

Provenance of metadata enrichments and translations in EDM *Internal*

EDM Extension

Version	0.9.1-beta
Author	Albin Larsson
Reviewers	Hugo Manguinhas, Valentine Charles, Antoine Isaac, Nuno Freire
Date	March 2022
Collaboration	Beta version was developed under the Europeana Generic Services Europeana XX Century of Change project ¹ . Funded under the Connecting Europe Facility Programme under the European Health and Digital Executive Agency (HaDEA).



European Commission

European Health and Digital Executive Agency (HaDEA)

Connecting Europe Facility

¹ <https://pro.europeana.eu/project/europeana-xx>

Introduction and background	2
Related documents	2
Specification	3
The new proxy/aggregation pair	3
New properties	3
Provenance with edm:provider and edm:dataProvider	4
Data quality annotations	5
Appendix A – Example record	5
Appendix B – Discussion and motivation	7
Appendix C – Relation to existing data	8
Document History	9

Introduction and background

In the current aggregation model, multiple aggregators do not only forward Cultural Heritage Institution (CHI) data to Europeana but also enrich it. The Europeana XX project² raised the need to distinguish original CHI data from aggregator-enriched data; similarly to how EDM currently distinguishes Europeana enrichments from the data provided by aggregators.

This EDM profile extends EDM and builds upon Europeana's approach with ORE proxies and aggregations to distinguish data from different sources as outlined in the EDM Primer³. Currently, in Europeana's internal EDM data, there are two aggregation/proxy pairs representing data *delivered* by a provider and data created by Europeana. This profile extends this with an additional set of proxies/aggregations for data known to originate from the aggregator while reserving the existing provider proxy for data originating from a CHI. Thus being able to retain provenance information.

Related documents

- [Provenance of metadata enrichments and translations in EDM *External*](#)

² <https://pro.europeana.eu/project/europeana-xx>

³

https://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation/EDM_Primer_130714.pdf

Specification

The new proxy/aggregation pair

The new proxy/aggregation pair should have a URI following the pattern outlined below.

```
http://data.europeana.eu/<proxy|aggregation>/aggregator/<dataset id>/<record id>
```

The new aggregator aggregation and proxy shall include properties mandatory to **ore:Aggregation** and **ore:Proxy** respectively (as defined by the EDM ORE schema⁴). This means that *edm:rights*, *edm:isShownBy*, etc will need to be present on several aggregations.

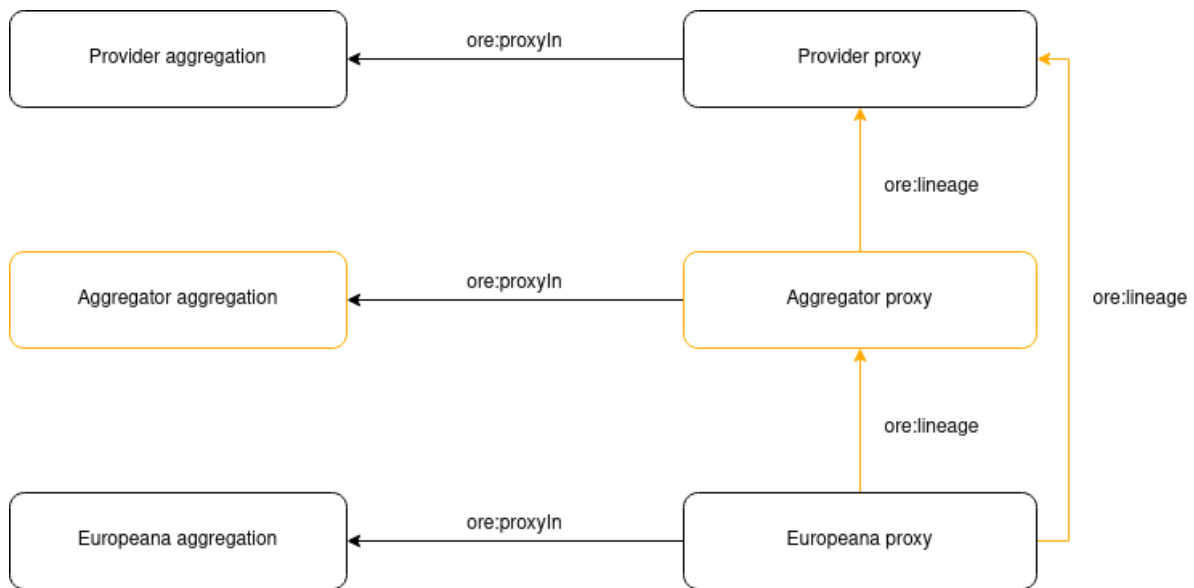
New properties

The only new property is *ore:lineage*. This addition is motivated by the need to be able to identify the source of a proxy similarly to how the *edm:europeanaProxy* has been used.

URI	ore:lineage
Rationale	ore:lineage indicates that a proxy originated or was sourced from another proxy. It allows the distinction between data provider's original metadata from aggregator-enriched metadata, by keeping track of the provenance of the proxies.
Obligation & Occurrence	Mandatory on each proxy pair.

⁴ <https://www.europeana.eu/schemas/edm/ORE.xsd>

The illustration below highlights the additional set of proxies/aggregations for CHI data as well as the usage of the *ore:lineage* property.



Because *ore:lineage* is not transitive, the last (above the Europeana proxy) proxy will need to have an *ore:lineage* property for all previous proxies.

Note: The OAI-ORE Abstract Model states that in [Section 5.5.5](#) that "One Proxy MUST NOT be the subject of more than one *ore:lineage* relationships expressed in a Resource Map.". It is not clear if this use of *ore:lineage* in EDM is a cause for concern, since this constraint is not mentioned in the [OAI-ORE vocabulary](#), and in EDM, *ore:lineage* is used without a *ore:ResourceMap*. A future revision of the model for provenance of enrichments in EDM should clarify this situation.

Provenance with *edm:provider* and *edm:dataProvider*

The usage of the properties *edm:provider* and *edm:dataProvider* has in this profile been extended. Not only should both of these properties be present on the new aggregation but they should also be present on the **edm:EuropeanaAggregation** class. To allow consumers to access this provenance information, while remaining compatible with the current semantics of these two properties, on each aggregation both *edm:provider* and *edm:dataProvider* should be set to the same value except in the provider one where it follows the same pattern as in EDM.

Aggregation	<i>edm:dataProvider</i>	<i>edm:provider</i>
provider	Name of CHI	Name of aggregator

aggregator	Name of aggregator	Name of aggregator
Europeana	"Europeana Foundation"	"Europeana Foundation"

Note that the related *edm:intermediateProvider* should go in the provider aggregation although there could be future scenarios where it would go on the aggregator aggregation.

Data quality annotations

Because *ore:lineage* is not transitive, the *oa:hasTarget* property on a potential **dqv:QualityAnnotation** would need to point to each aggregation it targets and not only to the proxy created in the last aggregation step.

Appendix A – Example record

The following record excludes the majority of required properties (*edm:type*, *edm:isShownBy*, etc) as well as technical properties. It illustrates an aggregator enriching a record with a translation of the *dc:description* field as well as resolving the given *dcterms:spatial* value to its Wikidata URI.

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:dcterms="http://purl.org/dc/terms/"
  xmlns:edm="http://www.europeana.eu/schemas/edm/"
  xmlns:ore="http://www.openarchives.org/ore/terms/"
  xmlns:oa="http://www.w3.org/ns/oa#"
  xmlns:dqv="http://www.w3.org/ns/dqv#">
  <edm:ProvidedCHO rdf:about="http://data.europeana.eu/item/1998/123"/>

  <ore:Aggregation rdf:about="http://data.europeana.eu/aggregation/provider/1998/123">
    <edm:aggregatedCHO rdf:resource="http://data.europeana.eu/item/1998/123"/>
    <edm:dataProvider>Järnvägmuseet</edm:dataProvider>
    <edm:provider>K-samsök</edm:provider>
  </ore:Aggregation>

  <ore:Proxy rdf:about="http://data.europeana.eu/proxy/provider/1998/123">
    <dc:description xml:lang="no">brus er ikke bra for tennene</dc:description>
    <dcterms:spatial xml:lang="sv">Stockholm</dcterms:spatial>
    <edm:europeanaProxy>false</edm:europeanaProxy>
    <ore:proxyFor rdf:resource="http://data.europeana.eu/item/1998/123"/>
    <ore:proxyIn
```

```
<rdf:resource="http://data.europeana.eu/aggregation/provider/1998/123"/>
</ore:Proxy>

<ore:Aggregation rdf:about="http://data.europeana.eu/aggregation/aggregator/1998/123">
  <edm:aggregatedCHO rdf:resource="http://data.europeana.eu/item/1998/123"/>
  <edm:dataProvider>K-samsök</edm:dataProvider>
  <edm:provider>K-samsök</edm:provider>
  <edm:rights rdf:resource="http://rightsstatements.org/vocab/InC-OW-EU/1.0"/>
</ore:Aggregation>

<ore:Proxy rdf:about="http://data.europeana.eu/proxy/aggregator/1998/123">
  <dc:description xml:lang="sv">läsk är inte bra för tänderna</dc:description>
  <dcterms:spatial rdf:resource="http://www.wikidata.org/entity/Q1754"/>
  <edm:europeanaProxy>false</edm:europeanaProxy>
  <ore:proxyFor rdf:resource="http://data.europeana.eu/item/1998/123"/>
  <ore:proxyIn
    rdf:resource="http://data.europeana.eu/aggregation/aggregator/1998/123"/>
  <ore:lineage rdf:resource="http://data.europeana.eu/proxy/provider/1998/123"/>
</ore:Proxy>

<edm:EuropeanaAggregation rdf:about="http://data.europeana.eu/aggregation/europeana/1998/123">
  <edm:aggregatedCHO rdf:resource="http://data.europeana.eu/item/1998/123"/>
  <edm:dataProvider>Europeana Foundation</edm:dataProvider>
  <edm:provider>Europeana Foundation</edm:provider>
  <dqv:hasQualityAnnotation
    rdf:resource="http://data.europeana.eu/item/1998/123#contentTier"/>
  <dqv:hasQualityAnnotation
    rdf:resource="http://data.europeana.eu/item/1998/123#metadataTier"/>
</edm:EuropeanaAggregation>

<ore:Proxy rdf:about="http://data.europeana.eu/proxy/europeana/1998/123">
  <edm:europeanaProxy>true</edm:europeanaProxy>
  <ore:proxyFor rdf:resource="http://data.europeana.eu/item/1998/123"/>
  <ore:proxyIn
    rdf:resource="http://data.europeana.eu/aggregation/europeana/1998/123"/>
  <ore:lineage rdf:resource="http://data.europeana.eu/proxy/provider/1998/123"/>
  <ore:lineage rdf:resource="http://data.europeana.eu/proxy/aggregator/1998/123"/>
</ore:Proxy>

<dqv:QualityAnnotation
  rdf:about="http://data.europeana.eu/item/1998/123#contentTier">
  <oa:hasTarget
    rdf:resource="http://data.europeana.eu/aggregation/provider/1998/123"/>
  <oa:hasTarget
    rdf:resource="http://data.europeana.eu/aggregation/aggregator/1998/123"/>
  <oa:hasBody rdf:resource="http://www.europeana.eu/schemas/epf/contentTier2"/>
</dqv:QualityAnnotation>
```

```
<dqv:QualityAnnotation
rdf:about="http://data.europeana.eu/item/1998/123#metadataTier">
  <oa:hasTarget
    rdf:resource="http://data.europeana.eu/aggregation/provider/1998/123"/>
  <oa:hasTarget
    rdf:resource="http://data.europeana.eu/aggregation/aggregator/1998/123"/>
  <oa:hasBody rdf:resource="http://www.europeana.eu/schemas/epf/metadataTierC"/>
</dqv:QualityAnnotation>

</rdf:RDF>
```

Appendix B – Discussion and motivation

Throughout the work on this profile, two approaches regarding the new aggregation/proxy pair have been discussed. One envisioned a new pair for data originating from the CHI and one (the final one) envisioned a new pair for data originating from the aggregator.

The first approach which envisioned a new “data_provider” aggregation/proxy pair for CHI data would suit Europeana’s “provider”/“data provider” terminology⁵ and not introduce additional confusion (see below). The major drawback of this approach is that it changes the purpose of the existing provider pair, and as a result the “main” pair for legacy data would be the provider one, while for records using this profile, it would be the data_provider pair. This would be particularly problematic for existing metadata consumers of Europeana data that would need to adapt and be able to distinguish between the profiles.

The second approach, which ended up being the final one, has the drawback of adding additional confusion regarding the terminology as the “provider” pair and its URIs now are intended for CHI data (the existing “provider” proxy/aggregation pair is indeed actually referring to the “data provider” in the official Europeana terminology).

A third approach that was envisioned, but considered unimplementable at this point in time, was to create local pairs/URIs for each organization in the aggregation process using its identifier from the Entity Collection. Below is the proxy URI of such a scenario conceptually illustrated:

```
http://data.europeana.eu/proxy/<dataset id>/<record id>#148225000004503474
```

Such an approach would go beyond a limited set of aggregation/proxy pairs and allow us to describe arbitrarily complex aggregation chains.

⁵ <https://pro.europeana.eu/page/glossary>

Appendix C – Relation to existing data

While only records using the attributes defined in the extension to EDM External shall be transformed into EDM Internal according to the profile defined by this document, the practice of adding *edm:dataProvider* and *edm:provider* to the Europeana aggregation should be applied to all Europeana records.

Enrichments that were done by aggregators might now appear to be in the “wrong” proxy/aggregation pair. This problem is not addressed in this document.

Document History

Version	Editor	Date	Comments
V0.9-beta	Albin Larsson,Hugo Manguinhas, Valentine Charles, Antoine Isaac	30/11/2021	Beta version developed under the Europeana Generic Services Europeana XX Century of Change project.
v0.9.1-beta	Nuno Freire, Antoine Isaac	21/03/2022	Editorial revision and alignment with the EDM Definition.