



**PSDI**  
PHYSICAL SCIENCES  
DATA INFRASTRUCTURE

# Connecting Infrastructures: The Physical Sciences Data Infrastructure (PSDI) in the UK

CoRDI Karlsruhe, 13<sup>th</sup> Sept 2023

| Lead Team                 |                           |
|---------------------------|---------------------------|
| STFC Scientific Computing | University of Southampton |
| Juan Bicarregui           | Simon Coles               |
| Vasily Bunakov            | Nicola Knight             |
| Brian Matthews            | Jeremy Frey               |
| Barbara Montanari         |                           |



Engineering and  
Physical Sciences  
Research Council



UK Research  
and Innovation

# Aim of PSDI

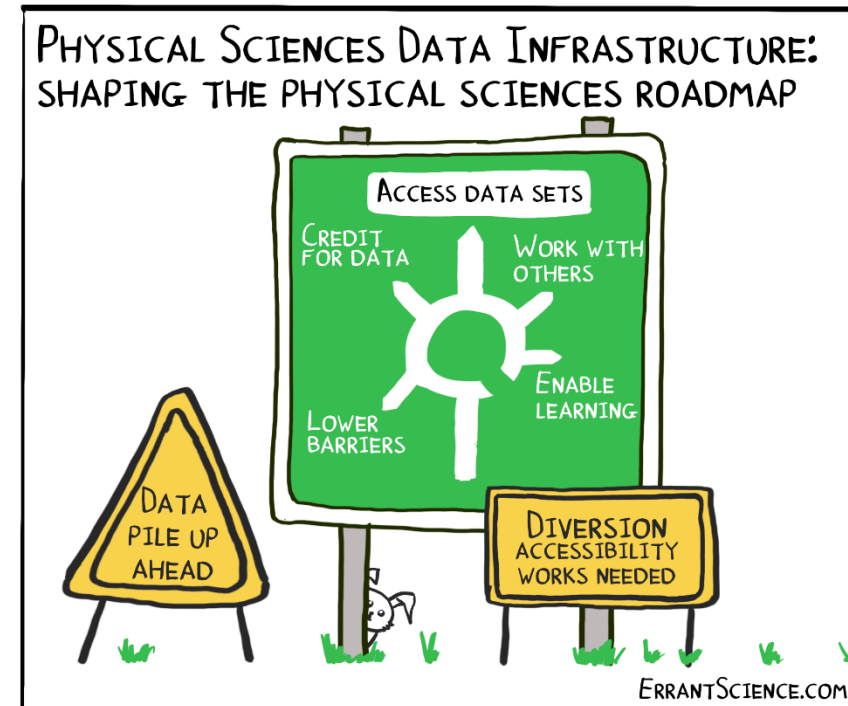
Data is a major driver of research in Physical Sciences, but it is fragmented

PSDI will provide

A data infrastructure that  
**connects existing**  
experimental and computational facilities  
within Physical Sciences and beyond

Through PSDI:

- ▶ 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



# PSDI: filling a Gap in Provision

- ▶ **Other countries** have initiatives underway in this domain, e.g.
  - ▶ USA: Materials Genome Initiative
  - ▶ Japan: NIMS
  - ▶ European data infrastructures, such as E-CAM, MaX and NOMAD
  - ▶ German National Research Data Infrastructure (NFDI)
- ▶ **Other domains** have initiatives underway in the UK, e.g.
  - ▶ EBI in Life Sciences
  - ▶ NERC Data centres in Environmental Science
  - ▶ UK Data Archive in Social Science

## We are building a UK, Physical Science, Data Infrastructure

- ▶ Supporting Chemistry, Materials and related disciplines
- ▶ Traversing to and interfacing with Life, Medical, Engineering and Environmental Sciences

# The bigger picture in the UK

## Facilities, Institutes & Hubs

### Examples:

- Catalysis Hub
- CCFE
- Central Laser Facility
- Diamond
- Future Manufacturing Hub
- ISIS
- Royce Institute
- ATI

## National Research Facilities

### Examples:

- HarwellXPS
- NXCT
- NCS
- PSDS
- SuperSTEM
- UK High Field Solid-State NMR
- XMaS

## Computational Initiatives

### Examples:

- CCP5++
- CCP9
- CCP Biosim
- CCP EngSci
- CCPi
- CCP NC
- CCP NTH
- CCP QC
- CCP SAS
- CCP Turbulence
- CCP WSI
- SSI
- UK society of RSE
- HEC Biosim
- HEC Plasma
- MCC
- UKCTRF
- UKCP
- UKTC
- CoSeC
- EPSRC Tier2
- ExCALIBUR
- STFC Hartree Centre
- ARCHER

## Research Institutions, Groups and Laboratories

### Examples:

- Equipment Infrastructures
- Equipment Facilities
- University Labs
- ELNs
- Repositories
- Local Computing Resources

**PHYSICAL SCIENCES DATA INFRASTRUCTURE**

# PSDI: Outline Timeline

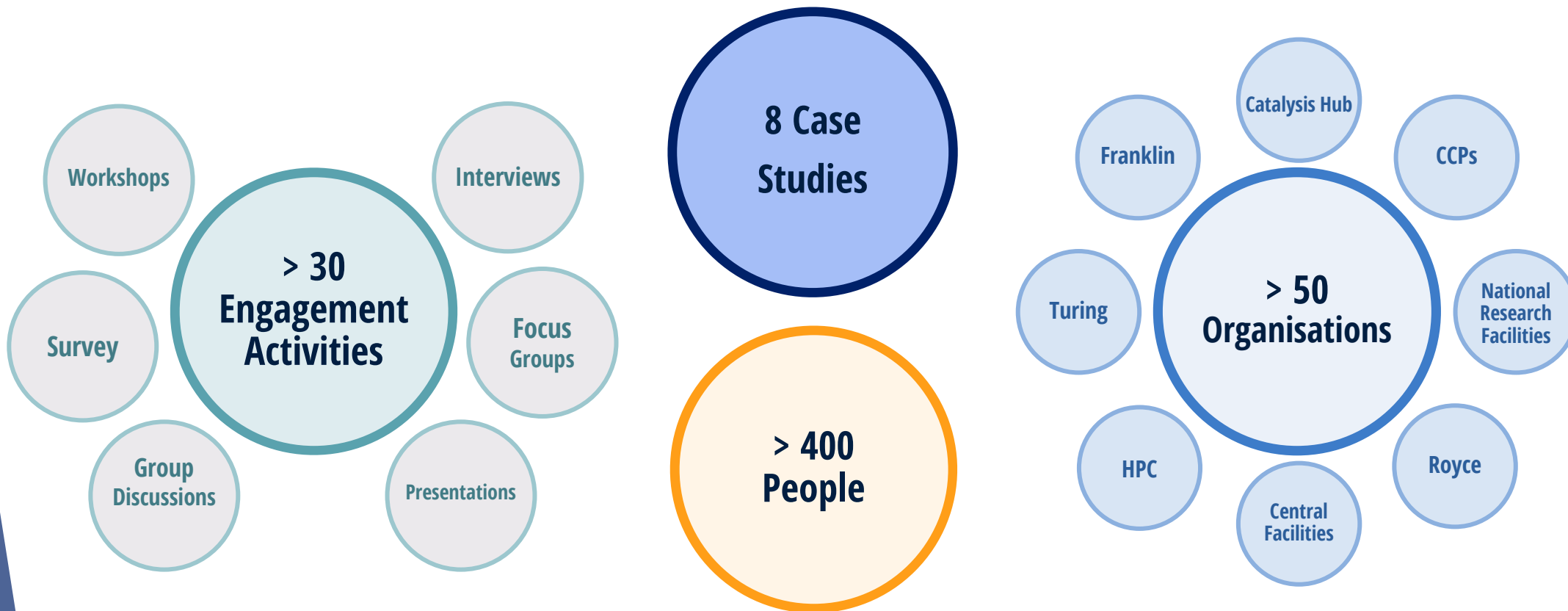
- ▶ Feb 2021: EPSRC statement of need for large-scale infrastructure
- ▶ Nov 2021 – Mar 2022: PSDI Pilot
  - ▶ Initial PSDI engagement & design including 8 case studies
- ▶ Oct 2022 – Sept 2023: PSDI Phase 1b
  - ▶ Funded as part of UKRI DRI Phase 1b
  - ▶ Begin initial development of PSDI Hub components
  - ▶ 5 pathfinders to create exemplar tools, data and services
- ▶ Oct 2023 – Mar 2025: PSDI (Phase 1c)
  - ▶ To be funded from next phase of UKRI DRI
  - ▶ Complete initial development and deploy Platform and 5 Pathfinders
  - ▶ Develop further Pathfinders
- ▶ Beyond Mar 2025 (Phase 2: Operational and Further development)



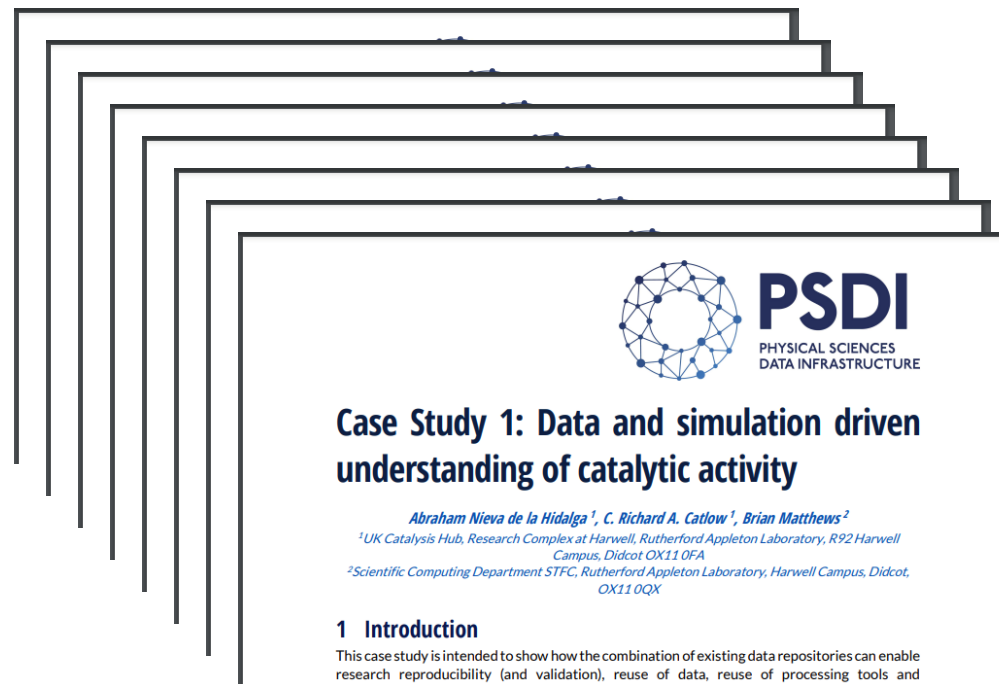
Single  
30 month  
development  
project

# Pilot Phase at a Glance

Catalysis      Spectroscopy      Structural Representation  
Materials      ELNs      Data Curation



# Pilot Outputs



You can find the PSDI *pilot* final report on the PSDI website....

...and the reports from each case study

Available on PSDI website: [psdi.ac.uk](https://psdi.ac.uk)  
And the [PSDI Zenodo Community](#)

## Acknowledgements

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

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

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

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

|  |   |   |   |
|--|---|---|---|
| Alan Turing Institute (ATI)<br>UK Catalysis Hub<br>Central Laser Facility<br>Diamond Light Source<br>ISIS Neutron and Muon Source<br>Rosalind Franklin Institute (RFI)<br>Henry Royce Institute<br>EPR National Research Facility (NRF)<br>Harwell/DXPS NRF<br>Ion Beam Facility NRF<br>NCS NRF<br>NXCT NRF<br>PSDS NRF<br>UK High-Field Solid-State NMR NRF<br>XMAS NRF | CCP Steering Panel (i.e., Chairs of CCPs, representatives of Research Councils, intl representatives from CECAM etc.)<br>CCP-Biosim<br>CCP-NC<br>CCP-NTH<br>CCPWSI+<br>CCP4<br>CCP5+<br>CCP7<br>CCP Turbulence<br>CCP-EM<br>European Psi-k Network AIDA | European Materials Modelling Council<br>MMMHub<br>High-End Computing Consortium (HEC) MCC<br>HEC UKCP<br>HEC UKCT<br>HEC UKCTRF<br>HEC UK-AMOR<br>HEC Plasma<br>HEC Biosim<br>HEC UKCOMES<br>RSE<br>SSI<br>IRIS | DAFNI<br>JASMIN<br>Tier 2 HPC Providers<br>NB-CIR<br>Wilder PS Community<br>CCDC<br>EMBL-EBI<br>NERC<br>EPSRC<br>STFC<br>NOMAD<br>CECAM |
|--|---|---|---|

Logo by Wearherd  
Report design & preparation by Nicola Knight

Physical Sciences Data Infrastructure

# Current Work - Phase 1b



# Platform, Pathfinders and Hub

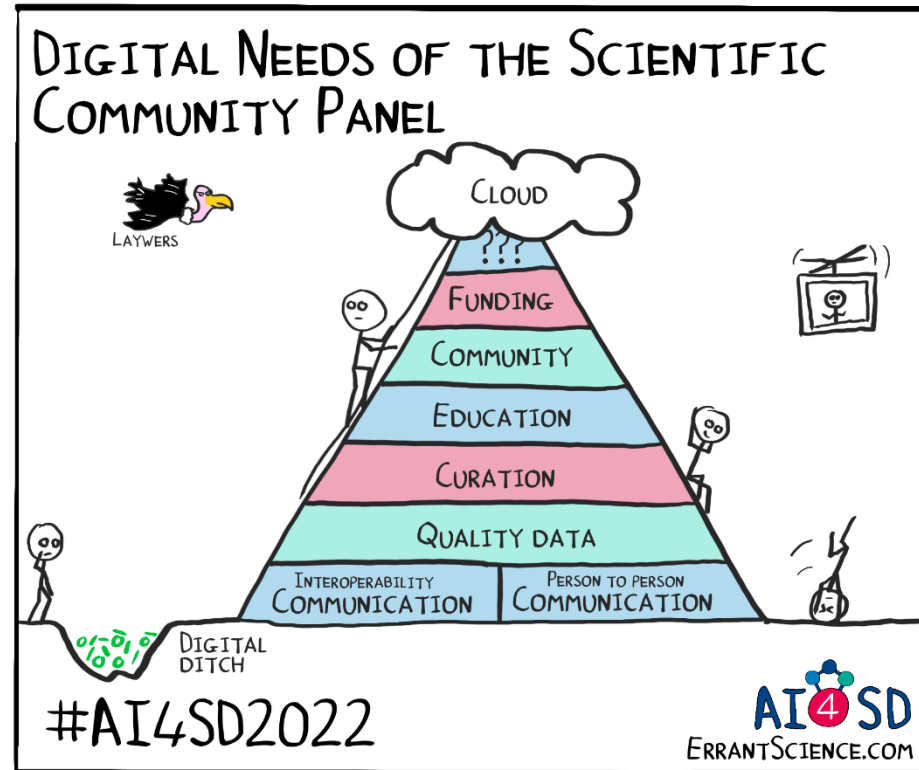
## ► Platform

- Requirements Analysis
- Capacity Planning
- System Architecture design
- Component testing
- Beginning Build

## ► Pathfinder project

- Initial development and implementation

## ► Hub: Communications, Governance, Planning,...

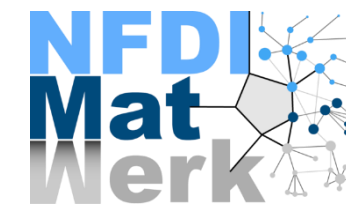


Cartoon created by ErrantScience.com for AI4SD: licensed under [CC-BY-NC](https://creativecommons.org/licenses/by-nc/4.0/)

# Our areas

- ▶ Physical Sciences is a broad area
- ▶ Initial focus areas:
  - ▶ Catalysis
  - ▶ Biomolecular simulation
  - ▶ Materials
  - ▶ Atomistic simulations
  - ▶ NMR
  - ▶ Process Recording
  - ▶ Data Collections

## NFDI Alignment



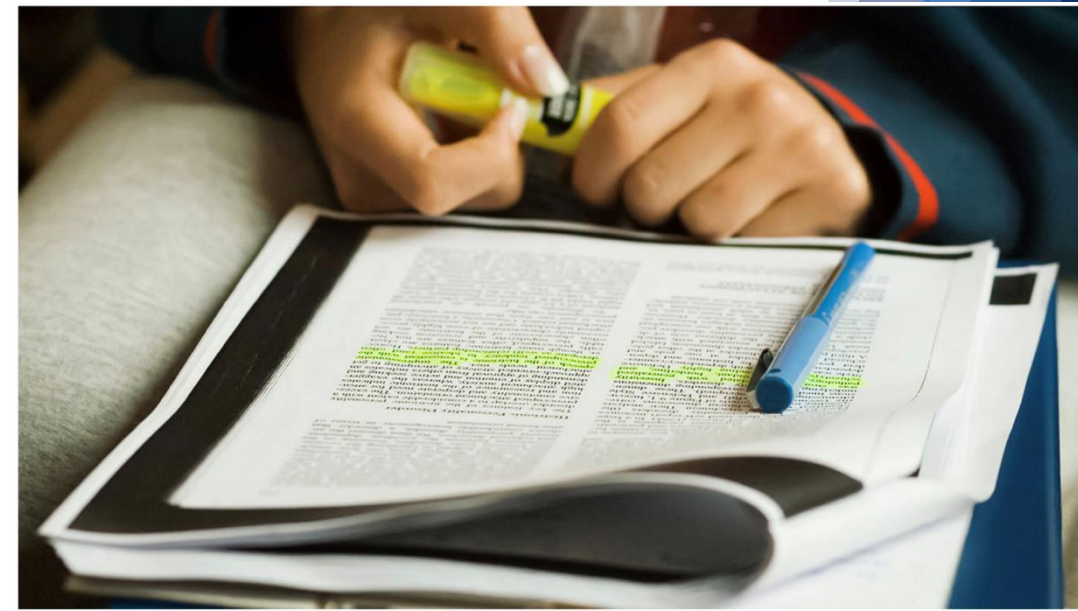
# An Example: Biomolecular Simulations



- Run 10s of simulations to generate data
- Apply know-how to extract science from data
- Publish paper

But ....

- Paper does not include all details needed to **repeat** simulation
- Citations do not give **credit** for *all* resources used



# Provenance map of a Single Command in a Simulation

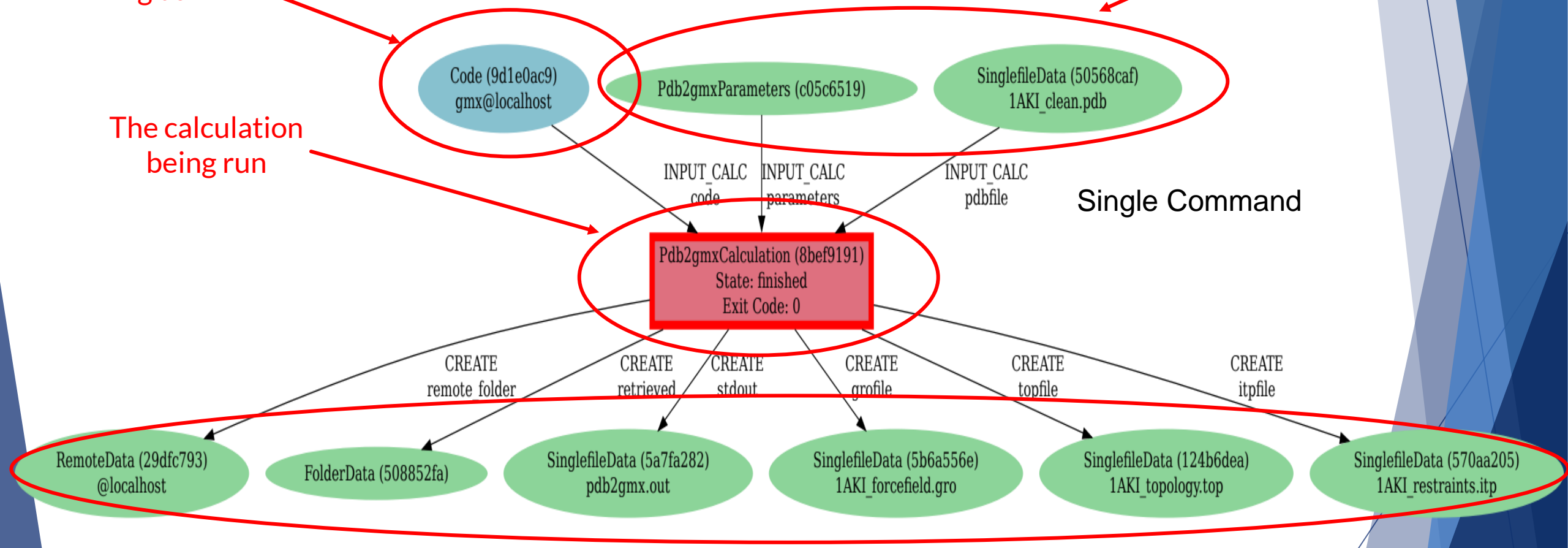
The computer  
being used

The inputs

The calculation  
being run

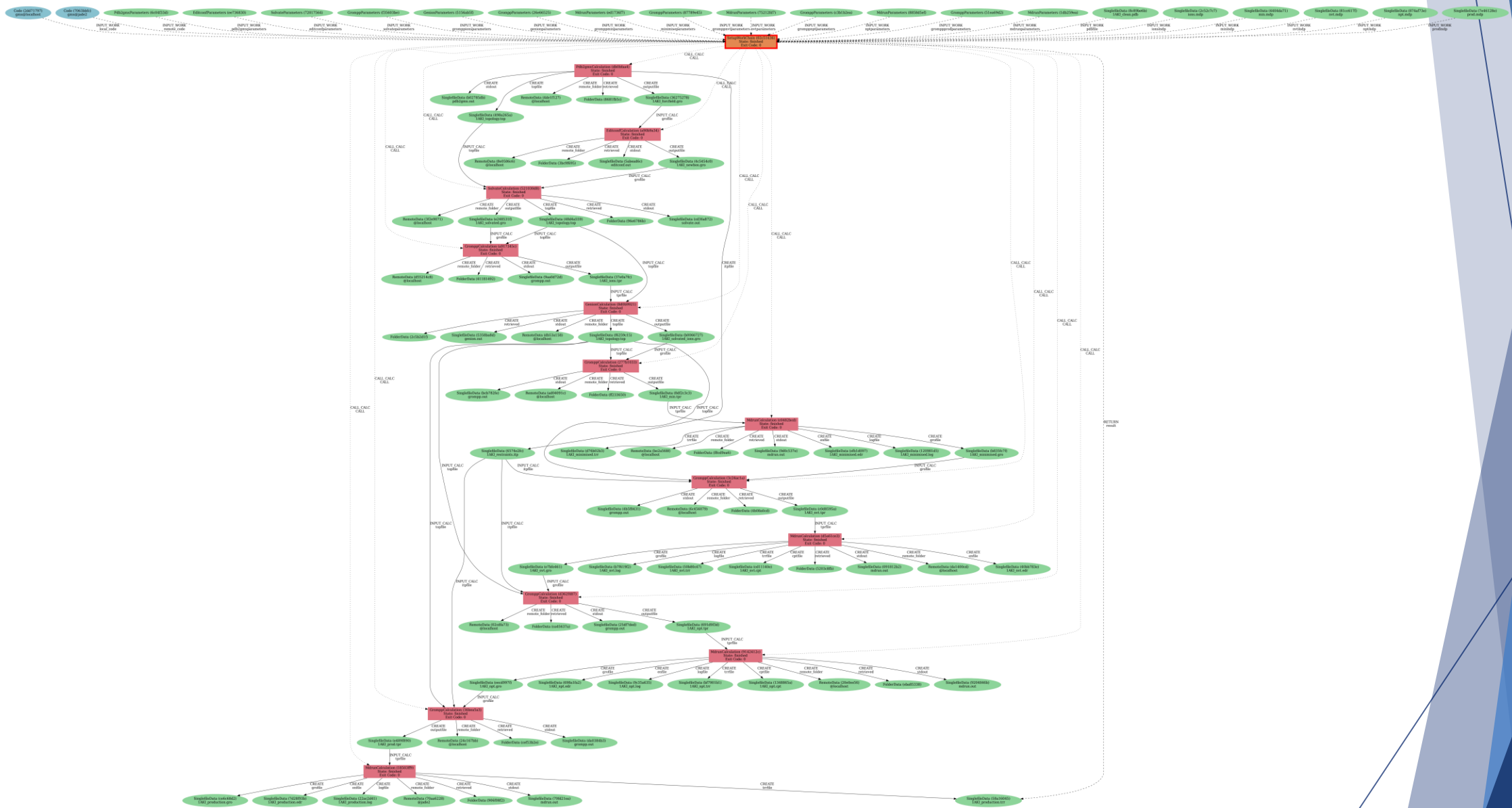
Single Command

The outputs

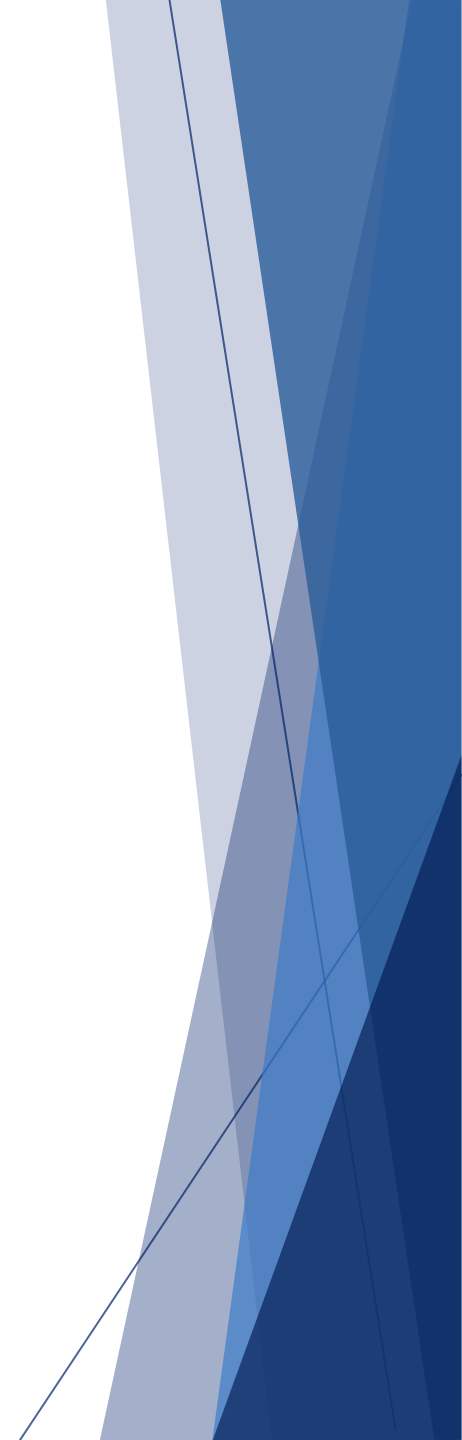
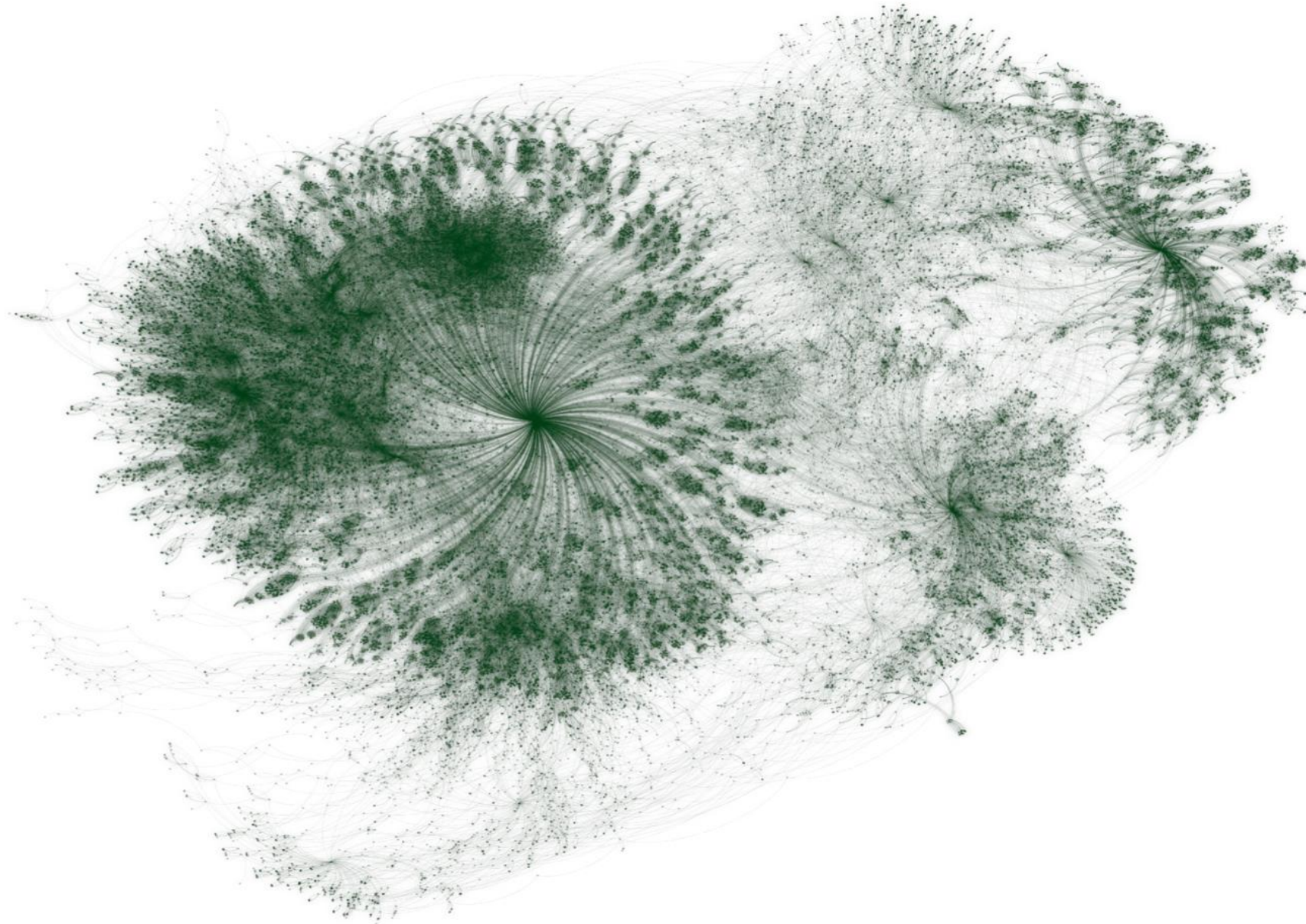




# Creating a model... (not yet a simulation)



# An Entire Study



# Process Orchestration PathFinder: User Environment Prototype

- Built prototype to capture full data provenance using AiiDA
- Building on GROMACS software (70% of users in UK HPC Biosim Consortium)
- Designed to mimic working with native package (command line driven)
- Simple to install and setup our plugin

## Normal command:

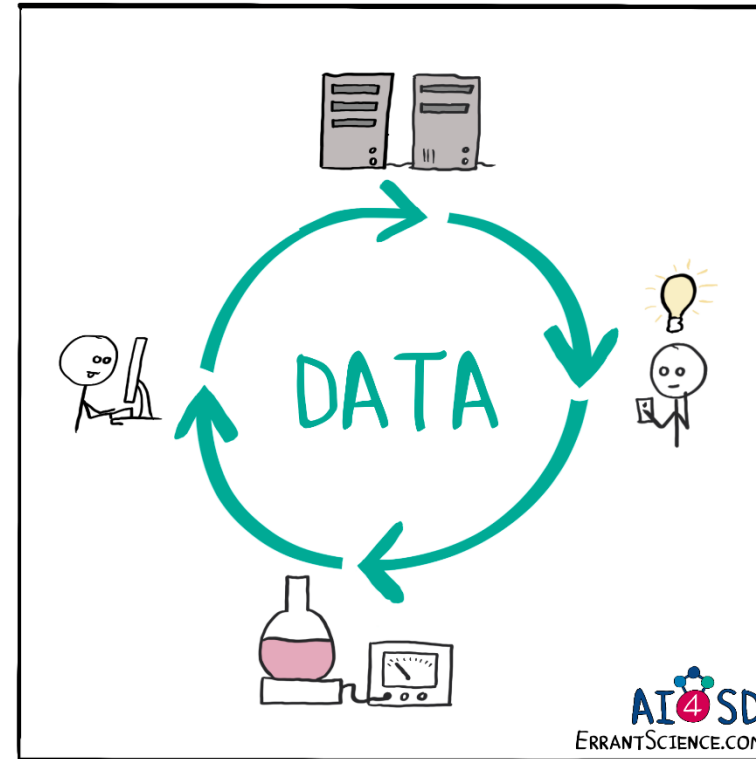
```
gmx pdb2gmx -f prot.pdb -ff oplsaa -water spce -o prot.gro -p prot.top -i prot.itp
```

## Capture provenance with AiiDA:

```
gmx_pdb2gmx -f prot.pdb -ff oplsaa -water spce -o prot.gro -p prot.top -i prot.itp
```

# Collaboration Topics

- ▶ Metadata
  - ▶ Standards
  - ▶ Collection
- ▶ Ontologies
- ▶ Standards
- ▶ Best practices
- ▶ Training
- ▶ Tool development
- ▶ Publishing & Supplementary Data



Cartoon created by ErrantScience.com for AI4SD: licensed under [CC-BY-NC](https://creativecommons.org/licenses/by-nc/4.0/)



# International Collaboration (among others)










**WorldFAIR**

# Collaboration with NFDI

Information sharing on key people within projects across NFDI / PSDI

Strengthening the connection and making it easier to collaborate

If you want to get involved with this please contact Nicola Knight (PSDI) or John Jolliffe (NFDI4Chem)

|  PSDI<br>PHYSICAL SCIENCES<br>DATA INFRASTRUCTURE   | Topic  |  NFDI <sub>4</sub> Chem  |  NFDI <sub>4</sub> at<br>NFDI for Chemistry Related Sciences |  DAPHNE<br>4NFDI   |  FAIRmat |  NFDI<br>Mat<br>Werk |  PUNCH<br>4NFDI |
|--|--|---|---|---|---|---|--|
| <a href="#">PSDI Website</a><br><a href="#">Twitter</a><br><a href="#">LinkedIn</a><br><a href="#">Research Community</a><br><a href="#">YouTube</a>                               | Online Resources   | <a href="#">NFDI4Chem Knowledge Base</a><br><a href="#">NFDI4Chem Database</a><br>Tutorials<br>Library<br>Literature<br>ELN/ELMS<br>Benchmarking Service<br>Database of NFDI<br>Ecosystem<br>Integration<br>Search Service  |   | <a href="#">DAPHNE/NETC website</a><br><a href="#">Ecosystem</a>  |   |   |  |
| <a href="#">Project Coordinator, Nicola Knight</a> <a href="#">@nknight</a><br>PSDI is managed by the main committee, comprised mostly of the workpackage leads and investigators. | Management and Coordination  | <a href="#">Project Coordinator, Francesca Giordani</a> <a href="#">@fagiordani</a><br><a href="#">Members, Christoph Birkbeck</a> <a href="#">@cbirkbeck</a><br><a href="#">Members, Christa Grottel</a> <a href="#">@cgrottel</a><br>NFDI4Chem is managed by its steering committee which is composed of the work area leads. You can find the corresponding leads of the work areas:<br>Lead area work package lead (see project manager) (PM) (see below table)   |   | <a href="#">Project coordinator, Lisa-Jessie DREIB</a> <a href="#">@ljdreib</a><br><a href="#">Members, Ralf W. G. (John) Brey</a> <a href="#">@rwbrey</a><br><a href="#">Members, S. J. (John) Brey</a> <a href="#">@sjbrey</a><br>DAPHNE/NETC is mainly managed by TM (project management) and the Executive Board which is composed of the lead area leads (see below for the corresponding leads of the lead areas).  |   |   |  |
| Corresponds to WP2 in Phase 1a<br>WP2 Lead: Nicola Knight <a href="#">@nknight</a>   | Skills & Training  | <a href="#">Coordinating NFDI4Chem Task Area</a> <a href="#">T1a</a><br><a href="#">T1a</a> leads: <a href="#">Christoph Birkbeck</a> <a href="#">@cbirkbeck</a><br><a href="#">T1a</a> PM: <a href="#">John Jolliffe</a> <a href="#">@jjolliffe</a><br><a href="#">T1a</a> PM: <a href="#">John Jolliffe</a> <a href="#">@jjolliffe</a><br>NFDI4Chem has a team of dedicated NFDI4Chem Systems for NFDI workshops. Workshops for the Chemistry ELN are also offered.<br><a href="#">https://www.nfdi4chem.de/en/infrastructure/</a><br>The training materials for the chemistry specific NFDI workshops will be made available on Zenodo in due course.<br>Contact for NFDI workshops: <a href="#">Nicola Knight</a> or <a href="#">John Jolliffe</a><br>Contact for Chemistry Workshops: <a href="#">Christoph Birkbeck</a><br>Chemists can also find us for <a href="#">knowledge base</a> |   | <a href="#">Coordinating DAPHNE/NETC Task Area</a> <a href="#">T1a</a><br><a href="#">T1a</a> workshop: <a href="#">Tobias Dohler</a> <a href="#">@tdohler</a><br><a href="#">T1a</a> leads: <a href="#">Jan Dirk Grunwald</a> <a href="#">@jdgrunwald</a> and <a href="#">Andri Schenkel</a> <a href="#">@schenka</a>  |   |   |  |
| Corresponds to WP2 in Phase 1a<br>WP2 Lead: Nicola Knight <a href="#">@nknight</a>   | Outreach & Networking + representation in international initiatives                | <a href="#">Coordinating NFDI4Chem Task Area</a> <a href="#">T1a</a><br><a href="#">T1a</a> leads: <a href="#">Christoph Birkbeck</a> <a href="#">@cbirkbeck</a><br><a href="#">T1a</a> PM: <a href="#">John Jolliffe</a> <a href="#">@jjolliffe</a><br><a href="#">T1a</a> PM: <a href="#">John Jolliffe</a> <a href="#">@jjolliffe</a><br>NFDI4Chem has many people involved in outreach in networking. Talks are given regularly. Internationally, conferences and other events. Performance, Outreach efforts to grow it's international network. NFDI4Chem works closely together with the German Chemical Society (GDCh) Contact <a href="#">John Jolliffe</a> for more details.<br><a href="#">Hans-Gregor Wang</a> <a href="#">@hgwang</a>  |   | <a href="#">Coordinating DAPHNE/NETC Task Area</a> <a href="#">T1a</a><br><a href="#">T1a</a> workshop: <a href="#">Tobias Dohler</a> <a href="#">@tdohler</a><br><a href="#">T1a</a> leads: <a href="#">Jan Dirk Grunwald</a> <a href="#">@jdgrunwald</a> and <a href="#">Andri Schenkel</a> <a href="#">@schenka</a><br><a href="#">DAPHNE/NETC Task Area</a> <a href="#">T1a</a><br><a href="#">T1a</a> leads: <a href="#">Andri Schenkel</a> <a href="#">@schenka</a> and <a href="#">Christian Goll</a> <a href="#">@cgoll</a> |   |   |  |
|  | Storing, Sharing and making Data available to others (Repositories, Archives etc.) | <a href="#">Coordinating NFDI4Chem Task Area</a> <a href="#">T1a</a><br><a href="#">T1a</a> leads: <a href="#">Felix Beck</a> <a href="#">@fbeck</a><br><a href="#">T1a</a> PM: <a href="#">Christoph Birkbeck</a> <a href="#">@cbirkbeck</a><br><a href="#">T1a</a> PM: <a href="#">Christoph Birkbeck</a> <a href="#">@cbirkbeck</a><br>A good overview and introduction to NFDI4Chem's work on repositories can be found here:<br><a href="#">https://www.nfdi4chem.de/en/infrastructure/</a>  |   | <a href="#">Coordinating DAPHNE/NETC Task Area</a> <a href="#">T1a</a><br><a href="#">T1a</a> leads: <a href="#">Robert Bausch</a> <a href="#">@rbausch</a> and <a href="#">Tobias Dohler</a> <a href="#">@tdohler</a>  |   |   |  |
| Corresponds mostly to Publications 23 in Phase 1a<br>Lead: <a href="#">Ramantha Kenna</a> <a href="#">@rkenna</a>  | ELNs / Smartlab / tools for capturing and managing data / tools / workflows        | <a href="#">Coordinating NFDI4Chem Task Area</a> <a href="#">T1a</a><br><a href="#">T1a</a> leads: <a href="#">Nicola Knight</a> <a href="#">@nknight</a><br><a href="#">T1a</a> PM: <a href="#">John Jolliffe</a> <a href="#">@jjolliffe</a><br>Contact for NFDI4Chem ELN/ELMS relating to the ELN Consortium. Chemists seek original designed for chemists and not vice versa. These tools are to allow researchers to capture their own data. The implementation of the FRK pipeline can be found here:<br><a href="#">https://www.nfdi4chem.de/en/infrastructure/</a><br>The ELN has direct integration to repositories like Zenodo, crossref, data publication, Pubmed etc.  |   | <a href="#">Coordinating DAPHNE/NETC Task Area</a> <a href="#">T1a</a><br><a href="#">T1a</a> leads: <a href="#">Wolfram Lohmeyer</a> <a href="#">@wlohme</a> and <a href="#">Bridget Murphy</a> <a href="#">@brimurphy</a>   |   |   |  |

# ELN Finder

<https://eln-finder.ulb.tu-darmstadt.de>

## ELN Finder

The ELN Finder helps you to search and select a suitable Electronic Lab Notebook (ELN) for your purposes.

- More than 40 filter criteria available.
- Filter criteria clearly divided into categories.
- Result list of the identified ELN tools displayed in an overview.
- Brief descriptions of the individual tools included.

 Find ELNs

- ▶ Collaboration of PSDI Process recording work with ELN Finder in NFDI4Chem
- ▶ Eliminates duplication of services and effort

# See what we are up to

- ▶ We are running a webinar series
  - ▶ Tues 3<sup>rd</sup> Oct 2pm – Experimental Data Capture in catalysis: producing publish ready data from processing and analysis processes
  - ▶ Oct – Exact date TBC - Process orchestration in Biomolecular Simulation
- ▶ Register now: <https://www.psd.ac.uk/events/>
- ▶ Check out our website
  - ▶ [www.psd.ac.uk](http://www.psd.ac.uk)
- ▶ Find us on social media
  - ▶ Twitter / X: @PSDI\_UK
  - ▶ LinkedIn: PSDIUK
- ▶ Follow our YouTube Channel: [youtube.com/@PSDI\\_UK](https://youtube.com/@PSDI_UK)
- ▶ Join our JISCmail mailing list via our website footer: PSDI



# PSDI

PHYSICAL SCIENCES  
DATA INFRASTRUCTURE

## Any Questions?

[www.psd.ac.uk](http://www.psd.ac.uk)



@PSDI\_UK



@PSDI\_UK



PSDIUK