Engineering and Technology Journal e-ISSN: 2456-3358

Volume 11 Issue 02 February-2026, Page No.-8914-8920

DOI: 10.47191/etj/v11i02.15, I.F. – 8.482

Development of Minimum Data Set and Data Model in Electronic Health Record System: Literature Review Article

Teddie Darmizal

Department of Informatics Engineering, Faculty of Science and Technology, UIN Sultan Syarif Kasim Riau, Indonesia

ABSTRACT: Electronic Health Record Systems (EHRs) require well-defined Minimum Data Sets (MDS) and robust data models to ensure data quality, interoperability, and effective system implementation. However, existing studies on EHR-related MDS and data modeling remain fragmented, often focusing on disease-specific or domain-specific contexts and stopping at conceptual or architectural levels. This systematic literature review aims to analyze and synthesize existing research on the development of MDS and data models for EHR systems, identifying prevailing approaches, limitations, and research gaps. A Systematic Literature Review (SLR) was conducted using the PICOC framework and a structured search strategy applied to academic databases, focusing on studies published within the last ten years. The review examined research related to EHR data sets, data elements, and data models, including conceptual, logical, and physical modeling approaches. The findings indicate that most MDS studies successfully standardize administrative and clinical data elements through expert consensus methods but lack comprehensive metadata specifications and integration with complete data modeling processes. Similarly, many data modeling studies emphasize conceptual or reference models (e.g., openEHR, HL7 RIM, ISO-based frameworks) without sufficient progression to logical and physical implementations or empirical validation in real clinical environments. In addition, limited adoption of newer interoperability standards, such as HL7 FHIR, and insufficient centralized metadata governance further hinder large-scale interoperability and reuse. This review highlights the need for a holistic, workflow-based MDS integrated with complete conceptual, logical, and physical data models to bridge the gap between theory and practice in EHR development. The study provides a consolidated reference for researchers, system developers, and policymakers in designing standardized, interoperable, and scalable EHR systems, particularly in contexts lacking national EHR data standards.

KEYWORDS: Minimum data Set**,** Data Model, Electronic Health Record System

INTRODUCTION

An Electronic Health Record System (EHR) is a data repository regarding the health and healthcare of a subject of care where all information is stored on electronic media' (ISO, 2011). Furthermore, EHR will be defined as an electronic application through which individuals can access, manage, and share their health information, and that of others for whom they are authorized, in a private, secure, and confidential environment (Plastiras & O'Sullivan, 2018).

Developing an Electronic Health Record (EHR) system requires a comprehensive database architecture that clearly defines the data set and data model to be implemented. The data set must represent all data needs in every business process [3]. Meanwhile, the data model must represent the data set in a more technical form, on the conceptual, logical, and physical levels of design [4], so that EHR system developers will later be guided by the existence of the EHR data model.

Data set or in medical research, usually called Minimum Data Set (MDS), is a collection of data elements to be grouped using a standard approach to allow the use of data for clinical and research purposes [5]–[8]. It specifies which data elements will be stored and how they will be stored, including their relationships and constraints [8].

A data model specifies features and relationships, such as data types, constraints, relationships, and metadata [9]. In the healthcare system, a data model is an abstract structure that organizes and standardizes data sets and data elements, defining their properties and relationships [10]. There are three levels of data models, which grow increasingly complex: conceptual, logical, and physical (Sherman, 2015).

The conceptual data model (CDM) in healthcare is to provide a high-level, abstract representation of the data that an organization uses or intends to use in its business operations. In the context of healthcare, a CDM helps bridge the knowledge gap between subject matter experts, IT architects, and designers by depicting the major business information objects and their relationships to each other using business terminology [11]. A logical data model (LDM) would define the structure of the data elements, their relationships, and the business rules that govern them. It would be used to develop a visual understanding of the data entities, attributes, and relationships specific to the EHR system. A physical data model (PDM) represents the actual

structure of a database: tables and columns, or the messages sent between computer processes. Here, the entity types usually represent tables, and the relationship type lines represent the foreign keys between tables [4], [9], [12].

This study aims to provide an overview of the development of MDS and Data Models in EHR Systems using systematic literature review techniques.

METHOD

Systematic Literature Review (SLR) is defined as the process of identifying, assessing, and interpreting all available research evidence to provide answers to specific research questions [13]. Search process uses the keywords (with a limit of the last 10 years):

TITLE-ABS-KEY (electronic AND health AND record AND data AND set OR data AND element)

The SLR process will be explained through the following figure 1:

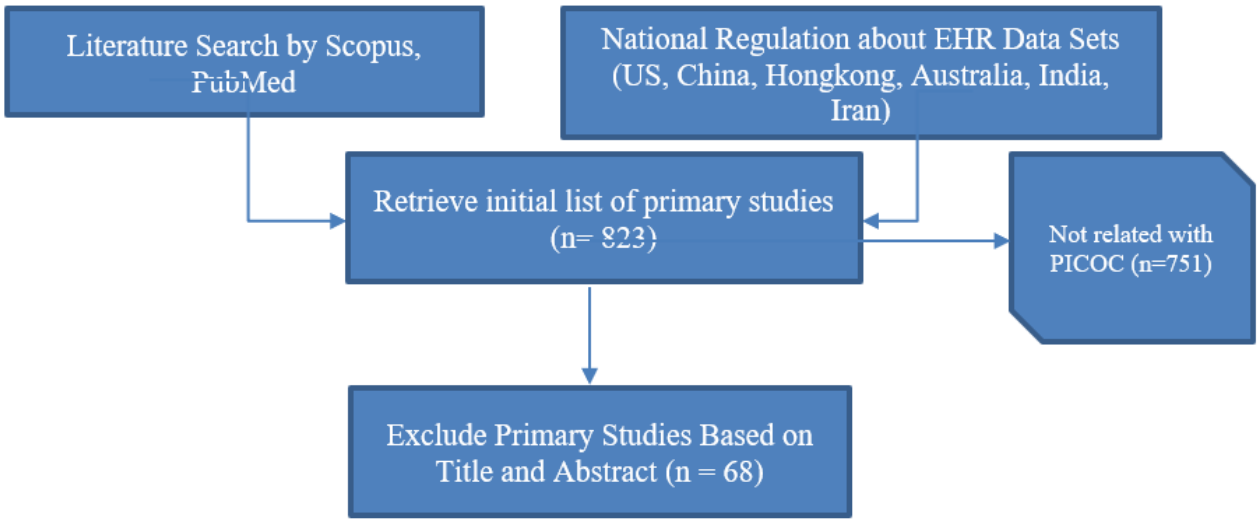

Figure 1. Systematic Literature Review (SLR) Process

In conducting research screening related to the use of the PICOC protocol, as follows:

Population (P)	Electronic health record data sets
Intervention (I)	Data sets, data elements, data model
Comparison (C)	-
Outcomes (O)	Electronic health record data sets, data elements, and data models
Context (C)	Studies in the academic and health industry

RESULTS & DISCUSSION

A. Previous Study in Minimum Data Set

The Minimum Data Set (MDS) is a set of data elements organized in a standardized way to facilitate their use in clinical and research contexts. It defines the specific data elements to be recorded, their storage methods, and the relationships and constraints among them [8]. The purpose of the Minimum Data Set, as core health data elements, is to standardize data items and their definitions [14]. A Common Data Element consists of a clearly defined question (variable) along with a specified range of responses that is shared across multiple data sets or utilized in various studies [15].

MDS is a front line in developing and implementing an information management system with the potential to improve health information and, thus, service quality [16]. The Minimum Data Set (MDS) is a tool for standardizing essential data within a specific area, enhancing information quality by defining necessary data elements consistently. As a result, MDS can ensure data quality and facilitate comparisons at both national and international levels by standardizing these elements [5].

By providing clear definitions of data elements, MDS aims to establish a shared language among all parties involved in registry and documentation. It also ensures the effective collection, analysis, reporting, and selection of critical data [17]. Additionally, MDS enhances medical history records, promotes data comparability, aids in the creation of a data repository, facilitates electronic data exchange across different healthcare systems, and ultimately improves data quality [18]. Based on their nature and use, disease data elements can be classified into administrative and clinical data. Administrative data typically encompasses demographic and socioeconomic information, as well as address, phone number, patient referral details, and the primary profession of healthcare providers [5]. In contrast, clinical data vary based on the disease type. Typically, they encompass diagnosis, medical history, laboratory results, medical imaging findings, treatment interventions, disease progression, and outcomes [7].

In the health care system, data sets identify data elements that should be collected for each patient and provide definitions for each element based on standards. Data comparison is used for different purposes, including external accreditation, evaluation of internal performance, and research and statistical studies. Therefore, it is important to

determine standards, data models, and minimum data sets to manage the clinical performance of health organizations in every country [19][20][21]. In addition to using the experiences of developed countries, considering national standards, needs, and opinions of domestic experts is essential in the development of MDS [22].

In recent years, significant research has focused on the MDS for advancing electronic health systems (EHRS). Most of these studies concentrate on developing MDS models tailored to specific diseases, injuries, or patient groups [23]. MDS for holistic health recording that combines general and specialist MDS is not widely available in the literature. Indonesia does not have an MDS standard to support the development of the EHRS in central and regional government hospitals. Thus, this study aims to develop a holistic MDS for the EHRS in Indonesia.

Research conducted by Yang et al. (2022) intended to develop a national health concept data model (HCDM) and develop a corresponding system to facilitate healthcare data standardization and centralized metadata management. Based on 55 data sets (4640 data items) from 7 health business domains in China, a bottom-up approach was employed to build the structure and metadata for HCDM by referencing HL7 RIM. According to ISO/IEC 11179, a top-down approach was used to develop and standardize the data elements. HCDM adopted a three-level architecture of class, attribute, and data type, and consisted of 6 classes and 15 sub-classes. Each class had a set of descriptive attributes, and every attribute was assigned a data type. 100 initial data elements (DEs) were extracted from HCDM, and 144 general DEs were derived from corresponding initial DEs.

Weissman et al. (2022) researched to facilitate data standardization across emerging studies. The National Heart, Lung, and Blood Institute (NHLBI) charged two groups with harmonizing data collection, and these groups collaborated to create a concise set of COVID-19 Common Data Elements (CDEs) for clinical research. The NHLBI COVID-19 CDEs are publicly available and being used for current COVID-19 clinical trials. CDEs are organized into domains, and each data element is classified within a three-tiered priority system.

Furthermore, Zarei et al. (2023) designed a minimum data set as the first step in creating a registry system for spinal canal stenosis and proposed that the MDS should be organized in two general sets of data: administrative and clinical. For the administrative data set, 40 data elements had been proposed, as five classes. Twenty-six of them were confirmed. In the clinical section, 95 data elements had been proposed in 14 classes; 94 of which were finally confirmed.

Similarly, Soleimani et al. (2021) developed a minimum data set for an information management system for aged care centers in Iran. Administrative and clinical data elements were finalized with 36 and 495 data elements using two and three Delphi rounds, respectively. The data sets of Iranian aged care centers could be a reliable tool for data standardization, leading to improved quality of data, thus care and service related to the elderly.

In another study, Zarei et al. (2021) conducted a study aimed at designing a minimum data set for the COVID-19 registry system. This study result is minimum data set of the COVID-19 registry system contains eight top groups including administrative (34 data elements), disease exposure (61 data elements), medical history and physical examination (138 data elements), findings of clinical diagnostic tests (101 data elements), disease progress and outcome of treatment (55 data elements), medical diagnosis and cause of death (12 data elements), follow-up (14 data elements), and COVID-19 vaccination (19 data elements) data, respectively.

Karbasi et al. (2020) established the objective of determining the comprehensive national minimum data set for the child abuse surveillance system (CASS), which resulted in 147 data elements being included in the Delphi survey. The data elements of the CASS were classified into seven categories as follows: demographic data, incident-related data, medical history, diagnostic tests, incident nature, therapeutic measures, and other required data.

Moreover, M. Ahmadi et al. (2019) developed a minimum data set of the information management system for disability in Iran, resulting in administrative and clinical categories with 130 and 345 data elements, respectively. 200 and 38 data elements were mandatory elements (administrative: 60, clinical: 178), and the rest were optional elements.

In addition, Kock-Schoppenhauer et al. (2019) conducted research to support the identification, collection, and provision of metadata in a predefined structured manner to foster consolidation. The Meta Data Repositories (MDR) Sheet is embedded in the processing pipeline MDR Bridge and delivers metadata in the CDISC ODM format for further use in MDRs. This approach is used to interactively unify core datasets, import existing standard datasets, and automatically extract all defined data elements from source systems.

Safdari et al. (2018) also researched developing a national antimicrobial resistance surveillance minimum data set. This research obtained 80 data elements in 8 axes. The resistance data categories comprised basic, clinical, electronic reporting, infection control, microbiology, pharmacy, World Health Organization-derived, and expert-recommended data. Relevance coding was extracted based on the Iranian electronic health record coding system.

Additionally, Jiang et al. (2016) developed and evaluated a data element repository (DER) for providing machine-readable Quality Data Model (QDM) data element service APIs to support phenotype algorithm authoring and execution, used the ISO/IEC 11179 metadata standard to capture the structure for each data element, and leveraged Semantic Web technologies to facilitate semantic representation of these metadata. This research proposed a harmonization with the models developed in HL7 Fast Healthcare Interoperability Resources (FHIR) and Clinical Information Modeling Initiatives (CIMI) to enhance the Quality Data Model (QDM)

specification and enable the extensibility and better coverage of the DER. This research also compared the DER with the existing QDM implementation utilized within the Measure Authoring Tool (MAT) to demonstrate the scalability and extensibility of our DER-based approach.

B. Previous Study in Data Model

Models are abstractions of real systems that are key artifacts during the engineering of software products [31]. Various data modeling paradigms are used at different levels of abstraction in the area of Data Management. Two main paradigms are used at the conceptual level: (i) Models based on entities (objects) and relationships (associations) among them, and (ii) models based on dimensions and measurable facts (dimensional modeling) [32].

At a lower level of abstraction, currently, most applications still rely on implementations based on the relational model. However, in some specific cases, non-relational paradigms provide good performance by supporting complex nested data types (aggregates) and large-scale distributed architectures. All those non-relational models also rely on the lack of a predefined schema, which brings advantages at data insertion and disadvantages at data querying. In particular, as applications cannot query the database catalog to get the schema, the schema must be encoded in the application's code, which is not the best place to be. Furthermore, the database cannot use the information of the schema to perform query optimization [33].

A data model is a specification of the data structures and business rules representing the business requirements. The business rules define the relationship between the various data entities or objects that are required. The data model represents the data portion of those rules and the relationships between those structures. The data model's primary job is to be the data design specification for an application [4].

A logical data model (LDM) for an electronic health record system (EHRS) would define the structure of the data elements, their relationships, and the business rules that govern them. It would be used to develop a visual understanding of the data entities, attributes, and relationships specific to the EHRS. Physical data model (PDM) for an EHRS would detail how the logical model is to be implemented in a specific database management system, including the specific data types, constraints, and other implementation details [34].

Data modelling is the process of determining how data is to be stored in a database. A data model specifies features and relationships, such as data types, constraints, and metadata. In the healthcare system, a data model is an abstract structure that organizes and standardizes data sets and data elements, defining their properties and relationships [10]. Common Data Element (CDE) is a combination of a precisely defined question (variable) paired with a specified set of responses to the question that is common to multiple datasets or used across different studies [15].

According to the research by Araújo et al. (2016) with the title *A conceptual data model for health information system*, the conceptual data model (CDM) in healthcare is to provide a high-level, abstract representation of the data that an organization uses or intends to use in its business operations. In the context of healthcare, a CDM helps bridge the knowledge gap between subject matter experts, IT architects, and designers by depicting the major business information objects and their relationships to each other using business terminology. However, the proposed CDM in this research is purely conceptual and is not directly implemented or validated within operational health information systems. Practical feasibility and performance are therefore not empirically evaluated. The model is derived from literature analysis and theoretical reasoning rather than extensive real-world case studies or large-scale empirical testing, which may affect its immediate applicability. The research stops at the conceptual modelling level. There is no formal or automated approach proposed for transforming conceptual schemas into logical database schemas or physical implementations, which is crucial for practical system development.

Min et al., (2018) stated the openEHR approach provides a robust and flexible framework for modelling Electronic Health Records by clearly separating clinical knowledge from technical data structures. Through the use of archetypes and templates, the openEHR dual-model architecture enables standardized representation of clinical data while allowing continuous evolution of medical knowledge without requiring changes to underlying software components. However, The proposed openEHR modelling approach is primarily conceptual and illustrative, with limited evaluation through real-world system implementation or large-scale clinical deployment.

Yang et al., (2022) on the research *Defining National Healthcare Concept Data Model (HCDM) for China* provide a unified metadata reference for multi-source health data standardization and management. By leveraging HL7 RIM and ISO/IEC 11179, the approach enables semantic interoperability across heterogeneous health systems. The accompanying web-based system has been proven feasible in practice, supporting centralized metadata management and facilitating healthcare information standardization and interoperability at a national level. However, the development process did not incorporate newer standards such as HL7 FHIR (Fast Health Interoperability Resources), which may pose challenges for future consistency with existing standards and interoperability with other international projects.

Tiase et al., (2024) with the research title *A Scalable and Extensible Logical Data Model of EHR Audit Logs for Temporal Data Mining (called by RNteract)* stated the proposed RNteract model successfully conceptualizes nurse–EHR interactions, nurse characteristics, patient context, and outcomes in a scalable and extensible manner. RNteract is capable of supporting temporal unsupervised machine

learning and a wide range of AI-based methods for analyzing nursing documentation workflows and workload patterns. However, there is a gap between the available EHR audit log data and their use in analysis. While EHR audit logs are widely available, there is a lack of structured, nurse-centric data models that make these data usable for advanced analytics and temporal data mining

Gamal et al., (2021) adopt a broader systems-level orientation, reviewing standardized data models and database persistence mechanisms to address interoperability and scalability challenges inherent in heterogeneous and distributed healthcare environments. Their work emphasizes the role of internationally recognized standards such as ISO 13606, HL7 CDA, OpenEHR, CIMI, OMOP, SNOMED CT, and FHIR, which enable semantic interoperability through a dual-level modelling paradigm that separates reference models from domain-specific archetypes. The proposed standardized data modelling approach is primarily conceptual and architectural. However, full-scale empirical validation using large EHR datasets is limited or not yet demonstrated.

Minimum Data Set (MDS) and a comprehensive data model for Electronic Health Record Systems (EHRS) in Indonesia, representing a foundational step toward standardized digital health information management at the national level. The proposed MDS was systematically developed based on the actual medical service workflow, covering the full patient journey from admission to discharge, and aligned with Indonesian healthcare regulations, resulting in 5 administrative data classes with 28 data elements and 21 clinical data classes with 858 data elements. This extensive structure ensures that both administrative and clinical information needs are comprehensively captured. Furthermore, the transformation of the validated MDS into conceptual, logical, and physical data models provides clear technical guidance for system developers in implementing EHRS databases. This structured modeling approach enhances data consistency, supports efficient data storage, and facilitates future system scalability [38].

CONCLUSION

Based on the synthesis of previous studies on Minimum Data Sets (MDS) and data models for Electronic Health Record Systems (EHRS), several critical research gaps can be identified.

First, although numerous studies have successfully developed MDS for specific diseases, patient groups, or particular clinical domains, the literature is dominated by fragmented and disease-specific MDS implementations. There is a clear lack of a holistic and integrated MDS that combines both general (administrative) and comprehensive clinical data elements into a unified framework applicable across multiple care settings. This limitation reduces data comparability, reusability, and scalability at national and cross-institutional levels.

Second, many existing MDS initiatives focus primarily on identifying and validating data elements through expert consensus methods such as Delphi techniques. However, detailed metadata specifications (including clear class definitions, attributes, data types, and constraints aligned with international metadata standards (e.g., ISO/IEC 11179)) are often insufficiently addressed. As a result, most MDS outputs remain at the list-of-elements level and are not fully ready for systematic implementation within interoperable EHRS.

Third, in the domain of data modeling, a substantial portion of prior research remains conceptual or architectural in nature. While conceptual data models (CDM) and dual-model approaches such as openEHR, HL7 RIM, and ISO-based frameworks provide strong theoretical foundations, many studies do not progress to logical data models (LDM) or physical data models (PDM). The absence of formal or automated transformation mechanisms from conceptual models to implementable database schemas creates a gap between theory and practice, limiting real-world adoption.

Fourth, empirical validation in real clinical environments is still limited. Several proposed data models and standardized frameworks lack large-scale implementation, performance evaluation, and maintainability assessment in operational hospital systems. This gap raises concerns regarding scalability, system integration, and long-term sustainability of the proposed models.

Fifth, despite the growing emphasis on interoperability, many national and institutional data models do not fully incorporate newer interoperability standards such as HL7 FHIR, or do so only partially. This results in challenges for future-proofing EHR systems and achieving seamless data exchange across heterogeneous platforms and international initiatives.

Sixth, there is an identified gap in centralized metadata governance and operational support systems. While some studies propose conceptual repositories or metadata models, only a limited number provide web-based, operational systems that support continuous management, evaluation, and evolution of national health metadata repositories.

REFERENCES

1. C. técnico ISO, "ISO 18308:2011 Health Informatics – Requirements for an Electronic Health Record Architecture," *Heal. (San Fr.*, vol. 2011, pp. 1–27, 2011.
2. P. Plastiras and D. O'Sullivan, "Exchanging personal health data with electronic health records: A standardized information model for patient generated health data and observations of daily living," *Int. J. Med. Inform.*, vol. 120, pp. 116–125, 2018, doi: 10.1016/j.ijmedinf.2018.10.006.
3. S. Daneshvari Berry, P. J. Kroth, H. J. H. Edgar, and T. D. Warner, "Developing the Minimum Dataset for the New Mexico Decedent Image Database,"

Appl. Clin. Inform., vol. 12, no. 3, pp. 518–527, 2021, doi: 10.1055/s-0041-1730999.
4. R. Sherman, "Foundational Data Modeling," *Bus. Intell. Guideb.*, pp. 173–195, 2015, doi: 10.1016/b978-0-12-411461-6.00008-3.
5. M. Ahmadi, T. Madani, and J. Alipour, "Development a national minimum data set (MDS) of the information management system for disability in Iran," *Disabil. Health J.*, vol. 12, no. 4, pp. 641–648, 2019, doi: 10.1016/j.dhjo.2019.05.008.
6. Z. Karbasi *et al.*, "Better monitoring of abused children by designing a child abuse surveillance system: Determining national child abuse minimum data set," *Int. J. Health Plann. Manage.*, vol. 35, no. 4, pp. 843–851, 2020, doi: 10.1002/hpm.2935.
7. J. Zarei, A. Mohammadi, M. R. Akrami, and A. Jeihooni Kalhori, "Designing a minimum data set for the information management system (registry) of spinal canal stenosis: An applied-descriptive study," *Heal. Sci. Reports*, vol. 6, no. 11, 2023, doi: 10.1002/hsr2.1671.
8. F. A. Bernardi *et al.*, "The Minimum Data Set for Rare Diseases: Systematic Review," *J. Med. Internet Res.*, vol. 25, pp. 1–13, 2023, doi: 10.2196/44641.
9. C. J. Date, H. Darwen, and N. A. Lorentzos, "Database Design I: Structure," *Time Relational Theory*, pp. 197–226, 2014, doi: 10.1016/b978-0-12-800631-3.50012-5.
10. M. G. Kahn, D. Batson, and L. M. Schilling, "Data model considerations for clinical effectiveness researchers," *Med. Care*, vol. 50, no. SUPPL. 1, 2012, doi: 10.1097/MLR.0b013e318259bff4.
11. A. M. C. de Araújo, V. C. Times, and S. C. B. Soares, "A Conceptual Data Model for Health Information Systems," *Steer. Comm. World Congr. Comput. Sci. Comput. Eng. Appl. Comput.*, no. July, pp. 236–242, 2016.
12. M. West, "Some Types and Uses of Data Models," *Dev. High Qual. Data Model.*, pp. 23–36, 2011, doi: 10.1016/b978-0-12-375106-5.00003-8.
13. B. Kitchenham, O. Pearl Brereton, D. Budgen, M. Turner, J. Bailey, and S. Linkman, "Systematic literature reviews in software engineering - A systematic literature review," *Inf. Softw. Technol.*, vol. 51, no. 1, pp. 7–15, 2009, doi: 10.1016/j.infsof.2008.09.009.
14. N. Hashemi, A. Sheikhtaheri, N. sadat Hashemi, and R. Rawassizadeh, "Electronic medical records for mental disorders: What data elements should these systems contain?," *Stud. Health Technol. Inform.*, vol. 260, pp. 25–32, 2019, doi: 10.3233/978-1-61499-971-3-25.
15. L. Lang, R. P. Moser, J. Odenkirchen, and D. Reeves, "Common Data Elements (CDEs)," vol. 13, no. 6, pp. 671–676, 2017, doi: 10.1177/1740774516653238.Improving.
16. E. Soleimani, M. Ahmadi, A. Mohammadi, and J. Alipour, "Development of minimum data set (MDS) for an information management system for aged care centers in Iran," *Informatics Med. Unlocked*, vol. 25, p. 100695, 2021, doi: 10.1016/j.imu.2021.100695.
17. R. Abbasi, R. Khajouei, and M. Mirzaee, "Evaluating the demographic and clinical minimum data sets of Iranian National Electronic Health Record," *BMC Health Serv. Res.*, vol. 19, no. 1, pp. 1–10, 2019, doi: 10.1186/s12913-019-4284-x.
18. Z. S. Nezamodini, Z. Rezvani, and K. Kian, "Identification of the necessary data elements to report AIDS: a systematic review," *Electron. Physician*, vol. 9, no. January, pp. 3592–3597, 2017.
19. Z. Rampisheh, M. E. Kameli, J. Zarei, A. V. Barzaki, M. Meraji, and A. Mohammadi, "Developing a national minimum data set for hospital information systems in the Islamic Republic of Iran," *East. Mediterr. Heal. J.*, vol. 26, no. 4, pp. 400–409, 2020, doi: 10.26719/emhj.19.046.
20. D. Definitions, *Victorian Emergency Minimum Dataset (VEMD) User Manual*. 2013.
21. Ministry of Health & Family Welfare, "EHR Standards for India," *India, M. H. F. W. G. eHealth Sect. (2016). Electron. Heal. Rec. Stand. India*, pp. 1–48, 2016, [Online]. Available: http://www.mohfw.nic.in/showfile.php?lid=4138
22. N. Davaridolatabadi, M. Shahi, F. Sadoughi, and M. Ahmadi, "The comparison of the minimum data set for elderly health in selected countries," *Acta Inform. Medica*, vol. 23, no. 6, pp. 393–397, 2015, doi: 10.5455/aim.2015.23.393-397.
23. S. Shahin *et al.*, "Multiple sclerosis national registry system in Iran: Validity and reliability of a minimum data set," *Mult. Scler. Relat. Disord.*, vol. 33, no. May, pp. 158–161, 2019, doi: 10.1016/j.msard.2019.06.009.
24. Z. Yang *et al.*, "Defining health data elements under the HL7 development framework for metadata management," *J. Biomed. Semantics*, vol. 13, no. 1, pp. 1–15, 2022, doi: 10.1186/s13326-022-00265-5.
25. A. Weissman *et al.*, "Development and implementation of the National Heart, Lung, and Blood Institute COVID-19 common data elements," *J. Clin. Transl. Sci.*, vol. 6, no. 1, 2022, doi: 10.1017/cts.2022.466.
26. J. Zarei *et al.*, "A study to design minimum data set of COVID-19 registry system," *BMC Infect. Dis.*, vol. 21, no. 1, pp. 1–13, 2021, doi: 10.1186/s12879-021-06507-8.
27. A. K. Kock-Schoppenhauer *et al.*, "One Step Away from Technology but One Step towards Domain Experts-MDRBridge: A Template-Based ISO

11179-Compliant Metadata Processing Pipeline," *Methods Inf. Med.*, vol. 58, no. 1, pp. E72–E79, 2019, doi: 10.1055/s-0039-3399579.

28. R. Safdari *et al.*, "National minimum data set for antimicrobial resistance management: Toward global surveillance system," *Iran. J. Med. Sci.*, vol. 43, no. 5, pp. 495–505, 2018.
29. P. Bruland *et al.*, "Common data elements for secondary use of electronic health record data for clinical trial execution and serious adverse event reporting," *BMC Med. Res. Methodol.*, vol. 16, no. 1, pp. 1–10, 2016, doi: 10.1186/s12874-016-0259-3.
30. G. Jiang *et al.*, "Developing a data element repository to support EHR-driven phenotype algorithm authoring and execution," *J. Biomed. Inform.*, vol. 62, pp. 232 – 242, 2016, doi: 10.1016/j.jbi.2016.07.008.
31. M. Claypool, M. Brambilla, J. Cabot, and M. Wimmer, *TBD, Series Editor C M & Model-Driven Software Engineering in Practice*. 2012.
32. D. Martínez, L. Po, R. Trillo-Lado, and J. R. R. Viqueira, "A conceptual data modeling framework with four levels of abstraction for environmental information," *Environ. Model. Softw.*, vol. 183, no. May 2024, p. 106248, 2025, doi: 10.1016/j.envsoft.2024.106248.
33. P. J. Sadalage, M. Fowler, and U. S. River, *NoSQL Distilled*. 2012.
34. US Public Health Service Centers, "Public Health Conceptual Data Model Premiere Edition," no. July, p. 91, 2000, [Online]. Available: http://www.ncmi.cn/UploadFile/2/6/651aba3517cfb717ba8acea2a4709662.pdf
35. L. Min, Q. Tian, X. Lu, and H. Duan, "Modeling EHR with the openEHR approach: An exploratory study in China Philip Payne," *BMC Med. Inform. Decis. Mak.*, vol. 18, no. 1, pp. 1–15, 2018, doi: 10.1186/s12911-018-0650-6.
36. V. L. Tiase, K. A. Sward, and J. C. Facelli, "A Scalable and Extensible Logical Data Model of Electronic Health Record Audit Logs for Temporal Data Mining (RNteract): Model Conceptualization and Formulation," *JMIR Nurs.*, vol. 7, 2024, doi: 10.2196/55793.
37. A. Gamal, S. Barakat, and A. Rezk, "Standardized electronic health record data modeling and persistence: A comparative review," *J. Biomed. Inform.*, vol. 114, no. September 2020, p. 103670, 2021, doi: 10.1016/j.jbi.2020.103670.
38. T. Darmizal, N. H. Ubaidullah, and A. Saad, "Designing Minimum Data Set and Data Model for Electronic Health Record Systems in Indonesia," *Int. J. Adv. Comput. Sci. Appl.*, vol. 16, no. 3, pp. 554–563, 2025, doi: 10.14569/IJACSA.2025.0160355.