JISTaP http://www.jistap.org
Journal of Information Science Theory and Practice
eISSN : 2287-4577 pISSN : 2287-9099

Research Paper
J Inf Sci Theory Pract 6(1): 18-30, 2018
http://dx.doi.org/10.1633/JISTaP.2018.6.1.2

Interlinking Open Government Data in Korea using Administrative District Knowledge Graph

Haklae Kim *

Korea Institute of Science and Technology Information
Yuseong-gu, Daejeon, Korea
E-mail: haklaekim@gmail.com

ABSTRACT

Interest in open data is continuing to grow around the world. In particular, open government data are considered an important element in securing government transparency and creating new industrial values. The South Korean government has enacted legislation on opening public data and provided diversified policy and technical support. However, there are also limitations to effectively utilizing open data in various areas. This paper introduces an administrative district knowledge model to improve the sharing and utilization of open government data, where the data are semantically linked to generate a knowledge graph that connects various data based on administrative districts. The administrative district knowledge model semantically models the legal definition of administrative districts in South Korea, and the administrative district knowledge graph is linked to data that can serve as an administrative basis, such as addresses and postal codes, for potential use in hospitals, schools, and traffic control.

Keywords: Open Data, Government Data, Knowledge Graph, Ontology

Open Access

Accepted date: March 6, 2018
Received date: November 9, 2017

***Corresponding Author:** Haklae Kim
Research Engineer
Korea Institute of Science and
Technology Information
Yuseong-gu, Daejeon, Korea
E-mail: haklaekim@gmail.com

All JISTaP content is Open Access, meaning it is accessible online to everyone, without fee and authors' permission. All JISTaP content is published and distributed under the terms of the Creative Commons Attribution License (http:/ creativecommons.org/licenses/by/3.0/). Under this license, authors reserve the copyright for their content; however, they permit anyone to unrestrictedly use, distribute, and reproduce the content in any medium as far as the original authors and source are cited. For any reuse, redistribution, or reproduction of a work, users must clarify the license terms under which the work was produced.

ⓒ Haklae Kim, 2018

1. INTRODUCTION

The White House report “Preparing for the Future of Artificial Intelligence” analyzes future changes in human society and public policies due to artificial intelligence technologies and proposes new directions (NSTC, 2016). Interestingly, the report emphasizes the importance of data as well as artificial intelligence technologies. It makes 23 recommendations, the second of which proposes open training data for artificial intelligence and open data standardization. Similarly, the UK government has emphasized the importance of data in a report on growth in the artificial intelligence industry (Hall & Pesenti, 2017). In particular, they emphasized the importance of data that can be processed by machines in an artificial intelligence environment. Artificial intelligence is inevitably connected to data (Section 3: Recommendations to improve access to data). However, securing the necessary data is a very difficult problem for data consumers. In such an environment, the movement of governments to open their data to the private sector can be a very important catalyst for the development of artificial intelligence industries.

The open government data policy started by the UK and USA has expanded to Europe, Asia, and Africa (Tran & Scholtes, 2015). The South Korean government is promoting an open data policy through the “Act on Promotion of the Provision and Use of Public Data” (MOIS, 2013) and has opened a large number of datasets through its open data portal.[1] As of October 2017, this site provides 20,797 files and 2,385 application programming interfaces (APIs) of 686 agencies. What are the ripple effects of open government data? South Korea is ranked first in the open data field hosted by OECD (Kang, 2017) and fifth in the Open Data Barometer (Web Foundation, 2017), which is an international open data evaluation index. However, the use of open government data is limited, and users have complained that it is difficult to create new values through open government data (Seo, 2017a; Seo, 2017b).

In general, effective use of open government data can be examined from a policy and technical standpoint. From a policy perspective, the government has delivered more demanding data in priority. For example, national core data is a high-quality and large-volume dataset that is selected by some criteria such as high demand for openness and urgency based on demand. However, datasets that are published without consideration of data quality are restricted for use in the private sector. Various technical issues are discussed in terms of open data such as data quality, data formats, or data integration (Matheus et al., 2014; Safarov et al., 2017). In particular, connecting of heterogeneous datasets should be taken seriously. Public data tends to be partially published from the source system possessed by the government, and information defined in the source system such as data structure and relationships between other tables is often not disclosed or is lost in the process of data publication. This presents a challenge for the data consumer to understand and use the data. In addition, the connection and integration of data is necessary to use fragmented and partial data. In this reason, a reference point is required to use common data between heterogeneous data sources. The more data is linked through common data, the higher the quality of the data. From a technical point of view, Linked Data can be utilized as a technique to connect data semantically using web standards (Berners-Lee, 2006; Bizer et al., 2007). In addition, by representing different public data as Linked Data, utilization in the field such as artificial intelligence, the Internet of Things, and smart cities can be enhanced (Henrique et al., 2017).

This paper proposes an ontology model for semantically expressing administrative districts with high utilization in public data, and then introduces a knowledge graph of all the objects constituting the administrative districts of the Republic of Korea. The administrative district knowledge graph can be used as basic data for making connections between public data. The contents of this paper are as follows. Section 2 examines the definition of public data with its similar terms and utilization issues. Section 3 examines the concept and status of administrative districts and proposes a knowledge model of administrative districts. Section 4 describes the current status of the administrative district knowledge graph and introduces use cases for interlinking some datasets with this graph. Section 5 concludes and discusses future research.

---

[1] http://data.go.kr

2. PROBLEMS WITH OPEN GOVERNMENT DATA UTILIZATION

2.1. Definition of Open Government Data

According to the Open Data Handbook,[2] open data are characterized by availability and access, reuse and redistribution, and universal participation (Open Knowledge, 2017). Open data must be usable as a whole and not by parts. It must be possible to download or modify via the Web. It must also be possible for open data to be reused and redistributed through a combination with other data. Finally, anyone must be able to use the data without discrimination against specific persons or groups. Public sector information made available to the public as open data is termed "Open Government Data." According to the Act on Promotion of the Provision and Use of Public Data (MOIS, 2013), open government data refers to "the data or information processed by optical or electronic methods, such as databases and electronic files, created or acquired by public agencies for purposes specified in laws." From an ownership perspective, open government data are originally possessed by the government but are provided to the private sector for government transparency and efficiency. "Provision" in the Public Data Act means "permission by public agencies for the users to access public data in forms that can be read by machines or transmission of public data in various ways" (MOIS, 2016a). "Machine readable" refers to a state in which data can be modified, converted, and distributed using software (MOIS, 2016a). In summary, open government data can be interpreted as a category of open data because they share the three characteristics that define open data.

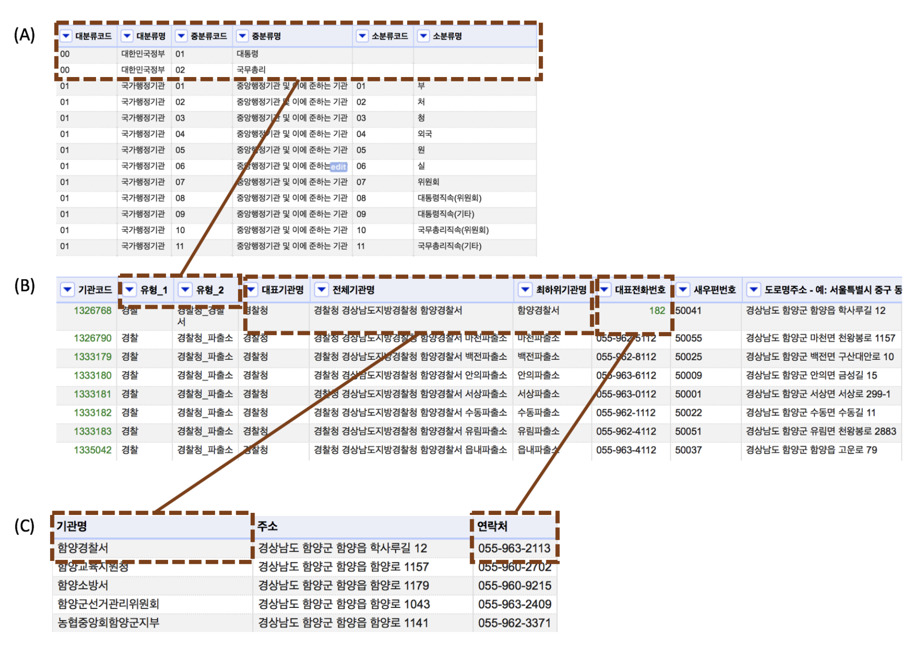
2.2. Open Government Data Utilization Issues

In South Korea, the Public Data Utilization Support Centre[3] and Public Data Strategy Committee[4] is in charge of practical support for public data release and utilization through the Act on Promotion of the Provision and Use of Public Data. However, impacts through open government data are insufficient. There are various factors that have reduced the utilization of open government data in South Korea, including performance oriented policies and a lack of shared culture. In addition, insufficient standardization of open government has been pointed out as an obstacle to data utilization.

Figure 1 shows an example of the data released as open government data. Three datasets have common information about administrative agencies. Figure 1-(A) shows government agencies, which are classified into three different levels from the administrative standard code management system.[5] Figure 1-(B) shows a dataset of government agencies provided by the open data portal (as of August 31, 2016),[6] which includes information on police substations, post offices, public health centres, and regional offices managed by the Ministry of the Interior and Safety (MOIS). Figure 1-(C) is the agency status data of Hamyang-gun, Gyeongsangnam-do (2017).[7] This example has two implications: 1) data quality and 2) data interlinking. The three datasets include information on administrative agencies, but their data fields and the values of each field are different. For example, "*대분류코드*" (*classification_large*) and "*유형_1*" (*type_1*) or "*전체*

---

[2] http://opendatahandbook.org/
[3] http://bit.ly/2v2avKO
[4] http://bit.ly/2v4hsL6
[5] http://code.go.kr
[6] http://bit.ly/2tBBuJk
[7] http://bit.ly/2tBNH0h



(A)

대분류코드	대분류명	중분류코드	중분류명	소분류코드	소분류명
00	대한민국정부	01	대통령		
00	대한민국정부	02	국무총리		
01	국가행정기관	01	중앙행정기관 및 이에 준하는 기관	01	부
01	국가행정기관	02	중앙행정기관 및 이에 준하는 기관	02	처
01	국가행정기관	03	중앙행정기관 및 이에 준하는 기관	03	청
01	국가행정기관	04	중앙행정기관 및 이에 준하는 기관	04	외국
01	국가행정기관	05	중앙행정기관 및 이에 준하는 기관	05	원
01	국가행정기관	06	중앙행정기관 및 이에 준하는edit	06	실
01	국가행정기관	07	중앙행정기관 및 이에 준하는 기관	07	위원회
01	국가행정기관	08	중앙행정기관 및 이에 준하는 기관	08	대통령직속(위원회)
01	국가행정기관	09	중앙행정기관 및 이에 준하는 기관	09	대통령직속(기타)
01	국가행정기관	10	중앙행정기관 및 이에 준하는 기관	10	국무총리직속(위원회)
01	국가행정기관	11	중앙행정기관 및 이에 준하는 기관	11	국무총리직속(기타)

(B)

기관코드	유형_1	유형_2	대표기관명	전체기관명	최하위기관명	대표전화번호	새우편번호	도로명주소 - 예: 서울특별시 중구 동
1326768	경찰	경찰청_경찰서	경찰청	경찰청 경상남도지방경찰청 함양경찰서	함양경찰서	182	50041	경상남도 함양군 함양읍 학사루길 12
1326790	경찰	경찰청_파출소	경찰청	경찰청 경상남도지방경찰청 함양경찰서 마천파출소	마천파출소	055-962-5112	50055	경상남도 함양군 마천면 천왕봉로 1157
1333179	경찰	경찰청_파출소	경찰청	경찰청 경상남도지방경찰청 함양경찰서 백전파출소	백전파출소	055-962-8112	50025	경상남도 함양군 백전면 구산대안로 10
1333180	경찰	경찰청_파출소	경찰청	경찰청 경상남도지방경찰청 함양경찰서 안의파출소	안의파출소	055-963-6112	50009	경상남도 함양군 안의면 금성길 15
1333181	경찰	경찰청_파출소	경찰청	경찰청 경상남도지방경찰청 함양경찰서 서상파출소	서상파출소	055-963-0112	50001	경상남도 함양군 서상면 서상로 299-1
1333182	경찰	경찰청_파출소	경찰청	경찰청 경상남도지방경찰청 함양경찰서 수동파출소	수동파출소	055-962-1112	50022	경상남도 함양군 수동면 수동길 11
1333183	경찰	경찰청_파출소	경찰청	경찰청 경상남도지방경찰청 함양경찰서 유림파출소	유림파출소	055-962-4112	50051	경상남도 함양군 유림면 천왕봉로 2883
1335042	경찰	경찰청_파출소	경찰청	경찰청 경상남도지방경찰청 함양경찰서 읍내파출소	읍내파출소	055-963-4112	50037	경상남도 함양군 함양읍 고운로 79

(C)

기관명	주소	연락처
함양경찰서	경상남도 함양군 함양읍 학사루길 12	055-963-2113
함양교육지원청	경상남도 함양군 함양읍 함양로 1157	055-960-2702
함양소방서	경상남도 함양군 함양읍 함양로 1179	055-960-9215
함양군선거관리위원회	경상남도 함양군 함양읍 함양로 1043	055-963-2409
농협중앙회함양군지부	경상남도 함양군 함양읍 함양로 1141	055-962-3371

Fig. 1 An example of data quality of open government data

기관명" (whole agency name) and "*기관명*" (agency name) are different data fields, and the values of "*함양 경찰서*" (Hamyang police station) are also different. In this case, people can interpret them as the same information, but machines cannot understand their identical relationship. In the case of the fields for telephone numbers, the field names are different ("*대표전화번호*" [representative telephone number] versus "*연락처*" [contact number]), and the values are not identical. Such problems frequently occur throughout open government data and lower the confidence in data quality, even though the government provides some guidelines for opening their data (MOIS, 2016a; MOIS, 2017). Second is a lack of reference information for interlinking data. It is essential to interlink datasets in situations where most open government data are available to administrative agencies or departments. The agency codes ("*기관코드*") in Figure 1-(B) can be important reference information for identifying data. If such agency codes are included in Figure 1-(C), it may be possible to interlink the data. The reference information plays a very important role in interlinking data that are currently released or to be released in the future and ultimately is a very effective factor for improving the use of open government data.

In summary, the effective use of open government data requires making it possible to interlink data by improving the quality of the released data and utilizing reference information for the values that are used in the released data. From a technical perspective, this can be realized by applying ontology and linked data (Kim, 2010; Kim, 2017a). In this paper, the proposed administrative district knowledge model presents semantic definitions of the administrative districts in South Korea and the relationship between the administrative districts, and a method is proposed for using data consistently based on code systems related to the administrative districts.

3. ADMINISTRATIVE DISTRICT KNOWLEDGE MODEL

3.1. Model for Administrative Districts in South Korea

Administrative districts are administrative units that compartmentalize the territory of a nation, which is a unit of politics, according to the purpose of the national administration. It is important in national administration, but they are particularly important in the local jurisdiction of local governments. From a data perspective, open government data includes a variety of areas of data on the country, including population, budget, maps, transport, and environment, where such information is typically divided into administrative divisions. Accordingly, administrative district information could be the common data that connects different public data.

As shown in Figure 2, administrative districts in South Korea currently consist of one special city, six metropolitan cities, one special autonomous city, eight provinces, and one special self-governing province. According to the Local Autonomy Law (MOIS, 2016c), administrative districts in South Korea are composed of upper-level local autonomies, lower-level local autonomies, and subordinate administrative districts. There are towns, townships, and neighborhoods. Towns and townships have rural villages as subordinate organizations. Almost all municipalities have urban villages and hamlets as subordinate organizations. Details on the administrative districts prescribed by the Local Autonomy Law are given below.

- Upper-level local autonomies: These include a special city, a special autonomous city, metropolitan cities, provinces, and a special self-governing province. Upper-level local autonomies are administrative organizations with higher autonomy than lower-level local autonomies and oversee larger areas.
- Lower-level local autonomies: These include cities, counties, and districts. In general, lower-level local autonomies are lower organizations than upper-level local autonomies and oversee smaller areas. The administrative cities and non-autonomous districts of the Jeju special self-governing province are not included in lower-level local autonomies.

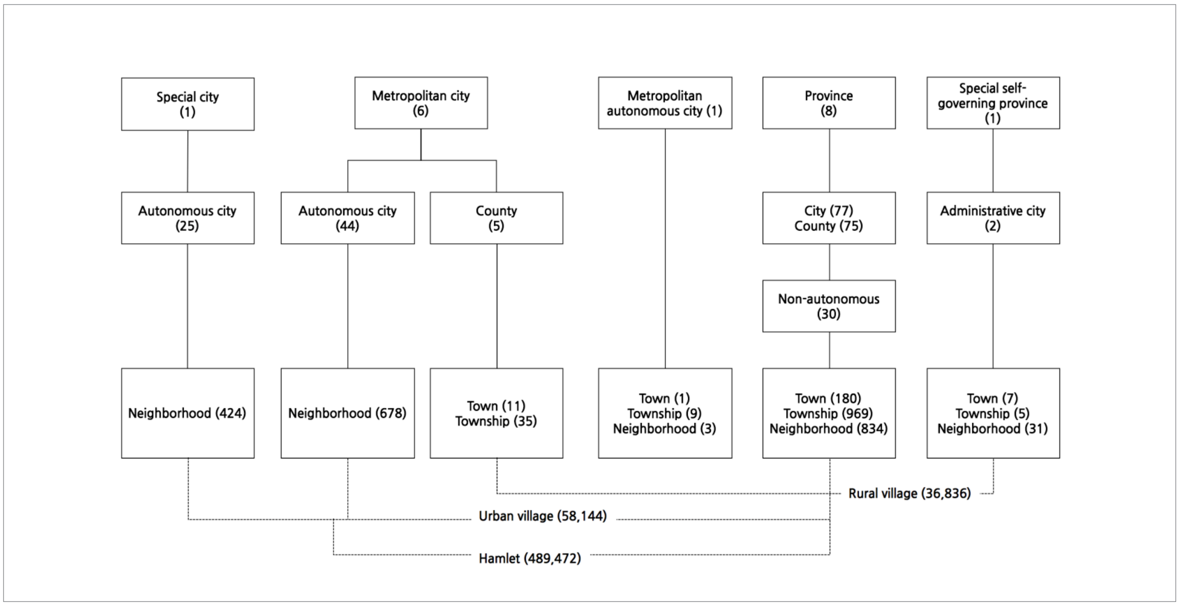

Special city (1)
Metropolitan city (6)
Metropolitan autonomous city (1)
Province (8)
Special self-governing province (1)
Autonomous city (25)
Autonomous city (44)
County (5)
City (77)
County (75)
Administrative city (2)
Non-autonomous (30)
Neighborhood (424)
Neighborhood (678)
Town (11)
Township (35)
Town (1)
Township (9)
Neighborhood (3)
Town (180)
Township (969)
Neighborhood (834)
Town (7)
Township (5)
Neighborhood (31)
Rural village (36,836)
Urban village (58,144)
Hamlet (489,472)

Fig. 2 Administrative district system and status in South Korea

- Subordinate administrative districts: The special city includes autonomous districts, the metropolitan cities include autonomous districts and counties, and the provinces include autonomous cities and counties.
- The special city and metropolitan cities may have autonomous districts as subordinate administrative districts, but the provinces have general districts (non-autonomous) as subordinate administrative districts of autonomous cities.
- The special autonomous city constitutes a basic local government.
- The special self-governing province has non-autonomous administrative cities. The administrative cities are directly governed by the governor of the special self-governing province and have no authority as lower-level local autonomies.
- Except for the special city and metropolitan cities, cities with a population of 500,000 or more may be designated as general districts. Because general districts are the subordinate administrative agencies of lower-level local autonomies, their leaders are not elected but appointed by the mayor of an autonomous city.
- Cities (autonomous and administrative) and districts (autonomous and general) have towns, townships, and neighborhoods, while counties have towns and townships as subordinate administrative districts. Towns and townships are divided into administrative rural villages, and neighborhoods are divided into urban villages. Urban villages and administrative rural villages are divided into hamlets, which are the lowest administrative districts.

The administrative district knowledge model semantically expresses the relationship between the administrative districts defined in the Local Autonomy Act (Kim, 2017b). The *KoreaAdministrativeDivision* class, which expresses the administrative districts in South Korea, is the highest class of all administrative districts, including upper- and lower-level local autonomies and non-autonomous divisions. Figure 3 shows the relationship between the upper and lower levels of the administrative district system in South Korea. The

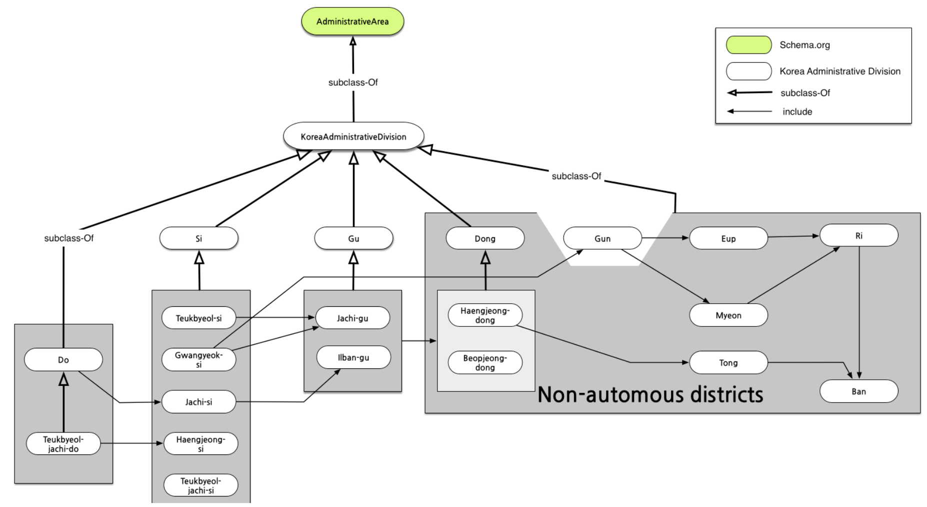

AdministrativeArea
subclass-Of
KoreaAdministrativeDivision
subclass-Of
Schema.org
Korea Administrative Division
subclass-Of
include
subclass-Of
Si
Gu
Dong
Gun
Eup
Ri
Teukbyeol-si
Gwangyeok-si
Jachi-si
Haengjeong-si
Teukbyeol-jachi-si
Jachi-gu
Ilban-gu
Haengjeong-dong
Beopjeong-dong
Myeon
Tong
Ban
Do
Teukbyeol-jachi-do
Non-automous districts

Fig. 3 Administrative district knowledge model. The *KoreaAdministrativeDivision* class is a sub-class of the *AdministrativeArea* of schema.org. It means that the terms of schema.org are reused in this model.

administrative district knowledge model applies the administrative units defined in the administrative district system and the relationships between such units. The relationship between administrative districts is represented by the *ad:include* and *ad:partOf* attributes, and the two attributes are inversely related (*owl:inverseOf*). For example, an upper-level administrative district may include a lower-level administrative district. Conversely, a lower-level administrative district would be included in an upper-level administrative district.

For municipalities, dong has both administrative and legal definitions. For example, *Myeong-dong* in *Jung-gu*, Seoul special city is the administrative dong and includes legal dong such as *Euljiro 1-ga, Euljiro 2-ga, Namdaemunro 1-ga, Samgak-dong, Suha-dong, Changgyo-dong, Hoehyeon-dong 3-ga, Chungmuro 2-ga, Myeong-dong 1-ga, Myeong-dong 2-ga, Namsan-dong 1-ga, Namsan-dong 2-ga, Namsan-dong 3-ga, Jeodong 1-ga, Mugyo-dong, Dadong*, and *Taepyeongro 1-ga*. In other words, a legal dong is the name of the area that is the base of the address, while an administrative dong is the name of the place where the resident centre (i.e. community centre) is located. While the administrative dong is used for distinguishing all administrative or electoral districts, it is common to mix the administrative dong and legal dong together. The administrative district knowledge model expresses the legal dong and administrative dong with an inverse relation. The *ad:isHaengjeongdongOf* attribute defines an administrative dong as associated with a specific legal dong. The *ad:isBeopjeongdongOf* attribute is used for the opposite case, i.e. to express a legal dong included in an administrative dong.

3.2. Design Characteristics

The administrative district knowledge model defines a minimum number of key terms and reuses the universally used ontology terms. In general, the reuse of ontology terms is understood as a basis for improving the accessibility of a knowledge base and realizing interoperability between knowledge bases. The administrative district knowledge model defines the hierarchical relationship between terms based on the schema (schema.org) terms and reuses the Dublin Core metadata terms. For example, all administrative district classes are subordinate classes of *schema:AdminstrativeArea*, and their class names use the *dc:title* attribute of Dublin Core.

A uniform resource identifier (URI) must be defined and extended consistently as a key element for connecting data. The administrative district knowledge model defines the URI system according to the class and instance levels. Classes are classified through *def* and define directories such as *ad* that can identify domains (administrative districts). The class (*Gu*) corresponding to an administrative district is located at the end as shown in Figure 4-(A). This instance uses *id* to distinguish it from *def*, which is the identifier of the class, and adds a directory to distinguish the type of the instance. The following example is for *Jung-gu*, Seoul special city. The autonomous district (*Jachi-gu*) type is defined, and the administrative classification code (1114000000) is defined as an identifier as follows:

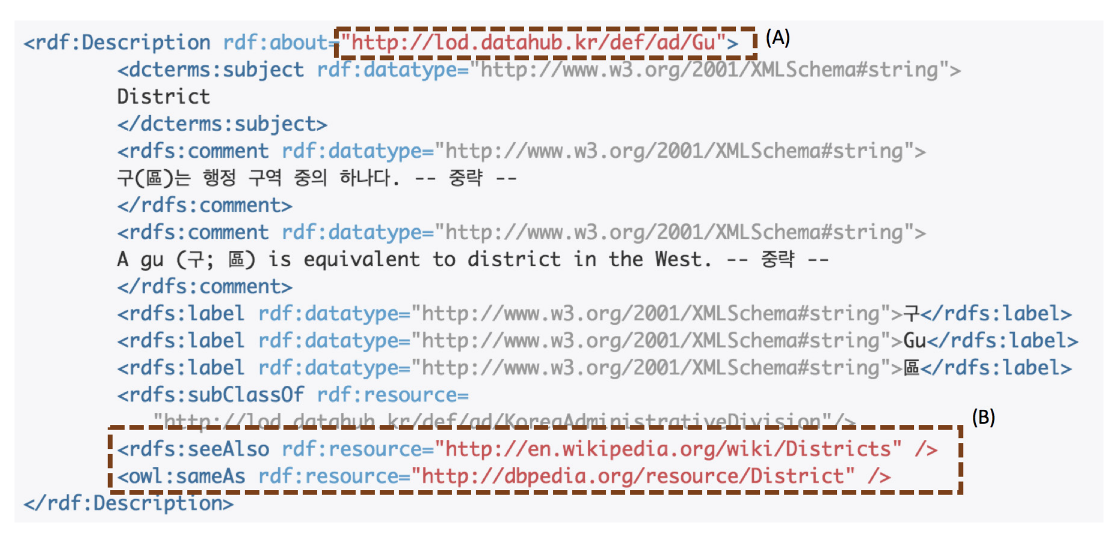
http://lod.datahub.kr/id/ad/Jachi-gu/1114000000

The vocabularies for ontology are generally defined in English. However, there is a limit to expressing the names of administrative districts in South Korea in English. For example, it is ambiguous whether a 'district' refers to a city or hamlet and is not suitable for defining the original meaning of a term. In addition, the Korean names are preferred, because specific terms for administrative districts are defined and used in Korean. Thus, the knowledge model generally uses Korean terms and Romanization notation, and the corresponding English names are written together. As shown in Figure 4-(B), the '*Gu*' class has its own URI (i.e. *http://lod.datahub.kr/def/ad/Gu*), and simultaneously it provides additional information for meaning in English using several properties such as *dcterms:subject or owl:sameAs*.

4. CONSTRUCTION AND APPLICATION OF THE KNOWLEDGE GRAPH

4.1. Status of the Administrative District Graph

Information related to administrative districts, addresses, and locations is commonly included in the data released by the government. In other words, consistent administrative district data can be effectively used to improve the quality of open government data and interlink such data. An administrative district knowledge graph was constructed by applying the administrative district data of South Korea to the administrative district knowledge model. Administrative district data



```
<rdf:Description rdf:about="http://lod.datahub.kr/def/ad/Gu">
        <dcterms:subject rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
        District
        </dcterms:subject>
        <rdfs:comment rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
        구(區)는 행정 구역 중의 하나다. -- 중략 --
        </rdfs:comment>
        <rdfs:comment rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
        A gu (구; 區) is equivalent to district in the West. -- 중략 --
        </rdfs:comment>
        <rdfs:label rdf:datatype="http://www.w3.org/2001/XMLSchema#string">구</rdfs:label>
        <rdfs:label rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Gu</rdfs:label>
        <rdfs:label rdf:datatype="http://www.w3.org/2001/XMLSchema#string">區</rdfs:label>
        <rdfs:subClassOf rdf:resource=
          "http://lod.datahub.kr/def/ad/KoreaAdministrativeDivision"/>
        <rdfs:seeAlso rdf:resource="http://en.wikipedia.org/wiki/Districts" />
        <owl:sameAs rdf:resource="http://dbpedia.org/resource/District" />
</rdf:Description>
```

Fig. 4 Example of the knowledge model

were constructed based on the resident registration address code[8] of MOIS and Korea administrative district classification[9] data of Statistics Korea. The former has information on the legal dong (45,957 cases), and the latter includes the relationship between the legal dong and administrative dong (21,695 cases). Therefore, individual open government data can be interlinked regardless of the legal dong and administrative dong. In addition, various open government data with administrative district information can be interlinked on a semantic level. For example, both the road name address and postal code use administrative district information; at the same time, they can be the base information for other data. In this manner, the data of elementary schools, junior high schools, high schools, universities, parking lots, subways, hospitals, and cultural assets were continuously interlinked in the administrative district graph. Table 1 summaries datasets that are interlinked to the administrative district knowledge graph.

4.2. Interlinking Administrative Agencies

The administrative district graph and administrative agency data were interlinked as follows: 1) the entire administrative agency data were converted into a graph, 2) the administrative agency graph and frontline administrative agency data were interlinked, and 3) the administrative agency graph and data of *Hamyang-gun, Gyeongsangnam-do* were interlinked.

---

[8] http://bit.ly/2xxQIVq

[9] http://bit.ly/1PQArf0

Table 1. Statistics of the Administrative District Graph and its Datasets

Dataset	Data Source	Data format	Size	Connected datasets
Administrative District	The Ministry of the Interior and Safety	CSV, XLS	45,957	Administrative District
Administrative Agency	The Ministry of the Interior and Safety	CSV, XLS	14,183	Administrative District
Road-name address	The Ministry of the Interior and Safety	CSV, XLS	347,242	Administrative District
Postcode	The Ministry of the Interior and Safety	CSV, XLS	302,125	Administrative District, Road-name address
Elementary school	Open Data Portal	CSV, XLS	6,247	Administrative District, Postcode
Middle school	Open Data Portal	CSV, XLS	3,204	Administrative District, Postcode
High School	Open Data Portal	CSV, XLS	2,304	Administrative District, Postcode
University	Open Data Portal	CSV, XLS	1,837	Administrative District, Postcode
Kindergarten	Open Data Portal	CSV, XLS	8,696	Administrative District, Postcode
Hospital	Open Data Portal	CSV, XLS	84,731	Administrative District, Postcode
Cultural Heritage	Open Data Portal	CSV, XLS	13,704	Administrative District
Parking lot	Seoul Data	CSV	471	Administrative District
Operation time of subways	Seoul Open Data Plaza	CSV	3,337	Administrative District
Lines of subways	Seoul Open Data Plaza	CSV	20	Administrative District
Subway stations	Seoul Open Data Plaza	CSV	581	Administrative District

- The administrative dong code and location code fields of the administrative agency data were interlinked with the administrative dong and legal dong agency codes of the administrative district graph. The postal code, remaining address, and parcel number fields were linked through the postal code and road name address graph. Because the type fields *classification_large, classification_medium*, and *classification_small* in the administrative agency data were represented as integers, the task of converting the administrative agency type classification data into a graph and linking them was added. The administrative agency type classification used a classification system for the government of South Korea, legislative/judicial/constitutional agencies, state administrative agencies, and municipalities. Here, 393 cases were constructed.
- The agency code, whole agency name, and lowest agency name fields of the frontline administrative agency data were consistent with those of the administrative agency data. Therefore, the two datasets were interlinked by identifying identical data based on the agency code. However, the expression values of the type information were different. As noted earlier, the administrative agency data had integer values, but the frontline administrative agency data had character values for the type classification. The type fields *classification_medium* and *classification_small* in the administrative agency data were mapped to the *type_1* and *type_2* fields in the frontline administrative agency data with the type classification graph data. For example, the Seoul regional office of the Ministry of Patriots and Veterans Affairs (MPVA) was classified into *type_1 = PV* and

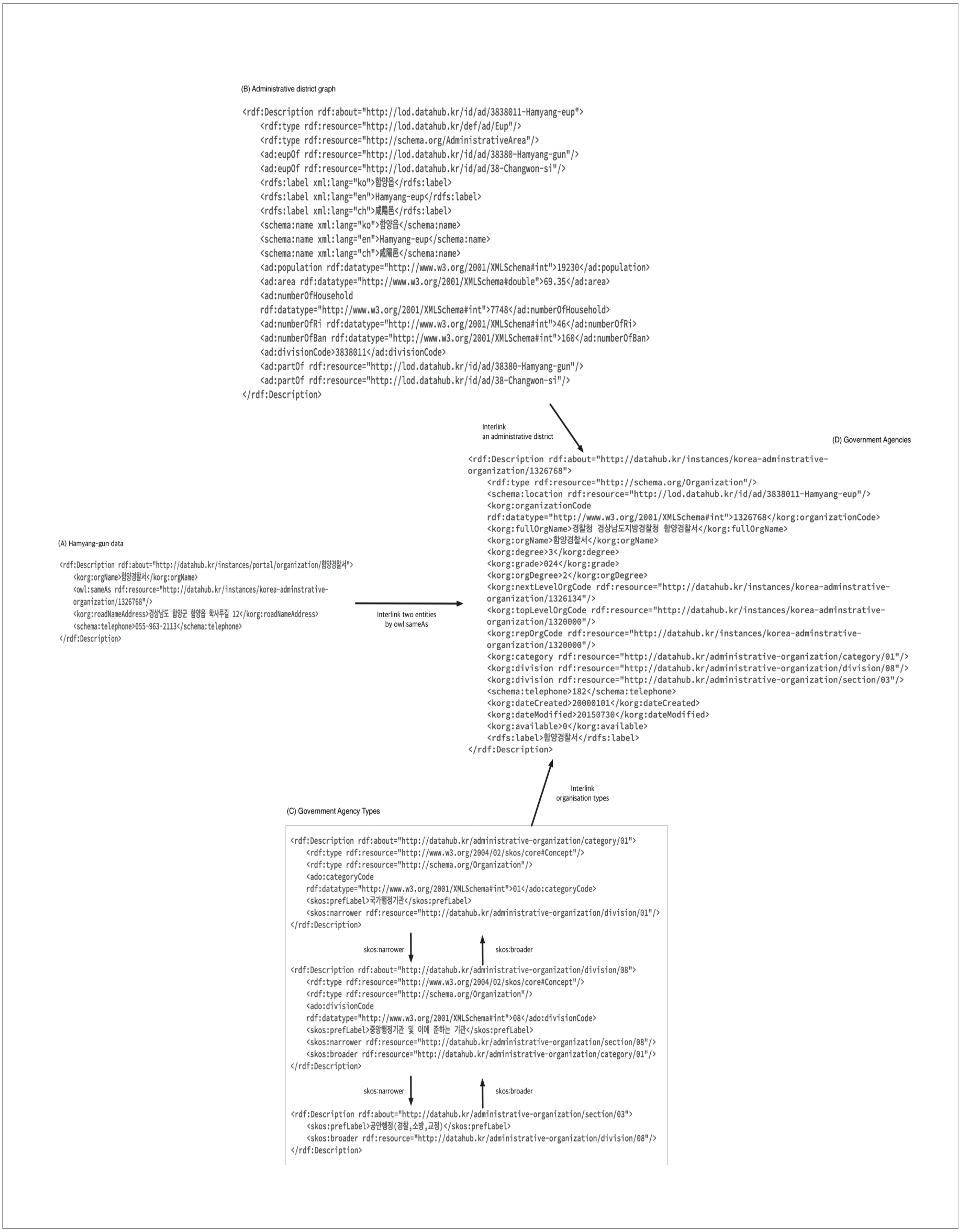
type_2 = MPVA_regional office in the administrative agency data. The administrative agency data were then mapped to type *classification_medium = 08* and type *classification_small = 01* in the administrative agency data and interlinked to "*http://datahub.kr/administrative-organization/division/0108*" and "*http://datahub.kr/administrative-organization/section/010801*" of the type classification graph. Although the frontline administrative agency data did not include information on the type field *classification_large*, the upper and lower relationships could be deduced because all types were defined by the *skos:broader/skos:narrower* relationship in the type classification graph. In addition, the data values were refined while the data were linked. For example, all spaces included in the telephone number field in the frontline administrative agency data were converted to "-" (298 cases), and fields without column names were modified to "*remarks*." In the case of telephone numbers, however, the representative numbers and regional numbers were included so that they could be used together in the graph.

- Because the data of *Hamyang-gun Gyeongsangnam-do* did not have the agency code field, identical data were identified by using the agency names as a reference. For the method of expressing the agency names, the whole agency names and lowest agency names were used in a mixed manner, and the classification of the orders of agencies through spaces and special characters differed from the administrative agency data. For example, "*Hamyang Register Office, Changwon Regional Court, Supreme Court*" and "*Hamyang and Sancheong Office, Yeongnam Branch, National Agricultural Products Quality Management Service, Ministry of Agriculture, Food and Rural Affairs*" in the administrative agency data were expressed as "*Hamyang Register Office, Changwon Regional Court*" and "*Hamyang and Sancheong Office, National Agricultural Products Quality Management Service*" in the data of *Hamyang-gun, Gyeongsangnam-do*. As a result, 17 out of 22 cases were identified as identical data by applying the whole agency name and lowest agency name information of the administrative district graph. Five inconsistent cases were fields that did not belong to the administrative agencies, such as "*KT Hamyang Branch*."

Applying the administrative district graph was found to be very effective at improving the quality of the open government data. The data quality can be improved directly by refining data values using standard data. In addition, data can be efficiently used without declaring redundant data values by semantically interlinking data. As shown in Figure 5, the data of *Hamyang-gun, Gyeongsangnam-do* could be interlinked with the administrative district graph to obtain information that was not available before. When "*Hamyang police station*" (i.e. the first value in Figure 5) was linked to the knowledge graph, new information such as the agency code, type, and postal code could be obtained. In other words, all of the information in the administrative agency graph could be linked, including "*agency code = 1326768*," "*whole agency name = Hamyang police station, Gyeongsangnam-do police department, National Police Agency*," "*highest agency code = 1320000*," "*postal code = 50041*," "*type classification_large = 01*," "*type classification_medium = 08*," and "*type classification_small = 03*." Furthermore, machines could automatically read the data and deduce their meaning, because all the data on the administrative districts and agencies were expressed through a semantic web language (Janowicz et al., 2014).

5. CONCLUSION

This paper proposes a knowledge model for administrative districts in South Korea and introduces a knowledge graph that interlinks administrative district data with relevant data. Although the use of open government data has been examined in various areas such as smart cities, autonomous vehicles, and artificial intelligence, their active use is still limited because of issues with data quality and linkage. The proposed knowledge model and knowledge graph effectively address such issues. First, an ontology model for the relationship between administrative districts as defined in the Local Autonomy Law was designed and the administrative district data published by the government were converted into a knowledge graph. Second, major open government data including the administrative district information were interlinked with the knowledge graph. The resulting administrative district knowledge graph can be used as a reference for improving the quality of open government data and semantically interlinking



(B) Administrative district graph

```xml
<rdf:Description rdf:about="http://lod.datahub.kr/id/ad/3838011-Hamyang-eup">
    <rdf:type rdf:resource="http://lod.datahub.kr/def/ad/Eup"/>
    <rdf:type rdf:resource="http://schema.org/AdministrativeArea"/>
    <ad:eupOf rdf:resource="http://lod.datahub.kr/id/ad/38380-Hamyang-gun"/>
    <ad:eupOf rdf:resource="http://lod.datahub.kr/id/ad/38-Changwon-si"/>
    <rdfs:label xml:lang="ko">함양읍</rdfs:label>
    <rdfs:label xml:lang="en">Hamyang-eup</rdfs:label>
    <rdfs:label xml:lang="ch">咸陽邑</rdfs:label>
    <schema:name xml:lang="ko">함양읍</schema:name>
    <schema:name xml:lang="en">Hamyang-eup</schema:name>
    <schema:name xml:lang="ch">咸陽邑</schema:name>
    <ad:population rdf:datatype="http://www.w3.org/2001/XMLSchema#int">19230</ad:population>
    <ad:area rdf:datatype="http://www.w3.org/2001/XMLSchema#double">69.35</ad:area>
    <ad:numberOfHousehold
    rdf:datatype="http://www.w3.org/2001/XMLSchema#int">7748</ad:numberOfHousehold>
    <ad:numberOfRi rdf:datatype="http://www.w3.org/2001/XMLSchema#int">46</ad:numberOfRi>
    <ad:numberOfBan rdf:datatype="http://www.w3.org/2001/XMLSchema#int">160</ad:numberOfBan>
    <ad:divisionCode>3838011</ad:divisionCode>
    <ad:partOf rdf:resource="http://lod.datahub.kr/id/ad/38380-Hamyang-gun"/>
    <ad:partOf rdf:resource="http://lod.datahub.kr/id/ad/38-Changwon-si"/>
</rdf:Description>
```

Interlink
an administrative district

(D) Government Agencies

```xml
<rdf:Description rdf:about="http://datahub.kr/instances/korea-adminstrative-
organization/1326768">
    <rdf:type rdf:resource="http://schema.org/Organization"/>
    <schema:location rdf:resource="http://lod.datahub.kr/id/ad/3838011-Hamyang-eup"/>
    <korg:organizationCode
    rdf:datatype="http://www.w3.org/2001/XMLSchema#int">1326768</korg:organizationCode>
    <korg:fullOrgName>경찰청 경상남도지방경찰청 함양경찰서</korg:fullOrgName>
    <korg:orgName>함양경찰서</korg:orgName>
    <korg:degree>3</korg:degree>
    <korg:grade>024</korg:grade>
    <korg:orgDegree>2</korg:orgDegree>
    <korg:nextLevelOrgCode rdf:resource="http://datahub.kr/instances/korea-adminstrative-
    organization/1326134"/>
    <korg:topLevelOrgCode rdf:resource="http://datahub.kr/instances/korea-adminstrative-
    organization/1320000"/>
    <korg:repOrgCode rdf:resource="http://datahub.kr/instances/korea-adminstrative-
    organization/1320000"/>
    <korg:category rdf:resource="http://datahub.kr/administrative-organization/category/01"/>
    <korg:division rdf:resource="http://datahub.kr/administrative-organization/division/08"/>
    <korg:division rdf:resource="http://datahub.kr/administrative-organization/section/03"/>
    <schema:telephone>182</schema:telephone>
    <korg:dateCreated>20000101</korg:dateCreated>
    <korg:dateModified>20150730</korg:dateModified>
    <korg:available>0</korg:available>
    <rdfs:label>함양경찰서</rdfs:label>
</rdf:Description>
```

(A) Hamyang-gun data

```xml
<rdf:Description rdf:about="http://datahub.kr/instances/portal/organization/함양경찰서">
    <korg:orgName>함양경찰서</korg:orgName>
    <owl:sameAs rdf:resource="http://datahub.kr/instances/korea-adminstrative-
    organization/1326768"/>
    <korg:roadNameAddress>경상남도 함양군 함양읍 학사루길 12</korg:roadNameAddress>
    <schema:telephone>055-963-2113</schema:telephone>
</rdf:Description>
```

Interlink two entities
by owl:sameAs

Interlink
organisation types

(C) Government Agency Types

```xml
<rdf:Description rdf:about="http://datahub.kr/administrative-organization/category/01">
    <rdf:type rdf:resource="http://www.w3.org/2004/02/skos/core#Concept"/>
    <rdf:type rdf:resource="http://schema.org/Organization"/>
    <ado:categoryCode
    rdf:datatype="http://www.w3.org/2001/XMLSchema#int">01</ado:categoryCode>
    <skos:prefLabel>국가행정기관</skos:prefLabel>
    <skos:narrower rdf:resource="http://datahub.kr/administrative-organization/division/01"/>
</rdf:Description>
```

skos:narrower
skos:broader

```xml
<rdf:Description rdf:about="http://datahub.kr/administrative-organization/division/08">
    <rdf:type rdf:resource="http://www.w3.org/2004/02/skos/core#Concept"/>
    <rdf:type rdf:resource="http://schema.org/Organization"/>
    <ado:divisionCode
    rdf:datatype="http://www.w3.org/2001/XMLSchema#int">08</ado:divisionCode>
    <skos:prefLabel>중앙행정기관 및 이에 준하는 기관</skos:prefLabel>
    <skos:narrower rdf:resource="http://datahub.kr/administrative-organization/section/08"/>
    <skos:broader rdf:resource="http://datahub.kr/administrative-organization/category/01"/>
</rdf:Description>
```

skos:narrower
skos:broader

```xml
<rdf:Description rdf:about="http://datahub.kr/administrative-organization/section/03">
    <skos:prefLabel>공안행정(경찰,소방,교정)</skos:prefLabel>
    <skos:broader rdf:resource="http://datahub.kr/administrative-organization/division/08"/>
</rdf:Description>
```

Fig. 5 Interlinking entities from heterogeneous data sources. (A) can be interlinked to (B) by the owl:sameAs properties, and the schema:location of (D) refers to an administrative location by using the administrative district graph of (B). In addition, (D) has government agency types from (C).

separately open data.

Studies on continuous expansion of the knowledge model and on data linkage are required, because open government data include subjects of various areas. In particular, studies on interlinking domains that are base data on a national level (e.g. national basic districts, postal codes, and spatial information) and closely related to administrative districts are required in the future. In addition, studies on practical applications of the knowledge graph interlinking different domains, such as question-and-answer services, visualization, and multidimensional analysis, are required in the future.

REFERENCES

Berners-Lee, T. (2006). Linked Data-World Wide Web design issues. Retrieved from https://www.w3.org/DesignIssues/LinkedData.html

Bizer, C., Cyganiak, R., & Heath, T. (2007). How to publish Linked Data on the Web? Retrieved from http://wifo5-03.informatik.uni-mannheim.de/bizer/HowtoPublishLinkedData.htm

Hall, W., & Pesenti, J. (2017). Growing the artificial intelligence industry in the UK. Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/652097/Growing_the_artificial_intelligence_industry_in_the_UK.pdf

Henrique S., Dantas, V., Furtado, V., Pinheiro, P., & McGuinness, D. L. (2017). From data to city indicators: A knowledge graph for supporting automatic generation of dashboards. *Proceedings of the 14th Extended Semantic Web Conference* (ESWC 2017), pp. 94-108.

Janowicz, K., Pascal, H., Benjamin, A., Kolas, D., & Vardeman, C. (2014). Five stars of linked data vocabulary use. *Semantic Web, 5*(3), 173-176.

Kang, W. (2017, July 25). Korea OECD leader in open government data. *Korea Times*. Retrieved from http://www.koreatimes.co.kr/www/tech/2017/08/133_233592.html

Kim, H. (2010). The semantic web and semantic connectivity: The starting point for web science. *Communications of the Korean Institute of Information Scientists and Engineers, 28*(3), 31-40.

Kim, H. (2017a). *Knowledge graph*. Seoul: Communication Books.

Kim, H. (2017b). Building knowledge graph of the Korea Administrative district for interlinking public open data. *Journal Of The Korea Contents Association, 17*(12), 1-11.

Matheus, R., Vaz, J., & Ribeiro, M. (2014). Open government data and the data usage for improvement of public services in the Rio de Janeiro City. In *Proceedings of the 8th International Conference on Theory and Practice of Electronic Governance* (ICEGOV '14) (pp. 338-341). New York: ACM.

Ministry of the Interior and Safety (2013). Act on promotion of the provision and use of public data. Retrieved from http://www.law.go.kr/eng/engLsSc.do?menuId=1&query=public+data&x=15&y=33

Ministry of the Interior and Safety (2016a). Public data open standard: Ministry of Public Administration and Home Affairs notice no. 2016-46. Retrieved from https://www.data.go.kr/information/PDS_0000000000000235/recsroom.do

Ministry of the Interior and Safety (2016c). Local Autonomy Act: Act No.14197, 29. May, 2016., amendment by other act. Retrieved from http://www.law.go.kr/eng/engLsSc.do?menuId=1&query=%EC%A7%80%EB%B0%A9%EC%9E%90%EC%B9%98%EB%B2%95&x=0&y=0#liBgcolor13

Ministry of the Interior and Safety (2017). Public data, open standards guide: MOGAHA notice no. 2016-46. Retrieved from https://www.data.go.kr/information/PDS_0000000000000255/recsroom.do

National Information Society Agency (2016b). Open government data quality management manual. Retrieved from https://www.data.go.kr/information/PDS_0000000000000404/recsroom.do

National Science and Technology Council (2016). Preparing for the future of artificial intelligence, technical report, United States. Executive office of the president and M. Holden, J. P. Smith., Washington D.C. 20502, October 2016.

Open Knowledge International, ed. (2017). *Open data handbook, open knowledge*. Retrieved from http://opendatahandbook.org/

Safarov, I., Meijer, A., & Grimmelikhuijsen, S. (2017). Utilization of open government data: A systematic literature review of types, conditions, effects and users. *Information Polity, 22*(1), 1-24.

Seo, H. (2017a). An empirical study on open government data: Focusing on ODB and OUR index. *In-*

formation Policy, 24(1), 48-78.

Seo. H. (2017b). A critical review on open, useful, reusable government data index by OECD with level of domestic open government data: Focusing on comparison with Open Data Barometer. Information Policy, 24(2), 43-67.

Tran, E., & Scholtes, G. (2015). Open data literature review. Retrieved from https://www.law.berkeley.edu/wp-content/uploads/2015/04/Final_OpenDataLitReview_2015-04-14_1.1.pdf

Web Foundation (2017). The open data barometer 4th ed., web foundation. Retrieved from https://webfoundation.org/2017/05/open-data-barometer-4th-edition-coming-may-23/.