

Article

Not peer-reviewed version

Conceptual Design of a Generic Data Harmonization Process for OMOP CDM

[Elisa Henke](#)*, Michele Zoch, Yuan Peng, Ines Reinecke, Martin Sedlmayr, Franziska Bathelt

Posted Date: 2 November 2023

doi: 10.20944/preprints202311.0104.v1

Keywords: OMOP; OHDSI; interoperability; data harmonization; clinical data; claims data



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

Conceptual Design of a Generic Data Harmonization Process for OMOP CDM

Elisa Henke ^{1,*}, Michele Zoch ¹, Yuan Peng ¹, Ines Reinecke ², Martin Sedlmayr ¹ and Franziska Bathelt ³

¹ Institute for Medical Informatics and Biometry, Carl Gustav Carus Faculty of Medicine, Technische Universität Dresden, 01307 Dresden, Germany; michele.zoch@tu-dresden.de (M.Z.); yuan.peng@tu-dresden.de (Y.P.); martin.sedlmayr@tu-dresden.de (I.R.)

² Data Integration Center, Center for Medical Informatics, University Hospital Carl Gustav Carus Dresden, 01307 Dresden, Germany; ines.reinecke@ukdd.de

³ Thiem-Research GmbH, 03048 Cottbus, Germany; f.bathelt@ctk.de

* Correspondence: elisa.henke@tu-dresden.de; Tel.: 0049351-458-87-7708

Abstract: To gain insight into the real-life care of patients in the healthcare system, data from hospital information systems and insurance systems are required. Consequently, linking clinical data with claims data is necessary. To ensure their syntactic and semantic interoperability, the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) was chosen. However, there is no detailed guide that would allow researchers to follow a consistent process for data harmonization. Thus, the aim of this paper is to conceptualize a generic data harmonization process for OMOP CDM. For this purpose, we conducted a literature review focusing on publications that address the harmonization of clinical or claims data in OMOP CDM. Subsequently, the process steps used and their chronological order were extracted for each included publication. The results were then compared to derive a generic sequence of the process steps. From 23 publications included, a generic data harmonization process for OMOP CDM was conceptualized, consisting of nine process steps: dataset specification, data profiling, vocabulary identification, coverage analysis of vocabularies, semantic mapping, structural mapping, extract-transform-load-process, qualitative and quantitative data quality analysis. This process can be used as a step-by-step guide to assist other researchers in harmonizing source data in OMOP CDM.

Keywords: OMOP; OHDSI; interoperability; data harmonization; clinical data; claims data

1. Introduction

The use of real-world data for research is becoming increasingly important in order to gain insights into the real-life care of patients in the healthcare system and, on this basis, to gain new knowledge for the diagnosis, treatment and prevention of diseases. To promote digitization in medicine for the areas of care and research in Germany, the Medical Informatics Initiative (MII) has been funded since 2018 by the German Federal Ministry for Education and Research (BMBF) [1]. The aim of the MII is to link data from patient care by providing digital infrastructures for the integration and harmonization of health data for research purposes. However, the developments of the infrastructures are currently focused on patient data of German university hospitals. Therefore, research in the MII is limited to clinical data from patients during hospitalization. The medical care of patients in university hospitals, in contrast, affects only a small percentage, since patients are usually hospitalized only when they already have a severe disease.

Green et al. [2] pointed out that a far greater number of patients are treated outside of the hospital. In comparison to inpatient data, outpatient data provide a more comprehensive overview of patients' medical histories. A relevant data source for outpatient data is claims data from the statutory health insurance funds in Germany. By linking claims data across institutions and sectors on a person-specific basis, longitudinal analyses of treatment histories can be realized. However, due to their billing focus, claims data lack depth of content, so that information on, for example, diagnostic

and laboratory data is not included. In order to integrate both detailed information on the respective inpatient stay of patients as well as the insured person-related course perspective for research with real-world data, the combination of clinical data with claims data is necessary.

To exploit the potential of linking clinical data with claims data, it is important to ensure the syntactic and semantic interoperability of both data sets. Achieving syntactic and semantic interoperability requires data harmonization in a standardized format [3]. For the unified representation of heterogeneous data sets, so-called standardized common data models (CDMs) are developed. In the last years, the Observational Medical Outcomes Partnership (OMOP) CDM of the Observational Health Data Sciences and Informatics (OHDSI) gained significant relevance for research with real-world data [4–8].

The main challenge for researchers is the harmonization of national and institution-specific terminologies, formats and structures into the standardized format of OMOP CDM. For this purpose, OHDSI introduces four major steps that should help to harmonize source data in OMOP CDM: Design the Extract-Transform-Load (ETL), Create the Code Mappings, Implement the ETL, Quality Control [9] (pp. 75-94). However, our own experience in harmonizing clinical data of the MII given in the Fast Healthcare Interoperability Resources (FHIR) format to OMOP CDM [10] has shown that these steps are not detailed enough. The literature demonstrates that many researchers are concerned with harmonization of source data in OMOP CDM [5]. Nevertheless, there is no detailed guidance that would allow researchers to follow a consistent process when transforming source data to OMOP CDM.

Thus, the aim of this paper is to conceptualize a generic data harmonization process for OMOP CDM. In this context, we focus on the following two research questions:

1. Which process steps need to be performed when harmonizing clinical data or claims data in OMOP CDM?
2. What sequence of identified process steps should be followed?

2. Materials and Methods

2.1. Literature Review

2.1.1. Paper Identification

To obtain a clearer perspective of the state of the art of methodological processes for data harmonization in OMOP CDM, we conducted a literature review on August 3, 2023. Our literature search included publications published in English between 2018 and 2023, focusing on the harmonization of clinical data or claims data in OMOP CDM. **Error! Reference source not found.** provides an overview of the search terms used in the literature databases PubMed and Web of Science.

The resulting publications were imported into the reference management software Zotero [11]. Afterwards, duplicates were removed using Zotero’s built-in duplicate detection feature.

Table 1. Search terms used for the literature search in PubMed and Web of Science.

Database	Search String
PubMed	((OMOP[Title/Abstract]) OR (OHDSI[Title/Abstract])) AND ((claims data[Title/Abstract]) OR (clinical data[Title/Abstract]))
Web of Science	((TI=(OMOP) OR TI=(OHDSI)) AND (TI=(claims data) OR TI=(clinical data))) OR ((AB=(OMOP) OR AB=(OHDSI)) AND (AB=(claims data) OR AB=(clinical data)))

2.1.2. Paper Exclusion

The process of paper exclusion consisted of a Title-Abstract-Screening and Full-Text-Screening performed by three reviewers (EH, FB, MZ). For this purpose, eight exclusion criteria (**Error! Reference source not found.**) were defined to categorize the excluded publications.

Table 2. Definition of exclusion criteria following to Reinecke et al. [5].

Criterion	Description of criterion
no "OMOP" or "OHDSI"	Publication does not mention "OMOP" or "OHDSI"
mentioned	Publication uses "OMOP" or "OHDSI" with other meanings
evaluated	Publication only mentions "OMOP" or "OHDSI"
vocabulary	Publication focuses on the evaluation of OMOP
extension	Publication focuses on vocabularies and their mapping in OMOP or use of OMOP vocabularies
usage	Publication focuses on an extension of OMOP or OHDSI tools
no full text	Publication focuses on the use of OMOP, e.g., for studies, data quality analyses, development of tools or frameworks (e.g., patient level prediction)
foreign language	Publication is not available as full text
wrong type of source data	Publication is written in other languages than English
	Publication focuses on types of source data other than clinical data or claims data

Next, all three reviewers performed the Title-Abstract-Screening for 20% of the publications found. Conflicts were then discussed and resolved. Afterwards, we utilized the kappa statistic to test the interrater reliability [12]. For this purpose, we used the Fleiss method [12] in the KappaM function of the R library *DescTools* [13]. For the analysis, an error probability of 5% was set. Depending on the result of the kappa value, we chose one of the two defined options for the further procedure of paper exclusion:

- Option 1: For a kappa value greater than 0.6 (substantial to almost perfect agreement (interpretation according to Fleiss [12])), the Title-Abstract-Screening for the remaining 80% of the publications found and afterwards, the Full-Text-Screening for the included publications should be divided as follows: 1) reviewer 1 (EH) should screen all publications; 2) reviewer 2 (FB) all included publications and 3) reviewer 3 (MZ) all excluded publications.
- Option 2: If the kappa value is less than or equal to 0.6 (poor to moderate agreement (interpretation according to Fleiss [12])), the remaining 80% of the publications and the full texts had to be screened by all three reviewers.

After both, Title-Abstract-Screening and Full-Text-Screening, conflicts between the reviewers were discussed and resolved.

2.1.3. Data Extraction

After the paper exclusion process, we focused on data extraction from the included publications. The data extraction was performed by reviewer 1 (EH) and subsequently verified by reviewer 2 (FB). The data extraction process consisted of three iterations. The first iteration focused on extracting the process steps from the publications used to harmonize the source data in OMOP CDM. In addition, the specification of the type of source data (clinical data and/or claims data) used for data harmonization and the countries from which the data was originated were documented for each publication.

In a second iteration, for each publication, we checked which of the extracted process steps were applied during the data harmonization in OMOP CDM. For this purpose, we created a matrix. The columns of the matrix represented the extracted process steps, while the rows represented the included publications. Within the matrix, we used crosses to indicate when a process step was mentioned in the corresponding publication.

In the final third iteration, the focus was on identifying the chronological order of the applied process steps per publication. For this purpose, we replaced the crosses in the matrix with ascending numbers. Afterward, the distribution of the given numberings per process step was calculated and the most frequent number was highlighted. This approach was performed for each of the publications that a) used clinical data as source data or b) used claims data as source data and c) for all publications

regardless of the type of source data. Publications that used both, clinical data and claims data, were categorized into all three groups.

2.2. Derivation of a Generic Sequence of Process Steps

In order to be able to apply the process steps to the harmonization of source data, it was necessary to establish a chronological classification of the process steps in an overall process. For this purpose, we compared the most frequent number(s) per process step. Our comparison started with the process step with the lowest number. Thereafter, we compared it to the number of the next process step. If the subsequent process step had a higher numbering (e.g., 1 vs. 2), both process steps remained in their position. In the case of two identical numberings (e.g., 1), the higher sum of the percentage of the most frequent numbering and the percentages of the preceding numberings decided the position in the comparison and all subsequent numberings were increased by the value 1. This method of comparing predecessor and successor numbers was used until a unique numbering and thus positioning in the overall process could be defined. The derivation of the sequence of process steps was done separately for groups a)-c) defined in section 2.1. Finally, the results of the three groups were compared with regard to their agreement to derive a generic sequence for the data harmonization process.

3. Results

3.1. Flow Diagram of the Literature Review

Based on the search terms, 162 publications (PubMed: 55, Web of Science: 107) were found. After removing duplicates in Zotero (34 publications), 128 publications remained for Title-Abstract-Screening. During the Title-Abstract-Screening, all three reviewers initially reviewed 20% (26/128) of the publications. Then we calculated the kappa value to check reviewer agreement. The kappa statistic resulted in a value of $\kappa = 0.764$ (substantial agreement). Since this value was greater than 0.6, we chose Option 1 (see section 2.1.2) for the further paper exclusion process. After Title-Abstract-Screening, 85 publications were excluded according to the definition of the exclusion criteria in **Error! Reference source not found.** In the subsequent screening of the full texts of 43 publications, we further excluded 20 publications. Finally, 23 publications were included for data extraction.

Error! Reference source not found. summarizes the process of the literature search, the subsequent screening of publications for inclusion and the remaining publications for the data extraction as PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses) flow diagram [15]. A detailed overview of the results of the literature search, including the data extraction matrix can be found in the Supplementary Materials “File S1”.

3.2. Process Steps

The 23 reviewed publications allowed us to extract a methodological process for data harmonization in OMOP CDM [16–38]. Clinical data were used in 18 of 23 publications (78%) [16,19–28,31–33,35,36,38,39], claims data were used in 3 of 23 publications (13%) [29,30,34], and both types of data sources were used in 2 of 23 publications (9%) [37,40] as source data for data harmonization in OMOP CDM. Clinical data originated from Belgium, Brazil, China, Denmark, Estonia, France, Germany, Italy, the Netherlands, Portugal, Serbia, Singapore, South Korea, Spain, Turkey, the United Kingdom and the United States. Claims data were used from Austria, France, and the United States (see Supplementary Material S1).

The methodological process extracted from the publications consists of the following nine process steps (in alphabetical order):

- Coverage analysis of vocabularies
- Data profiling
- Dataset specification
- Extract-Transform-Load (ETL)-process
- Qualitative data quality analysis

- Quantitative data quality analysis
- Semantic mapping
- Structural mapping
- Vocabulary identification

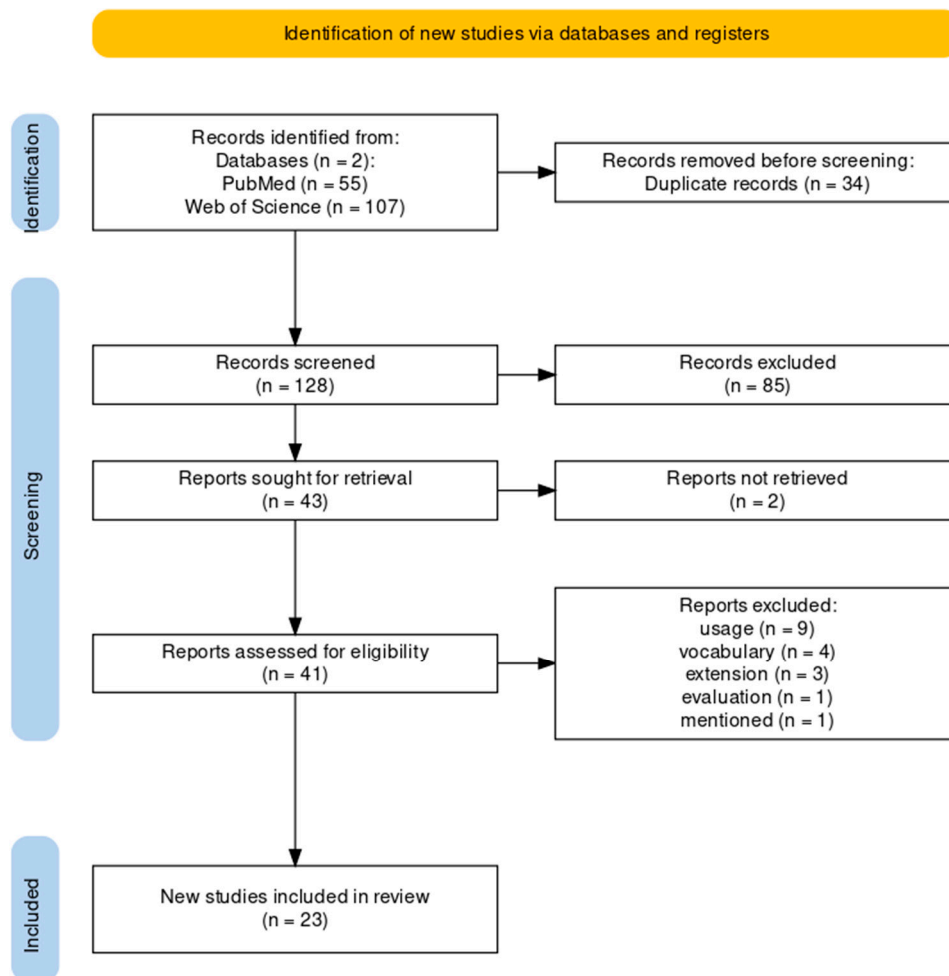


Figure 1. PRISMA flow diagram according to [14].

3.2.1. Frequency

The frequency of the process step occurrences in the literature is shown in **Error! Reference source not found.** in descending order. All publications included the implementation of ETL processes, semantic mapping and structural mapping. Furthermore, understanding the source data and mapping it to OMOP CDM was also important during data harmonization. This applies in particular to vocabulary identification (70%: 16/23), dataset specification (65%: 15/23), data profiling (39%: 9/23) and coverage analysis of vocabularies (35%: 8/23). About half of the publications dealt with qualitative (65%: 15/23) and quantitative data quality analyses (48%: 11/23) after successful transformation of the source data into OMOP CDM.

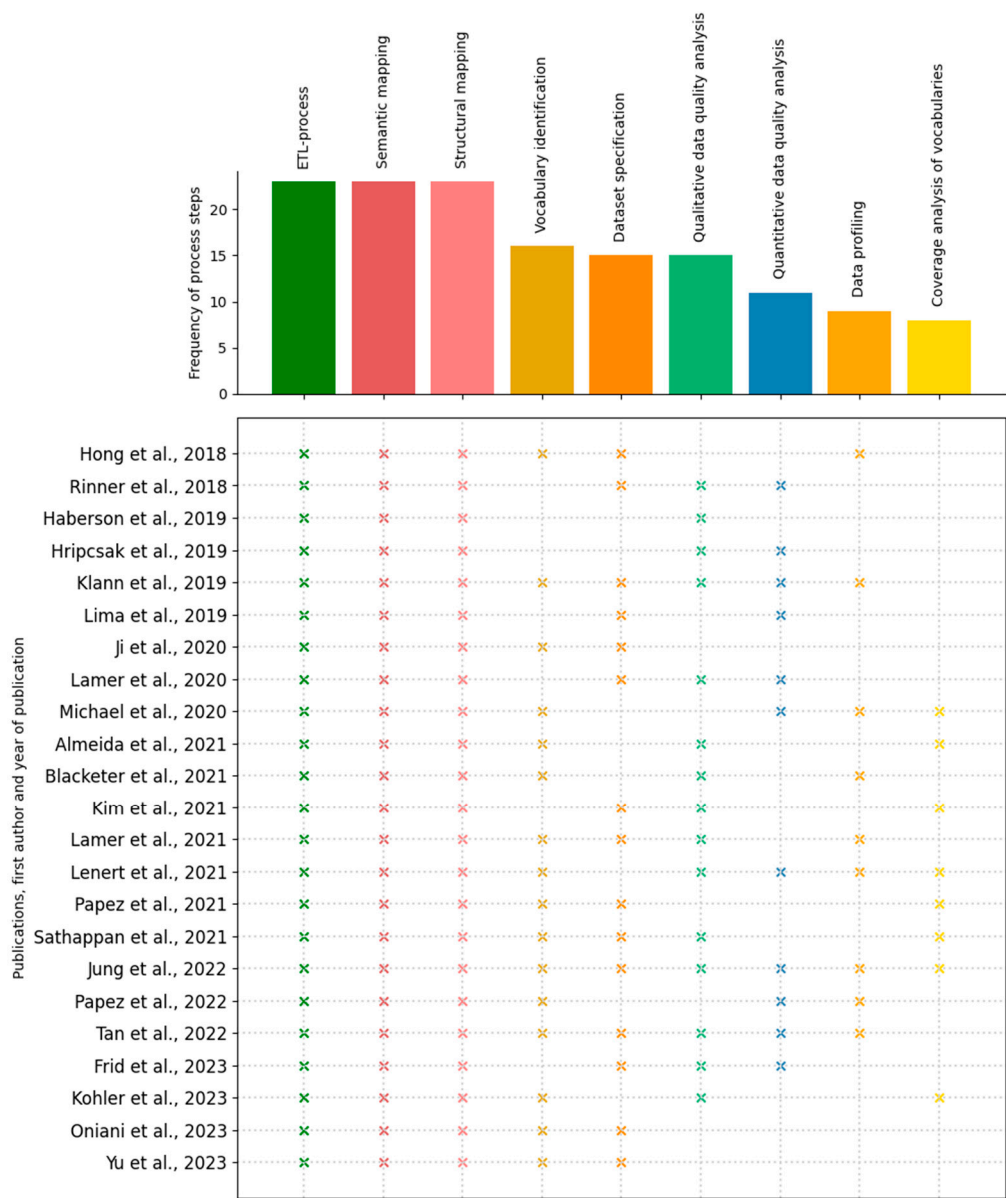


Figure 2. Frequency distribution of the extracted process steps and their assignment to the included publications.

3.2.2. Chronological Order

In order to define a generic sequence for the extracted process steps for data harmonization in OMOP CDM, the chronological order of the process steps was focused during the third iteration of data extraction. The chronological order of the process steps per publication can be found in the Supplementary Materials “File S1”.

To calculate the percentage distribution of the given numberings per process step, we assigned the included publications to the three groups a)-c). This resulted in the following number of publications per group: group a): 20 publications, group b): five publications and group c): 23 publications. **Error! Reference source not found., Error! Reference source not found. and Error! Reference source not found.** represent the percentage distribution as well as the indication of the most frequent numbering(s) for the groups a)-c).

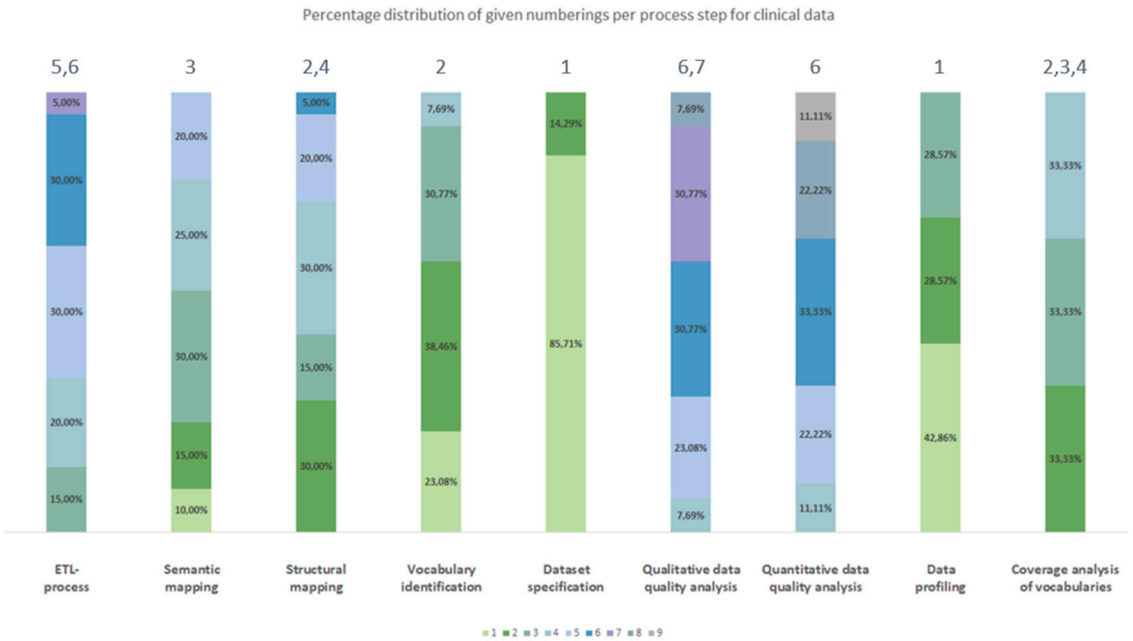


Figure 3. Percentage distribution of given numberings per process step for clinical data (group a)).



Figure 4. Percentage distribution of given numberings per process step for claims data (group b)).



Figure 5. Percentage distribution of given numberings per process step for clinical data and/or claims data (group c)).

The three diagrams show that the numbers 1 to 9 can occur several times as the most frequent numbering per group (e.g., number 1 in **Error! Reference source not found.**). Therefore, it was not yet possible to obtain a clear positioning of the identified process steps in an overall data harmonization process for each group.

3.3. Generic Data Harmonization Process

By comparing the most frequent number(s) per process step according to our approach in section 2.2, we were able to determine the chronological order of the data harmonization process for the groups a)-c). **Error! Reference source not found.** summarizes the results separated by the type of source data. The results show, that there are similarities (highlighted grey) and deviations of the chronological order of the process steps between the three groups. All three groups match completely for process numbers 1 and 7. For the remaining process numbers, there are deviations between the three groups, although two process steps per process number always match. This suggests that some process steps can also be interchanged with each other and the goal of harmonizing source data in OMOP has nevertheless been achieved by other researchers.

Table 3. Chronological order of process steps separated by type of source data.

Process number Type of source data	1	2	3	4	5	6	7	8	9
Clinical data	Dataset specification	Data profiling	Vocabulary identification	Coverage analysis of vocabularies	Structural mapping	Semantic mapping	ETL-process	Quantitative data quality analyses	Qualitative data quality analyses
Claims data	Dataset specification	Vocabulary identification	Coverage analysis of vocabularies	Data profiling	Semantic mapping	Structural mapping	ETL-process	Qualitative data quality analyses	Quantitative data quality analyses
Clinical data and/or claims data	Dataset specification	Data profiling	Vocabulary identification	Coverage analysis of vocabularies	Semantic mapping	Structural mapping	ETL-process	Qualitative data quality analyses	Quantitative data quality analyses

Finally, to conceptualize a generic process for data harmonization in OMOP CDM, the sequence of process steps that showed the most agreement was determined. **Error! Reference source not**

found. represents the resulting chronological order for the generic data harmonization process for OMOP CDM.

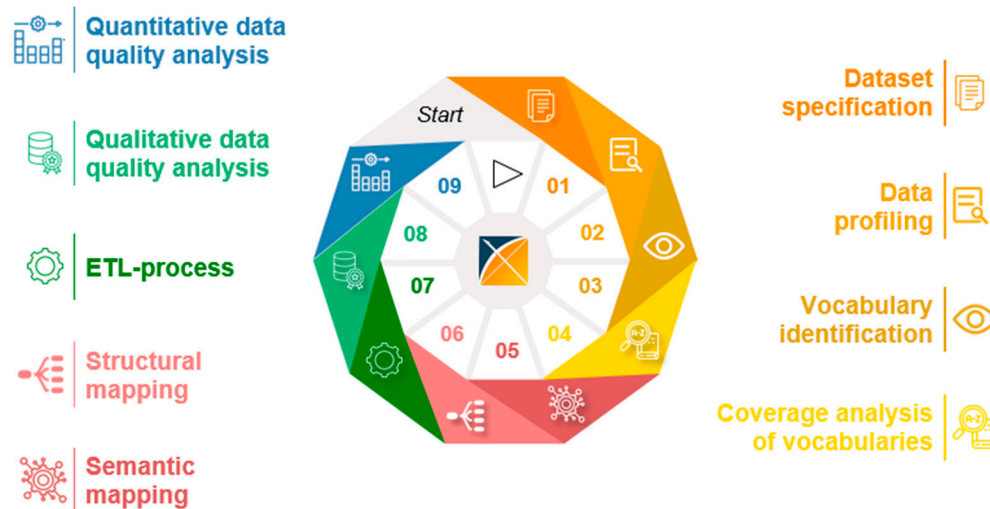


Figure 6. Generic data harmonization process for OMOP CDM; icons: Flaticon.com.

The generic data harmonization process consists of nine process steps in the following order:

1. **Dataset specification** refers to the definition of the scope of the source data for a specific use case. This is usually done by expert teams with clinical expertise. As a result, transformation to OMOP CDM is only performed on source data that is relevant to answering a specific research question.
2. To get an overview of the source data including their structure, formats and unique values, a **data profiling** is performed. For this purpose, OHDSI provides the tool WhiteRabbit [41] to analyzing source data.
3. **Vocabulary identification** focuses on providing a comprehensive compilation of the vocabularies found in the source data, including their scope of application.
4. In order to assess the extent to which the vocabularies found in the source data can already be mapped in OMOP CDM, a **coverage analysis of the vocabularies** of the source data is performed. The analysis helps to identify weaknesses (e.g., missing vocabularies) that would limit a full harmonization of the source data.
5. **Semantic mapping** refers to the mapping of local vocabularies to the standardized vocabulary of OMOP CDM [9] (pp. 55-74). This step is necessary to be able to uniquely identify source values by concepts in OMOP CDM and to transfer source values to standard concepts to enable research in an international context. The standardized vocabulary of OMOP CDM is provided by the OHDSI vocabulary repository Athena [42]. Furthermore, the OHDSI tool Usagi [43] supports researchers in semantic mapping of source values to OMOP CDM concepts.
6. The focus of the **structural mapping** is the conversion of the format of the source data into the standardized data model of OMOP CDM [9] (pp. 31-54). The structural mapping can be done by using the OHDSI tool RabbitInAHat [41].
7. The technical transformation of the source data into OMOP CDM is realized through the implementation of **ETL-processes**. ETL-processes enable the reading of source data (Extract), the practical implementation of semantic and structural mapping (Transform), and the final writing of OMOP-compliant source data to the target database (Load).
8. After a successful transformation of source data into OMOP CDM, the focus is on data quality analysis. The **qualitative data quality analysis** examines, in particular, the plausibility, conformity and completeness of the source data in OMOP CDM (according to Kahn et al. [44]). With the Automated Characterization of Health Information at Large-scale Longitudinal Evidence Systems (Achilles) [45] and the Data Quality Dashboard (DQD) [46], two OHDSI tools exist that perform data quality checks on the transformed source data in OMOP CDM.

9. The **quantitative data quality analysis** checks whether the number of data in the source matches the number of records in OMOP CDM.

4. Discussion

The results show that we have achieved our goal of conceptualizing a generic data harmonization process for OMOP CDM. By conducting a literature review, we were able to answer our first research question. The literature review revealed nine process steps that were used by other researchers to harmonize clinical data and/or claims data in OMOP CDM. However, the results per publication show that not all steps have to be relevant. For example, Hripcsak et al. [40], Lima et al. [25], Ji et al. [26], Oniani et al. [37] and Yu et al. [19] used only five from nine process steps, which also varied. Only the three process steps of semantic mapping, structural mapping and ETL-process are mentioned in all of the publications and thus form the mandatory part of the data harmonization process. Consequently, it is recommended to check individually for each data harmonization project whether process steps of the generic data harmonization process can be skipped if applicable.

In addition, our second research question was also addressed by looking at the chronological order of the identified process steps. A comparison of the results of the sequences of the three groups resulted in three different chronological orders of the nine process steps. This demonstrates that process steps can be interchanged if necessary while still achieving harmonization of source data in OMOP CDM (e.g., qualitative and quantitative data quality analysis). Despite the three different sequences, we were still able to derive a generic sequence of the nine process steps. Furthermore, it is important to mention that the generic data harmonization process should be considered as an iterative process. The last two steps of the qualitative and quantitative data quality analysis evaluate the correctness of the previous steps. If errors are identified, it may be necessary to perform the process again. In this context, it has to be checked individually whether all nine process steps have to be performed again or whether certain process steps do not contribute to the solution of the identified errors and can be skipped.

The present work is limited since only the harmonization of clinical data and claims data in OMOP CDM was focused. The literature review showed that there are also other types of data sources which are harmonized in OMOP CDM (e.g., registry data [47,48]). In the future, it is therefore necessary to check whether the addition of other types of source data has an impact on the amount and sequence of the process steps. Furthermore, there are limitations in the interpretation of the results from group b) (claims data). The reason for this is that only a small number of the included publications used claims data (5/23 publications), which limits a comparison with clinical data (20/23 publications) and a subsequent generalization of the results.

Notwithstanding the limitations listed above, our generic data harmonization process provides a major benefit. Compared to OHDSI's recommended 4-step process, our conceptualized 9-step process is more detailed. The reason for this is that the process description from OHDSI [9] is focused on the use of the OHDSI tools (e.g., WhiteRabbit, RabbitInAHat) for data harmonization in OMOP CDM. In contrast, our process can be carried out independently of the OHDSI tools and used as a guide for other researchers to follow.

5. Conclusions

Based on a literature review, necessary process steps for harmonizing clinical data or claims data in OMOP CDM were identified and placed in a chronological order. From these findings, a generic data harmonization process was derived. This process can be used as a step-by-step guide to assist other researchers in harmonizing source data in OMOP CDM. As future work, the applicability of the generic data harmonization process to German claims data will be investigated in practice. In this context, an evaluation will show whether further additional process steps need to be considered and to what extent the derived sequence is feasible in practice.

Supplementary Materials: The following supporting information can be downloaded at the website of this paper posted on Preprints.org, File S1: literature_review_results.xlsx.

Author Contributions: Conceptualization, Elisa Henke; Data curation, Elisa Henke and Ines Reinecke; Formal analysis, Elisa Henke, Ines Reinecke and Franziska Bathelt; Investigation, Elisa Henke; Methodology, Elisa Henke, Michele Zoch and Franziska Bathelt; Project administration, Elisa Henke; Resources, Martin Sedlmayr; Supervision, Martin Sedlmayr; Validation, Michele Zoch and Franziska Bathelt; Visualization, Elisa Henke, Yuan Peng and Ines Reinecke; Writing – original draft, Elisa Henke; Writing – review & editing, Michele Zoch, Yuan Peng, Ines Reinecke, Martin Sedlmayr and Franziska Bathelt.

Funding: This research received no external funding.

Data Availability Statement: The data presented in this study are available in the Supplementary Materials “File S1”.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Semler SC, Wissing F, Heyder R. German Medical Informatics Initiative - A National Approach to Integrating Health Data from Patient Care and Medical Research. *Methods Inf Med.* 2018;57(Suppl 1):e50-e56. doi:10.3414/ME18-03-0003
2. Green LA, Fryer GE, Yawn BP, Lanier D, Dovey SM. The Ecology of Medical Care Revisited. *N Engl J Med.* 2001;344(26):2021-2025. doi:10.1056/NEJM200106283442611
3. Kumar G, Basri S, Imam AA, Khawaja SA, Capretz LF, Balogun AO. Data Harmonization for Heterogeneous Datasets: A Systematic Literature Review. *Applied Sciences.* 2021;11(17):8275. doi:10.3390/app11178275
4. Garza M, Del Fiore G, Tenenbaum J, Walden A, Zozus MN. Evaluating common data models for use with a longitudinal community registry. *J Biomed Inform.* 2016;64:333-341. doi:10.1016/j.jbi.2016.10.016
5. Reinecke I, Zoch M, Reich C, Sedlmayr M, Bathelt F. The Usage of OHDSI OMOP - A Scoping Review. *Stud Health Technol Inform.* 2021;283:95-103. doi:10.3233/SHTI210546
6. EHDSN. European Health Data & Evidence Network. Published 2022. Accessed May 20, 2022. <https://www.ehden.eu/>
7. European Medical Agency. Data Analysis and Real World Interrogation Network (DARWIN EU). Published 2023. Accessed May 20, 2022. <https://www.ema.europa.eu/en/about-us/how-we-work/big-data/data-analysis-real-world-interrogation-network-darwin-eu>
8. Observational Health Data Sciences and Informatics. HL7 International and OHDSI Announce Collaboration to Provide Single Common Data Model for Sharing Information in Clinical Care and Observational Research. Published 2021. Accessed May 6, 2022. <https://www.ohdsi.org/ohdsi-hl7-collaboration/>
9. Observational Health Data Sciences and Informatics. The Book of OHDSI. In: *The Book of OHDSI.* ; 2021. Accessed April 19, 2022. <https://ohdsi.github.io/TheBookOfOhdsi/>
10. Peng Y, Henke E, Reinecke I, Zoch M, Sedlmayr M, Bathelt F. An ETL-process design for data harmonization to participate in international research with German real-world data based on FHIR and OMOP CDM. *International Journal of Medical Informatics.* 2023;169:104925. doi:10.1016/j.ijmedinf.2022.104925
11. Digital Scholar. Zotero. Zotero. Published 2023. Accessed March 31, 2023. <https://www.zotero.org/>
12. Fleiss JL. Measuring nominal scale agreement among many raters. *Psychological Bulletin.* 1971;76(5):378-382. doi:10.1037/H0031619
13. Signorell A, Aho K, Alfons A, et al. DescTools: Tools for Descriptive Statistics. Published online September 6, 2023. Accessed October 25, 2023. <https://cran.r-project.org/web/packages/DescTools/index.html>
14. Haddaway NR, Page MJ, Pritchard CC, McGuinness LA. PRISMA2020: An R package and Shiny app for producing PRISMA 2020-compliant flow diagrams, with interactivity for optimised digital transparency and Open Synthesis. *Campbell Systematic Reviews.* 2022;18(2):e1230. doi:10.1002/cl2.1230
15. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ.* 2009;339:b2700. doi:10.1136/bmj.b2700
16. Klann J, Joss M, Embree K, Murphy S. Data model harmonization for the All Of Us Research Program: Transforming i2b2 data into the OMOP common data model. *PLOS ONE.* 2019;14(2). doi:10.1371/journal.pone.0212463
17. Lamer A, Abou-Arab O, Bourgeois A, et al. Transforming Anesthesia Data Into the Observational Medical Outcomes Partnership Common Data Model: Development and Usability Study. *Journal of Medical Internet Research.* 2021;23(10):e29259. doi:10.2196/29259
18. Hripcsak G, Shang N, Peissig PL, et al. Facilitating phenotype transfer using a common data model. *JOURNAL OF BIOMEDICAL INFORMATICS.* 2019;96. doi:10.1016/j.jbi.2019.103253

19. Yu Y, Jiang G, Brandt E, et al. Integrating real-world data to assess cardiac ablation device outcomes in a multicenter study using the OMOP common data model for regulatory decisions: implementation and evaluation. *JAMIA OPEN*. 2023;6(1). doi:10.1093/jamiaopen/ooac108
20. Papez V, Moinat M, Payralbe S, et al. Transforming and evaluating electronic health record disease phenotyping algorithms using the OMOP common data model: a case study in heart failure. *JAMIA OPEN*. 2021;4(3). doi:10.1093/jamiaopen/ooab001
21. Tan H, Teo D, Lee D, et al. Applying the OMOP Common Data Model to Facilitate Benefit-Risk Assessments of Medicinal Products Using Real-World Data from Singapore and South Korea. *HEALTHCARE INFORMATICS RESEARCH*. 2022;28(2):112-122. doi:10.4258/hir.2022.28.2.112
22. Papez V, Moinat M, Voss E, et al. Transforming and evaluating the UK Biobank to the OMOP Common Data Model for COVID-19 research and beyond. *JOURNAL OF THE AMERICAN MEDICAL INFORMATICS ASSOCIATION*. 2022;30(1):103-111. doi:10.1093/jamia/ocac203
23. Jung H, Yoo S, Kim S, et al. Patient-Level Fall Risk Prediction Using the Observational Medical Outcomes Partnership's Common Data Model: Pilot Feasibility Study. *JMIR MEDICAL INFORMATICS*. 2022;10(3). doi:10.2196/35104
24. Almeida J, Silva J, Matos S, Oliveira J. A two-stage workflow to extract and harmonize drug mentions from clinical notes into observational databases. *JOURNAL OF BIOMEDICAL INFORMATICS*. 2021;120. doi:10.1016/j.jbi.2021.103849
25. Lima D, Rodrigues J, Traina A, Pires F, Gutierrez M. Transforming Two Decades of ePR Data to OMOP CDM for Clinical Research. In: OhnoMachado L, Seroussi B, eds. *Universidade de Sao Paulo*. Vol 264. ; 2019:233-237. doi:10.3233/SHTI190218
26. Ji H, Kim S, Yi S, Hwang H, Kim J, Yoo S. Converting clinical document architecture documents to the common data model for incorporating health information exchange data in observational health studies: CDA to CDM. *JOURNAL OF BIOMEDICAL INFORMATICS*. 2020;107. doi:10.1016/j.jbi.2020.103459
27. Kim J, Kim S, Ryu B, Song W, Lee H, Yoo S. Transforming electronic health record polysomnographic data into the Observational Medical Outcome Partnership's Common Data Model: a pilot feasibility study. *SCIENTIFIC REPORTS*. 2021;11(1). doi:10.1038/s41598-021-86564-w
28. Blacketer C, Voss E, DeFalco F, et al. Using the Data Quality Dashboard to Improve the EHDEN Network. *APPLIED SCIENCES-BASEL*. 2021;11(24). doi:10.3390/app112411920
29. Rinner C, Gezgin D, Wendl C, Gall W. A Clinical Data Warehouse Based on OMOP and i2b2 for Austrian Health Claims Data. *Stud Health Technol Inform*. 2018;248:94-99.
30. Haberson A, Rinner C, Schöberl A, Gall W. Feasibility of Mapping Austrian Health Claims Data to the OMOP Common Data Model. *J Med Syst*. 2019;43(10):314. doi:10.1007/s10916-019-1436-9
31. Sathappan SMK, Jeon YS, Dang TK, et al. Transformation of Electronic Health Records and Questionnaire Data to OMOP CDM: A Feasibility Study Using SG_T2DM Dataset. *Appl Clin Inform*. 2021;12(4):757-767. doi:10.1055/s-0041-1732301
32. Michael CL, Sholle ET, Wulff RT, Roboz GJ, Campion TR. Mapping Local Biospecimen Records to the OMOP Common Data Model. *AMIA Jt Summits Transl Sci Proc*. 2020;2020:422-429.
33. Hong N, Zhang N, Wu H, et al. Preliminary exploration of survival analysis using the OHDSI common data model: a case study of intrahepatic cholangiocarcinoma. *BMC Med Inform Decis Mak*. 2018;18(Suppl 5):116. doi:10.1186/s12911-018-0686-7
34. Lamer A, Depas N, Doutreligne M, et al. Transforming French Electronic Health Records into the Observational Medical Outcome Partnership's Common Data Model: A Feasibility Study. *Appl Clin Inform*. 2020;11(1):13-22. doi:10.1055/s-0039-3402754
35. Lenert LA, Ilatovskiy AV, Agnew J, et al. Automated production of research data marts from a canonical fast healthcare interoperability resource data repository: applications to COVID-19 research. *J Am Med Inform Assoc*. 2021;28(8):1605-1611. doi:10.1093/jamia/ocab108
36. Kohler S, Boscá D, Kärcher F, et al. Eos and OMOCL: Towards a seamless integration of openEHR records into the OMOP Common Data Model. *J Biomed Inform*. 2023;144:104437. doi:10.1016/j.jbi.2023.104437
37. Oniani D, Parmanto B, Saptano A, et al. ReDWINE: A clinical datamart with text analytical capabilities to facilitate rehabilitation research. *Int J Med Inform*. 2023;177:105144. doi:10.1016/j.ijmedinf.2023.105144
38. Frid S, Pastor Duran X, Bracons Cucó G, et al. An Ontology-Based Approach for Consolidating Patient Data Standardized With European Norm/International Organization for Standardization 13606 (EN/ISO 13606) Into Joint Observational Medical Outcomes Partnership (OMOP) Repositories: Description of a Methodology. *JMIR Med Inform*. 2023;11:e44547. doi:10.2196/44547
39. Lamer A, Abou-Arab O, Bourgeois A, et al. Transforming Anesthesia Data Into the Observational Medical Outcomes Partnership Common Data Model: Development and Usability Study. *JOURNAL OF MEDICAL INTERNET RESEARCH*. 2021;23(10). doi:10.2196/29259
40. Hripcsak G, Shang N, Peissig P, et al. Facilitating phenotype transfer using a common data model. *JOURNAL OF BIOMEDICAL INFORMATICS*. 2019;96. doi:10.1016/j.jbi.2019.103253

41. Observational Health Data Sciences and Informatics. White Rabbit. Published 2022. Accessed November 11, 2022. <http://ohdsi.github.io/WhiteRabbit/WhiteRabbit.html>
42. Observational Health Data Sciences and Informatics. Athena. Athena – OHDSI Vocabularies Repository. Published 2023. Accessed November 25, 2022. <https://athena.ohdsi.org/>
43. Schuemie M. Usagi. Usagi. Published 2021. <http://ohdsi.github.io/Usagi/>
44. Kahn MG, Callahan TJ, Barnard J, et al. A Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data. *EGEMS (Wash DC)*. 2016;4(1):1244. doi:10.13063/2327-9214.1244
45. Observational Health Data Sciences and Informatics. Achilles. Published online August 2, 2022. Accessed April 12, 2023. <https://github.com/OHDSI/Achilles>
46. Blacketer C, Defalco FJ, Ryan PB, Rijnbeek PR. Increasing trust in real-world evidence through evaluation of observational data quality. *Journal of the American Medical Informatics Association*. 2021;28(10):2251-2257. doi:10.1093/jamia/ocab132
47. Fischer P, Stöhr MR, Gall H, Michel-Backofen A, Majeed RW. Data Integration into OMOP CDM for Heterogeneous Clinical Data Collections via HL7 FHIR Bundles and XSLT. *Digital Personalized Health and Medicine*. Published online 2020:138-142. doi:10.3233/SHTI200138
48. Biedermann P, Ong R, Davydov A, et al. Standardizing registry data to the OMOP Common Data Model: experience from three pulmonary hypertension databases. *BMC Medical Research Methodology*. 2021;21(1):238. doi:10.1186/s12874-021-01434-3

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.