

D5.2 FAIR Research Data Management Workbench Operation Report

Lead Partner:	CNR
Authors	L. Candela (CNR-ISTI), N. Cazenave (CINES), L. Frosini (CNR-ISTI), Y. Le Fran (CINES), F. Mangiacrapa (CNR-ISTI)
Version:	1.0
Status:	Final
Dissemination Level:	Public
Document Link:	PDF version: https://repository.eosc-pillar.eu/index.php/s/PXjFRDFT4s6NcXw

Deliverable Abstract

EOOSC-Pillar is developing and integrating 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 EOOSC-Pillar toolset enacting the development of the EOOSC-Pillar data space. This first release of this typology of deliverable gives early indicators about the exploitation of the proposed services. The data space will be actually developed in the second period of the project.

COPYRIGHT NOTICE

This work by Parties of the EOSC-Pillar is licensed under a Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>). The EOSC-Pillar project is co-funded by the European Union Horizon 2020 programme under grant number 857650.

DELIVERY SLIP

<i>Date</i>	<i>Name</i>	<i>Partner/Activity</i>	<i>Date</i>
From:	L. Candela	CNR	
Moderated by:	Y. Le Franc	CINES	
Reviewed by:	M. Lorini		
Approved by:			

DOCUMENT LOG

<i>Issue</i>	<i>Date</i>	<i>Comment</i>	<i>Author</i>
v0.1	1 Dec 2020	First draft	L. Candela, N. Cazenave
V0.8	15 Jan 2021	First complete version ready for review	L. Candela
V0.9	27 Jan 2021	Review comments received	M. Lorini
V1.0	1 Feb 2021	Release version produced	L. Candela

TERMINOLOGY

<https://eosc-portal.eu/glossary>

<i>Terminology/Acronym</i>	<i>Definition</i>
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	12
2.4	Operation Activity Indicators	15
3	Conclusion and Remarks	18
4	References.....	19

Table of Figures and Tables

Figure 1. F2DS Overall Architecture	8
Figure 2. F2DS Metadata Repository Data Provisioning GUI.....	9
Figure 3. F2DS Data Catalogue Harvesting GUI.....	10
Figure 4. Data Catalogue Item Model.....	11
Figure 5. F2DS Metadata Repository Search GUI	12
Figure 6. F2DS Data Catalogue GUI	13
Figure 7. Catalogue Accesses	17
Figure 8. Catalogue Search & Browse	17
Figure 9. Catalogue Metadata Views	17
Figure 10. Catalogue Resource Views	17
Table 1. EOSC-Pillar Data Sources of interest.....	13
Table 2. F2DS Operation Activity Indicators.....	15
Table 3. F2DS Operation Activity Indicators up to December 2020.....	15

Executive summary

The goal of the EOSC-Pillar Work Package 5 “The Data layer: establishing FAIR data services at the national and transnational level” is 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 will provide data providers and data consumers with 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 to develop 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 some early 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 under development in WP6. A full development of the F2DS is expected to take place during the second period of the project (i.e., the initial period was dedicated to develop the toolset and demonstrate it). However, during the reporting period the following operation activities were performed: (a) a total of 4 out of 13 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) 3 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” is 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 (Wilkinson et al. 2016).

A lot of technologies, services and solutions have been developed and are being developed to support the implementation of the FAIR principles to some extent. A detailed and comprehensive record of existing solutions is given in EOSC-Pillar Milestone MS25 “Existing tools available for FAIRization identified and classified” (Cozzini, 2021) where the functionalities offered by 29 tools were described. Because of the fact that these existing solutions have been developed in diverse contexts and time frames they are mainly disconnected from each other and require human effort when expected to work together.

WP5 is proposing and developing a solution that 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. This 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 envisaged to implement the EOSC-Pillar Federated FAIR Data Space workbench by highlighting their major features and how these components have been integrated and are expected to work together.

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. Because of this it was decided to follow a *pragmatic and exploratory-oriented approach* for the first release of the tool set (Cazenave et al. 2020). In particular, the EOSC-Pillar F2DS tool set 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 technology¹ 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://ffds.eosc-pillar.eu/>

The Data Catalogue is based on gCube technology (in turn relying on CKAN technology for the catalogue) (Assante et al 2019b). 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. An instance of this components is available at <https://eosc-pillar.d4science.org/>

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>

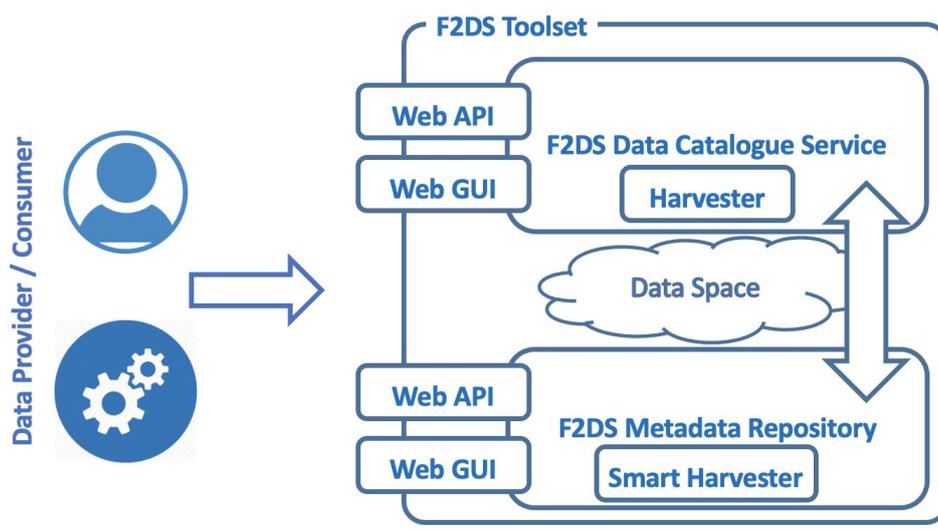


Figure 1. F2DS workbench Overall Architecture

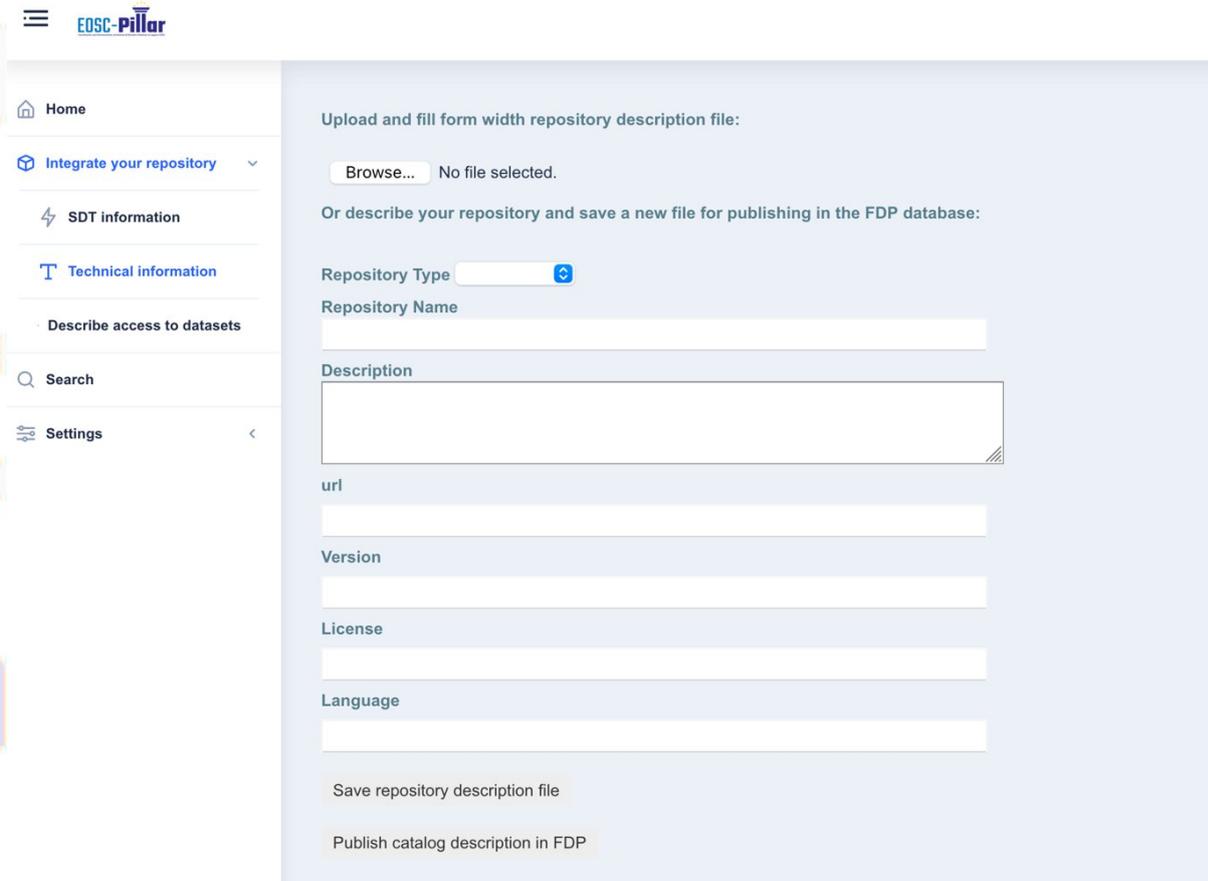
2.1 Supporting the data provision by F2DS facilities

The F2DS toolset 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 4-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. Launch the population of the FAIR Data Point (FDP) with metadata describing the available datasets of the catalogue / data source.



Home

Integrate your repository

SDT information

Technical information

Describe access to datasets

Search

Settings

Upload and fill form with repository description file:

Browse... No file selected.

Or describe your repository and save a new file for publishing in the FDP database:

Repository Type

Repository Name

Description

url

Version

License

Language

Save repository description file

Publish catalog description in FDP

Figure 2. F2DS Metadata Repository Data Provisioning GUI

Regarding the harvesting, the Data Catalogue offers a GUI (see Figure 2) 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)².

² At the time of writing this report, one gFeed harvesting plug-in has been developed to harvest content from OAI-PMH data sources.

[Home](#) / [Harvest Sources](#) / **Create Harvest Source**

Harvest sources

Harvest sources allow importing remote metadata into this catalog. Remote sources can be other catalogs such as other CKAN instances, CSW servers or Web Accessible Folders (WAF) (depending on the actual harvesters enabled for this instance).

URL:
This should include the http:// part of the URL

Title:
URL: `ckan-d-d4s.d4science.org/harvest/<harvest-source>` [Edit](#)

Description:
You can use [Markdown formatting](#) here

Source type: CKAN [?](#)

- CSW Server [?](#)
- Web Accessible Folder (WAF) [?](#)
- Single spatial metadata document [?](#)
- CSW server (GeoNetwork) [?](#)
- Generic DCAT RDF Harvester [?](#)
- DCAT JSON Harvester [?](#)
- OAI-PMH [?](#)

Figure 3. F2DS Data Catalogue Harvesting GUI

Regarding the publishing, the Data Catalogue offers a GUI (see Figure 3) enacting authorised users to register single datasets³ 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 depicted in Figure 4.

³ Actually, any item the user community defines by specifying the typology (i.e., a name and a set of metadata fields).

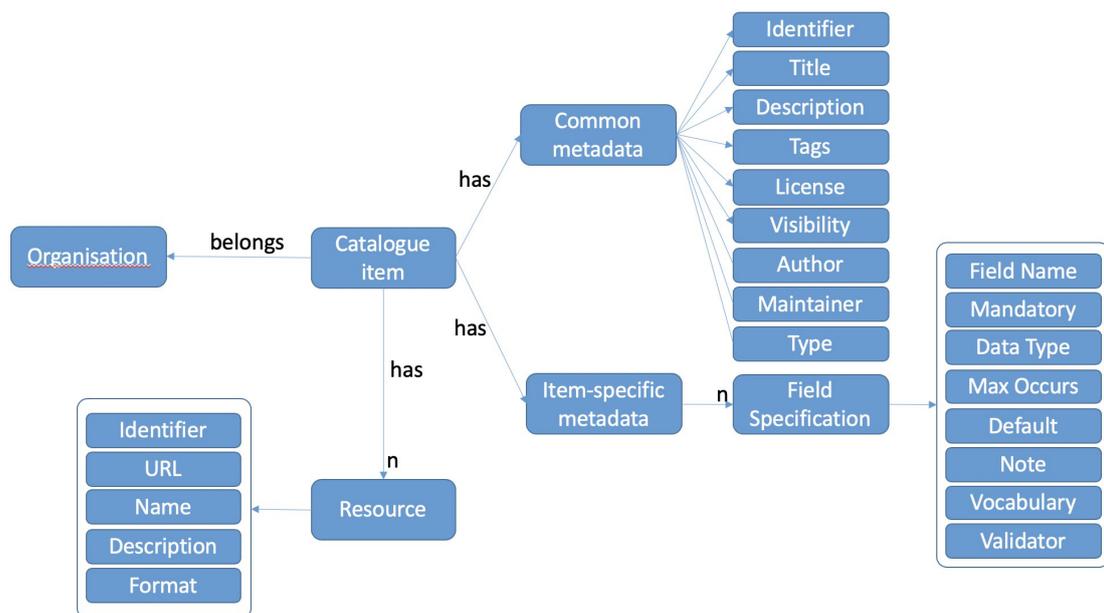


Figure 4. Data Catalogue Item Model

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).

These facilities enact 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 flow 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 API⁴;
- The Metadata Repository make it possible to search the content by SPARQL queries (by relying on the Blazegraph™ DB)⁵;
- The Data Catalogue supports a REST API (by the gCat service)⁶ that by the item collection operations make is 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)⁷.

Regarding the GUIs, the following ones are supported:

- The Metadata Repository offers a keyword-based search on its content (see Figure 4).

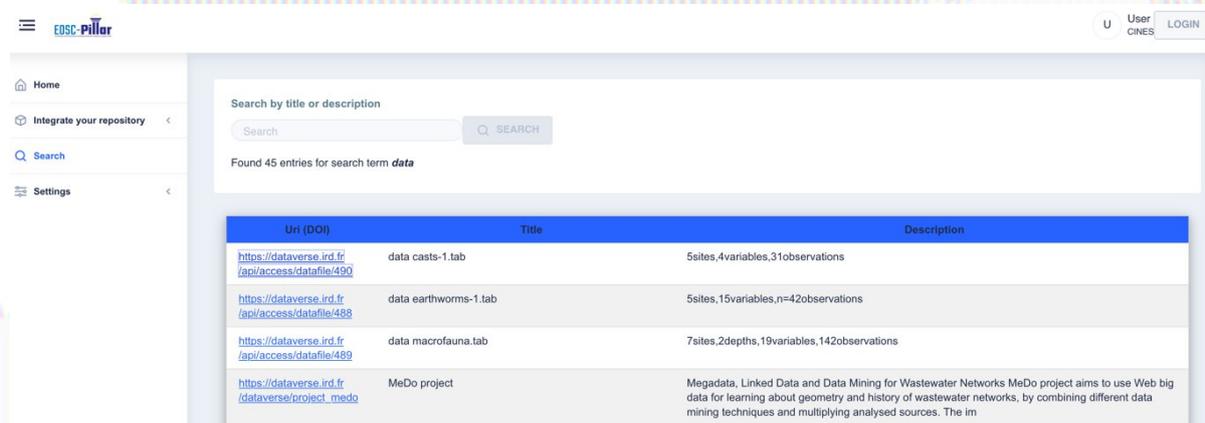


Figure 5. F2DS Metadata Repository Search GUI

- 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 5).

⁴ The swagger based description of the API is available by <http://51.178.69.147/swagger-ui.html>

⁵ F2DS Metadata Repository Blazegraph DB GUI <http://51.210.211.132:8888/blazegraph/#query>

⁶ F2DS Data Catalogue REST API https://wiki.gcube-system.org/gcube/GCat_Service

⁷ 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}>

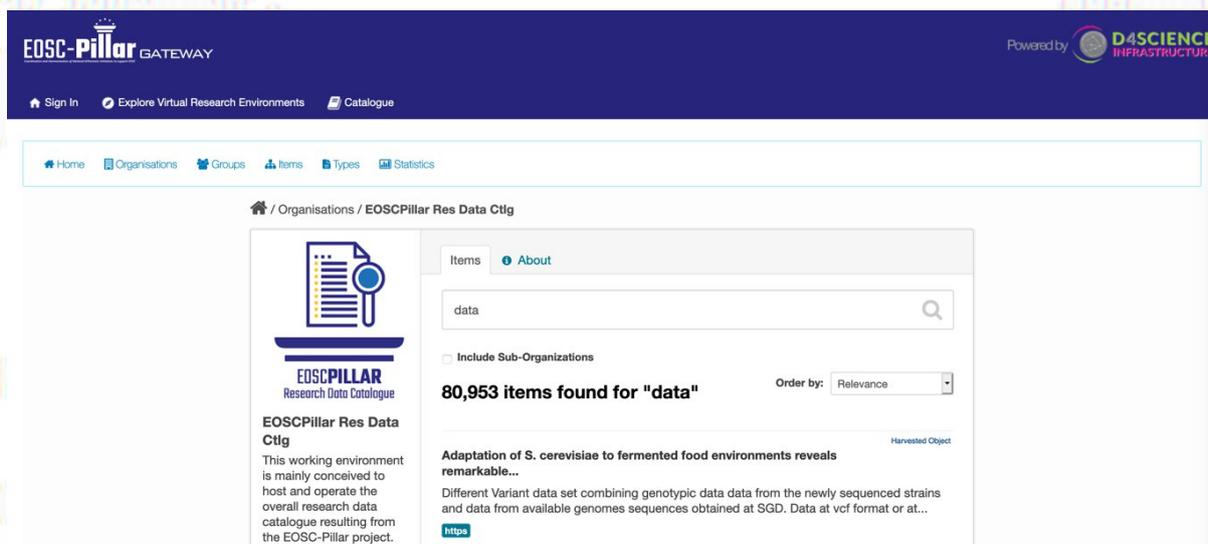


Figure 6. F2DS Data Catalogue GUI

These facilities enact 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 initial set of Repositories and Data Sources to be considered for the development of the EOSC-Pillar Federated FAIR Data Space are those stemming from the project use cases (Rizzo et al. 2020). In particular, the following ones are sources of data of interest for the use cases. For each data source, Table 1 reports the access (Protocols and APIs) as well as the metadata formats.

Table 1. EOSC-Pillar Data Sources of interest

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 ESGF Data Collection	T6.1: "Defining procedures/service to enforce data provenance for thematic"	See https://esgf-data.dkrz.de/search/cmip6-dkrz	REST API	Proprietary

	communities and beyond”			
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
GLOBAL OCEAN GRIDDED L4 SEA SURFACE HEIGHTS AND DERIVED VARIABLES REPROCESSED	T6.2: “Agile FAIR data for environment and earth system communities (ocean, atmosphere, continental surfaces, solid earth)”	See https://resources.marine.copernicus.eu/?option=com_csw&view=details&product_id=SEALEVEL_GLO_PHY_L4_REP_OBSERVATIONS_008_047	CSW, WMS, FTP	ISO-19115

SSED (1993-ONGOING)				
Data INRAE	T6.3: “Integration of data repositories into EOSC based on communities’ approaches”	The Data portal of INRAE	OAI-PMH	DublinCore
Software Heritage	T6.4: “Software source code preservation, reference and access”	Software Heritage is the largest public archive of software source code and its development history	Proprietary REST API	CodeMeta
HAL	T6.5: “FAIR principles in data life-cycles for humanities”	See https://hal.archives-ouvertes.fr/	OAI-PMH	DubliCore
NAKALA	T6.5: “FAIR principles in data life-cycles for humanities”	See https://www.nakala.fr	OAI-PMH	DublinCore
Data INSERM	T6.6: “Exploring reference data through existing computing services for the bioinformatics”	https://dataverse-test.ouvrirlascience.fr/dataverse/inserm/ (Under development)	n.a.	n.a.
GEOFON Seismological Archive	T6.7: “Suitable data formats for seismological big data provisioning via web services”	See https://geofon.gfz-potsdam.de/	FDSN API	StationXML
ARIADNE Catalogue	T6.9: “Integrating heterogeneous data on cultural heritage”	A portal of archaeological resources developed by the ARIADNE community. See http://portal.ariadne-infrastructure.eu/		

This table highlights the great variety of datasets the F2DS is called to deal with. The willingness to make this set of very diverse items seamlessly searchable and accessible require an intensive dialogue with the communities and data providers. In fact, the dialogue aiming at reshuffling the datasets so as to make them compliant with the FAIR principles and to benefit from the F2DS toolset

is planned to be performed more deeply in the second period. The first period was mainly dedicated to start collecting contents from some of these data sources and showcase how the collected content manifests in the F2DS.

2.4 Operation Activity Indicators

In order to quantify the operation activity related to the F2DS toolset the following indicators are collected.

Table 2. 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
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 (December 2020) the indicators in Table 3 have been collected.

Table 3. F2DS Operation Activity Indicators up to December 2020

Indicator	Value	Explanation
Aggregated data sources	4	The data sources CORA, SMOS, ARGO, and Data INRAE have been integrated by using the Metadata Repository service. Regarding CORA, SMOS, and ARGO the discussion is ongoing especially for what regards the concept of dataset.
Aggregated data sources	4	The data sources CORA, SMOS, ARGO, and Data INRAE have been integrated by using the Data Catalogue service. CORA, SMOS, and ARGO have been integrated each as a single entry.
Aggregated	80,954	This is the number of items made available by the Research Data Catalogue VRE ⁸ . This VRE has been

⁸ The Research Data catalogue VRE is available at <https://eosc-pillar.d4science.org/group/eoscpillarresdatactlg>

datasets		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 equipped with FAIR plug-ins and a link to the FAIR Evaluation Services (operated by FAIR sharing) that could be used to assess the FAIRness of datasets. The set of services this VRE offer might be extended in the future.
Catalogue Accesses	768	This is the total amount of access to the overall EOSC-Pillar catalogue ⁹ in the period January 2020 – December 2020. The overall catalogue contains the datasets resulting from the F2DS plus few items (85 out of 81k) that have been created to serve the needs of other communities and cases. A graph reporting the per month figures is in Figure 7.
Catalogue Item Metadata Views	98	This is the total amount of search and browse operations performed over the whole EOSC-Pillar catalogue in the period in the period January 2020 – December 2020. A graph reporting the per month figures is in Figure 8.
Catalogue Item Resource Views	1101	This is the total amount of views to catalogue items over the whole EOSC-Pillar catalogue in the period in the period January 2020 – December 2020. A graph reporting the per month figures is in Figure 9.
Catalogue search / browse tasks	146	This is the total amount of views to catalogue item resources over the whole EOSC-Pillar catalogue in the period in the period January 2020 – December 2020. A graph reporting the per month figures is in Figure 10.
Number of VREs equipped with F2DS facilities	3	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 under development environment stemming from use case 6.3, and (iii) the EOSC-Pillar 4 COVID-19, i.e. the under development environment for supporting the use case on drug discovery.

⁹ The overall EOSC-Pillar Catalogue is available at <https://eosc-pillar.d4science.org/catalogue-eoscpillar>

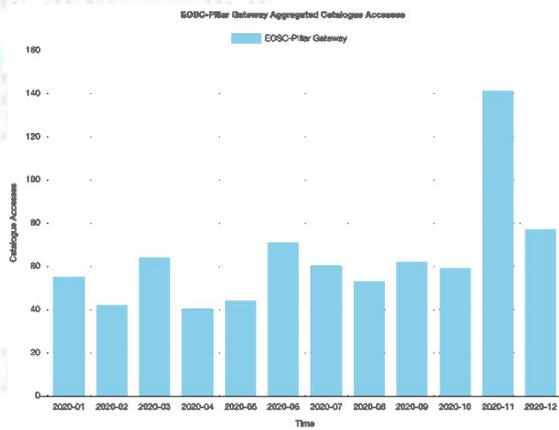


Figure 7. Catalogue Accesses

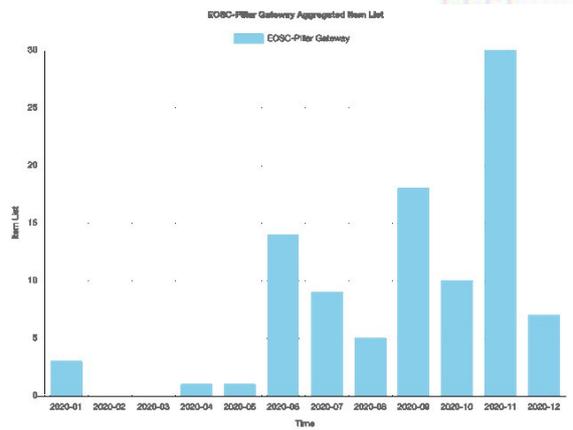


Figure 8. Catalogue Search & Browse

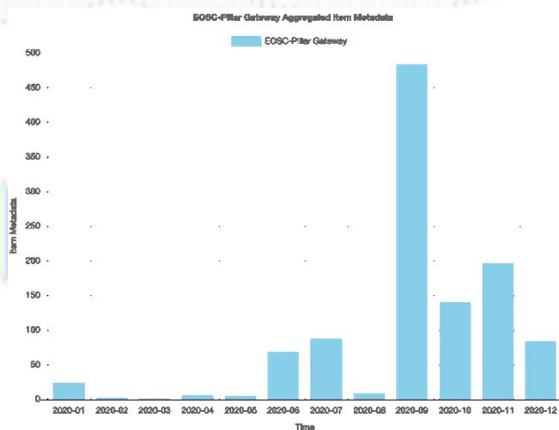


Figure 9. Catalogue Metadata Views

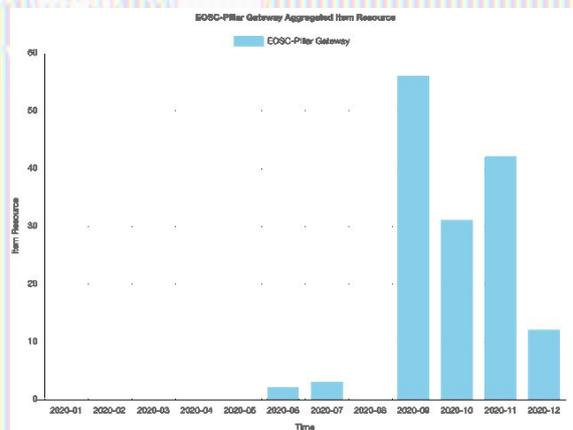


Figure 10. Catalogue Resource Views

These indicators confirm that the F2DS is still in a development stage. We do expect that the figures will grow a lot during the next period, i.e., once more data sources will be onboarded and the communities will start exploiting the aggregated contents and facilities to better support their use cases.

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 gives some early 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 under development in WP6.

The EOSC-Pillar F2DS tool set 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. Additional services will be added to this initial offering. In particular services helping to assess the FAIRness of datasets and services helping to enrich datasets by using ontologies are under consideration.

During the reporting period the following operation activities was performed: (a) a total of 4 out of 13 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) 3 virtual research environments have been deployed to provide the communities with F2DS service instances. These figures highlight how the operation activity of the period was mainly dedicated to showcase the capabilities of the under-development technology. A full development of the F2DS will take place during the second period of the project when the implementation of the use cases will be completed and the communities will start exploiting the resulting services and data.

4 References

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

Assante, M. et al. (2019b) 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](https://doi.org/10.1016/j.future.2018.10.035)

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>

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](https://doi.org/10.2777/1524)

Cozzini (2021) Existing tools available for FAIRization identified and classified. EOSC-Pillar Milestone MS25 <https://repository.eosc-pillar.eu/index.php/s/DYirN9oX68xZbmY>

Rizzo, A., Tanlongo, F., Galeazzi, F. & Pierkot, C. (2020). D6.1 State of the Art and Community Needs Report from Use Cases. EOSC-Pillar D6.1 Deliverable DOI [10.5281/zenodo.3894676](https://doi.org/10.5281/zenodo.3894676)

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](https://doi.org/10.1038/sdata.2016.18)