Engineering and Technology Journal e-ISSN: 2456-3358

Volume 11 Issue 02 February-2026, Page No.-8901-8909

DOI: 10.47191/etj/v11i02.13, I.F. – 8.482

Developing Data Model for Electronic Health Record System

Teddie Darmizal

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

ABSTRACT: The development of an Electronic Health Record (EHR) system requires a well-defined data set and a robust data model to support complex clinical workflows, interoperability, and data quality. Many existing EHR implementations lack standardized and integrated data models, resulting in redundancy, semantic inconsistency, and limited support for secondary data use. This study aims to develop a comprehensive EHR data model based on a Minimum Data Set (MDS) that systematically represents medical service processes across emergency, outpatient, inpatient, and surgical care. The proposed approach involves the design of a conceptual, logical, and physical data model. Business processes were first analysed to identify relationships among clinical and administrative data classes, which were then transformed into a conceptual data model. The logical data model was developed through database normalization and schema definition, while the physical data model was constructed to represent the actual database structure for implementation. The results demonstrate a structured EHR data model comprising 21 interrelated data classes that reflect real-world medical service workflows. The conceptual data model establishes clear entity relationships, supporting data consistency and interoperability. The logical data model provides a normalized SQL-based schema that minimizes redundancy and enhances data integrity, while the physical data model enables efficient data storage and access. A prototype EHR system was implemented to evaluate the feasibility of the proposed data model, showing stable performance, usability, and alignment with functional requirements. This study concludes that a standardized EHR data model derived from an MDS can effectively support comprehensive clinical documentation, interoperability, and future scalability. The proposed model provides a strong foundation for developing national or institutional EHR systems and can serve as a reference for further EHR data standardization efforts.

KEYWORDS: Data Model, Conceptual Data Model, Logical Data Model, Physical Data Model, Electronic Health Record System.

INTRODUCTION

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 [1]. Meanwhile, the data model must represent the data set in a more technical form, on the conceptual, logical, and physical levels of design [2], 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 [3]–[6]. It specifies which data elements will be stored and how they will be stored, including their relationships and constraints [6]. In a healthcare system, a data set represents the data elements that must be collected for each patient and defines each element based on standards. Therefore, it is important to determine standard data sets to manage the clinical performance of health organizations in every country [7].

EHR data sets are increasingly being used for patient registries and research [8]. EHRs provide various types of data, including structured data such as demographics, vital signs, laboratory test results, procedures, medications, and unstructured data like progress notes [9]. Some examples of data elements in EHR include patient identifiers, demographics, diagnoses, medications, procedures, laboratory results, vital signs, and utilization events [10]. Demographic and laboratory elements lead the list of available elements in hospitals' EHR systems [11].

EHR documentation often contains unstructured information as free text, and the availability of data elements in hospital EHR systems varies, with demographic and laboratory elements being more commonly available [10]. A systematic, structured, and comprehensive EHR data set is needed to analyze the health status of patients [9]. Standardization of the EHR data element facilitates sharing and exchanging information with other systems [12]. In every developing country, all data is generated and stored in EHR by different healthcare system providers. The main objective of formulating the EHR data set is to build a national database that can serve as an information management source to equip decision-makers and policy-makers with accurate and up-to-date information [13].

The standard health care data model usually indicates data elements that should be collected; a data set is a standard data collection tool [14]. A data model specifies features and relationships, such as data types, constraints, relationships, and metadata [15]. 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 [16]. There are three levels of data models, which grow increasingly complex: conceptual, logical, and physical [2].

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 [17]. 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 [2], [15], [18].

Designing a conceptual, logical, and physical data model for standardized data collection supports disease information management and leads to better quality of care [19]. The challenge in database design is to define a proper data set that will produce high value now and in the future. Designing highly useful minimal data sets is critical to maximize the value of research data over time [1].

Yang et al., (2022) stated the absence of a unified conceptual model linking data elements across domains is one of the problems in developing a national health concept data model in China. Existing national data element dictionaries in China primarily aggregate data elements from various business forms without a shared conceptual foundation. This results in redundancy and semantic inconsistency, revealing a gap in model-driven integration of multi-domain health data elements. Yang et al., (2022) also stated that the lack of practical systems for centralized metadata governance leads many studies to propose models or frameworks but not to provide operational, web-based systems that support real-world management, evaluation, and evolution of national health metadata repositories [21] [22].

METHOD

Designing a conceptual, logical, and physical data model for standardized data collection supports disease information management and leads to better quality of care [19]. Technically, the development of the EHR data model will be carried out in 3 ways, namely:

a. Developing EHR Conceptual Data Model - In this phase, the researcher analyzes business processes and data relationships, furthermore transforming data classes into a conceptual model.
b. Developing EHR Logical Data Model - In this phase, the researcher performs normalization, furthermore creating a logical data schema to transform the entity into a table.
c. Developing EHR Physical data model - In this phase, the researcher creates a physical data model as a representation of how data is stored, organized, and accessed in a database management system.

RESULTS & DISCUSSION

A. Developing Conceptual Data Model

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 creating a CDM, the first thing to do is identify the relationships between data classes and data elements from the EHR Data Set according to the business process of medical services. The following will explain the steps taken to create a CDM.

1. Analyzing Business Process and the Relationship between Data Classes

There are 5 phases in medical services, and there are 21 data classes (related to 21 processes running) in all of these phases. The processes in these 5 phases run in emergency, outpatient, inpatient, and surgery services. The initial phase, Registration, commences with the clinical service registration data class, which serves as the primary relational link to subsequent data classes within the electronic health record (EHR) framework. The next phase, Triage, Assessment, and Diagnosis, begins with the triage process—applied specifically to emergency care—followed by a structured nursing assessment. Within this phase, three distinct data classes are delineated: Nurse General Initial Assessment, Initial Assessment by the Psychiatric Nurse, and Initial Assessment by the Neonatal Nurse. Each nursing assessment is tailored to the patient category being managed, ensuring that the collected data accurately reflects the specific clinical context.

After completing the nursing assessment, it is then continued by the doctor, where there are 3 types of data classes, namely: Initial Assessment by General Doctor, Initial

Assessment by Psychiatrists, and Initial Assessment by Obstetricians. The assessment used by the doctor is in accordance with the type of patient they are treating. Before the doctor determines the diagnosis, it will first enter the supporting examination phase to support medical decisions. Suppose the patient is indicated to have other diseases and needs observation from other medical groups. In that case, the doctor will consult with other specialist doctors; in this process, there is a consultation data class.

As part of the process of establishing a diagnosis, doctors will generally also conduct laboratory and radiological examinations. In this process, there is a laboratory examination and a radiology examination data class. In hospitalized patients, there is a special assessment to assess the nutritional condition and dietary needs of patients by a nutritionist through a nutrition assessment data class. After all supporting examination data is obtained, the doctor can then establish a diagnosis and provide action (if necessary) through the diagnosis and procedure data class. Once the diagnosis is determined and medical action (if any) is taken, the doctor will then provide a prescription, and this process is handled by the pharmacist through the prescribing data class.

Patients with indications requiring surgery and surgical procedures will be forwarded to the surgery admission process through the surgery registration data class. In the surgery section, there is an assessment process through the pre-operative and pre-anesthesia data class. After the surgery is performed, they will continue to the reporting process through the post-operative and post-anesthesia data class. All examinations performed on patients by every medical service provider are documented within the Integration Patient Examination Record data class.

At the conclusion of the care episode—whether in emergency, outpatient, or inpatient settings—the process enters the resume phase, which includes two key data classes: the General Medical Resume, encompassing the overall clinical summary, and the Surgical Medical Resume, designated specifically for patients who have undergone operative procedures.

2. Transforming Data Class to Conceptual Data Model

After analysing the relationship between data classes in the entire series of business

processes of medical services, the next step is to describe the conceptual data model (CDM) for the EHR System. The CDM can be seen in Figure 1.

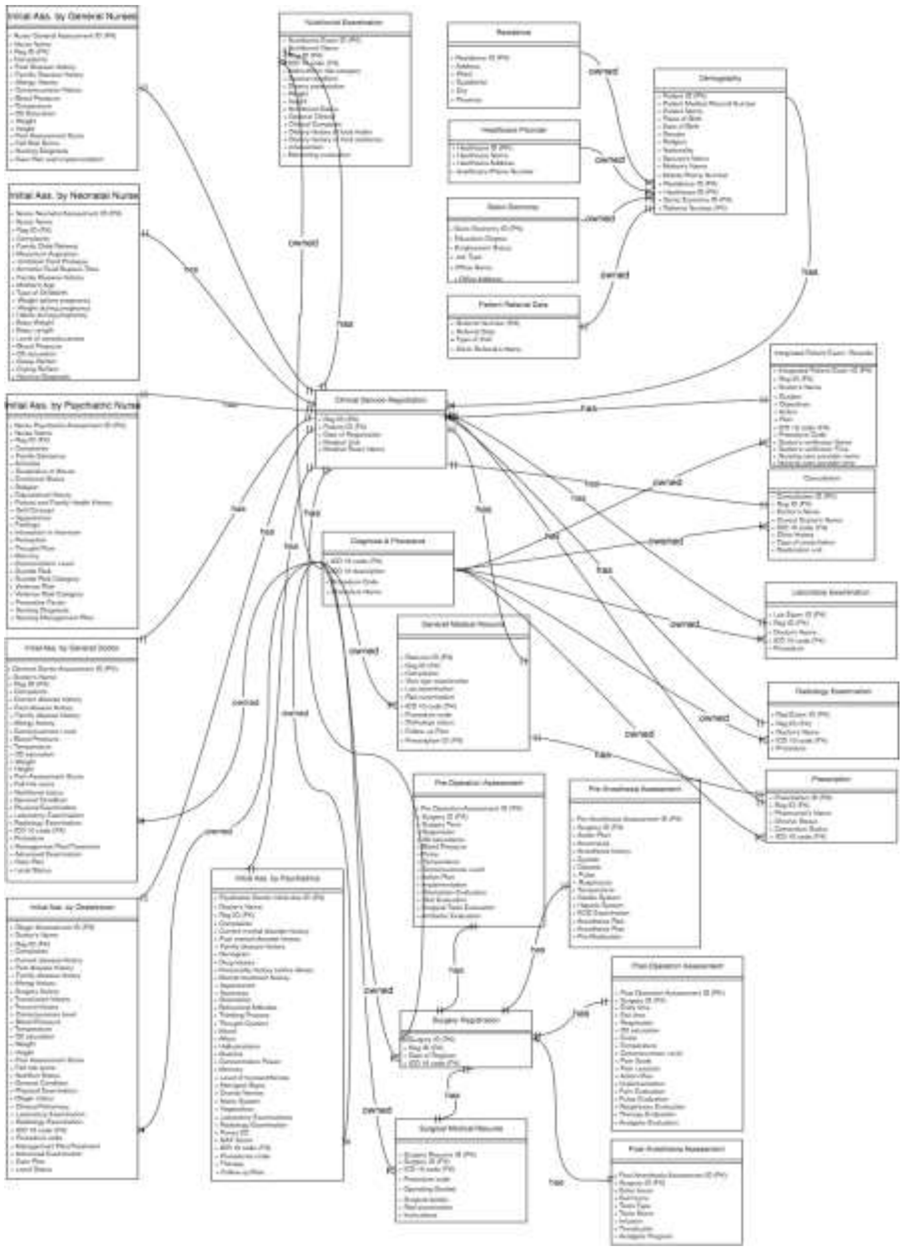

Figure 1. Conceptual data model of the EHR System

In the Conceptual Data Model (CDM), each data class derived from the EHR minimum data set (MDS) is represented as an entity, and the logical connections among these entities are established through relationships. The CDM highlights how clinical and administrative data classes interact to support the overall Electronic Health Record database architecture. Below are the key entities and the principal relationships defined within this model:

B. Developing Logical Data Model

After designing the conceptual data model, the next step is to design the logical data model. Logical data models (LDM) help to define the detailed structure of the data elements in a system and the relationships between data elements. They refine the data elements introduced by a Conceptual data model and form the basis of the Physical data model.

1. Normalization Process

Database normalization is the process of organizing data in a relational database to reduce redundancy and improve data integrity. It involves restructuring tables and defining relationships according to a series of rules, known as normal forms, which progressively minimize duplication and dependency issues. By breaking large, complex tables into smaller, related ones, normalization ensures that each piece of information is stored only once and is logically connected to related data through primary and foreign keys. This not only makes the database more efficient in terms of storage but also simplifies maintenance, improves query performance, and helps prevent anomalies during data insertion, updating, or deletion.

2. Logical Data Schema

The following is a proposed logical data schema for an Electronic Health Record (EHR) System, described in SQL script form. All entities mentioned in the Conceptual Data Model will be referred to as tables. This is one of example of logical data schema for “Initial Assessment General Nurse” Entity.

TABLE Initial_Assessment_ General_Nurse (

```
  Nurse_General_Assess_ID PRIMARY KEY,
  Employee_ID FOREIGN KEY,
  Reg_ID FOREIGN KEY,
  anamnesis_source,
  Complaints,
  Current_Disease_History,
  Past_Disease_History,
  Family_Disease_History,
  Allergy_History,
  Medication_History,
  Occupation,
  economy,
  emotional_status,
  history_violence_behavior,
  drug_dependencies,
  alcohol_dependencies,
  smoking_status,
  religion,
  ethnicity,
  immunization,
  gcs_Eye,
  gcs_Motor,
  gcs_Verbal,
  Consciousness_Level,
  body_posture,
  pulse,
  Blood_Pressure,
  respiration,
  Temperature,
  O2_Saturation,
  current_weight,
  previous_weight,
  height,
  upper_arm_circumference,
  head_circumference,
  BMI,
  triage,
  physical_exam_head,
  physical_exam_hair,
physical_exam_face,
physical_exam_eye,
physical_exam_nose,
physical_exam_mouth,
physical_exam_teeth,
physical_exam_ear,
physical_exam_neck,
physical_exam_skin,
physical_exam_respiratory,
physical_exam_cardiovascular,
physical_exam_sensory_perception,
physical_exam_neuro,
physical_exam_elimination,
physical_exam_genitalia,
physical_exam_gastro,
physical_exam_extremitis,
physical_exam_others,
growth_disorder_status,
development_disorder_status,
physical_status_self_image,
physical_status_identity,
physical_status_role,
physical_status_ideal_self,
physical_status_self_worth_code,
physical_status_appearance,
physical_status_speech,
physical_status_motor_activity,
physical_status_mood,
physical_status_effect,
physical_status_interview_interaction,
physical_status_perception,
physical_status_thought_content,
physical_status_thought_flow,
physical_status_memory,
physical_status_concentration_calculation,
physical_status_judgemnet_ability,
physical_status_insight,
pain_assessment_cause,
pain_assessment_worsening_factor,
pain_assessment_relieving_factors,
pain_assessment_quality,
pain_assessment_location,
pain_assessment_radiation,
pain_assessment_score,
pain_assessment_category,
pain_assessment_method,
pain_assessment_duration,
pain_assessment_frequency,
fall_risk_score,
fall_risk_category,
fall_prevention_steps,
```

```
    functional_status_score,
    functional_status_category,
    nutritional_risk_method,
    nutritional_risk_score,
    suicide_risk_score,
    suicide_risk_category,
    violence_risk_score,
    violence_risk_category,
    educational_need_topic,
    educational_need_translater,
    educational_need_barrier,
    nursing_diagnosis,
    management_plan,
    implementation,
    evaluation,
    care_plan
)
```

C. Developing Physical Data Model

A physical data model is a representation of how data is stored, organized, and accessed in a database system. It takes into account the specifics of the underlying hardware, software, and database management system (DBMS). Unlike a logical data model, which focuses on abstract relationships and structures, the physical model is concerned with how data is physically implemented and optimized for performance, storage efficiency, and retrieval. The physical data model diagram of the EHR System can be seen in Figure 2.

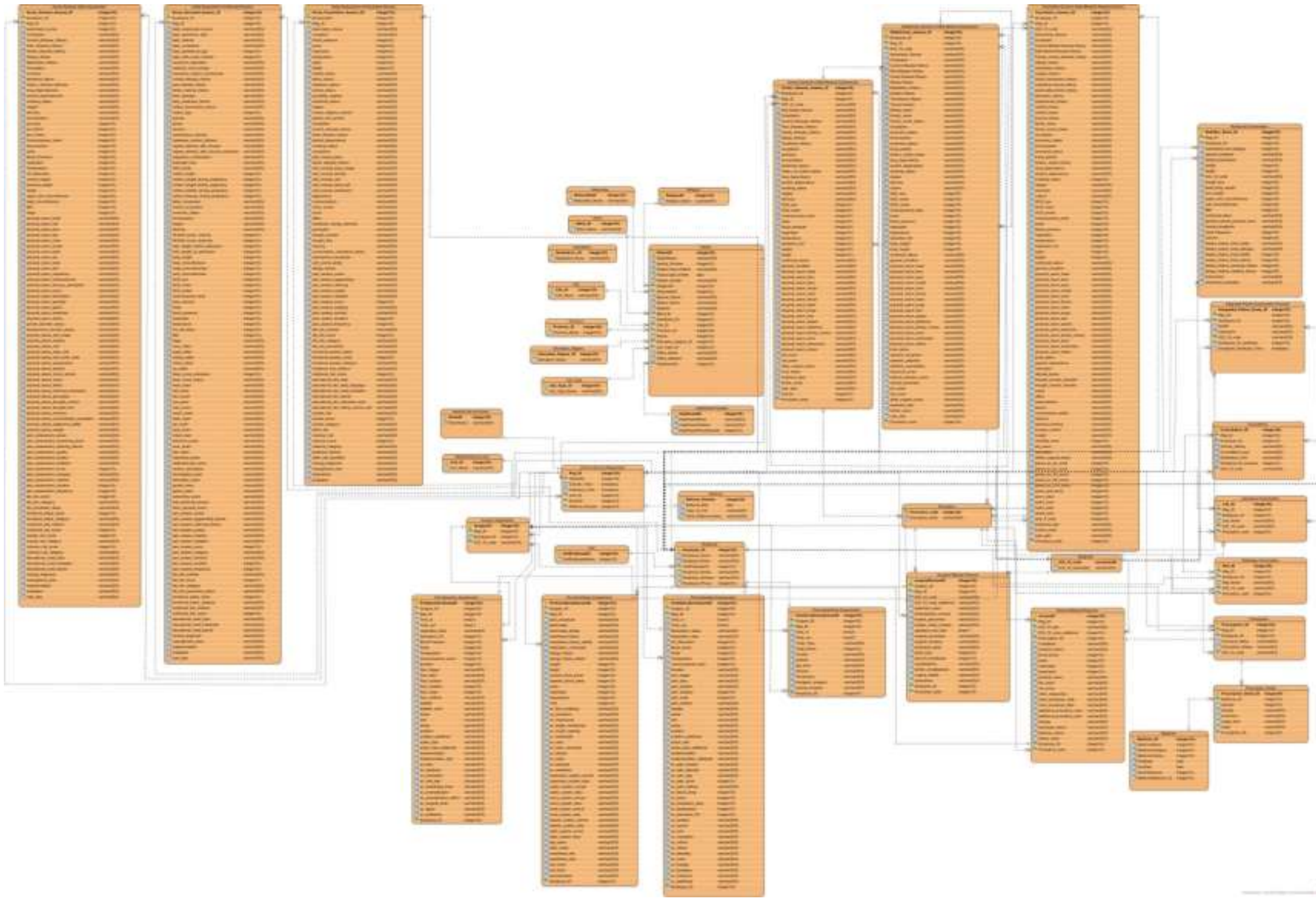

Figure 2. Physical data model of EHR System

CONCLUSION

The development of a data model and prototype EHR system, resulted in a conceptual and logical schema capable of accommodating complex medical service processes. The conceptual data model (CDM) demonstrates a strong foundation for interoperability by clearly establishing relationships between key entities. The transition from CDM to logical data model (LDM) is a crucial phase in the data model design stage. The normalization process plays an important role in refining the structure, reducing redundancy, and improving data integrity. The SQL-based schema proposed for the LDM shows a pragmatic approach to database implementation. The physical data model (PDM) developed for the EHR system represents the final stage in translating conceptual and logical models into a fully implementable relational database schema.

The EHR system was also built to support the EHR data model evaluation process, the goal of which was to ensure that the functional requirements derived from the MDS were accurately implemented. The system demonstrated stability, usability, and reliability, confirming the feasibility of translating the EHR data model into an operational application.

REFERENCES

1. 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.
2. R. Sherman, “Foundational Data Modeling,” *Bus. Intell. Guideb.*, pp. 173–195, 2015, doi: 10.1016/b978-0-12-411461-6.00008-3.
3. 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.
4. 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.
5. 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.
6. 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.
7. 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.
8. J. Owusu-marfo, Z. Lulin, H. A. Antwi, and M. O. Antwi, “Electronic Health Records Adoption in China ’ s Hospitals : A Narrative Review,” *Eur. J. Contemp. Res.*, vol. 8, no. 1, pp. 409–418, 2019, [Online]. Available: https://www.researchgate.net/publication/337293396_Electronic_Health_Records_Adoption_in_China’s_Hospitals_A_Narrative_Review
9. 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.
10. V. Ehrenstein, H. Kharrazi, H. Lehmann, and C. O. Taylor, “Obtaining Data From Electronic Health Records,” *Tools Technol. Regist. Interoperability, Regist. Eval. Patient Outcomes A User’s Guid.*, pp. 1–92, 2019.
11. 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.
12. D. Xiao, C. Song, N. Nakamura, and M. Nakayama, “Development of an application concerning fast healthcare interoperability resources based on standardized structured medical information exchange version 2 data,” *Comput. Methods Programs Biomed.*, vol. 208, p. 106232, 2021, doi: 10.1016/j.cmpb.2021.106232.
13. 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.
14. 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.
15. 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.

16. 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.
17. 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.
18. 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.
19. S. H. El-sappagh, S. El-masri, A. M. Riad, and M. Elmogy, "Electronic Health Record Data Model Optimized for Knowledge Discovery," *Int. J. Comput. Sci. Issues*, vol. 9, no. 5, pp. 329–338, 2012.
20. 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.
21. A. I. Paganelli *et al.*, "A conceptual IoT-based early-warning architecture for remote monitoring of COVID-19 patients in wards and at home," *Internet of Things (Netherlands)*, vol. 18, p. 100399, 2022, doi: 10.1016/j.iot.2021.100399.
22. 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.