DELIVERABLE

Project Acronym: Europeana Newspapers

Grant Agreement number: 297380

Project Title: A Gateway to European Newspapers Online

D4.4 Report on EDM for Newspapers

Revision: **1.0**

Authors: **Valentine Charles, The European Library**

Contributors **Reviewed by Robina Clayphan, Europeana Foundation and Anila Angjeli, Bibliothèque nationale de France.**

Project co-funded by the European Commission within the ICT Policy Support Programme

Dissemination Level		
P	Public	x
C	Confidential, only for members of the consortium and the Commission Services	

Revision History

Revision	Date	Author	Organisation	Description
V0.1	15-10-2012	Valentine Charles	The European Library	First draft outline of D4.4
V0.2	28-01-2012	Valentine Charles	The European Library	Version submitted for review
V0.3	30-01-2012	Valentine Charles	The European Library	Incorporation of the first reviewer's comments into a new version
V1.0	28-02-2012	Valentine Charles	The European Library	Incorporation of the second reviewer's comments into the final version

Statement of originality:

This deliverable contains original unpublished work except where clearly indicated otherwise.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

Table of Contents

1. Executive Summary

The deliverable *D4.4 Report on EDM for newspapers* has been written in the frame of Work Package 4 (WP4) of the Europeana Newspapers project. The aim of this report is to describe how newspapers metadata can be aligned on the *Europeana Data Model* (EDM).

The report begins by explaining the scope of the report and the methodology followed when writing it. The deliverable addresses newspapers as objects whether they are born-digital or digitized, and their content. Then the report highlights some key requirements addressed by newspapers. They are quite complex objects usually structured in a hierarchical way, the metadata usually contains very rich information which is not necessarily about the newspaper itself, as an object, but also relates to specific events, places, agents in relation to the intellectual content of the newspaper. The modeling to EDM needs to comply with these requirements in order to reveal the richness of newspaper content. The project also works with full-text which defines another set of specific requirements that need to be addressed.

The report continues by outlining the different solutions EDM provides to meet the requirements that have been listed. It lists the different classes and properties that have been selected from the main EDM specifications but also provides more details on some topics such as the representation of the hierarchical structure of newspapers.

This report has been written in parallel with the creation of the descriptive metadata, the specifications of the OCRisation practices and the definition of the Newspaper Content Browser developed by WP4 in the project. The last section of the report is therefore highlighting how this parallel work could affect the EDM model defined in the previous sections of the report. This future work does not go against the recommendations provided by this report but will probably define the areas where further work is required.

2. Introduction

This deliverable, D4.4 Report on EDM for newspapers, is delivered as part of the task Task 4.4: Aligning newspaper metadata to EDM:

"TEL will analyse the suitability of the Europeana Data Model (EDM) for use with newspaper content. This work will build on the work undertaken in the Europeana Libraries project to provide full-text materials to Europeana. A workshop will be held to facilitate this process. A report on providing recommendations of how to align newspaper content to EDM will be delivered by Month 12."

As described in the description of work of the project, the report will provide recommendations on how Newspapers content can be aligned on EDM.

3. Scope of the report and methodology

At the time of writing of this deliverable no newspaper content from the project has been provided to Europeana. Libraries partners in the project are still working on the OCRisation of their content[1]. It was therefore not possible to develop an EDM application profile based on real data and experiment with it.

The approach taken was to analyse first the different requirements that needed to be filled in order to support the complexity of newspaper data and the specificities of full-text. For each requirement an EDM modeling was proposed, detailing the required EDM classes and properties.

It was decided to build this work upon the achievements of the Europeana Libraries Project[2] and the work currently being done by the Europeana taskforce on hierarchical objects[3].
The workshop[4] organised in September 17th has allowed the validation of these requirements by a larger group of library experts.

This model will be further developed later in the project, according to the decisions made when achieving the task on the content browser[5].The content browser will improve the searchability of Newspapers content in Europeana and will offer a better way of navigating within these resources.

4. Main requirements or Newspapers

Libraries defined "***Newspapers***" in very different ways, but they agree on the fact that a newspaper is a serial publication issued at stated and frequent intervals and contains news on current events of special or general interest. The individual parts of the serial publication are listed chronologically or numerically and appear usually at least once a week. Newspapers are usually available in libraries under different forms: print, microfiche, digitized, and born-digital[6].
In addition libraries deal with newspaper content in very different ways, from cataloguing to digitisation, which makes the source data describing newspapers heterogeneous. It is however possible to identify main requirements the EDM model should take into account and be able to support.

4.1. Resources that are complex[7]

As highlighted in the main definition, newspapers can be considered as compound objects. They are indeed constituted of continuing resources issued in a succession of small issues or parts (usually numbered) that have no predetermined conclusion. Structured newspapers represent a hierarchy and a succession.

---

[1] The first delivery of content is planned M18 (August 2013)
[2] Report on the alignment of library metadata with the European Data Model (EDM) (D5.1) and Library domain metadata aligned with the Europeana Data Model (D5.2) at http://www.europeana-libraries.eu/web/guest/outcomes
[3] Europeana taskforce on hierarchical objects: http://pro.europeana.eu/web/network/europeana-tech/-/wiki/Main/Taskforce+on+hierarchical+objects
[4] MS 4.4 Workshop for aligning newspaper metadata to EDM
[5] Task 4.7: Newspaper Content Browser (M18 – M36)
[6] Definition inspired from the Online Dictionary of Library and Information Science: http://www.abc-clio.com/ODLIS/odlis_jk.aspx
[7] Hierarchical description of Provided CHOs from libraries by Stefanie Ruehle

It is possible to identify different levels for which descriptive metadata are potentially available. In the context of Europeana, each level could give rise to and could be the subject of a package of data submitted to Europeana[8]. Each level is a potential cultural heritage object that Europeana collects a description about and is searchable in the portal Europeana.eu.

For this report the following levels need to be considered:

- the article level. The article is the smallest 'standalone' information unit and the lowest level of granularity available. Libraries have reported during the workshop organised in September that descriptive metadata for this level are either poor or absent. However libraries would wish to 'rank' the article higher in importance as the smallest meaningful standalone unit in a newspaper.
- the issue: The issue is one of the successive parts of a newspaper.
- the volume. A volume has its own properties such as an independent pagination, foliation and signature. Individual volumes are usually numbered. This level is not always relevant for newspapers.
- the title of the newspaper. The title is the highest level of description. Most of the descriptive metadata are defined at the title level.

The **Figure 1** below describes this hierarchy. Some items and relations are highly important and relevant, in the diagram these are with continuous line. Other items and relations are optional or conditional, in the diagram these are with dotted line.

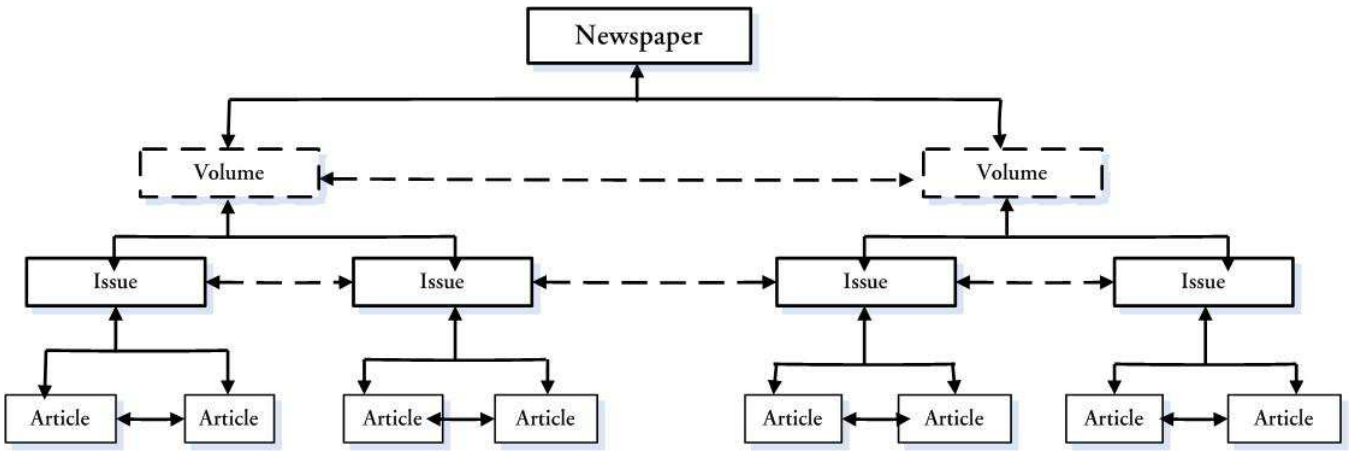

Figure 1 Multilevel description of a newspaper

Newspapers can also have supplements, annexes, articles that can either be described as one resource or described as part of the newspaper as shown in **Figure 2**.

[8] *Europeana Data Model Mapping Guidelines V1.0.1* http://pro.europeana.eu/web/edm-documentation

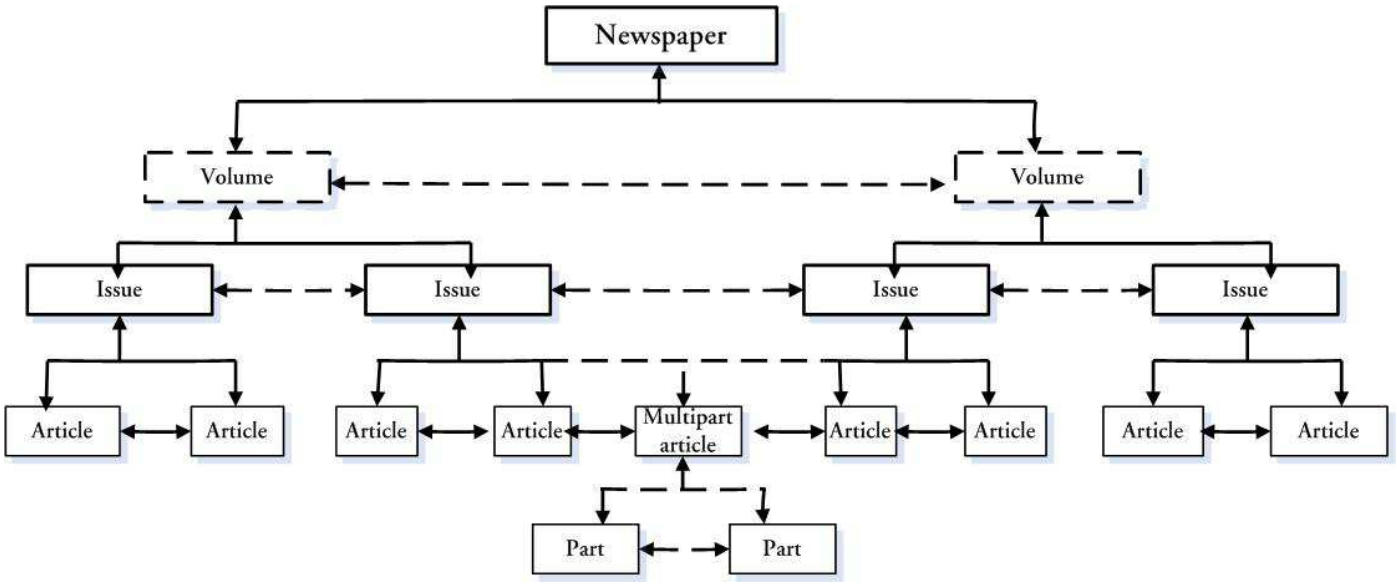

Figure 2 Multilevel description of a newspaper with a multipart article

The diagrams above are defining some mandatory relations and some optional ones. This difference comes from the fact that some levels are more likely to carry the descriptive metadata than others. For instance, libraries have in general lot of metadata for the title level but less for volumes and issues. Articles are also rarely described on their own. However, practices are very different when assigning metadata for these different levels. It is therefore important to keep the model as flexible as possible. The flexibility of the model is also key to ensure a good display and navigation within newspapers. The end-user needs to be able to grasp the complexity and richness of the resource. Information about the physical structure of a newspaper is needed for the building of viewers applications. They allow the user to browse through the digital resource in the same condition than for the physical resource.

4.2. Resources that contain rich contextual information

The definition of newspaper content highlights the fact that newspapers are “issued at stated and frequent intervals and contain news on current events of special or general interest”. Newspapers, since they are focused on events, potentially contain really rich information within their descriptive metadata. This information can be seen to be related not to the characterisation of the newspaper (or the other inferior hierarchical levels listed above) as objects of description, but to another resource related to these objects or to their intellectual content. Information specific to agents, time, places, and concepts can therefore be seen as potential new resources that can be described on their own.
The EDM model could greatly improve the description of contextual information by considering these types of information as additional resources.

4.3. Specific requirements for full-text[9]

[9] See Report on how the full-text content will be made available to Europeana (D4.3) from the Europeana Libraries project at http://www.europeana-libraries.eu/web/outcomes

One of the main tasks of the project is to deliver full-text material, which should also be taken into account when defining the model in EDM.

When analysing the requirements for full-text materials, a first distinction needs to be made for an object whether it is born-digital or has been digitized. When an object is born-digital, the file contains all the text, along with its visual attributes.

The situation is more complex for digitized full-text objects which are the result of a two step process. The digitization creates images from the analogue object; the OCRisation is then performed on the images to extract the text. The outcome of this process is the creation of two sets of files: the images and the textual information, which need to be combined.

Libraries have three strategies to do this as described in Table 1.

	Description	Example formats and standards
Single file	These solutions use specific file formats, which enable both the images and the text to be combined.	PDF, RTF
Files combined	The image and the text files exist separately, and are combined via external structural metadata that represent the complete logical and physical structure of the object (e.g. chapters, sections…)	Images: JPEG, PNG, a.o. Text: text/plain, HTML, TEI Metadata: METS
Files combined with coordinates	Similar to above, but with coordinates of every term in the corresponding image.	Images:JPEG, PNG, a.o. Text: text/plain, HTML, TEI Metadata: METS/ALTO

Table 1 Solutions for full-text representation after the digitisation process

In addition, libraries have different practices when creating the descriptive metadata for full-text at the volume, issue or article level. Libraries do not always create descriptive metadata because it may be too expensive. Hence, the retrieval of the digital object is based on the full-text and not the metadata. The metadata describing the full-text resource might be very poor and may consist only in a title and a date and an *IsPartOf* relationship pointing to another metadata record that describes the hierarchically superior level. Also, metadata about full-text digital objects may be hierarchical, requiring referencing to other metadata records.

The EDM model for full-text objects must accommodate these different solutions in terms of description of the full-text content.

Furthermore, additional requirements need to be created regarding the exchange of full-text. The Europeana context puts in place different types of actors:

- data providers, which hold the full-text objects and provide to end-users an access to them
- aggregators, which harvest the full-text objects and provide search and retrieval services.

The following requirements need to be defined in order to support the exchange of full-text contents between a data provider and an aggregator:

Requirement 1) The availability of full-text must be stated explicitly in the metadata.

Requirement 2) The URLs to the views of the digital objects must be explicitly stated in the data.

Requirement 3) The full-text resources must be referenced via direct URLs to one or more files.

Requirement 4) When more than one full-text resource is associated with a digital object, it should be possible to represent their sequential order.

Requirement 5) URLs to access specific parts of the digital objects (for example, to a section or page) may be provided in the data.

5. Main requirements of EDM

As shown in the European Newspaper Survey Report[10] very few libraries are currently applying Optical Character Recognition (OCR) techniques (36% or 17 out of the 47 respondents). In addition, most of the libraries are using Dublin Core as the metadata standard for their descriptive metadata. These practices lead to very poor metadata and therefore a poor representation of the newspapers content. These observations lead to the assumption that EDM might help libraries to more accurately describe their collections. In addition, aligning the descriptions of newspapers to EDM has the advantage of bringing these descriptions in a federated homogenous environment, which enhances the value of resources and increases their visibility.

As described in the *Europeana Data Model Primer V1.0*[11] , EDM has been designed around some key requirements for Cultural Heritage Objects.

- **R1**. distinction between “provided objects” (painting, book, movie, archaeology site, archival file, etc.) and their digital representations

- **R2**. distinction between objects and metadata records describing an object

- **R3**. multiple records for the same object should be allowed, containing potentially contradictory statements about this object

- **R4**. support for objects that are composed of other objects

- **R5**. compatibility with different abstraction levels of description (e.g. if a provider wishes to submit descriptions that follow the distinctions introduced in FRBR Group 1 [FRBR])

- **R6**. EDM provides a standard metadata format that can be specialized

- **R7**. support for contextual resources, including concepts from controlled vocabularies

Requirement 1 is very important in the context of the description of newspaper content since it is

[10] Deliverable 4.1 of the project.
[11] Available at http://pro.europeana.eu/web/edm-documentation

allowing for the distinction of metadata describing the a newspaper as a physical object from those related to its digital representation (full-text for instance). Then **requirement 4** allows the representation of complex objects such as newspaper; and finally **requirement 7** gives the opportunity to provide enriched data description for contextual resources.

6. Europeana Data Model for Newspapers

This section lists the EDM classes and properties that have been selected for representing newspapers content and then providing some details on specific requirements which have been formulated in the previous sections,

6.1. Selected EDM classes and properties

The three following tables show properties that have been selected from EDM[12] to represent the newspaper content in the three core classes of the model: edm:ProvidedCHO, edm:WebResource, and edm:Aggregation. The **Figure** 3 below shows how these different classes are organised and related to each other. This choice of properties and classes complies with the first EDM requirement (R1) aiming at the distinction between a CHO and its digital representation. The properties can be applied to every level to describe, according to the metadata available.

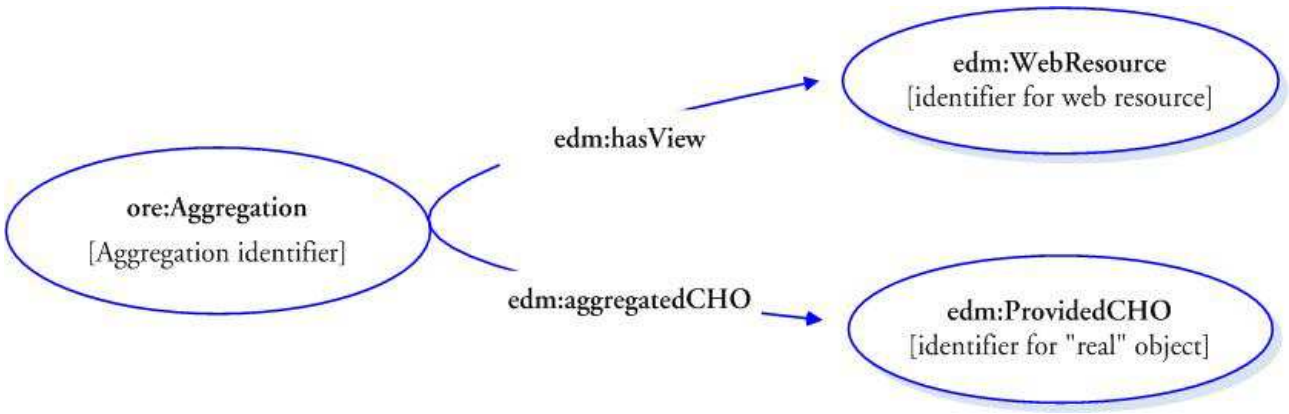

Figure 3 Main EDM classes used in the model for newspaper.

The rows shaded in grey are those properties Europeana Newspapers would like to include in the profile but which are not included in Europeana's current implementation of EDM. The tables have been adapted from the EDM specification for this report. The tables contain further information and comments on the use of the properties in the model for newspapers.

6.1.1. Properties for the ProvidedCHO

The ProvidedCHO, for newspapers, 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 original data.

[12] Classes and properties have been selected according to http://europeanalabs.eu/wiki/EDMObjectTemplatesProviders

Properties for the Provided CHO			
Property	**Definition**	**Obligation**	**Repeat-able**
dc:contributor	An entity responsible for making contributions to 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
dc:creator	An entity primarily responsible for making 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
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.)	optional	Yes
dc:format	The file format, physical medium, or dimensions 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. (Note: Includes the place of publication)	optional	Yes
dc:relation	A related 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
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	mandatory for Europeana Newspaper	No (different from EDM)

	granularity.	s Optional in Europeana	
dcterms:abstract	A summary of the resource.	optional	Yes
dcterms:alternative	An alternative name for the resource.	optional	Yes
dcterms:bibliographicCitation	A bibliographic reference for the resource.	optional	No
dcterms:extent	The size or duration of the resource.	optional	No (different from EDM)
dcterms:hasFormat	The described resource pre-existed the referenced resource, which is essentially the same intellectual content presented in another format.	optional	Yes
dcterms:hasPart	The described resource includes the referenced resource either physically or logically.	optional	Yes
dcterms:hasVersion	A related resource that is a version, edition, or adaptation of the described resource.	optional	Yes
dcterms:isFormatOf	A related resource that is substantially the same as the described resource, but in another format.	optional	Yes
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
dcterms:isReferencedBy	The described resource is referenced, cited, or otherwise pointed to by the referenced resource.	optional	Yes
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 (different from EDM)
dcterms:isVersionOf	A related resource of which the described resource is a version, edition, or adaptation.	optional	No
dcterms:medium	The material or physical carrier of the resource.	optional	Yes
dcterms:references	A related resource that is referenced, cited, or otherwise pointed to by the described resource.	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:tableOfContents	A list of subunits of the resource.	optional	No
dcterms:temporal	Temporal characteristics of the resource.	optional	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 strongly recommended to describe a succession of resources within a newspaper.	optional	No
edm:isSuccessorOf	This property captures the relation between the continuation of a resource and that resource. This applies a story, a newspaper, a journal etc. No content of the successor resource is identical or has a similar form with that of the precursor. The similarity is only in the context, subjects and figures of a plot. Successors typically form part of a common whole – such as a trilogy, a journal, etc	optional	No
edm:type	The Europeana material type of the resource	mandatory	No
owl:sameAs	Indicates that two URI references actually refer to the same thing.	optional	Yes

6.1.2. 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: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	optional	Yes
dc:format	The file format, physical medium, or dimensions of the resource. Encode as a MIME type.	mandatory	No different from EDM)
dc:rights	Information about rights held in and over the resource.	optional	Yes
dc:source	A related resource from which the described resource is derived.	optional	Yes
dcterms:conformsTo	An established standard to which the described resource conforms.	optional	Yes
dcterms:created	Date of creation of the resource. Encode at W3CDTF.	mandatory	No different from EDM)
dcterms:extent	The size or duration of the resource.	optional	No different from

			EDM)
dcterms: hasFormat	A related resource that is substantially the same as the pre-existing described resource, but in another format.	mandatory if a fullText resource class is used	Yes
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. This property is recommended to describe pagination information.	optional	No

6.1.3. Properties for the Aggregation

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

Properties for the Aggregation			
Property	**Definition**	**obligation**	**repeat-able**
ore:aggregates	Only stated in principle via edm:hasView and edm:aggregated CHO statements.	optional	Yes
edm:aggregatedCHO	This property associates an ore:Aggregation with the cultural heritage object it is about.	mandatory	No
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:dataProvider	The name or identifier of the organisation that contributes data to Europeana	mandatory	No
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: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	mandatory*	Yes

	not provided		
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 one of those listed in the **Europeana Rights Guidelines** (http://pro.europeana.eu/technical-requirements)	mandatory	Yes

6.2. Support of complex objects

EDM allows the representation on the horizontal and vertical relationships between the different levels constituting a newspaper. These relationships can be expressed in different classes in EDM whether the hierarchy is in the CHO or in the digital object. In both cases the hierarchies could be potentially different (differences may be introduced during the digitization).
The vertical relationships in a newspaper can be expressed with two specific EDM properties. They can be use to express the relation between the whole resource and its parts.

- The has-part relation (dcterms:hasPart property) which illustrates a top down relation:
- The is-part-of relation (dcterms:isPartOf property) which illustrates a bottom up relation.

In the figure thee relations are noted in blue.

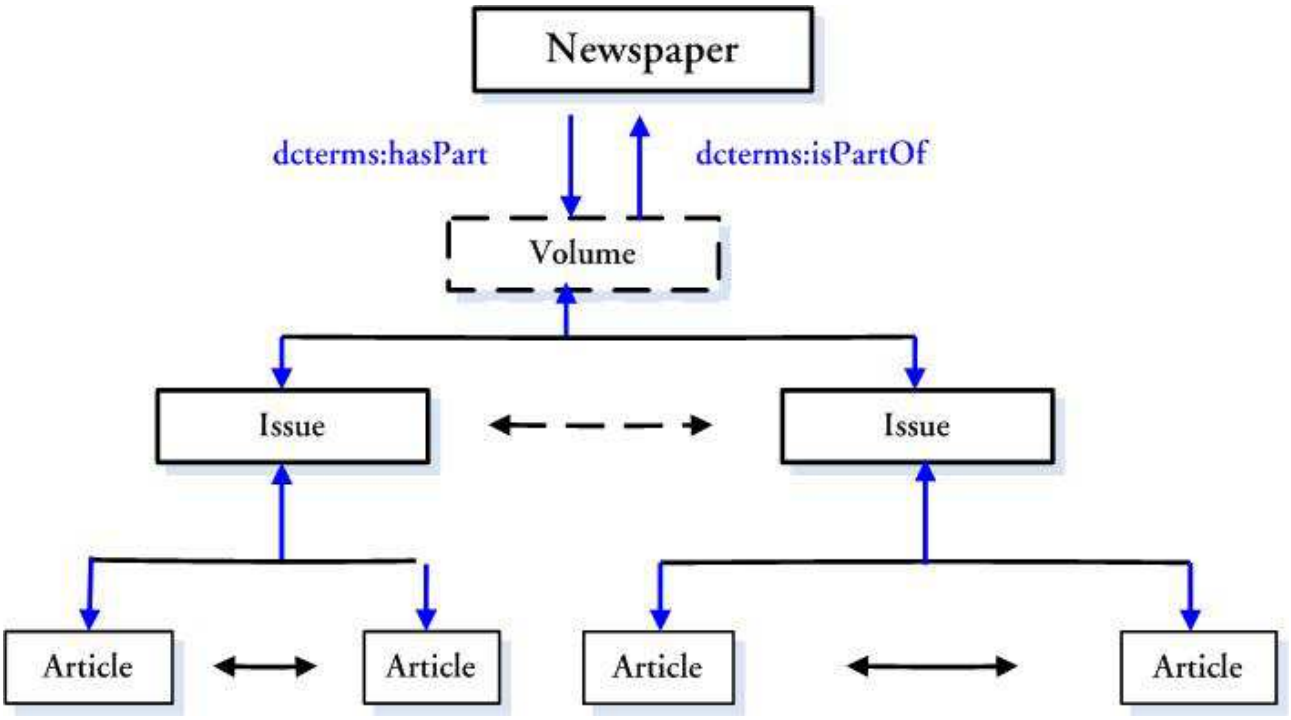

Figure 4 Representation of vertical relationships in a hierarchy of resources

The horizontal relationships in a newspaper can be expressed with the is-next-in- sequence property. This property describes relations between the parts of a resource given by the consecutive numbering of the parts or by pagination. It allows the navigation from the lower number to a higher number (e.g. issue 3 is the next in sequence of 2).

In the figure these relations are represented in blue.

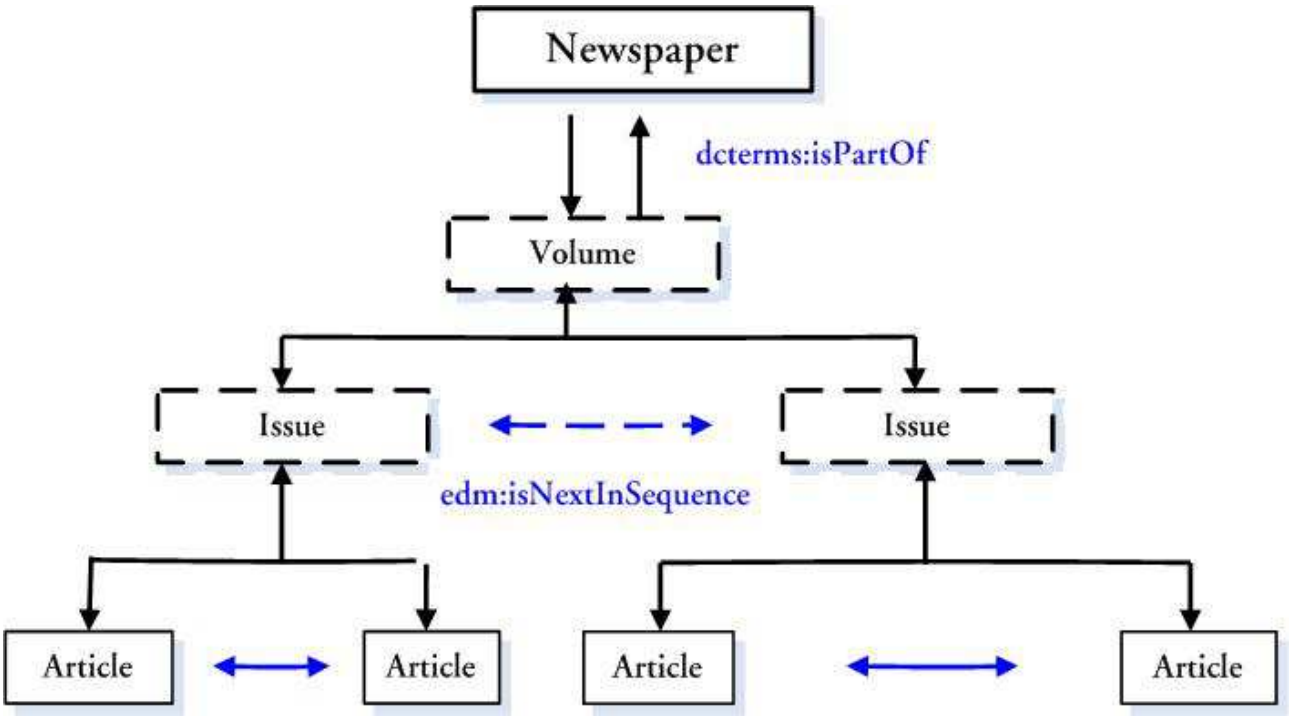

Figure 5 Representation of horizontal relationships in a hierarchy of resources

These properties allow the representation between different levels that can give rise to a ProvidedCHO. It is therefore important to also specify a type for each cultural heritage object. The model uses dc:type with a literal to describe this information. However it would be better practices to use a value from a controlled vocabulary. The MARC genre list[13], the ontology BIBO[14] or RDA are offering different solutions which could be interesting for later use in the project.

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

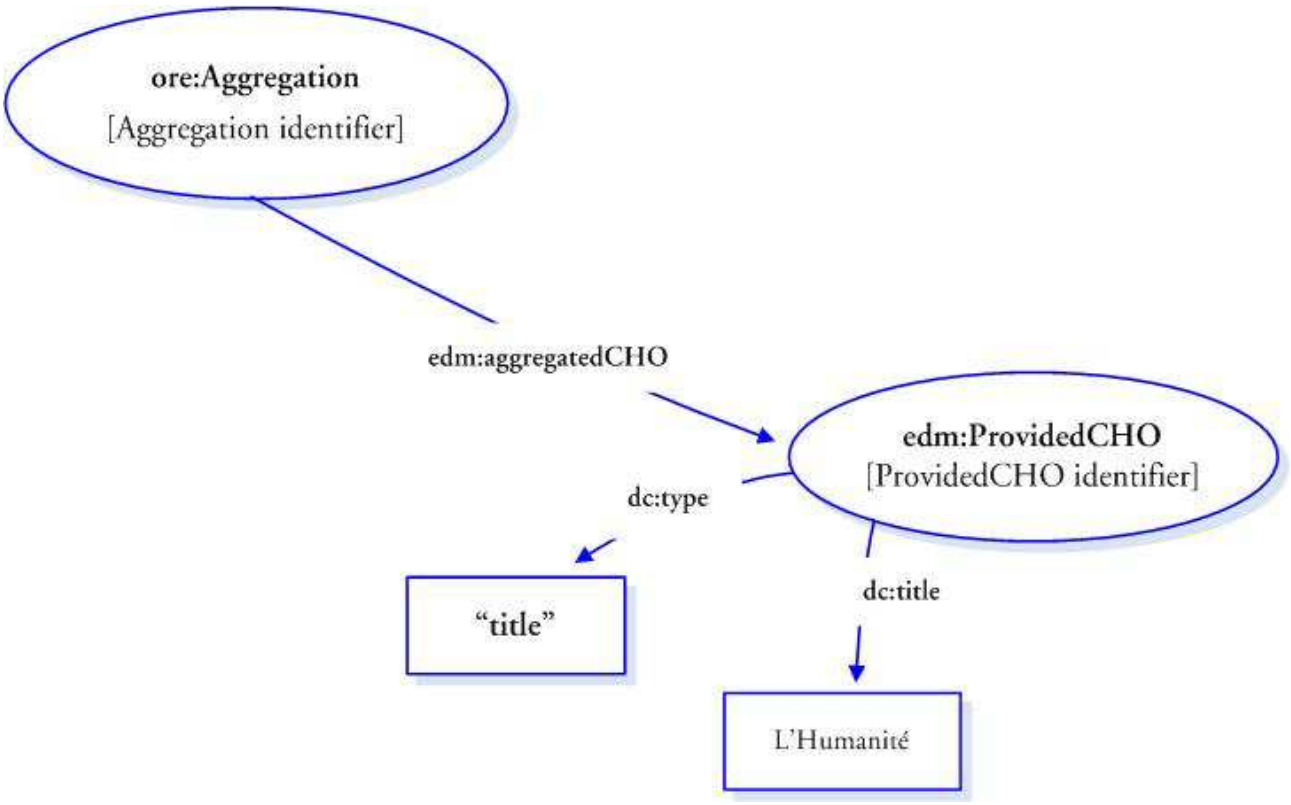

Figure 6 Representation of the type of a ProvidedCHO

6.3. Support of contextual information

EDM is offering a selection of properties which Europeana Newspapers would like to use to provide richer description for the contextual entities. Each of these contextual entities can be applied for every type of ProvidedCHO depending of the descriptive metadata available at a specific level. It is likely that most of the relevant information will be found in the title of the newspaper.

EDM defines four main classes to define Agent, Place, TimeSpan and Concept. A fifth class has been defined for event but is not currently implemented by Europeana. Each contextual resource is defined as a new class and is attached to the ProvidedCHO by an appropriate property as defined in the **Figure 7.** The properties describing the contextual entity may refer to a thesaurus or authority file which will link to further information related to that entity. For example, the identifier for an Author name in an authority file will give access to fuller information about that Author.

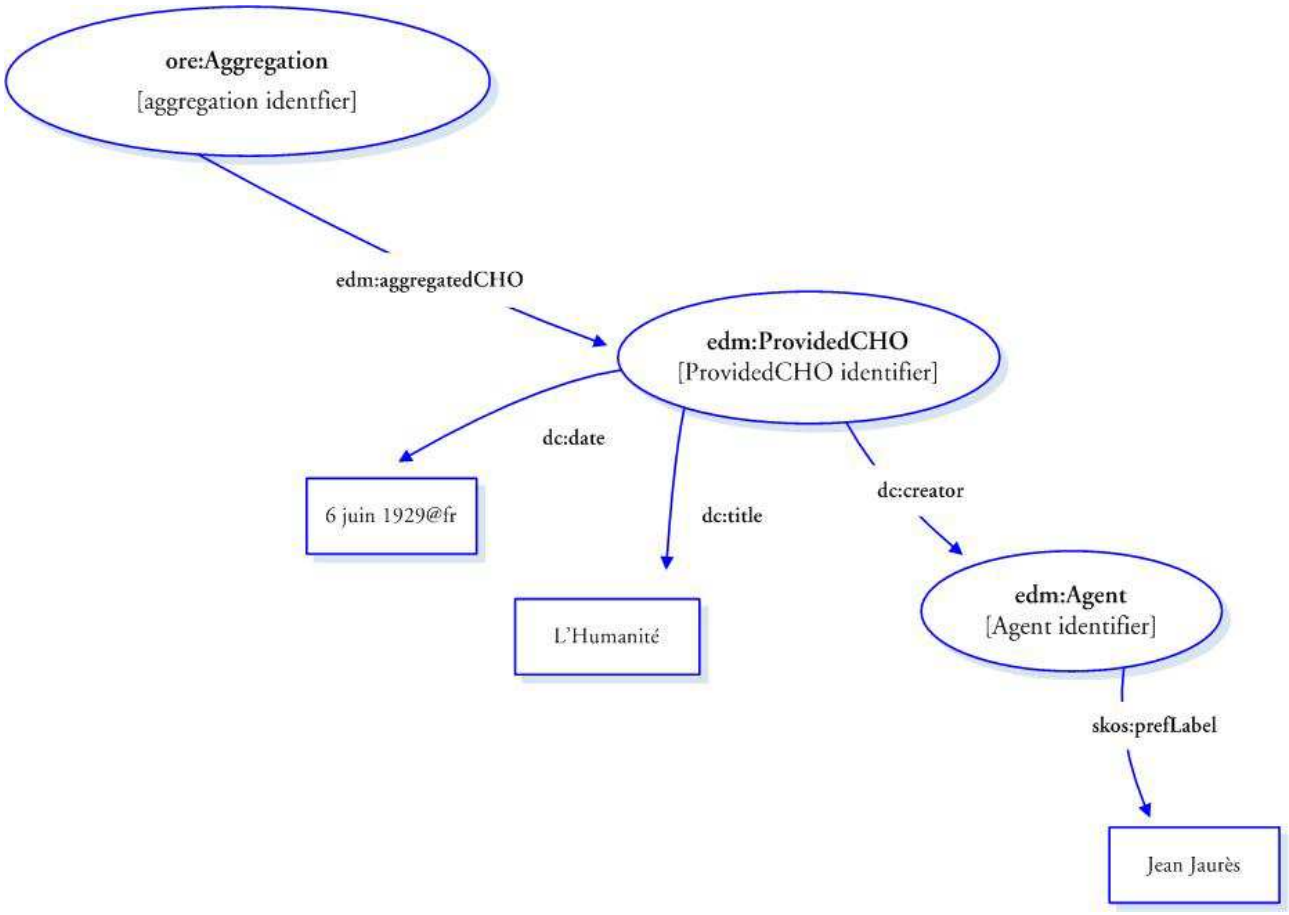

Figure 7 Representation of an agent (edm:Agent) in relation to a Cultural Heritage Object.

6.3.1. Properties for edm:Agent

The class edm:Agent describes people, either individually or in groups, who have the potential to perform intentional actions for which they can be held responsible.

Properties for Agent			
Property	**Definition**	**obligation**	**repeatable**
skos:prefLabel	The preferred form of the name of the agent.	optional	No
skos:altLabel skos:hiddenLabel	Alternative forms of the name of the agent.	optional	Yes
skos:note (for e.g., biographical notes)	A note about the agent e.g. biographical notes.	optional	Yes
dc:date	A significant date associated with the Agent.	optional	Yes
dc:identifier	An identifier of the agent.	optional	Yes

edm:begin	The date the agent was born/established.	optional	No
edm:end	The date the agent died/terminated.	optional	No
edm:hasMet	the identifier of another entity which the agent has "met" in a broad sense.	optional	Yes
edm:isRelatedTo ***(for generic relations to other agents, especially)***	The identifier of other entities, particularly other agents, with whom the agent is related in a generic sense.	optional	Yes
edm:wasPresentAt	the identifier of an event at which the agent was present.	optional	Yes
foaf:name	The name of the agent as a simple textual string.	optional	Yes
rdaGr2:biographicalInformation	Information pertaining to the life or history of the agent.	optional	Yes
rdaGr2:dateOfBirth	The date the agent (person) was born.	optional	No
rdaGr2:dateOfDeath	The date the agent (person) died.	optional	No
rdaGr2:dateOfEstablishment	The date on which the agent (corporate body) was established or founded.	optional	No
rdaGr2:dateOfTermination	The date on which the agent (corporate body) was terminated or dissolved.	optional	No
rdaGr2:gender	The gender with which the agent identifies.	optional	No
rdaGr2:professionOrOccupation	The profession or occupation in which the agent works or has worked.	optional	Yes
owl:sameAs	The URI for a set of data describing an agent. E.g. an authority record, a Wikipedia article, a DBpedia URI, or a VIAF cluster.	optional	Yes

6.3.2. Properties for edm:Place

A spatial location identified by the provider and named according to some vocabulary or local convention. In the context of Europeana Newspaper, this class could be used to provide further information on the place of publication.

Properties for Place			
Property	**Definition**	**obligation**	**repeatable**
wgs84_pos:lat	The latitude of a spatial thing (decimal degrees).	optional	No
wgs84_pos:long	The longitude of a spatial thing (decimal degrees)	optional	No
wgs84_pos:alt	The altitude of a spatial thing (decimal metres above the reference)	optional	No

wgs84_pos:lat_long	A comma separated representation of a latitude, longitude co-ordinate.	optional	No
skos:prefLabel	The preferred form of the name of the place.	optional	No
skos:altLabel skos:hiddenLabel	Alternative forms of the name of the place.	optional	Yes
skos:note	Information relating to the place.	optional	Yes
dcterms:hasPart	identifier of a place that is part of the place being described.	optional	Yes
dcterms:isPartOf	identifier of a place that the described place is part of.	optional	Yes
owl:sameAs	URI of a Place as given by a specific repository, data set, etc. E.g. Geonames	optional	Yes

6.3.3. Properties for edm:TimeSpan

A period of time having a beginning, an end and a duration.

Properties for Time Span			
Property	**Definition**	**obligation**	**repeat-able**
skos:prefLabel	The preferred form of the name of the timespan or period.	optional	No
skos:altLabel, skos:hiddenLabel	Alternative forms of the name of the timespan or period.	optional	Yes
skos:note	Information relating to the timespan or period.	optional	Yes
dcterms:hasPart	The identifier of a timespan which is part of the described timespan.	optional	Yes
dcterms:isPartOf	The identifier of a timespan of which the described timespan is a part.	optional	Yes
edm:begin	The date the timespan started.	optional	No
edm:end	The date the timespan finished.	optional	No
crm:P79F.beginning_is_qualified_by	Qualifying information about the start of the timespan – such as degree of certainty, precision, source etc.	optional	Yes
crm:P80F.end_is_qualified_by	Qualifying information about the end of the timespan – such as degree of certainty, precision, source etc.	optional	Yes
owl:sameAs	The URI of a timespan	optional	Yes

6.3.4. Properties for skos:Concept

A unit of thought or meaning that comes from an organised knowledge base (such as subject terms from a thesaurus or controlled vocabulary) where URIs or local identifiers have been created to represent each concept. This class would allow the description of concepts used in a newspaper and would allow us to have a better understanding of the different topics mentioned in an article for instance.

Properties for Concept			
Property	**Definition**	**obligation**	**repeat-able**
skos:prefLabel	The preferred form of the name of the concept.	optional	No
skos:altLabel**,** skos:hiddenLabel	Alternative forms of the name of the concept.	optional	Yes
skos:broader, skos:narrower, skos:related	The identifier of a broader concept in the same thesaurus or controlled vocabulary. The identifier of a narrower concept. The identifier of a related concept.	optional	Yes
skos:broadMatch, skos:narrowMatch, skos:relatedMatch	The identifier of broader, narrower or related matching concepts from other concept schemes.	optional	Yes
skos:exactMatch, skos:closeMatch	The identifier of close or exactly matching concepts from other concept schemes.	optional	Yes
skos:note	Information relating to the concept.	optional	Yes
skos:notation	The notation in which the concept is represented. This may not be words in natural language for some knowledge organisation systems e.g. algebra	optional	Yes
skos:inScheme (*URI should resolve to something meaningful*)	The URI of a concept scheme	optional	Yes

6.3.5. Properties for edm:Event

EDM offers the possibility of organising the descriptive information for object in an event based model which is focusing more on the event rather than on the object itself, a newspaper in this project.

Since newspapers mainly contain news about particular events, an event based model would be really interesting. With such a model it would be possible to create a network of entities which would reconstitute either the object history or some more abstract object such an historical event with resource. Such a model would show how different objects, place, agents are related to each others.

Even if the Event class is available and has been specified in EDM, this class won't be immediately implemented. The model and the list of properties below have been directly taken from the EDM specifications but could be subject to changes depending of future implementation choice.

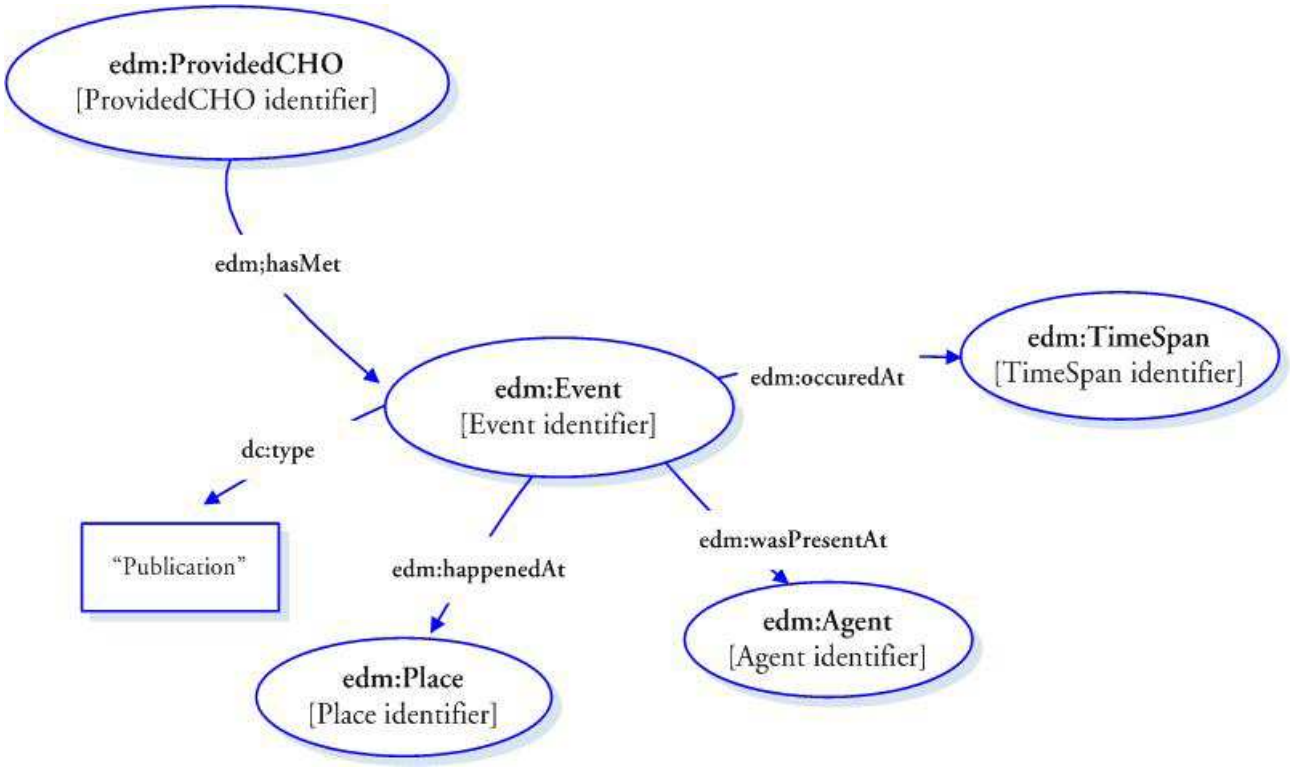

Figure 8 Event base model for newspaper

As described in the model above, the event class is acting as a "hub" that relates to other entities. The below properties can be used to describe the event but the most important ones are:

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

Properties for Event			
Property	**Definition**	**obligation**	**Repeat-able**
edm:happenedAt	This property associates an event with the place at which the event happened	optional	Yes
edm:occuredAt	This property associates an event to the smallest known time span that overlaps with the occurrence of that event	optional	Yes
owl:sameAs	URI of an Event	optional	Yes

skos:prefLabel	The preferred form of the name of the event	optional	No
skos:altLabel, skos:hiddenLabel	Alternative forms of the name of the event	optional	Yes
skos:note	Information relating to the event	optional	Yes
dc:identifier	Identifier of the event	optional	Yes
dcterms:hasPart	The identifier of an event which is part of the described event.	optional	Yes
dcterms:isPartOf	The identifier of an event of which the described event is a part.	optional	Yes
crm:P120F.occurs_before	The identifier of the relative chronological sequence of two temporal entities in an event	optional	Yes
edm:hasType	This property relates a resource with the concepts it belongs to in a suitable type system such as MIME or any thesaurus that captures categories of objects in a given field (e.g., the "Objects" facet in Getty's Art and Architecture Thesaurus). It does not capture aboutness.	optional	Yes
edm:isRelatedTo	The identifier of other entities, particularly other events, with which the event is related in a generic sense	optional	Yes

6.4. Potential support for full-text

As described in the previous section, Europeana Newspapers would need EDM to support full-text. The current EDM model is mainly based on the exchange of metadata about the digital objects, so it makes sense to build upon the existing framework to allow the exchange of digital object data to enable full-text aggregation.

Based of the requirements formulated for full-text, the following model defined in the frame of the Europeana Libraries project[15] is proposed for integration in EDM.

[15] D4.3 EDM for Full Text at http://www.europeana-libraries.eu/web/guest/outcomes

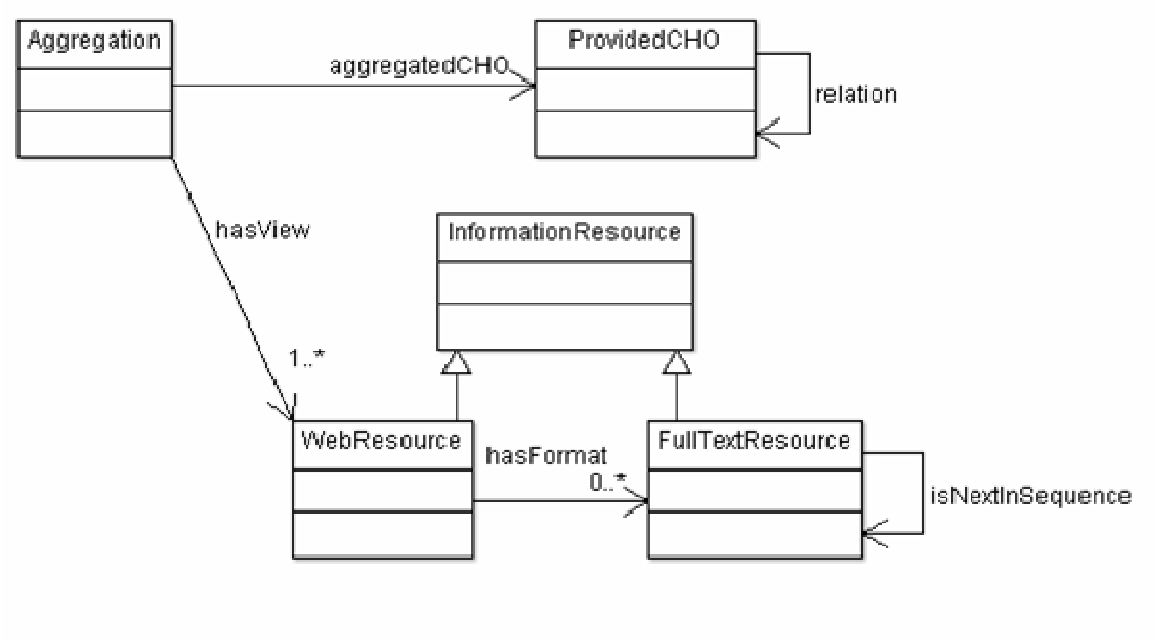

Figure 9 EDM model for full-text

This model is re-using the subset of classes and properties that is relevant to comply with the full-text requirements presented in the previous section. Even though most of the requirements are already met by the current EDM, the full support of full-text requires the creation of a new class named *FullTextResource*.

FullTextResource is a subclass of InformationResource. It allows the representation of the individual full-text content, separately from the end-user access versions, which are modeled in EDM in the class WebResource. A FullTextResource is included in the digital object metadata by using a hasFormat relationship from the WebResource to the FullTextResource. This allows the aggregators to index the full-text, and to provide the end-user with access to the views of the digital object.

Class name:	FullTextResource
Namespace	Europeana namespace most likely
URI	<to be assigned>
Label	Full-text Resource
Definition	InformationResources that have at least one WebResource and an URI and contain a full text representation of a digital resource.
Subclass of	edm:InformationResource
Obligation & Occurrence	The relation between Web Resources and the Full-text Resources is zero to many.
Example	A full-text resource containing the full-text of a book in machine-readable form.
Rationale	When full-text is available for a digital object, it must be represented in

	specific class, since in some cases, the Web Resource that provides the view of the digital object to the end-users, does not carry the full-text in machine-readable form. The full-text resources are associated with a Web Resource by the use of dcterms:hasFormat properties from the Web Resource to the Full-text Resource. The sequential order of Full-text resources within a cultural heritage object should be represented with edm:isNextInSequence.

Figure 10 Description of the FullTextResource class

The other properties mentioned in this model have been described in the section 6.1 Selected EDM classes and properties

The table below describes how the requirements for full-text are met by this EDM based model.

Requirement	Description
R1 – When full-text is available for a digital object, it must be explicitly stated in the data.	This requirement is met by the existence of *FullTextResource* class, which explicitly states the existence of full-text content.
R2 – The URLs to the views of the digital objects must be explicitly stated in the data.	The *WebResource* class, when associated with an aggregation through a *hasView* relation meets this requirement. The *WebResource*'s URIs explicitly states how to access the digital object's views.
R3 – The full-text resources can be referenced via direct URLs to one or more files with textual content.	This requirement is met by the URI of the *FullTextResource*. The existence of multiple *FullTextResource*'s for a digital object is met by allowing a *WebResource* to have more than one *hasFormat* relation with *FullTextResource*'s.
R4 – When more than one full-text resource is associated with a digital object, it should be possible to represent their sequential order.	The *isNextInSequence* relation between two *FullTextResource*'s meets this requirement.
R6 – URLs to access specific parts of the digital objects (for example, to a section or page) may be provided in the data.	This requirement is met by the *hasFormat* relation between *WebResource*'s and the *FullTextResource*'s. The *WebResource* provides the view URL for every *FullTextResource* it has a *hasFormat* relation with.
R7 – Metadata about full-text digital objects may be hierarchical, requiring referencing to other metadata records.	This requirement is met by EDM through the possible types of relations between *ProvidedCHO*s, in particular the *isPartOf* and the *hasPart* relations.

Table 2 Requirements matrix for full-text

7. Implementation of EDM in the Europeana Newspapers project

The model described above is not settled yet and could be subject to changes depending on the choices made by the project in the next two years. The model could have to be changed along three different lines: first the way descriptive metadata will be produced and delivered during the project, the requirements defined for end-users and then the requirements defined for the Newspaper Content browser.

First the properties of the source metadata submitted by libraries in the project may affect this model. All the levels constituting a newspaper might not be described in the same way from one library to another. Therefore a choice might be needed at the project level regarding the granularity of the descriptions sent for ingestion into Europeana. Since titles have the richest data, it is likely the project will focus on this level. Data Providers can also decide which level of granularity to provide. These choices are driven by the type of data they have but also the type of functionality they offer on their website. Lot of libraries are using a viewer on their original website which make possible to look at the images of digitized objects. Europeana by itself does not provide a viewer for this purpose, but links to the viewers of the data providers. For this reason it is not defined as necessary to model the full physical structure of newspapers into EDM. On the other hand a direct link to a semantic unit can only work, if there is a link to the first page of this unit in the context of the viewer. How this link looks depends on the viewer the data provider uses and should therefore be created by the data provider himself.

Secondly the model needs to take into account the end-users' needs when searching, browsing newspaper content in Europeana. All the levels might not be interesting for an end-user. For instance the volume level quite often only contains information relating to binding, pagination etc. which is useful for librarians in managing the resources but less important for end-users. Articles are really rich in information but having them as a search entry in Europeana might overwhelm the service with multiple hits. In this particular case clustering process would be required to make the content re-usable by end-users.

Finally Europeana currently does not have the capability to run a full-text service, and this is therefore being delivered by The European Library. As part of work package 4, a Newspaper Content Browser will be delivered in order to provide a better representation of newspapers to the end-users. The requirements for the browser will also influence the decision made in the EDM modelling. Initial requirements have been formulised for the content browser but have not been settled yet.

- The content browser will allow search at issue and article levels. This requirement will determine the level of granularity of descriptions provided to Europeana.
- It will include temporal and spatial browsing information. This requirement will justify the use of the contextual resource classes defined by EDM.
- It will support good navigation and browsing functionalities. This requirement will define which specific properties in EDM have to be used in the model for newspapers.

8. Conclusion

This deliverable should be considered as a set of specific requirements for newspaper content and a set of recommendations for modeling newspaper data in EDM. It considers the main issues when aligning newspaper data with EDM. Each section could give rise to a more detailed piece of work. Europeana is for instance currently working on the representation of hierarchical objects in EDM. The report which will be produced will bring additional details to this deliverable. The work done so far is the initial step in a longer process involving the creation of the descriptive metadata and the creation of the full-text content which necessarily implies the constant revision of this model during the course of the project.