Photon and Neutron Open Science Cloud - As part of the European Open Science Cloud

This paper presents the position of the Photon and Neutron Open Science Cloud (PaNOSC) H2020 cluster project of 6 ESFRI and ERIC organisations with respect to the European Open Science Cloud (EOSC). The 6 organisations are ESRF, XFEL.EU, ELI, CERIC-ERIC, ILL and ESS. These organisations are state of the art large scale instruments for photon, neutron and laser science. They are representative of the larger photon, neutron and laser community in Europe which serves more than 50,000 scientists and industrial users doing cutting edge experiments in many fields - spanning from life sciences and material sciences to cultural heritage [1]. The PaNOSC organisations generate (ESRF, ILL, CERIC-ERIC) or will generate (XFEL.EU, ELI, ESS) Petabytes of data every year. Access to the data is managed by the data policy of each organisation. The data policies are largely inspired by the PaNData data policy which respects the FAIR data principles [2]. The EOSC offers a unique opportunity to strengthen the best practices of data management and open access to simulated and acquired scientific data. It will show the way for the general adoption of these practices in the photon, neutron and laser communities and enable all social, legal and technical issues related to open data access to be addressed. This paper summarises the role of the EOSC and the PaNOSC consortium in realising an Open Scientific Data Cloud for finding and analysing data by answering the questions posed by the EOSC Stakeholders Meeting organisers below.

What are the services that your research collaboration would benefit from as consumer of EOSC?

The first essential task is the harmonization and interoperability of services required to provide secure access to data e.g. Authentication and Authorization Infrastructure (AAI), searching of meta-catalogs, data transfer, cloud computing resources. The EOSC can help evolve the best practices around data by defining standards for data repositories, providing training for scientists on publishing and using Open Data and by training data scientists on data management.

Do you envisage the need of shared services that EOSC could provide with central coordination, to take advantage of economies of scale and ensure harmonization across different providers?

Access to cloud computing resources, data storage, catalogs of software, AAI and long term archiving services will profit from being shared. Definition of standards for cloud services. Certification of these services to ensure their compliance with the standards. Efficient data transport across Europe to data centers and users sites. General services like a helpdesk and user feedback service, document sharing, editing etc.

3. What criteria should the EOSC services meet in order to be appealing to your research collaboration?

Most importantly the EOSC must consist of clearly defined sustainable services which are well documented and easy to use. The Quality of service must be of a high standard and reliable. It must be easy to link up to the EOSC via well defined APIs and services. The EOSC must be governed in an open and transparent manner. The role of the EOSC providers and partners must be clearly defined. Users must be consulted to determine which services are required.

4. What services and resources (e.g. data, publications, software and other digital artefacts) could you provide for sharing, third party access and use through the EOSC?

The Photon and Neutron community will provide access to Open Data from a large variety of experiments via searchable metadata catalogues. The PaN community provides scientific software for data analysis and scientific expertise to help users understand and analyse data. We are also moving towards the use of experiment simulation both in preparation and analysis of experiments, and anticipate to start providing such cloud hosted services.

5. How do you see your role in fostering the uptake of the FAIR principles in your environment?

The PaNOSC cluster will help introduce a new data culture to the user community – via training at each site and holding workshops on scientific data management and publishing practices. Best practices in data stewardship will be shared with other labs at least in the PaN community.

Experience and results will be shared openly via publications and meetings. The positive experience of implementing an Open Data policy will help convince other analytical facilities still struggling with adopting the FAIR principles.

Conclusion

The PaNOSC cluster and the Photon and Neutron community in general are a good example of communities where the EOSC can play a big role. They are aware of data stewardship and a number of them have started implementing data policies based on the FAIR principles. The EOSC can provide these sites with guidance to address the full range of scientific data stewardship and to integrate in the Open Science and Open Data landscape and play their role helping users adopt and benefit from Open Data policies. The PaNOSC consortium will provide data analysis services and scientific expertise close to the data sources for their users during or after their experiments in a wide range of domains. The EOSC will provide complementary resources for cases where the PaNOSC sites are

under resourced and for the wider community of Open Data users. The solutions, experiences and practices put in place to link PaNOSC to EOSC will be disseminated to the whole PaN community and beyond so that they can follow, thereby connecting 50000+ scientists in a wide variety of fields to the EOSC.

References

[1] http://pan-data.eu/node/105

[2] http://pan-data.eu/

Author(s)

Sune Rastad Bahn, Group Leader Data Systems & Technologies, European Spallation Source, ERIC