# Modelling principles

The following modelling principles have guided and informed the development of the CIDOC CRM.

### Reality and Knowledge Bases

The CIDOC CRM is a formal ontology in the sense introduced by N.Guarino [XXXX]. The CRM aims to assist sharing, connecting and integrating information about the past. In order to understand the function of a formal ontology for this aim, one needs to make the following distinctions:

1. The *material reality*. For the purpose of the CIDOC CRM, material reality is regarded as whatever has substance that can be perceived with senses or instruments. Examples are people, a forest or a settlement environment, sea, atmosphere, distant celestial or cellular micro structures, including what we assume that could be potentially or theoretically perceived if we could be there, such as the center of Earth or the sun, and all that is past. It is constraint to space and time. What goes on in *our minds* and produced by our minds is also regarded as part of the material reality, as it becomes materially evident to other people at least by our utterances, behavior and products.
2. The units of description or *particulars*, i.e., the things and relations which we refer to in order to distinguish parts of reality. Examples are Mount Ida, the Taj Mahal, the formation of China by emperor Qin Shi Huang (秦始皇) in 221BC, Tut-Ankh Amun and his embalmment, Prince Shotoku of Japan sending a mission to China in 607AD, the participation of Socrates in the Battle of Potidaea or the radiocarbon dating of the Iceman Ötzi[[1]](#footnote-1).

A formal ontology, such as the CIDOC CRM, constitutes a controlled language for talking about particulars. I.e., it provides descriptions of classes and properties for categorizing particulars as so-called “instances” in a way that their individuation, unity and relevant properties are as unambiguous as possible. For instance, Tut-Ankh Amun as instance of E21 Person *is* the real pharaoh from his birth to death, and not extending to his mummy, as follows from the specification of the class E21 Person and its properties in the CRM.

For clarification, the CRM does not take a position against or in favor of the existence of *spiritual* substance nor of substance not accessible by either senses or instruments, nor does it suggest a materialistic philosophy. However, for practical reasons, it relies on the priority of integrating information based on material evidence available for *whatever* human experience. The CRM only commits to a *unique material reality* independent from the observer.

When we *provide descriptions* of particulars, we need to refer to them by unique names, titles or constructed identifiers, all of which are instances of E41 Appellation in the CRM, if the reference should be independent from context. (In contrast, reference to particulars by pronouns or enumerations of characteristic properties, such as name and birth date, are context dependent). The appellation, and the relation between the appellation and the referred item or relationship, must not be confused with the referred item and its identity. For example, Tut-Ankh Amun the person (instance of E21 Person) is different to Tut-Ankh Amun the name (instance of E41 Appellation) and also different to the relationship between name and person (P1 is identified by). Instances of the CRM are the *real* particulars, not their names. Particulars are approximate individuations, like sections, of parts of reality. In other words, the uniqueness of reality does not depend on where one draws the line between the mountain and the valley.

In contrast, a CRM-compatible *knowledge base* (KB) is an instance of E73 Information Object in the CRM. It contains (data structures that encode) formal statements representing propositions believed to be true in a reality by an observer. These statements use appellations (e.g. [[[2]](#footnote-2)http://id.loc.gov/authorities/names/n79066005[[3]](#footnote-3)](http://id.loc.gov/authorities/names/n79066005The URI (instance of E41 Appellation) of the Library of Congress for Tut-Ankh-Amun, the pharaoh.)) of ontological particulars and of CRM concepts (e.g. *P100i died in*). Thereby users, in their capacity of having real-world knowledge and cognition, may be able to relate these statements to the propositions they are meant to characterize, and be able to reason and research about their validity. In other words, the formal instances in a knowledge base are the *identifiers*, not the real things or phenomena. Therefore, a knowledge base does not contain knowledge, but *statements that represent* knowledge of its maintainers, as long as there exist people that can resolve the used identifiers to their referents. (Appellations described in a knowledge base, and not used as primary substitute of other items, are of course explicitly declared as instance of E41 Appellation in the knowledge base.) Statements in a KB may be contradictory without the KB being broken or invalid in any sense. Contradictions in a KB merely document the fact that different maintainers of the KB hold incompatible beliefs about the same reality. A KB does not represent a slice of reality, but the beliefs of its maintainers about that reality. For simplicity, in what follows we will tacitly understand this fact and speak of a KB as representing some reality.

A KB in a computer system may contain statements about instances of E90 Symbolic Object. If it is possible to represent these instances solely through text, then this texts may reside within the same KB. In this case, the instance of E90 Symbolic Object and its representation are separate entities and they can be connected with the property *P190 has symbolic content.*

### Monotonicity

The CIDOC CRM’s primary function is to support the meaningful integration of information in an Open World. The adoption of the Open World principle means that the CIDOC CRM itself must remain fundamentally open and knowledge bases implemented using it should be flexible to receive new insights. At model level, new classes and properties within the CIDOC CRM’s scope may be found at any time in the course of integrating more documentation records or when new kinds of relevant facts come to the attention of its maintainers. At the level of the knowledge base, the need to add or revise information may arise due to numerous external factors. Research may open new questions; documentation may be directed to new or different phenomena; natural or social evolution may reveal new objects of study.

It is the aim of the maintainers of the CIDOC CRM to respect the Open World principle and to follow the principle of monotonicity. Monotonicity requires that adding new classes and properties to the model or adding new statements to a knowledge base does not invalidate already modelled structures and existing statements.

A first consequence of this commitment, at the level of the model, is that the CIDOC CRM aims to be monotonic in the sense of Domain Theory. That is to say, the existing CIDOC CRM constructs and the deductions made from them should remain valid and well-formed, even as new constructs are added by extensions to the CIDOC CRM. Additions to the model are known as extensions while the main model is known as CRM base. Any extensions should be, under this method, backwards compatible, with previous models. The only exception to this rule arises when a previous construct is considered objectively incorrect by the domain experts and thus subjected to corrective revision. Adopting the principle of monotonicity has active consequences for the basic manner in which classes and properties are designed and declared in the CIDOC CRM. In particular, it forbids the declaration of complement classes, i.e. classes solely defined by excluding instances of some other classes.

For example:

FRBRoo extends the CIDOC CRM. In version 2.4 of FRBRoo, F51 Name Use was declared as a subclass to the CIDOC CRM class E7 Activity. This class was added in order to describe a phenomenon specific to library practice and not considered within CRM base. F51 Name Use describes the practice of an instance of E74 Group adopting and deploying a name within a context for a time-span. The creation of this extension is monotonic because no existing IsA relationship or inheritance of properties in CRM base are compromised and no future extension is ruled out. By way of contrast, if, to handle this situation, a subclass “Other Activity” had been declared, a non-monotonic change would have been introduced. This would be the case because the scope note of a complement class like “Other Activities” would forbid any future declaration of specializations of E7 Activity such as ‘Name Use’. Just in case the need arose to declare a particular specialized subclass, a non-monotonic revision would have to be made, since there would be no principled way to decide which instances of ‘Other Activity’ were instances of the new, specialized class and which not. Such non-monotonic changes are extremely costly to end users, compromising backwards compatibility and long term integration.

As a second consequence, maintaining monotonicity also is also required during revising or augmenting data within a CIDOC CRM compatible system. That is, existing CIDOC CRM instances, their properties and the deductions made from them, should always remain valid and well-formed, even as new instances, regarded as consistent by the domain expert, are added to the system.

For example:

If someone describes correctly that an item is an instance of E19 Physical Object, and later it is correctly characterized as an instance of E20 Biological Object, the system should not stop treating it as an instance of E19 Physical Object. This is achieved by declaring E20 Biological Object as subclass of E19 Physical Object.

This example further demonstrates that the IsA hierarchy of classes and properties can represent characteristic stages of increasing knowledge about some item during the processes of investigation and collection of evidence. Higher level classes can be used to safely classify objects whose precise characteristics are not known in the first instance. An ambiguous biological object may, for example, be classified as only a physical object. Subsequent investigation can reveal its nature as a biological object. A knowledge base constructed with CIDOC CRM classes designed to support monotonic revision allows for seeking physical objects that were not yet recognized as biological ones. This is particularly important for large-scale information integration. Such as a system supports scholars to be able to integrate all information about potentially relevant phenomena into the information system without forcing an over or under commitment on knowledge of the object. Since large scale information integration always deals with different levels of knowledge of its relevant objects, this feature enables a consistent approach to data integration.

A third consequence, applied at the level of the knowledge base, is that in order to formally preserve monotonicity when it is required to record and store alternative opinions regarding phenomena, all formally defined properties should be implemented as unconstrained (many: many) so that conflicting instances of properties are merely accumulated. Thus integrated knowledge can serve as a research tool for accumulating relevant alternative opinions around well-defined entities, whereas conclusions about the truth are the task of open-ended scientific or scholarly hypothesis building.

For example:

King Arthur’s basic life events are highly contested. Once entered in a knowledge base, he should be defined as an instance of E21 Person and treated as having existed as such within the sense of our historical discourse. The instance of E21 Person is used as the collection point for describing possible properties and existence of this individual. Alternative opinions about properties, such as the birthplace and his living places, should be accumulated without validity decisions being made during data compilation. King Arthur may be entered as a different instance, of E28 Conceptual Object, for describing him as mythological character and accumulating possibly mythological facts.

The fourth consequence of monotonicity relates to the use of time dependent properties in a knowledge base. Certain properties declared in the CIDOC CRM, such as having a part, an owner or a location, may change many times for a single item during the course of its existence. Asserting that such a property holds for some item means that that property held for some particular, undetermined time-span within the course of its existence. Consequently, one item may be the subject of multiple statements asserting the instantiation of that property without conflict or need for revision. The collection of such statements would reflect an aggregation of these instances of this property holding over the time-span of the item’s existence. If a more specific temporal knowledge is required/available, it is recommended to explicitly describe the events leading to the assertion of that property for that item. For example, in the case of acquiring or losing an item, it would be appropriate to declare the related event class such as E9 Move. By virtue of this principle, the CRM achieves monotonicity with respect to an increase of knowledge about the states of an item at different times, regardless of their temporal order.

Time-neutral properties may be specialized in a future monotonic extension by time-specific properties, but not vice-versa. Also, many properties registered do not change over time or are relative to events in the model already. Therefore, the CIDOC CRM always gives priority to modelling properties as time-neutral.

However, for some of these properties many databases may describe a “current” state relative to some property, such as “current location” or “current owner”. Using such a “current” state means that the database manager is able to verify the respective reality at the latest date of validity of the database. Obviously, this information is non-monotonic, i.e., it requires deletion when the state changes. In order to preserve a reduced monotonicity, these properties have time-neutral superproperties by which respective instances can be reclassified if the validity becomes unknown or no longer holds. Therefore the use of such properties in the CRM is only recommended if they can be maintained consistently. Otherwise, they should be reclassified by their time-neutral superproperties. This holds in particular if data is exported to another repository.

### Extensions

Since the intended scope of the CRM is a subset of the “real” world and is therefore potentially infinite, the model has been designed to be extensible through the linkage of compatible external type hierarchies.

Of necessity, some concepts covered by the CIDOC CRM are less thoroughly elaborated than others: E39 Actor and E30 Right, for example. This is a natural consequence of staying within the model’s clearly articulated practical scope in an intrinsically unlimited domain of discourse. These ‘underdeveloped’ concepts can be considered as candidate superclasses for compatible extensions, in particular for disciplines with a respective focus.

Compatibility of extensions with the CRM means that data structured according to an extension must also remain valid as instances of CIDOC CRM base classes. In practical terms, this implies query containment: any queries based on CIDOC CRM concepts should retrieve a result set that is correct according to the model’s semantics, regardless of whether the knowledge base is structured according to the CIDOC CRM’s semantics alone, or according to the CIDOC 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 CIDOC CRM, regardless of how they are classified by the extension.

A sufficient condition for the compatibility of an extension with the CIDOC CRM is that its classes other than E1 CRM Entity 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, and that classes and properties of the extension can be well distinguished from those in the CRM. For instance, a class “tangible object” may be in conflict with existing classes of the CIDOC CRM. Obviously, such a condition can only be tested intellectually.

The CRM provides a number of mechanisms to ensure that coverage of the intended scope is complete:

1. Existing classes can be extended, either structurally as subclasses or dynamically using the type hierarchy (see section About Types below).
2. Existing properties can be extended, either structurally as subproperties, or in some cases, dynamically, using properties of properties which allow subtyping (see section About Types below).
3. Additional information that falls outside the semantics formally defined by the CIDOC CRM can be recorded as unstructured data using E1 CRM Entity. P3 has note: E62 String.
4. Extending the CIDOC CRM by superclasses and properties that pertain to a wider scope. They are called conservative extensions, if they preserve backwards compatibility with instances described with the CIDOC CRM.

Following strategies 1,2 and 3 will have the result that the CIDOC CRM concepts subsume and thereby cover the extensions. This means that querying an extended knowledge base only with concepts of the CIDOC CRM will nevertheless retrieve all facts described via the extensions.

**In mechanism 3**, the information in the notes is accessible in the respective knowledge base by retrieving the instances of E1 CRM Entity that are domain of P3 has note. Keyword search will also work for the content of the note. Rules should be applied to attach a note to the item most specific for the content. For instance, details about the role of an actor in an activity should be associated with the instance of E7 Activity, and not with the instance of E39 Actor. This approach is preferable when queries relating elements from the content of such notes across the knowledge base are not expected.

In general, only concepts to be used for selecting multiple instances from the knowledge base by formal querying need to be explicitly modelled. This criterion depends on the expected scope and use of the particular knowledge base. The CIDOC CRM models with priority for queries expected to retrieve related facts across heterogeneous content from different institutions.

**Mechanism 4**, conservative extension, is more complex:

With increasing use of the CIDOC CRM, there is also a need for extensions that model phenomena from a scope wider than the original one of the CIDOC CRM, but which are also applicable to the concepts that do fall within the CIDOC CRM’s scope. When this occurs, properties of the CIDOC CRM may be found to be applicable more generally to superclasses of the extension than to those of their current domain or range in the CIDOC CRM. This is a consequence of the key principle of the CIDOC CRM to model “bottom up”, i.e., selecting the domains and ranges for properties as narrow as they would apply in a well understood fashion in the current scope, thus avoiding making poorly understood generalizations at risk of requiring non-monotonic correction.

The fourth mechanism for extending the CIDOC CRM by conservation extension can be seen to be split into two cases:

1) A new class or property is added to an extension of the CIDOC CRM, which is not covered by superclasses other than E1 CRM Entity or a superproperty in the CIDOC CRM respectively. In this case, all facts described by such concepts only are *not* accessible by queries with CIDOC CRM concepts. Therefore, the extension should *publish* in a compatibility statement the additional relevant high-level classes and properties needed to retrieve all facts documented with the extended model. This case is a monotonic extension.

2) The domain or range of an existing property in the CIDOC CRM is changed to a superclass of the one or the other or both, because the property is understood to be applicable beyond its originally anticipated scope. In this case, all facts described by the extension are still accessible by querying with the concepts of the CIDOC CRM, but the extension can describe additional facts that the CIDOC CRM could not. This case is a monotonic extension and generally recommended, because it enables bottom-up evolution of the model. If this change is part of a new release of the CIDOC CRM itself, it is simply backwards compatible, and this has been done frequently in the evolution of this model.

If case (2) should be documented and implemented in an extension module separately to the CIDOC CRM, it may come in conflict with the current way knowledge representation languages, such as RDF/OWL, treat it, because in formal logic there is no distinction between changing the ontological meaning of a property and changing its range or domain. It is, however, similar to what in logic is called a conservative extension of a theory, and necessary for an effective modular management of ontologies.

Therefore, for the interested reader, we describe here a definition of this case in terms of first order logic, which shows how modularity can formally be achieved:

Let us assume a property P defined with domain class A and range class C also holds for a domain class B, superclass of A, and a range class D, superclass of C, in the sense of its ontological meaning in the real world. We describe this situation by introducing an auxiliary formal property P’, defined with domain class A and range class C, and apply the following logic:

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)

Then, P’ is a conservative extension of P if: A(x) ∧ C(y) ∧ P’(x,y) ≡ P(x,y)

In other words, a separate extension module may re-declare the respective property with another identifier, preferably using the same label, and implement the above rule.

### Minimality

Although the scope of the CIDOC CRM is very broad, the model itself is constructed as economically as possible.

* CIDOC CRM classes and properties are either primitive, or they are key concepts in the practical scope.
* Complements of CIDOC CRM classes are not declared, because, considering the Open World principle, there are no properties for complements of a class (see Terminology and first consequence of Monotonicity).

A CIDOC CRM class is declared when:

* It is required as the domain or range of a property not appropriate to its superclass.
* It serves as a merging point of two CIDOC CRM class branches via multiple IsA (e.g. E25 Human-Made Feature). When the branch superclasses are used for multiple instantiation of an item, this item is in the intersection of the scopes. The class resulting from multiple IsA should be narrower in scope than the intersection of the scopes of the branch superclasses.
* It is useful as a leaf class (i.e. at the end of a CIDOC CRM branch) to domain communities building CIDOC CRM extensions or matching key domain classes from other models to the CIDOC CRM (e.g. E34 Inscription).

### Shortcuts

Some properties are declared as shortcuts of longer, more comprehensively articulated paths that connect the same domain and range classes as the shortcut property via one or more intermediate classes. For example, the property *E18 Physical Thing. P52 has current owner (is current owner of): E39 Actor*, is a shortcut for a fully articulated path from E18 Physical Thing through E8 Acquisition to E39 Actor. An instance of the fully-articulated path always implies an instance of the shortcut property. However, the inverse may not be true; an instance of the fully-articulated path cannot always be inferred from an instance of the shortcut property.

The class E13 Attribute Assignment allows for the documentation of how the assignment of any property came about, and whose opinion it was, even in cases of properties not explicitly characterized as “shortcuts”.

### Disjointness

Classes are disjoint if they share no common instances in any possible world. That implies that it is not possible to instantiate an item using a combination of classes that are mutually disjoint or with subclasses of them (see “multiple instantiation” in section “Terminology”). There are many examples of disjoint classes in the CIDOC CRM.

A comprehensive declaration of all possible disjoint class combinations afforded by the CIDOC CRM has not been provided here; it would be of questionable practical utility, and may easily become inconsistent with the goal of providing a concise definition. However, there are two key examples of disjoint class pairs that are fundamental to effective comprehension of the CIDOC CRM:

* **E2 Temporal Entity is disjoint from E77 Persistent Item.** Instances of the class E2 Temporal Entity are perdurants, whereas instances of the class E77 Persistent Item are endurants. Even though instances of E77 Persistent Item have a limited existence in time, they are fundamentally different in nature from instances of E2 Temporal Entity, because they preserve their identity between events. Declaring endurants and perdurants as disjoint classes is consistent with the distinctions made in data structures that fall within the CIDOC CRM’s practical scope.
* **E18 Physical Thing is disjoint from E28 Conceptual Object.** The distinction is between material and immaterial items, the latter being exclusively human-made. Instances of E18 Physical Thing and E28 Conceptual Object differ in many fundamental ways; for example, the production of instances of E18 Physical Thing implies the incorporation of physical material, whereas the production of instances of E28 Conceptual Object does not. Similarly, instances of E18 Physical Thing cease to exist when destroyed, whereas an instance of E28 Conceptual Object perishes when it is forgotten or its last physical carrier is destroyed.

### Transitivity

CIDOC CRM is formulated as a class system with inheritance. A property P with domain A and range B will also be a property between possible subclasses of A and B. In many cases there will be a common subclass C of A and B. In these cases when the property restricted to C, that is, with C as domain and range, the restricted property could be transitive. For instance, an E73 Information Object can be incorporated in a E90 Symbolic Object and thus an information object can be incorporated in another information object.

In the definition of CIDOC CRM the transitive properties are explicitly marked as such in the scope notes. All unmarked properties should be considered as not transitive.

## Introduction to the basic concepts

The following paragraphs explain the most general logic of the CIDOC CRM. The CIDOC CRM is a formalized representation of historical discourse, a formal ontology. In this capacity, it is meant to support the (re)presentation of fact based, analytic discourse about what has happened in the past in a human understandable and machine-processable manner. It achieves this function by proposing a series of formalized properties (relations) and classes. The formalized properties support the making of semantically explicit statements relating classes of things. Their formal definition logically explicates the classes of things to which they may pertain. The CIDOC CRM properties thus enable a formal, logically explicit description of relations between individual, real world items, classified under distinct ontological classes. Encoding analytic data pertaining to the past under such a system of statements provides a standard representation for data and allows the uniform application of reasoning to large sets of data.

Grounding this high level logic is a hierarchical system of classes and relations, that provide basic ontological distinctions by which to represent historical discourse. Familiarity with the basic ontological distinctions made in the top level of the class hierarchy provides the basic entry point to understanding how to apply the CIDOC CRM for knowledge representation.

The highest level distinction in the CIDOC CRM is represented by the top level concepts of E77 Persistent Item, equivalent to the philosophical notion of endurant; E2 Temporal Entity, equivalent to the philosophical notion of perdurant and, further, the concept of E92 Spacetime Volume.

As an event-centric model, supporting historical discourse, the CIDOC CRM firstly enables the description of entities that are themselves time-limited processes or evolutions within the passing of time using E2 Temporal Entity and its subclasses. Their basic function is to capture the fact of something having happened over time. In addition to allowing the description of a temporal duration, the subclasses of E2 Temporal Entity are used to document the historical relations between objects, similar to the role of a verb in a natural language phrase. The more specific subclasses of E2 Temporal Entity enable the documentation of events pertaining to individually related/affected material, social or mental objects that have been described using subclasses of E77 Persistent Item. This precise documentation is enabled through the use of specialized properties formalizing the manner of the relation or affect. Examples of specific subclasses of E2 Temporal Entity include E12 Production, which allows the representation of events of making things by humans, and E5 Event which allows the documentation, among other things, of geological events and large scale social events such as a war. Each of these subclasses have specific properties associated to them which allow them to function to represent the specific, real world connection between instances of E77 Persistent Item, such as the relation of an object to its time of production through *p108 was produced by* (E12) or the relation of a place to a geological phenomenon through *p7 was place of* (E5). The entities that E2 Temporal Entity documents, being time limited processes / occurrences, are such that their existence can be declared only on the basis of direct observation or recording of the event, or indirect observation of its material outcomes. Evidence of such entities may be preserved on material objects that are permanently changed because of them. Likewise events may have been recorded in text or remembered through oral history. E2 Temporal Entity and its subclasses are central to the CRM and essential for almost all modelling tasks (e.g. in a museum catalogue one cannot consider an object outside its production event).

The real world entities which the event centric modelling of the CIDOC CRM aims to enable the accurate historical description of are captured through E77 Persistent Item and its subclasses. E77 Persistent Item is used to describe entities that are relatively stable in form through the passage of time, maintaining a recognizable identity because their significant properties do not change. Specific subclasses of E77 Persistent Item can illustrate this point. E22 Human Made Object is used for the description of discrete, physical objects having been produced by human action, such as an artwork or monument. An artwork or monument is persistent with regards to its physical constitution. So long as it retains its general physical form it is said to exist and to participate in the flow of historical events. E28 Conceptual Object is also used to describe persistent items but of a mental character. It is used to describe identifiable ideas that are named and form an object of historical discourse. Its identity conditions rely in having a carrier by which it can be recalled. The entities described by E77 Persistent Item are prone to change through human activity, biological, geological or environmental processes, but are regarded to continue to exist and be the same just as long as such changes do not alter their basic identity (essence) as defined in the scope note of the relevant class. Continuous matter, such as atmosphere, or things lacking sufficient stability or differentiation, such as clouds, are not instances of E77 Persistent Item. Discourse about such items may be documented with concepts of the CIDOC CRM as observations in relation to things of persistent identity, such as places.

The notion of identity is key in the application of CIDOC CRM. The properties and relations it provides are designed to allow the accurate historical description of the evolution of real world items through time. This being the case, classes and properties are created in order to provide a definition which will allow the accurate application of the classes or properties to the same real world items by diverse users. Identity in the sense of the CIDOC CRM, therefore, means that informed people are able to agree that they refer to the same, single thing, according to the scope note of the respective CIDOC CRM class it is regarded to be an instance of, regardless any other properties. For example, the Great Sphinx of Giza may have lost part of its nose, but there is no question that we are still referring to the same monument as that before the damage occurred, since it continues to represent significant characteristics and distinctness from an overall shaping in the past, which is of archaeological relevance. Things lacking sufficient stability or differentiation, such as atmosphere, soil, clouds, waves, are not instances of E77 Persistent Item, and not suited for information integration. Discourse about such items may be documented with concepts of the CIDOC CRM as observations in relation to things of persistent identity, such as places.

Learning to distinguish and then interrelate instances of E77 Persistent Item (endurants) and instances of E2 Temporal Entity (perdurants) using the appropriate properties is key to the proper understanding and application of CIDOC CRM in order to formally represent analytic historical data. In the large majority of cases, the distinction this provides and the subsequent elaboration of subclasses and properties is adequate to describe the content of database records in the cultural and scientific heritage domain. In exceptional cases, where we need to consider complex combinations of changes of spatial extent over time, the concept of spacetime (E92 Spacetime Volume) also needs to be considered. E92 Spacetime Volume describes the entities whose substance has or is an identifiable, confined geometrical extent in the material world that may vary over time, fuzzy boundaries notwithstanding. For example, the built settlement structure of the city of Athens is confined both from the point of view of time-span (from its founding until now) and from its changing geographical extent over the centuries, which may become more or less evident from current observation, historical documents and excavations. Even though E92 Spacetime Volume is an important theoretical part of the model, it can be ignored for most practical documentation and modeling tasks.

The key to the proper understanding of CIDOC CRM comes through the appropriation of its basic divisions and the logic these represent. It is important to underline that the CIDOC CRM is not intended to function as a classification system or vocabulary tool. The basic class divisions in CIDOC CRM are declared in order to be able to apply distinct properties to these classes and, in so doing, formulate precise, analytic propositions that represent historical realities The expressive power of CIDOC CRM comes not from the application of classes to classify entities but in the documenting the interrelation of individual historical items through well defined properties. These properties characteristically cover subjects such as relations of *identifying* items by names and identifiers; *participation* of persistent items in temporal entities; *location* of temporal entities and physical things in space and time; relations of *observation* and assessment; part-decomposition and *structural* properties of anything; *influence* of things and experiences on the activities of people and their products; *reference* of information objects to anything.

We explain these concepts with the help of graphical representations in the next sections.

Fig. 1 illustrates the minimal properties in the CIDOC CRM for documenting “what has happened”, the central pattern of the Model. Let us first consider the class *E1 CRM Entity*, the formal top class of the model. It primarily serves a technical purpose to aggregate the ontologically meaningful concepts of the model. It declares however two important properties of general validity and distinct features of the Model: *[P1](#_P1_is_identified) is identified by*, with range *E41 Appellation,* makes the fundamental ontological distinction between the identity of a particular and an identifier (see section “Reality and Knowledge Bases” above), and in practice allows for describing a discourse about resolving historical ambiguities of names and reconciliation of multiple identifiers. The property *P2 has type*, with range *E55 Type*, constitutes a practical interface for refining classes by terminologies, being often volatile, as detailed in the section “About Types” below.

All classes in figure 1 are direct or indirect subclasses of *E1 CRM Entity,* but for better readability, only the “subclass of” -link from *E2 Temporal Entity* is shown. The latter comprises phenomena that continuously occur over some time-span (*E52 Time-Span*) in the natural time dimension, but some of them may not be confined to specific area, such as a marriage status[[4]](#footnote-4). Further specializing, *E4 Period* comprises phenomena occurring in addition within a specific area in the physical space, which can be specified by *P7 took place at*, with range *E53 Place*. Instances of *E4 Period* can be of any size, such as the Warring States Period, the Roman Period, a siege or just the process of making a signature. Further specializing, *E5 Event* comprises phenomena involving and affecting certain instances of *E77 Persistent Item* in a way characteristic of the kind of process, which can be specified by the property *P12 occurred in the presence of*. This concept of presence is very powerful: It constrains the existence of the involved things to the respective places within the specified time and implies the potential of passive or active involvement and mutual impact. Via presence, events represent nodes in a network of **things meeting** in various combinations in the course of time at different places.

The most important specializations of *E77 Persistent Item* in this context are: *E39 Actor*, those capable of intentional actions, *E18 Physical Thing*, having an identity bound to a relative stability of material form, and *E28 Conceptual Object,* the idealized things that can be recognized but have an identity independent from the materialization on a specific carrier. The property *P12 occurred in the presence of* has 36 direct and indirect subproperties, relating these and many more subclasses of *E5 Event* and *E77 Persistent Item.* Regardless whether a CRM-compatible knowledge base is created with these properties only or with their much more expressive specializations, querying for the above presented five properties will provide answer to all “Who-When-Where-What-How” questions, and allow for retrieving potentially richly elaborated stories of people, places, times and things.

This pattern of “meeting” is complemented by two more subclasses of *E5 Event*: *E63 Beginning of Existence* and *E64 End of Existence*, which imply not only presence, but constitute the **endpoints of existence** of things and people in space and time, often in explicit presence and interaction with others, be they causal by producing or consuming or just witnessing, Note that the Model supports multiple instantiation. As a consequence particular events can be instances of combinations of these and others classes, describing tightly integrated processes of multiple nature. The representation of things connected in events by presence, beginning and end of existence is sufficient to describe the logic of *termini postquos and antequos*, a major form of reasoning about chronology in historical studies.

# 



*fig. 1 properties of basic concepts*

As a simple, real example of applying the above concepts we present a historical event, relevant for the history of art: Johann-Joachim Winkelmann (a German Scholar) has seen the so-called Laocoön Group in 1755 in the Vatican in Rome (at display in the Cortile del Belvedere). He described his impressions in 1764 in his “History of the Art of Antiquity”, (being the first to articulate the difference between Greek, Greco-Roman and Roman art, characterizing Greek art with the famous words “…noble simplicity, silent grandeur”). The sculpture, in Hellenistic "[Pergamene](https://en.wikipedia.org/wiki/Pergamon" \t "Pergamon) baroque" style, is widely assumed to be a copy, made between 27 BC and 68 AD (following a Roman commission) from a Greek (no more extent) original. Johann-Joachim Winkelmann was born 1717 as child of Martin Winkelmann and Anna-Maria Meyer and died in 1768 in Trieste.

Figure 2 presents a semantic graph of this event, as described above, using CIDOC CRM concepts. The facts in parentheses above are omitted for better clarity. Instances of classes are represented by informative labels instead of identifiers, in boxes showing the class label above the instance label. Properties are represented as arrows with the property label attached. After class labels and property labels we show in parenthesis the identifiers of the respective superclasses and superproperties from figure 1, in order to demonstrate that the story can be represented and queried with these concepts only. It also shows how concept specialization increases expressiveness without losing genericity. It is noteworthy that the transfer of information from the Greek original, to the copy, to the mind of Winkelmann and into his writings can be understood solely by this chain of things *being present* in different meetings. Note also that the degree to which a fact is believed to be real does not affect the choice of CIDOC CRM concepts for description of the fact, nor the reality concept underlying the Model.

Figure 2 represents in addition one more top-level property of the CIDOC CRM: *P67 refers to*, which describe an evidence-based fact that an information object makes reference to an identifiable item.

*fig. 2 CRM encoding example: Winkelmann seeing Laokoon*





*fig. 3 reasoning about spatial information*

The diagram above shows a partial view of the CIDOC CRM, representing reasoning about spatial information. Five of the main hierarchy branches are included in this view: E39 Actor, ~~E51 Contact Point~~, E41 Appellation, E53 Place and E70 Thing. All classes are shown as blue-white rectangles. Properties are shown as single arrows. In some cases the order of priority for property names has been reversed in order to facilitate reading the diagram from left to right. Double arrows indicate IsA relations between classes and their subclasses or between properties and their subproperties. 'Shortcuts' are indicated with light grey rectangles and their names are written in italics, such as the P55 has current location (currently holds) between E53 Place and E19 Physical Object, which is a shortcut of the path through E9 Move.

An instance of E53 Place *is identified by* an instance of E41 Appellation which includes things such as addresses and place names. It may *consist of* or *form part of* another instance of E53 Place, thereby allowing a hierarchy of geometric ‘containers’ to be constructed.

Instances of E18 Physical Thing are found on locations as a consequence of being created there or being moved there. Therefore the properties *P53 has former or current location (is former or current location of) (*and *P55 has current location (currently holds)* are regarded as shortcuts of the fully articulated paths through the respective events. *P55 has current location (currently holds)* is a subproperty of *P53* *has former or current location (is former or current location of)*. The latter is a container for location information in the absence of knowledge about time of validity and related events.

The model foresees that any instance of E19 Physical Object can serve as a frame of reference for an instance of E53 Place to be defined. The property *P1 is identified by (identifies)* can be used for such definition. For example, we may know that Nelson fell at a particular spot on the deck of H.M.S. Victory, without knowing the exact position of the vessel in geospatial terms at the time of the fatal shooting of Nelson. Similarly, a signature or inscription can be located “in the lower right corner of” a painting, regardless of where the painting is hanging.



*fig. 4 reasoning about temporal information*

This diagram shows how the CIDOC CRM handles reasoning about temporal information. Four of the main hierarchy branches are included in this view: E2 Temporal Entity, E52 Time-Span, E77 Persistent Item and E53 Place.

The E2 Temporal Entity class is an abstract class (i.e. it has no direct instances) that serves to group together all classes with a temporal component, such as instances of E4 Period, E5 Event and E3 Condition State.

An instance of E52 Time-Span is simply a temporal interval that does not make any reference to cultural or geographical contexts (unlike instances of E4 Period, which *took place at* a particular instance of E53 Place). Instances of E52 Time-Span are sometimes identified by instances of E41 Appellation~~E49 Time Appellation~~, for example as dates.

Both E52 Time-Span and E4 Period have transitive properties. E52 Time-Span has the transitive property *P86 falls within (contains),* denoting a purely incidental inclusion; whereas E4 Period has the transitive property *P9 consists of (forms part of)* that supports the decomposition of instances of E4 Period into their constituent parts. For example, the E52 Time-Span during which a building is constructed might *falls within* the E52 Time-Span of a particular government, although there is no causal or contextual connection between the two instances of E52 Time-Span; conversely, the E4 Period of the Chinese Song Dynasty *consists of* the Northern Song Period and the Southern Song Period.

Instances of E52 Time-Span are related to their outer bounds (i.e. their indeterminacy interval) by the property *P82 at some time within,* and to their inner bounds via the property *P81 ongoing throughout.* The range of these properties is the E61 Time Primitive class, instances of which are treated by the CIDOC CRM as application or system specific date intervals that are not further analysed.



*fig. 5 reasoning with spacetime volumes*

Carlo Meghini, Martin Doerr, 2018. *A first-order logic expression of the CIDOC*

*conceptual reference model.* In: *Int. J. Metadata, Semantics and Ontologies, Vol. 13, No. 2, 2018*

1. Kutschera, Walter. “Radiocarbon dating of the Iceman Ötzi with accelerator mass spectrometry.” (2002). [↑](#footnote-ref-1)
2. [↑](#footnote-ref-2)
3. The URI (instance of E41 Appellation) of the Library of Congress for Tut-Ankh-Amun, the pharaoh. [↑](#footnote-ref-3)
4. [↑](#footnote-ref-4)