A Realism-Based View on Counts in OMOP's Common Data Model

Werner CEUSTERS [a,b,1] and Jonathan BLAISURE[a,b]
[a] *Department of Biomedical Informatics, University at Buffalo, Buffalo NY, USA*
[b] *Institute for Healthcare Informatics, University at Buffalo, Buffalo NY, USA*

Abstract. Correctly counting entities is a requirement for analytics tools to function appropriately. The Observational Medical Outcomes Partnership's (OMOP) Common Data Model specifications were examined to assess the extent to which counting in OMOP CDM compatible data repositories would work as expected. To that end, constructs (tables, fields and attributes) defined in the OMOP CDM as well as cardinality constraints and other business rules found in its documentation and related literature were compared to the types of entities and axioms proposed in realism-based ontologies. It was found that not only the model itself, but also a proposed standard algorithm for computing condition eras may lead to erroneous counting of several sorts of entities.

Keywords. Common data models, OMOP, realism-based ontologies

1. Introduction

Realism-based ontology (RBO) design following the principles of Ontological Realism [1] is often criticized as being too difficult [2]. The difficulty is indeed witnessed by (1) the slow pace at which the Basic Formal Ontology (BFO) [3] evolves and (2) the few ontologies that become accepted as an Open Biomedical Ontologies (OBO) Foundry ontology [4] despite the large uptake of the BFO [3, p160-2]. One reason is that ontological commitment in BFO is exclusively towards what has (or once had) objective 'mind-independent' existence. This includes individual entities – '*particulars'* – such as Werner Ceusters and Jonathan Blaisure and the types – '*universals'* – they both belong to such as human being, mammal, and vertebrate. RBOs deal with the entities science commits to and provides an additional perspective based on the most general features of reality common to all specific domains such as unity, identity, difference, parthood and aboutness. This typically surpasses what scientists and information modelers work with. The second reason is that types must have an Aristotelean definition stating all necessary conditions that are also jointly sufficient and satisfiable for some entity to be an instance of the respective type. Terms for these types should also have general face validity independent of context [5].

Concept-based views such as the one endorsed by SNOMED CT [6] require mere adherence to formal logical criteria and an ontological commitment not to reality, but to a 'universe of discourse'. The notion of '*concepts as shared meanings*' seems to be

[1] Werner Ceusters, Department of Biomedical Informatics, Jacobs School of Medicine and Biomedical Sciences, University at Buffalo, 77 Goodell street, Suite 540, Buffalo NY 14203, USA; E-mail: ceusters@buffalo.edu.

perceived as more intuitive and closer to the language used in the domains modeled than the quite 'abstract' notions of 'universals' and 'portions of reality' as employed by realists. There is no place in RBOs for types such as 'prevented abortion' and 'absent nipple', while from a concept-based perspective their meaning as concepts seems clear. Metaphorically, while conceptualists can '*see'* both *absent nipples* and *persons with absent nipples*, realists can of these two only see *persons with absent nipples*. But as a consequence, conceptualists must be careful not to misclassify prevented abortions and absent nipples as special kinds of abortions, resp. nipples. The challenge for realists is to find ways to formally describe in what way persons with absent nipples are different from persons with prevented abortions without resorting in these descriptions to absent nipples and prevented abortions [7]. Such considerations are indeed hard for formal terminologists and information modelers without a solid education in philosophy.

2. Background

The University at Buffalo's Institute for Healthcare Informatics' (IHI) primary mission is to aggregate fully identified healthcare data sets from distinct sources (in- and outpatient electronic healthcare records (EHR), claims data, clinical studies, patient-reported outcomes, …) into a centralized secure environment where the data can be documented and appropriately distributed for secondary data use projects. Although the long-term goal is to have this data repository satisfy the principles of Ontological Realism [1, 8], various intermediate steps are explored to speed up the possibilities for advanced analytics. One such step is the use of common data models (CDM).

Although CDMs have been introduced in healthcare since the early nineties [9], the application thereof grew considerably after the Food and Drug Administration launched the *Sentinel Initiative* in 2008 to use data from multiple existing data systems for post-marketing surveillance [10]. Administrative data routinely collected by medical practices, hospitals, delivery systems, health plans, and insurers were the first sources. Medicare and Medicaid databases of prescriptions as well as disease and vital-statistics registries were to be brought in the mix later. CDM-based initiatives follow a distributed data approach in which participating partners keep the electronic data in their operational systems while the CDM provides standardization for, typically, administrative and clinical information across partners [11]. Since then, several such CDMs have been developed in parallel for various purposes, yet based on the same idea. Notable examples are the CDMs of the Observational Medical Outcomes Partnership (OMOP) [12], the Patient-Centered Outcomes Research Network (PCORnet) [13], the healthcare management organizations' research network (HMORN) virtual data warehouse [14] and the Study Data Tabulation Model (SDTM) of the Clinical Data Interchange Standards Consortium (CDISC) [15].

Several of these CDMs have been subjected to evaluation and comparison for pooling data from EHRs. In [16], it was found that the OMOP CDM best met the criteria for supporting data sharing from longitudinal EHR-based studies. The OMOP CDM also scored best when assessed for the purposes of comparative effectiveness research (CER) based on data extracted from EHRs [17]. Nevertheless, in several papers concerns have been expressed about the loss and distortion of information that may occur when EHR data are translated and pooled into CDM-based data repositories [18-20]. Attempts have therefor been made to streamline CDM evaluation methods [16, 21]. None, however, resorted to ontological methods of the sort applied at the IHI,

reason for which we started a project to investigate how the principles of Ontological Realism and the use of entities defined in realism-based ontologies, more specifically the Basic Formal Ontology (BFO) [3], the Ontology for General Medical Science (OGMS) [22], the Information Artefact Ontology (IAO) [23] and the Ontology for Biomedical Investigations (OBI) [24] can be used (1) to identify potential areas for improvement in CDMs, (2) to detect ambiguities and hidden assumptions in source data, and (3) to assist in the development of appropriate Extract-Transfer-Load (ETL) procedures to translate source EHR data into the syntax and semantics of a CDM.

The findings in favor of the OMOP CDM [16, 17] and the availability of open-source analytics tools that operate on data repositories built accordingly prompted us to select this CDM as an example for how to scrutinize CDMs on the basis of ontological principles. In this paper, we focus on just one ontological principle and the various ways in which this principle is violated not only in the design of the OMOP CDM, but also on how data sources are – and are suggested to be – translated following this CDM: the principle of appropriately counting individual entities, a principle we believe to be an absolute requirement for running analytics based on statistics. What follows is not an assessment of problems based on ontological misrepresentation of counts based on data, but an assessment highlighting where the OMOP CDM's specifications may lead to miscounting in data repositories that follow these specifications.

3. Methods

The methodology of our analysis rests on the distinction between data and data models on the one hand, and that what the data and data models *are about* on the other hand. That what they are about, is called in Ontological Realism a *portion of reality* (PoR). RBOs perceive PoRs as being composed out of *types* (such as HUMAN BEING, QUALITY and HEIGHT – types are standardly written in SMALL CAPS) and *particulars*, i.e. entities that carry identity (such as Donald Trump and his particular height). It is these particulars that should only be counted once when tallies are made in analytic algorithms [25]. Relationships obtain (1) between particulars such as the ***inheres-In*** relation between Donald Trump's height and Donald Trump, (2) between particulars and types, such as the ***instance-Of*** relation between Donald Trump and HUMAN BEING, and (3) between types, such as the **Is-A** relation between HEIGHT and QUALITY. RBOs view EHRs, CDMs, and their parts such as tables and fields as INFORMATION CONTENT ENTITIES (ICE). For example, OGMS recognizes a DIAGNOSIS as an ICE that, in the typical case, stands in an ***is-About*** relationship with a *configuration* formed by the HUMAN BEING about whom the DIAGNOSIS is made, the DISEASE that ***inheres-In*** that HUMAN BEING, the HUMAN BEING that made the DIAGNOSIS, and so forth [22, 23].

Our strategy consists thus of first identifying the PoR described by the OMOP CDM as well as the PoRs of data repositories referenced in the literature as being in a format conformant to the OMOP CDM. These PoRs need then be described using the types, relationships and axioms available in the RBOs listed above. The resulting representational artifacts, henceforth called the *RBO perspectives*, can then be compared with the original structures – the *OMOP perspectives* – of the OMOP CDM and the instantiations thereof in data repositories. Whereas the RBO perspective of the PoR derived from the CDM should primarily reference *types* and relationships between types, the RBO perspective on the data sources translated in terms of the OMOP CDM should reference particulars and what types these particulars instantiate. This is because

CDMs themselves, in contrast to the data repositories built according to these CDMs, represent what is general in the domain in a similar way as ontologies do.

The OMOP CDM's design is intended '*to accommodate data from the observational medical databases that are generally considered necessary for active safety analysis'* thereby being *'analyst-friendly*' so as to '*allow the analytic methods to execute quickly enough to be practical*' [26, p55]. The RBO perspective, in contrast, is purpose independent, thereby reflecting only how reality is structured [1]. It is therefore hypothesized that to allow the desired 'fitness for purpose' the OMOP perspectives would not only be reductionist, but perhaps also use constructs that do not correspond with a realist type. Differences between the two sorts of perspectives can then be classified along the forms of reductionism and deviant constructs encountered.

Two types of sources are used to obtain the OMOP perspectives. First, all files of the latest OMOP CDM version (v5) were downloaded from the GitHub site maintained by the Observational Health Data Sciences and Informatics (OHDSI) collaborative [27] and installed into a local instance of a Postgres SQL server. A script was run to document the relationships between the tables and how they were defined in the relational schema of the CDM. The corresponding OMOP documentation [28] (table and field descriptions, cardinality of relationships, etc.) was consulted to derive the informal semantics of the model, more specifically the types of entities existing in reality and how they relate to each other as perceived through the glasses of the CDM.

To detect issues with source data and the conversion thereof into the OMOP CDM, as well as recommendations to remediate or avoid them as reported in the literature, PubMed, as a second source, was queried with the strings 'OMOP' and 'Observational Medical Outcomes Partnership', what resulted after elimination of duplicates in 30 papers to be analysed. The goal was to obtain insight in the extent to which users of the CDM understood, and possibly encountered difficulties with its interpretation.

4. Results

We identified thus far three potential sources of error for inappropriately counting particulars in OMOP-compatible data repositories.

A first one is brought about by the cardinality and attribute restrictions in the OMOP CDM in contrast to the larger number of occurrences of specific sorts of configurations in which particulars of a type as denoted by the CDM can take part. As an example, the person table allows for each unique patient only one location, one gender, one primary care provider and one care site (the location of the primary care provider), although it is acknowledged in the documentation that patients over time can have distinct locations, genders, etc., whereby '*it is the responsibility of the data holder to select the one value to use in the CDM*' [28, p37]. What criteria to use to that end is left unspecified, which is awkward in light of the multiple observation periods that are allowed per patient. Other examples of this type of mistake are the impossibility to register the participation of multiple providers in the same visit, or the involvement of more than one care site in the same visit. The ETL programmer is in these cases instructed to either document the criterion used to pick one provider or care site, or otherwise leave the corresponding data fields blank [28, p41]. Mistakes of this sort lead to counting *less* particulars of specified sorts than there are in reality.

The opposite sort of mistake, counting more particulars than there are in reality, is introduced, for example, by keeping the person table exclusively reserved for *patients*

and a *provider* table which '[…] *contains a list of uniquely identified health care providers. These are typically physicians, nurses, etc.*' [28, p55], thus also *persons*, i.e. HUMAN BEINGS. It might thus very well be the case that a provider referenced in the provider table is also referenced in the person table as patient. For each such particular person there would be double counting whenever one would query the CDM repository for the number of people (= providers and patients) involved. This can only be avoided by assuring during ETL that whenever a specific person is referenced both as provider and as patient, the primary key generated for this person in the person table is exactly the same as the primary key generated for that person in the provider table. As a work-around within the constraints of the OMOP CDM, an alternative might be specifying identity by means of a dedicated relationship in the fact relationship table, but such relationship is currently not offered, and it would be a solution which nevertheless is not in line with realist ontology criteria.

A third mistake is brought about by the way in which conditions (that what is wrong with the patient, a 1st-order entity, i.e. an entity that not ***is-About*** something) are differentiated from diagnoses (a 2nd-order entity, i.e. that what is claimed about the corresponding condition as 1st-order entity), the cause being an insufficient appreciation of the distinction between data and what data are about. The basics of the OMOP approach are nevertheless laudable, but the problems are in the details. There is a *condition occurrence* table which '*captures records of a disease or a medical condition based on evaluation by a provider or reported by a patient*' [28, p10] and a *condition era* table, such era being defined as '*a span of time when the Person is assumed to have a given condition*' [28, p12]. It is this 2nd table which attempts to prevent double counting of medical conditions what would happen if counting would be done on the basis of the multiple statements made in the condition occurrence table about some particular condition: twice stating that a patient has a brain tumor, does indeed not make this patient have two brain tumors.

Unfortunately, the method proposed to prevent double counting seems flawed. The construction of the era table is said to be '*derived from the condition occurrence table using a standardized algorithm*' [28, p69] which according to [29, p656] is based on the following rule: '*Condition eras that represent the same condition concept from the Terminology Dictionary are aggregated if the start of the second era occurs within 30 days of the end of first era*'. This works fine for aggregating two condition occurrences in which the condition is represented in both cases by means of, for example, ICD-9 code 211.3 (Benign neoplasm of colon). However, the example – intended to be positive – given in [29, p656] is not convincing at all: '*a Condition Era representing ICD-9 code 410.01 (Acute Myocardial Infarction (AMI) of anterolateral wall, initial episode) would be aggregated to a Condition Era representing ICD-9 code 410.41 (AMI inferior wall, initial episode) occurring within 30 days as both of these ICD-9 codes annotate to the same Condition Concept, Acute Myocardial Infarction, within the MedDRA hierarchy'*. Obviously, something which is an anterolateral wall AMI is not an inferior wall AMI as much as something that is a tiger is not a panther. Moreover, that something which is an inferior wall AMI and something else that is an anterolateral wall AMI are both an AMI – as in the example is acknowledged by the MedDRA hierarchy – does not make these two things just one thing: although panthers and tigers are felines, a cage holding one panther and one tiger does not hold just one feline. So the decision to consider each of these two occurrences of a more specific type of AMI as being just one occurrence of a less specific type of AMI is wrong!

5. Discussion

One could argue that for the purposes of OMOP erroneously counting particulars of the sort mentioned above does not matter. Yet, cases in which more providers and/or care sites are involved, might be more complex cases while the OMOP constraints don't allow them to be recognized as such. The outcomes of certain types of AMI are different than other types of AMI, and lumping them all together blurs reality. But even if it would truly be the case that for the specified purposes it doesn't matter, it will matter for other types of studies. Should then a multitude of CDMs be developed for each such purpose? On the basis of, for example, [26, p60] we would argue against it: '*Converting data to the OMOP CDM required significant effort, a broad range of expertise, and extensive computational resources. We underestimated the level of effort and the breadth of skill sets (especially data mapping) required by the OMOP distributed research partners to convert their data to the OMOP CDM*'.

The alternative would be a CDM which purpose it is to represent reality as it is, based on types that do not leave room for interpretation, and specifications that allow appropriate counting. For example, RBOs distinguish ROLES, such as PATIENT-ROLE and PROVIDER-ROLE, from the HUMAN BEING(s) in which these ROLES ***inhere***. On this basis, we would suggest the introduction of three tables: one for HUMAN BEING(s), one for PROVIDER-ROLES and one for PATIENT-ROLES. RBOs recognize RELATIONAL-QUALITIES, particulars instantiating this type being the particular patient provider relations obtaining between particular HUMAN BEING(s) in which ***inheres*** a PATIENT-ROLE and HUMAN BEING(s) in which ***inheres*** a PROVIDER-ROLE. A similar approach would hold for care sites, locations, etc. The mechanism to do so exists in the OMOP CDM by means of the fact relationship table. Unfortunately, appropriate principles for using it correctly, at least from an RBO perspective, are lacking.

Similarly, issues with how to count conditions and condition eras can be solved by introducing in the model the distinctions made in OGMS between DISORDERS (the physical basis for a disease, e.g. a genetic defect, a pathogen, a tumor), DISEASES (the *disposition* for the occurrence of pathological processes because of some disorder), DISEASE-COURSES (the realization of the disease through the actual occurrence of pathological processes such as turning malignant, metastasizing, etc.) and DIAGNOSES (assertions about any of the former) [22]. The notion of 'condition' as used by OMOP blurs indeed these distinctions and it can be expected that the standard algorithm would create two condition eras for a benign tumor that turns malignant, whereas in reality there is in such case only one tumor. If this tumor when it was first diagnosed at time t_1 would have been given an instance unique identifier (IUI) [30], e.g. '#-1', then, from the RBO perspective, the assertion '#-1 ***instance-of*** BENIGN TUMOR ***at*** t_1' would serve as a diagnosis made about that tumor. If that tumor later would become malignant, then that would lead to a new diagnosis to the effect that '#-1 ***instance-of*** MALIGNANT TUMOR ***at*** t_2'. The tumor keeps its identity, i.e. #-1, when changing from being benign to malignant, just in the same way as the authors of this paper kept their identities while changing from children into adults. And there would only be one disease course, say '#-2 ***instance-of*** DISEASE COURSE', such that '#-1 ***participant-of*** #-2 ***at*** t_1' and '#-1 ***participant-of*** #-2 ***at*** t_2'. If that tumor at t_3 would develop a nearby satellite tumor '#-3', then that would be represented by three new assertions: (1) '#-3 ***instance-of*** MALIGNANT TUMOR ***at*** t_3', '#-3 ***participant-of*** #-2 ***at*** t_3' and (3) ''#-3 ***derives-from*** #-1 ***at*** t_3' [31]. When in the same patient a de novo malignant tumor #-4 would develop

independent of #-1 – whether before or after it turned malignant – then there would be a 2nd disease course related to #-4, and independent from disease course #-2.

Of course, a conversion of EHR data into such an RBO CDM would require even more effort and a larger skill set than required for the OMOP CDM. And, granted, the way most, if not all, EHR systems are currently designed and (mis-)used [18], would still leave room for error although also there appropriate measures can be taken to reduce misrepresentation despite inappropriate information models [32]. But such conversion would only need to happen once for any given type of EHR.

6. Conclusion

Research purpose-oriented CDMs that come with open-source analytics tools surely hold promises for secondary use of EHRs and other sorts of operational data. But when designed using traditional information modeling methodologies that focus on 'fit-for-purpose', the resulting representations do not allow the entities these representations are about to be counted correctly. When counting goes wrong, analytics goes awry. A realism-based approach is able to detect the root causes for such mistakes, and may contribute to remediate them. This requires however thorough education in the principles of ontological realism itself, as well as in how to apply them in traditional information modelling methods and technologies. This is not only relevant for post-hoc secondary data analysis, but also in the context of personal health and the growing number of devices that are able to communicate with each other and exchange information with the purpose to detect potentially dangerous situations. Realism-based modelling holds indeed much promises to increase data quality for devices collaborating within the Internet of Things [33].

Acknowledgement: This work was supported in part by Clinical and Translational Science Award NIH 1 UL1 TR001412-01 from the National Institutes of Health.

References

[1] Smith B, Ceusters W. Ontological realism: A methodology for coordinated evolution of scientific ontologies. Appl Ontol. 2010 Nov 15;5(3-4):139-88.
[2] Brochhausen M, Burgun A, Ceusters W, Hasman A, Leong TY, Musen M, et al. Discussion of "biomedical ontologies: toward scientific debate". Methods Inf Med. 2011;50(3):217-36.
[3] Arp R, Smith B, Spear AD. Building ontologies with basic formal ontology. Cambridge, Massachusetts: The MIT Press,; 2015.
[4] Smith B, Ashburner M, Rosse C, Bard J, Bug W, Ceusters W, et al. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. Nature Biotechnology. 2007 Nov;25(11):1251-5.
[5] Cimino JJ. Desiderata for controlled medical vocabularies in the twenty-first century. Methods of Information in Medicine. 1998;37(4-5):394-403.
[6] IHTSDO. International Health Terminology Standards Development Organization - SNOMED CT® Technical Implementation Guide - January 2015 International Release (US English). 2015. p. 757.
[7] Ceusters W, Elkin P, Smith B. Negative Findings in Electronic Health Records and Biomedical Ontologies: A Realist Approach. International Journal of Medical Informatics. 2007 March;76:326-33.
[8] Ceusters W, Smith B, Chiun Yu Hsu. Clinical Data Wrangling using Ontological Realism and Referent Tracking. Int. Conference on Biomedical Ontologies, ICBO 2014; Oct 6-9, 2014; Houston, TX.
[9] Hammond WE. The status of healthcare standards in the United States. International journal of bio-medical computing. 1995 Apr;39(1):87-92.

[10] Platt R, Wilson M, Chan KA, Benner JS, Marchibroda J, McClellan M. The new Sentinel Network--improving the evidence of medical-product safety. The New England journal of medicine. 2009 Aug 13;361(7):645-7.
[11] Ball R, Robb M, Anderson SA, Dal Pan G. The FDA's sentinel initiative--A comprehensive approach to medical product surveillance. Clinical pharmacology and therapeutics. 2016 Mar;99(3):265-8.
[12] FitzHenry F, Resnic FS, Robbins SL, Denton J, Nookala L, Meeker D, et al. Creating a Common Data Model for Comparative Effectiveness with the Observational Medical Outcomes Partnership. Appl Clin Inform. 2015;6(3):536-47.
[13] Califf RM. The Patient-Centered Outcomes Research Network: a national infrastructure for comparative effectiveness research. N C Med J. 2014 May-Jun;75(3):204-10.
[14] Ross TR, Ng D, Brown JS, Pardee R, Hornbrook MC, Hart G, et al. The HMO Research Network Virtual Data Warehouse: A Public Data Model to Support Collaboration. EGEMS (Wash DC). 2014;2(1):1049.
[15] Souza T, Kush R, Evans JP. Global clinical data interchange standards are here! Drug Discov Today. 2007 Feb;12(3-4):174-81.
[16] Garza M, Del Fiol G, Tenenbaum J, Walden A, Zozus M. Evaluating common data models for use with a longitudinal community registry. J Biomed Inform. 2016 Oct 28.
[17] Ogunyemi OI, Meeker D, Kim HE, Ashish N, Farzaneh S, Boxwala A. Identifying appropriate reference data models for comparative effectiveness research (CER) studies based on data from clinical information systems. Medical care. 2013 Aug;51(8 Suppl 3):S45-52.
[18] Hersh WR, Weiner MG, Embi PJ, Logan JR, Payne PR, Bernstam EV, et al. Caveats for the use of operational electronic health record data in comparative effectiveness research. Medical care. 2013 Aug;51(8 Suppl 3):S30-7.
[19] Rijnbeek PR. Converting to a common data model: what is lost in translation? : Commentary on "fidelity assessment of a clinical practice research datalink conversion to the OMOP common data model". Drug Saf. 2014 Nov;37(11):893-6.
[20] Yoon D, Ahn EK, Park MY, Cho SY, Ryan P, Schuemie MJ, et al. Conversion and Data Quality Assessment of Electronic Health Record Data at a Korean Tertiary Teaching Hospital to a Common Data Model for Distributed Network Research. Healthc Inform Res. 2016 Jan;22(1):54-8.
[21] Huser V, Cimino JJ. Desiderata for healthcare integrated data repositories based on architectural comparison of three public repositories. AMIA Annual Symposium proceedings / AMIA Symposium AMIA Symposium. 2013;2013:648-56.
[22] Scheuermann RH, Ceusters W, Smith B. Toward an ontological treatment of disease and diagnosis. Summit on translational bioinformatics. 2009;2009:116-20.
[23] Hogan WR, Ceusters W. Diagnosis, misdiagnosis, lucky guess, hearsay, and more: an ontological analysis. J Biomed Semantics. 2016 Sep 15;7(1):54.
[24] Bandrowski A, Brinkman R, Brochhausen M, Brush MH, Bug B, Chibucos MC, et al. The Ontology for Biomedical Investigations. PLoS One. 2016;11(4):e0154556.
[25] Ceusters W, Smith B. Strategies for referent tracking in electronic health records. J Biomed Inform. 2006 Jun;39(3):362-78.
[26] Overhage JM, Ryan PB, Reich CG, Hartzema AG, Stang PE. Validation of a common data model for active safety surveillance research. J Am Med Inform Assoc. 2012 Jan-Feb;19(1):54-60.
[27] Hripcsak G, Duke JD, Shah NH, Reich CG, Huser V, Schuemie MJ, et al. Observational Health Data Sciences and Informatics (OHDSI): Opportunities for Observational Researchers. Stud Health Technol Inform. 2015;216:574-8.
[28] Observational Medical Outcomes Partnership. OMOP Common Data Model Specification – Version 5. October 14, 2014. p. 69 pages.
[29] Reisinger SJ, Ryan PB, O'Hara DJ, Powell GE, Painter JL, Pattishall EN, et al. Development and evaluation of a common data model enabling active drug safety surveillance using disparate healthcare databases. J Am Med Inform Assoc. 2010 Nov-Dec;17(6):652-62.
[30] Ceusters W, Smith, B. Strategies for referent tracking in electronic health records. Journal Biomedical Informatics. 2006;39(3):362-78.
[31] Smith B, Ceusters W, Klagges B, Kohler J, Kumar A, Lomax J, et al. Relations in biomedical ontologies. Genome Biology. 2005;6(5).
[32] Blaisure J, Ceusters W. Business Rules to Improve Secondary Data Use of Electronic Healthcare Systems Informatics for Health 2017; 24-26 April 2017.; Manchester, UK 2017. (in press).
[33] Ceusters W, Bona J. Ontological Foundations for Tracking Data Quality through the Internet of Things. Stud Health Technol Inform. 2016;221:74-8.