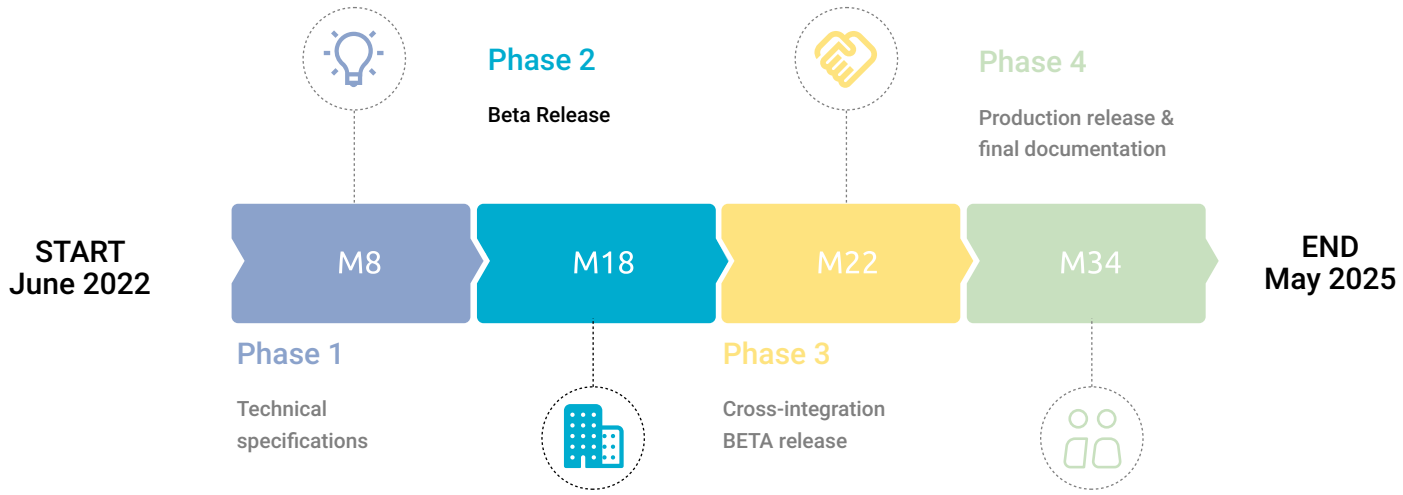
FAIRCORE4EOSC

Core Components Supporting a FAIR EOSC

Briefing #6

Top project achievements

1. RAiD (SAR): We have rebuilt the UI for the RAiD system web application using a more modern approach (https://app.demo.raid.org.au/home).
2. The API has also been revised and swagger implemented for documentation (https://api.demo.raid.org.au/swagger-ui/index.html#/raido-stable-v1 - focus on 'Stable V1', 'experimental' endpoints are being phased out).
3. The metadata schema is complete and stable; documentaiton is almost done (at https://metadata.raid.org). Additional documentaiton is being produced (at https://documentation.raid.org/).
4. Finally, acquiring DOIs via DataCite has been implemented in our test environment.



Technical updates

At the end of the last quarter the project reached its M18 Beta Release milestone. Therefore, most of the activities within the development teams have been working towards the beta release of the components. At end of the November the CAT development team released the beta release of the CAT component consisting of an API and a UI. Also automated tests have been developed to validate the CAT service. The CAT team also engaged with FAIR-IMPACT and the EOSC PID Task Force to collect user input. The team also participated in the FAIR-IMPACT Synchronisation Force and EOSC-A PID TF workshops to show the use of the CAT service.

The slides presented at the FAIR-IMPACT Synchronisation Force can be found via the following link, see https://fair-impact.eu/sites/default/files/2023-12/Synch%20Force%20Session%203%20PIDs_%20slides.pdf.

The CAT and the FAIR-IMPACT-related work is also embedded into the programme for the EOSC Winter School. The RDGraph development team has been progressing and is on track for the beta-release. They have been focusing on mapping from the internal RDGraph model towards a representation of the data suitable for the components implementing the NL Search and the Community Recommendation profiles, and on a workflow to automate this mapping process. Also, work has been done on enhancing the APIs to support impact-based ranking of research products, see https://graph.openaire.eu/docs/graph-production-workflow/indicators-ingestion/impact-indicators/.

The development team also created a demo interface to showcase the natural language search capabilities, see https://test.darelab.athenarc.gr/text-to-sql/docs. In collaboration with the mathematics community the RDGraph team has been working a collaborative filtering-based recommender system.

On the RAiD inference service, the development team has started on the service design, focusing on the design of the appropriate methodologies. The PIDGraph development team started the work with creation of the technical specification for the Data Dump service, followed by development work on the REST API for accessing the data dump files, and started the work for the creation and storage of the data dump files. Work also included improving the functionality of the underlying infrastructure and services that are the foundation for this component.

On the Data usage statistics service the work on this component focused on the validation of the Usage Tracker service with a repository (Dryad). In addition, the team worked on exposing usage statistics collected via Usage Tracker in DataCite Commons. The MSCR development team enabled a shared development version of the MSCR which has been deployed to the CSC's container platform, see https://mscr-test.rahtiapp.fi/. It contains both, the schema and crosswalk registry as well as a separate tool for vocabulary management. The current version allows the mapping of two schemas where users can map properties from the source to target to create crosswalk. The test instance has been integrated with the demo EOSC Core Infrastructure Proxy to test the integration of the AAI of the EOSC Platform and is accessible to anyone with proper credentials to support testing activities from demonstrators and case studies.

The RAiD development team has been rebuilding the UI of the RAiD system web application using a more modern approach, see https://app.demo.raid.org.au/home. The API has also been revised including a swagger implementation for documenting the API definition, see https://api.demo.raid.org.au/swagger-ui/index.html#/raido-stable-v1. The focus has been on 'Stable V1', 'experimental' endpoints are being phased out. The metadata schema has been completed and is stable, documentation is almost done and can be found at https://metadata.raid.org, additional documentation is being produced at https://documentation.raid.org/. Finally, acquiring DOIs via DataCite has been implemented in our test environment, implementation in the beta service is anticipated in January. While DOI acquisition implementation is ongoing, the climate change case study is using a RAiD demo service for testing purposes.

The PID Meta-resolver development team has setup a proxy service on basis of the Handle software which provides resolving service supporting currently PID from the following PID providers: ARK, arXiv, SWHID, German and Finnish URN:NBN, DOIs and Handles. The beta release of the PID Metaeresolver can be found at https://pidmr.devel.argo.grnet.gr/, but resolving is possible via any handle resolving service using the handle PID 21.T11973/MR@ followed by the PID to be resolved. The PID Metaresolver determines the structure of the information that can be expected from a resolution request and offers the possibility to request the landing page, metadata or resource of the digital object specified via the PID. The resolver can recognize the PID type entered for resolution and knows where to route the request to get the requested information.

RSAC development teams have made significant progress in aligning with the EOSC SIRS report and the 2021 SIRS report. The beta release, scheduled for the end of November 2023, represents a key milestone in integrating research software into the scholarly ecosystem. Across scholarly repositories, publishers, and aggregators, the project has achieved notable advancements, including integration with the Software Heritage archive and adherence to the CodeMeta standard. The RSAC developments teams released beta releases of the RSAC components, see:

- InvenioRDM - SWH APIs and connectors: https://sandbox.zenodo.org/
- DataVerse - SWH APIs and connectors: https://swh.dansdemo.nl/
- Dagstuhl - SWH APIs and connectors: https://faircore4eosc.dagstuhl.de/
- Episciences - SWH APIs and connectors: https://epijinfo.episciences.org/
- swMATH - SWH APIs and connectors: https://staging.swmath.org/wiki/Sagemath
- OpenAIRE - SWH APIs and connectors: https://beta.rdgraph.openaire.eu

Case Study Progress

Since the summer break, there has been good progress in the Case Studies work, as detailed below.

The Maths case study (CS) team met with the PIDgraph team to figure out how to integrate their service into the Maths use case. They have started developments to prepare ingestion of mathematics articles to be exposed in the DataCite services. A prototype has been achieved to display a part of the Mathematical Subject Classification (MSC) thanks to the new TaxonomyNode schema, with specific LaTeX code found in the titles of the MSC. The MediaWiki exposing the RSAC component has been enriched with more accurate schema mappings.

The Social Sciences and Humanities CS team has continued to work on the design for an internal API, part of the Digital Object Gateway. The internal API facilitates the integration with the two other services, the Switchboard and Virtual Collection Registry and already add support for some of the functionalities offered by the FC4E services. The aim for this effort is to already start with laying the groundwork for when the FC4E service beta versions are released. At the end of the reporting period work was started on the podcast interview scripts.

The Climate Change CS team prepared for the usage of the DTR by ingesting a first version of climate variable type definitions into the DTR and discussing open issues with the DTR developers. An initial skeleton of the provenance library supporting the climate case study was defined. A handover of all technical project work happened in November after our main CS developer left the project.

In the European Integration of National-level services CS, both the Research.fi and refactored CERIF mapping have been moving forward with weeklies of the MSCR component team to discuss implementation, in anticipation of the November release. The RAiD component usage was also discussed internally and the team plans to implement it for seed data within Research.fi for research projects. Members of the team participated in euroCRIS MM2023 meeting https://meeting.eurocris.org/ in Pamplona within FAIRCORE4EOSC session, where discussion with stakeholders continued in the CRIS community.

Based on the schedule of B2SHARE to MSCR integration, the EOSC service providers and RDM communities CS team has sketched initial plans on how first version of B2SHARE integration to MSCR could be used to ease management of metadata schemas for communities using B2SHARE. In the last week of November, the team participated in the EUDAT hackathon event to understand how DTR service affects metadata schema management in the B2SHARE context.

Finally, WP7 organised a 1-day in-person workshop in The Hague (5 October 2023) to facilitate the interaction between the CS teams and the components development teams, and to clarify expectations around the component beta release. Speed-dating sessions were organised in the afternoon to further discuss specific CS requirements handling. This meeting has been very beneficial to advance on both the component and CS developments.