

WHITE PAPER

Advancing Essential Services to Complement EOSC

Authors:

Bernard, L., Diepenbroek, M., Jander, M., Koepler, O., Lehmkuhler, D., Miller, B., Rettberg, N., Ritter, X., Schäfer-Neth, C., Schimmler, S., Vishen, N.

Date: 27 January 2025

Funded by **DFG** as part of **NFDI**. Grant Numbers:
521453681, 521460392, 521462155, 521463400, 521466146, 521471126, 521473512, 521474032, 521475185,
521476232

This short paper is aimed at policymakers for the NFDI (German National Research Data Infrastructure) and EOSC (European Open Science Cloud) and examines the Base4NFDI services in the context of the EOSC EU Node services. It is intended as the start of a dialogue on the Base4NFDI contribution to the federation.

NFDI and Base4NFDI

The NFDI (German National Research Data Infrastructure) aims to make valuable data from science and research systematically accessible, networked, and usable in a sustainable and qualitative manner for the entire German science system and beyond. It is comprised of 26 NFDI disciplinary consortia¹. Launched in 2023, Base4NFDI is a DFG-funded initiative which establishes a framework to provide cross-disciplinary services to further establish FAIR practices, supporting the transition to open science both nationally and internationally. Base4NFDI mirrors NFDI's bottom-up approach in its domain-agnostic service development by facilitating the community to identify their own cross-cutting needs within the sections of the NFDI association and to realise and to service them by providing interdisciplinary, largely generic services. The Base4NFDI process encourages services to be built on existing solutions, and to leverage the many well-established tools and components in various partner institutions, thus bringing together a diversity of national partners with different starting positions and institutional resources. The outlook is to develop a framework that works to harmonise and support service development which are to be used by the NFDI community and beyond. EOSC (European Open Science Cloud) was initiated in 2015 with the aim to establish a *federation* of connected infrastructures and services for open science to support FAIR data implementation for European researchers². Both the EOSC EU Node service offering and Base4NFDI work towards similar goals to facilitate open science and can complement each other in their service offering.

Basic Services

As of January 2025 Base4NFDI supports the development of eight basic services addressing different topics and the varying needs of the consortia³. Further services will follow. Each service addresses a critical need expressed by many communities in the German scientific RDM landscape and is built and operated directly and interdisciplinarily from within the NFDI research communities. A service in the Base4NFDI context is considered to be a technical-organisational solution, which is usually based on one or more existing services, and develops new features and technical applications. The service should be integrated within the existing landscape and accompanied with professional training and support.

¹ <https://www.nfdi.de/konsortien/>

² <https://eosc.eu/eosc-about>

³ [Overview submission rounds - Base4NFDI](#)

Each service is at a different level of maturity: Base4NFDI basic services move through three phases of development, starting in Initialisation phase (TRL 5-6) through Integration phase (TRL 7-8) to Ramp-Up phase, arriving as a fully developed ready-to-use service (TRL 9). Throughout this process, interoperability, scalability and training are always taken into account, further developed during each phase. While a thorough user requirements analysis is performed during the Initialisation phase, the focus shifts to building a sustainable governance model for each basic service during the Integration and Ramp-Up phases. During all phases, however, especially from Integration on, the services are modeled closely in line with the [EOSC Interoperability Framework](#) and [EOSC Rules of Participation](#)⁴, as described in our [Base4NFDI Policy Paper: "Base4NFDI Services and EOSC: Guidance for Interoperability"](#)⁵.

Alongside individual services, Base4NFDI develops a service catalogue with information on and access to the services. This will assure interoperability with EOSC in that the catalogue can be provided to EOSC as a community platform as has been described in the [EOSC Future Interoperability Guidelines](#)⁶. A team of [Service Stewards](#) work to connect the basic services to the consortia via engagement activities.

The following overview of basic services currently under development groups them into two categories relevant for a potential EOSC node: *core services* integral to the operation of the node, and *node resources*, which can be further divided into those accessible by node users only, or the whole EOSC community.

Table 1: Summary of Base4NFDI basic services.

Basic service	Summary	Ready for larger scale integration ⁷	Node core service (CS) or node resource (NR) ⁸
IAM4NFDI	NFDI-AAI based on Community AAI, enabling IAM for NFDI consortia/beyond	2025	CS
PID4NFDI	Persistent Identifier Coordination Hub	2026	NR
TS4NFDI	Development, curation, mapping of terminologies (semantic artefacts)	2025	NR

⁴ European Commission: Directorate-General for Research and Innovation, EOSC rules of participation, Publications Office, 2021, <https://data.europa.eu/doi/10.2777/30541>

⁵ Bernard, L., Jander, M., Manske, A., Miller, B., Rettberg, N., Ritter, X., Reißler-Pipka, N., Schäfer-Neth, C., Stein, R., & Zänkert, S. (2024). Base4NFDI Policy Paper: "Base4NFDI Services and EOSC: Guidance for Interoperability", Version 1 (Version 1). Zenodo. <https://doi.org/10.5281/zenodo.13946300>

⁶ Scardaci, D., Dietrich, M., & Manghi, P. (2024). EOSC Federation: Architecture and Federating Capabilities. Zenodo. <https://doi.org/10.5281/zenodo.13939396>

⁷ The service is likely developed to a stable and running prototype for the NFDI community, so that first steps towards a technical integration into EOSC can be considered.

⁸ This paper has taken these definitions from the Scardaci, D., Dietrich, M., & Manghi, P. (2024) paper: core node services enable the operation of the node and node resources are services offered to the community.

Basic service	Summary	Ready for larger scale integration ⁷	Node core service (CS) or node resource (NR) ⁸
KGI4NFDI	Central Knowledge Graph Registry	2027	NR
Jupyter4NFDI	Central JupyterHub	2026	NR
DMP4NFDI	Data and Software Management Plans	2026	NR
RDMT4NFDI	Training Material / Modules for RDM	Tbc	NR/CS
NFDI.software	Central Research Software Marketplace	Tbc	CS

IAM4NFDI – What it does: NFDI-AAI connects and expands Identity and Access Management (IAM) systems to enable researchers to use their home institution’s credentials to seamlessly access tools, data, and services within the NFDI, and manage virtual research organisations. **How:** By bridging eduGAIN and Community Authentication and Authorisation Infrastructures (AAIs), the NFDI-AAI is interoperable and will be federated with AAIs of other entities such as EOSC, NHR, and EGI and offer fine-grained authorisation to NFDI services. **What phase is it in:** Integration phase, runs until January 2026. All three phases are potentially completed by 2027. **Role in an EOSC node:** IAM4NFDI can be an integral EOSC *core service* of an EOSC node.

PID4NFDI – What it does: Integrates existing infrastructures for Persistent Identifiers (PIDs) to improve metadata quality and mapping, support the adoption and integration of PIDs in the NFDI consortia, and help scientists choose the PID system appropriate for their work. This addresses and fosters all aspects of FAIR and sustainable research and data management and is carried out aligned to the work of the German PID Network⁹. **How:** A central PID coordination hub for NFDI is to be established where these efforts come together. **What phase is it in:** Integration phase until December 2026. All three phases are potentially completed by 2028. **Role in an EOSC node:** PID4NFDI can be offered as an EOSC *node resource*, since its focus lies on the facilitation of good PID practices and harmonisation and less in hosting a technical infrastructure to generate and manage PIDs.

TS4NFDI – What it does: TS4NFDI provides harmonised and centralised access to the Terminology Services (Semantic Artefact Catalogues) of NFDI communities. It facilitates the provision, curation, development, harmonisation, and mapping of semantic artifacts (SAs), enabling annotation and semantic enrichment of data and metadata. **How:** TS4NFDI supports integration with existing services through an API Gateway¹⁰ and enhances applications with

⁹ <https://www.pid-network.de/en/>

¹⁰ <https://github.com/ts4nfdi/api-gateway>

widgets from the Terminology Service Suite¹¹. **What phase is it in:** Integration phase until December 2026. All three phases are potentially completed by the end of 2028. **Role in EOSC node:** TS4NFDI could become an EOSC *node resource*, to help curate and harmonise semantic artifacts among federated EOSC services, and is already involved in EOSC integration¹².

Jupyter4NFDI – What it does: Provides a web-based, interactive, and multi-user environment where users will find pre-defined or design their own environments to process data. **How:** Builds a central JupyterHub system for Jupyter notebooks in various programming languages such as Python, R, etc., providing core tools for data analysis, visualisation, and machine learning. Supports remote data access but also offers data storage on the spot. **What phase is it in:** Initialisation phase, which runs until September 2025. All three phases are potentially completed by the end of 2029. **Role in EOSC node:** It is built on federated heterogeneous infrastructures throughout Germany and could be federated in EOSC and linked to other HPC offers, in Germany only, according to the current funding regulations.

DMP4NFDI – What it does: Provides Research Data Management Organiser (RDMO)¹³ as a common tool for machine-actionable data and software management plans. **How:** Customisable RDMO clients to provide community-specific DMP/SMP services and templates based on a standardised and interoperable vocabulary, and integrated with other services via API and plugins for easier creation of DMPs/SMPs. **What phase is it in:** Initialisation phase, which runs until May 2025. All three phases are potentially completed by the end of 2029. **Role in EOSC node:** DMP4NFDI will be offered as an EOSC *node resource*, initially probably specific to the German research landscape.

KGI4NFDI – What it does: Builds a central and interoperable Knowledge Graph (KG) registry and consultancy. **How:** It provides access to metadata of domain-specific KGs across NFDI projects via a SPARQL endpoint. The guidelines¹⁴ for creating and hosting knowledge graphs and consultancy services support research communities to create their own customised KGs. **What phase is it in:** Initialisation phase, which runs until May 2025. All three phases are potentially completed by 2029. **Role in EOSC node:** KGI4NFDI can become an EOSC *node resource*, with the prospect of an EOSC-wide federation.

NFDI.software – What it does: Integrates existing software repositories, e.g. ELIXIR and Physics.tools, and builds a central marketplace for NFDI research software. **How:** This hub will integrate different scientific communities based on standardised software metadata to improve accessibility, compatibility, and coordinated software development. **What phase is it in:** Initialisation phase, which runs until November 2025. All three phases are potentially completed by 2029. **Role in EOSC node:** It can have the role of a *core service* in an EOSC node.

¹¹ <https://ts4nfdi.github.io/terminology-service-suite/comp/latest/>

¹² <https://fair-impact.eu/implementing-shared-api-for-semantic-catalogues>

¹³ <https://rdmorganiser.github.io/>

¹⁴ <https://kgi4nfdi.github.io/Guidelines/>

RDMTraining4NFDI – What it does: Develops a modular collection of RDM training materials, along with documented, proven training formats and methods. **How:** The training encompasses data, software, and machine-learning models, and will enable consortia to create their own community-specific training adaptations. **What phase is it in:** Initialisation phase, which runs until February 2026. All three phases are potentially completed by 2030. **Role in EOSC node:** RDMTraining4NFDI could be an EOSC *core service* for the training provided for this node, as well as a possible EOSC *node resource* when it comes to building an efficient and harmonized, possibly federated training infrastructure.

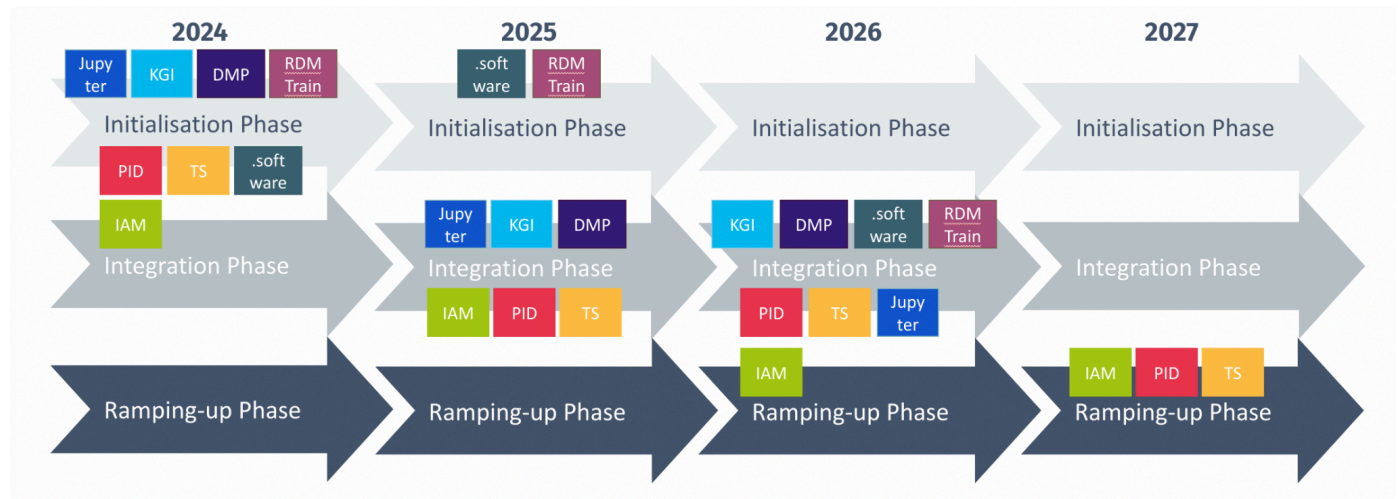


Figure 1: Projected development and timeline of currently funded Base4NFDI basic services until 2027 (status January 2025).

Base4NFDI services complement EOSC services

NFDI’s basic services can complement EOSC EU Node Services, others have the potential for future EOSC developments to provide some added value and to bring synergies.

Table 2: Where Base4NFDI services can complement EOSC services.

Service Name	Description (source: https://open-science-cloud.ec.europa.eu/)	Base4NFDI service
File Sync and Share	Storing, synchronising, and securely sharing data across teams and systems	<i>Under future consideration</i>
Virtual Machines	Virtual environments on-demand computing	<i>Under future</i>

		<i>consideration</i>
Interactive Notebooks	Dynamic, interactive platforms integrating live code, documentation, and visualisations	Jupyter4NFDI
Cloud Container Platform	A scalable and flexible system for deploying containerised applications	<i>Under future consideration</i>
Large File Transfer	A high-performance service for securely transferring datasets	<i>Under future consideration</i>
Bulk Data Transfer	An optimised infrastructure for moving massive datasets	<i>Under future consideration</i>
<u>Resource Hub</u>	A centralised platform that provides services and research outputs	Base services harvested

Takeaways

1) Services grounded in community needs: NFDI develops resources and infrastructures tailored to the research data management and research data analytic demands of disciplines. This also holds on the level of basic services and the service providers are distributed across Germany. Science has proven to adapt new workflows in a piecemeal fashion and a bottom-up approach **promises a more sustainable adoption of services**. All services are evaluated to assess quality, scalability, as well as **community buy-in and acceptance**. It should be noted that in the current Base4NFDI phase, the funding scheme does not allow to provide resources for running operational services: The aim is to get the services off the ground and provide resources for development and consortia integration.

2) Base4NFDI complements EOSC EU Node Services: Base4NFDI services have the potential to *complement* the EOSC EU Node services and some could be considered as future resources for a national EOSC Node. Until now **no services are duplicated**.

3) On-demand integration: We propose to allow for shallow integration of services into the EOSC early on, to allow for services to be integrated to the degree that usage patterns suggest it is prudent to pursue European interoperability. It might well be more resource-efficient to pursue interoperability **only to the extent that service adoption requires it**. The Base4NFDI framework development will ensure that EOSC interoperability is integrated from the start to ensure a pathway to their federation beyond Germany. Working together early on to **identify synergies** and added value aspects during development is the next step.