**HIR**
Healthcare Informatics Research

# Big Data Analysis Framework for Healthcare and Social Sectors in Korea

Tae-Min Song, PhD[1], Seewon Ryu, PhD[2]
[1]Statistics and Information Research Department, Korea Institute for Health and Social Affairs, Sejong; [2]Department of Health Policy and Healthcare Management, Inje Institute of Advanced Studies, Seoul, Korea

**Objectives:** We reviewed applications of big data analysis of healthcare and social services in developed countries, and subsequently devised a framework for such an analysis in Korea. **Methods:** We reviewed the status of implementing big data analysis of health care and social services in developed countries, and strategies used by the Ministry of Health and Welfare of Korea (Government 3.0). We formulated a conceptual framework of big data in the healthcare and social service sectors at the national level. As a specific case, we designed a process and method of social big data analysis on suicide buzz. **Results:** Developed countries (e.g., the United States, the UK, Singapore, Australia, and even OECD and EU) are emphasizing the potential of big data, and using it as a tool to solve their long-standing problems. Big data strategies for the healthcare and social service sectors were formulated based on an ICT-based policy of current government and the strategic goals of the Ministry of Health and Welfare. We suggest a framework of big data analysis in the healthcare and welfare service sectors separately and assigned them tentative names: 'health risk analysis center' and 'integrated social welfare service network'. A framework of social big data analysis is presented by applying it to the prevention and proactive detection of suicide in Korea. **Conclusions:** There are some concerns with the utilization of big data in the healthcare and social welfare sectors. Thus, research on these issues must be conducted so that sophisticated and practical solutions can be reached.

**Keywords:** Big Data, Delivery of Health Care, Social Welfare, Suicide, Public Health Surveillance

## I. Introduction

Big data analysis is considered a valuable area of study for both policymakers and researchers, as it can reflect the magnitude and impact of data-related problems within the realms of healthcare and social welfare [1]. Big data is created, stored, and disseminated through traditional and mobile Internet, smartphones, smart TV, sensor- and RFID-based ubiquitous networks, and social media [2]. Indeed, multi-method analysis and interpretation of big data is useful for enhancing business and social services. Healthcare professionals can analyze and integrate big data of hospital information systems, which can be connected to systems of

other hospitals. Social workers or researchers also can collect and analyze data about social welfare services from various online and offline sources, such as social network services (SNSs), blogs, metermen, closed circuit televisions (CCTV), and social worker and welfare organizations.

Big data analysis contains value-creating and problem-resolving possibilities. Developed and wired countries, such as the United States, the UK, Singapore, Australia, and even the OECD and the EU are proactively devising a strategic plan for big data, emphasizing its potential, and using it as a tool to solve long-sta nding problems or to uncover novel opportunities of value [3-9]. The Economist, Gartner, and McKinsey have provided examples of economic value-creating via the utilization of big data, such as predicting market changes and searching for new projects [10-12]. The Economist predicts that the proper utilization of big data might yield resolutions to problems concerning the environment, energy, food, and healthcare [10]. Moreover, Gartner projects that big data analysis is a monumental phenomenon that is destined to help organizations drive innovation by yielding new and faster insight into their customers. However, through 2015, 85% of Fortune 500 organizations will be unable to exploit big data to gain a competitive advantage [11].

Big data can enhance the health and social safety of citizens in the future. It has the potential to support complex and important social issues pertaining to healthcare, social safety, and welfare, such as the prevention and management of communicable diseases, person-centered collaborative care [13], terrorism, disaster, and global risks. For example, it is possible to prevent chronic disease by monitoring one's lifestyle and sensed data, and also to detect signs of social problems and the need for social services by monitoring data collected via multiple sources. Further, we can create policy agendas through monitoring and analyzing the big data of social media [14].

The purpose of this paper is to review trends in big data applications in some countries, consider and design a framework for big data in the areas of healthcare and social services in Korea, and suggest implications for the future and further research.

## II. Methods

We reviewed some cases of big data implementation in the areas of healthcare and social services in developed countries. We also reviewed strategies employed by the strategic policy of Ministry of Health and Welfare (MoHW) of Korea (Government 3.0) by projecting it onto big data characteristics. After reviewing the previous big data practices,

we formulated a conceptual framework for big data in the healthcare and social services sectors at the national level. As a specific case, we designed a process and method for social big data analysis on suicide buzz. Lastly, we discuss strategies and issues for big data analysis in the healthcare and social services sectors.

## III. Results

### 1. Review of Big Data Applications to Healthcare and Social Problems

Many countries are applying big data analytics to solve problems in healthcare and social services. In the United States, the Pillbox project results in an annual $500 million reduction in healthcare costs through the application of big data analytics [3,4]. The San Francisco Police Department has developed a big data system designed for crime prevention [3].

The UK is utilizing big data through establishment and management of the Foresight Horizon Scanning Centre, which serves as a countermeasure to various health and social problems such as obesity, potential risk management (coastal erosion, climate change), and epidemics [5]. The EU is dealing with uncertainty through the iKnow (Interconnect Knowledge) project, which provides opportunities for research on earthquakes, tsunamis, terrorism, networking, and global crisis [15]. The OECD adopted 'evaluating economic benefits of big data' as an agenda for the 15th Working Party on Indicators for the Information Society (WPIIS) by considering big data for business efficiency [8].

Moreover, the Australian Government Information Management Office has saved time and resources by developing an automated tool that can analyze, search, and reuse massive information through government 2.0 [7]. In 2004, Singapore established the Risk Assessment and Horizon Scanning (RAHS) to prepare for future uncertainty regarding terrorism and epidemics [6].

In Korea, the National Information Society Agency (NIA) demonstrated the potential for suicide prevention through analysis of online buzzwords. Following this, a 2012 study revealed a relationship between quantity of searching 'suicide' and suicide rate by analyzing Google search trends [2].

### 2. Big Data Strategies in Healthcare and Social Services in Korea

Big data strategies within the healthcare and social services sectors were formulated to seek effective ways of utilizing big data. We examined Government 3.0, an ICT-based policy devised by the current Korean government, and the Strategic Goals of the MoHW as a foundation into big data strategies

for healthcare and welfare services [2].

We matched three MoHW goals with the characteristics of big data (high-volume, high-velocity, high-variety, veracity, value-creating, and highly complex), and established three objectives for each goal, as showed in Figure 1. First, we matched the volume and complexity characteristics of big data with the goal of 'clearly communicating healthcare and welfare services.' To accomplish the goal of 'clearly communicating healthcare and welfare services,' public data, which were publicly accumulated and stored by the government and various public organizations and open to anyone who needed them, were used to proactively facilitate the use of big data. We established three objectives: meet the people's need to know, motivate active use of public data by the private sector, and strengthen collaborative governance between the private and public sectors. Each objective can be accomplished through the use of two more strategies.

Second, we matched the velocity and variety characteristics of big data with the goal of 'efficient and capable healthcare and welfare services.' To achieve this goal, it is necessary to integrate various data via scientific administration and improvement of the government operation system. Three objectives to support this goal were established: remove partitions within the government, improve systems to facilitate collaboration and communication, and realize administration by utilizing big data.

Last but not least, we matched the veracity and new value-creating characteristics of big data with the goal of 'people-

centered healthcare and welfare services.' To achieve this goal, we established three objectives: provide integrated services tailored to consumers, strengthen one-stop services for business start-ups and activities, and enhance accessibility for less-informed consumers.

## 3. Effective Big Data Utilization Measures in Healthcare and Social Services

Big data represents a new paradigm for healthcare and social services, serving as an engine to create new values. A large volume of big data in the domestic healthcare and social services sectors has been stored and maintained in public or private organizations according to the stable implementation of ICT-based legacy systems. We designed separate frameworks for big data analysis in the healthcare and social services sectors.

### 1) Health risk analysis center in the healthcare sector

In the healthcare sector, government organizations and health insurance agencies, such as the Korea Food and Drug Administration, National Healthcare Insurance Corporation, Health Insurance Review and Assessment Service, and National Cancer Center [2] have begun to apply big data analysis towards evaluating services.

In order to provide healthcare services tailored to a person's lifecycle, it is imperative to formulate a model for future prediction and policy decision by utilizing big data from various sources of healthcare data. Integrated data management and
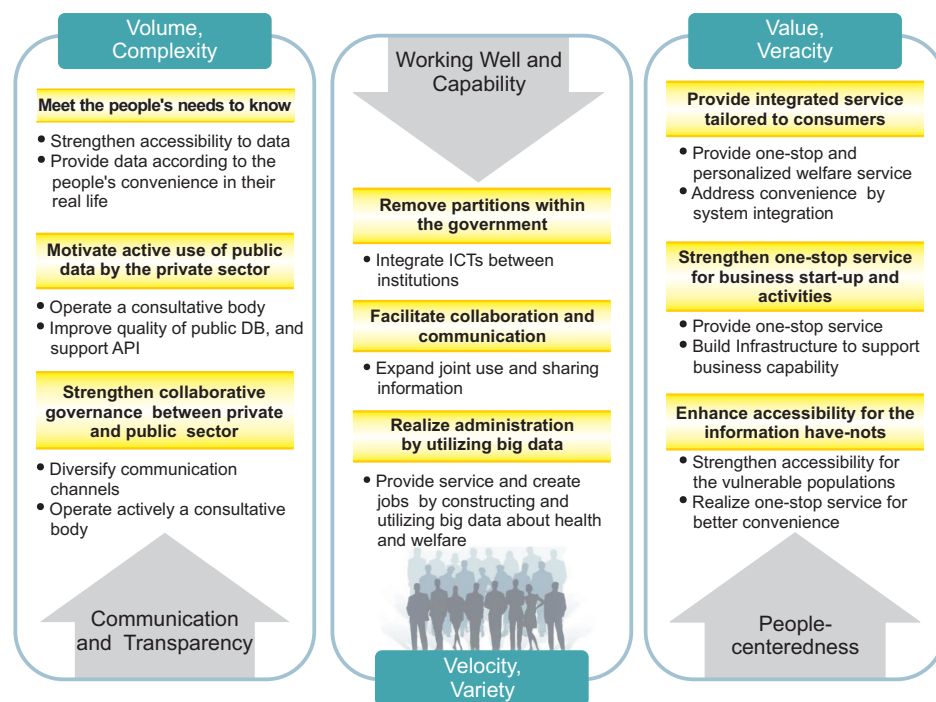


Figure 1. Strategic plan of big data for healthcare and social welfare service.

analysis of big data is necessary to achieve proactive public health management and healthcare services. Moreover, it is necessary to establish an organization to analyze health risk and services, establish policies, and evaluate on the individual, community, and country levels. We named the assumed organization the 'Health Risk Analysis Center' (Figure 2).

The health risk analysis center is expected to enable early response to diseases at the national level. It will predict the distribution and trends of major diseases through monitoring health behaviors and home environments, the utilization of medical services, and other health-related data and population statistics [16]. The center would perform the following

tasks:

- Establish a disease warning system based on characteristics and social stratum
- Provide health information to the local community
- Provide people with a tailored health and disease management program based on their lifestyle
- Provide information that can be applied towards the prevention and prediction of geriatric disease
- Provide information that can be used to diagnose individual health status
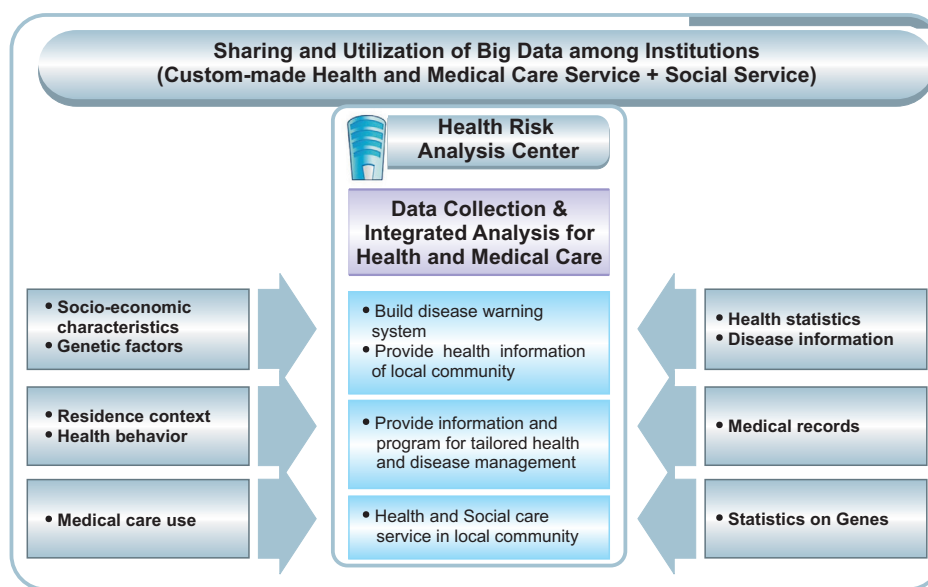- Perform social services within the local community (social care, home visiting care, rehabilitation, etc.)
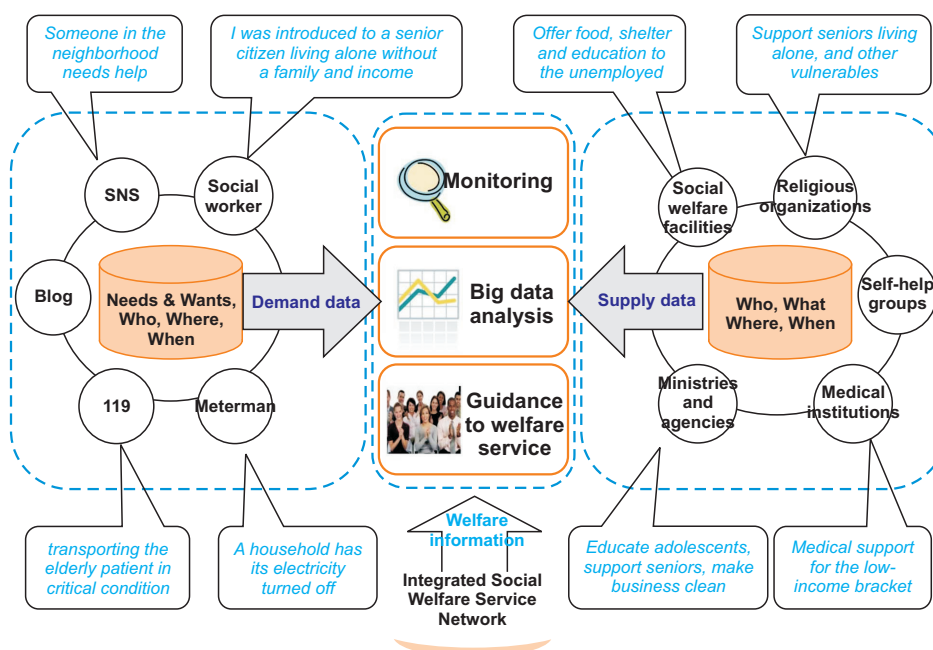


Figure 2. Health risk analysis center.



Figure 3. Integrated social welfare service network.

**2) Integrated social welfare services network in the social welfare sector**

In the social welfare sector, we need a more integrated social welfare service network. Since 2010, the 'Happiness e-Connection' system, a social welfare information center, is functioning to integrate and manage welfare services conducted by government ministries and agencies, including the MoHW, together with beneficiary information. The current form of the integrated social welfare service network should be expanded to form a national big data infrastructure by integrating with and linking to the information systems of all government ministries and agencies in order to remove blind spots within social welfare, and provide welfare services tailored to individuals based on their lifecycle (Figure 3).

**4. Social Big Data Analysis of Suicide**

As a prototype, we will present an analysis process of social big data in healthcare and social services. We established a framework of social big data analysis and applied it to the subject of preventing and proactively detecting suicide, as suicide cases have been recently increasing in Korea. In the social big data sector, social analytics quickly analyzes unstructured data collected from Facebook, Twitter, and other SNSs. The process of extracting and analyzing social big data from social media is illustrated in Figure 4. First, target social big data are collected. The target collection (unstructured big data on search portal or SNSs) and scope are defined, and collection is performed through collection engines such as a 'crawler' (i.e., a robot program).

Second, collected unstructured data are analyzed. Analysis of unstructured data is performed in the order of buzz analysis, keyword analysis, opinion analysis, and account analysis. The collected unstructured data are then subjected to text and opinion mining.

Third, collected unstructured data are classified using network analysis and subsequently converted to structured data. To convert data from an unstructured form to a structured form, each document on suicide buzz received an ID and was codified to various keywords and methods.

Fourth, the structured data are connected to offline statistics and surveys from government and public organizations. To perform analysis on the social phenomena, the structured big data is connected to the structured big data of public organizations. IDs (by date/month/year/region) that can be connected are checked and matched with the big data (offline statistics) of public organizations.

Finally, an analysis of the structured big data connected to the offline statistics and surveys can be performed through structural equation modeling, which enables a cause-effect analysis between factors or the tracking of time-based trajectory changes; the multi-level model, which enables the analysis of factors related to social phenomena by date/month/year and region; and data mining analysis, which allows for the discovery of new phenomena through classification of collected keywords.

## IV. Conclusion

In this study, we formulated a national-level framework for big data analysis of the healthcare and social service sectors in Korea based on a review of several big data applications in other countries. However, there are several issues that need

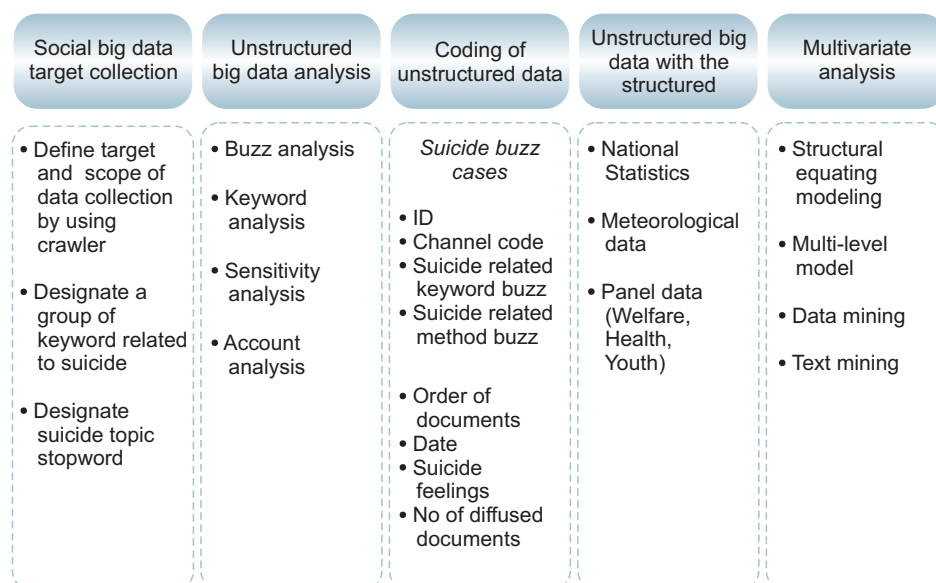| Social big data target collection | Unstructured big data analysis | Coding of unstructured data | Unstructured big data with the structured | Multivariate analysis |
|---|---|---|---|---|
| • Define target and scope of data collection by using crawler<br><br>• Designate a group of keyword related to suicide<br><br>• Designate suicide topic stopword | • Buzz analysis<br><br>• Keyword analysis<br><br>• Sensitivity analysis<br><br>• Account analysis | *Suicide buzz cases*<br><br>• ID<br>• Channel code<br>• Suicide related keyword buzz<br>• Suicide related method buzz<br><br>• Order of documents<br>• Date<br>• Suicide feelings<br>• No of diffused documents | • National Statistics<br><br>• Meteorological data<br><br>• Panel data (Welfare, Health, Youth) | • Structural equating modeling<br><br>• Multi-level model<br><br>• Data mining<br><br>• Text mining |

Figure 4. Social big data analysis process and method: application to suicide.

to be addressed before big data in the healthcare and social welfare sectors can be effectively utilized.

First, management commission at government ministries and agencies is required to properly operate big data for healthcare and welfare services because big data must be managed in an integrated manner. At the moment, big data on healthcare and welfare services are primarily managed and operated by government ministries and agencies, including the MoHW, Ministry of Employment and Labor, Ministry of Knowledge Economy (currently divided into the Ministry of Trade, Industry, and Energy, and the Ministry of Science, ICT and Future Planning), and Korea Food and Drug Administration, as well as public organizations, such as the National Health Insurance Corporation, the Health Insurance Review and Assessment Agency, and other national-level research organizations. A government-wide organization should be established in order to connect and share the information currently operated independently by each organization.

Second, it is imperative to establish a cooperative system with private organizations that maintain unstructured big data related to healthcare and welfare services. Considering that the unstructured big data related to healthcare and welfare services is stored and maintained through search portals or SNSs in the private sector, a close cooperative and communication system (tentatively named the 'Healthcare and Welfare Big Data Forum') is needed.

Third, an open application programming interface (API) at the national level should be made available. Most healthcare and welfare service-related big data are owned exclusively by the public sector. It is important to make information available on the Web in real-time, and simultaneously provide information collection and analysis; however, it is perhaps even more important to proactively consider opening the API at the government level in order to utilize big data of healthcare and welfare service effectively and efficiently. As of February 2014, there are 503 open APIs made public in the open data portal on shared resources [16], and among them, there are 79 and 24 big data sets pertaining to healthcare services and welfare services, respectively. The Presidential Council on Information Society proposes that the government must proactively utilize big data and build a national knowledge platform because the current explosive increase in data is an economic asset. With respect to the opening of big data related to healthcare and welfare services, big data can be categorized according to the needs of the government and citizens with the participation of relevant organizations and big data professionals, and the targets for open data can be stored in the national knowledge platform with strict security of personal information.

Fourth, it is necessary to develop the technologies related to analysis and processing of big data about healthcare and welfare services. In the 'smart-life' era, technologies governing the storage and analysis of non-relational, unstructured data, and the expansion of cloud services, semantic search services, inference-based situation recognition services, etc., will become critical. Accordingly, the development of technologies to enable big data 'collection → storage → analysis → inference' as well as technology standardization should be facilitated on a preferential basis.

Fifth, it is important to train data scientists who can detect information hidden in the large-volume data. In the big data era, people who are capable of maintaining and analyzing it are an invaluable asset. Global IT companies have already devoted great efforts to securing talented data scientists and strengthening competencies. Therefore, strategies to train data scientists in the healthcare and welfare sectors should be formulated in cooperation with the Ministry of Education.

Last but not least, security policies should be prepared to deal with personal information and confidential information related to big data on healthcare and welfare services. Big data of healthcare and welfare services contains almost all personal data about a person, but legal and institutional systems are not adequately prepared or discussed fairly. Utilization of big data is crucial, but the leak of excessive personal information can violate privacy and human rights in the cyberspace, or such information can be used for unlawful purposes. One of the most critical factors in protecting a person from big data is data anonymity, which renders individuals unrecognizable, and the control of information access and processing. However, as the control of information access and processing becomes stricter, the utilization of information becomes more inactive; thus, effective policies with respect to the 'utilization and protection' of healthcare and welfare big data should be developed.

It is certain that big data will be value-creating and promising within the healthcare and welfare sectors. To acquire more information about the healthcare and social welfare sectors, the above issues should be actively explored in research so that sophisticated and practical solutions can be reached.

## Conflict of Interest

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

## Acknowledgments

## References

1. Chen H, Chiang RH, Storey VC. Business intelligence and analytics: from big data to big impact. MIS Q 2012; 36(4):1165-88.

2. Song TM. Efficient utilization of big data on health and welfare. Health Welf Policy Forum 2012;(193):68-76.

3. Lee BH. Big data generation and new business opportunities [Internet]. Seoul: Hyundai Research Institute; 2012 [cited at 2015 Jan 5]. Available from: http://www.hri.co.kr/board/ reportView.asp?firstDepth=1&secondDepth=6%20&numIdx=18609.

4. Pillbox [Internet]. Washington (DC): National Library of Medicine; c2013 [cited at 2014 Dec 2]. Available from: http://pillbox.nlm.nih.gov/.

5. The Horizon Scanning Programme Team [Internet]. London: The UK Government; 2014 [cited at 2014 Aug 2]. Available from: https://www.gov.uk/government/groups/horizon-scanning-centre.

6. Risk Assessment and Horizon Scanning (RAHS) [Internet]. Singapore: National Security Coordination Secretariat; c2014 [cited at 2014 Dec 2]. Available from: http://app.nscs.gov.sg/public/home.aspx.

7. National Information Strategy Committee of Korea. Implementation of smart government using big data [Internet]. Seoul: National Information Strategy Committee; 2012 [cited at 2015 Jan 5]. Available from: http://mmlab.snu.ac.kr/courses/2011_information-universe/presentation/1121.pdf.

8. OECD. Exploring data-driven innovation as a new source of growth: mapping the policy issues raised by 'big data' [Internet]. Paris: OECD Publishing; 2013 [cited at 2014 Dec 2]. Available from: http://www.oecd-ilibrary.org/science-and-technology/exploring-data-driven-innovation-as-a-new-source-of-growth_5k47zw3fcp43-en.

9. Song YJ. Data based on national future strategy and implications of the developed countries. IT Future Strategy 2013;(2):1-18.

10. The Economist. Data, data everywhere [Internet]. London: The Economist; c2010 [cited at 2014 Dec 2]. Available from: http://www.economist.com/node/15557443/.

11. Gartner's 2012 hype cycle for emerging technologies identifies "tipping point" technologies that will unlock long-awaited technology scenarios [Internet]. Stamford (CT): Gartner Inc.; c2012 [cited at 2014 Dec 2]. Available from: http://www.gartner.com/newsroom/id/2124315.

12. Manyika J, Chui M, Brown B, Bughin J, Dobbs R, Roxburgh C, et al. Big data: the next frontier for innovation, competition, and productivity. Washington (DC): McKinsey Global Institute; 2011.

13. Koch S. Achieving holistic health for the individual through person-centered collaborative care supported by informatics. Healthc Inform Res 2013;19(1):3-8.

14. Song YJ. Big data world! Evolution of SNS and the public policy. IT Future Strategy 2012;(13):1-26

15. Go SJ, Jeoung YH. Health risk prediction using big health data. Health Welf Policy Forum 2012;(193):43-52.

16. Open data portal [Internet]. Seoul: National Information Society Agency; c2014 [cited at 2014 Dec 2]. Available from: https://www.data.go.kr.