ACCSCIENCE PUBLISHING

Artificial Intelligence in Health

REVIEW ARTICLE

Optimizing electronic health records to support artificial intelligence

Evelyn J. S. Hovenga[1,2]* and Koray Atalag[3]

[1]Department of Digital Health, Faculty of Health Sciences, Australian Catholic University, Fitzroy, Victoria, Australia

[2]eHealth Education Pty Ltd, Abbotsford, Victoria, Australia

[3]GALATA-Digital LLC-FZ, Dubai, United Arab Emirates

Abstract

Electronic health records (EHRs) provide the most important data sources for artificial intelligence (AI). Gaining access to quality data suitable for advanced analytics continues to be challenging. This rapid review documents the current state of available data; identifies foundational AI data/information needs; and explores the benefits of adopting new and emerging technologies to design and implement next-generation EHRs. Opportunities to optimize EHRs for AI purposes are identified. This review was informed by expert knowledge and shared experiences supported by the literature, including technical standards. Main findings include poor ecosystem-wide infrastructures due to the lack of adopting the right set of standards, and current data and knowledge governance no longer fit for purpose. While many jurisdictions are continuing the use of legacy systems, some forward-looking national health systems and health-care facilities are adopting transformational strategies by adopting a strong data and digital focus to transition to new-generation systems. New foundational-level national infrastructures with strong leadership and governance are essential to enhance the governance and quality of available data, from collection at source throughout the entire data supply chain. Secure and ubiquitous access to high-quality EHR data at scale will foster the evolution of more intelligent and trustworthy AI. Key characteristics of next-generation EHRs supported by currently available technologies and standards that are able to meet digital era demands are provided in this paper. We conclude that the use of generative AI in clinical settings can only be reliably achieved when EHRs are optimized throughout the entire global digital health ecosystem.

Keywords: Ontology; Standards; Terminology; System architecture; Models; Data; Electronic health records

***Corresponding author:**
Evelyn J.S. Hovenga
(e.hovenga@ehe.edu.au)

Citation: Hovenga EJS, Atalag K. Optimizing electronic health records to support artificial intelligence. *Artif Intell Health.* 2024;1(3):10-25. doi: 10.36922/aih.3056

Received: February 29, 2024

Accepted: June 5, 2024

Published Online: July 24, 2024

1. Introduction

Artificial intelligence (AI), which has progressed concurrently with the introduction and adoption of computers, has attained immense developments in recent years. The collective utilization of data science, AI, information, and communication technologies can potentially enhance or transform the health-care industry. Their successful use augments the potential for people to gain a greater insight into any health-related

problems and/or to consider how best to achieve desired health outcomes.

AI is used as an aid to problem solving. When used to support clinical decisions, AI cannot operate like a "black-box." For AI to be trusted by health-care professionals, these insights, inferences, and decision supports need to be explainable and repeatable. All types of data/information can be made use of by AI. In the health-care domain, all relevant legislative, regulatory requirements, and ethical principles need to be complied with for obtaining the authorization to use of these resources.

Many of AI data resources come from electronic health records (EHRs) and electronic medical records (EMRs). Based on the International Organization for Standardization (ISO) Technical Committee (TC) 215 definitions, an EHR is a health record in computer-processable format – such records have atomic data elements, representing the smallest possible concept such as genetic data, which can be analyzed and processed by a computer. Medical (healthcare) records are health records produced for and used within a health-care organization's enterprise system. Many health-care organizations begin their digital health transformation journey by scanning paper-based records. The content of EHR needs to enable continuous monitoring of people's health throughout a person's life span, facilitate continuity of care, and be informative and accessible to them anytime from anywhere, as well as accessible to others authorized to do so with consent.

EMRs represent records produced for and used by one health-care provider (independent or organizational). A growing number of EMR suppliers are providing access to other providers as well as their patients through a dedicated portal. Health information exchange (HIE) refers to the concept of exchanging clinical and administrative data across different systems and stakeholders, by connecting the platform to transactional systems. In most cases, EHR data can be gleaned from a multitude of proprietary EMR and HIE systems as well as a myriad of other clinical, administrative, and ancillary systems. This current fragmented landscape of health information systems' data acquisition, storage, and use is a barrier to the effective provision of continuous and person-centered care as well as trustworthy use of AI methods.

As clinicians, our focus is on ensuring that health data used for AI purposes are complete, accurate, trustworthy, and able to support person-centered precision medicine. This requires a major overhaul of this fragmented landscape by facilitating collaboration among diverse health service providers and across different types of care. Only then are we able to generate timely quality data to reliably support AI technologies. This paper explores what our future infrastructure needs are to ensure flexibility and agility to adapt to new requirements and shifting priorities. It is about informing politicians, top public officials, senior decision-makers, and procurement officers as well as data scientists working for the health industry.

1.1. Review aims and objectives

The aims and objectives of this review are as follows:

- To identify the current state of data availability, and quality for AI development use.
- To identify foundational data/information AI resource needs.
- To explore the benefits of new and emerging technologies used to design and implement next-generation EHR/EMRs.
- To identify new-generation capacity to optimize AI developments.

1.2. Research question

What next-generation technologies need to be in place to optimize EHRs for AI purposes?

1.3. Rapid review method

The scope of this modified scoping (rapid) review paper is confined to the use of data resources as stored for use by current and next-generation EHRs/EMR systems for AI purposes. This review was designed to benefit high-level decision-makers who need a conceptual understanding of current issues and how these are best addressed. The selection of references used was primarily informed by the authors' expert knowledge and shared experiences between standards development experts and known researchers working in the digital health space. This review was supplemented by targeted research methods based on literary information sources, including the most up-to-date gray literature and known published literature not retrieved by database searches.

2. Issues associated with the current state of health data

Individual health-care organizations began the development of EHR/EMR systems alongside their evolving computer science, information and communication technologies, and information systems research and development activities. Some of these early developed systems have evolved over more than 30 years by taking advantage of new insights, newly discovered technologies, new programming languages, and frameworks and evolving connectivity solutions, to become some of the mega-EMR enterprise systems still in use today. These are now legacy systems

based on proprietary system architectures requiring the adoption of transformation strategies.[1] The bigger the health ecosystem, the greater the difficulty to change foundational architectural system design. Ingram has documented these historical developments and evolutionary discoveries in great detail.[2] Ingram also explains the scientific foundations of new discoveries over time.

The literature included in this review has collectively exposed that most current health ecosystem governance infrastructures, including legislation, regulations, and policies, essentially determine not only any nation's health ecosystem infrastructure but also the governance of its data, information, knowledge, and wisdom assets. Such high-level infrastructures determine their strategic directions including the mandating of standards compliance.[3] Poorly informed policy decisions have resulted in numerous costly failures and continue to impede efficient progress.[4-8]

A lack of foundational technical knowledge in emerging developments and the continuing use of legacy systems has resulted in a large digital health ecosystem-wide architectural patchwork.[9] Natural language processing is compromised due to the prevalence of duplicate information in EMR systems;[10] secondary data use or advanced data analytics is compromised by incomplete data from EHRs.[11-13] We are also witnessing a continuing proliferation of applications (Apps). Most apps are standalone and unable to share data with EHR/EMR systems making most EHRs incomplete and unable to provide a comprehensive overview of a person's health status at any point in time and across different tiers of the health-care system. Incomplete EHRs represent a significant patient safety issue.

Software developers tend to focus on meeting procurement requirements, with a focus on how they can best meet market needs and be competitive. Consequently, we continue to have chaos, fragmentation, and data silos as very few of these decisions are being coordinated to best suit the digital health ecosystem. Its impact is that collectively we are generating large amounts of real-world data that cannot be used effectively. Yet, health data represent a valuable asset that needs to be well governed and managed.

The impact of the continuing fragmentation within any health ecosystems not only contributes to physician burnout[14] but also limits AI developments aiming to support clinical practice. This limitation is primarily due to health ecosystems' inability to manage all relevant data flows[15] required to compile a comprehensive and complete health record required to support any person's continuity of care. Incomplete health records, in turn, prevent the aggregation of quality data sets required to make "big health data sets" available for research and multiple other use cases. For example, Zhang *et al.*[16] found that multi-stage data flow chains in the UK do not fulfill recommended best practices for safe data access and that its existing infrastructure produces aggregation of duplicate data assets. Multi-stage data flow chains limit the diversity of data required to add value to end users.

> *"There are gaping holes in data platform infrastructure that supports deployment of data-driven tools (such as digitally/AI-enabled trials, or AI deployment)."*[16]

Information, Communication, and Technology (ICT) and Information Systems (IS) research tend to be undertaken as action research (through trial and error, providing local solutions to problems identified). The action research approach enables constant evaluation of implementation providing a process of checking for and affirming understanding that is specific, non-evaluative, manageable, and focused on the target of interest (assessments feedback loops). Standards development activities consist of collaborative problem solving, making use of international experts, and user feedback regarding issues encountered when testing new standards.

The ISO TC 215 is responsible for developing standards specifically to suit the health industry along with a few other Standards Development Organizations (SDOs), including Health Level Seven (HL7),[17] SNOMED International[18]), Digital Imaging and Communications in Medicine (DICOM), and Clinical Data Interchange Standards Consortium (CDISC).[19-22] A number of these SDOs are working collaboratively through the Joint Initiative Council (JIC) established in 2007.[23] However, it needs to be remembered that few governments have mandated compliance with any specific set of standards although this is changing.

2.1. Interoperability

The interoperability issue is primarily being addressed by ICT professionals making use of various versions of HL7 messaging standards for data exchange. Their implementation and use require extensive data mappings between proprietary data models and these standards. All health-related concepts need to be represented by data in a re-interpretable form to represent information in a formalized manner suitable for communication, interpretation, or processing by people or by automation.

Data mapping frequently results in a loss of information. Health data can be represented by any one of many terminologies. Anecdotally, we learned that many data mapping activities are undertaken by administrative staff not necessarily suitably qualified to accurately interpret the meaning of terms or codes to accurately retain meaning when mapping data from one data set to another. Some

terminologies represent the same knowledge domain but are structured and coded differently either as classifications, such as the International Classification of Diseases (ICD), or designed according to ontological principles to ensure each code is mutually exclusive of another, such as SNOMED CT (Clinical Terms), a global language representing clinical terms. Not only is mapping time consuming and costly to maintain given frequent terminology updates but also mapping errors introduce risks for patient safety. Few data maps are independently quality-assured, nor are those undertaking a mapping activity contracted to achieve a specified quality level. This data map quality issue has resulted in the development of an ISO standard[24] designed to address this data quality issue.

The shortcomings of using evolving messaging standards to represent clinical information have long been recognized. The continuing use of multiple versions of messaging standards, which focus on syntactic interoperability, has resulted in methods and standards going beyond the data level, such as the openEHR archetypes (data models)[25] and HL7 CDA (Clinical Document Architecture)[26] and HL7 FHIR (Fast Health-care Interoperability Resources).[27] Archetypes bring together relevant data items and clinical or health-care context to define composite clinical concepts such as blood pressure, laboratory results, medication lists, and prescriptions in a manner to suit any possible use case. These models may also contain terminology bindings where some of the data elements are linked with corresponding clinical terminology, such as SNOMED CT, ICD, or logical observation identifiers and codes (LOINC).[28] The application of conceptual modeling plus attribute binding to standard terminologies ensures that context and meaning are retained to guarantee a high degree of semantic interoperability within and between EHRs, significantly improving data quality. However, adoption of these standards by vendors has been slow due to a lack of effective regulatory or commercial mandates and incentives.

2.2. Data quality and interoperability

The value of common data models (CDM) was identified during the early 1990s. The adoption of a CDM empowered collaborative research across competing organizations.[29] This finding led to the establishment of the CDISC.[22] International collaborative research has demonstrated that semantic interoperability could be achieved by creating a CDM shared by all data contributors as these CDMs define central concepts, their attributes, constraints, and relations. CDM adoption allows for the pooling of information so that meaningful comparisons can be made.

Every EHR/EMR system is a potential data contributor and continues to make use of its own data reference model to structure its data repositories. The lack of widespread adoption of CDMs by EHR/EMR systems and issues with enforcement (governance) continues to be a major limitation. In the health-care domain, a CDM usually refers to a Clinical Data Model. openEHR has adopted a two-level modeling approach that separates its universal archetypes from applications,[30,31] as a means of optimizing semantic interoperability.

The lack of effective collaboration between ecosystem-wide stakeholders, including citizens, clients, patients, funders, health-care providers, researchers, and other institutions (data users) over time, has resulted in poor data access and data quality. Data use is limited to built-in system functionality, including reporting functionalities, as many multi-modal systems have difficulty interfacing with external systems.[32] Consequently, meaningful data aggregation to create large accessible databases or to ensure all health data pertaining to one individual is accessible through one record is limited. These represent major barriers for AI development.

As a consequence of poor quality and incomplete datasets, substantial research time, money, and effort is spent on "data cleansing" activities designed to improve data quality.[33] Data cleansing undertaken for medical AI systems can have negative effects on data quality if not performed carefully. Data cleansing can have dramatic harmful implications.[34] Stöger *et al.*[34] listed and described the following quality problems associated with the use of original data, which data cleansing activities are meant to mitigate. These are as follows:

- Absence of data – blank fields
- Dummy/default values – may be difficult to detect
- Noise (also known as the butterfly effect)
- Wrong data
- Inconsistent data
- Cryptic data
- Duplicate primary keys
- Non-unique identifiers
- Multipurpose fields
- Violation of (business) rules

Data processing sometimes requires conversion of numerical data to strings to represent a concept in words, representing another potential risk as it can lead to later issues.

2.3. Data sharing

This existing knowledge gap regarding data sharing and the need for quality data needs to be acknowledged and addressed by policy makers and research funders[35] as well as by those developing AI applications. The availability of a public library of terminology value sets enables clinical

information models and standard terminology value sets to work together to create a coherent data ecosystem.

The most accurate and adaptable method for representing computable clinical knowledge is through a dual information architecture model, which enables the development of clinical information models built from common reference components. Some existing strategies include data sharing through the use of cloud technologies and federated clinical data repositories (CDRs) to provide access to large amounts of data. CDRs need to enable reuse of data while preserving the data's original meaning and context.

Effective data sharing requires a strong data management strategy and framework including the creation of standardized, centralized processes around ingesting, classifying, storing, organizing, linking, and maintaining data. Centralization and linkage of health data on the cloud raises many security and privacy concerns as well. The use of cloud technologies to store data has the advantage of the ability to retrieve data using any type of device anytime. A major cultural shift is required to move to externally hosted services and the adoption of one set of compatible standards. CDRs need to be able to support timely health-care delivery, research, and public health initiatives as well as facilitate the creation and efficient implementation of decision-support tools. Many beneficial advances made to date are not necessarily visible to those providing frontline care.[36]

2.4. Continuing use of legacy systems

There is a desire to make the best possible use of our legacy systems to sustain existing profitable business models, to make the best possible use of significant investments made, and to maintain access to historical data. The market continues to be dominated by a few mega-EMR providers and numerous other legacy systems who are making their own data sharing arrangements, such as the HL7 Argonaut project, a private-sector initiative[37] designed to advance industry adoption of open interoperability standards. This represents a small step toward a digital transformation but is limited to users of the same enterprise-wide EMR system and its proprietary platform.

Recent collaboration managed by the Commonwealth Scientific and Industrial Research Organization's (CSIRO) Australian e-Health Research Center has resulted in the first release of the Australian Core Data for Interoperability (AUCDI) release for community comment. This collaborative consortium set out to build robust HL7 FHIR[27] profiles, extensions, and terminology value sets and bindings. This consortium's initiative (SPARKED) has launched a national FHIR Accelerator program to reinforce the move toward digital healthcare across Australia.[38] The consortium made use of universal computable clinical models (Archetypes)[39] mapped to SNOMED CT or LOINC,[40] *etc.*, which are utilized in these HL7 FHIR artifacts. The resultant AU core data set does not specify how and to what extent its elements are included in FHIR or other exchange standards. SPARKED represents another small evolutionary step toward improving data quality. While continuing to make use of legacy systems, these new initiatives need to be viewed as transitional arrangements.

3. Clinical data asset use

This review has identified a number of risk factors to be considered when extracting and collating data/information for the purpose of AI use from EHR/EMR systems. The New South Wales Government has identified these within their comprehensive AI Assurance framework[41,42] informed by groups of standards developed by the International Electrotechnical Commission (IEC)/ISO/and Joint TC (JTC1) family of SDOs. The New South Wales Government strategy includes the following key risks that need to be mitigated. These risks include:

- The use of incomplete or inaccurate data
- Having poorly defined descriptions and indicators of "fairness"
- Not ensuring ongoing monitoring of "Fairness Indicators"
- Decisions made to exclude outlier data
- Using informal or inconsistent data cleansing and repair protocols and processes
- Using informal bias detection methods
- The likelihood that re-running scenarios could produce different results (reproducibility)
- The inadvertent creation of new associations when linking data and/or metadata
- Differences between the data used for training compared to actual data
- Missing from this list was not ensuring that scenarios can be explained, which is a requirement for the generation of trustworthiness (explainability).
- Some of the questions to be answered by AI developers include:
- Is the data needed for the project in question available and of appropriate quality given the potential harms identified?
- Does your data reflect the population that will be impacted by your project or service?
- Have you considered how your AI system will address issues of diversity and inclusion (including geographic diversity)?
- Have you considered the impact regarding minority and disadvantaged groups?

- Do you have appropriate performance measures and targets?
- Do you have a way to monitor and calibrate the performance of your AI system?
- How will sensitive data be handled?

> *".…….development and implementation of AI technologies must be undertaken with appropriate consultation, transparency, accountability, and regular, ongoing review to determine its clinical and social impact and ensure it continues to benefit, and not harm, patients, health-care professionals, and the wider community."*[43]

A number of standards have been developed or are in development by the ISO/IEC/JTC1 Standards Committee 42, to assist all of us to responsibly develop, and make use of AI technologies including one for data quality for analytics and machine learning, data visualization to assess data quality and an AI data framework. The World Health Organization (WHO) has recently published its regulatory considerations on AI for health.[44]

There are numerous relevant standards for data sharing and use,[42] which cover data in general; these standards are not specific for health or clinical care data. A number of guidelines,[45-47] as well as a data governance framework,[48] have also been developed. Similar initiatives are being undertaken in other jurisdictions.[49,50] All of these measures are designed to improve data quality, streamline our use of data, and support AI development.

3.1. Data, information, and knowledge management requirements

Every known health-related terminology standard is based on an agreed categorial structure. Many of these do not identify as a formal ontology that represents a specific knowledge domain, thus resulting in ambiguities. A formal ontology consists of classes, instances, relations, functions, and axioms to reflect meaning by providing context. This allows for a clear digital understanding of concepts representing a defined knowledge domain. Terminologies were originally structured and developed to suit paper-based systems. An ontological design determines the knowledge domain's structure that enables semantic interoperability.[51]

In this digital era, it is important for standard terminologies in use to comply with the ISO standard[52] that specifies how categorial structures of terminologies need to be represented. The purpose of this ISO standard includes the need to support the development of specific standards of categorial structures for particular health-care subject fields with the minimum requirements to support meaningful exchange of information. The categorial structure approach recognizes the need for terminologies and classifications to be able to provide content related to a range of concepts and how those concepts impact the requirements of the terminology. One example is the categorial structure of nursing practice,[53] which may also be applicable to represent all types of clinical practice requiring the use of their own terminology.[54] The representation of concepts and characteristics need to especially be described in this manner for use in formal computer-based concept representation systems. Categorial structures also show relationships between categories and sub-categories.

The SNOMED CT terminology is an ontology-based comprehensive medical terminology used by many international members for standardizing the storage, retrieval, and exchange of electronic health data.[55] It is able to represent each data element and identify it together with a code. The WHO develops and updates a family of health-care terminologies including the ICD. Version 11 has an updated structure based on the use of ontology-driven tools[56] as one strategy designed to improve semantic interoperability.

Interoperability among systems requires the harmonization of such models; a project was undertaken by the Office of the National Coordinator for Health Information Technology[57] between 2017 and 2019 to advance the utility of observational data for Patient-Centered Outcomes Research (PCOR) and its interoperability across multiple networks. The PCOR project resulted in four clinical data models: (1) Sentinel, (2) PCOR Network (PCORnet), (3) Informatics for Integrating Biology and the Bedside (i2b2), and (4) Observational Medical Outcomes Partnership (OMOP).[57]

3.2. AI data quality objectives

This review has demonstrated that the design of these next-generation systems needs to be able to address the following key requirements to optimize data availability for AI development and applications:

- Ensure we have access and are able to make use of, the maximum number of data points at any required level of granularity as required to develop reliable accurate algorithms to suit AI application development. Accessing a maximum possible number of multiple desired data points needs to be achieved through linking and aggregating health-care data at scale and safely, across different tiers of care and multiple organizations, using interoperability standards and vendor-neutral data infrastructures.
- Maximize automation of routine reporting.
- Safeguard patient safety and ethical data use.

Every data point represents a single unit of information. For AI purposes, it is necessary to make use of a defined

collection of data points to determine if a pattern exists, or for algorithm development to make decisions or support decision-making or make predictions. Training any AI model requires large amounts of representative data. The number and types of accessible data points determine the accuracy of the model or a possible set of rules that can be identified. The delivery of health services is data centric where access to accurate and timely data is critical for decision-making. AI approaches making use of these data require the use of advanced analytics and access of a large amount of source data. Data-driven approaches are relevant for the provision of automated reporting as automation relies on pre-determined rules or assumptions.[58] There are significant limitations regarding access to source data collected and stored in legacy systems.

4. New and emerging technologies

This review's findings have confirmed that the interoperable and scalable ecosystem-wide architectures can be adopted, the knowledge about the health ecosystem's data supply chain, and the relationships between information models, terminologies, and ontologies with data exchange protocols. Health ecosystem-wide data supply chains need to:

(1) Include data/information flow requirements to support collaborative, person-centered life-long, and episodic continuity of care. Episodic events of multiple service episodes can also exist. Such episodes represent a treatment plan for one specific health issue such as for cancer care or a pregnancy as recorded by multiple systems over a period of time. Data collections able to meet all information needs associated with any treatment/care plan require identifiable data transfers between any number of individual and organizational health-care service providers as well as devices. Specific data needs will differ based on the individual's health status, treatment/care plans (life-long and episodic), and geographical location relative to service availability at any point in time.

(2) Facilitate the aggregation of de-identified data and identifiable data to classify any number of grouping protocols (populations) or individuals to suit specific data use cases. Data relationships will vary by use case and need to include data from systems other than data collected and stored by EHR/EMRs, such as clinical registries. Over time, such registries are expected to be generated from vendor/technology-neutral federated cloud-based health data repositories including CDRs. For some use cases, linkages may also need to include relationships between weather events or environmental status at a specific point in time or by geographical location, such as vaccination rates. CDR design needs to prioritize the separation of health information from citizen demographic or identification data by adopting a privacy-by-design approach. Every citizen needs to have control over their data and how it is used.

(3) Facilitate the linkage of health-care data with omics data, that is with the inclusion of data representing the various "omes" of an organism, to enable making sense of vast amounts of collected data to build next generations of clinical decision support and research methods and tools. At present, the use of genetic sequencing and variation information is not part of routine clinical practice because health-care professionals do not have the knowledge or skills. Most importantly, there is a lack of automated tools that can reliably associate phenotypic data from EHRs with many types of omics data to provide personal and precision care. Large-scale, well-annotated, and high-quality EHR data will have an immense impact on bringing omics and healthcare together.

(4) Facilitate the linkage of healthcare and data with the human physiome[59,60] comprising personal and mechanistic computational multi-scale models. Such models enable the provision of new types of insight into not only our understanding of human physiology and pathology but also predictions of disease and prognosis. Such insights are the result of using ontology-based EHR data linkage that parameterize these models that are able to run surprisingly reliable simulations at individual or population levels. Computational physiology and systems biology provide us with unprecedented precision to provide value-based and appropriate care as well as drive more effective drug and medical device development and faster compliance through *in silico* medicine and clinical trials.[61]

Data governance protocols, legislation, and regulations need to facilitate or enable these requirements to deliver optimal benefits of data use, including any type of effective reporting automation and AI adoption.

4.1. Next-generation EHR/EMR system characteristics

Next-generation EHR/EMR systems are designed to reduce or eliminate these gaps and improve the generation of quality data within a connected digital health ecosystem. New health platforms need to be engineered to integrate personal health information received from emerging technologies in the fields of personal health and well-being, including apps and wearables. EHR/EMR systems and CDRs should become a valuable computable data source for research and evaluation purposes as well as be enriched by data from external data sources while complying with

relevant regulatory frameworks. New systems are now adopting advanced technologies including cloud-based open (non-proprietary) ecosystem-wide platforms and openEHR's modeling approach to improve health data management. They are designed to enable plug-and-play of any number of new devices and niche applications, through architectural standards and frameworks like SMART-on-FHIR,[27] without losing the ability to share data.

Underpinned by open standards-based federated CDRs,[62] an effective separation of data and application becomes possible. The adoption of open standards enables secure access to vendor-neutral data by compliant third-party applications across the whole ecosystem. Fully standardized health data can be aggregated and utilized for many authorized purposes, including AI.

Ecosystem-wide architectural design is paramount to maximize these potential benefits. Semantic interoperability requires extensive use of ontologically structured knowledge domains[63] and ontology-driven architectures[64] as explained in details by Rector *et al.*[65] Its structure is based on the relationships between three resources: (1) Information models representing, for example, clinical concepts; (2) inference models; and (3) concept system models required to reliably undertake data abstraction – a process adopted to reduce a concept to a set of essential elements.[64,66-68]

Changing over from the legacy systems' data/information exchange paradigm to knowledge sharing at decreasing levels of abstraction requires the adoption of a reference architecture that starts at the IT concept level (semantic coordination), through the business domain concept level (agreed service function level cooperation), domain level (cross-domain cooperation), and up to individual context (skills-based end-user collaboration).[69] This architectural model supports ontology/knowledge harmonization to enable interoperability between, and integration of, systems, standards, and solutions at any level of complexity without the demand for continuous adaptation or revisions of those specifications.

Those marketing the next-generation systems have some difficulty gaining a foothold in this market as large vendors continue to protect their lucrative business models unless governments intervene. Most countries have established a national digital health framework, but these tend not to include the establishment of a suitable national supportive infrastructure designed to optimize data sharing and data quality.

4.2. Knowledge-driven architectural models and standards

The adoption of a knowledge-driven architectural model means that new-generation systems will have far greater capabilities to support life-long and person-centered care, ecosystem-wide safe data sharing through semantic interoperability, extensive automation of routine reporting, secondary data use, and advanced analytics. Widespread adoption of data standards and data governance protocols is expected to substantially reduce the need for data cleansing. Greater availability of timely, complete quality data is expected to reduce the costs of routine reporting, medical research, and other secondary data use including the development, training, and use of AI. The optimization of EHR data is expected to transform our AI capacity and the health-care industry generally.

A set of compatible standards enabling the establishment and maintenance of a well-connected national digital health ecosystem needs to be mandated. The adoption of data-driven digital health implementation strategies is now happening in some jurisdictions, such as the UK National Health Service (NHS),[70] Spain,[71] Netherlands,[72] Scandinavian countries,[73] United Arab Emirates, Kingdom of Saudi Arabia, and Jamaica,[74,75] as well as some health-care facilities. In 2019, the European Union published its Common Semantic Strategy for Health.[76]

Our work in the digital health space has identified an urgent need for new knowledge to be acquired regarding the scientific underpinnings of health data science among the health workforce.[77] Such knowledge and skills need to be applied to foster a data use culture to enable greater innovation and transformation. Only then are we in a position to improve the overall health system's performance.

This review found that essentially there are three relevant technical standards that need to be considered: openEHR,[62] HL7 FHIR,[27] and the ISO 13606:2019.[78] openEHR provides open standards for the structure, storage, and exchange of health-care information.[79] Core openEHR specifications[80] have been adopted by ISO,[78] making it a full international standard which underpins many national programs and vendor implementations worldwide.[81] The ISO 13606 standard consists of five parts and was based on the openEHR specification, making these two standards highly compatible.

4.3. openEHR

openEHR represents the evolutionary result of more than 20 years of research, innovative development, testing, implementation, and evaluation undertaken by a growing international community.

The openEHR archetypes represent health-care concepts (such as blood pressure measurement, laboratory results, and diagnoses) captured in clinical records and messages based on stable technical building blocks.[82] Its reference model defines generic but healthcare-specific

data structures, types, and value sets and a universal EHR architecture designed to process the high-level components of an EHR categorized as folder, composition, section, and entry. The entry category is further categorized into observations, evaluations, instructions, and actions. Its specifications include a platform model. Its open clinical data repository provides access to nearly 1000 archetypes to date. Collectively, these consist of the largest number of data points representing various levels of granularity in the world.[30] These archetype models are evidence-based, developed, and reviewed by well over 1000 multidisciplinary experts from 114 countries. The number of available archetypes is growing exponentially relative to openEHR-compliant implementations. The development of universal openEHR archetypes is undertaken by the potential data users who have a sound understanding of context that must be represented.

A further modeling layer is the openEHR templates which gather one or more archetypes and define use-case-specific constraints (*e.g.*, discharge summary, medication order, and clinical reports). Templates are used to drive information systems. Archetypes and templates, along with annotated clinical terminology and ontology concepts, define domain-specific information models and enable semantic interoperability in healthcare through this multi-level modeling approach. Their use for the management of clinical data especially optimizes the use of EHR data for AI purposes.

The Archetype Query Language allows formulation of portable queries using domain concepts unlike field or table names in a traditional relational database.[83] Several examples have been reported elsewhere.[84] At the knowledge level, openEHR also defines a formal clinical guideline specification (GDL) to drive decision support – all in a single standards stack.[85] There is also ongoing work to model and capture health-care processes and clinical workflows.

The openEHR Clinical Knowledge Manager (CKM) is an online clinical models repository (archetypes, templates, and clinical terminology subsets) and an advanced web-based distributed knowledge curation tool.[86] CKM supports an editorial process resembling peer-review process of a scientific journal where editors with the help of domain experts can conduct online reviews using the CKM's web interface and then publish models. Figure 1 shows the relationships between ontologies, models, and systems.

4.4. HL7 FHIR

The HL7 FHIR standard[27] represents an evolutionary result of ongoing developments of the HL7 messaging standards. Its implementation is gaining momentum. Pedrera-Jiménez *et al.*[71] explored if these three standards could work together. They found all three to be useful for the purposes for which they were designed but that they have limitations when used for different purposes. Selecting the most suitable set of standards able to best meet a defined purpose is critical.

Information models in FHIR are called resources. All clinical and other findings can only be represented using a single observation resource, which then requires adapting and bringing together many of them together as a profile to be able to represent even a simple concept like blood pressure. However, this HL7 FHIR design has been a deliberate compromise for the simplicity of technical implementation by developers at the expense of its expressivity. As a result, FHIR is being widely adopted by many vendors and health systems. There is now a common trend to use both FHIR and openEHR together, representing demographic, administrative and simple clinical data exchange using FHIR and rich clinical data using openEHR. Archetypes enable access to far more detailed clinical data points as they represent the maximal data elements for a given concept. The FHIR

Resources, on the other hand, include minimal data elements that have been adopted by current health information systems. Therefore, FHIR allows for rapid data exchange between legacy systems.

4.5. What's possible with next-generation EHR systems and data?

Current trends on building digital twins[88] exclusively using EHR data and AI without using atomic-level omics data and mechanistic knowledge and constraints of human physiology and anatomy will fall short of driving next-generation clinical decision support tools and research. Such limitations are due to not being able to train AI with reliable data. At present, available data are inevitably flawed by incorrect facts and associations as well as bias due to known shortcomings of most EHR systems in use and available data sets.

Atalag has defined an ontology mapping-based framework leveraging a multitude of existing and mature standards to bring together all these data sources (EHR, physiome, and omics) in a way that preserves the clinical, biological, physiological, and anatomical context and semantics that can drive these next-generation methods and tools,[89] as shown in Figure 2. This model represents an ontology mapping-based framework to show how compliance with various types of data exchange standards enable an openEHR-compliant clinical data repository to be populated and enable precision medicine supported by AI.

Current proprietary EMR/EHR systems and infrastructures are no longer considered "fit for purpose" due to their many shortcomings. When machine learning

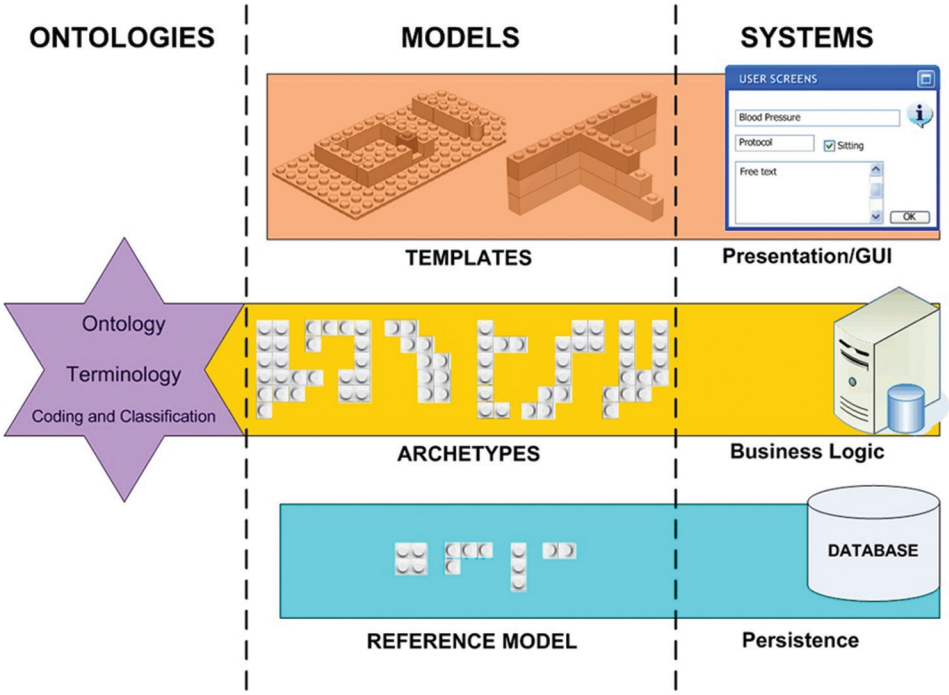

Figure 1. OpenEHR multi-level modeling[87]. Copyright © 2007 Author(s)

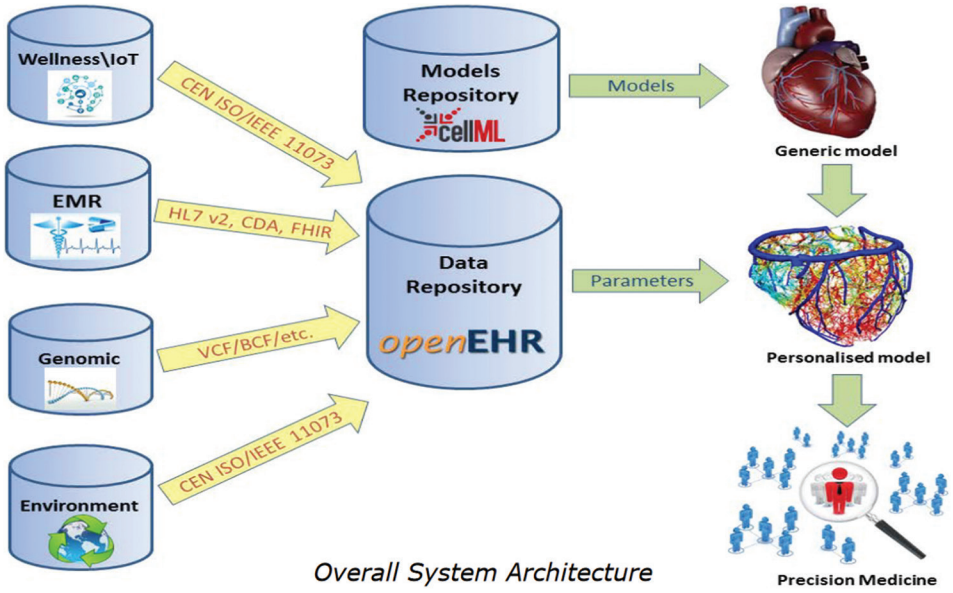

Figure 2. Ontology-based framework for an overall system architecture (adapted from a multitude of similar images developed to represent the openEHR architecture[80] and its relationships to external health data sources, exchange standards, and outcomes). Copyright © 2017 Author(s)

systems are applied to these medical data, there is an imminent and real danger of feeding AI algorithms with non-optimal EHR data which are highly likely to end up in the "garbage in garbage out" problem.[90]

There is an opportunity to capture high-quality and complete structured and computable health-care data as part of routine clinical practice. Post-ad-hoc data collection has been shown to be very expensive and error-prone, and at times, it is impossible to capture the clinical context in which the data were collected.[91] Data sources can be very diverse and range from operational EMR systems to well-structured longitudinal disease registries

Table 1. Ecosystem high-level building blocks enabling optimally connected digital health

Legislation, policy, and compliance[3]	Policy directions and associated legislation enabling the provision and compliance monitoring of health service funding mechanisms enabling universal, person-centered healthcare through an optimized primary care infrastructure supported by a well-specified and mandated digital health infrastructure. This includes privacy and cyber security legislative and regulatory requirements.
Ecosystem system architecture[20,92]	An ecosystem architecture needs to include an open platform able to meet the needs of all stakeholders and support the data supply chain. "*Better data and regular data use will create a data use culture, leading to better decisions, an improved health system, and improved health outcomes.*"[93]
Ecosystem data architecture[94]	A representation of concepts and their relationships. The data architecture defines concepts, constraints, and rules, which provide safe consistent data collection and use which retains meaning throughout the data supply chain. The domain or discourse contribute to the architectural requirements and select data from the data ecosystem based on their use case. Such data collections result in data that are structurally independent, simpler, and safer to share. The resulting lack of data silos enables advanced data analytics.
Concept representation standards[18]	Key health concepts need to be represented in the same manner throughout any digital health ecosystems to ensure data accuracy, enable consistent quality data collection at every level, optimizing data analytics, and reducing data collection burden. Consistent representation of key health concepts enables evidence-based decision-making at all levels and is best achieved by adopting a multilevel modeling approach and an open platform.
Data/information governance[48]	Specification of decision rights and an accountability framework to ensure appropriate behavior in the evaluation, creation, storage, processing, use, archiving, and deletion of information. Coordinated data governance applies to all points along the data supply chain.
Data access control	Legislation is required to indicate who can have access to identifiable and non-identifiable data for what purpose. Legislative mandates and regulatory requirements need to be considered in the light of ethical data use, and "use case" specific privacy and confidentiality, and continuity of care considerations.
Unique identifiers	An essential pre-requisite to ensure data collected can be linked to care recipients as well as to organizational and individual providers.
Cybersecurity	Minimizing risk of cyberattacks by protecting systems, servers, networks, and mobile devices. Adopt and maintain programs that educate the workforce, and manage and monitor unauthorized data access.
Vendor/technology-neutral federated data storage	The separation of systems and storage delivering scalable cost-effective data access and flexible systems for all users across the health-care network. Separating data from applications as used by the openEHR community were found to support persistent and transient data as well as real-time local and remote data access.
Electricity and broadband (Internet access) for everyone	A fundamental pre-requisite for all living in this digital era, irrespective of time, and location.

and biobanks. The patients' own contribution to their EHR, and the increasing use of mobile devices and sensors, are also important. They can add valuable insights about environmental and behavioral factors as well (*e.g.*, food, air quality, exercise, and mood).

Both data- and terminology-level standards are reasonably mature, although there is considerable overlap among certain terminology and ontology systems such as SNOMED CT and LOINC. Using fit-for-purpose data and ontology/terminology standards together can tackle most of the difficulties arising from the breadth, depth, complexity, variability, changeability, and longevity aspects of health data.

While openEHR specifications have been purposely engineered to cover all EHR data domains, including those that are intended to be exchanged by various systems, messaging standards such as HL7 v2 and HL7 FHIR have been designed to cover only data to be exchanged. These messaging standards were designed for the sake of simplicity for implementation by developers, but many of whom do not have full understanding about healthcare. Therefore, adoption of the openEHR standard is key for designing and building next-generation EHR/EMRs systems and other applications that deal with clinical data.

The emerging trend of using of HL7 FHIR beyond its purpose to represent the full breadth of clinical data in an EHR is not scalable and costly in terms of time required to develop and maintain FHIR profiles. It is far more cost-effective and safe to invest in the establishment of next-generation neutral EHR systems with vendor-neutral data repositories using openEHR and limit the use of FHIR to support simpler use cases for data exchange.

4.6. Governance and leadership

Our collective work over the last 20 plus years has highlighted the need for high-level governance leadership to maximize collaboration between all relevant stakeholders. Our collective findings over time, supported by this rapid review, have enabled us to identify the required building

blocks to make up any national foundation for successful digital health adoption. These are described in Table 1. Without such high-level government-focused ecosystem-wide collective initiatives and leadership, digital health transformation will continue to be compromised due to continuing fragmentation. The greater use of a compatible set of technology and data standards worldwide is translatable to the greater benefits and opportunities for advanced data analytics and reliable AI applications.

Every jurisdiction needs to determine how best to govern, manage, and provide strong leadership for each of these entities to facilitate the optimization of EHRs enabling AI and to meet desired health outcome objectives. Many health systems already have some of these building blocks in various forms.

5. Conclusion

This review of the current state of data availability, and data quality suitable for AI development and use, has revealed that we have a long way to go to achieve our aim of optimizing EHRs to serve as a data source for AI use. It became clear that most jurisdictions, mega-EMR vendors, and many newcomers are all tinkering at the edges by building on and working with current legacy systems and infrastructures. It is encouraging to see that some jurisdictions have bucked the trend of continuing to make incremental improvements by embarking on major digital health transformation strategies to build and implement next-generation systems and infrastructures. We have identified the need to transform high-level jurisdictional infrastructures. These infrastructural building blocks need to be designed to govern and provide strong leadership enabling ecosystem-wide compliance with mandated key sets of standards. Such standards need to enable flexibility at every point of care to ensure that data/information needs are able to be met in a timely manner for every stakeholder, and all users.

This review has identified the capacity of available next-generation technologies that need to be adopted to optimize EHR content enabling its use for AI purposes. We have the knowledge and skills required to make the best possible use of available innovative technologies to improve both the operational efficiency and effectiveness of every national health system. Many of us working in the digital health field continue to be frustrated by the lack of sufficient knowledge of the complexities associated with digital health by senior decision-makers driving investments, procurement, policy, and legislative solutions. The digital health knowledge domain is huge, both in depth and breadth. The only way we can move forward is through extensive international and multidisciplinary collaboration as made possible by the adoption of well-governed open-access standards, such as openEHR.

Multidisciplinary and multi-sector collaboration requires a change in mindset for many. The benefits of new and emerging technologies used to design and implement next-generation EHRs are huge. The use of generative AI in clinical settings can only be reliably achieved when EHRs are optimized throughout the entire global digital health ecosystem.

Acknowledgments

We thank Heather Grain, Co-Director, eHealth Education Pty Ltd for undertaking a final review and supporting this work.

Funding

None.

Conflict of interest

The authors declare that they have no competing interests.

Author contributions

Conceptualization. Evelyn Hovenga
Writing – original draft Evelyn Hovenga
Writing – review & editing: All authors

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data

Not applicable.

References

1. De Mello BH, Rigo SJ, Da Costa CA, *et al*. Semantic interoperability in health records standards: A systematic literature review. *Health Technol* (*Berl*). 2022;12(2):255-272.
doi: 10.1007/s12553-022-00639-w

2. Ingram D. *Health Care in the Information Society*. United Kingdom: OpenBook Publishers; 2023.
doi: 10.11647/OBP.0335

3. Hovenga E, Hullin C. Global collaborative leadership challenges and economic drivers. In: Hovenga E, Grain H, editors. *Roadmap to Successful Digital Health Ecosystems*. Ch. 3. United States: Academic Press; 2022. p. 35-63.
doi: 10.1016/B978-0-12-823413-6.00001-X

4. Heeks R. Health information systems: Failure, success and improvisation. *Int J Med Inform*. 2006;75(2):125-137.

   doi: 10.1016/j.ijmedinf.2005.07.024

5. Justinia T. The UK's National programme for IT: Why was it dismantled? *Health Serv Manage Res*. 2017;30(1):2-9.

   doi: 10.1177/0951484816662492

6. Sadoughi F, Kimiafar K, Ahmadi M, Shakeri MT. Determining of factors influencing the success and failure of hospital information system and their evaluation methods: A systematic review. *Iran Red Crescent Med J*. 2013;15(12):e11716.

   doi: 10.5812/ircmj.11716

7. Southon G, Sauer C, Dampney K. Lessons from a failed information systems initiative: Issues for complex organisations. *Int J Med Inform*. 1999;55(1):33-46.

   doi: 10.1016/S1386-5056(99)00018-0

8. Johnstone R. *Global Report Reveals Senior Officials Lack Understanding to Drive Digital Transformation of Government*. Singapore: Global Government Forum; 2022 Available from: https://www.globalgovernmentforum.com/global-report-reveals-senior-officials-lack-understanding-to-drive-digital-transformation-of-government [Last accessed on 2024 Mar 09].

9. Wallace M, Sharfstein JM. The patchwork U.S. Public health system. *N Engl J Med*. 2022;386(1):1-4.

   doi: 10.1056/NEJMp2104881

10. Steinkamp J, Kantrowitz JJ, Airan-Javia S. Prevalence and sources of duplicate information in the electronic medical record. *JAMA Netw Open*. 2022;5(9):e2233348.

    doi: 10.1001/jamanetworkopen.2022.33348

11. Wang EC, Wright A. Characterizing outpatient problem list completeness and duplications in the electronic health record. *J Am Med Inform Assoc*. 2020;27(8):1190-1197.

    doi: 10.1093/jamia/ocaa125

12. Madden JM, Lakoma MD, Rusinak D, Lu CY, Soumerai SB. Missing clinical and behavioral health data in a large electronic health record (EHR) system. *J Am Med Inform Assoc*. 2016;23(6):1143-1149.

    doi: 10.1093/jamia/ocw021

13. Weiskopf NG, Hripcsak G, Swaminathan S, Weng C. Defining and measuring completeness of electronic health records for secondary use. *J Biomed Inform*. 2013;46(5):830-836.

    doi: 10.1016/j.jbi.2013.06.010

14. Downing NL, Bates DW, Longhurst CA. Physician burnout in the electronic health record era: Are we ignoring the real cause? *Ann Intern Med*. 2018;169(1):50-51.

    doi: 10.7326/M18-0139

15. Zhang J, Symons J, Agapow P, *et al*. Best practices in the real-world data life cycle. *PLoS Digit Health*. 2022;1(1):e0000003.

    doi: 10.1371/journal.pdig.0000003

16. Zhang J, Morley J, Gallifant J, *et al*. Mapping and evaluating national data flows: Transparency, privacy, and guiding infrastructural transformation. *Lancet Digit Health*. 2023;5(10):e737-e748.

    doi: 10.1016/S2589-7500(23)00157-7

17. HL7. *Health Level Seven International*. Available from: https://www.hl7.org [Last accessed on 2019 Feb 06].

18. SNOMED-International. *Leading Healthcare Terminology, Worldwide*. Available from: https://www.snomed.org [Last accessed on 2024 Mar 10].

19. DICOM. *Digital Imaging and Communications in Medicine*. Available from: https://www.dicomstandard.org [Last accessed on 2019 May 23].

20. Beale T, Hovenga E. The knowledge-driven platform: Strategic technologies for a platform ecosystem approach. In: Hovenga E, Grain H, editors. *Roadmap to Successful Digital Health Ecosystems*. Ch. 6. United States: Academic Press; 2022. p. 115-141.

    doi: 10.1016/B978-0-12-823413-6.00006-9

21. Hovenga E, Grain H, Beale T. Fragmented global standards development organisations. In: Hovenga E, Grain H, editors. *Roadmap to Successful Digital Health Ecosystems*. Ch. 4. United States: Academic Press; 2022. p. 65-96.

    doi: 10.1016/B978-0-12-823413-6.00025-2

22. CDISC. *Clinical Data Interchange Standards Consortium*. Available from: https://www.cdisc.org [Last accessed on 2019 May 23].

23. JIC. *Joint Initiative Council for Global Health Informatics Standardisation*. Available from: https://www.jointinitiativecouncil.org [Last accessed on 2024 Mar 10].

24. ISO/TS_21564. *2019 Health Informatics-Terminology Resource Map Quality Measures (MapQual)*. Switzerland: Intenational Organisation for Standardisation; 2019. Available from: https://www.iso.org/standard/71088.html [Last accessed on 2019 Dec 06].

25. OpenEHR. *Archetype Technology Overview*. London: OpenEHR Foundation. Available from: https://specifications.openehr.org/releases/am/latest/overview.html [Last accessed on 2021 Jun 11].

26. HL7-International. *HL7 Version 3 Clinical Document Architecture (CDA) Release 2*. Available from: http://www.hl7.org/implement/standards/product_brief.cfm?product_id=7 [Last accessed on 2018 Sep 26].

27. HL7-International. *Fast Healthcare Interoperability Resources Specification (FHIR)*. Available from: https://www.hl7.org/fhir/overview.html [Last accessed on 2024 Feb 28].

28. Atalag Koray, Yang Hong Yul, Tempero Ewan, Warren Jim. Model Driven Development of Clinical Information Sytems

using openEHR. In: *Studies in Health Technology and Informatics*. IOS Press; 2011.

doi: 10.3233/978-1-60750-806-9-849

29. Weeks J, Pardee R. Learning to share health care data: A brief timeline of influential common data models and distributed health data networks in U.S. health care research. *eGEMS* (*Wash DC*). 2019;7(1):4.

doi: 10.5334/egems.279

30. OpenEHR/CKM. *Clinical Knowledge Manager*. Available from: https://ckm.openehr.org/ckm/# [Last accessed on 2019 Feb 21].

31. OpenEHR-Foundation. *OpenEHR Reference Model (RM)-Latest*. London: OpenEHR Foundation; 2017. Avaialble from: https://openehr.org [Last accessed on 2024 Jul 23].

32. Cresswell KM, Mozaffar H, Lee L, Williams R, Sheikh A. Safety risks associated with the lack of integration and interfacing of hospital health information technologies: A qualitative study of hospital electronic prescribing systems in England. *BMJ Qual Saf*. 2017;26(7):530-541.

doi: 10.1136/bmjqs-2015-004925

33. Guo M, Wang Y, Yang Q, *et al*. Normal workflow and key strategies for data cleaning toward real-world data: Viewpoint. *Interact J Med Res*. 2023;12:e44310.

doi: 10.2196/44310

34. Stöger K, Schneeberger D, Kieseberg P, Holzinger A. Legal aspects of data cleansing in medical AI. *Comput Law Secur Rev*. 2021;42:105587.

doi: 10.1016/j.clsr.2021.105587

35. Scott PJ, Dunscombe R, Evans D, Mukherjee M, Wyatt JC. Learning health systems need to bridge the 'two cultures' of clinical informatics and data science. *BMJ Health Care Inform*. 2018;25(2):126-131.

doi: 10.14236/jhi.v25i2.1062

36. Cresswell K, Domínguez Hernández A, Williams R, Sheikh A. Key challenges and opportunities for cloud technology in health care: Semistructured interview study. *JMIR Hum Factors*. 2022;9(1):e31246.

doi: 10.2196/31246

37. HL7. *Argonaut Project 2019*. Available from: http://docs.smarthealthit.org/argonaut [Last accessed on 2019 Feb 06].

38. CSIRO-eHealth-Research-Centre. *SPARKED-AU FHIR Accelerator Brisbane*; 2024. Available from: https://confluence.csiro.au/display/fhir/sparked+-+au+fhir+accelerator [Last accessed on 2024 Feb 17].

39. ISO-13606-2. *Health Informatics--Electronic Health Record Communication--Part 2: Archetype Interchange Specification: International Organsiation for Standardisation*; 2008. Available from: https://www.iso.org/standard/50119.html [Last accessed on 2019 May 25].

40. LOINC. *Logical Observation Identifiers Names and Codes from Regenstrief*. Indiana: Regenstrief. Available from: https://loinc.org [Last accessed on 2019 Feb 25].

41. NSW-Government DN. AI Guidelines: Using Artifical Intelligence in the NSW Public Sector; 2023. Available from: https://www.education.gov.au/schooling/announcements/australian-framework-generative-artificial-intelligence-ai-schools [Last accessed on 2024 Feb 16].

42. NSW-Government. *AI Assurance Framework;* 2022. Available from: https://www.digital.nsw.gov.au/policy/artificial-intelligence/nsw-artificial-intelligence-assurance-framework [Last accessed on 2024 Feb 16].

43. AMA. *Position Statement: Artifical Intelligence in Healthcare*; 2023. Available from: https://www.ama.com.au/sites/default/files/2023-08/artificial%20intelligence%20in%20healthcare%20-%20ama.pdf [Last accessed on 2024 Feb 16].

44. World Health Organisation. *Regulatory Considerations on Artificial Intelligence for health*. Geneva: World Health Organization; 2023. Available from: https://iris.who.int/handle/10665/373421 [Last accessed on 2024 Feb 16].

45. Australian-Government DoI, Science and Resources. *Australia's AI Ethics Principles*. Available from: https://www.industry.gov.au/publications/australias-artificial-intelligence-ethics-framework/australias-ai-ethics-principles [Last accessed on 2024 Feb 16].

46. Australian-Government SDA. *Guidelines for Secure AI System Development*; 2023. Available from: https://www.cyber.gov.au/about-us/view-all-content/advice-and-guidance/guidelines-secure-ai-system-development [Last accessed on 2024 Feb 17].

47. Australian-Government-DoH. *National Health Reform Agreement (NHRA) 2020-25*. Available from: https://www.health.gov.au/initiatives-and-programs/2020-25-national-health-reform-agreement-nhra [Last accessed on 2024 Feb 17].

48. Data-Governance-Institute. *The DGI-Data Governance Framework*. Available from: https://datagovernance.com [Last accessed on 2024 Feb 16].

49. The-White-House. *Delivering on the Promise of AI to Improve Health Outcomes*; 2023. Available from: https://www.whitehouse.gov/briefing-room/blog/2023/12/14/delivering-on-the-promise-of-ai-to-improve-health-outcomes [Last accessed on 2024 Feb 28].

50. Government-of-the-Netherlands. *Dutch Government Presents Vision on Generative AI*; 2024. Available from: https://www.government.nl/latest/news/2024/01/18/dutch-government-dutch-government-presents-vision-on-generative-ai#:~:text=the%20dutch%20government%20has%20announced,via%20the%20dutch%20ai%20coalition [Last accessed on 2024 Feb 28].

51. Elkin PL, Smith B. History of Health Terminology. In: *Health Informatics*. Springer International Publishing; 2023:1-7.

doi: 10.1007/978-3-031-11039-9_1

52. ISO17115. *Health Informatics-Representation of Categorial Structures of Terminology* (*CatStructure*); 2020. Available from: https://www.iso.org/standard/73695.html [Last accessed on 2023 Dec 05].

53. ISO18104. *Health Informatics-Categorial Structures for Representation of Nursing Practice in Terminological Systems*. Switzerland: International Organisation of Standards; 2023. Available from: https://www.iso.org/standard/81132.html [Last accessed on 2023 Dec 05].

54. Hovenga EJS. Integrating a Categorial Structure for Clinical Practice into EHRs. In: *Studies in Health Technology and Informatics*. IOS Press; 2024

    doi: 10.3233/SHTI230930

55. El-Sappagh S, Franda F, Ali F, Kwak KS. SNOMED CT standard ontology based on the ontology for general medical science. *BMC Med Inform Decis Mak*. 2018;18(1):76.

    doi: 10.1186/s12911-018-0651-5

56. Rodrigues Jean M, Kumar A, Bousquet C, Trombert B. Using the CEN/ISO Standard for Categorial Structure to Harmonise the. In: *Studies in Health Technology and Informatics*. IOS Press; 2009:255-259.

    doi: 10.3233/978-1-60750-044-5-255

57. ONC. *Common Data Model Harmonisation*. Available from: https://www.healthit.gov/topic/scientific-initiatives/pcor/common-data-model-harmonization-cdm [Last accessed on 2024 Feb 21].

58. Aldoseri A, Al-Khalifa KN, Hamouda AM. Re-thinking data strategy and integration for artificial intelligence: Concepts, opportunities, and challenges. *Appl Sci*. 2023;13(12):7082.

    doi: 10.3390/app13127082

59. McCormick KA. The digital needs of genomics resulting from pandemics. In: Hovenga E, Grain H, editors. *Roadmap to Successful Digital Health Ecosystems*. Ch. 15. United States: Academic Press; 2022. p. 363-374.

    doi: 10.1016/B978-0-12-823413-6.00018-5

60. Hassan M, Awan FM, Naz A, *et al*. Innovations in genomics and big data analytics for personalized medicine and health care: A review. *Int J Mol Sci*. 2022;23(9):4645.

    doi: 10.3390/ijms23094645

61. Viceconti M, Hunter P, Hose R. Big data, big knowledge: Big data for personalized healthcare. *IEEE J Biomed Health Inform*. 2015;19(4):1209-1215.

    doi: 10.1109/JBHI.2015.2406883

62. OpenEHR. *An Open Domain-Driven Platform for Developing Flexible e-Health systems*. London: OpenEHR Foundation. Available from: https://www.openehr.org [Last accessed on 2019 2019 Feb 06].

63. Blobel B, Kalra D, Koehn M, *et al.* The Role of Ontologies for Sustainable, Semantically Interoperable and Trustworthy EHR Solutions. In: *Studies in Health Technology and Informatics*. IOS Press; 2009:953-957.

    doi: 10.3233/978-1-60750-044-5-953

64. Rector AL, Rogers J, Taweel A. Models and Inference Methods for Clinical Systems: A Principled Approach. In: *Studies in Health Technology and Informatics*. IOS Press; 2004:79-83.

    doi: 10.3233/978-1-60750-949-3-79

65. Rector A, Schulz S, Rodrigues JM, Chute CG, Solbrig H. On beyond gruber: "Ontologies" in today's biomedical information systems and the limits of OWL. *J Biomed Inform*. 2019;100:100002.

    doi: 10.1016/j.yjbinx.2019.100002

66. Rector AL, Johnson PD, Tu S, Wroe C, Rogers J. Interface of Inference Models with Concept and Medical Record Models. In: *Lecture Notes in Computer Science*. Springer Berlin Heidelberg; 2001:314-323.

    doi: 10.1007/3-540-48229-6_43

67. Rector Alan L. The Interface between Information, Terminology, and Inference Models. In: *Studies in Health Technology and Informatics.* IOS Press; 2001:246-250.

    doi: 10.3233/978-1-60750-928-8-246

68. Rector A, Marley T, Qamar R. *Models of use and Model of Meaning Slideplayer*; 2017. Available from: https://slideplayer.com/slide/11210854 [Last accessed on 2024 Feb 29].

69. ISO-23903. *Health Informatics-Interoperability and Integration Reference Architecture-Model and Framework*. Geneva: International Organization for Standards (ISO); 2021. Available from: https://www.iso.org/standard/77337.html [Last accessed on 2024 Jun 11].

70. NHS-England. *Digital Transformation*. Available from: https://www.england.nhs.uk/digitaltechnology [Last accessed on 2024 Feb 28].

71. Pedrera-Jiménez M, García-Barrio N, Frid S, *et al*. Can OpenEHR, ISO 13606, and HL7 FHIR work together? An agnostic approach for the selection and application of electronic health record standards to the next-generation health data spaces. *J Med Internet Res*. 2023;25:e48702.

    doi: 10.2196/48702

72. OECD. *Towards an Integrated Health Information System in the Netherlands*. Paris: OECD Publishing; 2022.

    doi: 10.1787/a1568975-en

73. Eit-Health E. *Implementing the European Health Data Space in Sweden*; 2023. Available from: https://eithealth.eu/wp-content/uploads/2023/10/implementing-the-european-health-data-space-in-sweden.pdf [Last accessed on 2024 Feb 28].

74. Government-of-Jamaica. *Jamaica Poursues Digital Transformation of Health Sector: Ministry of Health and*

Welness; 2024 Available from: https://www.moh.gov.jm/jamaica-pursues-digital-transformation-of-health-sector [Last accessed on 2024 Feb 21].

75. Jones M. Caribbean/PAHO-Jamaican case study. In: Hovenga E, Grain H, editors. *Roadmap to Successful Digital Health Ecosystems*. Ch. 4. United States: Academic Press; 2022:523-535.

doi: 10.1016/B978-0-12-823413-6.00026-4

76. eHAction. *Common Semantic Strategy for Health in the European Union*; 2019 Available from: http://ehaction.eu/eu-common-semantic-strategy-open-consultation [Last accessed on 2024 Feb 28].

77. CODE-Center-for-Open-Data-Enterprise. *Sharing and Utilizing Health Data for AI Applications-Round Table Report*; 2019. Available from: https://www.hhs.gov/sites/default/files/sharing-and-utilizing-health-data-for-ai-applications.pdf [Last accessed on 2024 Feb 22].

78. ISO13606. *The ISO 13606 Standard Explained*; 2019. Available from: http://www.en13606.org/information.html [Last accessed on 2024 Feb 18].

79. OpenEHR. *Specifications*. Available from: https://specifications.openehr.org [Last accessed on 2021 Apr 25].

80. Atalag K, Beale T, Chen R, Gornik T, Heard S, McNicoll I. *OpenEHR-A Semantically-Enabled, Vendor-Independent Health Computing Platform-White Paper*; 2017. Avaialble from: https://www.openehr.org/resources/white_paper_docs/openEHR_vendor_independent_platform.pdf [Last accessed on 2017 Feb 17].

81. ISO-13606-1. *Health Informatics--Electronic Health Record Communication--Part 1: Reference Model*. Switzerland: ISO; 2019. Available from https://www.iso.org/standard/67868.html [Last accessed on 2017 Feb 18].

82. Beale T. Archetypes: Constraint-based domain models for future-proof information systems. In: Baclawski K, Kilov H, editors. *Eleventh OOPSLA Workshop on Behavioral Semantics: Serving the Customer*. Boston: Northeastern University; 2002.

83. Ma C, Frankel H, Beale T, Heard S. EHR query language (EQL)--a query language for archetype-based health records. *Stud Health Technol Inform*. 2007;129(Pt 1):397-401.

84. OpenEHR. *Archetype Query Language (AQL)*. Available from: https://specifications.openehr.org/releases/query/latest/aql.html [Last accessed on 2019 Feb 12].

85. Chen R, Valladares C, Corbal I, Anani N, Koch S. Early Experiences from a Guideline-based Computerized Clinical Decision Support for Stroke Prevention in Atrial Fibrillation. In: *Studies in Health Technology and Informatics*. IOS Press; 2013:244-247.

doi: 10.3233/978-1-61499-289-9-244

86. Garde S. Clinical Knowledge Governance: The International Perspective. In: *Studies in Health Technology and Informatics*. IOS Press; 2013:269-281.

doi: 10.3233/978-1-61499-291-2-269

87. Atalağ K, Bilgen S. Multi-level Modeling and the Role of Archetypes in the Design of Health Information Systems: A Modeling Example in Endoscopy. In: *Conference proceedings of the 2007 International Symposium on Health Informatics and Bioinformatics HIBIT07, Atalya, Turkey*. 2014.

doi: 10.13140/2.1.4762.2727

88. Viceconti M, De Vos M, Mellone S, Geris L. Position paper from the digital twins in healthcare to the virtual human twin: A moon-shot project for digital health research. *IEEE J Biomed Health Inform*. 2023;28:491-501.

doi: 10.1109/JBHI.2023.3323688

89. Nickerson D, Atalag K, De Bono B, *et al*. The human physiome: How standards, software and innovative service infrastructures are providing the building blocks to make it achievable. *Interface Focus*. 2016;6(2):20150103.

doi: 10.1098/rsfs.2015.0103

90. Teno JM. Garbage in, Garbage out-words of caution on big data and machine learning in medical practice. *JAMA Health Forum*. 2023;4(2):e230397.

doi: 10.1001/jamahealthforum.2023.0397

91. Institute of Medicine Committee on Quality of Health Care in America. *Crossing the Quality Chasm: A New Health System for the 21st Century*. Washington, DC: National Academies Press.

92. ISO/TR14639. *Briefing Note: Strengthening National Health Systems through a Capacity-Based eHealth Architecture*. Switzerland: International Organization of Standards ISO; 2014. Available from: https://www.iso.org/files/live/sites/isoorg/files/archive/pdf/en/14639-brochureversionv7.pdf [Last accessed on 2020 Jul 18].

93. PATH. *Data Use partnerships: Theory of Change*; 2016. Available from: https://www.path.org/our-impact/resources/data-use-partnership-theory-of-change [Last accessed on 2021 Jul 22].

94. Castro A, Machado J, Roggendorf M, Soller H. *How to build a data architecture to Drive Innovation-Today and Tomorrow*. *McKinsey Technology*; 2021. Available from: https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/how-to-build-a-data-architecture-to-drive-innovation-today-and-tomorrow [Last accessed on 2021 Apr 12].