

Short Documentation of the CIDOC CRM (4.2.4) Implementation in OWL-DL

Martin Oischinger, Bernhard Schiemann, Guenther Goerz
C.S.D., University of Erlangen-Nuremberg

May 10, 2008

The Erlangen OWL-DL implementation of the CIDOC CRM (version 4.2.4) has been designed as close as possible to the specifications in the CRM document. Therefore, the CRM document should serve as the primary reference for the implementation. Nevertheless, there are some features which could not be implemented as described or have not been implemented for certain reasons. This document describes just these differences by quoting the respective text sections and explaining why they could not be implemented as described. Quotations from “Scope Notes” in the CRM document are typeset in an *italic font*.

If CRM entities and properties have — to our best knowledge — been implemented conforming to the specification in the CRM document, they are not mentioned here. Inverse properties carry an “I” in its name, e.g., the inverse property of `P5.consists_of` is `P5I.forms_part_of`.

In the CRM document (ver. 4.2.4), the class **E55 Type** is described as a metaclass. For the sake of decidability, OWL-DL does not provide means to represent metaclasses. Therefore, **E55 Type** has been implemented as a class which — for the purpose of reasoning on the conceptual level — may serve as an interface to external concepts of formal domain ontologies (or thesauri) as subclasses or as constants (“individuals” in terms of the “one-of” language construct, i.e. an enumeration datatype). In the latter case, of course, the constants cannot have instances in turn. Because of this difference to the CRM document, the special properties with “.1” in its name have not yet been implemented; e.g., *P16.1 mode of use: E55 Type* connects an **E7 Activity** with an **E55 Type** to express a more detailed description of some other property.

Remark: The usual way to attach concepts of a domain ontology to the CRM is direct subclassing, e.g. the class “Artist” as an **E21 Person**; in that case representing it as a subclass of **E55 Type** would lead to contradictions. Instead, a constant “Artist” may be used, but it is up to the user to guarantee for semantic integrity.

In general, all “short cuts” described in the CRM document have not (yet) been included because there is no unique meaning of these abbreviations. E.g., the definition of **E4 Period** says: *P8 took place on or within (witnessed) is a short-cut of a path defining a E53 Place with respect to the geometry of an object. cf. E46 Section Definition*. Obviously, the property *P8 took place on or within* is used as a placeholder for the connection between **E4 Period** and the description of a geometry for an **E53 Place**. Unfortunately, there is no definition of a geometry at **E53 Place** such that it remains unclear how it should be implemented. Another kind is described in the

Scope Note of **E36 Visual Item**: *The property P62 depicts (is depicted by) between E24 Physical Man-Made Thing and depicted subjects (E1 CRM Entity) can be regarded as a short-cut of the more fully developed path from E24 Physical Man-Made Thing through P65 shows visual item (is shown by), E36 Visual Item, P138 represents (has representation) to E1 CRM Entity, which in addition captures the optical features of the depiction.* This definition describes a sequence of properties between three different CRM classes: the representing media (e.g. photographic paper), the visual item (e.g. a photograph itself) and the concept which represents the depicted object. If only the property **P62 depicts** would be used, it would have mutable semantics which could not be determined from the property itself but from the connected concepts. In favor of clear semantics the whole path should be composed of different properties. Both examples show that CRM short-cuts describe quite different abbreviations and it remains to be clarified how they should be modelled.

Whatever is underspecified or unspecified in the CRM document has been left open in the OWL-DL implementation as well. E.g., the definition of **P48 has preferred identifier (is preferred identifier of)** says: *Use of this property requires an external mechanism for assigning temporal validity to the respective CRM instance.* What this external mechanism would be remains unspecified; therefore, the OWL-DL implementation leaves the matter open, too. If there are exceptions, they are described explicitly in the following.

Many Scope Notes contain comments related to the application of the CRM, in many cases by giving recommendations or referring to best practice in documentation. Comments pertaining to the use of the CRM do not affect its implementation directly, but could give reason for (optional) constraints in future versions.

The following list enumerates those classes and properties which have not been implemented as described in the resp. Scope Notes; it obeys the order given in the CRM document.

E1 CRM Entity. P1 is identified by (identifies): *The property does not reveal anything about when, where and by whom this identifier was used. A more detailed representation can be made using the fully developed (i.e. indirect) path through E15 Identifier Assignment.* If **P48 has preferred identifier** is used — the definition of **E1 CRM Entity** enforces at most one of it —, **P1 is identified by** holds automatically, because **P48 has preferred identifier** is a subproperty of **P1 is identified by** in the property hierarchy.

P1 is identified by (identifies): *This property includes in particular identification by mathematical expressions such as coordinate systems used for the identification of instances of E53 Place.* This definition has not been implemented because **E53 Place** does not (yet) contain a property which includes mathematical expressions.

E2 Temporal Entity. P114 is equal in time to / P115 finishes (is finished by) / P116 starts (is started by) / P117 occurs during (includes) / P119 meets in time with (is met in time by) / P120 occurs before (occurs after): *This property is only necessary if the time span is unknown (otherwise the equivalence can be calculated).* The CRM document requires that these properties must be used if the time span is unknown. This condition cannot be verified.

E3 Condition State. *In general, the time-span for which a certain condition can be asserted may be shorter than the real time-span, for which this condition held.* The vague time specification (“may

be shorter”) does not allow to commit whether the **E3 Condition State** or the associated **E52 Time Span** is shorter.

The nature of that condition can be described using P2 has type. Due to the revised definition of **E55 Type** as a regular class (see below), **P2 has Type** is a “normal” property.

E4 Period. P10 falls within (contains): *The difference with P9 consists of (forms part of) is subtle. Unlike P9 consists of (forms part of), P10 falls within (contains) does not imply any logical connection between the two periods and it may refer to a period of a completely different type.* This property has been implemented as a regular OWL-DL object property which maps the logical relation between two **E4 Periods**.

E6 Destruction. P13 destroyed (was destroyed by): *Destruction implies the end of an item’s life as a subject of cultural documentation - the physical matter of which the item was composed may in fact continue to exist. A destruction event may be contiguous with a Production that brings into existence a derived object composed partly of matter from the destroyed object.* This definition has been implemented only partially: The properties of matter are not further specified and hence cannot be properly implemented. An **E81 Transformation** contains in turn an **E64 End of Existence**, such that the second part of the Scope Notes has been implemented by means of the concept **E81 Transformation**.

E7 Activity. P20 had specific purpose (was purpose of): *P20 had specific purpose (was purpose of) does not imply that an activity succeeded in achieving its aims. For example, dubious accounting practices may be carried out with the specific purpose of enhancing share values and enabling a take-over bid. The specific purpose remains the same even if the strategy fails and the company goes bankrupt instead.* and **P21 had general purpose (was purpose of):** *E7 Activity does not imply that an activity succeeds in achieving its general aims.* Both properties are described with an unspecified range, because the concept of “Purpose” is not defined in the CRM document. Therefore, their range has been defined to be **E1 CRM Entity** or **E62 String**.

P32 used general technique (was technique of): *Specific techniques may be further described as instances of E29 Design or Procedure.* and **P33 used specific technique (was used by):** *The property differs from P32 used general technique (was technique of) in that the E29 Design or Procedure referred to is specific and documented rather than simply being a term in the E55 Type hierarchy.* Because the description of **P32 used general technique** refers to **E55 Type**, it describes an unspecified relation between **E55 Type** and **E29 Design or Procedure**. The same holds for **P33 used specific technique**.

E11 Modification. P31 has modified (was modified by): *If a modification is applied to a non-made object, it is regarded as an E22 Man-Made Object from that time onwards.* This property describes a transition between two concept affiliations of an instance. Such a transition — first being an A and then becoming a B — cannot be represented in OWL-DL.

E18 Physical Thing. P51 has former or current owner (is former or current owner of): *The distinction with P52 has current owner (is current owner of) is that P51 has former or current owner (is former or current owner of) does not indicate whether the specified owners are current.* **P51 has former or current owner** has been implemented as a superproperty of **P52 has current owner**. In OWL-DL, only domain, range and subproperty expressions of properties can be

defined. Therefore, the feature to enforce the owner to be “actual” cannot be associated with the property.

E60 Number. *Identifiers in continua may be combined with numbers expressing distances to yield new identifiers, e.g., 1924-01-31 + 2 days = 1924-02-02.* The implementation of **E60 Number** defines the concept by means of integer and float numbers. Therefore, arithmetic expressions cannot be instances of **E60 Number**; arithmetic expressions cannot be evaluated and it is not possible in a logical language without equality to express “ $3+1 = 4$ ”. If there is a need to deal with arithmetic expressions a different representation must be used, e.g. **E73 Information Object**.

E61 Time Primitive. The definition of **E61 Time Primitive** contains the example 1994–1997, which cannot be represented with the chosen datatype (<http://www.w3.org/2001/XMLSchema#time>).

E81 Transformation. **P123 resulted in (resulted from):** *The physical continuity between the old and the new is expressed by the link to the common Transformation.* In OWL-DL, only domain, range and subproperty expressions of properties can be defined. Therefore, the feature to preserve “physical continuity” cannot be associated with the property.

P124 transformed (was transformed by) *It is replaced by the result of the Transformation, which becomes a new unit of documentation. The continuity between both items, the new and the old, is expressed by the link to the common Transformation.* This definition cannot be expressed in OWL-DL, because OWL-DL does not dispose of replacement rules (new for old). A new **E77 Persistent Item** (with **E63 Beginning of Existence**) comes up and an old **E77 Persistent Item** ceases to exist (with **E64 End of Existence**). In the OWL-DL implementation, both instances (of **E63 Beginning of Existence** and **E64 End of Existence**) are still present after the transformation.