FAIRCORE4EOSC

Developing EOSC-Core components to enable a FAIR EOSC ecosystem

05 | 04 | 2023 by Tommi Suominen (ORCID, LinkedIn), CSC – IT Center for Science

Funded by the European Union
Call title: Deploying EOSC-Core components for FAIR Research and Innovation Action

Budget: 10 million EUR

Duration: June 2022 – May 2025

Consortium: 22 partners, coordinated by CSC – IT Center for Science

Website: faircore4eosc.eu

Key results: In response to the gaps identified in the SRIA, the project will develop nine new EOSC-Core components aimed to improve the discoverability and interoperability of an increased amount of research outputs.
Position of EOSC according to the European Commission

"EOSC is the basis for a science, research and innovation data space that will bring together data resulting for research and deployment programmes and will be connected and articulated with the sectoral data spaces"

(European Data Strategy, COM(2020) 66 final)
The European Open Science Cloud (EOSC) is an ecosystem of research data and related services that will enable and enhance seamless access to and reliable re-use of FAIR research objects (including data, publications, software, etc.).

The Strategic Research and Innovation Agenda (SRIA) for EOSC was created in 2021, as a roadmap for future development. Priorities highlighted in the SRIA are the establishment of the Web of FAIR data and a Minimum Viable EOSC (MVE) by 2027, that is the core components and functions to enable EOSC to operate (the EOSC-Core).

Context

Enhancing FAIRness in the EOSC ecosystem
Challenges addressed

Developing the EOSC-Core

The EOSC-Core development has been initiated in the Horizon 2020 calls, but some of the challenges that require to be addressed are:

- **Identifiers**: Introducing new resource types; machine-actionable persistent identifiers (PIDs); establishing a PID meta-resolver; standardising PID graphs; PID compliance framework to ensure compliance to the EOSC PID policy and to ensure quality of service for PIDs;

- **Metadata and Ontologies**: Provide or embrace/stimulate existing registries of metadata schemas, ontologies and crosswalks, develop services that build on metadata registries and can facilitate the creation and sharing of crosswalks;

- **Interoperability**: Enable discovery of data sources available in different formats, making search tools available; Provide tools for quality validation of metadata records and of digital objects; Implement EOSC PID Policy;

- **Research Software**: metadata description standards for research software, automated deposit of new releases into a scholarly repository and Software Heritage.
The 9 FAIRCORE4EOSC components

RDGraph
EOSC Research Discovery Graph

EOSC Research Discovery Graph (RDGraph) is a flexible and federated EOSC search service across EOSC repositories that extends EOSC Research Catalogue.

PIDGraph
EOSC PID Graph

Services for providing access to the PID Graph, which is made up of links and records gathered from persistent identifier (PID) authority data sources.

MSCR
EOSC Metadata Schema and Crosswalk Registry

Support publishing, discovery and access of metadata schemas and crosswalks and provide functions to operationalise metadata conversion by combining crosswalks.
The 9 FAIRCORE4EOSC components

**DTR**
EOSC Data Type Registry

Provide user friendly and machine actionable Interfaces for the registration and usage of Data Types and Kernel Information Profiles.

**PIDMR**
EOSC PID Meta Resolver

Provides users with a common interface to resolve different types of PIDs regardless of their originating system. The PIDMR either resolves to a given URI or provides Kernel Information Profiles if available.

**CAT**
EOSC Compliance Assessment Toolkit

The Compliance Assessment Toolkit will support the EOSC PID policy with services to encode, record, and query compliance with the policy.
The 9 FAIRCORE4EOSC components

**RAiD**

EOCS Research Activity Identifier Service

The EOSC RAiD will mint PIDs for research projects, which will allow authorised EOSC users and services to manage information about project-related participants, services, and outcomes.

**RSAC**

EOCS Research Software APIs and Connectors

Ensure the long-term preservation of research software in different disciplines. APIs and connectors will be developed to interconnect research outputs infrastructures with the Software Heritage universal source code archive, using the CodeMeta standard, and the Software Heritage intrinsic identifiers (SWHID).

**SWHM**

EOCS Software Heritage Mirror

Equip EOSC with a mirror of the Software Heritage universal source code archive. In order to prevent information loss, a mirror of Software Heritage will be established by GRNET to serve the EOSC community and will be updated regularly to follow the growth of the universal source code archive.
FAIRCORE4EOSC | Tommi Suominen

Impact

**PROJECT OUTPUTS**

- New FAIRCORE4EOSC components fully integrated in the EOSC-Core

- EOSC CAT
- EOSC PIDGRAPH
- EOSC PIDMR
- EOSC DTR
- EOSC RAID
- EOSC RSA
- EOSC SWHM

**OUTCOMES**

- Contribution to the HE EOSC Partnership (i.e. establishment of the EOSC MVE)

- European researchers can find, access and re-use an increasing amount of research outputs across borders and disciplines

**SPECIFIC PROJECT IMPACTS**

- Improve FAIRness of Science
- Advance the establishment of the MVE
- Enhance the EOSC Interoperability Framework
- Increase uptake of FAIRCORE4EOSC components and EOSC-Core services

**DESTINATION HE WORK PROGRAMME MOST RELEVANT IMPACTS**

- Improving trust in science through increased FAIRness, openness and quality of scientific research in Europe
- Transforming the way researchers create, share and exploit research outputs within and across research disciplines
- Seamless access to and management of increasing volumes of research data following FAIR principles and other research outputs
Case Studies

The case study aims to meet domain-specific requirements of research communities for common data services that improve discovery, access and reusability of research data. Leveraging the EUDAT services, the case study will act as a rule model for other service providers to increase the adoption of the developed components.
Data Type Registry

Machine actionable standardized PID metadata

T4.3 H. Lienhop, S. Bingert
The Data Type Registry

How?

- Provide a hierarchical model of basic data type descriptions

- Allow users to register data types depending on community needs
  - Data types can be as simple or complex as necessary
  - Each type will be provided with a PID and a common set of metadata

- The **DTR Toolkit** will allow users to further work with the registered types:
  - Create validation schemas for types
  - Freely explore the registered types and relations between them
  - Precisely search for types to avoid duplications
The DTR and the MSCR will closely cooperate regarding interoperability and reusability.

Possible interaction upon needing to set a type for a field in the MSCR:

- Request a list of types based on a keyword
- Provide a list of suggested types
- Further specify search parameters
- Provide refined list of suggested types
- Send request to create new type
- Return PID of requested type
Metadata Schema and Crosswalk Registry

A (meta)data interoperability service

T4.2 Tommi Suominen, Joonas Kesäniemi
MSCR - Requirements

1. The MSCR is a registry to support publishing, discovery and access of metadata schemas and crosswalks.

2. Schemas and crosswalks can be hosted in the service or registered when hosting already takes place externally (including machine actionable capacity to access those schemas).

3. Provide a mechanism to operationalise metadata conversions by combining crosswalks.

4. Supporting minting (PIDs and) metadata descriptions for metadata schemas, crosswalks, and data types thus making resources more FAIR.

5. Develop the mechanism and guidelines for community or individual users to register metadata schemas and create crosswalks (facilitate projects and researchers to create and share crosswalks with others that can reuse and improve them – crosswalk versioning).

6. Create a GUI for visually creating crosswalks.

7. The MSCR is designed to facilitate conversion between metadata schemas and integrate DTR in the metadata schema registry using the DTR API. This means typing metadata schema’s elements and attributes and using registered data-types and data-type converters for format conversion.
MSCR will base on the Finnish Interoperability Platform

Production ready platform with a proper development road-map ([GitHub](https://github.com)), open source

Maintained and developed by the [Digital and population data services agency (DVV)](https://www.dvv.fi)

Five applications (Java SpringBoot + Typescript React/Next.js)

- Group management
- Codelists
- Terminologies
- Data vocabularies
- Commenting

[Shared UI components](https://www.dvv.fi) and design language

Data layer is a mixture of graph database and relational db

Elasticsearch/OpenSearch, you know, for search
MSCR extends the Interoperability platform

Customizing and extending the Finnish Interoperability platform
- Simplified user interfaces
- MSCR theme
- Group admin functionality
- New domain objects (schemas and crosswalks)
- PIDs for everything

New application: Crosswalks
- Dedicated for creating and reusing mappings
- Internal data model for describing crosswalks
- Custom data types via DTR integration

Tools for facilitating operationalization of crosswalks
- Implementation details are still open
  - Downloadable and actionable crosswalk definition, generated code, ready to run transformer images, job queue, synchronous service endpoint...
We are FAIRCORE4EOSC!
The HE INFRAEOSC Projects

AI4 EOSC

FAIR-EASE EOSC Focus
cancer

RAISE EOSC

FAIR-IMPACT EOSC

FAIRCORE4EOSC

EuroScienceGateway
faircore4eosc.eu
Twitter: @FAIRCORE4EOSC
LinkedIn: company/faircore4eosc
Youtube: FAIRCORE4EOSC

Scan to watch the webinar

Funded by the European Union