

FAIRCORE4EOSC

Developing EOSC-Core components to
enable a FAIR EOSC ecosystem

01 | 12 | 2022 by Tommi Suominen, CSC – IT Center for Science



**Funded by
the European Union**



Context

Enhancing FAIRness in the EOSC ecosystem

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).



2021



Minimum Viable 

Web of FAIR Data

Findable Accessible Interoperable Reusable



2027

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.



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

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.



eosc | FAIRCORE4EOSC

Amsterdam, Netherlands – Kick-off meeting, June 2022

The 9 FAIRCORE4EOSC components



EOSC Research Discovery Graph (RDGraph) to deliver advanced discovery tools across EOSC resources and communities.



EOSC PID Graph (PIDGraph) to improve the way of interlinking research entities across domains and data sources on the basis of PIDs.



EOSC Metadata Schema and Crosswalk Registry (MSCR) to support publishing, discovery and access of metadata schemas and provide functions to operationalise metadata conversions by combining crosswalks.



EOSC Data Type Registry (DTR) to provide user friendly APIs for metadata imports and access to different data types and metadata mappings.



EOSC PID Meta Resolver (PIDMR) to offer users a single PID resolving API in which any kind of PID can be resolved through a single, scalable PID resolving infrastructure.



EOSC Compliance Assessment Toolkit (CAT) to support the EOSC PID policy compliance and implementation.



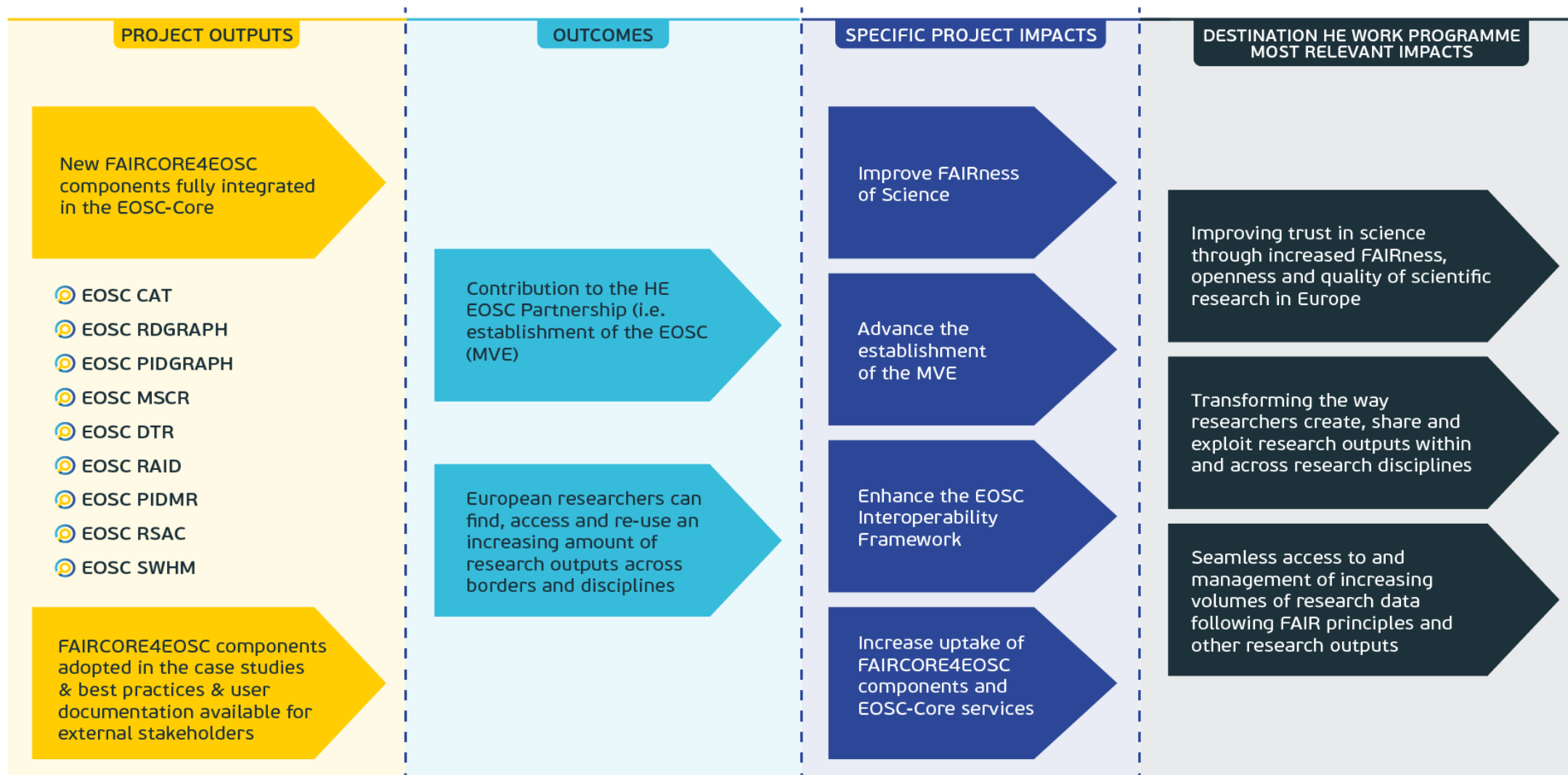
EOSC Research Activity Identifier Service (RAiD) to mint PIDs for research projects, allowing to manage and track project related activities.



EOSC Research Software APIs and Connectors (RSAC) to ensure the long-term preservation of research software in different disciplines.



EOSC Software Heritage Mirror (SWHM) to equip EOSC with a mirror of the Software Heritage universal source code archive.



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

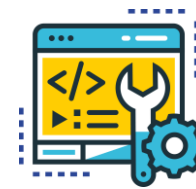
Provide user friendly and machine actionable Interfaces for the registration and usage of Data Types and Kernel Information Profiles.



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)

Case Studies

Social Sciences and Humanities

CLARIN
Common Language Resource Infrastructure

This case-study will focus on improving the discoverability of CLARIN data through the integration of the Digital Object Gateway (DOG), a crucial component for the interoperability of the CLARIN infrastructure, Language Resource Switchboard and Virtual Collection Registry tools.

Adopted components

DTR MSCR PIDGraph PIDMR RDGraph

Climate Change

DKRZ
DEUTSCHES KLIMARECHENZENTRUM

ENES supports climate modellers in their work, in particular in the area of data management. In this case study we demonstrate how the developed EOSC-Core components can improve the discoverability and re-use of research results from the ENES community.

Adopted components

DTR MSCR PIDGraph RAiD RDGraph

FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

Mathematics

zbMATH Open & swMATH projects aggregate significant scientific advances in mathematics and related disciplines supporting researchers in finding relevant publications and data. The case study will increase the discoverability of the zbMATH Open and swMATH data and services in the mathematical and EOSC community.

Adopted components

MSCR PIDGraph PIDMR RDGraph RSAC

CSC

European Integration of National-level Services

The case study will showcase how the developed components can enrich the content of the national research information systems displaying international connections to research objects and improve their interoperability.

Adopted components

DTR MSCR PIDGraph RAiD RDGraph RSAC SWHM

EUDAT Collaborative Data Infrastructure
Data shared and preserved across borders and disciplines

EOSC Service Providers

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.

Adopted components

CAT DTR MSCR PIDGraph RAiD RDGraph

European Integration of National-level services

- National systems for CRIS-like information are a valuable source of information for e.g. research assessment or to highlight EOSC-related contributions - if interoperability can be achieved
 - National systems provide both structured data AND usually offer high coverage on research outputs (e.g. publications, datasets, research profiles) and the connections between entities, with the possibility of sharing this information (compared to e.g. commercial databases)
 - National delimitation of data – the national CRIS can use RDGraph or PID graph to access information on international connections
 - Aggregating information from national services to RDGraph by OpenAIRE
- CERIF data model (*Common European Research Information Format*) maintained by [euroCRIS](https://europe.cerif.eu/)
 - A shared model for exchanging information between organizational and national CRIS systems & OpenAIRE
 - Supported to some extent by national services (traditional mappings done for some object types)
 - Currently under refactoring at <https://github.com/EuroCRIS/CERIF-Core>

This case study integrates 3 national services



Finland - National Research Information Hub project that was launched in 2017 and the portal www.research.fi was published in 2020 - Research.fi collects and disseminates information on Finnish research

The Netherlands – A business case is in development to collect, analyze, report, archive, disseminate information on Dutch research. See also: [feasibility study](#) OKB for a Dutch national service.

Greece – [HELIX \(Hellenic Data Service\)](#) is the national eInfrastructure for data-intensive research and [OS](#). Provides a catalogue/repo of open scientific data (central & federated) and access to interactive computing environments (Jupyter Lab)

Benefits:

- Realize the potential in combining CRIS-like information from national sources on European level, aggregating this information to RDGraph and OpenAIRE
 - Resolve issues with interoperability of CRIS-like metadata
 - Make it possible to utilize this well structured and high coverage metadata on research outputs and activities
 - Less manual input of *anything* research information related
- Showcase the CERIF data model for exchanging information between organizational and national CRIS systems & OpenAIRE
- Have a template for the integration of national research information systems with the EOSC Core

Components to be used:



Research Discovery
Graph



PID Graph

Utilise the MSCR to create crosswalks from organizational/national/European services to CERIF and utilizing these for data exchange. Data Type Registry support the concrete conversions



Research Activity
Identifier Service

Enriching the research metadata on outputs, exploring workflows and graphs and implementing them in national CRIS-systems and providing national information to OpenAIRE



Metadata Schema and
Crosswalk Registry



Data Type Registry

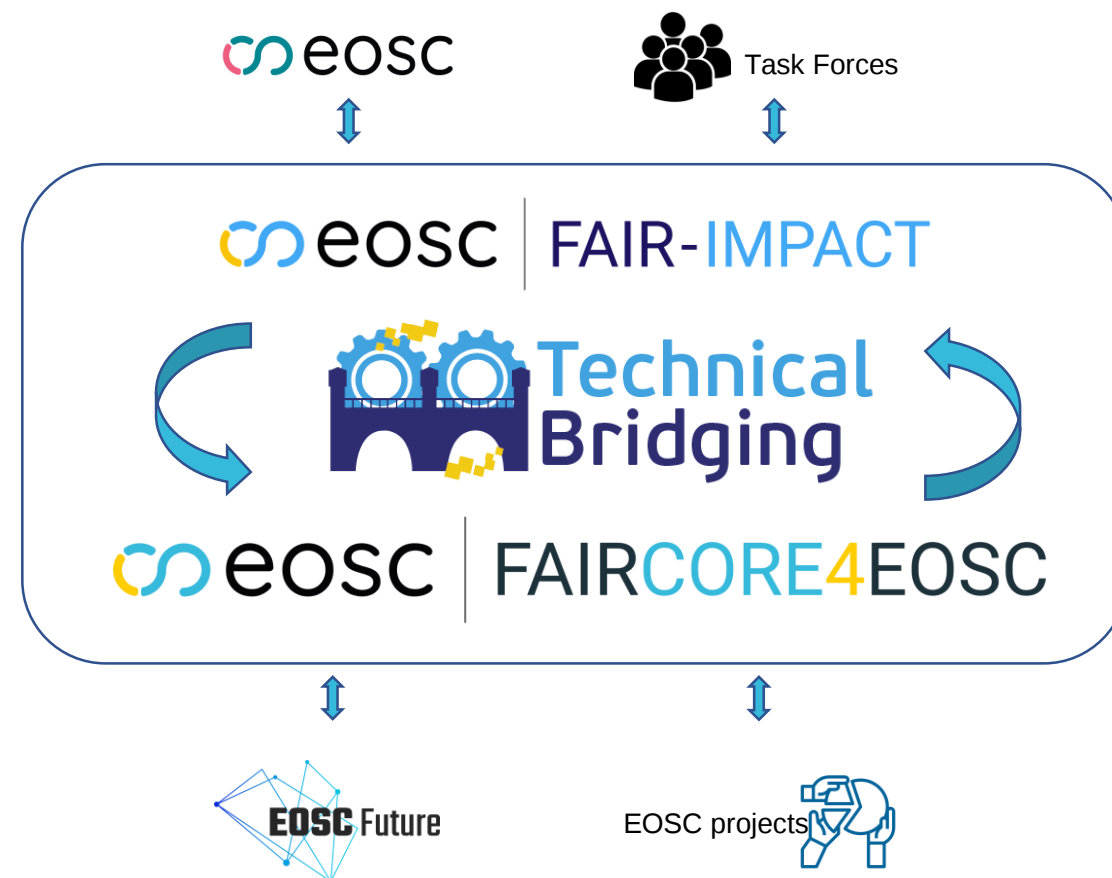
Explore and implement workflows for registering RAIDs through national CRIS' for projects - providing PID metadata envelopes for durable links between research entities

Technical Bridging Team

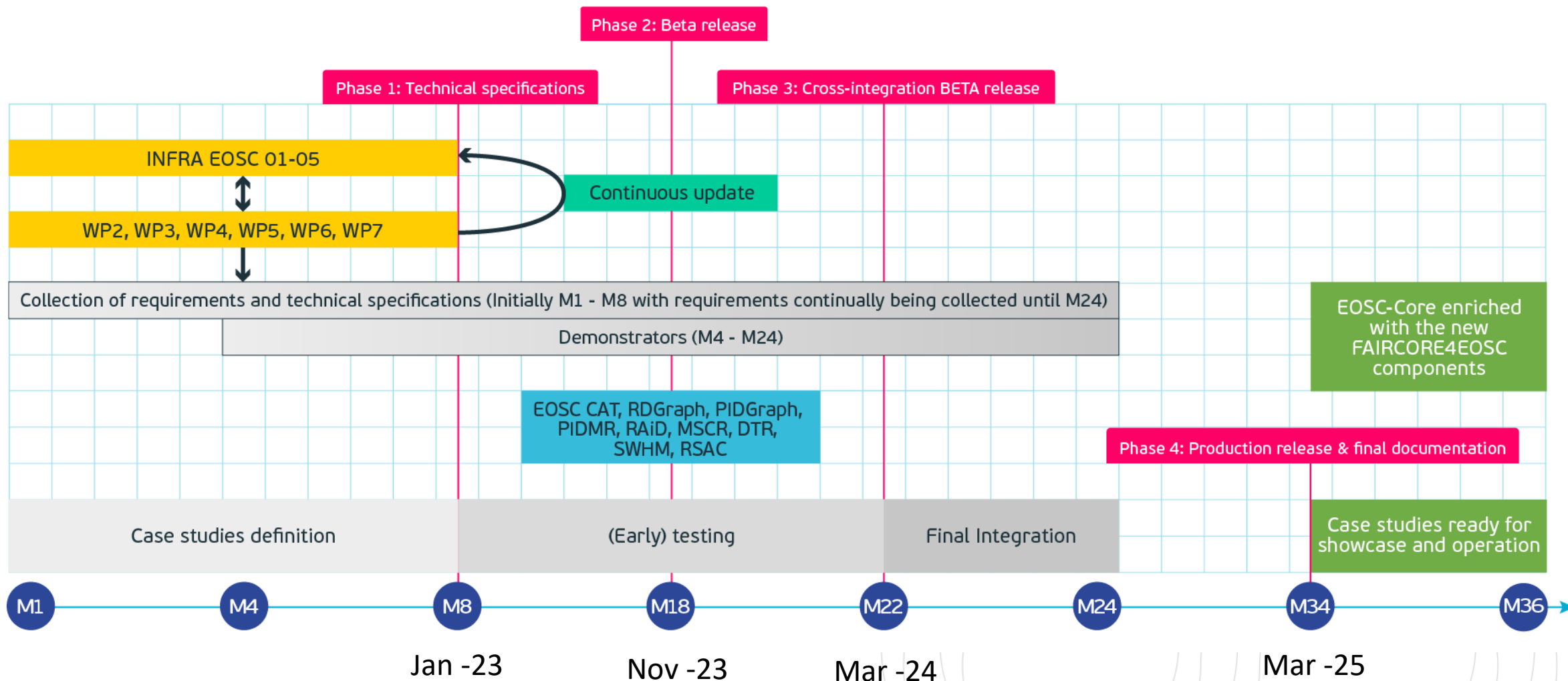
Aligning developments with the EOSC ecosystem

The Technical Bridging Team will provide a key link between the FAIRCORE4EOSC Technical Steering Group and the relevant interfaces in the Horizon Europe project FAIR-IMPACT.

FAIRCORE4EOSC will interact with the wider EOSC community, through dialogue and joint activities with the EOSC Association, its Task Forces, the EOSC Future project, and other key initiatives.



Technical implementation





SCHLOSS DAGSTUHL
Leibniz-Zentrum für Informatik



Consiglio Nazionale delle Ricerche

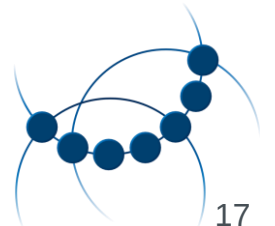


COMMpla
Multichannel Communication Platforms
and Online Solutions



Software Heritage
THE GREAT LIBRARY OF SOURCE CODE

CLARIN



Gesellschaft für wissenschaftliche
Datenverarbeitung mbH Göttingen



We are FAIRCORE4EOSC !



Australian Research Data Commons



FIND, ACCESS, AND REUSE DATA



The HE INFRAEOSC Projects

AI4 |  eossc

 eossc | FAIR-EASE  eossc | Focus

 eossc | cancer

 eossc | RAISE  eossc | FAIR-IMPACT

 eossc | FAIRCORE4EOSC

 eossc | EuroScienceGateway

faircore4eosc.eu

Twitter: @FAIRCORE4EOSC

LinkedIn: company/faircore4eosc

Youtube: FAIRCORE4EOSC

Watch the webinar [here](#)
or scan the QR code below



Funded by
the European Union

Webinar: Developing new EOSC-Core components for a FAIRer Open Science ecosystem

Speakers



Tommi Suominen
CSC, FAIRCORE4EOSC



Mark van de Sanden
SURF, FAIRCORE4EOSC

Moderator



Marieke Willems
Trust-IT, FAIRCORE4EOSC

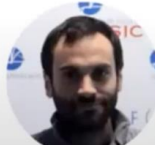
Panel discussion



Sophie Aubin
INRAE, FAIR-IMPACT



Maria Eskevich
CLARIN, FAIRCORE4EOSC
Case studies



Alvaro Lopez Garcia
CSIC, EOSC Technical
Interoperability of Data
and Services Task Force



Roksana Wilk
Cyfronet, EOSC Future

Developing new EOSC-Core components for a FAIRer Open Science ecosystem
14 October 2022 11:00-12:00 CEST

[Register now](#)

 **FAIRCORE4EOSC**