Improving the 'Fitness for Purpose' of Common Data Models through Realism Based Ontology

Jonathan C. Blaisure, BS, MS, PhD student[1,2], Werner M. Ceusters, MD[1,2]
[1]Institute for Healthcare Informatics, University at Buffalo, Buffalo, New York, USA;
[2]Department of Biomedical Informatics, University at Buffalo, Buffalo, New York, USA;

Abstract

Common data models are designed and built based on requirements that are aimed towards fitness for purpose. But when common data models are used as lenses through which reality is observed from the perspective according to which they are built, then they exhibit restrictions that distort such view. Realism-based ontology design, when done properly, does not have these limitations as its fitness for purpose is only determined by the degree to which reality is represented the way it is. Therefore, we can use the principles that realism-based ontologies adhere to, not only to design application ontologies serving some specific purpose, but also to assess whether and where common data models fall short in their representational adequacy and how they can be corrected. If a realism based ontological perspective on the portion of reality the some common data model is trying to represent is compared with the perspective of the common data model itself, it is possible to determine how the latter deviates from the former and to suggest solutions to correct the misrepresentations found. Applying this method to the common data model of the Observational Medical Outcomes Partnership, revealed two major categories of errors: one where relationships are restricted based on the constraints of the data model, and one where the representation of reality is oversimplified.

Introduction

The University at Buffalo's Institute for Healthcare Informatics' (IHI) has as primary function the aggregation of disparate distinct data sources such as in-patient, out-patient, claims and clinical study datasets into a centralized integrated data repository (CIDR). It is a design criterion of the IHI to create this resource for healthcare data in maximal compliance with the principles of Ontological Realism[1, 2] insofar doing so does not interfere with (1) the need for timely access and (2) the availability of resources. This is done in the spirit of not letting the perfect come in the way of the good. The IHI realizes this function by providing a secure environment where data analysis, cohort discovery and other secondary data use requirements can be evaluated and accomplished. Although the production of a CIDR is the primary principal mission of the IHI, the researchers at the University at Buffalo (UB) have a need for advanced analytics in the interim. One pathway forward is to adopt a Common Data Model (CDM).

Traditional CDMs are designed around a 'fitness for purpose' paradigm according to which the data are organized in a way that solves specific organizational necessities thereby allowing portability and integration or federation of other datasets. Some of these requirements can dictate the use of specific implementation designs. For example, Informatics for Integrating Biology and the Bedside (i2b2)[3] proposes a CDM that is designed specifically for *cohort discovery*. i2b2 achieves this by using a single fact table to represent observations about patients. The fact table is dimensionally described by what are called '*dimension tables*' – patient dimensions, provider dimensions, and encounter dimensions are examples. This data model implements a paradigm which is known as a *star schema* database. The i2b2 star schema database as applied to integrated data coming from, for example, electronic healthcare records (EHR) is highly efficient in identifying patient cohorts using inclusion and exclusion criteria in queries which run over observations in the fact table. Data coming from EHRs undergo a complex mapping process that makes use of standardized terminologies (or in-house grown terminologies) and which applies to each row that is loaded into the *observation_fact* table. In that way, the assertion in the EHR becomes transformed into what is called a 'fact' as perceived through the data model although, of course, what is the case in reality might be different from what is stated to be a 'fact'. Users can browse these terminologies and use them to identify in the data collection patient cohorts composed out of patients about whom a specific observation which maps to the specific terms of interest is asserted (or not). The determination of 'fitness for purpose' of i2b2 is to provide a high-speed query system to recognize patient cohorts for (mostly) clinical trials. Although i2b2 solves this particular obstacle elegantly, it falls short, in our opinion, when 'fitting' it to other evaluation criteria specifically those imposed by realism based ontology (RBO).

i2b2's data model is probably not the only one that falls short of the criteria imposed by RBO. Further examples of CDMs include the Observational Medical Outcomes Partnership (OMOP)[4], the Patient-Centered Outcomes Research Network (PCORnet)[5], the healthcare management organizations' research network (HMORN) virtual data warehouse[6] and the Study Data Tabulation Model (SDTM) of the Clinical Data Interchange Standards Consortium (CDISC)[7]. Several of these CDMs have been subjected to studies for their 'fitness for purpose' for storing data extracted from electronic medical records (EMRs) specifically for the purpose of secondary data use in research. The OMOP pilot program, for example, terminated June 2013, but development still proceeds at the Observational Health Data Sciences and Informatics (OHDSI)[9] collaborative. OHDSI has released version 5.0 of the CDM and has helped develop a software tooling chain to facilitate extract, transform and load (ETL) processes from a diversity of source systems. These tools are designed to aid mapping of data sources to terminologies, loading source data into CDM-compatible data repositories, and data analysis on these repositories. The intent of the CDM is thus to provide a 'common model' for data coming from all healthcare information systems to be transformed and loaded into CDM-compatible data stores for the purpose of research, analytics, and data integration. It is intended to do this with minimal transformations and data loss. We chose the OMOP CDM as our intermediary data model for (1) the way it can deal with what is generally called 'findings'[8, 9], (2) the variety of open source analytical tools available, and (3) its wider purpose. It has also been qualified in some studies as the 'least lossy approach' among several CDMs tested[8].

The work described here is the result of scrutinizing the OMOP CDM from a RBO perspective specifically on the ways the data model presents a distorted view on the reality of the world it is referencing. The primary principle of comparison is restrictions based on the 'fitness for purpose' of the CDM and how those restrictions inhibit referencing reality adequately. Several publicly available ontologies, and, perhaps more importantly, the principles that they adhere to, were used as references including the Basic Formal Ontology version 2 (BFO2) [10], the Information Artifact Ontology (IAO)[11] and the Ontology for General Medical Sciences (OGMS)[12].

Methods

We followed the approach outlined in ontological realism[13] which takes very serious the distinction between data and data models on the one hand versus what the data and data models are about on the other hand. This allows us then to determine the differences between the CDM's 'fitness for purpose' versus the 'fitness to reality'. One can for example, metaphorically, view the data model as a container with a defined shape and size and built out of a certain material that restricts in certain ways what it can be filled with. In the case of the CDM, the restrictions are brought about by the structure of the tables, the cardinality of relationships, and the constraints implemented in the model. The 'shape' and 'size' of the container determine the qualities of that model – how well the model represents reality and the technical requirements or the 'fitness for purpose'. The structure of a data model, specifically in this case, a relational data model, is a lens through which one can view certain portions of reality (PoR) – sometimes exactly the way they are, sometimes, as we will demonstrate, not without distortions – while others are shielded off. According to the RBO-perspective, PoRs are composed of types (PERSON, ROLE, PROCESS – types are standardly written in SMALL-CAPS, while particulars are written in *italics*) and *particulars* – instances of types that carry identity (the two authors of this paper are both particulars which instantiate the type PERSON; with respect to the work presented in this paper, they each had particular roles each one of which was an instance of ROLE, and so forth). Types relate to other types by virtue of the way all particulars of these distinct types relate to each other. Such relationships between types can be expressed by axioms in a variety of formal ways. For example, in the Basic Formal Ontology (BFO)[10] an axiom stating that an EXTENDED ORGANISM[14] *Isa* MATERIAL ENTITY[10] (relationships between types are written in italics, while relationships from particulars are written in bold) is an axiom about ***all*** instances of EXTENDED ORGANISM. Thus if a particular *John Doe* is an **Instance-Of** EXTENDED ORGANISM then *John Doe* is also **Instance-Of** MATERIAL ENTITY. Also *John Doe's Height* is a particular which carries identity and which **Inheres-In** *John Doe*. This *John Doe's Height* is a particular quality and is **Instance-Of** the universal QUALITY. It is not an **Instance-Of** the universal EXTENDED ORGANISM.

The OMOP CDM's 'fitness for purpose' is '*to accommodate data from the observational medical databases that are generally considered necessary for active safety analysis'* thereby being *'analyst-friendly*' to meet the requirement to '*allow the analytic methods to execute quickly enough to be practical*'[15, p55]. On the other hand, the RBO-perspective is purpose independent, with the exception of reflecting the structure of reality[13]. Therefore, it is hypothesized that the OMOP CDMs 'fitness for purpose' limits the OMOP-perspective to a reductionist representation of reality and thus that the fields used in the tables deviate from realist types. Or in other words: comparing the two perspectives might lead to a conclusion that the OMOP view is an oversimplification (reductionist view) of or an unfaithful (deviant) view to reality.

The PoR which is represented by means of the OMOP perspective can also be represented by means of an RBO view by using RBO compliant ontologies such as the *Basic Formal Ontology (BFO)*[10], the *Information Artifact Ontology (IAO)*[16], the *Ontology for Biomedical Investigations (OBI),* and the *Ontology for General Medical Science (OGMS)*[14]. These ontologies, and the principles that they adhere to, can be used to provide context in the comparison of the RBO-perspective vs the OMOP-perspective. By comparing these two perspectives we can qualify the accuracy of representation to reality and understand why and in what way certain design restrictions distort the representation of reality. The OMOP-perspective should primarily reference types and relationships amongst types so that the data which are stored in OMOP-compatible data repositories represent relationships between particulars in exactly the same way the RBO-perspective would reference particulars, relations between those particulars, and relations between these particulars and types.

From the RBO-perspective, relational data models, including those used in EMRs, practice management systems, and CDMs are composed of individual parts – tables, relationships, columns and so forth. Under an RBO-perspective, these components are INFORMATION CONTENT ENTITIES (ICE)[17]. A particular ICE **Is-About** some entity in reality. For example, a patient medical record number (MRN) is an ICE which **Is-About** some *person* with a *patient role* that **Inheres-In** that *person*. A *PatientID* column in a relational database based on the OMOP CDM represents the type PERSON whereby each particular cell of that column represents an **Instance-Of** PERSON. On the other hand, a *diagnosis* is an ICE that references some *output* of a *clinical diagnosis process* that **Is-About** a *disease,* while a *disease* **Inheres-In** some *person.*[14, 18] This may seem trivial – or perhaps overly complicated – to a clinician at the point of care, but the distinction is important to accurately represent reality: the *diagnosis* and the *disease* are separate and distinct entities but are often represented in CDMs as the same entity.

With this in mind, we explored whether there are design principles that are standardly used in information modeling approaches that may have had a negative impact on the implementation and development of the CDM. Our process took into consideration the 'fitness for purpose' as well as various projects' conformity to the CDM as described in the literature. Looking with the eye of a realist ontologist to the 'fitness for purpose', requirements and design goals of the OMOP CDM may provide insight to where the model references reality objectively and adequately, and where it falls short. For example, a data model that has a requirement to conform to a certain structure to allow business intelligence (BI) tools to be able to ingest and query the data can be a limiting design restriction in itself. BI tools are designed in certain ways for maximum efficiency and to answer queries of a specific kind. A component-based[19] ecosystem has been developed around specific versions of the OMOP CDM and the desire to take maximally advantage of the BI tools it brings with it provides an incentive to not stray away from the original design guidelines. The OMOP CDM has had noteworthy improvements from release to release but drastic changes to the CDM would cause decisions to be made about either, the immediate redesign of software – to bring it up to date with the changes in the current version of the CDM – or the acceptance of being out of version compliance. In fact, one of the design principles of the OHDSI consortium is '*Backwards Compatibility*'[20]. Additional issues to contemplate when designing a relational data model are the normalization of data – data duplication, restricting the number of joins required to traverse the data. Joins are expensive, and building constraints that provide data consistency but does not constrain the model in a way that clashes with the 'fitness for purpose' is a challenging problem.

We began this comparison by downloading the data definition language (DDL) for the OMOP CDM version (v5) from the GitHub repository maintained by the Observational Health Data Sciences and Informatics (OHDSI) collaborative[21]. The DDL scripts are instructions on how to create the tables, relationships, indices and constraints that makeup the OMOP CDM and were loaded into a locally installed PostgreSQL server. After these scripts finished running, we ran additional scripts on the database for the purpose of creating documentation such as an entity relationship diagram (ERD) that visually represents a relational database schema. ERDs allow the visual inspection of tables, their types, and the relationships to other tables (constraints) by means of table and field descriptions, cardinality of relationships, etc. Afterwards, we downloaded the documentation supplied by the CDM v5[22] and used it to compare the ERD with the purpose to derive the informal semantics of the OMOP CDM, specifically the lens representing OMOP CDMs PoR used to view entities and their relationships – the 'OMOP-perspective'. We then queried PubMed to identify information pertaining to source data conversion into the OMOP CDM to proactively avoid problems reported in the literature. We performed an analysis to decisively compare common downsides and gained knowledge from other organizations experiences and perspectives to improve our approach.

Results and discussion

We have identified thus far two different ways in which the OMOP CDM design goals conflict with an RBO-perspective on the data represented in OMOP CDM-compatible data stores (Table 1). One design principle of the CDM is reported as follows: '*The CDM aims at providing data organized in a way optimal for analysis, rather than for the purpose of operational needs of health care providers or payers*'[20]. Operational needs and data analysis needs differ indeed. Data normalization in operational systems is for instance focused on making transaction speeds satisfy the requirements of the operational environment, which in the case of EMRs comes down to the ability of quickly entering and retrieving data about a single patient. This constitutes a relatively small amount of data with respect to the totality of data hold over all patients in the EMR system so that the search space is quite small as well. Analyst queries on the other hand have to run over very large amounts of data while also returning large result sets. Although speed is an issue, it is not as severe as in EMR systems. For example, waiting hours for an analytics question to be answered using the data in a secondary use data store is for sure annoying, but does not need to disrupt the workflow of the data analyst. But it would be unacceptable to have the query run for days and weeks, what would be the case if a typical analytics question would be run over the back-end of the EMR system itself, rather than over the secondary use data store of which the model is optimized to handle these kinds of queries. By examining the CDM, specifically the *Person-table, Observation-table, Provider-table* and *Location-table* we start to see where some conflicts with the RBO-perspective as brought about by this sort of optimization arise from.

Table 1 – Identified problems.

Type of Problem	Description	Example
Cardinality	A problem where relationships are (incorrectly) restricted based on the constraints of the data model.	A person's address can change over time.
Reductionist	A problem where the representation of reality is over simplified.	A provider is a role that a person bears.

The *Person-table* is composed of many fields which hold data with specific data types, as for instance in the *person_id-column*. The OMOP specification documents this field as being *'A unique identifier for each person'*[23]. From the RBO-perspective we would phrase this that in relation to what occupies a specific cell in the *person_id-column* there exist a particular *person_id* which is an **Instance-Of** PERSON_ID, this type being itself a subtype of ICE. That particular *person_id* **Is-About** (at some point in time) a unique particular that is an **Instance-Of** PERSON. What is stored in the *person_id-column* of an OMOP-compatible data store is then a **Concretization-Of** that particular *person_id*. That concretization has 'in the database on disk' probably the form of a specific pattern of magnetization points. When that *person_id* is concretized on paper, it is most likely in the form of a (alpha-)numerical string. The paper can be destroyed separately from the database, or even both can be destroyed. But that would not result in the *person_id* to be destroyed, and for sure not the person about which this is the ID. There is in the RBO-perspective no 'death through nullification' for which the now abandoned HL7 RIM was once critiquized[24].

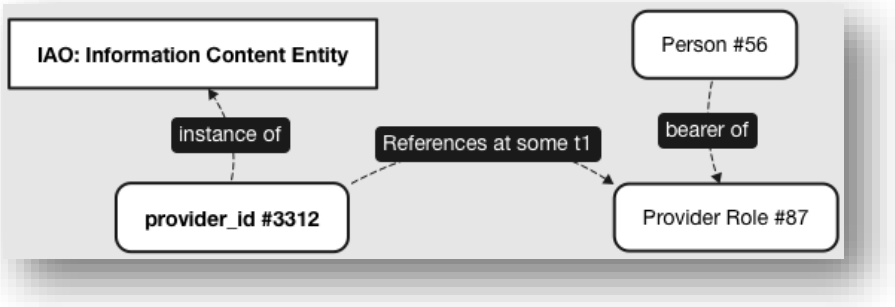

Figure 1. RBO-perspective Person vs Provider

On the other hand, examination of the *Provider-table* documentation reveals that the column *provider_id-column* is defined as '*A unique identifier for each Provider'* [25]. Under the RBO-perspective, a PERSON is a subtype of MATERIAL ENTITY which itself is a subtype of INDEPENDENT CONTINUANT while a PROVIDER is a subtype of ROLE which itself is a subtype of REALIZABLE ENTITY that **Inheres-In** a PERSON. PROVIDER is trivially not the same type as PERSON nor is it subsumed by PERSON. Figure 1 illustrates these relationships from the RBO-perspective (rectangles with rounded corners represent instances while rectangles with square corners represent types). The question now is what exactly is meant in the OMOP CDM with 'provider': should this be interpreted as meaning 'provider role' or 'person which has a provider role'? And it should make us wonder whether with 'person', OMOP really means what we typically understand under that term, or whether they mean 'patient'. Under the interpretation that the provider table is a lens that captures references to instances of PROVIDER ROLE but not to the particular person that **bears** the PROVIDER ROLE, one would expect the persons that bears that role to be represented in the person table as well. Although a valid distinction, the goal of this comparison is to compare the implementation of the CDM to its 'fitness for purpose'. Arguably a simple data analysis question: 'How many unique entities of type PERSON are referenced in the CDM?' would return too small a number if PERSONS who ***bear*** a PROVIDER ROLE are not included in the person table and only entries in the person table would be counted. On the other hand, if the same question would be answered by a query that returns the total number of both entries in the person table and entries in the provider table, then this would result in a number that is too high if there are persons represented in the dataset that have both a PATIENT ROLE and a PROVIDER ROLE. This is deviant of the RBO-perspective but also deviates from the OMOP-perspective's 'fitness for purpose' with that purpose being accurate *analysis of data*. Accuracy in this case is thus, as we assume, only expected for certain types of questions that according to the designers are from their perspective relevant to be asked, and not for all types of questions that can be asked over exactly the same domain: thus the two questions asked, although valid from a realist perspective, are not supposed to be asked to OMOP CDM-compatible data repositories.

The root cause of the problem just sketched can be described as a *confusion of* types. A possible solution to this mis-representation of reality used in the above example may be addressed by expanding the CDM to include a *Person-table, Provider-Role-table* and a *Patient-Role-table*. These tables would distinctly identify the realization of patient and provider roles throughout the dataset and allow unique and accurate counting of instances of type PERSON. Constraining each role by a start and end timestamp should be used to temporally qualify roles linearly – allowing deeper analysis and thus increasing the CDMs 'fitness' (figure 2).

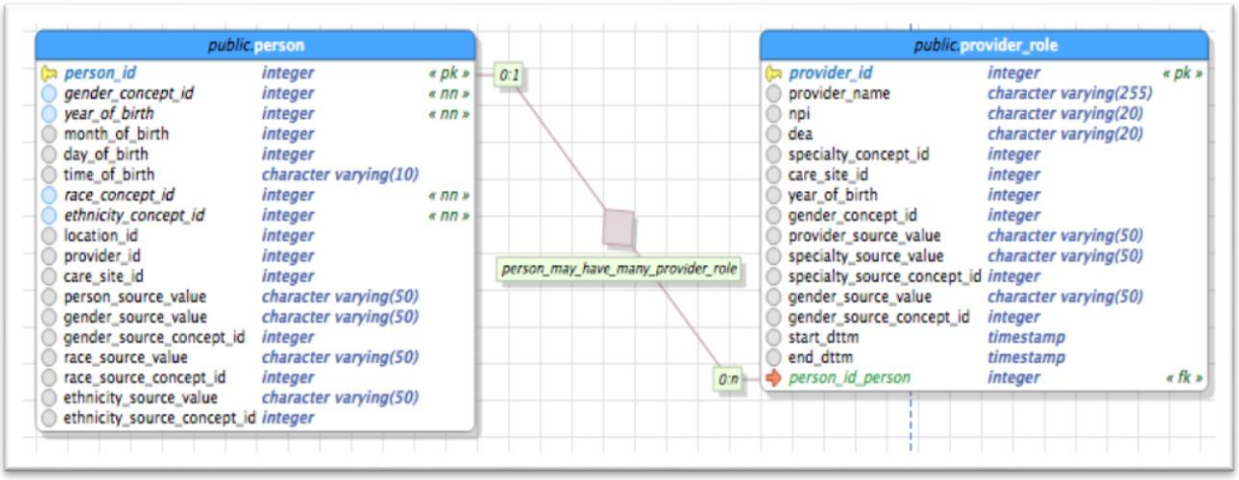

Figure 2. Provider Roles Example Solution (representation is not complete)

Another problem discovered is the opposite of the previous one. While the previous problem had to do with incorrect counting of particulars due to incorrect representation of uniqueness and confusion of types, this one has to do with cardinality constraints existing in the CDM. We have examined the *Location-table* and the *Person-table* to compare it with the RBO-perspective. As displayed in figure 3, the *Person-table* has a column named *location_id* which is defined as '*A foreign key to the place of residency for the person in the location table, where the detailed address*

information is stored'[23]. The model's definition (DDL) is designed in a way that will only allow one location to be assigned to a person at any given time. This is so is because of the structure of the data model which is designed in such a way that a row in the person table can only have one reference to a row in the location table (Figure 3). This type of *one-to-one* relationships form the least costly joins that can be made in a relational database.

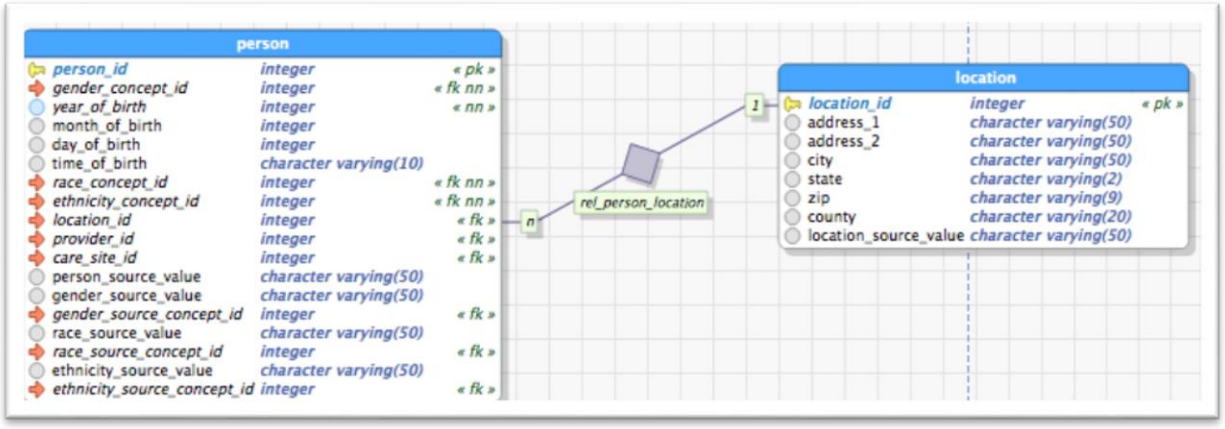

Figure 3 – Relationship between Person and Location table

From the RBO-perspective, relationships between particulars – the main relata – of which one is of type CONTINUANT always involve an extra particular which is an instance of TEMPORAL REGION. Assertions about such relationships should thus contain references to all three particulars: to both main relata and to the temporal region during which the relationship between the two particulars obtains. Entries in databases in which foreign keys are used as described above to indicate the address of a person, both persons and addresses being continuants, qualify as assertions about relationships that hold over a period of time. However, in the OMOP CDM, only two of the three required particulars are referenced: there is a reference to, say, *Person 1* and to the residency of *Person 1* (in this case an address – *Address A*) but there is no reference to the temporal region. Of course, at some point in time, the location_id may need to reference another address when *Person 1* moves to another city or even across town – *Address B* at some other time. A secondary data use question such as "How many instances of type PERSON could have been exposed to pollution from this *water source*?" would only include whichever single *location* is associated with that particular PERSON and not consider all the places the person lived in the past. The OMOP CDM documentation does acknowledge that patients over time can have distinct locations, genders, etc., but '*it is the responsibility of the data holder to select the one value to use in the CDM*'[22, p37]. A possible solution to correct the representation of reality in this case could be to change the cardinality of the *Person-table* to *Location-table* to allow multiple relationships between a person and a location. This can be accomplished by using a bridging table. In the example below we have added a *start_dttm* and *end_dttm* (figure 4) column to the *location-table* so we can provide a TEMPORAL REGION reference. With this addition, a person's locations can be tracked over time representing reality more accurately, and providing a stronger use case for secondary data use. The current CDM structure does not limit introducing a new location into the database, but it does not allow creating a historical transaction. If a person changes residency, the location of that residency would change removing the previous location and no record of this transaction would exist.

This is a reductionist problem involving *temporality*. Relationships amongst particulars one of which is a continuant obtain in some temporal region and it is important for the model to be able to capture this feature of reality. We can analyze other fields using this same logic such as the *Gender-column* which shows the exact same problem – gender can change over time and this reality is important for secondary data use and accurately representing reality.

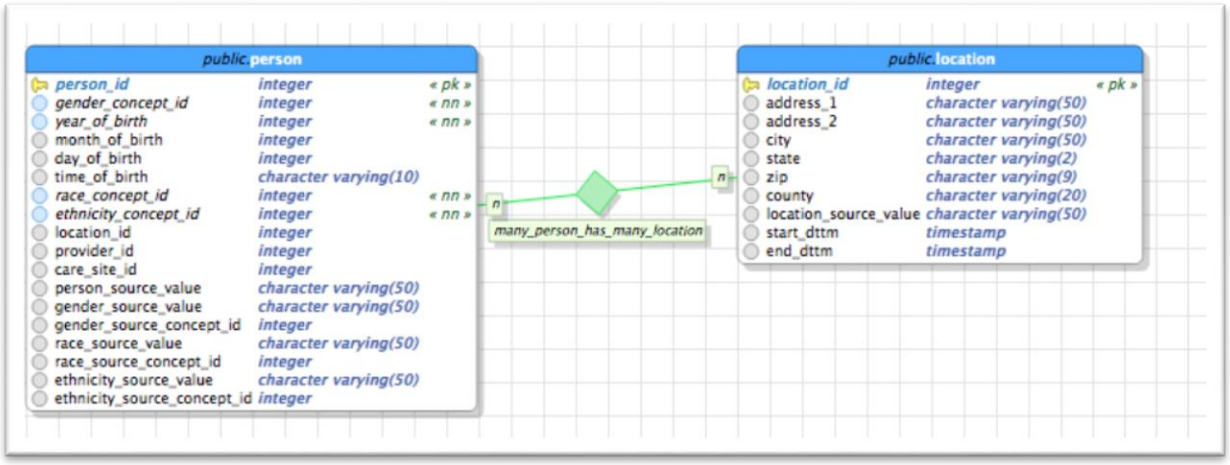

Figure 4. Person - Location relation with cardinality – An example Solution (representation is not complete)

Conclusion

It is our belief that we can compare data model designs to designs based on ontological realism to not only show their restrictions but improve their 'fitness for purpose'. Through an RBO-perspective we have identified thus far two examples of misrepresentation of reality – the confusion of types and cardinality problems related to temporal regions around particulars. Some may argue against the importance of these misrepresentations but as realism based ontologists we argue that representing data accurately to reality directly affects real world secondary data use requirements [26]. By applying realism based ontology, we can ultimately increase the 'fitness for purpose' of data models and their respective requirements. As such, a practical implementation of RBO-perspectives can be incorporated into the design of data models improving their accuracy to represent reality. Many aspects of the RBO field have been theoretical in the past and thus, have not been applied in a practical way to existing CDMs. We believe that the maturity of the work that is being done in the RBO fields, specifically in the biomedical domains, and the advances in database technologies, presents a unique opportunity to develop the next generation of data models.

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. Ceusters W, Chiun Yu Hsu, Smith B. Clinical Data Wrangling using Ontological Realism and Referent Tracking. CEUR Workshop Proceedings. 2014;1237:27-32.
2. Smith B, Ceusters W. Ontological realism: A methodology for coordinated evolution of scientific ontologies. Appl Ontol. 2010;5(3-4):139-88.
3. Kohane IS, Churchill SE, Murphy SN. A translational engine at the national scale: Informatics for integrating biology and the bedside. Journal of the American Medical Informatics Association. 2012;19(2):181-5.
4. 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.
5. Califf RM. The Patient-Centered Outcomes Research Network: a national infrastructure for comparative effectiveness research. N C Med J. 2014;75(3):204-10.
6. 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 (Washington DC). 2014;2(1):1049.

7. Souza T, Kush R, Evans JP. Global clinical data interchange standards are here! Drug Discov Today. 2007;12(3-4):174-81.
8. Garza M, Del Fiol G, Tenenbaum J, Walden A, Zozus MN. Evaluating common data models for use with a longitudinal community registry. Journal of Biomedical Informatics. 2016;64:333-41.
9. 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;51(8 Suppl 3):S45-52.
10. Arp R, Smith B, Spear AD. Building ontologies with basic formal ontology. Cambridge, Massachusetts: The MIT Press; 2015.
11. Ceusters W. An information artifact ontology perspective on data collections and associated representational artifacts. Studies in health technology and informatics. 2012;180:68.
12. Scheuermann RH, Ceusters W, Smith B. Toward an ontological treatment of disease and diagnosis. Summit on translational bioinformatics.
13. Smith B, Ceusters W. Ontological realism: A methodology for coordinated evolution of scientific ontologies. Applied Ontology. 2010;5(3-4):139-88.
14. Scheuermann RH, Ceusters W, Smith B. Toward an ontological treatment of disease and diagnosis. Summit on translational bioinformatics. 2009;2009:116-20.
15. 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;19(1):54-60.
16. Ceusters W. An information artifact ontology perspective on data collections and associated representational artifacts. Stud Health Technol Inform. 2012;180:68-72.
17. Smith B, Ceusters W. Aboutness: Towards Foundations for the Information Artifact Ontology. International Conference on Biomedical Ontology; July 27-30; Lisbon, Portugal2015. p. 47-51.
18. Hogan WR, Ceusters W. Diagnosis, misdiagnosis, lucky guess, hearsay, and more: an ontological analysis. J Biomed Semantics. 2016;7(1):54.
19. Vitharana P. Risks and challenges of component-based software development. NEW YORK: ACM; 2003. p. 67-72.
20. OMOP CDM – Design Principles 2015 [OMOP Common Data Model]. Available from: http://www.ohdsi.org/web/wiki/doku.php?id=documentation:cdm:design_principles.
21. 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.
22. Observational Medical Outcomes Partnership. OMOP Common Data Model Specification – Version 5. October 14, 2014. p. 69 pages.
23. cubarkthik. OMOP CDM – PERSON table 2017 [Available from: http://www.ohdsi.org/web/wiki/doku.php?id=documentation:cdm:person.
24. Smith B, Ceusters W. HL7 RIM: an incoherent standard. Stud Health Technol Inform. 2006;124:133-8.
25. cgreich. OMOP CDM – PROVIDER table 2014 [Available from: http://www.ohdsi.org/web/wiki/doku.php?id=documentation:cdm:provider.
26. Blaisure J, Ceusters W. Business Rules to Improve Secondary Data Use of Electronic Healthcare Systems. Informatics for Health 2017, Manchester Central, UK, April 24-26, 2017. Stud Health Technol Inform. 2017;235:303-307.