



## "Interconnected data worlds. Workshop on the implementation of CIDOC-CRM"

### SCHEDULE

Monday 23.11.2009

- 09.00-09.10 Begrüßung  
Felix Schäfer, Archäologisches Institut, Universität Köln
- 09.10-09.20 CIDOC-CRM aus Sicht von TOPOI  
Undine Lieberwirth, TOPOI, Freie Universität Berlin
- 09.20-09.30 CIDOC-CRM aus Sicht des DAI  
Ortwin Dally, Generalsekretär des DAI, Berlin

#### I. CIDOC-CRM APPLIED IN THE CONTEXT OF MUSEUMS

- 09.30-11.00 Einführung: CIDOC-CRM aus der Museums-Perspektive  
Siegfried Krause / Karl-Heinz Lampe

— Kaffeepause —

- 11.20-12.00 Using CRM and SKOS as a basis for the datamodelling of different collections  
and for the presentation of archaeological objects in PICA  
Axel Vitzthum / Frank Dührkohp
- 12.00-12.40 CRM in der Praxis: Das Harvestingformat museumdat  
Regine Stein

— Mittagspause —

#### II. CIDOC-CRM AS A TOOL FOR RESEARCH

- 14.00-14.40 CIDOC CRM and text data - The World of Thuc.  
Agnes Thomas / Reinhard Förtsch

- 14.40-15.20 From Questions to Answers - Travelling from Perseus to Arachne and Anywhere  
Robert Kummer
- 15.20-16.00 Prototype of a CIDOC-CRM implementation in a relational database with GIS interface  
Gerald Hiebel

— Kaffeepause —

### III. IMPLEMENTING THE CIDOC-CRM

- 16.20-17.00 CIDOC-CRM for sharing, CIDOC-CRM for use  
Graham Klyne
- 17.00-17.40 Das CIDOC CRM in der Praxis - Erfahrungen, Probleme und mögliche Lösungen  
Bernhard Haselhofer / Philipp Nussbaumer

— Abendempfang am DAI —

## Dienstag 24.11.2009

- 09.00-09.40 The CRM-English Heritage extensions of CIDOC CRM  
Keith May
- 09.40-10.20 Breaking down barriers to interoperability  
Douglas Tudhope / Ceri Binding

— Kaffeepause —

### IV. ALTERNATIVES TO CIDOC-CRM

- 10.40-11.20 The data exchange format ADeX of the Archaeological Heritage Management in Germany - Intentions and first thoughts about the mapping with CIDOC-CRM  
Axel Posluschny
- 11.20-12.00 Archaeology, Formality and the CIDOC-CRM  
Leif Isaksen
- 12.00-13.00 Abschlussdiskussion



# "Interconnected data worlds. Workshop on the implementation of CIDOC-CRM"

## ABSTRACTS

### I. CIDOC-CRM APPLIED IN THE CONTEXT OF MUSEUMS

Einführung: CIDOC-CRM aus der Museums-Perspektive

Siegfried Krause, Germanisches Nationalmuseum Nürnberg

Karl-Heinz Lampe, Zoologisches Forschungsmuseum Alexander Koenig Bonn

#### 1. CIDOC-CRM, an introduction (Krause)

- use in museums,
- history,
- methodological approach

#### 2. Mapping of knowledge in terms of the CIDOC-CRM (Krause, Lampe)

- from field names to events

#### 3. Transdisciplinary scenarios (Lampe)

- multiple instantiation,
- multiple verification,
- semantic information integration

#### 4. Scientific communication infrastructure (Lampe, Krause)

- presentation of the WissKI project (Wissenschaftliche KommunikationsInfrastruktur)

For more than 100 years the methodological background of documentation in scientific and scholarly disciplines hasn't substantially changed. The conceptual goals of structuring information within modern databases are still quite similar to those concepts used in traditional card index boxes. They provide, if at all, a limited support for a well structured documentation of information in form and content. So, the traditional practice of documentation in museums shares with analogous practises in archives and libraries a focus on classes entitled with simple field names such as painter, collector or author. A trend is now underway across scholarly and scientific disciplines in which the focus of documentation is expanding to include processes and events.

Thus the entities object/concept, person, place, time and activity or respectively what, who, where and when are now being related to each other through processes and events. Through a formal ontology, information handling and dissemination can be improved by semantic enrichment.

The formal specification of semantic concepts makes scientific activities understandable to a wider audience. Finally, in scientific as well as scholarly disciplines a formal ontology can give simple object documentation its full scientific depth and that beyond the limited purposes of collection management. In addition a documentation of processes and events seems to be a prerequisite for transdisciplinary information integration, which is needed for developing knowledge networks and knowledge representation tools on the internet.

<http://cidoc.ics.forth.gr/>

[http://de.wikipedia.org/wiki/CIDOC\\_Conceptual\\_Reference\\_Model](http://de.wikipedia.org/wiki/CIDOC_Conceptual_Reference_Model)

<http://forschung.gnm.de/index.htm>

<http://www.zfmk.de/web/Forschung/index.de.html>

---

## Using CRM and SKOS as a basis for the datamodelling of different collections and for the presentation of archaeological objects in PICA

Axel Vitzthum, Schleswig-Holstein Museum Kiel

Frank Dührkohp, Gemeinsamer Bibliotheksverbund Göttingen

digiCult is an interdisciplinary project for digital collection and publication of museum inventories. At present more than 70 museums in the German Federal States of Schleswig-Holstein, Hamburg and Saarland take part in it. We use a combined data management system which allows separate recording of data by each museum and integration of records into a central metadata pool. Metadata of this pool is made available to the public via the website "Museumsportal Nord" and the online Saarland museum portal. In addition, we transfer the data to national and international online portals like Europeana, BAM, Prometheus and GBV.

The paper is divided into four parts. First, we are going to show how CIDOC-CRM and museumdat/LIDO affected the cross-domain data modelling of Intranet Hamburg and the future web-based collection tool digiCULT.web. Then we will look briefly at the web-based digiCULT.xTree tool which allows collaborative work on vocabularies. It is mostly compatible with standards like ANSI/NISO Z39.19-2005 and BS8723-2/5 and uses SKOS XML/RDF as exchange formats.

In the third part we are going to explain the prototype of the digiCULT Accumulator. This modul combines the museumdat/LIDO metadata with the SKOS-based concepts and terms for metadata enrichment.

Finally, we will demonstrate the data transfer of a few archaeological items into Europeana, BAM-Portal and GBV. For instance, we will show how museumdat is transcribed into the bibliographic format PICA+. PICA is the software used for the union catalogue of the GBV. Once turned into PICA+, data mashups with other sources of the GBV can be made to discover related things depending on the museum objects. Thus, we give an outlook on how various archaeological data sources can be connected via the GBV.

<http://digicult.museen-sh.de/>

## CRM in der Praxis: Das Harvestingformat museumdat

Regine Stein, Deutsches Dokumentationszentrum für Kunstgeschichte, Marburg

Abstract not yet available

<http://www.museumdat.org/>

---

## II. CIDOC-CRM AS A TOOL FOR RESEARCH

CIDOC-CRM and text data - The World of Thuc.

Agnes Thomas, Archäologisches Institut, Universität Köln

Reinhard Förtsch, Forschungsarchiv für Antike Plastik, Universität Köln

The idea of this project has been developed from the Perseus Digital Library (Tufts University), administration Prof. Gregory Crane, and the Archaeological Object Database Arachne (German Archaeological Institute, University of Cologne), administration Prof. Reinhard Förtsch, with the aim to develop new modes to bridge together ancient literary data sources and archaeological data sources in the Semantic Web. Now with more and more information becoming available, there is an increasing demand for specific global research, comparative studies, data transfer, data migration and data mining in the heterogeneous sources of cultural contents.

The thematical background for the project is the Pentecontaetia, the almost 50 years between the Persian War and the Peloponnesian War in Greece in the 5th century B. C. (479-431 B. C.). The main literary source for this period is the text of the ancient greek historian Thucydides (Book 1 Chapter 89-118). To comprehend the history of this time segment using all available literary and archaeological sources as well as modern research literature from the Semantic Web has been the historical dimension of this project. One overarching quality of the project is that it provides a probably unusual view on well-known-sources that generally get analyzed in ways marked by hundres of years of routine which are hard to escape, even if one tries to reflect on this fact - which is not the regular case.

The work is structured in three main parts:

1. Encoding the Pentecontaetia (Thucydides 1,89-118)
  - 1.1. TEI markup
  - 1.2. TimeML markup
2. Creating a bibliography for further applications such as Text Mining
3. Event Modelling with CIDOC CRM to connect the different sources

<http://www.arachne.uni-koeln.de>

<http://www.perseus.tufts.edu/hopper/>

---

## From Questions to Answers - Travelling from Perseus to Arachne and Anywhere

Robert Kummer, Hist.-Kulturwiss. Informationsverarbeitung, Universität Köln

This contribution wants to provide some background for a discussion of what frustrates users that are collecting information on cultural heritage (CH) topics on-line and off-line, today. For that, it will introduce a user story that describes a historian working on a specific topic. It will be discussed how a system that has seamless access to multiple CH data sources can help this user to answer his questions. This will lead over to an illustration of the obstacles of seamless data integration across CH information systems by referring to ongoing data integration projects.

---

## Prototype of a CIDOC-CRM implementation in a relational database with GIS interface

Gerald Hiebel, Surveying and Geoinformation Unit, Universität Innsbruck

In the multidisciplinary research project HiMAT(History of mining activities in the tyrol) CIDOC CRM was used to structure Metadata in combination with spatial data in order to display metadata in a GIS and provide additional information to digital resources stored in a content management system. In a prototype to test the ability of CIDOC CRM classes to represent the desired information the metadata was entered in the content management system and attached to the digital resources. Spatial data was always stored in a relational database(oracle) and accessed with a GIS. In the next phase of the project the metadata was shifted to the relational database already containing the spatial objects. Some compromises have to be made in order to build a data structure in a relational database representing even part of an ontology like CIDOC CRM. Nevertheless we chose this approach because tools for maintenance, web access interfaces or connectivity to GIS are well established and standardized. A web based user interface for oracle is developed to enter, edit and display five upper level CIDOC CRM classes and their relations interactively. Data that have a relation to a spatial location can be displayed in a GIS and spatial objects can be entered and edited with a GIS interface. An essential part in the whole system is the creation and handling of a hierarchical thesaurus.

<http://www.uibk.ac.at/himat/>

---

## III. IMPLEMENTING THE CIDOC-CRM

### CIDOC-CRM for sharing, CIDOC-CRM for use

Graham Klyne, CLAROS-Project, Oxford

In this contribution, I would like to explore some modelling issues we have encountered in building our CLAROS data web, and in particular the tension between dual roles of Ontologies in general, and

CIDOC-CRM in particular, as means for sharing information and as a means for providing an information and data model for use within a specific application.

<http://www.clarosnet.org/index.htm>

---

## Das CIDOC CRM in der Praxis - Erfahrungen, Probleme und mögliche Lösungen

Bernhard Haselhofer, Research Group Multimedia Information Systems, Universität Wien  
Philipp Nussbaumer, Information Management Research Group, Universität Zürich

Das CIDOC Conceptual Reference Model (CRM) gilt als viel versprechende Lösung zur Herstellung von Interoperabilität zwischen heterogenen Metadaten in der Domäne des kulturellen Erbes. In der Praxis offenbaren sich beim Einsatz des CRM jedoch oftmals Schwierigkeiten: das CRM beschreibt eine formale Ontologie und definiert die Semantik der beinhalteten Konzepte; Abbildungsrichtlinien oder technische Spezifikationen - etwa wie ein konkretes Metadaten-Modell in das CRM zu überführen oder das CRM technisch zu repräsentieren ist - werden aber dezidiert ausgespart. Dies führt in konkreten Anwendungen oft zu abweichenden Abbildungen (Mappings) zwischen proprietären Quell-Modellen und dem CRM. In unserem Vortrag diskutieren wir die grundsätzlichen Probleme, die beim praktischen Einsatz des CRM typischerweise auftreten. Darüber hinaus stellen wir eine Methodik vor, welche eine konsistentere Abbildung (Mapping) von Modellen auf das CRM ermöglichen und damit die Interoperabilität der abgebildeten Metadaten erhöhen soll.

---

## The CRM-English Heritage extensions of CIDOC CRM

Keith May, English Heritage, Centre für Archaeology Portsmouth

This presentation will give some of the background to the most recent developments of the STAR project, to be discussed further by Tudhope and Binding. The STAR project has used the ontological model developed at English Heritage's Centre for Archaeology to develop a number of tools, web services and prototype interfaces to enable and demonstrate interoperability between a range of previously unconnected datasets derived from different database sources and platforms.

This paper will look in more detail at how and why the CRM-EH extensions for archaeology were developed and the degree to which they reflect a broader view of archaeological processes. It will also show how the CRM-EH is based on, and relates to, the CIDOC CRM and will discuss some of the key issues about the differences between ontological modelling and data mapping that have been encountered by our STAR project implementation. The aim will be to give an indication of how others could relate their archaeological records and scientific data to such an ontological framework.

<http://hypermedia.research.glam.ac.uk/kos/CRM/>

---

## Breaking down barriers to interoperability

Douglas Tudhope & Ceri Binding, Hypermedia Research Unit, University of Glamorgan

What does interoperability mean? Semantic interoperability at the data level implies the ability to seamlessly interchange, utilise and understand data from multiple sources. There are however successive levels of interoperability with many barriers to achieving them. Conforming or mapping to an overarching common conceptual model such as the CRM is an important first step but there are further issues to consider. Within the STAR project (Semantic Technologies for Archaeological Resources) we have achieved a level of semantic interoperability for effectively cross searching multiple archaeological datasets using the CRM with domain specific extensions. In this talk we will outline some important practical issues encountered such as granularity of modelling, interpretation and implementation of mappings, consistent data formats and controlled vocabularies. We will discuss the approach we have taken, our anticipated further work and ongoing issues.

<http://hypermedia.research.glam.ac.uk/kos/STAR/>

---

## IV. ALTERNATIVES TO CIDOC-CRM

The data exchange format ADeX of the Archaeological Heritage Management in Germany – Intentions and first thoughts about the mapping with CIDOC-CRM

Axel Posluschny, Römisch-Germanische Kommission des DAI, Frankfurt

ADeX (Archaeological DataExport-Standard) is an data exchange format that has been developed for the use of the Archaeological Heritage Management of the German federal states. It is a standard for the of archaeological sites, in future it will be used for exchanging and retrieval of data of any other Cultural Heritage objects. ADeX has been developed according to the precepts of the Core Data Standard for Archaeological Sites and Monuments (created by CIDOC, ICOM and the archaeology documentation group of the Council of Europe 1995). Unlike CIDOC CRM it is based on a very simple table structure. The paper will present an introduction to the aims of the format and its topology and will also present some still very preliminary aspects of mapping ADeX in CIDOC CRM.

[http://www.landesarchaeologen.de/kommissionen/komm\\_informationssysteme/ag\\_is\\_modellgruppe\\_00.html](http://www.landesarchaeologen.de/kommissionen/komm_informationssysteme/ag_is_modellgruppe_00.html)

---



## Archaeology, Formality and the CIDOC CRM

Leif Isaksen, School of Electronics and Computer Science, University of Southampton

The CIDOC CRM is the most sophisticated, best documented and well-known ontology in the Cultural Heritage domain. So much so, that it is frequently referred to as a 'miracle cure' and 'the only show in town'. Yet despite this perception, the rate of its adoption – like that of the Semantic Web with which it is frequently associated – has been glacial at best and almost exclusively by large, well-funded projects. What is hindering uptake and are there important lessons to be learned from it?

In their 1999 paper 'Formality Considered Harmful', Shipman and Marshall identify four barriers to user interaction with formal knowledge systems: (1) The cognitive overhead required to understand the formalism, (2) The need to elicit tacit knowledge, (3) enforcing premature structure on unstructured or poorly-understood source material, (4) the problems caused by situational structure, i.e. the different needs of different users. While they note that there is no 'silver bullet' that addresses all of these challenges they do propose several palliatives that can assist, and therefore encourage, the transition from free to structured information where beneficial.

This paper will discuss these principals in reference to current doctoral research being undertaken in archaeological data integration. While the work in question has elected to use ontologies other than the CIDOC CRM, the results derived are also likely to be of interest to the CRM community. In particular it focuses on means by which microproviders – owners of the small but important datasets that form the 'long tail' of excavation data – can participate in semantics-driven datasharing.