

PaNOSC-OSCARS projects mentorship meeting (23/05/2025) report

The PaNOSC-OSCARS Projects Mentorship Meeting was held remotely on May 23rd, 2025, via Zoom. The meeting brought together 14 successful projects funded under the OSCARS Open Call 2024 to **discuss specific mentorship needs** for adopting PaNOSC outcomes and best practices. The slides are available on indico: <https://indico.esrf.fr/event/203/timetable/#20250523>

List of participants

Jean-Francois Perrin (ESRF (Chair), Frieda Sorgenfrei (AI-SCOPE), Fabio Dall'Antonia (MC-ReDD), Leandro Liborio (CDIF-4-XAS), Bernd Hinrichsen (MatScatNet), Yohanna Yehudi (OSPARK Bootcamp), Oliver Knodel (mTeSS-X), Massimiliano Novelli (PaN-Finder), Ludovic Leroux (VISA), Marine Cotte (SHARE), Daniel Garijo (CODEMETASOFT) David Stansby (HEFTIE), Joerg Hammel (AMBCAT), Nicolas Bruyant (HiMAGNETOS), Sandor Brockhauser (Findable Big Data), Andy Gotz (ESRF), Jayesh Wagh (ESRF), Jordi Boderá Sempere (ESRF), Nicoletta Carboni (CERIC-ERIC), Melanie Nentwich (DESY), Simon Delcamp (ESRF), Nicolas Soler (ALBA), Heike Gorzig (HZB), Fredrik Bolmsten (ESS), Rolf Krahel (HZB), Loes M.J. Kroon-Batenburg (EuXFEL), Frank Schluenzen (DESY), Martin Voigt (HZDR)

AI-Scope

AI-Scope is dedicated to developing and integrating AI tools for scientific data analysis, with a strong focus on improving access to test data, training resources, and seamless integration with research infrastructures such as beamlines and data portals. **The project team is currently facing challenges in finding adequate test and training datasets for the AI tool, as well as in identifying the appropriate infrastructure for integration.** There is a clear need for mentorship in several areas: **guidance on who to contact and how to interface with data portals for depositing and accessing lab data (beyond beamline data), API development to enable the AI tool to interact with data portals and beamlines—including considerations for packaging and deployment and strategies for saving metadata alongside raw data.** The project is also interested in exploring collaboration opportunities with VISA for further integration.

Mc-ReDD

Mc-ReDD's core focus is the development of the imgCIF_Creator service, which supports advanced data management and interoperability for scientific research. **The project is currently seeking mentorship to facilitate the onboarding of this service to the PaNOSC Node (or EOSC) through the Competence Center, ensuring that it is accessible and beneficial to the broader research community. Sustainability is a key priority, with plans for the PaNOSC node to maintain and sustain the service over the long term.**

CDIF-4-XAS

CDIF-4-XAS is working to improve the transfer, interoperability, and reusability of X-ray spectroscopy (XAS) data and metadata. **The project is aligning its XAS metadata framework with the broader Cross-Domain Interoperability Framework (CDIF) and is exploring possible overlaps with data catalogues and portals. A key objective is to connect the CDIF-4-XAS framework to the XAS Galaxy tool and to metadata catalogues and data portals from PaNOSC. The team also intends to leverage the PaNOSC Common Search API for cross-facility data and metadata discovery. To further their goals, the project requires access to more experimental XAS data and metadata, highlighting the importance of collaboration and resource sharing.**

OSPARK

OSPARK is committed to delivering Open Science training through initiatives such as the OLS (Open Life Science) and the Digital Training Academy. These programs are designed to address the specific needs of participating organizations and are **open to collaboration for the training work package**. By fostering partnerships and sharing expertise, OSPARK aims to build capacity and promote best practices in Open Science across the research community.

m-Tess-X

m-Tess-X is developing a training e-support system (Tess) that merges the PaN training catalogue with the ELIXIR training catalogue, creating a unified resource for scientific training. **The project aims to streamline access to training materials and enhance the learning experience for researchers. However, there is a recognized risk of mild uptake for TessX, which the team is monitoring closely as they work to promote adoption and engagement.**

Findable Big Data from Various Material Characterisation Techniques

Findable Big Data is exploring the creation of an ontology for ESRF experimental techniques to improve the findability and interoperability of scientific data. **The project team is focused on enhancing data discovery and integration, with no immediate mentorship needs identified at this stage.**

PaN Finder

PaN Finder is focused on **expanding the information retrievable through the PaNOSC API or OAI-PMH protocol, using AI-based approach to enable the tool to operate on a broader range of data. The project is also working to enrich PaNET with additional definitions, synonyms, and keywords to improve search functionalities within PaN Finder. A key goal is to design a user interface that enhances the user experience and collects valuable feedback. Sustainability of these enhancements will be ensured through the PaNOSC data portal and node.**

VISA

VISA supports virtual research environments and software distribution, providing researchers with flexible and accessible tools for data processing and analysis. **The project currently reports no particular difficulties or mentorship needs. However, there is ongoing interest in discussions related to authentication and authorization infrastructure (AAI) within the PaNOSC node, as well as the potential for public CVMFS distribution of PaN software.**

SHARE

SHARE manages a large network focused on historical materials, with over 130 partners across the EU and beyond. **The project is building a comprehensive database of data and metadata collected through this network. SHARE has identified several mentorship needs: ensuring that cultural heritage data are discoverable through typical user searches, utilizing packaging for EWOKS and distributing software via Cern VM-File System (CVMFS), integrating training resources on the pan-training.eu platform, and creating customized VISA-VMs for data processing. Sustainability of the data and software will be managed by ESRF.**

HEFTIE

HEFTIE is focused on reviewing software using the Research Software Toolkit and integrating new training resources into the PaN training catalogue. The project is seeking mentorship to support these activities and ensure that best practices are followed. Long-term sustainability will be ensured by the software co-developers from the community, fostering ongoing collaboration and development.

CodeMetaSoft

CodeMetaSoft is dedicated to improving research software metadata practices across science clusters. **The project has utilized the PaN software catalogue as part of its landscape analysis and is actively promoting good practices in software metadata. No specific mentorship needs have been identified at this time.**

AMBCAT

AMBCAT is addressing the challenges of accessing amber fossils by developing a digital twin for these specimens. **The project will make data from 1,500 fossils openly available and FAIR, supported by visualization tools. Current user community involvement is serving as a proof of concept for the development of a general data catalogue, with SciCat being considered as a potential solution.**

HiMagnetOS

HiMagnetOS is interested in following the PaNOSC example to develop its own metadata catalogue and storage solutions. **The project has several mentorships needs: a deeper understanding and training on the NeXus metadata format, support for developing a software catalogue, assistance with authentication and authorization infrastructure (AAI) with particular interest in Umbrella ID, given the overlap with PaNOSC users and potential adoption of TESS for training and CodeMetaSoft for software quality. The team is also interested in collaborating on Horizon Europe EOSC calls, highlighting their commitment to advancing research infrastructure and interoperability.**