



Journal of Real Estate Analysis

<http://www.kab.re.kr>

November 2017, Vol.3, No.2, pp.57~76

국가건물에너지 통합DB의 데이터 신뢰성 검증 방법 개발 및 적용*

The Development and Application of Method for Validating the Data Reliability of National Building Energy Integrated Database

지창윤** · 최민석*** · 신성은**** · 오승엽*****

Ji, Chang Yoon · Choi, Min Seok · Shin, Sung Eun · Oh, Seung Youb

Abstract

The National Building Energy Integrated Database (NBEIDB) is used for determining buildings with high or low energy performance or analyzing the effectiveness of policies for reducing building energy consumption. To secure the reliability of analysis results based on NBEIDB and to use the results in policy making, it is very important to ensure data reliability in NBEIDB. This study presents a data validation method consisting of four steps. Steps 1-2 validate the accuracy of collected and integrated data, based on the error rate. Step 3 shows the completeness of data in NBEIDB in terms of matching rate. Finally, Step 4 shows data validity based on valid rate. As a result of case analysis using January 2015 data, there was no error found in the processes of data collection and integration (Steps 1 and 2). Finally, the valid rates for building, electricity, city gas, and district heating data were calculated as 89.6%, 76.6%, 91.7%, and 92.3%, respectively. The validation method can show not only information for ensuring data reliability (error rate, matching rate, and valid rate), but also improvements required to be made in the processes of data analysis and integration. It is expected that the method for validation proposed in this study will contribute to securing the reliability of NBEIDB and utilizing the integrated database.

Key Words : National Building Energy Integrated Database, Data reliability, Accuracy, Validity, Data verification

* 본 논문은 국토교통부 도시건축연구사업의 연구비지원(과제번호 : 17AUDP-C127876-01)에 의해 수행되었습니다.

** 한국감정원 녹색건축센터 부연구위원 (주저자 / 교신저자, k05765@kab.co.kr)

*** 한국감정원 녹색건축센터 부연구위원 (공동저자, k05395@kab.co.kr)

**** 한국감정원 녹색건축센터 책임연구원 (공동저자, k25843@kab.co.kr)

***** 한국감정원 녹색건축센터 차장 (공동저자, k25854@kab.co.kr)

I. 서론

대한민국은 대표적인 에너지 다소비 국가로 급부상하고 있다. 실제로, 2016년 기준 대한민국의 온실가스 배출량은 세계 8위에 달하며, 그 증가 속도는 OECD 국가의 2배 수준에 달한다. 이러한 이유로, 대한민국 정부는 2015년 파리협약을 통해 2030년 BAU 대비 37% 감축이라는 국가 온실가스 감축목표를 발표하였다. 그리고 이러한 감축목표의 달성을 위한 다양한 정책을 수립·시행하고 있다. 예를 들어, 온실가스 다배출 업체를 관리하여 온실가스 배출량을 줄이는 방안으로써 “온실가스·에너지 목표관리제”가 시행되고 있으며, 2015년부터는 보다 적극적인 시장의 참여를 유도하기 위하여 “배출권거래제”가 시행되고 있다.

건물은 우리나라 전체의 최종 에너지사용량 중 약 26%를 차지할 뿐만 아니라(에너지경제연구원, 2017) 타 분야에 비해 높은 온실가스 감축 잠재량을 갖기 때문에, 건물부문에 대한 온실가스 및 에너지사용량 감축정책이 적극적으로 시행되고 있다. 예를 들어, 건물의 단열성능 기준은 지속적으로 강화되고 있으며, 신재생에너지 설비나 차양의 설치 의무 기준 역시 강화되고 있다. 또한, 건물의 에너지성능 정보를 공개하는 등의 정책을 함께 시행함으로써, 신축 건물의 에너지사용량 저감을 유도하고자 한다. 기존 건물의 경우에는 앞서 언급한 목표관리제나 배출권거래제를 통해 관리하거나 건물의 에너지성능을 향상할 수 있도록 다양한 지원사업을 시행하고 있다. 그리고 개별 건물에 대한 에너지소

비량을 관리함과 동시에, 지역단위에서의 에너지소비총량을 제한·관리하도록 법령에 규정하기도 한다(국토교통부, 2017).

이러한 정책이 대중의 신뢰를 얻고 감축 효과가 극대화되기 위해서는, 정책의 타당성을 설명할 수 있는 명확한 근거자료의 제시가 필수적이다. 즉, “어느 지역, 어느 용도, 언제 준공된 어떤 규모의 건물의 에너지성능이 상대적으로 열등한가?”, “정부가 추진하는 정책의 감축효과는 기대에 부합하는가?”와 같은 물음에 대한 답이 제시되어야 한다.

이러한 물음에 대한 답은 건물의 에너지사용량에 대한 데이터 분석을 통해 확인될 수 있으며, 많은 국가에서는 이러한 분석을 위한 건물의 에너지사용량 데이터를 확보하고자 노력하고 있다. 미국의 Department of Energy의 지원을 받아 구축된 Building Performance Database(BPD)는 약 750,000동의 건물(주거용 및 상업용)에 대한 에너지사용량 데이터를 포함하고 있다(Mathew et al., 2015). 스페인의 Catalan Institute of Energy는 스페인에서 에너지성능인증(Energy Performance Certificates; EPCs)을 받은 건물의 약 20%에 해당하는 129,635동의 건물에 대한 데이터를 확보하였다(Gangoellis et al., 2106). 그리스에서 역시 EPCs DB가 구축되고 있으며, 2015년 8월까지 약 650,000건의 EPCs가 구축되었다(Droutsa et al., 2016). 대한민국 정부는 전국의 모든 건물에 대한 에너지사용량 데이터의 구축을 목표로, 「녹색건축물 조성지원법」 제10조에 따라 “국가건물에너지 통합DB”를 구축하였다. 국가건물에너지 통합DB는 개별 건물에 대한 전기, 도시가스, 지

역난방 에너지소비량 정보와 함께 건축물대장에 있는 모든 건물 정보를 포함하고 있으며, 2016년 12월까지 약 759만동의 건물과 3,958만개의 에너지수용가에 대한 데이터를 수집·구축한 상태이다.

이처럼 구축된 데이터는 다양한 연구에 활용되고 있다. 기존 건물 중에서 에너지 성능이 우수하거나 열등한 건물을 규정하거나, 건물의 에너지사용량 감축을 위한 정책의 효과 검증, 미래의 건물부문 에너지사용량 및 온실가스 배출량 예측 등은 대표적인 연구주제들이다(양시원·김선숙, 2014; 조상규·김영현, 2013; 최민석 외, 2017; Jeong et al., 2016; Jeong et al., 2018). 이러한 연구의 결과가 신뢰성을 확보하고 정책의 수립에 활용되기 위해서는 연구의 기본이 되는 통합DB의 신뢰성 확보가 무엇보다도 중요하다. 이에 본 연구에서는 통합DB의 구축 과정에 적용할 수 있는 데이터의 검증방법을 제안함으로써, 통합DB의 신뢰성 확보와 활용성 증대에 기여하고자 한다.

본 논문은 다음의 순서로 구성된다. 먼저, 2장에서는 국가건물에너지 통합DB의 구축 방법 및 현황을 파악하고 데이터의 신뢰성 검토와 관련한 기존 문헌을 검토하여, 데이터 신뢰성 검증을 위한 필요사항을 도출한다. 3장에서는 2장에서 도출된 필요사항을 모두 만족시키는 검증방법을 제시한다. 4장에서는 제시된 검증방법에 통합DB에 구축된 에너지사용량 및 건물 데이터를 적용하여, 검증방법의 타당성을 확인하고 현재 구축된 데이터의 오차율 및 유효율을 제시한다.

II. 기존 문헌 검토

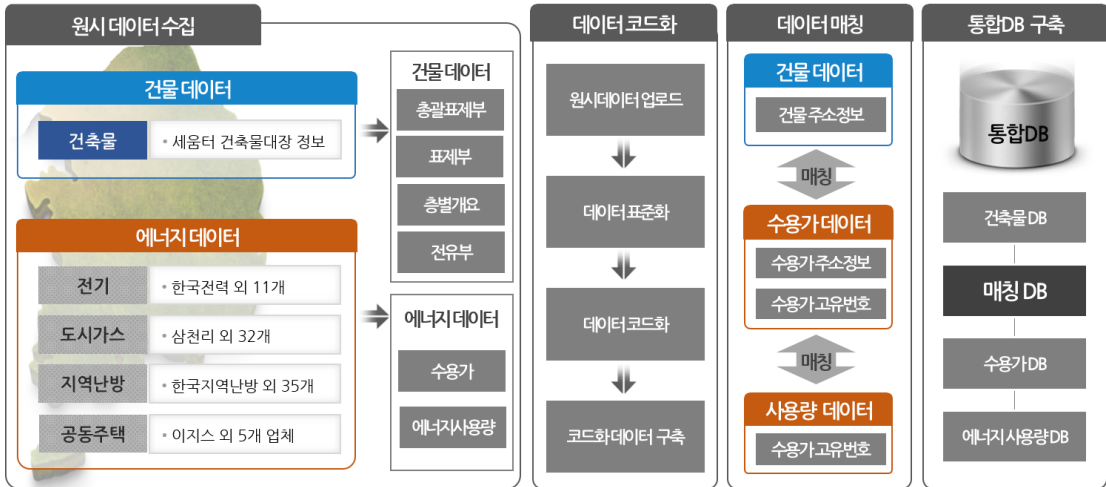
1. 국가건물에너지 통합DB

건축행정시스템(세움터)은 전국의 모든 건축물 정보DB를 구축하고 있다. 전기, 도시가스, 지역난방 에너지를 생산하고 각 사용처에 공급하는 에너지 공급기관은 판매한 에너지사용량에 대한 요금부과와 관리를 목적으로 에너지공급 정보DB를 구축하고 있다. 국가건물에너지 통합DB는 「녹색건축물 조성 지원법」을 근거로 건축행정시스템의 건축물 정보와 에너지 공급기관에서 제공하는 에너지 정보를 수집하여 통합·구축한 대규모 데이터베이스이다(박진영·정화미, 2015). 즉, 건축물의 온실가스 배출량 및 에너지사용량에 대한 분석과 통계의 개발을 위하여 2010년 6월부터 2015년 9월까지 국가건물에너지 통합DB의 구축이 진행되었으며, 2015년 9월부터는 한국감정원에 의해 운영·관리되고 있다.

국가건물에너지 통합DB의 구축방법은 다음〈그림 1〉과 같이 요약될 수 있다.

1) 원시데이터 수집

먼저, 원시데이터 수집 단계에서는 에너지 공급기관으로부터 개별 수용가 정보(계량기 번호, 주소 등)와 월별 에너지공급량(사용량) 데이터를 수집한다. 즉, 전기는 한국전력 외 10개의 구역전기 공급업체, 도시가스는 삼천리 외 32개의 공급업체, 지역난방은 한국지역난방 외 35개의 공급업체, 그리고 6개의 회계전산업체로부터 매



〈그림 1〉 국가건물에너지 통합DB 구축 절차

월 에너지 수용가(계량기) 및 공급량 데이터를 제공받는다. 그리고 건축행정시스템으로부터 매월 갱신된 건축물 데이터가 수집된다.

2) 데이터 코드화

에너지 수용가의 주소정보와 건물의 주소정보를 비교하여, 에너지사용량 데이터와 건축물 정보 데이터는 통합된다. 즉, “A” 수용가의 주소가 “가” 건물의 주소와 동일하면, “가” 건물에서 “A” 수용가로 공급된 에너지를 사용한 것으로 규정하고, 데이터를 통합한다. 그리고 이처럼 에너지 수용가와 건물의 주소정보를 비교하는 과정을 데이터 매칭이라고 정의한다. 이러한 데이터 매칭과 저장의 편의를 위하여, 2단계에서는 건물 및 에너지 정보 데이터의 코드화를 진행한다. 참고로, 지번까지의 주소정보는 PNU 체계에 따라 19자리로 정의되고, 건물 명칭 이하의 주소정보는 12자리의 숫자로 정의된다.

3) 데이터 매칭

앞선 언급과 같이, 코드화된 에너지 수용가의 주소정보와 건물의 주소정보를 바탕으로 데이터 통합이 이루어진다.

4) 통합DB 구축

매칭작업이 완료된 데이터는 통합DB에 적재된다. 2016년을 기준으로 7,592,880동의 건물과 39,579,853개의 에너지 수용가에 대한 정보가 구축되었다(〈표 1〉 참조). 비록, 건물에서 사용되는 모든 에너지원에 대한 데이터를 포함하지는 못하지만, 건물부문 전체 에너지사용량의 약 79%를 세 가지 에너지원(전기, 도시가스, 지역난방)이 차지한다는 점을 감안한다면(에너지경제연구원, 2017), 대단히 중요한 데이터라고 할 수 있다. 특히, 750,000동에 대한 데이터를 확보한 미국(Mathew et al., 2015)과 비교하더라도, 약 750만동 건물에 대한 정보를 포함하는 국가

〈표 1〉 국가건물에너지 통합DB 구축 현황

| 구분 | | 2012년 12월 | 2013년 12월 | 2014년 12월 | 2015년 12월 | 2016년 12월 |
|------|-------------|------------|------------|------------|------------|------------|
| 건물 | 동수(동) | 7,174,657 | 7,283,076 | 7,418,971 | 7,507,763 | 7,592,880 |
| | 연면적(천㎡) | 4,253,946 | 4,267,585 | 4,363,002 | 4,179,034 | 4,516,978 |
| 전기 | 수용가수 | 20,479,607 | 21,006,131 | 21,545,509 | 21,796,398 | 22,516,050 |
| | 에너지사용량(TOE) | 3,571,812 | 3,603,543 | 3,681,909 | 3,587,045 | 3,754,249 |
| 도시가스 | 수용가수 | 14,864,498 | 15,492,919 | 16,069,919 | 16,525,692 | 17,055,992 |
| | 에너지사용량(TOE) | 2,800,672 | 2,577,222 | 2,526,856 | 2,268,308 | 2,438,439 |
| 지역난방 | 수용가수 | 6,577 | 6,868 | 7,202 | 7,500 | 7,811 |
| | 에너지사용량(TOE) | 410,631 | 348,665 | 402,725 | 341,099 | 350,919 |

건물에너지 통합DB는 세계 최고의 DB라고 할 수 있다. 그리고 개별 건물단위의 월별 에너지 사용량 정보를 구축하고 있기 때문에, 데이터의 활용 가능성과 시의성 측면에서 매우 우수하다고 할 수 있다.

2. 데이터의 신뢰성 검토 방법

정보시스템을 통해 지속적으로 생산되는 데이터를 의사결정에 활용하고자 하는 관심은 공공뿐만 아니라 민간부문에서도 날로 높아져가고 있다. 데이터를 가공·분석하는 프로세스가 탁월하다 할지라도, 구축된 데이터가 신뢰할 만한 수준의 것이 되지 못한다면, 분석결과는 의미 없는 정보가 될 뿐만 아니라 잘못된 의사결정을 이끌 수 있다. 이러한 이유로 데이터의 신뢰성 확보에 대한 노력이 강화되는 추세이다(차경엽, 2010).

많은 연구자에 의해 데이터의 신뢰성 검토 기준이 제시되어 왔다(Ballou · Pazer, 1985;

Batini · Scannapieco, 2006). 특히, 공공부문 데이터에 대해서는 감사 측면에서 신뢰성을 점검하고 있다. 미국 감사원은 데이터의 신뢰성 검토 기준으로 완전성, 정확성, 일관성을 제시하고 있으며, 영국 감사원은 유효성과 적정성을 데이터의 신뢰성 기준으로 제시하였다. 차경엽(2010)은 이러한 해외 감사원의 사례를 분석하여 국내 실정에 맞는 공공부문 정보시스템 데이터에 대한 신뢰성 검토 기준을 제시하였다. 차경엽(2010)이 제시한 신뢰성 검토 기준은 4가지 항목(정확성, 완전성, 유효성, 일관성)으로 구성된다. 정확성은 데이터가 누락되거나 부정확한 데이터가 수집되었는지를 검토한다. 완전성은 모든 항목의 데이터를 포함하는지, 유효성은 수집된 데이터가 유효범위를 벗어나는지를 의미한다. 마지막으로, 일관성은 데이터의 구축이 일관된 형식과 규정된 절차에 따라 이루어지는지를 의미한다.

국가건물에너지 통합DB는 규정된 절차에 따라 건축물대장의 건물정보와 에너지정보를 모

두 포함하도록 구축이 이루어지기 때문에, 일관성 측면에서의 신뢰성은 확보되었다. 한편, 통합DB의 신뢰성을 확보하기 위하여 구축과정에 프로파일링 검사, 유형사례별 검사, 무작위 추출 검사 등이 시행된다. 이러한 검사과정은 개별 데이터의 정확성과 완전성 검토를 목표로 데이터의 누락이나 이상치 등의 결함을 찾을 수 있다. 하지만, 수집하고자 한 모든 데이터가 정확하게 수집되어 집계되었는지, 그리고 실제 분석에 활용 가능한 데이터는 얼마나 되는지에 대해서는 답을 주지 못한다. 이에 따라, 본 연구는 정확성, 완전성, 유효성 측면에서 국가건물에너지 통합DB에 수집된 모든 건물 및 에너지사용량 데이터를 검토하여 국가건물에너지통합DB의 신뢰성을 확보할 수 있는 검증방법을 제시하고자 한다.

III. 데이터 신뢰성 검증방법

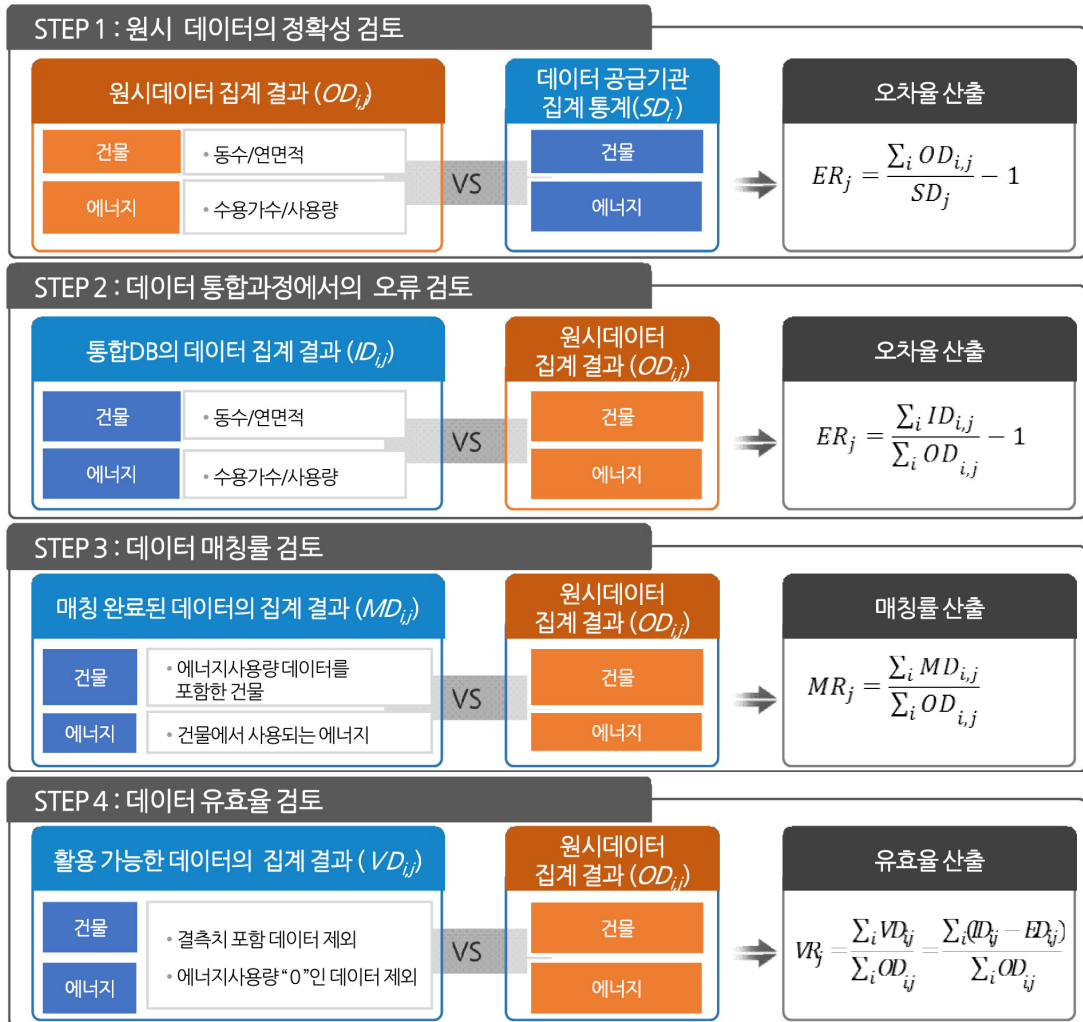
〈그림 1〉과 같이, 국가건물에너지 통합DB는 원시데이터의 수집, 데이터 코드화, 데이터 매칭, 통합DB 구축 등의 4단계 과정을 거쳐 구축된다. 데이터가 이동하고 가공되는 과정에서는 오류가 발생할 수 있기 때문에, 각 단계의 진행 전후에는 데이터 검토가 시행되어야 한다. 즉, 통합DB에 대한 신뢰성을 확보하기 위하여 데이터의 수집과정에서 오류가 있는 데이터를 포함하거나 가공 과정에 오류가 포함되지는 않았는지(정확성), 구축된 데이터는 전체 건물 중 어느

정도를 포함하고(완전성), 그 중 분석에 활용할 수 있는 데이터는 어느 정도인지(유효성) 등이 검토되어야 한다.

본 연구에서 제시하는 통합DB 데이터에 대한 검증방법은 〈그림 2〉와 같이, 4단계로 구성된다. Step 1과 Step 2는 구축된 데이터의 정확성에 대해 검토한다. 즉, Step 1은 데이터 수집 과정에서 오류를 검토한다. Step 2는 데이터 코드화와 데이터 매칭을 거쳐 구축된 통합DB의 오류를 검토한다. Step 3과 Step 4는 각각 구축된 통합DB 데이터의 완전성과 유효성에 대해 검토한다.

1. Step 1 : 원시데이터의 정확성 검토

앞선 언급과 같이, 통합DB를 구축하기 위한 첫 번째 단계는 건축행정시스템과 에너지 공급기관으로부터 건축물 정보 데이터와 에너지사용량 데이터를 수집하는 단계이다. 건축행정시스템과 에너지 공급기관에서는 통합DB 구축에 필요한 개별 건물 또는 개별 수용가에 대한 데이터를 제공하지만, 이와 함께 모든 건물 및 수용가의 데이터를 집계한 집계통계를 발표한다. 따라서 데이터 공급기관에서 제시하는 집계통계와의 비교를 통해, 원시데이터의 정확성을 검증하는 것이 필요하다. 원시데이터의 정확성은 〈식 1〉에 따라 산출되는 오차율을 바탕으로 검토된다. 그리고 원시데이터에 대한 정확성 검증은 건축물 정보 데이터와 에너지 정보 데이터를 구분하여 진행된다.



〈그림 2〉 통합DB의 데이터 신뢰성 확보를 위한 검증 방법

$$ER_j = \frac{\sum_i OD_{i,j}}{SD_j} - 1 \quad (\text{식 1})$$

ER_j : 검토항목 j에 대한 수집된 원시데이터의 오차율

$OD_{i,j}$: 공급기관으로부터 수집된 검토항목 j에 대한 건물 i(또는 에너지 수용가 i)의 원시데이터

SD_j : 데이터 공급기관에서 제시하는 검토항목 j에 대한 집계결과

j : 검토항목(건물동수, 연면적, 수용가수, 에너지 사용량)

1) 건물 데이터

건축행정시스템에서는 건축물대장에 대한 DB

를 구축하고, 이를 바탕으로 매년 건축물통계를 발표한다. 건축행정시스템과의 연계시스템을 통해 개별 건물에 대한 건축물대장 데이터가 자동적으로 수집되지만, 수집 및 집계과정에서 오류가 발생할 수 있다. 특히, 건축행정시스템에서 발표하는 결과는 대표적인 건축물 통계로 국가승인통계이다. 따라서 통합DB의 신뢰성과 적합성을 확보한다는 측면에서, 건축행정시스템에서 발표하는 집계결과와 통합DB 구축을 위해 수집된 데이터의 집계결과의 일치 여부가 파악되어야 한다.

건축행정시스템에서 발표하는 건축물 통계는 건축물 동수와 연면적의 집계결과를 포함한다. 따라서 건축물대장에 대한 원시데이터는 건물의 동수와 연면적이라는 두 가지 관점에서 비교·검토된다. 또한, 건축행정시스템은 17개 시도 및 건물용도에 따라 구분된 집계통계를 제시하기 때문에, 총량과 더불어 지역 및 건물용도에 따라 구분된 집계결과 역시 비교될 필요가 있다. 이때, 건축물통계는 주건물만을 대상으로 발표되기 때문에, 비교과정에서 부속건물은 제외되어야 한다.

2) 에너지사용량 데이터

한국전력¹⁾, 도시가스협회²⁾, 한국지역난방³⁾에서는 매월 공급한 에너지에 대한 집계결과를 발표한다. 그리고 에너지경제연구원에서는 에

너지 공급기관으로부터 에너지공급량 집계결과를 통보받아 에너지통계월보를 발간한다. 에너지통계월보는 대한민국의 에너지밸런스를 파악하는 것이 목적인 국가승인통계이지만, 최종 수요단계에서 소비되는 에너지사용량 정보를 포함한다. 즉, 전기, 도시가스, 지역난방의 경우에는 한국전력, 도시가스협회, 3개의 대표적인 지역난방 공급기관(한국지역난방, SH지역난방, GS와 위)으로부터 각각 제공받은 집계결과를 바탕으로 에너지통계월보는 발표된다(에너지경제연구원, 2017).

건물데이터와 마찬가지로 연계시스템을 통해 에너지 수용가 및 공급량 데이터가 자동적으로 수집되지만, 수집 및 집계과정에서 오류가 발생할 수 있다. 또한, 데이터의 수집 이후에 발생한 다양한 변동사항은 반영되지 못할 수 있다. 이러한 이유로, 통합DB의 신뢰성과 적합성을 확보한다는 측면에서, 각 에너지 공급기관에서 발표하는 집계결과와 통합DB의 구축을 위해 수집된 개별 수용가의 에너지사용량 집계결과가 일치하는지 여부가 파악되어야 한다.

에너지사용량에 대한 원시데이터는 두 가지 관점(에너지 수용가의 수, 에너지 사용량)에서 비교·검토되어야 한다. 앞선 언급과 같이, 현재 한국전력, 도시가스협회, 한국지역난방에서만 매월 에너지공급량의 집계결과를 발표한다. 따라서 수집된 원시데이터에서 공급기관이 한

1) 한전경제경영연구원에서는 한국전력의 전력통계속보를 발표한다.

2) 한국도시가스협회는 전국의 모든 도시가스 공급업체로부터 제공받은 결과를 집계하여, 도시가스 공급실적을 발표한다.

3) 한국지역난방공사는 매년 공급한 열관매량 정보를 제공한다.

국전력과 한국지역난방인 데이터만을 추출하여 집계한 후, 그 결과를 비교해야 한다. 전기와 도시가스는 17개 시도 및 사용용도에 따라 구분된 집계통계가 제시된다. 따라서 데이터의 신뢰성을 높이기 위하여 지역 및 사용용도에 따라 구분한 비교가 함께 시행되어야 한다.

2. Step 2 : 데이터 통합 과정에서의 오류 검토

수집된 원시데이터는 코드화와 매칭과정을 거쳐, 통합DB로 구축된다. 코드화나 매칭과 같은 데이터 가공과정은 오류를 포함할 수 있기 때문에, 통합과정 이후의 결과에 대한 정확성 검증 역시 필요하다. 따라서 2단계 검증과정에서는 (식 2)에 따라 원시데이터와 통합DB의 집계 결과를 비교하여, 통합 전·후의 데이터 일치 여부를 검토한다. 1단계에서와 같이, 2단계 검증 과정도 건축물정보 데이터와 에너지 정보 데이터로 구분하여 진행된다.

$$ER_j = \frac{\sum_i ID_{i,j}}{\sum_i OD_{i,j}} - 1 \quad (\text{식 } 2)$$

ER_j : 검토항목 j에 대한 통합데이터의 오차율

$ID_{i,j}$: 통합DB에 구축된 검토항목 j에 대한 건물 i(또는 에너지 수용가 i)의 데이터

$OD_{i,j}$: 공급기관으로부터 수집된 검토항목 j에 대한 건물 i(또는 에너지 수용가 i)의 원시데이터

j : 검토항목(건물동수, 연면적, 수용가수, 에너지 사용량)

3. Step 3 : 데이터 매칭률 검토

통합DB는 건물과 에너지 수용가의 주소정보 비교(데이터 매칭)를 통해 구축되지만, 여러 이유로 매칭이 되지 않는 데이터가 존재한다. 기존 연구에 따르면, i) 에너지 또는 건축물정보가 없는 경우, ii) 주소와 같은 속성 정보가 불일치하는 경우, iii) 건축물에 변경이 발생한 경우 등이 미매칭의 주요 원인으로 정의된다(정화미, 2015). 예를 들어, 가로등이나 신호등, 가판대에서 사용하는 에너지의 경우에는 일치하는 건물 주소가 존재하지 않을 것이다. 미매칭 유형의 약 58%가 건축물 정보의 부재로 인한다는 기존 연구의 결과(정화미, 2015)는 이러한 사실을 뒷받침한다고 할 수 있다.

통합DB를 활용한 분석결과가 신뢰성을 얻기 위해서는 매칭 데이터의 현황, 즉 완전성을 확보한 데이터의 현황이 파악되어야 한다. 여기서, 완전성을 확보한 데이터는 모든 항목의 데이터를 포함한다는 의미이기 보다는 건물정보와 에너지정보를 모두 포함하는 데이터를 의미한다. 따라서 건물 측면에서의 매칭률은 전국의 모든 건물 중에서 에너지 수용가를 가지는 건물의 비율을 의미한다. 에너지사용량 측면에서의 매칭률은 전체 에너지 수용가(또는 에너지사용량) 중 건물에서 사용하는 것으로 정의된 에너지 수용가(또는 에너지사용량)의 비율을 의미한다.

$$MR_j = \frac{\sum_i MD_{i,j}}{\sum_i OD_{i,j}} \quad (\text{식 } 3)$$

MR_j : 검토향목 j 에 대한 통합데이터의 매칭률
 $MD_{i,j}$: 건물데이터와 에너지데이터가 매칭 완료되어 통합DB에 구축된 검토향목 j 에 대한 건물 i 의 데이터
 $OD_{i,j}$: 공급기관으로부터 수집된 검토향목 j 에 대한 건물 i (또는 에너지 수송가 i)의 원시데이터
 j : 검토향목(건물동수, 연면적, 수용가수, 에너지 사용량)

4. Step 4 : 데이터 유효율 검토

일부 건물의 경우에는 계량기가 설치되어 있지 않을 수 있다. 또한, 계량기는 설치되어 있지만 전기와 도시가스, 지역난방을 사용하지 않는 건물이 존재할 수 있다. 에너지사용량 역시 건물에서 사용되지 않는 부분(가로등이나 가판대 등)을 포함한다. 3단계에서 산출되는 매칭률은 건물에서 사용되지 않는 에너지(또는 전기, 도시가스, 지역난방을 사용하지 않는 건물)를 포함하여 산출한 결과이기 때문에, 연구·분석에 활용이 가능한 유효한 데이터의 비율을 의미하지는 않는다. 이에 따라, 4단계에서는 실제 분석에 활용될 수 있는 데이터의 비율인, 유효율을 검토한다.

앞선 언급과 같이, 건물 정보와 에너지사용량 정보가 매칭되어 통합DB에 구축되었다고 하더라도, 일부 속성정보에 결측치가 있거나 오류를 포함하는 경우가 있다. 또한, 매칭 완료된 건물들 중에는 에너지사용량이 “0”인 건물도 포함되어 있다. 이러한 데이터는 집계 및 분석과정에서 적용되는 속성에 따라 다른 결과를 야기할

수 있다. 예를 들어, 에너지사용량이 “0”인 건물 역시 개별 사례로 고려되기 때문에, 단위면적당 에너지사용량과 같은 통계지표를 낮게 만들 수 있다. 따라서 결측치를 포함하거나 에너지사용량이 “0”인 건물을 제외하고, 분석과정에서 활용할 수 있는 데이터의 비율을 제시할 필요가 있다.

한편, 건축물대장은 총괄표제부(단지), 표제부(동), 층별개요(층), 전유부(호)로 구분되고, 에너지는 계약조건에 따라 단지, 동, 층, 호 단위로 계량기가 설치된다. 이로 인하여, 에너지 계량기마다 매칭되는 대상이 달라지며, 활용할 수 있는 데이터의 비율 역시 분석목적에 따라 달라질 수 있다. 예를 들어, 호(개별 세대)에 매칭된 에너지사용량은 동이나 단지로 합산하여 집계될 수 있기 때문에, 호단위의 분석뿐만 아니라 동단위나 단지단위의 분석에도 활용될 수 있다. 반면, 단지에 매칭된 에너지사용량은 동이나 호로 구분이 불가능하기 때문에, 세분화된 분석에 활용될 수 없다. 따라서 매칭 수준을 고려하여, 각각의 경우에 따른 유효한 데이터의 현황을 명확하게 제시해야 한다.

$$VR_j = \frac{\sum_i VD_{i,j}}{\sum_i OD_{i,j}} = \frac{\sum_i (ID_{i,j} - ED_{i,j})}{\sum_i OD_{i,j}} \quad (\text{식 4})$$

VR_j : 검토향목 j 에 대한 통합데이터의 유효율
 $VD_{i,j}$: 유효한 데이터를 포함하는 건물 i 에 대한 검토향목 j 의 데이터
 $OD_{i,j}$: 공급기관으로부터 수집된 검토향목 j 에 대한

건물 i (또는 에너지 수용가 i)의 원시데이터
 $ID_{i,j}$: 통합DB에 구축된 검토항목 j 에 대한 건물
 i (또는 에너지 수용가 i)의 데이터
 $ED_{i,j}$: 통합DB에 구축되었으나, 오류를 포함한
 검토항목 j 에 대한 건물 i (또는 에너지 수
 용가 i)의 데이터
 j : 검토항목(건물동수, 연면적, 수용가수, 에너지
 사용량)

IV. 사례적용

본 연구에서 제시한 검증방법의 타당성을 제
 시하고 기 구축된 통합DB의 정확성, 매칭률, 유
 효율을 파악하기 위하여 2015년 1월의 데이터를
 활용하여 사례분석을 시행하였다. 이를 위하여,
 한국전력, 도시가스협회, 한국지역난방에서 공
 개한 2015년 1월의 에너지공급량 집계통계를 수
 집하였다. 건축물통계의 경우 매년 12월 31일을
 기준으로 조사되어 발표되기 때문에, 건축행정
 시스템으로부터 2014년 건축물통계를 수집하였
 다. <그림 2>와 같이, 보다 정확한 결과를 얻기
 위해서는 지역 및 건물용도, 에너지사용 용도 등
 에 따른 상세 비교가 시행되어야 하지만, 본 사

례분석에서는 전국 단위로 건물의 연면적과 에
 너지사용량을 비교하는 것으로 범위를 한정하
 였다. 그리고 집합건물의 경우에는 단지단위로
 건물 데이터를 통합하여 사례분석을 시행하였
 다. 사례분석 결과는 다음과 같다.

1. 건물 데이터

<표 2>는 건물 데이터에 대한 검토 결과를
 보여준다. 먼저, 1단계 과정인 원시데이터와
 건축물통계를 비교한 결과, 건축물통계에 비해
 원시데이터의 집계결과가 상당히 큰 것으로 나
 타났다. 이는 일부 건물의 경우 연면적에 이상
 치를 포함하기 때문이었다. 이러한 오류는 수
 집과정에서 발생한 것이 아니라, 건축물대장에
 입력된 값이 잘못되어 있었던 것으로 확인되었
 다. 실제로 건축물대장을 확인한 결과, 표제부
 연면적의 합은 123,234,016.052m²로 표기된 반
 면, 전유부 연면적의 합은 88,014,027m²로 나
 타났다. 그리고 건축물통계 작성 담당자에게
 문의한 결과, 건축물통계를 작성할 때 이처럼
 건축물대장에 포함된 오류를 조정하지만, 건축
 물대장을 수정할 권한이 없기 때문에, 오류를

<표 2> 건물 데이터에 대한 검토결과

| 구분 | | 연면적 (백만m ²) | 오차율 | 매칭률 | 유효율 |
|--------|------------|-------------------------|--------|--------|--------|
| 기준 | 건축물통계 | 3,451 | - | - | - |
| Step 1 | 원시데이터 | 4,372 | 26.67% | - | - |
| Step 2 | 통합 데이터 | 4,372 | 0.0% | - | - |
| Step 3 | 매칭 데이터 | 3,831 | - | 87.64% | - |
| Step 4 | 유효 데이터 (1) | 3,442 | - | - | 78.74% |
| | 유효 데이터 (2) | 3,092 | | | 70.72% |

포함한 데이터를 제공하는 것으로 확인되었다.

2단계에서는 아무런 오류가 없이, 원시데이터와 동일한 결과가 도출되었다. 3단계 검증 결과, 에너지사용량정보와의 매칭률은 87.64%로 나타났다. 즉, 에너지사용량정보와 매칭된 건물의 연면적은 원시데이터에 비해 12.36% 작게 나타났다.

마지막으로, 분석에 활용 가능한 유효한 데이터의 비율인, 유효율을 검토하였다. 본 사례분석에서는 결측치나 이상치를 가진 건물 데이터와 에너지사용량이 “0”인 건물 데이터를 구분하기 위하여, 유효데이터를 2단계로 구분하여 파악하였다. 먼저, 매칭 데이터에서 결측치와 이상치를 제거한 결과, 연면적은 3,442백만m²로 건축물통계와 거의 유사한 것으로 나타났다(〈표 2〉의 유효데이터 (1) 참조). 이때, 650만m² 이상의 연면적을 임의로 이상치로 규정하고, 650만 m² 이상의 연면적을 가지는 건물을 대상에서 제외하였다.⁴⁾ 〈표 2〉의 유효데이터(2)는 에너지 사용량이 “0”인 건물까지 모두 제거한 결과를 보

여준다. 최종적으로 정리된 통합DB의 데이터 유효율은 70.72%인 것으로 나타났다.

〈표 2〉에 제시된 유효율은 전체 건축물 대비 유효한 에너지사용량 데이터를 가진 건물의 비율을 의미한다기 보다는, 오류를 가진 건물의 연면적까지도 모두 합한 결과에 대한 유효한 데이터의 비율을 의미한다. 건축물통계를 기준으로 유효율을 산출하면, 89.59%로 향상된다.

2. 에너지사용량 데이터

1) 전기

〈표 3〉과 같이, 통합DB 구축을 위해 수집된 원시데이터의 전기 사용량 집계 총량은 한국전력통계의 결과와 거의 차이가 없었다(오차율 0.03%). 이러한 결과는 전기 사용량 데이터는 건물 데이터와 달리 오류를 포함하지 않는 사실을 보여준다. 데이터 코딩화 및 통합과정(Step 2)에서는 아무런 오류가 발생하지 않았다(오차율 0.00%). Step 3의 결과에 따르면, 수집된 데

〈표 3〉 전기 사용량 데이터에 대한 검토결과

| 구분 | | 에너지사용량(TOE) | 오차율 | 매칭률 | 유효율 |
|--------|------------|-------------|-------|--------|--------|
| 기준 | 한국전력통계 | 3,906,404 | - | - | - |
| Step 1 | 원시데이터 | 3,907,525 | 0.03% | - | - |
| Step 2 | 통합 데이터 | 3,907,525 | 0.00% | - | - |
| Step 3 | 매칭 데이터 | 2,992,883 | - | 76.59% | - |
| Step 4 | 유효 데이터 (1) | 2,991,679 | - | - | 76.56% |
| | 유효 데이터 (2) | 2,991,679 | - | - | 76.56% |

4) 매칭 결과에 대하여 보다 정확하게 파악하기 위해서는 오류가 있을 것으로 여겨지는 개별 데이터를 일일이 검토하여 수정·개선해야 한다. 하지만 본 연구는 제시한 방법의 타당성을 보여주는 것을 목적으로 기 구축된 데이터의 현황을 개략적으로 파악하고자 하였기에, 임의의 기준인 650만m²를 제외 기준으로 적용하였다.

이터 중에서 건물정보와 매칭이 완료된 전기 데이터는 에너지사용량 기준 2,993천TOE로, 전체 전기사용량의 약 76.59%가 건물에 매칭된 것이 확인되었다. 하지만, 이러한 결과는 가로등이나 가판대와 같은 건물과 매칭이 불가능한 에너지까지 고려했을 때의 결과이므로, 모든 건물에서 사용하는 에너지 중에서 76.59%만이 완전한 데이터로 구축되었다는 의미는 아니다. 건물에서 사용되는 에너지만을 고려한다면, 데이터 매칭률은 높아질 것이다. 마지막으로 결측치와 연면적 오류를 포함한 건물을 제외한 결과 1,204TOE가 제외되어, 76.56%의 유효율을 가지는 것으로 도출되었다. Step 4의 유효 데이터 (2)는 매칭 데이터에서 에너지사용량이 0인 건물을 제외한 경우를 보여주기 때문에, 유효 데이터 (1)과 동일하다.

2) 도시가스

〈표 4〉는 도시가스 사용량에 대한 결과를 보여준다. 통합DB 구축을 위해 수집된 원시데이터의 도시가스 사용량 집계 총량은 도시가스협회에서 제시한 결과보다 약 0.11% 큰 것으로 나타났다. 그리고 데이터 코딩화 및 통합과정(Step 2)에서는 아무런 오류가 발생하지 않았다

(오차율 0.00%). Step 3의 결과에 따르면, 수집된 데이터 중에서 건물정보와 매칭이 완료된 도시가스 데이터는 에너지사용량을 기준으로 약 91.69%인 것으로 확인되었다. 전기와 달리, 대부분의 도시가스가 건물에서 사용되기 때문에 높은 유효율이 나타난 것으로 여겨진다. 마지막으로 결측치와 연면적 오류를 포함한 건물을 제외한 결과 1,551TOE가 제외되어, 91.64%의 데이터가 유효한 것으로 파악되었다. Step 4의 유효 데이터(2)는 에너지사용량이 0인 사례를 제외한 것이므로, 유효 데이터 (1)과 동일하였다.

3) 지역난방

〈표 5〉는 지역난방 사용량에 대한 결과를 보여준다. 앞선 언급과 같이, 매월 지역난방 에너지공급량 집계결과를 공개하는 공급업체는 한국지역난방이 유일하다. 따라서 사례분석에서는 통합DB로부터 한국지역난방에 대한 데이터만을 추출하여 비교하였다. 원시데이터의 지역난방 에너지사용량 집계 총량은 한국지역난방에서 제시한 결과보다 약 0.03% 작아 거의 차이가 없는 것으로 나타났다. 그리고 데이터 코딩화 및 통합과정(Step 2)에서는 아무런 오류가 발

〈표 4〉 도시가스 사용량 데이터에 대한 검토결과

| 구분 | | 에너지사용량(TOE) | 오차율 | 매칭률 | 유효율 |
|--------|------------|-------------|-------|--------|--------|
| 기준 | 도시가스협회 | 3,081,669 | - | - | - |
| Step 1 | 원시데이터 | 3,085,104 | 0.11% | - | - |
| Step 2 | 통합 데이터 | 3,085,104 | 0.00% | - | - |
| Step 3 | 매칭 데이터 | 2,828,840 | - | 91.69% | - |
| Step 4 | 유효 데이터 (1) | 2,827,289 | - | - | 91.64% |
| | 유효 데이터 (2) | 2,827,289 | - | - | 91.64% |

〈표 5〉 한국지역난방에서 공급한 지역난방 사용량 데이터에 대한 검토결과

| 구분 | | 에너지사용량(TOE) | 오차율 | 매칭률 | 유효율 |
|--------|------------|-------------|-------|--------|--------|
| 기준 | 한국지역난방 | 223,497 | - | - | - |
| Step 1 | 원시데이터 | 223,440 | 0.03% | - | - |
| Step 2 | 통합 데이터 | 223,440 | 0.00% | - | - |
| Step 3 | 매칭 데이터 | 209,234 | - | 93.64% | - |
| Step 4 | 유효 데이터 (1) | 206,321 | - | - | 92.34% |
| | 유효 데이터 (2) | 206,321 | | | 92.34% |

생하지 않았다(오차율 0.00%). Step 3의 결과에 따르면, 수집된 데이터 중에서 건물정보와 매칭이 완료된 지역난방 데이터는 에너지사용량을 기준으로 약 93.64%인 것으로 확인되었다. 마지막으로 결측치와 연면적 오류를 포함한 건물을 제외한 결과, 약 400TOE가 제외되어 92.34%의 데이터가 유효한 것으로 파악되었다.

한편, 〈표 6〉과 같이 모든 지역난방 공급업체로부터 수집한 지역난방 사용량은 382,736TOE로 집계되었다. 그 중, 건물과 매칭이 완료된 데이터는 348,339TOE(91.01%)이며, 분석에 활용될 수 있는 유효한 데이터는 347,654TOE(90.83%)로 나타났다. 이는 대표적인 3개 지역난방 공급업체 외에 나머지 33업체에서 약 20.8%(79,736TOE)의 지역난방을 공급하고 있음을 의미한다. 또한, 에너지통계월보에 비해 국가건물에너지 통합DB

데이터가 보다 포괄적이라는 사실도 보여준다. 이러한 결과를 고려할 때, 기존의 국가승인통계를 개선할 수 있다는 측면에서도 국가건물에너지 통합DB는 큰 역할을 할 것으로 기대된다.

3. 결과분석 및 논의

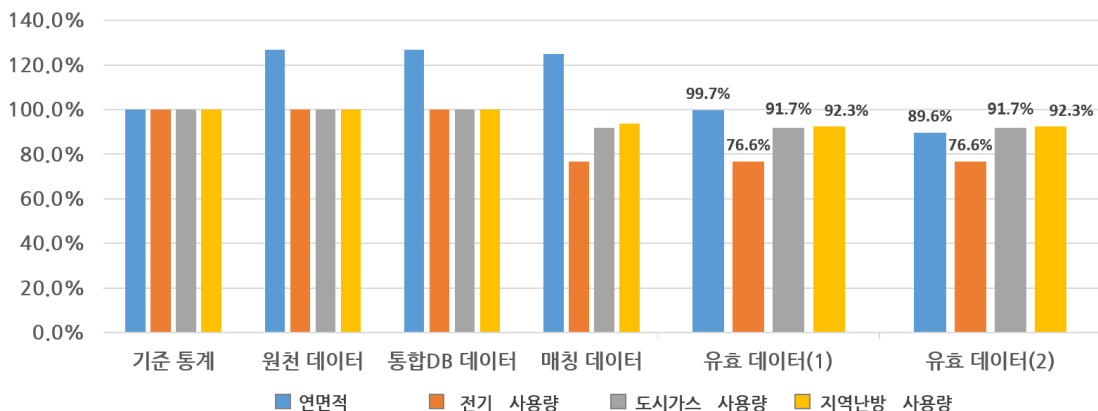
본 연구에서는 3장에서 제시한 검증방법의 타당성과 기존에 구축된 데이터의 신뢰성을 파악하고자 사례분석을 시행하였다. 4단계의 검증방법에 따라 2015년 1월에 대한 데이터를 분석한 결과, 건물 데이터는 연면적에서 상당한 오차를 가지는 것으로 확인되었다. 앞선 언급과 같이, 연면적이 1억m²를 넘는 건물도 확인되었으며, 이러한 오류로 인하여 2015년 1월의 모든 건물에 대한 연면적은 26.7%까지 높게 산출된 것이

〈표 6〉 지역난방사용량 데이터에 대한 검토결과

| 구분 | | 에너지사용량(TOE) | 오차율 | 매칭률 | 유효율 |
|--------|------------|-------------|------|--------|--------|
| Step 1 | 원시데이터 | 382,736 | - | - | - |
| Step 2 | 통합 데이터 | 382,736 | 0.0% | - | - |
| Step 3 | 매칭 데이터 | 348,339 | - | 91.01% | - |
| Step 4 | 유효 데이터 (1) | 347,654 | - | - | 90.83% |
| | 유효 데이터 (2) | 347,654 | | | 90.83% |

다(〈그림 3〉 참조). 이러한 오류는 데이터 수집 과정에서 발생한 것이 아니라, 건축물대장의 원시데이터가 오류를 포함하고 있었기 때문이다. 앞선 언급과 같이, 한 건물의 표제부 연면적 합은 123,234,016.052m²인 반면, 전유부 연면적의 합은 88,014.027m²였다. 이에, 본 연구에서는 연면적에 오류가 있는 데이터를 제외하기 위하여, 연면적 650만m² 이상의 건물단지 데이터는 분석 대상에서 제외하였다. 하지만, 연면적 650만m²는 건축물통계의 연면적과 오차율을 최소화하기 위하여 임의로 정의된 기준이다. 따라서 실제 통합DB의 구축과정에서는 개별데이터에 대한 오류를 검토할 수 있는 교차분석 등의 절차가 논의되어야 할 것이다. 예를 들어, 각각의 사례에 대하여, 총괄표제부에 제시된 연면적과 표제부에 제시된 연면적의 합, 전유부에 제시된 연면적의 합을 서로 비교하여 오류를 포함한 사례를 도출하고, 데이터를 개선하는 과정 등이 검토되어야 할 것이다.

건물 데이터와 달리, 에너지사용량 데이터는 거의 오차가 없었다. 하지만, 데이터 공급기관에서 발표하는 집계결과와의 오차율이 0.1% 이하로 작았을 뿐, 완전히 일치하는 것은 아니었다. 이러한 오차는 두 가지 측면이 원인이 될 수 있다. 첫 번째는 수집된 데이터를 TOE단위로 변환하는 과정이다. 예를 들어, 한국감정원에서는 도시가스 공급업체로부터 Nm³단위로 수집된 데이터에 에너지법 시행규칙에 제시된 에너지열량계수인 43.6kcal/Nm³을 적용하여, MJ 단위의 도시가스 사용량을 산출한다. 하지만, 실제로는 공급업체의 발전효율이나 해당 기간의 연료 비율과 같은 특성에 따라 변환계수는 달라진다. 두 번째 원인은 데이터를 수집한 이후 발생한 변동사항에 대해 반영하지 못한다는 점이다. 예를 들어, 에너지사용량에 따른 요금을 부과한 이후 잘못된 부분이 확인되어 부과요금을 환불하는 경우가 발생한다. 에너지통계월보는 이러한 변경 사항을 반영하지만, 통합DB는 반영하



〈그림 3〉 통합DB 데이터 검증결과 요약

지 못한다. 따라서 데이터 공급기관으로부터 데이터를 수집할 때, 데이터 정확성에 대한 검토 결과와 함께 다음의 내용도 지속적으로 수집·보고되어야 할 것이다. i) 데이터 수집 시점 및 데이터를 수집 한 이후에 발생한 변동사항. ii) 변환과정에서 사용한 방법 및 계수.

결측치나 연면적 오류를 포함하는 사례, 에너지사용량이 0인 사례를 제외하여 데이터 유효율을 산출한 결과, 전기 에너지에 대한 유효율이 가장 낮게 나타났다. 건물 데이터, 도시가스, 지역난방의 유효율은 각각 89.59%, 91.75%, 92.34%인 반면, 전기 데이터의 유효율은 76.58%에 머물렀다. 이처럼 전기가 상대적으로 낮은 유효율을 가지는 원인은 도시가스나 지역난방에 비해 전기가 건물 외에 다양한 곳에서 다양한 용도로 사용되기 때문인 것으로 여겨진다. 따라서 향후에는 건물에서 사용되지 않는 에너지사용량은 제외하고, 유효율을 산출하는 과정에 고려되어야 할 것이다.

그리고 사례분석 결과 에너지사용량이 0인 건물이 전체의 약 10%를 차지하였다. 이는 단위면적당 에너지를 산출하거나 벤치마크를 산출할 때, 결과에 큰 영향을 줄 수 있는 수준이라는 사실을 보여준다.⁵⁾ 특히, 대부분의 건물은 지역난방 에너지사용량이 0이기 때문에, 단순히 모든 건물을 포함하여 분석했을 때 지역난방 에너지사용량의 중앙값은 0에 수렴할 것이다. 그리고 지역이나 용도에 따라 세분화된 분석을 하는 경

우에는 더욱 큰 영향을 미칠 수 있다. 따라서 에너지원에 따라 구분된 결과를 분석하는 경우에는 반드시 각 에너지원별 사용량이 0인 건물은 분석 대상에서 제외되어야 한다.

이와 같이, 사례분석의 결과는 통합DB에 구축된 데이터가 완전하지 못하다는 사실과 분석과정에서 고려되어야 하는 사항을 보여주었다. 즉, 본 연구에서 제시한 검증 방법은 통합DB의 구축과정에서 발생한 오류와 구축된 통합DB의 현황을 명확하게 보여줄 수 있다. 따라서, 본 연구에서 제시한 검증 방법을 통해 데이터의 정확도와 유효율을 검토한다면, 국가건물에너지 통합DB의 신뢰성 향상에 기여할 것으로 기대된다.

V. 결론

국가건물에너지 통합DB는 대한민국의 모든 건물에 대한 정보 및 에너지사용량 정보를 포괄하는 세계 최초의 DB로서, 구축된 데이터는 건물부문에 대한 에너지 정책의 효과 분석, 개별 건물 또는 지역단위에서의 에너지사용량 관리, 에너지 계획, 건물부문에서의 미래 에너지사용량 예측 등에 활용될 수 있다. 이러한 분석에 대한 신뢰성 확보와 특히 모든 건물을 대상으로 하는 통계연구에서는 데이터의 신뢰성 확보가

5) 일반적으로 특정 지역, 용도에 따른 단위면적당 에너지사용량은 건물별 단위면적당 에너지사용량을 산출한 후 해당 지역 또는 용도에 포함되는 건물의 중앙값을 사용한다. 이로 인하여, 해당 지역 또는 용도에 에너지사용량이 0인 건물이 많이 포함되면, 중앙값이 0에 가까워지는 오류가 발생할 수 있다.

무엇보다도 중요하다. 이러한 측면에서, 본 연구에서 제시한 방법은 데이터의 정확성, 완전성, 유효성 측면에서 데이터의 신뢰성을 파악할 수 있는 매우 중요한 과정이라고 생각한다.

본 연구는 다음과 같이 요약될 수 있다.

- 기존 연구를 참고하여, 데이터의 신뢰성 확보를 위하여 필요한 요소로서 정확성, 완전성, 유효성을 정의하였다.
- 통합DB의 구축과정을 분석하여, 4단계로 구성된 통합DB 데이터의 신뢰성을 검토하는 검증방법을 제시하였다. i) 원시데이터의 정확성 검토, ii) 통합과정에서의 오류 검토, iii) 통합데이터의 매칭률 검토, iv) 통합데이터의 유효율 검토. Step 1과 Step 2에서는 오차율을 통해 각각 수집 및 통합과정을 거친 데이터의 정확성을 검토한다. Step 3과 Step 4를 거쳐 산출되는 매칭률과 유효율은 각각 완전성과 유효성 측면에서 구축된 통합DB를 설명한다.
- 4단계 검증방법에 기 구축된 데이터를 적용하여 사례분석을 시행하였다. 2015년 1월 데이터를 분석한 결과, 데이터의 수집과정과 통합과정에서는 거의 오류가 없는 것이 확인되었다. 다만, 건물 연면적의 경우에는 건축물대장에 입력된 정보가 오류를 포함하고 있어, 건축물통계의 연면적보다 약 26% 높게 집계되었다. 그리고 건물정보와 에너지사용량 정보가 매칭 완료되었음에도 불구하고, 특정 변수에 데이터가 없거나 에너지사용량이 0인 건물이 상당수 존재하였

다. 결측치와 에너지사용량 0인 사례, 연면적에 오류를 포함하는 사례를 제거하여 최종적으로 산출된 건물, 전기, 도시가스, 지역난방의 데이터 유효율은 각각 89.6%, 76.6%, 91.7%, 92.3%이었다.

이와 같이, 본 연구에서 제시한 검증방법은 데이터의 신뢰성 확보를 위한 정보는 물론, 개선이 필요한 부분을 도출하는 데에도 의미 있는 역할을 할 수 있다. 특히, 빅데이터 분석과 같은 4차 산업혁명에 대한 기대와 함께, 국가건물에너지 통합DB의 활용은 더욱 넓어질 것으로 기대된다. 이러한 점을 고려할 때, 통합DB의 신뢰성 확보는 무엇보다 중요하기에, 본 연구에서 제시한 검증방법의 의미는 매우 크다고 할 수 있다.

하지만, 본 연구는 기 구축된 데이터 중 일부(2015년 1월)에 한정하여 사례분석을 시행하였으며, 평가항목 역시 연면적과 에너지사용량으로 제한하였다. 또한, 본 연구에서는 지역이나 용도에 따른 상세 구분 없이 전국 단위의 총량만을 검토하였다. 이에 따라, 향후 연구에서는 기 구축된 모든 데이터를 대상으로 상세 구분에 따른 비교·검증이 시행되어야 할 것이다. 게다가, 본 사례분석에서는 건물의 연면적 측면에서의 유효율을 파악하기 위하여 650만 m^2 의 연면적을 제외 기준으로 사용하였다. 하지만, 이로 인하여 건물의 연면적 측면에서 제시된 사례분석 결과는 정확하다고 할 수 없다. 비록, 사례분석의 목적이 제시된 검증방법의 타당성을 제시하는 것이었다고 하더라도, 이는 분명한 본 연구

의 한계이다. 따라서 향후 연구에서는 연면적의 오류를 가진 건물을 파악하고 개선을 거쳐, 정확한 결과를 제시해야 할 것이다.

한편, 본 연구에서는 각 에너지공급기관에서 발표하는 집계통계 결과를 활용하여 사례분석을 시행하였으나, 대부분의 에너지 공급기관은 이러한 집계통계를 발표하지 않는다. 이에 따라, 검증방법의 지속적인 운영을 위해서는 통합 DB 구축 절차에 모든 에너지 공급기관과 건축행정시스템(세움터)으로부터 매달 집계결과를 함께 통보받는 절차가 추가되어야 할 것이다. 그리고 수집된 데이터에 검증방법을 적용하여, 매월 신규로 구축되는 데이터에 대한 현황 정보(정확성 및 매칭률, 유효율)가 함께 발표되어야 할 것이다.

참고문헌

1. 국토교통부, 「녹색건축물 조성지원법」, 법률 제 13790호.
2. 박진영 · 정화미, 2015, 「건축물 에너지 · 온실가스 정보체계 소개」, 『대한설비공학회 2015년도 동계 학술발표대회 논문집』, pp.62~64
3. 양시원 · 김선숙, 2014, 「에너지 사용량 평가방법 분석을 통한 건축물에너지소비증명제도 개선 방향」, 『대한건축학회 논문집 - 계획계』, 30(12): 307-314.
4. 에너지경제연구원, 2017, 『2017 에너지통계월보』, 33(04).
5. 정화미, 2015, 「건축물 에너지 · 온실가스 정보체계 현황과 활용방안」, 『건축환경설비』, 9(3): 12-19.
6. 조상규 · 김영현, 2013, 「녹색건축 정책수립을 위한 건축물 온실가스 배출량 통계 구축 및 분석」, 건축도시공간연구소.
7. 차경엽 · 심광호, 2010, 「공공부문 정보시스템 데이터의 신뢰성 점검기법 개발」, 『한국통계학회 논문집』, 17(5): 745-753.
8. 최민석 · 오승엽 · 지창윤 · 이동영, 2017, 「국가 건물에너지 통합 DB를 활용한 주거용 건물에너지 소비현황 분석 - 건물용도, 지역, 규모 그리고 준공년도를 중심으로 -」, 『부동산분석』, 3(1): 101-118.
9. Ballou, D.P., and Pazer, H.L., 1985, "Modelling data and process quality in multi-input, multi-output information systems", *Management Science*, 31: 150-162.
10. Batini, C. and Scannapieco, M., 2006, *Data Quality Concepts, Methodologies and Techniques*, Springer, New York.
11. Drousa, K.G., Kontoyiannidis, S., Dascalaki, E.G., and Balaras, C.B., 2016, "Mapping the energy performance of hellenic residential buildings from EPC(Energy performance certificate) data", *Energy*, 98: 284-295.
12. Gangolells, M., Casals, M., Forcada, N., Macarulla, M., and Cuerva, E., 2016, "Energy Mapping of existing building stock in Spain", *Journal of Cleaner Production*, 112: 3895-3904.
13. Jeong, J., Hong, T., Ji, C., Kim, J., Lee, M., and Jeong, K., 2016, "Development of an evaluation process for green and non-green buildings focused on energy performance of G-SEED and LEED.", *Building and Environment*, 105: 172-184.
14. Jeong, K., Hong, T., Kim, J., 2018, "Development of a CO2 emission benchmark for achieving the national CO2 emission reduction target by 2030", *Energy and Buildings*, 158: 86-94.
15. Mathew, P.A., Michael, L.N., Mercado, A., Custudio, C., and Walter, T., 2015, "Big-data

for building energy performance: Lessons from assembling a very large national database of building energy use", *Applied Energy*, 140: 85-93.

16. 건축행정시스템 세움터, <<http://www.eais.go.kr/>>
17. 한국도시가스협회, 공급실적, <http://www.citygas.or.kr/info/stats/index.jsp?sbranch_fk=2>
18. 한국지역난방공사, 열관매량, <https://www.kdlhc.co.kr/noticeList.do?srp=S22&siteCmsCd=CM3654&topCmsCd=CM3693&cmsCd=CM3711&pnum=1&cnum=1>
19. 한전경제경영연구원, 한국전력통계, https://home.kepco.co.kr/kepco/KO/ntcob/list.do?boardCd=BRD_000099&menuCd=FN05030103

논문접수일 : 2017년 10월 17일

심사(수정)일 : 1차 2017년 11월 13일

게재확정일 : 2017년 11월 30일

국문초록

에너지성능이 우수하거나 열등한 건물을 규정하거나 건물 에너지사용량 감축 정책의 효과를 분석하는 등 건물 에너지의 효율적인 관리와 정책 수립을 목적으로, 국가 건물에너지 통합DB는 구축되었다. 통합DB를 이용한 분석의 결과가 신뢰성을 확보하고 정책 수립에 활용되기 위해서는 통합DB의 신뢰성 확보가 무엇보다도 중요하다. 이에, 본 연구에서는 정확성, 완전성, 유효성 측면에서 데이터의 신뢰성을 보여줄 수 있는 4단계의 데이터 검증방법을 제시하였다. 첫 번째, 두 번째 단계는 오차율을 바탕으로 수집·구축되는 데이터의 정확성을 검증한다. 세 번째 단계와 네 번째 단계는 각각 매칭률과 유효율을 바탕으로 구축된 데이터의 완전성과 유효성을 보여준다. 사례분석으로 2015년 1월 데이터를 제시한 검증방법에 적용하였다. 그 결과, 데이터의 수집 및 통합과정에서는 오류가 없었다(1, 2단계). 최종적으로 건물, 전기, 도시가스, 지역난방의 데이터 유효율은 각각 89.6%, 76.6%, 91.7%, 92.3%로 산출되었다. 검증방법은 데이터의 신뢰성 확보를 위한 정보(오차율, 매칭률, 유효율)는 물론, 구축 및 분석 과정에 개선이 필요한 사항까지도 보여줄 수 있다. 본 연구에서 제시한 검증방법은 통합DB의 신뢰성 확보와 활용에 기여할 것으로 기대된다.

주제어 : 국가건물에너지 통합DB, 데이터 신뢰성, 정확성, 유효성, 데이터 검증

