

PaNOSC and Research Data Management

12 March, 2021

Presenter: Andy Götz (PaNOSC coordinator @ ESRF)

Meeting: Battery2030+ RDM workshop

Venue: Online



Outline of talk

1. What are the benefits of Research Data Management (RDM)?
2. How is PaNOSC approaching RDM?
3. Requirements for FAIR RDM
4. Lessons learned



What are the benefits of RDM

1. What do I get out of this as a researcher? Why is all this even worth the effort?
2. What do we achieve that could not be achieved without RDM, FAIR data and guidelines?
3. What can we learn from other disciplines (e.g., pharmaceutical research, astrophysics etc.) that have perhaps faced similar issues in connecting the community and have mastered those?
4. What can be practically done to move forward? What should be the steps the (European) battery community might want to take, in order to benefit from RDM, FAIR data, and from developing harmonized guidelines and standards for experimental and modelling work?



RDM = Making the most of data

“Cost of curating research data is a fraction of the cost of producing the data (in many cases)”

“The cheapest experiment is the one already in the database”



ELIXIR Europe
@ELIXIREurope

...

"The cheapest experiment is the one already in the database".

Our next [#KeynoteSpeaker](#) Francois Tardieu from [@EMPHASIS_EU](#) discusses how to move from [#sensor](#) data to [#FAIRdata](#). [@INRAE_France](#) [#ELIXIR4Innovation](#)
[Aistrigh an Tweet](#)

11:51 r.n. · 11 Márta 2021 · TweetDeck

2 Atweet 9



Reproducibility crisis in science

NEWS FEATURE

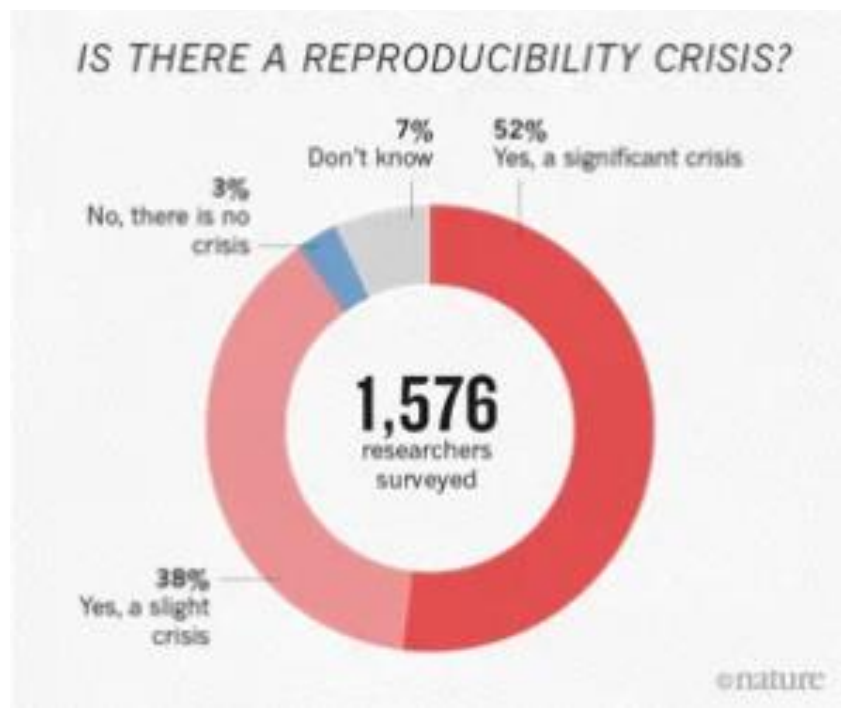
25 MAY 2016

Nature

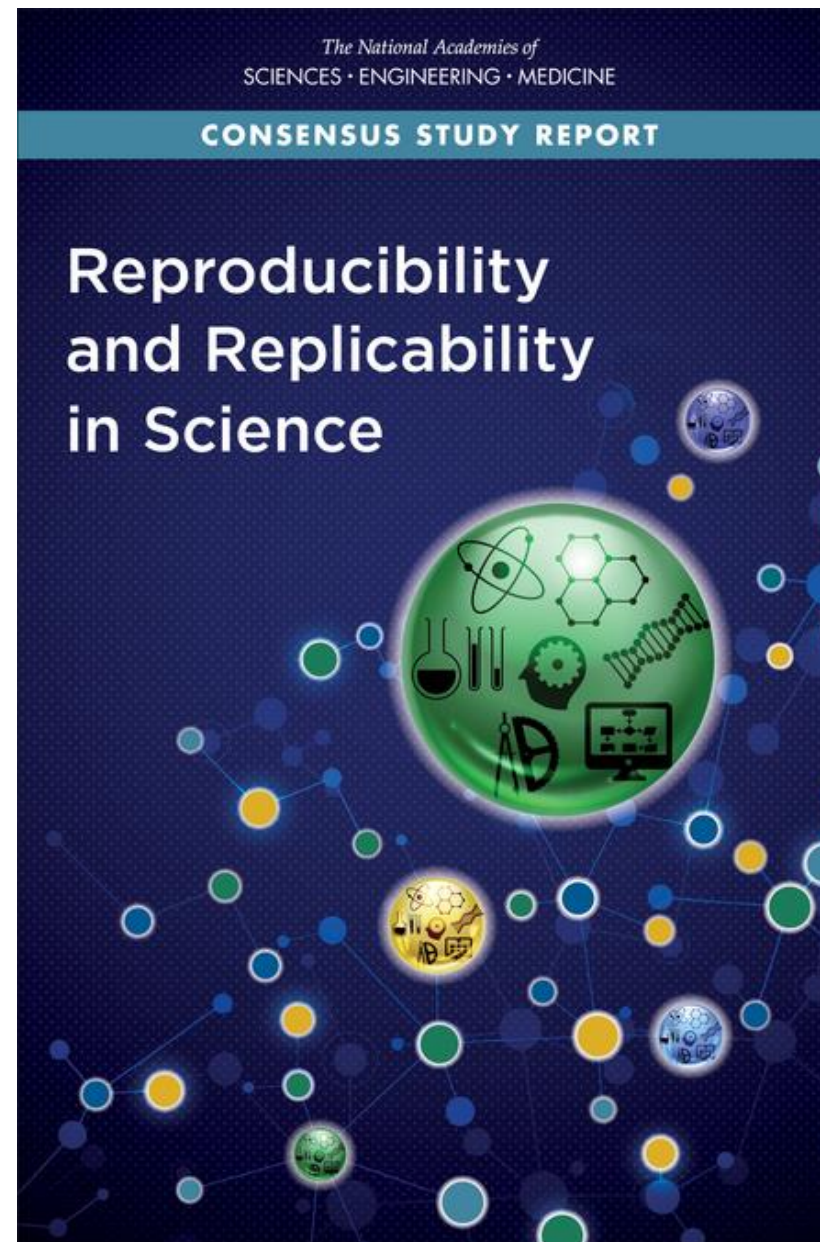
1,500 scientists lift the lid on reproducibility

Survey sheds light on the 'crisis' rocking research.

Monya Baker



Baker, M. 1,500 scientists lift the lid on reproducibility. *Nature* **533**, 452–454 (2016) doi:10.1038/533452a



PaNOSC factsheet

Call: Horizon 2020 InfraEOSC-04

Partners: ESRF, ILL, XFEL.EU, ESS, CERIC-ERIC, ELI-DC, EGI

Description: cluster of ESFRI Photon and Neutron sources

Observers/non-funded: GÉANT, EUDAT, national Ris

Co-funded project: ExPaNDS (national RIs)

Linked 3rd parties via EGI: DESY, STFC, CESNET

Status: Started 1/12/2018

Github: <https://github.com/panosc-eu>

Home page: <https://panosc.eu>

Twitter: @PaNOSC_eu #PaNOSC

Budget: 12 M€

Coordinator: ESRF

Started: 1/12/2018

Ends: 1/12/2022

Duration: 4 years

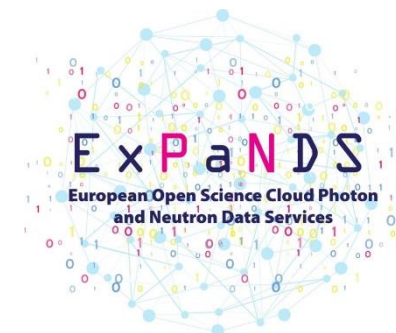
ESFRI



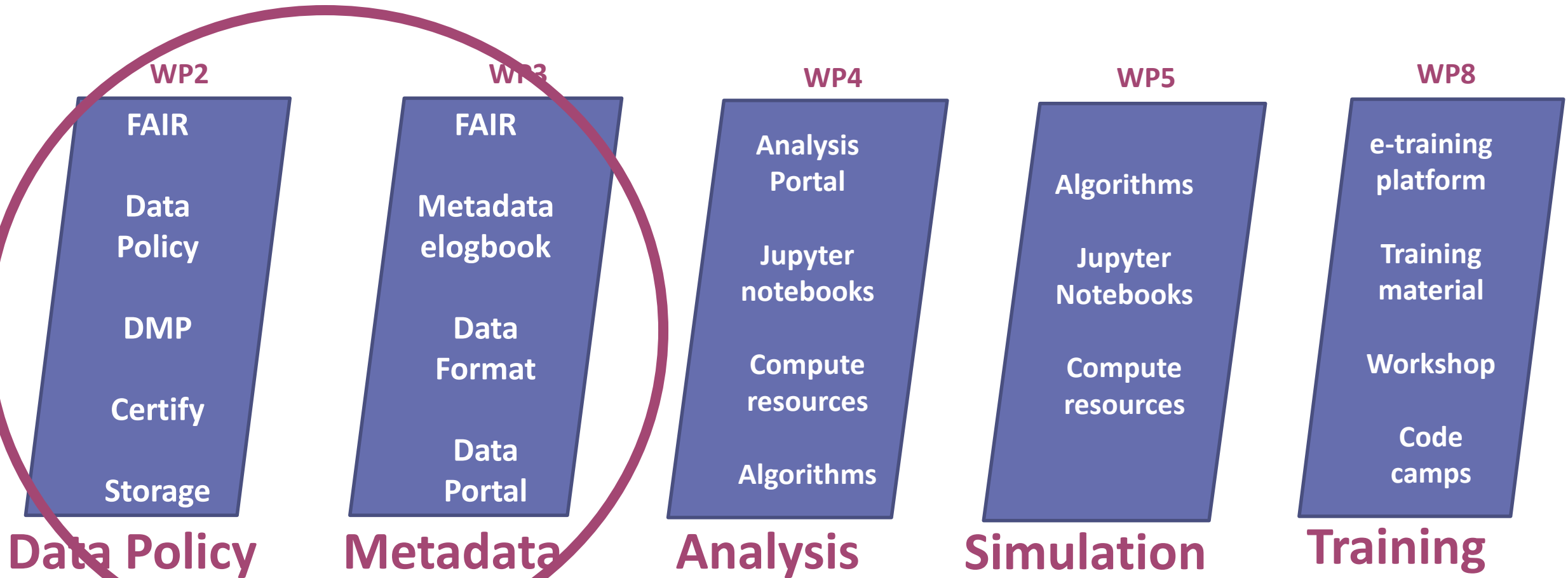
PaNOSC goals



PaNOSC works closely with the PaN sources in Europe to develop common policies, strategies and solutions in the area of FAIR data policy, data management and data services, integrating them into the EOSC.



PaNOSC = data policy + management + analysis + simulation + training



All the above services depend on RDM ...

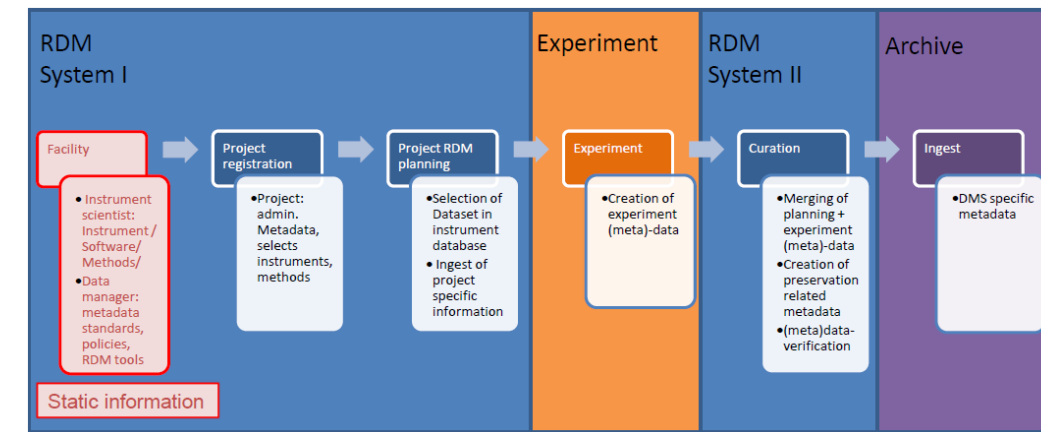
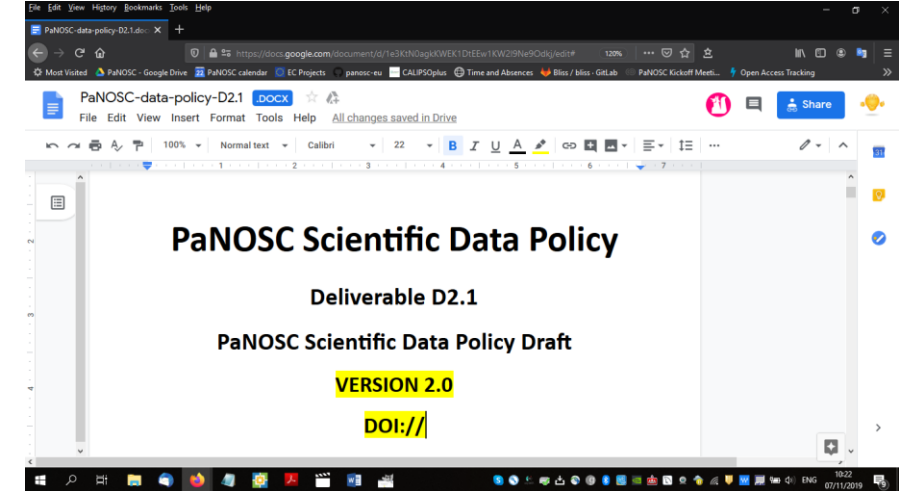
Data policy WP2

Updated the PaNdata data policy

Included the FAIR concepts to make it FAIR compliant (52 criteria)

Updating existing Data Policies to be compatible with the FAIR Data Policy framework

Current activity - adopting a tool for managing Data Management Plans together with



PaNOSC FAIR data policy framework

- First step in RDM is to define and adopt a data policy
- PaNOSC and ExPaNDS have published a FAIR data policy framework for research infrastructures

DOI 10.5281/zenodo.3862701

May 20, 2020

Project deliverable Open Access

PaNOSC FAIR Research Data Policy framework

 Gotz, Andy;  Perrin, Jean-Francois;  Fangohr, Hans;  Salvat, Daniel;  Gliksohn, Florian;  Markvardsen, Anders;  McBirnie, Abigail;  Gonzalez-Beltran, Alejandra;  Taylor, Jonathan;  Matthews, Brian

This paper presents the new photon and neutron research data policy framework based on the previous PaNData policy (<https://doi.org/10.5281/zenodo.3738497>) applicable to all photon and neutron facilities and scientific research data in general. The data policy framework is strongly aligned with the FAIR principles. The aim of the policy is to ensure that the FAIR principles are applied in research data policies. This deliverable has been prepared by the EOSC projects PaNOSC (<https://panosc.eu>) and ExPaNDS (<https://expands.eu>) together to ensure harmonisation of the updated data policies for the photon and neutron communities.

Preview

Page: 3 of 17 Automatic Zoom



APPENDIX 2 - FAIR Data Maturity Model analysis

The PaNOSC Data Policy Framework have been analysed according to the FAIR indicators in the RDA FAIR Data Maturity Model¹⁵. The results are presented in the table and in the plots below.

FAIR	INDICATOR_ID	INDICATORS	PRIORITY	DP METRIC	DP ITEM
F1	RDA-F1-01M	Metadata is identified by a persistent identifier	Essential	4	3.3.1
F1	RDA-F1-01D	Data is identified by a persistent identifier	Essential	4	3.3.1
F1	RDA-F1-02M	Metadata is identified by a universally unique identifier	Essential	4	3.3.1
F1	RDA-F1-02D	Data is identified by a universally unique identifier	Essential	4	3.3.1
F2	RDA-F2-01M	Sufficient metadata is provided to allow discovery, following domain/discipline-specific metadata standard	Important	4	3.2.5
F3	RDA-F3-01M	Metadata includes the identifier for the data	Important	3	IN-9

Main points of PaNOSC FAIR data policy framework

1. Explicitly mentions FAIR principles as guiding principles
2. Research Institute has the role to curate data
3. Defines different data types:
 - Raw, Reduced, Processed, Auxiliary, Results, Derived
4. Data Policy does not apply to derived data
5. During **Embargo Period of 3 years** (extendible on request) data access is reserved to the experimental team
6. Data become **Open Data** after Embargo Period
7. Strive to store Open Data for at least 10 years
8. Open Data can be accessed using a well defined protocol
9. Defines **DOIs** as preferred PID solution and must be cited

Citing Data DOIs in publications

PaNOSC, ExPaNDS, LEAPS and LENS contacted 9 major journals to incite them to request data DOIs be cited in publications – impact is not visible (yet)



Dear Editors,

Good data management and stewardship is an essential part of science. Making scientific data findable and accessible is one of the main objectives of the European Commission, as underlined by the [FAIR principles](#) and the [European Open Science Cloud \(EOSC\)](#).

Key representatives of the community of Photon and Neutron (PaN) sources in Europe are engaged in promoting the use of DOI for data, by stimulating their scientists and users to cite the DOIs of their experiments in their publications for the benefit of the whole scientific community. This short video-clip available [here](#) is an illustration of this effort.

The H2020 [FILL2030](#), [PaNOSC](#) and [ExPaNDS](#) projects, and [LEAPS](#) and [LENS](#), the largest initiatives federating European PaN sources, are calling you to support their commitment to FAIR data and open science by actively supporting the citation of data DOIs in the articles published in your journal.

The use of publication DOIs represented a major step forward to promote open science, networking and transparency. The use of DOIs for data presents the same advantages. They are persistent and stable references for scientific data, and enable data to be easily and reliably found, used and cited. The deposition of data in appropriate data repositories assures the quality of both data and metadata, which impacts the quality of science in general and of scientific articles in particular.

Citation of data DOIs in articles published in your journals would allow the access to data sets linked to the article, improve the readability of the data presented and make it easier for scientists to work on various data sets of the same subject. We believe that it will strengthen the trustworthiness of your articles and your journal as a whole.

The use of data DOIs would enable your reviewers to get background information and check open questions on the data used for the article easily. This may accelerate processes during your review phase and ensure data used for the publication are trustworthy.

We think that the use and citation of data DOIs would increase the potential of science to respond to new and old societal challenges.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852

DOI video explaining why use DOIs



<https://youtu.be/ekn0qicVFJM>

Data catalogue WP3

Develop an Application Programmers Interface (API) for searching for FAIR data

Integrate search API into EOSC portal

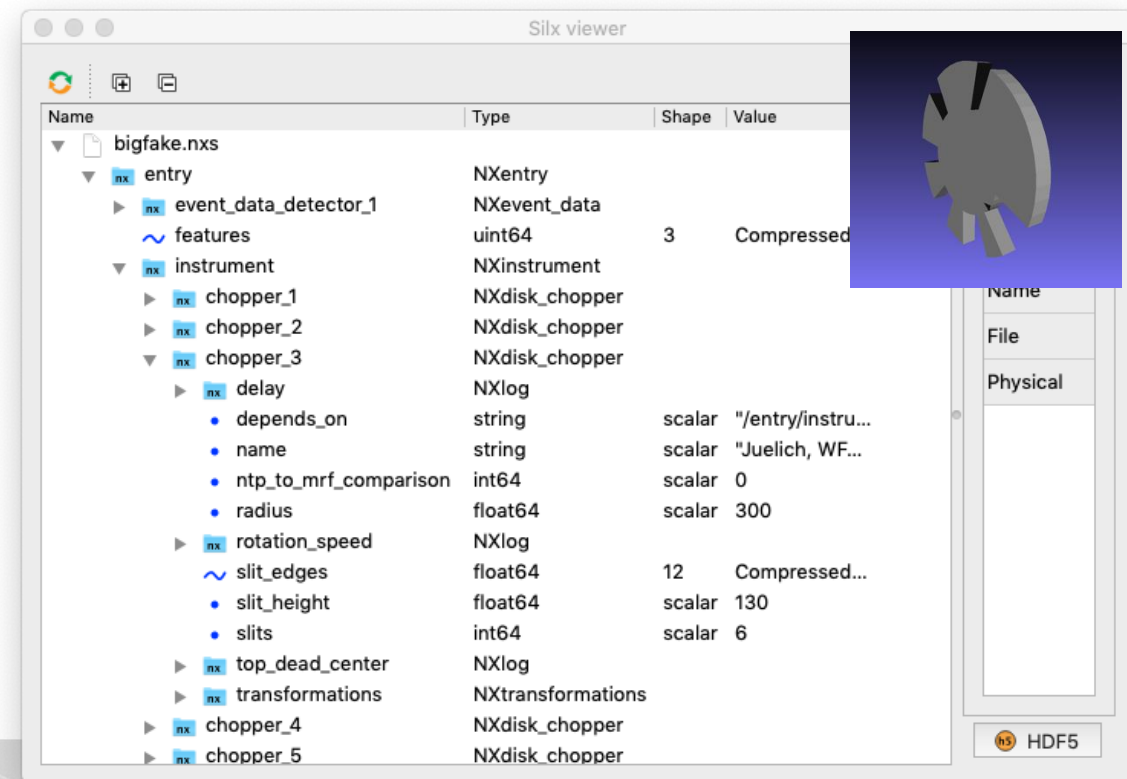
Using Nexus/HDF5 for metadata

Automating metadata collection on beamlines

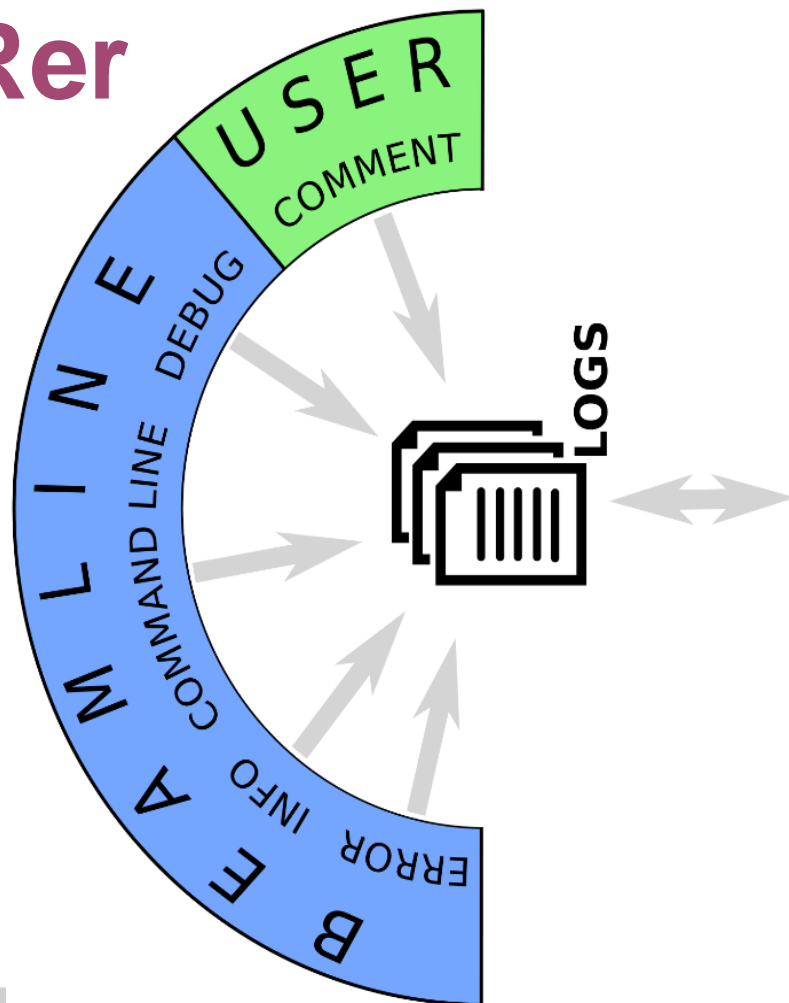
Using e-logbook to make data FAIRer

Long term storage (100s of Petabytes)

Dataset			Show/Hide	List Operations	Expand Operations
PATCH	/Datasets	Patch an existing model instance or insert a new one into the data source.			
GET	/Datasets	Find all instances of the model matched by filter from the data source.			
PUT	/Datasets	Replace an existing model instance or insert a new one into the data source.			
POST	/Datasets	Create a new instance of the model and persist it into the data source.			
PATCH	/Datasets/{id}	Patch attributes for a model instance and persist it into the data source.			
GET	/Datasets/{id}	Find a model instance by {{id}} from the data source.			
HEAD	/Datasets/{id}	Check whether a model instance exists in the data source.			
PUT	/Datasets/{id}	Replace attributes for a model instance and persist it into the data source.			
DELETE	/Datasets/{id}	Delete a model instance by {{id}} from the data source.			
GET	/Datasets/{id}/exists	Check whether a model instance exists in the data source.			



E-logbook makes experiments FAIRer

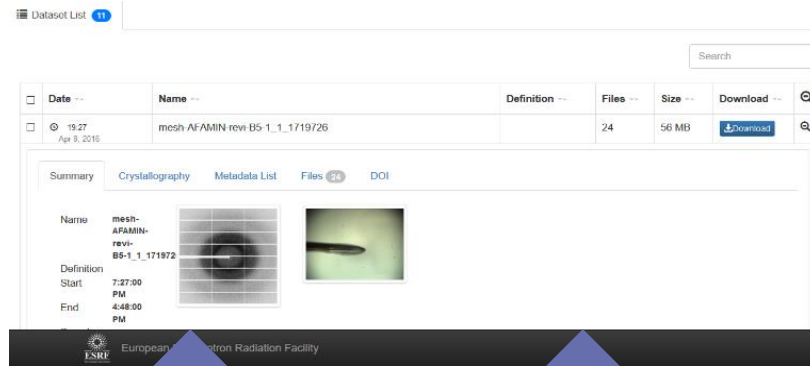


Web interface

Navigation bar: + New, Take a photo, View, PDF, Everywhere, Everywhere, Search

Time	Command Line	Info	Error
08:03:00	OPTICS> # io/optics/figa def measure1 '_ccd_set_concat(1)'		
08:01:40	OPTICS> snap flux		COMMENT
08:01:25	OPTICS> dt		COMMAND LINE
03:46:39	OPTICS> New dataset: cchof_root	INFO	
01:20:21	OPTICS> zapxiainage thg 6.65 10.69 10 vbg 54.8 27.6 27 0 10 (zapug: #6, spec: #3)		
01:20:20	OPTICS> New dataset: cFeo42-_root2		
01:19:51	OPTICS> prdef Maps_554		
00:54:47	no new data collected		ERROR
00:51:57	OPTICS> dt		

PaNOSC has 6 data catalogues with different APIs + UIs



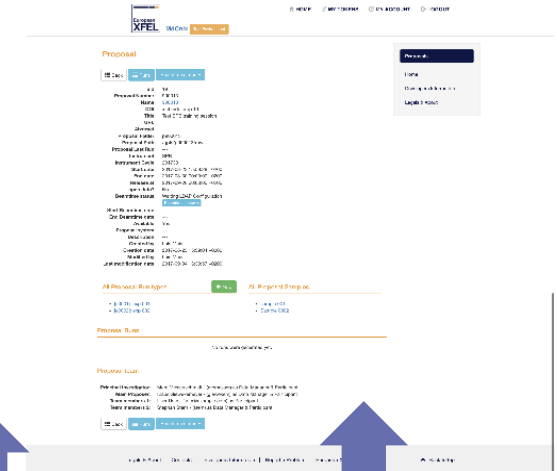
50 PB/yr

ESRF
(icat)
<https://data.esrf.fr>


<1 PB/yr

ESS
(SciCat)

.6 PB/yr

ILL
(local)
<https://data.ill.fr>


10 PB/yr

ELI
(tbd)

100 PB/yr

XFEL
(MyMdc)

PaNOSC common API across all sites



Search for Datasets



Common API to search across all PaNOSC catalogues

50 PB/yr

ESRF
(icat)

15 PB/yr

CERIC
(icat)

<1 PB/yr

ESS
(SciCat)

.6 PB/yr

ILL
(local)

10 PB/yr

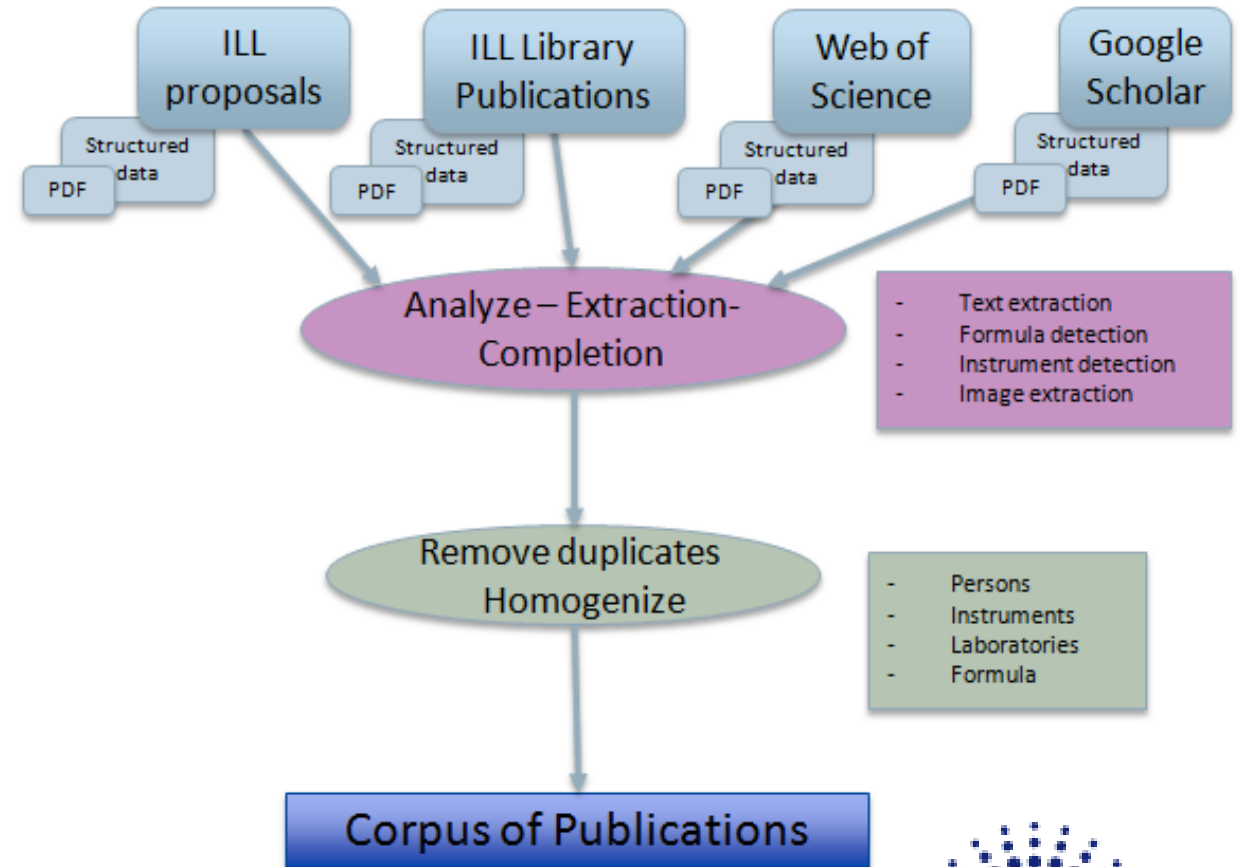
ELI
(tbd)

100 PB/yr

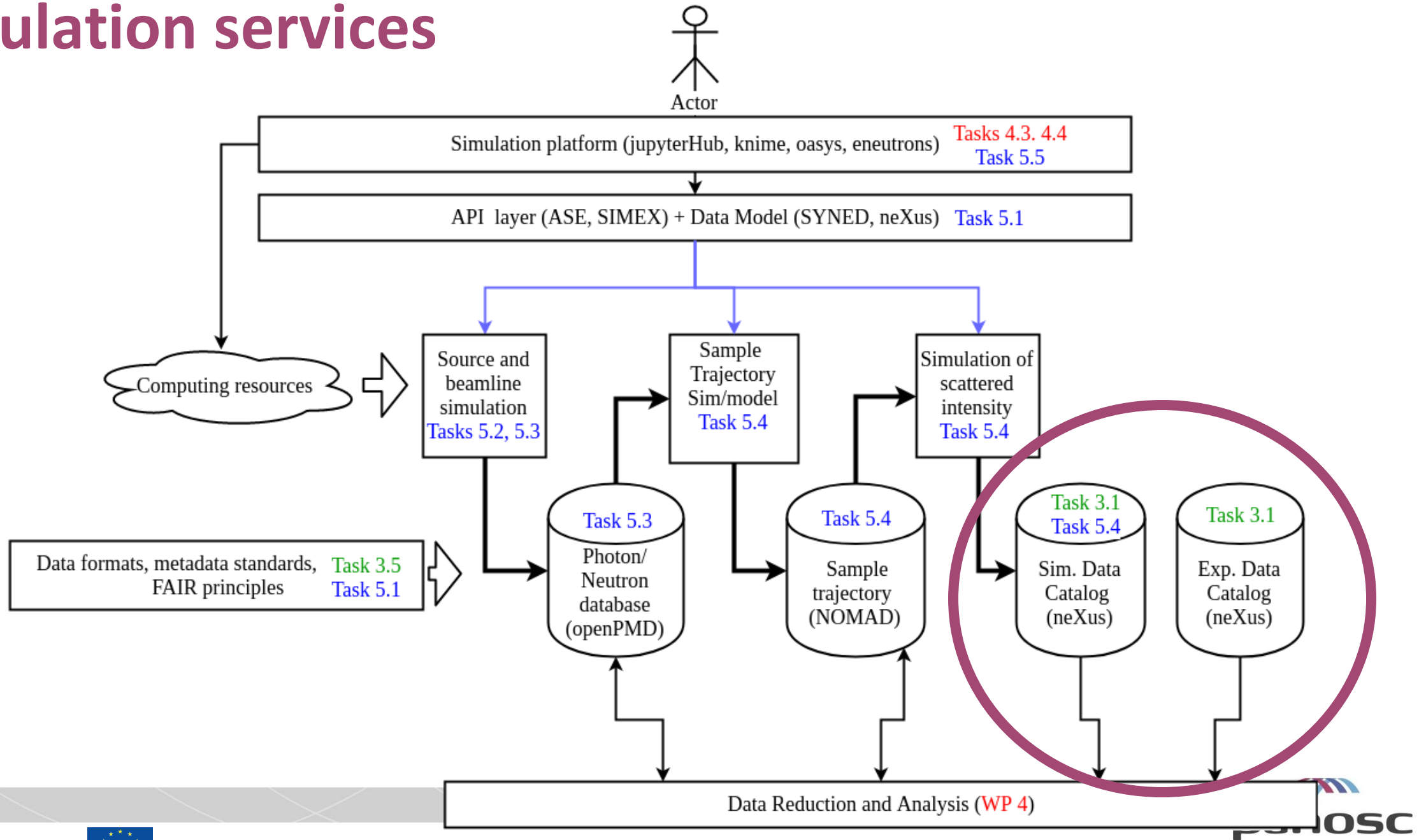
XFEL
(MyMdc)

PUMA publication search

- Publications and citations per instrument
- Delay between a proposal and a publication
- Countries and laboratories which use ILL (ESRF, ...) data
 - Look for potential new users
- Scientific trends
- Impact factor per science facility
- **F(AIR)inding data used in publications → DEMO**



Simulation services



PaNOSC + ExPaNDS contribution to EOSC



1. **Petabytes** of raw and processed data in a wide variety of scientific domains
2. **Meta-data** that will create **FAIR** raw and processed scientific data
3. **Software** for generic and specific data simulation and data analysis
4. **Workflows and expertise** for reducing and analysing data
5. **Reference training material** and **training platform** for understanding photon and neutron science and associated handling of data
6. Interface to large **user communities** of photon and neutron sources and their expectations for services



1. **Identify which data you need to curate.** Are they simulated, raw, reduced, processed, published or all of them?
2. **Define and adopt a FAIR Data Policy.** Consult the many resources out there
3. **Hire data managers (at least 2), identify experts for metadata.** Failure to do so means there is a strong risk that the data policy is not implemented.
4. **Choose and provide a single solution for software and infrastructure.** Needs to be sustainable in the long term.
5. **Implement above solutions to provide access to FAIR data.** Use an existing solution.
6. **Provide services on top of RDM.** Examples are search, metrics, download, analysis tools

PaNOSC lessons learned about FAIR

1. **It is a long but useful process.** Not to be underestimated but worth it – needs support
2. **Quality of data+metadata improves.** Metadata and curation improved a lot.
3. **Strengthens collaboration with community standards bodies.** Standards adopted for data formats + ontologies.
4. **Data services can be provided.** Remote access, search, analysis e.g. jupyter notebooks
5. **Data can be cited and accessible.** Still needs stronger support from journals.
6. **More and more Open Data are now available.** ESRF and ILL data embargo ending
7. **(FA)Interoperability and Reusability are still to be achieved.** The Holy Grail !



What is the situation for data in battery research?

Q: Is this example representative for battery research?

A story behind much needed data for battery research



Daniel Schröder

Junior Group Leader, Giessen University

Published Nov 13, 2020

Data Descriptor | [Open Access](#) | Published: 13 November 2020

Reproducible and stable cycling performance data on secondary zinc oxygen batteries

Saustin Dongmo, Julian Jakob Alexander Kreissl, Kohei Miyazaki, Takeshi Abe, Ting-Hsuan You, Chi-Chang Hu & Daniel Schröder

Scientific Data **7**, Article number: 395 (2020) | [Cite this article](#)

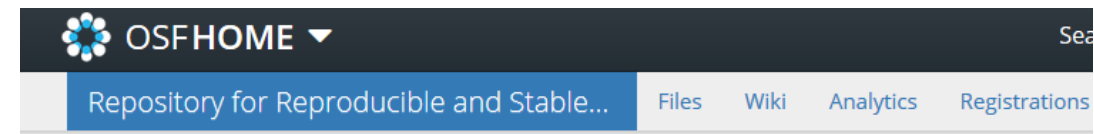
22

930 Accesses



2 Altmetric | [Metrics](#)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852



Repository for Reproducible and Stable Cycling Performance Data on Secondary Zinc-Oxygen Batteries

Contributors: [Daniel Schröder](#)

Date created: 2020-05-12 09:13 PM | Last Updated: 2020-09-08 07:35 PM

Identifier: DOI 10.17605/OSF.IO/7956W

Category: Project

Description: Data for use in model-based research and parameter fitting for simulations

License: CC-BY Attribution 4.0 International

e under grant agreement No. 823852



Conclusion

1. PaNOSC vision is to create a Scientific Data Commons for Photon and Neutron sources and make FAIR open data available as members of the EOSC
2. PaNOSC lessons learned in RDM have already been adopted by PaNOSC partners. Need to be sustainable.
3. PaNOSC+ExPaNDS partners can provide FAIR data for battery research → an example of data (FAI)Reuse

PaNOSC Resources

- <https://panosc.eu>
- <https://github.com/panosc-eu/panosc>
- “Enabling Open Science for Photon and Neutron sources”
presented at ICALEPCS 2019
<http://icalepcs2019.vrws.de/papers/tubpl02.pdf>
- PaNOSC Position Paper - https://github.com/panosc-eu/panosc/blob/master/Work%20Packages/WP6%20EOSC%20Integration/PaNOSC_EOSC_position_paper.pdf

