

Tracking Referents in Electronic Health Records

Werner Ceusters^a, Barry Smith^b

^a European Centre for Ontological Research, Saarbrücken, Germany

^b Institute for Formal Ontology and Medical Information Science, Saarbrücken, Germany and
Department of Philosophy, University at Buffalo, NY, USA

Abstract

Electronic Health Records (EHRs) are organized around two kinds of statements: those reporting observations made, and those reporting acts performed. In neither case does the record involve any direct reference to what such statements are actually about. They record not: what is happening on the side of the patient, but rather: what is said about what is happening. While the need for a unique patient identifier is generally recognized, we argue that we should now move to an EHR regime in which all clinically salient particulars – from the concrete disorder on the side of the patient and the body parts in which it occurs to the concrete treatments given – should be uniquely identified. This will allow us to achieve interoperability among different systems of records at the level where it really matters: in regard to what is happening in the real world. It will also allow us to keep track of particular disorders and of the effects of particular treatments in a precise and unambiguous way.

Keywords

Realist ontology; unique identifier; electronic health record; universals and particulars

1 Introduction

Rector *et al.* have claimed that the information in the medical record consists of a collection of statements “*not about what was true of the patient but [about] what was observed and believed by clinicians*” [1]. They distinguish statements about direct observations concerning the patient (i.e. of what was heard, seen, thought, and done), and statements concerning the decision-making process and the clinical dialogue. In this way the authors seek to define the requirements to be satisfied by the EHR in order that it satisfies the criterion of “*faithfulness to the clinical history and care of the patient*”. The requirements they list, however, refer exclusively to facilities for managing the statements inside EHR systems. Thus for instance they require that: “*the record should be capable of representing multiple instances of count nouns*”. Count nouns are for example “fracture”, “tumor”, “leg” – nouns that can be pluralized (as contrasted with mass nouns like “urine”). The requirement is, therefore, as put forward by Rector *et al.*, that the record should be capable of multiple representations not of fractures, tumors, legs themselves, but of the corresponding *nouns*, or rather (though we are not sure what meaning to assign to this phrase) of *instances* of such nouns. Indeed the insistence by Rector *et al.* that the record be a record of what was *said* rather than of what actually happened, positively rules out that it should contain representations of fractures, tumors or legs – or more generally of all those disorders, activities, symptoms, etc. which, as we believe, are of primary relevance to the health record. But this very fact, we believe, deprives the other requirements listed by

Rector *et al.* – such as the requirement that the record *allow conflicting statements* – of the possibility of coherent application to real cases, since there is no way in which the entities about which there is supposed to be a disagreement could be explicitly referred to.

Slightly more acceptable, in this respect, is the account proposed by Huff *et al.* [2], who take “*the real world to consist of objects (or entities)*”. They continue by asserting: “*Objects interact with other objects and can be associated with other objects by relationships ... When two or more objects interact in the real world, an ‘event’ is said to have occurred.*” They then, encouragingly, base their account upon the events themselves, rather than upon statements about events. Each event receives an explicit identifier, called an *event instance ID*, which is used to link it to other events (reflecting the goal of supporting temporal reasoning with patient data). This ID serves as an anchor for describing the event via a frame-representation, where the slots in the frame are name-value tuples such as *event-ID* = “#223”, *event-family* = “diagnostic procedures”, *procedure-type* = “chest X-ray”, etc. The framework of [2] incorporates also explicit reference via other unique IDs to the patient, the physician and even to the radiographic film used in an X-ray image analysis event. Unfortunately, they fail to see the importance of explicitly referring also to *what was observed*. This is in spite of the fact that the very X-ray report that they analyse begins with the sentences: “*PA view is compared to the previous examination dated 10-22-91. Surgical clips are again seen along the right mediastinum and right hilar region.*” Because they have no means to refer directly to those clips, they must resort to a complex representation with nested and linked event frames in order to simulate such reference. Even then, however, they are not able to disambiguate as between an interpretation in which (i) the surgical clips seen in the mediastinum are different from those seen in the hilar region and (ii) there is only one set of clips that extends from the mediastinum into the hilar region. The limitations of the event-based representation force them also to create a different event-frame for each location of the clips (though neither clips nor locations are themselves events). That this approach is questionable is seen in the fact that, while it is certainly possible, when looking at a chest X-ray, to see first a clip in the mediastinum (first event), and then by looking again (in a second event) to see a second clip, it is equally possible that just one glance suffices for the observer to apprehend all the clips at the same time, i.e. in one event. (To rule out this possibility would be tantamount to claiming that complex perceptions, e.g. of a tree, must involve the subject participating simultaneously in several thousands of distinct events of observation.)

Finally, we can mention Weed’s *Problem Oriented Medical Record*, the central idea of which is to organize all medical data around a problem list, thereby assigning each individual problem a unique ID [3]. Unfortunately Weed proposes to uniquely identify only problems, and not the various particulars that cause the problems, are symptomatic for them, or are involved in their diagnosis or therapy. This is exemplified in Barrows and Johnson’s model that suffers also from the ambiguity whether the unique IDs refer to the problems themselves or rather to statements about the problems [4].

2 Some ontological and epistemological aspects of introducing unique identifiers for particular entities in health records

We argue that the EHR should contain explicit reference, via unique identifiers, to the individual real world entities which are of relevance in such records – called “particulars” in what follows. These should be similar not only to the unique identifiers we already use to distinguish individual patients, individual physicians, individual healthcare organizations, individual invoices and credit card transactions, and even individual drug packages [5]. When I enter a hospital with a fracture of my left first metatarsal base, then I would like this

particular fracture, which prevents me from dancing and causes me a lot of pain, and of which the clinician examining me can see an image on an X-ray, to receive a unique ID for further reference. Only in this way can this fracture be distinguished not just from a second fracture in the same metatarsal, but also from the fracture in the very same place from which I suffered two years earlier. The bunion, in contrast, which the clinician observes when examining my foot, and of which there is also an image on the X-ray, should not receive at that time its own unique ID. Reference to it should be effected, rather, by means of the ID it received already two years ago, when it was diagnosed at the same time as the first fracture. True, the bunion is now much bigger; but it is still the same entity as it was when first observed.

There are good reasons for the above. Coding systems such as ICD have labels such as “multiple fracture occurring in the same limb”, “multiple fracture occurring in different limbs”, and so forth. Statistics concerning the incidence of disorders would be erroneous if, because of unconstrained use of such labels, multiple observations of the same entity came to be counted as separate incidences of disorder. Few patients will care about such statistics, but they will care if having two fractures would make them eligible for cost reimbursement because of the different ways each fracture occurred. We would even like to go further: it is not just the particular fracture which should get an explicit ID, but also the particular bone in which it occurred. For it might happen that this bone is later transplanted into another patient, and becomes the cause of a malignancy in the recipient because tumor material in the bone was transplanted together with it [6].

What does it mean to assign identifiers to particulars in an EHR? First, we note that ontologies and terminologies have focused hitherto on what is general in reality, allowing particularization to occur almost exclusively where there is explicit reference to the *human beings* who are the bearers of named general attributes or to the *times* at which observations occurred or statements were made. The realm of particulars relevant to the health record is however vastly broader than this. Indeed, there are both particulars (tokens, instances), such as *this bone of that person*, and universals (types, classes), such as *bone* and *person*, throughout the entire domain of medical science. Thus there are the two particular fractures: the one from which I am now suffering, and one which occurred two years ago. But there is also the universal *basal fracture of left first metatarsal*, which is of course just as real as the particulars in which it inheres. There is the particular pain in my foot that I had two years ago, pain that was caused by that first fracture. And there is the pain from which I am suffering now, in the very same place. The pains may feel the same, but they are distinct entities nonetheless, though both are instances of the same universal. Universals and particulars exist in reality independently of our use of language. As such, they are distinguished from the *concepts* which are said to provide meanings for terms.

Second, assigning a unique identifier tells us that the particular exists (or has existed in the past), and that nothing else is that particular. The particular does not come into existence because we assign an identifier to it: it must have existed before it became possible for such an assignment to be made. The assignment itself is then an act comparable to the act of naming a fetus or newborn child (or ship or artwork).

Third, we would find in such an advanced EHR also statements about these assignment acts (and – to avoid the confusions one finds in the HL7 RIM [7] – these statements would be clearly distinguished from the acts they are intended to describe). We could then assign truth values to such statements, and we note that for such a statement to be true the particular referred to must exist (or have existed) and have no ID already assigned.

Fourth, the *use* of an ID in a statement does not entail that an assignment has already been made. If an X-ray is ordered, then the X-ray event does not exist, and so it cannot be *assigned* an ID. But it is perfectly possible to *reserve* such an ID in advance. This

difference opens up interesting perspectives in the medico-legal context. If the X-ray that was ordered is carried out, and the images reveal a pathology, then the physician who issued the order can use this fact as a justification of the claim that his initial judgment about the case had been accurate. The mere fact of his having issued the order may protect him from a lawsuit, even when the X-ray is not carried out. Of course, the information relevant to such analyses can be extracted also from conventional records. The point here is that the framework here advanced would allow for the automatic analysis of such cases, and possibly even for the automatic prevention of associated medical mistakes or hazards.

Fifth, the mere fact of assigning an identifier to a particular does not imply that any statement is made about what *kind* of particular is involved. Such a statement is made only when one has grasped the relevant particular as an instance of some universal (such as “bone” or “fracture” or “pain”), and this often occurs only after the assignment or reservation of an ID. Statements of the given sort are then true only if the particular exists *and* is an instance of the claimed universal.

3 Towards and implementation of referent tracking

Let us go back to the emergency room of that modern hospital that I choose to be treated in because it has installed one of the new fancy EHR systems (EHRS) that allows careful and explicit reference to all my various problems and to the different kinds of entities associated therewith. Because of my story about what happened to my foot, and because of the pain the two attending physicians were able to induce by palpating my forefoot, both agreed that there was something wrong. That “something wrong” was given the ID #234, a meaningless consecutive number assigned automatically by the EHRS (and guaranteed to be unique according to some algorithm). The system at the same time also generated two statements, recording the assignment of #234 to that particular by each of the two physicians. These statements enjoy a high degree of positive evidence, since the referent-tracking database allows automatic checking to verify the absence of prior existing disorders of which my current problem might have been a continuation. It did find referent #15 for the left first metatarsal base fracture that I suffered from two years ago, but this – as witnessed by the X-ray image #98 taken half a year after the initial diagnosis – had since ceased to exist. The physicians also had good evidence that the referent-tracking database was complete in all relevant respects, since they knew that I never sought treatment elsewhere. The physicians’ statements concerning the assignment were each time-stamped both for occasion of utterance and for point of appearance in the EHRS. Note that these time-stamps do not necessarily imply assertions about when #234 itself began to exist. Also, at this stage, no statement has been made about which universal disorder #234 is an instance of.

The physicians ordered and received three X-ray photographs taken of my foot from different angles. They both looked at the first (identified by the EHRS as #235 and stated to be an instance of the universal referred to by SNOMED-CT as “257444003: photograph”), but they saw nothing abnormal. Of course, they saw an image of my left first metatarsal bone, this image being identified as #286 (they did not bother to look for a SNOMED-CT code for such an image, knowing by experience that they would find nothing that comes close). They were at the same time aware that entity #286 is clearly different from entity #221, which is my left first metatarsal bone itself, and which they declared to be (i) an instance of the universal referred to by the SNOMED-CT concept “182121005: entire first metatarsal”, further annotated with the side-modifier “left”, and (ii) a part of #2 (me). On the second photograph (#236), both saw a thin hypodense line appearing towards the top of my left first metatarsal bone. They assigned that line in the image the label #287, and both stated it to be the image of some corresponding particular #288, thereby agreeing on the

existence of #288 but disagreeing as to what universal it was an instance of – the one seeing it as a fracture line, the other as just a normal part of the bone somewhat less dense than the surrounding bony material. They agreed, however, that #287 was not an artefact, i.e. that it did indeed correspond to something in my body. On the third photograph (#237), both saw a clear fracture line, indisputably an image of a real fracture and identical with particular #288. They thereupon asserted that #234, i.e. the “something wrong” previously identified, was in fact an instance of the universal: *left first metatarsal base fracture*.

4 EHR architecture standards and health particulars

No current EHR architecture standards are to our knowledge able to deal with the tracking of all those types of referents that are required to support an implementation such as the one described above. The record architecture described in CEN prEN 13606-1:2004 draft [8] allows only a limited number of particulars to be referred to explicitly – i.e. without resorting to any external terminology – and as shown in Table 1 many of them (as we should expect, given what we observed in 1 above) are at the meta-level rather than at the level of direct care. Indirectly, it would be possible to use the “data item” construct as it is defined in CEN prEN 13606-1:2004 to refer to particulars – something that we strongly encourage, even though it would require an additional round of standardization.

| Direct care particulars | Meta-level particulars |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1) the <u>subject of care</u> from whose EHR an extract is taken. This does not need to be the patient since an extract might contain data about an unborn child | 4) the <u>EHR provider system</u> from which a record extract is being taken |
| 2) the <u>healthcare agent</u> which participates in some interaction with the subject of care | 5) the <u>EHR</u> from which the extract is taken |
| 3) a <u>geographical location</u> for any person (i.e. subject of care or human healthcare agent) or organization referred to | 6) the different sorts of <u>components</u> involved in a record extract |
| | 7) <u>International Coding Scheme Identifier</u> (ICSI) |
| | 8) <u>External Procedure Reference</u> : rather than referring to a real procedure that is carried out, this refers to a specific document in which a procedure is generically described |
| | 9) The <u>software</u> used in the EHR extract transmission process |

Table 1: The different sorts of particulars that can be identified in CEN prEN 13606-1:2004

The HL7-RIM [9] is much worse in this respect. Particulars that are listed as “entities” can, it is true, be referred to by using the “Entity.Id” attribute: living subjects (either human or non-human), geographical places, organizations, and manufactured materials including devices and containers. Great care needs to be taken, however, since these labels are notoriously prone to inaccurate use. As an example, the class Person can be used to refer either to individuals or to groups, depending on how the attribute “Entity.Quantity” is set. As an example, a group with 60% females is to be represented as “*Person(quantity = 100) has-part Person(quantity = 60; sex = female)*”. And similarly with HL7-RIM’s notoriously problematic Act class. Perhaps (we do not know) the “Act.Id” attribute can be used to refer to concrete events. But then the Act class allows for “mood” (possible, planned, ordered, etc.) and “negation” attributes, so that a reference ID would sometimes need to be understood (in our terminology) as a *reservation* and sometimes as an *assignment*, though even this simple distinction is masked by layers of confusion. And in HL7-RIM, too, there is no room to refer to instances of body parts, of disorders, and so forth. The world of HL7, after all, is a world of Acts (Acts themselves being artefacts of the HL7 regime, rather than entities we could encounter in our everyday reality). Only a thorough reworking of the HL7-RIM, taking into account all aspects of reality, might solve these problems.

5 Conclusion

Nowadays, the medical informatics community prefers to restrict the use of ontology to the development of more or less well-structured vocabularies of general terms. Hardly at all does it care about the correct representation of particulars in reality – without which, of course, such vocabularies would be entirely superfluous. This is strange in an era in which the clinical community, i.e. the community that should be served by medical informatics, is focused so emphatically on Evidence Based Medicine. For on what should evidence be based, if not on real cases? The focus on what is general was perhaps defensible in an era of limited computer resources. Now, however, we argue that only benefits would accrue from inaugurating EHR systems which are able to refer directly and systematically to concrete instances along the lines proposed in the above. If a hospital database is able to store all the SNOMED-CT codes that apply to a particular patient, then adding an additional reference ID to the particulars that are the relevant instances of the SNOMED classes will hardly create massive storage problems. CEN prEN 13606-1:2004, too can be adjusted quite easily along these lines, though of course adapting actual EHR systems in an appropriate way would involve a more substantial effort. Considerable organizational issues would above all still need to be resolved – as witnessed by the problems encountered in establishing a Unique Patient Identifier in a safe and secure manner [10]. But the benefits, in terms of better patient management, supporting advances in biomedical science, health cost containment and more reliable epidemiological data, can be expected to be enormous. And just as pseudonymisation is an effective approach to the collection of data about the same patient without disclosing his or her identity, so also the mechanism of ID assignment to disorders, body parts, etc. would provide additional avenues for supporting anonymity and thus promoting more and better HIPAA compliant research.

6 Acknowledgments

The present paper was written under the auspices of the Wolfgang Paul Program of the Alexander von Humboldt Foundation, and the Network of Excellence in Semantic Interoperability and Data Mining in Biomedicine of the European Union.

7 References

- [1] Rector AL, Nolan WA, and Kay S. Foundations for an Electronic Medical Record. *Methods of Information in Medicine* 30: 179-86, 1991.
- [2] Huff SM, Rocha RA, Bray BE, Warner HR, and Haug PJ. An Event Model of Medical Information Representation. *J Am Med Informatics Assoc.* 1995;2:116-134.
- [3] Weed L. Medical Records That Guide And Teach. *N Engl J Med* 1968: 278: 593-600.
- [4] Barrows RC, Johnson SB. A data model that captures clinical reasoning about patient problems. In: Gardner RM (editor). Proceedings of the 19th Annual Symposium on Computer Applications in Medical Care; New Orleans, LA; October 28 - November 1, 1995: pp 402-505
- [5] Bell J. Drug firms see future in RFID technology. In: *Cincinnati Business Courier*, December 27, 2004.
- [6] Collignon FP, Holland EC, Feng S. Organ Donors with Malignant Gliomas: An Update. *Am J Transplant.* 2004 Jan;4(1):15-21.
- [7] Vizenor L. Actions in health care organizations: an ontological analysis. *Medinfo.* 2004;2004:1403-10.
- [8] CEN. Health informatics - Electronic healthcare record communication - Part 1: Extended architecture
- [9] Case J, McKenzie L, Schadow G. (eds.) HL7 Reference Information Model. (<http://www.hl7.org/Library/data-model/RIM/C30202/rim.htm>)
- [10] <http://www.hipaonet.com/upin1.htm>.

Address for correspondence: Dr. W. Ceusters, European Centre for Ontological Research, Universität des Saarlandes, Postfach 151150, D-66041 Saarbrücken, Germany. www.ecor.uni-saarland.de.