**PaNOSC experience in Research Data Management presented at the Battery2030+ Initiative on RDM.**

On March 12th, 2021, PaNOSC coordinator, **Andy Götz**, attended with an invited talk the 2nd online workshop of the [Battery2030+ Initiative](https://battery2030.eu/), focused on the benefits of research data management (RDM) and guidelines, through the showcase of best practice examples.

PaNOSC vision is to create a Scientific Data Commons for Photon and Neutron sources and make the data available as members of the EOSC. The experience acquired by PaNOSC in achieving this goal, could help Battery2030+ Initiative in Research Data Management, and the partners of both PaNOSC and its sister project, [ExPaNDS](https://expands.eu/), can provide FAIR data for battery research, thus giving a concrete example of data (FAI)Reuse.

When highlighting the benefits of RDM for researchers, Andy pointed out that a state-of-the art RDM allows to make the most of data, and that the adoption of FAIR principles and practices increases the reproducibility of science, which is currently undergoing a crisis, as also shown in a 2016 paper published in *Nature* [1].

In this respect, PaNOSC, which works closely with the PaN sources in Europe to develop common policies, strategies and solutions in the area of FAIR data policy, aims at making photon and neutron (PaN) facilities’ scientific data fully compliant with the FAIR principles, at sharing best practices for open data policies, and at increasing the impact of research infrastructures by encouraging data reuse.

In the frame of PaNOSC’s [work package 2 – Data Policy and Stewardship](https://www.panosc.eu/work-packages/work-package-2-data-policy-and-stewardship/), project partners have been working towards these goals by:

* Updating and publishing the [PaN research data policy framework](https://doi.org/10.5281/zenodo.3862701) [2], which provides a common framework for management of scientific data at photon and neutron facilities, taking into account the FAIR principles;
* Adopting or aligning Data Policies at all PanOSC sites
* Adopting a managing data management together with ExPaNDS.

As highlighted by Andy in his presentation, these and a wider set of services (data storage, certified data catalogues, standard data formats, common data search and data analysis portal, compute resources, new algorithms, technologies such as Jupyter Notebook, e-learning platform and material, simulation services) developed and deployed in the project, strictly depend on RDM.

The use and citation of data DOIs in publications are also key to make data re-usable and science reproducible. To incite publishers to request data DOIs be cited in publications, a joint action was carried, out with the H2020 [FILL2030](http://www.fill2030.eu/) project, PaNOSC, ExPaNDS, and the [LEAPS](https://leaps-initiative.eu/) and [LENS](https://www.lens-initiative.org/) initiatives, through which 40+ journals have been contacted, and a video showcasing the benefits of the use of DOIs was also produced.

**WATCH THE VIDEO “The DOI for data” here**: https://youtu.be/ekn0qicVFJM

The main outcomes of other PaNOSC work packages were also presented. In particular, in [WP3 – Data Catalogue Services](https://www.panosc.eu/work-packages/work-package-3-data-catalog-services/), the activities carried out since the project’s start include:

* Development of an Application Programmers Interface (API) for searching for FAIR data;
* Integration of a search API into the EOSC portal;
* Use of Nexus/HDF5 standard format for metadata;
* Automation of metadata collection on beamlines;
* Use of e-logbook to make data FAIRer;
* Long-term storage (hundreds of Petabytes).

Andy concluded his talk by mentioning the lessons learnt in PaNOSC on RDM:

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

**Download Andy Götz’ presentation here**

During the event, editors of three scientific journals present at the event discussed questions on open data, and participants took part in the discussion about the next steps taken within the Battery2030+ Initiative towards concrete actions concerning RDM & guidelines.

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

[2] Götz, A., Perrin, J.-F., Fangohr, H., Salvat, D., Gliksohn, F., Markvardsen, A., McBirnie, A., Gonzalez-Beltran, A., Taylor, J., Matthews, B., PaNOSC FAIR Research Data Policy Framework, Zenodo, DOI: https://doi.org/10.5281/zenodo.3862701