Typed properties and negative typed properties: dealing with type observations and negative statements in the CIDOC CRM

Editor(s): Eero Hyvönen, University of Helsinki, Finland
Solicited review(s): Öyvind Eide, University of Cologne, Germany
Open review(s): Robert Sanderson, Yale University, USA

Athanasios Velios[a,*], Martin Doerr[b], Carlo Meghini[c] and Stephen Stead[d]
[a]*Ligatus, University of the Arts London, Chelsea College of Arts, 16 John Islip Street, London, SW1P 4JU, United Kingdom*
[b]*Information Systems Laboratory, Institute of Computer Science, Foundation of Research and Technology Hellas, Greece*
[c]*Networked Multimedia Information Systems Laboratory, Institute for Information Science and Technologies "Alessandro Faedo" – ISTI, Italian National Research Council (CNR), Italy*
[d]*Paveprime Ltd., United Kingdom*

Abstract. During condition and collection surveys in memory organisations, surveyors observe the absence of features on collection items. They also observe types of multiple components as single instances given that their large number makes them difficult to be captured as separate instances. Such observations are significant to researchers, documented in registration forms but are not easy to model in popular ontologies such as the CIDOC CRM. In this paper the nature of absence is explored from an ontology point of view alongside the role of the Open World and Close World Assumptions in knowledge bases. A proposal is then formulated for the use of special properties within the CIDOC CRM ontology, namely 'typed properties' and 'negative typed properties' which allow modelling the typology of multiple instances and the absence of instances. The nature of these properties is then explored in relation to their correspondence to longer property paths, their hierarchical arrangement and relevance to thesauri. First order logic statements are used to describe these properties. Examples from bookbinding structures are used given the significance of such observations in the field of bookbinding history. The paper concludes with reference to ongoing implementation work and a summary of findings.

Keywords: Open World assumption, Closed World assumption, existence, CIDOC CRM, bookbinding

1. Introduction

It can be reasonably claimed that the Conceptual Reference Model (CIDOC CRM) [6] has become the de facto standard for data integration in cultural heritage. The model offers classes and properties which allow expressing data produced in memory organisations such as museums, libraries, galleries and archives, including scientific record keeping. Such data are typically the result of observation of collection items or field observations, combined with knowledge about past events. The CIDOC CRM puts emphasis on events and activities and on the relations of events and activities [3] with:

- the time and place that these happened,
- the people that undertook them and participated in them,
- the things and ideas which were used or produced during them.

*Corresponding author. E-mail: editorial@iospress.nl. Check if the checkbox in menu *Tools/Options/Compatibility/Lay out footnotes like Word 6.x/95/97* is selected if you make a footnote for the corresponding author.

Datasets integrated with the CIDOC CRM can be jointly queried in a knowledge base based on such property relations. This makes sharing and relating observations or beliefs about the past possible. Therefore more accurate contributions to scholarship can be made because larger sets of data can be jointly examined and conflicting opinions from different researchers may even automatically be identified. For processing knowledge, such conflicting opinions may need either to be resolved or to be highlighted as co-existing in the knowledge base. Resolving conflicts of opinions is the core of scientific knowledge and the main motivation behind this paper.

The maturity and robustness of the CIDOC CRM is demonstrated, among others, by the apparatus that describes the materiality of objects; this apparatus has been critically reviewed for many years. For example, a frequently used property of the CIDOC CRM is 'P46 is composed of (forms part of)' which can be used to express the link between a physical thing (e.g. a collection item) and its component parts. This property is further extended by the more specific property 'P56 bears feature' which expresses the link between a physical thing and a feature found on it, such as an embossed decoration or a coloured decoration. These properties are used regularly in descriptions of the materiality of things and in this paper a representative case of describing bookbindings is highlighted.

The field of bookbinding history relies on the observation and recording of books as physical objects. Typical records of bookbindings outline material composition, techniques used during production or modification, and components contributing to the book structure. The benefits of integrating datasets of bookbinding descriptions using the CIDOC CRM have been described in detail elsewhere [18] and they may be summarised as: a) large sample: combining observations which any one researcher cannot make on their own results in a larger sample which can lead to more representative conclusions regarding the provenance of books and the development of the bookbinding craft, and b) cross-referencing: the capacity to combine provenance information from bibliographic records with provenance information from the materials and techniques found on particular books allows researchers identifying conflicts of opinion regarding provenance. These have significant impact on decision making for professionals in memory organisations, primarily conservators treating and repairing books but also curators and scholars consulting them, among others for understanding the history of the dissemination of the information content and its owners.

Books are complex objects which include several hundred components and many observations can be recorded for each one of them. However, often the resources available for surveying books do not allow the recording of all components in a knowledge base. A survey of manuscripts in the Saint Catherine Monastery in Sinai [13], Egypt, which was generally accepted as very detailed, resulted in more than a thousand observations on each book, but it was still not exhaustive. For example, page 7 of the survey only recorded basic information about endbands which are in themselves rather complex structures [1] and therefore more detailed data could have been captured about them to describe things such as production techniques. Limited resources require that a data registration form is used for ensuring that only observations which answer important research questions are recorded. Another example is documenting page markers (or leaf markers), small pieces of material attached to a leaf for marking an important part of the text. These are recorded on page 1 of the St. Catherine's data registration form [12]. A book may have many leaf markers but recording each one of them is not logistically possible. In the St. Catherine's form, only the materials and types of leaf markers are recorded without reference to individual leaf markers. In this article the concept of 'typed properties' is introduced to describe this process.

Another important set of questions expressed in data registration forms is that of existence. For example, when describing the decoration of a book cover, it is useful to know whether it features gold-tooled decoration as this may help the researcher estimate the cost of commissioning the binding and therefore the socio-economic profile of the owner. If a book cover does not feature gold-tooled decoration, then there is no component part of the binding to be described (i.e. there is no gold-tooled decoration). Intuitively, creating a record for something that does not exist in the context of a survey of physical things appears wrong. In this article the concept of 'negative typed properties' is introduced to explore and address this problem, which is also discussed in section 3.2 'Epistemology of TPs and NTPs and Open World assumption'.

1.1. Numerous components and non-existing components

An important distinction in documentation using formal ontologies like the CIDOC CRM is between ‘universals’ and ‘particulars’. A universal (type, class, category) corresponds to an idea which describes the nature of something. A particular (individual) corresponds to something identifiable in reality that fulfills or exemplifies the description of the universal by exhibiting the necessary and more or less of the typical traits described by that universal. Documenting this correspondence is called ‘instantiation’ and the recorded particular is called ‘instance’ of the universal. In this paper ‘class’ and ‘instance of class’ are used to denote universals and particulars, respectively. In the example of the previous section, the book component type of ‘leaf markers’ [24] is a class, whereas the record of an individual leaf marker on a specific book refers to an instance of that class. Likewise, ‘gold-tooled (decoration)’ [23] is a class whereas the specific gold-tooled decoration feature on a specific book is an instance of that class. The problem explored here is how to produce records about instances when these are too numerous to be observed (e.g. for a book with too many leaf markers) and how to produce records when instances do not exist (e.g. for a book without gold-tooled decoration).

1.2. Previous work

The problem of documenting non-existent instances has been mentioned before by Velios et. al. [20] but without offering in-depth analysis or potential solutions. Within the discourse of the Semantic Web the work by Razniewski and Nutt [14] and Darari et. al. [2] have explored the idea of completeness in knowledge bases which is relevant to documenting non-existence and in particular establishing a context based on which completeness can be checked. The problem of documenting types of multiple instances has been considered within the CIDOC CRM specification and it has also been discussed by Lin et. al. [8] and as part of the MetaCRM [22]. Within the discourse around the CIDOC CRM there has been no previous systematic attempt to address the problem of non-existent instances.

1.3. Logic and computation

In this paper the aforementioned problems are considered in logical terms and an attempt is made to describe statements about material observations following work that has been done for expressing the CIDOC CRM model in first order logic [10]. However, the computational aspects of our proposed solutions are not considered, but only their logical adequacy in addressing the problems at hand.

Two notations are used for expressing that individual a, instance of class A, and individual b, instance of class B, are linked by property P: a) the classical logical notation, i.e. P(a,b), and b) what is often used in the CRM community, i.e. a (A) → P → b (B). The same applies if variables x and y are used in place of the individuals.

1.4. Structure of this paper

Following the introduction, section 2 provides the theoretical foundation of considering non-existent individuals and the analysis on how previous work can inform our proposal. Section 3 explains the nature and use of typed properties and negative typed properties through articulation of property scope, first order logic statements and examples from bookbinding structures. Section 4 refers to practical implementation work underway and Section 5 summarises the findings and conclusions.

2. Non-existence and categorical statements

The philosophical discourse around non-existence is often introduced with a seemingly strange paradox. The Stanford Encyclopaedia of Philosophy [15] begins the relevant article on non-existence of objects with: “to be able to say truly of an object that it doesn’t exist, it seems that one has to presuppose that it exists, for doesn’t a thing have to exist if we are to make a true claim about it?”. The ‘thing’ in this case corresponds to an instance of the CIDOC CRM class ‘E77 Persistent Item’. A detailed review of philosophical discourse on non-existence is beyond the scope of this article, but this paradox is worth considering.

Hume considers non-existence of objects (instances of ‘E77 Persistent Item’) impossible, since making statements about something requires that it exists. In other words the characteristic property of existence is a requirement for talking about things.

Kant questions this requirement and disputes existence as an ontological criterion. This dispute leads to examining whether statements of non-existence are true or false. Stanford's article brings the example of the statement "Pegasus does not exist" and explains how by accepting this statement as true, the contradiction mentioned above is observed.

This contradiction is used as a starting point to explore fictitious and real things as well as the nature of the statements which lead to this contradiction.

2.1. Basic concepts

From the perspective of this paper which is describing the result of material observations, it appears that negating existence does not require statements about an individual not existing. For that purpose, let us distinguish three concepts: a) existence of an individual as an idea, b) a potentially real individual and c) the absence of any real individual.

2.1.1. An individual as an idea

The existence of an individual as an idea is an object of discourse and has virtual properties. Examples are mythological beings such as Pegasus and also literary characters. They are described in the CIDOC CRM as instances of 'E89 Propositional Object'. They may be a universal (dragon) or individual character (Asterix). Their only reasonable questioning of existence can be about their comparability and common identity across sources. For literary characters specifically, in 2017, IFLA's publication of the Library Reference Model changed the classification from 'Person' in FRAD [5] to 'Literary Character' as described in recent work [21]. This form of existence is not examined in this paper and it certainly does not apply to discourses similar to the development of historic bookbinding. For example fictional components of bindings structures are not considered as part of bookbinding history. Individuals as ideas (instances of 'E89 Propositional Object') are clearly different to instances of 'E18 Physical Thing'. The Pegasus paradox has no substance if Pegasus is considered as an instance of 'E89 Propositional Object' (and not 'E18 Physical Thing'). Often, the existence of an individual as an idea is confused with the existence of a potentially real individual which may be a prototype of the idea as explained next.

2.1.2. A potentially real individual

A potentially real individual may be the result of interpreting references and other sources of evidence or the result of observation. Examples include individual persons within the discourse of pseudonyms, non-identified suspects or murder weapons and prehistoric rulers, such as the prototype for King Arthur. In these cases, our knowledge only constitutes a finite set of constraints $A_1,...,A_n$ on properties relating the potential real individual to particulars and universals. Additional knowledge beyond these sets of constraints means that the individual is known to exist. As an example, let us consider the existence of sewing supports from the bookbinding discourse. Sewing supports [26] are used at the spines of books to help sewing gatherings and to provide additional strength. Sewing supports are often hidden under the book cover and when sewing is tight they are difficult to observe. Depending on the quality of the spine, one may be able to feel a raised area on the cover indicating the existence of a sewing support underneath. Correlating the location of that area with the location of sewing holes inside the gatherings makes this indication stronger. The set of constraints for the potential sewing support would be its location in relation to the sewing holes of the specific book and its length and possibly thickness. The existence statement matches a real individual with the given constraints, in our example "there is a sewing support of a specific length and thickness at the specific sewing hole location". The non-existence statement is: there exists no individual, instance of class X, fulfilling $A_1,..,A_n$. In our example: "there is no individual, instance of 'sewing support', fulfilling the constraints of length and thickness at the specific location". The latter does not make a statement about the counterfactual individual (i.e. the sewing support), but about the portion of reality covered by the constraints. Therefore it does not constitute the classical paradox of the Pegasus. Note however, that the counterfactual instance may need to be instantiated in its class in order to test whether it conforms with the applicable constraints, but this instantiation is a technical fact that occurs during the execution of an algorithm on the knowledge base and does not imply existence in the underlying reality.

2.1.3. Absence of any real individual

More generally, and abstracting from potentially real individuals, absence of any real individual is described by some constraints in a certain time and space. For instance the question from the field of biodiversity whether a certain species, like the legendary Dodo, is extinct. Often, current knowledge does not allow for deciding about non-existence. We maintain

that any such negation statement based on observation requires a complete observation of a region of reality, which is implicitly determined by the given constraints. Whereas the assumption is that the habitat of the Dodo has sufficiently been surveyed not to hide an animal that size, there are many cases of previously thought extinct species that have reappeared. For example the recent reappearance of Sea Otters in Monterey, believed to be extinct, proved the contrary [16]. This shows that absence of any individuals depends on the level of completeness of observation. If the cover of a book is torn at the spine and the sewing is loose, the existence or not of a sewing support can be observed, thus creating an environment where observations for sewing supports are complete. In the following sections of this paper, a description is made for alternative modelling approaches to the observation of non-existence of features in bookbindings, which pertain to things existing in limited areas of observation. These are manageable in a knowledge base generally considered having incomplete observations. This idea of completeness is again discussed alongside the Open World assumption later in this paper.

2.2. Non-existence in practice and compatibility of comparisons

Having established the three basic concepts around non-existence, in this section these concepts are used to propose a way of describing non-existence within the CIDOC CRM with the example about 'gold-tooled (decoration)'.

Based on the analysis of the three concepts above, one falls into contradiction when claiming that a physical thing does not exist since it has a physical manifestation. But there is no contradiction to say that a thing exists as an idea or a hypothesis. Therefore describing non-existence focuses on the discussion about real and hypothetical things. In practical terms for the problem of creating records of non-existing objects, this means that records about physical things should not be created but instead records about concepts - fictitious things describing hypothetical things can. In CIDOC CRM terms, instances of 'E18 Physical Thing' are replaced with instances of 'E28 Conceptual Object' or the more specific 'E89 Propositional Object'.

As mentioned in the introduction, the intention is to process knowledge by highlighting contradictions in statements. For example one opinion being that a gold-tooled decoration feature exists on the book and another opinion being that a gold-tooled decoration feature does not exist on the same book. Identifying the contradiction requires checking comparable entities. However, the first observation results in a statement about an instance of 'E18 Physical Thing' and the second about an instance of 'E89 Propositional Object'. A machine could not easily identify that they are contradictory since they refer to different types of entities. To establish a contradiction in this case would require human interpretation and would be difficult to automate. For example the following statements would not be identified as contradictory by a machine, i.e. without human interpretation:

book (E18 Physical Thing) → P46 is composed of (forms part of) → gold-tooled decoration feature (E18 Physical Thing)

"book does not have gold-tooled decoration" (E89 Propositional Object) → P129 is about → book (E18 Physical Thing)

Therefore, it seems that establishing instances of 'E28 Conceptual Object' or 'E89 Propositional Object' to mark the non-existence of physical things gets around the problem of creating records of non-existing physical things, but does not assist in the automatic detection of contradictory statements.

Previous work has highlighted the importance of categorical and cross-categorical knowledge [4] which is featured in the CIDOC CRM. The CIDOC CRM defines the class 'E55 Type' to mean any category. Instances of class 'E55 Type' are pseudo-particulars in that they instantiate 'E55 Type', so they are individuals, but also denote categories which could otherwise be considered extensions of CIDOC CRM classes. The property 'P2 has type' which connects an individual with its category, replaces that extension hierarchy, thus keeping the CIDOC CRM a compact model. This mechanism is used in the proposed solution.

Lin et al. [8] discuss issues around categorical knowledge using an example from the field of biodiversity: "The Kobra eats rodents and lives in India". This statement is expressed as if the category of 'Kobra snakes' is an instance of a snake (instance of 'E18 Physical Thing') although in reality it is an instance of 'E55 Type'. The example goes further mixing categories and individuals: "a specific snake of the type Kobra eats rodents". This is in parallel to our example that a specific book carries leaf markers or that a specific book carries gold-tooled decoration. In order to accommodate such statements a proposal for the MetaCRM [22] was established where all domains and ranges of CIDOC CRM properties were

replaced by ‘E55 Type’. In our bookbinding example this allows statements like:

the type book (T18 Type of Physical Thing) → CP46 is usually composed of → the type leaf markers (T18 Type of Physical Thing)

the type book (T18 Type of Physical Thing) → CP46 is usually composed of → the type gold-tooled decoration (T18 Type of Physical Thing)

and similarly for examples mixing individuals and categories:

book (E18 Physical Thing) → CP46 is usually composed of → the type leaf markers (T18 Type of Physical Thing)

book (E18 Physical Thing) → CP46 is usually composed of → the type gold-tooled decoration (T18 Type of Physical Thing)

These statements do not make sense strictly for the bookbinding discourse and they do not resolve the problem of forming statements about the existence or non-existence of things. They do, however, highlight the switch from statements about instances of things to statements about types of things and they show a path for making statements about numerous objects in the form of categorical properties: ‘typed properties’. If the range of all statements ended with a category, such as ‘E55 Type’, then comparisons for contradictions on these categorical statements would be possible. Therefore the solution may be located in the switch to categorical statements using ‘E55 Type’.

3. Typed properties and negative typed properties

Following the realisation that it is more practical to describe a) non-existent things and b) numerous similar things using categorical statements, a proposal for new properties to cover such cases is presented.

The proposal requires two steps, each producing a set of properties deriving from CIDOC CRM properties. Step 1 involves changing the range of each property to ‘E55 Type’. This ensures that the property links to a category instead of an instance and therefore allows us to describe things without referring to instances. For example for the property ‘P46 is composed of (forms part of)’:

CIDOC CRM: E18 Physical Thing → P46 is composed of (forms part of) → E18 Physical Thing

Step 1: E18 Physical Thing → TP46 is composed of physical thing of type (is type of physical thing which forms part of) → E55 Type

This allows expressing that a thing has many components of a type without having to describe the individual instances of these components and solves the problem of describing numerous individuals. Property TP46 is the typed property (TP) deriving from P46.

Step 2 involves using a negative label for the typed property to indicate negation.

CIDOC CRM: E18 Physical Thing → P46 is composed of (forms part of) → E18 Physical Thing

Step 1: E18 Physical Thing → TP46 is composed of physical thing of type (is type of physical thing which forms part of) → E55 Type

Step 2: E18 Physical Thing → NTP46 is not composed of physical thing of type (is not type of physical thing which forms part of) → E55 Type

This allows expressing that a thing is not composed by components of a specific type. Property NTP46 is the negative typed property (NTP) deriving from P46.

The significance of statements made with these properties is analysed in the next section.

3.1. First order logic of TPs and NTPs

3.1.1. Typed properties

The CIDOC CRM already includes a property which connects instances of ‘E7 Activity’ with types (instances of ‘E55 Type’). This is ‘P125 used object of type (was type of object used in)’ whose scope note reads: “This property associates an instance of E7 Activity to an instance of E55 Type, which defines the type of object used in an instance of E7 Activity, when the specific instance is either unknown or not of interest, such as use of ‘a hammer’.” This looks like a typical typed property as described above. The scope note is accompanied by the description of the class using first order logic (variables are always implicitly universally quantified):

$P125(x,y) \Rightarrow E7(x)$ (domain axiom)

$P125(x,y) \Rightarrow E55(y)$ (range axiom)

$P125(x,y) \Leftrightarrow (\exists z)\ [E70(z) \wedge P16(x,z) \wedge P2(z,y)]$

In this description, (x) is a variable standing for an instance of ‘E7 Activity’ and (y) is a variable standing for an instance of ‘E55 Type’. The third sentence above indicates that whenever this property is asserted of any two individuals x and y, it can be inferred that there is an instance (z) of the class ‘E70 Thing’ which is connected to (x) using ‘P16 used specific object’ and that this (z) can be categorised with ‘P2 has type’ as (y). In other words the use of P125 allows inferring the existence of at least one instance of ‘E70 Thing’. With appropriate instructions,

a machine can automatically return results of such instances whenever this property is used.

Following the same principle most CIDOC CRM properties can be the basis for typed properties. In our example of 'P46 is composed of (forms part of)' becomes 'TP46 is composed of physical thing of type (is type of physical thing which forms part of)' and the first order logic description would be:

$TP46(x,y) \Rightarrow E18(x)$ (domain axiom)

$TP46(x,y) \Rightarrow E55(y)$ (range axiom)

$TP46(x,y) \Leftrightarrow (\exists z)\ [E18(z) \wedge P46(x,z) \wedge P2(z,y)]$

In this description, (x) is a variable standing for an instance of 'E18 Physical Thing' and (y) is a variable standing for an instance of 'E55 Type'. The third sentence indicates that whenever this property is asserted of any two individuals x and y, it can be inferred that there is an instance (z) of the class 'E18 Physical Thing' which is connected to (x) using 'P46 is composed of (forms part of)' and that this (z) can be categorised with 'P2 has type' as (y).

The typed properties can also be considered as shortcuts within the CIDOC CRM. Again, on the same example, the statement:

a → TP46 is composed of physical thing of type (is type of physical thing which forms part of) → b

is a shortcut for the more complete set of statements:

a → P46 is composed of (forms part of) → z

z → P2 has type (is type of) → b

Finally it follows that if this set of statements exists:

a → P46 is composed of (forms part of) → z

z → P2 has type (is type of) → b

then this statement is also true:

a → TP46 is composed of physical thing of type (is type of physical thing which forms part of) → b

This last statement is needed for identifying contradictions as explained below.

3.1.2. Negative typed properties

The same principles are followed to describe negative typed properties. In the example of 'P46 is composed of (forms part of)' the negative typed property would be 'NTP46 is not composed of physical thing of type (is not type of physical thing which forms part of)' with range 'E55 Type'. The first order logic description would be:

$NTP46(x,y) \Rightarrow E18(x)$ (domain axiom)

$NTP46(x,y) \Rightarrow E55(y)$ (range axiom)

In this description, (x) is a variable standing for an instance of 'E18 Physical Thing' and (y) is a variable standing for an instance of 'E55 Type'.

Note that a statement like: $NTP46(x,y) \Leftrightarrow (\neg\exists z)\ [E18(z) \wedge P46(x,z) \wedge P2(z,y)]$ would indicate that whenever this property is used for (x) it follows that there is no instance (z) of the class 'E18 Physical Thing' which is connected to (x) using 'P46 is composed of (forms part of)' and that can be categorised with 'P2 has type' as (y). However this implies the use of (z) which is the very instance that should not exist in the knowledge base. As such, negative types properties do not correspond to shortcuts within the CIDOC CRM.

3.1.3. Identifying contradictions

In order to ensure that one is able to identify contradictory typed statements in a knowledge base first order logic statements need to be provided to allow such contradictions to emerge. The typed properties and the negative typed properties are exclusive and their co-existence for the same domain and range instances is contradictory. Therefore we can say:

$NTP46(x,y) \rightarrow \neg TP46(x,y)$, which is equivalent to:

$TP46(x,y) \rightarrow \neg NTP46(x,y)$

Either of these two sentences indicate that if the instances for variables (x) and (y) are connected by the property 'TP46 is composed of physical thing of type (is type of physical thing which forms part of)' then they are not connected with the property 'NTP46 is not composed of physical thing of type (is not type of physical thing which forms part of)'. If they are asserted to be, then a contradiction in the knowledge base is produced.

Based on what was discussed in the description of typed properties, it follows that the existence of an instance (z), such as:

a → P46 is composed of (forms part of) → z

z → P2 has type (is type of) → b

would indicate that:

a → TP46 is composed of physical thing of type (is type of physical thing which forms part of) → b

Therefore:

$P46(x,z) \wedge P2(z,y) \rightarrow \neg NTP46(x,y)$

In other words if there is an instance (z) of a specific type (y) and connected to the instance (x), then the negative typed property between (x) and (y) indicates a contradiction.

3.1.4. Hierarchy of negative typed properties

The CIDOC CRM is based on the idea of property inheritance where more specific classes can be described with all properties of their parent classes. The same principle remains unaltered for the typed prop-

erties and the negative typed properties. We maintain that the property hierarchy as described in the CIDOC CRM can be mirrored with the typed properties and negative typed properties. For example, the CIDOC CRM defined 'P46 is composed of (forms part of)' as a superproperty of 'P56 bears feature (is found on)'. Such hierarchical arrangement is also valid for:

'TP46 is composed of physical thing of type (is type of physical thing which forms part of)' being a superproperty for 'TP56 bears feature of type (is type of feature found on)' and 'NTP46 is not composed of physical thing of type (is not type of physical thing which forms part of)' being a superproperty of 'NTP56 does not bear feature of type (is not type of feature found on)'.

3.2. Epistemology of TPs and NTPs and Open World assumption

Having seen how the typed properties and negative typed properties can be used to query a knowledge base and retrieve contradictory statements, it is appropriate to consider principles that knowledge bases adopt when it comes to missing statements.

Some knowledge bases assume that if a statement about something cannot be inferred then this does not mean that the statement is false or corresponds to non-existent things. For example if a knowledge base of this kind does not include a statement about leaf markers for a book then we cannot conclude that this book does not have leaf markers, but consider that as a possibility and also that the knowledge base may be missing this statement either because making it was not considered important or that the surveyor was not able to observe leaf markers on the book. This approach conforms to the Open World assumption. The CIDOC CRM observes the Open World assumption as described in the introduction of the CIDOC CRM definition document (about "Monotonicity, Minimality and Open World") to ensure that different views of the world are not excluded through the definition of classes and properties. For example there are no CIDOC CRM classes declared as complement classes. Complement classes, i.e. classes including all instances which do not belong to another class, would mean that users of the CIDOC CRM would assume complete knowledge of a domain by being able to describe all that there could be in it (the class and the complement class). To avoid such problematic positions claiming absolute knowledge, the CIDOC CRM assumes that statements about things are incomplete and only provide a partial view of the world.

Some knowledge bases assume that if a statement about something does not exist then this does indeed mean that the statement is false. For example if a knowledge base of this kind does not imply a statement about leaf markers for a book then we should conclude that this book does not have any leaf markers. This approach conforms to the Closed World assumption. The negative typed properties described above point towards the Closed World assumption. For example making the statement:

a → NTP46 is not composed of physical thing of type (is not type of physical thing which forms part of) → b

implies that in the knowledge base we do not have statements like these:

a → P46 is composed of (forms part of) → z

z → P2 has type (is type of) → b

and that a possible instance (z) does not exist. By using a negative typed property we assume complete knowledge and capacity for observation which goes against the Open World assumption of the CIDOC CRM.

Razniewski and Nutt [14] have summarised the nature of partially-complete knowledge bases which follow neither the Open World assumption, nor the Closed World assumption. Instead these knowledge bases consider completeness in relation to querying. Queries need to be characterised based on completeness to allow users to understand whether the results assume an Open or Closed World. This characterisation can be done through providing contextual information about data completeness. In our case, this essentially reflects completeness of observation of a bookbinding structure. The conditions affecting completeness in observation can be considered as the contextual information for data completeness. Darari et al. [2] explore the Semantic Web as an Open World dataset with pockets of complete data assuming a Closed World. In a similar fashion they consider the certainty of answers as a metric to evaluate results of queries by comparing to a hypothetical complete dataset within a given context. Next we consider how the context of the Closed World for the negative typed property is established within an otherwise Open World knowledge base. By examining relatively small things like bookbindings (in contrast to large areas of potential observation which is the case for biodiversity studies) we are confident that such Closed Worlds can be established (in theory we

have a complete dataset) through a context which requires:

The domain instance of the property. Properties describe aspects of the domain instance and as such any observation for that description must be in relation to that instance. In our example the book (instance of ‘E18 Physical Thing’) that may or may not contain gold-tooled features is the domain instance providing the context.

The CIDOC CRM property from which the negative property derives. This is necessary context as it provides the type of observation that needs to be undertaken on the instance. For example the context of ‘NTP46 is not composed of physical thing of type (is not type of physical thing which forms part of)’ is given by ‘P46 is composed of (forms part of)’ which indicates the kind of observation required, i.e. whether a part of the book is a component of a specific type.

The instance of ‘E55 Type’. This is again necessary to examine the value that the property should take. The instance of the book may have component parts of one type but not another, so testing the constraint of the component type is necessary to establish a statement using the negative property.

The Closed World observation can therefore be described through the constraints imposed by these three parameters for every negative typed property statement. In the example with the hidden sewing supports the first and third of these constraints can be established since a) the specific book is available to us and b) we have complete understanding of what the component type ‘sewing support’ is. We cannot, however, establish the second constraint which is the property requiring the observation that a component is part of the book instance. In this case we are discussing potentially real individuals.

In the example of the gold-tooled decoration, all three constraints can be established since a) the book is available to us, b) we can observe its cover which is finite in its totality and c) we have complete understanding of what the component type ‘gold-tooled decoration’ is. In the case we are discussing absence of any real individuals.

3.3. Negative typed properties and thesauri

Having established an understanding of the kinds of statements that can be provided by typed properties and negative typed properties it is important to examine how the use of external thesauri is impacted by these properties. As mentioned before, the CIDOC CRM uses the class ‘E55 Type’ as a bridge to connect to expert thesauri. In some cases these thesauri are hierarchical using broader/narrower relationships provided by standards like ISO 25964-1:2011 [7] and SKOS [11]. In the field of bookbinding history the Language of Bindings Thesaurus [19] provides such relationships. The concept for ‘gold-tooled (decoration)’ has broader concept ‘tooled-decoration’ [27]. So we have:

gold-tooled (decoration) → broader → tooled-decoration

Also, a normal categorical statement about the type of a component would be:

cover feature (E25 Human-Made Feature) →P2 has type (is type of) → gold-tooled (decoration)

From these two statements it follows that:

cover feature (E25 Human-Made Feature) →P2 has type (is type of) → tooled decoration

Therefore searching based on the category ‘tooled decoration’ will also bring up results for ‘gold-tooled (decoration)’.

The same principle applies to typed properties. For example from the statement:

book cover (E18 Physical Thing) → TP46 is composed of physical thing of type (is type of physical thing which forms part of) → gold-tooled (decoration)

it follows that:

book cover (E18 Physical Thing) → TP46 is composed of physical thing of type (is type of physical thing which forms part of) → tooled decoration

This also confirms the fact that typed properties are shortcuts of CIDOC CRM longer paths.

If we test this for negative typed properties, the principle does not stand. For example from the statement:

book cover (E18 Physical Thing) → NTP46 is not composed of physical thing of type (is not type of physical thing which forms part of) → gold-tooled (decoration)

it does not follow that:

book cover (E18 Physical Thing) → NTP46 is not composed of physical thing of type (is not type of physical thing which forms part of) → tooled decoration

However, if we start with a statement using the broader type:

book cover (E18 Physical Thing) → NTP46 is not composed of physical thing of type (is not type of physical thing which forms part of) → tooled decoration

it follows that:

book cover (E18 Physical Thing) → NTP46 is not composed of physical thing of type (is not type of physical thing which forms part of) → gold-tooled (decoration)

Therefore, the reasoning based on broader/narrower relationships in thesauri can be considered reversed when using negative typed properties.

4. Test dataset and future work

Our initial incentive for this work was to enable the identification of contradictory statements about the existence or not of individuals. As a result of the preparatory work done for the development of typed properties and negative typed properties in the CIDOC CRM, we have also developed a dataset for testing the capacity of software applications to identify contradictory statements automatically based on the logic described above. The dataset is a hypothetical set of automatically produced statements about bookbinding features. It includes manually produced contradictory statements which coexist in the dataset. These contradictory statements are the target of the automatic identification exercise. The dataset is large enough (several thousands of records) to also test performance of software. We are reporting more extensively on this in a separate publication [17].

We are in the process of producing a formal CIDOC CRM extension for typed properties and negative typed properties, providing labels and scope notes for each. The development of that extension is undertaken as part of work for the Linked Conservation Data project [25]: a project which explores ways of sharing data produced by conservators with significant representation from book and paper conservators working with historic books. The progress of the development of the extension can be followed in the Linked Conservation Data GitHub repository [9].

5. Summary of conclusions

When documenting heritage, we often require making statements about a) the typology of numerous individuals without having to instantiate each one of them and b) the non-existence of individuals. In this work we explained the three types of statements about individuals and non-existence: a) individuals as ideas (with which we are not concerned), b) potentially real individuals and c) absence of any real individuals. We explained that neither case b) not case c) require instantiation of individuals, but instead relevant statements focus on the constraints based on which our observation is limited. When we are able to observe an environment in full, we establish statements in a Closed World assumption. When the type of individual that we are looking for cannot be observed then we can conclude that no individual of that type exists within the context of our Closed World. Therefore our statements are not about individuals but instead about types of individuals. We propose that CIDOC CRM properties can be used to derive typed properties and negative typed properties allowing us to make statements about the types of numerous individuals and the absence of individuals of specific type. The context of the Closed Worlds we establish is given by the domain and range of the negative typed property as well as the original CIDOC CRM property from which the negative typed property is derived. Discussing the epistemology of typed and negative typed properties we showed how typed properties are shortcuts in the CIDOC CRM whereas negative typed properties are not. The hierarchy of typed and negative typed properties mirrors that of the CIDOC CRM property hierarchy. When discussing reasoning about broader/narrower concepts from thesauri, we observe that statements using negative typed properties also apply to narrower terms of a thesaurus in contrast to typed properties where this is not the case.

6. Acknowledgements

The authors thank the CIDOC CRM special interest group. This work has been initiated by the Linked Conservation Data project and is partly funded by the Arts and Humanities Research Council in the UK. The problem was originally introduced by Prof. Nicholas Pickwoad during the condition survey of the manuscripts of the library of the St. Catherine's Monastery in Sinai, Egypt.

References

[1] G. Boudalis, Endbands in Greek-style bindings, *The Paper Conservator* **31** (2007) 29–49. doi:10.1080/03094227.2007.9638652.

[2] F. Darari, S. Razniewski, and W. Nutt, Bridging the Semantic Gap between RDF and SPARQL using Completeness Statements, in: *The 13th International Semantic Web Conference* ArXiv:1408.6395 [Cs], Riva del Garda, Trentino, Italy, 2014. http://arxiv.org/abs/1408.6395 (accessed February 12, 2021).

[3] M. Doerr, The CIDOC Conceptual Reference Module: An Ontological Approach to Semantic Interoperability of Metadata, *AI Magazine* **24** (2003) 75. doi:10.1609/aimag.v24i3.1720.

[4] M. Doerr, Modelling Learning Subjects as Relationships, in: G. Grieser, and Y. Tanaka (Eds.), *Intuitive Human Interfaces for Organizing and Accessing Intellectual Assets*, Springer, Berlin, Heidelberg, 2005: pp. 201–214. doi:10.1007/978-3-540-32279-5_14.

[5] International Federation of Library Associations and Institutions, Working Group on Functional Requirements and Numbering of Authority Records, and G.E. Patton, *Functional requirements for authority data: a conceptual model*, K. G. Saur, München, 2009.

[6] International Organization for Standardization, *ISO 21127: Information and documentation - a reference ontology for the interchange of cultural heritage information*, ISO, Geneva, 2006.

[7] International Organization for Standardization, *ISO 25964-2: Information and documentation - thesauri and interoperability with other vocabularies*, ISO, Geneva, 2013.

[8] C.-H. Lin, J.-S. Hong, and M. Doerr, Issues in an inference platform for generating deductive knowledge: a case study in cultural heritage digital libraries using the CIDOC CRM, *International Journal of Digital Libraries* **8** (2008) 115–132. doi:10.1007/s00799-008-0034-0.

[9] Linked Conservation Data Consortium, *linked-conservation-data/crmntp*, (2021). https://github.com/linked-conservation-data/crmntp (accessed February 24, 2021).

[10] C. Meghini, and M. Doerr, A first-order logic expression of the CIDOC conceptual reference model, *International Journal of Metadata, Semantics and Ontologies* **13** (2018) 131–149. doi:10.1504/IJMSO.2018.098393.

[11] A. Miles, and S. Bechhofer, *SKOS Simple Knowledge Organization System Reference*, World Wide Web Consortium. (2009). http://www.w3.org/TR/2009/REC-skos-reference-20090818/ (accessed December 30, 2015).

[12] N. Pickwoad, *Assessment Manual*, University of the Arts London, London, 2005. https://www.ligatus.org.uk/node/19.

[13] N. Pickwoad, Recording medieval bindings – The role of the conservation survey, with reference to work currently under way in the library of the monastery of St Catherine on Mount Sinai, in: G. Lanoë, and G. Grand (Eds.), *La Reliure Médiévale*, Brepols, Paris, 2008: pp. 47–59.

[14] S. Razniewski, and W. Nutt, Databases under the Partial Closed-world Assumption: A Survey, in: *Proceedings of the 26th GI-Workshop Grundlagen von Datenbanken, CEUR*, Bozen-Bolzano, Italy, 2014: p. 6.

[15] M. Reicher, Nonexistent Objects, in: E.N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy, Winter 2019*, Metaphysics Research Lab, Stanford University, 2019. https://plato.stanford.edu/archives/win2019/entries/nonexistent-objects/ (accessed February 1, 2021).

[16] M. Sakashita, *Sea Otter*, Centre for Biological Diversity. (n.d.). https://www.biologicaldiversity.org/species/mammals/sea_otter/index.html (accessed February 24, 2021).

[17] S.D. Stead, The Linked Conservation Data Semantic Test Data Set, in: *Semantic Web Journal*, forthcoming.

[18] A. Velios, Bookbinding Descriptions in a Linked Data World: How the CIDOC-CRM Can Improve Research in Bookbinding History, in: N. Golob, and J. Vodopivec Tomažič (Eds.), *Bookbindings, Bibliologia*, Brepols, Turnhout, 2017: pp. 13–26.

[19] A. Velios, and N. Pickwoad, The development of the Language of Bindings Thesaurus, in: A. Campagnolo (Ed.), *Book Conservation and Digitization - The Challenges of Dialogue and Collaboration*, ARC Humanities Press, 2020: pp. 157–168. https://www.aup.nl/en/book/9781641890533/book-conservation-and-digitization.

[20] A. Velios, N. Pickwoad, and A. Martin, Multilingual bookbinding terms: Complex conceptual issues, *Newsletter of the ICOM CIDOC* **1** (2014) 13.

[21] M. Žumer, and P. Riva, IFLA LRM - Finally Here, *International Conference on Dublin Core and Metadata Applications* (2017) 13–23.

[22] 11th Meeting on FRBR/CRM Harmonization together with 16th CIDOC CRM SIG Meeting Germanisches Nationalmuseum, Nuremberg 4 – 7 December 2007, CIDOC CRM Conceptual Reference Model (2007). http://www.cidoc-crm.org/Meeting/16th-cidoc-crm-and-11th-frbr-crm.

[23] gold-tooled (decoration), Language of Bindings (n.d.). https://w3id.org/lob/concept/1371 (accessed February 25, 2021).

[24] leaf markers, Language of Bindings (n.d.). https://w3id.org/lob/concept/5423 (accessed February 25, 2021).

[25] Linked Conservation Data, (n.d.) https://www.ligatus.org.uk/lcd/ (accessed February 24, 2021).

[26] sewing supports, Language of Bindings (n.d.). https://w3id.org/lob/concept/1585 (accessed February 24, 2021).

[27] tooled decoration, Language of Bindings (n.d.). https://w3id.org/lob/concept/1678 (accessed February 25, 2021).