

MS2: UPDATED PARTNER DEVELOPMENT PLAN, MS4: UPDATED COLLECTIONS AND DATA DEVELOPMENT PLAN, MS5: IMPROVED QUALITY OF METADATA (PARTNER AND DATA DEVELOPMENT PLAN)



MILESTONE

Project Acronym: Ev3

Grant Agreement number: 620484

Project Title: Europeana Version 3

MS2: UPDATED PARTNER DEVELOPMENT PLAN, MS4: UPDATED COLLECTIONS AND DATA DEVELOPMENT PLAN, MS5: IMPROVED QUALITY OF METADATA

(Merged and renamed to: Partner and Data Development Plan)

Revision	Final version
Date of submission	14/05/2015
Author(s)	Joris Pekel & Ylva Klaassen, Europeana
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
1	20/04/2015	Ylva Klaassen & Joris Pekel	Europeana	First draft
2	22/04/2015	Henning Scholz & Ylva Klaassen	Europeana	Feedback and comments
3	30/04/2015	Joris Pekel	Europeana	Second draft
4	04/05/2015	Henning Scholz	Europeana	Feedback
5	12/05/2015	Joris Pekel	Europeana	Final text version
6	13/05/2015	Imogen Greenhalgh	Europeana	Spelling and grammar check
7	14/05/2015	Joris Pekel	Europeana	Final version in template

Statement of originality:

This milestone contains original unpublished work except where clearly indicated otherwise.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

- Table of Contents..... 3**
- Executive summary 4**
- 1. Introduction 6**
- 2. Content and Collections 8**
 - 2.1 Quantitative content analysis..... 8
 - 2.1.1 Total number of objects 8
 - 2.1.2 Content by Member State..... 9
 - 2.1.3 Contribution by non-EU countries 11
 - 2.1.4 Content by domain and provider..... 12
 - 2.1.6 Audiovisual content 13
 - 2.1.7 20th century content 15
 - 2.2 Qualitative content analysis 15
 - 2.2.1 Implement the Content Re-use Framework 15
 - 2.2.2 Media File Checker..... 16
 - 2.2.3 Increase the number of objects in the database that can be freely re-used 16
 - 2.2.4 Channels and Thematic content 16
- 3. Metadata 18**
 - 3.1 Previews..... 18
 - 3.2 Direct links to content 19
 - 3.3 Native EDM 19
 - 3.4 Rights labelling 20
- 4. Partners 21**
 - 4.1 New partnerships 21
 - 4.2 Existing partnerships 21
 - 4.3 Events and knowledge sharing 22
 - 4.4 Aggregation infrastructure 22
- 5. Conclusions 24**

Executive summary

This combines the projected milestones *MS2 Updated Partner Development Plan*, *MS4 Updated Collection and Data Development Plan* and *MS5 Improved Quality of Metadata* of Europeana v3 into one document, covering the main tasks of the Europeana Data Partner Services team. It deals with Europeana's data (content+metadata) and partners (aggregators and data providers). It evaluates the accomplishment of earlier goals, provides an overview of the current situation, and defines aims for the coming years.

Europeana has brought together more than 42 million digital objects from the online collections of more than 3,300 galleries, libraries, museums and archives from across Europe. This is a great achievement and a shared effort, but we need to continue to build this open and trusted source for European cultural heritage. We would like to continue to work together with our contributing partners towards this common goal for mutual benefit. To do this, we have defined some key strategies and specific targets for the coming years.

This document consists of two parts. In the first section we discuss content¹ and metadata.² While Europeana's key activity is the harvesting and publishing of large sets of metadata, the actual digital cultural heritage objects that users find in the Europeana database have always been of great importance. We see metadata as the key product to help find the content the user is looking for. In previous years, we therefore worked not only on improving the quality of the metadata, and but on the likelihood that content gets found, while also improving the quality of the digital object stored on the server of the data provider. Over the last three years online users' expectations and demands have radically changed. To keep providing them a relevant product, Europeana and the data it gives access to need to live up to these demands. In this first chapter we'll reflect on the work completed over the previous years and, more importantly, looks at the plans and focus points for the future.

The second part describes the partner development plan. At the moment, over 3,300 cultural institutions from across Europe have made (parts of) their collection available via Europeana. This is a great achievement that could not have been done without a dedicated network of aggregators and projects. Europeana has been, and will always be, dependant on the work initiated by ministries of culture, national aggregators and the individual institutions. In 2015, though the number of partners we have will continue to grow, we want to make sure that we nurture and improve the relationship with existing partners too. This means there must be a strong focus on providing appropriate assistance for institutions to improve the quality of the data they contribute, but also, for example, to improve our own aggregation infrastructure.

Where relevant, this document will also discuss the implications and goals of the Connecting Europe Facility (CEF) program funded by the European Commission. In the next couple of years,

¹ The term 'content' here refers to a *physical or digital 'object' that is part of Europe's cultural and/or scientific heritage, typically held by the data provider or by an aggregator of the data provider.* (Description taken from the Europeana Data Exchange Agreement (DEA), article 1.)

² The term 'metadata' can be defined as *the textual information (including hyperlinks) that may serve to identify, discover, interpret and/or manage content.* The term 'preview' refers to *a reduced size or length audio and/or visual representation of content, in the form of one or more images, text files, audio files and/or moving image files* (descriptions from the Europeana Data Exchange Agreement (DEA), article 1.)

MS2: UPDATED PARTNER DEVELOPMENT PLAN, MS4: UPDATED COLLECTIONS AND DATA DEVELOPMENT PLAN, MS5: IMPROVED QUALITY OF METADATA (PARTNER AND DATA DEVELOPMENT PLAN)

Europeana will take a new direction to become a true core service platform and to achieve this, a number of goals with relation to data and partners have to be achieved.

As this report shows, there are several challenges that require our attention, ranging from improving the quality of the data to providing accurate information about the copyright status of the works. This report should therefore be seen as a starting point for further discussion and development to improve the Europeana repository.

1. Introduction

The Partner and Data Development Plan combines Milestones MS2, MS4 and MS5 of Europeana v3 into one document and covers the main tasks of the Europeana Data Partner Services team for 2015. It deals with Europeana's collections (content), metadata and partners (aggregators and data providers). It has three main purposes. First, it follows up on two previous reports (D2.2 Partner Development Strategy of January 2013 + D3.8 Collection and Data Analysis of March 2013), indicating the extent to which the goals defined in those reports have been accomplished in the last two years. Second, it provides an overview of the current status of Europeana's collections, metadata and partners. Third, it defines general aims and specific targets for future development, based on the Europeana Business Plan 2015³ and the Europeana Strategy 2015-2020.⁴

The first part of this report (chapter 2) will analyse the content, chapter 3 investigates the metadata, and chapter 5 looks at Europeana's partners. Finally, in chapter 6, we present some conclusions and make some recommendations for the further improvement of the Europeana repository, for further discussion with the Europeana aggregators and data providers.

Before discussing the main findings, the terms 'content' and 'metadata' must be clearly defined. For the most part, Europeana does not hold or stock any of the actual digital objects (the content)⁵ which belong to the partner institutions.⁶ Europeana assembles the metadata (information describing the object), the preview associated to an item and a link to the digital object on the partner's website. This has implications for the possibilities of effectuating improvements to our repository. Europeana has, for example, little influence over the quantitative and qualitative aspects of the content and collections made available by the partner institutions, although recommendations can be, and are, made to steer them in certain directions. We do, however, have a more direct responsibility to improve the quality of the repository through the improvement of the metadata, previews and links. Managing both content and metadata effectively will lead to improvements in the user experience.

With nearly 42 million objects in the repository, we can say that we have assembled a 'critical mass' of cultural artefacts. However, if the aim of Europeana and its partners is to represent the full breadth of European cultural heritage, then this is only the beginning. We will therefore continue to increase the amount of objects in the database with the aim of including over 45 million objects by the end of 2015. But that by itself is not enough. In parallel with these quantitative objectives, improving the quality of the database remains of crucial importance. End-users and re-users are increasingly critical of 'quality' in the broadest sense of the word. Not only should what they are looking for be easy to find on the web, but the information about the object has to be accurate and informative, and preferably in a number of languages too. They also

³ Business plan 2015:<http://pro.europeana.eu/publication/make-the-beautiful-thing-business-plan-2015>

⁴ <http://strategy2020.europeana.eu/>

⁵ The term 'content' here refers to "*a physical or digital 'object' that is part of Europe's cultural and/or scientific heritage, typically held by the data provider or by an aggregator of the data provider*" (description taken from the Europeana Data Exchange Agreement (DEA), article 1).

⁶ With the exception of user-generated content (UGC) which has been added to the database from 2011 onwards.

desire that the digital object is of high resolution, as well as re-usable and sharable across the web.

The improvement of the quality of the data, with a strong focus on accessibility, accuracy, consistency and re-usability of metadata and content, is key to fuel the creative industry and reach our end-users, which then gives as high return as possible to the memory institutions opening up their data. In early 2015 the recommendations of the Europeana Network Task Forces on metadata quality and content re-use were delivered and we started to implement these recommendations together with the *Europeana Publication Policy* that was published in December 2014.

To achieve these aims, we will establish new partnerships (chapter 4), with a focus on organisations who can provide high-quality content in the areas that are defined as high priority in the annual business plans. We will nurture the relationships with our current partners by facilitating knowledge transfer and the sharing of experiences, and by intensifying the collaboration and participation of the contributing organisations during the data provision cycle. Moreover, we will help our partners make the most out of their participation in Europeana and maintain a dialogue as to how we can help and simplify the work done by data providers and aggregators.

Moreover, the aggregation infrastructure will be remodelled and improved (4.4). The transformation of the aggregation process, now based on a shared technical infrastructure, will help to make it easier for institutions to participate and share their collections with the world. By spring 2016 we want to have identified the business requirements for the new aggregator model and a roadmap should be in place to implement the new model.

2. Content and Collections

2.1 Quantitative content analysis

The total number of objects represented in Europeana has, from the start, been an important indicator for measuring progress. This chapter analyses Europeana's content and collections in order to gain qualitative and quantitative insight into the the repository, and to identify existing gaps within it.

2.1.1 Total number of objects

With the advent of new businesses operating with large datasets, the increase in total objects remains a key goal. Over the past few years, Europeana has been very successful in achieving, indeed surpassing, the targets for the total number of objects. The target figure set in the 2013 business plan was 27 million objects by the end of 2013.⁷ This was achieved and with a wide margin: by the end of 2013, there were over 30 million objects in the Europeana database.

As of February 2015, the number of available objects is 40 million, which means that the 2014 KPI of 33 million objects has also been surpassed by a significant amount. Figure 1 below shows the growth of the amount of content (in number of objects in millions) in the Europeana portal over the last five years. This quick increase in numbers result from the great increase in new cultural institutions that wanted to share their data via Europeana, as well as more aggregators reaching a mature status. The increase has been particularly substantial in the past year.

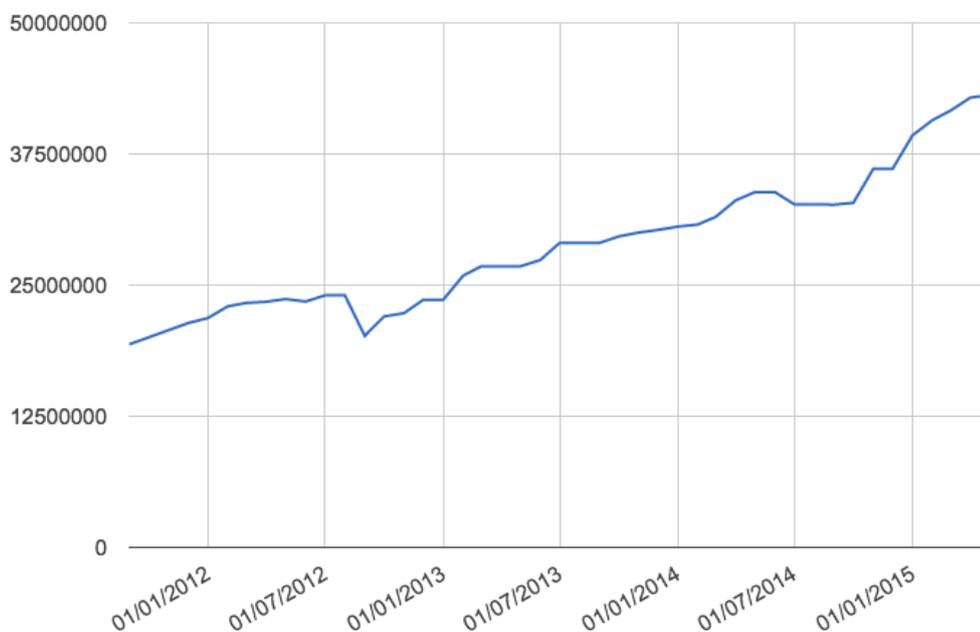


Figure 1. Number of records in Europeana 01/01/2012 - 01/05/2015⁸

⁷ Europeana Business Plan 2013

http://pro.europeana.eu/files/Europeana_Professional/Publications/Europeana%20Business%20Plan%202013.pdf

⁸ For the raw data see:

https://docs.google.com/spreadsheets/d/1H_LXEY7G9c8RjxtrKNW9V93XemBxLckzmJNd3JhQW9M/edit#gid=0

On the basis of these achievements, we can expect to successfully meet the target of 45 million objects in our dataset by the end of 2015.⁹ The good progress to date means that we now also have room to focus on the quality and consistency of the database, which are both addressed in the sections below.

2.1.2 Content by Member State

Europeana aims to offer access to the cultural heritage of all European countries. However, not all EU Member States are equally well-represented in the collections in Europeana's database. Therefore, the *Strategic Plan 2011-2015* stressed the need to ensure that the repository reflects the diversity of our respective cultural heritage.¹⁰ Indicative targets for the minimum contribution of content to Europeana by each Member State, to be achieved by the end of 2015, were set by the European Commission in a recommendation in 2011.¹¹ Table 1 shows for each Member State the original targets, the results at the end of 2013, the current situation, and the remaining content gap.

To work towards a more balanced representation of EU Member States, one of the key tasks outlined in Europeana's *Business Plan 2014* was to pay special attention to those countries with a different larger than 50% between their actual contribution and their content targets for 2015. These were: Belgium, Bulgaria, Cyprus, the Czech Republic, Greece, Italy, Latvia, Portugal, Romania, Slovenia, Slovakia, and the United Kingdom.¹²

The collection is more balanced in terms of Member State representation at the start of 2015 than one year ago. Table 2 shows the top 15 contributors to Europeana in terms of amount of records.

With regard to the content targets set by the EC, the countries can be grouped as follows:

- 13 Member States have already reached and even surpassed their targets for 2015 (Austria, Belgium, Czech Republic, Denmark, Estonia, Hungary, Latvia, Lithuania Luxembourg, the Netherlands, Poland, Spain and Sweden).
- A further 4 countries have reached 80% or more of their targets and need to deliver at most 20% in the remaining year of the EC's five-year period 2011-2015 (Tier 3 countries: France, Germany, Italy, Malta and Slovenia).
- Another 2 Member States are well over 50% of their targets and have between 20% and 50% remaining (Tier 2 countries: Greece, and the United Kingdom).
- The final 7 countries still need to supply over half of the content to reach their targets (Tier 1 countries: Bulgaria, Cyprus, Finland, Ireland, Portugal, Romania, and Slovakia).

⁹ *Europeana Business Plan 2015*, <http://pro.europeana.eu/publication/make-the-beautiful-thing-business-plan-2015>.

¹⁰ *Europeana Strategic Plan 2011-2015*, <http://pro.europeana.eu/publication/strategic-plan-2011-2015>

¹¹ European Commission recommendation 2011/711/EU of 27 October 2011.

¹² *Europeana Business Plan 2014*, 12. <http://pro.europeana.eu/publication/business-plan-2014>

Country Name	Achieved of Goal in % 25-03-2014	Achieved of Goal in % 27-03-2015
Austria	111.19	299.24
Belgium	69.54	142.63
Bulgaria	26.05	35.67
Cyprus	20.81	45.64
Czech Republic	46.08	108.73
Denmark	150.02	214.55
Estonia	535.58	418.53
Finland	75.85	21.18
France	75.45	98.41
Germany	83.73	92.38
Greece	44.09	77.82
Hungary	142.73	162.43
Ireland	88.43	19.11
Italy	66.56	89.25
Latvia	79.86	147.25
Lithuania	80.90	112.68
Luxembourg	131.10	230.47
Malta	89.85	85.33
Netherlands	179.74	346.70
Poland	92.47	117.98
Portugal	37.37	45.68
Romania	7.24	13.94
Slovakia	42.02	43.10
Slovenia	60.17	98.86
Spain	107.17	150.35
Sweden	115.06	172.25
United Kingdom	53.31	77.93

Table 1. Progress data providing per country according to the MSEG goals.¹³

¹³ For the raw data see:

https://docs.google.com/spreadsheets/d/1QeKsBmJVZvYntwonsShzUi_uGIhck7a23ScOHggqYTw/edit#gid=33857245

1	Netherlands	5,446,639
2	Germany	5,077,208
3	France	4,239,513
4	Spain	4,023,266
5	Sweden	3,334,681
6	Italy	3,306,825
7	United Kingdom	3,069,480
8	Poland	1,858,199
9	Norway	1,806,056
10	Austria	1,795,452
11	Belgium	1,082,571
12	Denmark	971,916
13	Europe	923,548
14	Hungary	677,316
15	Czech Republic	534,956

Table 2. The top 15 Member States contributing to Europeana, March 2015. The field 'Europe' represents data contributed from pan-European projects, where the data is not indicated at country level.¹⁴

In 2014 additional support was given to countries with a significant content gap. Workshops and roundtable discussion were organised in Romania, Bulgaria, Croatia, Slovakia, Hungary and Denmark. The goal of these meetings was to support the national aggregator in finding new partners, data modelling or copyright questions. For some countries, however, it has remained difficult to deliver substantial amounts of data to Europeana. This is, in many cases, due to a lack of governmental support from the individual member state. The top 5 countries all have dedicated financial and political support from their national ministries. In 2014, the amount of records contributed by Finland and Ireland dropped significantly due to the completion of the rights labelling campaign. In 2015 we expect both countries to return large amounts of data to Europeana with a valid rights statement.

Overall we are happy to see that in 2014 most countries moved up at least one tier. Nearly half of them have already reached their targets, and we expect at least another 5 to achieve them in 2015.

2.1.3 Contribution by non-EU countries

One of the difficulties has always been to establish the geographic boundaries of the 'European cultural heritage' represented in Europeana. Europeana has focused on contributions from EU Member States but has also collected data from other European countries outside of the European Union. Norway, Iceland, Switzerland, Russia, Serbia, Turkey and Croatia are seven data-providing countries which make significant contributions to the repository. Together they

¹⁴ For the full dataset see:

https://docs.google.com/spreadsheets/d/1QeKsBmJVZvYntwonsShzUi_uGIhck7a23ScOHggqYTw/edit#gid=75395365

provide more than 2.5 million objects (6% of the total number of objects in Europeana). Norway is the largest non-EU country in Europe provider with more than 1.8 million objects (4.3% of the total number of objects in Europeana) and Iceland is the second highest non-EU provider with more than 346,000 objects. While contributions from EU Member States will remain a priority, we welcome relevant objects from non-Member States and non-EU countries.

2.1.4 Content by domain and provider

One of the unique aspects of Europeana is that it brings together, in one repository, a wealth of material from several different domains: galleries, libraries, archives, audiovisual collections and museums. As well as aiming for a geographically balanced distribution of content, we also strive for an even representation of these different domains in our collection. At the start of 2013, 43% of the providers was cross-domain, 40% were libraries, 10% museums, 6% audiovisual archives, and 1% archives.¹⁵ As part of the key tasks set out in the Business Plan 2014 focused on creating a more balanced representation of domains, with a particular focus on increasing the amount of content from archives (incl. audiovisual archives), a domain which was lagging behind.¹⁶

In the past two years, we collaborated closely with aggregators, projects and institutions which could help provide more archival and museum content, or which were focused on material that is not usually available in a digital form through public heritage collections. These were, for instance, APEX (for archival material), EUscreenXL, EFG and Europeana Sounds (for audiovisual archives), Europeana Fashion (for fashion design), Europeana Newspapers (commercial newspaper publishing), Europhoto/Europeana Photography (commercial photography) and Digitising Contemporary Art (contemporary art). This resulted, in particular, in a large increase in the audiovisual domain, where now over 1.2 million objects can be found. At the same time the amount of text and image records have also substantially grown, for example through the Europeana Newspapers project which contributed about 18 million newspaper pages (3.6 million titles and issues in Europeana).

The distribution of data providers across the domains is closely connected to the means by which data is provided. Memory institutions can supply data to Europeana in four different ways:

- By delivering data to a national, usually cross-domain aggregator (such as Hispana for Spain or Digitale Collectie for the Netherlands).
- By delivering data to a domain-specific pan-European aggregator (such as The European Library for libraries).
- By participating in aggregating projects focusing on particular domains or themes (such as Europeana Fashion for fashion or EFG for film).
- By delivering data directly to Europeana (as a limited number of institutions such as the Dutch Rijksmuseum and the Institut national de l'audiovisuel in France do).

At the moment, Europeana has 113 direct providers, but not all of them are still active. These are mostly EU funded projects like Athena and Europeana Judaica where project funding has ended, or projects like EUscreen where the partners have moved to a new project, EUscreenXL.

¹⁵ D2.2 *Partner Development Strategy*, 24.

¹⁶ *Europeana Business Plan 2014*, 12.

	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

Table 3. 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. Table copied from D.1.1.¹⁷

All domain aggregators and some thematic aggregators (former EU funded projects) are direct partners of the Europeana DSI. These domain and thematic aggregators bring their networks in direct contact with Europeana. In the recent aggregator survey we held, the aggregators were asked to indicate how many records they expect to deliver to Europeana in 2015. Based on these answers, a very conservative count – taking the lowest number in each category – suggests that almost 10 million new objects can be expected this year. Of this estimate, around 1 million records will come from the archives via the Archives Portal Europe, and another million AV-objects from EUScreenXL and Europeana Sounds which will greatly help to balance the amount of records per domain.

2.1.6 Audiovisual content

There were (and still are) certain content types which are underrepresented in the Europeana dataset. At the start of 2013, only a small amount of content was audiovisual material. To create more balance in our collection, we devoted special attention to increasing the amount of certain types of content, with the aim of having 1.1 million audiovisual objects in the Europeana portal by the end of 2013.¹⁸

¹⁷ D.1.1: Recommendations to improve aggregation infrastructure

http://pro.europeana.eu/files/Europeana_Professional/Projects/Project_list/Europeana_Version3/Deliverables/EV3%20D1_1%20Aggregation%20Infrastructure.pdf

¹⁸ Europeana Business Plan 2013. <http://pro.europeana.eu/blogpost/europeana-business-plan-2013-published>

The past few years have seen concerted efforts to increase the amount of audiovisual material in the dataset. Europeana worked with two key projects to address this content gap: with EFG and EUScreenXL. In addition, we actively engaged with different audiovisual archives in collaboration with EFG for film and EUScreenXL for television archives, and we encouraged our network of aggregators to work with their local audiovisual archives. Nevertheless, it has been very difficult to achieve the target figures for several reasons. These include the commercial use of most existing material, which usually results in complex rights issues, as well as the fact that digitisation of audiovisual material is significantly more expensive and labour-intensive than that of other types of material.¹⁹ Another point to mention is that audiovisual archives do not tend to calculate the size of their archives through the amount of records they hold, but in how many hours of film they have. A 20 second clip and a 50 minute movie are both counted as 1 object in Europeana which does not accurately reflect the actual volume of audiovisual material that can be found.

The current distribution of content by type (or format) is as follows: 60% images, 38% texts, 1% sound recordings, 1% videos, and 0.05% 3D objects.

Type	Total number of objects	Percentage of the total
Image	24,159,765	58.03 %
Text	16,225,986	38.97 %
Sound	520,196	1.24 %
Video	701,811	1.68 %
3D	21,262	0.05 %

Table 4: Overview of the types of content in the Europeana repository on the 16th of April 2015.

As is clear, images and text make up the majority of the dataset, while sound, video and 3D material only makes up a very small portion of the total.

At the moment more than 1.2 million audiovisual records are available via Europeana. In 2015 we expect this number to greatly increase. Both EUScreenXL and EFG are partners of the Europeana DSI and will continue to work on providing more audiovisual material. Besides that, the Europeana Sounds project started in 2014, and aims to deliver another 500,000 sound recordings to Europeana in 2015. We therefore expect that, a year from now, the total number of records will be greater than 2 million and in line with the Commission recommendation on the digitisation and online accessibility of cultural material and digital preservation.²⁰ As mentioned above, the number of archival records will also greatly increase.

¹⁹ See also EUScreenXL D5.3 - Strategic Recommendations to Increase the Amount of Audiovisual Content in Europeana V1. <https://my.alfresco.com/share/s/VuUbxmEYSU6PIAifcU2hvw>

²⁰ EC Recommendation 2011/711/EU: <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2011:283:0039:0045:EN:PDF>

2.1.7 20th century content

Adding more content from the 20th century has always been a concern for Europeana. Copyright related issues make it very difficult for cultural heritage institutions to open up and submit their data to Europeana. This result in what is often referred to as the '20th century black hole'. Europeana will investigate how to close this gap and help cultural heritage institutions in making more 20th century content available online.

2.2 Qualitative content analysis

The demand for content that can be used directly by both portal and API users has increased greatly. Not only has the number of people with instant access to the internet via mobile devices radically grown in the last 5 years, these users' expectations have also changed when it comes to the quality of what they look for online. Not only should the desired content be easy to find on the web, the information about the object should be accurate and informative, and preferably available in a number of languages. The digital object should preferably be of high resolution, re-usable and sharable across the web.

This chapter analyses Europeana's content and collections in order to gain insight into the qualitative situation of the repository, and to identify existing gaps.

2.2.1 Implement the Content Re-use Framework

Europeana is evolving from an aggregator of metadata designed to be displayed on a portal, to a platform that facilitates the re-use of cultural heritage as held by the data providers in every shape or form. At the same time technology has evolved, and with it, the expectations for content display, which have increased significantly, both for simple viewing on the portal and for re-use elsewhere on mobile or tablet devices.

These changes force us to re-evaluate our content and display strategy and ask ourselves some fundamental questions: 'is our data fit for purpose?' and 'what should we change as a network of data providers in Europeana to remain relevant in the next 3-5 years?' We know that, in order to work with professionals in the Creative Industries and to make our heritage available on platforms such as Wikipedia, the demands on metadata and content need to be different. This is especially true if we want our cultural heritage to be found, viewed and re-used by visitors to a single-purpose website such as the Europeana Portal. For this reason the Content Re-use Framework has been developed.²¹ This framework aims to make these differences understandable and establishes a clearer relation between what a data provider shares with Europeana, and what the institution can expect in return.

This framework is in line with one of the main strategic aims of the our Strategy 2020, namely to improve the data. The framework presents four different tiers in which content Europeana can fall. The framework presents a model for re-use per tier and acknowledges that the higher the quality is, the more we, and crucially third parties, can do with it. In order to move content up to

²¹ D3.3: Documentation of the Extension of the Europeana Licensing Framework.

http://pro.europeana.eu/files/Europeana_Professional/Projects/Project_list/Europeana_Creative/Deliverables/eCreative_D3.3_KL_v1.0.pdf

higher tiers, we will actively approach data partners where we see potential and help them, together with the aggregators, improve their collection.

2.2.2 Media File Checker

In the course of 2015 the Media File Checker that was developed as part of the Europeana Creative project will be implemented. This tool will allow the users not only to search on metadata fields, but also on the technical aspects of the digital content, for example colours and resolution size. This will make it much easier for a user to find high quality material that can be re-used for their specific purposes, and easier for the data providers to have their material found by users.

Because the media file checker follows the link given by the data provider to the digital object on their website, this will also give us real time information about the amount of links that are not resolving. This will make it a lot easier to detect issues and address them with the aggregator.

2.2.3 Increase the number of objects in the database that can be freely re-used

In order to stimulate creative re-use of its content, Europeana had set a target of 8 million open-licence labelled objects (PDM, CC0, CC-BY, CC-BY-SA) to be reached in 2013.²² This has been achieved and currently more than 13 million objects are openly licensed in Europeana. Most of this material falls into the public domain due to its age and we actively encourage data providers to label this material as such and to not claim new copyright on digital representations of a work that is out of copyright. This is in line with the public domain charter²³ that was published in 2010 by Europeana.

It is more difficult to increase the amount of records with a Creative Commons license. This is due to the fact that only the rights holder can apply these licenses on their creative works. As institutions are very often not the rights holder of the material, it is not up to them to choose these licenses. It is also worth acknowledging the fact that a lot of material is simply still in copyright and can therefore not be opened up at this stage.

2.2.4 Channels and Thematic content

We want to reach more users and engage more deeply with the users of our portal, and to do this in collaboration with our network of data providing partners. In order to achieve this, we need to match our (partner's) content to identified interests and communities via specific thematic services. The Europeana Channels are therefore introduced in the Europeana Business Plan 2015. Channels are segmented sub-portals made available on the Europeana platform. Segmentation is based on themes of content and within a channel the user can search and browse within that theme of content specifically. Users will have access to advanced browsing capabilities based on thematically appropriate types, subjects and authorities (creators). Channels will be skinned to differentiate them from both the generic portal and other channels and can also offer functionalities specific to the theme of the channel.

²² *Europeana Business Plan 2013*, Key Performance Indicator 11.

²³ Europeana Public Domain Charter. <http://pro.europeana.eu/publications/the-europeana-public-domain-charter>

MS2: UPDATED PARTNER DEVELOPMENT PLAN, MS4: UPDATED COLLECTIONS AND DATA DEVELOPMENT PLAN, MS5: IMPROVED QUALITY OF METADATA (PARTNER AND DATA DEVELOPMENT PLAN)

In 2015 Europeana will work on the Art History and the Sounds Channels. Besides developing these two sub-portals, we will pro-actively work to get more high quality material in the Europeana database that would fit the channel. Therefore additional attention and support will be given to institutions that hold relevant material for the channels. This can either be an existing data partner and we will help them to improve their collection, or new partners that want to make their data available to the web.

In tandem, the Europeana Foundation will run a campaign with the ministries from all Member States that has the goal to make more important artworks available from each country. The Member States are asked to select 10 important pieces of art from their country and to submit them to Europeana. This total of 280 high quality artworks will be used to promote the launch of the Art Channel. The Art and Sound Channels are expected to be launched in early 2016.

3. Metadata

This chapter analyses the current status of the metadata in the Europeana repository by looking at several distinct but interrelated elements. Metadata are an important component of the user experience as they provide the user with information on, for instance, the author and subject of the work, the time and place of creation, and how the object can be (re-)used.

The first priority of the Europeana Strategy 2020 is to improve this data. As mentioned earlier, the Europeana Network did a great job collecting an incredible amount of records that are now available in Europeana. But we must also acknowledge that it is not always easy to find the material in a database with more than 42 million records. In order to improve this, the Europeana Foundation will work with the network and, as such, has made this a very explicit goal in the DSI program. All of the aggregating partners are asked to provide training and support to their data partners to make sure their metadata is published to the web in the best way possible. In 2014, we started work with the metadata quality Task Force, which researched how best to define what makes good quality metadata.²⁴

In the DSI plans, we focus less on the amount of records that the different partners will provide than in the previous years. We acknowledge that in order to make the Europeana database more fit for the current user needs, we need to improve the quality of the data first. For the next stage of our work, we will therefore aim for quality over quantity.

3.1 Previews

Findings from user research and experience with user engagement events indicate that the visual representation of the objects in the Europeana repository (the previews) has become increasingly important for the user experience.²⁵ The quality of previews mainly depends on the size of the preview, which is not always large enough to provide a preview of sufficient quality. Moreover, many objects do not have previews at all. It has proven difficult to establish exactly how many previews are available in the repository, as the current ingestion process does not record the existence or successful generation of previews.

At the moment there is no reliable way to check the exact amount of missing previews in the Europeana database. This will all change when the media file checker is implemented which will generate the thumbnails. Once this is implemented it will become clear where the gaps are and allow us to proactively approach data providers to solve this. The new method will also be much more thorough when generating the thumbnails. That said, for some types of content, in particular sound, generating a preview is more problematic as there is no visual object.

²⁴ Metadata Quality Task Force Report - pro.europeana.eu/publication/europeana-task-force-on-metadata-quality-report

²⁵ The term 'preview' refers to *a reduced size or length audio and/or visual representation of content, in the form of one or more images, text files, audio files and/or moving image files* (descriptions from the Europeana Data Exchange Agreement (DEA), article 1.

3.2 Direct links to content

Europeana is moving from being a single access point to cultural heritage via its portal, to a platform of cultural heritage data for others to build upon. The portal is only one representation of the database, and not fit for every user. If you are, for example, a manuscript scholar, you would need more elaborate tools than the Europeana portal can offer you, such as the means to annotate. Furthermore, it is unlikely you are interested in all 42 million records, but rather the ones relevant to your research interest. The Europeana API offers users and creative industries the opportunity for this. The material needed can be queried and new platforms can be build on top of that. For API users, it is essential however, that there is a direct link to media object provided. When an image is embedded in an html page or a viewer, it becomes impossible to gather digital objects from different sources together into one platform. At the moment, little over 50% of the metadata includes a direct link to the the content. While some institutions deliberately choose not to make direct links available in Europeana in order to redirect more traffic to their own institutional website, we also see a lot of institutions that want their material to be used by third party apps. We will therefore help them to map the direct links to the right metadata field (EDM:IsShownBy) to allow their material to be used.

3.3 Native EDM

In 2013 Europeana moved from the Europeana Semantic Elements (ESE) metadata model to the Europeana Data Model (EDM). This new model allows providing institutions to deliver richer metadata to Europeana. After the introduction of EDM, ESE was not instantly abolished as providers need time to make the internal transition to EDM. When ESE is delivered, Europeana maps this internally to EDM, but this will obviously never be as rich as native EDM delivered right from the source.

In the next period we want to move all data providing institutions to EDM. This is done by providing documentation and training to the aggregators. A big step forward was achieved when our biggest provider, The European Library, migrated completely to EDM. Besides that, all newly started projects have committed to supplying native EDM. We expect to have at least 25 million records in Europeana delivered as native EDM by the end of 2015.

3.4 Rights labelling

At the start of 2013, a significant share of all objects in the repository did not have a rights label, or had an incorrect one. In order to give the user a proper understanding of what is and is not allowed with the digital object, a rights statement is essential. We therefore set ourselves the goal to have no more unlabelled material in Europeana by the end of 2014. In order to achieve this, a Rights Labelling Campaign was launched in January 2013 to raise awareness of the importance of providing metadata with rights labels, but also to address datasets with missing or incorrect rights statements. The campaign was very successful, helping to increase the number of objects with a rights statement in their metadata from 51% in 2013 to 100% by the end of the campaign. In addition, to ensure that new objects or datasets will not be published without this information, rights labelling was made a mandatory field (edm:rights).

At the moment, every record in Europeana has a rights statement. At the same time, we acknowledge that the correct rights statement is not always used by the institution. We will therefore actively work together with the DSI partners to improve the level of competence of data providers when it comes to intellectual property rights. This will be done by 'training the trainers'. Europeana will organise a number of workshops together with the aggregators around copyright so the aggregators can confidently support and train their data providers. This should lead to a higher level of accuracy in the Europeana database as a whole.

4. Partners

This chapter looks at how Europeana will continue to develop its network of data partners. At the moment, over 3,300 cultural institutions from across Europeana have made (parts of) their collection available via Europeana. This great achievement could not have been done without a dedicated network of aggregators and projects.

In 2015 we will continue to grow the number of partners, but mostly make sure that we nurture and improve the relationship with the existing ones. This will mean a strong focus on providing the appropriate assistance to improve the quality of the data, but also, for example, improve our own aggregation infrastructure.

4.1 New partnerships

At the moment we have around 3,300 contributing institutions. 98% of these institutions do not contribute to Europeana directly, but via a national, domain or thematic aggregator. When a new potential data partner approaches us we look for the most appropriate route to share their data via Europeana. In this process we prefer national or domain aggregators over thematic, often project-based, aggregators as they are likely to be more sustainable.

In 2015 Europeana will continue to grow the amount of data partners that share their collection. Besides serving as a 'broker' between the institution and the aggregator, we will also take a more active approach towards institutions that are not yet working with Europeana. Previously, Europeana has been fairly neutral about the content being delivered. While we plan to remain inclusive and open to everybody and every type of cultural heritage collection, we will also focus our attention more on institutions that hold material that fits in the themes that have been prioritised. Together with the aggregators we will try to define potential partners and actively approach them to see if there is an interest in sharing their collection via Europeana. This way we will improve the offering to our end user in the thematic channels that will be launched by the end of the year.

4.2 Existing partnerships

We will foster the relationships with current partners in a spirit of cooperation and mutuality, emphasising the benefits of the partnership for each individual member and the ecosystem as a whole.

As mentioned earlier, Europeana has received data from 113 different providers, the majority of the material is from a few providers such as The European Library, the German Digital Library and Gallica. Without the network of aggregators that Europeana currently has we would not have come close to what we have achieved today. In order to ensure a sustainable relationship with the domain aggregators and thematic aggregators, many of them have become part of the DSI program.

In order to guarantee a more sustainable future for the national aggregators we count on the Member States. Only by having support from a local ministry of culture can a national aggregator function properly. We will keep actively supporting national aggregators by engaging in

discussions with the different Member States. It is important to stress that national aggregators are essential for not only Europeana, but for the European cultural heritage sector in general.

4.3 Events and knowledge sharing

Continuing our practice of the past years, we will facilitate interaction and the sharing of knowledge experiences through several meetings for and with our partners, in particular the meetings of the Aggregator Forum.

Over the years the Aggregator Forum meeting has grown from a relatively small get-together, to a full three day event with over 50 participants. It has become a crucial event where the aggregators can share their experiences, learn from each other and decide on next steps to be taken by them and Europeana. In 2015 we will therefore organise not just one, but two meetings. One is organised and hosted by the Europeana Foundation and will mainly be used to discuss upcoming plans and listen to the ideas of the aggregators. Later in the year one of the national aggregators (Cultura Italia in 2015) will organise another meeting, where Europeana will be the guest.

Besides the event we have also grown the virtual space of the Aggregator Forum where we share information with the aggregators. In 2014 we were able to have a number of fruitful discussions and new topics were also brought in by the aggregators itself. In 2015, for the first time, we held a virtual meeting with a number of aggregators to discuss a topic. This received good feedback and therefore the aggregation team will initiate more of these throughout the year when requested.

4.4 Aggregation infrastructure

The Europeana ecosystem depends on a network of national, thematic and domain aggregators which brings 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. 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 120 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 survey²⁶ agreed that the current aggregator model is slow, intransparent and chaotic. It can also lead to duplication and data loss.

For this reason one of the key tasks is to innovate the aggregation infrastructure in order to create 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.

²⁶ D.1.1: Recommendations to improve aggregation infrastructure

http://pro.europeana.eu/files/Europeana_Professional/Projects/Project_list/Europeana_Version3/Deliverables/EV3%20D1_1%20Aggregation%20Infrastructure.pdf

In 2014, the first attempts were made by consulting the network of aggregators, and various partners, about this. Early in 2015, we did more research via the aggregator survey. This has led to a clear overview of their needs and wishes of these partners. In March 2015 the first set of recommendations to improve the aggregation infrastructure were published. There, the different issues are further explored and solutions are proposed. In the coming years the Europeana Cloud will become more developed and, combined with the development of a set of aggregation tools that work on top of the cloud, also known as Metis, it will become much easier for data providers and aggregators to not only share their data, but also to edit, enrich and export it to the platforms that fit their needs.

As part of the Europeana DSI, 2015 will be used to further explore these recommendations and start working on a first working prototype of the Europeana Cloud and Metis. The results of this will feed into the roadmap for the innovation of aggregation infrastructure that will be published in early 2016.

5. Conclusions

In this final chapter, we will draw conclusions from the analysis and make recommendations for future strategies. As we have seen, most of the goals that were set in the strategic plan 2011-2015 and the previous partner and collection plan have been achieved. The Europeana database has grown substantially and big steps have been made by individual Member States. In a number of countries the national aggregators have reached a mature state and have become an essential part of the Europeana network. We have also welcomed a couple of new aggregators and a lot of European countries are well on their way to reach the target that was set by the MSEG group.

At the same time we must acknowledge that not all goals have been reached. For a number of countries it has not been possible to get support from the ministries of culture which has resulted in some discernible gaps. In order to create a truly European database of cultural heritage it is important that the material from these countries becomes online available as well. Europeana will therefore keep advocating for stronger support from these Member States.

In the last couple of years Europeana has shifted its main focus from quantity to quality. We acknowledge that in order to provide a usable product to our end users the quality of the data is essential. Having surpassed the 2015 goal of 40 million records already this allows us to focus more on this quality aspect. We will therefore put more effort into helping the data providers and aggregators to reach a higher level of data quality. The first step is to understand what good data quality actually means, and this has been research by the metadata quality task force. The next step is to then implement these recommendations in practice, and support our network to do so too. This is not an easy task but we believe that Europeana is well positioned to liaise with multiple parties and share different efforts happening around Europe to ensure we can learn from each other, rather than each trying to figure it out alone.

Finally, we as Europeana will start focusing more on thematic content. This is the direct result of consultation with our end users and the network about their requirements. We believe that in order to give a better user experience and to better surface the content of our partners, we should work on well described and curated datasets for our users. For this reason the Content Re-use Framework has been developed which describes in more detail how institutions can share their content via Europeana and what Europeana can do with that material in return. We are looking forward to the next phase developing Europeana into a truly networked organisation which delivers access to high quality metadata and content in an as open way as possible.