Organisation: EUDAT

Project Charter

EUDAT: an e-Infrastructure node to support Research Data Management

1. Project Summary

EUDAT Collaborative Data Infrastructure is a pan-European e-infrastructure with the vision to enable data sharing across domains and borders. EUDAT is a federation and collaboration of data and computing centres, that includes national computer centres and thematic organisations. Together they have been working to provide generic and community specific data and compute services for pan-European research. EUDAT is naturally positioned to be part of the EOSC Federation and provide a unique feature, transnational access to resources and essential data management services, for the other nodes of the federation and European wide research communities.

The contribution of EUDAT to the EOSC node federation is to provide a collaborative space for research projects and research communities enabling them cross border collaboration while supporting the creation of **high-value FAIR-by-design datasets**. This workspace will be composed of the existing and continuously developed EUDAT services suite which is federated through a common proxy (B2ACCESS), compliant with AARC blueprint specifications and compatible with the EOSC Federation AAI. EUDAT operates a Monitoring service that is fully compatible with EOSC EU Node Monitoring, ensuring reliable tracking of service performance and availability. The workspace will be supported by a collection of resources (data storage, data archiving and connection to compute), enhancing the service capability of the wider EOSC federation. The EUDAT workspace integrates and allows onboarding of complementary services operated by EUDAT members, thereby contributing to the FAIRification of datasets and construction of scientific workflows to enable the acceleration of scientific discovery in Europe.

To achieve this, we will rely on the strong expertise within the EUDAT in offering data management services to support research data life cycle and FAIR data and our strong involvement in defining interoperability frameworks and efficient and scalable technical infrastructure. Finally, all EUDAT services are operated and maintained in accordance with the FitSM framework, adhering to established procedures and best practices for effective IT service management.

The EUDAT federation covers 17 countries in Europe and supports scientific communities at different scales, and from a wide range of scientific domains since its inception. The activities of EUDAT will be focused toward extending our outreach to new scientific communities within/and outside of our geographical coverage. We will provide data management services essential for data provisioning and hosting, we will contribute to the development of the EOSC Interoperability Framework to enable the onboarding/integration of existing services provided by other nodes.

2. Value Proposition

As of now, the EOSC Federation is composed of a “single” node, the EOSC EU Node, which provides federating capabilities and generic services with a predefined amount of accessible resources that are allocated to researchers based on the credit system. The credit system allows access to a limited amount of resources which might be insufficient for the EOSC EU Node users. In addition, the EOSC EU Node provides only a few data services besides the sync and share service.

At the same time there is a need to have access to a larger amount of specific resources e.g. raw storage capacity, compute, extended storage for the sync and share service, to name a few. Also required is the ability to publish and share high quality FAIR datasets that will be preserved for a long period of time and are identified by globally unique and resolvable identifiers (e.g. handle, DOI,…). Furthermore, there is a clear need for easy sharing of data within the federation to support cross-border and cross-domain collaborations.

The current access provisioning to the services offered by the EOSC EU node is supported through a credit system that is linked only to the resources provisioned in the context of the EOSC tender. This credit system is local to the EOSC EU node and is not supported by a business model which would sustain access beyond the resources of the EOSC EU node and beyond the lifetime of the tender. For researchers, scientific communities and institutions, it is necessary to establish a business model that would sustain the access to a transnational pool of resources which will support collaboration across Europe. Business models with a long-term lifetime as compared to project and/or tender lifetimes are especially required for data services as high quality FAIR data needs to be stored and maintained even when not actively being used. Defining such a business model is currently a real challenge to be tackled in the build-up phase of the EOSC Federation.

- **Main Goal:**
  The main goal of this project is to provide a FAIR enabling data management collaborative workspace to complement the services provided by the EOSC EU node and the other nodes of the federation (e.g. data publication, data curation, data preservation, data replication,...) as well as infrastructure resources (storage, compute,...) to support communities and researchers with needs going beyond the current capacities of the EOSC EU Node. This platform should support seamless migration of data from Nodes to the EUDAT provided storage and vice versa. The provisioning of the EUDAT capacities should be backed by a sound business model which will support the long term sustainability of the platform and the node, provide guarantees for long term data preservation as well as support the maintenance and improvements of the services suite.

- **Needs addressed:**
  - Extension beyond the capacities of the EOSC EU Node
  - FAIR enabling tools: data repositories, data catalogue,...
  - Transnational pool of resources supported by a business model
  - Long term preservation of the high quality FAIR datasets

- **Key Benefits:**
  This project is a unique opportunity to:
  - strengthen data sharing across borders and disciplines
  - establish a business model to provide a transnational pool of resources independent from European Commission funding
  - interconnect with other nodes to provide an enriched data management and data science experience
  - investigate solutions for long term preservation of high quality FAIR data

- **Who Benefits:**
  - Research infrastructures with needs for scalability. The offer would enable them to focus their resources on the development of research specific services instead of infrastructure needs and maintenance.
  - Institutions: provide research data management services and associated infrastructure to reduce the burden of the management of the infrastructure.
  - Individual researchers and scientific communities
  - EU-funded and/or transnational research projects. The offer would provide an integrated environment for sharing and working with the data in the project and possibly offer solutions to manage the data after the lifetime of the project.

3. Use Case(s)

In the table below, we are presenting a short description of the use cases without providing information regarding what is in scope and out of scope as this will require to have more in-depth discussions with the potential involved nodes.

Use Case ID	*Use Case Description*	*Federation Contributions & Value to Users*
1	***Federating File Sync & Share Services*** EUDAT offers B2DROP, SURF offers Research Drive and the EOSC EU Node offers an Enterprise File Sync & Share service (EFFS). This use case is focussing on building up a federating capability on the EFFS services across the nodes. In addition, we will link this use-case with the scientific communities through the collaboration with the thematic multi-node use-case proposed by Data Terra. In this context, B2DROP will be used as a collaborative working space to share datasets from different EFSS used by the communities.	To support this use case, the Nodes offering an EFSS service enable the OpenCloudMesh API (https://github.com/cs3org/OCM-API) on their EFSS service and enable federated sharing. Through this use case, users can easily share data across the EOSC Federation from their own File Sync & Share service without the need to have access and/or an account on EFSS service on which the data is residing. This supports the collaboration between researchers across nodes within the EOSC Federation. Resources: B2DROP (https://eudat.eu/service-catalogue/b2drop) Access policies: The resources are accessible to any European researchers after review of their account request. The default free offer provides 20 GB of storage space and collaborative editing functionalities. Additional storage capabilities can be offered on demand based on the current B2DROP freemium business model. Multi-node interactions: As an infrastructure demonstrator, we will interact with service providers offering EFSS services (EOSC EU Node, SURF,...) to establish the federation of EFSS services As a scientific demonstrator, we will work with the thematic multi node use-case proposed by Data Terra and our communities to highlight the scientific value of this service.
2	**Enabling access to File Sync & Share Services across cloud computing resources** Multiple nodes provide access to cloud computing resources as virtual machines, containers and Jupyter notebooks. While in use case 1 a federation of File Sync & Share Services has been established, this use case	Most of the EFSS services support WebDav endpoints to allow remote mounts of user shares in computing environments. To enable this use case, nodes providing cloud computing resources as virtual machines, containers and jupyter notebooks need to enable access to remote mounts of EFSS endpoints in cloud computing environments.

	is focussing on enabling the access from cloud computing services to the EFSS services across the Federation. For example the EUDAT B2DROP service can be mounted within EOSC EU Node or SURF Research Cloud virtual machines, containers and/or jupyter notebooks. To provide a scientific demonstrator of this functionality, we will work with the thematic multi node use-case proposed by Data Terra to connect B2DROP to the Galaxy workflow engine to upload data for computation and deposit computing results in B2DROP	This allows easy access to data residing in a EFSS service from one node to computing resources offered through another node. The added value is that users do not need to copy first the data from one node to another to make the data available for computing. Resources: B2DROP (https://eudat.eu/service-catalogue/b2drop) Access policies: The resources are accessible to any European researchers after review of their account request. The default free offer provides 20 GB of storage space and collaborative editing functionalities. Additional storage capabilities can be offered on demand based on the current B2DROP freemium business model. Multi-node interactions: As an infrastructure demonstrator, we will interact with service providers offering Jupyter notebook services (EOSC EU Node, SURF, NFDI,...) to establish the connection between B2DROP and the different Jupyter notebook services. As a scientific demonstrator, we will work with the thematic multi node use-case proposed by Data Terra and our communities to highlight the scientific value of this connection with the computing services, extending our initial plan to connect with Galaxy, a widely used workflow engine.
3	**Support publishing within the EOSC Federation** One of the main aims of the EOSC is enabling the Web of FAIR data. The Web of FAIR data can only be achieved by allowing users to make FAIR research data and outputs available. Multiple nodes enable access to data repositories through which researchers can publish their research outputs. In this use case, we are focussing on allowing researchers to publish data from EFSS services to the user data repository of choice within the federation.	While one of the main aims of EOSC is the Web of FAIR data, at this moment no data repository service is being offered within the EOSC Federation, the EOSC EU node does not offer a publication and/or data repository service. Through this use case, users within the federation can make research outputs available in a FAIR way, for example by allowing researchers to publish research outputs in the EUDAT B2SHARE and/or Zenodo service offered by CERN. To enable this use case an EFSS app needs to be developed to allow the publication of data into a repository service (e.g. B2SHARE, Zenodo). Nodes offering an EFSS service need to enable this app within their EFSS service. Resources: B2DROP, B2SHARE (https://eudat.eu/service-catalogue/b2share-0) B2HANDLE (https://eudat.eu/service-catalogue/b2handle)/DOI

		(https://eudat.eu/service-catalogue/datacite) services Access policies: The service is accessible to any European researcher to deposit and document an unlimited number of datasets which will be identified with a handle and a DOI provided by the B2HANDLE/DOI service for free. Multi-node interactions: As an infrastructure demonstrator, we will interact with service providers offering EFSS services (SURF, EOSC EU Node, ...) to develop an application to publish datasets from the different EFSS service to different data repository services (B2SHARE, Zenodo, Dataverse,...). As a scientific demonstrator, we will work with the thematic multi node use-case proposed by Data Terra and our communities to highlight the scientific value of this connection with data repositories, supporting data publication.

4. External Dependencies & Key Risks

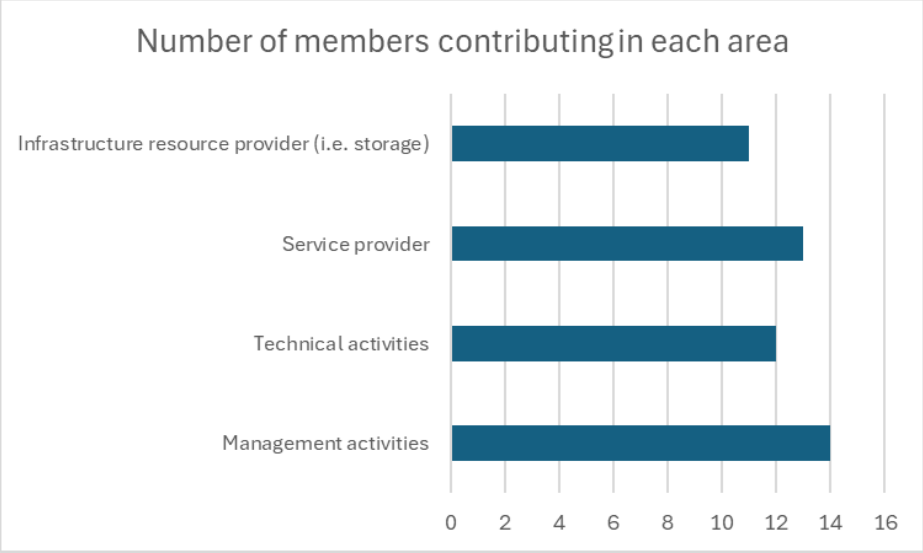
External Dependencies & Risks	Actions/mitigation measures	Type(s)
The basic components for building up the node platform are available as presented during the 2nd dialogue meeting. With these components EUDAT is able to implement the use cases. However, how these components are being offered to the federation depends on the requirements of the policies and procedures of the EOSC Federation. Depending on policies, procedure, requirements and required efforts needed to be implemented in the node platform potential delays can occur.	The creation of the node platform and the integration of the separate components within the node platform is one of the activities planned during the build up phase. An assessment of the requirements will be made at the start of the project. If the requirements change during the build up phase, the efforts and plans need to be revised leading to a less ambitious MVP at the end of the project.	External dependency/Risk
Integration with EOSC federated services. The EOSC Interoperability framework is still not completely defined and the integration may introduce complexities or service availability risks.	Follow the evolution of the interoperability framework and proactively provide feedback.	External dependency/Risk

Strong dependency on the other nodes' resources and capabilities for the multi-node use-cases	Investigate which other Nodes offer similar services and get in contact with those node operators and discuss/demonstrate mandatory steps to deploy the necessary parts and demonstrate the benefits for the users.	External Dependency (for Use case 1 and Use case 2)
Aligning EUDAT policies and procedures with the EOSC Federation policies and procedures and finding consensus among the EUDAT members will take time as it will require decisions on structural changes to be validated by the EUDAT Council.	The majority of the EUDAT Council members are active contributors to the EUDAT node build up phase; EUDAT council will be continuously kept up to date on the EUDAT node activities to reach a wider consensus in advance to the EUDAT council meetings.	Risk

5. Contributions

The EUDAT node development will be supported by 15 EUDAT organizations with the contributions summarised in the graph below. Each of the members will contribute to one or more areas, and all activities will be supported by different partners so as to rely on a wide range of expertises.

EUDAT ltd will be coordinating the activities among the partners and monitoring the progress with respect to the project plan.



- **Deliverables**

ID	Deliverable Name	Deliverable Description	Expected Deadline	Status

1	EUDAT node architecture	Definition of the architecture of the EUDAT Node	EOSC Symposium 2025	In progress
2	Description of the scientific use-cases and demonstrators	one or two use cases agreed and defined with other nodes	May 2025	DONE
3	EUDAT node - Mock-up	Interactive mock-up of the node User Interface and functionalities	EOSC Symposium 2025	In progress
4	EUDAT node - Proof of Concept	First demonstrable version of the EUDAT Node	May 2026	Not started
5	Scientific demonstrators implemented	Integration with scientific workflow demonstrators (1-2 multi node use cases)	EOSC Symposium 2025	In progress
6	EUDAT node - production roll-out	Production roll-out of the EUDAT Node	September 2026	Not started

- **Resources Provided**

Personnel

The EUDAT members have committed to provide a total of 8 FTEs for the development of the EUDAT node which will be internally organised in 2 working groups: one to address management and organisations issues (refine operational processes, define governance of the EUDAT federation, define business model, …) and one to discuss and coordinate the technical development (define the architecture, define service integrations and development,...).
These two groups will be coordinated by EUDAT ltd via the EUDAT secretariat and interact with the working groups of the EOSC Federation.

Working group	Committed Effort	Expertise
EUDAT Management WG	3 FTEs	• Infrastructure and organisational management • Service delivery processes • Business models • Legal/contractual aspects
EUDAT Technical WG	5 FTEs	• AAI • PIDs • Security • Metadata • Data discovery • FitSM processes

Infrastructure

The service offer of the EUDAT Node will be completely defined during the build-up phase but already some services were identified during the preparation.
The offer will include the B2services suite from EUDAT (B2FIND, B2SHARE, B2DROP, B2HANDLE, B2SAFE and B2ACCESS) which will be complemented by services provided by EUDAT members like Gitlab, Chat, Monitoring, Compliance Assessment Toolkit and by additional storage services (i.e. Data Archiving or Cloud Storage).
The service provisioning will be associated with relevant resources, in particular storage that can potentially be in the order of Petabytes.

6. Timing and Milestones

- **Start Date:** 01-04-2025
- **Expected Duration:** 18 months
- **Key Milestones:**

ID	Milestone Description	Target Delivery Date
1	Definition of the Minimum Viable EUDAT Node and core capabilities	01/10/25
2	Selection of services to offer for the EUDAT node	October 2025
3	Onboarding policies for external services	March 2026
4	Plan to connect the EUDAT Node to the EOSC Federation federating capabilities	September 2025
5	Legal framework between the EUDAT members refined in the context of the EOSC Federation (SDA, OLAs, SLAs,...)	February 2026
6	EUDAT Node operational guidelines and documentation	September 2026

ID	Use-case 1 - Federating File Sync & Share Services	Target Delivery Date
1	Description of the scientific use-case(s) and demonstrator(s)	May 2025
2	Enabling and testing the EFSS federation (infrastructure “use-case”)	October 2025
3	Enable and test the scientific demonstrator(s).	EOSC Symposium 2025

ID	**Use-case 2 - Enabling access to File Sync & Share Services across cloud computing resources**	**Target Delivery Date**
1	Description of the scientific use-case(s) and demonstrator(s)	May 2025
2	Enabling and testing the connection with Galaxy workflow engine (infrastructure "use-case")	October 2025
3	Enable and test the scientific demonstrator(s).	EOSC Symposium 2025

ID	**Use-case 3 - Support publishing within the EOSC Federation**	**Target Delivery Date**
1	Description of the scientific use-case(s) and demonstrator(s)	May 2025
2	Developing the EFSS application to connect and upload datasets to B2SHARE	October 2025
3	Developing the EFSS application to connect and upload datasets to different repositories (Zenodo, Dataverse,...)	June 2026
4	Enable and test the scientific demonstrator(s).	EOSC Symposium 2025

7. Contact & Submission

Role	Name	Email
Coordinator	*Yann Le Franc*	*ylefranc@esciencefactory.com*
Operation Manager	*Sander Apweiler*	*sa.apweiler@fz-juelich.de*
Security Officer		
Data Privacy Officer		*privacy@csc.fi*