

CIDOC CRM-SIG Workshop Day

This workshop will consist of three presentations and following discussions:

1) CIDOC CRM Tutorial (presenter Christian-Emil Ore):

The CIDOC Conceptual Reference Model (CIDOC CRM) and ISO21127, is a semantically rich ontology that delivers data harmonisation based on empirically analysed contextual relationships rather than relying on a traditional fixed field/value approach, overly generalised relationships or an artificial set of core metadata. It recognises that cultural data is a living growing resource and cannot be commoditised or squeezed into artificial pre-conceived boxes. Rather, it is diverse and variable containing perspectives that incorporate different institutional histories, disciplines and objectives. The CIDOC CRM retains these perspectives yet provides the opportunity for computational reasoning across large numbers of heterogeneous sources from different organisations, and creates an environment for engaging and exploration through its network of relationships. The core ontology supports the whole cultural heritage community including museums, libraries and archives and provides a growing set of specialist extensions. This tutorial will present requirements for the model, its form and selected concepts, including new concepts in version 6.0. It will further present a practical exercise how to transform data to a CRM-compatible form.

See also: www.cidoc-crm.org

2) CRMSci Scientific Observation Model and CRMarcheo
(presenter Chrysoula Bekiari)

The CRMSci is a formal ontology intended to be used as a global schema for integrating metadata about scientific observation, measurements and processed data in descriptive and empirical sciences such as biodiversity, geology, geography, archaeology, cultural heritage conservation and others in research IT environments and research data libraries. It uses and extends the CIDOC CRM (ISO21127) as a general ontology of human activity, things and events happening in spacetime. The Scientific Observation Model has been developed bottom up from specific metadata examples from biodiversity, geology, archeology, cultural heritage conservation and clinical studies, such as water sampling in aquifer systems, earthquake shock recordings, landslides, excavation processes, species occurrence and detection of new species, tissue sampling in cancer research, 3D digitization, based on communication with the domain experts and the implementation and validation in concrete applications. It takes into account relevant standards, such as INSPIRE, OBOE, national archeological standards for excavation, Digital Provenance models and others.

CRMarchaeo is an extension of CIDOC CRM with the aim to encode metadata about the archaeological excavation process. It is being developed in the framework of the ARIADNE European Research Infrastructure for Archaeology. The goal of this model is to provide the means to document excavations in such a way that the following functionality is supported: Maximize interpretation capability after excavation or to continue excavation; reason of excavation (research goals); possibility of knowledge revision after excavation; comparing previous excavations on same site (space) and all kinds of comprehensive statistical studies (“collective behavior”).

The presentation will give an introduction to both ontologies including real-life examples of use. Both ontologies have been proposed to CIDOC CRM SIG for revision and approval as CIDOC compatible extensions.

3) The Synergy Reference Model of Data Provision and Aggregation
(presenter Dominic Oldman)

The increased use of aggregation services and the growing use of the CIDOC CRM has necessitated a new initiative to develop a data provisioning reference model targeted at solving fundamental infrastructure problems ignored by data integration initiatives to date. These problems include data quality issues (the lack of adequate semantic and contextual meaning), the use of fixed field/value models that remove local perspectives, the lack of integration with data providers and their expertise and knowledge, and the lack of tangible benefits for providers and users alike. Instead the proposed model is designed to be distributed and collaborative and not divorced from providers as in more centralized systems. In reality it is the information providers that curate and understand the source resources and that update these resources at regular intervals. The provider is the one who can verify or falsify statements about the evidence in their hands. The role of the aggregator includes the responsibility for the homogeneous access and the synopsis of consistency, whereas any inconsistencies should be made known to, and can only be resolved by, the original providers. However, the process of data transformation to the aggregator’s target system requires a level of quality control that is often beyond the means of prospective providers. All current transformation tools fail to support integration of data from a large number of providers that inevitably undergoes continuous data, format and semantic changes at both the provider and aggregator side. The proposed reference models specifies particular business processes, S/W components and Open S/W interfaces for a comprehensive solution to this problem. The presentation will present the complete rationale and give an overview over the workflow and components.