

# Pushing the CIDOC-Conceptual Reference Model towards Linked Open Data by Open Annotations

Matthias Frank,<sup>1</sup> Stefan Zander<sup>1</sup>

**Abstract:** By using a novel modelling approach, we demonstrate how the Conceptual Reference Model (CRM) of ICOM's International Committee for Documentation (CIDOC) can be complemented using the Open Annotation Data Model (OADM) in order to create semantically rich annotations. We show that domain knowledge can be combined with meaningful and linked data exposed in the so-called Web of Data (aka semantic Web) by having the necessary provenance information for annotations. The combination of domain specific knowledge with existing Linked Open Data (LOD) requires well-designed modelling decisions for linking semantic data sets in a comprehensible way. We show that our combined approach is able to address the requirements of digital heritage in more sufficient ways than each model separately. We combine the advantages of a proven domain ontology with the flexibility and semantic richness of the OADM. In order to evaluate our approach, we show with a concrete example how a museum artifact is modeled in CIDOC-CRM and how these data can be interlinked with existing LOD in meaningful and machine-processable ways by encoding provenance information for new annotations using the OADM.

**Keywords:** Semantic Web, CIDOC-CRM, Open Annotation Data Model, Linked Open Data

## 1 Introduction

The digitization of our cultural heritage (CH), also known as digital heritage (DH), is one of the big challenges museums all over the world are faced with [KK13]. Therefore, the United Nations Educational, Scientific and Cultural Organization (UNESCO) has developed the *Charter on the Preservation of Digital Heritage* [Un04] to provide best practice guidelines for preserving DH. As stated by the UNESCO, DH "is inherently unlimited by time, geography, culture or format", which requires advanced data modeling approaches.

One approach to model DH data is the Conceptual Reference Model (CRM) [CI13] introduced by the ICOM's International Committee for Documentation (CIDOC)<sup>2</sup>. As of today, a growing number of museums and DH projects like the British Museum<sup>3</sup>, the Smithsonian American Art Museum (SAAM)<sup>4</sup> or the Classical Art Research Online Services<sup>5</sup> have started to publish semantically enriched data about their hosted objects using the CIDOC-CRM. However, from a data consuming point of view, modelling DH in this way has a limitation: The CIDOC-CRM does not provide means to encode provenance information

---

<sup>1</sup> FZI Forschungszentrum Informatik am KIT, Information Process Engineering, Haid-und-Neu-Str. 10-14, 76131 Karlsruhe, {frank, zander}@fzi.de

<sup>2</sup> <http://network.icom.museum/cidoc>

<sup>3</sup> <http://www.britishmuseum.org/>

<sup>4</sup> <http://americanart.si.edu/>

<sup>5</sup> <http://www.clarosnet.org>

for the annotations itself. Therefore, when using DH data modeled in CIDOC-CRM in an application, it is impossible to deduce implicit information like the trustworthiness or comparability of annotations. Without this information, data integration in cross-domain projects and the reuse and interpretation of annotations in different contexts is hardly possible.

In order to overcome these limitations of CIDOC-CRM, we introduce a novel modeling approach that extends CIDOC-CRM by combining its well-structured and proven domain-specific taxonomy with the flexibility of rich annotations using the Open Annotation Data Model (OADM) [SCV13].

Representing DH annotations using a combination of CIDOC-CRM and OADM encourage the following:

1. *Interoperability*: As the OADM is a draft of the World Wide Web Consortium (W3C) community, OADM-annotations can be consumed and interpreted correctly by varying applications.
2. *Adaption*: Semantically rich annotations like OADM-annotations allow to adapt the meta data contained within an annotation without cutting the link between body and target of the annotation.
3. *Reuse and interlinkage*: Once a semantically rich annotation is created, it can be reused by adding further bodies or targets to this annotation.
4. *Trustworthiness*: By stating the authorship of an annotation and add explanatory statements to it, annotations become reproducible and the trustworthiness of annotations can be justified.

As a consequence, the main outcome of this work is to provide the logical underpinning upon which CIDOC-CRM can be beneficially extended using the OADM in creating semantically rich annotations that can be shared, extended and utilized by related approaches and also adopted by other domains, e.g. incorporated in the *Educational Web of Data* [Br11].

## 2 Background

In this section we provide an overview of the current modelling approaches for CH and Linked Open Data (LOD). The focus is on CIDOC-CRM, OADM and LOD in general.

### 2.1 CIDOC-CRM

Museums as a stakeholder in the process of digital preservation of artifact descriptions perform archival functions, like building and maintaining reliable collections of well-defined digital objects. They preserve the features like content, fixity, reference, provenance and context which give these objects their integrity. To keep the integrity for digital information objects with long-term cultural value intact is a precondition in order to use them for referring, indexing, citing or any other purpose by the consumers of that data. [LT01, p. 47] For modeling ontologies in the knowledge domain of CH, CIDOC has defined a reference model for storing these digital objects: The CIDOC-Conceptual Reference Model. CIDOC-CRM defines the Terminological Box (TBox) for ontologies in the domain of CH that covers all concepts relevant to describe all types of material collected and displayed by museums and related institutions[BCT07, p. 255]. This terminology is also applied for Lightweight Information Describing Objects (LIDO), an Extensible Markup Language (XML) schema for CH provided by International Council of Museums (ICOM)[IC10]. An example for modeling DH data in CIDOC-CRM is shown in Figure 1. The marked nodes in Figure 1 represent the classes defined in CIDOC-CRM.

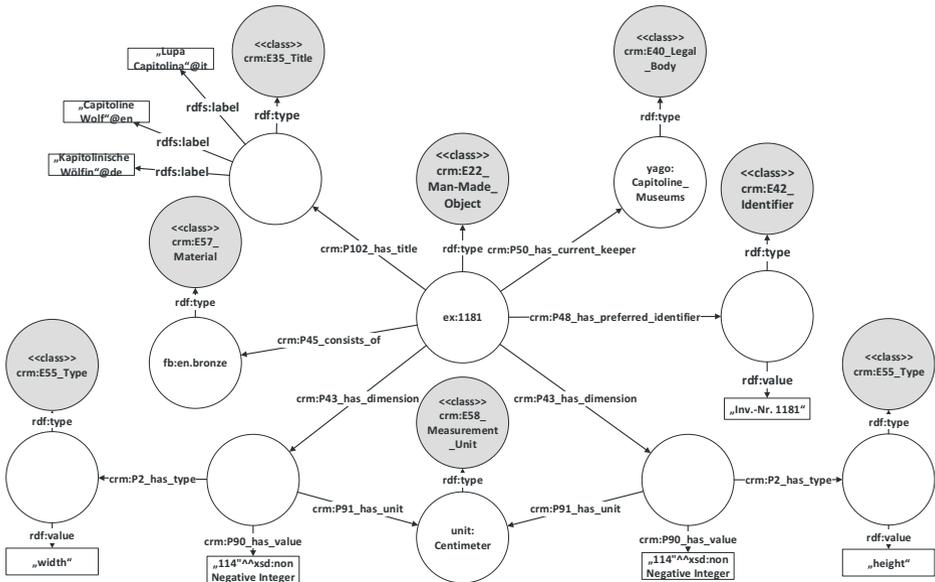


Fig. 1: Example for modelling in CIDOC-CRM

## 2.2 Open Annotation Data Model

Annotations created with Web Ontology Language (OWL)/Resource Description Framework (RDF) have a limited expressiveness by default. Basically, these annotations state that two resources are related to each other in a specific way. This relationship may be a predicate like `rdfs:seeAlso`, `rdfs:isDefinedBy` or any other predicate that expresses a relation of these resources. Within the scope of this work, the subject of the triple expressing the annotation is named *body* of the annotation, whereas the object of the triple is named *target* of the annotation. Such a basic annotation is depicted in Figure 2.



Fig. 2: Basic RDF Annotation

As there has been no uniform approach for creating annotations, in 2012 the W3C community introduced the OADM. This approach introduces a methodology for annotations that conforms to the architecture of the World Wide Web (WWW)[CSV12]. The OADM consists of a core which provides the basic functionality to create open annotations. The OADM core can be extended by several modules if necessary for a specific application. Rather than implementing an annotation as a simple triple pointing from a body resource to a related target, OADM creates a distinct resource for the annotation itself which then points to the body and the target of the annotation and also provides useful metadata. The idea of OADM is to reuse existing vocabulary wherever possible, for example the vocabulary defined in Friend of a Friend (FOAF) or Dublin Core (DC). The reuse of LOD-resources is therefore a major contribution of our approach. However, there are also classes and properties defined for the open annotation namespace in <http://www.w3.org/ns/oa#>, usually abbreviated with the prefix `oa:`. A depiction of the basic annotation model is shown in Figure 3, stating the same fact as in Figure 2.

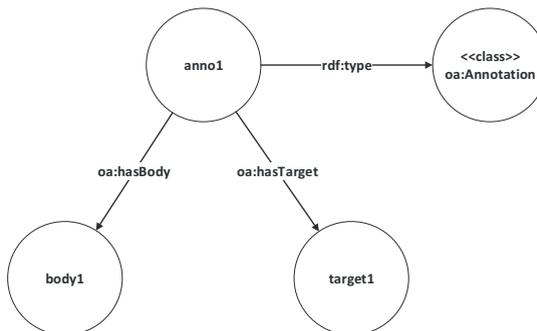


Fig. 3: Open Annotation Data Model

Metadata that can be included by OADM covers a person or organization that created the annotation, the motivation for the annotation and also the time when the annotation was created. This approach allows to express the authorship of an annotation and separate between annotations created by the museum's staff and annotations created by the community. In addition, OADM provides the possibility to state the software that was used for serialization and the time when the annotation was serialized, both can be taken into consideration when justifying the provenance and trustworthiness of an annotation. Especially the time of serialization is important for a proper version control [SCV13] when maintaining the annotations. With OADM, conservators of CH have the option to use linked data to augment user experience rather than only publishing their own linked open datasets. The OADM is therefore ideally suited for the creation of knowledge structures through semantic annotations in the field of CH.

### 2.3 Linked Open Data

The idea of linked data was introduced by Tim Berners-Lee in 2006 [BL06]. He defined expectations which apply to both, the conventional Web of Hypertext Markup Language (HTML) documents and the Web of linked data represented using the RDF. He also added a 5-Star scheme for LOD in 2010.

An early use of publishing linked data is the creation and publishing of personal profiles as some kind of business cards using the FOAF vocabulary. These FOAF-profiles have made an essential part of linked data in the early beginning of the semantic Web. However, the number of datasets in linked data, the number of triples within these datasets and also the RDF links interlinking these datasets have increased rapidly during the last years. Richard Cyganiak and Anja Jentzsch started an approach to visualize datasets of linked data available on <http://datahub.io/dataset?tags=lod> as a LOD cloud diagram in 2007. The first version of this diagram contained 12 datasets. This project was maintained over the years, in 2014 the LOD cloud diagram contained 570 datasets. 374 of them were described by the data providers themselves in the datahub.io dataset catalog, 196 more were discovered by a crawl of the Linked Data web conducted in April 2014 [CJ14]. A depiction of the resulting LOD cloud diagram is shown in Figure 4.

As can be seen from this diagram, FOAF-profiles, Geo-Names and especially DBpedia are highly interconnected datasets containing a huge amount of triples. For example, a SPARQL Protocol and RDF Query Language (SPARQL) query counting the triples contained in DBpedia in January 2015 returns a number of more than two billion triples.

## 3 Related Works

For the creation of annotations for museum objects in this work, we annotate the description of a museum artifact modeled in CIDOC-CRM with data from the LOD cloud. According to a research funded by SAAM, mapping the data of a museum to linked data involves three steps [Sz13, p. 1–2]:

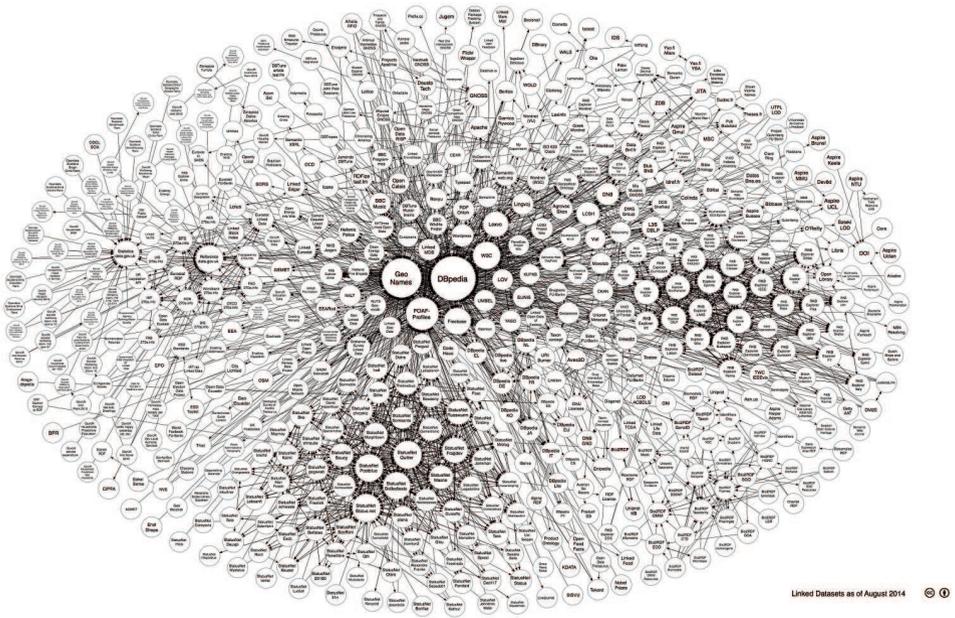


Fig. 4: Linked Open Data Cloud as of 2014

1. Map the Data to RDF.

The first step is to map the metadata about works of art into RDF. This involves selecting or writing a domain ontology with standard terminology for works of art and converting the data to RDF according to this ontology. De Boer et al.<sup>6</sup> note that the process is complicated because many museums have rich, hierarchical or graph-structured data. The data often includes attributes that are unique to a particular museum, and the data is often inconsistent and noisy because it has been maintained over a long period of time by many individuals. In past work, the mapping is typically defined using manually written rules or programs.

2. Link to External Sources

Once the data is in RDF, the next step is to find the links from the metadata to other repositories, such as DBpedia or GeoNames. In previous work, this has been done by defining a set of rules for performing the mapping. Because the problem is difficult, the number of links in past work is actually quite small as a percentage of the total set of objects that have been published.

<sup>6</sup> Boer, V., Wielemaker, J., Gent, J., Hildebrand, M., Isaac, A., Ossenbruggen, J., Schreiber, G.: Supporting Linked Data Production for Cultural Heritage Institutes: The Amsterdam Museum Case Study. In: Simperl, E., Cimiano, P., Polleres, A., Corcho, O., Presutti, V. (eds.) *Lecture Notes in Computer Science*, pp. 733-747. Springer Berlin Heidelberg (2012), cited by [Sz13]

### 3. Curate the Linked Data

The third step is to curate the data to ensure that both the published information and its links to other sources within the LOD are accurate. Because curation is so labor intensive, this step has been largely ignored in previous work and as a result links are often inaccurate.

We do also consider these three steps for our modeling approach of semantic annotations in the field of CH. As stated in the outcome of the SAAM project, for the first step there are already some successful approaches like the Europeana project<sup>7</sup>, the Amsterdam Museum<sup>8</sup>, the LODAC museum<sup>9</sup> or the KARMA approach<sup>10</sup>. Therefore, our work focuses on the linking of museums' RDF-data to external resources and also in maintaining these links. For structuring metadata of CH objects we extend the domain specific model with the OADM. This idea was also suggested for future work for structuring metadata of CH objects during the International Conference in DC and Metadata Applications in 2014 [Wi14].

For this work, we assume that data of a museum is modeled in CIDOC-CRM in order to describe museum objects. The goal is to ensure that on the one hand only the museum itself is able to publish authoritative data, but on the other hand the community is able to create annotations to this data in order to augment the knowledge structure of the museum's repository.

## 4 Approach

With our approach we show that the combination of the domain specific modeling of CIDOC-CRM and the OADM leads to an comprehensive model that covers the requirements of CH specialists and enables the modeling of provenance information for new annotations. This novel approach helps museums to use LOD in order to augment their visitors experience and also publish their data as LOD in a meaningful way by providing provenance data of annotations, which allows a collaborative annotation of museum objects.

### 4.1 Requirements

In order to contribute to the process of digital preservation of CH, we pose the following requirements:

- R1 The data published by a museum about their artifacts has to be modeled in a way that consumers can distinguish them from annotations created by the user community.

---

<sup>7</sup> <http://data.europeana.eu>

<sup>8</sup> <http://www.amsterdammuseum.nl/open-data>

<sup>9</sup> <http://www.ontotext.com/customers/lodac-museum-linked-open-data-academia/>

<sup>10</sup> <http://www.isi.edu/integration/karma/>

- R2 The user-created annotations have to be serialized and published in order to augment user experience when consuming DH-objects.
- R3 Methods that enables users to filter annotations by type, creator, annotated object, annotation time and version has to be provided.

All three requirements are on the approach in general, rather than for a specific prototype. The requirements are the result of our practical research work with partners from the CH-domain. To fulfill requirement R1, museums' data is published in CIDOC-CRM in read-only while annotations by users are added using OADM in order to clearly state the provenance of the created annotations. As an example, when the conservator of a museum has gathered the required information and added them to the system, any user can search for resources with similar properties on the LOD cloud and create personal annotations. Due to the use of the OADM, not only the semantic annotation itself will be created, but also some useful meta data. This meta data may include information about who created this annotation, what was his or her motivation doing that, when he or she actually created the annotation and also which software was used for the serialization and when was the serialization of the annotation performed.

## 4.2 Data Structure for Museum Objects

The description of museum objects is assumed to be modeled in CIDOC-CRM for this work. However, as the full CRM aims to define all classes and properties needed to describe knowledge in the domain of CH and not only museum objects, for this work only instances of the class *E22 Man-Made Object* defined by CIDOC in the CRM are considered. This comprises physical objects purposely created by human activity [CI13, p. 11] as artifacts of CH. This Section describes the properties which are applicable to the class *E22 Man-Made Object* and introduces the exemplary modeling of the description of a museum object as an instance of this class.

Figure 5 shows the Unified Modeling Language (UML) class diagram of the TBox for the class *E22 Man-Made Object*. The members of the classes are the respective properties of the classes, consisting of predicate (identifier P) and object (identifier E). In RDF, properties are not added to a class like in an object orientated modeling approach. In fact, the properties shown in Figure 5 result from the domain and range defined for each predicate. Although the class *E22 Man-Made Object* does not have any specified properties, it inherits all the properties from its superclasses.

## 4.3 Discover Resources in Linked Data for new Annotations

The core of our approach is to support the conservator of a museum in creating rich annotations for digitized artifacts modeled in CIDOC-CRM with resources available as LOD. This is done by providing suggestions for annotations and enable volunteers to review these suggestions.

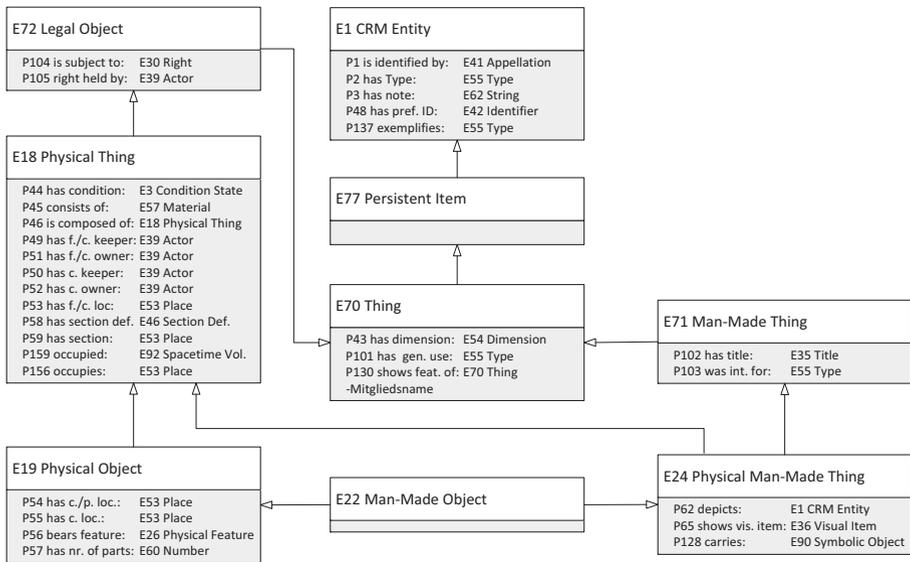


Fig. 5: UML class diagram for E22 Man-Made Object with superclasses

In order to find suitable resources for the semantic annotation of digitized artifacts, all resources which are related to the subject and are available as LOD should be discovered. Relatedness of resources in the semantic Web can be measured in different ways, for example the number of edges between two nodes. This approach does not consider the quality of the edges, therefore each predicate of an RDF-graph is treated similar. There are also approaches that weight the quality of edges, which means that a even a path with more edges could express a higher relatedness between two nodes if the quality of the edges is better, or in sense of RDF the predicates are semantically better suited to express a relation. In the context of our work, relatedness is measured in the number and quality of property matches, as a high accordance of properties does also indicate a high relatedness. Museum objects are objects of public interest, therefore it is likely that someone already published something with similar properties as LOD.

However, as the semantic Web is not a central database where properties are always defined in the same way, there may be relevant resources which are described with different properties which have the same meaning. To find these resources, the predicates have to be mapped to cover these different descriptions. An abstracted depiction of this mapping is shown in Figure 6.

Rather than just search for resources with similar properties of the subject in the example data introduced in Section 5.1, like for example the property `crm:P45_consists_of` `fb:en.bronze`, both, the predicate and the object have to be mapped to lists of equivalent predicates and objects. When the input data is enriched by lists of equivalent predicates and objects, resources with similar properties published as LOD can be queried. This is

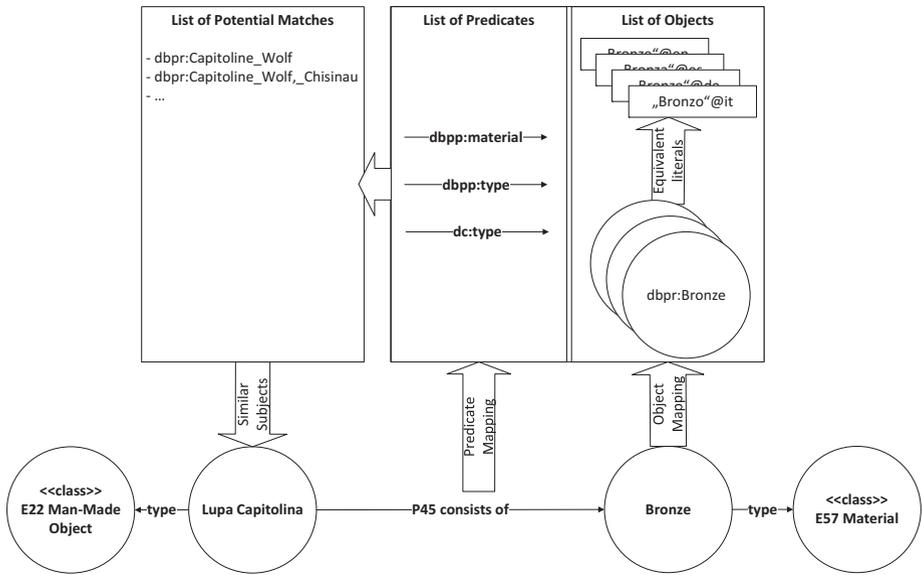


Fig. 6: Abstracted Mapping Process

done with the help of SPARQL using public SPARQL endpoints. The result of this query is then returned to the client for further processing.

The resulting graph of this new annotation is shown in Figure 7. However, in this graph, target and body of the annotation are not shown as actual resources. The abstracted body shows that the Internationalized Resource Identifier (IRI) resources used for the body of the annotation come from within the LOD cloud, whereas the abstracted target shows that the IRI-resource used for the target of the annotation is part of the museum’s repository modeled in CIDOC-CRM. By using OADM for user created annotations, R2 and R3 are fulfilled.

The OADM annotations created by our approach combine the digital description of CH artifacts modeled with CIDOC-CRM with information available as LOD. The result is therefore a knowledge structure that contains information of CH from both sources, including metadata about the annotations themselves. Figure 8 shows an example of an artifact modeled in CIDOC-CRM, the marked nodes represent an additional annotation about this artifact modeled in OADM.

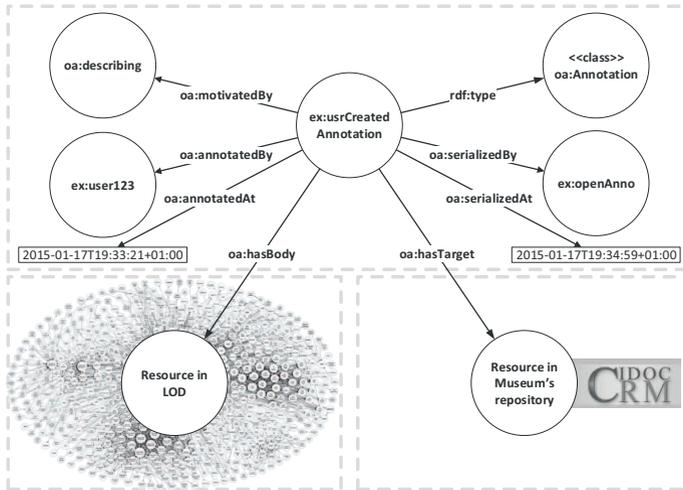


Fig. 7: RDF-Graph of OADM-Annotation

## 5 Use Case – Evaluation

### 5.1 Example data: Lupa Capitolina

Our approach presumes valid CIDOC-CRM ontology data. Therefore, in addition to the schema description of E22 Man-Made Object (TBox) given in Section 4.2, some instance data (Assertional Box (ABox)) has to be added. In our example, we assume that the Capitoline Museums in Rome, Italy, wants to publish a semantic descriptions of their hosted artifacts using CIDOC-CRM. First, the museum defines `http://museum.example.com/objects/` as the namespace for all museum objects. This namespace will be abbreviated with the prefix `museum:`. The museum artifact that is encoded in this example is the bronze sculpture “Capitoline Wolf” (Italian: Lupa Capitolina). A description of this sculpture is provided by the university of cologne<sup>11</sup>. The preferred local name for an artifact within the museums namespace is the inventory number of the corresponding object [CI11]. In case of the Capitoline Wolf, the Capitoline Museums assigned the inventory number 1181. The resulting IRI for the new object is therefore `http://museum.example.com/objects/1181`, abbreviated as `museum:1181`. This abbreviation is not a valid qualified name (QName) as the local name starts with a number, however, it is a valid Compact URI expression (CURIE).

As CIDOC-CRM does not foresee the description of instances of E58 Measurement Unit [CI13, p. 23], instances provided by the Quantities, Units, Dimensions and Data Types Ontologies (QUDT) are used. The namespace for this ontology is `http://qudt.org/vocab/unit#` (abbreviated with `unit:`). QUDT is developed by TopQuadrant and

<sup>11</sup> <http://arachne.uni-koeln.de/item/objekt/16611>

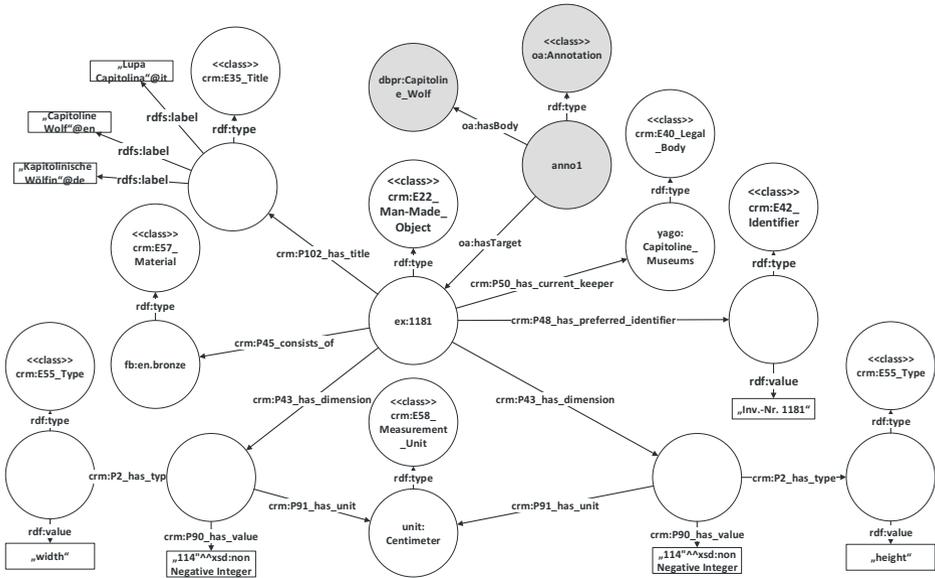


Fig. 8: CIDOC-CRM extended by OADM

the National Aeronautics and Space Administration (NASA) in order to provide interoperability between information systems. It is published under Creative Commons (CC) Attribution-Share Alike 3.0 United States License and can therefore be freely used, as long as the name of the creator is provided. For instances of other classes of CIDOC-CRM, for example E39 Actor or E53 Place, existing upper ontologies are used where applicable. In order to improve the reuse of resources of the semantic Web, we use upper ontologies which are published with an open license. In this example, we use the upper ontologies yago<sup>12</sup> (abbreviated with yago:) and freebase<sup>13</sup> (abbreviated with fb:).

## 5.2 Related Resources in LOD for new Annotations of Lupa Capitolina

The discovery of resources in LOD that are related to a fact which is described with a blank node (bnode) in the museum’s repository requires additional attention. Although it is possible to query bnodes due to the graph-oriented semantics of SPARQL, many subjects in LOD are annotated in a more simple way. As an example, the height of the artifact used here and modeled in CIDOC-CRM is stated with an dimension-bnode having the local Identifier (ID) \_:lh. This dimension has the type “height” and a value of 75 cm. The same fact is expressed in DBpedia<sup>14</sup>, a database that contains structured data from Wikipedia, with the properties dbpprop:heightMetric "75" and dbpprop:metricUnit "cm".

<sup>12</sup> <http://yago-knowledge.org/resource/>

<sup>13</sup> <http://rdf.freebase.com/ns/>

<sup>14</sup> [http://dbpedia.org/resource/Capitoline\\_Wolf](http://dbpedia.org/resource/Capitoline_Wolf)

Therefore, all properties modeled as a bnode have to be interpreted by the properties of the respective bnode. A first-degree (one edge distance from the subject) bnode-property may have bnode-properties as well. The same goes for the second-degree (distance of two edges), third-degree (distance of three edges) and so on, therefore this recursive procedure has to be limited to a particular distance level in order to answer queries in a reasonable time. A limit of 3 runs in example does ensure to get all properties up to a distance of three edges.

Our approach is implemented as a prototype which searches LOD for related resources to annotate a given CH artifact. When the search has finished, all results are listed, aggregated by instances and ordered by the number of matches as can be seen in Figure 9. The resource found with the most according properties is listed first. All results that fulfill the predefined requirements are preselected. For the prototypical implementation used for our work, the default parameters are  $P_1 = 2$  for the minimum number of congruent properties and  $P_{1=0.2}$  for the minimum rate of congruent properties in relation to the total number of properties. Therefore, all resources which have at least two properties that matches to properties of the target resource and the number of matching properties in relation to the total number of properties of the target resource equals to at least 20% are preselected for annotation. In this example, three resources are suggested to be annotated with `ind:Lupa_Capitolina` as can also be seen in Figure 9.

Title	Name Space	Local Name	Accordances	Select
Capitoline Wolf	http://dbpedia.org/resource/	Capitoline_Wolf	7 of 8 (87%, see tooltip)	<input checked="" type="checkbox"/>
Seated Yucatan Woman	http://dbpedia.org/resource/	Seated_Yucatan_Woman	2 of 8 (25%, see tooltip)	<input checked="" type="checkbox"/>
Capitoline Wolf, Chişinău	http://dbpedia.org/resource/	Capitoline_Wolf_museum	2 of 8 (25%, see tooltip)	<input checked="" type="checkbox"/>
Colossus of Constantine	http://dbpedia.org/resource/	Capitoline_Wolf lengthMetric "114"	1 of 8 (12%, see tooltip)	<input type="checkbox"/>
Romulus and Remus (Ru...	http://dbpedia.org/resource/	Capitoline_Wolf heightMetric "75"	1 of 8 (12%, see tooltip)	<input type="checkbox"/>
John the Baptist (Caravagg...	http://dbpedia.org/resource/	Capitoline_Wolf type dbpr:Bronze	1 of 8 (12%, see tooltip)	<input type="checkbox"/>
The Fortune Teller (Carava...	http://dbpedia.org/resource/	Capitoline_Wolf name "Capitoline Wolf"	1 of 8 (12%, see tooltip)	<input type="checkbox"/>
Esquiline Venus	http://dbpedia.org/resource/	Capitoline_Wolf title "Capitoline Wolf"	1 of 8 (12%, see tooltip)	<input type="checkbox"/>
The Burial of St. Petronilla	http://dbpedia.org/resource/	Capitoline_Wolf label "Capitoline Wolf"	1 of 8 (12%, see tooltip)	<input type="checkbox"/>
The Tower of Babel (Brueg...	http://dbpedia.org/resource/	The_Tower_of_Babel_(Br...	1 of 8 (12%, see tooltip)	<input type="checkbox"/>
Danza di contadini	http://dbpedia.org/resource/	The_Peasant_Dance	1 of 8 (12%, see tooltip)	<input type="checkbox"/>
Portrait of Innocent X	http://dbpedia.org/resource/	Portrait_of_Innocent_X	1 of 8 (12%, see tooltip)	<input type="checkbox"/>
Madonna and Child Enthro...	http://dbpedia.org/resource/	Madonna_and_Child_Enth...	1 of 8 (12%, see tooltip)	<input type="checkbox"/>
The Red Vineyard	http://dbpedia.org/resource/	The_Red_Vineyard	1 of 8 (12%, see tooltip)	<input type="checkbox"/>
Femme nue couchée	http://dbpedia.org/resource/	Femme_nue_couch%C3...	1 of 8 (12%, see tooltip)	<input type="checkbox"/>
Malle Babbe	http://dbpedia.org/resource/	Malle_Babbe	1 of 8 (12%, see tooltip)	<input type="checkbox"/>
Tributo de la moneda	http://dbpedia.org/resource/	The_Tribute_Money_(Titian)	1 of 8 (12%, see tooltip)	<input type="checkbox"/>
Tempi Madonna (Raphael)	http://dbpedia.org/resource/	Tempi_Madonna_(Raphael)	1 of 8 (12%, see tooltip)	<input type="checkbox"/>
Chrysalis (sculpture)	http://dbpedia.org/resource/	Chrysalis_(sculpture)	1 of 8 (12%, see tooltip)	<input type="checkbox"/>

Fig. 9: Positive matches

Ideally, all preselected resources are related to the target and therefore suited for an annotation. To check whether the suggestion does apply, a tooltip does show up when moving the cursor to the number of according properties of that resource. The tooltip indicates which properties exactly are in accordance with the target. Based on this information the user can decide whether this resource should be annotated or not. One example of a preselected

resource is shown in Figure 9. This resource is preselected as the number of properties that match properties of the target resource fulfill the requirements  $7 \geq 1$  and  $\frac{7}{8} \geq 0.2$ . The tooltip of Figure 9 indicates that this resource is in fact related to the resource introduced in Section 5.1. Therefore, the suggestion of our prototype was correct (true positive) in this case.

### 5.3 Creating new Annotations

Once the resources for annotation are reviewed, the annotation can be serialized. When serialized, an info message is displayed to show the result of the serialization including the ID for the newly created annotation, as shown in Figure 10. In order to fulfill R2, this serialization has also to be published. The provenance information of the new annotation encoded in this output can be used to filter annotations by type, creator, annotated object, annotation and version, which fulfills R3.

```

1  @prefix prov: <http://www.w3.org/ns/prov#> .
2  @prefix foaf: <http://xmlns.com/foaf/0.1/> .
3  @prefix agent: <http://www.example.org/agents/> .
4  @prefix dbpr: <http://dbpedia.org/resource/> .
5  @prefix oa: <http://www.w3.org/ns/oa#> .
6  @prefix ind: <http://www.example.org/individuals/> .
7  @prefix anno: <http://www.example.org/annotations/> .
8  @prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
9  @prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
10
11  anno:88e6667b-00f0-4591-89a9-618481c4f13a
12      a                oa:Annotation ;
13      oa:annotatedAt   "2015-01-27T16:02:03.309Z"^^xsd:dateTime ;
14      oa:annotatedBy   agent:e98eb2d5-fd26-4b29-9ab5-dddaed08c12f ;
15      oa:hasBody       dbpr:Capitoline_Wolf ,
16                      dbpr:Capitoline_Wolf,_Chisinau ;
17      oa:hasTarget     ind:1181 ;
18      oa:motivatedBy   oa:editing ;
19      oa:serializedAt  "2015-01-27T16:02:34.213Z"^^xsd:dateTime ;
20      oa:serializedBy  agent:a69f0971-ff45-4af0-a197-a3f61cfa163d .
21
22  agent:e98eb2d5-fd26-4b29-9ab5-dddaed08c12f
23      rdf:type         foaf:Organization ;
24      foaf:name        "FZI" .
25
26  agent:a69f0971-ff45-4af0-a197-a3f61cfa163d
27      rdf:type         prov:SoftwareAgent ;
28      foaf:name        "OpenAnno 0.3" .

```

Fig. 10: Output of OpenAnno in Turtle

## 6 Limitations and Conclusion

In this paper, we have shown that a domain ontology can be extended by OADM in order to provide meaningful, rich annotations. Our approach allows to include provenance information for new annotations for data modeled in a domain ontology without destroying the structure of the domain ontology. By including provenance information of annotations the annotation process may also be crowdsourced as an collaborative task for new annotations without decreasing the quality of museums' data. We have shown how this approach can be used in order to complement DH data by annotating them with existing resources in LOD while obtaining the provenance information of the new annotations.

However, there are limitations for our approach. In particular, when using the OADM for annotations, it is not possible to state the type of relation between the annotated resource and the resource used for the annotation explicitly, e.g. "is part of", "consists of" or "is same as". These specific relations have to be implemented separately, as they are not provided by the OADM. In addition, our approach does not include any cryptography that ensures the authenticity of the encoded provenance information. For a real justification of the trustworthiness of annotations, a cryptography module has to be added in order to ensure the authenticity of annotations.

**Acknowledgements.** This work was supported by the German Federal Ministry for Economic Affairs and Energy (BMWi) within the CultLab3D project (Ref. 01MT12022D).

## References

- [BCT07] Breitman, K. K.; Casanova, Marco Antonio; Truszkowski, Walt: *Semantic Web: Concepts, technologies and applications*. NASA monographs in systems and software engineering. Springer, New York and London, 2007.
- [BL06] Berners-Lee, Tim: , *Linked Data - Design Issues*. <http://www.w3.org/DesignIssues/LinkedData.html>, 2006.
- [Br11] Bratsas, Charalampos; Dimou, Anastasia; Alexiadis, Georgios; Kavargyris, Konstantinos; Parapontis, Ioannis; Bamidis, Panagiotis; Antoniou, Ioannis: , *Educational Semantic Wikis in the Linked Data Age: The case of MSc Web Science Program at Aristotle University of Thessaloniki*. *Linked Learning 2011: the 1st International Workshop on eLearning Approaches for the Linked Data Age*, 2011.
- [CI11] CIDOC: , *CIDOC-ICOM recommendation on Linked Open Data for museums –Draft*. [http://cidoc-crm.org/docs/LoD\\_For\\_Museums\\_v1.7-en.doc](http://cidoc-crm.org/docs/LoD_For_Museums_v1.7-en.doc), 2011.
- [CI13] CIDOC CRM Special Interest Group: , *Definition of the CIDOC Conceptual Reference Model*. [http://cidoc-crm.org/docs/cidoc\\_crm\\_version\\_5.1.2.pdf](http://cidoc-crm.org/docs/cidoc_crm_version_5.1.2.pdf), 2013.
- [CJ14] Cyganiak, Richard; Jentzsch, Anja: , *The Linking Open Data cloud diagram*. <http://lod-cloud.net/>, 2014.
- [CSV12] Ciccarese, Paolo; Sanderson, Robert; Van de Sompel, Herbert: , *Open Annotation Draft Data Model*. <http://www.openannotation.org/spec/core/20120328.html>, 2012.
- [IC10] ICOM: , *LIDO's background*. <http://network.icom.museum/cidoc/working-groups/lido/lido-overview/lidos-background/>, 2010.

- [KK13] Klimpel, Paul; Keiper, Jürgen, eds. Was bleibt? Nachhaltigkeit der Kultur in der digitalen Welt. iRights.Media, Berlin, 2013.
- [LT01] Lazinger, Susan S.; Tibbo, Helen R.: Digital preservation and metadata: History, theory, practice. Libraries Unlimited, Englewood, Colo., 2001.
- [SCV13] Sanderson, Robert; Ciccarese, Paolo; Van de Sompel, Herbert: , Open Annotation Data Model. <http://www.openannotation.org/spec/core/>, 2013.
- [Sz13] Szekely, Pedro; Knoblock, Craig A.; Yang, Fengyu; Zhu, Xuming; Fink, Eleanor E.; Allen, Rachel; Goodlander, Georgina: Connecting the Smithsonian American Art Museum to the Linked Data Cloud. 2013.
- [Un04] United Nations Educational, Scientific and Cultural Organization: Records of the General Conference: 32nd Session. United Nations Educational, Scientific and Cultural Organization, Paris, 2004.
- [Wi14] Wittenberg, Jamie: Retaining Metadata in Remixed Cultural Heritage Objects. Proc. Int'l Conf. on Dublin Core and Metadata Applications 2014, 2014.