



Comparative analysis of open government data topics and usability

Jane Cho¹

Accepted: 3 February 2023 / Published online: 16 February 2023
© The Author(s), under exclusive licence to Springer Nature B.V. 2023

Abstract

This study identifies differences in the content of open public data managed by the central government, local governments, public institutions, and the office of education in Korea through keyword network analysis. Pathfinder network analysis was performed by extracting keywords assigned to 1,200 data cases, open to the Korean Public Data Portals. Subject clusters were derived for each type of government and their utility was compared using download statistics. Eleven clusters were formed for public institutions with specialized information on national issues such as *Health care* and *Real estate*, while 15 clusters were formed for the central government with national administrative information, including *Crime* and *Safety policing*. Local governments and offices of education were assigned 16 and 11 topic clusters respectively, with data focusing on regional life such as *Local factories and manufacturing*, *Resident registration*, and *Lifelong education*. Usability was higher in public and central governments that deal with national-level specialized information than for regional-level information. It was also confirmed that subject clusters such as *Health care*, *Real estate*, and *Crime* showed high usability. Furthermore, there was a large gap in data utilization because of the existence of popular data that showed extremely high usage.

Keywords Open government data · Korean Government · Pathfinder network · Subject clusters · Usability

1 Introduction

As cutting-edge information technologies such as artificial intelligence and the Internet of things continue to be combined with various new technologies, most products and services in our society are becoming intelligent. As exemplars of such source data in the era of the 4th Industrial Revolution, public data produced by central ministries, local governments, and various government-affiliated organizations are becoming increasingly useful, creating economic and social value in various forms in both public and private sectors.

✉ Jane Cho
chojane123@naver.com; chojane@inu.ac.kr

¹ Department of Library and Information Science, Institute of Social Sciences, Incheon National University, 119 Academy-Ro, Songdo 1(il)-Dong, Yeonsu-gu, Incheon, South Korea

The term “public data” refer to information or data created or authorized by the government, provided free of charge to everyone, and reusable with machine-readable features (OECD 2018). The openness of public data is a policy representative of the global open government paradigm. In 2009, the Obama administration emphasized the importance of increasing private access and utilization of public data, and established a culture of transparency, participation, and collaboration (White House 2009). In Korea, alongside the global trend of open government, the Act on Promotion of Provision and Use of Public Data was enacted in 2013 to make data held by all public institutions, including national and local governments, open to the private sector; this created socio-economic value in various fields.

In the Korea public data portal, data open institutions are divided into the central government, local government, public institutions, and the office of education. There is a difference in the content of the data disclosed by each open body. Data from the central government and public institutions deal with administrative information on national issues in a field of specialization. In contrast, local government data, accumulated in close contact with people’s daily lives, deal with local innovation and current issues closely related to the local community. As a result, public data differ in subject scope depending on the type of government, leading inevitably to a difference in usability. Therefore, it is necessary to empirically analyze the differences in the intellectual structure of data made public by different government entities to determine which of the different data characteristics lead to usability. A prior understanding of the content and utilization of government data open to private consumers is very useful in terms of concept development and design in private business projects. Additionally, from the operator’s point of view, it is essential to enhance the utility of public data for better policy development.

This study aims to identify differences in content by grouping public data thematically via a network analysis on the relationship between keywords assigned to public data. We also aimed to identify the clusters that show high usability and the types of government that provide access to such data. Keyword network analysis is a methodology that can impose an intellectual structure on a network by extracting keywords from a dataset and calculating the similarity based on the simultaneous appearance frequency of each keyword pair. This study utilized the pathfinder analysis technique (PFNet), which is a method that removes paths having links with weights that violate the triangle inequality in networks. Thus, the thematic differences in the data released by the four different types of government can be understood and interpreted in relation to their utilization. To achieve this, the following research questions were established.

(RQ1) What is the difference between the subject clusters of data released by the central government, local governments, public institutions, and the office of education?

(RQ2) What is the difference in the utilization of data disclosed by the central government, local governments, public institutions, and the office of education? In addition, which subject clusters show a high utilization?

2 Prior research

Research on public data shows various interdisciplinary characteristics, which can be summarized as a discussion on usability, quality, and content as follows.

Recent studies have discussed the quality and usability of public data. Vetrò et al. (2016) evaluated the quality of Italian open government data (OGD), dividing samples into centralized and municipal level decentralized data. They identified the problem of insufficient metadata and incomplete information provision for data updates, arguing that policy guidelines for a decentralized environment should be supplemented. Kim et al. (2021) evaluated the utilization of a local government data portal open to the Korean public. They pointed out that although the size of open data in metropolitan cities such as Seoul and Busan is large, the number of downloads is not as high, arguing that it is necessary to identify themes that can satisfy demand rather than the quantitative scale of the provided data. In another study on the properties that affect the use of each type of public data (Lim 2021), Lim identified the attributing factors that should be considered relatively more important when opening data for each public institution with various purposes. He said that an influencing factor in local governments is whether to adopt an open format that allows users to easily and freely utilize the data. In contrast, in national administrative agencies, public corporations, and industrial complexes, the amount of open data is regarded as an influencing factor.

Research has also been published that analyzes the knowledge domain surrounding public data based on semantic analysis. Charalabidis et al. (2016) targeted research papers centered on open government data (OGD) to analyze the knowledge domain. They identified four major research areas in management, policy, infrastructure, interoperability, use, and value, as well as 35 detailed research topics. The study of the semantic analysis of the open public data directive (OPDD), which is a long-term plan for open data policy, is also noteworthy. Jung and Park (2015) performed a semantic network analysis targeting the open data policies of 34 government agencies. The results revealed that "open data", "open", and "openAPI" appeared as high appearance keywords. Moreover, the keyword "quality" was suggested to have a high correlation with "public data," "information," and "utilization." They noted that it is possible to identify the core agenda included in public policy documents using semantic analysis.

Finally, as in this study, the contents of public data have been analyzed through co-word-based semantic analysis. Shin (2021) collected the titles of datasets made available by 243 local governments through public data portals and analyzed the contents using text mining techniques. It was found that local governments most frequently open information on the status of local facilities and companies as well as other topics such as environmental inspection, public facilities, licensing and enforcement, and local tourism. Lee (2020) proposed a method for predicting public data demand in a timely manner through keyword analysis of the desired data applications submitted to public data portals. According to the results of the analysis, most public data belonging to topics with high demand are provided by the domestic public data portal (data.go.kr), but public data related to the actual needs of users are rarely provided. The keywords "disease" and "treatment" have recently surged in demand as a result of the COVID-19 pandemic, but the amount of open data is low. Similarly, Cho and Ha (2020) analyzed the desired data requested from the public data portal "Data 1st Avenue" and analyzed nine topics including transportation, medical welfare, administration, and real-time data as main topics of interest.

A study that analyzes the conceptual relationship of public data is also noteworthy: Jeong et al. (2017) proposed a concept pair that predicts future mashup demand at a time when the use of public data and its convergence are becoming important issues due to the 4th Industrial Revolution.

This study is differentiated from previous studies in that it provides empirical knowledge about which types of data were opened by each government type with different

characteristics by identifying the contents of public data. In addition, it is meaningful in that it analyzed not only the content characteristics but also the usability at the same time.

3 Research method

As of June 2021, data with the highest number of downloads from the public data portal of each government type (central government, local government, public institution, and office of education) were extracted. The analysis target data comprised a total of 1,200 cases, with 300 data items extracted from four types of government.

First, the average value of download statistics was compared through descriptive statistical analysis of each government type, and the gap between the number of downloads was compared using the Gini coefficient. The Gini coefficient is an index that explains the variance of wealth and is 0 for perfect equality and 1 for inequality. It is an index used to explain the gap in various fields such as economics and information science, and it was used here to identify government data types displaying large deviations in usability. The above analysis used SPSS 25 and the *Ineq* function of R.

Second, keyword network analysis was performed to determine the difference between the subject clusters of data released by the central government, local governments, public institutions, and educational institutions. The analysis consisted of a total of 3,600 keywords with 3 words assigned to each of the 1,200 metadata. Similar concepts were integrated by performing stemming for the extracted keywords, such as synonyms, broad-word, and narrow-word. The Pathfinder weighted network analysis was performed four times by selecting keywords that showed a frequency of four or more for each government type. Pathfinder weighted network analysis refers to a network created by removing paths created by weighted links that violate the triangle inequality (Lee 2006a). Like traditional methodologies such as multidimensional scale and cluster analysis, Pathfinder not only expresses the overall structure well, but has also been evaluated as more advantageous in expressing detailed structures using parallel nearest neighbor clustering (PNNC) (Lee 2006a, 2006b). Here, a weighted network analysis was performed using COOC and Wnet developed by Lee (<https://cafe.daum.net/wnets>), and a subject cluster was created for each government type.

Third, the number of downloads of the word selected as the analysis target was used to compare utilization by cluster. After backtracking multiple public data to which the selected keyword belonged using kutool (<https://www.extendoffice.com/product/kutools-for-excel.html>), the number of downloads of the public data was summed up. For example, if there were five public data items containing the keyword *crime* after backtracking, the sum of the download numbers of each five data item was considered to be the download number for the keyword *crime*. The number of times of use by cluster is the sum of the values obtained in the previous step for keywords included in the cluster, and this was used to compare relative utility among the subject clusters.

4 Results of the analyses

4.1 Government data usability analysis

First, the results of analysis for the data showing high usability included in the top 10 high usability data were considered. Most of the data published by public institutions showed the highest utilization, and data from the central government were also included. However, the data disclosed by local governments and education authorities were not included in the top 10 rankings. Of the openly accessible data from public institutions, business district information from the Small Enterprise and Market Service, health checkup information from the National Health Insurance Service, and news big data from the Korea Press Foundation showed the highest utilization. For the central government data, 3D printing status data from the Cultural Heritage Administration and the information on the import and export of seafood from the Ministry of Oceans and Fisheries showed high utilization. The business district data from the Small Enterprise and Market Service, which displayed high utilization, are used for self-diagnosis by start-ups and for marketing strategies by merchants. In addition, the data disclosed by the National Health Insurance Service are being actively used in mobile medical services, health and obesity management, and customized medical research (Table 1).

On the other hand, if the download statistics of open data are schematically represented, a long-tail phenomenon is observed. Only a few of the data items show high utilization, and most of the data show relatively low utilization, resulting in a long tail. In particular, data released by public institutions have the longest tails, while data released by the central government also show a similar pattern. In the case of data from local governments and the office of education, there were not many cases of extremely high utilization, and the overall utilization was low. In terms of the average number of downloads (Table 2), public institutions had an overwhelming 4,462 downloads, followed by the central government with 2,444 downloads. Local governments and offices of education had relatively low download numbers of 1,602 and 554, respectively. When the Gini coefficient was used to calculate the gap between public institutions and the central government, an extremely long tail was observed with $g=0.50$ and $g=0.44$, respectively, indicating a gap in utilization. In contrast, although the overall utilization of local government and office of education was low, open data were being used relatively evenly, with $g=0.30$ and $g=0.34$, respectively (Fig. 1)

By drawing a Fig. 2 scatter plot of the utilization and gaps by government type, it is possible to visually confirm the difference in utilization patterns of the open data according to each government type. Examining the download average on the y-axis and the Gini coefficient on the x-axis revealed that public institutions and central government data are located in the first quadrant and its boundary, while local governments and the office of education are located in the third quadrant, with the two types of institutions displaying opposite patterns. Although public institutions located in the upper right corner have the largest data utilization, they also show the largest data utilization gap. Central government data also have high utilization, but gaps still exist. In contrast, the office of education data exhibit the lowest utilization with a relatively low gap index. Finally, the local government data show the smallest gap in utilization, but in terms of utilization are lower than both the central government and the public institution data. In summary, although data from public institutions and central governments dealing with nationwide administrative data in specialized fields are highly utilized, there is a large gap in the utilization of each data case. The overall utilization of the local government and office of education data, that is, regional data, is low, making the gap in the utilization not so large. Subsequently, the data of each government entities were compared through keyword network

Table 1 Highly utilized public data

Government type	Data Name	Number of Down-loads	Institution Name
Public institutions	Commercial district information	114,167	The Small Enterprise and Market Service
Public institutions	Health check-up information	53,119	National Health Insurance Service
Public institutions	News big data analysis information	47,812	Korea Press Foundation
Public institutions	Medical history information	39,346	National Health Insurance Service
National administrative agency	3D Printing Status	35,640	Cultural Heritage Administration
Public institutions	Factory registration statistical information	30,982	Korea Industrial Complex Corporation
Public institutions	National industrial complex industry trend	30,608	Korea Industrial Complex Corporation
National administrative agency	Information on import and export of seafood	29,873	Ministry of Oceans and Fisheries
Public institutions	Details of workplaces subscribed to the National Pension	29,300	National Pension Service
Public institutions	Survey data status	29,099	Korea Institute of Public Administration

Table 2 Download average and gap index of data by type of government

Type of government	Average	N	Gini
Public Institution	4,462.83	300	0.50
Central Government	2,444.73	300	0.44
Local Government	1,602.80	300	0.30
Office of Education	554.36	300	0.35
Total	2,266.18	1,200	0.54

Fig. 1 Differences in the number of downloads of public data by government type

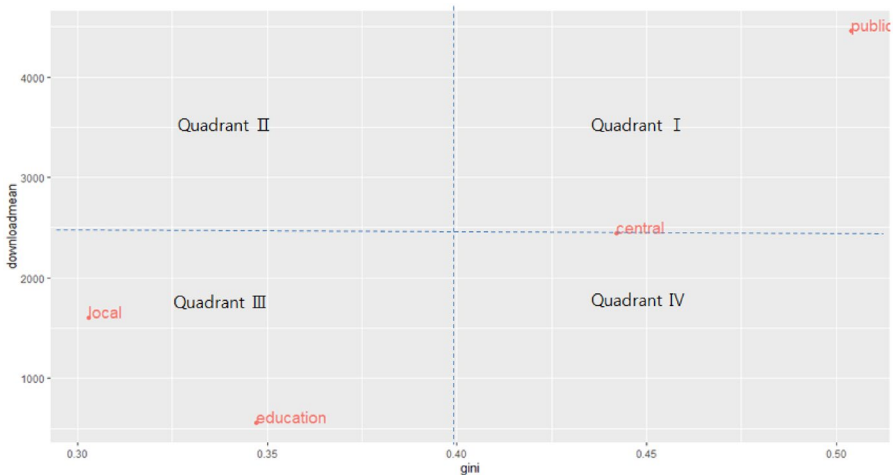
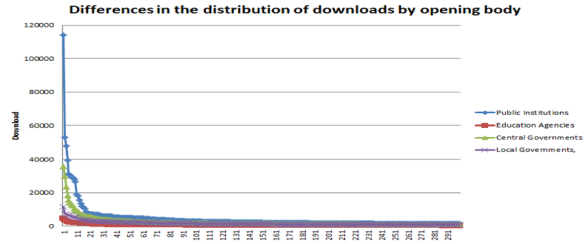


Fig. 2 Data utilization and utilization gaps by government type

analysis to identify the type of data that makes up the topic cluster and thus the cluster that shows high utilization.

4.2 Analysis of subject cluster differences

4.2.1 Public institutions

First, by integrating similar concepts through a stemming process for 900 keywords extracted from the open data of public institutions, 515 refined keywords were obtained. Among these, network analysis was performed on 40 keywords that appear four or more

Table 3 High-occurrence keywords and subject clusters (public institutions)

High-occurrence keywords	Clusters	Keywords
21 house 13 small business 11 traffic 10 guarantee 9 generation of electricity 9 statistics 8 tourism 8 technology 8 Jeju 7 economy	RealEstate (100,670) Small Business (90,873) Transportation (44,089) Energy (45,877) Statistics (117,844) Tourism (93,165) News (78,884) Bigdata (41,032) Airport (72,814) Other social issue (35,017) HealthCare (236,073)	House, realestate, public data, trust assets Small business, guarantee, technology R&D project Traffic, airport facilities Generation of electricity, new renewable energy, sale Statistics, electrical safety, information Tourism, Jeju, JDC, dutyfree Economy, news Bigdata, high frequency Transport, passenger, railroad, air route Flight section, aviation statistics Gambling problems, recycling information, nursery, insured financial company, sports Status statistics, nursing home, Medical information, health insurance, Medical treatment

times. The results of the analysis of the keyword occurrence frequency are shown in Table 3. Keywords implying specialized topics such as *house* (21), *small business* (13), *traffic* (11), and *guarantee* (10) had a high frequency of appearance. Houses, small businesses, generation of electricity, etc. are major issues in Korean society, representing topics that correspond to national issues. Network analysis of these keywords revealed 11 topic clusters (*Real estate, Small business, transportation, Energy, Statistics, Tourism, News, Bigdata, Airport, Other social issue, Healthcare*).

Among the 11 clusters in Fig. 3, the cluster with the highest utilization was *Healthcare* with an overwhelming number of downloads, much higher than that of the subject clusters of the other three government types. This category includes health checkup and treatment history information. Second, the *Real estate* cluster also showed high utilization. Collateral details, personal guarantees, and real estate public sales were included in this category. In addition, *Small businesses, Transportation, and Energy* clusters also showed high usability.

4.2.2 Central government

The results of the analysis of 48 keywords that appeared four or more times out of 436 keywords with similar concepts were integrated under central government data. First, assessment of the frequency of occurrence (Table 4) revealed that searches for *crime* were overwhelming with 61 occurrences, followed by *statistics* (29), *foreigners* (14), *national taxes* (11), and *industrial complexes* (10). In addition, keywords related to national administration, such as *land, nationality, and immigration*, are shown. Then, 15 clusters (*Crime,*

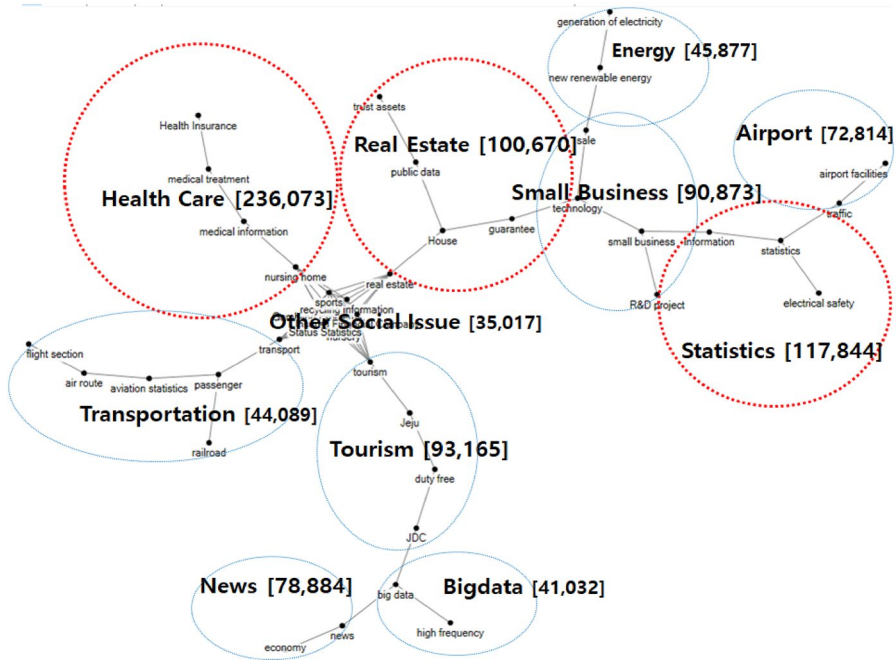


Fig. 3 Network maps and clusters (public institutions)

National statistics, Foreigner/immigration, Safety policing, Industry, Business work and other issues, Healthcare / public facility, Land, Disabled, Information, Livestock, School, Small business, Delinquent list, Accident) were formed. Excluding information and statistics that include comprehensive subject areas, the Crime cluster was the most frequently used, and the Safety policing and Foreigner/immigration clusters also showed high usability, as shown in Fig. 4. The Crime cluster and Safety policing cluster include the crime arrest status and the Foreigner/immigration cluster includes the nationality of domestic and foreign nationals, immigration status by port, and foreign residence data.

4.2.3 Local government

Analysis of 53 keywords that appeared more than four times in 515 data items with similar integrated concepts revealed that keywords closely related to local life, such as local factory and manufacturing, the categories of household and population, local transport, book catalog, and restaurant appeared most often. Among them, the most frequent were factory (31), household and population (28), and manufacturing (19), as shown in Table 5. Then, 16 clusters (Local factories and manufacturing, Resident registration status, Local transport, Local public library, Tourist attractions and commercial areas, Construction and building registration, Local industry, Local medical care, Restaurant catering, Car registration, Food hygiene, Local currency, Pollution, Livestock quarantine, Public facilities and policing, and Elderly welfare) were formed. The cluster with the highest utilization was Local factories and manufacturing, followed by Resident registration status and Car registration, as shown in Fig. 5. Local factories and manufacturing include keywords such

Table 4 High-occurrence keywords and subject clusters (central government)

High-occurrence keywords	Clusters	Keywords
61 crime	Crime	Crime, number of occurrences
29 statistics	(78,911)	
14 foreigner	National Statistics	Statistics, national rehabilitation center
11 national tax	(60,930)	Cancer
10 industrial complex	Foreigner/Immigration	Foreigner, Immigration, nationality
9 building	(53,327)	Airline
9 medical institution	Safety Policing	National tax, police, community, yearly
9 Immigration	(70,002)	
9 land	Industry	Industrial complex, industrial location
8 customs	(37,682)	
	Business, Work and Other Issues (54,955)	Building, customs, type of business, traditional food, social welfare, pollution source, working hours, nuclear power Postal statistics, 100 livelihood business
	Healthcare/ Public facility (50,827)	Medical institution, insurance, facility Books, mental illness
	Land (50,960)	Land, specific use, district
	Disabled (32,490)	Actual situation, disabled
	Information (111,047)	Information, satellite radio, seafood
	Livestock (11,827)	Livestock products, plant disease
	School (19,034)	School, education statistics
	Small Business (16,490)	Small business, technology commercialization
	Delinquent List (16,424)	Name list, seize and public sale, delinquent
	Accident (15,927)	Accident, traffic

as local companies, manufacturing, products, etc., and *Resident registration status* includes population and household status. In addition, data closely related to daily life, such as new library books, vehicle registration status, welfare facility status, tourist status, and building permit statistics are included. Thus, local governments mainly disclose data on permits, registrations, and status necessary for autonomous administration, consequently not showing high utilization as seen in data from central government and public institutions.

4.2.4 Office of education

Finally, among the data published by the office of education, the results of analyzing 38 keywords that appeared more than four times out of 253 keywords were further refined. Keywords related to school, academy, lifelong education, and library books mainly appeared. Among them, *school* (96), *school closing* (70), *academy* (49), *books* (47), and *lifelong education* (41) showed the highest frequency of appearance (Table 6). The data

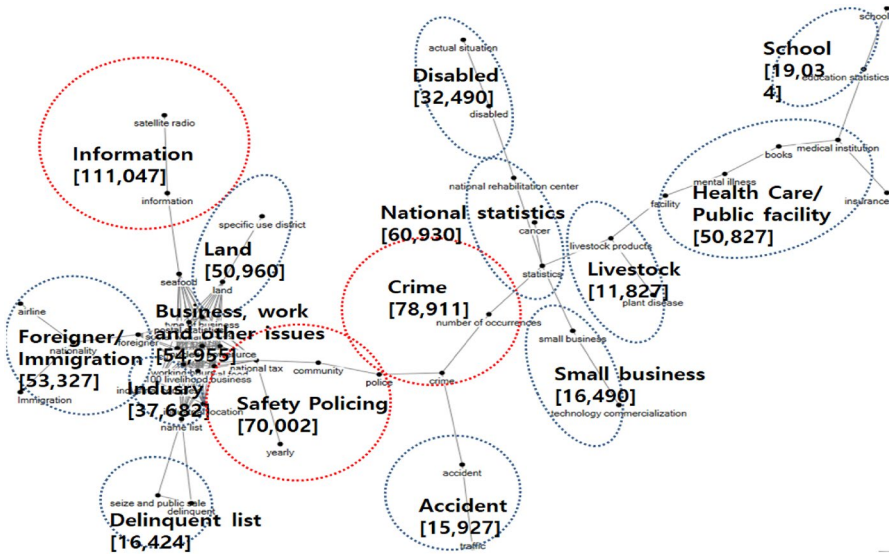


Fig. 4 Network map and clusters (central government)

were represented by 11 clusters (*Lifelong education, Office of education / school closing, Academics, Libraries, School district, Elementary and middle school, Education budget, Contact information, Kindergarten, Education statistics, and Others*). Among them, *Lifelong education* was found to be the most useful cluster, followed by *Academics* and *Libraries*, as shown in Fig. 6. The *Lifelong education* cluster includes lifelong education courses; the *Academics* cluster includes the current status of private academy and classrooms, and the *Library* cluster includes a list of new books. As with the local government data, the office of education mainly consists of education-related status or statistical data opened to the public corresponding to each educational district, as a result of which nationwide utility seems to be weak.

5 Discussion and conclusion

The results of this study have several implications. First, it was possible to accumulate empirical knowledge on the open data utilized by each government type for different purposes by identifying the content of public data, considered to be in an academic vacuum, through keyword network analysis. Such prior knowledge can increase understanding of and knowledge on data offered to users and can contribute to business planning in the private sector.

The findings based on each research question presented in this study are summarized as follows:

The first research question was whether there is any difference in the content of open data made available by different public institutions, namely, the central government, local governments, and the office of education. The identification of clusters through keyword network analysis revealed that public institutions provide specialized information on national issues; the central government provides nationwide administrative information;

Table 5 High-occurrence keywords and subject clusters (local governments)

High-occurrence keywords	Clusters	Keywords
31 factory	Local Factories and Manufacturing (97,698)	Factory, manufacturing, enterprise
28 house hold and population		Product, small business, status
19 manufacturing		
18 bus		
15 book catalog		
13 tourist destination		
12 construct		
12 industry		
12 medical treatment		
12 restaurant		
	Resident Registration Status (84,046)	Household and population, statistics Resident registration, administration Precipitation
	Local Transport (32,898)	Bus, traffic, bus stop
	Local Public Library (27,165)	Book catalog, library
	Tourist Attractions And Commercial Areas (19,076)	Tourist destination, commercial area, shopping
	Construct and Building Registration (19,070)	Construct, house
	Local Industry (20,252)	Industry, local business
	Local Medical Care (28,597)	Medical treatment, hospital, pharmacy
	Restaurant Catering (16,179)	Restaurant, catering
	Car Registration (41,329)	Car registration, registration statistics

Table 5 (continued)

High-occurrence keywords	Clusters	Keywords
	Food Hygiene (12,006)	Food processing, food
	Local Currency (17,332)	Local currency, market
	Pollution (25,493)	Pollution, air pollution, water quality
	Livestock Quarantine (21,214)	Agriculture, quarantine, livestock, industry
	Public Facility and Policing (30,135)	Floating population, disabled facilities School, seafood, day care Center CCTV, daily necessity
	Elderly Welfare (20,813)	Old man, town hall, welfare Senior facility

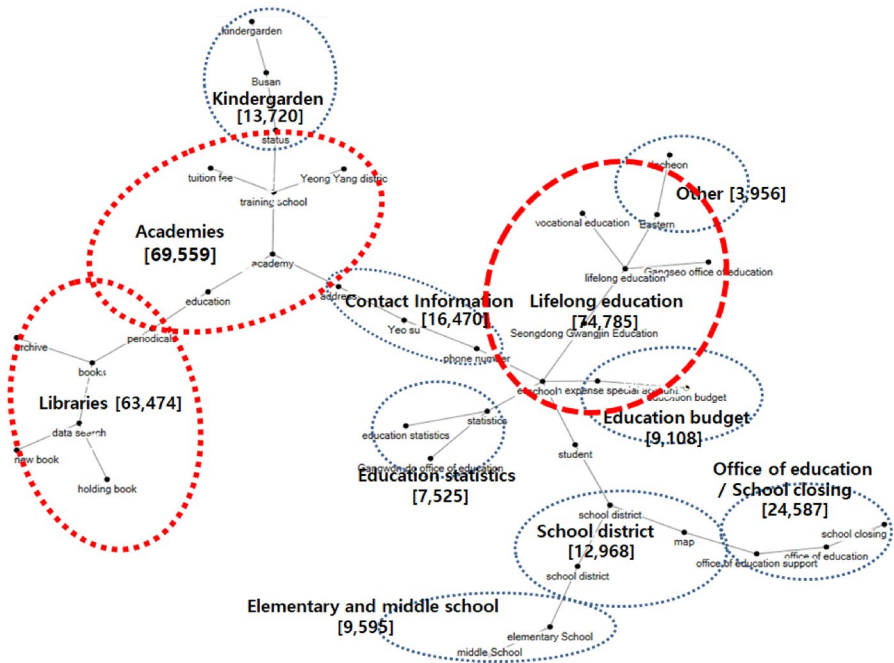


Fig. 6 Network map and clustering (Office of Education)

and local governments provide information on the status and registration of items closely related to local daily life. The Office of Education deals with current information on educational institutions and libraries of districts.

The second research question was whether there is a difference in the utilization of open data by each government type, and whether a cluster of subjects with prominent utility can be identified. The analysis revealed that the utilization of data from public institutions dealing with special information on national issues was the highest, followed by that from the central government with nationwide administrative information, characterized by a high Gini coefficient due to the existence of popular data showing extremely high utilization. In contrast, local data showed an overall lower utility than national data, and data from district offices of education had still lower utility than the local government data. On the other hand, backtracking the number of downloads of data to the assigned keywords and comparing them by clustering revealed that open data on *Health care* provided by public institutions showed overwhelming use. Data on *Real estate*, *Small businesses* in public institutes, and *National statistics* and *Crime / policing* in central government were also found to be highly useful. In addition, in the local government and office of education, relatively high utilization was observed for *Local factories and manufacturing*, *Resident registration status*, and *Lifelong education*.

However, the results of this study are based on the sampling and analysis of a total of 1,200 samples based on the number of downloads. This is a limitation of the study and follow-up studies using a larger sample are needed.

Understanding the content of public data by government type and interpreting it in relation to utilization, as in this study, is meaningful in that future utility can be predicted in preparation for developing open policies.

For future public data open policies, the following recommendations can be made. It is necessary to maximize the added value of private use, such as business creation, by opening public institution and central government data that show high usability based on open API, which is a convenient method for the private sector to use. In addition, because there is a large gap in the utilization of open data, it will be necessary to identify and develop specific strategies to improve the usability of data with low usability. If best practices for underutilized data or the possibility of mashup with other data are suggested, the use of underutilized data can be improved. However, data released by local governments and district offices of education have characteristics that are less useful than those of nationwide special data because of the quantitative limitations on data supply, low consumer awareness, and lack of human resources to promote start-ups. It is necessary to create demand from local users through continuous discovery of data with local characteristics and environment rather than recklessly increasing the scale of data (Korea Local Information Development Institute Editorial Office 2017). In addition, policies are needed that facilitate an unobstructed path between supply and demand, such as matching public data with local companies and trigger competition.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11135-023-01630-x>.

Author contributions The author confirms sole responsibility for the following: study conception and design, data collection, analysis and interpretation of results, and manuscript preparation.

Funding The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Declarations

Conflict of interest The author declares no conflict of interest.

References

- Charalabidis, Y., Alexopoulos, C., Loukis, E.: A taxonomy of open government data research areas and topics. *J. Organ. Comput. Electron. Commer.* **26**(1–2), 41–63 (2016). <https://doi.org/10.1080/10919392.2015.1124720>
- Cho, S., Ha, S.: Analysis of open Government data demand using structural topic modeling. *J. Inf. Technol. Archit.* **17**(2), 103–118 (2020)
- Jeong, D., Yi, D., Shin, D.: A study on the linkability of public information using social network analysis. *J. Korean Soc. Surv. Geod. Photogramm. Cartogr.* **35**(6), 461–470 (2017)
- Jung, K., Park, H.: A semantic (TRIZ) network analysis of South Korea's "Open Public Data" policy. *Gov. Inf. Q.* **32**(33), 353–358 (2015)
- Kim, D., Kim, H., Song, C., Yang, J., Kim, H.: Methods for utilising local Government's public data released to the public data portal. *J. Digit. Contents Soc.* **22**(3), 445–452 (2021)
- Korea Local Information Development Institute Editorial Office: Status of public data opening by local governments and measures to promote use. *Local Informatiz.* **104**, 26–32 (2017)
- Lee, J.: A study on the network generation methods for examining the intellectual structure of knowledge domains. *J. Korean Soc. Libr. Inf. Sci.* **40**(2), 333–355 (2006a)
- Lee, J.: A novel clustering method for examining and analyzing the intellectual structure of a scholarly field. *Korea Soc. Inf. Manag.* **23**(4), 215–231 (2006b)
- Lee, J.: Forecasting open Government data demand using keyword network analysis. *Informatiz. Policy* **27**(4), 24–46 (2020)
- Lim, J.: A Study on the Data Attribute Factors Influencing the use of the Open Data. Hanyang University, Seoul (2021)

- OECD. Open Government Data Report. Available at: <https://www.oecd.org/gov/digital-government/open-government-data-report-highlights-en.pdf>. (2018)
- Shin, S., Choi, H., Kim, D.: Analyzing contents of open data in Korean local governments: a text-mining approach. *Korean J. Public Adm.* **30**(2), 129–171 (2021)
- Vetrò, A., Canova, L., Torchiano, M., Minotas, C.O., Iemma, R., Morando, F.: Open data quality measurement framework: definition and application to Open Government Data. *Gov. Inf. Q.* **33**(2), 325–337 (2016)
- White House. Memorandum on Transparency and Open Government. Administration of Barack H. Obama, 2009. Available at: <https://obamawhitehouse.archives.gov/the-press-office/transparency-and-open-government>. (2009)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.