



DELIVERABLE

Project Acronym: Europeana Cloud

Grant Agreement number: 325091

Project Title: Europeana Cloud: Unlocking Europe's Research via The Cloud

D5.5 - Europeana Cloud Partner Roadmap

Authors: Victor-Jan Vos, Europeana Foundation
Alastair Dunning, Europeana Foundation
Harry Verwayen, Europeana Foundation

| | | |
|---|---|----------|
| Project co-funded by the European Commission within the ICT Policy Support Programme | | |
| Dissemination Level | | |
| P | Public | X |
| C | Confidential, only for members of the consortium and the Commission Services | |

Revision History

| Revision | Date | Author | Organisation | Description |
|----------|------------|-----------------------------------|--------------|--|
| 1 | 23-11-2016 | Alastair Dunning | EF | Initial version |
| 2 | 20-1-2016 | Victor-Jan Vos | EF | Revision of initial version |
| 3 | 19-4-2016 | Victor-Jan Vos; Harry Verwayen | EF | Revision, based on cloud business model |
| 4 | 22-4-2016 | Victor-Jan Vos | EF | Revision after review of Georg Eckes (DIF), Els Jacobs (Els Jacobs Advies) and Pavel Kats (EF) |
| 5 | 27-4-2016 | Victor-Jan Vos | EF | Final revision |

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This deliverable contains original unpublished work except where clearly indicated otherwise.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.

D5.5 - Europeana Cloud Partner Roadmap

Executive Summary

The roadmap outlined in this document identifies which partners will be the potential first users of the storage and data processing services. The roadmap analyses these partners' needs and explains how they may / will benefit from the Europeana Cloud outcomes.

As part of the outcomes of Europeana Cloud, several services were developed, which can be used by our partners. These services are known as Europeana Shared Services. To investigate partners' needs we interviewed several of Europeana's partners in a standardised format. The interviewed partners are all aggregators of Europeana, whether domain, thematic or national.

The partners indicated several items for improvement: 1) metadata mapping process; 2) ingestion and mapping tools; and 3) curation and data management.

The Europeana Cloud project has delivered solutions for all 3 issues.

This document describes in general lines which partners will use these tools and when. The planning includes an objective to develop a process for direct data delivery to Europeana (Operation Direct) from GLAMs, without interference of aggregators.

The research of partners indicates further that not storage should be a main focus of the Shared Services, but the focal point should lay on data processing and mapping tools. This has been delivered by the project and will continue to be developed.

Table of Contents

| | |
|--|----|
| Executive Summary | 2 |
| 1. Introduction | 4 |
| 2. Europeana Shared Services | 5 |
| 3. Selection and identification of partners..... | 6 |
| 3.1. Launching partners..... | 6 |
| 3.2. Domain and thematic aggregators..... | 6 |
| 3.3. National aggregators | 7 |
| 3.4. GLAMS and individual partners | 7 |
| 3.5. Selection and methodology..... | 8 |
| 4. Requirements analysis and partner needs | 9 |
| 5. Planning..... | 11 |
| 6. Conclusions and further development | 12 |
| Annex: overview of interview questionnaire..... | 13 |

1. Introduction

Europeana is the European platform for cultural heritage that streamlines sharing of cultural heritage content by data owners and delivers it to the end-user and re-user markets. It seeks to create force multipliers for the cultural sector by bringing together people and organisations who have heritage materials with people and businesses who want to view, use, and re-use them. The building blocks of this vision are a set of high quality digital data, a shared infrastructure that enables supply, enhancement and reuse of heritage content, and services that bring value for all stakeholders. In the Europeana Cloud Project a part of the shared infrastructure and shared services that operate on this infrastructure are developed.

The services are called Europeana Shared Services. These services are set up to the benefit of the Europeana ecosystem of data providers and aggregators. These services will allow partners to store both metadata and content, but also to perform operations on this data (such as data processing, data enrichment and data transformation).

The uptake of the services will be managed in the Europeana Digital Service Infrastructure (DSI) and with partners in the Europeana Network. The development and maintenance will be funded through European Union funding (Connecting Europe Facility - CEF) and other member-state based funding if applicable (e.g. for national organisations).

The first users of Europeana Shared Services are the three aggregator partners in the Europeana Cloud project, The European Library (TEL), the Poznan Supercomputing Center (PSNC) and Europeana Foundation (EF). They will use the Europeana Shared Services to store and access metadata and content.

In the Europeana Cloud Project, several other cultural heritage organisations were interviewed to investigate whether and how they would be using the services for data processing and transformation. This document reflects that research and provides a preliminary roadmap of partners that use the services. Overall, this document has three purposes:

1. It identifies possible future Europeana related aggregators (and other cultural heritage organisations) that will make use of the Europeana Shared Services.
2. It sets out how potential partners benefit from the Europeana Shared Services.
3. It describes and sets out an initial planning for the development and maintenance of Europeana Shared Services.

This document is closely linked to other outputs from Work Package 5, in particular the Europeana Cloud Business Model (D5.6.), of which an updated version will be delivered at the same time as this document.

2. Europeana Shared Services

The main objective of Europeana Cloud was to explore better ways for ingesting, storing, managing and sharing data through the Europeana Platform. In addition, other platforms for share and re-use of cultural heritage are a target of the project. The goal is to pilot for Europeana Foundation, TEL and PSNC a better means of aggregating and sharing data between the cultural heritage institutions of Europe and their potential users.

To that purpose, the Europeana Cloud Project has delivered in April 2016:

1. A storage service for metadata and digital media capable of scaling up to store massive amounts of data and to/from which data can be written and retrieved over an API. The storage service is based at the Poznan Supercomputing and Networking Centre (PSNC). The storage service includes fundamental internal services for its operation: *authentication*, *authorisation*, *identifier generation*, *data lookup*, *notifications*, and *logging*. These elements are essential components of the service to store and access data, but can also be used independently in other related services.
2. Generic data processing services that allow display, transformation and enrichment of large quantities of data. These include:
 - a. a generic IIIF Image Sharing Service, which transforms image files into JPEG2000 and makes them available in “zoomable” form;
 - b. a metadata transformation service that transforms metadata from one format into another;
 - c. an image transformation service to support the IIIF Image Sharing Service; and
 - d. a Search Service that supports the Europeana Data Model.
3. Europeana Research: a service assisting academic scholars to make use of the Europeana Platform of metadata and content for (digital humanities) research.

Additionally, the Europeana Research Service will be taken up by the Europeana Foundation as a separate product, with a separate advisory board and content strategy. Europeana Research is thus a part of the product portfolio in the Europeana DSI.

3. Selection and identification of partners

In the Europeana Cloud project Europeana Foundations fostered an excellent understanding of aggregator needs through interviews and qualitative research. Our analysis provides insight in how suppliers of data might wish to utilise the basic infrastructure and services for storage and data processing developed in the project. Europeana has interviewed potential partners (in both WP2 and WP5) with the aim to analyse each aggregators' requirements for use of the Europeana Shared Services. The interviews were conducted in a structured way. The set-up and questions of the interviews are attached as an annex to this document; the answers are available for the consortium, but, because of confidentiality reasons, are not publicly available. This market research was conducted between mid-2014 and mid-2015 through focussed interviews with potential data partners.

3.1. Launching partners

The Europeana Cloud project has three launching partners who will use the Europeana Shared Services, as specified in the project's description of work. They are the Europeana Foundation (EF), the European Library (TEL) and Poznan Supercomputing and Networking Center (PSNC). Their data has been transferred to storage of the Europeana Shared Services in April 2016, and the service is currently operational.

The European Library (TEL)

The European Library acts as the library domain aggregator for Europeana, providing access to data of national libraries and research libraries. It will use the Europeana Shared Services for storage of its data, media files and metadata.

Europeana Foundation (EF)

Europeana Foundation is the operator of the Europeana DSI and provides access to over 50 million records of cultural heritage data. These records have been stored on the Europeana Shared Services.

Poznan Supercomputing and Networking Center (PSNC)

Poznan operates the Polish Digital Library and is thus a national aggregator for Europeana. PSNC has a dual role in the project, as it was responsible for development and delivery of the services as well. It will continue this operation after the project ended, based on mutual agreement between Europeana Foundation (as operator of Europeana DSI) and PSNC. PSNC has migrated all its records to the Shared Services.

3.2. Domain and thematic aggregators

The following domain and thematic aggregators are considered to become partner to the Europeana Shared Services:

| Name | Purpose |
|------------------------|--|
| EUScreen | EUScreen is the data aggregator for television archives. |
| EFG | European Film Gateway is the aggregator for film archives. |
| Europeana Fashion | Europeana Fashion is the aggregator for fashion archives; it provides access to the data and files of fashion archives, both commercial and noncommercial. |
| Hope | HOPE is the aggregator for social history archives. |
| CARARE | CARARE provides access to data of archaeology. |
| Apex | Apex is the portal for European (national) archives and provides access to data of national archives to Europeana. It is a legal entity, Archives Portal Europe Foundation, support sustainability of the initial project that developed the portal. |
| OpenUp! | OpenUp! is the aggregator for natural history data. It is run by the Freie Universität Berlin. |
| MICHAEL/Museu | MICHAEL operates the aggregator for the museum domain. |
| Europeana Food & Drink | Europeana Food and Drink is a project aggregating food and drink heritage data. |
| Europeana Sounds | Europeana Sounds is a project that will create an aggregator in the Sounds-domain. |

3.3. National aggregators

There are several types of national aggregators: the types of national aggregators in the landscape differs considerably. For use of the Europeana Shared Services, a certain level of technical, operational development is required. Therefore, we considered the national aggregators for Italy (CulturalItalia), Spain (Hispana) and Germany (Deutsche Digitale Bibliothek) as candidates to include in this roadmap. In addition, under Europeana DSI-2, in efforts to enable direct data delivery to Europeana, national aggregators will be included as direct data providers.

3.4. GLAMS and individual partners

Europeana recognises the urgent need to modernise Europeana's outdated data supply approach and tools. Under the funding of *Europeana DSI-2-project* we plan to rebuild our current ingestion toolset (known as METIS). Europeana will research and prototype new and

different ways of publishing and sharing data via Europeana aimed not just at domain/thematic aggregators but also at individual GLAMs.

Under Europeana DSI-2, we will develop services that can serve as back-end services for new and shared tools and workflows that make it much easier and faster for aggregators and individual GLAMs to publish and share data via Europeana. As this was not the intention of Europeana Cloud, no interviews with individual GLAMs were conducted. As part of Europeana DSI-2, Europeana will start inventorying and identifying their needs from July 2016 onwards. All with the aim to make publication with Europeana much more easy than it currently is.

3.5. Selection and methodology

The domain and thematic aggregators in the Europeana ecosystem are the most logical partners to implement the Europeana Shared Services. These are long-time partners of Europeana and full partners in the Europeana Digital Service Infrastructure (DSI) as funded and non-funded partners. Europeana Shared Services is one of the main features of the Europeana DSI's services for data aggregating and data delivery partners (i.e. data partners).

Potential aggregator partners are divided into three distinct groups: domain aggregators that provide data from a specific cultural heritage domain (archives, libraries, etc.); thematic aggregators that provide data in certain themes or subjects and national aggregators that provide data from a country.

Not a specific target group in the Europeana Cloud Project, but nevertheless a very important group of stakeholders in the Europeana ecosystem are the individual cultural institutions (GLAMs). As they were outside of the scope of the Europeana Cloud project, they have not been interviewed under the project. Nevertheless, we expect the Europeana Shared Services to be beneficial for them as well.

The main topics of the interviews were hosting and storage needs, costs and the processes for data mapping, ingestion and delivery. In addition, we discussed performance, accessibility and availability of tools and the dissemination strategies for each of the aggregator partners. The questionnaire is added to this document at Annex 1.

4. Requirements analysis and partner needs

Based on interviews with 11 of the 13 aggregators¹ that the project has identified to use Europeana Shared Services, we will answer the question '**What are aggregators' current and future technical and strategic challenges**' below.

Metadata mapping process

The major challenge that interviewed aggregators identified is the process of mapping metadata from one format to another. To publish on Europeana, mapping (converting to) Europeana's metadata model EDM is required. Main issue with metadata mapping that it is a slow process with too many steps and discussions (e.g. between aggregators and technical partners, data partners and Europeana Foundation) involved.

Outcome of the interviews that support this notion:

- EFG does conceptual mapping of metadata, but uses a technical partner to do the actual mapping. This is a costly process, both in time and money.
- All aggregators indicate that publication of data could take between one and six months, which is unacceptable for today's user expectations.
- DDB uses different (types of) staff to map metadata from one format to another, both on a conceptual and technical level.
- EUScreen indicated that the workflow to map is very complex and could be improved.
- HOPE indicated that the workflow involves too many e-mail back and forth to iron out issues in metadata mapping.

Tools for ingestion and mapping

Another issue that we conclude from the interviews is that tools for ingestion and mapping are not as reliable as they could be. Aggregators worry that that data is not accessible enough and available for use. In addition, the usability of ingestion and mapping tools is too little, and need technical support from a technical service provider at a cost.

Outcome of the interviews that support this notion:

- TEL and PSNC both indicate that the data that they would like to use is not available all the time or the tools that they use do not seem reliable.
- Europeana Fashion, EFG both indicate that the interface and functionality of current tools (e.g. MINT) should be improved. Especially editing of records in bulk is a deficit in the current software.
- PSNC, Fashion, EUScreen and EFG argue that data partners (GLAMs) should work with data directly (mapping and ingestion), but current software does not support this.

¹ Two aggregators did not provide sufficient information for our analysis and were left out: Europeana Food&Drink and Europeana Sounds are EU-funded projects and not legal entities. Their main responsibility is therefore to find sustainability in the Europeana ecosystem first, before they could answer the question how they can use the Europeana Shared Services.

- Europeana Fashion, EUScreen and EFG specify that tools for ingestion are not always available; a cloud service would need to have 24/7 availability.

Curation / data management

Aggregators see the increasing need to curate (select, map, enrich, ingest) data with better mechanisms for assigning persistent identifiers and using authority files to enrich data. The usability and speed of current tools do not allow this, and decrease the efficiency and quality of data. Aggregators require more control of data in the cloud and see that the amount of data (and therefore the amount of data providers) processed by aggregators will increase significantly in the coming years and therefore the systems are not set up to support this.

Outcome of the interviews that support this notion:

- DDB, EUScreen, PSNC and EUScreen indicate that their data processing needs will increase in the coming years; mainly due to larger datasets and new data providers
- DDB indicate that (persistent) identification is a huge problem. Converting data to new systems require new identifiers; this does not support persistency of the identifiers.
- Fashion and EFG want to enrich data to increase the data quality by enriching the data, but the current tools do not support enrichment, not effectively enough.

In addition to these three main issues, two other subjects were discussed in the interviews:

Storage

Storage is not an urgent issue for any aggregator we interviewed. Most have the possibility to use different types of storage/hosting solutions (e.g. commercial providers such as Amazon or Dropbox) or do not store objects (metadata and media files) at all. Some aggregators make use of their data partners' storage. Also, it is common that aggregators receive storage facilities from host-institutions (e.g. AthenaRC for CARARE or The Netherlands Institute for Sound and Vision (NISV) for EUScreen). Costs for storage are therefore not a real issue with aggregators.

Outcome of the interviews that support this notion:

- TEL, EUScreen, Fashion, Hope DDB, EFG, Cultura Italia have own options for storage and use them without problem.
- All aggregators indicate that they are satisfied with the level of service of their hosting providers.
- DDB, EFG and CulturalItalia do not store objects themselves.

Dissemination

Interviewed aggregators use different strategies for disseminating data in- and outside of Europeana.

Outcome of the interviews that support this notice:

- All aggregators deliver data to Europeana and most see this as a major dissemination strategy.

- Aggregators ApeX, DDB, Fashion and TEL have different strategies for disseminating their data through APIs to other aggregators than Europeana.

From these interviews, we conclude that, contrary to initial assumptions of the project, the latter two issues should not be the main business proposition of Europeana Shared Services, The metadata mapping process, the tools for data transformation and (direct) data curation need much more attention in development of Europeana.

5. Planning

The first step of operation of the Europeana Cloud Services is an agreement between Europeana and PSNC describing both partners' responsibilities and requirements. This way, the Europeana Shared Services will be governed and sustained after the Europeana Cloud Project ends. Since The European Library is not a legal entity, its responsibility for the services will be governed through the Europeana Foundation and thus via this agreement. The agreement will be in the shape of a memorandum of understanding and will be in place by the end of the Europeana Cloud project (30 April 2016). Since PSNC is a full partner in the Europeana DSI, the roles and responsibilities will be carried out in the framework of the Europeana DSI-1 and Europeana DSI-2 projects as well.

As described above, new services for metadata mapping, data transformation will be operational for all aggregators (next to the initial three project partners) in the Europeana DSI before the end of 2016. Introduction of these will be gradual with concurrent user testing and evaluation with all aggregators in the Europeana ecosystem. Functionalities will be created and evaluated based on user demand and requirements. As part of the efforts to introduce direct data upload from GLAMs to Europeana ('Operation Direct'), a separate timeline will be created.

6. Conclusions and further development

Europeana has derived the following conclusions on further development and maintenance of the Europeana Shared Services (under Europeana DSI's projects):

1. The development of data partner services for aggregators will be focused on specific data processing services and position those as the main benefit and features of the Europeana Cloud, the focus will be less on shared storage service (however, this will stay a feature of the shared services, but as supportive service).
2. Europeana (and its partners in the Europeana DSI-2 project) will develop services that perform distinct tasks based on clear and identified customer demand. The IIIF Image Sharing Service is the example. It will first be used by Europeana Foundation and library partners (TEL) to serve their very large collection of hi-res newspaper scans (from Europeana Newspapers). It will then gradually be opened up to other aggregators, and individual GLAMs (as part of Operation Direct).
3. Europeana and PSNC will develop services that can serve as back-end services for new and shared tools and workflows that make it easier for aggregators and individual GLAMs to publish and share data via Europeana. The Metadata Transformation Service is one example. These services will be adopted by Europeana. Other aggregators and GLAMs are invited to become users/customers of a selection of them. The planning is based on user-needs.
4. Europeana will develop a new and shared tool and workflow (a client to the back-end services) to make it easier for aggregators to publish and share data via Europeana.
5. Europeana will research and prototype radically new and different ways of publishing and sharing data via Europeana aimed not at domain/thematic aggregators but at individual GLAMs.
6. The Improved Europeana API and Europeana Collections make it possible to search for and filter on the (technical) quality of digital media. In addition, Europeana Collections is already adapted to better support download of metadata and digital media.
7. Policies from Europeana Cloud (such as the Access and Re-user Framework and the governance model) are set up as flexible frameworks that can be adapted over the product's lifetime. These will support the further uptake of the Europeana Shared Services.

The development of products that address these requirements will be planned in the Europeana DSI-2 project (starting 1 July 2016), as part of the Development work package. The first milestone after the project will be the setting up of agreement between PSNC and EF on operation of the shared service. The services in operation from that moment will be the services described in chapter 2. The development of direct data delivery services to Europeana will start from July 2016 (with Europeana Foundation and Semantika as partners in Europeana DSI-2).

Annex: overview of interview questionnaire

| Question | Answer |
|---|--------|
| What software systems, tools and services are you using to fulfil your aggregation tasks? Please be specific and mention the purpose for which each one of these is used. When known, please also mention the version of the software and who developed it. | |
| Do you have available an architectural overview of your aggregation system? A description of the aggregation flow? Please share. | |
| Please describe also the process (both manual and automatic) of publishing to Europeana? Does it involve additional software? Do you also publish data to other super-aggregators besides Europeana? | |
| Do you have a web portal for users featuring your data? Please describe systems and technologies used for it. | |
| How long does it take on average from the moment a new dataset is made available by one of your content providers till this data is on your portal? On the Europeana portal? | |

Costs

- Efficient spending on IT; return on investment; costs and efficiency of IT services
- Satisfaction with the various parts of the IT

| Question | Answer |
|--|--------|
| Do you know how much you are paying for your IT infrastructure? Is it recurring payment or sunk (ie a one off) costs? Can you share with us the composition of your costs? (We promise not to share it with anyone and use only internally). | |
| Have you made any major decisions recently affecting your cost structure? The Long-term contracts? Could you share the details about it? | |
| Is your level of satisfaction the same with different components of the infrastructure? When not entirely happy, could you specify the reasons for this? | |

Hosting

- Hosting - taking advantages from development; general satisfaction with provider

| Question | Answer |
|-------------------------------------|--------|
| Who is hosting your infrastructure? | |

| | |
|--|--|
| Are you happy with the level of service you get from this provider? | |
| Are you happy with the price you are paying for these services? | |
| Does your hosting provider offer you anything of these: <ul style="list-style-type: none"> • Regular performance report and monitoring • Pay-as-you-go resource pricing schemes • Ability to easily scale-up/down your applications according to the need • Backup service | |
| Are there any kind of data interoperability scenarios which are currently not satisfied by your hosting infrastructure? That is, is there any use case of your data which you cannot implement now because limitations of your system? | |
| Are there any specific hosting-related functionality that you would like to have? | |
| Is your level of satisfaction the same with different the components of the infrastructure? When not entirely happy, could you specify the reasons for this? | |

Publication Process

- Speed of publishing process on own website; on other websites; human effort required

| Question | Answer |
|--|--------|
| Could you describe in detail the steps of the process to publish data on Europeana? (an overlap with a question above) | |
| Are you happy with how quickly you can publish new data on Europeana? | |
| What is average investment of human resources into the publication processes? What does it depend upon? | |

Overall Performance

- Scalability and performance; catering for redundancy
(there is some overlap with the questions above, please ignore or copy/paste in this case)

| Question | Answer |
|---|--------|
| Do you have periods of increased load on your system when the system does not cope with it? | |
| Do you have a backup strategy? Who is responsible for devising and running it? | |
| Have you ever suffered loss of data? Please specify details. | |
| Who is taking care of software development and architectural design of your system? Is it an in-house employee, a subcontractor or an external one? | |

Broader strategy

- More broadly, does your current IT infrastructure enable you to deliver your strategy ? Are there changes that need to be made in the long term to meet these challenges? Where are the problems that need to be altered?
- Do you have interested in storing or delivering content (as opposed to metadata). What is the type of content ? What are the demands on delivering it ?

| Question | Answer |
|---|--------|
| What is your long-term strategy about IT infrastructure and does it relate in any way to Europeana’s future plans around its aggregation network (aggregators as expert hubs etc.) | |
| Are you aggregating digital content as well or only metadata? If yes, please specify which kind of content, its volume and aggregation and storage technology. | |