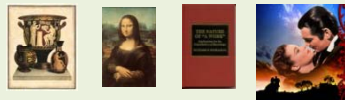


Mining Maps of Information Objects: An Exploratory Ontological Excursion: Sergey Zhrebchevsky, Nicolette Ceo, Michiko Tanaka, David Jank, Richard Smiraglia and Stephen Stead

What is an information object?

An artifact that is considered informative, and also is potentially retrievable

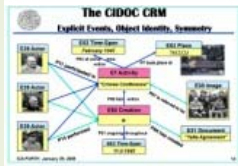
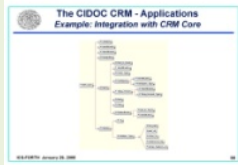
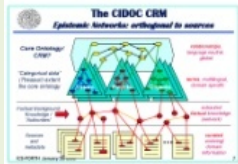
Examples:



document, painting, vase, audio, motion picture, etc.

What is the CRM?

A meta-level ontology for representing artifacts for information sharing



See, for example:

Doerr, Martin, and Athina Kritsotaki. 2006. Documenting events in metadata: 2006. In *Proceedings of the 7th international symposium on virtual reality, archaeology and cultural heritage (VAST)*, Aire-la-Ville, Switzerland: Eurographics. <http://www.ics.forth.gr/isi/publications/paperlink/fin-paper.pdf> (accessed 02 May 2008)

Lin, Chia-Hung, Jen-Shin Hong, and Martin Doerr. 2008. Issues in an inference platform for generating deductive knowledge: A case study in cultural heritage digital libraries using the CIDOC CRM. *International Journal on Digital Libraries*, 8: 115-132.5

Abstract:

The societal archive (virtual though it might be) is made up of information objects of all kinds—everything from stories to talismans to buttons to films to documents. To date the boundaries around domains within which these objects are stored have dictated both the ways in which we classify objects and the ways in which, therefore, they may be retrieved and used for scholarship. According to Hjørland (2003, 98) rationally deduced schema predetermine the potential use of intellectual content by limiting its retrieval. The empirical derivation of units of classification, on the other hand, particularly in developing or evolving systems, provides a basis upon which conceptual systems can be built.

We wish to suggest an approach to the mapping of information objects—that is, all artifacts that are considered informative and therefore might be retrievable through an information system such as a catalog, database, or search engine—that will reveal undiscovered semantic relationships. In this presentation we will use the *CIDOC Conceptual Reference Model* (hereafter *CIDOC CRM*) to map select sets of information objects. The *CIDOC CRM* is an ontology designed for the representation of artifacts and the integration of cultural information.

We hope to demonstrate the compatibility of multi-disciplinary (and therefore multi-epistemological) mappings of the components of information. That is to say, we will map the objects themselves, we will map them as artifacts resident in a repository, and we will map instantiations of their representations—the thing, the nature of its existence, and all searchable iterations of it.

The immediate objective of our project is to map a large set of instantiated information objects using the *CRM*, then to use *data-mining techniques* to discover as yet unknown patterns among the combinations of entities, properties, and relationships. Similar patterns will be grouped using the tools of naive classification. Our poster will demonstrate the mapping of selected objects, and the nascent classification of properties.

There is evidence that societal forces lead to the evolution of large sets of information objects, and that these sets display cultural imprints in their accumulated properties, attributes, and relationships. One way to enlarge the societal archive is to mine it for heretofore unrealized meaning. Our purpose is to discover epistemologically diverse groupings of characteristics of information objects. One might even hope that such a classification would reveal as yet unknown relationships among information objects that might help to create better pathways in the semantic web, or as Patrick Wilson (1968) might have said, to create greater exploitative power.

Mapping with the CRM

Satyr Head



(http://www.museum.upenn.edu/new/worlds_intertwined/etruscan/architecture.shtml).

Material {E22 (→P45: E57)}(← P108 : E12 → P126: E57))
Artifact {E22 → P2: E55}
Item Number {E22 → P48: E42}
Source {E22 (←P30: E10 → P29 : E40 → P131: E82)}(← P24: {E8 → P22: E40 → P131: E82}){E8/E10 → P7: E53 → P87: E44} + (: E53 → P27 (: E9 ← P25: E53 → P87: E44) + (: E9 → P14 : E40) + (→ P14.1: E55))
Date {E22 (→P108: E12 → P4: E52 → P78: E49)}(: E52 → P82: E61))
Image {E22 (→P138: E36)}(→P1: E75))

Deck Log



(United States Merchant Marine Academy, Class of 1942 Archives, Series IX)

Collection {E78 ← (P46: E78)}
Item {E22 ← (P46: E78)}
Creator {E22 ← ((P108: E12) → (P14: E82){P14.1: E55 → P131: E82})}
Image {E22 (→P138: E36)}(→P1: E75))

1. Mine the Mappings

Calculate the probability of occurrence of:

- Properties (forward, backward)
- Classes

Calculate the probability of co-occurrence of:

- Properties (forward, backward)
- Classes

Generate "conceptual" classes using probabilities as "IsA" set descriptors

Examples—Constitute the classes that include all objects for which the following are true:

- P45 Consists of *co-occurs with* ← P24 changed ownership
- P108 was produced by *co-occurs with* E10 transfer of custody

2. Match the Mappings

Discover heretofore unknown similarities among multi-epistemological matches by matching mappings

Such that the Satyr Head and the Deck Log both fall within the class:

E22 → P108 item was produced by Etc.

Potential List of Mapped Collections

- American National Standards Institute (Washington)
- Art Museums Image Consortium (AMICO) (Toronto) *ceased operation but maintains project data at www.amico.org*
- Cantabria (Spain) Cultural Heritage Project <http://www.fundacionmbotin.org>
- The Centre de Recherche et de Restauration des Musees de France* (Paris)
- Centre for Cultural Informatics (ICS-FORTH) (Heraklion, Crete, Greece)
- Chinese Military Academy of Management (Beijing)
- COVAX (European Contemporary Culture Virtual Archives) <http://www.covax.org>
- MALVINE (Manuscripts and Letters Via Integrated Networks in Europe) <http://www.malvine.org>
- Musee de Cherbourg *members of SCULPTEUR Project* (Paris)
- Museum of Finland (FinnONTO Project) (Helsinki)
- The Museum Project (Oslo, Norway)
- NASA (Washington)
- National Chi-Nan University (Taiwan)
- The National Gallery* (London)
- National Museum and Heritage Board (Singapore)
- Norwegian University Museums Database Project (Oslo)
- Uffizi Museum* (Florence)
- The Victoria and Albert Museum* (London)

Works Cited

- Beghtol, C. (2003). Classification for information retrieval and classification for knowledge discovery: Relationships between "professional" and "naive" classifications. *Knowledge organization* 30: 64-73.
- Dahlberg, I. (2006). Knowledge organization: a new science? *Knowledge organization* 33: 11-19.
- Doerr, Martin, and Stead, Steve. (2008). The CIDOC CRM: a standard for the integration of cultural information. Powerpoint for The Center for Cultural Information, Glasgow, Scotland, January 29, 2008. http://cidoc.ics.forth.gr/crm_for_glasgow_2008.ppt
- Wilson, Patrick. [1968] 1978. *Two kinds of power: an essay in bibliographical control*. California library reprint series ed. Berkeley: University of California Press.