[Research Paper]

한국과 일본의 공공데이터포털 재난안전 데이터셋 및 서비스 비교 분석

이해윤 · 정무현* · 박인선**†

중앙대학교 창의창업연구소 연구원, *한국소방안전원 선임연구원, **중앙대학교 다빈치교양대학 교수

Comparison of Disaster Safety Datasets and Service in Korean and Japanese Public Data Portals

Haeyoon Lee · Muheon Jeong* · Inseon Park**†

Member, Researcher, Creative & Entrepreneurship Institute, Chung-Ang University,

*Senior Researcher, Korea Fire Safety Institut,

**Professor, Da Vinci College of General Education, Chung-Ang University

(Received August 31, 2021; Revised October 27, 2021; Accepted November 3, 2021)

요 약

본 연구의 목적은 한국과 일본의 대표적인 공공데이터포털의 재난안전 데이터셋 및 서비스에 대한 비교 분석을 통해 한국 재난안전 공공데이터포털의 정책적 시사점을 도출하기 위한 것이다. 이를 위해 먼저 비교기준을 정립하고, 재난 유형 및 안전관리단계의 구성요소별 데이터셋 비중 분석, 데이터셋 설명에 대한 텍스트마이닝을 통한 동향 분석, 데이터 품질과 포털 서비스 분석 등을 수행하였다. 연구결과, 재난안전 공공데이터 셋은 한국이 수치적인 측면에서나 비중에서 모두 일본보다 낮았으며, 재난안전관리단계 측면에서 일본은 재난 대비 및 복구 데이터셋 비중이 높은 반면 한국은 예방 데이터셋 비중이 높았다. 또한 재난 대응 협업 측면에서는 한국은 대부분이 물자관리 및 자원 지원이 차지하고 있지만, 일본은 피해시설 응급 복구 및 상황관리 총괄 비중이 높았다. 데이터 품질 측면에서는 일본은 버너스리 평점 4단계 데이터셋이 많지만, 한국은 주로 평점 3단계 비중이 높았으나, 빅데이터 활용을 위한 데이터 형식은 한국이 더 유리한 것을 확인하였다. 포털 서비스는 일본은 자연재난 중심으로 편재되어 있으나, 한국은 사회 재난 중심으로 편재되어 있음을 알 수 있었다. 본 연구결과를 통해 한국의 재난안전 공공데이터 포털의 향후 운영 방향에 참고가 되기를 바란다.

ABSTRACT

The purpose of this study is to obtain implications through comparative analysis of disaster safety datasets and services of representative public data portals in Korea and Japan. Comparative standards were established first. Then, dataset weight analysis of disaster-type and safety-management -stage components, trend analysis through text mining on data-set descriptions, and data quality and portal services analysis were performed. As a result public data sets were lower in Korea both numerically and proportionally than in Japan. Japan had a high proportion of disaster preparation and recovery datasets in terms of disaster safety management, while Korea had a high proportion of prevention data-sets. In addition, in terms of disaster response collaboration, most of Korea has material management and resource support, but Japan has high proportion of emergency recovery and situation management of damaged facilities. In terms of data quality, Japan has many datasets with four levels of Berners-Lee rating. However Korea has a high proportion of datasets with three levels of Beners-Lee rating. However, Korea has a better data format for big-data utilization. Portal services are mainly centered on natural disasters in Japan, but in Korea, they are centered on social disasters. The results of this study provide a reference for the future direction of disaster safety public data portals in Korea.

Keywords : Open data, Disaster safety management, Text mining, Japan, Korea

† Corresponding Author, TEL: +82-2-820-5713, FAX: +82-2-826-6029, E-Mail: isp@cau.ac.kr

© 2021 Korean Institute of Fire Science & Engineering. All right reserved.

1. 서 론

21세기를 들어서면서 인터넷, ICT의 발달과 스마트 전자 기기의 확산과 SNS의 일반화된 활용으로 많은 빅데이터가 생성되고 있으며, 세계 각국은 정부 주도로 데이터 개방을 통한 빅데이터 플랫폼을 구축하고 미래 성장 동력 확보를 위한 노력을 경주하고 있다. 특히 이렇게 생성된 빅데이터는 재난안전관리분야에서도 활발하게 활용되고 있는 실정이다⁽¹⁾.

한국도 2011년 11월 교육과학기술부, 행정안전부, 지식경제부, 방송통신위원회, 국가과학기술위원회 등 5개 기관 합동으로 ‘스마트 국가 구현을 위한 빅데이터 마스터플랜’을 수립하고, 2013년 공공데이터법을 제정하는 등 데이터 활용도 제고를 위한 정책을 강화하고 있고, 3차에 걸친 ‘공공 데이터의 제공 및 이용 활성화 기본계획’과 한국판 뉴딜 사업의 중심인 ‘데이터 댐 사업’ 등을 통해 빅데이터 이용 활성화를 위한 정책을 적극적으로 추진하고 있다.

일본은 빅데이터를 국가경쟁력 강화에 기여할 수 있는 전략적 자원으로 평가하고 2011년부터 총무성을 중심으로 산학연 참여를 통한 빅데이터 추진체계를 재구축하고 그해 5월 ‘빅데이터 활용 기본전략’을 발표하고 「개인정보의 보호에 관한 법률」 개정, 세계 최첨단 ICT 국가창조 선언(빅데이터 이용 촉진), Active Japan ICT 등 3대 빅데이터 정책을 추진하는 등 빅데이터 구축과 활용을 미래 성장동력으로 보고 정부차원의 정책을 적극적으로 강구하고 있다⁽²⁾.

한국은 Open Data Barometer, Global Our Index, OECD OUR Index, 공공데이터 개방 역량에서 지속적으로 10위권 이내를 차지하고 있으며, IMD 세계 디지털 경쟁력 순위에서도 2017년 19위에서 2020년에 8위로 데이터 강국의 반열에 들어선 반면에 일본은 Open Data Barometer, Global Our Index, OECD OUR Index에서 20위 후반을 차지했으며, IMD 세계 디지털 경쟁력 순위에서 2018년 22위, 2019년 23위, 2020년 27위로 점차 하락세로 들어서, 한국이 일본보다 데이터 경쟁력을 갖고 있다고 볼 수 있다^(3,4).

그러나 Jung⁽¹⁾과 Seo⁽⁵⁾에 따르면 이런 평가들은 공공데이터 실증연구에서 상이한 결과가 제시되고 있으며, 한국과 학기과학기술기획평가원의 연구에 따르면 일본은 총무성(IT총괄) 산하 노무라연구소 전담인 재해대책본부시스템의 구축 운영으로, 재난 빅 데이터 솔루션을 중심으로 재난대응 및 예방 능력을 강화하고 있지만 한국 정부는 미비한 실정이라고 평가되고 있다. Shin과 Kim⁽⁶⁾은 국내 재난관리를 위한 효율적인 빅 데이터 활용 방안에 관한 연구는 아직 초기 단계이며, 재난안전 분야와 관련된 공공데이터는 양적, 질적으로 미흡하여 미국, 일본, 영국처럼 재난 중심의 데이터 과학자 양성이 필요하다고 하였다. 이처럼 서로 상이한 연구결과가 나오는 상황에서 본 연구는 한국과 일본의 정부 공공데이터 포털에서 제공하고 있는 재난안전관리 데이터 셋에 대한 양적인 비교 분석과 데이터 품질 및 서비스에

대한 비교 분석을 통하여 한국의 재난안전관리 중심의 빅 데이터 플랫폼의 현 운영 수준을 진단하고, 향후 한국의 재난안전관리 공공데이터 빅데이터 플랫폼의 구축과 활용 방향을 제시하고자 한다.

2. 본 론

2.1 공공데이터 의의

Kim⁽⁷⁾에 따르면 일반적으로 빅데이터는 민간과 공공부문의 전 분야에 걸쳐 생성된 데이터이고, 공공데이터는 공공부문에서 공개된 데이터(open data)를 지칭한다. 공공데이터는 데이터베이스, 전자화된 파일을 공공 기관이 법령 등에서 정하는 목적을 위하여 생성 또는 취득하여 관리하고 있는 광(光) 또는 전자적 방식으로 처리된 자료 또는 정보로서 그 범위와 관련하여 데이터는 크게 오픈데이터, 공공부문 정보, 공공 정부 데이터 등으로 구분할 수 있다.

공공데이터 개방의 목적은 국민의 공공데이터에 대한 이용권을 보장하고, 공공데이터 민간활용을 촉진하는데 목적이 있는데 특히 재난 분야에서의 이해관계자 정보 교환 및 소통을 위하여 필요한 정보로 그 중요성이 강조 되고 있다.

본 연구에서는 대표적인 공공데이터 포털인 한국 정부의 공공데이터 포털(data.go.kr)과 일본 정부의 공공데이터 포털(data.go.jp)의 재난 관련 데이터셋의 정보내용과 데이터 품질, 서비스를 연구 대상으로 하였다.

2.2 공공데이터 재난관리 분야 포털 및 활용에 관한 선행연구

재난 발생에 대한 전조 감지나 재난을 관리하는 영역에서 공공데이터를 활용하는 연구는 아직 초기 단계이다. 특히 선행 연구들을 보면 공공데이터를 활용한 재난관리의 효율성을 높이기 위한 연구보다 빅 데이터 산업의 육성을 위한 연구에 초점이 맞추어져 있다. 또한 빅 데이터 산업 육성에 대한 연구들도 국내뿐만 아니라 외국의 경우도 스마트폰을 활용한 소셜 네트워크 서비스 사용의 보편화로 인해 재난관리에 각종 SNS를 통해 만들어지는 빅 데이터들의 활용 방안에 대한 연구가 진행되고 있는 상황이다.

한편, 공공 데이터 포털의 상황과 활용에 대한 선행 연구들과 국제데이터 평가 기관인 Open Data Barometer, Global Open Data Index, OECD OUR Index를 토대로 공공 데이터 포털의 평가 항목을 비교하면 다음 Table 1과 같다. 여러 데이터 포털 평가항목 중에서 본 연구에서는 데이터 현황 관련해서는 데이터 범주(range), 유형(type)을 데이터 품질에서는 준비성(readiness), 일관성(consistency), 유효성(effectiveness)을 데이터 서비스에서는 사용자 지원(supportive)과 시각화(visualization)를 평가항목으로 활용하였다.

Table 1. Precedent Research Analysis Framework for Public Data Portal

Division		Contents	Kim ⁽⁷⁾	Kim ⁽⁸⁾	Data base Promotion Agency ⁽⁹⁾	Sylvain ⁽¹⁰⁾	Antonio ⁽¹¹⁾	This Study
Data status	Data range	Data availability	O	O	O	O	O	O
		Data range by subject	O					O
		Data range by administrative agency		O	O			
		Data range by standard code			O			
	Data type	Data extension format	O		O	O		O
		Reusable data types			O	O		
		Data standard code			O			
	Data update	New data update status		O		O	O	
		New data update cycle		O				
		Data reuse update status						
Data quality	Readiness	Content sufficiency			O		O	O
	Completeness	Logical model, identifier, physical structure, attribute			O		O	
	Consistency	Attributes, standards, linkage values			O			O
	Accuracy	Input value, business rule, range format, calculation formula			O	O	O	
	Security	Ownership, access restriction, DB protection	O		O			
	Timeliness	Response time, data delivery, latest value			O			
	Effectiveness	Sufficiency, accessibility, utilization						O
Data service	Supportive	User support						O
	Visualization	Data visualization						O

Table 2. Procedures for Diagnosing and Eliciting Public Data Portals

Derivation of diagnostic items	step 1	❖ public data portal and disaster safety sector data definition
	step 2	❖ confirmation of three key elements(safety management standards, data quality, and service) constituting the diagnosis area of the open disaster safety data set
	step 3	❖ the idea of 19 items representing three dimensions
Data analysis in public portal disaster safety sector	step 4	❖ collected data from 27,516 datasets in Japan(1,733 in the disaster safety sector) and 45,073 datasets in Korea(2,325 in the disaster safety sector)
	step 5	❖ basis of data analysis framework for public data portal disaster and safety sector √ among the disaster safety information classification criteria of the Ministry of Public Administration and Security, 5 stages of disaster management, 32 types of disasters, 13 analysis framework for disaster collaboration functions. √ reconstructing 3 areas through experts and panels. √ removal of duplicate data through validation and evaluation based on Berners-Lee star rating √ keyword and network analysis using text mining
	step 6	❖ check 8 items representing 3 dimensions
	step 7	❖ evaluate data service quality and service structure in the disaster safety sector
Policy evaluation and improvement plan	step 8	❖ seeking to improve public data services and data, quality in the disaster and safety sector

2.3 연구 설계

본 연구는 공공데이터 포털에서 제공하고 있는 재난 중심의 공공데이터 셋에 대한 양적인 비교와 데이터 내용 및 품질 분석, 데이터 서비스 분석을 위하여 데이터 진단항목에 대한 아이디어(1~3단계), 통계분석(4~6단계), 데이터 분석(7단계), 정책평가 및 개선계획(8단계) 등 총 8단계로

Table 2와 같이 연구를 진행하였다.

아울러 재난안전 데이터셋의 정보 내용을 분석하기 위하여 데이터의 분류체계 기준⁽⁹⁾을 재난관리단계, 재난 분류 유형, 재난대응 협업기능으로 분류하였다. 이를 나타낸 것이 Table 3이다.

데이터셋의 품질 평가 기준은 데이터 준비성과 일관성,

Table 3. Comparison Criteria for Public Data Sets in the Disaster Safety Sector

Classification			
Level 1	Level 2		
Disaster safety management stage	1.1 prevention	1.2 preparedness	1.3 response
	1.4 recovery	1.5 etc	
Disaster safety type (disaster type ³² , safety type ⁵ , etc1)	natural disaster		social disaster
	2.1 typhoon		2.17 collapse
	2.2 flood		2.18 explosion
	2.3 downpour		2.19 road traffic accident
	2.4 gale		2.20 airline accident
	2.5 storm		2.21 maritime accident
	2.6 tsunami		2.22 chemical fire accident
	2.7 heavy snow		2.23 environmental pollution accident
	2.8 thunderstroke		2.24 energy
	2.9 drought		2.25 communication
	2.10 earthquake		2.26 traffic
	2.11 dust storm		2.27 finance
	2.12 algal bloom		2.28 medical treatment
	2.13 assistant		2.29 water service
	2.14 volcanic activity		2.30 infectious diseases
	2.15 natural space object fall, crash		2.31 livestock disease
	2.16 complex disaster(common)		2.32 fire
	safety		
	2.33 life safety		2.36 crime safety
	2.34 traffic safety		2.37 social infrastructure safety
	2.35 health and safety		2.38 etc
Disaster response cooperation	3.1 general situation management		3.8 emergency life safety support
	3.2 disaster site environmental maintenance		3.9 emergency communication support
	3.3 emergency recovery of disaster facilities		3.10 energy function recovery
	3.4 disaster management promotion		3.11 material management and resource support
	3.5 transportation measures		3.12 medical, quarantine
	3.6 volunteer management		3.13 maintaining social order
	3.7 search, rescue, first aid		3.14 etc

유효성 속성에 초점을 맞추고, 사용자 중심에서 바라본 데이터 형식에 따른 공공데이터셋 품질 개방 지표인 팀 버너스리 별점 평가를 반영, 진행하였다. 그리고 Kim⁽¹²⁾의 ‘웹 기반 데이터 표준화 동향’을 참고하여 공공데이터 품질평가리스트를 재구성한 후, 전문가 패널 조사를 통하여 확정하고, 공공데이터 포털 서비스 중 서비스 지원성과 시각화 속성을 본 연구의 비교분석 기준에 포함하였다. 이상의 데이터 품질평가기준과 서비스 기준을 정리한 것이 Table 4이다.

2.4 연구 결과

2.4.1 재난안전 공공데이터셋 정보내용 분석

Figure 1은 한국과 일본의 재난 안전부문 공공데이터셋 정보내용을 분석하기 위하여 재난안전 데이터셋의 양과 관리단계별 데이터셋 구성내용을 비교한 것이다. 한국은 전체 공공데이터셋 45,012개 중 재난안전부문 데이터셋의 비중은 2,283개(5.07%)로 일본의 전체 데이터셋 27,526개 중 재난 안전부문 데이터셋 3,244개(11.78%) 비중 보다 낮았다. 재난안전관리 단계별 구성내용을 보면, 한국은 예방 1,157개(50.68%), 대응 486개(21.29%), 대비 402개(17.81%), 복

Figure 1. Dataset comparison analysis result based on disaster safety management system.

구 226개(9.90%). 기타 12개(0.53%) 순으로 예방 데이터셋 수가 가장 높은 비중을 차지한 반면 일본은 예방 884개(27.28%), 대응 897개(27.85%), 대비 197개(6.07%), 복구 1,243개(38.32%), 기타 23개(0.71%) 순으로 복구 데이터셋 수가 가장 높은 비중을 차지했다.

Table 4. Comparison Criteria for Quality and Service of Public Datasets in the Disaster Safety Sector

	classification based on machine reading	Japan	Korea	evaluation score
Data quality	OL, open license	GIF, DOC, PDF, PNG	PDF, TEXT, JPG, PNG	1
	OF, open Format	XLS, XLSX	HWP, XLS, XLSX	2
	Re, machine readable	CSV, XML	CSV, JSON, XML	3
	URI, uniform resource identifiers	HTML	RUL	4
	LD, linked data	RDF	RDF	5
Data service	supportability, visualization	Korea public data use cases	Japanese public data use cases	case comparative analysis

Figure 2. Dataset comparison analysis result based on the collaboration of disaster response. **Figure 3.** Dataset comparison analysis result based on disaster safety type.

재난대응 협업에 대한 구성내용을 비교한 결과, 한국은 재난 안전부문 공공데이터셋 2,283개 대비 물자관리 및 자원지원 1,183개(51.82%), 교통대책 262개(11.48%), 사회질서 유지 213개(9.33%), 피해시설 응급 복구 194개(8.50%), 상황관리 총괄 158개(6.92%) 순으로 물자관리 및 자원지원 부문 데이터셋이 전체 51.82%로 가장 높은 비중을 차지하였다.

이에 비해 일본은 전체 재난 안전부문 데이터셋 3,244개 대비 상황관리 총괄 1,121개(34.56%), 피해시설 응급 복구 1,018개(31.38%), 사회질서 유지 325개(10.02%), 재난수습 홍보 171개(5.27%)로 상황관리 총괄 데이터셋이 34.56%로 가장 높은 비중을 차지했다. 이를 나타낸 것이 Figure 2이다.

Figure 3은 재난 및 안전유형에 대한 정보내용을 비교한 그림이다. 한국은 전체 재난 안전부문 데이터셋 2,283개 대비 복합재난(공통) 738개(32.33%), 생활안전 422개(18.48%), 범죄안전 398개(17.21%), 사회재난(도로교통사고) 190개(8.32%), 사회재난(화재) 179개(7.84%) 순으로 복합재난(공통) 데이터셋이 전체 비중의 32.33%로 가장 많은 비중을 차지하였다. 일본은 재난 안전부문 데이터셋 3,244개 대비 자연재난(화산활동) 621개(19.14%), 생활안전 596개(12.21%), 범죄안전 594개(12.15%), 해상사고 270개(8.32%) 순으로 자연재난(화산활동) 데이터셋이 19.14%로 가장 많은 비중을 차지하였다. 즉, 일본은 화산활동, 태풍, 지진과 같은 자연재난에 대한 피해를 복구와 자연재난을 미리 감지하여 예방 및 대응을 통해 자연재난을 관리하고자 노력하고 있다고 볼 수 있다. 또한 생활안전과 범죄안전 부문의 데이터셋 수는 양 국가 간 차이가 없지만, 교통에 대한 데이터셋이 많은 한국과 달리 일본은 해상사고에 대한 데이터셋이 많은 것을 보아, 일본은 해상사고에 대한 관리가 한국보다 많이 이뤄짐을 알 수 있다.

2.4.2 텍스트 마이닝을 통한 키워드 분석 및 네트워크 분석 결과

재난안전 데이터셋 연관관계를 분석하기 위하여 데이터셋의 기술 내용을 토대로 텍스트 마이닝을 통하여 핵심 키워드를 분석한 후, 그 키워드를 토대로 네트워크 연관도를 분석하였다.

Figure 4. Results of text mining in Korea's public data disaster safety sector (keyword frequency analysis, word cloud analysis).

Figure 5. Results of text mining in Korea's public data disaster safety sector (co-occurrence network analysis).

Table 5. Results of Frequency Analysis of Public Data Keywords in Korea's Disaster Safety Sector using Text Mining

Disaster sector	Frequency ranking		Safety sector	Frequency ranking	
	1. police	6. survey results		1. police	6. data
	2. child	7. status		2. safety and health	7. media
	3. traffic safety	8. fire fighting		3. public enterprise	8. local
	4. facility	9. elementary		4. occupational Safety	9. occurrence
	5. online	10. fire station		5. health	10. sign

1) 한국 재난안전 데이터셋의 키워드 및 네트워크 분석 결과

먼저 키워드 분석을 위하여 한국 재난 안전부문 공공데이터 전체 데이터셋(2,283개)을 재난 안전유형 Table 3에 따라 분류한 후, 재난과 안전으로 구분하고, 데이터 제목을 대상으로 텍스트 마이닝을 통해 핵심 명사(단어)를 추출하여, 그 중 키워드 10개를 도출하였다. 도출된 키워드 중 ‘재난’과 ‘안전’은 Seo과 Lee⁽¹³⁾의 연구에서 실행한 것과 같이 본 연구의 내용 분석에 큰 영향을 미치지 않을 것으로 판단하여 분석에서 제외시켰다.

키워드 빈도분석 결과, Figure 4와 Table 5에서 보는 바와 같이 재난부문에서 가장 많이 출연한 키워드는 ‘경찰, 어린이, 교통안전, 시설, 온라인, 설문조사결과, 현황, 소방, 초등, 소방서’ 순으로 나타났다. 10개의 키워드를 보면, 도로 교통사고, 교통, 화재 등 사회재난 중심의 데이터 비중이 자연재난 데이터보다 높고, 이를 예방하고 관리하고자 이해관계자 중심의 설문자료와 도로교통사고 현황 같은 통계 데이터를 통해 사회재난에 대한 정보를 공유해 주고 있다고 볼 수 있다.

네트워크 분석 결과, Figure 5와 같이 ‘교통사고, 설문조사, 초등학생’ 등 키워드들 간의 연관성을 보았을 때, 키워드 빈도분석 결과를 더욱 뒷받침해 주는 결과가 도출되었다. 안전부문에서 가장 많이 출연한 키워드는 ‘경찰, 안전보건, 공단, 한국산업안전, 보건, 자료, 미디어, 지방, 발생, 표지’이었으며, 생활 안전부문은 작업환경 중심의 데이터로, 한국산업안전보건공단에서 작업장 안전관련 예방 미디어 자료의 빈도가 높음을 알 수 있으며, 경찰이라는 단어를 통해 범죄 안전에 대한 데이터 빈도가 높음을 알 수 있다. 네트워크 분석결과 중 안전보건, 근로자, 외국, 경찰, 검거, 현황, 정보, 등 키워드 간의 연관성을 보았을 때, 키워드 빈

Figure 6. Results of text mining in Japan's public data disaster safety sector (keyword frequency analysis, word cloud analysis).

도분석 결과를 더욱 뒷받침해 주고 있다고 본다.

2) 일본 재난안전 데이터셋의 키워드 및 네트워크 분석 결과

일본 재난 안전부문 공공데이터 전체 데이터셋(3,244개)에 대해서는 이를 한글로 번역한 뒤, 재난 안전유형 Table 3에 따라 분류한 후 데이터 제목을 대상으로 텍스트 마이닝을 통해 핵심 명사(단어)를 추출하였고, 그 중 키워드 10개를 도출하였다. 도출된 키워드 중 ‘재난’과 ‘안전’은 먼저 분석한 바와 같이 본 연구의 내용 분석에 큰 영향을 미치지 않을 것으로 판단되어 분석에서 제외시켰다. 네트워크 분석 역시 한국과 같은 방법으로 데이터셋에 대한 초록 설명 부분을 핵심 명사(단어)를 추출하여 네트워크 분석결과를 도출하였다.

키워드 빈도분석 결과, Figure 6, Table 6과 같이 재난부문에서 가장 많이 출연한 키워드는 ‘화산, 하천, 침수, 상황, 지진, 의지, 산업, 데이터, 예보, 기본도(지도)’ 순으로 나타났다. 10개의 키워드를 보면, 자연재난 중심의 데이터로 재난 상황을 파악하기 위한 기본도(지도)데이터와, 지리정보 중심의 상황 또는 예보에 대한 종합 통계 데이터가 많은

Table 6. Results of Frequency Analysis of Public Data Keywords in Japan's Disaster Safety Sector using Text Mining

Disaster sector	Frequency ranking		Safety sector	Frequency ranking	
	1. volcano	6. will		1. year	6. data
	2. river	7. business		2. labor	7. crime
	3. flooding	8. data		3. juniors	8. situation
	4. situation	9. forecast		4. survey	9. state of affairs
	5. earthquake	10. basic map		5. business	10. cyber

Figure 7. Results of text mining in Japan's public data disaster safety sector (co-occurrence network analysis).

비중을 차지하였다. 화산, 기본도, 데이터, 지도, 파일 등 키워드 연관성을 보면 이를 뒷받침해 주고 있다. 안전부문에 서 가장 많이 출연한 키워드는 연도, 노동, 후세대, 조사, 사업, 자료, 범죄, 상황, 정세, 사이버* 키워드로 일본은 생활 안전부문 중 작업 안전에 대한 통계 데이터와 범죄 안전에 대한 공공데이터가 많은 비중을 차지하고 있다. 이는 Figure 7의 네트워크 분석결과에서 보듯이 전년도, 연차, 보고서, 노동 등 핵심키워드 간의 연관성분석 결과에서 이를 뒷받침해 주고 있다.

이러한 분석결과를 토대로 보았을 때, 한국은 도로교통 사고, 교통, 화재 등 사회재난 중심의 데이터 비중이 자연재난 데이터보다 높고, 이를 예방하고 관리하고자 이해관계자 중심의 설문자료와 도로교통사고 현황 같은 통계 데이터를 통해 사회재난에 대한 정보를 공유해 주고 있지만 일본은 화산활동, 태풍, 홍수, 자연재난 중심의 지형도, 일기예보와 같은 상황 종합 데이터의 비중이 높았다.

안전부문 공공데이터셋은 한국과 일본 모두 생활 안전 부문 중 작업 안전에 대한 데이터와 범죄 안전에 대한 통계 데이터가 많은 비중을 차지하고 있어 안전부문의 공공 데이터는 한국과 일본이 비슷한 경향을 띄고 있음을 알 수 있다.

2.4.3 데이터 품질 및 서비스 비교분석

1) 데이터 품질 비교분석

재난 안전부문 공공데이터셋 데이터 품질 비교 분석결과, 한국 재난 안전부문 데이터 품질은 최고 수준의 정보공유기술인 URI, RDF와 같은 평점 4점, 평점 5점의 데이터

Figure 8. Comparative analysis of quality levels of datasets in disaster safety sector.

양은 미비하지만, 기본 정보공유 수준의 기계 판독에 유용한 평점 3점 수준인, CSV, JSON, XML 데이터는 공공데이터 전체 비중의 79.63%로 높게 나타났다. 반면 일본은 한국보다 1단계 높은 평점 4점인 HTML 데이터 비중이 일본 공공데이터 전체 비중의 50.03%를 차지하였으며, 화산 활동과, 호우, 태풍과 관련된 이미지 파일(JPEG, GIF, PNG 등)과 PDF 형식의 평점 1점의 데이터가 나머지 비중의 47.50%를 차지하였다. 이를 나타낸 것이 Figure 8이다. 일본의 재난안전 공공데이터에서 파일 유형의 비율을 보면 HTML, PDF, XLS 순으로 비율이 높았다. 이를 나타낸 것이 Table 7이다.

Table 8을 보면 한국의 재난안전 공공데이터에서 파일 유형의 비율을 나타낸 것으로 이를 살펴보면 XML, CSV, JSON, PDF 파일형식 순으로 비율이 높았다. 파일형식의 비율을 보건대 기계가 읽을 수 있는가의 관점과 데이터의 통계분석 등 빅데이터 활용관점에서 본다면 한국이 일본보다 활용도가 앞선다고 할 수 있다.

Table 9는 데이터 품질 및 서비스 평가리스트를 기준으로 한국과 일본 데이터셋 품질을 분석한 결과이다. Table 9를 살펴보면 한국은 가용성 전체 데이터의 수량은 총 2,283개이며 불용성 데이터는 0개였다. 한국의 주제별 데이터 범위로 행정에 따른 데이터 분류체계는 16종으로 분류하였으며, 서비스 유형은 7종, 제공기관 유형은 8종, 태그는 20종, 확장자별 15종으로 분류하여 사용자의 수요에 맞는 공공데이터를 제공하고 있었다. 데이터 품질평가 항목 중, 일

Table 8. Distribution of Public Data Filename Extensions in Korea's Disaster Safety Sector

Data type	Unit		Data type	Unit		Data type	Unit	
	Count	Ratio		Count	Ratio		Count	Ratio
AI	2	0.04	DOC	1	0.02	JPG	1	0.02
AVI	3	0.05	DWG	1	0.02	JSON	1,618	29.49
CSV	1,666	30.36	HWP	127	2.31	MP	3	0.05
PPT	1	0.02	PPTX	9	0.16	RUL	2	0.04
TEXT	1	0.02	TTL	18	0.33	TXT	1	0.02
XLS	20	0.36	XLSX	17	0.31	XML	1,701	31
MP4	10	0.18	PDF	269	4.90	PNG	1	0.02
SHP	7	0.13	WMV	2	0.04	XSML	1	0.02
ZIP	5	0.09						

Table 9. Assessment Lists of Public Data Quality and Service

Evaluation contents	Japan score	Korea score	Detailed evaluation criteria
1. Whether there is data available - Insoluble data rate, %			<ul style="list-style-type: none"> 1 point : 20% or more 2 point : 15~20% 3 point : 10~15% 4 point : 5~10% 5 point : less than 5%
2. Data coverage by subject - Evaluate data classification to meet user needs			<ul style="list-style-type: none"> 1 point : Very inconvenience 2 point : Inconvenience 3 point : Generally 4 point : Convenience 5 point : Very convenient
3. Data extension distribution			Refer to tables 4 and 8, which are data openness evaluation indicators
4. Management indicators, content fulfillment - Check data classification labeling and content match			<ul style="list-style-type: none"> √ Inconsistency rate 1 point : 20% or more 2 point : 15~20% 3 point : 10~15% 4 point : 5~10% 5 point : less than 5%
5. Attributes, standards, duplicate values, linkage values - Check data classification labeling and content match			<ul style="list-style-type: none"> √ Inconsistency rate 1 point : 20% or more 2 point : 15~20% 3 point : 10~15% 4 point : 5~10% 5 point : less than 5%
6. Sufficient, accessible, utilized - Evaluation of data download support by extension			<ul style="list-style-type: none"> 1 point : no download 2 point : Supports file format only 3 point : Supports API format only 4 point : Both support 5 point : All support
7. User support - Evaluate User custom data request service			<ul style="list-style-type: none"> 1 point : Very inconvenience 2 point : Inconvenience 3 point : Generally 4 point : Convenience 5 point : Very convenient
8. Data visualization - Comparative evaluation of data visualization use cases			<ul style="list-style-type: none"> 1 point : not supported 2 point : 1 support 3 point : 1~5 support 4 point : 5~10 support 5 point : More than 10 supported

○ : 1 point : 2 point : 3 point : 4 point : 5 point

관성과 유효성은 양호하나, 준비성에서 재난 안전부문과 불일치하는 데이터가 총 12개 있었다. 공공데이터 포털 서비스 중 데이터 시각화 및 사용자 지원 서비스는 사용자가 원하는 공공데이터를 시각화분석을 통해 더 유용한 정보를 제공해 주고 있었다.

반면 일본은 가용성 전체 데이터 수는 3,244개 대비 불용성 데이터는 총 398개(12.26%)였으며, 주제별 데이터의 행정예 따른 분류체계는 내각사무국, 내각 입법국, 국가 인사청, 내각 사무실, 궁내청, 공정거래위원회, 경찰청, 금융청, 소비자청, 개인정보보호위원회, 재건 에이전시, 총무성,

재무부, 법무부, 외무성, 후생노동성, 문부과학성, 농림수산업, 국토교통부, 경제산업성, 환경부, 국방부로 총 22종으로 분류하고, 그룹 유형 17종, search tag를 각각 지원하고 있었다. 일본 역시 데이터 품질평가 항목 중 일관성과 유효성은 양호하나, 준비성에서 재난 안전부문과 불일치한 데이터가 총 23개 있었다. 공공데이터 포털 서비스 중 데이터 사용자 맞춤 지원 서비스는 지원하고 있지만, 데이터 시각화 서비스는 따로 제공하고 있지 않았다.

Figure 9는 데이터 서비스 우수 활용사례를 비교한 것이다. 한국 공공데이터 포털에서는 사용자 맞춤 지원 데이터 분석 서비스를 기반으로 다양한 공공데이터 시각화 서비스를 운영하고 있으며 현재 COVID-19 백신 접종 지역별 통계와 백신 접종이 가능한 의료기관을 위치정보 시각화 서비스를 제공하고 있었지만 최고 수준의 정보 공유 기술수준인 URI, RDF와 같은 평점 4점, 평점 5점의 데이터양이 미비하여, 이를 활용한 실시간 포털 간의 공유 서비스가 다소 부족하였다. 반면 일본은 사용자 맞춤 공공데이터 활용 서비스를 제공하고 있지만, 공공데이터 시각화 서비스는 제공하고 있지 않았지만, 일본은 한국과 달리 공공데이터 활용을 무료로 실시간 제공하고 있으며 평점 4점인 HTML 공공데이터를 재난 관련 빅데이터 분석 전문회사들과 공유하여 재난 빅데이터 전문인력양성에 노력하고 있었다.

2) 데이터 서비스 활용사례 비교분석

Figure 9. Comparative analysis of best use cases for data service.

3. 결 론

한국과 일본이 공공데이터 포털 및 데이터셋 정보내용, 데이터 품질, 서비스 등에 대한 분석 결과를 토대로 정책 방안을 제시하면 다음과 같다.

첫째, 한국 공공데이터 포털은 일본 공공데이터 포털 대비 재난 안전부문 공공데이터셋 개수도 부족하지만 전체 공공데이터셋 대비 비중도 낮아 재난 안전부문 공공데이터셋에 대한 전반적인 양적 증대 방안이 필요하다.

둘째, 재난안전 공공데이터셋의 재난안전관리단계별 구성내역에 있어서도 일본은 대비와 복구에 대한 데이터셋이 가장 많이 있는 반면, 한국은 예방 데이터셋 비중이 높아 현장 활용도를 높이는 대비 및 복구 데이터셋 확보 방안이 필요하다.

셋째, 재난 안전유형별 구성내역에 있어서도 한국은 물자관리 및 자원지원 데이터와 복합재난(공통) 데이터셋의 비중이 높았고, 일본은 자연재난(화산활동, 태풍), 상황안전, 범죄안전 데이터셋 비중이 높아 향후 재난에 대한 균형적인 안전관리를 위한 데이터셋 확보 방안이 필요하다.

넷째, 데이터 확장자를 통한 데이터 품질을 평가한 결과, 한국은 CSV, JSON, XML과 같은 평점 3점의 데이터가 비중이 높은 반면, 일본은 한국보다 1단계 높은 수준의 HTML인 평점 4점의 데이터가 높은 비중을 차지하였다. 빅데이터 통계분석 등 활용관점에서는 한국의 데이터 유형이 유리할 수 있으나 데이터 품질은 높여 나가야 할 것이다. 아울러 수요자의 다양한 요구를 충족시키기 위해서는 링크드 데이터(linked data)를 확대할 필요가 있다.

다섯째, 포털 서비스 측면에서 일본은 사용자 맞춤 공공데이터 활용 서비스를 제공하고 있으며 공공데이터 활용 교육을 무료로 제공하고 있다. 한국은 사용자를 위한 시각화 서비스 제공에서 우수하지만, 향후 사용자 맞춤 서비스의 확대와 공공데이터 활용 교육 확대 방안이 필요하다. 예를 들어 등산로 지도 데이터와 등산로 상의 구급함 위치를 표시한 데이터를 결합하여 애플리케이션으로 시각화하여 제공한다면 안전한 등산을 즐기려는 사람들의 욕구를 충족시킬 수 있을 것이다. 그리고 재난안전 데이터 활용교육 확대를 위한 방안으로서 소방청, 재난안전 직렬의 공무원을 대상으로 빅데이터 활용교육을 실행한다면 국민이 보다 더 안전하게 될 것이라 보인다.

이상으로 본 연구는 일본과 한국의 공공데이터 포털 재난안전 데이터셋 및 서비스 비교분석을 통한 한국 재난 안전공공데이터 품질 및 서비스 방향을 모색하였다. 향후에는 데이터 양적, 품질의 변화와 사용자 관점에서의 욕구, 만족도, 성과 등에 대한 연구가 필요할 것으로 본다.

References

1. J. D. Jung, "The trend of disaster management ICT in major foreign countries and Implications", KISTEP Issue Paper 2014-11, Korea Institute of Science & Technology Evaluation and Planning (2014).
2. D. Y. Jung, M. C. Kim and J. H. Kim, "Current Status of Big Data Policy and Utilization Improvement Plan", National Assembly Research Service (2018).
3. B. Marcos, "OECD Open, Useful and Re-usable data (OURdata) Index: 2019" (2020).
4. IDRC, The World Bank, DFID, "public data Barometer 4th

- Edition — Global Report”, (2017).
5. H. J. Seo, “A Critical Review on Open, Useful, Reusable Government Data Index by OECD with Level of Domestic Open Government Data : Focusing on Comparison with public data Barometer”, *Informatization policy*, Vol. 24, No. 2, pp. 43-67 (2017). <https://doi.org/10.22693/NIAIP.2017.24.2.043>.
 6. D. H. Shin and Y. M. Kim, “The Utilization of Big Data’s Disaster Management in Korea”, *The Korea Contents Society*, Vol. 15, No. 2, pp. 377-392 (2015). <https://doi.org/10.5392/JKCA.2015.15.02.377>.
 7. G. Kim, “An Evaluation of Public Data Opening Policy: Focused on Public Data Portal”, *Public Policy Review*, Vol. 31, No. 2, pp. 57-82 (2017). <https://doi.org/10.17327/ippa.2017.31.2.003>.
 8. D. J. Kim, H. J. Kim, C. E. Song, J. W. Yang and H. L. Kim, “Methods for Utilising Local Government’s Public Data Released to The Public Data Portal”, *Journal of Digital Contents Society*, Vol. 22, No. 3, pp. 445-452 (2021). <https://doi.org/10.9728/dcs.2021.22.3.445>.
 9. C. H. Lee, S. Y. Kim, S. S. Sin, Z. S. Seo, S. J. Lim, D. H. Lee, S. Y. Park, J. H. Myoung, et al., “Data Quality Assessment Procedure Manual”, ISBN 978-89-88474-09-9 93560, Korea Data Agency (2009).
 10. K. Sylvain, R. Jérémie, N. Sebastian, U. Jürgen and L. T. Yves, “Comparison of metadata quality in public data portals using the Analytic Hierarchy Process”, *Government Information Quarterly* 35, Vol. 13, pp. 13-29 (2018). <https://doi.org/10.1016/j.giq.2017.11.003>.
 11. V. Antonio, C. Lorenzo, T. Marco, O. M. Camilo, L. Raimondo and M. Federico, “Open data quality measurement framework: Definition and application to Open Government Data”, *Government Information Quarterly*, Vol. 33, No. 2, pp. 325-337 (2016). <https://doi.org/10.1016/j.giq.2016.02.001>.
 12. H. R. Kim, “Web-based data standardization trends”, *Telecommunications Technology Association*, Vol. 2017, pp. 85-89 (2017).
 13. H. J. Seo and J. Y. Lee, “Analysis of Future Education Trends Using Semantic Network Analysis”, *Education Information Media Research*, Vol. 24, No. 4, pp. 649-678 (2018). <https://doi.org/10.15833/KAFELAM.24.4.649>.