

Review

# Use of Metadata-Driven Approaches for Data Harmonization in the Medical Domain: Scoping Review

Yuan Peng<sup>1</sup>, MSc; Franziska Bathelt<sup>2</sup>, Dr Rer Nat; Richard Gebler<sup>1</sup>, MSc; Robert Gött<sup>3</sup>, Dipl.-Ing.; Andreas Heidenreich<sup>4</sup>, Dipl.-Biol., Dr Rer Nat; Elisa Henke<sup>1</sup>, MSc; Dennis Kadioglu<sup>4,5</sup>, MSc; Stephan Lorenz<sup>1</sup>, MSc; Abishaa Vengadeswaran<sup>5</sup>, MSc; Martin Sedlmayr<sup>1</sup>, Dr Rer Nat, Prof Dr

<sup>1</sup>Institute for Medical Informatics and Biometry, Carl Gustav Carus Faculty of Medicine, Technische Universität Dresden, Dresden, Germany

<sup>2</sup>Thiem-Research GmbH, Cottbus, Germany

<sup>3</sup>Core Unit Datenintegrationszentrum, University Medicine Greifswald, Greifswald, Germany

<sup>4</sup>Department for Information and Communication Technology (DICT), Data Integration Center (DIC), Goethe University Frankfurt, University Hospital, Frankfurt am Main, Germany

<sup>5</sup>Institute for Medical Informatics, Goethe University Frankfurt, University Hospital Frankfurt, Frankfurt am Main, Germany

**Corresponding Author:**

Yuan Peng, MSc

Institute for Medical Informatics and Biometry

Carl Gustav Carus Faculty of Medicine

Technische Universität Dresden

Fetscherstraße 74

Dresden, 01307

Germany

Phone: 49 3514583648

Fax: 49 3514585738

Email: [yuan.peng@tu-dresden.de](mailto:yuan.peng@tu-dresden.de)

## Abstract

**Background:** Multisite clinical studies are increasingly using real-world data to gain real-world evidence. However, due to the heterogeneity of source data, it is difficult to analyze such data in a unified way across clinics. Therefore, the implementation of Extract-Transform-Load (ETL) or Extract-Load-Transform (ELT) processes for harmonizing local health data is necessary, in order to guarantee the data quality for research. However, the development of such processes is time-consuming and unsustainable. A promising way to ease this is the generalization of ETL/ELT processes.

**Objective:** In this work, we investigate existing possibilities for the development of generic ETL/ELT processes. Particularly, we focus on approaches with low development complexity by using descriptive metadata and structural metadata.

**Methods:** We conducted a literature review following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines. We used 4 publication databases (ie, PubMed, IEEE Explore, Web of Science, and Biomed Center) to search for relevant publications from 2012 to 2022. The PRISMA flow was then visualized using an R-based tool (Evidence Synthesis Hackathon). All relevant contents of the publications were extracted into a spreadsheet for further analysis and visualization.

**Results:** Regarding the PRISMA guidelines, we included 33 publications in this literature review. All included publications were categorized into 7 different focus groups (ie, medicine, data warehouse, big data, industry, geoinformatics, archaeology, and military). Based on the extracted data, ontology-based and rule-based approaches were the 2 most used approaches in different thematic categories. Different approaches and tools were chosen to achieve different purposes within the use cases.

**Conclusions:** Our literature review shows that using metadata-driven (MDD) approaches to develop an ETL/ELT process can serve different purposes in different thematic categories. The results show that it is promising to implement an ETL/ELT process by applying MDD approach to automate the data transformation from Fast Healthcare Interoperability Resources to Observational Medical Outcomes Partnership Common Data Model. However, the determining of an appropriate MDD approach and tool to implement such an ETL/ELT process remains a challenge. This is due to the lack of comprehensive insight into the characterizations of the MDD approaches presented in this study. Therefore, our next step is to evaluate the MDD approaches presented in this

study and to determine the most appropriate MDD approaches and the way to integrate them into the ETL/ELT process. This could verify the ability of using MDD approaches to generalize the ETL process for harmonizing medical data.

(*JMIR Med Inform* 2024;12:e52967) doi: [10.2196/52967](https://doi.org/10.2196/52967)

## KEYWORDS

ETL; ELT; Extract-Load-Transform; Extract-Transform-Load; interoperability; metadata-driven; medical domain; data harmonization

## Introduction

Multisite clinical studies are increasingly using real-world data to gain real-world evidence, especially during the COVID-19 pandemic [1]. However, not all clinics use the same hospital information system, resulting in heterogeneity of data produced by different hospital information systems. These heterogeneous data are not semantically and syntactically interoperable. Therefore, it is difficult to analyze such data in a unified way across sites. For this, the heterogeneous data need to be harmonized and standardized, for example, by using a common data model (CDM) [2]. For example, the European Medical Agency [3] set up the DARWIN EU (Data Analysis and Real World Interrogation Network European Union) [4] to provide real-world evidence on use and adverse events of medicines across the European Union. DARWIN EU uses the Observational Medical Outcomes Partnership (OMOP) CDM [5] as the base model, which is provided by the Observational Health Data Sciences and Informatics [6] community. To participate in such networks, a transformation of local data is needed. A common approach is to develop an Extract-Transform-Load (ETL) or Extract-Load-Transform (ELT) process. Both are used to harmonize heterogeneous data into the target systems. The only difference between them is the order of processing data. ETL transforms the data before loading them into the target systems, while ELT loads the data into the target systems first, and then transforms the data. Due to the different data formats and source systems, multiple ETL/ELT processes have to be implemented [7-10]. This work is time-consuming and hard to maintain [11].

Using a standard data exchange format can reduce the complexity of transforming heterogeneous data into CDMs. An example is the Fast Healthcare Interoperability Resources (FHIR) [12] format. FHIR is a communication standard and is provided by the Health Level 7 (HL7) [13]. In Germany, the Medical Informatics Initiative (MII) [14] provides a Core Data Set (CDS) [15] in FHIR format for enabling the interoperability of data across all university hospitals. Another German association “the National Association of Statutory Health Insurance Physicians” (KBV, German: Kassenärztliche Bundesvereinigung) [16] also provides a KBV CDS in FHIR format, which provides a stable foundation for the development of the medical information objects [17] (eg, immunization records and maternity records). Although both MII CDS and KBV CDS are based on the German HL7 Basis Profiles [18], the FHIR profiles defined in the 2 CDSs are not identical [19]. This is due to the different requirements of MII and KBV. For example, codes indicating departments within a clinic (eg, 0100 for internal medicine department) are defined in different

value-sets and therefore use different coding systems. This also complicates the implementation and maintenance of ETL/ELT processes.

Furthermore, most countries try to standardize their electronic health records (EHR) data for research and to improve the interoperability of the data. Consequently, country-specific FHIR profiles are developed, for example, German HL7 Basis Profiles [18] and the US CDS [20]. Due to different languages (ie, German vs English), different structure definitions (eg, extensions and cardinality) and different coding systems (eg, system URL for International Classification of Diseases, 10, Revision: German Modification [21] vs system URL for International Classification of Diseases, 10, Clinical Modification [22]) used in the FHIR profiles, different ETL processes need to be implemented [8,23]. Although these are just a few examples, it is conceivable that with the expansion of supported use cases, the time required for implementing an ETL/ELT process increases massively, while the maintainability decreases. Therefore, the implementation of a generic ETL/ELT process for harmonizing local health data can guarantee the semantic and syntactic interoperability of research data across sites and countries.

Using metadata for the implementation of ETL/ELT processes is a promising approach, as stated by David Loshin [24]: “in order to organize data for analytical purposes, it will need to be extracted from the original source (source metadata), transformed into a representation that is consistent with the warehouse (target metadata) in a way that does not lose information due to differences in format and precision (structure metadata) and is aligned in a meaningful way (semantic metadata).” A very broad definition of metadata is “data about other data” [25]. Depending on the specific context of use, metadata can be classified into 3 types [26]:

- **Descriptive metadata:** the metadata is used for discovery and identification purposes, for example metadata for source and target data.
- **Structural metadata:** the metadata is used for managing data in information systems, for example, column names and table names in a database.
- **Administrative metadata:** the metadata exists within a database that provides additional information, for example, the name of a person, who has changed the data in a database.

Metadata can be represented by metadata languages (eg, Resource Description Framework and Notation3) [27]. Such languages are also called ontology languages. For enabling the interoperability of data from different source and target systems, rule languages (eg, Rule Markup Language and Semantic Web

Rule Language) can be used to define the transformation rules between them [27]. Therefore, the use of metadata is expected to improve the development and maintenance for transforming FHIR resources to OMOP CDM.

As a side note, we understand any (descriptive and structural) metadata-based approach used for developing ETL/ELT processes as metadata-driven (MDD) approach. This work focuses on providing an overview of the types of MDD approaches and their use in different thematic categories. The overview aims to identify a suitable MDD approach to enhance the data transformation from FHIR to OMOP CDM. This will be achieved by answering the following questions:

- Q1: What are the themes of application for MDD approaches?
- Q2: What types of MDD approaches exist in the literature?
- Q3: What are the reasons for the usage of MDD approaches?
- Q4: What tool was used to implement the MDD approach?

## Methods

To answer our 4 research questions, we conducted a literature review. To ensure the transparency of the review process, we followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [28]. We used 4 publication databases (ie, PubMed, IEEE Explore, Web of Science, and Biomed Center) to search for relevant publications from 2012 to 2022 written in German or English (Textbox 1). The first search was performed on August 11, 2022, and the second one was on March 15, 2023, which in turn completed the search through December 31, 2022. The collected publications were loaded into the Zotero Citation Management

program (Corporation for Digital Scholarship) [29] and the duplicates were manually removed. To better categorize the publications to be excluded, we defined 8 exclusion criteria (Textbox 2).

This review was a 2-fold process consisting of Title-Abstract-Screening (TAS) and full-text screening (FTS). Both screening processes used the same exclusion criteria listed in Textbox 2. The unique publications were divided into 2 groups based on their publication dates and uploaded to a research collaboration platform, Rayyan (Qatar Computing Research Institute and Cochrane Bahrain) [30], as 2 separate projects. Each publication group was assigned with 4 reviewers. The corresponding author reviewed all publications. The TAS was performed under the blind-modus, so that each reviewer could label the publication independently. The blind-modus was turned off after all publications were tagged and the conflicts were discussed and resolved. After that, all included publications were randomly divided into 2 groups and reloaded into Rayyan as a new project for FTS. Similar to TAS, 4 reviewers were assigned to each publication group and the corresponding author reviewed all publications. The FTS was also conducted under the blind-modus and followed the same review process as the TAS.

We extracted the content of all included publications based on the categories listed in Textbox 3. The extraction of publication content was done by the corresponding author and validated by 4 coauthors. The extracted content was stored in a spreadsheet for further analysis and visualization.

The result of the literature review was visualized using an R-based tool, which was developed based on PRISMA 2020 [31].

**Textbox 1.** Search string and publication databases.

Search string
<b>PubMed</b>
<ul style="list-style-type: none"> <li>• ((meta data) OR (meta-data) OR (metadata) OR (ontology) OR (rules)) AND ((extract transform load) OR (ETL) OR (extract load transform) OR (ELT))</li> </ul>
<b>IEEE Explore</b>
<ul style="list-style-type: none"> <li>• (“All Metadata”:metadata) OR (“All Metadata”:meta-data) OR (“All Metadata”:meta data) OR (“All Metadata”:ontology) OR (“All Metadata”:rules)) AND (“All Metadata”:ETL) OR (“All Metadata”:extract transform load) OR (“All Metadata”:ELT) OR (“All Metadata”:extract load transform))</li> </ul>
<b>Web of Science</b>
<ul style="list-style-type: none"> <li>• (ALL=(metadata) OR ALL=(meta-data) OR ALL= (“meta data”) OR ALL=(ontology) OR ALL=(rules)) AND (ALL=(ETL) OR ALL= (“extract transform load”) OR ALL=(ELT) OR ALL= (“extract load transform”))</li> </ul>
<b>Biomed Center (BMC)</b>
<ul style="list-style-type: none"> <li>• (“meta data” OR meta-data OR metadata OR ontology OR rules) AND (“extract transform load” OR ETL OR “extract load transform” OR ELT)</li> </ul>

**Textbox 2.** Labels and descriptions of exclusion criteria.

<p><b>Wrong_abbreviation</b></p> <ul style="list-style-type: none"><li>• Publication does not contain Extract-Transform-Load (ETL) as “Extract-Transform-Load.”</li><li>• Publication does not contain Extract-Load-Transform (ELT) as “Extract-Load-Transform.”</li></ul> <p><b>Wrong_definition</b></p> <ul style="list-style-type: none"><li>• Publication does not use metadata in the context of “metadata of data in source or target.”</li><li>• Publication does not use rules in the context of “rules for data transformation.”</li></ul> <p><b>Only_etl_elt</b></p> <ul style="list-style-type: none"><li>• Publication describes only ETL/ELT.</li></ul> <p><b>Only_metadata</b></p> <ul style="list-style-type: none"><li>• Publication describes only metadata.</li></ul> <p><b>Wrong_focus</b></p> <ul style="list-style-type: none"><li>• Publication mentioned metadata and ETL/ELT, but the focus is not about data harmonization</li></ul> <p><b>Wrong_type</b></p> <ul style="list-style-type: none"><li>• Publication is not a conference paper or a journal publication</li></ul> <p><b>Foreign_language</b></p> <ul style="list-style-type: none"><li>• Publication is written in other languages than English and German</li></ul> <p><b>Wrong_content</b></p> <ul style="list-style-type: none"><li>• Publication does not mention ETL/ELT or metadata</li></ul>
---

**Textbox 3.** Categories for data extraction.

<p><b>Theme</b></p> <ul style="list-style-type: none"><li>• The main theme of the work.</li></ul> <p><b>Metadata-driven method</b></p> <ul style="list-style-type: none"><li>• The used metadata-driven method in the work.</li></ul> <p><b>Metadata-driven method tool</b></p> <ul style="list-style-type: none"><li>• Tool which was used to conduct the metadata-driven method.</li></ul> <p><b>Purpose</b></p> <ul style="list-style-type: none"><li>• The purpose of using the metadata-driven method.</li></ul>
---

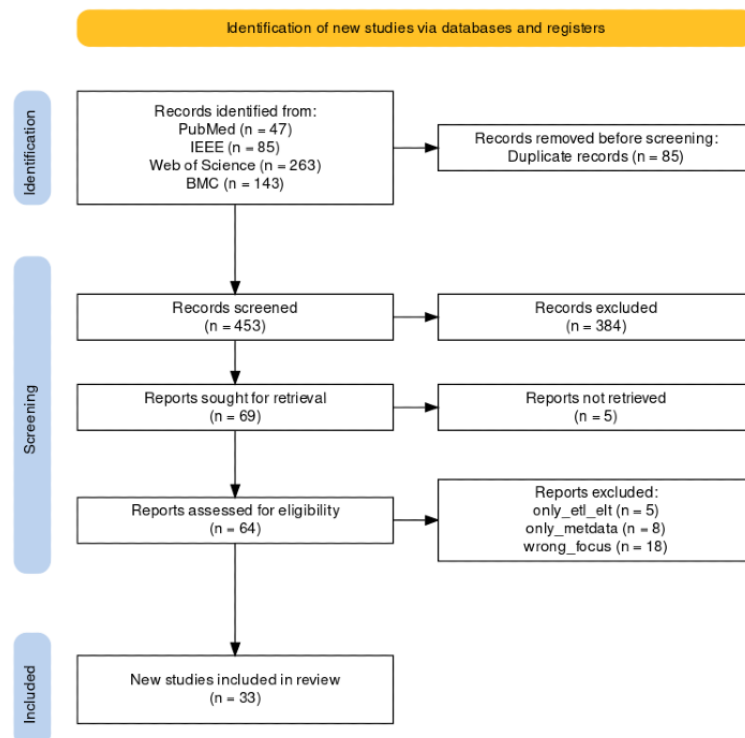
## Results

### Literature Search

The literature search resulted in 538 publications. After removing 85 duplicates, 453 publications were screened during the TAS phase. By using the exclusion criteria defined in

[Textbox 2](#) and excluding the publications, which have no full-text, 64 publications were included for FTS. Finally, we included 33 publications in this work. The screening process and results are structured using the PRISMA flow diagram 2020 ([Figure 1](#)). A complete list of included publications is available in [Multimedia Appendix 1](#).

**Figure 1.** PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram. Generated using an R-based tool (reproduced from Haddaway et al [31], with permission from Neal R Haddaway).



## Distribution of Publications

In order to gain an overview of the potential application focuses of MDDs (Q1) and thus an indication of where the approaches have proven beneficial, the focused theme of application was first evaluated. According to the extracted data, the focuses of all included publications are classified into 7 different categories, namely medicine (n=9) [10,32-39], data warehouse (n=13) [40-52], big data (n=4) [53-56], industry (n=4) [57-60], geoinformatics (n=1) [61], archaeology (n=1) [62], and military (n=1) [63]. This shows that data warehouse and medicine are the 2 categories that use the MDD approach the most.

## MDD Approaches Used for Various Thematic Categories

Different types of MDD approaches were used across the thematic categories. To gain knowledge about the use of these

types of MDD approaches in each category (Q2), the distribution of MDD approaches was investigated. Figure 2 shows the application of different types of MDD approaches in different thematic categories. The most frequently used type of MDD approach was ontology-based, where the ontology (using for example, resource description framework) of the source or target was applied in the ETL/ELT process. This approach was used in 6 categories, particularly in the categories of data warehouse [45-48,50,52] and medicine [10,32,35,37-39]. Another frequently used type of MDD approach was rule-based, which applied transformation rules generated based on the source and target to the ETL/ELT process. The rule-based approach was also widely used in the categories of data warehouse [40-43,49] and medicine [33,34,37,39]. All other MDD approaches besides the ontology-based and rule-based approaches were categorized as “other” (Table 1).

Figure 2. Metadata-driven approaches used in each thematic category.

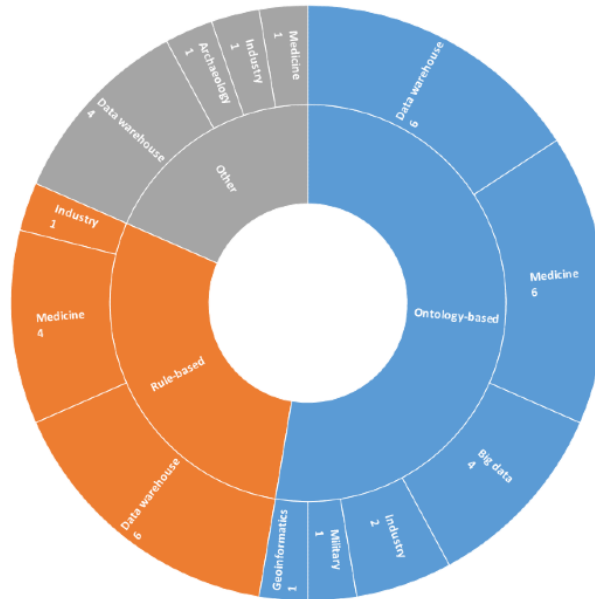


Table 1. MDD<sup>a</sup> approaches that are categorized as “other.”

MDD approach type and publication	Example
<b>UML<sup>b</sup>-based</b>	
Dhaouadi et al [46]	UML class diagram is used for modeling the transformation process
<b>Graphic-based</b>	
Dhaouadi et al [46]	BPMN <sup>c</sup> standard is used for modeling an ETL <sup>d</sup> process
<b>Ad hoc formalisms-based</b>	
Dhaouadi et al [46]	Entity Mapping Diagram is used for representing ETL tasks
<b>MDA<sup>e</sup>-based</b>	
Dhaouadi et al [46]	MDA is a multilayered framework with multiple submodules for separation of the specification of a functionality from its implementation
<b>Message-based</b>	
Novak et al [51]	“Normal message” contains information of mapping and transformation; “command message” configures the (execution) system
<b>Template-based</b>	
McCarthy et al [58]	A transformation template for each data source that manages the complex transformation process
Binding et al [62]	A template contains the mapping patterns which is then used for querying in database
<b>Metadata-based<sup>f</sup></b>	
Ozyurt and Grethe [36]	Implementing a generic data transformation language to transform heterogeneous data from multiple sources to a common format
Tomingas et al [44]	Metadata of the source and target stored in a knowledge and metadata repository
Suleykin and Panfilov [60]	Metadata of the mapping path stored in a metadata management framework

<sup>a</sup>MDD: metadata-driven.

<sup>b</sup>UML: unified modeling language.

<sup>c</sup>BPMN: Business Process Model Notation.

<sup>d</sup>ETL: Extract-Transform-Load.

<sup>e</sup>MDA: Model Driven Architecture.

<sup>f</sup>Metadata-based approach: approach uses metadata without any specification.

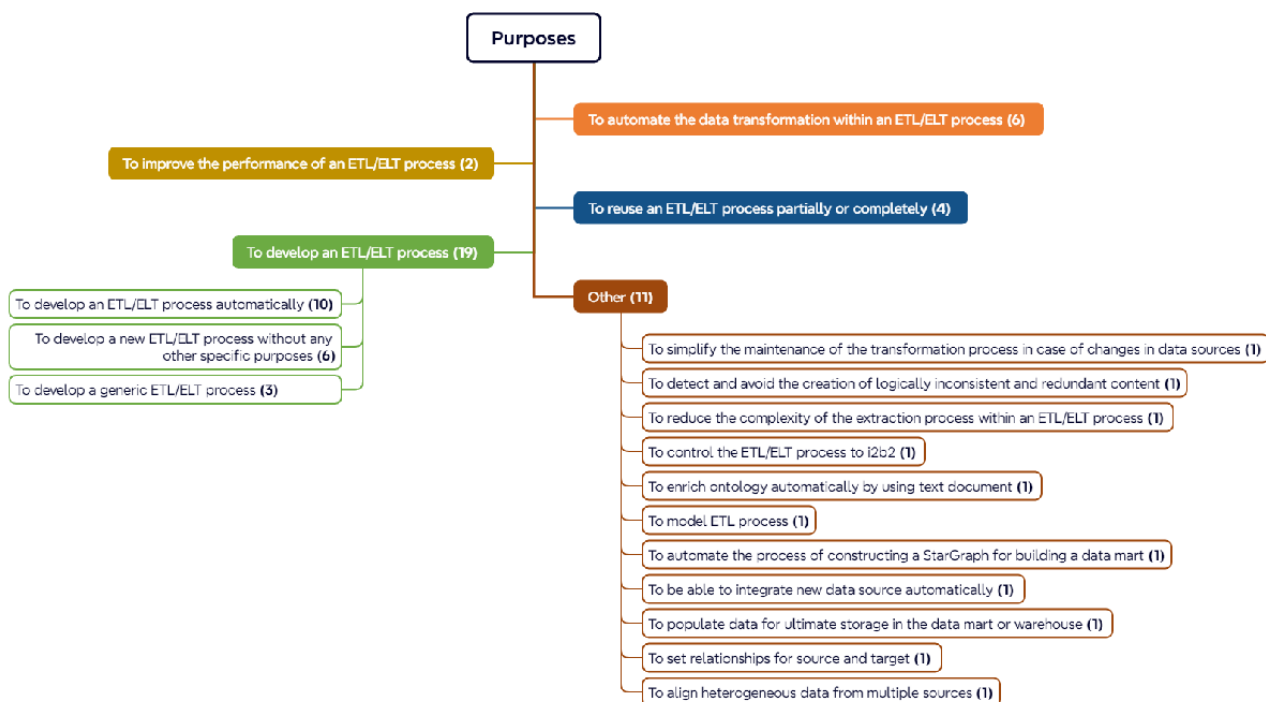


### Purposes of Using MDD Method for Data Harmonization

The purpose of using MDD approaches in each use case was then investigated to clarify the reasons why MDD approaches were used (Q3). Figure 3 shows different purposes of using MDD approaches in developing ETL/ELT processes based on the extracted data. The majority of publications describe the use of MDD approaches to develop an ETL/ELT process. This purpose can be divided into three detailed categories: (1) to automate the development of the ETL/ELT process [35,38,42,46,48-51,60], (2) to develop a generic ETL/ELT process [39,47,52], and (3) to develop a new ETL/ELT process

without any further technical specifications [40,45,46,55,57,61]. Additionally, the transformation part of the ETL/ELT process could also be automated by applying an MDD approach [34,37,41,44,58,63]. For example, Chen and Zhao [41] described an MDD approach for the automatic generation of SQL scripts for data transformation. Moreover, using MDD approaches can also help to improve the performance of ETL/ELT processes [43,46] or to partially or fully reuse the ETL/ELT process [10,33,43,62]. Other goals (categorized as “Others” in Figure 3), such as simplifying the maintenance of the transformation process [37] and reducing the complexity of the extraction process [53], can also be realized by using MDD approaches in ETL/ELT processes.

**Figure 3.** Purposes of using MDD approaches in ETL/ELT process. ELT: Extract-Load-Transform; ETL: Extract-Transform-Load; i2b2: Informatics for Integrating Biology and the Bedside; MDD: metadata-driven.



### Relationship Between Use Case and Used MDD Approach

As shown in the previous section, different MDD approaches were applied for different purposes. To further elucidate the reasons for choosing MDD approaches (Q3), the relationship between them was investigated. Table 2 lists the number of publications, which used a type of MDD approach to achieve a specific purpose. The ontology-based approach was used to achieve purposes (1) and (2), and (4)-(7). For example, Huang et al [63] created both local ontology (ontology based on the source data) and global ontology (ontology for the query processing) for the data transformation process, so that the data transformation from local ontology to global can be automated by applying ontology learning, ontology mapping, and ontology

rules. Additionally, the ontology-based approach was also used to achieve other goals, such as controlling the ETL process to Informatics for Integrating Biology and the Bedside [32] and reducing the complexity of the extraction process [53]. Similar to the ontology-based approach, the rule-based approach was used to achieve the purposes of (1)-(3) and (5)-(7). Due to the reusability of the transformation rules, it was also possible to simplify the maintenance of the ETL/ELT process by applying rules in the process [37]. Other MDD approaches such as template-based [58,62], message-based [51], and metadata-based [41,44,48] were used to achieve the goals of (1)-(3) and (5)-(7). A metadata-based approach (eg, metadata management framework) can be used to develop the ETL tasks automatically [60]. The detailed information of Table 2 is available in the Multimedia Appendix 1.

**Table 2.** Relationships between purposes and MDD<sup>a</sup> approaches used.

Purposes		MDD approaches		
Number	Description	Ontology-based, n/N (%)	Rule-based, n/N (%)	Other, n/N (%)
(1)	To automate the data transformation within an ETL <sup>b</sup> /ELT <sup>c</sup> process	2/6 (33)	3/6 (50)	1/6 (17)
(2)	To reuse an ETL/ELT process (partially or completely)	1/4 (25)	2/4 (50)	1/4 (25)
(3)	To improve the performance of an ETL/ELT process	0/2 (0)	1/2 (50)	1/2 (50)
(4)	To develop a generic ETL/ELT process	3/3 (100)	0/3 (0)	0/3 (0)
(5)	To develop an ETL/ELT process automatically	5/9 (56)	2/9 (22)	2/9 (22)
(6)	To develop a new ETL/ELT process (without any other specific purposes)	4/6 (67)	1/6 (17)	1/6 (17)
(7)	Other	5/11 (45)	2/11 (18)	4/11 (36)

<sup>a</sup>MDD: metadata-driven.

<sup>b</sup>ETL: Extract-Transform-Load.

<sup>c</sup>ELT: Extract-Load-Transform.

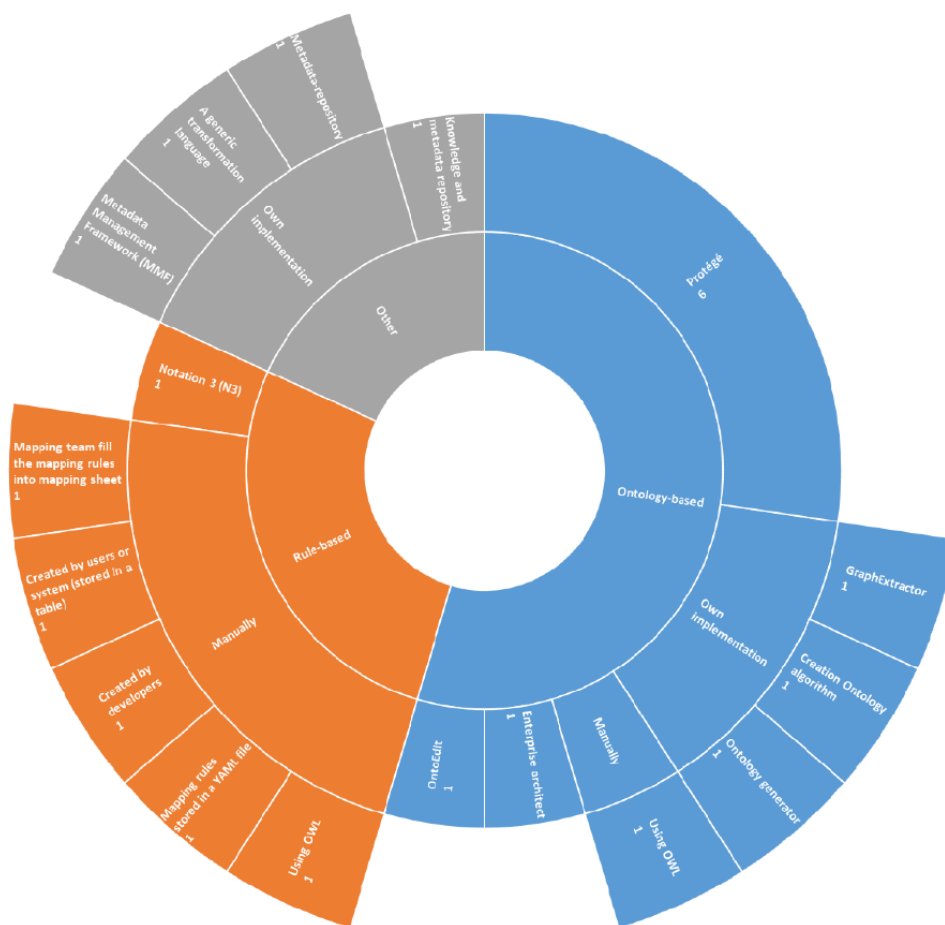
### Tools Used for Implementing MDD Approaches

Finally, we focused on the tools used to implemented MDD approaches (Q4). For achieving various purposes as shown in the previous section, different tools were used. As shown in Figure 4, each type of MDD approach can be implemented by using either an existing tool or a use case specific tool. Based on the included publications, the ontology-base approaches were mostly implemented using Protégé (Stanford Center for Biomedical Informatics Research) [64]. Protégé is an ontology editor, as well as OntoEdit (Institute AIFB, University of Karlsruhe and Ontoprise GmbH) [65]. The main reason for using an ontology editor is its ease of use and maintenance, as well as the various plug-ins. The use of case specific tools, such as ontology generator introduced by Kamil et al [45], generated ontologies based on the data definition language of the relational database. Both types of tools were used for creating and maintaining the ontology, which was then used to establish a

generic mapping logic in the ETL/ELT process [32,50,52,54,55,61]. Another type of frequently used MDD approach is rule-based, which is used for phrasing and storing the transformation rules. The transformation rules can be stored in a mapping sheet [49], a CSV file [34], a YAML (YAML Ain't Markup Language) file [33] or a table within a database [43], which were implemented manually. Afterwards, the transformation rules could be used in the ETL/ELT process, for example, to enable the automatic transformation. Other types of MDD approaches can also be implemented by using existing tools (eg, knowledge and metadata repository [66]) or use case specific tools (eg, metadata repository [41] and metadata management framework [60]). For example, Ozyurt and Grethe [36] implemented a generic transformation language using the bioCADDIE Data Tag Suite (bioCADDIE Project) [67] (a metadata schema) to align heterogeneous data from multiple sources, which provided a basis for further analytic queries.



**Figure 4.** Tools used for developing the metadata-driven approach. MMF: metadata management framework; OWL: Web Ontology Language; YAML: YAML Ain't Markup Language.



## Discussion

### Principal Findings

Our literature review on the topic “metadata-driven ETL/ELT” includes all publications listed on PubMed, IEEE Explore, Web of Science, and Biomed Center on MDD ETL/ELT process from 2012 to 2022. In some context, the use of metadata is represented specifically using “ontology” or “rules.” Therefore, we added “ontology” and “rules” into the search string to expand the search range.

With the review process presented, we were able to provide an overview of the thematic categories to which the MDD ETL/ELT processes were applied (Q1), the types of MDD approaches used in the ETL/ELT processes (Q2), the purposes of using MDD approaches (Q3), as well as the tools used to implement the MDD approaches (Q4).

Across all thematic categories, ontology-based and rule-based approaches are the most used approaches in the data warehouse and the medical thematic categories. In some cases, more than one MDD approach was used in the ETL/ELT process. For example, Del Carmen Legaz-García et al [39] used both ontology-based and rule-based approaches. Therefore, such publications were categorized as both MDD approach types.

Various tools can be used to implement MDD approaches. Unfortunately, we were not able to extract this information from

all included publications. The reason for that is that some publications used proprietary or nontransferable approaches (eg, data-specific ontologies [39,62] and rules from Data Vault [DataVaultAlliance] [42]). Some other publications did not explicitly mention or describe the tools they used. Therefore, these publications were not included in the analysis of MDD tools used.

The results indicate that it is promising to implement a generic ETL/ELT process to transform different FHIR profiles to OMOP CDM automatically by applying MDD approaches. However, the results do not provide a trivial solution for this. For example, Huang et al [63] used an ontology-based approach to be able to automate the data transformation in an ETL/ELT process, while Ong et al [34] used a rule-based approach to achieve the same purpose. In some cases, more than one MDD approach were used as complements in order to accomplish the data transformation. For example, Pacaci et al [37] chose an ontology-based approach to automate the data transformation and a rule-based to simplify the maintenance of the transformation process in case of changes in data sources. By applying these 2 approaches in combination, the authors were able to transform EHR data from heterogeneous EHR systems into OMOP CDM. Therefore, determining an appropriate MDD approach and tool to implement a generic ETL/ELT process to transform FHIR to OMOP CDM automatically remains a challenge.

This work aimed to provide an overview of different types of MDD approaches and their tools. Consequently, this review lacks an analysis of detailing the specific traits of each MDD approach. This gap underscores the importance of providing a comprehensive insight into the characterizations of the MDD approaches presented in this study. This analysis will be conducted in the future to provide solid evidence for selecting the most suitable MDD approach and tool, or for considering using multiple MDD approaches in combination to implement the generic ETL/ELT process for transforming FHIR to OMOP CDM.

## Conclusions

Our literature review shows that using MDD approaches to develop an ETL/ELT process can serve different purposes in

different focus groups (ie, medicine, data warehouse, big data, industry, geoinformatics, archaeology, and military). The results show that it is promising to implement an ETL/ELT process by applying MDD approach for automating the data transformation from FHIR to OMOP CDM. However, the determination of an appropriate MDD approach and tool to implement such an ETL/ELT process remains a challenge. This is due to the lack of comprehensive insight into the characterizations of the MDD approaches presented in this study. Therefore, our next step is to evaluate the MDD approaches presented in this study and to determine the most appropriate MDD approaches and the way of integrating them into the MII CDS FHIR to OMOP CDM ETL process [8]. This could verify the ability of using MDD approaches to generalize the ETL process for harmonizing medical data [11].

## Acknowledgments

This publication was partially funded by the German Federal Ministry of Education and Research (BMBF) Network of University Medicine 2.0: "NUM 2.0", Grant No. 01KX2121, Project: NUM-Data integration center – NUM-DIZ. The Article Processing Charge was funded by the joint publication funds of the Technische Universität, Dresden, including the Carl Gustav Carus Faculty of Medicine, and the Sächsische Landesbibliothek—Staats- und Universitätsbibliothek, Dresden, as well as the Open Access Publication Funding of the Deutsche Forschungsgemeinschaft.

## Authors' Contributions

All authors contributed substantially to this work. YP did the search string definition and publications for the review-process preparation. YP, FB, Robert G, AH, EH, DK, SL, and AV: screened the title and abstract. YP, FB, Richard G, Robert G, AH, EH, DK, SL, and AV screened the full text. YP did the data extraction. FB, DK, Robert G, and SL performed the data extraction validation. YP wrote the original draft. YP, FB, Richard G, Robert G, AH, EH, DK, SL, AV, and MS reviewed and edited the writing. MS handled the resources. All authors have read and agreed to the current version of the paper and take responsibility for the scientific integrity of the work.

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

Excel tables for extracted data from included publications.

[[XLSX File \(Microsoft Excel File\), 462 KB-Multimedia Appendix 1](#)]

## Multimedia Appendix 2

PRISMA-ScR checklist.

[[DOCX File , 85 KB-Multimedia Appendix 2](#)]

## References

1. Liu F, Panagiotakos D. Real-world data: a brief review of the methods, applications, challenges and opportunities. *BMC Med Res Methodol*. 2022;22(1):287. [[FREE Full text](#)] [doi: [10.1186/s12874-022-01768-6](https://doi.org/10.1186/s12874-022-01768-6)] [Medline: [36335315](#)]
2. 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. [[FREE Full text](#)] [doi: [10.1016/j.jbi.2016.10.016](https://doi.org/10.1016/j.jbi.2016.10.016)] [Medline: [27989817](#)]
3. European Medicines Agency. URL: <https://www.ema.europa.eu/en> [accessed 2022-08-18]
4. Data Analysis and Real World Interrogation Network (DARWIN EU). 2021. URL: <https://www.darwin-eu.org/> [accessed 2023-12-16]
5. The Book of OHDSI: Observational Health Data Sciences and Informatics. San Bernardino, CA. OHDSI; 2019. URL: <https://ohdsi.github.io/TheBookOfOhdsi> [accessed 2024-01-19]
6. 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;216:574-578. [[FREE Full text](#)] [Medline: [26262116](#)]

7. Klann JG, Joss MAH, Embree K, Murphy SN. Data model harmonization for the all of us research program: transforming i2b2 data into the OMOP common data model. *PLoS One*. 2019;14(2):e0212463. [FREE Full text] [doi: [10.1371/journal.pone.0212463](https://doi.org/10.1371/journal.pone.0212463)] [Medline: [30779778](https://pubmed.ncbi.nlm.nih.gov/30779778/)]
8. 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. *Int J Med Inform*. 2023;169:104925. [FREE Full text] [doi: [10.1016/j.ijmedinf.2022.104925](https://doi.org/10.1016/j.ijmedinf.2022.104925)] [Medline: [36395615](https://pubmed.ncbi.nlm.nih.gov/36395615/)]
9. Zoch M, Henke E, Reinecke I, Peng Y, Gebler R, Gruhl M, et al. Extract, transform and load German claim data to OMOP CDM—design and implications. *Ger Medical Sci*. 2022;153. [FREE Full text] [doi: [10.3205/22gmds057](https://doi.org/10.3205/22gmds057)]
10. Klann JG, Abend A, Raghavan VA, Mandl KD, Murphy SN. Data interchange using i2b2. *J Am Med Inform Assoc*. 2016;23(5):909-915. [FREE Full text] [doi: [10.1093/jamia/ocv188](https://doi.org/10.1093/jamia/ocv188)] [Medline: [26911824](https://pubmed.ncbi.nlm.nih.gov/26911824/)]
11. Peng Y, Henke E, Sedlmayr M, Bathelt F. Towards ETL Processes to OMOP CDM Using Metadata and Modularization. *Stud Health Technol Inform*. 2023;302:751-752. [FREE Full text] [doi: [10.3233/SHTI230256](https://doi.org/10.3233/SHTI230256)] [Medline: [37203486](https://pubmed.ncbi.nlm.nih.gov/37203486/)]
12. FHIR v4.0.1. HL7 International. URL: <https://www.hl7.org/fhir/> [accessed 2022-04-05]
13. Kabachinski J. What is Health Level 7? *Biomed Instrum Technol*. 2006;40(5):375-379. [FREE Full text] [doi: [10.2345/0899-8205-40-5-375.1](https://doi.org/10.2345/0899-8205-40-5-375.1)] [Medline: [17078369](https://pubmed.ncbi.nlm.nih.gov/17078369/)]
14. Semler SC, Wissing F, Heyder R. German Medical Informatics Initiative. *Methods Inf Med*. 2018;57(S 01):e50-e56. [FREE Full text] [doi: [10.3414/ME18-03-0003](https://doi.org/10.3414/ME18-03-0003)] [Medline: [30016818](https://pubmed.ncbi.nlm.nih.gov/30016818/)]
15. Ganslandt T, Boeker M, Löbe M, Prasser F, Schepers J, Semler SC, et al. Der Kerndatensatz der Medizininformatik-Initiative Ein Schritt zur Sekundärnutzung von Versorgungsdaten auf nationaler Ebene. *Forum der Medizin-Dokumentation und Medizin-Informatik*. 2018;20(1):17-21.
16. The National Association of Statutory Health Insurance Physicians and the regional Associations of Statutory Health Insurance Physicians. *Kassenärztliche Bundesvereinigung*. 2020. URL: [https://www.kbv.de/html/about\\_us.php](https://www.kbv.de/html/about_us.php) [accessed 2023-08-01]
17. Medizinische Informationsobjekte (MIO). *Kassenärztliche Bundesvereinigung*. 2021. URL: <https://www.kbv.de/html/mio.php> [accessed 2023-08-01]
18. Leitfaden Basis DE (R4). HL7 FHIR Implementierungsleitfäden. URL: <https://ig.fhir.de/basisprofile-de/stable/Home.html> [accessed 2023-08-01]
19. Koch M, Richter J, Hauswaldt J, Krefting D. How to Make Outpatient Healthcare Data in Germany Available for Research in the Dynamic Course of Digital Transformation. *Stud Health Technol Inform*. 2023;307:12-21. [FREE Full text] [doi: [10.3233/SHTI230688](https://doi.org/10.3233/SHTI230688)] [Medline: [37697833](https://pubmed.ncbi.nlm.nih.gov/37697833/)]
20. US Core implementation guide. HL7 International. URL: <https://www.hl7.org/fhir/us/core/> [accessed 2022-12-16]
21. System URL for ICD-10-GM. *Fast Healthcare Interoperability Resources*. URL: <http://fhir.de/CodeSystem/dimdi/icd-10-gm> [accessed 2023-12-30]
22. System URL for ICD-10-CM. HL7 International. URL: <http://hl7.org/fhir/sid/icd-10-cm> [accessed 2023-12-30]
23. OMOPonFHIR Project. URL: <https://omoponfhir.org/> [accessed 2022-04-05]
24. Loshin D. Chapter 9—metadata. In: Loshin D, editor. *Business Intelligence: The Savvy Manager's Guide*, 2nd Edition. Waltham, MA. Morgan Kaufmann; 2013;119-130.
25. Ulrich H, Kock-Schoppenhauer A, Deppenwiese N, Gött R, Kern J, Lablans M, et al. Understanding the nature of metadata: systematic review. *J Med Internet Res*. 2022;24(1):e25440. [FREE Full text] [doi: [10.2196/25440](https://doi.org/10.2196/25440)] [Medline: [35014967](https://pubmed.ncbi.nlm.nih.gov/35014967/)]
26. ISO/IEC TR 19583-1:2019: information technology: concepts and usage of metadata—part 1: metadata concepts. *International Organization for Standardization*. 2019. URL: <https://www.iso.org/standard/67365.html> [accessed 2023-05-15]
27. Breitman KK, Casanova MA, Truszkowski W. *Semantic Web: Concepts, Technologies and Applications*. London. Springer; 2007.
28. Moher D, Liberati A, Tetzlaff J, Altman D, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med*. 2009;6(7):e1000097. [FREE Full text] [doi: [10.1371/journal.pmed.1000097](https://doi.org/10.1371/journal.pmed.1000097)] [Medline: [19621072](https://pubmed.ncbi.nlm.nih.gov/19621072/)]
29. Zotero. 2022. URL: <https://www.zotero.org/> [accessed 2022-02-10]
30. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan—a web and mobile app for systematic reviews. *Syst Rev*. 2016;5(1):210. [FREE Full text] [doi: [10.1186/s13643-016-0384-4](https://doi.org/10.1186/s13643-016-0384-4)] [Medline: [27919275](https://pubmed.ncbi.nlm.nih.gov/27919275/)]
31. 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 Syst Rev*. 2022;18(2):e1230. [FREE Full text] [doi: [10.1002/cl2.1230](https://doi.org/10.1002/cl2.1230)] [Medline: [36911350](https://pubmed.ncbi.nlm.nih.gov/36911350/)]
32. Post AR, Pai AK, Willard R, May BJ, West AC, Agravat S, et al. Metadata-driven clinical data loading into i2b2 for clinical and translational science institutes. *AMIA Jt Summits Transl Sci Proc*. 2016;2016:184-193. [FREE Full text] [Medline: [27570667](https://pubmed.ncbi.nlm.nih.gov/27570667/)]
33. Quiroz JC, Chard T, Sa Z, Ritchie A, Jorm L, Gallego B. Extract, transform, load framework for the conversion of health databases to OMOP. *PLoS One*. 2022;17(4):e0266911. [FREE Full text] [doi: [10.1371/journal.pone.0266911](https://doi.org/10.1371/journal.pone.0266911)] [Medline: [35404974](https://pubmed.ncbi.nlm.nih.gov/35404974/)]

34. Ong TC, Kahn MG, Kwan BM, Yamashita T, Brandt E, Hosokawa P, et al. Dynamic-ETL: a hybrid approach for health data extraction, transformation and loading. *BMC Med Inform Decis Mak.* 2017;17(1):134. [FREE Full text] [doi: [10.1186/s12911-017-0532-3](https://doi.org/10.1186/s12911-017-0532-3)] [Medline: [28903729](https://pubmed.ncbi.nlm.nih.gov/28903729/)]
35. Mate S, Köpcke F, Toddenroth D, Martin M, Prokosch H, Bürkle T, et al. Ontology-based data integration between clinical and research systems. *PLoS One.* 2015;10(1):e0116656. [FREE Full text] [doi: [10.1371/journal.pone.0116656](https://doi.org/10.1371/journal.pone.0116656)] [Medline: [25588043](https://pubmed.ncbi.nlm.nih.gov/25588043/)]
36. Ozyurt IB, Grethe JS. Foundry: a message-oriented, horizontally scalable ETL system for scientific data integration and enhancement. *Database (Oxford).* 2018;2018:bay130. [FREE Full text] [doi: [10.1093/database/bay130](https://doi.org/10.1093/database/bay130)] [Medline: [30576493](https://pubmed.ncbi.nlm.nih.gov/30576493/)]
37. Pacaci A, Gonul S, Sinaci AA, Yuksel M, Erturkmen GBL. A semantic transformation methodology for the secondary use of observational healthcare data in postmarketing safety studies. *Front Pharmacol.* 2018;9:435. [FREE Full text] [doi: [10.3389/fphar.2018.00435](https://doi.org/10.3389/fphar.2018.00435)] [Medline: [29760661](https://pubmed.ncbi.nlm.nih.gov/29760661/)]
38. Haarbrandt B, Tute E, Marscholke M. Automated population of an i2b2 clinical data warehouse from an openEHR-based data repository. *J Biomed Inform.* 2016;63:277-294. [FREE Full text] [doi: [10.1016/j.jbi.2016.08.007](https://doi.org/10.1016/j.jbi.2016.08.007)] [Medline: [27507090](https://pubmed.ncbi.nlm.nih.gov/27507090/)]
39. Del Carmen Legaz-García M, Miñarro-Giménez JA, Menárguez-Tortosa M, Fernández-Breis JT. Generation of open biomedical datasets through ontology-driven transformation and integration processes. *J Biomed Semantics.* 2016;7:32. [FREE Full text] [doi: [10.1186/s13326-016-0075-z](https://doi.org/10.1186/s13326-016-0075-z)] [Medline: [27255189](https://pubmed.ncbi.nlm.nih.gov/27255189/)]
40. Gang H, Jin-Rong L, Xiu-Ying W. A kind of bidirectional mapping strategy of heterogeneous data model based on metadata-driven. Presented at: Proceedings of 2012 2nd International Conference on Computer Science and Network Technology; December 29-31, 2012, 2012;1023-1027; Changchun, China. [doi: [10.1109/iccnsnt.2012.6526100](https://doi.org/10.1109/iccnsnt.2012.6526100)]
41. Chen Z, Zhao T. A new tool for ETL process. Presented at: 2012 International Conference on Image Analysis and Signal Processing; November 09-11, 2012, 2012;269-273; Huangzhou, China. [doi: [10.1109/IASP.2012.6425038](https://doi.org/10.1109/IASP.2012.6425038)]
42. Puonti M, Raitalaakso T, Aho T, Mikkonen T. Automating transformations in data vault data warehouse loads. In: Thalheim B, Jaakkola H, Kiyoki Y, Yoshida N, editors. *Information Modelling and Knowledge Bases XXVIII.* Amsterdam. IOS Press; 2017;215-230.
43. Wang H, Zhang J, Guo J. Constructing data warehouses based on operational metadata-driven builder pattern. Presented at: 2015 International Conference on Logistics, Informatics and Service Sciences (LISS); July 27-29, 2015, 2015;1-4; Barcelona, Spain. [doi: [10.1109/liss.2015.7369630](https://doi.org/10.1109/liss.2015.7369630)]
44. Tomingas K, Kliimask M, Tammet T. Data integration patterns for data warehouse automation. In: Vakali A, Trajcevski G, Kon-Popovska M, Ivanovic M, Bassiliades N, Palpanas T, et al, editors. *New Trends in Database and Information Systems II.* Berlin. Springer; 2015;41-55.
45. Kamil I, Ingriani MM, Asnar YDW. Data migration helper using domain information. Presented at: 2014 International Conference on Data and Software Engineering (ICODSE); November 26-27, 2014, 2014;1-6; Bandung, Indonesia. [doi: [10.1109/icodse.2014.7062492](https://doi.org/10.1109/icodse.2014.7062492)]
46. Dhaouadi A, Bousselmi K, Gammoudi MM, Monnet S, Hammoudi S. Data warehousing process modeling from classical approaches to new trends: main features and comparisons. *Data.* 2022;7(8):113. [FREE Full text] [doi: [10.3390/data7080113](https://doi.org/10.3390/data7080113)]
47. Berkani N, Khouri S, Bellatreche L. Generic methodology for semantic data warehouse design: from schema definition to ETL. Presented at: 2012 Fourth International Conference on Intelligent Networking and Collaborative Systems; September 19-21, 2012, 2012;404-411; Bucharest, Romania. [doi: [10.1109/incos.2012.108](https://doi.org/10.1109/incos.2012.108)]
48. Nath RPD, Romero O, Pedersen TB, Hose K. High-level ETL for semantic data warehouses. *Semant Web.* 2022;13(1):85-132. [FREE Full text] [doi: [10.3233/sw-210429](https://doi.org/10.3233/sw-210429)]
49. Yu QC. Metadata driven data mapper development. *Appl Mech Mater.* 2013;411-414:403-407. [FREE Full text] [doi: [10.4028/www.scientific.net/amm.411-414.403](https://doi.org/10.4028/www.scientific.net/amm.411-414.403)]
50. Ta'a A, Abdullah MS. Ontology development for ETL process design. In: Ahmad MN, Abdullah MS, Colomb RM, editors. *Ontology-based Applications for Enterprise Systems and Knowledge Management.* Hershey, PA. Information Science Reference; 2013;261-275.
51. Novak M, Kermek D, Magdalenic I. Proposed architecture for ETL workflow generator. Presented at: Proceedings of the Central European Conference on Information and Intelligent Systems; October 2-4, 2019, 2019;297-304; Varaždin, Croatia.
52. Berkani N, Bellatreche L, Khouri S. Towards a conceptualization of ETL and physical storage of semantic data warehouses as a service. *Cluster Comput.* 2013;16(4):915-931. [doi: [10.1007/s10586-013-0266-7](https://doi.org/10.1007/s10586-013-0266-7)]
53. Hilali I, Arfaoui N, Ejbali R. A new approach for integrating data into big data warehouse. Presented at: Proceedings Volume 12084, Fourteenth International Conference on Machine Vision (ICMV 2021); March 4, 2022, 2022;120841M; Rome, Italy. [doi: [10.1117/12.2623069](https://doi.org/10.1117/12.2623069)]
54. Bansal SK, Kagemann S. Integrating big data: a semantic extract-transform-load framework. *Computer.* 2015;48(3):42-50. [doi: [10.1109/mc.2015.76](https://doi.org/10.1109/mc.2015.76)]
55. Bansal SK. Towards a semantic Extract-Transform-Load (ETL) framework for big data integration. Presented at: 2014 IEEE International Congress on Big Data; June 27-July 02, 2014, 2014;522-529; Anchorage, AK, USA. [doi: [10.1109/bigdata.congress.2014.82](https://doi.org/10.1109/bigdata.congress.2014.82)]



56. Boulahia C, Behja H, Louhdi MRC. Towards semantic ETL for integration of textual scientific documents in a big data environment: a theoretical approach. Presented at: 2020 6th IEEE Congress on Information Science and Technology (CiSt); June 05-12, 2021, 2020;133-138; Agadir-Essaouira, Morocco. [doi: [10.1109/cist49399.2021.9357280](https://doi.org/10.1109/cist49399.2021.9357280)]
57. de Cesare C, Foy G, Lycett M. 4D-SETL a semantic data integration framework. Presented at: Proceedings of the 18th International Conference on Enterprise Information Systems—Volume 1: ICEIS; April 25-28, 2016, 2016;127-134; Rome, Italy. [doi: [10.5220/0005822501270134](https://doi.org/10.5220/0005822501270134)]
58. McCarthy S, McCarren A, Roantree M. A method for automated transformation and validation of online datasets. Presented at: 2019 IEEE 23rd International Enterprise Distributed Object Computing Conference (EDOC); October 28-31, 2019, 2019;183-189; Paris, France. [doi: [10.1109/edoc.2019.00030](https://doi.org/10.1109/edoc.2019.00030)]
59. Scriney M, McCarthy S, McCarren A, Cappellari P, Roantree M. Automating data mart construction from semi-structured data sources. *Comput J*. 2019;62(3):394-413. [doi: [10.1093/comjnl/bxy064](https://doi.org/10.1093/comjnl/bxy064)]
60. Suleykin A, Panfilov P. Metadata-driven industrial-grade ETL system. Presented at: 2020 IEEE International Conference on Big Data (Big Data); December 10-13, 2020, 2020;2433-2442; Atlanta, GA, USA. [doi: [10.1109/bigdata50022.2020.9378367](https://doi.org/10.1109/bigdata50022.2020.9378367)]
61. Janecka K, Cerba O, Jedlicka K, Jezek J. Towards interoperability of spatial planning Data: 5-Steps harmonization framework. Presented at: 13th SGEM GeoConference on Informatics, Geoinformatics and Remote Sensing; June 16-22, 2013, 2013;1005-1016; Albena, Bulgaria. [doi: [10.5593/SGEM2013/BB2.V1/S11.051](https://doi.org/10.5593/SGEM2013/BB2.V1/S11.051)]
62. Binding C, Charno M, Jeffrey S, May K, Tudhope D. Template based semantic integration: from legacy archaeological datasets to linked data. *Int J Semantic Web Inf Syst*. 2015;11(1):1-29. [doi: [10.4018/ijswis.2015010101](https://doi.org/10.4018/ijswis.2015010101)]
63. Huang DM, Du YL, Zhang MH, Zhang C. Application of ontology-based automatic ETL in marine data integration. Presented at: 2012 IEEE Symposium on Electrical & Electronics Engineering (EEESYM); June 24-27, 2012, 2012;11-13; Kuala Lumpur. [doi: [10.1109/eesym.2012.6258574](https://doi.org/10.1109/eesym.2012.6258574)]
64. Musen MA, Protégé Team. The Protégé project: a look back and a look forward. *AI Matters*. 2015;1(4):4-12. [FREE Full text] [doi: [10.1145/2757001.2757003](https://doi.org/10.1145/2757001.2757003)] [Medline: [27239556](https://pubmed.ncbi.nlm.nih.gov/27239556/)]
65. Sure Y, Angele J, Staab S. OntoEdit: multifaceted inferencing for ontology engineering. In: Aberer K, March S, Spaccapietra S, editors. *Journal on Data Semantics I*. LNCS 2800. Verlag Berlin. Springer; 2003;128-152.
66. MMX metadata framework. Mindworks Industries. URL: [https://www.mindworks.industries/mmx\\_framework.html](https://www.mindworks.industries/mmx_framework.html) [accessed 2023-11-21]
67. Sansone SA, Gonzalez-Beltran A, Rocca-Serra P, Alter G, Grethe JS, Xu H, et al. DATS, the data tag suite to enable discoverability of datasets. *Sci Data*. 2017;4:170059. [FREE Full text] [doi: [10.1038/sdata.2017.59](https://doi.org/10.1038/sdata.2017.59)] [Medline: [28585923](https://pubmed.ncbi.nlm.nih.gov/28585923/)]

## Abbreviations

**CDM:** Common Data Model

**CDS:** Core Data Set

**DARWIN EU:** Data Analysis and Real World Interrogation Network European Union

**EHR:** electronic health record

**ELT:** Extract-Load-Transform

**ETL:** Extract-Transform-Load

**FHIR:** Fast Healthcare Interoperability Resources

**FTS:** full-text screening

**HL7:** Health Level 7

**KBV:** The National Association of Statutory Health Insurance Physicians (German: Kassenärztliche Bundesvereinigung)

**MDD:** metadata-driven

**MII:** Medical Informatics Initiative

**OMOP:** Observational Medical Outcomes Partnership

**PRISMA:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses

**TAS:** Title-Abstract-Screening

**YAML:** YAML Ain't Markup Language

*Edited by C Lovis; submitted 20.09.23; peer-reviewed by M Löbe, W Xu; comments to author 24.10.23; revised version received 01.12.23; accepted 03.12.23; published 14.02.24*

*Please cite as:*

*Peng Y, Bathelt F, Gebler R, Gött R, Heidenreich A, Henke E, Kadioglu D, Lorenz S, Vengadeswaran A, Sedlmayr M*

*Use of Metadata-Driven Approaches for Data Harmonization in the Medical Domain: Scoping Review*

*JMIR Med Inform 2024;12:e52967*

*URL: <https://medinform.jmir.org/2024/1/e52967>*

*doi: [10.2196/52967](https://doi.org/10.2196/52967)*

*PMID: [38354027](https://pubmed.ncbi.nlm.nih.gov/38354027/)*

©Yuan Peng, Franziska Bathelt, Richard Gebler, Robert Gött, Andreas Heidenreich, Elisa Henke, Dennis Kadioglu, Stephan Lorenz, Abishaa Vengadeswaran, Martin Sedlmayr. Originally published in JMIR Medical Informatics (<https://medinform.jmir.org>), 14.02.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Informatics, is properly cited. The complete bibliographic information, a link to the original publication on <https://medinform.jmir.org/>, as well as this copyright and license information must be included.