

# **DELOS Working Group on Ontology Harmonization, co-sponsored by the Harmony Project**

***Full report from the first meeting at CNR in Rome,***

***March 26-27 2001***

**Editor:**

**Martin Doerr, Jane Hunter**

**Creation date:**

**April 12, 2001**

**Meeting Place:**

The meeting took place at the facilities of the National Research Council in Rome, Italy. We wish to express our gratitude to Aldo Gangemi from ITBM-CNR, Rome for the technical organization and the hospitality.

**Beginning of the meeting:** Monday, March 26, 9:30

**End of the meeting:** Tuesday, March 27, 17:00

**Participants:**

Thomas Baker,	thomas.baker@gmd.de
Dan Brickley,	Daniel.Brickley@bristol.ac.uk
Donatella Castelli,	castelli@iei.pi.cnr.it
Nicholas Crofts,	nickcrofts@yahoo.com
Martin Doerr,	martin@ics.forth.gr
Aldo Gangemi,	gangemi@saussure.irmkant.rm.cnr.it
Nicola Guarino,	Nicola.Guarino@ladseb.pd.cnr.it
Jane Hunter,	jane@dstc.edu.au
Stephan Koernig,	Koernig@ifs.tu-darmstadt.de
Carl Lagoze,	lagoze@cs.cornell.edu
Wolfgang Meier,	meier@ifs.tu-darmstadt.de
Nikolay A.Skvortsov,	leonidk@synth.ipi.ac.ru

This is the technical report on the contents of each presentation and discussion.

## **Day 1:**

### **Presentations**

#### **1. Tom Baker**

gave an introduction to the background and goals of the Working Group. He defined the White Paper the Group is expected to produce.

#### **2. Martin Doerr:**

Welcome.

#### **3. Brief Introduction from each participant**

#### **4. Carl Lagoze**

gave an overview of Harmony and the ABC Model.

He explained the history and motivations behind the model. He presented its scope and goals and the approach taken for its construction and explained the approach the Harmony Group takes to solve metadata interoperability. He explained the entities and properties of ABC by graphical representations.

Details:

- ABC came out of DC (Dublib Core) qualification problems.
- There are conflicts between those who want to keep DC simple and those who needed to complexify it (often in uncontrolled ways).
- It is wrong to try complex modeling with simple DC Elements.
- The lack of a data model and first class objects etc. causes problems.
- The objective of ABC is attempting to understand random different metadata sets and
- trying to solve the mapping/common understanding problem by identifying commonalities in different metadata formats.
- There is a need to introduce linguistic entities.
- Completely automatic processing/understanding of mappings is impossible. The limits are close to the well-known problems of natural language understanding.
- The objective is building an additional semantic layer (the ABC vocabulary framework) on top of metadata descriptions,
- not developing the “grand” metadata scheme – but developing processes or approaches to consistent metadata modelling.
- We believe that modelling of the underlying semantics is the correct approach, models with explanatory power rather than prescriptive power.
- Another objective is developing search interfaces across graphs and RDF Databases using Squish. We regard as a proof of concept, "if you can ask questions about it",
- ABC is based on IFLA’s FRBR work – works, expressions, manifestations, items. It is formulated in RDFS.

- An overview of the ABC model contents was given showing graphical presentations from the Sirpac processor

## 5. Jane Hunter:

presented results from modelling CIMI examples using ABC. She explained difficulties with semantic mapping of poorly structured museum data and commented on specific logical situations the ABC model does not cover appropriately. She tried a first characterisation of the difference between ABC and the CIDOC CRM.

Details:

- The mapping exercise revealed presence of a lot of natural language text within “notes” explaining the museum object’s life cycle in detail
- It appears that accurate modelling needs to be done manually on an individual object-by-object basis – not machine-processable without lossiness
- The ABC Model has difficulties handling the property enhancement situation i.e., events in which properties are added or property values modified or deleted but object is unchanged
- The ABC Model also has difficulty handling the “identification” problem.. When is an object the “same” thing and when is it a new object. What set of properties define “sameness”? This set can change over time.
- It is clear that two major differences between ABC and CIDOC CRM are:
  1. ABC focuses on digital information objects whilst CIDOC CRM is more concerned with physical museum objects
  2. ABC is more event-focussed (snapshots in object’s lifecycle at times of major change) whilst CIDOC/CRM expresses an object’s state over “periods” and changes of state.

## 6. Nick Crofts

gave an overview of the CIDOC Conceptual Reference Model. He explained the historical background of the Model and described its objectives. He described how characteristics of museum data have influenced the design of the model. He finally presented its top-level entities.

In detail:

- One of the reasons to develop the object-oriented CRM was to overcome the problems (rigidity, complexity, incoherence) with the older CIDOC Relational Model developed from a database at the Smithsonian Institution until 1994.
- The CRM made a shift from a static data model to an extensible object-oriented semantic modelling in order to cope with the inherent diversity of cultural data.
- The objectives are to establish a neutral (institution independent) and implementation independent model for distributed access, and to develop a formal domain ontology for cultural heritage, suitable to explain the meaning of various data structures.

- The CRM is a product of reverse engineering of the conceptualisation of the museum universe of discourse from the CIDOC Relational Model, the CIDOC Information Categories and documentation guidelines and other data formats.
- It focuses on access to data from cultural heritage institutions, and the interoperability between their data and archives and libraries.
- Cultural heritage data are often incomplete or even contradictory - which requires working with multiple hypotheses and varying levels of detail.
- The CRM attempts to stay at a level of detail and granularity required by museums - this may be discipline-dependent, e.g. contemporary arts vs. fine arts.
- The CRM is designed for use in mediation systems - useful as a 'lingua franca' for mapping between data and metadata formats.
- It is further intended as a reference for good practice, clarification etc., and as an aid for schema design e.g. XML DTD, XML Schema, or RDBMS and ooDBMS schemata.
- Basic Entities of the CIDOC CRM are the *Temporal Entity*, *Physical Entity*, *Actor*, *Place*, *Time-Span*, *Dimension*, *Appellation*, *Type*.

## 7. Martin Doerr

gave an overview of CIDOC CRM contents and methodology. He presented the spatiotemporal concepts in the CRM, how they are specialised and interact with other entities. He described the methodology applied for the CRM, in order to use objective semantic criteria to (1) decide between modelling alternatives, and (2) to restrict the Model to a manageable unit of "core" concepts. Some key points:

- All documents about the CIDOC CRM are available at <http://cidoc.ics.forth.gr/>
- *Temporal Entity* is an abstract class which has no direct instances. Only its subclasses such as *Period* or *Condition State*, *Event* or *Activity* (which bring people in) may have direct instances.
- *Periods* bind related phenomena. *Periods* can contain *Periods*. *Events* are subclasses of *Periods* but have outcomes.
- The CIDOC CRM understanding of an event does not need atomic knowledge of its internal processes - knowledge is much more robust outside the event – often the long-term effect of an event is known but the details are not. The longer the temporal distance from an event, the more it can be dealt with as a discrete entity. Nevertheless an event can be analysed by any number of subevents.
- An example of subclassing: Entity *Attribute Assignment* has subclasses: *Measurement*, *Condition Assessment*, *Identifier Assignment*, *Type Assignment*. The idea is to create a new class in the CRM if it brings in a new attribute – otherwise subclasses are expressed by the "has type" property. This is not a semantic distinction, but a methodological choice to keep the CRM a manageable unit.

- There is a concept of the *Stuff* and which *was present at* and the *Actors* which *participated in* the event. This supports basic chronological reasoning.
- *Man-made Objects* can have specific purpose (if commissioned) or general purpose by creation, this intention is distinct from actual specific or general use. Symmetrically, *Activities* have general and specific purpose, and make use of objects (tools as well as plans).
- *Place* is a geometric extent in space. Currently, a primitive topology is supported: *Places* can fall within places.
- Many triangular relationships involving events have simpler shadow information, so-called "short-cuts" which do not include temporal information e.g. *has condition* as well as an *Condition Assessment* event. This is a non-formal way to document basic deductions, and to demonstrate an extension mechanism. This makes the CRM different from a documentation recommendation.
- Functional requirements for the CRM are to: import, transform, merge, mediate data or metadata.
- It needs to deal with incomplete knowledge. It must allow for contradictions and ambiguity, and support scholarly or scientific reasoning about possible pasts by correctly rendering the historical record. The monotony of information revision when knowledge is being completed is a major concern. The latter may make a major philosophical difference to foundations of other ontologies.
- Explicitly modelled attributes are restricted to those supporting relevant queries (but not to express directly queries). This is again a methodological choice to keep the CRM within a specific scope. Nicola Guarino regards this as a database feature - ontologies provide understanding of terminology semantics. Martin Doerr: The CRM supports terminology semantics by scope notes (free texts) and isA relations only.
- Comment by Nicola: one needs more than query relevance if trying to merge/harmonize metadata from heterogeneous domains.

## 8. Nicola Guarino

spoke about the ontology integration problem and the role of top-level ontologies. He presented a model of ontology sharing and a methodology to employ logical criteria to clarify differences and commonalities of concepts, e.g. to improve the quality of concept hierarchies. Basic in his methodology are ontological distinctions supported by notions of rigidity of properties, notions of identity and unity of particulars. Some key points:

- the ontology integration problem and the role of top-level ontologies
- bad vs good ontologies
- predicate logic language to express a conceptualisation
- ontology overlap - problem if two ontologies (approx models) do overlap but intended models do not overlap
- top down integration using top level ontology reduces the problem of unintended overlap

- common top level ontology – encourages reuse
- identity – how can an entity change but maintain its identity?
- Which properties are essential for identity?
- When are 2 entities the same?
- Need tools or formal language to express each view or conceptualisation e.g. person has a brain. Views may differ but need to be able to clearly express each view.
- Properties with rigid predicates are necessarily true for all instances
- <http://www.ladseb.pd.cnr/infor/ontology/ontology.html>
- identity and unity - where does the dog end and the collar start?
- Synchronic (occur at the same time) - or diachronic (same across time or at different times)
- Identity criteria – certain relationship between x and y implies  $x=y$
- Constitution and identity – its not true that 2 things are the same just because they have the same components – except for things like collections or sets
- May need relative importance of parts so if parts are replaced then in some cases the identity is unchanged, in other cases, the object has a new identity e.g. replacing the lid of a pen
- Identify identity criteria (IC) for 2 different objects,
- An entity is a whole if all the parts are linked by an equivalence relation = unity criteria (UC) i.e. object x is a whole
- Different kinds of wholes depending on the glue or equivalence relation e.g. topological wholes (lump of coal), morphological (constellation), functional wholes (bikini), social wholes (population)
- Singular, plural wholes (sum is also a whole), collections (sum is not a whole)
- Formal ontological analysis is very time consuming
- Useful for making modelling assumptions clear to resolve and recognize conflicts
- Taxonomic analysis – label properties with I (identity), O, U (unity), D(dependence), R(rigidity)
- problem that this is based on common sense physics – translation to the web, electronic or digital world may be problematic

## 9. Donatella Castelli

presented the methodology and experiences from the ECHO (European Chronicles Online) project creating the VandA common model for audiovisual libraries, which combines the IFLA FRBR model with multimedia metadata modelling requirements. Those were:

- to satisfy the modelling needs of the ECHO application domain (Historical documentary films);
- extensibility; and
- reusability.

In order to achieve these objectives ECHO started from an existing conceptualization (IOLA FRBR) that satisfied the general modelling needs and then refined it.

## 10. Nikolay A. Skvortsov

gave a brief description of the research and goals of the Institute for Problems of Informatics, Russian Academy of Sciences. He presented the approach SYNTHESIS model for merging of conceptual models and proposed its application for the CRM-ABC harmonization.

### ***Discussion at end of Day 1:***

The goal of this discussion was to clarify the fundamental common agreements and to refine the agenda for the following day. For that purpose, a set of reasons to share and the benefits of merging our ontologies was identified. An agreement on the general goals of the white paper was achieved and some steps in achieving a shared model were identified.

Details:

- We want to create a methodological framework for identifying the overlap between ontologies on the example of the CRM and the ABC, explore the degree of merging that can be achieved, and investigate methods for maintaining harmonized, semi-autonomous ontologies in long-terms.
  - There is a need to state explicitly what we fundamentally agree on, in terms goals, approach and terms.
  - What are our objectives and why do we want to share ontologies? Some reasons to share and the benefits of merging our ontologies in **general**:
    - To help people to design “good” metadata structures: make recommendations, foster quality, best practice:
    - provide a methodology for ontology sharing/merging so other domains can be covered in the future – now cover museums, libraries, archives
    - Sharing of equivalent propositions.
    - Retrieval of stories through chains of references; this application may require traversal of different domains with different needs through shared concepts in the overlapping areas.
    - Information extraction
    - Contextualization or integration of related data/information sets
  - **In particular: Harmony covers more digital objects and CIDOC CRM covers more physical objects so the merging of two worlds make sense.** Moreover there is hope to generate an ontology with a wider scope by combining other supplementary models.
  - How we can share ontologies? Some ideas:
    - Define the kind of questions that this would allow us to pose.
    - Need to identify identical concepts
    - Consider need for easy crosslinguistic translations of terms
-

- Representation questions should be addresses– RDF Schema, DAML – or richer languages?

## **Day 2:**

### **Discussion:**

The group discussed the following topics and agreed on position statements for each topic:

1. **Identification of the Common Ground of the Participants.**
2. **Why should we share common models?**
3. **Candidate ontologies & models for sharing or taking into account.**
4. **Design of the methodology to compare, share or merge ontologies and its steps.**
5. **Identification of representational alternatives - possible form, expressive power, encoding.**
6. **Management and administration – how to manage a shared ontology, how to manage future merging of ontologies from other domains or future diverging demands.**

## **Identification of the Common Ground of the Participants**

### **1. Identification of the Common Ground of the Participants**

We believe in the following approach:

**What:** We both are trying to develop formal representations which approximate the underlying conceptualization of a domain.

We want to encourage reuse of conceptual constructs and methods or guidelines by descriptive communities.

It may be possible to approximate a useful shared conceptualization across multiple domains.

**Why:** To make explicit the underlying semantics and semantic relationships across heterogeneous data. We believe it is an important step in enabling information interoperability, information comparison and reuse.

We want to identify common entities and relationships in heterogeneous data for information integration, comparison, cross-domain searching.

We do want to provide a framework for designing unambiguous well-structured metadata.

**How:** (Semi-) structured data become available from the communities. The Group's concern is not how (automatic, by hand etc.), but what do **once they exist**.

Neither ABC nor CIDOC CRM have been designed to provide a data format



or **prescription** by which users can create metadata records, but to **explain** existing metadata.

Moreover, we want to provide **methods** and guidelines for **designing** "good" metadata models to improve quality and interoperability of metadata records and for comparing and converging metadata models/ ontologies.

All this by identifying a mechanism for determining commonalities between models and by identifying common entities and relationships across heterogeneous data.

The idea is to create a **refinable** Conceptual Reference Model, an "Ontology with wings clipped".

## 2. Why should we share common models?

Problem: The report should specify:

- What would this allow us to do, which function can we achieve ?
- What queries we could ask?

Answers:

- Identification of commonalities and differences supports comprehension.
  - By combining/comparing both models we can enhance both.
  - We can provide better principles for designing "good" metadata.
  - By combining/comparing both models we achieve broader domain coverage
  - We can share equivalent propositions
  - We can answer specific (cross-domain) questions
  - We can **integrate** propositions via shared common entities
- CRM and ABC take similar approaches !

## 3. Candidate ontologies & models for sharing or taking into account

A list of candidate ontologies and models was begun, to be compiled by Dan Brickley, and subjected to a classification by Nicola Guarino. Those are:

- CYC
- Pat Hayes catalogue of Temporal Entities
- SIMPLE Project (starting from WordNet)
- WordNet/ EuroWordNet
- IndeCS
- SPIRT (Archival "Record Keeping" Community)
- FRBR
- IKF Project (EUREKA)
- release a Reference Ontology end of 2001
- IEEE Standard Upper Ontology Effort (suo.ieee.org)

- should be contacted!
- ONIONS

#### 4. Design of a methodology to compare, share or merge ontologies and its steps.

An introduction should clarify or define the terms we used to talk about the models, the way of representation, the notion of refinement etc.

##### Step1: Identification of context

We identify for each model:

- context of the model;
- scope;
- functional purpose;
- nature of the model.

Details:

We should distinguish the definite (current, undebatable) scope from the potential, envisioned scope.

The scope should additionally be defined in terms of examples and the envisaged "corpus" of resources the model is thought to cover.

We agree to define scope in terms of the 5 aspects mentioned in the Working Group invitation:

1. The **conceptual framework** (viewpoints) of the intended users (e.g. scholars, museum professionals and museum visitors, etc.)
2. The **activities** intended to be supported (e.g. collections management and conservation, research and analysis, promotion and communication)
3. The kinds of **objects** targeted at (e.g. normally museums don't collect viruses, gases, etc.).
4. The level of **detail** and **precision** required to provide an adequate level of quality of service.
5. Considerations of the necessary and manageable **technical complexity**.

Example for the functional purpose: "ABC is not necessarily meant as a data format, but to be used as a "single search interface"."

##### Step2: Identification of the nature of the model contents

We identify for each model, the ontological nature of the contents:

- entities;
- properties (in RDF terminology) or attributes (in TELOS term.)

and the representational contents:

- metamodel (e.g. RDF) and modelling primitives (e.g. relations)
- systematic patterns which occur and their justification (e.g. "short cuts" in the CRM)

### **Step3: Identification of common parts and achievable sharing**

We identify common parts and achievable sharing between entities, properties and metamodels and modelling primitives:

- Start with ontological commitment of entities and properties (attributes) by use of metaproperties (in the sense of N. Guarino)
- Compare their representation in terms of metamodels
- Compare expressiveness of employed representation mechanisms
- Try transformation into a useful common representation
- Evaluate differences
  1. differences due to different representation mechanisms
  2. differences due to different scope
  3. difference due to different ontological commitment
- Formulate overlap, with examples - a “corpus of resources”.
- Identify maximal achievable sharing.

### **Step4: Evaluation of the proposed methodology**

We apply the above proposed methodology to ABC and CIDOC/CRM (identify the set of common notions between the two models, at both ontological and representational levels; identify the differences between the two models and reasons for the differences; identify achievable sharing):

- We identify the problem issues which arise when trying to compare two models – try to identify/ eliminate the inadequacies of the method.
- We would also like to identify representational alternatives - possible form, encoding e.g. RDF Schema, DAML? And evaluate adequacy of expressive power and comprehensiveness.

## **5. Management and administration**

Basically the discussion on

- how to manage the merged ontology,
- how to manage future merging of ontologies from other domains – was postponed to the next meeting.

## **6. Research issues, future work**

The following preliminary topics were mentioned:

- To propose an analytic method for supporting convergence of "models"
- Does the method scale?
- Does it persist?

## **Decisions:**

The Group decided on the following:

### **Goals of the white paper:**

- 1) To propose a method for conceptual model comparison and sharing
- 2) To use the example of CRM /ABC to verify and demonstrate this methods
- 3) To identify at least to some common and some different notions in the CRM and the ABC model
- 4) To identify a maximal possible set of shareable notions and a minimal set of different notions.

It was agreed, that the goals of this group in terms of ontology sharing go beyond what can be promised for the white paper.

### **The group defined the complete structure of the white paper:**

The following outline for the white paper to be produced was agreed on (editor Tom Baker):

1. Intro – models are similar, describe objectives/mutual benefits of merging ontologies.
  - 1.a Definition of Terms/Glossary somewhere near the front e.g., what we mean by “model”, “schema”, “ontology”, “refinement”, “interoperability”
2. Goals of this workshop/whitepaper
  - a. Propose a method for conceptual model comparison and sharing
  - b. Identify the set of common notions and the differences between the two models
  - c. Identify the problem issues which arise when trying to compare/converge two models, the inadequacies of the method
3. Description of CIDOC/CRM – identify functional purpose, scope, nature of contents, examples
4. Description of ABC – identify functional purpose, scope, nature of contents, examples
5. Proposal of a formal analytical deterministic method(s) to compare/converge ontologies.
6. Application of the proposed method(s) to ABC and CIDOC/CRM to compare the two models and identify commonalities and differences:
  - Ontologically
  - Representationally
    - Compare modelling constructs
    - Compare metalevel constructs (patterns)

- Identify achievable sharing
- 7. Evaluation of the method used for ontology comparison
  - Does it scale?
  - Is it persistent?
  - What are its inadequacies?
- 8. Representation alternatives for the merged ontology – form, expressive power, encoding
  - compare expressiveness of different representation mechanisms
  - transform common/merged ontology into a useful representation
- 9. Management and administration – how to manage the merged ontology, how to manage future incorporation of other ontologies
- 10. Research issues, future work
- 11. Conclusion – prospective audiences for White paper – other metadata initiatives which may be interested in merging their ontologies with ours.
- 12. Bibliography

### Task List:

- Dan to set up email list;
- Dan to provide brief descriptions plus references for related ontologies/projects;
- Nicola to write a chapter for non-experts about the application of his formal analytical methods to ontology comparison for use in this Working Group and in the final report.
- Nicola to propose to the Working Group for Ontology Contents Standardization of the OntoWeb project concertation with this Working Group. Donatella to create a respective statement on behalf of DELOS.
- Everyone to look for other co-funding possibilities and concerted efforts.
- Dan to look into Semantic Web Action Line for potential common proposal
- Dan to make a statement for the VIth framework discussion to be presented by Nicola, April 27 at the CEC.
- Carl, Jane, Dan to complete the following for the ABC model and Martin, Nick Crofts for the CRM by June 1:
  - Identification of the functional purpose and scope using CIDOC/CRM's method of analysis:
    6. The **conceptual framework** (viewpoints) of the intended users (e.g. scholars, museum professionals and museum visitors, etc.)
    7. The **activities** intended to be supported (e.g. collections management and conservation, research and analysis, promotion and communication)
    8. The kinds of **objects** targeted at (e.g. normally museums don't collect viruses, gases, etc.).
    9. The level of **detail** and **precision** required to provide an adequate level of quality of service.
    10. Considerations of the necessary and manageable **technical complexity**.

- This in terms of the definite functional purpose and scope and of the potential ones.
  - Identification of the nature/content of the model
    - Definition of Entities
    - Definition of Properties
    - Clarification of the metamodel/data model
- For the CRM this is necessary only for items with overlap with ABC. Specific attention should be paid for identity criteria.
- Illustrative examples
  - Both reports will be distributed among the participants to support the preparation of the next meeting.
- 

## **Next meeting**

The next meeting is envisaged either in Darmstadt after the ECDL 01 conference – around Sept. 11-12 or in Crete at the same date. It will be clarified, if DELOS can fund rental of a meeting facility.