DELIVERABLE

Project Acronym:	**Europeana Cloud**
Grant Agreement number:	**325091**
Project Title:	**Europeana Cloud: Unlocking Europe's Research via The Cloud**

D2.6 - Metadata and Content Cloud Delivered

Authors:

Pavel Kats (EF)
Alastair Dunning (EF - Review)
Valentine Charles (EF - Review)

Revision: 1

Project co-funded by the European Commission within the ICT Policy Support Programme
Dissemination Level

P	Public	P
C	Confidential, only for members of the consortium and the Commission Services	

Revision History

Version	Status	Author	Partner	Date	Changes
0.1	First Version	Pavel Kats	EF	01/09/2015	
0.2	Revision	Alastair Dunning Valentine Charles	EF EF	02/09/2015	
1.	Publish Version	Pavel Kats	EF	Jan 2015	Adopting reviewers' comments

Statement of originality:

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

Abstract

This deliverable describes Europeana Cloud Services, the new suite of metadata and digital content services, developed by the Europeana Cloud project. It describes how the vision of the product was shaped, its scope (by the end of the project), user experience and some future plans.

Content

Introduction

The Europeana Cloud project started almost three years ago with a vision of creating a cloud-based infrastructure for European cultural heritage institutions to store and share data. During its course, the project has rethought its vision, the roadmap to achieve it, and the required organisational setup. Being agile, we needed to adjust our plans and priorities to reality.

One of the key insights here was that providing a cloud-based infrastructure for storing data is not enough. What organisations in our domain are sorely missing is efficient tools and services to operate and manipulate their data: improve its quality, convert to new formats and standards, support bi-directional flows and, ultimately, distribute it further to hubs, such as Europeana. Existing tools and services are disparate; they can not be easily integrated into continuous flows and are not suitable for growing demands for volume of data and complexity of flows. Not having powerful infrastructures is a part of the problem but not all of it.

The project has rethought its role in addressing the need from the sector for more efficiency in processing data. We decided to focus on two things: a reliable and open storage platform and a clear product vision for granular and independent metadata and content services operating on top of it. This vision is not abstract; it is put into practice by the project itself in 3-5 services that will be released before the project ends. The design of these services will guide the development of future ones, based on the same storage platform. In addition to the released services which are fully functional products on their own right we are pioneering the bigger idea of a service marketplace in the cultural sector, a popular concept in other sectors, such as mobile applications, e-commerce, translation services and more.

This is a new and bold vision for Europeana. It requires innovative thinking in many areas: technology, business models, marketing and sales, customer management and governance. The role of this deliverable is to describe technological and product-related aspects of how we plan to pursue this vision. The rest of the aspects are addressed by other deliverables due till the end of the project.

Cloud Services for Cultural Heritage

Over the last decade a major trend is reshaping how organisations build and use software: service-centered architectures. From complex monolithic systems, evolving over years, organisations move to dynamic stacks built of granular services, independently powered by specialised providers. Today, service-centered architectures are a necessity rather than an

innovation: constantly changing business landscapes require agile planning and short time to market, which are hardly achievable goals with slow-moving legacy software.

Service-oriented approaches got propelled by fast adoption of cloud technologies. In the cloud age services became reliable and scalable in addition to being granular and focused. And since the reliability of a complex workflow is crucially dependent on the reliability of each of its parts, it pays off to companies to invest in hosting individual parts on cloud.

These major changes are all around but not in the cultural heritage sector, which is still reigned by old-style waterfall projects, centralised budgets and large-scale tenders. Europeana sees its opportunity to enter the void and become the first provider of cloud-based services for the sector. The Europeana Cloud project - technical, business and organisational ones - are a perfect opportunity to test the new waters.

Our vision is to offer a suite of small, independent, and interoperable services addressing the needs of cultural heritage sector in managing metadata and services, for example search across metadata records, semantic enrichment, conversion between formats, or publishing high-resolution digital content. The services operate on top of a cloud-based storage platform where users store data to be used by services. The platform is accessible through a standard REST Developer API. We are working also on interactive interfaces for data upload. Users can control access to their data uploaded to the platform following the principles outlined in the Cloud Access Framework, Deliverable 5.3 in the project[1].

Europeana and its New Vision

Europeana is undergoing a transition period. It is moving from being funded by a series of thematic projects with a limited scope to becoming a DSI[2], supported through recurring funding rounds backed by CEF[3], with the remit to explore additional sustainability options. One of the core assets of Europeana as DSI is its technical infrastructure. It underpins business processes run by Europeana and is composed of software applications for aggregating cultural heritage data, improving its quality, and distributing it further on.

The transition from projects to DSI requires restructuring this infrastructure to make Europeana fit for the purpose of a DSI. Europeana Cloud is one of the means to reach this goal. Over the last months Europeana and the Europeana Cloud consortium have planned the integration of

---

[1]

http://pro.europeana.eu/files/Europeana_Professional/Projects/Project_list/Europeana_Cloud/Deliverables/D5.3%2520Europeana%2520Cloud%2520Access%2520and%2520Reuse%2520Framework.pdf&sa=D&ust=1443521295721000&usg=AFQjCNG1UHTnEShthrBLTYiVwCvZXFQ02Q

[2] https://ec.europa.eu/digital-agenda/en/news/connecting-europe-facility-cef-digital-service-infrastructures
[3] https://ec.europa.eu/digital-agenda/en/news/connecting-europe-facility-cef-digital-service-infrastructures

the outcomes of the project into Europeana's infrastructure, sustaining them under the umbrella of DSI, and making them the basis for services offered to the community.

Another part of this process is re-packaging Europeana's existing know-hows into a suite of services, reusable by other organisations. This requires overhaul of the entire stack because the parts which are farmed out to become independent services need to continue fueling Europeana's services today. In a way, this is about splitting a system working in production mode into two independent parts without interrupting the service.

Organisational restructuring is another aspect of the change. A possible outcome of Europeana Cloud will be a legal entity offering services to the members of the consortium, including Europeana itself. How this relationship will be governed to prevent conflict of interests will be addressed before the end of the project. Governance issues are being addressed in WP5 of the project.

Service Description

By the planned end of the project in January 2015 three core cloud services will be released. The services described below will use the cloud-based storage platform as the repository of data from which they retrieve input data and where they store output data. This storage platform is an open solution, accessible directly through a developer API. Internally, it uses popular open-source solutions OpenStack Swift and Apache Cassandra.

Metadata Transformation Service

A lion's share of various aggregation flows in the cultural heritage sector, and outside it, is about conversion from one metadata format to another. As volumes of aggregated data grow, complexity of conversion logic increases, and conversions are run more often, organisations struggle with building reliant workflows able to scale to accommodate the growth.

The soft belly of existing solutions, developed usually by academic institutions, less experienced in delivering commercial-grade software, is their inability to scale from datasets of hundreds to datasets of millions. Reliability and scalability is the unique service proposition of Europeana Cloud Services, and of this service in particular.

The service allows running XSLT transformations on XML datasets while using the computing power of the cloud. Its primary audience are developers and metadata specialists who need to run fast XSLT transformations on large volumes of data. The service will not provide an XSLT editor, like MINT[4], and will be targeted at users who create definitions of XSLT mapping

---

[4] http://mint.image.ece.ntua.gr/redmine/projects/mint/wiki/Mapping_Tool

themselves: with or without an editor. Users will need to provide the mapping file stored locally or online.

With many kinds of conversions possible, we chose the one which is probably the most popular one: conversion from one XML format to another using XSLT transformation standard. In the future we'll look at other conversion types and technologies as well.

Image Delivery Service

There is a strong need in the GLAM community today to allow re-use and ensure interoperability and accessibility of the world's image repositories in a transparent and open way. Inspired by this need, a group of institutions joined forces to specify a new open standard for image interoperability, [IIIF](). Several image software suites already support this format to different extents; some major cultural institutions are incorporating IIIF-compatible software into their stacks. Being at the focal point of sector's interoperability and standardisation efforts, Europeana obviously must join the game. By supporting IIIF in its own products, but more importantly, by offering an out-of-the-box service for other institutions to deliver images through IIIF-compatible software, Europeana will strengthen its standardising role in the sector, add expertise in content to its expertise with metadata, and allow thousands of institutions to enjoy this innovative technology without investing heavily into their own infrastructures.

The service allows content holders publishing high-resolution images using software compatible with the [IIIF]() standard. After uploading high-resolution images users will be able to view them in high-resolution in their own websites. To do this, they can either use the default IIIF viewer, supplied as part of the service, or develop a viewer independently.

EDM Search Service

To support metadata interoperability, Europeana is promoting the Europeana Data Model (EDM) format as a standard for the entire sector, including outside Europe. To help these efforts, Europeana should strive to support the standardisation effort by showcasing advantages of EDM, such as making it easy to develop applications around cultural data. Search is one of the core functions of applications dealing with data; however, efficient search across EDM records is quite a complex endeavour, as Europeana's own experience testifies. It is, thus, natural, that Europeana offers search across EDM as an out-of-the-box service for the community. It will lower adoption barriers as organisations will not need to develop their own one.

The service allows searching across EDM records and is based on Europeana's own search functionality. The full specification can be found [here](). The service will be backward compatible with Europeana's current search and add additional features. (Not to be confused with the

data-provider field in the EDM). A user will upload records to the storage platform in order to have them available for indexing, a prerequisite for search. This will then allow them to build their own search interface without the need for developing their own index.

How a Service Works?

There are several common principles in how user experience of Europeana Cloud Services is designed:

- Services get requests from users and issues responses in a standard way
- Users can track progress of services
- New data is stored by services and accessed by users in a standard way

We illustrate how a cloud service will work using the example of a generic enrichment service that enriches metadata. The service itself will be developed later on.
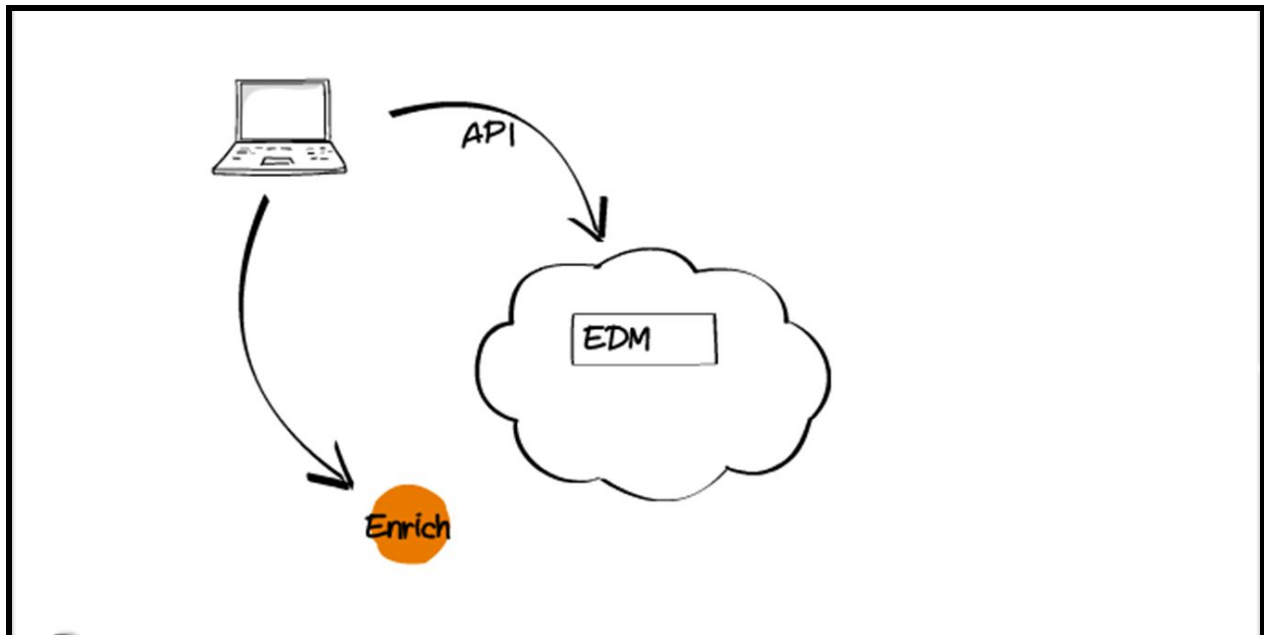
Stage 1

Using the Developer API, the user uploads a batch of EDM records to the storage platform of Europeana Cloud Services.
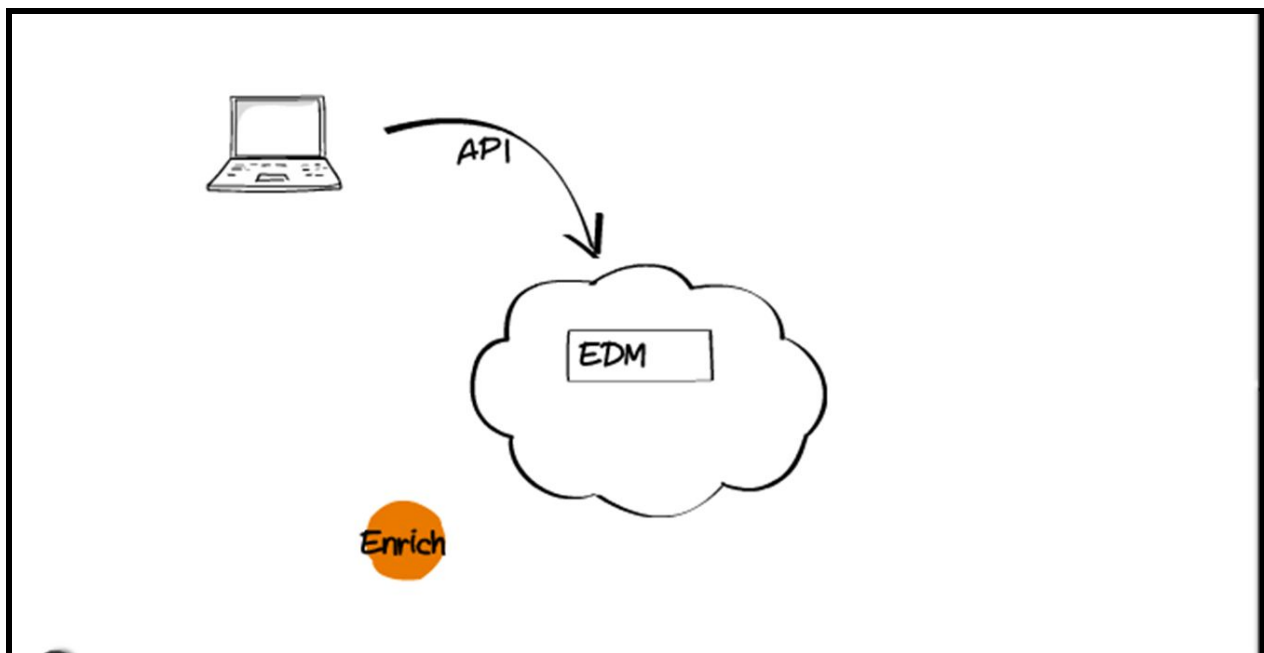

Stage 2

The user invokes the EDM Enrichment service on the uploaded records. The service starts running in the background.

Stage 3

The service finishes running and notifies the user about its completion. It adds semantically enriched fields to the original EDM records, uploaded by the user, and stores new versions of records next to the original ones. Both original and new versions can be accessed online by the user and other parties according to the access policy set by the user.
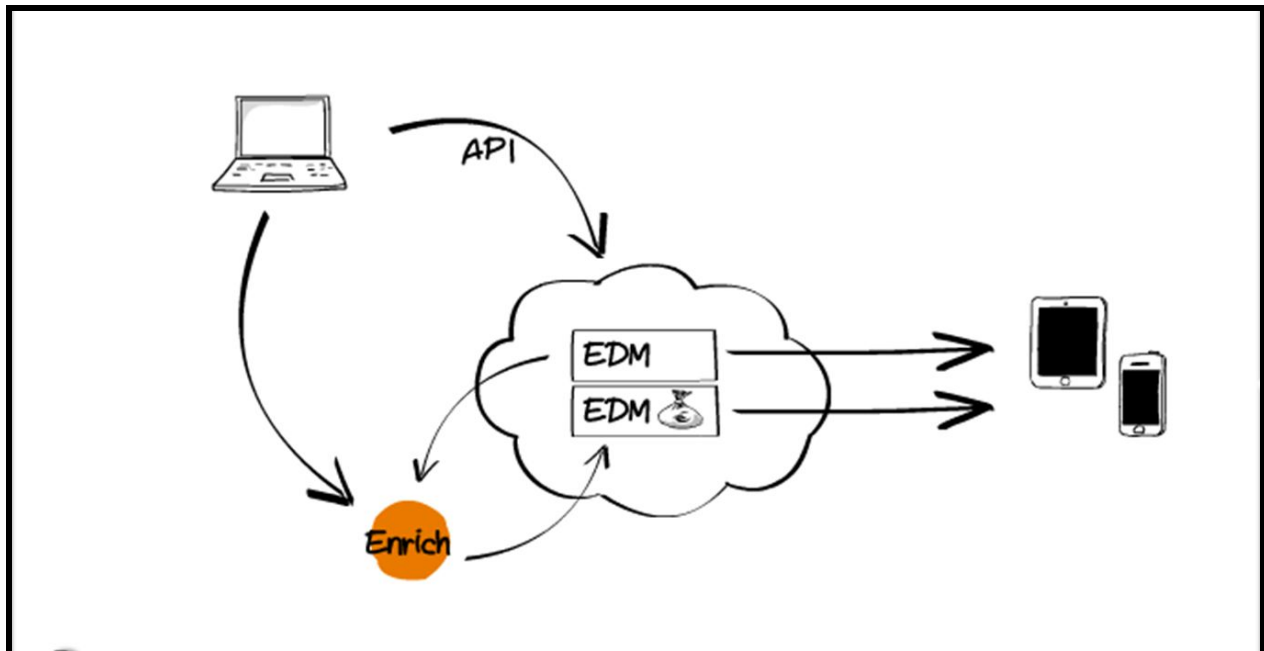
Each one of the Europeana Cloud Services is data-centric, that is it takes certain piece of data as input and usually creates another piece of data as output. An output can be a data record, as in the example above, or another kind of data that can be consumed by a client, such as, for example, a search index. Modeling services according to this principle resonates with the nature of aggregation processes which are often about deriving new pieces of data from incoming ones. Breaking down complex aggregation flows into a series of data-centric services, some of them are provided out of the box as Europeana Cloud Services, creates a standard user experience along the entire chain of aggregation logic.

Summary

The outcome of the Europeana Cloud project is a suite of robust and scalable metadata and content services operating on top of a cloud-based storage infrastructure. Such a service platform is a novel idea for the cultural heritage sector, yet it is long due. The project has helped the consortium led by Europeana build the technological foundation of the solution, identify market needs and develop product vision. Till the end of the project the consortium will release the first version of the product and test it on several internal use-cases; further development and maintenance will continue under the next round of funding of Europeana as a European Digital Service Infrastructure.