Project Acronym: EAwareness
Grant Agreement number: 297261
Project Title: Europeana Awareness

D5.1: Online Advocacy Kit for increasing the amount of content in Europeana that can be freely reused

Revision	3
Date of submission	03 January 2013
Author(s)	Paul Keller, Maarten Zeinstra (Kennisland), Julia Fallon (Europeana Foundation)
Dissemination Level	Public

REVISION HISTORY AND STATEMENT OF ORIGINALITY

Revision History

Revision No.	Date	Author	Organisation	Description
1	18 December	Paul Keller, Maarten Zeinstra, Julia Fallon	Kennisland/Europeana Foundation	Structure, selection of documents to annex
2	02 January	Paul Keller	Kennisland	Introduction and summary
3	03 January	Julia Fallon, Paul Keller	Europeana Foundation/Kennisland	Final review & Sign off

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This deliverable contains original unpublished work except where clearly indicated otherwise.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

1. Introduction & Context

This deliverable summarizes a 9 month effort to develop a strategy and an associated toolset to improve the quality of the rights statements contained in the metadata of the digital objects available via Europeana. The deliverable consists of a set of online tools, resources and procedures that will be employed by the Europeana Foundation in order to approach data providers whose collections lack rights statements for the objects made available via Europeana.

At the time we started work on this deliverable in March 2012 the 12.730.870 objects without rights statements made up 54% of all digital objects available via Europeana. As Chart 1 below shows, this number has since then declined to 34% (8.048.943 objects). This substantial decline can be attributed to a large re-harvesting effort that coincided with the transition to the new Europeana Data Exchange agreement. As part of this re-harvesting effort Europeana Awareness WP5 has worked with the ingestion team at Europeana to ensure that all re-harvested metadata contains the required rights statement. The remaining unlabelled metadata records in Europeana will need to be addresses separately and this deliverable creates the basis for doing so.
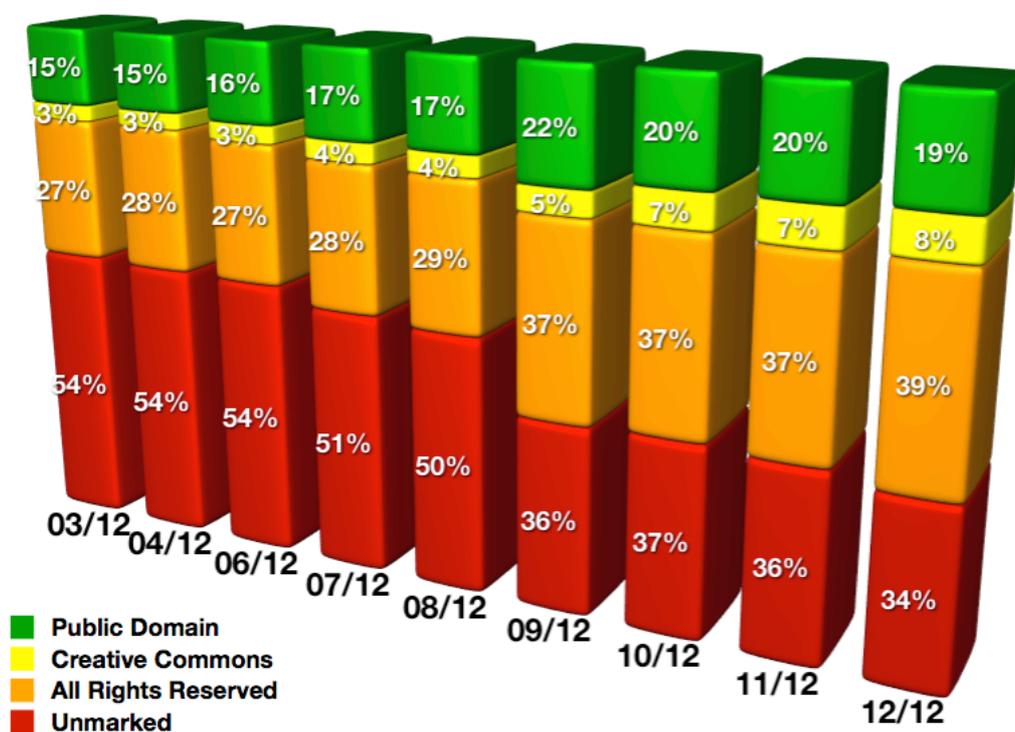

Chart 1: percentage of digital objects in Europeana per category of rights statement

Initial research into the existing rights statements in Europeana has shown that the majority of the objects available through Europeana were not labeled with rights statements and that a substantial amount of the rights statments provided were quite possibly wrong. Given this situation the partners in WP5 together with the management of the Europeana Foundation have agreed a strategy for improving the quality of rights statements associated with the digital objects available via Europeana. This strategy, the rights labelling campaign, puts less emphasis on advocacy to provide rights statements and instead focuses on procedures for

Europeana to directly reach out to all data providers with collections that lack rights statements.

This deliverable consists of a summary and 5 annexes that documents the different phases of the research as well as tools that will be employed for the rights labelling campaign that will start in January of 2013. Annex 1 represents the underlying research carried out by Kennisland. Annexes 2-4 document the elements of the toolset that will be used by the Europeana Foundation.

While this deliverable focuses on collections in Europeana that are lacking rights statements, WP5 has also developed a online resource for data providers that contribute new collections to Europeana. This web resource provides information about selecting an appropriate rights statement[1], an overview of all available rights statements[2] and an interactive tool that helps data providers to select a rights statement[3]. This online resource has been documented as Milestone 26 and will be continuously updated throughout the duration of the Europeana Awareness project.

2. Summary

This deliverable consists of two main parts.

- Research into the quality of the existing rights statements associated with collections that are available via Europeana. This research was carried out by Kennisland in Q2 of 2013. It categorises all collections within Europeana according to the status of their rights statement. If the rights statement is found to be either lacking or to be of questionable quality it additionally categorises the collections according to the proposed remedial actions. Annex 1 contains the analysis of the underlying data (the full dataset can be found at goo.gl/1H5DH).

- The Europeana strategy for approaching data providers and working with them on improving the quantity and quality of the rights statements related to existing collections. This strategy has been worked out together with the management of the Europeana Foundation. It consists of a strategy document (Annex 2), a rights statement module for Europeana's CRM system (documented in Annex 3) and a set of letter templates that will be used to approach data providers (Annex 4).

3. Implementation

The rights labelling campaign will start in January of 2013 and will be coordinated by Europeana Foundation with additional support from Kennisland. The campaign will initially focus on approaching the data providers with collections that either completely lack rights statements (action code 1 in Annex 3) or that are suspected of incorrectly identifying Public Domain content as copyrighted (action code 2 in Annex 3).

Based on the initial feedback to this campaign from the data providers we will gradually widen the campaign to also include collections that contain partially incorrect rights statements (action codes 3 and 4 in Annex 3). We expect to complete the initial phase of the rights labelling campaign (ensuring that all collections carry valid rights statements) in April 2013 and will conclude the rights labelling campaign (fixing obviously-wrong rights statements) before the end of 2013.

---

[1] http://pro.europeana.eu/web/guest/licensing
[2] http://pro.europeana.eu/web/guest/available-rights-statements
[3] http://pro.europeana.eu/web/guest/edm-rights-selection-tool

The progress of the rights labelling campaign will be communicated publicly through regular updates on the Europeana Professional blog. In addition the progress will also be reported to the Commission via the scheduled periodic reports. The rights labelling campaign is expected to positively impact all 4 Europeana Awareness KPIs that are related to the activities of WP5 (KPIs 1.4, 5.1, 5.2 and 5.3)

Project Acronym: EAwareness
Grant Agreement number: 297261
Project Title: Europeana Awareness

D5.1: Annex 1 Analysis on the quality and availability of EDM:rights

Revision	2
Date of submission	03 January 2013
Author(s)	Maarten Zeinstra, Kennisland
Dissemination Level	Public

REVISION HISTORY AND STATEMENT OF ORIGINALITY

Revision History

Revision No.	Date	Author	Organisation	Description
1	2012-07-11	Maarten Zeinstra	Kennisland	First version
2	2012-07-12	Paul Keller	Kennisland	Final Version

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 content

1. Introduction

With the launch of the Europeana Data Model (EDM) a intellectual property rights description field is required (EDM:rights). With this value end-users of Europeana can be informed what kind of permissions they have in re-using and distributing digital reproductions of cultural objects available through Europeana.

The predecessor of EDM:rights, ESE:rights was not mandatory. The transition between the non-mandatory ESE:rights statement to the mandatory EDM:rights statement resulted in legacy metadata that does not comply with the requirements of the EDM and does not comply with the Europeana Data Exchange Agreement (DEA).

The DEA states that all data providers will make their best efforts to provide Europeana with correct metadata on the intellectual property rights of the objects made available through Europeana. Furthermore, public domain material needs to be marked as such.

Both these requirements and the issue of legacy metadata have resulted in this deliverable with the main research question: what is the quality of the current intellectual property rights statements on Europeana and how can it be improved?

Kennisland researched the quality of the rights status of the whole collection on Europeana as part of work package 5 of Europeana Awareness starting March 2012. At the start of March Europeana had 23.407.051 items in its database. This document, describes the results and methodology of this research. The complete dataset on which these results are based can be found here goo.gl/1H5DH. The number of objects and collections mentioned reflects the status of these collections in March 2012 (and is thus slightly outdated).

2. Summary

Using an computer analysis based on the availability of metadata in the EDM:rights field and the EDM:year field we were able to create an index with statements of collection that are described correctly, incorrectly, or as having insufficient information to determine the correctness of EDM:rights. Maarten Zeinstra (Kennisland) developed the software for this part of the analysis. He also transformed the results in the format that is attached.

A human analysis went through 647 collections with more than 50 objects on Europeana and categorized collections based on the information they could find on the objects in these collections. They took a sample of objects to determine the rights status of individual works, regardless of the metadata in Europeana. This led to a numerical overview of the applicable rights in a collection. Lorine van Loon (Kennisland) and Claire Perrin (Kennisland) were responsible for sampling these large sets of metadata.

The results of these two analyses were combined into 14 categories. Only 15,47% of the collection does not need a correction. 25.40% need partial corrections or more metadata to confirm an accurate rights status. 52.75% does not have an EDM:rights value and a remaining 6.38% statements need to be changed to a public domain mark. Paul Keller (Kennisland) and Maarten Zeinstra were responsible for integrating the results of the computer research and the human research.

3. Methodology

The research consisted of two parts: a human analysis and a machine analysis. The human analysis determined the copyright status of collections on Europeana; the machine analysis gives an overview of the recorded copyright status on Europeana and an indication of the number of out of copyright works. A combination of these two overviews gives a list of 14 categories that indicate the status of a collection.

3.1 Human Analysis

Claire Perrin and Lorine van Loon did the human analysis. They sampled each collection on Europeana that contained more than 50 objects[1], going through many thousands of images and metadata records. Using these samples they determined the overall status of the quality of the rights status in Europeana using whatever metadata they could find, either on Europeana or on website of the data provider.

Eight categories were applied on all collections with more than 50 objects.

- MARKED INSUFFICIENT INFORMATION
- NO RIGHTS SHOULD BE PD
- NO RIGHTS SHOULD BE C/CC
- C/CC SHOULD BE PD
- PD SHOULD BE C/CC
- CORRECT
- NO RIGHTS INSUFFICIENT INFORMATION
- MIXED

Collections were either categorized fully into one category or they were categorized as MIXED. If they were categorized as MIXED an additional field was used to describe what mixture of categories is present in the collection.

The researchers used a rule of thumb for determining the copyright status of a work. A work is in the public domain when it was created before 1870. The European term of copyright protection of life + 70 years is covered assuming that all makers that created in 1870 died before 1940 (70 years later and 142 years ago). A lot of work with EDM:year before 1941 is also in the public domain, because the author has died over 70 years ago. We used this rule of thumb, because we do not have bibliographic information on all these creators to determine the exact year in which these works go out of copyright.

3.2 Machine Analysis

The Machine Analysis is the result of software created by Maarten Zeinstra that queried the Europeana API in March 2012. It queried, for each collection, how many objects were available given a certain rights statement. Creative Commons licenses were lumped together combining all versions and jurisdictions of a specific license. The difference between the sum of all objects that are marked with EDM:rights and the total objects in a collection yielded the number of unmarked items.

---

[1] 115 collections have 50 or fewer objects or 2.494 total objects. These collections are classified based on the machine analysis only.

Afterwards all objects with EDM:year before 1870 and EDM:right statements other than the Public Domain Mark were queried per collection[2]. This helped to give and indication that collections possibly have public domain material marked as in copyright material.

Reflection

This research was a combination of sampling a great number of objects per collection and automated research. There can be flaws on both parts of the research. The human researchers could have misjudged a collection and incorrectly could have placed it in a category that does not fit. Our researchers are not jurists and used a rule of thumb to determine the copyright status of a work. The machine analysis could be misled by bad metadata or a faulty feedback from the API.

4. Results

In March 2012 Europeana had metadata for 23.407.051 cultural objects. Our research indicates that 3.621.953 (15.47%) objects had correct data in their intellectual property field, EDM:rights. The remaining 19.785.098 (84,53%) objects need reviewing.

4.1 Categories and Actions

Four main categories and in total 14 categories are used to describe the quality of the copyright status of collections in Europeana.

Items per remedial action	Number of Objects	Percentage
No action required	3.621.953	15,47%
Request partial correction	5.945.971	25,40%
Request addition of EDM:rights value	12.346.101	52,75%
Request change of EDM:rights value to PD	1.493.026	6,38%
	23.407.051	100,00%

Copyright, Creative Commons, Mixed and Unmarked are the main categories that state whether or which rights status was found through the machine analysis. Copyright refers to all rights reserved statements of Europeana. Creative Commons to all Creative Commons licenses, Mixed is used on collections that mix copyright statements and licenses. Unmarked refers to those collections that have no EDM: rights statements.

Each of these statements is augmented with remarks that determine whether those collections need attention. These categories are a combination between the machine research and human research based on the dataset acquired in March 2012.

---

[2] Some false negatives could arise in this process. Instead of being able to query all EDM:year earlier than 1870 all EDM:years starting with 21, 19, 199, 198 and 197 were excluded. False negatives could arise here from objects with EDM:year in the first two centuries of the Common Era.

Per Collections (Combined results)	Number of Objects	Number of collections
COPYRIGHT	1.788.861	7,64%
COPYRIGHT (?)	588.156	2,51%
COPYRIGHT (all wrong)	1.493.026	6,38%
COPYRIGHT (min 15% wrong)	1.186.816	5,07%
CREATIVE COMMONS	393.073	1,68%
CREATIVE COMMONS (min 15% wrong)	100.938	0,43%
MIXED (?)	2.671.553	11,41%
MIXED	944.564	4,04%
MIXED (more than 15% PD)	178.326	0,76%
PUBLIC DOMAIN	495.455	2,12%
PUBLIC DOMAIN (min 15% C)	1.220.182	5,21%
UNMARKED (?)	4.767.368	20,37%
UNMARKED (100% PD)	107.920	0,46%
UNMARKED (C)	1.913.029	8,17%
UNMARKED (more than 15% PD)	5.557.784	23,74%
	23.407.051	100,00%

The sections below describe per category what we found and what actions are required.

4.1.1. COPYRIGHT

Both the machine research as well as the human research found that these collections had correct metadata concerning intellectual property. No action is required.

4.1.2. COPYRIGHT (?)

Although these collections are marked as being in copyright and rights are reserved, they have insufficient information to verify the validity of this statement. EDM:year was not available in the metadata to determine the age of the objects in these collections and human research did not find enough information to determine the rights status of these objects.

We recommend contacting the institutions that contributed these collections and request more metadata and previews to communicate these objects better.

4.1.3. COPYRIGHT (all wrong)

Using EDM:year and a rights reserved rights status we found that the objects in these collection pre-date 1860. This means that these works are most likely in the public domain. The institutions hosting the digital reproductions of these objects cannot legitimately claim copyright on these object. Our Human Analysis confirmed this.

We recommend that Europeana contacts these institutions and request them to change the EDM:rights to the Creative Common Public Domain Mark. Marking these objects as public domain is mandatory according to the DEA.

4.1.4. COPYRIGHT (min. 15% wrong)

Using EDM:year and the human analysis we found that these collections mix copyrighted material and public domain material but only apply rights reserved statements to their collections.

As some works are not copyright protected and need to be marked as public domain we recommend contacting these institutions to review their collection and mark all their public domain works as such.

At least 15% of objects in these collections are found to be in the public domain without being marked as such. However, this percentage is usually above 50%. We recommend that Europeana contacts these institutions and request them to review their collection and change the EDM:rights statements of those works that are in the public domain.

4.1.5. CREATIVE COMMONS

Our machine analysis found that these collections use CC licenses. Our human analysis did not find anything wrong with the rights statements of these collections. No action is required.

4.1.6. CREATIVE COMMONS (min. 15% wrong)

We treat Creative Commons works as regular in copyright works because we cannot verify the validity of a CC-license. We can only determine whether Creative Commons licenses are applied to works that are in the public domain.

Collections in this category have at least 15% of their collection mislabelled with a Creative Commons license. At least 15% of these objects are in the public domain. We recommend asking the institution to which these collection belongs to request a partial correction of their EDM:rights statements.

4.1.7. MIXED (?)

These collections have a mix of Creative Commons licensed material, public domain objects, and rights reserved works. Our machine analysis and human analysis showed insufficient information to determine a good indication of the copyright status of these works. This is usually caused by a lack of good metadata.

We recommend contacting the institutions that contributed these collections and request more metadata and previews to successfully communicate the rights status of these objects.

4.1.8. MIXED

These collections have a mix of Creative Commons licensed material, public domain objects and rights reserved works. Our machine analysis and human analysis showed no problematic rights statements. No action is required.

4.1.9. MIXED (More than 15% PD)

These collections have a mix of Creative Commons licensed material, public domain objects and rights reserved works. Our machine analysis and human analysis showed that 15% of these objects should be marked as public domain, but are marked as in copyright.

We recommend contacting the institutions that contributed these collections and request they review their collection and mark all public domain works with the Public Domain Mark.

4.1.10. PUBLIC DOMAIN

These collections are marked with the Creative Commons Public Domain Marked and were checked by the human analysis whether these objects are in the public domain. These collections seem to be correctly labelled. No action is required.

4.1.11. PUBLIC DOMAIN (min. 15% C)

Collections in this category are marked as public domain, but we have found a substantial number of works (more than 15% of the size of the individual collection) to be still in copyright. This means that either they are 2D representations of 3D works, created after 1941, or otherwise still copyright protected.

At least 15% of the objects in these collections are most likely to be still copyright protected, and a partial correction should be requested.

4.1.12. UNMARKED (?)

These collections have no EDM:rights value. Our human analysis and machine analysis was unable to determine whether these collections are public domain or still copyright protected.

Request the addition of the EDM:rights value and if possible additional metadata such as a year of publication.

4.1.13. UNMARKED (100% PD)

These collections have no EDM:rights value. Our human analysis and machine analysis, based on EDM:year, have found that these collections are most probably all public domain.

Request addition of EDM:rights value and indicate that according to our research the EDM:rights value should be the public domain mark.

4.1.14. UNMARKED (C)

These collections have no EDM:rights value. Our human analysis and machine analysis, based on EDM:year, have found that these collections are most probably all still protected by copyright.

Request addition of EDM:rights value. Indicate that according to our research this should be a rights value that indicates that the work is still copyright protected (all available values with the exception of PDM).

4.1.15. UNMARKED (more than 15% PD)

These collections have no EDM:rights value. Our human analysis and machine analysis, based on EDM:year, have found that these collections are a mix of public domain objects (at least 15% of the objects in these collections) and copyright protected works.

Request addition of EDM:rights value. Remark that public domain works need to be marked as such.

5. Datasets

5.1. Edmrights-analysis

The spreadsheet containing the full research data (available online at goo.gl/1H5DH) has five sheets. Each sheet represents a step in this research. Common in 'Analysis Machine', 'Analysis Human' and 'Combined Analysis' is the column called 'Collection name'. This is the unique key that connects these different sheets.

5.1.1. Analysis Machine

This is the result of a script that queried the Europeana API in March 2012. It queried, for each collection, how many objects were available given a certain rights statement as well as the total number of objects in that collection. Creative Commons licenses were lumped together combined all versions and jurisdictions of a specific license. The difference between the sum of all objects marked with an EDM:rights and the total objects in a collection yields the number of unmarked items.

Afterwards all objects with Europeana year before 1870 were queried per collection[3]. This helped to give and indication that collections possibly have public domain material marked as in copyright material.

Category
COPYRIGHT
COPYRIGHT (all wrong)
COPYRIGHT (no year)
COPYRIGHT (min 15% wrong)
PUBLIC DOMAIN
CREATIVE COMMONS
CREATIVE COMMONS min 15% wrong
MIXED
UNMARKED
UNMARKED (more then 15% PD)
UNMARKED (100% PD)
UNMARKED (no year)

5.1.2. Analysis Human

The human analysis sheet is an overview of the results of the analysis of the human research. Two researchers went through all collections with over 50 objects and determined the overall rights status of the work. Only works that were clearly public domain (i.e. created before 1870) were marked as needing to be public domain.

---

[3] Some false negatives could arise in this process. Instead of being able to query all EDM:year smaller than 1870 all EDM:years starting with 21, 19, 199, 198 and 197 were queried. False negatives could arise here from objects with EDM:year in the first two centuries of the Common Era.

They used the following categories to marked collections with the following categories:

Per Collections (Human analysis)

C/CC SHOULD BE PD
CORRECT
MARKED BUT INSUFFICIENT INFO
MIXED
NO RIGHTS SHOULD BE C/CC
NO RIGHTS SHOULD BE PD
PD SHOULD BE C/CC
UNMARKED/INSUFFICIENT INFORMATION
UNABLE TO ANALYSE

This sheet also includes a short description of the collections that have been analyzed. While these descriptions have not been used for our analysis we include them here, as they might be valuable for other purposes

5.1.3. Categories

The sheet Categories gives an overview of all categories with a per category description. For the human research shorthand was used that is not replaced by the actual categories in the column 'combined results' in the Analysis Human sheet.

5.1.4. Combined Analysis

The Combined Analysis sheet places the categories of the two research sheets next to each other. It then maps each category combination to one of 15 final categories:

Per Collections (Combined results)

COPYRIGHT
COPYRIGHT (?)
COPYRIGHT (all wrong)
COPYRIGHT (min 15% wrong)
CREATIVE COMMONS
CREATIVE COMMONS (min 15% wrong)
MIXED (?)
MIXED
PUBLIC DOMAIN
PUBLIC DOMAIN (min 15% C/CC)
UNMARKED (?)
UNMARKED (100% PD)
UNMARKED (C/CC)
UNMARKED (more then 15% PD)

5.1.5 Summary

The Summary sheet gives an overview of all other Sheets. It contains the numbers that this report is based on. It also serves as a page to check whether all collections are used in the research.

5.2. Actionlist edmrights

A second excel sheet is attached (actionlist edmrights) that organized collections by four main categories: No action required, Partial correction required, Corrections required and

Request addition of EDM:rights. Each of these has a column collection name, number of objects, action, and category. The category column has one of the 15 described categories above.

We recommend Europeana uses this file and goes through each sheet, except 'no action required', contacts all institution that have contributed a collection to Europeana to update their metadata.

6. Conclusions and General Recommendations

84.53% of the collections need to be updated. Europeana has stricter rules for the status of intellectual property marking, but Europeana failed to police their collection according to these new rules. The provided information and analysis brings together several ways of researching the collections.

Europeana will become better searchable and more valuable for end users and re-users when data providers update their metadata to the stricter standard of metadata and copyright descriptions.

We recommend that Europeana use the second attachment and this document to contact each of the institutions that provided the sets with metadata that needs to be updated. Some changes categories are recommendations and request for partial changes, other collections need be to change in order to remain in Europeana.

A sample analysis of all collections on Europeana is very time consuming. Machine analysis is much faster, but due to a lack of metadata and searchability of the Europeana API this has not yet reached a confidence level on human research.

Based on this experience we recommend making more metadata available and mandatory. Among the different metadata fields the EDM:year field stands out. Making EDM:year mandatory would greatly help determining whether or not objects are in the public domain. Bibliographic information about creators of cultural objects would also greatly increase the quality and confidence of an intellectual property analysis.

D5.1: Annex 2 Strategy on EDM:rights improvement

REVISION HISTORY AND STATEMENT OF ORIGINALITY

Revision History

Revision No.	Date	Author	Organisation	Description
1	2012-11-28	Yannick H'Madoun	Kennisland	First version
2	2012-12-02	Maarten Zeinstra	Kennisland	Final Version
4	2012-12-20	Julia Fallon	Europeana Foundation	Review
3	2012-12-18	Paul Keller	Kennisland	Sign-off

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.

1. Introduction

Europeana have ingested more than 700 different collections from various data providers over the past three years. During this time the requirements for submitting these collections for ingestion have changed. Notably the rights statements that were once required for each object have been updated and are now mandatory.

In the report "Analysis on the quality and availability of EDM:rights" (Annex 1) we have assessed the completeness and correctness of these rights statements in Europeana. The initial research provides a detailed explanation of the method and approach for this assessment. In short, we've applied a computer analysis and a human verification process to a representative sample of all datasets and combined these results to divide the collections into categories based on the completeness and correctness of their rights statements. The results of this research show that the rights statements for around 85% of the collections require some form of correction. This document describes the workflow and structure of the follow-up procedure for correcting the collections indicated in the assessment.

This documents defines a set of actions per collection and describes how these will be added to the Europeana CRM system (SugarCRM) to assist managing and reviewing the progress of the rights labelling campaign. Actions are based on 5 major categories, and in total 9 specific types of letters will be drafted. We will start by focusing on those collections that are determined to be completely missing rights statements and those that are believed to be 100% incorrect.

2. SugarCRM EDM:Rights improvement infrastructure

The following fields will be added to SugarCRM to provide status indicators for collections. These fields will also be used to follow the progress of the project.

Field Name	Option/description
Date Verified	free text (yyyymmdd)
Current Code	free text (1.1, 1.2, 1.3, .. 5)
Letter sent	tick box
Answer received	tick box
EDM right verified	tick box
Log	Note field with 255 character restriction

The SUGAR log field contains an overview of which statuses the collection has and an overview of all previous actions. The format for this field will be: <date in YYYYMMDD><action code> - <'letter sent' OR Empty>. Using this method allows us to

analyses the collections based on the action required, the verification status and determine if and when the designated action has been communicated to the data provider.

2.1 Date Verified

Date Verified indicates the date (YYYYMMDD) of the last EDM:rights check of this collection.

2.2 Current Code

The Analysis on the quality and availability of EDM:rights (see Annex 1) categorized over 700 collections in 14 categories. These categories have been grouped into 4 main categories with a number of corresponding actions per category. A description of each analysis code is provided beneath the table below.

Analysis Category	Action code
Copyright	5
Copyright (?)	4.1
Copyright (all incorrect)	2
Copyright (min 15% incorrect)	3.1
Creative Commons	5
Creative Commons (min 15% incorrect)	3.2
Mixed	5
Mixed (?)	4.2
Public Domain	5
Public Domain (min 15% incorrect)	3.3
Unmarked (?)	1.1
Unmarked (100% PD)	1.2
Unmarked (C/CC)	1.3
Unmarked (min 15% incorrect)	1.4

Action Code #1

Please add the appropriate rights statements and:

1.1 Unmarked (?): We don't have enough information to know which rights statement applies
1.2 Unmarked (100% PD): We think it's all Public Domain
1.3 Unmarked (C/CC): We think it's Copyright or Creative Commons

1.4 Unmarked (min 15% incorrect): We think there's at least some amount of Public Domain content

Action Code #2

Copyright (All incorrect): Your institution has marked this content as copyrighted, but we believe your collection is Public Domain content. Please update your collection accordingly.

Action Code #3

We've sampled your collection and concluded that at least some content was marked with incorrect rights statement:

3.1 Copyright (min 15% incorrect): You have at least some content marked as copyrighted that we believe should be Public Domain
3.2 Creative Commons (min 15% incorrect): You have at least some content marked as Creative Commons that we believe should be Copyright / PD.
3.3 Public Domain (min 15% incorrect): You have at least some content marked as PD that we believe should be C/ CC

Action Code #4.

Copyright (?)/Mixed (?): We believe there are some discrepancies in your rights statements, but we don't have enough information to know what the EDM rights statement should be.

Action Code #5

Copyright/Creative Commons/Mixed/Public Domain: The collection appears to be in order. No further action required.

2.3 Letter sent

This box is ticked when a letter connected to an action code is sent.

2.4 Answer received

This box is ticked when an answer is received from an institution.

2.5 EDM rights verified

This box is ticked when the collection is verified and deemed in order in accordance with the standards for rights statements set by Europeana.

2.6 Example

Field Name	Value
Date Verified	20121128
Current code	1.1
Letter sent	☒
Answer received	☐
EDM rights verified	☐
Log	20121128 1.1 - letter sent

3.1 Proposed implementation procedure

In order to systematically approach data providers with collections that lack rights statements the following steps need to be undertaken:

- Add the rights statements module to SUGAR CRM (Europeana foundation, December 2012)
- Draft template letters for approaching data providers covering all action codes outlined above (Kennisland & Europeana Foundation, December 2012)
- Assign action codes to collections in SUGAR CRM (Kennisland & Europeana Foundation, January 2013)
- Blog post announcing the rights labelling campaign (Kennisland & Europeana Foundation, January 2013)
- Start approaching data providers (Europeana Foundation, January 2013)

The campaign should initially focus on approaching data providers with collections that either completely lack rights statements (Action Code #1) or that are incorrectly identify Public Domain content as being copyrighted (Action Code #2).

Based on the initial feedback to this campaign from the data providers the campaign can be gradually widened to also include collections that contain partially incorrect rights statements (Action Codes #3 and #4).

Project Acronym: EAwareness
Grant Agreement number: 297261
Project Title: Europeana Awareness

D5.1: Annex 3 Rights module for SUGAR CRM

Revision	1
Date of submission	1 January 2013
Author(s)	Paul Keller (Kennisland)
Dissemination Level	Public

REVISION HISTORY AND STATEMENT OF ORIGINALITY

Revision History

Revision No.	Date	Author	Organisation	Description
1	2012-12-19	Paul Keller	Kennisland	Final Version

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.

1. Introduction

As part of the online toolkit for the rights labelling campaign (Europeana Awareness Deliverable D5.1) Europeana Foundation has added a rights statements module to their internal Customer Relationship Management System (SUGAR CRM). This module allows all individuals with access to the CRM system to track the progress of the rights labelling campaign on a per collection basis[1]. The ingestion team that is responsible for updating it will primarily use the rights statements module.

The screenshot below shows the fields contained in the rights statements module of sugar CRM. The functionality of the fields is specified and explained in the 'Strategy on EDM:rights improvement' document (Annex 2).

---

[1] Collections are the smallest possible aggregations in Europeana's database, and the quality of rights statements are being tracked on a per collection basis. One data provider can have multiple collections in Europeana.

Project Acronym: EAwareness
Grant Agreement number: 297261
Project Title: Europeana Awareness

D5.1: Letter templates for rights labelling campaign

Revision	4
Date of submission	3 January 2013
Author(s)	Paul Keller (Kennisland), Julia Fallon (Europeana Foundation)
Dissemination Level	Public

REVISION HISTORY AND STATEMENT OF ORIGINALITY

Revision History

Revision No.	Date	Author	Organisation	Description
1	2012-12-10	Julia Fallon	Europeana Foundation	First version
2	2012-12-10	Paul Keller	Kennisland	Review
4	2013-01-02	Julia Fallon	Europeana Foundation	Second version
3	2013-01-02	Paul Keller	Kennisland	Final version, Sign-off

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.

1. Introduction

This document contains template language to compose letters that can be sent to data providers as part of the rights labelling campaign. All letters consist of three parts: an introduction, an outline of observations and recommendations (based on the Action Codes identified in Annex 3) and a final section containing contact information. The following section of this document contains all of the elements needed to compose letters corresponding to all Action Codes identified in Annex 3. The final section of this document contains three sample letters. The letters will be via email by the Europeana Foundation.

2. Template elements

2.1 Introduction

Dear xxx,

Europeana have been undertaking a review of the metadata sets that are currently available through the portal. This review helps us ensure that published metadata meets the standards we set. We have paid particular attention to rights statements that are provided, as our goal is to deliver a robust service populated with good quality metadata.

2.2 Observations and recommendations

2.2.1 Action Code #1

From our review, we have found identified some collections that were not submitted with rights statements. Since May 2012, such rights statements are mandatory for every object submitted, so this now needs to be rectified.

    1.1   Collection ID: xxxxxxx

    We have undertaken a preliminary review of the above collection(s), and are unable to identify which rights statements are the most likely to apply.

    1.2   Collection ID: xxxxxxx

    From a preliminary review we believe, based on the information available to us, that the objects in the above collection(s) should be labeled as 'Public Domain'

    1.3   Collection ID: xxxxxxx

    From a preliminary review we believe, based on the information available to us, that the objects in the above collection(s) should be labeled with either a 'rights reserved' statement or a Creative Commons license.

1.4   Collection ID: xxxxxxx

From a preliminary review we believe some objects in the above collection(s) should be labeled as 'Public Domain'

We request that you review these collections and provide us with rights statements for these collections.

2.2.2 Action Code # 2

From our review, we have found that there are some collections that you have provided to us with a rights statement that indicates the works are in copyright. However, based on the information available to us, we believe that the correct rights statement for the following collections is 'Public Domain'.

Collection ID: xxxxxxx

We request that you review these collections and confirm or update the rights statements for these collections.

2.2.3 Action Code #3

From our review, we have found that there are some collections that you have provided which we, based on the information available to us, think might not be labeled with the correct rights statement.

3.1   From a preliminary review, we believe that the following collections should be labeled as 'Public Domain':

Collection ID: xxxxxxx

3.2   From a preliminary review, we believe that the following collections contain some objects that are falsely labeled with a Creative Commons mark and should instead either be labeled with either a 'rights reserved' statement or with the Public Domain Mark.

Collection ID: xxxxxxxxx

3.3   From a preliminary review we believe that the following collections contain some objects that are falsely labeled with a Public Domain Mark and should instead either be labeled with either 'rights reserved' statement or with a Creative Commons licenses

Collection ID: xxxxxxxxx

We request that you review these collections and confirm or update the rights statements for these collections.

2.2.4 Action Code #4

From this review, we have found that there are some collections that you have provided to us which we, based on the information available to us, think might not be labeled with the correct rights statement. We have undertaken a preliminary review of the following collections, and are unable to identify which rights statements are the most likely to apply.

Collection ID: xxxxxxx

We request that you review these collections and confirm or update the rights statements for these collections.

2.3 Final Section

You will find a list of the rights statements online at Europeana Pro here:

http://pro.europeana.eu/web/guest/available-rights-statements

For all other collections, we do not believe any further review of the rights statement is required. However, we are constantly working on increasing the quality of metadata in the Europeana Portal in order to guarantee a meaningful user experience. So we do recommend that the metadata submitted to Europeana is regularly reviewed to ensure it is kept up to date.

We would like to invite you to take this opportunity to improve the richness and the quality of your metadata. In particular we kindly ask you to review the presence, quality and consistency of the following fields:

- dc:description and/or dc:title
- dc:date (and/or dc:created and/or dc:issued)
- dc:coverage (and/or dc:spatial, and/or dc:temporal)
- Europeana:object/ edm:object

In addition, we ask you to check the consistency and persistency of the fields dc:identifier, Europeana:isShownBy/ edm:isShownBy and Europeana:IsShownAt/ edm:isShownAt, in order to prevent broken links directed to your portal.

To respond to this request to submit improved or additional metadata, please contact us at content@europeana.eu. We will then arrange re-harvesting your datasets.

If you have a question about this email, or require further guidance please contact the Europeana content team at content@europeana.eu.

3. Sample Letters

3.1 Sample Letter 1.1

Dear xxx,

Europeana have been undertaking a review of the metadata sets that are currently available through the portal. This review helps us ensure that published metadata meets the standards we set. We have paid particular attention to rights statements that are provided, as our goal is to deliver a robust service populated with good quality metadata.

From our review, we have found identified some collections that were not submitted with rights statements. Since May 2012, such rights statements are mandatory for every object submitted, so this now needs to be rectified.

Collection ID: xxxxxxx

We have undertaken a preliminary review of the above collection(s), and based on the information available to us we are unable to identify which rights statements are the most likely to apply.

We request that you review these collections and provide us with rights statements for these collections. You will find a list of the rights statements online at Europeana Pro here:

http://pro.europeana.eu/web/guest/available-rights-statements

For all other collections, we do not believe any further review of the rights statement is required. However, we are constantly working on increasing the quality of metadata in the Europeana Portal in order to guarantee a meaningful user experience. So we do recommend that the metadata submitted to Europeana is regularly reviewed to ensure it is kept up to date.

We would like to invite you to take this opportunity to improve the richness and the quality of your metadata. In particular we kindly ask you to review the presence, quality and consistency of the following fields:

- dc:description and/or dc:title
- dc:date (and/or dc:created and/or dc:issued)
- dc:coverage (and/or dc:spatial, and/or dc:temporal)
- Europeana:object/ edm:object

In addition, we ask you to check the consistency and persistency of the fields dc:identifier, Europeana:isShownBy/ edm:isShownBy and Europeana:IsShownAt/ edm:isShownAt, in order to prevent broken links directed to your portal.

To respond to this request to submit improved or additional metadata, please contact us at content@europeana.eu. We will then arrange re-harvesting your datasets.

If you have a question about this email, or require further guidance please contact the Europeana content team at content@europeana.eu.

3.2 Sample Letter 2

Dear xxx,

Europeana have been undertaking a review of the metadata sets that are currently available through the portal. This review helps us ensure that published metadata meets the standards we set. We have paid particular attention to rights statements that are provided, as our goal is to deliver a robust service populated with good quality metadata.

From this review, we have found that there are some collections that you have provided to us with a rights statement that indicates the work is in copyright. We, based on the information available to us, believe that the correct rights statement for the following collections is 'Public Domain'.

> Collection ID: xxxxxxx

We request that you review these collections and confirm or update the rights statements for these collections. You will find a list of the rights statements online at Europeana Pro here:

http://pro.europeana.eu/web/guest/available-rights-statements

For all other collections, we do not believe any further review of the rights statement is required. However, we are constantly working on increasing the quality of metadata in the Europeana Portal in order to guarantee a meaningful user experience. So we do recommend that the metadata submitted to Europeana is regularly reviewed to ensure it is kept up to date.

We would like to invite you to take this opportunity to improve the richness and the quality of your metadata. In particular we kindly ask you to review the presence, quality and consistency of the following fields:

- dc:description and/or dc:title
- dc:date (and/or dc:created and/or dc:issued)
- dc:coverage (and/or dc:spatial, and/or dc:temporal)
- Europeana:object/ edm:object

In addition, we ask you to check the consistency and persistency of the fields dc:identifier, Europeana:isShownBy/ edm:isShownBy and Europeana:IsShownAt/ edm:isShownAt, in order to prevent broken links directed to your portal.

To respond to this request to submit improved or additional metadata, please contact us at content@europeana.eu. We will then arrange re-harvesting your datasets.

If you have a question about this email, or require further guidance please contact the Europeana content team at content@europeana.eu.

3.2 Sample Letter 3.2

Dear xxx,

Europeana have been undertaking a review of the metadata sets that are currently available through the portal. This review helps us ensure that published metadata meets the standards we set. We have paid particular attention to rights statements that are provided, as our goal is to deliver a robust service populated with good quality metadata.

From our review, we have found that there are some collections that you have provided which we, based on the information available to us, think might not be labeled with the correct rights statement.

Collection ID: xxxxxxxxx

We request that you review these collections and confirm or update the rights statements for these collections. You will find a list of the rights statements online at Europeana Pro here:

http://pro.europeana.eu/web/guest/available-rights-statements

For all other collections, we do not believe any further review of the rights statement is required. However, we are constantly working on increasing the quality of metadata in the Europeana Portal in order to guarantee a meaningful user experience. So we do recommend that the metadata submitted to Europeana is regularly reviewed to ensure it is kept up to date.

We would like to invite you to take this opportunity to improve the richness and the quality of your metadata. In particular we kindly ask you to review the presence, quality and consistency of the following fields:

- dc:description and/or dc:title
- dc:date (and/or dc:created and/or dc:issued)
- dc:coverage (and/or dc:spatial, and/or dc:temporal)
- Europeana:object/ edm:object

In addition, we ask you to check the consistency and persistency of the fields dc:identifier, Europeana:isShownBy/ edm:isShownBy and Europeana:IsShownAt/ edm:isShownAt, in order to prevent broken links directed to your portal.

To respond to this request to submit improved or additional metadata, please contact us at content@europeana.eu. We will then arrange re-harvesting your datasets.

If you have a question about this email, or require further guidance please contact the Europeana content team at content@europeana.eu