



→ EDM Case Study: Europeana Libraries and EDM for libraries



Europeana Libraries was a two years project (January 2011 to December 2012) aiming to create a robust aggregation model to make digital content from research and national libraries across Europe available on both *Europeana*¹ and *The European Library*² portal. One of the objectives of the project was to work on how library data can be aligned to the *Europeana Data Model* (EDM). Europeana Libraries addressed to Europeana some keys issues which were useful for creating a EDM domain-specific application profile. This approach of making data compliant with a RDF model is in line with similar experimentation run within the library domain.

The first task of the Europeana Libraries Metadata working group was to start the mapping of traditional library metadata records to the three principle EDM structures (edm:ProvidedCHO, edm:WebResource, ore:Aggregation). The group considered, in the first instance, monographs, multi-volume works and serials, making a specific distinction between born-digital and digitised objects.

Since EDM is an RDF-based model, the records had to be viewed as being made up of a set of separate statements which could then be redistributed across the EDM classes. In order to support this process, a model was created, defining which EDM classes would be used in the library context, the relations between them, and the EDM properties chosen to describe these relations. Europeana had at that time implemented

¹ <http://www.europeana.eu>

² The website of The European Library – <http://theeuropeanlibrary.org> – was relaunched in early 2012 with a new set of functions and a new design. This improved website is referred to as the new *European Library* portal.



only a subset of the EDM properties defined in the main EDM specification.³ Therefore Europeana Libraries decided to provide two profiles, one taking into account the first implementation and one including the full EDM specification.

The main issue to have emerged during the modelling phase was the identification of the level of description to be chosen for a resource. The group identified the *edition* level as the cultural heritage object (CHO) to be delivered to Europeana. The CHO is therefore an abstract concept instead of a physical thing as defined by the museum community. However for some material (rare or unique books for instance) such modelling could lead to a loss of information. The group has also expressed the need to be able to handle these specialist cases. Use of the class `edm:PhysicalThing` was recommended to specify the nature of the work. These observations have, firstly, highlighted the need for further investigation on how to represent *Functional Requirements for Bibliographic Records* (FRBR) entities in EDM and, secondly, have shown that the approach taken by libraries will need to be checked against material from other domains for consistency reasons. The model used for serials has not raised the same issues since the definition of the CHO is really dependent on the practices of libraries in terms of digitisation and cataloguing. EDM contains properties that allow the expression of both the vertical and horizontal relationships that characterise the structure of serials. It is therefore possible for libraries to apply as many levels as they need to reflect their own practice (title, volume, issue and article).

Considering which EDM properties could be used to describe the CHO and its digital representation, the working group identified various issues which were directly addressed to Europeana.

One of the main issues was the fact that EDM allows non-compliance with the official range specifications for the Dublin Core (DC) properties. For legacy reasons, it allows the use of literals as values for some properties where DC specifies references only. Since most of the values that will be supplied by the library domain will be literals, the group considered different options to stay compliant with DC and EDM specifications. The chosen option is to continue to use literal values in order to avoid any information loss. However, the group also looked at other namespaces, which would legitimately allow the use of literals.

Another point was made on the need for extra properties for some classes. It is particularly true for the class `edm:WebResource` for which only two rights properties are available. The group wanted to have properties such as `dcterms:extent` or `edm:isNextInSequence` to allow the description of a sequence of digital files when a CHO has been digitised into separate files (e.g. pages in a book). `Dcterms:created` would also be useful to capture the date of creation of a digital resource or born-digital resource.

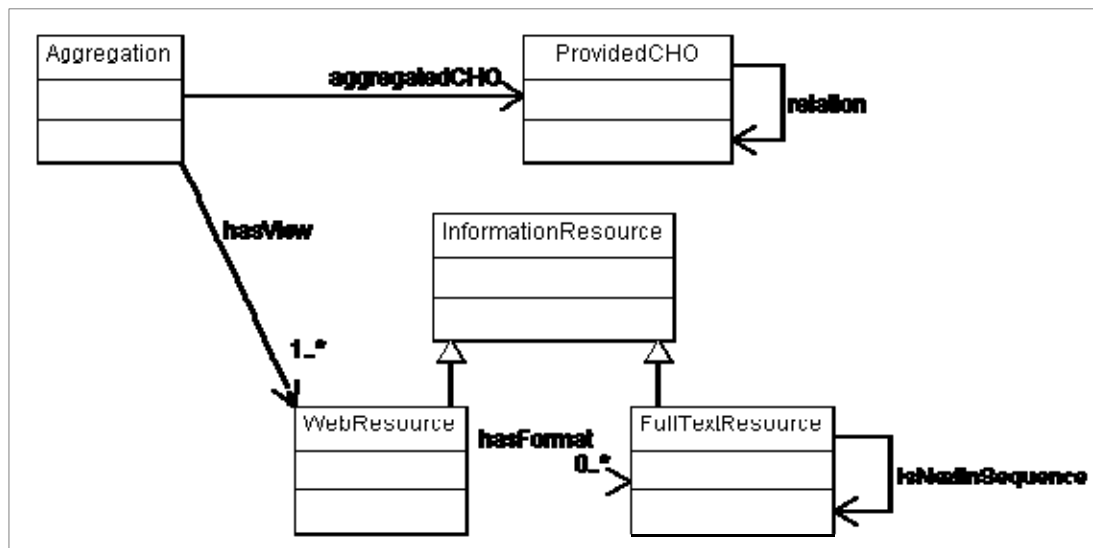
Library domain data, like museum data, is full of information describing the process of creation, publication, modification of a resource; the only difference being that

³ <http://pro.europeana.eu/edm-documentation>



libraries keep such information using a high level of granularity in their data. EDM, as a cross-domain model, allows event-based description. Even though the event class is not available in the first implementation of EDM, the working group looked at the possibility of using the `edm:Event` class to enable the inclusion of this information. It was actually the only means identified to describe statements such as the place of creation, place of publication etc.

Europeana Libraries also investigated how full-text could be represented in EDM. The current specification of EDM fulfils most of the requirements for representing full-text. However, Europeana Libraries recommended the creation of an additional class *FullTextResource* which would be a subclass of *InformationResource*. This class would allow the representation of the individual full-text content, separately from the digital representation (`edm:WebResource`) of the CHO. The relation between the `edm:WebResource` and the *FullTextResource* would be represented thanks to a new property- `hasFormat`.



Subset and extension of EDM for full-text

The model defined by the Europeana Libraries is tied to the complexity of library legacy data, which has introduced some uncertainties on some points. Further research was undertaken by the project in its second year in order to validate the recommendations made in its key deliverable D5.1.

The recommendations from Europeana Libraries were addressed to Europeana and were taken into account in future updates of the EDM documentation.

Further details are available in the deliverable *D5.1 Report on the alignment of library metadata with the European Data Model* available at <http://www.pro.europeana.eu/web/europeana-libraries-project>