

A data model for Co-Reference

Some notes towards a standard
Presented in June 2013 in Stockholm
Øyvind Eide, University of Oslo

Outline

- Refresh what co-reference is
- Outline a core model for explicit co-reference statements
- Suggest how the core model can be connected to CIDOC-CRM
- Includes a few question marks
- May be a possible starting point for developing a standard

What is co-reference?

- “The table by the window was...”

- “The beautiful table could be...”



What is co-reference

- The fact that two or more signs in a context and a reading refer to the same entity
- Exist in most if not all documents
- Co-references exist independently of our information systems
- Based on an identity definition for the types of entities in question

Co-reference, strings, and meaning

- The more similar two strings are, the more likely to be a co-reference
- ...but string similarity cannot be used as evidence
- Co-reference (as reference) must be identified by the meaning (extension) of an expression
- Thus, based on meaning, interpretation, context

Semiotics

- Co-reference works across all sign types (symbolic, iconic, indexical)
- Can be intentional or not. This has no practical consequences, but may be used in detection

Co-reference statements

- Implicit co-references exist in the world and can be detected
- This is not enough, we need explicit co-reference statements. They should include:
 - Two entities which can refer
 - The date of the statement
 - The responsible entity (person, software, etc.)
 - A third entity which is the target they refer to

Core outline for a model

C1 Co-Reference Assignment

- D1 assigned co-reference to C2 Referring Entity
- D2 reference target C3 Entity
- D3 carried out by C4 Actor
- D4 happened at time C5 Time-Span
- D5 has type C6 Type

Cn: co-reference entities

Dn: co-reference properties.

All elements are linkable (identified by URI)

C1 Co-Reference Assignment

- An event happening at specific time in history
- Establishes explicit information
- Do we need to record the place it happened?

C2 Referring Entity

- This can be anything that can refer
- A string in a text: “Paris”
- Depending on context and reading: a place, a person, a restaurant, a concept (mythical person)
- The co-reference statement fixes one interpretation, connected to a responsible person

C3 Entity

- This is the target of the references
- It can be anything a human can think about
- Examples: places, persons, dreams, ideas, love, the feeling I had when I saw the trees in the sun this morning

C4 Actor

- The responsible entity behind the co-reference statement
- Can be a human being
- Can be using a tool, such as a computer program, in the process

C5 Time-Span

- When the co-reference statement was made
- Straight forward

C6 Type

- Any kind of useful classification
- Should we also have a free text description?

D1 assigned co-reference

C1 Co-Reference Assignment D1 assigned co-reference C2 Referring Entity

- Need two or more of those for each co-reference assignment

D2 reference target

C1 Co-Reference Assignment D2 reference target C3 Entity

- This is the target of the references
- Should be included, but can be omitted?

D3 carried out by

C1 Co-Reference Assignment D3 carried out by C4 Actor

- Establishes the responsibility link.
- Vital to store this information, otherwise no system of trust can be built

D4 happened at time

C1 Co-reference Assignment D4 happened at time C5 Time-Span

- Must record when it happened
- Can be imprecise

D5 has type

C1 Co-Reference Assignment D5 has type C6 Type

- Connects a type to the co-reference
- Example: inferred

Example 1

Stating that Stockholm in two documents point to the same place:

C1 Co-Reference Assignment

- D1 assigned co-reference to C2 Stockholm (a part of <http://www.riksarkivet.se>)
- D1 assigned co-reference to C2 Stockholm (the topic of <http://en.wikipedia.org/wiki/Stockholm>)
- D2 reference target C3 "Stockholm" (the place in the real world)
- D3 carried out by C4 "Øyvind Eide"
- D4 happened at time C5 "2013-06-03"

Example 2

Stating that Paris in one source is co-referring with Lutetia in another source:

C1 Co-Reference Assignment

- D1 assigned co-reference to C2 Lutetia (a part of an Asterix album)
- D1 assigned co-reference to C2 Paris (a part of an Michelin Atlas)
- D2 reference target C3 "Paris" (the concept in human conceptualisation and communication)
- D3 carried out by C4 "Øyvind Eide"
- D4 happened at time C5 "2013-05-31"

Co-reference in CRM

- Co-references exist independently of our information systems
- Co-reference is modelled in CRM already.
 - Example: E48 Paris P87 identifies E53 Place A
 - E48 Lutetia P87 identifies E53 Place A

If Place A in the two examples is the same, then E48 Paris co-refers with E48 Lutetia in the context of a specific reading of the CRM model.

Co-reference statements and CRM 1

- **C1 Co-Reference Assignment** is a subclass of **E13 Attribute Assignment**
- **C2 Referring Entity** is **E89 Propositional Object** (to be discussed: precise placement in hierarchy)
- **C3 Entity** is **E1 Entity**
- **C4 Actor** may be an **E39 Actor**
- **C5 Time-Span** is **E52 Time-Span**
- **C6 Type** is **E55 Type**

Co-reference statements and CRM 2

- **D1 assigned co-reference to** is a subproperty of **P140 assigned attribute to**
- **D2 reference target** is **P67 refers to**
- **D3 carried out by** is **P14 carried out by**
- **D4 happened at time** is **P4 has time-span**
- **D5 has type** is **P2 has type**

Negative co-reference

- Is this just the inversion of a Co-reference assignment?
- Thus: **C7 Negative Co-reference Assignment** with the same properties?

Services

- Which services are needed?
- Should we have general services in the model?
- Or should it be implementation dependent?

This will be the topic of the next session

...but basically it is quite simple

C1 Co-reference Assignment

- D1 assigned co-reference to C2 Referring entity
- D2 reference target C3 Entity
- D3 carried out by C4 Actor
- D4 happened at time C5 Time-Span

Thank you!