

Design of Europeana Cloud Technical Infrastructure

Pavel Kats
Europeana Foundation
The Hague, Netherlands
pavel.kats@europeana.eu

Petr Knoth
The Open University
Milton Keynes, United Kingdom
petr.knoth@open.ac.uk

Georgios Mamakis
Europeana Foundation
The Hague, Netherlands
Yorgos.Mamakis@kb.nl

Marcin Mielnicki
Pozna Supercomputing and
Networking Center
Pozna , Poland
marcinm@man.poznan.pl

Markus Muhr
The European Library
The Hague, Netherlands
markus.muhr@theeuropeanlib
ary.org

Marcin Werla
Pozna Supercomputing and
Networking Center
Pozna , Poland
mwerla@man.poznan.pl

ABSTRACT

In this paper, we present the overview of Europeana Cloud system, which is a new undertaking of Europeana Foundation and partnering institutions aimed to provide shared, cloud-based infrastructure for aggregation and exchange of cultural heritage metadata and content for European institutions.

Categories and Subject Descriptors

D.2.11 [Software Engineering] Software architectures – *Domain specific architectures*; H.2.1 [Database Management] Logical Design - *Data models*; H.3.7 [Information Storage and Retrieval] Digital Libraries – *Collection, Dissemination, System issues*

General Terms

Performance, Design, Reliability.

Keywords

digital libraries, metadata aggregation, Europeana, cloud architecture

1. INTRODUCTION

Existing ICT tools and infrastructures are not sufficient to serve the vision of European integration in the cultural domain. They are poorly orchestrated, they usually support only unidirectional flow of information, often employ different technologies and standards, and, last but not the least, are very costly. Additionally, they are mostly oriented towards operating metadata, leaving the need of providers for efficient content storage and access solutions unanswered. Also humanities scholars still face an immense amount of dispersed resources and resort to manual methods in order to reach to these resources and use them in research. Europeana Cloud is a new project funded by the European Union scoped to address the above issues. It is coordinated by Europeana Foundation and has a vision of creating new digital infrastructure for cultural content that will be used by

Europeana and other entities from all over Europe, interested in sharing or reusing digital representations of cultural resources. This infrastructure aim is to provide new abilities for efficiently storing metadata and content, easily sharing cultural assets between institutions, improving abilities to access these assets and research them using innovative tools.

2. THE PURPOSE OF EUROPEANA CLOUD

Europeana Cloud system is intended for entities which are interested in storing, distributing and re using cultural data: digital files representing cultural objects as well as their metadata. These entities include cultural heritage institutions, data and metadata aggregators, scholars and creative industry companies. The initial purpose of the system, as defined by user stories gathered from metadata aggregators, is the following:

- To provide globally unique identifiers for cultural data records from diverse sources
- To provide storage and access capabilities for cultural data records, consisting of data and metadata streams in many formats and versions
- To provide annotation capabilities for cultural data records
- To provide cultural data records changes tracking capability
- To provide flexible, scalable and customizable cultural data processing capabilities

All the above should be done in a secure, reliable and scalable way, allowing to use the Europeana Cloud system as the underlying infrastructure for cultural applications and information systems – the backbone of digital ecosystem for cultural data.

3. SYSTEM ARCHITECTURE OVERVIEW

Europeana Cloud system will be a service-oriented infrastructure, consisting of one or more instances of a number of network services. Each service will be responsible for providing a particular set of functionality, giving together the full set of desired Europeana Cloud features. The Europeana Cloud Architectural Design [1] document defines the following frontend functional services:

- Unique Identifier Service – provides the mechanism to create mappings between local identifier (scoped with

the data provider) and the global identifiers inside the Europeana Cloud system scope.

- Metadata and Content Service – provides the create/read/update/delete operations for cultural data records in multiple representations and versions (see data model details in the next section).
- Data Lookup Service – allows to query technical metadata of data records stored in Metadata and Content Service.
- Notification Service – provides the communication mechanism between the internal services and external clients for notifications about data changes in the Europeana Cloud system.
- Data Annotation Service – allows to store and access any additional information (annotation) related to any data record or its components (e.g. versions).
- Data Processing Service – offers the possibility to process data in a three stages workflow: extracting the data for processing from the Europeana Cloud system, processing the data, loading the processing outcomes to specified output location.

Beside the above functional services, the Europeana Cloud system will also include three backend system services: Authentication/Authorization Service, Logging Service and Asynchronous Messaging Service.

4. DATA MODEL OVERVIEW

Cultural heritage institutions deal mostly with physical cultural heritage objects, like books, paintings, sculptures etc. Such objects usually have their identifiers, which are unique within particular scope (e.g. inventory numbers, unique for particular institution) and are a basis for creation of information resources which somehow represent these objects in digital form. These information resources are usually:

- outcomes of digitization process, called “[digital] content”, representing more or less precisely how the physical object looks like;
- outcomes of cataloguing, curation or similar process, called “metadata”, describing the content of the physical object (descriptive metadata) or its technical or administrative aspects etc.

One cultural heritage object can have many metadata information resources and many content information resources – as it can be many times digitized or described and the information resources can be also processed from one form to another (metadata schema mapping, content format migration). Beside single metadata information resource can describe several physical objects at least in two scenarios: it contains information which is common for a group of physical objects; it is a complex metadata record which contains descriptions of several different physical objects which are somehow composed together. For the design of the Europeana Cloud system it was assumed, that the system will not try to provide specific means to express the above complexities, as it would require to base the entire system on one particular metadata model. Instead the aim of the system is to provide scalable and reliable storage for various information resources, which are managed and exchanged by cultural heritage institutions, without limiting the scope of such functionality only to content or to metadata nor to specific formats of content or schemas of

metadata. These resources are called data records. The assumption is that the designed system should be flexible and if needed, the limitations or specific guidelines should be introduced by particular data providers and aggregators, according to their needs. Therefore further it is assumed that each data record consists of one or more files (data streams with content and metadata in many formats) and these files are grouped in data record representations, which can have several versions over time. Data record, and its versioned representations are described with basic technical metadata necessary to manage these entities.

5. EUROPEANA CLOUD DEPLOYMENT

The Europeana Cloud system, which from the end user point of view will look like SaaS cloud, should be also deployed in cloud environment, in order to be reliable, available, scalable and cost effective. In order to achieve that, two types of underlying cloud are needed: storage cloud (distributed database and file system) and computational cloud (virtual machines to deploy Europeana Cloud system services). These two types of underlying cloud will be constructed with the hybrid cloud approach. They will consist of a private, community-based part where the necessary hardware resources will be provided by voluntary, technically advanced institutional users of the Europeana Cloud system and a public part, based on resources leased from commercial providers. Such approach should allow to provide cost-effective service independent from any commercial provider, yet scalable to commercial resources if needed.

6. SUMMARY

This paper presented a short overview of the initial design of Europeana Cloud system, which development was started in the beginning of 2013. The system is developed in a 6 month releases and the first production-ready version is expected to be available in the 2015. The system will initially serve as shared infrastructure for three metadata aggregators: Europeana, The European Library and Polish Digital Libraries Federation. In parallel to technical developments, which initial stage was presented above, there are ongoing works in the Europeana Cloud project consortium which are focused on two aspects very important in the context of successful deployment and technical sustainability of such share infrastructure. These aspects are business/financial sustainability model and collaboration with researchers interested in using new possibilities provided by Europeana Cloud.

The source code of the system is open and made available on-line at <https://github.com/europeana/Europeana-Cloud>. Further information regarding the project can be found at the project website: <http://pro.europeana.eu/web/europeana-cloud>.

7. ACKNOWLEDGMENTS

“Europeana Cloud: Unlocking Europe’s Research via The Cloud” is a project co-funded by the European Commission within the ICT Policy Support Programme.

8. REFERENCES

- [1] Werla, M., Mamakis, G., Muhr, M., Knoth, P., Kats, P., Mielnicki, M. 2013. Europeana Cloud Architectural Design. <http://pro.europeana.eu/documents/1414567/1861920/D2.2+Europeana+Cloud+Architectural+Design>