Gathering Energy-Related Datasets of Public Buildings on a Nationwide Scale in South Korea

D.W. Kim[1,*], J.Y. Lim[2], and S.E. Lee[2]

[1]Convergence Research Planning Division, Korea Institute of Civil Engineering and Building Technology, Goyangsi 411-712, South Korea
[2]Building and Urban Research Institute, Korea Institute of Civil Engineering and Building Technology, Goyangsi, 411-712, South Korea

ABSTRACT
The government has collected a large set of energy-related data via online information systems over the last decade. This valuable information has been stored in large quantities; however, it is difficult for the public to access a necessary dataset because intragovernmental relationships are entangled. Different government departments separately manage the building energy datasets, which hinders data collection and thorough analyses. In this paper, an integrated data-gathering framework for the public building stock is introduced, while considering government agency coordination. The pilot database was constructed and its future application is discussed.

KEYWORDS
Energy data gathering, Energy epidemiology, Public building, Energy performance assessment

INTRODUCTION
Currently, about 6,800,000 buildings exist in South Korea, and the total energy consumption of the stock is about 20.5% of the total primary energy (MOTIE, 2015). For this reason, it is important to reduce and manage greenhouse gas emissions. To handle this, a large set of energy-related data has been collected via the government's online systems (EAIS 2016, GIR 2016, GreenTogether 2016) over the last decade. The primary targets are public buildings, mostly due to their political manageability.
However, the current online system still plays a partial role in data collection, given its regular release of simple statistical data (energy consumption) on a nationwide scale. Whole datasets stored in the systems are very important for evidence-based analysis and decision making; hence, the public needs to access this information. Different government departments separately manage the energy datasets, which hinders the effective collection of data, and the subsequent diagnosis of various issues, on a nationwide scale. The entangled intragovernmental relationship is currently slowing new developments. This situation has led to the circumstances in which retrofit and

* Corresponding author's email: deukwookim@kict.re.kr

remodeling solutions are conventionally similar, such as the replacement of obsolescent HVAC systems and LED lighting.

The current system is not appropriate for glancing over and diagnosing the energy efficiency of public buildings in less time. It is necessary to coordinate with the different government agencies to smoothly gather various datasets, and to develop more convenient and quicker methods to diagnose many buildings macroscopically and nationally based on big data. In this paper, a cooperative framework is introduced on how to gather energy-related datasets for public buildings on a nationwide scale, and future application of the gathered datasets is discussed.

DATA-GATHERING FRAMEWORK

Figure 1 shows the data-gathering framework and the general management bodies. Each dataset is managed by different government departments and agencies (e.g., the Korea Energy Agency, Korea Appraisal Board, National Geographic Information Institute, Korea Meteorological Administration, and Korea Land and Housing Corporation). Building ledgers, weather data, and information about a building's shape have been released to the public, but for any number of reasons other information has yet to be released or it has only been partially released. As such, one role of government-funded research institutes is their importance in persuading and enlightening each party to gather datasets smoothly and efficiently. Finally, the outcome of the integrated database will provide a fact hitherto unknown and it enhances the current modeling strategies and validation processes.

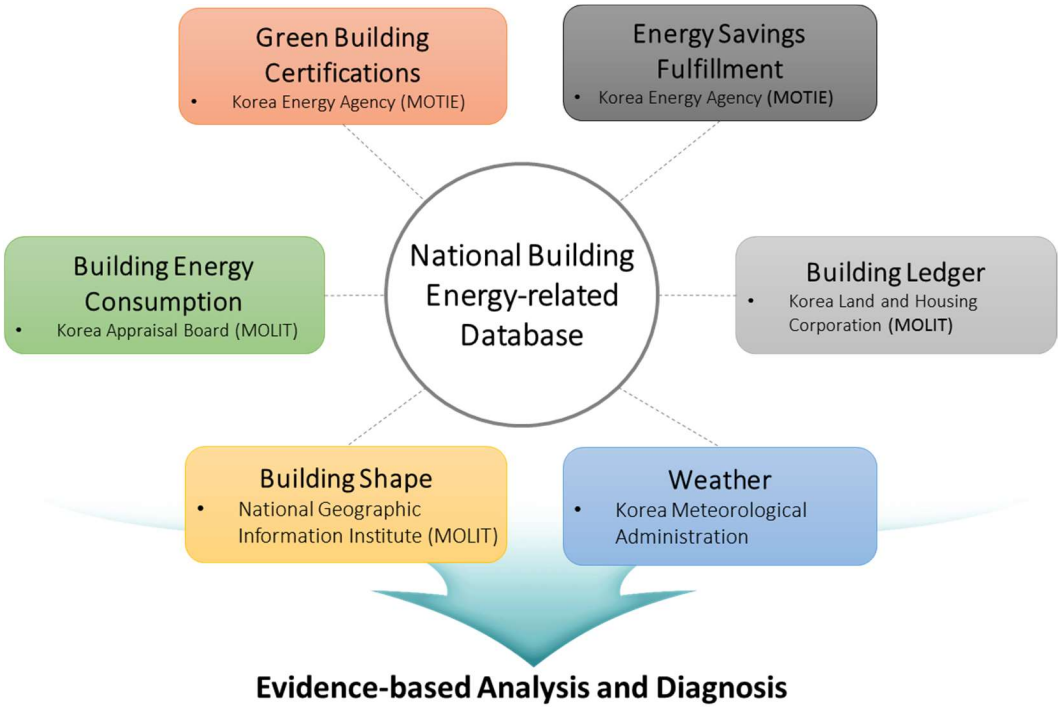

***Figure 1**. Framework for constructing the national building energy-related database with various government departments and agencies (MOTIE: Ministry of Trade, Industry and Energy, MOLIT: Ministry of Land, Infrastructure, and Transport)*

Table 1 provides an example of the gathered datasets for Seoul. The datasets consist of eight tables pertaining to the elements in Figure 1. The rows indicate the number of

samples, and the columns present the number of attributes. Each table provides valuable information to diagnose issues associated with building energy efficiency and compiling energy statistics. Based on this database, researchers can develop more practical/objective diagnostic models and simulation validation processes, or one can glance over a building stock's energy efficiency in less time on a nationwide scale.
As shown in Figure 1, it is necessary to coordinate with different government agencies to gather whole datasets. Currently, a working team is attending a conference with the Korea Energy Agency and the Korea Appraisal Board, and the datasets pertaining to green building certifications, energy savings fulfillment, and building energy consumption will be gathered in the near future for the entire region.

Table 1. *The gathered datasets (target area: Seoul)*

Table name	*No. of Rows*	*No. of columns*	*Main contents*
Building ledger I (site based)	5,748	26	Address, building name, completion date, building area, gross floor area, site area, parking accumulation, number of elevators, building height, number of stories, principal use, etc.
Building ledger II (building based)	493,998	39	
Building energy	477,445	26	Monthly electricity and gas consumption data with the site address
Green building certification	27	84	Information on U-value, SHGC of envelopes, type and capacity of heating/cooling/domestic hot water/lighting/renewable energy system
Building shape attributes	653,126	5	Polygon building shape (x–y coordinates)
Weather data	60	21	Weather station location, outdoor temperature (min/max/average), solar radiation, wind speed and direction, relative humidity, dew point temperature, cloud cover, etc.
Energy saving fulfillment	-	-	Results of energy performance measurement by the Energy Saving Company (in contemplation)
Post-occupancy evaluation	-	-	Survey of occupants' comfort, etc. (in contemplation)

DATA LINKING

Data linking (or data record matching) brings together information from intragovernmental databases. Linking building energy-related data (Table 1) from various sources can tell a bigger story than analyzing narrow-scope data. Hence, data linking can really make a difference to the current paradigm of energy performance assessment. This concept had already been initiated under the Energy Epidemiology approach (Hamilton et al 2013). As far as the authors know, energy consumption and its relevant data have not yet been linked on a nationwide scale in any country. In other words, this task is not easy to undertake, and many sensitive issues are involved, as mentioned above.
In this study, a shadow database system was constructed for local users, which covers only a partial area (Seoul). The aim is to test various basic functions such as querying

necessary information and visualizing the correlation between energy and its relevant data. At the present time, this approach supports 'table join' by primary key (PK) based on the address. The second aim is to determine the common patterns in the various energy consumption profiles and to classify them, in order to diagnose whether a given pattern is normal or not.

CONCLUSIONS AND FUTURE WORK

The government has collected a large set of energy-related data via online information systems over the last decade. This valuable information has been stored in large quantities; however, it has not been actively used due to entangled intragovernmental relationships and data accessibility issues faced by the public.

The research team has engaged in several efforts to handle this problem: (1) it has designed a data-gathering framework in consideration of the various government agencies; (2) it has constructed an integrated database; and (3) it has utilized energy-related data linking. The outcomes of this work will provide policymakers with enhanced evidence on how policies and technologies work, or do not work, in practice, and why (Hamilton et al. 2013). Feedback on the actual performance of the technologies will also support learning within this industry and academic area.

The current database system is temporary, and the following are now being investigated:

- Energy normalization: many factors influence building energy consumption (e.g., gross floor area, weather, occupants, system efficiency, etc). To determine any patterns and to compare them objectively, a normalization technique is important. Various methods (e.g., a standard score, quantile normalization, and rescaling) are now being studied.
- Outlier detection: After the datasets are properly normalized, outliers can be effectively detected. There are various detection algorithms (Kriegel et al 2010), and appropriate methods should be widely reviewed.
- Diagnosis algorithm: all energy-related data (Figure 1, Table 1) will be used as important materials to develop diagnostic algorithms. These will classify energy consumption patterns into several categories (good or not) on a nationwide scale.

ACKNOWLEDGMENTS

This work was supported by the Korea Institute of Energy Technology Evaluation and Planning (KETEP) and the Ministry of Trade, Industry & Energy (MOTIE) of the Republic of Korea (No. 2015210103180).

REFERENCES

EAIS. 2016, Electronic Architectural administration Information System, Ministry of Land, Infrastructure, and Transport (MOLIT), http://www.eais.go.kr (accessed 2016)

GIR. 2016. Greenhouse Gas Inventory & Research Center of Korea, Office for Government Policy Coordination, http://www.gir.go.kr (accessed 2016)

GreenTogether. 2016. Green architectural Information Portal, Ministry of Land, Infrastructure, and Transport (MOLIT), http://www.greentogether.go.kr (accessed 2016)

Hamilton I.G., Summerfield, A.J., Lowe, R., Ruyssevelt, P., Elwell C.A. and Oreszczyn, T. 2013. Energy epidemiology: a new approach to end-use energy demand research, *Building Research Information*, 41(4), 482-497

Kriegel, H. P., Kröger, P., & Zimek, A. 2010. Outlier detection techniques. *In Tutorial at the 16th ACM international conference on knowledge discovery and data Mining (SIGKDD)*, Washington, DC.

MOTIE. 2015, Energy Consumption Survey 2014 compiled by Korea Energy Economics Institute, Ministry of Trade, Industry and Energy (MOTIE)