DELIVERABLE

Project Acronym: Ev3
Grant Agreement number: 620484
Project Title: Europeana Version 3

D1.1: RECOMMENDATIONS TO IMPROVE AGGREGATION INFRASTRUCTURE

Revision	1.0
Date of submission	3 March 2015
Author(s)	Henning Scholz, Europeana Foundation (EF)
Dissemination Level	Public

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

REVISION HISTORY AND STATEMENT OF ORIGINALITY

Revision History

Revision No.	Date	Author	Organisation	Description
0.1	09/02/2015	Henning Scholz	EF	First draft
0.2	11/02/2015	Joris Pekel	EF	Reviewers comments
0.3	12/02/2015	Henning Scholz	EF	Comments integrated
0.4	20/02/2015	Stephan Bartholmei, Andra Patterson	DNB & DDB, BL & Europeana Sounds	Reviewers comments
0.5	23/02/2015	Henning Scholz	EF	Comments integrated
0.6	25/02/2015	Harry Verwayen, Nienke van Schaverbeke	EF, TEL	Reviewers comments
0.7	25/02/2015	Henning Scholz, Ylva Klaassen	EF	Comments integrated, Annex added
0.8	02/03/2015	Imogen Greenhalgh	EF	Copy editing and comments
1.0	02/03/2015	Henning Scholz	EF	Comments integrated, layout edits, minor changes

Organisation acronyms:

BL – British Library
DDB – Deutsche Digitale Bibliothek (German Digital Library)
DNB – Deutsche Nationalbibliothek (German National Library)
EF – Europeana Foundation
TEL – The European Library

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This deliverable contains original unpublished work except where clearly indicated otherwise.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

Table of Contents

Background and purpose

Europeana is the network for the cultural heritage sector in Europe, and shares a vision of the world where every citizen has access to all cultural heritage.[1] To make this vision come true, Europeana aggregates a comprehensive, reliable and authoritative collection of Europe's cultural and scientific heritage. The Europeana ecosystem depends on a network of national, thematic and domain aggregators which bring together, manage and provide access to data about Europe's cultural heritage. The concept of aggregation partners has been the cornerstone of Europeana's business model from the very beginning; Europeana does not currently have the resources to ingest metadata directly from the large number of organisations that already supply, or wish to supply, metadata to Europeana.[2] The aggregator model that Europeana established made it possible to obtain metadata from thousands of cultural heritage and scientific institutions while directly ingesting metadata from fewer than 150 organisations.

Until now, the current aggregator model has served Europeana and its ecosystem very well. However, the model has weaknesses and issues that need to be addressed.

The aggregators who completed the most recent aggregator survey agreed that the current aggregator model is slow, intransparent and chaotic. It can also lead to duplication and data loss. The sustainability of the data flow in the current aggregator model is another important issue. The need to streamline and innovate the aggregation infrastructure as a whole is therefore a key objective for the future development of Europeana as a Digital Service Infrastructure. The aggregators' feedback shows that a future aggregator model should maintain the strengths and address the weaknesses. We need a faster turnaround time from data creation to data publication, a more transparent approach, clearer incentives and the right tools to provide the best data. We must develop all the above in partnership with the aggregators that are, and will continue to be key, to the success of Europeana.

This document gives an overview of the first steps to achieve this goal. It provides an initial exploration of issues with the current aggregator model and reviews some aspects of the aggregator landscape. This includes feedback received from aggregators during a workshop and during the most recent aggregator survey. Potential pathways to an improved aggregation infrastructure are explored and these will be used as the basis for discussions with the aggregators during a series of workshops and meetings in 2015.

[1] Europeana Strategy 2020 http://pro.europeana.eu/documents/858566/640ac847-0dfc-4b01-9f36-d98ca1212ec9

[2] Taking Germany as an example, 30,000 cultural and scientific institutions are aiming to link up to the German Digital Library (and may then also be accessible through Europeana, https://www.deutsche-digitale-bibliothek.de/content/help/general-ddb).

Definitions

- **Aggregator:** A service (provided by an organisation) that collects, formats and manages metadata from multiple data providers, providing services such as offering their own portal and acting as data provider to Europeana.[3]
- **Aggregator model:** The underlying concept for building the Europeana database, i.e. memory institutions partner with aggregators that harmonise and process their data and submit it to Europeana.
- **Aggregation landscape:** The relationships between memory institutions, aggregators and their partners including technical and service providers.
- **Aggregation infrastructure:** The wider relationships between the aggregation landscape and the tools, technologies and services that are involved to make aggregation and collaboration work.

[3] Taken from the Europeana Glossary: http://www.pro.europeana.eu/web/guest/glossary.

The aggregation landscape

The Europeana aggregator model

An aggregator is an organisation that collects, formats and manages metadata from multiple data providing partners, offering services such as supplying their own portal and acting as a data provider to Europeana. An aggregator may or may not have a public interface or front-end (e.g. portal, API). An aggregator without a public interface or front-end is called a ‘dark aggregator’.

The Europeana ecosystem includes three main types of aggregators:

- National aggregators, whose scope is defined by a specific country or region and whose contributors are situated within that country or region;
- Domain aggregators, whose scope is defined by a particular industry sector (such as museums, archives or libraries) and whose contributors may be located in more than one country;
- Thematic aggregators, whose scope is defined by a particular topic or theme (such as fashion or food and drink) and whose contributors may be located in more than one country.

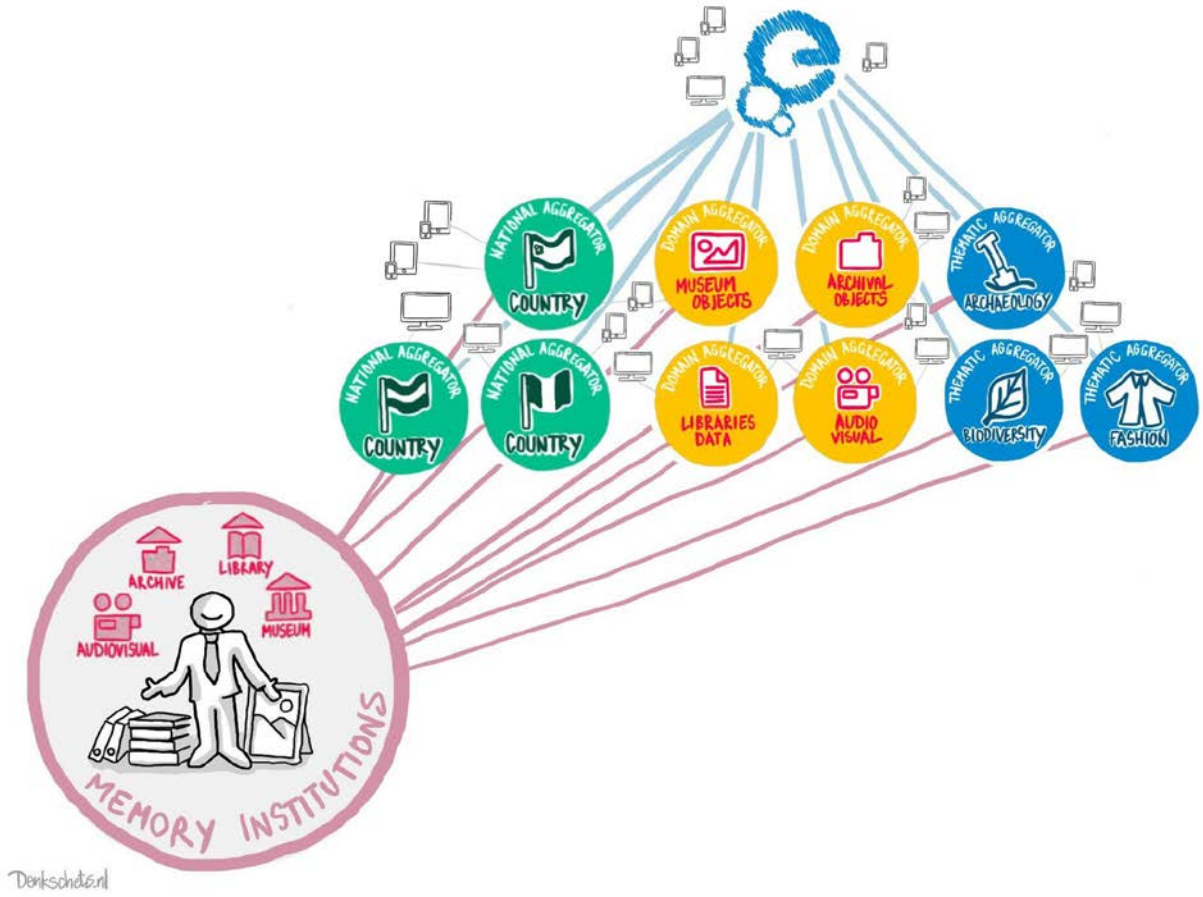

Fig. 1. Current (but simplified) aggregator model of Europeana: national aggregators (green), domain aggregators (yellow), thematic aggregators (blue). It is simplified in the sense that e.g. sub-aggregators are not included in this representation of the model.

Aggregators are key for the growth and success of Europeana. The concept of aggregation partners has been the cornerstone of Europeana's business model from the very beginning. In compliance with the aggregator model, a memory institution interested in providing data to Europeana would partner with a suitable aggregator (Fig. 1). Currently, the decision about the most suitable type of aggregator for a memory institution is based mainly on the scope of their digital collection, the expertise and technical infrastructures available to them, and their financial situation, geographical location and political commitments. In exceptional cases, a memory institution may be able to provide data directly to Europeana.

Data flow between partners

Following the aggregator model, data are flowing from a memory institution via an aggregator to Europeana. The aggregator cleans, harmonises and enriches data before submitting the data to Europeana. The data then becomes accessible to the end user via the Europeana Portal and Europeana API. However, a simple data flow such as this proves rare. A variety of different scenarios are possible for data to flow between partners across the aggregation landscape. The below is not designed to give a comprehensive analysis of all possible scenarios, but rather to provide examples of different data flow scenarios.

One common scenario is the existence of multiple aggregators between a memory institution and Europeana, and Europeana may not be the only public interface for the data. The institution may want, for instance, to reach the end user directly, via its own portal. In other cases it may need the help of another memory institution to set up data and technical infrastructures to join an aggregator. This aggregator may be a sub-aggregator and not directly connected to Europeana, but to another aggregator, which then acts as the direct provider to Europeana. It is also possible to reach the end user at other stages, for instance, if sub-aggregators or aggregators have their own portals and API as public interfaces to access the data (Fig. 2).

Given the choices facing an institution when providing data to Europeana, the data flow pattern can become complex. EU project funding is one way a memory institution can open up their digital collections and share them on the Web. Depending on the thematic scope of the project, only a subset of an institution's collections will become available through that project. The institution may then join another project for other parts of the collection or join a national aggregator to open up the remainder of the repository. Eventually, a memory institution may join several aggregators and provide data to Europeana via several routes (Fig. 3).

The clear advantage of the Europeana aggregator model (with the different types of aggregators: national, domain, thematic) and the availability of project funding is that it provides numerous opportunities for memory institutions to provide data to Europeana, which in turn opens up their valuable digital collections to a wider audience. It has worked well over previous years, and obtained millions of metadata records from thousands of cultural heritage and scientific institutions while directly ingesting metadata from fewer than150 organisations.

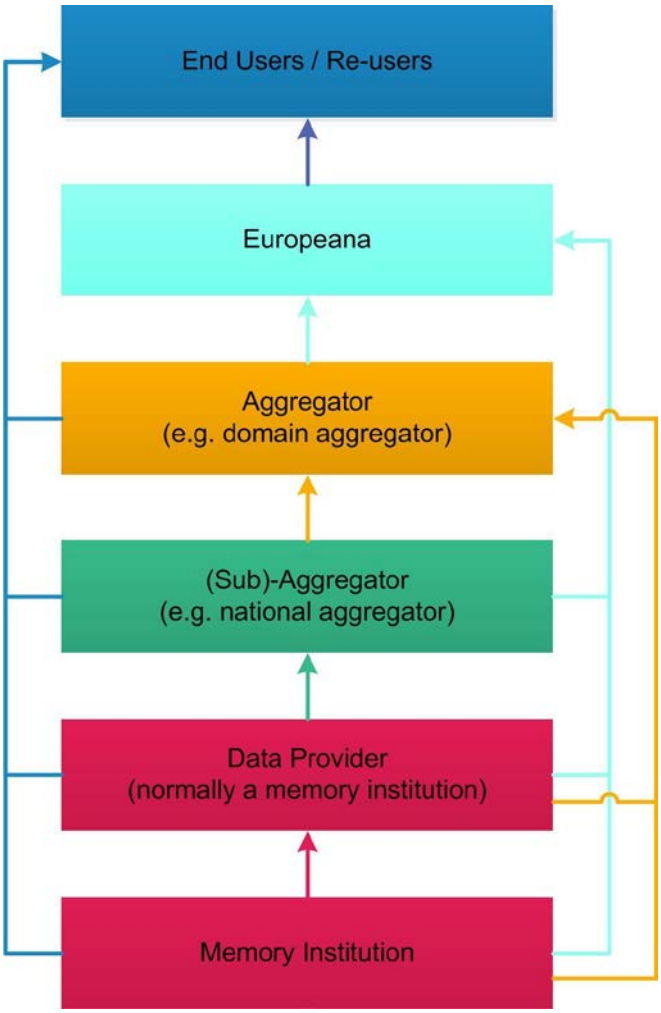

Fig. 2. Data flow from a memory institution to the end user using different aggregators and different levels of aggregation.

A significant weakness of the Europeana aggregator model is that it leads to duplication and data loss. Looking at the data flow illustrated in Fig. 3, it is obvious that, if collections are not clearly separated when prepared for different aggregators, the same data from one institution may end up in Europeana via different channels. Data loss can happen alongside data being mapped or cleaned up by aggregators. It can also happen if the use case changes along the aggregation chain. A curator in an institution. digitising and preparing the data. may have a different use case for this data than the person working with this data for an aggregator, to prepare it for Europeana.

The sustainability of the data flow is another important issue with the current aggregator model. Most of the domain and thematic aggregators were launched as EU funded projects. Projects usually last two to three years, with a clear start and end date. During a project, the data flows from data providers via projects acting as aggregators to Europeana. When the project finishes and no follow-up project continues the work, this data flow to Europeana stops. Therefore, considering options for the sustainability of project outcomes is important to ensure that the data flow between the memory institution and Europeana can continue beyond the end date of a project. Once a project ends and the technical infrastructure is no longer supported, new routes of data delivery need to be identified for the project partners. National aggregators are an

important means for sustaining the data delivery routes and absorb data and partners from projects which are ending. Europeana supports national aggregators because of their ability to coordinate a national infrastructure for digital information. Some domains have well established and sustainable aggregators too, providing suitable routes for sustaining data delivery routes for ending projects.

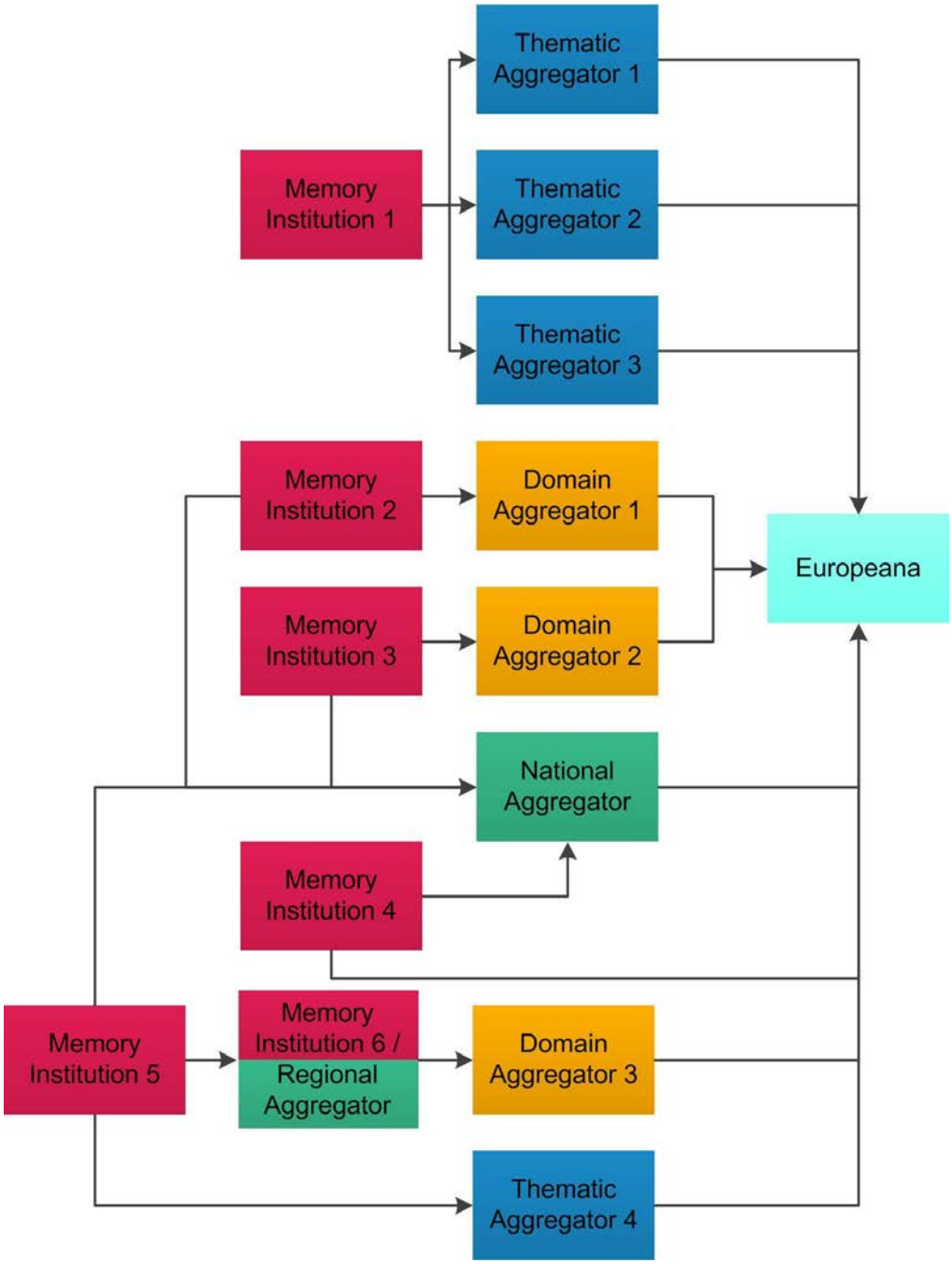

Fig. 3. Example of the data flow where different aggregators and different levels of aggregation are used, including different routes and multiple aggregation levels used by individual institutions (Memory Institution 5 in particular has quite a complex data flow and relationship with aggregators). This example is anonymised, but based on a real scenario.

Data providers and aggregators

For regular users of europeana.eu it is not currently easy to understand who provides data to Europeana or what type of organisations act as data providers to Europeana:

1. As explained in the previous section, memory institutions have several potential routes into Europeana depending on the type of data they deliver and the country they are based in. For a user, it is not necessarily obvious where to look for data from a specific institution.
2. Data provider information (organisation name) can currently be added to several metadata fields. Often this information is added to edm:provider and edm:dataProvider. In some cases this information is also added to dc:source, dcterms:provenance or dc:rights, particularly when sub-aggregators are involved before the data reaches Europeana. In these cases, the memory institution that provides the data is hidden and, sometimes, cannot be queried at all.
3. Organisation names are not normalised in all cases and duplicate entries or language versions exist based on the information Europeana has received from the data provider.
4. The current representation of data providers in Europeana also contains some outdated information, i.e. Europeana has now existed for over six years and some data providers have changed names or status over the years.

The above issues make it difficult, or even impossible, to give a consistent, clear overview of the aggregation landscape, or to show how many memory institutions or aggregators collaborate to provide data to Europeana. Therefore, neither the monthly content report nor the statistics provided via the Europeana Statistics Dashboard currently contain information about data providers and aggregators. To create this, an analysis of the current landscape was undertaken in December 2014, based on the aggregation landscape for this particular month. Europeana now features data from more than 3,300 memory institutions in Europe. This number is not precise due to the above complications, but it is still a good indicator of the range of memory institutions in Europeana.

Table 1. Overview of direct providers to Europeana. This is focused on the 113 unique direct providers that comply with the aggregator model (see text for further explanation). The numbers for each type of provider are the number of records available in Europeana in December 2014.

	Number of direct providers	Number of metadata records (as of Dec 2014)
National aggregators	34	12,528,959
Domain aggregators	4	9,441,012
EU funded projects (incl. ended projects)	33	13,097,791
Other aggregators	5	119,282
Memory institutions providing directly	37	935,722
Total	**113**	**36,122,766**

As of December 2014, europeana.eu features content from 149 direct providers. This means there are 149 unique names in edm:provider across the entire Europeana repository, which result when a user selects 'All providers' in the Europeana Portal.[4] This does not mean that all 149 direct providers represent unique entities providing directly to Europeana. Some duplicate entries exist. Some institutions that are listed as direct providers are, in fact, not direct providers but partners of an aggregator that is not named as a direct provider. This still needs to be cleaned up on a data level. Taking this into account for the analysis, there are 113 unique direct providers that comply with the Europeana aggregator model, i.e. aggregators or institutions providing directly to Europeana without other aggregation or facilitation between them and Europeana. The type and characterisation of these 113 unique direct providers to Europeana are shown in Table 1.

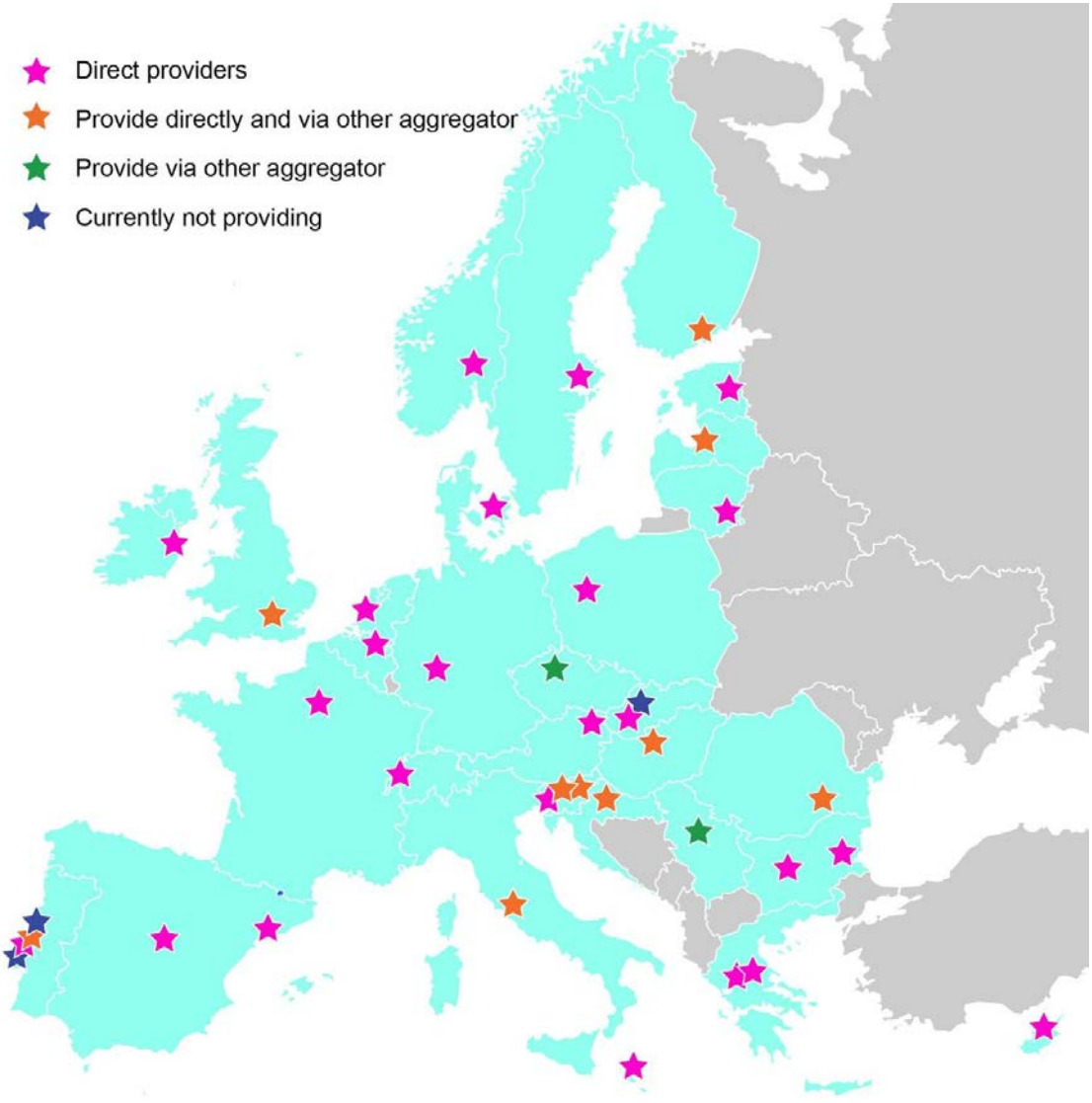

Fig. 4. National aggregators providing data to Europeana via various channels. This includes the 34 national aggregators (purple and orange) providing directly to Europeana that are also national aggregators in Table 1. The data of national aggregators that also provide under projects are included in the EU funded projects numbers in Table 1.

[4] http://www.europeana.eu/portal/europeana-providers.html

The vast majority of the 3,300 institutions that provide data to Europeana do so via a dedicated aggregator, as only 37 institutions are direct providers to Europeana (Fig. 4). In total, Europeana partners with 76 aggregators who act as direct data providers to Europeana. However, this does not mean that the aggregation landscape only includes 76 aggregators, as some aggregators are partnered with other aggregators to prepare data for Europeana. One example is the Bibliothèque nationale de France (BnF). The BnF runs Gallica, a digital library for the BnF, as well as a number of other institutions in France. The data available in Gallica is aggregated by The European Library for Europeana. As a result, it is important to recognise that The European Library is not only the largest domain aggregator to Europeana, but also the largest aggregator and data provider to Europeana. As of December 2014, The European Library provides more than eight million metadata records from national and research libraries across Europe.

The largest group of aggregators is the national aggregators group, particularly when acknowledging that the 33 projects include 16 completed projects which no longer provide data to Europeana. This was raised in the previous chapter, with regards to sustainability of the data flow. While project funding for EU funded projects usually lasts no longer than three years, national aggregators are set up by the EU member states with a far longer term perspective. National aggregators are, therefore, key to the success of the aggregator model. Fig. 4 gives an overview of the current distribution of national aggregators in Europe and their relationship with Europeana. It clearly shows that the vast majority of the EU member states have set up a national aggregator. Though not all are fully running at the moment, the progress is highly positive, and also highlighted by the number of data Europeana currently receives via national aggregators. This underlines why national aggregators are integral to the Europeana aggregation infrastructure.

Technical and service providers

The aggregation landscape described above is quite complex. Memory institutions partner with different types of aggregators, projects and Europeana while preparing to share their digital collections online. While many aggregators' are capable of all aspects of aggregation, it would not be possible to achieve data harmonisation and innovation for value-added services without the additional technical partners and other service providers.

A technical partner might offer support to one or more aggregators with technical solutions, software or infrastructures that are needed to harmonise, map, share or store data. One example of a technical partner supporting many aggregation projects is the National Technical University in Athens (NTUA) with the MINT platform. The MINT services provide a web-based platform designed and developed to facilitate aggregation initiatives for cultural heritage data in Europe.

All aggregators and even Europeana exist as service providers to help memory institutions make their digital collections available online, and to share them as widely as possible. However, the aggregation landscape also includes institutions that do not provide specific services such as mapping or publishing data. These institutions work in a different area of the aggregation landscape, and redesign aspects of the landscape or experiment with new value-added services. This might include new services for data enrichment and data roundtripping, which then provides enriched data for Europeana. This includes the innovation of API functionalities to enable both reading and writing data back into the database.

Aggregator Forum

A forum generally provides a physical or virtual space to ask and answer questions, exchange ideas and share knowledge. It is a place for people who share an interest in a topic to communicate with one another. In this sense, the Aggregator Forum provides a place for aggregators, their technical providers and any other service providers that are part of the aggregation landscape to communicate with each other. It is a place to discuss common issues and challenges and work towards resolutions. It aims to enable the transfer of knowledge, experience and advice between all aggregators, meaning more experienced aggregators can assist and support aggregators who are less established, or still working to fully establish themselves. The Aggregator Forum also aims to foster closer working relationships and two-way communication between organisations.

The participants in the Aggregator Forum meet at least twice a year, with one meeting taking place in spring and the second in autumn (Fig. 5). These meetings are organised and hosted by either the Europeana Office or an aggregator. Depending on the scope of a particular meeting, different participants in the Aggregator Forum may be invited to take part in the meeting. If an aggregator organises and hosts the meeting, the meeting may also focus on specific issues relating to the work of this particular aggregator.

Fig. 5. Participants of the Aggregator Forum in May 2014, which took place in The Hague (CC BY-SA Europeana).

It is important to emphasise that meetings are not only organised with larger groups of aggregators. Bilateral meetings, workshops or roundtables are also organised regularly with participants of the Aggregator Forum. The Europeana Office organises up to six workshops or roundtables per year with the Europeana Aggregation Team and representatives of other

aggregators, to identify ways to achieve joint goals. One example of these joint goals is the indicative targets for minimum content contribution to Europeana, set forth by the European Commission in the recommendation on the digitisation and online accessibility of cultural material and digital preservation (2011/711/EU).[5] The Europeana Office is working with national aggregators to achieve these goals by end of 2015.

The virtual component of the Aggregator Forum is a Basecamp project set up by the Europeana Office. This project is used to follow up on discussions and actions from the meetings, and to discuss common issues that were not discussed during the meetings. Anyone part of the intended audience described above and interested in joining the the Aggregator Forum Basecamp project can do so by emailing henning.scholz@europeana.eu.

Towards strategic planning for future aggregation

Aggregator workshop

A workshop was held on 3 September 2014, where representatives of seven domain and thematic aggregators and the Europeana Foundation worked on a vision and recommendations for the future of the Europeana aggregation landscape. As a first step, the services offered by the attending domain aggregators were reviewed. With these in mind, the four common themes for the future of aggregation were identified, and are summarised below.

Data

Harvesting, processing and submitting data to Europeana is the main task for aggregators. In the future, value-added services on top of high quality open data will become a higher priority. To achieve this, the richness of metadata needs to be preserved and data needs to be further enriched, contextualised and linked in meaningful ways. This requires lower barriers and higher incentives for institutions contributing data. The community and technological aspects of this are outlined below.

Communities

Building and maintaining well-defined communities were presented by all aggregators as key priorities for the aggregation service. These include expanding the network of contributing partners, building a knowledge- base among partners, enabling aggregators to benefit from the subject expertise in the community, and finding a working model for a community-based aggregator. Although every community shares common values, each aggregation community will differ depending on the domain the aggregator works in.

Technology

The development of new, shared tools and a shared technical infrastructure were both cited as key elements for the future of aggregation. The tools should reduce manual intervention by data providers and aggregators as much as possible, both for data processing and enrichment, and allow for more efficient data quality checks. A smoother, more seamless and semi-automated ingestion process with direct control of data (e.g. preview) would help to speed and scale up ingestion and aggregation. This would also help smaller institutions to contribute independently. The extension of services to end users (tagging, annotation, contextualisation) would help to

[5] http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2011:283:0039:0045:EN:PDF

reduce the workload for aggregators in improving data quality. Microservices for enrichment were mentioned as one way to improve data quality more efficiently.

Sustainability

Well defined communities and shared technologies should help to significantly reduce the costs for aggregation, and help make aggregation services sustainable. The OpenUp! project provides an example which demonstrates that this is achieveable for thematic aggregators. With excellent community support, in-house commitments from the contributing partners, a set of shared tools and services and a clear understanding of costs, a high quality aggregation service can be operated with a very low budget.

Aggregator survey

Europeana launched an aggregator survey to map the aggregation landscape for digital cultural content across Europe, both to provide an overview of the current situation and to inform future strategic planning to promote resilience, sustainability and innovation. This survey was open for four weeks between 8 January and 6 February 2015. It was completed by 33 aggregators, which together provide more than 80% of the data that is currently available in Europeana. The complete summary of the survey results is provided in the Annex to this document. The main outcomes with regards to the aggregation landscape are discussed below.

The aggregators who completed the survey agreed that the main weakness of the current aggregator model is that it is slow. It can take up to several months until data is available in Europeana. This is particularly a problem for aggregators with tight deadlines (e.g. EU funded projects). The aggregators also described the current model as being intransparent and chaotic. To quote one respondent, the “combination of thematic/domain/territorial aggregator models is confusing. It causes larger collections to be split up on their way to Europeana which is not very sustainable.”

The current aggregator model, with Europeana at the top, followed by aggregators, followed by data providers, indicates that, in the words of one respondent, the “aggregation workflow is designed around Europeana's needs, not as a service to aggregators”. This also creates the impression that data structures, guidelines or frameworks are developed centrally and not co-created by all partners for the benefits of all. A real partnership on equal terms between all players in the aggregation landscape and clear benefits for data providers are necessary to overcome these problems.

The main strength of the current aggregator model is that it has enabled the cultural sector to make millions of digital objects accessible via Europeana. The scalability for Europeana due to a shared workload with aggregators and the importance of domain or thematic expertise for memory institutions through aggregators were also acknowledged as strengths. A conclusion from the aggregator feedback is that a future aggregator model should maintain the strengths and address the weaknesses. We need a faster turnaround time from data creation to data publication, a more transparent approach, clearer incentives and the right tools to provide the best data. We need to develop all of the above in partnership with the aggregators that are and will be key to the success of Europeana.

The development of aggregation as an easy-to-use-service can help to give memory institutions a cost-effective way to distribute information about their collections. Issue trackers and automated data evaluation systems were mentioned as examples of tools that data providers and aggregators would appreciate being made available. It should give data providers more direct influence on the data representation in Europeana and also allow for a much shorter turnaround time before data get published in Europeana.

The aggregator survey described a vision for the future of aggregation: “Imagine all aggregators and Europeana are operating in a cloud-based environment using shared technologies and tools, where we are dealing less with the mechanics of aggregation but more with access to and enrichment of content. Every cultural heritage institution makes the data available in the cloud, specifies access rights and aggregators and/or Europeana process the data further, depending on their access rights (create once, publish everywhere).” For aggregators filling in the survey, this is an exciting but challenging scenario, which only the minority described as being desired. To quote one concern raised in the aggregator survey, “standardisation and data quality are huge requirements for improving the access, but the elements of the vision do not contribute to this. Instead they seem to suggest that everybody can participate on their own terms and make their own decisions.”

The values and benefits of the proposed cloud-based environment are also vague. Clear messages and good evidence about how cloud-based aggregation services could help memory institutions to achieve their goals will encourage them to support this vision. It is important for Europeana to develop and offer the services in close partnership and on equal terms with aggregators. Together, a clear value proposition has to be developed that proves cost-efficiency, ease-of-use, high data quality and standardisation. A governance model and clear terms of use are needed that support the value proposition and set clear conditions for data providers and aggregators to work on making cultural heritage objects more accessible.

Recommendations to improve the aggregation infrastructure

Improve the Europeana operational workflow

The tools and workflows currently in place to run Europeana operations are in constant need of improvement and optimisation. A particular area for improvement should be the communication processes between all levels of the existing aggregation landscape.

Analysis of the existing aggregation landscape

This document gives a few examples of the data flow in some areas of the aggregation landscape. More scenarios need to be analysed in order to get a broader understanding of the current landscape. This includes a better understanding of the current role of aggregators and service providers. It is also necessary to get a more clearcut overview of the strengths and weaknesses of the aggregator model. Having this overview should help to revise the aggregator model and restructure the aggregation landscape.

Develop a clear shared vision for the future of aggregation

The vision outlined in the last chapter to describe the future of aggregation has not been discussed with all the players in the aggregation landscape. It is thus not agreed what they all want aggregation to become over the next few years.

Investigate and develop the concept of expert and service hubs

Aggregators are currently dealing, to a large extent, with the mechanics of aggregation, e.g. harvesting, mapping, ingestion. If at some point in the future memory institutions can upload their data to a cloud infrastructure and use services working on top of this cloud infrastructure to prepare their data for Europeana, it is expected this will reduce the operational workload for aggregators and may change their role. Memory institutions will still need support by experts and trusted partners in order to create high quality data and publish them on the Web. Aggregators are expected to be these trusted partners, but possibly in different roles. Domain and thematic aggregators may be best positioned to develop into expert hubs. In developing communities and strong networks of domain and thematic experts, they can most effectively support memory institutions in curating and enriching their collections and applying best practices. National aggregators on the other hand may be best positioned to develop into service hubs to provide infrastructure and develop value-added services on top of the data. It is recommended that a series of workshops are set up in 2015 and 2016 to investigate, together with the aggregators (national, domain, thematic), if expert and service hubs are required (as separate entities, hybrid models) and, if so, what their roles would be.

Evolve the collection of Europeana's ingestion tools

Europeana currently uses the Unified Ingestion Manager (UIM) to orchestrate the ingestion workflow. This workflow needs to be standardised, become scalable, and layers of manual intervention need to be removed where possible. It is recommended that the current ingestion tools are developed into a set of EDM-centered, loosely coupled ingestion workflows, operating on top of a common data storage infrastructure. The new set of workflows will be collectively referred to as Europeana Metis.[6]

[6] Metis (μῆτις) meant wisdom, craft, skill in Ancient Greek. Metis may also refer in Greek mythology to the Titan goddess of good counsel, advise, planning, cunning, craftiness and wisdom.

Europeana Cloud as the central infrastructure for data storage

The development of a cloud infrastructure and all its business components is currently underway in the Europeana Cloud project (see Europeana Cloud Business Model and Product & Services Requirements for implementing Europeana Cloud Services).[7] This will be ready to use by the end of 2015 for the three pilot partners in the Europeana Cloud project (Europeana Foundation, The European Library, Poznan Supercomputing Centre). It is foreseen that more partners will be attracted to start using the cloud from 2016 onwards. This process will go through some stages of expansion to test the infrastructure and prove maturity, usability and scalability before the entire aggregation infrastructure comes to rely on the cloud as its backbone to store data (metadata and content) and run services on top of the cloud.

Roadmap for the innovation of the aggregation infrastructure

A roadmap and implementation plan needs to be developed that incorporates all recommendations highlighted above. A stepwise process is foreseen, although some steps can run in parallel without interfering with the further development of the aggregation infrastructure. The improvements of the operational workflow, for example, can run independantly to the analysis and restructuring of the aggregation landscape, and inform the development of the new ingestion tools. Also, the development, prototyping and testing of the Europeana Cloud can run in parallel to everything else, while benefitting from the outcomes of the other tasks. The roadmap is predicted to be delivered in spring 2016, incorporating the outcomes of the series of aggregator workshops which will take place over the next 12 months.

[7] http://pro.europeana.eu/files/Europeana_Professional/Projects/Project_list/Europeana_Cloud/Deliverables/D5.6%20Europeana%20Cloud%20Business%20Model.pdf & http://pro.europeana.eu/files/Europeana_Professional/Projects/Project_list/Europeana_Cloud/Deliverables/D5.7%20Product%20and%20Service%20Requirements.pdf

Annex: Europeana Aggregator Survey 2015

Survey designed by: Henning Scholz, Joris Pekel, Alastair Dunning, Julia Fallon
Survey sent out: 8 January 2015, Survey closed: 6 February 2015
Analysis ready: 27 February 2015 (Ylva Klaassen)

Introduction

This Aggregator Survey 2015 is the third survey conducted by Europeana among its aggregators, the previous two being held in 2009 and 2010-2011. The aim of this survey was to map the aggregation landscape for digital cultural content across Europe, both to provide an overview of the current situation and to inform future strategic planning to promote resilience, sustainability and innovation.

The survey consisted of 50 questions, divided into 8 parts (for the full list of questions, see the chapter at the end of this report):

1. General information
2. Partner relationships
3. IPR, content strategy and data quality
4. Technical infrastructure
5. Relationship with Europeana
6. Funding
7. Aggregator model
8. Additional comments

Most of the questions were multiple choice questions, often allowing for multiple answers simultaneously. In addition, many questions offered space for comments or elaboration, or were entirely open-ended.

The survey was sent out on 8 January 2015 via the Aggregators' Forum Basecamp and later in January to the Europeana Network via the Network newsletter too. The survey itself was open for four weeks between 8 January and 6 February 2015. It was completed by 33 aggregators, which together provide more than 80% of the data that is currently available in Europeana.[8] Fifteen of these also participated in the previous Aggregator Survey (#2, 2010-2011).[9]

This report analyses the results, summarising and contextualising the responses, focussing in particular on the information most relevant to external readers. When, in exceptional cases, the answers contained sensitive information, these are not disclosed herein.

General information

Overview of respondents

An overview of respondents is listed in Table 1. It specifies the name, location and portal URL of each aggregator, as well as the level of aggregation (the geographic coverage of their data providers: regional, national, European, worldwide, or otherwise) and type of aggregation (cross domain, single domain, thematic, or otherwise).[10]

[8] When the reports speaks about 'the aggregators', it refers to those aggregators who filled out the survey. When it refers to 'aggregators', it means all Europeana's aggregators in general.
[9] http://pro.europeana.eu/publications/aggregators-survey-2
[10] Cross Domain: including galleries, libraries, archives, museums, etc.; Single Domain: either galleries, libraries, archives or museums, etc.; Thematic: e.g. fashion, natural history, archaeology, etc.

Table 1. Respondents to the aggregator survey. The location for some projects is not applicable (N/A).

Name of the Aggregator	Location	Portal URL	Aggregation level	Aggregation type
AcrossLimits	MT	http://expo.acrosslimits.com/	National	Cross Domain
Archives Portal Europe (APEx)	NL	http://www.archivesportaleurope.net/	European	Single Domain (archives plus archival holdings of libraries, museums and related cultural heritage institutions)
AthenaPlus project	N/A	http://www.athenaplus.eu/	European	Cross Domain
Bulgariana	BG	http://bulgarianheritage.bulgariana.eu	National	Cross Domain
CARARE	N/A	http://www.carare.eu/	European	Other (Archaeology and Architecture - cross domain)
CULTURAITALIA	IT	http://www.culturaitalia.it /opencms/index.jsp?language=it	National	Cross Domain
Culture Grid	UK	http://www.culturegrid.org.uk/	National	Cross Domain
Cyprus Ministry of Education and Culture (CMOEC)	CY	http://www.cmoec.org.cy:8081/repox/	National	Cross Domain
Danish National Aggregator	DK	(none)	National	Cross Domain
Deutsche Digitale Bibliothek (DDB)	DE	http://www.deutsche-digitale-bibliothek.de/	National	Cross Domain
Digitale Collectie	NL	http://digitalecollectie.nl/	Other (national (domain, thematic) and regional)	Cross Domain
DISMARC	DE	http://www.dismarc.eu/	Worldwide	Other (Music and audio related material from all domains)
European Film Gateway (EFG)	DE	http://www.europeanfilmgateway.eu/	European	Single Domain (Audiovisual Archives)
Erfgoedplus.be	BE	http://www.erfgoedplus.be/	Regional	Cross Domain
eSbirky	CZ	http://www.esbirky.cz/	National	Cross Domain
Europeana Fashion	N/A	http://www.europeanafashion.eu	European	Thematic (fashion)
Europeana Sounds	N/A	(none)	European	Cross Domain
EuropeanaLocal Austria	AU	http://www.europeana-local.at	Regional	Cross Domain
EUscreenXL	N/A	http://euscreen.eu/	European	Thematic (public broadcasting)
Hellenic Aggregator for Europeana	GR	http://aggregator.libver.gr/	National	Cross Domain
HISPANA	ES	http://hispana.mcu.es/	National	Cross Domain
Hungarian National Digital Archive and Film Institute (MaNDA)	HU	http://mandadb.hu/tart/kereses/	National	Cross Domain
Judaica Europeana	FR/UK	http://www.judaica-europeana.eu/	Worldwide	Other (Jewish history and culture, cross domain)
Kamra	SI	http://www.kamra.si/	National	Cross Domain
National Heritage Institute	RO	(none)	National	Cross Domain
Norvegiana / Arts Council Norway	NO	http://norvegiana.no/	National	Cross Domain
OpenUp!		(none)	European	Thematic (Natural History)
Public Library - Varna	BG	(none)	Regional	Cross Domain
Registo Nacional de Objectos Digitais (RNOD)	PT	http://rnod.bnportugal.pt/	National	Single Domain (Libraries)
Slovenian national e-content aggregator	SI	http://www.agregator.si/	National	Cross Domain
Social History Aggregator	N/A	http://www.socialhistoryportal.org/	Worldwide	Other (cross domain and thematic (social history))
Swedish Open Cultural Heritage (SOCH) / K-samsök	SE	http://www.ksamsok.se/in-english/	National	Cross Domain
The European Library (TEL)	NL	http://www.theeuropeanlibrary.org/tel4/	European	Single Domain (Libraries)

From this list, it is clear that the largest group of respondents (52%) aggregates on a national level, often because they are the national aggregator in their country. This number has been relatively stable over the years: in the 2010-2011 survey, 51% of the respondents operated nationally, while it was 60% in 2009.[11] Most of the aggregators with a national scope work cross domain: they aggregate content from different types of institutions.

The second largest group operates on a European scale. These are often projects funded by the European Commission. As these projects often have a specific theme, many of them aggregate thematically, i.e. cross domain but with a particular focus, such as archaeology, fashion, public broadcasting, natural history or Jewish history and culture. There are also some single domain European-wide aggregators, which focus on archival material (Archives Portal Europe – APEx), film (European Film Gateway – EFG), or books (The European Library – TEL).

Two small groups work on the highest and the lowest geographic level: 9% of the respondents operate regionally, 9% worldwide. One respondent works on different levels simultaneously. All of the regional aggregators work cross domain, while the worldwide ones have a thematic focus.

When compared to the results of the previous survey in 2010-2011, there is a trend towards a broader kind of aggregation: the proportion of cross domain aggregators has increased from 51% in 2010-2011 to 67% in 2015, while the numbers of aggregators working thematically or within a single domain have decreased from 16% to 12% and from 33% to 12% respectively.[12]

Domains represented

The respondents aggregate across a broad range of domains, some of which are better represented than others. Archives (a domain included in 82% of the respondents' aggregated content), libraries (76%) and museums (67%) are the best-represented domains (Fig. 1).

Audiovisual archives (55%) and sound archives (39%) are the next most substantial groups. In light of Europeana's aim to include more audiovisual content in its repository, it is promising that so many aggregators are already aggregating AV-material. About one quarter of respondents (24%) aggregate content from Galleries.

Performing Arts (15%) and Publishers (9%) are less common domain types. The large category 'Other' (42%) represents a wide range of material and sources: heritage collections from research institutes, the educational sector, public cultural institutions, local heritage associations and NGOs (including societies); private collections; archival holdings of cultural heritage institutions; national databases for archaeological sites, monuments and historical buildings; photo and heritage agencies. The fact that almost half of all respondents aggregate material from providers which cannot be covered by the traditional set of cultural heritage domains indicates to the broad scope of the field, and suggests that our usual division might be too restrictive.

[11] See *Analysis of the Europeana and Athena Survey for Aggregators #2*, 6-8 (http://pro.europeana.eu/publications/aggregators-survey-2).
[12] See *Analysis of the Europeana and Athena Survey for Aggregators #2*, 6-8 (http://pro.europeana.eu/publications/aggregators-survey-2)

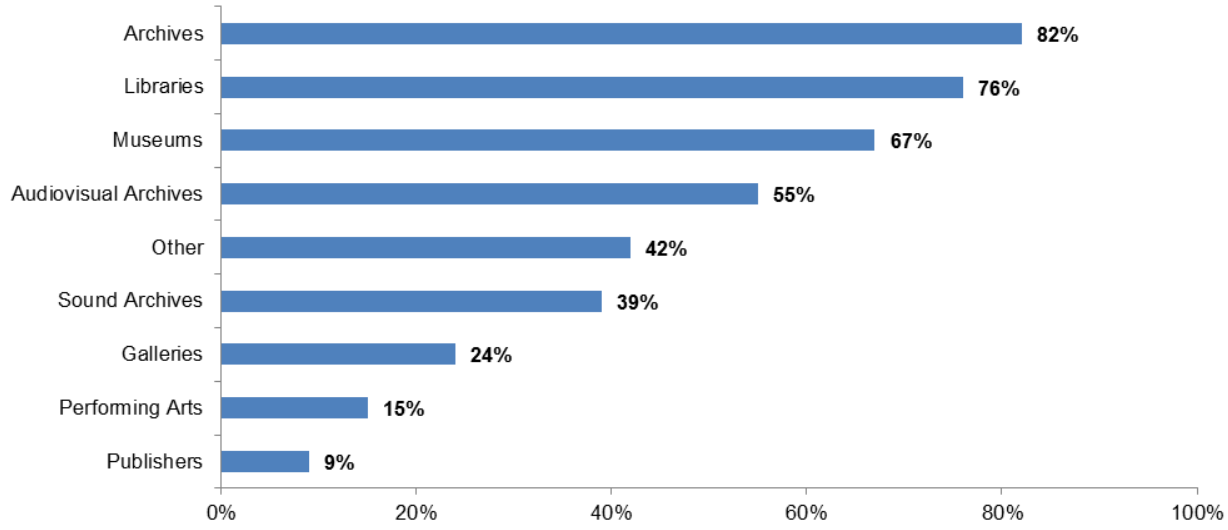

Fig. 1. Chart of the domains included in the aggregators.

Partner relationships

Aggregators' network of partners

Generally, the respondents are medium-sized: 61% of them aggregate content from 10 to 100 institutions, and 24% for more than 100 institutions. Only 15% have less than 10 contributors. Most of the respondents have a list of partner institutions on their website.

To increase their network of contributing partners, most aggregators rely on a combination of natural growth (being approached by interested parties; 70% of the respondents) and active and specific sourcing for new partners (79%). The fact that more than two thirds of the aggregators are approached by interested content providers means that they are well-known players in their field. However, only 36% of the respondents claim to have information about the number of cultural or scientific institutions in their country and in Europe which are potential data providers. General calls for participation (24%) and recruitment via Europeana and the Europeana Network (21%) are regular, but less common, methods for aggregators to extend their network of partners.

About half of the respondents (55%) are also partners themselves to other aggregators or projects with the aim of preparing and submitting data to Europeana. Such a partnership can take the form of a national aggregator contributing to a project such as AthenaPlus or LoCloud, or domain, regional and thematic aggregators submitting data to a national aggregator. This is not a new development and was reported back in the results of the previous survey.[13]

[13] See *Analysis of the Europeana and Athena Survey for Aggregators #2*, 9 (http://pro.europeana.eu/publications/aggregators-survey-2).

Communication with partners and data providers

The most common way for aggregators to communicate with their partners and data providers, for all but one respondent (97%), is personal e-mail. Phone or skype calls (79%) are also frequent, complemented by personal meetings at visits, workshops and conferences (30%). Basecamp, which is often used by Europeana and which is also used to set up the Aggregator Forum, is employed by 15% of the respondents only (Fig. 2).

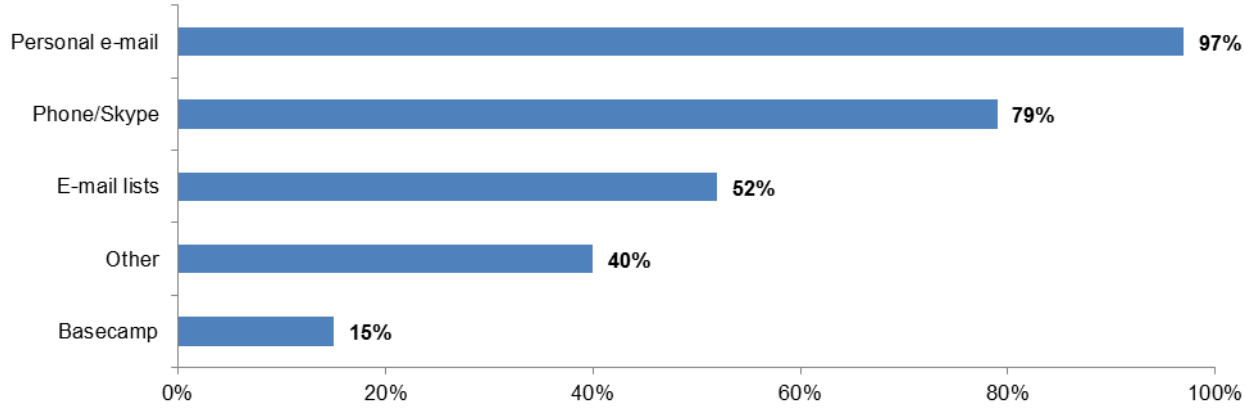

Fig. 2. Means of communication between the aggregators and their data providers and partners.

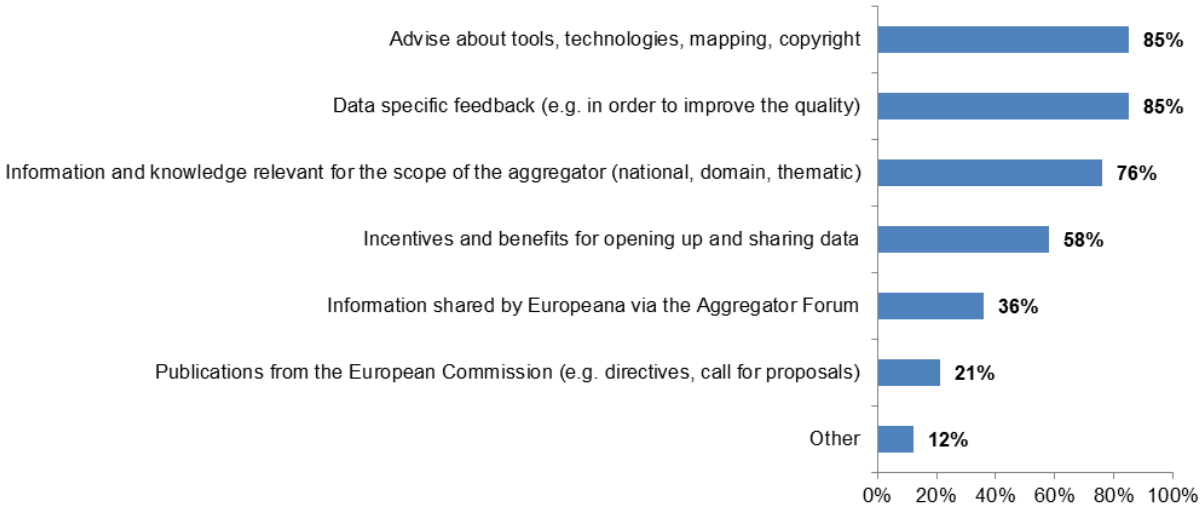

Fig. 3. Content of the aggregators' communication with their partners.

The content of the communication is mainly technical/operational (i.e. related to the actual aggregation of data). Information relating to strategy, politics and funding are topics which are communicated by a smaller number of the aggregators (Fig. 3).

85% report providing data-specific feedback to their partners in order to improve the quality of their data, and the same proportion of respondents also give advice about tools, technologies, mapping and/or copyright. 76% communicate information relevant to the scope of the aggregator (national, domain, thematic).

Just over half of all respondents (58%) communicate about incentives and benefits for opening up and sharing data. Information disseminated by Europeana (via the Aggregator Forum) or by the European Commission (e.g. directives, calls for proposals) is distributed by 36% and 21% only.

IPR, content strategy and data quality

IPR documentation

More than half (61%) of the respondents provide written documentation (guidelines and other kinds of information) relating to IPR (Intellectual Propery Rights) to their partners and data providers. The majority of these have made this documentation available on their portal, project page or institutional website, while some refer their data providers to publications on Europeana Pro. In addition, 46% organise training, and 14% give individual advice, while some are in the process of preparing these. Almost all aggregators, apart from the 5 who skipped the question, provide some form of documentation, training, guidance or advice.

Content and collection development strategy

Just over half of the respondents (55%) have a content strategy, or other policy or guideline documents. This is an increase from 2010-2011, where 44% reported having a content strategy.[14] In more than half of these cases (11 out of 18 respondents), these documents are publicly available on the aggregator, institutional or project website.

The focus of the collection development, however, varies among the respondents (Fig. 4). Five of the six possible answers are ticked by around one third or more of the respondents: curated collections (42%), outcomes of mass digitisation programmes (36%), enriched data (33%), other (33% – mainly material with a particular thematic focus),[15] and no focus (30%).

If we make a division between more specific strategies (aiming for a certain kind of content: curated collections, enriched data, masterpieces and/or thematic material), and a less specific approach (not seeking specific content, but taking data from mass digitisation programmes and/or not having a particular focus) the following pattern is discernible. Around one third of the respondents take a specific approach; one third a non-specific approach; and the last third have a mixed tactic. It is not possible to map these results directly onto the level of aggregation: there is no clear difference between thematic, single and cross domain aggregators.

[14] See *Analysis of the Europeana and Athena Survey for Aggregators #2*, 12 (http://pro.europeana.eu/publications/aggregators-survey-2).
[15] Based on the comments. 'Other' included also: more linked data (1), high quality (1), collections coming from new providers (1) and might change after the continuation of the project (1).

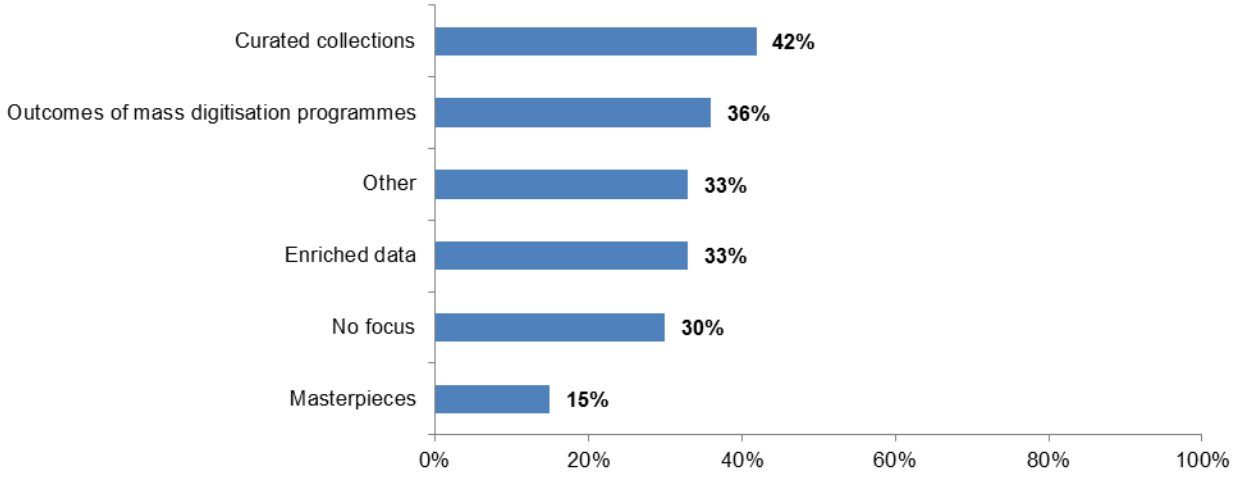

Fig. 4. Focus of the aggregators' collection development policy.

Almost two thirds (60%) of the aggregators have an ingestion plan, in which they schedule the amount of data to be processed. This, too, presents a slight increase compared to the 2011 survey, where 51% of the respondents indicated having an ingestion plan.[16] Most of the 2015 respondents did, nevertheless, answer the question of how many digital objects they plan to deliver to Europeana in 2015 (Fig. 5). The majority plans to make a substantial contribution this year: between 10,000 and over one million objects. Only 16% expects to be able to submit less. This means that Europeana can expect quite an increase in the size of its repository: based on these answers, a very conservative count – taking the lowest number in each category – suggests that a number of almost 10 million new objects can be expected.

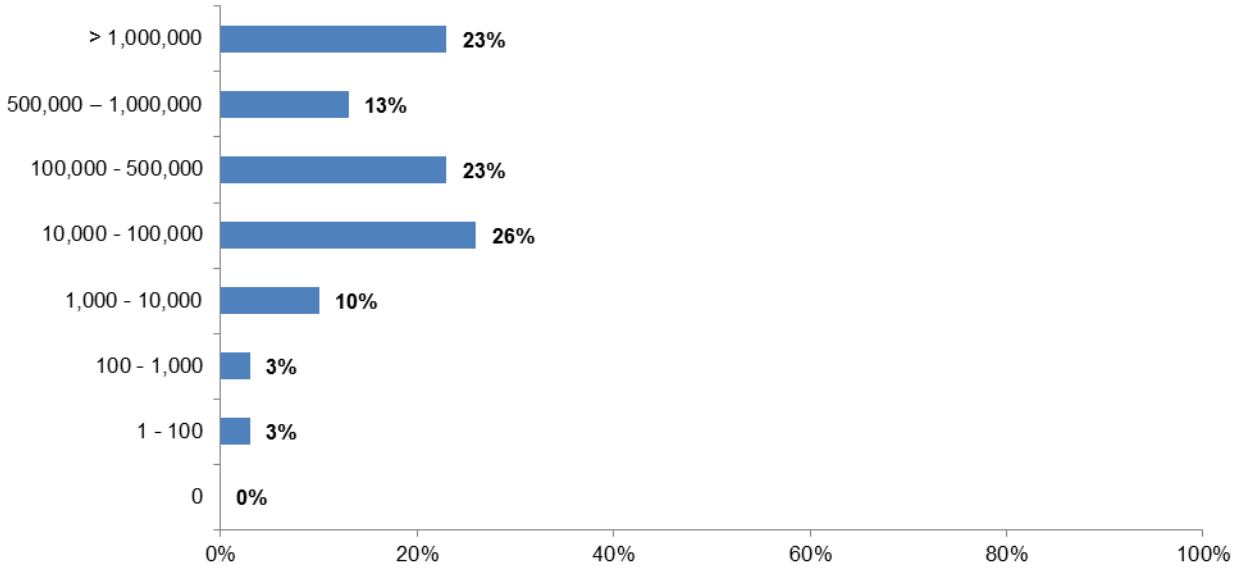

Fig. 5. Number of digital objects the aggregators plan to deliver to Europeana in 2015.

[16] See *Analysis of the Europeana and Athena Survey for Aggregators #2*, 12 (http://pro.europeana.eu/publications/aggregators-survey-2).

Direct links and persistent identifiers

Direct links to digital objects (not to viewers or websites) delivered with edm:isShownBy are an important mechanism to improve the user experience as it only takes one click from the Europeana portal to access the object. It also makes access to digital objects via the API much easier. For that reason, Europena is promoting the use of direct links.

The majority of the aggregators (82%) encourages and helps their partners to provide direct links to digital objects. However, they encounter difficulties in the process (Fig. 6). These are mainly related to institutional policies (58%) and technical issues (54%). Resource issues, such as additional workload, are encountered by over one third (38%) of the respondents, while only 17% report not being aware of the importance of direct links. Some of the answers submitted under 'other' were copyright issues, the problem that direct links are not appropriate for all kinds of content (e.g. 3D objects, which usually need a specialised viewer), and the fact that viewing an object through a direct link rather than on the portal of the data provider separates the object from its full context (metadata).

Despite the difficulties encountered, however, the survey suggests that most of the respondents are willing to push for direct links. If we provide more encouragement and tailored support to solve the most common problems, we may expect a positive development for the near future.

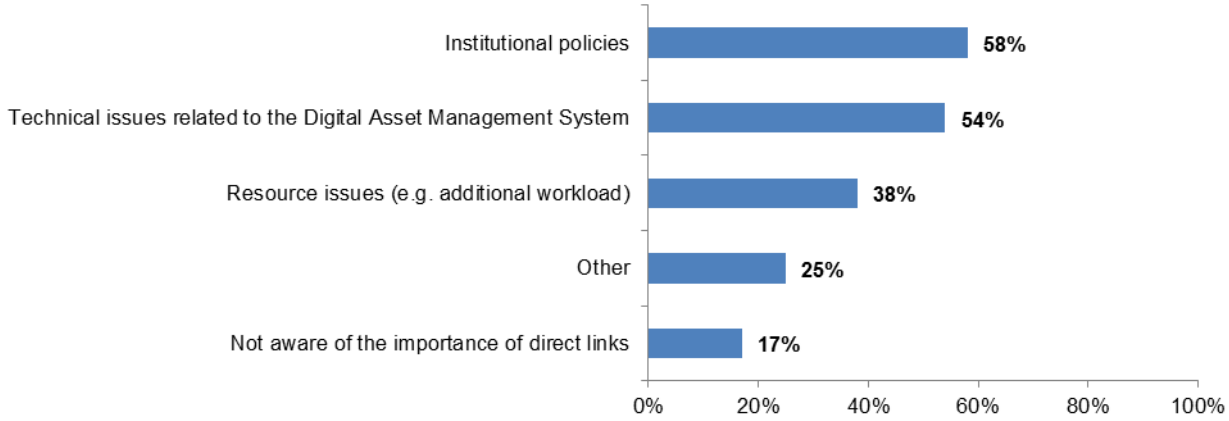

Fig. 6. Main difficulties experienced by the aggregators in getting direct links.

Two thirds of the respondents (67%) reported using persistent identifiers (PIs) for their objects. As in the previous survey, there is diversity in the kinds of identifiers in use: URNs, Handle, an internal PUID system generated upon aggregation, own constructions, HTTP URIs, persistent URIs, the URN-NBN-DE-Service provided by the German National Library, a PI based on original IDs, a PI based on LCSH (Library of Congress Subject Headings) PIs, and ID/INT (11).[17] Several aggregators allow content providers to continue using their own PI system, while one is in the process of developing its own system.

[17] See *Analysis of the Europeana and Athena Survey for Aggregators #2*, 15-16 (See *Analysis of the Europeana and Athena Survey for Aggregators #2*, 12 (http://pro.europeana.eu/publications/aggregators-survey-2).

Technical infrastructure

This part of the survey shows the diversity of technical systems and procedures among Europeana's aggregators. It also raises issues and concerns with systems currently in place and suggests future improvements.

Aggregating data

69% of respondents report that the stand-alone systems employed by them cover all their needs as aggregator (in terms of mapping, enrichment, cleaning processes etc.). The variety of systems that are in place across the aggregation landscape is great. Some aggregators use a well-established tool with the MINT platform, while other aggregators have developed their own systems based on various technology platforms. This is true for all steps of the ingestion workflow, including harvesting, mapping and enrichment.

Several aggregators indicate, however, that their system is in need of improvement to adapt to the changing needs of current and future aggregation. For instance, they want better tools for checking data quality and for more effectively enriching data and sharing such enrichment back to their data providers. Some point out that they need to adapt their systems to the increasing numbers of objects, collections or data providers they are working with, and to make them more user-friendly for data providers. Another issue which was raised was that it would be very useful for aggregators to be able to preview, check and test their datasets (including the mapping and harvesting) before actual publication to Europeana. They propose relaunching the Europeana Content Checker for this. Three respondents explicitly state that they are in the process of developing new systems or improving existing ones, but often aggregators lack the resources to invest in technical development, and try to find ways around this instead.

Sharing data

When it comes to sharing data – to deliver to Europeana and other aggregators, to re-publish data and to get back their materials enhanced through crowdsourcing – little more than one third (39%) of the respondents state that their current stand-alone systems essentially cover their needs. They either have no significant problems, or report having a working OAI/API in place. Problems mentioned include difficulties in harmonising data and in sending data to Europeana through OAI, working with EDM records rather than ESE, mapping data in MINT, a lack of support for data push by Europeana, the reliance on periodic harvesting, and difficulties for data providers to get back user-generated links for their collections. Some respondents indicate that they do not yet have a possibility for crowdsourcing, or that their system is currently under development.

General IT infrastructure and requirements

The survey further investigated whether their current IT infrastructure as a whole enables the aggregators to deliver their strategy, and which changes are required in the long term to meet any challenges. A little under half of the respondents (43%) answered that everything essentially works well at the moment and that nothing needs to be urgently changed. It is worrying that over

half of the respondents indicate that their IT infrastructure does not currently enable them to properly perform their duties as an aggregator.

Desired changes mentioned can be divided into four categories: hardware, software, workflow, and sustainability. First, investments in hardware, not just for better performance, but also for increased storage capacity are requested. Proposed concrete software improvements included investments in new technologies such as cloud services and Linked Open Data, the use of web-based tools, solutions for scalability issues, refinements to the annotation and enrichment process, a better CRM system to link providers to their collections, and creating a better user experience for data providers on the aggregators' websites. Several respondents indicated a wish to streamline the aggregation workflow, to work more efficiently, to be able to perform more tests before publication, to give data providers more control over the mapping and representation of their data, and to allow for the handling of increasing numbers of objects, collections and data providers. Last, respondents expressed the wish for investments in sustainability and durable storage facilities. Lack of resources was mentioned as the main reason precluding these changes.

Storing and delivering content

When asked whether they would be interested in storing or delivering content (as opposed to metadata), the respondents were divided in two. Half (13 respondents) answered no, while the other half were positive, indicating that they would be interested (3), they may be interested in the future (3), or that they already store/deliver content (7); 7 skipped the question in its entirety.

The type of content they would be interested in storing and/or delivering covered most of the regular types: full texts, images, and audio and video objects. As demands necessary for delivering this kind of material, the following were listed: high quality digitisation and specific metadata; bigger storage capacity and cost effective solutions for hosting/delivering digital content; IPR solutions; the possibility to map the aggregators' data between their platform and Europeana.

Relationship with Europeana

The aggregators are satisfied with their relationship with Europeana in general and with the Aggregation team in particular. They appreciate the revenues of participating in Europeana, but are also critical on some aspects of the aggregation and ingestion process.

Contact with the Europeana Aggregation team

Most of the respondents are in regular contact with the Europeana Aggregation team: 87% are in touch at least several times per year, of which 24% every month, and 30% several times per (monthly) publication. The majority (81%) indicates that they submit updates or new datasets at least once a year, with one fifth (21%) submitting every month. The frequency of contact means that most respondents are very active and have recent experiences with the Aggregation team and workflow.

The majority of the aggregators describe their relationship with the Europeana Aggregation team as good: 85% rate it as 'good' or 'excellent'. The majority is also satisfied with the communication between themselves and the team, with 75% describing it as 'good' or 'very good'. In the cases where the aggregators are not satisfied with the relationship or the communication, the reason often seems to be that the respondents see issues with the ingestion workflows. This may be deduced from the comments to this question as well as the answers to question 34 later in the survey (see below for further elaboration).

The aggregation workflow

Many different suggestions were submitted as the main issue which Europeana should focus on in order to improve the aggregation workflow. Although the communication with the Aggregation team was rated as generally good (previous section), particular aspects of the communication process are one the principal areas the respondents would like to adjust. More specifically, they would like to receive more (individual) feedback and support from Europeana, both upon first joining Europeana as a provider (basic information, procedures, contact details) and during the ingestion process (comments about data quality, validation problems, discarding records, errors etc.). This also includes compiling better guidelines, training and information (in the form of documentation or webinars on e.g. IPR or strategy) for data providers and aggregators. An example mentioned was a welcome pack for new aggregators explaining the basic workflow, main contacts, clear and concise documentation, acceptability criteria for data, FAQs etc. The Europeana Publication Policy and the new Europeana Professional pages for data providers and aggregators (incl. FAQs) should help to support new aggregators.[18]

The second key suggestion for change is the ingestion process as a whole: it needs to become simpler, more efficient and agile, more transparent and more flexible (for instance allowing for more frequent ingestion and publication). In addition, respondents recommend giving data providers more direct insight and influence on the publishing and representation of their data. This could be done in the form of a relaunched Europeana Content Checker, by providing more test possibilities, or by not working via an aggregator at all but dealing directly with the data providers. One respondent summarised it as follows: "the whole Europeana workflow needs to become a customer-focused service, not a Europeana-driven process".

Reasons and results of working with Europeana

The answers to this part of the survey show that aggregators start working with Europeana for many different reasons. Nevertheless, several common motives can be identified. First, aggregators want to share their data with Europeana to increase the visibility of their content (34%). In fact, for many of them publishing their data on Europeana, it is the main channel to make their content available online (28%). The second most frequent motive is the opportunity for networking and knowledge exchange that the participation in Europeana offers (31%). Respondents value the possibility to share ideas, experiences and knowledge with other professionals and organisations in the field, and profit from the inspiration, best practices and new acquaintances they gain.

[18] See http://pro.europeana.eu/publications/publication-policy and http://pro.europeana.eu/share-your-data/.

The fact that Europeana provides tools and expertise is stated by 17% of the aggregators as a reason for working with us. Several also mention shared ideals: they support Europeana's mission and goals and want to contribute to these (14%). Another 14% indicate that, as an aggregator in a particular country or domain, it is in their official mandate to aggregate data for Europeana. Some less frequent motives mentioned included the opportunity of connecting one's own collection to those of other institutions; to reach out to a different kind of audience; to enable interoperability across dispersed collections (in terms of rights labelling, shared vocabularies, multilingual access, etc.); because Europeana has an ability to influence sector standards and behaviours; because of the quality of the Europeana ingestion; and to enhance the position of one's own institutions through a collaboration with Europeana.

Interestingly, however, these motives cannot be mapped directly onto the actual benefits which aggregators identify as resulting from their work with Europeana. The main improvement the respondents note as a consequence of their participation (46%) is an increased awareness and knowledge of particular issues (such as rights labelling, metadata quality and the importance of open licences) as well as better skills and technical abilities. Indeed this was also identified as one of the chief benefits by the respondents of the previous Aggregator Survey.[19] The same proportion of respondents (46%) indicated that working with Europeana has been an incentive for innovation and development and has stimulated discussion and collaboration. 33% report that their professional network within the sector has improved. 21% answered that their institution and collections have gained a higher profile by publishing data on Europeana. Another 21% state that the quality of their metadata has improved through the use of the Europeana Data Model (EDM). Further positive results are better access to funding, and happier, more enthusiastic data providers and aggregators.

These answers suggest that Europena fulfills a key role within the cultural heritage sector by raising awareness, distributing knowledge, bringing people together and inspiring them to share their experiences and work together towards a common goal. Only one fifth of respondents highlight the increased visibility of their collections as an important benefit. This may be due to the current lack of possibilities for providers to gain insight into the (re)use of their data on Europeana. Europeana is developing a tool to solve this: the Statistics Dashboard, which is currently in an alpha phase of development.[20] The Dashboard will provide data providers and aggregators with an up-to-date overview of the content, traffic and usage statistics of their data in Europeana.

Funding

Of all the respondents, a substantial group is sustained by institutional funding (36%), receives money from a ministry or national government (30%), or is financed as part of a European-funded project (36%). An alarming 21% do not currently receive any funding at all. A smaller group relies on membership fees (12%), while some are sustained partly through own revenues and donations (Fig. 7).

[19] See *Analysis of the Europeana and Athena Survey for Aggregators #2*, 21 (http://pro.europeana.eu/publications/aggregators-survey-2).
[20] http://statistics.europeana.eu

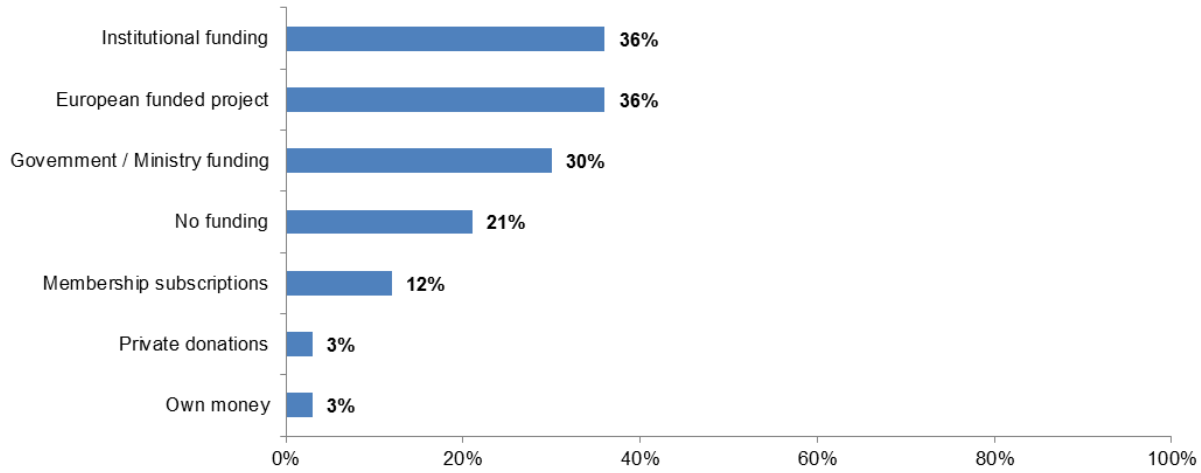

Fig. 7. Sources of funding for the aggregators.

Several aggregators combine different kinds of funding to support their organisations and operations: of the 33 respondents, 12 (36%) indicated having multiple sources of subsidy. Compared to the situation in 2011, when only 14% relied on a combination of funds, this is a sharp increase. The aggregators with mixed funding are both European projects and national aggregators.

For most respondents, the continuity of the funding is, for various reasons, still rather uncertain. For 16%, the period of funding runs out this year (2015), 6% are supported until at least 2016, 3% until 2017, and 3% already know that they will receive funding beyond 2020. The majority (71%), however, indicates that it is as yet unknown until when the aggregator will be funded. In some cases, this is because more structural or even permanent funding is still under discussion. While this certainly does not preclude the eventual financing and continuation of their work, the insecurity does obviously have consequences, such as complicating the planning of future data flow, both for the aggregators and for Europeana. Considering this uncertainty, Europeana needs to work with its aggregators even more to help them achieve more secure and long-term sustainability.

Despite the widespread insecurity regarding continued financial support, however, the survey shows a clear and considerable level of commitment towards Europeana on the part of the aggregators. 81% of respondents answered that they will continue working with Europeana after their current round of funding ends. 19% stated that they do not know yet if they will continue the collaboration; and none of the respondents indicated wanting to end the partnership after the termination of their funding.

Aggregator model

The current Europeana aggregator model

In the present aggregator model, Europeana works with various aggregators who first aggregate data from cultural heritage institutions, then process this data for submission, and finally deliver them to Europeana for publication (see the main document for more details). This model has both strengths and weaknesses, which this section of the survey sought to identify.

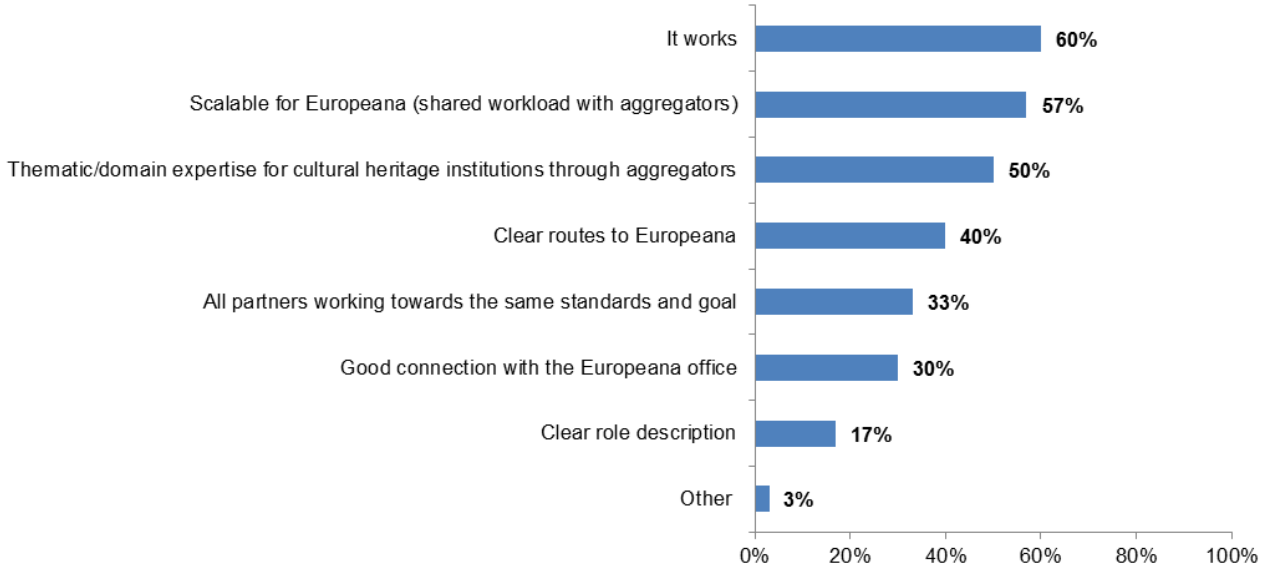

Fig. 8. Main strenghts of the current aggregator model.

The main strengths, as acknowledged by the respondents (Fig. 8), are, first, the simple fact that it works (mentioned by 60% of the aggregators); second, that it is scalable for Europeana due to the sharing of the workload with aggregators (57%); and third, that cultural heritage institutions benefit from the specific thematic or domain expertise offered by the aggregators (50%). In the comments section, several respondents further emphasise the scalability issue – as one of them wrote, “it is impossible for Europeana to satisfy the expectations and properly maintain a network of so many participants”. Other recognised strengths are that the model provides clear routes to Europeana (40%) as well as a good connection to the Europeana office (30%), that all partners work towards the same standards and end goal (33%), and that it entails a clear role description for all involved (17%).

As with the strengths, the main weaknesses identified by the respondents relate primarily to issues of workflow and relationships, but, in addition, they also include technical disadvantages.

Exactly half of the aggregators (50%) consider the main model too slow, 36% call it intransparent and chaotic, and another 36% states that it causes duplication and redundancy of data (Fig. 9). Respondents’ comments indicate that this relates to the two-tier ingestion process, the overlap between different kinds of aggregators and the subsequent division of collections, as well as the fact that feedback on metadata from Europeana is delivered to data providers through aggregators rather than to them directly. Solutions proposed are issue trackers or automated data evaluation system.

The fact that the current model is structured in a hierarchical way, with Europeana at the top, raises concerns among some respondents, who perceive a lack of partnership on equal terms. Respondents also identify disadvantages for cultural heritage institutions: they are not clearly visible on the portal europeana.eu (32%), and they have no direct contact with Europeana (11%). Moreover, there is a risk of potential competition between aggregators, while aggregators created by projects lack continuity.

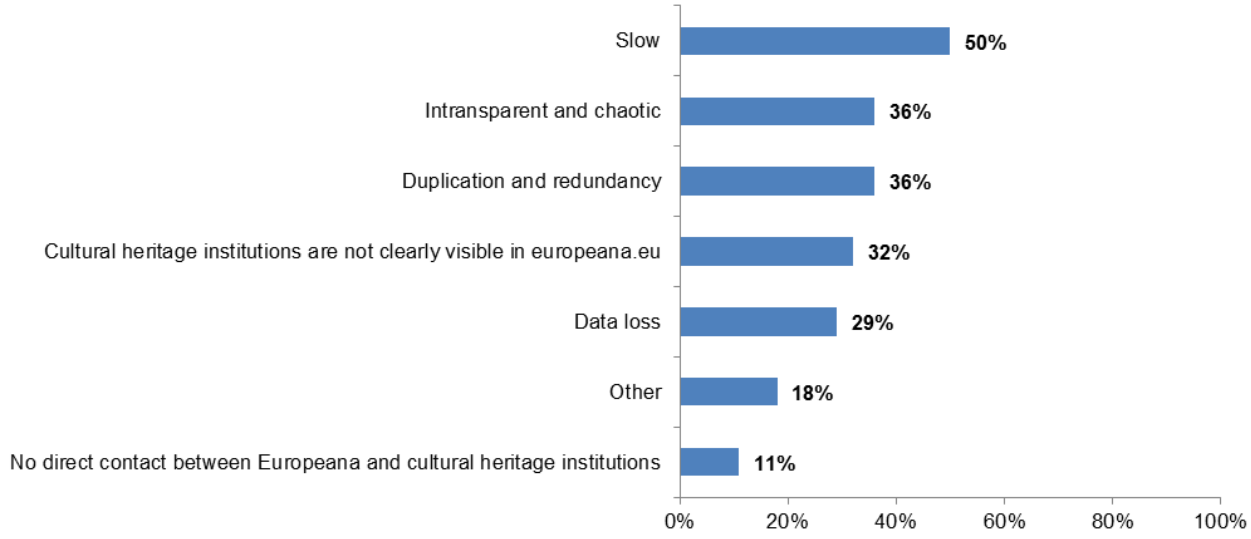

Fig. 9. Main weaknesses of the current aggregator model.

On the technical side, 29% of respondents mention the risk of data loss due to the potential double mapping/transformation of metadata; one aggregator introduces this as an argument for strengthening the links to data providers' own websites. Other problems mentioned include problems arising from multilingualism and the lack of an online tool for editing.

Further development of the aggregation landscape

The aggregation landscape is changing under the influence of the development of new technologies, interests and funding mechanisms. When asked what trends will have the biggest impact on the further development of this landscape (Fig. 10), the main answers were the increasing creative re-use of content (57%) as well as Europeana moving away from the portal into a service infrastructure (50%). Cloud computing (33%) and crowdsourcing to, for instance, help enrich metadata (33%) were further identified as influential developments, as was the European Commissions' Horizon 2020 research and innovation programme and its funding scheme (20%). Other trends which may impact the way data is aggregated included policy (European and national) and changes in political support, economic factors, the use of Linked Open Data, the transition of aggregators from projects to more permanent structures, and the emergence of tools which support digital curation and experience design at scale by larger technology players.

In the next questions, we described a vision for the future of aggregation: "Imagine all aggregators and Europeana are operating in a cloud-based environment using shared

technologies and tools, where we are dealing less with the mechanics of aggregation but more with access to and enrichment of content. Every cultural heritage institution is making the data available in the cloud, specify access rights and aggregators and/or Europeana is processing the data further, depending on their access rights (create once, publish everywhere).” Respondents were asked for their reaction to such a vision (Fig. 11).

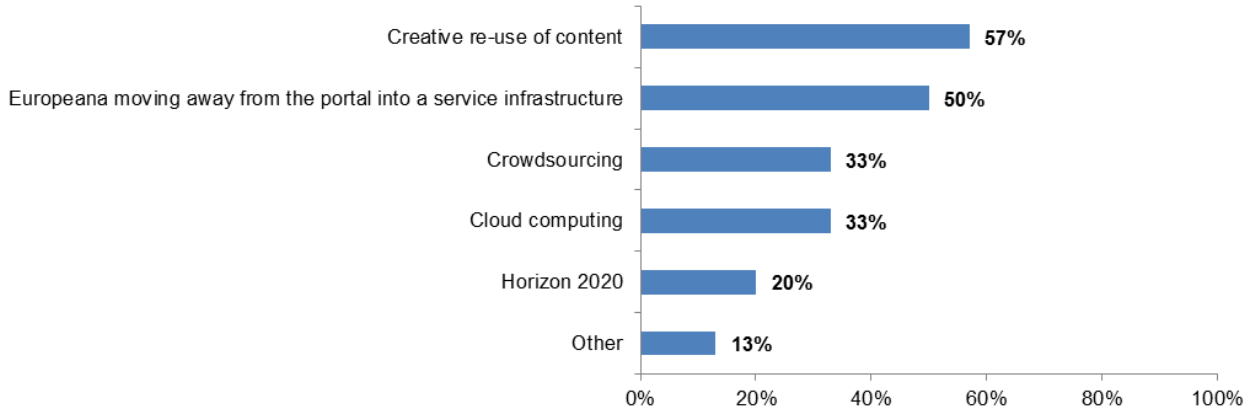

Fig. 10. Trends with the biggest impact on the further development of the aggregation landscape.

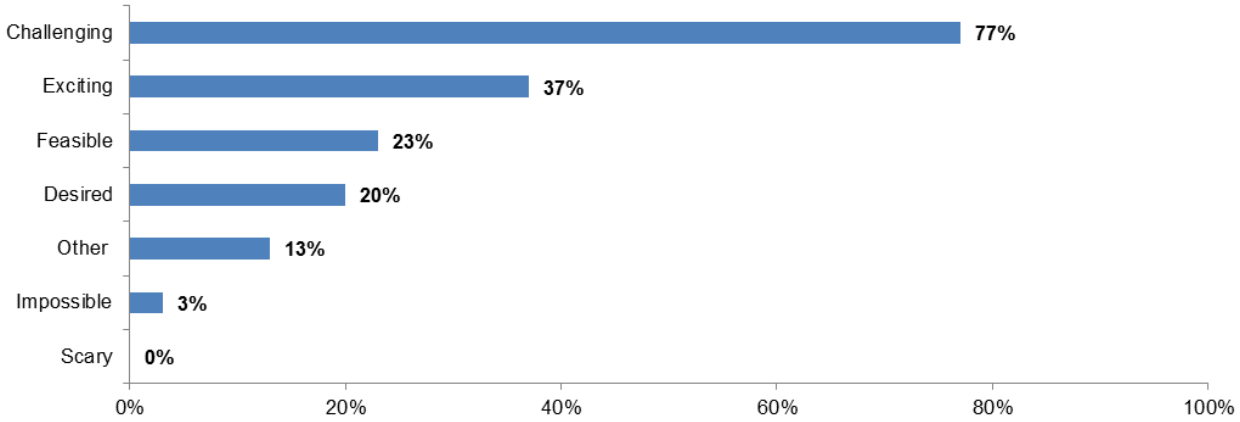

Fig. 11. Views on a cloud-based aggregation environment.

Most of the aggregators considered this scenario exciting (37%) but challenging (77%), while only a minority deemed it feasible (23%) or even desired (20%). To many, the benefits of such a cloud-based environment are still too vague, while several commentators raise concerns about the financial side of the proposal: who will pay for these cloud services? Other remarks relate to cultural heritage institutions’ lack of knowledge and resources for delivering their data to the cloud, and their fear of losing control over their data. Such a development may also create resistance among aggregators afraid of losing their position. The respondents who are excited about the proposal acknowledge that substantial investments in raising awareness are needed to gain support among the aggregators and data providers.

When asked how they expect a cloud-based way of aggregation will change current operations, the respondents indicated several different consequences (Fig. 12).

They think data sharing will become easier (57%) and feedback loops on data quality much more efficient (57%). They anticipate more value added services (e.g. data cleaning, enrichment, annotations) to become available (64%) and expect a greater focus on community building and partner relationships (including communication) (54%). In addition, they expect such a model to be more efficient: harvesting and mapping will be less time-consuming (39%), data loss will be diminished, and it will reduce overall costs and, as such, enable organisations to be sustained with less funding (43%). While some envisage a more streamlined and efficient method of aggregation, however, others point out the continued need for data mapping and support on the part of the aggregators. Overall, the answers suggest that the respondents acknowledge the potential value and benefits of a cloud-based aggregation model, but worry that it might not work, or be very difficult to implement and sustain.

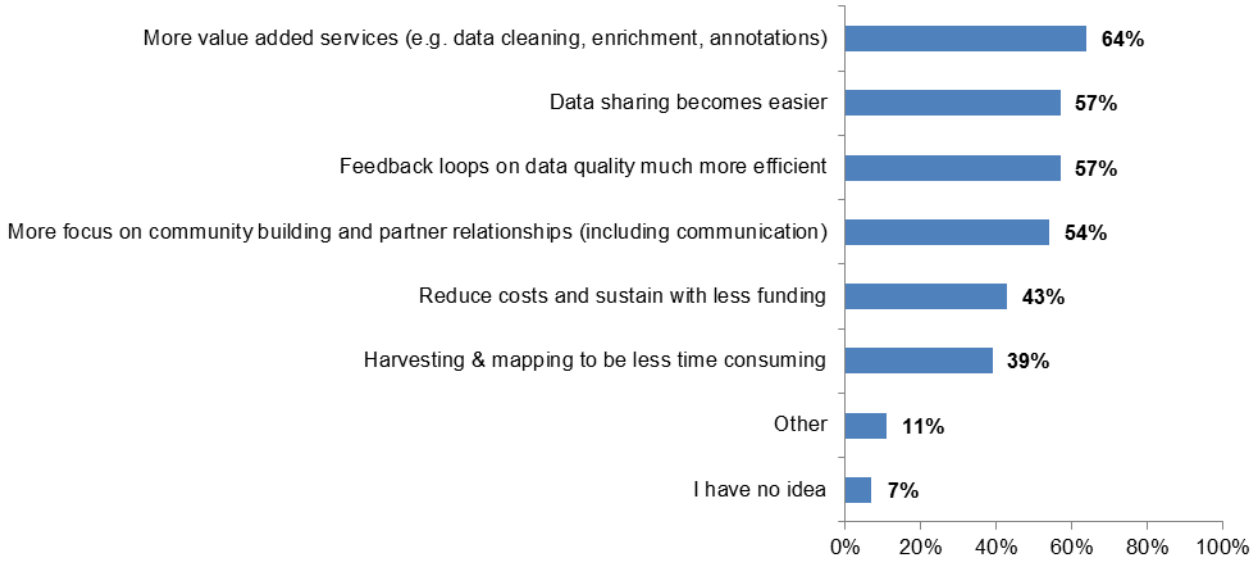

Fig. 12. Expected impact of a cloud-based aggregation model on current operations.

Conclusions

This survey has provided an overview of Europeana's aggregation landscape, including current challenges and expected developments. Some of the main findings can be summarised as follows:

- The aggregation landscape is very diverse, ranging from permanent national aggregators who aggregate cross domain and in large quantities, to thematic projects focusing on very specific content within a limited period of time. Various kinds of aggregators have different needs and face different challenges, so there is no 'one size fits all' solution.
- The aggregators who responded to the survey are active and well-established. They are aware of the values and drawbacks of the current model, competent in the technological aspects of aggregation, and informed about new developments.
- The majority of the aggregators value working with Europeana, to the extent that most of them want to continue to collaborate even after their funding ends. They particularly

appreciate the increased knowledge and skills they gain through working with Europeana, as well as the stimulation this collaboration provides in terms of development, innovation, collaboration and discussion. Networking and visibility are also identified as advantages.

- Aggregators primarily communicate with their data providers about technical issues. Many of them also communicate about the benefits of (openly) sharing data, while information about policy, strategy and funding is less frequently shared. Almost all aggregators provide their data providers with some form of documentation, training, guidance or advice on the topic of IPR.
- The respondents have very varied collection development plans; around half of the respondents have a content strategy, while almost two thirds have an ingestion plan.
- Despite the difficulties encountered, most of the respondents are willing to encourage the use of direct links.
- More than half of the respondents indicate that their IT infrastructure does not currently enable them to perform their duties as an aggregator properly, and will face even more challenges in the future. Improvements in hardware, software, workflow and sustainability are needed.
- Most of the respondents are in regular contact with the Europeana Aggregation team and satisfied with their relationship and communication with the team. The main things they would like to change in the aggregation process are more feedback and support from Europeana and a streamlined ingestion workflow.
- Several aggregators combine different kinds of funding to support their organisations and operations. The insecurity over future funding which many experience might complicate the planning of data flow both for aggregators and for Europeana.
- The main strengths of the current aggregation model – data providers delivering to aggregators, who submit data to Europeana – are that it works, that it is feasible in terms of scale, and that cultural heritage institutions benefit from the specific expertise offered by the aggregators. The main weaknesses are the slow, chaotic and intransparant nature of the process, and the risk of duplication and redundancy.
- The respondents acknowledge the potential value and benefits of a cloud-based aggregation model, but they see many challenges in making it work.

Survey questions

Europeana Aggregator Survey

Welcome to the Europeana aggregator survey

Dear Europeana partner,

Thanks a lot for filling in the Europeana aggregator survey.
It is your opportunity to influence and inform the future development of aggregation in general and the development of our collections. Please make use of this opportunity and complete this survey before the 30th of January 2015.

The main objectives of the survey are to:

- Refine the characterisation of different types of aggregators currently in operation or development in Europe;
- Map the current 'ecosystem' of aggregators across Europe;
- Understand the relationships of aggregators with their partners and with Europeana;
- Understand and model the different operational models by which these aggregators are supported and sustained;
- Assess the resilience of the aggregation ecosystem as a core priority of Europeana's planning as a core service platform under the Connecting Europe Facility.

If you have any comments or questions about the survey, please get in touch with joris.pekel@europeana.eu, or add them at the end of the survey. There is also a pdf available for you to have a look at before filling in the survey.

Thanks again, and looking forward to speaking again soon.

Kind regards,

The Europeana Aggregation Team

p.s: please note that while testing the survey, we sometimes got send back to the first page of the survey after clicking on the 'next page' button. This is an annoying bug in Surveymonkey that we can't fix right now. Your old results should be saved however and you can continue from the last page that was submitted successfully. We apologise for any inconvenience in advance.

General Information

1. Information about the Aggregator
   - Name
   - Portal URL
   - Legal entity running the Aggregator
   - Organisation website

2. Primary contact for the Aggregator
   - Name
   - Address
   - Email

3. What is the geographic coverage of partners providing data to the aggregator?
   a) Worldwide

b) European
c) National
d) Regional
e) Other (please specify)

4. Please specify the type of aggregation.
a) Cross Domain (including galleries, libraries, archives, museums, etc)
b) Single Domain (either galleries, libraries, archives or museums, etc)
c) Thematic (e.g. fashion, natural history, archaeology, etc)
d) Other (please specify)

5. Please specify the domains you include in the aggregator
a) Galleries
b) Libraries
c) Archives
d) Museums
e) Audiovisual Archives
f) Sound Archives
g) Performing Arts
h) Publishers
i) Other (please specify)

Partner Relationships

6. For how many institutions are you currently aggregating content?
a) 0
b) 1
c) 2-10
d) 10-100
e) 100+

7. Is an up-to-date list of your partner institutions accessible?
a) Yes
b) No
If yes, please indicate how to get access to that list.

8. How do you as an aggregator extend the network of contributing partners?
a) Natural growth (being approached by interested parties)
b) Active and specific sourcing for new partners
c) Regular but general call for participation
d) Via Europeana and the Europeana Network
e) Not at all
If not at all, please specify the reasons why.

9. Do you have information about the number of cultural or scientific institutions that exist in your country or in Europe that are potential partners for aggregators or Europeana?
a) Yes
b) No
If yes, please share this information.

10. Are you partnering with another aggregator or project to prepare and submit data to Europeana (e.g. a regional aggregator contributes to a national aggregator that contributes to Europeana)?
a) Yes

b) No

If yes, please specify.

11. How do you communicate with partners and data providers?
    a) Personal e-mail
    b) E-mail lists
    c) Basecamp
    d) Phone/Skype
    e) Other (please specify)

12. What are you communicating to all your partners and data providers?
    a) Data specific feedback (e.g. in order to improve the quality)
    b) Advise about tools, technologies, mapping, copyright
    c) Incentives and benefits for opening up and sharing data
    d) Information shared by Europeana via the Aggregator Forum
    e) Publications from the European Commission (e.g. directives, call for proposals)
    f) Information and knowledge relevant for the scope of the aggregator (national, domain, thematic)
    g) Other (please specify)

13. Do you have a dedicated partner agreement with all data providers?
    a) Yes
    b) No

Comments.

IPR, content strategy and data quality

14. How do you make sure that each data provider signs the Europeana Data Exchange Agreement (DEA) before data are provided to Europeana?
    a) DEA is covered by the partner agreement
    b) Every data provider is asked to sign the DEA individually
    c) No mechanism to sign the DEA
    d) Other (please specify)

15. Do you provide IPR documentation, training or other kind of IPR guidance four your partners and data providers?
    a) Documentation
    b) Training
    c) Other (please specify)

16. If you have IPR documentation, please indicate how to get access to the documents.

17. Do you have a content strategy or other policy or guideline documents?
    a) Yes
    b) No

18. Are these documents publicly available?
    a) Yes
    b) No

If Yes, please indicate how to get access to the documents

19. Do you have a specific focus for developing the collections and if this is the case, what is the focus of the aggregator's collection development?
    a) No focus

b) Masterpieces
c) Curated collections
d) Outcomes of mass digitisation programmes
e) Enriched data
f) Other (please specify)

20. Do you have an ingestion plan where you schedule the amounts of data you are processing?
a) Yes
b) No

21. How many digital objects approximately do you plan to deliver to Europeana in 2015?
a) 0
b) 1 - 100
c) 100 - 1000
d) 1000 - 10000
e) 10000 - 100000
f) 100000 - 500000
g) 500000 – 1000000
h) > 1.000.000

22. Direct links to digital objects (not to viewers or websites) delivered with edm:isShownBy are an important mechanism to improve the user experience as it only takes one click from the Europeana portal to access the object. It also makes access to digital objects via the API much easier. Are you encouraging and helping your partners to provide direct links to digital objects?
a) Yes
b) No

23. If you encourage providing direct links but still experience difficulties in getting them, what are the main reasons for that?
a) Not aware of the importance of direct links
b) Technical issues related to the Digital Asset Management System
c) Resource issues (e.g. additional workload)
d) Institutional policies
e) Other (please specify)

24. Do you use persistent identifiers?
a) Yes
b) No
If yes, please specify

Technical infrastructure

25. Do stand-alone systems employed and used by you cover all your needs as aggregator? Do you have all you need in terms of mapping, enrichment, cleaning processes?
a) Yes
b) No

26. Please elaborate.

27. Do stand-alone systems employed by you cover all your needs in terms of sharing? How easy is it for example to publish to Europeana? To other aggregators? To re-publish? To get back their materials enhanced through crowdsourcing? Please explain.

28. Does your current IT infrastructure enable you to deliver your strategy as an aggregator? 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? Please explain.

29. Are you interested in in storing or delivering content (as opposed to metadata). What is the type of content? What are the demands on delivering it?

Relationship with Europeana

30. How would you describe your relation with the Europeana aggregation team?
    a) Excellent
    b) Good
    c) Neutral
    d) Bad
    e) The worst
    f) Non existent

31. How often are you in touch with the Europeana aggregation team?
    a) Every publication multiple times
    b) Once every month
    c) 1-5 times a year
    d) Not in the last year
    e) Other (please specify)

32. How satisfied are you with the communication between you and the Europeana aggregation team?
    a) Very good
    b) Good
    c) Neutral
    d) Could be better
    e) Terrible
    f) Non existent

If you want, please explain.

33. How often do you submit updates or new datasets to Europeana?
    a) Every month
    b) 6 -11 times a year
    c) 1 - 5 times a year
    d) Still preparing for first submission
    e) Other (please specify)

34. What is the main thing that Europeana should focus on in order to improve the aggregation workflow? Please explain.

35. What are the most important reasons for you as an aggregator to work with Europeana? Please explain.

36. What improved in your organisation since you started working with Europeana? What is the most valuable for you as an aggregator? Please explain. If nothing improved yet or you are not getting the expected values, please specify.

Funding

37. How is the aggregator funded?
    a) Institutional funding
    b) Membership subscriptions
    c) Government / Ministry funding
    d) European funded project
    e) Private donations
    f) No funding
    g) Other (please specify)

38. Until when is the aggregator funded?
    a) 2015
    b) 2016
    c) 2017
    d) 2018
    e) 2019
    f) 2020
    g) > 2020
    h) Unknown

39. Will you continue working with Europeana after your current round of funding ends?
    a) Yes
    b) No
    c) Don't know yet

Please specify.

Aggregator model

40. What are the main strengths of the current aggregator model (Europeana partners with aggregators which aggregates from cultural heritage institutions)?
    a) It works
    b) Clear routes to Europeana
    c) Good connection with the Europeana office
    d) Clear role description
    e) Thematic/domain expertise for cultural heritage institutions through aggregators
    f) All partners working towards the same standards and end goal
    g) Scalable for Europeana (shared workload with aggregators)
    h) Other (please specify)

41. Optional: please elaborate.

42. What are the main weaknesses of the current aggregator model (Europeana partners with aggregators which aggregates from cultural heritage institutions)?
    a) Intransparent and chaotic
    b) Slow
    c) Data loss
    d) Duplication and redundancy
    e) Cultural heritage institutions are not clearly visible in europeana.eu
    f) No direct contact between Europeana and cultural heritage institutions
    g) Other (please specify)

43. Optional: Please elaborate.

44. What trends do you think have the biggest impact on the further development of the aggregation landscape?

a) Cloud computing
b) Creative re-use of content
c) Europeana moving away from the portal into a service infrastructure
d) Horizon 2020
e) Crowdsourcing
f) Other (please specify)

45. Optional: please elaborate.

46. Imagine all aggregators and Europeana are operating in a cloud based environment using shared technologies and tools, where we are dealing less with the mechanics of aggregation but more with access to and enrichment of content. Every cultural heritage institution is making the data available in the cloud, specify access rights and aggregators and/or Europeana is processing the data further, depending on their access rights (create once, publish everywhere). See this video to make this vision clearer. What attributes come up when you think about this vision?

a) Scary
b) Exciting
c) Challenging
d) Impossible
e) Feasible
f) Desired
g) Other (please specify)

47. Optional: please elaborate.

48. Having a vision like the one presented above in mind: How you expect this to change your current operations?

a) Harvesting & mapping to be less time consuming
b) Data sharing becomes easier
c) Feedback loops on data quality much more efficient
d) More focus on community building and partner relationships (including communication)
e) More value added services (e.g. data cleaning, enrichment, annotations)
f) Reduce costs and sustain with less funding
g) I have no idea
h) Other (please specify)

49. Optional: please elaborate.

Thanks

50. Thanks for filling in the survey. Please fill in here any additional comments and/or suggestions you would like to share.