D3.3 Final report on the EUreka3D services and resource hub: design and implementation

Due date: M22

Dissemination level: Public

Authors:
Lukasz Opiola (Cyfronet)
Ignacio Lamata Martinez (EGI Foundation)

HISTORY OF CHANGES			
Version	**Date**	**Author**	**Comments**
0.1	14/10/2024	Ignacio Lamata Martinez (EGI), Lukasz Opiola (Cyfronet)	First draft
0.2	23/10/2024	Valentina Bachi, Antonella Fresa, John Balean (Photoconsortium) Jolan Wuytz, (EF), Renato Santana (EGI)	Content contributions
0.3	24/10/2024	Drew Baker (CUT), Valentina Bachi (Photoconsortium)	Review
0.4	07/11/2024	Antonella Fresa (Photoconsortium)	General review
0.5	08/11/2024	Ignacio Lamata Martinez (EGI), Lukasz Opiola (Cyfronet)	Final revision
1.0	08/11/2024	Valentina Bachi (Photoconsortium)	Submitted version

Table of contents

List of acronyms

Acronym	Description
AAI	Authentication and Authorisation Infrastructure
ACL	Access Control List
API	Application Programming Interface
AR	Augmented Reality
CH	Cultural Heritage
CHI	Cultural Heritage Institution
CHO	Cultural Heritage Object
EDM	Europeana Data Model
EOSC	European Open Science Cloud
GUI	Graphical User Interface
HTTP	Hypertext Transfer Protocol
HTTPS	Hypertext Transfer Protocol Secure
IT	Information Technology
LOD	Linked Open Data
OAI-PMH	Open Archives Initiative Protocol for Metadata Harvesting
PID	Persistent Identifier
PKI	Public Key Infrastructure
OS	Operating System
QoS	Quality of Service
TLS	Transport Layer Security
VM	Virtual Machine
VO	Virtual Organisation
WP	Work Package
XR	Extended Reality

EXECUTIVE SUMMARY

The digital transformation of the Cultural Heritage (CH) discipline is becoming a key priority for the European landscape, but still this domain lacks efficient mechanisms and a proper infrastructure to take advantage of recent technological advances to support the digitisation, preservation and presentation of CH data, especially in 3D. EUreka3D contributes to the current digital transformation of the CH domain with a range of services that are immediately available and tested in real-life environments. Such services and tools, next to the availability of new, high quality 3D datasets and capacity building resources, support the implementation and enrichment of the European common data space for cultural heritage.

One of the significant achievements of the EUreka3D project, described in this deliverable, is providing a fully-functional, production-grade information system that serves as a virtual data space for Cultural Heritage Institutions (CHIs). It is built mostly around the EGI DataHub and Compute services and allows the CHIs to securely and efficiently manage their 3D assets using advanced storage and computing resources. This system offers an innovative, cost-effective solution for data storage and the online delivery of heritage assets, providing CHIs with a flexible and secure environment that can compete with existing solutions.

This deliverable is the main technical deliverable of the project and presents the design and implementation of the EUreka3D services and resource hub proposed by the EUreka3D project to CHIs .

The deliverable presents the general concepts behind the services such as Cloud computing and data management, followed by an analysis of functional and non-functional requirements that have been captured from the project partners. It continues to describe the service architecture and design, used technologies and hardware resources provisioned for deployment of the service, and the functionalities that have been implemented based on the technical offerings of Cyfronet and EGI partners. The deliverable also describes the technical integration conducted to facilitate compliance with the Europeana Portal, protocols and procedures, done in collaboration with Europeana Foundation and Coordinator PHOTOCONSORTIUM, accredited aggregator for Europeana. In addition, the deliverable illustrates reflections and work done to ensure the integration of EUreka3D services and 3D datasets in the European Open Science Cloud (EOSC), in consideration of the current redesign landscape that is ongoing for EOSC. Finally, reflections on future sustainability models for the EUreka3D suite of services and tools are provided, in consideration of the users needs, value proposition and competitive advantage, and a basic market analysis.

1. INTRODUCTION

The Digital Cultural Heritage (DCH) discipline is becoming increasingly important in the scientific world, being supported by EC Recommendations, such as 2011/711/EU or the most recent recommendation of 10.11.2021 on a common European data space for cultural heritage [1] on the digitisation and online accessibility of cultural material and its digital preservation of cultural assets such as monuments, sites, work artefacts, and other tangible heritage, especially those at risk. The EC has made this priority clear with the investment in the field of CH of more than €1 billion over the last two decades.

However, the European CH domain lacks efficient mechanisms and a proper infrastructure to take advantage of recent technological advances to support the digitisation, preservation, and presentation of CH data. There is no agreement in the sector for a common framework to support CH data interoperability and workflows, especially in the domain of 3D digitised cultural heritage. Initiatives such as Europeana play a leading role in the presentation of European cultural heritage assets to the different scientific communities and the general public, but still lack some 3D functionality, have limited storage features, and rely on commercial solutions that often operate outside of the EU to render 3D objects. This is why a new innovative and Europe-centred platform is needed. Acknowledging this need, the European Commission has launched a new initiative for a European Collaborative Cloud for Cultural Heritage (ECCCH) [2] under Horizon Europe calls [2]. However, at the time of writing this deliverable (October 2024), the ECCCH operational services are not yet ready, as the ECHOES project that is leading the creation of the ECCCH is in its initial phases and the implementation of the future interoperability of ECCCH with Europeana and the common European data space for cultural heritage is still under investigation.

EUreka3D aims to contribute to the current digital transformation of the CH domain with a range of services that are immediately available and tested in real-life environments. The infrastructure described in this deliverable has been established as a production-grade service and resource hub, serving as the foundation for the next generation of CH systems and services. This robust and innovative IT infrastructure supports the storage, management, and presentation of CH data in a secure manner. It has been integrated with the existing Europeana platform and has proven throughout the project to be extensible and adjustable to the evolving needs of the CH domain. This IT infrastructure has been tested and validated in a series of pilot use cases led by four project partners, that combine the work of WP2 (where digital contents are generated) and WP3 (where digital services are made available). In addition, several external organisations as associate partners in the project tested the platform by publishing their assets and providing feedback.

Specifically, the technical work of EUreka3D has focused on the following aspects:

- **Digitisation**, conducting a comprehensive approach to the needs and workflow of high-quality 3D digitisation of CH assets, based on the recommendations of EU-funded VIGIE Study 2020/654 [3], including guidelines and standards to promote digitisation skills, workflows, and methodologies to balance the knowledge and skills of the different European CH institutions. This work was implemented in WP2.
- **Preservation**, using recent technological advances to store and manage data securely and efficiently. The aim was to establish the basis for leading the CH domain worldwide in storage capabilities that can cope with the challenging data requirements of CH. This work was implemented in WP3.
- **Presentation**, enabling existing initiatives (such as Europeana) and future applications to use 3D content from the EUreka3D resource hub, for presenting data in a suitable form to the end user (e.g., raw data, rendered interactive 3D objects, etc.). This area is crucial in ensuring greater access to and use of cultural material. Content generated in WP2 and stored in WP3 has been delivered to

Europeana with the support of accredited aggregator PHOTOCONSORTIUM. Demonstrations of applications have been prepared in EUreka3D as a combination of the efforts of WP2, WP3, and WP4.

These aspects contribute to the creation of the European command data space for cultural heritage.

As part of the EUreka3D project's effort to support the digital transformation of Cultural Heritage Institutions (CHIs), we have taken significant steps toward rethinking underlying work processes and business models. This project is not only supporting the European common data space for cultural heritage, but is also helping CHIs of various sizes and requirements to adopt advanced digitisation techniques, holistic representation of digitised tangible objects, and re-use approaches. From a technical standpoint, this project provides an alternative for CHIs to move away from legacy ICT systems or commercial solutions toward a more integrated, cloud-based IT infrastructure in Europe, that extends beyond individual institutions. The EUreka3D project has established a knowledge centre and a service and resource hub built on a smart technical infrastructure, using services from the European Open Science Cloud (EOSC). This infrastructure allows CHIs to access a virtual data space and leverage storage and computing resources to manage their 3D assets. It supports the creation, management, archiving, preservation, and sharing of digitised objects, focusing on 3D digitisation and knowledge modelling. With these outcomes, the project contributes to improving the digital capacity of the cultural sector and lays the groundwork for creating high-value datasets that will directly benefit CHIs. Besides granting future availability of the services and datasets produced in this project, the potential for expansion and enlargement of such services is evident and is at the basis of next actions such as the continuation project EUreka3D-XR starting in February 2025.

In a nutshell, the EUreka3D services and resource hub are based on the EGI DataHub service powered by an open-source Onedata data access and management platform. It enables CHIs to create, manage, archive, preserve, and share digitised objects, in particular 3D digitisation for the semantically enriched 3D records. In this light, the use cases identified in the project can be grouped into three very general categories:

Integration with content providers

- Content upload through several interfaces (Web GUI, POSIX, S3, REST API)
- Metadata creation and curation in relevant formats, e.g. Europeana Data Model
- Content search and discovery based on the metadata

Training and capacity building

- Online access to stored objects using web-based clients, as well as standard protocols
- Easy content sharing between users

Content aggregation in Europeana

- Visualisation of the 3D models online via a viewer embeddable in Europeana
- Access to the stored data objects and their metadata through standard interfaces including POSIX, S3, REST and OAI-PMH

Main use cases (i.e. user stories) and requirements have been collected during the past months and have been reflected in this document. The documentation for the EUreka3D services and resource hub requirements has served as a bridge between the technical partners and Content Providers/CH community in the EUreka3D project. Its progress has been reported in separate documents in the project's knowledge space, which have been updated continuously and submitted to the EC with six-month interval reviews.

This report describes the final version of the EUreka3D services and resource hub that has built the capability needed for the project.

1.1 Role of this deliverable in the project

This report outlines the final design of the EUreka3D services and resource hub. The document is the main technical reference of the project, and it explains the outcome from the set of requirements that have been captured from the project partners, which include CHIs, Europeana Foundation and one Europeana accredited aggregator (i.e. PHOTOCONSORTIUM) - and on the technical offerings that are available from the technical partners, Cyfronet and EGI. Specifically, the deliverable deals with the following aspects:

- The **formal requirements**, initially collected among project partners during several meetings and sessions, and shared to the wider community of stakeholders, CHIs and Europeana partner organisations. These requirements have been derived from the set of use cases that have identified the actors and the tasks needed to be performed in the system. They are formalised in Section 3 of this document, serving as an "interface" between the technical partners and the content providers partners of the project.
- The description of the design and implementation of the requirements in a platform that supports and accelerates the digital transformation of the cultural heritage sector, offering a secure and efficient option for European CHIs.
- The connection between the technical infrastructure and the workflow to store and manage the 3D assets delivered by the **content providers**.
- The connection between the technical infrastructure and the workflow to serve the 3D assets to the **end users**, including the publication of items in the Europeana Portal according to the requirements of the Europeana Data Model that is currently in a process of upgrading and expansion to better accommodate the information (data, metadata and paradata) of 3D objects.

The design and the technical implementation of the EUreka3D service and resource hub has been completed and refined iteratively during the project life, through interviews and interactive sessions organised within the project consortium, and with the broader landscape of 3D cultural heritage preservation institutes. Additionally, the deliverable illustrates the work conducted to grant onboarding of EUreka3D datasets and services to EOSC, in consideration of its evolving landscape.

1.2 Relationship to other deliverables

This document is closely related to the following deliverables:

- **Deliverable 3.1** *"Report on the EUreka3D services and resource hub: design and implementation"* (Project Month 4, April 2023), which is the initial version of this deliverable at the start of the project, which has been fully updated in this document.
- **Deliverable 3.2** *"The EUreka3D AAI architecture"* (Project Month 22, October 2024), which describes the infrastructure and technologies implemented to perform the authentication and authorisation of users in the EUreka3D systems.

There is also an evident link with the intermediary technical progress reports (D1.3, D1.4, D1.5, D1.6) and the integration reports (D1.2 and D1.7), where updates on the progress of WP3 tasks, including the connection and interoperability with Europeana, are provided.

This deliverable also illustrates the work done in T3.4 toward Milestone MS9 "*Integration of the EUreka3D services in EOSC"*, also due in Month 22, October 2024.

1.4 Structure of the document

The rest of this document is organised as follows:

- **Section 2** explains some important concepts to understand the technology layer used in the project and references in the rest of the document.
- **Section 3** briefly defines the platform requirements that have been discussed in different meetings between the technical partners of the project and the content providers.
- **Section 4** describes the general architecture of the EUreka3D infrastructure and its main components.
- **Section 5** outlines EUreka3D in terms of external services for the general CH community and the current status of EOSC.
- **Section 6** discusses the future of EUreka3D and its potential sustainability.
- Finally, **Section** 7 provides some conclusions.

2. CONCEPTS AND TECHNOLOGIES

This section explains some of the concepts necessary to understand the rest of the document. It gives a basic understanding of the technologies that were finally chosen to implement the EUreka3D services and resource hub, as described later in Section 4.

One of the main enablers of the digital transformation of the CH sector is cloud technologies. They support the sector with a technical infrastructure that can work efficiently in a scalable manner. **Cloud computing** is discussed in Section 2.1. The **deployment** of hardware and software in the cloud is an important part of the development process, and some notions of it are explained in Section 2.2. By their nature, 3D data often require large amounts of space, so **storage systems** play an essential role. These, and the **distributed management of data**, are discussed in Sections 2.3 and 2.4 respectively. Section 2.5 describes the role of **public dat**a and the importance of **FAIR principles** (Findable, Accessible, Interoperable, Reusable), metadata and Persistent Identifiers (PID). Data management in distributed environments is quite challenging, and Section 2.6 describes **Onedata**, an open source distributed data management platform developed by Cyfronet. The core system used in EUreka3D, called **EGI DataHub**, is based on Onedata and is presented in Section 2.7. The role of **metadata and paradata** is very important in the description of CH objects. **Europeana**, the reference European platform to discover digital CH objects, provides EDM (Europeana Data Model), a metadata model specific for CH that is discussed in Section 2.8. Section 2.9 discusses 3D content, and outlines some of its challenges, including standardisation, 3D formats and visualisation. Finally, Section 2.10 briefly mentions data protection, giving a definition of authentication and authorisation. This information is not discussed at length, as it is covered in another deliverable (D3.2).

2.1 Cloud computing

Traditionally, the computing hardware infrastructure that supports software applications has run on independent physical servers that were purchased, installed and managed by the organisations behind such applications. This process highlighted several problems, including:

- It requires a **costly and time-consuming** procurement process, especially in highly bureaucratic organisations, and delivery and installation often take considerable time. Also, local infrastructures have an environmental impact that also generates costs.
- Servers **require expertise** to be managed and maintained.
- Servers **become outdated** over time and need to be upgraded.
- **Hardware failure** affects all computing components sooner or later, and replacing or fixing server components means unacceptable application downtime. The solution is to duplicate servers for high availability, which is more expensive and involves additional servers to maintain.

Cloud technologies are used to address these problems. They have been made possible by recent advances in computing and have become a common trend in recent decades, as organisations have transitioned from physical servers to cloud environments. Cloud-based services proved essential in the context of Covid pandemic, when many organisations were forced to boost remote working and/or online interaction with users, customers and other stakeholders, and this acted as an accelerator of the digital transformation process of the organisations in various sectors (including cultural heritage sector). Cloud technologies are based on the following principles:

- Servers and other hardware components are no longer in the organisation's premises, but are hosted by a cloud provider, who rents them for the organisation. This avoids the need to purchase hardware and allows organisations to dynamically use the exact amount of hardware they need at any given time.

- Servers are maintained by the cloud provider, so the client organisation does not have to devote effort and resources to this task.
- From the customer's perspective, servers and hardware components are nor perceived and managed as physical elements, but as virtual ones. This means that servers can be requested and served almost instantaneously and on demand, which helps to easily overcome hardware failures.

This virtual allocation of servers is done through **Virtual Machines (VM**), which are software components that run over physical hardware and emulate and provide the functionality of a physical computer system. One of the advantages of cloud technology is that the virtual infrastructure can be created directly through software instructions, without the need to physically access the servers. Moreover, this virtualisation greatly facilitates *elasticity*, a term used to describe the ability of a system to increase or decrease its resources to adapt to the current workload.

EUreka3D uses EGI's Cloud Computing to supply its infrastructure, which comprises a European Federation of national and intergovernmental computing and data centres. As described above, this infrastructure is virtual, not physical, being flexible to change according to the project needs.

2.2 Software deployment

The deployment of software is usually a complex task that involves different technologies and methodologies. Software is typically moved between different isolated spaces called *environments*, which have different purposes. For example, the Development environment is where the software is first created and tested by the developers. This is often the local computers of the developers. There could be one or more Testing environments, where the behaviour and functional aspects of the software is tested by dedicated personnel. Sometimes there is a Staging environment, which is a very close configuration of the final environment in terms of software and hardware. And finally we can find the Production environment, that is the actual installation of the software that is used by the end users. The number of environments involved, as well as the timing and methodologies followed to move software from one environment to the next one, greatly depends on the organisation developing the software and the nature of the software itself.

There are multiple factors that can make the same software behave differently and unexpectedly in different environments and many technologies and methodologies have been developed to solve this problem. Recent computing advances have benefited the emergence of technologies to address this and other issues. **Container technologies** is one of them.

A container is a mechanism to pack a software application together with its required dependencies. It is based on some features of the Linux kernel. Although these features are relatively old, the use of containers has become more popular in recent years, mainly thanks to technologies such as Docker[1], which have provided a comprehensive solution that facilitated the use of containers from a user-friendly approach. Software can be run in containers on a machine to provide resource isolation between different containers and, in many respects, with the host machine. EUreka3D uses containers for the deployment of some of its software.

To facilitate the installation and deployment of software, different auxiliary software can be used. In EUreka3D, the cloud deployment is facilitated by an EOSC service called **Infrastructure Manager** (IM [4]). The Infrastructure Manager is a free and open-source framework to assist users with the creation of infrastructure in a cloud environment and the deployment of software. The framework has been been

[1] https://www.docker.com/

developed by the Grid and High Performance Computing Group (GRyCAP)[2] at the Instituto de Instrumentación para Imagen Molecular (I3M)[3] from the Universitat Politècnica de València (UPV)[4], and it is also partially funded by the European Commission. Some of the key features of IM include:

- In a user-friendly manner, IM orchestrates the deployment of cloud resources, the installation of software packages, its configuration and the monitoring and potential update of the virtual assets created.
- It supports different cloud providers, such as EGI, AWS, Google Cloud[5] and Azure[6], amongst others. Also, it can create complex virtual infrastructures across multiple providers too.
- It uses TOSCA [5] and RADL [6] templates to define the virtual infrastructure, making it possible to manage virtual infrastructure through source code.

IM can be used from different interfaces, including the Command Line (CLI), a Web Graphical User Interface (Web GUI) and a REST Application Programming Interface (API). More information on IM can be found in its documentation[7]

For additional information on the deployment process of EUreka3D, refer to Section 4.1.4.

2.3 Storage systems in research institutions

Over the last decades, data management and storage systems have undergone significant advancements, reflecting shifts in computing platforms. Initially, high-performance computing relied on local file systems. However, as cluster computing emerged, local network file systems, like Network File System (NFS[8]), matured and became essential for data-intensive applications within individual data centres. With the rise of high-speed global networks and the need to handle and share data across multiple data centres, early data management frameworks, such as Globus Toolkit's GridFTP[9], were developed to enable data transfers between sites. These systems required users to manually pre-stage data to the data centre where computations were performed. The widespread adoption of cloud computing by enterprises and research communities has driven a shift toward object storage solutions that do not offer POSIX data access. Most common examples include AWS S3[10], Ceph[11], OpenStack Swift[12], and the Hadoop Distributed File System (HDFS[13]).

Object storage is ideal for storing large volumes of unstructured data, like images, videos, and log files. Instead of organising data as files, object storage treats it as discrete objects, offering scalability, durability, and cost efficiency. In contrast, POSIX-based storage systems such as NFS are optimised for traditional file-based storage, making them better suited for structured data, like databases or documents, where complex file hierarchies and access patterns are common. The key distinction between these two types of storage lies

2 http://www.grycap.upv.es/
3 http://www.i3m.upv.es/
4 http://www.upv.es/
5 https://cloud.google.com/gcp
6 https://azure.microsoft.com/
7 https://docs.egi.eu/users/compute/orchestration/im/
8 https://en.wikipedia.org/wiki/Network_File_System
9 https://www.globus.org/blog/gridftp-a-brief-history-of-fast-file-transfer
10 https://aws.amazon.com/s3/
11 https://ceph.com/en/
12 https://docs.openstack.org/swift/latest/
13 https://www.ibm.com/analytics/hadoop

in their architecture and data management approach. Object storage manages data as objects, while POSIX-based systems organise it into files and directories, providing fine-grained control over permissions and file operations. Additionally, object storage systems like S3 do not support modifying parts of a file — instead, files must be replaced entirely — whereas POSIX systems natively offer random-write functionality, allowing for partial file updates.

In addition to the abovementioned storage systems, there are several other solutions that address various data management needs. For instance, GlusterFS[14] is a scalable, distributed file system that aggregates storage resources from multiple servers into a single, unified system, providing redundancy and flexibility for large-scale data environments. WebDAV[15] (Web Distributed Authoring and Versioning) extends HTTP to support collaborative file management over the web, making it well-suited for document sharing and version control across distributed teams. Meanwhile, XRootD[16] is optimised for high-performance, distributed access to data repositories of many kinds in scientific computing environments.

All these different storage solutions come with their advantages and trade-offs, but from a user's perspective, the variety of interfaces and protocols can become overwhelming. Navigating different systems with distinct APIs and access models introduces unnecessary complexity. In light of this, POSIX remains the most familiar and user-friendly interface, offering a straightforward and consistent experience for managing files, making it the preferred choice for many, especially for day-to-day operations.

Nowadays, POSIX storage systems are often employed as "scratch" storage, providing short-term, high-performance access to data during intensive computational tasks. This approach allows users to benefit from the low-latency, high-speed capabilities of POSIX-compliant file systems, which are ideal for temporary data processing needs. However, when it comes to longer-term storage or managing larger datasets across multiple tiers, object storage systems are increasingly preferred. They offer greater scalability and cost efficiency, making them well-suited for archiving, backup, and less time-sensitive data access. However, their lack of transparent POSIX access and proprietary interfaces present a significant challenge for users, as existing applications must either be adapted to new cloud storage interfaces or rewritten entirely.

2.4 Distributed data management

The previous discussion focused on storage systems within a single data centre, where various distributed file system technologies like POSIX-based storage or object storage coexist to address different data management needs. However, modern data environments are increasingly distributed across multiple, often independent, data centres. This brings new challenges, as data must now be accessed, shared, and processed across geographically dispersed locations, each using diverse storage systems. The complexity of managing data across these heterogeneous environments — ranging from POSIX-compliant systems to cloud-based object storage — poses a significant challenge. As organisations and research institutions scale up, the need for unified access to distributed data becomes more pressing. Solving this requires new approaches that can seamlessly integrate and manage data, regardless of its location or storage type, while maintaining performance, accessibility, and security across the network.

Those challenges also apply to the EUreka3D project which brings together different CHIs. Here, distributed data management is necessary to provide an abstraction of various storage systems and facilitate easy access,

[14] https://www.gluster.org/
[15] https://en.wikipedia.org/wiki/WebDAV
[16] https://xrootd.slac.stanford.edu/

sharing, processing and publishing of the stored Cultural Heritage Objects (CHOs) in a distributed cloud environment.

From the perspective of a single data centre within the broader distributed data environment of EUreka3D, object storage technology presents itself as a highly suitable choice for storing CHOs. Its scalability, ease of deployment in cloud environments, and capacity to efficiently manage large volumes of data align well with the requirements of this project. Cultural heritage objects are typically accessed in a read-only manner, which complements this architecture. However, while object storage works well behind the scenes, many users, such as curators and researchers, may not be highly technical. To ensure ease of use, the interface for managing and interacting with these digital assets should mimic a familiar, POSIX-like file system. This will provide users with a human-friendly experience while leveraging the benefits of object storage in the background, ensuring both simplicity and efficiency in accessing and managing cultural heritage data.

Apart from the challenges of unified access to distributed data and user-friendliness, several additional issues arise in the context of data management and access for CHIs:

- *Lack of transparent access solutions*: there is a need for transparent data access solutions in hybrid cloud environments. Applications should enable users to deploy them in the public cloud while accessing data as if it were available locally.
- *Inadequate support for legacy applications:* many applications assume they can access data as if it were available through a local file system (POSIX) on virtual machines or containers in the cloud. However, cloud providers typically offer only object storage or local network file systems that aren't accessible outside their infrastructure.
- *Vendor lock-on*: applications often face vendor lock-in regarding data access, making it difficult to migrate between different cloud providers due to incompatible data access and management interfaces.
- *Simultaneous access across protocols:* complex applications, which may include legacy components reliant on file system access alongside modern components adapted to object storage like S3, need to access the same data across different clouds and protocols. Existing solutions struggle to facilitate this requirement.
- *Scalability issues:* some smaller, more affordable cloud storage providers may lack the scalability required by large-scale data processing applications. This can lead to bandwidth limitations when the aggregate network interfaces to the storage become saturated.
- *High throughput and low latency requirements*: applications operating in multi-cloud or hybrid cloud environments demand high throughput and low latency for data access. However, current solutions often necessitate cumbersome pre-staging of data, resulting in increased latency and complicating application logic.

2.5 Public data collections

Public data collections refer to data records that are openly available to the public without the need for authentication, yet may be associated with varying levels of licences and usage restrictions. These collections play a crucial role in the digital landscape of CHIs, as they facilitate the dissemination of valuable information and resources.

By making CH data accessible to a broader audience, institutions can promote transparency, foster collaboration, and enhance public engagement with cultural heritage materials. Europeana operates as a discovery platform for such public collections that represent CHOs and promotes their use and reuse by a variety of communities, as described in detail in Section 2.8.

EUreka3D services and resource hub were designed to take advantage of the following principles and standards.

2.5.1 FAIR principles

Public data collections align well with the FAIR[17] principles, promising better usability of CH data:

1) *Findable*: to enhance discoverability, institutions can ensure that datasets are indexed in search engines and repositories. Adding descriptive metadata and unique identifiers (such as DOIs) helps users locate datasets more easily.
2) *Accessible*: making datasets publicly available without authentication can promote accessibility.
3) *Interoperable*: utilising standard data formats and adopting widely accepted metadata schemas can facilitate interoperability.
4) *Reusable*: to support reuse, institutions should provide clear licensing terms and comprehensive documentation for each dataset. This helps users understand the context, limitations, and proper use of the data.

2.5.2 Persistent identifiers (handles)

Persistent identifiers (PIDs) are long-lasting references to digital objects — web pages, documents, files etc. They are sometimes referred to as handles as in a handle system (registry of handles to information resources). They are crucial for the accessibility and interoperability of public data collections. Some exemplary implementations include:

- DOI — stands for Digital Object Identifier and is an ISO standard. They are widely used to identify publications, journal articles, research reports and data sets. DOI example: 10.1000/182 (can be resolved here: https://doi.org/10.1000/182).
- Handle.net — every PID may be expressed as a URL through the use of a generic HTTP proxy server, for example, https://hdl.handle.net/20.1000/100.

2.5.3 Metadata

Metadata is essential for public data collections, as it ensures they are discoverable, understandable, and reusable. By providing detailed descriptions of data — such as its origin, structure, and licensing — metadata enables locating and interpreting datasets effectively. Most metadata formats are based on XML and RDF, and these machine-actionable formats facilitate the interoperability of different information systems. Thanks to standards like Dublin Core or Europeana Data Model (cf. Section 2.8), data collections can be aggregated and indexed for quick searching.

2.5.4 OAI-PMH

The Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH[18]) is a standard designed to facilitate the automated exchange of metadata between digital repositories. It enables different systems to expose and share metadata in a structured and consistent way, promoting interoperability across various archives and databases. This allows for linking digital collections from multiple institutions.

[17] https://www.go-fair.org/fair-principles/
[18] https://www.openarchives.org/pmh/

The protocol's primary function is metadata harvesting, where third-party services, known as harvesters, can collect metadata from participating repositories. This metadata, typically formatted according to standards like Dublin Core, provides essential information about digital objects such as titles, authors, and publication dates. The harvested metadata can be easily indexed to build search engines and discovery services.

OAI-PMH is widely used in the cultural heritage, academic, and library sectors, ensuring that digital collections can be made accessible to larger audiences through common platforms. Although it only handles metadata rather than full digital objects, the protocol plays a crucial role in increasing the visibility and discoverability of content across distributed repositories.

2.6 Onedata — a distributed data management platform

Onedata is a data management platform that provides easy and unified access to globally distributed storage resources, supporting a wide range of use cases from personal data management to data-intensive scientific computations. It is an open-source project implemented by the team from Cyfronet over the past 11 years. Onedata creates a virtual file system layer spanning geographically dispersed computing centres and data providers that host heterogeneous storage resources.

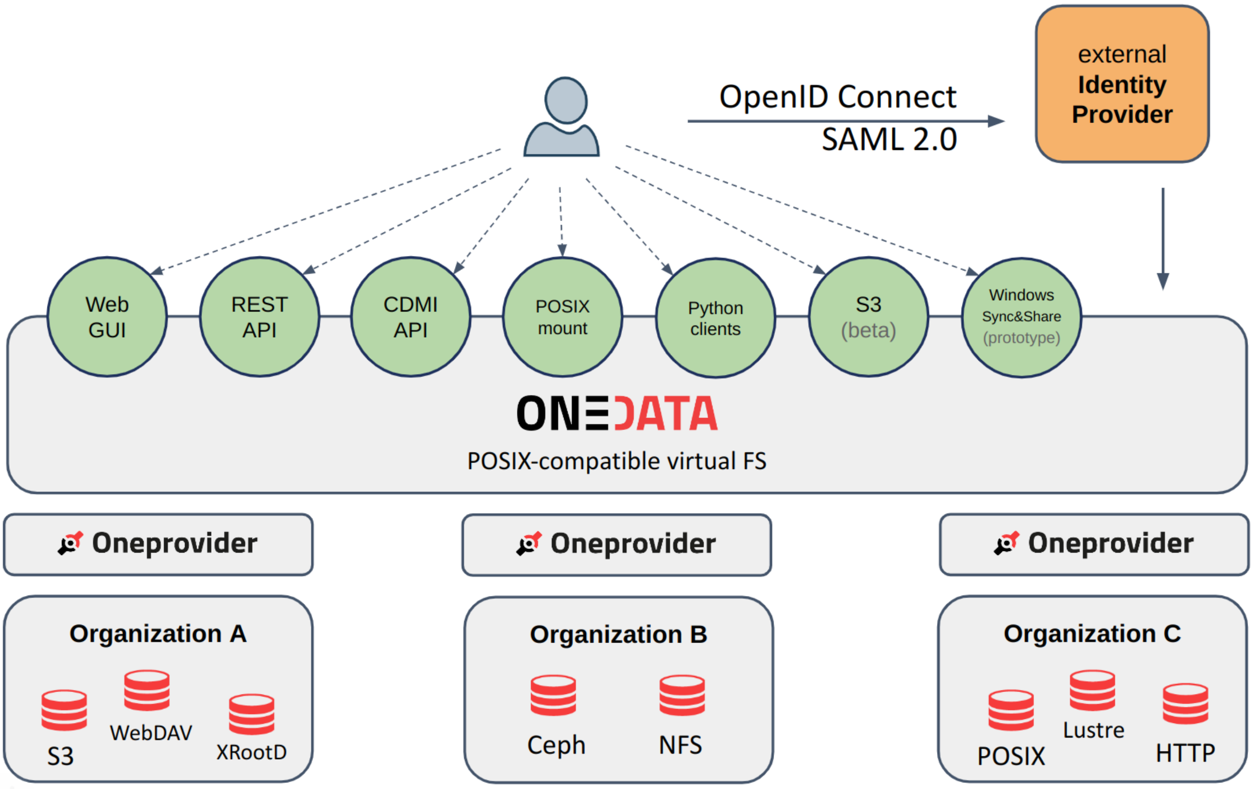


Figure 1: Data access virtualisation in Onedata

The Onedata virtual file system is POSIX compatible and uses a structure based on directories and files. Many interfaces can be used to access the virtualised data: Web GUI, REST API, CDMI API, fuse-based POSIX mount, Pythonic libraries, or S3. Regardless of the interface, a user gets a unified view of his data.

Onedata uses the concept of Spaces to organise the data. A Space is a logical data volume that appears as a monolithic file system from the user's point of view. Still, it virtualises the physical data stored on distributed storage systems of different data providers. Spaces facilitate collaborative data sharing between users and groups across organisational domains — using the Onedata interfaces, users can manage and access the data together in a unified namespace, while it is physically distributed. The system takes care of the replication management and on-the-fly data provision behind the scenes, but power users who understand the data distribution can embrace it and schedule data transfers, migrations, pre-staging, or define QoS (Quality of Service) rules. The QoS mechanism is used for policy-based replica management; users can define requirements regarding the replication ratio and storage data placement, meanwhile, the system will try to fulfil them automatically and reconcile the replicas whenever the data is modified.

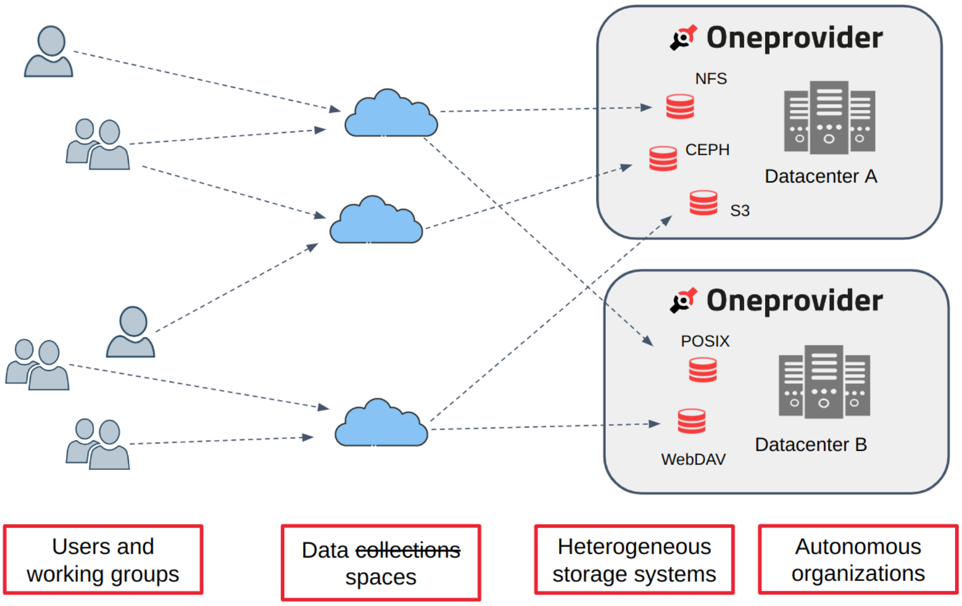


Figure 2: Spaces — logical data volumes used to organise data for users and working groups

Two main components of the Onedata architecture are Onezone and Oneprovider servers. Onezone is a coordinator of a Onedata ecosystem that implements AAI, managing the users, groups, and Spaces, and overseeing subject data providers. Oneprovider is installed in data centres and realises the data management and access tasks. It virtualises access to storage resources, synchronising the file metadata and managing transfers. Oneprovider provides a unified interface to multiple file systems used in the data centre, and the system can scale to improve performance. Thanks to the data import mechanisms, legacy storage resources can be indexed and exposed in a Onedata Space, without copying the data.

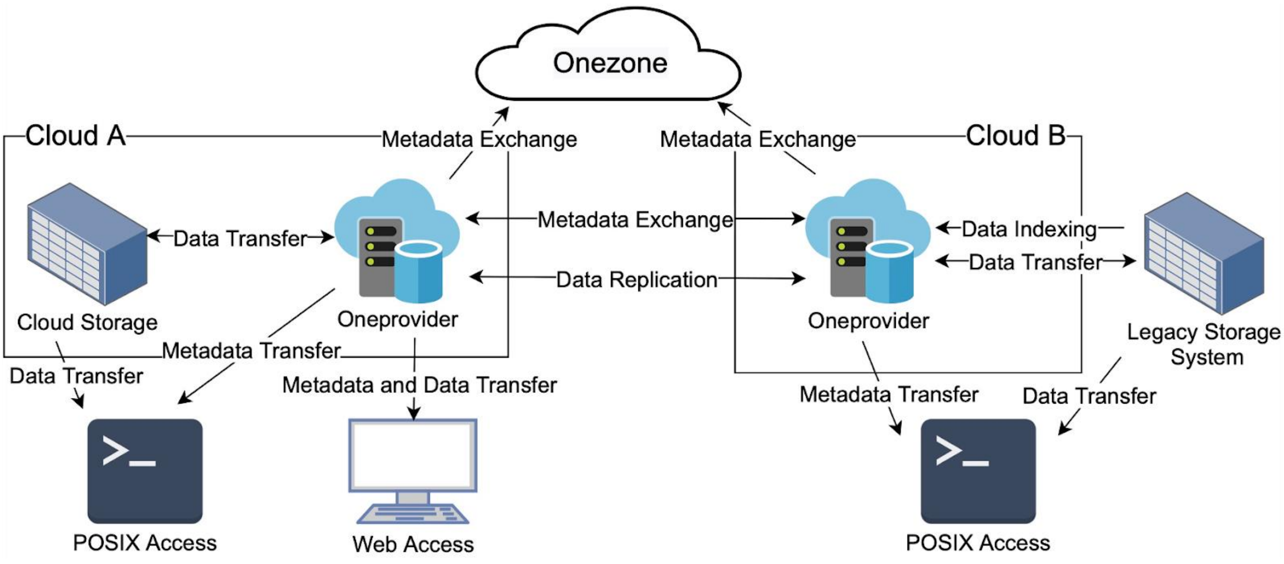


Figure 3: Overview of the Onedata architecture

Thanks to the direct storage access mode of Onedata client applications installed on computing nodes, it is possible to achieve data access performance suitable for HPC.

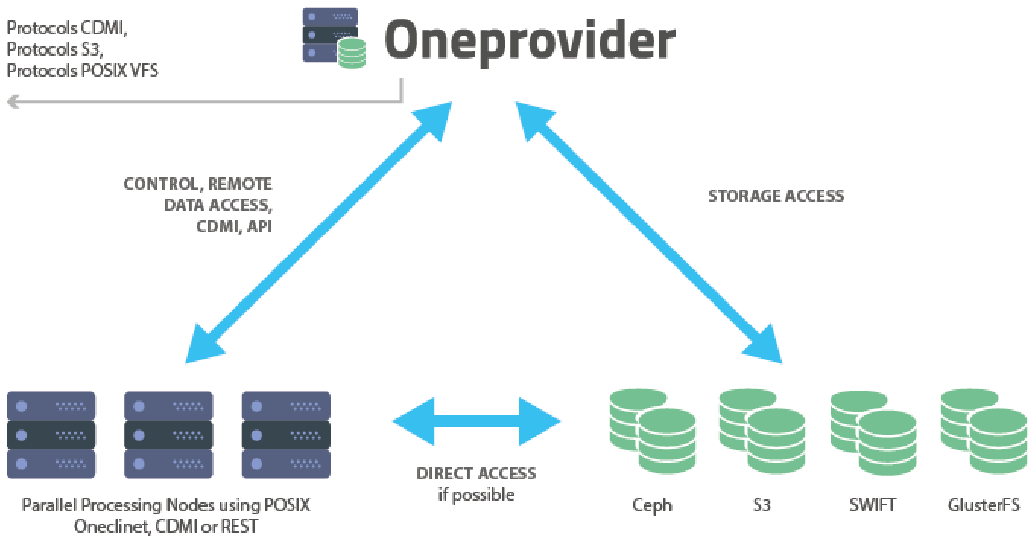


Figure 4: High-performance data access setup within a computing centre

Oneclient is a command line tool that enables users to access the virtual filesystem on a VM or their personal computers directly via a Fuse mount point.

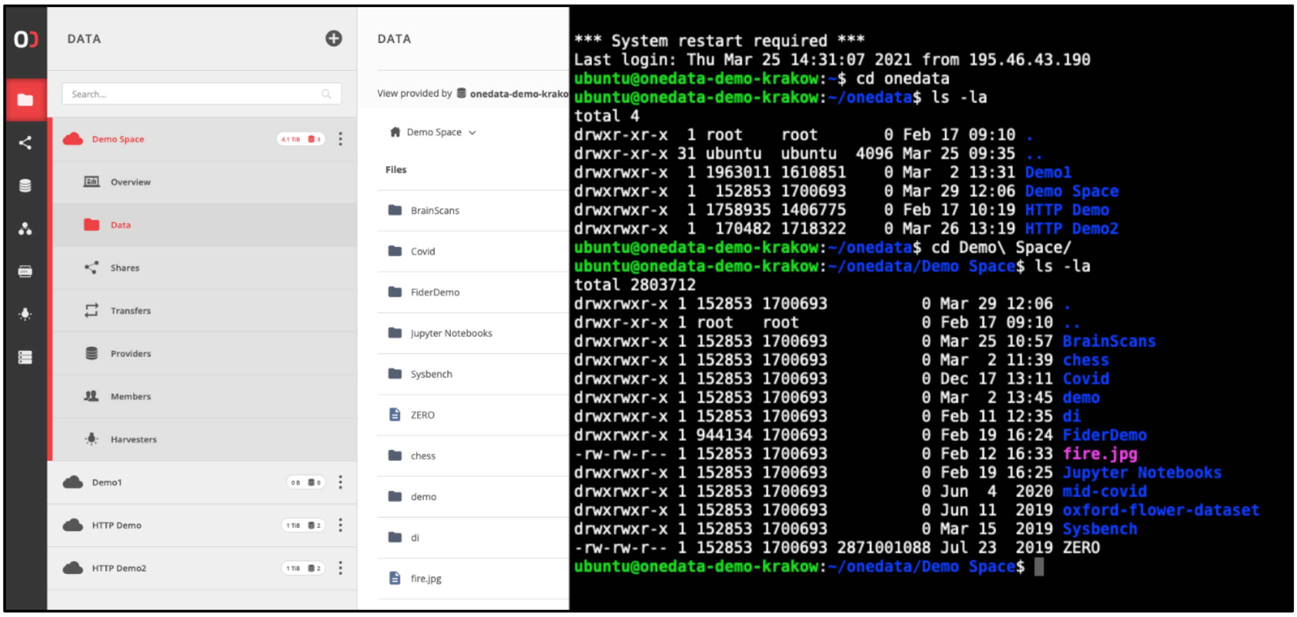


Figure 5: Onedata file system mounted locally with the use of Oneclient application

Onedata offers comprehensive support for custom metadata management: extended attributes, JSON, JSON-LD, and RDF. Any file or directory can be annotated using any of these metadata types. Later, this information can be used in the built-in data discovery engine, which indexes the metadata in an underlying Elasticsearch[19] service and offers a browser and query builder.

[19] https://www.elastic.co/elasticsearch

Figure 6: Data discovery view in Onedata; metadata is indexed to allow complex queries over datasets

Additionally, Onedata allows publishing data collections using its Share mechanism. A basic Share is a semi-public link that grants read-only access to a file or directory for non-authenticated clients. The link can be perceived as semi-public because it is not advertised and is impossible to guess. Each share can be converted to a public record, in this process it is assigned a persistent identified (DOI/PID), user-defined metadata conforming to a standard (e.g. Dublin Core), and advertised via the Onezone's built-in OAI-PMH server.

All Onedata components have REST APIs[20] defined using the OpenAPI[21] specification, enabling easy integration and automatic generation of client libraries for most existing programming languages and frameworks. The APIs provided by Onedata include:

- **Onezone API** to allow control and configuration of the Onezone service, in particular: management of users, groups, spaces, shares, providers, services, handles and clusters.
- **Oneprovider API** to access data through RESTful and CDMI-compatible endpoints and perform data management tasks such as data replication.
- **Onepanel API** to allow administrators to control Onezone and Oneprovider cluster deployments, modifying their configuration — e.g. adding more nodes or new storage resources.

The core system used in EUreka3D is based on Onedata.

For additional information, please refer to the Onedata documentation[22].

2.7 EGI DataHub

EGI DataHub[23] is a service for provisioning large reference public data collections, based on the Onedata[24] technology. As of October 2024, it brings together 17 data providers from around Europe. With over 800 registered users and 2200 Spaces (totalling ~2.16PB of allocated storage space), it caters to many scientific projects regarding their data storage and management needs.

[20] https://onedata.org/#/home/api
[21] https://www.openapis.org/
[22] https://onedata.org/#/home/documentation
[23] https://datahub.egi.eu
[24] https://onedata.org

EGI DataHub is essentially a Onezone instance, running as an EGI production service and managed by Cyfronet. It is tightly integrated with EGI Check-In (cf. Section 4.4.3) using OpenID Connect. DataHub embraces the information about group memberships that is propagated upon login from the Check-In service. Users' entitlements to specific VOs and working groups in EGI are reflected automatically as Onedata groups, which is crucial in organising seamless access to Onedata Spaces.

The users of EGI DataHub have access to all the features offered by Onedata, some of which have been described in the previous section. As one of the flagship Onedata ecosystems, EGI DataHub has its software regularly upgraded, keeping up to date with the latest stable Onedata releases (version *21.02.7* as of October 2024).

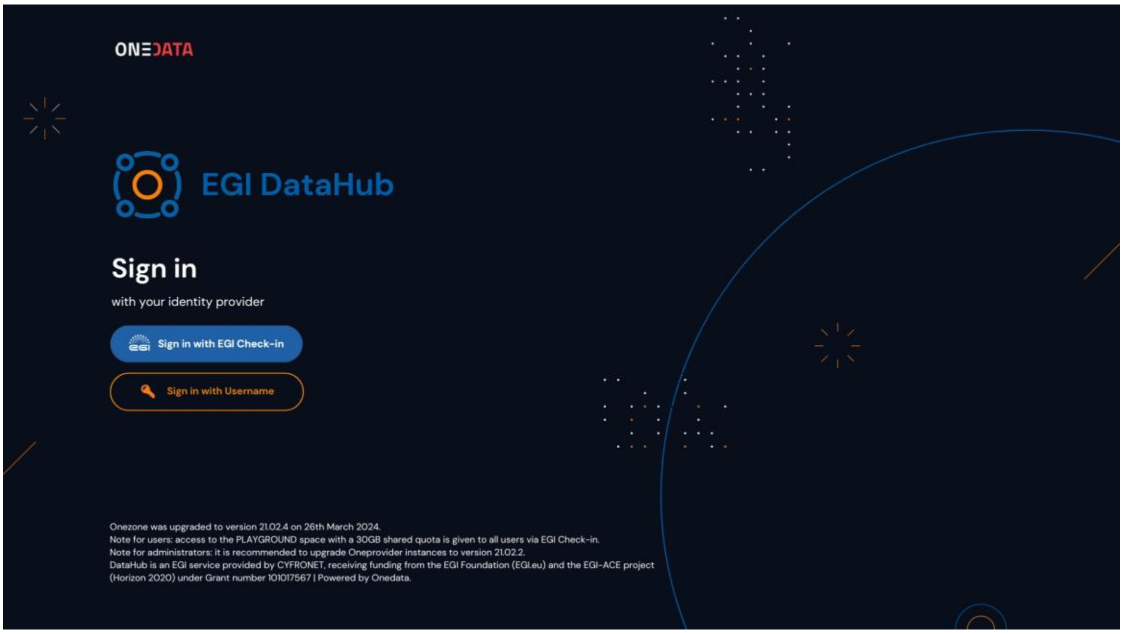


Figure 7: EGI DataHub login screen, with Check-in as a login option

For additional information, please refer to the EGI DataHub documentation[25].

2.8 Metadata and paradata in Cultural Heritage

Europeana[26] serves as the steward of the common European data space for cultural heritage, a flagship initiative of the European Union to support the digital transformation of the cultural heritage sector. One of the main goals of Europeana is to make digitised cultural heritage freely accessible and reusable online. Europeana aggregates metadata that describes cultural heritage records, offers discovery services, and provides information on how to access the corresponding digitised content, which is stored and maintained by various content providers who serve as repositories for these records. Currently, nearly 5,000 cultural institutions from all over Europe have shared their digitised collections in the europeana.eu website, via the provision of the metadata that describe such online content and a direct link to the digitised object that is made available in the institution's online repository.

[25] https://docs.egi.eu/users/datahub/
[26] https://www.europeana.eu

Europeana acts as an aggregation platform that hosts metadata files in the same metadata format, namely EDM: the Europeana Data Model[27]. These metadata files are organised in datasets, provided by CHIs and managed by aggregators. The data files themselves are generally hosted at the source, by the data-providing institutions. Europeana gives access to these metadata files and data files, collectively known as CHOs and sometimes also referred to as items or records) through item pages and a search and browse interface on europeana.eu. Europeana's technical stack, put simply, consists of a MongoDB database accessed through Solr. Users interact with Solr through a custom suite of APIs that allow for searching, browsing, filtering, fetching records and more. EDM is not built on any particular community standard but rather adopts an open, cross-domain Semantic Web-based framework that can accommodate the range and richness of particular community standards such as Dublin Core, LIDO for museums, EAD for archives or METS2 for digital libraries.[28]

EDM provides the ability of using Linked Open Data URLs for most of its data values, with the goal of connecting CHOs to the Semantic Web through the use of LOD vocabularies. EDM allows data and metadata to be part of this Semantic Web by providing ways to easily harvest and link data through their SPARQL and OAI-PMH endpoints. Using these open API standards for data exchange lies at the heart of Euroepana's open-source and open data policy. Europeana mainly uses OAI-PMH as the protocol through which it harvests data and metadata from data providers and aggregators. EDM is an RDF-based model[29], meaning it is best represented as a graph of triples containing subject, predicate and object:

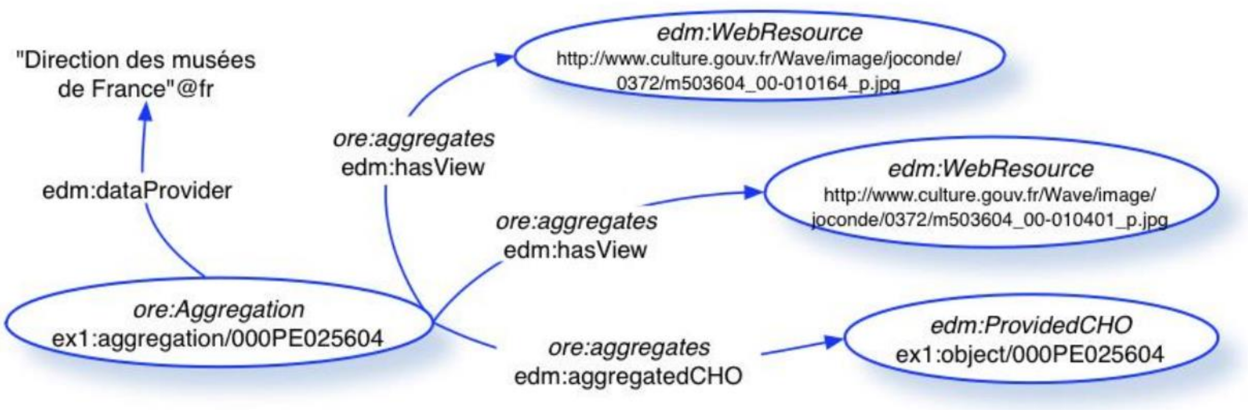


Figure 8: the Mona Lisa's metadata represented in an RDF graph, Provider's aggregation of web resources and provided CHO

This RDF data is most often provided to Europeana in JSON or XML format.

```
<rdf:RDF xmlns:...>
  <edm:ProvidedCHO rdf:about="#UEDIN:214">
    <dc:title>Flugelhorn</dc:title>
```

[27] https://pro.europeana.eu/page/edm-documentation

[28] https://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation/EDM_Primer_130714.pdf

[29] https://www.w3.org/RDF/

```
    <dc:type rdf:resource="http://www.mimo-db.eu/HornbostelAndSachs/356"/>
  </edm:ProvidedCHO>
  <skos:Concept rdf:about="http://www.mimo-db.eu/HornbostelAndSachs/356">
    <skos:prefLabel xml:lang="en">423.22 Labrosones with slides</skos:prefLabel>
  </skos:Concept>
</rdf:RDF>
```

Figure 9: example of EDM metadata as provided by a data provider.

The API suite is also completely freely accessible and reusable for development. End users interact with these APIs in a human-friendly way through Europeana.eu, which provides a user interface for searching, browsing, filtering, retrieving, downloading and reusing records. Cultural heritage data shared with Europeana through the network of aggregators is made publicly available through europeana.eu and disseminated in editorials, online collections and virtual exhibitions, giving educators, researchers, culture lovers and people everywhere the opportunity to explore Europe's cultural heritage.

Alongside data and metadata for CHOs paradata is increasingly seen as vital for the documentation of CHOs. The paradata record provides details about the provenance of the digitisation process essential in estimating the quality of the CHO (as defined by VIGIE 2020/654 [3]) and crucial for the reuse of CHOs beyond their original use-case intent. The correct documentation of CHOs paradata enables the re-user to establish confidence in the digitisation process(es) undertaken to produce the CHO as needed for their own (re)use case (e.g., a CHO produced for an online serious game may not be useful to a conservator requiring millimetric precision and high-resolution imagery to undertake their work, and vice versa). The paradata record also includes details on who, when, under what circumstances, by what method, and the authorisation/permissions (and the extent of such permits) for the CHO, providing a method of authentication of the digitisation and the resulting CHO, and details of alterations to the initial data acquired during post-processing to create the CHO (e.g., interpretation, filling of lacunae, combining of data sets, harmonisation of images to create textures, etc.).

The current profile and structure of EDM is evolving, in an iterative process led by the Europeana 3D Working Group of the data space project. The 3D Working Group, of which EUreka3D coordinator PHOTOCONSORTIUM is a member, works to review and expand the metadata fields that compose the EDM, in the light of accommodating richer and more descriptive information to accompany the 3D collections, and particularly to expose the paradata.

2.9 3D content

There are many challenges associated with the ingestion, processing, aggregation and delivery of 3D content. These challenges stem from the nature of 3D content, current hardware limitations and the quality target set by the VIGIE study [3] of the European Commission, as it promotes guidelines to ensure the highest level of quality and the best possible outcome. The current capabilities of consumer computers and networks impose limitations and challenges in the design of the EUreka3D platform. Quite often, processing of 3D data online is done on the client side, so the actual device used by the user plays a key role and must be considered when designing 3D experiences which are affected by network limitations, computer memory or processing capacity.

The 3D industry has greatly evolved over the years but still **lacks the standardisation level** that 2D content has. This lack of standards for the use of 3D data makes it a challenge to decide on a universal 3D format. Different users accessing an online 3D model have different use cases that may necessitate making the model available in multiple formats and file sizes. Someone wanting to 3D print a model will be better served with a lower resolution *.stl* file, and someone wanting to import a model into their game development engine is better served with a high quality glTF file. Herein, some content providers may use a format for the archival of 3D data, but this may not be the best choice for the visualisation or delivery to end users. For example, OBJ is a widely known format, commonly accepted by 3D software and 3D visualisation libraries, but in its text version it is less space-efficient for data, making it a poor choice if the data to be sent over a network are too large. Such cases can benefit from a binary format such as PLY. The Nexus multi-resolution format[30], created by CNR-ISTI (Italy), delivers 3D data more efficiently over the network, but it is not supported by common software and current technical challenges make it unsuitable for 3D CAD data. Some algorithms and 3D formats focus on compression ratio, while others focus on performance. It is usually a trade-off: compression makes more efficient use of space (benefiting, for example, the storage or transfer of a file) but increases processing effort (both for compressing and decompressing the data). These are not intrinsic problems for 3D, as 2D content also suffers from them, but they are more prominent in 3D as it is inherently more complex in its nature, and requires extensively larger amounts of space than 2D content, with implications for its storage, processing and transfer over a network. It is therefore necessary to consider the creation of the 3D model in different versions and formats, so to accommodate different users requirements (e.g. professional users may need a higher resolution model available for download and offline fruiting of raw data, while for dissemination purposes a compressed version shared with an embeddable viewer could be sufficient for the casual viewer to enjoy the content).

Another difference with the 2D industry is that the 2D industry has already solved basic aspects of 2D **content visualisation**. Even if new 2D formats emerge to solve new problems or improve certain aspects, the use of 2D content is basically transparent to developers, who for the most part do not have to deal with the burden of implementing or configuring 2D visualisation tools. For example, using an image on a Web page is a simple process that barely requires any effort from the developer, as the Web browser is already prepared to handle the situation in a globally expected way. However, the use of 3D content may still require developers to address some fundamental questions and make decisions on aspects such as the support library to be used, the data format to be accepted for efficient delivery and processing, and so on.

Therefore, a platform to deal with 3D content is not limited to providing a technically capable infrastructure but also has to address the challenges of the management of these three categories of assets: the **data** (3D models, raw data, audiovisual content, etc), the **metadata** (information about the models) and the **paradata** (information about the digitisation process). Many of the challenges associated with data have already been mentioned above. **3D metadata information** has been studied for a long time, and many of the challenges involved have been minimised with the help of EDM[31] (see Section 2.8), which provides a common framework for the understanding of systems that exchange CH metadata. However, the situation is not so favourable as far as the paradata are concerned. **Paradata information processing** is a necessary yet not widespread practice amongst 3D content providers and is not currently addressed by EDM. Delivering the paradata information associated with some data provides key insights into how the digitisation process was carried out to obtain the data. Although there are different initiatives and efforts focused on the description of paradata, the CH sector lacks a formal data model to express them, and this is one of the future enhancements planned for EDM.

EUreka3D had to address the challenges of managing 3D content.

[30] https://vcg.isti.cnr.it/nexus/

[31] https://pro.europeana.eu/page/edm-documentation

2.10 Authentication and authorisation

Every system needs protection against unauthorised access, from the visualisation of sensitive and confidential information to protecting data against unauthorised manipulation. There are two common mechanisms for protecting systems: authentication and authorisation.

Authentication is the process of identifying who the user is. Proving who the user is typically involves the knowledge of a password or the possession of a hardware key or a device.

Authorisation is the process of determining what the user can do in a system. This is sometimes implemented through user and resource permissions, such as the permission of a user to read or write on a system file. Sometimes, permissions are not only assigned to individual users but to a set of users that belong to the same "group".

EUreka3D uses a system called Check-in to implement its authentication and authorisation. For a more detailed discussion about this, refer to Deliverable 3.2 "*The EUreka3D AAI architecture*".

3. EUREKA3D SERVICES AND RESOURCE HUB REQUIREMENTS

Platform requirements have been collected in EUreka3D during multiple dedicated sessions between the technical partners and the content providers of the project. This section reports the outcome of the information gathered during these meetings:

- **Section 3.1** describes and characterises the different **users** of the system and how they will use the system.
- **Section 3.2** specifies the **added value** that the platform should provide for its users.
- Some **technical requirements** are listed in **Section 3.3**.
- **Section 3.4** defines requirements in terms of **data**, **metadata** and **paradata**, and how these will be used by users.
- Some requirements for the communication with external systems are discussed in **Section 3.5**.
- **Time** requirements are described in **Section 3.6.**

3.1 Users

In general terms, users can be divided into two categories:

1. Users that provide content (known as "Content Providers", mainly represented by CHIs):
    - Small CHIs with no infrastructure that need an infrastructure available for them to store and manage their digitised assets and associated data.
    - CHIs that do not want the trouble of managing their own technical infrastructure.
    - Institutions that want to publish their assets on the Web, especially in Europeana.
    - Dedicated CHIs that want their data safely stored in a European cloud (systems based in Europe).
2. Users that consume content (known as "data users"):
    - Professionals (architects, archeologists, etc) who want to download data to re-use it in their own job.
    - Researchers working in CH to access and re-use the data for their research.
    - Users (both humans and machines) that want to use the data as the basis for modification and exploitation in different fields (*e.g.* tourism, CH experiences, promotion of CH assets, printing business, educational institutions, *etc*).

The CHIs will use the system in this way:

- CHI will manage data,metadata, and paradata in the system (upload, modification, deletion, *etc*).
- CHI will control the access to their own data and should be able to share it.
- CHI will limit the audience of the objects (*e.g.* published in Europeana or not).
- Each CHI providing data can decide per each dataset if their data are downloadable or not and by whom.

Data users will use the system in this way:

- To explore and view objects (3D rendered objects), metadata and paradata.
- To download raw data if the owner of the data allows it.

3.2 Added value that must be provided by the platform

The platform should provide this value for its users:

- A European-based infrastructure.
- High-quality 3D assets to be used for different purposes (research, industry, *etc*).
- Access control for the data (as a content provider, provide access to a restricted group of users).
- Proposal for specific standards to enable interoperability (e.g. for the metadata).
- A user-friendly easy-to-use system (including documentation).
- Support and guidance.

3.3 Technical requirements

The technical requirements of the EUreka3D services and resource hub are:

- A service with a GUI and an API to upload, download, store and manage data.
- Authorisation rules that can be configured via GUI or API.
- A component for the visualisation of 3D objects, including basic operations such as rotation/zoom of the model and predefined views (side, front, etc).
- Documentation about how to use the system (text, images, video, *etc*).
- A component for interoperability and the integration with Europeana (*e.g.* MINT[32]).
- The platform should be compatible with Europeana requirements (*e.g. oEmbed*).
- The platform should be accessible via the Web.

3.4 Data, metadata and paradata requirements

- Content Providers should be able to decide what data they want to share (*e.g.* public access or with specific groups of people).
- Data will be used differently by each user group (*e.g.* architects, game development companies, etc).
- Users will download data and use it in their own fields.
- Paradata, metadata and data per Object are needed. Specific attributes should be defined.
- User-friendly interfaces must exist to input para/metadata and to import existing metadata that CHIs may already have.
- Most CHIs will use metadata in these formats: XML, CSV or DB (exported in JSON or Excel). Some may expose it through APIs (*e.g.* Museo della Carta). EDM uses RDF/XML.
- Metadata must be interoperable with Europeana.
- CHIs will need to update metadata over time (changes, added information, *etc*).
- If necessary, systems like MINT can be used for transformation of metadata to EDM, to store metadata and to publish objects in Europeana via OAI-PMH. However, this functionality could be covered also by the EUreka3D services and resource hub, which could be beneficial to store metadata together with the corresponding data.

3.5 External access

- The 3D models of the platform should be published in Europeana. Herein, there should be some technical data exchange between both systems.
- Europeana should be treated as a single user, so every user accessing EUreka3D metadata through Europeana can be considered the same user.

[32] MINT is a tool created and maintained by NTUA National Technical University of Athens to map source metadata to EDM

- Communication with other external systems is not foreseen at this time.

3.6 Time requirements

The following requirements in terms of timing for the EUreka3D project have been agreed among the partners:

- **M6-M12** of the project: Main project decisions resolved (decisions about metadata, object storage and basic 3D viewer).
- **M12-M18** of the project: Main developments and integration tests (authorisation implemented, integration tests with Europeana completed, 3D viewer embedded - using Europeana requirements, and collection of metadata). First release (Minimal Viable Product, MVP) should include storage and management of data and visualisation through Europeana (so mechanisms for integration with Europana are in place).
- **M18-20** of the project: Functional platform. 3D objects must be uploaded by the Content Providers of the project by M22 (CUT, CRDI, BIBRACTE and Museo della Carta).

- **M20-24** of the project: testing with additional content providers external to the EUreka3D consortium.

4. EUREKA3D SERVICES AND RESOURCE HUB

The EUreka3D project developed a piloting action, including four cases, to experiment with innovation in CHIs' workflows, especially to accommodate the needs and requirements concerning the management and sharing of 3D collections. The main result of this piloting action is a comprehensive suite of services and tools. It was created following an iterative process that saw the participation of all project partners, including content providers (who represent the category of users of the platform), PHOTOCONSORTIUM, and Europeana (who supported the development of the aggregation service in the EUreka3D system to be fully integrated and compliant with Europeana technical and procedural frameworks).

In the outreach communication of the project, the outcome of the deployment of the EUreka3D services and resource hub are referred with the brand name of EUreka3D Data Hub.

Figure 10 presents a graphical overview of the complete flow, from the physical CHO to be digitised in high quality, to the delivery of the digital object to end users.

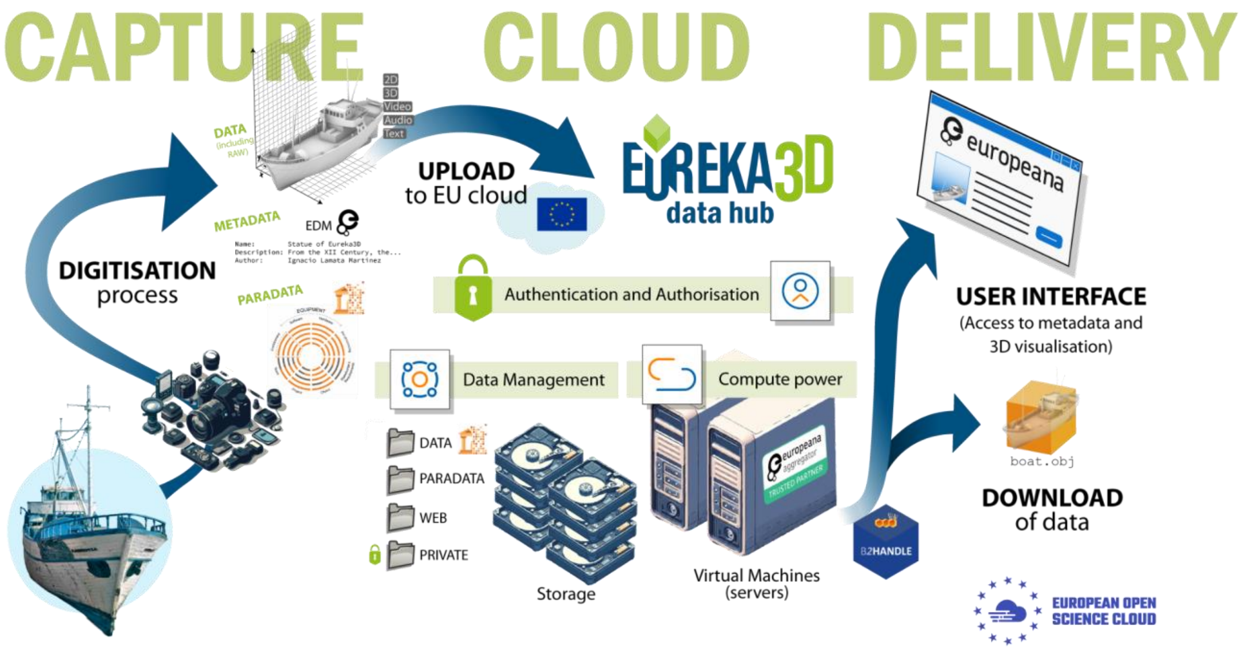


Figure 10: Overview of the EUreka3D flow, from digitisation to delivery

Three big phases are depicted in the figure:

- **Capture phase**. The flow starts with a physical CHO that must be digitised for both preservation and sharing purposes. The study of the object and its digitisation process is conducted with cameras, sensors, chemical analysis, and other means. As a result, three types of data are obtained: (i) Data (which includes the 3D models and the raw data), (ii) Metadata (information about both the physical and the digitised object) and (iii) Paradata (information about the digitisation process: cameras, sensors, personnel, weather conditions, budget, etc. How the digitisation is done is a crucial factor to assess the quality of the digitised object).
- **Cloud phase**. The data from the Capture phase is uploaded to the cloud. There are three main components that make up the EUreka3D infrastructure: (i) EGI Check-in, which is in charge of the authentication and authorisation of users (see Deliverable 3.2 "The EUreka3D AAI architecture" for

additional information), (ii) EGI DataHub, which is in charge of the data management, and constitutes the core of EUreka3D (see Section 4.2), and (iii) EGI Cloud Compute, which provides the servers that run auxiliary applications in EUreka3D, such as the code for the 3D viewer (see Section 4.1).

- **Delivery phase**. The last phase consists of the sharing of data externally, with users and other systems. It is supported by EUDAT's B2HANDLE[33]- a European initiative to release and manage PIDs (Persistent Identifiers) with which a collaboration was established by the EUreka3D project -, and the OAI-PMH capabilities of the EGI DataHub. The main system for publication is Europeana (see Section 2.8), which lists the published objects with their corresponding metadata, and a link for its 3D visualisation. Additionally, end users can use the DataHub to upload the 3D models, raw data and other data artefacts.

The actual architecture of the system is presented in Figure 11, which represents the three big groups of actors: (i) CHIs presented as Content Providers for the platform, (ii) Cloud Operators that develop and maintain the technical infrastructure and (iii) The general CH community. The arrows show the data interactions between these groups of actors and the system.

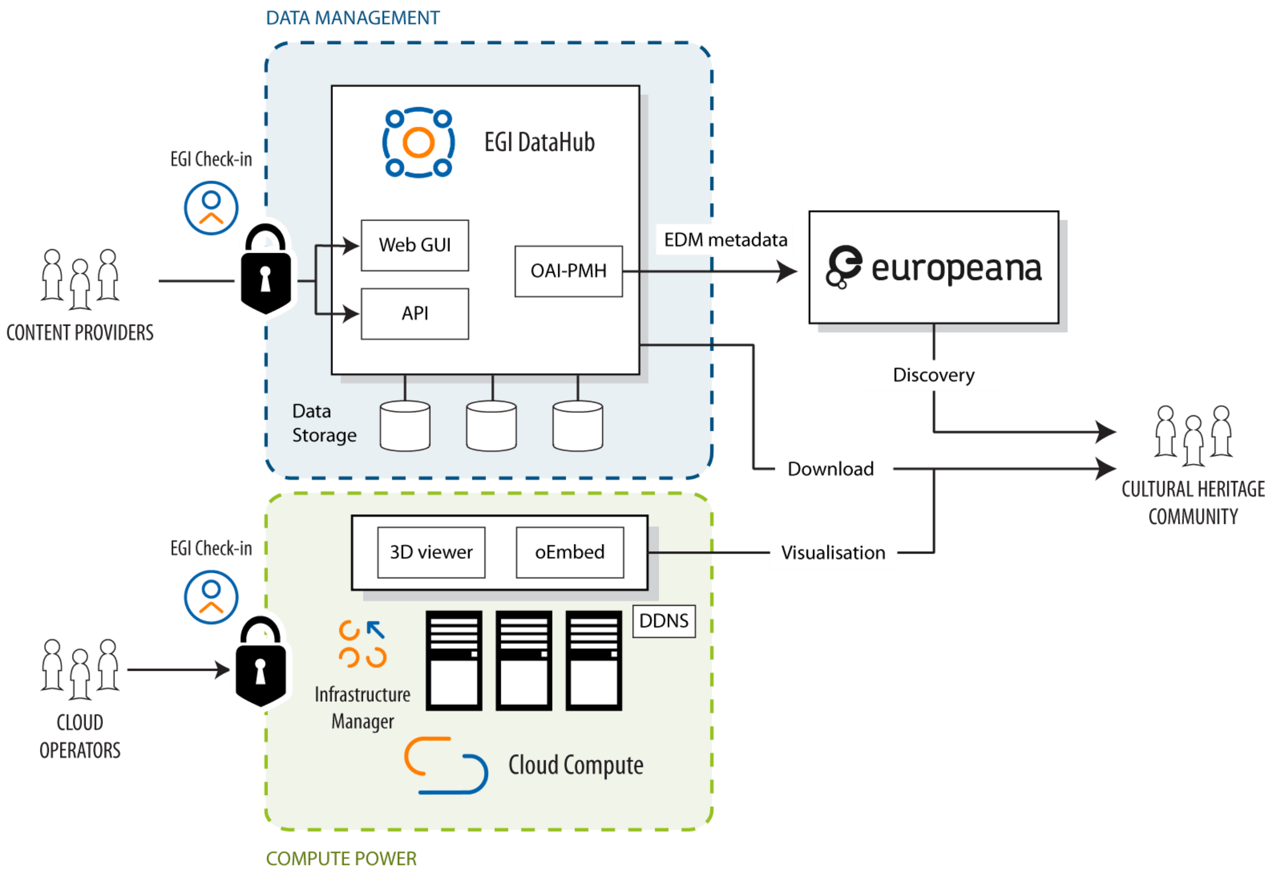


Figure 11: Overview of the architecture of the EUreka3D platform

[33] https://www.eudat.eu/services/userdoc/b2handle

We can find three main components in the figure:

- The **Data management**, which is protected by Check-in and implemented by the EGI DataHub. It provides two interfaces (GUI and API) and has an OAI-PMH endpoint that is used for the transfer of metadata to external systems. It is also supported by some data storage. This component is further discussed in Section 4.2.
- The **Compute power**, which is protected by Check-in and implemented by the EGI Cloud Compute. This component provides the main functionality to support the 3D visualisation and is further discussed in Section 4.1.
- **Europeana**, which is actually an external system of the platform, but acts as a support system for the discovery of EUreka3D objects and as a link for their visualisation.

4.1 Cloud Compute and Visualisation

EUreka3D relies on hardware in the cloud running specific software to support the 3D visualisation. Hardware allocation in the cloud benefits from *elasticity* (see Section 2.1), so that additional hardware resources can be allocated on demand if the workload demands it. They constitute the basic infrastructure to run the software necessary for the visualisation of 3D content. This software is necessary as Europeana only offers a catalogue of CHOs, providing metadata and references for end users, but the actual visualisation runs in each of the different external systems that are referenced from the Europeana Portal. Details on the hardware used and the deployment of software are given in the following sections.

4.1.1 Process for hardware allocation

The allocation of hardware resources is done within a cloud environment, as explained in Section 2.1. The servers used are *virtual*, allocated on demand as the project needs in a matter of minutes. This allows the infrastructure to be as agile as software components, being able to adapt to rapidly evolving environments.

The infrastructure of EUreka3D is deployed in **CYFRONET-CLOUD**, one of the resource providers of the EGI Federation, which uses OpenStack[34] as its cloud technology. Amongst other things, there are three main resources that are managed through the OpenStack platform:

- **Computing layer**. The different servers needed by EUreka3D are implemented as Virtual Machines, each of which is managed independently like a physical server.
- **Storage layer**. Disks are managed as virtual devices to provide the required storage capacity, and can be attached to the different Virtual Machines as if they were physical drives connected to physical servers. In EUreka3D, the main storage allocation is managed by the EGI DataHub[35] service described in Section 4.2.
- **Networking layer**. Network connectivity is managed and configured to allow Virtual Machines to communicate. This includes traffic routing (to move data packets from a source to a destination) and firewall rules (to block undesired network traffic).

Keeping the OpenStack platform operational is a complex process involving many challenges, and most of this burden is handled automatically by the EGI Federation resource provider (namely, in the EUreka3D

[34] https://www.openstack.org/

[35] https://www.egi.eu/service/datahub/

project this is by the EGI federation member Cyfronet, associate partner in the EUreka3D project), which facilitates the allocation of hardware for the project.

Figure 12 depicts an oversimplified version of this hardware allocation: There is a pool of physical hardware resources available in the cloud provider. When a user requests a server, a Virtual Machine is created by the cloud software (OpenStack in this case) and some hardware resources are assigned to it, making a virtual allocation of hardware resources. Resources can be added and removed on demand as needed, which enables a more efficient use of hardware resources.

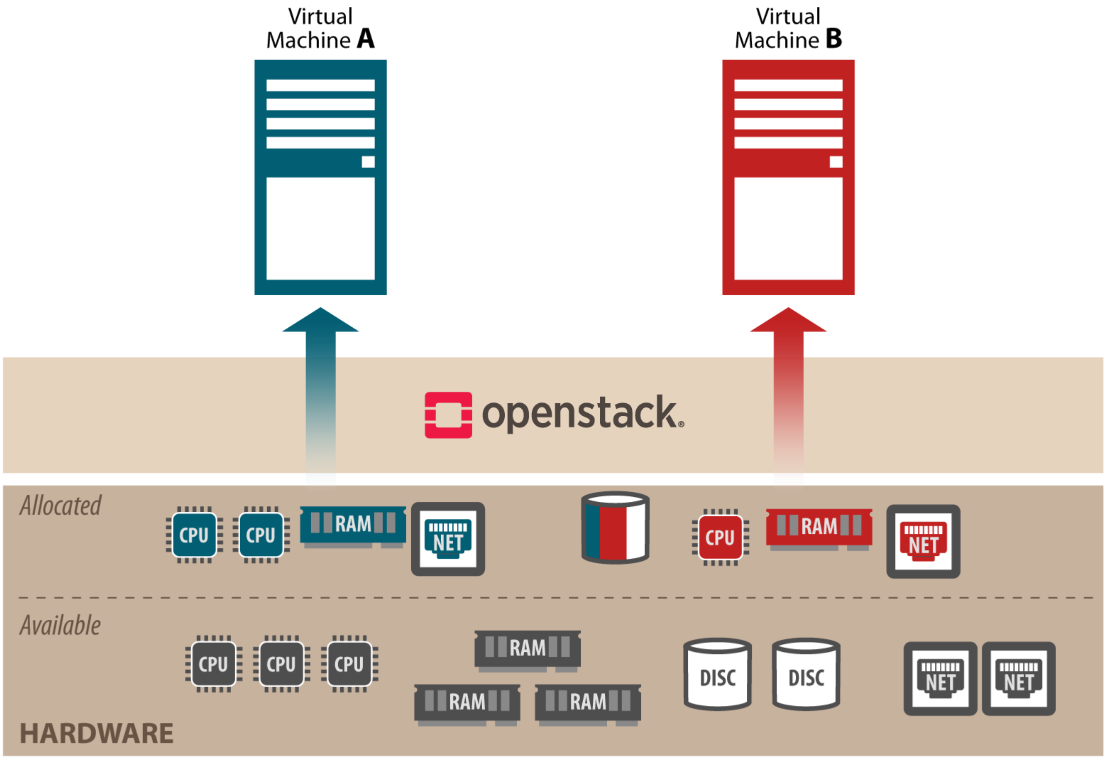


Figure 12: Simplified view of hardware allocation

In EUreka3D, the allocation of hardware is done by IM (see Section 2.2):

- It provides the virtual hardware according to the required specifications and the available images.
- It installs and configures the Operating System based on a predefined image of the EGI cloud.

4.1.2 Hardware Requirements

The EUreka3D platform uses the resources specified in Table 1. It shows the actual hardware allocation, but there is a maximum allocation planned for the project that can be used in case of increased workload.

Table 1: EUreka3D platform hardware allocation

NAME	TYPE	vCPU cores	RAM	STORAGE	OS
Servers	1 VM (up to 2)	16 (up to 32)	56 GB (up to 128 GB)	1 TB	Ubuntu 22.04.2 LTS
Data storage	S3 Object Store	-	-	30 TB (up to 300	-

				TB)	

4.1.3 Other infrastructure provisioning

In addition to the hardware allocation, two other elements are provisioned:

- **Domain Name System (DNS)**: The naming system for addressing machines in a human-friendly way. The EGI Dynamic DNS service[36] provides this functionality for EUreka3D.
- **Transport Layer Security (TLS) certificates**: The software artefacts containing cryptographic properties to enable secure network communication between users and applications. Certificates must be issued by a third-party that is globally trusted, and bespoke applications in EUreka3D use Let's Encrypt[37] for this.

4.1.4 Application deployment

There are multiple services, systems and applications used in EUreka3D. Some of them are existing tools that are already implemented and deployed. They may customise some features for EUreka3D, but they follow their own deployment procedures that are not under the direct control of EUreka3D. These mainly include Check-in, DataHub, Cyfronet's OpenStack, the Infrastructure Manager and the Europeana Portal.

However, there is a bespoke application under the control of EUreka3D that runs on EUreka3D cloud servers in order to assist in the visualisation of 3D models. Figure 13 presents the different elements involved in the execution of the application. The hardware (virtual server) is provisioned by the EGI Cloud Compute. Inside this server, the operating system, Ubuntu Server, constitutes the first layer of software. On top of it, the container runtime (Docker) is executed to allow the execution and management of containers (see Section 2.2). Both the virtual server and these two layers are installed with the support of the Infrastructure Manager, which greatly simplifies its deployment.

On top of the container runtime, a container is run with the application. It is based on two layers, one for the application server (NodeJS) and one for the final EUreka3D application written in JavaScript. The deployment is done through a container image stored in Docker Hub[38], a software repository widely used for container images in OCI[39] format. Since the application is not extensively developed, the deployment process of this element is conducted manually, when a new release is created.

---

36 https://docs.egi.eu/users/compute/cloud-compute/dynamic-dns
37 https://letsencrypt.org/
38 https://hub.docker.com/
39 Open Container Initiative

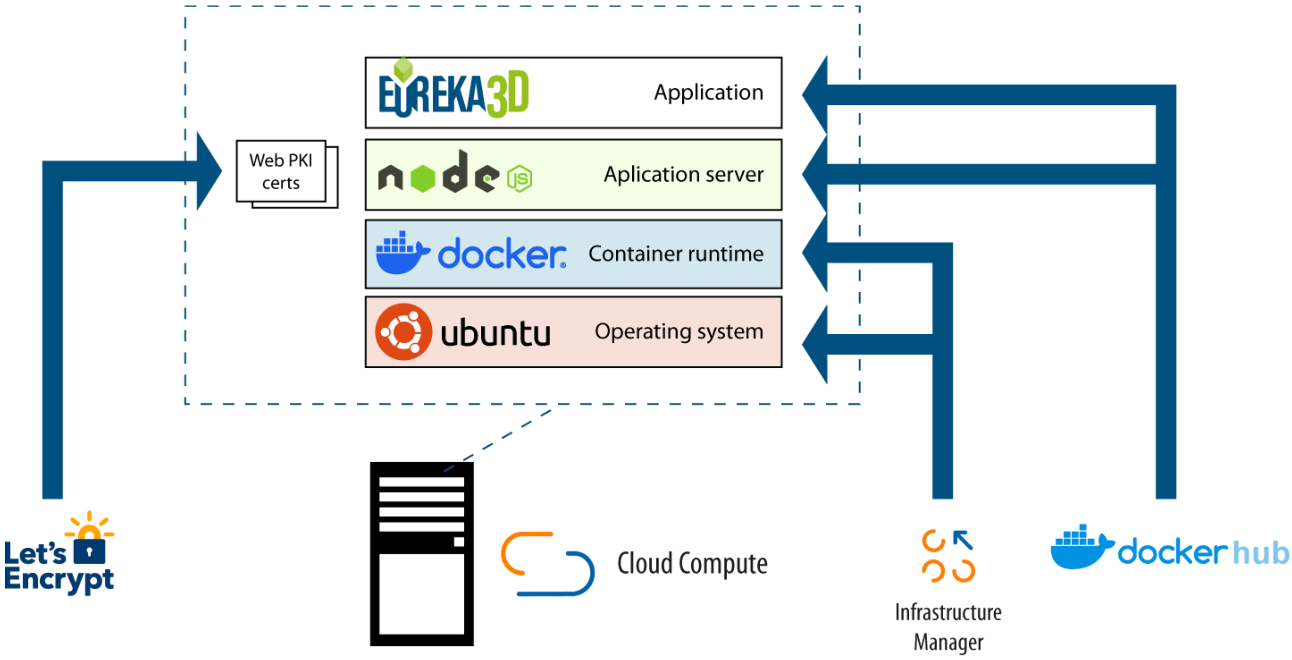


Figure 13: Deployment of the EUreka3D application to support 3D visualisation

The EUreka3D application provides an open source library for 3D visualisation, and an endpoint for the oEmbed protocol. These are used to support the 3D visualisation that takes place on the Europeana website.

Finally, the application server uses a certificate, delivered by Let's Encrypt, to secure the communication with users with the HTTPS protocol (basically, TLS on top of HTTP). This is also discussed in Sections 4.1.3 and 4.4.1.

4.2 Data Management

The data management aspects of the EUreka3D resource hub are realised by the EGI DataHub service — an instance of the Onedata distributed data management platform. While Onedata had already offered many crucial data management features before the EUreka3D kick-off, considerable work has been invested in improving it and adding functionalities catering for this project. Those developments have been integrated into the Onedata codebase, which means they can also be used in other ecosystems apart from EGI DataHub. This is an added value to the open-source Onedata software that will outlive the scope of this project and may be used for the good of the scientific communities.

In the scope of the EUreka3D project, EGI DataHub enables CHIs to access a virtual data space and use storage and computing resources to manage their 3D assets in a secure and easy-to-use manner. The range of Onedata data access interfaces and APIs makes it possible to quickly develop middleware for integration with information systems used in CHIs and automatisation of processes.

The data management solution that was built within the EUreka3D project, apart from serving a practical purpose, is intended to be an illustrative use case, validated with 3D data from CUT, CRDI, Bibracte, Museo della Carta, and additional collections from associated partners. It demonstrates that it is possible to offer an innovative, cost-effective, and competitive solution for data storage and online delivery of cultural heritage assets, providing CHIs with a secure and flexible environment.

4.2.1 Data organisation

The EUreka3D project has been assigned a single Onedata Space to store all the CH assets, coming from different organisations and working groups. This decision was argued by the need to oversee and harmonise the content produced by collaborating parties at the early stages of the project. From the technical point of view, the granularity and organisation of data collections can be arbitrary, one can imagine setting up multiple CH-related Spaces with the growth of content.

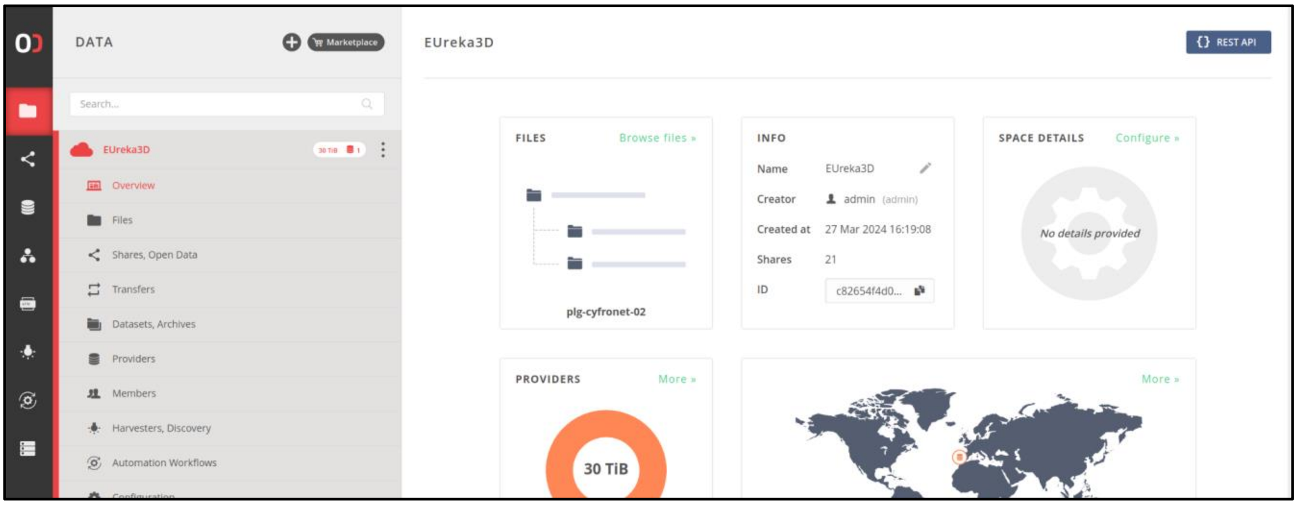


Figure 14: For EUreka3D community members, the Space is available immediately upon login.

User membership in the EUreka3D Space is set up automatically thanks to the group mapping mechanism. Upon login via EGI Check-In, user memberships are propagated and mapped to Onedata groups. If he/she belongs to an EGI group that's subject to the EUreka3D VO, he/she will inherit access to the space as a member of the corresponding Onedata group.

The Space structure is built around the concept of *projects*. Every *project* is responsible for curating and publishing a single CHO. The Onedata's automation engine helps keep a unified structure, as described in the "PLAN" lifecycle phase subsection below.

Figure 15: Contents of the EUreka3D Space as viewed using the Onedata Web file browser.

4.2.2 Data access control

For determining general access to resources, Onedata uses a graph of memberships with hierarchical inheritance, where groups can be nested. Each membership may be regulated by privileges that define what operations are allowed for a member. From the practical point of view, it means that, for example, a user of "EUreka3D-CUT" group inherits the group's membership in the EUreka3D Space, along with the group's privileges toward it. Users not belonging to the Space cannot see or access it in Onedata.

By default, all members of the same Space may access each other's data and collaboratively manage it. Several mechanisms allow further regulation of permissions: space-level privileges, POSIX permissions and ACLs (Access Control Lists). In EUreka3D, ACLs are typically used for this purpose. An ACL is a list of entries referring to users or groups and denying or allowing specific operations. If a group is assigned permissions, they are applied to every group member (including nested groups) on the terms of inheritance. Based on an ACL, all the users with granted access can collaboratively manage, modify and curate the data within a EUreka3D project, but other members of the community cannot access the data. Typically, the permission management granularity is kept on the organisational level (groups representing participating organisations are automatically mapped via EGI Check-In).

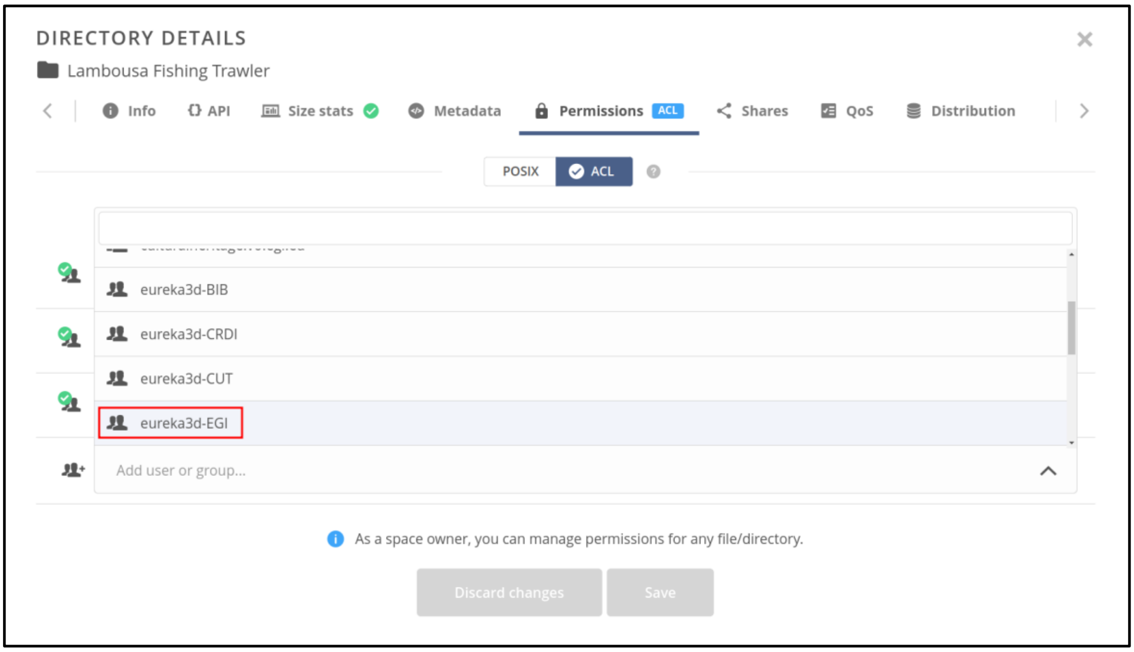


Figure 16: Onedata's ACL editor — assigning permissions to all members of the EGI organisation within EUreka3D.

4.2.3 Data lifecycle

EGI DataHub comprehensively covers the data lifecycle phases of a CHO, starting from the point when its representation is acquired as a digital object that can be stored in a filesystem. The lifecycle can be quite well visualised using the diagram popularised by RDMkit[40]:

[40] https://rdmkit.elixir-europe.org/

Figure 17: "Data life cycle diagram" by RDMkit, licensed under CC BY 4.0.

Plan

Thanks to Onedata's automation engine, it is possible to build workflows that execute arbitrary functions (typically implemented in Python) to run repetitive tasks that can interact with the Onedata filesystem. For the needs of EUreka3D, a new workflow has been implemented, which initialises a standardised project structure for a new CHO. This allows the imposition of a unified way of storing and organising CHO data within the Space. In addition, the workflow automatically sets proper ACL rules based on the user inputs regarding the managing group(s) of the project.

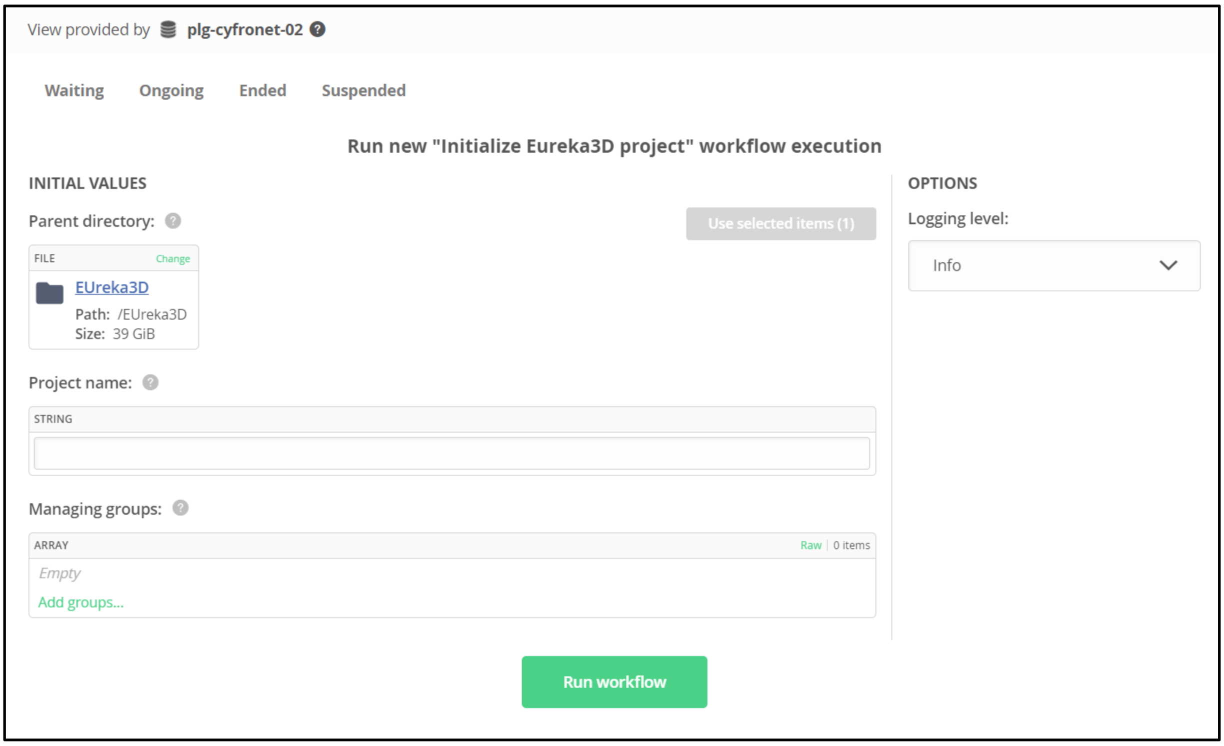


Figure 18: Automation GUI — collecting user input for the project creation workflow.

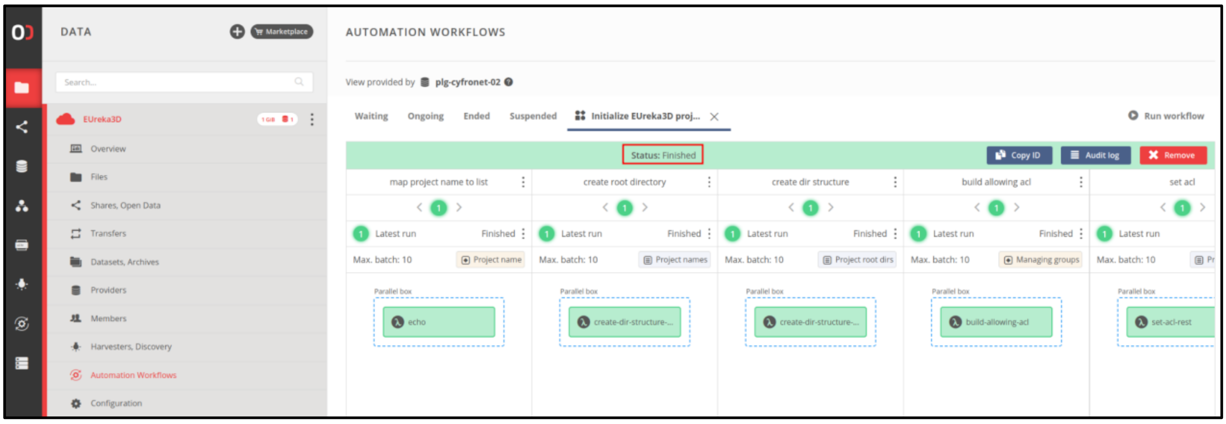


Figure 19: Workflow execution view after a successful run; the project has been set up.

Users responsible for the publishing and management of digitised CHOs are additionally given a handbook (cf. Annex A) that guides them through a well-established process.

Collect

Onedata offers a range of data access interfaces: Web GUI, REST API, CDMI API, fuse-based POSIX, Pythonic libraries, and S3. Regardless of the interface used, the users get a unified view of their data. In EUreka3D, Web GUI is the commonly used interface for uploading digital objects into the system, as most users in the community are not computer-savvy. It is intuitive and follows well-established patterns of a file browser UI, known to users from operating systems and cloud services. In the case of large collections of digital objects (over tens or hundreds of records), manual ingestion becomes infeasible. To address that, scripts using the REST API have been devised to automate the process (including also the subsequent data lifecycle phases).

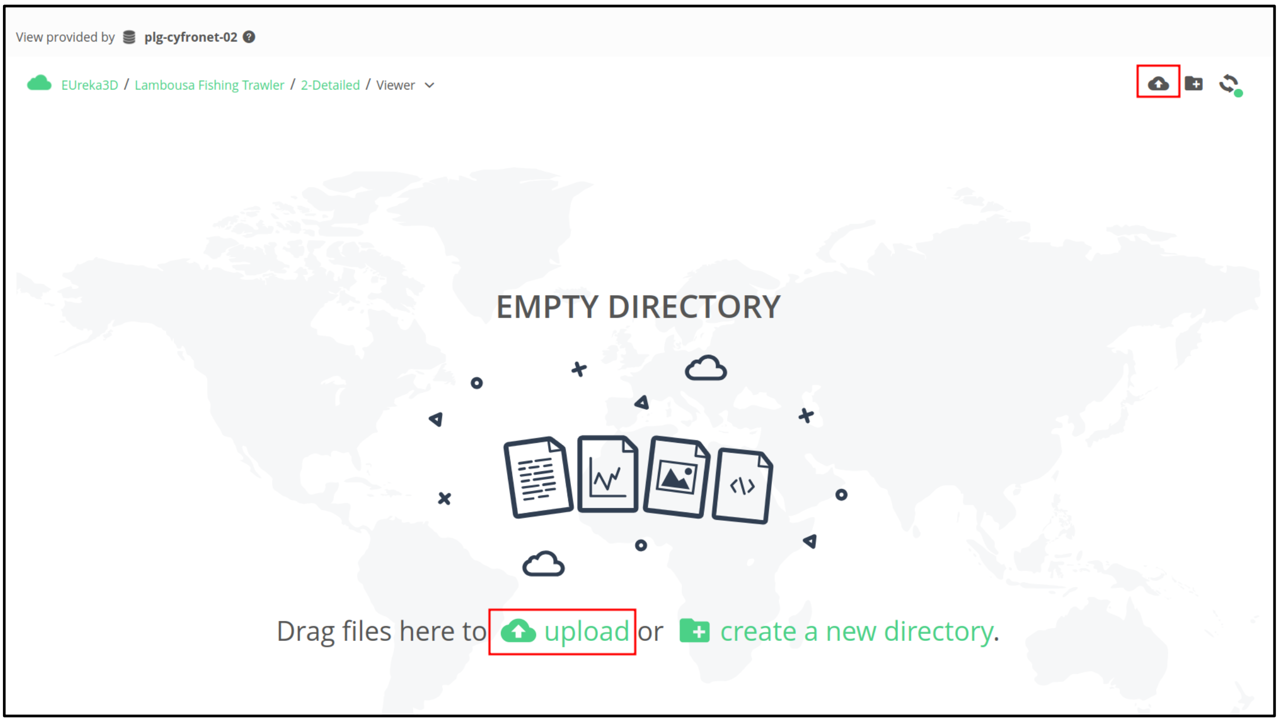


Figure 20: Uploading data using the Web GUI.

Process & Analyse

If needed, the CHO data can be transformed and analysed using the automation workflows. For example, it is possible to create tasks that will unify the 3D representation formats and quality of digital objects (*process*), sanitise the ingested data to make sure it conforms to guidelines (*analyse*), or annotate objects with metadata that can be inferred using automatic machine processes (*analyse*).

Due to the fact that the Onedata filesystem can be mounted on worker nodes of a computing centre and interacted with using interfaces like S3 or Python's *pyfilesystem* library, it is straightforward to run third party software and custom-tailored middleware that will perform required processing and analysis of the CH data.

Preserve

Onedata implements a concept of datasets and archives that can be used for the incremental preservation of evolving data. Datasets are essentially files or directories that contain relevant data collections. At any time, a dataset can be snapshotted as an archive — a persistent copy stored for long-term preservation. Archives are immutable and read-only. By applying a dataset structure over the EUreka3D Space and archiving curated digital objects, it is possible to control the preservation of CHO data in an organised, controllable, and granular manner.

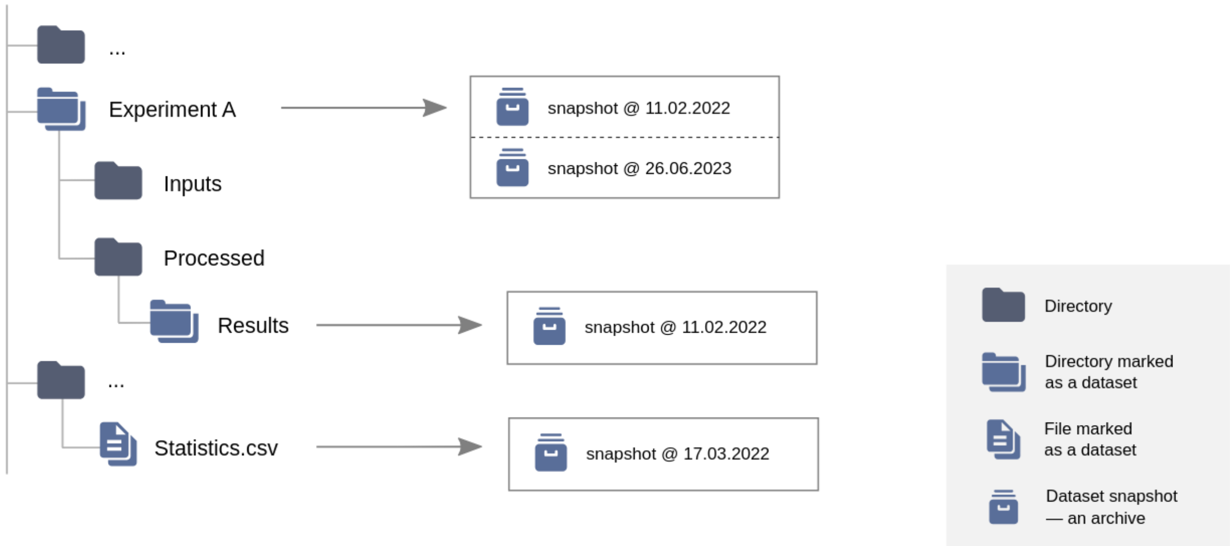


Figure 21: Datasets within the Onedata filesystem and archives created from them (snapshots).

Share

Sharing of the data can be done on multiple levels:

1) Collaboration with other members of the EUreka3D community — within the EUreka3D Space, users may regulate the permissions using ACLs, as described in Section 4.2.2.
2) Semi-public link sharing — allows creating a link that grants read-only access to a file or directory for non-authenticated clients. In EUreka3D, this mechanism is used to publish representative images of CHOs.
3) Public data collections — a public data record is an extended share that has been assigned a PID/DOI[41] persistent identifier (publicly accessible). It has a standardised XML metadata (e.g. Dublin Core or EDM) attached and is exposed via OAI-PMH for indexing, which makes it findable. These mechanisms are used

[41] EGI DataHub supports both PID and DOI which makes possible in the future to use DOIs in EUreka3D

to publish digitised CHOs and expose them for harvesting by Europeana (cf. section 4.3). In EUreka3D, the EUDAT B2HANDLE service is used to generate PIDs for the records[42].

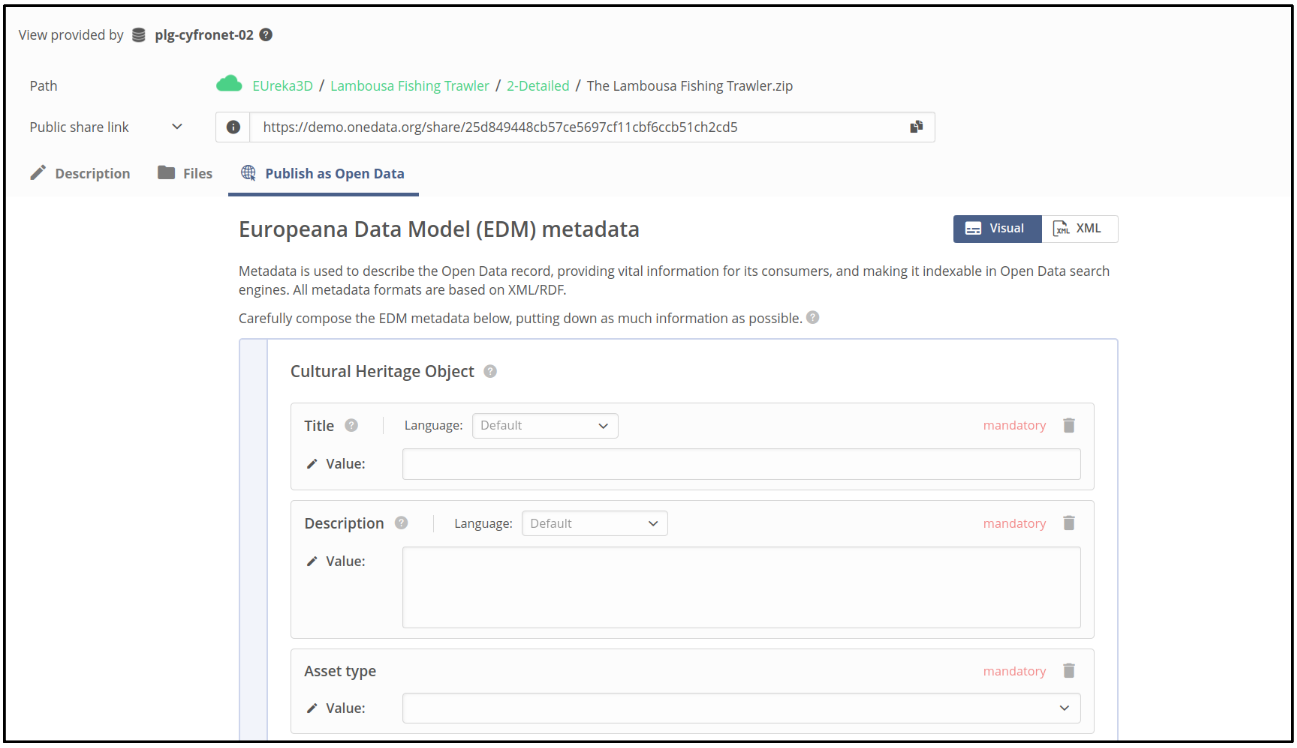


Figure 22: Onedata UI for data collection publishing — users are presented with an advanced form that guides them through metadata curation.

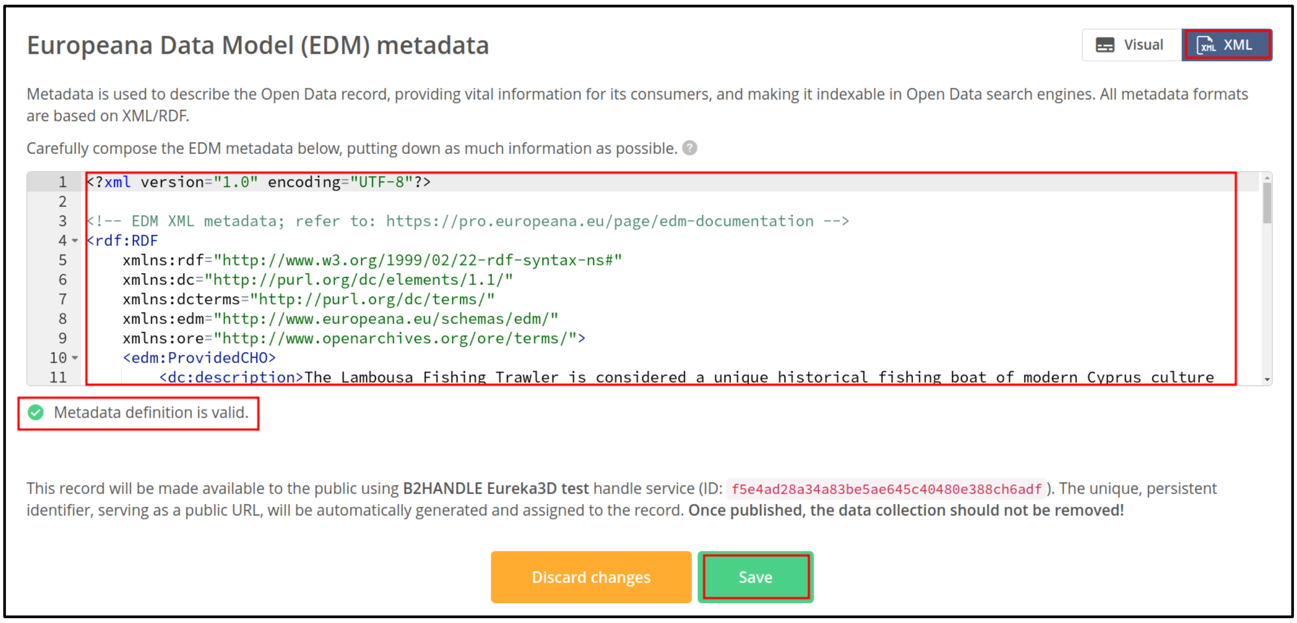


Figure 23: For advanced users, the metadata editor offers an XML source editor that is synchronised with the visual form.

[42] https://eudat.eu/service-catalogue/b2handle

Reuse

The reusability and accessibility of digitised CHOs stand at the foundations of the EUreka3D project. The digital objects and their metadata are indexed in publicly available services, which is possible thanks to their exposition via the OAI-PMH protocol. Users may view their visualisations and download the underlying models and related data using Onedata's data sharing and publication mechanisms. Full integration with the Europeana systems is provided, to grant a seamless publication of the records in europeana.eu (cfr. Section 4.3).

4.2.4 Summary of efforts

Table 2 summarises all features and aspects of the Onedata system (powering EGI DataHub) used within EUreka3D and classifies them into three categories:

- preexisting — existed before EUreka3D, did not require additional efforts,
- improved — have been improved over the course of this project,
- new — have been implemented specifically for the needs of EUreka3D.

Table 2: Onedata features and aspects used within EUreka3D

Category	Aspect	Comment / description
preexisting	OIDC/SAML integration	Onezone supports OpenID Connect and SAML 2.0 protocols, allowing for a quick setup of multiple login options. EGI DataHub is integrated with the EGI Check-In service.
preexisting	Group mapping	Group memberships can be transmitted as a part of user information over OIDC/SAML, and then automatically reflected as groups in Onedata. EGI DataHub embraces this mechanism to reflect EGI VOs and groups and organise seamless access to Spaces.
preexisting	Virtualised data access	Physical data stored on diverse storage systems is accessible via a unified POSIX-compatible virtual filesystem.
preexisting	Collaborative data sharing	Users and groups of different affiliations can collaborate on distributed datasets, working simultaneously in different geographical locations.
preexisting	Various data access interfaces	Onedata virtual filesystem can be accessed via Web GUI, REST API, CDMI API, fuse-based POSIX, Pythonic libraries, and S3 protocol.
preexisting	Custom metadata and discovery	Files and directories can be annotated with custom metadata (extended attributes, JSON, JSON-LD, RDF), which can be later used for data discovery.
preexisting	ACLs	Method for regulating permissions to data within a Space.
preexisting	Datasets and archives	Support structural organisation of data and the possibility to make incremental archives (snapshots) of the evolving data.
preexisting	Semi-public shares	Data collections can be shared using a link that allows unauthenticated read only access, but cannot be guessed or discovered.
preexisting	Public data collections	Shares can be converted to public data records that are annotated with standard XML-based metadata (e.g. DC, EDM).

preexisting	Integration with PID/DOI services	Automatic PID/DOI minting for public data collections; the integration is pluggable, i.e. adapters for different persistent identifier services can be implemented
preexisting	OAI-PMH repository	Onezone serves as an OAI-PMH repository and advertises the Public data collections via the protocol for harvesting and aggregation.
preexisting	Automation engine	Can be used to run repetitive tasks interacting with the Onedata filesystem on large data collections.
improved	Share and Public data management	Improved browser of Shares and Public data records, convenient functions for quick publishing and management.
improved	UX of the Public data view	Various user experience improvements to the Public data view.
improved	PID deregistration	Improved integration with the B2HANDLE service (for PID minting) that allows for deregistration (deletion) of identifiers.
improved	OAI-PMH — optimised performance	Quicker responses to OAI-PMH queries thanks to the minimised number of internal database queries and regulation of response sizes.
new	EDM metadata type support	Support for the Europeana Data Model metadata on the Public data collection level, in the system back-end and APIs.
new	EDM metadata editor	Rich metadata editor guiding users through curation of the EDM metadata for Public data collections (published CHOs).
new	Automatic fields in EDM metadata	Back-end logic for transformation of user-provided EDM metadata and enrichment with additional, automatically generated information. Uses as the basis for integrating visualisations (cf. Section 4.3.3).
new	Representative image for published CHOs	Support for displaying representative images for published CHOS.
new	Customised public data view	Embedding of EUreka3D and Europeana logotypes in the public (unauthenticated) view of published CHOs.
new	Project structure creation workflow	Corresponding workflow schema and lambda implementations.
new	Group data type in automation	Support for groups as the data type for automation tasks, required to automatically set ACLs for project's managing groups.
new	OAI-PMH — HTTPS endpoint	Exposition of the OAI-PMH endpoint on HTTPS secure channel (apart from the preexisting HTTP that's required by the OAI-PMH specification).
new	OAI-PMH — resumption tokens	Implementation of support for resumption tokens during large listings of records via OAI-PMH, crucial for repositories with millions of records. Results can be split into pages with limited size.
new	OAI-PMH — persistent deletions history	Implementation of the optional, but important for OAI-PMH repository integrity, support for listing deleted records.
new	Direct download URLs	Convenient feature to quickly obtain direct download links to shared and

	for shared files	publicly available files and directories.

4.3 Europeana integration

To make the digitised CHOs useful to the public and discoverable, the data management system (EGI DataHub), at the core of the EUreka3D infrastructure, has been integrated with Europeana. The backbone for this integration is the standard OAI-PMH protocol. Additional interoperability is achieved by embracing some of the fields of the EDM metadata, as described further on (cf. section 4.3.3).

Every CHO in the Onedata Space dedicated to EUreka3D, that is exposed as a public data collection (and so assigned EDM metadata and a PID), is advertised via OAI-PMH. Different groups in the EUreka3D community are assigned different OAI-PMH sets, making it possible to selectively harvest records from a single organisation and track its provenance. An exemplary OAI-PMH query listing EDM records in EGI DataHub can be done using a typical web browser[43] (the XML response should be rendered).

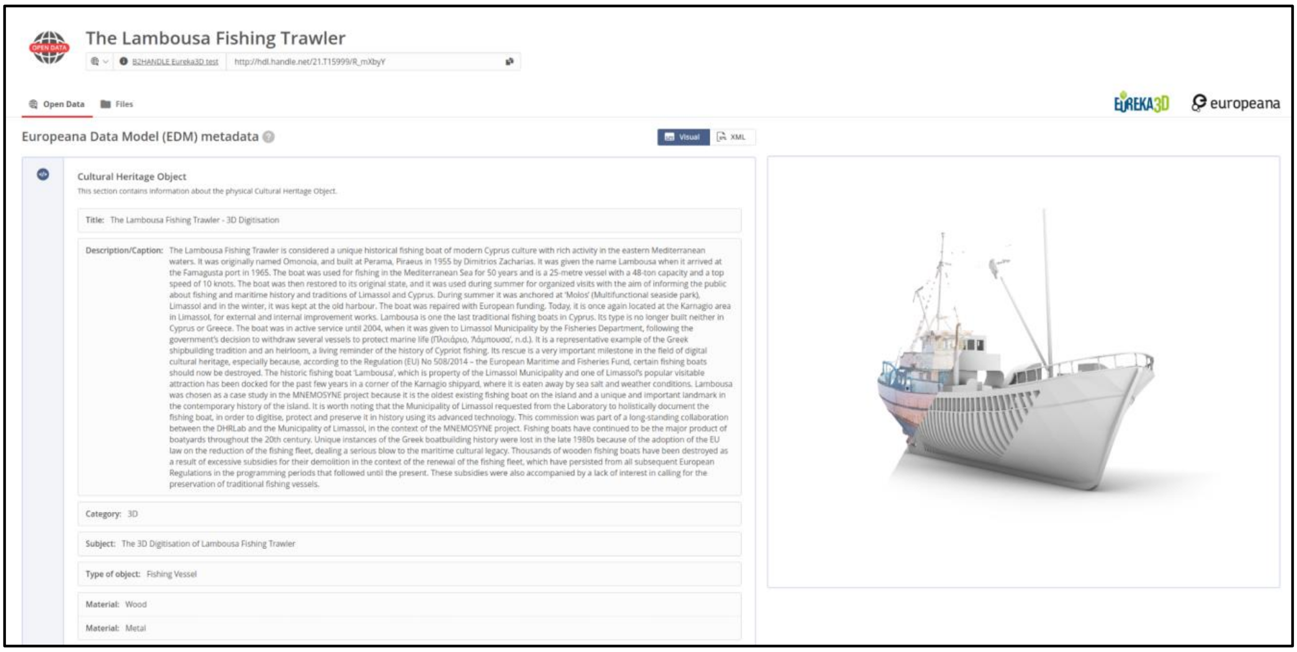


Figure 24: Public (unauthenticated) view of a CHO record published[44] in EGI DataHub.

4.3.1 Architecture

The OAI-PMH protocol defines the concepts of repositories and harvesters. Harvesters periodically query the repositories using so-called selective harvesting, defining parameters for filtering the records that should be returned. One of the basic parameters is the time window of the query, allowing for incremental scans that pick up only the changes that appeared since the last harvesting. In EUreka3D, the EGI DataHub acts as a repository and Europeana acts as a harvester, but this does not exclude the existence of different entities that interact with the services in the meantime.

[43] https://datahub.egi.eu/oai_pmh?verb=ListRecords&metadataPrefix=edm

[44] http://hdl.handle.net/21.T15999/R_mXbyY

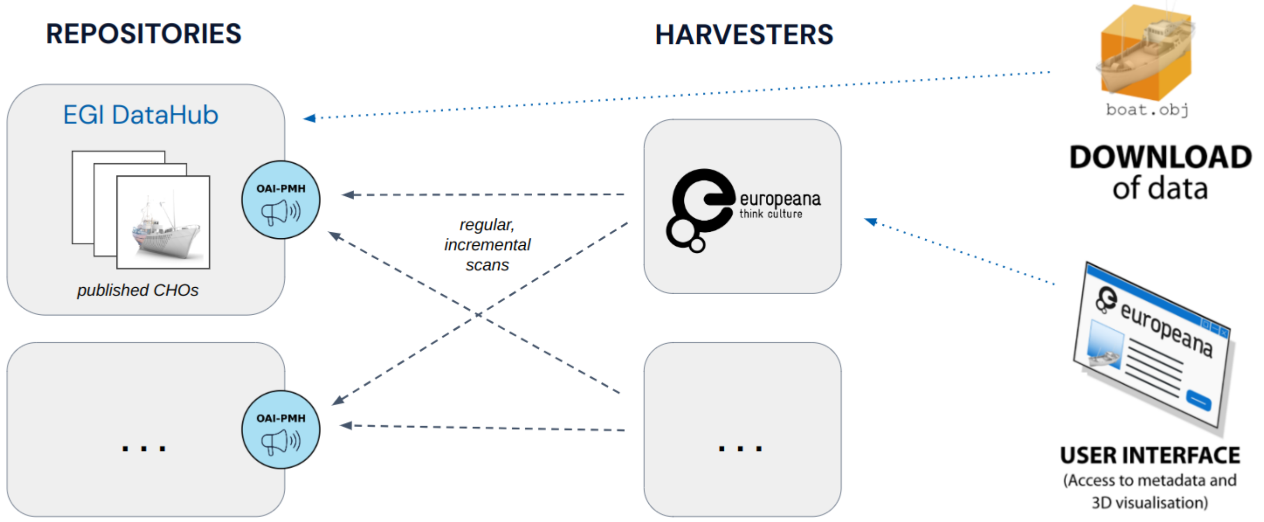


Figure 25: Architectural overview of the integration done for EUreka3D, based on OAI-PMH.

4.3.2 Searching and accessing CHOs

Once Europeana does a consecutive scan of the DataHub's OAI-PMH endpoint and picks up new records, they are incorporated into its indices. Then, the records can be found using the search engine.

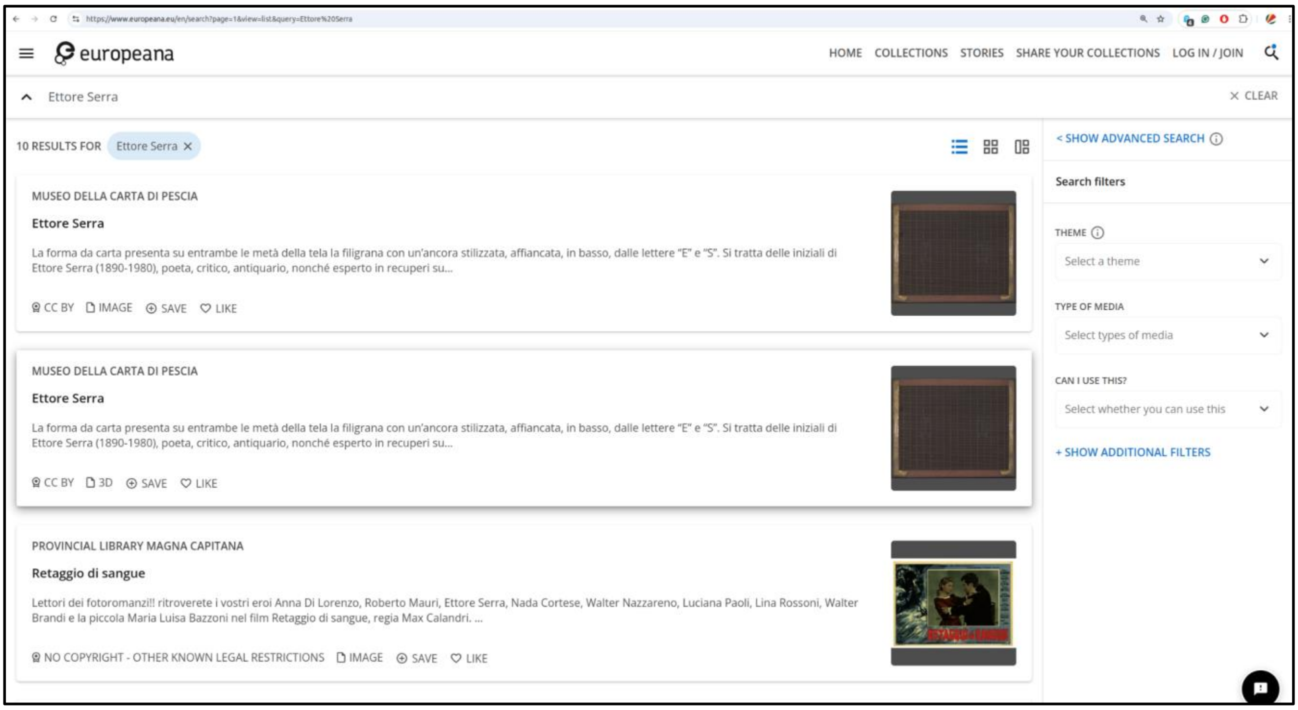
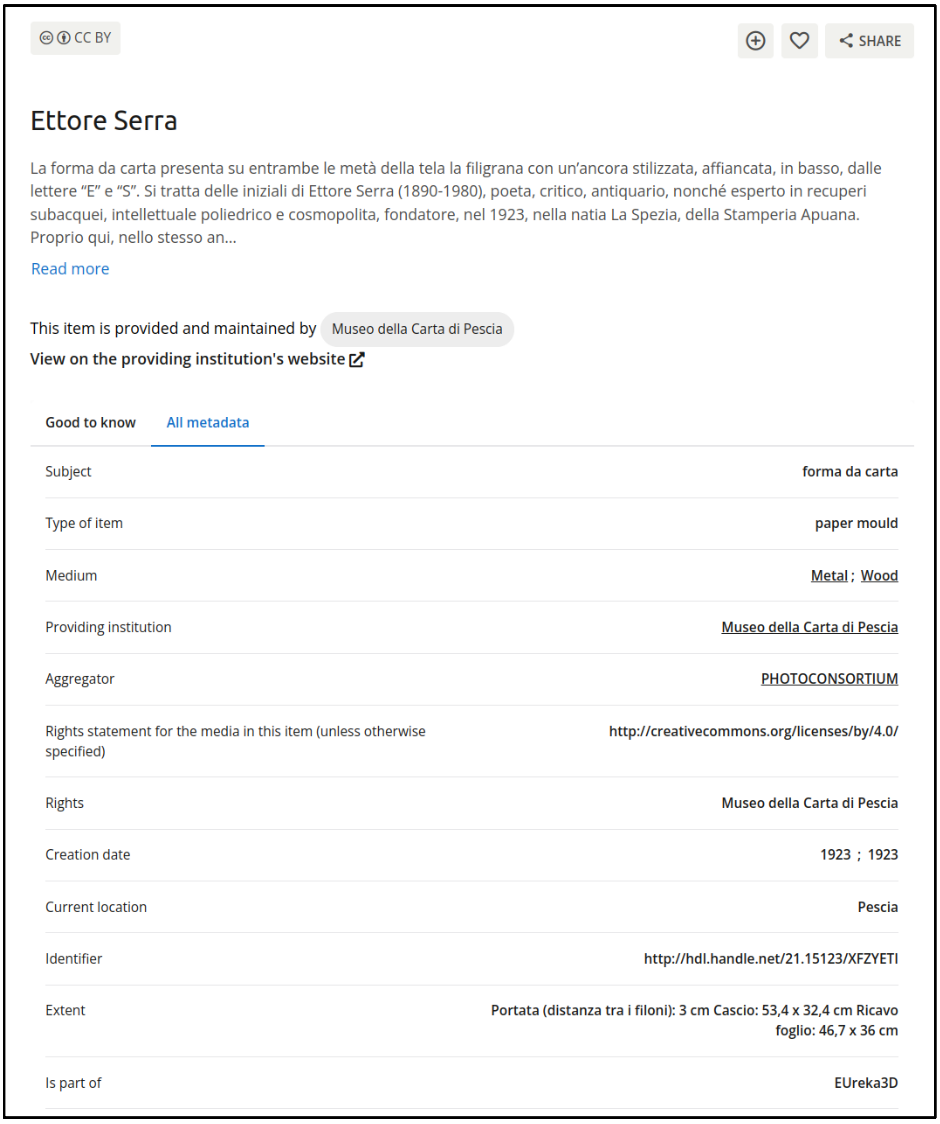


Figure 26: Exemplary search on the Europeana website

One can click on a desired record to view its details, which are essentially a human-friendly visual and textual representation of the EDM metadata that was attached to the record.


CC BY

SHARE

Ettore Serra

La forma da carta presenta su entrambe le metà della tela la filigrana con un'ancora stilizzata, affiancata, in basso, dalle lettere "E" e "S". Si tratta delle iniziali di Ettore Serra (1890-1980), poeta, critico, antiquario, nonché esperto in recuperi subacquei, intellettuale poliedrico e cosmopolita, fondatore, nel 1923, nella natia La Spezia, della Stamperia Apuana. Proprio qui, nello stesso an...

Read more

This item is provided and maintained by Museo della Carta di Pescia

View on the providing institution's website

Good to know | All metadata

Subject	forma da carta
Type of item	paper mould
Medium	Metal ; Wood
Providing institution	Museo della Carta di Pescia
Aggregator	PHOTOCONSORTIUM
Rights statement for the media in this item (unless otherwise specified)	http://creativecommons.org/licenses/by/4.0/
Rights	Museo della Carta di Pescia
Creation date	1923 ; 1923
Current location	Pescia
Identifier	http://hdl.handle.net/21.15123/XFZYETI
Extent	Portata (distanza tra i filoni): 3 cm Cascio: 53,4 x 32,4 cm Ricavo foglio: 46,7 x 36 cm
Is part of	EUreka3D

Figure 27: Details of a CHO as displayed on the Europeana website

4.3.3 Visualisation with the 3D Viewer

An important aspect of the integration is the embedding of the 3D viewer in the Europeana website. When a page with details of a CHO record is accessed, an interactive viewer is displayed. It loads the representation, accessing the publicly shared 3D model file in the EGI DataHub behind the scenes, and renders a preview. The user may control the rotation, viewpoint, and zoom with mouse controls.

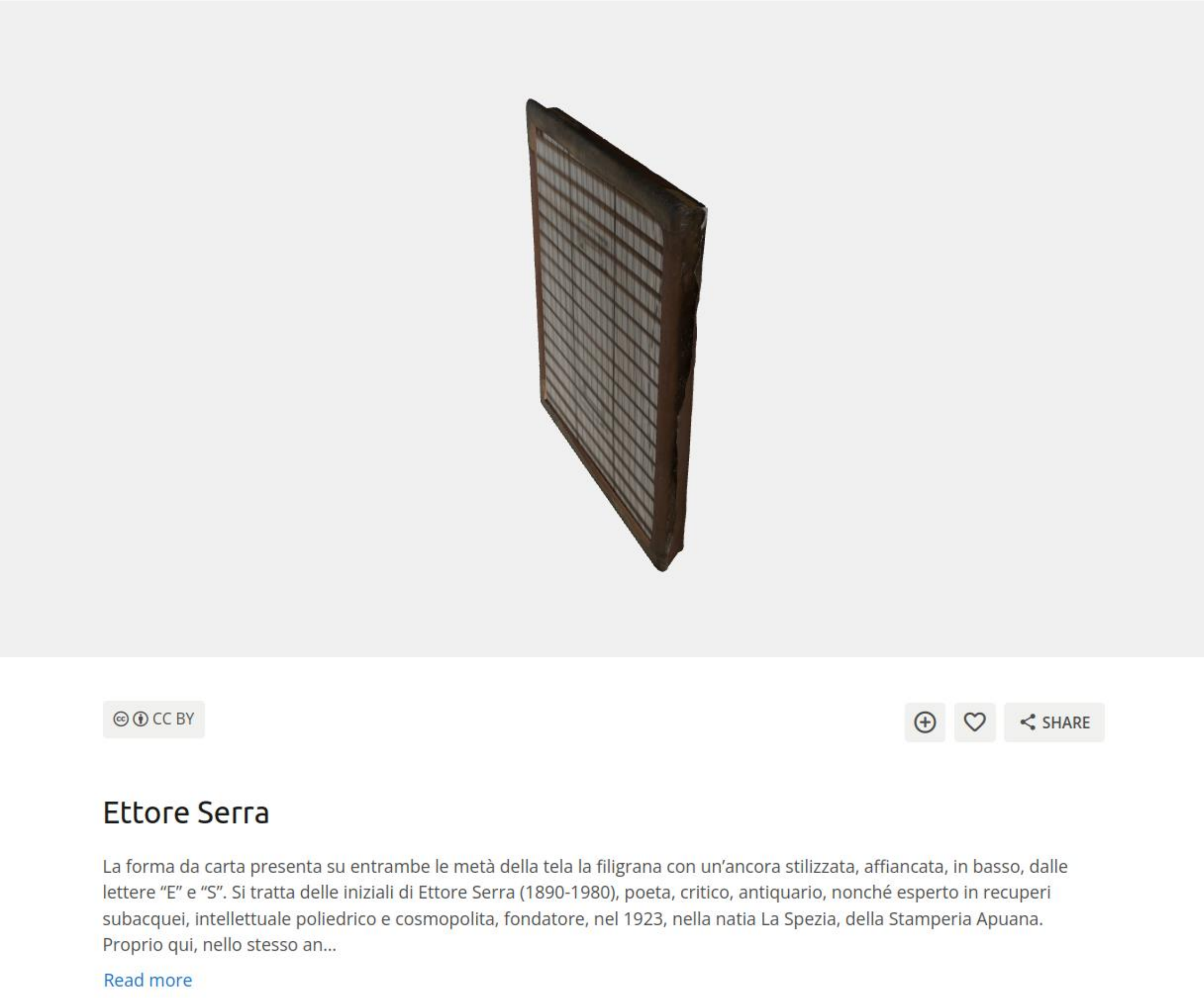


Figure 28: The 3D viewer visualisation embedded in the Europeana website.

This integration is achieved thanks to the `edm:isShownBy` field of the EDM metadata. This field is filled automatically by DataHub when a CHO record is published and contains a URL to the 3D viewer. The URL includes a public file ID of the shared model file stored in DataHub. This way, when the URL is visited, the information on what 3D file should be visualised is injected into the viewer application.

```
<edm:isShownBy rdf:resource="https://eureka3d.vm.fedcloud.eu/3d/0000000
0007E30F673686172654775696423346431303665396635313732396365666234656635
396333643537646135656463686332366523366566363735636632313434366132636666
562356538663461393932333331656368656632652364633537346665656130633736 35
3235353035326134363232316632396563646368373534 38"/>
```

Figure 29: Exemplary value of the `edm:isShownBy` field,
carrying the URL to the viewer application with the public ID of the 3D model file.

4.4 Security

Security is an important and centric part of the technology behind the EUreka3DEUreka3D services and resource hub, as important as the development, implementation, user experience and performance of the services provided. Security must be carefully considered from the design phase of the project, and will be monitored and developed, where necessary, throughout the project and during the lifetime of the data. This section does not intend to present a full security assessment, but to highlight some of the security mechanisms implemented to protect the major asset of the project: the data.

4.4.1 Data in transit

As usual on the Web, data in transit is protected by using the TLS protocol, which guarantees data confidentiality (the data cannot be read by an external actor) and integrity (the data cannot be altered by an external actor without being detected) during the transit of data over a network. Figure 30 shows an example of this communication. A user connects through a Web browser to a system within the EUreka3D ecosystem, such as the bespoke Web application running on a Virtual Machine or the EGI DataHub service. This data exchange is encrypted and protected in a TLS secure channel, with the help of certificates used by the respective server applications. To validate that these certificates are legitimate, they have to be certified by an external party, in this case *Let's Encrypt* (cf. Section 4.1.3). The validation process takes place in the user's environment, who is able to identify whether the certificates have been certified by a trusted third party.

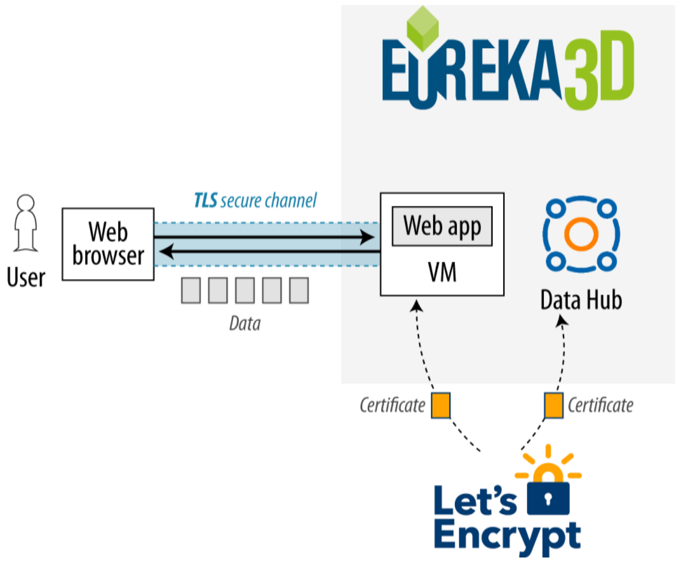


Figure 30: Encrypted data communication

4.4.2 Data at rest

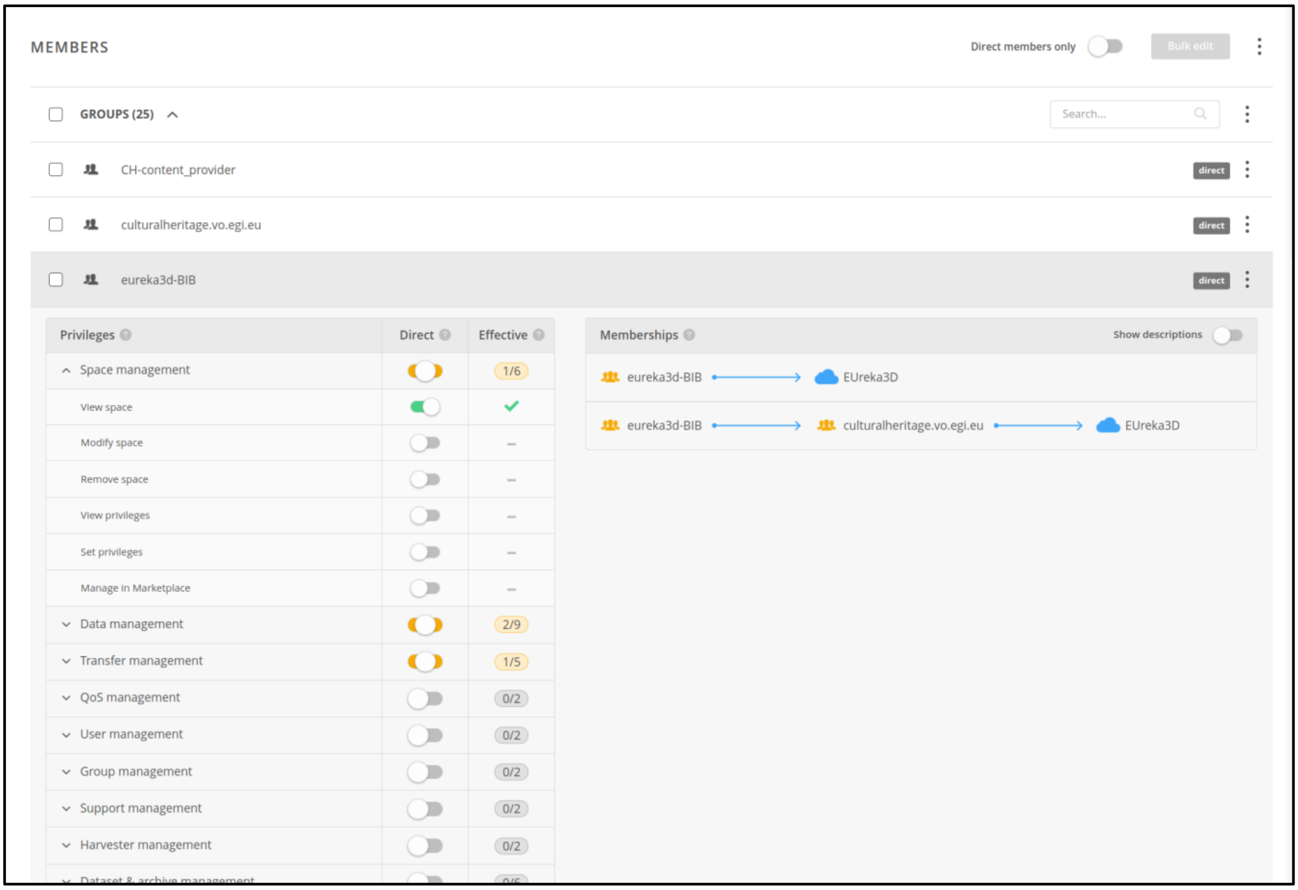
Data at rest is not encrypted by EGI DataHub, so the confidentiality and integrity of the data is delegated to the security limits presented by the access to the virtual machine(s) with access to the disk where the data is stored and to the applications that have the means to access the data. Users can upload data in an encrypted form, but this is not a requirement in EUreka3D and will just over complicate the system. However, due to the federated architecture of Onedata, users can explicitly specify QoS rules defining where the data can be stored (for instance by specifying geographical or network constraints), thus ensuring that data is only stored in trusted data centres.

4.4.3 Data access

Data access is protected with EGI Check-in[45], an identity management system that uses OIDC[46]/OAuth 2 (amongst other technologies) to protect Web resources. EGI Check-in acts as an *authentication proxy*[47] and as an attribute management system, delivering authorisation information to applications to enable them to make authorisation decisions. More details about this can be found in the Deliverable 3.2 "*The EUreka3D AAI architecture*".

A dedicated resource pool known as a Virtual Organisation (VO) has been created to configure the authorisation attributes and to organise the community of users. The resource pool allocated to the EUreka3D project is named ***culturalheritage.vo.egi.eu***[48], and is intended to be used to represent the Cultural Heritage community beyond the project.

Users authenticated in EGI Check-in can then proceed to EGI DataHub, where they can login using their credentials and access their storage resources through the Onedata Web interface, or generate an access token and use any of the supported data access mechanisms. The data access can be controlled using fine grained settings available in the web GUI or through a REST API. Figure 31 presents the authorisation settings as they are shown in the GUI.



[45] https://www.egi.eu/service/check-in/

[46] https://openid.net/connect/

[47] An intermediate system between the user and an organisation that authenticates the user.

[48] https://operations-portal.egi.eu/vo/view/voname/culturalheritage.vo.egi.eu

Figure 31. Onedata data access authorisation settings

5. EUREKA3D SERVICES FOR EOSC

5.1 EUreka3D research products and services

The development of EUreka3D has offered different research outcomes of interest for the CH community. Table 3 summarises these outcomes as exploitable results. The most noteworthy result is a stable platform providing services and resources oriented for the CH discipline and ready to be used by CHIs. This serves multiple purposes:

- It offers a solution for small CHIs that cannot afford to manage their own servers or storage.
- It contributes to the transformation of CH, enabling CHIs to use the cloud to store and manage their assets.
- It offers a European alternative, research oriented, to commercial products.
- It serves as an enabler for the publication of CH objects in Europeana.

The EUreka3D platform not only provides storage for the different 3D assets, but also metadata management, sharing, publication, and security to protect the data from unauthorised access. In addition to the suite of services and resources, other outcomes of EUreka3D can be of great interest to research communities.

Table 3: EUreka3D services and resources

Exploitable result	Type	Description
EUreka3D platform	Service	The platform that provides a solution for CHI to store and share their CH objects.
3D models with its associated metadata and paradata	Research products	The 3D models published in EUreka3D, which can be part of a catalogue of research products to be used by professionals and researchers in the area of cultural heritage
Capacity Building results	Training and Education	These are the training materials created from the Capacity Building activity of the project (WP2).
Publications, articles, case studies	Research products	Different publications and articles that have disseminated the experiences gathered during the development of EUreka3D, in the shape of papers and blog articles.

The 3D models created and managed in EUreka3D constitute a dataset. They can be published and shared from the Web, together with their metadata and paradata. This process can be done through different protocols, including OAI-PMH, facilitating the creation of a catalogue of CH objects, in the same way as Europeana does. The 3D models not only preserve the cultural heritage, but enable their study and reuse by professionals in different fields, from research, tourism, the gaming industry or simple citizens.

The activities of the Capacity building in EUreka3D have produced results that can be used by the CH community for training and education, including training about paradata and different digitisation procedures to ensure the correct quality of digitised 3D models. Part of the EUreka3D outcomes is to share the experiences gathered during the project that can help other professionals in the field.

Finally, the project has produced multiple articles, papers, case studies and blog publications that can be shared as examples of good practices.

Altogether, the results of EUreka3D can be conceived as products available to be shared for further research.

5.2 The European Open Science Cloud (EOSC)

The original core of EOSC was a catalogue of "research products", such as datasets relating to various disciplines, and "services", such as EGI Check-in and EGI DataHub, used in EUreka3D. These could be submitted in a process called "onboarding". However, the EOSC landscape has changed dramatically in the last year[49] [50], due to a process of redesign initiated by the European Commission. The EOSC Portal, which is where researchers could find services onboarded to EOSC, is no longer available, and has been replaced by the **EOSC Federation of Nodes**[51], as depicted in Figure 32. The EOSC Federation is re-developed in the form of multiple EOSC Nodes that are interconnected and can collaborate to share and manage scientific data, knowledge, and resources within and across thematic and geographical research communities.

The different EOSC Nodes will be entry points for users to the EOSC Federation, with each node offering its own and possibly third-party services, including data reposing and services. The **EOSC EU Node**[52] is the first of the EOSC Federation. However, detailed information on how one can contribute to the EOSC EU Node and Rule of Participation is not yet available.

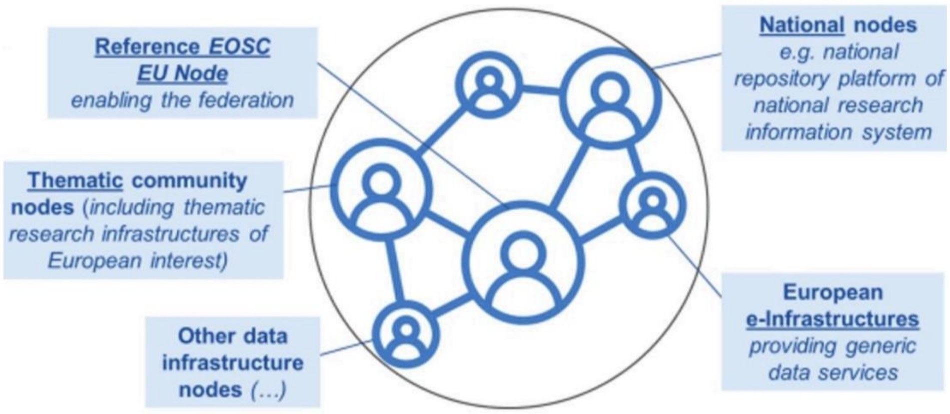


Figure 32: EOSC Federation of Nodes

Discussions over the individual EOSC Nodes that will complement the EOSC EU Node to initiate the formation of the EOSC Federation are still ongoing. The Tripartite Governance of EOSC recently launched a questionnaire[53] to gauge the scale and scope of general interest and readiness for the build-up phase of the EOSC Federation. Interested parties were invited to reply to the questionnaire as the first step in their

[49] https://open-science-cloud.ec.europa.eu/news/eosc-future-signs
[50] https://eoscfuture.eu/results/
[51] https://eosc.eu/building-the-eosc-federation/
[52] https://open-science-cloud.ec.europa.eu/
[53] https://eosc.eu/eosc-about/building-the-eosc-federation/questionnaire-for-parties-interested-in-contributing-to-the-build-up-phase-of-the-eosc-federation/

eligibility to contribute to a testbed of potential EOSC Nodes representing the different stakeholders involved in the research communities of Europe. EGI has replied to the questionnaire, focusing on providing a node with **core services** supporting **computing and storing**.

There have been several discussions about the future sustainability of EOSC, with a working group called the **EOSC Financial Sustainability Task Force**, which aims to develop a proposal for long-term financial sustainability of the main building blocks of EOSC (EOSC-Core, EOSC-Exchange and the Federation of Data and Data Services, as defined in the FAIR Lady report "Solutions for a Sustainable EOSC" [7]. The final report of the task force was published in April 2024 [8]. In the section "Key Principles for EOSC Financial Sustainability", the report concludes that brokered not-for-profit services, including horizontal and thematic services as part of EOSC Exchange, will constitute the majority of services and the true marketplace of EOSC. This is where the thousands of EOSC participants will offer services to each other. The report also notes that EOSC Marketplace needs to include mechanisms that will not only allow researchers to access these services but also cost recovery mechanisms for service providers, both for services that are "free at the point of use" for researchers as well as for those that require researchers to pay. These mechanisms should encourage service providers to join the marketplace and provide the best quality service possible at the best price. More importantly, the report notes that the usage of these services will not be centrally financed on the European level but via **national or institutional funds**, which should be taken into consideration for the sustainability of the services developed by the EUreka3D project.

Different EC initiatives, such as the project EOSC Beyond (2024-2027), are working in parallel to enrich EOSC Core capabilities, allowing scientific applications to find, compose and access multiple Open Science resources and offer them as integrated capabilities to researchers. As part of this process, the project is developing a testing environment for forthcoming EOSC Nodes, data and service providers and validating it with Pilot Nodes activities to build the EOSC Federation. A meeting between the EUreka3D and EOSC Beyond projects was held on 17/10/2024, to establish synergies and a plan of work that will eventually bring EUreka3D's (and the future EUreka3D-XR project starting in 2025) datasets and services to be onboarded in EOSC.

In Cultural Heritage, it is reasonable to consider the ECCCH [2] as the main entry point for a EOSC node implementing a thematic ecosystem of application services in cultural heritage. The future results from the ECHOES project[54] (2024-2029) will determine the directions to follow by the EUreka3D initiative. The rules of participation and acceptance criteria for the research products and services will be defined by the managers of this possible node.

As EOSC is currently in a redesigned phase, and with the scope of assessing the status of their development and how this affects EUreka3D, the project coordinator (PHOTOCONSORTIUM) attended the recent key events organised by the European Commission to present the current landscape of ongoing initiatives: "Shaping the Future of Cultural Heritage", on 27/9/2024, where the expected developments and interactions between the European Collaborative Cloud for Cultural Heritage (ECHOES project), the European common data space for cultural heritage (which EUreka3D serves) and the new configuration of EOSC were discussed and presented; and the Official launch of EOSC EU Node on 10/10/2024, where the main redesign, scopes, structure and organisation of the node was presented and also demonstrated.

What appeared clear from these informative events and overall analysis made in the context of EUreka3D task is that external providers will be allowed in the future to participate in the new EOSC, and thus the

[54] https://www.echoes-eccch.eu/

collection of EUreka3D datasets and services can still be "onboarded" by an EOSC infrastructure, and then be included in their marketplace CHI community services. What is at the moment unclear, and outside the control of EUreka3D, is the timeline and the modalities to implement such onboarding. The concrete action taken at the time of writing regarding the onboarding of services and datasets in EOSC is a collaboration agreement discussed with EOSC Beyond, which foresees various possible joint actions such as using the upcoming EOSC Core Innovation Sandbox[55] for publishing EUreka3D datasets on the EOSC marketplace. This publication can be done directly using the marketplace tools or through harvesting from Europeana, with the potential development of a connector. Moreover, another collaborative effort will focus on defining the future metadata set to be proposed by EOSC Beyond. Additionally, a close monitoring of the upcoming regulations and policy of participation for external Contributors of the EOSC EU Node[56] is being diligently performed by PHOTOCONSORTIUM and EGI so to initiate the procedures when at the earliest possible opportunity.

[55] https://www.eosc-beyond.eu/service/eosc-core-innovation-sandbox

[56] https://open-science-cloud.ec.europa.eu/contributors#contributor-categories

6. FUTURE OF EUREKA3D SERVICES AND TOOLS

Given the current efforts in the cultural sector towards 3D digitisation of heritage objects, sites and monuments, CHIs of any size and capacity are facing various challenges:

- first, to generate high quality digitisation data, that represent faithfully and accurately the CHO in question
- once the data file is created, to store safely such data in a way that various levels of access and editing rights are granted to CHI's staff members and possible partner organisations
- once the 3D models are created from raw data, to visualise them in the web, so that user communities can view the model via their own devices on the internet
- to accompany the 3D models with accurate metadata that describe the heritage object, and with in-depth information about the digitisation process that generated the 3D model, to converge in a paradata report
- to enable access and reuse of raw data and 3D models with their accompanying documentation, for user communities to use and reuse the content in different domains such as heritage research, education, cultural tourism and possible others
- if wanted, to associate a Persistent Identifier (PID) to the objects published as Public data collections, for granting digital preservation of such online content
- if wanted, to contribute the 3D collections in the common European data space for cultural heritage, via the publication in the Europeana Portal.

All these challenges are currently addressed by CHIs in an often disorganised manner, relying on in-house or outsourced services by different service providers, thus often resulting in duplication of efforts, redundancies and complex workflow management and orchestration. As a side, but not irrelevant point, cloud providers who also offer a 3D viewer are often private companies, and often set outside Europe (one example is the Sketchfab service), which raises concerns about safe data management and storage.

The absence of a EU based, non for profit, integrated solution for these challenges has created an evident need in the cultural heritage sector.

In response to these challenges, EUreka3D project developed a suite of services and resources expressly dedicated to CHIs who need to store, manage, document and share 3D models relating to digitised or reconstructed cultural heritage assets. It is specifically intended to facilitate the sharing of such 3D cultural collections in the common European data space for cultural heritage.

6.1 EUreka3D: the comprehensive solution and direct entry-gate to the common European data space for cultural heritage

As widely described in the previous sections, the EUreka3D suite of services and resources for the management and sharing of cultural 3D assets offer solutions for data, metadata and paradata management and storage, and for delivering 3D collections to users' platforms like Europeana. The publication process of the datasets in Europeana is coordinated by partner PHOTOCONSORTIUM, accredited aggregator for Europeana and partner in the common European data space for cultural heritage projects.

The EUreka3D platform was successfully developed and deployed as a pilot action and proof of concept, and the content partners successfully uploaded their models and associated metadata and paradata, available for aggregation in Europeana and as open access collections. The interest shown by a number of CHIs that are external organisations who became associated partners of the project, often with the particular scope of using the EUreka3D solution for their 3D objects, is a clear signal of the fact that the EUreka3D solution is an answer to CHIs' challenges. EUreka3D offers a very welcome "one-stop-shop" solution to these challenges, allowing to store, manage, visualise and share 3D objects in an integrated platform with secured access. Moreover, the project provides added value granting data storage in servers located in Europe, and permitting easier access to Europeana thanks to the full integration of the respective services.

Already in the course of the project, the EUreka3D platform is being used by associated partners to upload and share test collections, such as the Medelhavsmuseet (Museum of Mediterranean and Near Eastern Antiquities) in Sweden, who offered a selection of 3D models of Cypriot heritage from the museum's collection, or INSPAI Girona with a collection of five selected 3D objects, and the RAMS Regionaal Archeologisch Museum a/d Schelde in Belgium publishing in Europeana their first collection of thirteen models. Also the Department of Architecture of KU Leuven is currently developing a case study about sharing in Europeana 3D reconstructions of lost buildings from the XIII century. Another case of using EUreka3D to support a newly launched digitisation project is being carried on by the Basilica del Pi in Barcelona, where the statues adorning the cathedral are currently being digitised and will be made available in Europeana via the EUreka3D platform. Other interested organisations who are testing the EUreka3D platform for supporting their collections are the project Giravolt by GENCAT Barcelona, and the TEKNIKER technology centre in the Basque country. The experience and feedback of these external collaborations complement the iterative testing of the platform and workflow that is done by EUreka3D Content Providers in the project.

In addition to CHIs who are eager to use EUreka3D services, the network of collaborations and engaged stakeholders who participate in the project includes technology providers and research institutions, such as Archéosciences Bordeaux, the Digital Humanities Lab at the University of Basel, the NALIS (National Academic Library and Information System) Foundation in Bulgaria, the University of Cologne, and projects in education and cultural tourism, like ARTEST, Mnemosyne, INCULTUM and SECreTOUR. The interaction with these stakeholders helps the iterative development of solutions and tools at the service of CHIs to reuse and promote their heritage collections in 3D to a variety of audiences. Other large organisations with which EUreka3D collaborates to reach out to the community of CHIs in Europe include ICA the International Council on Archives, meemoo the Flemish Institute for Archives, and Museovirasto the Finnish Heritage Agency under the Ministry of education and culture. Finally, the close relationships with the Europeana initiative and satellite network of communities, projects, aggregators and content providers is further strengthened by presentations and participation in capacity building and outreach events that make the project and its services and tools known in the CH community. Through its innovative approach and resources, EUreka3D deployed the power of cloud services and tools in bringing together a wide range of organisations and professionals, who share a common goal and challenge of preserving CH in 3D for future generations. The project has by now definitely set the basis for moving to the next phase of deploying its services on the market.

6.2 From a research project to marketable product

In simple terms, what CHIs can do in EUreka3D platform is:

- Upload different versions/formats of 3D models, to be shared with different authorization of access depending on user's need/preferences.
- Visualise the model in a viewer that is compatible with Europeana.
- Input metadata, already in the Europeana Data Model, via a simple metadata input form.
- Link contextual information and paradata to each model, to be shared to the public.
- Assign PIDs to the public objects to grant their long-term preservation online.
- Publish such openly accessible objects in the Europeana Portal.

In this light, the EUreka3D platform integrates various components to create a flexible and scalable "product" that could eventually be placed on the market of not-for-profit solutions, according to a sustainable planning aimed at covering the costs with competitive and affordable revenue streams and mechanisms.

The value proposition identified in the project is to offer to CHIs an EU-based comprehensive solution for 3D data management and a direct entry-gate to the common European data space for cultural heritage. This clearly differentiates the EUreka3D platform from similar services for 3D data management, and showcases the competitive advantage of this solution, focusing more specifically on the needs of the "customers" (i.e. the European CHIs), in particular by making use of non-for profit cloud providers based in Europe, federated to the EGI European Grid Initiative, and offering safe data management mechanisms and integrated tools. All these features make the EUreka3D platform not only specialised and competitive, but also more resilient in terms of scalability, adaptability and flexibility to future developments of the digital transformation of the cultural sector. In terms of costs, initial investigations have been done as part of the project's impact assessment task, to estimate the running cost of the platform in its current shape, and a tentative cost of €1,000 per TB per year, that is in line with the price applied by other players like Sketchfab, has been considered. The comparison with Sketchfab is unavoidable given the fact that so far various CHIs relied so far on Sketchfab for visualising and sharing their 3D collections. However, as mentioned, the EUreka3D platform is not comparable to Sketchfab, being a much more specialised infrastructure addressing the needs of the specific community of the CH sector. In this light, EUreka3D Data Hub aims at offering a tailor made service to the target customers, with integrated tools, that CHIs would find much more suitable to their needs than Sketchfab. For this reason, in the competition of EUreka3D with Sketchfab is between two different products, one addressing specifically CHIs requirements, the other offering a generic platform for 3D contents, whatever is the target sector. Also from a user perspective in the domain of cultural heritage, the EUreka3D infrastructure is much more efficient being directly integrated with Europeana and the Data Space for Cultural Heritage.

In terms of revenue and sustainability systems, the mechanisms to develop a marketing model for the EUreka3D platform could be based on user profiling and the user's need for storage and services, similarly to the general *pay-per-use* approach of cloud and other service providers. An entry level with minimum services and functionalities can be established, and additional services can be available to the user upon payment.

The current system developed in the EOSC EU Node[57], based on a mechanism of virtual credits that the user can spend to acquire services and tools, could be used as an inspiration model for EUreka3D. For example, members of the PHOTOCONSORTIUM association get annual access to the EUreka3D platform and have by default a set amount of virtual credits (e.g. 500 virtual credits) that they can spend to use the services and tools. Virtual credits can be bought, they have been tentatively valued at €1 per credit.

As an example, Table 4 outlines a possible structure of the pay-per-use mechanisms that users of EUreka3D could be subject to. It consists of three levels of service, corresponding to three incremental plans. It is possible for the user to switch plans at any moment.

---

[57] https://open-science-cloud.ec.europa.eu/about/eosc-eu-node

Table 4: Exemplary pay-per-use pricing model for the EUreka3D platform

PRICING: 1 virtual credit = €1	
ENTRY LEVEL: the entry level corresponds to e.g. 500 virtual credits per year	
AAI services and EUreka3D user groups: to grant different access and editing rights to different types of users, all the data stored by the CHI in their cloud space on EUreka3D platform can be restricted, limited or open to different groups of users, in the most flexible manner.	service included in the entry level
Storage and data management facilities: as widely explained above, the EUreka3D platform offers storage, virtual machines and compute power based in the EU, hosted by the national providers of the European Grid Initiative (EGI) Federation.	(e.g.) 1 GB storage and functionalities included in entry level
3D viewer to visualise on the web 3D models in different formats: the viewer available in EUreka3D is a basic tool that is fully compatible with Europeana and embeddable on the web.	service included in the entry level
EDM-based metadata input form: the EUreka3D platform offers an easy and friendly way to users to create metadata in the format that is accepted by Europeana platform, via a simple input form with guidance, supported also with a XML editor for the more proficient users. LOD vocabularies such as Getty AAT and Wikidata are supported for relevant metadata fields. It is also possible for the user to upload their metadata already as an XML file, without the need of using the input form. The paradata file is included in the form as a link to enrich the metadata associated with the digital object.	service included in the entry level
PID service: the publishing of Public data collections is supported by the European initiative B2HANDLE, an EUDAT service to provide Persistent Identifiers (PID). The PID is automatically included in the metadata associated with the digital object.	(e.g.) 10 PIDs included in the entry level
Aggregation to Europeana: the EUreka3D platform, supported by Europeana accredited aggregator PHOTOCONSORTIUM, enables the metadata to be finalised, quality-checked against the Europeana Publishing Framework and collected for Europeana publication.	service included in the entry level

OAI-PMH endpoint: the datasets are exposed via an OAI-PMH endpoint for harvesting by harvesting tools of Europeana and other repositories	service included in the entry level

COMMUNITY LEVEL	
Storage and data management facilities	additional storage available upon payment (e.g. 100 credits = 100 MB)
PID service	additional PIDs available upon payment ((e.g. 10 credits = 1 PID)

PROFESSIONAL LEVEL	
API upload: the user can upload their files as a batch, via an API service	service and support available upon payment (e.g. 100 credits = 1 batch upload)
Individual, one-to-one technical support: in case the user needs a dedicated assistance with their models, metadata, paradata, or any other type of technical and procedural support, experts from PHOTOCONSORTIUM, EGI and Cyfronet are at disposal.	service and support available upon payment, to be agreed following a feasibility study

Despite the development of a full business and sustainability model for the EUreka3D suite of tools and services is not in the scope of the project, more investigations and reflections will take place and will be reported in the final D4.2 Impact assessment report. Appropriate collaboration and service agreements between the main actors operating the EUreka3D platform (EGI, Cyfronet, PHOTOCONSORTIUM) will be established as part of the sustainability action of the project in order to grant continuation of the services also beyond the end of the funding period.

6.3 Continuing the work: EUreka3D-XR

The legacy of EUreka3D lies heavily in the knowledge base that was created and disseminated via the project website and the blogs and pro-blogs published on Europeana (which will continue to serve as an information resource to CHIs leaving a lasting impact for the future), and in the EUreka3D suite of tools and services for CHIs that will be maintained and further developed in the light of creating a stable sustainability and business model. In addition, a consortium has been formed to take forward the work of EUreka3D into a project called EUreka3D-XR (starting 1 February 2025). This project will develop a set of new tools, offering shared web applications which are designed to help with the automatisation and standardisation involved with the generation of XR and AR experiences. The project will continue to engage stakeholder communities to share their knowledge and skills, giving access to datasets and transforming cultural contents (2D, 3D, video, texts, maps, stories) into innovative XR scenarios, thus resulting in more and better available digital cultural content for any type of re-use. The impact for CHIs as content users means they will be able to re-use the high-value datasets published by providers (other CHIs) and interact with XR contents co-created with and by colleague institutions and cultural professional, this will result in more culturally important content being available and used in more diverse ways. Both these impacts are expected to increase the rate, and enhance the quality, of digital 3D transformation by CHIs, supporting the EU’s recommendation to digitise all of the at-risk objects in 3D by 2030.

7. CONCLUSIONS

This deliverable presented the technologies used and the architecture of the EUreka3D platform, providing services and a resource hub for CHIs, and supporting publication of 3D datasets of cultural heritage in Europeana.

The document introduced the **technical background and concepts** necessary to understand the underlying technology layer, such as Cloud Compute or data management services. It discussed the **platform requirements** that have been refined over the course of the project as a result of cooperation between technical members and CH community representatives. With that background, it continued to describe the **EUreka3D services and resource hub**, which is based on EGI DataHub, an EGI Federation service deployed and managed by Cyfronet and based on the Onedata data access platform. A detailed architecture of the service has been presented, focusing on aspects relevant to EUreka3D requirements, such as data and metadata access and management, data sharing, data publishing, security, and the details concerning hardware requirements, software deployment and the integration with the Europeana Portal. Objects stored in EUreka3D receive a PID, which is registered and provided by the EC initiative B2HANDLE. With all these features, EUreka3D provides a comprehensive, user-friendly platform for CHIs.

The deliverable has also narrated about the work done in the light of onboarding EUreka3D datasets and services in **EOSC**, in consideration of the current redesign scenario. It has defined a list of services and research products for EUreka3D and analysed the ongoing developments of EOSC with the new EOSC Federation of Nodes. It has also explained the interactions between EUreka3D and the EOSC Beyond project, and the steps that have been taken to ensure compatibility with a future integration into a EU Node as soon as this is accepted.

In addition to the technical descriptions and the considerations about the publication of datasets and services in EOSC, the deliverable presents reflections on **sustainability of the EUreka3D platform**, in consideration of a requirement analysis of the needs of our target community (i.e. the Cultural Heritage sector), of the value proposition and competitive advantage that the EUreka3D solution offers to its current and potential users, and of a basic comparison with similar services. The work of EUreka3D will continue in EUreka3D-XR, a project starting in 2025 that will deliver a set of new tools for the generation of XR experiences in CH.

In summary, this deliverable described the final state of EUreka3D's architecture, which has been evolved and refined during the course of this project, to adapt to the needs of CH communities. At this final phase, the EUreka3D platform is an operational, production-grade system, integrated with the Europeana Portal, offering valuable support in the implementation of the European common data space for cultural heritage.

References

[1] Commission Recommendation of 10.11.2021 on a common European data space for cultural heritage https://digital-strategy.ec.europa.eu/en/news/commission-proposes-common-european-data-space-cultural-heritage (last accessed Oct 2024)

[2] A European Collaborative Cloud for Cultural Heritage https://research-and-innovation.ec.europa.eu/research-area/social-sciences-and-humanities/cultural-heritage-and-cultural-and-creative-industries-ccis/cultural-heritage-cloud_en (last accessed Oct 2024)

[3] VIGIE 2020/654 (2020) "Study on quality in 3D digitisation of tangible cultural heritage: mapping parameters, formats, standards, benchmarks, methodologies, and guidelines" https://digital-strategy.ec.europa.eu/en/library/study-quality-3d-digitisation-tangible-cultural-heritage (last accessed Oct 2024).

[4] Caballer M., Blanquer I., Moltó G., de Alfonso C. (2015) "Dynamic Management of Virtual Infrastructures" J. Grid Comput., vol. 13, no. 1, pp. 53–70, Mar. 2015, doi: 10.1007/s10723-014-9296-5.

[5] TOSCA (2016) "TOSCA Simple Profile in YAML Version 1.0". https://docs.oasis-open.org/tosca/TOSCA-Simple-Profile-YAML/v1.0/TOSCA-Simple-Profile-YAML-v1.0.html (last accessed Apr 2023).

[6] RADL (2023) "Resource and Application Description Language (RADL) — IM Documentation 1.0 documentation". https://imdocs.readthedocs.io/en/latest/radl.html (last accessed Apr 2023).

[7] European Commission: Directorate-General for Research and Innovation (2020) Solutions for a sustainable EOSC. A FAIR Lady (olim Iron Lady) report from the EOSC Sustainability Working Group. Publications Office, 2020, https://data.europa.eu/doi/10.2777/870770

[8] Robertson D., Meijer J., Roi A., Hacque-Cosson F., Klemeier J., Kaczmirek L., Rey Mazon M., De GIacomo O., Mergen P., Belsø R., Muscella S., Proudman V. (2024) Recommendations for a Financially Sustainable Post-2027 EOSC. 10.5281/zenodo.11317576. https://zenodo.org/records/11317576

ANNEX A: CONTENT PROVIDER HANDBOOK

Content Provider Handbook

v1.3

Dissemination level: Public

Authors:
Ignacio Lamata Martinez (EGI Foundation)
Łukasz Opioła, Katarzyna Such (Cyfronet)

HISTORY OF CHANGES			
Version	Date	Author	Comments
0.1	20/03/2024	Ignacio Lamata Martinez	First draft
1.0	18/07/2024	Łukasz Opioła, Katarzyna Such, Ignacio Lamata Martinez	Version ready for distribution
1.1	10/09/2024	Ignacio Lamata Martinez, Łukasz Opioła	Some changes in Section 3.6. Added VIGIE Study references and short description of directories. Added changes for v21.02.6. Other improvements based on user feedback
1.2	19/10/2024	Ignacio Lamata Martinez	Added comment about materials. Content updated. Cosmetic changes on Notes. Added section to test 3D visualisation before publishing.
1.3	28/10/2024	Ignacio Lamata Martinez, Łukasz Opioła	Added API section (automated upload), Annex A and Annex B.

Co-funded by
the European Union

Table of contents

1. Introduction

This **informal** document explains in a **practical** way the steps required to use EGI DataHub, focusing especially on the upload of 3D models by content providers. This can be considered as a manual for the upload of data that complements the official deliverables of the project.

1.1 Structure of the document

The rest of this document is organised as follows:

- **Section 2** explains how you can prepare your access to the applications available for the EUreka3D community.
- **Section 3** explains how you can upload and publish your 3D models through EUreka3D.
- **Section 4** discusses how to automate the process of uploading and publishing 3D objects, which is useful when hundreds of objects have to be uploaded and is unfeasible to do the work manually.
- Finally, **Section** 5 provides some conclusions.

1.2 Where to find additional information

If you need more context and information about the technologies used in EUreka3D, please refer to the official deliverables of the project, in particular:

- **Deliverable 3.2** "*The EUreka3D AAI architecture*" (October 2024), which describes the infrastructure and technologies implemented to perform the authentication and authorisation of users in EUreka3D.
- **Deliverable 3.3** "*Final report on the EUreka3D services and resource hub*" (October 2024), which describes the configuration of the cloud and data technologies used in the project, giving a deeper understanding of EGI DataHub, the applications and the compute side of the project.

2. Configuring your access

The access to applications and data in EUreka3D must be protected from unauthorised users. In order to implement this security mechanism, EUreka3D data hub and services are supported by **EGI Check-in**[1]. This section will explain to you the actions you need to take to configure your access to EUreka3D, which mainly consist of:

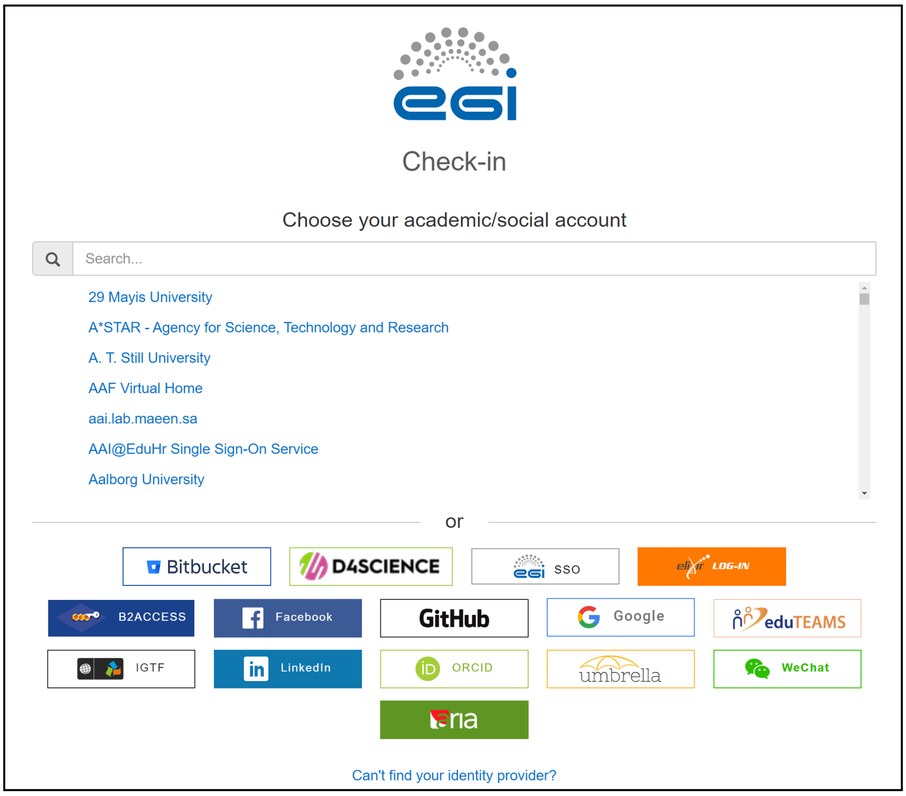
- **Register in Check-in**. You will need to register your account in Check-in. This is discussed in Section 2.1.
- **Join the EUreka3D community**. To obtain the required permissions to access EUreka3D applications and data, you will need to join the EUreka3D community. This is discussed in Section 2.2.

2.1 Registering an account in Check-in

Registering an account is a simple process. Just visit:

https://aai.egi.eu/signup

The process to register your account is as simple as just logging in with your usual account. As you can see in the image below, Check-in offers you multiple options to log in:



[1] https://www.egi.eu/service/check-in/

It is recommended that you use the account of your Home Organisation (your research institute, University, etc) by looking for it in the search box. If your institution is not listed here, you can use one of the social accounts available: Google, LinkedIn, ORCID, GitHub, etc. All of these have a specific button that will redirect you to their login page. Once you have completed the login process with your preferred account, you will be registered in Check-in.

As a last resource you can create a specific EGI account by clicking on the link “Can’t find your identity provider?” at the bottom of the page. Note that this is **not recommended**, as you will need to create a new account with a specific username and password (which is unique for this account) and you will lose the benefits of having a unique digital identity.

Additional documentation:

- https://docs.egi.eu/users/aai/check-in/signup/

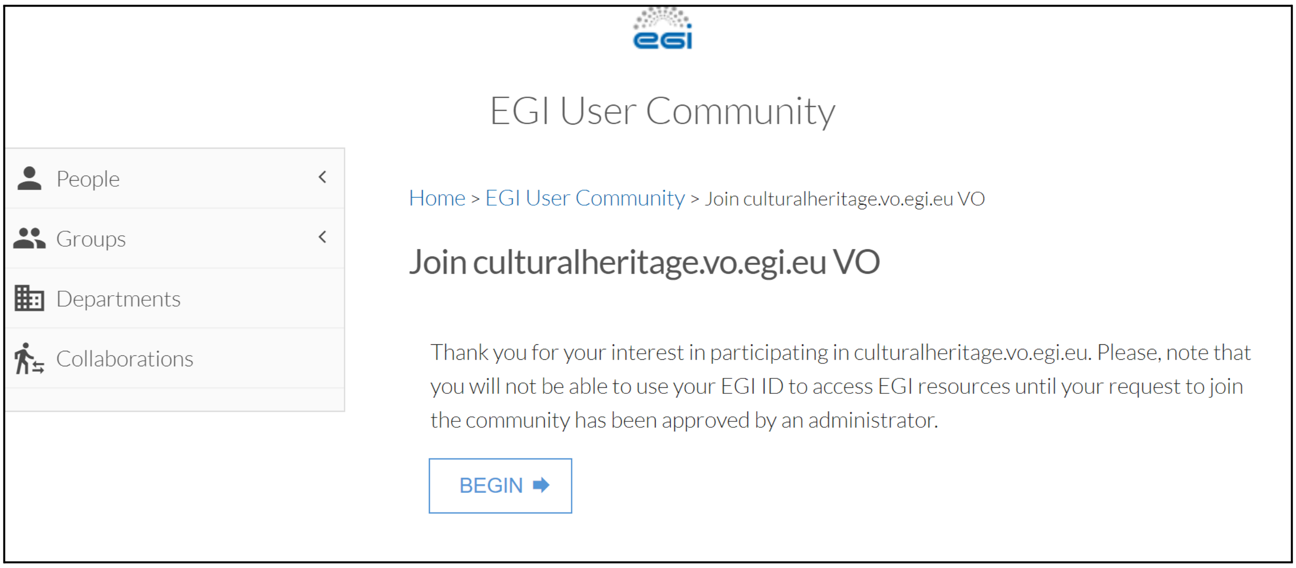
2.2 Joining the EUreka3D community

Once you are registered in Check-in you need to ask to be included in the EUreka3D community. This is done by requesting membership in a *Virtual Organisation* (VO), which is a group of users that belong to a community and have specific permissions assigned to access some resources. The VO of EUreka3D is called: **culturalheritage.vo.egi.eu**.

To ask for membership in the EUreka3D VO you need to visit the following link:

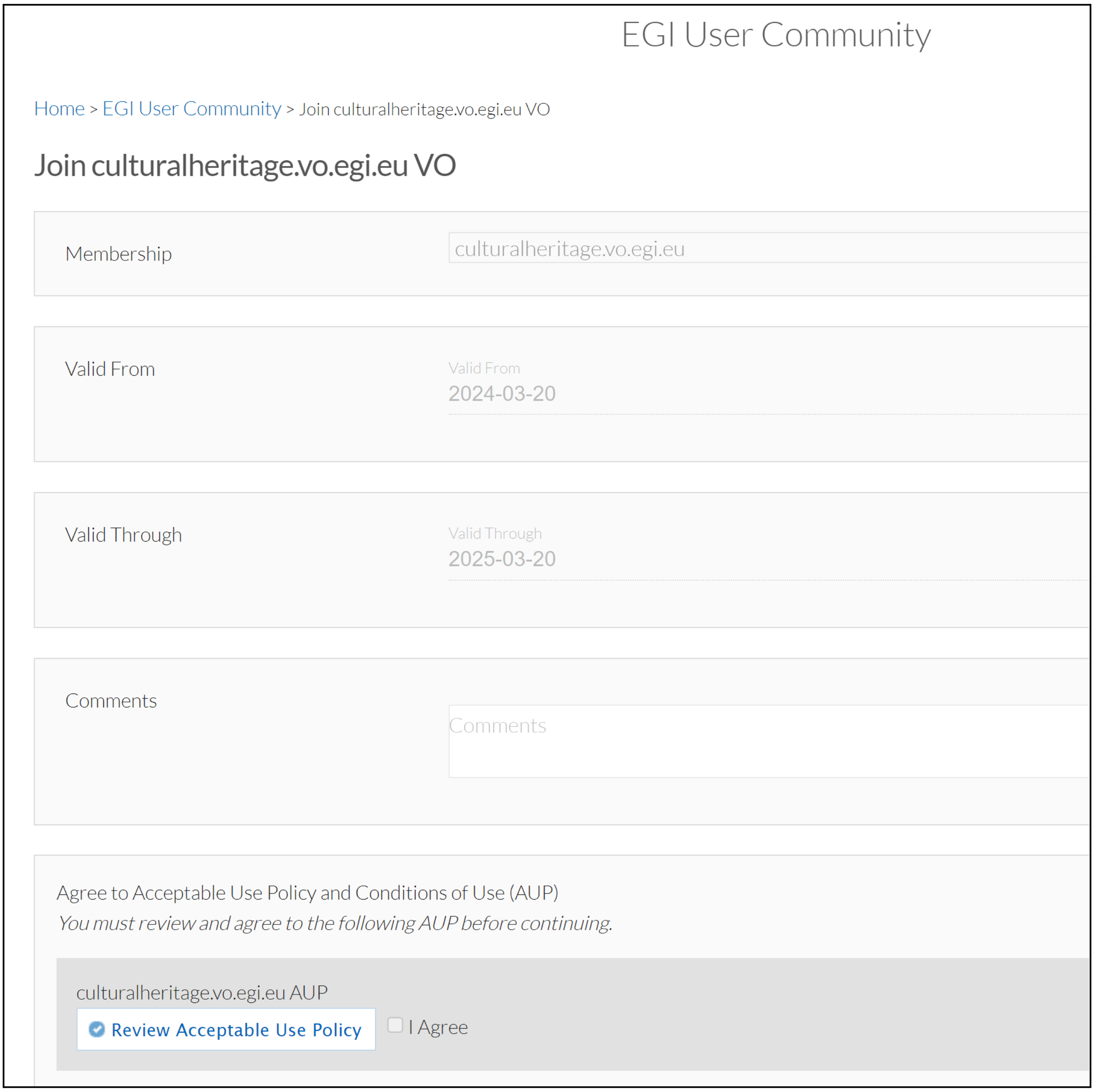
https://aai.egi.eu/registry/co_petitions/start/coef:632

which will present you with the following page:



Note that **you will need to be logged in** in Check-in, otherwise you will be requested to log in first.

Click on “*BEGIN*” and you will be presented with the following screen:



In this page you can provide some comments about why you need to request access to EUreka3D. Then, review the Acceptable Use Policy by clicking on the button and finally click on “*I Agree*” if this is the case. Click “*Submit*” to finish the request.

Your application will be received by an administrator who will evaluate it and manually accept or reject it. This process should only take a few hours, but please consider that this is a manual process and therefore subject to longer processing times. If you are also publishing in Europeana, some extra configuration is necessary: either create a specific group for your institution in the platform or assign you to an existing group. If the creation is necessary consider that the process to grant you access will take some extra time.

Membership in this VO **must be renewed every year**. This means that every year you will receive a notification email to remind you to renovate your membership. The process is exactly the same as when you join the VO for the first time: visit a link and follow the instructions.

Additional documentation:

- https://docs.egi.eu/users/aai/check-in/joining-virtual-organisation/

3. The data platform — DataHub

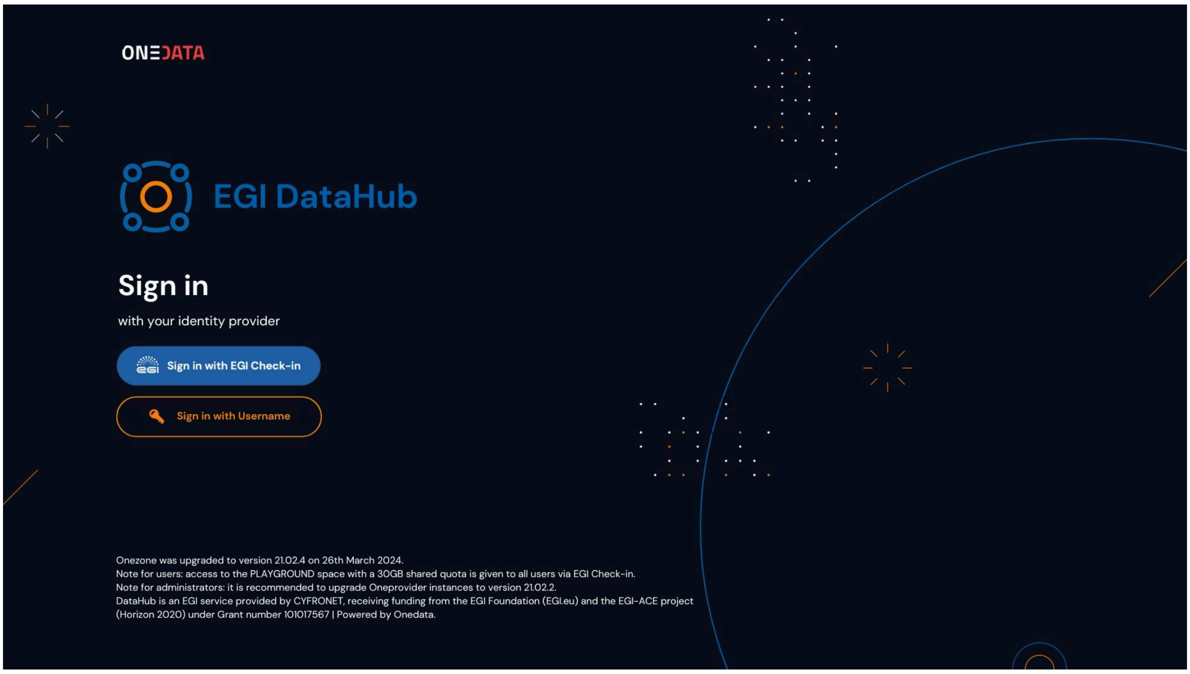
Once you have registered your account in Check-in and become a member of the EUreka3D community you will be able to access the data platform. The data management platform in EUreka3D is implemented with **EGI DataHub**[2].

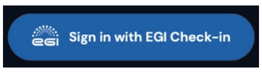
3.1 Accessing the DataHub

To access the data platform you just need to visit:

https://datahub.egi.eu

which will present the following login screen:



Click on the EGI button (

) to be redirected to Check-in, and proceed with the login process as usual.

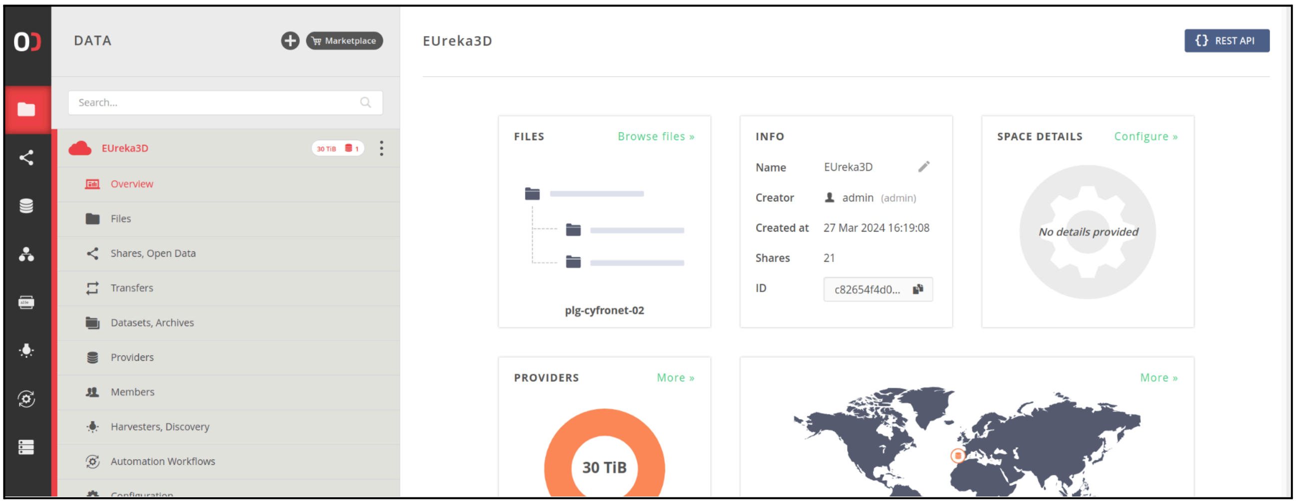
NOTES:

1. The login page will not appear if you are already logged in (then simply go to Section 3.2).
2. If this is the first login via Check-in, you will be asked to accept the terms and conditions and agree to release your basic user information to DataHub. Follow the instructions on the screen.

[2] https://www.egi.eu/service/datahub/

3.2 DataHub overview

After you have logged in, you will be presented with DataHub's home screen, where you can see a data space called "EUreka3D" in the left menu.



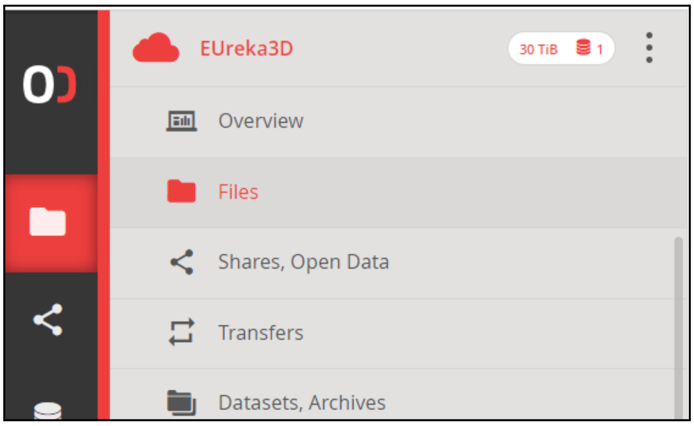
NOTE: If you do not see the EUreka3D data space depicted above, then:

1. Make sure you have completed the process described in Section 2.1.

2. If you had logged in before being accepted to the EUreka3D community, **you will have to log out and log in again**, so that your group memberships can be updated. Use the menu with the user icon in the bottom left corner to log out.

3. It is possible that you have not been accepted in the EUreka3D community or that you have not been assigned permission to access DataHub. If this is the case, contact an EUreka3D administrator via email.

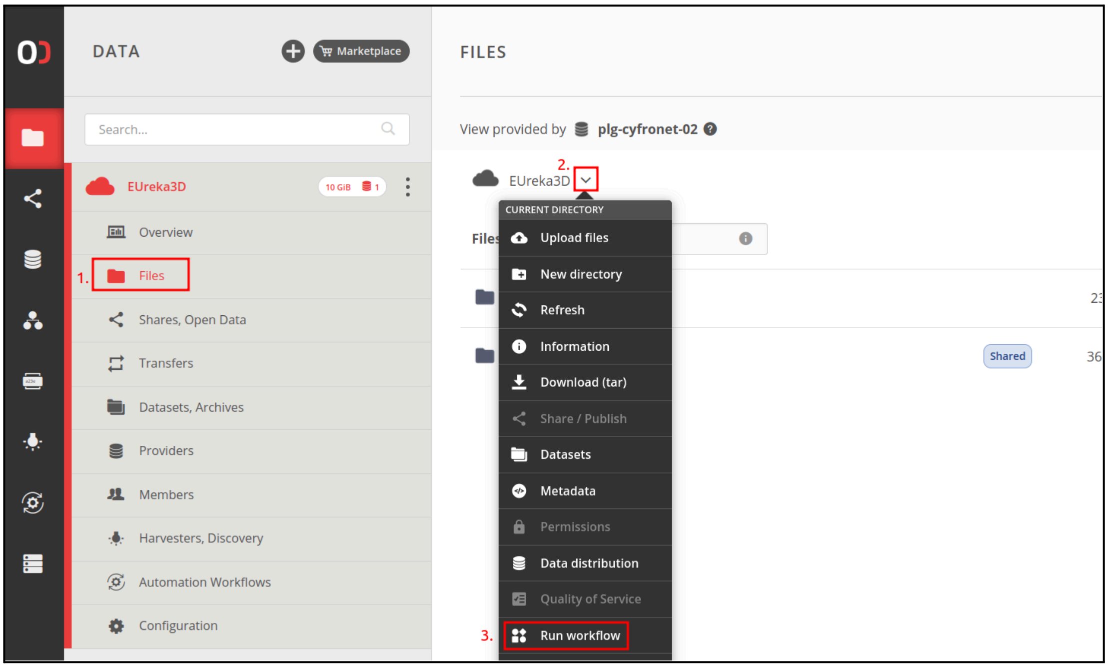
Every project in EUreka3D is organised into a single directory that lies directly in the root of the EUreka3D data space and has a predefined structure. Go to the left menu and click on the **Files** tab to open the file browser:



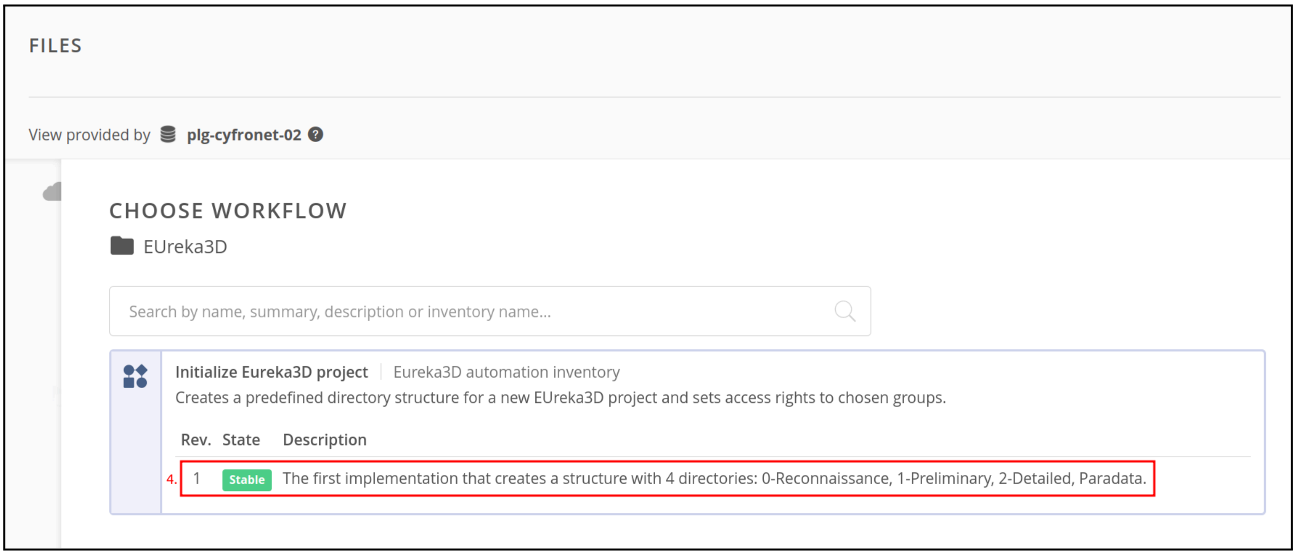
3.3 Creating a new project

To facilitate the creation of a project and its underlying directory structure, an automated procedure called "workflow" must be run. To do so, follow these steps:

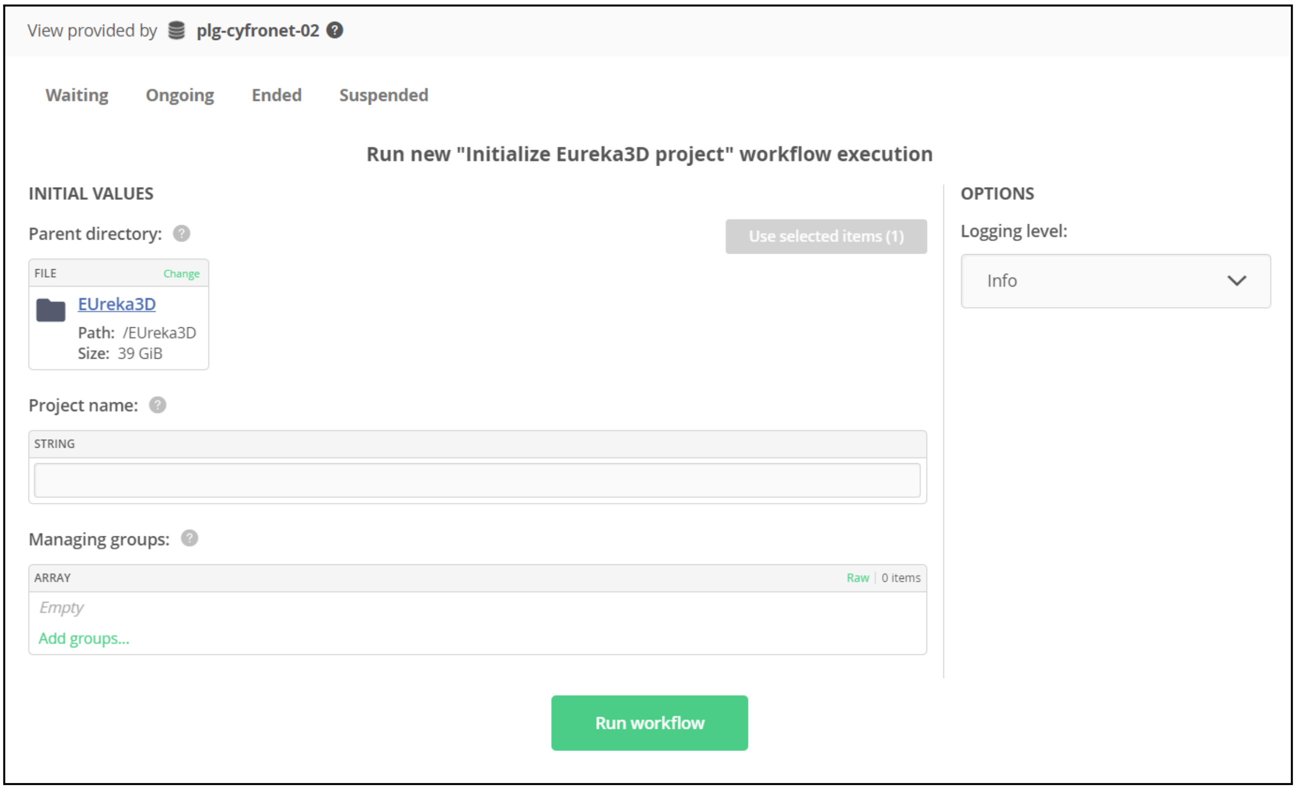
1. Click on the **Files** tab of the left menu, located in the **EUreka3D** space sidebar.
2. Open the **dropdown menu** for the main directory of the EUreka3D space by clicking on the button next to **EUreka3D**, as shown in the figure below.
3. Choose the **Run workflow** action.



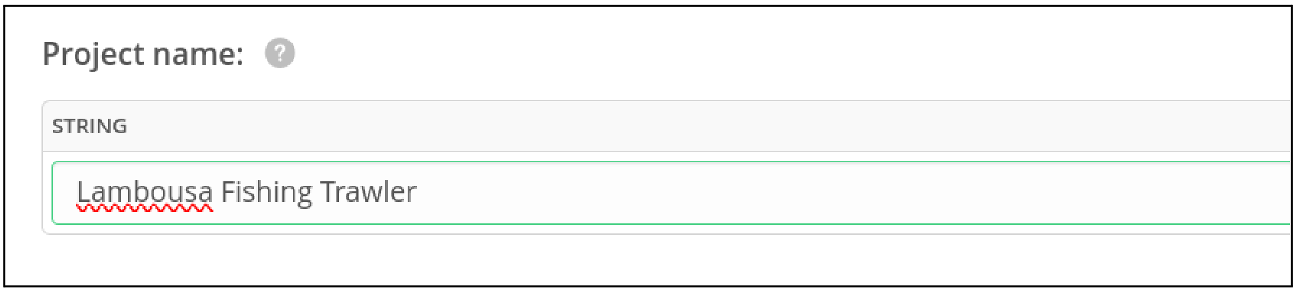
4. Locate the “**Initialize EUreka3D project**” workflow and click on the newest stable revision (e.g. Rev. 1), as shown in the figure below:



This will open the following screen:

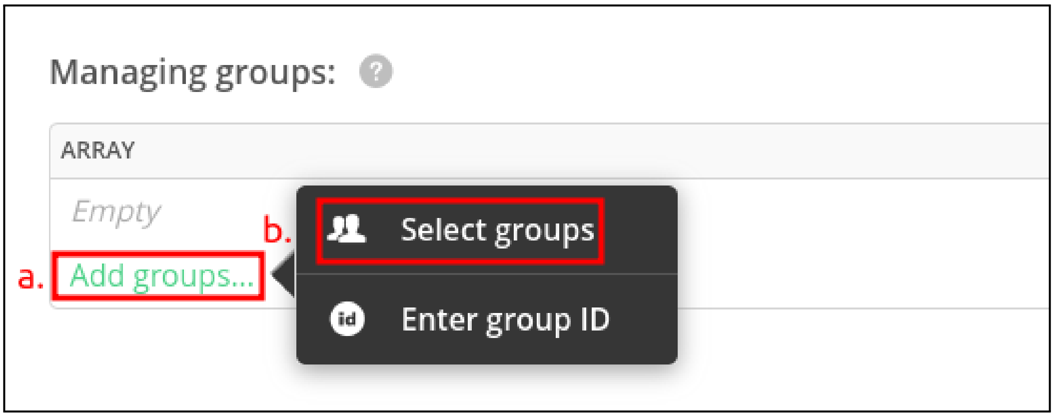


5. Enter the desired **project name**. Avoid using your institution name. Instead, use a meaningful name for the objects that will be stored.

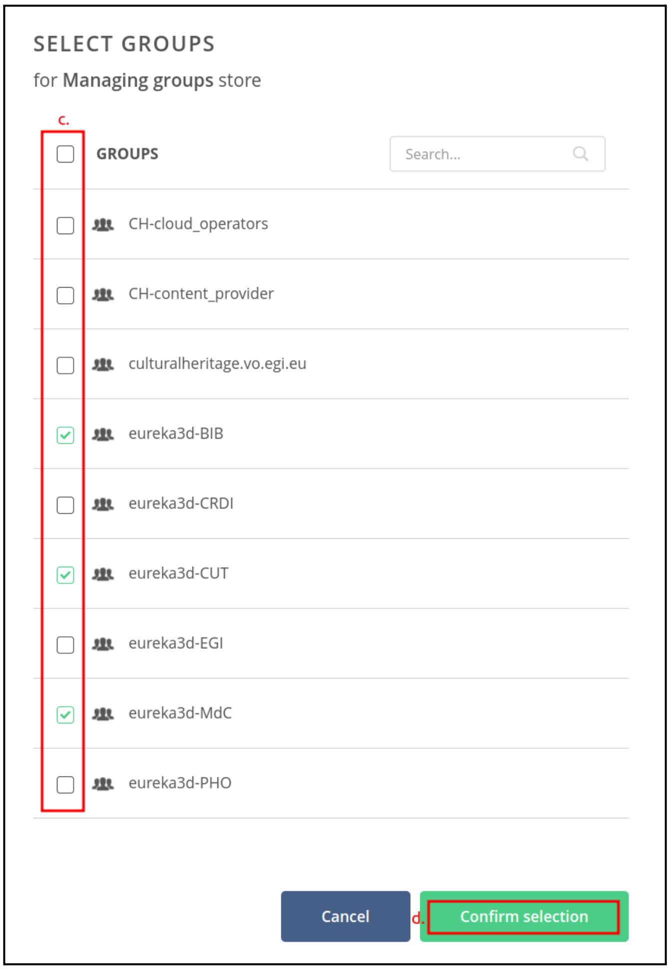


6. The section "**Managing groups**" lets you specify the groups inside EUreka3D that will have access to this project. For example, the "eureka3d-CUT" group represents the members of the Cyprus University of Technology, the "eureka3d-PHO" group represents the members of the Photocosortium, and so on. By default, no ACL will be established and **everyone in the EUreka3D community** will have access to your data so, if this is a problem, it is recommended that you limit the access to your data at this initial stage. Note that you do not need to assign access permissions to other groups at this time - you can also modify these permissions at any time after the project has been created.

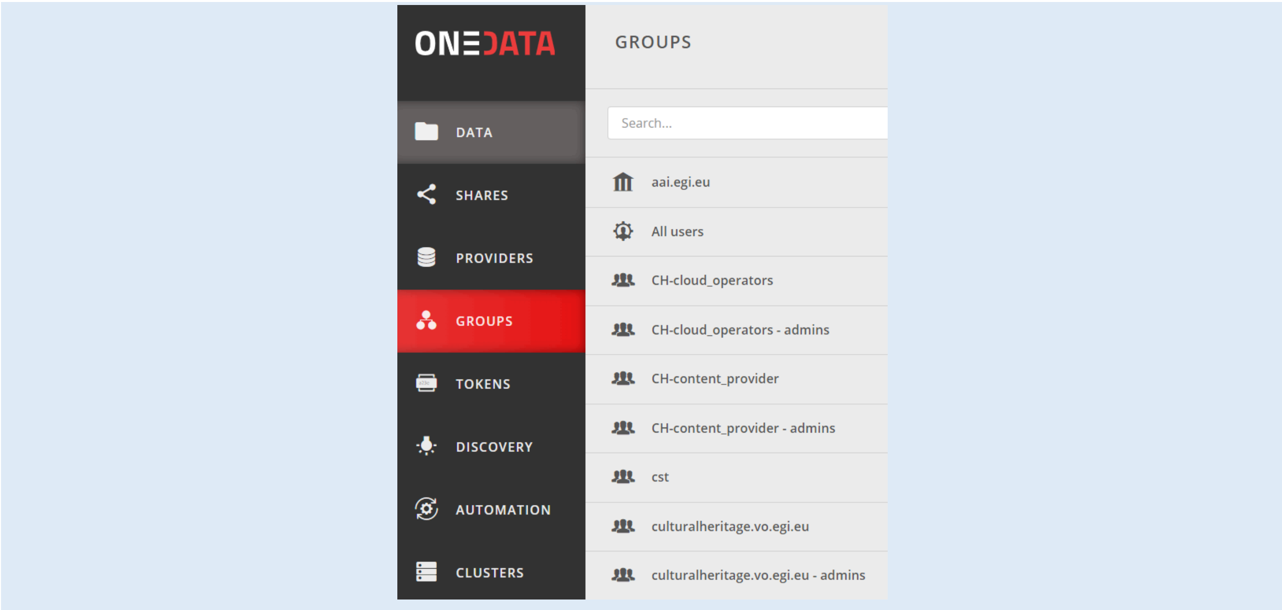
To specify the managing group(s):



a. click the **Add groups…** button,
b. click the **Select groups** action,
c. select the groups by checking the boxes,
d. click the **Confirm selection** button.

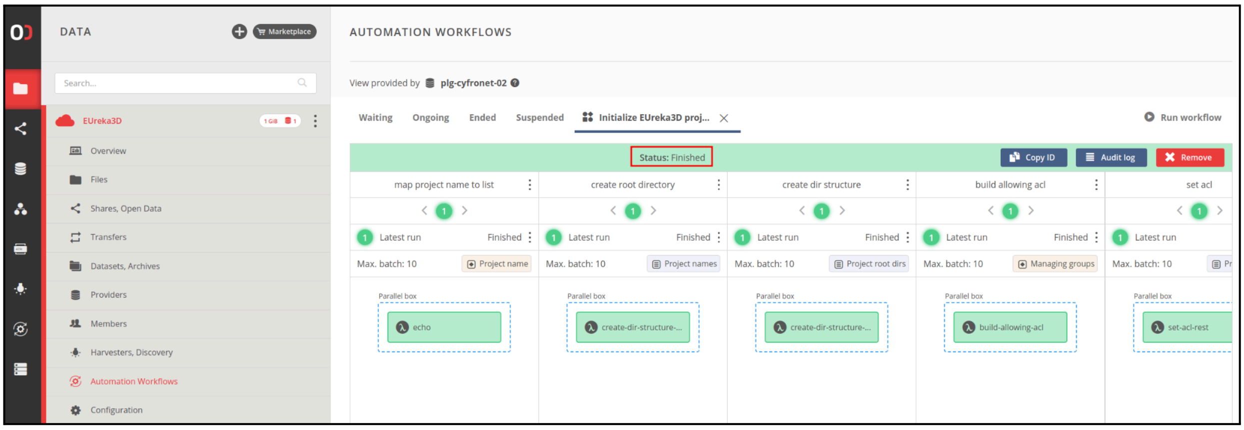


WARNING: If you configure an ACL, make sure you choose at least one group to which you belong or you will not be allowed to access the newly created project. If necessary, consult the **GROUPS** section in the left menu to check to which groups you belong — the list of your groups will be shown in the sidebar:



7. Click the **Run workflow** button.

8. Wait a moment for the workflow to finish.



9. Proceed to the **Files** tab in the **EUreka3D** space sidebar to see that the project directory was successfully created.

The new project will contain an initial structure where your data can be uploaded:

- 0-Reconnaissance
- 1-Preliminary

- 2-Detailed
- Paradata
- Viewer

Directories *0-Reconnaissance*, *1-Preliminary* and *2-Detailed* are intended for data in different quality levels:

- **0-Reconnaissance** is for data produced from a quick and easy digitisation, no complex and no expensive: (a) Very quick 2D/3D digitisation, (b) Not so accurate data set, when objects materials and their condition are not considered much, (c) Not expensive data acquisition and preprocessing of the data sets (when the most primitive equipment has been used), (d) The interest is only in the data formats used for your applications (and not about standards), (e) Data preservation and geometrical accuracy are not interesting.
- **2-Detailed** should contain the highest possible and most accurate digitisation, in order to achieve an outstanding record of data sets in the top possible quality. For example when you have high requirements in place, such as: (a) Enough time, (b) Enough budget, (c) The best infrastructure and professionals in place (for surveying and pre-processing), (d) No IPR issues, (e) You want to acquire all the information and data from the scanned objects (including accurate 2D textures/images, object's temperature and humidity/pixel, and 3D geometry, materials and their conditions, etc.), (f) Data long-term preservation is a must, (g) Data sets in standard formats is a precondition, (h) The availability of the unique results to be in Europeana.
- **1-Preliminary** is the intersection between Levels (0) and (2).

For more information about quality levels, refer to the EC's VIGIE Study[3].

Paradata is a directory intended to store all relevant paradata information (information about the digitisation process).

Viewer is a convenient directory for the storage of the 3D models that will be rendered by a viewer, such as those objects published in Europeana. This directory is intended to contain mainly public data that will be consumed by a 3D viewer.

3.4 Uploading files

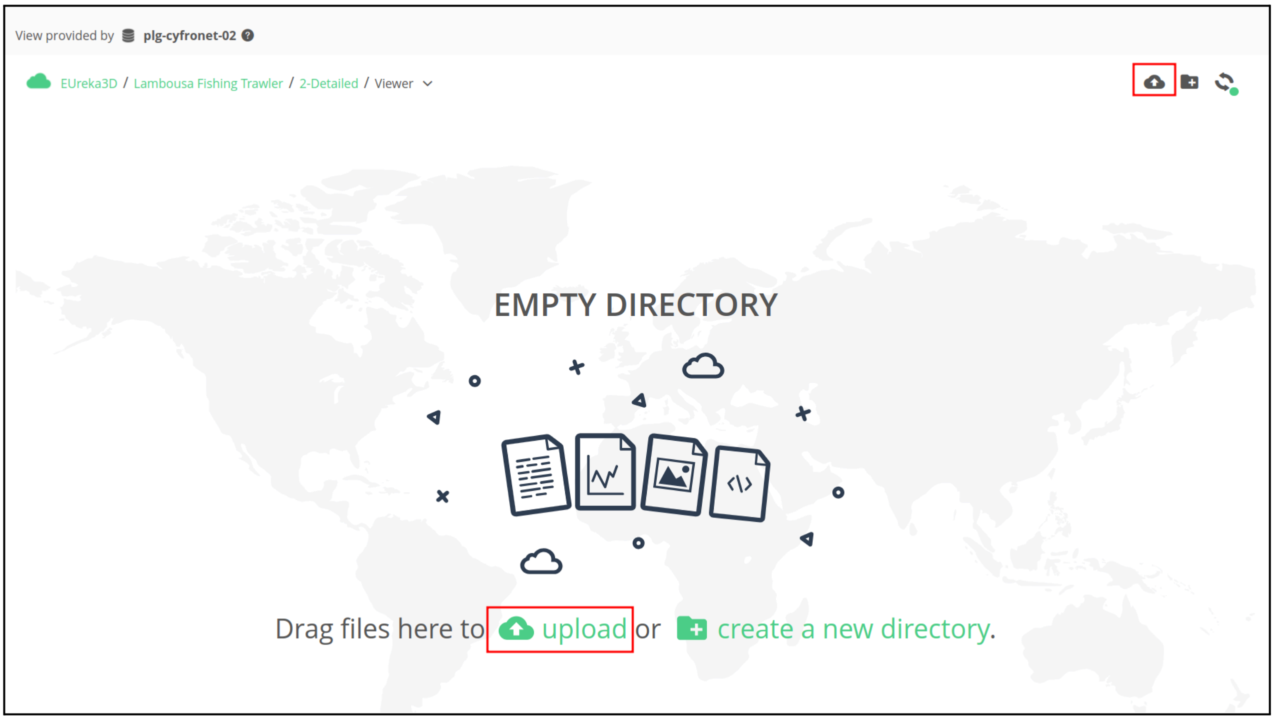
The process of uploading files is very easy, thanks to the user-friendly graphical interface.

To upload files, navigate to the target directory. Use **double-click** to enter a directory or the **breadcrumb navigator**, located on the top, to move back:

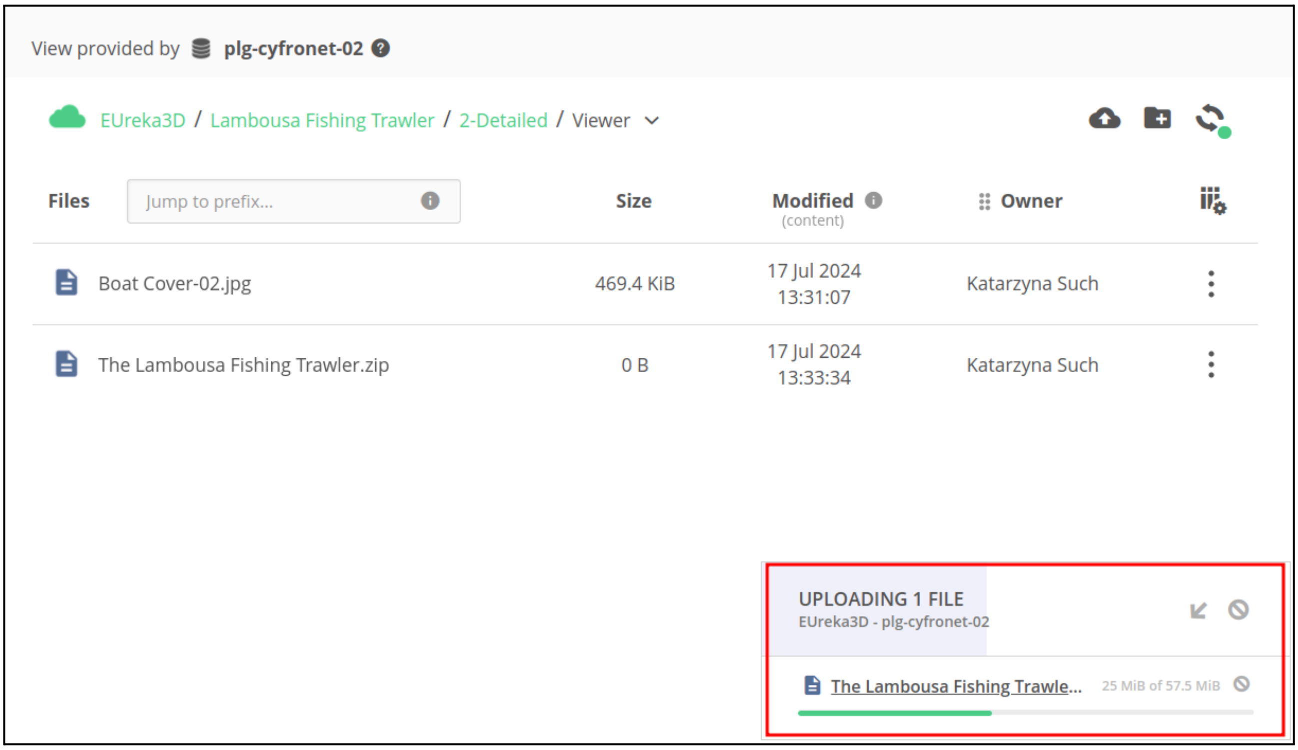
There are two main ways to upload files:

[3] *Study on quality in 3D digitisation of tangible cultural heritage*. Available at https://go.egi.eu/vigie

1. Files can be **dragged and dropped** over the directory and they will be uploaded
2. Alternatively, you can use the **upload action**, either in the centre of the page (if the directory is empty), in the top right corner, or invoking the context menu with a **right-click**.



Finally, choose the desired file from the computer. The upload progress bar is shown in the bottom right corner of the page.



3.5 Uploading 3D models for visualisation

3D models that are intended to be visualised by the EUreka3D viewer must be stored in a compressed ZIP file. You must pack your 3D model files in **a single ZIP file**. If the model has several files, such as an MTL, JPG and OBJ files, the three will be packed in a single ZIP file. If the model has only one file, such as a PLY, only this file should be packed in the ZIP file.

After the model has been packed, you should upload the ZIP file inside the "Viewer" directory.

IMPORTANT: Make sure that the initial view of the 3D model is in **its most natural position**, so the camera does not initially show the model in a strange position. If you are unsure, you can follow the steps of Section 3.7.2 to pre visualise the 3D model.

3.6 Sharing data

Once your data are uploaded, there are three main ways to share your data:

- Share data with other members of **the EUreka3D community**. All members of the community have access to DataHub, and you can grant them permissions to access your directories and data. These permissions are normally assigned to whole groups within the community. For example, the "eureka3d-CUT" group represents all members of the Cyprus University of Technology. To understand how to share data with other community members, refer to **Section 3.6.1**.

- Share data **publicly**. Data can be shared with any person that is not part of the EUreka3D community. This is a convenient way to provide specific files and even whole directories for users that have noy been allowed to use DataHub. This is explained in **Section 3.6.2**.
- Publish 3D models in **Europeana**. Another way to share data is to use Europeana, the European initiative that aims at facilitating cultural heritage for education, research, creation and recreation. This publishing process is discussed in **Section 3.7**.

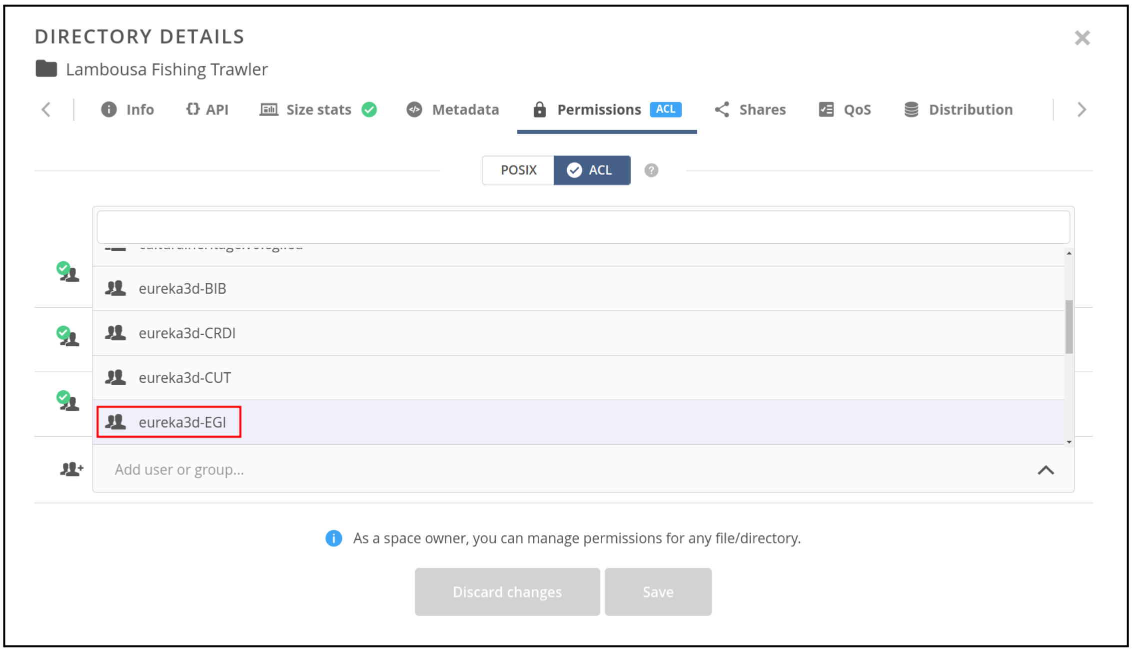
3.6.1 Configuring permissions: Collaboration with other members of EUreka3D

In order to configure the access of other users to your data, Access Control Lists (ACL) are used. These lists specify the permissions that each group of users have over the data (directories and files). Normally, permissions are assigned to preconfigured groups of users, rather than to individual users. If a new group is required, please contact an administrator (a "VO Manager") of the EUreka3D community.

The first configuration of access permissions is done during the creation of a new project, as explained in Section 3.3. The managing groups that are specified during the workflow execution grant rights to view and modify the project directory according to the specified ACL. Users who do not belong to any of the configured groups will not be allowed to view or modify the files inside the project.

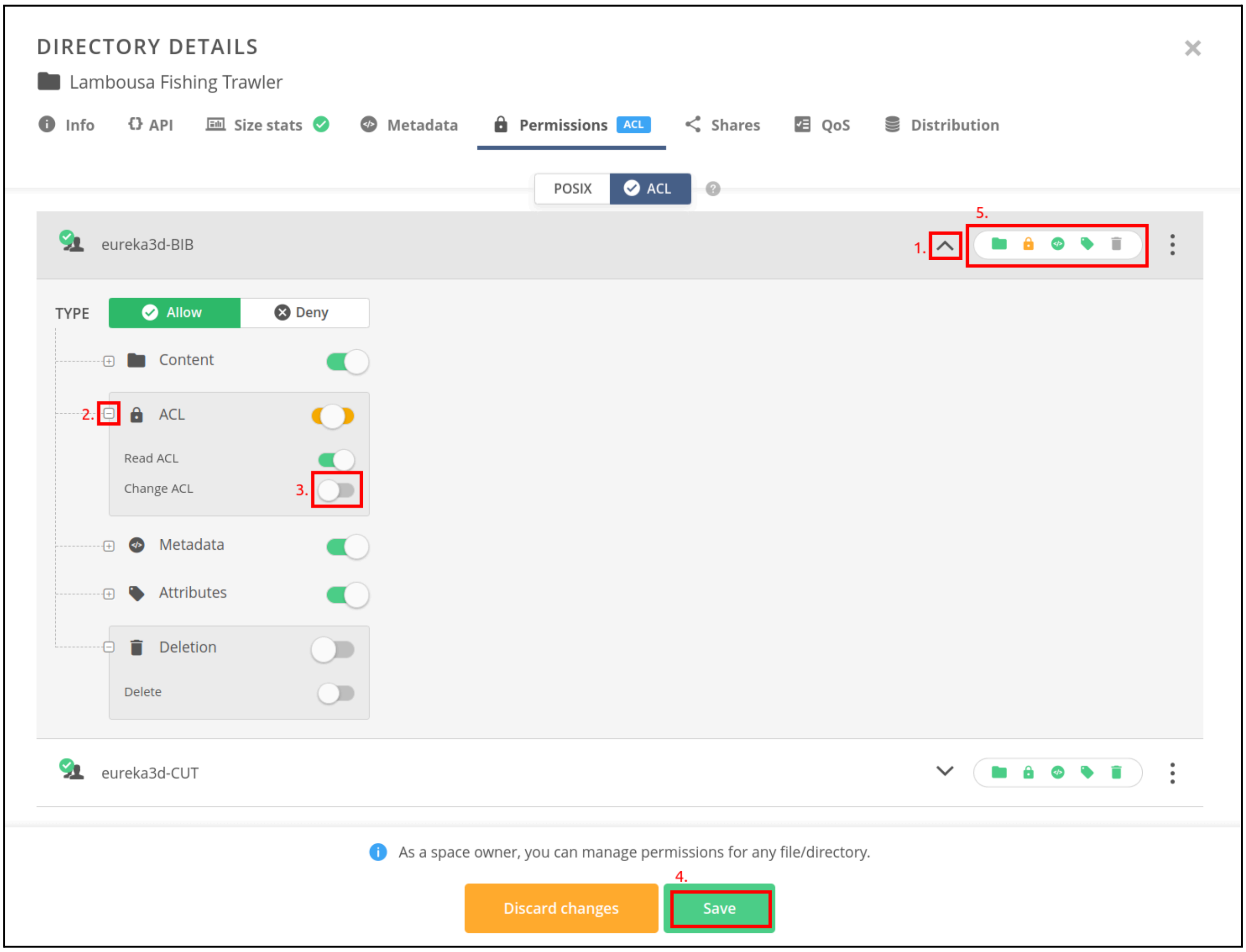
At any other time, the ACL can be modified to add or remove groups, or to configure specific aspects of the access to some data. Use the **ACL** badge or the **Permissions** action to view the ACL set for any directory:

The default ACL can be adjusted as needed. Use the editor to remove entries or add new ones, adding rules for a space member (user or a group of users).



To grant or deny granular operations for a specific user or group:

1. Open the drop-down menu on an Access Control Entry (ACE) for the user or group. The **TYPE** of the entry indicates if the permissions listed below will be allowed (green) or denied (red).
2. The permissions are grouped by category (Content, ACL, etc.). Use the **+/-** button to expand or collapse the list of permissions.
3. Toggle right to select a permission that will be impacted by this ACE. Inactive permissions (marked as grey) are neither allowed nor denied by this ACE.
4. Click the **Save** button to save the changes.
5. At the right end of the principal row, you can see icons indicating the status of the groups: green (fully granted), yellow (partially granted), or grey (not granted).



ACL provides control of access to such resources as:

Permissions	Description	Applies to
Read	Allows opening the file for reading	File
Write	Allows opening the file for writing	File
List files	Allows listing directory content (see files inside a directory)	Directory
Add files	Allows adding a file inside the directory	Directory
Add subdirectory	Allows creating a subdirectory inside this directory	Directory
Traverse directory	Allows navigating through a directory structure	Directory
Delete child	Allows deleting files or subdirectories inside a directory	Directory

Read ACL	Allows reading file/directory permissions	File/Directory
Change ACL	Allows writing file/directory ACL	File/Directory
Read metadata	Allows reading file/directory metadata	File/Directory
Write metadata	Allows writing file/directory metadata	File/Directory
Read attributes	Allows reading metadata associated with file/directory attributes	File/Directory
Write attributes	Allows writing metadata associated with file/directory attributes	File/Directory
Delete	Allows deleting the file/directory	File/Directory

Additional documentation:

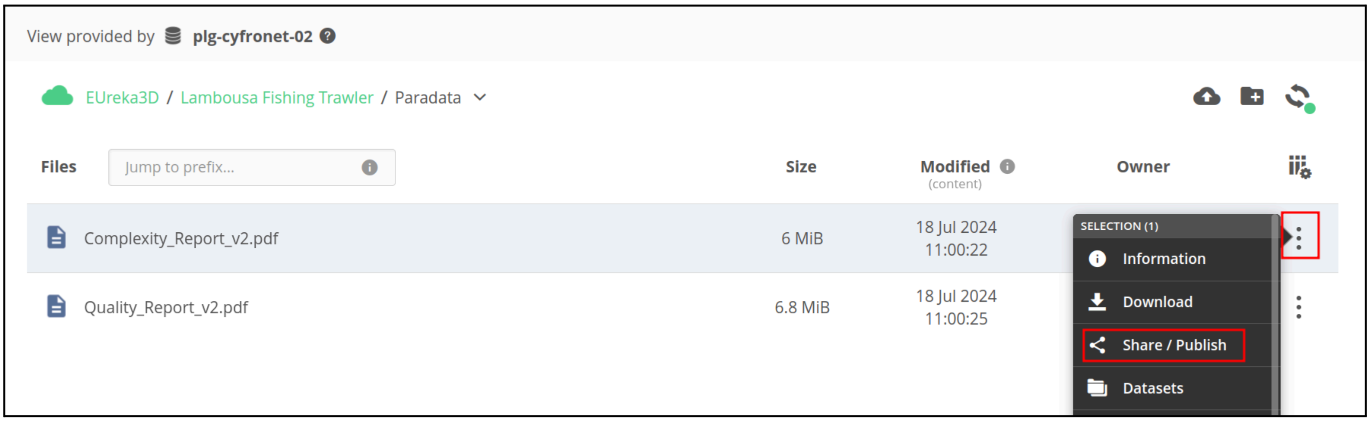
- https://onedata.org/#/home/documentation/21.02/user-guide/data[access-control-lists].html

3.6.2 Sharing data publicly

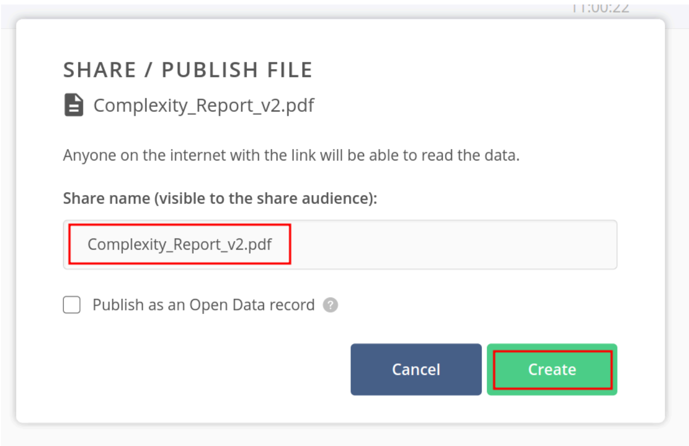
Sometimes it is useful to share data with other people, such as a paradata PDF, a raw 3D model or different interesting objects that have been produced during the digitisation process. For these cases, a **Share** can be created. Shares can be created on a single file, if only the file is to be shared, or on directories, if the whole content inside the directory is to be shared. Sharing directories is useful when many files have to be shared.

NOTE: Shares are semi-public, which means that anyone knowing the link to a share can access it. However, in practice, guessing the link is unfeasible in a timely manner.

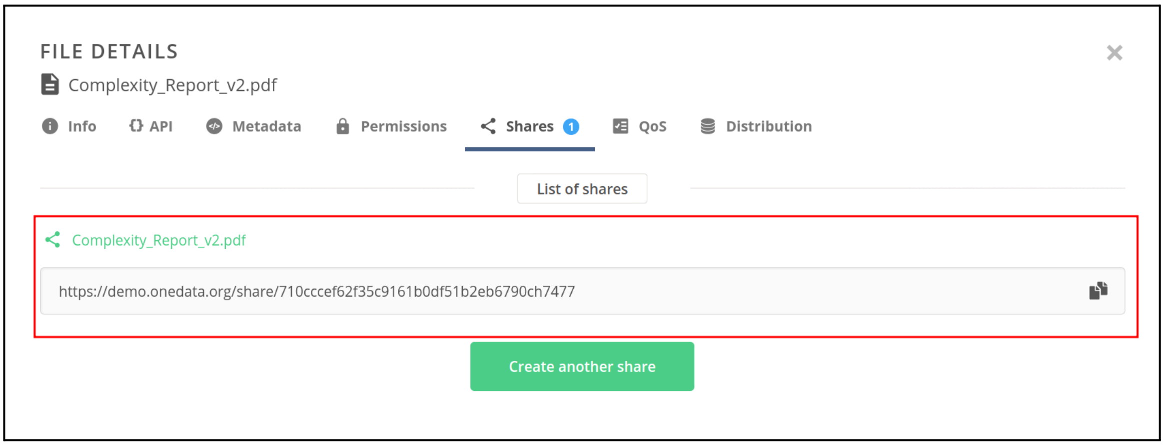
To create a Share click on the three dots button ⋮ at the far right of the file or directory, and choose the **Share/Publish** option.



Enter a Share name that will be visible to other users and click the **Create** button.



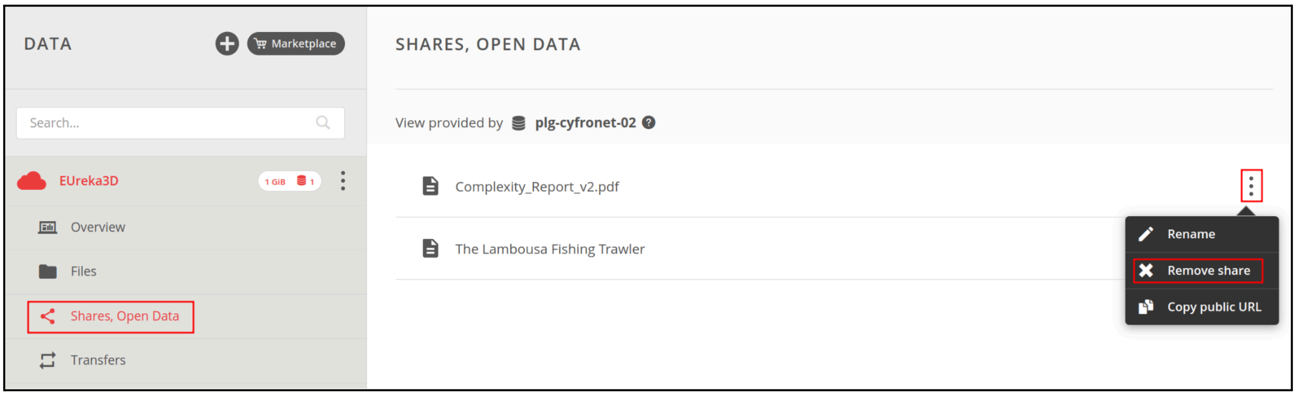
After the successful Share creation, a right-side details panel with the **Shares** tab will appear containing a list of Shares created for a selected file or directory.

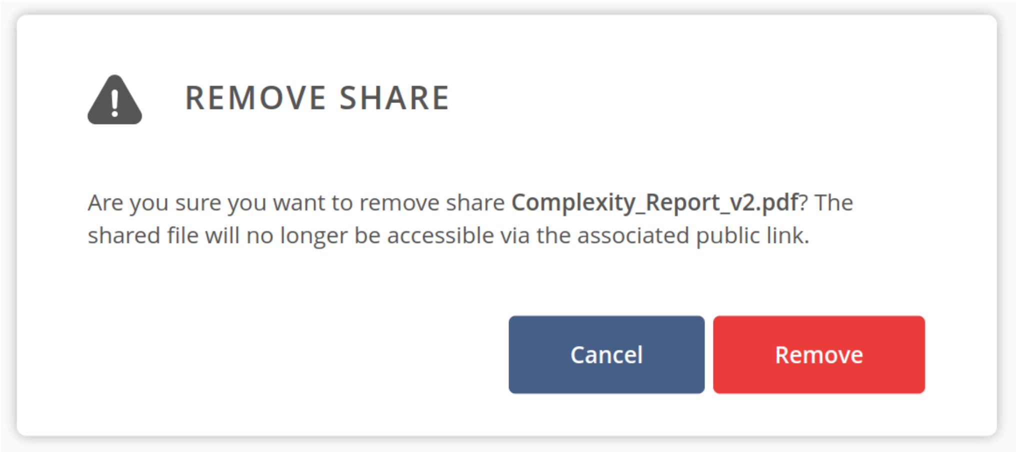


You can click on the green link (with the share icon and name) to navigate to the share details page (private view for share managers). Below you will find the public link for accessing the share (public view for anyone with the link). You can share this public URL with people interested in accessing this data.

To list all shares created in the space, navigate to the **Shares, Open Data** view. To open the action menu click the three dots button at the far right of the share row. The actions menu for Share provides the following operations: **Rename**, **Remove share**, and **Copy public URL**.

To remove a share, choose the **Remove share** option from the action menu and click the **Remove** button in the **Remove Share** modal.





Additional documentation:

- https://onedata.org/#/home/documentation/21.02/user-guide/shares.html

3.7 Publishing in Europeana

Updated 3D models can be published in Europeana through the EUreka3D platform. This process consists of three steps:

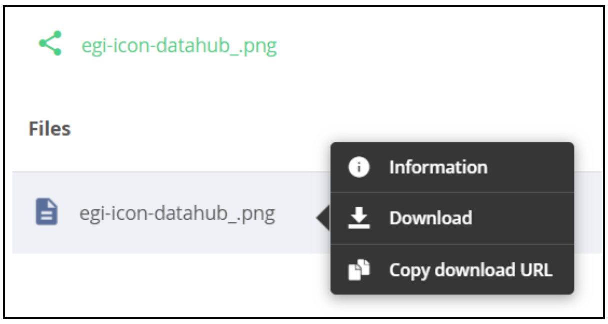
- **The request of a Persistent Identifier** (PID), which is a long-lasting reference to a digital object, a well-rooted standard in Open Data publishing. A PID is also a valid URL that can be visited to view the Open Data record in the public DataHub Web interface (publicly accessible without authentication). EUreka3D obtains PIDs from the external B2HANDLE service[4].
- **Creation of metadata**. A vital part of publishing is to include rich metadata that will make the record discoverable and meaningful for its consumers. All CH objects in EUreka3D use the EDM (Europeana Data Model) metadata format.
- **Data sharing with Europeana**. After a record is published, the Europeana service will automatically harvest its information (via the OAI-PMH protocol) and expose it in the Europeana Portal. The EDM metadata is ingested by Europeana and used to compile information about the CH object, so it must be well-curated.

The request of the PID and the data sharing are automatically done by DataHub, but content providers must provide the metadata for each object to be published. The following sections will guide you through the process of preparation of the data and their publishing in the Europeana Portal.

3.7.1 Preparing the model thumbnail

3D models can have a representative image that is used to give a first visual impression of the 3D object they will be seen in the 3D viewer. Herein, it is convenient to upload and share this static image of the 3D model before publishing it, by following these steps:

1. Create a representative image of the 3D model in JPG or PNG format.
2. Upload it in DataHub in the same directory as the ZIP file of the 3D model (directory **/Viewer** inside the project folder). Use the same file name suffixed by “thumbnail” for clarity.
3. Share the file publicly as explained in Section 3.6.2.
4. Go to the Share details.
5. Right-click on the file and click on “Copy download URL”. This will copy the necessary URL to your clipboard.



[4] https://www.eudat.eu/service-catalogue/b2handle

6. Note that you will need to use this URL for the “Representative image” attribute of the metadata.

NOTE: If the image is not shown in the metadata page once you publish the model, verify that **you are using the right URL** of the image. A simple test is to open the URL and see that you can visualise the image. One common error is to skip Step (4), so the right menu is not shown in Step (5).

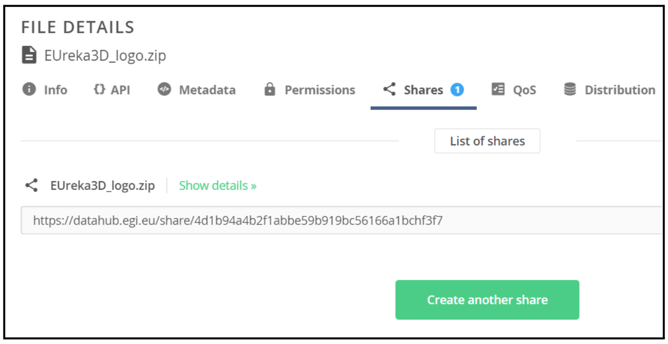
It is **strongly suggested** to upload this thumbnail as, not only DataHub, but other systems will use it to provide a first overview of the object. This is, for example, the listing of two EUreka3D objects where their corresponding images are shown:

3.7.2 Test the correct visualisation of your 3D model before publishing it

It is quite useful to test if a 3D model is visualised correctly before publishing it. If you need to do so, you can take a few steps:

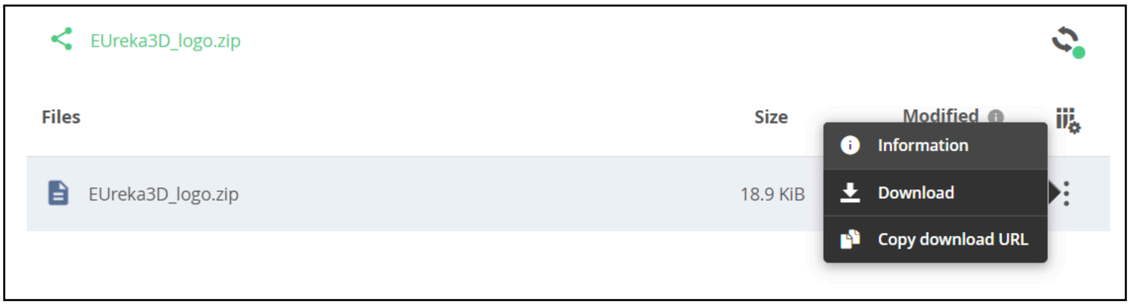
1. Upload the model normally, as explained in Section 3.5.
2. Then, provisionally share it as explained in Section 3.6.2. The file will present a “Shared” tag:

3. Click on the “**Shared**” tag: Shared This will open a screen like this one:

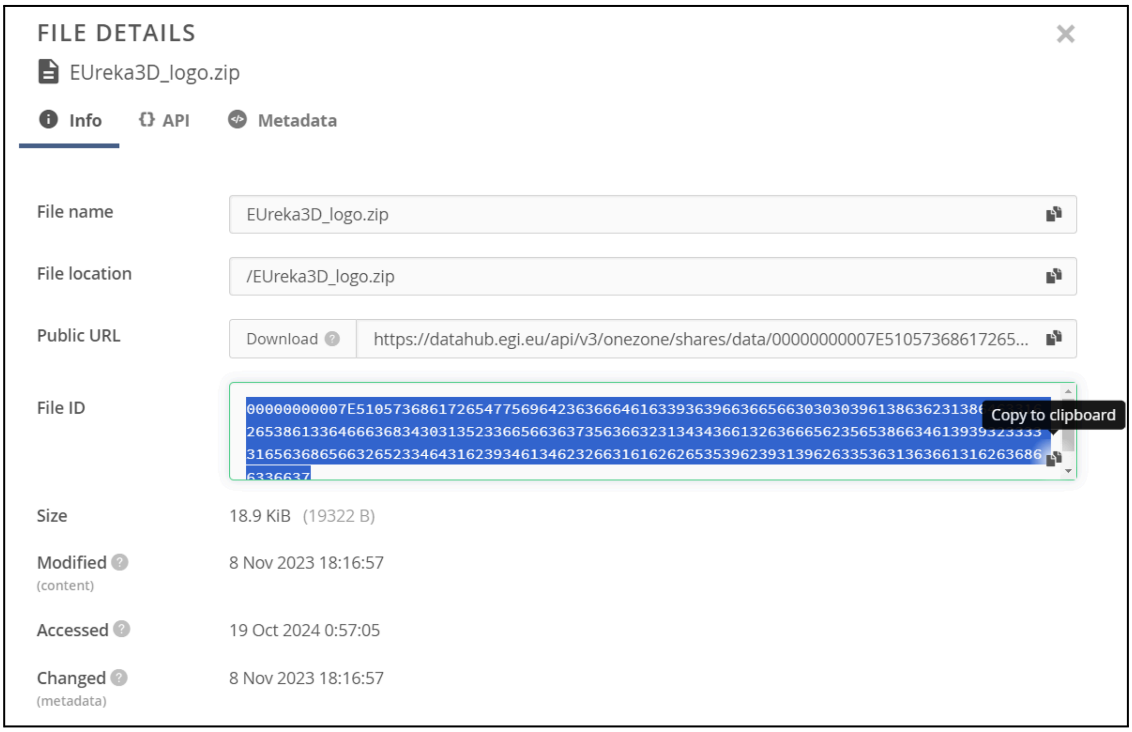


4. Click on “**Show details**”: Show details »

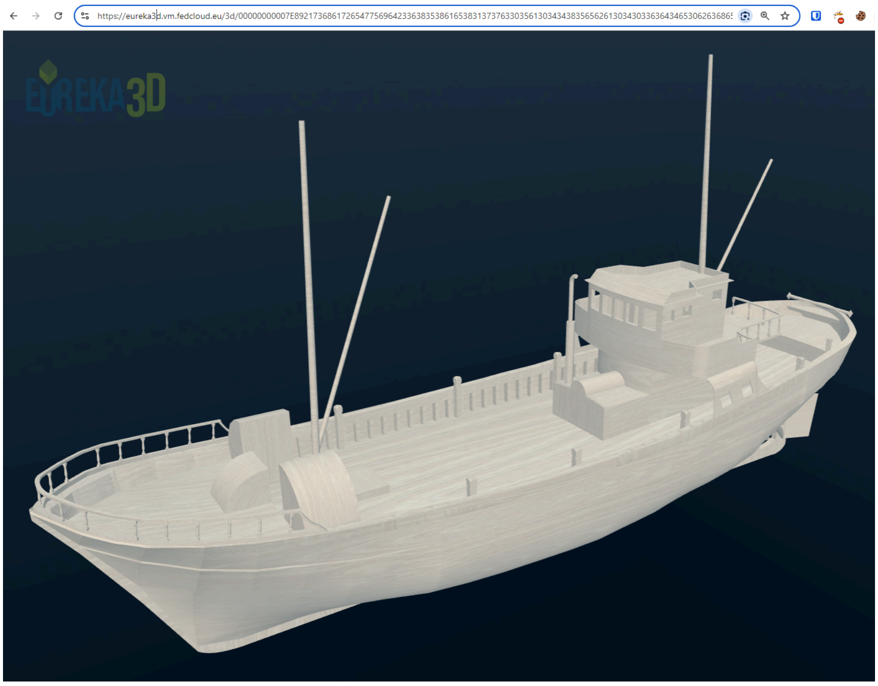
5. In the next screen click on the **three dots** (⋮) on the right side of the share to open the context menu, and then click on “**Information**”



6. In the next screen, copy the long number for the **File ID**:

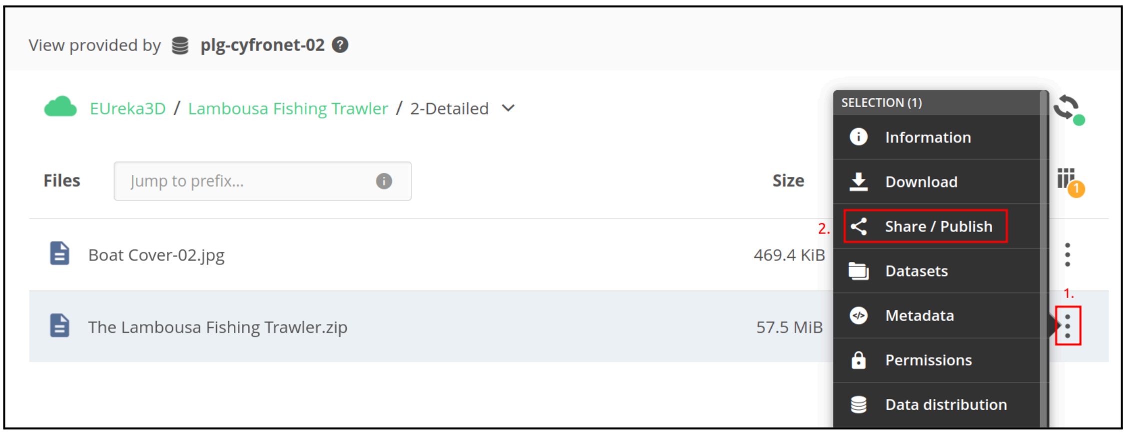


7. Finally, visit this URL:
   https://eureka3d.vm.fedcloud.eu/3d/<File ID>
   Where <File ID> is the long number you just copied.
   Example: *https://eureka3d.vm.fedcloud.eu/3d/00000000005276286775696423636664616339 3…*
8. If the 3D file is correct, this page will render the 3D model, so you can check how it looks. Once you finish your review, you can remove the Share you created in Step 2, as explained at the end of Section 3.6.2.

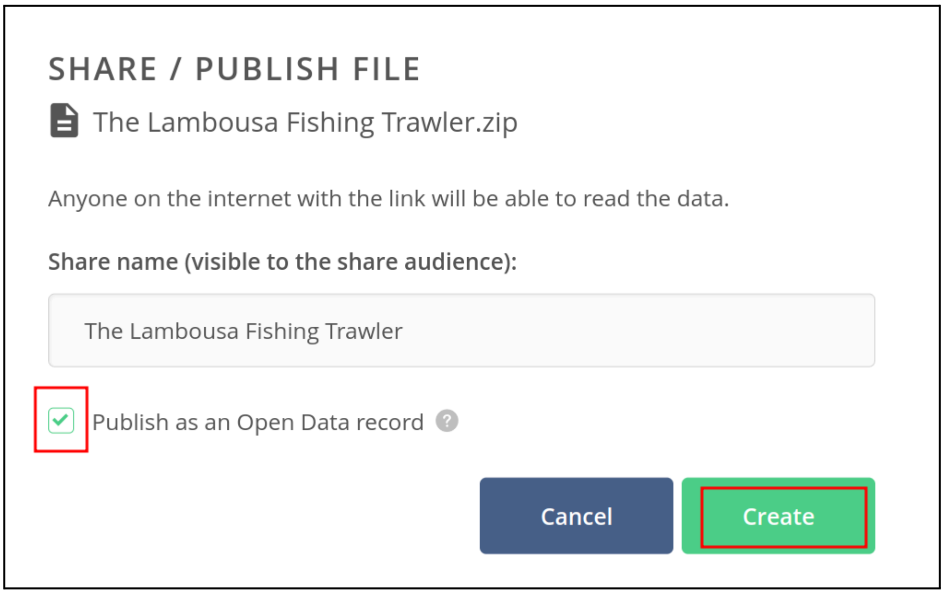


3.7.3 Create an Open Data record

To publish a 3D model click on the three dots button ⋮ at the far right of the file and click on **Share/Publish** from the dropdown menu, as shown below:

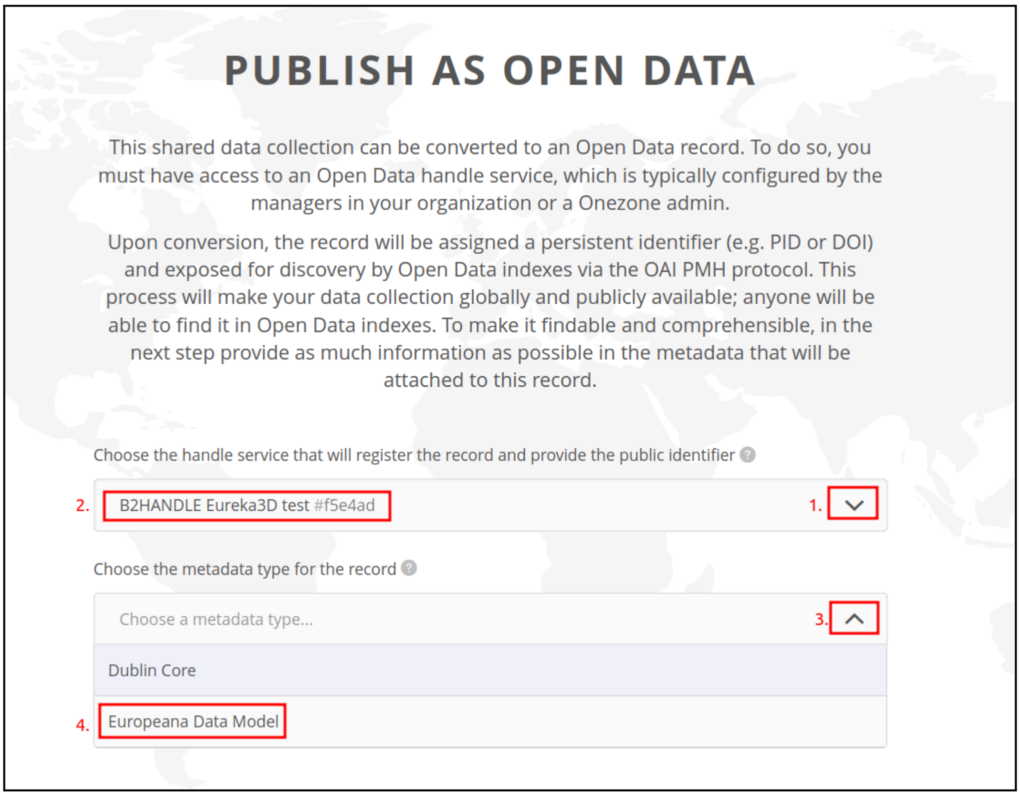


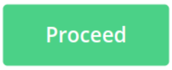
The process is similar to creating a share, but this time use the **Publish as Open Data record** option to create a share and convert it to an Open Data record in one go. Click the **Create** button to proceed:



NOTE: You can also take an existing Share and publish it as an Open Data record. Follow the instructions in Section 3.6.2 to open the share details. There, you will find the **Publish as Open Data** tab that will take you through the process, as described below. This way you can resume the Share+Publish procedure if you have not finished it, but the Share has already been created.

On the next page choose the **B2HANDLE** service. Next, select the **Europeana Data Model** as the metadata type:

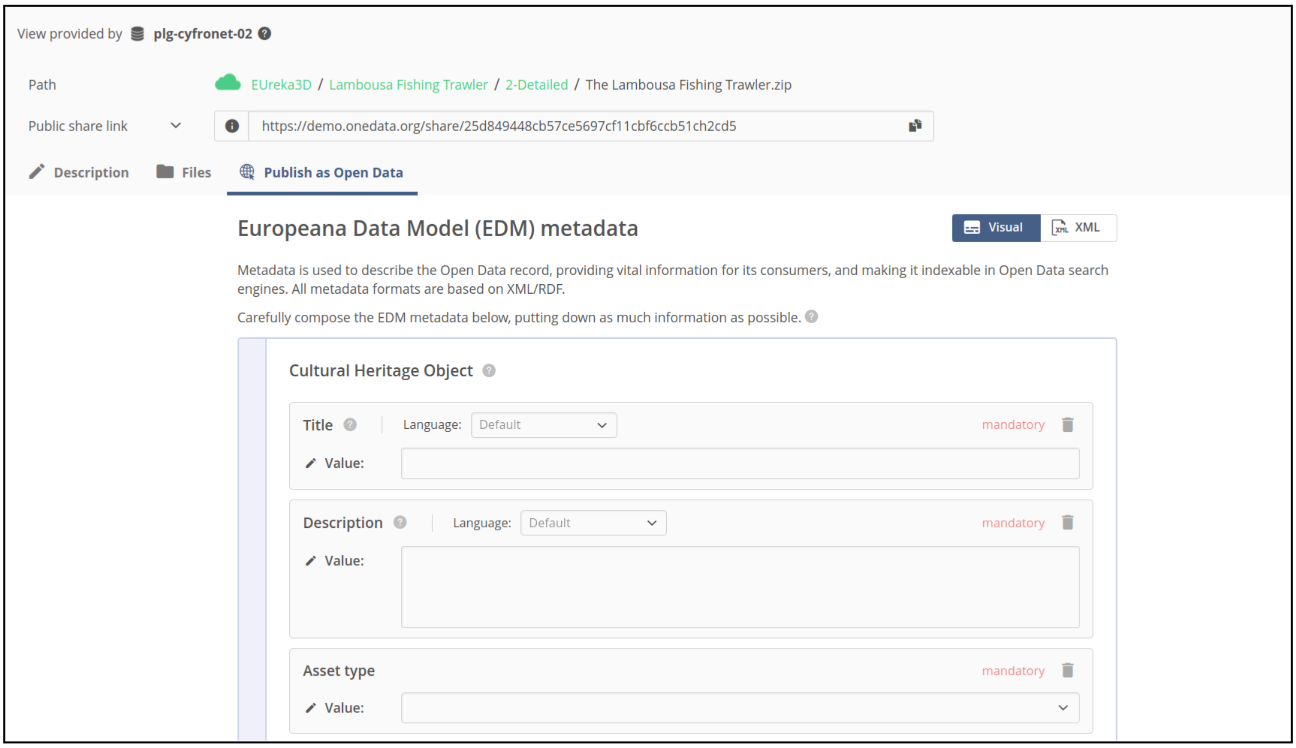


Click on the Proceed button 

.

IMPORTANT: This process will request a real PID (Persistent Identifier) from B2HANDLE. These identifiers are meant to be long-lasting references. If you want to do a test or you are not sure that the 3D model is correct, **do not publish it** in Europeana yet. For tests, ask a EUreka3D administrator to use the Testing environment. To verify the 3D model, you can proceed as explained in Section 3.7.2 to check its visualisation.

Pressing the Proceed button will lead you to a form to input metadata in EDM format.



There are two options to add the EDM metadata:

- By using a user-friendly **form**, which contains different fields for the different EDM attributes.
- By using **XML** directly to encode RDF data, which is a more advanced way to introduce EDM data.

To swap between these two modes, use the button at the top right of the form:

See Section 3.7.4 for a brief guide about the visual form and 3.7.5 for a brief guide about the XML editor.

NOTE: This document does not describe or explain EDM (Europeana Data Model). If you need to obtain more knowledge about this model, refer to the official documentation at: https://pro.europeana.eu/page/metadata

3.7.4 Adding metadata with the form

The form is a user-friendly way to introduce metadata, without needing any technical knowledge of EDM. Carefully complete the fields in the form, providing as much information as possible. Some fields are marked as mandatory and must be filled, but it is a good practice to consider and fill optional fields too.

The form is divided into three main sections to add metadata information, according to EDM:

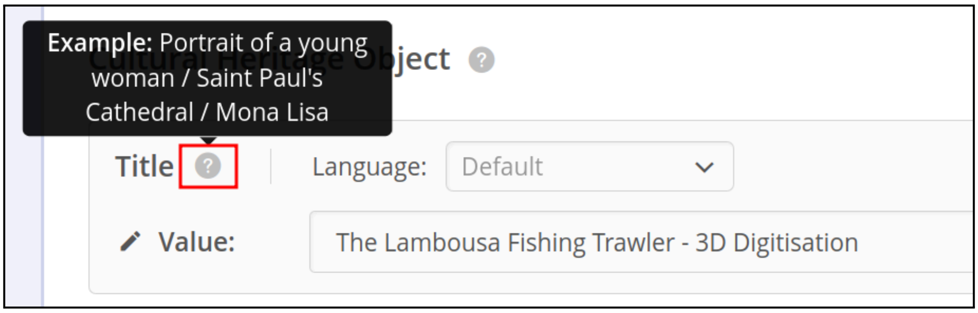
- **Cultural Heritage Object** (corresponding to *edm:ProvidedCHO* in EDM), which refers to information about the physical object in the real world.
- **Digital Object** (corresponding to *edm:WebResource* in EDM), which refers to information about the digitised 3D model of the object.
- **Aggregation** (corresponding to *ore:Aggregation* in EDM), which refers to information about the publication in Europeana and other external systems.

These three sections are mandatory to create a valid EDM record in EUreka3D.

A table with the current metadata fields can be found in Annex A.

The form presents the following features:

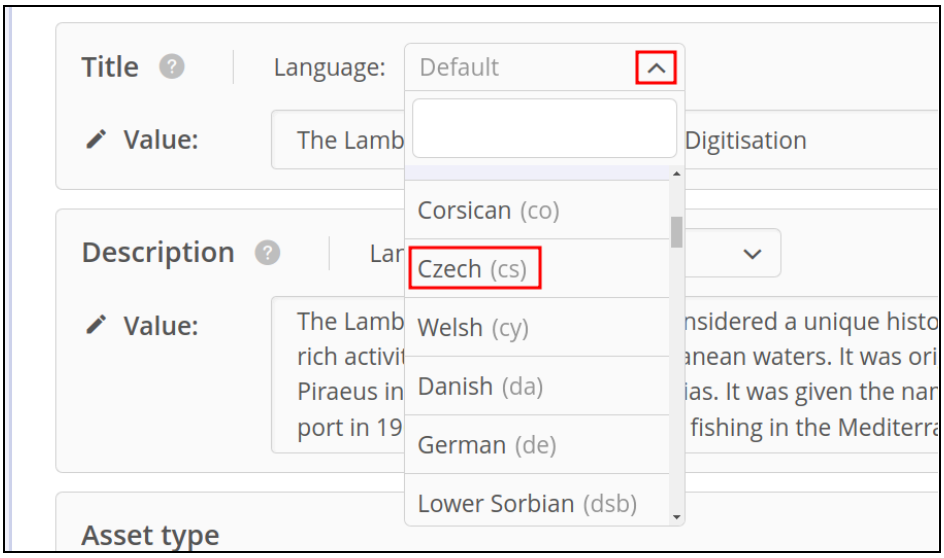
- If you do not know what a field means, you can access the field help, which may provide some examples. Just click on the question mark icon near the field name.



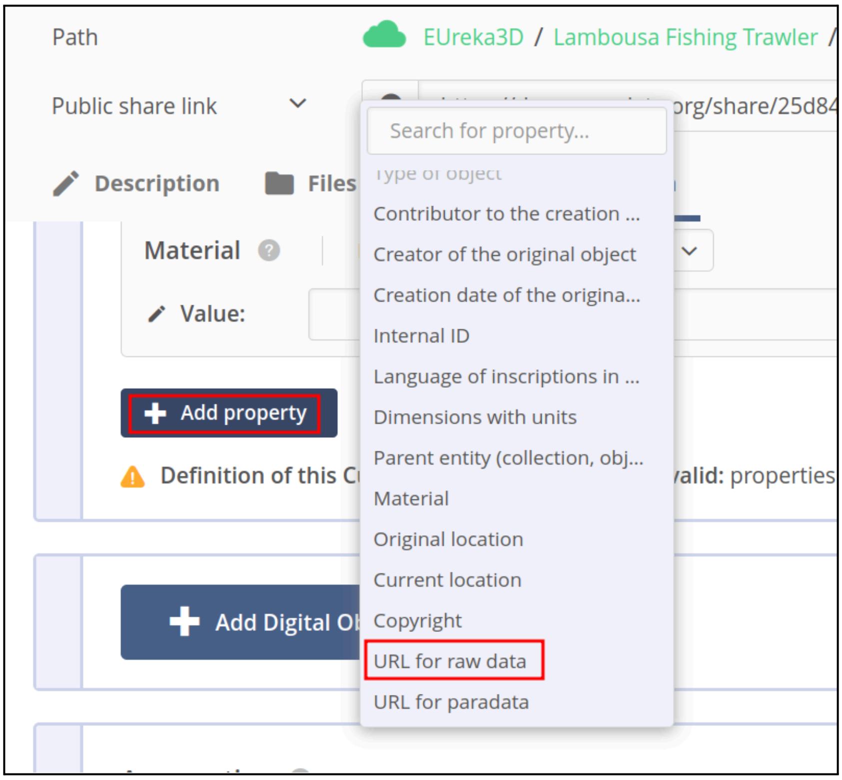
- Additionally, some fields will present an example to guide you about the sort of information that is expected for that attribute.

- To choose between **Literal value** and **Reference**, click the button near the field name. Literals represent text, such as “St. Paul's Cathedral”, whereas references refer to URIs. References can be used when a vocabulary is used (e.g. Getty) and is a convenient way to express interoperable data. It is suggested you use well-established references whenever possible.

- Some attributes allow information about the language in which they are written. For example, the name of an object (its “Title”) can be expressed in English or its original language (e.g. Greek). To change the field’s language, open the drop-down menu and select the desired language. By default, the language will be **English**.

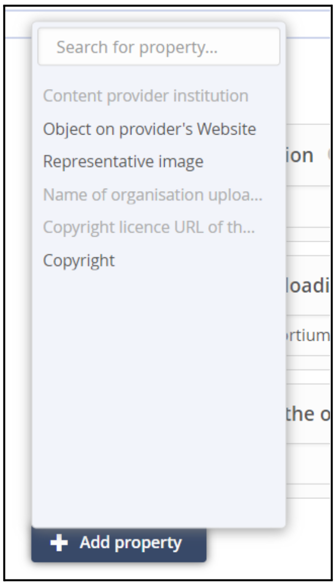


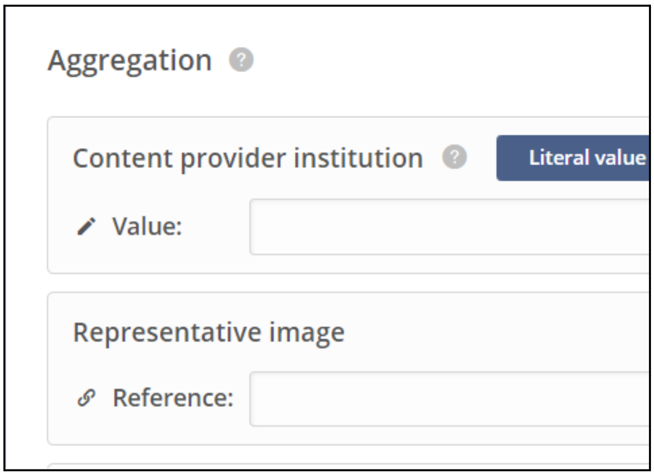
- Initially, not all fields are shown, only mandatory fields. This is done to simplify the form. To add optional fields, click the **Add property** button and choose the desired attribute.



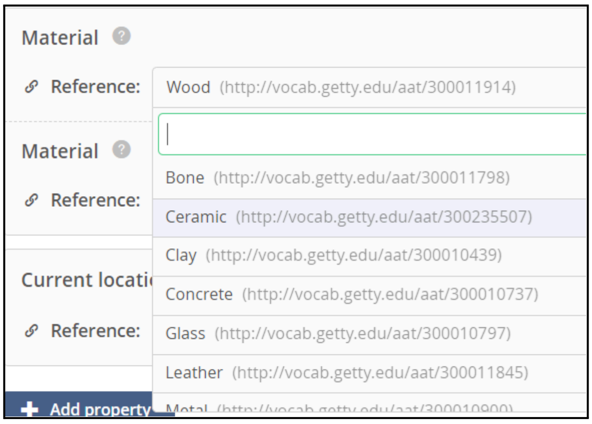
This option can be used to add existing attributes, in order to duplicate them. This allows, for example, the creation of multiple object titles that can have different languages (one in English and one in the original language).

- To add a representative image of the model, go to the **Aggregation** section and add a property **Representative image**. This field is optional, but it is **strongly recommended** you use it. To fill in the attribute, follow the steps listed in Section 3.7.1 and use the image public URL.

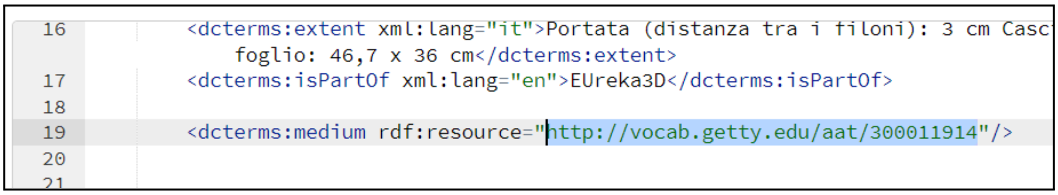




- The definition of Materials uses a predefined list of general groups (Bone, Ceramic, Clay, Concrete, Glass, Leather, Metal, etc):



Note that if you need to be more specific and use a material that is not in the predefined list, you can add it manually by using the RDF/XML editor (described in Section 3.7.5):



- To remove a field, click on the trash icon in the right corner.

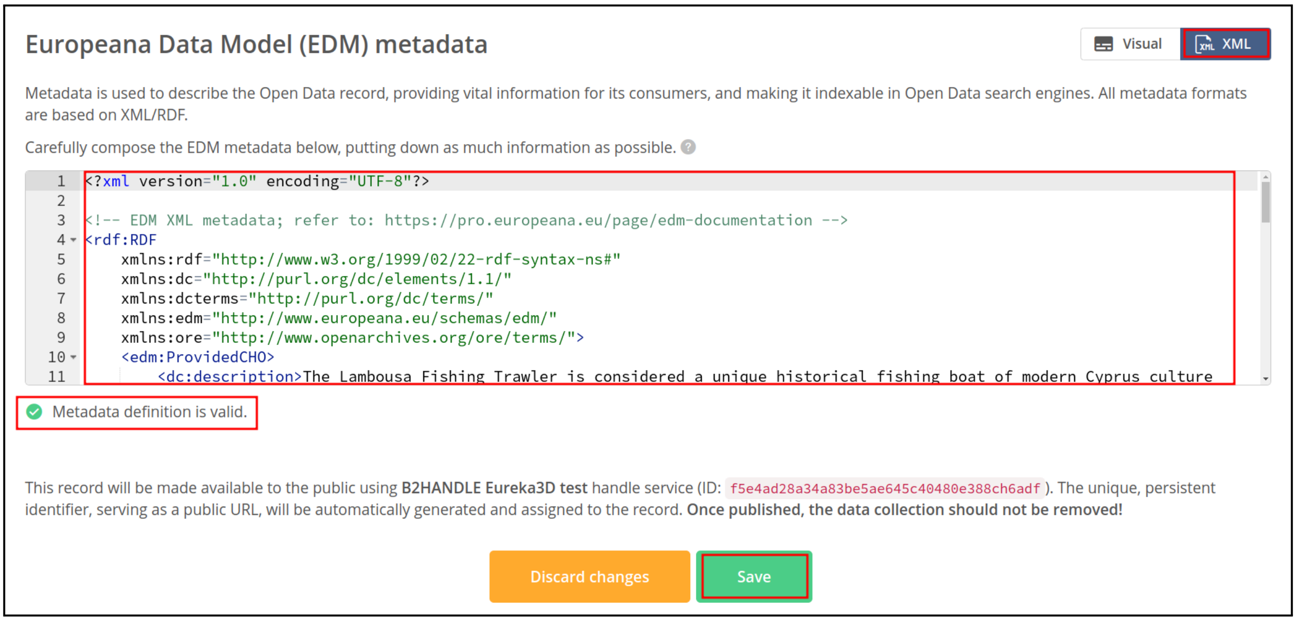
When you finish adding information, click on the **Save** button at the bottom of the page.

3.7.5 Adding metadata in RDF/XML

The XML editor mode displays the metadata expressed in the RDF/XML syntax. The carried information is equivalent to the information entered via the visual form, and is synchronised when the modes are switched.

You may edit and extend the metadata by inserting valid XML into the text field. It is also possible to paste in a whole XML document, as long as it is valid EDM metadata expressed in XML. The confirmation that the XML is valid will be shown on the page. See the Europeana documentation for detailed information about the EDM format and metadata mapping guidelines[5] [6] [7].

The XML mode lets you add extra fields that are not recognised in the visual form. They will be retained upon switching the modes — the visual form will display a note that the XML contains some extra information.



Once you finish editing the XML, click on the **Save** button at the bottom of the page.

[5] https://europeana.atlassian.net/wiki/spaces/EF/pages/987791389/EDM+-+Mapping+guidelines
[6] https://europeana.atlassian.net/wiki/spaces/EF/pages/1969258498/Metadata+Tier+A
[7] https://europeana.atlassian.net/wiki/spaces/EF/pages/2238447617/Examples+of+high+quality+data

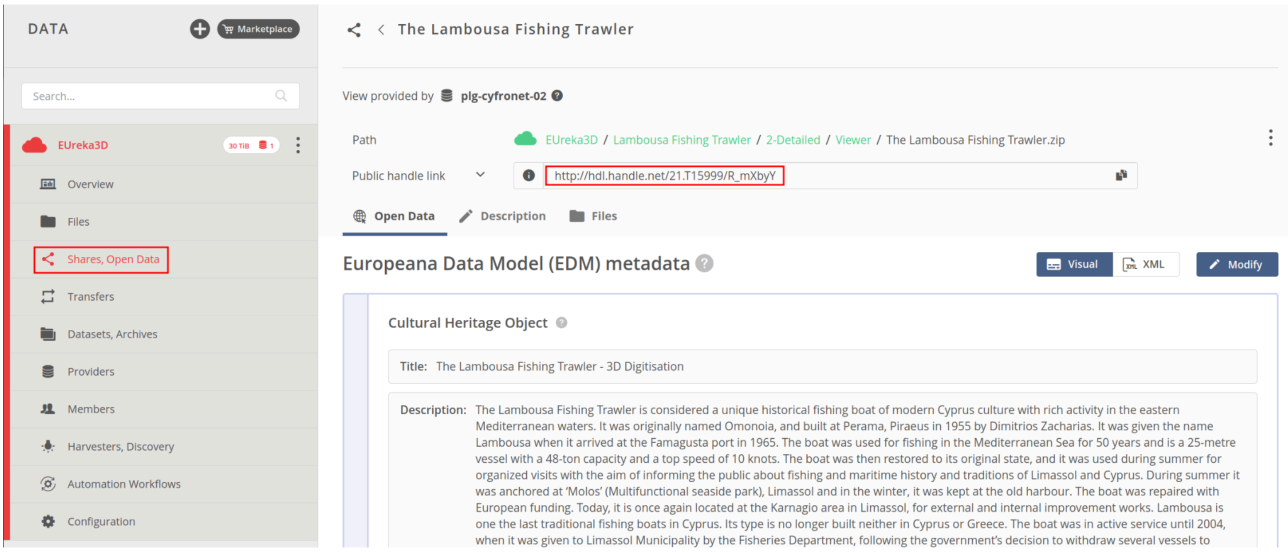
NOTE: Data introduced in the form will be shown in the XML, and vice versa. If the form does not accept some specific EDM attribute that you want to specify, use the XML mode to add it. The attribute will not be shown in the form but it will be sent to Europeana.

3.7.6 Publishing in the Europeana Portal

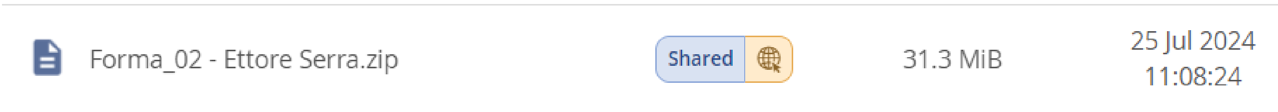
After you create the metadata, your model is published as Open Data with a PID, and it can be discovered by Europeana. Now, the Europeana systems will access your metadata at some point and ingest it in its database to be shown in its Web Portal.

NOTE: Europeana has not yet automated the process of detecting changes in the source systems, so they do some manual work to update their database Herein, **you may need to contact an EUreka3D administrator** to inform them that you have uploaded new models for publication in Europeana, so that they can contact Europeana administrators to trigger the updating process.

The metadata of the 3D model and its 3D visualisation will be public (no authorisation will be needed), and are accessible using the public handle link:



Published files will show a special icon next to the “Shared” tag to mark this situation:



You can use the **Shares, Open Data** menu in the sidebar to see the list of all shared data collections and Open Data records (an Open Data record is essentially a public Share that has been assigned a PID, metadata, and exposed to Europeana).

4. Automated publishing

The previous section has explained the process of managing your data with the GUI (Graphical User Interface) of DataHub, including uploading and sharing of models. However, when you have to upload many models (*e.g.* 500 models), working with the GUI to manually upload them one by one is a very tedious process. To solve this issue, you can use **DataHub's REST API** (Application Programming Interface). In simple terms, an API provides a way for a program to interact with a system, such as DataHub.

Using the API requires you to have some IT knowledge, at least to do some programming and understand basic Web concepts.

The process to use the API requires first to obtain a token, for security reasons. This is simply a string that gives your program permissions to do operations in DataHub in your name. This is discussed in Section 4.1.

Then, there are five steps that your program can conduct to upload and publish a model:

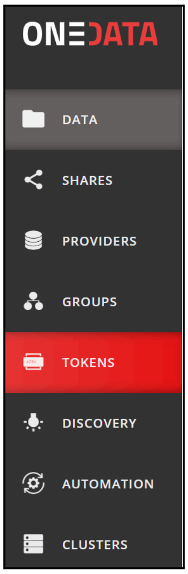
1. Create a project.
2. Check for the project creation before continuing with the rest of the steps.
3. Upload a 3D model.
4. Create a Share for the 3D model file.
5. Obtain a PID and upload the metadata.

Each of these steps will be discussed in Sections 4.2 to 4.6. Note that, strictly speaking, you do not need to perform all the steps with the API. You could for example create the project with the GUI, as explained in Section 3.3, and then automate the upload and publication of the models with the API.

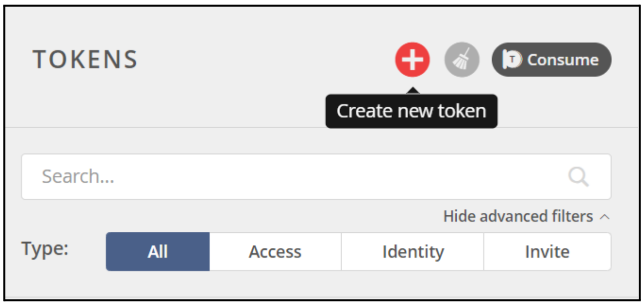
NOTE: For convenience, the examples given in this section use *curl*, but you are not expected to execute this manually. In real life, you will need to create a program or script in Python, Go, Java, Bash or any other programming language.

4.1 Obtaining access with the creation of a token

Log in to https://datahub.egi.eu and go to the menu option "Tokens":

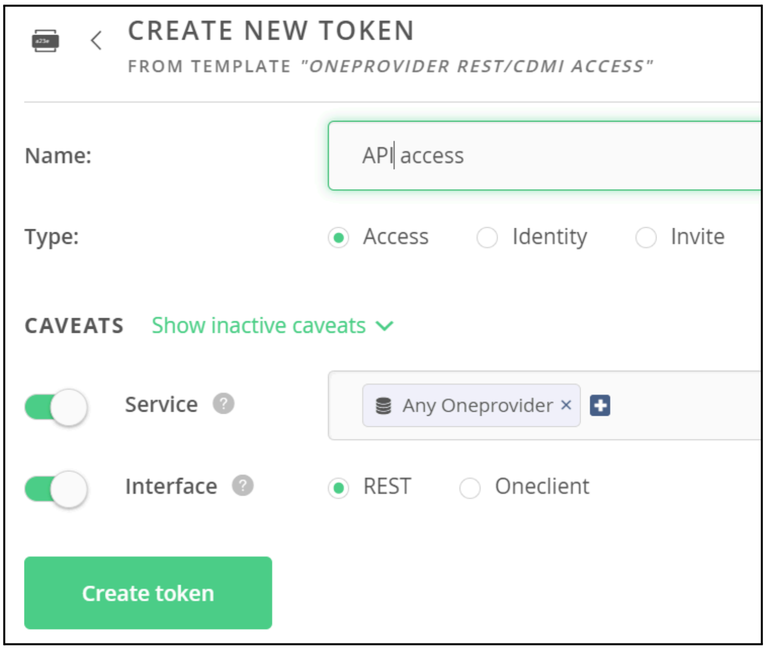


Click on the "+" sign to create a new token:



Choose the option "Oneprovider REST/CDMI access":

Give it a name, such as "API access" and click on "Create token":



The token is created:

You will need this token to call the different methods of the API.

4.2 Creation of a project

Similarly to what is explained in Section 3.3, you can create a project with the API.

Documentation:
https://onedata.org/#/home/api/stable/oneprovider?anchor=operation/schedule_workflow_execution

Example:

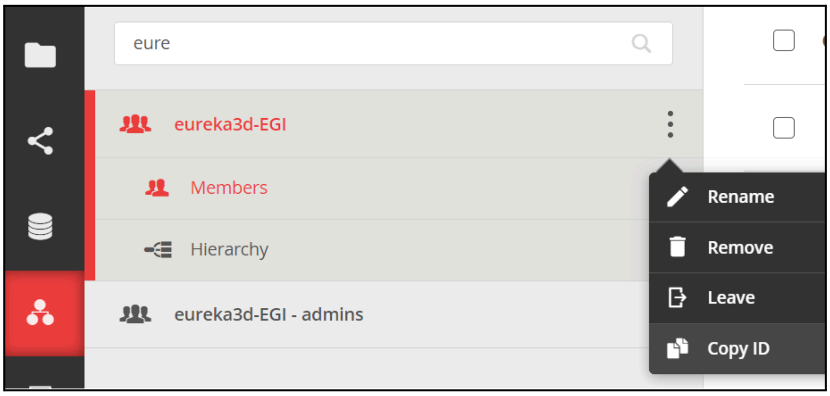
```
curl -H "X-Auth-Token: $TOKEN" -X POST
"https://plg-cyfronet-02.datahub.egi.eu/api/v3/oneprovider/automation/execution/wo
rkflows" -H "Content-Type: application/json" -d '{ "spaceId":
"6ef675cf21446a2cfeb5e8f4a992331echef2e", "atmWorkflowSchemaId":
"99ddd10bc160b73f65cb93af8444d465chb2ff", "atmWorkflowSchemaRevisionNumber": 1,
"storeInitialContentOverlay": { "cac81938bc14eb80041fe4ac51f13677abd874": "Example
project", "0fb9d0c480c2e5dbce39bc861b73ce6f53e4ca": { "fileId":
"000000000005855E5677569642373706163655F366566363735636632313434366132636665623565
3866346139393233333165636865663265233665663637356366323134343661326366656235653866
34613939323333316563686566326 5" }, "bb5c588454d4d79792964d18f019d478e596b3": [
{"groupId": "7ae89b417cf97219abe41c76a9a00f25chf972"}, {"groupId":
"13a7cc86a43e84328f74276a9a06b10031809"} ]}}'
```

Answer (you need this value for the next call):

```
{"atmWorkflowExecutionId":"fb15f9813710b2a62d1d621a85150860ch0cda"}
```

Notes:

- Marked in green are the values that you need to provide, in this case the name of the project and the “group ID” of the group(s) that you want to use for the ACL (see the part about “Managing Groups” in Section 3.3). To collect the group ID of one of your groups, you can access the GUI left menu: `Groups > click the three dots of the group > Copy ID`



- Currently, "atmWorkflowSchemaId" is always "99ddd10bc160b73f65cb93af8444d465chb2ff". This should stay like this, but you can always check the ID of a new workflow in the GUI left menu: `Automation > EUreka3D inventory > click the three dots of the workflow > Copy ID`
- Similarly, "atmWorkflowSchemaRevisionNumber" is currently “1”. If a new version is released you can find the revision number on the inventory page, through the GUI left menu: `Automation > EUreka3D inventory`

4.3 Checking the project has been created

Before moving to the upload of the 3D model you need to make sure that the step of Section 4.2 has finished successfully. This is necessary because running these steps in a program may run too fast, while the project creation workflow is run asynchronously in the background. This API call needs the “atmWorkflowExecutionId” value returned in the previous API call. You should periodically check for the *status*, and once it is “finished”, your program can proceed to the next call.

Documentation:

https://onedata.org/#/home/api/stable/oneprovider?anchor=operation/get_workflow_execution_details

Example:

```
curl -H "X-Auth-Token:  $TOKEN  " -X GET
"https://plg-cyfronet-02.datahub.egi.eu/api/v3/oneprovider/automation/execution/wo
rkflows/$WORKFLOW_ID" | jq .status
```

Answer:

```
"Finished"
```

Notes:

- WORKFLOW_ID is the value "atmWorkflowExecutionId" that was obtained in the previous step.

4.4 Uploading a 3D model

The next step is to upload the 3D model to the "Viewer" directory.

Documentation:

https://onedata.org/#/home/api/stable/oneprovider?anchor=operation/create_file_at_path

Example:

```
curl -H "X-Auth-Token: $TOKEN" -X PUT
"https://plg-cyfronet-02.datahub.egi.eu/api/v3/oneprovider/data/
6ef675cf21446a2cfeb5e8f4a992331echef2e/project_name/Viewer/file_name" -H
"Content-Type: application/octet-stream" -d "@model.zip"
```

Answer:

```
{"fileId":"00000000005855E5677569642373706163655F366566..."}
```

The file ID that is necessary for the next call.

Notes:

- The space ID for EUreka3D is always "6ef675cf21446a2cfeb5e8f4a992331echef2e"

4.5 Create a share for the 3D model

Once the file is uploaded, a share must be created.

Documentation:

https://onedata.org/#/home/api/stable/oneprovider?anchor=operation/create_share

Example:

```
curl -H "X-Auth-Token: $TOKEN" -X POST "https://
plg-cyfronet-02.datahub.egi.eu/api/v3/oneprovider/shares" \
-H "Content-Type: application/json" -d '{"name": "Example of share name",
"rootFileId": "'$FILE_ID'"}'
```

Answer:

```
{"shareId":"8f4a99221446a2cfe336ef675cfb5e1echef2e"}
```

The share ID that is necessary for the next call.

Notes:

- The share must have a meaningful name, which can be similar to the model name.
- "rootFileId" should have the *fileID* that was obtained in the previous step, for example: "rootFileId": "00000000005855E567756964237370616365 5F366566..."}
  (Note that the simple quotes are not necessary when a Bash variable is not used).

4.6 Publish the 3D model

Similarly to what is explained in Section 3.3, you can create a project with the API.

Documentation:

https://onedata.org/#/home/api/stable/onezone?anchor=operation/handle_service_register_handle

Example:

```
curl -H "x-auth-token: $TOKEN" -H "content-type: application/json" -X POST
https://datahub.egi.eu/api/v3/onezone/handles -d '{ "handleServiceId":
"'$HANDLE_SERV_ID'",  "resourceType": "Share",  "resourceId": "'$SHARE_ID'",
"metadataPrefix": "edm",  "metadata": "<?xml version=\"1.0\" encoding=\"utf-8\"
?>..."  }'
```

Notes:

- HANDLE_SERV_ID for the “handleServiceId” should contain the specific value of your EUreka3D group, for example: f5f5e13338851b6866b085fee019276ech038f.
  You can find your ID by calling: https://datahub.egi.eu/oai_pmh?verb=ListSets and looking for your group name (*e.g.* eureka-EGI for the EGI group).
- SHARE_ID for the “resourceId” is the *shareId* that was obtained in the previous API call.
- “metadata” should contain the EDM data in RDF/XML format. A draft for this template can be found in Annex B. Please check Annex A to see the list of metadata attributes that currently exist.

4.7 Final suggestions

The API is a very powerful feature that has to be used with care:

- When you work with the API, it is convenient to test your program with a single object first.
- Up to the last step (after you create the share as explained in Section 4.5), it is easy to revert what was created in case of any error, but the last step retrieves a PID from B2HANDLE, so it should not be executed for testing purposes.
- It may be appropriate to test the 3D visualisation after the share has been created, in a similar way as explained in Section 3.7.2.
- It is useful to use the XML Editor of the GUI (see Section 3.7.5) to validate your metadata RDF/XML. However, remember **NOT to save it**, as this step will request a real PID.
- Additionally, it is useful to publish one object with the GUI, so that you can see what metadata RDF/XML is generated, and what sort of data can be added.
- Once everything has been verified to work well, you can run your program to upload and publish multiple objects.

5. Conclusions

This handbook has explained the basics for Content Providers to upload, manage and share data in the EUreka3D platform. It first described the required steps to join the EUreka3D community through EGI Check-in, in order to gain access to the different services provided by the project. This consists mainly of two tasks:

- Registering an account in Check-in.
- Joining the EUreka3D Community.

Then, the handbook explained the DataHub service and the different tasks involved, including:

- Creation of new projects for 3D models.
- Upload of data.
- Assignment of permissions for data access.
- Sharing of data publicly.
- Introduction of metadata, through a user-friendly form and a more advanced RDF/XML editor.
- Final publishing in Europeana.

Thanks to the graphical user interface of DataHub, uploading data to the cloud is a simple process. In case multiple objects have to be uploaded at once, the API provides an efficient mechanism to tackle the task. The main basic steps have been explained in this handbook, with references to documentation for extended functionality.

Although the services supporting EUreka3D, such as EGI Check-in and DataHub, have been successfully used by many scientific communities in production for a long time, they are under continuous Improvements. Only the provisioning of a good user experience that remains useful for the Cultural Heritage community will determine the success of the EUreka3D platform.

Annex A. Metadata Fields

Note that this table may be outdated. Always refer to the metadata form found in DataHub and to the latest EDM version.

Name	Section	Literal/Reference	Cardinality	RDF mapping	Lang att	Notes
Title	Cultural Heritage Object	Literal	1..n	dc:title	yes	Lang attribute is mandatory in Europeana
Description / Caption	Cultural Heritage Object	Literal	1..n	dc:description	yes	Lang attribute is mandatory in Europeana
Category	Cultural Heritage Object	Literal	1	edm:type		
Subject	Cultural Heritage Object	Literal	1..n	dc:subject	yes	Lang attribute is mandatory in Europeana
Type of object	Cultural Heritage Object	Either	1	dc:type	yes	
Contributor to the creation of the original object	Cultural Heritage Object	Either	0..n	dc:contributor	yes	
Creator of the original object	Cultural Heritage Object	Either	0..n	dc:creator	yes	
Creation date of the original object	Cultural Heritage Object	Either	0..n	dcterms:created	yes	
Language of inscriptions in the object	Cultural Heritage Object	Literal	0..n	dc:language		
Dimensions with units	Cultural Heritage Object	Literal	0..n	dcterms:extent		

Parent entity (collection, object, site...)	Cultural Heritage Object	Either	0..n	dcterms:isPartOf	yes	
Material	Cultural Heritage Object	Reference	1..n	dcterms:medium		
Original location	Cultural Heritage Object	Either	0..n	dcterms:spatial		Entries of geonames vocabulary must end in a "/" for Europeana to accept it. Example: <dcterms:spatial rdf:resource="https://sws.geonames.org/2794576/"/>
Current location	Cultural Heritage Object	Either	0..1	edm:currentLocation		
Description of digital object	Digital Object	Literal	1..n	dc:description	yes	Lang attribute is mandatory in Europeana
Type of digital object	Digital Object	Literal	1	dc:type	yes	
Creator of the model	Digital Object	Either	0..n	dc:creator	yes	
Digitisation date	Digital Object	Either	0..1	dcterms:created	yes	
3D format	Digital Object	Literal	0..n	dc:format		
Internal ID	Digital Object	Literal	0..1	dc:identifier		
File size	Digital Object	Literal	0..1	dcterms:extent		
URL for raw data	Digital Object	Reference	0..1	dcterms:isFormatOf		
URL for paradata	Digital Object	Reference	0..1	dcterms:isReferencedBy		
Content provider institution	Aggregation	Either	1	edm:dataProvider	yes	

Object on provider's Website	Aggregation	Reference	0..1	edm:isShownAt		
Representative image	Aggregation	Reference	1	edm:object		
Copyright licence URL of the digital object	Aggregation	Reference	1	edm:rights		
Additional copyright information	Aggregation	Literal	0..n	dc:rights	yes	

ANNEX B. EXAMPLE OF EDM TEMPLATE

```xml
<?xml version="1.0" encoding="utf-8"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:dcterms="http://purl.org/dc/terms/"
xmlns:edm="http://www.europeana.eu/schemas/edm/"
xmlns:ore="http://www.openarchives.org/ore/terms/">
   <edm:ProvidedCHO rdf:about="">
       <dc:contributor>....</dc:contributor>
       <dc:creator>...</dc:creator>
       <dc:description>...</dc:description>
       <dc:language>...</dc:language>
       <dc:relation
rdf:resource="https://datahub.egi.eu/share/60261548ca50a1658ee35b1416f236c8ch9454"/>
       <dc:subject>...</dc:subject>
       <dc:title>...</dc:title>
       <dc:title xml:lang="it">....</dc:title>
       <dc:type>...</dc:type>
       <dcterms:created>...</dcterms:created>
       <dcterms:extent>...</dcterms:extent>
       <dcterms:isFormatOf rdf:resource="..."/>
       <dcterms:isPartOf>EUreka3D</dcterms:isPartOf>
       <dcterms:medium rdf:resource="http://vocab.getty.edu/aat/300235507"/>
       <dcterms:spatial rdf:resource="https://sws.geonames.org/2794576/"/>
       <edm:currentLocation rdf:resource="http://...."/>
       <edm:type>3D</edm:type>
   </edm:ProvidedCHO>
   <edm:WebResource rdf:about="">
       <dc:creator>...</dc:creator>
       <dc:description>...</dc:description>
       <dc:format>OBJ</dc:format>
       <dc:identifier>...</dc:identifier>
       <dc:type>3D</dc:type>
       <dcterms:created>...</dcterms:created>
       <dcterms:extent>1.5 MB</dcterms:extent>
       <dcterms:isFormatOf rdf:resource="http://...."/>
       <dcterms:isReferencedBy rdf:resource="http://...."/>
       <edm:rights rdf:resource="http://creativecommons.org/licenses/by-sa/4.0/"/>
   </edm:WebResource>
   <ore:Aggregation rdf:about="">
       <dc:rights>...</dc:rights>
       <edm:aggregatedCHO rdf:resource=""/>
       <edm:isShownAt rdf:resource="https://...."/>
       <edm:dataProvider>...</edm:dataProvider>
       <edm:object rdf:resource="https://datahub.egi.eu…."/>
       <edm:provider>Photoconsortium</edm:provider>
       <edm:rights rdf:resource="http://creativecommons.org/licenses/by-sa/4.0/"/>
   </ore:Aggregation>
</rdf:RDF>
```