

Towards an EOSC compliant research data repository in Hungary with ARP

Wednesday, 2 October 2024 15:50 (10 minutes)

The Hungarian Research Network's (HUN-REN) Data Repository Platform (ARP) is a national repository infrastructure that was opened to the public in March 2024. With ARP, we aim to create a federated research data repository system that supports the data management needs across its institutional network. Implementing ARP is our first step towards establishing an EOSC compliant research infrastructure.

Here we present the conceptualization, development, and deployment of this federated repository infrastructure, focusing on the ARP project's objectives, architecture, and functionalities.

The primary goal of the ARP project is to establish a FAIR focused, sustainable, continuously operational federated data repository infrastructure that not only supports the central storage and management of digital objects but also ensures the interoperability and accessibility of research data across various scientific domains.

The Hungarian Research Network (HUN-REN) currently comprises 11 research centers, 7 research institutes and 116 additional supported research groups, conducting research in the most varied disciplines of mathematics and natural sciences, life sciences, social sciences and the humanities.

ARP is built on a foundation of secure and scalable storage solutions, utilizing the HUN-REN's existing infrastructure to establish a resilient and redundant data storage environment. The system incorporates a hierarchical storage model with a capacity of 1.4 Petabytes, distributed across two sites for enhanced data security. This model supports triple replication of data, ensuring high availability and disaster recovery capabilities.

Central to the ARP's functionality is its suite of data management tools. The primary service of ARP is the data repository itself built on Harvard's Dataverse repository system. In ARP we addressed an important shortcoming of Dataverse, namely the difficulty to handle a diverse set of metadata schemas. As ARP's goal is to support the metadata annotation needs of researchers of various domains it was inevitable to provide a richer set of metadata schemas besides the ones built into Dataverse. To achieve this we added as a central component a Metadata Schema Registry, built on Stanford University's CEDAR framework and closely integrated with the ARP repository to manage diverse data types and standards, ensuring interoperability across different research disciplines.

Beside providing the possibility to author and use any domain specific metadata schema we also extended Dataverse with the import, export and authoring of datasets using RO-Crate via our custom AROMA tool. AROMA and RO-Crate facilitates the structured packaging and rich metadata annotation of research data, enhancing the granularity and usability of data curation. With RO-Crate it is possible to describe datasets or individual files in datasets in any detail that is not otherwise possible in Dataverse.

ARP as a federated service integrates disparate data management systems into a cohesive framework that supports a unified knowledge graph and query service for researchers nationwide. The ViVO based knowledge graph of ARP enables detailed, file-level search functionality and supports federated searches across a variety of national and international research databases, significantly improving data discoverability.

HUN-REN ARP project represents a significant advancement in the field of research data management for the Hungarian research community.

Topic

EOSC Developments and Open Science: Reproducible Open Science

Primary author: Mr PATAKI, Balázs (HUN-REN SZTAKI)

Co-author: Mr KOVÁCS, László (HUN-REN SZTAKI)

Presenter: Mr PATAKI, Balázs (HUN-REN SZTAKI)

Session Classification: National Perspectives: EGI Member Countries' Latest Developments and Future Initiatives