∞ eOSC | Federation

Scientific use case:
PaN-Finder

∞ eOSC Node | PaNOSC
Physical Sciences & Engineering

∞ eOSC Node | Germany

Massimiliano Novelli is a Senior Data Curation Scientist at the European Spallation Source (ESS), working within the Scientific Information and Management Systems (SIMS) group. His role focuses on ensuring FAIR data practices across research outputs from ESS's scientific instruments. Novelli collaborates closely with the PaNOSC and OSCARS projects to coordinate and integrate data catalogues across photon and neutron facilities.


The **PaN-Finder** science case presents an artificial intelligence−driven data discovery platform designed to enhance Open Science within Europe's research communities **by making data from photon and neutron facilities easier to find**. The initiative builds upon previous work undertaken through the PaNOSC (Photon and Neutron Open Science Cloud) project, which aimed to interconnect data catalogues from large-scale research facilities. While the initial federated portal provided a single access point to open data, it was limited by inconsistencies in metadata, domain-specific terminology, and the need for users to possess detailed technical knowledge to perform effective searches. PaN-Finder aims to **lower these barriers**, making the discovery of open research data more intuitive and inclusive.

Problem addressed

The core challenge is the **inconsistency and fragmentation of metadata** and terminology across facilities and communities. Similar data attributes—such as "sample," "sample name," or even abbreviated terms—are labelled differently, complicating federated searches. The initial PaNOSC federated search model also lacked semantic understanding and required prior user-awareness of this constraint. Consequently, even though valuable open data exists, it is often difficult to locate or interpret unless already cited in a publication.

Technical solution

PaN-Finder addresses these limitations by introducing a semantic, AI-enhanced search engine capable of interpreting natural language queries. Instead of requiring users to understand the

1
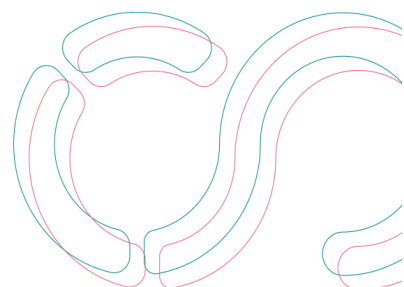
specific metadata structures of individual facilities, the system translates user intent—recognizing synonyms, context, and scientific terminology variations—to return relevant results across multiple repositories. The approach marks a shift from purely federated searching to a semi-centralized model in which metadata are harvested, harmonized, indexed and further curated before query processing. This allows for greater robustness, flexibility, and responsiveness. PaN-Finder leverages commercial AI tools with the long-term goal of migrating to open or even EOSC-provided AI services to ensure sustainability and European data sovereignty. The platform currently integrates open data from several major European facilities, including photon and neutron sources, via the PaNOSC and German national EOSC Nodes. PaN-Finder is designed to scale across the EOSC Federation and the broader EOSC ecosystem.

Scientific outcomes

By simplifying access to heterogeneous experimental datasets, PaN-Finder promotes data reuse and accelerates research cycles. PaN-Finder's accommodation of natural language prompts, such as, "Find datasets where lungs from a covid patient were scanned," enables scientists to formulate new hypotheses by locating existing datasets that cannot be found by existing search engines. This results in less duplication of experiments, lower costs and decreased time to publication. These benefits align with broader trends observed in astronomy and life sciences, where secondary analysis of open data already contributes substantially to scientific output. Moreover, the feedback loop between PaN-Finder's AI results and metadata quality drives continuous improvement of FAIR data curation standards across photon and neutron research infrastructures, fostering a cultural shift in the scientific community towards openness and collaboration.

Added value of EOSC

The integration of PaN-Finder with the EOSC Federation provides it with the governance, authentication framework, and infrastructural backbone that underpin PaN-Finder's long-term viability. It enhances long-term data stewardship, reduces dependence on commercial AI providers, and strengthens Europe's digital sovereignty. In the future, PaN-Finder and similar AI-powered discovery systems could form a cornerstone of EOSC's vision—making open research data not only available but also meaningfully usable by scientists across disciplines. By lowering barriers to discovery and fostering cross-domain collaboration, this work represents a crucial step toward a fully connected and intelligent European Web of FAIR Data.