An ontological analysis of diagnostic assertions in electronic healthcare records

Werner Ceusters[1,*] and William R. Hogan[2]
[1] Department of Biomedical Informatics, University at Buffalo, 921 Main street, Buffalo, USA
[2] Department of Health Outcomes and Policy, University of Florida, 2004 Mowry Rd, Gainesville, Florida, USA

ABSTRACT
We present a comparative analysis of two sets of Referent Tracking Tuples (RTT), which each author of this paper crafted independently from the other and which are about the same portion of reality that one could assume to be described faithfully through registered diagnoses in the problem list of an electronic healthcare record system. The analysis thereby focused on (1) the choice of particulars that each of the authors deemed necessary and sufficient for an accurate description, (2) what these particulars are instances of, (3) how they relate to each other, and (4) the motivations of each author for the choices made. It was found that despite the large variety in RTTs crafted, there was wide, though not total, agreement about the appropriateness of the choices made. Disagreements arose from various issues such as potential lack of orthogonality in the OBO Foundry and in some cases on what types the classes in the ontologies represent. The authors' main source of disagreement was due to different interpretations of the literature on Information Content Entities (ICEs).

1 INTRODUCTION

When during a clinical encounter a provider establishes a diagnosis for a patient under his care, he typically enters one or more diagnostic codes into that patient's electronic healthcare record (EHR). When the same patient is later seen by a distinct provider, for instance for a second opinion or for a reason not related to the first encounter, this second provider will also enter one or more diagnostic codes. Patients' records tend to accumulate many of these diagnostic assertions, specifically when the providers are working in the same EHR, or when such information is transferred from one to the other. They are also accumulated when records from various systems are merged into data warehouses equipped with master patient index facilities. There is a large variety amongst EHR systems and data warehouse interfaces in how they display such diagnostic information, a small hypothetical example being shown in Table 1.

1	**Patient ID**	**Diagnosis**		**Date entered**	**Entered by**
2	1234	274.9:	Gout, unspecified	9-1-2014	J. Doe
3	1234	715.97:	Osteoarthrosis, unspecified whether generalized or localized, ankle and foot	9-1-2014	S. Thump

Table 1: Two diagnoses provided on the same day, about the same patient, entered by two distinct EHR users.

A problem with information provided in this way, is that it is not possible to construct a completely accurate view on what is (and has been) the case in reality (Rector et al, 1991). A question which, in relation to the information in Table 1, cannot be answered reliably is whether the two diagnoses are about the very same disorder the patient suffers from (thereby highlighting different aspects of that disorder which cannot be expressed using a single ICD-code) or about two distinct disorders the patient suffers from simultaneously. Other questions are, for example, whether the persons that *entered* the diagnoses also *made* the diagnoses, how the dates when diagnoses were entered relate to the dates when the diagnoses were actually made, and so forth.

Referent Tracking (RT) is a methodology to avoid, resolve, and document these sorts of ambiguities in EHRs (Ceusters & Smith, 2006). This is achieved by building data stores composed of Referent Tracking Tuples (RTT). The core part of an RTT expresses a relationship that obtains between a particular—globally and singularly uniquely identified in the realm of the RT System used to generate (and track usage of) Instance Unique Identifiers (IUIs)—and either another particular or a universal (or defined class), representations of which are – ideally – taken from one or more ontologies that follow the principles of Ontological Realism (Ceusters & Manzoor, 2010; Smith & Ceusters, 2010). Whenever a continuant is referenced in an RTT, time indexing is used following the conventions outlined in (Smith et al., 2005). As an example, the following RTT—formulated in simplified abstract syntax—asserts that there exists a particular to whom IUI '#4' is assigned, and that this particular is an instance of human being during the time period to which the IUI 't5' is assigned:

#4 ***instance-of*** HUMAN BEING ***at*** t5 (Ex.1)

The methodology was expanded in (Ceusters et al., 2014) to translate datasets into assertions such that not only the portion of reality (POR) described by the dataset and the dataset itself are represented, but so also the relations between components of this dataset on the one hand and the corresponding PORs on the other hand.

The purpose of the work reported here was to assess to what extent the authors of this paper—two experts in RT—would be able to develop independently from one another a collection of RTTs that describe the same POR in a semantically-interoperable way. The analysis presented is the first

* To whom correspondence should be addressed: ceusters@buffalo.edu

step in this endeavor and focuses on (1) the choice of particulars deemed necessary and sufficient for an accurate description of the selected POR, (2) what these particulars are instances of, (3) how they relate to each other, and (4) the motivations of each author for the choices made.

2 METHODS

The POR selected for the experiment was the one ambiguously described in Table 1. Since the goal of the exercise was not to identify nor, when possible, resolve ambiguities, it was further specified that the diagnoses were about the same disorder, in the sense as formulated in the foundations for the Ontology of General Medical Science (OGMS) (Scheuermann et al., 2009). No instructions were given on what ontologies to use, or in what format to provide the RTTs. Results were exchanged in a password-protected file and the passwords disclosed after each author acknowledged receipt of the other's result. The authors then compared the original RTTs in stepwise fashion. The first step was to identify the particulars that both authors referred to in their assertions. Since both authors assigned IUIs independently, thereby assigning distinct IUIs to the very same particulars, a second step was then to re-assign IUIs as if the collection of RTTs was merged into one single RT system, thereby still keeping track of which RTT was asserted by which author. In a third step, this collection was then analyzed and differences in representations discussed, however without paying attention to the temporal indexing required for RTTs describing a POR in which a continuant is involved.

3 RESULTS

Table 2 lists the particulars and what they are instances of as originally—thus prior to comparison of the proposed representations—argued for by the author hereafter referred to as 'X'. Table 3 does so for author Y. Each row represents part of an RTT asserting that the particular denoted in the 'IUI'-column is an instance of the universal denoted by the representational unit (RU) in the 'Class'-column, drawn from the ontology named in the 'Ontology'-column. The description relates the particulars informally to the scenario analyzed. The column labeled 'Ind.' contains the IUIs of the Information Content Entities (ICE) of which the RTTs themselves are concretizations. The columns 'Y' and 'X' contain scores reflecting how Y, resp. X, after discussion considered the RTT appropriate, '0' meaning 'not at all', '1' 'ok, but', and '2' 'absolutely'. Table 4 and 5 list for X and Y respectively the RTTs involving non-instantiation relationships. An IUI or Index in bold indicates that the corresponding POR is referred to by both authors. Author X listed 21 particulars involving 23 instantiations; Y did so for 28 particulars involving 1 instantiation each, not counting in both cases as particulars the temporal regions related to the time-indexing required for certain RTTs.

Ind.	IUI	Description	Ontology	Class	Y	X
T1	**P1**	the patient	OBI	Homo sapiens	1	2
T2	**P2**	the doctor who made diagnosis #1	OBI	Homo sapiens	1	2
T3	**P3**	the doctor who made diagnosis #2	OBI	Homo sapiens	1	2
T4	**P4**	diagnosis #1	OGMS	Diagnosis	2	2
T5	**P5**	diagnosis #2	OGMS	Diagnosis	2	2
T6	P6	the disorder the patient has	OGMS	Disorder	2	2
T7	P6		DO	Gout	1	1
T8	P6		DO	Osteoarthritis	1	1
T9	P7	entry in problem list for diagnosis #1		Dataset record	2	2
T10	P8	entry in problem list for diagnosis #2		Dataset record	2	2
T11	**P9**	the process of doctor #1 making diagnosis #1	OGMS	Diagnostic process	2	2
T12	**P10**	the process of doctor #2 making diagnosis #2	OGMS	Diagnostic process	2	2
T13	P11	doctor #1's doctor role	OMRSE	Physician role	2	2
T14	P12	doctor #2's doctor role	OMRSE	Physician role	2	2
T15	P13	the patient's patient role	OMRSE	Patient role	2	2
T16	P14	EHR		EHR	2	2
T17	**P15**	patient ID cell of entry #1		Denotator	1	1
T18	**P16**	diagnosis cell of entry #1		Denotator	1	1
T19	P17	doctor cell of entry #1		Denotator	1	2
T20	**P18**	diagnosis cell of entry #2		Denotator	1	1
T21	P19	date cell of entry #2		Denotator	1	2
T22	P20	ICD-9-CM coding system			2	2
T23	P21	patient's afflicted foot	FMA	FMA:Foot	2	2

Table 2: Particulars and what they are instances of from the original perspective of author 'X'.

39 distinct particulars were identified, 10 of them by both authors. For only 2 (RTTs T2 and T3) did both authors select the same instantiating universal while for 2 others (T4 and T5) universals were selected from distinct ontologies, but with a close, nevertheless debatable, match. For the remaining 6, the universals chosen stand in is-a relations.

X drew 9 classes from 5 realism-based ontologies – the OGMS, the Ontology of Medically Related Social Entities (OMRSE), the Foundational Model of Anatomy (FMA), the Disease Ontology (DO) and the Ontology of Biomedical Investigations (OBI)—and identified the need for three more classes—'denotator', 'EHR' and 'dataset record', for which no realism-based ontology was found. Y used 9 classes drawn from 4 realism-based ontologies, 2 of which (FMA and OGMS) were also used by X, and 2 distinct ones: the Basic Formal Ontology (BFO) and the Information Artifact Ontology (IAO). He also identified the need for 2 classes currently without an ontological home: 'patient identifier' and 'ICD-9-CM code and label', as well as 2 classes (Gout and Osteoarthrosis, R49 and R51 in Table 5) for which he did not identify any particular as being required for an accurate description of the scenario.

53 particular-to-particular relationships in total were represented: 22 alone by X, 27 alone by Y and 4 (R14, R21, R22 and R23 in Table 4 and 5) by both authors, be it nevertheless through distinct, yet synonymous formulations. Y listed also two RTTs, each one expressing aboutness between a particular and a universal (R49 and R51, Table 5).

Ind.	IUI	Description	Ontology	Class	Y	X
T24	P22	the ICE which is concretized in the spreadsheet you might be looking at	IAO	ICE	2	2
T25	P23	the portion of chalk on the blackboard which make up what we call 'that spreadsheet'	BFO	Material entity	2	2
T26	P24	the pattern of chalk lines, spaces, characters, etc., in that portion of chalk	BFO	Quality	2	2
T27	**P1**	the material entity whose ID is ‘1234’ in the spreadsheet	BFO	Material entity	1	1
T28	**P15**	the patient identifier which is concretized in each first cell of the 2nd and 3rd row of the concretization of P22		Patient identifier	2	2
T29	P25	the portion of chalk making up the text string ‘1234’ in the first cell of the 2nd row	BFO	Material entity	2	2
T30	P26	the quality in P25 which makes P25 a concretization bearer	BFO	Quality	2	2
T31	P27	portion of chalk making up the string ‘1234’ in the 1st cell of the 3rd row of the spreadsheet	BFO	Material entity	2	2
T32	P28	the quality in P27 which makes P27 a concretization bearer	BFO	Quality	2	2
T4	**P4**	the diagnosis which is concretized in the first two cells of the 2nd row of the concretization of P22 in front of your eyes	OGMS	Diagnosis	2	2
T33	P29	the quality through which **P4** is concretized	BFO	Quality	2	2
T5	**P5**	the diagnosis concretized in the first two cells of the 3rd row of the concretization of P22 in front of your eyes	OGMS	Diagnosis	2	2
T34	P30	the quality through which **P5** is concretized	BFO	Quality	2	2
T2	**P2**	the person whose name is ‘J. Doe’ in the spreadsheet	FMA	Human being	1	1
T3	**P3**	the person whose name is ‘S. Thump’ in the spreadsheet	FMA	Human being	1	1
T35	P31	the clinical picture about **P1** available to **P2** and **P3**	OGMS	Clinical picture	2	2
T36	P32	part of the life of **P1** which is described in P31	OGMS	Bodily process	1	1
T37	**P9**	the interpretive process which resulted in **P4**	OGMS	Bodily process	1	2
T38	**P10**	the interpretive process which resulted in **P5**	OGMS	Bodily process	2	2
T39	P33	the disease in **P1**	OGMS	Disease	2	2
T40	**P16**	the ICE concretized in the 2nd cell of the 2nd row		Icd-9-cm code and label		
T41	P34	the quality through which **P16** is concretized	BFO	Quality	2	2
T42	**P18**	the ICE concretized in the 2nd cell of the 3rd row		Icd-9-cm code and label	2	2
T43	P35	the quality through which **P18** is concretized	BFO	Quality	2	2
T44	P36	the process of, as we say ‘entering’ diagnosis 1 in the EHR'	BFO	Process	2	2
T45	P37	the quality of some part of some hard disk which concretizes d1	BFO	Quality	2	2
T46	P38	the process of, as we say ‘entering’ diagnosis 2 in the EHR’	BFO	Process	2	2
T47	P39	the quality of some part of some hard disk which concretizes diagnosis 2	BFO	Quality	2	2

Table 3: Particulars and what they are instances of from the perspective of author ‘Y’.

Ind.	:	RTT in abstract syntax without time-component			Y	X
R1	:	**P1**	RO:bearer of	P13	2	2
R2	:	**P1**	RO:has part	P6	2	2
R3	:	**P1**	RO:has part	P21	2	2
R4	:	**P10**	RO:realizes	P12	2	2
R5	:	T7	corresponds with	**P16**	2	2
R6	:	T8	corresponds with	**P18**	2	2
R7	:	**P15**	RO:part of	P7	1	1
R8	:	**P15**	RO:part of	P8	1	1
R9	:	**P15**	IAO:denotes	**P1**	0	1
R10	:	**P16**	RO:part of	P7	1	1
R11	:	**P16**	IAO:denotes	**P4**	1	1
R12	:	P17	RO:part of	P7	1	1
R13	:	P17	IAO:denotes	**P2**	0	1
R14	:	**P2**	RO:agent of	**P9**	2	2
R15	:	**P2**	RO:bearer of	P11	2	2
R16	:	**P18**	RO:part of	P8	1	1
R17	:	**P18**	IAO:denotes	**P5**	0	1
R18	:	P19	IAO:denotes	**P3**	0	1
R19	:	P21	RO:has part	P6	1	1
R20	:	**P3**	RO:bearer of	P12	2	2
R21	:	**P3**	RO:agent of	**P10**	2	2
R22	:	**P4**	OBI:is specified output of	**P9**	2	2
R23	:	**P5**	OBI:is specified output of	**P10**	2	2
R24	:	P7	IAO:is about	P6	2	2
R25	:	P8	IAO:is about	P6	2	2
R26	:	P9	RO:realizes	P11	2	2

Table 4: particular to particular relationships listed by author X

Ind.	:	RTT in abstract syntax without time component			Y	X
R27	:	P24	inheres-in	P23	2	2
R28	:	P24	concretizes	P22	2	2
R29	:	**P15**	part-of	P22	2	2
R30	:	P25	bears-concretization-of	**P15**	2	2
R31	:	P26	inheres-in	P25	2	2
R32	:	P26	is-about	**P1**	2	2
R33	:	P27	bears-concretization-of	**P15**	2	2
R34	:	P28	inheres-in	P27	2	2
R35	:	P28	is-about	**P1**	2	2
R36	:	P29	concretizes	**P4**	2	2
R37	:	P29	is-about	**P1**	2	2
R38	:	P29	is-about	P33	2	2
R39	:	P30	concretizes	**P5**	2	2
R40	:	P30	is-about	**P1**	2	2
R41	:	P30	is-about	P33	2	2
R42	:	**P2**	agent-of	P36	2	2
R43	:	**P3**	agent-of	P38	2	2
R44	:	P32	has-participant	**P1**	2	2
R22	:	**P9**	creates	**P4**	2	2
R14	:	**P9**	has-agent	**P2**	2	2
R45	:	**P9**	has-input	P31	2	2
R23	:	**P10**	creates	**P5**	2	2
R21	:	**P10**	has-agent	**P3**	2	2
R46	:	**P10**	has-input	P31	2	2
R47	:	P33	inheres-in	**P1**	2	2
R48	:	P34	concretizes	**P16**	2	2
R49	:	P34	is-about	GOUT	2	2
R50	:	P35	concretizes	**P18**	2	2
R51	:	P35	is-about	OSTEOARTHROSIS	2	2
R52	:	P36	creates	P37	2	2
R53	:	P37	concretizes	**P4**	2	2
R54	:	P38	creates	P39	2	2
R55	:	P39	concretizes	**P5**	2	2

Table 5: relationships other than instance-of listed by author Y.

X indicated from which ontologies the relationships were drawn. Y used relations from the BFO 2.0 Draft Specifications, or under discussion in the context of the IAO.

4 DISCUSSION

Despite the large variation in RTTs crafted for what at first sight looks like a simple POR, there was after discussion wide, though not total agreement, about the appropriateness of the choices made (agreement is indicated by the same scores appearing in the X and Y columns of Tables 2-5). 'Appropriateness' is here to be measured in terms of what an optimal collection of RTTs for the POR under scrutiny would be since one could argue that the ground truth for what is expressed in EHR entries is largely unknown. The 'ground truth' is thus much broader than just what the patient had (this being part of the non-assertional part of the POR): it includes what the clinicians stated about what the patient had (these statements being part of the assertional part of the POR). If what the patient precisely had cannot be inferred from what was stated, it would be wrong to construct a collection of RTTs that states that the patient has such or such a specific type of disorder. To represent the non-assertional part of a POR that a collection of assertions is about, one has to resort to these assertions and to what has already been established to be the case through other means. The optimal collection of RTTs would be the one which satisfies the following criteria: (1) it consists of RTTs which describe the non-assertional part of the POR only to the extent to which there is enough evidence for what those RTTs themselves assert to be true (e.g. there is sufficient evidence that the patients are human beings, there is not sufficient evidence that the diagnoses are correct) and (2) it consists of other RTTs which describe the assertional part in relation to the RTTs referenced under (1).

We note here that the level of disagreement in the representations of X and Y do not invalidate the RT method, but rather reflect the need for uniform conventions on which ontologies and relations to use, as well as problems in the ontological theories, their implementations, and documentation that were available to represent the scenario. We return to this issue as we discuss the major sources of disagreement. Indeed, as will become clear, this work shows that the RT method is a stringent test of ontologies.

Although both authors agreed on the necessary existence of the patient (P1) and the two clinicians (P2, P3) for the analyzed scenario to be faithful to reality, they each selected distinct universals to assert instantiation. X represented P1 as a human with a patient role. Y represented P1 as a material entity without assigning a patient role, his choice of material entity being motivated by the fact that P1 has been a material entity all the time through its existence, but not a human (e.g., it was a zygote at a time prior to being human). This difference in representation is related to the temporal indexing that RT requires for continuants, an element not further discussed in this paper. But given the two authors' temporal indexing, both authors agree that each other's instantiations were correct.

Both authors disagreed though about how to interpret the representational units for the universal *Human being* from the selected ontologies. Y used 'human being' as synonym for the FMA's 'human body' class, although FMA does not list synonyms. Y argued against X's 'Homo sapiens' taken from OBI based on its linking to other ontologies in Ontobee, which altogether seem to confuse 'Homo sapiens' as an instance of the universal 'species' with those instances of organism that belong to – but are not instances of – the species 'Homo sapiens'. X counters that despite the use of species names, 'Homo sapiens' and similar classes in OBI all descend from a class called 'organism'. Also, the 'Homo sapiens' class in OBI has synonyms 'Human being' and 'human'. It would be an enormous task indeed to find non-taxonomic names for every type of organism in the world and refactor ontologies based on the NCBI Taxonomy on this basis. The problem here is the lack of face value of terms selected as class names in the respective ontologies.

Both authors agreed on the existence of a disorder and a disease resulting from it in the patient, as well as two diagnoses and the two distinct processes that generated each. Both authors also agree that none of these entities should be confused or conflated: nothing at the same time can be an instance of two or more of the following: disease, disorder, diagnosis, and diagnostic process.

A problem is that the Disease Ontology selected by X, confuses not only disorders and disease, but also disease courses. For instance 'physical disorder' in DO is a subtype of 'disease', in direct contradiction to OGMS. X agrees that DO is flawed, however it is the only ontology of disease that at least purports to strive for compliance with realist principles, and represents an improvement over flawed medical terminologies such as SNOMED-CT and NCI Thesaurus. If perfection were a requirement to use an ontology, we could make no progress. Nevertheless, the persistent, glaring flaws of DO from the perspective of OGMS give serious pause on using it accurately and precisely.

Both authors offered a different perspective on what parts of Table 1 actually constitute a diagnosis. They agreed that any such table—whether presented as a problem list on a video monitor or tablet as the scenario worked with by X, or as a spreadsheet drawn with chalk on a blackboard as envisioned by Y—is built out of continuants that are concretizations of instances of ICE reflecting a diagnosis. But whereas X identified the mere concretization of the ICD-code and label to be denoting the diagnosis, Y argued that also the concretization of the patient identifier is part of that which concretizes the diagnosis (a requirement for the diagnosis to stand in an 'is about' relation to the patient). This, and a large amount of other differences in representation, were due to distinct interpretations of the literature on the nitty-gritty of how to deal with ICE and concretizations thereof, how instances of ICE relate to other instances of ICE, and what exactly the relata are of relationships such as aboutness and denotation. Whereas, for instance, X committed to ICE being parts of other ICE, Y commits only to parthood of the independent continuants in which inhere the qualities that concretize the corresponding ICE, without making that clear in the representation however. Another key issue with ICE is that Y represented the qualities concretizing the ICE as being about something, whereas X followed the IAO where the ICE itself denotes or is-about something. After further analysis both authors agreed that Y's representation is better and that it advances the theory of ICE in IAO.

Although it was a priori agreed upon that the patient in the scenario would have only one disorder, an ambiguity that was left open was whether both diagnoses were actually correct: so Table 1 could be interpreted in distinct ways: (1) both diagnoses are correct from a medical perspective and describe distinct aspects of the same disease, or (2) at least one diagnosis is wrong. Also, because RT tuples contain provenance as to whom is the source of the statement contained therein – note that Ex.1 above is a simplified representation not containing the provenance information – X interpreted both (1) the RT tuple that instantiated the disease as gout (by Doe) and (2) the RT tuple that instantiated it as osteoarthritis (by Thump) as being faithful representations of what Thump and Doe believed at the time they formulated their diagnoses. X did not believe himself to be recognizing both diagnoses as straightforwardly accurate and therefore resorted in his representation to a mechanism offered in RT to craft RTTs about RTTs that are later found to have been based on a misunderstanding of the reality at the time they were crafted (Ceusters, 2007). Y crafted a representation that does not commit to what specific disease type(s) the patient's disease actually is an instance of. This was achieved by representing the diagnoses to be simultaneously about the patient on the one hand (in contrast to X who represents the diagnoses to be about the disorder/ disease itself), and about the disease universals – gout and osteoarthrosis resp. – denoted by the respective ICD-codes and labels on the other hand. This aboutness-relation between an instance of ICE and a universal can be represented in RT but of course cannot be represented in OWL without recourse to workarounds such as those discussed by Schulz et al (2014).

Although both authors resorted to OGMS for a large part of their RTTs, differences in representation were observed because of the source material consulted: X used the OGMS OWL artifact as basis, whereas Y used the definitions and descriptions in the paper that led to the development of OGMS (Scheuermann et al., 2009).

5 CONCLUSION

In representing a common scenario in healthcare which EHR data are about, the two authors agreed on key entities including the patient, doctors, diagnoses, and the processes by which the doctors generated the diagnosis. Although they agreed in general about the types instantiated by the particulars in the scenario, and how the particulars are related to each other, they chose different representational units and relations from different ontologies due to various issues such as potential lack of orthogonality in the OBO Foundry and in some cases disagreement on what types the classes in the ontologies represent. These distinctions exist, not because the authors entertained distinct competing conceptualizations, but because they expressed matters differently.

Differences in the choice of ontologies constitute a risk: distinct ontologies may represent reality from distinct perspectives and despite being veridical might not be derivable from each other because the axioms required to do so would be missing, for the simple reason that such axioms would fall outside the purpose of the specific ontologies. This would lead the representations by each of the authors not to be semantically interoperable unless additional ontology bridging axioms would be crafted. The authors' main source of disagreement was due to different interpretations of the literature on ICEs, which ultimately led to a planned reformulation of the theory of ICE and reference. Although this study is limited by the participation of only 2 subjects and the analysis of one report, it highlights the fact that the RT method and the clarity and precision it requires in representing reality is a powerful tool in identifying areas of needed improvement in existing, realism-based ontologies.

ACKNOWLEDGEMENTS

This work was supported by award UL1 TR000064 from the National Center for Advancing Translational Sciences. This paper is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

REFERENCES

Ceusters, W. (2007). Dealing with mistakes in a referent tracking system. In H. KS (Ed.), *Proceedings of ontology for the intelligence community 2007 (oic-2007)* (pp. 5-8). Columbia MA.

Ceusters, W., Chiun Yu Hsu, & Smith, B. (2014). Clinical data wrangling using ontological realism and referent tracking. *CEUR Workshop Proceedings, 1237,* 27-32.

Ceusters, W., & Manzoor, S. (2010). How to track absolutely everything? In L. Obrst, T. Janssen & W. Ceusters (Eds.), *Ontologies and semantic technologies for the intelligence community. Frontiers in artificial intelligence and applications.* (pp. 13-36). Amsterdam: IOS Press.

Ceusters, W., Smith, B. (2006). Strategies for referent tracking in electronic health records. *Journal Biomedical Informatics, 39*(3), 362-378.

Rector, A.L., Nowlan, W.A., & Kay, S. (1991). Foundations for an electronic medical record. *Methods Inf Med, 30*(3), 179-186.

Scheuermann, R.H., Ceusters, W., & Smith, B. (2009). Toward an ontological treatment of disease and diagnosis. *Summit on Translat Bioinforma, 2009,* 116-120.

Schulz, S., Martínez-Costa, C., Karlsson, D., Cornet, R., Brochhausen, M., Rector, A. (2014). An ontological analysis of reference in health record statements. In: *Proceedings of FOIS 2014,* 2014.

Smith, B., & Ceusters, W. (2010). Ontological realism: A methodology for coordinated evolution of scientific ontologies. *Applied Ontology, 5*(3-4), 139-188. doi: 10.3233/Ao-2010-0079

Smith, B., Ceusters, W., . . . Rosse, C. (2005). Relations in biomedical ontologies. *Genome Biology, 6*(5): R46. doi:10.1186/Gb-2005-6-5-R46