Indian Journal of Science and Technology, *Vol 10(13), DOI: 10.17485/ijst/2017/v10i13/112374, April 2017* *ISSN (Print) : 0974-6846*
ISSN (Online) : 0974-5645

Big Data in Health: New Challenges and New Solutions in Data Management (A Lifecycle Review)

Hamideh Ehtesham[1], Reza Safdari[1]* and Shahram Tahmasebian[2]

[1]Department of Health Information Management, School of Allied Medical Sciences, Tehran University of Medical Sciences, Tehran, District 6, Iran; h-ehtesham@razi.tums.ac.ir, rsafdari@tums.ac.ir
[2]School of Medicine, Shahrekord University of Medical Sciences, Shahrekord, Iran; stahmasebian@gmail.com

Abstract:

Background: Collecting and analyzing large volumes of disparate datasets are of major challenges arisen from the emergence of Big Data in the field of health. But the technology of Big Data is also associated with promising opportunities which can provide improvement of performance and facilitation of innovation in organizations. **Objective:** Since determining the lifetime is a practical approach regarding recognition of a phenomenon and its management, this paper aimed at identifying the challenges and opportunities of managing Big Data in the area of health at different stages of the lifecycle of Big Data. **Methodology:** This article is a structured review. After an initial review, 6 phases was detected in the lifecycle of Big Data, then the processes of traditional data were briefly reviewed in each phase, and the challenges associated with the emergence of Big Data and the solutions for their optimal management were discussed. **Results:** This study offers a broad over view of the advent of Big Data in the health sector, and provides a clear and accurate picture of the processes before and after its emergence through a comparative-based survey in each phase. The article points out innovations and modern methods of collection, pre-processing, and analysis of Big Data as well as the process of data extracting. It also describes cloud computing applications in the storage and release of Big Data. **Conclusions**: Our findings indicate that management of Big Data in health, based on its lifecycle, is resourceful for managers and policy-makers, in order to benefit from the technological features of Big Data with a managerial approach, to evaluate challenges, to apply innovative solutions at each phase of Big Data maturation, and to advance towards a new level of innovation, competitiveness, and productivity.

Keywords: Big Data, Lifecycle Stages, Health

1. Introduction

Big Data point out the exponential growth of data in both structured and unstructured forms whose timely storage, access, and analysis have impressive results in enhancing organizational performance; but their management, control, and processing are beyond the capabilities of traditional software in the expected time frame[1]. Prediction is the core of Big Data and is performed through mathematical algorithms[2]. New technologies of Big Data are dealing with integration of island information systems of organizations in order to reduce hardware and processing costs and proof the value of the data[1]. In this way, they set the stage for improvement of performance, facilitation of innovation in products and business models services, and decision support[3].

Big Data technology describes 3 Vs in the course of data including Volume, Velocity, and Variety of data[1]. In the field of health sciences, extensive volume refers to both the number of data (clinical documents) and the size of data (the entire genome). Velocity is related to the rate of

**Author for correspondence*

change in data and its immediate analysis in the course of information. Variety includes the complexity of different data sources. Some have added a fourth V to authenticate the Veracity of data extracted from incomplete and unreliable, or unequal resources. Finally, the fifth V emphasis the Value of health data in terms of time lines and accessibility to health care when needed[4].

However, beyond the concept of Big Data, dramatic changes in the business are associated with challenges, some of which are the same as for regular data, but the level of focus on Big Data has changed. Because Big Data is not merely data, it is further a complete framework including the data per se, storage, format, and modes of supply, processing, and analysis[5].

Management of data is a significant challenge for solving some key problems such as organizing Big Data for indexing, searching, and processing in an appropriate manner, and proper implementation of availability and security such as performance, scalability, and fault tolerance in health organizations; these challenges require logical solutions[6].

Identifying and prioritizing potential opportunities for Big Data, pay attention to privacy concerns, promoting data transparency and communicating results are important strategies in managing Big Data in health system[7].

Inclusion of the system lifecycle charter plan and emphasis on change management processes can improve prediction of path and provide reliable support for changes intheplan[8]. Lifecycle assessment studies, as a way to assess the current performance of management systems, and discussion on their strengths and weaknesses, as well as designing an evolutionary perspective are useful subjects[9].

When facing new challenges, the process should be evaluated at the components level in order to attain mastery in all their aspects and to make the best decision; therefore, this study draws the lifecycle of Big Data in health and then provides an overview of the problems and possible solutions in the path of Big Data maturation with an emphasis on health care so as to improve data management. Furthermore, prediction and visualization of future by designing and testing scenarios in Big Data analysis is explored; and the paper has also focused on the use of Big Data technology to transform management and to progress towards a new level of innovation, competitiveness, and productivity.

2. Methods

This study is a structured review of data management in dealing with big data in health. In response to the basic needs of Big Data management in the field of health and the need to identify the challenges and using their potential benefits, as well as to increase the demand for data-driven decisions, after an initial review, 6 phases of Big Data lifecycle were detected at first, and then the processes of traditional data were briefly reviewed in each phase, the challenges associated with the emergence of Big Data were pointed out, and a clear and accurate picture of effective actions and measures were presented for promoting management of Big Data in health providing organizations.

3. Results

The findings of the study will be presented in two parts:

1. Drawing Big Data lifecycle
2. Comparison of data processes before and after the advent of Big Data in the field of health.

3.1 Lifecycle of Big Data

Data are the bases of any investigation and decision-making, and other tasks are laid upon them. Therefore, it is necessary to be a dept at producing good data, properly processing and analyzing of data, and the way of their presentation[10]. The volume of enterprise data is rising with an extraordinary rate, given the continued use of communication, network technologies, marketing, and so on. Incorporating technology in a way that makes it possible to understand and manage the lifecycle of data for an organization is a promising approach[11].

Lifecycle thinking and lifecycle assessment are scientific methods that support basic policies and data-driven decisions, and inhibit the transfer of problems from one phase of the lifecycle to the next or resolve the problems in one phase to prevent problems from recurring in other phases[12]. Since the data are valuable assets beyond the immediate needs, data must be managed throughout the lifecycle. Data Lifecycle Management (DLM) is a policy-based approach for managing data flow in information systems throughout its lifecycle, from preliminary creation and storage until they become obsolete and deleted.

According to Demchenko, the general lifecycle of scientific data is a combination of successive phases including planning he experiment (project), data collection and processing, discussion, feedback, and archiving. He suggests a four-phase lifecycle for Big Data as Data collection and registration; Data filter/enrich, classification; Data analytics, modeling, prediction; and Data delivery and visualization[13]. Taylor has also presented the phases of data conceptualization, data collection, data distribution, data discovery, data analysis, and data repurposing for data lifecycle[14]. In another article, data lifecycle phases were outlined as concept, collection, processing, distribution, discovery, and analysis[15].

The general lifecycle proposed using Big Data technology and terminology includes Big Data accumulation, Big Data integration and cleaning, Big Data analysis, Big Data storing, Big Data sharing and publishing, and Big Data recovery and exploration1. Given its comprehensiveness, this framework is used in this paper for drawing the lifecycle of Big Data. Figure 1 display the lifecycle of Big Data in health.

Figure 1. Lifecycle of big data in health.

3.2 Comparative-Based Survey

This section is devoted to a review of various phases of Big Data lifecycle, a comparison of data processes before and after the advent of Big Data, and an introduction of promising opportunities created along with new challenges.

3.2.1 Big Data Accumulation

Different types of data which are currently being generated in daily life are important sources of Big Data; they include financial transactions, scientific models, spatial maps, emails, website clicks, documents, telemetry, medical images and records, climate records, and so on[1], nearly 70-80% of which are unstructured fororganizations[16] and possess common features such as large size, heterogeneous structure, and complex processes[6]. The spread of new technologies in organizations and changing or modification of processes necessitates redefining the architecture of data/information including data/information model, data dictionary, data forms, and appropriate standards, in addition to changes in the flow of data inside and outside the organization[17]. Creation and use of standard terminology and databases in accordance with these terms help dealing with the challenges associated with collecting data from multiple locations and the use of data in intelligent data analysis[18].

The sources and types of datain the field of health care are web, smart phones, social networks (Facebook, Twitter, LinkedIn), transferred data from sensors and other vital signs recording devices, large data transfers (health care follow-up and financial records), biometric data (fingerprints, genetics, medical images, blood pressure), and data generated by human (EMRs, physicians notes, emails)[19].

Providing an immediate feedback to patients, improving patient interaction with remote health care provider organizations, and raising incentives for organizations to provide services, can prevent the occurrence of risk in the future through focusing on abnormal data[20]. Improvement of healthcare systems quality, identifying groups at risk of diseases, control plans, and prevention and monitoring of diseases are feasible only through creation of information recording systems which collect complete and updated data[21]. The technology of Big Data paves the way to achieve new knowledge and insight through creation of tools for collecting and analyzing large sets from heterogeneous data sources. To develop the road map for Big Data technology and to take advantage of promising opportunities, it is essential to obtain a clear understanding of user needs and requirements of the various stake holders of healthcare including patients, medical team, physicians, care providers, payers, pharmaceutical industry, medical products manufacturers, and government[22]. Conventional studies, performed often through the sampling method, were focusing on proving previous assumptions and were not paying attention to find new relations. In the era of Big Data, and due to the increase of data analysis capability, it is possible to process all data related to a particular phenomenon, without relying on random sampling. This approach makes quicker and easier identification of problems and more detailed attention, which was not possible in traditional data studies[2].

3.2.2 Big Data Integration and Cleaning

Consistent with information revolution, the increased quantity of information has engulfed human like sea, and if not managed smartly, itwilldrown[23]. In fact, we have created Big Data from traditional data in recent years; now it is the time to remove the duplicates or unimportant data and to select the necessary ones in order to turn Big Data into traditional data[24]. The main challenges for many organizations consist of storage, searching, sharing, analysis, and mass visualization of available data. Big Data technology uses modern methods of data storage and mining to modify the nature of data generated by organizations[25]. Many plans have been provided to prevent data redundancy and to filter out unnecessary items in application server, because the transmission of redundant packet scan affect the ability of server, leading to over head transmission, reduced network life time, and delay in processing[26,27]. In the area of health, the scale and complexity of data has increased dramatically by increasing the diversity of data sources (wide-range resolution biosensors, smart phones, etc.). Some of these increments are systematized and can be removed or minimized with normalization methods. The remaining increments reflect the limitations of technology and should be controlled appropriately. Therefore, data preprocessing including normalizing and quality control is necessary before any-analys is and complex processes of data mining[28]. Many techniques are applying for preprocessing of data such as data cleaning, data integration, data transformation, and data reduction[29-31].

a) Data cleaning, sometimes known as data standardization, includes correction, deletion, and in some cases change in data field according to the predetermined values. In this method, invalid data are separated from valid ones. Errorful, irrelevant, and incomplete data are examples of superfluous data that must be refined[32]. There are two ways to deal with redundant data; detection and elimination of redundant data as a part of preprocessing stage[33] and providing a model which is resistant to these data[34].

b) Data integration is the cornerstone of modern business informatics which includes a combination of information from different sources and provides a comprehensive view of the data for users[35]. The purpose of merging or integration of data is to produce a heterogeneous data source from a large set of non-heterogeneous data sources. In this process, the data collected from various sources (databases, flat files, etc.) are stored and maintained in an integrated data repository, to be analyzed properly.

c) Data transformation include smoothing (removal of noise from the data through binning, regression, and clustering), aggregation (data summarization or aggregation), generalization (generalization of primary data to high-level concepts using hierarchical relationships), normalization (to uniform data with different scales in a particular area), and attribute construction (extraction of new features from a given dataset)[30].

d) Data reduction: data redundancy refers to duplicate or additional data which occur frequently in datasets. Therefore, in order to reduce unnecessary costs of data transfer and to avoid wasting of storage space and so on, various methods have been proposed to reduce redundancy, including data reduction, data filtering, and data compression. But reducing redundancy may have certain negative effects. For instance, data compression and decompression produce additional computational load, thus the costs and benefits of reducing redundancy should be carefully balanced[29].

3.2.3 Big Data Analysis

Big Data analys is of data controlling process is useful for discovering hidden patterns, unknown relationships, and other useful information and can lead to better decision-making[36]. Big Data analysis is considered a method for analyzing a particular type of data. Thus, many traditional methods of data analysis may be used for Big Data analysis, such as Cluster Analysis, Factor Analysis, Correlation Analysis, Regression Analysis, Statistical Analysis, and Data Mining Algorithms. But at the beginning of the era of Big Data, concerns have been arisen about the extraction of key information from large amounts of data and gaining of value for organizations and individuals.

Several studies have pointed out the critical applications of Big Data in health, such as finding various factors in the lifecycle of diseases[37], automatic processing of PHRs[38], designing and implementation of more effective Mobile Healthcare System[39], processing practitioners notes, medical images and monitoring[19], reproducible analysis, maximizing contribution field[40], predictive modeling, and optimization of decisions for organizations[4].

In general, Big Data analytics can help reduce waste and inefficiency in the fields of clinical practice, research and development, and public health. Additionally, it is used in areas of Evidence-based medicine, Genomic analytics, Pre-adjudication fraud analysis, Device/remote monitoring, and Patient profile analytics[19].

3.2.4 Big Data Storage

Data storage refers to storage and management of large-scale data as well as ensuring their reliability and availability. Data storage system consists of two parts: infrastructure and data storage mechanism. Traditionally, data storage systems were the only auxiliary equipment of servers and have been stored, managed, and analyzed with structured RDBMSs. But with the advent of Big Data, these methods were insufficient, and hence lots of storage systems emerged to meet the needs of Big Data[36]. In disease management programs, in order to demonstrate the effects of the treatment plan, report the results of medical interventions, and present the best care practices along with reduced costs, the health care organizations are required to carefully analyze the data, and they have gained successful experiences through the development of Data warehouse[41]. But with the arrival of the era of Big Data, extracting wealth from vast biological data created primary challenges in bioinformatics, resulting in an unprecedented demand for storage and retrieval of Big Data[40].

Although the costs of establishment of basic infrastructure for Big Data technology are compensable given its substantial benefits[4], with the continuing growth of data volume, development and maintenance of computer infrastructure for storing and processing data is daunting for small organizations or even large institutions. Cloud computing is a promising solution at present to address these challenges[40]. Cloud computing offers solutions for IT services in the form of rental[42] in which processing power and storage space are provided to users and organizations based on an on-demand delivery[43]. Cloud computing is currently an important technology in storage and analysis of Big Data; a revolution that affects the health sector. Biological and biomedical sciences are extensively involved in the Big Data revolution through the use of secondary data which are normally produced in the care, as well as new data sources such as social media[44]. In this context, personal care systems are formed based on cloud computing which automatically store personal useful information, such as templates and rules related to lifestyle and health information, through mobile devices in cloud[45].

In the area of health, cloud computing have gained more significant role in overcoming the various challenges posed by the rise of Big Data in various sectors including health monitoring system with high-volume of processed data, mobility of monitored users and the area covered by the network[46], recording high-scale and high spatiotemporal resolution electro physiological data[47], extracting knowledge from unstructured, semi-structured, and structured data of patients enrolled in the PHRs[48], mobile computing and storage, delivery, recovery and better management of medical files[49] smart homes, and processing of transmitted data independent of specialized environment[50] and many other cases.

3.2.5 Big Data Sharing and Publishing

Big Data has changed the basic culture of data confidentiality, in which the obtained results were shared through publishing, to a data-driven culture, in which both data and publications are shared in the scientific community. The key stakeholders in the biomedical Big Data ecosystem include data providers and users (researchers, practitioners, and citizens), data scientists, financial investors, publishers, and librarians. Implementation of such a biomedical Big Data ecosystem requires updating of the policies regarding budget, data sharing, and data referencing in the context of cultural change[51]. But since health data may be the most personal, provision of a legal frame work and effective supervision on the flow of data are of the main challenges in data sharing[52]. Given the significant benefits of data sharing and its re-use compared with centralization, the international policy of health IT tends to support cloud computing. Therefore, it is necessary to balance the need for dissemination and analysis of Big Data and confidentiality of patient's data[53]. Cloud computing provides a group of services according to demands. These services are consisted of various types and layers which are divided into 4 categories; Data as a service (Daas), Software as a service (Saas), Platform as a service (Paas), and Infrastructure as a service (Iaas).

Clouds are related to data, in particular, in the field of bioinformatics and health sciences, and data are important in the analysis and discovery of knowledge. Access to active data in Daas is performed according to web-based demand[40]. SaaS which deliver software on the Internet eliminates the need to install the software on the client computer and facilitates its maintenance and support. (Google Docs) Paas service provides strong basic features for development of applications (Google Apps Engine) and Iaas offers a virtual platform as a service. (Drop Box) Billing for these services is based on utility computing and the amount of resources consumed, and therefore costs reflect the level of activity[54].

The use of cloud computing will have significant positive impact on the cost of IT-using industries by reducing the total Cost of Ownership (TCO), resulting in the creation of business and macro-economic performance at the national and global levels. Therefore, it will have many benefits to the public and private sectors including healthcare (especially in the field of e-health). In the area of healthcare, the services will be provided as "Health Care as a Service" (HCaaS). HCaaS focuses on achieving two specific goals: availability of e-health applications and medical information at any place and time, and invisibility of computation[55].

3.2.6 Big Data Recovery and Exploration

Customer-oriented feature of IT, the demand for tools to simplify data collection, and expectations for availability has increased both data and information systems[56]. In the world of Big Data, traditional methods of data access (JDBC adapters from an RDBMS or unstructured data such as documents from Document Management Systems (DMS) using HTTP interfaces) is very time consuming and inefficient due to the excessive amount of data.

Clouds deliver services with 4 different modes in organizations: Public cloud, Private cloud, Hybrid cloud, and Community cloud. Public cloud is widely used by small and medium businesses and describes cloud computing in its traditional sense. Private cloud is mainly used by large businesses that need to protect their data center in a reliable manner. Hybrid cloud is composed of several domestic or foreign providers and is a good option for most businesses. Community cloud is used as a vertical market such as health care or vehicle in which users have some common features in their applications[42].

Since commercial clouds are still unable to provide data and certain software for complex analyses in bioinformatics, and despite the many benefits of clouds in the era of Big Data, yet a small fraction of bioinformatics data are loaded on clouds. However, keeping pace with the rapidly emerging needs arising from scientific research in bioinformatics is difficult in business clouds; on the other hand, open access of public to information and applications is a scientific necessity[40]. The potential benefits resulting from bioinformatics clouds involve facilitation of large-scale data integration, repeatable and reproducible analyses, maximization of sharing range, and harnessing collective intelligence for knowledge discovery. With the presence of multiple bioinformatics clouds, interoperability and standardization of the clouds will become important issues[57].

4. Conclusion

The present study refers to the emergence of Big Data in the field of health known as datasets which processing is not feasible with current technologies in a reasonable time, and requires cost effective and innovative methods of data processing, to improve in sight and decision-making.

Based on the review of current research in the lifecycle of Big Data, the challenges created by the emergence of Big Data have in fact set the stage for promising solutions in the management of Big Data revolution. So that increased data sources and exponential rise in health data were associated with development of new tools for collecting and classifying the data; and increased ability to analyze data has improved discovery of hidden patterns, and resulted in better decisions.

Evaluation of challenges and solutions generated at each phase of Big Data lifecycle showed that Big Data has created a substantial change in different phases of collection, integration, analysis, storage, and publication of health data; unfamiliarity with which in the competitive business environment is challenging. Today, the business of health services is growing rapidly and information technology and communications make this process easier. The application of data-driven approach; i.e. management of health based on detailed and measured evidence, is necessary to progress towards a new level of innovation, competitiveness, and productivity. Prospective lifecycle assessment studies that are mainly based on review of scientific literature are able to depict the future of products in the early stages of technology development[58].

According to this study, assessment and understanding the life cycle phases of Big Data is a new approach to dominate this process and to enjoy the benefits of Big Data technologies in order to achieve the business intelligent, and is promising for managers and policy makers in the field of health. According to other studies, in order to manage and protect large volumes of data in organizations and with stand the competitive pressures of business, it has been recommended to use work flow analysis and discovered knowledge analysis, and to compare them with the prescribed work flow, in order to highlight the danger zone[11].

5. References

1. Khan N, Yaqoob I, Hashem IA, Inayat Z, Ali WK, Alam M, et al. Big Data: Survey, Technologies, Opportunities, and Challenges, The Scientific World Journal. 2014(2014), 18p. ID: 712826. Crossref
2. Cui M, Li H, Hu X. Similarities between "Big Data" and Traditional Chinese Medicine Information, Journal of Traditional Chinese Medicine = Chung I Tsa Chih Ying Wen Pan / Sponsored by All-China Association of Traditional Chinese Medicine, Academy of Traditional Chinese Medicine. 2014; 34(4):518–22. Crossref
3. Kaisler S, Armour F, Espinosa JA, Money W. Editors. Big Data: Issues and Challenges Moving Forward. System Sciences (HICSS), 2013 46th Hawaii International Conference on; 2013. IEEE.
4. Peters SG, Buntrock JD. Big Data and the Electronic Health Record, The Journal of Ambulatory Care Management. 2014; 37(3):206–10. Crossref PMid:24887521.
5. Caballero I, Serrano M, Piattini M. Editors. A Data Quality in use Model for Big Data. International Conference on Conceptual Modeling; 2014: Springer.
6. Li B, Liao X. Peer-to-Peer in Big Data Management, Peer-to-Peer Networking and Applications. 2013; 6(4):361. Crossref
7. Stokes LB, Rogers JW, Hertig JB, Weber RJ. Big Data: Implications for Health System Pharmacy, Hospital Pharmacy. 2016; 51(7):599–603. Crossref PMid:27559194.
8. Alexander GL, Rantz M, Flesner M, Diekemper M, Siem C. Clinical Information Systems in Nursing Homes: An Evaluation of Initial Implementation Strategies, Computers, Informatics, Nursing: CIN. 2007; 25(4):189–97. Crossref PMid:17625399.
9. Rigamonti L, Falbo A, Grosso M. Improvement Actions in Waste Management Systems at the Provincial Scale Based on a Life Cycle Assessment Evaluation, Waste Management (New York, NY). 2013; 33(11):2568–78. Crossref PMid:23948052.
10. Smith D, Wood D. Data Management. Research in Clinical Practice: Springer; 2013. p. 59–64. Crossref. Crossref.
11. Korba L, Song R, Yee G, Patrick AS, Buffett S, Wang Y, et. al. Editors. Private Data Management in Collaborative Environments. International Conference on Cooperative Design, Visualization and Engineering; 2007: Springer.
12. Chomkhamsri K, Wolf M-A, Pant R. International Reference Life Cycle Data System (ILCD) Handbook: Review Schemes for Life Cycle Assessment. Towards Life Cycle Sustainability Management: Springer; 2011. p. 107–17.
13. Demchenko Y, Ngo C, Membrey P. Architecture Framework and Components for the Big Data Ecosystem, Journal of System and Network Engineering. 2013; 1–31.
14. Shaw-Taylor Y. Making Quality Improvement Programs More Effective, International Journal Of Health Care Quality Assurance. 2014; 27(4):264–70. Crossref PMid:25076601.
15. Ma X, Fox P, Rozell E, West P, Zednik S. Ontology Dynamics in a Data Life Cycle: Challenges and Recommendations from a Geoscience Perspective, Journal of Earth Science. 2014; 25(2):407–12. Crossref
16. Holzinger A, Stocker C, Ofner B, Prohaska G, Brabenetz A, Hofmann-Wellenhof R. Combining HCI, Natural Language Processing, and Knowledge Discovery-Potential of IBM Content Analytics as an Assistive Technology in the Biomedical Field. Human-Computer Interaction and Knowledge Discovery in Complex, Unstructured, Big Data: Springer; 2013. p. 13–24.
17. Ghazisaeidi M, Ahmadi M, Sadoughi F, Safdari R. A Roadmap to Pre-Implementation of Electronic Health Record: the Key Step to Success, Acta Informatica Medica: AIM: Journal of the Society for Medical Informatics of Bosnia and Herzegovina: Casopis Drustva za Medicinsku Informatiku BiH. 2014; 22(2):133–8. Crossref
18. Mohammadzadeh N, Safdari R, Baraani A, Mohammadzadeh F. Intelligent Data Analysis: The Best Approach for Chronic Heart Failure (CHF) follow up Management, Acta Informatica Medica: AIM: Journal of the Society for Medical Informatics of Bosnia and Herzegovina: Casopis Drustva za Medicinsku Informatiku BiH. 2014; 22(4):263–7. PMid:25395730 PMCid:PMC4216425.
19. Raghupathi W, Raghupathi V. Big Data Analytics in Healthcare: Promise and Potential, Health Information Science and Systems. 2014; 2:3. Crossref PMid:25825667 PMCid:PMC4341817.
20. Mohammadzadeh N, Safdari R, Rahimi A. Developing Framework for Agent- Based Diabetes Disease Management System: user Perspective, Materia Socio-Medica. 2014; 26(1):62–4. Crossref PMid:24757407 PMCid:PMC3990393.

21. Yazdanian A, Safdari R, Mahsoori N, Siamian H, Bagheri Nesami M, Haghshenas MR, et. al. Proposed Model for Iranian National System of Registration of Allergy and Asthma, Acta Informatica Medica: AIM: Journal of the Society for Medical Informatics of Bosnia and Herzegovina: Casopis Drustva za Medicinsku Informatiku BiH. 2013; 21(3):196–9. Crossref
22. Zillner S, Lasierra N, Faix W, Neururer SB. Editors. User Needs and Requirements Analysis for Big Data Healthcare Applications. MIE; 2014.
23. Jifa G. Some Issues on Big Data, Science and Technology for Development. 2014; 1:004.
24. Jifa G, Lingling Z. Data, DIKW, Big Data and Data Science, Procedia Computer Science. 2014; 31:814–21. Crossref
25. Cumbley R, Church P. Is "Big Data" Creepy? Computer Law and Security Review. 2013; 29:601–09. Crossref
26. Kadayif I, Kandemir M. Editors. Tuning in-Sensor Data Filtering to Reduce Energy Consumption in Wireless Sensor Networks. Proceedings of the Conference on Design, Automation and Test in Europe-Volume 2; 2004: IEEE Computer Society.
27. Yu H, Zeng P, Wang ZF, Liang Y, Shang ZJ. Study on Distributed Wireless Sensor Networks Communication Protocols, J. Commun. 2004; 25:102–10.
28. Wang X, Hessner MJ, Wu Y, Pati N, Ghosh S. Quantitative Quality Control in Microarray Experiments and the Application in Data Filtering, Normalization and False Positive Rate Prediction, Bioinformatics (Oxford, England). 2003; 19(11):1341–7. Crossref
29. Chen M, Mao S, Liu Y. Big Data: A Survey, Mobile Networks and Applications. 2014; 19(2):171–209. Crossref
30. Han J, Kamber M. Data Ining: Concepts and Techniques. 2006.
31. Winck AT, Machado KS, de Souza ON, Ruiz DD. Context-Based Preprocessing of Molecular Docking Data, BMC Genomics. 2013; 14(Suppl. 6):S6. Crossref PMid:24564276 PMCid:PMC3909228.
32. Randall SM, Ferrante AM, Boyd JH, Semmens JB. The Effect of Data Cleaning on Record Linkage Quality, BMC Medical Informatics and Decision Making. 2013; 13:64. Crossref PMid:23739011 PMCid:PMC3688507.
33. Patel ZH, Kottyan LC, Lazaro S, Williams MS, Ledbetter DH, Tromp H, et. al. The Struggle to Find Reliable Results in Exome Sequencing Data: Filtering out Mendelian Errors, Frontiers in Genetics. 2014; 5:16. Crossref PMid:24575121 PMCid:PMC3921572.
34. Ergun B, Kavzoglu T, Colkesen I, Sahin C. Data Filtering with Support Vector Machines in Geometric Camera Calibration, Optics express. 2010; 18(3):1927–36. Crossref PMid:20174021.
35. Lenzerini M. Editor Data Integration: A Theoretical Perspective. Proceedings of the Twenty-First ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems; 2002: ACM.
36. Chen M, Mao S, Zhang Y, Leung VC. Big Data: Related Technologies, Challenges and Future Prospects: Springer; 2014. Crossref
37. Fang Z, Fan X, Chen G. A Study on Specialist or Special Disease Clinics Based on Big Data, Frontiers of Medicine. 2014; 8(3):376–81. Crossref PMid:25186249.
38. Takeuchi H, Mayuzumi Y, Kodama N, Sato K. Personal Healthcare System using Cloud Computing, Studies in Health Technology and Informatics. 2013; 192:936. PMid:23920710.
39. Ji Z, Ganchev I, O'Droma M, Zhang X, Zhang X. A Cloud-Based X73 Ubiquitous Mobile Healthcare System: Design and Implementation, The Scientific World Journal. 2014; 2014.
40. Dai L, Gao X, Guo Y, Xiao J, Zhang Z. Bioinformatics Clouds for Big Data Manipulation, Biology Direct. 2012; 7(1):1. Crossref PMid:23190475 PMCid:PMC3533974.
41. Esposito D, Taylor EF, Gold M. Using Qualitative and Quantitative Methods to Evaluate Small-Scale Disease Management Pilot Programs, Population Health Management. 2009; 12(1):3–15. Crossref PMid:19216674.
42. Srinivasan S. Basic Cloud Computing Types. Cloud Computing Basics: Springer; 2014. p. 17–41. Crossref. Crossref.
43. Yang K, Jia X. Data Storage Auditing Service in Cloud Computing: Challenges, Methods and Opportunities, World Wide Web. 2012; 15(4):409–28. Crossref
44. Peek N, Holmes J, Sun J. Technical Challenges for Big Data in Biomedicine and Health: Data Sources, Infrastructure, and Analytics, Yearb Med Inform. 2014; 9(1):42–7. Crossref PMid:25123720 PMCid:PMC4287098.
45. Takeuchi H, Mayuzumi Y, Kodama N, Sato K. Personal Healthcare System using Cloud Computing, Studies in Health Technology and Informatics. 2012; 192:936-.
46. Almashaqbeh G, Hayajneh T, Vasilakos AV, Mohd BJ. QoS-Aware Health Monitoring System using Cloud-Based WBANs, Journal of Medical Systems. 2014; 38(10):1–20. Crossref PMid:25123456.
47. Brinkmann BH, Bower MR, Stengel KA, Worrell GA, Stead M. Large-scale Electrophysiology: Acquisition, Compression, Encryption, and Storage of Big Data, Journal of Neuroscience Methods. 2009; 180(1):185–92. Crossref PMid:19427545 PMCid:PMC2720128.
48. Mantas J. Machine Learning for Knowledge Extraction from PHR Big Data, Integrating Information Technology and Management for Quality of Care. 2014; 202:36.

49. Hsieh J-C, Li A-H, Yang C-C. Mobile, Cloud, and Big Data Computing: Contributions, Challenges, and New Directions in Telecardiology, International Journal of Environmental Research and Public Health. 2013; 10(11):6131–53. Crossref PMid:24232290 PMCid:PMC3863891.
50. Vimarlund V, Wass S. Big Data, Smart Homes and Ambient Assisted Living, IMIA Yearbook of Medical Informatics. 2014:143–9. PMid:25123734 PMCid:PMC4287073.
51. Margolis R, Derr L, Dunn M, Huerta M, Larkin J, Sheehan J, et. al. The National Institutes of Health's Big Data to Knowledge (BD2K) Initiative: Capitalizing on Biomedical Big Data, Journal of the American Medical Informatics Association. 2014; 21(6):957–8. Crossref PMid:25008006 PMCid:PMC4215061.
52. ComPUtING C. Cloud Computing Privacy Concerns on our Doorstep, Communications of the ACM. 2011; 54(1):36–8. Crossref
53. Currie W, Seddon J. A Cross-Country Study of Cloud Computing Policy and Regulation in Healthcare. 2014.
54. Yoon JP. Access Control and Trustiness for Resource Management in Cloud Databases. Grid and Cloud Database Management: Springer; 2011. p. 109–31. Crossref
55. Nur FN, Moon NN. Health Care System Based on Cloud Computing, Asian Transactions on Computers. 2012; 2(5):9–11.
56. Faught IC, Aspevig J, Spear R. New Means of Data Collection and Accessibility. Public Health Informatics and Information Systems: Springer; 2014. p. 375–98. Crossref
57. Parameswaran A, Chaddha A. Cloud Interoperability and Standardization, SET Labs Briefings. 2009; 7(7):19–26.
58. Arvidsson R, Kushnir D, Sandén BrA, Molander S. Prospective Life Cycle Assessment of Graphene Production by Ultrasonication and Chemical Reduction, Environmental Science and Technology. 2014; 48(8):4529–36. Crossref PMid:24646298