

Characteristics Desired in Clinical Data Warehouse for Biomedical Research

Soo-Yong Shin, PhD¹, Woo Sung Kim, MD, PhD^{1,2}, Jae-Ho Lee, MD, PhD^{1,3,4}

¹Department of Biomedical Informatics, Asan Medical Center, Seoul, Korea; ²Departments of Pulmonary and Critical Care Medicine, ³Department of Emergency Medicine, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Korea; ⁴Division of General Internal Medicine, Brigham and Women's Hospital, Boston, MA, USA

Objectives: Due to the unique characteristics of clinical data, clinical data warehouses (CDWs) have not been successful so far. Specifically, the use of CDWs for biomedical research has been relatively unsuccessful thus far. The characteristics necessary for the successful implementation and operation of a CDW for biomedical research have not clearly defined yet. **Methods:** Three examples of CDWs were reviewed: a multipurpose CDW in a hospital, a CDW for independent multi-institutional research, and a CDW for research use in an institution. After reviewing the three CDW examples, we propose some key characteristics needed in a CDW for biomedical research. **Results:** A CDW for research should include an honest broker system and an Institutional Review Board approval interface to comply with governmental regulations. It should also include a simple query interface, an anonymized data review tool, and a data extraction tool. Also, it should be a biomedical research platform for data repository use as well as data analysis. **Conclusions:** The proposed characteristics desired in a CDW may have limited transfer value to organizations in other countries. However, these analysis results are still valid in Korea, and we have developed clinical research data warehouse based on these desiderata.

Keywords: Information Storage and Retrieval, Research Design, Biomedical Research, Privacy, Research Ethics

I. Introduction

There are many ways to define a data warehouse (DW) due to its widespread adoption [1-4]; a good working definition

Submitted: January 13, 2014

Revised: March 22, 2014

Accepted: April 6, 2014

Corresponding Author

Jae-Ho Lee, MD, PhD

Department of Emergency Medicine, Asan Medical Center, 88 Olympic-ro 43-gil, Songpa-gu, Seoul 138-736, Korea. Tel: +82-2-3010-3350, Fax: +82-2-3010-8126, E-mail: rufiji@gmail.com

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

© 2014 The Korean Society of Medical Informatics

of a DW is a dedicated computer system or database that consolidates subject-oriented, time-variant, and non-volatile data from multiple sources to support decision-making processes. Recently, DWs have become invaluable resources in various domains, and they are used to analyze trends over time or to extract valuable information.

Based on the success of DWs in other fields, hospitals have started to adopt a DW system. A survey indicated that the adoption rate of DWs in Clinical and Translational Science Award (CTSA) institutions has increased from 64% (18 of 28 institutions) in 2008 to 86% (30 of 35) in 2010 [5]. DWs in hospitals, which are usually called clinical data warehouses (CDWs) [4,6,7], are used for various purposes, including administration, management, clinical practice, and research. These can be categorized as either conventional usage or hospital-specific usage. Conventional usage includes administration, operation, and management. Therefore, such a DW in a hospital is usually called an enterprise DW. It is

an earlier type of DW in hospitals [8]. The hospital-specific usage consists of clinical practice, quality improvement, and biomedical research [9]. However, research usage cannot be efficiently supported by conventional DW technology due to the complexity and heterogeneity of clinical and research data [10]. In addition, as electronic health records (EHRs) have been adopted in many hospitals, research using EHR data has been highlighted recently [11,12]. EHR is the legacy and live system that generates the raw data used to record clinical data of patients. Governmental regulations, such as the requirement of de-identification, limits the direct use of EHR data [13]. Also, EHR data must be extracted, transformed, and loaded to other databases for analysis. CDWs integrate and reconstruct raw data from EHRs and other legacy systems for analysis, and they can adopt several interfaces needed for research compliance. Therefore, the importance of CDWs in accessing and analyzing EHR data for research has been increasing.

Until now, CDWs have not been successful for hospital management compared to their promise because conventional DWs do not satisfy the needs of some unique hospital environments [9,10]. For example, an intensive care unit has many sets of continuous patient monitoring data and point-of-care device data, so their integration requires special concerns [14]. Radiology and other image data warehouses also require special features [4]. Recently, the term “big data” has also been introduced into DWs in the biomedicine field [15]. Therefore, we need to develop a special type of DW or CRDW for research to satisfy a hospital’s individual needs, not just incorporate a conventional DW technology. However, the characteristics of CDWs for research have not been discussed widely and have not been well differentiated from conventional CDWs, although the characteristics of CDW were well described by Huser and Cimino [16].

In this paper, we focused on how to build a CDW for research; we use the term, *clinical research data warehouse* (CRDW) because one of the most important reasons to build a CDW is to support research.

II. Methods

To clarify the key elements needed in a CRDW, we reviewed the various types of CRDW-related terms such as the CDW, the research data warehouse and the integrated data repository, defined CRDW-related terms, and compared them. First, we searched PubMed with the keywords “clinical data warehouse”, and we found 89 articles in total. We classified these into three types, namely, research usage of CDWs, multi-institutional research data warehouses, and single

institution research CDWs. From the three kinds of CDWs, two researchers selected the following three well-known CDWs based on their own knowledge to determine both the issues and benefits of current CRDWs: the Ohio State University Medical Center (OSUMC) Information Warehouse (IW), a research usage of CDWs in a hospital [17,18]; the Informatics for Integrating Biology and the Bedside (i2b2), a DW for an independent multi-institutional research [19,20]; and the Stanford Translational Research Integrated Database Environment (STRIDE), a CDW of single-institution research in a hospital [21,22].

From the literature review, CDW definitions, and a comparison of CDW cases, we propose some essential characteristics desired in a CRDW.

III. Results

1. Clinical Research Data Warehouse and Its Related Terminologies

Usually, the term CDW refers to an *enterprise data warehouse in a hospital*, which is used for administration, management, clinical practice, and research [23]. Here, we use the term CRDW to refer to *a data warehouse in a hospital or other organization that is used only for research* [24]. Therefore, a CDW is a place where healthcare providers can gain access to clinical data gathered during the patient care process [25] that may provide information for users in diverse areas [17-22]. The data in a CDW include any information related to patient care, such as specific demographics, vital signs, input and output data recorded for the patient, treatments and procedures performed, supplies used, and costs associated with the patient’s care.

The differences between a DW in a hospital and a DW in other domains were well described by Inmon [9]. He claimed that the information needs of medicine and healthcare are fundamentally different than those of other areas, and these fundamental differences in information gathering and storage make it difficult to implement successful data warehousing in hospitals [9]. The different perspectives for data warehousing between the healthcare domain and other domains are summarized in Table 1. First, each transaction or encounter in healthcare is relatively unique, as opposed to the business world in which each transaction is very repetitive. The data will even have different characteristics for each department in the hospital, including the emergency room, operation room, or the clinics. The second difference is in the types of data. Most healthcare data include textual descriptions of the various medical encounters of a patient. Additionally, data warehousing requires metadata or a common vocabulary, which

Table 1. Differences between a data warehouse of the healthcare domain and those of other domains

Category	Health care	Other domain (bank, finance)
Transaction	Unique	Repetitive
Data type	Mixed (text, code, number)	Number
Common vocabulary	Normalization required	Existing
Time value of information	Significant	Not significant
External category	Essential	Not significant

Summarized and modified from Inmon [9].

is already well defined and used in other domains, such as banking or finance. Although there are many common vocabularies and related standards in medicine, the usage rate of common jargon in a hospital is low.

Another major difference is that DWs in hospitals are primarily used as research data repositories. A recent CTSA survey reported that CDWs have shifted from a primarily administrative focus to a role incorporating more of the data contained in electronic medical records and the support systems for biomedical research [5]. Therefore, there is some precedent for CRDWs in ideas like research data warehouses and integrated data repositories [5]. The descriptions of both terms highlight the integration of multiple data sources, including hospital-generated data and genetic data, as essential for research.

2. Representative Examples of Clinical Research Data Warehouses

When we reviewed the current CDWs or CRDWs, we found that the current data warehouses that support research can be classified into three different categories, namely, research usage of CDW, multi-institutional research data warehouse, and single-institution research data warehouse. The first type of CDW can support research as well as clinical practice and management. Representative examples are the OSUMC IW and Emory Healthcare [17,18,26]. Usually, this type of CDW has little institutional conflict and is able to gather information from clinical data sources since it supports hospital administration and business. Data marts are implemented to allow for research data search and extraction.

Multi-institutional research data warehouses and single-hospital research data warehouses are designed for research purposes only, not for management. Lots of hospitals in the United States have adopted independent multi-institutional CRDW projects using the i2b2 platform [19,27]. There are over 60 hospitals that operate CDWs based on the i2b2 platform, including Cincinnati Children's Hospital [20,27]. This approach has several benefits by using an open-source platform, such as reducing implementation costs and guarantee-

ing the success of the project since there are many reference sites. Using the i2b2 architecture, their system integrates data from multiple sources, combines research data with clinical data, focuses on cohorts and patient populations, and has the potential for de-identified queries.

The representative example of single hospital research data warehouse is Stanford University's STRIDE [21,22]. Since STRIDE is implemented in a university, not a hospital, the STRIDE project itself was prioritized independently, regardless of the complexities of hospital IT, and it can easily implement the necessary regulations.

1) Ohio State University Medical Center Information Warehouse

The OSUMC IW seems to be a CDW drawing from diverse and disparate information systems throughout OSUMC [17,18]. Though it has been used for diverse areas including business, clinical, and research, we reviewed the OSUMC IW to determine the characteristics of a CDW that make it useful specifically for research. The OSUMC IW has little institutional conflict and is thus able to gather information from diverse clinical data sources. Data marts have been implemented to allow for research data extraction and getting the information out to users. In addition, data from several external sources are regularly incorporated into the OSUMC IW to assist in translational research. Therefore, this IW is a core asset that facilitates translational research and advances personalized healthcare. In 2006, the Ohio State University Institutional Review Board (IRB) approved a protocol recognizing the IW as an "honest broker" of clinical data, meaning that the IW can provide de-identified, limited, and coded data for use in research.

2) Cincinnati Children's clinical research data warehouse

Cincinnati Children's CRDW is based on the i2b2 architecture, which is a research project designed to build an institutional-independent research data repository [19,20]. The Cincinnati Children's Hospital adopted the open-source platform i2b2 to reduce their costs and guarantee the success of

the project. Using i2b2 architecture, their system integrates data from multiple sources, combines research data with clinical data, focuses on cohorts and patient populations, and has the potential for de-identified queries. The Cincinnati Children's CRDW includes patient demographics, diagnoses, procedures and medication orders for all inpatient and ambulatory encounters, including lab results, discharge summaries, and reports from pathology, cardiology, and radiology. This CRDW can integrate researchers' own data by serving as a platform for research registries. This approach

has several benefits. First, many of the queries asked of a registry are essentially forms of cohort identification. Second, integrating the registration information removes the need to load the data into multiple database systems or have users manually re-enter the relevant EHR data [19].

3) Stanford Translational Research Integrated Database Environment

STRIDE is a research and development project at Stanford University meant to create a standards-based informatics

Table 2. A comparison of CRDW and CDW

Category	CRDW	CDW (data warehouse in a hospital)
Aim	Clinical & translational research	General, hospital management
Need	Data extraction & review De-identification	Integrated report
System	Data extraction system Data (chart) review system Database	Data extraction system Database
Interface	IRB approval process Biologic specimen search	None
Essential function	Research design Cohort discovery De-identified data review Data extraction	Reporting & review Ad-hoc query
Subject area	Research data (e.g., disease, laboratory test, medication)	Hospital management (e.g., nursing practice)
User	Researchers	Administrative staffs
Internal data	EHR data CPOE data LIMS data Bio-specimen data	All in-hospital data
External data	e-CRF data Public research database Researcher-owned database	None
Metadata	Mandatory	Selective
Privacy	De-identifiable data Identifiable data (after IRB approval)	Identifiable data
Location	Hospital Research center Medical school	Hospital
Project priority	High	Low
Implementation easiness	Hard	Easy
Regulation	Easy to follow	Hard to follow

CRDW: clinical research data warehouse, CDW: clinical data warehouse, EHR: Electronic Health Record, CPOE: computerized physician order entry, LIMS: laboratory information management system, e-CRF: electronic case report form, IRB: Institutional Review Board.

platform that supports clinical and translational research [21,22]. Because STRIDE is implemented in a university rather than a hospital, the STRIDE project itself was prioritized independently, regardless of the complexities of hospital IT, and it can easily implement the necessary regulations. STRIDE consists of three main databases, including a clinical data warehouse, a bio-specimen database, and a research database. Working upon those database systems are an anonymous cohort identification tool, a patient cohort data review tool, clinical data extraction, research data management, and bio-specimen data management. STRIDE is an IRB approved project, and some processes, such as data extraction require IRB approval.

3. Comparisons between a Clinical Research Data Warehouse and an Enterprise Data Warehouse in a Hospital Setting

The results of a comparison between a CRDW and a CDW are summarized in Table 2. Essentially, CDWs are similar to conventional DWs, though there are some differences (as described in Table 1). However, a CRDW has significantly different characteristics. The purpose of a CRDW is to aid clinical and translational studies, not hospital management [5]. All data in a CRDW should be anonymized to protect the patients’ privacy. Additionally, IRB approval is required to process and search interfaces for ad-hoc queries. The most essential functions of a CRDW are research design, chart review, and data extraction, so it focuses on tasks like cohort identification and hypothesis generation and analysis [11,28]. Therefore, research data are the main subject area, though all related data and processes for clinical practice and research can also be incorporated. The primary sources of data for a CRDW are hospital information systems, such as EHR,

laboratory information management system (LIMS), and computerized physician order entry (CPOE). Clinical trial registries and other researcher-owned databases (i.e., cohort or genomic data) should be integrated into the CRDW as well [18-21]. Other public research databases should also be interfaced to promote research. Its users are researchers and clinicians, not hospital administrative staff members. However, metadata and structured formats are still required for the easy and accurate retrieval of data.

A CRDW can be located within a hospital, research center, or medical school, although a hospital should also contain a CDW for administration purposes. However, the location of a CRDW is problematic. A CRDW in a hospital is likely to rank lower in priority schemes than the more urgent hospital IT projects, making it likely to be neglected. A hospital also has to acquire additional funding for a CRDW project. However, a CRDW that is not located in a hospital requires a long developmental period and intra-institutional agreements must be made for clinical data to be obtained. Still, a CRDW outside of a hospital has some merits. It is much easier to incorporate non-hospital public data sources. Most importantly, the necessary regulations, such as the Health Insurance Portability and Accountability Act (HIPAA) compliance, IRB approval, and the honest broker system could be implemented more easily.

IV. Discussion

Based on the results of our comparisons, the ideal characteristics of a CRDW are defined in Table 3. The ultimate goal of a CRDW might be to serve as a biomedical research platform that is useful for data analysis in addition to functioning as a data repository. Therefore, interfaces for queries, honest bro-

Table 3. Desired characteristics of a clinical research data warehouse

Key element	Explanation	Remark
Honest broker	Protecting patient privacy based on hospital policy and HIPAA compliance	De-identification
Query interface	Direct ad-hoc queries Data analysis tools	Cohort discovery Hypothesis design and analysis
Chart review	Reviewing the de-identified EHR charts	
Data extraction	Extracting the necessary (de-identifiable) data	DRM module for access control Virtual desktop environment
IRB interface	Research approvals Waivers	

HIPAA: Health Insurance Portability and Accountability Act, EHR: Electronic Health Record, DRM: digital rights management, IRB: Institutional Review Board.

ker services, data extraction, chart review, and IRB approval may be the key elements.

The honest broker is an individual/organization/system which acts on the tissue bank and database [29]. The honest broker protects patients' privacy based on institutional policy and government regulations, such as HIPAA, through de-identification [13,17,29,30]. It de-identifies all of the necessary patient-related data and serves as an interface to extract the requested bio-specimen samples or clinical data. The identifiable data can be extracted with IRB approval. Therefore, an interface with the IRB system should also be prepared. If an electronic IRB system may be used, the research approval or waiver should be automatically transferred into the CRDW. In the case of a paper-based IRB, the necessary information should be entered into the database by the researchers. For research hypothesis design and analysis, an easy interface for queries and data review tools should be implemented. A query interface allows a user to find the candidate number of a study group and to search the necessary bio-specimen samples. By reviewing query results, researchers could design study hypotheses. A chart review tool is also needed to confirm the cohort size and analyze the hypothesis manually, displaying the de-identified patient data; however, if the IRB approves, the necessary identifiable data could be delivered. A data extraction tool is necessary to obtain desired data for further use. For the extraction of identifiable data, digital rights management tools, which control data access, should be considered to protect privacy. Alternatively, a virtual desktop environment could be used. If the virtual desktop environment used is based on cloud computing technology, several security concerns can be eas-

ily solved. The cloud can also offer the powerful computing resources required to handle genomic data processing, such as next-generation sequencing. Finally, a CRDW needs to integrate data analysis programs by supporting the seamless transfer of extracted data. The analysis programs could be statistical packages and machine learning toolkits.

Figure 1 shows a schematic diagram of a CRDW containing the key elements described above. Clinical and research data and processes can all be incorporated into this CRDW. The clinical data come from hospital information systems, including EHR, LIMS, and CPOE, and the research data are from electronic case reports or the users' research database. Requirements of patient safety, privacy and security should be implemented within the system. The CRDW is accessed by research tools for data extraction, a chart review system, data mining, and other analyses that allow the data to be better understood and used. The mandatory features of a CRDW are easy interfaces for queries and data extraction. An easy query interface helps a user design a hypothesis by finding the number of a study group and possible clinical data. Data extraction is important to test a hypothesis by analyzing extracted data. The remaining three characteristics, namely, an honest broker system, chart review, and IRB interface help a user to perform research more conveniently while obeying the necessary regulations.

Many older DWs in hospitals focus on hospital management, not on clinical research. However, the number of requests from researchers seeking access to CDWs has been increasing. Here, we describe a set of desired characteristics for a CRDW used for research purposes. A CRDW should include an honest broker system and an IRB approval interface

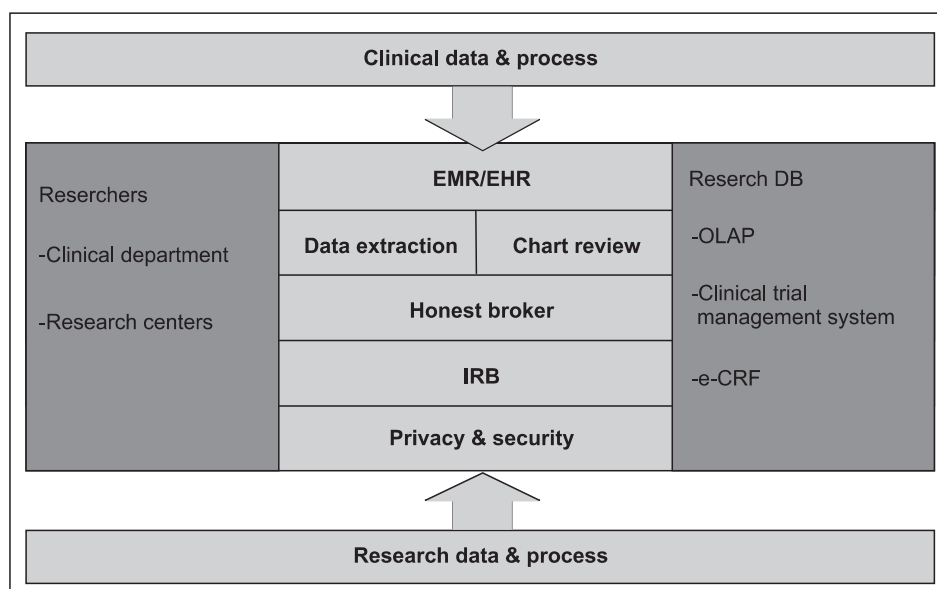


Figure 1. A diagram of an ideal clinical research data warehouse. EMR: Electronic Medical Record, EHR: Electronic Health Record, IRB: Institutional Review Board, DB: database, OLAP: online analytic processing, e-CRF: electronic case report form.

to comply with governmental regulations, as well as a simple query interface, an anonymized data review tool, and a data extraction tool. A CRDW should serve as a biomedical research platform for data analysis as well as a data repository.

However, CRDWs have diverse development obstacles, including funding and sponsorship, data ownership and access issues, and staffing issues [5]. To overcome these obstacles, open-source systems are gaining popularity over “in-house” systems [5]. The use of “in-house” developed front-facing business intelligence tools has decreased, while the adoption of open-source data warehouse tools, such as i2b2, has increased because it reduces costs and guarantees the success of a CRDW project. However, many CRDW projects are still based on “in-house” systems because open-source systems also need customization to satisfy the unique requirements of each hospital.

Conflict of Interest

No potential conflict of interest relevant to this article was reported.

References

- Inmon WH. Tech topic: what is a data warehouse. [place unknown]: Prism Solutions; 1995.
- Inmon WH, Derek S, Neushloss G. DW 2.0: the architecture for the next generation of data warehousing. Amsterdam: Morgan Kaufman; 2008.
- Chaudhuri S, Dayal U. An overview of data warehousing and OLAP technology. *ACM SIGMOD* 1997;26(1):65-74.
- Rubin DL, Desser TS. A data warehouse for integrating radiologic and pathologic data. *J Am Coll Radiol* 2008;5(3):210-7.
- 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;19(e1):e119-24.
- Prather JC, Lobach DF, Goodwin LK, Hales JW, Hage ML, Hammond WE. Medical data mining: knowledge discovery in a clinical data warehouse. *Proc AMIA Annu Symp* 1997;1997:101-5.
- Ledbetter CS, Morgan MW. Toward best practice: leveraging the electronic patient record as a clinical data warehouse. *J Healthc Inf Manag* 2001;15(2):119-31.
- Dewitt JG, Hampton PM. Development of a data warehouse at an academic health system: knowing a place for the first time. *Acad Med* 2005;80(11):1019-25.
- Inmon B. Data warehousing in a healthcare environment. Pittsburgh (PA): The Data Administration Newsletter; 2007 [cited at 2014 Mar 20]. Available from: <http://www.tdan.com/view-articles/4584/>.
- Evans RS, Lloyd JF, Pierce LA. Clinical use of an enterprise data warehouse. *Proc AMIA Annu Symp* 2012;2012:189-98.
- Embi PJ, Kaufman SE, Payne P. Biomedical informatics and outcomes research: enabling knowledge-driven health care. *Circulation* 2009;120(23):2393-9.
- Jensen PB, Jensen LJ, Brunak S. Mining electronic health records: towards better research applications and clinical care. *Nat Rev Genet* 2012;13(6):395-405.
- US Department of Health & Human Services. Guidance regarding methods for de-identification of protected health information in accordance with the Health Insurance Portability and Accountability Act (HIPAA) privacy rule. Washington (DC): US Department of Health & Human Service; c2013 [cited at 2013 Apr 12]. Available from: <http://www.hhs.gov/ocr/privacy/hipaa/understanding/coveridentities/De-identification/guidance.html>.
- 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;105(1):22-30.
- Bernstam EV, Hersh WR, Johnson SB, Chute CG, Nguyen H, Sim I, et al. Synergies and distinctions between computational disciplines in biomedical research: perspective from the Clinical and Translational Science Award programs. *Acad Med* 2009;84(7):964-70.
- Huser V, Cimino JJ. Desiderata for healthcare integrated data repositories based on architectural comparison of three public repositories. *Proc AMIA Annu Symp* 2013;2013:648-56.
- Liu J, Erdal S, Silvey SA, Ding J, Riedel JD, Marsh CB, et al. Toward a fully de-identified biomedical information warehouse. *Proc AMIA Annu Symp* 2009;2009:370-4.
- Kamal J, Liu J, Ostrander M, Santangelo J, Dyta R, Rogers P, et al. Information warehouse: a comprehensive informatics platform for business, clinical, and research applications. *Proc AMIA Annu Symp* 2010;2010:452-6.
- 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):124-30.
- Cincinnati Children's Hospital Medical Center. i2b2 Research data warehouse. Cincinnati (OH): Cincinnati Children's Hospital Medical Center; c2011 [cited at 2014

- Jan 8]. Available from: <https://i2b2.cchmc.org/>.
21. Lowe HJ, Ferris TA, Hernandez PM, Weber SC. STRIDE: an integrated standards-based translational research informatics platform. *Proc AMIA Annu Symp* 2009;2009:391-5.
 22. 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. *Proc AMIA Annu Symp* 2009;2009:244-8.
 23. Grant A, Moshyk A, Diab H, Caron P, de Lorenzi F, Bisson G, et al. Integrating feedback from a clinical data warehouse into practice organisation. *Int J Med Inform* 2006;75(3-4):232-9.
 24. The University of Chicago, Center for Research Informatics. Clinical Research Data Warehouse (CRDW). Chicago (IL): The University of Chicago; c2014 [cited at 2014 Mar 21]. Available from: http://cri.uchicago.edu/?page_id=772/.
 25. Gray GW. Challenges of building clinical data analysis solutions. *J Crit Care* 2004;19(4):264-70.
 26. Emory University. Clinical data warehouse - healthcare. Atlanta (GA): Emory University; c2014 [cited at 2014 Mar 21]. Available from: http://it.emory.edu/catalog/ehc_clinical_data_warehouse/.
 27. Kohane IS, Churchill SE, Murphy SN. A translational engine at the national scale: informatics for integrating biology and the bedside. *J Am Med Inform Assoc* 2012;19(2):181-5.
 28. Kahn MG, Weng C. Clinical research informatics: a conceptual perspective. *J Am Med Inform Assoc* 2012;19(e1):e36-e42.
 29. Dhir R, Patel AA, Winters S, Bisceglia M, Swanson D, Aamodt R, et al. A multidisciplinary approach to honest broker services for tissue banks and clinical data: a pragmatic and practical model. *Cancer* 2008;113(7):1705-15.
 30. El Emam K. Methods for the de-identification of electronic health records for genomic research. *Genome Med* 2011;3(4):25.