HL7 RIM: An Incoherent Standard

Barry Smith [a,b,c,1] and Werner Ceusters [c]
[a] *Department of Philosophy, University at Buffalo, Buffalo NY, USA*
[b] *Institute for Formal Ontology and Medical Information Science, Saarbrücken, Germany*
[c] *Center of Excellence in Bioinformatics and Life Sciences and National Center for Biomedical Ontology, University at Buffalo, Buffalo NY, USA*

from: *Studies in Health Technology and Informatics*, vol. 124 (2006), 133–138

Abstract. The Health Level 7 Reference Information Model (HL7 RIM) is lauded by its authors as a 'backbone' information model which can serve as 'the foundation of healthcare interoperability'. Like HL7 Version 3 as a whole, the RIM has met with enthusiastic response in many circles, purportedly rising to the level of adoption by standards bodies. Suspiciously, however, it has found few actual users, and even after some 10 years of development work it is still subject to a variety of logical and ontological flaws. As is becoming increasingly clear, these flaws place severe obstacles in the way of those who are called upon to develop RIM-compliant implementations. We here offer evidence for the thesis that these obstacles are insurmountable and that the time has come to abandon an unworkable paradigm.

Keywords: HL7, RIM, standardization, ontology, realism

1. Introduction

What follows is an initial chapter of a much needed exegesis and critique of the HL7 Reference Information Model (RIM) as described in the "HL7 Version 3 (V3) Normative Edition 2005" released by Health Level Seven, Inc. on November 22, 2005. We focus specifically on [1], the central RIM codex, in which the six back-bone classes of the RIM are set forth. In 1.1.1 of this document, the RIM itself is described, somewhat optimistically, as 'credible, clear, comprehensive, concise, and consistent'. In 1.1.3 it is asserted to be 'universally applicable' and 'extremely stable'. As is recognized in HL7's own internal forums, however, the RIM is marked by major flaws, which include at least the following (further documentation is provided at [2]):

- *problems of implementation*: the decision by HL7 to adopt the new RIM-based methodology was adopted already in 1996 (some years before the conception of Version 3); after ten years of effort, and considerable investment in the RIM itself and in the development of associated message types, DMIMs and RMIMs, the promised benefits of interoperability which were to have been engendered by its use remain elusive;

[1] Corresponding author: Barry Smith, Department of Philosophy, University at Buffalo, Buffalo NY 14260l, USA. Email: phismith@buffalo.edu; internet: http://ontology.buffalo.edu/smith.

- *problems of usability in specialist domains*: is the RIM an appropriate 'backbone' for producing new information models for each new domain (pharmacy, blood bank, lab observations, etc.)? The RIM methodology consists in defining a set of 'normative' classes (Act, Role, and so on), with which are associated a rich stock of attributes derived from the specific domain of US hospital billing practices; when the RIM is applied to each successive new domain (for example vocabulary, or clinical genomics), some of these attributes need to be deleted and replaced by others, rather as if one were attempting to create software for the manufacture of lawnmowers by deleting selected attributes from a template for manufacturing aircraft. Is there even one example where a deletion-and-replacement based methodology of this sort has been made to work successfully?
- *technical problems*: the decision to include this rich stock of attributes was made, it seems, in order to provide an environment which would reassure initial users; unfortunately, the resultant methodology breaks central rules of object-oriented software design [3], in ways which go far towards explaining why the RIM seems thus far to have led to the creation of incompatible, buggy and unworkable solutions; is V3 able to support the construction of good quality, re-usable software, capable of being integrated into modern computing infrastructures and development environments? does the RIM yield a coherent basis for constructing well-designed software artifacts for functions like Electronic Health Records and computerized decision support (and for the sorts of sophisticated tasks which will become necessary with the advance of genomics-based patient-centered informatics research)?
- *problems of antiquatedness*: is the message-based methodology of V3 appropriate to the world as it is today, with increasing migration towards web-centric service-based technologies?
- *problems of scope*: given that HL7 grew out of specific messaging needs of US healthcare institutions, many V3 terms relate not to healthcare processes as these exist in themselves, but rather to intentional actions of a type which are represented in billing documents; at the same time many healthcare processes themselves, including disease processes and drug interactions, are invisible to the RIM, because such processes are not addressed even indirectly in billing documents; to what degree is V3 capable of being applied in coherent fashion to those types of information exchange (for example in genomics research) where it is not intentional actions which play a central role, but rather processes and reactions on the side of the organism?
- *problems with the HL7 business model*: is there some conflict of interest involved in the fact that those (such as consultants) who stand to benefit from overly complex standards or unclearly formulated documentation, are also involved in the balloting process which determines what the standards and documentation should be?
- *problems with the HL7 governance process*: should decisions on modifications to HL7 standards be based on a slow and tedious balloting procedure involving a large number of distinct constituencies? are those members of the relevant constituencies working with new technologies able to influence the development of HL7 V3 in a timely fashion in an age of rapid technological change?

- *problems of documentation*: as inspection reveals, the RIM documentation is poorly integrated with those other parts of the HL7 V3 documentation for which the RIM itself is designed to serve as backbone; it is also subject to a series of internal inconsistencies and unclarities (for example in its syntactically sloppy use of terms such as 'act', 'Act', 'Acts', 'action', 'activity', 'ActClass' 'Act-instance', 'Act-object') of a sort which, one would imagine, should precisely be avoided in the context of work on messaging standards.
- *problems of learnability*: can HL7 V3 be taught, and therefore engaged with and used by a wider public, given the poor quality of its documentation, counterintutive and idiosyncratic use of singular and plural forms of terms like 'Act', 'Entity', the massive number of special cases which must be learned, and the complex hurdles that must be overcome in creating a message?
- *problems of marketing*: are the grandiose marketing claims made on behalf of V3 as 'the data standard for biomedical informatics' justifiable, given the many still unresolved problems on the technical side?

2. The RIM's Double Standards

The family of problems on which we shall focus here concerns the RIM's unsure treatment of the distinction between *information about an action* on the one hand and *this action itself* on the other. The former is what is *recorded* in a message or record. The latter is what *occurs*, for example within a hospital ward or laboratory. When challenged, RIM enthusiasts will insist that the RIM is concerned exclusively with the former – with information – and of course the very title of the RIM is in keeping with this conception. Interspersed throughout its documentation, however, we find also many references to the latter, often conveyed by means of the very same expressions. As example, consider the treatment of the RIM class LivingSubject which is defined, confusingly, as follows:

> A subtype of Entity representing an organism or complex animal, alive or not. (3.2.5)

Examples of this class are then stated to include: 'A person, dog, microorganism or a plant of any taxonomic group.' From this we infer that a person, such as you or me, is an example of a LivingSubject. At the same time in 3.2.1.13 (which is, oddly, the only subsection of 3.2.1) we are told that LivingSubjects can occupy just two 'states': *normal*, which is defined simply as 'the "typical" state', or *nullified*, which is defined as: 'The state representing the termination of an Entity instance that was created in error.'

Can it really be true, then, that LivingSubjects include persons as examples – so that we are here being invited to postulate a special kind of death-through-nullification in the case of those persons who were created in error? Or is it not much rather the case that by 'LivingSubject' the RIM means not (as is asserted at 3.2.5) 'mammals, birds, fishes, bacteria, parasites, fungi and viruses' but rather *information about* such entities? The answer, bizarrely, is that it means both of these things; for there are in fact, co-existing within its documentation, two competing conceptions of what the RIM is striving to achieve.

2.1. The Information Model Conception of the RIM

What we shall call the information model conception of the RIM is enunciated for example in 1.1 of [1]:

> The Health Level Seven (HL7) Reference Information Model (RIM) is a static model of health and health care information as viewed within the scope of HL7 standards development activities. It is the combined consensus view of information from the perspective of the HL7 working group and the HL7 international affiliates. The RIM is the ultimate source from which all HL7 version 3.0 protocol specification standards draw their information-related content.

The RIM, according to this first conception, is intended to provide a framework for the representation of the structures of and of the relationships between information, a framework that is independent of any particular technology or implementation environment. It is thus designed to support the work of database schema designers, software engineers and others by creating a single environment for messaging and other information-related purposes which can be shared by all healthcare institutions.

2.2. The Reference Ontology Conception of the RIM

What we shall call the reference ontology conception of the RIM is nowhere propounded explicitly within the RIM documentation. It can be inferred, however, from many programmatic statements describing the RIM's purpose, which is to facilitate consistent sharing and usage of data across multiple local contexts. For in striving to achieve this end of consistency (and thus to rectify problems affecting implementations of HL7 V2), the RIM cannot focus merely on healthcare messages themselves, as bodies of data. Rather it must provide a common benchmark for how such bodies of data are to be formulated by their senders and interpreted by their recipients. We can conceive of only one candidate benchmark for this purpose, namely the totality of things and processes themselves within the domain of healthcare, which is what such messages are about. That this is indeed the benchmark adopted by the RIM itself will become clear when we examine the many passages in its documentation in which definitions and examples are provided to elucidate the meanings of its terms.

2.3. Systematic Ambiguity

Rather than distinguishing the two tasks of *information model* and *reference ontology* in some coherent fashion, and addressing them in separation, the RIM seeks to tackle them both simultaneously, effectively through the device of ambiguous use of language. Expressions drawn from the vocabulary of healthcare are used, sometimes in alternative passages, both with their familiar meanings (which enable the RIM to secure a necessary relation to the corresponding activity in the healthcare domain) and with technical meanings (when the RIM is attempting to specify the relevant associated type of data). Thus for example 'stopping a medication' means both: *stopping a medication* and: *change of state in the record of a Substance Administration Act from Active to Aborted* (3.1.5).

3. 'Objects'

According to the information model conception, the RIM, and the HL7 messages defined in its terms, are about objects in information systems – hereafter called 'Objects' – which 'represent' things and processes in reality. Thus for example the particular human being named 'John Smith' is represented by an Object containing John Smith's demographic or medical data. This Object is thus different from John Smith himself. HL7's Glossary defines an 'Object' as:

> An instance of a class. A part of an information system containing a collection of related data (in the form of attributes) and procedures (methods) for operating on that data.

It defines an 'Instance' as: 'A case or an occurrence. For example, an instance of a class is an object.' Yet under the entry for the class 'Person' in the same Glossary we are told that 'Instances of Person include: John Smith, RN, Mary Jones, MD, etc.' For HL7, accordingly, 'John Smith, RN' is not the name of some human being; rather, it is the name of an Object which goes proxy for a human being in an information system.

Because Objects are bodies of data, and because different data about John Smith will be contained in the different information systems involved in messaging, 'John Smith, RN' will refer ambiguously to many John Smith Objects. How, then, is messaging about John Smith (the human being) possible, given that senders and recipients will associate distinct John Smith Objects with each given message?

The RIM can go some way towards alleviating this problem in the case of persons, by insisting that information objects include corresponding unique identifiers such as social security numbers. Indeed the HL7 Glossary *defines* a 'Person' as a 'single human being who … must also be uniquely identifiable through one or more legal documents (e.g. Driver's License, Birth Certificate, etc.).' In the ideal case, at least (which would require each institution to maintain a mapping of its own locally unique patient IDs to all the locally unique patient IDs used by the institutions it is communicating with), such identifiers could serve to bind the different Objects together in such a way that they would all become properly associated with what those of us living in the world outside the RIM are accustomed to think of as one and the same person. But what works for Persons will in almost every case not work for Objects representing things and processes of other types (for instance tumors, epidemics, adverse reactions to drugs), since unique identifiers for the latter are almost never available under present regimes for recording healthcare data [4].

In some cases it will be possible to infer that two messages refer to one and the same fracture, or pregnancy, or infection, by using temporal and other contextual information. This solution is however rendered once again problematic by the fact that the information in question is formulated within the RIM not as information about the tumors, fractures, blisters, noses, etc., themselves, but rather as information about Observations (indeed about what will standardly be a plurality of Observations associated with each given instance). Thus in the RIM we are forbidden to say: 'Swelling of tissue around tumor occurred due to accumulation of fluid', but must rather refer to 'Observation by physician of swelling of tissue around tumor' and to 'Observation by physician of accumulation of fluid', and have these Observation statements linked through some third statement involving a 'reason' code. The problem is that such statements create what is referred to by philosophers as

'referential opacity' [6], which means that, even where the same tumor or breast is referred to in a plurality of contexts by one and the same physician, an automatic reasoning system could not use this information to draw inferences from the corresponding statements.

4. 'Acts'

The term 'Act', too, refers within the framework of the RIM not to acts (intentional actions), but rather to associated information objects. Thus at 3.1.1 'Act' is defined as meaning: 'A record of something that is being done, has been done, can be done, or is intended or requested to be done.'

Confusingly, however, we are provided, in explication of this definition, with a list of examples not of *records* but rather of *intentional actions themselves* (referred to not as 'Acts' but as 'acts'):

> The kinds of acts that are common in health care are (1) a clinical observation, (2) an assessment of health condition (such as problems and diagnoses), (3) healthcare goals, (4) treatment services (such as medication, surgery, physical and psychological therapy), (5) assisting, monitoring or attending, (6) training and education services to patients and their next of kin, (7) and notary services (such as advanced directives or living will), (8) editing and maintaining documents, and many others.

At 3.1.14, similarly, the class PatientEncounter (a subclass of Act) is defined as follows:

> An interaction between a patient and care provider(s) for the purpose of providing healthcare-related service(s).

An instance of PatientEncounter, then, is not a *record* of an action, but rather this action itself – something that is made clear also by the examples provided to elucidate this definition, which include: emergency room visit, field visit (e.g., traffic accident), office visit, occupational therapy, telephone call.

Procedure the RIM defines as 'An Act whose immediate and primary outcome (post-condition) is the alteration of the physical condition of the subject,' providing examples such as: chiropractic treatment, balneotherapy, acupuncture, shiatsu, as well as straightening rivers, draining swamps, and building dams.

Observation it defines as:

> An Act of recognizing and noting information about the subject, and whose immediate and primary outcome (post-condition) is new data about a subject. Observations often involve measurement or other elaborate methods of investigation, but may also be simply assertive statements. (3.1.13)

From this we infer that Observations are processes of recognizing, noting, which have outcomes in the form of new data and can involve processes of measurement; again, therefore, we are here dealing not with *records* of actions, but rather with these actions themselves as they take place in reality, and this view is supported also by the RIM's assertion, still in 3.1.13, to the effect that Observations 'are professional acts … and as such are intentional actions', as also by its treatment of other subclasses of Act, such as DeviceTask, SubstanceAdministration, and Supply – all of which are similarly interpretable only against the background of a reference ontology conception of the RIM and thus as standing in conflict with the RIM's own definition of 'Act'. When, in contrast, we turn to other subclasses of Act, for example Account, or FinancialTransaction, then we find that the corresponding definitions revert to the information model conception, while yet others

(for example PublicHealthCase) are so incoherently defined as to make it difficult to understand how the corresponding passages are to be interpreted at all.

5. The Influence of Rector, Nolan and Kay

We can gain some clearer idea of the nature of the confusions underlying the RIM's treatment of 'Act' if we look at the analogous action-centered approach to healthcare information advanced by Rector, Nolan and Kay [5]. Just as the Act is conceived by the RIM, at least in some passages, not as an intentional action on the side of healthcare professionals but rather as a body of information pertaining thereto, so, according to the information model propounded in [5], the Electronic Health Record (EHR) should be conceived not as a record of disorders, symptoms, lesions, and so forth on the side of the patient, but rather as a record of clinicians' observations and associated decision-making processes and dialogue. It is not things and processes in the world that are recorded in the EHR, according to Rector, Nolan and Kay, but rather *what is said or thought about such entities*. The record must satisfy the requirement of 'Faithfulness to the clinician's observations of the patient,' i.e. 'to the direct observations of what was heard, seen, thought and done concerning the patient', not the requirement of faithfulness to what it is on the side of the patient which the clinician observes.

The distinction here is subtle, but important nonetheless, as we can see if we contrast the Rector proposal with what obtains in the normal context of scientific research, where records (for example of laboratory experiments) are standardly formulated in what is called the object-language. That is, the statements in such records are about the things subjected to measurement, the values obtained from such measurement, the changes in and the interactions between the items measured. Only in very special circumstances would scientific laboratory reports utilize meta-language statements in the style of: 'observer *A* described his experiences as being those of someone who was witnessing a change of temperature' and the like.

Certainly there is no objection to the inclusion of some meta-level observation-descriptions in an EHR, for example in order to account for observations marked by a high level of uncertainty or for negative observations ('no blisters observed'). For Rector *et al.*, however, this use of the meta-language would be made compulsory for *all* statements in the EHR. They provide three arguments for this proposal, turning (1) on the special medico-legal conditions prevailing in the healthcare domain, where what exists in reality is often for forensic reasons less important than what is claimed about what exists; (2) on the fact that observers may be in error, and there is an obvious concern to do all that we can to prevent false assertions from being entered into the record, and (3), on the fact that observers may disagree. The proposed move, which amounts to what philosophers call 'semantic ascent' [6,7], transforms an EHR in its entirety from an object-language to a meta-language document, thereby allowing the logically problematic effects of such disagreement to be neutralized by relativizing the corresponding contents to the authors of the respective observations. ('A said *p* and B said *not-p*' is not logically contradictory in the way in which '*p* and *not-p*' is.)

Unfortunately, the move in question comes at a price (bringing the same inference-blocking problems of referential opacity we referred to above). Moreover, none of the mentioned justifications is compelling. First, the forensic purposes of an audit trail can equally well be served by an object-language record if we ensure that meta-data are associated with each entry identifying by whom the pertinent data were entered, at what time, from what source, and so forth, together with facilities to quarantine erroneous entries as they are corrected. Second, the possibility of error confronts every empirical endeavor. Yet in the domain of scientific research we do not embargo the making of object-language assertions because there might be, among the totality of such assertions, some which are erroneous. Rather (and with good reason), we rely on the normal workings of science as a collective endeavor to weed out error over time, in an asymptotic process. And finally, the logical problems which arise when incompatible statements are included in the record can be solved in one or another of a variety of now familiar ways, which do not require the move of semantic ascent, for instance by chunking the relevant set of statements into maximally consistent subsets [8]. All of which means that there is nothing standing in the way of treating health records in just the way we treat empirical records in other scientific domains, which is to say as consisting, primarily, of object-level descriptions of real-world phenomena [9].

On the other side, moreover, even if we do move to meta-level assertions in the way advocated by Rector and, in its own way, by the RIM, this does not in fact solve the problems of error and logical conflict. For of course these problems arise not only when human beings are describing, on the object-level, fractures, or pulse rates, or symptoms of coughing or swelling, but also when they are describing, on the meta-level, what clinicians have heard, seen, thought and done. These matters, too, are subject to error, fraud, and disagreement in interpretation.

6. Speech Acts

An Act-instance, we are told at 3.1.1, 'represents a "statement" according to Rector and Nowlan'. We are also told, with characteristically confusing HL7 phraseology, that 'the Act class is this attributable statement'.

We can interpret this strange language as follows: each Act-instance is a statement describing what some clinician on some occasion has heard, seen, thought, or done. The Act class is the class of such coded, attributable statements. An Act-instance is, more precisely (in 3.1.1 at least), either an attributable statement or some similar attributable use of language such as an order or request, which serves as the coded accompaniment – the record – of what takes place in reality, for example when a surgical procedure is performed. There then follows a most peculiar passage, which has been criticized already in the literature on HL7 [10]:

> Act as statements or speech-acts are the only representation of real world facts or processes in the HL7 RIM. The truth about the real world is constructed through a combination (and arbitration) of such attributed statements only, and there is no class in the RIM whose objects represent "objective state of affairs" or "real processes" independent from attributed statements. As such, *there is no distinction between an activity and its documentation.* (3.1.1, emphasis added)

This passage captures what we might think of as the naked essence of the confusion at the heart of the RIM between reference ontology and information model.

We imagine (hope) that most of those involved in the authoring, dissemination, application and marketing of HL7 do in fact draw the distinction between an activity and its documentation in their everyday lives and in their clinical and coding practice. They are also (we hope) unaware of the fact that the RIM documentation contains a denial of this self-evident distinction.

The passage just-quoted gives us a glimpse of the solution to this puzzle that is advocated by at least one group of RIM authors: the given process, and indeed all other "objective states of affairs" and "real processes" involving Jane Smith, do not, in fact, exist. In fact, for the RIM, *there is no objective state of affairs or real process* which could make it true that, say, given oophorectomy records do indeed refer to some surgical process in the world. Rather, there is just documentation (records, data, attributable statements) in different information systems. Where parts of this documentation are in conflict, there occurs, presumably under the guiding hand of HL7, a process of arbitration through which, not only the truth about the real world, but indeed this real world itself – in a manner reminiscent of the philosophy of Kant – are constructed *de novo*.

7. Can the RIM be Saved?

The RIM, as we have seen, is constructed on the basis of a systematically ambiguous use of terms to refer both to information objects and to corresponding real-world things and processes. What is needed, however, if HL7 is to be in a position to satisfy its needs for consistent information representation, are two separate, though of course related, artifacts, which might be called 'Reference Ontology of the Healthcare Domain' and 'Model of Healthcare Information', respectively. The former would include those categories, such as *thing*, *process*, *person*, *anatomical structure*, *disease*, *infection*, *molecule*, *procedure*, etc., needed to provide a compact and coherent high-level framework in terms of which the lower-level types captured in vocabularies like SNOMED CT could be coherently organized [11].The latter would include those categories, such as *message*, *document*, *record*, *observation*, etc., needed to specify how information about the entities that instantiate the mentioned types can be combined into meaningful units and used for further processing. The CDA could then be related in an appropriate way to this Model of Healthcare Information, thereby avoiding the current counterintuitive stopgap, which forces a document to be an Act. And HL7's Clinical Genomics Standard Specifications could be related in an appropriate way to the Reference Ontology of the Healthcare Domain, thereby avoiding the no less counterintuitive stopgap which identifies an individual allele as a special kind of Observation.

This proposal would introduce at least a kernel of coherence to the HL7 endeavor. Disastrously, however, current moves within the HL7 community are pointing in exactly the opposite direction, which is to subject existing attempts to create coherent ontology-like treatments of the healthcare domain to the dead hand of the irredeemably confused RIM methodology [12, for example at 2.2.3.2].

Acknowledgements: This paper was written under the auspices of the Wolfgang Paul Program of the Alexander von Humboldt Foundation, the European Union Network of Excellence on Medical Informatics and Semantic Data Mining, the Volkswagen Foundation project "Forms of Life", and the National Center for Biomedical Ontology (the latter funded by the National Institutes of Health through the NIH Roadmap for Medical Research, Grant 1 U 54 HG004028). Thanks for helpful comments are due also to Thomas Beale and Walter Dierckx who however bears no responsibility for the positions here adopted.

References

[1] ANSI/HL7 V3 RIM, R1-2003: RIM Version V 02-11. Membership Normative Ballot Last Published: 11/22/2005 8:05 PM.
[2] http://ontology.buffalo.edu/HL7.
[3] Fernandez E, Sorgente T. An analysis of modeling flaws in HL7 and JAHIS, Proc 2005 ACM Symposium on Applied Computing, 2005;:216-223.
[4] Ceusters W, Smith B. Strategies for referent tracking in Electronic Health Records, J Biomed Inform, in press.
[5] Rector A, Nolan W, Kay S, Foundations for an electronic medical record. Methods Inf Med, 30 (1991): 179-86.
[6] Quine WVO. Word and Object. Cambridge, MA: MIT, 1970.
[7] Willard, D Why semantic ascent fails. Metaphilosophy, 1983;14: 276-290.
[8] Rescher R, Manor R On inference from inconsistent.premises, Theory and Decision, 1:1970;:179–217.
[9] Smith B. Reasoning with biomedical terminologies. J Biomed Inform, under review.
[10] Vizenor L. Actions in health care organizations: An ontological analysis. Proc MedInfo 2004, San Francisco, 1403-1407.
[11] Rosse C, Kumar A, Mejino JLV, Cook DL, Detwiler LT, Smith B. A strategy for improving and integrating biomedical ontologies. Proc AMIA Symp 2005;:639-643.
[12] Markwell D, et al. Using SNOMED CT in HL7 Version 3. Implementation Guide, Release 1.0. Terminfo Ballot Draft Oct 29, 2005.