily Bunakov

Work Package 3 Lead Vasily is a Researcher in the Scientific Computing Department at STFC. Vasily takes the lead on the technical development in WP3 - ...

# Brian Matthews

Technical Architect Brian Matthews is currently leader of the Data Science and Technology Group at STFC. He leads the DAFNI project, developing an ...

# **Current Contributors** Non-Executive and Partner Leads



# Samantha Pearman-Kanza

Pathfinder 2 Lead Samantha is a senior enterprise fellow at the University of



## Abraham Nieva de la Hidalga Pathfinder 1 Lead

Dr Abraham Nieva de la Hidalga currently collaborates as a



James Gebbie Pathfinder 4 Lead

James works for STFC at Daresbury Laboratory where his focus is supporting the High End

•••



# Alin-Marin Elena

Pathfinder 5 Lead Alin-Marin is a Computational Scientist at Daresbury Laboratory. He takes the lead on PF5 - ...

•••

## Leandro Liborio

Pathfinder 7 Lead Pathfinder 7: Reproducible Computational Workflows (lead Leandro Liborio) The ...

# Sathya Sai Seetharaman

#### Pathfinder 6 Lead

Sathya is a senior computational scientist in the Scientific Computation Department at STFC. ...

#### **Ian Bruno** Director of Data Initiatives -CCDC

Ian is the Director of Data Initiatives at our partner CCDC (The Cambridge Crystallographic

•••

# Kevin Ashley

Kevin Ashley is the Director of our partner DCC (The Digital Curation Centre). He joined DCC

•••



Royal Society University Research Fellow Chris Race is the lead at our partner, the Henry Royce Institute. Based at the University of ...

## **Other Contributors**

- Amanda Chapman STFC
- Victoria Hooper University of Southampton
- Georgia Lomas STFC
- Aileen Day University of Southampton
- Alexander Belozerov STFC
- Catherine Jones STFC
- Jo Grundy University of Southampton
- Mark Anderson University of Southampton
- Patrick Austin STFC
- Clara Lines Diaz University of Edinburgh
- Cheney Ketley STFC
- Rose Dickinson STFC
- Noel Vizcaino STFC
- Tom Allam University of Southampton
- Samuel Munday University of Southampton
- Elliott Kasoar STFC
- Jonathan Bathe STFC
- Noel Vizcaino STFC
- Philip Leadbitter University of Southampton
- Silvia Chiacchiera STFC
- Frazer Barnsley STFC
- Stephen Gow University of Southampton
- Silvia Chiacchiera STFC
- Jas Kalayan STFC
- Cerys Willoughby University of Southampton
- Amali Pawula Hewage STFC
- Federica Zanca STFC
- Pamela Slingsby STFC
- Ray Whorley University of Southampton
- Tom Underwood STFC
- Stavrina Dimosthenous University of Manchester

# **Previous Contributors**

- Alejandra Gonzalez Beltran STFC
- Gabin Kayumbi STFC

## • Sally Bloodworth – University of Southampton

#### Quick Links

Social

Join the email list | Twitter | LinkedIn | YouTube PSDI acknowledges the funding support by the EPSRC grants EP/X032701/1, EP/X032663/1 and EP/W032252/1

Terms of Use	Data	About us
Privacy	Services	Partners
Data protection	Communities	Support
Policy	Guidance and Training	Contact us
FAQ	News & Events	Sitemap



# The Pilot

Home » The Pilot

The content on this page is currently being updated

# **PSDI Pilot Activities**

PSDI pilot phase, was funded from the EPSRC Digital Research Infrastructure (DRI) programme, and ran from November 2021 – March 2022, following on from the large infrastructures SoN submitted by our project team. This pilot phase was a rapid scoping exercise, designed to expand on the ambitions of the project from the SoN, to engage broadly with the potential user community to gather and analyse requirements and to develop a plan for future phases.

Specifically, the objectives of the pilot were:

- Engage with the potential 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, elucidate the necessary functionality, and propose a technology architecture for PSDI
- Produce a detailed plan for future phases of PSDI
- Create a governance structure for future activities

## **Pilot Report**

The PSDI Pilot phase produced an overall high level report which summarised the activities that were carried out in the Pilot, as well as the recommendations and outlines for the future of PSDI.

#### Access the PSDI Pilot Report

## **Work Packages**

The pilot activity was split into four work packages, outlined below, to capture the different elements of the pilot activities. These WPs worked across the pillars to engage with different scales of data infrastructure and disciplinary areas to solicit a broad spectrum of requirements. However, there was collaboration between the different work packages due to the connected nature of the project and the activities in separate WPs fed into deliverables and outputs of others. You can find out more about each of the work packages on their individual pages.

#### **WP1. Coordination, Governance and Strategy**

WP1 directed the PSDI pilot project and developed the governance framework for later phases. It also collated results of the other WPs to produce a strategy and plan for PSDI going forward.

### **WP2. Stakeholder Engagement**

Community involvement was crucial to evaluating the potential of PSDI. A program of events and workshops were run to engage with as many identified stakeholders as possible. These engaged communities on both a group and individual level to elicit requirements that also fed into the other work packages.

## **WP3. Architecture & Technology**

WP3 explored and evaluated technology options and architecture designs for the PSDI underlying infrastructure. This also included undertaking small scale technology trials evaluating usability, scalability, security and interaction with other technologies.

### **WP4. Case studies**

Working with some of the key stakeholders identified earlier in this document we undertook case studies that scoped out or created prototype systems to demonstrate how proposed elements of PSDI could influence cutting-edge research. We also established a library of further use cases across our four pillars.

UK Research and Innovation	Engineering and Physical Sciences Research Council	Engineering and Physical Sciences	Quick Links		
		Research Council	<u>Terms of Use</u>	<u>Data</u>	<u>About us</u>
			<u>Privacy</u>	<u>Services</u>	Partners
loin the email list   Twitter   LinkedIn   VeuTube	PSDI acknowled	ges the funding support by the	Data protection	<u>Communities</u>	<u>Support</u>
<u>John the email list   Twitter   Linkedin   You lube</u>	EPSRC grants EI	P/X032701/1, EP/X032663/1	<u>Policy</u>	Guidance and Training	<u>Contact us</u>
	and EP/W0322	52/1	FAQ	<u>News &amp; Events</u>	<u>Sitemap</u>





# **PSDI Phase 1**

Home » PSDI Phase 1

Phase 1 of PSDI ran from October 2022 to September 2023 and built upon the scoping work carried out in the **Pilot Phase** and the recommendations and outputs that were generated through that work. This 12-month phase focused on activities towards the establishment of PSDI whilst continuing community engagement activities:

- PSDI-Hub creating core functionality and organisational structures to underpin PSDI operations
- Pathfinders early deployment activities to seed the infrastructure with exemplar tools, data and services

This phase was funded through EPSRC DRI funding (EP/X032701/1 and EP/X032663/1).

# **Work Packages**

The work in this phase was organised into 5 work packages, outlined below. You can find out more about each of the work packages on their individual pages.

## WP1. Management and Governance (lead Jeremy Frey)

WP1 oversaw the PSDI activities in this phase, undertaking the co-ordination and management activities across the project. It was also responsible for establishing governance and oversight mechanisms.

## WP2. Communications and Training (lead Nicola Knight)

Community awareness and engagement is critical to the establishment of PSDI. WP2 continued the engagement activities established in the pilot phase as well as coordinating with PSDI stakeholders and relevant interested parties. These engagement activities continued to feed into the development work being carried out in other work packages as well as communicating our work and development back out to the community.

## WP3. Platform Development and Operation (lead Vasily Bunakov)

WP3 was responsible for the development of the technical platform of PSDI. This WP worked closely with the Pathfinders in WP4 and community engagement activities throughout the development work.

## **WP4. First Pathfinders (lead Simon Coles)**

WP4 contained 5 pathfinders which were initial development activities focused around some key strategic areas. The first-round pathfinders were selected to have a mixture of domain focused pathfinders and cross-domain pathfinders.

# WP5. Future Pathfinders (lead Barbara Montanari)

WP5 focused on the development of additional candidates for pathfinder activities in the future. This builds on case studies and engagement activities from the pilot phase as well as wider community engagement. The pathfinder ideas will be developed so that they can be prioritised and planned as part of future PSDI phases.

Join the email list | Twitter | LinkedIn | YouTube

PSDI acknowledges the funding EPSRC grants EP/X032701/1, E and EP/W032252/1

operations s, data and services

		Quick Links	
	Terms of Use	<u>Data</u>	<u>About us</u>
	<u>Privacy</u>	<u>Services</u>	<u>Partners</u>
support by the	Data protection	<u>Communities</u>	<u>Support</u>
P/X032663/1	<u>Policy</u>	Guidance and Training	<u>Contact us</u>
	<u>FAQ</u>	<u>News &amp; Events</u>	<u>Sitemap</u>



# **PSDI Phase 2**

Home » PSDI Phase 2

#### The content on this page is currently being updated

The current phase of PSDI, Phase 2, which started in January 2024 continues to build upon the work carried out in <u>The Pilot</u> and <u>Phase 1</u> of the project. The PSDI initiative is funded through EPSRC DRI funding (EP/X032701/1 and EP/X032663/1).

This phase of PSDI will start with the second half of the initial 30-month development (started in Phase 1). This will lead to the launch of version 1 of PSDI in March 2025. PSDI will then continue in a dual mode, maintaining the operational services of version 1 and continued development to provide further services.

# **Work Packages**

This work is broken down into seven work packages, covering technical elements, content and social aspects of the project. The image below outlines the key aspects of each workpackage, with a more in-depth description given further down.

<b>Governance (WP5)</b> UKRI DRI Advisory Board International Policy Institutional	Management (WP6) Comms Partnerships Resources Project Standards Administration	<b>Community (WP7)</b> Engagement Training Funding call Requirements capture				
Content (WP4)						
External Open Data Internal Data Licensed Collections Knowledgebase / Tool						
Gateway (WP1)	Platform (WP2)	Data Services (WP3)				
User Gateway User Management System Administration	Hardware Operating systems Storage Databases Operations	Data Discovery Data workflows Publishing Notebooking				

## WP1. Gateway

This workpackage will be responsible for the front-end services available through PSDI. It will develop and operate the User Gateway through which users interact with PSDI, and the Administration Gateway through which the system is managed.

## **WP2. Platform**

This workpackage will be responsible for delivering and operating the scalable and secure hardware and system software platform upon which the PSDI services will run. It will specify, procure and install the hardware; install and maintain the operating systems and cloud environment; maintain the data storage and database systems, including virtualisation, object storage, caching, ingest and access APIs, data movement and data backup.

## **WP3. Data Services**

This workpackage will be responsible for services definition, development and extension. This includes services for Data Indexing and Search, Data Versioning, Data Sharing and Publication, Near-Storage Data Processing, scientific workflow systems connected to external data services and HPC, and auxiliary services such as ELN finder or data format conversion.

## **WP4. Content**

PSDI will surface and connect data from three different types of source: external (open), internal and licensed data. This workpackage is concerned with developing the partnerships, systems and processes to promote the growth and connectivity of these different types of data collections.

## **WP5. Governance**

The governance work package will ensure a cohesive project that meets the aims of the DRI programme, is responsive to the community, and adheres to the principles of public life in the use of public funds. The investigators, through the Management Board, will provide the executive role responsible for the PSDI functions, and an Advisory Board will provide national and international community perspectives for PSDI and, together with the Management Board, interact with the UKRI/EPSRC DRI overall programme.

## **WP6.** Management

Effective management is critical in coordination of this complex endeavour to deliver a functional national data infrastructure. This workpackage focuses on oversight and coordination of day-to-day project operations, ensuring that task work proceeds in an efficient and timely manner to produce deliverables.

## **WP7. Community**

In this and future phases, PSDI must grow from being an exemplar set of data resources, tools and services to become a fully-fledged ecosystem providing support across the physical sciences domain and integrate with DRI projects in other domains. This requires advocacy, engaging with existing communities and nurturing new communities, to define and promote the necessary coverage and utility.

## **Phase 2 Pathfinders**

Pathfinders are a means to focus some development activity in key strategic areas and explored and established exemplar approaches and systems that can be folded into the PSDI infrastructure and act as a template or starting point to bring in further domains, data types, techniques and user communities in future phases.

The Phase 2 pathfinders build upon or combining the ones from Phase 1, as well as adding new ones. Phase 2 currently has 6 pathfinders,

- PF1: Experimental Data Capture
- PF2: Process Recording and Data Collections
- PF4: Data Infrastructure and Tooling for Biomolecular Simulation
- PF5: Data to Knowledge
- PF6: Collaborative Computational Project for NMR Crystallography
- PF7: Reproducible Computational Workflows

More information about these pathfinders can be found on the **Phase 2 Pathfinders Page**.

Quick Links

News & Events

Sitemap

FAQ

Social		Terms of Use	Data	About us
	PSDI acknowledges the funding support by the	Privacy	Services	Partners
Join the email list   Twitter	LF SINC grants LF/X032701/1, LF/X032003/1	Data protection	Communities	Support
LinkedIn   YouTube	and EP/W032252/1	Policy	Guidance and Training	Contact us





# Phase 2 Pathfinders

Home » Phase 2 Pathfinders

## Pathfinder 1 – Experimental Data Capture (lead Abraham Nieva de la Hidalga)

The research supported by large scale facilities generates valuable and high-quality research data. However, data publishing practices seem to undervalue the significance of these data. There is limited support for reuse beyond the original research, reproducibility is challenging, and the provenance link between published results and supporting data is not always explicit. In this scenario, the PSDI can address these issues by supporting the development/adoption of tools that facilitate experimental data capture. The tools should (a) produce provenance data for every processing step, (b) provide references for all new and existing data used, (c) unambiguously map results to data objects, and (d) generate the necessary metadata to facilitate the publication of data supporting the presented results. We believe that this initiative will make a valuable contribution to physical sciences research. The current examples of the development of Galaxy tools and workflows for XAS processing can be extended to support other experimental areas including XPS, INS, QENS or PDF.

## Pathfinder 2: Process Recording and Data Collections (lead Samantha Pearman-Kanza)

Process recording is fundamental to driving validation, trust, reuse of data across the sciences and is a crucial aspect of data management and integration. It provides crucial support for the researcher in the laboratory whilst simultaneously structuring data for subsequent downstream management, publication, and reuse. It encompasses workflow support and data management systems, but more importantly tools for ingest that seamlessly fit into existing laboratory researcher practices and account for a range of processes from capturing handwritten observations to interfacing with instruments and recording data analysis. There are also socio-cultural aspects around disciplinary working practices that must be considered, and the support must seamlessly interface with other aspects of scholarly practice such as linking to published literature, database searching and writing reports/papers/theses. This pathfinder will investigate generic and domain specific tools for process recording and assess the landscape of semantic web technologies and metadata schemas for the physical sciences to aid with this process. This pathfinder will produce services to empower researchers to choose the best tools for them and provide exemplars and guidelines on best practices for publishing and sharing data.

Please note: due to a converging of interests this pathfinder now incorporates the work that was being done by pathfinder 3 in Phase 1

# Pathfinder 4: Data Infrastructure and Tooling for Biomolecular Simulation (lead James Gebbie-Rayet)

This pathfinder aims to create production data infrastructure services that will enable researchers to maximise the value and quality within their data and to enable the development of future novel methods and technologies, that are currently not possible due to an absence of centralised approaches to data. It will do this by focusing on providing a technological solution to enable and simplify the capture and sharing of data from within biomolecular simulation workflows, and in a way that does not require huge cultural shifts in ways of working thus lowering the barriers to adoption. The intention is to capture the full pipeline of data from experimental input through to analytical outputs to preserve the full data provenance in how scientific studies are performed. The work here will be in partnership with and informed by the biomolecular simulation community via existing relationships under the STFC CoSeC programme in the form of CCPBioSim (scientific) and HECBioSim (HPC) consortia along with external partner institutions such as the EBI.

## Pathfinder 5: Data to Knowledge (lead Alin Elena)

Pathfinder 5 focuses on the field of machine learning interatomic potentials, MLIP. Advances in this area need expensive calculations to be produced and used for training machine learning models. In addition the models resulting from these models are non trivial in terms of storage and distribution compared with previous generation interatomic potentials which tended to be analytical. This pathfinder aims are to design and deploy the hardware infrastructure to host both training data for ML and also act as a distribution centre within PSDI. There is also planned work to provide suitable benchmarks and validations for MLIP, building on work done Cambridge Crystallographic Metal Oraganics Frameworks data base to screen high volumes of compounds to extract relevant for specific work and providing training to the community through summer schools.

# Pathfinder 6: Collaborative Computational Project for NMR Crystallography (lead Sathya Sai Seetharaman)

Pathfinder 6 focuses on the Collaborative Computational Project for NMR Crystallography (CCP-NC). CCP-NC combines experimental NMR and computation to provide new insight, with atomic resolution, into structure, disorder, and dynamics in the solid state. The first aim of the pathfinder is to improve the existing Magnetic Resonance (magres) database at CCP-NC, which is an open repository of computational solid state NMR results (https://www.ccpnc.ac.uk/database/). The second is the to redevelop this database to future-proof the frontend whilst adding new features including but not limited to more complex search option and

interoperability with other material databases.

## Pathfinder 7: Reproducible Computational Workflows (lead Leandro Liborio)

The post-processing of experimental and simulation data, associated to large scale experiments performed at national facilities, requires that different software tools from various domains are connected into workflows. These workflows can be quite complex and, in this pathfinder, we will be using the open-source, web-based Galaxy platform to manage the software tools and data associated to these workflows. Galaxy is a platform for FAIR data analysis that enable users to: run code in interactive environments; share and publish results, workflows and their associated visualizations; and ensure the reproducibility of their results by capturing and packaging data, metadata and provenance models required for repeating and understanding their data analyses. We are initially working with muon science and catalysis experiments, but plan to expand the methodology to other large scale experiments.

	Quick Links		
Funding support by the 01/1, EP/X032663/1	Terms of Use Privacy Data protection Policy FAQ	Data Data Services Communities Guidance and Training News & Events	About us Partners Support Contact us Sitemap
f 7	funding support by the 701/1, EP/X032663/1	funding support by the 701/1, EP/X032663/1       Terms of Use         Privacy       Data protection         Policy       FAQ	Quick Linksfunding support by the O1/1, EP/X032663/1Terms of UseDataPrivacyServicesData protectionCommunitiesPolicyGuidance and TrainingFAQNews & Events