A Study of Evaluation Index Model Designed to Measure Open Data Outcome

Mi-Hwan Hyun, Dong-Gu Shin* and Ju-Yeon Shin
Korea Institute of Science and Technology Information, Yuseong-gu, Korea, 34141, South Korea; mhhyun@kisti.re.kr, lovesin@kisti.re.kr, jyshin0417@kisti.re.kr

Abstract

Objectives: As data opening becomes more common, assessment model and index designed to measure the concept and outcome of open data have been continuously upgraded. Methods/Statistical Analysis: For open-data assessment index development modeling, the Open Knowledge Foundation’s open definition and the Obama Administration’s definition of ‘open data’ were basically adopted. In particular, the model was designed with the open-data attributes of the Obama Administration which has systematically suggested open data in the most specific manner including the OKF’s definition as the important scale of the basic model. Findings: This study attempted to propose assessment indexes which can measure outcome by applying them to science and technology open data provided by the Korea Institute of Science and Technology Information (KISTI) after reviewing current leading assessment models and indexes as well as definition of open data. The purpose of this study is to reconfigure open-data assessment which has been performed on government data with new assessment model and index which are applicable to the science and technology open data provided by the KISTI. In terms of a scope of assessment index development for outcome measurement, this study was performed against science and technology open data which have been provided in a liked open data format by the KISTI. The results would be directly used or referred to in the assessment of open data services by the government agency or information center which has managed open data as well as the KISTI. Improvements/Applications: The results would be directly used or referred to in the assessment of open data services by the government agency or information center which has managed open data.

Keywords: Linked Open Data, Open Data, Open Data Index, Open Government, Web Index

1. Introduction

Recently, data have emerged as a key element in ‘New Economy’ and ‘Knowledge Information Society’. The terms which represent a data society such as linked data, open data, big data and gov2.0 focus on the possibility of using data in a previous stage, not on the processed information. Open data have spreads around the globe as major states declare ‘open government’. In addition, technologies have matured enough to back them up such as Semantic web and linked data. As a result, there has been a rising interest in open data and the amount of open data has rapidly increased. As more countries declare data opening, there has been assessment on if the government has properly promoted the open data policy. Specifically, there have been studies which assess the level of public data opening in each country by analyzing how much important data have been opened and if the available data are beneficial to users.

The typical assessment indexes include Web Index, Global Open Data Index and Open Data Barometer (ODB). With the announcement of these assessment indexes and results, each country’s open-data policy results have been unveiled and ranked, causing a positive effect on the importance and spread speed of open data. Even though the governments and public agencies have worked hard to assess the quality of open data and measure their outcome, there hasn’t been any internationally standardized assessment model or index yet. The purpose of this study is to reconfigure open data assessment which
A Study of Evaluation Index Model Designed to Measure Open Data Outcome

has been performed on government data with new assessment model and index which are applicable to the science and technology open data provided by the KISTI.

2. Material and Methods

2.1 Open Data

In terms of the definition of the open data of the Open Knowledge Foundation (OKF)\(^3\) and the Obama Administration, the features that open data should have are clearly described. This definition is directly applied to open-data indexes such as Open Data Barometer as well. This study aims to use the concept of open data defined by the OKF and the Obama Administration as a key scale to develop science and technology open-data assessment indexes. The OKF suggests the basic and important matters that must be known to promote open data and open the data, using the open data handbook as seen in Table 1.

It is government data which play a key role in an open-data ecosystem. The Obama Administration defines ‘open data’ as seen in Table 2.

| Features | Definition |
|----------|------------|
| Availability and Access | The data must be available as a whole and at no more than a reasonable reproduction cost, preferably by downloading over the internet. The data must also be available in a convenient and modifiable form. |
| Reuse and Redistribution | The data must be provided under terms (preferably license terms or legal provisions) that permit reuse and redistribution including the intermixing with other datasets. |
| Universal Participation | Everyone must be able to use, reuse and redistribute. For example, ‘non-commercial’ restrictions that would prevent ‘commercial’ use, or restrictions of use for certain purposes (e.g. only in education), are not allowed. |

Table 1. Definition of 'Open Data' in Open Knowledge Foundation (OKF)'s Open Data Handbook

| Features | Definition |
|----------|------------|
| Public | Among the data developed by government agencies and related organizations, all data except for minimum regulations such as personal information confidentiality and security have publicity. |
| Accessible | Open data should be easy to use and editable for search, download and indexing. In addition, they should be available in an open format or machine readable format. Open data should be designed for all users without any special benefits for a certain individual or organization, satisfying the purpose of use. As a result, they can be distributed in diverse formats. In addition, this format should be publicly accessible without any regulations. |
| Described | Open data should explain data strength, weakness, limit and security requirements as well as the information needed for users to handle data in a detailed and sufficient manner. For this, open data should provide diverse information such as samples as well as detailed documents and data dictionary. |
| Reusable | Open data should be distributed in an open license form, and they should not be limited in use by any regulations. |
| Complete | Open data should be available and provided in their original form. In case of secondary sources, the original data should also be open all the time. |
| Timely | Open data should be distributed as quickly as possible to maintain their value. The frequency is decided by major users’ and end-users’ needs. |
| Managed Post-Release | The manager’s contact information should be stated to respond to the use of open data and users’ needs and demand. |

Table 2. Definition of Open Data by the Obama Administration
2.2 Science and Technology Open Data
The KISTI has opened science and technology information such as papers, patents, reports, standards and trends which have been provided through the science and technology information portal National Digital Science Library (NDSL)\(^4\). Since 2015, a Resource Description Framework (RDF)-type linked open data have been established and open to the public to increase the usability of shared data and build a data-based knowledge ecosystem through semantic connection among the open data\(^5\). The types of science and technology open data established by the KISTI include the authors of domestic academic papers, domestic academic papers, national R&D reports, patents, journals, proceedings and bio-information. A total of 2 million data are open in a linked open data format. In terms of science and technology open data, metadata and authority data were developed in a linked data format based on journals, papers, authors, patents and other information\(^6\). To increase the usability of the data, in addition, the linked data from other agencies and interlinking data on papers, journals, authors, agencies, reference and others were developed. In terms of a scope of assessment index development for outcome measurement, this study was performed against science and technology open data which have been provided in a liked open data format by the KISTI.

3. Analysis of Research Trends on Open Data Index

3.1 Web Index
The Web Index is the world’s first multidimensional measure developed by the World Wide Web Foundation. It measures the political, economic and social influence of the Internet, ranks countries and opens the results to the public. The sub-indexes designed to measure the Web Index include universal access, relevant content, freedom and access and empowerment. Among these four sub-indexes, it is freedom and access which are related with open data. The structure of the Web Index assessment model is seen in Figure 1.

3.2 Open Data Barometer (ODB)
The ODB is a report developed by the World Wide Web Foundation (WWW) and Open Data Institute after analyzing and ranking each country’s open data around the globe\(^2\). While the Web Index measures the political, economic and social influence of the Internet broadly, the ODB concentrates on the measurement of each state’s open data. In the ODB, it is critical to assess open margin on the core dataset. The ODB-applied open data indexes include readiness, implementation and emerging impact. The ODB is applied to actual assessment with the background of the Web Index and 5-Star Open Data. In particular, it performs index development and assessment by all interested parts related with the data such as the government, businesses and citizens. Therefore, it is available as the latest and most reliable result. Based on the results of the ODB, top-ranked states’ current status is referred to. Then, it is available as the data which can back up a domestic shortage. Then, domestic studies can be performed by reducing the scope and applying the same index research method.

Figure 1. Structure of web index assessment model.

3.3 Global Open Data Index
The Global Open Data Index is a project operated by the OKF, which assesses and ranks the government’s data opening policies and implementations around the world. The Global Open Data Index ranks the data based on open data experts’ assessments on if the data annually created by the government in an open format can be easily accessed by citizens, mass media and citizens’ groups and well utilized. A key concept of the Global Open Data Index is that an open type is assessed, focusing on a key dataset.

The assessment procedure can be divided as follows:
1. Securing 10 core datasets,
2. Verifying if the information is accurate by local and data experts. Data collection is performed in two steps. In step 1, the dataset-related data are collected through volunteers. In step 2, the results are verified by experts (volunteers). Each dataset is assessed by 9 different questions. The biggest feature of
the Global Open Data Index is to have a systematic review system including a data proposer, reviewer and expert for the assessment of open data even though the assessments on the index differ. The index with the most important weight in the Global Open Data Index is a license-related item. The clear delivery of open data under a certain condition is considered as the element having the biggest effect on usability.

3.4 Open Data Readiness Assessment

The Open Data Readiness Assessment (ODRA) is a methodology designed to assess each country’s status and current situations in which the government’s open data are oriented by the World Bank’s working group on open government data. The ODRA is a single component of the Open Government Data Toolkit. The toolkit is comprised of Open Data Essentials, Technology Options, Demand and Engagement, Supply and Quality of Data and Readiness Assessment Tool. The ODRA is the last component of the total toolkit. It is in a draft stage until now and has been continuously updated. The World Bank’s Open Government Data Toolkit introduces all specific matters that the open government should prepare, instead of open data. However, the construction and procedures of the data toolkit are applicable to all types of agencies that prepare open data policies and services. In particular, the major indexes and specific questions which are prepared by the Readiness Assessment Tool which comprises the last part of the toolkit are more specific and detailed than any other cases. Another viewpoint of the total toolkit is that a guideline which is available as the standard can be supplied to the agencies which prepare open data prior to their assessment from the data preparation stage. Then, it is needed to distribute it for assessment based on the guideline.

3.5 Five Star Open Data

The most general type and level of open data are available in Tim Berners-Lee’s Five Star Open Data (http://5stardata.info/ko/). The Five Star Open Data explains the five stages of open data, using star rating and describes the examples and costs and effects in each stage. It is a key index which is cited and used directly in assessing the data quality of major open data projects such as the Web Index, the Open Data Barometer and the Global Open Data Index. Among the major features in each stage, machine readable format, open-format, URI and linked data except for open license are available as direct assessment indexes as the technical elements which can check actual open data directly. The open license is managed as a separate measure, an element that can perform diverse assessments according to the license level. In addition, it is available as an index which can assess directly by the license level.

4. Science and Technology Open Data Assessment Index Model

The science and technology open data assessment index model developed though the revision of the Web Index consists of six categories as seen in Figure 2: Accessible, Usable, Reusable, Complete, Management System and Community. After reflecting the features of science and technology open data, different weight was applied by category.

Accessible is an index which accounts for 20% in total assessment models. It assesses if open data and service are properly developed for all users and to meet the purpose of use. Availability (20%) is an index which assesses if data can be downloaded through the Internet and the costs which are consumed in the conversion of raw data are rea-
sonable. Reusability (20%) is an index which examines if the open data have the license for use, reuse and redistribution and they are limited in usage. Completeness is an index which accounts for 10% in total assessment models. It assesses if the data are provided in a readily executable and usable form and they are always provided with raw data in case of the secondary sources (when raw data are altered by the user or data proposer). The key assessment details are as follows: The Management System (15%) is an index which assesses if data opening-related policies, organizations and legal/institutional systems exist and are properly operated. Lastly, community (15%) is an index which examines if there is continued communication with the private developers or users who use the open data. The core assessment details by sub-index are seen in Table 3.

5. Conclusion

People have interested in data since a long time ago and their importance has been emphasized continuously. With the considerable growth and development of web technology, however, there has been reconsideration on raw data. In addition, each country’s open government-oriented new government system has established a turning point in opening the data era. The opening of government data is a critical domain which establishes the ground for a data world. As a result, its importance has been further emphasized over time. Therefore, the Web Index, Open Data Barometer and Global Open Data Index mentioned in this study all focus on assessment on the opening of government data. A dataset is a new information resource for the information management organization and it has emerged as a new service target which should be properly managed and preserved. In fact, previous data developed by many organizations have already been provided in diverse formats such as file, OpenAPI and LOD. However, there hasn’t been assessment on the information resource management organizations except for the government area.

The open data assessment model proposed in this study covers internal policies, progress and current status of the information system in addition to the aspects which can be intuitively confirmed on a user service page.

Table 3. Core assessment items by sub index

| Category               | Assessment Details                                                                                                                                 |
|------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------|
| Accessible             | - Are the data provided considering all users if possible?  
                      | - Are the data provided in an easily accessible and usable manner without any regulations?  
                      | - Are the data provided in an open format or machine readable format?                                                                          |
| Availability           | - Are all data downloadable through web?  
                      | - Are there any costs needed for data collection and processing?  
                      | - Are the data easily editable?  
                      | - Are data strength, weakness, limit and security requirements in addition to the information needed for data processing sufficiently explained?  
                      | - Is detailed metadata-related information provided, and are glossary of terms such as controlled vocabularies and sample data provided? |
| Reusability            | - Are the data provided in an open license format?  
                      | - Is there any problem in use, reuse and redistribution when licenses in addition to open license are provided?  
                      | - Are the data usable, reusable or redistributable by all people regardless of the purpose of use (either commercial or non-commercial)? |
| Completeness           | - Is data correction-related information provided?  
                      | - Are the data provide in a readily usable form?  
                      | - Are the raw data provided as well in case of secondary sources?  
                      |
| Management System      | - Are there special units, policies or documents from policy perspective?  
                      | - Is there any financial support or investment for the management of open data?  
                      | - Is there any PR activity for data opening?  
                      |
| Community              | - Can users’ opinions be reflected on open data policy planning?  
                      | - Are there diverse communication channels through which users are able to participate in open data services?  
                      | - Are there communication channels for the secondary sources and applied services?  
                      |
In an initial assessment, therefore, it is needed to perform the 2nd assessment through outsiders after carrying out the 1st assessment through the experts who are familiar with the internal situations or internal managers. The assessment models and indexes mentioned in this study have a limitation in that they were reconstructed with the government-targeted indexes. However, they may bring model advancement and improvement in index construction based on the research modes and indexes.

In current indexes developed for the assessment of open data, both qualitative and quantitative assessments are mixed. They are constructed in a complicated manner to assess the assessment indexes and data quality from policy and technical aspects. It is public data owned by government agencies, which can make the biggest quantitative contribution to the early-stage data ecosystem. The assessments on policy elements such as government commitment are essential. From the policy aspect, citizens' participation is an important index. The physical investments such as information communication technology infrastructure that citizens are able to make include comprehensive elements such as education for the support of data utilization. The biggest benefit data can bring lies in the creation of new values through the connection and interoperation of open data. Therefore, open data themselves should have technical maturation and they are generally included in the technical assessment elements.

In the typical cases such as the 5 Star Open Data, Web Index, Open Data Barometer and Global Open Data Index as well, assessment indexes are constructed by mutually accepting the indexes individually proposed. However, it is difficult to assess the quality of these assessment indexes. In fact, there should be continued observation and analysis on the latest assessment results (ODB) and outcome of the continuous assessment (Global Open Data Index).

Until now, for open data quality assessment, basic data have been formed with actual open data opening cases and documentary evidence (e.g. data dataset, data service system, etc.) and assessed by experts. In terms of documentary evidence, however, a system which assesses data opening (e.g. public transportation timetable, election results, etc.) on a certain domain according to importance and ripple effects and qualitative aspects is adopted. Therefore, there should be efforts to develop open data assessment indexes and discover and spread domestic data on the dataset which is deemed as the world's valuable data.

In practice, data are assessed by citizens or experts. Therefore, it is needed to secure professional staff from diverse fields and develop assessment procedures. Even though there is no demand for particular domain knowledge (e.g. technical completeness on the open data, fault, etc.), it is important to have professional manpower in data knowledge for data assessment in a qualitative aspect. From an innocent user's perspective, furthermore, there should personnel who can perform the assessment by dividing the staff into technicians and non-technicians. In case of machine readable data, the general public who have no technical background are able to assess spreadsheet-style issues. However, the data which observe the RDF technical rules and OpenAPI should be assessed by actual developers. For this, the backup personnel should be organized.

In open data, the government's contribution is critical. For the creation of visible outcome, however, citizens' participation and use and data use by a private sector and creation of results are needed. Therefore, it is necessary to keep observing open data movement and the world's open government policies as well as a study on open data assessments. Furthermore, basic infrastructure which can provide high-quality data and services should be established in advance after developing and distributing the preparation and assessment guidelines or instructions to handle open data prior to the development of assessment indexes.

6. References

1. Choi HR, Kwak CH, Cho M. A study to improve the public data management of the city of Busan. Indian Journal of Science and Technology. 2015; 8(15):1–6.
2. Hyun M, Shin S, Kim H. The method for quality evaluation of open data in Korea. International Conference on Convergence Technology. 2015 Jul; 5(1):878–9.
3. Open Knowledge Foundation. Open Data Handbook Documentation. 2012. Available from: http://opendatah andbook.org/en/
4. Yu S. A systematic review of studies about patient privacy information protection in the time of convergence. Indian Journal of Science and Technology. 2015; 8(35):1–5.
5. Hyunsook C, Jeongmin K. Utilization method of linked data profiles connecting curricula and digital content. Indian Journal of Science and Technology. 2015; 8(59):373–83.
6. Hyun MH, Shin SM, Kim HS. Application of linked open data on sharing library data in Korea. Indian Journal of Science and Technology. 2015; 8(27):1–4.
7. Davies T. Open Data Barometer: 2013 global report. World Wide Web Foundation and Open Data Institute; 2013.
8. Martin S, Foulonneau M, Turki S. 1-5 stars: Metadata on the openness level of open data sets in Europe. Research Conference on Metadata and Semantic Research. Springer International Publishing; 2013 Nov. p. 234–45.