

Electronic Health Records: Benefits and Challenges for Data Quality

Abdul Kader Saiod, Darelle van Greunen and Alida Veldsman

Abstract Data quality (DQ) issues in Electronic Health Records (EHRs) are a noticeable trend to improve the introduction of an adaptive framework for interoperability and standards to large-scale health Database Management Systems (DBMS). In addition, EHR technology provides portfolio management systems that allow Health Care Organisations (HCOs) to deliver higher quality of care to their patients than possible with paper-based records. The EHRs are in high demand for HCOs to run their daily services as increasing numbers of huge datasets occur every day. An efficient EHRs system reduces data redundancy as well as system application failures and increases the possibility to draw all necessary reports. Improving DQ to achieve benefits through EHRs is neither low-cost nor easy. However, different HCOs have several standards and different major systems, which have emerged as critical issues and practical challenges. One of the main challenges in EHRs is the inherent difficulty to coherently manage incompatible and sometimes inconsistent data structures from diverse heterogeneous sources. As a result, the interventions to overcome these barriers and challenges, including the provision of EHRs as it pertains to DQ will combine features to search, extract, filter, clean and integrate data to ensure that users can coherently create new consistent data sets.

Keywords Electronic Health Records (EHRs) • Data Quality (DQ) • Health Care Organisations (HCOs) • Database Management Systems (DBMS) • Kilometres (km) • Information and Communication Technologies (ICT) • Health Information Technology (HIT) • Quality Assessment Tools (QAT)

A.K. Saiod (✉) • D. van Greunen • A. Veldsman
Nelson Mandela Metropolitan University (NMMU), Port Elizabeth, South Africa
e-mail: saiodkader@gmail.com
URL: <http://www.nmmu.ac.za/DarelleVanGreunen>

D. van Greunen
e-mail: Darelle.vanGreunen@nmmu.ac.za

A. Veldsman
e-mail: Alida.Veldsman@nmmu.ac.za

1 Introduction

Electronic Health Records (EHRs) refer to implemented structured digital manifestations of real-time, patient-centred health records [1]. EHRs are considered as one of health care's innovation heuristic items and are widely adopted over HCOs and are becoming an important mechanism to perform their daily services [32, 40]. The secure EHR systems provide information available instantly and accurately to the authorized users, so user can coherently create new consistence of data sets [28, 61]. In general, the EHRs is the problem of combining health data that reside at different sources and providing accurate, comprehensive up-to-date patient history [26, 58]. Improvements in the data quality have brought about efficiency, scalability and safety in the implementation of a large scale healthcare DBMS [30]. Health data can therefore be composed, managed by authorised user and consulted by authorised providers from across multiple HCOs nation or global wide and can be shared across them. The EHRs include an enormous range of patient data set, including patient details, history, references, medication, immunisation, allergies, radiology report including images, laboratory data and test reports, admission and discharge details, personal statistics like Body Mass Index (BMI), blood pressure, sugar level etc. These datasets are electronically stored in database as narrative (free text) or encrypted data. The EHR databases are structured to store accurately and securely health information over the time. It can reduce data replication risk as all access points are retrieving data from the main data server and as well as reduces lots of paper work. The principle of data replication is to share information between multiple resources. The replication reduces fault tolerance, increase high accessibility and reliability. Many distributed database systems are using replication to avoid single access point failure and high traffic. It can be possible to dynamically improve load-spreading and load-balancing performance by providing replication [63]. Replication supports restoring replicated databases to the same server and database from which the backup was created [4]. Backup is one of the important processes of database server routine maintenance plans that, to copying and archiving data to an external device. So, backup data can be used to restore the original information after any data loss event. Now a day, electronic data is searchable even from heterogeneous sources and possible to combine them into a single data set. EHRs are even more effective when analysing long term patient medical history [22]. Due to EHRs data being tractable and easy to identify patient preventive visits or screening information, monitor the overall progress, effectively than the paper-based record in HCOs. EHRs improve patient care, increase patient participation, improve care coordination, improve diagnostics and patient outcomes, practice efficiencies for cost savings and allow more case studies for research purposes. Despite the many advantages and functionalities of EHR systems, there are still a considerable number of disadvantages associated with this technology [7]. One of the key concerns is the quality of the data, which includes inconsistency, privacy protection, record synchronisation, lack of standardised terminology, system architecture indexing, deficient standardised terminologies.

The productivity may drop temporally with associated EHRs adaptation as workflows have changed. Several long-standing consequences are emerging from the critical issue of EHRs adaptation [3]. Therefore, it is of utmost importance to advise healthcare organisations to choose the right EHR systems and provide proper setup to establish the complete system to become successful users of EHR systems [42]. Healthcare organizations that are using tangible, augmented EHR systems in their facilities can make better decisions based on the comprehensive information available to them. Improving in healthcare distribution systems are becoming the most consequential technology for medical innovation of all the times. EHR systems exhibit promising potential, which will play the crucial role in HCOs to ensure providing in excellent patient care service, quality management, accurate information, perfect diagnosis, patient information safety, disease management and investigation as advance innovation deftness [56]. In particular, the chapter focuses on large scale DBMS, the data quality introduction of smart interfaces and perfect data mapping in traditional EHR systems as well as mobile and cloud computing. This implies that integrated adoptive EHRs can show inconsistencies, because the data structure and standard from different HCOs are different.

2 Electronic Health Records Background

In our daily lives, data inconsistency may cause with uncertain incidents, if unstructured data composed in the data collection process [35]. An example a web healthcare domain page based on largely composed with free text data. Another example, medical data collections process methods are paper-based and/or archive information.

Information may collect by the data retrieval system and index them by non-text data, so that user can access and find data using special keywords to obtain accurate data sets [41].

Using the non-text data to indexing large text, may lead the data structure design in EHR systems of the other efficient way for accessing and searching information as there are a vast amount of non-text data available [60].

The four latest methods to detect and reduce data inconsistency are:

1. Rough set theory [20];
2. Logic analysis of inconsistent data method [13];
3. Corresponding relational variables of functional dependencies [8];
4. Fuzzy multi-attribute theory [2].

Fuzzy multi-attribute method has ideal performance of the inconsistency data and it can obtain the highest average level of correct information than other solutions [19]. A method for reducing data inconsistency has to be combined with a method for data integration to coherently solve the data inconsistency and the data integration problems simultaneously. The domain ontology may effectively

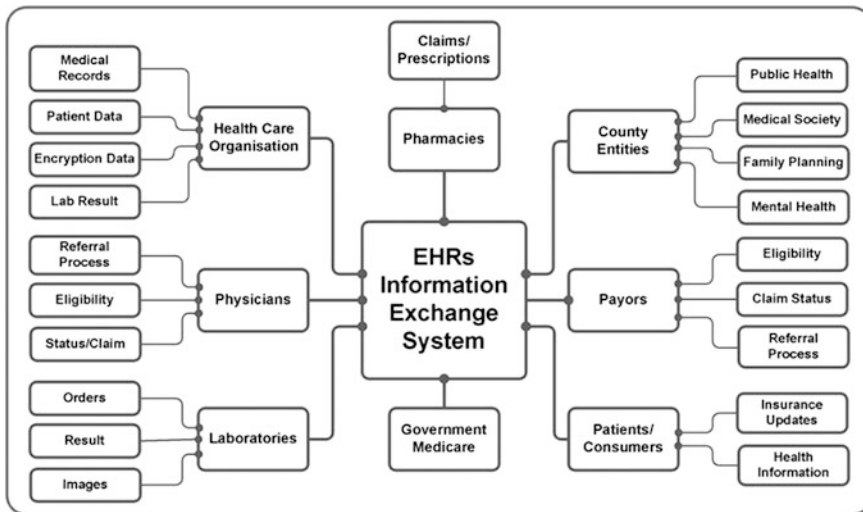


Fig. 1 EHRs information exchange systems architecture

combine data from diverse heterogeneous sources for data integration. The existing ontology data integration methods are, however, not sufficient to implement fuzzy-ontology [24]. The EHRs information exchange systems architecture is shown in Fig. 1.

The benefits of EHRs are numerous when compared to the physician's time and finances, the health benefits for patients and the impact on the environment. The sparse health data may have multi dimensions and it is practically challenging to investigate and analyse for a different reasons such as the heterogeneous features of the system, encompassing quantitative data as well as the categorical information. By the result the random systematic error effect badly and reduces the data quality. Most data integration methods are sufficiently robust to random systematic error for large data sets of input and process. This is commonly identical to bring them on a same scale when using pre-processing principal component analysis and data simplification algorithm [47]. The EHRs systems framework is shown in Fig. 2.

Data quality issues might include patient incorrect unique identification number. Other examples, misplace name, incorrect gender, incorrect date of birth, numeric diagnosis code written in text or saved wrong radiology image, incorrect inserting standard code, such as the National Drug Catalog (NDC) for drugs and derailing bulk analysis (e.g. ICD10 code: International Classification of Diseases Tenth Revision or CPT code: Current Procedural Terminology). Data quality refers to the concepts with immensely large-scale multi dimensional in DBMS, which include not only data search, validation, extract and verification, but also the appropriateness of use to take us even further beyond the traditional concerns with the accuracy of data. The EHR systems design, data structure, aggregation algorithm, simplification methodology and reporting mechanisms highly reflect on data quality.

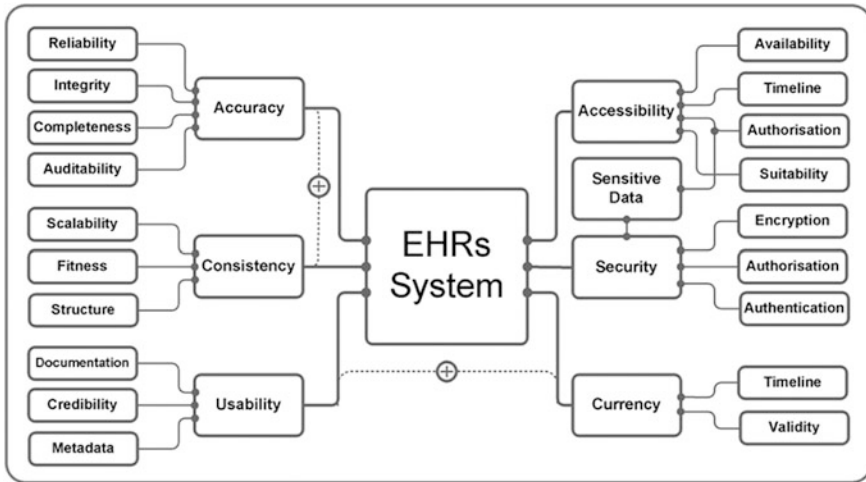


Fig. 2 EHRs systems framework

3 Different Network Architecture and Cloud Computing

In co-operating distributed health information systems and networks, the EHR Systems provided as a lifelong patient record advance towards core applications [6]. Several researchers have shown that only the arbitrary access to patient health information are proximal motive of accurate decision in health care during decision-making and the effective communication between patient care team members [14]. The number of hospitals and clinics are increasing every day, as well as increasing the health information. Health information has digitalised and archived their health record with the universal use of computer and information technology network. There are vast types of wired and wireless network layout, consisting of the type of device including hardware, software, connectivity protocols and communication mode of transmission. It also includes knowledge about the types of networks grouped according to types such as LAN, MAN and WAN. Such as cloud computing, that refers as fast computation and its capability to store large storage space. Now a day cloud computing is a convenient, on-demand network. It is also configurable computing resources to a share group network such as, application, service, server and archive. With the minimal managerial effort cloud computing can be rapidly provided and released the higher productivity. The EHR system can be integrated into cloud computing. Basically, smaller hospital and clinic has limited resources. Cloud computing is to facilitate for those smaller HCO with adequate electronic medical record storage space to provide the exchange and sharing of electronic medical records [12]. Cloud computing has high impact on parallel distributed grid computing systems. The flexibility for the further development of these techniques is recommendable. It is very effective location

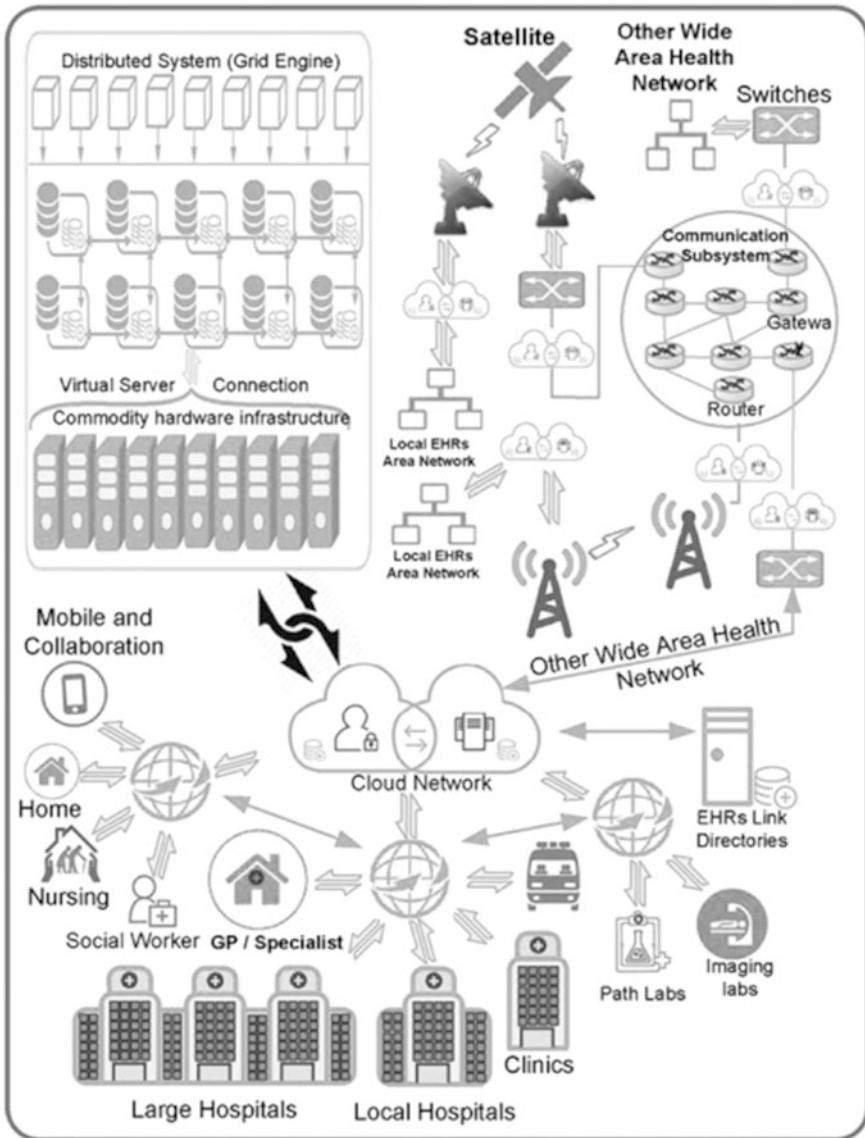


Fig. 3 Large-scale cross platform EHRs system architecture overview

independent technology as well as to enhance the user experiences over the Internet. Now a day, it can provide services for various application scenarios. More and more applications are migrated onto the cloud platform [25].

Mobile Pervasive Healthcare (MPH) service is another innovative technology in EHR system that can provide a wide range of location independent service.

Providing great benefits to both patients and medical personnel of MPH service includes monitoring, telemedicine, location independent medical services, emergency management and response, and pervasive access to healthcare information [39]. There are three specific goals in mobile healthcare systems:

1. The availability location independent of EHRs applications;
2. The location independent health information;
3. The anytime and the invisibility of computing.

There are two categories of EHRs systems: (i) Cloud based technology; (ii) Client server based technology;

Computer with internet connection in order to access via the web, the online data can be stored externally. The cloud computing application allows users on-demand access and provide by third party organisation using the internet. Large-scale cross-platform EHRs system architecture overview is shown in Fig. 3.

There are six benefits of cloud-based systems namely: security, privacy, cost effectiveness, accessibility, reduced IT requirements and it grows with you [29]. The significant risks of each HCO will face when transitioning to cloud-based hosting. The main disadvantage of cloud computing is all data, security, availability, maintenance and control sits to a third party so HCOs have absolutely no control of it. Trusting to the third party service provider is one of the important factors for cloud computing and it takes on a whole different meaning [51]. Despite all the barriers, it's important to remember that cloud health computing paradigms are still under development, but with a lot of chances of being a revolution in a lot of fields. In the near future there will be more services on offer and the development will be greater.

4 The Barriers and Threats of Electronic Health Records

The overarching barriers to EHRs framework are to tackle the indigent quality of data to provide a single, centralised and homogeneous interface for users to efficiently integrate data from diverse heterogeneous sources. Data quality issues may arise when capturing raw data into EHR systems. The data flow process has several factors that influence the quality of information obtained from such datasets at a later stage. The purpose of the data collection processes are data quality management functions include the data flow process application as well as data accumulate, warehousing process systems used to archive data and analyse the process of translating data into meaningful information. The data quality may seriously affect of patient care and even could lead to the death of the patient. This is the key challenges of eradicating treatment errors in the health service process. As patient safety is the key issue in health care service, using effective EHR systems integration and implementation can improve the data quality to reduce medical error. The main consideration for health data includes data accuracy and accessibility, as

well as data comprehensiveness, currency, consistency, granularity, precision, relevancy definition, and timeliness. Data quality will empower the tendency of EHRs systems, this emphasises the magnificence of implementing a design-oriented definition. The dimensions of the existing EHRs framework are basically based on historical reviews, understanding intuitive and comparative experiment.

The EHRs structures of an orientation usually vary from framework to framework. For example, the actual use of the data depends on the definition of data quality. It therefore data quality also depends on the application type and what may be deliberated in one application as good quality, but may not good for another. Data quality has emerged as a crucial issue in many application domains. The objective of data quality becomes even more important in the case of patients who need to be identified and notified about important changes in drug therapy or in the case of merging systems of different and similar organisations. The consolidation of information from diverse sources to provide a unified view of an organisation's data assets is technically challenging. The difficulty involves how to practically combine data from disparate, incompatible, inconsistent and typically heterogeneous sources. The other difficult objective in EHRs systems is that data has a structure, which is usually complex and cannot be treated as a simple string of bytes. Often data inconsistency occurs because the data structures may depend on other structures, therefore on a distributed system such data management is very difficult. Another significant aspect of a health data integration system is data mapping. The system must able to materialise data that are mapped from diverse source. Optimally using routinely collected data increases poor quality data, which automatic mechanism would raise the need of the semantic interoperability as well as quality data measurement [34]. Quality improvement and error reduction are two of the justifications for healthcare information technologies. Despite their concerns, HCOs are generally very interested in adopting and implementing EHR systems. A major concern of success of implementing is the large gap between planning for the introduction of EHR systems and medical maintenance system in hospitals. The primary purpose of the successful EHR systems implementation depends on these application systems and maintains the application significantly to achieve the benefit desired and expected. The real barriers causing this gap may not be the availability of technology to the HCOs, as information systems are actually becoming available almost everywhere, but the deficiency in providing proper support before, during and after implementation of the EHR system. The financial constrains are another important matter of the migration from the paper-based health record to an EHR system. Generally there are two principle barriers and challenges in the method of prosperous EHR system integration, namely:

1. Human barriers (e.g. professional and beliefs);
2. Financial barriers (e.g. available money or funding opportunities).

The human factors become even more important as the benefits are only anticipated after the successful integration and implementation of the EHRs systems. The information security is the most important for the quality health care

service. It improves the potential of EHRs as well as accuracy, accessibility, productivity, efficiency and to reducing the costs of healthcare and medical errors. Most HCO administrators are aware that it is time consuming to migrate from paper base record system to an EHR system. It is also important to change the provider behaviors and health care practitioner with regard to electronic healthcare systems, but time is also needed. Few things also need to be addressed regarding to successful implementation of an EHR system, such as attitudes, impressions and beliefs. The most important factor is essential to understand the reasons for and the purpose of the implementation of EHR systems in the whole subject [46]. Research and statistics showed EHRs estimated potential savings as well as the costs of the widespread adoption of EHRs systems. Important health and safety benefits were modelled and concluded that effective EHR system implementation and networking, could improve healthcare efficiency and safety. It also showed that Health Information Technology (HIT) could enhance the prevention and management of chronic diseases, which could eventually double the savings, while increasing health and other social benefits. The feasibility of introducing an EHR system to improve data quality is the meaningful association between the heterogeneous data source and the integration into HCOs to improve healthcare service. The integrity constraints are specified in the global scheme of data mapping, which can be used to promote EHRs data quality as well. The uncertainties are the other important integration aspect in EHRs that should be minimised to improve data quality. The most important barriers and constraints to high quality datasets in order to promote must solved the integration of EHR systems and electronic health record, to achieve maximum benefit of the healthcare services. Finally, it is noted that the query answer in the context of data exchange, contributes to data quality.

5 Literature Review

Existing literature shows that several techniques and major EHRs systems currently exist to deal with data quality issues, which historically have faced DBMS. After a profoundly analysis of various cutting-edge commercial accomplishment existing on the software market and an intensive review of literature, it appears there are still some limitations to practical tools for EHRs systems. Physically access of diverse information sources of robust support is provided, but only if these are standard database structure tables. At the moment, there are no automatic mechanisms to solve existing integration problem [43]. Peer-to-Peer (P2P) topology is used when system-individual participants contact a localised server to search other data and to contact other participants directly, to exchange information or share resources. However, Gribble et al. [23] stated that generic P2P systems often do not take care of the semantics of the data exchanged. This is a serious drawback, especially considering that when the network grows, it becomes hard to predict the location and the quality of the data provided by the system. Tania et al. [57] propose a mediated query service, which is a system that is used for configuring mediation

systems for building and maintaining multi-dimensional multimedia data warehouses. Considerable disadvantages are, however, involved in moving data from multiple, often highly disparate data sources, into a single data warehouse. This translates into long implementation time, high cost, lack of flexibility, outdated information and limited capabilities.

A subsequent representation by Gilson et al. [21] for data integration was proposed using middleware architecture. The middleware can encompass dynamic scheduling, performance management and transport services for distributing scientific visualisation tasks in a grid environment. Middleware, however, has a high development cost, the implementation thereof is time and resource consuming, few satisfying standards exist, its tools are not good enough and often threatens real-time performance of a system and middleware products are not very mature. Load-balancing issues, limited scalability, low levels of fault tolerance and limited programmer access area, for example, are some of the main disadvantages of middleware.

Combining Aggregation Operators (AO) and fuzzy Description Logics (DL), Vojtáš [59] presents a fuzzy DL with general AOs. The expressiveness of the logic is, however, very limited. An additional evaluation of this strategy was also done for data and multimedia sources using an ontology-based data integration system. A Mediator Environment for Multiple Information Sources (MOMIS) data integration system was proposed using a single ontology approach to overcome this limitation. The system combines the MOMIS framework with the STASIS framework. MAFRA [38] is an ontology mapping framework for distributed ontology, which supports an interactive, incremental and dynamic ontology mapping process in the semantic web context. Using ontology integration, the conflicts in the result can be solved by satisfying the consistence criterion. An approximation technique has been identified as a potential way to reduce the complexity of reasoning over ontology's in expressive languages such as OWL 1 DL and OWL 2 DL [49]. A vast amount of research concerning EHR mechanisms has been carried out over the last few years. Fuzzy-ontology is moving forward to express fuzzy properties, membership functions and linguistic hedges [11]. The fuzzy-ontology definitions that are found in the literature are quite naturally influenced by fuzzy set theory, fuzzy logic and existing ontology languages. Shaker et al. [55] performed an exercise using fuzzy-ontology integration to solve the problem of equivalently matching concepts to avoid pairs of mismatching concepts and conflicts regarding multiple entities to reduce data inconsistency. Another related work was performed by Sanchez et al. [52], which considers fuzzy-ontology with general quantifiers that could be used for some type of quantifier-guided aggregation.

Cristiane et al. [15] used a DISFOQuE system to analyse the fuzzy-ontology to perform semantic query expansions. This is an ontology-based data integration system for data and multimedia sources, which is essentially performed manually by the integration designer. A few studies handle fuzziness and give support for uncertainty in their conceptual models for multimedia materials. The studies by Aygün et al. [5] and Özgür [45] try to handle this uncertainty by supporting fuzzy attributes. Different types of databases exist, but the type most commonly used in

healthcare is the Online Transaction Processing (OTP) database. For the most part, healthcare databases are used as the foundation for running the many transactional system databases, which structures accommodate the creation of a wide range of transactional applications such as EHRs lab systems, financial systems, patient satisfaction systems, patient identification, data tracking, administration, billing and payment processing and research. The EHRs database servers are to replace the old paper-base documents, files, folders and filing cabinets. Data is therefore now more convenient and current. It's obvious that the benefits of EHRs are equal to the benefits of the applications that run on them. Significant advances in automation and standardisation of business and clinical processes can be attributed to these applications and databases. With EHR databases, data can also be stored externally and backed up in a secure place to prevent data loss. Because front-end software can provide tip text and enforce data integrity, the back-end data can therefore become more standardised and accurate. Lastly, because the data is electronic, it allows for quicker processing of typical transactions such as lab results and payment claims. One of the biggest benefits of all these databases is the amount of data healthcare organisations have been able to capture. They now have huge data stores that can be used to inform better and more cost-effective care. EHRs focus on strategies for combining data residing at different heterogeneous sources and providing the users with a unified view of the data. A vast amount of work has been developed in the EHRs area and some interesting results have shown the effectiveness of this approach. It has, however, not been extensively evaluated with regard to ease of data access to dynamic her systems and their widely implementation over HCOs. The aim is to avoid the theoretic pitfalls of monolithic ontologies, facilitate interoperability between different and independent ontologies and provide flexible EHRs capabilities. In addition, not all the existing EHRs integration techniques are sufficient, as many healthcare organizations are still capturing their data in spreadsheets and often mismatch information and formats, which cause incorrect report generation and reduce the quality of the data. There is thus a need to develop or use an efficient EHRs system, using a template screen, which is efficiently mapped to the online transaction-processing database. *An important objective of EHRs systems is the way it needs to be adapted to address the data quality issue to achieve all possible benefits and address, all problems described above.*

6 Electronic Health Records Data Structure

Most HCOs data are highly structured and heavily depend on claims data, but absence the prosperous scope provided by health data. Furthermore, leverage health data basically depend on vendor-delivered implementation, communication, such as a Continuity of Care Documents (CCDs) is the few analytics applications. They also stick limitations via both design and integration that make them insufficient for populating health and productivity analytics, until CCDs offer a consolidated and

expedient way to implement electronic health data. Methods for data capturing in EHR systems include direct capturing, capturing on the screen template, scanning hand written documents or importing transcribed data from other information systems in different data exchange formats, such as JSON, XML, CSV, TXT, REBOL, Gellish, RDF, Atom, YAML and other data exchange technologies. Each one of these methods has strengths and weaknesses that may have an impact on data quality. Only the quality of the data can ensure that healthcare providers have confidence in EHRs systems to deliver the best service possible. To capture, store as well as develop and implement structured health data to avoid data quality gap must use the integrated analysis program. In order to effectively solve the challenge of data quality gaps, there is a need to discuss further relevant points. Valid data capturing techniques require that when a clinical encounter takes place and provider and/or automated systems insert information into an EHR system, it is for example, captured accurately into the EHR of a patient. Valid data structures need to be used both in the way in which data is captured as well as storage of the data in an appropriate format and location. If an integer is captured in a VARCHAR field, its feasibility for reporting, analysis and quality will be reduced, even if it is captured in a structured field. If the template or screen structure is not properly mapped or configured in the database, the value may still be stored in an incongruous location. The analysis or reporting purposes for this information is extracted from the server and made instant and available to authorised user. There is no needing include extracting all pertinent information when it is back end database connection. How data will extract or how data will select from the query table, are the key factors in how the exchange of create the data set in how the exchange tackle impacts on the application quality of the outgoing data quality. It is of importance to identify the point at which data quality gaps are introduced. This will in turn lead to focused initiatives to eliminate such gaps.

Data security is a key concern in healthcare interoperability whether paper-based or electronic health records. According to the human rights, every individual can keep personal data confidential and not being disclosed for surveillance or interference from another organisation or even to the government. All confidential information should be protected and encrypted while data that is shared as a result of the clinical relationship [50]. Patient data can only be released when the patient gave his/her consent or when stipulated by law. Information may disclose information sharing only if the patient is unable to do so because of age or mental incapacity, data sharing decision should be made by the legal representative or legal guardian of the patient. The information is considered confidential and must be protected when a result the query result of clinical cooperation. The identity of the patient cannot be ascertained when information is populated, for instance, the number of patients with HIV in a government hospital, does not fall in this denomination [50].

7 Electronic Health Records Data Synchronisation

The gradual harmonisation of the data overtime, so called data synchronisation is the procedure of establishing consistency between information from a diverse source to the destination data server and vice versa. Considering large-scale computing, dataflow between multi-user clinicians and one central server, definitely entails a multi-way synchronisation model. Since, the server must send a client the data previously created by other clients. The patient is not conscious of the entire data structure as it is the server data structure algorithm to figure out the modifications so called he rights management rules. In many cases, data should be available in more than one directory server using three different techniques for achieving this. This includes a directory replication protocol, direct synchronisation between a pair of directory servers and indirect synchronisation between two or more servers. Replication has the best operational characteristics but the lowest functionality, which can differ between the various techniques. The majority of replication techniques entail indirect synchronisation, which has the highest functionality and the poorest operational characteristics, while direct synchronisation is intermediate. There are numerous reasons why data in one directory server needs to be made available in another directory server. These include availability, load sharing, locality, reaching data on other servers, data access restrictions as well as data mapping. Figure 4 depicts the two-way data synchronisation workflow model:

Generally synchronisation between a client and a server follows five steps:

1. The data administrator rules prepares the data for a “go/no-go” response when the authorised user initialises the request;

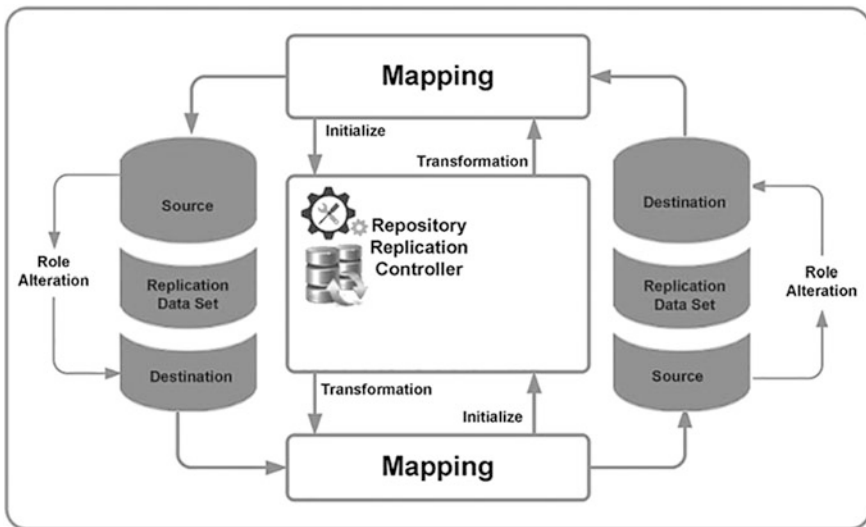


Fig. 4 Two way data synchronisation workflow

2. The server algorithm rules checks the user authentication to accomplish whether synchronisation is required and finally checks for all possible conflicts;
3. The authorised user submit the data trees;
4. Before to nodes and stores data the server assigns new IDs to trees;
5. The server uniquely identify the data in the network to these collective IDs and the collective database.

It should be noted that only the authorised user allows viewing the sent to the client according to the accurate management rules. Finally, before replacing the local IDs with collective ones and storing the new trees to the server, the authorised user updates it's to local database [17]. The overall practice shows that EHRs and the ability to exchange health information electronically can help HCOs to provide higher quality and safer care for patients, while creating tangible enhancements for healthcare. EHR systems thus enable healthcare providers to not only improve the care management plan for their patients, but to also provide improved healthcare through accurate, up-to-date and complete information sets about patients. This enables quick access to patient records for an improved and coordinated care plan. This is achieved by securely sharing electronic information with patients and other clinicians that in turn helps providers to diagnose more effectively and thus reduce the medical errors. It contributes to the provision of safer care, improved patient and provider interaction and communication. Add to this healthcare convenience, more reliable prescribing, promotion of legible and complete documentation supported by accurate, streamlined coding and billing. Other improvements include enhancing privacy and security of patient data, helping providers improve productivity and work-life balance, enabling providers to improve efficiency and meet their business goals, reducing costs through decreased paperwork, improved safety, reduced duplication of testing and improved healthcare services. Figure 5 depicts a typical model of the data synchronisation architecture of a healthcare risk manager's organisation:

It therefore became necessary, to implement adaptive, interoperable EHR systems to improve the quality of data, which addresses the current EHRs challenges. As a result, the proposed solution will focus on a novel approach based on different methods and existing systems, to reduce the challenges of EHRs and data quality. EHRs technology will be applied to not only perform the function of receiving and displaying information, but to automatically and accurately extract information from diverse heterogeneous data sources that use makes use of healthcare services. For data to equivalently match two concepts across different data sources and automatically resolve any inconsistency arising from multiple data entities is the challenge of EHRs. The important expected contribution of this study will be to realise a method to improve on EHRs data quality from heterogeneous and inconsistent data sources. The key outcome of EHRs will be to discover a new merged concept by finding consensus among conflicting data entries.

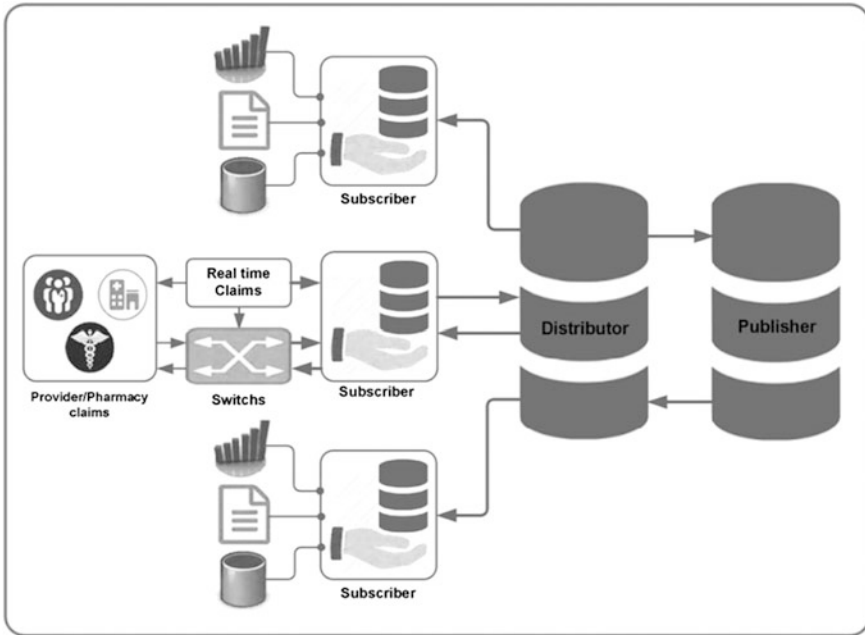


Fig. 5 A healthcare risk manager's organisation data synchronisation architecture

8 Electronic Health Records Data Collection

Data collection is defined as the on-going, systematic assembling and measuring of information, analysis and illustration of health data necessary for integration, implementing, designing, and evaluating public health prevention programmes, which then enables one to answer relevant questions and evaluate outcomes [62]. The HCOs collect data to observe health to handle subsidies and services as well as inform bankroll and resource allocation, identify and appraise healthcare services, inform the development of health policies and interventions, assist clinical decisions about patient care and meet legislative requirements. Surveillance is undertaken to inform disease prevention and control measures, identify health emergencies as an early warning system, guide health policies and strategies and measure the impact of specified health interventions. Few people are, for example, dying from infectious diseases, but for instance, due to changing patterns of physical activity and expenditure of drug, tobacco, alcohol and food more people are suffering from chronic diseases. The survey process is conducted to maximise accuracy and participation to generate statistics. Using a different data collection algorithm, these statistics are generated from diverse sources, including household surveys, routine reporting by health services, public registration and censuses and disease observation systems. HCOs involve a different civil data set and private data collection systems, including clinical surveys, administrative enrolments, billing

records and medical records used by various entities, including hospitals, physicians and healthcare plans. The possibility of each to facilitate data on patients or enrollees data on race, ethnicity and language are also collected to some extent by all of these entities suggesting [16]. Data breaches in healthcare come in a variety of forms such as different healthcare capturing and storing methods as well as technology used such as excel, access, SQL and Oracle. Manual data collection from ward-based sources captured only 376 (69%) of the 542 in-patient episodes, captured by the hospital's administrative electronic patient management programme. Administrative data from the electronic patient management programme had the highest levels of agreement with in-patient medical record reviews for both length of stay (93.4%) data and discharge destination (91%) data [53]. Currently, fragmentation of dataflow occurs because of silos of data collection. In HCOs data are often collected by clinical assistants, clinical nurses, clinicians and practice staff.

Compare the completeness of data capturing and level of agreement between three data collection methods in prospective observational studies:

1. Manual data collection from ward-based sources or paper based: Paper & Pencil, Surveys, Chart abstraction and Weekly return card;
2. Administrative data from an electronic patient management program: Dedicated electronic data collection systems, EHRs-based, Images and Audio and video recording (qualitative research);
3. Inpatient medical record review for hospital length of stay and discharge destination.

With specific diseases in clinical and genomic research, the objective of EHRs is to generate large cohorts of patients. The electronic phenotype selection algorithms are to find such cohorts a rate-limiting step is the development.

This study evaluated the portability of a published phenotype algorithm to identify Rheumatoid Arthritis (RA) patients from electronic health records at three different institutions, using three different EHR systems. EHR systems are seen by many as an ideal mechanism for measuring the quality of healthcare and monitoring ongoing provider performance. It is anticipated that the availability of EHRs-extracted data will allow quality assessment without the expensive and time-consuming process of medical record abstraction. A review of the data requirements for the indicators in the Quality Assessment Tools (QAT) system, suggests that only a third of the indicators would be readily accessible from EHRs data. Other factors such as the complexity of the required data elements, provider documentation habits and the EHRs variability make the task of quality assurance more difficult than expected. Accurately identifying eligible cases for quality assessment and validly scoring, those cases with EHRs extracted data will pose significant challenges, but could potentially lower costs and therefore expand the use of quality assessment. Improving the data collection process across the healthcare system is one of the key challenges to improve data quality.

9 Improving Data Collection Processes

Overreaching opportunities abound for increased quantities of EHRs, for improved quality of data and for new data elements that were once considered too burdensome or expensive to capture. This wealth of EHRs can be used to validate or calibrate health demand models, for inpatient care information systems analysis and for modelling mobile source emissions across a healthcare network. These data collection and processing advancements are however, costly and should be implemented with caution. The focus of healthcare is on data accuracy issues pertaining to the mechanism chosen for data collection and data processing, using EHRs technology. Vast amounts of technology spreads exist throughout the transportation field, which are automating numerous manual data collection processes. These advances generally reduce labour costs and manual capturing errors. Automation of survey data collection allows EHR systems to collect new data streams without increasing respondent burdens. When data are combined from diverse heterogeneous sources, the data that are syntactically identical (same format, same units) can show important inconsistencies, as data elements that supposedly represent the same concept, actually represent different concepts at each site. The term semantic variability expresses the data variability caused by differences in the meaning of data elements. Differences in data collection, abstraction and extraction methods, or measurement protocols can result in semantic variability. Figure 6 depicts the dataflow control system in large-scale DBMS.

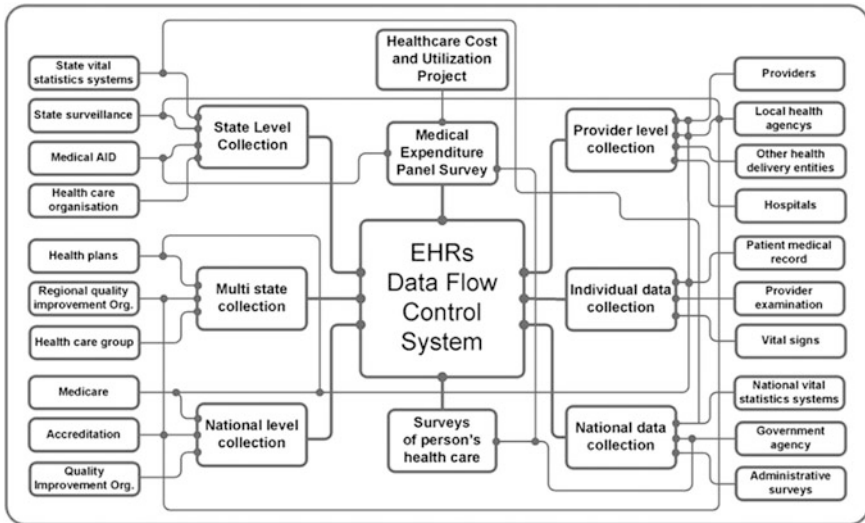


Fig. 6 The data flow control system in large scale DBMS

Failure to distinguish between fasting and random blood glucose, finger-stick or venipuncture sampling or serum or plasma measurements would, for example, result in glucose values that do not represent the same concept. Semantic variability is difficult to detect using single-site data alone, because data semantics tend to be consistent within an institution. Only when the data are combined from multiple heterogeneous sources can such semantic differences be detected. The above discussion regarding the challenges faced by various healthcare professionals and healthcare institutions, highlights the importance of accurate data capturing and data quality in order to overcome HIT constraints and minimise respondent and organisational resistance. Integration of data systems has the potential to streamline collection processes, so that data can be reported on easily and that an individual would not need to self-identify race, ethnicity, and language requirements during every health encounter. Integrating the various data systems, enhancing legacy HIT systems, implementing staff training, and educating patients and communities about the reasons for, and importance of collecting these data, can help improve data collection processes. Not all data systems capture the method through which the data were collected, and some systems do not allow for data overrides. The interoperability of data systems may, for example, prohibit a provider from updating a patient's data, which were provided by the patient's healthcare plan. Self-reported data should therefore trump indirect, estimated data or data from an unknown source. Ways of facilitating this process logistically warrant further investigation. Data overriding should be used with caution, as overriding high-quality data with poor-quality data reduces the value for analytic processes.

Currently, one specific data collection effort under evaluation for automation is the patient update survey, which traditionally has been administered to obtain a comprehensive up-to-date patient history. The fundamental concern associated with the need to change medical practice tendencies and the way of interacting with patients, created barriers to EHRs implementation and use. The adaptation to EHR systems were also considered a major threat to practitioner professionalism, because of the corresponding requirements for providers to adhere to the requirements of the EHRs, including electronic documentation and compliance with standardisation guidelines. Even though current data collection methods are subject to numerous errors, the survey data collected are used to forecast regional health data such as demographics, hospital admissions and discharge notes, medical history of patients, improvement notes, outpatient clinical health notes, medication prescription records, medication and allergies, immunisation statuses, radiology reports and images, laboratory data and test reports, essential symptom, personal statistics like BMI, blood pressure, age and weight information etc. In addition, the availability of EHRs databases makes the automated processing of such data feasible. With the application of these technologies, however, care and caution should be applied when using and interpreting the datasets obtained from the data collection method used.

10 Data Quality Issues in Electronic Health Records

Quality data, appropriate for use, comprise of characteristics that include completeness, uniqueness, consistency, accuracy, validity, correctness and accurate timelines. The quality of data can be analysed from multiple dimensions. One such dimension is a measurable data quality property that represents some aspect of the data accuracy and consistency that can be used to guide the process of understanding quality [44]. Though EHRs data quality is often only considered within the narrow scope of data verification and validation, it should also concern equally critical aspects of assuring that EHRs data is appropriate for a specific use. Alternatively, the quality of data is comprehended as in high demand, according from this denomination, as the volume of data increases and the question of internal consensus within data become significant, regardless of its appropriateness for use for any particular external purpose. Even when discussing the similar set of data used for the same intention, confluence's prospect on data quality can often be in uniqueness. Some information quality problems may arise from when the raw data is collected until it becomes useful information. The majority of EHRs data is captured by a large number of individuals from heterogeneous sources and data exchange accessed these days to index text object use data rescue systems devised. Due to unit measurement without different definitions and may be captured in the EHR system. It will absolutely impossible or may not comparatively and assessment to interpret what is being reported by other clinician when validated psychometric scales to assess patient status are not used. The objective deficiency that these problems, classify idiosyncratic data quality features.

The data inconsistencies can be identified directly, which can lead to inaccuracies and bias, as the data is collected geographically and over time and might be adjusted differences over to account for unequal measures over time [9]. The main concern of EHRs is the feasibility of introducing EHR systems in HCOs to improve the data quality in order to achieve all possible benefits pertaining to healthcare services. To effectively address this concern, the following issues need to be taken in account:

1. What are the most meaningful associations among heterogeneous health data sources that can be explored to improve EHRs data quality?
2. What kinds of integrity constraints are specified in the global scheme of data mapping that can be explored to improve EHRs data quality?
3. What are the uncertainties in data integration that when minimised, results in an improved EHRs data quality?

Schaal et al. [54] motivate the adoption of accessible data based on its definition of data that comprises of clarity and consistency. The intimidation posed during data storage and transmission, EHRs are seen as a hopeful accomplishment to problems in EHR managements and despite them. One of the key barriers is to optimally use routinely collected data, as the increasingly poor quality remains in the data. This raises the need for automating the mechanisms used to measure data

quality and semantic interoperability. This framework is a result of filtering the existing data quality dimensions in many research sources and checking its suitability to the nature of e-health systems. In many research sources and verifying its praiseworthiness to the behaviour of e-health systems, this skeleton is a outcome of percolation the subsist data quality dimensions.

11 Methodology

The main contribution of EHR systems consists of mathematical modelling of a heuristic methodology based on a combination of the perfect matching and similarity measurement for distributed concepts unanimity techniques for inconsistency and conflict resolution performance to improve data quality. Creating a consensus between perfect matching and similarity measurement can be resolved unite of data inconsistency, mismatches and conflict ontology entity regarding diverse data sources. Data quality, consistency, reliability, validity, accuracy, completeness and timeliness are the eventual significant list of EHR systems. Matching is a process of finding alignment between sets of correspondences with a semantic verification output of the matching process. This merging is a process of creating a new set of possibly overlapping data. However, the aim of the main task is to determine best illustrate object to find semantically fundamental equivalent motive in EHR systems. This provides a strong theoretical and practical framework to work with heterogeneous, complex, conflicting and automatic consensus methods for EHRs. A concept detection method could be semantically more meaningful, if multimedia data integration process is a subset of the annotation process for trustable automatic data mapping, such as image, sound or video indexing by special code. This means that each and every EHRs amalgamated data associate to a distinguishable adumbration characteristic. So, conflict may happen on EHRs data integration, if a diverse amalgamated data associate to the same apprehension in the diverse EHR system. Discovering the interrelation in EHR systems is a significant phenomenon among entities manifested in diverse EHRs values. The similarity measurement often discovered that those conflicting entities are approximately identical among EHRs entities. Specially, with a chronic health circumstance, the EHRs statistic predicts individual development and effectuation, since EHRs adoption better meets the needs of the growing modern community. Currently, there are five principles that contribute to data quality. These principles are listed as:

1. Formal Concept Analysis;
2. Conceptual Clustering;
3. Generation;
4. The Grid-File for Multi-attribute Search;
5. Semantic Representation Conversion.

This will be done to ease data access, extract information, search mechanisms, synchronises and establish semantic connections, filter data and provide different levels of security, provide data inconsistency solutions, resolve equivalently matching or conflicting information in multiple entities, resolve queries and achieve data compression and automatic EHRs integration simultaneously. Conflicts based on the occurrence of the same names or the same structures for different concepts, were solved by using the concept of Potentially Common Parts (PCP) propagation. Other aforementioned conflicts such as associated value conflicts and conflicts on a concept level, were also resolved using consensus methods. Specific criteria can be attributed to the representation. The criteria comprise of:

Comparability consists and specify designate of the analytical data quality, in observing the variation in performance critical measures to the Triad as the Triad emphasises collaborative data sets. Various analytical methods show that there are three different types of divers can be identified in collaborative data sets. These variations are individual-level variations (for example age, sex, co-morbidities), provider level variations and random/residual variations. It is obvious that challenges in data availability and comparability issues are numerous national and international comparisons.

Completeness is defined as the extent to which all data elements are integrated. One of the most important aspect for integration result that the completeness is the guarantee of appearance of all ingredients when integrating. Each data element should be captured in an EHR system, so that a provider could create a data set for a patient's characteristics. Either, it will not possible for a provider to create an accurate diagnosis for a patient characteristic. Incomplete data can't provide accurate diagnosis even when the corresponding information supplied by EHR systems. The data will remain incomplete, until providers are not completed with the approximate group value associating to semantic categories. Finally, the inconsistencies of the patient diagnosis with appropriate values, with the inconsistency of any value in these appropriate values are the intention acceptability of total designation for the EHRs completeness.

Consistency is defined as the absence of any inconsistencies in the EHRs result and been solved all appearing conflicts among elements when integrated. When stream data appended dependent interpretative variables from diverse heterogeneous sources, often integrate data refers to the similar subject, but apprehend different inconsistency information. The time (T) dimensional observation data could be large and are asymptotically valid for certain time. The data could be repeat approximately same value over the time such a situation is called a conflict.

Identification is defined as structure similarity among entity source and the EHRs result. The EHRs domain apprehensions often contain entities that have interrelated among the property value. This defines that each EHRs entity accompanied to certain concepts. If identical characteristic associating to the identical apprehension in various ontologies, the conflict in the EHRs are also associated with different associated EHRs value, it also manifested as conflict.

Timeliness is defined as associate between the registration and diagnosis entity and the determination time to the observation diagnosis of the occurrence statistical

report. The EHRs statistics reports improve provider observation of patient outcome. Overall statistic indicated that the use of EHR systems could sustain improves productivity of health care service, such as timeliness statistic report or invoices etc.

12 The Similarity Measurement

The importance of EHRs data integration is to find the correlation among entities manifested in diverse EHRs. The similarity measure technology may discover those conflicting entities are approximately identical among EHRs entities. With the tremendous growth in the adoption of EHRs, heterogeneous sources of patient health information are becoming available. It is practically challenging to discover significant similarity entity and how to measure and leverage providers inputs. It is a very important aspect to identify accurate subsidiary uses of EHRs data to achieve the goal [31]. Figure 7 shows the workflow and architecture of the similarity detection service as follows:

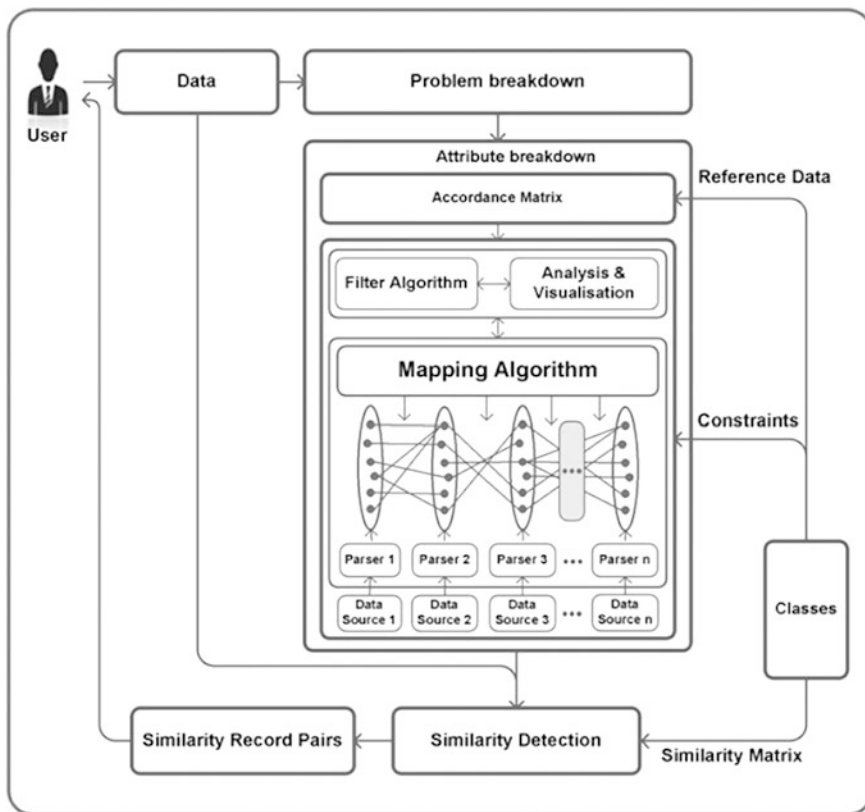


Fig. 7 Workflow and architecture of the similarity detection service

The objective of similarity becomes even more important, when measuring the similarity among equivalent patient entity based on their EHRs data. There are three effective similarity measurement types are appropriate in many applications, such as:

1. Case-based retrieval of similar;
2. Treat similarity between the batch similarity;
3. Cohort comparison and comparative effectiveness.

According to the aforementioned EHRs integration, the similarity measurement technology assorted in four different groups:

Instance-Based Similarity: The similarity between concepts is determined by common instances, as well as comparing new problem instances with instances and stores in memory, instead of performing explicit generalisation. Instance-based ontology mapping is a promising solution to a class of ontology alignment problems. The similarities between concepts define as ordinary instance and matching new entity issues and store them, but not executing the exact generation. The common entity is the key value of similarity among two concepts. The promising solution is instance-based ontology mapping for a class of ontology classification. Measuring among similarity and annotated entity sets crucially depends on it. A set of abstractions evolved from significant entities do not maintain by Instance-based algorithms. If it has the large storage capability, then this approach reaches the nearest neighbour algorithm. The classification accuracy significantly reduced the large storage requirements, but its performance degrades rapidly.

The instance-based similarity equation is:

$$\text{Similarity}(x, y) = - \sqrt{\sum_{a=1, n} (x_a - y_a)^2} \quad (1)$$

where, x and y are instances in an n -dimensional instance space [18].

1. **Lexical-Based Similarity:** The similarity between two concepts is based on the analysis of the linguistic interpretation of associated names. An example, let's find a most similar words for word W , to combine an estimate. Weight the evidence provided by word W' by a function of its similarity to W . Combining information among similar words is a similarity measuring function between words. Which word pairs require a similarity-based measurement is determined by a scheme method. If word W'_1 is "alike" to word W_1 , then W'_1 can participating entity about the probability of invisible word pairs involving W_1 [27]. The Lexical-Based Similarity methods for language modelling of combining evidence evaluated as:

$$P_{sim}(W_2|W_1) = \sum_{w'_1 \in S(w_1)} \frac{W(w_1, w'_1)}{N(w_1)} P(w_2|w'_1) \quad (2)$$

$$N(w_1) = \sum_{w'_1 \in S(w_1)} W(w_1, w'_1) \tag{3}$$

$S(W_1)$ —The set of words most similar to W_1 ;

$W(W_1, W_1')$ —Similar function.

2. **Scheme-Based Similarity:** The similarity among amalgamated characteristic is the analysis of similarity among two intentions. There are two types of structure based similarity:

1. The internal structure-based similarity;
2. The external structure-based similarity.

It can be the alteration of $P_{sim}(W_2|W_1)$ in the back of Eq. (2), such as interpolating with the unigram probability $P(w_2)$:

$$P_r(w_2|w_1) = \gamma P(w_2) + (1 - \gamma)P_{SIM}(w_2|w_1) \tag{4}$$

The linear combination illustrate in yielding between the similarity estimate and the back-off estimate:

if $\gamma = 1$, Then possibly to make γ depended on w_1 .

So, the similarity allotment for achievement could vary between words [33].

3. **Taxonomy-Based Similarity:** The structural relationship breakdown is the base of similarity among two concepts in Taxonomy-Based Similarity. It considers the relations as links connecting concepts. If two concepts are already matched, their neighbours (concepts are collected along with the links from the already matched concepts) may also be somehow similar.

Let's consider two gene products W_1 and W_2 , and being represented by collections of terms $W_1 = T_{11}, \dots, T_{1i}, \dots, T_{1n}$ and $W_2 = T_{21}, \dots, T_{2i}, \dots, T_{2n}$.

Based on the two sets, the goal is to define a natural similarity between

The main goal is to determine a natural similarity, based on two sets among W_1 and W_2 , denoted as $S(W_1, W_2)$ [48]. Considerable two principle approaches are:

1. **First approach:** The similarity is computed pair-wise, say $S_{ij}(T_{1i}, T_{2j})$ and then the aggregation is performed using, for example, the average as:

$$S_a(W_1, W_2) = \frac{\sum_{i=1}^n \sum_{j=1}^m S_{ij}}{mn} \tag{5}$$

It is an interesting factor, when the objects T_{1i}, T_{2j} , belong to a given ontology.

Here, the pair wise similarity can be determined as in [37] using shortest paths and information theoretic constructs. The problem rise, only if the average is used with this approach.

Even when the two sets are very similar, $S_a(W_1, W_2)$ may not be 1.

When W_1 and W_2 have only one common entity, then the similarity is 1 and it will ignore the other. Then real trouble is to choose the maximum.

13 The Perfect Matching

Every apex of the nodes is incident and to exactly one edge of the matching is an assignment of nodes, called perfect matching. The concept of perfect matching is $n/2$ edge, it means that perfect matching only possible on nodes with the even vertices number only. Complete matching or 1-factor is the other name of perfect matching. Here, the dynamic Hungarian algorithm is presented, appropriate to optimally solve the assignment task in condition with changing edge costs, time and as well as improving the healthcare service. The combinatorial optimisation algorithm of Hungarian method is to solve the assignment issues in polynomial time and which expected later primal-dual methods. The assignment problem is widely-studied and exists in many application domains, known as the maximum weighted bipartite matching problem [10]. The Hungarian algorithm undertakes the existence of a bipartite graph, $G = (H, P; E)$ that have illustrated in Fig. 8, where E is the set of edges and H and P are the sets of nodes in each baffler of the diagram.

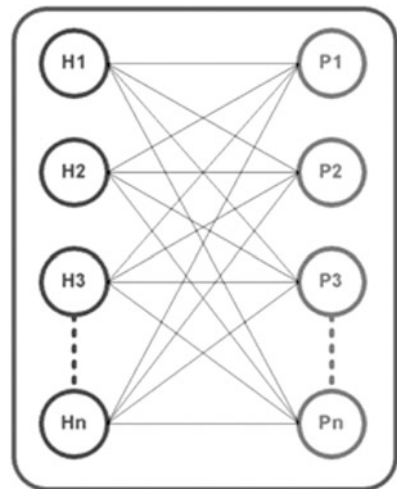
Let us call a function $y: (H \cup P) \rightarrow R$ a potential if, $y(i) + y(j) \leq c(i,j)$ for each $i \in H, j \in P$. The potential value of y is $\sum_{v \in H \cup P} y(v)$.

The time of each perfect matching is the latest value of each potential.

The perfect matching of tight edges discover by the Hungarian method: an edge ij is called tight for a potential y , if $y(i) + y(j) = c(i,j)$. Let us denote the subgraph of the tight edges by G_y . The time of a perfect matching in G_y (if there is one) equals the value of y .

Suppose, there are four hospitals (same group hospital) in a big city to which a model is assigned tasks on a one-to-one basis. The time of assigning a given resource to a given task is also known. Figure 8 shows a bipartite graph of Hungarian algorithm as follows:

Fig. 8 Bipartite graph of the Hungarian algorithm



An optimal assignment needs to be found, which minimises the total time to better service. The central hospital call centre receives emergency phone calls and decides from which hospital to send the ambulance and to which hospital the minimum distance is between the hospitals (H) and the patient call location (P). A decision-making model is designed using the Hungarian algorithm to calculate how the ambulance should respond to the patient for emergency services to minimise the total time, as time is the biggest factor which can save the life of patients. The distance in kilometres (km) between the hospital H (ambulance location) and the patient call location P are given below:

H \ P	H1	H2	H3	H4
P1	110	95	95	100
P2	55	105	75	85
P3	145	115	110	125
P4	165	130	115	135

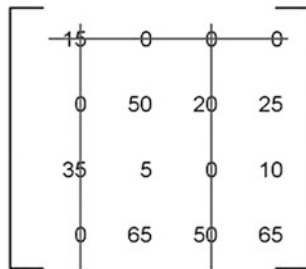
Step 1: Subtract row minima: This step is to determine the lowest element. Then in that row, subtract it from each element. In row 1 subtract 95, in row 2 subtract 55, in row 3 subtract 110, and in row 4 subtract 65 as the lowest element.

$$\begin{bmatrix} 110 & 95 & 95 & 100 \\ 55 & 105 & 75 & 85 \\ 145 & 115 & 110 & 125 \\ 65 & 130 & 115 & 135 \end{bmatrix} \sim \begin{bmatrix} 15 & 0 & 0 & 5 \\ 0 & 50 & 20 & 30 \\ 35 & 5 & 0 & 15 \\ 0 & 65 & 50 & 70 \end{bmatrix}$$

Step 2: Subtract column minima: In this step similarly as step 1 for each column let's determine the lowest element then, subtract it from each element in that same column. In column 1 subtract 0, in column 2 subtract 0, in column 3 subtract 0 and in column 4 subtract 10.

$$\begin{bmatrix} 15 & 0 & 0 & 5 \\ 0 & 50 & 20 & 30 \\ 35 & 5 & 0 & 15 \\ 0 & 65 & 50 & 70 \end{bmatrix} \sim \begin{bmatrix} 15 & 0 & 0 & 0 \\ 0 & 50 & 20 & 25 \\ 35 & 5 & 0 & 10 \\ 0 & 65 & 50 & 65 \end{bmatrix}$$

Step 3: Cover all zeros with a minimum number of lines: This step is to cover all zeros in the resulting matrix. Minimum number of horizontal and vertical lines should be use to cover all zeros. An optimal assignment exists between the zeros, if n lines are required. The algorithm stops.



The step 4 will continue, as it required less than n th lines.

Step 4: Create additional zeros: This step is to determine the smallest element (call it k) in step 3. This smallest element didn't cover by a line. All uncovered elements must subtract by k , here is the smallest element. Then if the element covered twice then add k to all elements. We have to proceed to Step 5, as we have the minimal number of lines is less than 4.

Step 5: This step is to determine the smallest entry (5) that is not covered by any line. So, in each uncovered row subtract 5.

$$\begin{bmatrix} 15 & 0 & 0 & 0 \\ 0 & 50 & 20 & 25 \\ 35 & 5 & 0 & 10 \\ 0 & 65 & 50 & 65 \end{bmatrix} \sim \begin{bmatrix} 15 & 0 & 0 & 0 \\ -5 & 45 & 15 & 20 \\ 30 & 0 & -5 & 5 \\ -5 & 60 & 45 & 60 \end{bmatrix}$$

Now add 5 to each covered column.

$$\begin{bmatrix} 15 & 0 & 0 & 0 \\ -5 & 45 & 15 & 20 \\ 30 & 0 & -5 & 5 \\ -5 & 60 & 45 & 60 \end{bmatrix} \sim \begin{bmatrix} 20 & 0 & 5 & 0 \\ 0 & 45 & 20 & 20 \\ 35 & 0 & 0 & 5 \\ 0 & 60 & 50 & 60 \end{bmatrix}$$

And now let's, return to the **Step 3**.

Step 3: Again cover all the zeros in the resulting matrix. Minimum number of horizontal and vertical lines should be use to cover all zeros.

$$\begin{bmatrix} \text{---} 20 \text{---} 0 \text{---} 5 \text{---} 0 \\ 0 & 45 & 20 & 20 \\ \text{---} 35 \text{---} 0 \text{---} 0 \text{---} 5 \\ 0 & 60 & 50 & 60 \end{bmatrix}$$

Step 4: Since the minimal number of lines is less than 4, return to Step 5.

Step 5: Note that 20 is the smallest entry not covered by a line. Subtract 20 from each uncovered row.

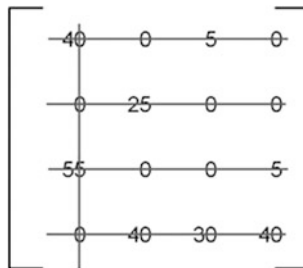
$$\begin{bmatrix} 20 & 0 & 5 & 0 \\ 0 & 45 & 20 & 20 \\ 35 & 0 & 0 & 5 \\ 0 & 60 & 50 & 60 \end{bmatrix} \sim \begin{bmatrix} 20 & 0 & 5 & 0 \\ -20 & 25 & 0 & 0 \\ 35 & 0 & 0 & 5 \\ -20 & 40 & 30 & 40 \end{bmatrix}$$

Then add 20 to each covered column.

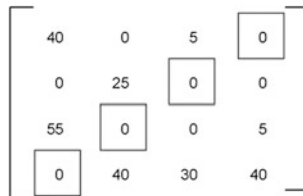
$$\begin{bmatrix} 20 & 0 & 5 & 0 \\ -20 & 25 & 0 & 0 \\ 35 & 0 & 0 & 5 \\ -20 & 40 & 30 & 40 \end{bmatrix} \sim \begin{bmatrix} 40 & 0 & 5 & 0 \\ 0 & 25 & 0 & 0 \\ 55 & 0 & 0 & 5 \\ 0 & 40 & 30 & 40 \end{bmatrix}$$

Now return to *Step 3*.

Step 3: Cover all the zeros in the matrix with the minimum number of horizontal or vertical lines.

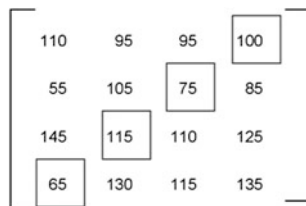


Step 4: Again determine the smallest element of lines is 4. This smallest element didn't cover by a line. The calculation is finished as an optimal assignment of zeros is possible.



We have found the zero as the total cost for this assignment. So, it must be an optimal assignment.

Now, let's return to the original time matrix of the same assignment.



So the hospital should send ambulance H4 to Site P1, ambulance H3 to Site P2, ambulance H2 to Site P3, and ambulance H1 to Site P4.

14 Analysis of Results

Information and Communication Technologies (ICT) in healthcare organisations can be used in a beneficial way to address the key benefits and challenges faced by EHR systems, and policy makers increasingly recognise this potential. ICT enabled solutions support the provision of effective, efficient and good quality services, when implemented on a larger scale DBMS.

Healthcare policy makers and strategists inevitably will have to find some way in which to deliver more and more complex services to meet the increasing demand and expectations for promotion and maintenance of health, treatment and care. It is a significantly essential component to confirm that expected benefits must actualize by healthcare professionals to ensure EHRs adoption. There is associating the specific impact of isolating and organisational factors in designating EHRs adoption called a knowledge base gap. Therefore need to be assessed on the adoption of EHRs in healthcare settings, the unique contributions of isolating and organisational factors, as well as the possible interrelations between these factors.

All experimental measured units such as time, distance and motion are a continuous variable and calculated in standard deviations and units in standard time formats. Time is estimated using a count of the incidences of an activity within a certain time period and reported as proportions. To facilitate comparisons across studies, taking in account the different sampling units, such as ambulance encounter versus ambulance total emergency service time, a relative time difference was calculated. The relative time difference was determined for each, considering the time it took to document using a computer, minus the time it took to document on paper, divided by the time it took to document on paper, producing a negative value if the EHRs was time efficient. 95% Confidence intervals were calculated for differences in means and proportions to assess the significance of reported differences, when there was insufficient information to compute 95% confidence intervals. The weighted averages were calculated for both types of sampling unit ambulance encounter and emergency service time, to accumulate for the changeability across our test studies. The following formulas have used to calculate the weighted averages [36]:

$$WA = \frac{\sum_{i=1}^n [SW(i) * RTD(i)]}{\sum_{i=1}^n SW(i)} \quad (6)$$

In which,

$$(SW) = (n_{group1} + n_{group2}) \quad (7)$$

$$(RTD) = \frac{(documentation\ time_{group2} - documentation\ time_{group1})}{documentation\ time_{group1}} \quad (8)$$

where *WA*—*Weighted Average*, *SW*—*Sampling Weight*, *RTD*—*Relative Time Difference*.

The overall research identified that to achieve the benefits, depends on successful EHRs system implementation and use. Only data quality can provide confidence about the EHRs data to providers so that the benefits of using EHRs, such as best service, data accessibility, quality and safety measurement, improvement and reporting can be seen.

15 Conclusions

The main contribution of this chapter is the improvement of a novel framework for an effective method for electronic health records to achieve its maximum benefits and reduce data quality challenges in healthcare organisations to the minimum. A consensus method was also applied to solve the matching conflicts in EHRs integration. In practice, a dynamic Hungarian algorithm-matching tool was implemented by combining PCP and consensus techniques. The EHRs consist of the following essential steps to achieve the goal: the formal concept analysis the conceptual clustering, the ontology generation, the Grid-File for multi-attribute search and the semantic representation conversion. EHR technology became even more essential for modern healthcare services as increasing the communication network (Internet) and ICT technologies. The aim of EHR systems are not only to improve the healthcare service and wellbeing, it is indispensable demand to design a novel framework for EHRs services to reach beyond independent towards sustainability of our modern society and adaptation. Introducing EHRs systems in healthcare service can, however, offer vast benefits to HCOs and society. The social and ethical acceptance is an important factor for the EHR systems adoption, such as services relies to trust between patients and providers have towards. In this chapter, we discussed the possible benefits and challenges of data quality by introducing efficient EHRs systems in HCOs.

The dynamic Hungarian algorithm shows how a decision-making system for the assignment problems with emergency services saves time and reduces service costs and the results showed that both accuracy and completeness have a large impact on the approach effectiveness. In real-time scenarios, the goal of the method algorithm is to efficiently integrate health data and repair inconsistency data instantly and accurately, when changes in the edge time and costs appear. The overall scenario and challenges discussed above shows data quality in EHRs systems, which demonstrate the method to be effective with regard to accurate performance of the

provider's service. This method presents the result of the principle theoretical characteristics that are considered to tackle any thereafter theoretical and practical problems for both qualitative and quantitative methodologies of implementing EHRs. The EHRs system will not have any limits and the system can be modified efficiently to benefit scalability.

References

1. Abdel, N. H. Z., Mohammed, E. and Seham, A. E. (2015), Electronic Health Records: Applications, Techniques and Challenges; International Journal of Computer Applications (0975 – 8887) Volume 119 – No. 14, June 2015
2. Abdolhadi, N., Mohammad, M. S., Abbas, A. B. (2012). Landfill site selection by decision-making tools based on fuzzy multi-attribute decision-making method. DOI [10.2007/s12665-011-1137-2](https://doi.org/10.2007/s12665-011-1137-2)
3. Andrea, H. and Barbara, P. (2005), Software Quality by Misuse Analysis, Case study in section 6 was removed because it is confidential
4. Andrew, D. B., Scott, M. A., Scott, B. and Brandon, J. R. (2015). Data replication tracing; Publication number: US8935612 B2
5. Aygün, R. S. and Yazici, A. (2004). Modeling and Management of Fuzzy Information in Multimedia Database Applications. Multimedia Tools Appl. vol. 24, pp. 29–56
6. Bernd, B. (2006). Advanced and secure architectural EHR approaches, International Journal of Medical Informatics (2006) 75, 185—190
7. Brent A. Metfessel (2014) Financial Management Strategies for Hospitals and Healthcare Organizations; ISBN-13: 978-1466558731
8. Bernhard, T. (2013). Dependencies in Relational Databases, December 1, ISBN: 9783663120186
9. Bruce, BK., Tom, B., Lucy, S., Andrew, L., Masica, M., Nilay, S., Neil, S. (2013). Fleming. Challenges in Using Electronic Health Record Data. CER. ISSN: 0025-7079/13/5108-0S80
10. Burkard, R. E. and Ela, E. (1999). Handbook of Combinatorial Optimization. pp. 75–149
11. Carlos, B., Roberto, Y., Fernando, B., Sergio, I., Jorge, B., Eduardo, M., Raquel, T., Ángel, LG. (2015). Emerging Semantic-Based Applications; DOI: [10.1007/978-3-319-16658-2_4](https://doi.org/10.1007/978-3-319-16658-2_4)
12. Charalampos, D., Thomas, P., and Ilias M. (2010). Mobile Healthcare Information Management utilizing Cloud Computing and Android OS. 978-1-4244-4124-2/10/\$25.00 ©2010 IEEE
13. Cheng, H. (2014). Analysis of panel data, Third Edition, ISBN: 978-1-107-65763-2
14. Chery, R., Mary, DW. and Suzanne, AB. (2012). Patient-Centered Care and Outcomes: A Systematic Review of the Literature. DOI: [10.1177/1077558712465774](https://doi.org/10.1177/1077558712465774)
15. Christian, B., Sabine, H. S., Silke R. (2010). Data Bases, the Base for Data Mining, Data Mining in Crystallography, Volume 134. pp 135–167, DOI: [10.1007/430_2009_5](https://doi.org/10.1007/430_2009_5)
16. Citro, C. F., Martin, M. E. and Straf, M. L. (2009). Principles and Practices for a Federal Statistical Agency: Fourth Edition. ISBN: 978-0-309-12175-0, DOI: [10.17226/12564](https://doi.org/10.17226/12564)
17. Consuela, H., Dominique, S., Philippe, A., Marie, C. J., and Patrice, D. (2005). Author information, AMIA AnnuSymp Proc.: 296–300. PMCID: PMC156070
18. David, W. A.(1992). Tolerating noisy, irrelevant and novel attributes in instance-based learning algorithms. Journal—International Journal of Man-Machine Studies—Special issue: symbolic problem solving in noisy and novel task environments archive, Volume 36 Issue 2, Feb.1992, Academic Press Ltd. London, UK, UK, DOI: [10.1016/0020-7373\(92\)90018-G](https://doi.org/10.1016/0020-7373(92)90018-G)
19. Evangelos, T. (2013). Multi-criteria Decision Making Methods: A Comparative Study. ISBN: 9781475731576
20. Ewa, O. (2013), Incomplete Information: Rough Set Analysis. ISBN: 9783790818888

21. Gilson, A., Giraldi, Fabio, P., Bruno, S., Vinicius, F., and Dutra, M. L. (2005). Data Integration Middleware System for Scientific Visualization., RJ, ZIP 25651-070, Brazil
22. Gombert, B. G. and Rochester, (2015). Systems and Methods for Attaching Electronic Versions of Paper Documents to Associated Patient Records in Electronic Health Records; United States Patent Application 20150339441 Kind Code: A1
23. Gribble, S., Halevy, A., Ives, Z., Rodrig, M. and Suci, D. (2001). What can databases do for peer-to-peer? In Proceedings of the Fourth International Workshop on the Web and Databases, WebDB
24. Hai, B. T., Trong, H. D. and Ngoc, T. N. (2013). *A Hybrid Method for Fuzzy Ontology Integration*. An International Journal, 44, pp. 133–154
25. Hsu, J. T., Hsieh, S. H., Cheng, P.H., Chen, S. J. and Lai, F. P. (2011) “Ubiquitous mobile personal health system based on cloud computing,” in Proc. IEEE Region 10 Conf. TENCN, Bali, Indonesia, Nov. 2011, pp. 1387–1390
26. Ibrahim, AFS. (2016). New Secure Solutions for Privacy and Access Control in Health Information Exchange; <http://dx.doi.org/10.13023/ETD.2016.307>
27. Ido, D. and Lillian, L. (1997). Fernando Pereira. Similarity-based methods for word sense disambiguation. DOI: [10.3115/979617.979625](https://doi.org/10.3115/979617.979625)
28. Illhoi, Y., Patricia, A., Miroslav, M., Keila, PH., Rajitha, G., Jia-Fu, C. and Lei, H. (2012). Data Mining in Healthcare and Biomedicine: A Survey of the Literature, DOI: [10.1007/s10916-011-9710-5](https://doi.org/10.1007/s10916-011-9710-5)
29. Jasmine, P. (2013). 6 Advantages of Using a Cloud-Based EHR System for Physician Practices, <http://hitconsultant.net/2013/11/19/6-advantages-of-using-a-cloud-based-ehr-system-for-physician-practices/> (Accessed 12 November 2016)
30. Jens, M., Stefan, O., Daniel, F., Christoph, H., Janina, K. and Wolfgang, H. (2012). Efficient data management in a large-scale epidemiology research project, Computer Methods and Programs in Biomedicine; <http://dx.doi.org/10.1016/j.cmpb.2010.12.016>
31. Jimeng, S., Fei, W. and Jianying, H. (2016) Shahram Edebollahi Supervised Patient Similarity Measure of Heterogeneous Patient Records, IBM TJ Watson Research Center, http://www.kdd.org/exploration_files/V14-01-03-Sun.pdf (Accessed 23 August 2016)
32. Jinyuan, S., Yuguang, F. and Xiaoyan, Z. (2010). Privacy and emergency response in e-healthcare leveraging wireless body sensor networks; DOI: [10.1109/MWC.2010.5416352](https://doi.org/10.1109/MWC.2010.5416352)
33. Karov, Y. and Edelman, S. (1996). Learning similarity-based word sense disambiguation from sparse data. pp. 42–55. Somerset, New Jersey: Association for Computational Linguistics
34. Liaw, S.T. (2012). Towards an ontology for data quality in integrated chronic disease management: A realist review of the literature. International journal of medical informatics, pp. 1–15
35. Leonardo, C. B., Jéssica, O. de S., Fábio, R. J., Caio, S. C., Márcio, R. de C., Vânia, P. de A., N. and Regina, B. de A. (2015). Methodology for Data and Information Quality Assessment in the Context of Emergency Situational Awareness, DOI: [10.1007/s10209-016-0473-0](https://doi.org/10.1007/s10209-016-0473-0)
36. Lise, P., Jennifer, P., Robyn, T., and Yuko, K. (2005). The Impact of Electronic Health Records on Time Efficiency of Physicians and Nurses: A Systematic Review, doi:[10.1197/jamia.M1700](https://doi.org/10.1197/jamia.M1700)
37. Lord, P.W., Stevens, R.D., Brass, A. and Goble, C. A. (2003). Semantic similarity measure as a tool for exploring the gene ontology, pp. 601–612. 2003
38. Maedche, A., Motik, B., Silva, N. and Volz, R. (2002). Mafra - a mapping framework for distributed ontologies. DOI: [10.1007/3-540-45810-7_23](https://doi.org/10.1007/3-540-45810-7_23), Print ISBN: 978-3-540-44268-4, Online ISBN: 978-3-540-45810-4
39. Maglogiannis, I., Doukas, C., Kormentzas, G. and Pliakas, T. (2009). Wavelet-Based Compression With ROI Coding Support for Mobile Access to DICOM Images Over Heterogeneous Radio Networks, vol. 13, no. 4, pp. 458–466, July 2009
40. Matthew, G. C., Russel, K., Marisa, R. and Richard, S. (2016). Utility of Daily Mobile Tablet Use for Residents on an Otolaryngology Head & Neck Surgery Inpatient Service; DOI: [10.1007/s10916-015-0419-8](https://doi.org/10.1007/s10916-015-0419-8)

41. Moon, F., Juan, M., Campos, I. and Luis, M. (2016). Personalizing xml information retrieval; ISBN: 9788491250548
42. Nir, M. and Taleah, H. (2011). Risk Management and Healthcare Policy. Department of Health Care, Organisation and Policy, USA; DOI: [10.2147/RMHP.S12985](https://doi.org/10.2147/RMHP.S12985)
43. Nirase, F. A., Azreen, A., Shyamala, D., Masrah, A. and Azmi, M. (2016). An integrated method of associative classification and neuro-fuzzy approach for effective mammographic classification. DOI: [10.1007/s00521-016-2290-z](https://doi.org/10.1007/s00521-016-2290-z)
44. Nuno, L., Seyma, N. S. and Jorge, B. (2015). A Survey on Data Quality: Classifying Poor Data. Conference Paper. DOI: [10.1109/PRDC.2015.41](https://doi.org/10.1109/PRDC.2015.41)
45. Özgür, N. B. (2007). An Intelligent Fuzzy Object Oriented Database Framework for Video Database Applications. Computer Engineering Department, METU, Ankara, Turkey
46. Pagliari, C., Donnan, P., Morrison, J., Ricketts, I., Gregor, P. and Sullivan, F. (2005). Adoption and perception of electronic clinical communications in Scotland. 13 (2), 97–104
47. Peter, B. Jensen, L. J. and Søren B. (2012). Mining electronic health records: towards better research applications and clinical care. DOI: [10.1038/nrg3208](https://doi.org/10.1038/nrg3208)
48. Raychaduri, S. and Altman, R.B. (2003). A literature-based method for assessing the functional coherence of a gene group, *Bioinformatics*, 19(3), pp. 396–401
49. Reb, Y., Pan, J. Z., and Zhao, Y. (2010). Soundness Preserving Approximation for TBox Reasoning. In the Proc. of the 25th AAAI Conference (AAAI2010)
50. Rinehart, T. L. A., Harman, L. B. (2006). Ethical Challenges in the Management of Health Information. Jones and Bartlett. Privacy and confidentiality; p. 54. Chapter 3
51. Rodney, L., MD (2012). Cloud Computing in the Health Care Setting: Advantages and Disadvantages. ISSN 1559-4939
52. Sanchez, E. and Yamanoi, T. (2006). Fuzzy Ontologies for the Semantic Web. 7th International Conference on Flexible Query Answering Systems (FQAS 2006), Milan, Italy. pp. 7–10
53. Sarkies, M. N., Bowles, K. A., Skinner, E. H., Mitchell, D., Haas, R., Ho, M., Salter, K., May, K., Markham, D., O'Brien, L., Plumb, S. and Haines, T. P. (2015). Data Collection Methods in Health Services Research, Hospital Length of Stay and Discharge Destination. DOI: [10.4338/ACI-2014-10-RA-0097](https://doi.org/10.4338/ACI-2014-10-RA-0097)
54. Schaal, M., Smyth, B., Mueller, R. M. and MacLean R. (2012). Information Quality Dimensions for the Social Web, in Proceedings of the International Conference on Management of Emergent Digital EcoSystems, ser. MEDES '12. New York, NY, USA: ACM, pp. 53–58
55. Shaker, E. S., Mohammed, E. and Riad A.M. (2015). A fuzzy-ontology-oriented case-based reasoning framework for semantic diabetes diagnosis. Volume 65, Issue 3, Pages 179–208, DOI: <http://dx.doi.org/10.1016/j.artmed.2015.08.003>
56. Sumit Chakraborty (2014). Healthcare System Innovization through Digital Transformation. Reference: HSIDT/ V1.0/ CR 15082014
57. Tania, C., Genoveva, V., Jose, L. and Zechinelli, M. (2004). Building Multimedia Data Warehouses from Distributed Data. *e-Gnosis* [online], Vol. 2, Art. 10
58. Terence, H. L. (2015). Regression Analysis of Cloud Computing Adoption for U.S. Hospitals; <http://scholarworks.waldenu.edu/dissertations> (Accessed 31 Oct 2016)
59. Vojtáš, P. (2006). A fuzzy EL Description Logic with crisp roles and fuzzy aggregation for web consulting. *Knowledge-Based Systems (IPMU 2006)*
60. Wang, C., Zhang, J. and Qin, L. (2016). Design & Research of Legal Affairs Information Service Platform Based on UIMA and Semantics. ISSN: 2233-7857 IJFGCN
61. Weng, C. C., Hao, H., Carolina, O. L. U. and Yong, C. (2013). Benefits and Challenges of Electronic Health Record System on Stakeholders: A Qualitative Study of Outpatient Physicians, DOI: [10.1007/s10916-013-9960-5](https://doi.org/10.1007/s10916-013-9960-5)
62. WHO: http://www.who.int/violence_injury_prevention/surveillance/en/ (Accessed 24 July 2016)
63. Yeturu, A., Michael, P. and Naeem, A. (2016). Methods, devices and systems enabling a secure and authorized induction of a node into a group of nodes in a distributed computing environment; Publication number US9264516 B2