**Introduction**

This document is the formal definition of the**CIDOC Conceptual Reference Model (“CRM”),**a formal ontology intended to facilitate the integration, mediation and interchange of heterogeneous cultural heritage information and similar information from other domains. The CRM is the culmination of more than two decades of standards development work by the International Committee for Documentation (CIDOC) of the International Council of Museums (ICOM). Work on the CRM itself began in 1996 under the auspices of the ICOM-CIDOC Documentation Standards Working Group. Since 2000, development of the CRM has been officially delegated by ICOM-CIDOC to the CIDOC CRM Special Interest Group, which has been collaborating soon after with the ISO working group ISO/TC46/SC4/WG9 to bring the CRM to the form and status of an International Standard. This collaboration has resulted in ISO21127:2004 and ISO21127:2014, and will be continued to produce the next update of the standard. This document belongs to the series of evolving versions of the formal definition of theCRM, which serve the ISO working group as community draft for the standard. Eventual minor differences of the ISO standard text from the CIDOC version in semantics and notation that the ISO working group requires and implements are harmonized in the subsequent versions of the CIDOC version.

**Objectives of the CIDOC CRM**

The primary role of the CRM is to enable the exchange and integration of information from heterogeneous sources for the reconstruction and interpretation of the past at a human scale [specify *time*-scale], based on all kinds of material evidence, including texts, audiovisual material and even oral tradition. It starts from, but is not limited to, the needs of museum documentation and research based on museum holdings. It aims at providing the semantic definitions and clarifications needed to transform disparate, localised information sources into a coherent global resource, be it within a larger institution, in intranets or on the Internet, and to make it available for scholarly interpretation and scientific evaluation. Its perspective is supra-institutional and abstracted from any specific local context [or language]. This goal determines the constructs and level of detail of the CRM.

More specifically, it defines, in terms of a formal ontology, the **underlying semantics** of database **schemata** and **structured** documents used in the documentation of cultural heritage and scientific activities. In particular it defines the semantics related to the study of the past and current state of our world, as it is characteristic for museums, but also or other institutions and disciplines. It does **not** define any of the **terminology** appearing typically as data in the respective data structures; however it foresees the characteristic relationships for its use. It does **not** aim at proposing what cultural institutions **should** document. Rather it explains the logic of what they actually currently document, and thereby enables **semantic interoperability.**

It intends to provide a model of the intellectual structure of the respective kinds of documentation in logical terms. As such, it is not optimised for implementation-specific storage and processing aspects. Implementations may lead to solutions where elements and links between [delete relevant] elements of our conceptualizations are no longer explicit in a database or other structured storage system. For instance, the birth event that connects elements such as father, mother, birth date, birth place may not appear in the database, in order to save storage space or response time of the system. The CRM allows us to explain how such apparently disparate entities are intellectually interconnected, and how the ability of the database to answer certain intellectual questions is affected by the omission of such elements and links.

**Scope of the CIDOC CRM**

The overall scope of the CIDOC CRM can be summarised in simple terms as the curated, **factual knowledge** [do you want to exclude beliefs and inferences here?] about the past at a human scale.

However, a more detailed and useful definition can be articulated by defining both the **Intended Scope**, a broad and maximally-inclusive definition of general application principles, and the Practical Scope, which is expressed by the overall scope of a growing reference set of specific, identifiable documentation standards and practices that the CRM aims to encompass, however restricted in its details to the limitations of the Intended Scope. --> too difficult for me

The reasons for this distinction are twofold. Firstly, the CRM is developed in a “**bottom-up**” manner, starting from well-understood, actually and widely used concepts of domain experts, which are disambiguated and gradually generalized as more forms of encoding are encountered. This allows for avoiding the misadaptations and vagueness often found in introspection-driven attempts to find overarching concepts for such a wide scope, and provides stability to the generalizations found. Secondly, it is a means to identify and keep a focus on the concepts most needed by the communities working in the scope of the CRM and to maintain a well-defined agenda for its evolution.

The **Intended Scope** of the CRM may be defined as all information required for the exchange and integration of heterogeneous scientific and scholarly documentation about the past at a human time-scale and its evidence that has come upon us. This definition requires further elaboration:

* The term “scientific and scholarly documentation” is intended to convey the requirement that the depth and quality of descriptive information that can be handled by the CRM should be sufficient for serious academic research. This does not mean that information intended for presentation to members of the general public is excluded, but rather that the CRM is intended to provide the level of detail and precision expected and required by museum professionals and researchers in the field.
* As “evidence that has come upon us” are regarded all types of material collected and displayed by museums and related institutions, as defined by ICOM[[1]](https://mail.google.com/mail/u/0/#m_3886930029748061135__ftn1), and other  collections, in-situ objects, sites, monuments and intangible heritage relating to fields such as social history, ethnography, archaeology, fine and applied arts, natural history, history of sciences and technology. Does this cover, e.g., a soil MS measurement? Not clear to me...
* The documentation includes the detailed description of individual items, in situ or within collections, groups of items and collections as a whole, as well as practices of intangible heritage. It pertains to their current state as well as to information about their past. The CRM is specifically intended to cover contextual information: the historical, geographical and theoretical background that gives cultural heritage collections much of their cultural significance and value.
* The exchange of relevant information with libraries and archives, and the harmonisation of the CRM with their models, falls within the Intended Scope of the CRM.
* Information required solely for the administration and management of cultural institutions, such as information relating to personnel, accounting, and visitor statistics, falls outside the Intended Scope of the CRM.

The Practical Scope[[2]](https://mail.google.com/mail/u/0/%22%20%5Cl%20%22m_3886930029748061135__ftn2%22%20%5Co%20%22%22%20%5Ct%20%22_blank) of the CRM is expressed in terms of the set of reference standards and de facto standards for documenting factual knowledge that have been used to guide and validate the CRM’s development and its further evolution. The CRM covers the same domain of discourse as the union of these reference standards; this means that for data correctly encoded according to these documentation formats there can be a CRM-compatible expression that conveys the same meaning.  --> too difficult for me

**Coverage and Extensions**

The intended scope of the CRM is a subset of the “real” world and is therefore potentially infinite [what does that mean?]. Further, the strategy to develop the model bottom-up from a practical scope has the consequence that the model will always miss some areas of relevant application or, on the other hand, some parts may not be developed in sufficient detail for a specialized field of study, such as *E30 Right*. Therefore, the CRM has been designed to be extensible by different mechanisms in order to achieve an optimal coverage of the intended scope without losing compatibility with the CRM.

Strict **compatibility of extensions** with the CRM means that data structured according to an extension must also remain valid as a CRM instance. In practical terms, this implies *query containment:*any queries based on CRM concepts should retrieve a result set that is correct according to the CRM’s semantics, regardless of whether the knowledge base is structured according to the CRM’s semantics alone, or according to the CRM plus compatible extensions. For example, a query such as “list all events” should recall 100% of the instances deemed to be events by the CRM, regardless of how they are classified by the extension.

A sufficient condition for the compatibility of an extension with the CRM is that CRM classes subsume all classes of the extension, and all properties of the extension are either **subsumed** by CRM properties, or are **part of a path** for which a CRM property is a shortcut. Obviously, such a condition can only be tested intellectually.

The mechanisms for extensions are:

1. Existing classes and properties can be extended dynamically using thesauri and controlled vocabularies with CRM properties having as range *E55 Type*, as further elaborated in the section “About Types”. This approach is preferable when specializations of classes are independent from specializations of properties [don't follow this...], and for local, non-standardized concepts.
2. Existing classes and properties can be extended structurally by adding subclasses and subproperties respectively. This approach is particularly recommended to communities of practice needing well-established properties specific to classes that are not present in the CRM.
3. Additional information that falls outside the semantics formally defined by the CRM can trivially be recorded as unstructured data using *E1 CRM Entity. P3 has note: E62 String* to attach such information to the most adequate instance in the respective knowledge base. This approach is preferable when detailed, targeted queries are not expected; in general, only those concepts used for formal queryingneed to be explicitly modelled.

**Conservative Extensions of Scope**

Extensions may be incorporated in **new versions** of the CRM, or become **semi-independent modules** maintained in parallel to the CRM by communities of practice. In mechanisms 1 and 2 above, the CRM concepts subsume and thereby cover the extensions. This specialization as only method of extension would mean that the CRM from the beginning has foreseen all necessary high-level classes and properties. This comes in conflict with the very successful bottom-up methodology of evolution of the CRM itself and the development of extensions more peripheral to the current practical scope.

Extensions that are the result of widening the scope, rather than elaborating it in more detail, may quite well find a class “C” not covered by the CRM so far and even a superclass “B” of class C that must be regarded as a superclass of an existing CRM class “A”. From a logical-theoretical point of view, we precisely regard such extensions as compatible, if the CRM classes subsume all classes and all properties of the extension as long as instances are **restricted to the not extended scope** of the CRM.

In this case, an existing property p of class A may also hold for the new superclass B. We call the latter a **conservative extension**. That is, when restricted to the original class A, the extended property, p’, is identical to the original property p. In general, a superproperty is said to be a conservative extension of a subproperty when it is identical to the subproperty when restricted to its domain and range. In first order logic, the conservative extension of a property can be expressed as follows. Assume that A and C are subclasses of B and D respectively and  that p, p’ are properties between A,C and B, D respectively:

                               A(x)  ⊃ B(x)
                               C(x)  ⊃ D(x)
                               P(x,y) ⊃ A(x)
                               P(x,y) ⊃ C(y)
                               P’(x,y) ⊃ B(x)
                               P’(x,y) ⊃ D(y)

If p’ is a conservative extension of p then

                               A(x) ∧ C(y) ∧ P’(x,y) ≡  P(x,y)

[the above is beyond me, although I must confess I have a cold and am not at my sharpest]

This is similar to what in logic is called a conservative extension of a theory. This construct is necessary for an effective modular management of ontologies, but is not possible with the current way RDF/OWL treats it. It has very important **practical consequences**:

1.       Taken on its own, the CRM is not affected by such an conservative extension of scope, since it is not concerned with instances of class B that are not in class A.

2.       If a conservative extension is incorporated into a **new version** of the CRM, the new version becomes **backwards compatible** with the previous one (therefore it is conservative).

3.     The **bottom-up** development of ontologies encourages to find as domain and range [useful to define these here] of a property not the most general ones for all future, but the **best understood** ones, and leave it to conservative extensions to find more general ones in the future.

4.     Extensions of the CRM maintained in separate modules that declare classes and/or properties not covered by superclasses and/or superproperties of the CRM **should** **clearly** **mark** the highest-level ones to be used by a respective query system in order to retrieve all instances described in terms of the CRM and the extension modules. --> lost me again...

5.     Extensions of the CRM maintained in separate modules must be harmonized with the CRM: All ontologically justified relationships of **subsumption** between the CRM and the extension should **explicitly** be declared and contained in the extension, or, if indicated, be submitted for the CRM to consider their inclusion.

It is the hope that over time the CRM and its compatible extension modules will provide a more and more complete coverage of the intended scope as a coherent logical and ontologically adequate theory of widest practical use. Besides others, this will require a collaboration of the involved communities based on a continuous effort of mutual understanding and respect.

[[1]](https://mail.google.com/mail/u/0/#m_3886930029748061135__ftnref1) The ICOM Statutes provide a definition of the term “museum” at <http://icom.museum/statutes.html#2>

[[2]](https://mail.google.com/mail/u/0/#m_3886930029748061135__ftnref2) The Practical Scope of the CIDOC CRM, including a list of the relevant museum documentation standards, is discussed in more detail on the CIDOC CRM website at <http://cidoc.ics.forth.gr/scope.html>