DELIVERABLE

Project Acronym: Europeana Cloud

Grant Agreement Number: 325091

Project Title: Europeana Cloud:

Unlocking Europe's Research via The Cloud

5.6 UPDATED Europeana Cloud Business Model

Authors: **Victor-Jan Vos, Europeana**
Harry Verwayen, Europeana

Revision History

Revision	Date	Author	Description
0.1	19 DEC 14	Harry Verwayen Jill Cousins Julia Fallon Pavel Kates Alastair Dunning Marcin Werla	First draft for review
0.2	12 JAN 15	Georg Eckes (European Film Gateway)	Review
1.0	26 JAN 15	Harry Verwayen Julia Fallon Alastair Dunning	Edits to respond to review comments. Updated links in document.
Revision 2.0	20 April 2016	Victor-Jan Vos Harry Verwayen	Updated version to reflect outcomes of the Europeana Cloud Project
Final Revision 2.0	27 April 2016	Jill Cousins	Overall Europeana Strategy and needs of partners

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This deliverable contains original unpublished work except where clearly indicated otherwise.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

D5.6 UPDATED Europeana Cloud Business Model

Table of Contents

1. Introduction

This document shows how Europeana proposes to update the [Europeana Cloud business model](#) written at the end of January 2015. During the past year we have gained many insights into the perceived value of the services developed during the Europeana Cloud project. These learning points have been incorporated into our strategy and we have adjusted our activities accordingly.

This document highlights the main learning points and necessary pivots, and outlines the vision for the future development of Europeana's services[1] for both data partners and aggregators after the end of the Europeana Cloud project in May 2016. The previous business model is still valid, this document shows where pivots have been made and why.

---

[1] See also [this presentation](#) given at the Cloud final Conference Roundtable, April 2016

2. Europeana Cloud and its place in Europeana's strategy

2.1. Europeana strategy

Europeana's mission is to *transform the world with culture* and its strategy is based on the premise that Europeana needs to make it very easy and rewarding for cultural institutions to share their data through Europeana and more widely, on the web. The Europeana platform has three markets: end-user services, re-user services and data partner services.
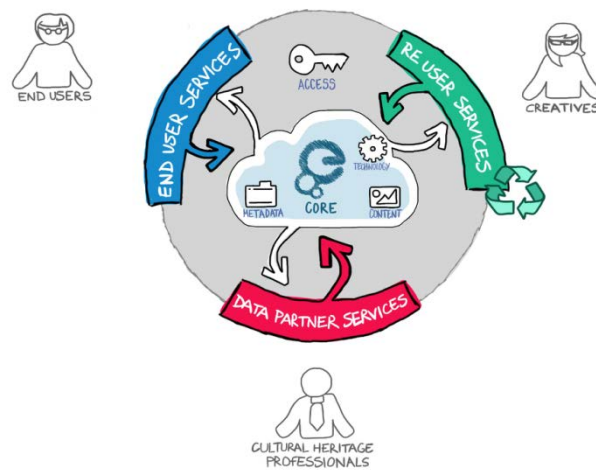

Fig 1: Europeana Platform

The Europeana Cloud project, begun in 2013, supports the work to improve the services for the data partners: make it easier to store, manage and share their data, using cloud technologies.
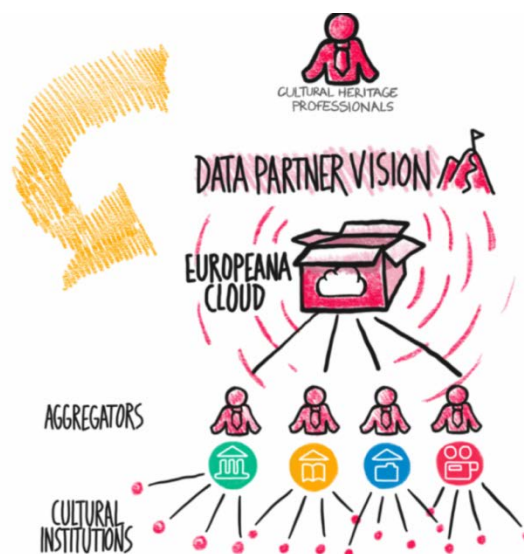

Fig 2: Europeana Cloud project

2.2. Europeana Cloud founding hypotheses and goals

In 2011-2012 when the project was conceived, we knew that Europeana, The European Library (TEL) and the Polish Digital Library (PSNC) needed to find better ways of ingesting, storing, managing and sharing data. We thought cloud technologies could provide some of the answers. The goal was to pilot a better means of aggregating and sharing data between the cultural heritage institutions of Europe and their potential users. The underlying assumption was that Europeana aggregators needed a much cheaper infrastructure capable of storing and retrieving both metadata and content and that this could be extended to Data Partners. A part-assumption was that the cost of technical operations in general and storage in particular was a major part of aggregators' Total Cost of Ownership (TCO). Another part-assumption was that infrastructure developed for aggregators would also be useful and valuable to researchers and research infrastructures. The assumptions were predicated on the existing model of aggregation that had been put into place by Europeana in 2008, encouraging both national and domain aggregation as the means of reaching individual cultural heritage institutions.

The project therefore set out to develop:

1. A storage service for metadata and digital media that would be capable of scaling up to massive amounts of data[2] and to/from which data could be written and retrieved over an API. This would include a generic data processing service to serve as the basis for the development of specific data processing services (see 2).

2. A small number of specific data processing services to demonstrate the capability of the generic data processing service.

In rough terms: 1. The storage service would form the Infrastructure as a Service layer (IaaS); and 2. The data processing service, functions as the beginning of a Software as a Service portfolio (SaaS). Both were not common terms for the cultural heritage sector in 2012.

The vision of the Europeana Cloud project was that this infrastructure could become the basis of a shared service, co-owned by the partners (i.e. aggregators) in the Europeana ecosystem. A part reason for this construct was that for political and financial reasons this should become a separate entity from the Europeana Foundation, run on 'commons' principles. Any profits from this entity would then be used to subsidise the development of new services in the entity, and perhaps even subsidise the Foundation.

---

[2] To begin with all of the metadata held by the three partner aggregators: Europeana, TEL, PSNC.

The project shared this vision in the form of an [animation] with the Europeana Cloud consortium early in 2015

2.3. Project outcomes

At a top-level the outcomes have been achieved:

1. A storage service for metadata and digital media capable of scaling up to store massive amounts of data[3] and to/from which data can be written and retrieved over an API. This is based at the Poznan Supercomputing Centre. Fundamental internal services to operate this storage service were developed: *authentication*, *authorisation*, *identifier generation*, *data lookup,notifications*, and *logging*. The data from all three aggregators, PSNC, The European Library and Europeana have been added to this cloud storage.

---
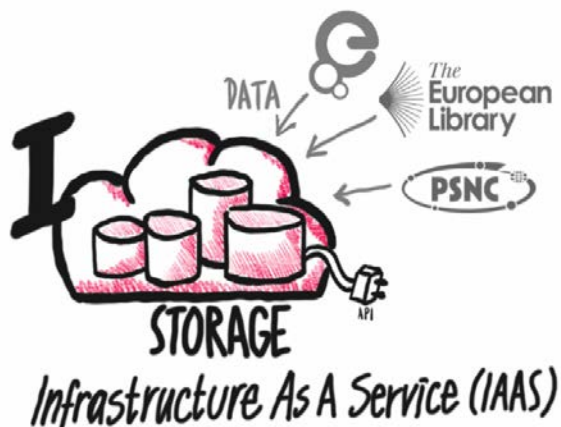
[3] To begin with all of the metadata held by the three partner aggregators: Europeana, TEL, PSNC.

2. Data processing services to demonstrate the capability of the generic data processing service, the most prominent being the IIIF Image Sharing Service (which transforms image files into JPEG2000 and makes them available over IIIF in "zoomable" form)[4].

3. Europeana Research: a service aimed at creating a shop window on Europeana data for for researchers and research infrastructures to access and use together with tools and services to manipulate the data.

In addition, the framework conditions for these services have been designed and documented: The legal framework for the access and re-use of the material (Cloud Access and Re-use Framework - D5.3), the governance framework and legal entity concept (D5.4), the partner roadmap (D5.5), and this business model (D5.6). A Handbook / User Guide explaining the practical use of the features and functionalities of the new cloud based system is also available for the core users of the infrastructure (D5.2).

However, we have since changed our views on some fundamental aspects of the services and its product positioning. These changes have been made in response to improved business intelligence (primarily improved user/customer and market insight, as well as changed technology conditions during the project timeframe) and on general changes and trends in the Cloud services industry.

---

[4] Others: Metadata Transformation, Image Transformation (primarily in support of the IIIF Image Service), EDM Search Service

2.4. What we learned during the project lifetime

1. Conceived in 2011-2012, the project's assumption that storage was a major pain point (in terms of money and time spent) for aggregators has been proven false. We know this from having interviewed our aggregator partners as well as from the general trend of ever reduced storage costs on the commercial market; the cost of storage has become cheaper and cheaper against what in big data terms is a relatively very small amount of data produced by cultural heritage organisations.

   Conclusion: **Cheap storage will not be positioned as the main outcome of Europeana Cloud.**

2. The project assumption that the cost of maintaining aggregation tools and services constitute a major part of aggregators overall budgets has proven to be a partial-truth. The total cost is circa 20% which is not considered onerous and any reduction is likely to be in single figures. However, an issue identified is that most of the aggregators are dependent on one technical provider (NTUA) and that the main tool used (MINT) is in the later stages of its product lifecycle with no replacement from NTUA in development.

   Conclusions: **Developing alternative solutions to MINT should be considered. Europeana Cloud IaaS should not be positioned as primarily being a cost-saver for aggregators.**

3. The process of aggregation is cumbersome, slow, work intensive and opaque and leads to friction and increased costs. The investment of work effort and total turnaround time from when a GLAM has decided to publish their data on Europeana to when it is published, is counted in months, when it should be counted in days or even hours. Equally aggregators and providers have to repeat similar processes to publish to other distributors of their data, such as Wikipedia or into education systems etc.

   Conclusion: **Mechanisms that remove layers of processing and make the publishing of cultural heritage data to the web simple are needed.**

4. Researchers with advanced (coding) skills in applying digital methodologies can and do use the current Europeana APIs and documentation on Europeana Labs.[5] Researchers who are not interested in applying digital methodologies, but are looking for cultural heritage material (from GLAMs) can use Europeana Collections.[6] For researchers with intermediate skills in applying digital methodologies there is a need for better functions to download metadata and digital media from the Europeana platform - both individual items and batches of items - and to annotate and curate metadata and digital media in Europeana. Europeana Research is a researcher driven shop window on the same data and same tools but targets primarily the audience of digital humanities academic scholars and serves up the data and tools in ways they need. Additionally researchers need to know what tools and services already exist that make data mining and manipulation easier and are largely poorly educated in this area.

    Conclusions: **Developing researcher specific APIs is not needed, but improvements to the existing content and technologies - Europeana API(s) and Europeana Data - are desirable to make them more useful to researchers. It should also be made easier for researchers to identify and then access/download the existing content that Europeana does have and is of the quality their research requires. The showcasing of tools and services (on Europeana Research) for the manipulation of data is useful for Digital Humanities Researchers.**

5. During the course of the project the political environment changed. The requirement for the DSIs to become fully self-sustainable has moderated. Creating a separate entity is therefore less urgent and may prove superfluous, particularly as the need to make money from the market has diminished in favour developing a model based on shared costs - which can be done through the Europeana Foundation. The project also discovered that the pain points were more to linked to providing simple publishing solutions directly to data partners, pointing towards the need to evolve a different business model on sustainability.

    Frameworks developed to start operating the originally conceived, independent service were developed and can be activated should partners in the Europeana ecosystem decide to set it up. An additional factor has been the effective disappearance of TEL (in 2015) as a proposed independent entity. The two remaining technology partners (Europeana Foundation and Poznan Supercomputing and Networking Centre) will create an MoU (or similar agreement) to ensure that the services that have been built can continue to deliver and develop.

---

[5] And research infrastructures (e.g. CLARIN and EUDAT) and services (EBSCO) who wish to mirror subsets of Europeana data can do so via those same Europeana APIs.

[6] Europeana's latest research on users of Europeana Collections show that a significant proportion of respondents identify as researchers or students; indicated that Europeana Collections serves a purpose for this target group.

Conclusions: splitting of a separate entity with the aim to generate income from the market is an unlikely option at this stage. Much more viable from an economic impact perspective is the development of shared services under the Europeana Foundation. As a business model the Foundation will provide infrastructure services under the central EU funding charging for any value added services. Efforts will be put into the development of a direct publishing service to Europeana (and the web), using the technologies developed under Europeana Cloud.

2.5. Pivots and re-prioritisation

Based on the conclusions above we re-focussed the project and made the following pivots:

- Increase the focus on developing specific data processing services and position those as the main benefit and features of the Europeana Cloud, decrease the focus on the storage service as a main feature of the service

- Develop services that perform distinct tasks based on clear and identified customer demand ((begun in Cloud, to be continued in Europeana DSI 1 and 2))
    - The IIIF Image Sharing Service is the example. It will first be used by Europeana and library partners to serve their very large collection of hi-res newspaper scans. It will then gradually be opened up to other GLAM users/customers.

- Shift strategy to make it simple and easy to deliver and extract cultural heritage data at institutional level, removing friction of process and giving control back to the institutions.
    - Development of better, frictionless data processing services for the short to medium term – Metis.
    - Piloting of systems that giving publishing control back to the data providers – Operation Direct.

These pivots have lead to the following three areas of focus for Europeana and a consequent modification of the Europeana Cloud Business Model from 15 months ago:

● 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 (begun in Cloud and Europeana DSI-1, to be continued in Europeana DSI-2)
  ○ The Metadata Transformation Service is one example developed within the Europeana Cloud project.
  ○ Within Europeana DSI-1 we have begun the development also of services for Validation[7], Preview[8] and Statistics.[9]
  ○ These services will be adopted by Europeana with other aggregators and GLAMs invited to become users/customers of a selection of them.

● 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.[10]

● 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 (begun at the tail-end of Europeana DSI-1, to be continued in Europeana DSI-2)

---

[7] Of any XML-data given a specific XML Schema (e.g. EDM)
[8] Of what the data would look like if/when published on Europeana Collections
[9] Statistics on number of records, fields present and to which proportion and other metrics to support QA
[10] This effort is often referred to as "Metis".

2.6. Europeana Research

Lastly, Europeana is now in a better position to deliver a service that is truly of use to Digital Humanities scholars and other researchers interested in cultural heritage data.

We will therefore:
● Develop the Europeana Research service and website further to make it easier for academic researchers and research infrastructures to use Europeana metadata and partner content; to connect to this vast resource of cultural heritage, using existing Europeana technology. We will direct its audience to use the Europeana APIs; encourage use of Europeana Collections; and provide them with data sets, tools and services and standardised research methodologies to allow them to find source and make use of cultural heritage data for the betterment and understanding of society and to connect Europe.

Europeana Research is live since February 2015. To improve the service we will:
● Improve the dedicated user interface.
● Improve the Europeana API.
● Encourage use of Europeana Collections via Europeana Research, explaining and exposing different aspects to Digital Humanities scholars. Make it possible to search for and filter on the technical quality of digital media (begun in Europeana Creative and continued in Europeana DSI-1).
● Develop a Europeana Annotations API and exploit that API in Europeana Research (begun in Europeana Sounds, to be continued in DSI-2).
● Develop Europeana Platform to better support download of metadata and digital media (begun in Europeana DSI-1, to be continued in Europeana DSI-2).

2.7. Europeana Shared Services (Lean Canvas)

The pivots of the Europeana Cloud project create an improved perspective for the roll-out of what we have come to think of as the **'Europeana Shared Services'.** The Europeana Cloud project has fostered an excellent understanding of aggregator needs, and how as suppliers of data to Europeana they might wish to utilise such an infrastructure. It has delivered some basic infrastructure and services. This has helped create a the roadmap for Europeana Cloud to advance once the project is completed, thereby serving a much wider body of stakeholders. But they have also led us to change focus from Aggregators to the Cultural Heritage Institutions themselves, partly because the sharing and storing technologies related to cloud services now permit this to be created.

Seen from a business model perspective the proposed experiments will mean a significant change for our partners. The precise implications have yet to be seen but we have sketched out our hypotheses in a lean (business model) canvas.

Problem · Insert

Slow
It takes months to publish

Opaque — Edit [x]
No control or direct feedback

Chaotic
Many intermediaries and different paths (direct, national, domain, thematic) to Europeana

Complex
Data Model, DEA, et.

Solution · Insert

Improve Metis
4 Europeana / TEL end DSI-2
4 Aggregating partners end of DSI-3

Europeana Direct
Radical experimentation

Develop valuable services
IIIF

Change aggregation organisation
Expert Hubs

Key Metrics · Insert

Statistics
Customers should have very clear insight into the quality evaluation of their data and the reach&use; of their data post-publication.

User surveys
Customer satisfaction will be measured via user surveys on a bi-annual basis

Unique Value Proposition · Insert

Improved data processing services for Aggregators
Moving to a singular shared infrastructure

Quick, direct publishing for GLAMS
The data would appear online (almost) instantly on Europeana, without mediators or complex data mapping

Transparent
Publishing process is open and controlled by partners

Direct
Data partners can publish data without mediators

Simple
No need for complex data mapping in order publish

Unfair Advantage · Insert

Europeana as the Platform for CH
Europeana already has the largest distributed network of Memory Institutions (3500+)

Channels · Insert

Via aggregators
Aggregators will focus much more on the relational aspects and less on maintaining a separate technical infrastructure

Direct
This is the main pivot of the model: GLAMS should have much more direct control over the publication of their data. This model aims at removing friction.

Customer Segments · Insert

GLAMS
All 3500+ current data partners and an estimated 60.000 cultural heritage institutions across Europe

Aggregators
Around 10 Domain aggregators and 30 national aggregators

Cost Structure · Insert

Metis development
4 dedicated FTEs (PO + 2 developers EF, 1 developer PSNC)

Operation Direct
3 dedicated FTEs (1 internal and 2 Semantika) until December 2016 (evaluation)

LoCloud Collections development
1 dedicated FTE from PSNC

Revenue Stream · Insert

Improved data quality
Partners will be more engaged and committed. As a result they will share higher quality data (tier 2/3/4) and metadata.

Lower cost overall structure
By developing shared services (domain) aggregators will be able to lower their individual costs of development

More visibility for the data
Higher quality will result in higher visibility

Empowerment
Partners will feel more empowered and intrinsically engaged (instead of 'publish and forget')

D5.6 UPDATED Europeana Cloud Business Model1

Following on from the above are some changes to our process and planning.

2.8. Process

One of the most important learning points has been that a waterfall approach to developing shared services is bound to lead to disappointing results, as technologies and user needs change too rapidly. We will therefore adopt more lean and agile methods following the lean startup principles with frequent iterative cycles of development (build, measure, learn).
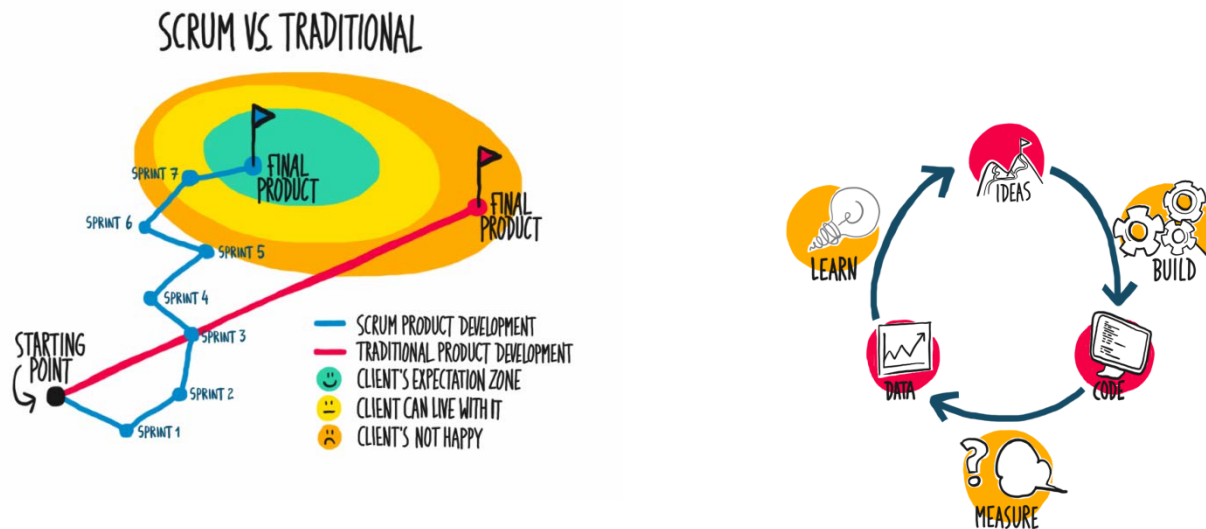
2.9. Planning

The outcomes of the Europeana Cloud project (Europeana Shared Services and Europeana Research) will continue to be operated by PSNC and Europeana Foundation. For the former this will be in the framework of an MoU (May 2016) under the Europeana DSI and Europeana DSI-2 projects (until at least September 2017). In the timeframe of DSI-2, Europeana will research the possibility for direct data delivery with GLAMs, while collaborating with domain aggregators to develop their positions as expert hubs in the Europeana ecosystem. In addition, the services and for ingestion (Metis) and display of cultural heritage material (IIIF and IxIF) will be standarised in cooperation with the IIIF-consortium.
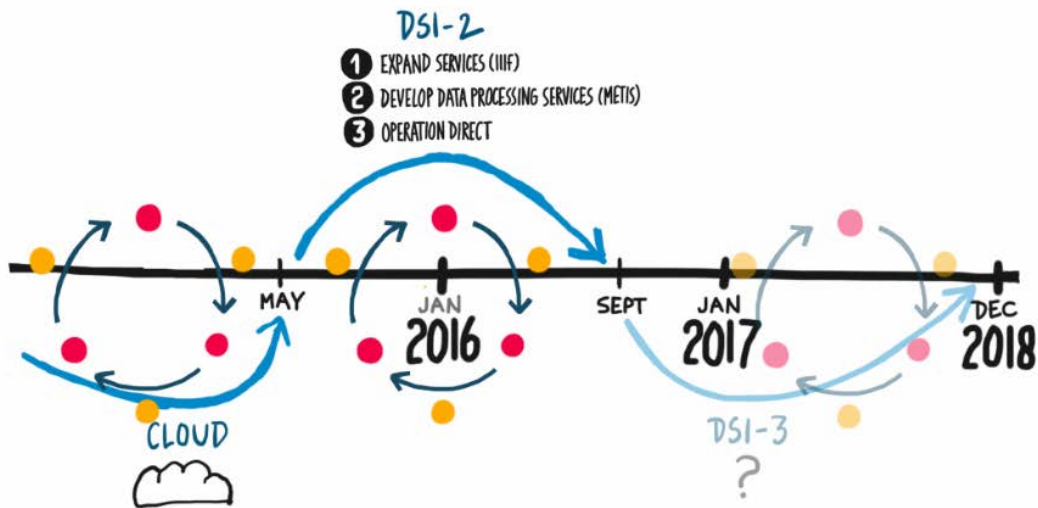
2.10. Governance

After the end of the Europeana Cloud project the coordination, maintenance and development of the infrastructure will be principally operated by the Europeana Foundation as part of its role as a Digital Service Infrastructure (DSI). The Europeana Digital Service Infrastructure belongs, via the Europeana Foundation, to all the cultural heritage organisations contributing to its data and its development.

To support the operation of services from mid-2016 and further, Europeana Foundation, Poznan Supercomputing and Networking Centre have established an in-principle agreement. In addition, PSNC and EF and several domain and thematic aggregators are partners in the Europeana DSI and Europeana DSI-2 projects.As the customer base grows, the governance structure of Europeana Cloud Services needs to be able to scale to include new partners. In Europeana Foundation governance structure we have the legal basis on which to build a cooperative governance of the service that belongs to all its contributors.

2.11. Finance and sustainability

As a European Digital Service Infrastructure, funded by the European Union for the foreseeable future, the running costs of the Europeana Platform and Services, using cloud technologies will be funded from the Europeana DSI (in subsequent projects) and will be provided to partners (Aggregators and individual GLAMs) under a shared additional cost model. The main aim of the DSI is to provide the best possible services at the lowest costs for the community.

It is the intention of Europeana that ongoing development makes sure that the costs for maintenance will stay the same for all stakeholders involved. Of course, division between partners can change over time.

The below table provides an estimate of the current costs of the Europeana Shared Services. They can be divided into development costs, which are budgeted in subsequent projects, and yearly maintenance costs (for data processing, and technical operation). The costs based on the budget of the Europeana DSI-2 project are as follows:

		FTE	Out-of-pocket Costs	Total (1 FTE = 66,000 euro per year)
Development (14 months: July 2016- August 2017)	IIIF METIS Operation Direct LoCloud Collections	1 FTE 4 FTE 3 FTE 1 FTE	20,000 euro 15,000 euro	9 FTE + 35,000 euro =728,000 euro
Maintenance (Per year)	Domain aggregators Europeana Foundation PSNC	12 FTE 4 FTE (data officers) 4 FTE (software maintenance and system administration) 1 FTE	35,000 euro 150,000 euro	21 FTE + 315,000 euro = 1,571,000 euro per year (or roughly 130,000 euro per month)