EGI-Engage

Open Data Platform First Prototype

D4.7

Date	06 December 2016
Activity	WP4
Lead Partner	CYFRONET
Document Status	FINAL
Document Link	https://documents.egi.eu/document/2976

Abstract

This document presents the work related to the preparation and release of the first public prototype of the EGI Open Data Platform (ODP) solution provided by EGI-Engage. The EGI ODP is new data management solution enabling users to access in a transparent manner high-performance storage resources of federated storage providers with heterogeneous storage platforms. The EGI ODP allows users to manage and access their data from anywhere in an easy and secure manner, and supports users with custom features for open data publishing, discovery and access. The EGI ODP is also the basis for the EGI DataHub offering.

The EGI-Engage project is co-funded by the European Union (EU) Horizon 2020 program under Grant number 654142 http://go.egi.eu/eng

COPYRIGHT NOTICE

DELIVERY SLIP

	Name	*Partner/Activity*	*Date*
From:	Bartosz Kryza	ACC Cyfronet AGH	18/11/2016
Moderated by:	Malgorzata Krakowian	EGI Foundation/NA1	
Reviewed by	Tiziana Ferrari Peter Solagna Kostas Koumantaros	EGI Foundation/NA1 EGI Foundation/SA1 GRNET/JRA1	08/11/2016
Approved by:	AMB and PMB		6/12/2016

DOCUMENT LOG

Issue	*Date*	*Comment*	*Author/Partner*
v.1	8/11/2016	First Draft	Bartosz Kryza, Lukasz Dutka/AGH
v.2	8/11/2016	Suggestions for improvement and corrections	Matthew Viljoen/EGI Foundation
v.4	10/11/2016	Corrections after suggestions	Bartosz Kryza/AGH
v.5	11/11/2016	Improvements after internal review	Matthew Viljoen/EGI Foundation
v.6	14/11/2016	Final corrections after internal review	Bartosz Kryza/AGH, Matthew Viljoen/EGI
v.7	18/11/2016	Added exploitation and dissemination table	Bartosz Kryza/AGH
v.1.0	1/12/2016	Final version after AMB/PMB comments	Bartosz Kryza/AGH, Matthew Viljoen/EGI

TERMINOLOGY

A complete project glossary and acronyms are provided at the following pages:

- https://wiki.egi.eu/wiki/Glossary
- https://wiki.egi.eu/wiki/Acronyms

Contents

Executive summary

This document presents the status of the first prototype of the EGI Open Data Platform (ODP) solution. A brief overview of the EGI ODP architecture aspects related to open data is discussed along with current status of implementation of requirements identified as part of initial requirement analysis in *M4.1 Open Data Platform: Requirements and Implementation Plans*. Furthermore an example user workflow related to open data publishing and accessing is presented based on functionality already available in the Open Data Platform prototype.

The prototype has been already deployed as the basis for EGI DataHub serving over 10 TB of Sentinel-2 data products to users who can be authenticated using EGI OpenID Connect protocol.

Future work is discussed in terms of further improvements of the EGI ODP based on the requirements as well as from the perspective of integrating the EGI ODP service providers with EGI operations.

1 Introduction

The EGI Open Data Platform (ODP) is a new technology solution for federated data management with inherent support for sharing and accessing Open Data.

The overall goal of the EGI ODP within EGI is to extend the current FedCloud offering for user communities with advanced mechanisms for data management, high performance data transfers, direct POSIX protocol for immediate data access without pre-staging and advanced metadata capabilities. The EGI ODP can be deployed by Virtual Organizations, who would like to have a high performance data management solution for experiment or research data management, as well as to allow federation with other communities for sharing data and running advanced research projects.

The EGI ODP is focused on collaboration between users, enabling seamless and secure data sharing based on access tokens, which can be used for fine-grained authentication and authorization control, sharing data with other users and requesting storage support.

Furthermore, the EGI ODP supports open data access communities by providing implementation of interfaces for open data publishing and access such as the management and registration of handle based identifiers management as well as OAI-PMH[1] protocol for metadata harvesting.

Based on the EGI ODP, EGI has recently deployed a prototype service called EGI DataHub, which aims at providing open science community high performance access to large reference open data sets.

The Onedata[2] distributed data management platform, on which EGI Open Data Platform is based, is developed as an open source project under Apache 2.0 license. It is currently part of the Polish National Grid infrastructure and is being further development as part of the EU INDIGO-DataCloud project. Within the framework of EGI-Engage project, open data related features are built and integrated into Onedata in order to provide a unified solution for large-scale research data management with supporting open science requirements.

[1] https://www.openarchives.org/pmh/

[2] https://onedata.org

2 EGI Open Data Platform architecture

EGI Open Data Platform reflects the distributed nature of federate storage providers in infrastructures, where on the one hand users need single-sign on authentication and authorization mechanism while administrators need require full control over who can access their storage resources and in what capacity. EGI ODP builds on top of Onedata data management systems and adds open science specific functionalities and interfaces.

2.1 User interfaces

The EGI ODP provides several interfaces both for users as well as for integration with other services:

- **GUI** – web-based Graphical User Interface with responsive design which can be used for basic data management tasks,
- **REST** – comprehensive REST API which allows for integration with 3rd party applications and community portals,
- **CDMI**[3] – SNIA standard for Cloud data management systems allowing object-based access to data,
- **POSIX** – native file system protocol for many operating systems allowing to directly access distributed data sets from local desktops or Virtual Machines or containers running in the Cloud without the need for prestaging,
- **OAI-PMH**[4] - Open Data Platform instances can be easily integrated with Open Data content aggregators such as OpenAIRE, as all data sets published via ODP are exposed through it’s OAI-PMH Data Provider protocol,
- **HTTP** – furthermore, datasets can be shared using simple HTTP protocol for users who do have means for accessing or previewing the datasets.

[3] Cloud Data Management Interface - http://www.snia.org/cdmi

[4] Open Archive Initiative Protocol for Metadata Harvesting - https://www.openarchives.org/pmh/

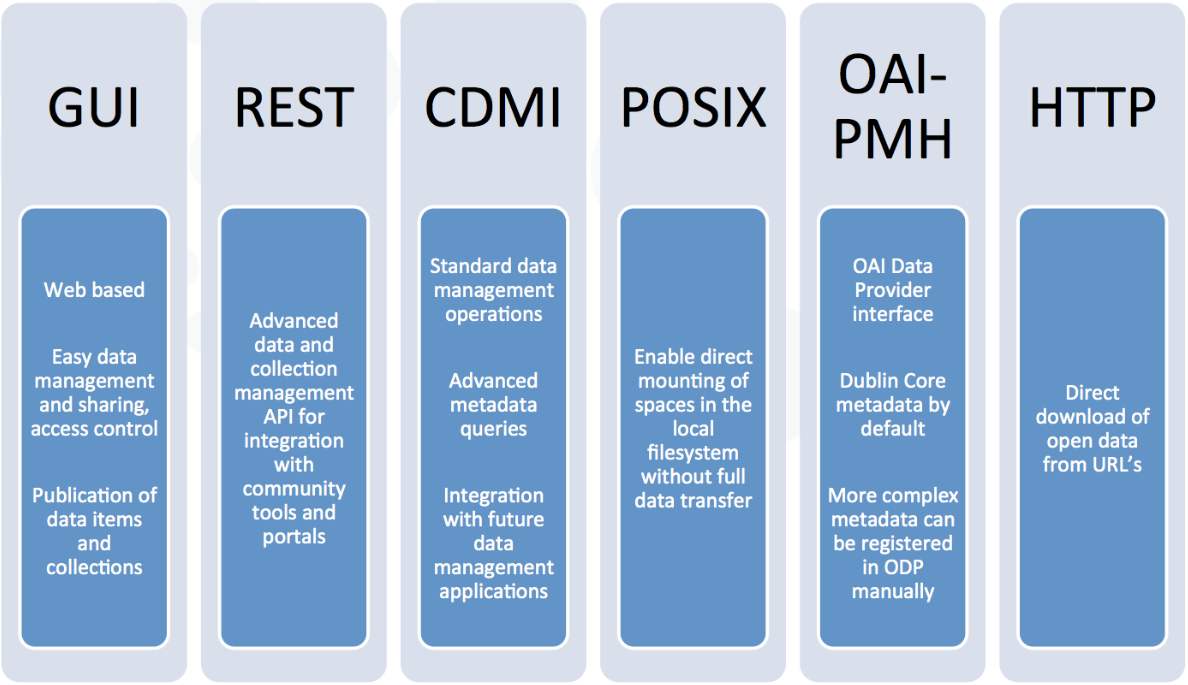

GUI
Web based
Easy data management and sharing, access control
Publication of data items and collections
REST
Advanced data and collection management API for integration with community tools and portals
CDMI
Standard data management operations
Advanced metadata queries
Integration with future data management applications
POSIX
Enable direct mounting of spaces in the local filesystem without full data transfer
OAI-PMH
OAI Data Provider interface
Dublin Core metadata by default
More complex metadata can be registered in ODP manually
HTTP
Direct download of open data from URL's

Figure 1 Main interfaces of Open Data Platform

2.2 Architecture

The EGI ODP architecture reflects the architecture of its underlying technology, i.e. Onedata platform, and is comprised of 2 main services:

- Onezone – each community, user group or federation can deploy their own Onezone service and use as main authentication, authorization and data discovery point for users and storage providers,
- Oneprovider – storage providers can deploy one or more instances of Oneprovider service and by registering it in selected Onezone service, give their users transparent access to storage resources, while keeping the complete control over storage access terms (e.g. quotas) in hands of local administrators.

User authentication to the Onezone service is possible using OpenID Connect (including EGI OIDC[5]) as well as basic authentication based on usernames and passwords. After authenticating to the Onezone service, users are able to generate personal access tokens, which can be used for authorization of all data access requests, among all storage providers.

[5] https://aai.egi.eu/oidc

With respect to storage administration, Oneprovider services currently support several backends including POSIX storage, Ceph, Amazon S3 and OpenStack Swift, which gives administrators big flexibility while deploying Oneprovider on their premises.

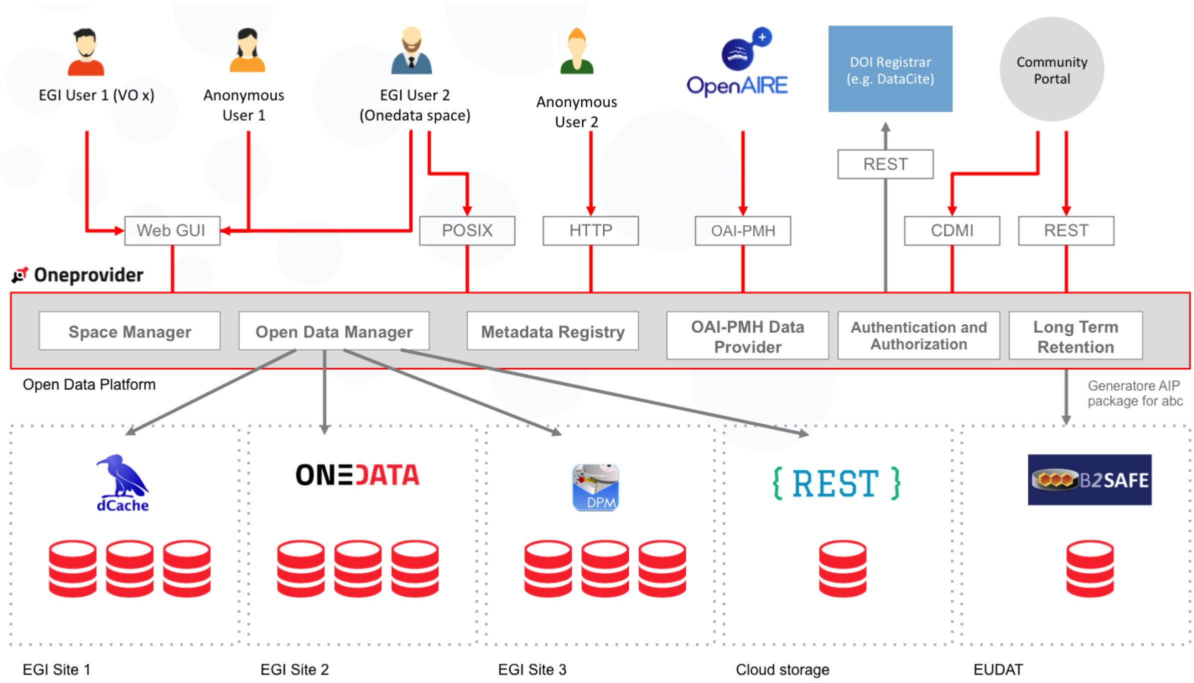


Figure 2 Overall architecture of the Open Data Platform

With respect to Open Data requirements, the following functionality is available from Onezone or Oneprovider services directly:

Onezone

- Authentication
- DOI and PID registration and resolution
- OAI-PMH Data Provider interface
- Web user interface

Oneprovider

- Data access (downloading and POSIX access)
- Metadata queries

2.3 Status of requirements implementations

As part of the first EGI ODP milestone, a review and analysis of requirements coming from EGI-Engage communities was prepared and compiled within the *M4.1 Open Data Platform: Requirements and Implementation Plans*[6]*.* The table below summarizes the main requirements identified within this document and their current status.

#	Requirement name	Implementation status
1	Publication of open research data based on policies	Publication of open data sets is currently supported along with automatic registration of identifiers such as DOIs or PIDs, as well as publication of OAI-PMH metadata records via the OAI-PMH Data Provider interface. Public open data sets can be downloaded by users, even without an ODP account based on a public URL or handle. **Status: in progress** **Future work and timeline.** Future work will include conditional publication based on embargos (time, IP range, etc.).
2	Make large data sets available without transferring them completely	This requirement is fulfilled now completely by means of Onedata virtual filesystem functionality allowing users to directly mount data sets using POSIX protocol. **Status: complete**
3	Enabling complex metadata queries	Metadata in ODP can be defined using either simple key-value pairs as well as arbitrary JSON or RDF documents. Currently metadata queries can be performed by defined index functions on key-value or JSON metadata. **Status: in progress** **Future work and timeline.** Future work will include the possibility of attaching external metadata backends, such as triple stores for RDF metadata.
4	Integration of the open data access data management with community portals	Integration with community portals as well as other services is now possible via ODP comprehensive and fully documented REST API, which gives access to all functionality of Open Data Platform. Authentication

[6] https://documents.egi.eu/public/ShowDocument?docid=2547

		between user portals and EGI ODP can be achieved using OpenID Connect and authorization using authorization tokens, which can be generated using the REST API or from the Onezone service user interface. **Status: in progress** **Future work and timeline.** Future work includes the actual integration of EGI ODP with selected community portals (e.g. LifeWatch).
5	Data identification, linking and citation	Users can easily assign handles (e.g. DOI or PID) to published data sets directly within the graphical user interface. The EGI ODP provides a flexible mechanism for adding various Handle compatible registrars by deploying a simple REST service which translates API calls between ODP and the registration service and optionally transforms metadata from community schema to that required by registration services (e.g. to DataCite metadata schema[7]). **Status: complete.**
6	Enabling sharing of data between researchers under certain conditions	Users of Open Data Platform can now use the Onedata 'share' feature to easily and securely share datasets (which are not necessarily public) among each other. **Status: complete.**
7	Sharing and accessing data across federations	This requirement implementation is in progress. It will be possible to establish trust between multiple Onezone deployments allowing several communities or infrastructures to allow users to share their data securely across the federations in a transparent manner. For infrastructures, which will not deploy EGI ODP, migration of data will be possible by means of supported transport backend storage types including POSIX, OpenStack Swift, Amazon S3 or Ceph. **Status: In progress.** Future work will include

[7] http://schema.datacite.org/

		the functionality for establishing trust between multiple zones (e.g. communities or infrastructures).
8	Long term data preservation	Long-term preservation will not be directly supported by ODP. **Status: in progress** **Future work and timeline.** Future work will include integration with other infrastructures or services allowing users to store important data sets on cold storage e.g. using EUDAT infrastructure as well as commercial solutions (e.g. Amazon Glacier).
9	Data provenance	In the current prototype, data provenance can be achieved manually by attaching required tracking metadata with the datasets. **Status: in progress** **Future work and timeline.** By the end of EGI-Engage project, the automated generation and recording of data sets evolution (e.g. actions performed by users on the data and metadata, origin of data and its ownership transfer) will be possible.

2.4 Open Data functionality provided by the prototype

The Open Data Functionality is presented in this section based on a typical open data user workflow, presented in the figure below. The workflow steps are explained in the consecutive subsections below.

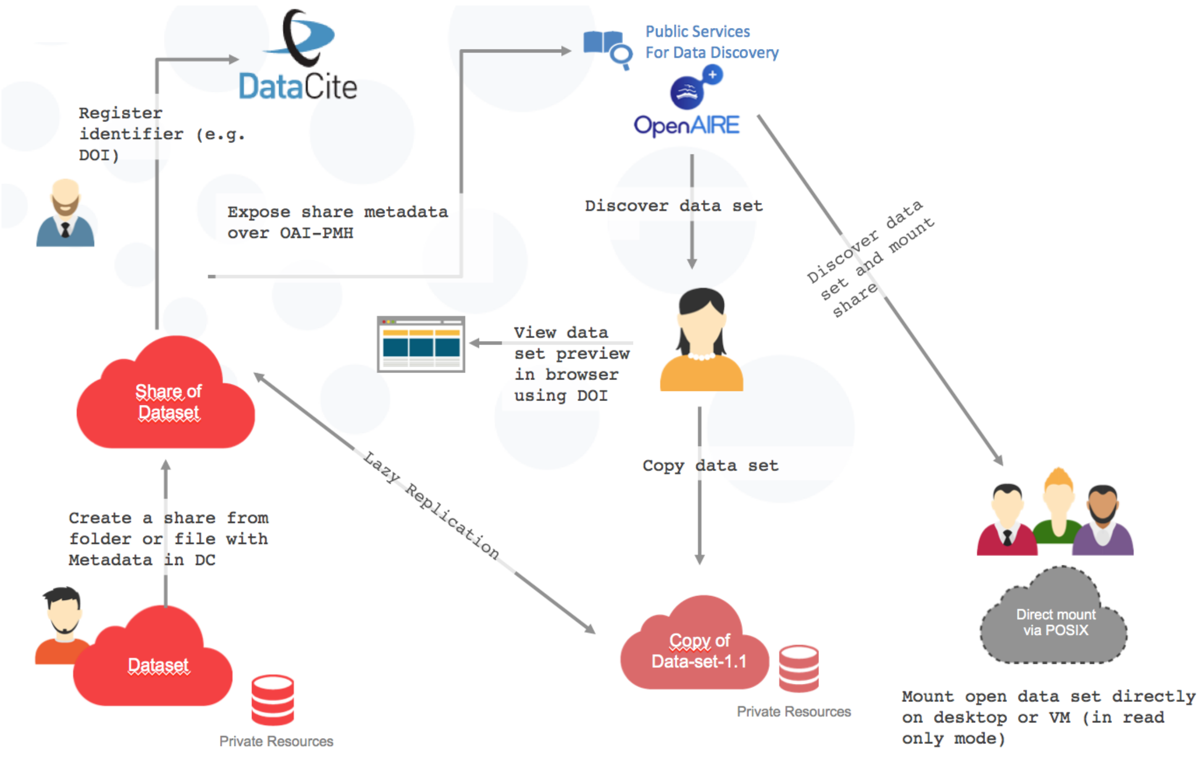
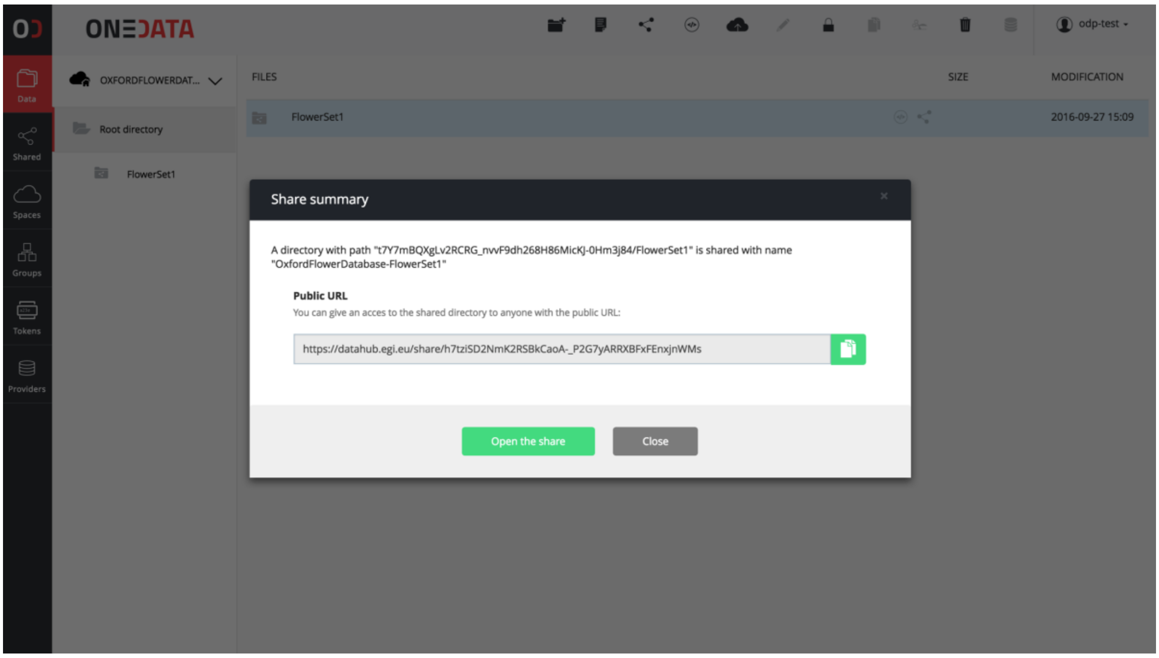

DataCite
Register identifier (e.g. DOI)
Expose share metadata over OAI-PMH
Public Services For Data Discovery
OpenAIRE
Discover data set
Discover data set and mount share
View data set preview in browser using DOI
Share of Dataset
Lazy Replication
Copy data set
Create a share from folder or file with Metadata in DC
Dataset
Private Resources
Copy of Data-set-1.1
Private Resources
Direct mount via POSIX
Mount open data set directly on desktop or VM (in read only mode)

Figure 3 ODP user workflow

2.4.1 Create a share from a folder or a file

The EGI ODP allows for the easy creation of shareable data sets from data managed by users in the data spaces. Entire data spaces as well as single folders or files can be shared. By creating a share from a single resource, the users get a URL, which can be shared with other users. This is the first step in preparing an open data set as only shares can be published as open data sets and assigned Handle based identifiers such as DOIs or PIDs.

At this point the share is not yet an open dataset, as the share can be exchanged securely between users based on the specified Access Control Lists.



Share summary

A directory with path "t7Y7mBQXgLv2RCRG_nvvF9dh268H86MicKJ-0Hm3j84/FlowerSet1" is shared with name "OxfordFlowerDatabase-FlowerSet1"

Public URL

You can give an acces to the shared directory to anyone with the public URL:

https://datahub.egi.eu/share/h7tziSD2NmK2RSBkCaoA-_P2G7yARRXBFxFEnxjnWMs

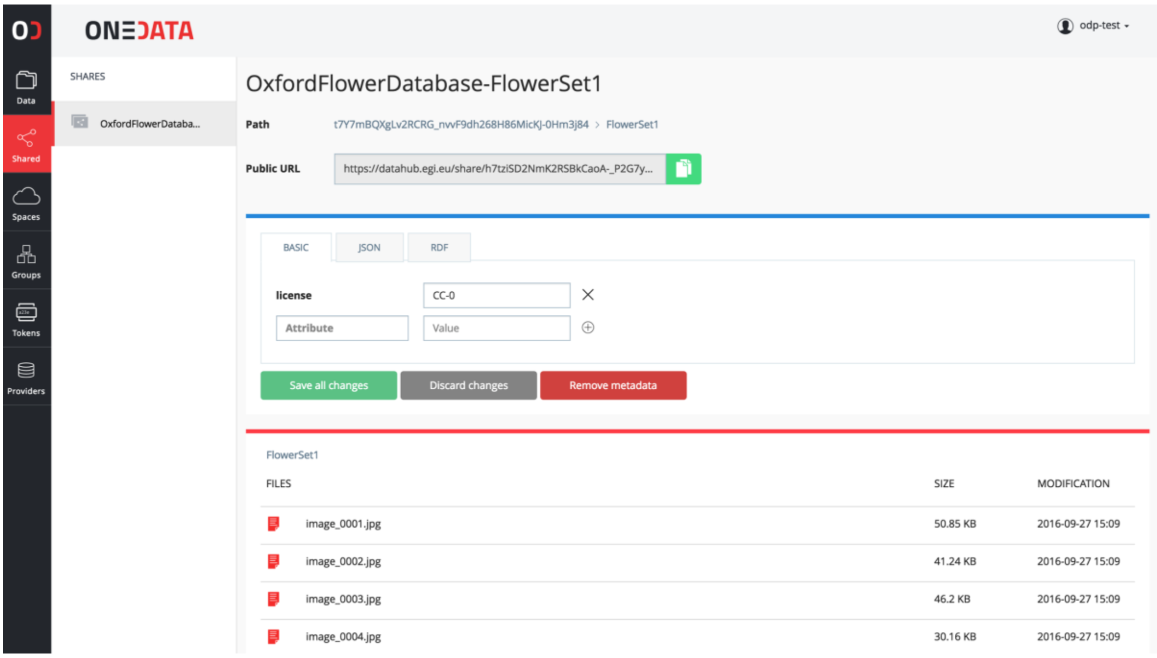
Figure 4 Share creation

2.4.2 Assign metadata

The next step is to make sure that the data set contains appropriate metadata. Metadata in Open Data Platform can be specified in several forms including:

- Key-value pairs
- JSON documents
- RDF documents

Currently indexing and complex queries are available only on key-value and JSON metadata backends.



OxfordFlowerDatabase-FlowerSet1

Path t7Y7mBQXgLv2RCRG_nvvF9dh268H86MicKJ-0Hm3j84 > FlowerSet1

Public URL https://datahub.egi.eu/share/h7tziSD2NmK2RSBkCaoA-_P2G7y...

license CC-0

Attribute Value

FlowerSet1

FILES	SIZE	MODIFICATION
image_0001.jpg	50.85 KB	2016-09-27 15:09
image_0002.jpg	41.24 KB	2016-09-27 15:09
image_0003.jpg	46.2 KB	2016-09-27 15:09
image_0004.jpg	30.16 KB	2016-09-27 15:09

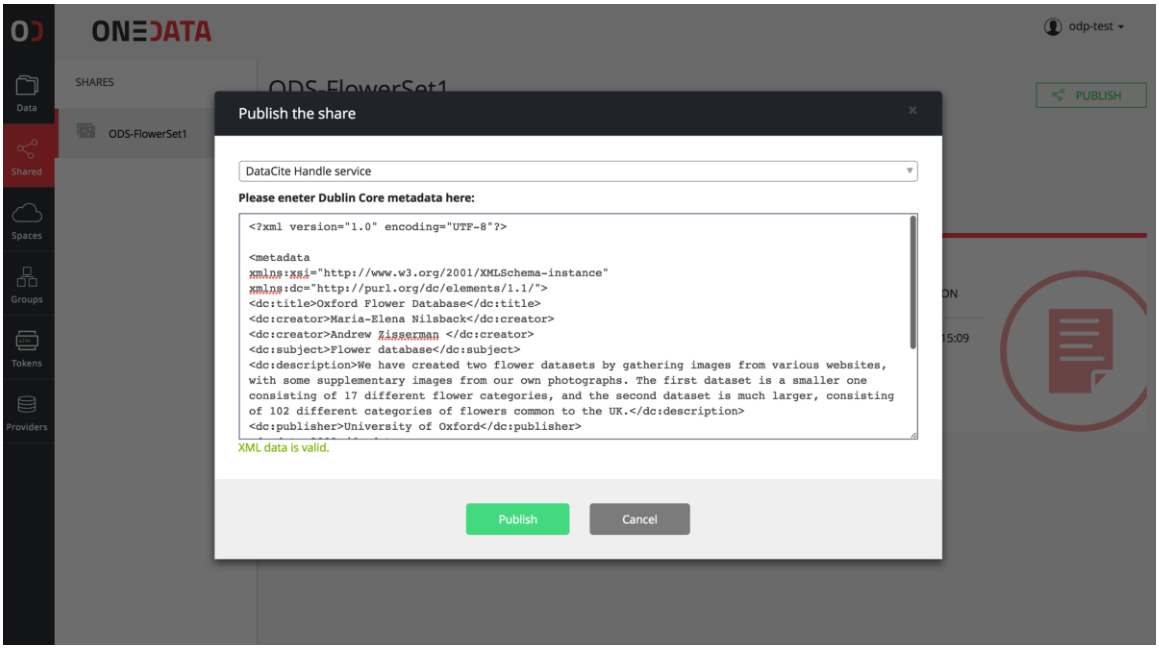
Figure 5 Metadata editor in Open Data Platform

2.4.3 Publishing a share as open data

When a share is created, it can be turned into an open data set with few simple steps. EGI ODP design includes the possibility for integrating various Handle.net based services such as DOI or PID registrars. These services allow for the automatic assignment of identifiers to the data sets before they are published via ODP.

In order to configure new identifier service and prefix, the user community can register their DOI account (e.g. from DataCite). Users with specific roles (such as community managers) can control which users can assign identifiers using this account.

For specific handle services it is possible to add custom metadata as required by each service.



DataCite Handle service

Please eneter Dublin Core metadata here:

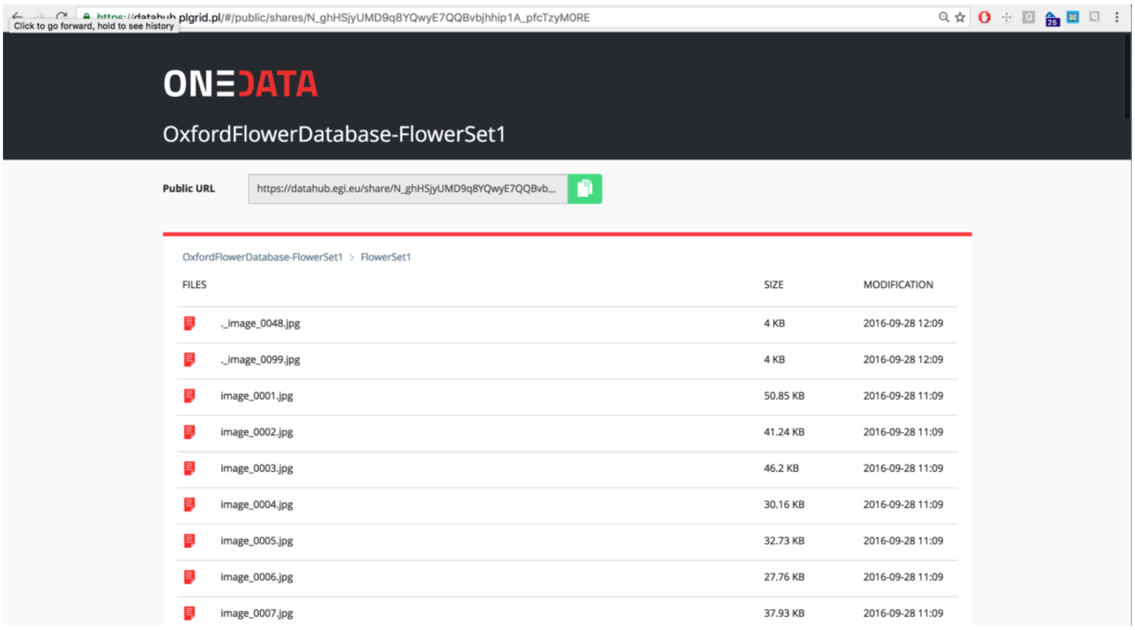
```
<?xml version="1.0" encoding="UTF-8"?>

<metadata
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:dc="http://purl.org/dc/elements/1.1/">
<dc:title>Oxford Flower Database</dc:title>
<dc:creator>Maria-Elena Nilsback</dc:creator>
<dc:creator>Andrew Zisserman </dc:creator>
<dc:subject>Flower database</dc:subject>
<dc:description>We have created two flower datasets by gathering images from various websites,
with some supplementary images from our own photographs. The first dataset is a smaller one
consisting of 17 different flower categories, and the second dataset is much larger, consisting
of 102 different categories of flowers common to the UK.</dc:description>
<dc:publisher>University of Oxford</dc:publisher>
```

XML data is valid.

Figure 6 DataCite DOI assignment request example.

2.4.4 Accessing open data sets using web interface

Once the open data set is published and indexed by some open data discovery service, such as OpenAIRE, users who want to access the data using the identifier URI only need to open the link in the browser. The EGI ODP presents the users a preview of the data set with direct access to files and metadata browser.


https://datahub.plgrid.pl/#/public/shares/N_ghHSjyUMD9q8YQwyE7QQBvbjhhip1A_pfcTzyM0RE
Click to go forward, hold to see history
ONEDATA

OxfordFlowerDatabase-FlowerSet1

Public URL https://datahub.egi.eu/share/N_ghHSjyUMD9q8YQwyE7QQBvb...

OxfordFlowerDatabase-FlowerSet1 > FlowerSet1

FILES	SIZE	MODIFICATION
._image_0048.jpg	4 KB	2016-09-28 12:09
._image_0099.jpg	4 KB	2016-09-28 12:09
image_0001.jpg	50.85 KB	2016-09-28 11:09
image_0002.jpg	41.24 KB	2016-09-28 11:09
image_0003.jpg	46.2 KB	2016-09-28 11:09
image_0004.jpg	30.16 KB	2016-09-28 11:09
image_0005.jpg	32.73 KB	2016-09-28 11:09
image_0006.jpg	27.76 KB	2016-09-28 11:09
image_0007.jpg	37.93 KB	2016-09-28 11:09

Figure 7 Open Data set preview

2.4.5 Accessing open data sets using virtual file system

The EGI ODP enables users to directly access open data using POSIX protocol and the Onedata command line client – **oneclient**[8]. This may be done from one or many virtual machines or Docker containers running on the EGI Cloud Federation.

Naturally, this access of data may be done as part of an application by a community using EGI cloud computing resources, and paves the way for large-scale data intensive applications by integrating data and computing services on EGI.

[8] https://onedata.org/docs/doc/using_onedata/oneclient.html

3 Open Data Platform prototype deployment

In order to evaluate the EGI ODP in the user community, the first EGI ODP prototype has been deployed and forms the basis of the EGI DataHub open data dissemination service.

3.1 Dissemination and evaluation

The EGI DataHub is a Data as a Service (DaaS) offering from EGI, which aims to collect and give access to reference data sets. Based on the EGI ODP, the EGI DataHub will connect several storage providers giving users high performance and reliable access to large datasets, using multiple protocol options. EGI DataHub will be promoted to users as a service offering large reference open data sets, which using this technology can be accessed in an efficient manner.

The EGI DataHub is available at the following address: https://datahub.egi.eu and access is granted to all users who can be authenticated in EGI AAI via OpenID Connect protocol. It has been currently deployed in the Cyfronet data centre, with the plan to extend the deployment to multiple storage providers federated within EGI.

The EGI DataHub prototype currently gives users access to a subset (~10TB) of Sentinel-2 data, using GUI, CDMI and POSIX protocols. The goal of this use case is to give users the ability to access and process Earth Observation data remotely without the need to copy the data to storage local to their computing resources. In practice, the EGI DataHub allows users to access the data using the virtual filesystem POSIX protocol on any remote machine either in the cloud or hosted internally and run computations on the Sentinel-2 products.

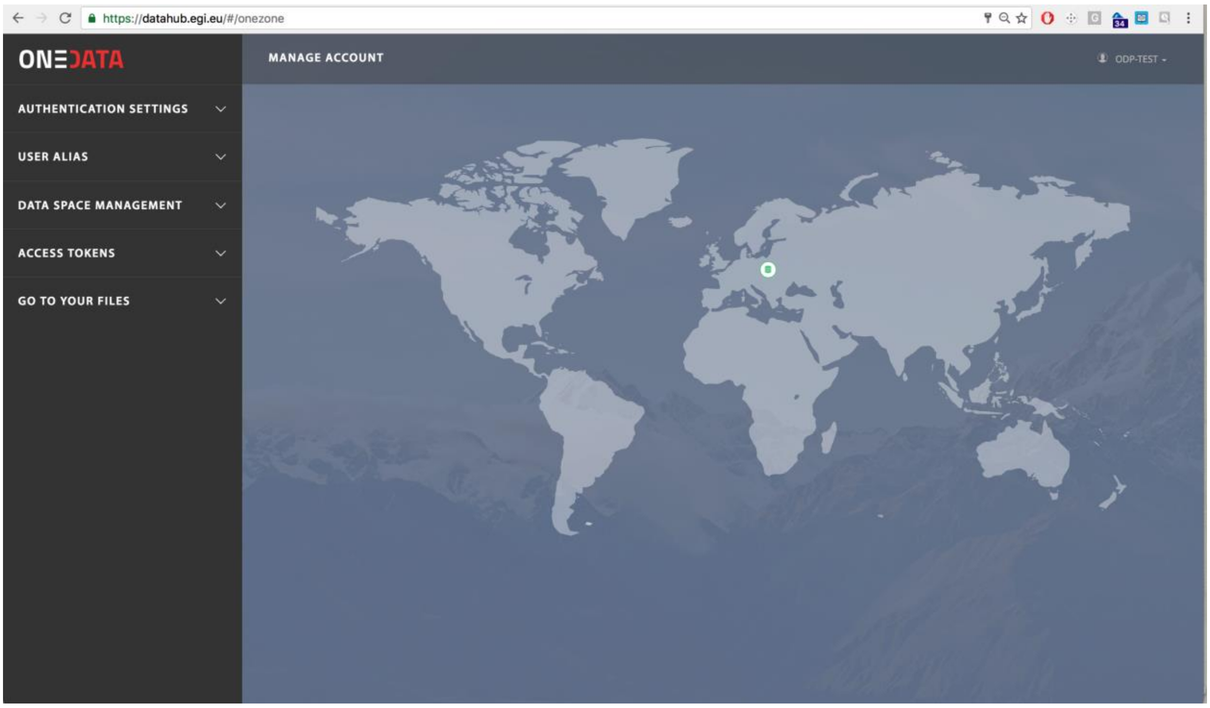


Figure 8 Main DataHub portal

The EGI DataHub service has been demonstrated during the DI4R conference in September in Krakow[9], by presenting a live demonstration of creating and publishing an example open data set, registering it via DataCite[10] and processing the data set in the cloud. EGI DataHub has been integrated with EGI OIDC[11] authentication mechanism, which allows users to access data using their preferred identity provider to login. During the hands-on session, users have started and configured their instances of Oneprovider service on EGI FedCloud, registered them in EGI DataHub and accessed data remotely using Oneclient POSIX protocol.

In addition to EGI DataHub service deployment, Open Data Platform prototype is being offered to user communities as a platform for their data. In particular, within LifeWatch use case EGI ODP is used as the data management platform for uploading files via the Water Reservoir Platform portal and metadata management. The initial prototype evaluation was successful in terms of functional requirements (POSIX data access, metadata management) and performance.

EGI DataHub has also been demonstrated as part of INDIGO-DataCloud where it was integrated with EGI AppDB service enabling EGI users to easily share job outputs between Virtual Machines deployed on FedCloud resources. Although at this within EGI-Engage scalability and performance of the Open Data Platform has not yet been assessed, as this is the goal of the remaining project

[9] https://www.digitalinfrastructures.eu/

[10] https://www.datacite.org/

[11] https://aai.egi.eu/oidc

period, the Onedata, which is the basis of the EGI ODP has been deployed and tested by several data centers including Cyfronet, INFN, DESY, PCSS, LIP, UPV and CSIC. Initial evaluations proved that the solution is scalable and performance is suitable for high performance computations.

The remaining part of the project will be dedicated to evaluation of the EGI Open Data Platform among the user communities, and the results will be reported in the D4.9 deliverable at M29.

3.2 Integration with EGI production processes

The EGI ODP is currently in the process of integration with the EGI services catalogue through the process PROC19[12].

This mainly involves security evaluation by EGI operations team, integration with GOCDB and implementation of Nagios probes for monitoring.

[12] https://wiki.egi.eu/wiki/PROC19

4 Plan for Exploitation and Dissemination

Name of the result	EGI Open Data Platform
DEFINITION	
Category of result	Software & service innovation
Description of the result	Open Data Platform aims at providing a novel solution for open data management, giving the researchers similar experience and ease of use as with commercial data management and file synchronization solutions, while providing means for seamless publication and access to open data from any location, either from personal laptop or virtual machine running in the cloud.
EXPLOITATION	
Target group(s)	The main target groups of the EGI ODP are: • RIs • international research collaborations • long-tail of science • industry/SMEs
Needs	The main community requirements fulfilled by ODP have been identified within M4.1 milestone deliverable and include: • Publication of open research data based on policies • Make large data sets available without transferring them completely • Enabling complex metadata queries • Integration of the open data access data management with community portals • Data identification, linking and citation • Enabling sharing of data between researchers under certain conditions • Sharing and accessing data across federations • Long term data preservation • Data provenance
How the target groups will use the result?	EGI Open Data Platform will be used by target communities to manage, process, share and disseminate open data, which are input or output essential to their research activities.
Benefits	The main benefits for the user communities include: • Unified high-performance data access and management system • Easy access to remote data sets from Virtual Machines and containers via standard POSIX protocol

	• Direct support for open data publishing • Secure data sharing between users and across federations
How will you protect the results?	EGI Open Data Platform as well as its underlying technology, Onedata, are fully open-source components licenses under Apache 2.0 license.
Actions for exploitation	Currently Open Data Platform is being integrated into the EGI operational services via procedure PROC19[13]. Once this is complete, selected data centers federated in EGI infrastructure will deploy Onedata and register it with EGI DataHub service to provide a distributed open data environment for researchers.
URL to project result	https://onedata.org https://datahub.egi.eu https://github.com/onedata/onedata
Success criteria	The main success criteria for this product are: • Number of storage providers provisioning storage space to the EGI DataHub service • Number of open data sets published via the EGI ODP platform • Number of communities use cases integrated with the EGI ODP platform • Number of open data sets accessed by external users
DISSEMINATION	
Key messages	• With EGI Open Data Platform, users can access, store, process and publish data using global data storage backed by computing centres and storage providers worldwide, • EGI ODP focuses on instant, transparent, POSIX-compliant access to distributed data sets, without unnecessary staging and migration, allowing access to the data directly from your local computer or worker node, • EGI ODP makes the process of open data publishing effortless by supporting Handle based identifiers (e.g. DOI) and OAI-PMH protocols
Channels	The EGI Open Data Platform will be disseminated through several channels including: • EGI website and newsletter • Scientific publications • Open science conferences
Actions for dissemination	• Live demonstration during DI4R conference in Krakow, September 2016 • Hands-on session during DI4R conference in Krakow, September 2016 • Hands-on session during 3rd ENVRI+ Week in Prague, November 2016 • Presentation during Workshop on Cloud Services for Synchronization and

[13] https://wiki.egi.eu/wiki/PROC19

	Sharing (CS3), January 2017 - EGI Community Forum, May 2017
Cost	No additional costs except for the already allocated within EGI-Engage will be necessary.
Evaluation	The dissemination results will be evaluated using several means: - Number of unique visits to the EGI DataHub website - Number of new users registering on the EGI DataHub service - Number of citations of publications related to EGI Open Data Platform

5 Future work plan

The main future work plan for the EGI ODP within the framework of EGI-Engage project includes:

Action	Planned completion
Enable migration of database schema between the EGI ODP release upgrades. This will ensure that all metadata is retained between subsequent upgrades, thus reducing the impact on users of upgrades.	Dec'16
Integration of the EGI DataHub upgrades with the EGI Change Management procedure. This will ensure high quality upgrades where the risk of adverse impact on service delivery is reduced by adherence to IT Service Management good practice, according to the FitSM standards family[14].	Dec'16
Integration with EGI services catalogue and fulfilment of PROC19 procedure requirements	Aug'17
Integration of X.509 Grid certificates for authentication and authorization. Implementation of this feature is being planned under the INDIGO-DataCloud project" and will assist in compatibility with software components using X509 AAI such as gridftp and voms.	Sep'17
Solution of data provision policy issues (e.g. related to national legislations, community specific publishing constraints, acceptable data, embargos, etc.)	Sep'17
Addition of further storage providers to the EGI DataHub service. This will diversify the amount of open data available on the Open Data Platform hosted by EGI.	Sep'17
Implementation of user activity logging. This information will include who performed what action on which dataset and is necessary for security and data provenance purposes.	Sep'17

[14] http://fitsm.itemo.org/

6 Conclusions

This document presents the status of the first official prototype of the EGI Open Data Platform. The requirements identified in the first phase of the project have been mostly fulfilled and the remaining functionality and integration issues will be resolved during the remainder of the project. EGI ODP has been already successfully demonstrated to various user communities (e.g. during the DI4R conference in Krakow in September 2016) and has been used as the basis for EGI DataHub public offering of reference open datasets, currently serving around 10TB of Sentinel-2 data products.

Future work will include be focused on further evaluation and development of the platform with user communities and the implementation of remaining functionalities. This will include further scalability tests with stress tests, more users and larger datasets, as well as validation against different use cases.

Furthermore, a Service Design and Transition Plan (SDTP) is being compiled for the EGI DataHub service and will be made available to EGI-ENGAGE project reviewers at the end of the project, when the prototype is fully ready to go into production.

By allowing easy and scalable access to data storage providers, the launch of the ODP prototype marks a significant step forward in enabling large-scale data intensive applications using EGI computing and data resources.