PSDI

PHYSICAL SCIENCES
DATA INFRASTRUCTURE

UK Physical Sciences Data Infrastructure (PSDI) initiative

24th November 2023 - NFDI4Chem Stammtisch

NFDI4Chem

Dr Nicola Knight
&
Dr Samantha Pearman-Kanza

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

Aim(s) of PSDI

Support Data as a major driver of research in Physical Sciences

PSDI will provide
A data infrastructure that
connects existing
experimental and computational facilities
within Physical Sciences and beyond

Building Bridges

- Sustaining data resources beyond lifespan of individual research projects

PSDI
PHYSICAL SCIENCES
DATA INFRASTRUCTURE

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)

UK catch up

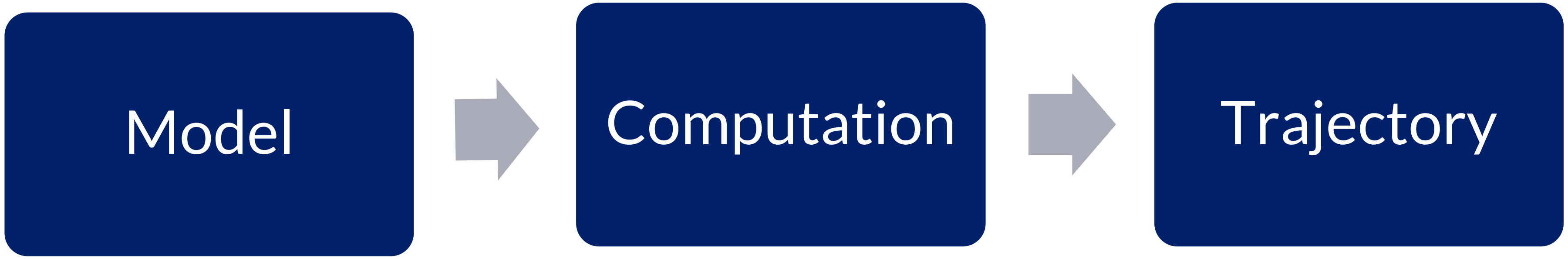
- **Other domains** have initiatives underway **in the UK**, e.g.
  - EBI in Life Sciences
  - NERC Data centres in Environmental Science
  - UK Data Archive in Social Science

Physical Sciences
catch up

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 through federated systems

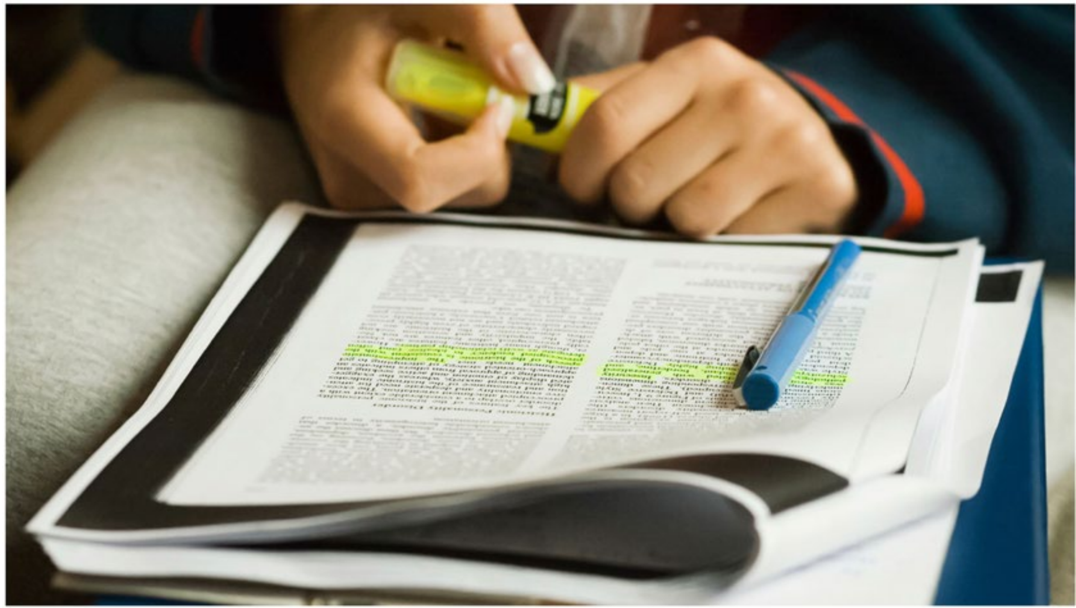
An Example:
Biomolecular Simulations



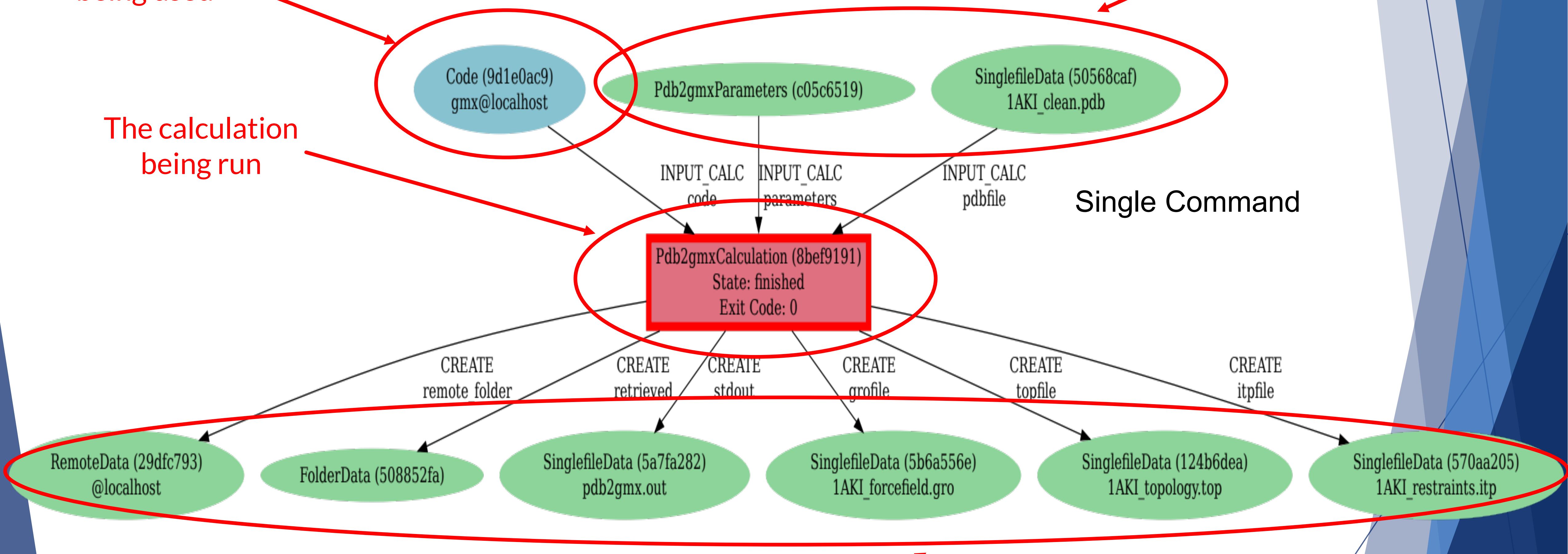
- Run 10s of simulations to generate data
- Apply know-how to extract science from data
- Publish paper

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

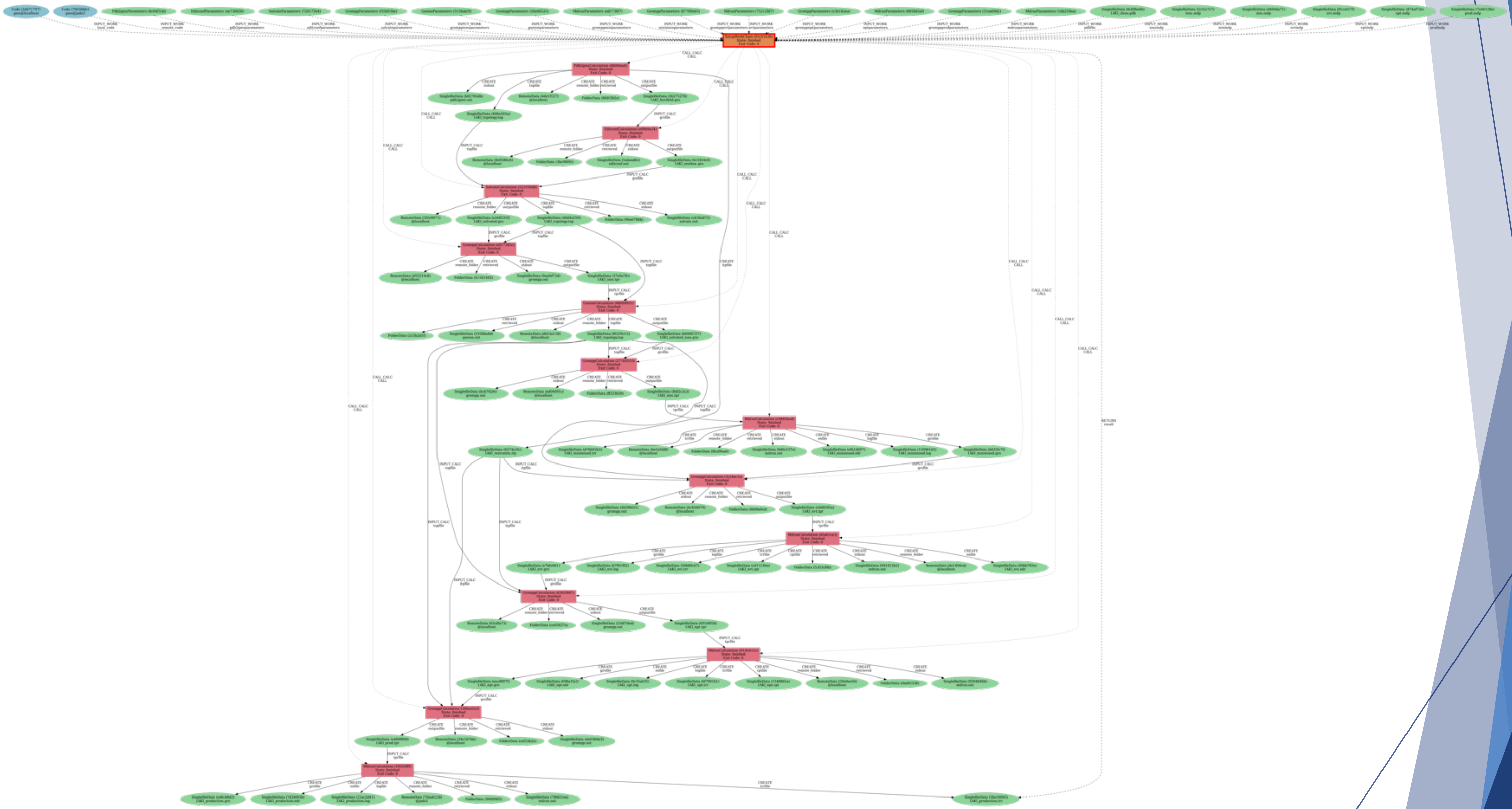
The inputs

The calculation being run

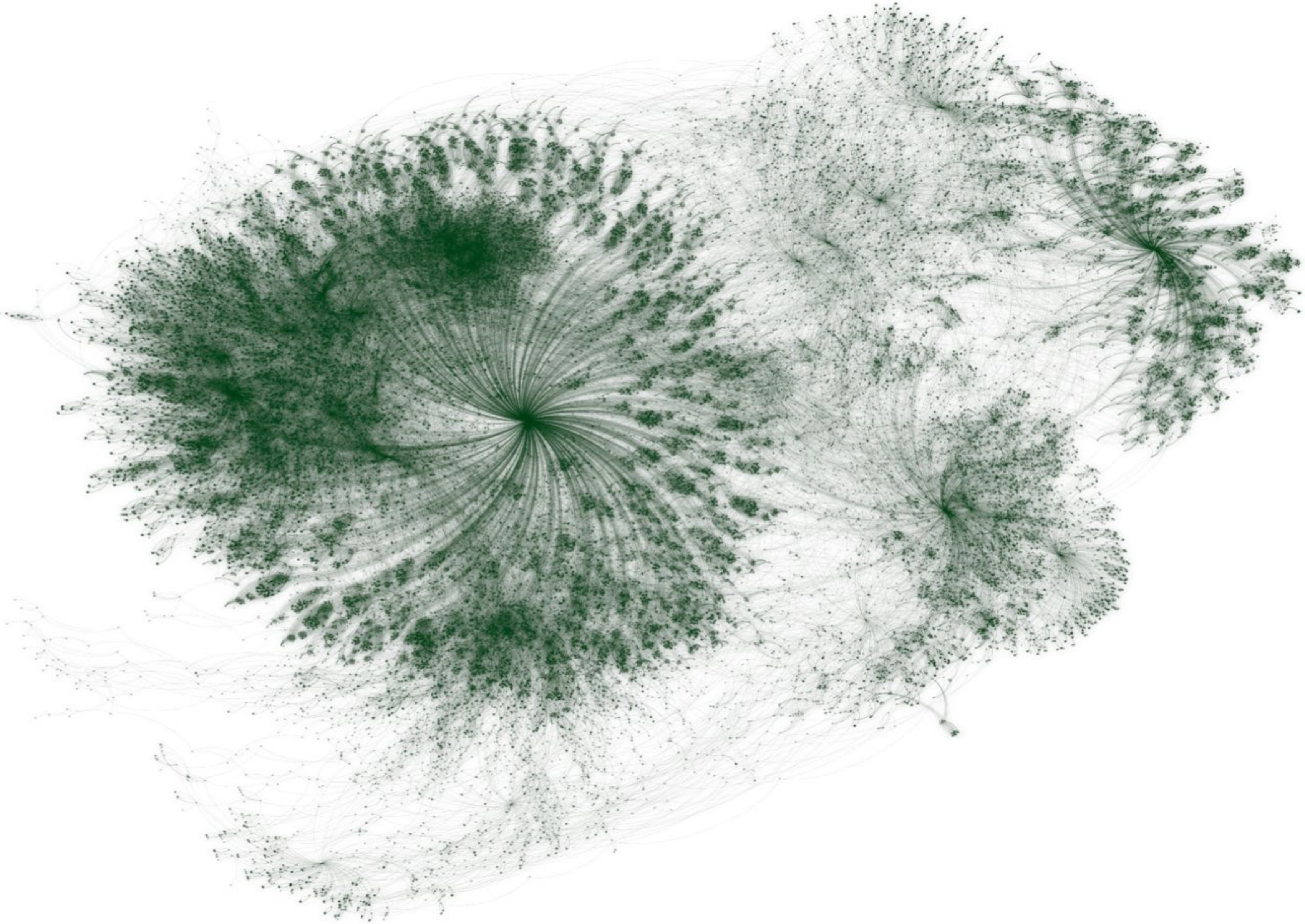
Single Command

The outputs

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



An Entire Study



PSDI
PHYSICAL SCIENCES
DATA INFRASTRUCTURE

PSDI PathFinder on Research Process Orchestration

Main aim is to improve data practices in domain – align with FAIR principles

- Prototype tools to **capture full data provenance** for model creation, simulation and analytics (FAIR)
- Prototype infrastructure tools to **store, access, find and share** data (FAIR)
- **Collect** and Integrate existing small scale, disparate data sources
- Maintain **compatibility** with other data initiatives (EBI, EU and US)
- Link **computational and experimental** data sources
- "I" (FAIR) **Integrations** *not yet in scope* of this pathfinder (excellent projects in CCPBioSim)

James Gebbie & Jas Kalayan

PSDI
PHYSICAL SCIENCES
DATA INFRASTRUCTURE

Process Orchestration PathFinder: User Environment Prototype

- Building on GROMACS software (70% of users in UK HPC Biosim Consortium
- Designed to mimick working with native package (command line driven)
- Simple to install and setup our plugin "`pip install aiida-gromacs`" – available through AiiDA

Normal command:

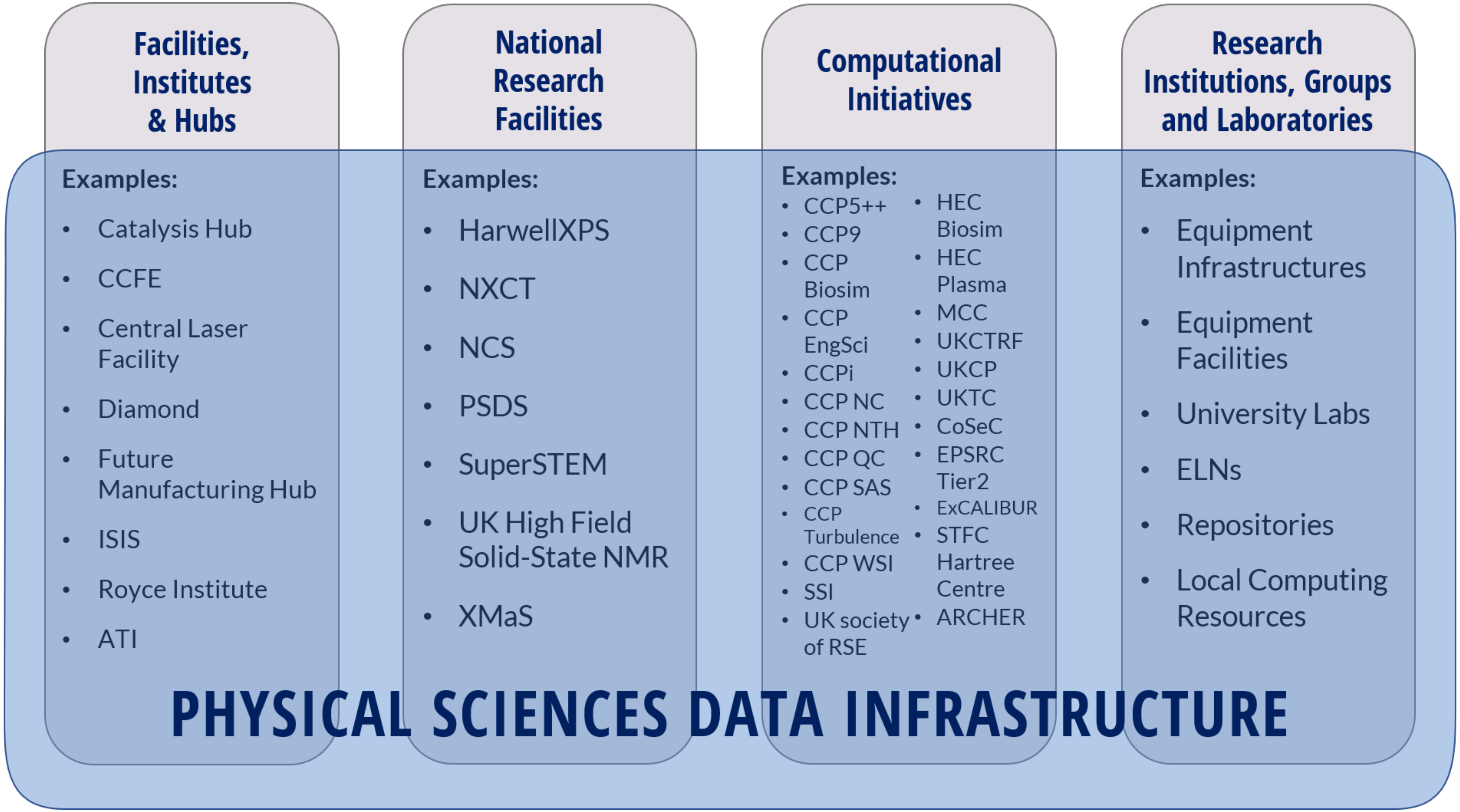
gmx pdb2gmx -f prot.pdb -ff oplsaa -water spce -o prot.gro -p prot.top -i prot.itp

Capture provenance with AiiDA:

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

PSDI
PHYSICAL SCIENCES
DATA INFRASTRUCTURE

The bigger picture (just 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

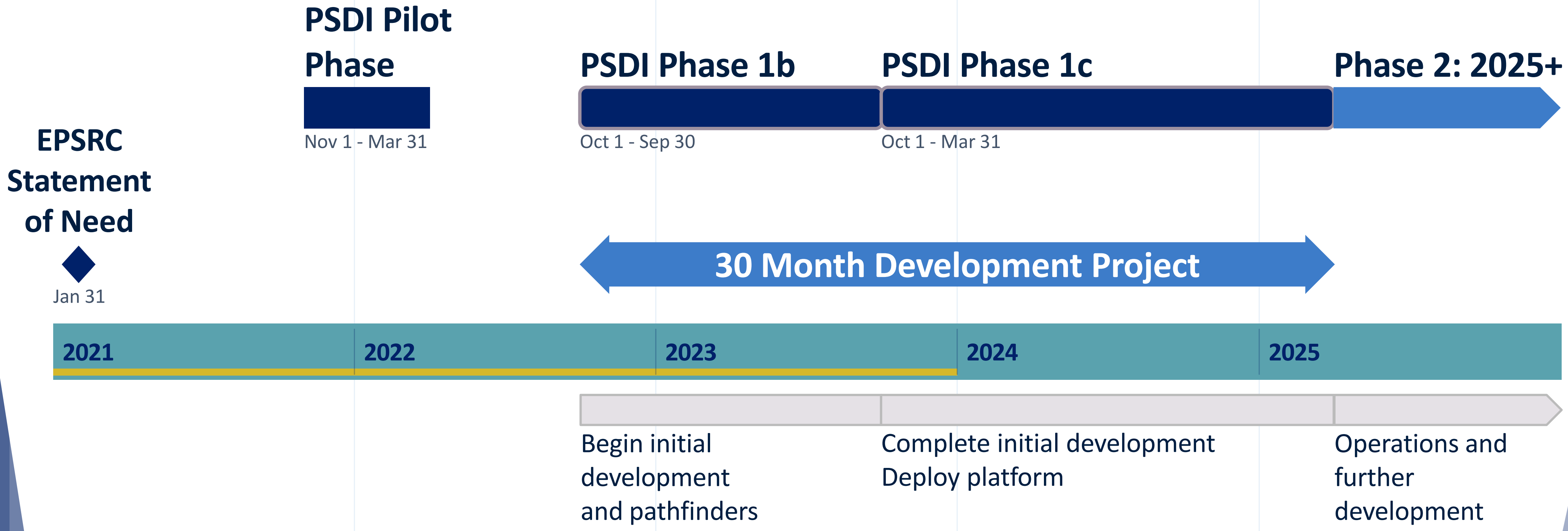
Examples:

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

PHYSICAL SCIENCES DATA INFRASTRUCTURE

PSDI
PHYSICAL SCIENCES
DATA INFRASTRUCTURE

PSDI: Outline Timeline



EPSRC Statement of Need

Jan 31

PSDI Pilot Phase

Nov 1 - Mar 31

PSDI Phase 1b

Oct 1 - Sep 30

PSDI Phase 1c

Oct 1 - Mar 31

Phase 2: 2025+

30 Month Development Project

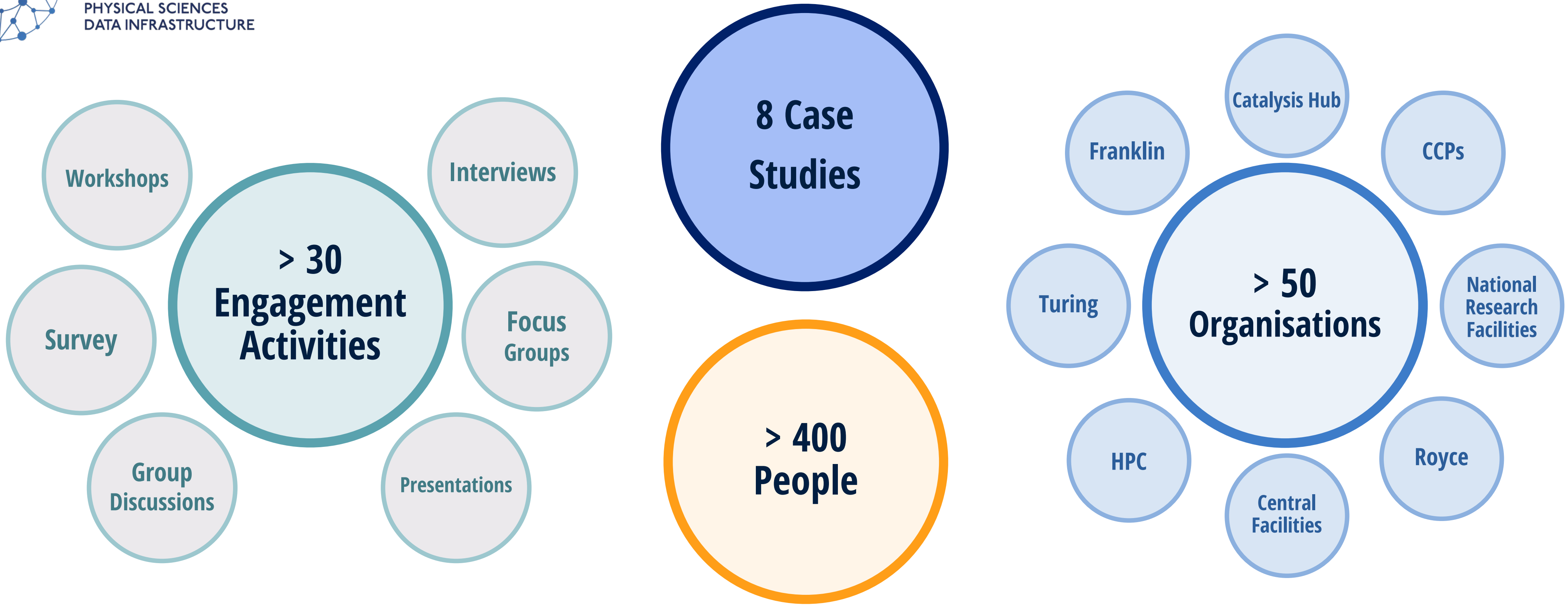
Begin initial development and pathfinders

Complete initial development
Deploy platform

Operations and further development

PSDI
PHYSICAL SCIENCES
DATA INFRASTRUCTURE

Pilot Phase at a Glance



> 30 Engagement Activities

Workshops

Interviews

Survey

Focus Groups

Group Discussions

Presentations

8 Case Studies

> 400 People

> 50 Organisations

Catalysis Hub

Franklin

CCPs

Turing

National Research Facilities

HPC

Royce

Central Facilities

CS1: Data and simulation driven understanding of **catalytic** activity
CS2: Simulations driven **materials discovery**
CS3: Combining data sources in **Materials Physics**
CS4: **Spectroscopy** data infrastructure
CS5: **Data curation** and availability at instrument-based facilities
CS6: Process Recording and **Electronic Laboratory Notebooks**
CS7: Data **trust**, sharing & **preservation**
CS8: The **role of structure** in Physical Sciences data management

PSDI
PHYSICAL SCIENCES
DATA INFRASTRUCTURE

Pilot Recommendations

13 recommendations in 4 areas:

Connecting existing infrastructures

3 Recommendations: connecting existing research data services, beyond the lifespan of individual projects, co-operation and co-creation between all stakeholder organisations

Best Use of Data

4 Recommendations: developing a toolkit for publishing, access to provenanced data, tools for reproduceable data processing, support for transforming data to knowledge

Best Use of People

4 Recommendations: co-ordination for community activities and input, community training and support, professionalisation for data roles, governance structure for PSDI

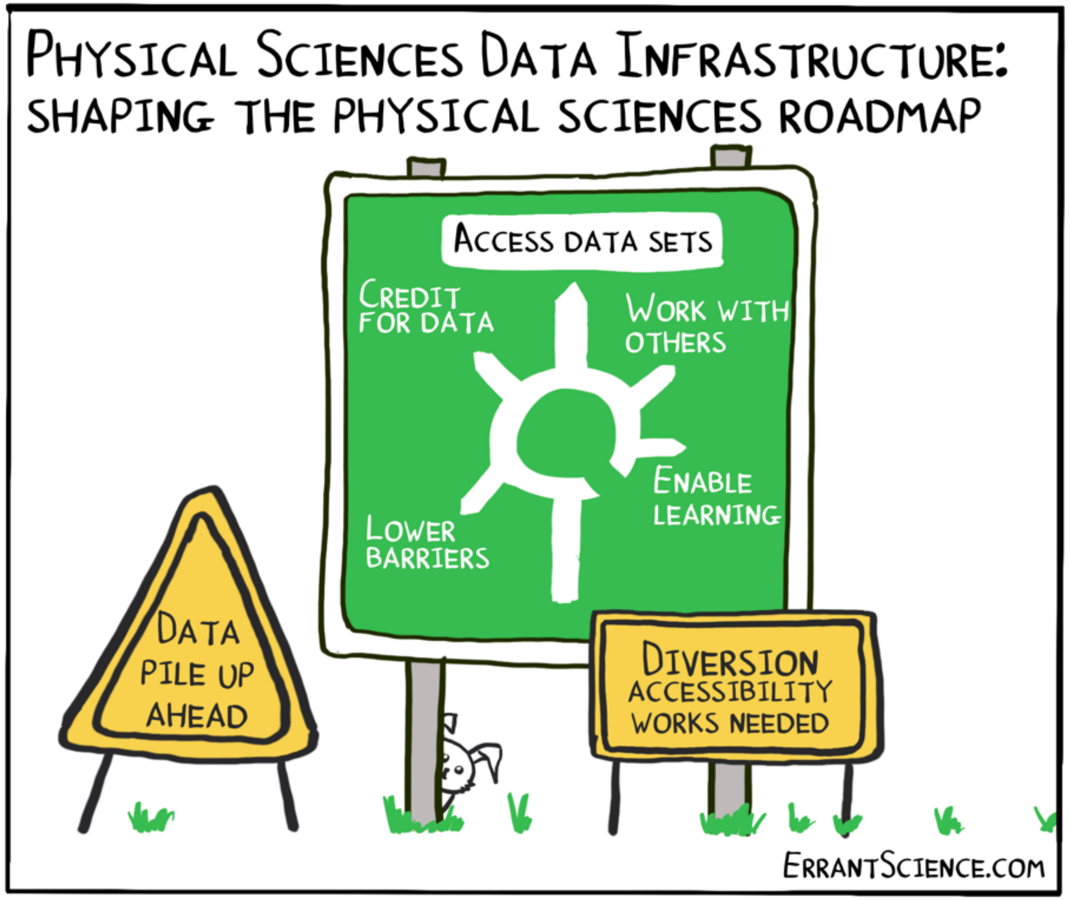
Best Use of Technology

2 Recommendations: services to connect existing provision (data and services), adopt existing technologies

Full recommendations at: https://www.psdi.ac.uk/the-pilot/recommendations
Outputs available via www.psdi.ac.uk and PSDI zenodo community

Current Work –
Platform, Pathfinders and Hub

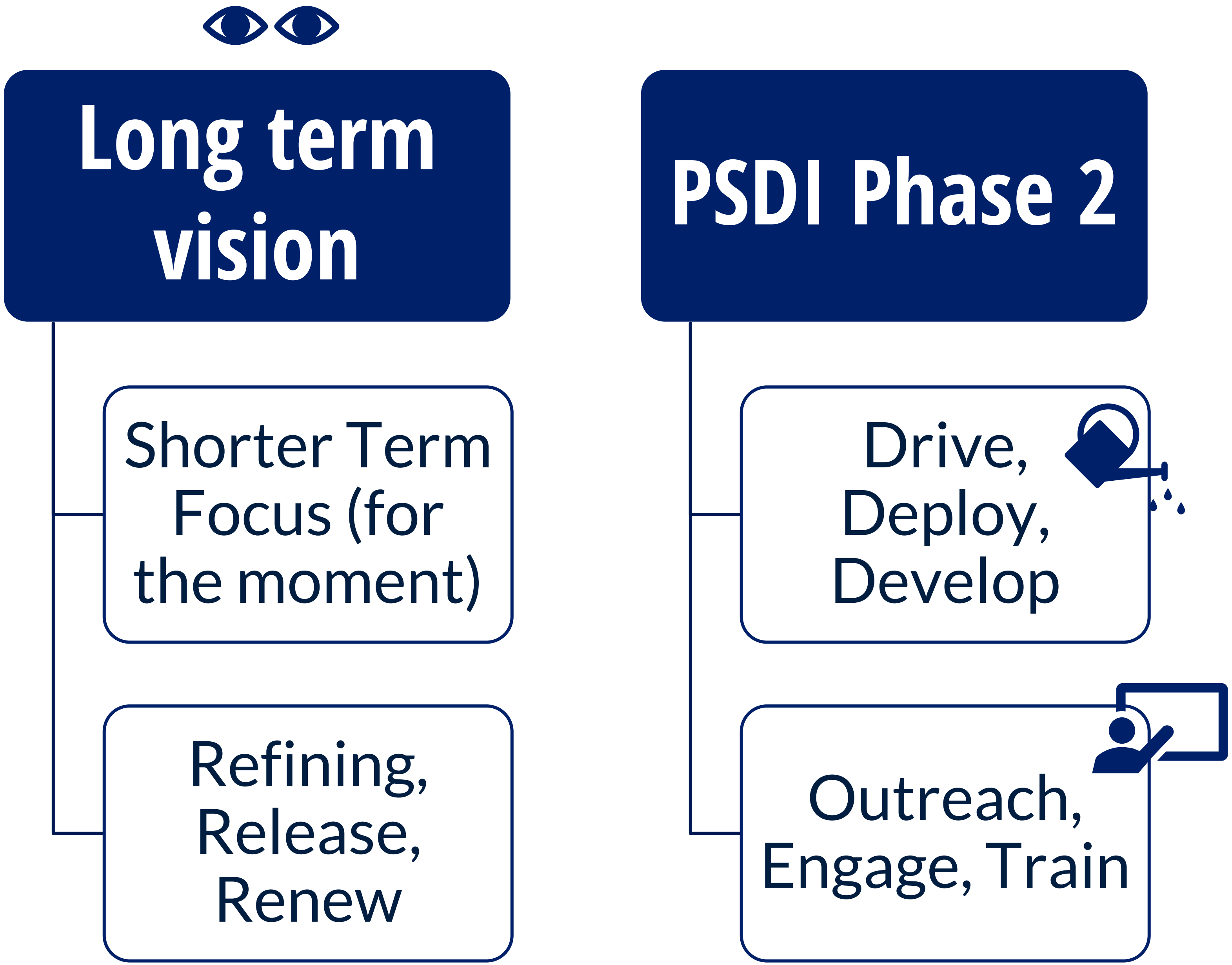
- **Platform**
  - Requirements Analysis
  - Capacity Planning
  - System Architecture design
  - Component testing
  - Beginning Build
- **“Pathfinders”**
  - PF1: Experimental data capture
  - PF2: Process Recording
  - PF3: Building Data Collections
  - PF4: Process Orchestration
  - PF5: Data to Knowledge
  - PF6: CCP-NC Database
  - PF7: Reproducible Computational Workflows
- **Hub**: Communications, Governance, Planning,…


PHYSICAL SCIENCES DATA INFRASTRUCTURE:
SHAPING THE PHYSICAL SCIENCES ROADMAP
ACCESS DATA SETS
CREDIT FOR DATA
WORK WITH OTHERS
ENABLE LEARNING
LOWER BARRIERS
DATA PILE UP AHEAD
DIVERSION ACCESSIBILITY WORKS NEEDED
ERRANTSCIENCE.COM

PSDI
PHYSICAL SCIENCES
DATA INFRASTRUCTURE

PSDI Hub
Core Activities & Services

- Management, Governance &coordination
- Core data infrastructure components
- Communications and Engagement
- Training



Long term vision

Shorter Term Focus (for the moment)

Refining, Release, Renew

PSDI Phase 2

Drive, Deploy, Develop

Outreach, Engage, Train

PSDI
PHYSICAL SCIENCES
DATA INFRASTRUCTURE

International Collaboration

Research and data is not bounded by international borders!

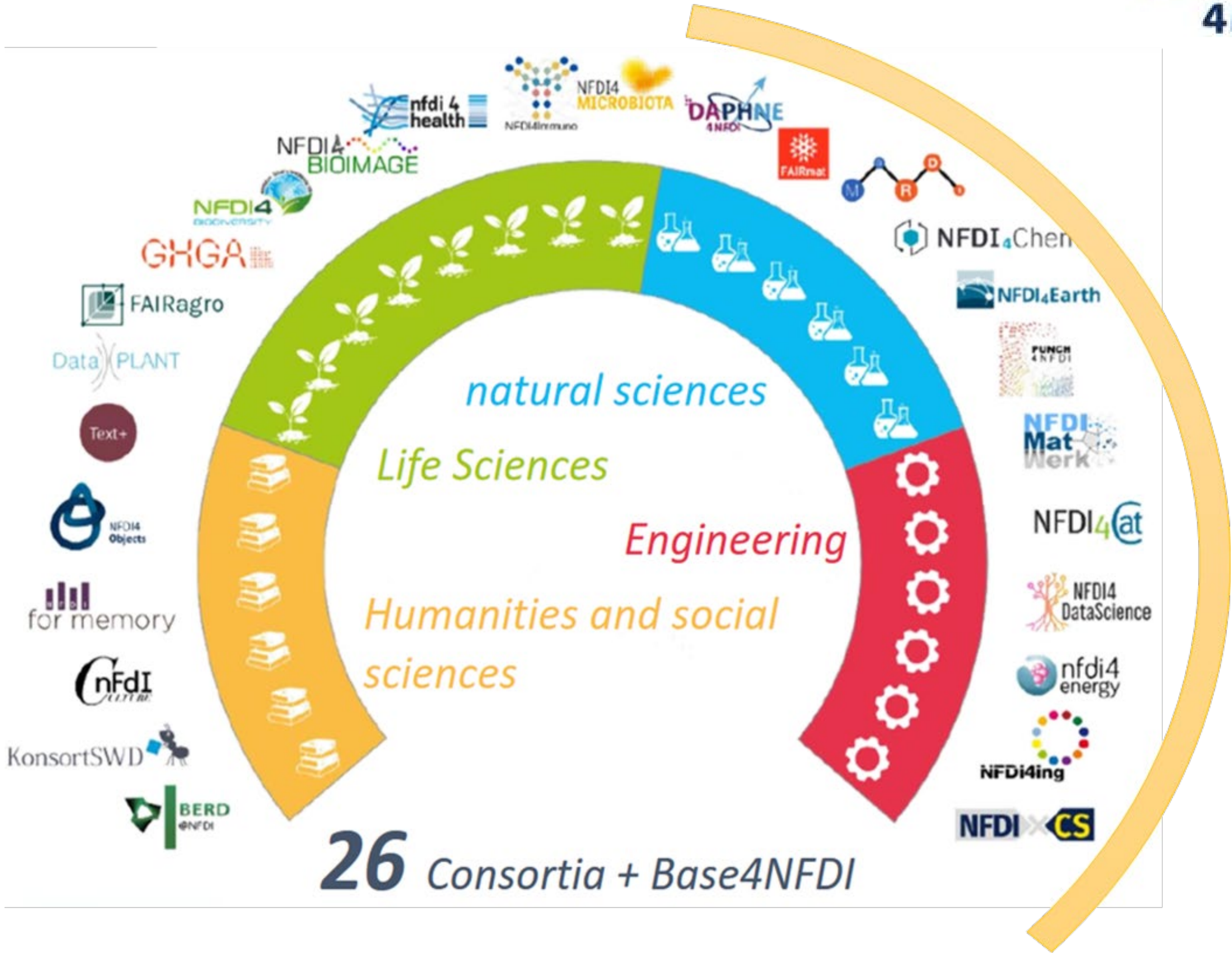
Alignment with other ongoing and developing international projects

CODATA, RDA, WorldFAIR engagement (among others)

How might we align / collaborate with NFDI?

PSDI
PHYSICAL SCIENCES
DATA INFRASTRUCTURE

Physical Sciences in nfdi


natural sciences

Life Sciences

Engineering

Humanities and social sciences

26 Consortia + Base4NFDI

DAPHNE 4NFDI

FAIRmat

NFDI4Chem

PUNCH 4NFDI

NFDI MatWerk

NFDI4Cat
NFDI for Catalysis-Related Sciences

NFDI4ing

PF1: Experimental data capture

Lead: Abraham Nieva de la Hidalga

Goal

Improve data publishing practices to promote better use of this valuable resource

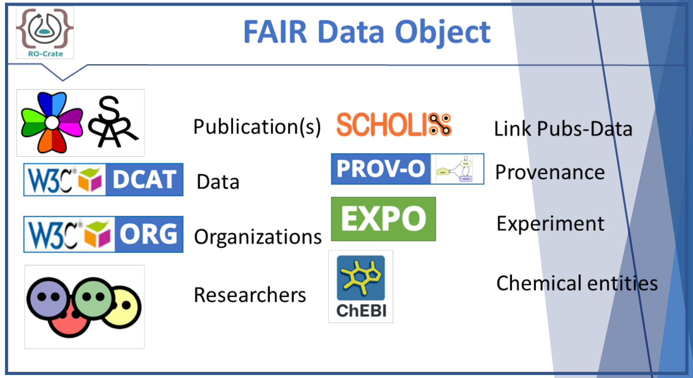
Requirements

- track data provenance
- reference all data used
- link data to results
- generate publish ready data

Proposed approaches

Create FAIR data objects

- Enable reproducibility and replicability
- Promote data reuse



Scientific workflows

- Create custom processing/analysis tools
- Combine tools into workflows
- Share and publish workflows
- Generate FAIR digital objects

Webinar recording available on @PSDI_UK YouTube focusing on this PF

PSDI
PHYSICAL SCIENCES
DATA INFRASTRUCTURE

CatalysisHub

PF1 - Opportunities for collaboration

NOMAD — Data management platform

NFDI4Cat
NFDI for Catalysis-Related Sciences

voc4cat
reac4cat

Vocabulary and ontology tools tailored to the needs of a research community

LabMotion — ELN for documenting custom experimental workflows

ROCK-IT DESY. KIT
Karlsruher Institut für Technologie

Automate beamline experiments and accelerate operation

PSDI
PHYSICAL SCIENCES
DATA INFRASTRUCTURE

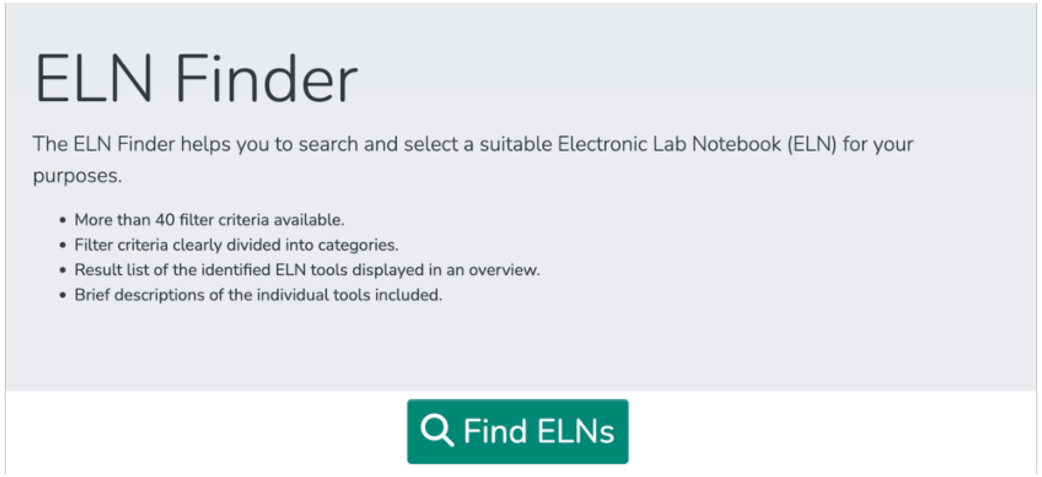
PF2: Process Recording

Lead: Dr Samantha Pearman-Kanza

Investigating routes for recording research process, as well as developing metadata and ontology layers to enable processing and analysis tools

Research focus/service areas:

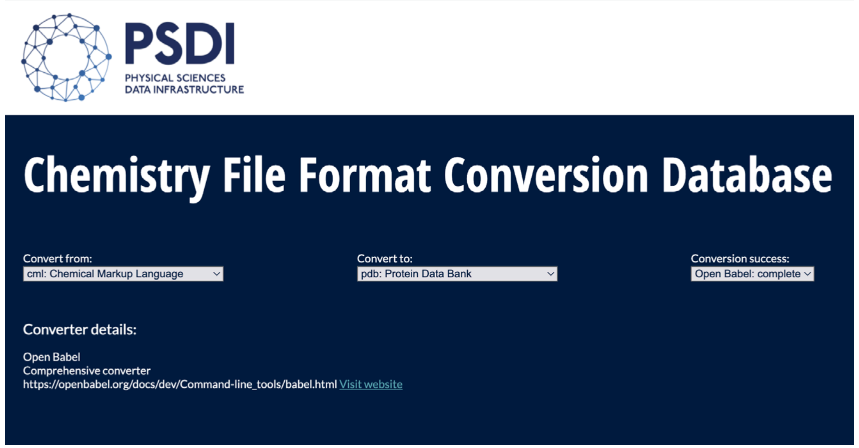
- Process recording tools
  - ELNS & generic notebooks
  - Investigating the data trail from Lab Notebook to Thesis/Paper to Supplementary Information
- Exemplars for FAIR data/software/research
- Data format conversion service
- Converting paper lab notebooks into machine-readable data using Data Revival
- Metadata & semantics research



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.



PSDI
PHYSICAL SCIENCES
DATA INFRASTRUCTURE

Chemistry File Format Conversion Database

Convert from: cml: Chemical Markup Language

Convert to: pdb: Protein Data Bank

Conversion success: Open Babel: complete

Converter details:

Open Babel
Comprehensive converter
https://openbabel.org/docs/dev/Command-line_tools/babel.html Visit website

Webinar recording available on @PSDI_UK YouTube focusing on this PF

PSDI
PHYSICAL SCIENCES
DATA INFRASTRUCTURE

PF2: Process Recording – Collaborations & NFDI Alignment

Current Collaborations:

- ELNFinder – enabling scientists to choose between ELNs

NFDI Alignment:

NFDI4Chem

- NFDI4Chem
  - FAIR data publishing – this aligns with our research for exemplars on FAIR data/software/research
  - Chemotion – opportunities to use Chemotion for case studies – aligns with ELN research
  - Terminology Service – aligns with semantics research
- FAIRMAT – FAIR data
- DAPHNE - metadata

FAIRmat

DAPHNE
4NFDI

PSDI
PHYSICAL SCIENCES
DATA INFRASTRUCTURE

PF3: Building Data Collections

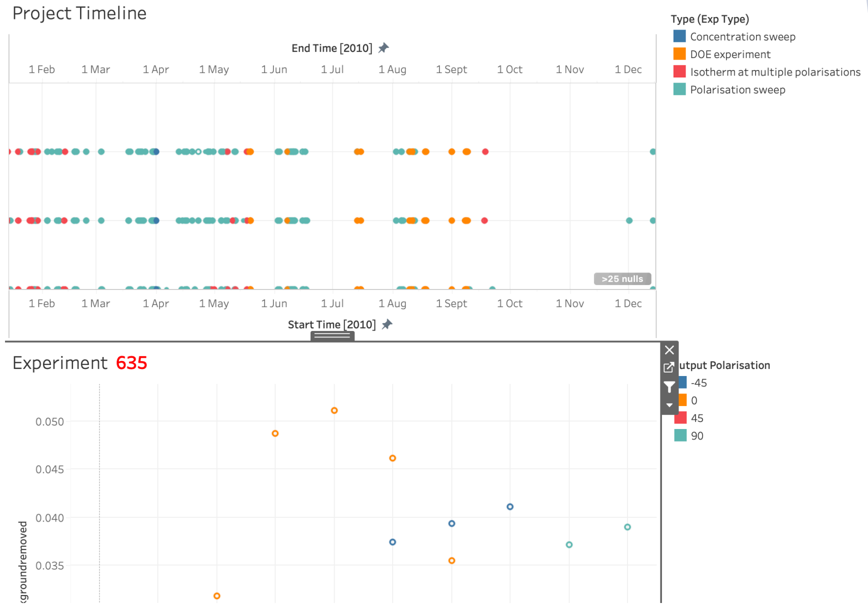
Lead: Professor Jeremy Frey

Explore and develop methods to **build**, **store**, **manage** and access collections for types of data, such as:

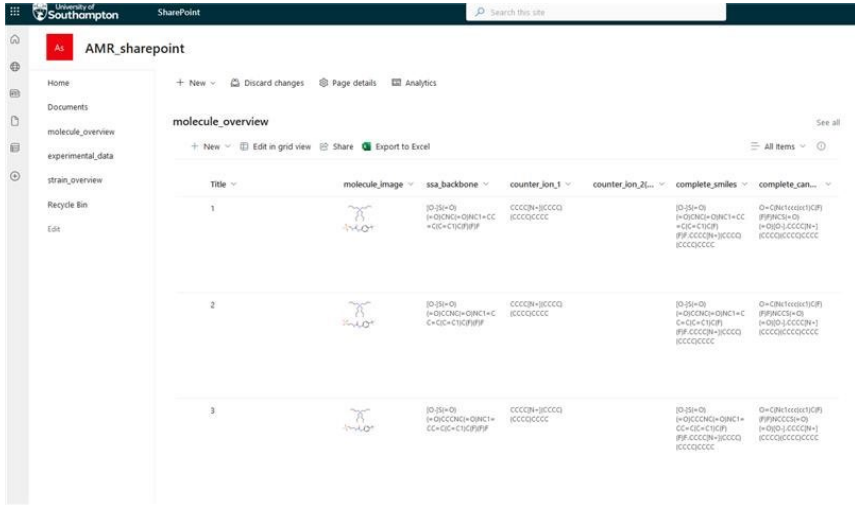
- institutional data
- facilities data
- legacy data
- orphaned data

Several use cases are being worked on:

- Multiple data types (legacy/active/paper/pdf/electronic/structured)
- How best to manage, curate and store data
- Working with a range of tools and technologies
- These use cases will enable us to provide guidelines on data practices e.g.
  - Database creation
  - Chemical identifiers
  - Data publishing
- Production of high quality curated datasets
- Investigating the available repositories (institutional and domain based)



Legacy Second Harmonic Generation data now curated and on Tableau



AMR_sharepoint

molecule_overview

Title	molecule_image	ssa_backbone	counter_ion_1	counter_ion_2(...	complete_smiles	complete_can...
1		[O-]S(=O)(=O)CNC(=O)NC1=CC=C(C=C1)C(F)(F)F	CCCC[N+](CCCC)(CCCC)CCCC		[O-]S(=O)(=O)CNC(=O)NC1=CC=C(C=C1)C(F)(F)F.CCCC[N+](CCCC)(CCCC)CCCC	O=C(Nc1ccc(cc1)C(F)(F)F)NCS(=O)(=O)[O-].CCCC[N+](CCCC)(CCCC)CCCC
2		[O-]S(=O)(=O)CCNC(=O)NC1=CC=C(C=C1)C(F)(F)F	CCCC[N+](CCCC)(CCCC)CCCC		[O-]S(=O)(=O)CCNC(=O)NC1=CC=C(C=C1)C(F)(F)F.CCCC[N+](CCCC)(CCCC)CCCC	O=C(Nc1ccc(cc1)C(F)(F)F)NCCS(=O)(=O)[O-].CCCC[N+](CCCC)(CCCC)CCCC
3		[O-]S(=O)(=O)CCCNC(=O)NC1=CC=C(C=C1)C(F)(F)F	CCCC[N+](CCCC)(CCCC)CCCC		[O-]S(=O)(=O)CCCNC(=O)NC1=CC=C(C=C1)C(F)(F)F.CCCC[N+](CCCC)(CCCC)CCCC	O=C(Nc1ccc(cc1)C(F)(F)F)NCCCS(=O)(=O)[O-].CCCC[N+](CCCC)(CCCC)CCCC

AMR data curated from spreadsheets, stored in FAIR database & Sharepoint Site

PF3: Building Data Collections – Collaborations & NFDI Alignment

NFDI Alignment:

- NFDI4Chem
  - Chemistry repositories – this aligns with our research on available repositories, and guidelines towards data publishing
- DAPHNE
  - Community repositories
- NFDI4Cat
  - Data Management in Catalysis – links to guidelines on data

NFDI4Chem

DAPHNE
4NFDI

NFDI4Cat
NFDI for Catalysis-Related Sciences

PF4: Biomolecular Simulations

Lead: Dr James Gebbie-Rayet

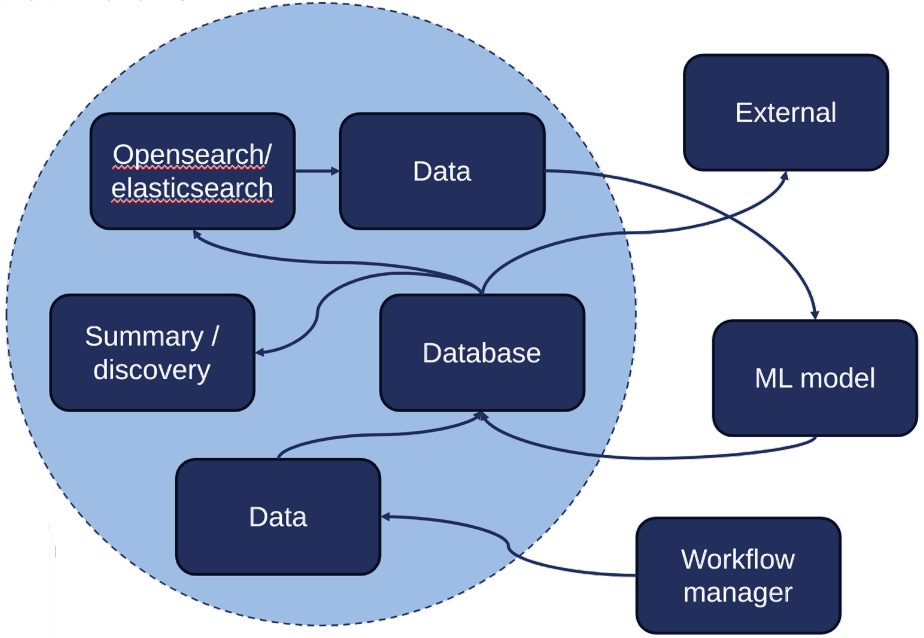
- Exemplar shown earlier in presentation
- This pathfinder aims to establish tools and an infrastructure prototype for capturing the full data provenance for biomolecular simulations, starting from experimental input data through to eventual publications.
- Most closely aligned with **NFDI4Chem** and **FAIRMAT**
- Will be working with MDDB (Molecular Dynamics Data Bank) so there will be collaboration through EBI with European partners

Webinar recording available on @PSDI_UK YouTube focusing on this PF (publishing soon)

PF5: Data to Knowledge

Lead: Dr Alin Marin Elena

- **Design and deploy hardware infrastructure** to host both training data for machine learnt interatomic potentials and the potentials
- Proof of concept database & app:
  - installable centrally within PSDI
  - blueprints for local installation
  - users can interrogate, download and deposit data
  - API and web interface
- Advanced search features: elastic search techniques
- Integrate into ML workflows



Alin Elena, Elliott Kasoar, Federica Zanca, collaboration with Gábor Csányi

Upcoming Webinar focusing on this PF
Dec 14th 1400 UTC
www.psdi.ac.uk/event/webinar-psdi-pf5/

PF6: CCP-NC Database

Lead: Dr Sathya Sai Seetharaman

Collaborative Computational Project for NMR Crystallography (https://www.ccpnc.ac.uk/)

- Supports the multidisciplinary experimental NMR community with computational tools
- Recognised by International Union of Crystallography
- Created .magres file format for unified representation of crystal NMR parameters.

Pathfinder Objectives

- Improving the current CCP-NC magres database (https://www.ccpnc.ac.uk/database/)
- Development of a state-of-the art database (version 2)

CCP-NC – FAIRmat

Productive discussion with NOMAD about potential collaboration:

- Magres database support
  - Initial discussions about magres file parser support
  - Add functionality for QE/CASTEP -> magres workflows
  - CCP-NC magres database – support to legacy data
- Adding NMR specific searchability within NOMAD
- Potential extension of services to NMR experimental community
- Data visualisation support in-line with CCP-NC tool standards
- Potential CCP-NC – NOMAD collaborative development

NOMAD

FAIRmat

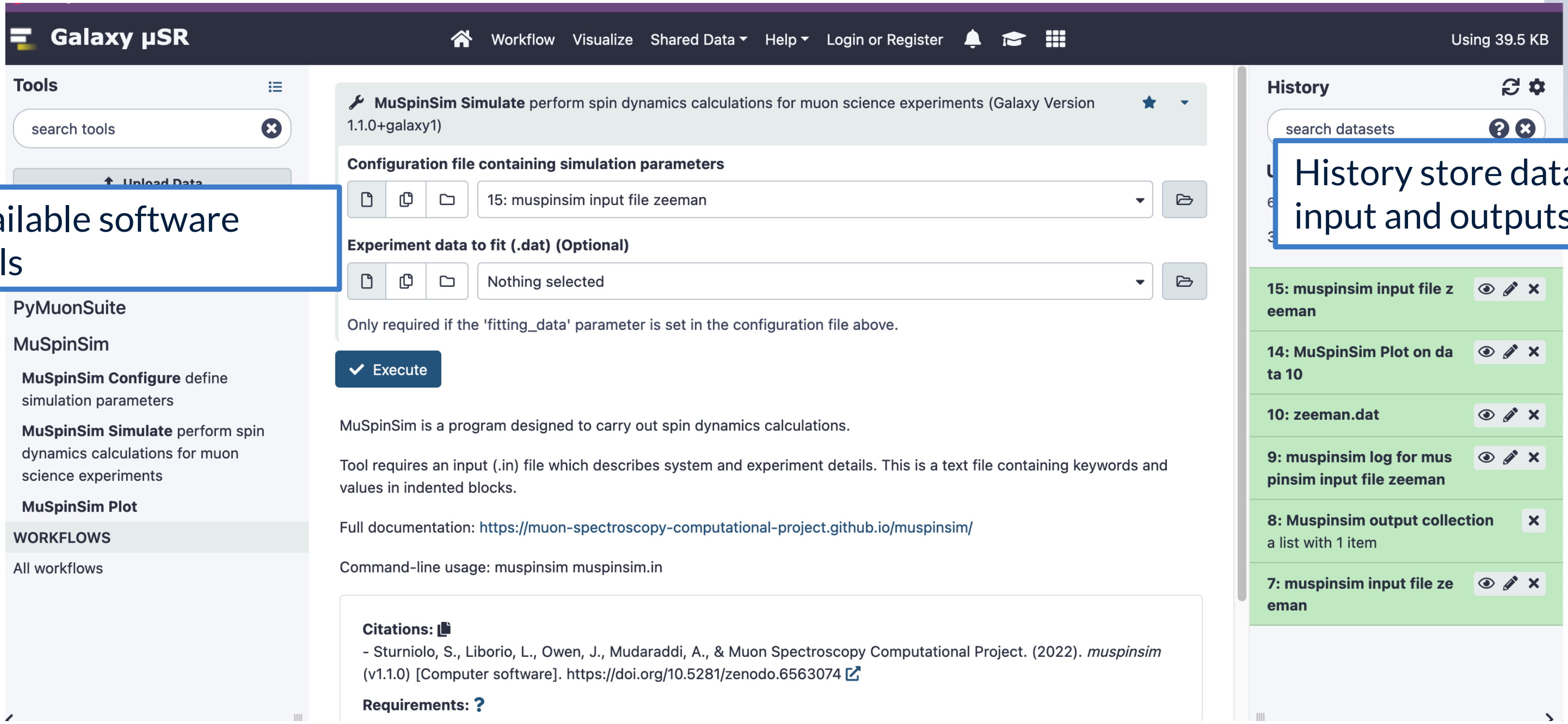
PSDI
PHYSICAL SCIENCES
DATA INFRASTRUCTURE

PF7: Reproducible Computational Workflows

Lead: Dr Leandro Liborio

Galaxy
PROJECT

https://galaxyproject.org/



Available software tools

History store data files: input and outputs

MuSpinSim Configure define simulation parameters

MuSpinSim Simulate perform spin dynamics calculations for muon science experiments

MuSpinSim Plot

MuSpinSim Simulate perform spin dynamics calculations for muon science experiments (Galaxy Version 1.1.0+galaxy1)

Configuration file containing simulation parameters

15: muspinsim input file zeeman

Experiment data to fit (.dat) (Optional)

Nothing selected

Only required if the 'fitting_data' parameter is set in the configuration file above.

MuSpinSim is a program designed to carry out spin dynamics calculations.

Tool requires an input (.in) file which describes system and experiment details. This is a text file containing keywords and values in indented blocks.

Full documentation: https://muon-spectroscopy-computational-project.github.io/muspinsim/

Command-line usage: muspinsim muspinsim.in

Citations:
- Sturniolo, S., Liborio, L., Owen, J., Mudaraddi, A., & Muon Spectroscopy Computational Project. (2022). *muspinsim* (v1.1.0) [Computer software]. https://doi.org/10.5281/zenodo.6563074

Requirements: ?

15: muspinsim input file zeeman

14: MuSpinSim Plot on data 10

10: zeeman.dat

9: muspinsim log for muspinsim input file zeeman

8: Muspinsim output collection
a list with 1 item

7: muspinsim input file zeeman

Muon Experiments

- Muon experiments are performed at the Rutherford Appleton Lab, STFC, UK.
- Develop software tools for interpretation of those muon experiments.
- Tools based on computing simulations, i.e.: DFT.
- Created associated Galaxy tools and Galaxy instance.
- Use the Galaxy platform to manage the workflows resulting from the tools.

X-ray Absorption Spectroscopy (XAS) Experiments

- XAS experiments are performed at the Rutherford Appleton Lab, STFC, UK. Catalysis-related experiments.
- Software tools for processing experimental data already available.
- Created associated Galaxy tools and Galaxy instance.
- Use the Galaxy platform to manage the workflows resulting from the tools

Potential Collaborations/Alignment Between PF7 and NFDI

NFDI4Cat

- Present galaxy tools as a complementary method for processing workflows. Currently working with A. Nieva de la Hidalga (Pathfinder 1) on galaxy for catalysis experiments.

DAPHNE4NFDI

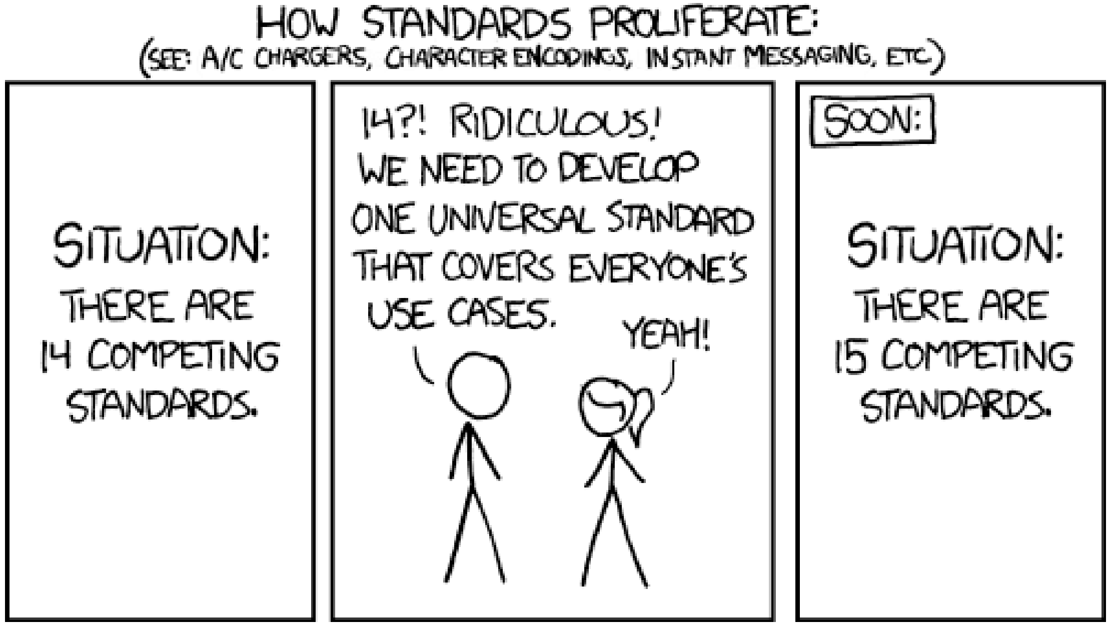
- Task area 3 from DAPHNE4NFDI refers to “Infrastructure for Data and Software Reuse”. Galaxy is an open, web-based platform for accessible, reproducible, and transparent computational research.
- We are collaborating with colleagues from Oak Ridge National Lab on Galaxy tools for neutron science.
- We are working with the Diamond Light Source on Galaxy tools for x-ray experiments.

NFDI MatWerk

- IUC05 Digital infrastructure and workflows for labs.
- PP13 Tomography and Microstructure-based Modelling.
  “a workflow for the transfer of tomography data to related multi-scale simulations will be established”.
- IUC15 Method- and scale-bridging workflows and data structures for tomography.
  “Materials tomography methods and resulting data vary strongly depending on the method used, the experimental approach and the workflow for post-processing. Currently, there is no established protocol which would allow to conduct all necessary steps in a well-defined manner. The resulting data from different methods are therefore not interconnected and workflows are intransparent.”

Cross Project Topics

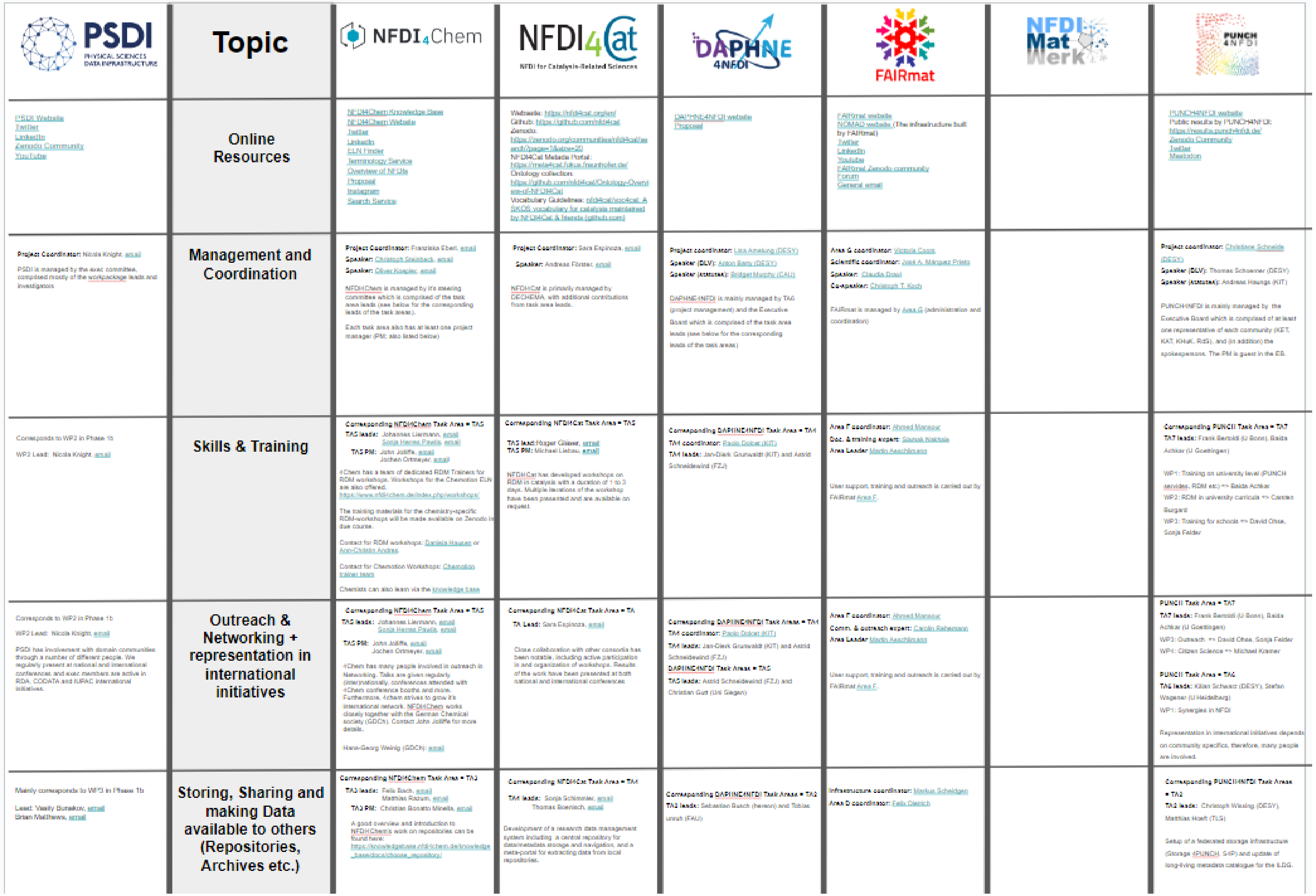
- Best practices
- Skills / Training
- Standards
  - Files formats
  - Metadata
- Publishing / Sharing
- Semantics / Ontologies


HOW STANDARDS PROLIFERATE:
(SEE: A/C CHARGERS, CHARACTER ENCODINGS, INSTANT MESSAGING, ETC.)
SITUATION: THERE ARE 14 COMPETING STANDARDS.
14?! RIDICULOUS! WE NEED TO DEVELOP ONE UNIVERSAL STANDARD THAT COVERS EVERYONE'S USE CASES.
YEAH!
SOON:
SITUATION: THERE ARE 15 COMPETING STANDARDS.

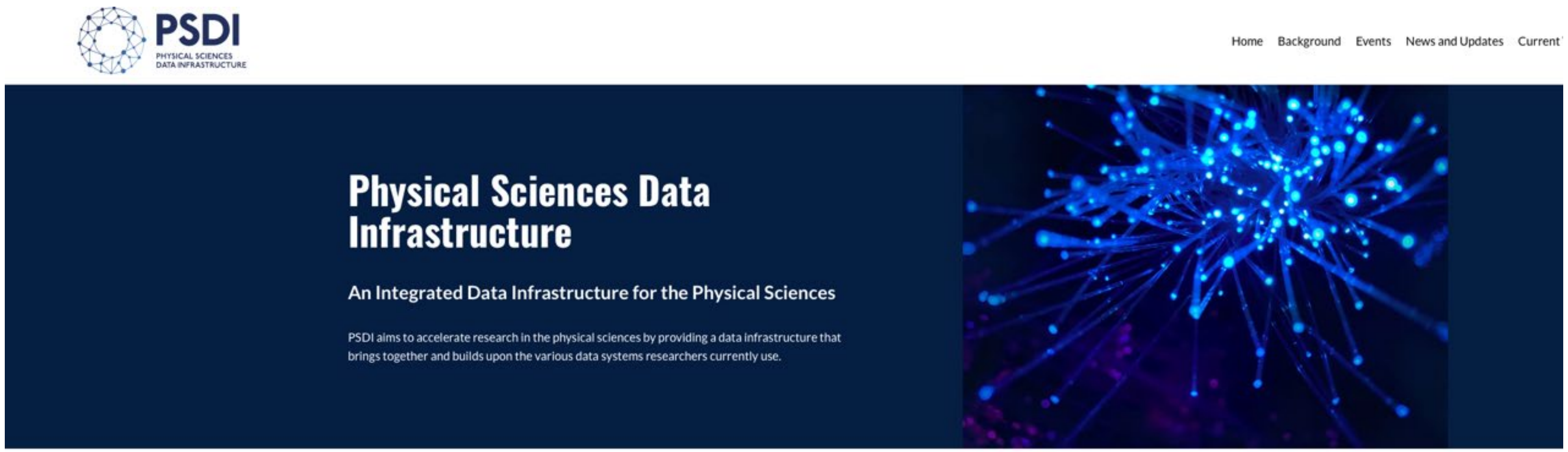
XKCD – Standards, https://xkcd.com/927/ -Creative Commons Attribution-NonCommercial 2.5 License

PSDI
PHYSICAL SCIENCES
DATA INFRASTRUCTURE

Enabling communication



PSDI PHYSICAL SCIENCES DATA INFRASTRUCTURE	Topic	NFDI4Chem	NFDI4Cat NFDI for Catalysis-Related Sciences	DAPHNE 4NFDI	FAIRmat	NFDI MatWerk	PUNCH 4NFDI
PSDI Website Twitter LinkedIn Zenodo Community YouTube	Online Resources	NFDI4Chem Knowledge Base NFDI4Chem Website Twitter LinkedIn ELN Finder Terminology Service Overview of NFDIs Proposal Instagram Search Service	Website: https://nfdi4cat.org/en/ Github: https://github.com/nfdi4cat Zenodo: https://zenodo.org/communities/nfdi4cat/search?page=1&size=20 NFDI4Cat Metadata Portal: https://meta4cat.fokus.fraunhofer.de/ Ontology collection: https://github.com/nfdi4cat/Ontology-Overview-of-NFDI4Cat Vocabulary Guidelines: nfdi4cat/voc4cat: A SKOS vocabulary for catalysis maintained by NFDI4Cat & friends (github.com)	DAPHNE4NFDI website Proposal	FAIRmat website NOMAD website (The infrastructure built by FAIRmat) Twitter LinkedIn Youtube FAIRmat Zenodo community Forum General email		PUNCH4NFDI website Public results by PUNCH4NFDI: https://results.punch4nfdi.de/ Zenodo Community Twitter Mastodon
Project Coordinator: Nicola Knight, email PSDI is managed by the exec committee, comprised mostly of the workpackage leads and investigators	Management and Coordination	**Project Coordinator:** Franziska Ebert, email **Speaker:** Christoph Steinbeck, email **Speaker:** Oliver Koepler, email NFDI4Chem is managed by it's steering committee which is comprised of the task area leads (see below for the corresponding leads of the task areas). Each task area also has at least one project manager (PM; also listed below)	**Project Coordinator:** Sara Espinoza, email **Speaker:** Andreas Förster, email NFDI4Cat is primarily managed by DECHEMA, with additional contributions from task area leads.	**Project coordinator:** Lisa Amelung (DESY) **Speaker (DLV):** Anton Barty (DESY) **Speaker ([illegible]):** Bridget Murphy (CAU) DAPHNE4NFDI is mainly managed by TA6 (project management) and the Executive Board which is comprised of the task area leads (see below for the corresponding leads of the task areas.)	**Area G coordinator:** Victoria Coors **Scientific coordinator:** José A. Márquez Prieto **Speaker:** Claudia Draxl **Co-speaker:** Christoph T. Koch FAIRmat is managed by Area G (administration and coordination)		**Project coordinator:** Christiane Schneide (DESY) **Speaker (DLV):** Thomas Schoerner (DESY) **Speaker ([illegible]):** Andreas Haungs (KIT) PUNCH4NFDI is mainly managed by the Executive Board which is comprised of at least one representative of each community (KET, KAT, KHuK, RdS), and (in addition) the spokespersons. The PM is guest in the EB.
Corresponds to WP2 in Phase 1b WP2 Lead: Nicola Knight, email	Skills & Training	**Corresponding NFDI4Chem Task Area = TA5** **TA5 leads:** Johannes Liermann, email Sonja Herres-Pawlis, email **TA5 PM:** John Jolliffe, email Jochen Ortmeyer, email 4Chem has a team of dedicated RDM Trainers for RDM workshops. Workshops for the Chemotion ELN are also offered. https://www.nfdi4chem.de/index.php/workshops/ The training materials for the chemistry-specific RDM-workshops will be made available on Zenodo in due course. Contact for RDM workshops: Daniela Hausen or Ann-Christin Andres. Contact for Chemotion Workshops: Chemotion trainer team Chemists can also learn via the knowledge base	**Corresponding NFDI4Cat Task Area = TA5** **TA5 lead:** Roger Gläser, email **TA5 PM:** Michael Liebau, email NFDI4Cat has developed workshops on RDM in catalysis with a duration of 1 to 3 days. Multiple iterations of the workshop have been presented and are available on request.	**Corresponding DAPHNE4NFDI Task Area = TA4** **TA4 coordinator:** Paolo Dolcet (KIT) **TA4 leads:** Jan-Dierk Grunwaldt (KIT) and Astrid Schneidewind (FZJ)	**Area F coordinator:** Ahmed Mansour **Doc. & training expert:** Siamak Nakhaie **Area Leader** Martin Aeschlimann User support, training and outreach is carried out by FAIRmat Area F.		**Corresponding PUNCH Task Area = TA7** **TA7 leads:** Frank Bertoldi (U Bonn), Baida Achkar (U Goettingen) WP1: Training on university level (PUNCH services, RDM etc) => Baida Achkar WP2: RDM in university curricula => Carsten Burgard WP3: Training for schools => David Ohse, Sonja Felder
Corresponds to WP2 in Phase 1b WP2 Lead: Nicola Knight, email PSDI has involvement with domain communities through a number of different people. We regularly present at national and international conferences and exec members are active in RDA, CODATA and IUPAC international initiatives.	Outreach & Networking + representation in international initiatives	**Corresponding NFDI4Chem Task Area = TA5** **TA5 leads:** Johannes Liermann, email Sonja Herres-Pawlis, email **TA5 PM:** John Jolliffe, email Jochen Ortmeyer, email 4Chem has many people involved in outreach in Networking. Talks are given regularly (inter)nationally, conferences attended with 4Chem conference booths and more. Furthermore, 4chem strives to grow it's international network. NFDI4Chem works closely together with the German Chemical society (GDCh). Contact John Jolliffe for more details. Hans-Georg Weinig (GDCh): email	**Corresponding NFDI4Cat Task Area = TA** **TA Lead:** Sara Espinoza, email Close collaboration with other consortia has been notable, including active participation in and organization of workshops. Results of the work have been presented at both national and international conferences.	**Corresponding DAPHNE4NFDI Task Areas = TA4** **TA4 coordinator:** Paolo Dolcet (KIT) **TA4 leads:** Jan-Dierk Grunwaldt (KIT) and Astrid Schneidewind (FZJ) **DAPHNE4NFDI Task Areas = TA5** **TA5 leads:** Astrid Schneidewind (FZJ) and Christian Gutt (Uni Siegen)	**Area F coordinator:** Ahmed Mansour **Comm. & outreach expert:** Carolin Rehermann **Area Leader** Martin Aeschlimann User support, training and outreach is carried out by FAIRmat Area F.		**PUNCH Task Area = TA7** **TA7 leads:** Frank Bertoldi (U Bonn), Baida Achkar (U Goettingen) WP2: Outreach => David Ohse, Sonja Felder WP4: Citizen Science => Michael Kramer **PUNCH Task Area = TA6** **TA6 leads:** Kilian Schwarz (DESY), Stefan Wagner (U Heidelberg) WP1: Synergies in NFDI Representation in international initiatives depends on community specifics, therefore, many people are involved.
Mainly corresponds to WP3 in Phase 1b Lead: Vasily Bunakov, email Brian Matthews, email	Storing, Sharing and making Data available to others (Repositories, Archives etc.)	**Corresponding NFDI4Chem Task Area = TA3** **TA3 leads:** Felix Bach, email Matthias Razum, email **TA3 PM:** Christian Bonatto Minella, email A good overview and introduction to NFDI4Chem's work on repositories can be found here: https://knowledgebase.nfdi4chem.de/knowledge_base/docs/choose_repository/	**Corresponding NFDI4Cat Task Area = TA4** **TA4 leads:** Sonja Schimmler, email Thomas Boenisch, email Development of a research data management system including a central repository for data/metadata storage and navigation, and a meta-portal for extracting data from local repositories.	**Corresponding DAPHNE4NFDI Task Areas = TA2** **TA2 leads:** Sebastian Busch (hereon) and Tobias unruh (FAU)	**Infrastructure coordinator:** Markus Scheidgen **Area D coordinator:** Felix Dietrich		**Corresponding PUNCH4NFDI Task Areas = TA2** **TA2 leads:** Christoph Wissing (DESY), Matthias Hoeft (TLS) Setup of a federated storage infrastructure (Storage4PUNCH, S4P) and update of long-living metadata catalogue for the ILDG.



Physical Sciences Data Infrastructure

An Integrated Data Infrastructure for the Physical Sciences

PSDI aims to accelerate research in the physical sciences by providing a data infrastructure that brings together and builds upon the various data systems researchers currently use.

www.psdi.ac.uk @PSDI_UK @PSDI_UK PSDIUK

Any Questions?

?

Please do contact our researchers directly

They just love to talk about our work!