D5.7 - FAIR Research Data Management Workbench Operation Report Update

Lead Partner:	CNR
Authors	L. Candela (CNR-ISTI), L. Frosini (CNR-ISTI), Y. Le Franc (CINES), F. Mangiacrapa (CNR-ISTI), O. Rouchon (CINES), B. Toulemonde (CINES)
Version:	1.1
Status:	Final
Dissemination Level:	Public
Document Link:	doi:10.5281/zenodo.6341732

Deliverable Abstract

EOSC-Pillar developed and integrated a set of tools and services overall supporting the construction and maintenance of an aggregated data space implementing the FAIR principles. This deliverable documents the activities and results (e.g., indicators on integrated data providers, and datasets, indicators on datasets accesses) of the operation of the EOSC-Pillar toolset enacting the development of the EOSC-Pillar data space. This is the revised and final release of this typology of deliverable offering information up to November 2022.

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement No. 857650.

COPYRIGHT NOTICE

DELIVERY SLIP

Date	*Name*	*Partner/Activity*	*Date*
From:	L. Candela	CNR	
Moderated by:	Y. Le Franc	CINES	
Reviewed by:	V. Breton, F. Galeazzi	CNRS / GARR	
Approved by:			

DOCUMENT LOG

Issue	*Date*	*Comment*	*Author*
v0.1	1 Nov 2022	First draft	L. Candela, B. Toulemonde
v0.8	6 Dec 2022	Ready for internal review version	L. Candela, B. Toulemonde
v0.9	14 Dec 2022	Annotated version with internal reviewers comments	V. Breton, F. Galeazzi
V1.0	16 Dec 2022	Final version	L. Candela
V1.1	05 May 2023	Improved resolution figures 12-15	F. Galeazzi

TERMINOLOGY

https://eosc-portal.eu/glossary

Terminology/Acronym	*Definition*
API	Application Programming Interface
FAIR Data Point	A software enabling the implementation of a metadata repository providing access to metadata according to the FAIR principles.
Federated FAIR Data Space	A unifying data space aggregating datasets scattered across several data sources and repositories with the aim to give access to them according to the FAIR principles.
FDP	see FAIR Data Point
F2DS	see Federated FAIR Data Space
RDM	Research Data Management
Virtual Research Environment	A web-based working environment conceived to provide a community of practice with services and data of interest
VRE	see Virtual Research Environment

Contents

1 Introduction 5
2 Operating the EOSC-Pillar Federated FAIR Research Data Management Workbench 6
2.1 Supporting the data provision by F2DS facilities 7
2.2 Supporting the data consumption by F2DS facilities 11
2.3 Populating and Exploiting the Federated FAIR Data Space 13
2.4 Operation Activity Indicators 19
3 Conclusion and Remarks 22
4 References 23

Table of Figures and Tables

Figure 1 F2DS workbench Overall Architecture 7
Figure 2 F2DS Metadata Repository Data Provisioning GUI 8
Figure 3 F2DS Data Catalogue Harvesting GUI 9
Figure 4 F2DS Data Catalogue Publish Item GUI 10
Figure 5 Data Catalogue Item Model 11
Figure 6 F2DS Metadata Repository Search GUI 12
Figure 7 F2DS Data Catalogue GUI 13
Figure 8 EOSC-Pillar for Agrifood VRE: Catalogue Service 15
Figure 9 EOSC-Pillar for Earth Science VRE: Catalogue Service 16
Figure 10 EOSC-Pillar for COVID-19 VRE: Catalogue Service 16
Figure 11 EOSC-Pillar for Research Data Catalogue VRE: Catalogue Service 17
Figure 12 Catalogue Accesses 21
Figure 13 Catalogue Search & Browse 21
Figure 14 Catalogue Metadata Views 21
Figure 15 Catalogue Resource Views 21

Table 1 EOSC-Pillar Data Sources integrated into F2DS 14
Table 2 Mapping between Metadata Repository and Catalogue 17
Table 3 F2DS Operation Activity Indicators 19
Table 4 F2DS Operation Activity Indicators up to November 2022 20

Executive summary

EOSC-Pillar Work Package 5 “The Data layer: establishing FAIR data services at the national and transnational level” was called to establish the settings for an effective sharing, exploitation and reuse of data across initiatives and communities partaking to EOSC-Pillar and beyond. In order to attain this challenging goal, the project leverages and builds upon results from previous and ongoing projects as well as on the experience of the partners in the project.

This combined expertise offers to data providers and data consumers a dedicated set of services (and accompanying training) supporting the creation of a ***Federated FAIR Data Space (F2DS)*** where multiple datasets from scattered data sources are virtually joined by combining their metadata and are subsequently published and made available in accordance with the FAIR principles.

This F2DS offers a rich array of tools for both data providers and data consumers. The tool for data providers makes it possible to make data more compliant with the FAIR principles and any other specific policies, as well as to integrate them with other data across disciplines, thus enabling the development of a unifying data space. The tools for data consumers facilitate the discovery and access to the datasets populating the unifying data space.

This deliverable provides the readers with an up-to-date brief description of the services contributing to the EOSC-Pillar F2DS workbench and gives indicators on the exploitation of this technology when dealing with datasets and data sources of interest for the communities involved in EOSC-Pillar via the use cases developed in WP6. In particular, the deliverable documents the operation activity up to November 2022. The following operation activities were performed: (a) a total of 6 data sources of interest have been integrated to showcase the early implementation of the F2DS, (b) a total of 81k datasets resulted from this initial integration, (c) 4 virtual research environments have been deployed to provide the communities with F2DS service instances.

1 Introduction

The primary goal of the EOSC-Pillar WP5 "The Data layer: establishing FAIR data services at the national and transnational level" was to create the conditions for an effective sharing, exploitation and reuse of data across initiatives and communities partaking to EOSC-Pillar and beyond. To pursue this challenging goal, the project planned to leverage and build upon the wealth of past and ongoing projects, initiatives and experiences to provide *data providers* and *data consumers* with the tools they need to develop a *shared data space* where datasets of interest are collected from scattered data sources and providers and published following unifying strategies thus to become seamlessly and easily findable, accessible, interoperable and reusable in accordance with the FAIR principles [10].

The overall tool set proposed and developed by WP5 was documented by a specific deliverable [4]. This tool set integrates and complements existing tools and approaches thus to realize an end-to-end integrated solution providing both (a) ***data providers*** with services facilitating the integration and publishing of existing datasets into a shared data space matching the FAIR principles and (b) ***data consumers*** with services facilitating the seamless discovery and access to the datasets contributing to the shared data space. In a first step, the tools stemming from EOSC-Pillar Tasks T5.1 and T5.2, namely the Metadata Repository (based on FDP) and the Data Catalogue (based on D4Science) were seamlessly integrated into a single solution. In a second step, a third facility was added that stems from EOSC-Pillar Task T5.5 [7] and enacts the semantic annotation of datasets during the onboarding phase. The overall solution is named ***EOSC-Pillar Federated FAIR Data Space workbench*** and consists of a set of interoperating services and approaches enabling human and machine users to populate the EOSC-Pillar Federated FAIR Data Space and subsequently discover and access the published datasets seamlessly.

The rest of the deliverable provides an overview of the set of components and tools developed to implement the EOSC-Pillar Federated FAIR Data Space workbench by highlighting their major features and how these components have been integrated and exploited.

2 Operating the EOSC-Pillar Federated FAIR Research Data Management Workbench

The EOSC-Pillar Federated FAIR Data Space (F2DS) is a unifying data space that is built by aggregating and enriching datasets from a set of scattered repositories/data sources and communities with the aim to facilitate the steps ranging from data discovery up to re-use in accordance with the FAIR principles and practices.

While datasets are the primary item typology of the resulting data space, other typologies of items might be managed including repositories and data sources, APIs, metadata schemas and ontologies.

The implementation of the F2DS concept is leveraging on existing tools and services that have been developed to deal with similar issues so far. The EOSC-Pillar F2DS tool set [4] has two focal points: (i) a ***metadata repository*** aggregating the dataset of interest and offering them via APIs and protocols adhering to FAIR principles, and (ii) a ***data catalogue*** offering search and browse on top of the aggregated datasets as well as the possibility to implement and integrate "views" of the whole data space to be included into virtual research environments.

The Metadata Repository is primarily based on the FAIR Data Point technology[1] extended by a GUI for, e.g., registering data sources and discovering aggregated datasets plus tools taking care of the harvesting of metadata and allowing the FAIRification of metadata. The instance of this component is available at https://f2ds.eosc-pillar.eu/login while the source code is available at https://git.garr.it/eosc-pillar/ffds/.

The Data Catalogue is based on gCube technology (in turn relying on CKAN[2] technology for the catalogue) [2]. This technology has been extended to support the definition of specific metadata profiles for the catalogue items. Moreover, it has been extended to be nicely integrated within Virtual Research Environments thus to provide the specific community served by the VRE with a custom view of the data space of interest. Finally, it was extended to interface with the Metadata Repository and systematically collects the datasets onboarded into it. An instance of this component is available at https://eosc-pillar.d4science.org/ while its source code is available at https://code-repo.d4science.org/ (by several repositories[3]).

Figure 1 depicts how the two focal points interact with each other as well as the fact that the two components provide a rich array of solutions, made available via GUI and APIs, for supporting the data provision (cf. Sec. 2.1) and the data consumption (cf. Sec. 2.2) phases. In particular, the two components will exchange metadata about the datasets aggregated into the data space thus to make it possible for both of them to populate internal data structures, and to develop dedicated views on top of the overall data space.

[1] https://github.com/FAIRDataTeam/FAIRDataPoint-Spec

[2] https://ckan.org/

[3] It consists of diverse components including portlets like gcube-ckan-datacatalog, connectors like ckan-connector, software libraries like ckan-util-library, ad-hoc harvesters like the FDP-harvester, and web services like gcat and gFeed.

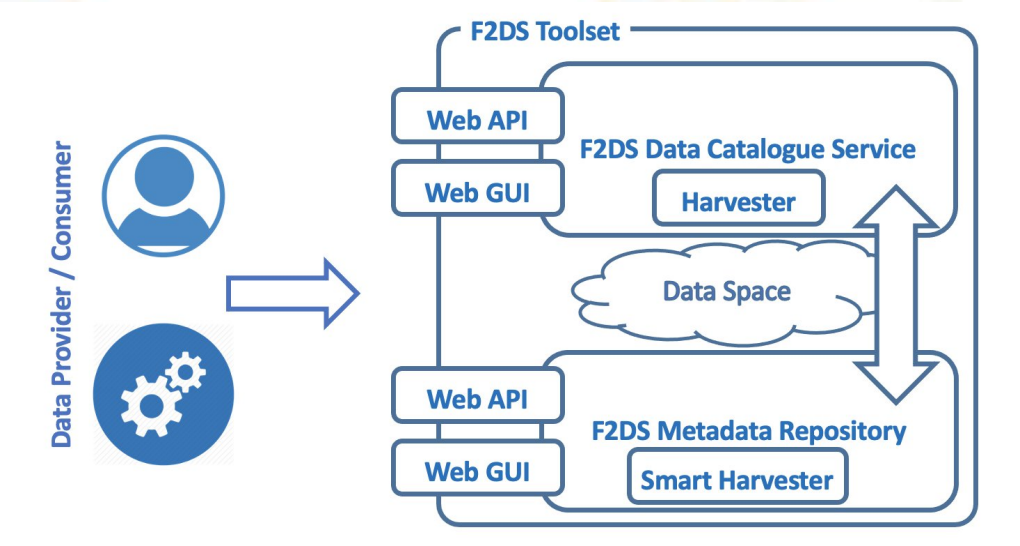

Figure 1 F2DS workbench Overall Architecture

2.1 Supporting the data provision by F2DS facilities

The F2DS tool set offers two approaches for populating the data space:

- by *harvesting,* i.e., both the Metadata Repository and the Data Catalogue implement a method to collect metadata about datasets from existing data sources;
- by *publishing*, i.e., both the Metadata Repository and the Data Catalogue implement a method to add metadata about single datasets directly into the data space they expose.

Regarding the harvesting, the Metadata Repository offers a GUI enabling a data provider to contribute a catalogue / data source via a simple 5-step workflow (see Figure 2):

1. Describe the catalogue / data source;
2. Describe the catalogue API / data source API to access the data;
3. Carry out a metadata mapping in DCAT format;
4. Annotate the dataset with semantic artefacts;
5. Launch the population of the FAIR Data Point (FDP) with metadata describing the available datasets of the catalogue / data source.

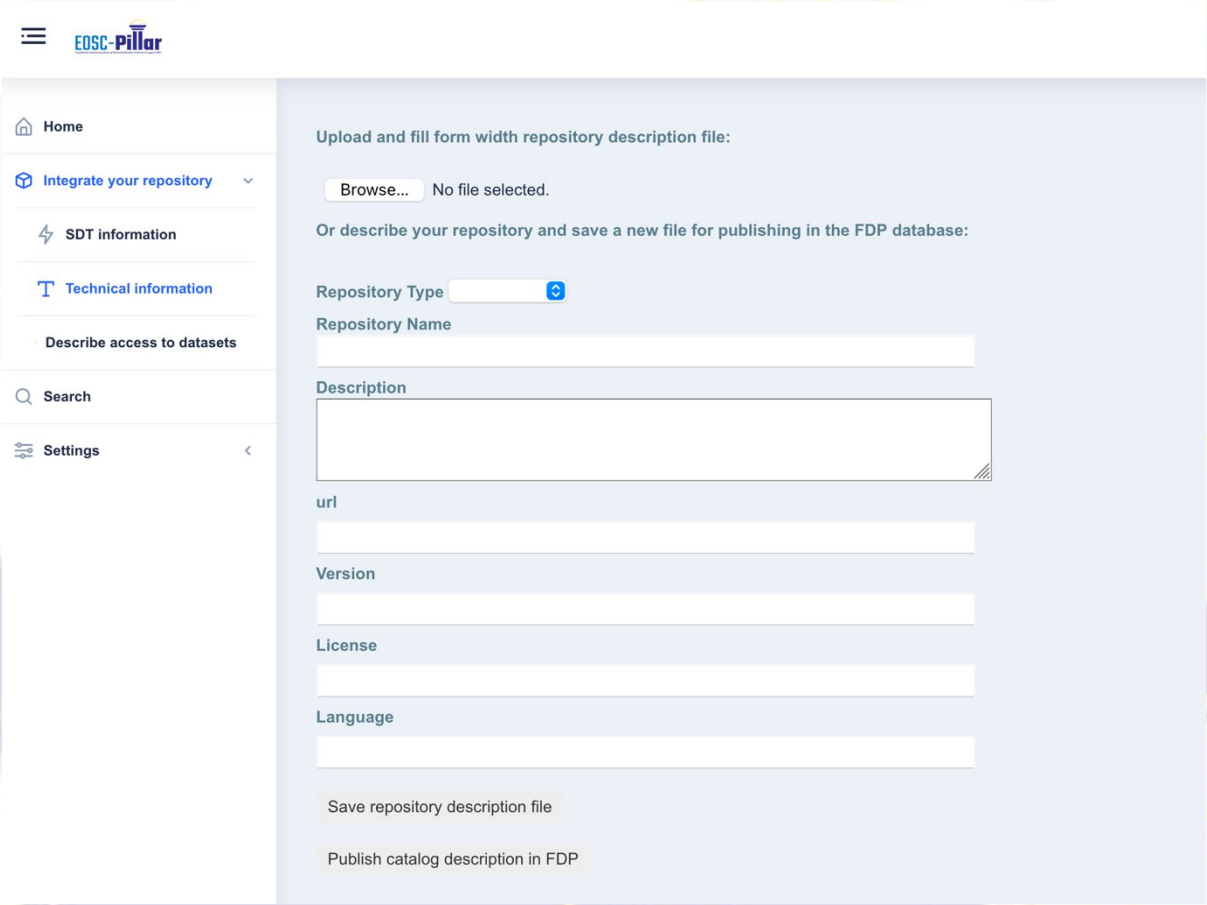

Figure 2 F2DS Metadata Repository Data Provisioning GUI

Regarding the harvesting, the Data Catalogue offers a GUI (see Figure 3) where authorised users can register data sources of interest and specify the protocol to use to collect metadata (including DCAT, OAI-PMH, and CSW). Besides the GUI, the Data Catalogue offers a RESTful service (gFeed) that can be exploited to run a harvesting task by counting on specific plug-ins (i.e., specific components developed to interact with the data source)[4].

[4] At the time of writing this report, two gFeed harvesting plug-in have been developed to harvest content from OAI-PMH data sources and to harvest content from the Metadata Repository by DCAT.

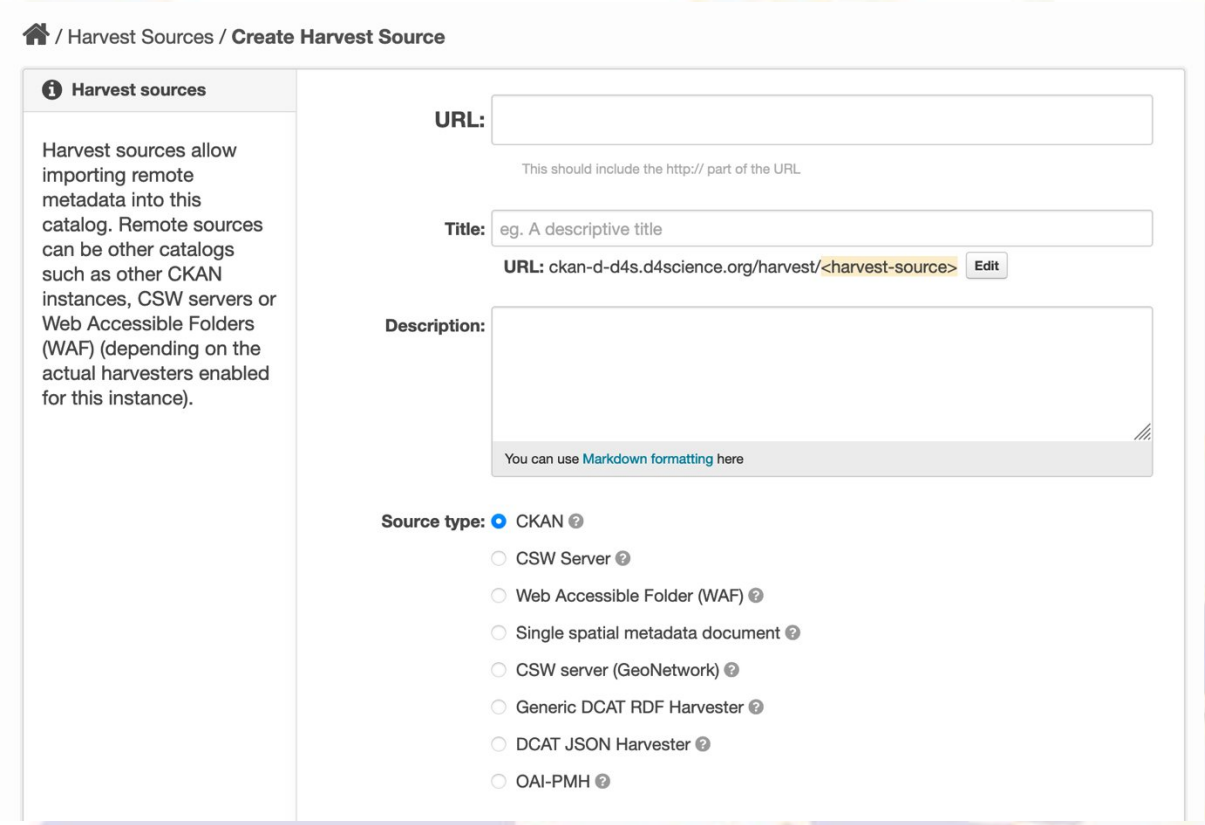

Figure 3 F2DS Data Catalogue Harvesting GUI

Regarding the publishing, the Data Catalogue offers a GUI (see Figure 4) enabling authorised users to register single datasets[5] by simply compiling a form. The metadata collected by the form can be configured for every single domain, i.e., catalogue item profiles can be defined by properly instantiating and exploiting the catalogue item model discussed below.

[5] Actually, any item the user community defines by specifying the typology (i.e., a name and a set of metadata fields).

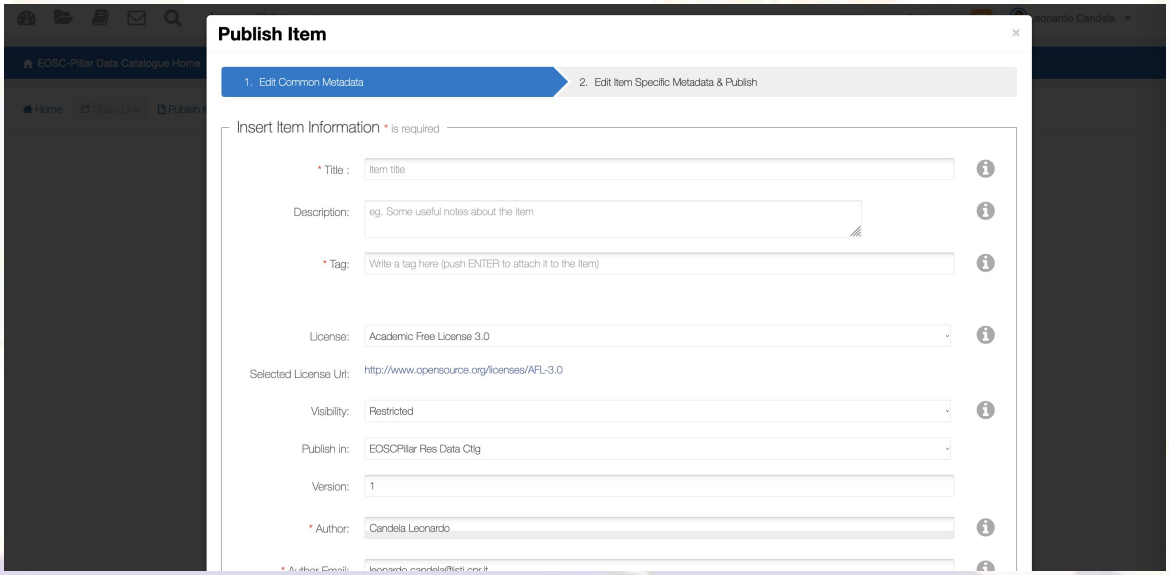

Figure 4 F2DS Data Catalogue Publish Item GUI

The model characterising every catalogue item (see Figure 5) is quite powerful and flexible. Every catalogue item:

- has a number of *common metadata* (independently of the typology of the item) including (i) a unique identifier, (ii) a title, (iii) a description, (iv) a number of tags, (v) a license, (vi) a visibility flag (to indicate whether the item is public, i.e. visible to everyone, or private, i.e. the item is visible to VRE members only), (vii) an author, (viii) a maintainer, and (ix) a typology. A typology adds additional metadata (see item-specific metadata);
- has a number of *item specific metadata*. Item specific metadata can be driven by a profile specifying a number of field specifications each characterising a metadata field by specifying: (i) its name, (ii) the mandatory flag (whether the field is mandatory or optional), (iii) the type of field values (i.e. String, Text, Boolean, Number, Geometry in GeoJSon, Time, Time interval, List of Times), (iv) the maximum number of occurrences (i.e. whether the field is repeatable or not), (v) any default value to be proposed, (vi) any accompanying note to facilitate the data entry, (vii) any controlled vocabulary to facilitate the selection of suitable values, (viii) any validator to check the inserted value correctness;
- has a number of specific *resources*, i.e., objects representing identifiable item payloads. Every resource has (i) an Identifier, (ii) the URL where the payload is stored, (iii) a name, (iv) a description, and (v) a format (e.g., CSV, XML).

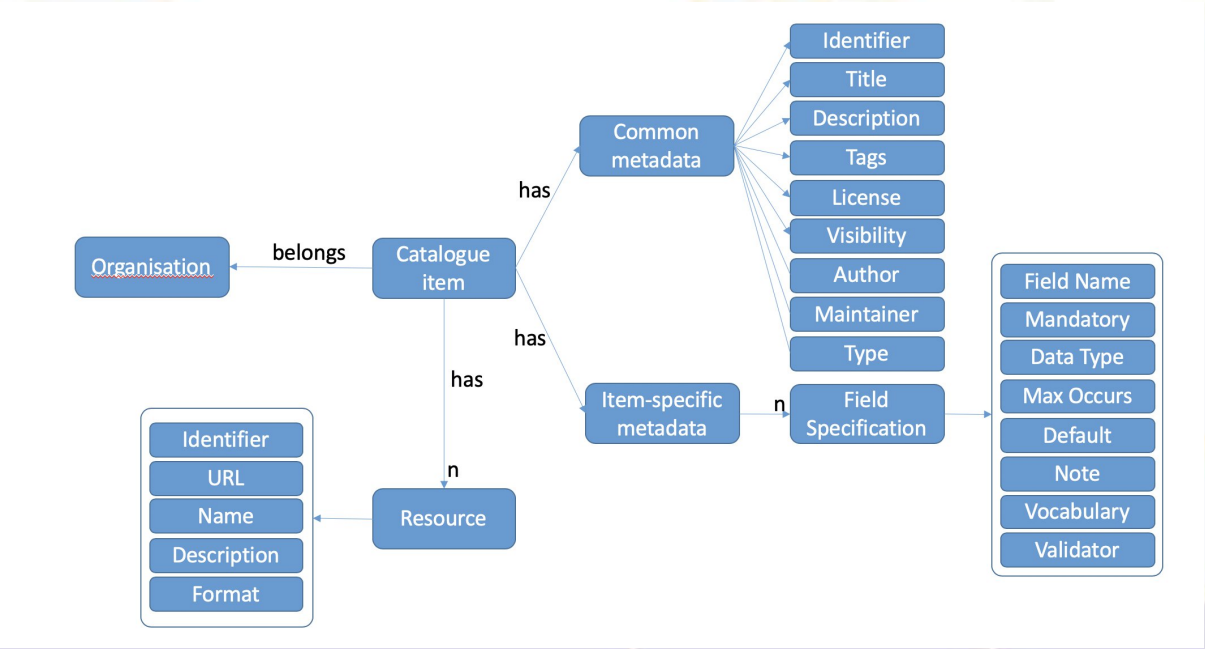

Figure 5 Data Catalogue Item Model

These facilities (i.e. the Metadata Repository and Data Catalogue) open a number of exploitation scenarios, including:

- data sources are integrated into the F2DS via the Metadata Repository facilities and subsequently made available for data consumption via both the Metadata Repository itself and the Data Catalogue. One or more Data Catalogue instances, each conceived to serve the needs of a designated community, can be created on top of the content aggregated and FAIRified by the Metadata Repository.
- Datasets can be created into a Data Catalogue instance by either harvesting existing data sources or publishing facilities. Once into the Data Catalogue, content can be manipulated / curated by the community and subsequently flows into the Metadata Repository via its harvesting facility.

2.2 Supporting the data consumption by F2DS facilities

The F2DS offers a rich array of facilities for discovery and access to the metadata of the datasets aggregated into the F2DS. Depending on the entry point used by the users to access the F2DS content, diverse views are implemented and supported including both GUIs and APIs.

Regarding the APIs, the following ones are supported:

- The Metadata Repository exposes its content via the FAIR Data Point REST API[6];
- The Metadata Repository make it possible to search the content by SPARQL queries (by relying on the Blazegraph™ DB)[7];

[6] The swagger based description of the API is available by https://f2ds.eosc-pillar.eu/smart-harvester/open-api.html
[7] F2DS Metadata Repository Blazegraph DB GUI https://f2ds.eosc-pillar.eu/blazegraph/#query

- The Data Catalogue supports a REST API (by the gCat service)[8] that by the item collection operations makes it possible to programmatically access its content;
- The Data Catalogue exposes DCAT RDF endpoints for both the whole catalogue and the single item (by using a CKAN plugin)[9].

Regarding the GUIs, the following ones are supported:

- The Metadata Repository offers a keyword-based search on its content (see Figure 6).

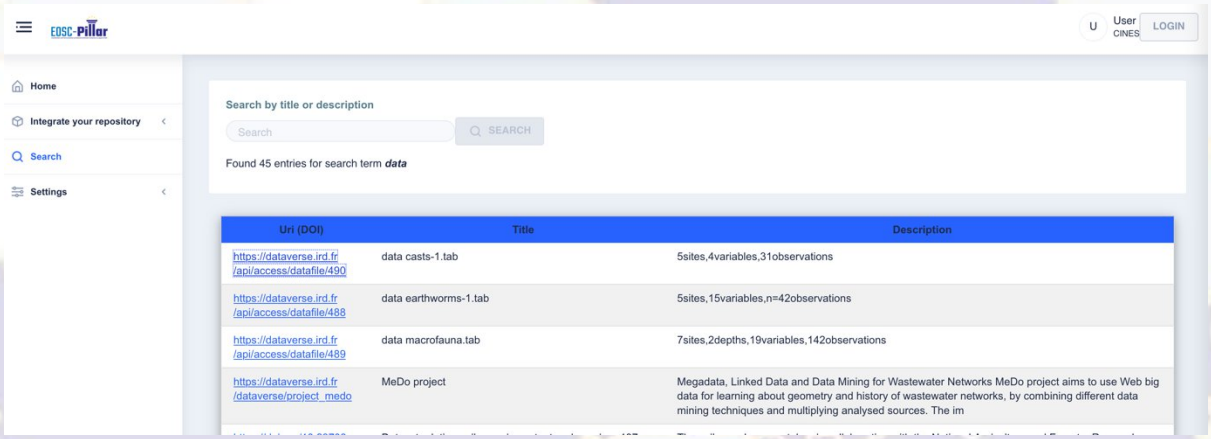

Figure 6 F2DS Metadata Repository Search GUI

- The Data Catalogue offers a GUI enabling users to execute keyword-based search as well as to browse and filter content via faceted-search and spatial extent (see Figure 7).

[8] F2DS Data Catalogue REST API https://wiki.gcube-system.org/gcube/GCat_Service

[9] The F2DS Data Catalogue DCAT RDF Endpoint https://ckan-eoscpillar.d4science.org/catalog.rdf (it can also be serialized in RDF/XML, Turtle, Notation3, and JSON-LD). There is also an endpoint for every single catalogue item via the following URI schema https://ckan-eoscpillar.d4science.org/dataset/{datasetid}.{format}

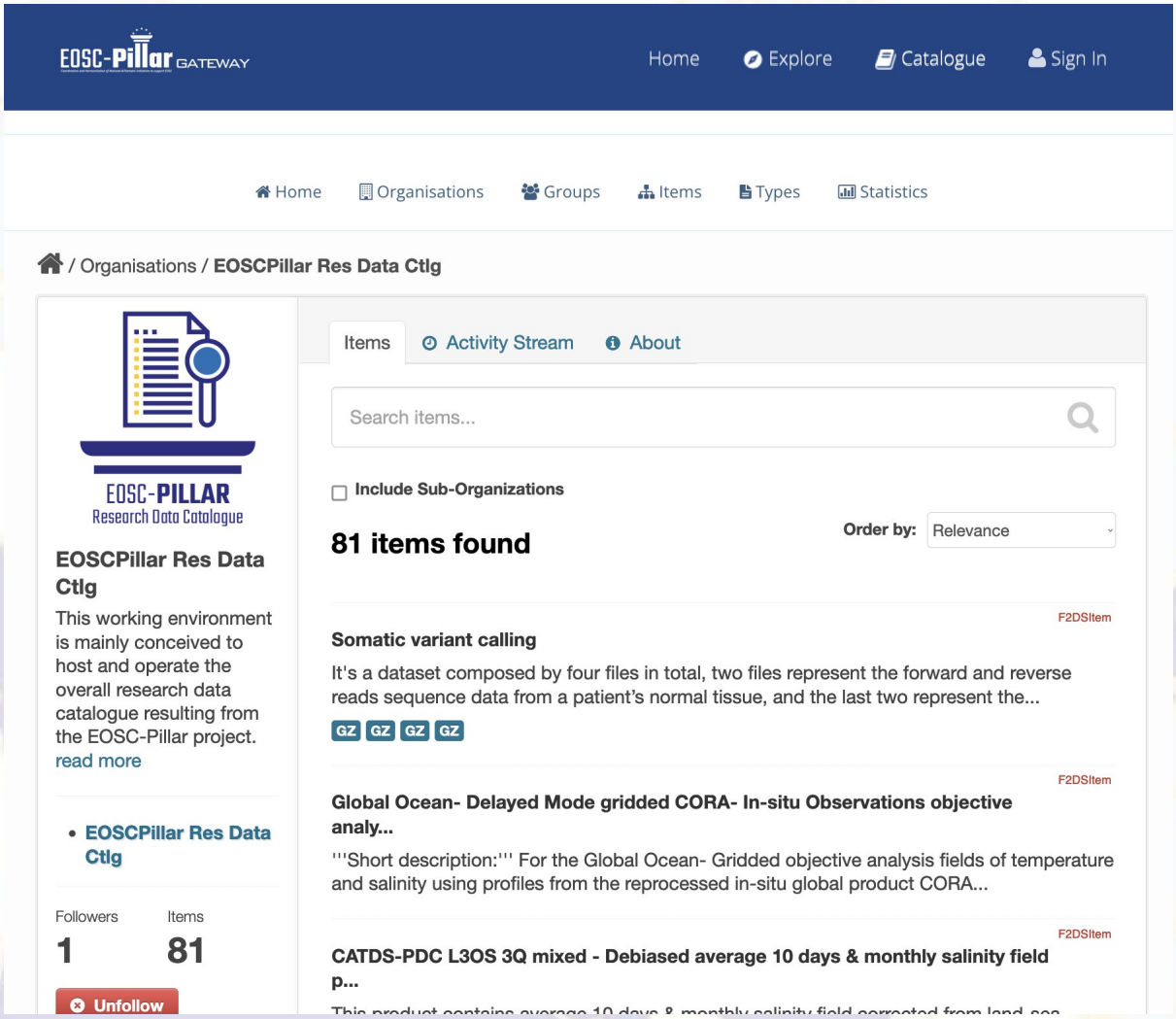

Figure 7 F2DS Data Catalogue GUI

These facilities open a number of exploitation scenarios including:

- Data consumers can use any of the available endpoints and APIs for programmatically accessing the data space or part of it;
- Communities can embed an F2DS Data Catalogue instance in a Virtual Research Environment thus to have a focused access on community defined part of the whole F2DS.

2.3 Populating and Exploiting the Federated FAIR Data Space

The F2DS Metadata Repository was made available as a service on its own at https://f2ds.eosc-pillar.eu/login as well as from a dedicated virtual research environment (see below). This web interface allows to register and describe the mapping in order to harvest the dataset to a DCAT schema. The metadata harvested are then published to an instance of the Fair Data Point available through https://f2ds.eosc-pillar.eu/app.

The Repositories and Data Sources exploited to populate the EOSC-Pillar Federated FAIR Data Space originate from WP6 use cases [9]. In particular, the following ones were successfully integrated.

Table 1 EOSC-Pillar Data Sources integrated into F2DS

Name	Use Case	Description	Protocols and APIs	Metadata Format(s)
CMIP5 ESGF Data Collection	T6.1: "Defining procedures/service to enforce data provenance for thematic communities and beyond"	See https://esgf-data.dkrz.de/search/cmip5-dkrz	REST API	Proprietary
CMIP6 ESFG Data Collection	T6.1: "Defining procedures/service to enforce data provenance for thematic communities and beyond"	See https://esgf-data.dkrz.de/search/cmip6-dkrz	REST API	Proprietary
CORA Dataset (Global Ocean-Delayed Mode gridded CORA)	T6.2: "Agile FAIR data for environment and earth system communities (ocean, atmosphere, continental surfaces, solid earth)"	See https://sextant.ifremer.fr/Donnees/Catalogue#/metadata/7691f8b8-1193-4aef-8e7e-0d7a9c88c057	CSW, OAI-PMH, proprietary APIs	ISO-19115, DublinCore
SMOS Dataset (CATDS-PDC L3OS 3Q mixed - Debiased average 10 days & monthly salinity field product from SMOS satellite (mixed orbits)	T6.2: "Agile FAIR data for environment and earth system communities (ocean, atmosphere, continental surfaces, solid earth)"	See https://sextant.ifremer.fr/Donnees/Catalogue#/metadata/0f02fc28-cb86-4c44-89f3-ee7df6177e7b	CSW, OAI-PMH, proprietary APIs	ISO-19115, DublinCore

ARGO GDAC	T6.2: "Agile FAIR data for environment and earth system communities (ocean, atmosphere, continental surfaces, solid earth)"	See https://sextant.ifremer.fr/Donnees/Catalogue#/metadata/3df904de-e47d-4bf9-85a0-7c0942aff8b6	CSW, OAI-PMH, proprietary APIs	ISO-19115, DublinCore
Data INRAE	T6.3: "Integration of data repositories into EOSC based on communities' approaches"	The Data portal of INRAE	OAI-PMH	DublinCore

The F2DS Data Catalogue was made available from the following four Virtual Research Environments.

EOSC-Pillar 4 Agrifood This virtual research environment was created to support the implementation of UC6.3: "Integration of Data repositories into EOSC based on communities' approaches" [9]. In particular, the catalogue was there to make it possible to publish new research artefacts stemming from the analysis of data published by the INRAE repository.

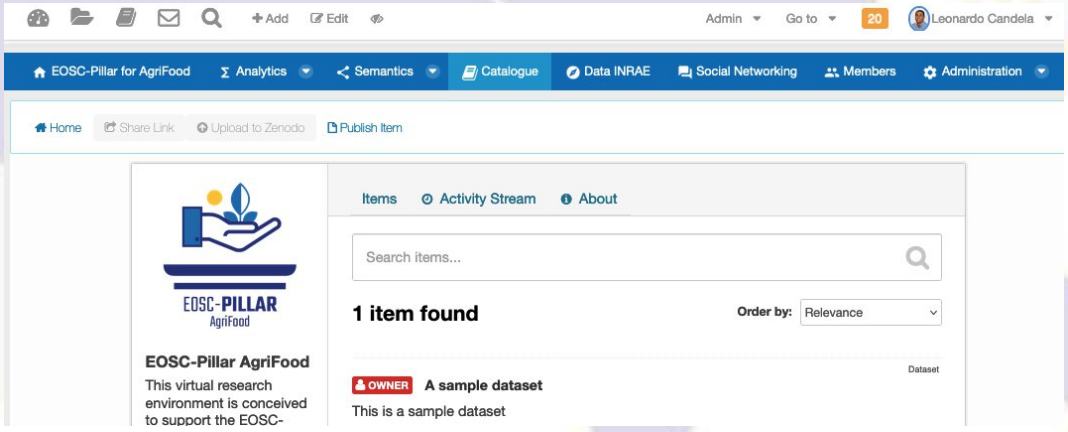

Figure 8 EOSC-Pillar for Agrifood VRE: Catalogue Service

EOSC-Pillar 4 Earth Science This virtual research environment was created to support the UC6.1: "Defining Procedures/Services to enforce Data provenance for thematic communities and beyond" and UC6.2: "Agile FAIR data for Earth Environment and Geosciences communities" [9]. In particular, the catalogue was configured to give access to the CMIP data set integrated into the Metadata Repository.

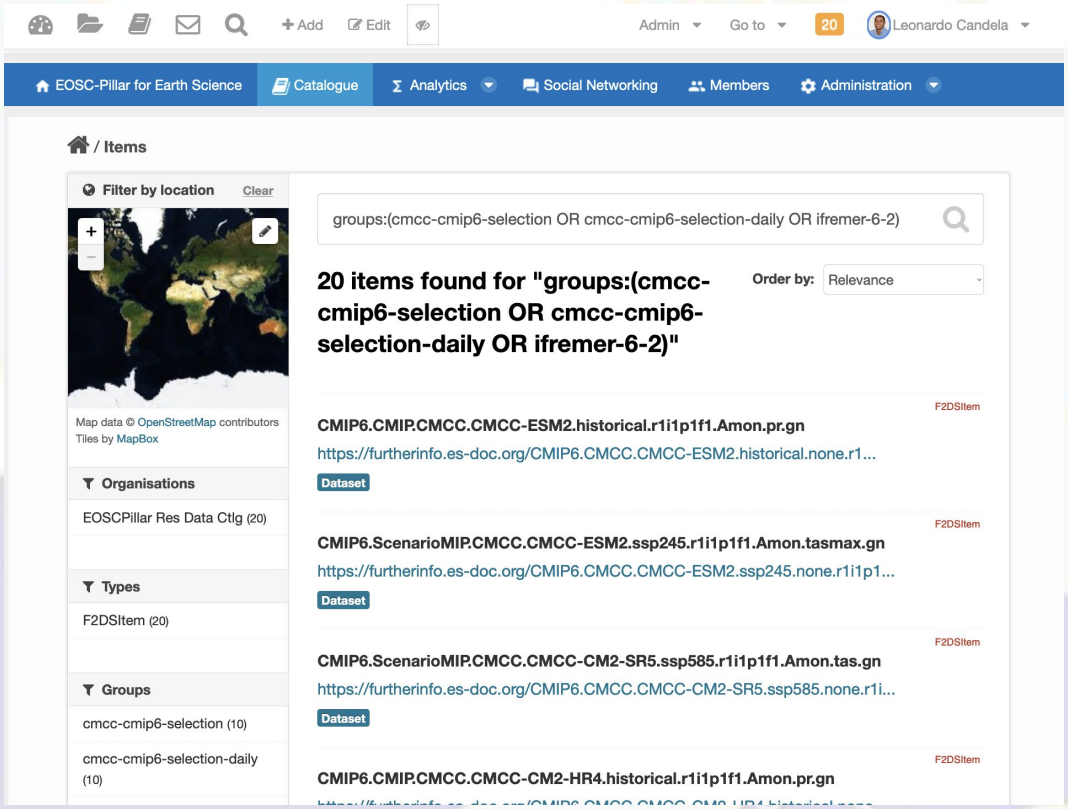

Figure 9 EOSC-Pillar for Earth Science VRE: Catalogue Service

EOSC-Pillar 4 COVID-19 This virtual research environment was created to support the use case on COVID-19 [11]. In particular, the catalogue was there to make it possible to publish data sets and research artefacts stemming from the analysis performed by the tools made available.

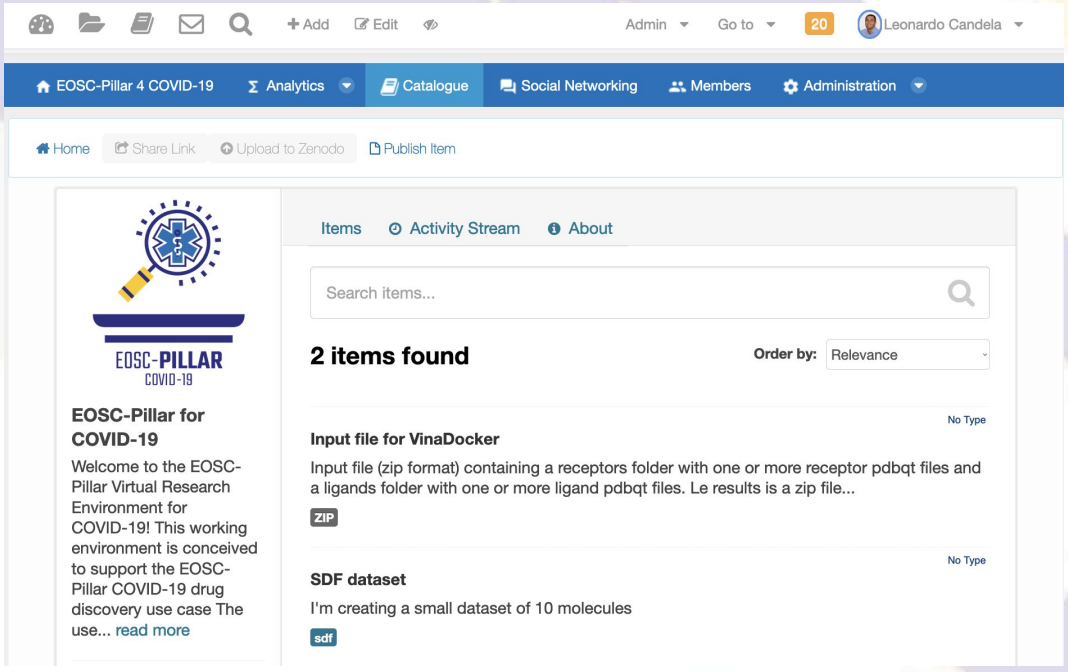

Figure 10 EOSC-Pillar for COVID-19 VRE: Catalogue Service

EOSC-Pillar Research Data Catalogue This virtual research environment was created to realise a working environment where the tools forming the F2DS toolset can be showcased. The catalogue here is daily populated by harvesting all the contents onboarded into the Metadata Repository.

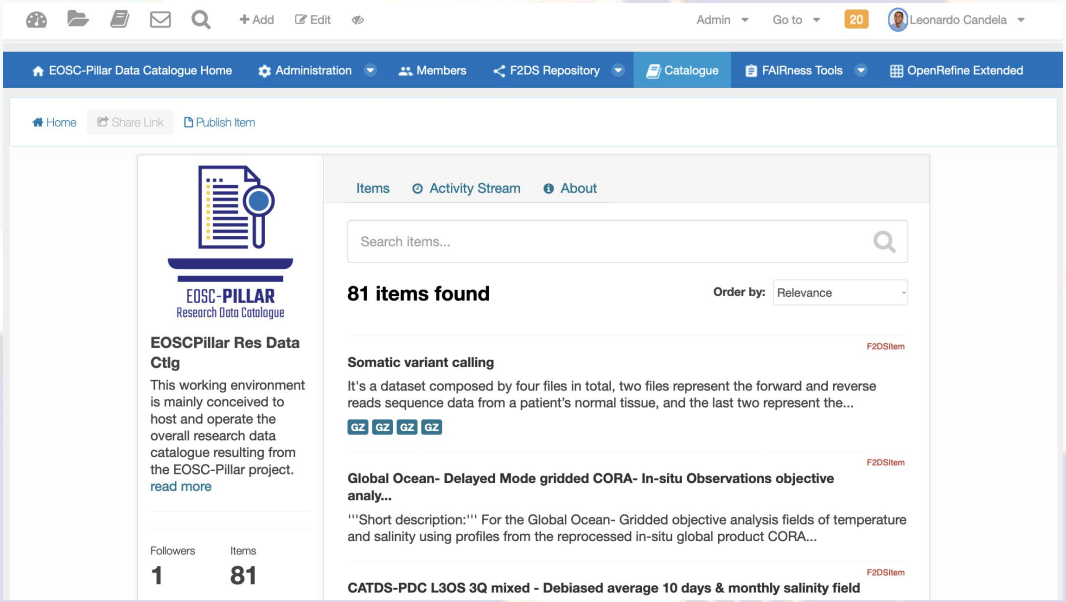

Figure 11 EOSC-Pillar for Research Data Catalogue VRE: Catalogue Service

The mapping governing the transformation of Metadata Repository objects into Catalogue items is documented below.

Table 2 Mapping between Metadata Repository and Catalogue

Metadata Repository (DCAT Class – Property)	Catalogue Field
dcat:Dataset	extra:uri
dcat:Dataset - dct:title	title
dcat:Dataset - dct:description	notes
dcat:Dataset - dct:keyword	tags
dcat:Dataset - dct:theme	extra:theme
dcat:Dataset - dct:identifier	extra:identifier
dcat:Dataset - adms:identifier	extra:alternate_identifier
dcat:Dataset - dct:issued	extra:issued
dcat:Dataset - dct:modified	extra:modified
dcat:Dataset - owl:versionInfo	version
dcat:Dataset - adms:versionNotes	extra:version_notes
dcat:Dataset - dct:language	extra:language
dcat:Dataset - dct:landingPage	url

dcat:Dataset - dct:accrualPeriodicity	extra:frequency
dcat:Dataset - dct:conformsTo	extra:conforms_to
dcat:Dataset - dct:accessRights	extra:access_rights
dcat:Dataset - foaf:page	extra:documentation
dcat:Dataset - dct:provenance	extra:provenance
dcat:Dataset - dct:type	extra:dcat_type
dcat:Dataset - dct:hasVersion	extra:has_version
dcat:Dataset - dct:isVersionOf	extra:is_version_of
dcat:Dataset - dct:source	extra:source
dcat:Dataset - adms:sample	extra:sample
dcat:Dataset - dct:spatial	extra:spatial_uri
dcat:Dataset - dct:temporal	extra:temporal_start + extra_temporal_end
dcat:Dataset - dct:temporalResolution	extra:temporal_resolution
dcat:Dataset - dct:spatialResolutionMeters	extra:spatial_resolution_in_meters
dcat:Dataset - dct:isReferencedBy	extra:is_referenced_by
dcat:Dataset - dct:publisher	extra:publisher_uri
foaf:Agent - foaf:name	extra:publisher_name
foaf:Agent - foaf:mbox	extra:publisher_email
foaf:Agent - foaf:homepage	extra:publisher_url
foaf:Agent - dct:type	extra:publisher_type
dcat:Dataset - dcat:contactPoint	extra:contact_uri
vcard:Kind - vcard:fn	extra:contact_name
vcard:Kind - vcard:hasEmail	extra:contact_email
dcat:Dataset - dcat:distribution	resources
dcat:Distribution	resource:uri
dcat:Distribution – dct:title	resource:name
dcat:Distribution – dcat:accessURL	resource:access_url
dcat:Distribution – dcat:downloadURL	resource:download_url
dcat:Distribution – dct:description	resource:description
dcat:Distribution – dcat:mediaType	resource:mimetype
dcat:Distribution – dct:format	resource:format
dcat:Distribution – dct:license	resource:license

dcat:Distribution – adms:status	resource:status
dcat:Distribution – dcat:byteSize	resource:size
dcat:Distribution – dct:issued	resource:issued
dcat:Distribution – dct:modified	resource:modified
dcat:Distribution – dct:rights	resource:rights
dcat:Distribution – foaf:page	resource:documentation
dcat:Distribution – dct:language	resource:language
dcat:Distribution – dct:conformsTo	resource:conforms_to
dcat:Distribution – dcatap:availability	resource:availability
dcat:Distribution – dcat:compressFormat	resource:compress_format
dcat:Distribution – dcat:packageFormat	resource:package_format
spdx:Checksum – spdx:checksumValue	resource:hash
spdx:Checksum – spdx:algorithm	resource:hash_algorithm

2.4 Operation Activity Indicators

In order to quantify the operation activity related to the F2DS toolset the following indicators were identified and collected.

Table 3 F2DS Operation Activity Indicators

Indicator	Description
Aggregated data sources	The total number of distinct data sources integrated into the F2DS
Aggregated datasets	The total number of datasets integrated into the F2DS
Metadata Repository Users	The number of users registered in the Metadata Repository to onboard data sources
Catalogue Accesses	The number of accesses (working sessions) to the catalogue service by the GUI
Catalogue Item Metadata Views	The number of accesses to the single catalogue item
Catalogue Item Resource Views	The number of accesses to a catalogue item resource
Catalogue search / browse tasks	The number of search / browse tasks done by the catalogue service
Number of VREs equipped with F2DS facilities	The number of virtual Research Environments provided with F2DS facilities

At the time of writing this deliverable (November 2022) the value of the above indicators are documented in Table 4.

Table 4 F2DS Operation Activity Indicators up to November 2022

Indicator	Value	Explanation
Aggregated data sources	6	The data sources CMIP, CORA, SMOS, ARGO, and Data INRAE have been integrated by using the Metadata Repository service.
Aggregated datasets	81	This is the number of items made available by the Research Data Catalogue VRE[10]. This VRE has been created to experiment and showcase how a collaborative working environment offering F2DS services may look like. The environment offers both the GUI of F2DS facilities (Metadata Repository and Catalogue) as well as other services including an OpenRefine instance[11] equipped with FAIR plug-ins and a link to the FAIR Evaluation Services[12] (operated by FAIR sharing) that could be used to assess the FAIRness of datasets. The set of services this VRE offers might be extended in the future.
Metadata Repository Users	34	This is the number of users registering into the Metadata Repository to onboard data sources.
Catalogue Accesses	1373	This is the total amount of access to the overall EOSC-Pillar catalogue[13] in the period January 2021 – November 2022. Figure 12 displays the number of catalogue access per month.
Catalogue search / browse tasks	563	This is the total amount of search and browse operations performed over the whole EOSC-Pillar catalogue in the period January 2021 – November 2022. Figure 13 displays the search and browser operations per month.
Catalogue Item Metadata Views	4272	This is the total amount of views to catalogue items metadata over the whole EOSC-Pillar catalogue in the period in the period January 2021 – November 2022. Figure 14 displays the number of views to catalogue items metadata per month.
Catalogue Item	950	This is the total amount of views to catalogue item resource over the whole EOSC-Pillar catalogue in the

[10] The Research Data catalogue VRE is available at https://eosc-pillar.d4science.org/web/eoscpillarresdatactlg

[11] OpenRefine is an open-source technology for data cleanup and transformation to other formats, an activity commonly known as data wrangling. It is available in the Research Data Catalogue VRE at https://eosc-pillar.d4science.org/web/eoscpillarresdatactlg

[12] FAIR Evaluation Tools available in the Research Data Catalogue VRE at https://eosc-pillar.d4science.org/web/eoscpillarresdatactlg are: FAIR-Aware, FAIR Evaluation Services, and F-UJI. These tools are operated by third-party providers.

[13] The overall EOSC-Pillar Catalogue is available at https://eosc-pillar.d4science.org/catalogue-eoscpillar

Resources Views		period in the period January 2021 – November 2022. Figure 15 displays the number of views to catalogue item resource per month.
Number of VREs equipped with F2DS facilities	4	F2DS facilities have been integrated into the following environments: (i) the Research Data environment, i.e. the environment created to showcase F2DS facilities, (ii) the EOSC-Pillar 4 Agrifood, i.e. the environment stemming from use case 6.3, (iii) the EOSC-Pillar 4 Earth Science, i.e. the environment stemming from use case 6.1 and 6.2, and (iv) the EOSC-Pillar 4 COVID-19, i.e. the environments supporting the use case on drug discovery.

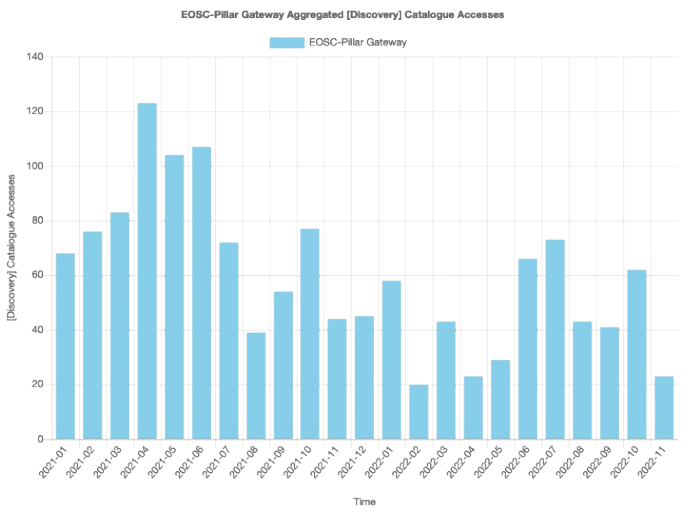

Figure 12 Catalogue Accesses

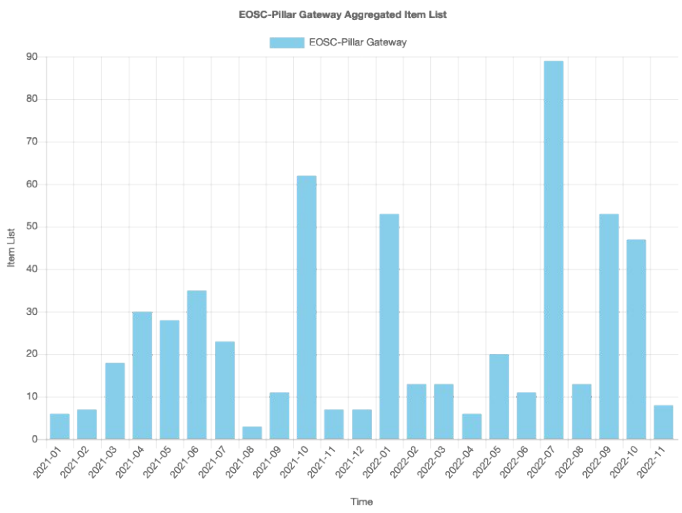

Figure 13 Catalogue Search & Browse

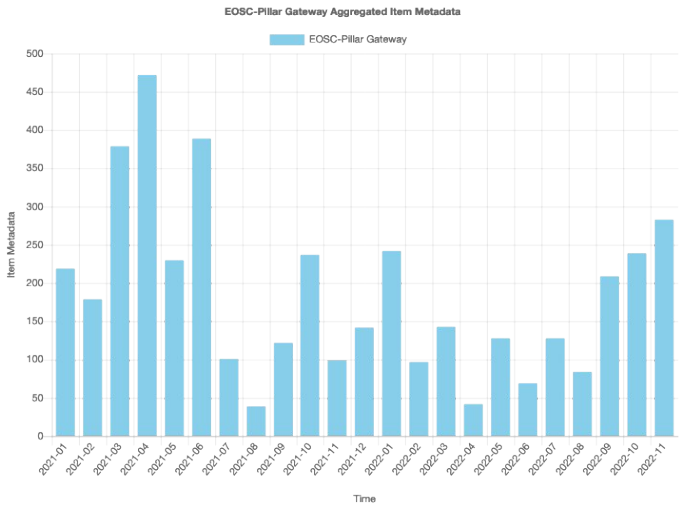

Figure 14 Catalogue Metadata Views

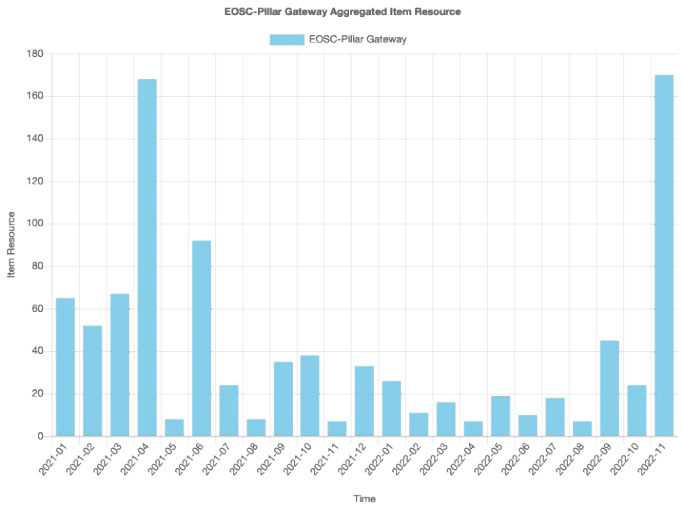

Figure 15 Catalogue Resource Views

3 Conclusion and Remarks

This deliverable provides an up-to-date brief description of the services and facilities contributing to the EOSC-Pillar F2DS workbench and offers indicators on the exploitation of this technology when dealing with the diverse datasets and data sources of interest to the communities involved in EOSC-Pillar via the use cases developed by WP6.

The EOSC-Pillar F2DS tool set has two interoperating focal points: (i) a ***metadata repository*** aggregating the dataset of interest and offering them via APIs and protocols adhering to FAIR principles, and (ii) a ***data catalogue*** offering search and browse on top of the aggregated datasets as well as the possibility to implement and integrate "views" of the whole data space to be included into virtual research environments.

During the reporting period (January 2021 – November 2022) the following operation activities were performed: (a) a total of 6 data sources of interest have been integrated to showcase the F2DS facilities, (b) a total of 81 diverse datasets resulted from this integration, (c) 4 virtual research environments have been deployed to provide the communities with F2DS service instances. These figures prove that the overall solution is suitable for supporting the development of data spaces compliant with the FAIR principles and suitable for serving diverse communities and scenarios.

4 References

[1] Assante, M. et al. (2019) *Enacting open science by D4Science*. Future Gener. Comput. Syst. 101: 555-563 DOI: 10.1016/j.future.2019.05.063

[2] Assante, M. et al. (2019) *The gCube system: Delivering Virtual Research Environments as-a-Service*. Future Gener. Comput. Syst. 95: 445-453 DOI: 10.1016/j.future.2018.10.035

[3] Candela, L., Cazenave, N., Frosini, L., Le Fran, Y., Mangiacrapa, F. (2021) *FAIR Research Data Management Workbench Operation Report*. EOSC-Pillar Deliverable D5.2. DOI 10.5281/zenodo.5513794

[4] Candela, L., Frosini, L., Mangiacrapa, F., Rouchon, O., Le Franc, Y., Toulemonde, B. (2021) *FAIR Research Data Management Tool Set Update*. EOSC-Pillar Deliverable D5.2. DOI 10.5281/zenodo.5720410

[5] Cazenave, N., Candela, L., Berberi, L., van Wezel, J., Hashibon, A., Le Franc, Y. (2020) *FAIR Research Data Management Tool Set*. EOSC-Pillar Deliverable D5.1 https://repository.eosc-pillar.eu/index.php/s/MBgkN5kexKMbMoq

[6] Collins, S. et al. (2018) *Turning FAIR into reality - Final report and action plan from the European Commission expert group on FAIR data*. DOI 10.2777/1524

[7] Domínguez, J. M., Huschle, T., Toulemonde, B., Claude, A., Sarramia, D., Lombardo, F. Alexandru-Aurelian Todor, A.-A. (2022) *Service architecture providing fast access to the concepts held within semantic artifacts*. EOSC-PIllar Deliverable D5.5 https://repository.eosc-pillar.eu/index.php/s/Mo7La99xpywmFYP

[8] European Commission, Directorate-General for Research and Innovation (2022) *European Research Data Landscape: final report*. Publications Office of the European Union, https://data.europa.eu/doi/10.2777/3648

[9] Alessandro Rizzo, Elda Osmenaj, Fabrizio Antonio, Marco Kulüke, Marine Vernet, Dimitri Szabo, David Douard, Adeline Joffres, Gilles Mathieu, Marwa Belhaj-Salem, Javier Quinteros, Alessandro Bombini, Franco Niccolucci, Lisana Berberi (2022). *Final Report on Use Cases and community involvement*. EOSC-Pillar D6.3 Deliverable (to be submitted, at time of writing)

[10] Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. (2016) *The FAIR Guiding Principles for scientific data management and stewardship*. Sci Data 3, 160018. DOI: 10.1038/sdata.2016.18

[11] de Andres, P., Berberi, L., Breton, V., Candela, L., Franc, A., Frigerio, J.-M., Gaido, L., Mandreoli, P., Lechaudel, D., Lorini, M., Di Fazio, M., Orro, A., Pansanel, J., Pisa, C., Pop, S., Porquet, T., Romier, G., Tangaro, M. A., Vianello, E., van Wezel, J. (2022) *Transnational usage of EOSC-Pillar services*. EOSC-Pillar Deliverable D7.4 (to be submitted, at time of writing)