

Research on Quality Characteristics of Public Open Data

Go-Eun Park¹ and Chang-Jae Kim²

¹*The Graduate School of Software, Soongsil University
Seoul - South Korea*

²*The Graduate School of Software, Soongsil University
Seoul - South Korea*

**Corresponding author: Chang-Jae Kim*

¹hnnbmn@naver.com, ²winchang@ssu.ac.kr

Abstract

The object of opening a public data is to make them freely used by public so as to improve quality of civil life and activate new industry and job creation. Opening a public data is a policy that has been importantly emphasized around the world and active policy for public open data is creating successful cases. Public open data needs proper quality to achieve public objects. However, quality management and a lack of standardization cause error data found and a lack of availability, and a lack of guideline about quality. This research figures out quality characteristics necessary for public open data through analysis and verifies establishment of model through survey of professionals. Quality characteristics of public open data verified by survey are publicity, availability, credibility, and suitability. It has its significance for suggestion of quality characteristics necessary for improvement of quality and availability of public open data.

Keywords: *Public Open Data, Data Quality, Data Quality Characteristics*

1. Introduction

The objects of opening public data has anyone freely reuse public data so as to improve quality of life and create new industry and job creation [1].

Opening public data is an important policy emphasized around the world, and successful cases thereof have been made through active policy in USA and EU.

On September 2015, opened public data in the portal (data.go.kr) are: file data (12,680), open API (1,796), data visualization (251), post (3,600); total of 18,327. Types of data are transportation logistics (112), environment meteorology (90), industry employment (70) and safety and health (54). Korean government first set institutional basis and minimum guidelines through 『Quality Management Guideline of Public Institute's Database』 (2011.7) for quality management of public data, and National Information Society Agency (NIA) provided quality management and evaluation and improvement of public data with 『Public Data Quality Management Manual』 (2012.1). NIA's public data quality management support center has been active for opening public data doing quality improvement of national important database, portal operation, education and promotion, guideline development and performance of professionals. However, use of public data has not been made as expected and its output has not achieved its goal yet. The main problem is error data found and low availability occurred from quality management of public data and a lack of standardization [2].

Guidelines about quality management of opening public data helpful for direction of management in every institute are, not to mention, absent. Opening public data is mainly done focusing on providers so that its low availability arises as problems.

The object of this research is to find out quality characteristic and additional characteristic necessary for opening public data. It has its significance by suggesting quality characteristics as a direction of improvement of opening public data.

This research executed prior research on data quality, public data quality and public service quality to find out quality characteristics of opening public data. Also, definition and standard of opening public data in USA among foreign countries' cases was analyzed.

Next, quality and additional characteristic of opening public data were deduced through data quality characteristics in public service quality and prior research.

Survey on current supervisors and executives of public data in public institute was done to verify quality characteristics found out in this research. Through survey whether current public institute's quality management complies with standard and the thesis was formed reflecting survey modifying model and state conclusion and future research in fifth chapter.

2. Related Work

2.1. Data Quality and Quality Characteristics

This chapter deals with definition about analyzed data quality. English was defined in data quality to the level that satisfies users' various purpose of use and satisfaction [3]. Wang *et al.* defined focusing on whether it is fit for use [4]. Larry defined data quality's object is to improve satisfaction between user and relevant parties to the level of user's convenience [5], and Redman defined it requirement satisfaction of user [6].

In this research, data quality is defined as "a level of data to satisfy purpose and requirement of user" integrating details above.

Data quality could be defined by a group of quality guideline and data quality guideline is a group of data quality property constantly required by majority users [7].

Data quality characteristics are also defined as data quality property or quality standard and quality component, but according to international standard ISO/IEC25000 and as quality related property is defined as quality characteristics, this research also defines it as quality characteristics.

Korea Database Agency(KoDB) defined details classifying validity and availability with the data quality standard in DQC (Data Quality Certification) [7], and suggested public data classifying them as structured and unstructured data; quality standard of structured data is completeness, uniqueness, validity, consistency, accuracy; data quality for unstructured data is functionality, reliability, usability, efficiency and portability [8].

25012, Quality Model Division among 5 quality evaluation model of ISO/IEC, has a data quality evaluation model simultaneously functioning system software in computer. Data property suggested here are 15 types: accuracy, completeness, consistency, credibility, currentness, accessibility, compliance, confidentiality, efficiency, precision, traceability, understandability, availability, portability and recoverability [9].

2.2. Public Data Quality

"Public Data" are database, electronic form of documents or materials in electronics or light form for the purpose of legislation and public institute [10].

Data quality in "Administrative guideline of public data" of Ministry of the Interior has useful value securing latest, precise and mutual data [11], and "Public Information Quality Management Manual" (Ver2.1) is top 8 guideline for measuring public institute's data quality that suggests 24 detailed object reflecting preparation, usefulness, readiness, security, completeness, accuracy, timeliness, and consistency [12].

Initial research on public data's quality is likely to have a microscopic view on accuracy. However, influence in problem in actual use cannot guarantee quality only by accuracy. Moreover, opening public data has characteristics that need various requirements. In examples of opening public data overseas, Office of Management and Budget in USA

established Data Quality Act (DQA) in December, 2000 and suggested guaranteeing information's quality, objectivity and integrity of federal government, and establishing data quality system proper for federal and state government became mandatory [13]. Quality component related with this is stated in Table 1.

Table 1. Data Quality on Data Quality Act of USA

Quality Characteristics	Definitions
Utility	Usefulness of the information to its intended users
Integrity	Security of information — Protection of the information from unauthorized access or revision, to ensure that the information is not compromised through corruption or falsification
Objectivity	Focuses on whether the disseminated information is being presented in an accurate, clear, complete, and unbiased manner, and as a matter of substance, inaccurate, reliable, and unbiased

This index shows a definition of concept and standard stays at high level compared to general quality index, so detailed and proper index are made to apply to business and environment of each institute.

Meanwhile, the model of WWW (World Wide Web) foundation's Tim Berners Lee's 'Five Stars of Linked Open Data' designates important point of opening public data [14].

5 Stars model suggested 5 level of development of open data as star 1 to star 5. This model is based on a form of data that effect of obtaining relevant knowledge gets higher. A level of star 5 means relevant data could be find and maximum of relevant knowledge and use due to direct learning about data schema. Through this model the important point of reprocess and connection of characteristics of opening public data could be learned.

3. Quality Characteristics of Public Open Data

Figure 1.shows the deduced public open data's quality characteristics model of this research. This research deduced quality characteristics of public open data through point of existing characteristics of public data and public institute's guideline of data quality management.

The referenced standard to deduce quality characteristics of public open data is DQC's "DQI(Data Quality Indicator)" existing standard of data quality, and international standard - ISO/IEC 25012, NIA's "Public information quality management manual" and USA's data quality of "Data Quality Act" [8-9,12-13].

Through mapping referenced characteristics, quality characteristics of public open data in accordance with requirement of public open data including basic quality characteristics is deduced.

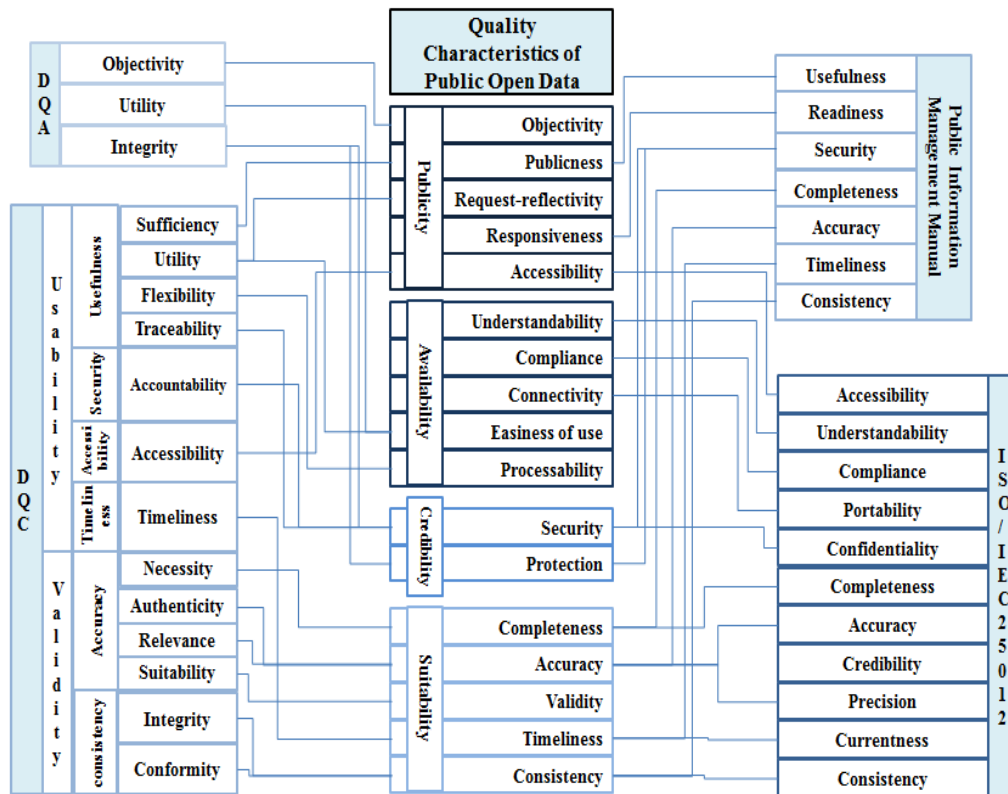


Figure 1. Deduced Quality Characteristics of Public Open Data

Opening public data means making ‘public data available to anyone’, and opened data should be fairly provided to requesting public for public benefits [11].

Focusing on the fact that opening public data means ‘data that creates public benefits by free of use and access’ reflecting object of opening public data, 4 aspects are dealt with for necessary quality of public data: publicity, availability, credibility and suitability, and its definition is shown in Table2.

Table 2. Main Characteristics of Public Open Data Quality

Main Characteristics	Definition
Publicity	Provided fairly to public for public benefits, and the data are partial or do not represent a part of benefit
Availability	Easy to create value through process and use of data
Credibility	Data are protected with technical security and confidential info are processed with proper protection
Suitability	Figures of data are complete and accurate so that proper to use

3.1 Publicity

Publicity is defined as characteristics of public open data to satisfy quality of public data necessary as public service. Publicity contains fairness of provided data and provided target. They do not represent a benefit of some parties and are set for improvement of public benefits. Publicity’s additional characteristics and definition are accessibility, objectivity, publicness, request-reflectivity and responsiveness.

Table 3. Definition of “Publicity”

Publicity	Definition	References
Accessibility	<ul style="list-style-type: none"> ·No subject discrimination on the data provide. · User's data acquisition is constituted by a official portal-based set. 	ISO 25012 (Accessibility), DQC (Accessibility), Public Data Management Guidelines (Usefulness)
Objectivity	<ul style="list-style-type: none"> ·The source of the data is proven by scientific or statistical techniques. ·The source of the data is recorded and managed. ·Raw Data is provided preferentially - free from third-party judgment and opinion. 	DQA (Objectivity), DQC(Flexibility)
Publicness	<ul style="list-style-type: none"> ·The data is not biased nor represent specific interest. ·Provided data is meet the quantity and quality sufficient of the user's purpose. 	Public Data Management Guidelines (Usefulness), DQC (Sufficiency, Utility)
Request-reflec tivity	<ul style="list-style-type: none"> ·User's requirements for data modification or offer requests are systematically managed and reflected. ·The data meet the purpose of use and value sufficient. ·Business rules are established and performed of nondisclosure information. 	DQC (Utility)
Responsivene ss	<ul style="list-style-type: none"> · Establish for open data standards through regular survey and analysis of user demand, satisfaction, and data utilization status. 	Public Data Management Guidelines (Readiness), ISO25012 (Compliance)

Accessibility is defined as fairness of provided target and easiness of obtaining data. Public data become more available based on specified portal and shall not have discrimination to provided target. To have data satisfy users, necessary interface and infrastructure to make data available are necessary, not to mention improvement of own quality.

Objectivity is about whether source of data are scientific and proved statistically, and raw data that is free of judgement and opinion of the third party are first given.

Publicness is about whether is improves public benefit obtaining enough value for object not representing benefits of some parties. This is a characteristic not shown in existing data quality characteristics. Most data belong to specific institute or company and are processed for benefits of specific parties, so this characteristic was not focused. However, public open data is a data for public benefits, not for benefits of specific group or owner, so this research suggests this characteristic as those data have different meanings in different group maintain neutral.

Request-reflectivity is about whether reflectivity about data modification and provision for users quick and satisfied. It could be seen as extension of accessibility suggested above, and characteristic of public open data that provides value after managing and processing accuracy of data.

Responsiveness is a level of gathering and reflectivity of opening standard through demand/satisfaction/current state of use. Request-reflectivity is a quality made from active behavior of users' request and modification to data itself, and responsiveness is a quality

made from active survey of party of opening data.

3.2. Availability

Availability is a necessary characteristic for public open data to recreate value through free processing. Opening public data needs availability for its recreation of value through application of data and ICT technology beyond satisfying right-to-know of people and transparent government. Details of availability consist of understandability, processability, connectivity, easiness of use and compliance.

Table 4. Definition of “Availability”

Usability	Definition	References
Understandability	·Understandable for various users without confusion or distortion.	ISO 25012 (Understandability)
Processability	·No limitations for data processing, such as hardware or software limitations. ·No need for special support or configurations for using data.	DQC(Flexibility)
Connectivity	·The leverage is available in variety environments and situations. Also, can be used in conjunction with other data(different from context, systems, institutions, and countries.) · Redundant data inside the institution or associated data have a consistent management standard.	ISO25012 (Portability)
Easiness of use	·Data can be provided for users with a familiar format. ·User satisfaction is high.	DQA (Utility), DQC(Utility), Public Data Management Guidelines(Usefulness)
Compliance	·Compliance with the administrative standard terms and codes. Easy for Integrated government data linkage and common use of administrative information ·Compliance with government data standards and institution data standards.	Public Data Management Guidelines (Readiness), ISO25012 (Compliance)

Understandability is defined as recognizable data without distortion or confusion when being translated by various users. Its value should be translated same meanings and described clear and consistent.

Processability is a characteristic of data that does not require specific condition or technology such as hardware or software. Public open data are different in different places by environment and infrastructure, so technological limit should not exist. It is important that provision of data should be done both in format of machine-readable and maximum of availability not consuming additional cost and procedure in processing. This reuse of availability of data is called processability. Processability also contains easiness of use applicable to other businesses.

Connectivity is a necessary characteristic for use in different condition and environment (different data in context, system, institute and nations). A character that can support and create new value through easy mash up among data is called connectivity of data. Nonexistence of disturbance in connecting of data with other structure of data or technological component is important character of public open data quality.

Easiness of use is a character that guarantees users' familiarity and satisfaction of use. Data format should be provided in familiar format without disturbance in processing. Also,

inconvenient or limiting component for use by users should be eliminated.

Compliance is defined complying with standard and guideline of nation and institute and administrative standard. Public open data should guarantee matching and connecting firstly complying with standard and guideline of nation. Public data should basically comply with government's data standard and be available for connection and use for government-wide administrative information.

3.3. Credibility

Credibility is a necessary character to protect personal or confidential information kept by public data. Public open data should be protected with security. Details of credibility consist of security and protection.

Table 5. Definition of "Credibility"

Reliability	Definition	References
Security	<ul style="list-style-type: none"> ·Data administration is being managed. ·Data access is being controlled in accordance with Permission. 	Public data Management Guidelines (Security), DQA(Integrity), DQC(Accountability, Traceability), ISO 25012 (Confidentiality)
Protection	<ul style="list-style-type: none"> ·Personal information or confidential data is being encrypted by adequate security measures and appropriate process. ·Being acquired / possessed, used / provided / destroyed by compliance with the Privacy Act. Backed up by a data backup policy. 	DQA(Integrity), Public Data Management Guidelines (Security)

Security is defined as a party of data management is managed and limited so that it could be protected from illegal addition, deletion and modification, and pollution of data is prevented through backup. Management should be done with physical access including hardware, software and network, and limit of use also means management for maintaining data security.

Protection is defined that personal or confidential information are processed and managed with code, and gathering/maintenance of personal information, use/provision/termination should comply with personal information act.

Important issues related with public open data are privacy and security problems. Privacy is deemed as basic human right so it should be dealt with careful manner for its risk of outflow. So improvement of public data, process, management and legal act is needed. Public data is originally set to achieve administrative object managing a number of personal information. The risk of leak leaves implication related with public open data. The original purpose of public open data, fair availability, conflicts with protection of personal information, so expansion of public data protecting personal information not only needs legal act but technological support. So public open data's quality should be individually set for each institute and credibility and availability should be inspected to satisfy right-to-know of citizen.

3.4. Suitability

Suitability is about quality itself to achieve object of availability of public data. Details of suitability consist of completeness, accuracy, validity, timeliness and consistency.

Table 6. Definition of “Suitability”

Suitability	Definition	References
Completeness	·Not missing a required field.	DQC (Necessity), DQA(Objectivity), Public Data Management Guidelines(Completeness), ISO25012(Completeness)
Accuracy	·Accurately reflect the value of real-world objects. ·The data is accurately reflect the value of intended concept or situation. ·The data is precise and identifiable.	ISO 25012(Accuracy, Credibility, Precision), DQC (Authenticity, Relevance), Public Data Management Guidelines(Accuracy), DQA(Objectivity)
Validity	Data entry meets specified data scope, format, and the domain.	DQC (Suitability)
Timeliness	·The data is adequate and available as seasonally. ·Update period is being defined and implemented in accordance with the data characteristics. ·The value of the data reflect the latest value closest to reality.	ISO 25012(Currentness), Public Data Management Guidelines(Timeliness), DQC(Timeliness)
Consistency	·Data having same meanings have consistent name and type in the logical property unit and physical column unit. ·Structure of the data, values, forms of expression are consistently defined and correspond to each other. ·Keep data's consistency when moving or creating data, without data missing, generated error, or redundancy created.	ISO 25012(Consistency), DQC(Integrity, Conformity) Public Data Management Guidelines (Consistency)

Completeness is defined as integrity without omission of necessary information in data. Also whether depth of data for business use is enough is judged.

Accuracy is a character that shows a value of data is consistent with real fact. In other words, it clearly reflects a value of real world without error. Data should be precise in terms of meaning and rule. Consistency of values saved in DB with real world means semantic accuracy, and accuracy of rule confirms whether the value exists in scope of domain. Also data should be identifiable having precise and proper details. If used information is not accurate, it can bring risk and damage. Also inaccuracy of information itself can bring low availability of public data.

Validity is about whether data items satisfy set scope of validity, format and domain.

Timeliness means whether use of data is proper and valid for current circumstance. To satisfy object of use reflecting latest value near to real world, proper update in accordance with data property should be done and defined.

Consistency is defined that inconsistency among data will not happen. In other words, data of same meaning exist with consistency of name and format, and it should be maintained even though its process or transmittance without omission, error and duplication. To evaluate consistency of information, compliance of rule, data format, saving format of data value and description are included.

4. Model Verification

This research proceeded survey targeting 53 manager and executive of public data to verify characteristic of public open data quality. Each surveyed item has semantic differentials from 1 ‘Surely not’ to 7 ‘Surely yes’, making respondents designate relevant

value and they are analyzed and modification to characteristics of public open data quality is applied. As reflecting opinions of professionals managing real public data, it proves that characteristics dealt with in this research are applicable to real work.

Survey was done from May 10, 2015 to May 24, 2015 (for 15 days) distributing it through mail and online (non face-to-face) securing 53 valid responds from 80 respondents. 4 respondents seen as outlier were excluded. 4 cases were excluded because whole questions were answered by one question. SPSS Statistics Ver. 20 was used as analytical tools.

To verify validity, exploratory factor analysis was done in this research. Through factor analysis, each character and additional character was identifiable whether they are classified meaningfully. Through this factor analysis, it can be verified that whether survey is credibly formed to classify factor in accordance with character. Decision of number of factor for factor analysis was used Varimax among orthogonal rotation methods.

The result showed that 17 questions out of 21 that form “Publicity” were classified as 5 factors (accessibility, objectivity, publicness, request-reflectivity, responsiveness) by optimal cut-off value (0.5) shown in Table 7. 17 questions out of 21 that forms “Availability” were classified as 5 factors (understandability, processability, connectivity, easiness of use and compliance) shown in Table 8, 9 questions out of 11 that forms “Credibility” were classified as 2 factors (protection and security) shown in Table 9, 15 questions out of 17 that forms “Suitability” were classified as 5 factors (completeness, accuracy, validity, timeliness and consistency) shown in Table 10.

Also credibility of each factor shows that Cronbach’s Alpha value exceeds 0.5, which means it is reliable [15].

Table 7. Factor Analysis and Reliability of “Publicness”

Factors						
	1	2	3	4	5	Cronbach's α
Accessibility1	0.815	0.377	0.158	0.174	-0.071	0.790
Accessibility2	0.770	-0.145	0.146	0.322	0.164	
Accessibility3	0.693	0.232	0.527	-0.162	0.186	
Objectivity1	0.078	0.832	0.190	0.118	0.317	0.820
Objectivity2	0.095	0.818	0.154	0.118	0.281	
Objectivity3	0.469	0.579	0.304	0.286	-0.073	
Publicness1	0.115	0.093	0.918	0.060	0.102	0.875
Publicness2	0.225	0.201	0.795	0.265	0.070	
Publicness3	0.357	0.202	0.650	0.348	0.260	
Request-reflectivity1	0.263	0.163	0.226	0.792	0.005	0.810
Request-reflectivity2	0.104	0.540	-0.144	0.664	0.166	
Request-reflectivity3	0.190	-0.079	0.471	0.643	0.390	
Request-reflectivity4	0.009	0.218	0.265	0.613	0.459	
Responsiveness1	0.052	0.116	0.054	0.145	0.909	0.934
Responsiveness2	0.042	0.162	0.071	0.157	0.902	
Responsiveness3	0.099	0.109	0.022	0.135	0.855	
Responsiveness4	0.008	0.202	0.267	-0.018	0.850	
Eigen Value	3.818	2.818	2.469	2.387	2.295	
Factor extraction method: principal component analysis. a. On 7 iterations factors have converged.						

Table 8. Factor Analysis and Reliability of “Availability”

Factors						
	1	2	3	4	5	Cronbach's α
Understandability1	0.932	0.182	0.040	0.099	0.091	0.947
Understandability2	0.910	0.174	0.169	0.094	-0.014	
Understandability3	0.907	0.145	0.069	0.054	0.228	
Understandability4	0.828	0.264	0.127	-0.111	0.068	
Processability1	0.480	0.770	0.133	0.233	0.148	0.947
Processability2	0.497	0.757	0.199	0.081	0.200	
Processability3	0.516	0.737	0.166	0.223	0.011	
Connectivity1	0.137	0.211	0.861	0.118	0.318	0.938
Connectivity2	0.181	0.096	0.812	0.187	0.457	
Connectivity3	0.201	0.180	0.739	0.365	0.326	
Easiness of use1	0.269	0.137	0.252	0.750	0.383	0.875
Easiness of use2	-0.129	0.379	0.302	0.658	0.410	
Easiness of use3	-0.065	0.518	0.400	0.560	0.339	
Compliance1	0.122	0.119	0.223	0.193	0.931	0.984
Compliance2	0.109	0.074	0.260	0.163	0.925	
Compliance3	0.081	0.093	0.250	0.205	0.918	
Compliance4	0.098	0.122	0.239	0.156	0.896	
Eigen Value	4.177	2.424	2.636	1.765	4.341	
Factor extraction method: principal component analysis. a. On 8 iterations factors have converged.						

Table 9. Factor Analysis and Reliability of “Credibility”

Factors			
	1	2	Cronbach's α
Security1	0.871	0.423	0.959
Security2	0.868	0.421	
Security3	0.819	0.350	
Security4	0.754	0.560	
Protection1	0.357	0.887	0.953
Protection2	0.421	0.871	
Protection3	0.411	0.863	
Protection4	0.582	0.718	
Protection5	0.468	0.709	
Eigen Value	6.006	6.853	
Factor extraction method: principal component analysis. a. On 10 iterations factors have converged.			

Table 10. Factor Analysis and Reliability of “Suitability”

Factors						
	1	2	3	4	5	Cronbach's α
Completeness1	0.837	0.235	0.242	0.271	0.060	0.926
Completeness2	0.832	0.317	0.304	0.168	0.169	
Completeness3	0.787	0.130	0.340	0.071	0.350	
Accuracy1	0.271	0.857	0.087	0.036	0.290	0.919
Accuracy2	0.162	0.810	0.172	0.207	0.360	
Accuracy3	0.292	0.670	0.246	0.214	0.470	
Validity1	0.288	0.120	0.882	0.110	0.253	0.981
Validity2	0.395	0.209	0.801	0.182	0.268	

Validity3	0.437	0.153	0.781	0.139	0.325	
Timeliness1	0.238	0.051	0.397	0.834	0.108	0.813
Timeliness2	0.233	0.265	-0.158	0.731	0.393	
Timeliness3	0.045	0.131	0.660	0.692	-0.008	
Consistency1	0.228	0.295	0.155	0.202	0.848	0.938
Consistency2	0.140	0.352	0.240	0.135	0.816	
Consistency3	0.151	0.365	0.384	0.067	0.758	
Eigen Value	2.835	2.529	3.242	2.034	2.956	
Factor extraction method: principal component analysis. a. On 10 iterations factors have converged.						

After confirming factor analysis of each additional character is reliable and applicable with survey, second factor analysis was done to confirm whether sub characteristics suggested in the model is classified. Items representing each quality were extracted and factor analysis was done. 5 items of publicity (accessibility, objectivity, publicness, request-reflectivity and responsiveness), 5 items of availability (understandability, processability, connectivity, easiness of use, timeliness and compliance), 2 items of credibility (protection and security), 5 items of suitability (completeness, accuracy, validity, timeliness and consistency) were selected and analyzed whether they are classified as common factor.

The result showed that ‘accessibility’ of existing model’s sub characteristic of “publicness” is near factor to sub characteristics of “availability”. So ‘accessibility’ is modified as sub character of the “availability”. The result of factor analysis is shown in Table 11.

Table 2. Factor Analysis and Reliability of Representative Items of Sub-Characteristics

Characteristics	Sub-Characteristics	1	2	3	4
Publicity	Objectivity1	0.862	-0.053	0.019	0.235
	Publicness1	0.853	-0.020	0.035	0.245
	Request-reflectivity1	0.822	0.228	0.117	0.141
	Responsiveness1	0.811	0.082	0.135	0.285
Availability	Accessibility1*	0.328	0.569	0.122	0.546
	Understandability1	0.037	0.867	0.035	0.203
	Processability1	0.032	0.866	0.087	0.198
	Connectivity1	-0.005	0.818	0.248	0.288
	Easiness of use1	0.050	0.817	0.176	0.143
	Compliance1	0.057	0.786	0.061	0.246
Credibility	Security1	0.261	0.164	0.833	0.051
	Protection1	0.259	0.192	0.724	0.017
Suitability	Completeness1	0.356	0.347	0.099	0.820
	Accuracy1	0.194	0.416	0.284	0.745
	Validity1	0.113	0.293	0.466	0.693
	Timeliness1	0.079	0.336	0.525	0.691
	Consistency1	0.286	0.384	0.378	0.668
Eigen Value		6.998	6.307	2.828	3.076

T Verification on each value of responds using modified model through factor analysis was done. A level of significance was 95% and value of verification was set 5. The result was adopted for value of 2/3 of responds did not exceed 0.05 excluding responsiveness of publicness and timeliness of integrity. Also the ‘responsiveness’ and ‘timeliness’ that average of value of responds shows fewer than 4 were excluded judging that they are not

meaningfully managed and administered in current institutes.

Figure 2 shows the modified public open data quality characters reflecting the result of survey, and it has 4 main characters (publicity, availability, credibility and suitability) and 15 sub characteristics. Sub characteristics of publicity are 3 things: objectivity, publicness, request-reflectivity; sub characteristics of availability are 6 things: accessibility, understandability, processability, connectivity, easiness of use, compliance; sub characteristics of credibility are 2 things: security and protection; sub characteristics of suitability are 4 things: completeness, accuracy, validity, and consistency.

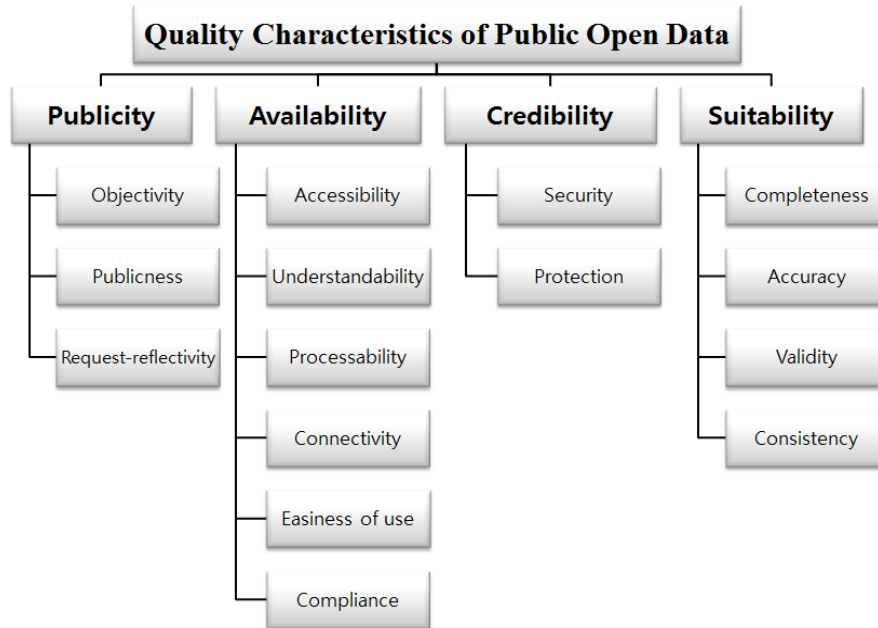


Figure 2. Quality Characteristics of Public Open Data

5. Conclusion and Tasks for Future

Opening public data, its importance and necessity being emphasized as world trend, is not actively used compared to expectation. Problems related with quality of opened public data arise as this factor.

This research suggested necessary characters for public open data referencing existing analysis and international standard. Also suitability of the research was verified from survey on manager and administrator of public data.

The result shows that necessary characters for public open data are 4 main character including 15 sub characteristics: publicity (objectivity, publicness, and request-reflectivity), availability (accessibility, understandability, processability, connectivity, easiness of use, and compliance), credibility (security and protection), and suitability (completeness, accuracy, validity and consistency).

Public data has much valuable information and value for administrative work, but problems from dealing with personal information and opening data that need security lowers open-available public data and this is becoming disturbance of use of data.

To develop public open data that have both values and risks, each data-owning institute should establish standard of data quality character and provide valuable data and prevent risks with preemptive act, which are pointed in this research through characters shown in public data quality.

This research has significance from suggestion that characters necessary for public data were insufficient in existing research. However, it has limits from the point that it is not

verified by applying it to public data available to public, and to solve this, empirical study and research applicable to real public data should be a task in future.

References

- [1] National Information Society Agency, "National Informatization White Paper", (2014).
- [2] Ministry of the Interior, "Completion report for diagnosis and improvement of public data quality", (2014).
- [3] L. P. English, "Improving Data Warehouse and Business Information Quality: Methods for Reducing Costs and Increasing Profits", John Wiley & Sons, (1999).
- [4] R. Y. Wang and D. M. Strong, "Beyond accuracy: what data quality means to data consumers", *Journal of Management Information Systems*, vol. 12, no. 4, (1996), pp. 5-34.
- [5] L. P. English, "Improving Data Warehouse and Business Information Quality", New York: Wiley, (1999).
- [6] T. C. Redman, "Data Quality for the Information Age", Norwood, MA: Artech House, (1996).
- [7] R. Y. Wang, "Data quality", Springer Science & Business Media, vol. 23, (2001).
- [8] Korea Database Agency, "Data Quality Management Maturity Model, ver. 1.0", (2006).
- [9] ISO/IEC 25012:2008, "Software engineering - Software product Quality Requirements and Evaluation (SQuaRE) - Data quality model," (2006).
- [10] Ministry of the Interior, "Act On Promotion of the Provision and Use Of Public Data", Article 2 (Definitions), (2013).
- [11] Ministry of the Interior, "Administrative guideline of public data", (2014).
- [12] National Information Society Agency, "Public Information Quality Management Manual", (2012).
- [13] US Office OF Management and Budget, "Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by Federal Agencies", (2001).
- [14] D. Tim, "Open data barometer: 2013 global report", World Wide Web Foundation and Open Data Institute, (2013).
- [15] H. Y. Lee, "Data analysis using the SPSS", Cheong Ram, (2016).
- [16] S. H. Jung and D. H. Jeong, "A Study on the Influence Factors in Data Quality of Public Organizations", *Journal of Korea information processing society*, vol. 3, no. 4, (2013), pp. 251-266.
- [17] A. Parasuraman, V. A. Zeithaml and L. L. Berry, "SERVQUAL: A Multiple-item Scale for Measuring Consumer Perceptions of Service Quality", *Journal of Retailing*, (1998), pp. 15-29.
- [18] Y. H. Woo and M. Y. Hong, "A Feasibility Study on quality measurement of public service", *Journal of Korean Public Administration Quarterly*, vol. 16, no. 3, (2004).

