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

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

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

List of Figures

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.

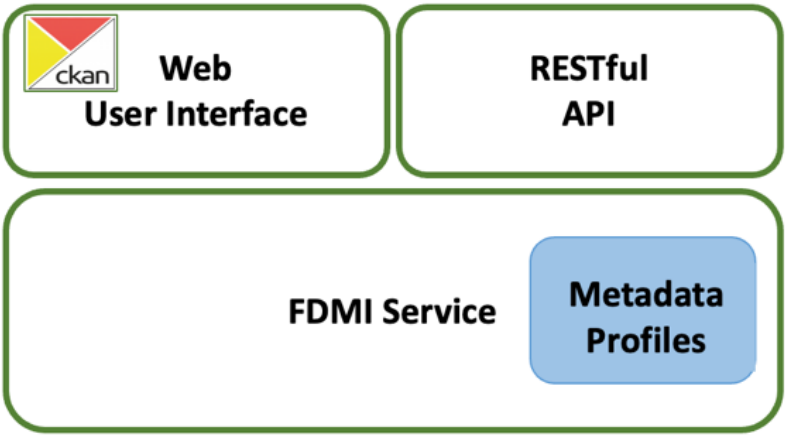

Figure 1. FDMI reference architecture.

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.

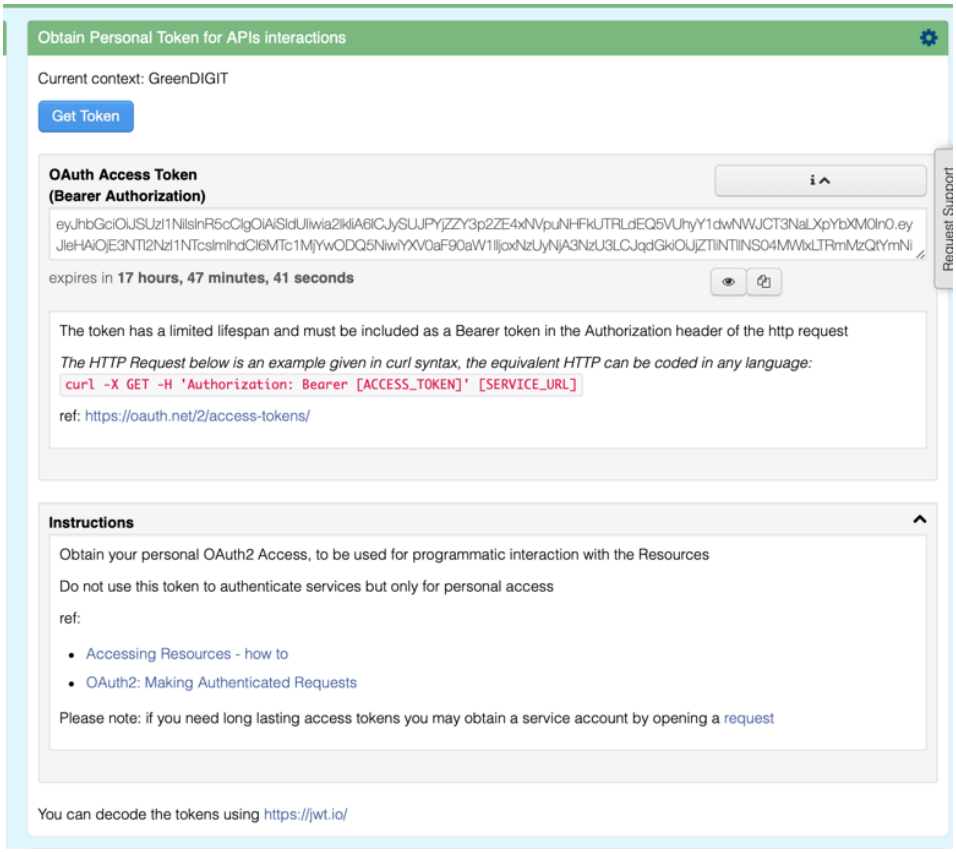

Figure 3. The OAuth Access Token.

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.

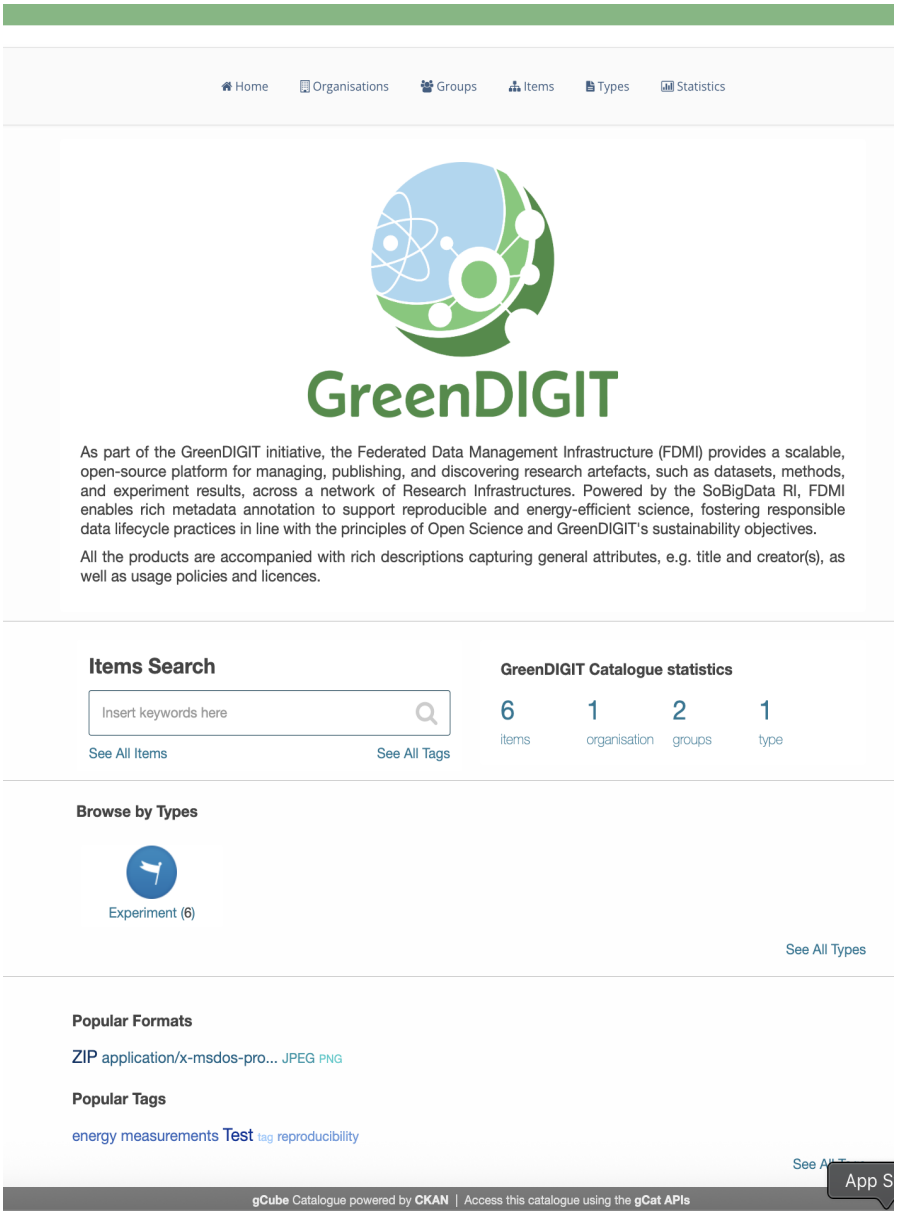

Figure 4. The FDMI Catalogue interface.

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

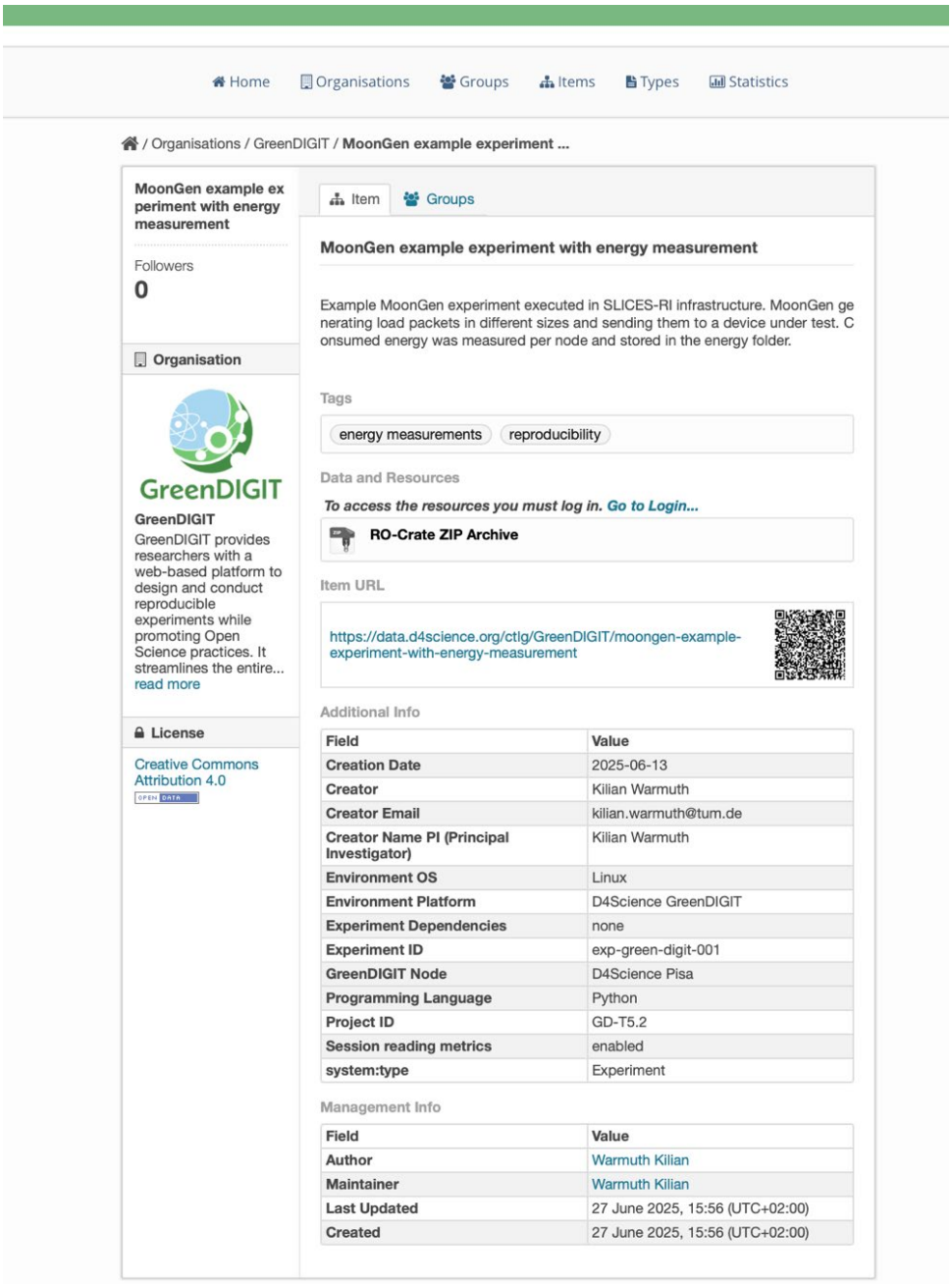

Figure 5. Example of Experiment published in the FDMI.

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.

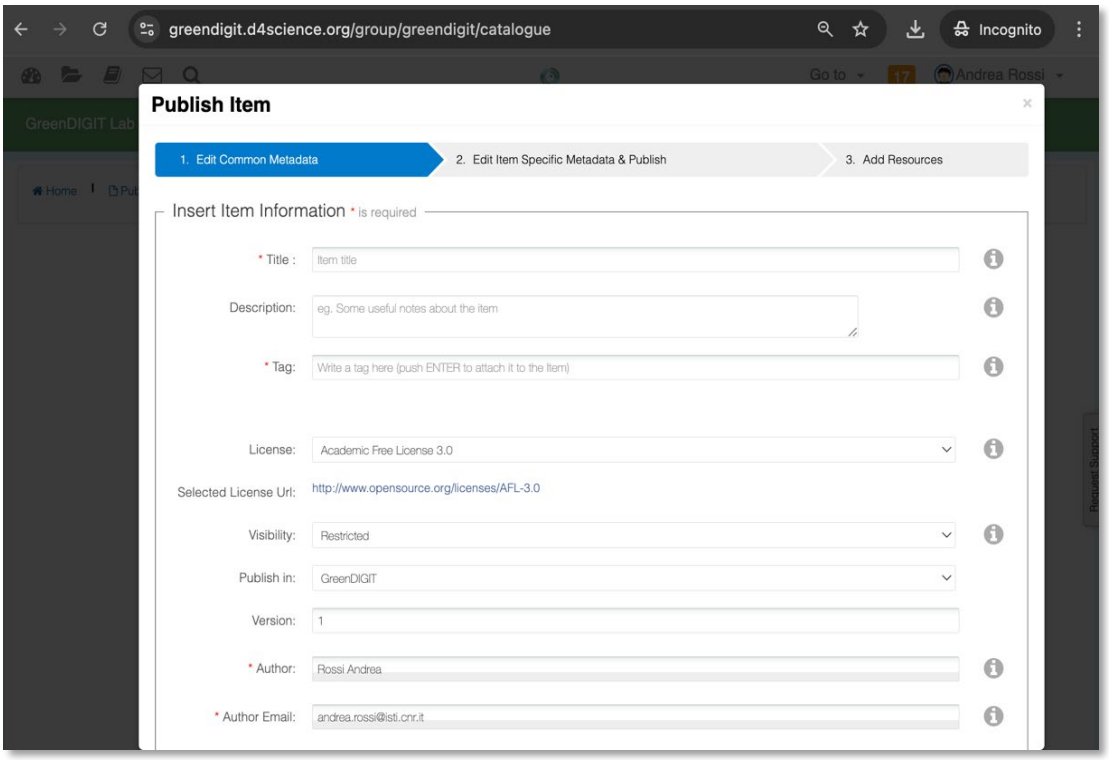

Figure 6. Publishing of a Research product in the FDMI via Web User Interface.

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.