# PUBLIC OPEN-DATA COLLECTION AND VISUALIZATION BASED ON HTML5: MOBILE WEB APPROACH

Kwangseob Kim[1] and Kiwon Lee[2]

[1]Graduate Student, Dept. of Info. Sys. Eng., Hansung University, Seoul, Korea, 02876,
Email: lovekph@nate.com
[2]Professor, Dept. of Info. Sys. Eng., Hansung University, Seoul, Korea, 02876,
Email: kilee@hansung.ac.kr

**KEY WORDS:** HTML5, Mobile Web, Public Data, Open Source

**ABSTRACT:** A mobile web application performs specific tasks with digital contents on mobile web browsers in any place and any time. Mobile web technologies for service development or target-oriented application system implementations are mostly based on HyperText Markup Language5 (HTML5) to perform core functionalities from client-side. Among several useful APIs of HTML5, indexedDB provides off-line data storage capacity to support dual-mode of off-line and on-line processing. While, public open data policy in government organizations or agencies in most countries provides Open API or of downloading numerous types of geo-based data sets for mobile mashup or fusion applications. These data sets contain base map of Open Street Map (OSM) or other mapping servers, multiple types of vector layers and multi-resolution satellite images. Data sets generated from both natural environment and man-made environment, if they are freely accessible public open data, is considered as significant geo-based data contents to solve many real world problems. This study is to design and implement system about public open data collection and visualization, including geo-based data. As well, mobile visualization processor also provides through dashboard in an integrated environment. This implementation was carried out by fully open source stack and linkage approach such as Spring framework for common components and Bootstrap for client interface, without any commercialized software or proprietary tools. Conclusively, the implemented product and unique results by this approach can play a role to develop geo-based mobile application fields using many kinds of open sources and multiple types of public open data.

## 1. INTRODUCTION

Most mobile devices have web browsers embedded in them operating independently from the platform, but there are technical problems and inconveniences in the direct use of the existing desktop web services on mobile devices, such as a resolution issue. As the mobile users are increasing and the said issues are surfacing, technologies providing the same web services on mobile devices, as for the desktop, are being announced. The basis of such technologies is HTML5. HTML5 has an application programming interface (API) that allows the use of some of the existing technologies without plug-ins by installing the plug-ins on the desktop web browser. Among the representative technologies of HTML5 API are storage technologies that allow the activation of applications offline (e.g., Application Cache, IndexedDB, SQL Database) and the bidirectional communication technology (WebSocket), which is an improvement of the HyperText Transfer Protocol (HTTP) technology. There are various other technologies, such as location-based services, the touch user interface, and graphics. These technologies enable the provision of applications to the users instead of simple web services.

Guidelines and laws are being established to disclose the data possessed by national governments so that they can be used in various ways around the world. The U.S. established public data policies in 2009 (The White House, 2009), and Taiwan did the same in 2012 (Wang and Lo, 2016). South Korea also began to establish policies to open its public data in 2013 (Jung and Park, 2015). The data are downloadable as files or are provided through open API. Especially, the data closely related to everyday life are receiving much attention. According to the statistics on the public data portal (http://data.go.kr), an integrated website for the provision of public data in South Korea, the file data that have so far been used the most between 2011 and the present year (2016) include “major business districts database”, “status of education institutions”, and “number of floating population in Seoul”. The popular open APIs include “road name address query service”, “tour information service”, and “forecast information query service”. Most of these data are updated and added periodically rather than disclosed one time only. Therefore, more diverse information can be provided to the users by collecting the disclosed data.

Public data include geographic information. Some general data contain location or address information, which can be linked to geographic information. There are many different means of providing public data. One of these is to use easily accessible mobile web services to provide better services. The most basic method of service provision is

visualization, regarded as an important means of showing data to the users.

For the studies related to the mobile web, Kim et al. (2015) conducted a study related to the visualization of public data and geo-spatial information real-time data on the mobile web. Furthermore, a study was conducted on the editing of geo-based properties disclosed on the mobile web using the HTML5 technology (Kim & Lee, 2016). Some data provided through open API from public data confirmation agencies are updated at specific times. New information can be provided by collecting and showing or analyzing these data. Various factors need to be considered, however, to visualize the collected data on the mobile web, unlike the visualization of real-time data. In this study, the system was designed and implemented by considering the collection method, the first stage of visualization from the technical aspect.

## 2. COLLECTION OF PUBLIC OPEN DATA FOR MOBILE WEB

A separate server and a separate storage space are needed to collect and manage public data provided through open API. The Java-based Spring framework was used for the server designed in this study. According to the statistics of the web technology survey company W3Techs (https://w3techs.com), Java is one of the languages that are showing an upward trend as a server-side programming language. Numerous types of public data have been disclosed, and most of the data have different update times. The data collection efficiency can be improved by separately scheduling the public data update time. One of the open source libraries that manage schedules according to each job is Quartz. This can be used to automatically request data in line with the update time of each public data.

Another important means for data collection is the database, which is a storage space. Two types of databases were used in this study. The first was the relational database PostgreSQL. PostgreSQL is an open-source relational database that includes geo-spatial information storage and analysis feature (PostGIS). Some public data provide geo-spatial information in addition to general data, which can be merged and visualized when public data services are provided to help decision-making users. Therefore, a database that can store and analyze geo-based information was used. Furthermore, the database is used for the storage space to collect and manage general public data. Table 1 lists the schema for public data collection management.

**Table 1. Schema of collection management**

| Column name | Column type | Allow null | Comment |
|---|---|---|---|
| Name | Character | NO | Primary key. Always set a real name for open data.<br>It applies the collection name. |
| URL | Character | YES | Open-data providers' URL |
| EP | Character | NO | Open-data request end-point. |
| Time | Integer | NO | Request period |
| Status | Boolean | NO | The default is true. Set open-data collection conditions. |
| Comment | Character | YES | Comment about open data |
| Keys | Character | YES | Additional parameters for request (except end-point) |

Table 1 was designed with a total of seven columns. Most open API services have an end-point and variables for request. The column names of this table stored for requests based on such information are EP and Keys. The Time column allows the request time for each public data to be determined. PostgreSQL can also be used as a storage space for data collection. When public data are provided through the open API service, however, they are provided in Extensible Markup Language (XML) or JavaScript Object Notation (JSON), which are web-friendly. They are converted to the relational database format before being stored, and this can decrease the speed of data provision in the future. Especially, when considering services through the mobile web, a technique for processing and showing only the required data instead of all data is needed.

The NoSQL is an appropriate storage space technique for this method. NoSQL offers the advantage of faster queries for large-volume data compared to SQL, and is often used in the big-data field (Abbes & Gargouri, 2016). Various software applications have been released, and MongoDB, which has been disclosed as an open source, was used in this study. Figure 1 shows the design diagram for collecting public data. The system is driven by the Tomcat server. It was designed to collect and manage data while simultaneously using both relational and non-relational databases to collect and manage data. Furthermore, a user interface that users can easily manage was provided through

the management web interface.

The design in Figure 1(a) shows a technology that processes only the data needed in MongoDB in line with the visualization request. Furthermore, a user interface that can conveniently manage users through the management web interface was provided. The design part in Figure 1(a) shows the development of a technology that can quickly process only the data required from MongoDB in line with the visualization request. Therefore, the data can be efficiently provided on mobile web, which has a relatively lower performance than the desktop.

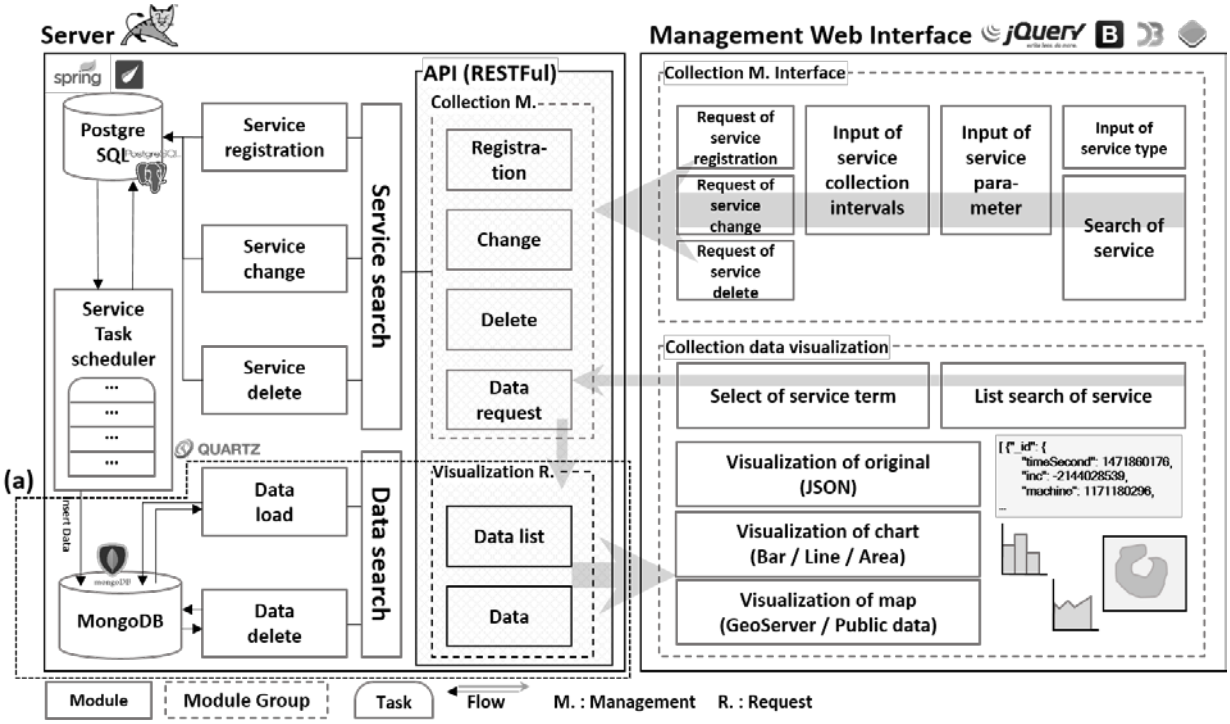


**Figure 1. System architecture for the collection of public open data for mobile web visualization.**

## 3. MANAGEMENT SYSTEM OF PUBLIC OPEN DATA

The public data collection management server is focused on user convenience and supports requests for the representational state transfer (REST) principle. Developers can utilize data using the REST principle, but it has difficult aspects for general users. Therefore, the Dashboard service for the management system user interface was implemented to allow general users to conveniently use the services.

Dashboard shows the public data collection status. Furthermore, data can be added, modified, and deleted, and the collected data can be visualized. The environment for running Dashboard is outlined in Table 2. Application Server is a web application server that contains a public data collection feature, and the client is the Dashboard service using HTML5. The collected data can be visualized as graphs and charts through D3.js, and the stored geo-spatial information is visualized through OpenLayers.

Figure 2-6 show the visualization results for the collected data on Dashboard. Figure 2 shows the direct visualization results for the collected raw data. Figure 2(a) shows the data identification ID, Figure 2(b) shows the public data that responded at the requested time, and Figure 3 shows the visualization of the environmental data PM25 in a chart, among the collected data. It is possible to visualize data for a single day and to group data based on a period. The data collected in the server can be visualized in various ways based on appropriate criteria. Figure 4 is the result of chart visualization: bar chart and area chart. Figure 5 shows the visualization results for geo-based information. The geo-spatial information stored in PostgreSQL can be visualized through the mapping library OpenLayers. The information can be instantly visualized on the mobile web because the mobile web was considered in the designs of the data and technology. Figure 6 is the result of weather observation position mapping using public open data.

**Table 2. System environment of Dashboard for open data management**

| Environment | | Name | Version |
|---|---|---|---|
| Application Server | Web application server | Tomcat | 8.0.36 |
| | Web framework | Spring boot (Spring framework) | 1.3.3 (4.2.5) |
| | Scheduler | Quartz | 2.2.1 |
| | Template engine | Thymeleaf | 2.1.4 |
| | Database connector | PostgreSQL | 9.1-901-1 |
| | | MongoDB | 2.13.13 |
| Client | JavaScript library | jQuery | 1.11.2 |
| | User interface | BootStrap | 3.3.6 |
| | Chart library | D3 | 3.5.16 |
| | Mapping library | OpenLayers | 3.14.2 |


**Figure 2. Result of source visualization in Dashboard: (a) unique identification key; and (b) public open data.**

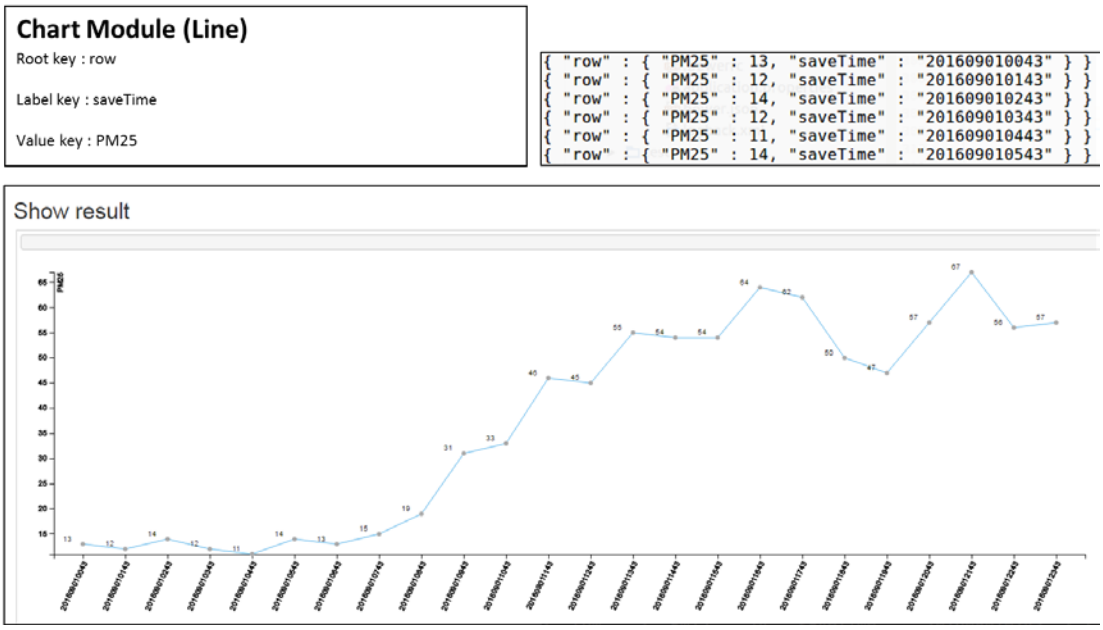


**Figure 3. Result of chart visualization in Dashboard with regard to the period (2016/09/01 00:00-24:00, PM25).**

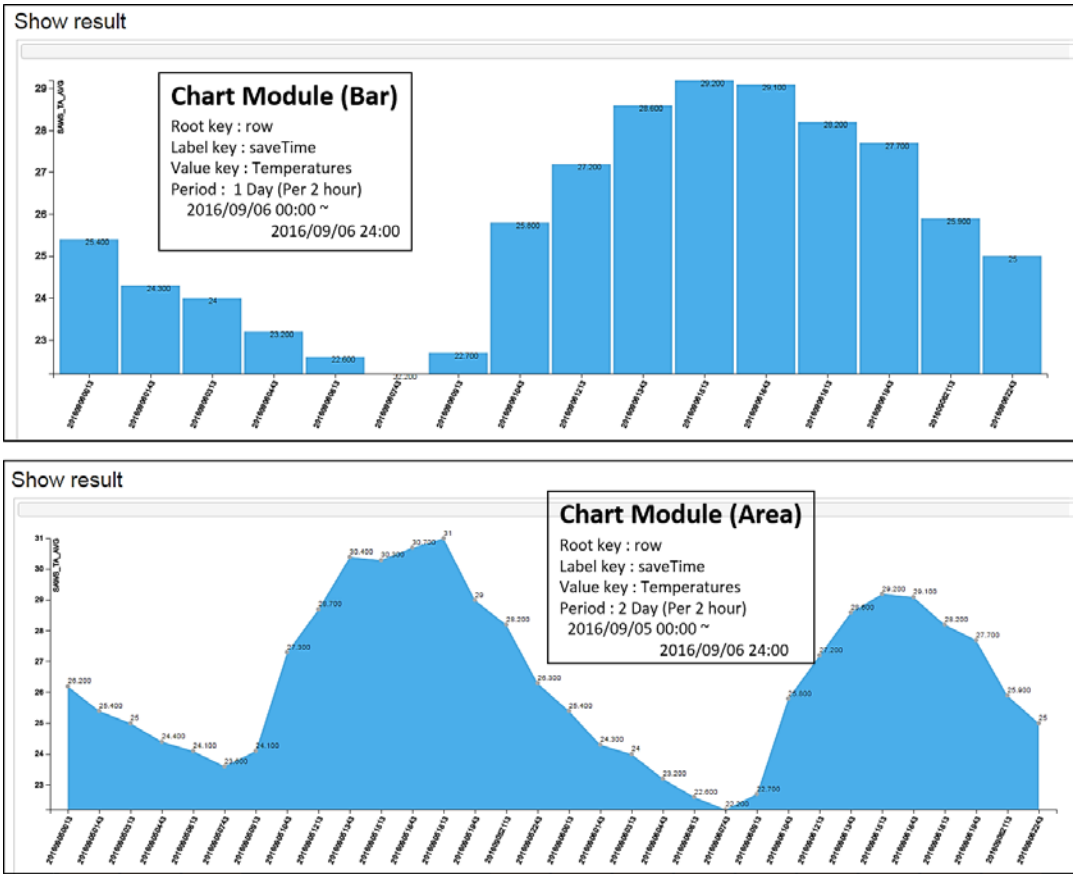


**Figure 4. Result of chart visualization: bar chart and area chart.**

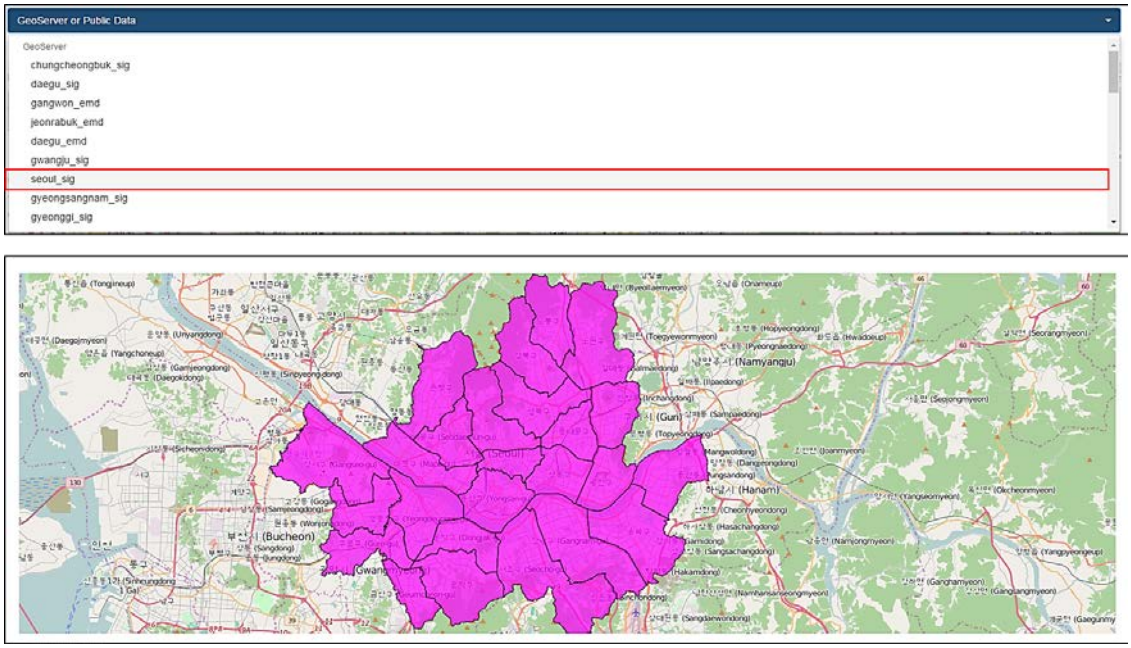


**Figure 5. Result of geo-based visualization in Dashboard: administrative district in Seoul**

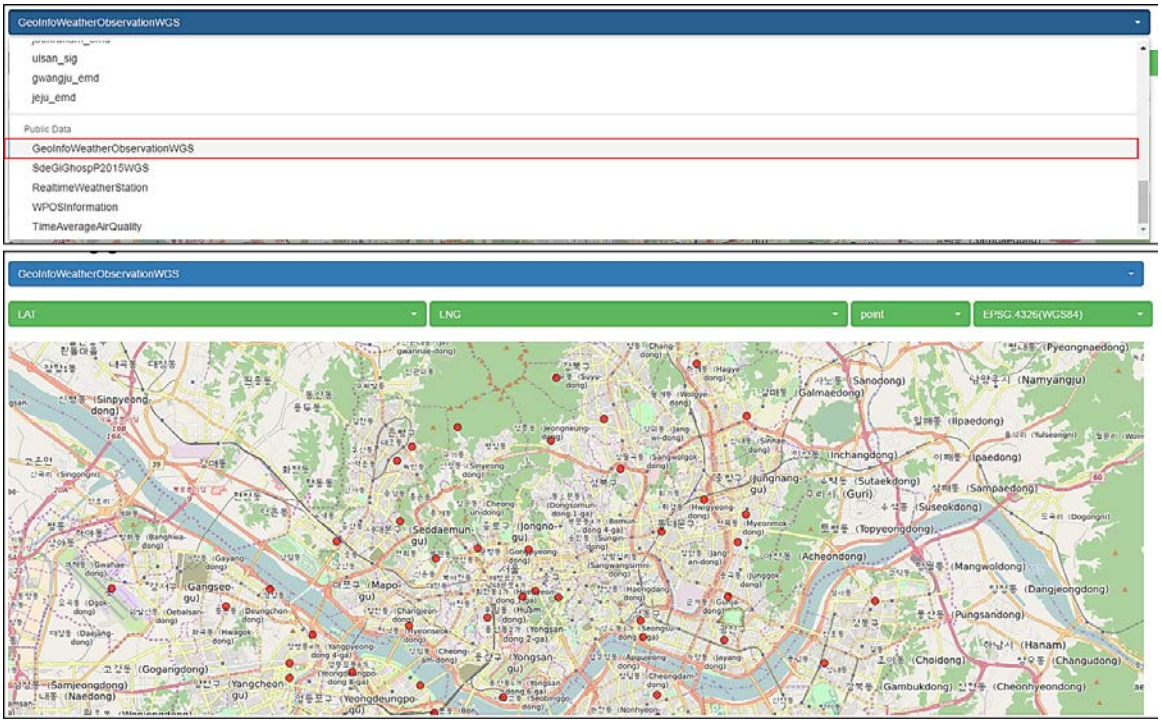


**Figure 5. Result of weather observation position mapping using public open data in Seoul.**

## 4. CONCLUDING REMARKS

With the development of mobile communication technologies along with information and communication technology (ICT), users want to obtain various kinds of information anywhere they go. Therefore, the development of services using smart devices is expected to continue in the future. In particular, web services that have good accessibility because they are already installed in most smart devices will play a key role in service provision in the future. Furthermore, the data possessed by the country have added public benefits, in contrast to personally collected and processed data. Thus, reliable information can be provided to users. In the future, policies for efficient data disclosure are expected to be established in a better direction, and the number of open data is expected to increase. It

is not easy to provide services, however, considering all these suggestions, and the diversification of mobile devices can make this more and more difficult. The public data collection and visualization techniques have been designed considering the above-mentioned issues. In particular, the addition of geo-spatial information will help users make better decisions compared to general visualization by adding geo-spatial information. In the future, more diverse visualization methods that merge mobile web public data and geo-based data are expected to be established, and the number of various visualization methods that have merged public data and public geo-spatial data will be investigated.

## ACKNOWLEDGEMENTS

This research was supported by the National Land Space Information Research Program from the Ministry of Land, Infrastructure and Transport, Korea (No. 14NSIP-B080144-01).

## REFERENCES

Abbes, H. and F. Gargouri, 2016. Big data integration: A MongoDB database and modular ontologies based approach. International Conference on Knowledge Baed and Intelligent Information and Engineering Systems, Sep. 2016.

Jung, K. and H. Park, 2015. A semantic network analysis of South Korea's "Open Public Data" policy. Government Information Quarterly, 32 (3), pp.353-358.

Kim, K., H. Park and K. Lee, 2015. Mobile application of open source stack to geo-based data visualization on e-government web framework. Free and Open Source Software for Geospatial 2015, Sep. 2015.

Kim, K. and K. Lee, 2016. Real-time processing of spatial attribute information for mobile web based on standard web framework and HTML5. Spatial Information Research, 24 (2), pp.93-101.

The White House, 2009. Open government directive, Retrieved Sep. 06, 2016, from http://www.whitehouse.gov/sites/default/files/omb/assets/memoranda_2010/m10-06.pdf.

Wang, H. and J. Lo, 2016. Adoption of open government data among government agencies. Government Information Quarterly, 33 (1), pp. 80-88.