OMOP CDM compared to ContSys (ISO13940) to make data FAIR

Rowdy de Groot, MSc, rowdy.degroot@amsterdamumc.nl[1], Ronald Cornet, PhD, r.cornet@amsterdamumc.nl[1], Nicolette de Keizer, PhD, n.f.keizer@amsterdamumc.nl[1], Nirupama Benis, PhD, n.benis@amsterdamumc.nl[1], Ferishta Raiez, PhD, f.raiez@amsterdamumc.nl[1]
[1]Department of Medical Informatics, Amsterdam UMC, Location AMC, Amsterdam Public Health research institute, Amsterdam, The Netherlands

Abstract

Quality registries contain valuable information for research. However, every registry uses its own data structure and data definitions. This makes the data hard to share with other organizations. The data should support a wider use and should therefore be represented in a common information model. OMOP CDM and ContSys (ISO13940), commonly used information models, were compared with regard to making quality-registry data FAIR.

A sub dataset of a quality registry was represented both in OMOP CDM and ContSys. OHDSI provides tools and guides, ContSys does not. All source values were represented with SNOMED CT for ContSys, as SNOMED CT has a close alignment with ContSys. OMOP CDM was mapped to ContSys to determine the interoperability between the two models.

All of the necessary data items could be represented in both models and 94.6% of the data values could be represented in OMOP CDM and 93.5% in ContSys. Of the two models, OMOP CDM was the easier model to map. The OMOP domains and columns are more specific than the clauses and concepts from ContSys. This made it much clearer to which OMOP attribute a data item should map. Due to the lack of guidance and general concepts in ContSys, different users that map the same dataset could represent the dataset differently in ContSys. This is far more unlikely to happen in OMOP CDM. OMOP CDM is therefore more viable to use for FAIR data. The OMOP to ContSys mapping showed that the models are interoperable from one organization to another.

Research Category (please highlight or circle which category best describes your research)

Methodological research

Introduction/background

With the increasing need to improve the infrastructure supporting the reuse of data, more attention is given to make data FAIR (Findable, Accessible, Interoperable, Reusable)[1]. The information model is an important component in the process of making data FAIR. It is a representation of concepts, relationships, constraints, rules and operations that are used to specify data semantics in a certain domain[2]. Information models clearly define the data items in the database and the relationships between them. Therefore, information models are useful for interoperability (the 'I' of FAIR) between databases.

Two generally used information models are OMOP CDM and ContSys (ISO13940). It is not clear whether both models are applicable to observational data, for example for observational data from an ICU quality registry.

NICE (National Intensive Care Evaluation) is a Dutch quality registry that collects data on Intensive Care Unit (ICU) admissions from all the ICUs in the Netherlands[3]. Quality registries such as NICE collect valuable information for research. However, every registry uses its own data structures and definitions. This makes it hard to use the data outside of the registry. NICE intends to make their data FAIR to support wider use of the data, not only nationally for the participating Dutch ICUs but also internationally with other quality registries. Therefore, NICE data should be represented using a common information model. OMOP CDM and ContSys (ISO13940) were compared as options to make NICE data FAIR.

ISO13940 or ContSys is a conceptual model developed by the International Organization for Standardization (ISO). It defines a system of concepts for different aspects of the provision of healthcare. The model has been designed to include professional healthcare, self-care, care by a third party and all aspects of social care, so the model was intentionally kept broad[4]. Due to this broad scope, ContSys is devised to support all (health)care-related data. Because ContSys is a conceptual model it is not oriented towards implementation, in contrast to OMOP CDM.

The aim of this study was to compare OMOP CDM and ContSys (ISO13940) by applying both models to a dataset from an observational ICU quality registry and determining the interoperability between the two models.

Methods

We used a small subset of the NICE registry, the Severe Acute Respiratory Infections (SARI) set. It contains eighteen variables for physiological information, lab results and diagnoses in the first 24 hours after ICU admission.

The book of OHDSI was used to conduct the process of making the Extract Transform, Load (ETL) design for the OMOP CDM[5]. The tools White Rabbit, Rabbit-in-a-Hat and Usagi from OHDSI were used[6,7]. The contsys.org website[8] was used to guide the process of representing the SARI dataset items in ContSys. SNOMED CT has a very close alignment with ContSys. SNOMED CT represents clinical statements within a record while ContSys establishes the relationship within the healthcare domain[4]. Therefore, the decision was made to map all data values of the data items to SNOMED CT codes. ContSys does not provide any tools for this mapping process while OHDSI offers the Usagi tool.

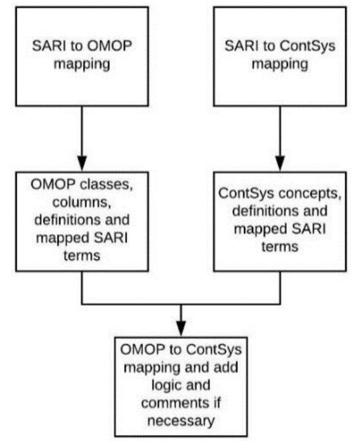

Figure 1. Process to map OMOP to ContSys to check the interoperability.

Models were compared on the amount of data items and data values that could be mapped. The models were also cross-mapped to evaluate the interoperability of the two models, see Figure 1.

Results

All of the necessary variables from the dataset were mapped to the OMOP CDM tables and columns in Rabbit-in-a-Hat. In Usagi 94.6% of the source data values could be mapped to a concept id. All of the source data items could also be represented in ContSys and 93.5% of the source values could be mapped. All of the used OMOP CDM fields were mapped to a ContSys concept. All of the source values that are represented with a concept ID from the vocabulary SNOMED CT in OMOP CDM are interoperable with ContSys.

Conclusion/discussion

Although a similar amount of source items and values could be represented in both models, the usability of the models differed much. Of the two models, OMOP CDM was the easier model to map. OHDSI provides a guidebook, forum, and tools to help users apply OMOP CDM[5]. In contrast, ContSys being a conceptual model, is not implementation-oriented. Consequently, ContSys does not provide any guidebook or tools to help use ContSys. The OMOP domains and attributes are also more specific than the clauses and concepts from ContSys. This made it much clearer to which OMOP attribute each data item should be mapped. Deciding which data item should be mapped to a certain concept was sometimes problematic in ContSys, because the definitions for the concepts are very general. However, an advantage of ContSys is that it can represent negative findings and OMOP CDM is not designed to do that. The OMOP to ContSys mapping made apparent that OMOP and ContSys are highly interoperable. Only data values that represent negative findings will be lost when mappings are made from ContSys to OMOP CDM.

Both models could be used to represent observational data from an ICU quality registry. However, different users that map the same dataset could represent it differently in ContSys due to the lack of guidance and the general concepts in ContSys. This is far more unlikely to happen in OMOP CDM. OMOP CDM is therefore far more viable to use for realizing FAIR data.

References

1. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. Scientific data. 2016;3.
2. Lee YT, editor Information modeling: From design to implementation. Proceedings of the second world manufacturing congress; 1999: International Computer Science Conventions Canada/Switzerland.
3. van de Klundert N, Holman R, Dongelmans DA, de Keizer NF. Data resource profile: the Dutch National Intensive Care Evaluation (NICE) registry of admissions to adult intensive care units. International journal of epidemiology. 2015;44(6):1850-h.
4. O N. Background of this system of concepts 2019 [cited 2020 20th of January]. Available from: https://contsys.org/page/Background.
5. OHDSI. The Book of OHDSI 2019 [cited 2020 10th of January]. Available from: https://ohdsi.github.io/TheBookOfOhdsi/
6. OHDSI. WhiteRabbit 2020 [cited 2020 27th of January]. Available from: https://github.com/OHDSI/WhiteRabbit.
7. OHDSI. Usagi 2020 [cited 2020 27th of January]. Available from: https://github.com/OHDSI/Usagi.
8. O N. A system of concepts for the continuity of care 2019 [cited 2020 29th of January]. Available from: https://contsys.org/page/default.