



**Project Acronym: Europeana v2**  
**Grant Agreement number: 270902**  
**Project Title: Europeana Version 2**

**D3.8: Content Strategy Update**  
**MS12: Content Acquisition & Development Plan**  
**Merged and renamed to:**  
**Collections and Metadata Analysis, Strategy and Plan 2013**

|                            |                                 |
|----------------------------|---------------------------------|
| <b>Revision</b>            | Hans de Haan                    |
| <b>Date of submission</b>  | 29 March 2013                   |
| <b>Author(s)</b>           | Annette Friberg, Harry Verwayen |
| <b>Dissemination Level</b> | 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            | 19-26 March 2013 | Hans de Haan    | Europeana    | updating, revising |
| 2            | 25 March         | Annette Friberg | Europeana    | review             |
| 3            | 26 March         | Beth Daley      | Europeana    | editing            |

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This deliverable contains original unpublished work except where clearly indicated otherwise.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

|  |           |
|--|-----------|
| <b>1. Executive summary .....</b>                        | <b>4</b>  |
| <b>2. Introduction .....</b>                             | <b>6</b>  |
| <b>3. Content and collection analysis .....</b>          | <b>8</b>  |
| 3.1. Total number of objects .....                       | 9         |
| 3.2. Content by member state .....                       | 10        |
| 3.2.1. Contributions from Non-EU countries .....         | 14        |
| 3.3. Content by domain and providers .....               | 15        |
| 3.4. Content by type .....                               | 17        |
| 3.5. Representation of content by year of creation ..... | 20        |
| 3.5.1 20th century content and orphan works .....        | 21        |
| 3.6. Masterpieces .....                                  | 23        |
| 3.7. User-generated content .....                        | 23        |
| <b>4. Data analysis .....</b>                            | <b>24</b> |
| 4.1. Previews .....                                      | 25        |
| 4.2. Descriptions .....                                  | 25        |
| 4.3 Rights labelling .....                               | 25        |
| 4.4. Accuracy of links .....                             | 28        |
| 4.5 European Data Model .....                            | 29        |
| 4.6 Metadata Licence .....                               | 29        |
| <b>5. Conclusions and recommendations .....</b>          | <b>30</b> |
| 5.1. Content and collections .....                       | 31        |
| 5.2. Europeana Cloud .....                               | 40        |
| 5.3. Specific actions .....                              | 41        |
| <b>References .....</b>                                  | <b>42</b> |

# 1. Executive summary

As stated in the *Europeana Strategic Plan 2011-2015*, Europeana aims to be the ‘trusted source’ of European cultural heritage information. In an increasingly crowded landscape of digital cultural heritage providers (such as Google, Wikipedia, etc.), Europeana and its partners offer users very specific and unique qualities such as authenticity and trust. As we move from a ‘supply-driven’ structure towards a more ‘demand-driven’ structure in which accessibility is key, these qualities will become the differentiating factors of our collective value proposition to end-users. Indeed, Europeana’s central tenet is to facilitate the creation of a landscape in which users can find what they want, when and where they want it.

With well over 26 million objects in the repository right now, it is possible to say that we have assembled a critical mass of cultural artefacts. But clearly, if the aim of Europeana and its partners is to represent the full breadth of European cultural heritage, this is only the beginning. We will therefore continue to increase the amount of objects in the database with the aim of including over 30 million objects by 2015. But that by itself is not enough. In parallel with the quantitative objectives, improving the quality of the database remains of crucial importance. End-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, the information about the object has to be accurate and informative, of high resolution, re-usable and sharable across the web.

In order to achieve this, a lot of work remains to be done by Europeana and its contributing partners.

This study draws the following conclusions and recommendations for the further development of the Europeana repository:

## Content and Collections

- **Increase in objects:** with the advent of new businesses operating on large datasets, the increase in total objects remains a key goal. In terms of progress made against targets, we are on the right track but fostering the aggregation and projects landscape carefully remains a crucial priority. We intend to pursue the strategies as planned (see chapter 5 for conclusions and recommendations).
- **Attention to representation by EU member state:** focus on the contribution by member states has recently resulted in more equally distributed representation of

numbers of objects per country. Targets for the representation per country have been set (see table 1 for the estimates per country by 2015). Therefore, the focus will be on supporting member states to reach the target recommended by the European Commission (2011/711/EU).

- **Ensuring good representation for each of the domains:** all domains (galleries, libraries, archives, museums, audiovisual collections) are represented but the distribution seems uneven across the domains with the archives lagging behind. Investigate what can be done to improve this situation.
- **Ensure that content from all ages is represented:** analysis shows that early modern history is not as well represented as it could be, especially as content from this period is all in the public domain. Focus on getting more of this important material in. Also investigate what can be done to increase the amount of 20<sup>th</sup> century 'orphan' works in the repository.
- **Ensure inclusion of user-generated content (UGC):** ensure the inclusion of UGC without compromising the mandate of Europeana to create a trusted source of information. Ensure close cooperation with the 'UGC' Task Force of the Europeana Network.
- **More focus on audiovisual (AV) content:** AV material currently makes up less than 3% of the database, while research shows that this material gets most attention from end-users. Actively pursue both small and large institutions to contribute in order to reach the goal of two million AV objects by 2015.
- **Ensure the inclusion of masterpieces:** masterpieces are difficult to define, but the inclusion of the most important works from each country (in particular the ones in the public domain) is an expectation of most users. Work with the Member States Experts Group and Aggregator Forum in particular to raise awareness of the importance of contributing masterpieces in each member state.

## Metadata

- **Increase the amount of previews:** it is estimated that currently 75% of the objects in the repository have previews. Although this is still too low, it does show a 25% increase over the 2012 number. The ingestion team is working with partners to increase these numbers further. The target is to get 90% previews by 2015.
- **Decrease the amount of dysfunctional links:** the amount of 'dysfunctional' links is 3% on average. Although this percentage has not changed, it is, in fact, a success considering the enormous growth of objects. We are certain the number will decrease to the target percentage of 1% before 2015. In order to improve the user experience,

this number should at least stay stable but preferably decrease to less than 1%. Investigate possibilities for permanent identifiers.

- **Improve the amount of objects with rights labels:** currently less than half of the objects have rights statements in their metadata and a significant proportion are mislabelled. With the introduction of the Europeana Data Model (EDM), rights labelling became a mandatory field (edm:rights), but a concerted effort is needed in order to attach rights labels to all objects. A 'Rights Labelling Campaign' has been launched to improve the presence and quality of the rights statements. Currently, 30% of objects are missing rights statements, and 27% of objects are marked with open licences (PDM/CC0/CC-BY/CC-BY-SA).
- **Manage the transition to EDM and CC0 for metadata:** in the autumn of 2012, the Data Exchange Agreement (DEA) took effect, releasing Europeana's metadata under CC0 (Creative Commons Zero Public Domain Dedication). The implementation of EDM and the conversion of the Europeana Semantic Elements (ESE) data is now taking place. Data received from partners is already being ingested through EDM in several projects, helping to fine tune the EDM. We plan to have the ingestion process ready to handle EDM objects before July 2013. In March 2013, the new Europeana portal, based on EDM, was launched. In accordance with the implementation of the new CC0 waiver, the Data Exchange Agreement (DEA) was introduced to partners last year. This was a major operation conducted by Europeana during 2012/13 and the full repository of 26.7 million objects is now all covered by the DEA. Interestingly, the Digital Public Library of America (DPLA) is re-using large portions of the EDM for its own system implementation, making it easy to use and re-use both the DPLA's and Europeana's objects.

Conclusions and recommendations from this work will be discussed with partners at the Europeana Network Annual General Meeting in autumn 2013 in order to be validated and included in the *Europeana Business Plan* for 2014.

## 2. Introduction

This report provides an update from the report *Collections and Data Analyses, Strategy and Plan* (D3.7 & MS11) and incorporates results from the recently developed report *D2.2 Partner Strategy and Development Plan, January 2013*. It provides an update on Europeana's content, collections and data repository and follows the planned direction for development. As the update will show, there are many issues that require our attention,

ranging from the equal representation of all European countries to accurate rights labelling in the provided metadata. This report should therefore be seen as a starting point for further discussion with the members of the Europeana Network to establish the order of priorities to improve the Europeana database.

In order to carry out and discuss a collection and data analysis, the terms 'content' and 'metadata' must be clearly defined. Until recently, Europeana did not hold or stock any of the actual digital objects (the content) which belong to the partner institutions. 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.

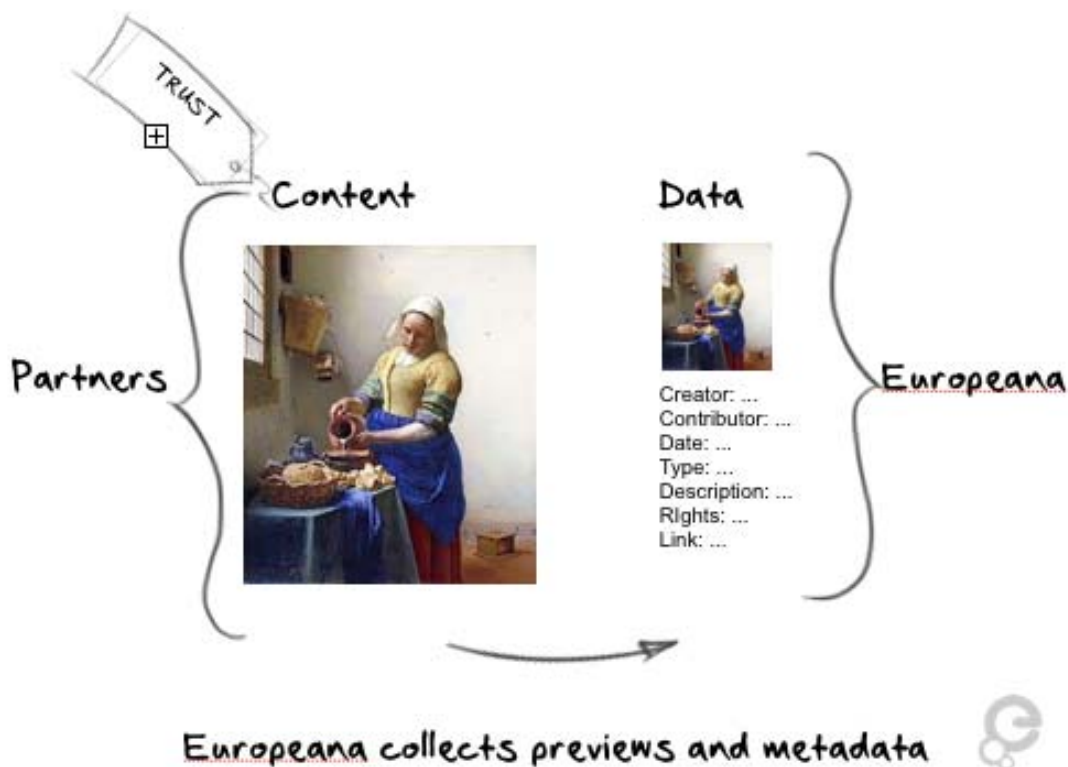


Figure 1: Europeana collects previews and metadata

In 2011, Europeana started collecting user-generated content and added this data to its repository as part of the Europeana 1914-1918 project. This change led to specific opportunities and issues that continue to be dealt with thoroughly in consultation with the Network and following the *User-Generated Content Policy*.<sup>1</sup> However, it does not change the mission of Europeana to make Europe's cultural heritage accessible for all.

<sup>1</sup> See: [pro.europeana.eu/documents/844813/851957/D1.4+UGC+policy.pdf](http://pro.europeana.eu/documents/844813/851957/D1.4+UGC+policy.pdf)

The difference between ‘content’ and ‘metadata’ has implications for the organisation of improvements to the repository. The Europeana Office has, for example, little influence over the content and collections that are made available by the partner institutions (either qualitatively or quantitatively), although recommendations can be made to steer in certain directions. The Europeana Office does, however, have a more direct responsibility to improve the quality of the Europeana repository through the improvement of the metadata, previews and links. Managing both content and metadata effectively will lead to improvements in the user experience. What type of material will users be able to find thematically and in what format (image, audio, video, 3D)? How exhaustive will the material in particular subject areas need to be, and will all countries and languages be represented? Will these users be able to re-use the material, particularly if the original was in the public domain? These are the types of questions that we want to address in this document.

The first part of this report (**chapter 3**) will analyse the content and collections. **Chapter 4** will investigate the metadata, and more specifically the conditions under which the current data is made available in the repository. Finally, in **chapter 5**, we present some conclusions and make some recommendations for the further improvement of the Europeana repository, for further discussion with the Europeana Network and aggregators.

This report is based on several sources,<sup>2</sup> including information displayed in the Europeana portal and information available in previous Europeana reports.

### 3. Content and collection analysis

As stated above, it is important that we acknowledge what we are identifying by the term ‘content’. As described in the Data Exchange Agreement (DEA), the term ‘content’ 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. By ‘metadata’ we mean the textual information (including hyperlinks) that may serve to identify, discover, interpret and/or manage digital objects. The term ‘preview’ refers to a reduced size or length visual and/or audio representation of digital objects, in the form of one or more images, text files, audio files and/or moving image files.

In this chapter, we will investigate specific aspects of Europeana content and collections in order to gain insight into the qualitative and quantitative situation of the repository.

---

<sup>2</sup> See references’ list pages 41- 42





*Figure 2: Content & Collections*

After a series of consultations, Europeana has, in cooperation with its Network partners, decided to start an experiment, creating a data repository in a new way. The Europeana Cloud project kicked off in March 2013 to experiment with a more sophisticated, cheaper and easily sustainable repository using a cloud structure. Content providers and aggregators across the European information landscape urgently need a cheaper, more sustainable infrastructure that is capable of storing both metadata and content. Researchers require a digital space where they can undertake innovative exploration and analysis of Europe's digitised content. Europeana needs to get closer to the target of 30 million items by 2015.

### 3.1. Total number of objects

The total number of objects represented in Europeana has, from the start, been an important metric for the success of the project. Getting these objects has been rather successful. As of March 2013, the number of available objects is 26,787,154 and so the 2013 KPI target of 27m objects has almost been achieved.

Figure 3 below shows Europeana's content growth (number of objects in millions), which has regularly surpassed expectations.

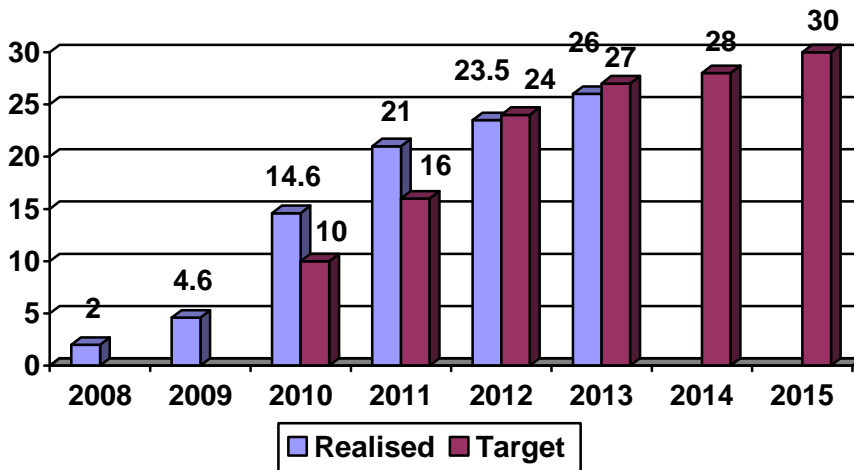


Figure 3: Europeana content development<sup>3</sup> (March 2013)

### 3.2. Content by member state

In total there are 26.7 million objects available in Europeana.eu. The contribution by countries is now more equally distributed. Germany, France, the Netherlands, Sweden, Spain, United Kingdom and Norway provided most of the content to the portal.

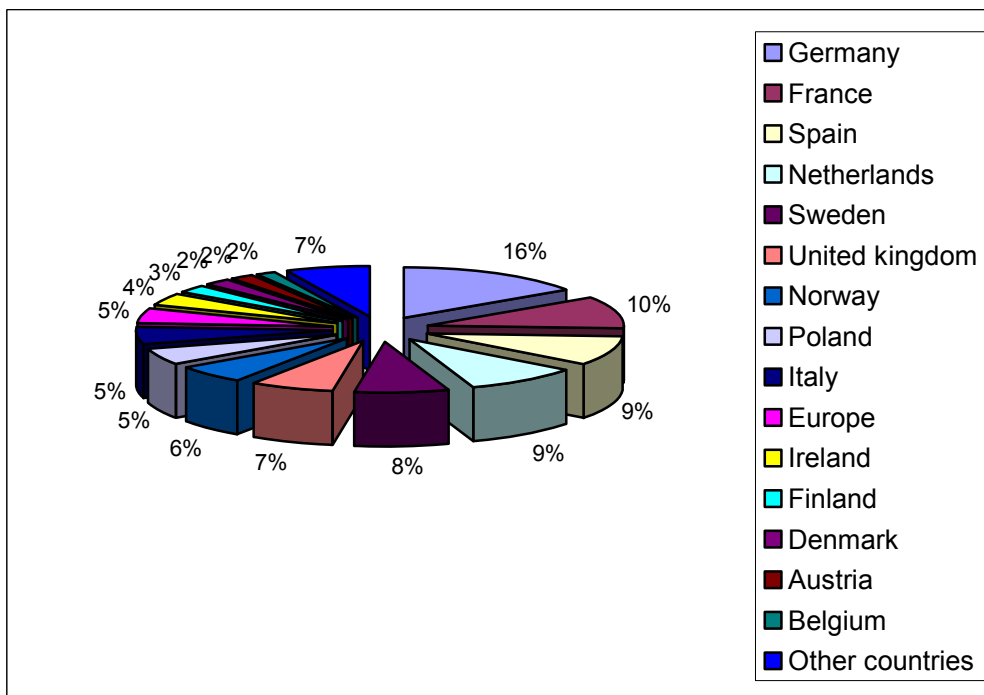


Figure 4: Representation by country of the top 15 contributors to Europeana, March 2013. The field 'Europe' represents data contributed from pan-European projects, where the data is not indicated at country level.

<sup>3</sup> Source: Europeana Partner Strategy and Development Plan, 2013

The *Strategic Plan 2011-2015* stresses the need to make sure that the repository reflects the diversity of our respective cultural heritage. The rich culture of Estonia, Lithuania and Latvia for example, though still underrepresented, has shown a considerable growth, with a total of 170,897 objects by the beginning of 2013. In a recommendation by the European Commission (2011/711/EU) dated 27 October 2011,<sup>4</sup> a list of indicative targets was made for minimum content contribution to Europeana per member state by 2015.

| <b>Country</b> | <b>Material supplied by country, March 2013</b> | <b>2015 estimate</b> | <b>content gap</b> | <b>Gap in %</b> |
|----------------|---|----------------------|--------------------|-----------------|
| AUSTRIA        | 501,914   | 600,000              | 98,086             | 16              |
| BELGIUM        | 424,472   | 759,000              | 334,528            | 44              |
| BULGARIA       | 54,538  | 267,000              | 212,462            | 80              |
| CYPRUS         | 7,591   | 45,000               | 37,409             | 83              |
| CZECH REPUBLIC | 123,842   | 492,000              | 368,158            | 75              |
| DENMARK        | 605,012   | 453,000              | -152,012           | -34             |
| ESTONIA        | 75,956  | 90,000               | 14,044             | 16              |
| FINLAND        | 698,183   | 1,035,000            | 336,817            | 33              |
| FRANCE         | 2,727,349                                       | 4,308,000            | 1,580,651          | 37              |
| GERMANY        | 4,175,464                                       | 5,496,000            | 1,320,536          | 24              |
| GREECE         | 247,633   | 618,000              | 370,367            | 60              |
| HUNGARY        | 336,175   | 417,000              | 80,825             | 19              |
| IRELAND        | 983,503   | 1,236,000            | 252,497            | 20              |
| ITALY          | 1,440,759                                       | 3,705,000            | 2,264,241          | 61              |
| LATVIA         | 36,898  | 90,000               | 53,102             | 59              |
| LITHUANIA      | 78,270  | 129,000              | 50,730             | 39              |
| LUXEMBOURG     | 86,524  | 66,000               | -20,524            | -31             |
| MALTA          | 56,233  | 73,000               | 16,767             | 23              |
| NETHERLANDS    | 2,518,520                                       | 1,571,000            | -947,520           | -60             |
| NORWAY         | 1,563,243                                       |                      |                    | 0               |
| POLAND         | 1,457,931                                       | 1,575,000            | 117,069            | 7               |
| PORTUGAL       | 121,544   | 528,000              | 406,456            | 77              |
| ROMANIA        | 55,868  | 789,000              | 733,132            | 93              |
| SLOVAKIA       | 89,485  | 243,000              | 153,515            | 63              |
| SLOVENIA       | 254,574   | 318,000              | 63,426             | 20              |
| SPAIN          | 2,521,787                                       | 2,676,000            | 154,213            | 6               |
| SWEDEN         | 2,128,680                                       | 1,936,000            | -192,680           | -10             |
| SWITZERLAND    | 126,491   |                      |                    | 0               |
| UNITED KINGDOM | 1,825,601                                       | 3,939,000            | 2,113,399          | 54              |

*Table 1: Content provided per country (status March 2013) in relation to estimated content in 2015.*

<sup>4</sup> Source Europeana.eu, *European Commission Recommendation 2011/711/EU* of 27 October 2011

| <b>Country</b> | <b>National aggregator</b>   | <b>Content gap in objects</b> |
|----------------|--|-------------------------------|
| AUSTRIA        | Kulturpool.at  | 98,086                        |
| BELGIUM        | Erfgoedplus.be   | 334,528                       |
| BULGARIA       | Баpна  | 212,462                       |
| CYPRUS         | SMOEC  | 37,409                        |
| CZECH REPUBLIC | eSbirky (electronic collections)   | 368,158                       |
| DENMARK        | The Royal Library- DK Aggregation Service                                | -152,012                      |
| ESTONIA        | Muis - Museums Public Portal   | 14,044                        |
| FINLAND        | KDK, National Digital Library  | 336,817                       |
| FRANCE         | Culture.fr   | 1,580,651                     |
| GERMANY        | German Digital Library   | 1,320,536                     |
| GREECE         | Hellenic Aggregator  | 370,367                       |
| HUNGARY        | MaNDA (National Digital Data Archive, NDDA)                              | 80,825                        |
| IRELAND        | The Irish Manuscripts Commission   | 252,497                       |
| ITALY          | Cultura Italia (Athena)  | 2,264,241                     |
| LATVIA         | Latvian National Digital Library   | 53,102                        |
| LITHUANIA      | ePaveldas  | 50,730                        |
| LUXEMBOURG     | No Initiative  | -20,524                       |
| MALTA          | Across limits  | 16,767                        |
| NETHERLANDS    | Digitale Collectie   | -947,520                      |
| NORWAY         | Kulturnett.no  |                               |
| POLAND         | Digital Libraries Federation   | 117,069                       |
| PORTUGAL       | Portal arquivos, Instituto dos Museus e da Conservação, RNOD, Cinemateca | 406,456                       |
| ROMANIA        | CIMEC  | 733,132                       |
| SLOVAKIA       | Slovakiana   | 153,515                       |
| SLOVENIA       | dLib.si  | 63,426                        |
| SPAIN          | Hispana  | 154,213                       |
| SWEDEN         | K-samsök (Kringla)   | -192,680                      |
| SWITZERLAND    | Musées lausannois  |                               |
| UNITED KINGDOM | CultureGrid  | 2,113,399                     |

*Table 2: Countries, national initiatives/aggregators, and the gap in numbers between objects supplied to Europeana by March 2013 and target for 2015*

Four member states have already surpassed their target for 2015: Sweden, Luxembourg, Denmark and the Netherlands (Tier 3 countries). A further seven member states, (Austria, Estonia, Ireland, Poland, Slovenia, Spain, Hungary) have reached 80% or more of their targets (Tier 2 countries) and need to deliver at the most 20% in the remaining two years of the EC's five-year period (2011-2015). A further six member states (Belgium, Finland, France, Germany, Lithuania, Malta) have around 20% – 50% remaining. The final 10 Member States still need to supply over half of the content to reach their targets (Tier 1

countries). Based on the March 2013 numbers, the focus will be on helping these 'Tier 1' member states (gap >50%): Bulgaria, Cyprus, Czech Republic, Greece, Italy, Latvia, Portugal, Romania, Slovakia and the UK.

In summary;

Tier 1 countries (gap >50%): Bulgaria, Cyprus, Czech Republic, Greece, Italy, Latvia, Portugal, Romania, Slovakia and the UK

Tier 2 countries (gap between 20%-50%): Belgium, Finland, France, Germany, Lithuania and Malta

Tier 3 countries (gap <20% or over target/no gap): Austria, Estonia, Ireland, Poland, Slovenia, Spain, Hungary, Denmark, Sweden, the Netherlands and Luxembourg.

The three figures below (fig. 5a, b, c) show the gaps between the amount of content provided and the target that has been agreed upon per member state. Three target categories in object supply have been defined to do justice to the variety in member states:

- member states with targets to supply under 400,000 objects
- member states with targets to supply between 400,000 and one million objects
- member states with targets to supply over one million objects.

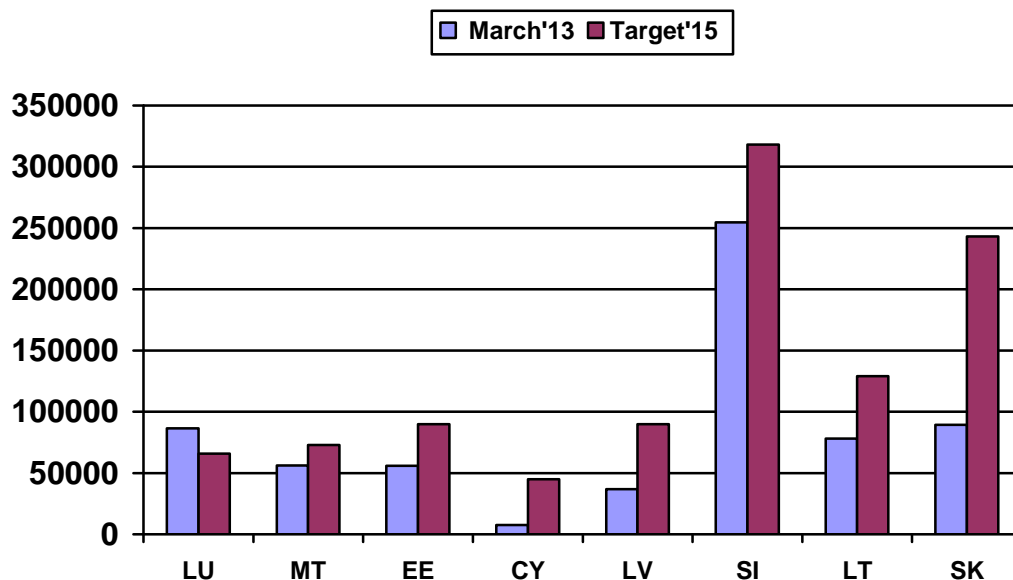


Figure 5a: Member states with a target of under 400,000 objects

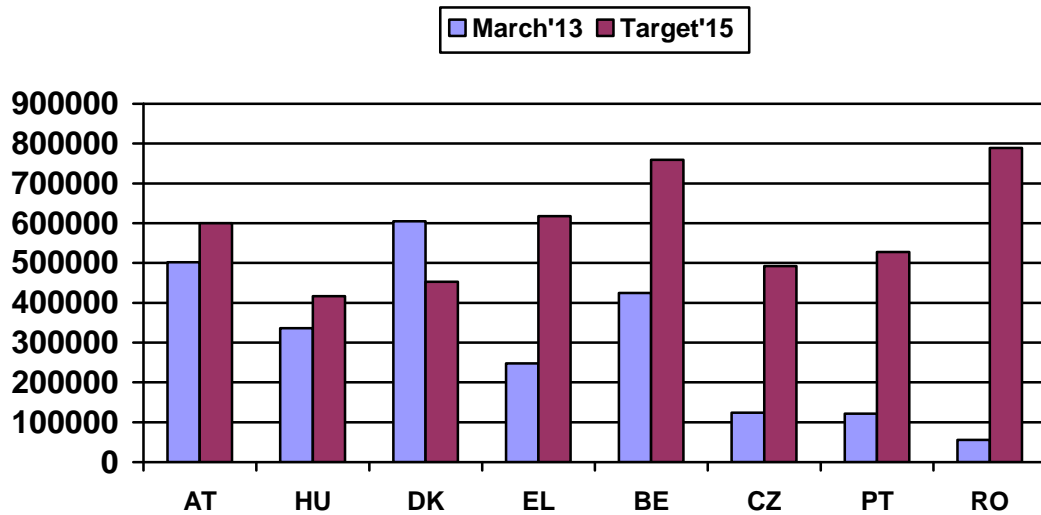


Figure 5b: Member states with a target of 400,000-1,000,000 objects

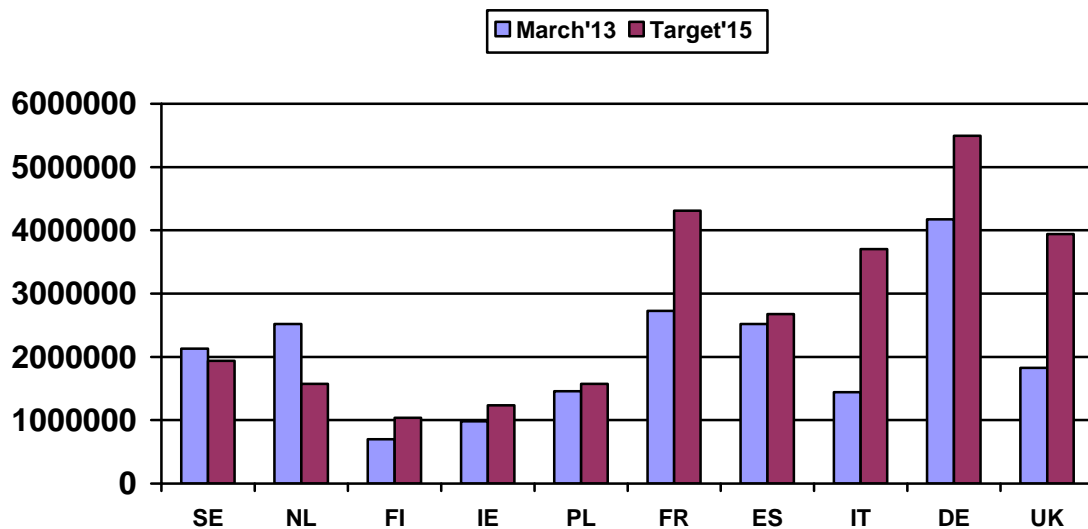


Figure 5c: Member states with a target above 1,000,000 objects

### 3.2.1. Contributions from 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 EU. Norway, Switzerland, Turkey, Iceland, Russia and Serbia are six data-providing countries contributing significantly to the repository.

Together they provide almost two million objects. Norway is the largest non-EU country provider with more than 1.5 million objects (6% of the total number of objects in Europeana) and Switzerland is the second highest non-EU provider with more than 122,000 objects.

### 3.3. Content by domain and providers

Europeana has, at the moment, approximately 132 direct metadata providers. They represent the following domains: galleries, libraries, museums, archives, audiovisual archives, monuments and archaeology. Bringing the wealth of these domains together into one repository is one of the unique aspects of Europeana.

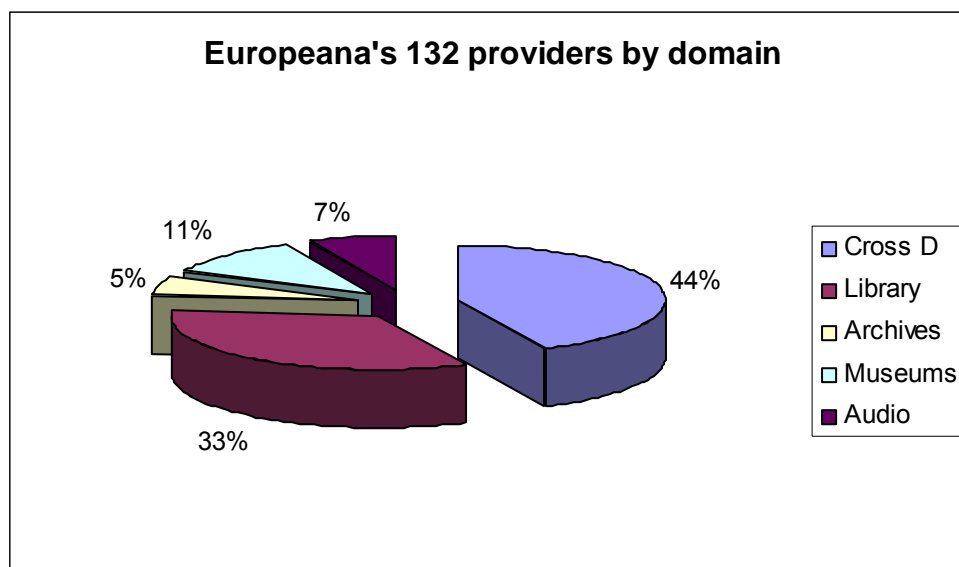


Figure 6: Institution representation by domain

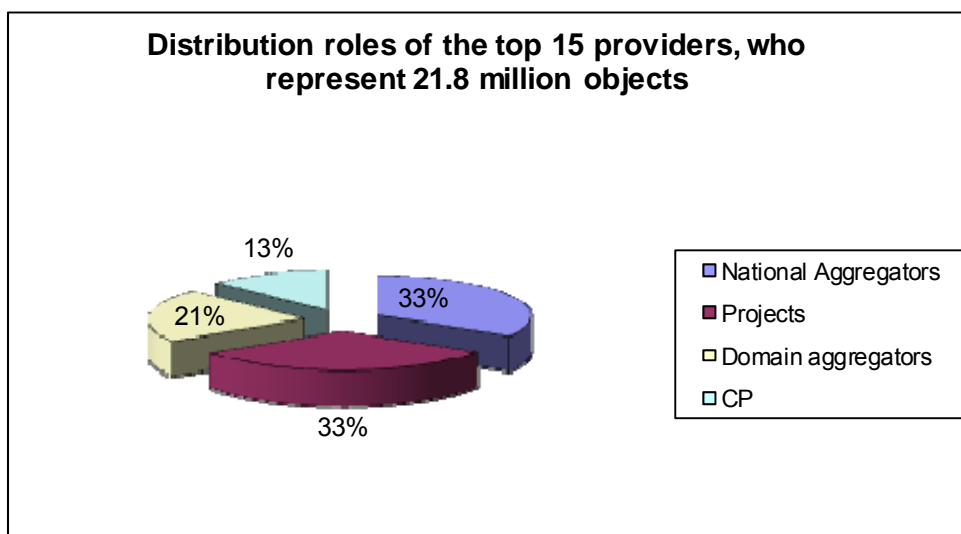
Figure 6 shows that most of the direct content providers are cross-domain institutions. This is a result of member states developing their digitisation institutions and operational structures. As of March 2013, out of the 27 member states, 26 have a national aggregator in place representing either cross or single domains. Two non-EU partners, Switzerland and Norway, have a national aggregator in place (see table 2 for all national aggregators/initiatives). By 2015, all countries are expected to have an operational aggregator.

The *Europeana Partner Strategy and Development Plan 2013* establishes again that Europeana will support all forms of aggregation as it is the only way to make Europeana scalable and sustainable. There is no 'one size fits all' solution for dealing with aggregation. National and domain initiatives are at different stages of their lifecycles. At a national level, each country has different approaches to both digitisation and to ways of collaborating among the four main domains (museums, libraries, archives and audiovisual collections). It is

difficult to encourage best practice for aggregation structures as various factors in each country determine the framework for the national aggregator. A national initiative may result in an aggregator being appointed by a country's Ministry to work with single or multiple domains, aggregating from diverse native cultural heritage institutions.

Projects set up by various single and cross-domain institutions in Europe enable large amounts of content provision to Europeana; they enable aggregation, improve data quality, solve language issues and develop new technologies. These project aggregators increase the speed and volume of results and promote knowledge transfer within a European context, which they can bring back to their own national aggregation initiatives. As stated above, the Europeana Cloud project will try to find new ways to cater for these needs in a cost-effective way.

Countries can supply data by means of participation in aggregating projects (Athena, CARARE, APEnet, EFG and Judaica) or by going through National aggregators or domain specific pan-European aggregator such as the The European Library.



*Figure 7: The top 15 providers represent 21.8 million objects in Europeana. This figure shows their distribution roles; CP stands for direct Content Provider to Europeana (March 2013)*



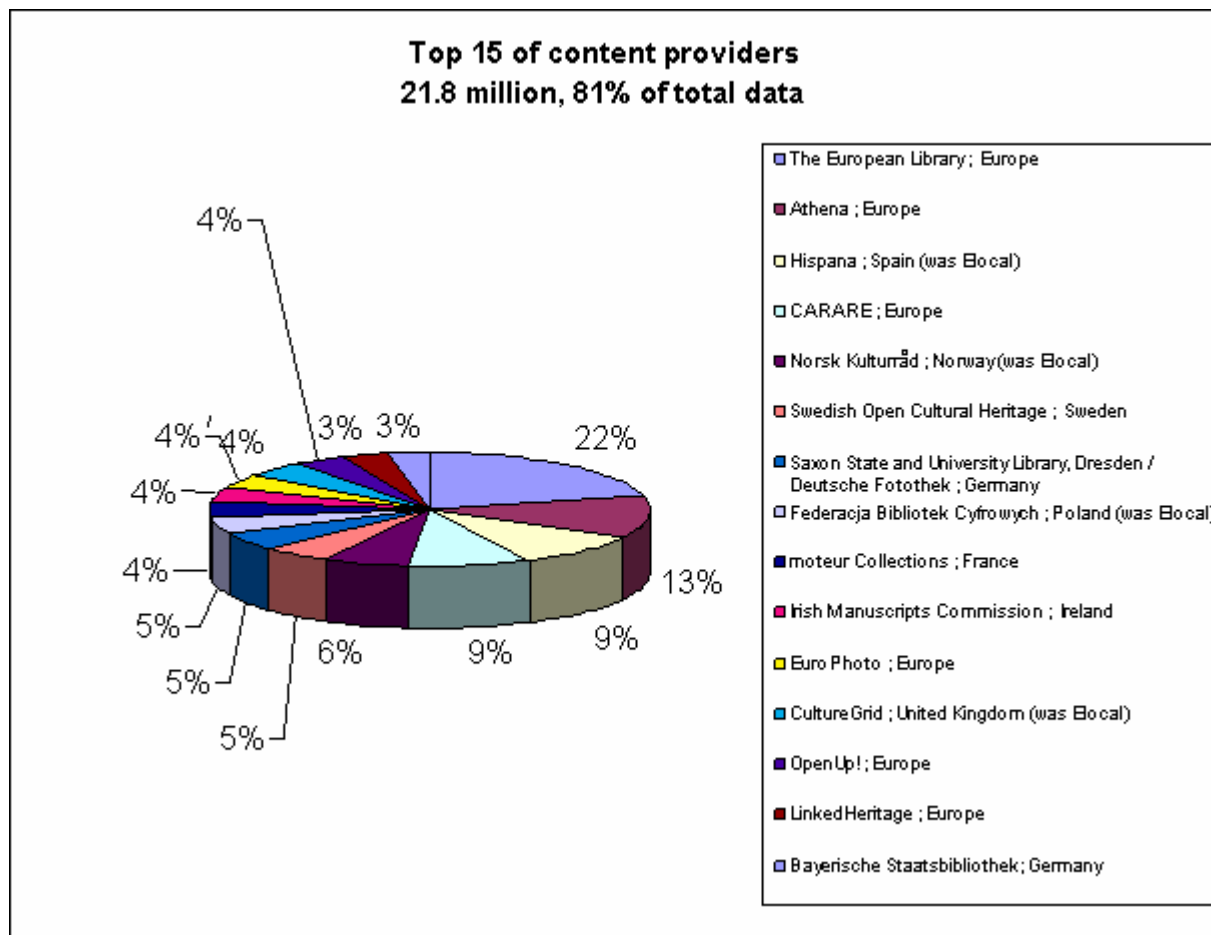


Figure 8: Top 15 provider representation (March 2013)

### 3.4. Content by type

The current distribution of content by type (or format) is as follows: 57% images, 40% texts, 2% sound recordings, and 1% videos.<sup>5</sup>

| Type                    | Total             | Percentage |
|-------------------------|-------------------|------------|
| <b>Total of records</b> | <b>26.787.154</b> |            |
| IMAGE                   | 15,549,838        | 57%        |
| TEXT                    | 10,602,935        | 40%        |
| SOUND                   | 448,653           | 2%         |
| VIDEO                   | 172,195           | 1%         |
| 3D                      | 13,533            | 0%         |

Table 3: Type of objects in Europeana (March 2013)

<sup>5</sup> Source: March 2013, Ingestion Department, internal document

Strong and coordinated efforts have to be made to increase the proportion of audiovisual content significantly to reach the agreed targets. Although content growth is satisfactory in general, we have to acknowledge that the amount of audiovisual content is below the target that has been set.

In the *Business Plan 2012*,<sup>6</sup> the quantity of audiovisual content targeted was a total of one million (550,000 sounds and 450,000 videos). At the moment, the total number of objects is 620,848 (448,653 sounds and 172,195 videos). The objective has been met neither for sounds nor for video. However, one should bear in mind that getting audiovisual objects into Europeana is hard because of the usually complex rights issues involved and the commercial use of most existing material.

Audiovisual archives and audiovisual content are underrepresented in Europeana. Figures 9 and 10 below break down this content.

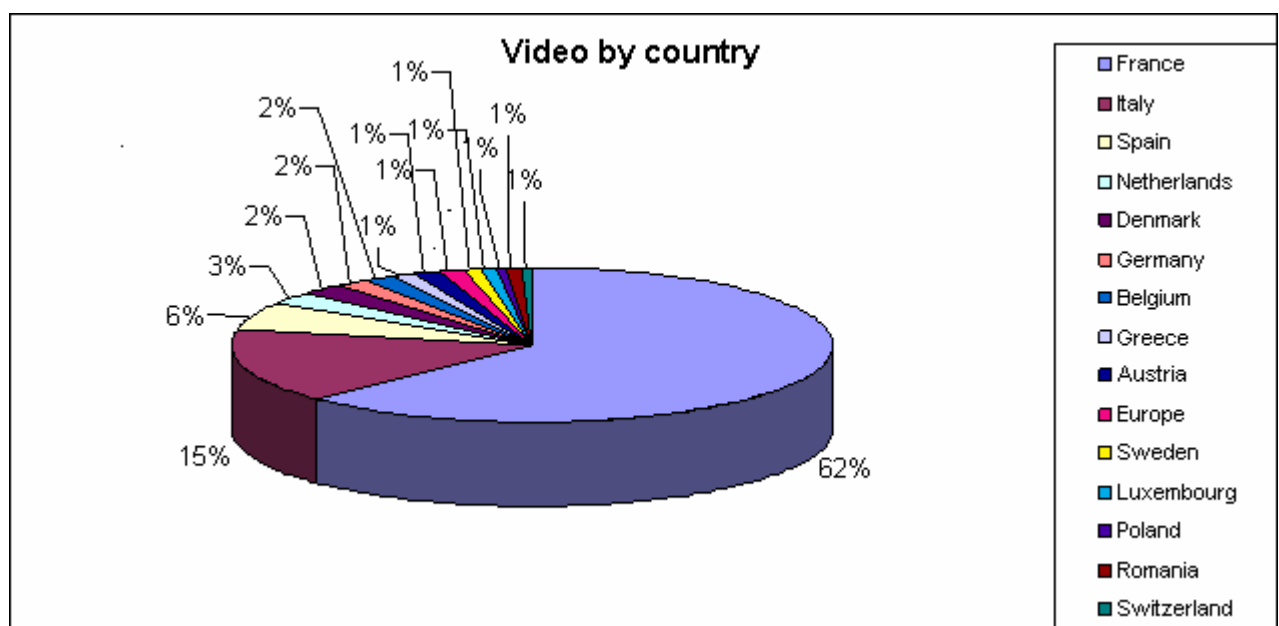


Figure 9: Video content broken down by country

France and Italy are the top providers of audiovisual content.<sup>7</sup> As shown, France provides 63% of the video content, and Italy 15%. The most likely explanation for France providing more than half of the videos is that the French partner, INA (Institut National de l'Audiovisuel), holds rights to a good part of the French public television archives and has a strong will to collaborate with Europeana. As digitisation of audiovisual material is significantly more expensive and labour-intensive than other sources, and the rights issues

<sup>6</sup> Key Performance Indicator number 5, *Europeana Business Plan 2012*

<sup>7</sup> Source: europeana.eu

more complicated, we can assume that adding AV material to the repository will remain a challenge in the near future.

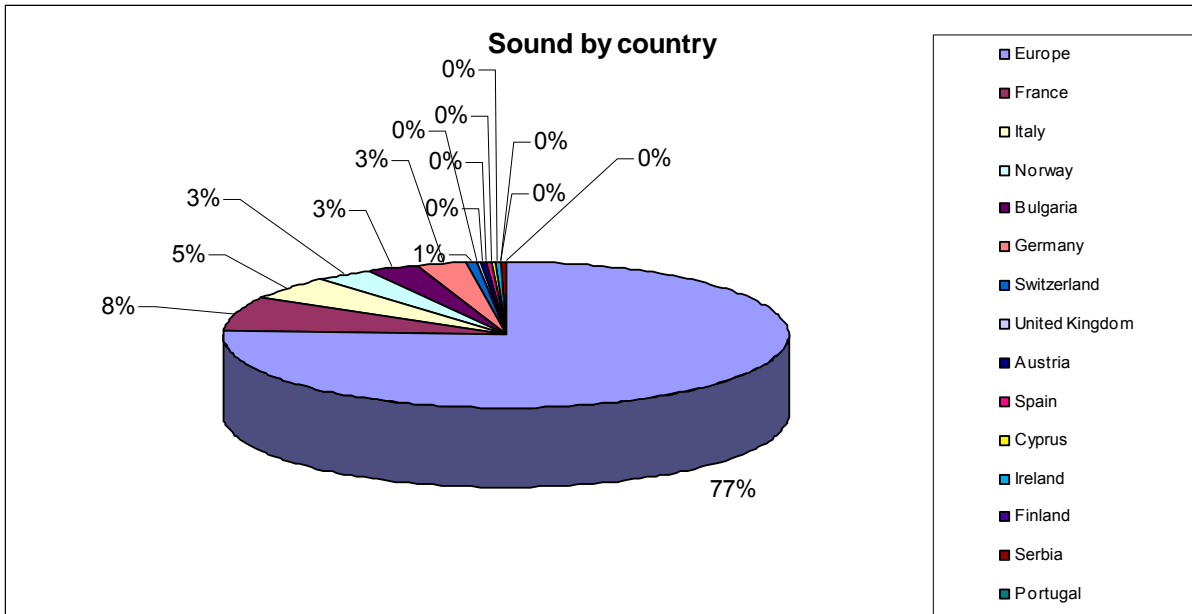


Figure 10: Sound content broken down by country

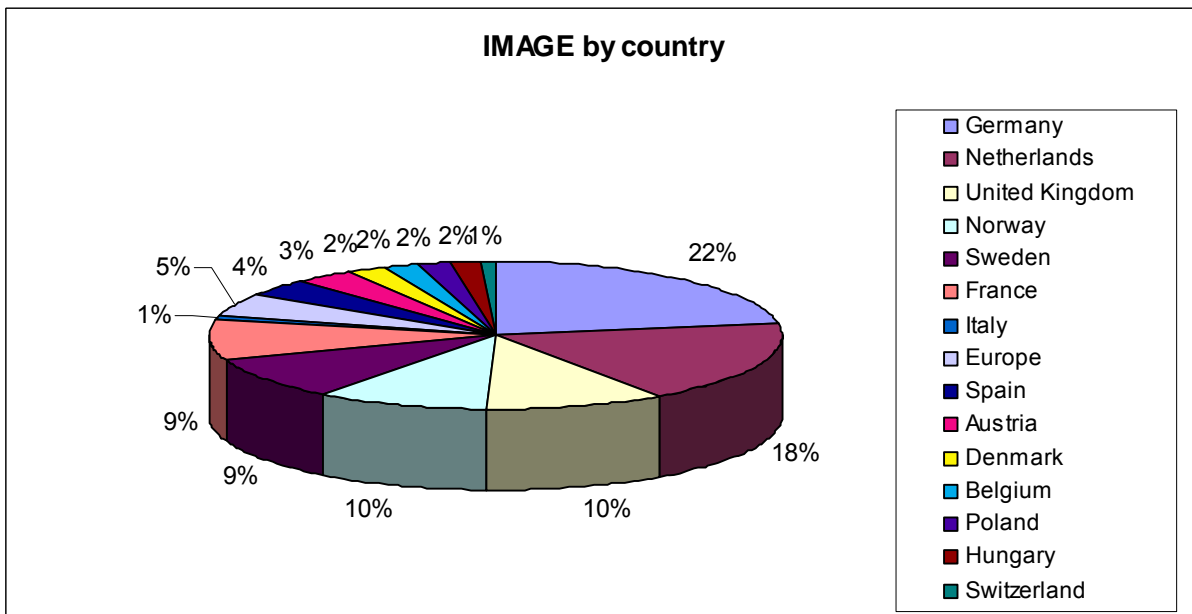


Figure 11: Image content broken down by country

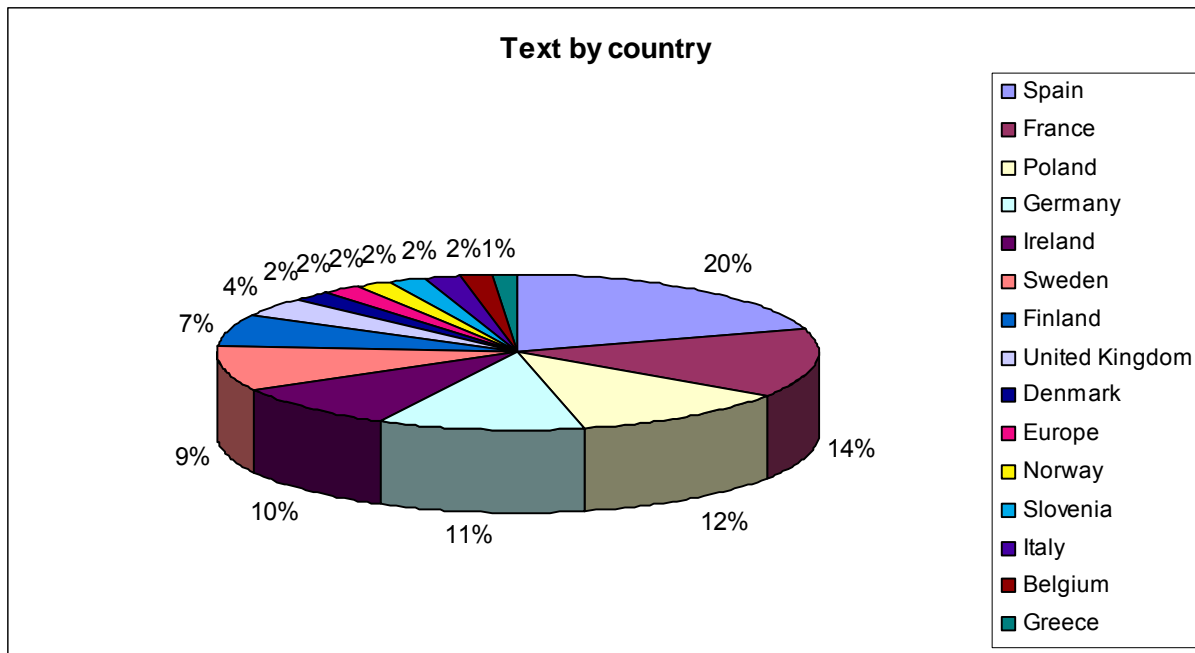


Figure 12: Text content broken down by country

### 3.5. Representation of content by year of creation

The distribution of digital objects in Europeana over the years is another interesting metric to investigate. These numbers have to be interpreted with some reservation because they do not take into account all 26 million objects in Europeana as the information is missing in many cases, and its accuracy cannot be guaranteed. This ambiguity is due to the way data providers state the date related to an object: some use the creation date or the digitisation date, and some use an elusive period of time (e.g. 1974-1987, second part of the 15<sup>th</sup> century). However, these numbers can still be considered as a good indicator of all the content as date information is recorded for 16.3 million objects.

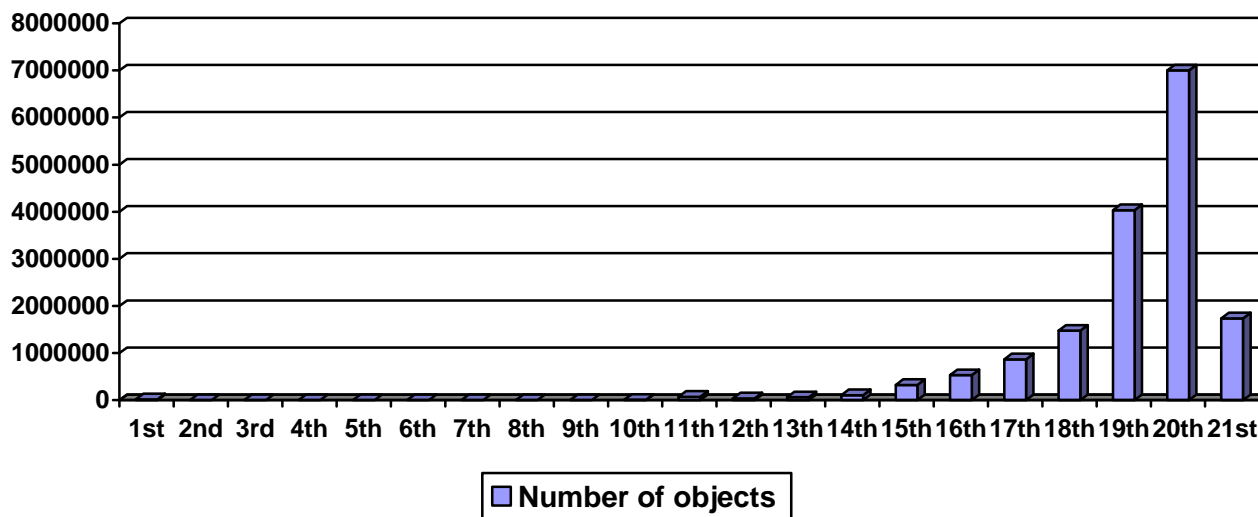


Figure 13: Distribution of Europeana content over centuries<sup>8</sup>

From Figure 13, we can conclude that the best represented centuries are the 19<sup>th</sup> and the 20<sup>th</sup> centuries. Together they represent more than two-thirds of the content for which we have year indications, that is to say about 11 million out of 16.3 million objects. This seems to indicate that the Middle Ages and early modern history (17<sup>th</sup> and 18<sup>th</sup> centuries) are underrepresented, but there are obviously several factors to take into account. The most obvious being that the amount of works produced has grown exponentially over the years.

### 3.5.1 20th century content and orphan works

With regards to items from the 20th century, Europeana is facing a mixed picture. In absolute terms, the 20th century accounts for the biggest amount of items from a single century (c.17 million out of a total of 26 million items in Europeana). However, if we look at the distribution of these items over the 11 decades that make up the 20th and the 21st century, we are presented with another picture:

<sup>8</sup> Source: Europeana year distribution, Technical and Operations Department, internal document

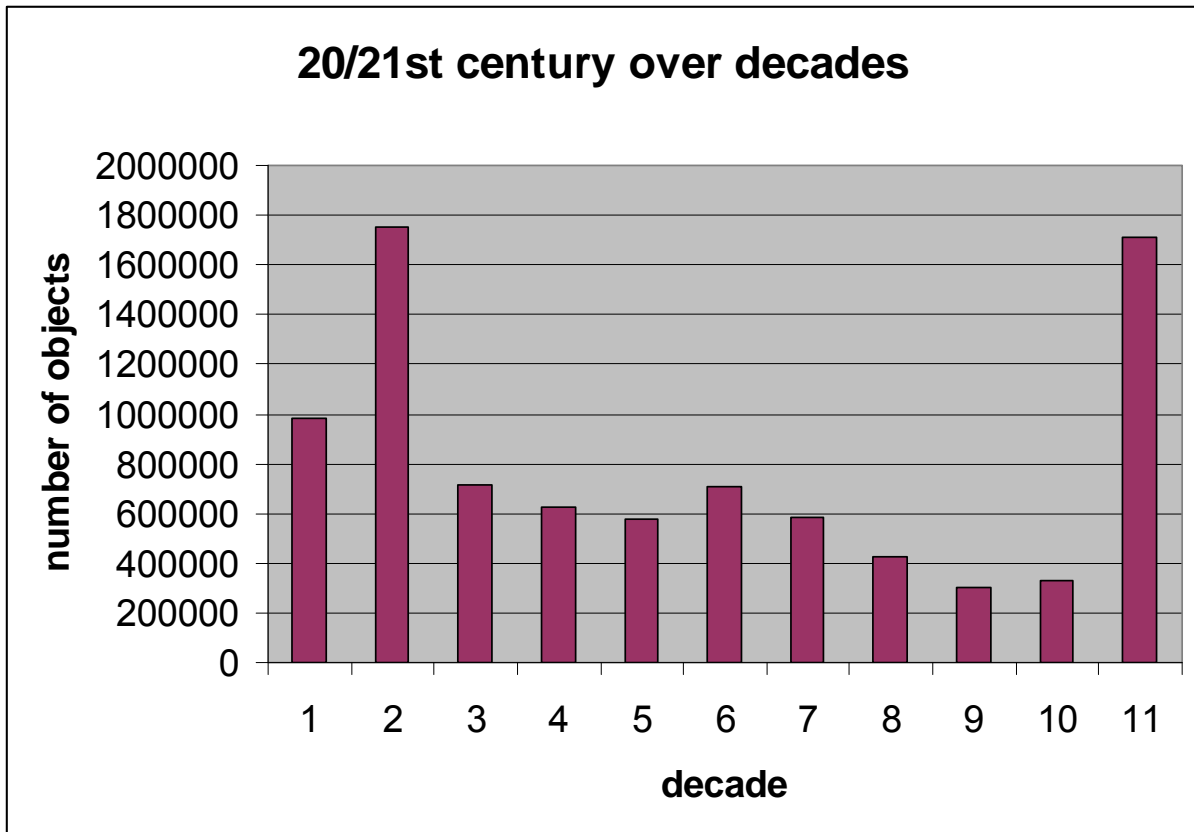


Figure 14: Distribution of Europeana content in the 20<sup>th</sup>/21<sup>st</sup> century, by decade.

This distribution shows that the spread of content in this period is uneven, with the number of items between the years 1930 and 2000 much lower than preceding and subsequent decades. This is at odds with the explosion of creative expression and historical documentation that happened during this period.

The absence of works from this period is sometimes called the '20th century black hole', and is primarily due to the effects of copyright protection. In Europe, the duration of copyright is 70 years after the death of the author or, in the case of institutional or anonymous authors, 70 years after first publication. This means that all works published after 1943 and a large share of the works in the decades immediately preceding the year 1943 are still protected by copyright. Only a tiny fraction of these works is still available commercially. At the same time, Europe's cultural heritage institutions hold large amounts of works from this period in their collections.

### 3.6. Masterpieces

The European Commission made a recommendation to the member states that by 2015, all their public domain masterpieces should be in Europeana.<sup>9</sup>

As mentioned during the ICT PSP (Information and Communication Technologies Policy Support Programme) information day in Brussels on 28 February 2011, a masterpiece can be defined as ‘a work of outstanding creativity, skill or workmanship or piece of work that has been given much critical praise [...] representing legacy of physical and intangible attributes of the past of a group or society’ and ‘an object inherited, maintained in the present and worthy of preservation for the future’.<sup>10</sup>

With such a loose definition, one has to remain careful when referring to the concept of masterpieces. In the foreword of its report, the Comité des Sages sets this rhetorical question: ‘On what grounds would we have the right to lay down selection criteria for what should or should not deserve to be protected?’<sup>11</sup> This question also applies to masterpieces.

Some research is being conducted regarding masterpieces by Europeana. Having considered both the list of suggested masterpieces and the guidelines, our research consists of checking which masterpieces from each country can be found in the repository. The list of suggested masterpieces has an average of 51 works per country, although this varies from 13 for the smallest country to 116 for the largest. Rather than including every great piece of work from the very best-known artists in every country, the list suggests a notable work from a greater number of artists. This way, more attention can be given to artists who are locally very well-known but who do not necessarily have a high profile internationally.

### 3.7. User-generated content

User-generated content (UGC) is playing an increasingly important role in Europeana, as highlighted in the ‘Engage’ track of the *Strategic Plan 2011-2015*. Europeana has, for example, successfully rolled out ‘Community Collections’ projects such as Europeana 1914-1918, in close collaboration with Oxford University. While it is felt strongly that UGC can make an important contribution to the overall quality of the repository, it also foregrounds the tensions inherent in notions of quality, authenticity and with that, trust. In November 2010,

---

<sup>9</sup> Source: *European Commission recommendation 2011/711/EU* of 27 October 2011

<sup>10</sup> [http://ec.europa.eu/information\\_society/activities/ict\\_psp/documents/11\\_theme\\_2\\_call5\\_infoday.pdf](http://ec.europa.eu/information_society/activities/ict_psp/documents/11_theme_2_call5_infoday.pdf)

<sup>11</sup> Source: *The New Renaissance* report, p. 1

Europeana published a policy that laid out Europeana's principles in dealing with UGC.<sup>12</sup> This policy establishes a number of high-level principles such as enabling Europeana users to clearly distinguish UGC metadata from trusted institutional metadata, and the necessity for open licensing of both metadata and content contributed to Europeana by its users.

The terms for user contributions have been developed to achieve three objectives:

1. To ensure that copyrighted content contributed to Europeana by end-users can be used by Europeana and can be published under terms that allow re-use by others such as Wikipedia (CC-SA: Creative Commons Attributions Share Alike)
2. To ensure that metadata contributed to Europeana by end-users can be published by Europeana under the same terms as the metadata that has been provided by Europeana's institutional data providers (CC0 waiver: Creative Commons Zero Public Domain Dedication).
3. To provide Europeana and its partners with the authority to remove infringing, illegal or out-of-context content that has been contributed by end-users.

User-generated content is a promising kind of content for Europeana. UGC refers to a large set of media that are created by the user as opposed to content assembled and curated by institutions. So far, UGC in Europeana refers to the personal collections and/or single objects that people own, and which are catalogued and digitised by professionals. The Europeana 1914-1918 project is a great success. It allows end-users to directly contribute content by bringing their own cultural heritage pieces from the Great War to be digitised at Family History Roadshows, or to upload their material online at home. Europeana has so far benefited in quantitative terms with 49,846 items being added to its collections, but also and especially, qualitatively by enlarging the Great War collection with outstanding new objects.

## 4. Data analysis

In this chapter, we will analyse the other important component of the user experience: the current conditions of metadata in the Europeana repository. We will do this by looking at several distinct but interrelated elements.

---

<sup>12</sup> Source: *The Europeana Licensing Framework*, 10 January 2011, p. 14



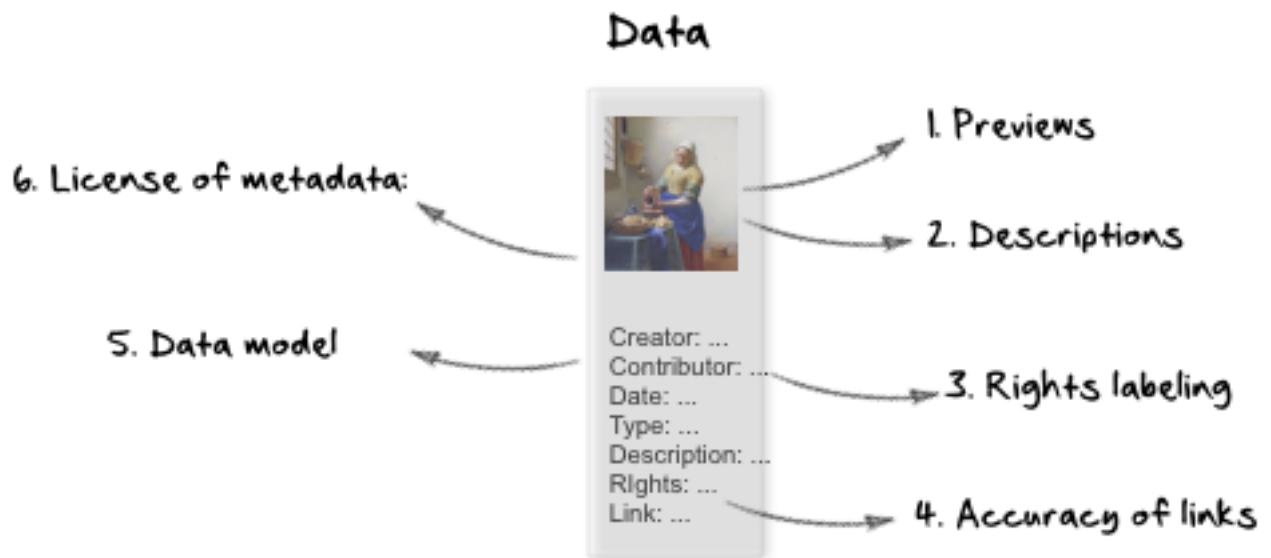


Figure 15: The make-up of Europeana metadata

#### 4.1. Previews

Through user research and the experiences we have had with several user engagement projects and, for example, the Europeana hackathons, we have established that the visual representation of the objects (previews) has become increasingly important for the user experience. The quality of previews is largely established by the size of the preview but in many cases, the preview is not available in Europeana at all. It proves hard to establish exactly how many previews are available, as the ingestion process does not record the existence or successful generation of the preview. Based on a number of cross-disciplinary samples, however, we can roughly estimate that the repository has previews for 60% of the content.

#### 4.2. Descriptions

An important part of the contextual information for the objects in Europeana is what is entered for the 'Description'. The extent and quality of these descriptions varies widely from collection to collection, some are (very) extensive while, in other cases, the descriptions leave a lot to be desired. The standardisation of description quality would be an important contribution to the overall user experience.

#### 4.3 Rights labelling

Understanding what one can and cannot do with the content represented in Europeana is important if users are to use the repository to its greatest advantage. Currently, there are 12

rights statements that data providers may use to provide rights information about the digital objects (and correspondingly the previews), which should help users to understand the conditions of re-use for each object. These rights statements are explained in the *Europeana Licensing Framework*<sup>13</sup> and in the *Guidelines for the Rights in Objects Submitted to Europeana*.<sup>14</sup>

The available rights statements are:

Public domain: the public domain mark (PDM) must be applied to all content that is in the public domain. Works that are labelled as being in the public domain can be re-used by anyone without any restrictions.

The Creative Commons Zero Public Domain Dedication: if a data provider wants to waive all its rights in a digital object, they can apply this CC0 waiver to the works in question. By applying this waiver, all rights to the content are waived and it can then be re-used by anyone without restrictions.

Creative Commons licences: these provide options for copyright holders to allow others to re-use digitised objects exposed through Europeana under certain conditions. These conditions range from relatively open (only requiring attribution in the case of re-use or redistribution) to relatively restrictive (only allowing the non-commercial redistribution of verbatim copies). Creative Commons licences can only be applied by the rights holder or an entity that has been authorised by the rights holder to apply them. In general, this means that data providers will only be able to CC-license digital objects when they are also the rights holders for these objects.

The following six Creative Commons licences can be used as rights statements for digital objects that are described in Europeana:

- Creative Commons – Attribution (BY);
- Creative Commons – Attribution, Share Alike (BY-SA);
- Creative Commons – Attribution, No derivatives (BY-ND);
- Creative Commons – Attribution, Non-Commercial (BY-NC);
- Creative Commons – Attribution, Non-Commercial, Share Alike (BY-NC-SA);
- Creative Commons – Attribution, Non-Commercial, No Derivative (BY-NC-ND).

---

<sup>13</sup> Source: *Europeana Licensing Framework*

<sup>14</sup> Source: *Guidelines for the Rights in Objects Submitted to Europeana*

Rights Reserved: Europeana has also developed three standardised rights statements that can be applied by data providers to indicate that digital objects can be accessed on their website but may not be re-used by third parties. By applying one of the Europeana rights statements, the data provider is indicating that rights to the digital object are reserved and that it may not be re-used without additional permissions by the data provider or rights holder. These statements are:

- Right Reserved-Free Access (applicable when users have free, direct and full access to the digitised object);
- Rights Reserved-Paid Access (applicable when users need to pay to gain access to the digitised object itself);
- Rights Reserved-Restricted Access (applicable when users are limited in accessing a digitised object by other means than the need to pay for it, for example when registration is required).

Unknown: the Unknown rights statement can be applied to digital objects when the data provider does not have conclusive information pertaining to the rights status of the digital object. This label may only be used when the copyright status of the work described remains unknown after a diligent search has been carried out.

Europeana has set a target of eight million labelled open-licence objects (PDM, CC0, CC-BY, CC-BY-SA) to be reached in 2013.<sup>15</sup> In order to achieve this, a rights labelling campaign has been launched to raise awareness of the importance of providing metadata with rights labels, but also to address datasets with missing or incorrect rights statements. In accordance with the EDM guidelines, the 'edm:rights' field is now mandatory so objects or datasets will not be published without this information. For objects and datasets already published but missing this information, Europeana is working with data providers to update their datasets with this information.

The rights labelling campaign, launched in January 2013, has already returned good results even at this early stage. Within the first quarter of 2013, the amount of unmarked objects has fallen from 34% to 30% and the percentage of openly licensed objects (PDM, CC0, CC-BY and CC-BY-SA) is 27%, which accounts for over 6.9m objects.

---

<sup>15</sup> Source: *Europeana Business Plan 2013*, Key Performance Indicator 11

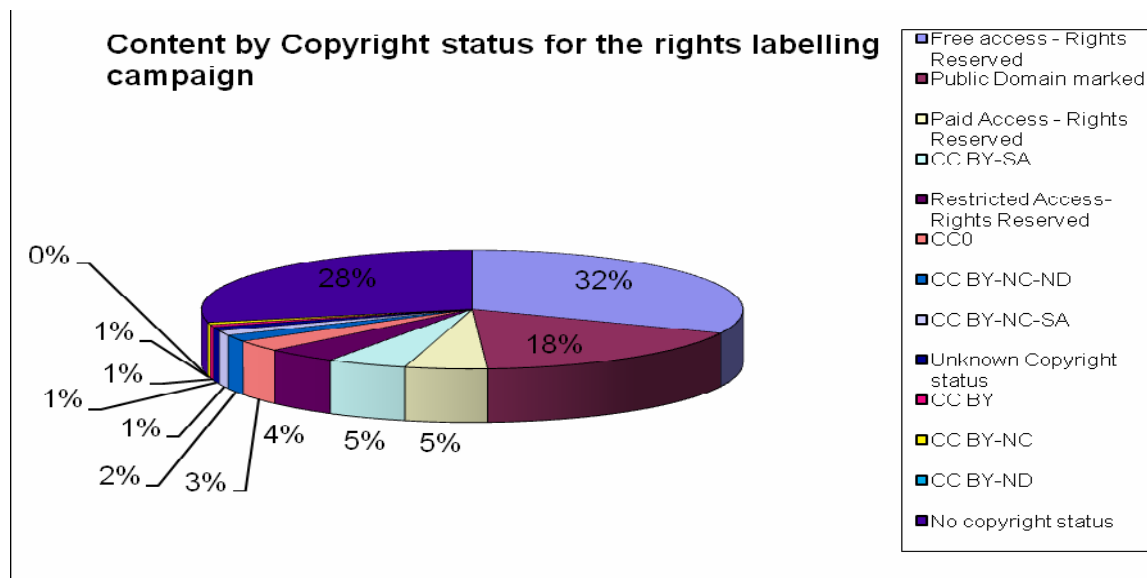


Figure 16: Content by rights status (March 2013)

| Rights statement                    | Total             | % of total  |
|-------------------------------------|-------------------|-------------|
| Free access-rights reserved         | 8,455,389         | 32%         |
| Public domain marked                | 4,693,674         | 18%         |
| Paid access –rights reserved        | 1,310,456         | 5%          |
| CC-BY-SA                            | 1,278,284         | 5%          |
| Restricted access – rights reserved | 1,131,110         | 4%          |
| CC0                                 | 810,182           | 3%          |
| CC BY-NC-ND                         | 483,259           | 2%          |
| CC BY-NC-SA                         | 348,249           | 1%          |
| Unknow copyrights status            | 276,105           | 1%          |
| CC BY                               | 172,680           | 1%          |
| CC BY-NC                            | 167,552           | 1%          |
| CC BY-ND                            | 16,067            | 0%          |
| No copyright status                 | 7,639,985         | 28%         |
| <b>Total of objects</b>             | <b>26,787,154</b> | <b>100%</b> |

Table 4: Content by rights status (March 2013)

#### 4.4. Accuracy of links

The link leading users to an object on a partner's website is an essential element of the Europeana business model. Europeana uses a tool that checks the links on a daily basis. It checks three objects per collection every day. The resulting estimation of dysfunctional links in Europeana has been established at 3% on average.

#### 4.5 Europeana Data Model

In autumn 2012, Europeana and its partners moved from the Europeana Semantic Elements (ESE) model towards the richer Europeana Data Model (EDM), as the formal specification for the classes and properties used in Europeana. The EDM details all the classes and properties of an object, not just the smaller subset used by the initial ESE model. This new model improves data quality by allowing a finer level of granularity in the description. In general, this means richer metadata that makes the linking between objects and contextual resources possible.

The EDM specifies how data needs to be formatted so that Europeana can use it. The specification covers many things including rights information. Europeana uses the contents of the 'edm:rights' field to indicate to end-users under which terms they can use the previews and digital objects that they find in Europeana.

Europeana accepts only one rights statement per digital object. Based on the value in the 'edm:rights' field, Europeana displays a badge alongside the preview indicating the rights situation. Clicking the badge takes the user to a web page explaining the rights statement and informing the user about the terms under which the digital object and the corresponding preview can be used. Europeana also allows its users to filter search results based on the 'edm:rights' information. This allows users to restrict their searches to objects that they can re-use.

#### 4.6 Metadata Licence

Perhaps the most far-reaching change over the recent years is the transition of the licensing conditions for metadata from the more restrictive CC-BY-NC to the open CC0 licence, which is provided for under the Data Exchange Agreement (DEA).

The DEA (Data Exchange Agreement) is the central element of the *Europeana Licensing Framework*. It structures the relationship of Europeana and its data providers, and deals with the provision of metadata and previews by data providers to Europeana.

Firstly, the DEA establishes that Europeana publishes metadata it receives from its data providers under the terms of the Creative Commons Zero Universal Public Domain Dedication (CC0). This means that anyone can use the metadata published by Europeana for any purpose without any restrictions whatsoever. Effectively, the metadata becomes part of the public domain and is no longer subject to intellectual property rights. However, the

DEA places a number of restrictions on the way Europeana itself can use the metadata. Indeed, whenever Europeana publishes metadata that can be attributed to one or more data providers, it has to give attribution to these providers. This ensures that whenever users of Europeana encounter metadata on the website, this information will clearly be attributed to the data provider(s) who have provided it. As a whole, the DEA guarantees that the metadata provided to Europeana can be re-used as freely as possible, while at the same time supporting the data providers' attribution, visibility and professional reputations.

Secondly, the DEA specifies how previews provided by data providers can be used by Europeana and third parties. Previews do not currently fall under the scope of the CC0 waiver and cannot be re-used by third parties unless specified. Previews may only be published together with the metadata that they pertain to, which ensures that they will be attributed to the data provider. Third-party users can only re-use the previews in accordance with the rights statement displayed alongside them. Data providers are therefore required to provide Europeana with correct metadata on the intellectual property rights of the content (digital objects) as fully as possible. This requirement is important because it allows Europeana to communicate rights information about the previews that are displayed on the portal to end-users. It also enables Europeana to properly communicate to its users what they will be able to do with the content that they have found via Europeana. For ease of use, rights statement icons are displayed alongside search results.

Thirdly, the DEA suggests that data providers decide for themselves how rich the metadata they provide to Europeana should be. This allows data providers to keep certain types of metadata to themselves, for example, because of copyright restrictions.

## **5. Conclusions and recommendations**

In this final chapter, we will draw conclusions from the analysis and make recommendations for future strategies. The consequences of these conclusions have been presented in the *Business Plan 2013* after consultation with our Europeana Network partners in September 2012.

Based on this analysis, we can conclude that a tremendous amount of work has been done in the period 2008-2012 at both the collection and data level. Over 26 million objects representing the richness of our cultural heritage have been collected in the data repository

and made available through various interfaces in the Europeana.eu domain. However, a lot remains to be done. From an end-user perspective, the depth and breadth of the collections are not yet up to the level many expect from a trustworthy, authoritative database representing the cultural heritage of Europe. On the content level, the CARARE project has produced a large number of objects from archaeology and architecture which has widened the scope of Europeana beyond museums, archives and libraries. Equally, on the level of data representation, an end-user would have some legitimate claims on quality improvements: previews are often lacking, as are important contextual data such as the rights status of the work, accurate geospatial information and informative descriptions.

It has to be stressed that now is an important time for working on quality improvements. With the release of the data under CC0, the opportunities for wide distribution will increase tremendously. Ensuring that the content and data are of sufficient quality is crucial if we claim to be a trustworthy source of information.

## 5.1 Content and collections

As has been argued in the previous chapters, the content and collections in Europeana are largely the responsibility of the partners in Europeana. Currently, links to 26 million works from over 2,200 partners are represented in the Europeana repository. Every partner and every country will, of course, want to be represented as fully as possible on the international stage. The Europeana Office, however, has to look at the complete picture from an end-user perspective and will try to ensure, where possible, that the breadth and depth of the collection is well-balanced across countries, domains, types and time.

Many of Europeana's partners take part in one or more projects related to aggregating material for Europeana. The European Commission has co-funded a number of European projects mainly through its Competitiveness and Innovation Framework Programme (CIP) and Information and Communication Technologies Policy Support Programme (ICT-PSP).

EU-funded projects relating to Europeana generally fall into two broad categories:

- Projects that work to increase the amount of metadata relating to digitised collections available via the Europeana dataset, and
- Technological projects which assist Europeana's work in other ways.

Europeana has a wide range of engagement with various EU projects. Europeana can either be a primary leader, consortium member or a subcontractor for any one project. The primary leader role is specifically for projects that Europeana is leading such as Europeana Cloud

and Europeana Awareness. Europeana also works with other organisations and content providers as part of a central team undertaking a project as a consortium partner. Because of resource limitations, Europeana is a consortium partner in limited cases only. More often, Europeana acts as a subcontractor in a project, undertaking specific set tasks such as ensuring any metadata is made available via the European datasets.

Of the 29 projects running at this moment, 21 are projects which will result in more metadata being available via the Europeana dataset.

In 2013, there are eight key projects addressing Europeana's content gap and increasing available metadata:

- For audiovisual material - European Film Gateway (EFG) and EUScreenXL
- For archive material - APEX
- For 20<sup>th</sup> century material - Europeana Fashion, Europeana Newspapers, EURO-Photo, Europeana Photography and Digitising Contemporary Art (DCA).

More details on these projects see chapter 5.1.3

#### *5.1.1 Total numbers*

With almost 27 million objects in March 2013, Europeana is well on target. As stated in the *Strategic Plan 2011-2015*, the aim is to reach 30 million by 2015. We are confident that we will reach this number by continuing to foster good relationships with partner projects and aggregators. Good progress to date means that we now also have room to focus on the quality and consistency of the database.

#### *5.1.2 Country*

While a lot has been accomplished and Europeana can claim to have representation from 33 countries, we still need to improve the balance between countries if we are to fulfil our aim of representing the full diversity of our cultural heritage. With regard to underrepresentation by country, we have focused on the target number of objects per country to be delivered by 2015, based on the recommendation of the European Commission (27 October 2011). This is an objective way, and respects each country's specific situation. It also gives each country a clear view of progress against their 2015 target. The gap between the 2015 target and the current achievement varies per country; countries have been divided into three tiers (see also Table 1):



Tier 1 countries (gap >50%): Bulgaria, Cyprus, Czech Republic, Greece, Italy, Latvia, Portugal, Romania, Slovakia and the United Kingdom

Tier 2 countries (gap between 20%-50%): Belgium, Finland, France, Germany, Lithuania and Malta

Tier 3 countries (gap <20% or over target/no gap): Austria, Estonia, Ireland, Poland, Slovenia, Spain, Hungary, Denmark, Sweden, the Netherlands and Luxembourg.

Tables 1 and 2 give an overview of the existing gaps in content delivery per country. We will support Tier 1 countries with any aggregation issues and developments. Tier 2 countries need to be approached actively, with a strong focus on the potential to contribute audiovisual material. Europeana will nurture Tier 3 countries in order to reach a higher contribution of audiovisual material and to ensure that their masterpieces are included (in the case of the Netherlands, Luxembourg and Sweden).

The Tier 1 countries will require the most direct engagement by Europeana. Along with raising awareness and promoting the digitisation of content gap materials, Europeana will work with them to increase the amount and range of organisations working with the aggregators to ensure regular and increasing amounts of metadata are made available. This may also include hands-on workshops and other meetings to increase the face-to-face interaction with Europeana as well as a concerted effort to increase engagement via the Aggregator Forum and Basecamp discussions. Basecamp is a web-based project management and collaboration tool which partners are invited to use to share files and have discussions.

We will support the Tier 2 group through a mixture of raising awareness of the content gap and working directly with aggregators to promote the digitisation of content gap material by the content holders. This group will require considerable input from Europeana but will also have the most potential to increase metadata and address the content gap. Again, work with this group includes support to determine the parameters for masterpieces but also for increasing the general amount of metadata available to create a more balanced dataset.

To assist the Tier 3 group, we will be contacting them, discussing their content gaps and establishing where areas of possible development could be. For example, the Swedish aggregator does not aggregate audiovisual material but over 1,400 Swedish video items are available via EUscreen. There will be value in working with the Swedish national aggregator to find more sources of film, video and audio.

All aggregators have established networks of content-holders, so disseminating information to content-holders will be more efficient via the aggregator than directly from Europeana. Aggregators will also be able to play a significant role in the development of solutions to issues that may arise in relation to the content gap, such as being the focal point for the creation of the lists of masterpieces from any one country.

While contributions from EU member states will remain a priority, we welcome relevant objects from non-member states and non-EU countries.

To increase content from underrepresented countries, we will focus on building stronger and new partnerships with aggregators and content providers in individual countries. The aim is to include more cultural content from these countries, but with a specific focus on masterpieces and audiovisual material.

### *5.1.3 By domain*

Single-domain portals with both aggregation and distribution functions for their sectors are very important in the Europeana ecosystem. By aggregating content from their domains, e.g. all library material into The European Library or film material into the European Film Gateway, it ensures that their professional audiences gain access in a format-specific context. This material is then ingested into Europeana, which has a broader appeal than the single-domain aggregators. The benefits realised by single-domain aggregations such as The European Library (national and research libraries), the European Film Gateway (audio), APEX (national archives) and EUscreen (audio), include targeted expertise on particular content such as library, film or archival material. Such aggregators have, over many years, developed metadata schemas that bring the best out of the material they host; this expertise has led to the development of the Europeana Data Model.

Europeana also gathers metadata from single content providers such as the Institut National de l'Audiovisuel in France or the Rijksmuseum in the Netherlands. For efficiency and cost effectiveness, Europeana prefers content providers to aggregate via a domain or national aggregator, or a national initiative. If one of these three alternatives is not feasible, Europeana will take data in directly to ensure that material is accessible.

As well as having a more balanced representation of material at a national level, it is also necessary to have a balanced EU range of content across the domains.

The numbers in table 3 demonstrate that domain-specific aggregators and projects such as The European Library benefit Europeana by facilitating access to certain content types. It also shows the underrepresentation of certain domains. These areas of underrepresentation need to be proactively addressed by Europeana and its partners if we are going to achieve the most value for our end-users. With such a wide base of end-users accessing digitised content via Europeana, an overemphasis on one domain could create a bias and increase the difficulty of finding the right material quickly and easily.

As an example, currently the primary sources of audiovisual archive material are two projects, EUscreen and European Film Gateway (EFG). Both of these projects are well-established and are, in fact, in the early stages of initiating their second phases. In the gap between the end of the first phases and initiation of the second, no audiovisual material was aggregated to Europeana. In tandem to working with these specific projects, more work needs to be done with national aggregators to ensure they are able to aggregate metadata relating to audiovisual collections and have active and ongoing relationships with their respective audiovisual collections.

APEX is the second phase of the APENET project, which is a pan-European project providing access to archival records. In 2012, Europeana and representatives of APEX met to discuss future closer collaboration with the aim of increasing the amount of archival content made available. In 2013, this will be continued and work will be done to forecast the amounts of archive-related metadata that will be aggregated over the year. There will also be discussions to see whether there are possibilities of increasing the amount of archival material further.

The projects Europeana Fashion, Europeana Newspapers, Europhoto, Europeana Photography and Digitising Contemporary Art specifically cover the 20<sup>th</sup> century gaps and were selected for particular attention because of their focus on a defined area of content which is not well-represented. In most cases, they are also focused on material that is not usually available in a digital form through public heritage collections, such as commercial photography, fashion design or commercial newspaper publishing. It is hoped that working with these specific projects, Europeana will be able to learn the process these organisations undertake to reach their published goals and will be able to assist using its own network.

#### 5.1.4 *By type*

Looking at the composition of the database by type, a striking fact is that images contribute close to two-thirds of the database, texts one-third and that audio and video are underrepresented with only 3% of the total. The current analysis falls short of detailed explanation for this, although it would be fair to say that the expense of digitising audiovisual material and the very complicated rights structure relating to them are two important factors. The commercial use of audiovisual material by businesses creates another substantial blockage to general usage. From an end-user perspective, moving image is an increasingly important part of the user experience and should therefore be stimulated. Our strategy will be directed at actively approaching the large audiovisual archives in Europe in order to meet the target of one million audiovisual objects in the Europeana repository during 2013, (the unreached target set for 2012), and two million objects by 2015. But taking into account the abovementioned serious issues, these numbers are ambitious to say the least.

Across Europe, there are around 100 national and regional film archives and even more sound collections that are either stand-alone organisations or are part of a wider collection within a GLAM (galleries, libraries, archives, museums) organisation (e.g. the British Library). In addition to this, there are private archives which are part of national or regional broadcasters or footage libraries which hold large amounts of audiovisual material which is often already digitised.

The reasons for the lack of availability of digitised audio, film and video items can be traced back to common themes of digitisation across the GLAM sector. Associated rights, digitisation infrastructure, funding, hosting and metadata mapping are all factors and these are amplified with film, video and audio due to the time-based nature of the content. But Europeana's own research has shown that audiovisual material is an area in which end-users are very interested – AV items are 10 times more likely to be clicked on than other records.

In 2012, Europeana developed a direct approach to address the lack of audiovisual content. With EFG in the very early phases of its second phase and working predominantly on First World War content, and with EUscreen wrapping up its first project, Europeana assessed and established possible priority organisations and developed a strategy for contacting film archives both directly and via the network of aggregators.

Although EFG1914 has already begun, one of the issues that arose in 2012 was that there was a perception in the wider audiovisual archive sector that it was only selecting First World

War era material. This was not the case and so in 2013 there will be a concerted joint effort by both Europeana and EFG to continue to make non-First World War material available via EFG.

EUScreenXL, the project for television archive material, will be an important partner for Europeana. The work EUScreen has done in building relations with broadcast archives is a good example of the technical expertise that a project can bring to an area of digitised content that the more generalist aggregation model sometimes is unable to address.

In 2013, a different approach will be adopted taking into account the current environment and the new phases of the respective AV projects. There will be active engagement with large and small audiovisual archives but this will be undertaken in conjunction with EFG for film and EUScreenXL for television archives. There will also be a concerted effort to engage the network of aggregators and initiatives to work with their local audiovisual archives to increase the amount of material available or to work through plans to ensure future availability via Europeana.

There will also be concerted work to better understand the landscape of audio archives and, if it is possible, to increase the engagement of audio archives and collections with Europeana. This will entail working with professional associations, aggregators and existing audio collections who currently supply metadata to Europeana.

#### *5.1.5 By year*

The final variable that Europeana wishes to address with the national and domain imbalance is the date range of material. Europeana's own research of its dataset has shown that there is under-representation of the 20<sup>th</sup> century when compared to earlier time periods. For Europeana, this is a risk because it limits the value of our services. For example, as Europeana moves to engage with the education sectors, if the dataset does not have appropriate content to meet the requirements of the curriculum then it will not be of any value and so will not be adopted.

There are some obvious reasons why more modern material is less available than earlier material. Undoubtedly, rights and the expense of digitisation of more modern media (such as film and video tape) have had a direct influence on the dates represented. It could also be reasoned that earlier material like paper records and art have been higher in the preservation priorities of cultural organisations which may have led to the earlier digitisation of certain collections.

To address this gap, Europeana will have to work with its strategic partners to first fully understand why 20<sup>th</sup> century material has not been digitised and then work with initiatives to highlight this with organisations digitising collections. Specific projects are currently working to address this gap, for example PartagePlus, which is working to digitise works in the Art Nouveau style popular in the late 19<sup>th</sup> and early 20<sup>th</sup> century.

Another area of work Europeana is focusing on which will also bridge the 20<sup>th</sup> century gap is the inclusion of user-generated content (UGC). The term UGC can be used to describe many areas of material created by users but in the context of Europeana this usually means the addition of metadata (e.g. tags, geo-location information, and descriptive information) or the creation of content through digitisation. The digitisation of material by users and its inclusion in the Europeana dataset has been the focus of the Europeana 1914-1918 project. This project has a website through which users can upload their digitised artefacts from the First World War, and add descriptive information and context. Their items are then also aggregated into Europeana. In conjunction with the website, a series of Family History Roadshows have been taking place across Europe, at which members of the public can bring their artefacts to have them digitised and have their stories (metadata) recorded. These roadshows are very successful and have resulted in a large amount of material being digitised and made available. The success of this project has led to another similar project which will begin in 2013 and have as its focus the fall of the Iron Curtain in the Warsaw Pact countries in 1989.

UGC projects expose a large amount of content that is of great interest to the public and would not necessarily be available from a more traditional collecting institution such as a museum or an archive. They are also specifically focused on 20<sup>th</sup> century events and so go some way to covering Europeana's content gap. With significant anniversaries of the First World War and the fall of the Iron Curtain occurring in the coming years, these projects will be an important area of development.

Further investigation is required to create a better balance across the centuries, with particular attention paid to the 20<sup>th</sup> century and the problem of so-called 'orphan works'. Orphan works are 'works and other subject-matter which are protected by copyright or related rights and for which no rightholder is identified or for which the rightholder, even if identified, is not located'.<sup>16</sup>

---

<sup>16</sup> Source: European Commission, *Orphan Works Directive 2012/28/EU*, p. 1

In this situation, it would appear sensible for cultural heritage institutions to digitise and make available those works from this period that are not available commercially any more ('out-of-commerce' works). However, clearing rights for these works is often very resource-intensive or even impossible in the case of 'orphan' or 'hostage' works.

As part of its mission to make Europe's cultural and scientific heritage available to the citizens of Europe, Europeana and its contributing partners need to be enabled to make 20<sup>th</sup> century out-of-commerce and orphan works available. This is especially relevant since material from this period relates to historic events that shape the social and political realities of Europe today (the First and Second World Wars, the Cold War and the fall of the Iron Curtain, the process of European unification, migrations, the intra-European migration movements between the north and the south of Europe, etc.).

This ambition requires changes to policy to enable heritage institutions contributing to Europeana to deal with the challenges posed by mass digitisation projects of 20<sup>th</sup> century material and out-of-commerce and 'orphan' works in particular. Europeana is concerned that the recent policy initiatives on these issues fail to adequately address these questions. The proposed orphan works directive does not provide sufficient incentives for heritage institutions to engage in large-scale digitisation of orphan works and the Memorandum of Understanding on out-of-commerce works lacks legislative backing and still has to be translated into concrete projects. We are also worried that in spite of efforts to harmonise policy on the European level, member states are reverting to national solutions (see the recent French legislation and the proposals made by the UK government). This will severely hamper Europeana's ability to provide pan-European access to 20th century content from all over Europe to all Europeans.

#### *5.1.6 Masterpieces*

Masterpieces play a significant role in nation-building and thus they are an integral part of a country's cultural heritage. Therefore, recognising European masterpieces and facilitating their introduction to Europeana remains a priority.

In 2012, Europeana also initiated activities in response to the recommendation of the *New Renaissance Report* to ensure the availability of European masterpieces via Europeana. This work included presentations at conferences and workshops across Europe and the use of the Aggregator Forum to raise awareness.

Initial research by Europeana to define masterpieces had been undertaken in 2011-12. This gave a list of suggested masterpieces, with an average of 51 works per country, although varying dramatically from 13 for the smallest country to 116 for the largest. The list does not include every great piece of work but rather a suggestion of a notable work from a greater number of artists. This way, more attention can be given to internationally lesser-known but locally very well-known artists.

Making these lists available to the national aggregators for feedback raised many more fundamental issues regarding masterpieces and their availability. What constitutes a masterpiece? Who should make the selection? Each domain has its own masterpieces so how would the priority be set? These issues have still to be fully investigated, discussed and resolved so in 2013, in addition to awareness-raising, Europeana plans to increase the engagement of the Member States Expert Group to advise on the wider questions regarding the selection of masterpieces.

The Europeana Office could actively support the MSEG and ministries with the research they are conducting but ultimately the selection would be signed off by the individual ministries. The next steps would then be that the selected masterpiece content would need to be either digitised, if this has not happened before, or made available to Europeana. In the case of digitisation, the Europeana Office could help the institutions involved to find the right projects for them to make the content digitally available.

#### *5.1.7 UGC*

Europeana will be collecting more information created by users in the future, in particular through initiatives such as the Family History Roadshows. This is a valid addition to the repository, but we need to be alert that the inclusion of UGC does not compromise the mandate of Europeana to create a trusted source of information. To develop a real strategy requires more investigation on the possibilities.

## 5.2 Europeana Cloud

The Europeana Cloud project kicked off in March 2013 to experiment with a more sophisticated, cheaper and easily sustainable repository. Content providers and aggregators across the Europeana information landscape urgently need a cheaper, more sustainable infrastructure that is capable of storing both metadata and content. Researchers require a digital space where they can undertake innovative exploration and analysis of Europe's digitised content. This project aims to provide solutions for both providers and researchers.



### 5.3 Specific actions

**Increase the amount of previews:** it is estimated that currently 75% of the objects in the repository have previews. Visitors want and expect previews. Although increased by 20% from last year it is still too low. Target is set on 90% previews before 2015.

**Improve the amount of objects with rights labels:** currently over 70% of the objects have rights statements in their metadata. With the introduction of EDM, this has become a mandatory field but a concerted effort is needed in order to attach rights labels to all objects. Special emphasis is required to ensure correct rights statements are attached to objects and to increase the amount of openly licensed content. Currently, over 7 million objects are available for free creative re-use.

**Decrease the amount of dysfunctional links:** the amount of 'dysfunctional' links is 3% on average. In order to improve the user experience, this number should at least stay stable but preferably decrease to less than 1%. Investigate possibilities for permanent identifiers.

## References

Comité des Sages New Renaissance Report

[http://ec.europa.eu/information\\_society/activities/digital\\_libraries/doc/refgroup/final\\_report\\_cds.pdf](http://ec.europa.eu/information_society/activities/digital_libraries/doc/refgroup/final_report_cds.pdf)

DISMARC website

<http://www.dismarc.org/info/index.html>

Europeana Business Plan 2012

<http://pro.europeana.eu/documents/858566/c0c6e31f-5174-4898-9771-f9b9a8d1d4d7>

European Commission, Orphan Works Directive 2012/28/EU

<http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2012:299:0005:0012:EN:PDF>

European Commission Recommendation 2011/711/EU of 27 October 2011 on the digitisation and online accessibility of cultural material and digital preservation

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

Europeana Content Development Strategy

<http://pro.europeana.eu/documents/866205/0/EV1-ContentDevStrategySummary.pdf>

Europeana Licensing Framework

<http://pro.europeana.eu/documents/858566/7f14c82a-f76c-4f4f-b8a7-600d2168a73d>

Europeana Policy on User-Generated Content

[pro.europeana.eu/documents/844813/851957/D1.4+UGC+policy.pdf](http://pro.europeana.eu/documents/844813/851957/D1.4+UGC+policy.pdf)

Europeana Partner Strategy and Development Plan 2012

Available soon at pro.europeana.eu

Europeana Strategic Plan 2011-2015

[http://www.pro.europeana.eu/web/guest/search?p\\_p\\_auth=E2riqNBe&p\\_p\\_id=20&p\\_p\\_lifecycle=1&p\\_p\\_state=exclusive&p\\_p\\_mode=view&p\\_p\\_col\\_id=column-2&p\\_p\\_col\\_count=1&\\_struts\\_action=%2Fdocument\\_library%2Fget\\_file&\\_groupId=10602&\\_folderId=370691&\\_name=13464](http://www.pro.europeana.eu/web/guest/search?p_p_auth=E2riqNBe&p_p_id=20&p_p_lifecycle=1&p_p_state=exclusive&p_p_mode=view&p_p_col_id=column-2&p_p_col_count=1&_struts_action=%2Fdocument_library%2Fget_file&_groupId=10602&_folderId=370691&_name=13464)

D3.8: Content Strategy Update  
MS12: Content Acquisition and Development Plan

Guidelines for the Rights in Objects Submitted to Europeana

<http://pro.europeana.eu/documents/900548/0d423921-23e0-45fa-82a4-2ac72b3e6f38>

Linked Open Data pilot description

<http://pro.europeana.eu/linked-open-data>

Numeric project: Developing a statistical framework for measuring the progress made in the digitisation of cultural materials and content, Final Report, 2010

[http://cordis.europa.eu/fp7/ict/telearn-digicult/numeric-study\\_en.pdf](http://cordis.europa.eu/fp7/ict/telearn-digicult/numeric-study_en.pdf)