

## Design of Open Knowledge Platform Based On Knowledge Base Utilization Model And Service Scenario To Support Solutions Of Regional Issues

**ChulSu Lim, Hak-Su Park, Hyoung-Seop Shim, Ji-Sun Kang, and Jiseong Son** \*

Researcher, Korea Institute of Science and Technology Information, 34141, South Korea

\*Corresponding author: mobile phone: +82-010-6385-0724; Email address: jsson@kisti.re.kr

**Article History:** Received:11 november 2020; Accepted: 27 December 2020; Published online: 05 April 2021

**Abstract:** Open knowledge platform can provide a purified knowledge base. Thus, we build a platform for several application areas in a cloud computing that supports APIs for various data based on a knowledge utilization model. The goal of this platform is to maximize the utilization of the knowledge base. In order to achieve this goal, we designed the structure of this platform as an open knowledge platform. The targets of the design are to maximize the utilization of data linkage, to expand it to national common knowledge and to increase its usability by providing services with knowledge graphs. In order to design the platform we identified users, information sources, and infrastructures. In the process, we found it is crucial to specify roles and services to the users of the platform. The requirements are induced from a utilization model and scenario of the service based on the knowledge graph. With the service scenario, stakeholders of the platform started narrow down function modules needed to support the service. One of the modules is a national common knowledge in the knowledge base, which provide essential connected knowledge to support solving regional problems of government such as earthquake, flooding. To increase the usability of data scattered by departments and agencies, data linkage, and knowledge between fragmented data sets is included in this platform. Subsequently, we designed modules to support the effective utilization of this knowledge information. Also, we found that a cloud infrastructure instead of in-house hardware and software could provide flexible and compatible services for the platform. Moreover, the cloud system has advantages on big data analysis and distributed system interconnection. Utilization model and scenario-based process modeling provide a systematic approach to design an open knowledge platform that supports many required components enabling interoperability, compatibility, and connectivity among other knowledge bases..

**Keywords:** Linked open data, Knowledge graph, Open platform, Data driven solution, Triple

### 1. Introduction

Korea has secured the world's highest level of public data openness by ranking first in the world in the OECD Public Data Opening Index recently, but the actual use and value creation of data is insufficient. And, due to social and environmental changes such as climate change, urbanization, industrialization, and population density increase, large-scale disasters frequently occur in which various risk factors are interconnected. Thus needs are increasing to use these open data of the government to confront and prepare for the disasters. Due to the large-scale, diversification, and complexity of disasters and disasters, national resources have increased interest in disaster disasters, and as a countermeasure for pre-existing national safety, various disaster-related organizations have been established and related systems have been established. However, in order to increase the usability of data scattered by departments and agencies, data linkage and knowledge between fragmented data sets is required, and a knowledge data platform is required for effective utilization of this knowledge information. In Korea, there has been continuous R & D investment in linked data, but it is limited to specific fields and fails to produce visible results. On the other hand, major companies that provide intelligent services such as Google, Apple, and IBM are building intelligent data in the form of knowledge graphs[1]. They are built as knowledge graphs based on encyclopedia data such as Wikipedia, and provided in a format that can be reused in various domains [2-4]. In this paper, we discuss the design method of an open knowledge platform that can utilize knowledge data that connects various data to support data-based decision-making for solving community problems, and detailed functional elements that the platform must support. Proposed platform supports application programs, and refers to an ecosystem in which multiple groups of users, customers, and partners participate, and each group reasonably exchanges values.

### 2. Materials and Methods

In order to design a platform to enable users to utilize required functions on this platform, we need to identify users and applications. The users and applications could specify services and functions they use. The services and functions could designate or in favor of internal structure and underlying infrastructure of the platform. Based on users and application analysis, we defined technology architecture for the platform. There are many knowledge bases which provide structured information. Main sources of the information are derived from Wikipedia, WordNet and Geonames[5]. YAGO enriched knowledge from Wikipedia info box for entity, WordNet for word and Geonames for location information[6]. In addition to the triple information(subject,

\*Corresponding author: Jiseong Son

Researcher, Korea Institute of Science and Technology Information, 34141, South Korea

Email address: jsson@kisti.re.kr

predicate and object), there are time, location and context information in YAGO[7]. Korean DBpedia was built from English DBpedia in which the triples are extracted from English Wikipedia[8]. The subject-object-relation triples are a basis of the knowledge base or Linked Open Data (LOD). In case where there are no identifiers, the subjects and objects need to be identified to make relations between them. Entity recognition and disambiguation technologies could be applied in this step. After the entity recognition, we need to connect these entities with proper relation defined in ontology[9-11]. The relations on the ontology need to be fine tuned to each application and problem domain[12, 13].

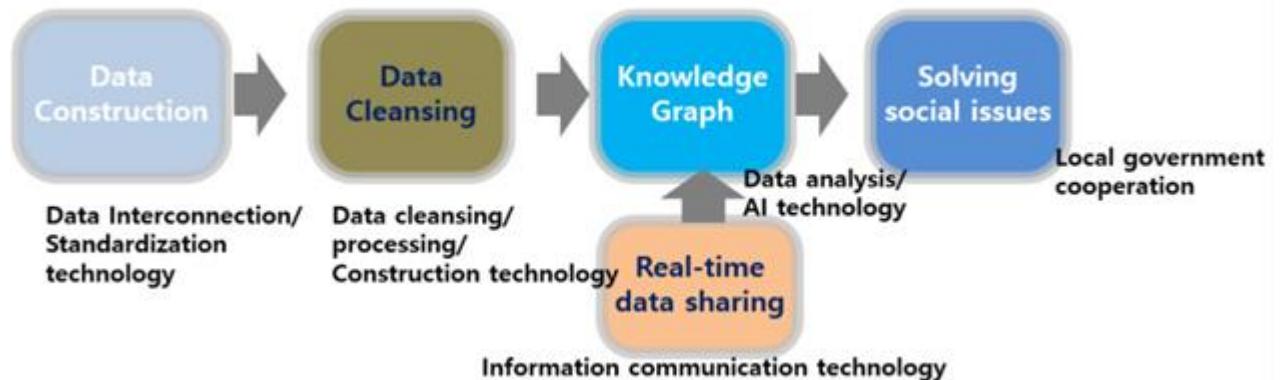
The steps to design the open platform are as follows:

1. Define users and applications
2. Define a technology architecture of the system
3. Design of utilization model and service scenario of the knowledge base
4. Identify modules and functions to support the service
5. Identify information sources and data processing plans to construct a knowledge base using data interconnection
6. Define components of the infrastructure of the platform considering HW, SW, networking, Storage, security

### 2.1 A technology architecture for Data Driven Solutions (DDS) system

To develop a data driven solution system, we need to setup a technology architecture which will guide overall steps for each modules of the system.

Figure 1 shows technology architecture for DDS system. The first technology involved in this platform is data interconnection and standardization to collect data from various information sources. The following step is the data processing and construction technology to process the collected data which include data cleansing, mapping and transformation. One unique application of this platform is real-time data utilization based on Named Data Networking(NDN) for Internet of Things(IoT) generated data contents. Connected data and real-time data are merged into a knowledge graph in which each data is identified as a subject or object with a relation. The relations are predefined for this domain as ontology. This knowledge graph could be the basis data for the social problem solving module. Data visualization, searching, navigation are the major services of this module.



**Figure 1** Technology architecture for DDS

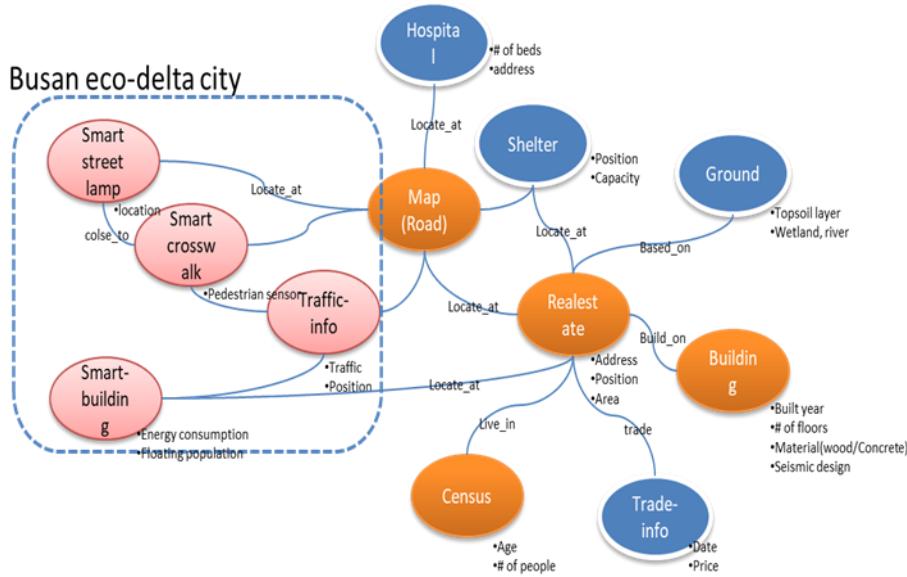
## 3. Results and Discussion

### 3.1 Knowledge graph utilization models and scenario

In order to identify and get concrete concept of functions and modules for the knowledge platform we developed a utilization model and scenario of the platform as shown in Figure 2.

#### 3.1.1 Knowledge graph utilization model

Location datasets(map, real-estate) plays major role in this model. Other datasets(hospital, shelter, ground, trade-info) are connected to the major datasets using appropriate relations like locate\_at, live\_in, build\_on. And a region specific datasets which are collected from a smart city including smart street lamp, real-time traffic information is connected to a map dataset.

**Figure 2** Knowledge graph utilization model and scenario

### 3.1.2 Service scenario and information interconnection

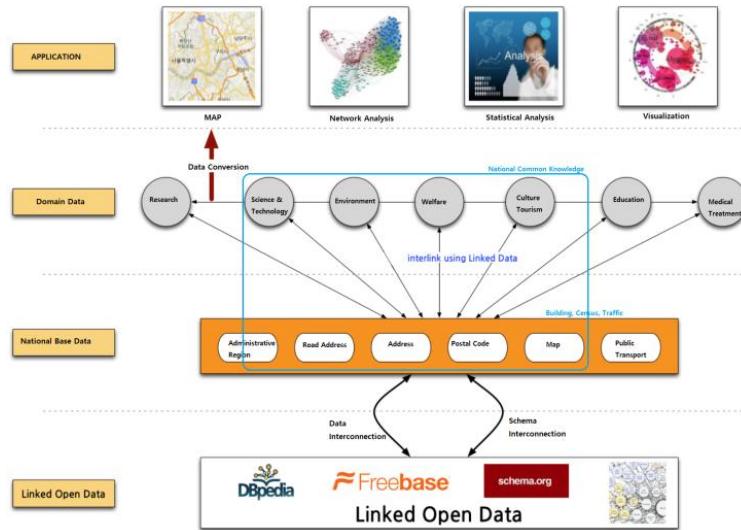
**Table 1** Utilizing interconnected information in a knowledge graph for each scenarios

Scenario	dataset involved	details of information interconnection	comment
Estimation of damage to buildings based on actual transaction price when a magnitude 5.4 earthquake occurs	building, real trade transaction	location-> building(year, material, seismic design) -> real trade information(date, price)	2 datasets
Find a three-story earthquake-prone wooden house built in a wetland with a large elderly population.	building, population, ground	population(age, number) -> building(wooden/concrete) -> ground(wetland)	3 datasets
Find hospitals which have more than 100 beds within 20 minutes by car using real-time traffic information?	hospital, real-time traffic information	Hospital(# of beds) -> real-time traffic(road traffic)	2 datasets(include IoT real-time information)

With this scenario, we could confirm the usefulness of interconnected information represented in a knowledge graph and checked core knowledge as we call ‘national common knowledge’ as shown in Figure 3(based on LOD data hub([lod.datahub.kr](http://lod.datahub.kr))).

### 3.2 A national common knowledge

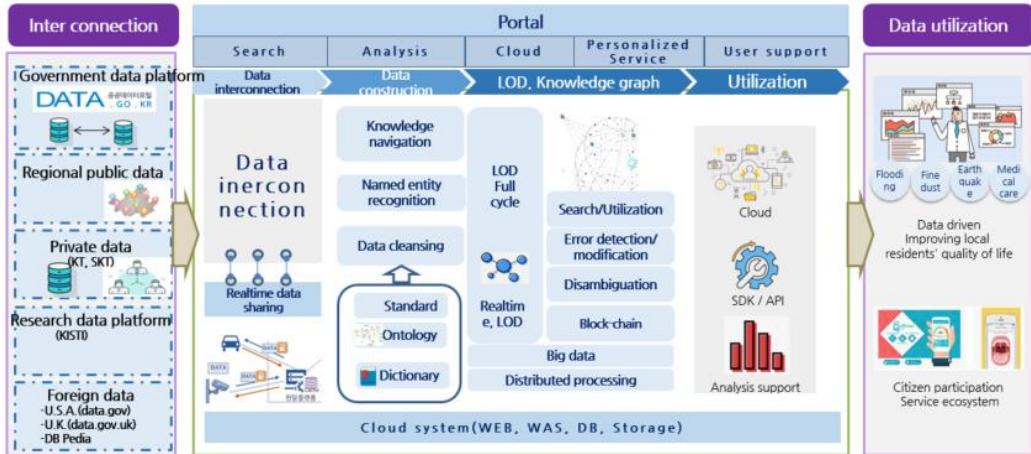
Components of the national common knowledge are interlinked with widely used knowledge bases such as DBpedia, Freebase. The schema used in the national base data and national common knowledge should be compatible with other knowledge bases. Thus the schema should be compatible with schema.org. In addition to the ‘Address’, ‘Postal code’ information in the national base data, ‘Building’, ‘Census’, ‘Traffic’ information need to be included to the national common knowledge. Furthermore, some domain specific data such as ‘safety shelter’, ‘safety facilities’ data to support solving social problems should be included as shown in the diagram.



**Figure 3** A structure of national common knowledge

### 3.3 A design of the knowledge platform

Information sources such as government data platform, regional public data are interconnected to the platform. The collected data from the information sources need to be refined with a standard. The refined data is checked for similar existing data for possible interconnection. A connection is made between the two data items if they refer to the same object. We could identify the objects with named entity recognition module. The connection could have relations such as ‘same-as’, ‘kind-of’, ‘belongs-to’[12, 13]. Dictionary is constructed for the platform to assist standardization, cleansing and named entity recognition. Components like named entity recognition, ontology, dictionary are included as a part of data construction and linked open data, graph searching, and block-chain technology are included as knowledge processing steps as shown in Figure 4. This figure depicts all the platform modules to support functions of the DDS system as a result.



**Figure 4** An open knowledge platform components

## 4. Conclusion

In this paper, we discuss the design method of an open knowledge platform that can utilize knowledge data that connects various data to support data-based decision-making for solving community problems, and detailed functional elements that the platform must support. In order to design the platform we identified users, information sources and infrastructures. A utilization model and scenario are defined to induce services and modules which users the knowledge graph. There are many function modules to support the service. One of the modules is a national common knowledge in the knowledge base, which provide essential connected knowledge to support solving regional problems of government such as earthquake, flooding. To increase the usability of data scattered by departments and agencies, data linkage and knowledge between fragmented data sets is included in this platform. Subsequently we designed modules to support effective utilization of this knowledge information. Also, we found that a cloud infrastructure instead of in-house hardware and software could provide flexible and compatible services for the platform. Moreover, the cloud system has advantages on big data analysis and distributed system interconnection. Utilization model and scenario based service process modeling

provide a systematic approach to design an open knowledge platform which supports many required components enabling interoperability, compatibility, and connectivity among other knowledge bases.

## 5. Acknowledgment

This research was conducted with the support of Data Driven Solutions (DDS) Convergence Research Program funded by the National Research Council of Science and Technology "Development of solutions for region issues based on public data using AI technology "(1711101951).

## 6. References

1. Xin Luna Dong, Evgeniy Gabrilovich, Jeremy Heitz, Wilko Horn, Ni Lao, Kevin Murphy, Thomas Strohmann, Shaohua Sun, and Wei Zhang. (2014). Knowledge Vault: A Web-scale approach to probabilistic knowledge fusion. In SIGKDD.
2. Eun-kyung Kim, Matthias Weidl, Key-Sun Choi, Soren Auer. (2010). Towards a Korean DBpedia and an Approach for Complementing the Korean Wikipedia based on DBpedia, Proceedings of the 5th Open Knowledge Conference K. Bollacker, C. Evans, P. Paritosh, T. Sturge, and J. Taylor. (2008). "Freebase: A collaboratively created graph database for structuring human knowledge." In Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data, pages 1247–125.
3. Lehmann Jens, Isele Robert, Jakob Max, Jentzsch Anja, Kontokostas Dimitris, Mendes Pablo, Hellmann Sebastian, Morse Mohamed, Van Kleef Patrick, Auer Sören, Bizer Christian. (2014). DBpedia - A Large-scale, Multilingual Knowledge Base Extracted from Wikipedia. Semantic Web Journal. 6. 10.3233/SW-140134.
4. Fellbaum, Christiane. WordNet: An electronic lexical database (Language, Speech, and Communication). Cambridge, MA: The MIT Press; 1998.
5. Rebele Thomas, Suchanek Fabian, Hoffart Johannes, Biega Joanna, Kuzey Erdal, Weikum Gerhard. (2016 ). YAGO: A Multilingual Knowledge Base from Wikipedia, Wordnet, and Geonames. 177-185. 10.1007/9 78-3-319-46547-0\_19.
6. Hoffart Johannes, Suchanek Fabian, Berberich Klaus, Lewis-Kelham Edwin, de Melo Gerard, Weikum Gerhard. (2011). YAGO2: Exploring and Querying World Knowledge in Time, Space, Context, and Many Languages. Proceedings of the 20th International Conference Companion on World Wide Web, WWW 2011. 229-232. 10.1145/1963192.1963296.
7. Kim Eun-Kyung, Choi Donghyun, Lee Jihye, Ahn Jinhyun, Choi KeySun. (2020). An Approach for Supplementing the Korean Wikipedia based on DBpedia.
8. Gupta Rahul, Halevy Alon, Wang Xuezhi, Whang Steven, Wu Fei. (2014). Bipedia: An ontology for search applications. Proceedings of the VLDB Endowment. 7. 505-516. 10.14778/2732286.2732288.
9. Fader Anthony, Soderland Stephen, Etzioni Oren. (2011). Identifying Relations for Open Information Extraction. EMNLP2011Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference. 1535-1545.
10. Wick Michael, Singh Sameer, Pandya Harshal, McCallum Andrew. (2013). A joint model for discovering and linking entities. AKBC 2013 Proceedings of the 2013 Workshop on Automated Knowledge Base Construction, Co-located with CIKM 2013. 67-72. 10.1145/2509558.2509570.
11. Jain Sarika, Mehla Sonia, Mishra Tiwari Sanju. (2016). An ontology of Natural Disasters with exceptions. 232-237. 10.1109/SYSMART.2016.7894526.
12. Noy N., McGuinness Deborah. (2001). Ontology Development 101: A Guide to Creating Your First Ontology. Knowledge Systems Laboratory. 32.