# Principles for Modelling Ontologies: a short reference guide

##

## Introduction

*Narrative here*

(General introduction for both parts)

Document has two parts:

New Modelling Construct Proposal Checklist and Procedure

New Extension Proposal Checklist and Procedure

Terms to be explained here:

Modelling construct

…

Audience

Modellers who want to submit new constructs or extensions to the CRM-SIG

##

## New Modelling Construct Proposal Checklist and Procedure

*Narrative here*

*The process of developing new properties and classes for official submission to the CRM-SIG has three main steps. The first step involves establishing the need, relevance and distinctiveness of the new additions. This is established through actual practice and evidence from that practice. Step two, involves the development of appropriate properties that* encapsulate *the new process and the decision as to whether new classes are needed, and at what level of the CIDOC CRM hierarchy. The third step involves a re-evaluation of steps one and two but through wider consultation with others. It is highly likely that steps one and two must then be revisited and therefore the process should not be considered linear. This is all underpinned by the need that the proposal is not just considered real, but is adequately supported by evidence. Creating new properties and classes involves research but is practitioner led research.*

### Step ONE: Need, relevance and distinctiveness

Proposals for new modelling constructs or modifying existing ones should start with a “how to model” question. The only exceptions to this rule are relevant research questions, that are already represented by multiple different modelling constructs, but do justify a more abstract property, such as “P12 was present at” or “P92 brought into existence”, which in turn justify a very abstract class such as “E77 Persistent Item”.

| **what you need** | **what this is** | **what it should take into consideration [criteria for approval or rejection]** |
| --- | --- | --- |
| A ”how to model” question | This question should refer to a phenomenon in the domain of interest of the user.Example: “how to model the modification of an art object by an artful contribution, possibly by a different artist” | Adding the envisaged model to the CRM or a local or approved extension should allow for documenting instances of the phenomenon in a digital form that will allow for contextual integration and answering specific research questions on the (integrated or not) data. |
| examples | real-world, individual examples of the referred phenomenon, from a database or other documented source | examples must be real, not fictional and should be documented somewhere. They must be representative for a larger amount of such documentation.  |
| Research questions | These are questions which can either be answered directly by querying the envisaged knowledge base, or which can be answered by synopsis of a reasonable (recall-precision ratio) subset retrieved by queries. Example: “all art objects created by person X / at time Y / at place Z/ on order by U/ using style W?” | The questions can be simple and straight-foward, They will be used, based on the examples, to decide if the current CRM constructs are already adequate or sufficient (in precision) to answer such questions. In particular if the required information will be a detail inside one document that (the whole document) can easily be found with simpler means, such as keyword search or other obvious metadata and does not need a new model for information integration. Example: the likelihood that different parameters of the same intentional action of the same dog will appear in different sources without characteristic keywords about types of actions and species is low, as well as data about the same dog in different sources without its name. |
| Relevance and feasibility | The expected answers about the phenomenon must have a sufficient frequency and/or cultural-historical bearing to justify a model addition to rank among the “top thousand” constructs across the subdomains of the wider CRM scope. | The default way to prove relevance is that sufficient people are spending considerable time and/or money to already encode descriptions of the phenomenon or parameters of its context in data structures or in systematic comments within data structures. If the phenomenon has not yet been formally encoded in any form, but is found in free texts, feasibility of larger scale extraction should be assessed. The research questions are valid for a discipline that is within scope of the CRM. |

#### Decision if the modelling should continue:

If the existing models are deemed sufficient or the question not sufficiently relevant, A) either

the issue is closed, or B) more evidence will be required or C) the discussion may have shed new light on the initial question.

In case B) and C) we may reformulate the problem. Case C) may also induce a new issue.

If the existing models are not deemed sufficient and the question is sufficiently relevant, we enter step TWO. The initial question may not be new. New research questions may have arised, new relevance or feasibility of data extraction encountered,

If in this process existing constructs may be found to be inconsistent, in particular between extensions and CRMbase without a new construct to be proposed, we may go to step THREE directly.

##

### Step TWO: Modelling constructs

Development of a modelling construct for addition and possibly modification of CRMbase or compatible extension. A class can never come alone. It needs to be the domain or range of at least one new or redefined property. In the same way as shown below, modifications of existing constructs may be proposed, such as only a scope note change of an existing class.

| **what you need** | **what this is** | **what it should take into consideration [criteria for approval or rejection]** |
| --- | --- | --- |
| One or more initial properties | Properties that can explain, at least in a naive way, the examples and respective research questions. Labels and scope notes may be still tentative, rendering the idea mainly. |  If there are more than one property proposed, they should share some domain or range classes and thereby constitute a propositional context. Graphics are very helpful for that. A minimal proposal may only widen the domain or range of an existing property.  |
| Initial proposals for domain and range classes  | The initial properties need some proposed domain and range classes. Labels and scope notes may be still tentative, rendering the idea mainly. | These classes will often be existing ones. However, in the initial design, it may be often better to propose a new class than overstretching an existing one.  |
| A proposal for integration in CRMbase or extension constructs | Superproperty or superclass proposals for new classes and properties.  | In the initial modelling construct, a superclass for at least one of the domains or ranges can be sufficient. |
| Adequacy assessment | List here the questions that instantiating this class with its related properties or property either by itself or together with other properties would make possible.Refer to the relevance statement and the relevant examples. |  |

### Step THREE: Integration into CRMbase and all finished approved extensions

The result of this final step should be a decidable, consistent proposal. This is obtained by intensive collaborative discussion. Properties and classes must be distinguishable in an unambiguous way from all other classes and properties. Normally, it requires several revisions of step TWO, and often of step ONE. The discussion normally converges. If not, the issue may be abandoned until further evidence is provided.

| **what you need** | **what this is** | **what it should take into consideration [criteria for approval or rejection]** |
| --- | --- | --- |
| A label | a heuristically helpful human readable name that gives a useful general indication of what the class or property might be used for | The existing conventions for declaring labels of properties or classes in CRM |
| A scope note | a description of the intension of the class or property, that is to say, a human readable definition of the kinds of real world things or relationships that would be instances of this class or property | The scope note should serve to guide a reasonable user in the domain to an understanding of what would count as an instance of the class or property. The CRM SIG has guidelines for writing scope notes, specifying how one should indicate the identity, substance, unity and existence conditions of a class or property. These can be consulted to help organize a scope note according to the CRM best practices. |
| examples | real world, individual examples of the class or property proposed, from a database or other documented source | examples must be real, not fictional and should be documented somewhere |
| subclasses or subproperties | Note if the class or property proposed is considered to be a subclass or subproperty of an existing CIDOC CRM class or property | logical consistency with existing classes and properties |
| superclasses or superproperties | Note if the class or property proposed is considered to be a superclass or superproperty of an existing CIDOC CRM class or property | logical consistency with existing classes and properties |
| domain | For properties, it is important to define the scope of their applicability within the ontology, therefore a domain class must be given for a proposed property | logical consistency with existing classes and properties |
| range | For properties, it is important to define the scope of their applicability within the ontology, therefore a range class must be given for a proposed property | logical consistency with existing classes and properties |
| questions this class or property helps answer | list here questions that instantiating this class or property either by itself or together with other properties would make possible | is the task of answering questions of this sort aided by the introduction of the above classes or properties? Are existing classes and properties sufficient to do so without this addition?  |

## New Extension Proposal Checklist and Procedure

*Narrative here*

| **what you need** | **what this is** | **what it should take into consideration [criteria for approval or rejection]** |
| --- | --- | --- |
| Theoretical scope | A textual outline of the overall intended scope of phenomena meant to be covered by the extension | That it has a discernible function and does not overlap with other extensions should be lucidly explained |
| Practical scope | A textual outline of the actual set of phenomena having been considered and modelled based on empirical evidence | Takes into account relevant real world kinds of data and modes of research |
| Research questions | A list of the kinds of research questions that the extension aims to be able to answer | Validated / supported by scholars or scientists in the field |
| Instance check | A list of real-world instances independent of existing modelling constructs (or sources) which are covered by the extension (as proof of its functionality?) | ... not sure. Caution should be applied to taking examples from existing sources since the abstractions created in systems may well have not used the same process as the one documented here and in traditional models where artificial abstraction are used.  |
| Defined classes | see new class or property | see new class or property |
| Defined properties | see new class or property | see new class or property |
| Bibliographic references | appropriate reference literature from the theoretical scope indicating major theories, and approaches of necessary consideration | Academic bibliography about relevant theories related to this model, possibly compatible, similar or explicating logically some constructs of the model and related standards. Note that a CRMextension should be neutral to specific subject matter theories except for disciplinary basics as are taught in beginners courses, and this only to the degree they are cross-disciplinarily acceptable. This is because data collected in information integration must serve verification and falsification of theories |