

8.5 EPOS-ORFEUS (CC)

- [Ambition](#)
- [User stories](#)
- [Use cases](#)
- [Architecture & EOSC-hub technologies considered/assessed](#)
- [Requirements for EOSC-hub](#)
 - [Technical Requirements](#)
 - [Capacity Requirements](#)
- [Validation plan](#)

| | |
|-------------------|---|
| Short description | EPOS-ORFEUS |
| Type of community | Competence Center |
| Community contact | Luca Trani , Javier Quinteros |
| Meetings | |
| Supporters | |

Ambition

The CC drives collaboration between EOSC-hub and the ORFEUS-EIDA federation of EPOS. The CC collects and assesses the requirements of the solid-Earth science community, with a specific focus on Seismology, and addresses them by leveraging the EOSC-hub technical offerings. The CC delivers a software platform that facilitates access and exploitation of computational resources; it supports and fosters harmonisation of best practices for data management at ORFEUS-EIDA; and it enables the generation of seismological products customised on user requirements. By the end of the EOSC-hub project the CC aims to have a pre-production quality, modular software platform that could be deployed at (selected) data centres. However, the actual deployments will depend on agreements for service provisioning and operation.

User stories



Instruction

Requirements are based on a user story, which is an informal, natural language description of one or more features of a software system. User stories are often written from the perspective of an end user or user of a system. Depending on the community, user stories may be written by various stakeholders including clients, users, managers or development team members. They facilitate sensemaking and communication, that is, they help software teams organize their understanding of the system and its context. Please do not confuse user story with system requirements. A user story is an informal description of a feature; a requirement is a formal description of need (See section later).

User stories may follow one of several formats or templates. The most common would be:

"As a <role>, I want <capability> so that <receive benefit>"

"In order to <receive benefit> as a <role>, I want <goal/desire>"

"As <persona>, I want <what?> so that <why?>" where a persona is a fictional stakeholder (e.g. user). A persona may include a name, picture; characteristics, behaviours, attitudes, and a goal which the product should help them achieve.

Example:

"As provider of the Climate gateway I want to empower researchers from academia to interact with datasets stored in the Climate Catalogue, and bring their own applications to analyse this data on remote cloud servers offered via EGI."

| No. | User stories |
|-----|--|
| US1 | As a provider of an EIDA data centre I want to provide users with an authentication and authorisation service in order to enable them to securely access restricted and embargoed data. |
| US2 | As a seismological researcher I want to search for datasets offered by EIDA and stage them on the available cloud infrastructure offered by EOSC providers. |

| | |
|------------|---|
| US3 | As a seismological researcher I want to analyse my data in a Jupyter environment, pre-populated with my preferred libraries and with access to my pre-staged datasets. I want to store results in my personal workspace/storage area and eventually share them with my colleagues. |
| US4 | As an EIDA data manager I want to define my data management (DM) policies and share them with my colleagues at EIDA data centres. I want to enable them to understand, adjust and apply DM policies at their data centres. |

Use cases



Instruction

A use case is a list of actions or event steps typically defining the interactions between a role (known in the Unified Modeling Language as an actor) and a system to achieve a goal.

Include in this section any diagrams that could facilitate the understanding of the use cases and their relationships.

| Step | Description of action | Dependency on 3rd party services (EOSC-hub or other) |
|------------|---|--|
| UC1 | A researcher requests to access the services of the EPOS-ORFEUS CC. He is redirected to the CC Authentication Service (relying on B2ACCESS in the background) where he can log in at his home institution or create a local account if needed. He receives a token. Depending on his profile he might be authorised to use the services of the CC. Profiles include information about the groups he belongs to (e.g. read permission of particular restricted data). | B2ACCESS |
| UC2 | A researcher (authenticated and authorised) searches for datasets of her interest, by using an API. She selects one or more staging nodes from the available ones received, by interrogating the API. Finally she initiates the data movement by calling a dedicated method of the same API. | B2STAGE, B2SAFE, EIDA WFCatalog (with Dublin Core extension) |
| UC3 | A seismologist (authenticated and authorised) wants to perform an analysis on a datasets previously selected and staged. He logs in the Jupyter environment close to the staged datasets. He selects and launches a kernel containing his preferred seismological libraries. When the correspondent Jupiter notebook is up and running the datasets are available in a local directory and he can perform his analysis. He might choose to pause his work and save it for later. Finally he can download results on his PC, move them to his personal cloud storage folder or made them available on a local folder. | EGI Notebook, B2DROP |
| UC4 | A data centre acquires and stores waveform data by connecting to servers or devices. A network operator indicates data publicity policies to the data centre. At a next phase, a check of the expected data is performed as well as computation and ingestion of waveform data quality metrics (e.g. percentage availability). Meanwhile, manually data maintenance (e.g. gap filling) and replication is being implemented. Concerning replication, the data are transferred from the data archive to external resources by using B2SAFE. Finally, data requests via services are being traced and summed up to statistics regularly. | B2SAFE-DPM |
| UC5 | A seismologist wants to analyse data that is available (previously staged) at different distributed compute centres. After accessing one the available Jupyter environments, he writes and test his analysis code. When he is satisfied with the results he might decide to run such an analysis code on a selected number of compute centres. | EGI Notebook |
| UC6 | Despite that EPOS works with and provide access to open data, some seismic networks related to temporary experiments need to keep data embargoed for a short period of time. The management of the Access Control List (ACL) is in charge of the Network Operator or project PI. Originally, the PI contacted the data centre to include/exclude users from the ACL. EPOS would like to give permissions to PIs, so that they can manage the ACL by themselves using B2ACCESS groups from the B2ACCESS GUI. This way, data centres can configure their systems to grant access to groups (not individuals) and the PIs manage group members in a decentralised way. | B2ACCESS |



DaSta Framework Dockers schema



Sequence 4 endpoint: `/api/airods/stage` :: make a request for data and stage the results remotely

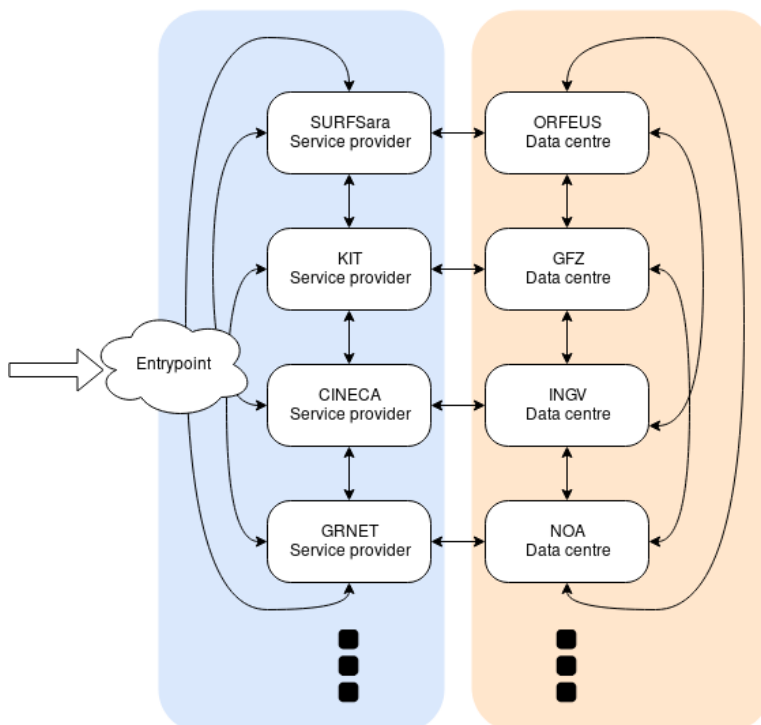


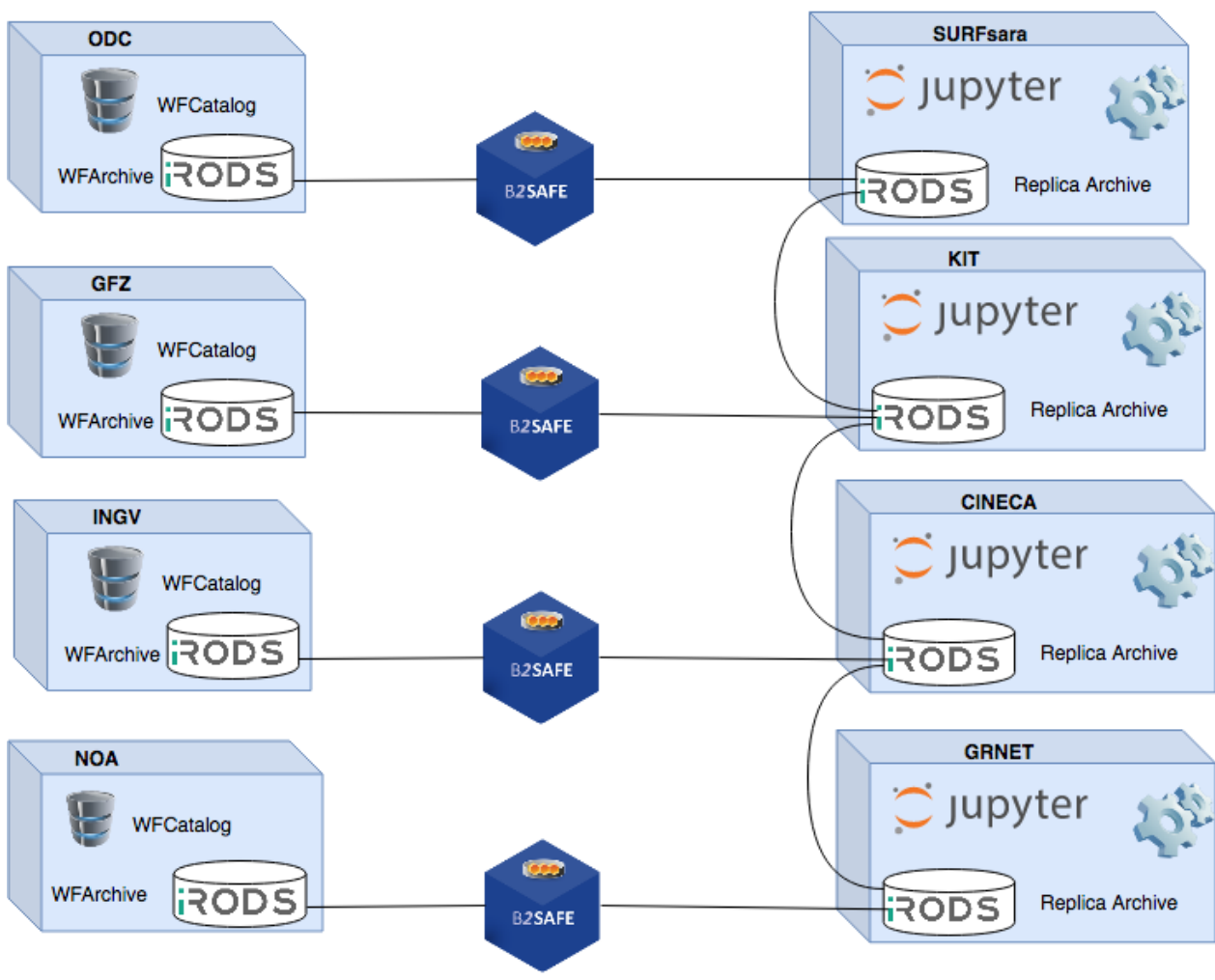
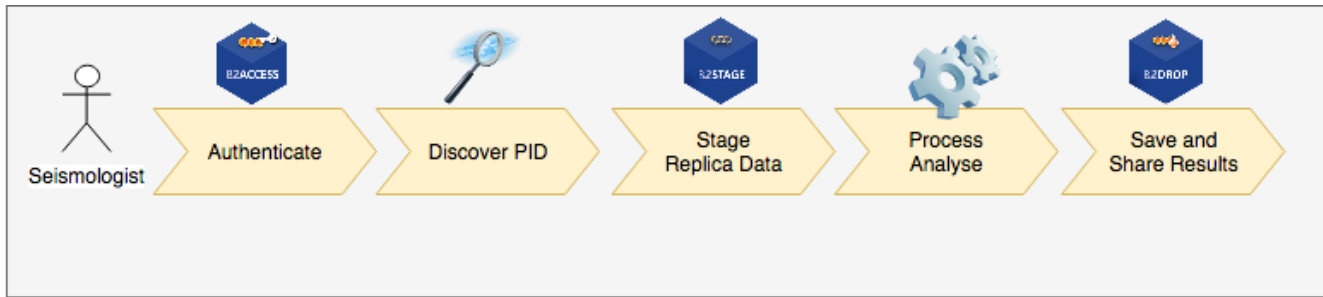
Architecture & EOSC-hub technologies considered/assessed

Data centre view



Federation overview







Requirements for EOSC-hub

Technical Requirements



Instruction

- Requirement number: Use numbers RQ1, RQ2, RQ3, ...
- Requirement title: Use a short but descriptive title. Use the same title in the Jira ticket 'Summary' field
- Link to requirement JIRA ticket: Open a ticket in <this JIRA queue <https://jira.eosc-hub.eu/projects/EOSCWP10/issues/EOSCWP10-4?filter=allopenissues>> (click on 'CREATE' button in the middle-top of JIRA)
- Source use case: Refer back to the use cases above (UC1, 2, ...)

| Requirement number | Requirement title | Link to Requirement JIRA ticket | Source Use Case |
|--------------------|-------------------|---------------------------------|-----------------|
|--------------------|-------------------|---------------------------------|-----------------|

| | | | |
|---------|--|--|----------|
| Example | EOSC-hub to provide an FTS data transfer service | EOSCWP10-21 - ... <input type="text"/> | UC1 |
| RQ1 | iRODS instance accessible from the Jupyter environment and federated with local B2SAFE/iRODS instances EOSCWP10-66 - ... <input type="text"/> | | UC2, UC3 |
| RQ2 | Customisable and permanent kernels in Jupyter (EGI Notebook) EOSCWP10-65 - ... <input type="text"/> | | UC3 |
| RQ3 | Personal data folder with staged data available for mounting in the Jupyter notebook EOSCWP10-64 - ... <input type="text"/> | | UC3 |
| RQ4 | Operating SeedLink slarchive and rsync for data acquisition, while ArcLink and FDSNWS for data exposal. WFCatalog for quality metrics collection and distribution, as well as B2SAFE for data replication and Webreqlog for statistics sum-ups | | UC4 |
| RQ5 | Execution of distributed Jupyter notebooks | | UC5 |
| RQ6 | A centralised catalogue of policies. It should collect descriptions of data management policies and make them available (via API and metadata) | | UC4 |

Capacity Requirements

| EOSC-hub services | Amount of requested resources | Time period |
|-----------------------|--|---|
| EGI Notebooks service | Ideally we need this service deployed at the providers hosting our replicated archives i.e. SURFsara, KIT, CINECA and GRNET. It should be coupled with a scratch space where data from the archive can be staged to be processed and results temporarily stored. Size ~1TB | Duration of the project |
| B2SAFE | It is already in place and running at the 4 nodes. The requirements vary from site to site and there are local agreements. | It should be sustained after the project ends |
| B2ACCESS | | It should be sustained after the project ends |
| | | |

Validation plan

The AAI based on B2ACCESS was validated in a test case targeting the AlpArray community. Future validation will include the whole EIDA user community.

The integration of the staging and pocessing services will be validated via pilots targeting selected EIDA users. The extent of the pilots depends quite substantially on the resources available.

The execution and management of the policies will be validated at the 4 data centres: KNMI, GFZ, NOA and INGV. Results will be presented to the remaining EIDA partners