Streamlining and Standardization Proposal of Electronic Health Record Encoding and Storage

Raneille Quirsten Baligod, Angela Felicia Marie Reyes, Heidi Lindt Sy

De La Salle University
2401 Taft Avenue, Manila 0922, Philippines
E-mail: raneille_baligod@dlsu.edu.ph, angela_felicia_reyes@dlsu.edu.ph, heidi_sy@dlsu.edu.ph

Abstract

The COVID-19 pandemic has amplified the development of telemedicine, where patients are diagnosed and monitored through remote healthcare services, often involving digital technologies. One of the crucial components of telemedicine is electronic health records (EHRs), which allow healthcare providers to access patients' information using a digital device. EHRs provide better access, storage, and analysis than paper-based records, however, these still present challenges, especially regarding their efficiency and interoperability. The manual encoding of data can be tedious and also prone to human errors. In using EHRs, data privacy breaches could expose confidential patient information. Additionally, healthcare databases are not similarly organized and synced in different systems, causing challenges in terms of data sharing. Other systems and data are needed in various healthcare fields, and having different protocols and terminologies would lead to incompatibility. Technologies such as artificial intelligence (AI), machine learning, and blockchain have been explored to address the stated challenges. Artificial intelligence has offered features such as speech-to-text transcription and Natural Language Processing (NLP), which can automate documentation. Blockchain technology has offered secure and decentralized data management, contributing to the privacy, security, integrity, and accessibility of patient records. For interoperability, patient records are stored in the cloud, increasing accessibility. This study aims to tackle these challenges from an encoding and storing perspective. These challenges often go hand in hand, therefore an appropriate solution would be to standardize how data is inputted and stored. The fields to be filled in per patient would be according to common parameters across different databases, such as the personal information of the patient and the nature of their ailment. Sensitive data of the patient would need to be encrypted and protected. While the use of speech-to-text AI would be a hands-free method of encoding data, many factors such as noise and the speaker's enunciation, could cause errors. Therefore, the interface and method of inputting these fields should be designed intuitively, leaving as little friction as possible for the user.

Keywords — (e.g., linear programming, …)

I. Introduction

The COVID-19 pandemic has dramatically accelerated its adaptation to telemedicine, where its goal is to allow healthcare professionals to reach patients during a time wherein physical interaction is restricted. With the help of telemedicine, the patient can be diagnosed, treated, and monitored to reduce the risk of viral transmission [1]. Electronic Healthcare Record (EHR) serves as the backbone of telemedicine, wherein it enhances the accessibility of patient records, allowing the healthcare provider to receive the patient record information in real-time, regardless of the location [2]. Moreover, EHR allows systematic storage and analysis of health data, which enables much better decision-making and results in a much better patient outcome [3]. The impact of EHR on healthcare safety and quality would improve data accessibility and help avoid medical errors due to mixed-up clinical reports. However, not all healthcare professionals agree that EHR is effective, as nurses think it's effective for documenting each patient. Still, physicians disagree; they often view it as time-consuming [4]. Additionally, in terms of the security and interoperability of EHRs, they become vulnerable due to data breaches and ineffective integration across the EHR platforms which could result in a breach in patient information resulting in possible threat and also hinder the smooth exchange between data on the side of healthcare providers [5].

Some challenges limit the effectiveness of the EHR system. One of the significant issues is the manual encoding of data, which is also prone to human error and time-consuming. Furthermore, with this error, a data privacy breach would happen wherein the sensitive patient information would be accessed by an unauthorized person. This study addresses EHR encoding and storing challenges to prevent human error. And also to develop a standardized data input field to ensure compatibility and interoperability throughout the EHR system by adding an encryption method to protect the patient's sensitive information. The study aims to develop a standardized format to facilitate effective data transfer between databases in the healthcare setup. This emphasizes ensuring that the encoding process is accurate and compatible. In that format, it would use a standardized parameter to ensure consistency in the system's interoperability, allowing manual encoding and not automated, and lastly, creating a format that will enable the transferring of data from one database to another.

II. Review of Related Literature

A. *Electronic Health Record*

The Electronic Health Record (EHR) is a digital version of the patient's medical history. This would include comprehensive patient health-related information, such as diagnosis, medication, treatment plans, laboratory results, etc. EHR is designed to have efficient data storage and access and to share medical data across the healthcare institute, improving patient care [6]. With it being interoperable, it is usually seen on frameworks like HL7 or FHIR, enabling healthcare providers to collaborate by ensuring that the patient's information is precise [7].

The EHR system would collect, store, and manage the patient health data by distributing the data into categories. Thus, the primary function of EHR is to provide a manual input of patient data into the system, which would then be stored securely in the cloud of the database. Data in EHRs can only be accessed by authorized users, such as healthcare professionals and patients, through user-friendly interfaces. EHR would also be interoperable, allowing a seamless sharing of patient data between specialists, doctors, and patients with the standard framework of HL7 or FHIR. This seamless patient information sharing would allow the specialists to access their patient's medical history, thus enabling a collaborative decision-making tool for much more precise patient care [8].

It would play an important role in healthcare by enhancing patient care, making it a much more efficient operation, and having a data-driven insight. Allowing healthcare providers access to patients' complete medical histories would reduce the risk of medical errors [6]. With EHR being applied in health care, it has improved the communication between healthcare providers and patients to be satisfied by ensuring that records are accurate and safe. In the Philippines, in rural areas, EHR addresses problems such as misplaced patient files and delayed retrieval of files, and EHR can lead to better healthcare service delivery [9].

B. *Challenges in EHRs*

One of the challenges of EHR interoperability is that it lacks standardization in frameworks like HL7 and FHIR, which hinders the effective integration and medical data exchange. EHRs also struggle with maintaining accuracy, consistency, and completeness, whereas manual data input can cause errors such as duplication and incomplete patient records. Multiple variations in how data is categorized and stored make it difficult for healthcare providers to access and comprehend patient histories [10]. In the study of Issa et al. [11], Data breaches are being considered due to unauthorized access. These breaches indicated that the EHR lacks weak encryption and a lack of collaboration between healthcare providers. In 48% of the electronic health records with the administered connection of security, it has been observed that data breaches jeopardize the work data integrity and health quality.

C. *Innovations on EHRs*

There have been advancements in EHR technology to address the existing interoperability, privacy, security, and efficiency challenges. These include Artificial Intelligence (AI), Machine Learning (ML), Natural Language Processing (NLP), Blockchain, and Cloud-based EHRs. These advancements focus on the automation, accessibility, and security of EHRs.

Artificial Intelligence (AI) is beneficial for a variety of applications. In healthcare, AI models are used for diagnosis, predicting patient outcomes, and organizing EHRs. The study of Wang et al. [12] has shown that AI models can be trained to streamline data processing. The model that they have developed are specifically for the EHR data of COVID-19 ICU patients. Their approach standardizes how raw clinical data is handled, which makes it more understandable by AI models to accurately analyze and predict health outcomes of patients. This then improves decision making for COVID-19 cases as well as creates a foundation for EHR handling for other applications. The study of El-Rashidy et al. [13] further shows the significance of using AI-driven Remote Patient Monitoring (RPM) systems where real-time data exchange between patients and healthcare providers is involved. Findings include that RPM reduces hospital readmissions by 35% for chronic diseases. A clinical decision support system (CDSS) was also proposed in the study, which employs AI and ML for interoperability.

Blockchain-based EHR frameworks are also explored as they offer a promising solution regarding interoperability and security concerns. A development and implementation of a blockchain-based EHR framework was performed by Reegu et al. [14]. The framework provides secure, immutable, decentralized storage compiled with HIPAA and HL7. Thus, blockchain technology enhances data exchange while ensuring the security of EHR against data breaches. Moreover a study conducted by Keshta & Odeh [15] suggests the usage of smart contracts for automating access control. This grants patients with greater autonomy over their data. In contrast, improved accessibility and cost efficiency are the benefits of cloud-based EHR systems. According to the study of Omboni et al. [16], cloud platforms enable real-time EHR access from any location, which further supports telemedicine and remote care for patients. Issa et al. [11] has also reported that hospitals integrated with cloud-based systems have 40% lower infrastructure costs compared to on-premise servers. However, data sovereignty laws and internet dependency limit cloud adoption in regions that lack resources [15].

D. *Standardization and Benchmarking*

Standardization is critical for ensuring interoperability and data consistency across EHR platforms. Widely adopted standards include HL7 Fast Healthcare Interoperability Resources (FHIR), which enables real-time data exchange through application programming interfaces (APIs), and SNOMED CT, which is a comprehensive clinical terminology system that reduces semantic errors in diagnosis coding [17]. Other standards such as LOINC for laboratory data and DICOM for medical imaging have further harmonized components of EHR to secure compatibility between existing healthcare systems.

A study explored 21 hospitals that provide palliative care, using a unified framework for reporting patient activities such as diagnoses, referrals, and symptom reports and using a standardized format called the standard Format for Reporting Hospital PCT Activity. It is applied in different hospitals, proving the feasibility of standardizing these reports across multiple heterogeneous institutions [18]. Therefore, regulatory frameworks have a significant role in EHR standardization. The World Health Organization (WHO) promotes the International Classification of Diseases (ICD-11) to unify diagnostic reporting globally. Policies like the Health Insurance Portability and Accountability Act (HIPAA) in the U.S. and the General Data Protection Regulation (GDPR) in the EU enforce stringent data privacy measures [11].

Benchmarking efforts have also been done to evaluate EHR performance using metrics such as interoperability, fidelity and privacy-utility trade-offs. For example, FHIR-based systems demonstrate higher data accuracy in cross-platform exchanges, compared to older protocols like HL7v2 [19]. Synthetic EHR data are also used to preserve patient anonymity. The study of Yan et al. [20] benchmarked synthetic EHRs. The findings indicate that synthetic EHR data often sacrifices 15–20% of data utility to meet privacy thresholds.

III. Methodology

This paper aims to create a database format that would act as a base and standard for digital health records. With this, the transfer of data from one source to another would be more seamless and minimize the loss of data. The project infrastructure must be built with this in mind to ensure a format that is both informative and intuitive.

A. Basis of Design

Electronic health record databases will serve as the main source of this project. These will provide both data to test with as well as real-world examples of how current databases are formatted. Such databases can be found in the IEEE Dataport as well as Kaggle, which are both open research data platforms with focus on engineering and machine learning [21, 22]. While this project will not be utilizing intelligent systems, datasets such as these are very valuable sources of raw, and often real-world, data that can be used as a basis for other studies.

Currently, there are 6 datasets of interest to be further investigated. These are SynSUM, a synthetic medical record database with both structured and unstructured entries [23], a clinical dataset compiled by Mohamadreza Momeni [24], the clinical dataset of tbe Cytochrome Pyschotropic Genotyping Under Investigation for Decision Support (CYP-GUIDES) trials [25], synthetic electronic medical records generated by Sina Sheikholeslami [26], electronic medical records compiled by David Victor [27], and synthetic medical records generated for instructional activities at the Samford University uploaded by user Mexwell [28]. The specific usage of each database, whether it be using the data itself for format testing or using the database as inspiration for the design of the format is to be determined. Given that this data will not be used for machine learning, the usability score of these sources is not a property of concern.

Another important source would be protocols and standards for electronic medical records, whether these be official standards or proposed standards found in research papers. The World Health Organization published a manual on the creation and maintenance of electronic health records [29]. In the manual, these records should contain and cover all personal health information of the individual, is able to be accessed and altered by healthcare providers, and also be able to be accessed by the patient themselves and any future healthcare provider attending to the patient. Several papers have discussed, reviewed, and designed methods and protocols for the digital encoding of medical data [10, 12, 30]. The proposed layout will be based on these recommendations and the noted limitations and challenges found in these papers.

As for the software, Python will be utilized as the main language. Python has many database libraries and modules that make it ideal for this project. The most popular of these modules are MySQL, SQLite, and JSON [31]. Depending on how each method can satisfy the needs of the format to be developed, the program will be written using that module.

B. Work Distribution

The database development is divided into four phases: (1) dataset collection and preparation, (2) Database design and development, (3) Implementation, and (4) Testing and Validation.

First, the datasets are gathered from sources such as Kaggle and IEEE DataPort. The data structures must be verified to ensure efficiency, especially in implementation. The members of the group will be assigned to gather information about EHR encoding and storage, as well as the accepted practices and standards for EHRs, before designing the database.

The second phase of the database development involves the designing the architecture of the database. Programming tools such as Python programming along with database technologies such as MySQL, SQLite, and JSON will be used to construct the database. In this phase, different data fields are defined to allow smooth retrieval and storage of patient information. Shown in Figure 1 is the data flow diagram of the proposed database. This shows how the data will be processed and implemented.

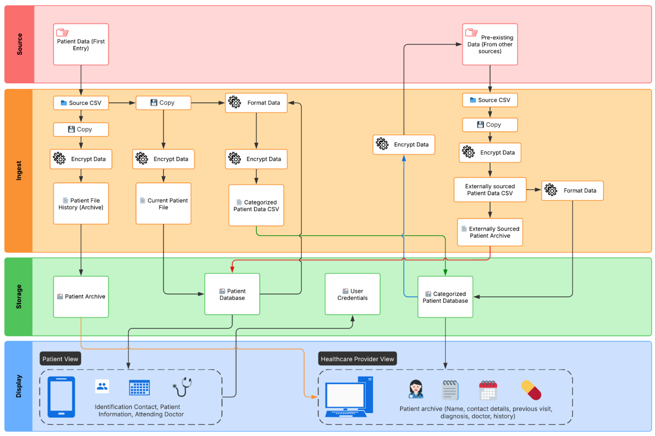

Figure 1. Data flow diagram.

After programming and designing the system, the third phase is dedicated to the implementation and testing of the proposed database. The prototype system will be developed based on the defined structure of the data and testing will be conducted using the EHRs gathered from the online sources. Moreover, in this phase, the ability of the system to handle large data with accuracy, security, and integrity, as well as its interoperability is assessed. With this, improvements to the system will be done to enhance efficiency.

The last phase involves validation and quality checks, where the effectiveness of the proposed system will be evaluated. One way is to compare it with existing interoperability frameworks and benchmarks. The comparison will include compliance with healthcare data-sharing regulations, security performance, and ease of use.

C. Quality Check

There are numerous challenges and requirements for a patient portal and EHR systems to be interoperable. These include data sharing standards and regulations, heterogeneous systems, data management and storage, information availability, data formats, identification of individuals, user access, and security [32]. Our proposed system then prioritizes these factors through adhering to established interoperability standards and frameworks to ensure that data is efficiently transferred between systems without compromising quality.

An important component of the quality of the database is maintaining data integrity and reducing inconsistencies within EHRs. The benchmarking of synthetic EHR generation models emphasize the significance of structured and unstructured data validation in evaluating the accuracy of generated records [20]. For our system, a structured approach to data standardization will be employed. This performs careful validation checks on the EHRs before storage. The process involves cross-checking with standardized medical terminologies, such as SNOMED CT and LOINC, to ensure that encoded information remains clear and universally understandable [19]. Additionally, automated data correction tools will be used to aid in identifying and correcting potential errors. This can reduce the risk of discrepancies in patient records.

The proposed system will also use encryption and access control mechanisms to protect patient data. User authentication and role-based access will be employed on the database system to prevent unauthorized access and modifications. This allows authorized users to edit patient records while other healthcare professionals or unauthorized users can use and collaborate through the database. Blockchain-based systems may also be explored for certain issues since they have the ability to provide decentralized and tamper-proof records [14].

Usability testing will also be conducted to evaluate the efficiency of the system. The design will focus on minimizing redundant data entry, speeding up retrieval times, and improving the overall user experience. Benchmarking studies on synthetic EHR generation models emphasize the importance of usability and accessibility. This ensures that generated records closely resemble patient data in structure and format in existing systems [20]. Lastly, our system will undergo iterative refinements to enhance its practicality and adaptability in healthcare environments.

IV. Results and Discussion

A research paper and a total of 5 databases were analyzed with all the fields, as well as any additional notes, recorded into a Google Sheet, which is uploaded as a supplementary file. A total of 144 fields were compiled. The research paper *Data Flow Construction and Quality Evaluation of Electronic Source Data in Clinical Trials: Pilot Study Based on Hospital Electronic Medical Records in China* [10] used 29 different fields for their study, with 17 of those relating to vital signs or laboratory tests and results. The clinical dataset by Mohamadreza Momeni [24] had two different datasets, and therefore two formats, included. For simplicity, these were combined in the compilation, coming to a total of 19 unique fields. The number of fields from the CYP-GUIDES database [25] came to a total of 43 fields, however, 30 of these fields were letter keys representing different medications. This distinction between each medication makes sense due to the database coming from a clinical trial, but for the purposes of this paper, this specification only creates more tedium. This was later generalized. The two synthetic medical record datasets, created by Sina Sheikholeslami [26] and Mexwell [27] 17 and 28 fields respectively. The latter came with 18 different tables of data, each one containing different types of information such as allergies, healthcare plans, and transactions. The fields extracted for this study came from the table with patient information. This dataset deals with more specific entries, with the first name, middle name, and last name for instance all being separate fields instead of a single "name" field. Finally, the dataset of electronic medical records compiled by David Victor [28] had 8 total fields, two of which were "medical condition" and "symptoms." For each value of medical conditions, there would be a corresponding set of symptoms to match. All of these fields, once compiled in the Google Sheet, was divided with each source having its own row, as seen in Figure 2 below:

B	C	D	E	F	G	H	I	J	K
Source	Notes								
Clinical Dataset	Comes in two formats (combined here for simplicity)	Patient ID	Specimen date	Dead or Alive	Survival time (days)	Date of Death	Date of last Follow	Sex	Age
Clinical Dataset of the CYP-GUIDES Trial	"COLUMNS A through AD – Coded columns for the medications and the number of administrations for each during the hospital stay. Each Column represents a different medication from the formulary as prescribed and comprised of the following drugs: Amitriptyline, Aripiprazole, Asenapine, Bupropion, Chlorpromazine, Citalopram, Clomipramine, Clonidine, Doxepin, Duloxetine, Escitalopram, Fluoxetine, Fluphenazine, Fluvoxamine, Guanfacine, Haloperidol, Imipramine, Lithium, Methylphenidate, Mirtazapine, Nortriptyline, Olanzapine, Paliperidone, Perphenazine, Quetiapine, Risperidone, Sertraline, Trazodone, Venlafaxine, Ziprasidone. A key for these columns is included in the dataset. Each number corresponds to the number of times that the given medication was administered to the patient while in the hospital. # ADMINISTRATIONS – The count of the times a given psychotropic medication was administered to each patient during the participant's hospital stay. In the example above, the 2 administrations of Aripiprazole, 3 administrations of Fluoxetine, and 4 administrations of Trazodone, would be counted and reported in the respective column for the specific drug for each patient."	ID	GENDER	AGE	RACE/ETHNICITY	Diagnosis	MD	Assignment	EMR
Synthetic Electronic Medical Records		patient id	age	gender	race	heigh	weight	bmi	blood pressure systolic
electronic medical records EMR	Conditions and their corresponding symptoms: 'Hypertension': ['headache', 'dizziness', 'shortness of breath', 'nosebleeds'], 'Diabetes': ['increased thirst', 'frequent urination', 'fatigue', 'blurred vision'], 'Asthma': ['wheezing', 'shortness of breath', 'chest tightness', 'coughing'], 'COPD': ['chronic cough', 'wheezing', 'shortness of breath', 'chest tightness'], 'Heart Disease': ['chest pain', 'shortness of breath', 'fatigue', 'irregular heartbeat'], 'None': ['healthy', 'no symptoms'], 'Malaria': ['fever', 'chills', 'headache', 'muscle pain'], 'Typhoid': ['fever', 'weakness', 'abdominal pain', 'rash'], 'Ulcer': ['stomach pain', 'bloating', 'nausea', 'vomiting'], 'HIV': ['fever', 'night sweats', 'weight loss', 'swollen lymph nodes']	patient id	name	age	gender	ethnicity	medical condition	symptoms	clinical notes
Synthetic AL Medical Records	Has many different tables for different categories of information (allergies, careplans, organizations, transcactions, etc.) The fields written here come from the patient table. Note that many of the fields are pieces of a general field (ex. name)	id	birthdate	deathdate	ssn	drivers	passport	prefix	first
JMIR Medical Informatics - Data Flow Construction and Quality Evaluation of Electronic Sou		Subject ID	Date of birth	Sex	Ethnicity	Race	Age	Combined drug name	Whether for the treatment of adverse reactions

Figure 2. Field compilation.

A compilation of all fields was put in a column, with all entries being converted to uppercase for uniformity, as Google Sheets differentiates lowercase and uppercase letters. The unique() function of Google Sheets was then applied to this column, extracting all unique entries from the list to remove duplicate entries so fields such as gender or age only appear once in the list. This reduced the list of fields to 127. The amount of times each field was used for the whole dataset was also measured using the countif() function, serving as the main basis of what medical databases usually record. From there, the fields were filtered even more manually as some unique words have the same meaning, such as birthdate and date of birth. Other fields were combined and generalized, such as the letter keys for the 30 different medications being grouped together into a single "medication" field. This then reduced the count to 45 unique fields, as seen in Figure 3 below.

COMPILATION OF ALL FIELDS:	
Patient ID	PATIENT ID
Specimen date	SPECIMEN DATE
Dead or Alive	DEAD OR ALIVE
Survival time (days)	SURVIVAL TIME (DAYS)
Date of Death	DATE OF DEATH
Date of last Follow	DATE OF LAST FOLLOW
Sex	SEX
Age	AGE
Race	RACE
Event	EVENT
Tumor size	TUMOR SIZE
Grade	GRADE
Stage	STAGE
Cigarette	CIGARETTE
Pack per year	PACK PER YEAR
Type Adjuvant	TYPE ADJUVANT
Batch	BATCH
EGFR	EGFR
KRAS	KRAS
ID	ID
GENDER	GENDER
AGE	AGE
RACE/ETHNICITY	RACE/ETHNICITY
Diagnosis	DIAGNOSIS

(a)

Raw Unique Categories	Count	Manually Filtered
PATIENT ID	3	Patient ID
SPECIMEN DATE	1	Date of entry
DEAD OR ALIVE	1	Dead or alive
SURVIVAL TIME (DAYS)	1	Survival time
DATE OF DEATH	1	Date of death
DATE OF LAST FOLLOW	1	Date of last follow-up
SEX	2	Sex
AGE	5	Age
RACE	4	Race
EVENT	1	Event/Reason for admission
TUMOR SIZE	1	Tumor size
GRADE	1	Grade
STAGE	1	Stage
CIGARETTE	1	Cigarette use
PACK PER YEAR	1	Pack per year
TYPE ADJUVANT	1	Type adjuvant
BATCH	1	Batch
EGFR	1	EGFR
KRAS	1	KRAS
ID	2	Diagnosis
GENDER	4	MD
RACE/ETHNICITY	1	Assignment
DIAGNOSIS	1	EMR

(b)

Figure 3. Conversion to uppercase (a) and filtering (b)

The final chosen fields were picked based on the number of times they were used (as mentioned above) as well as more grouping. An example would be the combination of all the smoking related fields, a single risk factors field, which also accommodates other risks such as drinking habits and exposure to toxins. The final count of fields was equal to 27. These were then classified according to the type of field – namely administrative, current ailment, healthcare details, personal information, biological data, and history, sorted, and given descriptions as needed. Below is the resulting table of fields with the description omitted to make better use of space and to increase legibility:

Picked Fields	Type
Date of entry	Administrative
Reason for admission	Administrative, Current Ailment
Symptoms	Current Ailment
Assignment	Administrative, Healthcare Details
Status	Administrative
Date of last status update	Administrative
Treatment plan	Current Ailment, Healthcare Details
Diagnosis	Current Ailment
Name	Personal Information
Address	Personal Information
Sex	Personal Information, Biological Data
Age	Personal Information, Biological Data
Risk Factors	History
Co-morbities	History
Past diagnoses	History, Current Ailment, Healthcare Details, Biological Data
Past procedures	History, Current Ailment, Healthcare Details
Medications	History, Healthcare Details
Height	Biological Data
Weight	Biological Data
Blood type	Biological Data
Blood pressure	Biological Data, Current Ailment
Body temperature	Biological Data, Current Ailment
Pulse	Biological Data, Current Ailment
Date of vital signals measurement	History, Biological Data, Current Ailment
Laboratory results	History, Current Ailment, Healthcare Details, Biological Data
Date of laboratory test	History, Healthcare Details, Biological Data
Clinical notes	History, Current Ailment, Healthcare Details, Biological Data

Table 1. Final fields to be used with categories.

The Patient ID field was omitted despite its prevalence as these are unique identifiers specific to the healthcare institution the data is logged in, making the transfer of these useless and unnecessary.

Using this as a basis, a trial run was done using an entry from each of two different databases by simply laying it out in a Google Sheet, to serve as a proof of concept for this format. This was entry 5 in the Synthetic Electronic Medical Records by Sina Sheikholesami and entry 6 from the Clinical Dataset of the CYP-GUIDES Trial. These are to be referred to as set A and set B respectively for the discussion of this test. These were used for this test as they had extensive and complete fields. Set A's data was first fitted into the proposed format. From there, only the proposed format's data would be transferred to set B's format. This would ensure that set A and set B are "blind" to each other to simulate the actual transfer and processing of data. The same was done for Set B's data into Set A. The process and result can be seen in the figure below:

	A -> B										
A	patient_id	age	gender	race	height_cm	weight_kg	bmi	blood_pressure_systolic	blood_pressure_diastolic	cholesterol_total	cholesterol_hdl
	5	71	M	White	178	95	30	150	95	220	3
OURS	Date of entry	Reason for admission	Symptoms	Assignment	Status	Date of last status update	Treatment plan	Diagnosis	Name	Address	Sex
	-	-	-	-	-	-	-	-	-	-	M
B	ID	GENDER	AGE	RACE/ETHNICITY	Diagnosis	MD	Assignment	EMR	LOS	RAR	A
	-	M	71	-	-	-	-	E	-	-	0

Figure 4. Partial table of conversion set A to B

	B -> A										
B	ID	GENDER	AGE	RACE/ETHNICITY	Diagnosis	MD	Assignment	EMR	LOS	RAR	A
	6	F	24	L	MDD, Single Episode,Severe Without Psychotic Features	T	S	C	120	0	0
OURS	Date of entry	Reason for admission	Symptoms	Assignment	Status	Date of last status update	Treatment plan	Diagnosis	Name	Address	Sex
	12/3/2024	-	-	S	Discharged	4/2/2025		MDD, Single Episode, Severe Without Psychotic Features	-	-	F
A	patient_id	age	gender	race	height_cm	weight_kg	bmi	blood_pressure_systolic	blood_pressure_diastolic	cholesterol_total	cholesterol_hdl
	-	24	F	-	-	-	-	-	-	-	-

Figure 5. Partial table of conversion set B to A

Some fields, marked in red, were not used as they were not applicable to the format they were being transferred to. On the other hand, the fields in green were successfully transferred. Fields that could not be filled used a dash (-) to indicate that they were empty. At first glance, the results seem like they lose a lot of data, however, this could be due to the two datasets needing and having different information.

V. Conclusion

While the study was successful in its main objective of creating a framework for easier data sharing, the actual implementation was absent. Different coding languages were not able to be compared, the data flow was not actualized, and the format was not encoded into a proper digital table. The future iteration of this study plans to include actual implementation in a simple program to demonstrate the applicability of this format into an actual system. The choice of data must also be reconsidered, possibly trying different combinations of sets to get a better idea of how flexible this format is. This would allow for actual feasibility testing and quality control.

References

[1] Singh, N., Birla, S., & Shukla, N. K. (2024). Recent trends in biomedical informatics. In *Machine Learning Models and Architectures for Biomedical Signal Processing* (pp. 3–17). Elsevier. https://doi.org/10.1016/B978-0-443-22158-3.00001-6

[2] Zhang X, Saltman R. Impact of Electronic Health Record Interoperability on Telehealth Service Outcomes. JMIR Med Inform. 2022 Jan 11;10(1):e31837. doi: 10.2196/31837. PMID: 34890347; PMCID: PMC8790688.

[3] Popescu C, El-Chaarani H, El-Abiad Z, Gigauri I. Implementation of Health Information Systems to Improve Patient Identification. Int J Environ Res Public Health. 2022 Nov 18;19(22):15236. doi: 10.3390/ijerph192215236. PMID: 36429954; PMCID: PMC9691236.

[4] Upadhyay S, Hu H. *A Qualitative Analysis of the Impact of Electronic Health Records (EHR) on Healthcare Quality and Safety: Clinicians' Lived Experiences*. Health Services

Insights. 2022;15. doi:10.1177/11786329211070722

[5] Keshta, I., Odeh, A. (2021). Security and privacy of electronic health record: Concerns and challenges 22(2). https://doi.org/10.1016/j.eij.2020.07.003

[6] Electronic Health Records | CMS. (n.d.). https://www.cms.gov/priorities/key-initiatives/e-health/records

[7] What are the advantages of electronic health records? | HealthIT.gov. (n.d.). https://www.healthit.gov/faq/what-are-advantages-electronic-health-records

[8] Robinson, S., & Lee, K. (2024, October 21). What is an electronic health record (EHR)? Health IT and EHR. https://www.techtarget.com/searchhealthit/definition/electronic-health-record-EHR

[9] Healthcare Readers. (2025, March 5). Understanding the Electronic Health Record (EHR): examples, benefits, and best practices. https://healthcarereaders.com/medical-devices/electronic-health-record-examples-and-benefits

[10] Y. Yuan *et al.*, "Data flow construction and quality evaluation of electronic source data in clinical trials: pilot study based on hospital electronic medical records in China," *JMIR Medical Informatics*, vol. 12, p. e52934, Apr. 2024, doi: 10.2196/52934.

[11] Issa, W. B., Akour, I. A., Ibrahim, A., Almarzouqi, A., Abbas, S., Hisham, F., & Griffiths, J. (2020). *Privacy, confidentiality, security and patient safety concerns about electronic health records. International Nursing Review*, 67(2), 218–230. https://doi.org/10.1111/inr.12585

[12] Z. Wang *et al.*, "Protocol for processing multivariate time-series electronic health records of COVID-19 patients," *STAR Protocols*, vol. 6, no. 1, p. 103669, Mar. 2025, doi: 10.1016/j.xpro.2025.103669.

[13] El-Rashidy, N., El-Sappagh, S., Islam, S. M. R., M. El-Bakry, H., & Abdelrazek, S. (2021). *Mobile Health in Remote Patient Monitoring for Chronic Diseases: Principles, Trends, and Challenges. Diagnostics*, *11*(4), 607. https://doi.org/10.3390/diagnostics11040607

[14] Reegu, F. A., Abas, H., Gulzar, Y., Xin, Q., Alwan, A. A., Jabbari, A., Sonkamble, R. G., & Dziyauddin, R. A. (2023). *Blockchain-Based Framework for Interoperable Electronic Health Records for an Improved Healthcare System. Sustainability*, 15(8), 6337. https://doi.org/10.3390/su15086337

[15] Keshta, I., Odeh, A. (2021). Security and privacy of electronic health record: Concerns and challenges 22(2). https://doi.org/10.1016/j.eij.2020.07.003

[16] Omboni S, Padwal RS, Alessa T, Benczúr B, Green BB, Hubbard I, Kario K, Khan NA, Konradi A, Logan AG, Lu Y, Mars M, McManus RJ, Melville S, Neumann CL, Parati G, Renna NF, Ryvlin P, Saner H, Schutte AE, Wang J. *The worldwide impact of telemedicine during COVID-19: current evidence and recommendations for the future*. Connect Health. 2022 Jan 4;1:7-35. doi: 10.20517/ch.2021.03.

[17] Palojoki, S., Lehtonen, L., & Vuokko, R. (2024). *Semantic Interoperability of Electronic Health Records: Systematic review of alternative approaches for enhancing patient information availability*. JMIR Medical Informatics, 12, e53535. https://doi.org/10.2196/53535

[18] T. Sasahara *et al.*, "Assessment of reasons for referral and activities of hospital palliative care teams using a standard format: a multicenter 1000 case description," *Journal of Pain and Symptom Management*, vol. 47, no. 3, pp. 579-587.e6, Aug. 2013, doi: 10.1016/j.jpainsymman.2013.04.009.

[19] Chatterjee, A., Pahari, N., & Prinz, A. (2022). *HL7 FHIR with SNOMED-CT to Achieve Semantic and Structural Interoperability in Personal Health Data: A Proof-of-Concept Study*. *Sensors*, *22*(10), 3756. https://doi.org/10.3390/s22103756

[20] Yan, C., Yan, Y., Wan, Z. et al. *A Multifaceted benchmarking of synthetic electronic health record generation models*. Nat Commun 13, 7609 (2022). https://doi.org/10.1038/s41467-022-35295-1

[21] "IEEE DataPort," *IEEE DataPort*. https://ieee-dataport.org/

[22] "Kaggle: your machine learning and data science community." https://www.kaggle.com/

[23] P. Rabaey, "SynSUM – Synthetic Benchmark with Structured and Unstructured Medical," *IEEE DataPort*, Jan. 21, 2025. https://ieee-dataport.org/documents/synsum-%E2%80%93-synthetic-benchmark-structured-and-unstructured-medical-records

[24] "Clinical Dataset," Kaggle, Oct. 05, 2023. https://www.kaggle.com/datasets/imtkaggleteam/clinical-dataset

[25] "Clinical dataset of the CYP-GUIDES trial," *Kaggle*, Feb. 20, 2021. https://www.kaggle.com/datasets/shashwatwork/clinical-dataset-of-the-cypguides-trial

[26] "Synthetic Electronic Medical records," *Kaggle*, Apr. 20, 2024. https://www.kaggle.com/datasets/sinasheikholeslami60/synthetic-electronic-medical-records

[27] "electronic medical records EMR," *Kaggle*, Jul. 31, 2024. https://www.kaggle.com/datasets/davidvictor297/electronic-medical-records-emr

[28] "mexwell | Expert," *Kaggle*. https://www.kaggle.com/mexwell

[29] W. R. O. for the W. Pacific and W. H. Organization. R. O. for the W. Pacific, Electronic Health records: A Manual for Developing Countries. WHO, 2006.

[30] A. Gamal, S. Barakat, and A. Rezk, "Standardized electronic health record data modeling and persistence: A comparative review," *Journal of Biomedical Informatics*, vol. 114, p. 103670, Dec. 2020, doi:

10.1016/j.jbi.2020.103670.

[31] GeeksforGeeks, "Python Database Tutorial," *GeeksforGeeks*, Mar. 15, 2023. https://www.geeksforgeeks.org/python-database-tutorial/

[32] O. Fennelly, D. Moroney, M. Doyle, J. Eustace-Cook, and M. Hughes, "Key interoperability Factors for patient portals and Electronic health Records: A scoping review," *International Journal of Medical Informatics*, vol. 183, p. 105335, Jan. 2024, doi: 10.1016/j.ijmedinf.2023.105335.