Recording uncertainty

# Scope

## Inaccuracy and imprecision

This well-known diagram describes accuracy and precision of measurements.



This is not the kind of uncertainty we are referring to here.

## Types of uncertainty

In many cases an observer is not certain about an observation and they wish to give a degree of confidence alongside their observation. This is a self-judgment about the quality of their observation, i.e. an assessment of the assessment by the same Actor.

Nicola has [looked into this in detail](https://github.com/swiss-art-research-net/vocab) and he has identified the following types of uncertain data (examples and commentary are mine, apologies Nicola if I have misinterpreted things):

* Imprecision: for example, the height of this book is 10cm ±2cm. This is the same as the diagram above and therefore out of scope for this exercise. E54 Dimension is being revised in relation to this (see recent [thread](http://lists.ics.forth.gr/pipermail/crm-sig/2019-October/004010.html) in the SIG list). Imprecision are interval which are not precisely recorded but can be verified
* Ambiguity: for example the statement “This is a bound book” is too generic to tell us anything about the type of binding which is my research interest. I.e. the level of description does not match the level of query.
* Incompleteness: for example, recording the metal furniture of a book implies that the book has boards which however may not be recorded. Incompleteness should be considered in relation to the CRM chains. Otherwise incompleteness is subjective.
* Uncertainty: for example, I know that the board of a book is attached to it in some way, but because the attachment components are hidden, I do not know what they are.

I think that it is only the last type that is in scope here. Nicola may have arguments for addressing the above in combination.

# Proposals about structure

## Solution 1:

There is a proposal by Martin (<http://www.cidoc-crm.org/Issue/ID-349-belief-values>) for a .2 property to describe confidence/uncertainty.

### Pros

* This can be relatively easily introduced to the CRM following the .1 model.

### Cons

* Any implementation will have the same limitations of the .1 properties, e.g. for RDF: reification.
* Another concern is interpreting the statements missing the .2 property: is the confidence level “certain” or “unknown”?

## Solution 2:

Robert argues for Attribute Assignment and explains as follows:

“We use AttributeAssignment to be able to record:

* Who was uncertain and when about the data
* To have a real activity, to which other assertions can be made (e.g. to have statements about the uncertainty, the process that led to the Attribute Assignment, etc etc)
* The information separate from the main graph, without introducing either new classes (PC14 etc), named graphs (which we use for other purposes), or non-RDF-possible syntax (.1 .2 etc). It can be queried via a regular graph search.
* The certainty could be quantified as part of the Attribute Assignment
* The new assigned\_attribute\_type property (P170?) gives the relationship, so we're now able to use P2 for more classification oriented tasks … e.g. a certainty typology.
* We could then use the (not CRM due to scope discussion) ordered types to say that unlikely is less than uncertain is less than probable is less than almost certain to be able to compare uncertainties in order, without a false precision of "0.6" certainty - the order is important, not the interval.“

### Pros

* Rob made the case above

### Cons

* The main disadvantage appears to be confusion around agency. I.e. normal statements in the knowledge base are implicitly attributed to the keeper of the KB while the statements with extra AttributeAssignment statements for confidence are explicitly attributed to the keeper.
* Another disadvantage appears to be the fact that you cannot directly assign confidence to a statement (a triple) in the KB, but only assign confidence to an attribute assignment event, which is slightly different, i.e. querying specific KB statements will not return confidence level, the query will have to include an Attribute Assignment event.

## Solution 3:

Franco’s paper (<https://doi.org/10.1007/s00799-016-0195-1>) makes use of a subclass of *E14 Measurement* to express the level of uncertainty/reliability (Franco’s image from [here](http://www.cidoc-crm.org/Resources/expressing-reliability-with-the-crm-are-you-sure)):

I think this is similar to Solution 2 in that it makes use of Attribute Assignment and then it extends Measurement and Dimension to specify confidence level. See also “Proposals about confidence values” below.

## Solution 4:

In CRMinf, property *J4 that* and *J5 holds to be* can be used alongside *I4 Proposition Set* to assign a belief value to an observation.

### Pros

* Existing and robust model.

### Cons

* It appears that agency for the explicit versus implicit statements will still be a problem.
* The implementation of this solution requires named graphs.

# Proposals about confidence values

Numerical (Franco) versus hierarchical/ordinal. Under development.