

Report on the NEDIMAH WG 3 meeting, Stockholm 3-4 June 2013

Christian-Emil Ore, University of Oslo, co-chair NEDIMAH WG 3
01-12-2013

1 Short summary

The workshop ran over two days. In the first day we focused on the CultureCloud initiative and museums and archives and on how core ontologies like CIDOC-CRM and FRBRoo can be used as a common vocabulary for discourse in museums, libraries and archives. The first day was attended by the participants in the K-samsøk group and members of the staff of the Swedish National Archives (ca 15 extras). The second day focused on theoretical and practical topics connected to linked data and the problem of co-reference. The workshop was co-organized with two of the WG of ICOM-CIDOC: CRM-SIG and Co-reference working group. The meeting of these two groups continued on Jun 5th to June 7th.

2 Background for the workshop

The WG 3 of the NEDIMAH ESF network deal with ontological methods and linked data in the field of Digital Humanities. These subjects can be somewhat difficult to delimit. On the first day of the workshop we decided to focus on two core ontologies for memory institutions, CIDOC CRM and FRBRoo and their applications. On the second day the focus was on problems and challenges connected to linked data. We especially wanted to focus on the so called co-referential problem.

2.1 CultureCloud and ontological methods for linking data in archives and museums

CultureCloud is an open initiative developed over the last 6 years with the objectives to enable collaboration and interlinking of cultural heritage data on all levels from the small village museum to the national and international level. This is done by designing a generic; standards based information architecture and specifying a set of tools for sharing, linking and enriching cultural heritage content. The basic interlinking mechanisms underlying the Semantic Web or portals like the Europeana are necessary to link data about our cultural heritage. However, these mechanisms are in themselves not sufficient to establish a cultural heritage information space as formulated in for example the strategic documents of Europeana. The reason is that in most cases the metadata in the memory institutions is not collected for the purpose of interlinking.

2.2 Co-reference and linked data.

The origin of the concept of co-reference is in linguistics but subsequently its scope has broadened. A variety of representations can co-refer to the same object or topic. Some authors talk about co-reference chains. The term co-reference networks may be preferred to denote co-referring representations. Within these networks we may have transitivity. In this case it is warranted to speak about co-reference chains.

Co-reference networks can be modelled for information systems. This is where a CIDOC CRM compatible schema becomes a necessity. Identifier creation and management is necessary for the handling of co-reference networks.

Co-reference networks can be expressed as Linked Data. Languages are formed of nested structures. In natural language phonology, morphology, syntax, semantics and pragmatics form complementary layers. It may be beneficial to view Linked Data in a similar manner. The mere creation of a Linked Data mesh is useful but insufficient to serve communication about cultural heritage.

Theoretical work as well as experimentation is required. The following questions need to be clarified:

1. the relation of Linked Data to rich representations and
2. the creation expressions that carry the intent of a speaker.

Linked Data is just an apparently infinite but in fact a finite number of expressions that have been formed out of data sets and other information resources by the use of triples. The Linked Data cloud carries information in the sense that it reduces the entropy of the source. However, there is no receiver that can grasp the information content of the totality of this information source. There is some relief to this in that each query forms a projection over this near-infinite cloud and thus results in a selection. One might say that we interact with Linked Data in an oracle mode, where the Linked Data oracle provides us with seemingly rational but actually fragmentary and often irrelevant if not irrational statements.

Information systems may indicate what relations are coupled to some item. These systems usually do not indicate what explicit links are missing. If we would have a well defined, domain specific schema in the background we might ask, what is typically said about this type of objects? In that case a linked data structure might indicate what knowledge we still are lacking about some specific object, in addition to what is currently known about it.

These are only random examples. An additional layer of distinctions has to be formed on top of the Linked Data layer. The reason for why very little thought is given to this, is that development is currently very much technology driven. It has been easier for people in cultural heritage organizations to jump on the Linked Data bandwagon instead of trying to grasp its theoretical underpinnings. We are convinced that such theories exist, but they are not widely known or accepted. Additionally, Linked Data produces something tangible, although of limited use for the above mentioned reasons. It seems so advanced that little attention is paid to that in its current phase it is quite useless. So this is a challenge we have to face.

3 Presenters

Peder Andrén, Swedish National Archives, Sweden

Zhanna Belik, Andrey Rublyev Museum, Moscow, Russia

John Bradley, King College, London, UK
Martin Doerr, FORTH, Heraclion, Greece
Øyvind Eide, University of Oslo, Norway
Toivo Flink, University of Jyväskylä, Finland
Heikki Hanka, University of Jyväskylä, Finland
Faith Lawrence, King College, London, UK
Mika Nyman, Synapse Computing Oy, Finland
Christian-Emil Ore, University of Oslo, Norway
Stephen Stead, PavePrime Ltd, UK
Mikael Vakkari, Finland
Thomas Wikman, Delving B.V., Sweden/Netherlands

4 List of participants

Peder Andrén, Swedish National Archives, Sweden
Zhanna Belik, Andrey Rublyev Museum, Moscow, Russia
Chrysoula Bekiari, Forth, Heraklion, Greece
John Bradley, King College, London, UK
Martin Doerr, Forth, Heraklion, Greece
Øyvind Eide, University of Oslo, Norway *
Toivo Flink, University of Jyväskylä, Finland
Heikki Hanka, University of Jyväskylä, Finland *
Börje Justrell, Swedish National Archives, Sweden
Rolf Källman, Swedish National Archives, Sweden
Faith Lawrence, King College, London, UK, *
Mika Nyman, Synapse Computing Oy, Finland,
Christian-Emil Ore, University of Oslo, Norway *
Sebastian Rahtz, University of Oxford, UK *
Stephen Stead, PavePrime Ltd, UK
Mikael Vakkari, Finland
Thomas Wikman, Delving B.V., Sweden/Netherlands

Additional participants from Sweden K-samsøk (Swedish common search in museums, archives and libraries) and National Archives, ca 15 on Monday.

Members of NEDIMAH WG3 marked with * and WG1 **.

5 Participants supported by the ESF grant:

Zhanna Belik, Russia
John Bradley, UK
Øyvind Eide, Norway
Toivo Flink, Finland
Faith Lawrence, UK,
Mika Nyman, Finland,
Christian-Emil Ore, Norway,
Sebastian Rahtz, UK



Riksarkivet



EUROPEAN
SCIENCE
FOUNDATION



NEDIMAH WG 3, Culturecloud, Co-reference, Archive workshop and CRM/FRBROO SIG week

Hosted by DIGISAM (www.digisam.se)
and Riksarkivet (www.ra.se)

3-7 June, National Archives, Stockholm

Agenda

Monday 3/6 (Lecture room): NEDIMAH/CultureCloud/Ontologies

09:00	Coffee & registration
09:30	Welcome, Börje Justrell, Riksarkivet, Rolf Källman, DIGISAM
09:45	Introduction to NEDIMAH and WG3, Christian-Emil Ore
10:00	CultureCloud, overview, Christian-Emil Ore, Thomas Wikman
11:00	Core standards: CIDOC CRM, Martin Doerr
11:30	Core standards: FRBROO, Martin Doerr
12:00	Core standards: APEX EAC-CPF etc, Karin Bredenberg
12:30	Lunch
13:30	<ul style="list-style-type: none"> • 13:30: Co-reference (LoD), Mika Nyman • 14:00: Aggregation, mapping Martin Doerr • 14:30: Training and CIDOC Summer Schools, Stephen Stead • 15:00: NFDLL: New Forms of Collaboration within the Digital Learning Landscape, Ambjörn Naeve
15:30	Coffee
15:45	<ul style="list-style-type: none"> • Current status of CultureCloud implementations, Thomas Wikman • Possible collaborations (archives, libraries, museums) • What's next
17:00	End
19:00	Dinner

Tuesday 4/6: NEDIMAH/CRM and Co-referencing

09:00	Registration		
	CRM (Lecture room)		Co-reference (Seminar room)

09:15	CIDOC CRM Workshop, Stephen Stead	09:15	TOWARDS A SHARED UNDERSTANDING OF CO-REFERENCE <ul style="list-style-type: none"> ● Introduction to Co-reference and the Co-reference Working Group ● Introduction to NEDIMAH and WG3, Christian-Emil Ore ● Reference and representation, Mika Nyman
10:15	Coffee	10:15	Coffee
10:30	CRM Workshop, Stephen Stead	10:30	CO-REFERENCE ASPECTS IN PROJECTS <ul style="list-style-type: none"> ● Linked data and digital narrative, Faith Lawrence ● Co-reference and prosopography, John Bradley ● Iconrussia, Zhanna Belik, Comment: Heikki Hanka ● Presentations: Ingria archival prototype, Toivo Flink
12:30	Lunch	12:30	Lunch
14:00		14:00	FORMALIZATION <ul style="list-style-type: none"> ● Co-reference and networks of identity, Martin Doerr ● Pre-study for a Co-reference schema compatible with CIDOC-CRM, Øyvind Eide, Mika Nyman
15:30		15:30	Coffee
15:50		15:50	INFRASTRUCTURES Thomas Wikman: <ul style="list-style-type: none"> ● Services crossing project cycles, institutional and national borders ● X-Road (Estonia), Svensk Nationell Datatjänst / Swedish National Data Service ● Closing discussion and way forward
17:00		17:00	End

Wednesday 5/6: Archive workshop (Seminar room)

09:00	Welcome, Björn Jordell, General Director Riksarkivet
09:15	Introduction to the Ådalen -31 and Ingermanland content for today's workshop, Peder Andrén, Mika Nyman & Toivo Flink
10:00	Integrating, modelling and mapping archival, museums and library material, Martin Doerr
10:30	Coffee
11:00	Integrated access to archive material supporting research processes, Martin Doerr
11:30	Practical mapping
12:30	Lunch
13:30	Practical mapping cont.
14:30	Coffee
14:45	Practical mapping cont.
16:00	Summary. What have we learned and how do we proceed

17:00	End
-------	-----

Thursday 6/6 and Friday 7/6 CIDOC CRM/FRBRoo

09:00	<ul style="list-style-type: none"> ● Mapping Reference Model ● Spatial Model ● meta CRM
12:30	Lunch
13:30	<ul style="list-style-type: none"> ● Mapping Reference Model ● Spatial Model ● meta CRM
14:30	Press oo Europeana
14:45	<ul style="list-style-type: none"> ● CIDOC CRM Issues ● FRBRoo Issues
17:00	End

Thursday 6/6 and Friday 7/6 CIDOC CRM/FRBRoo

09:00	FRBRoo Issues
12:30	Lunch
13:30	Discussions, closing meeting
17:00	End

Dinner on Monday and lunches at own expense, except Thursday.