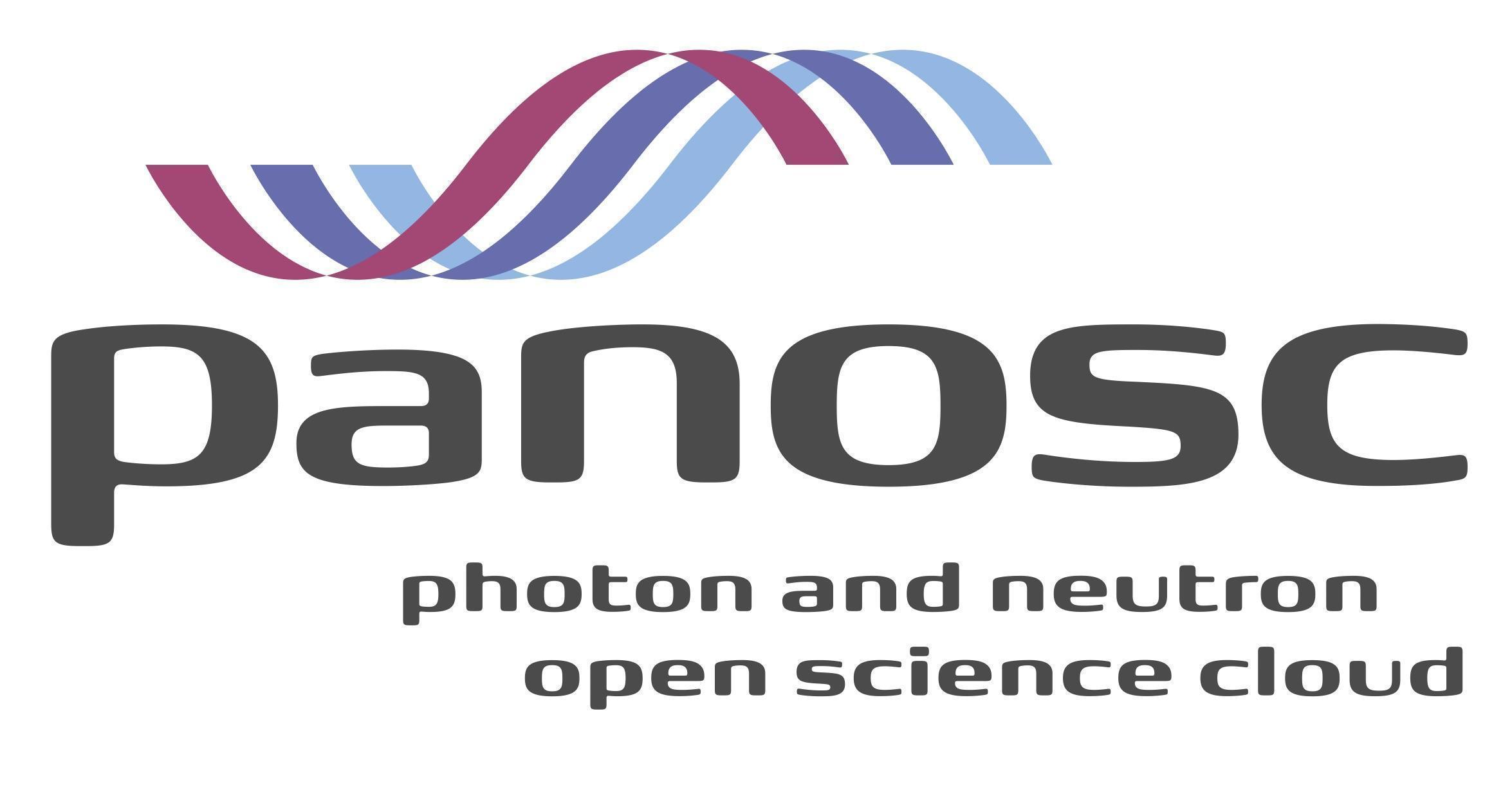
PaNOSC

Photon and Neutron Open Science Cloud

H2020-INFRAEOSC-04-2018

Grant Agreement Number: 823852



Work Package 3

MS – 3.3

Project Deliverable Information Sheet

|  |  |
| --- | --- |
| Project Reference No. | 823852 |
| Project acronym: | PaNOSC |
| Project full name: | Photon and Neutron Open Science Cloud |
| H2020 Call: | INFRAEOSC-04-2018 |
| Project Coordinator | Andy Götz (andy.gotz@esrf.fr) |
| Coordinating Organization: | ESRF |
| Project Website: | www.panosc.eu |
| Dissemination Level | Public |
| Contractual Delivery Date: | 30/05/2021 |
| Actual Delivery Date: | 01/02/2021 |
| EC Project Officer: | Simona Misiti |

Document Control Sheet

|  |  |
| --- | --- |
| Document | Title: MS3.3 Catalogue Integration Best Practices Meeting |
| Version: 1.0 |
| Available at: https://github.com/panosc-eu/panosc |
| Files: 1 |
| Date | 23/06/2021 |
| Authorship | Written by: L. Schrettner (ELI ALPS), T. Richter (ESS), T. Ivănoaica (ELI-DC) |
| Contributors: …… |
| Reviewed by: A. Götz(ESRF) |
| Approved: J. Bodera Sempere (ESRF) |

List of participants

|  |  |  |
| --- | --- | --- |
| Participant No. | Participant organisation name | Country |
| 1 | European Synchrotron Radiation Facility (ESRF) | France |
| 2 | Institut Laue-Langevin (ILL) | France |
| 3 | European XFEL (EuXFEL) | Germany |
| 4 | The European Spallation Source (ESS) | Sweden |
| 5 | Extreme Light Infrastructure Delivery Consortium (ELI-DC) | Belgium |
| 6 | Central European Research Infrastructure Consortium (CERIC-ERIC) | Italy |
| 7 | EGI Foundation (EGI.eu) | The Netherlands |

Catalogue Integration Best Practices Meeting

# Summary

The Catalogue Integration Best Practices meeting took place on three consecutive afternoons between 18 May 2021 and 20 May 2021. Fourteen presentations were scheduled to cover best practices mainly from PaNOSC partner institutions with three main focus points: user portal and proposal management, data catalogue systems and their integration, and electronic logbook solutions. An indico site had been set up (https://indico.eli-laser.eu/event/3/) for the event, there were 65 registered participants and all presentations have been uploaded and are publicly available. Several questions and issues have come up during the workshop which will be further pursued by the interested parties in separate follow-up meetings and in PaNOSC WP3 as well.

# Main discussion points

The agenda of the meeting was as follows:

18 May 2021

|  |  |  |
| --- | --- | --- |
| **Start** | **Title** | **Speaker** |
| 13:00 | Welcome message from WP3 Leader | Tobias Richter |
| 13:15 | User Portal and Proposal System @ ESS | Fredrik Bolmsten |
| 13:55 | CERIC User Office - VUO Current Status | Emiliano Coghetto |
| 14:25 | User Portal and Proposal System @ ALBA | Daniel Sánchez |
| 14:55 | SciCat and data ingestion @ PSI | Carlo Minotti |
| 15:20 | Round Table Discussions Closing Day1 |  |

19 May 2021

|  |  |  |
| --- | --- | --- |
| **Start** | **Title** | **Speaker** |
| 13:00 | Survey of cataloguing systems for ELI | Balázs Bagó |
| 13:40 | ICAT recent developments, mapping, providers | Stuart Pullinger |
| 14:10 | Cataloguing + logbook solution @ XFEL | Luis Maia |
| 14:50 | ICAT catalogue integration @ CERIC | Emiliano Coghetto |
| 15:05 | eLogBook @ ELI-ALPS | Lajos Schrettner |
| 15:25 | Logbook adapted from ELETTRA @ CERIC | Emiliano Coghetto |
| 15:40 | Logbook @ ESRF | Alex de Maria |
| 16:10 | Round Table Discussions Closing Day2 |  |

20 May 2021

|  |  |  |
| --- | --- | --- |
| **Start** | **Title** | **Speaker** |
| 13:00 | Data Ingestion @ ILL | Stuart Caunt |
| 13:20 | Photon and Neutron Ontologies from design to implementation | Heike Görzig et al. |
| 14:10 | Nexus data formats @ XFEL | Yury Kirienko |
| 14:30 | Round Table Discussions Closing Day3 |  |

As reflected in the agenda, the meeting approached catalogue integration issues from three main directions, based on preliminary discussions with interested parties, especially CERIC-ERCI and ELI, the two distributed partner facilities of PaNOSC. Regarding the use of cataloguing systems, partners started from different conditions, developed or applied their own solutions over the years and need to carefully plan any modifications and updates to their systems. In this respect a best practice not unknown in software engineering is that large monolithic systems should be avoided as much as possible. Building from smaller, functionally self-contained components leads to clearer designs and offers greater flexibility in deployment and upgrades. This approach is also used in defining the data portal architecture in WP4.

Proposal management was selected as the first focus point. It is the first phase of the experiment life cycle where base metadata are collected about a prospective experiment and the users associated with it. There were three presentations, covering user portal development activities at ESS, the CERIC User Office portal VUO (Virtual User Office), and the User Portal and Proposal system at ALBA (an ExPaNDS member). All speakers emphasized the difficulties of implementing the management process which is often complex and has to provide many options. One way of coping with this is to provide extensive configuration capabilities so that users or administrators can tailor the process without the need to develop extra functionality into the system. Other important factors are the choice of technologies that are used to implement and operate the system, and the related issues of documentation and long term sustainability of a deployed system.

The second focus point was a survey and current status of available cataloguing systems, their internal organisation and components, with a special interest in data ingestion techniques and capabilities. Seven presentations contributed to this, including a status report from both ICAT and SciCAT developers, the two main catalogue systems in use by PaNOSC and ExPaNDS partner facilities. There was a survey report enumerating the most important aspects of catalogue systems. The survey concentrated on ICAT, SciCAT and Invenio for evaluation against the set of predefined aspects that covered seven categories: data model, authentication and authorization, ingestion, search, GUI, deployment/operation, and documentation. The survey results show that besides ICAT and SciCAT, Invenio may also worth to be considered as data catalogue solution. Data ingestion techniques were discussed in the case of ICAT and SciCAT, as this is one of the most important technical challenges faced by research infrastructures who would like to introduce or develop their data management systems towards better compliance with FAIR principles. ELI is going to set up test environments where it can prototype data ingestion techniques of existing data sources.

A third series of talks concentrated on electronic logbooks as they serve as another important source of data and metadata for cataloguing systems. Several institutions have developed or adopted their own implementation of the electronic logbook despite the fact that there is a substantial overlap in the functionality that users expect from such a system. Standardisation of the electronic logbook application programming interface (API) would allow catalogue systems to interoperate with any of conforming logbooks and thus would make it easier for institutes to adopt existing solutions instead of developing their own.