Review

What You Need to Know Before Implementing a Clinical Research Data Warehouse: Comparative Review of Integrated Data Repositories in Health Care Institutions

Kristina K Gagalova[1,2,3], MSc; M Angelica Leon Elizalde[3,4], MSc; Elodie Portales-Casamar[3,5], PhD; Matthias Görges[3,6], PhD

[1]Canada's Michael Smith Genome Sciences Centre, BC Cancer, Vancouver, BC, Canada

[2]Bioinformatics Graduate Program, University of British Columbia, Vancouver, BC, Canada

[3]Research Institute, BC Children's Hospital, Vancouver, BC, Canada

[4]School of Population and Public Health, University of British Columbia, Vancouver, BC, Canada

[5]Department of Pediatrics, University of British Columbia, Vancouver, BC, Canada

[6]Department of Anesthesiology, Pharmacology and Therapeutics, University of British Columbia, Vancouver, BC, Canada

Corresponding Author:
Matthias Görges, PhD
Department of Anesthesiology, Pharmacology and Therapeutics
University of British Columbia
Rm V3-324
950 West 28th Avenue
Vancouver, BC, V5Z 4H4
Canada
Phone: 1 875 2000 ext 5616
Email: mgorges@bcchr.ca

Abstract

Background: Integrated data repositories (IDRs), also referred to as clinical data warehouses, are platforms used for the integration of several data sources through specialized analytical tools that facilitate data processing and analysis. IDRs offer several opportunities for clinical data reuse, and the number of institutions implementing an IDR has grown steadily in the past decade.

Objective: The architectural choices of major IDRs are highly diverse and determining their differences can be overwhelming. This review aims to explore the underlying models and common features of IDRs, provide a high-level overview for those entering the field, and propose a set of guiding principles for small- to medium-sized health institutions embarking on IDR implementation.

Methods: We reviewed manuscripts published in peer-reviewed scientific literature between 2008 and 2020, and selected those that specifically describe IDR architectures. Of 255 shortlisted articles, we found 34 articles describing 29 different architectures. The different IDRs were analyzed for common features and classified according to their data processing and integration solution choices.

Results: Despite common trends in the selection of standard terminologies and data models, the IDRs examined showed heterogeneity in the underlying architecture design. We identified 4 common architecture models that use different approaches for data processing and integration. These different approaches were driven by a variety of features such as data sources, whether the IDR was for a single institution or a collaborative project, the intended primary data user, and purpose (research-only or including clinical or operational decision making).

Conclusions: IDR implementations are diverse and complex undertakings, which benefit from being preceded by an evaluation of requirements and definition of scope in the early planning stage. Factors such as data source diversity and intended users of the IDR influence data flow and synchronization, both of which are crucial factors in IDR architecture planning.

XSL•FO
RenderX

Introduction

Background

An electronic health record (EHR) is a system for the input, processing, storage, and retrieval of digital health data. EHR systems have been increasingly adopted in the United States over the past 10 years [1], and their use is spreading worldwide in both hospital and outpatient care settings [2,3]. An EHR is typically organized in a patient-centric manner and has become a powerful tool to store data in a time-dependent and longitudinal structure. EHR data can also be integrated into an enterprise data warehouse or integrated data repository (IDR). IDRs collect heterogeneous data from multiple sources and present them to the user through a comprehensive view [4]. Unlike EHRs, IDRs offer specialized analytical tools for researchers or analysts to perform data analyses.

An IDR is a significant institutional investment in terms of both initial costs and maintenance, but it offers the advantage of clinical data reuse beyond direct clinical care, such as for research and quality improvement studies. Secondary use of clinical data is a rapidly growing field [5,6]; an increasing number of institutions have implemented in-house IDRs and several others are developing IDRs for future research endeavors.

Unlike clinical practice, which focuses on enhancing the well-being of current patients, the purpose of an IDR is to produce generalized knowledge that can be extended to future patients. Typical applications of IDRs include retrospective analysis and hypothesis generation [7]. Some IDRs also support clinical applications, such as clinical decision support systems (CDSSs), that work alongside clinical practice to estimate risk factors or predictive scores associated with clinical treatments. CDSSs help to avoid medical errors and deliver efficient and safer care by assisting the provider with diagnosis, therapy planning, and treatment evaluation decisions [8]. All these applications are valuable resources that have the potential to improve the quality of health care [9] and reduce health costs if implemented appropriately [10].

Objective

Our study is motivated by the need to develop a pediatric IDR at our institution and by the lack of literature providing practical recommendations to apply during the initial development stages. Reviews by Shin et al [11] and Huser et al [12] highlighted the recommended characteristics when designing an IDR; however, they include only a small set of examples and a limited number of example IDRs. Since 2014, the IDR landscape has evolved rapidly, and thus, we felt more recent developments needed to be better addressed as well. A 2018 review by Hamoud et al [13] provided a comprehensive description of most recent data warehouses, including information about their data content, processing, and main purpose; it also provides general recommendations for the implementation of an IDR, but no practical considerations to guide the planning stages.

This study compares the features of contemporary IDRs and presents some guiding principles for the design and implementation of a clinical research data warehouse. Our research objective was to identify the major features of contemporary IDRs and obtain a list of established architectures used in the field of health informatics. We expect that this review will be useful for other small- to medium-sized institutions that plan to implement an institutional IDR and have no extensive experience in the field.

Methods

We conducted a literature review and a targeted web-based search to identify the major existing IDRs and synthesized the retrieved information around key themes.

Literature Review Search

We performed a narrative review following the procedure described below. First, a literature search was conducted using Ovid MEDLINE (Medical Literature Analysis and Retrieval System Online) and IEEE Xplore (Institute of Electrical and Electronics Engineers Xplore), queried in March 2020 (Figure 1). Articles were identified in 2 iterative phases. The first phase used an initial list of keywords querying for infrastructure purposes (data integration, such as linkage and harmonization) as well as infrastructure type and hospital setting (Multimedia Appendix 1: A1). The second phase search used additional keywords identified from the titles and abstracts of articles retrieved in the first phase (Multimedia Appendix 1: A1). Second, Google Scholar was queried for major article keywords (Integrated Data Repository) OR (Clinical Data Warehouse), and the first 150 retrieved hits were screened. The query was executed in a single search stage because the traditional search methods using Ovid MEDLINE and IEEE Xplore already produced exhaustive results.

Figure 1. Article selection process. The diagram shows the number of articles at each stage of selection for each of the 3 databases: MEDLINE (Medical Literature Analysis and Retrieval System Online), IEEE Xplore (Institute of Electrical and Electronics Engineers Xplore), and Google Scholar.
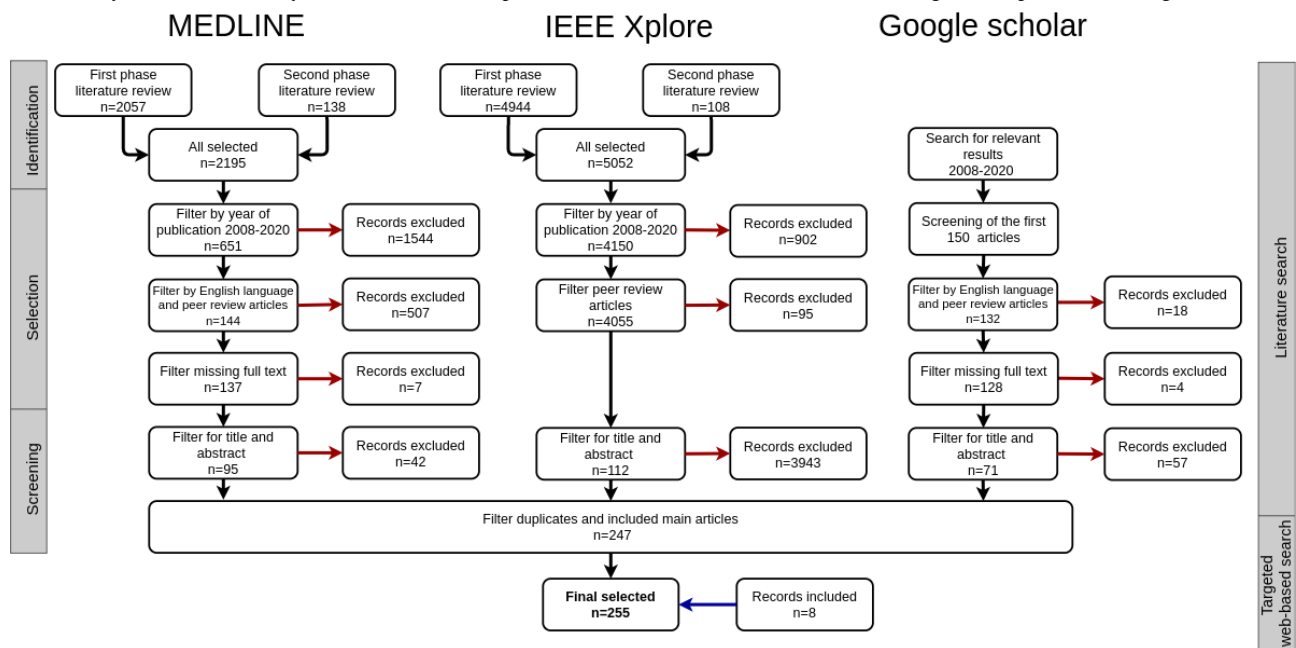
We selected peer-reviewed articles, published in the English language between January 2008 and March 2020, to include the most current data warehouse features. Non-English articles were excluded because of a lack of resources for translation. We retained articles for which the full text was available and removed duplicates. KG read the abstracts, and the articles describing specific data integration strategies, describing architecture structures, or providing more information about the data models were included. When it was unclear whether an article should be included, the authors EPC and MG were consulted. Duplicated articles were removed using EndNote reference management software (Clarivate Analytics). Additional articles providing the most up-to-date information about selected IDRs or cited by the selected articles were included in the selection process because they were considered relevant for the IDR definition. Targeted Web-Based Search of Known Institutional IDRs

We manually queried nonpublished resources with the goal of adding contemporary data warehousing practices implemented in large North American hospitals. A convenience sample of hospitals known to be leaders in these types of data warehousing was suggested by EPC and MG.

Additionally, we browsed publicly available information on each of the targeted institutional websites (Multimedia Appendix 1: A2). This was complemented with relevant peer-reviewed articles cited in these websites related to the design, implementation, and applications of such repositories.

Manual Shortlisting for a Comparative Review Analysis

For the comparative review analysis, we performed a manual selection to shortlist articles specifically describing IDR architectures. The shortlisting considered the major focus of the article and the presence of significant details describing data integration, data processing, or database services. The selected articles were searched for related IDR projects and further web-based resources (Table 1 and Multimedia Appendix 1: A3).

Table 1. Institutions and major features of the integrated data repositories.

IDR[a,b]	IDR scope	Architecture model	Standard common data model	Standard terminologies	Primary references
The National Institutes of Health Clinical Center-Small-sized institution					
Biomedical Translational Research Information System (BTRIS)	General care	General	N/A[c]	RED[d]	[14]
Deceased subjects (dsIDR)	Deceased subjects	General	N/A	RED	[15]
University of Kansas Medical Cent er-Medium-sized institution					
Healthcare Enterprise Repository for Ontological Narration (HERON)	General care	General	i2b2[e]	ICD[f]-9/ICD-10, CPT[g], RxNorm[h], SNOMED-CT[i], NDFRT[j], NCI[k], FDB[l]	[16,17]
Stanford University Medical Center- Medium-sized institution					
Stanford Translational Research Integrated Database Environment (STRIDE)	General care	General	i2b2, OMOP[m]	ICD-9, CPT, RxNorm, SNOMED-CT	[18,19]
STAnford Research Repository (STARR)	General care	General	i2b2, OMOP	ICD-9, CPT, RxNorm, SNOMED-CT	[20]
The Georges Pompidou University Hospital (HEGP)-Medium-sized institution					
HEGP CDW[n] platform	General care, cardio-vascular, cancer	General	i2b2	ICD-10, LOINC[o], SNOMED-CT	[21-23]
Hanover Peter L. Reichertz Institute- Large-sized i nstitution					
Hanover Medical School Translational Research framework (HaM-STR)	General care	General	i2b2	ICD-10, LOINC	[24]
Erlangen University Hospital-Large-sized institution					
Clinical data warehouse	General care	General	I2b2	LOINC, NCI	[25]
Seoul St. Mary Hospital-Large-sized institution					
Prostate cancer research database	Cancer	General	N/A	N/A	[26]
Lead partner: Cincinnati Children's Hospital Medical Center-Collaborative project					
Maternal and Infant Data Hub (MIDH)	Perinatal	General	OMOP	ICD-9/ICD-10, SNOMED-CT	[27]
Georges Pompidou, Cochin and Necker Hospitals- Collaborative project					
CAncer Research for PErsonalized Medicine (CARPEM)	Cancer	General	Variant of i2b2 (tranSMART[p])	ICD-9/ICD-10, SNOMED-CT, ATC[q], GO[r], HPO[s]	[28]
Learning Healthcare System (LHS) across South Carolina-Collaborative project					
Health Science South Carolina (HSSC) clinical data warehouse	General care	General	i2b2	N/A	[29]
Windber Research Institute-Collaborative project					
Data Warehouse for Translational Research (DW4TR)	Cancer	General	N/A	MeSH[t], SNOMED-CT, NCI, caBIG VCDE[u]	[30,31]
Veterans' Health Administration-Collaborative project					
VA EHR (Veterans Administration's electronic health records)	General care	General with CDSS	N/A	ICD-9	[32]
Coordinated by Medtronic Iberica SA- Collaborative project					
Models and simulation techniques for discovering diabetes influence factors (MOSAIC)	Diabetes	General with CDSS	i2b2	ICD-9, DRG[v], ATC	[33]
National collaboration- Collaborative project					

IDR[a,b]	IDR scope	Architecture model	Standard common data model	Standard terminologies	Primary references
China Stroke Data Center (CSDC)	Cerebrovascular	General with CDSS	N/A	N/A	[34]
Houston Methodist Hospital-Large-sized institution					
Methodist Environment for Translational Enhancement and Outcomes Research (METEOR)	General care	General with CDSS	Extension of i2b2	ICD-9, CPT	[35]
Mayo Clinic- Large-sized institution					
Mayo Enterprise Data Trust (MEDT)	General care	General	i2b2	LexGrid[w]	[36]
Ovarian cancer registry	Cancer	General	i2b2	LexGrid	[37]
Translational Research Center (TRC)	Cancer	Biobank-driven	i2b2	LexGrid	[38]
Vanderbilt University Medical Center- Large-sized institution					
Synthetic Derivative	General care	General	N/A	FDB, ICD-9, CPT	[39]
BioVU	General care	Biobank-driven	N/A	FDB, ICD-9, CPT	[40]
The Children's Hospital of Philadelphia- Medium-sized institution					
Biorepository Portal (BRP)	Cancer, pediatric	Biobank-driven	Harvest	N/A	[41]
University of São Paulo-Large-sized institution					
BioBankWarden (BBW)	Cancer	Biobank-driven	N/A	ICD-10, SNOMED-CT, LOINC, GO	[42]
University of Pavia and Fondazione S. Maugeri- Large-sized institution					
onco-i2b2	Cancer	Biobank-driven	i2b2	SNOMED-CT	[43]
Leon Berard Cancer Center- Small-sized institution					
CLB-IT[x]	Cancer	User-controlled application layer	N/A	ADICAP[y], ICD-O	[44]
Lead partner: University of Utah-Collaborative project					
Federated Utah Research and Translational Health electronic Repository (FURTHeR)	Several	Federated	i2b2, OMOP, OpenMRS[z]	ICD-9/ICD-10, LOINC, SNOMED-CT, RxNorm	[45]
OpenFurther	Several	Federated	i2b2, OMOP, OpenMRS	ICD-9/ICD-10, LOINC, SNOMED-CT, RxNorm	[46]
Lead partner: University of Utah-Collaborative project					
Pediatric Health Information System (PHIS+)	Pediatric	Based on FURTHeR	i2b2	LOINC, SNOMED-CT	[47]
@neurIST European Project-Collaborative project					
@neurIST platform	Cerebrovascular	Federated as in FURTHeR	N/A	@neurIST ontology[aa]	[48]
University Clinics in Northern Germany- Collaborative project					

IDR[a,b]		IDR scope	Architecture model	Standard common data model	Standard terminologies	Primary references
	Research Data Management System (RDMS)	Cancer	Federated as in FURTHeR	i2b2	ICD-10, SNOMED-CT	[49]

[a]IDR: Integrated Data Repository

[b]The IDRs are defined by their data scope, architecture model (as defined by the major design class represented in Figure 2), standard common data model, standard terminology, and primary reference.

[c]N/A: not applicable, n=4 in Standard Terminology.

[d]RED: Research Entities Dictionary, n=1.

[e]i2b2: Informatics for Integrating Biology and the Bedside

[f]ICD-9/ICD-10/ICD-O: International Classification for Diseases, version 9/10, O for oncology, n=14.

[g]CPT: Current Procedural Terminology, n=4.

[h]RxNorm: standardized nomenclature for clinical drugs, n=3.

[i]SNOMED-CT: Systematized Nomenclature of Medicine-clinical terms, n=11.

[j]NDFRT: National Drug File Reference Terminology, n=1.

[k]NCI: National Cancer Institute, n=2.

[l]FDB: First Databank, n=2.

[m]OMOP: Observational Medical Outcomes Partnership

[n]HEGP CDW: Hôpital Européen Georges Pompidou Clinical Data Warehouse, n=1.

[o]LOINC: Logical Observation Identifiers Names and Codes, n=5.

[p]tranSMART: Open-source data platform for translational research, n=1.

[q]ATC: Anatomical Therapeutic Chemical Classification, n=2.

[r]GO: Gene Ontology, n=2.

[s]HPO: Human Phenotype Ontology, n=1.

[t]MeSH: Medical Subject Headings; n=1.

[u]caBIG VCDE: the cancer Biomedical Informatics Grid Vocabulary and Data Elements Workspace, n=1.

[v]DRG: Diagnosis Related Group, n=1.

[w]LexGrid: Lexical Grid, n=1.

[x]CLB-IT: Léon Bérard Cancer Center-IT.

[y]ADICAP: Association pour le Développement de l'Informatique en Cytologie et en Anatomie Pathologique, n=1.

[z]OpenMRS: Open Medical Record System, n=1.

[aa]@neurIST ontology, n=1.

Figure 2. Architecture models identified from selected integrated data repositories (IDRs). Arrows indicate data output because of a query (blue) and data input (orange) because of data integration or update. Continuous lines show data query and integration applied by research users, whereas dashed lines are data queries performed by operational or clinical users.
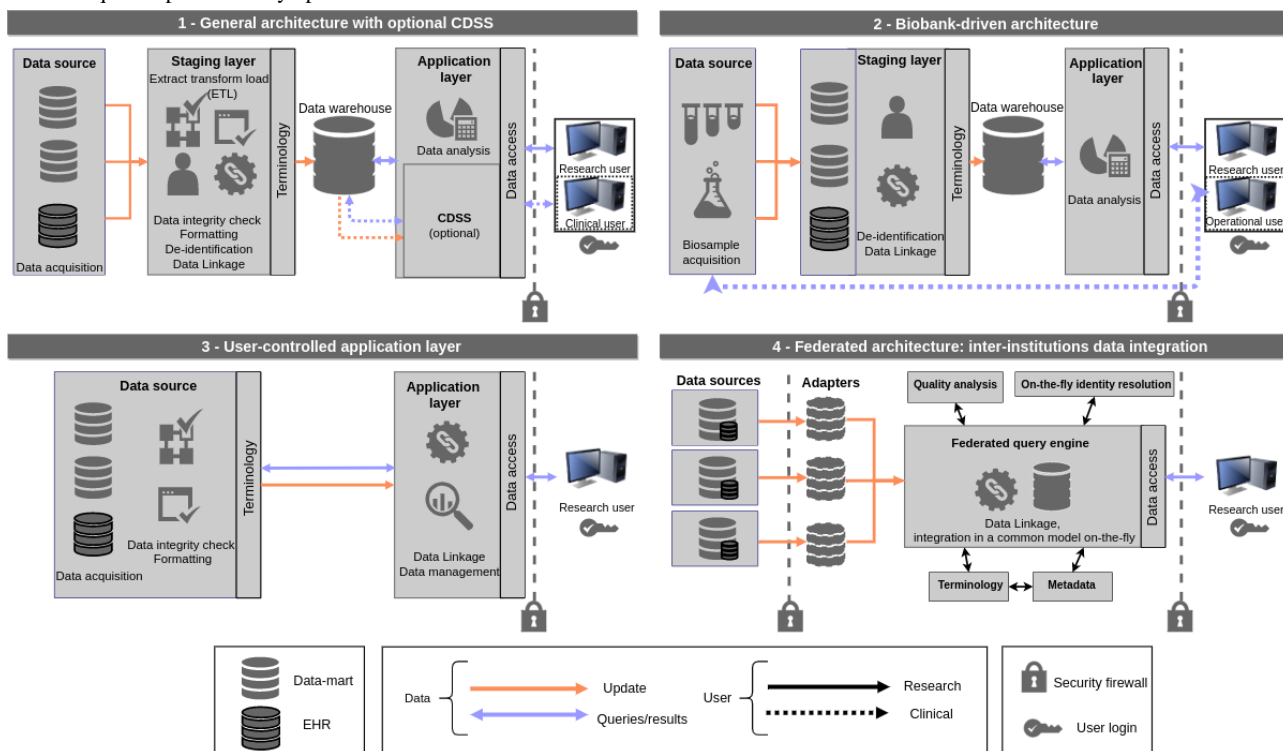
Literature Synthesis and Institution Characterization

Information from the literature was aggregated through thematic analysis and collapsed into 4 classes of IDR architectures. We evaluated the main features of the identified IDRs, such as data processing components, data characteristics, common terminologies, and data models. Features were summarized, compared, and contrasted. We extracted information about host institutions and divided them into small (≤500 beds), medium (500-1000 beds), and large (>1000 beds) institutions based on the number of beds listed on the institution's websites.

Analysis of Word Content

Selected articles were uploaded into NVivo 12 (QSR International LLC) for qualitative analysis, specifically to count the word frequency in the selected papers. The words with a minimum length of 5 in the full text were counted, excluding stop words, and grouped by synonyms. The word frequency is represented as a word cloud, generated with R (R Foundation for Statistical Computing) and *wordcloud* package 2.6.

Citations Analysis

The references of the articles describing IDRs were downloaded in a semiautomated manner using Content Extractor and Miner software [50] to parse the full-text PDF files. References to web resources, video-cast meetings, and software were removed, and partial references were manually corrected. The references were grouped by *first author* and *year of publication* and loaded in R (R Foundation for Statistical Computing) and plotted with UpSetR [51].

Results

Overview

A total of 241 articles were identified in the literature search [11,13-19,21-29,31,33-35,37,43,44,47-49,52-264]; the largest number of articles were identified in IEEE Xplore (n=112), followed by MEDLINE (n=95), and Google Scholar (n=71). After removing duplicates (n=24), we added 3 articles that were frequently cited in the selected articles but were missing from our search results [30,36,42]. Three articles [38,40,45] were further added that provided additional details relevant to the review topic. Finally, 1 article was replaced by a more updated publication [265]. These 247 articles were combined with the targeted web-based search [32,39-41,266-269]; hence, we identified a total of 255 articles (Figure 1). The most frequent words in the articles were *system,information*, *study*, *project*, and *design* (Multimedia Appendix 1: A4.1). A total of 79 of these 255 articles were published between 2014 and 2016, and 34 were published in 2019; this date range covers the full range of initially identified articles in this domain area (Multimedia Appendix 1: A4.2 and A4.3).

A total of 116 articles were presented in proceedings of international scientific conferences, particularly those published in the book series Studies in Health Technology and Informatics (n=23); this included the World Congress of Medical and Health Informatics and Medical Informatics Europe. The second most frequent proceedings were the American Medical Informatics Association annual symposium and joint summits on translational science (n=12). The most frequently observed journals were the Journal of the American Medical Informatics Association (n=9) and BioMed Central (BMC; n=8), with BMC

Bioinformatics being the most common. More details about the individual conferences and journals can be found in Multimedia Appendix 1: A4.4 and A4.5.

For this review, we focused on the 34 articles describing 29 IDRs for which sufficient design details were presented. The additional web resources describing 2 IDRs, Stanford Translational Research Integrated Database Environment (STRIDE) and Federated Utah Research and Translational Health Electronic Repository (FURTHeR), referred to novel projects STAnford Research Repository (STARR) [20] and OpenFurther [46], respectively, which increased the number of IDRs to 31 from 25 different institutions or collaborative projects (Table 1). In reviewing the references in these 34 articles, we observed only a small overlap, with 1 reference [270] being found in common in a maximum of 11 articles (Multimedia Appendix 1: A5.1). The most frequently cited among the 34 are onco-Informatics for Integrating Biology and the Bedside (i2b2) [43,271], STRIDE [18], and the Mayo Clinic [36] IDRs, cited in 8, 5, and 4 articles, respectively (Multimedia Appendix 1: A5.2).

IDRs represent a variety of applications of health data warehousing for research. Although they share common characteristics, as described in detail below, they also demonstrate the many different purposes they can serve. For example, BioVU [40] and the Synthetic Derivative [39] at Vanderbilt University Medical Center are examples of a biobank-driven database that automatically couples patients' clinical information to biological samples (biosamples). The power of this system is its connection between genotype and phenotype and its large number of biosamples (>50,000), which allows a rich set of cohort research studies. The Maternal and Infant Data Hub (MIDH) at Cincinnati Children's Hospital Medical Center [27] is a regional perinatal data repository that integrates a large and diverse set of data from different institutions. The strength of the project is the combination of delivery and postdischarge hospital data and the linked mother and child data sets. The pilot database contains approximately 70,000 newborns and 42,000 pediatric postnatal visits. Another example is the Hanover Medical School Translational Research Framework (HaMSTR) framework at the Hanover Peter L. Reichertz Institute [24], which was developed to automatically load data from a clinical data repository into a standard data model that researchers can query; it is a successful example of fast data upload and query using data structures designed from standard data models available for clinical research.

Characteristics of the Institutions in the Selected IDR Sample

We identified 2 types of IDRs: those developed for use in a single institution (n=19) and those implemented for a collaborative project (n=12). The latter typically integrate patient data and provide project-specific tools. The median number of different institutional partners in a collaborative IDR is 6, with one of the partners acting as an organizational hub. The partners range from research institutes, laboratories, and private institutions to university medical centers.

The IDRs were further divided by their scope (Table 1), which were classified as general or specialized medical care (cancer, pediatrics, perinatal, cerebrovascular, or cardiovascular). Seven of the 10 IDRs containing specialized data were collaborative projects, likely indicating the need to pool data from several institutions when dealing with smaller but more focused patient populations.

Four Major Architecture Models Used in Our Selected IDR Sample

We identified 4 overarching conceptual architectures that summarize the data layers in the selected IDRs (Figure 2). Different institutions can implement multiple architectures for different purposes; we assigned each IDR to a category considering the major features of the IDR, as described in their respective articles.

The general architecture model is the most common model, with 19 identified IDRs structured around medical data mining (Figure 2, *General architecture with optional CDSS*). In outline, different data marts are transferred to a staging layer that harmonizes the input to a common data view; data are loaded into a common data warehouse and queried through an application layer that communicates with the user; a CDSS tool can provide added functionality. Hence, in this architecture, each data source is originally stored in an independent data mart, collecting data from a separate research or clinical source within the same institution. Data are processed in the staging layer, which reshapes the input to an integrated view through several steps of data linkage, transformation, and harmonization. The next stage of processing is loading the data into a single database connected to an application layer that provides the tools for end users, typically researchers, to access and analyze the data securely with different services. An example of an IDR providing multiple services is the STRIDE architecture stack [18], which includes several services for data analysis or research data management. The articles describing METEOR [35], CSDC [34], models and simulation techniques for discovering diabetes-related factors (MOSAIC) [33], and Veterans Administration's EHRs [32] provide further details about the integration of CDSS tools in the architecture. In these cases, the architecture model is divided into CDSS and data analysis modules, both of which communicate with the common database. The CDSS allows clinical staff to retrieve real-time individual patient records and to use analytical models to make risk prediction. The CDSS tools described by METEOR and MOSAIC, for example, learn from the clinical data stored in the data warehouse and estimate risk factors predicting hospital readmission or long-term complications.

The Health Science South Carolina (HSSC) [29] IDR gathers data from different clinical systems implemented in various institutions, all of which are party to a data collaboration agreement that authorizes data aggregation in a single data warehouse. This data warehouse contains a longitudinal record for each individual across all institutions. Data processing and terminology mapping occur in a conceptual staging layer, as in the case of the general architecture model.

In the case of the Erlanger University Hospital IDR [25], terminology is mapped using vocabularies that are manually curated and mapped through an automatic workflow that processes the raw data to the final data warehouse format. Other

IDRs that make use of multiple terminologies are health care enterprise repository for ontological narration [16], Research for PErsonalized Medicine (CARPEM) [28], and STRIDE [18], but further details of their mapping processes were not available.

The biobank-driven architecture model is built around a particular application, in this case, biobanking (Figure 2, *Biobank-driven architecture*). This model is similar to the general architecture model but, in this case, the IDR is built around the biosamples database. The biosample data integration occurs at the staging layer. The main feature is that the model allows the biosample operational user to access the raw and identified biobank data source for quality control and biosample management. An example of a biobank-driven structure is the biorepository portal (BRP) [41,266], which allows for the automatic integration of biosamples with clinical data, while maintaining unrestricted access to the biorepository for the operational team. The Mayo Clinic and Vanderbilt University adopt the general and biobank-driven architecture models in parallel.

The user-controlled application layer architecture model does not have a specific staging layer (Figure 2, *User-controlled application layer*). This architecture does not include a central data warehouse; the data are preprocessed and integrated from the original data sources only when the users query the data. Hence, data are processed in 2 stages: the first stage preprocesses the original data to a common format. The user query then carries out the final data integration function for the output delivery. In this architecture, a common data warehouse is not implemented, but rather the data are dynamically queried. An example is the text mining technology at the Léon Bérard Cancer Center (CLB) [44], which indexes text documents during the preprocessing stage and in which the users' queries return the exact documents matched.

The federated architecture is implemented for heterogeneous data retrieval and integration across multiple institutions (Figure 2, *Federated architecture*, adapted from OpenFurther). In this case, institutions selectively share their data through an adaptor system that applies common preprocessing, with data integrated on-the-fly in a virtual data warehouse. The FURTHeR federated query platform [45] builds a virtual IDR that responds to the needs of the user and calls several services for data resolution on-the-fly and upon query. The architecture model is flexible and operates using several services for data integration. An application of FURTHeR is the Pediatric Health Information System+ project [47], which combines data from 6 institutions. The IDR uses a federation component, which aggregates and stores translated query results in a temporary, in-memory database for presentation and analysis by the researcher for the duration of the user's session. Federated data integration was also proposed using a research data management system (RDMS) [49], which integrates clinical and biosample data from several institutions in Germany. The @neurIST [48] is a large IDR dedicated to translational research that includes data, computing resources, and tools for researchers and clinicians. Data are located across different sites and are securely shared with a grid infrastructure that allows federated data access.

The 4 types of architecture present different analytics tools, data presentation logic, and query interface based on the type of user they serve, which can be classified into 2 major groups: the first group, such as researchers and operational or business analysts, uses the IDR to identify important clinical features that occur at the level of patient cohorts. The second type of user, such as physicians and other health care professionals, uses the IDR to make decisions at an individual patient level, for example, to plan specific therapeutic interventions or predict risk. The first type of user is served by all the architecture models (*Research user* in Figure 2). The general architecture model that incorporates a CDSS presents a clear separation of both user types who have different applications for IDR data, with CDSS queries being made by clinical users (Figure 2, *General architecture with optional CDSS*). Similarly, the biobank-driven architecture model includes operational users who can directly query the information regarding patient biosamples for clinical applications (Figure 2, *Biobank-driven architecture*).

Data Retrieval and Update Are Influenced by the IDR Architecture Model

Both data update and integration schedules in an IDR are important features that define the timeliness of data. Here, we describe some of the key limiting steps and their occurrence in the different IDR architecture models.

Data Retrieval

The data processing involved in extraction, transformation, and loading (ETL) is described in detail in the articles of biomedical translational research information system (BTRIS) [14], HaMSTR [24], Mayo Translational Research Center (TRC) [38], CARPEM [28], onco-i2b2, Vanderbilt's Synthetic Derivative [39] and BioVU [40], and BRP [41]. These IDRs represent the general and biobank-driven architecture models, which implement a staging layer for the ETL process. A temporal sequence of the ETL steps is as follows:

1. Data extraction from source(s): The source data are extracted by an automatic (or manual) process.
2. Deidentification: Identifiable patient features, such as demographics or localization, are removed before loading into the IDR. The biobank-driven IDRs implement an automated process of this step without the need for extensive institutional reviews. In addition to the deidentified data, BTRIS [14] and Vanderbilt's Synthetic Derivative [39] maintain a parallel database with original identifiable patient entries for research purposes where appropriate.
3. Assignment of unique identifiers: Deidentified data are assigned unique patient identifiers that are used as a reference for linking.
4. Data transformation and standardization: Data are first checked for possible errors or missing values and are then transformed into a common format that is standard for all cohorts. Data may be subjected to transformation, such as the derivation of new values from the existing ones (pseudonymization) for maintaining privacy.
5. Standard terminology and ontology mapping: Data types are labeled with standard terminologies.

6. Data linkage: If the data are derived from multiple sources, they are linked and combined in the IDR.
7. Loading into the data warehouse: This is performed by either an update of existing data or a complete data re-import into the data warehouse.

The CLB [44] IDR (user-controlled application layer architecture model) uses specialized software to manipulate the content from unstructured data without using an ETL process. IDRs representing architecture model 4 do not provide additional information on the ETL process in their respective articles.

Data Update

Five of the selected articles provide additional information about the frequency of data updates in their IDRs. BTRIS [14] and Vanderbilt's Synthetic Derivative [39] argue for daily IDR updates as new source data accumulate daily. Onco-i2b2 [43] performs more frequent data synchronization, as frequent as every 15 min. A real-time data update is presented by METEOR

[35] and MOSAIC [33], which also integrate a CDSS in their architecture model and thus need this frequency to make actionable decisions. MOSAIC presents an example with asynchronous data update; although the CDSS is updated in real time, the demographics are synchronized only every 6 months. The general architecture model combined with a CDSS may require real-time data updates, whereas the general or the biobank-driven architecture models, without a CDSS, may have periodic updates that vary widely in frequency.

Major IDR Features: Data Type, Standard Terminology, and Common Data Model

Data Type

We have listed the data types in 19 of the selected IDRs based on information in the articles (Figure 3). The most common types of data are those extracted from EHR that include patient demographics, diagnoses, procedures, laboratory tests, and medications.

Figure 3. Common data types across IDRs. Columns show the main types of data collected in the selected IDRs. Gray-filled cells denote feature presence, with colors classifying the IDRs based on the examined architectures. Only 19 IDR articles contained enough information in their articles to be included in this figure. BRP: biorepository portal; BTRIS: biomedical translational research information system; CARPEM: cancer research for personalized medicine; CLB-IT: Léon Bérard Cancer Center Information Technology; DW4TR: Data Warehouse for Translational Research; EHR: electronic health record; HEGP: Hôpital Européen Georges Pompidou; HERON: health care enterprise repository for ontological narration; HSSC: Health Science, South Carolina; IDRs: integrated data repositories; Mayo Clinic-TRC: Mayo Clinic – Translational Research Center; METEOR: Methodist Environment for Translational Enhancement and Outcome Research; MIDH: Maternal and Infant Data Hub; MOSAIC: models and simulation techniques for discovering diabetes-related factors; Onco-i2b2; PHIS+: Pediatric Health Information System+; STARR: STAnford Research Repository; VUMC-BioVU: Vanderbilt University Medical Center–BioVU; VUMC-SD: Vanderbilt University Medical Center–Synthetic Derivative.
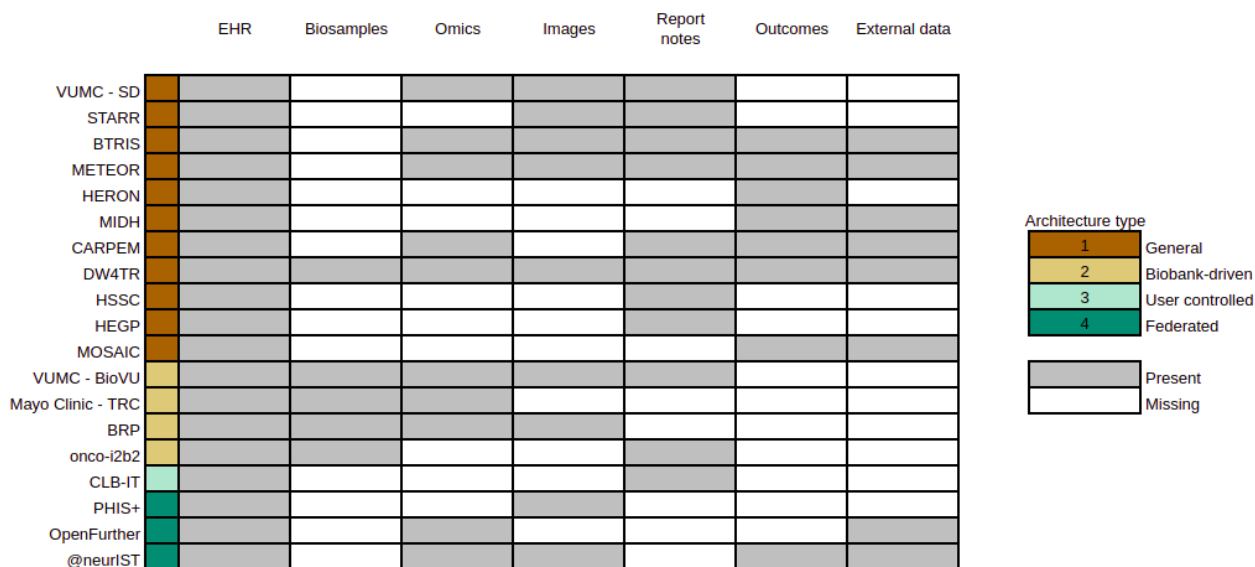
Several IDRs incorporate data from biosamples and their omics characterization, especially those based on the biobank-driven architecture model such as TRC [38], BRP [41,266], and BioVU [40]. Other examples of omics-based IDRs are CARPEM [28], Data Warehouse for Translational Research (DW4TR) [30], and @neurIST [48], which are dedicated to specific domains of research, namely cancer and cerebral aneurysm research.

Several types of images are part of modern IDRs, such as radiographic images in BTRIS [14] and document images in Methodist Environment for Translational Enhancement and Outcome Research (METEOR) [35]. In addition, medical reports are integrated in the IDRs. Clinical documents can be processed using natural language processing (NLP) algorithms to extract

clinical conditions, medication types, and other features from common hospital procedures, which increases their utility through transformation into structured data. NLP modules are integrated in CLB-IT [44], which is specifically built for text processing entries, as well as BTRIS [14], METEOR [35], and onco-i2b2 [43].

IDRs including CDSS include outcome data types, which are relevant for calculating risk factors or predictive values in clinical domains. External data can also be integrated into the IDRs, including genomics data from disease model organisms (BTRIS [14]), patients from external sources (BTRIS [14] and DW4TR [30]), or environmental indices and geolocation (MIDH) [27].

Standard Terminology

Health information technology uses controlled terminologies to condense the information to a set of codes that can be manipulated more easily and automatically in data processing. We observed the adoption of both common [272,273] and specialized terminologies (eg, Anatomical Therapeutic Chemical Classification [274], human phenotype ontology [275], Gene Ontology [276]). The most broadly used were International Classification of Diseases (ICD)-9 and 10 for the classification of diseases, systematized nomenclature of medicine-clinical terms (SNOMED-CT) for a variety of medical domains, Logical Observation Identifiers, Names, and Codes for laboratory observations, and current procedural terminology for common procedures (Table 1). These terminologies were utilized within the EHR and further integrated into the IDRs.

Common Data Model

A common data model (CDM) is a standard data schema that enables data interoperability and sharing. Contemporary data warehouses propose an analytical platform built around the CDM that provides all the software components to construct and manage the data in a CDM. A few different CDMs have been developed and adopted by the wider clinical research community, although some institutions still favor using a custom data schema tailored to their specific needs. In our study, a standard CDM was adopted by 18 of the 29 IDRs. The most frequently applied CDM, found in 16 instances, is Informatics for Integrating Biology and the Bedside (i2b2) [277]. METEOR [35] applies i2b2 with an expanded schema, and CARPEM [28] applies tranSMART [278], which is a framework layered on top of i2b2, dedicated to integrating omics data with EHR data. Another popular CDM that has been used more frequently in recent years is the Observational Medical Outcomes Partnership (OMOP) [279], adopted by 3 IDRs, namely MIDH [27], OpenFurther [46], and STARR [20]. OpenFurther uses OpenMRS [280], which is an open-source software and CDM that delivers health care in low- and middle-income countries. The BRP [41] is the only example using Harvest as their CDM.

Discussion

Principal Findings

Our review identified several institutions of various sizes and scopes that utilize an IDR. These IDRs contain data used for both research and clinical decision-making purposes. The use of structured data from natural language processing of clinical notes, clinical imaging, and omics data are the most recent big data types to be integrated with standard clinical observations. Owing to the large heterogeneity, however, integration is complex and tailored to the specific needs during the IDR implementation and maintenance, as ETL necessitates a significant effort in both the initial modeling and the ongoing updates.

As a novel contribution, we proposed and classified IDR architectures into 4 major models that highlight the processing and integration steps. The most common architecture model employs a staging layer implemented before the data are loaded into the data warehouse.

A set of common features are applied across most IDRs: IDRs commonly use standard terminologies such as ICD-9/10 and SNOMED-CT, which are often already part of the EHR data. Several IDRs use an open-source translational research framework to model their data, as described by Huser et al [12]. We observed extensive use of i2b2 CDM and the emergent adoption of OMOP CDM, which has the possibility to map additional domain-specific terminologies. Interestingly, PCORnet is one of the newest CDMs, but its application was not discussed in the sample of IDRs reviewed. The PCORnet is the most recently implemented CDM that borrows from several other CDMs and is organized around patient outcomes [261].

To safeguard the data in the IDR, data security and privacy need to be ensured from the initial steps of development. Data security is an important factor in all architecture types, with a particular need in collaborative projects that share data across jurisdictions. For example, in the general architecture of HSSC [29], data need to be stored in physically and logically secure facilities, where data management is extended to all the parties involved, and data need to be transmitted between the participating institutions through private high-speed networks. In the case of federated data warehouses, such as @neurIST [48], there is a tight control of data flow between different institutions and clinical and research domains, following policies aligned with recommendations from the Legal and Ethics Advisory Board. Privacy, referring to the protection of patient's personal information, emerged as an important feature, especially in the biobank-driven architecture; here, identifiable patient information is deleted from both the biosamples and the patient clinical data. Developers at the Children's Hospital of Philadelphia and the Children's Brain Tumor Tissue Consortium created an electronic Honest Broker (eHB) and Biorepository Portal (BRP) eHB [41], which provides a method for patient privacy protection by removing all the exposure of the research staff to patient identifiers and automating the deidentification process. Following a different privacy-preservation approach, Vanderbilt's Synthetic Derivative database [39] alters the patient data by obfuscating the true entries while preserving their time dependence.

Guiding Principles

The implementation of an IDR is subject to several factors that must be considered before development. We identified 2 major factors: (1) the data stored in the IDR and (2) the scope of the IDR, either being exclusively used for research purposes or in combination with clinical or operational purposes, as shown in the general and biobank-driven architecture models. Data types, heterogeneity, and volume greatly influence system load, update, and query of the database. The scope of the IDR influences its primary end users, researchers, clinical users, or operational users, who have different needs and, thus, need access to different sets of tools to extract, analyze, and visualize the data. All the features influence both the data latency and the data synchronization, which are major elements in the model architecture. Moreover, available funding plays an important role in architecture decisions, as are considerations for future expansions.

XSL•FO

RenderX

Among the set of selected IDRs, we observed a number of collaborative projects that work within specialized medical domains, such as cancer or pediatrics. Collaborative IDRs are likely to integrate their data to increase the number of patients, thus increasing the statistical power of their respective cohorts.

On the basis of our analysis, we highlight the following guiding principles for small- to medium-sized institutions planning to implement an IDR:

1. The general architecture model, with or without CDSS, is the most straightforward to implement; the data staging layer facilitates ETL and data processing before loading into the data warehouse.

2. Select a standard CDM already in use by other institutions; both i2b2 and OMOP provide server and client services in a single unique platform that serves the user with all the necessary tools to set up a structured IDR.

3. Wherever possible, adopt standard terminologies; we listed the most common terminologies derived from the integration with EHR data (Table 1). One promising approach is that common terminologies are applied in the first phases of the IDR development with other, more specialized terminologies, added later as the project scope expands.

4. Finally, the data update requirements and ETL process design should be carefully considered, the level of automation, as these are the limiting stages in data integration and update.

Commercial electronic medical record platforms such as Epic, Cerner, Meditech, and Allscripts are dominant in large institutions. However, although some information about how to query underlying databases and application programming interfaces to communicate with these systems are available, little information on transforming such data into IDR is available in the literature, most likely because of their proprietary nature. Most vendors also sell tools for analysts to query and make use of data from these clinical production systems; however, they are not IDRs themselves and are not targeted toward secondary use for research.

As for lessons learned in the field, Epstein et al [281] demonstrate the feasibility of transferring the development of a perioperative data warehouse (schemas and processes) built on top of Epic's database from one institution to another.

Comparison With Prior Work

In their review, Hamoud et al [13] provided general requirements for building a successful clinical data warehouse, recommending a top-down approach to the initial stages of development. They recommended considering all the individual components of the final system to decrease integration obstacles when dealing with heterogeneous data sources.

Three major factors contributing to the success of IDRs were identified by Baghal [231] when developing their in-house IDR: (1) organizational, enhancing the collaboration between different departments and researchers; (2) behavioral, building new professional relationships through frequent meetings and communication between departments; and (3) technical improvements to deploy new self-service tools that empower

researchers. Collectively, these factors increase the utility and adoption of IDRs in clinical research.

In addition, the report by Rizi and Roudsari [282] on lessons and barriers from their development of a public health data warehouse, which IDR developers might want to consider, specifically, not to underestimate technical challenges such as those related to extracting data from other systems, difficulties in modeling and mapping of data, as well as data security and privacy. Other considerations include leveraging the IDR to improve data quality at the source, implementing a data governance framework from the beginning, and ensuring that key organizational stakeholders endorse the project early and strongly [282].

Limitations

Our search was not intended to be a systematic search; therefore, we may have missed some articles. An example of missing articles is those describing raw and unstructured data repositories, also referred to as *data lakes*, as these did not appear in our search results although we knew they exist. One of the *data lakes* was presented by Foran et al [207] as a file reservoir, integrated in the data warehouse schema. For researchers to access those data, it was necessary to use a *feeder* database before their upload to the final data warehouse.

Furthermore, we were able to report on the IDRs and IDR features described in the literature, possibly omitting smaller institutions that are not actively publishing in peer-reviewed journals. In an attempt to mitigate this issue, we searched the representative institutional websites to retrieve additional details about the IDR architectures. As shown in Multimedia Appendix 1 [283-288]: A2, several organizations provide further details about their architecture in GitHub repositories or institutional Wiki pages, which can be explored for additional information besides the published literature.

This review includes articles and web resources shortlisted according to aspects of the IDR architectures that were considered relevant. Providing an exhaustive coverage of all aspects of IDR implementation, such as tools designed to interact with the IDR, are better left for a dedicated review. An example of such tools is the *Green Button* project, which provides critical help in treating patients [289-292]. Examples of CDM-based tools, built around an application, are the @neurIST platform [48], @neurLink, and @neurFuse application suites that consist of research-oriented modules dedicated to knowledge discovery and image processing. CDSS tools such as Green Button, @neurIST applications, or many other existing frameworks are essential in providing sophisticated analyses to support clinicians, but are beyond the scope of our review.

Conclusions

There is significant potential in the implementation of IDRs in health institutions, and their importance is evident from the growing number of projects developed in the past 10 years. Despite the common trends in IDR implementation observed in this study, there are also many variations. There are 2 major design factors, namely data heterogeneity and IDR scope, which

XSL•FO

RenderX

need to be carefully considered before embarking on the IDR design and planning process.

Finally, we aim to apply the knowledge presented in this study for the implementation of a pediatric IDR at our institution. By sharing our experience of planning and designing our IDR with those joining the field or planning to implement an IDR for research purposes, we hope to contribute to future IDR endeavors.

Acknowledgments

Conflicts of Interest

None declared.

Multimedia Appendix 1

Supplementary methods and results.
[PDF File (Adobe PDF File), 1649 KB-Multimedia Appendix 1]

References

1. Adler-Milstein J, Holmgren AJ, Kralovec P, Worzala C, Searcy T, Patel V. Electronic health record adoption in US hospitals: the emergence of a digital 'advanced use' divide. J Am Med Inform Assoc 2017 Nov 1;24(6):1142-1148. [doi: 10.1093/jamia/ocx080] [Medline: 29016973]
2. Lau F, Price M, Boyd J, Partridge C, Bell H, Raworth R. Impact of electronic medical record on physician practice in office settings: a systematic review. BMC Med Inform Decis Mak 2012 Feb 24;12:10 [FREE Full text] [doi: 10.1186/1472-6947-12-10] [Medline: 22364529]
3. Schoen C, Osborn R, Doty MM, Squires D, Peugh J, Applebaum S. A survey of primary care physicians in eleven countries, 2009: perspectives on care, costs, and experiences. Health Aff (Millwood) 2009;28(6):w1171-w1183. [doi: 10.1377/hlthaff.28.6.w1171] [Medline: 19884491]
4. MacKenzie SL, Wyatt MC, Schuff R, Tenenbaum JD, Anderson N. Practices and perspectives on building integrated data repositories: results from a 2010 CTSA survey. J Am Med Inform Assoc 2012 Jun;19(e1):e119-e124 [FREE Full text] [doi: 10.1136/amiajnl-2011-000508] [Medline: 22437072]
5. Anderson N, Abend A, Mandel A, Geraghty E, Gabriel D, Wynden R, et al. Implementation of a deidentified federated data network for population-based cohort discovery. J Am Med Inform Assoc 2012 Jun;19(e1):e60-e67 [FREE Full text] [doi: 10.1136/amiajnl-2011-000133] [Medline: 21873473]
6. Meystre SM, Lovis C, Bürkle T, Tognola G, Budrionis A, Lehmann CU. Clinical data reuse or secondary use: current status and potential future progress. Yearb Med Inform 2017 Aug;26(1):38-52 [FREE Full text] [doi: 10.15265/IY-2017-007] [Medline: 28480475]
7. Murphy SN, Dubey A, Embi PJ, Harris PA, Richter BG, Turisco F, et al. Current state of information technologies for the clinical research enterprise across academic medical centers. Clin Transl Sci 2012 Jun;5(3):281-284 [FREE Full text] [doi: 10.1111/j.1752-8062.2011.00387.x] [Medline: 22686207]
8. Kawamoto K, Houlihan CA, Balas EA, Lobach DF. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. Br Med J 2005 Apr 2;330(7494):765 [FREE Full text] [doi: 10.1136/bmj.38398.500764.8F] [Medline: 15767266]
9. Buntin MB, Burke MF, Hoaglin MC, Blumenthal D. The benefits of health information technology: a review of the recent literature shows predominantly positive results. Health Aff (Millwood) 2011 Mar;30(3):464-471. [doi: 10.1377/hlthaff.2011.0178] [Medline: 21383365]
10. Belard A, Buchman T, Forsberg J, Potter BK, Dente CJ, Kirk A, et al. Precision diagnosis: a view of the clinical decision support systems (CDSS) landscape through the lens of critical care. J Clin Monit Comput 2017 Apr;31(2):261-271. [doi: 10.1007/s10877-016-9849-1] [Medline: 26902081]
11. Shin S, Kim WS, Lee J. Characteristics desired in clinical data warehouse for biomedical research. Healthc Inform Res 2014 Apr;20(2):109-116 [FREE Full text] [doi: 10.4258/hir.2014.20.2.109] [Medline: 24872909]
12. Huser V, Cimino JJ. Desiderata for healthcare integrated data repositories based on architectural comparison of three public repositories. AMIA Annu Symp Proc 2013;2013:648-656 [FREE Full text] [Medline: 24551366]
13. Hamoud A, Hashim A, Awadh W. Clinical data warehouse: a review. Iraqi J Comput Inf 2018 Dec 31;44(2):16-26. [doi: 10.25195/ijci.v44i2.53]
14. Cimino JJ, Ayres EJ, Remennik L, Rath S, Freedman R, Beri A, et al. The national institutes of health's biomedical translational research information system (BTRIS): design, contents, functionality and experience to date. J Biomed Inform 2014 Dec;52:11-27 [FREE Full text] [doi: 10.1016/j.jbi.2013.11.004] [Medline: 24262893]

XSL•FO
RenderX

15.   Huser V, Kayaalp M, Dodd ZA, Cimino JJ. Piloting a deceased subject integrated data repository and protecting privacy of relatives. AMIA Annu Symp Proc 2014;2014:719-728 [FREE Full text] [Medline: 25954378]

16.   Liu M, Melton BL, Ator G, Waitman LR. Integrating medication alert data into a clinical data repository to enable retrospective study of drug interaction alerts in clinical practice. AMIA Jt Summits Transl Sci Proc 2017;2017:213-220 [FREE Full text] [Medline: 28815131]

17.   Adagarla B, Connolly DW, McMahon TM, Nair M, VanHoose LD, Sharma P, et al. SEINE: Methods for Electronic Data Capture and Integrated Data Repository Synthesis with Patient Registry Use Cases. KU ScholarWorks. 2015. URL: http://hdl.handle.net/2271/1303 [accessed 2020-08-09]

18.   Lowe HJ, Ferris TA, Hernandez PM, Weber SC. STRIDE--an integrated standards-based translational research informatics platform. AMIA Annu Symp Proc 2009 Nov 14;2009:391-395 [FREE Full text] [Medline: 20351886]

19.   Hernandez P, Podchiyska T, Weber S, Ferris T, Lowe H. Automated mapping of pharmacy orders from two electronic health record systems to RxNorm within the STRIDE clinical data warehouse. AMIA Annu Symp Proc 2009 Nov 14;2009:244-248 [FREE Full text] [Medline: 20351858]

20.   STARR Tools COVID-19 Research Support. Stanford Medicine. 2019. URL: http://med.stanford.edu/starr-tools.html [accessed 2020-07-10]

21.   Jannot A, Zapletal E, Avillach P, Mamzer M, Burgun A, Degoulet P. The Georges Pompidou university hospital clinical data warehouse: a 8-years follow-up experience. Int J Med Inform 2017 Jun;102:21-28. [doi: 10.1016/j.ijmedinf.2017.02.006] [Medline: 28495345]

22.   Boussadi A, Caruba T, Zapletal E, Sabatier B, Durieux P, Degoulet P. A clinical data warehouse-based process for refining medication orders alerts. J Am Med Inform Assoc 2012;19(5):782-785 [FREE Full text] [doi: 10.1136/amiajnl-2012-000850] [Medline: 22523345]

23.   Zapletal E, Rodon N, Grabar N, Degoulet P. Methodology of integration of a clinical data warehouse with a clinical information system: the HEGP case. Stud Health Technol Inform 2010;160(Pt 1):193-197. [Medline: 20841676]

24.   Haarbrandt B, Tute E, Marschollek M. Automated population of an i2b2 clinical data warehouse from an openEHR-based data repository. J Biomed Inform 2016 Oct;63:277-294. [doi: 10.1016/j.jbi.2016.08.007] [Medline: 27507090]

25.   Zunner C, Ganslandt T, Prokosch H, Bürkle T. A reference architecture for semantic interoperability and its practical application. Stud Health Technol Inform 2014;198:40-46. [Medline: 24825683]

26.   Choi IY, Park S, Park B, Chung BH, Kim C, Lee HM, et al. Development of prostate cancer research database with the clinical data warehouse technology for direct linkage with electronic medical record system. Prostate Int 2013;1(2):59-64 [FREE Full text] [doi: 10.12954/PI.12015] [Medline: 24223403]

27.   Hall ES, Greenberg JM, Muglia LJ, Divekar P, Zahner J, Gholap J, et al. Implementation of a regional perinatal data repository from clinical and billing records. Matern Child Health J 2018 Apr;22(4):485-493 [FREE Full text] [doi: 10.1007/s10995-017-2414-9] [Medline: 29275460]

28.   Rance B, Canuel V, Countouris H, Laurent-Puig P, Burgun A. Integrating heterogeneous biomedical data for cancer research: the CARPEM infrastructure. Appl Clin Inform 2016;7(2):260-274 [FREE Full text] [doi: 10.4338/ACI-2015-09-RA-0125] [Medline: 27437039]

29.   Turley CB, Obeid J, Larsen R, Fryar KM, Lenert L, Bjorn A, et al. Leveraging a statewide clinical data warehouse to expand boundaries of the learning health system. EGEMS (Wash DC) 2016;4(1):1245 [FREE Full text] [doi: 10.13063/2327-9214.1245] [Medline: 28154834]

30.   Hu H, Correll M, Kvecher L, Osmond M, Clark J, Bekhash A, et al. DW4TR: a data warehouse for translational research. J Biomed Inform 2011 Dec;44(6):1004-1019 [FREE Full text] [doi: 10.1016/j.jbi.2011.08.003] [Medline: 21872681]

31.   Maskery S, Bekhash A, Kvecher L, Correll M, Hooke JA, Kovatich AJ, et al. Aggregated biomedical-information browser (ABB): a graphical user interface for clinicians and scientists to access a clinical data warehouse. J Comput Sci Syst Biol 2014;7:20-27. [doi: 10.4172/jcsb.1000134]

32.   Rajeevan N, Niehoff KM, Charpentier P, Levin FL, Justice A, Brandt CA, et al. Utilizing patient data from the veterans administration electronic health record to support web-based clinical decision support: informatics challenges and issues from three clinical domains. BMC Med Inform Decis Mak 2017 Jul 19;17(1):111 [FREE Full text] [doi: 10.1186/s12911-017-0501-x] [Medline: 28724368]

33.   Dagliati A, Sacchi L, Tibollo V, Cogni G, Teliti M, Martinez-Millana A, et al. A dashboard-based system for supporting diabetes care. J Am Med Inform Assoc 2018 May 1;25(5):538-547. [doi: 10.1093/jamia/ocx159] [Medline: 29409033]

34.   Yu J, Mao H, Li M, Ye D, Zhao D. CSDC: a nationwide screening platform for stroke control and prevention in China. Conf Proc IEEE Eng Med Biol Soc 2016 Aug;2016:2974-2977. [doi: 10.1109/EMBC.2016.7591354] [Medline: 28268937]

35.   Puppala M, He T, Yu X, Chen S, Ogunti R, Wong S. Data Security and Privacy Management in Healthcare Applications and Clinical Data Warehouse Environment. In: IEEE-EMBS International Conference on Biomedical and Health Informatics. 2016 Presented at: BHI'16; February 24-27, 2016; Las Vegas, NV, USA. [doi: 10.1109/bhi.2016.7455821]

36.   Chute CG, Beck SA, Fisk TB, Mohr DN. The enterprise data trust at mayo clinic: a semantically integrated warehouse of biomedical data. J Am Med Inform Assoc 2010;17(2):131-135 [FREE Full text] [doi: 10.1136/jamia.2009.002691] [Medline: 20190054]

37. Hong N, Li Z, Kiefer R, Robertson MS, Goode EL, Wang C, et al. Building an i2b2-Based Integrated Data Repository for Cancer Research: A Case Study of Ovarian Cancer Registry. In: Lecture Notes in Computer Science. 2017 Presented at: LNCS'17; September 2-6, 2017; New Delhi, India. [doi: 10.1007/978-3-319-57741-8_8]

38. Horton I, Lin Y, Reed G, Wiepert M, Hart S. Empowering mayo clinic individualized medicine with genomic data warehousing. J Pers Med 2017 Aug 22;7(3):- [FREE Full text] [doi: 10.3390/jpm7030007] [Medline: 28829408]

39. Danciu I, Cowan JD, Basford M, Wang X, Saip A, Osgood S, et al. Secondary use of clinical data: the Vanderbilt approach. J Biomed Inform 2014 Dec;52:28-35 [FREE Full text] [doi: 10.1016/j.jbi.2014.02.003] [Medline: 24534443]

40. Roden DM, Pulley JM, Basford MA, Bernard GR, Clayton EW, Balser JR, et al. Development of a large-scale de-identified DNA biobank to enable personalized medicine. Clin Pharmacol Ther 2008 Sep;84(3):362-369 [FREE Full text] [doi: 10.1038/clpt.2008.89] [Medline: 18500243]

41. Felmeister AS, Masino AJ, Rivera TJ, Resnick AC, Pennington JW. The biorepository portal toolkit: an honest brokered, modular service oriented software tool set for biospecimen-driven translational research. BMC Genomics 2016 Aug 18;17(Suppl 4):434 [FREE Full text] [doi: 10.1186/s12864-016-2797-9] [Medline: 27535360]

42. Ferretti Y, Miyoshi NS, Silva WA, Felipe JC. BioBankWarden: a web-based system to support translational cancer research by managing clinical and biomaterial data. Comput Biol Med 2017 May 1;84:254-261 [FREE Full text] [doi: 10.1016/j.compbiomed.2015.04.008] [Medline: 25959800]

43. Segagni D, Tibollo V, Dagliati A, Zambelli A, Priori SG, Bellazzi R. An ICT infrastructure to integrate clinical and molecular data in oncology research. BMC Bioinformatics 2012 Mar 28;13(Suppl 4):S5 [FREE Full text] [doi: 10.1186/1471-2105-13-S4-S5] [Medline: 22536972]

44. Biron P, Metzger MH, Pezet C, Sebban C, Barthuet E, Durand T. An information retrieval system for computerized patient records in the context of a daily hospital practice: the example of the Léon Bérard cancer center (France). Appl Clin Inform 2014;5(1):191-205 [FREE Full text] [doi: 10.4338/ACI-2013-08-CR-0065] [Medline: 24734133]

45. Livne OE, Schultz ND, Narus SP. Federated querying architecture with clinical & translational health IT application. J Med Syst 2011 Oct;35(5):1211-1224. [doi: 10.1007/s10916-011-9720-3] [Medline: 21537849]

46. OpenFurther. 2019. URL: http://openfurther.org/ [accessed 2019-08-10]

47. Narus SP, Srivastava R, Gouripeddi R, Livne OE, Mo P, Bickel JP, et al. Federating clinical data from six pediatric hospitals: process and initial results from the PHIS+ consortium. AMIA Annu Symp Proc 2011;2011:994-1003 [FREE Full text] [Medline: 22195159]

48. Benkner S, Arbona A, Berti G, Chiarini A, Dunlop R, Engelbrecht G, et al. NeurIST: infrastructure for advanced disease management through integration of heterogeneous data, computing, and complex processing services. IEEE Trans Inf Technol Biomed 2010 Nov;14(6):1365-1377. [doi: 10.1109/TITB.2010.2049268] [Medline: 20435543]

49. Ulrich H, Kock A, Duhm-Harbeck P, Habermann JK, Ingenerf J. Metadata repository for improved data sharing and reuse based on HL7 FHIR. Stud Health Technol Inform 2016;228:162-166. [Medline: 27577363]

50. Tkaczyk D, Szostek P, Fedoryszak M, Dendek PJ, Bolikowski A. CERMINE: automatic extraction of structured metadata from scientific literature. Int J Doc Anal Recognit 2015 Jul 3;18(4):317-335. [doi: 10.1007/s10032-015-0249-8]

51. Conway JR, Lex A, Gehlenborg N. UpSetR: an R package for the visualization of intersecting sets and their properties. Bioinformatics 2017 Sep 15;33(18):2938-2940 [FREE Full text] [doi: 10.1093/bioinformatics/btx364] [Medline: 28645171]

52. Bortis G. Experiences With Mirth: an Open Source Health Care Integration Engine. In: Proceedings of the 30th International Conference on Software Engineering. 2008 Presented at: ICSE'08; May 10-18, 2008; Leipzig, Germany. [doi: 10.1145/1368088.1368179]

53. Kang B, Kim D, Kim H. Two-Phase chief complaint mapping to the UMLS metathesaurus in Korean electronic medical records. IEEE Trans Inf Technol Biomed 2009 Jan;13(1):78-86. [doi: 10.1109/TITB.2008.2007103] [Medline: 19129026]

54. Lhotska L, Aubrecht P, Valls A, Gibert K. Security Recommendations for Implementation in Distributed Healthcare Systems. In: 42nd Annual IEEE International Carnahan Conference on Security Technology. 2008 Presented at: CCST'08; October 13-16, 2008; Prague, Czech Republic. [doi: 10.1109/ccst.2008.4751280]

55. Maragoudakis M, Lymberopoulos D, Fakotakis N, Spiropoulos K. A hierarchical, ontology-driven Bayesian concept for ubiquitous medical environments--a case study for pulmonary diseases. Conf Proc IEEE Eng Med Biol Soc 2008;2008:3807-3810. [doi: 10.1109/IEMBS.2008.4650038] [Medline: 19163541]

56. Pruulmann J, Willemson J. Implementing A Knowledge-Driven Hierarchical Context Model in a Medical Laboratory Information System. In: The Third International Multi-Conference on Computing in the Global Information Technology (iccgi 2008). 2008 Presented at: ICCGI'08; July 27-August 1, 2008; Athens, Greece. [doi: 10.1109/iccgi.2008.19]

57. Riedl B, Grascher V, Fenz S, Neubauer T. Pseudonymization for Improving the Privacy in E-Health Applications. In: Proceedings of the 41st Annual Hawaii International Conference on System Sciences. 2008 Presented at: HICSS'08; January 7-10, 2008; Waikoloa, HI, USA. [doi: 10.1109/hicss.2008.366]

58. Sung T, Hung F, Chiu H. Implementation of an integrated drug information system for inpatients to reduce medication errors in administering stage. Conf Proc IEEE Eng Med Biol Soc 2008;2008:743-746. [doi: 10.1109/IEMBS.2008.4649259] [Medline: 19162762]

59. Taddei A, Dalmiani S, Vellani A, Rocca E, Piccini G, Carducci T, et al. Data Integration in Cardiac Surgery Health Care Institution: Experience at G Pasquinucci Heart Hospital. In: Computers in Cardiology. 2008 Presented at: CIC'08; September 14-17, 2008; Bologna, Italy. [doi: 10.1109/CIC.2008.4749034]

60. Zamboulis L, Poulovassilis A, Roussos G. Flexible Data Integration and Ontology-Based Data Access to Medical Records. In: 8th IEEE International Conference on BioInformatics and BioEngineering. 2008 Presented at: BIBE'08; October 8-10, 2008; Athens, Greece. [doi: 10.1109/bibe.2008.4696774]

61. Agorastos T, Koutkias V, Falelakis M, Lekka I, Mikos T, Delopoulos A, et al. Semantic integration of cervical cancer data repositories to facilitate multicenter association studies: the ASSIST approach. Cancer Inform 2009 Feb 3;8:31-44 [FREE Full text] [doi: 10.4137/cin.s963] [Medline: 19458792]

62. Amoretti M, Zanichelli F. The Multi-Knowledge Service-Oriented Architecture: Enabling Collaborative Research for E-Health. In: 42nd Hawaii International Conference on System Sciences. 2009 Presented at: HICSS'09; January 5-8, 2009; Big Island, HI, USA. [doi: 10.1109/hicss.2009.441]

63. Archer N, Cocosila M. Improving EMR System Adoption in Canadian Medical Practice: A Research Model. In: World Congress on Privacy, Security, Trust and the Management of e-Business. 2009 Presented at: CONGRESS'09; August 25-27, 2009; Saint John, NB, Canada. [doi: 10.1109/congress.2009.15]

64. Bradshaw RL, Matney S, Livne OE, Bray BE, Mitchell JA, Narus SP. Architecture of a federated query engine for heterogeneous resources. AMIA Annu Symp Proc 2009 Nov 14;2009:70-74 [FREE Full text] [Medline: 20351825]

65. Chung P, Afzal F, Hsiao H. A Software System Development for Probabilistic Relational Database Applications for Biomedical Informatics. In: International Conference on Advanced Information Networking and Applications Workshops. 2009 Presented at: WAINA'09; May 26-29, 2009; Bradford, UK. [doi: 10.1109/waina.2009.99]

66. Costanzo D. Biomedical Data Acquisition and Processing in the Decision Support Services of HEARTFAID Platform. In: International Workshop on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications. 2009 Presented at: IDAACS'09; September 21-23, 2009; Rende, Italy. [doi: 10.1109/idaacs.2009.5342976]

67. Dong J, Zhou D, Hu X, Zhang Z, Jiang K. Analysis and Design on Standard System of Electronic Health Records. In: Proceedings of the 2009 First International Workshop on Education Technology and Computer Science - Volume 01. 2009 Presented at: ETCS'09; March 7-8, 2009; Italy. [doi: 10.1109/etcs.2009.223]

68. Milano F, Eijo J, Gomez A, de Quiros F, Risk M. MedSiGRe: Medical Signal Grid Repository, an Integration to Italica Project. In: Latin American Network Operations and Management Symposium. 2009 Presented at: LANOMS;09; October 19-21, 2009; Punta del Este, Uruguay. [doi: 10.1109/lanoms.2009.5338792]

69. Ongenae F, Dupont T, Kerckhove W. Design of ICU Medical Decision Support Applications by Integrating Service Oriented Applications With a Rule-Based System. In: 2nd International Symposium on Applied Sciences in Biomedical and Communication Technologies. 2009 Presented at: ISABEL'09; January 8, 2009; Bratislava, Slovakia. [doi: 10.1109/isabel.2009.5373663]

70. Patra D, Ray S, Mukhopadhyay J, Majumdar B, Majumdar A. Achieving E-Health Care in a Distributed EHR System. In: 11th Internation11th International Conference on e-Health Networking, Applications and Services al Conference on e-Health Networking, Applications and Services. 2009 Presented at: Healthcom'09; December 16-18, 2009; Sydney, NSW, Australia. [doi: 10.1109/health.2009.5406205]

71. Rusu M, Saplacan G, Sebestyen G. Distributed e-Health system with Smart Self-Care Units. In: 5th International Conference on Intelligent Computer Communication and Processing. 2009 Presented at: ICCP'09; August 27-29, 2009; Cluj-Napoca, Romania. [doi: 10.1109/iccp.2009.5284744]

72. Siddiqi J, Akhgar B, Rahman F. Towards an Integrated Platform for Improving Hospital Risk Management. In: Sixth International Conference on Information Technology: New Generations. 2009 Presented at: IRNG'09; April 27-29, 2009; Las Vegas, NV, USA. [doi: 10.1109/itng.2009.204]

73. Wah T, Sim O. Development of a data warehouse for lymphoma cancer diagnosis and treatment decision support. WSEAS Transactions on Information Science and Applications 2009;6(3):530-543. [doi: 10.37394/23209]

74. Wu F, Williams M, Kazanzides P, Brady K, Fackler J. A Modular Clinical Decision Support System Clinical Prototype Extensible Into Multiple Clinical Settings. In: 3rd International Conference on Pervasive Computing Technologies for Healthcare. 2009 Presented at: Pervasive Health'09; April 1-3, 2009; London, UK. [doi: 10.4108/icst.pervasivehealth2009.6078]

75. Yang L, Tuzel O, Chen W, Meer P, Salaru G, Goodell LA, et al. PathMiner: a web-based tool for computer-assisted diagnostics in pathology. IEEE Trans Inf Technol Biomed 2009 May;13(3):291-299 [FREE Full text] [doi: 10.1109/TITB.2008.2008801] [Medline: 19171530]

76. Yu J. Distributed Data Processing Framework for Oral Health Care Information Management Based on CSCWD Technology. In: First International Conference on Information Science and Engineering. 2009 Presented at: ICISE'09; December 26-28, 2009; Nanjing, China. [doi: 10.1109/icise.2009.511]

77. Ceusters W, Smith B. A unified framework for biomedical terminologies and ontologies. Stud Health Technol Inform 2010;160(Pt 2):1050-1054 [FREE Full text] [Medline: 20841844]

78. Chen P, Freg C, Hou T, Teng W. Implementing RAID-3 on Cloud Storage for EMR System. In: International Computer Symposium. 2010 Presented at: ICI'10; December 16-18, 2010; Tainan, Taiwan. [doi: 10.1109/compsym.2010.5685395]

79. Couderc J. A unique digital electrocardiographic repository for the development of quantitative electrocardiography and cardiac safety: the Telemetric and Holter ECG Warehouse (THEW). J Electrocardiol 2010;43(6):595-600 [FREE Full text] [doi: 10.1016/j.jelectrocard.2010.07.015] [Medline: 20863512]

80. Dangl A, Demiroglu SY, Gaedcke J, Helbing K, Jo P, Rakebrandt F, et al. The IT-infrastructure of a biobank for an academic medical center. Stud Health Technol Inform 2010;160(Pt 2):1334-1338. [Medline: 20841901]

81. Duennebeil S, Sunyaev A, Leimeister J, Krcmar H. Strategies for Development and Adoption of EHR in German Ambulatory Care. In: 4th International Conference on Pervasive Computing Technologies for Healthcare. 2010 Presented at: PervasiveHealth'10; March 22-25, 2010; Munich, Germany. [doi: 10.4108/icst.pervasivehealth2010.8887]

82. El Fadly A, Lucas N, Rance B, Verplancke P, Lastic P, Daniel C. The REUSE project: EHR as single datasource for biomedical research. Stud Health Technol Inform 2010;160(Pt 2):1324-1328. [Medline: 20841899]

83. Frank L, Andersen S. Evaluation of Different Database Designs for Integration of Heterogeneous Distributed Electronic Health Records. In: International Conference on Complex Medical Engineering. 2010 Presented at: ICCME'10; July 13-15, 2010; Gold Coast, QLD, Australia. [doi: 10.1109/iccme.2010.5558844]

84. Frize M, Bariciak E, Weyand S. Suggested Criteria for Successful Deployment of a Clinical Decision Support System (CDSS). In: International Workshop on Medical Measurements and Applications. 2010 Presented at: MEMEA'10; April 30-May 1, 2010; Ottawa, ON, Canada. [doi: 10.1109/memea.2010.5480227]

85. Jiang L, Cai H, Xu B. A Domain Ontology Approach in the ETL Process of Data Warehousing. In: 7th International Conference on E-Business Engineering. 2010 Presented at: ICEBE'10; November 10-12, 2010; Shanghai, China. [doi: 10.1109/icebe.2010.36]

86. Kataria P, Juric R. Sharing E-health Information Through Ontological Layering. In: 43rd Hawaii International Conference on System Sciences. 2010 Presented at: HICSS'10; January 5-8, 2010; Honolulu, HI, USA. [doi: 10.1109/hicss.2010.338]

87. Kildea J, Evans M, Parker W. A Framework for Comprehensive Electronic QA in Radiation Therapy. In: Ninth International Conference on Machine Learning and Applications. 2010 Presented at: ICMLA'10; December 10-14, 2010; Washington, DC, USA. [doi: 10.1109/icmla.2010.157]

88. Mougiakakou SG, Bartsocas CS, Bozas E, Chaniotakis N, Iliopoulou D, Kouris I, et al. SMARTDIAB: a communication and information technology approach for the intelligent monitoring, management and follow-up of type 1 diabetes patients. IEEE Trans Inf Technol Biomed 2010 May;14(3):622-633. [doi: 10.1109/TITB.2009.2039711] [Medline: 20123578]

89. Payne PR, Borlawsky TB, Stephens W, Barrett MC, Nguyen-Pham T, Greaves AW. The TRITON project: design and implementation of an integrative translational research information management platform. AMIA Annu Symp Proc 2010 Nov 13;2010:617-621 [FREE Full text] [Medline: 21347052]

90. Ping H, Xin-Lei W. Health Information System Grid Based on WSRF. In: Second International Conference on Information Technology and Computer Science. 2010 Presented at: ITCS'10; July 24-25, 2010; Kiev, Ukraine. [doi: 10.1109/itcs.2010.133]

91. Podchiyska T, Hernandez P, Ferris T, Weber S, Lowe HJ. Managing medical vocabulary updates in a clinical data warehouse: an RxNorm case study. AMIA Annu Symp Proc 2010 Nov 13;2010:477-481 [FREE Full text] [Medline: 21347024]

92. Roelofs E, Persoon L, Qamhiyeh S, Verhaegen F, de Ruysscher D, Scholz M, et al. Design of and technical challenges involved in a framework for multicentric radiotherapy treatment planning studies. Radiother Oncol 2010 Dec;97(3):567-571. [doi: 10.1016/j.radonc.2010.08.009] [Medline: 20864198]

93. Sachdeva S, Mchome S, Bhalla S. Web Services Security Issues in Healthcare Applications. In: 9th International Conference on Computer and Information Science. 2010 Presented at: ICIS'10; August 18-20, 2010; Yamagata, Japan. [doi: 10.1109/icis.2010.134]

94. Spitzer AR, Ellsbury DL, Handler D, Clark RH. The pediatrix babysteps data warehouse and the pediatrix qualitysteps improvement project system--tools for 'meaningful use' in continuous quality improvement. Clin Perinatol 2010 Mar;37(1):49-70. [doi: 10.1016/j.clp.2010.01.016] [Medline: 20363447]

95. Tohouri R, Asangansi I, Titlestad O, Braa J. The Change Strategy Towards an Integrated Health Information Infrastructure: Lessons From Sierra Leone. In: 43rd Hawaii International Conference on System Sciences. 2010 Presented at: HICSS'10; January 5-8, 2010; Honolulu, HI, USA. [doi: 10.1109/hicss.2010.367]

96. Vcelák P, Klecková J, Rohan V. Cerebrovascular Diseases Research Database. In: 3rd International Conference on Biomedical Engineering and Informatics. 2010 Presented at: BMEI'10; October 16-18, 2010; Yantai, China. [doi: 10.1109/bmei.2010.5639845]

97. Yaowen Z, Wei X, Yuwan H. Cerebrovascular Diseases Research Databaseresearch on Healthcare Integrating Model of Medical Information System Based on Agent. In: International Conference on Computational and Information Sciences. 2010 Presented at: ICISS'10; December 17-19, 2010; Chengdu, China. [doi: 10.1109/iccis.2010.157]

98. Zhao J, Wang T. A General Framework for Medical Data Mining. In: International Conference on Future Information Technology and Management Engineering. 2010 Presented at: FITME'10; October 9-10, 2010; Changzhou, China. [doi: 10.1109/fitme.2010.5654718]

99. Zhou X, Chen S, Liu B, Zhang R, Wang Y, Li P, et al. Development of traditional Chinese medicine clinical data warehouse for medical knowledge discovery and decision support. Artif Intell Med 2010;48(2-3):139-152. [doi: 10.1016/j.artmed.2009.07.012] [Medline: 20122820]

100. Apte M, Neidell M, Furuya EY, Caplan D, Glied S, Larson E. Using electronically available inpatient hospital data for research. Clin Transl Sci 2011 Oct;4(5):338-345 [FREE Full text] [doi: 10.1111/j.1752-8062.2011.00353.x] [Medline: 22029805]

101. Chazard E, Băceanu A, Ferret L, Ficheur G. The ADE scorecards: a tool for adverse drug event detection in electronic health records. Stud Health Technol Inform 2011;166:169-179. [Medline: 21685622]

102. Cossu M, Furfori P, Taddei A, Mangione M, del Sarto P. Anesthesia Information Management System in Cardiac Surgery. In: Computing in Cardiology. 2011 Presented at: CIC'11; September 8-11, 2011; Hangzhou, China URL: https://ieeexplore.ieee.org/document/6164631

103. Crichton D, Mattmann C, Hart A. An Informatics Architecture for the Virtual Pediatric Intensive Care Unit. In: 24th International Symposium on Computer-Based Medical Systems. 2011 Presented at: CBMS'11; June 27-30, 2011; Bristol, UK. [doi: 10.1109/cbms.2011.5999031]

104. Cuggia M, Garcelon N, Campillo-Gimenez B, Bernicot T, Laurent J, Garin E, et al. Roogle: an information retrieval engine for clinical data warehouse. Stud Health Technol Inform 2011;169:584-588. [Medline: 21893816]

105. Hatakeyama Y, Kataoka H, Nakajima N, Watabe T, Okuhara Y, Sagara Y. An Education Support System with Anonymized Medical Data Based on Thin Client System. In: International Conference on Internet of Things and 4th International Conference on Cyber, Physical and Social Computing. 2011 Presented at: CPSCom'11; October 19-22, 2011; Dalian, China. [doi: 10.1109/ithings/cpscom.2011.37]

106. Kanagaraj G, Sumathi A. Proposal of an Open-source Cloud Computing System for Exchanging Medical Images of a Hospital Information System. In: 3rd International Conference on Trendz in Information Sciences & Computing. 2011 Presented at: TICS'11; December 8-9, 2011; Chennai, India. [doi: 10.1109/tisc.2011.6169102]

107. Kiong Y, Palaniappan S, Yahaya N. Health Ontology System. In: 7th International Conference on Information Technology in Asia. 2011 Presented at: CITA'11; July 12-13, 2011; Kuching, Sarawak, Malaysia. [doi: 10.1109/cita.2011.5999506]

108. Murphy SN, Gainer V, Mendis M, Churchill S, Kohane I. Strategies for maintaining patient privacy in i2b2. J Am Med Inform Assoc 2011 Dec;18(Suppl 1):i103-i108 [FREE Full text] [doi: 10.1136/amiajnl-2011-000316] [Medline: 21984588]

109. Rajala T, Savio S, Penttinen J, Dastidar P, Kähönen M, Eskola H, et al. Development of a research dedicated archival system (TARAS) in a university hospital. J Digit Imaging 2011 Oct;24(5):864-873 [FREE Full text] [doi: 10.1007/s10278-010-9350-1] [Medline: 21042830]

110. Suapang P, Yimmun S, Puditkanawat A. Web-based Medical Image Archiving and Communication System for Teleimaging. In: 11th International Conference on Control, Automation and Systems. 2011 Presented at: CAS'11; October 26-29, 2011; Gyeonggi-do, South Korea.

111. Tanioka T, Osaka K, Chiba S. PSYCHOMS, an Electronic Nursing Management System to Facilitate Interdisciplinary Communication and Improve Patient Outcomes in Psychiatric Hospitals. In: 7th International Conference on Natural Language Processing and Knowledge Engineeringtional Conference on Natural Language Processing and Knowledge Engineering. 2011 Presented at: NLPKE'11; November 27-29, 2011; Tokushima, Japan. [doi: 10.1109/nlpke.2011.6138241]

112. Tenenbaum JD, Whetzel PL, Anderson K, Borromeo CD, Dinov ID, Gabriel D, et al. The biomedical resource ontology (BRO) to enable resource discovery in clinical and translational research. J Biomed Inform 2011 Feb;44(1):137-145 [FREE Full text] [doi: 10.1016/j.jbi.2010.10.003] [Medline: 20955817]

113. Teodoro D, Choquet R, Schober D, Mels G, Pasche E, Ruch P, et al. Interoperability driven integration of biomedical data sources. Stud Health Technol Inform 2011;169:185-189. [Medline: 21893739]

114. Zheng L, Wang L, Wang D, Deng N, Lu X, Duan H. A Clinical Omics Database Integrating Epidemiology, Clinical, and Omics Data for Colorectal Cancer Translational Research. In: 4th International Conference on Biomedical Engineering and Informatics (BMEI). 2011 Presented at: BMEI'11; October 15-17, 2011; Shanghai, China. [doi: 10.1109/bmei.2011.6098677]

115. Antoniades A, Georgousopoulos C, Forgo N. Linked2Safety: A Secure Linked Data Medical Information Space for Semantically-interconnecting EHRs Advancing Patients' Safety in Medical Research. In: 12th International Conference on Bioinformatics & Bioengineering. 2012 Presented at: BIBE'12; November 11-13, 2012; Larnaca, Cyprus. [doi: 10.1109/bibe.2012.6399767]

116. Armstrong A, Reddy S, Garg A. Novel approach to utilizing electronic health records for dermatologic research: developing a multi-institutional federated data network for clinical and translational research in psoriasis and psoriatic arthritis. Dermatol Online J 2012 May 15;18(5):2 [FREE Full text] [Medline: 22630572]

117. Bernal JG, Lopez DM, Blobel B. Architectural approach for semantic EHR systems development based on detailed clinical models. Stud Health Technol Inform 2012;177:164-169. [Medline: 22942049]

118. Blechner M, Saripalle R, Demurjian S. A Proposed Star Schema and Extraction Process to Enhance the Collection of Contextual & Semantic Information for Clinical Research Data Warehouses. In: International Conference on Bioinformatics and Biomedicine Workshops. 2012 Presented at: BIBMW'12; October 4-7, 2012; Philadelphia, PA, USA. [doi: 10.1109/bibmw.2012.6470242]

119. de Mul M, Alons P, van der Velde P, Konings I, Bakker J, Hazelzet J. Development of a clinical data warehouse from an intensive care clinical information system. Comput Methods Programs Biomed 2012 Jan;105(1):22-30. [doi: 10.1016/j.cmpb.2010.07.002] [Medline: 20728956]

120. Holford ME, McCusker JP, Cheung K, Krauthammer M. A semantic web framework to integrate cancer omics data with biological knowledge. BMC Bioinformatics 2012 Jan 25;13(Suppl 1):S10 [FREE Full text] [doi: 10.1186/1471-2105-13-S1-S10] [Medline: 22373303]

121. Hsu W, Taira RK, El-Saden S, Kangarloo H, Bui AA. Context-based electronic health record: toward patient specific healthcare. IEEE Trans Inf Technol Biomed 2012 Mar;16(2):228-234 [FREE Full text] [doi: 10.1109/TITB.2012.2186149] [Medline: 22395637]

122. Jayapandian CP, Zhao M, Ewing RM, Zhang G, Sahoo SS. A semantic proteomics dashboard (SemPoD) for data management in translational research. BMC Syst Biol 2012;6(Suppl 3):S20 [FREE Full text] [doi: 10.1186/1752-0509-6-S3-S20] [Medline: 23282161]

123. Liu D, Görges M, Jenkins SA. University of Queensland vital signs dataset: development of an accessible repository of anesthesia patient monitoring data for research. Anesth Analg 2012 Mar;114(3):584-589. [doi: 10.1213/ANE.0b013e318241f7c0] [Medline: 22190558]

124. Majeed RW, Röhrig R. Automated realtime data import for the i2b2 clinical data warehouse: introducing the HL7 ETL cell. Stud Health Technol Inform 2012;180:270-274. [Medline: 22874194]

125. Meyer J, Ostrzinski S, Fredrich D, Havemann C, Krafczyk J, Hoffmann W. Efficient data management in a large-scale epidemiology research project. Comput Methods Programs Biomed 2012 Sep;107(3):425-435. [doi: 10.1016/j.cmpb.2010.12.016] [Medline: 21256617]

126. Pan X, Zhou X, Song H, Zhang R, Zhang T. Enhanced Data Extraction, Transforming and Loading Processing for Traditional Chinese Medicine Clinical Data Warehouse. In: 14th International Conference on e-Health Networking, Applications and Services (Healthcom). 2012 Presented at: HealthCom'12; October 10-13, 2012; Beijing, China. [doi: 10.1109/HealthCom.2012.6380066]

127. Sfakianakis S, Sakkalis V, Marias K, Stamatakos G, McKeever S, Deisboeck TS, et al. An architecture for integrating cancer model repositories. Conf Proc IEEE Eng Med Biol Soc 2012;2012:6628-6631. [doi: 10.1109/EMBC.2012.6347514] [Medline: 23367449]

128. Tamersoy A, Loukides G, Nergiz ME, Saygin Y, Malin B. Anonymization of longitudinal electronic medical records. IEEE Trans Inf Technol Biomed 2012 May;16(3):413-423 [FREE Full text] [doi: 10.1109/TITB.2012.2185850] [Medline: 22287248]

129. Vcelak P, Kratochvil M, Kleckova J, Rohan V. Metamed-Medical Meta Data Extraction and Manipulation Tool Used in the Semantically Interoperable Research Information System. In: 5th International Conference on BioMedical Engineering and Informatics. 2012 Presented at: BMEI'12; October 16-18, 2012; Chongqing, China. [doi: 10.1109/bmei.2012.6513038]

130. da Silva KR, Costa R, Crevelari ES, Lacerda MS, de Moraes Albertini CM, Filho MM, et al. Glocal clinical registries: pacemaker registry design and implementation for global and local integration--methodology and case study. PLoS One 2013;8(7):e71090 [FREE Full text] [doi: 10.1371/journal.pone.0071090] [Medline: 23936257]

131. Farley T, Kiefer J, Lee P, von Hoff D, Trent JM, Colbourn C, et al. The biointelligence framework: a new computational platform for biomedical knowledge computing. J Am Med Inform Assoc 2013 Jan 1;20(1):128-133 [FREE Full text] [doi: 10.1136/amiajnl-2011-000646] [Medline: 22859646]

132. Fraccaro P, Dentone C, Fenoglio D, Giacomini M. Multicentre clinical trials' data management: a hybrid solution to exploit the strengths of electronic data capture and electronic health records systems. Inform Health Soc Care 2013 Dec;38(4):313-329. [doi: 10.3109/17538157.2013.812648] [Medline: 23957714]

133. Gokulakannan E, Venkatachalapathy K. Survey on Privacy Preserving Updates on Unidentified Database. In: Fourth International Conference on Computing, Communications and Networking Technologies. 2013 Presented at: ICCCNT'13; July 4-6, 2013; Tiruchengode, India. [doi: 10.1109/icccnt.2013.6726507]

134. Hamoud AK, Obaid T. Building data warehouse for diseases registry: first step for clinical data warehouse. SSRN J 2013;4(7):636-640. [doi: 10.2139/ssrn.3061599]

135. Heath J. A Privacy Framework for Secondary Use of Medical Data. In: International Symposium on Technology and Society: Social Implications of Wearable Computing and Augmediated Reality in Everyday Life. 2013 Presented at: ISTAS'13; June 27-29, 2013; Toronto, ON, Canada. [doi: 10.1109/istas.2013.6613116]

136. Kovacevic A, Dehghan A, Filannino M, Keane JA, Nenadic G. Combining rules and machine learning for extraction of temporal expressions and events from clinical narratives. J Am Med Inform Assoc 2013;20(5):859-866 [FREE Full text] [doi: 10.1136/amiajnl-2013-001625] [Medline: 23605114]

137. Miyoshi NS, Pinheiro DG, Silva WA, Felipe JC. Computational framework to support integration of biomolecular and clinical data within a translational approach. BMC Bioinformatics 2013 Jun 6;14:180 [FREE Full text] [doi: 10.1186/1471-2105-14-180] [Medline: 23742129]

138. Post AR, Kurc T, Cholleti S, Gao J, Lin X, Bornstein W, et al. The analytic information warehouse (AIW): a platform for analytics using electronic health record data. J Biomed Inform 2013 Jun;46(3):410-424 [FREE Full text] [doi: 10.1016/j.jbi.2013.01.005] [Medline: 23402960]

139. Rüping S, Anguita A, Bucur A, Cirstea T, Jacobs B, Torge A. Improving the Implementation of Clinical Decision Support Systems. In: 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2013 Presented at: EMBC'13; July 3-7, 2013; Osaka, Japan. [doi: 10.1109/embc.2013.6610225]

XSL•FO
RenderX

140. Seah B. An Application of a Healthcare Data Warehouse System. In: Third International Conference on Innovative Computing Technology. 2013 Presented at: INTECH'13; August 29-31, 2013; London, UK. [doi: 10.1109/intech.2013.6653698]

141. Spyropoulos B, Tzavaras A, Zogogianni D, Botsivaly M. Adapting the design of anesthesia information management systems to innovations depicted in industrial property documents. Conf Proc IEEE Eng Med Biol Soc 2013;2013:890-893. [doi: 10.1109/EMBC.2013.6609644] [Medline: 24109831]

142. Tsafara A, Tryfonopoulos C, Skiadopoulos S. CloudStudy: A Cloud-based System for Supporting Multi-centre Studies. In: 13th IEEE International Conference on BioInformatics and BioEngineering. 2013 Presented at: BIBE'13; November 10-13, 2013; Chania, Greece. [doi: 10.1109/bibe.2013.6701549]

143. Zhao X, Dong T. Design and experimental approach to the construction of a human signal-molecule-profiling database. Int J Environ Res Public Health 2013 Dec 9;10(12):6887-6908 [FREE Full text] [doi: 10.3390/ijerph10126887] [Medline: 24351788]

144. Bernsmed K, Cruzes D, Jaatun M, Haugset B, Gjære E. Healthcare Services in the Cloud--Obstacles to Adoption, and a Way Forward. In: Ninth International Conference on Availability, Reliability and Security. 2014 Presented at: ARES'14; September 8-12, 2014; Fribourg, Switzerland. [doi: 10.1109/ares.2014.28]

145. Cano I, Tényi A, Schueller C, Wolff M, Huertas Migueláñez MM, Gomez-Cabrero D, et al. The COPD knowledge base: enabling data analysis and computational simulation in translational COPD research. J Transl Med 2014 Nov 28;12(Suppl 2):S6 [FREE Full text] [doi: 10.1186/1479-5876-12-S2-S6] [Medline: 25471253]

146. Chen F, Wang S, Mohammed N, Cheng S, Jiang X. PRECISE:privacy-preserving cloud-assisted quality improvement service in healthcare. IEEE Int Conf Systems Biol 2014 Oct;2014:176-183 [FREE Full text] [doi: 10.1109/ISB.2014.6990752] [Medline: 26146645]

147. Chouvarda I, Philip N, Natsiavas P, Kilintzis V, Sobnath D, Kayyali R, et al. WELCOME – innovative integrated care platform using wearable sensing and smart cloud computing for COPD patients with comorbidities. Conf Proc IEEE Eng Med Biol Soc 2014;2014:3180-3183. [doi: 10.1109/EMBC.2014.6944298] [Medline: 25570666]

148. Dagliati A, Sacchi L, Bucalo M. A Data Gathering Framework to Collect Type 2 Diabetes Patients Data. In: International Conference on Biomedical and Health Informatics. 2014 Presented at: BHI'14; June 1-4, 2014; Valencia, Spain. [doi: 10.1109/bhi.2014.6864349]

149. Dalpé G, Joly Y. Opportunities and challenges provided by cloud repositories for bioinformatics-enabled drug discovery. Drug Dev Res 2014 Sep;75(6):393-401. [doi: 10.1002/ddr.21211] [Medline: 25195583]

150. Dietrich G, Fette G, Puppe F. A Comparison of Search Engine Technologies for a Clinical Data Warehouse. In: Proceedings of the 16th LWA Workshops. 2014 Presented at: LWA'14; September 8-10, 2014; Aachen, Germany.

151. Fiehe C, Litvina A, Tonn J. Building a Medical Research Cloud in the EASI-CLOUDS Project. In: 6th International Workshop on Science Gateways. 2014 Presented at: IWSG'14; June 3-5, 2014; Dublin, Ireland. [doi: 10.1109/iwsg.2014.13]

152. Gavrielov-Yusim N, Friger M. Use of administrative medical databases in population-based research. J Epidemiol Community Health 2014 Mar;68(3):283-287. [doi: 10.1136/jech-2013-202744] [Medline: 24248997]

153. Ghane K. Healthcare information exchange system based on a hybrid central/federated model. Conf Proc IEEE Eng Med Biol Soc 2014;2014:1362-1365. [doi: 10.1109/EMBC.2014.6943852] [Medline: 25570220]

154. Jaja BN, Attalla D, Macdonald RL, Schweizer TA, Cusimano MD, Etminan N, et al. The subarachnoid hemorrhage international trialists (SAHIT) repository: advancing clinical research in subarachnoid hemorrhage. Neurocrit Care 2014 Dec;21(3):551-559. [doi: 10.1007/s12028-014-9990-y] [Medline: 24865271]

155. Laohakangvalvit T, Achalakul T. Cloud-Based Data Exchange Framework for Healthcare Services. In: 11th International Joint Conference on Computer Science and Software Engineering. 2014 Presented at: JCSSE'14; May 14-16, 2014; Chon Buri, Thailand. [doi: 10.1109/jcsse.2014.6841874]

156. Mohammed RO, Talab SA. Clinical data warehouse issues and challenges. Int J u-e-Serv Sci Technol 2014 Oct 31;7(5):251-262. [doi: 10.14257/ijunesst.2014.7.5.22]

157. Parane K, Patil N, Poojara S, Kamble T. Cloud Based Intelligent Healthcare Monitoring System. In: International Conference on Issues and Challenges in Intelligent Computing Techniques. 2014 Presented at: ICICICT'14; February 7-8, 2014; Ghaziabad, India. [doi: 10.1109/icicict.2014.6781365]

158. Pecoraro F, Luzi D, Ricci F. A Clinical Data Warehouse Architecture based on the Electronic Healthcare Record Infrastructure. In: Proceedings of the International Conference on Health Informatics. 2014 Presented at: CHI;14; March 3-6, 2014; Angers, France. [doi: 10.5220/0004764502870294]

159. Pennington JW, Ruth B, Italia MJ, Miller J, Wrazien S, Loutrel JG, et al. Harvest: an open platform for developing web-based biomedical data discovery and reporting applications. J Am Med Inform Assoc 2014;21(2):379-383 [FREE Full text] [doi: 10.1136/amiajnl-2013-001825] [Medline: 24131510]

160. Rosenbloom ST, Harris P, Pulley J, Basford M, Grant J, DuBuisson A, et al. The mid-south clinical data research network. J Am Med Inform Assoc 2014;21(4):627-632 [FREE Full text] [doi: 10.1136/amiajnl-2014-002745] [Medline: 24821742]

161. Scalone L, Cesana G, Furneri G, Ciampichini R, Beck-Peccoz P, Chiodini V, et al. Burden of diabetes mellitus estimated with a longitudinal population-based study using administrative databases. PLoS One 2014;9(12):e113741 [FREE Full text] [doi: 10.1371/journal.pone.0113741] [Medline: 25470484]

XSL•FO
RenderX

162.   Schreiweis B, Schneider G, Eichner T, Bergh B, Heinze O. Health information research platform (HIReP)--an architecture pattern. Stud Health Technol Inform 2014;205:773-777. [Medline: 25160292]

163.   Setareh S, Rezaee A, Farahmandian V, Hajinazari P, Asosheh A. A Cloud-based Model for Hospital Information Systems Integration. In: 7th International Symposium on Telecommunications. 2014 Presented at: IST'14; September 9-11, 2014; Tehran, Iran. [doi: 10.1109/istel.2014.7000792]

164.   Thorogood A, Joly Y, Knoppers BM, Nilsson T, Metrakos P, Lazaris A, et al. An implementation framework for the feedback of individual research results and incidental findings in research. BMC Med Ethics 2014 Dec 23;15:88 [FREE Full text] [doi: 10.1186/1472-6939-15-88] [Medline: 25539799]

165.   Tsumoto S, Hirano S. Healthcare IT: Integration of Consumer Healthcare Data and Electronic Medical Records for Chronic Disease Management. In: International Conference on Granular Computing. 2014 Presented at: GrC'14; October 22-24, 2014; Noboribetsu, Japan. [doi: 10.1109/grc.2014.6982855]

166.   Yamaguchi H, Ito Y. Improving the Effectiveness of Interprofessional Work Teams Using EHR-Based Data in the Treatment of Chronic Diseases: An Action Research Study. In: International Conference on Management of Engineering & Technology. 2007 Presented at: PICMET'07; August 5-9, 2014; Portland, OR, USA. [doi: 10.1109/picmet.2007.4349302]

167.   Adler-Milstein J, DesRoches CM, Kralovec P, Foster G, Worzala C, Charles D, et al. Electronic health record adoption in US hospitals: progress continues, but challenges persist. Health Aff (Millwood) 2015 Dec;34(12):2174-2180. [doi: 10.1377/hlthaff.2015.0992] [Medline: 26561387]

168.   Agbele K, Oriogun P, Seluwa A, Aruleba K. Towards a Model for Enhancing ICT4 Development and Information Security in Healthcare System. In: International Symposium on Technology and Society. 2015 Presented at: ISTAS'15; November 11-12, 2015; Dublin, Ireland. [doi: 10.1109/istas.2015.7439404]

169.   Ahmadi M, Ghazisaeidi M, Bashiri A. Radiology reporting system data exchange with the electronic health record system: a case study in Iran. Glob J Health Sci 2015 Mar 18;7(5):208-214 [FREE Full text] [doi: 10.5539/gjhs.v7n5p208] [Medline: 26156904]

170.   Amato F, Cozzolino G, Maisto A, Mazzeo A, Moscato V, Pelosi S, et al. ABC: A Knowledge Based Collaborative Framework for E-health. In: 2015 IEEE 1st International Forum on Research and Technologies for Society and Industry Leveraging a better tomorrow (RTSI). 2015 Presented at: RTSI'15; September 16-18, 2015; Turin, Italy. [doi: 10.1109/rtsi.2015.7325107]

171.   Combi C, Pozzani G, Pozzi G. Design, Development, Deployment of a Telemedicine System in a Developing Country: Dealing With Organizational and Social Issues. In: International Conference on Healthcare Informatics. 2015 Presented at: ICHI'15; October 21-23, 2015; Dallas, TX, USA. [doi: 10.1109/ichi.2015.103]

172.   Delamarre D, Bouzille G, Dalleau K, Courtel D, Cuggia M. Semantic integration of medication data into the EHOP clinical data warehouse. Stud Health Technol Inform 2015;210:702-706. [Medline: 25991243]

173.   Girardi D, Dirnberger J, Giretzlehner M. An ontology‐based clinical data warehouse for scientific research. Saf Health 2015 Jul 20;1:6. [doi: 10.1186/2056-5917-1-6]

174.   Huang Z, Duan H, Li H. TCGA4U: a web-based genomic analysis platform to explore and mine TCGA genomic data for translational research. Stud Health Technol Inform 2015;216:658-662. [Medline: 26262133]

175.   Kämpgen B, Werner H, Deeb R, Bornhövd C. Towards a Semantic Clinical Data Warehouse: A Case Study of Discovering Similar Genes. In: Proceedings of the 4th Workshop on Knowledge Discovery and Data Mining Meets Linked Open Data co-located with 12th Extended Semantic Web Conference (ESWC 2015). 2015 Presented at: KDD'15; May 31, 2015; Portoroz, Slovenia.

176.   Kantorovitch J, Giakoumaki A, Korakis A. Knowledge Modelling Framework. In: 2nd International Conference on Information and Communication Technologies for Disaster Management. 2015 Presented at: ICT-DM'15; November 30-December 2, 2015; Rennes, France. [doi: 10.1109/ict-dm.2015.7402037]

177.   Kong G, Xiao Z. Protecting privacy in a clinical data warehouse. Health Informatics J 2015 Jun;21(2):93-106. [doi: 10.1177/1460458213504204] [Medline: 25301198]

178.   Lee H, Kim H. Ehealth Recommendation Service System Using Ontology and Case-Based Reasoning. In: International Conference on Smart City/SocialCom/SustainCom. 2015 Presented at: SmartCity'15; December 19-21, 2015; Chengdu, China. [doi: 10.1109/smartcity.2015.217]

179.   Lu J, Keech M. Emerging Technologies for Health Data Analytics Research: A Conceptual Architecture. In: Proceedings of the 2015 26th International Workshop on Database and Expert Systems Applications (DEXA). 2015 Presented at: DEXA'15; September 1-4, 2015; Washington, DC, USA. [doi: 10.1109/dexa.2015.58]

180.   Maaroufi M, Choquet R, Landais P, Jaulent M. Towards data integration automation for the French rare disease registry. AMIA Annu Symp Proc 2015;2015:880-885 [FREE Full text] [Medline: 26958224]

181.   Martin-Sanchez F, Turner M, Johnstone A, Heffer L, Rafael N, Advisory Group, et al. Personalised medicine possible with real-time integration of genomic and clinical data to inform clinical decision-making. Stud Health Technol Inform 2015;216:1052. [Medline: 26262351]

182.   Mohyuddin. Bridging the gap from bench to bedside--an informatics infrastructure for integrating clinical, genomics and environmental data (ICGED). Stud Health Technol Inform 2015;216:1054. [Medline: 26262353]

183. Mou X, Wang X, Wu Z, Wang X, Zhou M. An Automatic Ehealth Platform for Cardiovascular and Cerebrovascular Disease Detection. In: International Symposium on Bioelectronics and Bioinformatics. 2015 Presented at: ISBB'15; October 14-17, 2015; Beijing, China. [doi: 10.1109/isbb.2015.7344924]

184. Puppala M, He T, Chen S, Ogunti R, Yu X, Li F, et al. METEOR: an enterprise health informatics environment to support evidence-based medicine. IEEE Trans Biomed Eng 2015 Dec;62(12):2776-2786. [doi: 10.1109/TBME.2015.2450181] [Medline: 26126271]

185. Sanz-Requena R, Mañas-García A, Cabrera-Ayala J, García-Martí G. A Cloud-Based Radiological Portal for the Patients: IT Contributing to Position the Patient as the Central Axis of the 21st Century Healthcare Cycles. In: Proceedings of the 2015 IEEE/ACM 1st International Workshop on TEchnical and LEgal aspects of data pRivacy and SEcurity. 2015 Presented at: TELERISE'15; May 18, 2015; Florence, Italy. [doi: 10.1109/telerise.2015.18]

186. Schnell R, Borgs C. Building a National Perinatal Data Base Without the Use of Unique Personal Identifiers. In: International Conference on Data Mining Workshop. 2015 Presented at: ICDMW'15; November 14-17, 2015; Atlantic City, NJ, USA. [doi: 10.1109/icdmw.2015.19]

187. Sharghi H, Ma W, Sartipi K. Federated Service-Based Authentication Provisioning for Distributed Diagnostic Imaging Systems. In: 28th International Symposium on Computer-Based Medical Systems. 2015 Presented at: CBMS'15; June 22-25, 2015; Sao Carlos, Brazil. [doi: 10.1109/cbms.2015.85]

188. Skripcak T, Just U, Simon M, Buttner D, Luhr A, Baumann M, et al. Toward distributed conduction of large-scale studies in radiation therapy and oncology: open-source system integration approach. IEEE J Biomed Health Inform 2016 Sep;20(5):1397-1403. [doi: 10.1109/jbhi.2015.2450833]

189. Soudris D, Xydis S, Baloukas C. AEGLE: A Big Bio-Data Analytics Framework for Integrated Health-Care Services. In: International Conference on Embedded Computer Systems: Architectures, Modeling, and Simulation. 2015 Presented at: SAMOS'15; July 19-23, 2015; Samos, Greece. [doi: 10.1109/samos.2015.7363682]

190. Weber GM. Federated queries of clinical data repositories: scaling to a national network. J Biomed Inform 2015 Jun;55:231-236 [FREE Full text] [doi: 10.1016/j.jbi.2015.04.012] [Medline: 25957825]

191. Yang C, Liu C, Tseng T. Design and Implementation of a Privacy Aware Framework for Sharing Electronic Health Records. In: International Conference on Healthcare Informatics. 2015 Presented at: ICHI'15; October 21-23, 2015; Dallas, TX, USA. [doi: 10.1109/ichi.2015.92]

192. Yao Q, Tian Y, Li P, Tian L, Qian Y, Li J. Design and development of a medical big data processing system based on Hadoop. J Med Syst 2015 Mar;39(3):23. [doi: 10.1007/s10916-015-0220-8] [Medline: 25666927]

193. Andrew NE, Sundararajan V, Thrift AG, Kilkenny MF, Katzenellenbogen J, Flack F, et al. Addressing the challenges of cross-jurisdictional data linkage between a national clinical quality registry and government-held health data. Aust N Z J Public Health 2016 Oct;40(5):436-442. [doi: 10.1111/1753-6405.12576] [Medline: 27625174]

194. Bauer CR, Ganslandt T, Baum B, Christoph J, Engel I, Löbe M, et al. Integrated data repository toolkit (IDRT). A suite of programs to facilitate health analytics on heterogeneous medical data. Methods Inf Med 2016;55(2):125-135. [doi: 10.3414/ME15-01-0082] [Medline: 26534843]

195. Chelico JD, Wilcox AB, Vawdrey DK, Kuperman GJ. Designing a clinical data warehouse architecture to support quality improvement initiatives. AMIA Annu Symp Proc 2016;2016:381-390 [FREE Full text] [Medline: 28269833]

196. Denney MJ, Long DM, Armistead MG, Anderson JL, Conway BN. Validating the extract, transform, load process used to populate a large clinical research database. Int J Med Inform 2016 Oct;94:271-274 [FREE Full text] [doi: 10.1016/j.ijmedinf.2016.07.009] [Medline: 27506144]

197. García-de-León-Chocano R, Muñoz-Soler V, Sáez C, García-de-León-González R, García-Gómez JM. Construction of quality-assured infant feeding process of care data repositories: construction of the perinatal repository (part 2). Comput Biol Med 2016 Apr 1;71:214-222. [doi: 10.1016/j.compbiomed.2016.01.007] [Medline: 26950399]

198. Gupta S, Tripathi P. Big Data Lakes Can Support Better Population Health for Rural India - Swastha Bharat. In: International Conference on Innovation and Challenges in Cyber Security. 2016 Presented at: ICICCS-INBUSH'16; February 3-5, 2016; Noida, India. [doi: 10.1109/iciccs.2016.7542361]

199. Koppad S, Kumar A. Application of Big Data Analytics in Healthcare System to Predict COPD. In: International Conference on Circuit, Power and Computing Technologies. 2016 Presented at: ICCPCT'16; March 18-19, 2016; Nagercoil, India. [doi: 10.1109/iccpct.2016.7530248]

200. Liaw S, De Lusignan S. An 'integrated health neighbourhood' framework to optimise the use of EHR data. J Innov Health Inform 2016 Oct 4;23(3):826 [FREE Full text] [doi: 10.14236/jhi.v23i3.826] [Medline: 28059689]

201. Olaronke I, Oluwaseun O. Big Data in Healthcare: Prospects, Challenges and Resolutions. In: Future Technologies Conference. 2016 Presented at: FTC'16; December 6-7, 2016; San Francisco, CA, USA. [doi: 10.1109/ftc.2016.7821747]

202. Padula WV, Blackshaw L, Brindle CT, Volchenboum SL. An approach to acquiring, normalizing, and managing EHR data from a clinical data repository for studying pressure ulcer outcomes. J Wound Ostomy Continence Nurs 2016;43(1):39-45. [doi: 10.1097/WON.0000000000000185] [Medline: 26727681]

203. Vishnyakova D, Gaudet-Blavignac C, Baumann P, Lovis C. Clinical data models at university hospitals of Geneva. Stud Health Technol Inform 2016;221:97-101. [Medline: 27071885]

XSL·FO
RenderX

204. Bellini F, Gutierrez-Zorrilla J, Anza L, Ferreira E, Deneault L, Vanerio G. MDi: acquisition, analysis and data visualization system in healthcare. In: 2017 IEEE URUCON. 2017 Presented at: URUCON; 23-25 Oct. 2017; Montevideo, Uruguay p. 1-4. [doi: 10.1109/urucon.2017.8171879]

205. Dogaru D, Dumitrache I. Holistic Perspective of Big Data in Healthcare. In: E-Health and Bioengineering Conference (EHB). 2017 Presented at: EHB'17; June 22-24, 2017; Sinaia, Romania. [doi: 10.1109/ehb.2017.7995450]

206. Ewing A, Rogus J, Chintagunta P, Kraus L, Sabol M, Kang H. A Systems Approach to Improving Patient Flow at UVA Cancer Center Using Real-time Locating System. In: Systems and Information Engineering Design Symposium (SIEDS). 2017 Presented at: SIEDS'17; April 28, 2017; Charlottesville, VA, USA. [doi: 10.1109/sieds.2017.7937727]

207. Foran DJ, Chen W, Chu H, Sadimin E, Loh D, Riedlinger G, et al. Roadmap to a comprehensive clinical data warehouse for precision medicine applications in oncology. Cancer Inform 2017;16:1176935117694349 [FREE Full text] [doi: 10.1177/1176935117694349] [Medline: 28469389]

208. Khan SI, Hoque AS. Development of National Health Data Warehouse Bangladesh: Privacy Issues and a Practical Solution. In: 18th International Conference on Computer and Information Technology. 2017 Presented at: ICCIT'17; December 21-23, 2017; Dhaka, Bangladesh. [doi: 10.1109/ICCITechn.2015.7488099]

209. Lai J, Lee D, Yang M. Constructing the Cloud Computing System for Advanced Data Analysis of Biomedical Research. In: International Conference on Machine Learning and Cybernetics. 2017 Presented at: ICMLC'17; July 9-12, 2017; Ningbo, China. [doi: 10.1109/icmlc.2017.8108969]

210. Poenaru C, Merezeanu D, Dobrescu R, Posdarascu E. Advanced Solutions for Medical Information Storing: Clinical Data Warehouse. In: E-Health and Bioengineering Conference. 2017 Presented at: EHB'17; June 22-24, 2017; Sinaia, Romania. [doi: 10.1109/ehb.2017.7995355]

211. Sáez C, Moner D, García-De-León-Chocano R, Muñoz-Soler V, García-De-León-González R, Maldonado JA, et al. A standardized and data quality assessed maternal-child care integrated data repository for research and monitoring of best practices: a pilot project in Spain. Stud Health Technol Inform 2017;235:539-543. [Medline: 28423851]

212. Vuppala S, Dinesh M, Viswanathan S, Ramachandran G, Bussa N, Geetha M. Cloud Based Big Data Platform for Image Analytics. In: International Conference on Cloud Computing in Emerging Markets. 2017 Presented at: CCEM'17; November 1-3, 2017; Bangalore, India. [doi: 10.1109/ccem.2017.11]

213. Weng W, Wagholikar KB, McCray AT, Szolovits P, Chueh HC. Medical subdomain classification of clinical notes using a machine learning-based natural language processing approach. BMC Med Inform Decis Mak 2017 Dec 1;17(1):155. [doi: 10.1186/s12911-017-0556-8] [Medline: 29191207]

214. Brisimi TS, Chen R, Mela T, Olshevsky A, Paschalidis IC, Shi W. Federated learning of predictive models from federated electronic health records. Int J Med Inform 2018 Apr;112:59-67. [doi: 10.1016/j.ijmedinf.2018.01.007] [Medline: 29500022]

215. Deshpande P, Rasin A, Brown E. Big Data Integration Case Study for Radiology Data Sources. In: Life Sciences Conference. 2018 Presented at: LSC'18; October 28-30, 2018; Montreal, QC, Canada. [doi: 10.1109/lsc.2018.8572185]

216. Fette G, Kaspar M, Liman L, Dietrich G, Ertl M, Krebs J, et al. Exporting data from a clinical data warehouse. Stud Health Technol Inform 2018;248:88-93. [Medline: 29726423]

217. Garcelon N, Neuraz A, Salomon R, Bahi-Buisson N, Amiel J, Picard C, et al. Next generation phenotyping using narrative reports in a rare disease clinical data warehouse. Orphanet J Rare Dis 2018 May 31;13(1):85 [FREE Full text] [doi: 10.1186/s13023-018-0830-6] [Medline: 29855327]

218. de Quirós FG, Otero C, Luna D. Terminology services: standard terminologies to control health vocabulary. Yearb Med Inform 2018 Aug;27(1):227-233 [FREE Full text] [doi: 10.1055/s-0038-1641200] [Medline: 29681027]

219. Guo S, Jin Z, Gotz D, Du F, Zha H, Cao N. Visual progression analysis of event sequence data. IEEE Trans Vis Comput Graph 2018 Aug 20;25(1):417-426. [doi: 10.1109/TVCG.2018.2864885] [Medline: 30136953]

220. Kundalwal M, Singh A, Chatterjee K. A Privacy Framework in Cloud Computing for Healthcare Data. In: International Conference on Advances in Computing, Communication Control and Networking. 2018 Presented at: ICACCCN'18; October 12-13, 2018; Greater Noida (UP), India, India. [doi: 10.1109/icacccn.2018.8748480]

221. Maji G, Debnath N, Sen S. Data Warehouse Based Analysis with Integrated Blood Donation Management System. In: 16th International Conference on Industrial Informatics. 2018 Presented at: INDIN'18; July 18-20, 2018; Porto, Portugal. [doi: 10.1109/indin.2018.8471988]

222. Mande R, JayaLakshmi G, Yelavarti K. Leveraging Distributed Data Over Big Data Analytics Platform for Healthcare Services. In: 2nd International Conference on Trends in Electronics and Informatics. 2018 Presented at: ICOEI'18; May 11-12, 2018; Tirunelveli, India. [doi: 10.1109/icoei.2018.8553827]

223. Mullin S, Zhao J, Sinha S, Lee R, Song B, Elkin PL. Clinical data warehouse query and learning tool using a human-centered participatory design process. Stud Health Technol Inform 2018;251:59-62 [FREE Full text] [Medline: 29968601]

224. Patiny L, Zasso M, Kostro D, Bernal A, Castillo AM, Bolaños A, et al. The C6H6 NMR repository: an integral solution to control the flow of your data from the magnet to the public. Magn Reson Chem 2018 Jun;56(6):520-528. [doi: 10.1002/mrc.4669] [Medline: 28981966]

225. Pletcher MJ, Forrest CB, Carton TW. PCORnet's collaborative research groups. Patient Relat Outcome Meas 2018 Feb;Volume 9:91-95. [doi: 10.2147/prom.s141630]

226. Raisaro JL, Troncoso-Pastoriza JR, Misbach M, Sousa JS, Pradervand S, Missiaglia E, et al. MedCo: enabling secure and privacy-preserving exploration of distributed clinical and genomic data. IEEE/ACM Trans Comput Biol Bioinform 2019;16(4):1328-1341. [doi: 10.1109/TCBB.2018.2854776] [Medline: 30010584]

227. Rinner C, Gezgin D, Wendl C, Gall W. A clinical data warehouse based on OMOP and i2b2 for Austrian health claims data. Stud Health Technol Inform 2018;248:94-99. [Medline: 29726424]

228. Solbrig HR, Hong N, Murphy SN, Jiang G. Automated population of an i2b2 clinical data warehouse using FHIR. AMIA Annu Symp Proc 2018;2018:979-988 [FREE Full text] [Medline: 30815141]

229. Sylvestre E, Bouzillé G, Chazard E, His-Mahier C, Riou C, Cuggia M. Combining information from a clinical data warehouse and a pharmaceutical database to generate a framework to detect comorbidities in electronic health records. BMC Med Inform Decis Mak 2018 Jan 24;18(1):9 [FREE Full text] [doi: 10.1186/s12911-018-0586-x] [Medline: 29368609]

230. Ye B, Basdekis I, Smyrlis M, Spanoudakis G, Koloutsou K. A Big Data Repository and Architecture for Managing Hearing Loss Related Data. In: International Conference on Biomedical & Health Informatics. 2018 Presented at: BHI'18; March 4-7, 2018; Las Vegas, NV, USA. [doi: 10.1109/bhi.2018.8333397]

231. Baghal A, Zozus M, Baghal A, Al-Shukri S, Prior F. Factors associated with increased adoption of a research data warehouse. Stud Health Technol Inform 2019;257:31-35. [Medline: 30741168]

232. Black M, Wallace J, Rankin D. Meaningful Integration of Data, Analytics and Services of Computer-Based Medical Systems: The MIDAS Touch. In: 32nd International Symposium on Computer-Based Medical Systems. 2019 Presented at: CBMS'19; June 5-7, 2019; Cordoba, Spain. [doi: 10.1109/cbms.2019.00031]

233. Boujdad F, Gaignard A, Südholt M, Garzón-Alfonso W, Navarro L, Redon R. On Distributed Collaboration for Biomedical Analyses. In: 19th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGRID). 2019 Presented at: CCGRID'19; May 14-17, 2019; Larnaca, Cyprus. [doi: 10.1109/ccgrid.2019.00079]

234. Conte B, Fortin E, Singh P. Health Information Management System for a Rural Medical Clinic in Nicaragua. In: 2019 IEEE 39th Central America and Panama Convention. 2019 Presented at: CAPC'19; November 20-22, 2019; Guatemala City, Guatemala. [doi: 10.1109/concapanxxxix47272.2019.8977099]

235. Daniel C, Kalra D, Section Editors for the IMIA Yearbook Section on Clinical Research Informatics. Clinical research informatics: contributions from 2018. Yearb Med Inform 2019 Aug;28(1):203-205 [FREE Full text] [doi: 10.1055/s-0039-1677921] [Medline: 31419833]

236. Deshpande P, Rasin A, Furst J, Raicu D, Antani S. DiiS: a biomedical data access framework for aiding data driven research supporting fair principles. Data 2019 Apr 20;4(2):54. [doi: 10.3390/data4020054]

237. Duncan D, Vespa P, Pitkänen A, Braimah A, Lapinlampi N, Toga AW. Big data sharing and analysis to advance research in post-traumatic epilepsy. Neurobiol Dis 2019 Mar;123:127-136 [FREE Full text] [doi: 10.1016/j.nbd.2018.05.026] [Medline: 29864492]

238. DuVall SL, Matheny ME, Ibragimov IR, Oats TD, Tucker JN, South BR, et al. A tale of two databases: the DoD and VA infrastructure for clinical intelligence (DaVINCI). Stud Health Technol Inform 2019 Aug 21;264:1660-1661. [doi: 10.3233/SHTI190584] [Medline: 31438280]

239. Fette G, Kaspar M, Liman L, Dietrich G, Ertl M, Krebs J, et al. Query translation between openEHR and i2b2. Stud Health Technol Inform 2019;258:16-20. [Medline: 30942705]

240. Gaies M, Anderson J, Kipps A, Lorts A, Madsen N, Marino B, Cardiac Networks United Executive Committee and Advisory Board. Cardiac networks united: an integrated paediatric and congenital cardiovascular research and improvement network. Cardiol Young 2019 Feb;29(2):111-118. [doi: 10.1017/S1047951118001683] [Medline: 30567622]

241. Gardner BJ, Pedersen JG, Campbell ME, McClay JC. Incorporating a location-based socioeconomic index into a de-identified i2b2 clinical data warehouse. J Am Med Inform Assoc 2019 Apr 1;26(4):286-293 [FREE Full text] [doi: 10.1093/jamia/ocy172] [Medline: 30715327]

242. Gerl A, Meier B. Privacy in the Future of Integrated Health Care Services - Are Privacy Languages the Key? In: International Conference on Wireless and Mobile Computing, Networking and Communication. 2019 Presented at: WiMob'19; October 21-23, 2019; Barcelona, Spain. [doi: 10.1109/wimob.2019.8923532]

243. Juárez D, Schmidt EE, Stahl-Toyota S, Ückert F, Lablans M. A generic method and implementation to evaluate and improve data quality in distributed research networks. Methods Inf Med 2019 Sep;58(2-03):86-93 [FREE Full text] [doi: 10.1055/s-0039-1693685] [Medline: 31514209]

244. Khalique F, Khan SA, Nosheen I. A framework for public health monitoring, analytics and research. IEEE Access 2019;7:101309-101326. [doi: 10.1109/access.2019.2930730]

245. Klangprapunt P, Seresangtakul P. An Information Integration System to Continuing of Care Case study Nongsung Hospital, Mukdahan THAILAND. In: 16th International Joint Conference on Computer Science and Software Engineering. 2019 Presented at: JCSSE'19; July 10-12, 2019; Chonburi, Thailand. [doi: 10.1109/jcsse.2019.8864214]

246. Krithara A, Aisopos F, Rentoumi V. iASiS: Towards Heterogeneous Big Data Analysis for Personalized Medicine. In: 32nd International Symposium on Computer-Based Medical Systems. 2019 Presented at: CBMS'19; June 5-7, 2019; Cordoba, Spain. [doi: 10.1109/cbms.2019.00032]

247. Lee JS, Darcy KM, Hu H, Casablanca Y, Conrads TP, Dalgard CL, et al. From discovery to practice and survivorship: building a national real-world data learning healthcare framework for military and veteran cancer patients. Clin Pharmacol Ther 2019 Jul;106(1):52-57 [FREE Full text] [doi: 10.1002/cpt.1425] [Medline: 30838639]

248. Liu T, Liu X, Fan Y. Constructing a Comprehensive Clinical Database Integrating Patients' Data from Intensive Care Units and General Wards. In: 12th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics. 2019 Presented at: CISP-BMEI'19; October 19-21, 2019; Suzhou, China. [doi: 10.1109/cisp-bmei48845.2019.8966055]

249. Phan-Vogtmann LA, Helhorn A, Kruse HM, Thomas E, Heidel AJ, Saleh K, et al. Approaching clinical data transformation from disparate healthcare IT systems through a modular framework. Stud Health Technol Inform 2019;258:85-89. [Medline: 30942720]

250. Miller JB. Big data and biomedical informatics: preparing for the modernization of clinical neuropsychology. Clin Neuropsychol 2019 Feb;33(2):287-304. [doi: 10.1080/13854046.2018.1523466] [Medline: 30513257]

251. Motema J, Appiah M. Factors Affecting the Adoption of Cloud Computing in a South African Hospital. In: International Conference on Advances in Big Data, Computing and Data Communication Systems. 2019 Presented at: icABCD'19; August 5-6, 2019; Winterton, South Africa. [doi: 10.1109/icabcd.2019.8851015]

252. Offia CE, Crowe M. A theoritical exploration of data management and integration in organisation sectors. Int J Database Manag Syst 2019 Feb 28;11(01):37-56. [doi: 10.5121/ijdms.2019.11103]

253. Opali?ski A, Regulski K, Mrzygóod B. Medical Data Exploration Based on the Heterogeneous Data Sources Aggregation System. In: Federated Conference on Computer Science and Information Systems. 2019 Presented at: FedCSIS'19; September 1-4, 2019; Leipzig, Germany. [doi: 10.15439/2019f258]

254. Post A, Chappidi N, Gunda D, Deshpande N. A method for EHR phenotype management in an i2b2 data warehouse. AMIA Jt Summits Transl Sci Proc 2019;2019:92-101 [FREE Full text] [Medline: 31258960]

255. Raebel MA, Quintana LM, Schroeder EB, Shetterly SM, Pieper LE, Epner PL, et al. Identifying preanalytic and postanalytic laboratory quality gaps using a data warehouse and structured multidisciplinary process. Arch Pathol Lab Med 2019 Apr;143(4):518-524 [FREE Full text] [doi: 10.5858/arpa.2018-0093-OA] [Medline: 30525932]

256. Safavi KC, Driscoll W, Wiener-Kronish JP. Remote surveillance technologies: realizing the aim of right patient, right data, right time. Anesth Analg 2019 Sep;129(3):726-734 [FREE Full text] [doi: 10.1213/ANE.0000000000003948] [Medline: 31425213]

257. Sekeres MA, Gore SD, Stablein DM, DiFronzo N, Abel GA, DeZern AE, et al. The national MDS natural history study: design of an integrated data and sample biorepository to promote research studies in myelodysplastic syndromes. Leuk Lymphoma 2019 Dec;60(13):3161-3171. [doi: 10.1080/10428194.2019.1616186] [Medline: 31111762]

258. Seneviratne M, Kahn M, Hernandez-Boussard T. Merging heterogeneous clinical data to enable knowledge discovery. Pac Symp Biocomput 2019;24:439-443 [FREE Full text] [Medline: 30864344]

259. Shaanika I, Nehemia M. Developing an Integration Architecture to Manage Heterogeneous Data by Private Healthcare Practitioners: A Case of Namibia. In: Open Innovations. 2019 Presented at: OI'19; October 2-4, 2019; Cape Town, South Africa. [doi: 10.1109/oi.2019.8908188]

260. Shah C, Shaikh M, Shah D, Samdani K. A Review on Big Data Practices in Healthcare. In: International Conference on System, Computation, Automation and Networking. 2019 Presented at: ICSCAN'19; March 29-30, 2019; Pondicherry, India. [doi: 10.1109/icscan.2019.8878687]

261. Weeks J, Pardee R. Learning to share health care data: a brief timeline of influential common data models and distributed health data networks in US health care research. EGEMS (Wash DC) 2019 Mar 25;7(1):4 [FREE Full text] [doi: 10.5334/egems.279] [Medline: 30937326]

262. Lian W, Xue T, Lu Y, Wang M, Deng W. Research on hierarchical data fusion of intelligent medical monitoring. IEEE Access 2020;8:38355-38367. [doi: 10.1109/access.2019.2958854]

263. Yau Y, Khethavath P, Figueroa J. Secure Pattern-Based Data Sensitivity Framework for Big Data in Healthcare. In: International Conference on Big Data, Cloud Computing, Data Science & Engineering. 2019 Presented at: BCD'19; May 29-31, 2019; Honolulu, HI, USA. [doi: 10.1109/bcd.2019.8885114]

264. Zhang C, Ma R, Sun S, Li Y, Wang Y, Yan Z. Optimizing the electronic health records through big data analytics: a knowledge-based view. IEEE Access 2019;7:136223-136231. [doi: 10.1109/access.2019.2939158]

265. Johnson AE, Pollard TJ, Shen L, Lehman LH, Feng M, Ghassemi M, et al. MIMIC-III, a freely accessible critical care database. Sci Data 2016 May 24;3:160035 [FREE Full text] [doi: 10.1038/sdata.2016.35] [Medline: 27219127]

266. Felmeister AR, Masino A, Resnick A, Pennington J. Scalable Biobanking: a Modular Electronic Honest Broker and Biorepository for Integrated Clinical, Specimen and Genomic Research. In: International Conference on Bioinformatics and Biomedicine. 2015 Presented at: BIBM'15; November 9-12, 2015; Washington, DC, USA. [doi: 10.1109/bibm.2015.7359732]

267. Kheterpal S, Woodrum DT, Tremper KK. Too much of a good thing is wonderful: observational data for perioperative research. Anesthesiology 2009 Dec;111(6):1183-1184. [doi: 10.1097/ALN.0b013e3181c1496c] [Medline: 19934857]

268. Raman PW, Storm P, Lilly J, Mason J, Heath A, Felmeister A, et al. Cavatica- a pediatric genomic cloud empowering data discovery through the pediatric brain tumor atlas. In: Neuro-Oncology. 2017 Presented at: 4th Biennial Conference on

Pediatric Neuro-Oncology Basic and Translational Research; June 15-16, 2017; New York City, NY, USA p. iv21. [doi: 10.1093/neuonc/nox083.086]

269. Raman P, Resnick AC, Storm PB, Mueller S, Schultz N, Cerami E, et al. PedcBioPortal: a Cancer Data Visualization Tool for Integrative Pediatric Cancer Analyses. In: 21st Annual Scientific Meeting and Education Day of the Society for Neuro-Oncology. 2016 Presented at: Neuro-Oncology'16; November 17-20, 2016; Scottsdale, Arizona. [doi: 10.1093/neuonc/now212.337]

270. Murphy SN, Weber G, Mendis M, Gainer V, Chueh HC, Churchill S, et al. Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2). J Am Med Inform Assoc 2010;17(2):124-130 [FREE Full text] [doi: 10.1136/jamia.2009.000893] [Medline: 20190053]

271. Segagni D, Tibollo V, Dagliati A, Perinati L, Zambelli A, Priori S, et al. The ONCO-I2b2 project: integrating biobank information and clinical data to support translational research in oncology. Stud Health Technol Inform 2011;169:887-891. [Medline: 21893874]

272. Bodenreider O, Cornet R, Vreeman DJ. Recent developments in clinical terminologies - SNOMED CT, LOINC, and RxNorm. Yearb Med Inform 2018 Aug;27(1):129-139 [FREE Full text] [doi: 10.1055/s-0038-1667077] [Medline: 30157516]

273. Cimino JJ. High-quality, standard, controlled healthcare terminologies come of age. Methods Inf Med 2011;50(2):101-104 [FREE Full text] [Medline: 21416108]

274. Structure and Principles. WHO Collaborating Centre for Drug Statistics and Methodology. 2020. URL: https://www.whocc.no/atc/structure_and_principles/ [accessed 2020-01-01]

275. Robinson PN, Köhler S, Bauer S, Seelow D, Horn D, Mundlos S. The human phenotype ontology: a tool for annotating and analyzing human hereditary disease. Am J Hum Genet 2008 Nov;83(5):610-615 [FREE Full text] [doi: 10.1016/j.ajhg.2008.09.017] [Medline: 18950739]

276. The Gene Ontology Consortium. Expansion of the gene ontology knowledgebase and resources. Nucleic Acids Res 2017 Jan 4;45(D1):D331-D338 [FREE Full text] [doi: 10.1093/nar/gkw1108] [Medline: 27899567]

277. Murphy SN, Mendis M, Hackett K, Kuttan R, Pan W, Phillips LC, et al. Architecture of the open-source clinical research chart from informatics for integrating biology and the bedside. AMIA Annu Symp Proc 2007 Oct 11:548-552 [FREE Full text] [Medline: 18693896]

278. Scheufele E, Aronzon D, Coopersmith R, McDuffie MT, Kapoor M, Uhrich CA, et al. TranSMART: an open source knowledge management and high content data analytics platform. AMIA Jt Summits Transl Sci Proc 2014;2014:96-101 [FREE Full text] [Medline: 25717408]

279. Hripcsak G, Duke JD, Shah NH, Reich CG, Huser V, Schuemie MJ, et al. Observational health data sciences and informatics (OHDSI): opportunities for observational researchers. Stud Health Technol Inform 2015;216:574-578 [FREE Full text] [Medline: 26262116]

280. Mohammed-Rajput NA, Smith DC, Mamlin B, Biondich P, Doebbeling BN, Open MRS Collaborative Investigators. OpenMRS, a global medical records system collaborative: factors influencing successful implementation. AMIA Annu Symp Proc 2011;2011:960-968 [FREE Full text] [Medline: 22195155]

281. Epstein R, Hofer I, Salari V, Gabel E. Successful implementation of a perioperative data warehouse using another hospital's published specification from epic's electronic health record system. Anesth Analg 2020 Apr 22:- epub ahead of print. [doi: 10.1213/ANE.0000000000004806] [Medline: 32332291]

282. Rizi SA, Roudsari A. Development of a public health reporting data warehouse: lessons learned. Stud Health Technol Inform 2013;192:861-865. [Medline: 23920680]

283. STARR Tools. Stanford Medicine. URL: http://med.stanford.edu/starr-tools.html [accessed 2018-03-26]

284. Wharton Research Data Services. URL: https://wrds-www.wharton.upenn.edu/ [accessed 2018-03-27]

285. MPOG – Multicenter Perioperative Outcomes Group. URL: https://mpog.org/ [accessed 2018-04-03]

286. Boston University Medical Campus and Boston Medical Center. URL: http://www.bumc.bu.edu/ohra/using-bmc-and-chc-data-for-research-purposes/ [accessed 2018-04-12]

287. Children's Hospital of Philadelphia® Center for Data-Driven Discovery in Biomedicine. URL: https://d3b.center/aboutd3b/history/ [accessed 2018-04-17]

288. Lex A, Gehlenborg N, Strobelt H, Vuillemot R, Pfister H. UpSet: visualization of intersecting sets. IEEE Trans Vis Comput Graph 2014 Dec;20(12):1983-1992 [FREE Full text] [doi: 10.1109/TVCG.2014.2346248] [Medline: 26356912]

289. Gallego B, Walter SR, Day RO, Dunn AG, Sivaraman V, Shah N, et al. Bringing cohort studies to the bedside: framework for a 'green button' to support clinical decision-making. J Comp Eff Res 2015 May;4(3):191-197. [doi: 10.2217/cer.15.12] [Medline: 25959863]

290. Longhurst CA, Harrington RA, Shah NH. A 'green button' for using aggregate patient data at the point of care. Health Aff (Millwood) 2014 Jul;33(7):1229-1235. [doi: 10.1377/hlthaff.2014.0099] [Medline: 25006150]

291. Gombar S, Callahan A, Califf R, Harrington R, Shah NH. It is time to learn from patients like mine. NPJ Digit Med 2019;2:16 [FREE Full text] [doi: 10.1038/s41746-019-0091-3] [Medline: 31304364]

292. Schuler A, Callahan A, Jung K, Shah NH. Performing an informatics consult: methods and challenges. J Am Coll Radiol 2018 Mar;15(3 Pt B):563-568 [FREE Full text] [doi: 10.1016/j.jacr.2017.12.023] [Medline: 29396125]

Abbreviations

BMC: BioMed Central
BRP: biorepository portal
BTRIS: biomedical translational research information system
CARPEM: CAncer Research for PErsonalized Medicine
CDM: common data model
CDSS: clinical decision support system
CLB: Léon Bérard Cancer Center
DW4TR: Data Warehouse for Translational Research
EHR: electronic health record
ETL: extraction, transformation, and loading
FURTHeR: Federated Utah Research and Translational Health Electronic Repository
HaMSTR: Hanover Medical School Translational Research Framework
HSSC: Health Science, South Carolina
ICD: International Classification of Diseases
IDR: integrated data repository
IEEE Xplore: Institute of Electrical and Electronics Engineers Xplore
i2b2: Informatics for Integrating Biology and the Bedside
MEDLINE: Medical Literature Analysis and Retrieval System Online
METEOR: Methodist Environment for Translational Enhancement and Outcome Research
MIDH: Maternal and Infant Data Hub
MOSAIC: models and simulation techniques for discovering diabetes-related factors
NLP: natural language processing
OMOP: Observational Medical Outcomes Partnership
SNOMED-CT: systematized nomenclature of medicine-clinical terms
STARR: STAnford Research Repository
STRIDE: Stanford Translational Research Integrated Database Environment
TRC: Translational Research Center

XSL•FO
RenderX