Referent tracking for Digital Rights Management

Werner Ceusters*

Ontology Research Group,
NY State Center of Excellence in Bioinformatics and Life Sciences,
701 Ellicott Street, Buffalo, New York 14203, USA
E-mail: ceusters@buffalo.edu
*Corresponding author

Barry Smith

Department of Philosophy,
University at Buffalo,
135 Park Hall, Buffalo, NY 14260, USA
E-mail: phismith@buffalo.edu

Abstract: Digital Rights Management (DRM) covers the description, identification, trading, protection, monitoring and tracking of all forms of rights over both tangible and intangible assets. The Digital Object Identifier (DOI) system provides a framework for the persistent identification of entities involved in this domain. Although the system has been very well designed to manage object identifiers, some important questions relating to the creation and assignment of identifiers are left open. The paradigm of a Referent Tracking System (RTS) recently advanced in the healthcare and life sciences environment is able to fill these gaps. This is demonstrated by pointing out inconsistencies in the existing DOI models and by showing how they can be corrected using an RTS.

Keywords: referent tracking; digital rights management; DRM; digital object identifier system; DOI.

Reference to this paper should be made as follows: Ceusters, W. and Smith, B. (xxxx) 'Referent tracking for Digital Rights Management', *Int. J. Metadata, Semantics and Ontology*, Vol. x, No. x, pp.xxx–xxx.

Biographical notes: Werner Ceusters studied Medicine, Neuropsychiatry, Informatics and Knowledge Engineering in Belgium. Since April 2006, he has been Health Sciences Professor and Coordinator of Bioinformatics for the Health Science Faculties in the University at Buffalo, NY, and Director of the Ontology Research Group of the New York State Center of Excellence in Bioinformatics and Life Sciences in Buffalo NY.

Barry Smith is Julian Park Distinguished Professor of Philosophy in the University at Buffalo (New York, USA) and Director of the Institute for Formal Ontology and Medical Information Science in Saarbrücken, Germany. His current research focus is ontology and its applications in biomedicine and biomedical informatics, where he is working on projects relating to biomedical terminologies and electronic health records.

1 Introduction

Digital Rights Management (DRM) is a paradigm that holds much promise for novel business models related to the distribution of and access to digital content such as text, music, movies and software. It also tries to manage the interests of the holders of rights in non-digital assets, such as ideas or performances, and in modifications thereof, for example when a fragment of a performance is incorporated into a TV broadcast (Cunard et al., 2003). In order to make DRM systems semantically interoperable, several initiatives for standardisation have been proposed (Wegner, 2002). Initially, work was focused on a syntactic approach, for example through the formalisation of XML Document Type Definitions and Schemas to form what were called Rights Expression Languages (XML-DRM). More recently, ways are being explored to make DRM systems suitable for operation under the Semantic Web paradigm (García et al., 2004).

An essential element in DRM is the unique identification of certain key entities such as persons and organisations, the property or other rights they enjoy, and the assets to be protected (Rust and Bide, 2000). In 1998, the International DOI Foundation was created to support the development and promotion of a DRM system centred on the notion of a *Digital Object Identifier* (DOI). A DOI is a single, unambiguous and persistent string that

references a single entity and that is generated on the basis of a consistent syntactic frame (a '*numbering scheme*' as defined in the NISO standard ANSI/NISO Z39.84), in such a way that it can be expressed in a form suitable for use in an automated system (The DOI Handbook, 2005). The DOI system is a specific implementation of the Uniform Resource Identifier paradigm advanced by W3C, further complemented with management policies for use in the domain of DRM as defined by the DOI foundation.

The initial focus for DOIs was on '*creations*' that is, resources made by human beings such as art works, scientific papers, or theatre plays, rather than natural objects, people, places, or events. However, since the latter are also involved in intellectual property transactions, they, too, may be identified by DOIs where appropriate. *Creations* exist in tangible (pictures, paintings) and intangible forms (games, work-out routines, diets). The totality of creations is moreover to be interpreted as including not only *manifestations* (such as books) but also *expressions* such as performances. Especially for the latter, it is extremely important to be able to distinguish, and hence uniquely identify, various secondary derivations. For instance, a performance by the Buffalo Chamber Orchestra on a specific day of a work by David Felder might be recorded by two different music or video/DVD publishing companies, with different sorts of equipment and subsequent editing, thereby producing recordings that capture the original performance in different ways. The recordings will be copied many times and sold to various individual customers. Some recordings may be used for public broadcasting, others be restricted for domestic use. In such cases, digital watermarking, a technique used to hide unique identifiers in copies of digital recordings of music or digital images (Hartung and Kutter, 1999), allows copyright owners to identify such copies, but only on the condition that an asset rights management system is in place from the beginning.

To bring some form of organisation into the wealth of entities to which DOIs can be assigned, the DOI system relies on an analysis method and data dictionary maintained by Indecs Framework Ltd. (Indecs). Unfortunately, as we shall see, the work of the latter is marked by a lack of formal rigour and this gives rise to several flaws in its definitions and leaves many questions unanswered. As an example, the description of '*creator*' and of the other entities that are used in this description (see Table 1) does not seem to allow painters to be included under this heading. This is because their work, obviously, requires at least the use of a brush, and the description of '*input*' suggests that a brush is to be regarded as an input of their work. At the same time, however, an '*original creation*' – that to which a '*creator*' contributes – is required to be without '*input*'. For the same reason, Michelangelo's Mona Lisa would not qualify as an '*original creation*', since Michelangelo used a certain lady as '*input*'.

Table 1 Description of the entity 'creator' and of the entities that are further used in its description (see also Table 2 for entities not further described here)

Element	*Description*
creator	A *party* contributing to the making of an *original creation*, in whole or in part
party	An *agent* undertaking an *activity* or task in a creative or commercial *relation*
agent	An *entity* acting in an *event* or sustaining a *situation*; a characteristic active *role* undertaken by an *entity*
creation	The *output* of creative activity
original creation	A *creation* without a source *input*
input	A pre-existing *entity* which participates in a *relation* in a passive, qualifying or supportive *role*
output	An *entity* created or changed through an *event*
role	A part played or function fulfilled by an *entity* in relation to another *entity* or entities; a classification of an *entity* in terms of its external *relations*; an extrinsic classification

Source: Adapted from Rust and Bide (2000)

Another problem is the level of detail that must be taken into account when entities are to be identified as instances of specific types in accordance with the provisions of the Indecs framework. A DOI can be applied at any level of granularity – for example to an entire document or to any part or version thereof. Which entity is at issue should in every case be specified before a DOI is assigned. Following the Indecs principle of '*Functional Granularity*', however, the decision as to what a DOI identifies is to be taken by the registrant in a purely *ad hoc* functional way on the basis of the (surely over-optimistic) assumption that "*it should be possible to identify an entity whenever it needs to be distinguished*" (Rust and Bide, 2000). Unfortunately, neither Indecs nor DOI provide clear answers to questions such as:

- in what way are '*creations*' different from '*expressions*' or '*manifestations*'? (an ontological issue)
- how are these different sorts of entities to be differentiated from each other by users of the system? (an epistemological issue).

As an example, the DOI handbook states that

> "A publisher could consider the English and Spanish to be different 'versions' of the same underlying 'work' or 'creation' (similar to having both a pdf and html version) in which case one DOI. Or a publisher could consider them two separate underlying works, hence two DOIs. These could perhaps be related in one or more applications using the Indecs entities and relationships or they could be grouped together under a third DOI for the work." (The DOI Handbook, 2005, par 1.6.4.)

It is here, we believe, that the referent tracking approach can contribute much needed additional clarity – in ways which will have practical consequences not just for the adequate

application of the intended DOI principles, but also for querying of the DOI system or for linking it to other systems such as digital libraries, multimedia archives, etc., for example within the framework of the Semantic Web.

This paper is organised as follows: we first briefly describe referent tracking, a paradigm designed to refer explicitly to entities in reality and to build composite representations out of these references. Because referent tracking is based on Basic Formal Ontology (BFO), we also summarise that theory and show how it is contrasted to the sort of analysis that has been used in building the Indecs and DOI models which are the subject of our critique. We then suggest some directions in which these models can be improved, first in terms of a better upper ontology, and second by introducing appropriate relationships.

2 Referent tracking

In Ceusters and Smith (2005), referent tracking was introduced as a new paradigm for entry and retrieval of data in the Electronic Health Record (EHR). Its purpose is to avoid the ambiguity that arises when statements in an EHR refer to disorders, lesions and other entities on the side of the patient exclusively by means of generic terms from a standard terminology or ontology. Suppose that two different physicians are treating the same patient A, and that each enters into A's EHR a statement to the effect that A suffers (i) from diabetes or (ii) from a fracture of the right lower arm. Then it is in either case left unspecified whether they are referring to the same or to different entities on the side of the patient. In case (i), medical science makes clear that only one answer is possible, since one can suffer from only one instance of the disease diabetes; yet the ambiguity as to whether each of the two physicians is referring to the same diabetes will still cause problems for software agents programmed to make inferences from the data. In case (ii) this ambiguity causes problems even for human beings, since the physicians in question might have been referring either to the same or to different fractures.

Referent tracking avoids such ambiguities by introducing, for each concrete individual entity relevant to the accurate description of a patient's condition and of associated therapies and outcomes, a unique identifier (called 'IUI', for '*Instance Unique Identifier*').

Referent tracking distinguishes between *IUI creation*, a matter of appropriate software, IUI *assignment* (allowed only in relation to entities that exist or have existed in the past), and *IUI reservation*, which is a provision made for entities that are expected to come into existence in the future. When a physician orders an X-ray, the order itself does already exist, and can thus be assigned a IUI. When the order is accepted by the radiology department, then a plan to execute the order exists also, and that plan, too, can be assigned a IUI, though this will of course be different from the IUI assigned to the initial order. The X-ray itself, however, cannot be assigned a IUI until it has been performed. At best, a IUI for that X-ray can be reserved at the time of the order, so that statements of the form "I expect to see the tumour on the X-ray that will be taken tomorrow" can be encoded.

IUIs are to be used in a *Referent Tracking System* (RTS), which implements the following requirements (Ceusters and Smith, 2006):

- a mechanism for generating IUIs that are guaranteed to be unique strings
- a procedure for deciding what particulars should receive IUIs
- protocols for determining whether or not a particular has already been assigned a IUI (except for some exceptional configurations that are beyond the scope of this paper, each particular should receive maximally one IUI)
- rules governing the use of IUIs in other systems such as EHR systems and digital library systems, which are designed to resolve issues concerning the syntax and semantics of statements containing IUIs
- methods for determining the truth values of propositions that are expressed through descriptions in which IUIs are used
- methods for correcting errors in the assignment of IUIs, and for investigating the results of assigning alternative IUIs to problematic cases
- methods for taking account of changes in the reality to which IUIs get assigned, for example when particulars merge or split.

3 Realist ontology makes the difference

Although the DOI and RTS paradigms were developed independently, they share a number of common features. Most prominent is the recognition of the need for persistent and unique identifiers referring in unambiguous fashion to particular entities in reality, whether material (*books*, *tumours*) or immaterial (*works*, *treatment plans*). Another is to have identifiers be supported by a system that implements certain policies. But it is precisely in the nature of these policies that major differences between DOI and RTS arise, the most fundamental concerning the policies proposed for distinguishing the kinds of entities to be identified and the ways these entities are to be described. Whereas DOI is based on the (in some areas rather superficial) concept-based analysis of the Indecs Framework, RTS relies on an ontological analysis that is grounded in a thorough-going realism, so that its application demands a careful analysis of the types of entities by which the relevant domain is populated, and of the relations between them.

3.1 DOI: the world through 'models'

Although the Indecs developers did a much better job in the DRM domain than did most of the terminology and model builders in the domain of healthcare informatics

(Smith and Ceusters, 2006), their work, because it is based on the ISO 11179 standard (Rust and Bide, 2000, p.11), exhibits the confusions typical of what (Smith et al., 2005b) have called '*Wüsteria*' – the main feature of which is that terms in a terminology are asserted to correspond not to entities in reality but rather to concepts in what are called 'concept systems', while at the same time no facility is provided for the establishment of some benchmark in relation to which a postulated concept system developed could be established as correct. Often, concepts are confused with the corresponding entities in reality, in a way which fosters also further confusions, for example between entities and the data pertaining thereto.

Several statements in the documentation of the Indecs Framework description exhibit this latter confusion:

- "*The <indecs> model elaborates a logical and semantic framework for describing entities, their attributes and, where appropriate, values of each. Entities, attributes and values are referred to as types of metadata elements*" (Rust and Bide, 2000, p.11). Bear in mind that '*entity*' is defined by Indecs as '*something which is identified*' (see Table 2), and thus may refer to persons of flesh and blood. It is then hard to see how persons such as you and me can be considered to *be* metadata (or, if one prefers, types of metadata, or types of metadata elements). This statement violates the standard (and surely correct) view according to which data are *about* entities in reality, and metadata are *about* data.
- "*A thing [i.e., an entity] must be both thought about or perceived and identified before it exists in a metadata framework*" (Rust and Bide, 2000, p.12). This makes sense only if it is to be interpreted in such a way that data about a thing can exist *in a metadata framework* only if the thing has been perceived or thought about.
- "*All metadata relationships are either events in themselves, or rely on events to establish them*" (Rust and Bide, 2000, p.13). We are at a loss as to what this statement might mean, and the definition of 'event' (provided below) does not add further light.
- "*Electronic trading depends to a far greater extent than traditional commerce on the way in which things are identified (whether they are people, stuff or deals) and the terms in which they are described (metadata, or data about data)*" (Rust and Bide, 2000, p.4). The use of 'they' here (in '*they are described*') thus identifies people, stuff and deals with data about people, stuff and deals.

These last two statements are, in addition, difficult to line up with the definition that: "A*n item of metadata [for example that something is an 'entity' or 'attribute'] is a relationship that someone claims to exist between two entities*", and with the additional comment that "*It raises the question of authority: the identification of the person making the claim is as significant as the identification of any other entity*" (Rust and Bide, 2000, p.11). And this raises an even more basic question, namely: what are the relationships that obtain amongst the mentioned entities as they are in themselves, independently of those relationships that are claimed to obtain? Is this relationship ever relevant to the tasks of DRM, for example because careful distinctions between *claimed* and actual relationships would provide us with the facility to deal with those cases where claims of relationships are made in error?

Table 2 Top-level ontology of the INDECS framework

Element	*Definition*	*Hierarchy*
Entity	Something which is identified	Concept
Percept	An *entity* which is perceived directly with at least one of the five senses (derived)	Entity
Being	An *entity* which has the characteristics of animate life (derived); anything which lives and dies	Percept
Thing	An *entity* without the characteristics of animate life (derived)	Percept
Relation	The interaction of *percepts* and/or *concepts*; a connection between two or more *entities*	Entity
Event	A dynamic *relation* involving two or more *entities* (derived); something that happens; a *relation* through which an attribute of an *entity* is changed, added or removed	Relation
Situation	A static *relation* involving two or more *entities* (derived); something that continues to be the case; a relation in which the attributes of *entities* remain unchanged	Relation
Concept	An *entity* which cannot be perceived directly through the mode of one of the five senses (derived); an abstract *entity*, a notion or idea; an abstract noun; an unobservable proposition which exists independently of time and space	Entity

We accept, with Indecs, that a representation system needs to represent only what is relevant to its intended purpose. Where we differ is in regard to the number and kinds of distinctions which need to be drawn if an information system in the realm of DRM is to realise the very intended purpose for which the Indecs framework was designed. Thus Indecs does not seem to recognise that the nature of people, stuff or deals in reality is not a matter of choice or decision. A given portion of reality does not become something different just because it is 'analysed' from a different perspective. That a performance necessarily occurs somewhere in space-time is a matter of reality, not of analysis. Yet we find:

> "Stuff may be analysed, for example, in terms of molecular entities (chemistry), particles such as electrons, quarks or superstrings (physics), spatial co-ordinates (geography), biological functions (biology, medicine), genres of expression (creations), price categories (commerce), and so on"

which is taken as an argument for the thesis that "*The basic 'elements' of a resource* [i.e., of 'stuff'] *may be entirely different according to your purpose*" (Rust and Bide, 2000, p.10). Molecules, prices, or other 'basic elements' do not become different according to the different purposes we have in mind when building information systems. Rather they are *viewed* differently in light of such different purposes, and again, confusion can only result, if the way things *are* is confused with the way things are viewed in one or other context. Such confusion is indeed nicely encapsulated in Indecs' own 'Fifth axiom': "*everything is a view*" (Rust and Bide, 2000, p.12).

The model-based approach adhered to by Indecs (Rust and Bide, 2000, p.4) is responsible for some further dubious features of what could be seen as its top-level ontology (Table 2).

First, there is the circular 'subtype' relationship between 'entity' and 'concept'.

Second, is the strange condition (again derived from ISO 11179) that for an animate or inanimate entity to exist, it must have been perceived; and even worse: that "*nothing exists in any useful sense until it is identified*" (Rust and Bide, 2000, p.12). On a sensible realist view, in contrast, perception itself is recognised as a process which relates entities perceived to a perceiving entity, and the latter always pre-exist the process of perception itself.

3.2 RTS: the world through ontology

Referent tracking is based on BFO, a theory proposed in the recent literature of ontological realism (Grenon and Smith, 2004). BFO rests on the idea that it is necessary to develop an ontology that remains as close as possible to our widely shared and continuously tested intuitions about the objects and processes in reality. It consists in a number of sub-ontologies, corresponding to the fundamental division between continuants (persons, manuscripts, videotapes, CDs) on the one hand, entities which endure self-identically through time), and occurrents (performances, perceptions, events of broadcasting), entities which can be divided along the temporal axis into successive phases.

Each continuant ontology is a partition of the totality of objects and their continuant qualities, roles, functions, etc., existing in a given domain of reality at a given time. Each occurrent ontology is a partition of the totality of processes unfolding in a given domain across a given temporal interval.

Continuants and occurrents themselves exist only in mutual dependence on each other. Continuants require processes in order to be maintained in existence; processes require continuants as their bearers or carriers. Like the Indecs Framework, BFO serves as the basis also for a series of sub-ontologies at different levels of granularity, reflecting the fact that the same portion of reality can be apprehended in an ontology at a plurality of different levels of coarser or finer grain (from whole symphonies to individual notes). BFO recognises that such partitions will be determined in each case by the purposes for which an ontology is developed. In contrast to Indecs, however, it does not conclude from this that objects are to be identified with views or with associated purpose-specific levels of granularity.

Another difference between DOI and RTS is the careful treatment in the latter of both universals and their particular instances, the former (also called 'types' or 'kinds') being determined by the fact that there are intrinsic features which the latter share in common. Where DOI refers only to particulars ("*something which is identified*") such as *this essay by these authors*, RTS thus takes account also of universals, i.e., generic entities such as *essay* or *author*, which according to the philosophy of realism are as real as the particulars by which they are instantiated.

4 A realist's view of DOI

A disadvantage of working with models, rather than with the entities themselves as they exist in reality, is the absence of any reliable method for testing whether or not a model corresponds to anything that is real, or is a faithful representation of reality. Thus we are not surprised to find statements in the Indecs Framework such as:

> "it is meaningful, for example, to say that John Williams, Marilyn Monroe, the London Philharmonic Orchestra and Mickey Mouse are all performers, even though one is a "real" human being, one is using a stage persona, one is a name that represents a constantly changing group of individuals, and one is a fictional cartoon character." (Rust and Bide, 2000, p.25) (Note how the authors refer to the name 'the London Philharmonic Orchestra', where the other items on the list are used to refer to the corresponding bearers. This reveals once again the confusion between a name and what it designates.)

Even less are we surprised to read that, as part of the DOI policies, "*Reverse look-up (from metadata to a DOI) is not a function of the DOI system itself*" (The DOI Handbook, 2005, par. 6.3). For in the haphazard way the Indecs data dictionary is currently built, it would indeed be a very tricky endeavour to perform meaningful queries.

We argue that DOI would benefit considerably from a principles-based revision of its underlying framework guided by BFO, along the same lines already demonstrated in the biomedical domain in the improvements realised in systems such as the OBO Ontologies (Smith et al., 2005a) and the Foundational Model of Anatomy (Rosse and Mejino, 2003).

This effort would consist in

- building a coherent ontology of the various types of entities referred to in the Indecs data dictionary
- giving a formal and logical account of the relevant relationships between these entities in reality, and doing this in ways which reflect the separate roles of universals and particulars in the specific domain of intellectual products.

4.1 Towards a realist version of the DOI upper ontology

4.1.1 Three levels of entities

In a realist view of the world, an *entity* is anything which exists, including objects, processes, qualities and states. A first distinction can made between three levels of entities (Smith et al., 2006):

- *Level 1*: the objects, processes, qualities, states, etc., in reality
- *Level 2*: cognitive representations of this reality on the part of cognitive beings
- *Level 3*: concretisations of these cognitive representations in textual and graphical artifacts.

Level 1 reflects the assumption that (as we hope) those who see themselves as building for example 'data models' in the domain of right managements are attempting to create artifacts which stand ultimately in some representational relation to entities in the real world. Level 2 reflects the fact that a crucial role is played in ontology and terminology development by the cognitive representations (including 'views') of human subjects. Level 3 reflects the fact that cognitive representations can be shared only when they are made communicable in a form whereby they can also be subjected to criticism and correction. Note that the three levels overlap; thus the textual and graphical artifacts distinguished in Level 3 are themselves objects on Level 1.

A *representation* is for example an idea, image, record, model, or description which refers to (is *of* or *about*) some entity or entities external to the representation. Most representations are built out of constituent sub-representations in the way in which paragraphs are built out of sentences and sentences out of words. The smallest constituent sub-representations are called *representational units* such as icons, names, simple word forms, or the sorts of alphanumeric identifiers we might find in patient records, including IUIs. IUIs, in general, are thus Level 3 entities that refer to the Level 1 entities out there in reality, and not to data about these entities. IUIs are also not *the entities themselves*. This might seem obvious, but use-mention confusions, in which an entity in reality and its digital representation are confounded, are abundantly present in the literature (Smith, 2004) – and we have encountered examples of such confusions already in the above.

Recognising this distinction would allow the entities that currently fit the various descriptions of '*concept*' in Table 2 to be categorised at the right level.

4.1.2 Dependent and independent entities

A second distinction is between *dependent* and *independent* entities. Independent entities (such as violins or keyboards) do not depend on any other entity in order to exist, while dependent entities (such as the *shape* of a violin, the *click-rate* of a keyboard) cannot exist without the existence of some other entity which serves as bearer or carrier.

At first sight, DOI's '*thing*' and '*being*' seem to comprehend independent entities, at least on the basis of the definitions provided (see Table 2).

Some caution is however required because of the subtype relationships from '*thing*' and '*being*' to '*percept*'. Since, presumably, percepts exist only if there are perceiving subjects, this would seem to employ that both things and beings are dependent entities after all – namely dependent on some perceiving act of a perceiving subject. We take it, however, that this is a mistake in DOI (following a parallel mistake in those terminology standards which are associated with the work of ISO TC37).

Additional caution is required in relation to '*thing*', because of its asserted disjointness from '*being*'. An entity is defined by Indecs as anything that is identified, and a percept as any entity which is perceived. The only difference between 'beings' and 'things', according to Indecs, is that the former are animate, the latter not. Yet entities such as weights, temperatures and colours are not animate and thus have to be qualified as 'things'. But in contrast to entities such as violins and keyboards, they clearly depend on their bearers (the entities which *have* weights or temperatures or colours). For this reason, whether or not 'thing' subsumes only independent entities could be assessed only by human inspection of the entire DOI data dictionary. This excludes the DOI framework from being used for automatic reasoning, and this in spite of the fact that it is claimed to have "*been validated against the W3C ontology language OWL-DL*" (DOI Fact sheets). As has been shown in the domain of biomedicine, validation against a description logic is by no means a sufficient guarantee against mistakes, and not even serious mistakes (Ceusters et al., 2004).

All other entities in the DOI upper ontology are to be categorised as dependent entities: a '*concept*' (specifically under its reading as '*idea*') depends on a cognitive agent. '*Relations*', '*events*' and '*situations*' clearly depend on those entities which serve as their relata or participants.

4.1.3 Continuants and occurrents

A third distinction is that between continuants and occurrents. DOI's '*thing*' and '*being*' seem, on the basis of an inspection of the DOI metadata dictionary (which includes terms such as '*audience*', '*creator*', '*organisation*'), to refer primarily to continuant entities,

which is to say to entities that are wholly present at any time of their existence even while they undergo changes of various sorts. An '*idea*', too, is a continuant entity (as contrasted with the occurrent process of someone's conceiving the idea). DOI's '*event*' and '*situation*', in contrast, are occurrents: they are only partially present at any given time. We do not endorse DOI's confusing claim that events are relations. Rather, we argue that some events (kisses, hits) are relational processes, and that all events are such as to stand in relations of dependence to the continuant entity or entities that partake in them.

4.1.4 Particulars, universals and defined classes

A fourth distinction is between particulars and universals on the one hand, and universals and defined classes on the other. As proposed in Smith et al. (2006), we use the term *portion of reality* to comprehend both single universals and particulars and their more or less complex combinations. A '*domain*' is a portion of reality that forms the subject-matter of a single science or technology or mode of study; in this case therefore: DRM. A DOI system keeps track of individual creators, their rights, the pieces of art they contribute to, recordings, books and so forth, thus of *particulars*. It is therefore an 'inventory': a representational artifact (Level 3) in which particulars are represented by means of representational units called 'DOIs', just as IUIs would be used in EHRs faithful to the referent tracking paradigm.

But where referent tracking uses exclusively realism-based ontologies for describing what sorts of entities given particulars are (i.e., ontologies whose representational units refer to *universals*, which are entities that are multiply located in space and time through their particular instances), the representational units under the DOI paradigm refer to concepts – which means, when the latter are analysed from a realist perspective, that the units in question rather refer either to *universals* or to *defined classes*. The latter belong to the realm of particulars; they are collections of particulars to which some general term refers.

An example of such a defined class would be a *percept* in Indecs parlance. A realist ontology would not accept 'percept' as designating a universal, for the same reason that it would not accept a universal instantiated by: *entities referred to in my diary*, since there is no intrinsic feature which the latter share in common. We argue that the DOI model would benefit considerably by specifying which representational units refer to universals, and which merely to defined classes or to particulars of other sorts. This would be of value also in preventing mistakes in logical reasoning when the DOI model is implemented in some form of executable logic. This is because the properties of the relationships used in reasoning (such as being transitive or symmetric) differ whether according to they are asserted to obtain between universals or particulars (Donnelly et al., 2006; Smith et al., 2005).

4.2 Towards a realist version of DOI relations

The Indecs framework does not provide formal definitions for the relations (such as *using*, *creating*, *modification*, etc) proposed in its ontology. This makes it hard to understand what exactly its authors are attempting to represent. Furthermore, there is an inconsistent use of what in the knowledge representation community is known as 'reification': "*Any entity fulfilling a role in a relation may then be said to be of the type described by the role*" (Rust and Bide, 2000, p.21). Indecs' '*percept*' entity is a typical example of such reification. Under the realist view, an entity such as a painting may participate in a perception event that ontologically depends on both

- a person who enjoys the agent role with respect to the perception
- the painting that is perceived (Smith, 1984).

The painting itself then stands in a relation to the perception that might be described as one of *being perceived*. This 'being perceived', however, does not warrant the introduction of a new universal (*percept*) of which the painting then would become an instance, or – as it is phrased in the Indecs documentation – of which the painting would "*be said to be of the type [being perceived]*". Being perceived reflects no intrinsic features of the corresponding object, any more than does: being more than 100 km from Cambridge, or: being not identical with my brother.

Specifically problematic for the DRM domain are the vague specifications provided of two foundationally distinct relations: transformation and modification. Thus 'transformingEvent' is defined as:

> "an event which results in the making of a new creation including elements of at least one existing creation; an event in which both creating and using occur." (Rust and Bide, 2000, p.22)

while '*modification*' is defined as: "*a creation made by changing a pre-existing creation of the same type (aka version)*" (Rust and Bide, 2000, p.30). How, on this basis, are we to establish which of these two relations is to be applied in any given case? Does the result of *modification* also constitute a new creation? Only by using formal definitions such as are supplied in Smith et al. (2005) do such questions become answerable. The idea behind DOI's '*modification*' is, we believe, captured formally in BFO's '*transformation_of*':

> "the universal A is a transformation of the universal B if and only if every instance of A is at some earlier time an instance of B and there is no time at which it is an instance of both A and B."

The way to go forward here is by concentrating on genuinely ontological relations, i.e., relations that obtain between entities in reality, independently of our ways of gaining knowledge about such entities and independently of our ways of representing or processing such knowledge in computers. It requires also that we distinguish three major families of relations between entities:

- from particular to particular (for example: the *Mona Lisa created_by* Michelangelo)
- from particular to universal (for example: Michelangelo *instance_of* the universal person; the *Mona Lisa instance_of* the universal painting)
- from universal to universal (for example: painting *is_a (meaning: is a subkind of)* work of art).

Against this background we would then be able to use the formal machinery provided by referent tracking to set up a DRM system which not only has a much clearer semantics but which is also such that we can use reality itself as a benchmark of its correctness, by following the rules set forth in Ceusters and Smith (2006b) for defining measures of the quality of successive versions of the system and of the accuracy of mappings to other systems (Ceusters, 2006).

As an example, that Michelangelo is the creator of the *Mona Lisa* would under the referent tracking paradigm be represented by a series of formal statements in which we can distinguish a number of structural elements:

- an authorised user observes one or more objects which have already been assigned IUIs in the RTS in hand
- the user apprehends that these objects stand in a certain relation, which is represented in some ontology *o*
- the user asserts that this relation obtains and publishes this assertion by entering corresponding data into the DRM system.

This relationship (R-) data will then take the form of ordered sextuples:

$$R_i = < IUI_a, t_a, r, o, P, t_r >$$

to be interpreted according to the key:

IUI_a: IUI of the author asserting that the relationship referred to by r holds between the particulars referred to by the IUIs listed in P

t_a: time-stamp indicating when the assertion was made

r: designation in o of the relationship obtaining between the particulars referred to in P

o: ID of the ontology from which r is taken

P: an ordered list of IUIs referring to the particulars between which r obtains and containing as many IUIs as are required by the arity of the relation r

t_r: a time-stamp representing the time at which the relationship was observed to obtain.

By means of these and associated types of expressions, referent tracking allows the various entities that are to be represented in a DRM system to be described at the level of granularity most appropriate to the tasks which such a system is designed to fulfil. This is because it provides a framework for such representations which remains both faithful to the salient portions of reality *and* to the views of those who are charged with the task of managing or reasoning about this reality. The explicit representation of the author of assertions (IUI_a) and of the times assertions are made (t_a) meets precisely the requirements put forward by the Indecs framework to keep track of conflicting views and to be able to re-assess the veracity of assertions made.

5 Conclusion

DOI is establishing itself as an important asset in the world of DRM. The orientation of the underlying Indecs Framework towards particular entities in the real world, entities which are able to preserve their identity over time, rather than towards generic or conceptual entities, exhibits a clear understanding of what is at stake. Yet the framework lacks any clear ontological underpinning of this orientation. We argue that, by subjecting Indecs to a deep ontological analysis based on philosophical realism, and by adjusting its data dictionary accordingly, we can make the system more useful, more robust, more easily understandable, and, as a by-product of all of this, such as to satisfy more adequately the requirements of the Semantic Web.

Acknowledgement

This work was supported by the Wolfgang Paul Program of the Humboldt Foundation, and the Volkswagen Foundation, and by the National Center for Biomedical Ontology under roadmap-initiative grant U54 HG004028 from the National Institutes of Health.

References

Ceusters, W. and Smith, B. (2005) 'Tracking referents in electronic health records', in Engelbrecht, R. *et al*. (Eds.): *Medical Informatics Europe*, IOS Press, Amsterdam, pp.71–76.

Ceusters, W. and Smith, B. (2006a) 'Strategies for referent tracking in electronic health records', *J. Biomed. Inform*., Vol. 39, No. 3, pp.362–378.

Ceusters, W. and Smith, B. (2006b) 'A realism-based approach to the evolution of biomedical ontologies', Forthcoming in *Proceedings of the Annual AMIA Symposium*, Washington DC, November 11–15, pp.121–125.

Ceusters, W. (2006) 'Towards a realism-based metric for quality assurance in ontology matching', in Bennet, B. and Fellbaum, C. (Eds.): Formal Ontology in Information Systems, IOS Press, Amsterdam, pp.321–332; *Proceedings of FOIS-2006*, November 9–11, Baltimore, Maryland.

Ceusters, W., Smith, B., Kumar, A. and Dhaen, C. (2004) 'Mistakes in medical ontologies: where do they come from and how can they be detected?', in Pisanelli, D.M. (Ed.): *Ontologies in Medicine. Proceedings of the Workshop on Medical Ontologies*, Rome October 2003, IOS Press, Studies in Health Technology and Informatics, Vol. 102, pp.145–164.

Cunard, J.P., Hill, K. and Barlas, C. (2003) 'Current developments in the field of digital rights management', *World Intellectual Property Organisation*, SCCR 10/2, August 1, Geneva.

Donnelly, M., Bittner, T. and Rosse, C. (2006) 'A formal theory for spatial representation and reasoning in biomedical ontologies', *Artif. Intell. Med*., Vol. 36, No. 1, pp.1–27.

García, R., Gil, R. and Delgado, J. (2004) 'Intellectual property rights management using a semantic web information system', *Lecture Notes in Computer Science*, Springer-Verlag, Berlin, Vol. 3291, pp.689–704.

Grenon, P. and Smith, B. (2004) 'SNAP and SPAN: towards dynamic spatial ontology', *Spatial Cognition and Computation*, Vol. 4, No. 1, pp.69–103.

Hartung, F. and Kutter, M. (1999) 'Multimedia watermarking techniques', *Proc. IEEE*, Vol. 87, pp.1079–1107.

Rosse, C. and Mejino Jr., J.L.V. (2003) 'A reference ontology for bioinformatics: the foundational model of anatomy', *J. Biomed. Informatics*, Vol. 36, pp.478–500.

Rust, G. and Bide, M. (2000) *The <Indecs> Metadata Framework: Principles, Model and Data Dictionary*, WP1a-006-2.0, June, http://www.indecs.org/pdf/framework.pdf. Last visited: July 8th, 2005.

Smith, B. (1984) 'Acta cum fundamentis in re', *Dialectica*, Vol. 38, pp.157–178.

Smith, B. (2004) 'Beyond concepts, or: ontology as reality representation', in Achille, V. and Vieu, L. (Eds.): *Formal Ontology and Information Systems. Proceedings of the Third International Conference (FOIS 2004)*, IOS Press, Amsterdam, pp.73–84.

Smith, B. and Ceusters, W. (2006) 'Ontology as the core discipline of biomedical informatics: legacies of the past and recommendations for the future direction of research', Forthcoming in Gordana, D.C. and Susan, S. (Eds.): *Computing, Philosophy, and Cognitive Science*, Cambridge Scholars Press, Cambridge.

Smith, B., Ceusters, W., Klagges, B., Köhler, J., Kumar, A., Lomax, J., Mungall, C., Neuhaus, F., Rector, A. and Rosse, C. (2005a) 'Relations in biomedical ontologies', *Genome Biology*, Vol. 6, No. 5, R46.

Smith, B., Ceusters, W. and Temmerman, R. (2005b) 'Wüsteria', in Engelbrecht, R. *et al*. (Eds.): *Medical Informatics Europe*, IOS Press, Amsterdam, pp.647–652.

Smith, B., Kusnierczyk, W., Schober, D. and Ceusters, W. (2006) 'Towards a reference terminology for ontology research and development in the biomedical domain', in Bodenreider, O. (Ed.): *Proceedings of KR-MED 2006*, November 8, Baltimore, MD, pp.57–66.

The DOI Handbook (2005) Version 4.2.0, released February 2005, http://www.doi.org/hb.html, Last visited: July 11.

Wegner, S. (2002) *Overview of State-of-the Art DRM Systems and Standardisation Activities*, Technical Report, http://www.eurescom.de/public/projectresults/P1200-series/P1207-TI.asp. Last visited: July 10th, 2005.

Websites

DOI Fact sheets: http://www.doi.org/factsheets/DOIDataDictionaries.html

Indecs: Indecs Framework Ltd, http://www.indecs.org/.

The DOI system, developed by the International DOI Foundation (IDF), www.doi.org.

W3C Uniform Resource Identifier paradigm, http://www.gbiv.com/protocols/uri/rfc/rfc3986.html.

XML-DRM: XML and Digital Rights Management, http://xml.coverpages.org/drm.html.