Referent Tracking for Digital Rights Management

Werner Ceusters[1], Barry Smith[1,2]

[1] ECOR, Universität des Saarlandes, Postfach 151150, D-66041 Saarbrücken, Germany.
Werner.Ceusters@ecor.uni-saarland.de
[2] Department of Philosophy, University at Buffalo, 135 Park Hall, Buffalo, NY 14260
phismith@buffalo.edu

Abstract. Digital Rights Management covers the description, identification, trading, protection, monitoring and tracking of all forms of rights over both tangible and intangible assets, including management of relationships between rights holders in a digital environment. The Digital Object Identifier (DOI) system provides a framework for the persistent identification of content in its broadest interpretation. Although the system has been very well designed to manage object identifiers, some important questions related to the 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 DOI models and by showing how they can be corrected using an RTS.

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 modifications thereof [1]. In order to make DRM systems semantically interoperable, several initiatives for standardization have been proposed [2]. Initially, work was focused on a syntactic approach such as the formalisation of XML Document Type Definitions and Schemas that define Rights Expression Languages. Recently, ways are being explored to make DRM systems suitable for operating under the Semantic Web paradigm [3].

An essential element in DRM is the unique identification of certain key entities: persons and organizations, the property or other rights they enjoy, and the assets to be protected [4]. In 1998, the International DOI Foundation was created to support the development and promotion of the DOI system [5] resting 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), so that it can be expressed in a form suitable for use in an automated system [6]. The DOI system is a specific implementation of the Uniform Resource Identifier paradigm advanced by W3C [7] 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, 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 [5]. *Creations* exist in tangible (pictures, paintings) and intangible forms (physical work-out methodologies, games). The totality of creations is moreover to be interpreted as including not only *manifestations* (such as books) but also *expressions* such as performances.

In order to bring some form of organization 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 [8]. Unfortunately, the work of the latter is marked by a lack of formal rigour (as explained further) and this gives rise to several flaws in its definitions, leaving many questions unanswered. As an example, the definition of '*creator*' (and further of '*party*' and '*agent*') does not allow my laserprinter to be excluded as an instance. Another problem is the level of detail that must be taken into account when entities are identified. 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*' [4]. Unfortunately, neither Indecs nor DOI provide clear answers to questions such as in what way '*creations'* are different from '*expressions*' or '*manifestations*' (this is an ontological issue) or how they are to be differentiated from each other (which is 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*' [6, par 1.6.4.]. It is here, we believe, that the referent tracking approach can contribute valuable additional clarity – in ways which will have practical consequences not just for 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.

2   Referent tracking

In [9], referent tracking was introduced to refer explicitly to all of the concrete individual entities relevant to the accurate description of each patient's condition, as well as to therapies and outcomes, through the assignment to each of a unique identifier, called a IUI, for '*Instance Unique Identifier*'. IUIs refer to the real entities themselves 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 [10]. Referent tracking distinguishes between *IUI assignment* (possible 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.

IUIs are to be used in a *referent tracking system* (RTS), which implements the following requirements [11]:

- 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 only one IUI);
- rules governing the use of IUIs in other systems such as electronic healthcare record systems, digital library systems, … (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 DOI and RTS: realist ontology makes the difference

Although the DOI and RTS paradigms were developed independently, they share a number of features in common. 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*, *tumors*) 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 a deep ontological analysis that is grounded in realism.

3.1 DOI: the world through '*models*'

Although the Indecs developers did a much better job in its own domain than did most of the terminology and model builders in the domain of healthcare informatics [12], their work, because it is based on ISO 11179 [4, p. 11], exhibits the confusions typical of what Smith *et al*. have called '*Wüsteria*' – the main feature of which is that there is no independent benchmark in relation to which the concept-system developed could be established as correct [13].

Several statements in the Indecs Framework description exhibit the characteristic confusion between entities in reality and data relating thereto:

- '*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*' [4, p. 11];
- '*a thing* [i.e. an entity] *must be both thought about or perceived and identified before it exists in a metadata framework*' [4, p. 12];
- '*all metadata relationships are either events in themselves, or rely on events to establish them*' [4, p. 13].

This last statement is, in addition, difficult to line up with the thesis that: '*An item of metadata is a relationship that someone claims to exist between two entities. It raises the question of authority: the identification of the person making the claim is as significant as the identification of any other entity*' [4, p. 11]. And this raises an even more basic question, namely: what *are* the relationships that obtain amongst the entities themselves, independently of those relationships that are claimed to obtain.

Table 1. Top-level ontology of the Indecs Framework. Adapted from [4, p13]

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 find also a confusion as to whether terms refer to the things described or to data about those things: '*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)*' [4, p. 4]. The use of 'they' here (in '*they are described*') thus identifies people, stuff and deals with *data about* people, stuff and deals.

Although we accept the justifiable need to represent in a representation system only what is relevant to its intended purpose, the Indecs paradigm does not seem to recognize that the nature of 'stuff' 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: '*The basic "elements" of a resource [i.e. "stuff" as used above] may be entirely different according to your purpose. Stuff may be analysed, for example, in terms of molecular entities (chemistry), particles such as electrons, quarks or superstrings (physics), spatial coordinates (geography), biological functions (biology, medicine), genres of expression (creations), price categories (commerce), and so on*' [4, p. 10]. Even more obviously confused is Indecs' '*Fifth axiom*': '*everything is a view*' [4, p. 12]).

The model-based approach adhered to by Indecs [4, p. 4] is also responsible for the dubious structure of what could be seen as its top-level ontology (Table 1). 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*' [4, p. 12]. On a sensible realist view, in contrast, perception itself is recognized as a process which relates entities perceived to a perceiving entity, and the latter always pre-exists the process of perception itself.

3.2. RTS: the world through ontology

Referent tracking is based on Basic Formal Ontology (BFO), a theory proposed in the recent literature of ontological realism [14]. 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 or videotapes which endure self-identically through time), and occurrents (performances or perceptions which can be divided along the temporal axis into their 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 themselves in a given domain across a given temporal interval. Continuants and occurrents themselves then exist only in mutual dependence on each other. 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). In contrast to Indecs, however, BFO 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 particulars and universals. Where DOI refers only to particulars ('*something which is identified*') such as *this paper by these authors*, RTS takes account also of universals, i.e. generic entities such as *paper* and *author* that 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 '*So 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*' [4, p. 25] (Note how the name '*the London Philharmonic Orchestra*' is here used differently as compared to the other names, revealing 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*' [6, par. 6.3]. In the haphazard way the data dictionary is currently built, it would indeed be a very tricky endeavor 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 realized in systems such as the OBO Ontologies [15] and the Foundational Model of Anatomy [16]. This effort would consist in 1) building a coherent ontology of the various types of entities referred to in the Indecs data dictionary, 2) 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

A first 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. DOI's '*thing*' and '*being*' seem to comprehend entities of the former type, at least on the basis of the definitions provided (see Table 1). Some caution is however required for '*thing*' because of its asserted disjointness from '*being*': entities such as weights, temperatures and colours are also inanimate, but they depend on their bearers. 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, despite the fact that it is claimed to have '*been validated against the W3C ontology language OWL-DL*' [17]. As has been shown in the domain of biomedicine, validation against a description logic is by no means a sufficient guarantee against mistakes, not even serious ones [18]. Additional caution is required because of the subtype relationships from '*thing*' and '*being*' to '*percept*' explicitly asserted to obtain. Since, presumably, percepts exist only if there are perceiving subjects, this would seem to employ that both things and beings are dependent after all. We take it, however, that this is a mistake in DOI (following a parallel mistake in the ISO standards). All other entities in the DOI upper

ontology are to be categorized as dependent entities: a '*concept*' (specifically under its reading as '*idea*') depends on a cognitive agent. '*Relations*', '*events*' and '*situations*' clearly depend on those other entities which serve as their relata or participants. As argued already earlier, a realist ontology does not accept a class of entities called '*percept*'.

A second distinction is that between continuants and occurrents. DOI's '*thing*' and '*being*' refer 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 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 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. Events fall under BFO's 'processual entities', and situations under 'settings'.

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*' [4, p. 21]. The '*percept*' issue mentioned before is a typical example. Specifically problematic for the DRM domain are the vague specifications provided of two foundationally distinct relations: transformation and derivation. 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*' [4, p. 22], and '*modification*' as: '*a creation made by changing a pre-existing creation of the same type (aka version)*' [4, 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 [15] 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*'.

5 Conclusion

DOI is establishing itself as an important asset in the world of DRM. The orientation of the underlying Indecs Framework towards particular, identity-bearing entities in the real world, rather than to 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 fit better the requirements of the Semantic Web.

References

1. Cunard JP, Hill K, Barlas C. Current developments in the field of digital rights management. World Intellectual Property Organisation, SCCR 10/2, Geneva, August 1, 2003.
2. Wegner S (ed). Overview of state-of-the art DRM systems and standardisation activities. Technical Report, September 2002. http://www.eurescom.de/public/projectresults/P1200-series/P1207-TI.asp. Last visited: July 10th, 2005
3. García, R., Gil, R. and Delgado, J. Intellectual Property Rights Management using a Semantic Web Information System. Lecture Notes in Computer Science, Vol. 3291, pp 689 - 704. Springer-Verlarg, 2004.
4. Rust G and Bide M. The <Indecs> metadata framework: principles, model and data dictionary. WP1a-006-2.0. June 2000. http://www.indecs.org/pdf/framework.pdf. Last visited: July 8th, 2005.
5. http://www.doi.org. Last visited: July 11th, 2005.
6. The DOI Handbook (Version 4.2.0, released February 2005). http://www.doi.org/hb.html. Last visited: July 11th, 2005.
7. http://www.gbiv.com/protocols/uri/rfc/rfc3986.html. Last visited: July 13th, 2005.
8. http://www.indecs.org/. Last visited: July 12th, 2005.
9. Ceusters W, Smith B. Tracking Referents in Electronic Healthcare Records. Accepted for MIE 2005.
10. Smith B. Beyond Concepts, or: Ontology as Reality Representation. In: Achille Varzi and Laure Vieu (eds.), Formal Ontology and Information Systems. Proceedings of the Third International Conference (FOIS 2004), Amsterdam: IOS Press, 2004, 73.84
11. Ceusters W, Smith B. Strategies for Referent Tracking in Electronic Health Records. Proceedings of IMIA WG6 Conference on "Ontology and Biomedical Informatics". Rome, Italy, 29 April - 2 May 2005. (in press).
12. Smith B, Ceusters W. Ontology as the Core Discipline of Biomedical Informatics: Legacies of the Past and Recommendations for the Future Direction of Research. Forthcoming in Proceedings of Forum Engelberg, Lucerne, 2005.
13. Smith B., Ceusters W, Temmerman R. Wüsteria. Accepted for MIE 2005, Geneva, 28-31 Augustus 2005.
14. Grenon P and Smith B 2004 "SNAP and SPAN: Towards Dynamic Spatial Ontology", Spatial Cognition and Computation, 4: 1, 69–103.
15. Smith B, Ceusters W, Klagges B, Köhler J, Kumar A, Lomax J, Mungall C, Neuhaus F, Rector A, Rosse C. 2005 "Relations in Biomedical Ontologies", Genome Biology, 2005, 6 (5), R46.
16. Rosse, C. and Mejino, J. L. V. Jr. 2003 "A Reference Ontology for Bioinformatics: The Foundational Model of Anatomy", J Biomed Informatics, 36:478–500.
17. http://www.doi.org/factsheets/DOIDataDictionaries.html. Last visited: July 13th, 2005.
18. Ceusters W, Smith B, Kumar A, Dhaen C. Mistakes in Medical Ontologies: Where Do They Come From and How Can They Be Detected? in Pisanelli DM (ed) "Ontologies in Medicine. Proceedings of the Workshop on Medical Ontologies, Rome October 2003" IOS Press, Studies in Health Technology and Informatics, vol 102, 2004.