eosc | FAIRCORE4EOSC
Core Components Supporting a FAIR EOSC

# FAIRCORE4EOSC

## Developing EOSC-Core components to enable a FAIR EOSC ecosystem

30 | 03 | 2023 by Tommi Suominen, CSC – IT Center for Science

Funded by
the European Union

# FAIRCORE4EOSC in a nutshell

**Call title:** Deploying EOSC-Core components for FAIR Research and Innovation Action

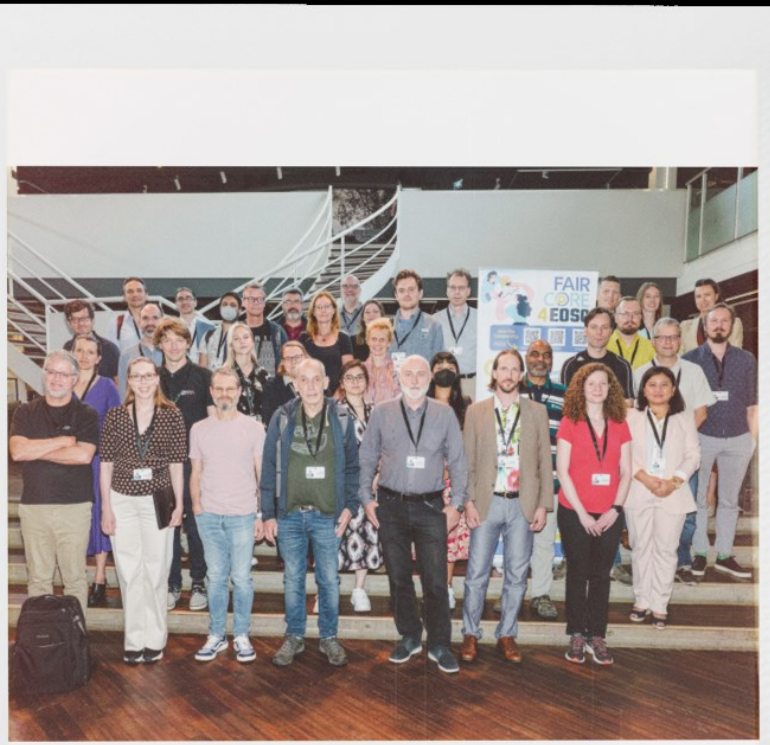
**Budget**: 10 million EUR

**Duration:** June 2022 – May 2025

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

**Website:** [faircore4eosc.eu](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.




Amsterdam, Netherlands – Kick-off meeting, June 2022

# 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

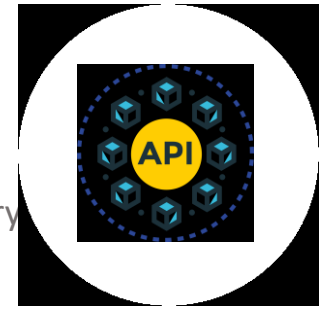
Support publishing, discovery and access of metadata schemas and crosswalks and provide functions to operationalise metadata conversion by combining crosswalks.

# 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

EOSC 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

EOSC 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

EOSC 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.


# Case studies

## How do the components benefit communities?

Components are co-developed and tested within domain-specific communities:

- Climate Change (DKRZ)
- European Integration of National-level Services (CSC)
- Mathematics (FIZ)
- Service Providers and Research Data Management Communities (EUDAT)
- Social Sciences and Humanities (CLARIN)

eosc | FAIRCORE4EOSC
Core Components Supporting a FAIR EOSC

# Data Type Registry

## Machine actionable standardized PID metadata

T4.3 H. Lienhop, S. Bingert

Funded by
the European Union

# 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


- 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

eosc | FAIRCORE4EOSC
Core Components Supporting a FAIR EOSC

# Metadata Schema and Crosswalk Registry

## A (meta)data interoperability service

T4.2 Tommi Suominen, Joonas Kesäniemi

Funded by
the European Union

# 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), open source

Maintained and developed by the Digital and population data services agency (DVV)

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

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

Shared UI components and design language

Data layer is a mixture of graph database and relational db

Elasticsearch/OpenSearch, you know, for search

# Development speed through reuse of an existing solution

Basic idea: Reuse as much as possible on all layers of the application

Same UI components

Same OWL/SHACL based internal model for data models

Same SKOS(–XL) based internal model for vocabularies

Same libraries - Spring, Jena, React, Next.js (Java + Typescript)

Same build tools

Same architecture - RDF database, ElasticSearch, API, JS Frontend

Same deployment - Docker, Docker-compose, Kubernetes

- Development and test instances in Rahti

# 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…

# Transformations between …

## … Schemas

- XML schema description (XSD)
- JSON schema description (JSON Schema)
- CSV schema description (CSV on the Web, Simple table)
- RDF schema description (RDFS, SHACL, ShEx)

## … Codelists

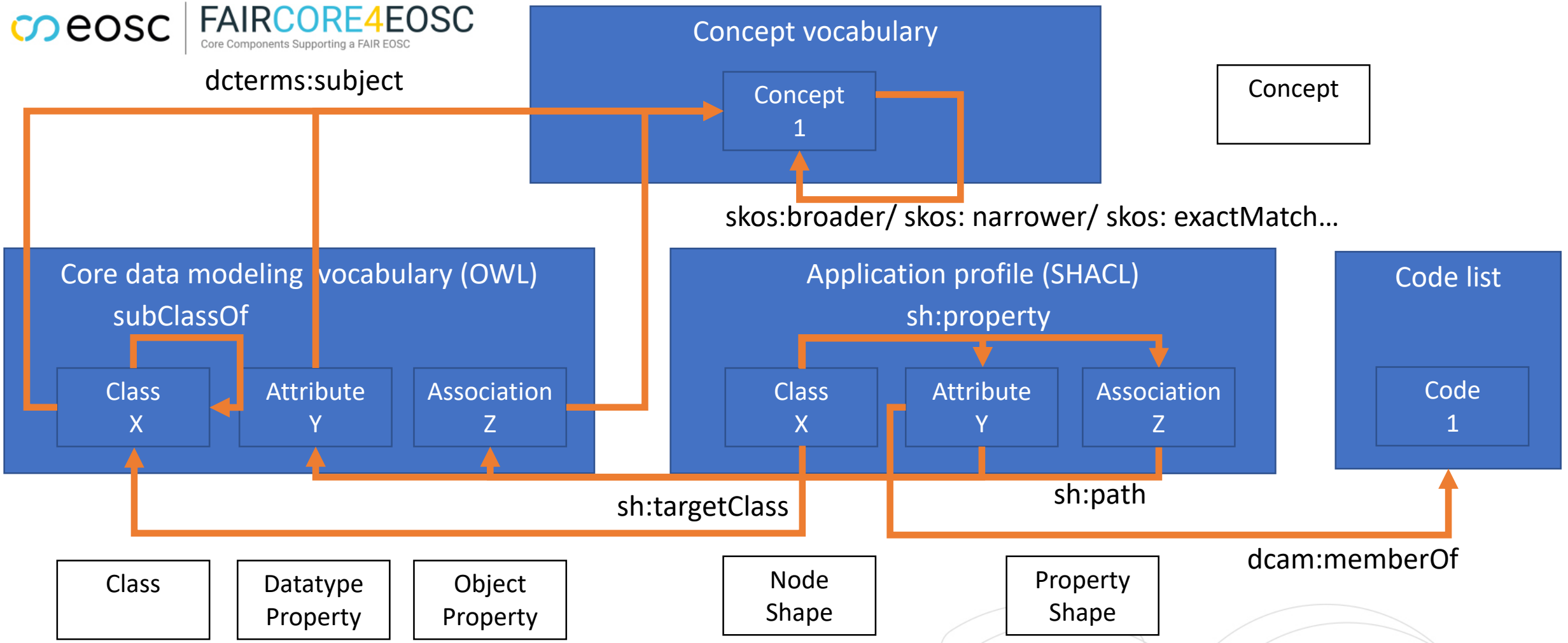
- XML schema enumeration
- JSON schema enum
- Simple list
- SHACL (sh:in)

## … Data types

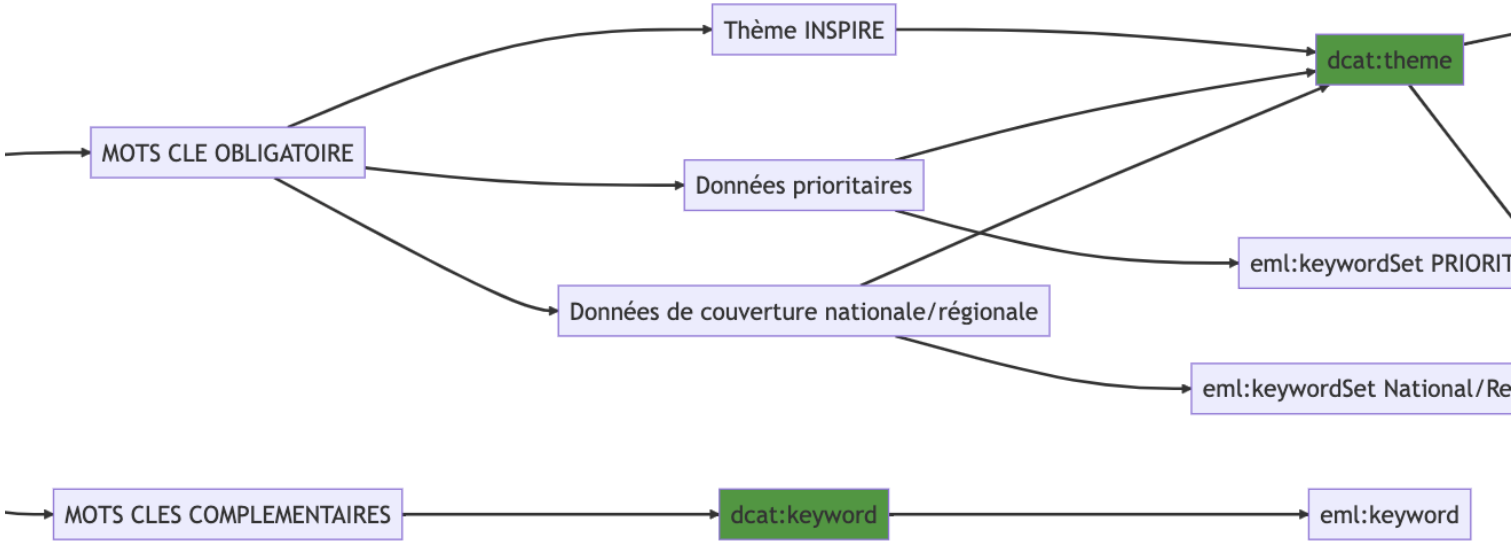
- Predefined list of XSD data types
  - RDF1.1 compatible types
- Imported data type vocabularies
- Custom DTR data types

## … Concept vocabularies

- SKOS(-XL) vocabularies


eosc
FAIRCORE4EOSC
Core Components Supporting a FAIR EOSC
dcterms:subject
Concept vocabulary
Concept 1
Concept
skos:broader/ skos: narrower/ skos: exactMatch…
Core data modeling vocabulary (OWL)
subClassOf
Class X
Attribute Y
Association Z
Application profile (SHACL)
sh:property
Class X
Attribute Y
Association Z
Code list
Code 1
sh:targetClass
sh:path
dcam:memberOf
Class
Datatype Property
Object Property
Node Shape
Property Shape

# Further thinking




Brokering WG RDA17:

- Not all mediations are reversible

- Some mediations are non-commutative

Daan Broeder RDA20:

- Information loss due to aggregation (lossy and non-reversible) of metadata may harm data discovery as fitness for purpose may be ” more difficult for user to evaluate due to the information loss.”

SSSOM: ”*Mappings are directional, i.e. they are defined in one direction (from ‘subject’ to ‘object’). Whether a mapping can be interpreted back (from the ‘object’ to the ‘subject’) is purely defined by the semantics of the predicate (e.g. owl:equivalentClass is symmetric as defined by the OWL specification).*”

Matentzoglu et al., A Simple Standard for Sharing Ontological Mappings (SSSOM), Database, Volume 2022, 2022, baac035, https://doi.org/10.1093/database/baac035

Figure: https://github.com/cnigfr/metadonnee/tree/main/MappingINSPIRE-DCAT__EML

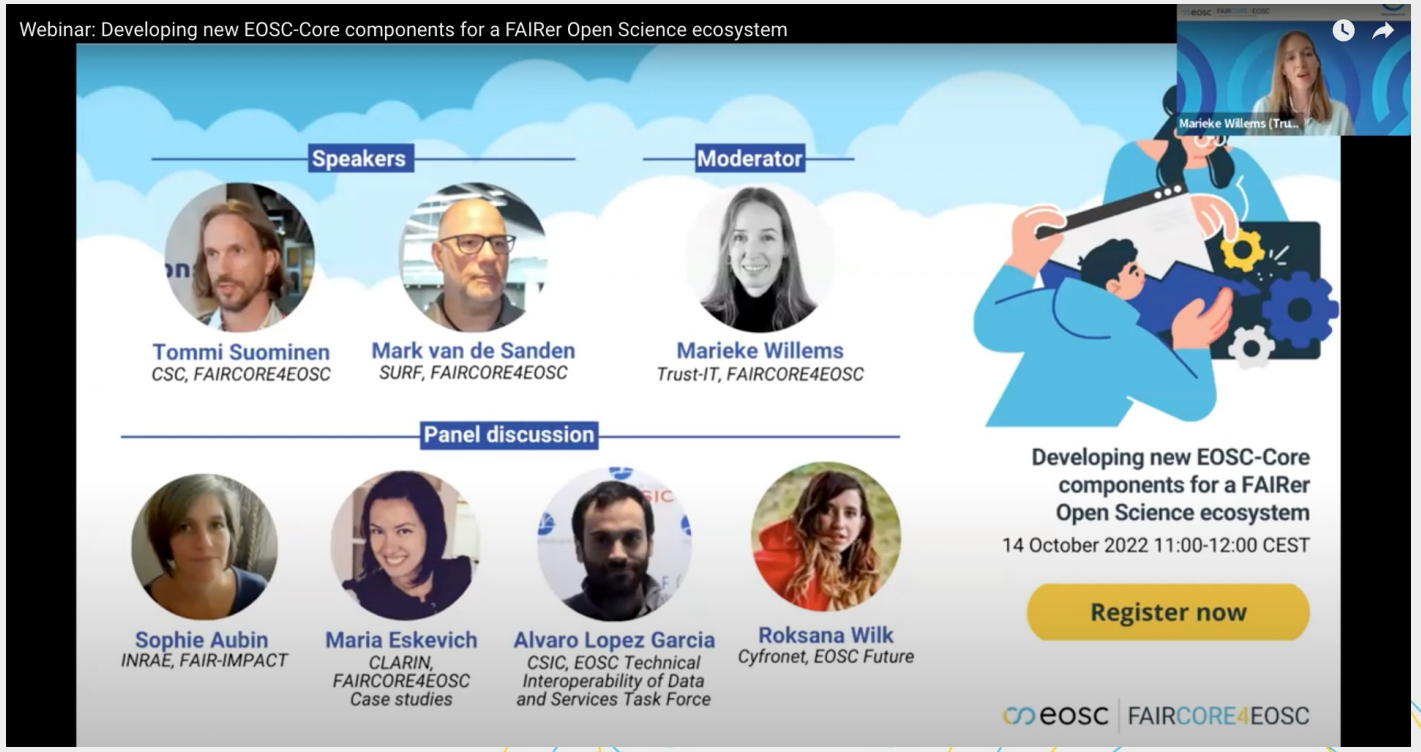

GWDG
Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen
FIZ Karlsruhe
grnet
CERN
DataCite
FIND, ACCESS, AND REUSE DATA
SURF
We are FAIRCORE4EOSC !
CYFRONET
SCHLOSS DAGSTUHL
Leibniz-Zentrum für Informatik
CSC
HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI
16 40
THE NATIONAL LIBRARY OF FINLAND
ARDC
Australian Research Data Commons
OpenAIRE
Consiglio Nazionale delle Ricerche
DKRZ
DEUTSCHES KLIMARECHENZENTRUM
Data Archiving and Networked Services
DANS
ATHENA
Research & Innovation
Information Technologies
Trust-IT Services
Communicating ICT to markets
Inria
Software Heritage
THE GREAT LIBRARY OF SOURCE CODE
CLARIN

# FAIRCORE4EOSC

Core Components Supporting a FAIR EOSC

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

Scan to watch the webinar




Funded by the European Union