



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

D5.1: Content Quality specifications

Revision	[2.0)
Date of submission	29 Oktober 2012
Author(s)	Annelies van Nispen - The Europeana Foundation
Dissemination Level	[Public]

REVISION HISTORY AND STATEMENT OF ORIGINALITY

Revision History

Revision No.	Date	Author	Organisation	Description
1.0	28 June 2012	Annelies van Nispen	The Europeana Foundation	Quality Assurance specifications
2.0	21 October 2012	Annelies van Nispen	The Europeana Foundation	Revised Quality Assurance specifications

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This deliverable contains original unpublished work except where clearly indicated otherwise.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

Introduction

Chapter 1 Aligning the content strategy and content quality

Chapter 2 Metadata Quality

- 2.1 Quality Assurance of Europeana Semantic Elements (ESE)
- 2.2 Moving from Europeana Semantic Elements (ESE) to Europeana Data Model (EDM)
- 2.3 Mandatory Fields of EDM/ESE
- 2.4 Correct rights statements for EDM/Europeana rights element
- 2.5 Quality assurance of other mandatory elements

Chapter 3 Linking to the digital objects

- 3.1 Taskforce Persistent identifiers
- 3.2 Revising the technical infrastructure, identifiers and implementing link check functionality in the United Ingestion Manager (UIM)
- 3.3 Data Quality of EDM/ESE elements: IsShownBy, isShowAt and Object

Conclusion

Appendix

Introduction

This deliverable D5.1: Content Quality Specifications builds on earlier work on data quality improvement from the period 2008-2012, will give an overview of the work done and will describe the plans for further improvement set for EuropeanaV2.

Europeana currently (June 2012) holds 23,4 million digital objects from libraries, audio-visual institutions, museums and archives across Europe. The strategy of Europeana has been to focus on increasing the number of content to create a critical mass of digital cultural heritage for users. With the focus on quantity, quality of metadata was subject to the goal of critical mass. This was addressed in the evaluation of EuropeanaV1 and Europeana's focus is shifting towards quality.

Europeana has addressed data quality in the [Strategic Plan 2011-2015](#) in Strategic Track 1 Aggregate:

Improve the quality of the metadata

Europeana holds descriptive metadata rather than digitized objects. Improvements to this data offer better search results to users and increase the research opportunities that are enabled by Europeana. We are developing guidelines, tools and training to help content providers improve and standardize data quality and the validity of persistent identifiers. We are enforcing the clear attribution of rights information so that users know how they can legitimately use the content.

The web works on the basis of things being found; scholarship functions on traceable reference. These require that the cultural heritage sector ensures the persistency of its digital heritage in the same way that it has preserved and made accessible our physical and intellectual heritage. Agreement must be reached on the use of persistent identifiers so that links are not broken, users are not alienated and scholarship can proceed on the basis of secure citation.

The full potential of interoperability will be unlocked by the implementation in 2011 of the Europeana Data Model (EDM), a new way of structuring data. EDM will enable the use of Semantic Web Technology, support Linked Open Data, maintain more domain-specific rich information and allow digital objects from providers to be shown alongside authoritative and curated information from other domains. The change will benefit not only Europeana but also our providers who will be able to use the enriched data to upgrade to their own users.

Multilingual access to Europeana's content is known to be a priority for users. in association with experienced partners in the field, including Humboldt University and Google, we are investigating solutions for the forthcoming releases of Europeana.

Improving the quality of the metadata is not only a task of Europeana. The Aggregators and data provider network are the most important players in improving the quality of the metadata that they deliver to Europeana. Data quality improvement starts at their end. The institutions are the creators and controllers of the metadata that they deliver to Europeana.

D5.1: Content Quality Specifications

The construction of aggregators has created an extra layer where the data is processed and often converted to another data format from the original data format. Improving or guaranteeing the quality of the original metadata has to be addressed at the aggregator level and the data provider level. In all levels there is potential for duplication and loss of information.

Improvement of data quality needs to be seen as an interactive process between all levels of the Europeana network. Europeana has to facilitate and give support and guidance to all its suppliers. Europeana has done a lot of work on writing guidelines and defining best practices. This work will continue during the Ev2 and the current Strategic Plan.

Quality assurance of the metadata in the Europeana repository will focus in 2012- on improving:

- Aligning the content strategy, aggregation landscape and content quality
- Quality of the metadata
- Linking to the digital objects.

The appendix gives an overview of available support tools and guidelines for providers and aggregators.

Chapter 1 Aligning the content strategy and content quality

In 2012-2013 the Content Strategy of Europeana and the data quality assurance need to be further aligned.

The Content Strategy (in draft) addresses Data Quality in the following:

As stated in the Strategic Plan 2013-2015, Europeana aims to be the “trusted source” of cultural heritage information. In an increasingly crowded landscape of providers of digital cultural heritage (such as Google, Wikipedia, etc) Europeana and its partners offer users very specific and unique qualities such as authenticity and trust. As we move from a “supply driven” towards a more “demand driven” structure where accessibility is key, these qualities will become the differentiating factors of our value proposition to end-users. Indeed, Europeana’s central tenet is to create a landscape in which the user can get what he wants when and where he needs it. With 23.4 million objects in the repository we can state that we have assembled a critical mass of cultural artifacts. We will continue to increase the amount of objects in the database with the aim of including over 30 million objects by 2015. But that by itself is not enough. In parallel with the quantitative objectives it remains crucial to improve the quality of the database as well.

A collection strategy can not be elaborated without a definition of what is understood by “content”. Until recently, Europeana did not hold nor stock any of the actual digital items (the content), which belong to the partner institutions. Europeana assembles the metadata (information describing the object), the preview associated to an item and a link to the content (digital object) on the partner’s website. This has changed when Europeana started assembling User Generated Content such as the material harvested with the Europeana 1914-1918 project. While this is a significant change to the previous strategy our primary aim will remain to make the content of Europe’s memory institutions accessible for all. Europeana has little influence over the content that is made available through the partner institutions (either qualitatively or quantitatively), although recommendations can be made to steer into certain directions. Europeana does have a responsibility to improve the quality of the Europeana repository through the improvement of metadata, previews and links, and works with its partners to do so. What type of material will users be able to find thematically and in what format (image, audio, video, 3D)? How exhaustive will the material in particular subject areas need to be, and will all countries and languages be represented? Will these users be able to re-use the material, in particular if the original was in the public domain? These are the types of questions that we want to address with this strategy and plan.

And focuses on the following quality conditions

It is essential to set as an introductory comment that the ideal environment for the partners’ digital object starts with the quality of the Europeana portal. The metadata needs to be accurate, the hyperlinks functioning, and previews of objects available in the best possible quality.

D5.1: Content Quality Specifications

When browsing on the portal, one can see that the metadata is not always as accurate as one would wish, nor displaying the same kind of information. For example the date of the object, its description, or other fields of information may or may not be available in metadata. The links leading the user to the partner's website, where the digital object is held are also part of the essential elements. Europeana uses a tool that checks the links on a daily basis. It checks 3 objects per collection everyday. The resulting estimation of dysfunctional links in Europeana is therefore of 3%, although this number is a little over estimated as some links considered as dysfunctional by this simple link-checking script can actually be read by a browser.

In parallel to this, the preview (or thumbnail) for an object can be available or not. It is hard to say exactly how many previews are available or not because the ingestion process does not record the existence or successful generation of the preview. According to the total number of previews we have accumulated we can roughly estimate that we hold previews for 60% of the content.

D5.1 Content Quality Specifications will describe in the following chapters the Data Quality Assurance.

Chapter 2 Metadata Quality

Metadata quality is very important for Europeana and the users. The [Technical requirements](#) inform providers about the basic technical requirements to be fulfilled.

Until recently the metadata must conform to the Europeana Semantic Elements (ESE), a Dublin Core-based set of fields with additionally 12 specific Europeana elements, which are necessary for records to display correctly in Europeana. Aggregators have developed tools with which they can help providers to map their metadata to the Europeana Semantic Elements.

Europeana is currently moving towards the European Data Model (EDM) and this is currently implemented at Europeana. The first providers have started to deliver data in EDM.

2.1 Quality Assurance of Europeana Semantic Elements (ESE)

Europeana has made many changes to ESE to support the goals of Europeana. These are visible in the different versions of ESE. Version 2.1 was a major change. The latest version is 3.4.1.

The last changes in ESE have been mainly to implement new elements and the next step was to make elements such as Europeana:rights mandatory. These alterations were necessary to support the goals as defined in the [Europeana Licensing Framework](#).

In 2012 changes have been to include the element User Generated Content and 3D is added as a value for type. The Europeana portal/api is now able to distinguish both the 3D collections and User Generated Content.

Overview of last changes to ESE version 2.1 and version 3.4.1) in the [Metadata Mapping and Normalization Guidelines for the Europeana Semantic Elements](#).

Main changes last version (v2.1)

The document has been rearranged to reflect the introduction of more mandatory elements. europeana:rights	Changed to Mandatory. Clarification of use of this element and dc:rights
europeana:dataProvider	Changed to Mandatory.
dc:language	Changed to Mandatory for objects with type TEXT, strongly recommended for other types where appropriate.
dc:title and dc:description	Mandatory to provide one of these two
dc:subject, dc:type, dc:coverage; dcterms:spatial	Mandatory to provide one of these four
europeana:hasObject	Removed, now redundant.
europeana:country	Revision of note to reflect current practice.
europeana:language	Revision of note to reflect current practice.

Main changes in latest version 3.4.1

- Addition of the element europeana:UGC to indicate that content is user generated.
- Addition of the value "3D" for the europeana: type element

List of elements by group

Overview of all elements grouped into 4 categories:

- Mandatory for Providers
- Recommended for Providers
- Additional for Providers
- Europeana supplies by Europeana. These elements are added to the metadata of Providers by Europeana

Mandatory elements
dc:title or dc:description
dc:language for text objects
Europeana:dataProvider
Europeana:isShownAt or Europeana:isShownBy
Europeana:provider
dc:subject or dc:type or dc:coverage or dcterms:spatial
Europeana:rights
Europeana:type
Europeana:UGC (when applicable)

Recommended elements
dcterms:alternative
dc:creator
dc:contributor
dc:date
dcterms:created
dcterms:issued
dcterms:temporal
dc:publisher
dc:source
dcterms:isPartOf
Europeana:object

Additional elements
dc:format
dcterms:extent
dcterms:medium
dc:identifier
dc:rights
dcterms:provenance
dc:relation
dcterms:conformsTo
dcterms:hasFormat
dcterms:isFormatOf
dcterms:hasVersion
dcterms:hasPart
dcterms:isReferencedBy
dcterms:references
dcterms:isReplacedBy
dcterms:replaces
dcterms:isRequiredBy
dcterms:requires
dcterms:tableOfContents
Europeana:unstored

Elements supplied by Europeana
Europeana:country
Europeana:language
Europeana:uri
Europeana:usertag
Europeana:year

2.2 Moving from Europeana Semantic Elements (ESE) to Europeana Data Model (EDM)

As described ESE is a flat data model that provides the common denominator for ensuring interoperability between metadata standards. The disadvantage of this is that richness is lost in the process. Europeana Data Model (EDM) is a data model that will improve the way metadata can be provided and used in Europeana. It has the following benefits:

- it will allow a greater degree of granularity in describing objects: distinguishing between the original object and its digital representation
- the original object and the record describing that object
- objects that are composed of other objects
- it will support the aggregation of representations of the same object from different sources with different and possibly contradictory statement in the metadata

To exploit the richness of data from many cultural heritage domains, EDM adopts a cross-domain semantic web-based framework that will also allow enrichment of data from third party sources by linking.

EDM is currently implemented in Europeana and the first data providers are delivering their metadata in EDM. ESE is a subset of EDM and is therefore fully compliant with EDM to assure backward compatibility.

The first providers have started to deliver data in EDM. To support this task Europeana has created documentation, case studies and guidelines on Europeana Pro for providers. Europeana currently works intensively to support providers (projects and aggregators) in applying EDM to their metadata.

This gives valuable feedback on implementing EDM. Europeana will evaluate this and use this to create updated documentations, guidelines, FAQ's and workshops.

Overview of available EDM documentation

The [EDM object templates](#): this working document is a simple wiki listing that shows which properties apply to which class and states the data types and obligation (cardinality – how often they can occur within one EDM 'object') of the values. These templates should be regarded as a work in progress however and may be out of step with the Guidelines.

The [XML schema](#): this is the XML schema for the first implementation of EDM.

The **EDM Factsheet** and an **EDM presentation** sum up the rationale and expected benefits of EDM.

Available documentation on Europeana Pro

[Europeana Data Model Definition v5.2.3](#)

[Europeana Data Model Primer](#)

[Europeana Data Model Mapping Guidelines](#)

[Europeana Data Model Factsheet](#)

EDM case studies

This section features implementations of EDM in different domains. It aims to demonstrate how EDM can be used for richer data provision, how the main EDM specifications can be specialized to support the needs of specific domains and how EDM is suitable for the publication of Linked Data.

The following pages present cases studies on:

- [MIMO and EDM](#)
- [Europeana Libraries and EDM for Libraries](#)
- [Europeana Linked Open Data pilot](#)
- [Mapping EAD data to EDM](#)
- [The Polymath Virtual Library and EDM](#)

Europeana currently works intensively to support providers (projects and aggregators) in applying EDM to their metadata and case studies are added regularly.

2.3 Mandatory Fields of EDM/ESE

As described earlier Europeana changed the number of mandatory elements in the metadata of providers to enhance the metadata quality on Europeana. As this was done over the course of years, the quality level of the metadata of collections differs in Europeana.

With the current set of mandatory elements data quality is improved and guaranteed to this level. Some providers give very rich metadata and other provider's only give the mandatory elements. Therefore Europeana has and will have different levels of quality of metadata.

All mandatory elements should apply the current guidelines and best practices. This is part of the workflow for the current collections coming in to Europeana. What the effect of the DEA and EDM will be on the quality of metadata needs to be monitored and evaluated over the next year.

An effect of the DEA and EDM expected is that many data providers will resubmit their collections to Europeana; this will improve the quality of the metadata. For the remaining legacy data a Data Quality Improvement Plan needs to be set in place in 2012- 2013.

2.4 Correct rights statements for EDM/Europeana: rights element

The highest priority for Europeana is to improve the correct rights labeling in the mandatory element EDM: rights (in ESE europeana: rights). The rights information in this element describes the rights information about the digital objects (and the previews/thumbnails). Only one rights statement can be provided per resource. Rights statements are encoded as URLs referring to web pages that contain information about the applicable rights. For providers the difference between the DEA, which licenses the metadata standard as Creative-Commons-Zero Public Domain Dedication (CC0) and the edm: rights element is not always clear. The Europeana Ingestion Team quality checks the edm: rights statement, but is relying on the aggregators and data providers to provide Europeana correct rights statements about their content.

Europeana has published the [Europeana Licensing Framework](#) in December 2011. This Framework is set to govern the relationships of Europeana, its data providers and its users.

Europeana uses standard, interoperable and machine-readable licenses that can be understood by both humans and machines. There are 12 rights statements made available by Europeana for data providers made available by Europeana. These are the values used for the edm:rights element.

The Europeana Licensing Framework gives an overview of the available rights statements and these are available on the [Legal Requirements page](#) on Europeana Pro:

Types of Rights Statements

There are four different types of rights statement. These are listed here and explained more fully below:

1. Objects that are not protected by copyright and can therefore be freely re-used must be marked as being in the public domain by applying the Public Domain Mark (see 1 below).
2. When the data provider is also the rights holder and wants to make the digital object available for re-use (or has been authorised by the rights holder to do so) the data provider can apply a Creative Commons Licence or the CC0 1.0 Universal Public Domain Dedication (see 2 - 8 below).
3. When the data provider is also the rights holder and wants to make the digital object available without authorising re-use by third parties (or has been authorised by the rights holder to do so), the data provider can apply one of the three standardised Rights Reserved statements developed by Europeana (see 9 - 11 below).
4. Objects with a copyright status that is unclear (for example because no rights holder could be identified) can be marked with an 'unknown' copyright statement. This should only be used if absolutely necessary (see 12 below).

Within these four types, there are 12 separate statements that can be applied to a particular object.

Available Rights Statements

1 The Public Domain Mark (PDM)

All content that is in the public domain must be labeled accordingly. Europeana has worked with Creative Commons to develop a simple mark that indicates that a work is in the public domain – the Public Domain Mark. Works that are labeled as being in the public domain can be used by anyone without any restrictions. When showing the Public Domain Mark, Europeana will also link to [usage guidelines](#) for end-users.

The URL to use in the metadata is: `http://creativecommons.org/publicdomain/mark/1.0/`

`<europeana:rights>http://creativecommons.org/publicdomain/mark/1.0/</europeana:rights>`

See examples of objects with this rights statement on Europeana [here](#).

2 The Creative Commons CC0 1.0 Universal Public Domain Dedication (CC0)

If a rights holder wants to waive all the rights in a digital object, they can apply a CC0 waiver to the works in question. By applying this waiver, all rights in the content are waived and the objects can be used by anyone without any restrictions. CC0 can only be applied with the authority of the rights holder.

The URL to use in the metadata is: `http://creativecommons.org/publicdomain/zero/1.0/`

`<europeana:rights>http://creativecommons.org/publicdomain/zero/1.0/</europeana:rights>`

See examples of objects with this rights statement on Europeana [here](#).

Creative Commons Licences

The Creative Commons (CC) licenses provide options for copyright holders to allow others to re-use digitised objects under certain conditions. The CC licenses can only be applied with permission from the rights holder.

The following six CC licenses can be used as rights statements for digital objects in Europeana. The licenses are summarized below using the words from the CC website and are presented in order from the most open to the most restrictive. It is strongly recommended that you refer to the Creative Commons [website](#) to see the full definitions and legal code when making your decision.

The CC licenses also exist as jurisdiction-specific licenses. Europeana supports all available jurisdiction versions, however, the examples given below use the URLs of the 'universal' licenses. If you want to make objects available under a jurisdiction-specific Creative Commons license, you are advised to use the [selection tool](#) on the Creative Commons website. This will give you the correct license URL for the edm:rights field.

3 Creative Commons - Attribution (BY)

This license lets others distribute, remix, tweak, and build upon your work, even commercially, as long as they credit you for the original creation. This is the most accommodating of licenses offered. Recommended for maximum dissemination and use of licensed materials.

The URL to use in the metadata is: <http://creativecommons.org/licenses/by/3.0/>

<europeana:rights><http://creativecommons.org/licenses/by/3.0/><europeana:rights>

See examples of objects with this rights statement on Europeana [here](#).

4 Creative Commons - Attribution, ShareAlike (BY-SA)

This license lets others remix, tweak, and build upon your work even for commercial purposes, as long as they credit you and license their new creations under the identical terms. This license is often compared to 'copy left', free and open source software licenses. All new works based on yours will carry the same license, so any derivatives will also allow commercial use. This is the license used by Wikipedia, and is recommended for materials that would benefit from incorporating content from Wikipedia and similarly licensed projects.

The URL to use in the metadata is: <http://creativecommons.org/licenses/by-sa/3.0/>

<europeana:rights><http://creativecommons.org/licenses/by-sa/3.0/><europeana:rights>

See examples of objects with this rights statement on Europeana [here](#).

5 Creative Commons - Attribution, No Derivatives (BY-ND)

This license allows for redistribution, commercial and non-commercial, as long as it is passed along unchanged and in whole, with credit to you.

The URL to use in the metadata is: <http://creativecommons.org/licenses/by-nd/3.0/>

<europeana:rights><http://creativecommons.org/licenses/by-nd/3.0/><europeana:rights>

See examples of objects with this rights statement on Europeana [here](#).

6 Creative Commons - Attribution, Non-Commercial (BY-NC)

This license lets others remix, tweak, and build upon your work non-commercially, and although their new works must also acknowledge you and be non-commercial, they don't have to license their derivative works on the same terms.

The URL to use in the metadata is: <http://creativecommons.org/licenses/by-nc/3.0/>

<europeana:rights><http://creativecommons.org/licenses/by-nc/3.0/><europeana:rights>

See examples of objects with this rights statement on Europeana [here](#).

7 Creative Commons - Attribution, Non-Commercial, ShareAlike (BY-NC-SA)

This license lets others remix, tweak, and build upon your work non-commercially, as long as they credit you and license their new creations under the identical terms.

The URL to use in the metadata is: <http://creativecommons.org/licenses/by-nc-sa/3.0/>

<europeana:rights><http://creativecommons.org/licenses/by-nc-sa/3.0/></europeana:rights>

See examples of objects with this rights statement on Europeana [here](#).

8 Creative Commons - Attribution, Non-Commercial, No Derivatives (BY-NC-ND)

This license is the most restrictive of our six main licenses, only allowing others to download your works and share them with others as long as they credit you, but they can't change them in any way or use them commercially.

The URL to use in the metadata is: <http://creativecommons.org/licenses/by-nc-nd/3.0/>

<europeana:rights><http://creativecommons.org/licenses/by-nc-nd/3.0/></europeana:rights>

See examples of objects with this rights statement on Europeana [here](#).

Europeana Rights Reserved Statements

For data providers who do not want or cannot allow object re-use, Europeana has developed three standard rights statements. These statements express the conditions under which objects can be accessed on the data provider's website. Use of these statements means the data provider is reserving the rights in the digital object and that the object may not be used without additional permissions.

9 Rights Reserved – Free Access

This rights statement is applicable when users have free (as in gratis), direct and full access to the digitised object on the data provider's website.

The URL to use in the metadata is: <http://www.europeana.eu/rights/rr-f/>

<europeana:rights><http://www.europeana.eu/rights/rr-f/></europeana:rights>

See examples of objects with this rights statement on Europeana [here](#).

10 Rights Reserved – Paid Access

This rights statement is applicable when users need to pay data providers to gain access to the digitised work on the data provider's website. This may be the case if only a preview is accessible through the data provider's portal and registration and payment is required to gain access to the digitised object itself. In this case, the link from Europeana should give access to the metadata and (ideally) a low-resolution preview. Europeana will not link directly to a payment page.

The URL to use in the metadata is: <http://www.europeana.eu/rights/rr-p/>

<europeana:rights>http://www.europeana.eu/rights/rr-p/</europeana:rights>

See examples of objects with this rights statement on Europeana [here](#).

11 Rights Reserved – Restricted Access

This rights statement is applicable when there are limitations other than the requirement to pay a fee for accessing a digitised object on the data provider's website. For example, when a registration is required or only snippets or previews are available to users. In this case, the link from Europeana should give access to the metadata and (ideally) a low-resolution preview. Europeana will not link directly to a registration page.

The URL to use in the metadata is: <http://www.europeana.eu/rights/rr-r/>

<europeana:rights>http://www.europeana.eu/rights/rr-r/</europeana:rights>

See examples of objects with this rights statement on Europeana [here](#).

12 Unknown

The Unknown rights statement can be applied to objects for which the data provider does not have conclusive information pertaining to the rights status of the digital object (e.g. orphan works). This value is only to be used when the copyright status of the work described is unknown. This statement may be used by Europeana to exclude items from display and should therefore not be used without consultation with the Europeana Ingestion team.

The URL to use in the metadata is: <http://www.europeana.eu/rights/unknown/>

<europeana:rights>http://www.europeana.eu/rights/unknown/</europeana:rights>

See examples of objects with this rights statement on Europeana [here](#).

The value of the EDM: rights statement is used for the rights badge displayed under the preview. This badge is both human and machine readable. See example:



De bocht van de Herengracht bij de Nieuwe Spiegelstraat te Amsterdam

Creator: [schilder: Berckheyde, Gerrit Adriaensz.](#) | ▶

Contributor: [Aankoop met steun van het Ministerie van CRM en de Rijksmuseum-Stichting](#)

Date: [derde kwart 17e eeuw](#) ; 1672 - 1672

Type: [schilderij](#) | ▶

Description: De bocht van de Herengracht bij de Nieuwe Spiegelstraat te Amsterdam, gezien vanaf de Leidsestraat.

Data provider: [Rijksmuseum](#) | ▶

Provider: [Rijksmuseum](#) | ▶ [Netherlands](#) | ▶

Language: nl

Publisher: [Rijksmuseum, Amsterdam](#) | ▶

Identifer: SK-A-4750 ; RM0001.COLLECT.6000

Format: breedte 63 cm ; hoogte 40.5 cm ; image/jpeg ; olieverf ; paneel

View item at [Rijksmuseum](#)

PUBLIC DOMAIN

Explore further!

Similar content

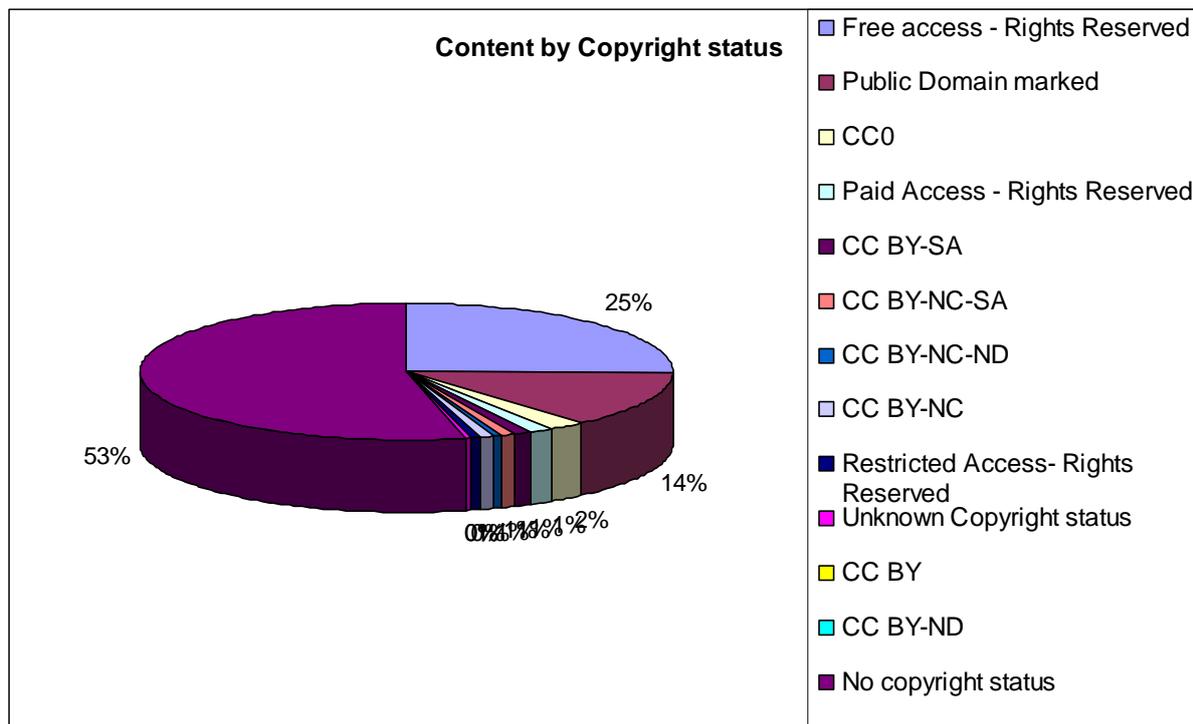


Europeana has commissioned Kennisland to research and analyze the quality of rights labeling: missing rights statements and % (in) correct rights labeling.

The results of this research will be available for Europeana soon. On basis of the analysis a Quality Improvement Plan and Rights Information Awareness Campaign will be planned for 2012-2013. A special emphasis will be on Public Domain.

Europeana has set as a target of 5 million labeled open-license objects (PDM, CC0, CC-BY, CC-BY-SA) to be reached in 2012.

The content report of June gives the latest overview of Content in Europeana by Copyright Status



Europeana content by Copyright status

Rights statement	Total	Percentage
Total of records	23.431.250	
Free access - Rights Reserved	5.896.508	25%
Public Domain marked	3.167.252	14%
CC0	503.332	2%
Paid Access - Rights Reserved	317.370	1%
CC BY-SA	278.848	1%
CC BY-NC-SA	176.113	1%
CC BY-NC-ND	158.328	1%
CC BY-NC	156.201	1%
Restricted Access- Rights Reserved	127.169	1%
Unknown Copyright status	70.570	0%
CC BY	41.556	0%
CC BY-ND	3.163	0%
No copyright status	12.534.840	53%

A revised Rights Guidelines and an online tool to support providers to correctly label their digital objects have been planned and are currently under development.

2.5 Quality assurance of other mandatory elements

In the [metadata mapping and normalization guidelines](#) **best practices** have been set for the (mandatory) elements. Europeana relies on the (data) providers and aggregators to deliver the mandatory elements.

Europeana currently checks samples for quality of the data. When the quality is insufficient, Europeana asks the aggregator to deliver higher quality.

Europeana communicates best practices to all aggregators and data (providers) and will advocate this in a series of Aggregator/data provider workshops.

Data Quality of the elements EDM/Europeana: isShownBy, EDM/Europeana: isShownAt and EDM/Europeana: object is addressed in *Chapter 3 Linking to the digital objects*

For a complete overview consult the metadata mapping and normalization guidelines.

Example with best practices from the mapping guidelines for element EDM/ESE: Title

3.3.1 Title

`dc:title`. A value for this element **or** [dc:description](#) must be provided.

The title or name by which the digital object is known. This is likely to be the title or name also applied to the original physical object in the case of a digitisation. Use the refinement [dcterms:alternative](#) for any title variants, translations etc.

Although we hope to show translations of titles using the `xml:lang` attribute in future, the portal can only display one instance of `dc:title` at present. Translations and variations should therefore be provided using the `dc:alternative` property as well.

Where many objects share the same title, as may be the case where there are many issues of a periodical, consider aggregating the title values with information from another element to give a more useful, non-ambiguous display.

Example:

```
<dc:title>A-Z : Luxemburger illustrierte Wochenschrift</dc:title>
<dc:description>1933-12-24 (Numéro 01)</dc:description>
```

Aggregate the values using a full stop as a separator to give:

```
<dc:title>A-Z : Luxemburger illustrierte Wochenschrift. 1933-12-24 (Numéro 01)</dc:title>
```

Dc example:

```
<dc:title xml:lang="en">Eight weeks</dc:title>
<dc:title xml:lang="it">Otto settimane</dc:title>
```

MODS crosswalk⁸ example

```
<mods:titleInfo><mods:title>Florida Environments Online</mods:title></mods:titleInfo>
becomes
```

```
<dc:title>Florida Environments Online</dc:title>
```

Note that MODS defines several Type attributes and sub-elements for the `<titleInfo>` element which may more accurately be mapped to `dcterms:alternative9`.

EAD 2002 XML example:

```
<unittitle>The Golden Gate entrance to the bay of San Francisco: sunrise</unittitle>
```

becomes

```
<dc:title>The Golden Gate entrance to the bay of San Francisco: sunrise</dc:title>
```

⁸ The MODS examples for dc elements are gratefully quoted from <http://www.loc.gov/standards/mods/dcsimple-mods.html>. The examples for dcterms are not taken from this source.

⁹ See also <http://www.loc.gov/standards/mods/v3/mods-userguide-elements.html>

Chapter 3. Linking to the digital objects

Europeana links to digital object on the side of the aggregators and (data) providers. There is a mutual interest for Europeana and the (data) providers to keep the links as stable as possible for users.

Alas it often happens that (data) providers change their structuring and therefore change their links. If they do not have a resolver/redirect in place this will generate broken links on Europeana.

Europeana has taken action to get stable persistent identifiers in place

- by installing a Taskforce Persistent Identifiers,
- and by revising the technical infrastructure,
- in implementing link checker functionality in Unified Ingestion Manager,
- and is working with data providers on improving the quality of the EDM/ESE elements isShownby, isShownat and object.

3.1 Taskforce Persistent identifiers

The taskforce investigates with network partners the feasibility of Persistent Identifiers. Technical and organizational aspects of a PID infrastructure are on the agenda of the taskforce.

The taskforce is scheduled to report in fall 2012. Europeana will follow-up on the recommendations of the taskforce

3.2 Revising the technical infrastructure, identifiers and implementing link check functionality in the United Ingestion Manager (UIM)

The revision of the technical infrastructure of Europeana had the following requirements:

1. Europeana must not break its own links
2. Existing links must work also in the future
3. IDs/links must be backwards resolvable

These requirements are currently under development as the technical infrastructure needed to be revised.

Europeana has revised the creation of the Europeana URI.

EDM is stricter on providing unique record IDs on a collection level by (data) providers. Validation of data with EDM schema sets unique record IDs mandatory. Therefore Europeana can create more persistent IDs. These consist of two parts

- collection ID (later collection_ID) and
- record ID (later local_ID), to guarantee that they are unique in Europeana scope.

Europeana ID format

Europeana persistent ID format is /collection_ID/local_ID where

- Slash (/) character before collection_ID
- collection_ID is Europeana collection ID.

- Although it is called collection ID, it contains both data provider ID and collection identifier. It may also contain a letter code in the split collection case. E.g. 54321 , where provider ID is 543, collection identifier part is 21, or letter code 12345A, 12345B... , where provider ID is 123, collection identifier part is 45, letter code is A, B...
- Slash (/) character separating collection_ID and local_ID parts
- local_ID is the local id (rdf:about attribute of the ProvidedCHO resource in EDM) provided by data provider either as it is or slightly modified.
- <edm:ProvidedCHO rdf:about="local_ID">

Link checker functionality in United Ingestion Manager (UIM)

Europeana is implementing structural link checking functionality in the United Ingestion Manager (UIM).

This will enable the ingestion team to run frequent checks on the collections. Europeana will define link checking rules at the last stage of implementation.

The full implementation of UIM is planned for fall 2012.

Europeana will finalize the implementation of the revised technical infrastructure, including taking UIM in production in fall 2012.

The guidelines will be updated with best practices to instruct the (data) providers on how they can provide Europeana with stable identifiers and links.

In addition we will continue to do a daily collection level spot check (checking three random records from each collection) to identify structural problems with individual collections, such as servers moving etc.

3.3 Data Quality of EDM/ESE elements: *isShownBy*, *isShownAt* and *Object*

Europeana has defined 3 elements in EDM/ESE to link to the digital object:

- EDM/Europeana: *isShownAt*
- EDM/Europeana: *isShownBy*
- EDM/Europeana: *Object*

It is **mandatory** to use at least one of the two elements, or both. They can be used both to link to the actual digital object (*isShownBy*) and to the digital object in context (*isShownAt*).

The element EDM/Europeana: *object* is used to generate the preview/thumbnail that is visible on Europeana.

As described above these links should be as stable and persistent as possible, and we recommend that Data Providers use an established Persistent Identifier framework, such as Handle, DOI and URN/NBN etc.

EDM/Europeana:*isShownAt*

This element will be active in the portal and will provide the link to the digital object in full information context on the provider website. It is a complementary element to EDM/Europeana: *isShownBy* (and it is mandatory to provide a URL link in one of these elements.)

Europeana focuses on giving access to digital objects and this two-fold URL strategy is

useful when ESE is not able to properly represent the original metadata model. If the digital object is displayed with local metadata, with a header or banner, or if the object is only accessible by clicking another icon on the local page then EDM/Europeana:isShownAt should be used.

Records that do not have a valid link in either EDM/Europeana:isShownAt or EDM/Europeana:isShownBy will be excluded from the portal.

EDM/Europeana: isShownBy

This element will be active in the portal and will provide a link to the digital object on the provider website. This is a complementary element to EDM/Europeana:isShownAt (and it is mandatory to provide a URL link in one of these elements).

To map to EDM/Europeana:isShownBy the object should be directly accessible by the URL and reasonably independent at that location. The inclusion of, for example, short copyright information or minimal navigation tools would be acceptable in this element. The link should be to the object in its best available resolution/quality.

(Note that it may be possible to construct URLs to objects that are embedded in HTML by right-clicking on the object and using the metadata identifier. In this case please provide the direct URL in EDM/Europeana:isShownBy as well as the link provided in EDM/Europeana:isShownAt.)

Europeana focuses on giving access to digital objects and records that do not have a valid link in either Europeana:isShownBy or Europeana:isShownAt will be excluded from the portal.

Europeana allows data providers to limit access to their digital object, e.g. when they are behind a pay wall or a log-in area when rights statements rights reserved - restricted access is applicable. As Europeana is in the pilot phase in establishing collaborations with commercial partners, e.g. publishers Europeana might get more content under this conditions.

The business team will evaluate the user experience for this content in 2012-2013 and will define what is acceptable or not for Europeana and its users.

EDM/Europeana: object

EDM/Europeana:object supports the process of creating small preview images for use in the portal. There are two situations:

- 1) A suitable image can be provided for image generation so its URL is mapped to Europeana:object. This may be the same URL as in europeana:isShownBy.
- 2) There is no suitable image with a URL. In this case nothing can be mapped to Europeana:object so a default icon for the object type will be used in the portal. The specifications for suitable source images and further details of how small images are used and processed in Europeana can be found in the Europeana Portal Image Policy.

If a record contains several pages in a PDF, the front page is used as the source of the images. If a record contains several image files, the first file that appears is used as the

source. Please ensure that the page that will be used is an appropriate image to represent the object e.g. it does not contain the colour scale used in the digitisation process.

Quality of the preview/thumbnail

The quality of the digital object is defined in the [Europeana Portal Image Policy](#)

This policy has been developed to as part of the initiative to improve the features and functionality of the Europeana portal by improving aspects of the data submitted. This document specifies the standards for the source images that should be provided for the creation of small images for use in the portal and explains how they are used.

To give users a preview of the objects they have found and to make an attractive interface to the portal, Europeana creates images in two sizes from the objects providers submit. If the user is not to be disappointed these small images must be of a reasonable quality. They are generated from the source images whose links are provided in the metadata and the small images resulting are cached in the Europeana system. Note that no high quality or large format source objects are stored in the Europeana system.

What sort of images should be provided as the source?

The URL of a source image from which Europeana can generate the required small images should be provided in the europeana:object element. This may be the same URL as given in europeana:isShownBy.

This URL must link directly to an object which:

- is an image file (e.g. <http://www.server.org/image.jpeg>) and NOT an image embedded within a webpage.
- has a width of at least 200px
- is ideally a jpg file (or another image format supported by ImageMagick)
- alternatively is a PDF in which case the images will be created from the first page of this PDF. (Providers should ensure that the first page is a suitable image and not a blank page or a page containing the colour scales etc.)

Providers should not supply a link to an image that is itself already the size of a thumbnail as this produces very poor quality results when it is used in the portal functions described. Similarly, the source image should not have a watermark nor should it be a local default thumbnail image as these do not give a good result.

Future work on improving quality of previews linking to Audio, Video and 3D

For Audio/Video and 3D material best practices and guidelines to create meaningful previews need to be defined. This can be done in collaboration with the European projects as e.g. CARARE, Euscreen.

Conclusion and work plan

Quality assurance of the metadata in the Europeana repository will focus in 2012-2013 on improving:

- Aligning the content strategy and content quality. In cooperation with the business team the alignment between the content strategy and data quality will take place.
 - Evaluations of pilots with commercial data providers,,
 - Establishing objective (technical and content-related) criteria for accepting/rejecting content for Europeana,
 - Evaluate the current workflow procedure and optimize where possible
 - Plan a series of workshops for both teams

- Quality of the metadata
 - Focus on mandatory elements, highest priority on correct rights labels.
 - Monitoring implementation of DEA
 - Support and monitoring implementation of EDM
 - Best practices on (mandatory) elements is advocated
 - Documentation, guidelines and best practices will be updated.
 - Active campaign and advocacy in 2012-2013 to contact providers that have missing mandatory fields (all older versions of ESE) to upgrade their data to ESE3.4.1 or EDM.
 - series of workshops for aggregators and data providers on metadata quality and EDM

- Linking to the digital objects
 - Follow-up on recommendations of Taskforce on Persistent Identifiers
 - Implementing technical revision to support stable Europeana Identifiers
 - Future work on image portal display of Audio, Video and 3D content

Data Ingestion Work plan 2012-2013

Q3 2012 - Follow through of DEA, implement UIM and EDM ICT-infrastructure with more persistent identifiers

Q4 2012 - Focus on communicating EDM, revised guidelines and updated schema to aggregators and data providers. Follow-up on the analysis of (in)correct rights statements and older versions of ESE. Start of workshops. Align Annual Plan, Content Strategy and Metadata Quality

Q1 2013 - Evaluate data quality of first EDM data. Follow-up on the analysis of (in)correct rights statements and older versions of ESE. Workshops on metadata quality.

Q2 2013 - Evaluate data quality of first EDM data. Follow-up on the analysis of (in)correct rights statements and older versions of ESE. Workshops on metadata quality.

Appendix

[Strategic Plan 2011-2015](#)
[Europeana Licensing Framework](#)
[Guidelines for the europeana:rights metadata element](#)
[Technical requirements](#)

ESE3.4.1 documentation

[Europeana Semantic Elements specifications v3.4.1](#)
[Metadata mapping and normalisation Guidelines v2.1.1](#)
[Europeana Semantic Elements v3.4 XML Schema \(valid for 3.4.1\) | About the Europeana Semantic Elements v3.4 XML Schema \(valid for 3.4.1\)](#)
[Europeana Content Checker version 2.0 User Guide](#)

Useful resources

[Europeana Rights Guidelines](#)
[Technical Report: Archival Digital Object Ingestion into Europeana \(ESE-EAD harmonisation\) v1.0](#)
[Europeana Portal Image policy](#)

EDM documentation

[Europeana Data Model Definition v5.2.3](#)
[Europeana Data Model Primer](#)
[Europeana Data Model Mapping Guidelines](#)
[Europeana Data Model Factsheet](#)
[Europeana Data Model presentation \(for v5.2.3\)](#)

The [XML schema](#): this is the XML schema for the first implementation of EDM.

The [EDM object templates](#)