9th CIDOC CRM Special Interest Group Meeting

Venue: ICS Forth

Wednesday 21st April

Present

Jon Holmen
Matthew Stiff   matthew.stiff@english-heritage.org.uk
Martin Doerr   martin@ics.forth.gr
Jen-Shin Hong   jshong@csie.ncnu.edu.tw
Juha Inkhari   inkari@iki.fi
Juha Kotipelto   Juha.Kotipelto@fng.fi
Kari Peiponen
Stefano Mazzocchi   stefanom@mit.edu
Patrick Sinclair   pass99r@ecs.soton.ac.uk
Athanasis Nikolaos   athanasi@ics.forth.gr
Maria Theadoridou   maria@ics.forth.gr
Christos Georgis   georgis@ics.forth.gr
Juha Makkonen   jamakkon@cs.Helsinki.FI
Tony Gill   tg@artstor.org
Nick Crofts   nicholas.crofts@etat.ge.ch
Siegfried Krause   s.krause@gnm.de
Patrick Le Boeuf   patrick.le-boeuf@bnf.fr
Tyler Bell   tyler.bell@oxarchdigital.com
Stein Olsen

14:30 Discussion session

Issues for discussion:
Following the presentation made Tuesday 20 April – Wednesday 21 April (see separate programme) the workshop entered discussion phases.

- Mappings
  - Technical challenge
  - Tools development
  - Consistency of implementation of comparable constructs
  - Best practices of meaning – a repository of all mappings as examples.
  - Best practices of encoding
  - How to structure mappings
  - Mapping dictionary

Action: FAQs to be written by Tony and Matthew
Action: Martin to facilitate creation of mappings repository at FORTH

TG suggested setting up a CIDOC CRM Wiki so we could collaborate on producing documentation such as mappings. Wikipedia is a specific example of a wiki that uses the software to develop and maintain a collaborative encyclopaedia. SM suggested looking at the Moin Moin software.

.
Action: Some one to set up a collaborative editing tool.
Decision: All to use CRM SIG mailing list for discussion

Issue: Problem with "reply to" when responding to messages on CRM SIG mailing list (replies to sender rather than list – discussions then go off list).
Action: MD to check list configuration to try to solve this problem.

Mappings

TG made the point that each mapping should start with the class that corresponds to the entity that the record describes, and should end with the class that corresponds to the field being described. This may require the direction for the property link to be made explicit (e.g. F – forwards or B – backwards).

A discussion followed on the problems associated with representing complex CRM mappings.

PLB described his methodology, working as suggested by TG to identify the subject most closely matching what is being described and using three columns to describe the path to the object. It may be necessary to repeat this process with the object becoming the subject.

SM asked if thought had been given to expressing the CRM in OWL.
Action: SM to look into possibility of doing this. MD suggested contacting Detlev Balzar.
MD offered to look into this but stressed need for mapping language.

MS & MD demonstrated the multiple column format used by SDS for the MIDAS mappings.

Conclusions
MD asked for a proposal for a mapping methodology. NC urged use of a simple approach.

Action: TG and NC will propose a simple mapping format. MS will make RL's mapping tool available ASAP. Consideration will also be given to the more structured approach used by SDS. SIG will then take a vote.

Issue: TG mentioned the currency problem of existing mappings to the CRM.

The group broke for coffee at 15:35.

The group reconvened at 16:15.

Progress of CRM Standardisation

 NC gave a presentation of the CRM standardisation process and the progress to date. TC46 SC4 WG9 is effectively a subset of the CRM SIG. The model is being developed under the accelerated procedure. The CRM has been through the first two stages and is now at the Inquiry stage, registration fo the draft international standard. The ballot stage will last five months. The subsequent Approval stage may be

omitted if there are no negative votes. ISO 21127 DIS is based on CRM version 3.4.6. The English version has already been submitted to NISO. Final comments from NISO are pending. The French version is nearing completion. NC has requested that there are no alterations to the British English spellings without changes to the grammar. The DIS cannot go to vote without the French version.

Entities and property declarations are expressed in a tabular format. Revisions have also been made to wording to meet ISO requirements. Other sections have also been removed.

If negative votes are received then the Final Draft International Standard phase will have to be gone through. Otherwise it can go straight to publication and diffusion.

Changes that can be made between the DIS and the Final Draft International Standard have to be very minor. Structural changes must be kept to an absolute minimum. This is so that it cannot be argued that the standard is not finished. The scope notes rewriting should not be a problem. The main issue is structural changes. The removal of one of the properties should not be a problem. Appendixes can be added with some examples. NC offered to make copies of the ISO DIS available in pdf format.

Action: NC to compare changes to from 3.4.6 to 3.4.10.

MD asked when the DIS would go to ballot. NC replied that this would be once the French version is finished. The balloting process will take 5 months. It is hoped that this will be completed by mid October.

Other issues

Issue: There is a need for greater explanation for the examples in the text, e.g. – do we need to explain who Michael Jackson is?
Decision: It was agreed that we should stop here (a line has to be drawn somewhere!

Decision: The group does not propose any further rewording to 3.4.10.

Decision: 3.4.10 to be reissued as version 4.0.

Issue: Italian Ministry of Culture wishes to become a member of the SIG on behalf of themselves and the MINERVA consortium.
Decision: Agreed
TB suggested persuading them to add reference to the CRM in the BRICKS project.

FRBR/CRM Harmonisation
PLB explained FRBR to the group – an intermediate ER model between a conceptual and data model. Widely accepted as a model for bibliographic data. He also mentioned Functional Requirements for Names and Authority Records (FRNAR). MD gave an account of the CRM/FRBR meeting. A decision had been taken to create an integrated OO model. So far there has been no need to make any changes to the CRM. A meeting took place from March 22-March 25. The result of this meeting was an elaborated conceptual model closer to the intended conceptualisation. MD gave a brief overview of development to date.

Action: MD to send out copy of the current document.

Another meeting is expected this year. This should help to substantially extend the user community for the CRM. MD invited anyone interested in joining this group to attend.

There will be a one-day meeting at Seoul. SDS will give a presentation to the whole ICOM assembly to raise awareness.

MD stressed the importance of using the CRM successfully – to keep in contact and to seek advice. NC clarified: "If you are happy, tell your friends, if not, tell us".

TG asked if the issues log would be checked on Thursday. Agreed.

Graphical representations

MD talked of mapping form CIDOC Information Categories onto the CRM (version 2). The CHIOS project gave support to the updating of this document to version 4 (formerly known as 3.4.10). The resulting diagrams include some new categories. SDS volunteered to describe application cases as a combination of several functionalities that are then described in these sub-models (e.g. phases of buildings, multimedia objects etc.).

MD asked for assistance in this work. Shared writing of texts with SDS would be helpful.

The graphs have now been updated. Every class and every property appears at least once. The graphs are a complete representation of the CRM in functional groups.

Action: MD to put document on the website.

Issue: The diagrams require proof reading.

The diagrams allow repeating of classes for the sake of clarity. This is made obvious by the different shading used in the class boxes. Shortcuts are shown with a grey background.

There was discussion of graphic representing Image Information. It was not possible to create a single graphic for Location Information (hence two diagrams).

The group broke for the day at 17:50.

Thursday 22 April

10:15 – Definition of minimal level of compatibility of the CRM

MD suggested that discussion should start with a definition of the requirements for a simplified subset of the CRM that was raised in Oxford. In the past we have discussed the level of detail at which someone should instantiate the CRM. With the creation of a P0 property it would be possible to make all datasets CRM compatible. There is a question of the level of analysis that is required.

SM mentioned the level 1 and level 2 approach. He pointed to the need to simplify the presentation of the CRM to new users. There is a need to present a global view. MS asked if this is a functional question or a presentation question. TG asked if the question also relates to scope. Is the proposed core attempting to address too wide an audience? The problem of Dublin Core aspiring to cover too wide an audience ws mentioned.

MD said that the standard itself refers to a minimal level of compatibility which needs definition. NC talked about the need to avoid loss of meaning. MD gave an example of incompatibility in a system that has a unit of information that combines Site and Place. If these cannot be disambiguated then there is a clear problem. But is it always going to be necessary to disambiguate to the lowest level of classes in the CRM? NC talked about inversions of CRM hierarchies or classes that connect to wrong classes – These could be valid examples of incompatibility. SM felt that these could be validated. SM wondered whether there are any objective ways of measuring this.

The classes represented in this diagram were considered to be particularly useful:
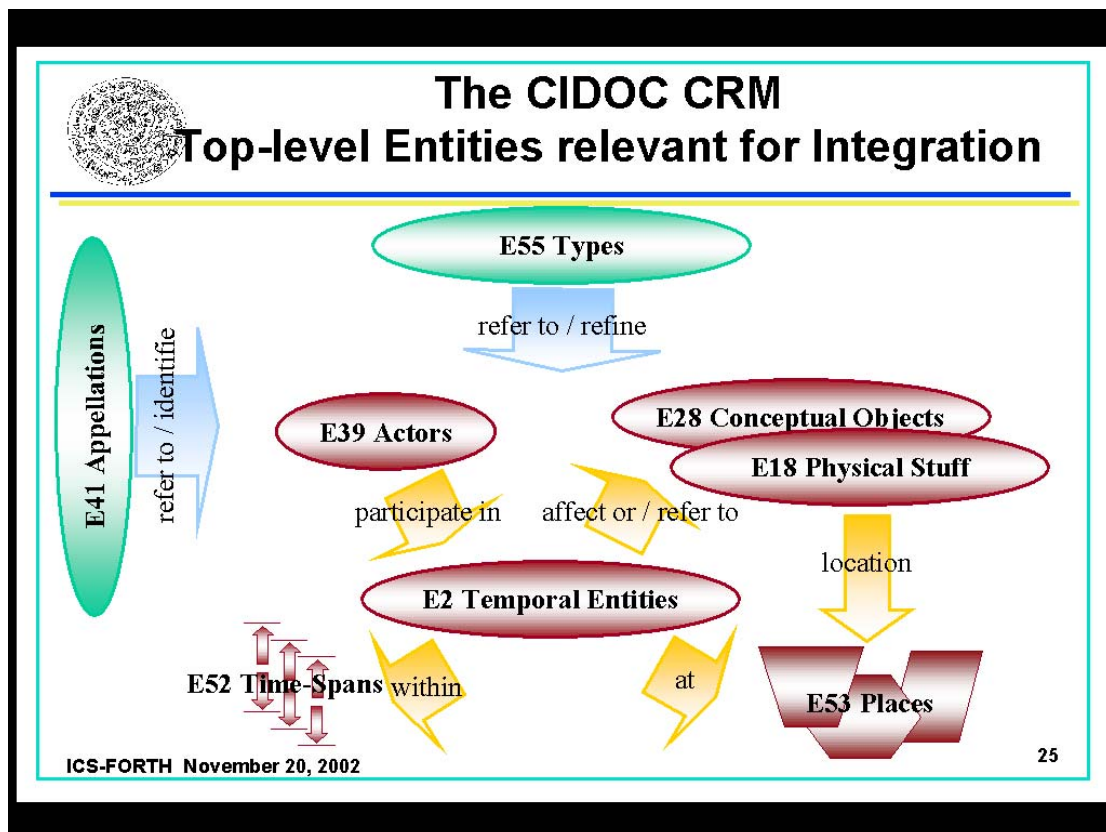
MD began to look at core levels of analysis:

A minimal system should be able to distinguish between physical stuff and conceptual objects.

- Physical Stuff
- Conceptual Object
- Event
- Actor

- Type
- Appellation
- Place
- Timespan
- Dimension

In Geneva the CRM is implemented in four tables:
  Actor
  Physical Object
  Document (Conceptual Object)
  Event
The Type hierarchy is represented using a thesaurus.

MD felt that there is a limited number of properties within the property hierarchies that would be sufficient for a minimal implementation. May be there should be more generalizations of the current properties that are not part of a property hierarchy.

NC wondered if such generalised properties are required for use in a functional way or whether they are also useful in a pedagogically to help to make the model more comprehensible.

MD wondered whether the properties identified so far can be categorised under the following classes
- Identification – P1
- Classification – P2
- Part decomposition
- Participation – P12
- Location
- Influence
- Reference

Are there any properties that don't fit? NC suggested P3 "Has note". This was dealt with by adding
- Characterisation – P3

This approach may well help in training newcomers to the CRM. NC felt that this was getting at what he meant by a lack of guide terms in the property hierarchy.
There was the general agreement, that there is no need for a core model of the CRM. The problem of complexity of the CRM is purely pedagogical, and not functional, since a user can select anyhow the level of abstraction s/he prefers.

The group broke for coffee at 11:20

11:40

MD felt that there was little need to discuss this further. Tasks:
- Create a global view
- Identify the minimal level of analysis required by someone building a new system

It is agreed that the properties are not sufficiently structured. There is a need to check that the existing properties fit into the suggested classifications and to identify the existing super properties that already provide structure. This can then be written up to provide clearer introductory material.

Decision/Action: MS, NC and MD to work on this.

Examination of outstanding Issues in the Issues Log

Issue 22: How to deal with implementation guidelines.
The idea of writing an implementation guideline has been abandoned. Information exchange using SIG members is probably the best approach for this. MS felt that this issue can probably be closed. There is the question of where information on uses and applications should best be gathered on the web site.
Proposal: Presentations and papers about CRM applications should be placed under the Implementation Guidelines section.
Issue: TG pointed out that much content on the website can fit under more than one category. Perhaps the website needs its own ontology! TB felt that this is where a content management system is required. CAA has the same problem. TB offered to share his findings with the SIG. MD will investigate possibilities with FORTH
Decision: All implementation papers will appear under implementation guidelines. MD will create a page about CRM Use which will be limited to contact information. Issue closed

Action: MD will make all the diagrams available on the website as diagrams rather than PowerPoint presentations.
Action: TG to create unzipped version of version 4 for the website
Issue: TB mentioned the problem of the header graphic on the CRM website.
Action: TB to resolve this problem in conjunction with Lida.
Action: All to check content of website for currency.

Issue 44: Modelling states
Decision: Leave the issue open until a real application case emerges.

Issue 54: FAQs
Action: NC to propose FAQs derived from the Issues log
Action: MS to identify FAQs in recent lists of minutes.

Issue 57: Effort to teach use of the CRM
What is the material required? How long does it take to get a global view of the CRM? About a day. To teach mapping skills? Perhaps three days. To digest, absorb, gain confidence? Several months. There is a minimum time period required. NC suggested analogies – learning a programming language, driving a car, learning golf….
Action: Someone to formulate this
Action: MD to update tutorial on the CRM website.
Issue closed.

New Issue: Define a comprehensive list of training materials

New Issue: FAQ required to deal with availability of the standard. Possible use of core standards. Local publication. Keep it available in a slightly different format on the website.

Issue 126: Explanation of Allen Operators.
Action: MS to complete paper on way home from this meeting. Issue remains open.

Action: MD to check if the CRM text mentions multiple instantiation

Action: MD to remove remove second instance of the definition of Instance from the CRM. The second of these needs to be removed. Martin to do this. The CRM still contains two

Presentation on Sets, Collections and Composites

Martin repeated a presentation given in Paris to the group.

The group broke for lunch at 13:00.

Sites and Monuments Software Presentation

Following lunch there was a presentation of the Greek sites and monuments inventory software.

Functional Specifications for CRM-based retrieval interfaces

MS was absent during much of this session so minutes are incomplete.

Issue: Different introductions to the CRM are required based on end-user needs.

There was discussion concerning whether there should be URIs for the entities and properties in the CRM to facilitate compatibility with RDF.

User interface suggestions for integrated public access to collection data to facilitate implementation.

Action: MD to write a short paper and to receive feedback from TB, NC and TG (and anyone else on the list).

NC suggested that negative examples should also be included (things that don't work).

Venue and topics for next meeting

A meeting will be held in Seoul.

An ISO TC46 meeting will take place in Washington on October. It is possible that a meeting could be hosted by NISO or in New York (TG).

Siegfried Krause offered to host a meeting in December in Nuremberg.

Topics to include
- Are there useful extensions to the CRM that are in scope – e.g. those required by
  - FRBR (ongoing)
  - more specific archaeological reasoning (possible)
  - handling social relationships (planned)
  - social norms
  - etc.
- How would multiple CRM-compatible systems talk to each other?

Decision: Meeting to be held in Nuremberg in December.