

# 주거부문 행정등록자료의 인구주택총조사 활용가능성에 대한 경험적 검토: 건축물대장을 중심으로

이명진\* · 서우석\*\* · 이진\*\*\*

## < 요약 >

이 연구는 인구총조사에서 행정자료의 활용을 확대해 나아가기 위한 기초적인 작업으로서, 인구총조사에서 주택관련 항목과 밀접한 관련이 있는 건축물대장의 활용정도를 살펴보았다. 2004년 7월에 실시한 5차 시험조사 중 3개 지역에서 수집된 조사자료를 바탕으로 가구명부-거처정보 자료와 건축물대장이 얼마나 잘 상응하는지를 분석하였다. 지역별, 지역내 조사구별 매칭 결과의 비교 분석을 바탕으로 건축물대장의 활용 정도와 정보의 정확성 등을 평가하였다. 끝으로 건축물대장의 활용 정도를 높이기 위한 방안으로서 자료 매칭을 효과적으로 수행하기 위한 기술적인 방안들을 제시하였고, 건축물대장과 새주소체계의 개선 방안을 제시하였다.

**핵심단어 :** 인구주택총조사, 행정자료 활용, 건축물대장

\* 국민대 사회학과 교수(email: [mjlee@kookmin.ac.kr](mailto:mjlee@kookmin.ac.kr)) - 공동교신저자

\*\* 서울시립대 도시사회학과 교수(email: [usseo@uos.ac.kr](mailto:usseo@uos.ac.kr)) - 공동교신저자

\*\*\* 서울시립대 도시사회학과 교수(email: [kunlee@uos.ac.kr](mailto:kunlee@uos.ac.kr))

## 머리말

거의 대부분의 나라에서 국가통계의 기본이 되는 인구주택총조사(census, 이하 인구총조사)를 정기적으로 전수조사방식에 의해 시행해 왔다. 그러나 최근 들어 선진국을 비롯한 일부 국가에서 조사환경이 나빠지고 있다. 생활양식이 변하여 조사대상을 접촉하기 어렵고, 응답에 대한 거부감이 증가하였다. 더불어 조사비용이 빠르게 증가하여 전통적인 조사방법으로 국가통계를 작성하기가 점점 어려워지고 있다. 이러한 조사환경의 악화는 조사의 포괄성과 신뢰성에 대한 문제로 이어졌다. 심지어 네덜란드는 1970년대, 독일은 1990년대 이후 인구총조사를 중단한 상태이다(Bierau, 2000). 선진국들과 마찬가지로 우리나라에서도 조사환경이 빠른 속도로 악화되고 있다. 더욱이 우리의 경우 읍면동사무소 기능축소로 말미암아 과거 인구총조사에서 실질적 도움을 주었던 행정지원이 없어짐에 따라 앞으로 조사의 어려움은 더욱 커질 것으로 보인다.

인구총조사를 수행하는 데 어려움을 겪으면서 많은 국가는 최근의 사회적 변화를 전통적 조사방법에 대한 '근본적인 도전'으로 인식하게 되었다(Jensen, 2000). 이러한 변화에 대응하여 선진국들은 전통적으로 전수조사에 의존해왔던 인구총조사를 보완하거나 대체할 수 있는 방법들을 모색하고 있다. 예를 들면, 인구총조사 방식을 살펴보면 일반적으로 5년이나 10년마다 정기적으로 실시하는 전수조사의 전통적 센서스 이외에도 순환형 센서스(rolling census)와 행정적으로 등록된 여러 형태의 행정등록자료(register data, 이하 행정자료)를 이용하는 방법이 있다. 순환형 센서스는 현재 프랑스와 미국에서 일부 활용되고 있는 방식으로 인구총조사의 주기에 따라 전체 기간 동안 지속적으로 조사하여 전체 자료를 확보하는 방법이다. 따라서 기존의 인구총조사와 마찬가지로 조사환경 악화에 따른 조사의 포괄성과 신뢰성에 대한 문제가 발생한다. 이런 이유에서 많은 나라

들은 순환형 센서스보다는 궁극적으로 행정자료를 인구총조사에 활용하는 방안을 모색하고 있다. 특히 덴마크나 핀란드 등 일부 국가에서는 이미 실제 조사를 하지 않고 행정자료로 대부분의 인구총조사 자료를 생산하고 있다. 따라서 행정자료를 기반으로 하여 국가통계를 생산하는 방법에 대한 기대가 높다.

많은 나라들이 행정자료를 활용한 통계생산방식을 선호하는 데는 또 다른 이유가 있다. 행정자료를 활용할 경우 비교적 저렴한 비용으로 원하는 통계를 빠른 시간 내에 확보할 수 있다.<sup>1)</sup> 행정자료를 이용할 경우 인구총조사와 유사한 자료를 매년 생산할 수 있다. 또한 이러한 자료를 바탕으로 지역통계 수요에 즉각 대처할 수 있다. 더 나아가 이와 같은 통계는 전 국민에 대한 패널자료이기 때문에 통계적 활용의 범위가 방대하다. 특히 개인, 가구, 사업체 등 사회 활동의 주체들이 어떻게 변화하는지를 추적할 수 있는 자료를 생산함으로써 다양한 인과적 통계분석을 할 수 있다.<sup>2)</sup>

이렇듯 행정자료를 기반으로 하는 인구총조사가 많은 장점을 가졌지만, 모든 국가가 당장 행정자료로 인구총조사자료를 대체할 수 있는 것은 아니다. 행정자료로 인구총조사자료의 조사항목을 생산하기 위해서는 각 행정부서별로 사용하는 행정자료를 연계·통합할 수 있도록 국가사회 전반에 걸쳐 행정 체제가 갖추어져야 하기 때문이다. 특히 모든 국민 개개인에 관한 기본 정보, 개인들이 거주하며 생

---

1) 예를 들어 핀란드에서는 질문지 조사를 사용하였던 1980년의 인구총조사에 3천4백만 유로의 비용이 소요되었는데, 행정자료를 사용한 2000년의 인구총조사에는 소요된 비용이 84만 유로에 불과했다(Statistics Finland, 2002). 특히, 최근 모든 행정자료들이 IT 기술로 데이터베이스화하고, 네트워크를 통해 행정정보들을 손쉽게 한군데 모을 수 있게 되었다. 따라서 각 부처별로 행정을 위해 축적한 정보를 IT 기술로 연계·통합하면 막대한 조사비용을 들이지 않더라도 인구총조사자료를 적은 비용으로 생산할 수 있는 근간이 자연스레 마련되었다고 할 수 있다.

2) 물론 행정자료에 기반한 인구총조사가 장점만을 가지고 있는 것은 아니다. 행정정보가 다양한 부처에서 생산되는 경우 동일한 항목을 중복하여 조사하는 경우가 많은데, 이때 개념 규정이 약간씩 차이를 갖게 되는 경우 이 차이를 조화시켜서 새로운 변수를 만들어야 한다(Spieker, 1999). 또한 법령 개정이나 정책의 변화에 따라 수집되는 자료에 관해 변화가 발생할 경우 통계의 연속성을 확보하기 어렵다는 문제가 있다.

활하는 단위인 개별 주거단위에 관한 정보가 행정부에 등록되어 있고, 잘 정비되어 있어야 하며, 정보의 형태 또한 서로 연계가 가능하도록 표준화되어 있어야 한다. 이와 더불어, 행정력이 사회 곳곳에 골고루 잘 퍼져 모든 구성원들의 사회·경제·정치·문화 활동과 관련된 각종 정보들이 해당 기관과 관련 행정 기관에 잘 기록·정비되어 있어야 한다.

우리나라의 경우 행정자료를 인구총조사에 활용할 수 있는 몇 가지 중요한 기반을 갖추고 있다. 첫째, 1962년부터 시행한 주민등록제도가 있다. 주민등록제도는 모든 국민 개개인을 파악할 수 있는 유일한 개인식별번호인 주민등록번호를 기반으로 하고 있다. 그리고 정부는 40년 이상 강력하게 이 주민등록제도를 시행해 오고 있다. 아울러 적어도 행정부서에서는 오류가 거의 없는 편으로 알려져 있다. 둘째, 세계 10위권 내에 들 정도로 높은 우리나라의 정보화 수준과 2000년부터 시작된 전자정부사업으로 행정자료를 연계·통합할 수 있는 기반이 잘 갖추어져 있다.

반면, 우리나라 행정자료 가운데 주거(생활)단위와 사업체를 파악할 수 있는 자료는 매우 불완전하다. 대표적으로 인구총조사통계의 주요한 단위인 가구를 파악할 수 있는 수준으로 주소체계가 정비되어 있지 않다. 사업체를 식별하는 사업자등록번호 같은 식별번호와 관련정보가 엄격하게 관리되고 있지 못하다. 특히 많은 소규모 사업체가 등록되어 있지 않거나, 등록오류가 많은 편이다. 이외에도 과세대장, 토지대장 등 많은 행정자료가 아직은 불완전하여 이들을 직접 연계하기 어렵다. 행정자료를 연계하기 위해서는 모든 자료를 정비하고 표준화하여 실제 행정에 활용하여야 하기 때문에 행정적으로 많은 노력과 시간이 필요하다. 따라서 현재는 손쉬운 부분에서부터 인구총조사에 행정자료를 활용하고, 앞으로 활용 과정을 거치면서 행정자료를 정비하고 표준화하는 장기적인 방안을 마련할 필요가 있다.

이 연구는 인구총조사에서 행정자료의 활용을 확대해 나아가기

위한 기초적인 작업이다. 특히 여러 형태의 행정자료 중에서 인구총조사에서 주택관련 항목과 밀접한 관련이 있는 건축물대장의 활용정도를 살펴보고자 한다. 현재 건축물대장의 자료는 전산화되어 건설교통부가 운영하는 건축행정정보시스템(AIS: Architectural administration Information System)에서 통합·관리되고 있다. AIS는 건축허가부터 착공, 사용승인에 이르는 건축행정업무를 전산처리하며, 행자부의 G4C시스템, 부동산등기시스템, 토지정보관리시스템, 4대 보험망 등 유관시스템과 연계하여 정보의 공동활용 범위를 확대하는 것을 목표로 삼고 있다.

현재 인구총조사는 조사수행의 측면에서 새로운 국면을 맞이하고 있다. 이전에는 조사대상가구들에 관한 대략의 정보를 읍·면·동사무소의 협조로 사전에 파악하여 조사구를 설정하고, 조사를 수행하였다. 그러나 2000년에 읍·면·동사무소의 기능이 축소되면서 2005년부터 읍·면·동사무소의 지원을 받을 수 없게 되었다. 이러한 문제를 해결하기 위해 통계청은 조사구 설정에 과거의 인구총조사자료, 건축인허가대장, 건축물대장, 한국감정원 자료 등을 활용하여 거처와 가구의 수를 추정하고 이를 바탕으로 기본적인 조사구를 설정하였다. 2005년도 인구총조사를 위한 시험조사(2004년 3, 7월)에서는 이렇게 만들어진 조사구를 실제 조사에 적용하였다.

이 연구는 시험조사에서 수집된 조사자료를 바탕으로 건축물대장의 활용도를 높이는 방안을 마련하는 데 목적을 둔다. 이를 위해 조사자료와 건축물대장이 얼마나 잘 상응하는지를 점검하고, 서로 상응하지 않는 경우를 줄일 수 있는 방안을 마련하고자 한다. 더욱이 두 자료가 서로 잘 상응하면 조사구 설정뿐만 아니라 인구총조사의 주택부문 항목에 건축물대장의 활용이 수월해지 때문에 건축물대장의 활용도가 높아지는 효과도 함께 있다. 이런 취지에서 이 연구는 2004년 7월에 실시한 5차 시험조사 결과와 이 지역의 건축물대장자료를 매칭(대응)시킨 자료를 분석의 주된 대상으로 한다. 이와 함께 인구총조사에서 행정자료를 폭 넓게 활용할 수 있도록 행정자료의

개선 방안도 함께 모색해보고자 한다.

## I. 행정정보 활용에 관한 기존 연구 검토

### 1. 국외연구

#### 1) 통계목적 행정정보 활용의 기본 모형

덴마크와 핀란드는 행정자료를 기반으로 각종 국가통계를 성공적으로 생산하는 대표적인 국가들로 알려져 있다. 덴마크와 핀란드의 센서스 방식은 행정정보를 활용하여 전통적 센서스의 여러 문제들을 해결할 수 있는 하나의 방안을 제시하였다. 국민들의 응답 부담은 사라진 반면, 과거 질문지를 사용해 조사되었던 거의 모든 항목들에 대해서 행정정보에 의한 통계를 산출하게 되었다. 자료의 질에서도 질문지 조사에 비해 나아진 것으로 평가되고 있다(Nordberg et al., 2001).<sup>3)</sup>

덴마크와 핀란드의 행정정보의 활용은 전통적 방식의 센서스가 가지고 있던 문제점을 해결했을 뿐만 아니라 전수에 바탕을 둔 자료를 연도별로, 지역별로 생산해 낼 수 있게 되었다는 점에서 새로운

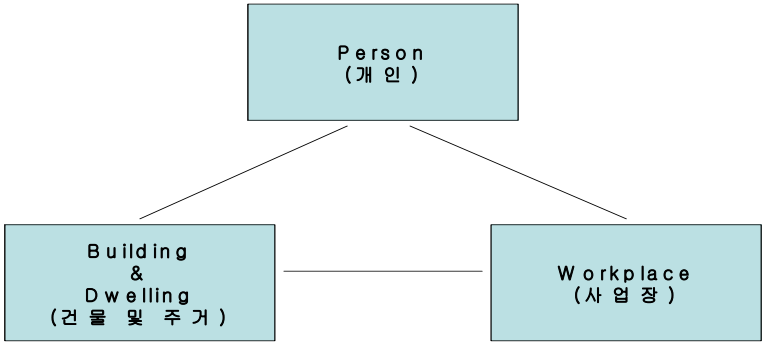
---

3) 덴마크는 1981년에 센서스를 전통적 전수조사에서 행정정보로 완전히 대체함으로써 세계에서 가장 먼저 행정정보 기반의 센서스를 실시한 국가가 되었다. 또한 인구총조사뿐만 아니라 모든 국가통계를 행정정보를 이용하여 생산하는 국가이기도 하다. 덴마크는 오랜 센서스의 역사를 가지고 있다. 덴마크는 1743년에 징병의 목적으로 출생과 사망 통계를 작성하였고, 1769년 이래 지속적으로 센서스를 실시해 왔다. 센서스에서 행정정보를 활용하는 기반은 1924년에 시행된 지방별 지방인구등록(Local Population Register)사업으로 마련되기 시작하였다. 핀란드 역시 인구총조사의 오랜 역사를 가지고 있다. 스웨덴의 식민지배 기간인 1749년에 스웨덴 정부는 식민통치의 목적으로 최초의 인구총조사를 실시하였다. 러시아 식민지배 기간인 1865년에는 통계국이 개설되었고, 1879년에 통계연감이 발간되었다. 이러한 역사는 핀란드 국민에게 통계의 중요성을 인식하는 바탕이 되었다. 이러한 배경에서 식민지배로부터 벗어난 1948년에 핀란드 정부는 최초로 국민총소득을 집계하였고, 1990년에는 덴마크에 이어 세계에서 두 번째로 행정등록 기반 센서스를 실시하였다(덴마크 통계청 내부자료, 핀란드 통계청 내부자료).

연구 가능성을 만들어 주었다. 소지역 추정 가능성이 되고, 변동에 관한 통계 산출이 용이해진 것이다(Myrsjykam 1999). 따라서 이들 국가들이 기본으로 삼고 있는 연구들을 살펴보는 것이 중요하다. 행정정보를 바탕으로 통계를 생산하기 위해서는 국민 개개인을 식별할 수 있어야 한다. 그들의 이름, 나이(생년월일), 국적, 성별, 혼인상태, 가족관계, 교육정도를 파악하여야 한다. 또한 지역별로 통계를 생산하기 위해서는 각 개인의 주거지를 알아야 한다. 이와 더불어 각 건물 및 주거지에 관한 센서스 항목을 파악할 수 있어야 한다. 또한 센서스를 통해 각 개인의 경제활동에 관한 정보를 알아야 한다.

이러한 정보를 획득하기 위해서 덴마크 통계청은 행정정보를 크게 개인(person), 주거 및 건물(dwelling and building), 사업장(workplace)에 관한 부분으로 나뉘어 파악하고 있다(<그림 1> 참조).

<그림 1> 덴마크 행정정보 구성체계의 기본모형



(출처 : 덴마크 통계청 내부자료)

## 2) 덴마크와 핀란드의 사례

먼저 덴마크의 경우를 살펴보면 다음과 같다. 덴마크 통계청은 위에 제시한 모형에 따라 행정정보를 크게 개인, 주거 및 건물, 사업장

에 관한 부분으로 나뉘어 파악하고 있다. 개인에 관한 정보는 개인의 이름, 주소, 주소변경, 가족관계, 출생, 사망, 결혼, 이혼, 이민, 입양, 시민권 취득, 인구등록 날짜 등으로 중앙인구등록에 포함되어 있다. 중앙인구등록은 내무부 주민등록청(Ministry of Interior, the Central Office of Civil Registration)이 총괄하여 관리하고 있으며, 자료의 유지·보수는 275개의 지방 인구등록사무소가 담당하고 있다.

주거에 관한 정보는 주로 건물/주거등록에 포함되어 있다. 건물/주거등록에는 건물의 주소, 크기, 토지 면적, 용도, 각 주거지의 주거 설비, 난방, 소유 등 건물과 주거시설에 관한 구체적인 정보가 있다. 건물/주거등록의 개별 건물과 주거지 관련 정보는 각 지자체가 수시로 갱신하며, 이를 중앙정부의 주택도시부(Ministry of Housing and Urban Affairs)가 종합하여 총체적으로 관리한다.

사업장(workplace)에 관한 정보는 사업등록(The Business Register)에 있다. 사업등록은 1970년대 초에 세금원천징수를 위해 만들어진 다양한 행정등록체계를 바탕으로 국세청(Tax Administration)과 중앙세관(Central Customs)이 관리·운영한다. 그러나 세금징수 목적의 행정등록은 기업(법인)을 중심으로 구성되어 있어 기업에 속한 각 사업장별 자료를 확보할 수 없다. 이런 이유에서 덴마크 통계청은 각 사업장별로 자료를 관리한다. 사업체행정 자료는 자영업자와 사업자별 정보로 구성되어 있고, 사업체 내에 개인에 대한 고용/피고용 상태, 상시/임시 고용형태, 직업, 산업, 소득 등이 포함되어 있다.

이러한 다양한 행정정보를 함께 사용하기 위해서는 이들을 서로 연계할 수 있도록 체계화된 등록번호체계가 필요하다. 먼저 전 국민을 식별하기 위해서는 우리의 주민등록번호와 비슷한 개인별 고유번호가 있어야 한다. 덴마크에서는 개인별 고유번호를 중앙인구등록번호(Central Population Register Number)라 부른다. 이 번호는 생년월일과 함께 무작위로 주어지는 번호, 논리적으로 번호를 점검할 수 있는 점검번호(check number) 등 11개의 숫자로 구성되어 있다.<sup>4)</sup>



전 국민의 주거지와 국가의 모든 건물들을 식별하기 위해서 건물 및 주거별 번호가 있어야 한다. 일반적으로 주소는 건물과 주거지의 식별을 위한 것이다. 그러나 주소는 개인의 신상정보를 보여줄 뿐 아니라 통계적으로 처리하기도 불편하다. 이런 이유에서 중앙인구등록이나 건물/주거등록 등의 행정자료는 부호화된 주소번호를 사용하고 있다. 주소번호는 지역번호 3자리, 도로번호 4자리, 건물번호 3자리, 건물 입구 1자리, 층 2자리, 주거입구 4자리 등 16자리의 체계적인 정보로 구성되어 있다.

또한 일에 관한 정보를 처리하기 위해서는 일하는 곳, 즉 사업체의 대한 식별번호가 있어야 한다. 행정적으로 사업체는 국세청이 관리하고 있다. 사업체는 민간기업, 중앙정부, 지방정부로 분류되며 각각이 고유한 번호체계를 갖고 있다. 통계청은 이를 통합하여 상응하는 통계청 자신의 고유번호체계를 만들어 관리한다.

덴마크 통계청은 개인, 주소, 사업체의 번호체계를 바탕으로 여러 종류의 행정정보를 연계하여 항목별 통계를 만든다. 예컨대 중앙인구등록번호로 국가 전체의 인구통계를, 주소로 행정구역별 인구통계를 산출한다. 또한 중앙인구등록의 주소를 바탕으로 가구를 구성하여 가구관련통계를, 가족관계 자료를 바탕으로 가족을 구성하여 가족관련 통계를 산출한다. 특히 정확한 가구구성을 위해 개인별 주소의 정확성을 확인하고 오류를 수정하는 과정에서 건물/주거행정자료를 사용한다. 고용관련 통계는 개인의 중앙인구등록번호와 사업체등록에 속한 개인의 중앙인구등록번호를 연계하여 산출한다.

한편 핀란드에서도 덴마크와 유사하게 인구총조사 자료를 산출하고 있다. 우선 행정정보를 연결하여 통계를 산출하기 위해서는 개별 행정정보를 서로 연결할 수 있는 번호체계가 필요하다. 그러한 번호체계는 개인과 개인이 속한 가구의 주거지, 개인이 속한 사업체의 번호와 연결하여 주거관련, 경제활동 관련 통계를 산출한다. 따라서 개

---

4) 등록번호의 구성요인과 구체적인 예는 <부록>에 제시되어 있다.

인별 중앙인구등록번호, 중앙인구등록번호가 포함되는 건물/주거 번호 및 법인/사업체 번호에 대한 체계가 있어야 한다. 중앙인구등록번호는 우리나라의 주민등록번호, 덴마크의 중앙인구등록번호와 매우 유사하게 생년월일, 일련번호, 점검번호로 구성되어 있다.

건물/주거번호(Domicile Code)체계는 건물번호와 주거번호 두 가지 종류의 번호로 이루어져 있다. 사업체번호는 국세청이 부여하고 있다. 하지만 통계청에서도 자체적으로 사업체등록번호를 부여한다. 국세청은 법인에게만 관리번호를 부여하기 때문에 개별 사업장에 대해서는 통계청이 자체적으로 번호를 부여하여 이용한다.

## 2. 국내연구

### 1) 행정자료와 인구총조사의 항목 비교

실제 통계자료를 생산하는 덴마크와 핀란드의 경우와는 달리 국내에서는 행정자료의 활용가능성을 탐색하는 수준의 연구들이 진행되어오고 있다. 특히 개인과 사업장과의 연계에 대한 정보를 얻기가 현재 힘들기 때문에, 개인과 거주지의 연계를 살펴보려는 연구가 시도되고 있다. 그 중에서도 건축물대장이 적어도 항목으로 볼 때, 인구총조사의 조사항목과 많이 중복되고 있다는 점을 주목하고 있다. 예를 들면, 김설희·강계화(2005)는 각종 행정정보와 인구총조사의 조사항목이 어떻게 구체적으로 연결될 수 있는지를 인구총조사 조사표의 항목과 그 항목이 있는 행정자료를 연계하여 제시하였다. 구체적으로 주민등록초본, 호적등초본, 이산가족 방북신청자 현황, 교육행정정보시스템, 장애인 증명서, 근로소득원천징수영수증, 건축물대장총괄표제부, 등기부 등본, 자동차등록증 등과 인구총조사의 조사항목을 연결하여 각 항목의 조사를 대치할 수 있는 방안을 제시하였다(<표 1> 참조).

물론 연구자들의 노력에도 불구하고, 현재 행정자료의 항목과 조

사항목의 일대일 대응으로 행정정보의 활용가능성을 파악하는 것은 단지 이론적 수준에서 가능성을 제시한 것뿐이다. 이런 의미에서, <표 1>에서 제시한 활용가능성은 조사항목과 동일한 내역의 항목이 주민등록과 같은 행정자료에 있다는 사실만을 확인한 것에 불과하다. 실제로 행정자료의 정보가 부실하거나 부정확한 경우, 행정기관이 실제 개별 등록 대상을 제대로 파악하지 못하는 경우 행정자료의 활용가능성은 표에서 제시한 수치보다 더 낮을 수 있다.

<표 1> 2005년 센서스 항목의 행정정보 활용가능성(%)

항목	활용 가능성	항목	활용 가능성	항목	활용 가능성
<b>가구에 관한 사항</b>		경제활동상태	50	거주 층	30
성명	100	종사상 지위	0	자동차대수	100
성별	100	산업	50	주차시설	0
나이	70	직업	0	난방시설	0
가구주와 관계	100	근로장소	50	접유형태	0
남북이산가족	70	혼인상태	80	임차료	0
교육정도	70	혼인 년 월	80	주인가구/소유	80
종교	0	총 출생아 수	80	<b>주택에 관한 사항</b>	
아동보육실태	0	추가계획 자녀 수	0	거처종류 및 층수	80
활동제약	70	고령자 생활비	0	연건평	80
5년 전 거주지	80	<b>가구에 관한 사항</b>		대지면적	80
통근 통학 여부	50	가구구분	100	충방수	80
통근 통학 장소	50	거주기간	100	건축년도	100
이용교통수단	0	사용 방수	40	편익시설수	80
통근통학시간	0	주거시설형태	20	전체단순평균	51.6

출처 : 김설희·강계화(2005)

## 2) 서울시 건축물 활용도 조사

건축물대장의 활용가능성을 다른 측면에서 검토해 본 연구는 서울시정개발원의 연구진에 의해 수행되었다. 이 연구에서 연구자들은 서울시의 건축물대장의 자료 효용성에 대해 검토하였다. 일반적으로 서울의 행정력이 다른 도시들이나 지역보다 더 발달된 것으로 알려

져 있다. 연구자들은 서울시 강남구 역삼동, 종로구 창신동, 서대문구 창천동 건축물대장의 항목별 누락률을 살펴보았다. <표 2>에서 제시되어 있듯이, 건축물대장의 누락률이 높은 편이다. 또한 전반적으로는 건축면적, 대지면적, 용적률, 건폐율 등 측정이 필요한 항목에서 누락률이 높다. 지역별로 비교하면 건축물대장의 항목별 누락 양상이 매우 다르다. 일반주택이나 다세대, 다가구 주택이 많이 밀집해 있는 서대문구 창천동 지역이나 종로구 창신동 지역이 대지면적, 구조(코드), 건물용도(코드)에서 아주 높다. 특히 창신동은 건축면적, 대지면적, 용적률, 건폐율 등 측정과 관련된 항목에서 누락률이 80% 정도로 매우 높다. 이에 반해 아파트가 많은 강남구 역삼동 지역은 대체로 잘 정비되어 있는 편이다(서울시정개발연구원, 2002).

<표 2> 서울시 건축물대장의 주요 항목별 누락 비율 (2002년)

동 항목	강남구 역삼동*		종로구 창신동**		서대문구 창천동**	
	누락수	비율	누락수	비율	누락수	비율
연면적	18	0.4	20	0.7	1	0.1
건축면적	1,803	38.9	2,484	81.7	407	24.0
대지면적	1,791	38.6	2,360	77.6	1,145	67.5
용적률	1,814	39.1	2,486	81.7	410	24.2
건폐율	1,812	39.1	2,486	81.7	410	24.2
구조(코드)	7	0.2	2,312	76.0	1,205	71.0
구조(명)	7	0.2	4	0.1	0	0.0
용도(코드)	18	0.4	2,480	81.5	1,505	88.7
용도(명)	18	0.4	7	0.2	0	0.0
준공년도	44	0.9	620	20.4	160	9.4
지상층수	17	0.4	38	1.2	5	0.3
전체수	4,636		3,042		1,696	

\* 건설교통부 건축행정정보시스템 기반의 건축물대장 자료 기준(2002.6.10)

\*\* 행정자치부 건축물대장전산화사업 기반의 건축물대장 자료 기준

출처: 서울시정개발연구원(2002).

## II. 행정정보 활용 가능성 확인을 위한 경험적 분석 결과

### 1. 분석 자료와 연결방법

#### 1) 분석자료

이 연구에서는 2004년 7월에 실시한 2005년도 인구총조사를 위한 5차 시험조사의 조사결과 중 3개 지역(A동, B면, C동)의 자료를 사용하였다. 자료는 시험조사 결과인 가구명부와 거처정보자료 및 해당 지역의 건축물대장으로 구성되었다. A동은 광역시의 고속도로 IC와 지하철 시발점이 위치하여 광역시의 관문 기능을 하는 지역으로서 자연부락과 아파트단지가 공존하는 도농 복합지역이다. B면은 농어업의 1차 산업을 중심으로 형성된 지역으로서 농어업인구가 전체 총인구수에서 차지하는 비중이 80% 이상이다. C동은 광역시의 노후 주택이 많은 주거지역으로서 아파트 밀집지역으로의 변화가 진행 중이었다.

<표 3> 지역별 가구명부, 거처정보, 건축물대장의 자료 수

지역	가구명부	거처정보	건축물대장
A동	4,135	4,137	1,571
B면	1,569	1,566	2,221
C동	3,150	3,143	7,040

#### 2) 자료의 연결방법과 문제점

분석을 위해서는 우선 가구명부와 거처정보자료 및 건축물대장의 자료를 서로 연결시키는 작업이 선행되어야 했다. 통계청에서 처음에 사용한 방법은 가구명부와 거처조사 결과를 연결한 후 이 자료에 건축물대장 자료들을 연결한 것이었다. 자료의 연결에는 데이터베이

스의 쿼리(query) 기능을 사용하였는데<sup>5)</sup>, 여러 가지 해결해야 할 문제점들이 있었다.

우선 통계청에서 사용한 쿼리 방식의 매칭으로 산출된 자료에는 중복자료가 존재하는 문제점이 있다. 특히 아파트 단지와 같이 동일한 지번에 여러 채의 건물이 존재할 경우 중복자료가 생성된다. 한 예로, 한 지번에 건물 5동이 있는 아파트 단지의 경우 이 아파트의 가구명부-거처 자료와 건축물대장을 지번으로 연결하면 한 거처에 그 지번의 주거용 건물 5개가 모두 연결되어 중복자료가 만들어지면서 자료수가 5배로 증가한다.

실제 거처 데이터베이스의 테이블에서 중복자료를 삭제하려면 두 단계가 필요하다. 먼저, 중복된 자료가 있는 테이블 구조의 복사본을 만들고 중복자료가 있는 모든 필드의 기본 키 (기본 키: 테이블의 각 레코드를 고유하게 식별하는 값을 갖는 하나 이상의 필드 혹은 열)를 정한다. 여기에서는 가구명부 ID가 기본 키로 사용된다. 그 다음으로 원래 테이블에서 기존 테이블의 끝에 쿼리 결과 집합의 레코드를 추가하는 실행 쿼리인 추가쿼리를 만들어 새 테이블로 실행한다. 이런 과정을 거치면 기본 키 필드에는 중복자료가 들어갈 수 없으므로 중복자료가 없는 테이블이 만들어진다.

그러나 이러한 방식의 자료 연결도 문제를 완전히 해결하지 못한다. 데이터베이스의 연결(join)기능으로 중복자료를 제거하면, 중복된 매칭 항목 중 임의의 매칭 항목만이 남게 된다. 중복자료 가운데 어느 하나를 임의로 골라 매칭하기 때문이다. 또한 매칭되지 않은 거처 항목들이 배제된다. 중복자료의 경우 매칭된 자료는 실제 거처와 건

---

5) 데이터베이스의 쿼리 기능을 이용한 자료 연결에서는 가구명부와 거처조사결과자료를 왼쪽 연결(join)하고 다시 오른쪽 연결을 하였다. 두 자료를 연결할 때 사용한 매칭값은 시도코드, 시군구코드, 행정동코드, 조사구 번호, 거처번호이다. 연결의 과정에서, 일단 왼쪽 연결을 통해 가구명부에 있는 모든 자료에 대해 이와 상응하는 거처자료의 개별 자료들이 모두 연결되었으며, 오른쪽 연결을 통해 나머지 거처자료가 추가되었다. 이렇게 만든 자료에 다시 건축물대장을 연결하였다. 여기에서 가구명부/거처정보를 기준으로 왼쪽 외부연결 방식을 사용하여 자료를 매칭하였다. 매칭값은 시군구코드, 법정동코드, 본번, 부번을 사용하였다.

축물이 정확히 연결된 것이 아니며, 건축물대장과 매칭되지 않은 경우는 자료 자체에서 사라지기 때문에 자료의 정확성이나 포괄성이 떨어진다는 점이다.

가구명부-거처 자료와 건축물대장을 연결하는 데 또 다른 문제가 있다. 이것은 가구명부와 거처 자료는 행정동을 사용하는데 반해 건축물대장은 법정동을 사용한다는 점이다. 하나의 행정동에 여러 법정동이 포함된 경우도 있고, 하나의 법정동에 다른 행정동이 함께 섞여 있는 경우도 많다. 따라서 가구명부와 거처자료에는 있으나 건축물대장에 없는 건물이 있으며, 반대로 건축물대장에는 있으나 가구명부와 거처자료에는 없는 건물이 있기 마련이다. 이 때문에 가구명부와 거처자료를 건축물대장과 연결할 때 연결되지 않는 부분이 발생할 수밖에 없다.

현재 데이터베이스의 기능을 기반으로 통계청 자료를 연결할 때 발생하는 문제를 해결할 수 있는 알려진 방법은 없다. 따라서 먼저 이러한 문제들의 양상이 어떠한지를 살펴보고, 그러한 문제의 원인들을 찾아보며, 문제들을 유형화하여 논리적으로 해결할 수 있는 기반을 마련한 다음, 궁극적으로는 정확한 매칭을 위한 알고리즘을 개발하는 것이 문제를 해결하는 순서이다.

이 연구에서는 먼저 자료 연결에서 발생하는 다양한 문제를 파악하기 위해 수작업을 통해 매칭 자료를 만들었다. 수작업 매칭에서는 가구명부와 거처조사결과를 연결한 자료와 건축물대장 자료를 직접 하나하나씩 대조하고 연결하여 자료를 생성하였다. 수작업을 통해 개별 자료마다 발생하는 문제를 파악하고, 그러한 문제의 원인을 찾아보며, 논리적으로 의미 있는 유형으로 구분하고자 하였다.

## 2. 가구명부-거처정보와 건축물대장의 매칭 유형

지번을 매칭값으로 사용하여 가구명부-거처 정보와 건축물대장 정보를 매칭할 때 발생 가능한 여러 유형들을 나누어 볼 수 있다. 우

선 거처가 소재하는 지번에 건축물대장이 존재하는지에 따라 매칭이 된 경우와 매칭이 안 된 경우를 구분할 수 있다. 매칭이 된 경우에도 가구명부 및 거처조사 자료와 건축물대장 정보가 매칭이 되는 유형에 따라 다음의 <표 4>와 같이 세분화할 수 있으며 이에 따라 건축물대장 정보의 활용 정도와 방법이 크게 달라진다.

<표 4> 가구명부-거처와 건축물대장이 지번으로 매칭 되는 유형: 거처기준

매칭 여부	유형	내용	건축물대장 정보의 활용 가능성
매칭	유형1	거처 1곳과 건축물정보 1건이 매칭	추가 작업 없이 활용
	유형2	거처 다수와 건축물정보 1건이 매칭된 대단위 공동주택	건축물 동수 구별 후 활용
	유형3	대단위 공동주택이 아니면서 거처 다수와 건축물정보 1건이 매칭된 경우	공동주택 확인 후 활용
	유형4	거처 1곳에 1개의 주택과 다수의 기타용도 건축물정보가 매칭된 경우	주택 이외의 건축물이 주택의 부속건물인 경우 활용
	유형5	거처 1곳에 다수의 주택이나 시설에 대한 건축물정보가 매칭된 경우	실제 조사에서 대조자료로서 활용
	유형6	한 지번에 다수의 거처와 다수의 건축물정보가 동시에 매칭된 경우	실제 조사에서 대조자료로서 활용
비매칭	유형7	거처에 대한 건축물정보가 없는 경우	활용 불가능

유형1은 거처 1곳과 건축물대장 1건이 일대일로 매칭된 경우이다. 이 경우 이론적으로는 별다른 추가 작업 없이 건축물대장에 있는 정보를 거처에 그대로 적용할 수 있다. 하지만 이 경우에도 건축물대장의 건물과 지번에 소재하는 거처가 동일 건물인지에 대한 검증이 필요하고, 건축물대장 정보의 정확성에 대한 평가가 필요하다.

유형2는 다수의 거처에 건축물대장 1건이 매칭되는 대단위 공동주택의 경우이다. 한 지번에 여러 동의 아파트와 상가동이 존재하는



경우 한 거처에 여러 동의 건축물대장 자료와 상가동의 건축물대장 자료가 모두 매칭 된다. 데이터베이스 프로그램의 연결 기능을 사용할 때 다수의 거처에 다수의 건축물대장 자료가 매칭 되는 것이다. 따라서 한 거처에 해당하지 않는 다른 건축물대장 항목들까지 중복되어 매칭 된다. 하지만 대단위 공동주택의 경우 거처에 동수가 표시되어 있고, 건축물대장도 동별로 구별이 가능하기 때문에 원칙적으로는 거처에 해당하는 건축물대장 정보를 사용할 수 있다. 이 경우 효과적인 매칭을 위해 현재 사용한 것과 다른 매칭 방법이 필요하다. 예컨대 거처의 동 명칭과 이에 해당하는 건축물대장 항목을 매칭값으로 사용하면 좀 더 나은 매칭 결과를 얻을 수 있다.

유형3은 다수의 거처와 1개의 건축물대장 자료가 매칭 된 경우로 대단위 공동주택이 아닌 경우이다. 이 경우 다세대주택이나 다가구주택일 수 있으나, 일부 거처에 대한 건축물대장 자료가 누락된 경우도 배제할 수 없다. 다세대주택, 다가구주택의 경우 건축물대장 자료를 해당 지번에 있는 모든 거처에 적용할 수 있다. 하지만 일부 거처에 대한 건축물대장 자료가 누락됐을 때는 이러한 방법이 가능하지 않다. 이 경우 건축물대장의 정보를 사용하기 위해서는 다세대주택이나 다가구주택인 경우와 일부 거처에 대한 건축물대장 자료가 누락된 경우를 구별할 수 있는 방법이 개발되어야 한다.

유형4는 거처 1곳에 다수의 건축물대장 자료가 매칭 된 경우로 매칭 된 건축물대장 항목들이 주택과 다른 용도의 건물(변소, 창고 등)로 구성된 경우이다. 이 경우에 다른 용도의 건물과 주택의 관계를 파악하는 것이 필요하다. 다른 용도의 건물이 변소인 경우 주택에 딸린 부속건물일 가능성이 크다. 하지만 창고나 생활근린시설 등인 경우 다른 용도의 건물이 주택에 딸린 부속건물일 수도 있고 주택과는 별개의 건물일 수도 있다. 따라서 이에 대한 구분은 실제 조사에 의존할 수 밖에 없다.

유형5는 거처 1곳에 다수의 주택이나 다수의 시설에 대한 건축물대장 자료가 매칭 된 경우이다. 이 경우 유형4처럼 거처가 실제 다수

의 주택으로 구성되었을 가능성을 배제할 수 없다. 예컨대 주택에 별채가 따로 있는 경우 거처 1곳에 다수의 주택에 대한 건축물대장 자료가 존재할 수 있다. 하지만 해당 지번에 여러 개의 독립된 주택이나 여러 개의 독립된 시설들이 있는 경우에도 이러한 매칭 유형이 나타날 수 있다. 실제 매칭 유형이 이 중 어느 경우에 해당하는지를 알기 위해서는 건축물대장에 있는 정보만으로는 판정할 수 없으며 실제 조사가 필요하다<sup>6)</sup>.

유형6은 한 지번에 거처와 건축물대장 자료가 모두 다수로 매칭된 경우이다. 이 경우 거처에 해당하는 건축물대장 항목을 매칭하기 위해서는 실제 조사가 필요하다.

유형7은 거처에 해당하는 건축물대장 자료가 존재하지 않는 경우이다. 이 유형은 건축물로 등록이 되지 않은 무허가 건물인 경우, 등록된 건축물대장 정보가 전산입력시 누락된 경우, 법정동이나 지번이 잘못 입력되어 건축물대장 항목을 매칭하지 못한 경우 등에 의한 것일 수 있다.

건축물대장 정보의 활용이라는 점에서 볼 때 유형1과 유형2은 비교적 활용 가능성이 높은 경우라 할 수 있다. 이 경우 건축물대장 정보를 그대로 활용하거나 매칭 알고리즘의 개발을 통해 사용할 수 있다. 유형3과 유형4은 실제 거처가 공동주택이나 부속건물을 포함하는 주택인지를 확인할 수 있다면 건축물대장의 정보를 적극 활용할 수 있다. 실제 조사의 결과를 바탕으로 할 때 조건부로 건축물대장 정보를 활용할 수 있는 경우이다. 유형5와 유형6에서도 실제 조사의 결과에 대한 보완자료로서 건축물대장 정보를 활용할 수 있지만, 건축물대장과 실제 거처의 매칭 관계가 복잡한 경우로서 건축물대장의 정보 활용이 훨씬 제한된다. 유형7에서는 거처에 해당하는 건축물대장 자료가 없기 때문에 건축물대장 정보를 활용할 수 없다. 하지만

---

6) 유형4와 유형5는 거처 1곳에 다수의 건축물대장 자료가 대응한다는 점에서 유사하다. 하지만 유형4의 경우 주택 이외 다른 건물의 기타 용도가 부속건물로 추정되는 반면 유형5에서는 이러한 추정이 가능하지 않다.

자료의 매칭이 되지 않는 이유가 자료 입력이나 조사과정의 오류인 경우에는 오류 수정 후 건축물대장의 정보를 활용할 수 있다. 이 경우에도 최소한 조사과정의 오류를 검토해 보는 계기로서 건축물대장 정보를 활용할 수 있다.

### 3. 매칭 결과 분석

#### 1) 지역별 비교

지역별로 매칭 유형을 비교한 결과를 <표 5>에 제시하였다. 지역별로 매칭 유형의 차이가 뚜렷하게 나타났다. 거처들 중 매칭이 되지 않은 경우인 유형7을 제외한 다른 유형들은 일단 거처정보와 건축물대장이 매칭이 된 경우들이다. B면에서는 매칭률이 60%에 못 미친 반면, 광역시의 주거지역인 A동과 C동에서는 90% 이상의 높은 매칭률을 보였다. 이는 농촌지역에 비해 아파트, 연립주택, 다세대주택 등 공동주택이 많은 대도시 주거지역에서 건축물대장 자료의 사용가능성이 높다는 사실을 보여준다. 농촌지역에서는 건축물대장 자료의 사용가능성이 현저히 떨어졌는데, 이는 건축물대장 정보와 실제 건축물 사이에 괴리가 크기 때문인 것으로 보인다.

<표 5> 가구명부-거처조사자료와 건축물대장의 지역별 매칭유형 분포

지역 유형	A동		B면		C동	
	빈도	퍼센트	빈도	퍼센트	빈도	퍼센트
유형1	257	6.2	449	28.7	406	12.9
유형2	3,439	83.1	10	0.6	1,803	57.4
유형3	61	1.5	181	11.6	9	0.3
유형4	4	0.1	16	1.0	659	21.0
유형5	11	0.3	67	4.3	14	0.4
유형6	69	1.7	203	13.0	11	0.3
유형7	295	7.1	640	40.9	241	7.7
계	4,136	100.0	1,566	100.0	3,143	100.0

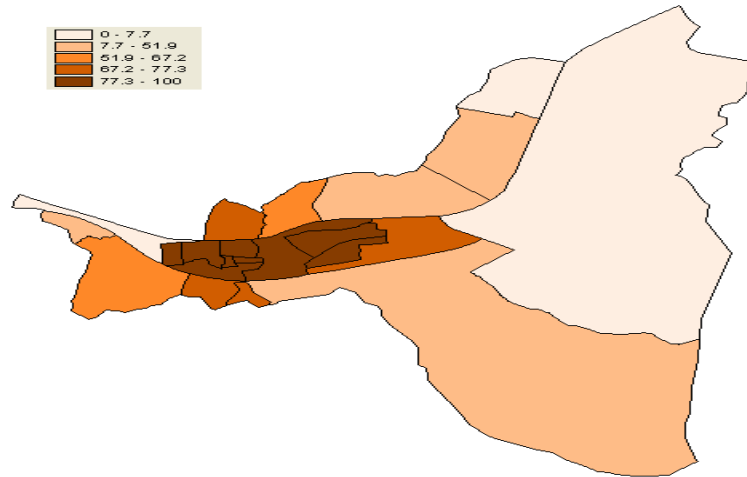
대단위 공동주택의 매칭유형인 유형2는 B면에서 두드러지게 적었던 반면, 다수의 거처에 한 개의 건축물대장 자료가 매칭 되는 경우인 유형3은 B면에서 상대적으로 많았다. 하나의 거처에 주택과 부속건물로 추정되는 건축물대장 자료가 매칭 된 경우인 유형4는 C동에서 두드러지게 많았고 다수의 거처에 다수의 건축물대장 자료가 매칭 되는 경우인 유형6은 B면에서 가장 많았다.

## 2) 조사구별 심층 비교 분석 : A동의 사례

매칭 유형의 차이는 세 지역의 내부에도 존재한다. 지역 내부에서 나타나는 매칭 현황의 차이를 알아 보기 위해 A동의 조사구별 매칭 현황을 비교해 보았다. 특히 다른 조사구에 비해 현저하게 매칭률이 떨어지는 경우는 그 이유를 파악하는 것이 중요하다. 이러한 목적을 위해 A동에서 총 73개의 조사구 중에 5개의 조사구에 대하여 심층적인 분석을 진행하였다. 심층분석을 통해 매칭률이 낮은 이유에는 여러 가지 유형이 있음을 알 수 있었다<sup>7)</sup>.

7) 조사구별 상세 정보는 지면의 제약으로 생략한다. 필요한 경우 요청에 따라 상세 결과를 별도 제시할 수 있다.

<그림 1> A동 매칭률



먼저 조사구 a의 경우, 매칭이 되지 않는 대부분 경우에 건축물대장에서 해당 지번에 대한 정보가 누락되어 있다. 많은 경우에 1950년 이전에 사용허가가 난 건물이라는 점에서 건축물대장 정보의 정확성과 관련이 깊은 것으로 추측된다.

조사구 b의 경우는 매칭률이 10%미만으로 다른 조사구에 비해 현저하게 낮다. 여기에는 다양한 이유가 존재한다. 첫째, 매칭할 때 사용한 정보에 문제가 있다. 시군구코드, 법정동, 본번, 부번 등 네 가지 변수로 이루어졌기 때문에 하나의 정보라도 일치하지 않으면, 매칭이 되지 않는다. 둘째, 거처정보의 지번 정보에 문제가 있다. 물론 이것은 조사자의 오류일 수도 있고, 입력자의 오류일 수도 있다.

조사구 c의 경우는 8개 거처에서 해당 지번의 건축물대장에는 건축물이 없었다. 또한 지도에서도 해당하는 건축물이 없기 때문에 거처정보와 매칭 되지 않는다고 판단하였다.

조사구 d의 경우도 모든 거처가 건축물과 매칭이 되지 않았다. 다섯 개의 거처가 존재하지만 거처정보에 있는 해당 지번에 어떠한 건

축물 정보도 없다. 거처정보의 지번 정보에 문제가 있다는 점을 추측할 수 있다.

다음으로는 거처 특성별로 매칭 유형 차이를 살펴보았다. 아파트의 경우는 100% 매칭이 되어 건축물대장의 활용 가능성이 가장 높다. 연립주택의 경우는 약 80.2%가 공동주택으로 다대일 매칭이 되고 있다. 단독 주택의 경우는 절반 정도가 건축물대장과 일대일 매칭이 된다. 이러한 차이는 아파트와 연립주택 등 공동주택이 84.5%에 이르는 A동에서 매칭률이 92.6%, 공동주택이 62.4%인 C동에서 92.6%인 반면, 공동주택이 없는 B면에서는 58.9%에 그친 것과도 궤를 같이 한다. 또한 서울시가 건물도형과 건축물대장을 매칭 했을 때도 비슷한 패턴으로 나타나 등록된 자료가 도형이든 속성이든 규모가 큰 공동주택일수록 잘 정비되어 있음을 시사해준다.

건축물대장의 빈집과 건축 연도가 매칭률과 관련이 있는 것으로 보인다. 대부분의 빈집은 공동주택으로 다대일 매칭을 보이고 있다. 그리고 건축연도가 1959년 이전인 거처의 경우는 일대일 매칭률이 상대적으로 낮다. 거처의 종류와 건축 연도 같은 거처 특성에 따라 매칭률과 매칭 유형이 다르기 때문에 일률적으로 건축물대장을 적용하는 것은 현재로서는 무리가 있을 수 있다. 1959년 이전의 노후 건물과 단독주택, 상가건물에 존재하는 거처에 관해 실상을 파악하기 위해서는 조사원에 의한 직접적인 조사가 바람직하다.

#### 4. 건축물대장 활용에 대한 평가

##### 1) 자료의 매칭에서 건축물대장 정보의 활용 정도

자료의 매칭에서 건축물대장이 활용된 정도를 평가해 보면 전체 건축물대장 자료 중 일부분만이 가구명부 및 거처조사결과와 매칭되었다. A동과 C동에서는 각각 건축물대장 자료 중 27.3%, 28.1%만이 매칭 되었다. B면에서는 매칭률이 다소 높았으나 36.3%에 불과했다.

건축물대장 자료 중 매칭 되지 않는 건축물이 많은 것은 주택용도가 아닌 건축물의 비율이 매우 높기 때문일 가능성이 있다. 하지만 매칭 되지 않은 건축물대장 자료들 중 주택용도의 건축물이 차지하는 비율이 상당히 높음을 볼 수 있다<sup>8)</sup>.

또한 가구명부와 거처정보가 행정동을 사용하는데 반해 건축물대장은 법정동을 사용하여 주소가 일치하기 않기 때문에 나타나는 결과일 가능성 역시 있다. 그렇지만 행정동과 법정동의 불일치 문제가 나타나지 않는 B면의 경우에도 매칭률이 높지 않다는 사실은 낮은 매칭률에 대한 다른 원인이 있음을 시사한다.

첫째, B면의 경우 대도시에 비해 건축물대장이 잘 정비되지 않았을 가능성이 있다. 둘째, 건축물대장에 등록되어 있는 주택 소재 거처들에 대한 조사가 불충분하게 진행되었을 가능성이 있다. B면이 대도시 행정구역에 비해 상대적으로 넓고 외딴 곳에 있는 거처 등 접근이 어려운 곳이 많아 실제 조사가 제대로 이루어지지 못했을 가능성이 있다. 이런 경우 건축물대장의 자료를 유용한 준거 자료로서 사용할 수 있다.

## 2) 건축물대장의 건축물 주용도에 따른 매칭 유형

건축물대장에 명기된 건축물의 주용도에 따라 자료의 매칭유형이 어떻게 나타나는지를 살펴보았다. 여기에서는 용도코드에 따라 1000 단위의 용도코드를 ‘단독주택’으로 분류하였으며, 2000 단위를 ‘공동주택’으로 포괄하였고 그 밖의 코드를 ‘기타시설’로 구분하였다. 분석의 대상으로 삼은 것은 매칭이 된 건축물대장 자료들이다. 이들 매칭된 건축물대장 자료들이 세 가지로 구분한 건축물의 주용도에 따라 매칭 유형에 차이가 어떻게 나타나는지를 살펴보았다.

여기에서 특히 주목할 것은 기타시설 용도의 건축물대장 자료들이었다. 건축물대장에 명기된 주용도를 바탕으로 거처 수를 추정할 경우 단독주택이나 공동주택 이외의 목적으로 용도가 분류된 건축물

---

8) A동의 경우 주택용도의 건축물이 매칭 되지 않은 건축물대장 자료들 중에서 차지하는 비중이 82.7%에 달하고, B면에서도 60.8%에 달하며, C동에서는 88.1%에 달한다.

들은 거처가 아닌 건축물로 추정된다. 하지만 거처와 매칭 된 기타시설 용도의 건축물대장 자료들 중에서 일대일 매칭의 유형1에 해당하는 경우가 상당히 많았다. A동에서는 매칭 된 기타시설 용도의 건축물대장 자료의 73.8%가 일대일 형태로 거처 정보와 매칭 되었다. B면에서도 기타시설 용도의 건축물대장 자료 중 유형1에 속하는 비율이 50.0%나 되었다. C동에서는 기타시설 용도의 건축물대장 자료 중 유형1에 속하는 비율이 34.0%로 비교적 적었으나, 이는 주택이 부속건물을 가진 것으로 추정되는 유형4의 비율이 57.9%로 많았기 때문이다.

이러한 사실은 건축물대장의 주용도를 바탕으로 거처수를 추정할 경우 그만큼 오차의 가능성이 클 수 있음을 시사한다. 또한 이러한 현상이 발생하는 원인에 대해 다음과 같은 추정을 가능케 한다. 첫째, 주용도가 기타시설이지만 기타용도에 주택이 포함되어 있을 수 있다. 둘째, 지번이 같아서 건축물대장이 거처와 매칭은 되었으나 실제 거처와는 상관없는 다른 건축물일 가능성이 있다. 셋째, 기타시설 용도로 건축물대장에 명기된 건축물들이 실제 거처로 활용되고 있을 가능성이 있다. 이러한 여러 가지 가능성들을 좀 더 정확히 추정하기 위해서는 건축물대장에서 기타용도까지 고려한 자료의 매칭방법이 필요하다.

### 3) 거처정보와 건축물대장 정보 비교

거처정보와 건축물대장의 정보를 비교함으로써 앞으로 행정자료를 이용하는데 유용성을 평가하였다. 이를 위해서 A동의 경우 일대일 매칭이 된 287개 건물만을 선택하여 평가하였다<sup>9)</sup>. 거처와 매칭 된 건축물대장의 건축물에 대한 동연면적 정보가 없는 약 50여개의 건물을 제외하고 최종적으로 233개 거처만을 비교하였다.

9) 일대일 매칭이 된 건물만을 선택한 이유는 거처정보와 건축물대장의 정보를 비교하기 위해서는 동일한 건물이라는 전제 조건이 있어야 하기 때문이다. 그런데 적어도 현재의 건축물대장에서는 공동주택의 각 호수별 정보를 구하기 힘들다. 따라서 여기에서는 두 가지 자료가 정확하게 일대일 매칭이 되는 단독 주택만을 선택하였다.



거처정보에는 모두 21개의 변수(필드)가 포함되어 있다. 한편 건축물대장에는 건축물에 대한 상세한 정보를 포함하여 47개의 변수가 포함되어 있다. 그런데 이렇게 다양한 변수가 있음에도 불구하고 정작 비교할 수 있는 변수는 한정되어 있다. 더욱이 변수의 정의 자체로는 비교가 가능하지만, 실제로 자료가 없는 경우가 많다. 여기서는 이러한 문제를 감안하여 거처정보의 연건평과 건축물대장의 동연면적을 선택하여 비교하였다.

A동의 경우 거처정보의 연건평과 건축물대장의 동연면적을 대조해본 결과 차이가  $\pm 4$  평 이내에 드는 거처가 22.3%(52/233),  $\pm 9$ 평 이내에 드는 거처가 40.8%(95/233)에 그쳐 전체적으로 조사한 자료와 행정자료 사이에는 많은 차이가 있는 것으로 나타났다. 이러한 차이는 어느 쪽으로든 해석이 가능하다. 만일 거처정보의 정보가 사실에 가깝다면, 건축물대장의 정보를 매우 제한적으로 활용할 수밖에 없다. 반면에 조사에서 오류가 발생한 것이라면 건축물대장의 활용도는 높아질 수 있다. 만약 건축물대장의 정보가 좀 더 사실에 근접한 정보라면, 건축물대장의 정보를 조사에 직접 활용할 수 있을 것이다. 활용에는 여러 가지 방법이 있다. 조사원들이 조사대상에 대한 건축물대장의 관련 정보를 조사에서 확인하는 방식을 취한다면 조사시간이나 비용을 줄일 수 있을 것이다.

이렇듯 두 가지 상반된 가능성으로 말미암아 건축물대장의 정보를 조사에 직접 활용할 수 있는지에 대한 판단은 실제조사와 건축물대장의 정보 가운데 어느 것이 더 정확한지를 파악한 후에야 내릴 수 있다. 이런 측면에서, 이 연구는 건축물대장과 조사 결과의 정보에서 차이가 많이 나는 거처를 다시 한번 조사하여 연구할 필요가 있음을 확인하였다. 후속 연구에서는 그러한 차이의 현황과 원인을 파악하고, 더 나아가 건축물대장과 조사 정보의 정확성에 대한 판단을 내릴 근거를 마련하는 것이 주된 목적이 될 것이다.

### III. 결론 및 제언

#### 1. 가구명부-거처정보와 건축물대장 자료 매칭에 대한 평가와 시사점

데이터베이스의 매칭을 활용하여 건축물대장 정보를 효과적이고 정확하게 활용하기 위해서는 기술적인 측면에서 다음과 같은 사항들의 개선이 필요하다.

첫째, 건축물대장에 있는 정보를 매칭 작업에 좀 더 많이 활용할 수 있는 알고리즘의 개발이 필요하다. 현재 가구명부 및 거처 조사의 자료와 건축물대장 정보를 매칭 하는데 사용하는 정보는 시군구코드, 법정동코드, 본번, 지번의 네가지 항목이다. 효율적인 매칭을 위해서는 거처의 종류를 고려하여 거처의 종류에 따라 다른 매칭값을 사용하거나, 주소와 함께 공동주택명, 동번호, 동명칭을 동시에 매칭값으로 사용하면 수작업과 유사한 매칭의 효과를 얻을 수 있을 것이다. 자료 매칭의 방법을 개선하기 위해서는 국내외에서 시도된 다양한 데이터 매칭 기법과 시스템을 파악하는 것이 바람직하다. 예컨대 캐나다 통계청에서는 지난 1980년대 Fellegi와 Sunter의 이론(Fellegi and Sunter, 1969)을 바탕으로 데이터 매칭 시스템을 개발한 이래 다양한 이론들을 기반으로 한 시스템을 개발해 왔다(Newcombe, Fair and Lalonde, 1992).

둘째, 조사 과정 및 자료 입력에서 나타나는 오류를 방지할 수 있는 시스템 구축이 필요하다. 거처정보 자료를 작성할 때 여러 가지 단계에서 오류가 발생할 수 있다. 조사 단계에서 뿐만 아니라 입력 단계에서도 오류가 발생할 수 있다. 따라서 앞으로 조사에서는 이러한 오류를 최소화하는 방식으로 조사와 입력 과정을 재구조화할 필요가 있다. 예를 들면, 해당 조사구에 존재하지 않은 정보를 입력할 때, 오류 발생 등을 알려줄 수 있는 시스템을 구성하여, 개인적인 오류의 발생을 최소화해야 할 것이다.

## 2. 행정자료 활용을 위한 개선방안

### 1) 건축물대장

자료간 매칭이 개선되더라도 행정자료들의 오차 수준과 유형이 정확히 분석되지 않고 조사에 활용할 경우, 각각의 자료가 가진 질적인 문제점들이 종합적으로 나타날 수 있다. 이 연구에서 살펴본 주된 행정자료인 건축물대장에서도 여러 가지 문제점을 발견할 수 있었다. 실제로 존재하는 건축물이 등록되지 않은 경우, 반대로 이미 허물어졌는데 대장에서는 말소되지 않은 경우, 신축되거나 변경되었으나 대장에는 반영되지 않은 경우, 건축물대장의 항목들 가운데 누락이 많은 경우, 기재된 정보가 부정확한 경우 등 다양한 형태의 오류들을 발견하였다. 또한 그러한 오류가 상당히 많은 자료에서 발생하며, 지역별로 발생 빈도의 차이가 큰 것도 알 수 있었다.

이러한 문제를 개선하기 위해서는 전국적으로 건축물대장을 정비해야 한다. 간단한 입력오류에 대한 수정 작업은 물론 입력 내역에 대한 표준화 문제를 해결해야 한다<sup>10)</sup>. 또한 건축물대장의 입력과 갱신에 대한 표준 지침이 있어야 한다. 변동이 발생할 경우 언제 입력할 것인지, 지자체단체에서 데이터베이스를 갱신하는 시기는 언제이며 중앙부처에서 갱신하는 시기는 언제인지에 대한 정해진 규칙을 만들어야 한다.

이런 제도적 정비와 함께 전체적인 정비를 위해서는 인구센서스와 같은 방식으로 짧은 기간 동안 일시에 전국의 건물을 조사하여 건축물대장의 내역을 정비하는 것이 효율적이며 그 효과 또한 클 것이다. 그렇지 않으면, 시간 차이에서 오는 오류들로 대장을 정비하는데 혼란이 있어 건축물대장을 정비하는 효과가 감소할 위험이 있기 때문이다. 이런 측면에서, 연구자들은 건물센서스(building census)를 제안한다. 건물센서스의 목적은 건물에 대한 파악 뿐 아니라 건물

---

10) 예를 들면 여러 필지를 묶어 한 건물이 들어선 경우 어떤 지번을 사용할 것인지에 대한 표준 지침이 있어야 할 것이다.

내 사업장 공간(establishment unit)과 거처(dwelling unit)를 파악하고자 하는 것이다.

## 2) 새주소체계

우리나라의 주소체계는 현재 행정자료를 활용하고 있는 나라뿐만 아니라 대부분의 유럽과 미국에서 사용하는 주소체계와 매우 다른 구조를 갖고 있다. 일단 우리나라의 주소체계는 행정동과 법정동으로 이분화 되어있어 혼란을 일으킨다.

이런 혼란에 대응하여 정부가 만들기 시작한 새주소체계는 토지와 건물을 분리하여 지번 중심의 구주소는 토지에만 사용하고, 건물에는 길 이름에 건물별로 번호를 부여하여 사용하는 것을 골자로 한다. 새주소체계는 찾기 쉬운 주소체계를 제공하는 것 이외에 부가적으로 발생하는 이점들을 가지고 있다.

도로를 따라 건물마다 번호를 붙이기 때문에 자연스럽게 건물의 고유번호가 만들어진다. 만약 건축물대장에서 구주소 대신 새주소를 사용하면 기존의 지번보다 상세하고 정확한 건물에 관한 자료를 관리할 수 있다. 또한 새주소사업의 하나로 진행된 수치지도를 이러한 건물 정보와 함께 사용하면 지리정보시스템(Geographical Information System) 기반의 행정서비스 뿐만 아니라 통계자료를 파악하기 쉽게 가공하고 자료를 시각적으로 확인하는 부수적 효과를 가질 수 있다.

새주소체계에서의 유용성을 높이기 위해서는 건물 안의 거처까지 번호를 부여해서 확인 수준을 확대할 필요가 있다. 이와 함께 현 새주소시스템의 번호체계에서 사용하는 숫자가 너무 많다<sup>11)</sup>. 따라서 새주소체계의 번호체계를 새로 설계할 필요가 있으며, 통계적 목적을 고려하여 숫자의 단위를 가능하면 적게 하는 것이 바람직하다.

---

11) 건물까지만 주소를 부여하는 데도 27개의 숫자로 이루어져 있다. 여기에 거처 수준, 가구 수준까지 확대된다면 숫자의 단위는 족히 30개를 넘어설 것이다. 반면 외국의 경우 가구 수준까지 파악할 수 있도록 주소체계가 되어 있음에도 덴마크는 18개, 핀란드는 23개의 숫자로 이루어져 있다.

## 참고문헌

- 김설희·강계화(2005), “GIS DB에 연계시킨 행정정보를 활용한 현장조사 품질제고 방안,” 한국조사연구학회 2005년도 춘계학술대회 발표논문집. 서울시정개발연구원(2002), 『공간데이터웨어 하우스 구축 시범사업(조사)』.
- Bierau, D.(2000), “Testing a Register-Based Census”, Insee-Eurostat Seminar on Census after 2001, Paris.
- Fellegi, I.P. and Sunter, A.B.(1969), “A Theory of Record Linkage,” Journal of the American Statistical Association, 40, 1183-1210.
- Jensen, L.(2000), “Opening speech”, Insee-Eurostat Seminar on Census after 2001, Paris.
- Myrskylä, P.(2000), “Effects of Changes to Registers: Experiences from Finnish Register-based Censuses”, Insee-Eurostat Seminar on Census after 2001, Paris.
- Newcombe, H.B., Fair, M.E. and Lalonde, P.(1992) “The Use of Names for Linking Personal Records,” Journal of the American Statistical Association, 87, 1193-1208.
- Nordberg, L., Irmeli P., and S. Sandstrom(2001), “A Study on the Effects of Using Interview versus Register Data in Income Distribution Analysis with an Application to the Finnish ECHP-survey in 1996”, CHINTEX Working Paper #1.
- Spieker, F.(1999), “Formation of Central Variables in a Decentralised Statistical System”, Working Paper No. 25. Joint ECE/Eurostat Work Session on Registers and Administrative Records in Social and Demographic Statistics. Geneva.
- Statistics Finland(2002), “Use of Administrative Data in Population Censuses-Definition of Main Type of Activity as an Example”, Working Paper No. 19. UNECE Work Session on Statistical Data Editing. Helsinki.

## < 부 록 >

<부록 표 1> 덴마크의 중앙인구등록번호 체계

번호	태어난 날짜	태어난 달	태어난 해	태어난 세기	무작위 번호	점검번호 (성별포함)	총 자릿수
자릿수	2	2	2	1	2	1	10

<부록 표 2> 덴마크의 주소번호 체계

내역	지역 Municipality Code	도로 Road Code	건물 House Number	건물입구 House Letter	층 Floor	주거입구 Side/Door Number	총 자릿수
자릿수	3	4	3	1	2	4	17

<부록 표 3> 핀란드의 중앙인구등록번호 체계

번호	0	5	0	1	5	9	-	5	0	4	2
내역	태어난 날짜	태어난 달	태어난 해	태어난 세기	일련 번호				점검번호		

출처 : Myrskylä(2000)

<부록 표 4> 핀란드의 건물/주거 번호 체계

내역	건물번호(Building Code)					주거번호 (Dwelling Code)	
	지역 Municipal	소지역 Village	토지 Real Estate	점검번호 Check Code	건물 Building	입구 Door	주거 Dwelling
자릿수	3	3	4	1	3	1	3

출처 : Myrskylä(2000)

# Applicability of Housing Register to Population and Housing Census: The Case of Register of Building Management

Myoung-Jin Lee\* · U-Seok Seo\*\* · Kun Lee\*\*\*

## < Abstract >

We investigated the applicability of housing register to population and housing census. Its empirical assessment was done by comparing the register of building management with the housing information from the census (the 5th pretest for 2005 Census) in three different types of areas. Our results suggest that considerable housing information register of building management can be applied to housing census, especially in the case of newly-built apartments. Yet our findings also indicate that it is necessary to introduce various policies and systems, e.g., new address system, in such a way to improve data quality and applicability of housing register.

**Key words : population and housing census, housing register, register of building management**

\* Professor, Department of Sociology Kookmin University(email: [mjlee@kookmin.ac.kr](mailto:mjlee@kookmin.ac.kr)) -  
Corresponding Author

\*\* Professor, Department of Urban Sociology University of Seoul(email: [usseo@uos.ac.kr](mailto:usseo@uos.ac.kr))  
- Corresponding Author

\*\*\* Professor, Department of Urban Sociology University of Seoul(email: [kunlee@uos.ac.kr](mailto:kunlee@uos.ac.kr))