
# Europeana Data Model Primer

10/02/2011 – **Editor's Draft**

## Europeana v1.0

co-funded by the European Union

The project is co-funded by the European Union, through the **eContent*plus*** programme
http://ec.europa.eu/econtentplus

## Disclaimer

This document and the companion EDM Definition available from the Europeana v1.0 website (http://version1.europeana.eu/web/europeana-project/technicaldocuments/) reflect the consensus reached in discussions in the Europeana v1.0 Work Package 3 meetings in 2009 and the first half of 2010.

These documents are complete and consistent but the model described may undergo changes as a result of testing in the prototype.

It is the intention that these documents will remain stable until the end of the prototyping phase. At that time, any changes that need to be made as a result of prototyping and further discussions with technical experts and content providers will be included in the next version of the Europeana Data Model.

Users of these documents should be aware that changes may occur between this release and the next one, which are not backwardly compatible. The team will only make such changes if they are absolutely necessary. However, developers that are working with these versions will need to be able to modify their software and re-convert their data to reflect potential changes.

Please also note that, as of February, the RDF schema for the namespace http://www.europeana.eu/schemas/edm/ is not yet in place. There is however ongoing effort as part of the prototyping activities to create such a schema. For further information, please contact info@europeana.eu (using “EDM” as subject) or look for more technical details on the EDM Prototyping Wiki at http://www.europeanalabs.eu/wiki/EDMPrototyping.

## Credits

This document has been edited by Antoine Isaac.
Carlo Meghini, Makx Dekkers, Stefan Gradmann, Robina Clayphan, Jan Molendijk and Jonathan Purday have reviewed it and made valuable improvements.
Martin Doerr provided the Tuthmosis III amphora example; Victor de Boer and Steffen Hennicke contributed to the EDM representation used as illustration for hierarchical objects.
The many discussions on EDM with other participants of WP3 in Europeana v1.0 have greatly helped to define the motivation, scope and content of this document.

# Table of Contents

# 1 Introduction

## *1.1 Overview*

The Europeana Data Model (EDM) is a new proposal for structuring the data that Europeana will be ingesting, managing and publishing. The Europeana Data Model is a major improvement on the Europeana Semantic Elements (ESE), the basic data model that Europeana began life with.

Each of the different heritage sectors represented in Europeana uses different data standards, and ESE reduced these to the lowest common denominator. EDM reverses this reductive approach and is an attempt to transcend the respective information perspectives of the sectors that are represented in Europeana – the museums, archives, audiovisual collections and libraries. EDM is not built on any particular community standard but rather adopts an open, cross-domain Semantic Web-based framework that can accommodate the range and richness of particular community standards such as LIDO [LIDO] for museums, EAD[1] for archives or METS[2] for digital libraries.

EDM not only supports the full richness of the content providers' metadata but also enables data enrichment from a range of third party sources. For example, a digital object from Provider A may be contextually enriched by metadata from Provider B. It may also be enriched by the addition of data from authority files held by Provider C, and a web-based thesaurus offered by Publisher D. EDM supports this richness of linkage, while clearly showing the provenance of all the data that links to the digital object.

EDM also supports more complex objects than ESE is able to. In terms of a digitised book, the individual chapters, illustrations and index can be understood both individually and collectively; in terms of an archival finding aid or *fonds*, the constituent letters, deeds, manuscripts or other items can be similarly understood.

EDM is still under development, and will continue to be refined until the end of 2010. It will be implemented during 2011, in the lead up to the Danube release of Europeana. Before, during and after the implementation of EDM, data that is compliant only with ESE will continue to be accepted. EDM is compatible with ESE and no data will need to be resubmitted. Europeana will make available a convertor, and any provider who wishes to resubmit data, in order to increase its richness within Europeana, will be able to do so if they wish but will be under no obligation.

EDM will let users browse Europeana in revealing new ways. It answers the 'Who?', 'What?', 'When?', 'Where?' questions, and makes connections between the networks of stories that will animate Europeana's content. This linking of data is supported by the open structure of the EDM, and will put Europeana in the vanguard of semantic web developments.

[1] http://www.loc.gov/ead/
[2] http://www.loc.gov/mets/

## 1.2 The rationale behind the Europeana Data Model (EDM)

EDM is a qualitative change in the way Europeana deals with the metadata gathered from data providers and aggregators. It is aimed at solving some of the issues observed with the current ESE, by providing extra expressivity and flexibility.

In particular, it makes a distinction between the intellectual and technical creation that is submitted by a provider (a bundle of resources about an object curated by the provider), the object this structure is about, and the digital representations of this object, which can be accessed over the web.

Also, EDM adheres to the modelling principles that underpin the approach of the Web of Data ("Semantic Web"). In that approach, there is no such thing as a fixed schema that dictates just one way to represent data. A common model like EDM can be seen instead as an anchor to which various finer-grained models can be attached, making them at least partly interoperable at the semantic level, while the data retain their original expressivity and richness. It does not *require* changes in the local approaches, although any changes in local practice that increase the cross-domain usefulness of the data is *encouraged*, such as the use of publicly accessible vocabularies (for persons, places, subjects etc.).

In this sense, EDM is an attempt to transcend the respective information perspectives of the various communities constituting Europeana, such as museums, archives, audio-visual collections and libraries. EDM is not built on any particular community standard but rather adopts an open, cross-domain Semantic Web-based framework that can accommodate particular community standards such as LIDO, EAD or METS.

## 1.3 How to read this Primer

EDM is more difficult to grasp than ESE and similar approaches to interoperability. This primer is a complement to the normative specification of EDM elements [EDM-Definition]. It tries to lay out clearly the main EDM features, and discuss the motivations behind them. It also provides examples illustrating how EDM can be used, either by a provider who submits EDM data to Europeana or by Europeana.eu itself.

In more detail, EDM enables the representation and accessing of objects provided to Europeana, via the packages of digital representations submitted by Europeana providers (sect. 4.1). It also provides support for ingesting the descriptive metadata submitted by various providers, possibly for the same object, and representing new information added by Europeana (sect. 4.2 to 4.5). In addition, EDM accommodates various description paradigms for the ingested objects (sect. 5.1), and paves the way for enriching objects by connecting them to (networks of) semantically enriched resources (sect. 5.2), Crucially, EDM does this while still allowing for different levels of granularity in the descriptions, using the possibilities of semantic mapping (sect. 5.3). This allows Europeana to retain compatibility with existing description approaches, including the simpler Europeana Semantic Elements (ESE) currently used for data submission at Europeana (sect. 5.4). Advanced EDM features will be discussed in sect. 6.

It is perfectly possible for a reader experienced with EDM matters to jump to a specific section of interest. The reader new to EDM is however recommended to read it in sequential order, as one section tends to elaborate on aspects that are introduced in the previous one,

providing a "story" about EDM features. For example, Section 5 presents various aspects related to the representation of descriptive metadata in EDM. It can be read separately, but the reader should be aware of the motivations for using “proxies”, which appear in all examples of this section. Section 6 presents information on more specific or optional details of Europeana, and may thus be considered independently of the rest of the document.

*RDF graphs*

The Primer features a number of graphs. These graphs have been created to provide the reader with a more intuitive view of examples. The reader should however be aware that these graphs are meant to represent data expressed in RDF, adapting the conventions used, e.g., by the RDF Primer [RDF-Primer]. This implies that they correspond exactly to a set of RDF “statements” (or “assertions”), using the following rules:

- a circled URI in normal font denotes a standard RDF resource. Two URIs’ being in a single circle indicates that one resource has been given two identifiers. Such situation may typically result from asserting an `owl:sameAs` statement between the two URIs.[3]
- a string enclosed with quotes denotes an RDF literal. It can carry a language tag, as for `"Example"@en`.
- an arrow between two resources (or between a resource and a literal) indicates an RDF statement ("triple") between these two resources. The object of the statement is the origin of the arrow; its subject is the target of the arrow. The predicate of the statement is the property indicated by the URI in normal font next to the arrow
- a URI in italic font denotes:
    - a type for the resource, if appearing in a "resource circle".
    - a super-property of the property, if appearing next to an "property arrow"

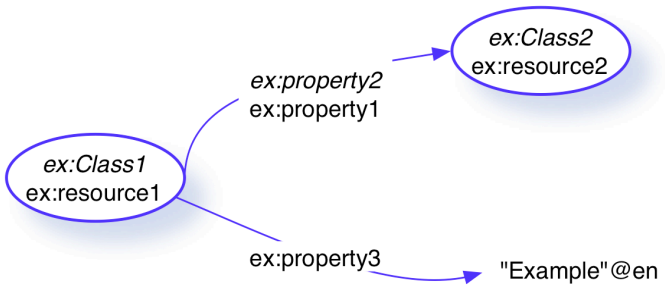



*Fig. 1 Simple RDF graph*

As an example, the graph above corresponds to the following RDF statements:

```
ex:resource1 rdf:type ex:Class1 .
ex:resource2 rdf:type ex:Class2 .
ex:resource1 ex:property1 ex:resource2 .
ex:resource1 ex:property3 "Example"@en .
ex:property1 rdfs:subPropertyOf ex:property2 .
```

*RDF syntaxes*

This document often uses the concise Turtle syntax [Turtle] for examples. Readers should be aware that these examples could very well have been given in the normative but much more verbose RDF/XML syntax [RDF/XML]. Interested readers can find a conversion tool at http://www.rdfabout.com/demo/validator/.

[3] Cf. http://www.w3.org/TR/owl-ref/#sameAs-def

*URI abbreviations*

For the sake of brevity a number of namespace declarations are omitted from the examples. This applies to standard namespaces RDF/RDFS [RDF-PRIMER], OWL [OWL], SKOS [SKOS], and Dublin Core [DC]) and namespaces for other ontologies (ORE [ORE], FOAF [FOAF], EDM-specific elements [EDM-Definition]) but also to namespaces coined for the examples.
Generally, these namespaces could be declared in the preamble of RDF (Turtle) files as in the following code:

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .
@prefix dc: <http://purl.org/dc/elements/1.1> .
@prefix dcterms: <http://purl.org/dc/terms/> .
@prefix ore: <http://www.openarchives.org/ore/terms/> .
@prefix ens: <http://www.europeana.eu/schemas/edm/> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix viaf: <http://viaf.org/viaf/> .
@prefix ex: <http://www.example.com/> .
@prefix ex1: <http://www.example.com/1/> .
@prefix ex2: <http://www.example.com/2/> .
@prefix ex3: <http://www.example.com/3/> .
@prefix ex-eu: <http://example.europeana.eu/> .
```

*Other editorial features*

Elements in Courier font, such as `ore:Aggregation`, refer to classes and properties introduced or re-used by EDM. For most of these, the readers can access definitions or documentary notes in the normative specification of EDM elements [EDM-Definition].

# 2 Reminder on EDM requirements and design principles

A number of requirements and principles have been formulated at various times, which have strongly influenced the design of EDM as it stands now. The reader should be aware of these while trying to figure out the motivation for some modelling choices in EDM.

Noticeable requirements are:
- R1. distinction between “provided object” (painting, book, movie, archaeology site, archival file, etc.) and digital representation
- R2. distinction between object and metadata record describing an object
- R3. multiple records for the same object should be allowed, containing potentially contradictory statements about an object
- R4. support for objects that are composed of other objects
- R5. compatibility with different abstraction levels of description (e.g. if a provider wishes to submit descriptions that follow the distinctions introduced in FRBR Group 1 [FRBR])
- R6. EDM provides a standard metadata format that can be specialized
- R7. EDM provides a standard vocabulary format that can be specialized

Also, a basic motivation for EDM is to support the integration of the various models used in Cultural Heritage data, so that all original descriptions could be collected and connected through higher-level concepts. This motivation, derived for the general goal of Europeana to exploit the richness of all available data in order to support the richest possible functionality, justifies three fundamental design principles:

- D1. EDM allows data integration in an *open* environment: it is impossible to anticipate all data contributed
- D2. EDM allows for rich functionality, possibly via extensions
- D3. EDM should re-use existing (standard) models as much as possible

These design principles are the basis for the choice of Semantic Web representation languages—RDF(S), OWL—for EDM. These allow flexible re-use and articulation of existing models, as demonstrated by the conception of the EDM model itself, and by the mapping approach to data integration which underlies the way EDM should be used in practice (cf. Sect. 5.3). Further, the Linked Data approach[4] emphasizes the re-use and linkage of richly described resources over the web. This really fits the EDM ambition of making use of existing resources as well as supporting their enrichment, notably via the establishment of new relations between them. Whether these resources belong to one Europeana provider's information space, to different providers' spaces, or to external spaces used as knowledge references.

# 3 Examples used in this document

This document uses examples from Europeana related to the Mona Lisa painting by Leonardo da Vinci. There are two records that are about the painting itself: one comes from the Joconde database,[5] the other from the Louvre database.[6] There is also a third record that describes a work inspired by the Mona Lisa painting, *Mona Lisa - 2000*.[7] The screenshots below show how these objects are on their providers' sites, together with their various digital representations and their metadata. In all cases, the metadata displayed on these screenshots closely reflect the source metadata. Another example is used for event-centric metadata, which is documented further in Annex 1.

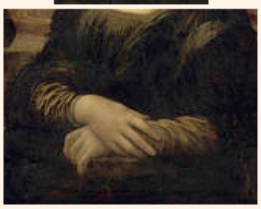
[4] http://linkeddata.org/

[5] http://www.europeana.eu/portal/record/03901/B69913DBC34A320BAC4B64EF64E08A9440CFE1E2.html, original record at http://www.culture.gouv.fr/public/mistral/joconde_fr?ACTION=CHERCHER&FIELD_1=REF&VALUE_1=000PE025604

[6] http://www.europeana.eu/portal/record/03902/B357C5A7F169D3DAF39F9A11C722574D5CBF0EEC.html, original record at http://cartelfr.louvre.fr/cartelfr/visite?srv=car_not_frame&idNotice=14153

[7] http://www.europeana.eu/portal/record/01004/AC2B3AA843934B18E804DD40BF6E7BDD9C04067F.html, original record at http://www.deutschefotothek.de/obj30131760.html

Réponse n° **1**



| | |
|---|---|
| Domaine | peinture |
| Type d'objet | tableau |
| Titre | **PORTRAIT DE MONA LISA (1479-1528) ; DITE LA JOCONDE** |
| Auteur/exécutant | LEONARDO DI SER PIERO DA VINCI ; VINCI Léonard de (dit) |
| Précision auteur/exécutant | Vinci, 1452 ; Amboise, 1519 |
| Ecole | Italie |
| Période création/exécution | 1er quart 16e siècle |
| Millésime création/exécution | 1503 entre ; 1506 et |
| Genèse | oeuvre en rapport ; reproduit en gravure |
| Historique | commandé par le florentin Francesco del Giocondo, époux de Mona Lisa entre 1503 et 1506 ; nombreuses copies dont une conservée au Louvre ; gravé par Fauchery, par Filhol, par Landon |
| Matériaux/techniques | peinture à l'huile ; bois |
| Mesures | 77 H ; 53 L |
| Sujet représenté | portrait (Mona Lisa, femme, à mi-corps, de trois-quarts, assis, accoudé, loggia, Italien) ; fond de paysage (montagne, rocher, cours d'eau, pont, plaine, route) |
| Date sujet représenté | 1479-1528 |
| Lieu de conservation | Paris ; musée du Louvre département des Peintures |
| | *Musée de France*<br>*au sens de la loi n°2002-5 du 4 janvier 2002* |
| Statut juridique | propriété de l'Etat ; musée du Louvre département des Peintures |
| Anciennes appartenances | François Ier ; Couronne de France |
| Numéro d'inventaire | INV 779 |
| Commentaires | légère diminution du tableau sur les côtés (environ 7 mm) ; acheté vraisemblablement vers 1519, après la mort de l'artiste |
| Bibliographie | HEYDENRICH 6 ; OTTINO DELLA CHIESA 31 ; VILLOT I 484 ; HAUTECOEUR 1601 ; C.S.I. 1981, P 192 |
| Copyright notice | © Musée du Louvre, © Direction des Musées de France, 1999 |
| Crédits photographiques | © Réunion des musées nationaux ; © Hervé Lewandowski ; © Thierry Le Mage |
| | commande reproduction et/ou conditions d'utilisation |
| | renseignements sur le musée |
| | 000PE025604 |

*Fig. 2. Mona Lisa at the Joconde website*

© Musée du Louvre/A. Dequier - M. Bard

Denon
1 e étage
Salle de la Joconde
Salle 6

détail de la bouche
détail des mains
détail des yeux
détail du buste

## Leonardo di ser Piero DA VINCI, dit Léonard de Vinci

Vinci, 1452 - Amboise, 1519

***Portrait de Lisa Gherardini*, épouse de Francesco del Giocondo, dite *Monna Lisa*, la *Gioconda* ou la *Joconde***

**Vers 1503 - 1506**
Peint à Florence
Bois (peuplier)
H. : 0,77 m. ; L. : 0,53 m.

**Acquis par François Ier en 1518**
**Département des Peintures**

**INV. 779**

*Fig. 3. Mona Lisa at the Louvre website*


*Fig. 4. Mona Lisa - 2000 at the Deutsche Fotothek website*

# 4 Walking through aggregations and proxies in EDM

## *4.1 Representing provided objects and their digital representations as aggregations*

EDM considers two basic classes of resources provided to Europeana:

- the "provided object" itself (a painting, a movie, a music score, a book…) and
- a (set of) accessible digital representation(s) of this object, some of which will be used as previews (e.g., a thumbnail of a painting's digital picture).

This allows capturing the distinction between "works", which are expected to be the focus of users' interest, and their digital representations, which are the elements manipulated in information systems like Europeana.

Following the ORE approach, EDM considers that the provided object, together with the digital representations that are contributed by one provider, form an *aggregation*.[8] This aggregation is the result of this provider's activity and is represented using the `ore:Aggregation` class.

In the Europeana information space, each instance of `ore:Aggregation` is related to:

- one resource that stands for the provided object, using the `ens:aggregatedCHO` property;
- one or more resources that are digital representations of the provided object, using the `ens:hasView` property.

[8] http://www.openarchives.org/ore/1.0/datamodel#Aggregation

Note that both `ens:aggregatedCHO` and `ens:hasView` are sub-properties[9] of `ore:aggregates`, representing the fact that the aggregation indeed aggregates the "real" object and its digital views.
As an example, Fig. 5 shows an EDM representation of the Mona Lisa painting, as described in the Joconde database.[10] We see that Joconde, which is maintained by the Direction des musées de France, provides an aggregation that consists of one "real" object, modelled by the provider using a `Painting` class (from an internal vocabulary or a domain-specific standard, like AAT), and two digital views. These views are declared as instances of the class `ens:WebResource`, as they are digital resources made available over the Web, and connected to the aggregation using the `ens:hasView` property.

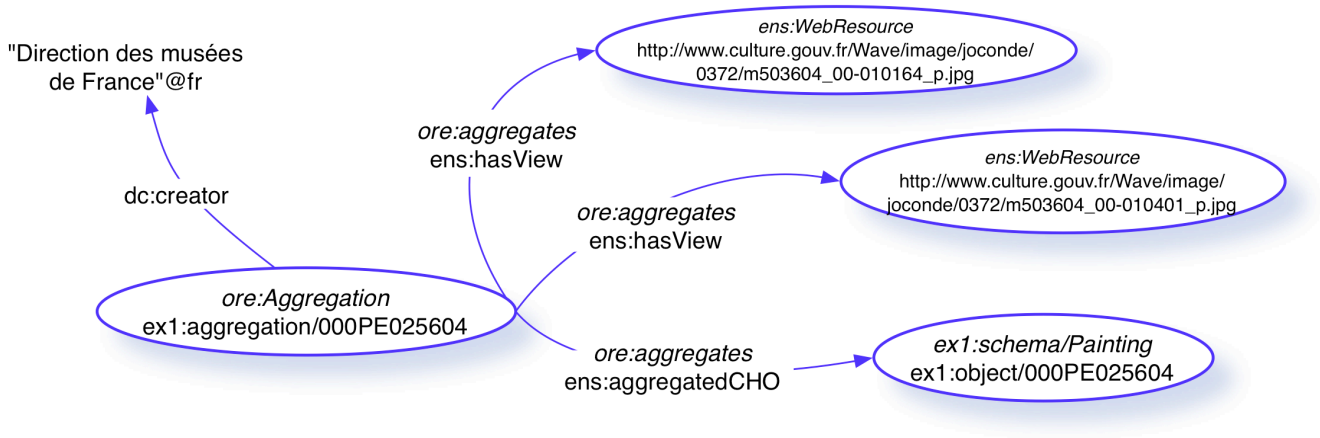



*Fig. 5. Provider's aggregation and provided object*

Note that other sub-properties of `ore:aggregates` may be used to relate the aggregation to other resources. EDM in fact introduces one such property itself, `ens:landingPage`, which can be used to link an aggregation to some reference web page for the object.

Note also that descriptive metadata can be represented for the provided object, e.g., the creator. To represent such descriptions, EDM uses dedicated properties that it either introduces or re-uses, such as `ens:hasMet`, `dcterms:creator` or `dcterms:title`. It also allows use of specializations of these properties, or any other property that providers judge relevant for describing the characteristics of the object. Section 5.3 further details those description options.
The reader should however be aware that the statements describing the provided object are not directly attached to the resource standing for the object. This will be further explained in the following section.

The reader should finally be aware that though there will very often be a one-to-one relationship between an aggregation, a provided object and a metadata record in the original provider's information system, there is no rule enforcing it. In fact, there are situations where a record can give rise to several aggregations, as in the case of records describing complex, hierarchical digital aggregations (see Section 6).

[9] For example, for each `ens:hasView` statement between an aggregation and a digital representation, an `ore:aggregates` statement can thus also be inferred between these resources.
[10] For the sake of readability, we focus on a relevant subset of the whole data contributed by the provider. A complete EDM representation would include more descriptive information as well as more digital resources linked to the aggregation. Please also note that the `ex1:` namespace is a toy namespace, not intended to represent any recommendation on what the actual URIs should be.

## *4.2 Why proxies?*

Requirement R3 raises the need for handling cases where Europeana takes data from many providers and this data may be about the same real world resource, thus giving multiple views on the same resource. In addition, Europeana will attempt to add its own data about that resource giving yet another view on the same resource.

These views will not be merged however. In such cases, it is indeed very likely that the metadata will differ, e.g., different names may be used for the same creator. So mechanisms are needed to keep the different views distinct. To this end, Europeana leverages the *proxy*[11] mechanism from the Object Re-use and Exchange (ORE) model, which is meant to enable the representation of different views on the same resource.

Let us consider our Mona Lisa example. We have two records available for it, respectively from the Joconde database and the Louvre. As represented in Fig. 7 each data submission to Europeana will give rise to a specific instance of the ore:Aggregation class, used to group all the elements related to one resource that come from one provider. Both providers indeed contribute a different set of digital representations, e.g., different resolutions, different file types and, of course, different locations for the representations. An aggregation can be seen as one provider's contribution for an object.

But each metadata record provided to Europeana also gives raise to a specific *proxy* for the object described, modelled using the `ore:Proxy` resource. This proxy is specific to a given provider, and is used to represent the description of the provided object, *as seen from the perspective of that specific provider*. With proxies it is possible to represent different, possibly conflicting pieces information on provided objects, while still keeping track of the provenance of this information. For instance, the title of Mona Lisa for Joconde could be "Portrait de Mona Lisa" while for Louvre it could be "Portrait de Lisa Ghirardini."[12]

A proxy is connected to the resource it is a proxy of using the `ore:proxyFor` property. It is connected to its provider's aggregation using the `ore:proxyIn` property, as in Fig. 6.

[11] http://www.openarchives.org/ore/1.0/datamodel#Proxy

[12] Note that in our examples we use the `dcterms:` namespace whenever it features a property with appropriate semantics and for which there is no constraint that conflicts with our intended usage. Otherwise, we use the "corresponding" property from the legacy `dc:` namespace. For example, `dcterms:creator` does not fit well cases where the creator is given as a mere string, and not as a fully-fledge resource. For such cases we use `dc:creator`.


*Fig. 6. Provider's aggregation, provided object and proxy—simple case with only one provider for the object*

Note that an aggregation can have only one proxy per provided object that it aggregates, since it results from the activity of only one provider. Where two providers have each generated a proxy for the same "real" object both proxies would ideally be linked to a resource that represents that object independently of either description context, using the `ore:proxyFor` property. The birth of this resource can be seen in Fig 6 as the object of the `ore:proxyFor` property and this is expanded upon in the next figure. Proxies are thus essential for representing and relating the different views of the same resource from various providers, including Europeana itself, as will be shown in the next section.

At this stage, how and when identical objects should be recognized remains open. Fig. 7 reflects that two providers may well have contributed two different URIs for the same resource. In such case, some identification mechanism has to be applied to infer an `owl:sameAs` link between the two URIs, which enables the "merging" of the resource. In the Mona Lisa case, the Louvre inventory number (INV 779) may be exploited for this.

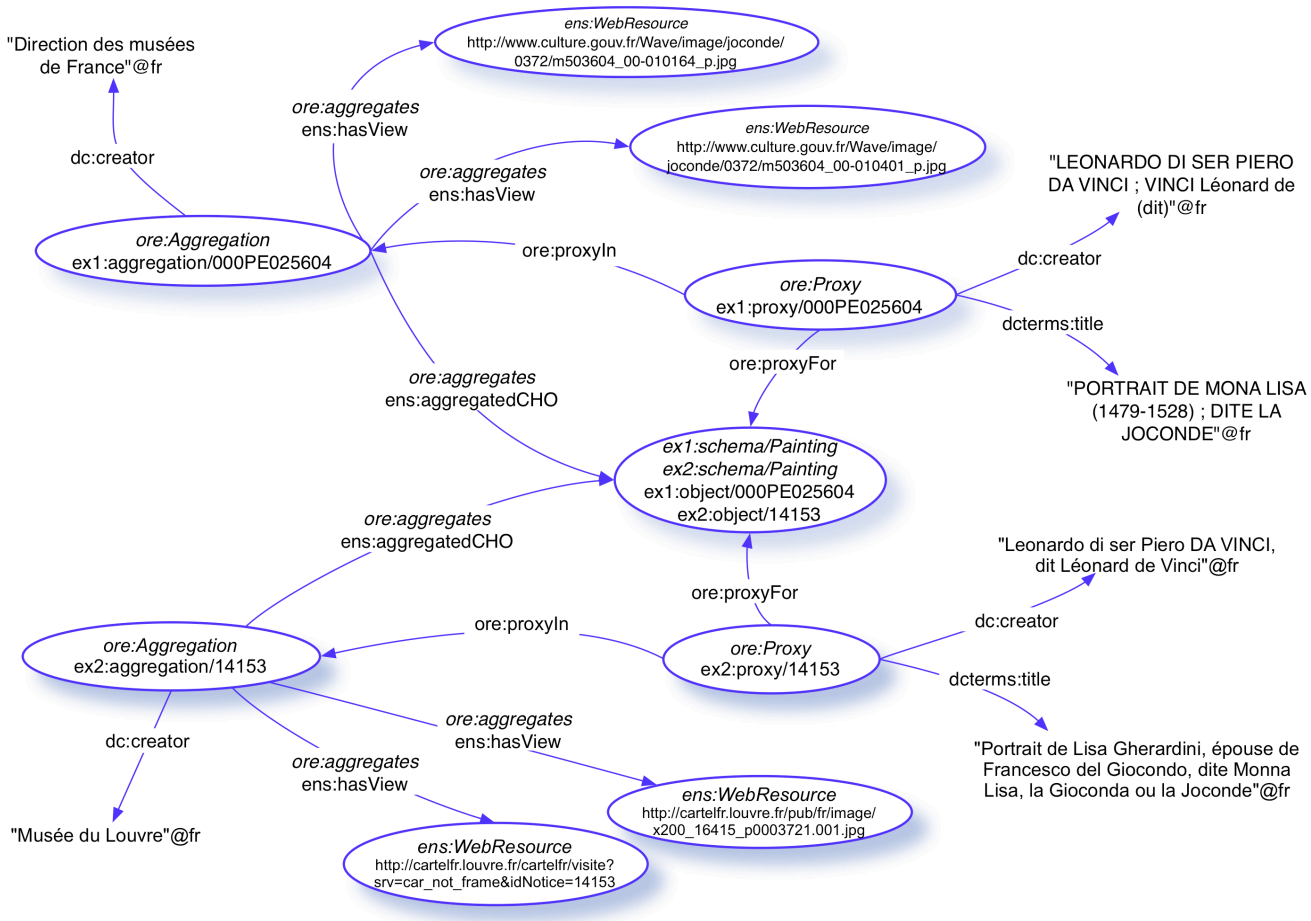



*Fig. 7. Providers' aggregations, provided object and proxies—complex case with two providers for the object*

One can expect cases where two providers submit data on a same object to be relatively numerous, once Europeana receives data from a complex network of providers. Moreover, such cases are very difficult to anticipate: Europeana aggregators cannot readily know whether the providers they aggregate data for are already providing data through another aggregator. Additionally, there is always a second information source on the provided object beyond its original provider: Europeana itself.

## *4.3 Europeana proxies and data enrichment*

Europeana creates new data for the object it ingests so as to provide more value to its users. At the time of writing, this information results mostly from the process of *normalizing* data formatted using the Europeana Semantic Elements (ESE). Europeana massages some of the metadata fields, so that they can be used seamlessly for specific purposes.

In the future, Europeana foresees more such information creation, notably by *semantic enrichment*. Objects ingested by Europeana often use simple strings as values for the metadata field. Europeana hopes to update that information by linking objects to fully-fledged resources that are thoroughly described and are themselves connected to other resources, such as authority files for persons and thesauri for subjects. These resources enable richer functions, such as query expansion (e.g., using alternatives for a creator's name),

recommendation of objects using semantic relations between them (objects created by connected artists), etc.

Let us consider the Mona Lisa example again. The Joconde database provides for this painting one single string value as the creator: "`Leonardo di ser Piero da Vinci ; Vinci, dit Léonard de (dit)`". This is valuable, but there is no direct way to know for instance that an artist who was born in Italy and who died in France created this painting. This can be enhanced, at the Europeanea level, by creating an explicit link between Mona Lisa and a carefully curated resource that stands for Leonardo as a person, and provides much more information about him: the VIAF authority record for Leonardo, identified by `http://viaf.org/viaf/24604287`. This is a crucial aspect, and Europeana intends to proceed with such semantic enrichment on a large scale, using classes that EDM introduces for this specific purpose (see section 5.2).

The bottom part of Fig. 8 shows how the result of such enrichment can be represented using EDM. This is shown by the presence of the `ens:Agent` resource attached to the Europeana proxy. Thanks to the proxy mechanism, Europeana can maintain the original metadata alongside the new, richer metadata that it generates, allowing it to serve (or display) one or the other, depending on a given information need. Note that the `ore:aggregates` link between the two aggregations will be explained in the next section.

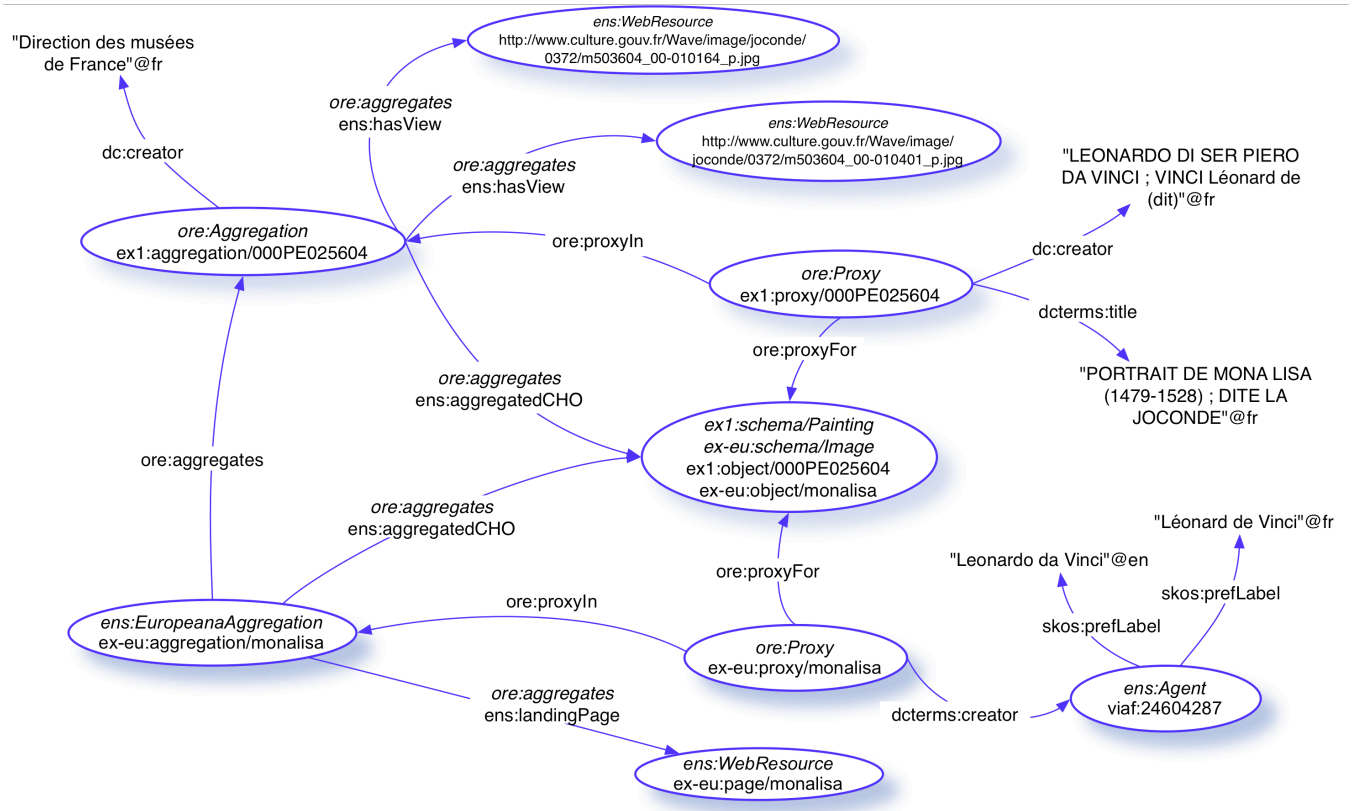



*Fig. 8. Europeana aggregation—simple case with only one provider for the object*

## 4.4 Europeana aggregations and proxies

As shown in Fig. 8, Europeana creates its own aggregation and proxy for the provided object. This enables the connection of new information (derived from normalization or

enrichment) to the original object description, while still keeping the distinction between what is provided and what is added.

This new Europeana aggregation is modelled using `ens:EuropeanaAggregation`, a specific subclass of `ore:Aggregation`. It captures the fact that such aggregations are the result of Europeana's own work. Europeana can use them to manage its own IPR, access restrictions, and so on. They also hint that an extra aggregation layer is being introduced here.

Like providers' aggregations, a Europeana aggregation is indeed currently linked to the provided object using `ore:aggregates`. It can also aggregate other resources, especially digital representations of the object, or a reference landing page for it, using the `ens:landingPage` property. One crucial point, though, is that in EDM the Europeana aggregation is considered to aggregate each specific provider's aggregation that is about the same object. Fig. 9 extends the example of Fig. 8 by introducing a more complete version of this extra aggregation layer, which enables the introduction of new descriptive information via the Europeana proxy but also reflects Europeana's crucial role of bundling (digital) resources together.

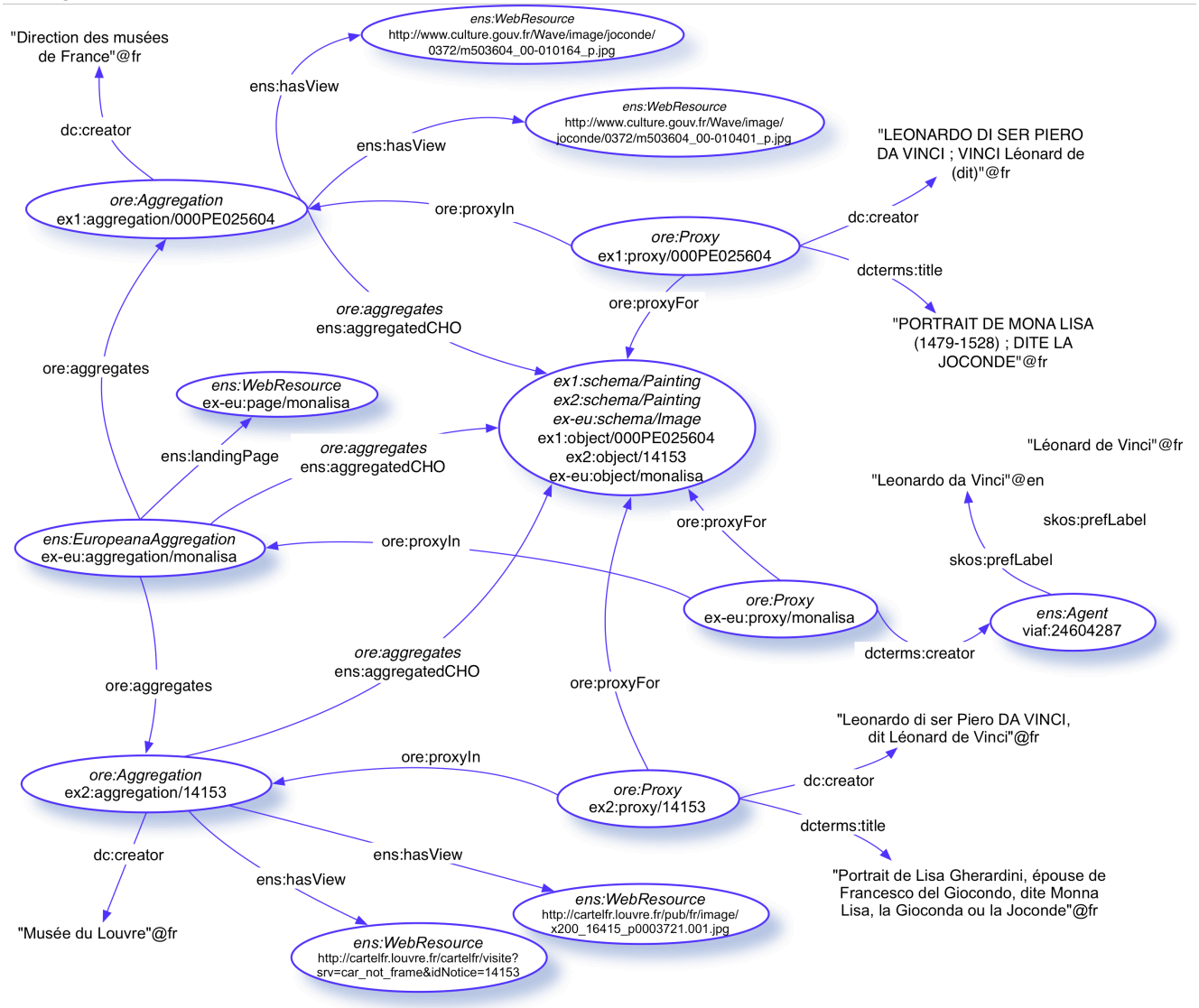



*Fig. 9. Europeana aggregation—complex case with two providers for the object*

## 4.5 Why manage central nodes for provided objects?

A question the reader will possibly ask is: why would Europeana care about the "central" resource that denotes the provided object? After all, the provider-specific proxies seem to be enough to carry the descriptive metadata. And the Europeana aggregations are enough to glue together all digital representation and data contributed by various providers, including Europeana itself.

The first answer simply derives from technical motivations: according to the ORE model, an ORE proxy has to be *proxy for* some "view- independent" resource that is aggregated by the aggregation. But this technical constraint is dictated by more strategic data access considerations in ORE, which also apply in the Europeana context. Consider a user needs to access information for a provided object, say, the Mona Lisa painting. It is very likely that this user cannot anticipate which are the specific views that apply to it, or even if there are any such views. In most cases, the information needs of this user won't even require to access a specific view. These needs will rather be focused on "real" objects—in Europeana, a painting, a book, etc. In our example, users are expected to search first for information and objects for *the* Mona Lisa, not for one specific view on Mona Lisa.

This has important consequences in Linked Data scenarios, where data consumers will expect to access information via the identifiers of real objects. Proxies cannot act as independent access points, i.e., without prior knowledge of the resource they stand for. Europeana thus needs to manage resources for the real objects that are at the core of its users' interests—as well as of its providers' business.

## 4.6 What parts of the EDM core pattern should providers provide?

The pattern presented in the previous sections is quite complex, compared to the existing ESE practice. In particular, it is clear that this complexity arises from requirements that are not shared by all Europeana data providers. The question is thus whether Europeana could accept, for its data ingestion process, simpler input than the complex data exemplified in this document, e.g., in Fig. 9, and take care of the rest itself.

This question will certainly not be definitively answered until a first prototyping phase for EDM has been carried out. This document tries however to discuss now some aspects that might not be so relevant for providers to focus on.

Consider for instance the case of proxies. These entities are introduced to represent the point of view of a specific actor (provider, user, Europeana) on an object. This corresponds to a specific need: the one for distinguishing between different sources of information, when several such sources concern the same object.

First, it is obvious that providers should not be concerned by the creation of Europeana proxies: they can't be asked to anticipate what Europeana will create for the object.
Second, it is quite clear that providers should not create proxies, when they don't have themselves to organize information coming from different perspectives on the same object. In that case, they may attach information directly to the resource that stands for the submitted object, as in the simplified version of Fig. 6 shown in Fig. 10. One of the most important pieces of information expected from providers is the distinction between the metadata that

applies to the object itself, and the metadata that applies to the digital representations (and the package that holds them together). Proxies are not strictly necessary for this.

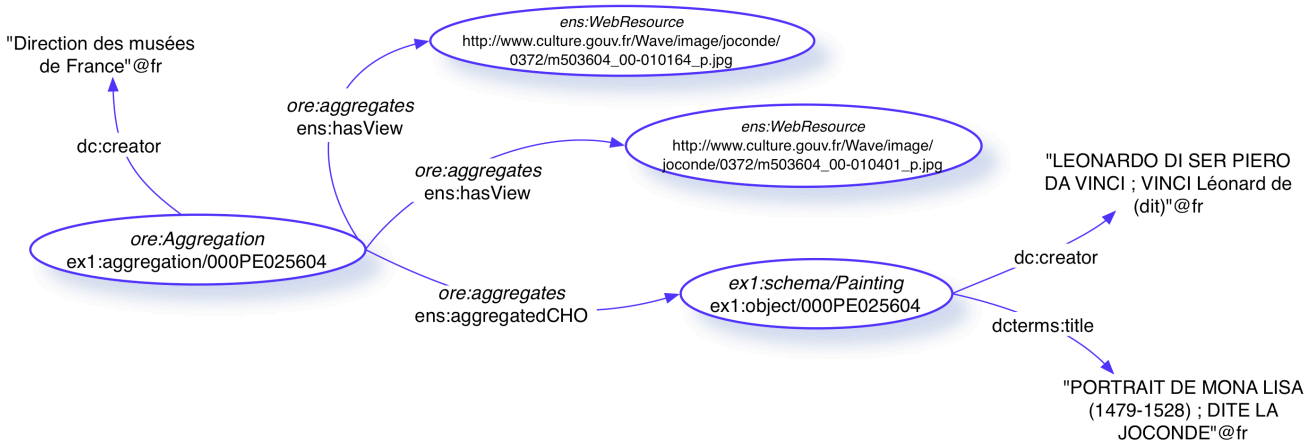



*Fig. 10. Simple data submission to Europeana, without proxies*

In fact, the submission of proxy-based representation would be expected in only two cases:

- for Europeana aggregators who already own several records pointing to the same thing.
- for providers that want to link their data submission to objects already ingested in Europeana or curated by other institutions, when they know they have records about these objects as well, and wish to help Europeana to detect this.

The next issue regards the provision of (URI) identifiers for the various objects that appear in the pattern. In the previous sections, we assumed that all resources have been provided (HTTP) URIs. Providers may not be expected to provide all this, though.

A first suggestion is that providers will submit URIs for web-accessible digital representations (e.g., pictures) and for the provided objects or aggregations that already have permanent identifiers. Europeana itself would take care of assigning (or re-assigning) URIs for the proxies it creates and for the aggregations that don't have URIs yet. It will also create URIs for all objects, so as to implement a linked data publication strategy that relies on Europeana's own (HTTP) services.

Note that not all URIs submitted by providers need to be HTTP URIs, In fact, only the ones minted by Europeana will be.

# 5 Descriptive metadata in EDM

Aggregations and proxies enable capturing a description of the "digital environment" of an object submitted to Europeana, and attaching descriptive information to the various resources that take part in this environment. This mechanism remains however agnostic with regard to *which* descriptive data that should be provided. EDM therefore includes a set of "descriptive" and "contextual" properties that capture the different features of a resource, as well as relate it to the other entities in its context.

## 5.1 Object-centric vs. event-centric metadata: two "descriptive cores" in EDM

Among the possible approaches for metadata, one can distinguish *"object-centric"* and *"event-centric"* approaches. The former focus on the object described: information comes in the form of statements that provide a direct linking between the described object and its features—be they simple strings or more complex resources denoting entities from the real world. Most metadata practices making use of the Dublin Core metadata set [DC] can be seen as an application of such an approach. The records corresponding to the objects in Fig. 6, 7, 8 and 9 correspond to an object-centric approach, too, insofar as they directly relate the provided object to all its features.
Event-centric approaches, on the other hand, consider that descriptions of objects should focus on characterizing the various events in which objects have been involved. The idea is that it will lead to establishing richer networks of entities—by representing the events that constitute an object's history—than with the object-centric approach. This approach underlies models such CIDOC-CRM.[13] A typical example of event-centred description, which shows how different places and actors can be unambiguously related to one object via the events these entities participated in, can be found in Annex 1.

EDM provides constructs that allow representing metadata that follow both approaches. Fig. 11 extends the graph of Fig. 6 to provide a more complete example of an object-centric description for the painting in Fig. 2. In this example, the Mona Lisa is directly attached to its creator (represented by a simple string), its title, its creation date(s), its former owner, etc.

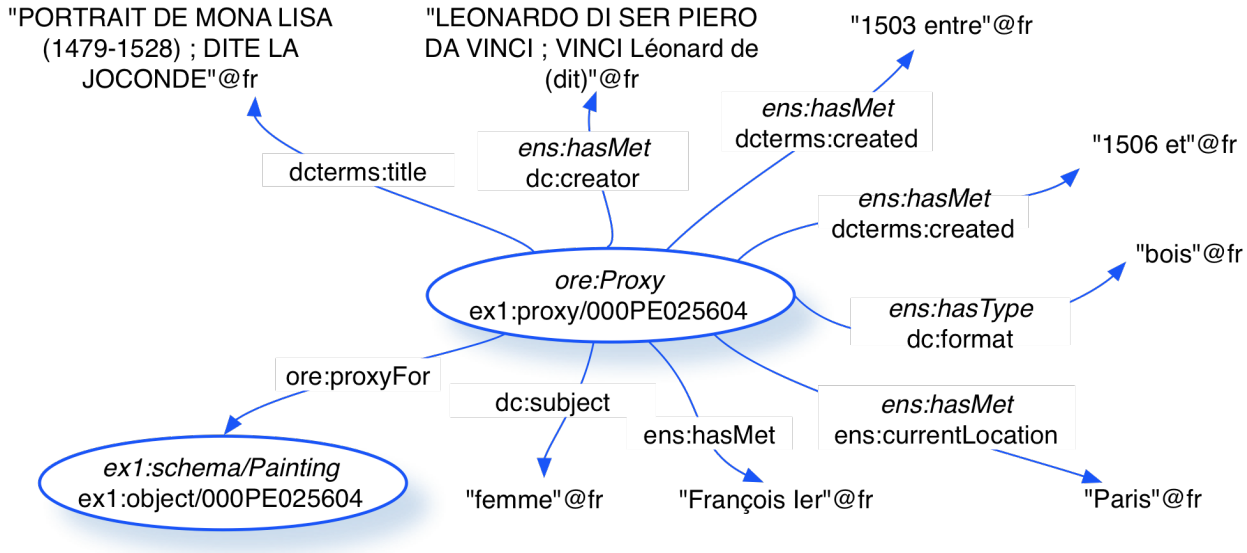



*Fig. 11. Mona Lisa – an object-centric description*

The property `ens:isRelatedTo` and its specializations `ens:hasMet` and `ens:hasType` are especially important here: they provide anchors via which more specialized properties can be connected to the core EDM model—an issue further discussed in section 5.3. `ens:isRelatedTo` can be used to link an object to virtually any entity that belongs to its "context": agents involved in its life cycle, places it has been associated with, subjects it is about, etc. `ens:hasMet` is used to relate more precisely a given object to the various things (persons, places, etc.) that have participated to the same events as this object. For example,

[13] http://cidoc.ics.forth.gr/

the creator of an object is an agent that participated in the creation event of that object. Note that the current location of an object can be expressed using the specific `ens:currentLocation` property, which is a sub-property of `ens:hasMet`. `ens:hasType` connects an object to a concept from a type system to which that object belongs—excluding "aboutness" links, in particular.

Note that the object-centric approach does not dictate one specific level of "semantic richness" for the resources attached to objects or events. An enrichment of Fig. 11 by replacing a number of strings by instances of `ens:Place`, `ens:Agent`, `skos:Concept` or `ens:TimeSpan` (see next section) would still fall in the object-centric category.

Fig. 12 provides an example of how the same object could be represented in an event-centric fashion.[14] Two new events related to the object's lifecycle—`000PE025604-c` denoting the creation of the painting and `000PE025604-c` denoting its acquisition—have been introduced, using the class `ens:Event`. These events are now the "hubs" that relate the object to other entities that were directly connected to it in the previous object-centric approach. These relations are represented in EDM using the three following properties:

- `ens:wasPresentAt`, holding between any resource and an event it is involved in;
- `ens:happenedAt`, holding between an event and a place;
- `ens:occurredAt`, holding between events and the time spans during which they occurred.

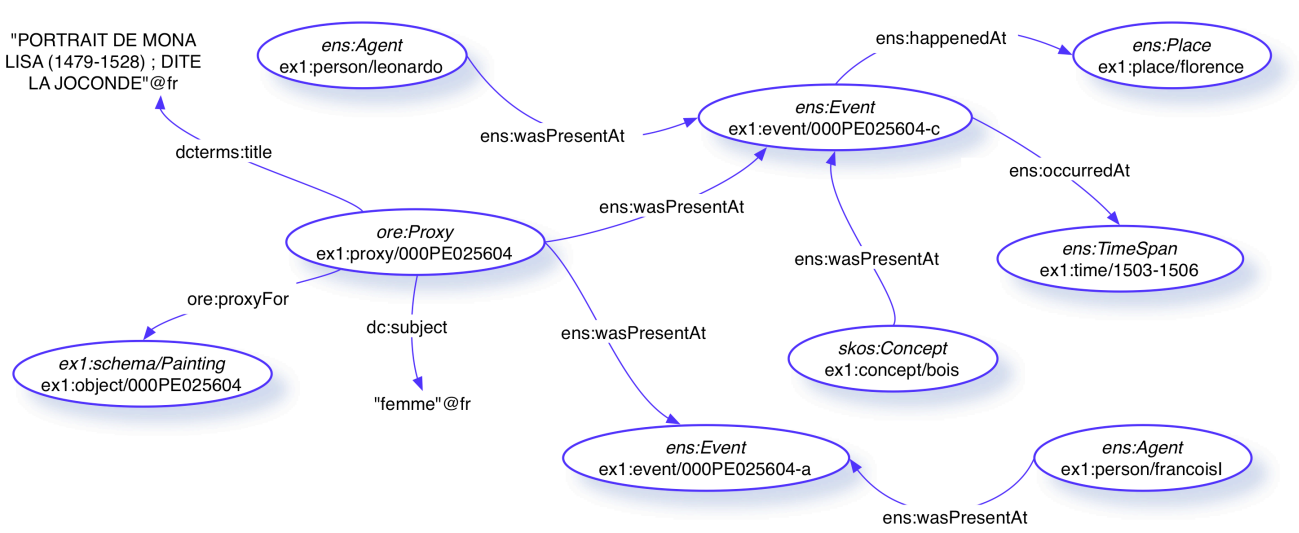



*Fig. 12. Mona Lisa – an event-centric description*

Fig. 13 shows how this approach to using EDM can be used for representing the more complex, real example of Annex 1. While Fig. 12 is a straightforward adaptation of a simple example, Fig. 13 hints that more can be achieved with event-based representations. This is especially true when these events are related together, e.g., through happening in a same location, or if events help distinguishing between the different contexts, e.g., dates, that can be related to one same object.

[14] For the sake of simplicity, we omit a number of statements that should apply to resources attached to the event, such as the link between `ex1:person/francoisI` and the string `"François Ier"`.

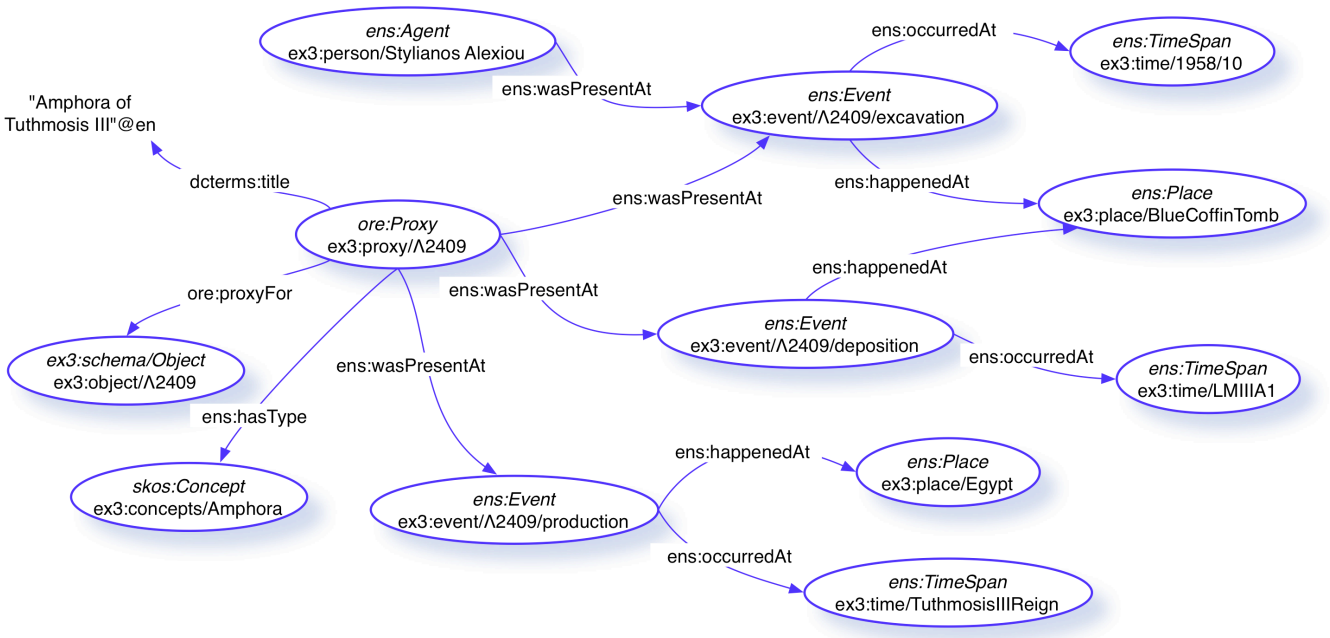



*Fig. 13. Amphora of Tuthmosis III – an event-centric description (without aggregation and related digital representation entities)*

The reader should be aware that EDM perfectly allows both object-centric and event-centric approaches to co-exist seamlessly for the same object. As a matter of fact, taking benefit of the RDF approach, EDM allows any kind of network to be attached to a provided object, be it event-centric, person-centric, place-centric, etc.

Note finally that it is not the aim of EDM to capture the full complexity of a model like CIDOC-CRM. Nor can it capture the full diversity of all object-centric models. Rather, it provides a small set of properties and classes to which more specialized constructs can be "attached," following the approach discussed in the next section.
One can however notice that the "core" corresponding to the event-centric approach (`Event`, `happenedAt`, `occurredAt` and `wasPresentAt`) is much less developed than the object-centric "core", which is based on Dublin Core elements [DC] as detailed in the EDM specification [EDM-Definition].

This basically boils down to two reasons. First, the object-centric approach is much more widespread. Second, there is a simple, commonly used standard for object-centric approach—Dublin Core. This standard can be re-used almost out-of-the-box without forcing providers to adopt a whole new conceptual framework.
In fact, while event-centric descriptions will be ingested and exploited as much as possible, it is likely that Europeana will still request the submission of a basic, object-centric core next to it. This will allow full compatibility of the new model with the legacy ESE data, and supporting consistent and coherent indexing for elementary search functionalities at a relatively low cost and risk.

Yet, by introducing basic compatibility between EDM and event-centric representations now, we hope to accommodate initiatives such as CIDOC-CRM and LIDO [LIDO], which are aimed at making descriptions of events more interoperable and more widespread. If a simple "event-centric core" gets widely used by Europeana providers and gives clear added value over the current simple core, it may be considered as a refinement to be included in a next version of EDM. Future versions of this document will include corresponding examples.

## 5.2 EDM and enriched descriptive metadata

The previous section uses in Fig. 11 a number of classes that can be used to represent the entities referred to in metadata fields as fully-fledged resources, that are potentially subjected to richer descriptions themselves.

Section 4.3 has introduced a first example, where Leonardo da Vinci is no longer represented via a simple string, but as a resource to which several name variants can been attached. This example could be further extended by adding to it birth and death dates, places where he lived, links to his masters, etc.

Such semantic enrichment can bring huge benefits to current search processes. Europeana intends to proceed with it at a large scale. This is the reason for EDM featuring a number of classes devoted to the representation of "contextual" entities:

- `ens:Agent`, to be used for representing persons or organizations
- `ens:Place`, for spatial entities
- `ens:TimeSpan`, for time periods or dates
- `skos:Concept`, for all entities from knowledge organization systems like thesauri, classification schemes (including some place gazetteers or person authority files)…

These classes can be used to produce enriched descriptions, especially when controlled vocabularies curated by providers or other organizations are available for them. The example of Fig. 11 could lead to a new representation, as in Fig. 14 below.

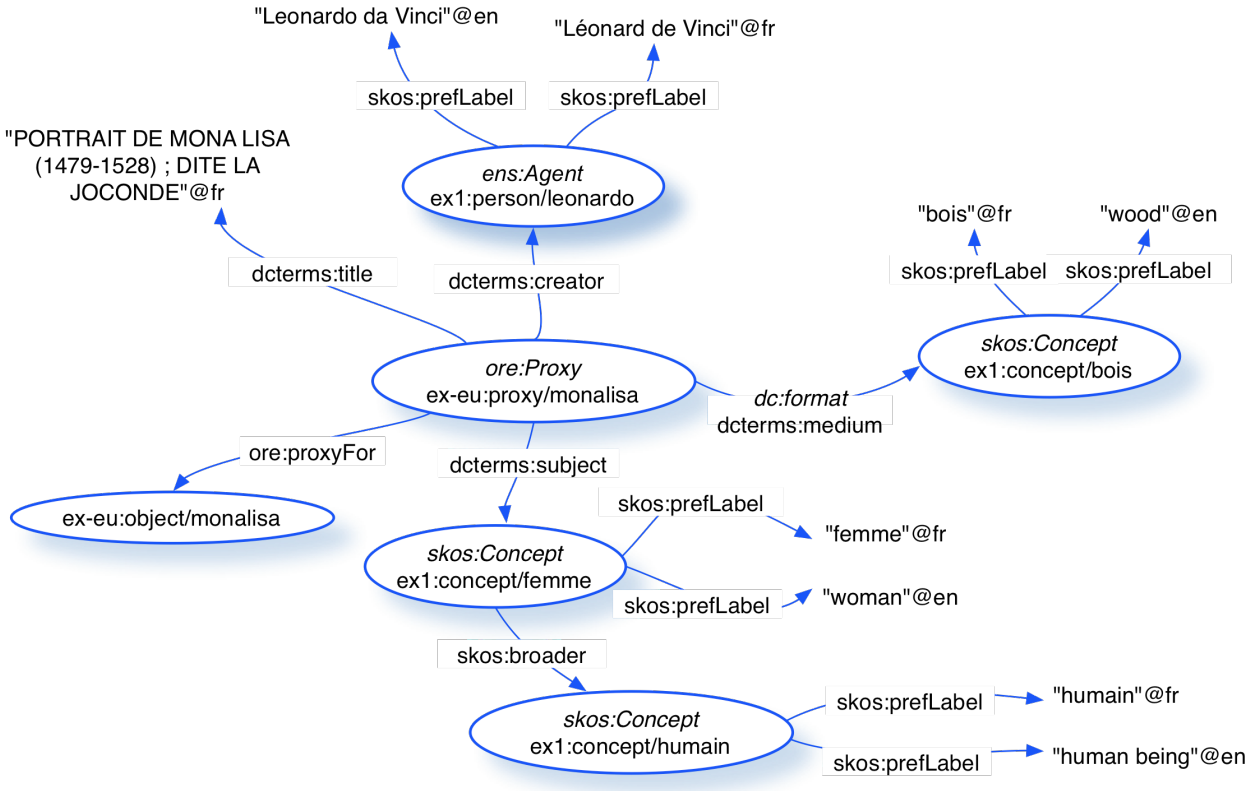



*Fig. 14. Mona Lisa – an enriched description*

Note here the various statements attached to the new resource (more precisely, the new Europeana proxy for the object described), including the links from the specific concept of

*woman* to the more general one of *human being*. Note also that once they are represented as full-fledge resources, the entities linked to the objects can be themselves connected to other entities, for instance as a result of *semantic alignment*. For instance, `ex1:concept/femme` could be matched to `http://dbpedia.org/resource/Femme`, enabling to use (possibly in a linked-data fashion) all the information available on Wikipedia for this specific subject, including a quite rich description of the topic and translations for the term as well as that description.
These crucial features allow bringing in more information to enhance access to the original objects, They can also enable a complete change of paradigm in the way these objects are accessed, by allowing the user to browse through a semantic space of contextual entities before getting to the actual objects.

## *5.3 EDM as a flexible data model*

As presented in the previous sections, EDM provides a number of constructs (classes and properties) that can be used by providers when submitting metadata to Europeana. ***It is however expected that in most cases these constructs will be used indirectly, via RDF assertions using more specialized constructs.***

EDM is indeed aimed at providing a much more flexible description framework than the existing ESE. We expect that providers, while submitting data to Europeana, will submit descriptions that fit their own specific level of interest. The key to ensure interoperability at the semantic level is *mappings*, following common practice in the Semantic Web framework.

Let's consider an example. Joconde provides an historical note for the Mona Lisa painting. A straightforward representation of it in RDF could yield the following statements:

```
ex1:proxy/000PE025604 ex1:schema/historicalNote "commandé par le florentin Francesco del Giocondo, époux de Mona Lisa entre 1503 et 1506"@fr .
```

As such, however, this information cannot be fully exploited in an environment like Europeana, which has to deal with hundreds such specific schemas. Mapping to a semantic interoperability core is required to ensure that a general tool can exploit at least a part of the intended semantics for these specific properties.
Such mappings are typically achieved in RDF by asserting semantic relationships between the specific constructs and the core ones. Those can take the form of statements using `rdfs:subClassOf` or `rdfs:subPropertyOf`, as in the following:

```
ex1:schema/historicalNote rdfs:subPropertyOf dcterms:description .
```

If the provider submitting the above Mona Lisa metadata also provides this mapping, then a tool able to exploit the mapping will be able to derive from the original description a new generalized statement:

```
ex1:proxy/000PE025604 dcterms:description "commandé par le florentin Francesco del Giocondo, époux de Mona Lisa entre 1503 et 1506 "an alternative title"@fr .
```

This co-existence between the generic and the specific level allows for example:

- to search for the painting using a generic description-based index

- to display the information for that painting using the finer-grained distinctions made by the provider.

Note that this mechanism is already at play within the various levels of descriptive data of EDM. `ens:hasMet`, for instance, is meant as a super-property of various other properties re-used in EDM, such as `dcterms:creator`, `dcterms:contributor` and `dcterms:publisher`. This property will thus allow users to find the objects that are related to a given person, whether they have "met" this person as their main creator, a secondary contributor, or a publisher.

To sum up, using the full potential of EDM requires providers to provide descriptive data according to their most specific interest, but also at the more general interoperability level EDM defines. The Mona Lisa metadata as expressed in the object-centric view of Fig. 11 should thus be submitted in the more complete form expressed in Fig. 15, following the original metadata presented at the Joconde site. Fig. 16 shows how a more precise description could also be provided for the event-centric view of Fig. 12.

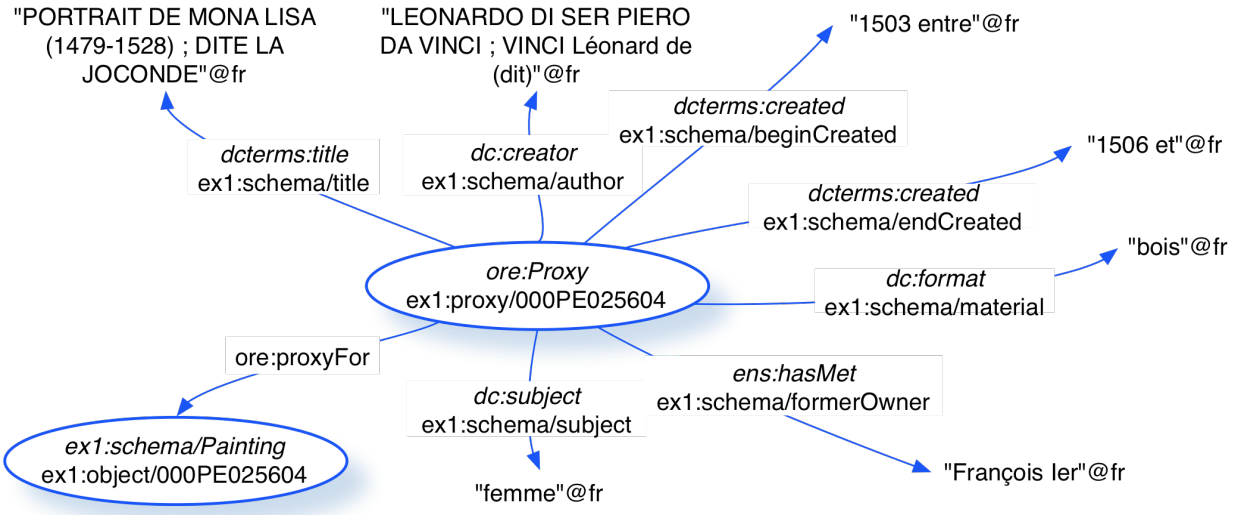



*Fig. 15. Mona Lisa – object-centric description at both specific and interoperability levels*

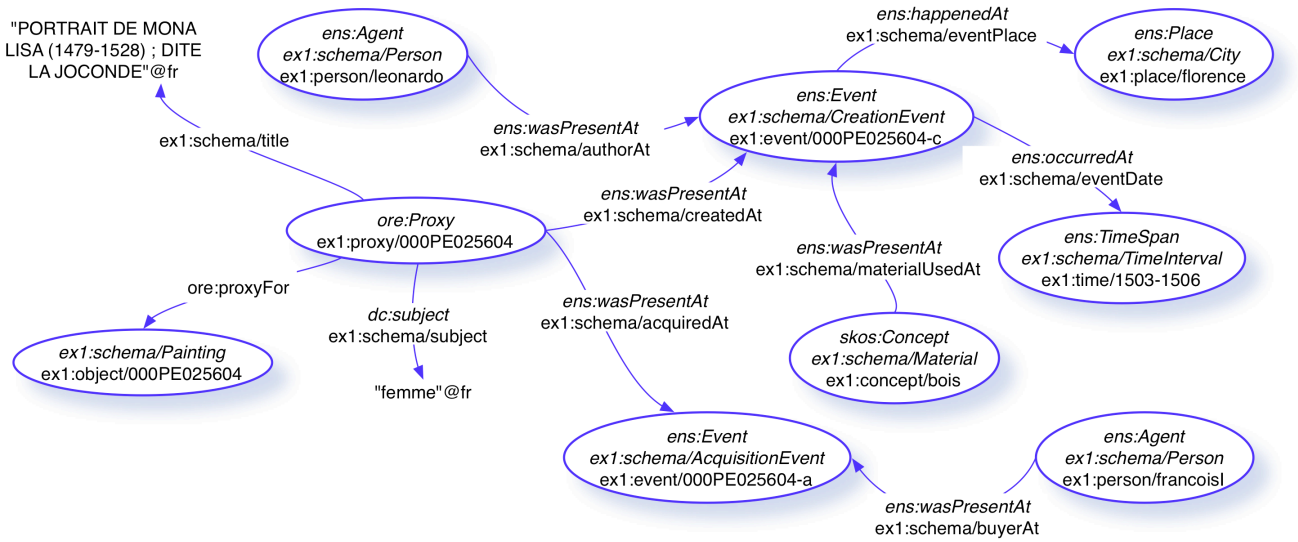



*Fig. 16. Mona Lisa – event-centric description at both specific and interoperability levels*

The reader should notice that at the time of writing, the practical details on how to organize the submission of precise metadata together with its mappings are still to be worked out. This will be investigated during the EDM prototyping phase, as well as in the EuropeanaConnect project.

Finally, one may argue that presently some of these required mappings may bridge too wide a conceptual gap to be useful. Consider for instance the sub-property link between `ex1:schema/buyerAt` and `ens:hasMet` in Fig. 16. In that case, EDM only allows capturing a minimalistic part of the original property's meaning.

This highlights the value of "interoperability cores", such as Dublin Core, which allow to better capture the intended semantics of metadata fields in a cost-effective way.[15] As was already hinted in the previous section, it could be that a large number of providers agree on a common set of other interoperability-level constructs that provides the basis for richer functionality. Europeana will then naturally consider including that set as a "reference extension" to EDM, and implementing functions that exploit it properly.

## *5.4 Relationship between Europeana Semantic Elements (ESE) and EDM*

Most of the properties used in ESE (the Dublin Core ones) actually constitute the "semantic interoperability core" of EDM, as presented in the section of the EDM Definitions presenting the mappings between ESE properties and EDM ones.

The main difference lies in the way these properties can be used. To remain compatible with legacy data, and data that will come in "not enriched" ways, ESE/DC properties can be used with simple strings as values. But EDM recommends, for the properties that can be used that way (e.g., `dcterms:creator`, etc.) to use fully-fledged resources, as with the VIAF example first introduced in Fig. 8.

The reader should be aware, as already noted, that one should use `dcterms:` namespace whenever it features a property with appropriate semantics and for which there is no constraint that conflicts with our intended usage. Otherwise, the "corresponding" property from the legacy `dc:` namespace should be used. For example, `dcterms:creator` does not fit well cases where the creator is given as a mere string, and not as a fully-fledge resource. For such cases we use `dc:creator`.

Note that some of the elements that are currently in ESE will not appear as such in EDM for the moment. For instance, some non-Dublin Core elements generated by Europeana (e.g., `europeana:year`). Status is unclear at the time of writing.

# 6 Advanced EDM

EDM allows for even more complex representations—still in a flexible fashion, as discussed in Section 5.3. The following are especially of interest:

[15] This is of course caused by the mass of data available in Dublin Core format, which maximizes the return on implementing functions that exploit this data.

- hierarchies of aggregations, grouping digital resources together;
- relations between provided objects, for instance artistic derivation between works;
- explicit representation of data packages via ORE resource maps.

In this document we only present examples of the hierarchical objects. Other aspects will be detailed in a coming version.

## *6.1 Representing hierarchical objects*

To illustrate how EDM enables representing hierarchical (part-of) links between objects, we consider an example from the archive domain—a register of the archive of the counts of Holland.[16] This object can be considered as a simple one: it has some physical unity, and its content is addressing to one general subject. However, it can also be considered as a grouping of individual pages, each of them being digitized and potentially answering a user's information need, for instance, a request for information over a specific town. This complex situation is appropriately described in the metadata Europeana aims at harvesting. These archive objects are indeed described in EAD files, which can represent hierarchical containment between different "levels" of archive resources.

EDM aims at tackling this sort of relation, by allowing one to use:

- `dcterms:hasPart` and `dcterms:isPartOf` for representing inclusion links between the represented objects;
- `ore:aggregates` links between different `ore:Aggregation`s that correspond to these represented objects, for representing that some groupings of digital resources about one cultural object can be decomposed into smaller sub-groupings of digital resources about parts of this object;
- `ens:isNextInSequence` to express order among the parts of the object, when such ordering is applicable.

A possible representation of the first two pages of the above example, using only the more general properties EDM recommends and a subset of available metadata,[17] is shown in Fig. 17.

Note that the proxy mechanism allows several hierarchical views to be deployed on the same objects. One book may be viewed as a set of component pages for one provider, while it would remain one simple entity for another, or even be decomposed in a different way. This will be especially useful when Europeana has to aggregate such different views, producing new hierarchies without messing up the original ones.

[16] The object was provided by the APEnet project (http://apenet.nac.kei.pl/) and is accessible from the site of the Dutch National Archives at http://beeldbank.nationaalarchief.nl/na:col1:dat515829. Individual pages are also available at Europeana, for example http://www.europeana.eu/portal/record/92105/D6BF76BA92CA1FC0C5BA6B595AB328D09291022D.html .

[17] A more complete version is available through the semantic layer at http://semanticweb.cs.vu.nl/europeana/browse/list_resource?r=http://purl.org/collections/apenet/proxy-3_01_01-5-5_3-2149&raw=true .

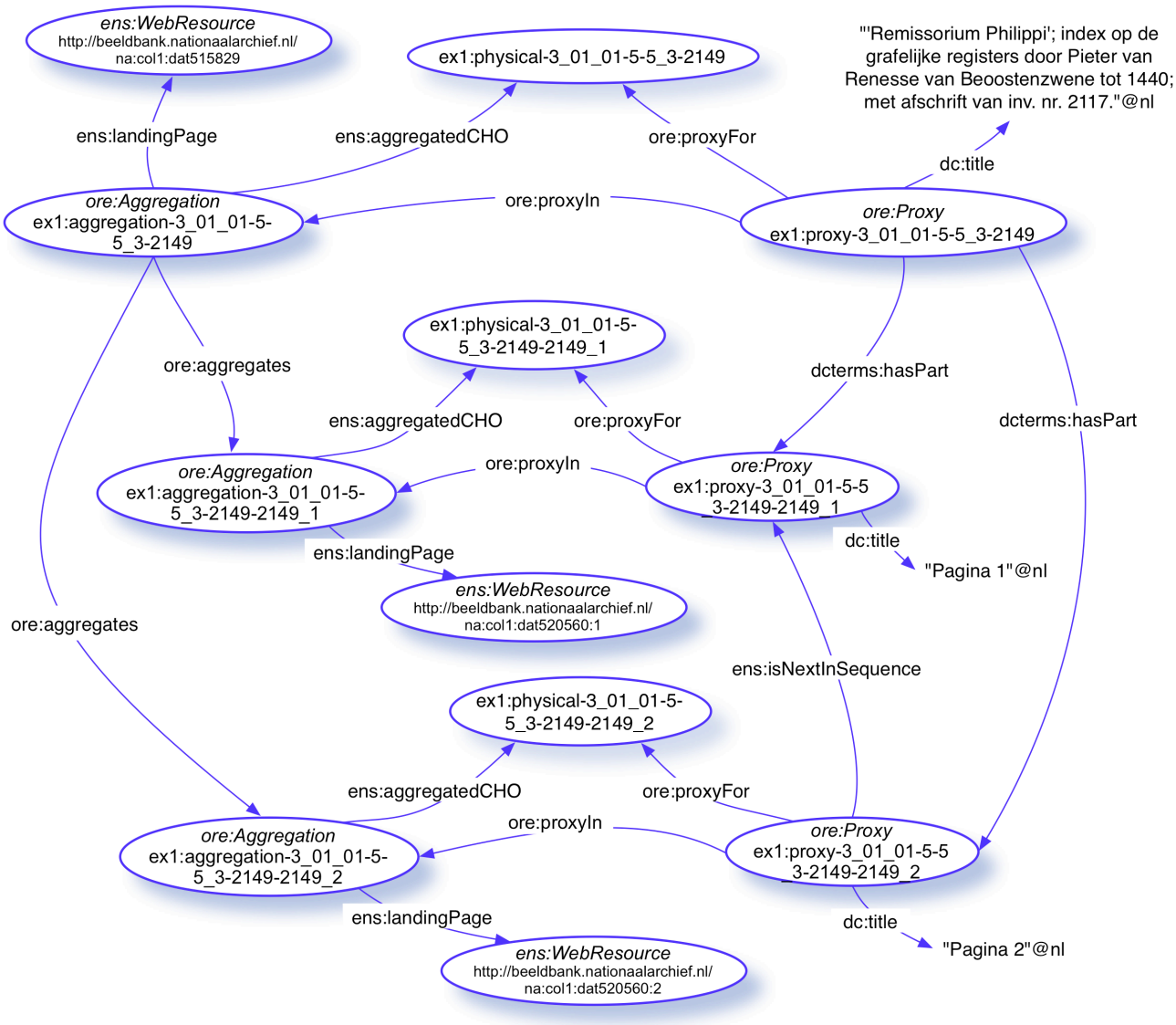



*Fig. 17. Representation of a the first components of a hierarchically structured object*

# References

[DC] *DCMI Metadata Terms*, 14 January 2008. Latest version available at http://dublincore.org/documents/dcmi-terms/ .

[EDM-Definition] *Definition of the Europeana Data Model elements version 5.2*, Europeana v1.0. Available at http://version1.europeana.eu/web/europeana-project/technicaldocuments/.

[FOAF] *FOAF Vocabulary Specification*, Dan Brickley, Libby Miller, 1 January 2010. Latest version available at http://xmlns.com/foaf/spec/.

[FRBR] *Functional Requirements for Bibliographic Records: Final Report*, IFLA Study Group on the Functional Requirements for Bibliographic Records, 1997 (amended and corrected through 2009). http://www.ifla.org/VII/s13/frbr/frbr_2008.pdf.

[LIDO] *LIDO - Lightweight Information Describing Objects*, Richard Light, Gordon McKenna, Regine Stein, Axel Vitzthum, ATHENA project deliverable D3.3, November 2009, http://www.athenaeurope.org/getFile.php?id=535

[ORE] *ORE Specification - Abstract Data Model*, Carl Lagoze, Herbert Van de Sompel, Pete Johnston, Michael Nelson, Robert Sanderson, Simeon Warner, Editors, 17 October 2008. Latest version available at http://www.openarchives.org/ore/datamodel .

[OWL] *OWL Web Ontology Language Reference*, Mike Dean, Guus Schreiber, Editors, W3C Recommendation, 10 February 2004. Latest version available at http://www.w3.org/TR/owl-ref/ .

[SKOS] *SKOS Reference*, Alistair Miles, Sean Bechhofer, Editors, W3C Recommendation, 18 August 2009. Latest version available at http://www.w3.org/TR/skos-reference .

[RDF-Primer] *RDF Primer*, Frank Manola, Eric Miller, Editors, W3C Recommendation, 10 February 2004. Latest version available at http://www.w3.org/TR/rdf-primer/ .

[RDF/XML] *RDF/XML Syntax Specification* (Revised), Dave Beckett, Editor. W3C Recommendation, 10 February 2004. Latest version available at http://www.w3.org/TR/rdf-syntax-grammar/ .

[Turtle] *Turtle - Terse RDF Triple Language*, David Beckett, Tim Berners-Lee. W3C Team Submission, 14 January 2008. Latest version available at http://www.w3.org/TeamSubmission/turtle/ .

# Change history from previous version (05/08/2010)

| Date | Change | Author |
|---|---|---|
| 10/02/2011 | Corrections following comments from Paul Hermans:<br>• changed into dc:format the occurrences of dcterms:medium with literal as object<br>• removed dcterms:created with an ens:TimeSpan as object<br>Corrections following comments by Doug Tudhope:<br>• p.11: precision on the provider-supplied class Painting<br>• p.13: precision on link between proxies for a same object<br>• p.14: added note on expectation of cases where two providers submit data on a same object.<br>• p.15: precision in the text about Fig. 8<br>• p.20: precision in the text about the two events in Fig. 12, added note on the potential benefits of event-based representation.<br>• p.22: added a note on using skos:Concept to represent person or location authority files<br>• p.23: added some extra motivation for mapping to dbPedia<br>Minor editorial modifications and updates | Antoine Isaac |

# Annex 1

## Amphora of Tuthmosis III

**Identifier:** Λ2409

**Classification:** Amphora

**Event: Type:** Excavation

**Agent:** Stylianos Alexiou

**Date:** 1951, October

**Place:** Katsampas, Tomb of the ”blue coffin”, Heraklion

**Event: Type:** Deposition

**Place:** Katsampas, Tomb of the “blue coffin”, Heraklion

**Period:** LMIII A1 (14th century BC)

**Event: Type:** Production

**Place:** Egypt

**Period:** 18th Dynasty, reign of Tuthmosis III (15th century BC)

**Current Location:** Archaeological Museum of Heraklion Crete

**Current Owner:** Archaeological Museum of Heraklion Crete

**Description:** Intact, veined, Egyptian alabaster jar. It has a piriform body, short neck, flat everted rim, foot of biconcave profile, defined by a ring with hollow underside, imitating a slightly asymmetrical base. Two vertical strap handles separate the shoulder from the top of the belly. On one side of the belly is a rectangular frame enclosing a hieroglyphic inscription with the name of Tuthmosis in two cartouches. The inscription reads:

“1.The virtuous god
2. Men-Heper-Re
3. Son of the Sun
4. Tuthmosis, the Fair One in the transformations
5. Blessed with eternal life”.

This imported Egyptian vase of the 18th Dynasty was found at Katsampas, in the tomb of the “blue coffin”, together with other Egyptian stone vessels. The name Men-Heper-Re refers to the pharaoh of the dynasty of Tuthmosis III, who reigned from about the beginning to the middle of the 15th century BC. The vase was probably imported to Crete in the years when Egypt was strongest at sea. […]