Learning Ontology & CIDOC CRM

12/10/2019 - G Bruseker

Learning Ontology & CIDOC CRM **1**

Ontology 2

Bibliography 3

Online Resources 3

CIDOC CRM 5

Bibliography 5

Online Resources 5

Modelling / Mapping Resources 7

Bibliography 7

Online Resources 7

Ontology

Formal ontology, conceptual modelling, knowledge representation, knowledge engineering, conceptual engineering… oh my?

In the mid-90s the idea of ‘knowledge representation’ takes off. Short things short, people keep building databases for the same purpose in different ways. Information systems sell the possibility of efficient and comprehensive analysis of data, but the manner of building information systems themselves, in ad hoc and unprincipled manner (barring formal methods for the structure e.g.: ‘third form’ rules etc.), means that data is incompatible and sharing information or moving it from one system to another is burdensome and expensive. In principle, the more data you bring together, the more knowledge, you would hope, you might derive from its comparison.

Thus is born the idea to marry computer science and philosophy through knowledge representation to create ‘formal ontologies’. The idea borrows from the philosophical notion of ‘ontology’, the ‘word’ of ‘being’. The notion is that philosophers have actually been thinking about how to categorize and organize the world for several millennia and are, furthermore, responsible for the formal logic on which computer science is based. Ergo a collaboration with philosophers to understand good epistemological principles for approaching the fundamental description of the world is sought. Lofty ideas of a new discipline is born (see B. Smith especially). Philosophers will turn to collaborations with specialists and computers scientists to build not ontologies but formal ontologies. The key difference? The philosophical tradition aimed for ontologies that would capture the world as such, the ontology produced would be a True description of the world (at the categorical level) with a capital T. Formal ontologies were to be functional. They remain ontological because they aim for a coherent and consistent account of the empirical world based on expert knowledge. They are merely ‘formal’, however, in that they are only meant to be a consistent and coherent account of the world relative to an empirical domain of discourse. They are not committed, initially, to creating a coherent single picture of the world.

The range and function of what a formal ontology is for is debated and debatable. Early interest in the field was generated by selling its massive capacity for uniting data and generating knowledge across disciplines and borders. The power of the formal ontology to allow formal reasoning over the network of data created with it was sold as a means to make new discoveries just by traversing data graphs. Significant investment goes in to biomedical research. Top players in the early game are Smith, Guarino, Gruber, Davis and Doerr. Much of their work is highly theoretical. Different base formal ontologies are generated by research groups associated to these leading figures. Many of the ontologies created are suggested as ‘top level’ ontologies which aimed to provid highest level abstract classes and relations under which more precise, domain level knowledge can be modelled consistently.

CIDOC CRM is among the early ontology model contenders. It is developed under the patronage of the International Council of Museums. The aim is to develop an ontology capable of integrating museum collections data. CIDOC CRM is built from the ground up using event based modelling. It does not adopt any of the proposed top level ontologies. It stays close to the museum world. As a result it is viewed as more practical than its competitors but often understood to be only of use for museum data. The abstraction it arrives at is useful for describing meso-scopic human level events, meetings of things, ideas, and actors in time and space. It therefore occupies an ambiguous space between a domain ontology (only appropriate to its domain, ie museums) and a top level ontology (applicable to any scenario).

The interest of the cultural heritage community to adopt and implement an ontology gives strength to CIDOC CRM as a practically usable ontology. This leads to many experiments in its use and significant development work in creating 'extensions' of the base model to cover particular domains of research. These extensions are harmonized to the base CIDOC CRM, adding to its new use and interpretation as a 'top level' ontology.

The following selection of texts are some early texts around conceptual modelling and formal ontology, what it is, what it is for, how it can be done.

Bibliography

Davis, Randall, Howard Shrobe, and Peter Szolovits. 1993. What is a knowledge representation? *AI magazine* 14: 17. Click
Doerr, Martin. 2003. The CIDOC conceptual reference module: an ontological approach to semantic interoperability of metadata. *AI magazine* 24: 75. Click
Gruber, Thomas R. 1995. Toward principles for the design of ontologies used for knowledge sharing? *International journal of human-computer studies* 43: 907–928. Click
Guarino, Nicola. 1997. Understanding, building and using ontologies. *International Journal of Human-Computer Studies* 46: 293–310. Click
Smith, Barry. 2003. Ontology. Click

Online Resources

General:

Formal Ontology Basics: (George Bruseker)
http://training.parthenos-project.eu/sample-page/formal-ontologies-a-complete-novices-guide/

Video:

What is a formal ontology? - George Bruseker Click
What is data heterogeneity? - George Bruseker Click

CIDOC CRM

CIDOC CRM is a formal ontology with the scope of cultural heritage data, with the aim of being used for data integration. The ontology is encoding agnostic but it typically expressed in RDF or OWL documents. The model is meant to be used in order to translate existing data into a common form, integrating datasets across institutions and disciplines. The model is managed by a volunteer community that meets 3 times yearly. The model is expressed primarily in a specification document delivered as a doc/pdf. It is also delivered in an encoding for practical use. The official encoding is done in both RDF and OWL. Because of the rules of RDF and OWL, there may be additional elements in these encodings than exist in the base specification.

Learning CIDOC CRM entails understanding the function and form of an ontology and then a close study of the model itself, its specific classes and relations and the

Bibliography

Bruseker, George, Nicola Carboni, and Anaïs Guillem. 2017. Cultural heritage data management: the role of formal ontology and CIDOC CRM. In *Heritage and Archaeology in the Digital Age*, 93–131. Springer. Click

Doerr, Martin. 2003. The CIDOC conceptual reference module: an ontological approach to semantic interoperability of metadata. *AI magazine* 24: 75. Click

Doerr, Martin, Christian-Emil Ore, and Stephen Stead. 2007. The CIDOC conceptual reference model: a new standard for knowledge sharing. In *Proceedings Tutorials, posters, panels and industrial contributions at the 26th International Conference on Conceptual Modeling*, 51–56. Auckland, New Zealand: Australian Computer Society, Inc. Click

Online Resources

Encodings Basics

What is XML: https://www.w3schools.com/xml/
What is RDF: https://www.w3schools.com/xml/xml_rdf.asp
What is OWL: http://w3schools.sinsixx.com/rdf/rdf_owl.asp.htm

CIDOC CRM Basics

The CIDOC CRM website: http://www.cidoc-crm.org/

The CIDOC CRM specification documents: http://www.cidoc-crm.org/versions-of-the-cidoc-crm
Current RDF Official Community Version of CRM:
http://www.cidoc-crm.org/sites/default/files/cidoc_crm_v6.2.1-2018April.rdfs
Current OWL Official Community Version of CRM:
http://erlangen-crm.org/current-version
The official CIDOC CRM extensions list: http://www.cidoc-crm.org/collaborations

CIDOC CRM Learning Tools:

OntoMe (navigate and search classes easily): http://ontome.dataforhistory.org/
CRM Online Search: http://www.cidoc-crm.org/Version/version-6.2.1
Visual Graphs of Common Modelling Scenarios: http://www.cidoc-crm.org/functional-units
Downloadable ZIP of CRM as 'hypertext': http://www.cidoc-crm.org/use_and_learn

CRM Modelling Patterns:
Swiss Art Research Infrastructure Standard Models: https://docs.swissartresearch.net/
Linked.art Models: https://linked.art/

Video:

CIDOC CRM Tutorial - Stephen Stead Click
Introduction to CIDOC CRM - George Bruseker Click

Modelling / Mapping Resources

Once you have learned the basic ideas behind formal ontologies, CIDOC CRM and XML and RDF/OWL encoding you are ready to move on to data modelling and mapping. In data modelling and mapping you create a conceptual model for a data structure using CIDOC CRM and then use different mapping tools in order to translate existing data sets into this model. The outcome of a mapping and transformation process is RDF encoded data which can be searched and navigated using semantic queries. Data in RDF is typically stored in a type of database known as a triple store or graph database. These databases do not usually have native interfaces. For this reason, several organizations have worked on developing user friendly interfaces for navigating and querying CIDOC CRM encoded RDF data. Information about the modelling and mapping process, different tools for mapping and platforms for storing and exploring data are found below.

Bibliography

Doerr, Martin, Maria Theodoridou, Edeltraud Aspöck, Anja Masur, and ÖAW–Österreichische Akademie der Wissenschaften. 2015. Mapping archaeological databases to CIDOC CRM. In *CAA 2015 KEEP THE REVOLUTION GOING. Proceedings of the 43rd Annual Conference on Computer Applications and Quantitative Methods in Archaeology*, 443–452.

Oldman, Dominic, Martin de Doerr, Gerald de Jong, Barry Norton, and Thomas Wikman. 2014. Realizing Lessons of the Last 20 Years: A Manifesto for Data Provisioning and Aggregation Services for the Digital Humanities (A Position Paper) System. *D-Lib Magazine* 20. https://doi.org/10.1045/july2014-oldman.

Online Resources

Mapping tools:

3M Mapping Tool Online Free Service: http://139.91.183.3/3M/
Karma Tool: https://usc-isi-i2.github.io/karma/

Semantic Data Platforms:

Arches Project [Getty]: https://www.archesproject.org/
Arches Demo: http://v4demo.archesproject.org/

ResearchSpace [Mellon]: https://www.researchspace.org/
ResearchSpace Demo: https://demo.researchspace.org/resource/rsp:ExampleResources
Wisski [German National Museum]: http://wiss-ki.eu/
Wisski Demo: http://gesichter-des-dka.gnm.de/50JahreDKA

Video:

Setting up a CIDOC CRM Modelling/Mapping Project - George Bruseker Click
Mapping to CIDOC CRM - George Bruseker Click