

Research Paper:

The Accuracy of Rights Statements on Europeana.eu

Judith Blijden (Kennisland)



Kennisland

The research was led by Judith Blijden (Kennisland). The script for this research has been written by Maarten Zeinstra (Kennisland) who also conducted the data retrieval. The individual assessments have been conducted by Judith Blijden, Lisette Kalshoven, Maarten Zeinstra, Paul Keller (Kennisland) and Pablo Uceda Gomez (Europeana). This research paper was written by Judith Blijden.

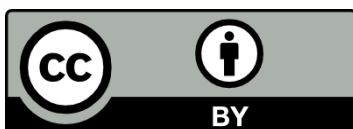
Europeana DSI-2 is co-financed by the European Union's Connecting Europe Facility. The sole responsibility of this publication lies with the author. The European Union is not responsible for any use that may be made of the information contained therein.



Co-financed by the European Union

Connecting Europe Facility

Unless otherwise noted everything in this report is licensed under a [Creative Commons Attribution 4.0 license](#). This means the document can be reused for any purpose, build upon and adapted - even for commercial purposes - when proper attribution is given to the authors. We kindly ask you to use the following method of attribution: Judith Blijden (Kennisland), (Amsterdam 2018), [CC BY 4.0](#). In case of offline use of the report, include the URL of the license in the attribution: Judith Blijden (Kennisland), Amsterdam 2018), <https://creativecommons.org/licenses/by/4.0/>.



Cover picture:

Girl in front of a mirror (1652) | Gerard ter Borch | Rijksmuseum | Public domain

Contents

1. Research question	3
1.1 The problem	3
1.2 Research question	4
2. Preliminary research	5
2.1 Preliminary qualitative research	5
2.2 Preliminary quantitative research	6
2.3 Unexpected outcomes	8
3. Methodology	10
3.1 Data retrieval	10
3.2 Individual assessment	11
3.3 Limitations	12
3.4 Analysis of results	12
4. Results	14
4.1 Composition of the sample	14
4.2 Accuracy of rights statements	15
4.3 Reasons for inaccuracy	18
4.4 Summary	20
5. Recommendations	23
5.1 The public domain and non-original and non-digital objects	23
5.2 Training, materials and community	24
5.3 Dynamic and complicated nature of copyright	24
5.4 CC licenses and PDM	25
5.5 Further research	25
Addendum 1 Scripts	26
1.1. Get 1500 random urls	26
1.2. Use regular expressions to transform URLs to list of Europeana IDs	27
1.3. Get additional relevant information	27
Addendum 2 Individual assessment of digital objects from the Europeana Database	29
2.1 Fill in the document by following these steps	29
2.2 Basic rules for assessing the accuracy of rights statements	31

1. Research question

Europeana publishes over 50 million records describing cultural heritage objects. It supports [14 rights statements](#) that communicate the copyright status of an object to the user and, if possible, the reuse conditions of the digital objects they find through the database. There are currently 11 rights statements actively being used in the Europeana database:

- Public Domain Mark (PDM)
- No Copyright - non commercial reuse only (NoC-NC)
- Creative Commons CC0 1.0 Universal Public Domain Dedication (CC0)
- Creative Commons - Attribution License (CC BY)
- Creative Commons - Attribution, ShareAlike License (CC BY-SA)
- Creative Commons - Attribution, No Derivatives License (CC BY-ND)
- Creative Commons - Attribution, Non-Commercial License (CC BY-NC)
- Creative Commons - Attribution, Non-Commercial, License ShareAlike (CC BY-NC-SA)
- Creative Commons - Attribution, Non-Commercial, No Derivatives License (CC BY-NC-ND)
- In Copyright (InC)
- Copyright Not Evaluated (CNE)

There are three available rights statements that have not yet been applied to any objects:

- No Copyright - Other Known Legal Restrictions (NoC-OKLR)
- In Copyright - Educational Use Permitted (InC-EDU)
- In Copyright - EU Orphan Work (InC-EU-OW)

1.1 The problem

When searching through Europeana, one may encounter objects that have a rights statement that seems inaccurate. [Europeana promotes the application of free reuse statements](#) when this is accurate and possible.¹ Europeana employs a '[clean-hands policy](#)' when it comes to rights statements. The data partners apply the rights statements according to their internal policies and relevant local (intellectual property) laws. Europeana generally assumes that the rights statements provided by the data partners are correct. However, during ingestion Europeana does take steps to verify the accuracy of rights statements in the cases where there are reasons to doubt the veracity of the statements provided as written out in the [Europeana Publishing Guide](#).

Europeana is committed to improving the quality of the metadata it publishes, including rights statements. Therefore it is important to have an evidence-based estimate of the accuracy of rights statements so that opportunities to improve their accuracy can be identified. Europeana makes sure their policy documents and other informational resources correspond with the needs of data providers.

¹ The free reuse rights statements are PDM, CC BY, CC BY-SA and CC0.

In this context Kennisland undertook research into the accuracy of the rights statements contained in the Europeana database. Verifying the accuracy of rights statements is a difficult process. Because Europeana contains over 50 million objects, it is impossible to individually assess every object. Due to lack of sufficient correct metadata concerning the status of the creator of the object, it is also not possible to automatically assess the rights statements of objects. Scripts that would be able to automatically retrieve data depend on metadata that has been provided to Europeana. To be able to assess the rights statements automatically, a minimum amount of data such as on the dates of creation and the death of the author(s), would be required. In many cases the data is incorrect or missing, or not available. Therefore, not only the inaccurate rights statements are a problem in the database, finding a method that can assess the accuracy of rights statements is challenging as well.

1.2 Research question

We have conducted research in which the accuracy of a representative sample of the objects in the database is assessed to be able to answer the central question: what is the accuracy of rights statements in the Europeana Database? This research paper will set forth the findings of the research.

In the next chapter we will explain our preliminary research which resulted in the methodology used for this research. In the third chapter we will explain the methodology itself. In the fourth chapter we will present and analyse the results on the accuracy of the rights statements. Lastly, in the fifth chapter we will present a set of recommendations for improving the accuracy of the rights statements.

2. Preliminary research

We have conducted two preliminary studies to make sure our assumptions on common inaccuracies and our methodology to find these inaccuracies were correct. The results of these researches are described below.

First, we conducted preliminary qualitative research to confirm whether our assumptions about common errors were correct. Based on our experience of working with copyright issues in relation to Europeana, we assumed we could divide common errors concerning rights statements into two categories. The first category contains public domain works with an In Copyright rights statement or a Creative Commons (CC) License. The second category concerns in copyright works with a Creative Commons license or with a Public Domain Dedication (CC0) of which we suspected that there was no permission given by the rights holder.

Second, we conducted preliminary quantitative research for which we retrieved large amounts of data from Europeana that could tell us more about the two earlier mentioned categories. This process proved to be more difficult than expected, because the results from the quantitative research differed from the results from our preliminary qualitative research.

In this chapter we will discuss both preliminary researches and their results which enabled us to construct a methodology for assessing the accuracy of the rights statements of Europeana.

2.1 Preliminary qualitative research

For the preliminary research we individually examined the 200 most-viewed digital objects on Europeana in 2016 according to Europeana's webanalytics. We retrieved data on those 200 objects. These make up for 3% of Europeana's pageviews of 2016. We collected the following information of each digital object:

- the page views
- the data partner from which the object originated
- the rights statement attached to the object
- the date of creation

Based on this information we determined the accuracy of the rights statement manually and, when applicable, what rights statement should have been applied to the digital object. We included relevant laws of the jurisdiction of the country that applied to the object in this assessment.

For category 1, public domain works with an In Copyright rights statement or a CC License, this meant that we started looking at the copyright term. The copyright term is based on the life of the author. In the European Union the term is harmonised and extends to 70 years after the death of the creator. When this happens, copyright expires and the work enters the public domain.² The work can then be freely

² For more information about the public domain see '[The Europeana Public Domain Charter](#)'.

distributed and used.³ Rights holders can only license a work when it is in copyright. When a CC license is inaccurately applied to a public domain work, a restriction of the use of that work is mistakenly communicated. A public domain work with an In Copyright rights statement or a CC License is incorrect.

For the second category, in copyright works with a CC license or with a Public Domain Dedication (CC0), we again looked at the copyright term as well as whether it was likely that the author had given permission for a CC license or a CC0 dedication.

Our expectations were confirmed. 22% of 182 digital objects had an incorrect rights statement (chart 1). The larger majority of rights statements fell under the two earlier mentioned categories. The examined dataset was too small to be able to deduct any conclusions on the entire database. We decided to follow up with preliminary quantitative research in which we wanted to retrieve large amounts of data on the two categories.

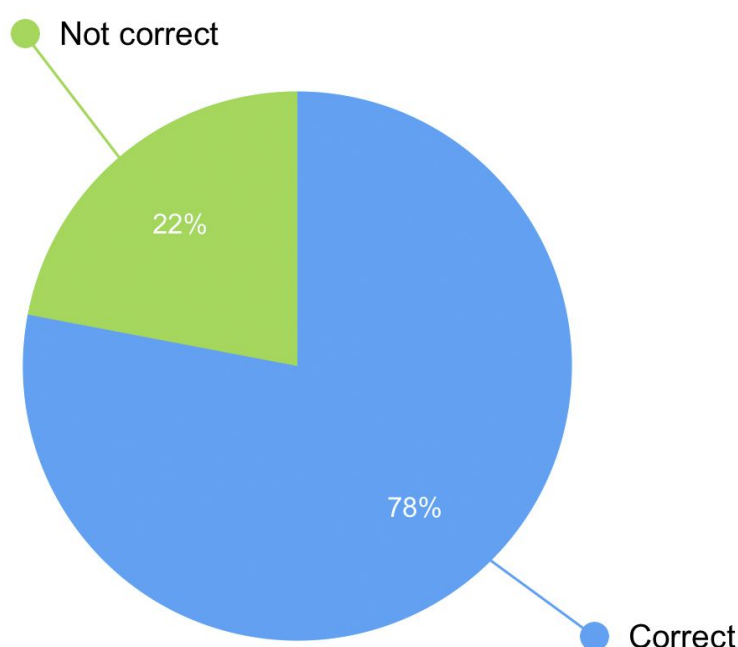


Chart 1: Accuracy results of the qualitative research.

2.2 Preliminary quantitative research

We retrieved a large pool of data that would be more indicative for the entire collection of Europeana. We wrote automated scripts that could bring in large pools of data for further analysis. For the first category we gathered data with a couple of specifications. These were works created before 1860 of which we can reasonably assume that copyright term has since ended.

³ See [Public Domain Usage Guidelines](#).

Additionally, these had to be works with a CC license or an In Copyright statement. We would manually filter out works that received any rights through digitisation⁴ when analysing the data as well as special national provisions.

For the second category we gathered data on collections that were (almost) completely licensed under the same license or with public domain dedication. A collection often contains many different authors. It is unlikely that all these different authors have given consent for the exact same license or dedication. The results for the first category were as follows:

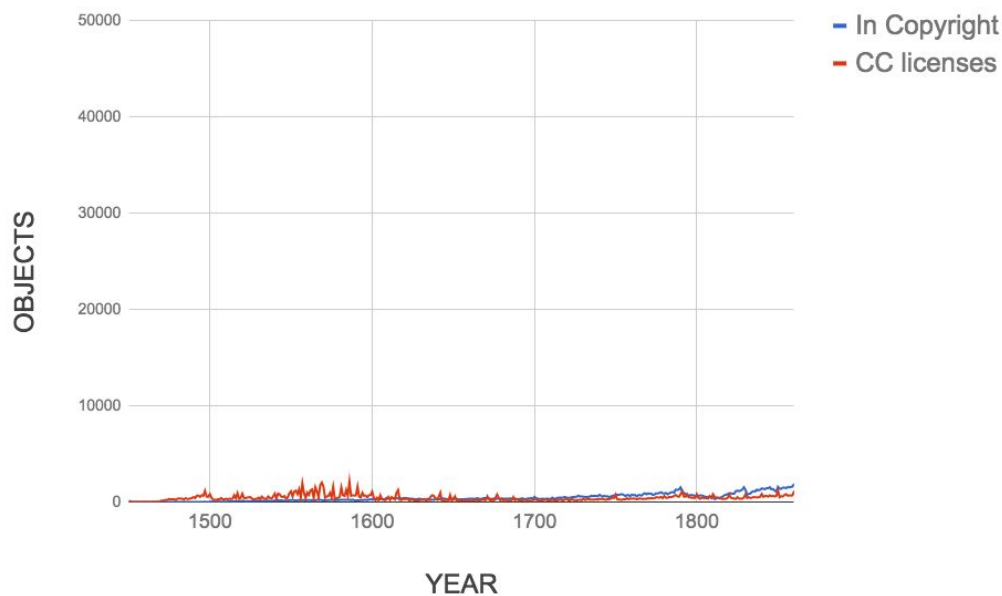


Chart 2: The number of objects before 1860 with with an IC or CC rights statement, category 1. x-axis time from 1500 - 1860, y-axis number of objects.

When we looked at the results we found hardly any works that fulfilled the above mentioned criteria of category 1. The results are shown in Chart 2. Due to our preliminary research we expected to retrieve data that would suggest that works older than 1860 had an In Copyright rights statement or were licensed under a CC license. There is an increase in CC-licensed works and in copyright statements in the twentieth century, as it should be. As time goes on, it becomes less likely that a work is in the public domain as the copyright term might not have ended yet.

The results from the second category showed that only a six institutions had used almost entirely one rights statement for all objects in their collection (in this period). Each colour in chart 3 represents one rights statement. The number of works can be found on the vertical axis. The ones that did have one rights statement for large amounts of works were collections that included works in the public domain. In this case it is expected that they all have the same rights statement. Or they were collections that consisted out of public domain works that had been digitised. The digital objects had received CC0 statements, most likely because the institution wanted

⁴ Rights can be acquired after digitising a work for a various reasons. For instance, because the digitisation involved creative choices or skilled labour. For more information see: <http://outofcopyright.eu/rights-after-digitisation/>.

to make sure to waive possible rights acquired after digitisation. There was only one institution that licensed its entire collection under CC BY-NC-SA.

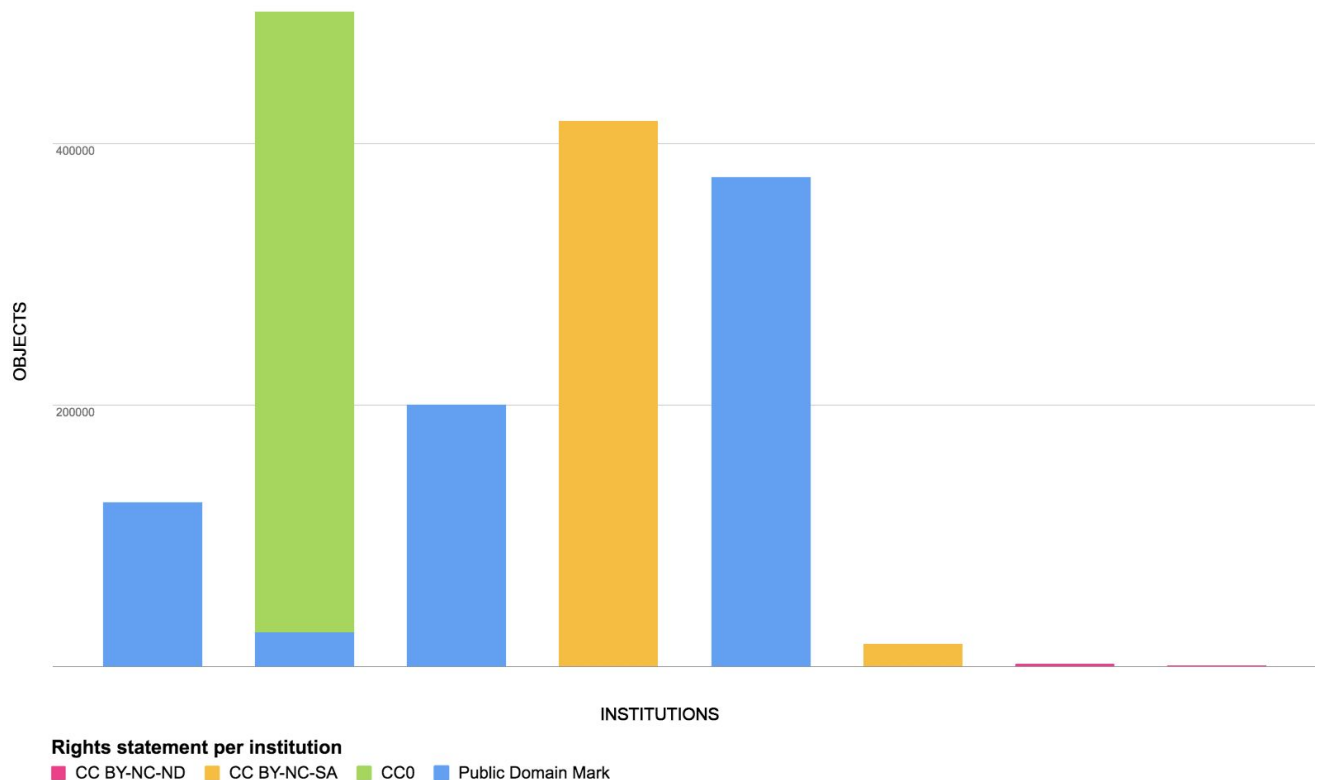


Chart 3: Category 2, collections that were (almost) completely licensed under the same license or with the public domain dedication. X-axis institutions, y-axis number of objects.

2.3 Unexpected outcomes

We have conducted two preliminary researches with contradictory results. Qualitative research that suggested our assumptions were correct, but was too small of a sample to be indicative for the entire Europeana database. And quantitative research that suggested our assumptions were incorrect, but was based on a large and one-sided amount of (possibly incomplete) metadata. We think there are two reasons for the different results.

The Europeana dataset is not complete (enough) in terms of relevant metadata to be able to automatically determine the rights status of a work through scripts. When writing a script we depended on metadata available in the Europeana Database. However, there are some data fields in the Europeana Data Model which various data partners have interpreted differently. For example, there are institutions that provide 'date of creation' in the date field, while others include 'digitisation date' or 'acquisition date'.

We could only retrieve data from institutions that have filled in metadata fields (edm:issued) in a similar (often correct) way. EDM refers to [Europeana Data Model](#). This has vastly restricted the amount of data we could retrieve and the accuracy of the

results. It narrowed the sample down to data that was of relatively high quality in comparison with the data we could not retrieve. Therefore, the results cannot be used as an indication of the accuracy of rights statements for the Europeana database as a whole. The method used is unsuitable for determining the accuracy of rights statements of the entire database.

We ended up with datasets which appear to be accurately labelled. We noticed that larger institutions such as national museums and archives would fill in metadata in a similar and complete fashion. Larger institutions often have experts dealing with the digitisation of their collection. While it is good news that such institutions have a similar and correct understanding of dealing with metadata, it is not feasible to conclude that other organisations of which we could not retrieve data because of their incorrect metadata, would have correct rights statements.

The results of the quantitative research mostly confirm the idea that digital objects with correct metadata have correct rights statements. The fact that correct metadata seems to correspond with correct rights statements, makes it difficult to systematically examine Europeana's database on incorrect data via data retrieval. At the same time the preliminary qualitative research shows there are good reasons to further examine and analyse the Europeana database on its accuracy.

Our preliminary research showed that one cannot depend on scripts to determine the accuracy of a rights statement. The results give a distorted display of the accuracy of rights statements in Europeana and cannot be used as an indication of the accuracy of rights statements for the Europeana database as a whole.

A remaining possibility to determine the accuracy of the Europeana Database is to individually assess the accuracy of the rights statements of a representative sample. A qualitative research of a (representative) random sample would be able to give an indication of the accuracy of the rights statements of the entire database of Europeana.

3. Methodology

We have constructed a methodology for qualitative research in which we individually assess the accuracy of the rights statements of a sample of Europeana that takes the findings of our preliminary researches into account.

The qualitative research on the accuracy of rights statements consists out of three parts: data retrieval, individual assessment and the analysis of the results. First we needed to retrieve data on a sample of objects that fulfilled the requirements of random sampling so that it would give an indication of the accuracy of the entire database. Secondly, we needed to individually assess the object and its data to determine whether the rights statement was accurate. And finally, we analysed the results of the individual assessment. In this chapter we will discuss the methodology of each part of the research.

3.1 Data retrieval

We collected a sample of data⁵ that is large enough to give an estimation about the accuracy of the rights statements of the entire collection. A requirement of such a sample is that the objects are randomly collected.⁶ It is important that the random sample is unbiased in order to accurately represent the population.

Random sampling is not a feature supported by the Europeana API. However, we found an alternative approach by exploiting the [sitemaps of Europeana](#). Sitemaps provide an index of all Europeana pages to be indexed by search engines. Europeana has around 1,200 sitemaps, each containing 45,000 links to the digital objects.

We created a script that outputs a list of unique random numbers between one and the total objects on Europeana at the time of running.⁷ The list is sorted and numbers are matched to an index between 1 and 45,000 in one of the 1,200 files. Relevant objects are stored and the relative link is outputted. Using the Europeana Identifier, which is part of the link, additional metadata is gathered in a separate Europeana API call. These include title, country, IsShownBy, provider, dataProvider, rights, media type, and author.⁸

The size of the sample depends on different factors such as the size of the population (the complete collection of items that are to be researched), the minimum acceptable level of precision, variability within the population and the sampling method. We calculated the size of the sample according to common practices. The minimum required size of the random sample is 1,068 digital objects to have an indication of the accuracy of the entire database.⁹ For an absolute determination of the accuracy a larger sample would need to be examined. Taking the resource considerations such as time limitations and the technological possibilities into account, we have gathered 1,500

⁵ The data:

<https://docs.google.com/spreadsheets/d/1IGWXySVSHC0uwwR0aF2ODIO688gI7P73W6xjSc0kVrw/edit#gid=0>.

⁶ A random sample is a subset of individuals (the digital objects) chosen from the larger set (the population).

⁷ Using a simple random methodology.

⁸ [Definition of the Europeana Data Model v5.2.7](#).

⁹ See also: <https://www.surveymonkey.com/mp/sample-size-calculator/>.

objects that can also be used for other examinations such as broken links or the quality of metadata.

3.2 Individual assessment

The digital objects in the data sample have been individually examined on their accuracy by a researcher with a knowledge of both the Europeana Licensing Framework and European Copyright law in its variations. The individual assessments have been conducted by Judith Blijden, Lisette Kalshoven, Maarten Zeinstra and Paul Keller from Kennisland, and Pablo Uceda Gomez from Europeana, each of whom assessed 300 objects. The researchers checked whether the conditions for applying certain rights statements were met within the limitation of the research. Their assessment was based on the available data – such as creation date, publication date and the date of the death of the author – on Europeana and other sources such as the website of the data partner. The conditions on the use of the available rights statements are as follows:

In Copyright (InC):

- In Copyright statement must be applied to works that are still in copyright and have not been licensed.

Public Domain Mark (PDM):

- The Public Domain Mark (PDM) must be applied to digital objects that are out of copyright.

Public Domain Dedication (CC0):

- CC0 can only be used to waive all rights for a digital object and data when either the data provider holds these rights, or if the data provider has permission by the rights holders to do so.

Creative Commons license (CC):

- CC BY, CC BY-SA, CC BY-ND, CC BY-NC, CC BY-NC-SA and CC BY-NC-ND must be applied to a digital object that is in copyright with consent of the rights holder.

Other¹⁰:

- No Copyright - non commercial reuse only (NoC-NC) can only be applied to public domain digital objects which have been digitised as an outcome of a public-private partnership, where the terms of the contractual agreement limit commercial use for a certain period of time.
- Copyright not evaluated (CNE) is intended to be applied to digital objects of which the copyright status has not been evaluated.

We ensured accuracy and consistency of the results by a predetermined list of steps based on the information presented on the Europeana Database and possible additional information on the website of the Data Partners.¹¹ The researchers that conducted the individual assessment checked the first 30 assessments with each other to make sure the list of steps is followed up in a consistent way.

¹⁰ NoC-OKLR, InC-EDU and InC-EU-OW have not been applied to objects in Europeana yet.

¹¹ Addendum 2.

This list includes all relevant steps a researcher has to take and sets up basic rules to follow when determining the accuracy of a rights statement. We assess the accuracy of the following rights statements: PDM, CC0, CC BY, CC BY-SA, CC BY-ND, CC BY-NC, CC BY-NC-SA, CC BY-NC-ND, InC and NoC-NC.

The rights statements InC-EU-OW, NoC-OKLR and InC-EDU are not yet used in Europeana and therefore could not be found in the sample. We expected all NoC-NC statements to be accurate since Europeana is involved in applying this rights statement.

We divided the accuracy into five categories: accurate, inaccurate, questionable, not evaluated and unknown. In case of the first two categories, the researcher could determine the accuracy status based on the information provided by the data providers, as well as other available sources.

A rights statement was questionable when there was reasonable doubt on the accuracy based on the data provided. For example, a photograph of an object that seemed to be made for commercial purposes such as a fashion photograph of a famous photographer with a Creative Commons license applied to it. The researchers have justified why there was reasonable doubt on the accuracy of the rights statement by writing down their suspicion in the comment section.

Our sample also included 43 digital objects that are marked with an CNE rights statement. This rights statement signifies that the copyright status still needs to be evaluated, and as such, the digital objects with this rights statement cannot be evaluated in the same way as the digital objects with other rights statements.

Therefore, all CNE rights statements do not give any information on the copyright status of the work. The rights statements are accurate by their nature, if the rights holder had evaluated the copyright status of the work, it had applied a different statement. Thus, all CNE are categorised as 'not evaluated'. Unknown is applied when there is a lack of data to confirm anything about the rights status of the object.

3.3 Limitations

The researchers were only able to check whether the rights statement was accurately applied based on the information provided by the data provider. As a consequence, researchers were not always able to determine the accuracy. Especially in relation to Creative Commons and CC0 statements they were often unable to confirm whether the rights holder had given consent based on provided data.

It has been taken into account that the properties to express time in EDM are sometimes not specific enough to determine the date of creation. For example dc:date could indicate the date of creation or any other relevant date in the history of the object. Therefore, it was important to check in the source website to assess whether the date indicated is really the date of creation.

3.4 Analysis of results

After the individual assessments of the 1,500 digital objects was completed, the results were compiled and converted to graphs that display the accuracy. Subsequently we

analysed the results of the assessment of every category (accurate, inaccurate, questionable, not evaluated and unknown) examining the most common reasons for positioning a rights statement in every category.

4. Results

We gathered the data on 6 February 2017. In this chapter we will start by discussing the composition of the random sample of digital objects. Then we will look into the accuracy of these objects. We will follow this up by looking at the (possibly) inaccurate rights statements and their suggested alternatives. Lastly, we summarise our findings.

4.1 Composition of the sample

We were able to assess 1,462 digital objects between 13 April 2017 and 6 July 2017. We planned to assess 1,500 objects, but were unable to do so because we encountered 38 non working Europeana URLs at the time of the individual assessment. This concerns the URLs on the Europeana website that were automatically gathered using the script. This can occur for several reasons, mostly due to the fact that Europeana removed the web page (404 error).¹² We had to eliminate these objects from our research. In some cases we were still able to retrieve the rights statements of these objects automatically, however, it was not possible to verify this information since we could not visit the web page.

The division between the rights statements in Europeana and in the sample can be seen in the graph below. Most objects in the sample originated from the Netherlands, France, Germany and Norway which is also the case in Europeana. The rights statements in the sample correspond with the division of rights statements in the Europeana database as illustrated by Chart 4.

Sample and Europeana

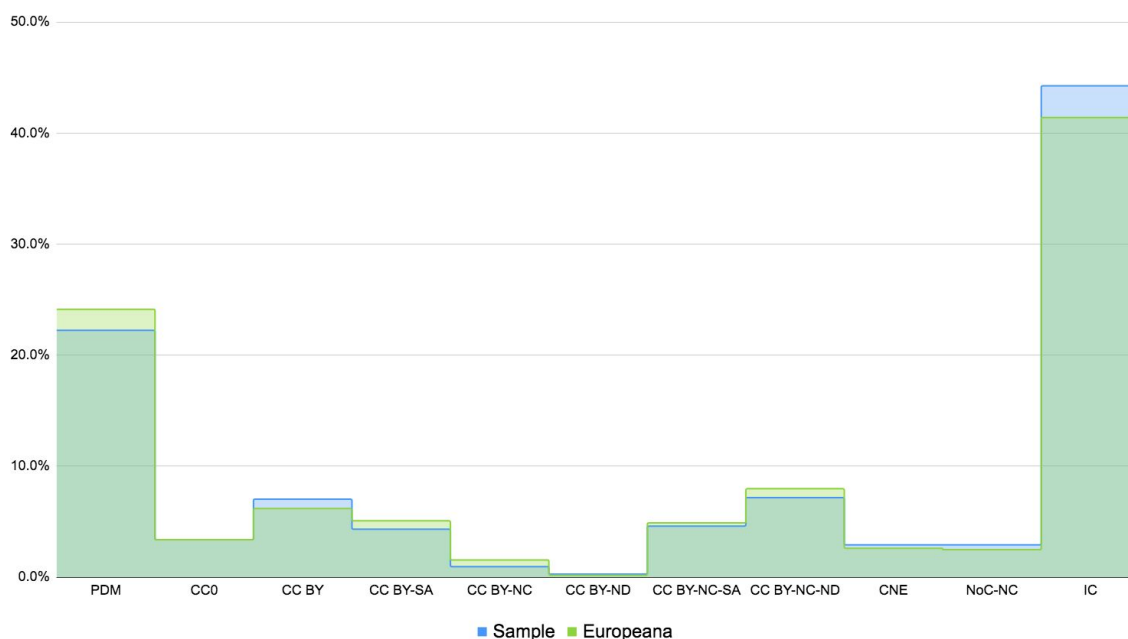


Chart 4. The division of rights statements in the sample (blue) and in Europeana (green).

¹² On 17 July 2017 some of the URLs worked again. Since the individual assessment period was already over, the object linked to these URLs have been discarded nonetheless.

4.2 Accuracy of rights statements

4.2.1 General remarks

After assessing the sample we found that the majority, 61.8%, of the rights statements was accurately applied, 9% of the statements were inaccurate based on the available information. Both of these numbers could be higher since the researchers were unable to determine the accuracy of the remaining 29% of the rights statements. The researchers were unable to determine the accuracy of 9% of the rights statements due to a lack of information about the digital objects. The accuracy of 17% of the rights statements was questionable, meaning that the researchers had reasonable doubt that a rights statement was inaccurate but were unable to confirm this based on the provided data. 3% of the rights statement have not been evaluated.

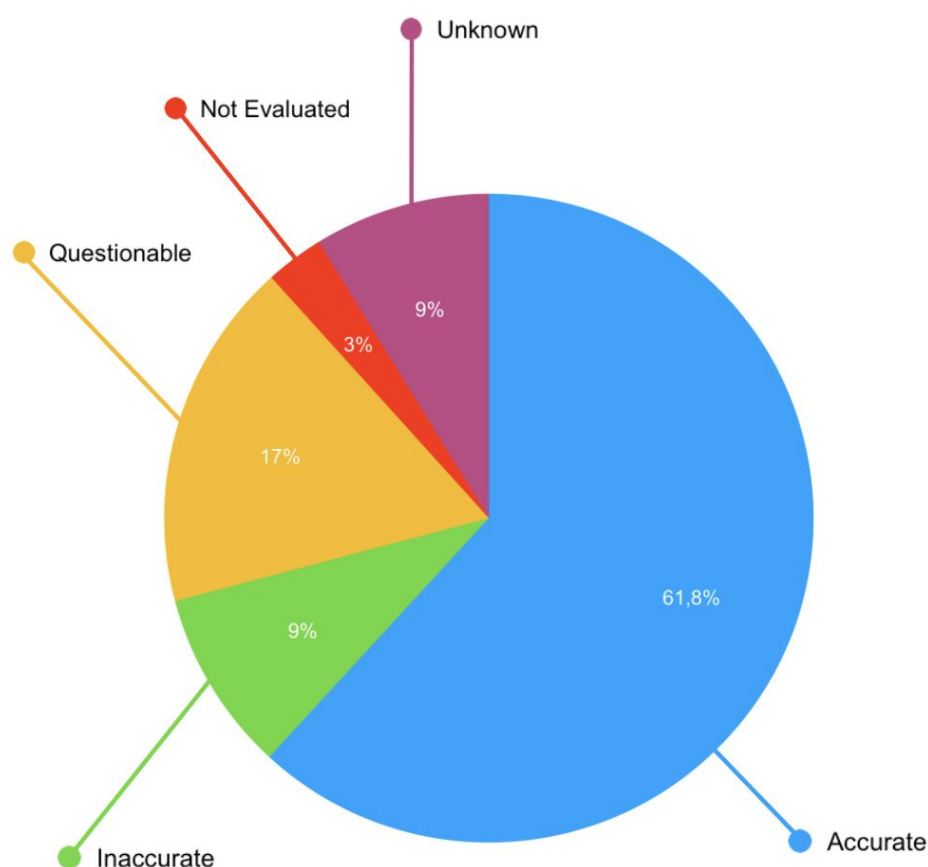


Chart 5. Results that display the accuracy of the rights statements.

4.2.2 Rights acquired through digitisation

The digitised version of works that were previously in the public domain are sometimes denied free use by the public as they are protected by (related) rights in Austria, Denmark, Germany, Iceland, Italy, Norway, Spain and Sweden.¹³ These rights can arise in a digitisation process, even when the process lacked any creative input. Rights

¹³ This list is based on the research of Thomas Margoni that can be found here: <http://outofcopyright.eu/>.

acquired through digitisation arise in any country when it concerns a 3D object, since the process of digitising 3D objects always results in making creative choices. Almost half of the digital objects that were analysed had received new rights through digitisation.¹⁴ The researchers took these rights into account when determining the accuracy of the rights statement.

As a matter of policy Europeana discourages data providers from exercising rights created during digitisation of Public Domain works.¹⁵ Europeana does, however, allow data partners to use In Copyright rights statements (including CC licenses) as such rights labels accurately reflect the situation in the above mentioned countries.

4.2.3 Category of statements

When examining the results of the individual assessment, it was helpful to divide the rights statements into six categories, because the reasoning behind the application of these rights statement(s) is similar. This enabled us to highlight large differences between the accuracy of rights statements. The six categories were:

1. In Copyright statements
2. Creative Commons Licenses
3. Public Domain Mark statements
4. Public Domain Dedication statements
5. No Copyright - non commercial reuse only
6. Copyright not evaluated

The first category contains In Copyright statements that need to be applied to works that are still in copyright and have not been licensed. The second category contains Creative Commons statements including CC BY, CC BY-SA, CC BY-ND, CC BY-NC, CC BY-NC-SA and CC BY-NC-ND. Creative Commons licenses can only be applied to works in copyright with consent of the rights holder. The third category contains Public Domain Mark statements (PDM). The fourth category contains Creative Commons CC0 1.0 Universal Public Domain Dedication (CC0) statements. The fifth category contains NoC-NC statements. And the last category contains CNE statements. The results are different when looking at the accuracy of objects per category (see chart 6).

¹⁴ 44,4% of the objects received rights through digitisation.

¹⁵ See the [Europeana Public Domain Charter](#).

Accurate, Inaccurate, Questionable, Not Evaluated and Unknown

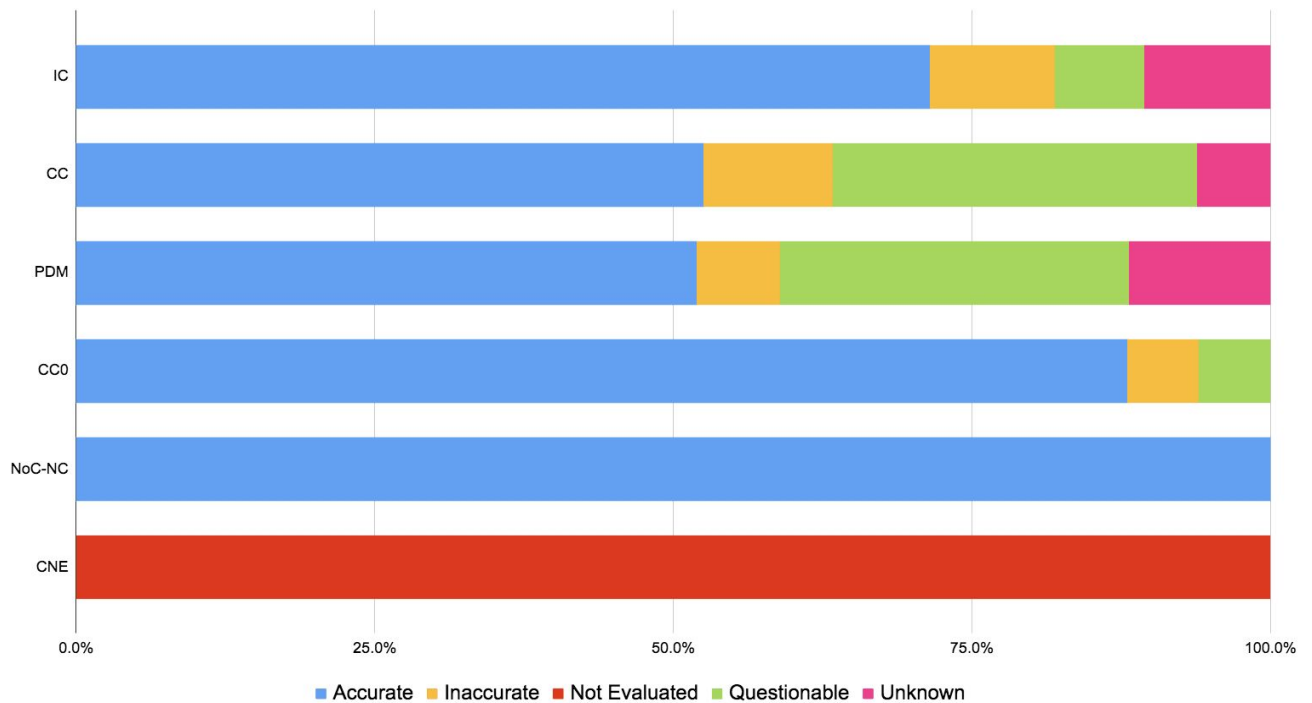


Chart 6. Accuracy of rights statements per category.

	Accurate	Inaccurate	Not Evaluated	Questionable	Unknown
IC	71.5%	10.5%	0.0%	7.5%	10.5%
CC	52.5%	10.8%	0.0%	30.6%	6.1%
PDM	52.0%	7.0%	0.0%	29.2%	11.9%
CC0	88.0%	6.0%	0.0%	6.0%	0.0%
NoC-NC	100%	0.0%	0.0%	0.0%	0.0%
CNE	0.0%	0.0%	100.0%	0.0%	0.0%

Table 1. Percentage of accurate, inaccurate, questionable, not evaluated and unknown rights statements.

Around 50% of rights statements of the objects with a Creative Commons license and a PDM statement are accurate, whereas 70% of the objects with an In Copyright statement and almost 90% of the objects with a CC0 dedication was accurately applied. Also, as expected, 100% of the Noc-NC statements are accurate. The copyright status of the digital objects with a CNE statement had not been evaluated.

The amount of objects with a rights statement of which the researchers were unable to assess its accuracy, labelled unknown in the graph, is roughly the same in the In Copyright and PDM group. 7.5% of the In Copyright statements and 6.0% of the CC0 statements were questionable, while around 30% of the CC and PDM statements were questionable.

When looking at a rights statement with a PDM or CC statement, users are as likely to encounter an accurate rights statement as they are to encounter a possibly inaccurate rights statement. Because of Europeana's policy of promoting the possibility to reuse these results are especially alarming. The PDM and CC categories are supposed to enable reuse and communicate the conditions hereof.

The IC group makes up 41.1% of the Europeana database, CC 25.8%, PDM 24.1% and CC0 3.4%.¹⁶ The accuracy is far better in the CC0 group, however, this is by far the smallest group.

4.3 Reasons for inaccuracy

Per inaccurate rights statement the researchers recorded why that statement was inaccurate and, where possible, what rights statement should have been applied. We have analysed these results to see whether certain rights statements have been inaccurately applied more often than others and to identify common mistakes.

4.3.1 Group 1: Inaccurate rights statements

There were 141 objects (9.6%) with an inaccurate rights statement. The researchers provided additional information and reasons for inaccuracy on these objects. They were able to suggest an alternative rights statement in 124 cases. After analysing this data we have found the following common reasons for the inaccuracy.

Creative Commons Licenses were most often inaccurately applied (10.8%). The most recurrent reason for this inaccuracy was the fact that the objects fell outside the scope of copyright, often meaning that the objects contained a presentation of factual information or a registry database.¹⁷ In these cases the researcher suggested to use the Public Domain Mark.

Other times there was no actual digital object. The 'object' consisted out of factual information or a link to factual information. Such material cannot be licensed at all, because it is never been in copyright.

Another common reason for the inaccuracy of CC licenses was that the object had entered the public domain. In this case the object had been in copyright. It is possible in some cases that the objects became part of the public domain after the rights statement was applied. Meaning that at the time of application the rights statement was applied accurately.

¹⁶ Based on the data in Europeana on 7 July 2017.

¹⁷ For example see:

http://www.europeana.eu/portal/en/record/2058621/loccloud_preg_b199b9f7_f635_44a6_b263_333eead0480d.html.

10.5 % of the In Copyright rights statements were applied inaccurately. The main reason for the inaccuracy is that the object had entered the public domain. This could have been the case when the rights statement was applied but it is also possible that the copyright term ended after the rights statement was applied and added to Europeana.

7% of the Public Domain statements have been found to be inaccurate. For this category one reason for the inaccuracy stood out: the copyright term of the object had not ended, therefore the object was still in copyright. 6% of the public domain dedications (CC0) was applied inaccurately. It was not possible to deduct recurring reasons.

	CC0	IC/CC	PDM
CC	1	0	37
CC0			2
IC			62
PDM	4	18	0

Table 2. Researchers were able to suggest alternative rights statements for 124 of the 141 statements. Inaccurately applied rights statements can be found on the left. Above the suggested rights statements.

4.3.2 Group 2: questionable rights statements

There were 17.4% rights statements of a questionable nature. The researchers provided additional information on these objects and reasons for the questionable nature. After analysing this data we have found the following recurrent reasons for the assumed inaccuracy.

Questionable Creative Commons Licenses

Around 30% of the CC statements were of questionable nature. It was to be expected that one of the reasons to qualify a rights statement as questionable was the probable lack of consent of the rights holder to license the work under the terms of a CC license. It is unlikely that permission was given when it concerns relatively old objects, especially when the rights holder(s) died before Creative Commons existed, or when the objects are from well-known commercial companies or professionals.¹⁸ In many cases there was not enough information available to determine whether consent was given, and given the above it is likely that a large proportion of CC licenses marked as questionable are in fact inaccurate.

The second most common reason for the researchers to consider a CC license as questionable was that copyright did not apply to the object as they were presented in

¹⁸ For example it is highly unlikely that the British American Tobacco group has given consent to license this image under a CC BY license: http://www.europeana.eu/portal/en/record/2021648/0247_7885.html.

Europeana.¹⁹ It was difficult to determine whether this was really the case because of number of reasons:

- The digital object presented consisted of multiple objects of which some could possibly obtain copyright. However, it was unclear what object or what element of an object the rights statement applied to.
- An object contained factual information and could be presented as such, but it could also be presented as an art composition, such as a photo of a dried plant with an academic description. From the object alone it is impossible to determine the intention of the author (if there is one). In these cases it could be very well that the creator of the object intended the object to be solely factual, but it also could be that the creator made creative choices and therefore had obtained copyright.
- The (non-digital) 'object' was a link, such as links to a place on a geographical map in another database.

The third most common reason why the researchers marked CC licenses as questionable was because it was likely that the copyright term had ended although this could not be determined with certainty because of the limitations of this research.

Questionable In Copyright rights statements

A small number, 7.5% of the objects with an In Copyright statement, was questionable. The most common reason was that the copyright term had ended. The second most common reason was that copyright did not seem to apply because the works were purely factual.

Questionable PDM statements

Almost 30% of the PDM statements were questionable. There were two recurring reasons for this. First, due to national legislation in some countries, non-original photographs receive rights through digitisation automatically, even when the original object was already in the public domain. If such legislation applies, the digital non-original photographs are in copyright, meaning that a rights statement should reflect that. If the rights holder wants to give these photos to the public domain, it should communicate so by giving it the public domain dedication (CC0). Second, some objects are possibly still in copyright.²⁰

4.4 Summary

The results show that at least 61.8% of the rights statements were accurately applied and at least 9.1% were inaccurate based on the available information. The accuracy of 17.4% of the rights statements is questionable, while for 8.8% of the rights statements it was not possible to determine the accuracy.

The results from the sample give a reasonable indication of the accuracy of the rights statements of the entire database of Europeana. The results show that there are two

¹⁹ Meaning that it was either unclear to what (digital) object the rights statement was attached to or that it was unclear whether the (non-)digital object presented in Europeana was copyrightable.

²⁰ For example, the author of this work was created in 1902 and it could be still in copyright:

http://www.europeana.eu/portal/record/9200365/BibliographicResource_3000094920085.html

categories – objects with a PDM statement and objects with a CC license – where the accuracy is significantly lower in comparison to the other categories. The accuracy in these two groups is around 50%. There are three common themes when looking at the results on the (presumed) inaccurate rights statements.

4.4.1 The term of copyright

The copyright term is often inaccurately applied resulting in works that are in the public domain with an In Copyright or CC statement, and in copyright works with an PDM statement. This can be due a number of reasons. First of all, copyright – including the copyright duration – is complicated and difficult to understand. Even though the copyright term is harmonised in the European Union, rules concerning authorship, publication and national exceptions vary among member states. This makes it difficult for institutions who are not specialised in copyright to accurately apply copyright rules.

In addition, copyright is not static. Copyright laws change over time, and time itself can change the copyright status of an object. This can cause incorrect rights statements even though the data provider has accurately applied the rights statement at the time.

4.4.2 Copyright not applicable

The results showed that some rights statements were inaccurate because the object was not covered by copyright.

These are either digital objects that represent facts or series of digital objects that are presented as one object, such as catalogues, databases or collections of geographical locations. These objects have always been in the public domain and should always be labelled with the Public Domain Mark. CC or IC statements applied to these objects are inaccurate.

Or, these are objects of which it can be questioned whether they belong in Europeana at all. In this case the 'objects' are links to collections of information. The Europeana Licensing Framework has not been designed with such 'objects' in mind and applying any of the rights statements supported by Europeana to them does not make sense semantically.

4.4.3 Presumable lack of consent

The results show that it is likely that there is a large amount of digital objects with CC Licenses that have been applied without consent from the rights holders. Creative Commons licenses can only be granted by (or with consent of) the rightsholders and as such a substantial proportion of the CC licences applied to objects available via Europeana are likely invalid. As a result the digital objects cannot be reused without prior permission from the copyright holders even though the rights statements imply otherwise.

These three themes should be placed in the context of the reality of the organisations that provide data and apply rights statements. Applying a rights statement is not an easy process. There are multiple steps a data provider needs to take. Not every data provider has the required procedure in place to acquire all necessary data on every object. Nor does every data provider have the required knowledge to be able to accurately interpret the complex rules that determine the duration of copyright and

related rights in the EU. Objects with a questionable rights statement or a statement of which the accuracy could not be determined were often accompanied by incomplete data.

The inaccurate application of the copyright term and Creative Commons Licenses are likely the result of a lack of reliable metadata on the objects which will often be connected to a lack of resources, relevant knowledge on copyright and the understanding of the benefits of applying rights statements accurately.

5. Recommendations

The accuracy of the rights statements in Europeana is a metadata-quality issue and should be seen in the wider context of discussions about metadata quality related to Europeana. Europeana has a strong interest in delivering reliable, trustworthy information about the digital objects available on its platform. This is not limited to rights information but also includes other metadata such as publication dates, object descriptions, geodata and information about authors and other creators.

Europeana can undertake activities that can help improving the quality of the rights statements through its policy documents and training materials, but cannot by itself completely solve the underlying problems concerning the inaccuracy of the rights statements. The accuracy of rights statements is in first instance the responsibility of the data providers.

We have set forth a number of recommendations for Europeana to improve the accuracy of the rights statements, and they can help open up a dialogue with data providers about how to improve the accuracy of rights statements. These recommendations focus on copyright issues. However, it should be noted that improving the accuracy of rights statements will need to go hand in hand with improving the quality and availability of other metadata such as information on publication dates and reliable information about authors and other creators.

5.1 The public domain and non-original and non-digital objects

A number of [Europeana's IPR resources](#) primarily focus on objects to which copyright is applicable.^{21 22} For instance, the Europeana selection tool started off with the question 'Is the cultural heritage object still protected by copyright?' assuming that there is a (digital) object and that this object has been in copyright. These assumptions do not always apply and are not in line with the [Public Domain Charter](#).

Europeana wants to be inclusive and 'in principle publish all types of content that museums, libraries, archives, and audiovisual institutions based on their digital strategies consider to be important enough for publication in Europeana'.²³ This includes objects that only contain information. Europeana acknowledges as well that it is possible that cultural heritage institutions sometimes will want to publish information on Europeana that is not necessarily connected to a physical or digital object. These are defined by Europeana as non-digital objects.^{24 25}

The results from this research show that these objects, factual information in the form of a digital object and non-digital objects, are often given the wrong rights statements,

²¹ The [Europeana Publishing Guide](#) (p. 25) does refer to the [Public Domain Charter](#) but does not specifically include the relation between copyright and facts, ideas and information.

²² As well as the description of the [available rights statements](#).

²³ [Europeana Content Strategy](#), p. 19.

²⁴ [Europeana Publishing Guide](#), p. 16.

²⁵ Although one could argue that it would be better to call these "digital non-objects", as that what is on Europeana is not an object but the 'element' the rights statement is applied to.

because they are given an in copyright rights statement, and/or in the case of non-digital objects they are not connected to any digital or physical objects.

Europeana should consider making sure that all policy documents and information resources clearly indicate that copyright does not apply to facts, information and ideas - they are part of the public domain. Objects (analog or digital), that are made up exclusively of information and/or facts may lack originality and would thus not be protected by copyright.

In addition, the available information and policy on what rights statement to apply to non-digital objects is unclear. Europeana should construct a clear policy on non-digital objects in which it becomes clear how they should be labelled.

5.2 Training, materials and community

Information on how to apply the rights statements and relevant copyright and related rights legislation alone is not enough. It is not possible to improve the quality of rights statements without also improving the quality of the other metadata relating to objects in the collections of cultural heritage institutions. This kind of information should be accompanied with how data providers can implement this in their internal procedures when dealing with digitised content and why it is important to make IPR part of their digital strategy.

In order to achieve this Europeana could include trainings and workshops on applying rights statements and understanding IPR challenges as an integral part of events. As well as raising awareness on copyright challenges during events and sessions for example on data quality where copyright and related rights is not the main topic.

The current training materials should be complemented with information on how to apply rights statements to non-digital and unoriginal objects. Most current training materials of Europeana do not explain in detail what kind of objects fall outside the scope of copyright. It seems that data providers do not understand that objects like these have never been in copyright.

Data providers could benefit from a well-developed copyright community. Europeana could look at providing more support for this community to help it develop into a network where its members can discuss recurring issues and share relevant information among each other. Data providers can use the copyright community as a place for dialogue with Europeana. Europeana can gather input from data providers, especially on whether they are aware of Europeana's IPR resources and how they perceive these resources. Such a dialogue could enable Europeana to improve the accessibility and findability of its existing resources, as well as give Europeana insight on how to improve their resources. For instance it would be useful to understand the policy choices of data providers concerning the application of rights statements of data providers as it can also be that data providers have specific reasons to apply certain rights statements that do not correspond with Europeana's policy on the application of rights statements.

5.3 Dynamic and complicated nature of copyright

As mentioned in the previous chapter copyright is intricate and complex. Even though it is important to apply the current rules accurately, it is equally important to critically examine the current copyright framework. Therefore, Europeana should continue advocating for a more simplified copyright with a shorter harmonised copyright term and continue to oppose legislation that creates new rights during digitisation of public domain works.

In addition to copyright being complicated, it is also dynamic. A certain number of rights statements will become inaccurate over time, because they will enter the public domain. Europeana should develop a long-term policy how to deal with this group of objects. Europeana should consider, with cultural heritage institutions, requiring data providers to provide the date the object will enter the public domain. This would make it possible to keep track of objects that will enter the public domain after ingestion. It should be noted that national copyright laws can change as well.

Another possibility to deal with the dynamic nature of copyright, is to enable users to file errors. Other websites such as www.rijksmuseum.nl have made it possible for its visitors to communicate errors as well as contact the Rijksmuseum if they have additional information about the objects. The current feedback option in Europeana could be placed differently and changed in such a way that it is clear on what users can give feedback about. Such a mechanism would be able to improve all information on Europeana.

5.4 CC licenses and PDM

We recommend that Europeana together with its data partners make efforts to strengthen the ability of the data partners to correctly apply Creative Commons licenses to Digital Objects that they make available via Europeana. This should include developing procedures that allow for the detection of incorrect rights statements prior to publication in Europeana. The procedures for PDM should be reviewed and strengthened where possible.

5.5 Further research

To fully understand the reasons for the submission of inaccurate rights information it would be beneficial to conduct further research on the underlying reasons of the inaccuracy of the rights statements. In particular to the accuracy of the application of Creative Commons statements and how data providers deal with the need to obtain consent of the authors and other rightsholders for publication under a CC license.

In addition, further research should be conducted into the issues posed by non-digital objects and how they can be dealt with in the context of the Europeana Licensing Framework which has not been designed with these types of objects in mind. It is currently unclear whether a rights statement can be applied to something that is not an object, and/or whether when a rights statement is applied, this rights statement is tied to the object the non-digital object refers to or to the non-digital object itself. This ambiguity needs to be resolved.

Addendum 1 | Scripts

Kennisland developed three scripts to gather the information that was processed in the research. The scripts are in the sections below. They are written in Python 3 and released under an [EUPL V1.2 license](#).

1.1. Get 1500 random urls

```
#-*- coding: utf-8 -*-
# encoding=utf8
import requests, os, urllib, xmltodict
from random import randint

#Fill in these data items to determine how many objects it is randomly going to
select

#TOTAL OBJECTS in EUROPEANA
totalEuropeanaObjects = 54153588
#AMOUNT OF OBJECTS TO GATHER
amount = 1500

#STATIC, unless XML files change
objectsPerFile = 45000

objects = []
for count in range(0, amount):
    item = randint(1,totalEuropeanaObjects)
    while (item in objects):
        item = randint(1,totalEuropeanaObjects)
    objects.append(item)
#order the list to minimise the amounts of file needed to be download.
objects.sort()

#Print objects so in case of problems we can just use the export and data
gathered so far.
print(objects)

for count in range(0, totalEuropeanaObjects // objectsPerFile) :
    start = objectsPerFile * count
    end = start + objectsPerFile

    #print page, XML file that is being progressed/
    print ("PAGE", start, end);

    #reset file
    data = None

    #while still objects to be processed and the objects can be found in the
current 'page'
    while len(objects) > 0 and objects[0] >= start and objects[0] <= end:

        #Only download the XML if it is the first time you need this file
        if data is None:
            query =
"http://www.europeana.eu/portal/europeana-sitemap-hashed.xml?from=" +
str(start) + "&to=" + str(end)
            result = requests.get(query)
            data = xmltodict.parse(result.content)

        #take the first object of the list
```

```

        item = objects[0]

#OPTIONAL print number in range 0 - total objects
        print ("ITEM", item)

#OPTIONAL print progress 0- amount
        print ("COUNT", amount - len(objects))

#itemnumber is the item number mapped to 0-45000
        itemNumber = item % objectsPerFile

#print url of that number
        print(data['urlset']['url'][itemNumber]['loc'])

#delete the object from the
        del objects[0]

os.system('say "your program has finished"')

```

1.2. Use regular expressions to transform URLs to list of Europeana IDs

On the output of the above script do the following regular expression. We used grep in TextWrangler:

[http://www.europeana.eu/portal/record\(.*\)\.html](http://www.europeana.eu/portal/record(.*)\.html)

and replace with

\1

To obtain a list of Europeana IDs, use these in the next script.

1.3. Get additional relevant information

```

#-*- coding: utf-8 -*-
# encoding=utf8
import sys, requests, json, re, os, time, urllib

# SET YOUR API KEY HERE
key = "<KEY>"
api = "http://www.europeana.eu/api/v2/search.json?"

#List of all Europeana IDs (regexped from the previous script)
objects = []

for object in objects:
#get information for one object with Europeana ID (id)
query = api + "query=europeana_id%3A\"" + object + "\"&rows=10&wskey=" + key
    result = requests.get(query)
    result = result.json()
    count = result['totalResults']

# Test whether the ID is unique or that I made a mistake getting the ID from
the URL.
    if (count == 1):
        number = object
        object = result['items'][0]

# A series of test to see if the information exists. Otherwise us an empty
string
    if 'title' in object:
        title = object['title']
    else:

```

```

        title = ' '
    if 'country' in object:
        country = object['country']
    else:
        country = ' '
    if 'edmIsShownBy' in object:
        edmIsShownBy = object['edmIsShownBy']
    else:
        edmIsShownBy = ' '
    if 'provider' in object:
        provider = object['provider']
    else:
        provider = ' '
    if 'dataProvider' in object:
        dataProvider = object['dataProvider']
    else:
        dataProvider = ' '

    if 'rights' in object:
        rights = object['rights']
    else:
        rights = ' '
    if 'type' in object:
        type = object['type']
    else:
        type = ' '

#print all found information with tabs in between for easy copying into a
spreadsheet.
    print (number, "\t", title, "\t", country, "\t", edmIsShownBy, "\t",
provider, "\t", dataProvider, "\t", rights, "\t", type )
    else:
#if error in unique id print 'not singular'
    print(object, "Not singular")

#Communicate that the script is done.

    os.system('say "your program has finished"')

```

Addendum 2 | Individual assessment of digital objects from the Europeana Database

The sample size of 1,500 URLs to digital objects in Europeana database can be found in this [document](#). It contains the following information: the URL to the Europeana record, title, country, edmIsShownBy, provider, dataProvider, rights, edm:type, and author. You need to assess whether the rights statement (RS) applied to the digital object is correct or incorrect. It is not always possible to determine the accuracy of the rights statement. In this case you should state that the accuracy of the rights statement cannot be determined. Be consistent in your wording in order to facilitate the analysis and use the dropdown menu whenever available.

2.1 Fill in the document by following these steps

Step 1

Click on the link and go to the digital object in this document.

Step 2 Column J

Take note of the rights statement that is applied to this digital object in the Europeana Database.

Step 3 Column K

Fill in the date of creation of the object. The date of creation is the date the (physical or digital) object was created. There is an exception: if new rights were created through the digitisation process of physical object, the date of creation is the date the digitisation rights were acquired. If the date of creation is not mentioned on Europeana, you can follow through to the website of the provider or other relevant websites. Mention the websites you have visited in column S 'remarks'. If you are unable to find a date, but can find the period such as the Renaissance or time frame such as '16th century' in which the object is created, write it down. If the date of creation remains completely unknown, write 'UN'.

Step 4 Column L

Fill in the date of the death of the author. The author is the creator of the physical work. There is one exception: if new rights were created through the digitisation process, the author is the author of the digital object. If the death of the author is not mentioned on Europeana, you can follow through to the website of the provider or other relevant websites. Mention the websites you have visited in column S 'remarks'. If the date remains unknown, write UN.

Step 5 Column M

Choose whether there is a possibility that rights are created by digitisation in case of semi-automated digitisation or human operated digitisation of the object in the drop-down menu or a digitised 3D object. If there is a possibility rights were created

through digitisation, specify the country of the heritage institution where the digitisation took place resides. In case of a 3D object, fill in 3D object.
Helpful tools in deciding if such rights were created are:

- <http://outofcopyright.eu/rights-after-digitisation/>
- <http://outofcopyright.eu/wp-content/uploads/2015/02/Thomas-Margoni-ExecSummary-Rights-created-during-ditigisation.pdf>

Step 6 Column N

Choose whether the rights statement is correct taking into account the basic rules for assessing the accuracy of a rights statement. You can choose 'yes', 'no', 'unknown' or 'questionable'. In case of 'yes' you are done and can proceed to the next object. If you have chosen 'no' continue to step 7. In case of lack of information you have to choose 'unknown', proceed to step 8 and explain why you were unable to determine the accuracy of the rights statement. If you feel that it is possible to interpret the information of the object in different ways, you have to choose 'questionable' and add a remark on column N in which you explain the multiple interpretations. Place a comment in this column if you think someone else needs to have a look at it.
Helpful tool in calculating whether a work is in the public domain:

- <http://outofcopyright.eu/calculators/>

Step 7 Column O

Even though some objects have digitisation rights attached to them, sometimes objects with digitisation rights are used as an excuse to restrictively license objects where this is not needed or wanted. Therefore, you need to indicate in column O whether the chosen rights statement is in line with the Public Domain Charter. You can choose between 'yes', 'no' or 'unknown'.

The principles of the public domain charter are:

1. Copyright protection is temporary.
2. What is in the Public Domain needs to remain in the Public Domain. Exclusive control over Public Domain works cannot be re-established by claiming exclusive rights in technical reproductions of the works, or by using technical and/ or contractual measures to limit access to technical reproductions of such works.
3. The lawful user of a digital copy of a Public Domain work should be free to (re)use, copy and modify the work.

You can find the Public Domain Charter here:

http://pro.europeana.eu/files/Europeana_Professional/Publications/Public%20Domain%20Charter%20-%20EN.pdf

Step 8 Column P

Suggest the correct rights statement taking into account the basic rules for assessing the accuracy of a rights statement.

Step 9 Column Q

Add additional remarks such as which website you have consulted, whether there was important information lacking, why you were unable to determine the accuracy of the rights statement and/or other remarks.

Step 10

In case of recurrent questionable objects that fall within the same category, send Judith a message concerning these objects. She will gather relevant information about the cases and formulate how we will consistently assess these objects.

2.2 Basic rules for assessing the accuracy of rights statements

- If the last surviving author has died more than 70 years ago and there are no digitisation rights in the country of which the object originates from, the object is in the public domain (PDM).
- If the last surviving author has not yet died or has not died more than 70 years ago, the object is in copyright. It is possible that the object has a CC, CC0 or IC statement. Without a specific reason indicating otherwise you can assume this is correct.
- If an author has died more than 70 years ago and there are digitisation rights in the country of which the object originates from it is possible that the object has a CC license or CC0. Without a specific reason indicating otherwise you can assume this is correct.
- If an author has died more than 70 years ago and there are no digitisation rights in the country of which the object originates from, a CC0 dedication is incorrect.
- If the object, such as a video, is protected by related rights or neighbouring rights, the object is protected for 50 to 70 years after creation. It is possible that the object has a CC, CC0 or IC statement.
- NoC-NC statements can only be applied with approval of Europeana. Therefore, to check the validity of the object, ask Europeana whether the list below corresponds to their list:
 - Bayerische Staatsbibliothek
 - Österreichische Nationalbibliothek - Austrian National Library
- CNE should be applied to digital objects where the copyright status has not been evaluated. We can only establish whether rights statements are correct after they have been evaluated. You do not have to fill in anything other than column L.