

Europeana Cloud User Documentation

This is a document that will guide you through what Europeana Cloud Services. First there is a description of Europeana Cloud itself and its data model. In the next pages you will find a more detailed description of some of the services as well as the user tutorial.

- [Introduction](#)
- [Data Model](#)
- [Services](#)
 - [Storage Service](#)
 - [Data Processing Service](#)

Introduction

Europeana Cloud is Europeana's new cloud-based infrastructure for storing and sharing cultural heritage data. The goal of this tutorial is to introduce its Developer API.

The data stored by Europeana Cloud is structured according to the Europeana Cloud Data Model (NOTE: Please be aware that Europeana Cloud Data Model is different than the Europeana Data Model (EDM). Differences between these models are presented later on in this tutorial). The API of Europeana Cloud consists of standard REST services over HTTP which allow structuring data according to the data model and storing it in the system. To use the API, it is essential first to understand the data model, the subject of the next section.

Data Model

The goal of Europeana Cloud is to allow storing, sharing and processing data of various types. The data model was designed to capture relationships between various pieces of data which support data aggregation workflows. Several types of entities make up the data model.

[blocked URL](#)

Digital objects and metadata records, most common types of cultural heritage data, are stored in physical files. Often several files represent the same object, for example a master file of a scanned artwork and a metadata record of this artwork. We call these different ways to represent a cultural heritage object *representations*. In the last example we'll then have two representations: one for the scan and one for the metadata record. It is possible that one representation will contain several physical files, for example when a newspaper scan is spread over several files, one for each page.

As a whole, a representation, can change over time. For example, a metadata record can be edited or a scan can be replaced with a better one. To reflect these changes, representations have *versions*. That is, more than one version of a representation can exist and consecutive versions of a representation correspond to its consecutive changes. Versions will always be numbered. We will usually look at the most recent version but we'll sometimes want to keep the old ones, e.g. for provenance needs, or when it is already in use by someone.

Altogether, various representations of the same object form a *record*. Hence, a record is a collection of all available representations of a cultural object upon its various versions and files.

Finally records can be grouped for various reasons. These groups of records are called *datasets*. A single record can belong to many datasets.

This data model can look complicated, but it is really not. We'd like to demonstrate it with a simple example from real life.

Consider, for example, a data provider which has a metadata record about a painting in an XML file, a scan of the painting in a TIFF file and a preview version of the scan in a JPG file. To store this data in Europeana Cloud the data provider could create a record consisting of three representations conveniently named after the file types - XML, TIFF and JPG. But such an approach may be misleading if for some reason both the full and preview version of the image will be in JPG format or if XML format will be also used to store metadata in another schema. Therefore the suggestion is to avoid representation names limited to the technical file format and use more meaningful names e.g. "master", "preview", "EDM", "METS" etc. especially that the information about technical file formats can be retrieved from Europeana Cloud as technical metadata associated with versions (NOTE: It is also planned to develop guidelines regarding the shared representations naming to facilitate the reuse of representations in the aggregation workflow. This development will be part of the process of migration of Europeana data to Europeana Cloud storage services.). Each representation will have one file and the entire structure will look like this:

[blocked URL](#)

Our data provider can decide that to one of the representations a new file needs to be added:

[blocked URL](#)

One day a better scan of the painting arrives and another version is added to the TIFF representation:

[blocked URL](#)

Finally, our data provider can add more records to the system and divide them into two groups. For this two datasets are created and individual records are assigned to these datasets:

[blocked URL](#)

Finally, it is important to mention the relation between the described data model and [EDM](#), Europeana Data Model. EDM is a high-level conceptual model for describing cultural heritage objects. As you have just seen, the Europeana Cloud Data Model is a low-level model for storage and aggregation purposes. The models serve different needs and are easily compatible. Europeana Cloud Data Model can store EDM representations in parallel to representations in other formats.

Services

Europeana Cloud is a SaaS cloud, which is provided on top of IaaS cloud. It consists of several microservices. Taking into account the responsibilities of the services they can be divided into two groups:

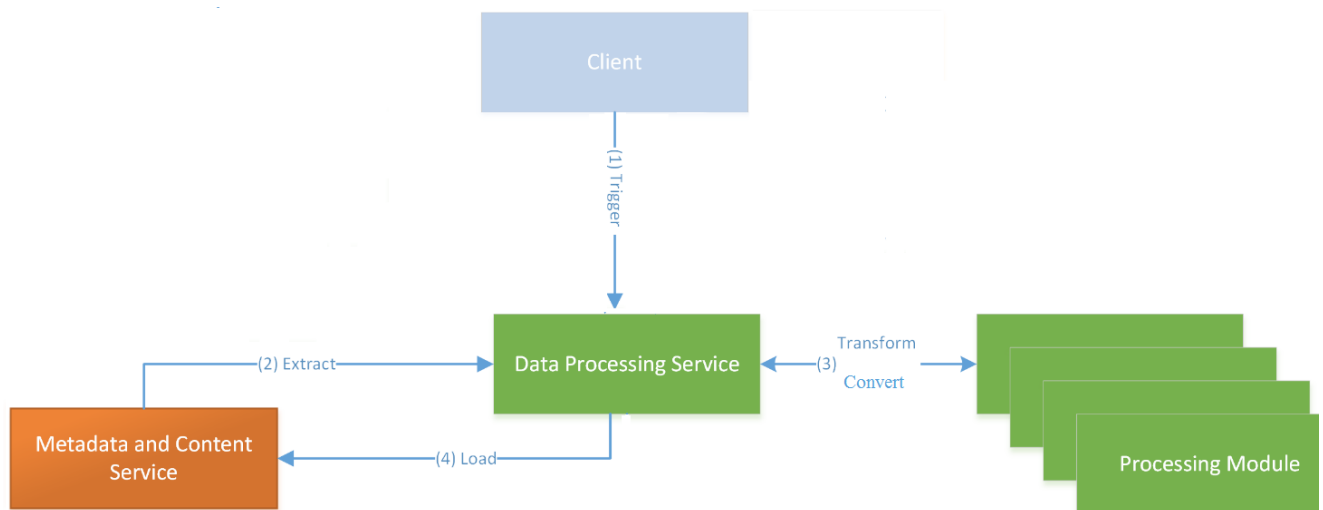
Storage Service

Services that let you store and retrieve your data in the Cloud within the data model described above. Their functionality covers:

- provide global identifiers for data records
- provide the mechanism to create mappings between local identifier (scoped with the data provider) and the global identifiers inside the eCloud system scope. storing, retrieving etc.
- provide the create/read/update/delete operations for cultural data records in multiple representations and versions.
- provide the ability to restrict access rights to the data stored in eCloud
- retrieve images stored in the storage using IIIF protocol with [Image Service](#)

Data Processing Service

A part of eCloud dedicated to provide data processing facility. The client triggers the interaction with DPS by sending a task containing list of files in a purpose of processing them. This processing is done by build-in processing services (topologies) created for specific processing purposes. DPS task should be formulated to have the necessary data needed to point to specific files inside Europeana Cloud Storage (Metadata and Content Service - MCS), those files will be processed by DPS topologies and then uploaded back to the Storage Cloud.



Right now two processing services are available:

- XSLT topology: used to transform XML files based on XSL transformation file.
- Image conversion topology: used to convert TIFF type images into JP2 type images.

Data Processing Service was designed to be pluggable. Extending the current processing services was a major part of its design. To integrate your necessary processing services please contact eCloud Team.