DELIVERABLE

Project Acronym:	Europeana Libraries
Grant Agreement number:	270933
Project Title:	Europeana Libraries: Aggregating digital content from Europe's libraries

D5.2 Library domain metadata aligned with the Europeana Data Model

Version 1.0

Authors:

Anila Angjeli	Bibliothèque nationale de France
Martin Baumgartner	Bayerische Staatsbibliothek
Valentine Charles	The European Library / Europeana
Robina Clayphan	The European Library / Europeana
Corine Deliot	The British Library
Jörgen Eriksson	Lunds Universitet
Nuno Freire	The European Library
Alexander Huber	University of Oxford
Alexander Jahnke	Consortium of European Research Libraries
Gilberto Pedrosa	Instituto Superior Ténico, Lisbon
Vicky Phillips	National Library of Wales
Natalie Pollecutt	Wellcome Library
Glen Robson	National Library of Wales
Wolfram Seidler	Universität Wien
Stefanie Rühle	Consortium of European Research Libraries

With the participation of

Ana Barbeta	Historic Library of the University of Valencia
Audrey Drohan	Universty College Dublin
Justyna Walkowska	Poznańskie Centrum Superkomputerowo-Sieciowe
Hans Michelsen	Roskilde Libraries
Gill Hamilton	National Libray of Scotland
Radka Kalcheva	Public Library of Varna
Sarantos Kapidakis	Veria Central Public Library
Stefanie Gehrke	Herzog August Bibliothek

Project co-funded by the European Commission within the ICT Policy Support Programme		
DisseminationLevel		
P	**Public**	**P**

Revision History

Revision	Date	Author	Organisation	Description
0.1	15.10.12	Valentine Charles	The European Library	First draft outline of D5.2
0.2	20.12.12	Valentine Charles and Robina Clayphan	The European Library	Version submitted for external review.
V1.0	29.12.12	Valentine Charles	The European Library	Incorporation of the reviewer's comments into the final version

Statement of originality:

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

Table of Contents

1. Executive Summary

This report, *D5.2 Library domain metadata aligned with the Europeana Data* Model*,* has been written by Work Package 5 (WP5) of the *Europeana Libraries* project. The aim of the report is to describe how library metadata were converted to the Europeana Data Model (EDM) based on the recommendations provided by the first report *D5.1: Report on the alignment of library metadata with the Europeana Data Model.*

The report begins by explaining the methodology followed by the group against the reviewed model described in its first report. WP5 has decided to opt for two phases of validation. This first phase was aimed at the validation of the recommendations reported in D5.1. This validation was concretely designed as a mapping exercise. The outcomes of this exercise were implemented to perform the conversion of "real" data in EDM. The second phase of the validation aimed at identifying issues more related to data conversion rather than modelling issues.

The deliverable addresses all the comments, issues detected during this long validation process and proposes solutions for further developments. This practical work was accompanied by the production of a large documentation. The report continues by outlining how the comments addressed during this validation have pushed the group to review its first abstract model. A new version of the initial report has been published in addition to this more technical report.

This deliverable should be considered as a milestone in an ongoing process to improve the quality of an EDM which is already answering the Europeana requirements. The work was so far focused on the definition of mappings and conversion process to EDM. Future work will now be embedded in a strategy for improving the quality of EDM data. It will be also supported by a continuous dialogue with libraries to encourage better practices. The work of the group provided useful feedback for the Europeana team at a time where EDM was being implemented.

2. Introduction

This deliverable, D5.2 Library domain metadata aligned with the Europeana Data Model, is delivered as part of the task 5.2 Aligning library-domain metadata with The Europeana Data Model.

"*5.2.1 KB and Europeana will validate the recommendations of D5.1 by aligning a sample of metadata from different types of libraries (regional, municipal, research, university) that has been provided directly to Europeana or via a network project (e.g. Europeana Local) (…)*

5.2.2 KB will align the metadata of 10 libraries (regional, municipal, research, university) that has already been provided directly to Europeana or via a network project (e.g. Europeana Local) in Europeana Semantic Elements (ESE) format with the Europeana Data Model (EDM) (…)"

The description of work initially planned two separate deliverables, one for the national libraries: (D5.2 National Library metadata aligned with the Europeana Data Model) and the second for the research libraries (D5.3 Metadata from the wider library domain aligned with the Europeana Data Model). While proceeding to the validation of the recommendations of D5.1, WP5 realised that the conclusions of the work would be similar for the national and research libraries, the data described following the same metadata standards. It has been therefore agreed to merge the deliverables D5.2 and D5.3 into one deliverable.

This deliverable describes how the abstract model and the recommendations provided in the *Report on the alignment of library metadata with the Europeana Data Model* have been validated against real data following different metadata standards.

3. Initial statements

In the first year of the project, WP5 analysed EDM in detail and defined some specifications and requirements for library data. The initial focus was on monographs, multi-volume works and serials. Different issues were raised and addressed during this phase and a report with recommendations produced[1]. The objective of the second year of the project was twofold: first of all to validate the recommendations and then to transform data samples into EDM. These two phases of the validation aimed at separating the issues raised by the model itself and the issues raised by the implementation of EDM in the main aggregation infrastructure (WP4 activities).

4. Validation first phase

In its first deliverable, WP5 defined some recommendations on the use of the EDM classes and properties with monographs, multi-volumes and serials materials. The objective was then to validate these recommendations against real library data. Experts from national, research and university library partners in the Europeana Libraries project were asked to map their data to EDM according to the recommendations provided in D5.1 and focus on the issues raised by the report.

4.1. Main issues to be validated

During the course of the work to define the profiles for the different types of library materials, several issues arose. Some of these are to do with differing library practice or the significance of differing materials, some to the nature of current library data and some to the constraints of EDM. In particular, there are limitations due to the first implementation of EDM which comprises a subset of the full EDM specification.

4.1.1. Model for monographs

4.1.1.1. edm:ProvidedCHO

In the context of published textual resources, WP5 has defined the Cultural Heritage Object (the ProvidedCHO in EDM terminology) as being the edition, that is to say, the entirety of all identical copies of a text produced in the same process of publishing. This modelling has the advantage of allowing separation between the data coming with the process of publication and the data about the process of digitisation.

This was seen as a problem for those partners who principally catalogued individual books (rare books) at the item level (in FRBR terms). Initially it was thought a separate model would be needed for such objects and a separate “Rare Book Model” was proposed in the original version of D5.1. After feedback and discussion however a better understanding was reached and this is reported in Section 5 of this document and in the updated verion of D5.1 (D5.1v2)

4.1.1.2. Properties requested for implementation by Europeana

4.1.1.2.1. For the edm:ProvidedCHO

Europeana Libraries recommends the implementation of the property *edm:isSuccessorOf* to capture the implicit relation between the continuation of a resource and that resource.

4.1.1.2.2. For the edm:WebResource

Europeana Libraries recommends the implementation of the properties *dc:format, dc:source, dcterms:extent, dcterms:conformsTo, dcterms:isFormatOf, edm:isNextInSequence, edm:isRepresentationOf.*

[1] D5.1 Report on the alignment of library metadata with the Europeana Data Model (EDM): http://www.europeana-libraries.eu/documents/868553/1eade085-34ac-487f-82af-d5cd2545e619

The use of all the above properties would provide a much more useful description of the resource and allow sequencing of digital files where one original CHO had been digitised into many separate files (e.g. pages in a book).

4.1.1.2.3. dcterms:created

The group wants to use dcterms:created to capture the date of creation of a digital resource by digitisation of an existing object or the creation of a digital resource. This property will be added by Europeana in the next iteration of the EDM schema.
Since the library community uses many dates in relation to library materials, including those now born digital (e.g. theses); the group will have to pay a particular attention to this issue during the validation.

4.1.1.2.4. edm: isRepresentationOf

Europeana Libraries is interested in using edm:isRepresentationOf to indicate the source of the digital object. This solution might be redundant with the class ore:Aggregation which links the edm:WebResource and the edm:ProvidedCHO.
A use case will have to be identified to justify the use of edm:isRepresentationOf.

4.1.2. Model for Serials

4.1.2.1. edm:ProvidedCHO

Europeana Libraries identified different levels in the serials model (article, issue, volume, title) which could give rise to and could be the subject of a package of data submitted to Europeana. The group raised the importance of specifying the type of each level when submitted as a Provided CHO. The model would used dc:type to describe this information at each level. The short term solution is to provide a literal value for this property. However the group recommended the use of controlled terms (MARC genre list[2], Ontology bibo[3], etc)

Europeana Libraries recommends dcterms:isPartOf to represent vertical relationships existing between the resources. edm:isNextInSequence is recommended to describe horizontal relationships between resources.

4.1.2.2. Properties requested for implementation by Europeana

4.1.2.2.1. For edm:ProvidedCHO

Europeana Libraries recommends the implementation of the properties *dcterms:abstract, dcterms:bibliographicCitation* to capture specific information particular to serials.

4.1.2.2.2. For edm:WebResource

Europeana Libraries recommends the implementation of the property edm:isNextInSequence to allow description of pagination within a digital resource.
The project also recommends the implementation of dcterms:hasFormat since this property will be used in the model for full-text[4] .

4.1.3. Range of the properties to use

DCMI specifies non-literal values as the range of many of its dcterms properties, but EDM allows the use of literals for these properties. Most of the values that will be supplied from the library community will be literals and will therefore need to be accommodated in the schema.
Europeana Libraries suggests to use literal values in order to ensure the values are represented in the portal.

---

[2] Marc genre list available at http://www.loc.gov/standards/valuelist/marcgt.html
[3] Bibo ontology http://bibliontology.com/content/complex-series-proceeding-article-use-case and http://bibliontology.com/content/article
[4] Report on how the full-text content will be made available to Europeana (D4.3) http://www.europeana-libraries.eu/documents/868553/0ffa3e2f-5e38-4507-bd34-88dba2b03040

The use of properties from other namespaces such as ISBD which will permit use of literals could be a solution at a later stage.

4.2. Results and analysis of the results

Each WP5 member was asked to select typical records for the types of materials covered by D5.1. The selection contained 5 to 10 representative records in any library metadata format, and from different collections.

Considering the issues mentioned above, each WP5 member was asked to map a small number of records manually by using a spreadsheet created to help the exercise[5]. It contains the listing of classes and properties for the recommended profiles and additional pages showing examples of other formats mapped to EDM. It has five named worksheets:

1. **'ELibs monos properties'** lists the classes and properties recommended for monographs and multi-volume works. It includes some that are not going to be implemented at first but are included to show the need for them.
2. **'ELibs serials properties'** lists the full set for serials.
3. **'HOPE mappings'** is the mapping produced by the Hope project[6]. This is an extensive mapping from different formats based on the principles of library cataloguing. MODS and MARC-XML are shown. This should be seen as valuable guidance, but bear in mind that HOPE had different objectives. Many columns are hidden as they related to HOPE's requirements but you can unhide them if you would like.
4. **'TEL-AP'** shows the mapping from The European Library Application Profile. This is largely Dublin Core-based but has several additional elements used in the TEL portal. This page shows all the EDM elements but those that are not going to be implemented at first are shown in a different colour.
5. **'METS-MODS'** from Wales is the mapping provided by the National Library of Wales

EDM allows the modelling of additional data about entities that are distinct from the CHO and then produces new contextual resources. Although the use of the contextual resources has not been extensively discuss in the report D.5.1, partners were encouraged to use them. The classes and properties referring to them have been added in the documentation.

In total, around 15 mappings have been produced by the group. They illustrate the diversity of materials and metadata formats used in libraries. This mapping exercise has shown how EDM is proposing new ways of thinking about library data. The group has reached a consensus on the mapping from ESE, MODS, MARC21 and UNIMARC to EDM. The mapping rules defined during the exercise have been used to process the conversion of the data.

The first issues were related to the distribution of the EDM properties across the different classes. This difficulty is also highlighted by the fact that there are not always identifiers to describe these different classes. A paradoxical situation can be observed regarding the provision of contextual resources. Identifiers are sometimes available in the data but not always thought as a potential contextual resource.

The mapping exercise has reiterated the limits of EDM when there is the need to describe a person role information and the place of publication. In these two situations the recommendations of D5.1 have been considered a good solution before further developments of EDM.

[5] Validation mappings document available at http://www.theeuropeanlibrary.org/confluence/display/downloads/Europeana+Libraries-+WP5+documentation
[6] http://www.peoplesheritage.eu/

5. Update of D 5.1 *Report on the alignment of library metadata with the European Data Model (EDM)* [7]

D5.1 was reviewed by a group of experts just before it was delivered and some important issues were raised. There was not time to address these in D5.1 but it was decided to tackle them as part of the validation phase and to produce an updated version of D5.1 (D5.1 V2).

5.1. Comments raised during the review

The first issue related to the "Rare Book Model" that had been proposed in D5.1 as a solution to having defined the ProvidedCHO as representing the edition level of an object. The reviewers wondered why the edition level had been selected rather than the Work level. Europeana clarified the nature of the Provided CHO class however, and this resolved the problem: edm:ProvidedCHO works in Europeana as a functional type. It is not supposed to say anything on the nature of the objects. A ProvidedCHO can be any object interest for Europeana, that can appear as one item in a result list of a query on Europeana.eu. It is therefore possible to define an edition as a ProvidedCHO in one case and an item as a ProvidedCHO in another even from the same domain..

The second comment from the reviewers was related to the use of blank nodes where URIs are required but library data only has literals (e.g. hasMet for place of publication or to model an Agent.). Blank nodes allow the creation of an anonymous resource and therefore don't require a URI and literal when they are not available. It is right that DCMI specifies non-literal values as the range of many of its dcterms properties, but EDM allows the use of literals for these properties. However using blank nodes when URIs are not available is not the best solution. EDM allows local identifiers if no URIs are available. Local identifiers or "fake" identifiers should be used instead.

5.2. Changes included in the second version

A new version of the deliverable D5.1 answering the questions raised by the internal reviewers has been produced[8]. This new version consists of many relatively minor changes such as the clarification and further explanation of diagrams and some statements or assumptions.

The comment about the definition of the ProvidedCHO needed a more extensive consideration as it affected a major aspect of the report. In light with the reviewers recommendations it has been decided that a separate model for monographs and rare books was not necessary anymore. The group decided to consider the ProvidedCHO at the item level as well as at the edition level. Therefore in the current Europeana Libraries model for text resources all information concerning the manifestation, expression and work will be added to the ProvidedCHO the same as for an item. The distinction between them will lie only in the metadata used and in the relationships expressed. For example, an Item level ProvidedCHO could have an edm:realises link to a ProvidedCHO that represents the edition level. The initial "Rare Books Model" has been kept for reference as annex[9].

Ideally, the description of items and bibliographic objects should be compliant with FRBR, a framework that identifies the entities relevant to find, identify, select and obtain resources. The group therefore recommend that Europeana continues the work to extend EDM using FRBRoo entities for the description of the relations between item, manifestation, expression and work. This extension cannot be part of the first implementation of EDM by Europeana.[10]

---

[7] The version 2.0 of D5.1 is available at http://www.theeuropeanlibrary.org/confluence/display/downloads/Europeana+Libraries-+WP5+documentation

[8] D5.1v2 http://www.theeuropeanlibrary.org/confluence/display/downloads/Europeana+Libraries-+WP5+documentation

[9] This is shown in Annex 4 in D5.1.v2

[10] To ensure that it is possible to apply the data to the FRBR entities at a later date a provisional distribution of the the properties in the ProvidedCHO to FRBR entities has been made. This is shown in Section 9 – Future Work of D5.1. V2.0.

The definition(s) of the ProvidedCHO for library materials, particularly in relation to the FRBR group 1 entities (Work, Expression, Manifestation and Item), was the area that provoked most discussion during the Metadata Working Group meetings. The review of the draft deliverable by the Europeana Libraries Appraisal group highlighted that, despite the pragmatic approach of the group, further discussion will be needed, both within the library-domain and with the other cultural heritage domains, as part of the validation process and in the future work suggested. The work done by Europeana Libraries has been submitted for feedback as input for the work of the Europeana Taskforce on FRBRoo[11].

WP5 has worked in strong collaboration with the Europeana office. The list of issues addressed to Europeana has been in addition summarized into a case study published in the Europeana professional website (**Annex 1**).

6. Validation second phase

The final objective of WP5 was to be able to actually convert libraries data to the Europeana Data Model for libraries. After having tested the model from a theoretical perspective, it was necessary to validate it on a more technical level. The model for libraries has been therefore included in the main Europeana Libraries aggregation infrastructure and the output has been again submitted to the full WP5.

6.1. Implementation of EDM into the United Ingestion Manager

After the production of its key deliverable D5.1 *Report on the alignment of library metadata with the Europeana Data Model (EDM)*[12], the next step for WP5 was to produce some data samples in EDM. These data samples are a first experiment on real library records existing in The European Library, it was based on the recommendations made in D5.1V2.

WP5 and WP4 worked in strong collaboration on the implementation of EDM. It was necessary to integrate the work done by WP5 into the current workflow developed by WP4 to be able to transform data from any library source format into EDM. The first step in this process was to implement the EDM mapping defined in D5.1 in the Unified Ingestion Manager (UIM)[13] developed by WP4 in this project. The Unified Ingestion Manager is a central platform for managing the ingestion of metadata in The European Library. It is a framework in which the mapping, the normalisation and the enrichment of data are performed. The UIM has been developed further within WP4 task 4.2.

In order to implement EDM in the UIM an extra step was needed, which consists in the mapping of the Internal Object Model of The European Library to EDM. The Internal Object model is a technical solution chosen by The European Library to represent data of aggregated bibliographic records and digital objects within the UIM and the portal. It allows the representation of entities with different levels of semantic detail and structure. It also keeps the richness of the original data coming from libraries. The different classes of this internal object model have been mapped to the EDM classes and properties, and the mapping was implemented into the UIM, as shown in Figure 1.

[11] More information on the taskforce on FRBRoo- EDM at http://pro.europeana.eu/web/network/europeana-tech/-/wiki/Main/Task+Force+EDM+FRBRoo

[12] D5.1 Report on the alignment of library metadata with the Europeana Data Model (EDM): http://www.europeana-libraries.eu/documents/868553/1eade085-34ac-487f-82af-d5cd2545e619

[13] For more details on the UIM, see The European Library Standards Handbook at http://www.europeana-libraries.eu/documents/868553/4f7aaf72-2822-42a0-b8a5-f31ef6fcc5bb

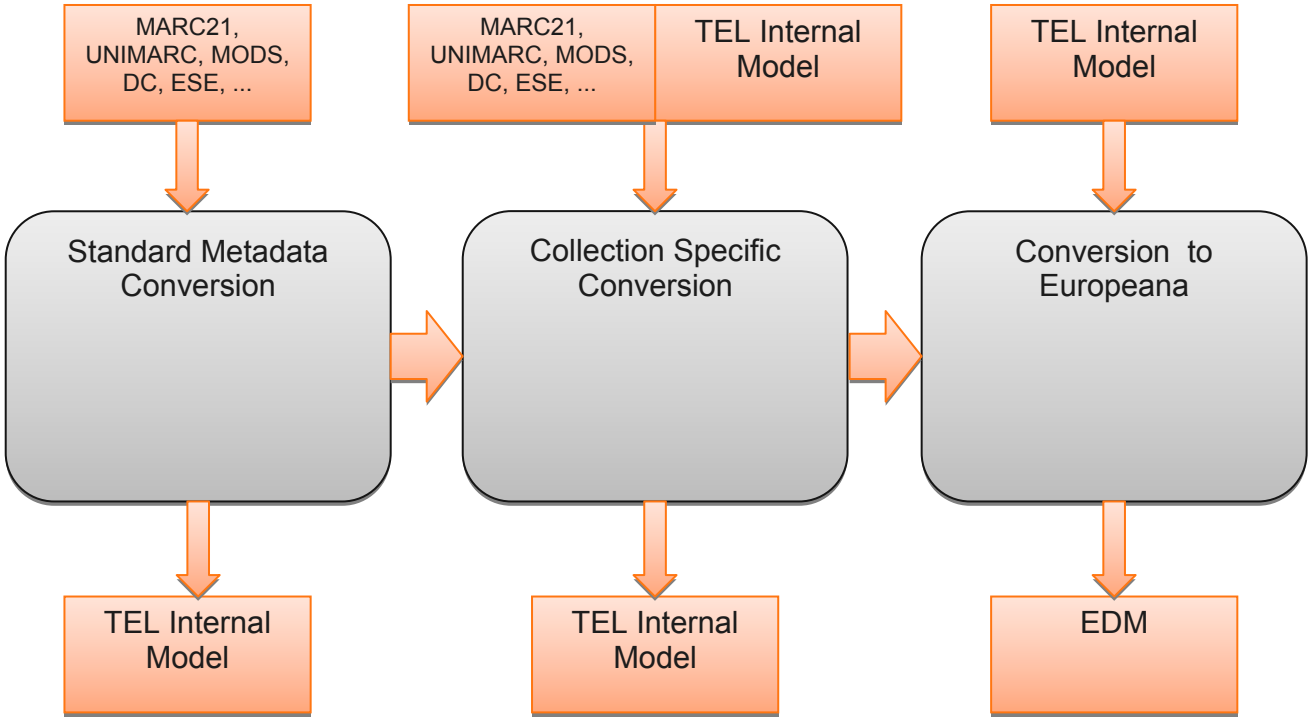


Figure 1 Conversion workflow within the Europeana Libraries aggregation infrastructure

6.1.1. Samples of data produced for validation

From June 2012, a series of data samples have been created and then continuously improved based on the feedback from WP5 participants.

6.1.1.1. First set of samples

In the first round, fourteen different datasets[14] of ten records each, were transformed. They came from National Library and research library partners in the Europeana Libraries project. They illustrate the diversity of materials and metadata formats used in libraries.

Datasets identifiers	Library name	Original data format
a1012 http://www.theeuropeanlibrary.org/tel4/collection/a1012	Universidad Complutense Madrid (ES)	Marc21
a1016 http://www.theeuropeanlibrary.org/tel4/collection/a1016	Universidad Complutense Madrid (ES)	Marc21
a1017 http://www.theeuropeanlibrary.org/tel4/collection/a1017	Bayerische Staatsbibliothek (DE)	Europeana Semantic Elements (ESE)
a1018 http://www.theeuropeanlibrary.org/tel4/collection/a1018	The Wellcome Library (UK)	Marc21
a1023 http://www.theeuropeanlibrary.org/tel4/collection/a1023	Library of Lucian Blaga, University of Sibiu (RO)	Dublin Core
a1024 http://www.theeuropeanlibrary.org/tel4/collection/a1024	National Library of Wales (UK)	Europeana Semantic Elements (ESE)

[14] The data samples are available at: http://www.theeuropeanlibrary.org/confluence/display/downloads/Europeana+Libraries-+WP5+documentation

a1027 http://www.theeuropeanlibrary.org/tel4/collection/a1027	National Library of France (FR)	TEL Application Profile
a1033 http://www.theeuropeanlibrary.org/tel4/collection/a1033	The Romanian Academy Library (RO)	Dublin Core
a1034 http://www.theeuropeanlibrary.org/tel4/collection/a1034	University of Tartu (EE)	TEL Application Profile
a1036 http://www.theeuropeanlibrary.org/tel4/collection/a1036	University of Leuven (BE)	Europeana Semantic Elements (ESE)
a1040 http://www.theeuropeanlibrary.org/tel4/collection/a1040	Universität Bern (CH)	Marc21
a1047 http://www.theeuropeanlibrary.org/tel4/collection/a1047	National Library of Wales (UK)	Europeana Semantic Elements (ESE)
a1050 http://www.theeuropeanlibrary.org/tel4/collection/a1050	Universidad Complutense Madrid (ES)	Marc21
a1060 http://www.theeuropeanlibrary.org/tel4/collection/a1060	University of Uppsala (SE)	TEL Application Profile

Each folder contains ten records from several collections. Several files are available per record:

- The file ending by .edm.xml is the output record in EDM
- The file ending by .telom.xml is the The European Library Internal Object Model xml record.
- The file ending by <originalFormat>.xml (e.g. marc21.xml, ese.xml) is the original metadata from the library in its source format.
- The file edm.validation.txt is the result of the validation of the edm record against the XML schema. This file has been useful to identify the following issues.

6.1.1.2. Second set of samples validated by partners

Some issues where detected in the first round and were fixed in the second versions that were provided back to partners. More datasets were converted for this new version[15]. This version primarily includes datasets from partners involved in WP5.

Datasets identifiers	**Library name**	**Original data format**
a0444 http://www.theeuropeanlibrary.org/tel4/collection/a0444	The British Library (UK)	Customed data format
a0142 http://www.theeuropeanlibrary.org/tel4/collection/a0142	The National Library of France (FR)	OAI_DC
a1114 http://www.theeuropeanlibrary.org/tel4/collection/a114	Bayerische Staatsbibliothek (DE)	MARC 21
a1018 http://www.theeuropeanlibrary.org/tel4/collection/a1018	The Wellcome Library (UK)	MARC 21
a1030 http://www.theeuropeanlibrary.org/tel4/collection/a1030	Historic Library of the University of Valencia (ES)	UNIMARC
a1028 http://www.theeuropeanlibrary.org/tel4/collection/a1028	Herzog August Bibliothek (DE)	ESE
a1051	Universität Wien (AT)	ESE

[15] The data samples are available at http://www.theeuropeanlibrary.org/confluence/display/downloads/Europeana+Libraries-+WP5+documentation

http://www.theeuropeanlibrary.org/tel4/collection/a1051		
a1009a http://www.theeuropeanlibrary.org/tel4/collection/a1009a	University of Oxford (UK)	TEL Application Profile
a1037 http://www.theeuropeanlibrary.org/tel4/collection/a1037	Lunds Universitet (SE)	OAI_DC
a1024 http://www.theeuropeanlibrary.org/tel4/collection/a1024	National Library of Wales (UK)	Europeana Semantic Elements (ESE)
a0005 http://www.theeuropeanlibrary.org/tel4/collection/a0005	National Library of Portugal (PT)	UNIMARC

6.2. Creation of a questionnaire

The datasets converted to EDM were submitted to WP5 for validation. The objective of this second validation was to address issues more related to data conversion rather than to the abstract model as done previously. In order to facilitate this validation work, a questionnaire (**Annex 2**) has been created. Each of the questions asked to partners highlighted a specific feature or requirement of EDM:

- **Loss of data:** it was important to identify whether or not some data were lost during the conversion process. This issue can be both related to the conversion process itself but also to the mapping that has been implemented.
- **Loss of semantic or incorrect semantics:** this issue is directly related to the main requirement of EDM which is to respect of the "One-to-One principle[16]". EDM allows the distinction between descriptive metadata related to a "real-world" object and its digital representation(s).
- **Identification of resources:** EDM requires URIs for the identification of the main classes and for lot of properties. It is important for each of these classes to have an identifier which is unique.

This questionnaire also pushed WP5 to think their data in a less traditional way, EDM being closer to a semantic web approach rather than the library well-known standards.

6.3. Issues detected

The validation of the data samples by WP5 has highlighted some issues that will be crucial for further development of the model. These issues don't always occur depending of the source format of the data. For instance it is easier to introduce noise in the data when they are originally described with Marc 21. Data originally described with Dublin Core-like application profiles are less likely to introduce errors.

6.3.1. Distribution of properties across the EDM classes.

The first issue is related to the distribution of properties across classes in EDM. The figure 2 shows how distinct resources are described separately in EDM. For libraries data (e.g. a typical Marc21 record), it is difficult to separate the information related to the Cultural Heritage Object (class named edm:ProvidedCHO) from the ones related to the digital representation of this CHO (edm:WebResource).

[16] Miller, Steven J. (2010). The One-To-One Principle: Challenges in Current Practice. In Proc. of the Intl. Conf. on Dublin Core and Metadata Applications (DC2010), Pittsburgh, Pennsylvania.

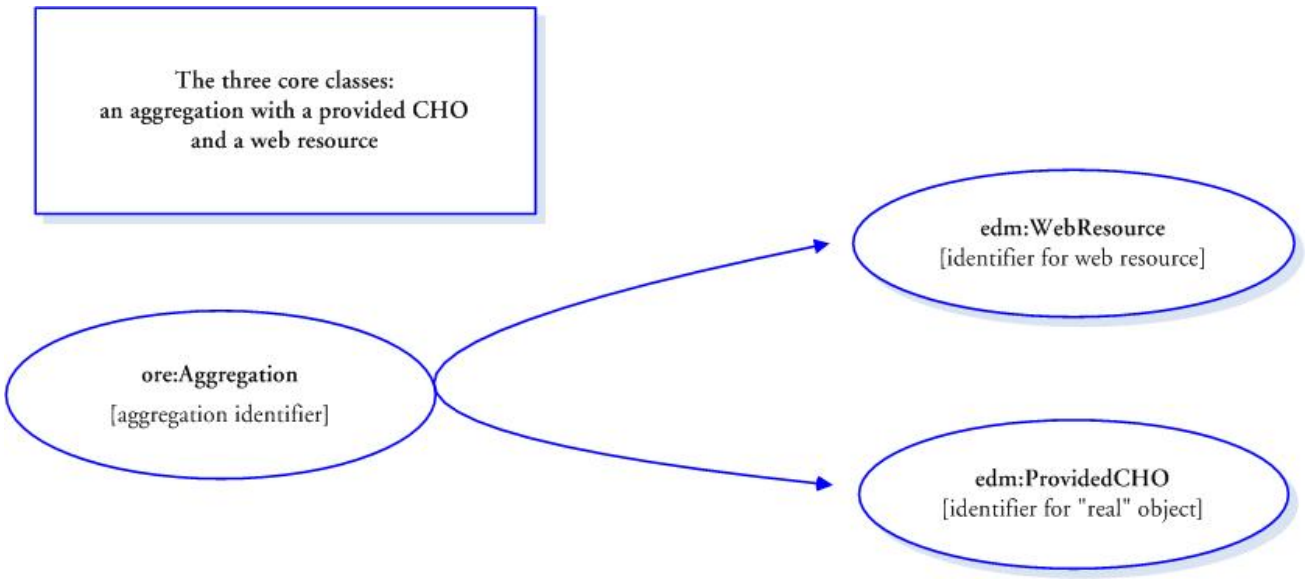


Figure 2 Core EDM classes

This issue is crucial when providing identifiers for these specific classes. This challenge doesn't only happen at the classes level but also when defining specific properties. For instance EDM requires different kind of links to a digital resource for different purposes, and these are represented by different properties: edm:isShownBy, edm:isShownAt and edm:object. It is difficult to identify from the original record which link corresponds to which properties based only on the fact of identifying a link. For instance a particular link may be used to express a hierarchical relation and should be used in dc:relation instead. This issue can be found in data described with MARC21 where a datafield can be used to store more than one URL.
This issue is reinforced by the fact that lot of EDM properties require a URI which are usually not in the legacy data.

6.3.2. Support of multiple statements per resources

Another issue is related to the support of multiple statements per resource. Since EDM allows multiple web resources per object, it is possible to have many and different metadata statements per web resource. A typical situation would be the support of different rights statements per Web resource, or the differentiation between the rights applied to a 'real' object and the rights applied to the digital resource. The mapping to express these cases is at the moment not optimal for the library data even though we note that the feature that would allow it is not part of the first implementation of EDM by Europeana. The difficulty of this situation is also recognised by Europeana: it explains that the rights information is currently supported only at the Aggregation level so it is necessary to choose only one rights statement for the whole "package of data".

Rights information (dc:rights element) is not the only challenge. Difficulties can also be encountered in the fields describing date (dc:date) and format (dc:format) information. Similar issues will apply also for the definition of contextual resources. It will be difficult from the original data to identify whether metadata applies to an Agent, a Place related to a certain CHO. In MARC 21 data for instance this information is quite often encoded and distributed across a record.

6.3.3. Improve the semantic of the data

The initial mapping to EDM has been kept simple on purpose. The output EDM data have highlighted some situations where the semantic of the data could be improved by refining the data in EDM. Quite often data have been mapped to a generic property which semantic could be

improved by choosing the appropriate refinement. For instance the current dc:description could be split in more properties such dcterms:abstract, dcterms:tableOfContents…

6.3.4. Improve the mapping of data

The validation of the conversion has shown that the mapping implemented could be refined by the addition of some properties (**Annex 3**). This comment is especially relevant for data from MARC21 and UNIMARC which are richer. The mapping currently doesn't support very well attributes present in the data such as the xml:lang attributes. This information is usually very important for the interpretation of coded values in MARC21 and UNIMARC and should be supported.

6.3.5. Issue related to source data in Marc21 or similar formats

Original data described using bibliographical standards family present the difficulty of encoded values. It is quite difficult for a conversion process to interpret all the coded values. Even if the mapping is currently supporting it quite well, it is possible to identify some noise in the data coming from the interpreted labels.

6.3.6. Generation of contextual resource

EDM allows finer grained data descriptions by providing ways of modelling contextual resources. Library legacy data are rich in identifiers and references to controlled vocabularies, authority files which could be modelled as an Agent, Place…However these data are often either encoded or described as simple strings. Another issue comes from the fact that controlled identifiers pointing to a vocabulary or authority file are not always identified in the data itself. It is then difficult to re-use them as appropriate by attributing them to an agent, time, place entity.

6.3.7. Summary

To summarise the main issues considered during the validation of these samples are:

- **the granularity of the description.** On one hand, if the original data are rich, EDM will have to keep the same level of granularity. On the other hand the quality of the EDM record might be affected by the lack of metadata in the original data.
- **The respecting the "one-to-one" principle**. Information related to a specific resource will need to be attached to this resource. The differentiation between the CHO, the Web Resource and the Aggregation are key in EDM.
- **The identification of available identifiers** which will trigger their re-use.

7. Solutions and future work

Improvement of the conversion of the data and the mapping is an ongoing effort which will go beyond the time-frame of the Europeana Libraries project.

7.1. Distribution of properties across the EDM classes and Support of multiple statements per resources

This issue is mainly linked to the library original data. However a better interpretation of the coded values and attributes available in the data would help to identify the kind of information which is described. The identification would ease the process of data distribution across the EDM classes. A longer term effort would consist in replacing the literal value by resolvable URIs.
An automatic conversion is not optimal and data quality policy would be necessary to really make sure resources are properly described across the EDM properties. Manual operations are likely to be the best solution to reduce the ambiguity and provide a better interpretation of the data. It also belongs to libraries to be more rigorous in the way they provide identifiers.

7.2. Improve the semantic, mapping of the data

A continuous improvement of the mapping is the solution for improving the quality of the data. The European Library as the aggregator for libraries is responsible for providing richer library data to Europeana. On the other hand Europeana by raising its expectations in terms of data quality with EDM will participate in this effort.

7.2.1. Generation of contextual resources

Lots of identifiers present in library data are good candidates to create new contextual resources. However the source of these identifiers should be always described in the appropriate attributes. This practice would easier the process of disambiguation of the resources. For instance identifying a series of numbers as being a code from the Library of Congress Subject Headings[17] would allow the identification of the resource as a potential concept. After the identification of the identifiers present in the data, the solution is to select key vocabularies to which data could be matched. The validation exercise as shown that an important number of libraries are using VIAF[18] identifiers or GND[19] identifiers in their data. These could be the first targets Europeana Libraries could link the data to.

8. Conclusion

The validation of the records by partners has shown that the EDM produced by Europeana Libraries reaches the quality level currently expected by Europeana. This work also proved the efficiency of the aggregation infrastructure developed by WP4 by allowing the implementation of another data model alongside other library metadata standards already supported by the infrastructure. The mapping will from now continuously improve in the frame of the aggregation infrastructure sustained by The European Library.
The Metadata Working group created during Europeana Libraries will continue working on the challenges raised by the paradigms proposed by the Semantic Web. Because EDM evolves in the same perspective, it can be considered as a means for libraries to rethink their cataloguing practices and own standards. This work will be done collaboratively with other projects currently working on similar topics. The outcomes of the Digital Manuscripts to Europeana (DM2E)[20] project or the Europeana taskforce on EDM-FRBRoo for instance, will need to be embedded in this future work.
It will be also key for The European Library to follow the work that is being led by the Library of Congress as part of the Bibliographic Framework Transition Initiative. The first draft of this new data model[21] was unfortunately issued too late for the group to incorporate it into the work of D5.2, but some aspects of it, such as the different organization of the classes with respect to the FRBR model and the creation of contextual resources, have already been identified as interesting for future work.

9. Annexes

[17] http://id.loc.gov/authorities/subjects.html
[18] The Virtual International Authority File http://viaf.org/
[19] Gemeinsame Normdatei (GND) ontology http://d-nb.info/standards/elementset/gnd
[20] Digital Manuscripts to Europeana (DM2E) http://dm2e.eu/
[21] http://www.loc.gov/marc/transition/pdf/marcld-report-11-21-2012.pdf

Annex 1

Case study Europeana Libraries and EDM for libraries available on the Europeana Professional website at *http://pro.europeana.eu/europeana-libraries-edm*

Europeana Libraries is a two years project aiming at the creation of a robust aggregation model, which will make digital content from research and national libraries across Europe available on both *Europeana*[22] and the new *European Library*[23] portal. One of the objectives of the project is to work on how library data can be aligned to the Europeana Data Model (EDM). In its latest report, Europeana Libraries has addressed to Europeana some keys issues, which are 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 is currently implementing only a subset of the EDM properties defined in the main EDM specification.[24] Therefore Europeana Libraries has 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 modeling 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 modeling could lead to a loss of information. The group has also expressed the need to be able to handle these specialist cases. The class edm:PhysicalThing has been recommended to specify the nature of the work. These observations have, firstly, highlighted the need for further investigation on how to represent FRBR entities in EDM and, secondly, 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 have been 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

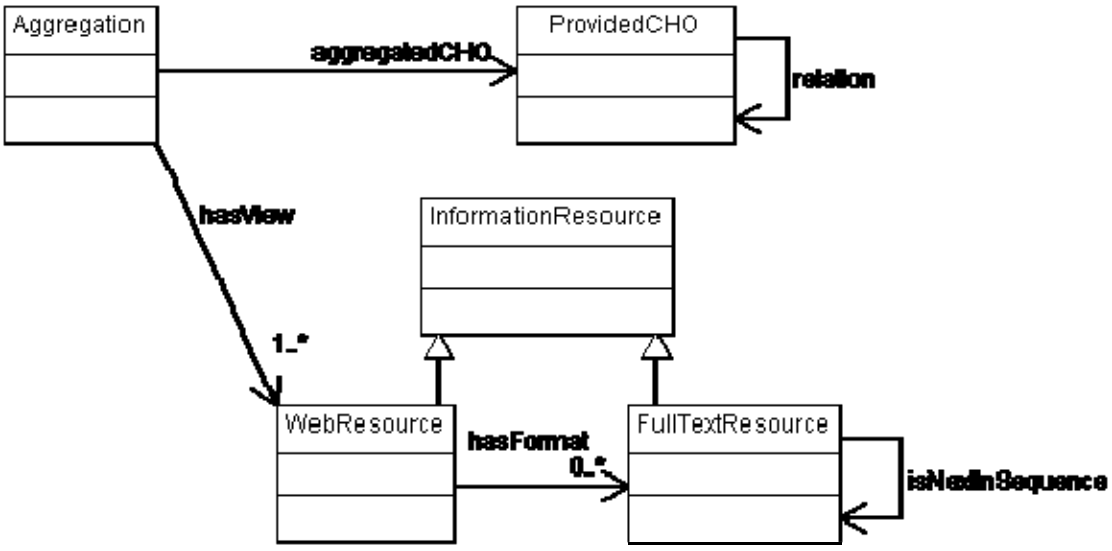
[22] http://www.europeana.eu
[23] The website of The European Library – http://theeuropeanlibrary.org – will be relaunched in early 2012 with a new set of functions and a new design. This improved website is referred to as the new *European Library* portal.
[24] http://pro.europeana.eu/edm-documentation

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 is also looking 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 would like 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 be also 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 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 has 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 has 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 is recommending 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 will therefore be undertaken by the project in its second year in order to validate the recommendations made in its key deliverable D5.1. If needed, the project will also provide further developments which could be useful for the creation of other EDM domain profile.

The recommendations from Europeana Libraries have also been addressed to Europeana and will be 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

Annex 2

Questionnaire to validate the EDM conversion
Example filled by Martin Baumgartner, 12.11.2012

This questionnaire is meant to help partners to judge the conversion of their original data to EDM. This questionnaire is reproducing the main classes of EDM and their properties. When comparing your orignal data with the output to EDM, select your answer in the list when a question is asked, and add your comments in the last column of the table.

Note: the properties that have not been implemented by Europeana have not been included in this questionnaire.

1) The ProvidedCHO

Property	Definition	Is this information available in the original data?	Is this information available after conversion to EDM?	Other questions		Remarks on the property/class conversion
edm:ProvidedCHO	The ProvidedCHO is the cultural heritage object which has given rise to and is the subject of the package of data that has been submitted to Europeana	no	yes	Is this resource properly identified?	yes with a functional HTTP URI	Why not a working URL instead of this fake URL? (http://data.theeuropeanlibrary.org/source_resource/2000081699439)
				Was there any other relevant data for the end user, present in the original but not in the EDM record?	if yes, please enter a short description of the fields missing?	identifier: VD-Numbers, WorldCat URL Shelf mark
				Was any data incorrectly assigned to properties of this class?	If yes, please indicate in which properties	I'm not sure what "<dc:identifier>a092c9c4-edaa-44b8-bece-64bc1f7a5a2d" is about? Should publication year be part of <dc:publisher>?

owl:sameAs	Indicates that two URI references actually refer to the same thing.	no	no			when we can differentiate between item and manifestation, then there will be identifiers for sameAs
dc:contributor	An entity responsible for making contributions to the resource.	yes	yes	Was there any other relevant data that could be mapped to a contextual resource (Agent, Place…)	yes	There is a loss of information from MARC to EDM MARC: <datafield ind1="0" ind2=" " tag="700"> <subfield code="a">Heinrich</subfield> <subfield code="b">II.</subfield> <subfield code="c">Römisch-Deutsches Reich, Kaiser</subfield> <subfield code="d">973-1024</subfield> <subfield code="0">(DE-588)118548255</subfield> </datafield> EDM: <dc:contributor>Heinrich</dc:contributor> Use GND record to build Agent
dc:creator	An entity primarily responsible for making the resource.	yes	yes	Was there any other relevant data that could be mapped to a contextual resource (Agent, Place…)	yes	see remark above Use GND record to build Agent
dc:coverage (mandatory**)	The spatial or temporal topic of the resource, the spatial applicability of the resource, or the jurisdiction under which the resource is relevant. (Note: Mandatory in EDM to supply one of dc:subject or dc:coverage or dc:type or dcterms:spatial)	yes	yes	Was there any other relevant data that could be mapped to a contextual resource (Agent, Place…)	yes	Use GND records to build Agents, Places, Times

dcterms:spatial (mandatory**)	Spatial characteristics of the resource. (Note: Mandatory in EDM to supply one of dc:subject or dc:coverage or dc:type or dcterms:spatial)	yes	yes	Was there any other relevant data that could be mapped to a contextual resource (Agent, Place…)	yes	MARC 651
dcterms:temporal	Temporal characteristics of the resource.	yes	no	Was there any other relevant data that could be mapped to a contextual resource (Agent, Place…)	yes	MARC 648
dc:date	Use for a significant date in the life of the CHO. Consider the subproperties of dcterms:issued or dcterms:created.	yes	yes	The values in this property are related to the digital object	no	
dcterms:issued	Date of formal issuance (e.g., publication) of the resource. (Encode as W3CDTF)	yes	yes	The values in this property are related to the digital object		should be used instead of dc:date (TEL)
dcterms:created	The date of the creation of the CHO	no	no	The values in this property are related to the digital object		not used
dc:description (mandatory*)	Description may include but is not limited to: an abstract, a table of contents, a graphical representation, or a free-text account of the resource. (Note: Mandatory in EDM to supply one of dc:title or dc:description. Dc:title is mandatory in this specification.)	yes	yes	Do you think information in this property could be refined by using other properties?	no	currently filled with MARC 500, 533 should be filled with MARC 245 $c, 250, 520, too (TEL)
				Do you notice "noise"in the data caused by the presence of labels coming from MARC21 labels?	no	
				Was there any other relevant data that could be mapped to a contextual resource (Agent, Place…)	no	

dcterms:tableOfContents	A list of subunits of the resource	no	no			
dcterms:provenance	A statement of changes in ownership and custody of the CHO since its creation. Significant for authenticity, integrity and interpretation.	no	no			perhaps these information will be added in the future; what about current provenance like holding institution, shelf mark?
dc:format	The file format, physical medium, or dimensions of the resource	yes	yes	The values in this property are related to the digital object	no	MARC 300 $b should not go into <dc:format>, but into <dcterms:extent>, together with the other subfields
dcterms:extent	The size or duration of the resource.	yes	yes			Subfields of MARC 300 should be put together in one element dcterms:extent, separated by ' ; ' e.g. MARC <datafield ind1=" " ind2=" " tag="300"> <subfield code="a">1 Mikrofilm</subfield> <subfield code="c">35 mm</subfield> </datafield> EDM <dcterms:extent>1 Mikrofilm ; 35 mm</dcterms:extent>
dcterms:medium	The material or physical carrier of the resource.	yes	no			e.g. microform is coded at MARC 007, pos. 0 = h
dc:identifier	An unambiguous reference to the resource within a given context.	yes	no			Identifier should not be changed or expanded MARC <controlfield tag="001">BDR-BV021681192-43915</controlfield> EDM <dc:identifier>a1114 - BDR-BV021681192-43915</dc:identifier>
dc:language(mandatory for objects of EDM type "TEXT")	A language of the resource. Encode as ISO 639-2. (Mandatory in EDM for objects of EDM type “TEXT”)	yes	yes			Language code is not available in all records

dc:publisher	An entity responsible for making the resource available.	yes	yes			Publication year should not be part of dc:publisher
dc:relation	A related resource.	yes	no			There are a lot of relations expressed in the MARC format, e.g. to DDC, other bibliographic records, country codes, authority and subject data, provenance ...
dcterms:hasFormat	The described resource pre-existed the referenced resource, which is essentially the same intellectual content presented in another format.	no	no			
dcterms:isFormatOf	A related resource that is substantially the same as the described resource, but in another format.	no	no			
dcterms:hasPart	The described resource includes the referenced resource either physically or logically.	yes				there is no example in the small dataset used

dcterms:isPartOf	The described resource is a physical or logical part of the referenced resource.	yes	no			**Part of a series** MARC <datafield ind1=" " ind2="0" tag="830"> <subfield code="a">Bad Wiesseer Tagungen des Collegium Carolinum</subfield> <subfield code="v">10</subfield> <subfield code="w">(DE-604)BV004255563</subfield> </datafield> **Part of a multi-part work** MARC <datafield ind1="1" ind2="0" tag="245"> <subfield code="a">Reise nach China durch die Mongeley</subfield> <subfield code="n">1</subfield> <subfield code="p">Reise nach Pekin : Mit einem Kupfer, einer Charte und einem Grundrisse</subfield> <subfield code="c">Aus d. Russ. übers von J. A. E. Schmidt</subfield> </datafield> <datafield ind1="0" ind2="8" tag="773"> <subfield code="w">(DE-604)BV001700520</subfield> <subfield code="g">1</subfield> </datafield> EDM <dc:title>Reise nach China durch die Mongeley</dc:title>
dcterms:hasVersion	A related resource that is a version, edition, or adaptation of the described resource.	no	no			
dcterms:isVersionOf	A related resource of which the described resource is a version, edition, or adaptation.	no	no			

dcterms:isReferencedBy	The described resource is referenced, cited, or otherwise pointed to by the referenced resource.	no	no			
dcterms:references	A related resource that is referenced, cited, or otherwise pointed to by the described resource.	no	no			
dc:rights	Information about rights held in and over the resource.	no	yes			
dc:subject(mandatory**)	The topic of the resource. (Note: Mandatory in EDM to supply one of dc:subject or dc:coverage or dc:type or dcterms:spatial)	yes	yes	Was there any other relevant data that could be mapped to a contextual resource (Agent, Place…)	yes	DDC, SSGN, RVK
dc:title (mandatory*)	A name given to the resource.	yes	yes			Information is incomplete in EDM; see remarks on dcterms:isPartOf 245 $a and $b are combined with ' '; should be with ' : '
dcterms:alternative	An alternative name for the resource.	yes	yes			
dc:type (mandatory**)	The nature or genre of the resource. (Note: Mandatory in EDM to supply one of dc:subject or dc:coverage or dc:type or dcterms:spatial)	yes	yes			
edm:isNextInSequence	edm:isNextInSequence relates two resources that are ordered parts of the same resource where the later part uses this property to point back to the former.	yes	no			The NextInSequence information can be accessed indirect, looking up all resources related to the higher hierarchical resource and than choosing the one with a numeration next to the given resource
edm:type (mandatory)	The Europeana material type of the resource	yes	yes	The element should contain one of this value: TEXT, VIDEO, SOUND, IMAGE, 3D?	yes	TEXT or IMAGE

2) The Aggregation

Property	Definition	Is this information available in the original data?	Is this information available after conversion to EDM?	Other questions		Remarks on the property/class conversion
ore:Aggregation	The ore:Aggregation class is the pivotal object between the edm:Provided and the edm:WebResource(s) associated with it. It is also the place where the metadata relating to this whole object will be recorded.	no	yes	Is this resource properly identified?	yes with a "fake" HTTP URI	
				Was there any other relevant data for the end user, present in the original but not in the EDM record?	no	
				Was any data incorrectly assigned to properties of this class?	If yes, please indicate in which properties	for edm:dataProvider, MARC 845 should be used, not "CL Bavaria"
edm:hasView	This property relates an ore:aggregation with a web resource providing a view of the associated edmProvidedCHO. This may be the source object itself in the case of a born digital cultural heritage object. Use where one CHO has several views of the same object. (e.g. a shoe and a detail of the label of the shoe)	no	no	Does the link point to the object image?	no	not used
edm:aggregated CHO (mandatory)	This property associates an ore:Aggregation with the cultural heritage object it is about.	no	yes	Does the link point back to the edm:ProvidedCHO?	yes	
edm:dataProvider (mandatory)	The name or identifier of the organisation that contributes data to Europeana	yes	no			MARC 845 should be used, not "CL Bavaria"

edm:isShownBy (mandatory*)	An unambiguous URL reference to the digital object on the provider's web site in the best available resolution/quality. * if edm:isShownAt is not provided	yes	yes	Does the link point directly to the object?	yes	
edm:isShownAt (mandatory*)	An unambiguous URL reference to the digital object on the provider's web site in its full information context. * if edm:isShownBy is not provided	no	no	Does the link point to the object in its context?	no	is not used
				Does the link point to the object mage?	no	is not used
edm:object	The URL of a thumbnail representing the digital object or, if there is no such thumbnail, the URL of the digital object in the best resolution available on the web site of the data provider from which a thumbnail could be generated. This will often be the same URL as given in edm:isShownBy.	yes	no	Does the link point to the object image?	no	use MARC 982 $t
edm:provider (mandatory)	The name or identifier or the organization that sends the data to Europeana, and this is not necessarily the institution that holds or owns the original or digitised object.	no	yes			

3) The WebResource

Property	Definition	Is this information available in the original data?	Is this information available after conversion to EDM?	Other questions	Remarks on the property/class conversion

edm:WebResource	A edm:WebResource is a digital representation of the edm:ProvidedCHO	yes	yes	Is this resource properly identified?	yes with a functional HTTP URI	
				Was there any other relevant data for the end user, present in the original but not in the EDM record?	if yes, please enter a short description of the fields missing?	MARC 856 $3, Digital publication year MARC 856 $3, Digital publisher (Provider)
				Was any data incorrectly assigned to properties of this class?	no	
dc:rights	Information about rights held in and over the resource.	no	no			
edm:rights (mandatory)	Information about copyright of the digital object as specified by isShownBy and isShownAt	no	yes	Does this statement apply to the digital object?	yes	

4) Agent class

Property	**Definition**	**Is this information available in the original data?**	**Is this information available after conversion to EDM?**	**Other questions**	**Remarks on the property/class conversion**	
edm:Agent	The class edm:Agent comprises people, either	yes	no	Is this resource properly identified?		not filled

	individually or in groups, who have the potential to perform intentional actions for which they can be held responsible.			Was there any other relevant data for the end user, present in the original but not in the EDM record?	if yes, please enter a short description of the fields missing?	GND data linked to providedCHO, MARC 100, 110, 111, 600, 610, 611, 700, 710, 711, 800, 810, 811
				Was any data incorrectly assigned to properties of this class?		
skos:prefLabel	The preferred form of the name of the agent.	yes	no			
skos:altLabel	Alternative forms of the name of the agent.	yes	no			
skos:note	A note about the agent e.g. biographical notes.	yes	no			
edm:begin*	The date the agent was born/established.	yes	no			
edm:end*	The date the agent died/terminated.	yes	no			

5) Place class

Property	**Definition**	**Is this information available in the original data?**	**Is this information available after conversion to EDM?**	**Other questions**		**Remarks on the property/class conversion**
edm:Place	A spatial location identified by the provider and named according to some	yes	no	Is this resource properly identified?	yes with a functional HTTP URI	Analysis of MARC 260 $a

	vocabulary or local convention.			Was there any other relevant data for the end user, present in the original but not in the EDM record?	if yes, please enter a short description of the fields missing?	GND data linked to providedCHO, MARC 651 Analysis of MARC 044 $a and $c
				Was any data incorrectly assigned to properties of this class?		
wgs84_pos:lat	The latitude of a spatial thing (decimal degrees).	no	no			can be deduced
wgs84_pos:long	The longitude of a spatial thing (decimal degrees)	no	no			can be deduced
skos:prefLabel	The preferred form of the name of the place.	yes	no			
skos:altLabel	Alternative forms of the name of the place.	yes	no			
skos:note	Information relating to the place.	yes	no			
dcterms:isPartOf	identifier of a place that the described place is part of.	yes	no			

6) Time class

Property	**Definition**	**Is this information available in the original data?**	**Is this information available after conversion to EDM?**	**Other questions**		**Remarks on the property/class conversion**
skos:Concept	A unit of thought or meaning that comes from	yes	no	Is this resource properly identified?		not filled

	an organised knowledge base (such as subject terms from a thesaurus or controlled vocabulary) where URIs or local identifiers have been created to represent each concept.			Was there any other relevant data for the end user, present in the original but not in the EDM record?	if yes, please enter a short description of the fields missing?	GND data linked to providedCHO, MARC 650, 630
				Was any data incorrectly assigned to properties of this class?		
skos:prefLabel	The preferred form of the name of the concept.	yes	no			
skos:altLabel	Alternative forms of the name of the concept.	yes	no			
skos:broader,	The identifier of a broader concept in the same thesaurus or controlled vocabulary.	yes	no			
skos:note	Information relating to the concept.	yes	no			

7) Concept class

Property	**Definition**	**Is this information available in the original data?**	**Is this information available after conversion to EDM?**	**Other questions**		**Remarks on the property/class conversion**
edm:TimeSpan	A period of time having a beginning, an end and a	yes	no	Is this resource properly identified?		not filled

	duration.			Was there any other relevant data for the end user, present in the original but not in the EDM record?	if yes, please enter a short description of the fields missing?	MARC 648, is to be interpreted
				Was any data incorrectly assigned to properties of this class?		
skos:prefLabel	The preferred form of the name of the timespan or period.	yes	no			
skos:altLabel	Alternative forms of the name of the timespan or period.	no	no			
skos:note	Information relating to the timespan or period.	no	no			
dcterms:isPartOf	The identifier of a timespan of which the described timespan is a part.	no	no			
edm:begin*	The date the timespan started.	yes	no			
edm:end*	The date the timespan finished.	yes	no			

Annex 3
Europeana Libraries WP5- Analysis of the validation questionnaire

1) ProvidedCHO

Property	Remarks on the property/class conversion		
	MARC 21	**UNIMARC**	**DC like formats**
edm:ProvidedCHO	The URI for all the data has been created automatically by The European Library and is not functional. Partners are wondering why we couldn't make this URI functional. The URI for the ProvidedCHO could be created based on an existing identifier in the data (the most persistent one in particular [Comment form BSB]		
owl:sameAs			
dc:contributor	The MARC data are providing rich data per contributors that are not fully exploited even at the contextual level. We need to improve the way we deal with these information Some MARC rules such as the inversion of person names are not applied. For instance: Dates (birth, death) + relation codes of contributor + the inversion of names of persons are not being applied. There is a loss of information from MARC to EDM MARC: <datafield ind1="0" ind2=" " tag="700"> <subfield code="a">Heinrich</subfield> <subfield code="b">II.</subfield> <subfield code="c">Römisch-Deutsches Reich, Kaiser</subfield> <subfield code="d">973-1024</subfield> <subfield code="0">(DE-588)118548255</subfield> </datafield> EDM: <dc:contributor>Heinrich</dc:contributor> For some set of data like the Bavarian State Library the agent	Some improvements are necessary for the conversion from UNIMARC to EDM: *addition of UNIMARC subfields $b $c $d $j $q $4 from 700 and 710. *addition of UNIMARC subfields $e and $4 for the contribution type/contributor role type [Comment from a1030] Unimarc is not processed into sensible information [Comments for a0005]	

	could give rise to an Agent entity based on the linked data version of GND. The identifier of the resource in GND is already available in the original Marc data. [Comment for a114]		
dc:creator	As for contributor there is a data loss in the conversion from MARC to EDM. In some case we could improve the way we deal with authority records. GND could be interesting to experiment since a linked data version exists. For instance: Dates (birth,death) + relation codes of contributor + the inversion of names of persons is not being applied. There is a loss of information from MARC to EDM MARC: <datafield ind1="0" ind2=" " tag="700"> <subfield code="a">Heinrich</subfield> <subfield code="b">II.</subfield> <subfield code="c">Römisch-Deutsches Reich, Kaiser</subfield> <subfield code="d">973-1024</subfield> <subfield code="0">(DE-588)118548255</subfield> </datafield> EDM: <dc:contributor>Heinrich</dc:contributor> [Comment for a114]	Unimarc is not processed into sensible information [Comment for a0005] Some improvements are necessary for the conversion from UNIMARC to EDM: *Addition subfields from UNIMARC 100 (subfields $b $c $d $j $q $4) and 110 (subfields $b $c $d $n) Example: Tomàs,|cd'Aquino, sant,|d1225?-1274 --> is simplified just to "Tomas" (Ex b18498590) [Comment for a1030]	dc:creator contains mainly literal strings that can't be used directly for the creation of contextual entity. However it could be used to match against external data sources as has been done for edm:Place (but see caveat in edm:Agent). [comment for a444]
dc:coverage (mandatory**)		*Some improvements are necessary for the conversion from UNIMARC to EDM: 752 [Comment for a1030 ,a0005]	
dcterms:spatial (mandatory**)	This information is quite often available in the data but is not mapped properly to EDM. The spatial information is often mapped as a subject. Since EDM allows a finer grain of description it would be important to use more specific properties instead of generic ones. For instance <dc:subject>Wales [lctgm]</dc:subject><dc:subject>Amlwch		The contextual class edm:Place has been created but not in all cases. For example, in records referring to Gloucester City (006GLO10370CC35U00005000, 006GLO10370CC35U00006000, etc.), Barbados

	[lctgm]</dc:subject>. [Comment for a 1024] Some improvements are necessary for the conversion from MARC to EDM: *Addition of MARC 651 [Comment for a114]		(010BB0001861S70U00001001, 010BB0001861S70U0000100, etc.), United Kingdom (010GB0001913S19U00001001, 010GB0001913S19U00001002) but not Durham City (006ENG10370CC35U00011000) nor York City (006ENG10370CC35U00008000), etc. It is not clear whether edm:Place is generated with dcterms:spatial. [Comment for a444].
dcterms:temporal	The time information available in this element in not always related to the ProvidedCHO. Information are sometimes mixed with information about the digital object. Different variations of the same dates are sometimes available. It would be better to select only one. Some improvements are necessary for the conversion from MARC to EDM: *Addition of MARC 648 [Comment for a114]		
dc:date	There is usually no issue related to the differentiation of data related to the digital object and the ProvidedCHO. There is just sometimes a mix between the date of creation and the date of publication. We should try to always use the refinement when possible.		
dcterms:issued	This information is quite often mapped as a date. This date is often provided in the original data but is not used in EDM. [Comment for a1030]		
dcterms:created	This information is quite often mapped as a date. This date is often provided in the original data but is not used in EDM		

dc:description (mandatory*)	Information coming from MARC 546, 561, 541, 585 has been mapped to description but don't have a specific label which makes the information useless for a user. Some improvements are necessary for the conversion from MARC to EDM: *Addition should be filled with MARC 245 $c, 250, 520 [Comment for a114]	Some improvements are necessary for the conversion from UNIMARC to EDM: 561 can be mapped to dcterms:provenance but in most of the situation this filed doesn't have a label and should be removed from the mapping. [Comment for a1030]	The description field contains lot of information which would be better described by using another property such as dcterms:abstract…
dcterms:tableOfContents	In mapping from ESE to EDM this property is not mapped in EDM.	Some improvements are necessary for the conversion from UNIMARC to EDM: *Addition of UNIMARC field 327$a *Addition of UNIMARC 505	
dcterms:provenance	Some improvements are necessary for the conversion from MARC to EDM: *Addition of Marc 541	Some improvements are necessary for the conversion from UNIMARC to EDM: *Addition of 561 [Comment for a1030]	In mapping from ESE to EDM this property is not mapped in EDM.
dc:format	Some improvements are necessary for the conversion from MARC to EDM: *Addition <datafield ind1=" " ind2=" " tag="300"> [Comment for a1018] *MARC 300 $b should be removed from format. should not go into <dc:format>, but into <dcterms:extent>, together with the other subfields [Comment for a114]	Some improvements are necessary for the conversion from UNIMARC to EDM: *Remove from <dc:format> II. (MARC 300$b) [Comment for a1030]	Data related to the digital object and the ProvidedCHO are mixed. The problem does not necessary come from the mapping but also from the original data
dcterms:extent	Some improvements are necessary for the conversion from MARC to EDM: *Addition of MARC 300 $b * Subfields of MARC 300 should be put together in one element dcterms:extent, separated by ' ; ' e.g. MARC <datafield ind1=" " ind2=" " tag="300"> <subfield code="a">1 Mikrofilm</subfield> <subfield code="c">35 mm</subfield> </datafield> [Comment for a114] [Comment for a444] In some case extent contains a set of		

	different values Length, Width and Unit of measurement. These three elements were all mapped to three separate dcterms:extent in ESE and consequently all mapped to three separate dcterms:extent in EDM. e.g. <dcterms:extent>41</dcterms:extent> <dcterms:extent>26.3</dcterms:extent> <dcterms:extent>Centimetres</dcterms:extent>. For display for the user, if possible, it would be more useful to concatenate all data in one instance of dcterms:extent, e.g. <dcterms:extent>41 x 26.3 centimetres</dcterms:extent>. This is part of a more generic problem on how to record extent and dimensions in a machine-readable form. An ALA Taskforce on Machine-Actionable Data Elements in RDA Chapter 3 is looking into these issues (see http://www.rda-jsc.org/docs/6JSC-ALA-17.pdf)		
dcterms:medium	There are few issue with the MARC labels e.g. microform is coded at MARC 007, pos. 0 = h [Comment for a114] Some improvements are necessary for the conversion from MARC to EDM: *could use <datafield ind1=" " ind2=" " tag="300"> values from original [Comment for a1018]	Some improvements are necessary for the conversion from UNIMARC to EDM: *Add 340 [Comment for a1030]	
dc:identifier	MARC records contain different types of identifiers. The conversion to EDM is keeping all the identifiers, even duplicating some. It seems that the internal object model should better address the different type of identifiers and their mappings. The different types of identifiers need to be distinguished. For instance: Source contains 4 such elements (one related to item bibligoraphic identifier, another relates to collection identifier and the other two relate to the digital object). EDM contains 5 such elements (one related to item bibliographic identifier, another relates to collection identifier and the other three relate to the digital object (one of which is repeated twice (<dc:identifier>llgc-id:1123576</dc:identifier>)). [Comment for a1024] The identifier should also not be changed or expended	There are other information that can be considered as identifier, specially the shelfmark (UNIMARC 945$a), esential for manuscripts. Another identifier can be the url (UNIMARC 962$u)[Comment for a1030]	

	For instance MARC<controlfield tag="001">BDR-BV021681192-43915</controlfield> EDM<dc:identifier>a1114 - BDR-BV021681192-43915</dc:identifier> [Comment for a114] Some improvements are necessary for the conversion from MARC to EDM: *<datafield ind1=" " ind2=" " tag="907"> should not be used. This refers to the digital object. [Comment for a1018]		
dc:language(mandatory for objects of EDM type "TEXT")	Language code is not available in all the records		
dc:publisher	The publisher year has been mixed with the publisher information. Is it intentional? Data necessary for dc:publisher have been moved to dc:description. Some improvements are necessary for the conversion from MARC to EDM: *In MARC the whole 260 field shouldn't be mapped but just 260|a [Comment for a1018]	Some improvements are necessary for the conversion from UNIMARC to EDM: * The EDM conversion considersconsiderig the full 210 UNIMARC field (210$aplace of publication$cPublisher$dData of publication); it should consider only 210$c.	
dc:relation	Very few relations have been expressed in EDM. There are a lot of relations expressed in the MARC format, e.g. to DDC, other bibliographic records, country codes, authority and subject data, provenance … but they are not expressed in EDM.		
dcterms:hasFormat			
dcterms:isFormatOf			
dcterms:hasPart	A solution for having a value here would be to generate these properties from the table of content		
dcterms:isPartOf	The mapping needs to be refined. Only a small subset of the source data is mapped (only tile instead author+ identifier) . Part of a series MARC <datafield ind1=" " ind2="0" tag="830"> <subfield code="a">Bad Wiesseer Tagungen des Collegium Carolinum</subfield> <subfield code="v">10</subfield> <subfield code="w">(DE-604)BV004255563</subfield>	Some improvements are necessary for the conversion from UNIMARC to EDM: *Addition of all 830+ subfields. [Comment for a1030]	

	</datafield> Part of a multi-part work MARC <datafield ind1="1" ind2="0" tag="245"> <subfield code="a">Reise nach China durch die Mongeley</subfield> <subfield code="n">1</subfield> <subfield code="p">Reise nach Pekin : Mit einem Kupfer, einer Charte und einem Grundrisse</subfield> <subfield code="c">Aus d. Russ. übers von J. A. E. Schmidt</subfield> </datafield> <datafield ind1="0" ind2="8" tag="773"> <subfield code="w">(DE-604)BV001700520</subfield> <subfield code="g">1</subfield> </datafield> EDM <dc:title>Reise nach China durch die Mongeley</dc:title> [Comment for a114] The range of dcterms:isPartOf is an RDF resource but most of the time libraries data doesn't support this, i.e. it is a literal string [Comment 0444]		
dcterms:hasVersion			
dcterms:isVersionOf			
dcterms:isReferencedBy			
dcterms:references			
dc:rights	The licence is sometimes wrong : Some improvements are necessary for the conversion from MARC to EDM: *Addition of <datafield ind1=" " ind2=" " tag="593"> [Comment for a1018]		
dc:subject(mandatory **)	Attributes for subjects are not interpreted and can’t be used to generate a concept entity [Comment a444]		

dc:title (mandatory*)	Title information is incomplete when coming from MARC. Some improvements are necessary for the conversion from MARC to EDM: *245 $a and $b are combined with ' '; should be with ' : '		
dcterms:alternative	In mapping from ESE to EDM this property is not mapped in EDM.	Some improvements are necessary for the conversion from UNIMARC to EDM: *Addition of subfields $l $n $p from UNIMARC 240 and 730 *Addition of UNIMARC 700 (subfields $t $n $p) and 740 (subfields $a $n $p [Comment for a1030]	
dc:type (mandatory**)	Some improvements are necessary for the conversion from MARC to EDM: *Addition MARC 655 [Comment for a1018]		In conversion from ESE to EDM, occurrence of dc:type are duplicated. One contains an xml lang:attribute. Two elements are redundant and we should keep one. instances <dc:type>StillImage</dc:type> and <dc:type xml:lang="en">Still Image</dc:type>. [Comment for a444]
edm:isNextInSequence	This property is not available in EDM but could be deduced from the numeration for related resources.		
edm:type (mandatory)			

2) WebResource

Property	Remarks on the property/class conversion
edm:WebResource	We should improve the way we select the WebResources. The selected resources are not always relevant. WebResource are sometimes different from the hasView in aggregation.
dc:rights	Some improvements are necessary for the conversion from MARC to EDM: *Addition <datafield ind1=" " ind2=" " tag="593"> [Comment for a1018] The <dc:rights>Copyright@Britih Library Board</dc:rights> has wrongly been attached to the edm:ProvidedCHO instead of the edm:WebResource. But this would be really hard to automatised since dc:rifghts cvould be used in different places.

edm:rights (mandatory)	Some improvements are necessary for the conversion from MARC to EDM: *Addition <datafield ind1=" " ind2=" " tag="593"> [Comment for a1018]

3) Aggregation

Property	Remarks on the property/class conversion
ore:Aggregation	
edm:hasView	edm:HasView not always contain the same URL than the ones described in isShownBy or At.
edm:aggregatedCHO (mandatory)	The aggregated CHO properties always points back to the CHO
edm:dataProvider (mandatory)	It seems that the values available in this property come from a kind of controlled vocabulary. They are however not understandable from a user or a provider perspective. Some improvements are necessary for the conversion from MARC to EDM: *Addition of MARC 845
edm:isShownBy (mandatory*)	The URL mapped are not always pointing to the digital resource neither relevant.
edm:isShownAt (mandatory*)	In mapping from ESE to EDM this property is not mapped in EDM.
edm:object	Some improvements are necessary for the conversion from MARC to EDM: *Addition of MARC 982 $t
edm:provider (mandatory)	
edm:rights	This element is missing from the data and is mandatory. The value in the field should be the same than the right of the WebResource chosen in edm:IsShownBy and not the licence of the metadata.

4) Agent

Property	Remarks on the property/class conversion

edm:Agent	In general we have very few agent entities in the converted in the EDM data, and this mainly because either information is not in the original data but in another authority file or because of the lack of identifier. It would be necessary to improve the way we use the following elements: MARC 100, 110, 111, 600, 610, 611, 700, 710, 711, 800, 810, 811 [Comment for a114] <dc:creator> contains a literal string: for a person, surname, first name/initials; for a corporate body, name string. Data was transcribed from the resource, and not authority controlled. The data string could be matched to headings in id.loc.gov and the id.loc.gov URI could be used as was done for edm:Place and Geonames. However, the data contained in dc:creator would need additional processing before being matched. The original BL metadata used a custom list of roles, e.g. <item_creator>Yeakell, Thomas Jr</item_creator> <creator_role>Draughtsman</creator_role>. When mapping to ESE, it was decided to include the creator role within the dc:creator element so as not to lose useful information, i.e. <dc:creator>Draughtsman : Yeakell, Thomas Jr</dc:creator>. Whilst this is useful from a display perspective, it is less so from a data and semantics perspective. [Comment for a444]
skos:prefLabel	
skos:altLabel	
skos:note	
edm:begin*	
edm:end*	

5) Place

Property	Remarks on the property/class conversion
edm:Place	Spatial information are available in the source data but are spread within the dc:subject properties along the subject terms. When a edm:Place is created the EDM version has often a broader geographical area than the origina l record: e.g. London instead of Bermondsey, UK instead of England Some improvements are necessary for the conversion from MARC to EDM: *Interpretation of MARC 044 $a and $c and MARC 260 $a and MARC 651 [Comment for a114] As mentioned above in edm:ProvidedCHO, the edm:Place class has only been created in certain cases. Matches are sometimes wrong. The mapping should either take the value of the dcterms:spatial property or use the identifiers available in the data such as TGN.
wgs84_pos:lat	

wgs84_pos:long	
skos:prefLabel	When available this property doesn't have a language tag. It is very important to have one. i.e. França, Napoli, Florence...). The label for the place in the original metadata in dcterms:spatial is sometimes different from the skos:prefLabel (maybe this is because the original metadata creators used a different gazetter than Geonames). So sometimes the two labels match, e.g. Barbados; sometimes they don't, e.g. <dcterms:spatial>United Kingdom</dcterms:spatial> and <skos:prefLabel>United Kingdom of Great Britain and Northern Ireland</skos:prefLabel>
skos:altLabel	
skos:note	
dcterms:isPartOf	

6) Concept

Property	**Remarks on the property/class conversion**
skos:Concept	Some improvements are necessary for the conversion from MARC to EDM: *Interpretation of MARC 650, 630 [Comment for a114] The original record contains <dc:subject xsi:type="dcterms:LCSH">. This could allow for the creation of a skos:Concept, using the URI (either for the full LCSH string or partially) defined at id.loc.gov. The original record contains <dc:subject xsi:type="dcterms:DDC">. This could in theory allow for the creation of a skos:Concept (in theory because our metadata doesn't include the Dewey edition, which makes things trickier). However, should other datasets include that type of information (e.g. DDC edition 22 or 23, UDC, etc), a skos:Concept could be created. However, this would require the addition of the property skos:notation to record the classification number. [Comment for a444]
skos:prefLabel	
skos:altLabel	
skos:broader,	
skos:note	

7) Time

Property	**Remarks on the property/class conversion**

edm:TimeSpan	Some improvements are necessary for the conversion from MARC to EDM: *Interpretation of MARC 648 [Comment for a114]
skos:prefLabel	
skos:altLabel	
skos:note	
dcterms:isPartOf	
edm:begin*	
edm:end*	