Abstract

The CIDOC Conceptual Reference Model (CRM) is regarded as an interoperability solution for integrating heterogeneous metadata in the cultural heritage domain. The major problem developers are confronted with when applying this model in real-world applications is that the CRM constitutes a formal ontology model on a semantic level but lacks technical specifications and guidelines on how to integrate that model with other, data source-specific models. This leads to divergent mappings between proprietary metadata models and the CRM ontology model. The aim of this paper is to provide *Model Implementation Guidelines*, including a CIDOC CRM mapping methodology, which are based on our practical experience and can serve as a recommendation for future CIDOC CRM adopters. Besides explicitly pointing out the problems the current standard entails, we propose a methodology for reducing the divergence of CRM mappings, which should bring CRM-based systems one step further towards metadata interoperability.

Towards Model Implementation Guidelines for the CIDOC Conceptual Reference Model

PHILIPP NUSSBAUMER
University of Zurich, Switzerland

BERNHARD HASLHOFER and WOLFGANG KLAS
University of Vienna, Austria

June 29, 2010

1 Introduction

Many institutions in the cultural heritage field, such as museums, libraries, or archives, are confronted with the need to integrate their metadata with those of other institutions. The obstacle they are facing is that their metadata are not interoperable because, even within a single application domain, they follow different metadata models. Two archaeological institutions, for instance, both located in the same geographical region, store metadata about the same type of digital asset (e.g., coins) by using distinct, institution-specific metadata models[1]. The CIDOC Conceptual Reference Model (CIDOC CRM) [8], which recently became an ISO standard, is a potential solution for achieving metadata interoperability. It defines a global ontology as semantic basis for describing and structuring metadata models in the cultural heritage domain.

The goal of the CRM model is to serve as a unified model as well as an instrument for integrating information into a global knowledge network, as described in [13]. Making third-party metadata and domain models accessible via some interoperable layer requires the mapping of these models to the CRM in order to resolve the heterogeneities among them and to achieve interoperability.

Although CIDOC CRM offers high-level concepts for structuring an application domain, we have encountered some technical and conceptual difficulties in establishing interoperability with data source-specific models. The main problem is that the degree of freedom in interpreting the CRM concepts is very high, with the result that semantically related metadata could be mapped to different CRM concepts, which would cause the goal of achieving interoperability to fail. Currently, the CIDOC CRM standard defines the model on a purely conceptual level and does not impose guidelines defining how to integrate and map third-party models to the CRM.

[1] e.g., the Portable Antiquities Scheme (`http://www.finds.org.uk`), and the Archaeology Data Service (`http://ads.ahds.ac.uk`)

As essential contributions of this paper, we first point out the problems the current CIDOC CRM standard entails when being implemented in real-world application scenarios. Secondly, based on the experience we have made with the CRM in the BRICKS[2] project, we propose *Model Implementation Guidelines*, which include a methodology for mapping source models to the CRM. This methodology should assist future adopters of the CRM, steer the application of CIDOC CRM modeling primitives when third party models are integrated, and produce more consistent and less divergent mappings.

The structure of this paper is as follows: in Section 2, we give a general introduction to the CIDOC CRM. Subsequently, we describe the problems that arise when applying the CRM in real-world application scenarios and analyze how existing CRM solutions solve these problems. In Section 3, we present our *Model Implementation Guidelines*, which should lead systems adopting the CRM further toward interoperability. After giving an overview of the related work in Section 4, we summarize and conclude this paper in Section 5.

2 Background

2.1 The CIDOC CRM

The CIDOC Conceptual Reference Model is an object-oriented ontology designed for the cultural heritage domain. It has been developed to meet the needs of integrating, mediating, and exchanging heterogeneous information from museums, libraries, and archives. Version 4.2.1 of the ontology consists of a set of 81 classes and 132 properties to describe things, concepts, people, places and time, and their relationships. The CRM's main purpose is to provide the semantic basis for describing data models and metadata schemes already in use within the cultural heritage domain. Both the entities (classes) and relationships (properties) are arranged in multiple *isA* hierarchies — Figures 1 and 2 show small sections of these hierarchies.

The CIDOC CRM is property centric, which means that classes were introduced to describe the domain and range of properties, such "that any other ontological refinement of the classes can be done as additional 'terminological distinction' without interfering with the system of relationships" [12]. Since it was designed to support alternative opinions and incomplete information, the CIDOC CRM does not impose cardinality constraints on its properties and recommends to implement properties as optional and repeatable, i.e., an object can have multiple properties of the same type (e.g., *P3 has Note*) with different values.

The design of the model encourages extension by inheritance, i.e., introducing more specialised entities and properties if the existing concepts' semantics are not sufficient. An example for such an extension could be the introduction of two classes "Sculpture" and "Vase" as sub-classes of *E22 Man-Made Object*.

[2]The BRICKS project: `http://www.brickscommunity.org/`

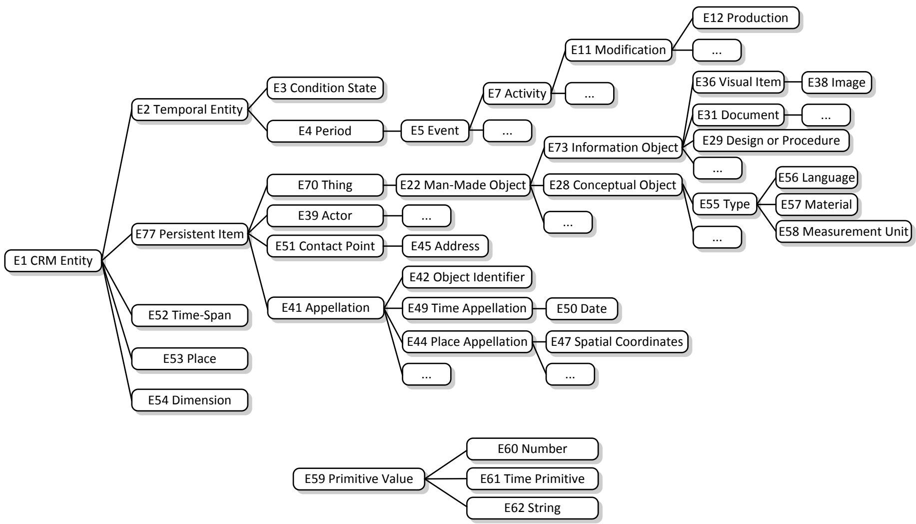


Figure 1: Small section of the CIDOC CRM classes hierarchy (version 4.2.1)

Property id	**Property Name**	**Entity – Domain**	**Entity - Range**
P1	is identified by (identifies)	E1 CRM Entity	E41 Appellation
P47	- is identified by (identifies)	E19 Physical Object	E42 Object Identifier
P48	- - has preferred identifier (is preferred identifier of)	E19 Physical Object	E42 Object Identifier
	...		
P2	has type (is type of)	E1 CRM Entity	E55 Type
P3	has note	E1 CRM Entity	E62 String
	...		

Figure 2: Example properties of the CIDOC CRM properties hierarchy (version 4.2.1)

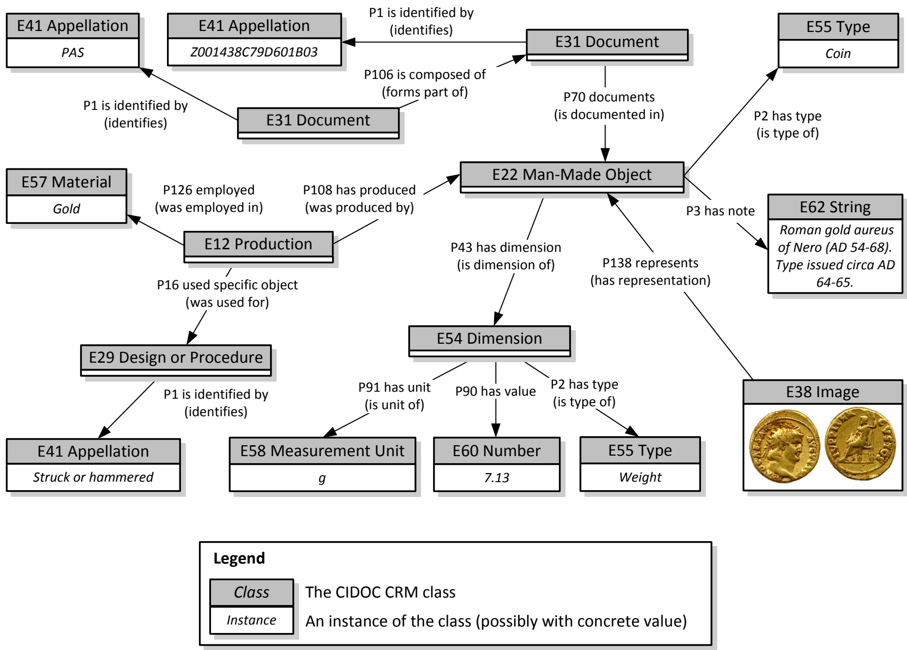


Figure 3: A coin from the Portable Antiquities Scheme described using the CIDOC CRM

To illustrate how the CIDOC CRM can be applied in the domain of archaeology, Figure 3 depicts metadata describing a single coin in a CRM-compliant form. The entire graph represents the description and depicts the CRM classes as well as the respective application-specific instances. Using only the high-level concepts provided by the CRM, the semantics of the graph is as follows:

- The described object is a *Man-Made Object* of *Type* “Coin”.
- It *is documented in* a *Document identified by* an *Appellation* “Z0014[...]”, forming part of another *Document identified by* the *Appellation* “PAS”.
- The coin has a documented *Dimension* of *Type* “Weight”, having the *Number* value “7.13” and a *Measurement Unit* of “g”, as well as a *String* note of “Roman gold aureus [...]”.
- Regarding the *Production* of the object we learn that the *Material employed* in the production activity was “Gold” and the *Design or Procedure* used is identified by an *Appellation* “Struck or hammered”.
- Finally, the object *has* a *representation* in the form of an *Image*.

The defined properties’ domains and ranges are restricted to specific classes, which means that properties are restricted to relate entities of fixed classes.

Property *P2 has type*, for instance, has a domain of *E1 CRM Entity* and a range of *E55 Type*. This class offers a second alternative to semantically extend and specialise the CRM: using instances of *E55 Type* allows to declare that an object is a member of a class defined outside the CRM, possibly linking to external vocabularies, thesauri, or ontologies and thereby allowing to refine and specialise the existing class hierarchy. In the above example, we have used this typing mechanism to declare that the object of discourse is of some type "Coin", which might be elaborated in a vocabulary outside the CRM. As property *P2 has type* is inherited by all sub-classes of *E1 CRM Entity* (any class of the model excluding *Primitive Value* and its sub-classes), links to external type hierarchies may be used to refine virtually every concept of the model (including properties).

Since the CRM is meant to cover the whole cultural heritage domain, it defines very abstract concepts. This leads to an ample scope of interpretation and gives users high degrees of freedom in representing their metadata. Figure 4 exemplarily shows how metadata may be represented differently depending on the interpretation of CRM concepts: Representation A models the concepts "Material" and "Method of Manufacture" different from Representation B, which does not model the type of the object and the procedure used at its production, but relates the material to the object itself instead of the production activity. Nevertheless, both representations are perfectly valid in terms of correct application of CRM primitives.

Due to the high degree of freedom resulting from the very high level semantics of concepts and their imprecise definitions, the same facts can easily be modelled in various ways — each alternative featuring slightly different semantics for actually identical concepts.

2.2 Problem Areas

Having introduced the CIDOC CRM, here we discuss in detail the major issues that arise when applying it in data integration scenarios. They can be categorized according to three subsequent tasks, which must be performed when applying the CRM in real-world applications: (i) lifting the source models and metadata to a common technical representation, (ii) mapping the concepts in these models to the CRM, and (iii) processing the mappings and visualizing the integrated metadata at the application-level.

2.2.1 Lifting and Normalization

Institutions organize their cultural metadata using different systems (e.g., relational databases), which are based on different data models. Accessing digital assets in distributed repositories requires uniform representation of metadata and schema definitions in all involved data sources. This guarantees that uniform access to the integrated data can be provided via a certain query language. A uniform model and schema representation is a necessary prerequisite for defining semantic correspondences between their concepts.

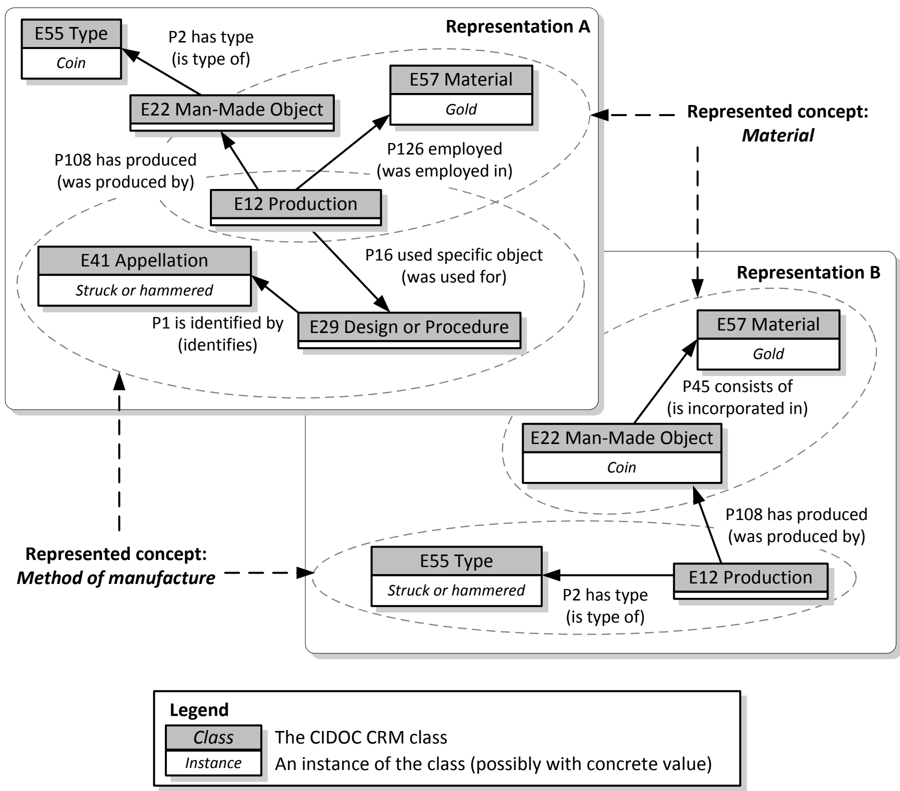


Figure 4: Different valid CRM representations for equal metadata attributes

Although lifting and normalization is not a CIDOC CRM specific task, some issues arise because it abstracts from any technical representation or implementation. There exist model definitions expressed in OWL and RDFS, but some open issues remain. The OWL definition, for instance, lacks properties which allow to store instance values of different data types. When lifting external schemes to a common representation, a precise technical specification of the target data model is required. Lacking specifications result in workarounds which lead to incompatibilities and impede metadata exchange.

2.2.2 Mapping

As any global ontology approach for achieving interoperability, the CIDOC CRM requires each source model to be mapped. In the following we discuss aspects that complicate such a mapping.

Semantic correspondences between the concepts defined in the CRM and those in the local schemes are identified by domain experts that precisely know the semantic definitions of both schemes. To define a mapping they must create a *mapping chain* (or mapping *path*) for each source concept. We define a mapping chain as a sequence of semantically associated classes and properties, representing a specific concept (for example: the concept of an object's *material* may be represented by the mapping chain *E22 Man-Made Object – P45 consists of – E57 Material*). We further make *validity* a condition of a mapping chain presupposing that a chain is valid in terms of correctly using the CRM's modelling primitives. This means that the chain follows the properties' domain and range restrictions.

However, the CRM provides no guidance on how to define such a mapping path, i.e., which classes and properties to use for representing a concept. We have experienced that domain experts have great difficulties in mapping their models to the CRM. The main reasons are its complexity, the high semantic abstraction level of its concepts, and the lack of a mapping methodology. Consequently, the mapping process had to be supported by a CRM expert.

If several institutions map their source schema independently in absence of mapping guidelines it is likely that the domain experts create divergent mapping chains. It can occur, for instance, that semantically equivalent metadata are mapped to different chains or, reversely, that identical chains are applied for semantically different metadata attributes.

In Figure 5, we illustrate possible mapping inconsistencies when creating equivalent chains for equivalent metadata. The metadata fields "SecUID" (PAS) and "SampleRef" (ADS) are semantically equivalent, denoting an (internal) item identifier. However, mappings created by different experts resulted in differing mapping chains:

```
PAS:SecUID       E22-P47-E42
ADS:SampleRef    E22-invP70-E31-P1-E41
```

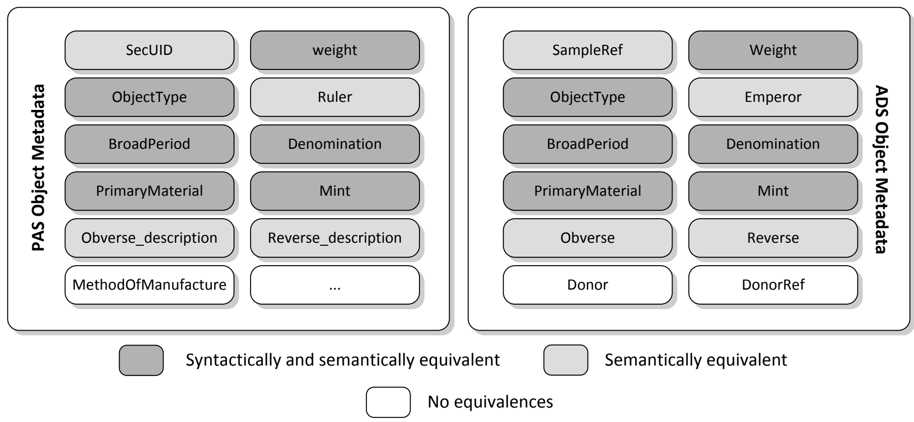


Figure 5: Correspondences between the PAS and ADS metadata fields

While the PAS identifier (PAS:SecUID) is encoded as "(E22) *Man-Made Object* (P47) *is identified by* (E42) *Object Identifier*", the semantically equivalent ADS identifier (ADS:SampleRef) is represented by "(E22) *Man-Made Object* (invP70)[3] *is documented in* (E31) *Document* (P1) *is identified by* (E41) *Appellation*". As a result, the semantic equivalence fails to be preserved from the CRM perspective.

2.2.3 Processing and Visualization

The benefit of a global ontology is that client applications can transparently access metadata from heterogeneous sources without dealing with the semantic definitions in the source models. However, as it is the case with the CIDOC CRM, a global ontology can be very complex. Therefore it is necessary to hide this complexity from the user and provide an easy-to-use Graphical User Interface (GUI) for efficient search and retrieval. One possibility is to provide a configurable faceted search interface that takes into account different mapping structures.

In case of the CIDOC CRM, formulating queries respecting the (semantic) correct combination of classes and properties of the model exhibits high complexity. The CRM makes no propositions on which metadata information to map or which combination of classes and properties to use, such that the structure (and therefore the semantics) of the actual graph that represents the metadata depends on the mapping expert's interpretation. Consequently, querying for specific aspects (e.g., a coin's diameter) requires the incorporation of mapping information such as the different chains and vocabularies used for mapping the metadata to the CRM.

[3] *inv*: abbr. for *inverse*, meaning that the property's semantics changes from "documents" to "is documented in".

The CIDOC CRM's structure suggests a graph-like presentation of metadata, so that a user can browse through the associated concepts. However, such a graph representation — with nodes and edges labelled using the CIDOC CRM's conceptualization terms — may lack sufficient comprehensibility for certain user groups (i.e., users that are not familiar with the CIDOC CRM). A common understanding of the information in such a graph is further impeded by possibly inconsistent sub-graphs for equivalent metadata information, caused by divergent mappings.

2.3 Existing Solutions

In this section, we discuss existing applications that aim to establish interoperability amongst heterogeneous data sources by using the CIDOC CRM. We will analyze them in terms of their approaches on lifting and normalizing the source data, mapping the source models to the CRM, and how they process and visualize the integrated data.

2.3.1 Lifting and Normalization

Though the CIDOC CRM abstracts from technical representations of the model, there are OWL and RDFS definitions available. While [14], [23], and [24] make use of the RDFS definition and store the CRM-compliant metadata as RDF instances, the Sculpteur project [2] keeps the original data structures and associates existing metadata attributes with CIDOC CRM mapping chains [37]. In the eCHASE project [38], the unified metadata repository consists of different areas for various application functionalities: the legacy data is stored in its original structure for searching and displaying purposes, whereas a subset of the legacy metadata is mapped into a highly structured database schema (influenced by the CIDOC CRM) which supports structured browsing.

In the BRICKS application context, the OWL definition of the CIDOC CRM has been applied. Since the CRM itself is not concerned with representing instance data, also the OWL definition lacks a mechanism for storing instance data of different data types. Therefore, the CRM model had to be extended by properties based on XML Schema data types.

2.3.2 Mapping

For discussing the existing approaches in mapping discovery, we separately describe approaches to *discover* the mappings between source models and the CRM, and solutions to *represent* the mappings for further processing.

Mapping Discovery In the BRICKS project, the involved institutions had great difficulties in mapping their models to the CRM, due to its complexity and the lack of a mapping methodology. Consequently, the mapping process had to be supported by a CRM expert. In spite of that, creation of the mappings proved to be time-consuming and error-prone. Other projects applying the CRM share this experience: in regard to the Sculpteur project [2], Sinclair et al. [37] report that defining mappings to the CRM is a complex and time-consuming task, which requires a "good understanding of both ontological modelling as well as the source metadata system." Also in the eCHASE project [38], completion and validation of the mappings required assistance of a CRM expert — an experience also shared by the works reported in [39] and [7]. The Perseus digital library and the Arachne archaeological database, as described in [23, 24], map only the most important fields in order to reduce the mapping complexity and effort.

A tool that aims to support both the mapping discovery and representation has been developed within the EPOCH initiative[4]: the Archive Mapper for Archaeology (AMA) tool provides a Web-based user interface that allows to map source models to the CIDOC CRM and export the mappings to XML. It does not, however, provide guidance on how to map source concepts to the corresponding CRM chains or validation of the created CRM statements.

Mapping Representation In the BRICKS context, experts may define mappings using pre-formatted spreadsheets — an approach that has also been used for, e.g., mapping the MIDAS standard to the CRM[5] and the efforts reported in [7]. These spreadsheets are automatically transformed to XSL stylesheets used for transforming and ingesting the source metadata into the system. To allow such transformations, it is assumed that each institution's data are available in XML or can be exported as XML.

In the Sculpteur project, the result of the mapping process is a table that links database fields to the respective CRM chains. At query-time, the fields are dynamically mapped to their CRM representations. In the project reported in [24], mappings are semi-formally documented along with annotations for iterative refinement. These mappings are then manually converted into XSL stylesheets that are used for transformation.

The approach discussed in [14] involves different specialized programs to transform the source data to a common CRM-compliant XML representation, which may then be ingested from a source into the target system.

Another approach on mapping and mapping representation is presented in [22]: the authors present a mapping language for information integration, also including a mapping annotation format. We will base our mapping methodology presented in Section 3 on parts of this work and discuss its main ideas there.

2.3.3 Processing and Visualization

While in [37] and [38] search and retrieval on legacy data relies on standardized protocols such as SRU/SRW[6], and [24] only considers browsing the metadata, the work presented in [14] allows to query a CRM-compliant RDF database;

[4]European Network of Excellence in Open Culture Heritage: http://www.epoch.eu

[5]see `http://cidoc.ics.forth.gr/crm_mappings.html`

[6]Search/Retrieval via URL (SRU/SRW), available at `http://www.loc.gov/standards/sru/`

however, it is not clear how the complexity of such queries is handled, especially in regard to possible mapping ambiguities.

BRICKS provides a configurable faceted search interface, taking into account the possibility of different mapping structures. This solution involves creation of SPARQL[7] queries from a set of known mapping chains. These chains are used to query along the "metadata attributes" encoded implicitly in the CRM graph, whereas the "attributes" refer to the source schemes' concepts initially mapped to the CIDOC CRM. For example, when querying for all coins from the Roman period, the mapping chains of the "period" attributes of both the ADS and PAS schemes have to be combined in a single query. In the Sculpteur project, a concept browser allows users to navigate to concepts of interest. However, the complexity and terminology of the CRM hinders an intuitive presentation. Therefore, simplifications based on the legacy metadata structures can increase familiarity for the museum users. They allow, for instance, to group and merge CRM concepts and to associate them to comprehensible terms, such as "who" or "when", thus making the CRM's structures and terms transparent for the users. Similar to the interface created in the eCHASE project, the concept browser makes use of an mSpace[8] interface to visualize and query instances.

The work presented in [24] visualizes CRM-compliant data structures using the Longwell browser[9]. Thereby it provides a faceted-based search interface for CRM entities and properties. However, users of such functionality must inevitably be familiar with the CRM and its structures.

In BRICKS, an *application-specific vocabulary* (cf. Section 3.2.1) has been introduced to associate equivalent metadata mappings to unambiguous terms. To retrieve metadata the CRM mappings are iterated, the according information is extracted, and the results are rendered as attribute–value pairs (e.g., "diameter: 8.17"). Given that the metadata mappings are unambiguously associated to vocabulary terms, such a representation is semantically well-defined and easier to comprehend.

2.4 Observations

Having discussed the problem areas and existing approaches that aim at solving these shortcomings, we can observe that achieving one of the main goals of the CIDOC CRM — providing interoperability — is hindered by the following issues:

1. The CIDOC CRM lacks a standardized and complete technical representation. Consequently, applications that integrate data sources based on the CRM are very heterogeneous — not only in terms of their technical implementations and interfaces but also regarding their application-specific extensions of the standard.

[7] SPARQL is a query language and protocol for RDF; see `http://www.w3.org/TR/rdf-sparql-query/`

[8] http://www.mspace.fm

[9] http://simile.mit.edu/wiki/Longwell

2. Due to the CRM's complexity, discovering mappings is time-consuming and requires assistance of CRM experts in order to correctly apply and validate the model. Furthermore, its unrestrictive nature results in a high degree of freedom regarding the modeling of semantic concepts, causing semantically incompatible mappings. This is further aggravated by the absence of tools providing support for mapping identification and validation.

3. Finally, the aforementioned shortcomings (or rather the diversity in solving them) culminate in complicated processing and visualization of CRM-encoded data.

Though the discussed problems are partially solved in the various application-specific approaches, allowing for creation of "intra-operability" between heterogeneous data sources on application level, the technical and — most notably — *semantical* interoperability is hindered when it comes to the integration of individual data sources that are already encoded in a CIDOC CRM-compliant form.

3 Model Implementation Guidelines

Previously, in Section 2.2, we have outlined the major issues application developers are facing when integrating metadata using the CIDOC CRM: first they must bring the source models and metadata to a common technical representation (*lifting and normalization*), then they must determine and represent the semantic correspondences between the concepts in the source models and the CRM (*mapping*), and finally, they must process the mappings and visualize the integrated metadata (*processing and visualization*). In this section, we describe our *Model Implementation Guidelines*, which comprise a potential solution for the first two issues.

3.1 Lifting and Normalizing Source Models to the CRM Level

Usually data sources employ different systems and data models for managing their metadata. In order to map metadata from a source model to a target model, it is helpful to lift the source model to a representation that features the same modeling primitives as the target model: [22] present a mapping approach that lifts the source schema to a semantic model residing on the same conceptual level as the CIDOC CRM. Following this approach, metadata stored in a relational database can be interpreted as follows:

- the relation names, e.g., "coin", describe entities (source domain)
- attribute names, e.g., "SecUID", are interpreted as properties connecting source domain entities with source range entities

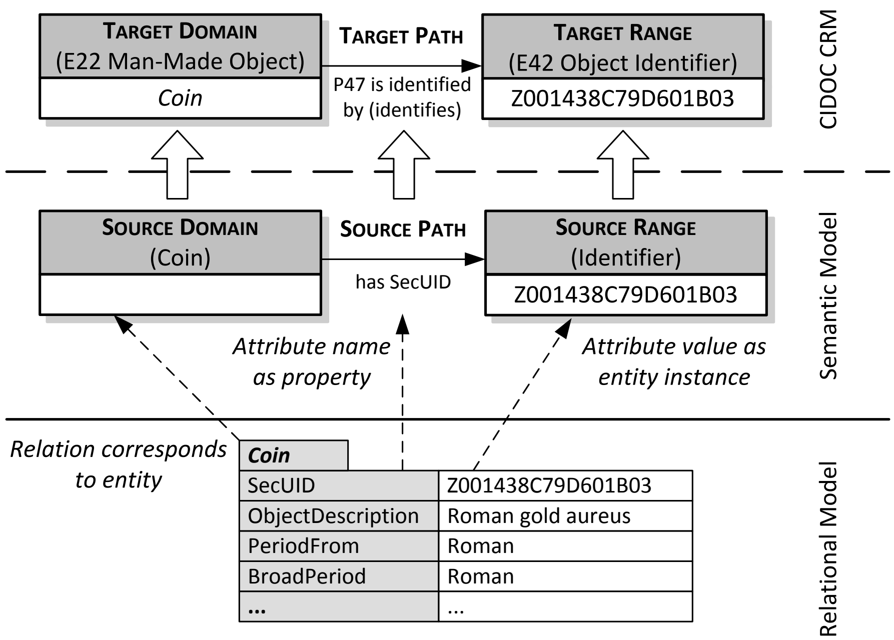


Figure 6: Lifting a relational model to the CRM via a semantic model

- attribute values are instances of the newly created source range entities

Figure 6 illustrates the lifting of the PAS database schema to the proposed semantic level: the relation is described as the *Source Domain*, the attribute name is lifted to the *Source Path*, and the attribute value is depicted as the *Source Range*. When mapping the model to the CIDOC CRM, the defined source entities and properties are mapped to their respective *target* pendants.

3.2 A Generic CIDOC CRM Mapping Methodology

A major goal of our model implementation guidelines is to provide a methodology for discovering mappings among the concepts in source models and those in the CRM. In the following we will gradually develop and describe this methodology. First, we illustrate a simple approach that assigns terms from an *application-specific vocabulary* (e.g., "identifier" for object identifiers) to a set of CRM mapping chains. Since such an approach contradicts the idea of the CRM, we propose a generic mapping methodology that guides expert users through the process of mapping their source models to the CRM.

3.2.1 A First Simple Approach

To reduce the divergence of the mappings one can introduce an *application-specific vocabulary*, defining the terms of an application domain together with the corresponding CIDOC CRM mapping chains. One could specify, for example, that an attribute "identifier" is always mapped against the chain "*E19 Physical Object – P47 is identified by – E42 Object Identifier*".

For systematically designing such a vocabulary one can follow the dimensions of a narrative world — a methodology which has also been applied in other metadata standards such as MPEG-7[10] and is partly reflected also by the CIDOC CRM. A narrative world describes the semantics of real-world entities together with their attributes and relationships. The entities of a narrative world can, according to [5], be categorized as follows:

- *objects and events*: perceivable entities that exist or take place in time and space (e.g., a coin; roman period)
- *agent objects*: objects that are persons, group of persons, or organizations
- *concepts*: entities which cannot be perceived in the narrative world (e.g., "friendship")
- *semantic states*: properties of semantic entities at a specific point in time and space (e.g., age, weight, height)
- *semantic places and times*: locations and times in the narrative world

The example presented in Figure 7 shows an excerpt of the terms that serve as *application-specific vocabulary* in the BRICKS context for defining the common terms for the ADS and PAS schema. For each term it also defines the corresponding CIDOC CRM mapping chain.

Introduced a priori — hence before the source models are mapped against the CIDOC CRM concepts — such a vocabulary serves as "mapping index" where the domain experts can find CRM mapping chains for common terms in an application domain. If the index covers most of the concepts used in the source models to be integrated, it will lead to more consistent mappings. When divergent mappings are already in place, the introduced terms can serve as semantic glue and re-link the mapping chains.

The drawback of this simple approach is that the introduced vocabulary in fact serves as an intermediate layer between the source models and the CIDOC CRM. This implies that the vocabulary takes the role the CIDOC CRM should play, which arguably is not the intended goal.

[10] The MPEG-7 Multimedia Content Description Interface. Further information available at `http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm`

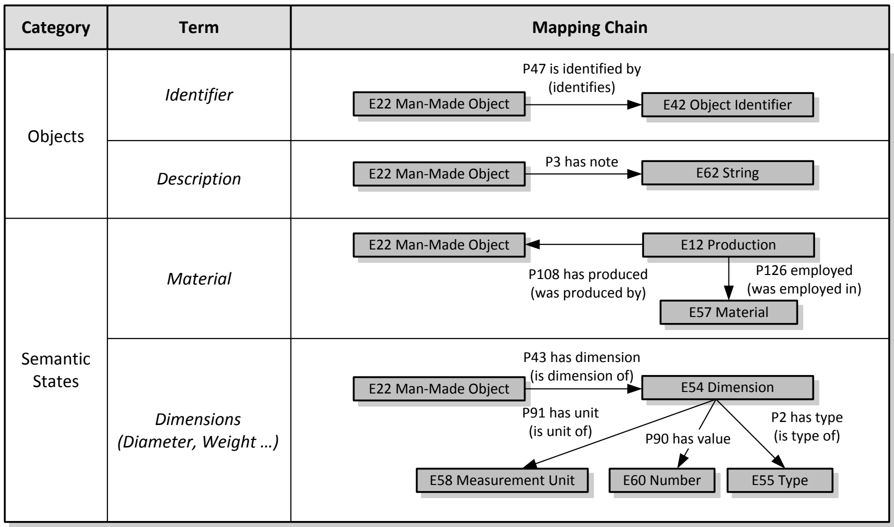


Figure 7: Example terms from the vocabulary used in BRICKS and the corresponding mappings proposed to use

3.2.2 The Generic CRM Mapping Methodology

After the source schemes have been lifted from their data source-specific data model representation (e.g., relational model, XML) to the level of CIDOC CRM, one can start mapping them against the CRM ontology. The advantage of a common representation is that one need not further deal with data model-specific mappings, e.g., how to map relational tables to the graph-based CRM-specific model, but can concentrate on the semantics of the concepts to be mapped. The goal of the following methodology is to convert each instance of an element of the source model to a valid instance of the CRM model while preserving its meaning. It is designed to:

1. Map the Source Domain to the Target Domain.
2. Map the Source Range to the Target Range.
3. Find the (Target) Path between the Target Range and the Target Domain.

Before describing the details of the methodology, we need to introduce the following notations: a mapping is always established between a source model M^{src} and the CRM model M^{crm}. Each model has a set of entities $e_0, e_1, \ldots, e_n \in E$ and a set of properties $p_0, p_1, \ldots, p_n \in P$. If A is an arbitrary set, then $\mathcal{P}(A) = \{x \mid x \subseteq A\}$ denotes the powerset of A. For the mapping methodology, we require the following functions:

- isA: $E \times E \longrightarrow BOOLEAN$ is defined as $isA(e_x, e_y) = (true \mid false)$ and returns true if entity e_x is a direct or indirect subclass of entity e_y. Each entity is also subclass of itself.

- getDomain: $P \longrightarrow \mathcal{P}(E)$ is defined as $getDomain(p) = \{e_i \in E\}$ and returns the set of all direct and indirect entities that are defined as domain of p. If p reads inversely, this function returns $getRange(p)$.
- getRange: $P \longrightarrow \mathcal{P}(E)$ is defined as $getRange(p) = \{e_i \in E\}$ and returns the set of all direct and indirect entities that are defined as range of p. If p reads inversely, this function returns $getDomain(p)$.

The mapping methodology, which is presented in Mapping Rule 1, works as follows: the inputs are a source specific model M^{src} lifted to a conceptual level and the CIDOC CRM model M^{crm}. The output of is a set of mapping chains, one chain for each property of the source specific model. The domain expert iterates over each entity in the source model ($e \in M^{src}$) and in a first step defines this entity as instance-of the most specific CIDOC CRM entity that describes its semantics (e_{start}). This entity also represents the beginning of a mapping chain. In each loop the expert regards all properties of the source model ($p \in M^{src}$) that have entity e defined as domain. In order to build the mapping chain for that property, the expert first regards the most specific CRM entity e_{end} that most accurately describes its semantics. This entity, marked by the variable x, represents the end of a mapping chain. In order to connect the end with the start of a mapping chain, the expert traverses the CIDOC CRM graph and repeatedly determines the most specific CRM property p_{crm}, which has e_{end} as range, and the most specific CRM entity e_{crm}, which is defined as domain of p_{crm}. Both parts of the chain link — p_{crm} and e_{crm} — are then added to the mapping chain. This traversal continues until e_{start} is the same class or a direct or indirect subclass of e_{crm}. Finally, e_{start} is added to c, which then represents the mapping chain in a reverse order. The expert then continues with the next property $p \in M^{src}$ with a proposed definition of e_{start}.

In Figure 8, we give an example for a mapping created between the PAS model lifted to a conceptual level (M^{src}) and the CIDOC CRM (M^{crm}) using the previously described methodology. Specifically, we regard the PAS entity *Coin* and the property *hasSecUID*, whose domain is the entity *Coin*. In the first step, we search the CRM model M^{crm} to find the most specific entity e_{start} describing the semantics of the PAS entity *Coin* (e). Regarding the entities in the CRM standard, it turns out that *E22 Man-Made Object* most appropriately describes the concept *Coin*. Therefore we denote *E22 Man-Made Object* as e_{start} and define *Coin* as instance of *E22 Man-Made Object*. In the second step we regard the properties of entity *Coin*, which in this case is only the property *hasSecUID*. For this property p we search the CRM standard for the most specific entity that describes the semantics of that property. It turns out that the corresponding entity is *E42 Object Identifier*, so we denote this entity as e_{end} and define the instance of the range of p, which is the value *Z00014...*, as instance of e_{end}. Now having the beginning (E22) and the end (E42) of the mapping chain, the goal is to connect both ends via a chain of CRM

Mapping Rule 1: CIDOC CRM mapping methodology

Data: a source specific model $M^{src} \in \mathcal{M}$ and the CIDOC CRM $M^{crm} \in \mathcal{M}$

Result: a set of mapping chains $c_0, c_1, \ldots, c_n \in C$ denoting the semantic correspondences between M^{src} and M^{crm}

forall the $e \in M^{src}$ *being source domain entities* **do**
    find the most specific entity $e_{start} \in M^{crm}$describing the semantics of e;
    define e as instance of e_{start};
    forall the $p \in M^{src}$ *with* $getDomain(p) = e$ **do**
        $c \longleftarrow \emptyset$;
        find the most specific entity $e_{end} \in M^{crm}$ describing the semantics of p;
        define the instance of the range of p as instances of e_{end};
        add e_{end} to the mapping chain c;
        $x \longleftarrow e_{end}$;
        repeat
            find the most specific property $p_{crm} \in M^{crm}$ such that $getRange(p_{crm}) = x$;
            add p_{crm} to the mapping chain c;
            find the most specific entity $e_{crm} \in M^{crm}$ such that $getDomain(p_{crm}) = e_{crm}$;
            add e_{crm} to the mapping chain c;
            $x \longleftarrow e_{crm}$;
        until *isA*(e_{start},x);
        invert the mapping chain c;
        $e_{start} \longleftarrow x$;

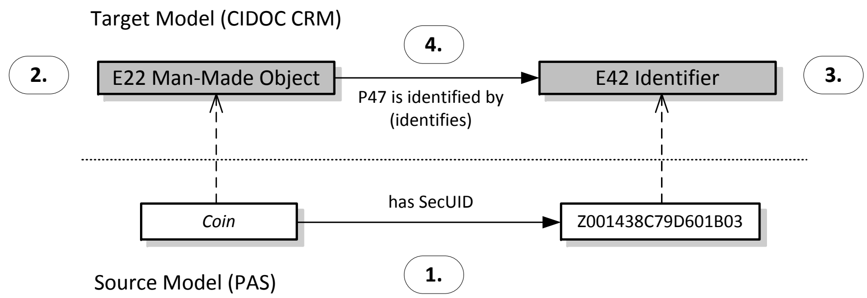


Figure 8: Mapping the PAS coin identifier to the CRM

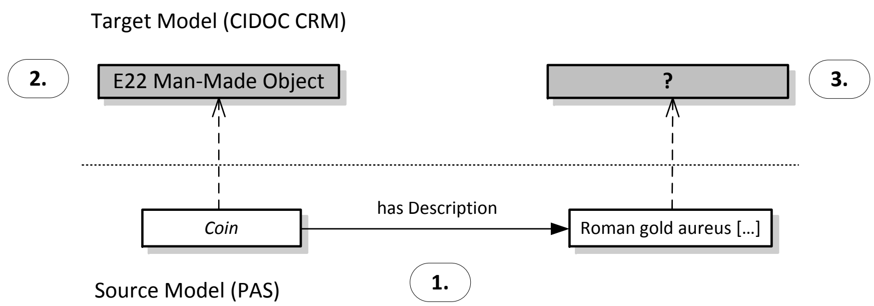


Figure 9: Mapping the PAS coin description to the CRM

properties and entities. We can achieve that by traversing the path backwards and searching for the most specific CIDOC CRM property p_{crm} that has e_{end}, hence *E42 Object Identifier*, as range. We find that *P47 is identified by* matches this criteria and add it to the mapping chain. Finally we regard the domain of p_{crm}, hence *P47 is identified by*, find out that this (e_{crm}) is *E19 Physical Object*. Since *E19 Physical Object* is a superclass of *E22 Man-Made Object* we can determine the mapping rule and add e_{start} to the mapping chain which is then, after being inverted, *E22:P47:E42*.

However, if we apply our generic mapping methodology for other properties, we encounter its limitations. In Figure 9 for instance, we try to find an appropriate mapping chain for the PAS property *hasDescription*. While the first and second step work analogously to the previous example, the third step fails because in the CIDOC CRM standard there is no "most specific entity that describes the semantics of *hasDescription*". One could of course choose a very abstract entity definition such as *E1 Entity*, which is also valid according to the CRM, but this would imply a loss of semantics.

Regarding the previous example it becomes obvious that a generic mapping methodology is sufficient for structuring the process of mapping source models to the CIDOC CRM, but insufficient for covering possible problems that can occur during the mapping process. Therefore it must be possible to refine the mapping rule and provide additional guidelines for specific categories of mapping problems.

3.2.3 Refining the Generic Mapping Methodology

The mapping methodology presented in the previous section can be improved by outsourcing the critical steps of the mapping rule into functions that can then be refined according to domain-specific needs. If we let $C \in \mathcal{C}$ be a single mapping chain, and $cl \in CL$ be a chain-link $cl =< e, p >$ with $e \in M^{crm}$, $p \in M^{crm}$, we can define the following functions:

- findTargetDomain: $E \longrightarrow E$ is defined as $findTargetDomain(e_x) = e_y$, where $e_x \in M^{src}$ and $e_y \in M^{crm}$, and returns the most specific entity in the CRM model describing the semantics of a given entity of the source model.
- findTargetRange: $P \longrightarrow E$ is defined as $findTargetRange(p) = e$, where $p \in M^{src}$ and $e \in M^{crm}$, and returns the most specific entity in the CRM model describing the semantics of a given property of the source model.
- findChainLink: $E \times E \longrightarrow CL$ is defined as $findChainLink(e_x, e_y) = cl$, where $e_x, e_y \in M^{crm}$ and $cl \in CL$, and returns the last chain link being part of a mapping chain between two CRM entities.

Mapping Rule 2 illustrates an excerpt of the methodology presented in the previous section but has delegated certain steps of the methodology to the functions defined above. If any of these functions returns no entity or property (e.g., $findTargetDomain = \emptyset$), the methodology skips a loop and proceeds with the next entity or property.

Mapping Rule 3 outlines how the function *findTargetRange* can provide guidance for situations when there is no “most specific entity that describes the semantics of a property”, as it is the case in the example presented in Figure 9. It could provide a set of hard-wired domain solutions and propose, for instance, to map all object descriptions to the CRM chain *P3:E62* as illustrated in Figure 10. In an if-else manner, the function should check for specific conditions and provide appropriate mapping solutions. If the function still determines without providing a solution, a source entity or property can be considered as non-mappable against the CIDOC CRM using our methodology. For *findTargetDomain* an analogous function can be defined.

Frequently, a domain expert can find the beginning and the end of a mapping chain but cannot determine the connecting chain links. Figure 11(a) illustrates such a situation: from the source path one can conclude that a *Coin* corresponds to *E22 Man-Made Object* and that the range of *hasEasting* denotes a *E47 Spatial Coordinate*. However, it is not obvious how to connect these two CRM entities. To solve that problem, the function *findChainLink*, as illustrated in Mapping

Mapping Rule 2: Refined CIDOC CRM mapping methodology

forall the $e \in M^{src}$ *being source domain entities* **do**
  $e_{start} \longleftarrow findTargetDomain(e)$;
  define e as instance of e_{start};
  forall the $p \in M^{src}$ *with* $getDomain(p) = e$ **do**
    $c \longleftarrow \emptyset$;
    $e_{end} \longleftarrow findTargetRange(p)$;
    define the instance of the range of p as instances of e_{end};
    add e_{end} to the mapping chain c;
    $x \longleftarrow e_{end}$;
    repeat
      $cl \longleftarrow findChainLink(e_{start}, x)$;
      add cl to the mapping chain c;
      $x \longleftarrow$ take the first element e_{crm} of cl;
      invert the mapping chain c;
      $e_{start} \longleftarrow x$;
    until $isA(e_{start}, x)$;

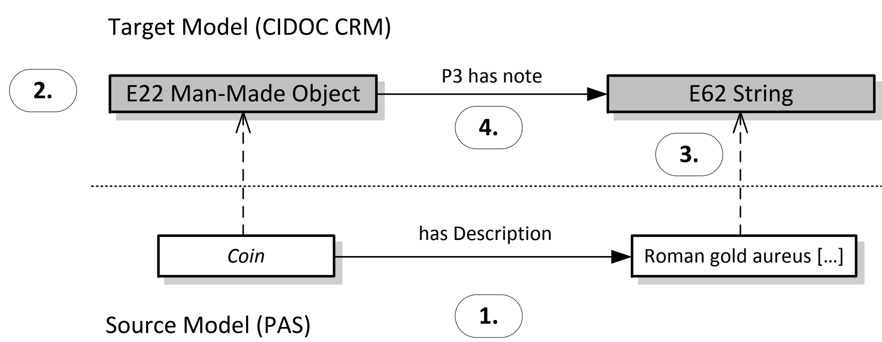


Figure 10: Hard-wired mapping of the PAS coin description to the CRM

Mapping Rule 3: domain-specific findTargetRange function

Data: a property $p \in M^{src}$

Result: an entity $e_{end} \in M^{crm}$

$e_{end} \longleftarrow$ the most specific entity $e_{end} \in M^{crm}$ describing the semantics of p if no semantic appropriate entity can be found, set $e_{end} = \emptyset$;

if $e_{end} = \emptyset$ **then**

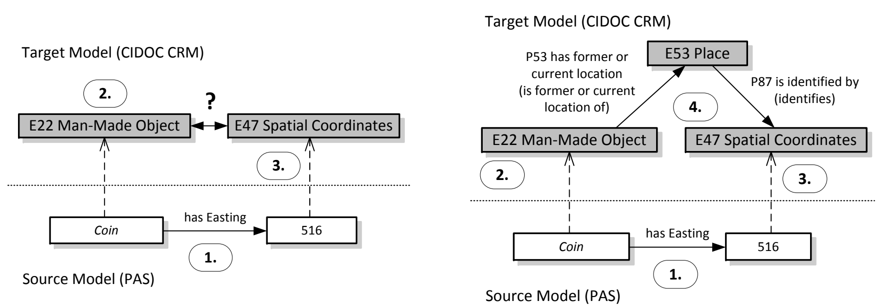
  if *getRange(p) is a textual description of an object within a narrative world* **then**

    $e_{end} \longleftarrow$ E62 String;

  if *getRange(p) describes the date of the production of an object* **then**

    $e_{end} \longleftarrow$ E61 Time Primitive;

  . . . //implementation of further individual domain-specific mappings



(a) (b)

Figure 11: Mapping the easting coordinate of (the finding place of) a PAS coin

Rule 4, provides the necessary guidance and proposes a single chain link to the domain expert. Having got a missing chain link, the methodology can further process the source path and provide the remaining links for a mapping chain. Figure 11(b) shows the mapping resulting from applying the function *findChainLink*.

As the caption of Figure 11 already hints, in this scenario the domain expert does not seem to have properly lifted the source schema to the conceptual level: the easting coordinate does not refer to the coin itself but to the *finding place* of the coin. In such cases the model may not be mapped to the CRM in an intuitive way. If the source path in the semantic model would read "Finding Place — has easting — 516", the source domain could easily be mapped to the target domain of "E53 Place", which then could be linked to the *E22 Man-Made Object* it refers to. As we can see, functions like *findChainLink* might not be necessary if the semantic model's path has been created correctly, however, such functions prove useful when the source path is not well-defined.

Mapping Rule 4: domain-specific findChainLink function

Data: two entities $e_{start}, e_{end} \in M^{crm}$
Result: a chain link $cl \in CL$

```
cl ⟵ ∅;
find the most specific property p_crm ∈ M^crm such that
getRange(p_crm) = x, if no such property can be found, set p_crm = ∅;
find the most specific entity e_crm ∈ M^crm such that
getDomain(p_crm) = e_crm, if no such entity can be found, set e_crm = ∅;
if (p_crm ≠ ∅) ∧ (e_crm ≠ ∅) then
    add p_crm and e_crm to the mapping chain cl;
else if e_start denotes an object and e_end the coordinates of a place in a
narrative world
then
    add P87 is identified by (identifies) and E53 Place to the mapping
    chain cl;
else if ... then
    ... //further domain-specific mapping chain definitions
```

3.3 Summary and Limitations

As part of our proposed *Model Implementation Guidelines*, we have presented potential solutions for lifting third party data models to the level of CIDOC CRM and for discovering and representing mapping relationships among model concepts.

The mapping methodology is not meant to be an automatic algorithm but rather a structured process intending to guide domain experts through the mapping task. For each link in the mapping chain, they must refer to the standard specification and determine the valid and semantically most appropriate CRM concept. An automated algorithm could of course pre-compute all possible chains and let the user select a chain. However, this would produce a large number of possible results and require the domain expert to analyze each link of each proposed mapping chains. This would make the whole mapping task more complex for the user than iteratively following the above mapping rule.

Our generic CRM mapping methodology, presented in Section 3.2, resembles the original idea of the CRM because it does not restrict the scope of possible mappings to a fixed set of chains listed in an application-specific mapping index. It leaves the freedom to map any source path to the CRM and guides the domain expert through a structured mapping process.

A major limitation of our approach is — as it is the nature of any generic solution — that it does not automatically fit for all possible scenarios. If a mapping iteration does not provide any result, domain-specific, hard-wired guidelines — such as those presented in section 3.2.1 — must be introduced in terms of functions. Although this reduces the freedom of a completely generic approach, we believe that this is a good trade-off between flexibility and restriction to a set of predefined mapping chains.

4 Related Work

Besides global ontology approaches such as the CIDOC CRM, there exist also other techniques for establishing metadata interoperability [6]: agreement on a single metadata standard, usage of application profiles [17], and the definition of bilateral crosswalks (mappings) between incompatible metadata schemes. Another metadata interoperability technique is the application of a metadata frameworks (e.g., MPEG-7 [20], MPEG-21 [19]).

Global ontologies have also been defined for other domains: the Functional Requirements for Bibliographic Records (FRBR) [18] is an entity-relationship model which should serve as a generalized view of the bibliographic universe, intended to be independent of any cataloguing standard or implementation [40]. The Suggested Upper Merged Ontology (SUMO)[11] is another example for a global ontology that "will promote data interoperability, information search and retrieval, automated inferencing, and natural language processing" [31]. The Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE) [45] is yet another example for a global ontology.

Reports on the quality and applicability of global ontologies for establishing interoperability are manifold: [32] states that the communities do not yet have enough experience to claim that global ontology approaches are a success. She refers to two reports, one describing the success [34] and another the difficulties [41] of using global ontologies for information integration. [43] asserts that no global ontology can be defined in such a way that it fulfills all ontological requirements of all possible information systems that are integrated in a certain domain. [16] argue that in large scale environments a global ontology becomes the bottleneck in the process. It must be designed and maintained very carefully and cannot change significantly without violating existing mappings from data sources. In general, global ontologies only work well in integration scenarios where the sources to be integrated share nearly the same view of a domain [42].

Basically there are two architectural possibilities for integrating the data sources: centralized or decentralized. In a centralized approach the metadata are converted according to the structural and semantic definitions in the CIDOC CRM and transferred into a central data store. In a decentralized approach the metadata reside in the data sources and are virtually integrated using a mediator-wrapper architecture [44]. The choice of the interoperability technique directly affects the architectural properties of such an architecture [42]: when using a single global ontology approach, the mediator exposes that ontology and the wrappers relate the encapsulated information to that ontology (e.g., SIMS [3], Ontobroker [9]). If bilateral mappings are the chosen interoperability technique, the roles of mediators and wrappers conflate and each information

[11] `http://ontology.teknowledge.com/`

source is described by its own ontology and semantic interoperability is established on a peer-to-peer basis (e.g., Observer [29], Edutella [30], GridVine [1]).

Finding mappings among the elements of incompatible schemes or ontologies is a crucial task in any integration scenario. Many (semi-)automatic mechanisms are available and have widely been discussed in the literature and summarized in several surveys: [21] provide a survey of (semi-)automatic ontology mapping (alignment) techniques, [35] a survey of schema matching approaches, [36] analyze both schema and ontology mappings, and [10] analyze existing mapping discovery solutions from a data integration perspective. Most of the approaches cited in these surveys are based on heuristic algorithms searching for lexical (e.g., using the Levenshtein distance [25]) and/or structural similarities between models (e.g., PROMPT [33]), or employ machine learning techniques to find mappings (e.g., GLUE [11]). Some approaches operate either on the schema level (e.g., Cupid [27]), on the instance level (e.g., SemInt [26]), or include both levels (e.g., COMA++ [4]) in order to (semi-)automatically discover mappings between schemes or ontologies. Recent developments (e.g., [46]) propose a public, community-driven approach for mapping discovery where end users, knowledge engineers, and developer communities take part in the process of establishing mappings.

According to [32], mappings can be represented using three different types of formalisms: (i) representing them as instances of a defined mapping model (e.g., MAFRA [28]), (ii) defining bridging axioms or rules to represent transformations, and (iii) using views to define mappings between a source and a target ontology (see [15]). In distributed environments, after being formally declared, mappings serve as input for a process commonly referred to as query reformulation.

5 Summary and Conclusion

The CIDOC CRM has gained great attention in the cultural domain as a possible means for establishing interoperability among heterogeneous metadata. A major limitation of the current standard is the lack of guidelines specifying how to employ the CRM model in real-world applications, i.e., how to integrate and map source-specific models with the CRM and how to process CRM metadata afterwards. Especially the high degree of freedom in creating mappings is likely to produce valid but divergent mapping chains for semantically corresponding source concepts.

We have pointed out the limitations of the CIDOC CRM and proposed *Model Implementation Guidelines* as an initial step to lead CRM-based systems towards interoperability. These guidelines are not meant to be automated mapping algorithms or ultimate solutions covering all possible mapping problems. They should rather be seen as an assistance for future CRM adopters that are likely to encounter similar problems when using the CIDOC CRM model.

So far we have applied them on the source-specific models that have been integrated in the BRICKS context. The application we have implemented on top of the integrated metadata demonstrates the applicability of the guidelines in the domain of archaeological findings. From the difficulties we have encountered during our work, we can conclude that global ontologies are well suited for an intellectual analysis of an application domain. However, for being implemented in practice, they require detailed guidelines and technical specifications.

Future work on this topic must include an evaluation across several application scenarios in the cultural heritage domain, which requires several participating institutions and domain experts familiar with the CRM concepts. Reusing the results of other related CRM projects is currently hardly possible because, as described in Section 2.3, each has applied the CRM differently. Further, since the guidelines rely on the knowledge of human experts, it requires application scenario independent visualization techniques and user interfaces that can handle the CRM concepts and mappings between them. Such interfaces are not available yet. With our work we intend to lead future integration projects, which use the CRM, into a common direction, which in turn can provide the basis for a broader evaluation.

References

[1] K. Aberer, P. Cudre-Mauroux, M. Hauswirth, and T. van Pelt. Gridvine: Building Internet-Scale Semantic Overlay Networks. In *International Semantic Web Conference (ISWC)*, volume 3298 of *LNCS*, pages 107–121, 2004.

[2] Matthew Addis, Mike J. Boniface, Simon Goodall, Paul Grimwood, Sanghee Kim, Paul H. Lewis, Kirk Martinez, and Alison Stevenson. SCULPTEUR: Towards a New Paradigm for Multimedia Museum Information Handling. In *International Semantic Web Conference*, pages 582–596, 2003.

[3] Yigal Arens, Chun-Nan Hsu, and Craig A. Knoblock. Query Processing in the SIMS Information Mediator. In Michael N. Huhns and Munindar P. Singh, editors, *Readings in Agents*, pages 82–90. Morgan Kaufmann, San Francisco, CA, USA, 1997.

[4] David Aumueller, Hong Hai Do, Sabine Massmann, and Erhard Rahm. Schema and Ontology Matching with COMA++. In Fatma Özcan, editor, *SIGMOD Conference*, pages 906–908. ACM, 2005.

[5] Ana B. Benitez, Di Zhong, Shih-Fu Chang, and John R. Smith. MPEG-7 MDS Content Description Tools and Applications. In *CAIP '01: Proceedings of the 9th International Conference on Computer Analysis of Images and Patterns*, pages 41–52, London, UK, 2001. Springer-Verlag.

[6] Lois Mai Chan and Marcia Lei Zeng. Metadata Interoperability and Standardization – A Study of Methodology Part I + II. *D-Lib Magazine*, 12(6), June 2006.

[7] Paul Cripps, Anne Greenhalgh, Dave Fellows, Keith May, and David Robinson. Ontological Modelling of the Work of the Centre for Archaeology. Technical report, September 2004.

[8] Nick Crofts, Martin Doerr, Tony Gill, Stephen Stead, and Matthew Stiff. Definition of the CIDOC Conceptual Reference Model, October 2006. `http://cidoc.ics.forth.gr/docs/cidoc_crm_version_4.2.1.pdf`.

[9] Stefan Decker, Michael Erdmann, Dieter Fensel, and Rudi Studer. Ontobroker: Ontology Based Access to Distributed and Semi-Structured Information. In *DS-8: Proceedings of the IFIP TC2/WG2.6 Eighth Working Conference on Database Semantics-Semantic Issues in Multimedia Systems*, pages 351–369, Deventer, The Netherlands, 1998. Kluwer, B.V.

[10] AnHai Doan and Alon Y. Halevy. Semantic-Integration Research in the Database Community. *AI Magazine*, 26(1):83–94, 2005.

[11] AnHai Doan, Jayant Madhavan, Pedro Domingos, and Alon Halevy. Learning to map between Ontologies on the Semantic Web. In *WWW '02: Proceedings of the 11th international conference on World Wide Web*, pages 662–673, New York, NY, USA, 2002. ACM Press.

[12] Martin Doerr. The CIDOC Conceptual Reference Module: An Ontological Approach to Semantic Interoperability of Metadata. *AI Magazine*, 24(3):75–92, 2003.

[13] Martin Doerr and Dolores Iorizzo. The dream of a global knowledge network—a new approach. *J. Comput. Cult. Herit.*, 1(1):1–23, 2008.

[14] Martin Doerr, Kurt Schaller, and Maria Theodoridou. Integration of Complementary Archaeological Sources. In *Computer Applications and Quantitative Methods in Archaeology Conference, CAA 2004*, April 2004.

[15] Alon Y. Halevy, Zachary G. Ives, Peter Mork, and Igor Tatarinov. Piazza: Data Management Infrastructure for Semantic Web Applications. In *WWW '03: Proceedings of the 12th international conference on World Wide Web*, pages 556–567, New York, NY, USA, 2003. ACM Press.

[16] Y. Halevy, G. Ives, Dan Suciu, and Igor Tatarinov. Schema Mediation For Large-Scale Semantic Data Sharing. *The VLDB Journal*, 14(1):68–83, 2005.

[17] Rachel Heery and Manjula Patel. Application Profiles: Mixing and Matching Metadata Schemas, September 2000. Available at: `http://www.ariadne.ac.uk/issue25/app-profiles/`.

[18] IFLA. Functional Requirements for Bibliographic Records, September 1997. Available at: `http://www.ifla.org/VII/s13/frbr/frbr.htm`.

[19] ISO/IEC JTC 1/SC 29. *MPEG-21 Multimedia Framework — ISO 21000-1-7:2003-2007*. International Standardizaton Organization (ISO), 2007.

[20] ISO/IEC JTC 1/SC 29. *MPEG-7 Multimedia Content Description Interface — ISO 15938-1-11:2002-2007*. International Standardizaton Organization (ISO), 2007.

[21] Yannis Kalfoglou and Marco Schorlemmer. Ontology Mapping: The State of the Art. *Knowl. Eng. Rev.*, 18(1):1–31, 2003.

[22] Haridimos Kondylakis, Martin Doerr, and Dimitris Plexousakis. Mapping Language for Information Integration. Technical Report 385, ICS-FORTH, December 2006.

[23] Robert Kummer. Integrating data from the Perseus Project and Arachne using the CIDOC CRM: An examination from a software developer's perspective. In *Exploring the Limits of Global Models for Integration and Use of Historical and Scientific Information-ICS Forth Workshop, Heraklion, Crete*, October 2006. `http://www.perseus.tufts.edu/~rokummer/KummerCIDOC2006.pdf`.

[24] Robert Kummer. Towards Semantic Interoperability of Cultural Information Systems — Making Ontologies Work. Master's thesis, 2007.

[25] V. I. Levenshtein. Binary Codes Capable of Correcting Deletions, Insertions and Reversals. *Soviet Physics Doklady*, 10, February 1966.

[26] Wen-Syan Li and Chris Clifton. Semint: a tool for identifying attribute correspondences in heterogeneous databases using neural networks. *Data Knowl. Eng.*, 33(1):49–84, 2000.

[27] Jayant Madhavan, Philip A. Bernstein, and Erhard Rahm. Generic Schema Matching with Cupid. In Peter M. G. Apers, Paolo Atzeni, Stefano Ceri, Stefano Paraboschi, Kotagiri Ramamohanarao, and Richard T. Snodgrass, editors, *VLDB*, pages 49–58. Morgan Kaufmann, 2001.

[28] Alexander Maedche, Boris Motik, Nuno Silva, and Raphael Volz. MAFRA - An Ontology Mapping Framework in the Semantic Web. In *Proceedings of the ECAI Workshop on Knowledge Transformation, Lyon, France, 2002*, 2002.

[29] Eduardo Mena, Arantza Illarramendi, Vipul Kashyap, and Amit P. Sheth. OBSERVER: An Approach for Query Processing in Global Information Systems Based on Interoperation Across Pre-Existing Ontologies. *Distrib. Parallel Databases*, 8(2):223–271, 2000.

[30] Wolfgang Nejdl, Boris Wolf, Changtao Qu, Stefan Decker, Michael Sintek, Ambjörn Naeve, Mikael Nilsson, Matthias Palmér, and Tore Risch. EDUTELLA: A P2P Networking Infrastructure Based on RDF. In *WWW '02: Proceedings of the 11th International Conference on World Wide Web*, pages 604–615, New York, NY, USA, 2002. ACM Press.

[31] Ian Niles and Adam Pease. Towards a Standard Upper Ontology. In *FOIS '01: Proceedings of the international conference on Formal Ontology in Information Systems*, pages 2–9, New York, NY, USA, 2001. ACM Press.

[32] Natalya F. Noy. Semantic Integration: A Survey Of Ontology-Based Approaches. *SIGMOD Rec.*, 33(4):65–70, 2004.

[33] Natalya F. Noy and Mark A. Musen. The PROMPT Suite: Interactive Tools for Ontology Merging and Mapping. *Int. J. Hum.-Comput. Stud.*, 59(6):983–1024, 2003.

[34] Stephen T. Polyak, Jintae Lee, and et al. Applying the Process Interchange Format (PIF) to a Supply Chain Process Interoperability Scenario. In *Workshop on Applications of Ontologies and Problem Solving Methods, ECAI'98*, Brighton, England, 1998.

[35] Erhard Rahm and Philip A. Bernstein. A Survey of Approaches to Automatic Schema Matching. *The VLDB Journal*, 10(4):334–350, 2001.

[36] Pavel Shvaiko and Jérôme Euzenat. A Survey of Schema-Based Matching Approaches. *Journal of Data Semantics*, 3730:146–171, 2005.

[37] P. Sinclair, S. Goodall, P.H. Lewis, K. Martinez, and M. J. Addis. Concept browsing for multimedia retrieval in the SCULPTEUR project. In *Proceedings of The 2nd Annual European Semantic Web Conference*, 2005.

[38] Patrick Sinclair, Paul Lewis, Kirk Martinez, Matthew Addis, Adrian Pillinger, and Daniel Prideaux. eCHASE: Exploiting Cultural Heritage using the Semantic Web. 2005.

[39] R. Stein, J. Gottschewski, R. Heuchert, A. Ermert, M. Hagedorn-Saupe, H.-J. Hansen, C. Saro, R. Scheffel, and G. Schulte-Dornberg. *Das CIDOC Conceptual Reference Model: Eine Hilfe fuer den Datenaustausch?* Staatliche Museen zu Berlin, 2005.

[40] Barbara Tillett. What is FRBR — A Conceptual Model for the Bibliographic Universe, 2004. Available at: `http://www.loc.gov/cds/FRBR.html`.

[41] Andre Valente, Thomas Russ, Robert MacGregor, and William Swartout. Building and (Re)Using an Ontology of Air Campaign Planning. *IEEE Intelligent Systems*, 14(1):27–36, 1999.

[42] H. Wache, T. Vögele, U. Visser, H. Stuckenschmidt, G. Schuster, H. Neumann, and S. Hübner. Ontology-Based Integration of Information — A Survey of Existing Approaches. In H. Stuckenschmidt, editor, *IJCAI–01 Workshop: Ontologies and Information Sharing*, pages 108–117, 2001.

[43] Holger Wache. *Semantische Mediation für heterogene Informationsquellen.* PhD thesis, University of Bremen, 2003.

[44] Gio Wiederhold. Mediators in the Architecture of Future Information Systems. *Computer*, 25(3):38–49, 1992.

[45] WonderWeb Consortium. DOLCE: A Descriptive Ontology for Linguistic and Cognitive Engineering, 2003. Available at: `http://www.loa-cnr.it/DOLCE.html`.

[46] Anna V. Zhdanova and Pavel Shvaiko. Community-Driven Ontology Matching. In York Sure and John Domingue, editors, *ESWC*, volume 4011 of *Lecture Notes in Computer Science*, pages 34–49, Berlin, Heidelberg, 2006. Springer.