

# 공공개방데이터 품질 특성에 관한 연구

박고은, 김창재  
송실대학교 소프트웨어특성화대학원

## Quality Characteristics of Public Open Data

Go-Eun Park, Chang-Jae Kim

The Graduate School of Software, Soongsil University

**요 약** 공공데이터 개방은 민간을 포함한 누구나 공공데이터를 자유롭게 재이용하여 국민 삶의 질을 높이고 신(新)산업, 일자리 창출로 창조 경제 활성화에 기여하고자 하는 목표를 가진다. 공공데이터 개방은 전 세계적으로 중요성이 강조되고 있는 정책이며, 개방의 성공 사례들이 만들어지고 있다. 공공개방데이터는 공공의 목적을 달성하기 위해 이에 적합한 품질을 갖추어야 한다. 그러나 공공데이터 품질 관리와 표준화의 미흡으로 인한 오류데이터 발견 및 활용성 저하 문제가 제기되며, 품질에 관한 가이드라인이 미흡하다. 이에 본 연구에서는 기존의 데이터 품질과 공공데이터 품질, 공공 서비스 품질에 관한 복합적 시각을 적용한 연구를 통해 개방 공공데이터가 갖춰야 할 품질 특성에 대해 도출하고 전문가 설문을 통해 모델을 수정 및 검증하였으며, 그 결과 공공개방데이터의 품질 특성으로 공공성, 활용성, 신뢰성, 적합성을 도출하였다. 공공개방데이터의 품질 향상과 활용 활성화를 위해 갖춰야 할 품질 특성을 제시함에 본 연구의 의의가 있다.

**주제어** : 공공데이터, 공공개방데이터, 데이터 품질, 데이터 특성, 공공개방데이터 품질, 품질 복합적 특성

**Abstract** Public data open is one of the important tasks of Korea Government 3.0. By making open data available to the private sector, the goal is to create jobs, increase innovation and improve quality of life. Public data open is a policy that emphasized its importance worldwide. Open data should have adequate quality in order to achieve the object of the public. However, there are open data's quality problems due to the lack of data quality management and standardization. The purpose of this study is to derive data characteristics of public open data from existing researches. In addition, the model was modified and verified through a survey targeting the experts on public open data. The study indicates that public open data's quality characteristics as publicity, usability, reliability, suitability. This study is significant in that it suggests quality characteristics to improve the data quality and promote utilization of the open data.

**Key Words** : public data, public open data, data quality, data characteristics, open data quality, quality characteristics complex

### 1. 서론

공공데이터 개방은 민간을 포함한 누구나 공공데이터

를 자유롭게 재이용하게 하여, 이를 통해 국민 삶의 질을 높이고 신(新)산업, 일자리 창출로 창조 경제 활성화에 기여하고자 하는 목표를 가진다[1].

Received 12 August 2015, Revised 21 September 2015  
Accepted 20 October 2015  
Corresponding Author: Changjae Kim  
(The Graduate School of Software, Soongsil University)  
Email: winchang@ssu.ac.kr  
ISSN: 1738-1916

© The Society of Digital Policy & Management. All rights reserved. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

공공데이터 개방은 전 세계적으로 중요성이 강조되고 있는 정책이며, 미국을 비롯한 EU 회원국들도 적극적인 정책을 펼치며 공공데이터 개방의 성공 사례들을 만들어 내고 있다.

‘15년 9월 기준 공공데이터 포털(data.go.kr)에 개방된 데이터는 파일데이터(12,680), 오픈API(1,796), 데이터 시각화(251), 게시물(3,600)로 전체 18,327건이다. 데이터 분야는 교통물류(112건), 환경기상(90건), 산업고용(70건), 보건의료(54건)등이다.

대한민국 정부는 공공데이터 품질관리를 위해 행정자치부(구 행정안전부)에서 『공공기관의 데이터베이스 품질관리지침』(2011.7)을 통해 각 기관들의 데이터 품질관리 수행을 위한 제도적 근거와 최소 준수 사항들을 우선적으로 규정하였고, 한국정보화진흥원(NIA)에서 『공공정보 품질관리 매뉴얼』(2012.1)으로 공공데이터 품질관리와 진단 및 개선에 대한 해설을 제공하였다. 한국정보화진흥원의 공공정보 품질관리지원센터는 국가 중요 데이터베이스의 품질 개선, 포털 운영, 교육과 홍보, 지침 개발 등 전문기술 지원을 수행하고 있는 등 정부의 공공데이터 개방을 위한 노력이 활발히 이루어지고 있다.

그러나 공공데이터 활용은 기대만큼 활발히 이루어지고 있지 않으며, 데이터 활용 성과들은 기대에 미치지 못하는 상황이다. 그 주요한 원인으로 제기되는 것으로 공공데이터의 품질 관리와 표준화의 미흡으로 인한 오류데이터 발견 및 활용성 저하 문제를 들 수 있다[2].

더욱이 각 기관의 개방 데이터 품질 관리 방향 정립에 도움이 되는 공공개방데이터 품질에 관한 가이드라인이 부재하다. 공공데이터 개방이 공급자 중심의 개방 계획에 따라 추진되고 있어 그 활용도가 높지 않다는 문제 또한 공공데이터 관련 연구에서 제기되고 있는 실정에 품질 가이드라인의 부재는 문제점으로 지적된다.

본 연구의 목적은 공공개방데이터가 갖춰야 할 품질 특성 및 부특성의 도출을 목적으로 한다. 기존 연구에서 논의가 부족했던 공공개방데이터의 품질 향상과 활용 촉진을 위한 방향으로써의 품질 특성을 제시 한다는 것에 본 연구의 의의가 있다.

본 연구에서는 공공개방데이터 품질특성을 도출하기 위해 먼저 데이터 품질, 공공데이터 품질, 공공 서비스 품질에 관해 사전 연구를 실시하였다. 또한 해외의 공공데이터 개방 사례를 조사하여, 그 중 미국의 공공개방데이

터 품질에 관한 정의 및 기준에 대해 조사하였다.

다음으로, 공공 서비스 품질과 사전연구 및 사례에서의 데이터 품질 특성을 통해 공공개방데이터 품질 특성 및 부특성을 도출해 냈다.

본 연구에서 도출한 품질 특성을 검증하기 위해서 공공기관에서 현재 공공데이터를 관리하는 관리자 및 실무자를 대상으로 설문을 실시하였다. 설문을 통해 본 연구의 품질특성들이 현 공공기관에서의 데이터 품질 관리 기준과 부합하는지 분석하고, 설문 결과를 반영하여 모델을 정제 및 수정하고, 5장에서 결론 및 추후 연구에 대해 기술하는 것으로 본 논문을 구성한다.

## 2. 관련 연구

### 2.1 데이터 품질과 품질 특성

본 장에서는 그 동안 연구된 데이터 품질에 관한 정의를 살펴보았다. 데이터 품질에 관해 English는 데이터를 활용하는 사용자의 다양한 활용 목적과 만족도를 충족시킬 수 있는 수준으로 정의 하였다. Wang et al.은 사용자가 사용하기에 적합한지(fit for use) 중심으로 정의하였다[3,4]. Larry는 데이터 품질은 사용자 편의에 부합하는 수준으로 사용자와 이해 관계자 만족을 증대시키는 것을 목표로 한다고 하였고, Redman은 데이터 사용자의 요구사항 충족 정도로 정의하였다[5,6].

이를 종합하여 본 논문에서는 데이터 품질을 ‘사용자의 활용 목적과 요구사항을 충족시키기 위한 데이터의 수준’이라 정의한다.

데이터 품질은 품질지표의 모음에 의해 정의될 수 있으며 데이터 품질지표는 대다수의 정보 수요자가 지속적으로 요구하는 데이터 품질 속성들의 집합이다[7].

데이터 품질 특성은 데이터 품질 속성 또는 품질 기준, 품질 요소라고도 혼용되어지나, 국제표준인 ISO/IEC 25000을 비롯한 품질 모델에서는 품질관련 속성을 품질 특성이라고 명명함에 따라, 본 연구에서도 품질 특성으로 정의한다.

한국 데이터베이스 진흥원은 DQC(Data Quality Certification)[7]에서 데이터 품질 기준으로 유효성과 활용성을 나누어 각 세부 기준을 정의하였으며, 공공 데이터를 정형데이터와 비정형 데이터로 구분하여 정형데이

터의 품질 기준을 완전성, 유일성, 유효성, 일관성, 정확성, 기능성으로, 비정형 데이터의 품질 특성을 기능성, 신뢰성, 사용성, 효율성, 이식성으로 제시한 바 있다[8].

소프트웨어 품질평가 모델인 ISO/IEC 25000의 5개의 모델 구조 중 품질모델(Quality Model Division)인 25012는 컴퓨터 시스템에 소프트웨어와 함께 구동되는 데이터 품질평가의 모델을 가지고 있다. 여기서 제시한 데이터 속성은 15가지로 정확성, 완전성, 일관성, 신뢰성, 현재성, 접근성, 준수성, 기밀성, 효율성, 정밀성, 추적성, 이해성, 가용성, 이식성, 복구성이다[9].

## 2.2 공공데이터 품질

“공공데이터”란 데이터베이스, 전자화 된 파일 등 공공기관이 법령 등에서 정하는 목적을 위하여 생성 또는 취득하여 관리하고 있는 광(光) 또는 전자적 방식으로 처리된 자료 또는 정보를 말한다[10].

행정자치부의 공공데이터 관리지침에서의 데이터 품질이란 데이터의 최신성, 정확성, 상호연계성 등을 확보하여 사용자에게 유용한 가치를 줄 수 있는 수준이며, 공공정보 품질관리 매뉴얼(Ver2.1)은 공공기관의 데이터 품질을 측정하는 기준인 8대 지표로 준비성, 완전성, 일관성, 정확성, 보안성, 적시성, 유용성과 지표별 세부 특성을 반영한 24개의 세부 지표를 제시한다[11,12].

공공데이터의 품질에 관한 초기 연구들도 대부분 정확성만을 보는 미시적인 관점을 갖추고 있다. 그러나 실사용에서의 문제나 활용에서의 영향은 정확성만으로는 품질을 보증하기가 어렵다. 더욱이 공공개방데이터는 민간에 개방되어 두루 다양한 사용자들이 다양한 요구를 가지고 활용되어지는 특성을 가지고 있다.

해외의 공공데이터 개방 사례를 살펴보면, 미국 예산관리국(Office of Management and Budget)이 2000년 12월 데이터 품질법(Data Quality Act, 이하 DQA)을 발효하고 연방정부 기관들에게 정보의 품질, 객관성, 유용성, 통합성을 보장하고 최대화하도록 정책과 절차 가이드라인을 제시하여, 연방기관과 주정부 등이 이를 준수하여 각 기관에 적합한 품질관리체계를 마련하는 것이 의무화되었다. 이에 관련한 품질요소는<Table 1>과 같다[13].

본 지표는 일반적인 품질 지표에 비해 개념과 기준의 정의가 상위 수준에 머물러 있어, 각 기관의 업무와 정보화 환경에 맞추어 적합한 세부지표를 개발·적용할 수 있

도록 되어있다.

한편 WWW(World Wide Web) 재단 의 팀버너스리(Tim Berners Lee)의 ‘Five Stars of Linked Open Data’ 모델은(이하 ‘5 Stars 모델’) 공공데이터 개방에서 중요한 점을 시사한다[14].

5 Stars 모델은 오픈데이터의 다섯 단계 발전방향을 별1개에서부터 별 5개까지 5단계로 제안하였다. 이 모델은 사용자의 데이터 활용 및 연관지식의 획득 효과가 높아질 수 있도록 데이터의 형식에 기반을 둔다. 별 5 단계는 관련 데이터를 찾을 수 있으며 데이터 스키마에 대해 직접 배울 수 있어 연관지식 획득 및 활용효과가 극대화되는 단계이다. 이 모델을 통해 공공데이터의 개방 품질 특성 중 자유로운 재가공성과 연계의 중요성에 대한 시사점을 얻을 수 있다.

<Table 1> Data Quality on Data Quality Act of USA

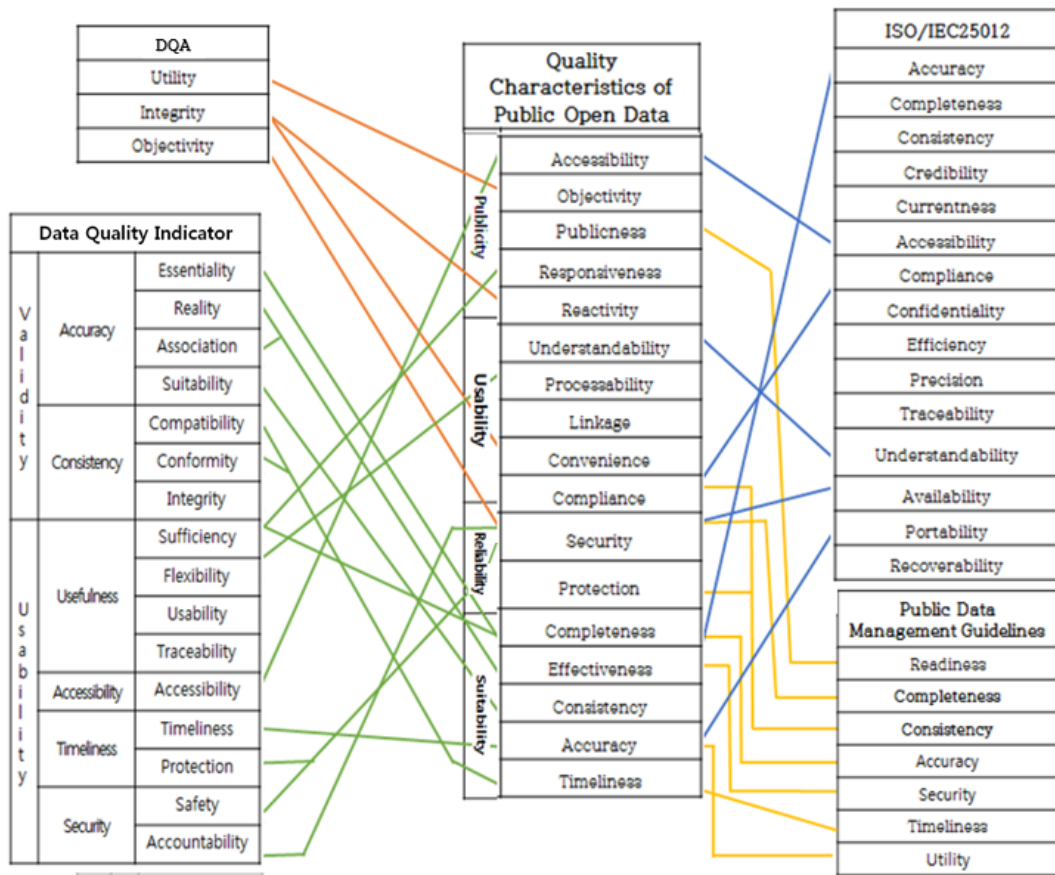
Quality Characteristics	Definitions
Utility	Usefulness of the information to its intended users
Integrity	Security of information – protection of the information from unauthorized access or revision, to ensure that the information is not compromised through corruption or falsification.
Objectivity	Focuses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

## 3. 공공개방데이터 품질 특성

본 연구는 기존의 공공데이터 특성에 관한 기존 연구와 시사점을 통해 개방 공공데이터 품질 특성을 도출하였다[Fig 1].

참고된 기준은 DQC의 DQI(Data Quality Indicator)(이하 DQI), 행정자치부의 공공데이터 관리지침 상 품질 지표(이하, 공공데이터 관리지침), ISO/IEC 25012의 데이터 품질속성(이하, ISO 25012), 미국의 Data Quality Act의 품질요소(이하, DQA)이다[8,9,11,12].

공공데이터의 개방은 ‘공공데이터를 누구나 쉽게 이용하도록 하는 것’이며, 개방 데이터는 공익을 위하여 제공 대상에 차별 없이 요청하는 대중에게 평등하게 제공되어야 한다[11].



[Fig. 1] Quality characteristics of public open data

이에 본 연구에서는 공공개방데이터는 ‘누구나 자유롭게 접근하고 재사용 하여 새로운 서비스와 공공의 가치를 창출하는 데이터’임에 초점을 맞추어, 공공개방데이터가 갖추어야 할 품질에 관해 공공성, 활용성, 신뢰성, 완전성의 4가지 관점으로 접근하고, 이에 따른 세부 특성들을 도출하였다.

### 3.1 공공성

공공성은 대중에게 제공되는 공공서비스로써 개방 공공데이터가 갖춰야 할 특성이다. 편향되거나 일부의 이익을 대변하지 않고 공익 증진을 목표로 해야 하는 공공개방데이터의 특성을 공공성이라 한다.

공공성의 세부 특성에 대해서는 접근성, 객관성, 공익성, 요구 반영성, 대응성이 있다.

접근성은 제공 대상의 평등성과 데이터 획득의 용이

성에 대한 정도이다. 공공데이터는 명시된 포털을 기반으로 데이터 획득이 용이하게 이루어지고, 제공 대상에 차별이 없이 제공되어야 한다.

객관성은 데이터의 출처가 과학적이고 통계적으로 증명되며 관리되는지, 제3자의 판단이나 의견이 개입되지 않은 원자료(Raw data)의 제공이 우선적으로 이루어지는지에 관한 것이다.

공익성은 특정 대상의 이익을 대변하지 않으며 목적 충족에 충분한 가치를 가져 공공의 이익을 증진시키는가에 대한 것이다.

요구반영성은 사용자의 데이터 수정 및 제공 요구에 관한 반영이 신속히 이루어져 목적을 충족시키는가에 관한 것이며, 대응성은 사용자 수요/만족도/활용현황 조사를 통해 기준이 수립 및 반영되는 정도이다.

〈Table 2〉 Definition of "Publicity"

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

### 3.2 활용성

활용성은 자유로운 가공을 통해 가치를 재창출하기 위해 공공개방데이터가 갖춰야 할 특성이다. 공공데이터의 개방은 투명한 정부와 국민의 알권리를 충족시키는 것을 넘어서 데이터와 ICT 기술의 접목을 통한 가치 재창출에 있으므로 활용성을 갖추는 것이 필요하다.

활용성의 세부 특성에는 이해성, 가공성, 연계성, 이용 편의성, 표준준수성이 있다.

이해성은 다양한 사용자가 읽고 해석하는데 혼란 또는 왜곡의 소지가 없이 인식이 가능한 데이터의 특성이다. 같은 값에 관한 해석이 사용자에게 따라 달라지지 않고 동일하게 인식이 가능함을 의미한다. 정보의 표현은 이해하기 쉽게 간결하고 일관된 형식으로 표현되어야 한다.

가공성은 하드웨어나 소프트웨어 등 환경의 제약이 없고 특별한 기술지원이나 설정을 요구하지 않는 데이터의 특성이다. 공공개방 데이터는 사용자가 민간에 걸쳐 넓기 때문에 데이터의 자유로운 활용 저해하는 기술적 제약이 없어야 한다. 데이터의 제공은 기계가독형 포맷

임과 동시에 활용성이 최대한으로 보장되고 프로세싱에 추가적인 과정과 비용이 소모되지 않는 형식으로 제공되는 것이 중요하다. 이러한 포맷의 재활용성을 데이터의 가공성이라 한다. 가공성은 정보를 쉽게 응용해서 다른 업무에 쉽게 활용할 수 있는 조작 용이성을 포함한다.

연계성은 환경과 상황이 다른 데이터(문맥, 시스템, 기관, 국가가 다른 데이터)와의 활용을 위해 갖추어야 할 특성이다. 공공개방데이터는 정부 데이터 표준을 준수하고 범정부적 행정정보 연계 및 공동 활용이 용이해야 한다. 같은 값이라도 표현 형식이 다르게 저장되어 있는 경우(예: 날짜 1956/07/27을 56.07.27 또는 27.07.1956으로 저장 한 경우) 데이터의 활용이 어렵거나 추가 변환과정을 거쳐야 하는 번거로움이 발생한다. 이러한 문제를 해결하기 위해 통일된 용어 또는 표준 표기 방식을 지킴으로써 데이터 간 연계성을 보장해야 한다. 또한 기관 내 중복 데이터나 외부와의 연결 데이터는 일괄적인 관리 기준을 마련하여 데이터 간 연계성이 보장되어야 한다.

이용편의성은 사용자 친숙성 및 이용 만족도를 보장하는 특성이며, 표준 준수성은 국가 차원 및 기관 차원의 표준 및 지침과 행정표준용어 표준을 준수함을 뜻한다.

〈Table 3〉 Definition of "Usability"

Usability	Definition	References
	Adequate for value creation through processing or re-using the open data.	
Understandability	<ul style="list-style-type: none"> <li>Understandable for various users without confusion or distortion.</li> </ul>	ISO 25012 (Understandability)
Processability	<ul style="list-style-type: none"> <li>No limitations for data processing, such as hardware or software limitations.</li> <li>No need for special support or configurations for using data.</li> </ul>	DQI(Flexibility), ISO 25012 (Accessibility)
Linkage	<ul style="list-style-type: none"> <li>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.)</li> <li>Redundant data inside the institution or associated data have a consistent management standard.</li> </ul>	DQI (Flexibility)
Convenience	<ul style="list-style-type: none"> <li>Data can be provided for users with a familiar format.</li> <li>User satisfaction is high.</li> </ul>	DQA (Utility)
Compliance	<ul style="list-style-type: none"> <li>Compliance with the administrative standard terms and codes.</li> <li>Easy for Integrated government data linkage and common use of administrative information</li> <li>Compliance with government data standards and institution data standards.</li> </ul>	Public Data Management Guidelines (Readiness), ISO25012 (Compliance)

### 3.3 신뢰성

신뢰성은 공공데이터가 가진 개인정보나 기밀사항을 보호하기 위해 개방 공공데이터가 갖춰야 할 특성이다. 공공개방데이터는 기술적 보안으로 보호되어지며, 기밀사항이 적절한 처리로 보호되어야 한다. 신뢰성에 포함되는 세부 특성으로는 보안성, 보호성, 공유성이 있다.

〈Table 4〉 Definition of "Reliability"

	Definition	
Reliability	Data is protected by technical security and treated with adequate confidential protection.	References
Security	<ul style="list-style-type: none"><li>• Data administration is being managed.</li><li>• Data access is being controlled in accordance with Permission.</li></ul>	Public Data Management Guidelines (Security), DQA(Integrity), DQI(Accountability), ISO 25012 (Confidentiality)
Protection	<ul style="list-style-type: none"><li>• Personal information or confidential data is being encrypted by adequate security measures and appropriate process.</li><li>• Being acquired / possessed, used / provided / destroyed by compliance with the Privacy Act.</li><li>• Backed up by a data backup policy.</li></ul>	DQA(Integrity), DQI(Protection, Safety), ISO 25012 (Confidentiality), Public Data Management Guidelines (Security)
Share	<ul style="list-style-type: none"><li>• Prior action is prepared to avoid future copyright disputes.</li></ul>	Public Data Management Guidelines (Readiness)

보안성은 데이터 관리 주체가 관리되고 접근이 통제되어 불법적인 추가, 삭제, 수정으로부터 보호되고, 백업을 통해 데이터의 오염을 방지하고 안전하게 관리됨을 뜻한다. 관리는 하드웨어, 소프트웨어, 네트워크를 포함하여 물리적 접근 통제가 적절히 이루어지며, 사용 통제 또한 데이터의 보안을 유지하기 위한 적절하고 일관된 규정에 의해 관리되는지를 의미한다.

보호성은 개인정보 또는 기밀 정보가 암호화 및 적절한 처리로 보안 조치되고 개인정보의 수집/보유, 이용/제공/파기에는 개인정보 보호법을 준수해야 하는 것을 의미한다. 또한 적절한 백업정책에 의해 데이터는 소실 및 훼손으로부터 자유로워야 한다.

공공데이터의 특성 상 본래 행정적 목적을 달성하기

위해 다량의 개인정보가 취득되어 관리되고 있다. 이러한 정보의 유출 위험성은 공공데이터 개방과 관련하여 시사점을 남긴다.

공공데이터 개방의 본래의 목적인 정보의 평등한 활용과 개인정보의 보호는 충돌하는 문제이기 때문에 이를 인지하고, 개인정보를 보호하면서 활용 가능한 공공데이터를 확대시키는 것은 법·제도적 측면뿐만 아니라 기술적·관리적 보완을 통한 해결 사항이다.

이에 각 기관별로 업무 규칙에 맞게 공공개방데이터 품질 특성을 설정하고 신뢰성과 활용성의 점검을 통해 국민의 권리를 충족시키고 지키는 것은 공공개방데이터의 필수적 관리 과제이다.

공유성은 민간이 이용함에 있어 데이터의 저작권에 관한 선 처리가 이루어지고 명확히 제시되어, 활용 단계에서는 저작권 분쟁의 우려가 없이 활용이 가능함을 의미한다.

### 3.4 적합성

적합성은 공공데이터가 활용목적을 이루기 위해 값 자체가 갖춰야 할 품질에 관한 내용이다. 완전성의 세부 특성은 완전성, 정확성, 유효성, 적시성, 일관성이 있다.

완전성은 필수 항목에 누락이 없이 완전한 데이터의 특성이다. 또한 목적하는 업무에 활용되기 위해 데이터가 충분히 깊고, 넓고, 범위를 갖추었는지에 대한 특성이 다.

정확성은 데이터의 값이 실제 사실과 일치하는가에 대한 특성이다. 즉, 데이터가 실세계의 값을 수집과 입력 과정에서 오류 없이 정확히 반영함을 의미한다. 데이터는 의미상, 규칙상의 두 측면에서 정확해야 한다. DB에 저장된 값이 실세계의 값과 일치하는 것이 의미적인 정확성(Semantic Accuracy)이며 규칙상의 정확성(Semantic Accuracy)은 해당하는 도메인 범위에 값이 있는지 여부를 보는 것이다. 또한 데이터는 적정 수준의 상세화를 통해 정밀함을 갖추어 여러 다른 데이터와 엄밀한 식별이 가능해야 한다.

유효성은 데이터 항목이 정해진 유효범위, 형식, 도메인을 충족하는가에 대한 것이다.

적시성은 데이터의 나이가 활용 목적에 적합하며 시기적으로 유효하고 적절함을 뜻한다. 데이터 값이 현실에 가까운 최신 값을 반영하여 활용목적을 충족시키기

위해서 데이터 성격에 따른 적정 업데이트 시기가 정해지고 이행되어야 한다.

일관성은 동일한 데이터 간에 불일치가 발생하지 않는 것이다. 즉, 같은 의미의 데이터는 일관된 이름과 형식을 가지며, 데이터 생성과 가공으로 인한 이동에도 누락이나 오류·중복 생성이 없이 일관성을 유지해야 한다.

정보의 일관성의 평가를 위해서 명명규칙의 준수, 데이터 형식, 데이터 값의 저장 형식, 표현 형식을 측정 범위에 포함시킨다.

<Table 5> Definition of "Suitability"

Suitability	Definition		References
	The data value is complete, appropriate and accurate for application.		
Completeness	• Not missing a required field.		DQI (Essentiality, Sufficiency), Public Data Management Guidelines (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), DQI (Reality, Association), Public Data Management Guidelines (Accuracy)
Effectiveness	• Data entry meets specified data scope, format, and the domain.		DQI (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), DQI (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), DQI (Compatibility, Conformity) Public Data Management Guidelines (Consistency)

#### 4. 모델 검증

본 논문에서 제안한 개방 공공데이터 품질 특성을 검증하기 위하여 53명의 공공데이터 관리자 및 실무자를 대상으로 설문을 진행하였다.

각 설문항목은 '전혀 그렇지 않다'를 1로 '매우 그렇다'를 7로 하는 7단계의 의미차별척도 (Semantic Differentials)로, 응답자가 속한 기관의 데이터 관리 사항에 해당하는 수치에 대해 표시하도록 하였고, 응답자의 답변을 분석하여 앞서 제안한 공공개방데이터 품질특성에 수정을 가하였다.

이는 제시된 품질 특성이 현장에서 실제로 얼마나 중요하게 인식되고 관리되어지는지에 대해 실제 공공데이터를 관리하고 있는 전문가들의 의견을 반영함으로써, 본 논문에서 제시하는 데이터 품질 특성은 실제로 현장에서 유효하게 활용이 가능한 속성임을 목표로 두고 연구되었다는 점을 검증하는 단계이다.

설문은 2015년 5월 10일부터 2015년 5월 24일까지 15일간 실시하였으며, 설문지 배포는 메일 등 온라인을 이용한 비대면 설문 방식으로 80명에게 배포하여 53건의 유효 응답을 확보하였다. 이 중에서 아웃라이어(outlier)로 판단되는 4건을 배제하였다. 배제된 4건은 전체 답변을 한 번으로 일관되게 기입한 응답이다. 분석 도구는 SPSS Statics Ver. 20을 사용하였다.

본 연구에서는 타당성(Validity) 검증을 위해 탐색적 요인분석을 실시하였다 요인분석을 통해 각 특성 및 부특성을 구성하는 문항끼리 서로 의미 있게 구분되어지는지를 알 수 있다. 본 연구에서 요인분석을 위한 요인 수의 결정은 직각회전 방법 중 하나인 베리맥스(Varimax) 회전방식을 사용하였다.

분석 결과 공공성을 구성하는 21개의 문항 중 17개의 값이 요인 적재량의 절단값(0.5)에 의하여 5개의 요인(접근성, 객관성, 공식성, 요구반응성, 대응성)으로 분류됨을 확인하였다<Table 6>. 활용성의 21개 문항 중 17개의 문항이 5개의 요인(이해성, 가공성, 연계성, 이용편의성, 표준준수성)<Table 7>으로, 신뢰성의 11개 문항 중 9개의 문항이 2개의 요인(보호성, 보안성)<Table 8>, 적합성의 17개 문항이 15개 문항을 통해 5개의 요인(완전성, 정확성, 유효성, 적시성, 일관성)<Table 9> 값을 가짐을 확인하였다.



또한 각 요인의 신뢰도가 크론바흐 알파(Cronbach's Alpha) 값이 0.5를 상회하는 것으로 신뢰성이 있다고 볼 수 있다[15].

〈Table 6〉 Factor analysis and reliability of "Publicness"

Factors						Cronbach's α
	1	2	3	4	5	
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	
Responsiveness1	0.263	0.163	0.226	0.792	0.005	0.810
Responsiveness2	0.104	0.540	-0.144	0.664	0.166	
Responsiveness3	0.190	-0.079	0.471	0.643	0.390	
Responsiveness4	0.009	0.218	0.265	0.613	0.459	0.934
Reactivity1	0.052	0.116	0.054	0.145	0.909	
Reactivity2	0.042	0.162	0.071	0.157	0.902	
Reactivity3	0.099	0.109	0.022	0.135	0.855	0.934
Reactivity4	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 7〉 Factor analysis and reliability of "Usability"

Factors						Cronbach's α
	1	2	3	4	5	
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	0.947
Processability1	0.480	0.770	0.133	0.233	0.148	
Processability2	0.497	0.757	0.199	0.081	0.200	
Processability3	0.516	0.737	0.166	0.223	0.011	0.938
Linkage1	0.137	0.211	0.861	0.118	0.318	
Linkage2	0.181	0.096	0.812	0.187	0.457	
Linkage3	0.201	0.180	0.739	0.365	0.326	0.875
Convenience1	0.269	0.137	0.252	0.750	0.383	
Convenience2	-0.129	0.379	0.302	0.658	0.410	
Convenience3	-0.065	0.518	0.400	0.560	0.339	0.984
Compliance1	0.122	0.119	0.223	0.193	0.931	
Compliance2	0.109	0.074	0.260	0.163	0.925	
Compliance3	0.081	0.093	0.250	0.205	0.918	0.984
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 8〉 Factor analysis and reliability of "Reliability"

Factors			Cronbach's α
	1	2	
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	0.953
Eigen Value	6.006	6.853	

Factor extraction method: principal component analysis.  
a. On 3 iterations factors have converged.

〈Table 9〉 Factor analysis and reliability of "Suitability"

Factors						Cronbach's α
	1	2	3	4	5	
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	
Effectiveness1	0.288	0.120	0.882	0.110	0.253	0.981
Effectiveness2	0.395	0.209	0.801	0.182	0.268	
Effectiveness3	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.

각 부특성들의 요인분석을 실시한 후 부특성을 구성하는 요인들이 하나의 주특성 아래 요인들로 분류되는지를 확인하기 위한 2차 요인 분석을 실시하였다.

이를 위해 각 품질특성을 대표하는 항목을 추출하여 요인분석을 실시하였다. 공공성의 5개의 항목(접근성, 객관성, 공익성, 요구반영성, 대응성), 활용성의 5개 항목(이해성, 가공성, 연계성, 이용편의성, 표준준수성), 신뢰성의 2개의 항목(보호성, 보안성), 적합성의 5개의 항목(완



전성, 정확성, 유효성, 적시성, 일관성)을 대표하는 문항을 각 하나씩 선별하였으며, 이들이 공통된 요인으로 구분되어 지는지 요인 분석하였다.

그 결과 기존의 모델에서 공공성을 구성하는 부특성 중 접근성이 활용성의 요인으로 구분되어졌으므로 기존 모델에 수정을 가했으며, 요인분석의 결과 값은 아래의 표와 같다<Table 10>.

〈Table 10〉 Factor analysis and reliability of representative items of sub-characteristics

Characteristics	Sub-Characteristics	1	2	3	4
Publicness	Objectivity1	0.862	-0.053	0.019	0.235
	Publicness1	0.853	-0.020	0.035	0.245
	Responsiveness1	0.822	0.228	0.117	0.141
	Reactivity1	0.811	0.082	0.135	0.285
Usability	Accessibility 1*	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
	Linkage1	-0.005	0.818	0.248	0.288
	Convenience1	0.050	0.817	0.176	0.143
	Compliance1	0.057	0.786	0.061	0.246
Reliability	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
	Effectiveness1	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

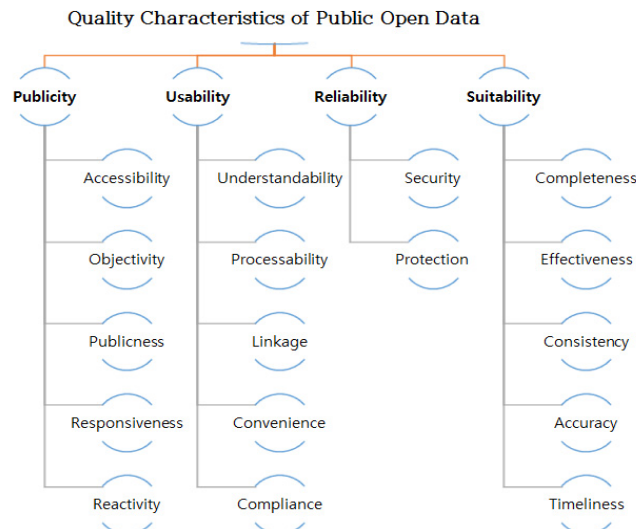
요인분석을 통해 수정된 모델을 활용하여 각 응답 값의 T검정을 실시하였다. 유의수준은 95%, 검정 값은 5로 진행하였다. 분석 결과, 공공성의 대응성과 완전성의 적시성을 제외한 모든 특성의 2/3이 넘는 문항의 응답 값의 유의확률이 0.05를 넘지 않았으므로 채택하였다. 또한 응답 값의 평균이 4에 미달한 대응성과 적시성은 현 공공기관에서 의미 있게 관리 되고 있지 않은 기준이라 판단하고 모델에서 제외하였다.

설문 분석 결과를 반영하여 수정된 개방 공공데이터 품질 특성은 [Fig. 2]와 같으며, 4가지 주특성(공공성, 활용성, 신뢰성, 적합성)과 그에 따른 16가지 부특성을 갖춘다.

공공성의 부특성은 객관성, 공익성, 요구반영성 3가지이며, 활용성의 부특성은 접근성, 이해성, 가공성, 연계성, 이용 편의성, 표준준수성 6개, 신뢰성의 부특성은 보안성과 보호성 2가지이고, 적합성의 부특성은 완전성, 정확성, 유효성, 적시성, 일관성의 5가지로 정의한다.

## 5. 결론 및 향후 과제

세계적 추세이며 그 중요성과 필요성이 강조되어지고 있는 공공데이터 개방은 기대에 비해 활발히 활용되고



[Fig. 2] Modified Quality Characteristics of Public Open Data

〈Table 11〉 Result of One-Sample t-test

One sample test / Values = 5								
Char acter	Sub-Character	t	DOF	P-value	Mean Difference	95% Confidence Interval of the Difference		Adoption/Reje ction
						Lower	Upper	
Publicity	Objectivity	2.948	48	0.005	0.531	0.170	0.890	Adoption
		0.976	48	0.334	0.184	-0.190	0.560	Rejection
		4.753	48	0.000	0.796	0.460	1.130	Adoption
	Publicness	4.129	48	0.000	0.694	0.360	1.030	Adoption
		1.769	48	0.083	0.306	-0.040	0.650	Adoption
		6.149	48	0.000	0.939	0.630	1.250	Adoption
	Responsiveness	0.590	48	0.558	0.122	-0.290	0.540	Rejection
		1.803	48	0.078	0.347	-0.040	0.730	Adoption
		1.667	48	0.092	0.265	-0.050	0.590	Adoption
		2.987	48	0.004	0.531	0.170	0.890	Adoption
	Reactivity	-1.366	48	0.178	-0.388	-0.960	0.180	Rejection
		-2.002	48	0.051	-0.551	-1.100	0.000	Adoption
-0.633		48	0.530	-0.163	-0.680	0.360	Rejection	
-1.674		48	0.101	-0.408	-0.900	0.080	Rejection	
Usability	Accessibility	3.981	48	0.000	0.633	0.310	0.950	Adoption
		0.771	48	0.444	0.184	-0.300	0.660	Rejection
		5.437	48	0.000	0.878	0.550	1.200	Adoption
	Understandability	1.429	48	0.159	0.245	-0.100	0.590	Rejection
		2.603	48	0.012	0.408	0.090	0.720	Adoption
		2.203	48	0.032	0.367	0.030	0.700	Adoption
	Processability	2.808	48	0.007	0.449	0.130	0.770	Adoption
		3.225	48	0.002	0.490	0.180	0.800	Adoption
		3.167	48	0.003	0.490	0.180	0.800	Adoption
	Linkage	1.995	48	0.052	0.347	0.000	0.700	Adoption
		2.423	48	0.019	0.510	0.090	0.930	Adoption
		2.191	48	0.033	0.388	0.030	0.740	Adoption
	Convenience	0.912	48	0.366	0.184	-0.220	0.590	Rejection
		1.720	48	0.092	0.327	-0.060	0.710	Adoption
		3.807	48	0.000	0.510	0.240	0.780	Adoption
	Compliance	0.468	48	0.642	0.082	-0.270	0.430	Rejection
		2.109	48	0.040	0.367	0.020	0.720	Adoption
		2.109	48	0.040	0.367	0.020	0.720	Adoption
1.968		48	0.055	0.347	-0.010	0.700	Adoption	
3.136		48	0.003	0.510	0.180	0.840	Adoption	
Reliability	Security	3.756	48	0.000	0.653	0.300	1.000	Adoption
		5.694	48	0.000	0.898	0.580	1.220	Adoption
		5.838	48	0.000	0.939	0.620	1.260	Adoption
		5.894	48	0.000	0.980	0.650	1.310	Adoption
	Protection	6.125	48	0.000	1.041	0.700	1.380	Adoption
		4.972	48	0.000	0.878	0.520	1.230	Adoption
		5.582	48	0.000	0.918	0.590	1.250	Adoption
		5.367	48	0.000	0.857	0.540	1.180	Adoption
Suitability	Completeness	4.399	48	0.000	0.714	0.390	1.040	Adoption
		0.622	48	0.537	0.122	-0.270	0.520	Rejection
		3.911	48	0.000	0.612	0.300	0.930	Adoption
	Accuracy	2.227	48	0.031	0.347	0.030	0.660	Adoption
		-0.101	48	0.920	-0.020	-0.430	0.390	Rejection
		2.273	48	0.028	0.408	0.050	0.770	Adoption
	Effectiveness	1.741	48	0.088	0.327	-0.050	0.700	Adoption
		1.968	48	0.055	0.327	-0.010	0.660	Adoption
		1.856	48	0.070	0.327	-0.030	0.680	Adoption
	Timeliness	2.085	48	0.042	0.347	0.010	0.680	Adoption
		-0.649	48	0.520	-0.143	-0.590	0.300	Rejection
		-0.876	48	0.385	-0.184	-0.610	0.240	Rejection
	Consistency	0.781	48	0.438	0.163	-0.260	0.580	Rejection
		0.922	48	0.361	0.184	-0.220	0.580	Rejection
1.906		48	0.063	0.367	-0.020	0.750	Adoption	
1.700		48	0.096	0.327	-0.060	0.710	Adoption	

있지 못한 실정이다. 이의 요인으로써 개방되는 공공데이터의 품질에 관한 문제가 제기되고 있다.

이에 본 연구는 기존 연구와 국제 표준에서 제시한 공공데이터의 품질특성을 참조하여 공공개방데이터가 갖추어야 할 품질 특성에 대해 제시하였다. 또한 공공데이터 관리자 설문문을 통해 현장에서 공공데이터 관리에 중점을 두고 있는 사항을 도출하여 본 모델의 적합성을 검증하였다.

그 결과, 공공개방데이터가 갖추어야 할 특성으로 공공성(객관성, 공익성, 요구반영성), 활용성(접근성, 이해성, 가공성, 연계성, 이용편의성, 표준준수성), 신뢰성(보안성, 보호성), 적합성(완전성, 정확성, 유효성, 적시성, 일관성)의 4가지 주특성과 이에 속하는 16개의 부특성을 도출하였다.

공공데이터의 특성 상 행정업무를 위해 수집되어 많은 양질의 정보와 가치를 지니지만, 개인정보 또는 보안이 필요한 데이터의 개방 문제와 그로 인한 파급력에 관한 문제는 개방 가능한 공공데이터의 수를 줄이고 이는 데이터 활용에 걸림돌이 되고 있다.

이렇게 가치와 위험요소를 동시에 지니고 있는 공공데이터의 개방 발전을 위해 각 데이터 보유 기관에서는 데이터 품질 특성 기준을 세우고 세부 특성에 따른 관리 기준을 수립하여 양질의 데이터 제공과 위험요소의 선제적 관리가 필요하다는 것이 본 논문의 공공데이터 품질 특성을 통해 시사된다.

기존의 연구에서 미흡했던 개방 공공데이터가 갖추어야 할 특성에 대해 제시했다는 점에서 본 연구의 의의를 찾을 수 있다. 그러나 실제 개방되는 공공데이터에 적용시켜 검증하지 못한 점이 한계로 남으며, 이 한계점을 극복하기 위해 실제 공공데이터에 적용 가능한 공공데이터 품질 특성 및 품질 관리의 실증적 연구를 본 논문의 향후 과제로 한다.

## REFERENCES

- [1] National Information Society Agency Republic of Korea, "National Informatization White Paper", 2014
- [2] Ministry of Government Administration and Home Affairs, "Completion report for diagnosis and improvement of public data quality", 2014
- [3] English, L. P., "Improving Data Warehouse and Business Information Quality: Methods for Reducing Costs and Increasing Profits", John Wiley & Sons, 1999
- [4] Wang, R. Y. and D. M. Strong, "Beyond accuracy: what data quality means to data consumers," *Journal of Management Information Systems*, Vol.12, No.4, pp.5 - 34, 1996
- [5] Larry P. English, "Improving Data Warehouse and Business Information Quality.", New York : Wiley, 1999
- [6] Redman, T. C, *Data Quality for the Information Age*, Norwood, MA: Artech House, 1996
- [7] Wang, R. Y. et al. "Data quality", Vol. 23. Springer Science & Business Media., 2001
- [8] Korea Database Promotion Center, *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 Security and Public Administration, "Act On Promotion Of The Provision And Use Of Public Data", Article 2 (Definitions), 2013
- [11] Ministry of Government Administration and Home Affairs, "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] Davies, Tim, "Open data barometer: 2013 global report." World Wide Web Foundation and Open Data Institute, 2013
- [15] Hun Young Lee. "Data analysis using the SPSS.", Cheong Ram, 2016
- [16] Seung Ho Jung, 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, pp.251~266, 2013
- [17] Parasuraman, A., Zeithaml, V. A., & Berry, L. L., "SERVQUAL: A Multiple-item Scale for Measuring Consumer Perceptions of Service Quality", Journal of Retailing, pp.15-29, 1998
- [18] Yang Ho Woo, M.Y. Hong, "A Feasibility Study on quality measurement of public service.", Journal of Korean Public Administration Quarterly, Vol.16 No.3, 2004

#### 박 고 은(Park, Go Eun)



- 2014년 2월 ~ 현재 : 숭실대학교 소프트웨어특성화대학원(석사)
- 2014년 7월 ~ 현재 : 한국지역정보개발원 선임
- 관심분야 : SW공학, SW아키텍처, 데이터베이스, 정보 시스템, 데이터 품질, 공공데이터 등
- E-Mail : hnnbmn@ssu.ac.kr

#### 김 창 재(Kim, Chang Jae)



- 2005년 2월 : 숭실대학교 정보과학대학원(석사)
- 2009년 8월 : 숭실대학교 컴퓨터학부(박사)
- 2013년 4월 ~ 현재 : 숭실대학교 소프트웨어특성화대학원 교수
- 관심분야 : SW공학, SW아키텍처, 데이터베이스, 빅 데이터 등
- E-Mail : winchang@ssu.ac.kr