



# DELIVERABLE

**Project Acronym:** Europeana Cloud  
**Grant Agreement number:** 325091  
**Project Title:** Europeana Cloud: Unlocking Europe's Research via The Cloud

---

## D4.4 Recommendations for enhancing EDM to support research-oriented content

**Revision: Version 1.0**

---

### Authors:

**Valentine Charles, Europeana Foundation**  
**Nuno Freire, Europeana Foundation**  
**Hugo Manguinhas, Europeana Foundation**  
**Peter Vos, VU Amsterdam Library**  
**Glen Robson, National Library of Wales**

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
1	06/10/2015	Valentine Charles	European Foundation	First outline and definitions of the case studies
2	11/12/2015	Valentine Charles, Nuno Freire	European Foundation	Additions and completion of Annex 2
3	22/12/2015	Valentine Charles, Nuno Freire	European Foundation	Additions and completion of Annex 1
4	06 to 18/01/2016	Valentine Charles	European Foundation	Additions and revisions based on the authors listed above feedback.
5	21/01/2016	Valentine Charles	European Foundation	Draft version sent for review
6	28/01/2016	Karen Estlund Antoine Isaac	Penn State University Libraries, USA European Foundation	Internal and External review of the document
7	29/01/2016	Valentine Charles	European Foundation	Final version V1.0

### **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This deliverable contains original unpublished work except where clearly indicated otherwise.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

<b>INTRODUCTION</b>	<b>3</b>
<b>1. METHODOLOGY AND CASE STUDIES</b>	<b>3</b>
1.1 Case studies 1 and 2: Linking (a) digital object(s) to its corresponding CHO(s) (case provided by the VU Amsterdam)	4
1.2 Case study 3: Representing complex digital objects (case provided by the National Library of Wales)	4
<b>2. REPRESENTING CONTENT IN EDM</b>	<b>5</b>
<b>3. LINKING CHO TO THEIR DIGITAL REPRESENTATIONS</b>	<b>8</b>
3.1 Simple linking between a CHO and its digital representation	9
3.2 Complex linking between a CHO and its digital representations	12
<b>4. REPRESENTING FULL-TEXT OBJECTS IN EDM</b>	<b>13</b>
4.1 Functional requirements for full-text objects	13
4.2 Modelling solutions for full-text	16
4.2.1 Representing full-text	16
4.2.2 EDM extension for representing full-text	17
<b>CONCLUSION</b>	<b>22</b>
<b>ANNEXES</b>	<b>23</b>
Annex 1: Case study 2: Complex linking between content and metadata at the VU Amsterdam	23
Annex 2: Case study 3: Newspapers content in the IIF service of the National Library of Wales	28

## List of figures

Figure 1. EDM core classes for data providers .....	5
Figure 2. Different types of digital representations that could be captured for a Newspaper. ....	6
Figure 3. The Europeana Cloud data model.....	7
Figure 4. Different levels of a Newspaper with the corresponding IIF resources (as discussed in the IIF Newspapers Interest Group).....	8
Figure 5. A CHO provided by the VU University Amsterdam having two different digital representations: a landing page and a thumbnail. The digital objects are served from the Europeana Cloud.....	9
Figure 6. Proposed model to allow the interoperability between IIF and EDM for simple cases. ....	10
Figure 7. Representation of hierarchical relationships between a Journal Issue and its articles. Note that in this example the URIs don't correspond to URIs to real resources. More details are available in Annex 2. ....	13
Figure 8. EDM model to link a full-text resource to the WebResource it is extracted from.....	17
Figure 9. EDM model based on the Web Annotation Data Model to represent fragment of full-text on a WebResource .....	18
Figure 10. Simpler version of the Figure 9 where the oa:SpecificResource is represented as a WebResource. ....	19

## **Introduction**

The milestone *MS21: Report with recommendations for enhancing EDM*<sup>1</sup> highlighted the fact that the scope of the deliverable D.4.4 might be changed to better accommodate the needs raised by WP4 in the course of the project.

The first round of metadata aggregation from the Europeana Cloud project into Europeana has been done without triggering new requirements in terms of data modelling. However while starting to aggregate content in the Europeana Cloud storage service during the second round of aggregation, new requirements gaps have been identified. This deliverable highlights the requirements for representing digital objects (such as images) and full-text. We propose modelling solutions using the Europeana Data Model (EDM).

In addition as Europeana investigates new ways to serve digital content via the Europeana Cloud API, and through existing platforms for streaming audio and video (e.g. SoundCloud), or via open content protocols, this deliverable will explore how to support interoperability between these services and Europeana. The case studies in this deliverable are the Europeana Cloud<sup>2</sup> content services, and services based on the International Image Interoperability Framework (IIIF)<sup>3</sup>, a service that is currently being developed in Europeana Cloud.

### **1. Methodology and case studies**

WP4 has identified a clear need for guidance on how to interact with the Europeana Cloud services both from a metadata and content perspective. The representation of structured relationships between metadata and content is not a simple issue which required investigations from WP4. It was therefore decided to investigate those issues around real case studies brought by some of the Europeana Cloud project data providers.

The work has been based on case studies provided by the Vrije Universiteit (VU) Amsterdam Library and the National Library of Wales (NLW) with the following objectives:

- to define different use cases for linking metadata and content: from the simple case of adding one link to the content in the metadata to more complex cases.
- to extract requirements and propose solutions using the Europeana Data Model (EDM).
- To propose ways to interact with two solutions to handle content: the Europeana Cloud storage service and IIIF protocol.

---

1

[http://pro.europeana.eu/files/Europeana\\_Professional/Projects/Project\\_list/Europeana\\_Cloud/Milestones/ECLOUD\\_MS21\\_Recommendations\\_EDM.pdf](http://pro.europeana.eu/files/Europeana_Professional/Projects/Project_list/Europeana_Cloud/Milestones/ECLOUD_MS21_Recommendations_EDM.pdf)

<sup>2</sup> <http://pro.europeana.eu/project/europeana-cloud>

<sup>3</sup> <http://iiif.io/>

## **1.1 Case studies 1 and 2: Linking (a) digital object(s) to its corresponding CHO(s) (case provided by the VU Amsterdam Library)**

In this case study the content and the descriptive metadata<sup>4</sup> corresponding to the CHO have been delivered to Europeana separately. While the metadata have been delivered using the traditional aggregation route (in this case The European Library<sup>5</sup> as an aggregator delivered the data to Europeana) the digital objects (or content) have been made available via the Europeana Cloud infrastructure.

Two different situations have been studied:

- The first case study consists in a dataset of CH objects, where each CH object's content exists as a single file. In this case each CHO is linked to one digital representation. For this case, the VU Amsterdam uploaded a sample of content in the Europeana Cloud storage service using the API<sup>6</sup>. The identifiers provided by the Europeana Cloud API were added to the metadata already aggregated by The European Library.
- The second case study is a dataset of CH objects where each object's content is constituted by multiple files, organised hierarchically, and with two different representations of the CH object: one for end-user presentation purposes and one containing the full-text in computer readable form (plain-text or marked-up text as opposed to image format that can't be processed by a machine). The genre of the CH objects studied is "issues of a journal". More details on the case study 2 are provided in Annex 1.

## **1.2 Case study 3: Representing complex digital objects (case provided by the National Library of Wales)**

This case study is based on the Historical Newspapers Collection from The National Library of Wales<sup>7</sup>. Each dataset consists of hierarchically structured CH objects whose content exists in multiple files. These files are themselves hierarchically structured. Note that the hierarchical structure of the digital objects might not fully correspond to the one described at the level of the CHO (e.g. the digital version could be produced from a microfilm having a different sequencing than the original print).

In this case the digital representations available for a CHO are scanned images and full-text content<sup>8</sup>.

In this case the content is made available through the IIIF protocol<sup>9</sup> and requires a mapping between EDM and IIIF to be done to capture the required metadata.

More details on the National Library of Wales case are available in the Annex 2. This case will be also referred to throughout the report.

---

<sup>4</sup> The descriptive metadata in this case were provided in Qualified Dublin Core and then delivered in EDM by The European Library to Europeana.

<sup>5</sup> <http://www.theeuropeanlibrary.org/tel4/>

<sup>6</sup> <https://github.com/europeana/Europeana-Cloud/wiki/Europeana-Cloud-API>

<sup>7</sup> <http://welshnewspapers.llgc.org.uk/>

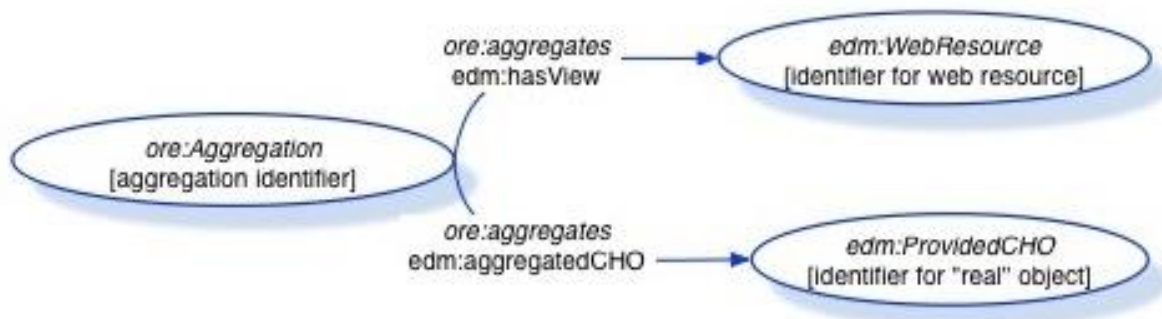
<sup>8</sup> In this particular example the full-text is derived from the digitised images by optical character recognition (OCR).

<sup>9</sup> <http://iiif.io/>

## 2. Representing content in EDM

As explained in the EDM documentation<sup>10</sup>, the model articulates metadata description around three main types of resources:

- The Provided Cultural Heritage Objects or CHOs (*edm:ProvidedCHO*) which describe the original objects—either physical (painting, book, etc.) or born-digital (3D model), which are the focus of description and search in Europeana.
- Web Resources (*edm:WebResource*) are a digital representation of the provided cultural heritage object (CHO), published on the web.
- Aggregation (*ore:Aggregation*) groups the Provided CHO and the Web Resource(s) into one bundle, and information on the aggregation process is also recorded (e.g., the provider of the metadata).



**Figure 1. EDM core classes for data providers**

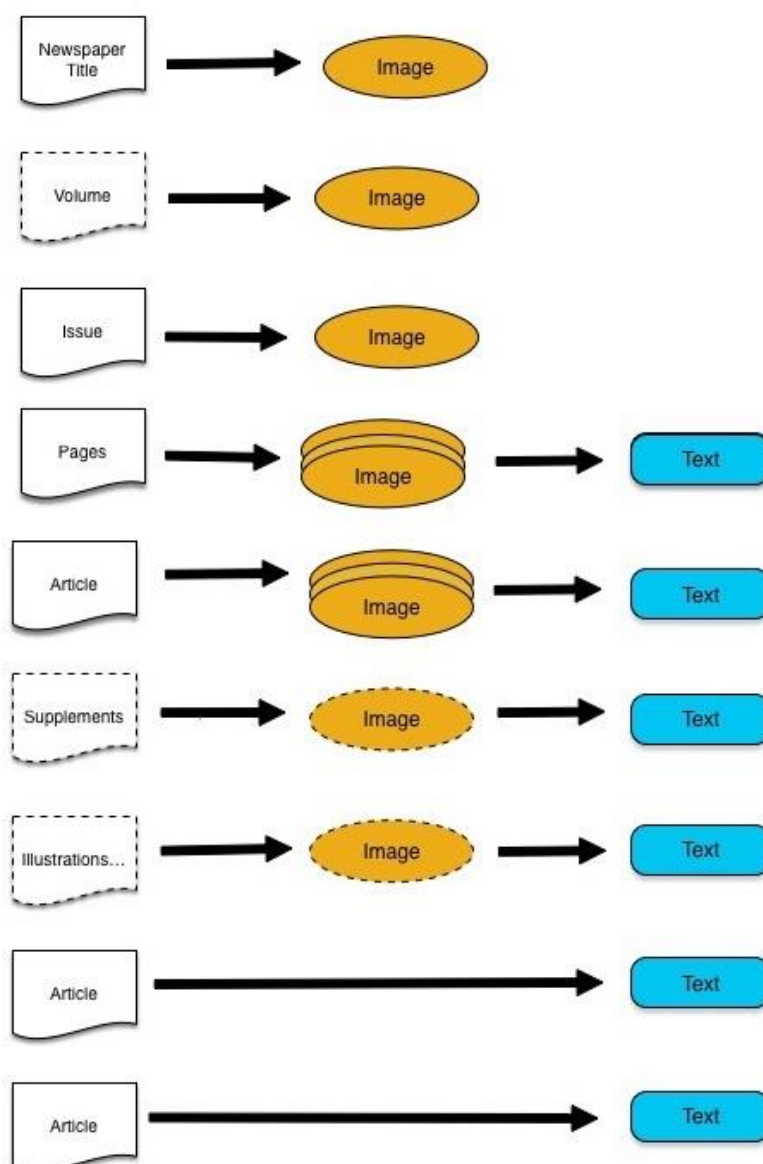
Europeana allows analogue objects, born-digital objects as well as digital objects to be represented as CHO as long as they have the minimal required amount of metadata and a digital representation attached to it. The decision as to what to provide as a CHO belongs to the data provider unless Europeana specifies it otherwise in its content strategy.

The diversity of levels of description applicable to CHOs is also true for WebResources:

- The most common case will be a one-to-one relation between a CHO and a file whether it is a text, an image, a video or audio file.
- Then each type of digital content can have many instances of the same “master” file: an image, a video or an audio file can exist in different formats and in different resolutions.
- The structure between different WebResource representing the same CHO could be complex. For instance hierarchical relationships may link WebResources together and this relation will have to be expressed in the data (i.e. a file and its components).
- Finally the granularity of representation for the digital objects might vary a lot. The degree of digitisation and OCR performed on the provider side will determine the granularity of representation that will need to be retained. For instance digitised Newspapers could have an image representation for the title, issue, page, and article level, a non OCR text representation at the level of the issue (e.g. a PDF file); while the full-text could be at page

<sup>10</sup> <http://pro.europeana.eu/edm-documentation>

level, article, lines or even words level. The following diagram provides examples of the different types of digital representations that could be produced for a Newspaper.



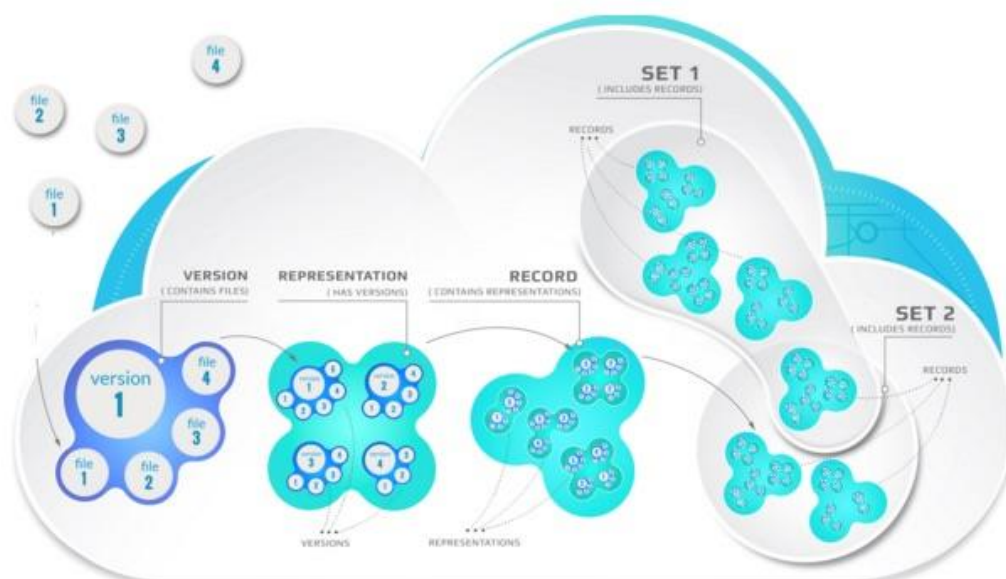
**Figure 2. Different types of digital representations that could be captured for a Newspaper.**

*A Newspaper Title entity represents a series of publications under the same name over time, in this structure represented as Issues. Issues consist generally of a sequence of pages. Within the pages, the articles can be found. The full-text is derived from the digitized images by optical character recognition (OCR). It can be found at different levels as shown here. In the case of the National Library of Wales it can be found at the page level or at the article level.*

The degree of granularity will be determined by the data providers when proceeding to the digitisation and the generation of text (e.g. OCR, corrected OCR, full transcript) of their materials. However Europeana might not need all the levels that can be represented in EDM. Europeana will have to clarify its strategy in terms of content ingestion while data provider will have to decide whether all the levels are worth describing as a CHO (if they have enough metadata and a digital representation) or if they should be described as WebResources.



In the Europeana Cloud storage service, the nature of the uploaded content can be represented using a Representation as defined by the Europeana Cloud data model<sup>11</sup>. It is possible to create as many representations as files.



**Figure 3. The Europeana Cloud data model**

In other services such as IIF the different content types or levels of digital representation are captured in the metadata describing the content.

<sup>11</sup> <https://github.com/europeana/Europeana-Cloud/wiki/Europeana-Cloud-API#data-model>

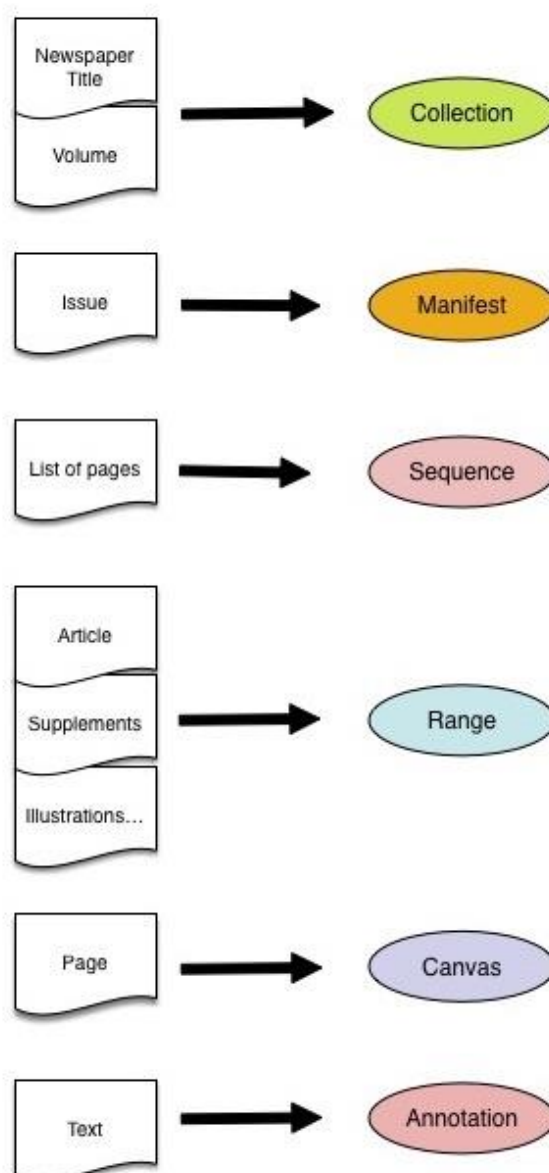


Figure 4. Different levels of a Newspaper with the corresponding IIF resources (as discussed in the IIF Newspapers Interest Group<sup>12</sup>).

### 3. Linking CHO to their digital representations

EDM has a solution for the representation of one or multiple digital representations per CHO via the class *edm:WebResource*. For each digital representation of a CHO an individual WebResource can be created and characterised by a set of descriptive and/or technical metadata.

Until now, Europeana's data providers add links to their digital representations in the metadata they provide to Europeana while the content is stored in the national and local repositories.

<sup>12</sup> The IIF Newspapers Interest Group had its first meeting in December 2015 [http://bit.ly/iif\\_ghent\\_newspaperig](http://bit.ly/iif_ghent_newspaperig)

New open storage solutions such as the Europeana Cloud storage service or IIF offer new challenges to Europeana as it will have to maintain the interoperability between the metadata and its related digital files.

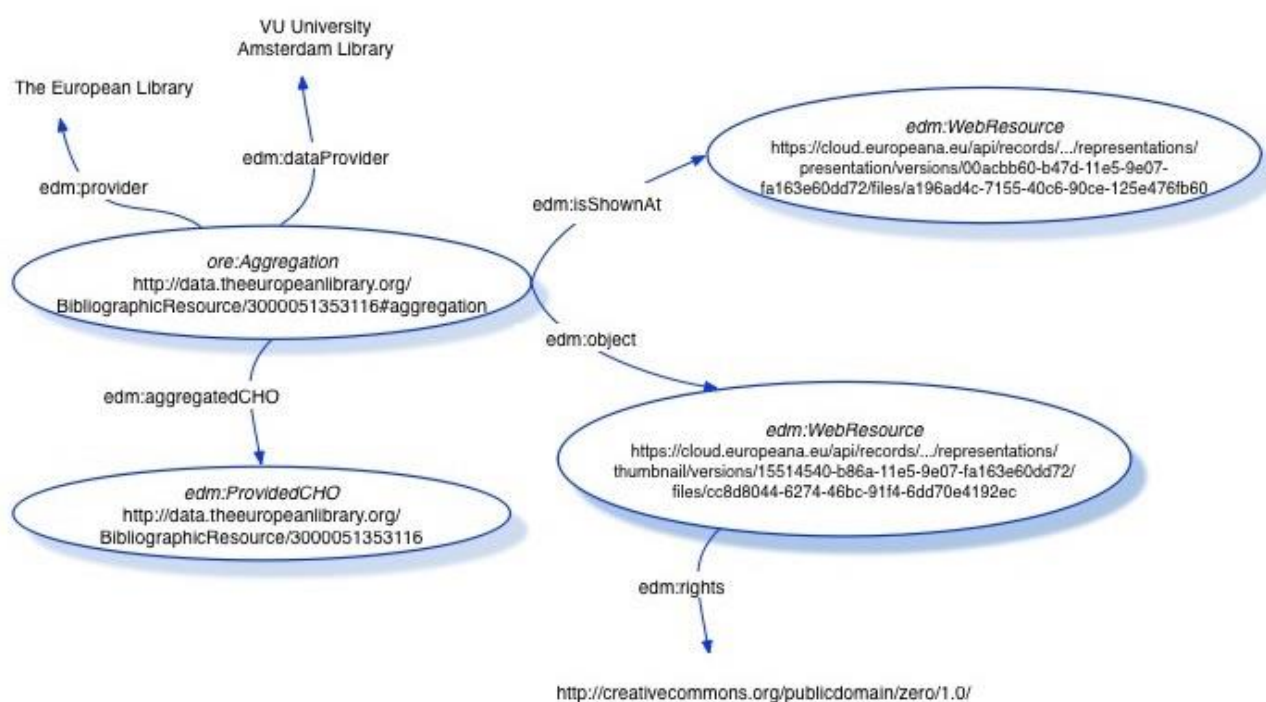


Figure 5. A CHO provided by the VU Amsterdam Library having two different digital representations: a landing page and a thumbnail. The digital objects are served from the Europeana Cloud.

### 3.1 Simple linking between a CHO and its digital representation

Materialising the link between a CHO and its digital representation(s) as it is currently defined in EDM using *edm:isShownBy*, *edm:isShownAt*, *edm:object* or *edm:hasView*<sup>13</sup> won't be sufficient for allowing Europeana to "consume" the digital files.

Each type of media files might be defined against a protocol Europeana will have to support in order to be able to interpret and consume the media files.

The main requirements are therefore the following:

- **R1)** the digital representation need to be referenced by a unique and persistent identifier
- **R2)** the technical specifications and standards required for the export and processing of text, images, audio or video need to be clearly identified in the metadata describing the digital representations (in the *edm:WebResource*).
- **R3)** the interoperability between services using these specifications and Europeana will need to be supported by EDM, which might necessitate mappings or extensions of the model.

<sup>13</sup> Refer to the EDM Mapping Guidelines for more details on the individual properties: <http://pro.europeana.eu/edm-documentation>

**Example with a IIIF resource**

For instance some data providers such as the National Library of Wales have started to use the International Image Interoperability Framework (IIIF) to serve their images to Europeana. IIIF provides a standardised method for describing and sharing images over the web. However Europeana should be able to recognize and ingest IIIF images from EDM metadata.

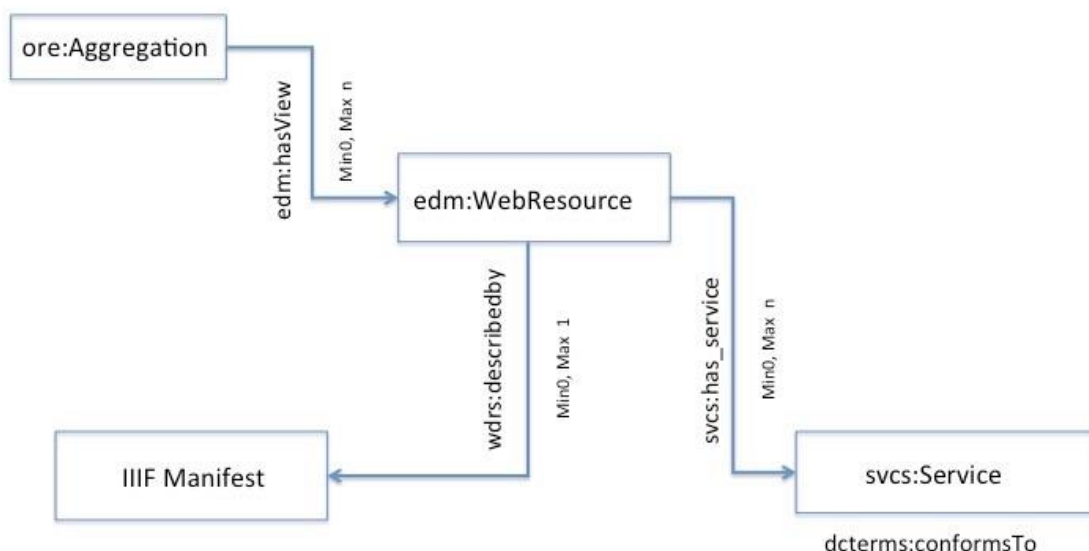
The proposed model is the following:

A CHO will be linked via its Aggregation to a WebResource using the properties *edm:isShownBy*, *edm:object* or the more generic *edm:hasView*. The WebResource should be identified by a URI allowing Europeana to access a media file.

The WebResource can be further linked:

- to a Service describing a Web Service that conformsTo a particular standard.
- to another resource describing the WebResource it relates to (in IIIF, this role is played by 'manifests').

The proposed model can be summarised as following:



**Figure 6. Proposed model to allow the interoperability between IIIF and EDM for simple cases.**

Requirement	Description
<b>R1) the digital representation need to be referenced by a unique and persistent identifier</b>	A <i>edm:WebResource</i> has a unique identifier that allows Europeana to access a media file.
<b>R2) the technical specifications and standards required for the export and processing of text, images, audio or video need to be clearly identified in the metadata describing the digital representations.</b>	A <i>edm:WebResource</i> has a Service ( <i>svcs:has_service</i> ). The Service ( <i>svcs:Service</i> ) describes a WebService that conforms to ( <i>dcterms:conformsTo</i> ) a standardised protocol that will allow the consumption of the media file defined as WebResource. A WebResource can also be linked to

	another resource (with <i>wdrs:describedby</i> ) that brings additional information required to consume the media file properly.
<b><i>R3) the interoperability between services using these specifications and Europeana will need be supported by EDM which might necessitate mappings or extensions of the model.</i></b>	The class <i>svcs:Service</i> as well as the properties <i>svcs:has_service</i> and <i>wdrs:describedby</i> will need to be added to the EDM schema.

The proposed model is generic enough that it can be applied for WebResources served by any clients (for instance SoundCloud<sup>14</sup>).

A WebResource described against the IIIF protocol is described in EDM as follows:

- The WebResource is identified by a URI allowing Europeana to access a IIIF image.
- The IIIF image is related to the Manifest<sup>15</sup> describing it via *wdrs:describedby*. This link allows Europeana to access descriptive metadata if needed.
- The WebResource points to a Service via *svcs:has\_service* which indicates that the resource is a IIIF resource.
- The Service has a *dcterms:conformsTo* indicating the IIIF protocol and making sure the resource is identified as IIIF. Note that *dcterms:conformsTo* can be repeated if more than one link is required to describe the conformance to the protocol (IIIF has 'levels' of conformance). The use of *dcterms:conformsTo* is a proposal made by Europeana to the IIIF community to ensure interoperability between the models. It is still pending a final agreement<sup>16</sup>.

```
<edm:WebResource
rdf:about="http://dams.llgc.org.uk/iiif/image/2.0/1294670/full/512,/0/default.jpg">
  <wdrs:describedby
rdf:resource="http://dams.llgc.org.uk/iiif/2.0/1294670/manifest.json"/>
  <svcs:has_service rdf:about="http://dams.llgc.org.uk/iiif/image/2.0/1294670"/>
</edm:WebResource>

<svcs:Service rdf:about="http://dams.llgc.org.uk/iiif/image/2.0/1294670">
  <dcterms:conformsTo rdf:resource="http://iiif.io/api/image"/>
  and/or
  <dcterms:conformsTo rdf:resource="http://iiif.io/api/image/2/level1.json">
</svcs:Service>
```

### ***Example with Europeana Cloud storage service as storage layer***

A IIIF server will be also made available from the Europeana Cloud services and in this case the modelling from IIIF to EDM will follow the same pattern than above.

<sup>14</sup> <https://soundcloud.com/>

<sup>15</sup> The manifest resource represents a single object and any intellectual work or works embodied within that object <http://iiif.io/api/presentation/2.0/#manifest>

<sup>16</sup> <https://github.com/IIIF/iiif.io/issues/558>

The Europeana Cloud services have also a storage layer where Europeana data providers will be able to store metadata and/or digital representations. The Europeana Cloud API provides a unique identifier to every resource uploaded (fulfilling this way **R1**). The snippet below shows how the digital representations uploaded in the Europeana Cloud storage service are linked to the descriptive metadata. The pattern of the URIs is likely to change as the specifications of the Europeana Cloud API are still being revised.

```
<ore:Aggregation
rdf:about="http://data.theeuropeanlibrary.org/BibliographicResource/3000051353116#aggregation">
  <edm:aggregatedCHO
rdf:resource="http://data.theeuropeanlibrary.org/BibliographicResource/3000051353116"/>
  <edm:dataProvider>VU University Amsterdam Library</edm:dataProvider>
  <edm:isShownAt
rdf:resource="https://cloud.europeana.eu/api/records/YEG2KMUPKB5GNDNOFZBCTWGYDN4KU3ACUYWNWNBPD
QLKCZLMCOQ/representations/presentation/versions/00acbb60-b47d-11e5-9e07-
fa163e60dd72/files/a196ad4c-7155-40c6-90ce-125e476fb602"/>
  <edm:object
rdf:resource="https://cloud.europeana.eu/api/records/YEG2KMUPKB5GNDNOFZBCTWGYDN4KU3ACUYWNWNBPD
QLKCZLMCOQ/representations/thumbnail/versions/15514540-b86a-11e5-9e07-
fa163e60dd72/files/cc8d8044-6274-46bc-91f4-6dd70e4192ec"/>
  <edm:provider xml:lang="en">The European Library</edm:provider>
  <edm:rights rdf:resource="http://creativecommons.org/publicdomain/zero/1.0/" />
</ore:Aggregation>
```

The Europeana Cloud data model (see Figure 3) provides additional information that will allow identifying the format of the Europeana Cloud resource (**R2**). If a digital resource requires a specific protocol to be consumed the same EDM pattern than above can be used. The property *dcterms:conforms* identifies the protocol required for Europeana to “interpret” the media.

### 3.2 Complex linking between a CHO and its digital representations

Lot of work has been done in the past on how to represent hierarchical structures in EDM and we will mainly refer to it in this deliverable.

As explained in the Task Force report Recommendations for the representation of hierarchical objects in Europeana<sup>17</sup>, EDM enables the representation of hierarchies between the digitised representations of a CHO using the EDM class *edm:WebResource*.

So far Europeana data providers have mainly represented hierarchies at the level of the CHO and not represented the hierarchies of the corresponding WebResources.

Digitised objects can be represented in a more granular way in the Europeana Cloud storage service or in the IIIF protocol. Because of digitisation practices and constraints or user requirements (e.g a service like the Historic Newspaper browser<sup>18</sup>) the digital objects might have a more granular structure than the one represented at the level of the CHO. For example, a book may not be digitised in a way that there is an exact match with its physical structure, or a single piece of music may be divided between two audio files for reasons of size. It might be therefore interesting to represent both horizontal and vertical relationships between the different levels constituting digitised objects.

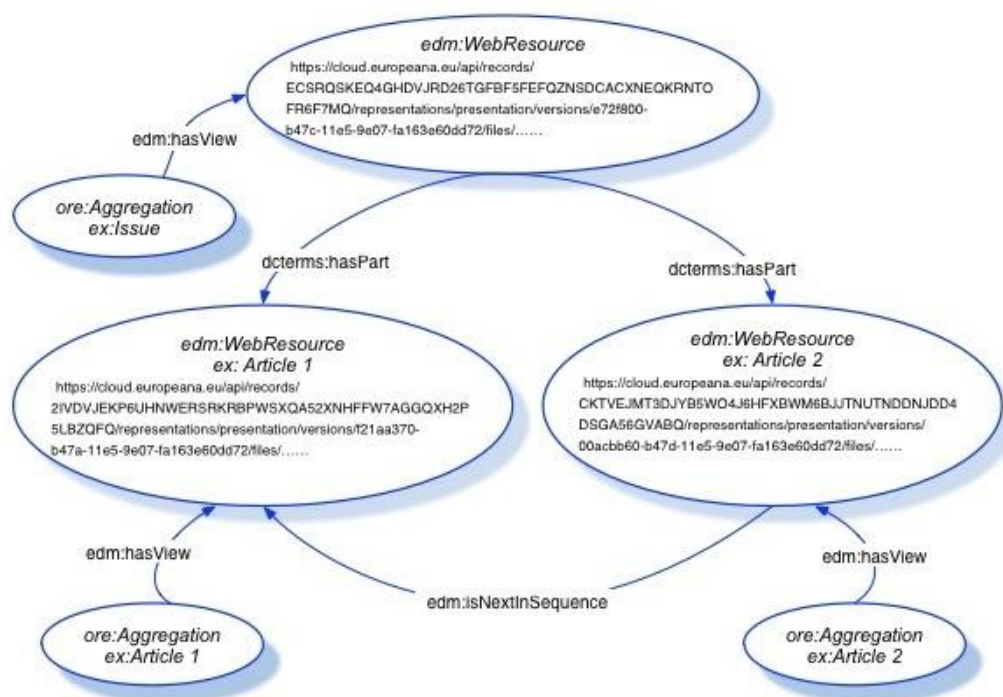
<sup>17</sup> <http://pro.europeana.eu/taskforce/hierarchical-objects>

<sup>18</sup> <http://www.theeuropeanlibrary.org/tel4/newspapers>



In EDM the properties *dcterms:hasPart* and *dcterms:isPartOf* can be used to represent top down or bottom up relationships.

The property *edm:isNextInSequence* 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 (in the Figure 7 below Article 2 is the next in sequence of Article 1).



**Figure 7. Representation of hierarchical relationships between a Journal Issue and its articles. Note that in this example the URIs don't correspond to URIs to real resources. More details are available in Annex 2.**

Representation of hierarchies is especially interesting in the context of digitised Newspapers as it easier the representation of the complex structure illustrated in Figure 2. Different examples can be found in the Task Force report Recommendations for the representation of hierarchical objects in Europeana and the deliverable Report on EDM for Newspapers<sup>19</sup> written in the context of the Europeana Newspapers project.

## 4. Representing Full-Text objects in EDM

### 4.1 Functional requirements for full-text objects

Specific attention is brought in this deliverable on the representation of full-text objects in EDM. The Europeana Cloud services provide a way to handle content in a more granular way and it will help Europeana to better serve content. The first use case regards Newspaper content. The work within

<sup>19</sup> [http://www.europeana-newspapers.eu/wp-content/uploads/2015/05/D4.4\\_EDMforNewspapers\\_V1.0\\_final.pdf](http://www.europeana-newspapers.eu/wp-content/uploads/2015/05/D4.4_EDMforNewspapers_V1.0_final.pdf)

the Europeana Libraries and Europeana Newspapers projects have brought lot of insight of the representation of digitised newspapers and have allowed the definition of initial requirements.

### ***Strategies for representing full-text***

The work around the aggregation of Newspapers in the Europeana Libraries and Europeana Newspapers projects has allowed the identification of three different strategies for digitising Newspapers<sup>20</sup>.

	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, ePub
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, other formats. Text: text/plain, HTML, TEI Metadata: METS
Files combined with coordinates	Similarly to above, but with textual coordinates in the according image.	Images: JPEG, PNG, ... Text: text/plain, HTML, TEI, Web Annotation Model. Metadata: METS/ALTO

These different strategies result in the production of different objects as summarised in the Figure 2 used in the previous section.

### ***User-requirements***

As explained earlier, Europeana might not want to represent all the digitisation levels of a Newspaper as it might not be required by the services using Newspaper content. At the time of writing, no detailed user-requirements specifying how to search, browse and navigate in a dedicated service for digitised newspaper have been defined. The requirements taken into account in this deliverable are those currently defined by the current Newspaper browser in The European Library portal and considering the simplest structure of Newspapers (Title, Issue, Article).

We are aware that more complex situations will need to be handled such as a fragment of full-text spanning across different pages images but we won't tackle them in this deliverable. We will continue to explore the issue while improving our data interoperability with IIIF. We will also closely follow the discussions of the IIIF interest group on Newspapers on these challenges.

<sup>20</sup>

[http://pro.europeana.eu/files/Europeana\\_Professional/Projects/Project\\_list/EuropeanaLibraries/Deliverables/D](http://pro.europeana.eu/files/Europeana_Professional/Projects/Project_list/EuropeanaLibraries/Deliverables/D)



## **Functional requirements for full-text objects**

The requirements for representing full-text are derived from the previous work mentioned already.

### **R1) The availability of full-text must be stated explicitly in the metadata.**

The presence of a full-text resource associated with a digital object should be clearly stated in the metadata describing a CHO. Depending on the strategy chosen by the data provider when digitising the Newspapers, the different files should be identified (corresponding to the title level, page level...).

### **R2) The structure of the Newspapers should be represented in the descriptive metadata**

The digitisation of Newspapers can be done at various level of granularity. The structure of the digitised Newspapers should be identifiable from the metadata describing it.

### **R3) The URLs to the views of the digital objects must be explicitly stated in the data.**

Each digital object or WebResource for a Newspaper object should be accessible via a URL from the metadata. The WebResource's URIs explicitly state how to access the digital object's views. This requirement will enable Europeana to access the digital files referred to in the metadata.

### **R4) The full-text resources must be referenced via direct URLs to one or more files.**

If multiple full-text resources are available for a digital object, these different resources should be accessible via direct links.

### **R5) Resources requiring a specific protocol to be served need to be clearly identifiable**

If a digital object requires a specific Web service to be served and consumed, the protocol it needs to conform to should be indicated in the metadata. It will allow Europeana to interpret and display the digital object.

### **R6) When more than one full-text resource is associated with a digital object, in a hierarchical relationship, it should be possible to represent their part-whole relationship.**

If a digital object has a complex structure made of hierarchical relationships between a parent resource and a child resource, those relationships should be explicitly stated in the metadata. It will allow Europeana to offer a hierarchical view of these resources in its services.

### **R7) When more than one full-text resource is associated with a digital object, it should be possible to represent their sequential order.**

If an object has been digitised as a sequence of digital objects (for instance all the pages of a book), the sequence between the different digital objects should be represented in the metadata. It will allow Europeana to display the sequence in its services.

### **R8) When a full-text resource is available as a fragment of text, the URI or the literal identifying the specific text fragment may be provided in the data.**

A full-text fragment should be identified via a URI.

---

[4.3%20Report%20on%20how%20the%20full-text%20content%20will%20be%20made%20available%20to%20European.pdf](#)

**R9) When a full-text fragment is available, the image area it refers to should be identified (via coordinates).**

Coordinates for the image must be available to identify a portion of text on an image.

## 4.2 Modelling solutions for full-text

### 4.2.1 Representing full-text

As shown in the corpus of digitised newspapers collected as part of the Europeana Newspapers project as well as the case studies collected for this deliverable, full-text can be provided in many different ways within the data.

Putting aside the issues of granularity detailed in section 2, we identified key elements to be addressed in the EDM modelling:

- Full-text is represented as a fragment of text issued from the OCR process. Within the data, this fragment of text can refer to the whole text (i.e., the entirety of an article), or as a fragment (a paragraph, a line or even a word)
- The full-text has an image file for source (or WebResource)
- The full-text can refer to the whole image or a particular area of an image. In the latter situation, coordinates are provided in the data to identify a particular area of the image.

The case study from the National Library of Wales (more details in Annex 2) provides an example on how full-text can be structured in IIIF.

The full-text content is represented as annotations made on the images (or digitised representations of Newspapers issues). The annotations make reference to their respective position within the canvas of the image of the page. The annotations are present in the IIIF manifests of the Articles, as a list of annotations, each one containing a line of text of the article, as follows:

```
...
{
  "@context": "http://iiif.io/api/presentation/2/context.json",
  "@id": "http://dams.llgc.org.uk/iiif/3100023/annotation/list/ART4.json",
  "@type": "sc:AnnotationList",
  "resources": [
    {
      "@type": "oa:Annotation",
      "motivation": "sc:painting",
      "resource": {
        "@type": "cnt:ContentAsText",
        "format": "text/plain",
        "chars": "placed in the consentaneous declarations of the thousands"
      },
      "on": "http://dams.llgc.org.uk/iiif/3100021/canvas/3100023#xywh=303,1201,1054,57"
    },
    {
      "@type": "oa:Annotation",
      "motivation": "sc:painting",
      "resource": {
        "@type": "cnt:ContentAsText",
        "format": "text/plain",
        "chars": "and tens of thousands who have tested its properties m"
      },
      "on": "http://dams.llgc.org.uk/iiif/3100021/canvas/3100023#xywh=304,1248,1054,57"
    },
    ...
  ]
}
```

Representing full-text as an annotation seems the best solution to represent full-text as it allows the support of more complex scenarios such as the positioning of a text fragment on a image or OCR correction scenarios for instance.

The IIF community as well as the Web Annotation community<sup>21</sup> are still refining their specifications. The models proposed in the section below will need to be aligned with the conclusions of these working groups.

## 4.2.2 EDM extension for representing full-text

### *EDM extension addressing the defined requirements*

The extension of EDM for representing full-text follows the recommendations of IIF and is also based on the Web Annotation Data Model<sup>22</sup>. Full-text is represented as a text annotation on a specific image (*edm:WebResource*). The text annotation can refer to a file containing the full-text which is identified as a FullTextResource (*edm:FullTextResource*); or directly to the OCR text which is a literal value.

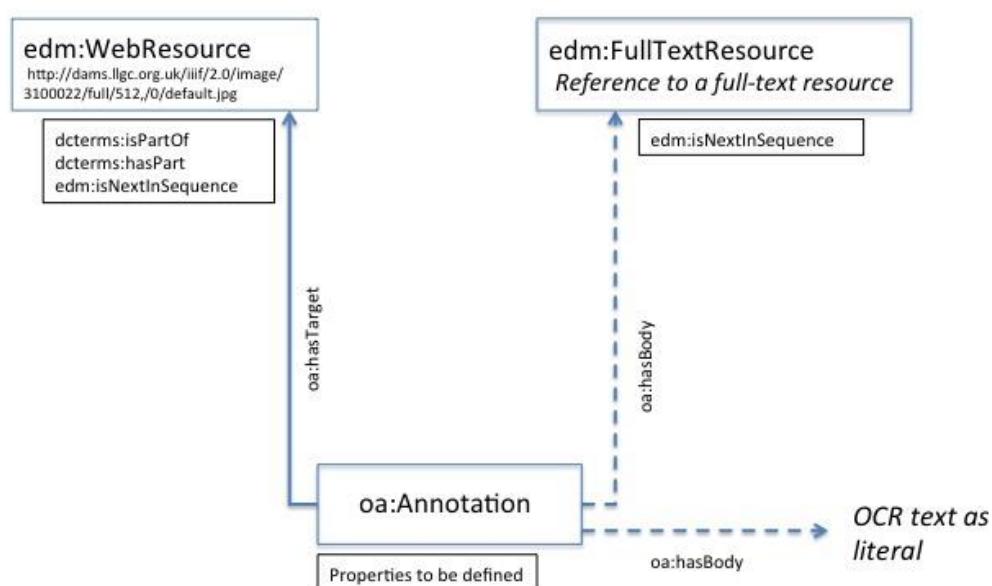


Figure 8. EDM model to link a full-text resource to the WebResource it is extracted from.

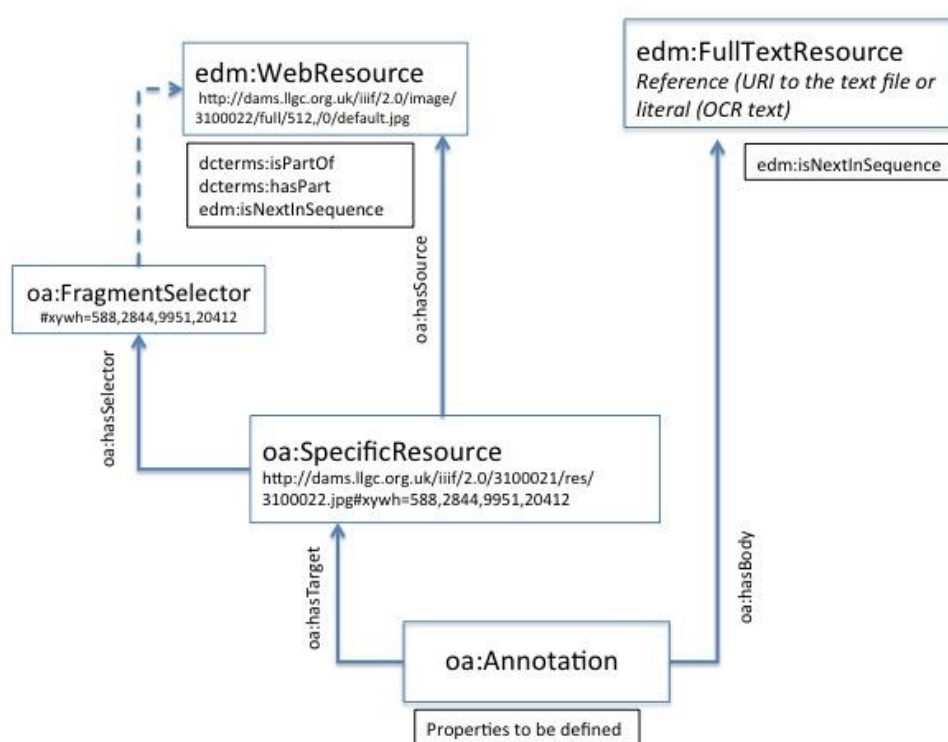
<sup>21</sup> <https://www.w3.org/annotation/>

<sup>22</sup> <https://www.w3.org/TR/annotation-model/>

Note that in the future the *edm:FullTextResource* could be extended to support the representation of the full-text as a literal. A new property would need to be created as part of the class *edm:FullTextResource* to contain such literal.

The National Library of Wales case study shows that the full-text is sometimes represented as several fragments of text, each referring to a specific area of an image (an article, a specific line in the text or a word). In this case, the full-text fragment is accompanied with coordinates indicating its position on the image.

To support this requirement, we introduce in the model a *SpecificResource*<sup>23</sup> (*oa:SpecificResource*). As described in the Web Annotation model the *SpecificResource* “is used in between the Annotation and the Body or Target, as appropriate, to capture the additional description of how it is used in the Annotation”. It is used with a *FragmentSelector*<sup>24</sup> (*oa:FragmentSelector*) which allows the description of the specific section to which the text fragment is applied.



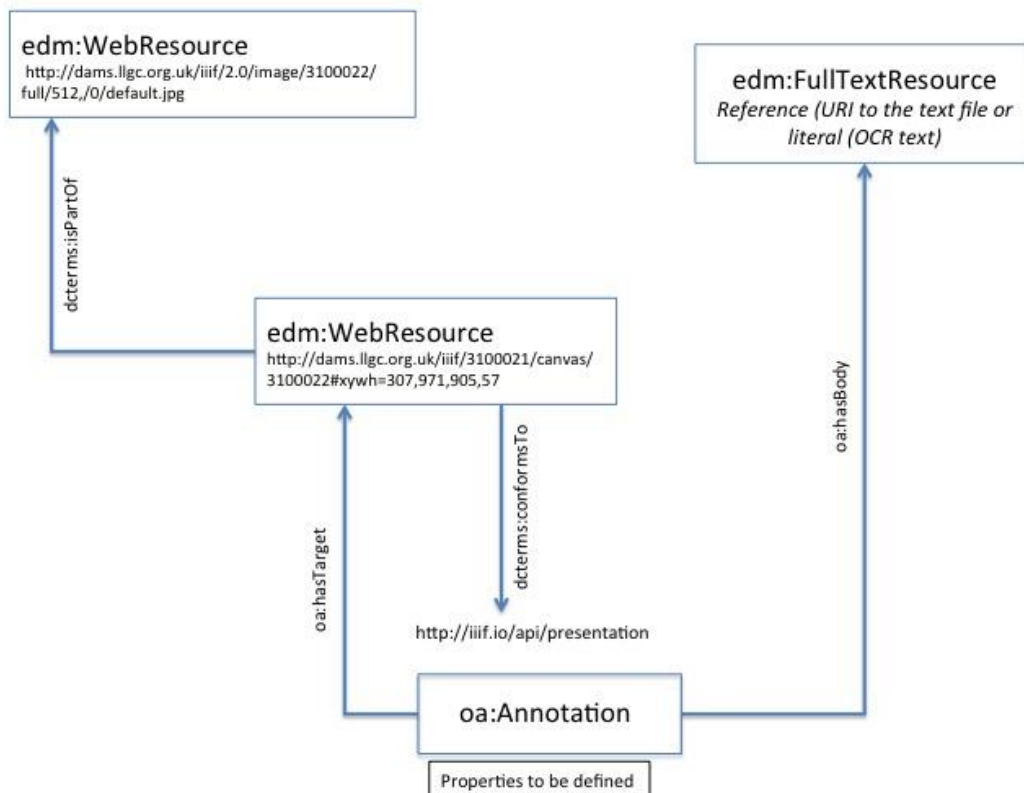
**Figure 9. EDM model based on the Web Annotation Data Model to represent fragment of full-text on a WebResource**

The second model follows the Web Annotation Data Model to the letter and would require the additions of new classes and properties in EDM.

Since the Web Annotation Data Model specifications are still under discussion, we might want to adopt a simpler version of the model proposed above for the time being. Europeana would therefore implement the new required classes or properties from the Web Annotation Data Model in EDM step by step.

<sup>23</sup> <http://www.openannotation.org/spec/core/specific.html#Specific>

In this simpler model, the SpecificResource would be represented as another WebResource pointing to a particular area of an image. The protocol used to serve this URI should be indicated via a *dcterms:conformsTo* property.



**Figure 10. Simpler version of the Figure 9 where the oa:SpecificResource is represented as a WebResource.**

We can summarize our defined requirements and the way they are addressed in the model as following:

Requirements	Description
<b>R1) The availability of full-text must be stated explicitly in the metadata.</b>	Full-text is represented as an Annotation ( <i>oa:Annotation</i> ) having as a target a WebResource. When full-text is available as a resource (URI to a file) it must be represented as a FullTextResource. This requirement is met by the introduction of the <i>edm:FullTextResource</i> . If the text is directly available as OCR text, it will be modelled as a literal as the body of the Annotation. Note that at a later stage the <i>edm:FullTextResource</i> might be used to capture the text literal.

<sup>24</sup> <http://www.openannotation.org/spec/core/specific.html#FragmentSelector>

<p><b>R2) The structure of the Newspapers should be represented in the descriptive metadata</b></p>	<p>It is important that the different levels and therefore the different resources constituting a digitised Newspaper are represented.</p> <p>There are two options to address this requirement:</p> <ul style="list-style-type: none"> <li>-either representing the different levels as a concept using a controlled vocabulary such as the MARC genre list<sup>25</sup>, the ontology BIBO<sup>26</sup> or RDA.</li> <li>-or representing the different levels as resources re-using classes from existing standards (using <code>rdf:type</code>). The IIF community is currently discussing this issue and is looking into the definition of a series of resources types. Our group prefers to wait for further recommendations from the IIF community.</li> </ul> <p>One key aspects as the discussion will be to decide whether these types are aligned with the current resources described by IIF (Manifest, Sequence, Canvas...) or closer to the semantic structure of Newspapers (title, issue, pages...).</p>
<p><b>R3) The URLs to the views of the digital objects must be explicitly stated in the data.</b></p>	<p>The digital objects are represented as <code>WebResource</code> (<code>edm:WebResource</code>) associated with an aggregation (<code>ore:Aggregation</code>) through a <code>hasView</code> relation (including specializations). The <code>WebResource</code>'s URIs explicitly state how to access the digital object's views.</p>
<p><b>R4) The full-text resources must be referenced via direct URLs to one or more files.</b></p>	<p>The existence of multiple <code>FullTextResource</code>'s for a digital object is met by allowing multiple <code>WebResources</code> for a given CHO and multiple <code>FullTextResources</code> per <code>WebResource</code>.</p>
<p><b>R5) Resources requiring a specific protocol to be served need to be clearly identifiable</b></p>	<p>As explained in section 3: A <code>edm:WebResource</code> has a <code>Service</code> (<code>svcs:has_service</code>). The <code>Service</code> (<code>svcs:Service</code>) describes a <code>WebService</code> that conforms to (<code>dcterms:conformsTo</code>) a standardised protocol that will allow the consumption of the media file defined as <code>WebResource</code>.</p>
<p><b>R6) When more than one full-text resource is associated with a digital object, in a hierarchical relationship, it should be possible to represent their part-whole relationship.</b></p>	<p>Hierarchies can be represented at the level of the <code>WebResource</code> (for instance the relation between a full image and the images fragments) or at the level of the <code>FullTextResource</code> using the <code>dcterms:isPartOf</code> and <code>dcterms:hasPart</code> relationships.</p>
<p><b>R7) When more than one full-text resource is associated with a digital object, it should be possible to represent their sequential order.</b></p>	<p>Sequences between <code>WebResources</code> or <code>FullTextResources</code> can be represented using the <code>edm:isNextInSequence</code> property as shown in Figure 7.</p>

<sup>25</sup> <http://www.loc.gov/standards/valuelist/marcgt.html>

<sup>26</sup> <http://biblontology.com/content/complex-series-proceeding-article-use-case-and>  
<http://biblontology.com/content/article>

<p><b>R8 When a full-text resource is available as a fragment of text, the URI or the literal identifying the specific text fragment may be provided in the data.</b></p>	<p>Following the Web Annotation Data Model, a full-text fragment is represented as a SpecificResource (<i>oa:SpecificResource</i>).</p>
<p><b>R9) When a full-text fragment is available, the image section it refers to should be identified (via coordinates).</b></p>	<p>Following the Web Annotation Data Model, a full-text fragment is linked to the image section via a SpecificResource with a FragmentSelector (<i>oa:FragmentSelector</i>) specifying the coordinates of the image section. A simplified modelling would represent a specific image section as a WebResource.</p>

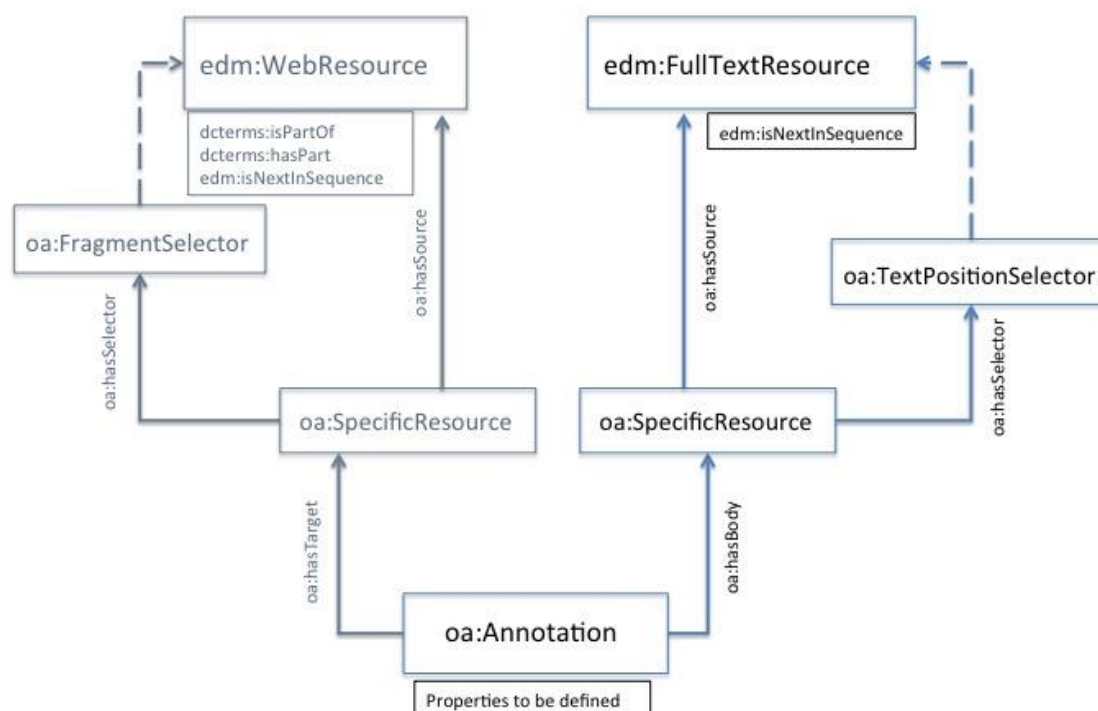
These models will be shared with the IIF and Web Annotation communities for discussions and validation.

### **Extended model to cover new requirements**

Future user-requirements might raise the need to support text annotation on text. For example, an end-user service based on digitised Newspapers may allow users to correct the OCR provided for a given article or provide the full-text for a text object (html or pdf document). In this case, we need to be able to represent a text annotation on a specific area of an image (as covered in the previous model), but also the same text annotation on the text it annotates.

The Web Annotation model offers ways to address this requirement. Similarly to the previous model, a SpecificResource is created to capture additional information about the text annotation. A TextPositionSelector<sup>27</sup> (*oa:TextPositionSelector*) resource is created to describe the annotated fragment of text. The selector records the start and end positions of the range of text.

<sup>27</sup> <https://www.w3.org/TR/annotation-model/#text-position-selector>



**Figure 11. Complex EDM model based on the Web Annotation Data Model to represent an annotation on both a part of the image and a portion of the text.**

As for the other models, this model will be submitted to the Web Annotation community for validation.

## **CONCLUSION**

The case studies analysed for this deliverable provide an overview of the challenges offered by services like the Europeana Cloud services or IIF in terms of data modelling.

In this deliverable we have identified the main interoperability requirements and have tried to address them by re-using the modelling patterns developed by other communities mainly the IIF and the Web Annotation communities. These communities are themselves still working on their specifications, which explain why our modelling proposals are still unstable. Europeana and the National Library of Wales will continue the discussions around modelling of full-text on the IIF and Web Annotation model forums.

The development of the Europeana Cloud services is also ongoing and changes will be made which will require the update of our recommendations. Europeana will continue testing the Europeana Cloud storage service and API by uploading more data and providing recommendations on how to link content and metadata.



## ANNEXES

### ***Annex 1: Case study 2: Complex linking between content and metadata at the VU Amsterdam***

#### Journal issues content from VU Amsterdam Library in Europeana Cloud

Europeana Cloud Project - Case Study 2: Metadata and content structure analysis

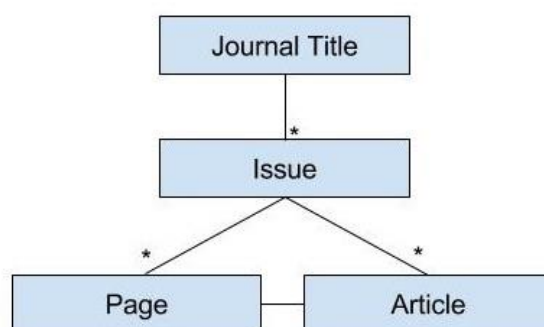
### Introduction

This case study addresses the aggregation of content that a data provider makes available in the Europeana Cloud Infrastructure. In this case, the kind of content studied was a dataset of CH objects where each object's content is constituted by multiple files, organized hierarchically, and with two different representations of the CH object: one for end-user presentation purposes; and a second one containing the full-text in computer readable form. The genre of the CH objects studied is Issues of a Journal.

### Logical structure of the digital objects

The logical structure of the content of a journal issues, as it appears in the studied case, is represented in the figure below. Although a more detailed logical structure of the content of journal issues exists in reality, for the purpose of this study, this level of detail is adequate.

A Journal Title entity represents a series of publications under the same name over time, in this structure represented as Issues. Issues consist of a sequence of Pages that contain Articles. In Journals, there is a strong association between the Articles and the Page number of its parent issue, since it is used for bibliographic references.



The following sections present how this structure is described, and its content represented, in the case of the VU Amsterdam Library.

## Metadata and Content associated with components of the logical structure

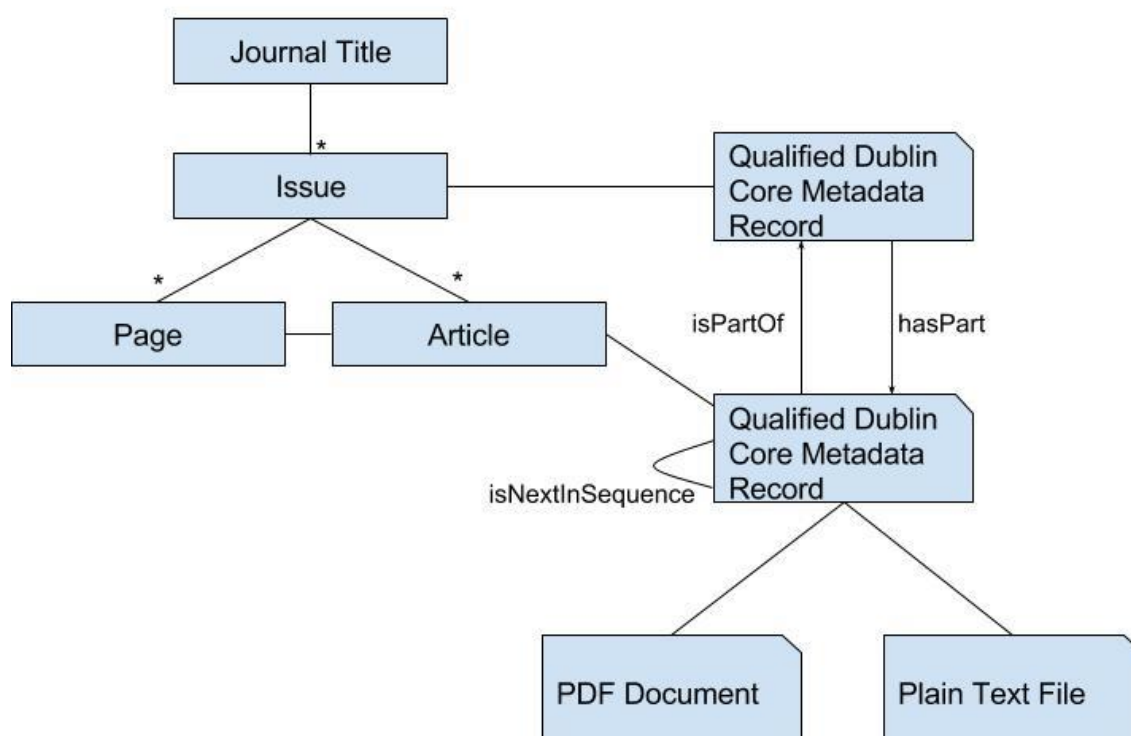
In the case of the VU Amsterdam Library, there is no metadata record available for the whole Journal. Instead, VU Amsterdam Library manages all the issues of the Journal in a single collection named according to the Journal's Title.

Issues and Articles are described individually in their respective metadata records. The VU Amsterdam Library uses, and makes available for aggregation to Europeana, qualified Dublin Core metadata records for these digital objects.

The logical structure of the Journals is shown, in the following figure, along with the metadata records and the content of the digital objects.

The figure depicts also the structural relations between these resources. The metadata records of issues are linked to the metadata records of articles by Dublin Core *hasPart* properties, and in the inverse direction, the metadata records of articles are linked to the metadata records of issues by Dublin Core *isPartOf* properties. The order of the articles within an issue, is represented with EDM *isNextInSequence* properties between metadata records of the articles.

The content of the CH digital objects is referenced from the metadata records of the articles. In this case study, each article has content available in two alternative forms: a PDF Document for end-user presentation; and a plain text file with the full-text.



An excerpt of the descriptive metadata in a Journal Issue in Qualified Dublin Core is shown below.

```
<metadata xmlns:dc="http://purl.org/dc/elements/1.1/"
xmlns:dcterms="http://purl.org/dc/terms/"
  <dc:title>Architectuur en opdrachtgever, Kunstlicht jrg. 05 (1984), nr. 14</dc:title>
  <dc:description>nr. 14</dc:description>
  <dcterms:issued>1984</dcterms:issued>
  <dcterms:extent>30 p.</dcterms:extent>
  <dc:identifiser>
    http://imagebase.ubvu.vu.nl/cdm/ref/collection/klct/id/2135
  </dc:identifiser>
  <dc:identifiser>

https://cloud.europeana.eu/api/records/UMGT5HT4ECMQ3BMJRCANDFNCRXDN2FXLGG23KDRP2JMR3IBKBIHA
  </dc:identifiser>
  ...
</metadata>
```

An excerpt of the descriptive metadata in a Journal Article in Qualified Dublin Core is shown below.

```
<metadata xmlns:dc="http://purl.org/dc/elements/1.1/"
xmlns:dcterms="http://purl.org/dc/terms/" xmlns:edm="http://www.europeana.eu/schemas/edm/"
  <dc:title>Bouwen voor de Bijenkorf. Een commercieel bedrijf als opdrachtgever</dc:title>
  <dc:creator>Groot, A. de</dc:creator>
  <dc:title>Kunstlicht jrg. 05 (1984), nr. 14</dc:title>
  <dc:date>05</dc:date>
  <dc:description>nr. 14</dc:description>
  <dc:identifiser>0408060013002</dc:identifiser>
  <dc:identifiser>http://imagebase.ubvu.vu.nl/cdm/ref/collection/klct/id/2132</dc:identifiser>
  <dcterms:issued>1984</dcterms:issued>
  <dcterms:extent>p. 9-15</dcterms:extent>
  <dc:identifiser>0408060013002</dc:identifiser>
  <dc:identifiser>
    http://imagebase.ubvu.vu.nl/cdm/ref/collection/klct/id/2132
  </dc:identifiser>
  <dc:identifiser>

https://cloud.europeana.eu/api/records/KREYTXSFYYJYQRA6EMSWMQEVAFB6KZ2NND7L6YIR2P2WEW47I5EQ
  </dc:identifiser>
  ...
</metadata>
```

An excerpt of the structural links within metadata in a Journal Issue is shown below.

```
<metadata xmlns:dc="http://purl.org/dc/elements/1.1/"
xmlns:dcterms="http://purl.org/dc/terms/"
  <dc:title>Architectuur en opdrachtgever, Kunstlicht jrg. 05 (1984), nr. 14</dc:title>
  ...

  <dc:identifiser>
https://cloud.europeana.eu/api/records/UMGT5HT4ECMQ3BMJRCANDFNCRXDN2FXLGG23KDRP2JMR3IBKBIHA
  </dc:identifiser>
  <dcterms:hasPart>
https://cloud.europeana.eu/api/records/7ROSNEBAM5VTDYE3W640LLB3ZHVRMBFE6CO43YI5ZFP66QEIMQ
  </dcterms:hasPart>
  <dcterms:hasPart>
https://cloud.europeana.eu/api/records/KREYTXSFYYJYQRA6EMSWMQEVAFB6KZ2NND7L6YIR2P2WEW47I5EQ
  </dcterms:hasPart>
  <dcterms:hasPart>
https://cloud.europeana.eu/api/records/LWW6YPN6QSDXCA465272ZM3HP3M4KWV7Y34VNMHEKVRZWFH3ZLA
  </dcterms:hasPart>
  <dcterms:hasPart>
https://cloud.europeana.eu/api/records/4CC4BWKPVVTPMDOY7JBLQX5CODOV5MHKQ3I2OTXEB6QP7AV2SVQ
  </dcterms:hasPart>
  <dcterms:hasPart>
https://cloud.europeana.eu/api/records/PZPW2SP4KWFTNIR6FC4WQCH5PIYQFXFL3VTF7WNDADQ3RJEDSSPQ
  </dcterms:hasPart>

</metadata>
```

An excerpt of the structural links within metadata in a Journal Article is shown below.

```
<metadata xmlns:dc="http://purl.org/dc/elements/1.1/"
xmlns:dcterms="http://purl.org/dc/terms/" xmlns:edm="http://www.europeana.eu/schemas/edm/">
  <dc:title>Bouwen voor de Bijenkorf. Een commercieel bedrijf als opdrachtgever</dc:title>
  ...

  <dc:identifier>
https://cloud.europeana.eu/api/records/KREYTXSFYYJYQRA6EMSWMQEVAFB6KZ2NND7L6YIR2P2WEW47I5EQ
  </dc:identifier>

  <dcterms:isPartOf>
https://cloud.europeana.eu/api/records/UMGT5HT4ECMQ3BMJRCANDFNCRXDN2FXLGG23KDRP2JMR3IBKBIHA
  </dcterms:isPartOf>

  <edm:isNextInSequence>
https://cloud.europeana.eu/api/records/7ROSNBEWAM5VTDYE3W64OLLB3ZHVRMBFE6CO43YI5ZFP66QEIMQ
  </edm:isNextInSequence>

</metadata>
```

The metadata of the Articles does not contain any link to the content. The content is associated with the Article through the mechanism of the Europeana Cloud data model. An example of the output from the Europeana Cloud API, is shown below. The example contains excerpts of the Europeana Cloud record of an Article, where the metadata is contained (as a representation named *qdc*) as well as the files that contain the content for end-user access (as a representation named *presentation*), and the full-text (as a representation named *fulltext*)

```
<record>
  <cloudId>KREYTXSFYYJYQRA6EMSWMQEVAFB6KZ2NND7L6YIR2P2WEW47I5EQ</cloudId>
  ...
  <representations>
    <representationName>qdc</representationName>
    <files>
      <contentLength>1045</contentLength>
      <contentUri>
        https://195.216.97.95/api/records/KREYTXSFYYJYQRA6EMSWMQEVAFB6KZ2NND7L6YIR2P2WEW47I5
        EQ/representations/qdc/versions/86ef0ad0-b93b-11e5-9e07-fa163e60dd72/files/a2d8b6e6-a19b-
        4987-974e-572ba671332c
      </contentUri>
      <mimeType>application/xml</mimeType>
    ...
  </files>
  ...
</representations>
<representations>
  <representationName>presentation</representationName>
  <files>
    <contentLength>5740838</contentLength>
    <contentUri>
      https://195.216.97.95/api/records/KREYTXSFYYJYQRA6EMSWMQEVAFB6KZ2NND7L6YIR2P2WEW47I5
      EQ/representations/presentation/versions/84b28850-b93b-11e5-9e07-
      fa163e60dd72/files/70d73389-7929-4e49-a8dd-ea12035d304f
    </contentUri>
    <mimeType>application/pdf</mimeType>
    ...
  </files>
  ...
</representations>
<representations>
  <representationName>fulltext</representationName>
  <files>
    <contentLength>30567</contentLength>
    <contentUri>
      https://195.216.97.95/api/records/KREYTXSFYYJYQRA6EMSWMQEVAFB6KZ2NND7L6YIR2P2WEW47I5
      EQ/representations/fulltext/versions/88bfd330-b93b-11e5-9e07-fa163e60dd72/files/e9cfa82c-
      cd6e-47cd-b9f4-2ae10e91f2f5
    </contentUri>
    <mimeType>text/plain</mimeType>
    ...
  </files>
  ...
</representations>
  ...
</record>
```

## ***Annex 2: Case study 3: Newspapers content in the IIIF service of the National Library of Wales***

Newspapers content in the IIIF service of the National Library of Wales

Europeana Cloud Project - Case Study 3: Metadata and content structure analysis

### **Introduction**

This case study addresses the aggregation of content that a data provider makes available through the IIIF protocol. One kind of content will be studied: CH objects whose content exists in multiple files. The files represent several representations of the same object, each representation possibly containing one or more files.

The publication genre analysed in this case study is the digitized newspaper. One particular collection of digitized newspapers was selected from the data providers of the Europeana Cloud project: The Historical Newspapers Collection from the National Library of Wales<sup>28</sup>. The newspapers in this collection are accessible through the IIIF protocol, in a way that makes available metadata about the digital objects, and two different representations of the content: the scanned image and the textual content. The textual content was extracted by the data provider through an Optical Character Recognition (OCR) of the scanned image.

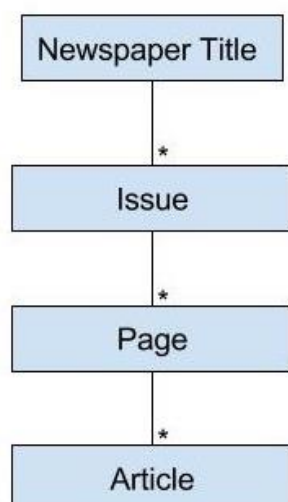
### **Logical structure of the digital objects**

The logical structure of the content of a newspaper publication is represented in the figure below. Although a more detailed logical structure of the content of newspapers exists in reality, for the purpose of this study, this level of detail is adequate.

A Newspaper Title entity represents a series of publications under the same name over time, in this structure represented as Issues. Issues consist of a sequence of pages. Within the pages, the Articles can be found.

---

<sup>28</sup> <http://welshnewspapers.llgc.org.uk/>

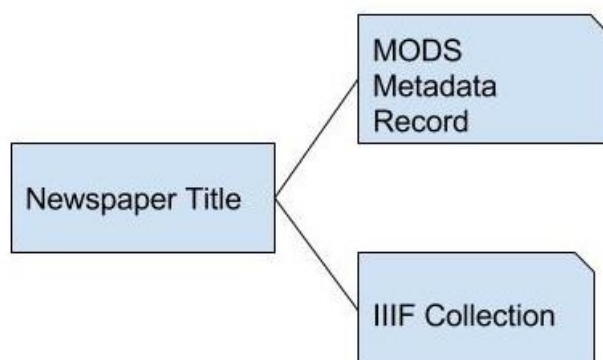


The following sections present how this structure is described, and its content represented, in the case of the National Library of Wales.

## Metadata associated with components of the logical structure

In collections of newspapers, before their digitization, the main entry point for end-users has traditionally been the newspaper title. It is at this level that libraries have been providing finding aids for their users, therefore detailed MARC metadata is often available.

In the case of the newspapers from the National Library of Wales, MODS metadata, a format close to MARC, is available for harvesting through OAI-PMH. In the IIF service, the newspaper titles are described as a IIF Collection resources, which provides descriptive metadata regarding the newspaper title, and structural information regarding the relations between the newspapers titles and the respective issues.



The descriptive metadata existing in a IIF Manifest is represented as labeled fields, therefore it is focused on the presentation of the information for end-users, and not on machine readable semantics. An excerpt of the descriptive metadata in a Newspaper Title IIF Manifest is shown below.

```

{
  "@context": "http://iiif.io/api/presentation/2/context.json",
  "@id": "http://dams.llgc.org.uk/iiif/newspapers/310020.json",
  "@type": "sc:Collection",
  "label": "Potter's electric news",
}
  
```

```

"metadata": [
  {
    "label": [
      { "@value": "Publisher", "@language": "en"},
      { "@value": "Cyhoeddwr", "@language": "cy-GB"}
    ],
    "value": "Joseph Potter"
  },
  {
    "label": "Place of Publication",
    "value": "Haverfordwest"
  },
  {
    "label": [
      { "@value": "Frequency", "@language": "en"},
      { "@value": "Amllder", "@language": "cy-GB"}
    ],
    "value": "Weekly"
  },
  {
    "label": "Language",
    "value": "eng"
  },
  ...

```

The IIIF Collection resource describing the newspaper title also contains structural information regarding the relations between the newspapers titles and the respective issues, which is represented as a list of references to the IIIF Manifests describing the individual issues. These references include also a human readable label for the issue and an associated date. An excerpt of the structural metadata in a Newspaper Title IIIF Manifest is shown below.

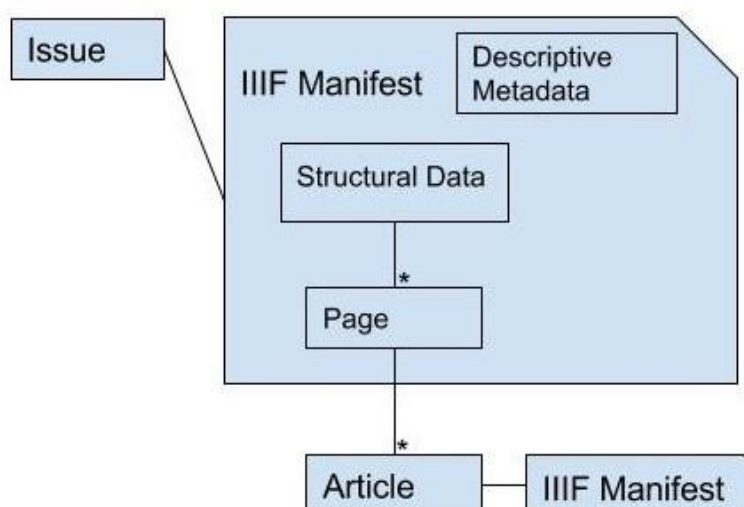
```

...
"manifests": [
  {
    "@id": "http://dams.llgc.org.uk/iiif/newspaper/issue/3100021/manifest.json",
    "@type": "sc:Manifest",
    "navDate": "1859-01-05T00:00:00Z",
    "label": "Potter's electric news (1859-01-05)"
  },
  {
    "@id": "http://dams.llgc.org.uk/iiif/newspaper/issue/3100026/manifest.json",
    "@type": "sc:Manifest",
    "navDate": "1859-01-12T00:00:00Z",
    "label": "Potter's electric news (1859-01-12)"
  },
  {
    "@id": "http://dams.llgc.org.uk/iiif/newspaper/issue/3100031/manifest.json",
    "@type": "sc:Manifest",
    "navDate": "1859-01-19T00:00:00Z",
    "label": "Potter's electric news (1859-01-19)"
  },
  ...

```

At the Issue level, an IIIF Manifest is available that provides descriptive metadata regarding the newspaper issue, and structural information regarding the relations between the issue and its pages.





Similarly to the title manifests, the descriptive metadata existing in the IIF Manifests of issues is represented as labeled fields, having as focus the presentation of the information for end-users. An excerpt of the descriptive metadata in a newspaper Issue IIF Manifest is shown below.

```

{
  "@context": "http://iiif.io/api/presentation/2/context.json",
  "@id": "http://dams.llgc.org.uk/iiif/newspaper/issue/3100021/manifest.json",
  "@type": "sc:Manifest",
  "label": "Potter's electric news (1859-01-05)",
  "navDate": "1859-01-05T00:00:00Z",
  "metadata": [
    {
      "label": [
        { "@value": "Publisher", "@language": "en"},
        { "@value": "Cyhoeddwr", "@language": "cy-GB"}
      ],
      "value": "Joseph Potter"
    },
    {
      "label": "Place of Publication",
      "value": "Haverfordwest"
    },
    {
      "label": [
        { "@value": "Frequency", "@language": "en"},
        { "@value": "Amllder", "@language": "cy-GB"}
      ],
      "value": "Weekly"
    },
    {
      "label": "Language",
      "value": "eng"
    },
    ...
  ]
}

```

The IIF Manifest describing an issue also contains structural information regarding the englobing Newspaper Title, and also the relations to constituting pages of the issue. In this case of the National Library of Wales, the page structure is present within the manifest of the issue and no individual manifest for each page exists. Some further details on the structure of the issue exists, with the description of the Newspaper Articles present within the page. The articles are then described in their particular IIF manifest.

Three excerpts of the structural metadata in an Issue's IIIF Manifest are shown below. The first presents the relation to the Newspaper Title, the second the sequence of Pages, and the third, the articles available within a page.

The following corresponds to the relation with the newspaper title.

```
...
  "within": "http://dams.llgc.org.uk/iiif/newspapers/3100020.json",
  ...
```

The following corresponds to the sequence of Pages:

```
...
  "sequences": [{
    "@type": "sc:Sequence",
    "canvases": [{
      "@id": "http://dams.llgc.org.uk/iiif/2.0/3100021/canvas/3100022.json",
      "@type": "sc:Canvas",
      "label": "[1]",
      "height": 8015,
      "width": 4850,
      "images": [{
        "@id": "http://dams.llgc.org.uk/iiif/2.0/3100021/annotation/3100022.json",
        "@type": "oa:Annotation",
        "motivation": "sc:painting",
        "resource": {
          "@id": "http://dams.llgc.org.uk/iiif/2.0/3100021/res/3100022.jpg",
          "@type": "dctypes:Image",
          "format": "image/jpeg",
          "service": {
            "@context": "http://iiif.io/api/image/2/context.json",
            "@id": "http://dams.llgc.org.uk/iiif/2.0/image/3100022",
            "profile": "http://iiif.io/api/image/2/level1.json",
            "height": 8015,
            "width": 4850,
            "tiles": [{
              "width": 256,
              "scaleFactors": [ 1, 2, 4, 8, 16, 32 ]
            }
          ]
        },
        "height": 8015,
        "width": 4850
      },
      "on": "http://dams.llgc.org.uk/iiif/2.0/3100021/canvas/3100022.json"
    }],
    "otherContent": [
      ...
    ]
  },
  {
    "@id": "http://dams.llgc.org.uk/iiif/2.0/3100021/canvas/3100023.json",
    "@type": "sc:Canvas",
    ...
```

The following IIIF manifest of an Issue presents how the articles within a page are represented in the IIIF Manifest:

```

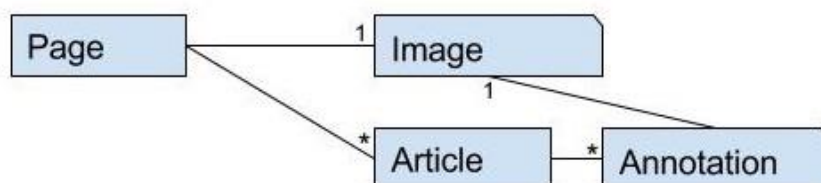
...
"sequences": [{
  "@type": "sc:Sequence",
  "canvases": [{
...
  }],
  "otherContent": [{
    "@id": "http://dams.llgc.org.uk/iiif/3100023/annotation/list/ART4.json",
    "@type": "sc:AnnotationList",
    "label": "[No title]"
  },
  {
    "@id": "http://dams.llgc.org.uk/iiif/3100023/annotation/list/ART5.json",
    "@type": "sc:AnnotationList",
    "label": "HAVERFORDWEST MARKET."
  },
  {
    "@id": "http://dams.llgc.org.uk/iiif/3100023/annotation/list/ART6.json",
    "@type": "sc:AnnotationList",
    "label": "TO OUR CORRESPONDENTS."
  },
  {
    "@id": "http://dams.llgc.org.uk/iiif/3100023/annotation/list/ART7.json",
    "@type": "sc:AnnotationList",
    "label": ", LLANSTADWELL.-CHRISTMAS BOUNTY"
  },
  {
    "@id": "http://dams.llgc.org.uk/iiif/3100023/annotation/list/ART8.json",
    "@type": "sc:AnnotationList",
    "label": "COMMISSIONS BY THE LORD LIEUTENANT."
  }
  ]
},
...

```

## Content associated with components of the logical structure

The content that is available in the IIIF service of the National Library of Wales is associated with the Pages and with the Articles. The content consists of images resulting from a digitization process, and full-text that is derived from the digitized images by optical character recognition.

The associations between the logical structure of Newspapers and the content resources is represented in the following figure:



The pages, are associated with a single Image. Within the IIIF Manifest of the Issues, the images are referenced within the sequence of canvases that compose the complete Issue:

```

...
"@id": "http://dams.llgc.org.uk/iiif/2.0/3100021/canvas/3100022.json",
"@type": "sc:Canvas",
"label": "[1]",
"height": 8015,
"width": 4850,
"images": [{
  "@id": "http://dams.llgc.org.uk/iiif/2.0/3100021/annotation/3100022.json",
  "@type": "oa:Annotation",

```

```

"motivation":"sc:painting",
"resource": {
  "@id":"http://dams.llgc.org.uk/iiif/2.0/3100021/res/3100022.jpg",
  "@type":"dctypes:Image",
  "format":"image/jpeg",
  "service": {
    "@context": "http://iiif.io/api/image/2/context.json",
    "@id":"http://dams.llgc.org.uk/iiif/2.0/image/3100022",
    "profile":"http://iiif.io/api/image/2/level1.json",
    "height":8015,
    "width":4850,
    "tiles" : [{
      "width": 256,
      "scaleFactors": [ 1, 2, 4, 8, 16, 32 ]
    }]
  },
  "height":8015,
  "width":4850
},
"on":"http://dams.llgc.org.uk/iiif/2.0/3100021/canvas/3100022.json"
}],
...

```

For each article, a IIIF Range is created that contains the article metadata and links the image with the article text.<sup>29</sup>

The full-text content is represented as annotations made on the images. The annotations make reference to their respective position within the canvas of the Image of the Page. The annotations are present in the IIIF manifests of the Articles, as a list of annotations, each one containing a line of text of the article, as follows:

```

...
{
  "@context":"http://iiif.io/api/presentation/2/context.json",
  "@id":"http://dams.llgc.org.uk/iiif/3100023/annotation/list/ART4.json",
  "@type":"sc:AnnotationList",
  "resources":[{"
    "@type":"oa:Annotation",
    "motivation":"sc:painting",
    "resource": {
      "@type":"cnt:ContentAsText",
      "format":"text/plain",
      "chars":"placed in the consentaneous declarations of the thousands"
    },
    "on":"http://dams.llgc.org.uk/iiif/3100021/canvas/3100023#xywh=303,1201,1054,57"
  },
  {
    "@type":"oa:Annotation",
    "motivation":"sc:painting",
    "resource": {
      "@type":"cnt:ContentAsText",
      "format":"text/plain",
      "chars":"and tens of thousands who have tested its properties m"
    },
    "on":"http://dams.llgc.org.uk/iiif/3100021/canvas/3100023#xywh=304,1248,1054,57"
  },
  ...
}

```

<sup>29</sup> This level hasn't been detailed to keep the analysis simple.