



## → Report and Recommendations from the Task Force on Metadata Quality

Task Force Chair: Marie-Claire Dangerfield (EF)

Task Force Co-Chair: Lisette Kalshoven (KL)

December 2013 - May 2015.



## Contents

<b>Executive Summary</b>	<b>3</b>
<b>Background</b>	<b>5</b>
<b>Task Force members</b>	<b>6</b>
<b>Task Force Scope</b>	<b>7</b>
<b>Importance of Good Quality Metadata to Europeana</b>	<b>8</b>
<b>Key Elements of Good Quality Metadata</b>	<b>8</b>
1. Trusted process	9
1) Findable	14
3. Human and Machine Readable	21
4. Standardised	23
5. Meaningful to users	24
6. Clear on re-use	27
7. Visible	28
<b>How Metadata Quality is assessed in Europeana</b>	<b>32</b>
2.1 Mandatory metadata elements	32
2.2 Metadata quality checks by the Europeana Aggregation team	34
2.3 The Role of Partners in Metadata Quality Checks	40
<b>Blockers to good metadata quality submission</b>	<b>42</b>
3.2 Metadata as a by-product of the preservation process	42
3.3 Lack of Digital Appraisal	43
3.4 Limited resources	43
3.5 Type of Material Digitised	43
3.6 Lack of Understanding of Technical Documentation	44
<b>Task Force Recommendations: Future Steps</b>	<b>51</b>
<b>Technical</b>	<b>52</b>



## Executive Summary

Metadata is the lifeblood of Europeana. It is only by describing digital collections with rich metadata that they can become part of the shared cultural heritage landscape of Europe.

Metadata quality is controlled by a set of processes which ensures that cultural heritage objects can be identified, discovered and seen in context by audiences, in a manner appropriate to the context in which the data provider created them. Metadata must include information on the potential re-use of cultural heritage objects.

This report looks at how data partners' motivation, the technical requirements and the content of the metadata affect overall metadata quality. We believe this document to be relevant to the entire Europeana Network.

### **We have defined good metadata quality as:**

1. Resulting from a series of trusted processes
2. Findable
3. Readable
4. Standardised
5. Meaningful to audiences
6. Clear on re-use
7. Visible

### **The Task Force's recommendations are as follows:**

1. The metadata processes undertaken by Europeana and aggregators should be more transparent - to create more trust, all parties should be involved in the development of documentation and organisational processes.
2. Existing documentation on the Europeana Data Model available on the 'Share your documentation' section of Europeana Pro should be revisited and, where necessary, rewritten or expanded, to make sure it works for its intended audience.
3. Metadata crosswalks need to be documented so that institutions can learn from past experiences and domain-specific metadata issues.
4. Metadata submitted to Europeana should have relevant, human-readable fields, or be structured in a way that makes logical sense to a human being (in the case of hierarchical structured datasets) and objects should be listed as part of a collection.
5. The Task Force recommends that the minimum quality standard be raised, with richer data provided in the mandatory elements, to ensure that the digital cultural heritage objects submitted by partners can be discovered.
6. The Task Force encourages the curation of datasets, and recommends that smaller curated datasets with greater metadata quality are submitted and should be given priority in ingestion and display via Europeana.



7. The Task Force recommends greater use of the Europeana Publication Policy to communicate the processes and requirements needed in relation to metadata in Europeana.
8. There should be greater use of the Europeana Publication Policy to communicate the processes and requirements needed in relation to metadata in Europeana.
9. Data should be enriched where possible (e.g. through multilinguality or vocabularies). The recommendations made by the Europeana Task Force on Semantic Enrichment and Multilinguality should be implemented.
10. Open vocabularies such as Iconclass and Getty Art and Architecture Thesaurus should be used where possible and the pertinent URI should be added to the metadata to ensure its exposure online.
11. The Task Force recommends that the Europeana Aggregation team be more available to the data partners.
12. Data should be exported and checked as early as possible and a Europeana Data Model (EDM) content checker should be implemented (using a preview portal as a means of data validation).
13. Increased metadata checks should be taken up by the aggregator prior to submission.
14. A thumbnail should be made available for all digital cultural heritage objects.
15. Templates should be used to ensure a certain metadata standard is met.
16. Further areas to be explored include the following:
  - a. Motivating factors in the provision of good metadata from data providers
  - b. Impact of curation of collections on traffic and dissemination
  - c. Data quality metrics
  - d. EDM validation



## Background

Europeana aggregates metadata submitted by over 3,000 cultural heritage institutions in Europe. Quality regulation is a constant and pressing issue, despite both aggregating partners and Europeana conducting checks on metadata on a regular basis.

The Task Force on Metadata Quality was established to see why there is such a variety in the quality of data submitted to Europeana, and to look at the key obstacles surrounding the submission of high quality metadata.

The Task Force also aimed to answer questions raised by the following: the Europeana Task Force on Multilingual and Semantic Enrichment;<sup>1</sup> the deliverables on rights statements from the Europeana Awareness project;<sup>2</sup> the development of the Europeana Content Re-Use Framework; discussion with data providers.

These questions included:

- How does Europeana define high quality metadata?
- What is preventing the submission of good quality metadata?
- How should data providers add information that is not in the original metadata but is required by Europeana?
- What is the difference between domain-specific metadata and what is required for Europeana Data Model metadata?
- What can be done to help the Europeana Network to increase the submission of high quality metadata?
- Who is responsible for checking data quality - data providers, aggregators or Europeana?

This report answers these questions and discusses the importance of good metadata in relation to the functions of the Europeana repository and related services.

The Task Force on Metadata Quality was established in Q4 of 2013, and held its first exploratory meeting during the Europeana Network AGM in Rotterdam. Prior to a face-to-face meeting in The Hague in April 2014, there were several Skype calls to discuss the matters at hand. The subject was also discussed during the Europeana Aggregator Forum meeting in May 2014.

---

<sup>1</sup> EuropeanaTech Task Force on a Multilingual and Semantic Enrichment Strategy: final report <http://pro.europeana.eu/get-involved/europeana-tech/europeanatech-task-forces/multilingual-and-semantic-enrichment-strategy>

<sup>2</sup> Europeana Awareness Deliverable 5.3: Evaluation report on the Europeana Licensing Framework [http://pro.europeana.eu/files/Europeana\\_Professional/Projects/Project\\_list/Europeana\\_Awareness/Deliverables/EA%20D5\\_5%20EAwareness\\_IPR.pdf](http://pro.europeana.eu/files/Europeana_Professional/Projects/Project_list/Europeana_Awareness/Deliverables/EA%20D5_5%20EAwareness_IPR.pdf)



## Task Force members

The Europeana Metadata Quality Task Force was made up of 11 members representing a variety of institutions, aggregators, data providers and metadata experience. The group's task was to accumulate, assess and propose solutions to some of the key issues raised by data providers and end-users during the creation, submission and display of metadata in Europeana. The members of the Task Force were:

- Marie-Claire Dangerfield, Europeana Foundation.
- Lisette Kalshoven, Kennisland
- Francesca Schulze, Deutsche Digitale Bibliothek (DDB)
- Martin Reisacher, Landesarchiv Baden-Württemberg
- Grace Toland, Irish Traditional Music Archive
- Eve-Marie Oesterlen, British Universities Film & Video Council (EUScreenXL)
- Juliane Stiller, Max Planck Institute
- Nacha van Steen, Koninklijke Musea voor Kunst en Geschiedenis, Brussel (KMKG-MRAH)
- Adina Ciocoiu, The Europeana Library (TEL)
- Andra Patterson, The British Library (EuropeanaSounds)
- Valentine Charles, Europeana Foundation
- Cecile Devarenne, Europeana Foundation
- Robina Clayphan, Europeana Foundation
- Chiara Latronico, Europeana Foundation

With additional commentary from Antoine Isaac (Europeana Foundation), Henning Scholz (Europeana Foundation) and Joris Pekel (Europeana Foundation). Many thanks to all involved, as well as Dimitra Atsidis and Francesca Morselli, and special thanks to Yorgos Mamakis (Europeana Foundation).



## Task Force Scope

The information in this document is primarily designed to ensure that good quality metadata is submitted for inclusion in the Europeana repository. It can also be used as a general guide for the creation of metadata for use elsewhere.

This report is not designed to criticise individual institutional databases in which the format used appropriately describes and catalogues a given collection. It will not outline quantitative measures to improve metadata quality; we recommend that such measures could be investigated in future Task Forces.

The Europeana Data Model (EDM) was designed to provide a baseline for a variety of types of metadata, and functions as a cross-domain metadata standard which allows the aggregation of material from a variety of institutions across Europe. EDM requires that one webpage represents a single digital representation of a cultural heritage object, but this can create issues when directly converting an existing database or finding aid to EDM.

Each form of cultural heritage metadata will have its own structure with its own specific criteria (e.g. MARC21 for libraries, EAD for Archives, LIDO for museums, as well as METS/MODS for digitised print etc). The information provided in each metadata field is a result of the system of rules applied within a specific domain or for specific kinds of work. What is appropriate for a library may not be appropriate for an archive or a museum. However, many of the smaller institutions do not apply even general guidelines and instead develop their own rules around metadata creation.

Some metadata types, in particular Encoded Archival Description (EAD), may provide richer metadata in the original format than after it has been transformed into EDM. Mapping to EDM before transforming all the data is useful as it shows where the metadata needs to be adjusted to achieve the best display in Europeana.

The recommendations from this Task Force relate to EDM data quality. The report outlines what 'good' metadata quality is to Europeana, its portal and API, and how data providers can achieve this. It is hoped that the take-up of these recommendations will result in a marked improvement in metadata created by institutions and material made available via Europeana in the future.



## Importance of Good Quality Metadata to Europeana

Europeana collects metadata which represents cultural heritage and has been created within the cultural heritage sector.

In the world of digital cultural heritage, good quality metadata is vital.

Metadata allows digital cultural heritage objects to be described, disseminated and found. It should be considered as the key product of digitised cultural heritage. To undervalue the importance of metadata is to devalue the work of the data creator, the cultural heritage object, the institution and the audience for our cultural heritage.

To produce good quality metadata, there must be a series of processes in place which ensure that metadata is as rich as possible while remaining true to the spirit in which it was created.

Without a certain standard of metadata quality, digitised objects remain hidden and unfindable, and therefore of little use to either the contributing institution or its audiences. A lack of defined metadata quality elements, in terms of technical and descriptive criteria, also creates a digital legacy issue - providers may publish with an intent to fix problems later but never re-submit a fixed version. This approach may work within an institution's own systems, but attempts at publishing first and fixing later have not been successful for Europeana because of the considerable additional investment of both time and financial resources required to revisit already published collections.

We must not think about metadata solely in the context of an end product - something which is delivered and then published on Europeana. We must also think about its quality and purpose at the moment of its creation, considering why it is being created and for whom.

### Key Elements of Good Quality Metadata

The Task Force identified seven elements that are necessary for the creation of good quality metadata. These elements balance human-created/readable metadata requirements with technical ones. These areas could be broken into further granularity (and we hope they will be in future) but this was beyond the scope of the Task Force.





## 1. Trusted process

# Europeanana Ingestion Workflow

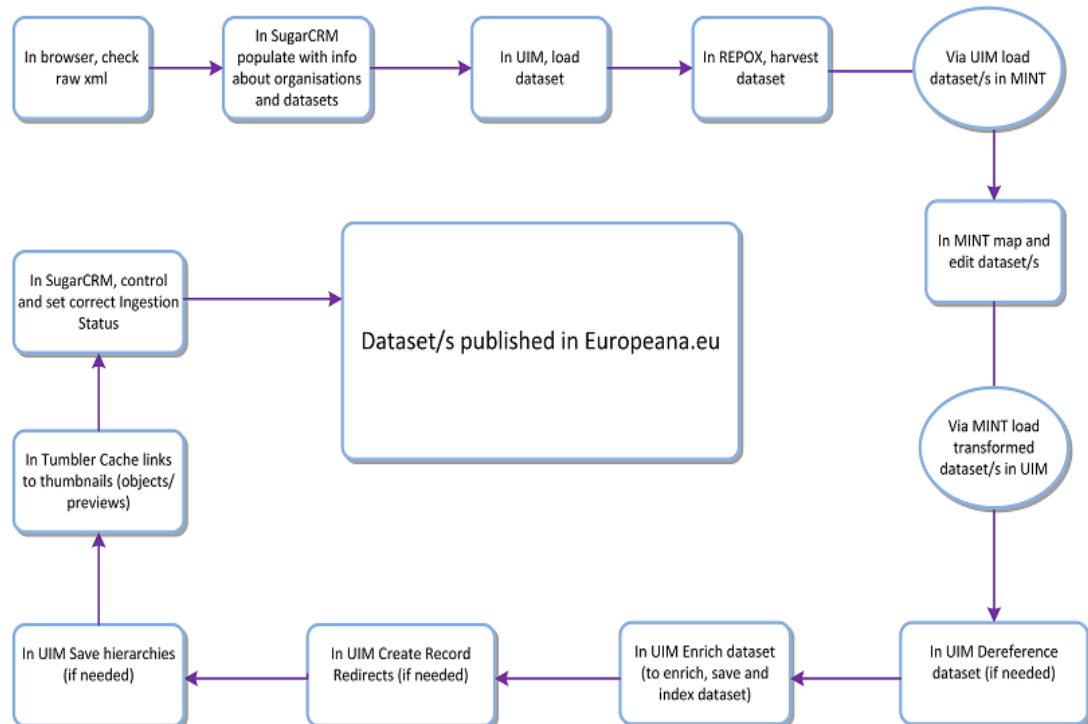


Fig. 1.1 The workflow processes used in publishing metadata to Europeanana

To achieve high quality metadata, there must be a series of trusted processes in place between the data creator, data provider, aggregator and Europeanana. The diagram above indicates the number of processes a dataset goes through between being submitted to Europeanana and published through the portal/API.

The systems of transformation used in any process need to be trusted to lose as little data (richness) as possible when metadata is converted from one format to another (e.g. from MARC to EDM). The same applies to the systems used before a dataset reaches Europeanana, for example, when metadata is converted from an institutional database to a format used by an aggregator.

### 1.1 How do these processes work?

When data partners submit collections to Europeanana, their digital datasets have either been created specifically for this purpose or have been exported directly from their own repositories. This means that the metadata comes from an authoritative source that the partner trusts. The partner needs to be sure that their metadata will resemble this original data after mapping to EDM for use in Europeanana.

If partners do not trust the transformation process, then the motivation for producing good quality metadata is reduced. This is especially true when institutions create data specifically for Europeanana (e.g. they add additional metadata for the sole purpose of making their authoritative records fit the EDM format).



The diagrams below show the workflows of three of Europeana's data partners- The European Library (TEL) (fig 1.2); the Deutsche Digitale Bibliothek (DDB) (fig 1.3) ; and the Landesarchiv Baden-Württemberg (fig 1.4). Add the aggregators workflows to Europeana's and you can see that metadata passes through around 25 processes. The Landesarchiv Baden-Württemberg diagram shows additional processes for export or enrichment.

The Task Force was unable to source a concrete diagram of the transformations a dataset goes through on the data creator side (i.e. before the data reaches the aggregator), which would be useful in further investigations of metadata quality. The number of steps, and the possible steps during which a loss in metadata quality could occur, serve to highlight the importance of creating a series of trusted processes between all parties.

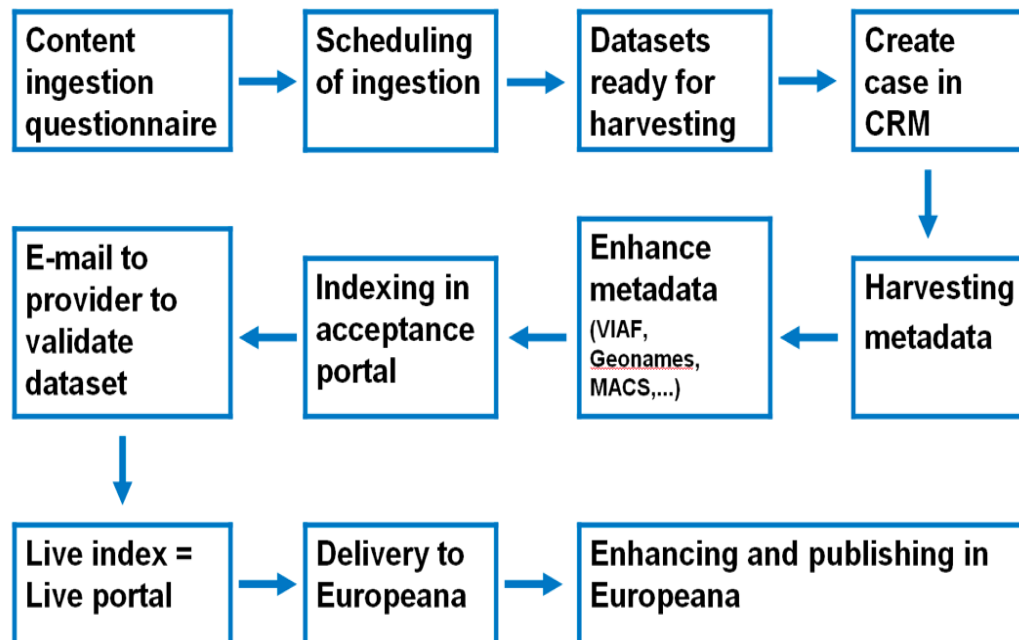


Fig. 1.2 depicts the ingestion process of The European Library

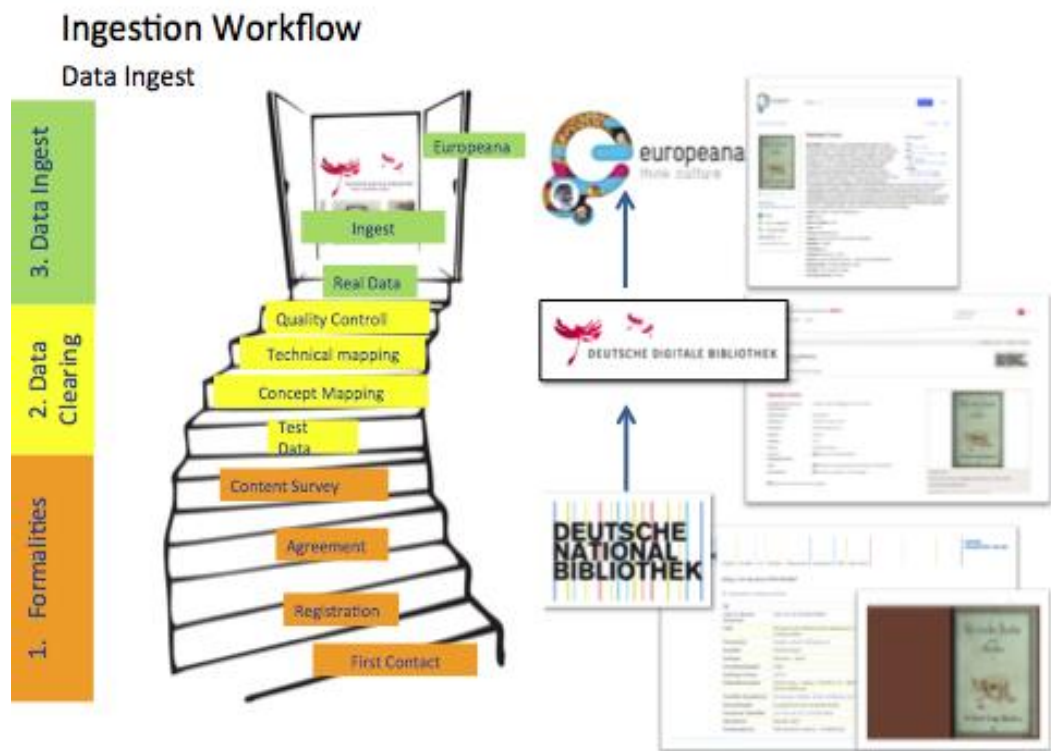


Fig 1.3 The ingestion workflow of the Deutsche Digitale Bibliothek

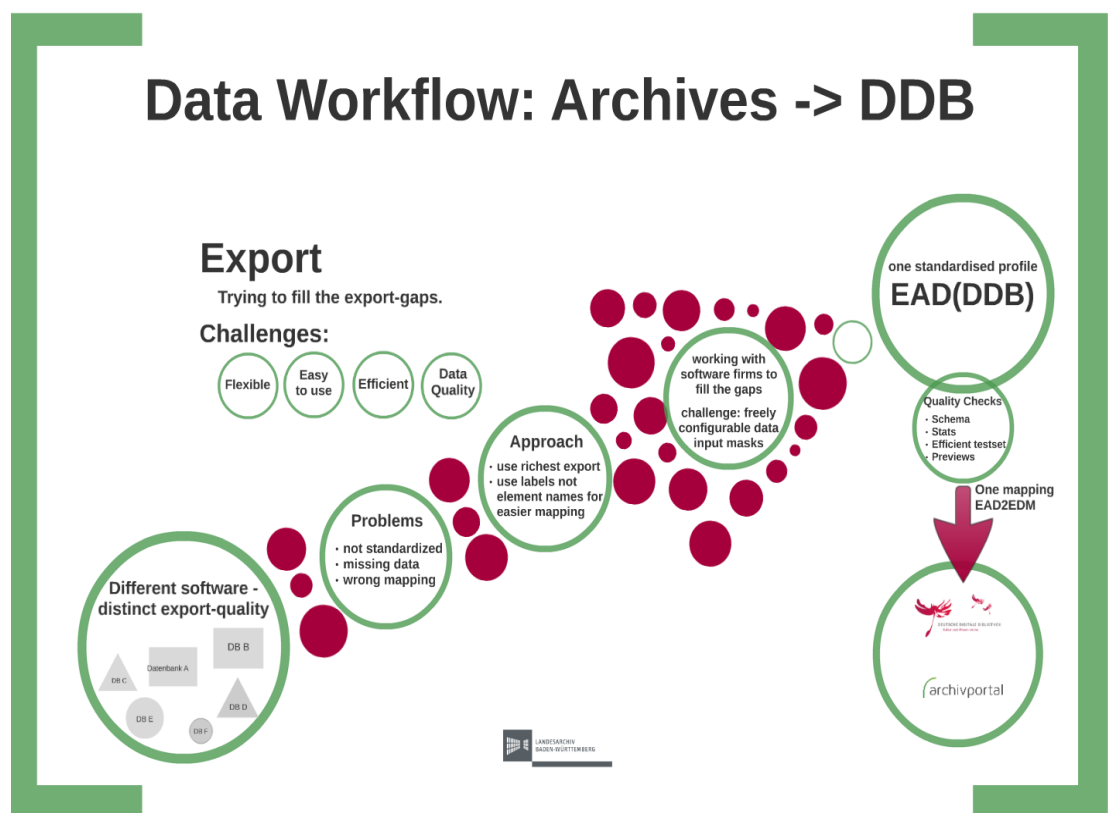


Fig 1.4 shows data export from Landesarchiv Baden-Württemberg to Deutsche Digitale Bibliothek DDB and Archives Portal Europe (ApeX)



Europeana aims to publish the richest metadata possible. However, metadata loss during transformation from one format to another is possible and common, as has been outlined in previous research.<sup>3</sup> Metadata loss may occur at any point during export to a project or aggregator and again from the aggregation point to publishing on Europeana. Since a metadata record may go through many transformations before reaching Europeana, all stages need to be trusted.

Members of the Task Force discussed how their data providers felt isolated from the ingestion process. In order to change that, projects and aggregators may have to increase their level of communication with their data providers and maintain a greater transparency of process.

As well as trusting that no metadata will be lost, there is also a need to trust that any augmentation of data which occurs outside of the mapping process does not alter the fields in the metadata. For example, automatic enrichment using the terms supplied in dc:subject may produce additional and confusing elements like multiple artists with the same name, or geo-spatial data which, when enriched, points to a different geographical area to the one intended. This process too must be trusted so that partners can see that their metadata is not corrupted.

- 1 The metadata processes undertaken by Europeana and aggregators should be more transparent - to create more trust, all parties should be involved in the development of documentation and organisational processes.

## 1.2 How can we create trust in the processes?

Trust in these processes can be improved by creating or promoting additional EDM or mapping guidelines that are relevant to their target audience. The Task Force reviewed the existing documentation and concluded that it could be improved by greater attention to the needs of the target audiences. For example, information on the technical issues surrounding ingestion should be aimed at technical staff, while documentation about which metadata fields to include in the mapping process should be aimed at a more non-technical audience.

- 2 The Task Force recommends that existing documentation on the Europeana Data Model available on the 'Share your documentation' section of Europeana Pro should be revisited and, where necessary, rewritten or expanded, to make sure it works for its intended audience.

Furthermore, it is important to document metadata crosswalks, so they can be shared with the wider cultural heritage community. Crosswalks match the named variables of one metadata format and align them with another metadata format for conversion or output purposes. The more information that is available and shared, the better equipped partners will be to assess and tackle issues coming from their various

---

<sup>3</sup> Realizing Lessons of the Last 20 Years: A Manifesto for Data Provisioning & Aggregation Service for the Digital Humanities (A Position Paper) <http://www.dlib.org/dlib/july14/oldman/07oldman.html>



institutions or domain-specific metadata.

- 3 Metadata crosswalks need to be documented so that institutions can learn from past experiences and domain-specific metadata issues.

Two EuropeanaTech Task Forces<sup>4</sup> have sought to gather such information. Europeana has also brought together several case studies on mapping domain-specific metadata to EDM<sup>5</sup>, and a number of projects have created their own EDM profiles for their data, including Europeana Fashion and Europeana Sounds. Similarly, the Deutsche Digitale Bibliothek (DDB) and the Digital Public Library of America (DPLA) have adapted EDM to suit their own institutional as well as providers' needs. Current EDM can be extended to suit most types of objects submitted to Europeana, without creating an individual profile for a chosen thematic or format type.

The Archives Portal Europe (ApeX) project has published a crosswalk for converting Encoded Archival Description (EAD) to apeEAD, the version of EAD required for the project.<sup>6</sup> While their document covers the entire standard, an archivist may actually only use a few elements to create a record within a hierarchical structure. The apeEAD format then needs to be converted to EDM for publishing, and so a further crosswalk will be required.

Metadata creation is affected by several factors: institutional requirements; the information available to the creator; time constraints; and to a certain extent, personal discretion. It then usually goes through a technical intermediary, then to a project, and finally to Europeana. It is easy to see the difficulty in creating end-to-end trust in such a long and complex process. This is why it so important to ensure efficient and effective communication between data creators, projects and aggregators and larger digital repositories as well as creators and publishers of authority vocabularies. This communication could happen through open discussion channels, e.g. digital or physical forums, or through explicit and appropriate documentation.

To improve cultural heritage metadata, institutions must believe in the value of the information housed within metadata and see it as an integral part of the digital cultural heritage landscape. By working together as a cultural heritage network and sharing knowledge of EDM experiences with the wider metadata community, we can build trust in, and therefore compliance with, our trusted processes.

---

<sup>4</sup> Report from the Task Force on EDM mappings, refinement and extensions

<http://pro.europeana.eu/documents/468623/bca65b72-fb8f-4b4f-802d-1072690ae33a>

Task Force on an EDM profile for Sound <http://pro.europeana.eu/documents/468623/9cbba0d6-c802-4d76-86e6-4298b0424458>

<sup>5</sup> EDM case studies <http://pro.europeana.eu/case-studies-edm>

<sup>6</sup> Mapping guide for using EAD in the Archives Portal Europe [http://www.apex-project.eu/images/docs/D2.6D4.8\\_mapping-guide\\_apeEAD.pdf](http://www.apex-project.eu/images/docs/D2.6D4.8_mapping-guide_apeEAD.pdf)



## 2 Findable

To be useful to audiences, cultural heritage objects have to be findable and presented in context. Good metadata allows for this. If people can find an institution's collections on Europeana, that institution benefits in terms of increased web traffic. The more comprehensive the metadata, the more likely it is to be viewed and used by someone either via the Europeana portal or the API.

It is important to understand that metadata records may comprise of both human-created and machine-created elements (e.g. elements of automatic enrichment). Depending on the circumstances of the cultural heritage object's creation, these two types of metadata may have been created at different times, sometimes years apart. When this report discusses findable records, it takes both human and machine-created elements into account, as it views them both as part of the metadata output found in Europeana.

- 4 Metadata submitted to Europeana should have relevant, human-readable fields, or be structured in a way that makes logical sense to a human being (in the case of hierarchical structured datasets) and objects should be listed as part of a collection.

### 2.1 What increases findability?

For objects to be found easily, they should have relevant, human-readable fields, or be structured in a way that makes logical sense to a human being (in the case of hierarchical structured datasets). Accessing records via Europeana's API can be useful, but if the metadata output is illogical at any juncture then it is a wasted effort for all parties (memory institutions, Europeana and the end-user).

There are a number of ways memory institutions can make their metadata more findable. These elements are important to take forward into future requirements for digitisation projects. The techniques outlined below are currently seldom used, possibly because the metadata creators are unaware of how the metadata could be used in future.

If the source metadata is poor, or if the dataset is created from an old inventory and there is limited information available at source, or if the dataset is a new digital collection but the digitisation was done rapidly with no checks, then the metadata may only have a limited number of fields containing minimal information. In this case, nothing will make the records more findable.

- **Standardised mapping**

It is important that metadata crosswalks are developed and shared by institutions and metadata specialists as mentioned in *1.2 How to create trusted processes*.

Crosswalks allow for a standardised mapping of fields when metadata is transformed from one format to another. The use of appropriate metadata structuring is crucial when mapping to EDM to avoid issues with the retrieval of records. For example, if



the 'title' of an object is mapped to the 'creator' field and the 'creator' field is mapped to a different field again, the object is unlikely to be found.

- **Keywords**

There are a number of ways in which items become more easily discoverable, including the use of relevant keywords and unique metadata elements. As with all internet searches, it helps to have relevant keywords in either the descriptive fields (dc:title or dc:description) or relevant subject fields (e.g. dc:subject or a dc:subject connected with skos:Concept class).

- **Unique elements**

Providing as much unique information as possible will increase the likelihood of the object's discovery.

There is currently an overall lack of unique metadata elements. Many records in Europeana have the same title, e.g. 'photograph'. In 2014, Europeana removed c.900,000 records from the portal for this reason as there was no useful accompanying metadata to explain context, so despite technically being correct metadata, they were removed for quality reasons. Many other records lack information to make them visible to end-users. Without unique elements, the records may become digitally invisible to their desired audience. This defeats the purpose of digitising records for discovery and is not beneficial to preservation because the material cannot be found easily in the future.

This issue of digital invisibility can be nicely illustrated with the use of the search term 'photograph'<sup>7</sup> which results in c. 36,000 matches when the title 'photograph' 'photographie' or 'fotograph' is searched for on the Europeana Portal. This makes records using the title 'photograph' without supplementary information in the metadata difficult to find. This is similarly true of collections where physical objects have been digitised and the type of object is used as its title e.g. 'Vase', 'Chair', 'Jug'. There may be no way around this for titles, particularly for museum collections, but in such cases, increased metadata for other elements should be provided e.g. creator, period created, materials used. If such information is not available then it may not be appropriate to include the object in a digital collection, as it is essentially undiscoverable.

---

<sup>7</sup> Search term Photograph

[http://www.europeana.eu/portal/search.html?query=proxy\\_dc\\_title%3A%28photograph+OR+Photograph+OR+photographie+OR+Photographie+OR+Fotograph+OR+fotograph%29&rows=24](http://www.europeana.eu/portal/search.html?query=proxy_dc_title%3A%28photograph+OR+Photograph+OR+photographie+OR+Photographie+OR+Fotograph+OR+fotograph%29&rows=24)



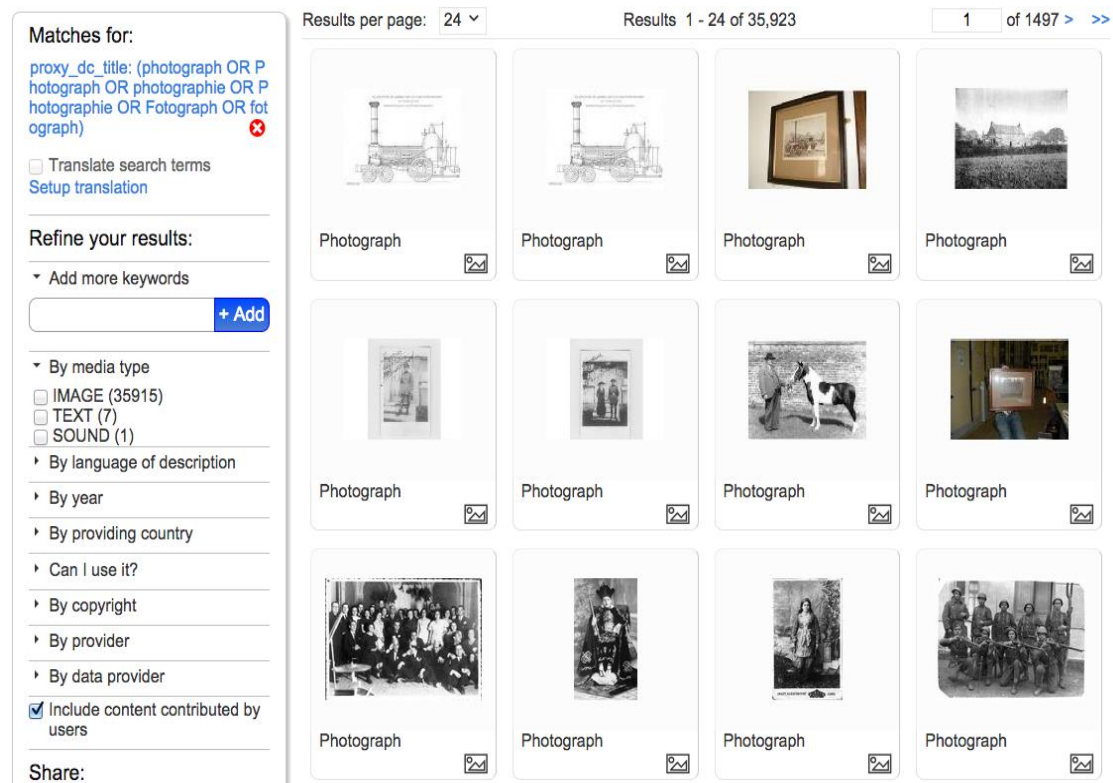


Fig. 2.1 example of search term “photograph/photographie/fotograph”

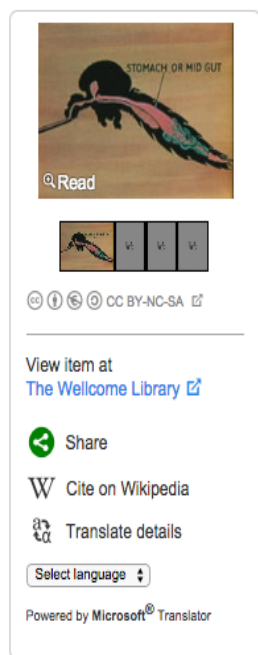
- **Unique Identifiers**

Objects will be more visible if a URI (Unique Resource Identifier) is provided to a relevant online vocabulary.

Where machine readable elements are concerned, repetition should be limited as much as possible. This means that the identifiers, both in terms of the dc:identifier and the rdf:about fields for both the ProvidedCHO and the Aggregation, should be unique and stable/persistent, so an object can be found under the same link after various update processes.

Europeana recommends including a field containing a unique value e.g. the identifier provided in the rdf:about of the ProvidedCho, particularly in records which may have near-duplicated metadata e.g. two jugs titled ‘jug’ created by the same artist. This identifying element is commonly taken from internal catalogues, and may not make sense to the end-user, but will be useful in terms of machine readability. As an identifier is often set by the data creator, this cannot be changed by the aggregator or Europeana. In the example below, the unique identifier is ‘0055-0000-3651-0202-0-0000-0000-0’. The identifier is not a human-readable element but it makes the record unique.





## The dissection of a mosquito for malaria parasite

### Description:

Wellcome Film Unit with Mr P.G. Shute and Dr. R.S. Bray of the Medical Research Council's former Malaria Reference Laboratory, Horton Hospital, Epsom, Surrey;

A detailed cinematographic record of the classic laboratory technique for dissecting a mosquito to detect malaria parasites. The film describes the oocyst and sporozoite phases of the life-cycle of the malaria parasite within the mosquito before demonstrating how the mosquito's salivary glands are pressed out and prepared for examination so that sporozoites can be identified, and the way in which the stomach is removed and the oocysts are detected.;

Audience: General public;

Segment 1; Opening credits. There is a shot of a female anopheles atroparvus species of mosquito feeding on blood from a human arm, the means whereby malaria is transferred. Diagrams show the anatomy of the mosquito and the lifecycle of the malaria parasite. Intertitle: Technique of Dissection: 1) removal of salivary glands. The technique is shown. Special dissection needles are shown; the mosquito is killed and made ready for dissection.;

Segment 2; The salivary glands are pressed out into a drop of saline; the sporozoites re-examined under the microscope. 2) removal of stomach. This is done on a slide and oocysts are seen. Commentary explains the oocyst rate and sporozoite - the latter is the strongest indicator of field malaria.

### Geographic coverage:

City, City of London, London, London City, The City  
Latitude: 51.50853; Longitude: -0.12574

### Publication date:

1953

### Type:

Monograph

### Format:

sd., col.; 1 encoded moving image (9.30 min.) :

### Subject:

Malaria.; Tropical Medicine.; Dissection.; Encoded moving images.

### Identifier:

0055-0000-3651-0202-0-0000-0000-0 .jpg

Fig 2.2 dc:identifier in context of a record from The Wellcome Library via The European Library<sup>8</sup>

- **Part of a collection**

As the Task Force recommends data providers focus on submitting *curated* datasets, it can be said here that ideally, digital cultural heritage objects should be listed as part of a collection. This can be achieved by using a hierarchical display or by using elements such as dcterms:isPartOf, edm:isRelatedTo, dc:relation, edm:isNextInSequence etc, or the soon to be implemented EDM Collection Profile<sup>9</sup> which expresses the context of the digital cultural objects represented.

It has been noted by Task Force members that data creators may be worried about their metadata being taken out of the context of its creation. While this is a concern with any digital material, by maintaining a context via a curated or collection dataset, it is hoped that this risk would be minimised.

<sup>8</sup> 'The dissection of a mosquito for malaria parasite' via The Wellcome Library  
[http://www.europeana.eu/portal/record/92086/BibliographicResource\\_1000086165671.html](http://www.europeana.eu/portal/record/92086/BibliographicResource_1000086165671.html)

<sup>9</sup> EDM Collection Profile document  
<https://www.dropbox.com/s/ers9toi0hivpcze/EDM%20Collection%20Profile.pdf>



Using a collection-based dataset or a dataset with hierarchical structure provides greater context for audiences, and may encourage them to explore the collections or institutions further. It also benefits records which may not have as much information at the record level but gain more in the context of a collection. For example 'Photograph 1723' may make sense in the context of a collection of 1,723 photographs of houses despite there being limited information in the record itself.

Linking records to collections is especially recommended for collections that are intended for use by the research community as researchers can then see if further material from the same dataset is pertinent to their work.

This Task Force notes that collection context is important and largely missing from Europeana. For this reason, we recommend that a further Task Force be set up to investigate how data partners could be motivated to provide better context in an explicit/formalised way.

5 The Task force recommends further investigation into how data partners could be motivated to provide better context in an explicit/formalised way



Fig 2.3 hierarchical display of Bundesarchiv collection submitted by the Archives Portal Europe<sup>10</sup>

<sup>10</sup> 'Beratungen des Sekretariats des ZK der SED mit den Ersten Kreissekretären, Abteilungsleitern des ZK und Funktionären des Parteiapparates' submitted by the Bundesarchiv Deutschland [http://www.europeana.eu/portal/record/2048346/providedCHO\\_dy30bho\\_DY\\_30\\_2152.html](http://www.europeana.eu/portal/record/2048346/providedCHO_dy30bho_DY_30_2152.html)



## 2.2 How poor metadata makes collections invisible

Collections are digitised for a reason, often as part of an organisation's preservation process. Whatever the reason for digitisation, the act of publishing data on the internet presumes that we want the information to be found. But beautiful digital images, texts, films or sounds without relevant accompanying metadata will sadly remain undiscovered.

With the recent focus and importance on digitising cultural heritage, the expectation is that increasing numbers of digitised objects available on the internet correlates with an increase in the amount of good quality metadata supplied, but this does not appear to be the case. A lack of metadata makes digitised cultural heritage invisible globally, not just in repositories like Europeana.

- 6 The Task Force recommends that the minimum quality standard be raised, with richer data provided in the mandatory elements, to ensure that the digital cultural heritage objects submitted by partners can be discovered.

To illustrate this, we will very briefly look at the `dc:title` field and `dc:identifier` fields using our example of 'Photograph 1723' again. As noted above, the title 'Photograph 1723' or 'image' is not particularly useful to someone trying to find a specific image, unless 'Photograph 1723' forms part of a greater and connected collection within the dataset. A much better alternative might be to give it a unique title such as 'Photograph of Art Deco House on Mainstreet number 18 in Leuven'.

- **Linked Open Data Vocabularies**

Europeana performs automatic semantic enrichment on top of the data delivered by its data providers. This process implies the addition of information to the data about certain resources such as agents, places, concepts and timespans. It also creates new links between these enriched resources and other reference datasets.

Europeana currently performs enrichment using open and multilingual vocabularies such as [Geonames](#), [Dbpedia](#) and [Gemet](#).

The Task Force on Multilingual and Semantic Enrichment Strategy has shown in its report that good quality data is crucial for preventing errors and flaws when enriching metadata.

Automatic enrichments can be very beneficial for enabling retrieval across languages and adding context to the resources beyond the scope of the Europeana platform. If automatically added enrichments are incorrect or ambiguous, the benefits are reversed, propagating the errors to several languages, impacting on the retrieval performance and giving audiences a bad experience. This Task Force recommends that data should be enriched when possible and that the recommendations made by



the Europeana Task Force on Semantic Enrichment and Multilinguality<sup>11</sup> should be implemented.

- |  |
|--|
| 7 Metadata should be enriched where possible using Linked Open Data vocabularies to ensure its exposure online and that the recommendations made by the Europeana Task Force on Semantic Enrichment and Multilinguality should be implemented. |
|--|

The addition of URIs to online vocabularies in metadata allows for digital cultural heritage objects to become findable for a wider audience, and can assist in providing multilingual access.

Other ways of making a record more discoverable are to include appropriate information for dc:subject, or to provide a concept class. This allows records featuring a dc:subject field to be enriched and linked to larger vocabularies such as the Getty Art and Architecture Thesaurus<sup>12</sup> PartagePlus vocabulary<sup>13</sup>, Gemeinsame Normdatei (GND)<sup>14</sup>, IconClass<sup>15</sup> and Virtual International Authority File (VIAF)<sup>16</sup>. Adding URIs to the metadata in either the dc:subject or a connected skos:Concept class gives further information and may enable multilingual translations of the dc:subject terms.

Another way to make records more visible is to use the edm:Place Class, which allows memory institutions to provide geospatial coordinates for the digital cultural heritage object. However, many institutions currently use this to submit the location of the institution in which the object is held, which is not exceptionally useful to audiences and is information already provided in the metadata via the field edm:dataProvider. A good example of information to use in this field is the location of a sculpture or a monument. Ideally, the latitude and longitude of the object should be submitted in the edm:Place class.

Similarly, the edm:TimeSpan Class can be used for date-related information, although this may be harder as there are not many vocabularies to choose from. Europeana has previously used Semium.org.<sup>17</sup> Another suggestion for formatting time comes from the International Organisation for Standardisation (ISO) 8601 which recommends dating things for the Georgian calendar. It would be recommended to at least normalise time formats across datasets.

<sup>11</sup> Europeana Task Force on Semantic Enrichment and Multilinguality Final Report [http://pro.europeana.eu/files/Europeana\\_Professional/EuropeanaTech/EuropeanaTech\\_taskforces/MultilingualSemanticEnrichment/Multilingual%20Semantic%20Enrichment%20report.pdf](http://pro.europeana.eu/files/Europeana_Professional/EuropeanaTech/EuropeanaTech_taskforces/MultilingualSemanticEnrichment/Multilingual%20Semantic%20Enrichment%20report.pdf)

<sup>12</sup> Getty Art and Architecture Thesaurus <http://vocab.getty.edu/aat/>

<sup>13</sup> PartagePlus Vocabulary <http://partage.vocnet.org/>

<sup>14</sup> Gemeinsame Normdatei (GND) <http://d-nb.info/gnd>

<sup>15</sup> IconClass <http://iconclass.org/>

<sup>16</sup> Virtual International Authority File (VIAF) <http://viaf.org/viaf/>

<sup>17</sup> Semium.org available on Europeana Github

[https://github.com/europeana/tools/blob/master/annocultur\\_solr4/converters/vocabularies/time/time.historical.rdf](https://github.com/europeana/tools/blob/master/annocultur_solr4/converters/vocabularies/time/time.historical.rdf)



Fig 2.4 example of non-normalised time, as displayed on the Europeana Portal May 2015

The edm:Agent Class is pertinent for gaining more information on the person that created an artwork. Europeana generates enriched data using all DBpedia artists for this class. However, this may create some false matches, for example, if the word 'Anonymous' is used in the Agent Class, DBpedia will match this to Anonymous, a band, and the Eurovision entry for Andorra 2007.<sup>18</sup>

### 3. Human and Machine Readable

By readable data, the Task Force means that the content of the metadata fields makes coherent sense not just to a human user but also to a machine. This is also something Europeana encourages in metadata submission.

In order for the fields contained within the metadata to be readable for the widest audience, Europeana already recommends that they include the correct encoding in the header so that subsequent special characters used in the data may be displayed correctly, e.g. Ὶ (Greek script), Ў (Cyrillic), æ, ô, ä (extended Latin alphabets), but this is not mandatory. Readable EDM metadata should feature a UTF-8 encoding in the header field.

```
<?xml version="1.0" encoding="UTF-8" ?>
<rdf:RDF xmlns:ore="http://www.openarchives.org/ore/terms/"
  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xmlns:rdaGr2="http://rdvocab.info/ElementsGr2/"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:skos="http://www.w3.org/2004/02/skos/core#"
  xmlns:crm="http://www.cidoc-
  crm.org/rdfs/cidoc_crm_v5.0.2_english_label.rdfs#"
  xmlns:wgs84="http://www.w3.org/2003/01/geo/wgs84_pos#"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:xalan="http://xml.apache.org/xalan"
```

<sup>18</sup> This issue will be resolved in Q2 2015



```
xmlns:dcterms="http://purl.org/dc/terms/"
xmlns:edm="http://www.europeana.eu/schemas/edm/"
xmlns:foaf="http://xmlns.com/foaf/0.1/">
```

Fig 3.1 Line one of the header field for EDM XML shows where the UTF-8 coding should be

Correct UTF-8 encoding is only one way of making the metadata readable to both human and machine users. Another is the addition of the xml-language tag. This increases an object's findability because it allows metadata fields, and therefore whole records, to be translated into the correct language for the end-user when an automatic translation tool is applied, which is possible on Europeana.

The screenshot shows the Europeana search interface. At the top left is the Europeana logo with the tagline 'think culture'. A search bar contains the text 'nightwatch' and a blue 'Search' button. To the right of the search bar is a 'Help' link. Below the search bar, the interface shows 'Matches for: nightwatch' with a red 'x' icon. There are links for 'Translate search terms' and 'Setup translation'. A 'Refine your results:' section includes a dropdown for 'Add more keywords' and a '+ Add' button. Below this are filters for 'By media type' (IMAGE (9), TEXT (3), SOUND (2)), 'By language of description', 'By year', 'By providing country', 'Can I use it?', 'By copyright', 'By provider', 'By data provider', and 'Include content contributed by users' (checked). A 'Share:' section has a 'Share' button. The main results area shows 'Results per page: 24' and 'Results 1 - 14 of 14'. The results are displayed in a grid of 14 items. Each item has a thumbnail image and a title. The titles are: 'Militia Company of District II under t...', 'De nachtwacht', 'Restaurateur aan het werk aan de 'Na...', 'Sigarenblik van "ERNST CASIMI...', 'Nightwatching', 'Nightwatch', 'Nightwatching', 'Vigiliæ; or, Night-watches, a poe...', 'Alle die so jnn und kreiß der Maure...', 'Ant aukšta kalnelia', 'Dar aušrelei neišaušus', and 'Escutcheon of Frans Banninck Cocq, ...'. Each item also has a small icon indicating its media type (e.g., image, text, sound).

fig 3.2 a search result exposing multilingual results from an English language query





## 4. Standardised

Good metadata is compliant metadata, meaning that the information supplied in the metadata fields corresponds to the appropriate categories, and that these fields also correspond to appropriate industry standards.

The Task Force believes that an outline should exist to show providers how to fill in metadata fields. This outline should include things like what is considered a meaningful title and how we can prevent repetition of titles. This information relates to how metadata is displayed in the Europeana portal and API. What works for the provider in their own domain or metadata system may not be suitable for the Europeana portal. Partners should not edit out domain-specific metadata for EDM; the Europeana Aggregation team will attempt to map all fields if possible.

### 4.1 Model Compliant Metadata

Metadata usually conforms to a standard e.g. Dublin Core, EAD, MARC 21, METS/MODS, LIDO, EDM, or a set of criteria established by an institution e.g. which fields are available and filled in on cataloguing software such as CALM or Adlib. For the purposes of the Task Force, the model to be complied with is the Europeana Data Model.

To be compliant with EDM, all mandatory elements must be filled in correctly. It should be noted that completion of mandatory fields does not eliminate the risk of bad metadata, as fields are often filled automatically with generic values which might not be correct, and can easily go unnoticed. However, mandatory elements are in place to ensure that a certain standard is met, and that providers consider their metadata before submitting it to Europeana or elsewhere.

### 4.2 Use of controlled vocabularies

To make collections readable, the Task Force recommends using open vocabularies such as Iconclass and the Getty Art and Architecture Thesaurus and adding the pertinent URI to the metadata. This is outlined in section 1.4 as the use of vocabularies can make digital cultural heritage objects more findable as well as more readable.

Another standard which could be followed is ISAAR (CPF).<sup>19</sup> Its naming conventions should be followed where possible to standardise the names of people, corporations and families in submitted metadata.

---

<sup>19</sup> ISAAR (CPF) 2nd edition <http://www.icacds.org.uk/eng/ISAAR%28CPF%292ed.pdf>



## 5. Meaningful to users

Many things contribute to good metadata quality, but probably the most obvious is whether or not the metadata quality is meaningful to the person looking at it. This means that metadata has to do slightly more than simply describe the object, it has to make the record and object make sense in context. In the example below, the metadata does describe the object technically, but has little meaning to a person reading it as it gives no descriptive information that can be seen on the page.

**Title pages**

**Alternative Title:**  
Title pages

**Creator:**  
[Neophilologica, Ed.](#)

**Date:**  
22 ENG/2013

**Type:**  
[periodical](#)

**Identifier:**  
<http://www.ceeol.com/aspx/getdocument.aspx?logid=5&id=750BC107-A023-469F-B6F0-3C46C53C0921>

**Is part of:**  
<http://data.theeuropeanlibrary.org/Collection/a1143>

**Language:**  
Multilingue (mul)

**Publisher:**  
Neophilologica

**Data provider:**  
[Central and Eastern European Online Library](#)

**Provider:**  
[The European Library](#)

**Providing country:**  
Europe

View item at  
[Central and Eastern European Online Library](#)

Share

Cite on Wikipedia

Translate details

Select language

Powered by Microsoft® Translator

Fig 5.1 example of not so meaningful metadata, despite having EDM mandatory elements

### 5.1 - Rich Descriptive Content

Meaningful metadata means that people can identify the object they are viewing, learn more about it or seek further information from the data provider.

Often, two or more EDM fields display the same information, for example, the object represented is a jug and EDM shows *Title: Jug*, and *Description: Jug*. Sometimes there is simply no more information than this available.

A good metadata record is one containing findable information and information that is meaningful to people. In our jug example, better metadata might show *Title: Brown Roman jug found in Ostia*, *Description: Brown jug with etchings of Emperor Augustus being crowned*. The descriptive metadata does not have to be lengthy, indeed, it should be succinct enough to be findable when people enter their search term(s) into Europeana.





© Rights Reserved - Free Access

View item at  
[Ateneumin taidemuseo](#)

Share

Cite on Wikipedia

Translate details

Select language

Powered by Microsoft Translator

## Harakoita kuolleen koppelon ympärillä

### Alternative Title:

Skator omkring en död tjäderhöna; Magpies round a Dead Woodgrouse; Elstern um ein totes Hasselhuhn

### Description:

Inscription: oik.alh.: Ferd.v.Wright.1867.

### Creator:

[Von Wright, Ferdinand, Ferdinand Von Wright](#) (Von Wright, Ferdinand; Ferdinand Von Wright) To: 1906+02:00 Date of birth: 1822+02:00 Date of death: 1906+02:00

### Date of creation:

1867

### Type:

[Physical Object](#); [maaläus](#)

### Format:

66 x 86 cm; öljy kankaalle

### Subject:

[wood](#); [ELÄIN](#); [lintu](#); [harakat](#); [kuollut koppelo](#); [talvi](#)

### Identifier:

A I 62

### Source:

[Ateneumin taidemuseo](#)

### Data provider:

[Ateneumin taidemuseo](#)

### Provider:

[National Library of Finland](#)

### Providing country:

Finland

Fig.5.2 Example of succinct discoverable metadata from Ateneumin Taidemuseo for the National Library of Finland<sup>20</sup>

The records shown above and below show how good quality metadata serves as an access point to further information on the data partner's own portal. The record above displays enough metadata to be discoverable, and a clear image so that it is visible. Also, the search was conducted in English but multilingual elements within the metadata return a result in Finnish.

<sup>20</sup>Von Wright, Ferdinand *Harakoita kuolleen koppelon ympärillä* Ateneumin Taidemuseo for the National Library of Finland [http://www.europeana.eu/portal/record/2021002/e\\_IsShownAt\\_F5901F89\\_01D7\\_4C11\\_9E35\\_9142C96C605A.html](http://www.europeana.eu/portal/record/2021002/e_IsShownAt_F5901F89_01D7_4C11_9E35_9142C96C605A.html)



FINNISH  
NATIONAL  
GALLERY



Photographer: Finnish National Gallery / Pakumies, Hannu  
Copyright and related rights © Artists, photographer, other copyright holders, the Visual Artists' Copyright Society Kuvasto and the Finnish National Gallery.  
Pictures of the works of art can be ordered for publication use from the Photographic Service.



ART COLLECTIONS • ARTWORKS ARTISTS ATENEUM KIASMA SINEBRYCHOFF

Wright, Ferdinand von  
Magpies round a Dead Woodgrouse  
1867

#### Keywords

ELÄIN (21), KOHTAUS (16), harakat (1), metsä (1), kuollut koppelo (1), lintu (1)

#### Exhibition information

Taiteilijajäikeet von Wright, Ateneum 5.3.-3.5.1982 - Kuopion taidemuseo 21.5.-8.8.1982 - Tampereen taidemuseo 18.8.-3.10.1982 - Turun taidemuseo 10.10.-28.11.1982

Valjekeet von Wright / Bröderna von Wright, Suomen Taidesäätiön (Järjestyksellisesti) 30/1971 (16.4.1971-28.1.1973, 28 palkkumallia)

Taitehalli 1930 ja 1.6.-4.8.1. Suomalaisista taidetta, Valkeana Asemen Taidesäätiön / Finns Konst, Ett unval ur Konstnärernas i Ateneum / Finnsche Kunst, Ein Auswahl aus den Sammlungen im Ateneum.

von Wright, Taiteilijajäikeet / Konstnärbröderna, 31.5.-27.8.2000, Taidesäätö Retretti, Punkaharju / Konstnärret Retretti, Punkaharju.

Konstnärslits och Finns Konstnärernas exposition i Waas 1894 / Nya Föreläsningar i Suomen Taidesäätiön näyttely Väasassa 1894 uuden kansakoulun huoneustissa.

Finns Konstnärernas Exposition [7 maj-17 jun] 1877. Konstnärernas lokal, Unionsgatan 20, 2 trappor upp.

Eläinten aika 16.2.-27.5.2007 / djuren tecken, 16 februari - 27 maj 2007 / Age of the Animal 16 Feb - 27 May 2007, Ateneum.

Suomen Taidesäätiön ensimmäinen kertomus (Järjestyksellisesti) 1936, Suomalaisista maalaustaidetta vuoteen 1900 / Finns Konstnärernas första vandringsutställning 1936, Finns måleri till år 1900.

#### ATENEUM

Öljy kankaalle  
Korkeus 86,00 cm  
Leveys 86,00 cm  
Tähtäinmerkintä 1876, A I 62  
Hörsing  
Victor Hovings testamenttikoelma

Fig 5.3 The same record on the data partner's web service contains additional metadata on the object <sup>21</sup>

## 5.2 - Curated information

- 8 The Task Force encourages the curation of datasets, and recommends that smaller curated datasets with greater metadata quality are submitted and should be given priority in ingestion and display via Europeana

Smaller, curated collections are more likely to showcase the cultural heritage institution's holdings, be more discoverable, more shareable, and more likely to achieve end results for the submitting institution. For the future of digital collections in cultural heritage, it is essential to adopt the attitude of curation and quality over quantity of material digitised.<sup>22</sup>

The Task Force recommends and encourages the submission of curated datasets which contain fewer objects but richer descriptive and technical metadata than we are currently seeing submitted. We also recommend that Europeana's Aggregation team works more closely with its aggregators and data providers to encourage this. It would also be beneficial for Europeana Network partners to instigate a community effort to enable and encourage the submission of better quality metadata.

## 5.3 Information in Context

Clearly, large-scale collections can showcase an institution's holdings very well. The problem is that when they are exported to become part of a larger repository like

<sup>21</sup> Von Wright, Ferdinand *Harakoita kuolleen koppelon ympärillä* Finnish National Gallery  
[http://kokoelmat.fng.fi/app?lang=en&si=http%3A%2F%2Fwww.muusa.net%2Fteos\\_F5901F89-01D7-4C11-9E35-9142C96C605A](http://kokoelmat.fng.fi/app?lang=en&si=http%3A%2F%2Fwww.muusa.net%2Fteos_F5901F89-01D7-4C11-9E35-9142C96C605A)

<sup>22</sup> This has been more thoroughly discussed in the following article. Karen M. Wickett, Antoine Isaac, Katrina Fenlon, Martin Doerr, Carlo Meghin, Carole L. Palmer, and Jacob Jett (2013). Modeling Cultural Collections for Digital Aggregation and Exchange Environments. CIRSS Technical Report 201310-1, University of Illinois at Urbana-Champaign.  
<http://hdl.handle.net/2142/45860>



Europeana, they may lose the context they require for discovery and dissemination, unless the metadata provided is of high quality.

We recommend that context is added to curated collections. Context can be expressed in multiple ways within a dataset e.g. in terms of objects, places, subject.<sup>23</sup> If item records have contextual information, we should see a better level of data quality both technically and descriptively across the entire dataset.

Encouraging institutions to think about context gives them the opportunity to consider the information contained in their metadata before exporting records en-masse to a larger repository. Publishing all, or a majority of records, is an option, but it is better suited to an institution's own web service, rather than an aggregator or data repository. There can be both quantity and quality, and many institutions have achieved this and published such data on Europeana. However the Task Force has noted a correlation between larger datasets and lower data quality.

Europeana as a repository would like to receive more curated datasets with connected collection information so that the context is maintained as it was created and intended by the institution.

## **6. Clear on re-use**

Another important aspect of metadata is that it should provide clarity on what people can do with the digital representation of the object they have found. This is why the edm:rights field is essential. To complete this field, the data provider chooses one of 13 rights statements that apply to the object.<sup>24</sup>

The Task Force noted that many data providers approach rights statements as an afterthought and lack sufficient know-how to apply the appropriate statement. They therefore choose a restrictive rights statement as a default. The Task Force finds that the benefits of open culture need to be advocated more, perhaps by data providers receiving training from aggregators in this area.

---

<sup>23</sup> Gradmann, Stefan 2010 *Knowledge = Information in Context: on the Importance of Semantic Contextualisation in Europeana* [http://pro.europeana.eu/files/Europeana\\_Professional/Publications/Europeana%20White%20Paper%201.pdf](http://pro.europeana.eu/files/Europeana_Professional/Publications/Europeana%20White%20Paper%201.pdf)

<sup>24</sup> Europeana Available Rights Statements <http://pro.europeana.eu/web/guest/available-rights-statements>



## 7. Visible

The digitisation of cultural heritage is so exciting because it means that objects housed in memory institutions become accessible to people and communities everywhere. It connects people and objects. But for that object to be viewed, it must have a good digital rendering.



*Fig 7.1 Actual size of object image submitted to Europeana.eu*

The image above was submitted as the main digital representation of the object. The image is too small to be viewed without magnification, and the quality when this object is expanded is too poor to be of use. What would be useful to audiences, particularly people hoping to use digitised cultural heritage for personal or academic research, is a high resolution image depicting the object clearly, with appropriate accompanying metadata.

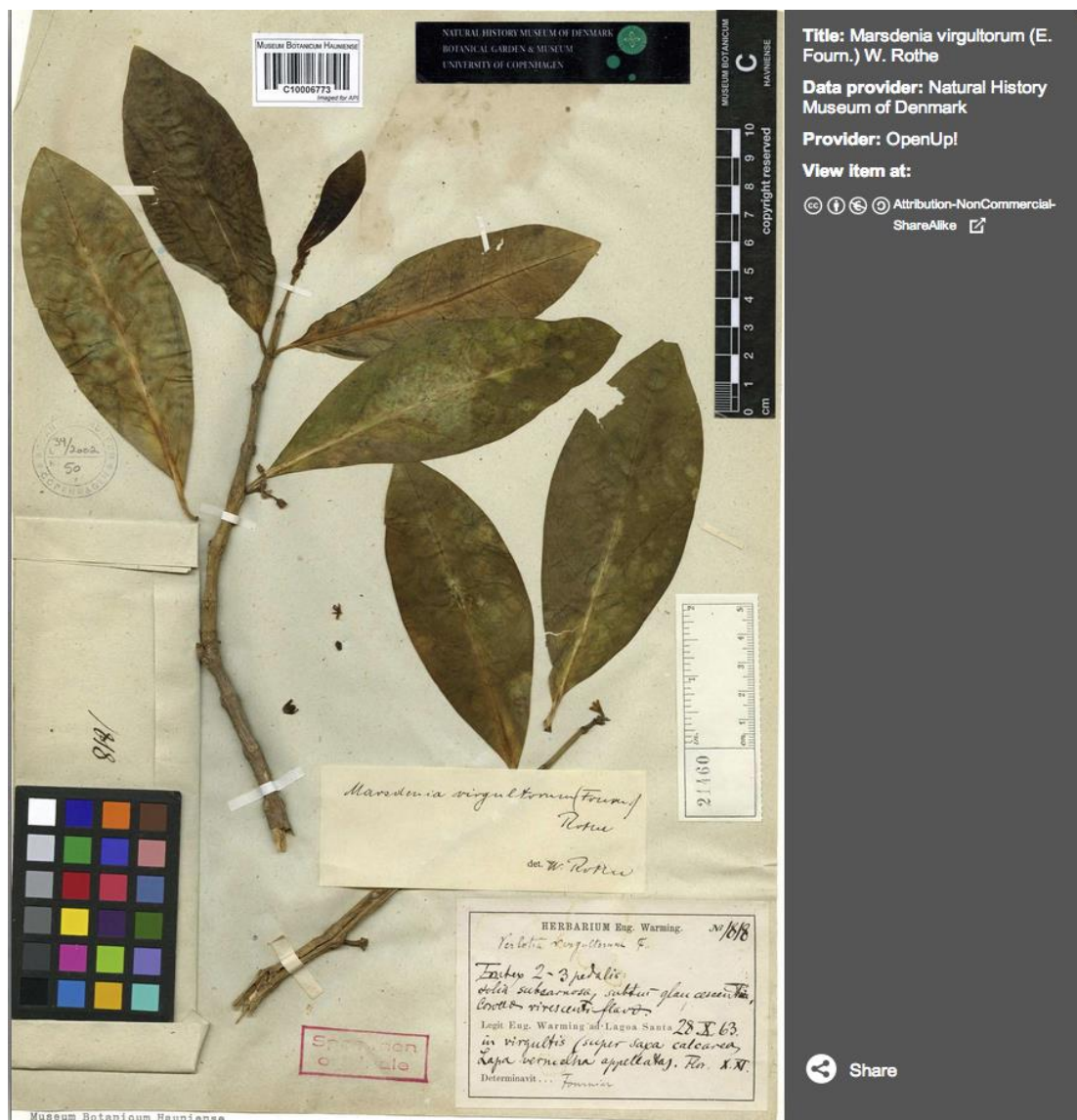
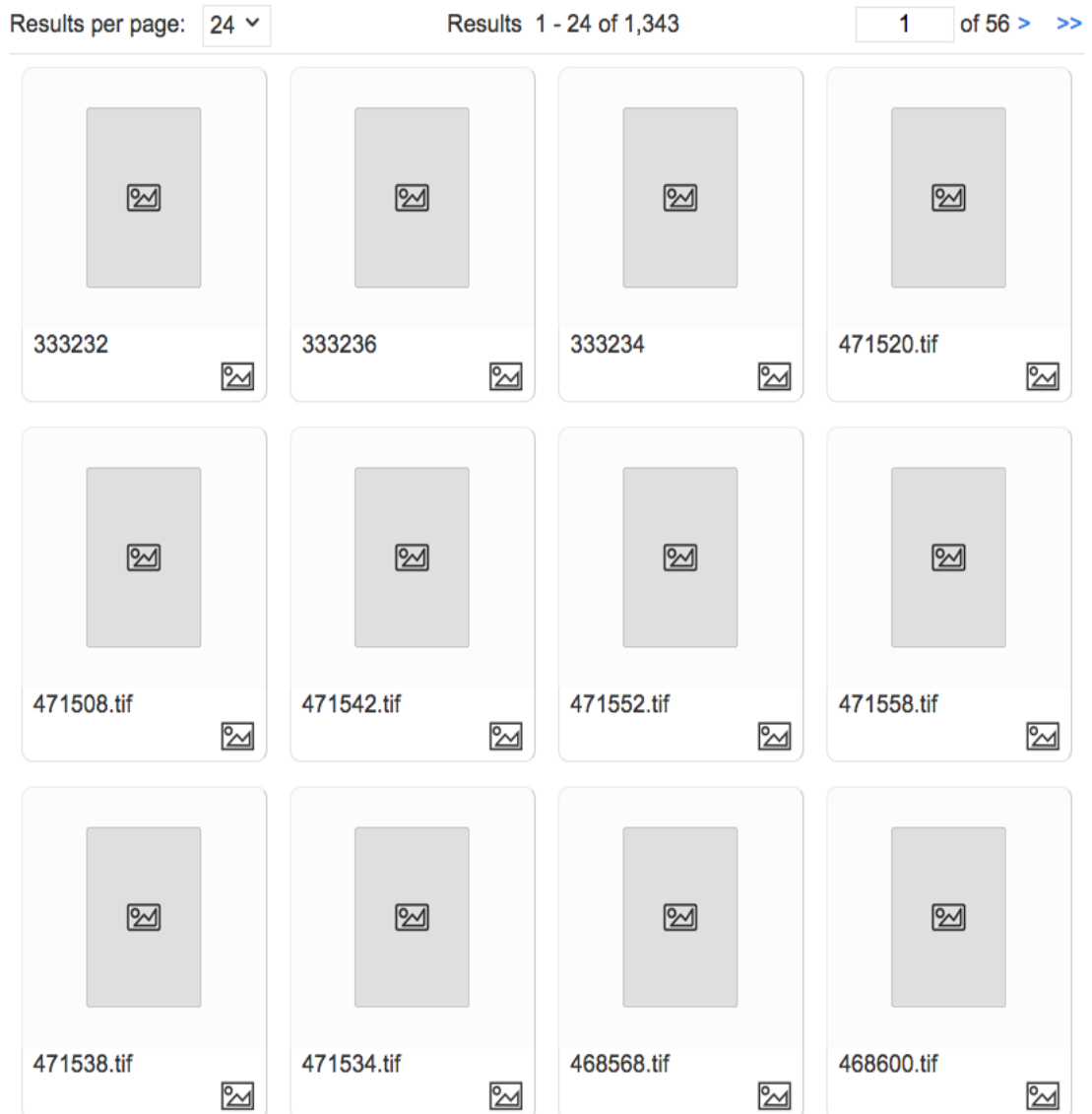


Fig 7.2 High resolution image submitted by the Natural History Museum of Denmark

In this example, a high resolution image allows for the digital cultural heritage object to be viewed clearly in context of the collection. A thumbnail image can be generated from an image this size. Thumbnails (or previews) make it easier for records to be identified and accessed from a set of search results.

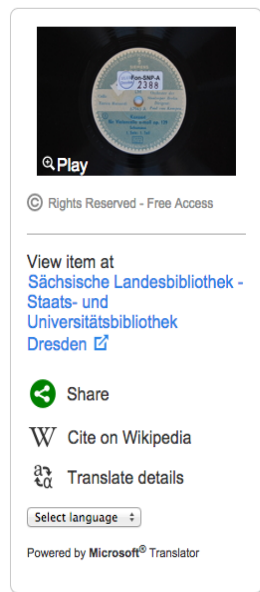


*Fig.7.3 A lack of preview images and metadata makes cultural heritage objects invisible*

Each record should include an image provided in the edm:object field. This makes the digital cultural heritage object more visible in search results and provides essential information to people about what they can expect from the digital cultural heritage object. This applies to all media types including sound and video as well as images and text.

- 9 A thumbnail image should be made available for all digital cultural heritage objects.





## Konzerte, Vc Orch / op. 129 / Auswahl

### Description:

1. Satz, 1. Teil ; 1. Satz, 2. Teil., Konzert für Violoncello a-moll op. 129, Konzerte, Vc Orch / op. 129 / Auswahl

### Creator:

Schumann, Robert; Mainardi, Enrico; Kempen, Paul van

### Contributor:

Schumann, Robert; Mainardi, Enrico; Kempen, Paul van

### Coverage:

Dresden; Dresden(Drážd'any, Dresde, Drezden, Drezdzany, Drezno, Dreßden, Haupt- und Residenzstadt Dresden, Königliche Haupt- und Residenzstadt Dresden, Landeshauptstadt Dresden, Stadt Dresden, State Capital Dresden); Same as:

- (DE-588b)37172-5
- (DE-588b)5556921-3
- (DE-588)37172-5
- (DE-588c)4012995-0
- 4012995-0
- <http://sws.geonames.org/2935022>

### Geographic coverage:

Dresden(Drážd'any, Dresde, Drezden, Drezdzany, Drezno, Dreßden, Haupt- und Residenzstadt Dresden, Königliche Haupt- und Residenzstadt Dresden, Landeshauptstadt Dresden, Stadt Dresden, State Capital Dresden); Same as:

- (DE-588b)37172-5
- (DE-588b)5556921-3
- (DE-588)37172-5
- (DE-588c)4012995-0
- 4012995-0
- <http://sws.geonames.org/2935022>

Latitude: 51.05089; Longitude: 13.73832

### Subject:

Schellackplatte; Konzert

### Identifier:

<http://mediathek.slub-dresden.de/ton70900205.html>

Fig. 7.4 Example of a sound recording and accompanying thumbnail image via the Deutsche Digitale Bibliothek

The Task Force recommends making the availability of an image mandatory for all digital cultural heritage objects. This, in addition to correct metadata such as good descriptive elements in all mandatory fields and persistent working links, will aid visibility and accessibility. The quality of the digital cultural heritage object itself should also be as high as possible. The Task Force acknowledges that the Task Force on the Content Re-Use Framework will further discuss this issue.<sup>25</sup>

<sup>25</sup> Documentation of the Extension of the Europeana Licensing Framework  
[http://pro.europeana.eu/files/Europeana\\_Professional/Projects/Project\\_list/Europeana\\_Creative/Deliverables/eCreative\\_D3.3\\_KL\\_v1.0.pdf](http://pro.europeana.eu/files/Europeana_Professional/Projects/Project_list/Europeana_Creative/Deliverables/eCreative_D3.3_KL_v1.0.pdf)



## How Metadata Quality is assessed in Europeanana

Members of the Task Force, and data partners contributing to the Europeanana Aggregators' Forum, raised issues of feeling excluded from Europeanana. They felt that they did not understand what happened to their data once it left their own repositories.

Below, we demonstrate how the Europeanana Aggregation team assesses metadata quality within submitted datasets. This way, we hope to frame the discussion of metadata quality in the context of the Europeanana repository and its value for audiences, and also take into account the various issues that data providers encounter in terms of data creation and export.

Europeanana ingests metadata submitted by over 3,000 cultural heritage institutions across Europe. Data providers representing individual cultural heritage institutions submit metadata through one of three channels: via a national or domain aggregator (1), a project (2), or directly to Europeanana (3). This ensures that the maximum number of data providers can contribute to Europeanana.

The issue of metadata quality regulation is a constant and pressing one, with both aggregating partners and Europeanana conducting regular checks. However, due to the volume of records submitted every month, these checks can only be applied to a sample of the records.

### 2.1 Mandatory metadata elements

The quality of metadata submitted to Europeanana varies widely between institutions and domains, as each provider must adapt their own standards to the Europeanana Data Model. There are nine mandatory elements in EDM (and an additional one for user-generated content),<sup>26</sup> including information on re-use potential.

Applicable class	Mandatory Properties (or alternatives)
Aggregation	edm:dataProvider
Aggregation	edm:isShownAt or edm:isShownBy
Aggregation	edm:provider
Aggregation	edm:rights
Aggregation	edm:aggregatedCHO
ProvidedCHO	dc:title or dc:description
ProvidedCHO	dc:language for text objects
ProvidedCHO	dc:subject or dc:type or dc:coverage or dcterms:spatial
ProvidedCHO	edm:type
ProvidedCHO	edm:ugc (when applicable)

Fig. 2.1.1 mandatory elements of EDM

<sup>26</sup> Further information on mandatory and recommended elements can be found in the EDM documentation <http://pro.europeanana.eu/share-your-data/data-guidelines/edm-documentation>






These mandatory elements were implemented as a way of achieving a minimum standard of quality across the diverse institutions represented in the Europeana repository. In practice, the mandatory elements are sometimes misinterpreted, accidentally misused or contradictory.


## Description: description


Fig 2.1.2 A frequent metadata submission to Europeana




© Rights Reserved - Free Access


View item at  
[Fitzwilliam Museum](#)

 Share

 Cite on Wikipedia

 Translate details

Select language ▾

Powered by  Microsoft® Translator

### unknown, unknown

**Description:**  
; Classification(s): unknown, unknown; Acquisition: given by British School of Archaeology, 1913 [E.152.1913]  
Description: circular slab, with incised decoration on both sides, table top ?

**Creator:**  
[The Fitzwilliam Museum, Cambridge, UK](#)

**Type:**  
[Image](#)

**Format:**  
text/html

**Subject:**  
[unknown](#)

**Identifier:**  
<http://www.fitzmuseum.cam.ac.uk/opacdirect/51641.html>

**Is part of:**  
Fitzwilliam Museum

**Language:**  
en-GB


**Publisher:**  
The Fitzwilliam Museum, Cambridge, UK

**Data provider:**  
[Fitzwilliam Museum](#)

**Provider:**  
[CultureGrid](#)

Fig 2.1.3 The use of 'unknown' as a default element








[View](#)

© Rights Reserved - Free Access


View item at  
[FylkesFOTOarkivet i Møre og Romsdal](#)

 Share

 Cite on Wikipedia

 Translate details

Select language ▾

Powered by  Microsoft® Translator

## Portrett 1904-1955

**Description:**  
(Tekst manglar)

**Creator:**  
[Storegjerde, Olaf Andreas Knutsen](#)

**Time period:**  
1904 - 1955

**Date of creation:**  
1904 - 1955

**Type:**  
[Stillimage](#)

**Format:**  
12 x 16; 222 - Negativ s/k tørrpl.

**Subject:**  
[Portrett](#)

**Identifier:**  
FAKf-100100.113352

**Is part of:**  
FAKf-100100

**Provenance:**  
Sande Sogelag

**Data provider:**  
[FylkesFOTOarkivet i Møre og Romsdal](#)

**Provider:**  
[Norsk Kulturråd](#)

Fig 2.1.4 shows the use of a default description 'tekst manglar' meaning 'text missing' in Norwegian

The Task Force recommends that the minimum quality standard be raised, with richer data provided in the mandatory elements, to ensure that the digital cultural heritage objects submitted by partners can be discovered. The Task Force suggests that data providers and aggregators look at their internal digitisation policies and review their metadata using the ideas and techniques outlined in section one of this report, before submitting to a further digital repository. The Task Force recommends greater use of the Europeana Publication Policy to communicate the processes and requirements needed in relation to metadata in Europeana.

## 2.2 Metadata quality checks by the Europeana Aggregation team

To date in 2015, 1,809 datasets (April Publication 2015) have been processed by the Europeana Aggregation team. The ingestion procedure is undertaken by a team of four Operations Officers who manually process all datasets submitted by data providers to Europeana. There is usually a period of two weeks during the month during which EDM datasets are harvested, assessed, edited and enriched to be published on the Europeana portal and API.

The Aggregation team primarily relies on three main tools to assess metadata quality, the REPOX - for harvesting from partners' OAI-PMH servers, Unified Ingestion Manager (UIM) - an ingestion management tool governing import, dereferencing and



enrichments of metadata, and MINT - for quality assessment, mapping, editing, transformation and validation.

Prior to the ingestion of datasets, the Aggregation team uses a customer relationship management tool called Sugar CRM to assign a unique provider number and subsequent running dataset number. This makes it easier for all parties to track and process the metadata in Europeana as the aggregator and data provider can use their assigned number to check their datasets after publication, but this dataset ID is not used for checking metadata quality.

Europeana's Aggregation team have a standard process for conducting sample checks. By outlining this process below we hope to make the process transparent. The steps involved are:

1. **Check a sample of the dataset in raw EDM XML form.** This is the step which will highlight any missing mandatory elements. The way XML is displayed in its raw form makes it easy for the Aggregation team to locate possible holes or missing mandatory elements. It also makes it possible to identify elements to investigate later such as where fields might be populated with default values, e.g. the submitted title may be simply 'Title', or may use incorrect characters. Fields flagged at this point are checked in stage 2.
2. **Check dataset statistics.** Once sets are added to Sugar CRM, then harvested and uploaded to the MINT mapping tool, the dataset statistics are checked to see if they are as expected. The columns indicate the metadata field supplied (Xpath), the number of times that field appears in the dataset (Count), the number of times this field is unique (Distinct) and how many



characters have been entered on average in that field (Length).

Xpath	Count	Distinct	Length
<input type="checkbox"/> repoxWrap	258494		
<input checked="" type="checkbox"/> @oaipmh	258494	258494	59
<input type="checkbox"/> rdf:RDF	258494		
<input type="checkbox"/> edm:ProvidedCHO	258494		
<input checked="" type="checkbox"/> @rdf:about	258494	258494	34
dc:identifier	258494	257541	24.9
dc:title	258494	61169	30.9
dc:description	251128	11	38.2
dc:relation	258494	61169	73.9
dc:source	258494	2	8.9
dc:type	258494	3	17.9
dcterms:temporal	258494		
<input type="checkbox"/> edm:hasType	258494		
<input checked="" type="checkbox"/> @rdf:resource	258494	3	47.9
edm:type	258494	1	5
dc:subject	1773939	100356	19.2
<input type="checkbox"/> edm:WebResource	516988		
<input checked="" type="checkbox"/> @rdf:about	516988	516035	39

Fig. 2.2.1 Shows the dataset statistics as seen in Europeana's version of MINT

Given time constraints, this is the only way the Aggregation team can identify ambiguities in the metadata. They can assume that 1,000 records with just one unique title are more likely to be of lower quality than 1,000 records with 1,000 titles.

The above table shows that a set of 258,494 records has the same number of rdf:abouts (the number of unique identifiers or unique URIs representing the submitted digital cultural heritage objects within a dataset) but only 61,169 unique titles (dc:title), and only 11 unique entries for dc:description. This serves as a warning sign that the dataset will require further investigation to see if it is suitable for publication. A set with such statistics is likely to be returned to the provider to see if it could be improved, but the set is so large it is unlikely the provider will be able to put in the manual effort required to bring it up to standard, leaving the set published as is if the mandatory elements are there.

The Operations Officer also checks that any links to the digital cultural heritage objects work, including those in edm:object, edm:isShownAt and edm:isShownBy. These links are needed to ensure that the object can be



made visible to audiences (via `edm:isShownAt` and `edm:isShownBy`), and that a preview image can be created for the record (via `edm:object`). The lack of a link to a viable digital object is a common reason for a dataset to be delayed or not published.

The team also checks that WebResource URIs and URIs from vocabularies such as Getty Art and Architecture Thesaurus (AAT), VIAF, PartagePlus, GND, IconClass lead (or de-refer) to appropriate conceptual web content or data. It is necessary to check links to online vocabularies because UIM uses them to fetch data (including human-readable labels) to provide audiences with additional context.

**Place Term:**

<http://sws.geonames.org/3175395/>

**Place Label:**

[Italia] (no); [Itaalje] (fy); [Italia] (nn); [An Eadailt] (gd); [An Iodáil] (ga); [Itàlia] (oc); [Italia] (fi); [Italy] (om); [Italie] (fr); [Italia] (fo); [Итали] (os); [יִיטאַליע] (he); [Italia] (gl); [Włochy] (pl); [ਭਾਰਤ] (lo); [Italia] (ln); [Italia] (Y, Y) (vi); [ਭਾਰਤ] (dz); [Itālija] (lv); [Italia] (lt); [Litaliyān] (vo); [Italien] (de); [Италија] (uz); [IT, Italia, Italian Republic, Italie, Italien] (def); [Италија] (mk); [इटाली] (ml); [Итали] (mn); [Италија] (uk); [Italia] (eu); [इटाली] (mr); [Italiya] (mt); [Itali] (ms); [إيطاليا] (fa); [إيطاليا] (ur); [Itària] (ty); [Italy] (na); [Ιταλία] (el); [Italia] (nb); [इटाली] (ne); [Italo, Italujo] (eo); [Italian Republic, Italy, Republic of Italy] (en); [Itaalia] (et); [Italia, República Italiana] (es); [Italië] (nl); [Itali] (to); [Itàlia] (ca); [Italya] (tl); [Italya] (tr); [Итолиë] (tg); [ਭਾਰਤ] (te); [Italiya] (bs); [Italia] (br); [ປະເທດອິຕາລີ, ອິຕາລີ] (th); [इटाली] (bn); [ਭਾਰਤ] (pa); [ਭਾਰਤ] (bo); [இத்தாலி] (ta); [იტალია] (ka); [Italien] (sv); [Италија] (bg); [Italia, Uitaliani] (sw); [Италија] (be); [Italiya] (sl); [Itali] (kw); [Taliensko] (sk); [Italien] (da); [Talyaani, Talyaaniga] (so); [Italya] (ku); [Италија] (sr); [Itali, Italia] (sq); [Italiya] (ko); [Itàlia] (sc); [Yr Eidal] (cy); [Itàlia] (se); [Italiya] (sh); [Итали] (cv); [Итали] (cu); [ਭਾਰਤ] (km); [ਭਾਰਤ] (kn); [Italie] (cs); [Italië] (li); [Italia] (co); [इटाली] (sa); [Italia] (la); [Italien] (lb); [Италија] (ru); [Italiya] (hr); [意大利] (zh); [Italia] (ro); [Italia] (rm); [Itali] (ht); [Olaszország] (hu); [इटाली] (hi); [Itali, Italia] (id); [Italia] (ia); [Italya] (qu); [Италия] (hy); [Italiya] (az); [Itàlia] (is); [Italia, Repubblica Italiana, il Bel Paese] (it); [ਭਾਰਤ] (ii); [إيطاليا, إيطاليا] (ar); [Italia] (io); [Ἰταλῖν, Ἰταλῖν] (am); [Italia] (an); [Itàlia] (pt); [Italy] (aa); [イタリア, イタリア共和国] (ja); [إيطاليا, إيطاليا] (ps); [ਭਾਰਤ] (yi); [Italië] (af)

**Geo Space:**

null; null

Part of: [Arrondissement Antwerpen](#)

Part of: [Florence](#)

**What**

**Concept Term:**

<http://vocab.getty.edu/aat/300014109>

**Concept Term:**

<http://vocab.getty.edu/aat/300041273>

**Concept Label:**

[druckgraphischen Blättern] (de); [gravures (général, œuvres visuelles)] (fr); [prints (visual works)] (en); [estampas] (es); [prenten] (nl)

Fig 2.2.2 A record with dereferenced online vocabulary URIs

### 3. Check using MINT.

The MINT mapping tool is used to check the type of information entered in various elements. Members of the Aggregation team are limited by the number of languages and alphabets known to them to decipher fields such as `dc:title` or `dc:description`. There is a translation tool on the Europeana portal but this is not available to the Aggregation team until after publication.



At this stage, the Aggregation team can make decisions about the quality of a dataset and ask for clarification from the data providers via their aggregators prior to transforming the dataset into publication-ready EDM.

The Task Force recommends that part of this activity be taken up by the aggregator prior to submission.

#### 4. Check EDM mapping.

After the sets have been mapped, they are transformed and then validated against the EDM schema. Checks show if a mandatory element such as `dc:type`, `dc:subject`, `dcterms:spatial` or `dcterms:temporal` is missing, or if there is an inappropriate number of instances of `edm:isShownAt` or `edm:isShownBy`. The system flags invalid records, showing which elements are missing, and creates a file to gather all invalid records together.

fig. 2.2.5 shows invalid elements within a dataset in Europeana's version of the MINT tool. Records which do not have valid EDM elements appear as invalid and are removed from the set. These records are not published but sent back to the relevant projects/aggregators to be forwarded to the data provider to be fixed. Once fixed, they can be resubmitted in a subsequent publication cycle.

#### 5. Check for duplicate records.

The valid EDM records are returned to the UIM where any duplicate records are discarded from the datasets. This only occurs when the identifiers (`rdf:about` of the `edm:ProvidedCHO`) are duplicates or near duplicates in the dataset. Unfortunately, it does not apply duplication checks to all metadata





fields and so cannot identify multiple records containing the same titles or descriptions. Records may have the same title but be of different objects such as 'Postkarte aus Wien' or have different identifiers. The tools used by Europeana cannot filter for this to flag it as a problem.



*Fig 2.2.6 Different objects with different identifiers showing near duplicate information*

Datasets which do not meet the minimum mandatory requirements outlined by EDM are immediately rejected for publication and the appropriate aggregator or project is contacted. This is so the data provider can make the necessary changes. These issues are usually resolved relatively quickly because of external factors (e.g. project deadlines, funding) which require the datasets to be published on Europeana.

While datasets may meet the requirements for EDM by having information in the mandatory fields, Europeana does not make any stipulations about the quality of that information because it is difficult to do so objectively. As there are no such quality stipulations, and because it is difficult to assess and compare metadata quality across sets, providers may never check their metadata quality, nor how it displays on the Europeana portal, and they may not realise how the metadata they create affects the experience audiences have with their objects. The Task Force recommends that the



Europeanana Aggregation team be more available to the data partners, so that these issues can be resolved earlier in the publication process.

10 The Task Force recommends that the Europeanana Aggregation team be more available to the data partners

### 2.3 The Role of Partners in Metadata Quality Checks

Europeanana assumes that partners submitting datasets are conducting data quality checks of their own prior to considering them suitable for publication. Ideally these checks would include EDM schema validation or other technical implementation such as checking for working links and file size of images. Perhaps more realistically, it is presumed that partners are checking that their links work, that they are persistent and provide redirects when their servers change, that their digitised media are of sufficient resolution and that their descriptions, if provided, make sense.

These checks can be done on the same sample basis that Europeanana uses, and do not require a vast technical knowledge or multiple tools. Some partners already use the MINT tool and so could easily perform the same statistics check that the Aggregation team does prior to mapping.

The partners who conduct these checks are the projects and aggregators that are less likely to have issues with their datasets.

Ultimately, metadata quality is better when the Aggregation team is able to work with projects and aggregators before the datasets are submitted. The later issues are spotted in the ingestion/project process, the more difficult it is to encourage providers to change or provide better quality metadata.

The Task Force notes that metadata quality benefits when Europeanana can enter into a clear and consistent dialogue with the data providers as early as possible. By the time a dataset reaches the point for export to a larger repository, it is too late to improve its quality, particularly with medium to large collections (anything over 300 digital cultural heritage objects). The work and resources required to increase the quality significantly will often be deemed as too great.

The lack of early dialogue between data partners, aggregators and Europeanana rules out the chance of conducting any metadata control at a later stage. It also increases the likelihood of adding false or inconsistent information through normalisation of information in datasets e.g. formalisation of spelling of provider names. As such, there is an issue with balancing what data providers can provide under their current budgets with what audiences might require in order to discover the submitted digital cultural heritage objects.





The Task Force recommends that data is exported and checked as early as possible and that an EDM content checker (using a preview portal as a means of data validation) be implemented.

As multiple metadata issues are reported to providers about their (many) datasets on a monthly basis, it is clear that the level of metadata quality checks undertaken by aggregators and projects should be increased overall.

11 Data should be exported and checked as early as possible and a Europeana Data Model (EDM) content checker should be implemented (using a preview portal as a means of data validation). Also that increased metadata checks should be taken up by the aggregator prior to submission.



## Blockers to good metadata quality submission

Metadata is what makes objects discoverable. High quality metadata is required in order for material to be found, disseminated and re-used. Some of the problems caused by poor metadata are outlined in section one. What is stopping data partners providing better quality metadata? The Task Force identified several blockers:

### 3.1 Context of Metadata Creation

Cultural heritage objects may be digitised without a plan for online exposure, and the metadata may be created without future thought to digital use outside an institution's reading room. Members of the Task Force noted that metadata creation is often carried out by a mix of professional cataloguers, librarians, archivists and interns, all of whom may have different concepts of how the metadata will be used in the future. In short, at the time of metadata creation there may be a lack of understanding regarding its future use.

### 3.2 Metadata as a by-product of the preservation process

Metadata is often regarded as a by-product of digital preservation, rather than an end in itself. This means that little attention is paid to its quality. Metadata which comes from the export of an older database may not have all the elements required to produce valid EDM records, but will nevertheless have metadata in an analogue form that suits the needs of the institution.

The older the collection is, the more likely it is that it does not have all of the elements required for digital representation. Additional metadata may have been created at some stage but not added to the record entry submitted to Europeana. Older records may be improved during the digitisation/export process but it is also possible that while more metadata fields become available in new digital records, they lack the richness of older records, as the data creators may not have the time and resources to provide richer data, particularly in larger scale digitisation projects.

Metadata which is automatically generated in digitisation projects may be useful but may not meet the requirements of audiences and researchers. Metadata fields such as dc:title or dc:description may provide more information and greater context, but cannot be standardised as they are free text fields and will depend on the type of object discussed, the information available and the type of record being created e.g. item level records in an archival finding aid such as in Fig 2.3 will have fewer details than an individual item from a gallery.

When the emphasis in an organisation is solely on digital preservation, metadata is created but it is not the focus and so the quality is different and may not be suitable for online publishing. For example, when digitising large-scale collections, the cataloguer will have to include certain mandatory elements to satisfy their own digitisation and organisational needs, but may not have the resources available to provide richer



unique descriptions or titles and other accompanying information. Common examples of this in Europeana include 2D renderings of 3D objects given the name of the object as their title e.g. “Glass” or “Vase” or “Photograph”. Here, the criteria for a title has been satisfied but with little accompanying metadata, which restricts the findability of the object. This is also an example of little to no curation taking place in terms of which collections are most valuable to be shared through Europeana.

### **3.3 Lack of Digital Appraisal**

Viewing metadata as only a by-product of the digital preservation process may be a mistake. With this mind-set, there is no appraisal of the standard of the digital material, either prior to its digitisation or afterwards. However, data providers are submitting high volumes of digitised material, often at the expense of metadata quality.

Metadata creation must be perceived as an essential part of the digitisation process and policy-makers need to be made aware of its importance.

Task Force members expressed concern about the legacy issues that come from high volume digitisation, and that funding for resources is favoured for mass digitisation over metadata creation, leading to higher quantity that does not always translate into higher quality. It seems that the emphasis is on getting as much material as possible digitised, rather than appraising collections and putting greater time and detail into those that will yield traffic and increase an institution's profile. In an analogue archive or museum environment, the appraisal process may take a long time, depending on the size of the collection. This practice does not appear to have translated into the digitisation process, which affects the quality of metadata created.

### **3.4 Limited resources**

Financial resources are one of the core reasons for poor metadata output cited by members of the Task Force and the data providers they represent. Limited financial capital to fund the digitisation process, in terms of technology, equipment and staff, severely limits the output of cultural heritage institutions. This is particularly true for smaller institutions who have to prioritise running costs over digitising material.

This clearly limits what institutions can achieve and what will be digitised. However, this does not have to be a disadvantage where metadata quality is concerned. It has been noted that institutions with smaller digitised collections produce higher quality metadata than those with larger digital collections, particularly those that are exported on a large scale from cataloguing software.

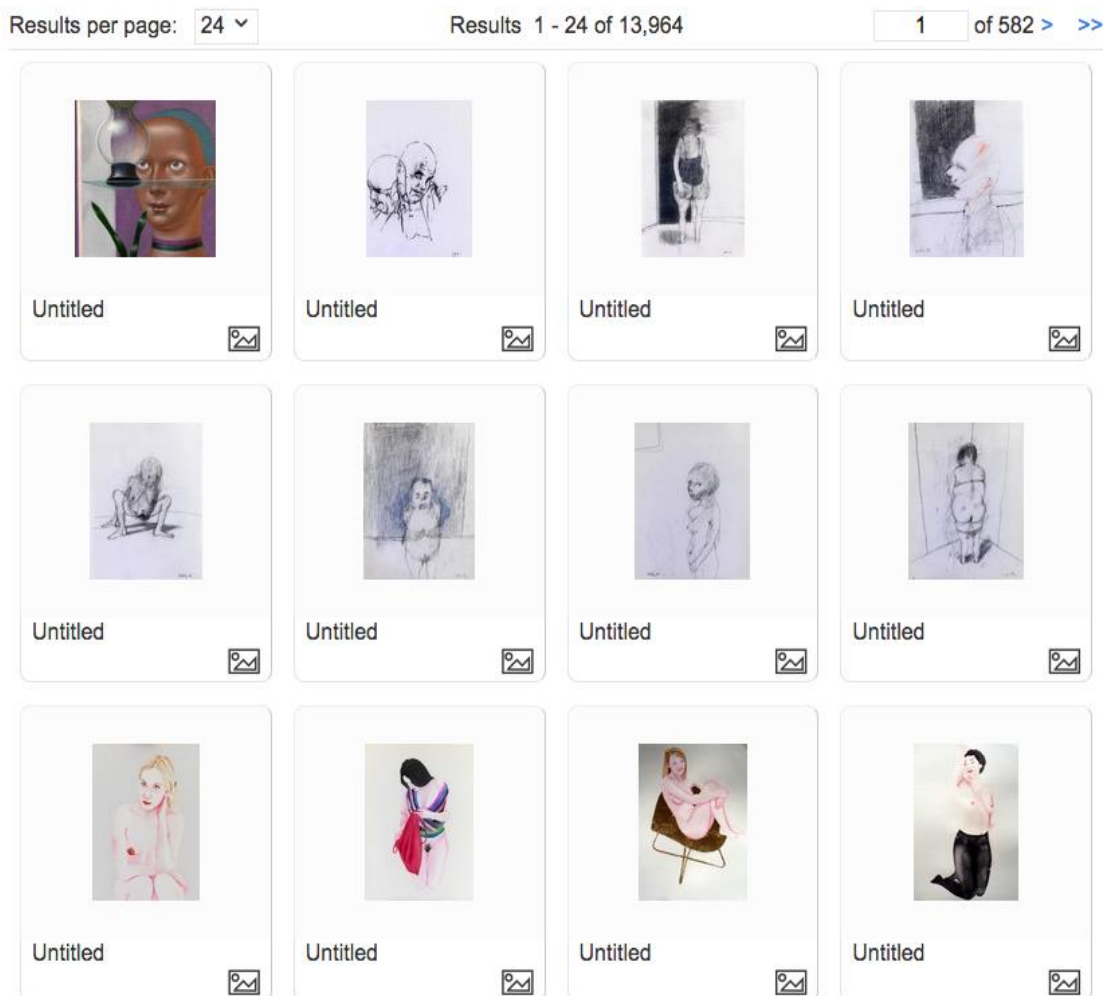
### **3.5 Type of Material Digitised**

Some materials that are digitised may not have much metadata. There is no way around this unless data providers make a conscious decision to provide and maintain digital curated content. The metadata required for say modern art or ceramics will be less rich than metadata for political papers in an archive, for which there is more



information available to start with. This is not to say that the metadata that does exist is not appropriate, but that its digitisation and publishing online may not generate the desired result for the institution. Thus the type of material digitised needs to be considered in terms of the metadata it will create.

There are cases when the term 'Untitled' is used to name an object. This should only be acceptable in limited cases, such as modern art collections. In these cases, the accompanying metadata should be of higher quality in order to fill this metadata gap - information on the creator, the collection and country should be provided as a minimum.



*Fig 3.1 Untitled cultural heritage objects do not work on a large-scale non-institutional repository like Europeana and must be in context.*

### 3.6 Lack of Understanding of Technical Documentation

Sometimes, the partners Europeana is in contact with about specific collections are not those with technical backgrounds. This means that while projects/aggregators want to provide quality feedback to their data providers, they might not fully



understand the requirements for aggregation to Europeanana. This has already been flagged as a recommendation in this report.

Initially, the Task Force looked towards the technical language of the EDM documentation as a blocker for submitting high quality metadata. The following section outlines some of the key points of misunderstanding raised by data providers, Task Force members and the Aggregators' Forum 2014. The Task Force recommends existing documentation on the Europeanana Data Model available on the 'Share your documentation' section of Europeanana Pro should be revisited and, where necessary, rewritten or expanded, to make sure it works for its intended audience, but also acknowledges that this might be difficult for Europeanana staff to maintain in the long term.

Europeanana encourages the use of templates showing data partners what information is required for the type of object they are submitting.<sup>27</sup> The templates include a brief explanation of how the metadata fields correspond to the Europeanana portal display. This information can help the data provider decide which elements they would like to include. The Task Force recommends that Europeanana encourages the use of existing templates and that these are further developed.

12 Templates should be used to ensure a certain metadata structure and standard is met.

The table below is an elaboration of the object templates available on GitHub with the descriptions slightly expanded. The example table best represents an IMAGE type record. The rows in blue are Europeanana's mandatory elements, while the fields in white are standard recommended fields in the EDM guidelines. This table represents the current Europeanana display and structure as of April 2015, which may be subject to change in the future.

EDM field	Where it shows up on the Europeanana portal	Mandatory	Recommended	Description of element
<b>DC properties - ProvidedCHO/Proxy</b>				
dc:title	Title on the item page + Title in the lightbox if it exists	x	x	Mandatory element if dc:description is not supplied. Both are recommended
dcterms:alternative	Alternative Title		x	Recommended use if available, also if the object has multiple titles
dc:description	Description = The description of the submitted object	x	x	Existing mandatory element if dc:title is not supplied. Both are recommended

<sup>27</sup> EDM Object Templates <https://github.com/europeanana/corelib/wiki/EDMObjectTemplatesProviders>



dc:creator	Creator on the item page + Creator in the lightbox if it exists		x	Recommended if available for attribution purposes
dc:publisher	Publisher		x	Recommended if available for attribution purposes
dc:subject	Subject		x	Recommended preferably using appropriate vocabulary and links
dc:type	Type		x	The type of digital object represented. Preferably using appropriate vocabulary and links
dc:identifier	Identifier		x	Recommended to provide a numeric persistent identifier
dc:coverage	Coverage		x	Recommended for geographical coverage
dc:date	Date		x	Date related to the creation of the physical object. Recommended if available for IPR purposes, and to make objects more discoverable
dcterms:created	Date of creation		x	Preferably for the digital copy
dcterms:issued	Publication date		x	
dcterms:spatial	Geographic coverage		x	Recommended for geographical coverage
dcterms:provenance	Provenance		x	Particularly if multiple providers are present in one dataset or collectors feature in the dataset, as this is useful to researchers
dc:format	Format		x	Makes it easier for end-users to find particular objects
dc:source	Source			Recommended if images come from a particular collector or a collection in an institution
dc:relation	Relation		x	Similar justification as dcterms:provenance. Relation to an item in the same or other collection




dcterms:hasPart	Has Part		x	Recommended if images are part of a set e.g. related to other objects or have multiple web:resources
dcterms:isPartOf	Is part of			Recommended if images are part of a set e.g. related to other objects or have multiple web:resources
<b>EDM properties - ProvidedCHO/Proxy</b>				
edm:isNextInSequence	Next in sequence		x	Strongly recommended. Must then be related to links supplied in webResource, and to the links supplied in edm:hasView
edm:isRelatedTo	Related to			Recommended if it is part of a hierarchy shows relation to another object
edm:type	Influences the default thumbnail + generates the Label for isShownBy link if it exists			Existing mandatory element
<b>IPR Information:</b> This shows the Intellectual Property Right elements found in an EDM record				
edm:rights (in Aggregation)	Shown under the thumbnail on the item page + in the lightbox if it exists	x		Existing mandatory element
dc:rights (in Aggregation)		x		Recommended element for attribution purposes, can be institutional e.g. if for CC BY SA it can link to the person or institution which requires the attribution
<b>Content providers - Part of Aggregation Class</b>				
edm:provider	Provider on the item page + Provider in the lightbox if it exists	x		Existing mandatory element
edm:dataProvider	Data provider + Label for "View item at" link on the item page + in the lightbox if it exists	x		Existing mandatory element
<b>Digital resources - Aggregation</b>				
edm:isShownAt	"View item at" link on the item page + in the lightbox if it exists	x	x	Both isShownBy or isShownA are recommended but only one is mandatory
edm:isShownBy	Link for "View", "Read", "Play" at the bottom of the	x	x	Both isShownBy or isShownA are





	thumbnail			recommended but only one is mandatory
edm:object	used for thumbnail caching		x	Recommended element to create the thumbnail. Creates greater visibility of digital collections
edm:hasView	Extra images available in the image carousel		x	Recommended if there is a webResource class
edm:UGC	Shows if material is user-generated content	x		Existing mandatory element if applicable to the material

Records such as the one below would be considered by the Task Force to be of good metadata quality and feature all mandatory elements, extended recommended elements, links to an online vocabulary, with the correct rights statement for the work, and with a high resolution image (clicking on the thumbnail will generate a larger image supplied by the institution).



View

Public Domain marked

View item at [Skoklosters slott](#)

Share

Cite on Wikipedia

Label format

Translate details

Select language

Powered by Microsoft Translator

### Greve Wrangel af Salmis, barn ( Augustus Gideon?). Son till fältmarskalken Karl Gustav Wrangel. Oljemålning på duk.

**Description:**  
Stående helfigur, fas höger. Röd kolt med vitt spetsförkläde röd hatt med vit plym, i högra handen käpp. I vänster hand nejlika. Grönt draperi i bakgrunden. Ny spikad spännram, smalare kant än 702, 703, 704, 707. Försilvrad ram med grön marmorering. Listens bredd 100 mm. 1800-talslapp: " Wrangel son af Riksamiralen Grefve Carl Gustaf Wrangel".

**Creator:**  
[Jacob Heinrich Elbfas \(Tilverkare, \)](#); [Elbfas, Jacob Heinrich](#), [Jacob Heinrich Elbfas](#) (Elbfas, Jacob Heinrich; Jacob Heinrich Elbfas) To: 1664+02:00 Date of death: 1664+02:00

**Contributor:**  
[Carl Gustaf Wrangel \(Tidigare ägare\)](#)

**Date of creation:**  
1650 cirka

**Type:**  
[child](#); [Oljemålning](#)

**Format:**  
770 mm,; 940 mm,

**Subject:**  
[Porträtt \(Konst och konsthantverk, Måleri\)](#); [Augustus Gideon Wrangel af Salmis\(Avbildning\)](#)

**Identifier:**  
(Inv. nr. 705); 21997

**Data provider:**  
[Skoklosters slott](#)

**Provider:**  
[AthenaPlus](#)

**Providing country:**  
Sweden

[Auto-generated tags](#)

[What](#)

Fig 3.2 Example of rich metadata record from Skoklosters Slott via AthenaPlus



There is a technical element to the creation of metadata, and this is where all parties involved in metadata creation, digital management, aggregation, publication and dissemination, need to be working from the same knowledge base. As has been pointed out in several places in the report, the technical documentation for EDM can be confusing, but the table above should go some way to explaining the elements as they are in relation to the portal, and add to the knowledge base of all parties involved in the Europeana Network.

The table above should be used to decide which elements are relevant to the institution and helpful to their dissemination goals. The desired elements should then be used to generate a crosswalk, and be applied via an XSLT conversion to create valid EDM XML. This should resemble something like the XML below but with valid values in the metadata fields. A data provider can provide many more metadata elements and classes and it is preferable that they do so to make the digital cultural heritage object more discoverable.

Below is how the metadata should look in XML. It can be considered as a template as the metadata fields have been clearly explained so that non-technical partners can comprehend more thoroughly what is needed. As mentioned previously, EDM documentation is not very accessible for non-technical partners and this is something that needs to be addressed.

```
<?xml version="1.0" encoding="UTF-8" ?>
<rdf:RDF xmlns:ore="http://www.openarchives.org/ore/terms/"
  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xmlns:rdaGr2="http://rdvocab.info/ElementsGr2/"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:skos="http://www.w3.org/2004/02/skos/core#"
  xmlns:crm="http://www.cidoc-crm.org/rdfs/cidoc_crm_v5.0.2_english_label.rdfs#"
  xmlns:wgs84="http://www.w3.org/2003/01/geo/wgs84_pos#"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:xalan="http://xml.apache.org/xalan"
  xmlns:dcterms="http://purl.org/dc/terms/"
  xmlns:edm="http://www.europeana.eu/schemas/edm/"
  xmlns:foaf="http://xmlns.com/foaf/0.1/">
  <edm:ProvidedCHO rdf:about="choose own identifier here: can be a unique URI or
  catalogue identifier"/>
  <edm:WebResource rdf:about="Include valid URI for the chosen web resource">
    <dc:format>Include format type</dc:format>
    <dc:rights xml:lang="en">Copyright information or desired attribution goes
    here</dc:rights>
```



```
<dc:rights xml:lang="nl">Copyright information or desired attribution goes
here in a different language. Can be different to rights in the Aggregation
class</dc:rights>
<dcterms:created>when the item was digitally created</dcterms:created>
<edm:rights rdf:resource="list one of the rights statement
http://pro.europeana.eu/web/guest/available-rights-statements"/>
</edm:WebResource>
<ore:Aggregation rdf:about="use a unique identifier that was not used in the
edm:ProvidedCHO ">
  <edm:aggregatedCHO rdf:resource="can be the same identifier as
ProvidedCHO"/>
  <edm:dataProvider>Name of Data Provider</edm:dataProvider>
  <edm:isShownAt rdf:resource="URI of the digital cultural heritage object
with metadata"/>
  <edm:isShownBy rdf:resource="URI of non embedded hi-resolution digital
object to create light box "/>
  <edm:object rdf:resource="Can be the same or lesser quality of the
isShownBy link generates thumbnail"/>
  <edm:provider>Name of Provider - Project or Aggregator names go
here</edm:provider>
  <dc:rights xml:lang="en">Copyright information or desired attribution
goes here</dc:rights>
  <dc:rights xml:lang="nl">Copyright information or desired attribution goes
here in a different language. Can be different to rights in the Web
Resource class</dc:rights>
  <edm:rights rdf:resource="list one of the rights statement
http://pro.europeana.eu/web/guest/available-rights-statements"/>
</ore:Aggregation>
<ore:Proxy rdf:about="choose own identifier here can be a unique URI or
catalogue identifier">
  <dc:creator>Name the creator</dc:creator>
  <dc:description xml:lang="en">Succinct description of object in chosen
language </dc:description>
  <dc:format rdf:resource=" use links to appropriate vocabulary e.g
http://vocab.getty.edu/aat/300014109"/>
  <dc:identifier>unique object identifier</dc:identifier>
  <dc:language>insert the language of the metadata</dc:language>
  <dc:subject rdf:resource="Use a URI from one of the vocabularies
mentioned above e.g. http://iconclass.org/rkd/32B2\(ROMANSY\)"/>
  <dc:subject>Free text form of subject <dc:subject/>
  <dc:subject xml:lang="en">Free text form of subject in chosen
language</dc:subject>
  <dcterms:alternative>Common or alternative name of the
object</dcterms:alternative>
  <dc:type rdf:resource="URI for appropriate online vocabulary e.g.
http://vocab.getty.edu/aat/300041273"/>
  <dc:type xml:lang="en">Free text of object type in chosen
language</dc:type>
  <dc:date>When original object was created</dc:date>
  <dcterms:extent xml:lang="en"> dimensions or object
length</dcterms:extent>
  <dcterms:spatial>The country of the object's origin</dcterms:spatial>
  <edm:type>From Europeana's type list
AUDIO/IMAGE/VIDEO/SOUND/3D</edm:type>
</ore:Proxy>
</rdf:RDF>
```



## Task Force Recommendations: Future Steps

The Europeanana Task Force on Metadata Quality was a short general scope Task Force which raised more questions than it answered. It suggested many more areas to explore in order to get a fuller picture of how the Europeanana Network can help improve metadata quality submission. The Task Force looked at metadata quality in a more global context in order to both motivate institutions to provide better metadata, and to improve the service provided to our audiences. The Task Force recommendations are divided into three sections: motivational, technical and contextual.

### Motivational

The Task Force recommends that data providers provide *curated* datasets to Europeanana. This hopefully means that providers and data providers submit datasets with greater metadata quality and better context, instead of large data dumps of institutional records. It is also hoped that they consider what they want for their institution by publishing records in Europeanana and that they focus on digital appraisal. If they want to showcase their treasures or the most important objects in their collection, it might be better for the institution, the aggregator, Europeanana and the audience to submit a smaller curated dataset. These curated datasets should be given priority in ingestion and display in Europeanana.

This kind of focus would require a greater effort on behalf of the data provider, but the return on investment may be greater in terms of representation, increased institutional profile and increased online and offline visitors.

Data providers have voiced that they do not feel involved in the ingestion process, or do not trust the processes of ingestion as outlined in the first section of this report. To maintain standards and outline the metadata requirements, it is recommended that Europeanana make greater use of the Europeanana Publication Policy.<sup>28</sup> The Publication Policy details the information data partners need to provide to publish their collections on Europeanana, and it is split into three areas: content, technical and legal. The Publication Policy also explains how Europeanana will query the submission of poor quality metadata, for example, where mandatory elements contain a single repeated word making a digital cultural heritage object undiscoverable. The policy communicates to partners what is acceptable in terms of data quality, and makes the ingestion process more transparent for both Europeanana and data partners.

---

<sup>28</sup> Europeanana Publication Policy  
[http://pro.europeana.eu/files/Europeana\\_Professional/Aggregation/Europeana%20-%20Publication%20Policy.pdf](http://pro.europeana.eu/files/Europeana_Professional/Aggregation/Europeana%20-%20Publication%20Policy.pdf)



It is recommended that the Europeana Aggregation team be more available to the data partners so that they feel that they are a fuller part of the Europeana Network. This availability could take the form of a fortnightly or monthly open hour (depending on demand) for data providers to bring questions directly to the Aggregation team. This was attempted in Europeana Version 1 and again in Europeana Version 2 in the form of an online forum but with little participation. It is hoped that changing the format of the forum to video will make it more accessible and encourage both data providers and aggregators/projects to participate.

A Task Force on the impact of the curation of collections would collaborate with a few institutions and their collections in order to see if curation results in better quality metadata and has an impact on the access and re-use of a collection, compared to larger datasets.

## **Technical**

The Task Force sees a need for existing documentation on EDM available on Europeana Pro to be rewritten or expanded. There are multiple audiences for EDM and different people may have different understandings of some of the documentation. However, it is possible that providing increased access to the Europeana Aggregation team, in addition to the explanations given in this report, may suffice.

The Task Force recommends that data providers and aggregators document their metadata crosswalks to EDM and then make them available to all partners. By doing this it is hoped that creating EDM-compliant metadata will become easier and standardised across formats, and that a sense of community will be created, with all partners working towards a specific metadata quality goal.

The Task Force also recommends that data providers and aggregation services make use of linked open data vocabularies such as VIAF, Iconclass and the Getty Art and Architecture Thesaurus and provide relevant URIs in the metadata submitted to Europeana. The Task Force similarly advises that parties enrich their data prior to submitting to Europeana to provide the best possible data. Doing this should also satisfy the recommendations from the Task Force on Semantic Enrichment and Multilinguality. If data providers and aggregation services implement these recommendations, the findability of records on Europeana will be increased, giving audiences better access to cultural heritage.

There was not enough scope for this Task Force to investigate elements such as metrics for metadata quality or how EDM schema validation could affect metadata quality. It is recommended that Task Forces be set up to investigate these areas. These Task Forces would give measurable outcomes regarding metadata quality which could prove useful for all Europeana Network parties.



## Contextual

Putting records in context begins before the metadata is created. All parties involved with the publishing of metadata on Europeana, from the data creator to the Europeana Aggregation team, need to have trust in each other's processes. The Task Force recommends that both Europeana's and data partners' metadata processes should be made more transparent through improved documentation and discussion. This transparency should create more trust and help ensure that the submitted metadata does not lose quality anywhere along the way.

The Task Force notes that building and maintaining this trust should not be the sole responsibility of Europeana. The Task Force recommends that providers increase the metadata checks undertaken at both the data creator level and the aggregator/project level. Aggregation partners, such as project EUScreen and EUScreenXL, who have encouraged this type of activity, have a higher level of quality across their datasets.

Such checks should raise the awareness of the need for context. Submitted metadata should not only conform to the EDM model but also provide within it the appropriate context for the object represented. This is best expressed via a curated dataset which maintains the collection. Increased context reduces the likelihood of records becoming invisible.

It is also recommended that the minimum standard within the mandatory elements be raised. This means that the records should be more findable via searches on Europeana and via the Europeana API, give more exposure to the data providers and aggregators, and properly showcase the digital cultural heritage object.

By combining all of the recommendations of the Task Force, it is hoped that there will be an improvement in the quality of metadata submitted to Europeana in the future, and that the digital cultural heritage is shown to its fullest potential online.



## Conclusion

The Europeana Metadata Quality Task Force was able to take a brief look into what metadata quality means in relation to Europeana and the Europeana Network. Having first established what metadata quality is in this context, we can now go further, investigating how to improve or measure metadata quality in the future.

As the Task Force was organised by Europeana - an organisation known for pan-European metadata aggregation - the questions were skewed to the creation of metadata for widespread dissemination online. However, as members of the Task Force worked closely with data providers or are data providers, the next steps and actions have been designed to reach a mutually beneficial solution for improving metadata quality for all parties.

The Task Force raised many questions regarding a permanent standard for good metadata quality, which will hopefully be taken on, researched and resolved in the future.

Den Haag, May 2015