DELIVERABLE

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

D5.1 Report on the alignment of library metadata with the European Data Model (EDM)

Version 2.0

Authors:

Anila Angjeli	Bibliothèque nationale de France
Martin Baumgartner	Bayerische Staatsbibliothek
Sally Chambers	The European Library
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

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	24.11.11	Sally Chambers	The European Library	First draft outline of D5.1
0.2a	14.12.11	Valentine Charles, Robina Clayphan	The European Library	Text added to Section 7. Sections 5.3 and 8 added.
0.2b	18.12.11	Stefanie Rühle, Alexander Jahnke	CERL	Chapter 5: Europeana Libraries data model for monographs
0.3	19.12.11	Sally Chambers	The European Library	Version prepared for the Europeana Libraries Appraisal Group
1.0	30.12.11	Sally Chambers	The European Library	Final version incorporating feedback from the Europeana Libraries Appraisal Group: Rodolphe Bailly, Emmanuelle Bermès and Antoine Isaac.
1.1	04.05.12	Robina Clayphan	The European Library	Revised version with major reviewer issues addressed
2.0	08.11.2012	Robina Clayphan, Stefanie Rühle	The European Library, CERL	Revised version following discussion at May meetings and feedback from Europeana and further comment.

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This deliverable contains original unpublished work except where clearly indicated otherwise.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, *D5.1: Report on the alignment of library metadata with the Europeana Data Model,* has been written by Work Package 5 (WP5) of the *Europeana Libraries* project[1]. The aim of the report is to describe how library metadata can be aligned with the Europeana Data Model (EDM).

The report begins by positioning this work within the current library metadata context, in particular in relation to the recent developments related to Resource Description and Access (RDA), The Library of Congress's Bibliographic Framework Transition Initiative and the work of the W3C Library Linked Data Incubator Group on the global interoperability of library data on the Web. It continues by outlining the methodology used to prepare the groundwork of this report, from the creation of the Europeana Libraries Metadata Working Group, to the definition of the workplan. Originally focusing on mapping from the key metadata formats to EDM, it was decided to change approach to define a model showing how library materials could be represented in EDM.

The deliverable addresses monographs, multi-volume works and serials and also distinguishes between born digital and digitised objects. However, the need to address other library materials such as still images, videos, manuscripts/incunabula, theses, maps, scores, plates' materials has been noted, however, primarily due to time restrictions, is beyond the scope of this report.

For the key library materials selected - monographs, multi-volume works and serials – the EDM classes and the related properties for each class are described. Optimal properties, for the ideal implementation solution, as well as those available for Europeana's first implementation of EDM are described. In addition, specific case studies for rare books, event-based description of textual resources and the clustering of similar textual resources are also considered.

This deliverable should be considered as a milestone in an ongoing process to define how EDM can be used with library resources. The work so far focused on the modelling of the data. Even though the group worked with real data, the model needs to be validated at a technical level in year 2 of the project. Following the validation of the deliverable, due for completion in June 2012, the alignment of library metadata to the validated model can begin.

An important element of the process of creating this deliverable was the need to work closely with the EDM documentation. The work of the group provided useful feedback for the Europeana team and will contribute to the ongoing development of EDM. A specific section of this deliverable is dedicated to the provision of feedback on EDM to Europeana.

The need for further research is highlighted. In addition to undertaking the modelling of additional library materials, with the next priorities being theses, the manuscripts and maps, it is important to ensure that 'EDM for libraries' remains interoperable with the other domain-level of implementations of EDM.

[1] http://www.europeana-libraries.eu/

2. Introduction

This deliverable, *D5.1 Report on the alignment of library metadata with the Europeana Data Model* is delivered by WP5 as part of the task 5.1.1 Aligning national library metadata with the Europeana Data Model:

"5.1.1 KB and Europeana will investigate how national library metadata can be aligned with the Europeana Data Model (EDM). This investigation will be validated using a sample of metadata from national libraries that has already been provided to Europeana in Europeana Semantic Elements (ESE) format. A report recommending best practice of aligning library metadata to the Europeana Data Model (EDM) will be produced by Month 12."

As described in the description of work of Europeana Libraries, the report on the alignment of library metadata with the Europeana Data Model will describe how library metadata can be aligned with the Europeana Data Model (EDM). This report will describe best practices and will address recommendations to Europeana.

3. Library metadata context

The Europeana Data Model for Libraries is being developed at an exciting time within the library metadata landscape. During 2011, three major reports that impact on the library metadata context have been published.

Following the publication of the new metadata content standard, Resource Description and Access (RDA)[2] in June 2010, RDA, which builds on the Anglo-American Cataloguing Rules, 2nd Edition (AACR2), is intended to provide 'a comprehensive set of guidelines and instructions on resource description and access covering all types of content and media'[3]. RDA represents a huge shift in how items in the bibliographic universe will be described. Following the criticisms the Working Group on the Future of Bibliographic Control issued its report, *On the Record*, on January 9, 2008[4]. In it one recommendation--3.2.5--was notable in that it called for a suspension of work underway on RDA. The Working Group suggested that further development work on Resource Description and Access (RDA) be suspended until a business case had been articulated, benefits demonstrated, and there had been better testing of FRBR (Functional Requirements for Bibliographic Records) as it relates to RDA.[5] This recommendation led to the Library of Congress, the National Agricultural Library and the National Library of Medicine to charge ' the U.S. RDA Test Coordinating Committee 'to devise and conduct a national test of *Resource Description and Access (RDA)*[6]. The report of this committee, which was publically released in June 2011[7], stated *'the RDA Test Coordinating Committee recommended that the national libraries adopt RDA with certain conditions and that implementation will not occur before January 1, 2013. '*

[2] Announcement of the launch of Resource Description and Access: http://www.rda-jsc.org/rdapublish.html

[3] Further information about Resource Description and Access (RDA) is available at: http://www.rdatoolkit.org/

[4] On the Record, Report of The Library of Congress Working Group on the Future of Bibliographic Control http://www.loc.gov/bibliographic-future/news/lcwg-ontherecord-jan08-final.pdf

[5] Response of the Library of Congress, the National Agricultural Library, and the National Library of Medicine to the RDA Test Coordinating Committee: http://www.loc.gov/bibliographic-future/rda/rda-execstatement-13june11.pdf

[6] Further information about the US RDA Test is available at: http://www.loc.gov/bibliographic-future/rda/

[7] Testing Resource Description and Access: http://www.loc.gov/bibliographic-future/rda/

The second significant announcement of 2011, was The Library of Congress launching the Bibliographic Framework Transition Initiative, in May 2011.[8] The aim of the initiative is to undertake 'a review of the bibliographic framework to better accommodate future needs. A major focus of the initiative will be to determine a transition path for the MARC 21 exchange format in order to reap the benefits of newer technology while preserving a robust data exchange that has supported resource sharing and cataloging cost savings in recent decades.' In October 2011, the Library of Congress issued an initial plan and time schedule[9] for this work to be undertaken. According to the indicative schedule it will be late 2013 or early 2014, before this work is completed.

The third and final report of the year was the Final Report of the W3C Library Linked Data Incubator Group[10] which was published in October 2011. The mission of the group, was "to help increase global interoperability of library data on the Web, by bringing together people involved in Semantic Web activities — focusing on Linked Data — in the library community and beyond, building on existing initiatives, and identifying collaboration tracks for the future." [11]

In addition to these high-level international initiatives, similar initiatives at the institutional level have taken place. For example, the creation of the British Library Data model[12] to support the creation of a linked data version of the British National Bibliography[13] is one such initiative.

Within the Europeana family, liaison with projects such as Heritage of the People's Europe (HOPE), who developed a common HOPE Metadata Structure[14], and in particular the HOPE library profile, which is fully-interoperable with the Europeana Metadata Framework, has been undertaken.

4. Methodology

EDM is a complex data model which differs from the other well-known libraries metadata formats in that it was designed from the start to be semantic web compliant. In order to handle the complexity of the task, WP5 has formed a group of experts. This group has managed to reach a consensus and has created a work methodology in order to produce this deliverable.

4.1. The Europeana Libraries Metadata Working Group

The first task of the WP5 was to create a Metadata Working Group which would ensure that Europeana Libraries can produce a high quality profile for libraries data using the Europeana Data Model. This group undertakes the different tasks of WP5 as described in the description of work. The terms of reference (**see Annex 1**) define the framework of WP5.

---

[8] Further information about the Bibliographic Framework Transition Initiative is available at: http://www.loc.gov/marc/transition/

[9] The initial plan for the Bibliographic Framework Transition Initiative is available at: http://www.loc.gov/marc/transition/news/framework-103111.html

[10] W3C Library Linked Data Incubator Group Final Report is available at: http://www.w3.org/2005/Incubator/lld/XGR-lld-20111025/

[11] Further information about the W3C Library Linked Data Incubator Group is available at: http://www.w3.org/2005/Incubator/lld/

[12] At the time of publication, the latest version of the British Library Data Model is v 1.2, 1 December 2011, which is available online at: http://www.bl.uk/bibliographic/pdfs/bldatamodelbook.pdf

[13] Further information about the creation of a linked data version of the British National Bibliography is available in a presentation by Neil Wilson, Establishing the connection: creating a linked data version of the BnB, is available online at: http://www.slideshare.net/nw13/establishing-the-connection-creating-a-linked-data-version-of-the-bnb

[14] Further information about the Common Hope Metadata Structure is available at: http://www.peoplesheritage.eu/pdf/D2_2_Metadata%20Structure.pdf

The Metadata Working Group comprises one representative from each of the organisations participating in WP5. Members of The European Library Metadata Working Group have joined the group. They represent namely the French National Library, the British Library and the German National Library. This group brings together a group of specialists who have expertise in library metadata standards and who are familiar with the issues raised by aggregation of data for Europeana and with the diversity of resources in libraries.

One of the first tasks of the Metadata Working Group leading to the production of this deliverable was to understand the key concepts of EDM and define how to use them within the library environment.

Since Europeana was still in the process of refining the EDM documentation, the group had to keep track of Europeana's latest developments. The communication with the Europeana Scientific team has been crucial.

Over the year the group had several physical and remote meetings[15] during which a consensus was reached. Always working with data samples, the group has decided to tackle issues with pragmatic solutions.

Since the group also has the task of promoting and building consensus on the adoption of standards related to the aggregation of digital content within the Europeana network and beyond, several presentations have been done during conferences (**see Annex 2**). WP5 has also been in contact with other projects, which have brought expertise to the group, such as the Heritage Of People of Europe (HOPE) project that is working on a library profile for their cross-domain portal.

4.2. Definition of the workplan

The first approach of the group was to start from the most common metadata standards used in libraries and the legacy data available and see how they could be mapped to EDM. This first ''metadata to metadata approach'' was based on the analysis of MARC21, UNIMARC, MODS and Dublin Core based formats records. With this approach, the discussions seemed to remain rooted in discussions related to FRBR, so we decided that another approach was needed.

The group agreed that the best way to move forward was to define a model showing how library materials could be represented in EDM first. This approach has been chosen as being the best one to identify the gaps within the current EDM specifications which could be then addressed as recommendations for Europeana at the end of this deliverable. The model defined by Europeana Libraries may differ according to the different types of library materials considered.

The initial requirements for EDM library profile are therefore based and specified on the type of materials held by libraries. In the first instance, this deliverable will address monographs, multi-

[15] The Europeana Libraries Metadata Working Group had several meetings in 2011:
Physical meetings: 31st May, 6th July, 29th September, 16th November.
Remote meetings: 16th August, 2nd September, 24th September, 8th November.

volume works and serials. A specific distinction can be made between born digital and digitised objects. This deliverable will not address still images, video, manuscripts/incunabula, theses, maps, scores, plates' materials which will be the focus of this group in a second instance (within the Europeana Libraries project or another one).

The EDM profile for libraries is described as an output model for Europeana Libraries aggregation infrastructure to Europeana. This model is compliant with EDM, meets the requirements of the Europeana Libraries mapping process developed and implemented by WP4 and ensures that the mapping from different libraries metadata standards is consistent.

We define the following requirements for the EDM library profile data within Europeana Libraries:

- The European Library is the domain aggregator for libraries to Europeana. The EDM library profile should embed the specificities of library materials and data.
- Europeana Libraries is delivering metadata describing digital resources to Europeana. Therefore the EDM library profile will use the EDM classes and properties described in the EDM specifications 5.2.2.
- Europeana Libraries metadata shall be used in a linked data context. Therefore we will use a linked data compliant model namely RDF.
- Information about digital resources shall be interlinked in and across library domains. The EDM profile for libraries should use properties supporting the usage of URIs.
- Europeana Libraries will extend the model for further types of resources. We should then use a flexible entity-relationship model that can be reconfigured for the different types of resource. (Recognising that RDF can be seen as such a model.)

The next step in this deliverable is to define the EDM classes and properties which can be used in the library profile.

5. Europeana Data Model for monographs

In the traditional library world, mapping is understood as "the literal translation of elements from one schema to another".[16] In this world the description of library objects (books, journals etc.) is record based. Such a record, which can be seen as a closed box, is a defined set of description elements that can include information about:

- the real world object or item (e.g. the book in a shelf)
- the bibliographic object (e.g. the edition representing the entirety of all identical copies of a publication)
- the digital representation of the real world object (e.g. a digital copy of the book)
- the record itself (e.g. information when the record was created and by whom).

In contrast to this closed box model, the Europeana Data Model is a statement based model where the description of library objects covers separate entities and their relationships on different abstract levels. In this world we therefore differentiate between:

[16] Dunsire G. et al. (2011): A Reconsideration of Mapping in a Semantic World, p. 26 , see:. http://dcevents.dublincore.org/index.php/IntConf/dc-2011/paper/view/52/6.

- the description of the information (the entirety of all identical copies of a book) and the information carrier (the book in the shelf)
- the description of the real world object (the book) and its digital representation (a digital copy of this book)
- the description of the object described (the book) and the object describing it (the metadata)

And we interlink these entities and express these links with RDF statements.[17]

Mapping a traditional metadata record to EDM therefore means that the closed box has to be opened, the record will be taken apart and the different entities described in one record will be identified as EDM Classes and related to each other. Following this, the metadata elements used in the record will be assigned to the entities they belong to. After which, these elements will be mapped to the EDM Properties that are allowed to be used with the used EDM Classes. To support this process we developed a model that defines what EDM Classes should be used in the Europeana Libraries context, what relations need to be expressed and what EDM Properties should be used to describe these relations.

5.1. Definition of the scope

In order to define the basic scope we use an Entity-Relationship Model. According to the Singapore Framework such a model defines the basic entities and the fundamental relationships between these entities.[18] As EDM is the basis for our model we must select entities compliant to the following EDM Classes:

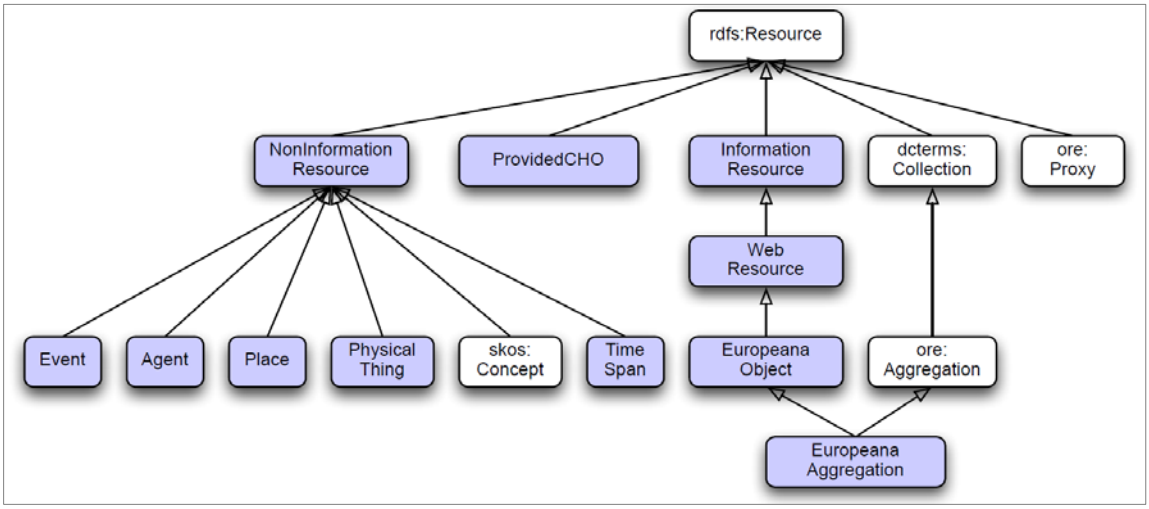

Figure 1: EDM Classes (Source: EDM v5.2.2 http://pro.europeana.eu/edm-documentation)

However, it is worth noting that only seven of these Classes will be included in the first EDM implementation by Europeana. These are the core classes – edm:ProvidedCHO, edm:WebResource

[17] See http://www.w3.org/TR/2004/REC-rdf-concepts-20040210/
[18] See: http://www.dublincore.org/documents/singapore-framework/.

and ore:Aggregation – and the following contextual classes: edm:Agent, edm:Place, edm:TimeSpan, skos:Concept.[19]

5.1.1. Definition of the entities used to describe text resources

In compliance with the first EDM implementation, the **basic entities** used for the mapping of monograph metadata descriptions in Europeana Libraries are:

edm:providedCHO

An edm:ProvidedCHO is "the cultural heritage object that Europeana collects descriptions about". This class "was especially intended to fit the cases where edm:PhysicalThing cannot be used as the type of the resource" (see EDM v.5.2.2. Page 13). But as a result of the heterogenity of objects described in museums, archives and especially libraries - with their traditional distinction between editions and their constituent items - there are diverging conceptions of what a cultural heritage object might be. Therefore we consider it necessary to refine the definition of edm:ProvidedCHO in context of the objects that it is standing for.

For published textual resources (books, articles, journals, newspapers etc.) we recommend the following refinement:

An edm:ProvidedCHO is the cultural heritage object Europeana collects descriptions about. In the context of published textual resources (library material) this can be interpreted as the Item in the FRBR[20] model - a concrete entity (like a book or a web resource if it is born digital) that carries the text. But traditionally for published textual resources the library world distinguishes between the real world object (the item) and the the bibliographic object (the edition). The edition represents the entirety of all identical copies of a publication and, in FRBR terms, can be seen as the union of the Work, Expression and Manifestation levels. The working group initially proposed to adopt the edition level as the focus of the edm:ProvidedCHO because most library materials are described that way. This led to the problem of how to handle descriptions that are actually at the item level. This was an serious issue for many participants whose collections contain many rare books that are only described at the item level.

To accommodate this, a separate model was proposed that could be used for such resources. Feedback from Europeana has clarified the situation however: the ProvidedCHO is a functional type that implies nothing about the nature of the resource. It can represent any object of interest that would appear as one item in a result list from a query on the Europeana portal. Even from one domain ProvidedCHOs can be an edition level resource or an item level resource. 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,

[19] See: Europeana Data Model Mapping Guidelines from 27.10.2011, http://pro.europeana.eu/edm-documentation.
[20] See: Functional Requirements for Bibliographic Records http://www.ifla.org/files/cataloguing/frbr/frbr_2008.pdf

an Item level ProvidedCHO could have an edm:realises link to a ProvidedCHO that represents the edition level. The original rare book model has been preserved for reference in **Annex 4**.

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. Therefore we 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.[21]

The notion of edition applies to the primary publication process and does not include the process of digitisation (true to the original layout) printed material. Large parts of the library domain treat a digitisation as a different edition, which is not the case here. Any information about the digital representation(s) of an edition – whether it is born digital or digitised - can be provided using the class edm:WebResource. Information about the physical data carriers of an edition cannot be provided in the first EDM implementation and is currently accessible only by linking to the metadata description on the provider's website.

edm:WebResource

EDM defines edm:WebResource as "Information Resources that have at least one Web Representation and at least a URI." (see EDM v.5.2.2. Page 14). In this model the edm:WebResource is a digital representation of an item (i.e. of a printed resource or a born digital). All information about this representation – whether it is born digital or digitised – has to be provided using the class edm:WebResource.

ore:Aggregation

"A set of related resources (Aggregated Resources), grouped together such that the set can be treated as a single resource. This is the entity described within the ORE interoperability framework by a Resource Map." (see EDM v.5.2.2. Page 6) In compliance with the EDM Model the **fundamental relationships** between edm:WebResource and edm:ProvidedCHO are realised by using the ore:Aggregation class. Therefore, following the basic EDM pattern, the core of the Europeana Libraries Entity Relationship Model defines two relations:

- The edm:aggregatedCHO between the ore:Aggregation and the edm:ProvidedCHO
- The edm:hasView between the ore:Aggregation and the edm:WebResource

[21] 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

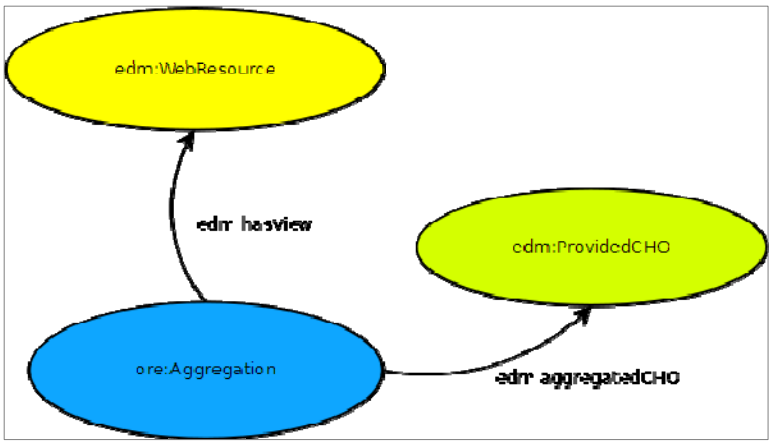

Figure 2: Europeana Libraries Entity-Relationship Model for monographs

The descriptive properties appropriate to each entity (shown in the tables in section 5.2) are then attached to each of these nodes.

This model is really simple and flexible and has the following advantages:

- It can be used for both digitised and born digital objects.
- It can be easily expanded for further entities – e.g. the classes of non-information resources like edm:Event, edm:Agent, edm:Place, edm:PhysicalThing, edm:TimeSpan and skos:Concept may be included in this model in a next step.
- It can be used to describe the sometimes really complex hierarchical structure of multi-volume works. This is achieved by connecting this core group of entities with a small set of properties created to express structural relationships (edm:isNextInSequence and dcterms:isPartOf) according to the structure of the object being described. This is demonstrated in the multi-volume model that follows.

It should be noted that in the models for multi volume works and for serials some assumptions have been made about using a web resource as the target of more than one aggregation. [22] Such assumptions need to be tested in the validation work to follow and in discussion with Europeana about implementation and with the other cultural heritage sectors who will have similar needs for describing complex objects.

[22] For example, as shown in the following diagram, the top level of a multi-volume work (e.g. The Complete Works of Shakespeare) has no physical existence, it is made up of the many separate volumes it groups together. To be represented in EDM it needs to link to a Web Resouce and therefore links to the first volume in the group.

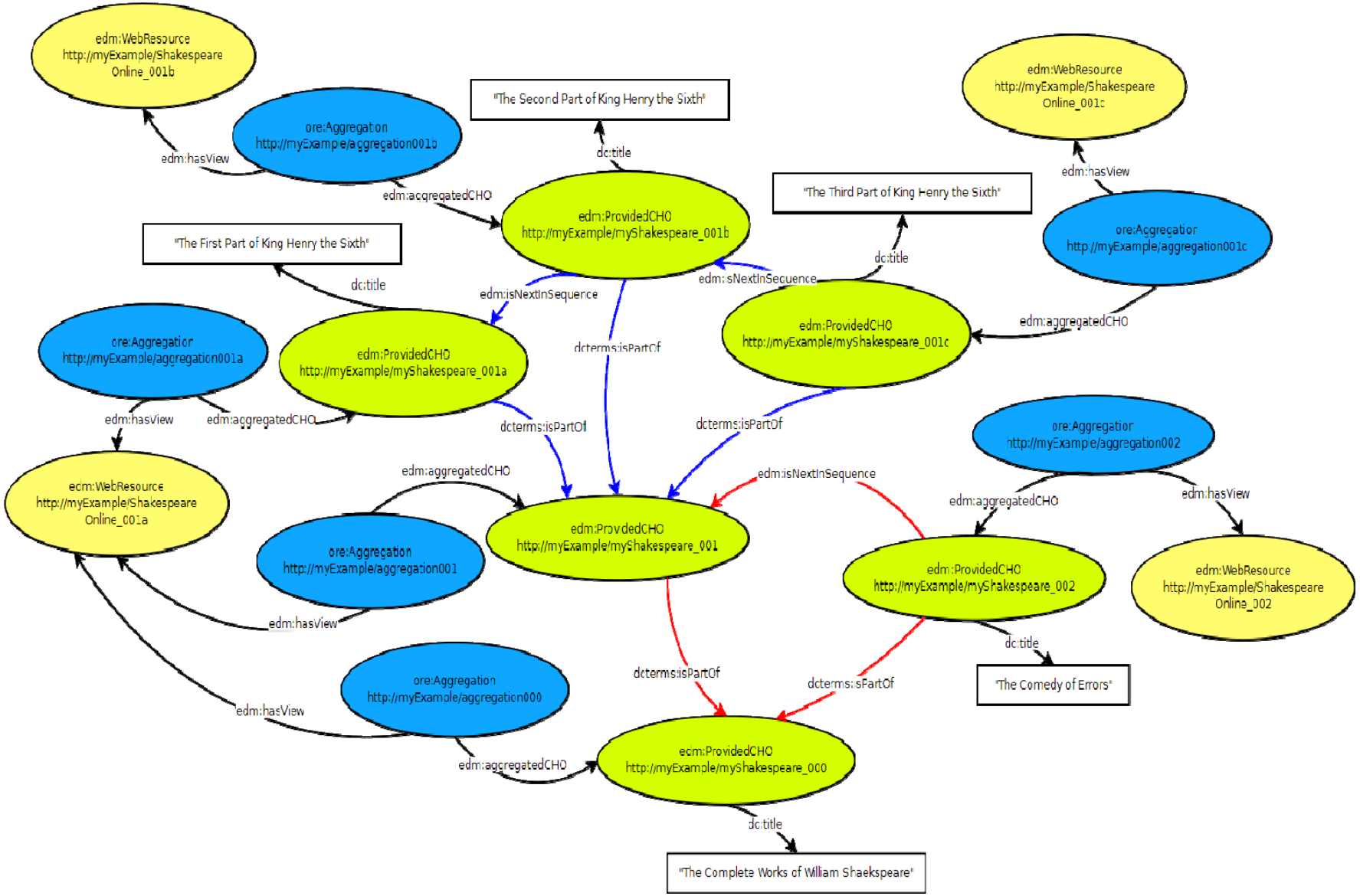

Figure 3: Example of a multi-volume work

5.2. Optimal properties and those available for the Europeana first implementation

The three following tables show the properties that have been selected from EDM to represent library monographs in the three core classes of the EDM model: edm:ProvidedCHO, edm:WebResource and edm:Aggregation. The rows shaded grey are those properties that Europeana Libraries would like to include in the profile but which are not included in Europeana's first implementation of EDM.

For this deliverable all these tables have been adapted from the spreadsheets that contain the specification. The spreadsheets contain further information and comments plus an indication of the ranges of the properties. The latter needs further work and possibly experimentation during the validation task 5.2.1 when the profile will be applied to sample data from different types of library that has already been provided to Europeana.

5.2.1. Properties for the 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. Its properties are those of the original cultural heritage object. Strictly speaking, the ProvidedCHO is the resource in Europeana that *represents* the real cultural heritage object as, clearly, it cannot *be* the real cultural heritage object itself.

Properties for the Provided CHO			
Property	**Definition**	**Obligation**	**Repeatable**
owl:sameAs	Indicates that two URI references actually refer to the same thing.	optional	Yes
dc:contributor	An entity responsible for making contributions to the resource.	optional	Yes
dc:creator	An entity primarily responsible for making the resource.	optional	Yes
dc:coverage	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)	optional	Yes
dcterms:spatial	Spatial characteristics of the resource. (Note: Mandatory in EDM to supply one of dc:subject or dc:coverage or dc:type or dcterms:spatial)	optional	Yes
dcterms:temporal	Temporal characteristics of the resource.	optional	Yes
dc:date	Use for a significant date in the life of the CHO. Consider the sub-properties of dcterms:created or dcterms:issued.	optional	Yes
dcterms:issued	Date of formal issuance (e.g., publication) of the resource. (Encode as W3CDTF)	optional	No
dc:description	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	optional	Yes

	dc:title or dc:description. Dc:Title is mandatory in this specification.)		
dcterms:tableOfContents	A list of subunits of the resource.	optional	No
dc:format	The file format, physical medium, or dimensions of the resource.	optional	Yes
dcterms:extent	The size or duration of the resource.	optional	No
dcterms:medium	The material or physical carrier of the resource.	optional	Yes
dc:identifier	An unambiguous reference to the resource within a given context.	optional	Yes
dc:language	A language of the resource. Encode as ISO 639-2. (Mandatory in EDM for objects of EDM type "TEXT")	optional	Yes
dc:publisher	An entity responsible for making the resource available.	optional	Yes
dc:relation	A related resource.	optional	Yes
dcterms:hasFormat	The described resource pre-existed the referenced resource, which is essentially the same intellectual content presented in another format.	optional	Yes
dcterms:isFormatOf	A related resource that is substantially the same as the described resource, but in another format.	optional	Yes
dcterms:hasPart	The described resource includes the referenced resource either physically or logically.	optional	Yes
dcterms:isPartOf	The described resource is a physical or logical part of the referenced resource.	optional	Yes
dcterms:hasVersion	A related resource that is a version, edition, or adaptation of the described resource.	optional	Yes
dcterms:isVersionOf	A related resource of which the described resource is a version, edition, or adaptation.	optional	No
dcterms:isReferencedBy	The described resource is referenced, cited, or otherwise pointed to by the referenced resource.	optional	Yes
dcterms:references	A related resource that is referenced, cited, or otherwise pointed to by the described resource.	optional	Yes
dc:rights	Information about rights held in and over the resource.	optional	Yes
dc:subject	The topic of the resource. (Note: Mandatory in EDM to supply one of dc:subject or dc:coverage or dc:type or dcterms:spatial)	optional	Yes
dc:title	A name given to the resource.	mandatory	Yes
dcterms:alternative	An alternative name for the resource.	optional	Yes
dc:type	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)	optional	Yes
edm:isSuccessorOf	This property captures the relation between the continuation of a resource and that resource. This applies a story, a serial, a journal etc. No content of the successor resource is identical or has a similar form with that of the presursor. The similarity is only in the context, subjects and figures of a plot. Successors typically form part of a common whole – such as a triology, a journal, etc.	optional	No
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.	optional	No
edm:type	The Europeana material type of the resource	mandatory	No

Apart from feeling the need to define what "ProvidedCHO" meant in the context of library material there were three particular issues that arose during discussion of the properties for this class. Where appropriate the questions were referred to Europeana for clarification. In addition, the working group will report back to Europeana any areas where changes or additions are desired. The issues are summarised in Section 8 of this document.

- 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. In Europeana this is a situation inherited from the ESE format and it will be allowed to continue for the time being. The working group considered several solutions to try and be compliant with both DC and EDM:
  a) Use literal values in order to ensure the values are represented in the portal. Although not DC compliant, any other solution risks losing the data in the short term.
  b) Use dc namespace properties instead of dcterms where the values are literals. This would mean using one of the top level DC properties instead of the dcterms refinements, for example, dc:format instead of dcterms:extent. This would mean a significant loss of granularity.
  c) Use blank nodes instead of URIs where there are literal values. Blank nodes are not useful in the linked data community and not acceptable to the XML schema used for the first implementation. (Local or "fake" identifiers can be used for this purpose however.)
  d) Use properties from other namespaces such as ISBD which will permit use of literals. This has great potential for the future when Europeana can interpret models that extend EDM but will result in loss of data in the first (and maybe later) implementations of EDM.
  e) Create Europeana Libraries properties. This has the advantages and disadvantages of the previous solution but would require a greater amount of work to define and maintain the properties.

For the present, we should use solution a) as a pragmatic choice and reconsider solution d) when this becomes possible. There are several namespaces in the library domain that may be re-usable for this purpose and this would fit well with the intentions of EDM.

- How to map the place of publication. This is a significant piece of information for library material but neither DC nor ESE offer a suitable property. EDM offers the possibility of using edm:event but further investigation revealed that this would require the value to be a non-literal (edm:happenedAt). It was conceded that library values for this property would be literals so this solution was not feasible at the moment. In addition, the Event class is not included in the first implementation of EDM. Solutions d) and e) above were suggested but have the same drawbacks. The probable solution is to add the place of publication and the publisher name in dc:publisher. This does not fit the semantic of the property but has been the normal course of action in libraries for ten years. This should be referred to Europeana.

- There was discussion about the use of edm:hasType. The definition of edm:hasType in EDM v5.2.2 implies that it is for "what" questions in the portal by acting as an umbrella super-property for dc:type and format (and possibly others). It therefore has the domain of Europeana CHOs and range of Concepts. However, it is not anticipated that it will be included in the data

created by providers. It has also been included in the edm:Event class to allow the typing of Event entities (C.f. CIDOC-CRM P2-has-type). Therefore, there is some looseness in the EDM specification that should be tightened up.

5.2.2. Properties for the WebResource

An edm:WebResource is a digital representation of the edm:ProvidedCHO.

Properties for the Web Resource			
Property	**Definition**	**obligation**	**repeat-able**
dc:format	The file format, physical medium, or dimensions of the resource. Encode as a MIME type.	mandatory	no
dc:rights	Information about rights held in and over the resource.	optional	no
dc:source	A related resource from which the described resource is derived.	optional	no
dcterms:extent	The size or duration of the resource.	optional	no
dcterms:created	Date of creation of the resource. Encode at W3CDTF.	mandatory	no
dcterms:conformsTo	An established standard to which the described resource conforms.	optional	no
dcterms:isFormatOf	A related resource that is substantially the same as the described resource, but in another format.	optional	yes??
edm:rights	Information about copyright of the digital object as specified by isShownBy and isShownAt	mandatory	no
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.	optional	no
edm:isRepresentationOf	[Not in candidate property list for edm:WebResource]		

As can be seen from the table, Europeana Libraries would like many more properties for the web resource than are currently proposed for the first implementation. 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).

The group particularly wanted 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. dcterms:issued was in the candidate property set but is not the most appropriate for digitisation purposes. The initial candidate EDM properties were close to being finalised during this discussion period and at the request of the Europeana Libraries group the property was able to be added. It was too late to have it included in the first implementation but it will be available in future. The library community uses many dates in relation to library materials, including those now born digital (e.g. theses) but the group accepts that these will have to be put forward in the recommendations to Europeana for the time being.

It would be preferable to use edm:isRepresentationOf to indicate the source of the digital object instead of the over-used dc:source. At the moment this is not available for the web resource class. In fact it may be a completely redundant property because the ore:Aggregation links the edm:WebResource and the edm:ProvidedCHO: the former is by definition a representation of the latter. It is still possible for the property to be added however if a use case can be discovered. However, this will not be in the first implementation of EDM.

There was some uncertainty about the best place to attach the edm:rights statement. This property is a candidate for the first implementation for both the edm:WebResource and the edm:Aggregation. It may well be the case that different web resources have different rights attached to them (for example, an image and a sound recording of the same thing could have different rights). Ideally, there should be a separate edm:rights statement attached to each web resource to allow different functions and features to be implemented accordingly. At the moment Europeana can only make use of the rights statement at the Aggregation level and this property is mandatory at that level. It is also strongly recommended at the level of each web resource by Europeana and has been made mandatory in this library profile.

5.2.2.1. Properties for the Aggregation

The **ore:Aggregation** class is the pivotal object between the edm:ProvidedCHO and the edm:WebResource(s) associated with it. It is also the place where the metadata relating to this whole object will be recorded.

Properties for the Aggregation			
Property	**Definition**	**obligation**	**repeat-able**
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)	optional	yes
edm:aggregatedCHO	This property associates an ore:Aggregation with the cultural heritage object it is about.	mandatory	no
ore:aggregates	Only stated in principle via edm:hasView and edm:aggregated CHO statements.	optional	yes
edm:dataProvider	The name or identifier of the organisation that contributes data to Europeana	mandatory	no
edm:isShownBy	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	mandatory*	yes
edm:isShownAt	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	mandatory*	yes
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.	mandatory	no
edm:provider	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.	mandatory	no
edm:rights	This is a **mandatory** property and the value given here should be the rights statement that applies to the digital representation at the URL given in edm:object or edm:isShownAt/By. The value should be taken from	mandatory	

	one of those listed in the **Europeana Rights Guidelines** (http://pro.europeana.eu/technical-requirements)		

Only one issue arose out of this class. The specification in EDM v5.2.2 for edm:dataProvider and edm:provider states that they have a range of edm:Agent and therefore should have an agent class node and an identifier as object. In the templates however, the value is shown as "literal or ref" which would in fact suit most current library data better. This anomaly was raised with Europeana who acknowledged that the error lay in the EDM specification. For the time being these values may be simple literals and it is hoped to upgrade them to real resources in due course.

6. Europeana Data Model for Serials[23]

Note: This deliverable will not address specific issues linked to full-text materials. These issues are addressed in deliverable D4.3 Report on how full-text content will be made available to Europeana.

6.1. Definition of the scope

Europeana Libraries describes as serials different types of resources which have in common the fact that they are often composed of multiple parts. The model described below addresses the following materials:

- Journals/Newspapers/Periodicals: These terms describe different types of serials issued at stated and frequent intervals, usually daily, weekly or semi weekly; and that report events and discuss topics of general current interest. These serials contain separate articles. They constitute a continuing resource issued in a succession of small issues or parts, usually bearing numbering, that has no predetermined conclusion
- An issue of a serial has to be considered as a resource dependent on another i.e. an issue is one of the successive parts of a serial.
- A volume is a major division of a work, distinguished from other major divisions of the same work by having its own properties such as an independent pagination, foliation and signature. Individual volumes are usually numbered.
- A journal article is considered and is working as an independent unit when a journal is representing the whole.

These different materials raise specific issues which will need to be taken into account in the EDM model defined.

A serial is a structured type of resource representing both a hierarchy and a succession. While some resources can be considered as independent units and can be understood standing alone e.g. a journal article; others can be understood only thanks to their relationships with other resources e.g.

[23] Source: ISBD(CR) : International Standard Bibliographic Description for Serials and Other Continuing Resources: http://www.ifla.org/node/900

different issues of the same journal. Supplements, annexes, articles can either be described as one resource or described as part of a serial.

The model is based on the most common access point for serial in a library catalogue and also from a user point of view.

The structure issue is visible for instance at the title level. A serial can bear a title for the individual available entities (e.g. article) but also a collective title linking to the parent entity. We might also find the case when different journal articles can also be published under the same title.

The format and extent of serials also need to be considered in detail. Serials can be in a print form and a non-print form, with and without chronological or numerical designations. Issues published in the same year might be bound in one volume. Some newspapers might have been converted to a microfilm or micro fiche.

Finally, serials have a strong temporal dimension which needs to be reflected in the model.
A serial is a resource which can change in time (e.g. new issues of a journal are published, an erratum is associated to an already published article). We might want to track the history of a journal (e.g. change of name, closing of a journal which is continued under another name).

6.2. Definition of the model for the serials

As a result of the different ways in which libraries deal with serials, from digitisation to the cataloguing, Europeana libraries was not able to define a clear model to describe serials. However some needs and issues have been raised and will be explained in the following sections.

The validation of the model against real data will define whether or not the model fits to the serial type material.

Europeana Libraries has identified in the serials different levels which could give rise to and could be the subject of a package of data submitted to Europeana. Each level is a potential cultural heritage object Europeana collects descriptions about. In the context of serials the levels that need to be considered in the model are:

-the article level
-the issue
-the volume
-the title level of the serials[24]

We consider the article as the smallest entity and the lowest level of granularity available, when the title is the highest level. These four levels can potentially carry the three main classes proposed by EDM, that is to say a ProvidedCHO, one or more WebResource and an Aggregation.

[24] In the following diagrams the title level is represented by the type journal.

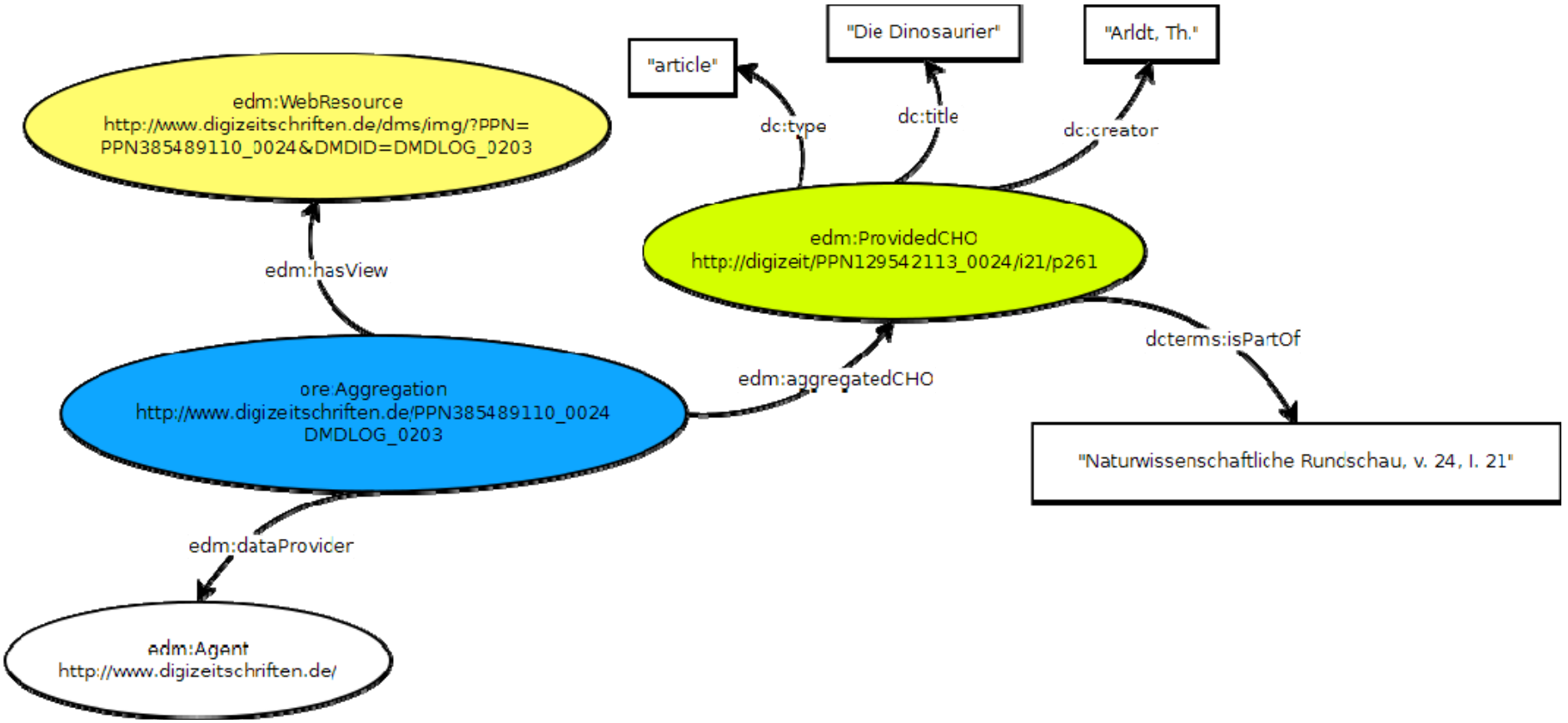

Figure 4 : Entity-Relationship Model for an article

To make sure such a model works, the group has raised the importance of specifying the type of each cultural heritage object. The model would use dc:type to describe this information. The short term solution is to provide a literal with this property. However the group has recommended the use of controlled terms if we want to be able to track the granularity of our mapping. The group has suggested the use of the MARC genre list[25] to standardise the values to indicate as a type. The drawback of this solution is that the term "volume'' is not available in the proposed vocabulary. The group is also looking at the possibility of using the Ontology bibo[26] to describe serials depending on the use cases which will be analysed during the validation process. The group will also stay aware of the development of RDA. The group will work on further developments during the next year of the project.

Having specified the level of description of our model, it is then needed to represent the relationships between resources. These relationships are, in the first instance, vertical.

Since each level of the serials model can be considered as a ProvidedCHO, it is possible to represent the model as a hierarchical representation.

In the following diagram, the different levels of the serial are linked together using the property dcterms:isPartOf. This property allows us to represent the relationships existing between the resources.

[25] http://www.loc.gov/standards/valuelist/marcgt.html
[26] http://bibliontology.com/content/complex-series-proceeding-article-use-case
http://bibliontology.com/content/article

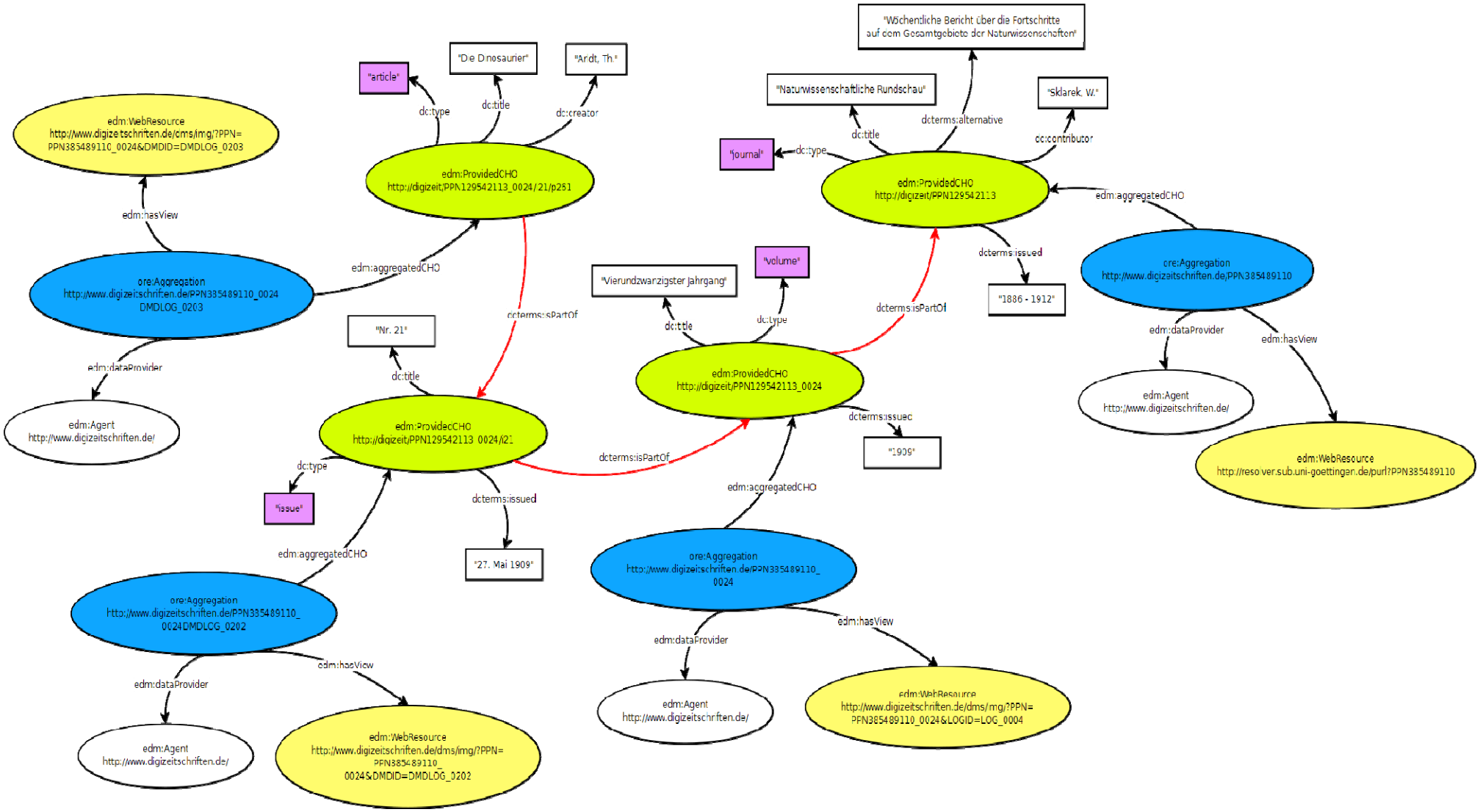

Figure 5: Entity-Relationship Model for a serial with hierarchical relationships

As mentioned in the introduction of the model for serials, this type of resource can also be described through its horizontal relationships. To describe these relations, the model will be using the property edm:isNextInSequence. This property could be the basis of a specific functionality with the Europeana portal which will allow a user to browse through the resource.

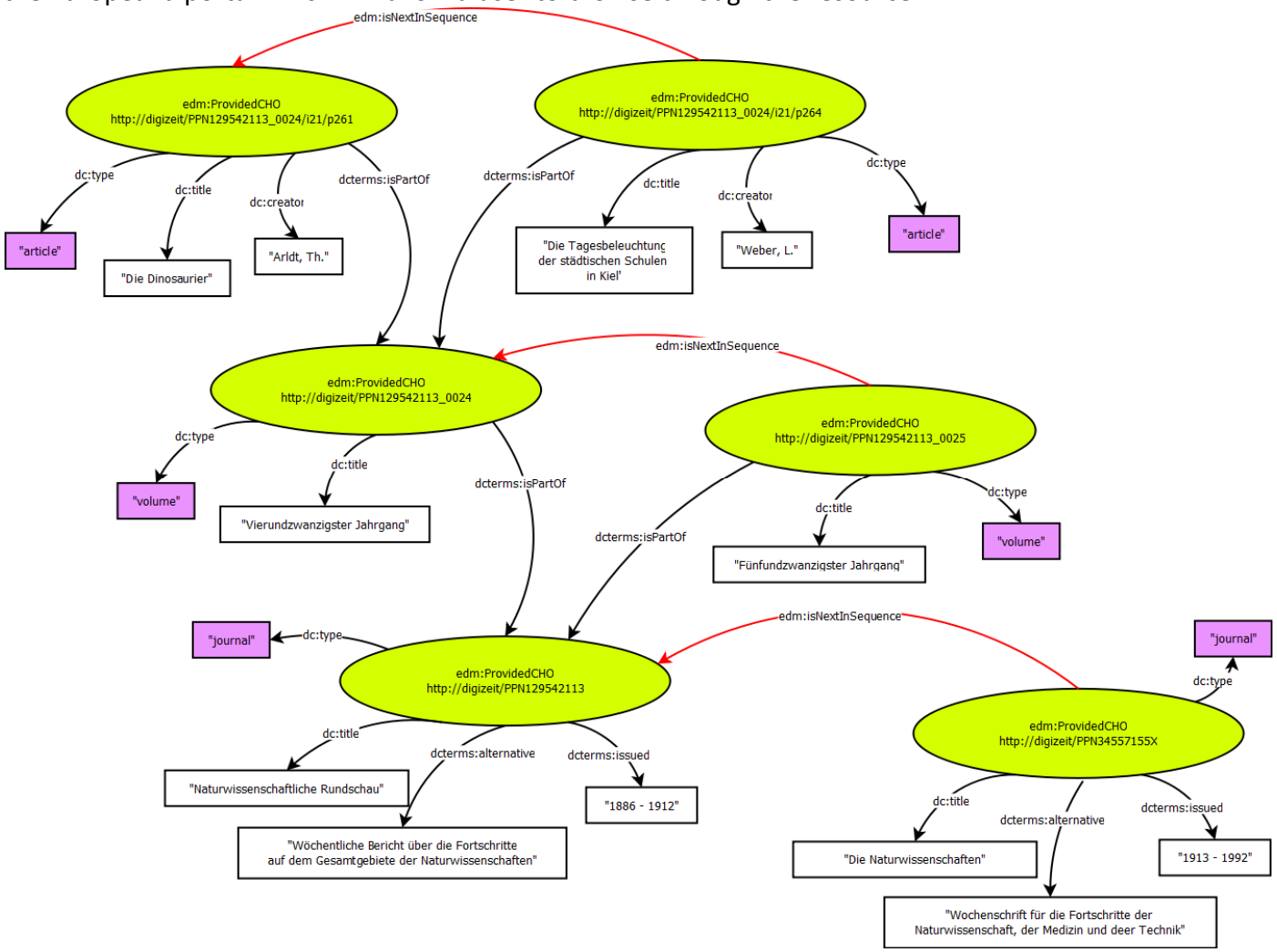

Figure 6 : Entity-Relationship Model for a serial with succession relationships

6.3. Optimal properties and those available for the Europeana 1st implementation

As described earlier, the three following tables show the properties that have been selected from EDM to represent serials materials. Since a lot of properties proposed for monographs are also applicable to the serials, we will describe only the properties specific to serials or having a specific use in this context.
These tables are proposing one general set of properties for every level to describe. These properties can be applied to each ProvidedCHO according to what is available.

6.3.1. Properties for the ProvidedCHO

The ProvidedCHO, for serials in Europeana Libraries, the cultural heritage object can be either an article, an issue, a volume or a title. The choice of the properties will depend of the richness of the data.

Properties for the Provided CHO			
Property	**Definition**	**Obligation**	**Repeatable**
dcterms:issued	Date of formal issuance (e.g., publication) of the resource. (Encode as W3CDTF). This property is highly recommended for the issue and volume level.	mandatory	No
dcterms:abstract	A summary of the resource.	optional	Yes
dcterms:bibliographicCitation	A bibliographic reference for the resource.	optional	No
dcterms:isPartOf	The described resource is a physical or logical part of the referenced resource. This property is recommended to link hierarchical resources.	Optional (mandatory when a resource is part of a hierarchy)	Yes
dc:type	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). The type will be used to define the level of granularity.	mandatory for Europeana Libraries Optional in Europeana	No (different from EDM)
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. This property is strongly recommended to describe a succession of resources within a serial.	optional	No

In addition to the properties proposed in the model for monographs, the group would like to see the properties dcterms: abstract and dcterms: bibliographicCitation added to the model for serials. The discussion on how to map the place of publication has been already held in the section on monographs. The same conclusions apply for the model for serials.

6.3.2. Properties for the WebResource

An edm:WebResource is a digital representation of the edm:ProvidedCHO.

Properties for the Web Resource			
Property	**Definition**	**obligation**	**repeat-able**
dcterms: hasFormat	A related resource that is substantially the same as the pre-existing described resource, but in another format.	mandatory if a fullTextresource class is used	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. This property is recommended to describe pagination information.	optional	no

In this model, Europeana Libraries recommends the use of edm:isNextInSequence to describe the pagination of WebResource(s) with serials.

The group would like to use the property dcterms: hasFormat for the integration of Full-Text materials with Europeana. The deliverable *D4.3 Report on how the full-text content will be made available to Europeana* [27] suggests the creation of a new EDM class which would be called FullTextResource. This class would be linked to the WebResource class using the property HasFormat. It would then provide the end-users with access to the views of the digital object.

6.3.2.1. Properties for the Aggregation

The **ore:Aggregation** class is the pivotal object between the edm:ProvidedCHO and the edm:WebResource(s) associated with it. It is also the place where the metadata relating to this whole object will be recorded.

The serials model does not require additional properties within the Aggregation class.
The issue which arose in the monographs model about the specification of edm:Agent is also valid for this model.

7. Next steps

The work on the EDM profile for libraries has been so far focused on the modelling of the data. Even if the group has always worked with real data, the model will need to be validated at a technical level. The outcome would be the alignment of the data to EDM.

[27] *D4.3 Report on how the full-text content will be made available to Europeana* is due to December 2011 and is therefore not yet publicly available.

7.1. Validation

This deliverable provides recommendations for the EDM modelling. Although this model is based on the analysis of real data sample, a further analysis of the libraries legacy data might reveal some inconsistencies which will need to be checked.

As part of the task 5.2 Aligning library-domain metadata with the Europeana Data Model, the deliverable D5.1 will be submitted to the libraries partners in the WP5 for validation. In order to enlarge the scope of the validation, ten extra libraries[28], who have already contributed content to Europeana, will also participate in the validation process.

When discussing the model on monographs, the group worked on a MODS example and tried to select the EDM properties relevant to describe the monographs **(See Annex 3)**. Another aspect of this work was to describe the different relationships existing between resources. A similar exercise will be untaken by the libraries involved in the validation process.

The results of the validation will be summarized in a reviewed version of the deliverable D5.1. This final version of the deliverable D5.1 will be the reference document which will be used for the technical validation of the EDM model for libraries.

7.2. Technical issues

The discussions and mapping exercises undertaken by WP5 have raised several issues which will have a crucial role when WP5 will start working on the alignment of the data on the EDM profile and the practical implementation of the model.

The first range of issues is related to the Europeana Data model itself. EDM has been developed to be interoperable at the semantic level and is based on relationships between different kinds of resources. The logic behind EDM based on the Resource Description Framework RDF, is then requiring the use of identifiers and URI to define the resource and can imply specific technical choices. Libraries have become XML orientated and the RDF approach is therefore providing new challenges.

As mentioned in the section 5.3, an issue for Europeana Libraries will be to adapt the EDM profile in line with the legacy data held in libraries. This issue will be particularly applicable for properties, which, according to the DCAM (Dublin Core Abstract model [29]), should contain non-literal values. During the different meetings of Europeana Libraries a second issue directly related to the legacy data available in libraries was discovered. Metadata standards used in libraries such as MODS, TEI, MARC21 are using attributes quite extensively in their data. These attributes are sometimes

[28] The libraries for the wider library domain are: Herzog August Bibliothek, Germany, Historic Library of the University of Valencia, Spain, Roskilde Libraries, Denmark, Veria Central Public Library, Greece, Public Library of Varna, Bulgaria, National Library of Scotland, Scotland, Poznan supercomputing and networking center, Poland, Irish virtual Research Library, Ireland.

[29] See http://dublincore.org/documents/abstract-model/index.shtml

providing useful information which allows to keep a quite high level of granularity within the data. The question is how to keep these attributes during the mapping process.

For instance: In TEI:

```
<publisher>
        <name type="org">Herzog August Bibliothek Wolfenbüttel</name>
        <ptr target="http://www.hab.de"/>
```

In MODS:

```
</name>
<name authority="koba" type="personal" xlink:href="oai:ub.uu.se:koba.aut-1940"/>
<namePart type="family">Gimberg </namePart>
<namePart type="given">Fredrika</namePart>
<role><roleTerm type="text">fotograf</roleTerm>
```

The first and pragmatic solution to this issue will be the use of blank nodes instead of URIs. The model would then still support the granularity required and would still be Dublin Core compliant. This solution could have only implication if EDM is used as an XML schema as it will be in the first implementation of EDM done by Europeana.

For instance: edm:ProvidedCHO dcterms:isPartOf [rdf:value “My Collected Works’’]
The drawback of this solution is that it will not be possible to track back the attributes and their specificities.

Another solution to this issue would be to not respect the Dublin Core recommendations and accept literals within properties. In the longer term, Europeana Libraries will be able to provide URIs for these values using an enrichment process. The WP4 of European Libraries is already working on enrichment using VIAF and Geonames, and subjects from the MACS project and is therefore able to provide identifier and URI for places, agents and subjects terms.

7.3. Alignment of the data

The model and technical issues raised during the modelling activities will then be tested with real data. WP5 will gather data samples representing the diversity of libraries legacy data and metadata formats and will apply it the EDM model. The group is looking at the possibility to execute this work using the aggregation infrastructure developed in WP4. The WP5 libraries partners and the ten extra libraries will validate the output of the data transformation.

8. Feedback to Europeana

This section summarises the various issues in relation to EDM and its implementation that arose during the discussions about the creation of a profile of EDM for use by libraries. As the earlier sections of this document show, there was a steep learning curve involved in working out which classes and properties to consider. Many issues were resolved by discussion and in the light of additional EDM documentation advice.

1. A clear definition of what an edm:ProvidedCHO is should be provided. In the library domain the guiding model is FRBR and, in this case, the group 1 entities (Work, Expression,

Manifestation and Item.) There is a desire to describe both the edition level (as defined in Section 5.2.2) and Item level for rare and unique objects (e.g. a copy of a book but it was owned by a significant person, or it has significant notes in the margins.) It would be useful if Europeana could define a standardised way to represent the FRBR entities in EDM. (*Note: feedback was provided from Europeana on this point and the rare book model is no longer included.)*

2. EDM allows non-compliance with the DCMI specifications of ranges for the dcterms properties which caused some lengthy discussion in the group. The group recognises that Europeana is aware of this position, and also that many other organisations ignore the specification as well, but would recommend that it should be rectified as soon as it is feasible. (See section 5.2.1 – Properties for the ProvidedCHO)
3. There are several credible namespaces in the library domain from which properties could be reused to extend EDM properties. Europeana Libraries would encourage Europeana to move towards an implementation of EDM that would incorporate such developments. (See section 5.2.1 – Properties for the ProvidedCHO)
4. Advice on how to include the place of publication. It could be modelled using edm:Event but this would require a URI as the value in edm:happenedAt and is therefore not suitable for most current library data which has simple literals for place names. (See section 5.2.1 – Properties for the ProvidedCHO)
5. Europeana has already responded to the request to include the dcterms:created property in the edm:WebResource set of properties. It would be useful if this could be implemented as soon as possible. (See section 5.2.2 – Properties for the WebResource)
6. There is a valuable set of data that could be provided about the web resources (dates, format etc) and it would be helpful to be able to sequence the digital files where digitising one object has resulted in many files. Libraries cannot provide this for the first implementation however as so few properties have been selected for this class. The group would like to see this expanded as soon as possible as being fairly central to describing digitised and born-digital resources. (See section 5.2.2 – Properties for the WebResource)
7. In the portal Europeana currently uses the the rights statement attached to the edm:Aggregation, but it would prefer this to be at the level of the web resource. It is highly desirable that rights statements could be attached to each web resource as there may be several of them and they are likely to carry different rights. Europeana Libraries supports Europeana in its intention to use rights statements at this more granular level and has made this a mandatory property in the profile. (See section 5.2.2 – Properties for the WebResource)
8. Many dates are used in relation to typical library material and some of these are significant for researchers. A good example is in the process of producing a thesis. Europeana Libraries would like an unambiguous way to incorporate many more of these in descriptions.
9. dc:source has been much abused in the past and it would be better to use edm:isRepresentationOf to indicate the original resource in the case of a digitisation. (See section 5.2.2 – Properties for the WebResource).
10. There is a mismatch in the Europeana documentation regarding the use of edm:hasType for both edm:ProvidedCHO and edm:Event and that they would be used for different purposes in each of these. Europeana has acknowledged that there is potential for misunderstanding over this but does not anticipate providers using the property in data creation. They will seek to improve it when possible. (See section 5.2.1 – Properties for the ProvidedCHO)

11. The group would like to be able to use edm:isSuccessorTo in relation to the edmProvidedCHO. (See section 5.2.1 – Properties for the ProvidedCHO)
12. There is an mismatch between the templates and EDM 5.2.2 regarding the acceptable values for edm:provider and edm:dataProvider. EDM 5.2.2 states that the range must be an edm:Agent node whereas the templates state that a literal or reference can be used. Europeana has acknowledged this mismatch and will amend it when possible. (See section 5.2.2 – Properties for the Aggregation).

9. Future Work

Further research for the Europeana Libraries metadata working group would be to develop new EDM models for library resources the current profile presented in this deliverable does not cover. These resources would be in priority the thesis, the manuscripts and maps. A close collaboration with other domain specific or cross domain projects could help Europeana Libraries to extent its EDM profile to audio and video resources also held by libraries. Europeana Libraries has already started working in this direction inviting experts from the cross- domain aggregator Hope. Such synergies would be really useful in the development of domain oriented EDM profiles.

FRBR entities

As described above in the library world the providedCHO as an item is part of a complex structure of FRBR entities and their relationships. We recommend to integrate the FRBR entities in EDM using FRBRoo terms in the future, where the edm:ProvidedCHO is the same as a F5 Item of the FRBRoo specification.

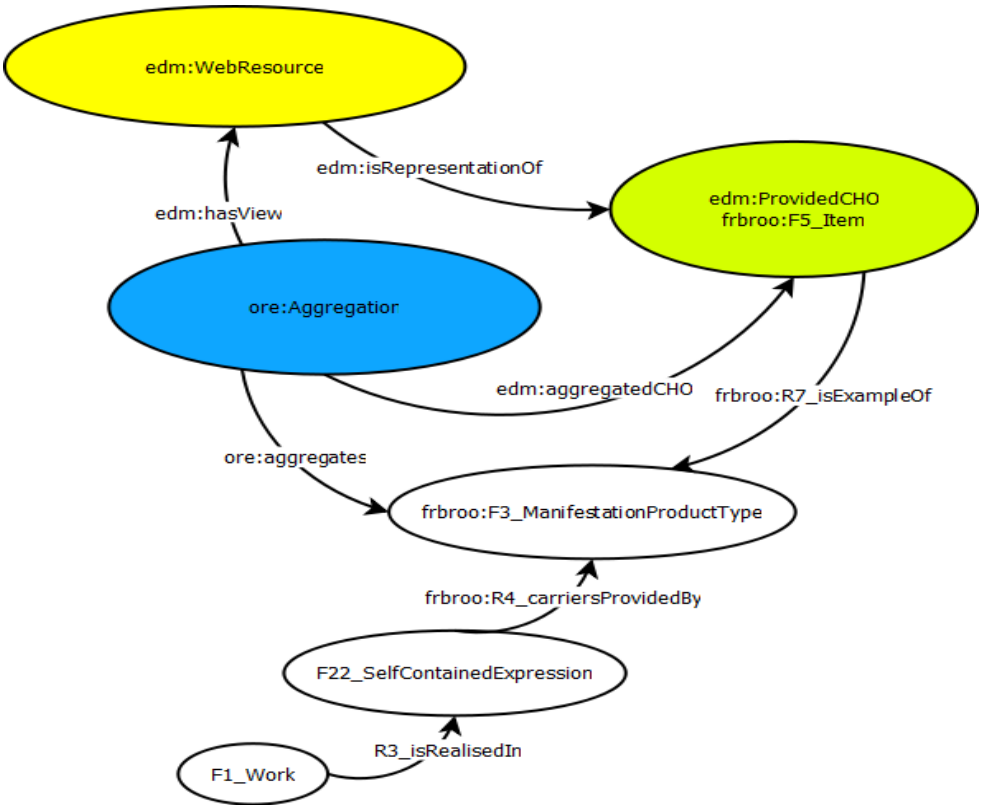

Figure 7: Europeana Libraries Entity-Relationship Model with FRBRoo entities

Properties for edm:ProvidedCHO distributed to FRBR Group 1 entities

Table 1: Properties of the edm:ProvidedCHO

Property	Definition	Elib Comment	Obligation	Repeatable
owl:sameAs	Indicates that two URI references actually refer to the same thing.	The thing referenced to must be the same **item** as the item represented by the ProvidedCHO. For relationship between the item and another item of the same expression use dc:relation	optional	yes
dc:contributor	An entity responsible for making contributions to the resource.	Describes the contributors of the **manifestation**	optional	yes
dc:creator	An entity primarily responsible for making the resource.	Describes the creators of the **work**	optional	yes
dc:coverage	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)	Describes the topic of the **work**	optional	yes
dcterms:spatial	Spatial characteristics of the resource. (Note: Mandatory in EDM to supply one of dc:subject or dc:coverage or dc:type or dcterms:spatial)	Describes the topic of the **work**	optional	yes
dcterms:temporal	Temporal characteristics of the resource.	Describes the topic of the **work**	optional	yes
dc:date	Use for a significant date in the life of the CHO. Consider the subproperties of dcterms:issued or dcterms:created.	Use only if dcterms:issued or dcterms:created cannot be used.	optional	yes
dcterms:issued	Date of formal issuance (e.g., publication) of the resource. (Encode as W3CDTF)	Describes the date the **manifestation** was published.	optional	no
dcterms:created	The date of the creation of the CHO	Describes the date the **expression** was created	optional	no
dc:description	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.)	A description of the **item**. For the description of the expression use dcterms:abstract	optional	yes
dcterms:tableOfContents	A list of subunits of the resource	The subunits of the **expression**	optional	no

dcterms:provenance	A statement of changes in ownership and custody of the CHO since its creation. Significant for authenticity, integrity and interpretation.	The provenance of the **item**	optional	yes
dc:format	The file format, physical medium, or dimensions of the resource	The format of the **manifestation**	optional	yes
dcterms:extent	The size or duration of the resource.	The size of the **manifestation**	optional	no
dcterms:medium	The material or physical carrier of the resource.	The material of the **item**	optional	yes
dc:identifier	An unambiguous reference to the resource within a given context.	The identifier of the **manifestation** (e.g. ISBN, ISSN). The identifier of the item is the URI identifying the ProvidedCHO	optional	yes
dc:language	A language of the resource. Encode as ISO 639-2. (Mandatory in EDM for objects of EDM type “TEXT”)	The language of the **expression**	mandatory for objects of EDM type "TEXT"	yes
dc:publisher	An entity responsible for making the resource available.	The publisher of the **manifestation**. Includes the place of publication (e.g. Berlin : Springer)	optional	yes
dc:relation	A related resource.	A resource that has the same **expression**	optional	yes
dcterms:hasFormat	The described resource pre-existed the referenced resource, which is essentially the same intellectual content presented in another format.	Relation between two **manifestations**	optional	yes
dcterms:isFormatOf	A related resource that is substantially the same as the described resource, but in another format.	Relation between two **manifestations**	optional	yes
dcterms:hasPart	The described resource includes the referenced resource either physically or logically.	Relation between two **manifestations**	optional	yes
dcterms:isPartOf	The described resource is a physical or logical part of the referenced resource.	Relation between two **manifestations**	optional	yes
dcterms:hasVersion	A related resource that is a version, edition, or adaptation of the described resource.	Relation between two **expressions**	optional	yes
dcterms:isVersionOf	A related resource of which the described resource is a version, edition, or adaptation.	Relation between two **expressions**	optional	no
dcterms:isReferencedBy	The described resource is referenced, cited, or otherwise pointed to by the referenced resource.	Relation between the **expression** and another resource	optional	yes
dcterms:references	A related resource that is referenced, cited, or otherwise pointed to by the described resource.	Relation between the **expression** and another resource	optional	yes

dc:rights	Information about rights held in and over the resource.	Copyright of the **manifestation**	optional	yes
dc:subject	The topic of the resource. (Note: Mandatory in EDM to supply one of dc:subject or dc:coverage or dc:type or dcterms:spatial)	Describes the topic of the **work**	Optional	yes
dc:title	A name given to the resource.	The title of the **manifestation**	mandatory	yes
dcterms:alternative	An alternative name for the resource.	The alternative title of the **manifestation**	optional	yes
dc:type	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)	The nature or genre of the **expression**	optional	yes
edm:isSuccessorOf	The identifier of a resource to which the described CHO is a successor. E.g. issue 53 of "Le Temps" isSuccessorOf issue 52	Relation between two **manifestations**	optional	no
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.	Relation between two **manifestations**	optional	no
edm:type	The Europeana material type of the resource	For text resources the edm:type will be "TEXT"	mandatory	no

Event based description of the life cycle

One fundamental difference between library-domain and museum-domain data is the description of the creation, modification, publication, etc. of an object. Where the museum-domain builds on events (e.g. creation, publication, modification) interlinked with information about actors, places, dates etc. relevant to this event, the library-domain distinguishes the different types of actors, places, dates etc. by using the granularity of metadata terms (e.g. dateCreated, dateIssued, dateModified – creator, publisher, contributor).

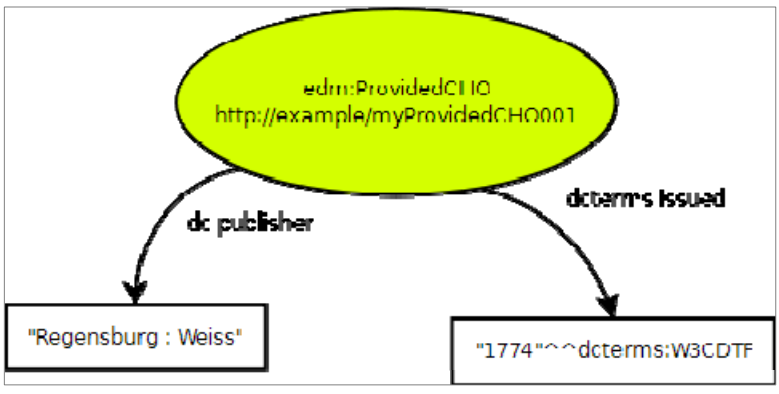

Figure 9: Library-domain description of publication information

As a cross-domain model, EDM allows both ways to describe such statements and the description of the place of creation, publication, modification etc. can not satisfactorily be described except by using an event-based description.

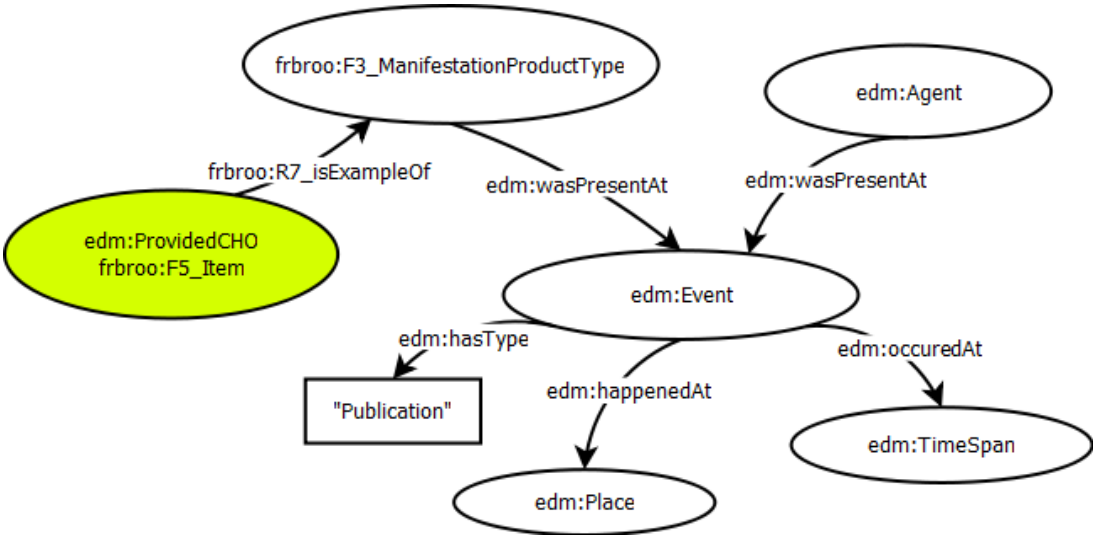

Figure 10: Model of an event-based description of publication information

As the first implementation does not allow the use of edm:Event and the related EDM properties, Europeana Libraries decided to go the traditional way and to use dc:publisher for both the publisher's name and the place of publication. However, the discussion will be resumed as soon as edm:Event has been implemented.

Clustering similar textual resources

In addition to the interlinking of hierarchical structured entities the model can also be used for the clustering of different resources carrying similar content. This invented example shows how an aggregator could cluster different versions of the same resource: library A has a born digital monograph, library B has a book of the same monograph and publishes a digitised version of it.

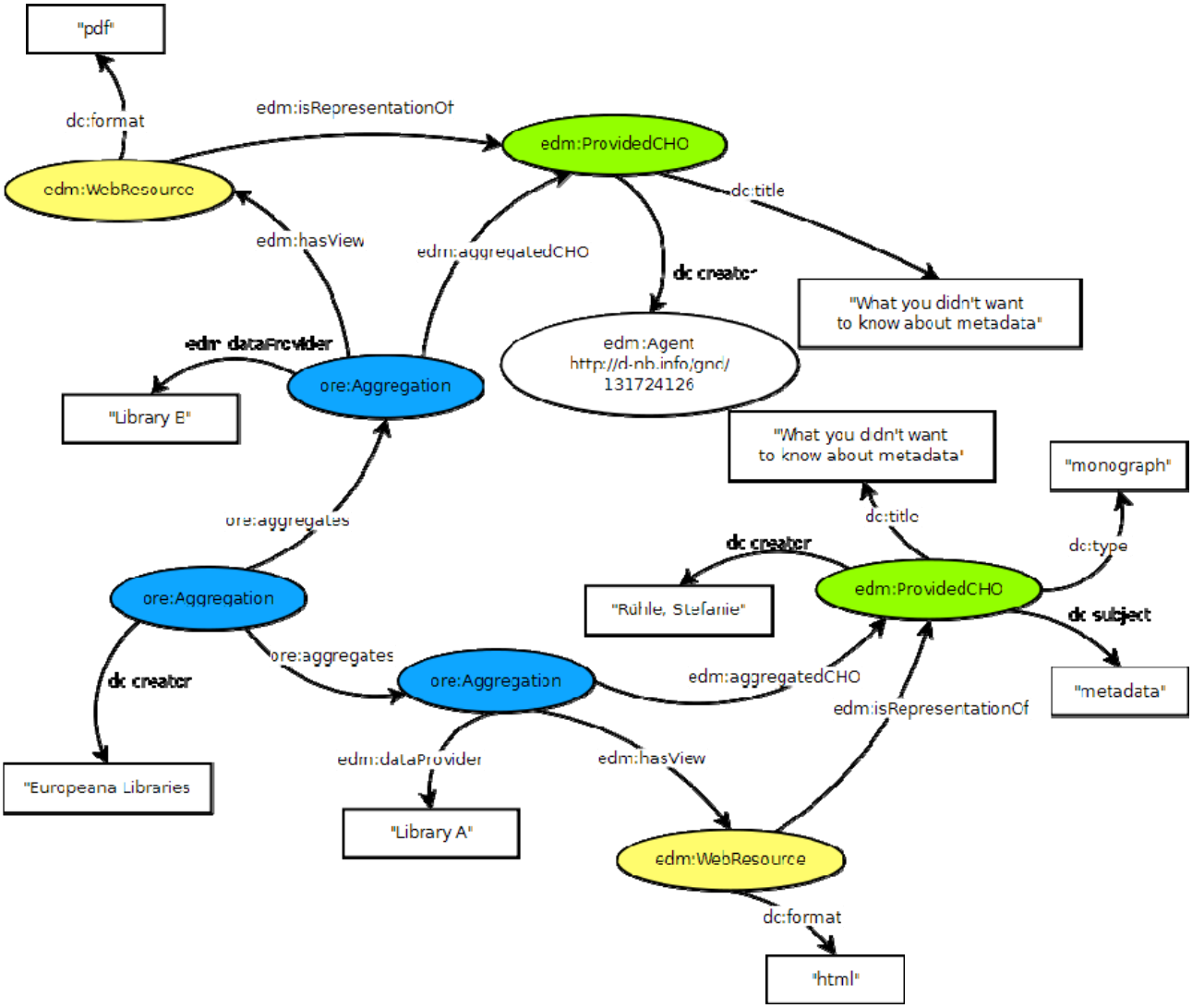

Figure 11: Clustering of similar text resources with different information carriers

The group discussed this issue with Europeana and decided not to address this level of complexity at this stage in Europeana Libraries as there is no requirement in the Europeana Libraries Description of Work to undertake clustering as part of the project. However, it was noted that creating links between similar items in different libraries is a feature that a library-domain aggregator could offer as an added value. It was noted that links between similar items could be made by building on the clustering algorithms that have been developed as part of the ARROW project. For that reason Europeana Libraries will return to this topic, in year two of the project.

10. Conclusion

This deliverable should be considered as a milestone in an ongoing process to define how EDM can be used with library resources. The work so far focused on the modelling of the data and, as highlighted in the further research section, the modelling of additional library materials, with the next priorities being theses, the manuscripts and maps, is needed.

The modelling undertaken to date, needs to be validated at a technical level in year 2 of the project. Following the validation of the deliverable, due for completion in June 2012, the alignment of library metadata to the validated model can begin.

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.

11. Annexes

Annex 1 : Metadata Working Group - Terms of reference

- **Task 5.1.1:** to study the metadata formats used by Europeana with particular focus on understanding the principles and usage of the Europeana Data Model (EDM).

- To use their library domain metadata expertise to develop the most appropriate alignment of library domain metadata with the EDM. To validate this in the first instance based on ESE metadata already supplied to The European Library by national libraries.

- To examine any work on transforming library metadata to EDM in other Europeana-related projects and recommend adoption of any outputs if appropriate.

- To be responsible for producing D5.1 – Report on the alignment of library metadata with EDM.

- **Task 5.1.2:** to support the work of KB and Europeana in the transformation of existing national library metadata to EDM, particularly by having an evaluation role.

- **Task 5.2.1** to support the work of KB and Europeana taking forward the recommendations of D5.1 by applying them to metadata from different types of library.

- **Task 5.2.2 :** to support the work of KB and Europeana in the transformation of existing ESE metadata from research, university to EDM by having an evaluation role.

- To maintain awareness of other developments in the transformation and use of metadata including semantic web developments such as Linked Open Data ensure the project is informed accordingly.

Annex 2
List of conferences at which the work of the WP5 has been presented to the network:

- Presented *Iets korts over het Europeana Data Model (EDM) voor bibliotheken* at the Data Salon 6, Metadata Barcamp, Brussels, Belgium, 21 January 2011
- Represented WP5 at the Europeana Libraries Kick-off Meeting, 26 – 27 January
- Presented Mapping the European(a) metadata landscape at the Library Association of Ireland, Cataloguing and Indexing Group Annual seminar, Describing the digital object, 25 February 2011, Dublin, Ireland
- Facilitated a session on Europeana and the metadata landscape at Bibcamp, Hamburg, Germany, 11 – 12 March, 2011
- Presented *Aligning library-domain metadata with the Europeana Data Model* at ELAG 2011 Conference, Prague, Czech Republic, 25 – 27 May 2011.
- Europeana Libraries Workshop, Den Haag, Netherlands, 30-31 May 2011
- Presentation at DCMI 2011, Den Haag, Netherlands, 21-23 September 2011
- Europeana Libraries Workshop, Belgrade, Serbia, 14-16 November 2011

Annex 3: Properties used for the description of monographs

"The RDF Concept and Abstract Syntax specification describes the concept of an RDF property as a relation between subject and object resources".[30] EDM properties thus describe:

- the relationship between EDM Entities
- the relationship between EDM Entities and Literals.

A first exercise has been to identify the EDM properties applicable to their classes depending on the existing relationships. Similar work has been done using a MODS data sample.

Properties describing the relations between an aggregation and its aggregated resources:

DOMAIN	PROPERTY	RANGE	OBLIGATION
ore:Aggregation	edm:aggregatedCHO	edm:ProvidedCHO	mandatory
ore:Aggregation	edm:hasView	edm:WebResource	optional
ore:Aggregation	edm:isShownAt	edm:WebResource	mandatory if edm:isShownBy is not provided
ore:Aggregation	edm:isShownBy	edm:WebResource	mandatory if edm:isShownAt is not provided
ore:Aggregation	edm:object	edm:WebResource	mandatory

Properties describing relations between a resource and an agent or agent literal

DOMAIN	PROPERTY	RANGE	OBLIGATION
edm:ProvidedCHO	dc:contributor		optional
edm:ProvidedCHO	dc:creator		optional
edm:ProvidedCHO	dc:publisher		optional
ore:Aggregation	edm:dataProvider	edm:Agent	mandatory
ore:Aggregation	edm:provider	edm:Agent	mandatory

Properties describing relations between a resource and a time literal

DOMAIN	PROPERTY	RANGE	OBLIGATION
edm:WebResource	dcterms:created	rdfs:Literal	mandatory
edm:ProvidedCHO	dcterms:issued	rdfs:Literal	mandatory

Properties identifying a resource

DOMAIN	PROPERTY	RANGE	OBLIGATION
edm:ProvidedCHO	dc:identifier	rdfs:Literal	optional
edm:ProvidedCHO	dc:title		mandatory
edm:ProvidedCHO	dcterms:alternative	rdfs:Literal	optional

[30] See: http://www.w3.org/TR/rdf-schema/

Properties describing the content of a resource

DOMAIN	PROPERTY	RANGE	OBLIGATION
edm:ProvidedCHO	dc:coverage		optional
edm:ProvidedCHO	dc:description	rdfs:Literal	optional
edm:ProvidedCHO	dc:subject		optional
edm:ProvidedCHO	dcterms:spatial		optional
edm:ProvidedCHO	dcterms:tableOfContents		optional
edm:ProvidedCHO	dcterms:temporal		optional

Properties describing the nature of a resource

DOMAIN	PROPERTY	RANGE	OBLIGATION
edm:ProvidedCHO	dc:format		optional
edm:ProvidedCHO	dc:language	rdfs:Literal	optional
edm:ProvidedCHO	dc:rights		optional
edm:WebResource	dc:rights		optional
edm:ProvidedCHO	dcterms:extent		optional
edm:ProvidedCHO	dcterms:medium		optional
edm:WebResource	edm:rights		mandatory

Properties describing the type of a resource

DOMAIN	PROPERTY	RANGE	OBLIGATION
edm:ProvidedCHO	dc:type		optional
edm:ProvidedCHO	edm:type	rdfs:Literal	mandatory

Properties describing the relation between a resource and another resource

DOMAIN	PROPERTY	RANGE	OBLIGATION
edm:ProvidedCHO	dc:relation		optional
edm:ProvidedCHO	dcterms:hasFormat		optional
edm:ProvidedCHO	dcterms:hasPart		optional
edm:ProvidedCHO	dcterms:hasVersion		optional
edm:ProvidedCHO	dcterms:isFormatOf		optional
edm:ProvidedCHO	dcterms:isPartOf		optional
edm:ProvidedCHO	dcterms:isReferencedBy		optional
edm:ProvidedCHO	dcterms:isVersionOf		optional
edm:ProvidedCHO	dcterms:references		optional
edm:ProvidedCHO	edm:isNextInSequence		optional
edm:ProvidedCHO	owl:sameAs		optional

An example of mapping to EDM

The Entity-Relationship Model was validated during its development particularly by mapping MODS to EDM on the basis of this model. In contrast to EDM, the MODS model is a record orientated XML-based model distinguishing and differentiating cataloguing units (e.g. mods:titleInfo is sub-divided into mods:title, mods:subtitle etc; mods:originInfo is sub-divided into mods:dateCreated,

mods:dateIssued etc.) instead of separate entities (e.g. edm:ProvidedCHO, edm:WebResource). What become clear during this mapping exercise is that the terms used with edm:type are not sufficient to map the respective data encoded in MODS. Some of the terms used in MODS cannot be mapped to an adequate term in EDM. The extension of the EDM terms by an "other" term could solve this problem. This could be by specialising it with a sub-property (although implementation of these in Europeana may be some time in the future).

Finally we proved the usability of this mapping by a test mapping of a mods dataset published by the Göttingen State and University Library.

```
<?xml version="1.0" encoding="UTF-8"?>
<MODS:mods
xmlns:xlink="http://www.w3.org/1999/xlink"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns="http://www.loc.gov/mods/v3"
xsi:schemaLocation="http://www.loc.gov/mods/v3 http://www.loc.gov/standards/mods/v3/mods-3-3.xsd"
xmlns:MODS="http://www.loc.gov/mods/v3 http://www.loc.gov/standards/mods/v3/mods-3-3.xsd">

  <MODS:identifier type="purl">http://resolver.sub.uni-goettingen.de/purl?PPN563016310</MODS:identifier>
  <MODS:location>
    <MODS:url>http://resolver.sub.uni-goettingen.de/purl?PPN563016310</MODS:url>
  </MODS:location>
  <MODS:titleInfo>
    <MODS:title>Elementa Ornithologica, Iconibus Vivis Coloribvs Expressis Illvstrata</MODS:title>
  </MODS:titleInfo>
  <MODS:language>
    <MODS:languageTerm type="code" authority="iso639-2b">lat</MODS:languageTerm>
  </MODS:language>
  <MODS:originInfo>
    <MODS:dateIssued encoding="w3cdtf">1774</MODS:dateIssued>
    <MODS:place>
      <MODS:placeTerm type="text">Regensburg</MODS:placeTerm>
    </MODS:place>
    <MODS:publisher>Weiss</MODS:publisher>
    <MODS:dateCaptured encoding="w3cdtf">2008-06-02</MODS:dateCaptured>
  </MODS:originInfo>
  <MODS:physicalDescription>
    <MODS:extent>233 pages</MODS:extent>
    <MODS:digitalOrigin>reformatted digital</MODS:digitalOrigin>
  </MODS:physicalDescription>
  <MODS:physicalDescription>
    <MODS:extent>[114] Bl.</MODS:extent>
  </MODS:physicalDescription>
  <MODS:classification authority="gdz">Zoologica</MODS:classification>
  <MODS:name type="personal">
    <MODS:role>
      <MODS:roleTerm type="code" authority="marcrelator">aut</MODS:roleTerm>
    </MODS:role>
    <MODS:namePart type="family">Schaeffer</MODS:namePart>
    <MODS:namePart type="given">Jakob Christian</MODS:namePart>
  </MODS:name>
  <MODS:identifier type="oai">gdz.sub.uni-goettingen.de:PPN563016310</MODS:identifier>
</MODS:mods>
```

Figure 1:XML file of the mapped MODS dataset

The result of this mapping was a RDF graph that includes properties that are not supported in the first implementation of EDM:

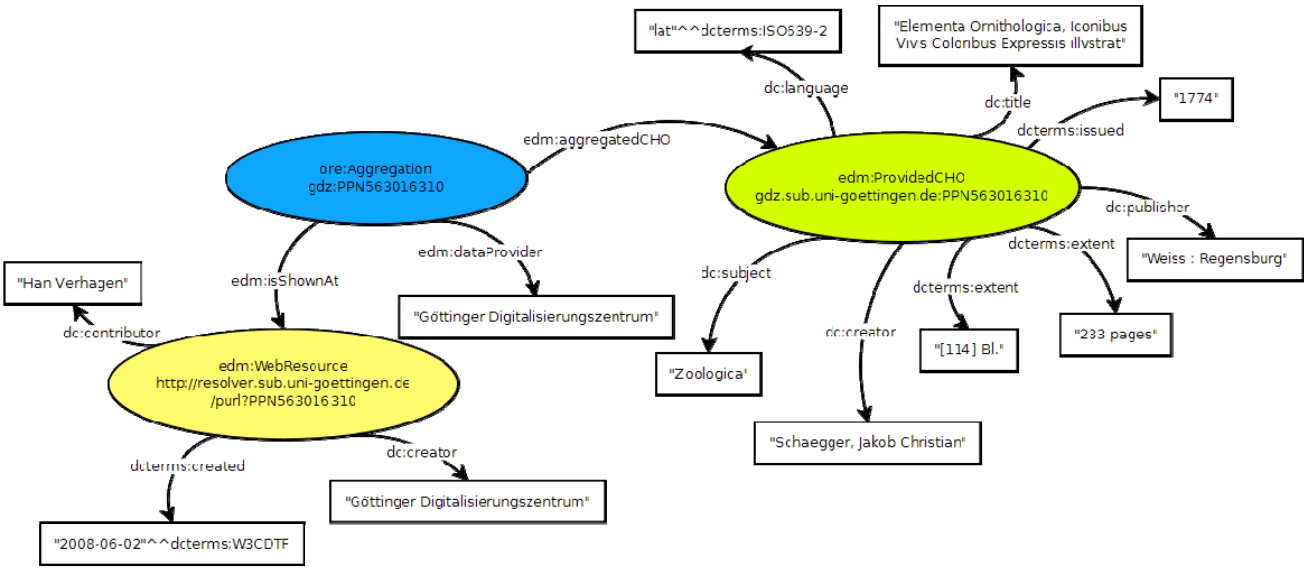

Figure 12: RDF graph of the mapped MODS dataset

A detailed version of this mapping is available on request.

The main findings of this mapping tests where the following:

- The Europeana Libraries model for monographs fits the mapping requirments of library-domain metadata describing monographs
- The entities and properties used in the first EDM implementation are however not sufficient for the description of monographic objects. In particular, more properties describing the web resource and a class for the item as a physical object are needed.

Annex 4 The Rare Book Model

This annex contains the section from the first version of D5.1 that related to the cataloguing at the itme level as opposed to the edition level). It has been removed from the body of the text as feedback from Europeana indicated that it was an unnecessary distinction. It is kept here purely for reference purposes.

Definition of the Cultural Heritage Object for Rare Books

While the edition of a work is the conceptual linchpin in the cataloguing of contemporary library material, the description of old and rare books requires an additional viewpoint. The standard approach to literature gathered in a library collection is the retrieval of the knowledge that is contained in the books, journals and articles etc. that form the collection. This makes all items of the same edition fully equivalent to each other. The approach of the historian, though, may differ in a way that the material evidence of production, ownership and use of the book – from which conclusions about the history of ideas or the social conditions at a certain time or similar might be drawn – receives the same amount or even more attention than its intellectual contents. This material angle is covered (in addition to the description of the edition) by most libraries in their

descriptions of old and rare books and must not be lost, when the metadata goes into Europeana. This is particularly important as Europeana has the great advantage of presenting the description of books as material objects in the context of the description of other material objects coming from the museum domain. Losing this information would mean that it would not be possible to answer a number of historic questions using Europeana data, which the data would have been capable of answering in its original library context; thus lessening the usefulness of Europeana for a whole scholarly community.

Therefore a conceptual model of how library data should be mapped to EDM must take the material viewpoint into consideration and must cater for recording this particular kind of data. This is not a side issue, for the vast majority of digital content produced by libraries nowadays are digitisations of old and rare books.

In order to fit into an overall model for library material that represents contemporary material in the same way as historic material, the backbone of the ORE Aggregation that describes a library item (book, journal etc.) in EDM – the edm:ProvidedCHO class – should not be used to describe the item itself, for this would create a situation where the same element would represent different things according to different use contexts that are not clearly distinguishable from the data itself. Instead, the Aggregation should comprise of a specific class representing the physical object (edm:PhysicalThing) that itself can be the subject for statements about the original copy from which a digitisation has been made. While the description of the edition the digitised item is a part of remains with edm:ProvidedCHO. This approach of adding an (optional) class for the physical object would meet the requirements of those users interested in the material aspects of a book, while, at the same time, fitting in perfectly with the edition-centred description of library material. This would further facilitate the mapping of existing library metadata to EDM, since most library metadata is actually organised this way. Therefore we recommend that Europeana shall include the edm:PhysicalThing Class in the next EDM implementation at the latest.

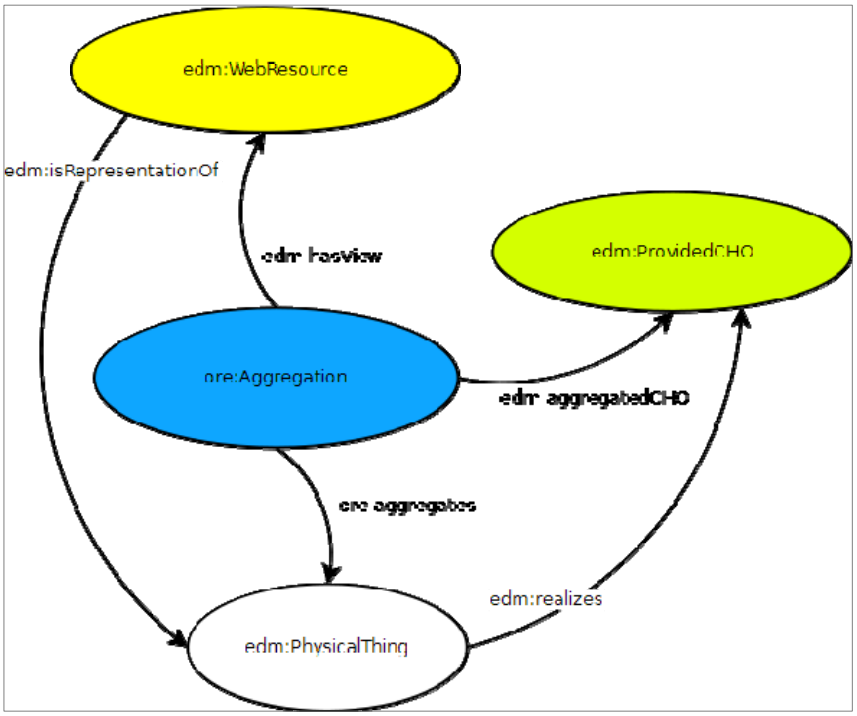

Figure 13: Entity-Relationship Model for rare books

Change history from previous versions

Date	Change	Author
04/05/2012	Many relatively minor changes were made following comments from the internal reviewers. These were mostly for clarification and further explanation: for example, more textual explanation was added to some diagrams and some statements or assumptions were expanded for clarity. There were no substantial changes to the content. There was one piece of feedback from Europeana that needed a more extensive consideration as it affected a major aspect of the report. This was discussed at the May WP meeting.	R Clayphan
10/07/2012	Feedback from Europeana led to the reconsideration of the "Rare Books Model" that had been proposed to overcome the "item vs edition level" discussion. Chapter five was rewritten to reflect this and the earlier conclusions were added to an Annex. A "Future Work" section has been added to take account of the desire to map to FRBR entities and related issues.	S Ruhle and R Clayphan
05/09/2012 and 08/11/2012	Further edits to clarify the changes made and finalise the document.	Robina Clayphan