

# A comparison of the OpenGIS™ Abstract Specification with the CIDOC CRM 3.2

*Draft*

Martin Doerr  
ICS-FORTH, Heraklion, Crete  
Oct 4, 2001

## 1 Introduction

This Mapping has the purpose to identify, if the OpenGIS specifications, to the degree they are on the same conceptual level, can be taken as a compatible extension of the CIDOC CRM. This makes in particular sense for archeological data, but also for any other geo-referenced cultural information. Most of OpenGIS deals with the encoding level of geometric information. This is out of the scope of the CRM, and can be regarded as extensions of the CRM entity E41 Appellation – or better as the creation and comparison of instances of E41. On the other side, OpenGIS contains features to encode semantic information, providing a data model to encode a kind of ontology. It does rarely commit to specific real world notions.

It is indeed most relevant to identify, which are the extension points of the CIDOC CRM to OpenGIS, as we suppose that both describe the same world but with different levels of detail and different kinds of information. The CRM in particular tries to capture the notions by which historical resources refer to place and time, and the actual transition between historical notions and mathematical spatiotemporal formulations is a scientific challenge by itself. In this study we can only proceed the mapping to a level of comparison where our intuition says, that the transition seems feasible to the degree the available knowledge is sufficiently rich to yield such results, or to indicate areas where changes or extensions of definitions on either side could allow for reaching the optimal transition.

An example to make this clearer. Let's assume a historical reference like: "In the third year of his reign the King Y decided to cross the borders to punish the X for their atrocities. A large host of the enemy was defeated right at the border near mount Z. The King proceeded until their capital...". How can the integration of historical and GIS data contribute to approximate the battle field, given that borders and the mountain are only indirectly known by other references, that borders have changed frequently at these times etc.?

## 2 Mappings

The OpenGIS™ Abstract Specification is organized in a series of topics, which are unfortunately not completely consistent, and are at a different stage of development. In contrast, the CRM is globally consistent (of course it is by far smaller than OpenGIS). Here we provide mappings Topic by Topic, throughout based on the Abstract Specification version 4 and CIDOC CRM version 3.2. We follow the text of the OpenGIS specifications, referring to its section numbers on the left. We use the equal sign (=) to denote equivalence, and  $\subset$  (subclass of) to denote a broader

equivalence. On the left always OpenGIS terms, on the right CRM terms in bold. OpenGIS terms are identified, where possible, with section identifiers. All mappings are provided on our naïve understanding of the given text - explanations by experts may quite well lead to revisions.

## **Topic: 0 Abstract Specification Overview**

### **4.2 Basic Data Types**

Basic Data Types = **Value Type** (this is a CRM metaclass not further defined)

Number = **E60 Number**

4.2.1-6 Real, Float, Double, int32, uint32, Byte  
⊂ (subclass of) **E60 Number**

4.2.7 Boolean = no equivalent, ontological nature probably out of scope, as it seems to be used on the encoding level.

4.2.8 CharacterString = **E62 String**.

4.2.9 Vector = no equivalent. Encoding form of certain physical quantities. CRM ends before reaching this level, e.g. with the notion of “Spatial Coordinates”.

4.2.10 Name ⊂ (subclass of) **E41 Appellation**.

### **4.3 Unit of Measure Package**

4.3.8 Measure = **E54 Dimension**

Measure.value: Number = **Dimension.value: Number**

Measure.uom: ..... = **Dimension.unit: Measurement Unit**

4.3.3,4.3.4,4.3.6,4.3.7 ⊂ (subclass of) **Measurement Unit**

All subclasses of 4.3.8: 4.3.9-4.3.12 Length,Angle,Area,Velocity, are refinements of **E54 Dimension** encoded by the **has type** property. The CRM does not enforce consistency of **Dimension** type with unit of measurement as OpenGIS does. The CRM is in general not designed to enforce consistency constraints.

4.3.14 Scale, 4.3.2 UomScale seems to be out of scope of the CRM – encoding feature.

4.3.13 Time = **E52 Time-Span**

Here the OpenGIS notion of time seems to be badly designed as measure, which it is not, as times do not add, besides other features. This is inconsistent with Topic 2, 1.2.4, page 2. The notion of temporal extent and indeterminacy of the CRM is not there. For time-spans transition from CRM to OpenGIS would be difficult due to lack of expressive power in OpenGIS. The CRM puts temporal durations under

**Dimension**, equal to OpenGIS Measure. May be "duration" was the intended meaning of "time" in this paragraph of OpenGIS.

#### 4.4 Accuracy Package

This package has a far equivalent to the CRM: It is based on error distribution and mean values, (best fits), whereas the CRM focuses on recall and precision. Consequently, CRM proposes to deal accuracy with intervals of indeterminacy or extent. Even though only elaborated for **Time-Span** in the CRM, the concept applies equally to space, and indeterminacy intervals apply to all **Dimensions**. The CRM should better specify accuracy of space and **Dimensions**. Indeterminacy interval computation is feasible for retrieval in very large data sets.

Error distributions can be roughly mapped to intervals, never achieving 100% recall or precision. Typically 95% can be regarded as a sufficient approximation for information retrieval. On the other side, the potentially infinitely large error bounds of a Gaussian error distribution are often out of reality, i.e. there is no scientific base to claim that very large, very rare errors are systematically produced by instruments read by humans. So intervals can be more realistic for the wider bounds. (Without going into detail on that issue).

#### **Topic: 1: Feature Geometry**

This chapter deals with the geometry of objects: “the spatial characteristics of features, including dimension, position, size, shape, and orientation.” The OpenGIS does not formalize “position, size, shape” as such, but goes in detailed definitions for geometric objects like Points, Curves, Surfaces, Solids, simple and composite. So any matching depends on the individual matching of specific object types (roads, bridges, swords, diamonds, birds) with suitable geometric objects. The CRM obviously deals also with objects out of the GIS concern like birds, cloths etc. So a transition from OpenGIS to the CRM is relatively simple: CRM positions (**P55 has current location** etc.) and sizes (**P43 has dimension: Dimension**) can easily be produced from OpenGIS data, whereas a transition from CRM to OpenGIS needs a matching of certain object types with suitable OpenGIS morphological classes.

The CRM has no model for dimension, shape and orientation.

Dimension in the sense of OpenGIS, 1, 2, 3-dimensional objects, is out of the scope of the CRM. All CRM **E19 Physical Objects** are real, 3-dimensional. **E26 Physical Features** may be relatively “flat”, pseudo-2-dimensional. Points do not exist in the CRM.

“Shape” is typically a question of morphological classification in the CRM, i.e. the **P2 has type** property.

Orientation is not modelled at all in the CRM, i.e. it can be kept in a textual description in CRM instances. It is not relevant for global information integration, it is

relevant however for local geometric reasoning, and for statistics on culturally relevant orientation types (the latter can be part of a classification).

Position is modelled in the CRM as a wider approximation of the area occupied by a **E19 Physical Object** or **E4 Period**. It is modelled in OpenGIS as the position of a set of characteristic points on the surface of earth, i.e. the “corners”. So, OpenGis definitions can be seen as refinement of CRM notions of position.

“Size” is modelled in the CRM as the “**P43 has dimension: Dimension**” properties of an object, i.e. as object type specific measures, without defining specific ones however. E.g. length of a curve could be a CRM dimension, but also “life wingspan” of a bird. The OpenGIS and CRM notions are basically compatible, but some models on either side are out of scope of the other.

1.1 “The Essential Model is a description of how the world works (or should work)”... = corresponds basically to the ontological level of the CRM, even though CRM contains some epistemological features and minimal encoding features.

1.2 “A feature is an abstraction of a real world phenomenon”  
Feature =  $\cup$  (union of) **E77 Existence, E2 Temporal Entity**  
("Feature" is is a typical “underspecification”, many kinds of objects that fall under this definition are by sure not of concern for OpenGIS)

“it is a geographic feature if it is associated with a location relative to the Earth”  
Geographic Feature =  $\cup$  (union of) **E4 Period, E18 Physical Stuff**

“Topology deals with the characteristics of geometric figures which remain invariant if the space is deformed....”

The rest of the Topic 1 deals with the representation geometric transformations, which is not of ontological nature but on the encoding level – or it can be regarded as on the identifier creation level.

## **Topic: 2: Spatial Reference Systems**

The key phrase is on page 5:

“As far as the user of GIS is concerned, spatial coordinates are names for locations...”

This is precisely the abstraction the CRM makes.

1.2.4 states:

"Background of the Notion Time and Place:

Place and time in the geodata model do not correspond to software entities. They are part of the boundary between the software and the 'real world'. Place is a measurable piece of the real world. Time is a point, an interval or collection of points and intervals in what we perceive as the time continuum. Time and place can be measured and surveyed, and their coordinates in a particular spatial, temporal reference system can be derived. Because the approach in the model can unify place and time, we will use 'location' to mean both. "

So

Location =  $\cup$  (union of) **E53 Place, E52 Time-Span**

For Places:

A "SpatialTemporalReferenceSystem" corresponds to the CRM properties

**P87 is identified by (identifies)** and "CoordinateGeometry" to **E47 Spatial Coordinates**.

For Time-Span:

A "SpatialTemporalReferenceSystem" corresponds to the CRM properties :

**P81 at least covering, P82 at most within, P83 had at least duration, P84 had at most duration** and "CoordinateGeometry" to **E61 Time Primitive** or **E54 Dimension** for the durations.

The match is not perfect. In the case of time, the CRM uses indeterminacy intervals, and not statistical errors. The CRM has no notion of a "space-time box". Period in CRM are the real world phenomena themselves, not their spatiotemporal container. OpenGIS seems to imply the definition of space-time boxes, even though I could not detect any 4-D constructs. This seems to be no important issue for the CRM. It may however be useful to denote alternative opinions about space-time boxes associated with Periods.

A Time-Span in CRM is one interval, not a collection of them. The CRM currently fails to express, if an object occurs in disjoint time-intervals. This seems not to be a realistic case. Assignment of multiple time-intervals for a Temporal Entity in the CRM should be interpreted as alternative opinions (OR conditions).

The OpenGIS refers to Time "points". For the CRM, there is no such thing. Reality can not be pinned down neither to time points nor to space points. Time points make sense only for querying, e.g.: which processes or periods were active at this point in time, which items did exist? Querying is out of scope of the CRM definition.

There seems to be no specific construct in the OpenGIS to declare mobile-object-relative coordinates.

I cannot decide, if the transformation mechanisms offered in this chapter can capture reasoning of cases where the relative coordinates of objects are partially indeterminate, which are frequent cases in historical reference. No explicit reference to moving frames seems to be done in OpenGIS.

### ***Topic: 3: Locational Geometry Structures***

“Locational Geometry refers to structures that map from one locational system to another. The domain and range of such mappings can be one of the following:

1. Feature geometry
2. World coordinate systems
3. Map coordinate systems
- .....”

Mapping between point 1 and 2 is the concern of “**E46 Section Definitions**” in the CRM, i.e. definition of object positions wrt the geometry of other objects. Unfortunately, Topic 3 is not elaborated. Moving reference could be introduced here.

### ***Topic: 8: Relationship Between Features***

“Topic 5 introduces the concepts of feature and feature type, but does not seek to define particular feature types. Similarly, this (8) Topic introduces the concepts of relationship and relationship type. But it does not dictate particular relationship types”.

As such, this definition becomes a metamodel, which cannot be compared to the CIDOC CRM. Ternary relationships are introduced. some examples (page 5, “book, person, library”) are comparable to CRM events. Without any specific relationship, more comparison makes no sense. Here ontological work on specific notions seems to be useful, and may reveal enough short-comings and open issues in a so general approach, as e.g. what it means that a 2-dimensional river crosses a 2 dimensional river etc.

The CRM is concerned with basic topological relationships between places, in particular those, which preserve recall if replaced by broader ones. E.g. “at the northern border of” can be replaced by “at the border of”. Topological theory teaches that inclusion is the most basic topological primitive. Therefore the CRM provides inclusion relationships for all entities of topological nature. If the CRM develops more such relationships in the future (like above “at the border of”), it may be useful to harmonize formal semantics with OpenGIS.

## **3 Conclusions**

We could not go into any depth in this study. There are two key phrases for me: “...spatial coordinates are names for locations” and “Topic 5 introduces the concepts of feature and feature type, but does not seek to define particular feature types.

Similarly, this (8) Topic introduces the concepts of relationship and relationship type. But it does not dictate particular relationship types”.

So it seems that OpenGIS is absolutely consistent with the CRM on the intellectual level, but deals with notions mostly out of the CRM concern. There is a very small overlap in the Measures, actually with a small inconsistency of OpenGIS wrt the definition of time, a relationship depending on complex details between CRM dimensions and positions, and OpenGIS notions of size and position.

As such, OpenGIS could actually be combined with the CRM. (and other ontologies).

Ontological systems like the CRM could fill OpenGIS feature types and relationship types with a useful “life”. E.g. the general topological relationships as “before”, “after”, “within” etc. seem to appear in OpenGIS as a result of computation rather than as primary knowledge. Even though argued for, Topic 8 does not define any of them.

CRM has a strong need of “intuitive” or “naïve” geometric terms and relationships as found in cultural sources. The transitions for “feature geometry” need individual work and cannot be defined in general. E.g. it is usual to measure a painting with two measures, which relate clearly to a respective “Solid”,

OpenGIS may regard interval arithmetic as a more scalable approach to error bounds than Gaussian errors as an additional alternative.

OpenGIS does not offer any new useful notion to the CRM within the scope of the CRM. There could be an “OpenGIS awareness” of CRM instances, making additional definitions explicit to allow for example to communicate object positions and shapes in a manner exploitable by OpenGIS.

Algorithms to interpret historical qualitative relationships in GIS terms would be an important step forward to scientific reasoning in the cultural area.