# Co-reference in CIDOC-CRM and the semantic web

*Øyvind Eide, with contributions from Martin Doerr and Arianna Ciula*

In the scope notes for E91 Co-reference Assignment in CIDOC-CRM the following is found:

The use of *P155 has co-reference target* is limited to entities within the knowledge base in which the E91 Co-Reference Assignment is found. This is because the E91 Co-Reference Assignment is making explicit the worldview of the E39 Actor carrying out the assignment and this world view is expressed as such only within that specific knowledge base. For a further discussion of the relationship from E91 Co-Reference Assignment via P155 has co-reference target to E1 CRM Entity, see the technical paper … .

This document is a draft for the technical paper describing the rationale behind the short text in the scope note in some more detail. While understanding this solution is not that difficult it adds a level of complexity to what was planned to be a simple event for adding explicit co-reference. The co-reference target is needed within CIDOC-CRM, but we will try to limit the complexity behind it as much as possible. This is a pedagogical issue and a standard writing issue.

Co-reference is an inherent in all human communication. We use the fact that names and other references to persons, places and other entities co-refer every day. The E91 Co-reference Assignment is a mechanism to make statements about such co-references explicit in an information system. It aims mainly to document author intention in historical documents, specifically the problem of interpreting to what particular entity a name, pronoun or equivalent kind of expression was intended to refer within a particular passage.

In other words, a co-reference assignment expresses the interpretation of the creator of the assertion about the meaning of information provided by another person, that is, the original author of the document. Such an interpretation can only be documented with respect to another reference. This can be found in another text by the same author or in a text by another author. It can also be found by referring to the world known to the creator of the co-reference assignment herself. To do the latter, the property P155 has co-reference target (is co-reference target of) allows for referring to an instance of CRM Entity of the creator’s world.

In a sense, an instance of E91 Co-Reference Assignment using the property P155 has co-reference target (is co-reference target of) in a knowledge base forms a propositional object referring to the creator’s target entity. After all, a knowledge base as a whole can be seen as a propositional object. Consequently, if in a Semantic Web implementation the target entity is instantiated by a URI, the meaning of this identifier must be unambiguous to the creator of the co-reference assignment. Similarly, a URI of another authority, such as an author catalogue of a library, can be interpreted as a referring proposition of this catalogue, and be referred to by the property P153 assigned co-reference to (was regarded to co-refer by) or P154 assigned non co-reference to (was regarded not to co-refer by): E89 Propositional Object in order to express that it does not immediately represent the creator’s known world. In this case, the authority that knows the meaning of this URI must be unambiguous by the form of the URI itself. Such a URI of an external authority can never be referred to using the property P155 has co-reference target (is co-reference target of).

In contrast, the meaning of the property ‘owl:same\_as’ of the OWL knowledge representation language cannot specify whose knowledge it represents and cannot express propositional attitudes. Therefore it is not adequate to model the progress of scholarly co-reference research .

## My man is your URI: the game of make-believe

In the co-reference assignment event we have pointers to the propositional objects co-referring (or not, in the case of negative co-reference). That is fine, no problem. The propositional objects are anything which can refer and thus, they are information objects living happily without our documentation system.

The co-reference target, however, is a tricky animal. In the context of a model it is easy. We have Hans Jæger on the title page of a book and we have Hans Jæger as the motive of Munch’s painting. These two signs both refer to the same entity. They are different signs, but in ordinary human communication (and in museum documentation) it is obvious that they refer to the same entity. That entity is the now dead person we refer to by the name Hans Jæger, which was given to him in a baptising event and was constantly re-established in the social environment in which he lived. This re-establishment still goes on; when I talk about him to knowledgeable people they know to whom I refer even if they are born long after he died.

In my information system, which we for simplicity see as a straightforward material implementation of CIDOC-CRM, I include an E21 Person. This person is identified by a link to an E82 Actor Appellation, which has a string attached to it: “Hans Jæger”. This name in itself does not identify the person, but other things (such as birth and death dates, links to works he created, etc.) makes it clear to any knowledgeable human who this person was. The computer does not and cannot know or care; physical persons is out of scope for the computer, it only manipulates information objects.

So this person entity is the target of the co-reference. All clear and nice. An event pointing to a historical person and to two or more information objects referring to him. Clear and nice except for the fact that it is not. This is just a game. There is no dead person in my information system. There are only information objects. Hans Jæger is a box on a whiteboard, or a URI.

This is fine within the information system. The problem is the moment I integrate different information systems. The co-reference target can only be an entity within the local information system. While the co-referring entities can be anywhere, the target must be local. An entity—also a Person entity—is an information object seen from the outside. The game is broken when we leave our playground.

This implies that if one wishes to use the P155 has co-reference target (is co-reference target of) property to link to a target only available outside the information system of the co-reference assignment the only way of doing it is to import (or re-create) this target entity inside the local information system.

## Co-reference example

The problem outlined above and the solution we have chosen will here be illustrated by an example. If we take these three sentences about one well known historical person:

Then they all stabbed Caesar.

The first emperor is claimed to be Julius, but…

The consul went to…

The three red things from three different sources (information systems, models) all co-refer. This is not a problem. As was pointed out already in Ciula et.al. (2008) such references taken from documents only gives us the evidence that this person was referred to in texts, not additional facts about his life, such as date of birth or family relationships. Such information can be added, for instance, as authority data linked to a person element in the information system.

Then we make a co-reference assignment based on my understanding of the historical facts:

E91 Co-Reference Assignment

P14 carried out by E39 Actor (ØE)

P153 assigned co-reference to: E89 Propositional Object (Caesar)

P153 assigned co-reference to: E89 Propositional Object (co-refers: Julius)

P153 assigned co-reference to: E89 Propositional Object (co-refers: The consul)

This all operates at information object (sign) level.

Also: if we add a note “They co-refer because they all refer to a historical person who…” it is still fine, as this is another information object.

The problem is the statement in blue:

E91 Co-Reference Assignment

P14 carried out by E21 Person (ØE)

P153 assigned co-reference to: E89 Propositional Object (Caesar)

P153 assigned co-reference to: E89 Propositional Object (co-refers: Julius)

P153 assigned co-reference to: E89 Propositional Object (co-refers: The consul)

P155 has co-reference target: E21 Person: (Historical Person Caesar)

This is when we enter a link to an entity which can only live within the same information system as the co-reference statement. The reason for this is that "Historical Person Caesar" is not a link to the historical person but a representation, a stand-in, for this person in the system. The entity is not referring to the person but is included in the system as if it was the person.

# References

Ciula, Arianna, Paul Spence, and José Miguel Vieira. "Expressing complex associations in medieval historical documents: the Henry III Fine Rolls Project." *Literary and Linguistic Computing* 23, no. 3 (2008): 311–25.