

GreenDIGIT

Greener Future Digital Research Infrastructures

Deliverable 5.3: Design and Operation of a Federated Data Management Infrastructure for Open Science Workflows

GRANT AGREEMENT NUMBER: 101131207

Lead Beneficiary:	CNR
Type of Deliverable:	Demonstrator
Dissemination Level:	Public
Submission Date:	31.08.2025
Version:	1.0



This project has received funding from the European Union's HE [research](#) and innovation programme under the grant agreement No. 101131207

Disclaimer: "Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them."



Versioning and Contribution History

Version	Description	Contributions
0.1	Deliverable Structure and Content	Massimiliano Assante, Luca Frosini, Francesco Mangiacrapa, Elisa Molinaro, Alfredo Oliviero, Giancarlo Panichi
0.2	Content updated, user access and metadata structure added	Massimiliano Assante, Luca Frosini, Francesco Mangiacrapa, Elisa Molinaro, Alfredo Oliviero, Giancarlo Panichi
1.0	Formatting for submission	Naomi van der Most, Ineke Brouwer

Authors

Author	Partner
Massimiliano Assante	CNR
Luca Frosini	CNR
Francesco Mangiacrapa	CNR
Elisa Molinaro	CNR
Alfredo Oliviero	CNR
Giancarlo Panichi	CNR

Reviewers

Name	Organisation
Ciprian Ciulpan	EBRAINS



Disclaimer

The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the European Commission. The Commission does not guarantee the accuracy of the data included in this study. Neither the Commission nor any person acting on the Commission's behalf may be held responsible for the use, which may be made of the information contained therein.



Executive Summary

This short report accompanies the deliverable “Design and Operation of a Federated Data Management Infrastructure for Open Science Workflows”, classified as a “DEM” (Demonstrator). It provides a concise overview of the design and deployment of the key software components constituting the Federated Data Management Infrastructure (FDMI), developed to meet the current operational needs of the GreenDIGIT community in support of open science workflows.



Table of Contents

1	Introduction and overview	8
2	FDMI Architecture, Access Mechanisms, and User Interfaces.....	9
2.1	Single Sign-On (SSO) and Federated Identity and Access Management Service (IAM)	9
2.2	FDMI Access via APIs	10
2.3	FDMI Access via Web Interfaces.....	11
3	Conclusion	14



List of Figures

Figure 1. FDMI reference architecture.8

Figure 2. The GreenDIGIT FDMI Login flow.9

Figure 3. The OAuth Access Token.10

Figure 4. The FDMI Catalogue interface.12

Figure 5. Example of Experiment published in the FDMI.12

Figure 6. Publishing of a Research product in the FDMI via Web User Interface.13



List of Abbreviations

Abbreviation	Description
FDMI	Federated Data Management Infrastructure
IAM	Identity and Access Management
IdP	Identity Provider
JWT	JSON Web Token
OIDC	OpenID Connect
SSO	Single Sign-On
UMA 2	User-Managed Access 2.0



1 Introduction and overview

The **Federated Data Management Infrastructure (FDMI)** is a flexible, extensible system tailored to the needs of the GreenDIGIT project. It has been enhanced to support a wide and evolving range of community-defined data asset types, such as **experiments**, **datasets**, and **scientific workflows**. Designed to enable the **publication**, **discovery**, and **reuse** of research outputs, the FDMI leverages rich metadata and standards-compliant interfaces to promote FAIR and reproducible science.

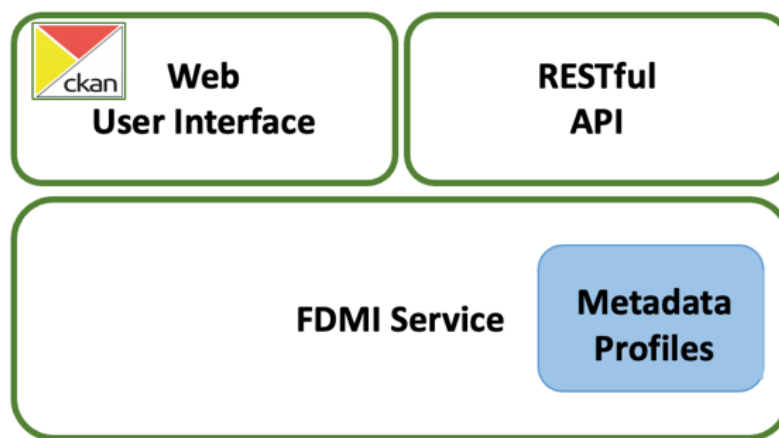


Figure 1. FDMI reference architecture.

Figure 1 illustrates the FDMI’s reference architecture, composed of four main components:

- **(a) Web User Interface** – an intuitive environment for browsing, searching, and interacting with catalogue contents;
- **(b) RESTful API** – provides programmatic access for querying and publishing metadata;
- **(c) FDMI Service** – handles metadata storage, indexing, and coordination across infrastructure components;
- **(d) Metadata Profiles** – define the structure, semantics, and validation rules for each resource type.

The actual content referenced in the catalogue may be stored via **persistent URLs** or directly within the SoBigData e-infrastructure, ensuring interoperability and long-term support for reproducible research.



2 FDMI Architecture, Access Mechanisms, and User Interfaces

FDMI is a flexible, extensible system built on top of the SoBigData e-infrastructure. It leverages the open-source platform CKAN exclusively for its Web User Interface, which enables users to interact with and explore metadata and research products in a user-friendly manner.

FDMI has been extended to support a diverse, community-driven, and expandable range of data asset types, such as custom metadata profiles. In particular, it was designed to accommodate the metadata models defined within the GreenDIGIT project, including the Experiment metadata profile.

All software components composing the FDMI have been successfully deployed and are operational. The FDMI Catalogue, accessible via a public Web Interface, provides streamlined access to the research products managed within the infrastructure and is available at: <https://greendigit.d4science.org/fdmi>.

2.1 Single Sign-On (SSO) and Federated Identity and Access Management Service (IAM)

The **GreenDIGIT Federated Identity and Access Management (IAM)** system allows users to securely access services using their institutional or other Identity Provider (IdP) (e.g. Google) credentials without the need to register a new account (Figure 2). Standards such as **OpenID Connect (OIDC)** and **SAML** ensure interoperability with federated research institutions.



Figure 2. The GreenDIGIT FDMI Login flow.



Single Sign-On (SSO) functionality enables users to log in once and access all services, simplifying workflows while maintaining strong security and centralised role-based access control.

Obtain Personal Token for APIs interactions

Current context: GreenDIGIT

Get Token

OAuth Access Token (Bearer Authorization)

eyJhbGciOiJIUzI1NiIsInR5cCI6IkpzZW50L3p2ZS4xNjpuNH-FkUTRLdEQ5VUhyY1dwNWJCT3NaLXpYbXM0In0.eyJleHAiOiJlMjI2NTZlNTcsmihndCl6MTc1MjYwODQ5NiwwaF90aWV1Ij0xNzUyNjA3NzU3LjQqdGkiOiJZTIINTIINSO4MMVxLTRmVzQlYmNi

expires in 17 hours, 47 minutes, 41 seconds

The token has a limited lifespan and must be included as a Bearer token in the Authorization header of the http request

The HTTP Request below is an example given in curl syntax, the equivalent HTTP can be coded in any language:

```
curl -X GET -H 'Authorization: Bearer [ACCESS_TOKEN]' [SERVICE_URL]
```

ref: <https://oauth.net/2/access-tokens/>

Instructions

Obtain your personal OAuth2 Access, to be used for programmatic interaction with the Resources

Do not use this token to authenticate services but only for personal access

ref:

- [Accessing Resources - how to](#)
- [OAuth2: Making Authenticated Requests](#)

Please note: if you need long lasting access tokens you may obtain a service account by opening a request

You can decode the tokens using <https://jwt.io/>

Figure 3. The OAuth Access Token.

To obtain an OAuth 2.0 Access Token, GreenDIGIT relies on the SoBigData e-infrastructure, which adopts industry-standard protocols for secure authentication and authorisation. Specifically, it supports OpenID Connect (OIDC) for authentication and User-Managed Access 2.0 (UMA 2) for authorisation. The system issues JSON Web Tokens (JWTs) as Access Tokens, used to securely access services. Figure 3 shows an example of a JWT that can be used to interact with the FDMI APIs.

2.2 FDMI Access via APIs

GreenDIGIT's FDMI access are provided via APIs through the gCat (gCube Catalogue) service, part of SoBigData. This RESTful web service exposes a complete set of operations using standard HTTP methods, supporting full CRUD (Create, Read, Update, Delete) functionality. Access requires a valid JWT Bearer token, ensuring secure and standards-compliant interactions. The service returns appropriate HTTP status codes (e.g. 200 OK, 201 Created, 401 Unauthorized, 404 Not Found, 500 Internal Server Error) in line with REST best practices. Users can integrate gCat into their workflows using any REST client or application.



Documentation is available at: <https://api.d4science.org/gcat/docs/index.html>

Its **source code is released under the European Union Public Licence v.1.2 (EUPL v.1.2)** and is publicly available at: <https://code-repo.d4science.org/gCubeSystem/gcat>

2.3 FDMI Access via Web Interfaces

Figure 4 shows the FDMI Catalogue interface, built on top of CKAN, a widely adopted open-source data management system. The interface allows users to search, access, and manage metadata records for datasets registered within the GreenDIGIT infrastructure. It provides a user-friendly environment for publishing and discovering marine and climate-related resources in alignment with FAIR principles.

Home Organisations Groups Items Types Statistics

GreenDIGIT

As part of the GreenDIGIT initiative, the Federated Data Management Infrastructure (FDMI) provides a scalable, open-source platform for managing, publishing, and discovering research artefacts, such as datasets, methods, and experiment results, across a network of Research Infrastructures. Powered by the SoBigData RI, FDMI enables rich metadata annotation to support reproducible and energy-efficient science, fostering responsible data lifecycle practices in line with the principles of Open Science and GreenDIGIT's sustainability objectives.

All the products are accompanied with rich descriptions capturing general attributes, e.g. title and creator(s), as well as usage policies and licences.

Items Search

See All Items See All Tags

GreenDIGIT Catalogue statistics

6 items 1 organisation 2 groups 1 type

Browse by Types



Experiment (6)

See All Types

Popular Formats

ZIP application/x-msdos-pro... JPEG PNG

Popular Tags

energy measurements Test tag reproducibility

See All Tags






Figure 4. The FDMI Catalogue interface.

The FDMI Catalogue interface is available at <https://greendigit.d4science.org/fdmi>.

The screenshot shows the FDMI Catalogue interface for an experiment titled "MoonGen example experiment with energy measurement". The page includes a navigation bar with links for Home, Organisations, Groups, Items, Types, and Statistics. The breadcrumb trail is "/ Organisations / GreenDIGIT / MoonGen example experiment ...".

Experiment Details:

- Title:** MoonGen example experiment with energy measurement
- Followers:** 0
- Organisation:** GreenDIGIT. Description: GreenDIGIT provides researchers with a web-based platform to design and conduct reproducible experiments while promoting Open Science practices. It streamlines the entire... [read more](#)
- License:** Creative Commons Attribution 4.0 [OPEN DATA](#)
- Tags:** energy measurements, reproducibility
- Data and Resources:** To access the resources you must log in. [Go to Login...](#)
 - RO-Crate ZIP Archive**
- Item URL:** <https://data.d4science.org/ctlg/GreenDIGIT/moongen-example-experiment-with-energy-measurement> 

Additional Info Table:

Field	Value
Creation Date	2025-06-13
Creator	Kilian Warmuth
Creator Email	kilian.warmuth@tum.de
Creator Name PI (Principal Investigator)	Kilian Warmuth
Environment OS	Linux
Environment Platform	D4Science GreenDIGIT
Experiment Dependencies	none
Experiment ID	exp-green-digit-001
GreenDIGIT Node	D4Science Pisa
Programming Language	Python
Project ID	GD-T5.2
Session reading metrics	enabled
system:type	Experiment

Management Info Table:

Field	Value
Author	Warmuth Kilian
Maintainer	Warmuth Kilian
Last Updated	27 June 2025, 15:56 (UTC+02:00)
Created	27 June 2025, 15:56 (UTC+02:00)

Figure 5. Example of Experiment published in the FDMI.



Figure 5 illustrates the metadata structure and the corresponding payload(s) associated with the example experiment titled “MoonGen example experiment with energy measurement.”

Researchers can programmatically access the FDMI through the dedicated RESTful APIs, which are fully documented and available at: <https://api.d4science.org/catalogue/api-docs/ui/index.html>.

Research Products can be published either via these programmatic interfaces (gCat service) or through user-friendly Web Forms, as shown in Figure 6.

Publish Item

1. Edit Common Metadata | 2. Edit Item Specific Metadata & Publish | 3. Add Resources

Insert Item Information * is required

* Title : ⓘ

Description: ⓘ

* Tag: ⓘ

License: ⓘ

Selected License Url:

Visibility: ⓘ

Publish in: ⓘ

Version:

* Author: ⓘ

* Author Email: ⓘ

Figure 6. Publishing of a Research product in the FDMI via Web User Interface.



3 Conclusion

The GreenDIGIT Federated Data Management Infrastructure (FDMI), developed on the SoBigData platform, provides a robust environment for publishing, discovering, and reusing datasets, workflows, and experiment results. Through secure Single Sign-On, standards-based APIs, and a user-friendly web interface, researchers can easily register and share their outputs in line with FAIR and reproducibility principles. The infrastructure is fully operational and designed to support integration with widely used open repositories such as Zenodo, ensuring that research data and results can be disseminated broadly and preserved for long-term accessibility.