

EOSC Data Commons

Project Overview

Contribute to establishing EOSC as the **European Research Commons by federating repositories** from national, institutional and thematic initiatives with a **metadata warehouse** of paired data and software for **use and reuse of datasets** across the compute continuum



Call topic:

HORIZON-INFRA-2024-EOSC-01-05

Innovative and customizable services for EOSC Exchange

Project Officer:

Ourania SKONDRA
European Research Executive Agency
Unit REA.C.4

Budget: 7.000.000€

Start: 01/04/2025

Duration: 36 months

7

Work packages

29

Deliverables

26

Milestones

Consortium: 22, among those:

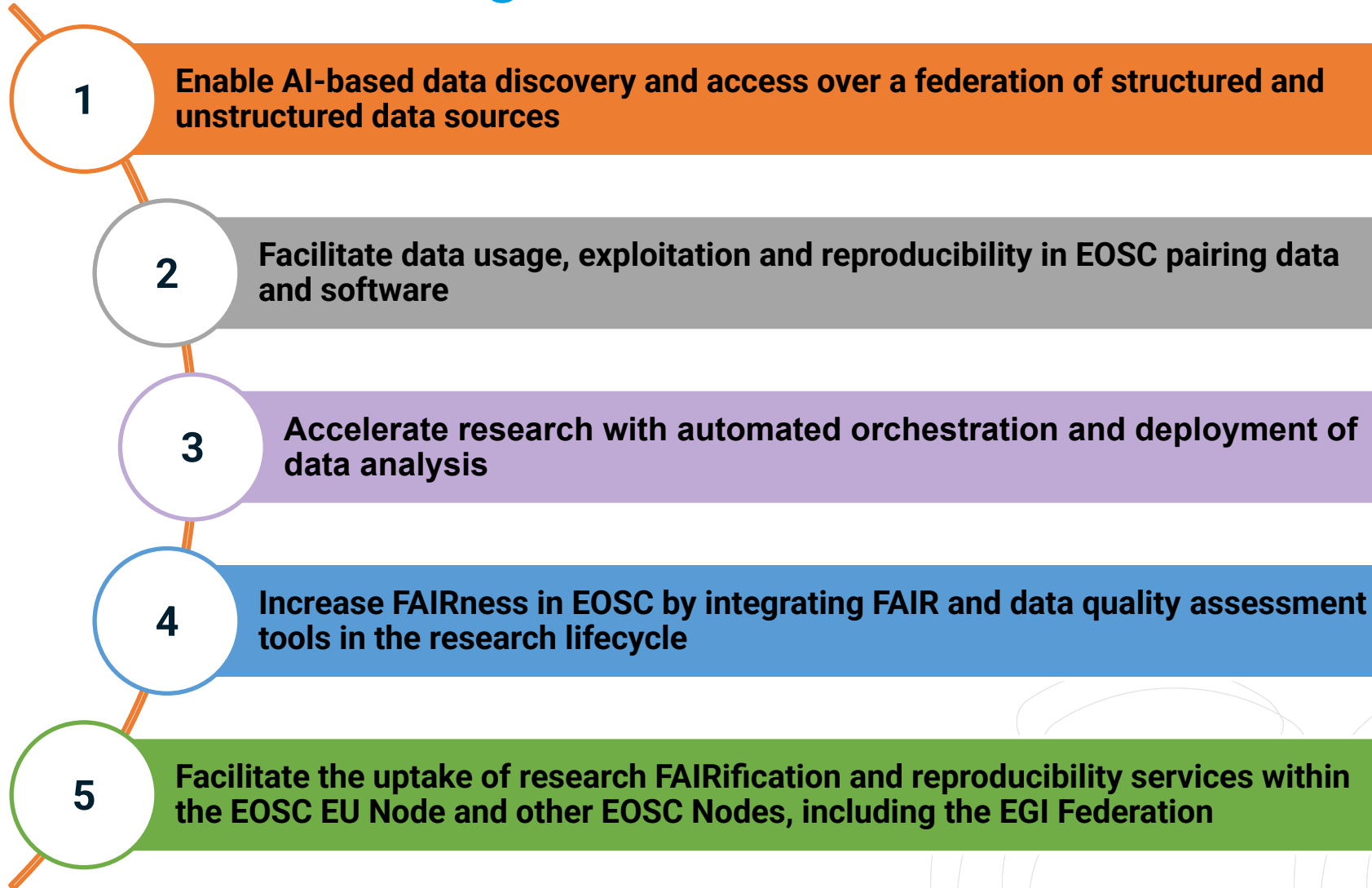
- National/international/thematic repositories owners: CNRS, SRCE, UZH, KNAW, EODC, TUD, SIB, FZJ, UH-Luomus, CERN
- Technology providers: SWITCH, UST/AGH, UPV, UFR, ETH, INFN ...
- Research communities (brought by): UZH, CERN, CESNET, CIRMMP, CSIC, UU, CNRS, EODC ...

Coordinator: EGI Foundation



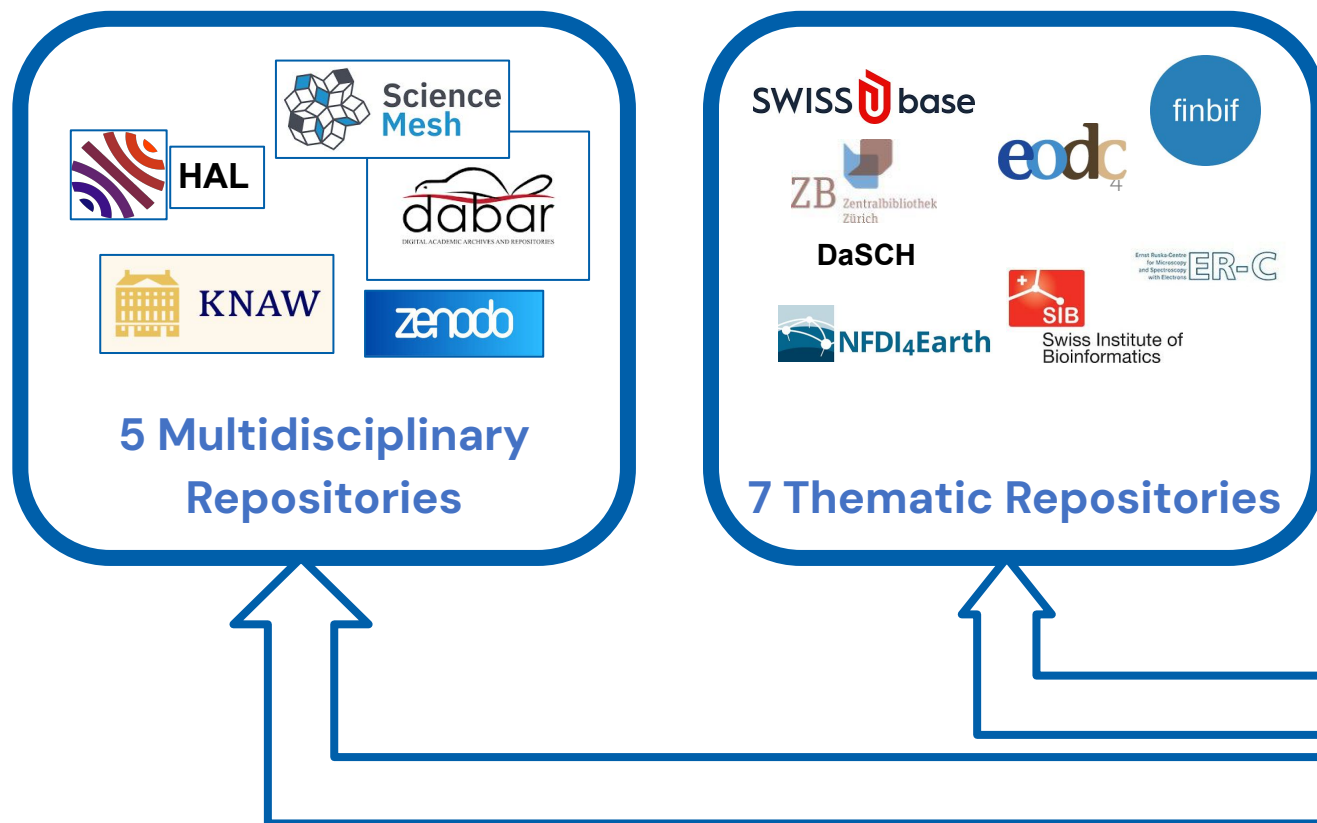
Project Objectives

Key contributions to EOSC



Make data available in e-Infrastructures

Partnership with Repository Owners and Data Providers



National Approach to bring data in EGI & EOSC

5 Countries as Data Providers



France



Finland



Croatia



Switzerland

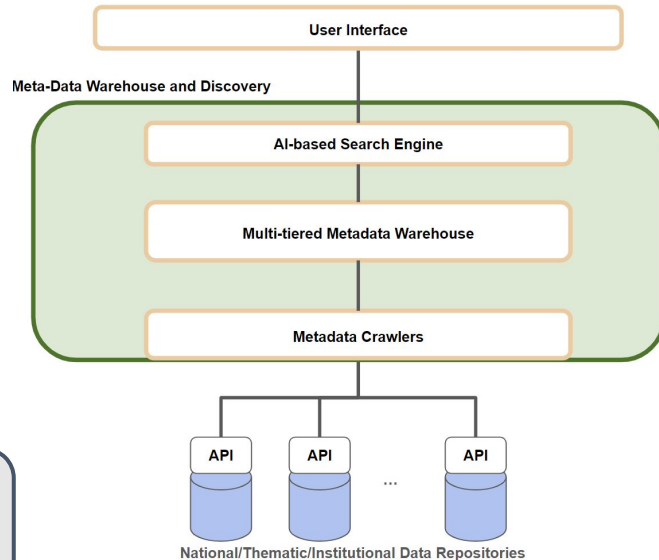


Germany

**Common discovery channel
for Data from 12 National
Repositories in EGI Infra**

Data Discovery and Access

AI-based natural language search tools



KER1
Metadata Warehouse and
Discovery Service

5

Pairing Data & Analytics

AI mechanisms that suggest
compatibility between Data & Tools
&
Automated Orchestration



Datasets



Analytics

**Package for processing
datasets**

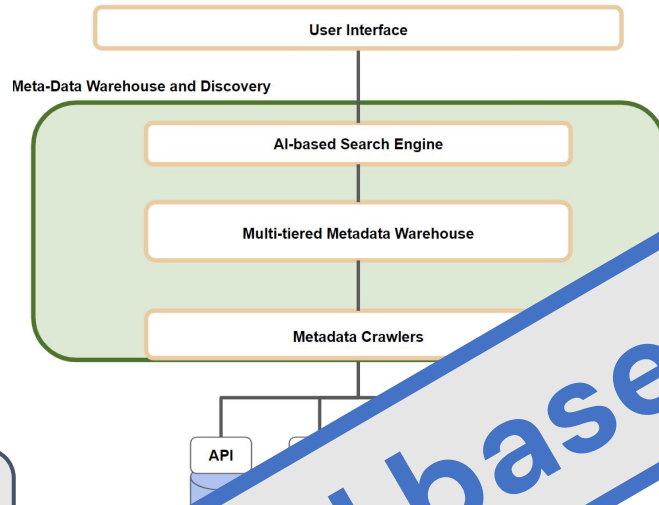
KER5
Metadata specification for
processing datasets

KER3
Catalogue of Analytics

KER4
Execution Service

Data Discovery and Access

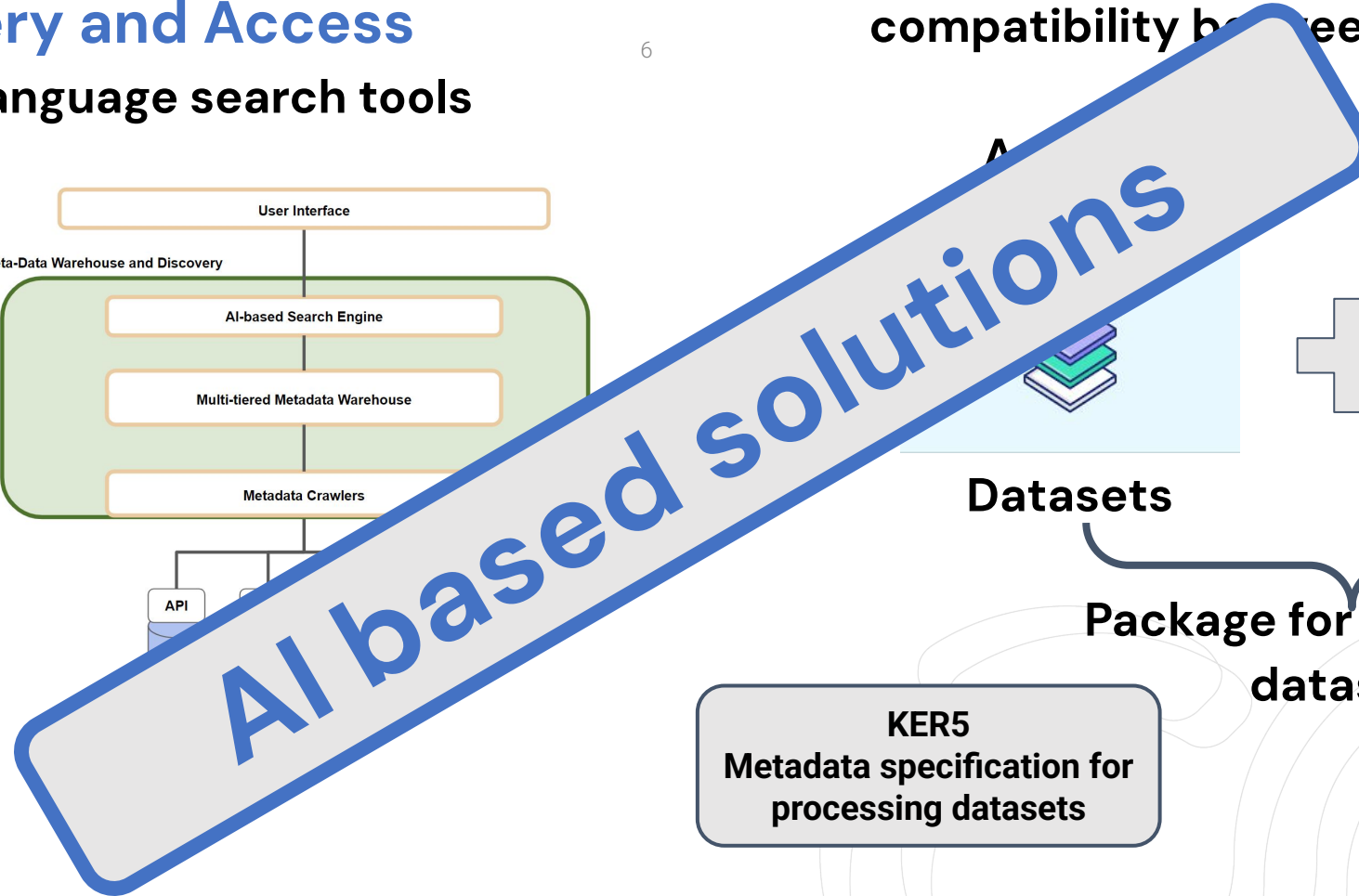
AI-based natural language search tools



KER1
Metadata Warehouse and
Discovery Service

Pairing Data & Analytics

AI mechanisms that suggest
compatibility between Data & Tools



Datasets



Analytics

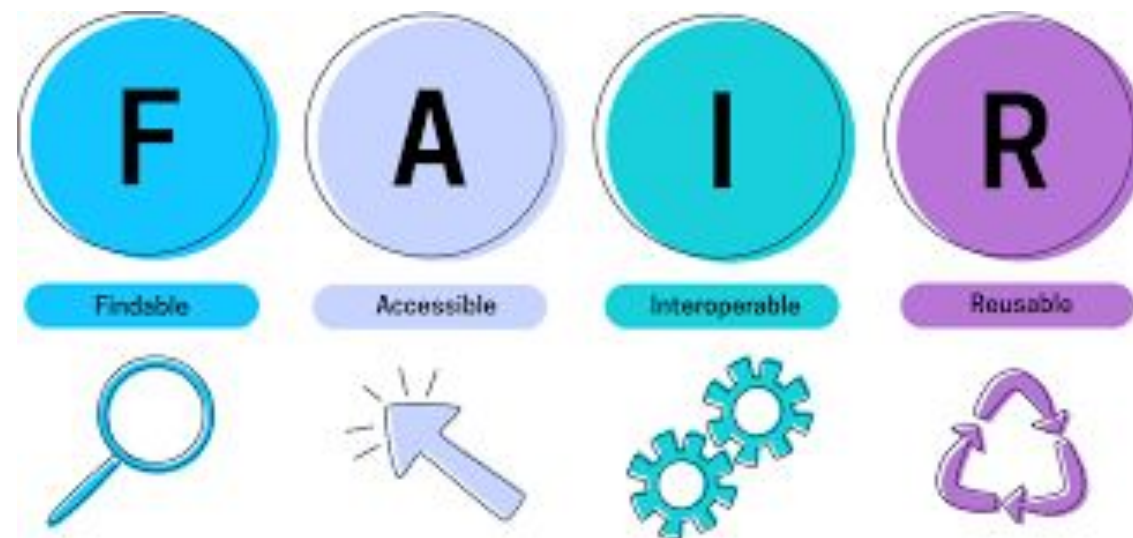
Package for processing
datasets

KER5
Metadata specification for
processing datasets

KER3
Catalogue of Analytics

KER4
Execution Service

- More data in EOSC with increased quality → **Integration of FAIR assessment Infrastructure & Tools**
- Increase **FAIRness** for:
 - recently produced data (e.g. from instrumentation or as a result of analysis)
 - output of research analysis
 - provenance metadata that aligns with reproducibility expectations (e.g. by providing an RO-Crate record)
 - composing a basic metadata record in a convertible format (D-CAT, DataCite...)
 - and/or create a BagIt archive
- FAIR assessment Tools offered to participating infrastructures and repositories



KER6
FAIRness assessment and reproducibility toolset and policies

EOSC Data Commons will build two services:

- **EOSC Matchmaker: AI-Based Analytics-Oriented Metadata Warehouse and Discovery Service**
 - Aggregates and enriches metadata from repositories in a Metadata Warehouse for the discovery of datasets and linking of these with analytics tools in a format that can be processed by orchestration services for their execution.
- **EOSC Data Player: Data Analysis Orchestration and Deployment Service**
 - Runs analytics tools and accesses data through unified data access interfaces. A plugin-based dispatcher is capable of running tasks on heterogeneous compute resources with automated on-demand orchestration of computing services.

Impact Canvas

Needs

N1: Availability of efficient data discovery and seamless access over a multitude of structured and unstructured data from different scientific domains (P01)

N2: Streamlined data usage, exploitation and reproducibility in EOSC (P02)

N3: Availability of automated orchestration and deployment of data and software (P03)

N4: Increased availability of FAIR data in EOSC (P04)

N5: Services for research lifecycle available in the EOSC EU Node and other EOSC Nodes (P05)

Key Exploitable Results

KER1: AI-based Analytics Oriented Metadata Warehouse and Discovery Service

KER2: Federation of Data Repositories

KER3: Catalogue of Analytics for Research Communities

KER4: EOSC Data Commons Execution Service

KER5: Harmonised metadata specification for processing datasets

KER6: FAIRness assessment and reproducibility toolset and policies to assist Data FAIRness

Outcomes

- > **20** Data Repositories Federated
- > **2000+** tools in the catalogue
- > **12** analytics tools linked to datasets
- > **1** Interoperability Guideline > 6 Execution Engines
- > **3** Data Management Engines supported
- > **8** National repositories with datasets that can be analysed in the EOSC EU Node via EOSC Data Commons
- > **6** Use cases that execute data analysis with EOSC EU Node > 3 Tools and services for FAIRness assessment integrated
- > **10** research communities adopting the Data FAIRness toolset
- > **100** datasets with improved FAIRness quality
- > **2** Policies for FAIR data
- > **+3000** AI-based data discovery/month
- > **10** domain-specific metadata schema supported
- > **13** integrated use cases

Impacts

IMPACT1:

- > More innovation and higher productivity of research are enabled
- > New discoveries, breakthroughs, and innovations are facilitated

IMPACT2:

- > More collaboration across scientific disciplines enabled
- > Increased sharing and reuse of data
- > Systematic barriers hindering data and service sharing across disciplines, institutions, and countries removed

IMPACT3:

- > Wide ecosystem of multidisciplinary datasets and tools that will enable seamless discovery, access and reuse of the research outputs available in EOSC

IMPACT4:

- > Progressive improvement of the level of FAIRness in EOSC
- > Increased quality of research
- > Reproducibility of science greatly simplified
- > EOSC Data Commons services recognised as high-quality services by the research communities thanks to the successful use cases

EOSC Data Commons will develop building blocks for EOSC nodes

- Federation of data repositories with advanced AI-based discovery
- Facilitating data re-use and increasing FAIRness of available data
- Pairing data and analytics with automated deployment of workloads on the compute continuum

Interoperable services

- Usage of the Core Federating Capabilities: AAI, Monitoring, Accounting, Helpdesk, AWM
- OAI-PMH for harvesting
- Aiming at creating new ways to interlink EOSC Nodes with the Package for datasets processing

and with possible integration with the EOSC EU Node

- Enrich EU Node FAIR data feeding the EOSC Resource Catalogue with the metadata from the Data Commons repositories
- Publish Data Commons tools in the EU Node Tool Market and enable their execution through the EU Node AWM service
- Run analysis jobs and notebooks on the EOSC EU Node
- Make EU Node File Sync & Share and FTS services accessible for processing and move data



Get in touch with us

www.eosc-data-commons.eu (Spring 2024)

Email eosc-data-commons-po@mailman.egi.eu

