

Research Report:

A look at the accuracy of edm:rights in the Europeana Database

How accurate are the rights statements for the images of the highest quality?



Kennisland

Table of Contents

1. Introduction	2
2. Methodology	4
3. Overview of analysis results	6
4. Closer look	7
Accurate collections	7
Inaccurate collections	7
Mixed, mostly accurate collections	9
Mixed, mostly inaccurate collections	10
Unable to determine	10
Principles of the Public Domain Charter	11
5. Conclusion	12
Sharing Information	13
Contact information	14

1. Introduction

[Europeana](#) is the digital platform for Europe's heritage. It collects over 58 million digital objects that come from thousands of heritage institutions across the continent. These 58 million objects have a rights statement assigned to them, making clear what a user can and cannot do under the relevant copyright laws. These statements can be chosen from a list of 14, as described in the [Europeana Licensing Framework](#). This list includes the industry-standard [Creative Commons licenses](#) as well as a selection of the rights statements developed in the [rightsstatements.org](#) consortium, which included Europeana, the [Digital Public Library of America](#) and Kennisland.

When searching through Europeana, one may encounter objects that have a rights statement that seems inaccurate. The platform 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 improve 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 partners.

In this context Kennisland undertook [research in 2017 into the accuracy of the rights statements contained in the Europeana database](#). Verifying the accuracy of rights statements is, and was, a difficult process. Because Europeana contains over 58 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 at this time 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 unavailable. 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.

The 2017 study by Judith Blijden (Kennisland) revealed an indication of the rights statement accuracy in the database by generating a list of 1,500 random digital objects which were individually assessed for accuracy by a team of experts. While the outcomes of the 2017 study was welcome and worthwhile, the outcomes were difficult

to follow up on because the digital objects were random and came from many different collections.

This follow-up study is more limited and aims to create a benchmark for the accuracy of an important subset of the Europeana database: the image collections in tier 3 and 4. These are the images that have the highest quality in terms of size, and have rights statements that allow for some form of reuse. They most commonly use either Creative Commons licenses or one of the Creative Commons Public Domain Tools. Our assumption is that these are the digital objects that are interesting for reuse by end users of the Europeana platform and other platforms that share open culture that can make use of the Europeana API.

In this research paper we first describe the methodology used to determine the accuracy of the images in the tiers 3 and 4 of the Europeana dataset in December 2018. Then we give an overview of the results, afterwards we delve into the specific categories and collections. And, finally we end with a list laying out the most common mistakes made when applying rights statements to the dataset researched.

2. Methodology

The accuracy of edm:rights in image collections of tier 3 and 4¹ of the Europeana dataset in November 2018 was determined by Kennisland following research on a [dataset](#) provided by the Europeana Foundation. The dataset was created based on an analysis of the largest 500 collections in the Europeana dataset. The tier estimate was based on checking between one and three items per collection. Since collections can be very heterogeneous it was noted by Europeana that it is possible that some tier 1 and 2 items were part of this dataset, as well as missed collections with tier 3 and 4 material. The set as received adds up to 13,129,742 digital objects of the 58,024,364 digital object Europeana database as of early December 2018.

[During analysis several collections were removed](#) from the initial collections list. First collections that held only text files and collections that had issues with edm:rights values that changed upon viewing were marked with red. Secondly, marked in blue, other collections that were determined to hold no digital objects in tier 3 or 4 were removed from the list, as well as collections that held no images.² These changes were expected due to the nature of the initial dataset as described above.

The total amount of tier 3 and 4 images in the Europeana collection is 3,986,338. We analysed the collections that accounted for 67.59% of the total amount of possible digital objects in those tiers, the final number of digital objects analysed being 2,694,219.

The method of analysing is an adaptation from the method used for the [previous research](#) on the accuracy of rights statement in Europeana. That research was performed on an per-item basis, this research was performed on a per-collection basis. Therefore, the accuracy determination began with the question whether the collection is homogenous or not. If the collection was homogenous, the accuracy was determined based on viewing between three to five items, if the collection was not homogenous the accuracy was determined by viewing between five to ten items per collection.

The second significant deviation from the previous research undertaken was the option to not only mark a collection as the edm:rights value being not only 'accurate', 'inaccurate', or 'unable to determine' but also as 'mixed, mostly inaccurate' and 'mixed, mostly accurate'.

¹ For a digital object to qualify for tier 3 and/or 4 it needs to have a direct link to the object, it needs to be of a certain size and it needs to have an edm:rights value that allows for some reuse. For more detailed information, please refer to the [Europeana Publishing Framework](#).

² Note that this method excluded digital objects that did not have the technical metadata to be filtered as tier 3 and 4 on the front end of the platform (our reference point for this research). If the technical metadata was to be added it could contain tier 3 and 4 images.

Please note that determination of accuracy of the edm:rights value on a collection basis was made by researcher [Lisette Kalshoven](#) (Kennisland) based on her knowledge of the EU copyright framework, the existing metadata by digital objects and a semi-random selection of digital objects within collections. The researcher's knowledge of the EU Copyright Framework was supported by the [OutOfCopyright.eu](#) website and underlying research. Because of the fragmented nature of copyright and related rights within the European Union the determination of accuracy is an indication, and should be treated as such.

3. Overview of analysis results

In determining accuracy we look at two different values: accuracy in terms of the amount of collections and accuracy in terms of amount of the digital objects those collections represent. These values have been expressed in a pie chart and table below.

	# collections	Percentage of total digital objects
Accurate	23	37.66%
Inaccurate	10	19.11%
Mixed, mostly accurate	5	16.07%
Mixed, mostly inaccurate	3	2.18%
Unable to determine	15	24.98%

Figure I: Accuracy of edm:rights values of tier 3 and 4 images expressed in number of collections as well as the percentage of digital objects these collections represent.

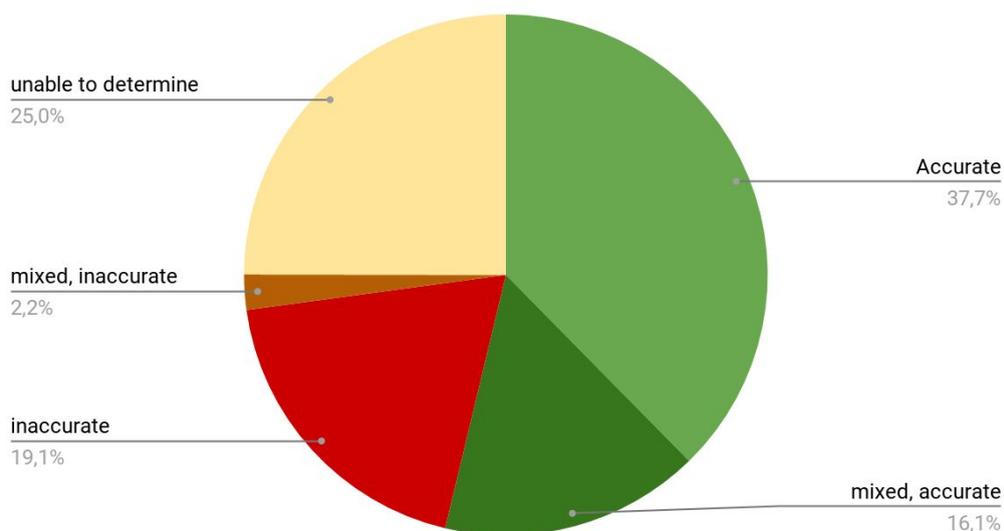


Figure II: Accuracy of edm:rights values of tier 3 and 4 images as expressed in the percentage of digital objects they represent.

4. Closer look

In this section we will delve into the separate accuracy indicators used above, and see whether any significant points can be deduced from those selections.

Accurate collections

Of the twenty-three collections deemed to be 'accurate' three of the five largest are from the OpenUp! collections. In total nine of the twenty-three are from this project. Of the just over a million digital objects that were marked as accurate, almost half a million come from these collections (458,456 digital objects out of 1,014,666).

Of the other collections in this subset it is notable that they all appear to be very old collections of which the objects are likely in the public domain, and therefore more easy to clear rights for. This is not to say that these collections are always marked with the PDM, as several of the collections come from countries with some form of digitisation rights.

Inaccurate collections

When we take a closer look at the ten collections determined to have inaccurate edm:rights values, none of the European countries stand out. The only country that can be found twice is the Netherlands, and since the country is a big contributor to the Europeana Database this is not surprising. Interestingly the only collection that has "Europe" as a country of origin (the 1914-1918 collection) was also determined to be inaccurate. The collection was created through crowdsourcing, and given the [mandatory edm:rights value for User Generated Content \(UGC\)](#) on Europeana, CC BY-SA. It appears this approach can lead to inaccurately labelled collections.

Interesting to note is that with the exception of the two smallest collections, all collections in this list were determined to be non-homogenous. Which suggests that the more complex the collection, the less likely the edm:rights value is correct.

Collection	Size of collection	Reason for determining as inaccurate
The Royal Botanic Garden (Edinburgh, Scotland)	397,875	Objects never in copyright (plant material) and no digitisation rights in the UK.
Forum Hungaricum (several institutions) (Hungary)	61,508	Several objects that should be in the PD with CC licenses, several CC licenses unlikely to have been used appropriately.
Architecture and Design Centre (Sweden)	24,943	Mislabelled still in copyright works as Public Domain.
Archief Eemland (The Netherlands)	13,605	Mixed set with unlikely CC licenses, on their website they claim copyright has lapsed.
Deventer Musea (The Netherlands)	8,685	Unlikely that the CC licenses are correct due to the content matter photographed (such as Disney material).
Folklore and Ethnological Museum of Macedonia-Thrace (Greece)	4,207	Mixed collection in which it is unlikely that a CC license is applied correctly given the multiplicity of rightsholders. Also probable PD material licensed under CC.
Europeana 1914-1918	2,500	Unlikely correct use of CC licenses due to the complexity of the set and the multiplicity of potential rightsholders.
Biblioteca Digital Memoria de Madrid (Spain)	1,570	Unlikely correct use of CC licenses due to the complexity of the set and the multiplicity of potential rightsholders.
Digital Library of UMCS (Poland)	7	Unlikely PDM as they are photographs from the 1970's.
Map Collection of the Faculty of Science, Charles University in Prague (Czech Republic)	6	19th-century maps under a CC license, should likely be labelled as Public Domain.

Figure III: Table of collections determined to use the incorrect edm:rights value.

Mixed, mostly accurate collections

This subsection only consists of five collections and just over half a million digital objects. Three of these five originate from Sweden. Most interestingly is perhaps the Rijksmuseum collection, whose subset of CC BY-NC-ND material (661 objects) is questionable due to unknown authors of mid-20th-century photography but is known to be a reliable partner with regards to Public Domain material.

In the other collections in this subset is worthy to note that the inaccurate parts of the collection also mostly have to do with photography from the mid to late 20th century with a Creative Commons license, while the author is unknown or unlikely to have been able to give permission for the license.

Collection	Size of collection	Reason for determining as 'mostly accurate'
Rijksmuseum (Netherlands)	315,186	Most of the collection is correctly labelled as Public Domain. The CC BY-NC-ND material (661 objects) is questionable as sometimes the author is unknown.
Länsmuseet Gävleborg (Sweden)	104,422	Collection likely claims digitisation rights over the mostly PD material. The more recent material is unclear as it is unlikely that the author was able to give permission for a CC license.
Flygvapenmuseum (Sweden)	12,581	Unknown photographers with photos from the 1940's with a PDM. Unlikely that rights could have been cleared to confirm PD status.
National Library of Romania, National and University Library St. Kliment Ohridski, Library of the "Lucian Blaga" University of Sibiu (diverse countries)	732	Mixed collection from several countries. The PD material seems to be accurate. Unlikely use of CC license application on material from the 1970's. Cyrillic script makes it difficult to analyse.
Etnografiska museet (Sweden)	4	Maps from early 20th century to mid-20th century. One of the maps will be Public Domain starting January 2019, Public Domain for the later maps unlikely.

Figure IV: Table of collections determined to mostly use the accurate edm:rights value.

Mixed, mostly inaccurate collections

This subsection also only consists of three collections and just over 58,731 digital objects. The mistakes seem to stem - just like the closer look above - through the use of CC licenses for objects while permission for that use is unlikely such as 20th-century photographs.

Collection	Size of collection	Reason for determining as 'mostly inaccurate'
Sjohistoriska (Sweden)	38,327	Photos from the 1950's with a PDM.
Marinmuseum (Sweden)	11,012	Photos from the 1960's with a PDM (possibly from a government agency which could make them accurate).
KU Leuven, Europeana Photography, (Belgium)	9,392	Majority of the PDM part of the collection is accurately labelled, several anonymous photos from the 1920 with unlikely CC licenses.

Figure V: Table of collections determined to mostly use the inaccurate edm:rights value.

Unable to determine

This is the most interesting subsection of the analysis. It consists of fifteen collections, and the reason for marking as 'unable to determine' is more often than not that CC licenses were applied while it is possible they were applied correctly but not enough information was made available. This is the case for the largest collection in this set (Uppsala Museet, Sweden, 148,493 digital objects). What differs these collections from those labelled as 'inaccurate' is that the collections mostly consist of a few named photographers who could have been alive to have given permission for CC licenses.

It would be interesting to take these fifteen collections and enquire after the reasoning behind the choices for edm:rights values to be able to put them into the other categories such as accurate or inaccurate.

Of the fifteen collections in this subsection nine come from Sweden, mostly the Swedish National Heritage Board. What complicates the matter in the Swedish collections is that it is also unclear whether they use CC licenses on material because of digitisation rights (that are possible in Sweden). Or because they believe the underlying material is still in copyright, and either they hold the copyright or they have been given permission to license under CC. The Swedish collections in this subsection have 409,992 digital objects in them.

Other notable collections in this subsection are two regional archives from the Netherlands (Gooi en Vecht and 's Hertogenbosch), which both apply CC BY-SA on collections in which it is questionable, but not impossible. It would be advisable to verify the reasoning with these collections and better categorise the just over 80,000 digital objects they represent.

Lastly the Wellcome Trust collection is worth noting, as it is perhaps the most complex set from all collections analysed because different CC licenses and the PDM have been used without apparent logic with digital objects that span centuries.

Principles of the Public Domain Charter

Europeana - and Kennisland as well - support a healthy and thriving public domain. It supports the [Public Domain Charter](#) with the principle that the digitisation of Public Domain content does not create new rights over it: works that are in the Public Domain in analogue form continue to be in the Public Domain once they have been digitised.

Of all fifty-six collections fourteen were determined to follow the principles of the Public Domain Charter, sixteen were unable to determine and twenty-six appear to not follow the principles of the charter. Of those not following the principles twenty still had accurate edm:rights values, but mostly had some form of digitisation rights in their countries (like several from the OpenUp! Collection). The other collections are a mix between collections of which the accuracy was unable to determine (but some of the underlying material had never been in copyright) or were very diverse collections in which at least in some of the collection the charter was not applied.

5. Conclusion

The determined accuracy rate of 37.66% appears quite low. This number can give a skewed image, however, since the category of 'unable to determine' is high at almost 25%. Despite this, it can be clearly stated that the accuracy of edm:rights in the image collections in tier 3 and tier 4 of the Europeana database is not as high as preferred. From the analysis above we can conclude that the most common mistakes made by data providers are:

- Applying a Creative Commons license to material while it is highly unlikely that permission for such a license has been given;
- Applying a Creative Commons license to material that was never in copyright and when there are no digitisation rights in the relevant country;

To a lesser degree:

- Applying a Creative Commons license while the material is Public Domain;
- Applying a Public Domain Mark while material is still in copyright.

Furthermore, in order to bring the Europeana collection more in line with the [Europeana Public Domain Charter](#) the common mistake is:

- Applying digitisation rights over material that was either never in copyright or of which copyright has lapsed.

Sharing Information

This publication was developed under the DSI-4 programme which was co-financed by European Union's Connecting Europe Facility and European Union Member States.



Co-financed by the European Union

Connecting Europe Facility

Kennisland created this publication and advocates sharing.

The sole responsibility for this publication lies with Kennisland. Neither the European Commission, nor any person acting on the European Commission's behalf, is responsible or liable for the accuracy or use of the information in this publication.

Unless differently stated, everything in this work is licensed according to the Creative Commons Attribution 4.0 International license (CC BY 4.0). This means that you can copy, edit and distribute this work, including for commercial purposes, as long as you use the correct attribution to the maker, provide a link to the license (<http://creativecommons.org/licenses/by/4.0/>) and indicate whether the work has been edited.

Cover picture: '[Young woman with a fan](#)' | Simon Maris | Rijksmuseum | Public Domain



Please use the following literature reference:

Kennisland (Lisette Kalshoven), A look at the accuracy of edm:rights in the Europeana Database - How accurate are the rights statements for the images of the highest quality? (2019) [CC BY 4.0], <http://creativecommons.org/licenses/by/4.0/>.

Contact information

Kennisland

De Ruijterkade 128
1011 AC Amsterdam
+31205756720

www.kl.nl

IBAN: NL62RABO0168780976

Swift: RABONL2U

KvK: 33304517

Btw: NL 807961395 B01