

보고서

주거부문 행정자료의
인구주택총조사 현장적용방안

2004. 12

수탁기관 : 한국조사연구학회

통 계 청

제 출 문

통계청장 귀하

본 보고서를 「주거부문 행정자료의 인구주택
총조사 현장적용방안」의 최종보고서로 제출
합니다.

2004. 12.

사단법인 한국조사연구학회
회 장 이 계 오

연 구 진

책임연구원 : 이 건 (서울시립대 도시사회학과 부교수)

공동연구원 : 변미리 (서울시정개발연구원 도시정보센터 부연구위원)

이명진 (국민대 사회학과 부교수)

서우석 (서울시립대 도시사회학과 전임강사)

결론 및 요약

1. 현행 조사방법에 대한 평가

○ 건물, 거처, 가구 등 주거에 관한 자료가 불완전한 상황에서 수치지도를 바탕으로 이들 자료를 서로 보완적으로 조사구 설정에 활용하는 것은 효과적인 방법이다.

○ 행정자료를 통계적으로 활용하기 위해서는 자료의 정비가 필요하며, 이들을 서로 연계하여 활용하기 위해서는 자료연결방법에 관한 후속 연구가 필요하다.

2. 가구명부-거처정보와 건축물대장 자료 매칭에 대한 평가

○ 건축물대장에 있는 정보를 매칭 작업에 좀 더 많이 활용할 수 있는 알고리즘의 개발이 필요하다.

○ 조사 과정 및 자료 입력에서 나타나는 오류를 방지할 수 있는 시스템 구축이 필요하다.

3. 자료 매칭의 시사점

○ 등록된 자료가 도형이든 속성이든 규모가 큰 공동주택일수록 잘 정비되어 있다.

○ 1959년 이전의 노후 건물과 단독주택, 상가건물에 존재하는 거처에 관해 실상을 파악하기 위해서는 조사원에 의한 직접적인 조사가 바람직하다.

4. 건축물대장 활용에 대한 평가

○ 조사구 지도에서 건물의 존재유무와 거처의 존재유무를 확인하는 데 도움이 된다.

○ 앞으로 조사구 설정에서 공동주택의 건축물대장 활용은 적극적으로 추진되어야 한다.

5. 건축물대장의 정보 활용

- 건축물대장과 조사 결과의 정보에서 차이가 많이 나는 거처를 다시 조사하여 연구할 필요가 있다.
- 후속 연구에서는 그러한 차이의 현황과 원인을 파악하고, 더 나아가 건축물대장과 조사 정보의 정확성에 대한 판단을 내릴 근거를 마련해야 한다.

6. 행정자료 활용을 위한 제도적 개선방안

- 전국적으로 건축물대장을 정비해야 한다. 특히 자료의 구성과 유지, 보수 과정을 체계화하고 표준화해야 한다.
- 제도적 정비와 함께 전체적인 정비를 위해서는 인구센서스와 같은 방식으로 짧은 기간 동안 일시에 전국의 건물을 조사하여 건축물대장의 내역을 정비하는 건물센서스(building census)를 제안한다.
- 수치지도에 대한 표준화 작업도 있어야 한다. 지방자치단체의 개별사업추진으로 인한 편차를 줄여야 한다. 또한 기준점, 지도 작성 방식, 갱신 방법 및 절차에 대해서도 구체적인 지침을 만들어 제시하고 교육하여야 할 것이다.
- 앞으로 통계적 목적으로 고려하여 새주소체계는 적어도 거처수준까지는 파악이 될 수 있도록 확인 수준을 확대할 필요가 있다.

제 목 차 례

I. 연구개요	1
1. 연구 배경 및 목적	1
2. 연구내용 및 방법	4
1) 행정자료의 연계	4
2) 시험조사결과와 건축물대장 분석	5
II. 인구주택총조사의 조사방법	6
1. 인구주택총조사의 흐름	6
2. 조사구 설정	8
1) 거처	8
2) 조사구	10
3. 준비조사와 조사구 확정	15
III. 행정자료의 연계	18
1. 행정자료 연계와 번호체계	18
1) 기본 틀	18
2) 번호체계	20
(1) 개인번호체계	20
(2) 우리나라의 주소체계	21
(3) 주소번호체계	23
2. 도형자료 기반의 행정자료 연계 : 서울시 사례	24
1) 자료 연계의 배경	25
2) 건물도형과 건축물대장의 연결사례	28
3) 도로와 건물 데이터의 중첩 사례	29
IV. 분석 자료 및 자료 연결방법	32
1. 주거 부문의 행정자료 현황	32
2. 분석 자료	36
1) 지역 특성	37
2) 자료 내역 : 자료수와 필드(변수)	39
3) 건축물대장 자료	41

3. 자료의 연결	44
1) 통계청 제공 자료의 연결 방법과 문제점	44
(1) 가구명부, 거처, 건축물대장의 연결방법	44
(2) 데이터베이스 기반 매칭의 문제점	47
2) 수작업을 통한 자료연결	48
V. 지역별 비교 분석	53
1. 개요	53
2. 가구명부-거처정보와 건축물대장의 매칭 유형	53
3. 지역별 자료의 매칭 유형 비교	56
1) 지역별 유형 비교	56
4. 건축물대장 정보의 활용에 대한 평가	62
1) 건축물대장의 매칭률	62
2) 건축물대장의 건축물 주용도와 자료의 매칭유형	64
5. 매칭 오류의 사례 및 가능성	66
1) 매칭 오류의 사례	66
2) 법정동코드 사용의 문제점	68
6. 소결 및 정책대안	73
VI. 조사구별 매칭률 심층 분석 - 대전시 판암 1동	75
1. 개요	75
2. 주요 조사구별 매칭 현황 분석	79
3. 거처 특성과 매칭 현황	87
4. 거처정보와 건축물대장 정보 비교	92
5. 소결	94
1) 조사구별 매칭률	94
2) 주요 조사구의 매칭 현황	95
3) 거처 특성과 매칭 현황	96
4) 거처정보와 건축물대장 비교	96
6. 정책적 제안	97
1) 거처정보의 오류 최소화	97
2) 매칭방법	97

3) 거처 특성과 건축물대장	98
4) 건축물대장 활용	98
VII. 결론 및 제안	100
1. 현행 조사방법에 대한 평가	100
2. 가구명부-거처정보와 건축물대장 자료 매칭에 대한 평가	101
3. 자료 매칭의 시사점	103
4. 건축물대장 활용에 대한 평가	103
5. 건축물대장의 정보 활용	106
6. 행정자료 활용을 위한 제도적 개선방안	106
1) 건축물대장	106
2) 새주소체계	107
3) 데이터 오류	123
4) 사례	124
(1) 건물도형과 건축물대장의 연결사례 (서울시, 2002)	124
(2) 도로와 건물 데이터의 중첩 사례 (서울시정개발연구원, 2002)	125
(3) 입력오류	134

표 차례

<표 2-1> 인구주택총조사의 거처 구분	9
<표 2-2> 전국의 유형별 주택수 현황	9
<표 2-3> 조사구 분류	11
<표 2-4> 공동주택 기본정보 수집표(양식) 예시	17
<표 2-5> 가구명부 예시	17
<표 3-1> 덴마크, 핀란드, 한국의 개인번호 체계	21
<표 3-2> 우리나라의 주소체계	21
<표 3-3> 덴마크와 핀란드의 주소번호 체계	24
<표 3-4> 주택자료들과 특징	26
<표 3-5> 주택 및 건물 분야 데이터	26
<표 3-6> 건물도형과 건축물대장 연계 형태	29
<표 3-7> 건축물대장과 건물도형의 연계율	29
<표 3-8> 건물 및 도로면 레이어의 정합성 오류 집계 - 중구, 송파구	30
<표 3-9> 정합성 오류의 발생원인	31
<표 4-1> 인구주택총조사의 주거부문 조사 항목	32
<표 4-2> 주거부문 행정자료의 자료원과 항목	33
<표 4-3> 건축물대장의 건물 주용도 구분	35
<표 4-4> 등기부등본과 건축물대장의 고유번호 체계	36
<표 4-5> 지역별 주택유형 분포	38
<표 4-6> 지역별 농어업인구 비율	38
<표 4-7> 지역별 가구명부, 거처정보, 건축물대장의 자료 수	39
<표 4-8> 가구명부와 거처정보의 필드	40
<표 4-9> 건축물대장의 필드	41
<표 4-10> 건축물대장의 주요 항목별 누락률	42
<표 4-11> 서울시 건축물대장의 주요 항목별 누락률 (2002년 기준)	43
<표 4-12> 가구명부와 거처 매칭 및 비매칭 자료수	45
<표 4-13> 가구명부-거처와 건축물대장 매칭률 - 전체자료와 중복제거후 자료	46

<표 4-14> 조사지역에 해당하는 법정동명	48
<표 4-15> 지역별 가구명부-거처정보와 건축물대장의 매칭률 비교	49
<표 4-16> 동일 지번의 거처정보 일부와 건축물대장 예시	50
<표 4-17> 데이터베이스 매칭 결과 - 중복 허용	51
<표 4-18> 데이터베이스 매칭 결과 - 중복 삭제	52
<표 4-19> 수작업 매칭 결과	52
<표 5-1> 가구명부-거처와 건축물대장이 지번으로 매칭 되는 유형: 거처기준	54
<표 5-2> 가구명부-거처조사자료와 건축물대장의 지역별 매칭유형 분포	57
<표 5-3> 조사구별 매칭 유형 : 대전 판암 1동	59
<표 5-4> 조사구별 매칭 유형 : 강화 화도면	60
<표 5-5> 조사구별 매칭 유형 : 광주 백운1동	61
<표 5-6> 지역별 건축물대장의 매칭률(수작업 매칭)	62
<표 5-7> 매칭 되지 않은 건축물대장 자료 중 주택용도 건축물	64
<표 5-8> 지역별 건축물 주용도에 따른 자료의 매칭유형	65
<표 5-9> 동일 지번에 여러 주택유형이 동시에 존재할 때 발생하는 매칭 오류	67
<표 5-10> 대전 동구 판암1동에서 판암동과 삼정도 사이의 중복 지번	71
<표 5-11> 광주 남구 백운1동에서 서동과 백운동 사이의 중복 지번	72
<표 6-1> 판암 1동 조매칭률 현황 (중복 포함)	76
<표 6-2> 판암 1동 순매칭률 현황 (직접매칭, 중복배제)	78
<표 6-3> 판암1동 조사구 1의 정보 매칭 요약	80
<표 6-4> 판암1동 조사구 1의 매칭 상세정보	80
<표 6-5> 판암1동 조사구 6의 정보 매칭 요약	82
<표 6-6> 판암1동 조사구 6의 매칭 상세정보	83
<표 6-7> 판암1동 조사구 29의 정보 매칭 요약	85
<표 6-8> 판암 1동 조사구 29의 매칭 상세정보	85
<표 6-9> 판암1동 조사구 47의 정보 매칭 요약	86
<표 6-10> 판암1동 조사구 47의 매칭 상세정보	86
<표 6-11> 판암1동 조사구 71의 정보 매칭 요약	87
<표 6-12> 판암 1동 조사구 71의 매칭 상세정보	87
<표 6-13> 거처 종류별 매칭 현황	88

<표 6-14> 거주유무별 매칭 현황	89
<표 6-15> 건축년도별 매칭 현황	90
<표 6-16> 건축년도별 매칭 된 거처수 (관암 1동 조사구 1)	90
<표 6-17> 건축년도별 매칭 된 거처수 (관암 1동 조사구 2)	91
<표 6-18> 연건평별 매칭 현황	92
<표 부록-1> SDW 데이터 목록	117
<표 부록-2> SDW 기초/블록 DB 상세 내용	118
<표 부록-3> 주택자료 원천과 특징	121
<표 부록-4> 주택 및 건물 분야 분석을 위한 필수데이터	121
<표 부록-5> 데이터 오류발생 원인	123
<표 부록-6 > 건물도형과 건축물대장 연결 형태	124
<표 부록-7> 건축물대장과 건물도형의 연계율	125
<표 부록-8> 도로면과 건물 자료 갱신방법	126
<표 부록-9> 건물 및 도로면 레이어의 정합성 오류 집계-중구, 송파구	129
<표 부록-10> 건물·도로면 레이어의 정합성 오류 유형-중구, 송파구	129
<표 부록-11> 정합성 오류 발생 원인	133
<표 부록-12> 건축물대장의 주요 항목별 누락 비율 (2002년 기준)	135

그림 차례

<그림 2-1> 인구주택총조사 조사부분의 흐름	7
<그림 2-2> 조사구 경계 설정	13
<그림 2-3> 1/25,000 수치지도(Bessel) 연도별 제작현황	14
<그림 2-4> 조사구 지도의 예시	16
<그림 2-5> 조사구 지도 기입 예시	16
<그림 3-1> 덴마크와 핀란드 행정자료 구성체계의 기본모형	19
<그림 3-2> SDW 데이터베이스 연계 및 구축 현황	25
<그림 3-3> 서울시 기본수치지도(base map)	28
<그림 4-1> 가구명부와 거처정보 매칭	45
<그림 4-2> 가구명부-거처정보와 건축물대장 매칭	46
<그림 5-1> 매칭이 된 경우에서 각 유형이 차지하는 비율	58
<그림 6-1> 판암1동 조매칭률	77
<그림 6-2> 판암1동 순매칭률	79
<그림 6-3> 거처정보의 연건평과 건축물대장의 동연면적 차이	94
<그림 6-4> 데이터베이스 매칭 방식 개선 방안 - 공동주택의 경우	98
<그림 부록-1> 수치지도 : 서울시 기본지도	114
<그림 부록-2> SDW 시스템 구성도	115
<그림 부록-3> SDW 시스템 내 기초/블록 DB 구축의 변화과정	116
<그림 부록-4> SDW 기초/블록 DB 구축 현황	119
<그림 부록-5> SDW 기초/블록 DB 유지관리 방안	119
<그림 부록-6 > 지도 뷰어 어플리케이션 화면	120
<그림 부록-7> 도로관리시스템 도로면 갱신과정	127
<그림 부록-8> 새주소전산통합시스템 건물 갱신과정	128

I. 연구개요

1. 연구 배경 및 목적

인구주택총조사는 국가통계의 가장 기본이 되는 자료를 생산하는 조사로 거의 대부분의 나라에서 전수조사방식으로 정기적으로 시행해왔다. 그러나 최근 들어 일부 국가, 특히 선진국에서 응답거부가 늘고, 조사대상을 접촉하기 어려운 등 조사환경이 나빠지고 있으며, 조사비용이 빠르게 증가하고 있어, 이를 인구센서스에 대한 '근본적인 도전'으로 간주하고 있다(Jensen, 2000). 심지어 독일이나 네델란드에서는 조사환경의 악화로 1990년대 이후 인구센서스를 중단한 상태이다(Bierau, 2000).

조사환경의 악화는 조사의 포괄성과 신뢰성에 대한 문제성으로 나타나기 시작하고 있다. 선진국들과 마찬가지로 우리나라에서도 조사환경이 빠른 속도로 악화되고 있다. 더욱이 우리의 경우 읍면동사무소 기능축소로 말미암아 과거 인구주택총조사에서 실제 조사에 도움을 주었던 행정지원이 없어짐에 따라 앞으로 조사의 어려움은 더욱 커질 것으로 보인다.

이렇듯 악화되는 조사환경변화에 대응하여 선진 국가에서는 다양한 형태의 인구센서스방식들이 모색되고 있다. 일반적으로 인구주택총조사 방식에는 5년이나 10년마다 정기적으로 하는 전통적 센서스 이외에도, 순환형 센서스(rolling census)와 행정 등의 목적으로 등록된 여러 형태의 자료를 이용하는 방법이 있다. 순환형 센서스는 인구센서스의 주기에 따라 전체 기간 동안 지속적으로 조사하여 전체 자료를 확보하는 방법이다. 따라서 기존의 인구센서스와 마찬가지로 조사환경에 따른 조사의 포괄성과 신뢰성에 대한 문제가 발생한다. 이런 이유에서 많은 나라들이 순환형 센서스보다는 행정자료를 인구주택총조사에 활용하는 방안을 모색하고 있으며, 덴마크나 핀란드 등 일부 국가에서는 이미 전혀 조사를 하지 않고 행정자료로 대부분의 인구센서스 통계를 생산하고 있다.

많은 나라들이 행정자료를 활용한 인구센서스 방식을 선호하는 데는 또 다른 이유가 있다. 자료의 측면에서 보면, 행정자료를 활용할 경우 매년 인구센서스 통계를 생산할 수 있다. 실제로 현재 덴마크와 핀란드는 인구센서스에 준하는 통계를

매년 생산하고 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. 행정자료를 활용한 인구센서스의 이러한 특징은 국가의 교육정책, 노동정책, 복지정책 등 다양한 정책을 정확한 자료를 근거로 수립할 수 있는 기반을 제공한다.

이와 더불어 행정자료 기반의 인구센서스는 비용이 적게 드는 장점이 있다. 예를 들어 덴마크나 핀란드에서는 조사로 자료를 생산하던 때의 1/20 정도 비용으로 행정자료로 인구센서스의 모든 자료를 생산하고 있다. 특히, 최근 모든 행정자료들이 IT 기술에 의해 데이터베이스 형태로 바뀌고, 인터넷을 근간으로 한 컴퓨터네트워크가 발달함에 따라 각 부처별로 행정을 위해 축적한 자료를 IT 기술로 연계·통합하면 막대한 조사비용을 들이지 않더라도 인구센서스자료를 적은 비용으로 생산할 수 있는 근간이 마련되었다.

이렇듯 행정자료 기반의 인구센서스가 많은 장점을 가졌지만, 그렇다고 모든 국가가 당장 행정자료로 인구센서스를 대체할 수 있는 것은 아니다. 행정자료로 인구센서스통계를 생산하기 위해서는 각 행정부서별로 사용하는 행정자료들을 연계·통합할 수 있도록 국가사회 전반에 걸쳐 행정 체제가 갖추어져야 하기 때문이다. 특히 모든 국민 개개인에 관한 기본 정보, 개인들이 거주하며 생활하는 단위인 개별 주거단위에 관한 정보가 행정부에 등록되어 있고, 잘 정비되어 있어야 하며, 정보의 형태 또한 서로 연계가 가능하도록 표준화되어 있어야 한다. 이와 더불어, 현재 인구센서스에서 표본조사를 통해 부가적으로 생산하는 경제활동통계를 생산하기 위해서는 개인이 속한 사업체를 파악할 수 있도록 모든 사업체가 등록되어 있고, 개인의 경제활동과 관련된 각종 정보들이 사업체에 잘 기록 및 정비되어 있어야 한다. 따라서 행정자료 기반의 인구센서스통계생산은 단지 국가의 통계뿐만 아니라 행정조직과 행정체계를 정비하고, 개인과 사업체의 등록체계를 정비하며, 사업체의 개인에 관한 정보를 정비하여 표준화하는 막대한 작업을 수반한다. 이런 이유에서 대부분의 국가들은 장래에 행정자료 기반의 인구센서스통계생산을 목표로 하되, 당장은 행정자료를 인구센서스에 보조적 수단을 사용하는 데 노력을 기울이고 있다.

우리나라의 경우 행정자료를 인구주택총조사에 활용할 수 있는 몇 가지 중요

한 기반을 갖추고 있다. 첫째, 1962년부터 시행한 주민등록제도가 있다. 주민등록제도는 모든 국민 개개인을 파악할 수 있는 주민등록번호가 있을뿐더러, 40년 이상 제도화되어 오류가 거의 없는 편이다. 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 인구센서스통계의 주요한 단위인 가구를 파악할 수 있는 수준으로 주소체계가 정비되어 있지 않으며, 많은 사업체, 특히 소규모 사업 가운데 등록되어 있지 않거나 등록오류가 많은 편이다. 이외에도 과세대장, 토지대장 등 많은 행정자료가 아직은 불완전하여 이들을 직접 연계하기에 어렵다. 행정자료를 연계하기 위해서는 모든 자료를 정비하고 표준화하여 실제 행정에 활용하여야 하기 때문에 행정적으로 많은 노력과 시간이 필요하다. 따라서 현재는 손쉬운 부분에서부터 인구주택총조사에 행정자료를 활용하고, 앞으로 활용 과정을 거치면서 행정자료를 정비하고 표준화하는 장기적인 방안을 마련할 필요가 있다.

이 연구는 인구주택총조사에서 행정자료의 활용을 확대해 나아가기 위한 첫 번째 연구이다. 현재 인구주택총조사는 조사수행의 측면에서 새로운 국면을 맞이하고 있다. 인구주택총조사에서는 면접원이 가구를 방문하여 가구원 및 주택에 관한 사항을 면접하여 조사한다. 따라서 조사를 실시하기 면접원이 조사할 지역의 범위를 정해야 한다. 면접원이 조사할 지역의 범위, 조사수행, 조사관리 등을 위해 통계청은 60가구를 기준으로 근접한 지역의 가구를 모아 조사구를 설정한다. 이전에는 조사할 지역의 가구들에 관한 대략의 정보를 읍·면·동사무소의 협조로 사전에 파악하여 조사구를 설정하고, 조사를 수행하였다. 그러나 2000년에 읍·면·동사무소의 기능이 축소되면서 2005년의 인구주택총조사부터는 읍·면·동사무소의 지원을 받을 수 없게 되었다.

이러한 문제를 해결하기 위해 통계청은 조사구 설정에 과거의 인구주택총조사자료, 건축인허가대장, 건축물대장, 한국감정원 자료 등을 활용하여 거처와 가구의 수를 추정하고 이를 바탕으로 기본적인 조사구를 설정하였다. 2005년도 인구주택총조사를 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 조사 자료와 건축물대장이 얼마나 잘 상응하는지를 점검하고, 서로 상응하지 않는 경우를 줄일 수 있는 방안을 마련하고자 한다. 더욱이 두 자료가 서로 잘 상응하면 조사구 설정뿐만 아니라 인구주택총조사의 주택부문 항목에 건축물대장을 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 인구주택총조사에서 행정자료를 폭 넓게 활용할 수 있도록 행정자료의 개선 방안도 함께 모색해보고자 한다.

2. 연구내용 및 방법

1) 인구주택총조사 방법과 행정자료의 연계

먼저 우리 나라의 인구주택총조사에서 행정자료를 어떻게 활용하는지, 현장조사에서는 어떤 행정자료를 적용하는지에 관한 현황을 살펴볼 필요가 있다. 이미 2000년에 읍면동의 통계기능이 시군구로 이관되면서 통계청은 읍면동에서 제공받던 행정서비스를 대체할 정보의 출처를 찾기 시작했다. 이 과정에서 조사구 설정에 건축물대장, 건축인허가대장, 한국감정원 자료 등을 활용하여 조사구를 설정하는 방법들을 모색하였다. 여기서는 그러한 모색의 결과를 소개하고 관련된 인구주택총조사의 흐름을 정리한다.

다음은 이 연구의 큰 지향점인 행정자료 기반의 인구센서스라는 기본틀을 찾아본다. 인구주택총조사에서 주거부문에 행정자료를 적용하기 위해서는 이미 인구센서스에 행정자료를 활용하는 유럽의 여러 국가를 살펴볼 필요가 있다. 덴마크는 1980년부터, 핀란드는 1991년부터 모든 인구센서스 항목을 행정자료로 대체하였으며, 싱가포르도 2000년부터 행정자료 기반의 인구센서스를 본격적으로 실시하기 시작하였다. 이러한 사례들은 우리나라 인구주택총조사에서 행정자료를 활용하는 데 발생하는 문제를 예측하고, 앞으로 그러한 자료의 활용도를 높이는 데 많은 시사점을 줄 수 있다. 이런 측면에서, 이 연구는 먼저 덴마크와 핀란드 인구센서스의 기본틀을 살펴보고 우리나라의 인구주택총조사에 시사하는 바를 찾아본다.

이와 더불어 최근 서울시가 통계 기반의 행정을 위해 서울형서베이 등 다양한 형태의 조사를 수행하기 시작하였다. 이 과정에서 서울시는 모집단 설정, 현황 파악 등을 위해 여러 가지 행정자료를 조사에 활용하였다. 이 연구는 그 가운데 건물 및 주거 관련 행정자료를 기반으로 구축하는 서울시의 공간데이터웨어하우스 (Spatial Data Warehouse) 구축 과정을 검토하여 이 연구에 주는 시사점을 찾아본다.

2) 시험조사결과와 건축물대장 분석

5차 시험조사에서는 과거 인구주택총조사 자료와 건축인허가대장 및 감정원 자료의 아파트 호수 등을 이용하여 조사대상인 거처를 파악하고 이를 바탕으로 조사구를 설정하고 조사구 지도를 작성하였다. 만약 조사가 철저히 이루어지고 모든 대장자료가 완벽하다면, 시험조사 자료와 건축물대장은 상당 부분 일치하여야 한다. 그러나 실제로는 건축물 등록 방식과 등록 오류, 조사의 오류 등으로 두 자료 사이에는 차이가 발생한다. 여기서는 먼저 두 자료가 얼마나 잘 상응하는지를 점검한다. 이 과정에서 그러한 차이의 원인을 찾아보고 이에 대한 대응 방안을 마련한다. 좀 더 구체적으로는 지역의 특성, 지역의 건물 구성, 각 건물의 용도에 따라 그러한 차이는 어떻게 나타나는지를 살펴본다. 더 나아가 이를 바탕으로 인구주택총조사에서 건축물대장을 좀 더 효과적으로 사용할 수 있는 방안을 찾아본다. 또한 인구센서스 주택부문의 항목 조사에 건축물대장의 정보가 얼마나 활용 가능한지를 살펴본다.

이 연구에서 사용하는 자료는 5차 시험조사 지역 가운데 대전시 동구 판암1동, 광주시 남구 백운 1동, 인천시 강화군 화도면의 세 개 지역이다. 먼저 지역별로 조사 자료와 건축물대장이 얼마나 잘 일치하는지를 살펴본 다음, 지역의 특성과 일치도의 관계를 살펴본다. 다음은 대전시 판암 1동을 좀 더 깊이 있게 다루어 건물 구성, 용도, 분포 등 건물의 특성에 따른 조사 자료와 건축물대장이 얼마나 일치하는지를 살펴본다.

II. 인구주택총조사의 조사방법

1. 인구주택총조사의 흐름

인구주택총조사는 0년과 5년으로 끝나는 해에 전국의 거주인구와 주택을 대상으로 11월 1일부터 15일 사이에 실시한다. 워낙이 대규모 조사이며 단기간에 조사를 완료해야하기 때문에 조사설계, 조사관리, 조사방법, 자료처리의 모든 과정을 기획하고 관리하는 것이 매우 중요하다. 통계청은 2년여의 기간에 걸쳐 인구주택총조사를 준비하고, 1년 이상 자료를 처리하고 분석하여 공표한다. 2005년도 인구주택총조사의 크게 5 단계로 나누어 진행되고 있다. 그 다섯 단계는 사전준비단계(2004년 1월~12월), 본조사 추진단계(2005년 1월~11월), 자료 처리·심사단계(2005년 11월~2006년 4월), 결과 분석 및 공표단계(2005년 12월~2006년 12월), 최종평가단계(2006년 6월~2007년 3월)이다(통계청, 2004b).

이 가운데 조사와 관련된 부분은 사전 준비단계와 본조사 추진단계다. 사전 준비단계는 대체로 본조사가 있기 약 2년 전에 시작되는데 새로운 조사환경에 맞춰 본조사 수행에 필요한 각종 사안들을 준비하고 점검한다. 본조사를 위해 지방자치단체와 협조체제를 구축하고 전체적인 관리를 위한 실시본부 등의 조사관리 조직을 편성한다. 이때 지역별로 조사와 조사표 입력을 위한 전산장비를 점검하고 이에 대한 대책을 마련하기도 한다. 또한 이 기간에 여러 차례의 회의를 통해 조사항목을 확정하고, 조사표와 결과표를 설계하며, 시험조사를 거쳐 조사항목과 조사표를 확정하며, 새로운 조사방법들을 시험해보기도 한다. 2005년 인구주택총조사를 위한 시험조사는 2002년부터 2004년까지 5차례에 걸쳐 시행되었는데, 이를 바탕으로 자료 입력방식, 자료처리방식, 조사항목과 조사표 형태에 대한 점검이 있었으며, 조사원의 업무량, 업무배정, 인력동원, 조사체계 등에 관한 조정이 있었고, 다양한 행정자료를 조사의 보조적 수단으로 활용하는 방안들도 검토되었다.

인구주택총조사를 준비하는 과정 가운데 가장 중요한 작업 중 하나가 조사구를 설정하는 것이다. 인구주택총조사는 전국의 모든 가구를 15일 동안에 방문하여 조사하기 때문에 신속하면서도 정확한 조사를 위해서 원활한 조사관리가 필수적인 요소이다. 이를 위해 통계청은 전국의 모든 조사대상 가구를 일정한 규모(60가구

기준)로 표준화하여 조사구를 설정한다.1) 이렇게 설정한 조사구를 지역별로 지도에 표시하여 조사구 지도(조사구요도)를 만든다. 조사구 설정과 조사구 지도 작성은 상당히 기간동안 여러 차례의 자료 대조 및 확인을 통해 수정·보완된다. 본조사 직전에 준비조사가 있는데, 이때 조사원이 조사구 지도를 갖고 조사구를 방문하여 조사구 지도에 표시된 도로 및 건물과 그 건물에 사람이 사는지 여부와 건물 안의 가구 수 등을 확인한다. 이를 토대로 통계청은 조사구와 조사구 지도를 수정 및 확정하고, 가구명부를 작성한다. 이렇게 만들어진 가구명부와 조사구 지도는 본조사에서 가장 중요한 조사도구로 조사원은 이들을 바탕으로 가구를 방문하여 조사표를 작성한다.



<그림 2-1> 인구주택총조사 조사부분의 흐름

출처 : 통계청 내부자료 및 통계청(2004a)

1) 2000년 인구주택총조사에서 실제로 60±10 가구의 조사구는 66.8%, 60±20 가구의 조사구는 92.8%였다.

2. 조사구 설정

1) 거처

조사구 설정에서 가장 먼저 고려할 사항은 사람이 사는 건물을 파악하는 것이다. 사람들은 다양한 형태의 건물에서 살고 있다. 단독주택이나 아파트와 같은 공동주택의 경우는 거주여부가 명확한데 반해, 상가나 창고 등 거주 이외의 목적으로 지어진 건물에도 사람들이 거주하는 경우가 종종 있어 이러한 건물들을 조사할 것 인지에 대해서도 판단을 해야 한다. 또한 다세대주택이나 다가구주택과 같이 여러 가구가 함께 살고 있는 경우도 조사구 설정에서 고려해야 할 사항이며, 여인숙이나 여관과 같은 곳에서 장기적으로 거주하는 경우도 고려해야 한다.

이와 같이 복잡한 거주 현황을 파악하기 위해 통계청은 사람이 살고 있는 장소를 거처라는 개념으로 파악한다. <표 2-1>은 통계청이 분류한 거처의 종류이다. 거처에는 주택과 주택 이외의 거처가 있다. 주택은 한 가구가 살림을 할 수 있도록 지어진 집으로 영구건물이어야 하며, 부엌, 방, 독립된 출입구를 갖추고 있어야 하고, 분리해서 소유 또는 매매가 가능해야 한다. 주택은 다시 단독주택, 공동주택, 비거주용 건물내의 주택으로 나뉜다. 단독주택에는 일반 단독주택이 있으며, 등록 상으로는 단독주택 형태이지만 실제로는 여러 가구가 사는 다가구 주택이 있고, 1층에 가게가 있고 2층에는 사람이 거주하는 형태의 영업을 겸한 단독주택이 있다. 공동주택에는 아파트, 연립주택, 다세대주택이 있다. 아파트는 5층 이상의 공동주택이며, 연립주택은 4층 이하로 건축 당시 ‘연립주택’으로 허가받은 주택이고, 다세대주택은 4층 이하로 ‘다세대주택’으로 허가 받은 주택이다. 비거주용 주택은 상가, 학원 등 영업을 목적으로 지은 집이지만 방, 부엌, 독립된 출입구를 갖추어 주택의 요건을 갖춘 경우이다.

주택 이외의 거처는 주택의 요건을 갖추지 못한 거주단위이다. 이러한 거처에는 오피스텔처럼 사무실의 용도로 지었지만 숙식이 가능하며 실제로 숙식을 하는 사람들이 살고 있는 경우, 호텔이나 여관과 같은 숙박업소에서 대실료를 내고 오랜 기간 동안 사는 사람이 있는 경우, 기숙사, 고아원, 양로원, 요양원, 보육원 등 집단 수용을 위한 건물에 사람이 살고 있는 경우, 판잣집, 비닐하우스, 움막 등에 사람이

사는 경우, 이외에도 업소의 잠자는 뒷방이나, 공사장의 임시 막사와 같은 임시 거주를 위한 시설에 사람이 사는 경우 등이 해당된다.

<표 2-1> 인구주택총조사의 거처 구분

주택	1) 단독주택(일반단독, 다가구단독, 영업겸용단독) 2) 아파트 3) 연립주택 4) 다세대주택 5) 비거주용 건물 내 주택(상가, 공장, 점포, 여관 등에 주택시설)
주택 이외의 거처	1) 오피스텔(사무실겸용 주택) 2) 호텔·연관 등 숙박업소의 객실 3) 기숙사 및 특수사회시설 4) 판잣집, 비닐하우스, 움막 5) 기타

출처 : 통계청(2004a)

1970년대까지는 우리나라 사람들의 대부분이 단독주택에 살았다. 그러나 지난 30년 동안 주택의 유형이 많이 달라졌다(<표 2-2>). 단독주택뿐 아니라 다양한 형태의 공동주택이 증가하였으며, 특히 아파트는 크게 증가하여 거의 모든 주택의 반을 차지하게 되었다. 이렇듯 사람들이 다양한 형태의 주택에서 살고 있기 때문에 조사현장에 대해 상당히 잘 알고 있지 않으면 실제 조사에서 조사원이 거처를 제대로 파악하기란 그다지 쉽지 않다.

<표 2-2> 전국의 유명별 주택수 현황

연도	1970		1975		1980		1985		1990		1995		2000	
	천호	%	천호	%	천호	%	천호	%	천호	%	천호	%	천호	%
단독주택	4,155	95.3	4,382	92.6	4,652	87.5	4,719	77.3	4,727	66.0	4,337	47.1	4,069	37.1
아파트	33	0.8	89	1.9	374	7.0	822	13.5	1,328	22.7	3,455	37.5	5,231	47.7
연립/다세대	146	3.4	165	3.5	162	3.0	350	5.7	603	8.4	1,071	11.6	1,266	11.6
비거주용	25	0.6	98	2.1	131	2.5	213	3.5	202	2.8	313	3.7	393	3.6
전체	4,360	100.0	4,734	100.0	5,319	100.0	6,104	100.0	7,160	100.0	9,205	100.0	10,959	100.0

출처 : 김호철(2002: 52)

이러한 이유에서 1995년도의 인구주택총조사까지는 거주 현황을 잘 알고 있는 읍면동사무소에서 통계를 담당한 공무원이 거처를 파악할 수 있는 정보를 제공

함으로써 조사구 설정이 수월하였다. 그러나 1999년부터 시행된 행정자치부의 읍면동 기능전환 사업으로 통계기능이 시군구로 이관되면서 인구주택총조사에서 거처 파악이 점점 어려워지고 있다.

2000년 인구주택총조사 때는 다행스럽게도 아직 읍면동의 기능 이전이 시작된지 얼마 지나지 않아 지방자치단체의 통계담당자를 어느 정도 활용할 수는 있었으나, 2005년도 인구주택총조사부터는 읍면동 단위의 지방자치단체로부터 거처파악 등 통계관련업무에 관한 공식적 협조를 받을 수가 없는 실정이다. 현재 행정자치부의 인구주택총조사 관련 읍면동의 사무처리 지침(2004)을 보면 구는 조사원 모집 공고, 조사원의 선발 및 채용, 조사구역 배정, 조사원 교육, 조사결과 내용 검토를 담당하고, 각 동은 담당 공무원을 지정하여 조사원을 관리하며, 지역별로 통합관리할 수 있는 공간을 마련해주는 등 주로 지방자치단체는 조사원 관리와 관련된 실무를 협조하도록 되어 있다.

이런 행정체제의 변화에 따라 통계청은 거처 파악을 위한 다양한 방법을 모색하고 있다. 통계청은 2000년 인구주택총조사에서 파악된 건물과 거처에 대한 자료를 보유하고 있기 때문에 이 자료를 활용할 수 있다. 2000년 이후 변동이 없는 건물의 거처는 2000년과 큰 차이가 없을 것이지만 기존의 건물이 없어지거나 새로운 건물이 들어선 경우는 거처의 차이가 발생한다. 이런 경우, 특히 신축건물이 있을 경우 건축물대장과 건축인허가대장을 이용하여 거처를 확인하고 있다.

2) 조사구

거처확인을 바탕으로 거처별로 몇 가구가 사는지를 파악하여 60가구 내외로 조사구를 (가)설정한다. 조사구는 크게 일반조사구와 특별조사구로 나뉜다(<표 2-3>). 일반조사구는 일반인이 거주하는 곳으로 조사원이 출입할 수 있는 지역에 설정된 조사구이며, 특별조사구는 조사원의 출입이 제한되거나 조사가 불가능한 지역에 설정된 조사구이다. 일반조사구에는 보통조사구, 섬조사구, 아파트조사구, 기숙시설조사구, 특수사회시설조사구, 관광호텔 및 외국인 거주지역조사구가 있다. 이들 가운데 아파트조사구는 아파트의 지속적 증가추세에 따라 2000년에 신설된 조사구이다(<표 2-2> 참조). 특별조사구에는 재외주재공관, 교도소 및 소년원, 육·해·공군, 전투경찰대 등이 있다.

가구 수를 기준으로 보면 보통조사구, 아파트조사구, 섬조사구가 총조사구의 대부분을 차지한다. 참고로 2000년도 인구주택총조사의 경우 보통조사구가 62.9%(159,923 가구), 아파트조사구가 34.8%(88,516 가구), 섬조사구가 0.4%(1,127 가구)로 이들 세 조사구가 전체의 98.2%를 차지하였다(임명선, 2003). 한 가지 부연할 사항은 현재도 아파트가 계속 증가하는 추세이기 때문에 2005년도 조사에서도 아파트조사구는 더욱 증가할 것으로 예상하고 있다.

<표 2-3> 조사구 분류

대분류	소분류	비고
일반조사구	아파트조사구	60 가구 기준으로 지역을 분할해 설정
	보통조사구	
	섬조사구	
	기숙사시설조사구	개별시설단위로 설정
	특수사회시설조사구	
	관광호텔 및 외국인 거주지역조사구	호텔과 외국인 거주지역
특별조사구	재외주재공관(외무부)	조사원으로는 조사가 불가능한 지역
	교도소, 소년원 등(법무부)	
	육·해·공군(국방부)	
	전투경찰대 등(경찰청)	
	전투경찰대 등(해양경찰청)	

출처 : 통계청(2002)

2000년 인구주택총조사에서는 조사구를 군부와 시부로 나누어 단계적으로 설정하였다. 그 당시에는 먼저 읍면동에서 조사구를 가설정한 다음, 통계청에서 이를 심사하여 조사구를 확정하였다. 군지역은 다른 지역에 비해 조금 일찍 1999년 9~10월에 조사구가 설정되었으며, 일반시 지역은 2000년 2~4월에, 특별 및 광역시 지역은 2000년 4~6월에 조사구가 설정되었다.

지난 인구주택총조사와 달리 이번에는 읍면동에서 조사구를 설정할 수 없기 때문에 조사구 가설정 및 설정 작업을 모두 통계청에서 해야 한다. 그렇지만 가설정된 조사구와 변동 조사구를 확인하여 조사구를 수정·보완하기 위해서는 지방자치단체의 협조를 받는 것은 불가피하다. 이에 조사구 설정에서는 시군구 통계담당 직원의 업무 협조를 받는다(임명선, 2003). 또한 조사구 설정을 지역별로 나누기보다는 업무별로 나누어 2004년 하반기에 가구 수를 파악하였고, 2005년 상반기에 조사구를 설정할 계획이다.²⁾

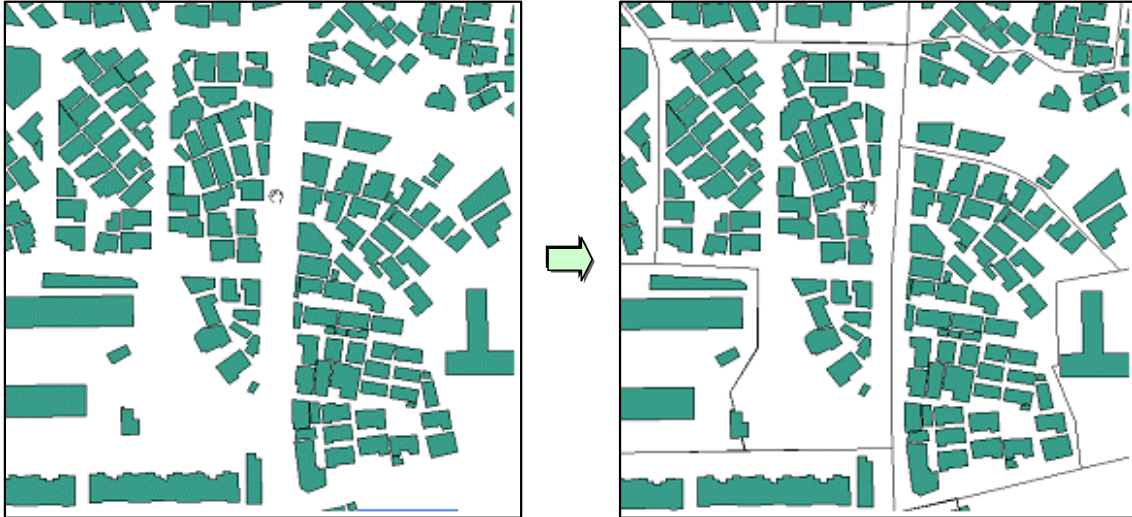
조사구 설정은 조사할 지역의 구획을 정하는 것으로 공간적으로 고려해야 할 사항들이 많다. 조사구는 조사구 안의 모든 거처를 빠지지 않고 가능하면 짧은 시간 안에 방문할 수 있도록 이동경로가 단순·명료하며, 장애물이 없어야 하고, 총 이동거리가 또한 길지 않은 것이 바람직하다. 따라서 조사구 설정에는 도로와 건물이 표시된 지도가 필수적인 도구이다. 2000년 인구주택총조사부터 통계청은 행정자치부의 수치지도를 활용하여 지도 작업을 GIS(Geographical Information System)를 기반으로 전산화하여 조사구 설정용 기본도 및 조사구 지도를 작성하였다. 당시 그렇게 작성한 조사구 지도의 활용도가 조사에서 98.7%로 매우 높아 수치지도로 작성한 조사구 지도의 유용성은 입증되었다고 할 수 있다.

수치지도란 항공사진 등을 바탕으로 지형·지물 등을 표현한 도형자료와 지명·수치 등을 표현한 속성자료를 컴퓨터에서 처리 할 수 있도록 디지털 자료로 변환한 지도이다. 그림과 문자를 컴퓨터에서 처리하기 위해서는 점, 선, 면과 같은 좌표와 기하학적 요소의 형태로 표현하여 선추적형으로 변환하여 수치자료 좌표(X, Y, Z)로 입력한 지도이다. 행정자치부의 수치지도에는 도형자료로 도로와 건물이 표시되어 있고, 속성자료로 건물명과 새주소가 있으며, 이에 함께 구주소가 포함되어있다. 도형자료로 표현된 수치지도는 <그림 2-2>의 왼쪽에 있는 지도이다. 이 지도에서 어두운 부분은 건물이며 비어 있는 부분이 대체로 길이라고 보면 된다. 조사구 설정은 지도 위에 조사구의 경계선을 그려 넣는 작업이다. 이미 파악된 거처와 거처별 가구 수를 바탕으로 주로 길을 따라 경계선을 그려 넣는다. 그렇게 만들어진 조사구를 표시한 것이 <그림 2-2>의 오른쪽 그림이다.

이렇게 만든 지도가 항상 정확한 것은 아니다. 가장 보편적인 오류가 지도의 제작 시점의 차이에서 오는 것이다. 이러한 오류도 크게 제작 시점과 사용 시점의 차이와 지역별로 다른 제작 시점으로 인한 오류로 나눌 수 있다. 전자의 경우는 제작 시점이 오래 전일수록 현재의 지형지물과 차이가 많이 발생한다. 후자의 경우 같은 지도에 제작 시점이 다른 경우 어떤 지역은 지도와 잘 맞지만 어떤 지역은 잘 맞지 않아 혼동이 발생할 수 있다(<그림 2-3>). 특히 한 행정구역 안에서 그런 문제가 나타날 경우 오류의 근원을 쉽게 발견하지 못할 수도 있다. 예를 들어 대전시

2) 한때 조사구 설정에 기초단위구를 활용할 계획이었으나 조사구별 가구수 편차와 변동, 조사구 경계설정의 어려움 등으로 기초단위구를 활용하기 어려웠다. 그러나 앞으로 안정된 기초단위구가 확보되면 이의 활용도가 높아질 것으로 보인다.

1/1,000 수치지도의 경우 어떤 지역은 1995년에 촬영을 하고 1997년에 조사·제작하였으며, 어떤 지역은 2003년에 촬영·조사·제작을 함께 하였다.

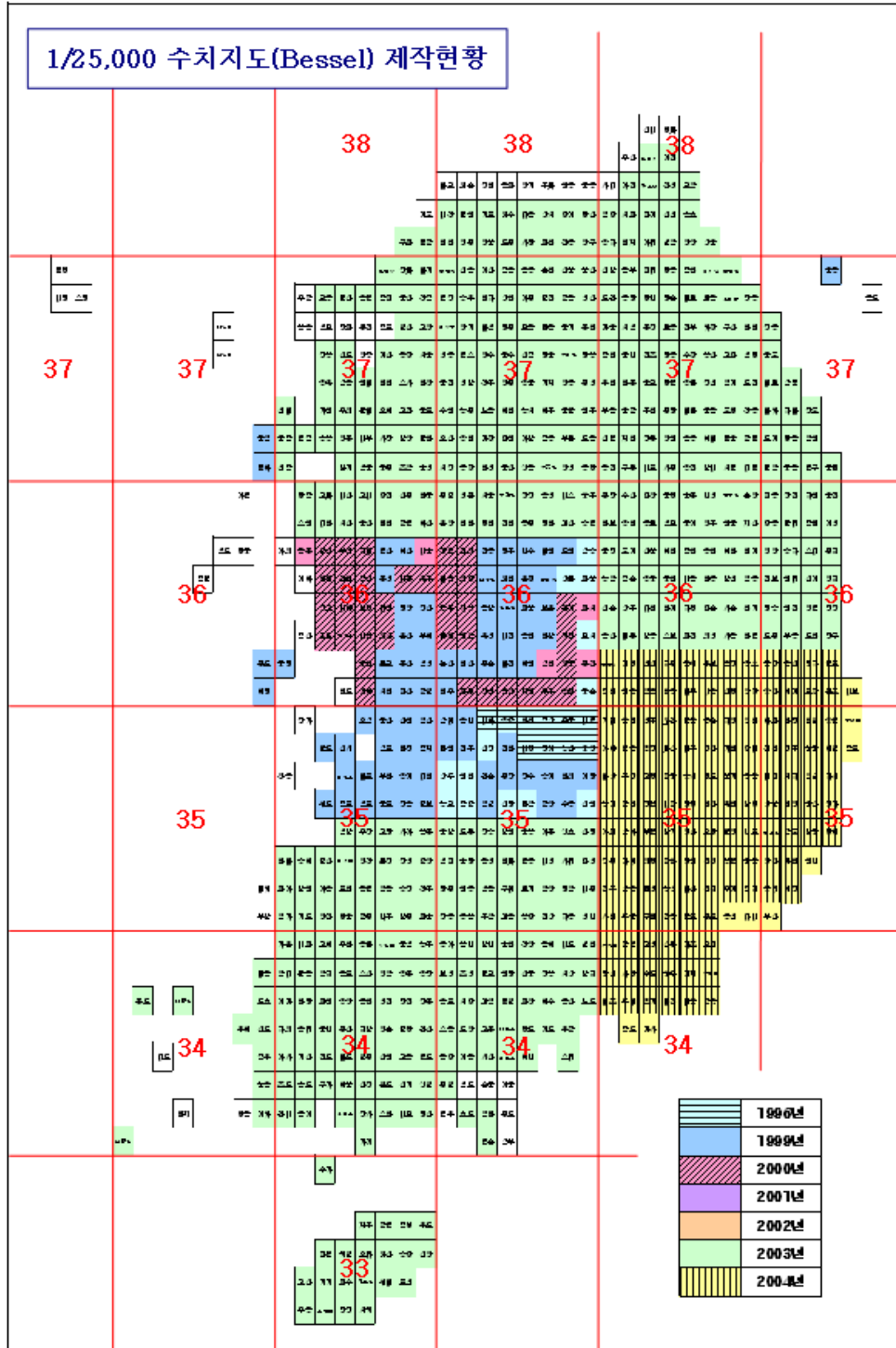


<그림 2-2> 조사구 경계 설정

모든 지역의 수치지도가 모든 축척으로 제작된 것은 아니다. 특히 조사구 설정에 적합한 1/1,000 지도는 아직 제작되지 않은 지역이 많이 있다. 또한 수치지도가 모든 지역에서 같은 수준의 정확도로 제작되는 것은 아니다. 농촌이나 산간 지역은 지적 측량이 오래 되어 지도가 정확하지 않은 경우가 많다. 더 나아가 지도 자체의 문제는 아니지만 주택이 아닌 부속 건물들도 모두 지도에 표시되기 때문에 현장조사에서 어려움이 있다. 이런 경우도 도시보다는 창고, 축사, 외부 변소 등 부속 건물이 많은 농촌이나 산간지역에서 더 많이 나타난다. 이런 문제로 조사구 설정을 완전히 지도상의 지형지물에만 의존할 수 없으며, 실제로 현장조사를 통해 확인하는 절차가 필요하다.

통계청은 그동안 통계청이 확보한 자료를 활용하여 완벽하지는 않으나 이러한 문제를 어느 정도 해결하는 방안을 마련하였다. 통계청은 2000년 인구주택총조사에서 수치지도를 사용하였기 때문에 조사 과정에서 조사구 지도를 수정하였다. 이 지도는 그 당시로는 상당히 정확하며, 적어도 조사구 설정을 위해서는 가장 정확하다고 할 수 있다. 앞서 설명한 거쳐확인작업과 마찬가지로, 통계청은 이 자료에 2000년 11월 이후에 새로 제작된 수치지도를 붙이고, 더 나아가 허물거나 새로 건축한 건물을 파악하여 지도를 수정하였다. 이때 허물거나 새로 신축한 건물은 행정

자치부와 지방자치단체가 보유하고 있는 건축인허가대장과 건축물대장의 도면을 사용하여 지도를 수정하였다.



<그림 2-3> 1/25,000 수치지도(Bessel) 연도별 제작현황

출처 : 국토지리정보원(2004)

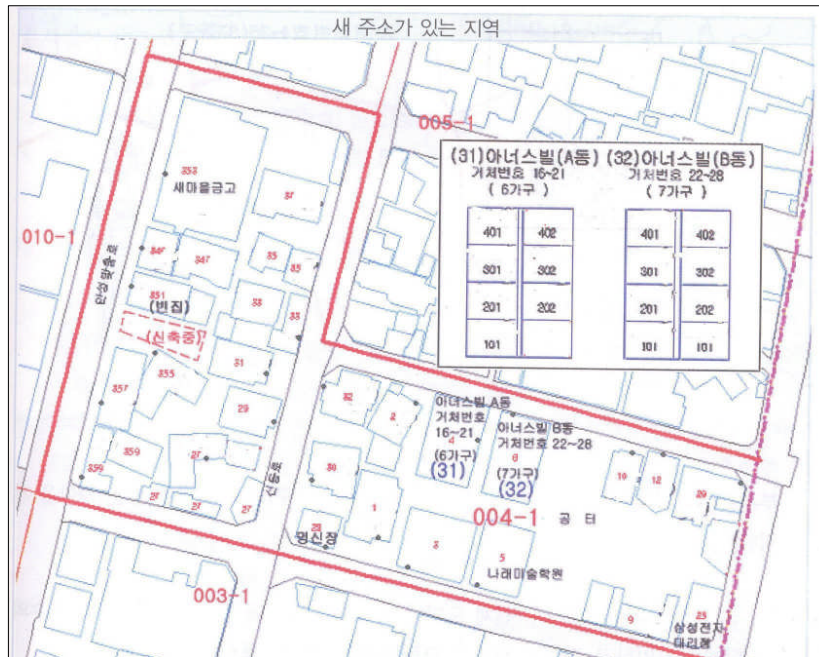
3. 준비조사와 조사구 확정

준비조사는 조사구, 조사구 지도, 조사가구를 확정하기 위한 조사로 본조사 직전에 모든 조사원이 조사현장에 나가 실시한다. 지난 2000년 조사의 경우 10월 30, 31일 이틀 동안 준비조사가 있었다. 조사원은 설정된 조사구에 나가 실제 가구 수를 파악하고, 조사구 지도의 도로와 건물 및 조사구의 경계를 확인하고, 조사구와 조사구 지도를 수정·보안한다.

조사 현장에서 직접 지도 위에 잘못된 경계를 수정하고, 실제와 다른 건물이나 지형지물을 표시하며, 신축 또는 누락된 건물을 그려 넣고, 신축 중인 건물은 점선으로 표시하며, 건물마다 사람이 사는지 여부를 확인하여 거처가 아닌 건물은 'X' 표한다(<그림 2-4, 5>). 이와 함께 건물에 거처번호를 부여하고, 거처별로 실제 거주하는 가구 수를 파악하여 조사구 지도에 거처번호와 함께 기입한다.

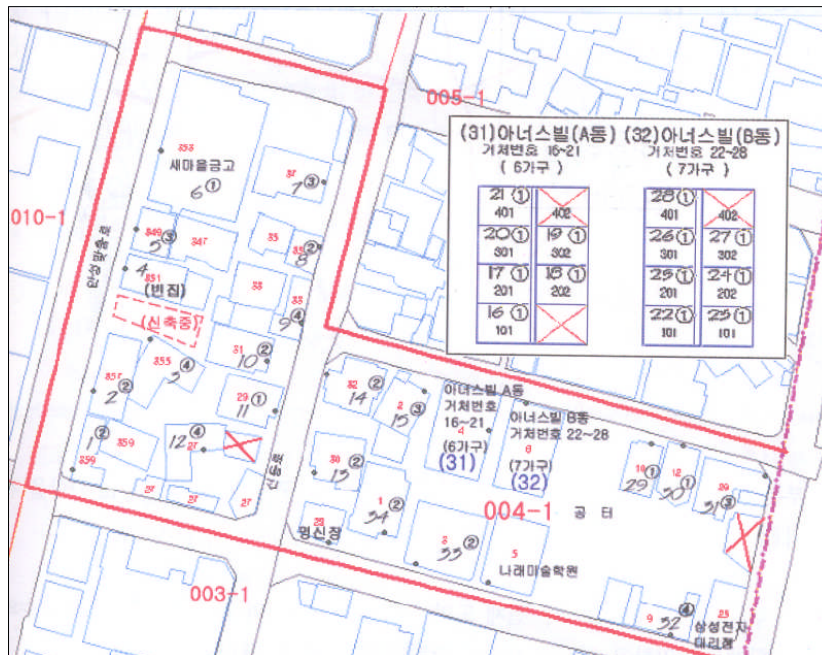
준비조사에서 아파트, 연립주택, 다세대주택 등의 공동주택의 주택 관련 정보를 파악한다. 공동주택에서는 거주자 개개인을 통해 주택에 관한 사항을 파악하기 보다는 일괄적으로 파악하는 것이 손쉬우며 더욱 정확하기 때문이다. 조사원은 자신의 조사구에 있는 공동주택에 대해 기본정보를 수집하여 '공동주택 기본정보 수집표(<표 2-4>)'에 기재하고 조사표에 옮겨 적는다. 공동주택의 기본정보는 주택분류, 전용면적, 총 방수, 건축년도, 편의시설수 등이다. 기본정보는 아파트의 경우는 주로 관리사무소를 활용하며, 연립주택이나 다세대주택의 경우 주인집이나 부동산소개소 등 그 집에 대해 잘 아는 주민에게 문의해 수집한다.

준비조사에서 조사원의 가장 중요한 업무 가운데 하나가 가구명부 작성이다. 가구명부는 본 조사에서 조사원이 조사할 대상의 명단이다. 가구명부에는 각 건물의 주소, 지정된 조사구 번호, 공동주택 명칭 등이 포함되어 있다(<표 2-5>). 가구명부작성에서 조사원은 조사구 지도에 기입한 거처번호와 가구번호를 기입하는데, 이렇게 작성한 번호는 조사구번호와 함께 가구의 고유한 ID가 된다.



<그림 2-4> 조사구 지도의 예시

출처 : 통계청(2004a)



<그림 2-5> 조사구 지도 기입 예시

출처 : 통계청(2004a)

<표 2-4> 공동주택 기본정보 수집표(양식) 예시

조사구 번호	공동주택 분류			명칭 및 동·호수	연건평 (전용면적)	총방수			건축 년도	편의 시설수			비고
	아 파 트	연 립	다 세 대			방 수	거 실 수	식 당 수		부 역 수	화 장 실 수	출 입 구 수	
032-A	○			대전아파트 201동 101호 ~ 2002호까지	26평	3	1	0	1998	1	2	1	
.....													

출처 : 통계청(2004a)

<표 2-5> 가구명부 예시

시·도	시·군·구	읍·면·동	조사구 번호	조사표 부수	가구명부 매수	조사구 지도 매수	조사원	다세미
대전 광역시	서구	가수원동	31-1		매중 매	매	조사관리자 총관리자	박인구 김통계

거 처 번 호	가 구 번 호	주소					가구주 성명	가구 원수		가구 종류		거처 종류				농림어가 여부			비 고				
		법정 읍면동	동리	번지	공동주택			남	여	일 반 가 구	집 단 가 구	외 국 인 가 구	단 독 주 택	아 파 트	연 립 주 택	다 세 대 주 택	비 거 주 용 건 물	주 택 이 외 거 처		농 가	임 가	어 가	
					명칭	동· 호수																	
001	001	괴곡동	3통	121-1	신성 타운	A동 101호																	
002	001	괴곡동	3통	122																			
	002	괴곡동	3통	122																			
003	001	“	“	123																			

출처 : 통계청(2004a)

III. 행정자료의 연계

인구주택총조사에서 여러 가지 어려운 점이 있으나 그 중에서도 현재 가장 큰 어려운 점은 조사대상 가구를 파악하는 것이다. 가구를 파악하기 위해서는 지역 사회에 관한 정보가 필요하지만 이를 확보할 방법이 매우 부족한 실정이다. 이전에는 읍면동사무소에서 지역사회의 거주에 관한 상세한 정보를 갖고 있었으나 이들의 기능축소로 말미암아 이러한 정보를 얻을 길이 막연하다. 이러한 문제에 봉착하여 통계청은 수치지도로 기반으로 과거의 인구주택총조사 자료를 활용하고, 건축인허가대장과 건축물대장을 활용하여 부족한 정보를 최대한 보완하고 있다. 통계청의 이러한 노력은 현재로서는 가장 최선의 방법으로 보인다. 하지만 현재 사용하고 있는 방법은 일일이 손으로 작업하는 것이 많고, 준비조사에 의존하는 부분이 너무 많다. 앞으로 이러한 부분을 줄이기 위해 관련된 행정자료를 체계적으로 활용하는 방안을 마련할 필요가 있다. 이런 취지에서 여기서는 장기적인 안목에서 행정자료의 연계와 관련하여 자료연계의 주요 근간인 주소체계를 살펴보고, 건물 및 주거에 관한 행정자료를 살펴본 다음, 또 다른 방법으로 수치지도 기반의 자료 연계를 서울시의 사례를 통해 살펴보고자 한다.

1. 행정자료 연계와 번호체계

1) 기본 틀

인구주택총조사에서 행정자료로 통계를 생산하기 위해서는 관련 행정자료를 서로 연계할 수 있어야 한다. 이미 오래 전부터 행정자료로 인구센서스를 대체한 덴마크와 핀란드의 경우를 살펴보면 개인, 건물 및 주거, 사업장에 관한 정보의 연계를 기본으로 하여 모든 자료를 생산하고 있다(<그림 3-1>). 예를 들어 인구센서스의 가구 및 가구원과 주택에 관한 통계는 주거정보와 개인정보를 연계하여 생산하며, 경제활동에 관한 통계는 개인정보와 사업장정보를 연계하여 생산한다.

<그림 3-1> 덴마크와 핀란드 행정자료 구성체계의 기본모형

출처 : 덴마크와 핀란드 통계청 내부자료

행정자료의 연계는 개인, 건물 및 주거, 사업장 번호체계(coding system)를 바탕으로 이루어지며, 행정자료 활용의 가장 중요한 기반이 된다(Borchsenius, 2000). 모든 국민 개개인을 식별하기 위한 개인번호체계, 국가의 모든 건물 및 주거를 식별하기 위한 주소번호체계, 모든 사업체를 식별하기 위한 사업체 번호 체계가 없으면 그에 해당하는 정보를 수집하여 일관되게 연계할 수 없기 때문이다. 예를 들어 전국이나 지역별 가구 수를 추정하기 위해서는 행정적으로 모든 국민 개개인의 주거지 주소가 있어야 한다. 이때 주소는 모든 주거단위별로 식별이 가능한 주소이어야 하며, 그렇게 하기 위해서는 전국의 모든 주거를 포괄하는 주소체계가 있어야 한다. 이런 체계가 갖추어진 경우, 가구 수는 각 개인을 주거별 주소로 묶어서 산출 할 수 있다. 그러나 개인식별체계와 주거별 주소체계가 글자로 되어 있으면 전산기술을 바탕으로 자료 연계 및 자료 점검을 하기가 쉽지 않다. 이런 이유에서 행정자료를 연결하기 위해서는 먼저 개인, 건물 및 주거, 사업장에 대한 번호체계를 정립해야 한다.

일단 그러한 번호체계가 확보되면 모든 행정부서 및 관련 기관들이 그 번호체계를 사용해야 한다. 행정을 통해 생성되는 자료는 따로 별도의 과정을 거쳐 생산하는 것이 아니다. 행정의 결과로 자연스럽게 만들어지는 자료이다. 따라서 모든 기관에서 같은 번호체계를 사용해야만 개별 자료들을 연결할 수 있다.

이와 더불어 정확한 자료를 확보하기 위해서는 실생활에서도 같은 번호체계를 사용해야 하고, 그러한 체계의 사용이 생활과 직접 연관이 있어야 하며, 당사자의 자발적 신고가 이루어져야 한다. 그렇지 않으면 실생활에서 변동이 발생할 경우 실제로 행정등록의 갱신이 이루어지지 않아 정확한 자료를 확보할 수 없다. 예를 들어 주민이 이사를 오면, 주민등록을 갱신하는 데 이때 행정자료가 갱신되고, 이렇게 갱신된 자료를 모으면 그 행정 단위의 주민 집계도 갱신된다. 따라서 정확한 집계자료가 생성되기 위해서는 이사를 오는 주민은 반드시 주민등록을 갱신하여야 한다. 현재 주민등록법에는 이사 후 14일 이내에 주민등록의 변경을 신고하도록 하고 있으나 실생활에서 주민등록 이전 신고가 그다지 필요하지 않아 신고하지 않는 경우가 종종 있다. 이럴 경우 행정자료는 실제와 일치하지 않으며, 그런 오차가 크면 행정자료로 생성한 통계자료의 유용성은 크게 떨어진다. 이런 오류를 막기 위해 선진국에서는 자동으로 이사 간 주소지에 우편물을 배달해 주는 서비스나, 주거지 기반의 의료보험 서비스 등의 제도를 갖고 있다.

2) 번호체계

(1) 개인번호체계

이런 배경에서 먼저 현재 우리나라 행정에서 사용되는 번호체계를 파악할 필요가 있다. 우리나라에도 여러 가지 번호체계가 인구주택총조사와 관련해서는 개인번호체계와 주소번호체계를 살펴보기로 한다. 먼저 우리나라의 개인식별번호를 살펴보면 주민등록번호가 있다. 아래 <표 3-1>은 덴마크와 핀란드의 개인번호체계와 우리나라의 주민등록번호체계를 함께 제시한 것이다. 자릿수를 제외하면 덴마크, 핀란드와 우리나라의 개인번호체계는 매우 유사하다. 두 나라의 개인번호 자릿수가 우리나라보다 적은 것은 두 나라의 인구가 500만 명 정도로 우리나라 인구의 1/10 수준이기 때문이다. 우리나라의 주민등록번호체계는 40년 동안 사용되어 보편화되어 모든 국민 개개인을 식별하는 데 큰 지장이 없었으며, 특히 50세 이하의 연령층에서는 거의 오류가 없는 것으로 알려져 있다. 또한 공식적인 행정서류나 금융에서도 주민등록번호를 사용하고 있기 때문에 자료 연계를 위한 번호체계로서의 여건을

잘 갖추고 있다고 할 수 있다.

<표 3-1> 덴마크, 핀란드, 한국의 개인번호 체계

덴마크	내역	출생 일	출생 월	출생년도	출생세기	무작위 번호	점검번호 (성별포함)	계
자릿수	2	2	2	2	1	2	1	10

출처 : 덴마크 통계청 내부 자료

핀란드	내역	출생 일	출생 월	출생년도	출생세기	일련번호	점검번호	계
자릿수	2	2	2	2	1	3	1	11

출처 : Myrskylä(2000)

한국	내역	출생년도	출생 월	출생 일	성별	출생지역	추가번호	점검번호	계
자릿수	2	2	2	2	2	4	1	1	13

(2) 우리나라의 주소체계

개인번호체계에 비해 우리나라의 주소체계는 아직 일관된 체계를 갖고 있지 못하며 여러 가지 주소체계가 혼용되고 있어 혼란스러운 편이다. 우리나라에는 크게 구주소와 새주소가 있는데, 현재 우리가 사용하는 주소가 구주소이다. 구주소는 다시 법정동과 행정동으로 이원화되어 있다. 법정동은 법규정에 따라 관공서의 장부에 사용하고, 행정동은 행정업무에는 사용한다. 구주소는 본번과 부번을 합한 지번체제로 토지 중심이며, 새주소는 도로와 건물 중심으로 구성되어 있다(<표 3-2>).

<표 3-2> 우리나라의 주소체계

구주소	내역	법정동	본번	부번	계
	자릿수	10	4	4	18
새주소	내역	법정동	도로	건물번호	계
	자릿수	10	9	8	27

우리나라의 주소체계가 이원화된 데는 90 여년의 역사적 배경이 있다. 우리나라의 주소체계는 원래 일제시대에 조세징수를 목적으로 한 토지조사사업(1910~

1918)을 바탕으로 만들어졌다. 이 사업은 지적법에 기초를 두고, 각 필지에 대해 지적도상에서 경계선을 확정하고, 지번을 부여하였다. 이렇게 만들어지는 주소체계는 토지의 소유가 중요했던 농업사회에서 토지 중심의 재산을 관리하기 위한 중요한 수단이었다. 행정적으로 필지에 대한 자료는 토지대장과 임야대장으로 나누어 관리하고, 건물에 대한 자료는 건축물대장으로 관리하게 되었다.

법정동은 그 당시 인구규모, 도로, 하천, 능선 등을 기준으로 권역을 나누는 것으로 그 이후 우리 사회의 인구와 도시의 변화를 반영하지 못하고 있다. 비근한 예로 서울시 종로구에는 87개의동이 있으며, 중구에는 74개의동이 있는데 반해, 관악구에는 3개의 동밖에 없다(박현주, 2000). 해방 이후 도시의 변화에 따라 행정업무를 위해 법정동을 여러 권역으로 나누었고, 그 결과 하나의 법정동에 여러 개의 행정동이 생기게 되었다. 가장 대표적인 경우는 동에 번호를 붙인 것으로, 예를 들어 서울시립대학교의 법정동 주소는 서울특별시 동대문구 전농동 90번지에 위치해 있지만 행정동 주소는 전농3동에 있다.

또한 도시 내부의 재편과 도시의 팽창 과정에서 도시구조의 변화로 동 경계가 달라졌다. 그러나 법정동 주소체계는 이러한 변화들을 반영할 수 없었기 때문에 골목길 사이로 동이 나뉘기도 하고, 도로가 없어진 경우에는 조밀한 집 사이로 경계가 설정되기도 하였다. 그 결과 지번이 불규칙하게 부여되면서 우리나라의 주소체계는 매우 혼란스러운 형태를 갖게 되었다.

한편 도시로 인구가 집중하여 일반 사람들의 생활이 도시 중심으로 재편됨에 따라 토지 사용의 형태가 달라지고, 토지보다는 건물이 더 중요한 재산의 가치를 갖게 되었다. 또한 농업시대에는 일상의 활동이 토지를 중심으로 이루어졌으나 도시에서는 건물 및 건물과 건물을 연결하는 도로를 중심으로 이루어진다. 그러므로 지번 중심의 주소체계는 생활과 그다지 연관이 없다. 예를 들면 어떤 지역에서 건물이나 장소를 찾는 데 구주소는 그다지 도움이 되지 않을뿐더러 우편통신에 불편하고, 화재와 같은 재난구조의 어려움을 주기도 한다.

실생활의 불편함 이외에도 지번 기반의 구주소체계는 건물을 관리하는 데 적합하지 않다. 토지 중심의 사회에서 건물은 토지의 부속물이었기 때문에 건물은 그 건물이 위치한 토지의 지번으로 관리하였다. 그 결과 지금도 건축물대장에는 개개의 건물을 식별할 수 있는 건물의 고유번호가 없다. 지번으로 건물을 관리하는 경우 한 필지에 한 건물이 1:1로 일치하는 경우에는 큰 문제가 없다. 그러나 한 필지

에 여러 개의 건물이 존재하거나 여러 개의 필지에 한 건물이 존재하는 경우 또는 여러 필지에 여러 개의 건물이 동시에 존재하는 경우가 많아 등 지번으로 건물번호를 사용하는 경우 건물정보를 관리하기가 어렵다. 실제로 용산구 3가 40번지는 본번 하나에 약 3,000개의 부번이 부여되어 있으며, 신림동 1449-30에는 48채의 집이 존재한다(박현주, 2000).

이렇듯 구주소체계가 도시의 일상생활과 괴리되어 있고, 건물에 대한 행정관리에 적합하지 않은 부분이 점점 커지면서 도로와 건물 중심의 새로운 주소체계가 필요해졌다. 이에 행정자치부가 주관하여 1996년부터 각 지방자치단체에서 새주소사업이 시행되어 2009년 완료될 예정이다. 2003년 6월 현재 20개 시, 45개 구, 1개 군 등 66개 지역에 새주소사업이 완료되었으며, 도시는 전체 143개 지역 중 133개 지역(93.0%), 농촌은 전체 89개 지역 중 14개 지역(15.7%)이 사업에 착수하여 진행 중에 있다.

새주소사업은 토지와 건물을 분리하여 지번 중심의 구주소는 토지에만 사용하고, 건물에는 길 이름에 건물별로 번호를 부여하여 사용하는 것을 골자로 한다. 새주소의 일차적 목적은 찾기 쉬운 주소체계이다. 새주소사업에 부가적으로 발생하는 이점들이 몇 가지 있다. 1) 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. 특히 건물별로 식별이 가능한 번호가 매겨지면 건물의 속성, 소유자, 사용자 등에 관한 건물 관련 정보를 건물별로 데이터베이스화 하여 행정에 활용하거나 통계를 생산하기에 수월하다. 또한 새주소사업에 하나로 진행되는 수치지도는 이러한 건물정보와 함께 사용하면 지리정보시스템(Geographical Information System) 기반의 행정, 행정서비스 뿐만 아니라 통계를 파악하기 쉽게 가공하고 자료를 시각적으로 확인하는 부수적 효과를 가질 수 있다.

(3) 주소번호체계

우리나라의 주소체계를 좀 더 체계적으로 살펴보기 위해 <표 3-3>에 덴마크, 핀란드의 주소번호체계를 제시하였다. 덴마크의 주소는 도로를 중심으로 구성되어 있어 우리의 새주소와 비슷하며, 핀란드의 주소는 토지를 중심으로 구성되어 있어

우리의 구주소와 비슷하다. 다시 말해, 우리나라는 덴마크와 핀란드의 주소체계와 비슷한 두 가지 체계를 모두 가지고 있다. 그렇지만 이들 주소체계의 기반만 비슷할 뿐 체계성이나 주소정보의 상세한 수준에서 큰 차이가 있다. 덴마크와 핀란드의 주소는 모두가 개별 거처수준까지 식별이 가능하여 가구별 주거지를 확인할 수 있도록 되어 있다. 이에 반해 우리나라의 주소체계는 구주소의 경우 지번(토지)까지, 새주소의 경우 건물까지만 식별이 가능하다. 따라서 한 건물에 여러 가구나 업체들이 있으면 그 각각의 대상에 대해 식별이 가능하지 않다. 물론 지역의 동사무소나 우체국 등에서는 이들을 따로 파악하고 있는 경우가 대부분이기는 하지만 모든 지역에 공히 적용되는 일관된 체계는 없기 때문에 행정자료를 활용하여 개별 가구를 파악하기는 어렵다. 또한 덴마크의 주소는 번호로 환산하면 17 자리이며, 핀란드는 건물 18자리, 주거 5자리로 총 23자리이다. 우리나라의 경우 구주소의 번호는 18자리이고 새주소는 27자리로 자릿수가 조금 많은 편이다(<표 3-2> 참조). 특히 새주소의 경우 건물번호까지밖에 없어 자릿수로 치면 덴마크의 체계에서는 10자리, 핀란드의 체계에서는 18자리 수준의 정보 밖에는 표현하지 못하고 있다.

<표 3-3> 덴마크와 핀란드의 주소번호 체계

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

출처 : 덴마크 통계청 내부 자료

핀란드	내역	건물(Building)						주거(Dwelling)			계
		지역 Municipal	소지역 Village	토지 Real Estate	구역 Block	점검번호 Check Code	건물 Building	입구 Door	층 Flat	주거 Dwelling	
	자릿수	3	3	4	4	1	3	1	3	1	23

출처 : Myrskylä(2000)

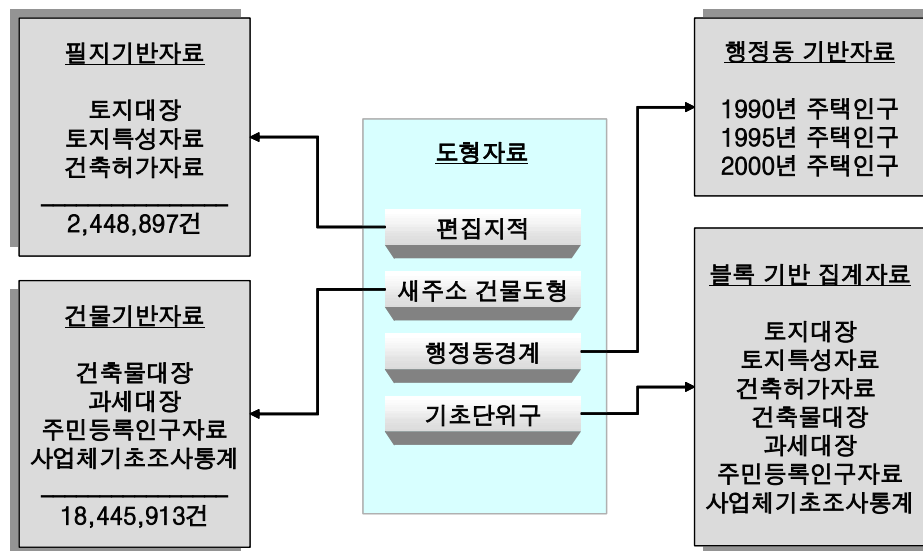
2. 도형자료 기반의 행정자료 연계 : 서울시 사례

지금까지 살펴본 자료 연계는 숫자자료와 문자자료를 연계하는 방식이다. 이와 다른 방식으로는 수치지도를 이용하여 도형자료를 중심으로 숫자 또는 문자자료

들을 서로 연계하는 방법이 있다. 수치지도를 이용하는 방식은 행정자료들의 연계성이 미흡하더라도 조사에서 지도를 이용할 수 있으며, 현장 파악이 용이하다는 이점이 있다. 통계청은 조사구 설정에 수치지도를 이용하고 있으나 관련 자료는 매우 제한적으로 활용하고 있다. 여기서는 수치지도를 좀 더 폭 넓게 이용한 서울시의 사례를 살펴봄으로써 행정자료의 통계적 활용에 주는 시사점을 찾아보고자 한다.

1) 자료 연계의 배경

서울시는 2001년부터 공간데이터웨어하우스(Spatial Data Warehouse)를 구축하는 작업을 통해 부처 내부의 자료 공유와 부처간 자료 유통체계를 확립하고자 하였다. 이 과정에서 행정자료를 연계하여 통계자료를 생산하는 시도가 있었으며, 자료 연계 과정에서 여러 가지 문제점들을 파악할 수 있었다. 여기서는 서울시의 SDW의 건물 및 주택 부분의 자료 구축 과정을 살펴보고자 한다. SDW의 전체적인 연계 구조를 나타낸 <그림 3-2>에서는 건물기반자료 구축에 해당되는 부분이다.³⁾



<그림 3-2> SDW 데이터베이스 연계 및 구축 현황

출처 : 서울시 GIS 개발팀 내부자료(2004)

서울시의 주택 수를 파악할 수 있는 자료는 통계청의 인구주택총조사, 서울시

3) 서울시의 SDW 사례에 관한 상세한 내용은 부록 참조

주택기획과와 주택재개발과, 한국주택은행의 도시주택가격동향조사 등이 있다(<표 3-4>). 서울시는 주택 수를 인구주택총조사년도의 경우 인구주택총조사에서 파악하며, 나머지 연도에는 추정치를 활용한다. 그러나 통계청의 인구주택총조사의 자료는 구별로 집계되기 때문에 동별 자료는 사용할 수 없다. 서울시 주택기획과와 주택재개발과, 한국주택은행이 다양한 주택관련 표본조사를 수행하고 있으나 조사의 표본 규모가 작기 때문에 여전히 동 단위 자료를 확보하기 어렵다. 또한 조사주관 부서와 조사목적에 따라 조사대상이나 집계기준이 달라 이들을 함께 사용하여 통계를 생산하기도 어렵다.

<표 3-4> 주택자료들과 특징

조사기관	원시자료	특징
통계청	인구주택총조사	전수조사이나 구별단위로 취합
주택기획과	서울시 내부자료	주택건설촉진법에 의해 승인된 건축물(자치구 집계)
주택재개발과	서울시 내부자료	도심재개발법에서 승인된 건축물, 자치구 단위 집계
한국주택은행	도시주택가격동향조사	샘플조사, 집계단위가 강북/강남

다른 방안으로 서울시의 행정자료를 이용하여 동별 주택 수와 건물 수를 파악하는 방안을 생각해 볼 수 있다. 이론적으로는 건축물대장과 재산세 과세대장에는 건물과 주거 관련 자료가 있기 때문에 이들을 이용하여 동별 주택 수와 건물 수를 파악할 수 있다(<표 3-5>). 그러나 건축물대장은 해당 항목의 누락률이 높아 실질적으로 이용하기 어려우며, 과세자료에는 비과세 건물이 누락되어 건물 집계에 이용하기 어렵다. 따라서 서울시의 행정자료를 직접 활용하는 것도 동별 주택 및 건물 수를 파악할 수 있는 방법이 아니다.

<표 3-5> 주택 및 건물 분야 데이터

가능 통계		자료원
주택	동별 주택 수	건물분 재산세 과세대장(추정)
	용적률	건축물대장 / 건물분 재산세 과세대장(추정)
건물	동별 건물 수	건축물대장 / 건물분 재산세 과세대장(추정)
	용적률	건축물대장 / 건물분 재산세 과세대장(추정)
	층수	건축물대장

이런 상황에서 서울시는 건축물대장과 수치지도의 건물도형을 연계하여 주택

과 건물 수를 추정하는 방법을 시도했다. 서울시에서 작성한 수치지도는 1/5,000 수치지도와 1/1,000 수치지도가 있다. 서울시의 1/5,000 수치지도는 국가지리정보시스템(National Geographical Information System) 사업의 일환으로 1997년에 제작된 것이다. 내용과 정확도에서 그다지 훌륭한 품질은 아니지만 1/1,000 수치지도가 포괄하지 못하는 산림이나 서울시 외곽을 포괄하고 있는 유일한 수치지도라는 점에서 유용하다. 항공사진 측정은 대략 1994년부터 1996년 사이에 이루어졌으며, 현지조사는 1995년에서 1997년 사이에 시행되었다. 한 가지 문제점은 1/1,000 수치지도의 동일 지역과 중첩시킬 경우 지형/지물이 일치하지 않는 경우가 많다는 점이다.

서울시의 1/1,000 수치지도는 서울시 지리정보센터에서 2003년 말에 완성한 지도다. 386개의 레이어(layer)를 갖고 있다. 1996년 처음 수치지도 작업이 시작된 이후 2001년부터 2003년까지 계속 수정·보완되었다. 현 수치지도는 서울시 전 지역을 포괄하지 못하는데, 그러한 지역은 대부분이 산악지역으로 필요할 경우 1/5,000 수치지도와 연계하여 사용한다. 수치지도에 표기된 지적은 도시계획도(1/3,000)의 지적을 사용하나 실제와 다르기 때문에 참조용으로만 사용할 뿐이다. 또한 수치지도의 행정동과 법정동 경계가 잘못 구축된 곳이 많다. 따라서 이럴 경우 다른 행정구역도(구경계도, 행정동 경계도 등)를 사용한다. 서울시의 기본도형(base map)은 <그림 3-3>과 같다.

수치지도와 연계하기 위해서는 건축물대장은 전산화가 되어 있어야 한다. 서울시의 건축물대장은 행정자치부 건축물대장 전산화 기본계획(1997-2001)에 의해 2000년에 전산화가 완료되었다. 행정자치부의 전산화된 건축물대장을 건설교통부 건축행정정보시스템으로 변환하는 작업이 2002년 7월에서 2003년 2월까지 진행되었으며, 2003년 말에 데이터베이스 변환작업이 완료되었다. 애초에는 은평구와 강남구에는 건설교통부 건축행정정보시스템이, 나머지 구에는 행정자치부 건축물대장 전산화 자료 기반의 건축물대장이 구축되었으나 2004년 현재 25개 구가 건설교통부 건축행정정보시스템으로 통합되어있다. 건축행정정보시스템은 본청의 별도 통합DB서버 없이 각 자치구에 서버를 두고 직접 관리하는 체계로 운영된다.

<표 3-6> 건물도형과 건축물대장 연계 형태

건물 : 대장	1 : 1	1 : N	N : 1	N : N
비율(%)	39.1	1.9	9.7	49.4

2002년 기준 건축물대장과 건물도형 연계율을 파악한 결과, 강남구 역삼동의 경우 건축물대장수와 건물도형을 연계할 수 없는 비율이 11.9%, 종로구 창신동은 24.8%, 서대문구 창천동은 17.6% 등으로 나타났다. 건물도형자료란 서울시 새주소 시스템에서 실시간으로 관리하고 있는 시스템으로 2002년 7월부터 서울시와 자치구청에 설치되어 활용되고 있는 상황이다. 이때 강남구 역삼동, 종로구 창신동, 서대문구 창천동 등을 시범조사 하였는데, 건축물대장과 건물도형을 연계할 수 없는 비율이 강남구 역삼동 11.9%, 종로구 창신동 24.8%, 서대문구 창천동 17.6%로 나타났다(<표 3-7>).

<표 3-7> 건축물대장과 건물도형의 연계율

연계항목 동	건축물대장수 (A)	개별연계 (B)	지번 ≠ 편집지적 (C)	미연계 (D)	미연계율 {(C+D)/A}
강남구 역삼동	4,637	4,084	30	523	11.9
종로구 창신동	3,042	2,286	227	529	24.8
서대문구 창천동	1,696	1,396	89	211	17.6

3) 도로와 건물 데이터의 중첩 사례⁴⁾

발생 원인에 따라 데이터의 오류는 정합성 오류(logical consistency error)와 기계적 오류(mechanical error)로 구분된다. 정합성 오류는 입력오류, 자료갱신 오류, 법적 기준과 현실의 괴리에서 오는 오류가 있으며, 기계적 오류는 시스템에서 발생하는 오류다. 서울시정개발연구원(2002)은 도로관리시스템에서 생성된 도로면 데이터와 새주소전산통합시스템에서 생성된 건물 데이터를 중첩하였을 경우의 오류 발생 원인을 살펴보았다. 이 두 시스템의 도로 데이터와 건물데이터는 다음과 같다.

4) 서울시정개발연구원(2002)에서 발췌

- 도로 및 도로면 데이터 : 도로관리시스템의 도로면 레이어, 1/1,000 수치지도의 도로(AD001) 레이어
- 건물 데이터 : 새주소전산통합시스템의 건물 레이어, 1/1,000 수치지도의 건물(AA001) 레이어

중구와 송파구를 선정하여 건물 데이터와 도로면 데이터를 중첩시켜 일반적인 공간관계(면대면) 검사하였다. 그 결과 두 데이터 레이어의 중첩에서 데이터의 입력 오류 및 갱신절차의 차이 등에 의한 오류가 나타났으며, 이들에 대한 구분이 불가능하였다. <표 3-8>은 중구와 송파구별로 건물 데이터와 도로면 데이터의 중첩 결과 나타난 오류의 비율을 제시한 것이다.

<표 3-8> 건물 및 도로면 레이어의 정합성 오류 집계 - 중구, 송파구

구	오류비율(%) (오류건물수 / 총건물수)	총 오류면적(m ²)	호당 오류면적(m ²) (총 오류면적/오류건물수)
중구	2,688/25,047 = 10.73	10,797	4.02
송파구	583/29,283 = 1.99	4,746	8.14

건물과 도로에서 오류는 주로 도로관리시스템이나 새주소전산통합시스템과 같은 단위업무시스템별로 1) 데이터 갱신 주기가 다르거나, 2) 지형지물 입력방법이 달라서 발생한다(<표 3-9>). 이러한 오류는 전형적인 정합성 오류다. 먼저 갱신 주기를 보면, 새주소전산통합시스템의 건물데이터는 새 건축물을 짓거나 건물의 보수 또는 증축이 있을 경우 건축물관리대장에 등재하는데 이때 갱신된다. 반면 도로관리시스템의 데이터는 도로의 인정 공고가 있을 경우 갱신된다. 이렇듯 두 데이터의 갱신 주기가 달라서 두 데이터를 연계할 때 오류가 발생할 수 있다.

또 실제측량의 부재로 데이터에 오류가 발생할 수 있다. 건물의 현황도면 데이터는 일반적으로 준공도면을 지도화한 것으로 편집지적과 연속지적의 좌표를 활용하여 기준점을 잡는다. 그런데 실제 측량을 하지 않은 상태에서 자료의 갱신이 이뤄지면 현실을 정확하게 반영하지 못하여 오류가 발생할 수 있다.

<표 3-9> 정합성 오류의 발생원인

원인		데이터 종류	도로면 데이터 (도로관리시스템)	건물 데이터 (새주소전산통합시스템)
		데이터 갱신주기 상이	인정 공고시 속성갱신 (개통 및 미개통)	건축물관리대장에 등재시 (신축 및 개보수)
단위 시스템별 데이터 갱신절차 상이	실측(측량)의 부재	도로 : 도시계획도 (1/3,000)의 도로계획선 디지털타이징	건물 : 황도면(준공도면) 디지털타이징 - 기준점 : 편집지적 및 연속지적 좌표 활용	
	갱신대상 지형지물의 기준선 상이	도로 : 도시계획선 건물 : 건물처마선 (정사투영면)	도로 : 준공도면 건물 : 건물 처마선 또는 건물 외벽	
단위 시스템별 지형지물 입력방법 상이	배경도면 지형지물 갱신방법 부재	도로 대한 체계적 갱신방법 부재	건물에 대한 체계적 갱신방법 부재	

또한 단위시스템별 지형지물을 입력하는 방법이 달라서 발생하는 오류가 있다. 이러한 오류는 갱신 대상 지형지물의 기준선이 다르거나 배경도면 지형지물의 갱신방법이 다르기 때문에 나타난다. 예를 들어 도로관리시스템의 도로의 기준선은 도시계획선이지만 새주소시스템에서 도로의 기준선은 준공도면이다. 건물 역시 도로관리시스템의 기준선과 새주소통합시스템의 기준선이 다르다. 이러한 차이로 인해 데이터의 정합성 오류가 발생한다.

이 같은 문제점을 해결하기 위해서는 단위시스템에서 데이터 갱신주기 및 절차에 관한 표준화 작업과 서로 자료를 연계할 경우의 상세한 지침이 필요하다. 이와 함께 법적으로 모든 데이터에서 공통으로 사용하는 기준을 일관되게 하는 작업들도 이뤄져야 한다. 더불어, 건축물대장 작성과 관련해서는 부서별로 사용하는 주소체계도 통일시켜야 한다.

IV. 분석 자료 및 자료 연결방법

이 연구는 통계청에서 대전시 판암1동, 인천시 강화군 화도면, 광주시 백운1동 지역의 시험조사를 통해 작성한 가구명부와 거처정보 및 해당 지역의 건축물대장을 연결한 자료를 분석한다. 통계청에서는 먼저 가구명부와 거처조사결과를 연결한 다음 여기에 건축물대장자료를 연결하였다. 마지막으로 개인정보보호를 위해 연결이 완료된 파일에서 필요한 필드만을 남기고 나머지는 삭제하였다. 여기서는 건물 및 주거와 관련 있는 행정자료의 현황과 분석 대상 지역의 특성을 살펴보고, 분석에 사용된 자료의 내역과 자료 연결 방법에 대해 논의하고자 한다.

1. 주거 부문의 행정자료 현황

먼저 인구주택총조사의 주거(주택)부문과 관련 있는 행정자료의 현황을 살펴볼 필요가 있다. 앞서 살펴보았듯이 조사구 설정에서 2000년 11월 15일 이후에 신축한 건물을 파악해야 하기 때문에 행정자료에서 건축 연도가 필요하다. 또한 행정자료에서 건물 용도, 층수, 층별 용도 등은 거처를 파악하는 데 유용한 항목이다. 더 나아가 인구주택총조사의 주거부문의 조사항목을 직접 파악할 수 있으면 행정자료의 활용도를 더욱 높일 수 있다. 인구주택총조사의 주거부문에서는 11개 항목을 조사하고 있으며, 가구부문에서도 주거시설과 관련 있는 항목을 10개 내지 12개 조사하고 있다(<표 4-1>).

<표 4-1> 인구주택총조사의 주거부문 조사 항목

주택에 관한 사항	가구에 관한 사항
1) 사는 집의 종류	1) 침실 수
2) 사는 집의 건물 층수	2) 침실 이외의 방수
3) 연건평	3) 거실 수
4) 대지면적	4) 식당 수
5) 방	5) 거주하는 층(지하, 지상, 옥상)
6) 거실	6) 부엌 형태(입식, 재래식)
7) 식당	7) 수도 형태(상수도, 간이수도)
8) 건축 연도	8) 화장실 형태(수세식, 재래식)
9) 부엌 수	9) 목욕 시설(온수, 비온수)
10) 화장실 수	10) 난방 시설
11) 독립된 출입구 수	11) 주차 시설

이러한 항목들과 관련 있는 행정자료를 살펴보면 등기부등본, 건축물대장, 건축허가대장, 한국감정원 자료 등이 있다(<표 4-2>). 한 가지 참고할 점은 행정자료의 항목들을 한마디로 무엇 무엇이라고 규정하기 어렵다는 것이다. 이러한 자료들은 지방자치단체에서 수집되어 중앙부처에 갱신되는 경우가 많은데 지방자치단체와 중앙부처에서 취급하는 항목들이 같지 않다. 또한 지방자치단체의 경우 지역마다 다루는 항목도 같지 않다. 뿐만 아니라 개인정보보호 차원에서 공개되지 않는 것들도 있어 항목의 내역을 일목요연하게 정리하기 어렵다. 여기서는 이런 한계점을 감안하고 각 행정자료에 대해 간단히 소개하고자 한다.

먼저 등기부등본부터 살펴보면, 등기부등본은 부동산의 표시와 그 부동산에 관한 권리관계를 기재해 일반인에게 공시하는 서류로 법원에서 관리하고 있다. 누구든지 등기부를 열람하거나 등본을 발급받을 수 있어서 부동산 매매에 사용된다. 등기부등본은 표제부와 갑구 및 을구로 구성되어 있다. 구체적인 내용을 살펴보면, 표제부는 토지 또는 건물의 표시와 그 변경에 관한 사항을 기재한다. 토지 관련으로는 소재지, 지번, 지목, 면적이, 건물 관련으로는 소재지, 지번, 건평, 층수, 구조, 용도 등이 포함되어 있다. 갑구는 소유권에 관한 사항(변동 및 변경사항, 압류, 가압류, 경매신청, 가등기, 가처분, 환매등기, 예고등기 등)이 포함되어 있다. 을구는 소유권 이외의 권리(지상권, 지역권, 전세권, 저당권, 임차권)에 관한 변동 및 변경사항이 포함되어 있다.

<표 4-2> 주거부문 행정자료의 자료원과 항목

자료원	항목
등기부등본(법원)	지번, 건물 명칭 및 번호, 건물내역(토지사용목적, 건물형태, 소유권 관련 사항)
건축물대장(건설교통부)	대지면적, 연면적, 주용도, 층수, 층별 면적, 층별 용도, 부속건축물, 건축 연도, 관련 지번
건축인허가대장(건설교통부)	건축용도, 지상층수, 지하층수, 허가일자, 준공일자, 연면적 등
주택감정평가 자료(한국감정원)	아파트명, 총세대수, 준공일자, 면적, 시공회사, 용적률, 입주일자, 총동수, 평형 등

건축물대장(가옥대장)은 건물의 소재, 번호, 종류, 구조, 건평, 소유자의 주소·성명 등을 등록하여 건물의 상황을 명확하게 하는 장부이다. 건축물대장은 과세의

기본이 되며, 구청과 시·군에 비치되어 있다. 토지대장 및 임야대장에 관하여는 지적법에 그 규정이 있으나, 건축물대장에 관하여는 아직 그 근거가 되는 법률이 제정되어 있지 않다. 건축물대장은, 등기소에 비치되어 가옥에 관한 권리관계를 공시(公示)하는 건물등기부와 구별된다. 건축물대장에 등록된 부동산에 관한 사항은 등기부에 기재되는 등기명의인 표시의 기초가 된다. 따라서 건물의 상황에 변동이 생겼을 때에는 먼저 건축물대장을 변경한 후에 등기변경을 신청해야 한다.

현재 정부는 전자정부 사업의 일환으로 전국 건축물대장에 대한 일괄정비 사업을 추진할 예정이다. 건설교통부는 2004년 9월부터 2005년 3월까지 7개월 동안 전국 건축물대장 1,650만 건의 소유권 및 지번정보를 일괄 정비해 데이터베이스로 구축하는 사업을 추진하고 있다. 이러한 시스템은 정부의 종합부동산세 도입을 위하여 건축물과 주택에 대한 개인별·세대별 보유현황을 실시간으로 파악할 수 있는 부동산정보관리센터 프로젝트의 일환으로 추진되고 있다. 현재 건축물대장은 등기·영업허가·매매 등에 이용되는 자료로 발생 빈도가 매우 높은 민원서류이다. 그러나 내용의 오류와 누락률이 높고 시군구마다 자료의 정확도가 달라 부동산 정책 수립을 위한 기초 자료로 활용하기에는 무리가 있다는 지적이 많았다. 이렇다 보니 등기 등 연관업무와의 시스템 연계가 힘들어 상당부분이 수작업으로 처리되어 왔다는 지적이 있다(안경애, 2004).

건축물대장은 11개의 항목으로 구분되어 있는데, 소유구분, 연면적/용적률산정용 연면적, 건축면적, 대지면적, 용적률, 건폐율, 주구조, 주용도, 준공연도, 지상층수, 지하층수 등이다. 이 11개 항목 중 인구주택총조사 관련 항목은 주용도, 대지면적, 연면적, 건축 연도 등이다. 건축물대장에서 건물의 주용도는 21개로 구분되고 있다(<표 4-3>).

건축인허가대장은 건축물을 신축할 때 시군구 지방자치단체장에게 제출하는 건축허가신청서를 토대로 만들어진다. 건축물대장보다 신속하게 처리되고 있다는 점에서 현재 조사구 설정에서 활용되고 있다. 인구주택총조사와 관련 있는 건축인허가대장의 항목은 건축물대장의 항목과 동일하다.

<표 4-3> 건축물대장의 건물 주용도 구분

번호	용도	번호	용도
1	단독주택	12	위락시설
2	공동주택	13	공장
3	제1종 근린생활시설	14	창고시설
4	제2종 근린생활시설	15	위험물저장, 처리시설
5	문화 및 집회시설	16	자동차 관련시설
6	판매 및 영업시설	17	동식물 관련시설
7	의료시설	18	분뇨쓰레기 처리시설
8	교육연구 및 복지시설	19	공공용시설
9	운동시설	20	묘지관련 시설
10	업무시설	21	관광휴게시설
11	숙박시설		

한국감정원은 토지, 건물, 기계/기구, 동산, 영업권 등 유, 무형의 재산에 대한 경제적 가치를 판정하여 그 결과를 가액으로 표시하는 역할을 하고 있다. 따라서 각 지역별 건물의 특성에 대한 많은 자료를 보유하고 있다. 아파트를 예를 들면, 아파트명, 총세대수, 준공일자, 면적, 시공회사, 용적률, 입주일자, 총동수, 평형 등 일반적인 정보뿐만 아니라 평형정보(평형, 타입, 전용면적, 방·거실·주방·욕실·현관구조 등), 주변정보(주변 아파트, 역세권 아파트 등), 주소정보(행정동, 법정동 구분), 동호정보(동 번호, 호 번호 <층+호>, 평형, 평형타입 등) 상세정보를 포함하고 있다. 부수적으로 우편번호, 용적률, 편의시설, 주차대수, 분양일자, 시세 등의 정보도 포함하고 있다. 따라서 한국감정원 자료는 부동산 거래와 같은 실생활에 가장 밀접한 자료이다. 그러나 감정평가에 대한 기초 자료를 등기부등본과 건축물대장에 근거한다는 점에서 다른 자료와 크게 다르지 않다고 볼 수 있다. 다만 인구주택총조사와 관련하여 아파트의 경우 동호 정보, 평형, 전용면적, 거실 주방, 욕실이 있어 건축물대장보다 좀 더 많은 정보를 갖고 있기 때문에 조사구 설정에서 활용되고 있다.

마지막으로, 행정자료의 연계와 관련하여 번호체계를 살펴볼 필요가 있다. 현재 건축물에 대한 번호체계는 두 개가 있는데, 이들을 <표 4-4>에 제시하였다. 첫 번째는 등기부 등본에서 사용하고 있는 번호체계이다. 이 번호는 모두 14자리로 이루어져 있으며 관할 등기소에 대한 정보, 건축물의 종류와 일련번호 등의 정보를 갖고 있다. 두 번째 번호 체계는 건축물대장의 고유번호이다. 건축물대장의 고유번호는 19자리로 이루어져 있다. 번호체계의 구조를 살펴보면, 행정구역코드(10자리)

- 대장구분(1자리) - 대표지번(8자리) 등으로 나눌 수 있다. 건축물대장의 행정구역 코드번호는 법정동·리 코드번호 사용한다. 대장구분은 토지, 임야(산), 무지번별로 각각 일반건축물은 1, 2, 0, 집합건축물은 3, 4, 5의 번호를 부여한다. 등기부등본의 번호체계는 법원의 독립적인 번호체계로 다른 자료와 연계할 수 있는 체제를 전혀 갖추지 못하고 있다. 건축물대장은 지번 중심의 번호체계를 갖고 있어 건물에 대한 고유번호를 갖지 못하고 있다. 서울시의 사례에서 논의하였듯이 건축물대장에 새주소를 사용하면 건물의 고유번호를 부여한다는 측면에서 조금 더 자료를 연계하기 수월해질 것이다.

<표 4-4> 등기부등본과 건축물대장의 고유번호 체계

등기부등본	내역	관할 등기소	건축물 종류	일련번호
	자릿수	4	4	6
건축물대장	내역	행정구역코드	대장구분	대표지번
	자릿수	10	1	8

2. 분석 자료

이 연구에서는 대전시 관암1동, 인천시 강화군 화도면, 광주시 백운1동 지역에 대한 시험조사 결과인 가구명부와 거처정보자료 및 해당 지역의 건축물대장과 이들을 연결(매칭)한 자료를 사용한다. 2005년도 인구주택총조사를 위한 시험조사는 사전 준비단계 시작되기 훨씬 전부터 시행되었다. 1차 시험조사는 2002년 10월 30일부터 11월 9일까지 경북 경주시의 54조사구 3,250가구를 대상으로, 2차 시험조사는 2003년 6월 15일부터 26일까지 서울, 경남, 전남, 충남에서 412 조사구를 대상으로, 3차 시험조사는 2003년 11월 2일부터 13일까지 부산, 강원, 충북, 전북의 4개 읍면동에 걸쳐 244 조사구를 대상으로 실시되었다. 좀 더 본격적인 시험조사는 사전 준비단계 기간인 2004년 3월 2일부터 17일까지 서울, 경기도 안성 및 보개면의 3개 지역의 94개 조사구를 대상으로, 2004년 7월 1일부터 15일까지 5개 광역시 및 6개 구·군의 9개 면·동 679개 조사구의 40,600 가구를 대상으로 실시되었다. 이 연구에서 사용하는 자료는 2004년 7월에 있었던 시험조사의 조사결과

가운데 대전시 동구 판암1동, 인천시 강화군 화도면, 광주시 남구 백운1동의 것이다.

1) 지역 특성

조사 과정은 조사대상 지역의 특성에 많은 영향을 받는다. 건축물대장 역시 지방자치단체에서 작성하기 때문에 작성방법이나 정보의 정확성이 조사대상 지역별로 편차를 보일 수 있다. 이런 측면에서 지역의 자료를 분석하여 비교하기에 앞서 각 지역의 특성에 대해 살펴볼 필요가 있다.

대전시 동구 판암1동은 대전군 외남면에 편입되어 있다가 1935년 11월 1일 대전부 신설에 따라 대덕군 외남면에 편입되었으며, 1940년 대전부 구역확장 때 판암정이 되었고, 1946년 왜식 동명 변경에 따라 판암동이 되었으며, 1992년 9월 1일 판암1, 2동으로 분동되어 현재에 이르고 있다. 판암1동은 판암 IC와 지하철 1호선의 시발점이 있어 대전의 관문으로서 기능을 하고 있다. 삼정골, 망태골, 집너머마을 등의 자연부락과 아파트단지가 공존하는 도농 복합지역이다.

광주시 남구 백운1동은 1923년 4월 1일 광주읍의 확장으로 효천면 벽도리와 주월리가 백운정이 되었으며, 1947년 8월 15일 백운동으로 개칭되었다. 1973년 7월 1일 구제 실시로 서구 관할이 되었다가, 1985년 11월 1일 백운1동과 백운2동으로 분동이 되었고, 1995년 3월 1일 서구의 분구와 함께 남구 소관으로 관할이 변경 되었다. 광주시 서남부 방면의 관문이며 도심부 주거지역의 노후주택이 많고, 도시 기반 시설이 미비하나 아파트 밀집지역으로 변모하고 있다. 2002년 주거환경개선지구로 지정되었다.

화도면이 위치한 강화군은 1995년 경기도에서 인천광역시로 통합되었으며, 현재 1읍 12면 1출장소의 행정구역으로 이루어져 있다. 마니산이 면 중심부에 위치하고 있으며 마니산을 중심으로 마을이 남북방향으로 형성되어 있다. 화도면에는 내리, 상방리, 문산리, 덕포리, 사기리, 동막리, 흥왕리, 여차리, 장화리 등 9개의 리가 존재하고 있다.

먼저 2000년도 인구주택총조사를 바탕으로 조사지역별로 주택유형의 분포를 살펴보았다(<표 4-5>). 대전시 동구 판암1동은 아파트가 80.2%로 주택의 거의 대부분을 차지 하며, 광주시 남구 백운1동은 아파트가 61.8%로 아파트의 비중이 높다.

반면 강화군 화도면은 단독주택이 93.0%로 주택의 대부분을 차지하며, 광주시 백운 1동의 경우 주택은 33.4%로 적지 않은 부분을 차지하고 있다. 이러한 주택유형의 분포는 지역의 산업유형 및 분포, 주민의 생활양식 및 문도 등과 관련이 있다. 이런 측면에서 여기서는 농어업인구 비율도 함께 살펴보고자 한다.

<표 4-5> 지역별 주택유형 분포

주택유형	지역		대전 판암1동		강화군 화도면		광주 백운1동	
	주택수	%	주택수	%	주택수	%	주택수	%
단독주택	596	14.6	1,163	93.0	996	33.4		
아파트	3,274	80.2	0	0.0	1,840	61.8		
연립주택	174	4.3	0	0.0	17	0.6		
다세대주택	0	0.0	0	0.0	0	0.0		
영업용건물내주택	40	1.0	88	7.0	126	4.2		
계	4,084	100.0	1,251	100.0	2,979	100.0		

출처 : 통계청(2002)

지역이 얼마나 농촌사회의 모습을 보전하고 있는지를 알아보기 위해 지역별 농어업 종사자 비율을 살펴보았다(<표 4-6>). 화도면의 경우 2,416명이나 되는 농가 인구 이외에 어가인구도 591명이나 되었다. 이들 농어업인구가 전체 총인구수에서 차지하는 비중은 82.6%나 되었으며 화도면이 농어업의 1차 산업을 중심으로 형성된 지역사회라는 점을 알 수 있다. 반면 다른 지역에서 농어업 인구의 비율이 낮았다. 백운1동은 0.4%로 거의 무시해도 좋을만큼 적었으며, 판암1동은 2.9%로 매우 낮은 편이었다. 비록 작은 차이지만 판암1동의 농가인구 비율이 백운1동보다 높은 것은 판암1동의 삼정골이나 망태골과 같은 일부 지역이 아직 농업지역으로 남아있기 때문인 것으로 보인다.

<표 4-6> 지역별 농어업인구 비율

지역	대전시 판암1동	강화군 화도면	광주시 백운1동
인구			
농가인구 ¹⁾	433	2,416	54
어가인구 ²⁾	--	379	--
어업종사자 ³⁾	--	212	--
농어업인구계	433	3,007	54
지역의 총인구	14,777	3,641	12,706
농어업인구비율(%)	2.9	82.6	0.4

출처 : 1) 농업총조사(2000), 2) 어업총조사 (2000), 3) 인구주택총조사(2000)

2) 자료 내역 : 자료수와 필드(변수)

대전시 판암1동, 인천시 강화군 화도면, 광주시 백운1동 지역에 대한 시험조사 결과인 가구명부 및 거처정보와 건축물대장의 자료 수는 <표 4-7>과 같다. 가구명부와 거처정보는 시험조사에서 확인 과정을 거쳐 만들어진 자료이기 때문에 자료에서 큰 차이가 있을 수 없다. 반면, 건축물대장은 행정자료로 그 내역에서 조사 시점과 차이가 있을 수 있다. 더욱이 점은 가구명부와 거처정보의 자료는 행정동 단위로 작성되었으며, 건축물대장 자료는 법정동 단위로 작성되었다. 따라서 가구명부와 거처정보의 지역별 자료 수는 비슷하지만, 건축물대장의 것과는 많은 차이가 있다.

<표 4-7> 지역별 가구명부, 거처정보, 건축물대장의 자료 수

지역	통계청 코드	가구명부	거처정보	건축물대장
1) 대전 동구 판암1동	25-010-55	4,135	4,137	1,571
2) 인천 강화군 화도면	23-310-34	1,569	1,566	2,221
3) 광주 남구 백운1동	24-030-62	3,150	3,143	7,040

가구명부와 거처정보 자료에 포함된 필드(변수)들은 <표 4-8>과 같다. 통계청에서는 가구명부와 거처정보를 연결하기 위해서 먼저 가구명부를 각 조사지역별로 입력한 다음 그들을 합쳤다. 가구명부의 각 파일별로 자료연결을 위해 매칭값을 일관성 있게 정리하였다. 입력시 일관성이 없는 법정동명을 일괄적으로 정리하여 ‘[수정]법정동명’(〈표 4-6〉의 7번)이란 필드명으로 저장하였다. 건축물대장과 매칭을 하기 위해 행정자치부의 5자리 코드인 ‘[추가]시군구코드’(8번)를 추가하였다. 같은 이유에서 행정자치부의 5자리 코드인 ‘[추가]법정동코드’(9번)를 추가하였다. 또한 거처정보와 매칭을 위해 조사구번호 앞자리의 ‘0’을 없애고, 뒷자리의 아파트와 주택구분인 ‘A’와 ‘1’을 없애 ‘[수정]조사구번호’로 저장하였다.

<표 4-8> 가구명부와 거처정보의 필드

일련 번호	가구명부		거처정보	
	필드명	설명	필드명	설명
1	가구명부ID	편의상 별도 추가	SIDO	시도코드(통계청)
2	시도코드(통계청)	통계청코드 2자리	SIGUN	시군구코드(통계청)
3	시군구코드(통계청)	통계청코드 3자리	DONG	행정동코드(통계청)
4	행정동코드(통계청)	통계청코드 2자리	JOSAGU	조사구번호
5	행정읍면동명칭		GEONO	거처번호
6	법정읍면동명칭		ADDR	
7	[수정]법정동명	일관되게 수정	GEOGBN	거처구분
8	[추가]시군구코드	행자부코드 5자리	NOKAGU	빈집
9	[추가]법정동코드	행자부코드 5자리	KEONPYUNG	연건평
10	조사구번호		DAEJIAREA	대지면적
11	[수정]조사구번호	거처정보와 통일	ROOMSU	총방수
12	거처번호		LIVINGSU	거실수
13	본번		DININGSU	식당수
14	공동주택명칭		BULDYEAR	건축연도
15	동		KITSU	부엌수
16	호		CRSU	화장실수
17	관련본번	한 거처에 여러 번지	DOORSU	출입구수
18	관련부번	한 거처에 여러 번지	KEOFLOOR	
19	재확인	불확실한 부분	INFLAG	
20			GEOGBN1	
21			WDATE	

출처 : 통계청 내부자료

건축물대장은 건설교통부가 전산화된 파일의 형태로 제공한 것을 그대로 사용하였다. 제공된 건축물대장의 필드는 <표 4-9>와 같다. 이 가운데 매칭에 필요한 필드는 시군구코드, 법정동코드, 본번, 부번이다. 인구주택총조사의 주택부문 항목과 관련 있는 필드는 연면적, 대지면적 등이 있다. 또한 조사구 설정에 유용한 필드는 주용도, 기타용도, 지상층수, 지하층수, 부속건물수 등이 있다.

<표 4-9> 건축물대장의 필드

일련번호	필드명	일련번호	필드명	일련번호	필드명
1	시군구코드	17	동기타용도	33	동승용승강기대수
2	법정동코드	18	동세대수	34	동비상용승강기대수
3	행정동코드	19	동가구수	35	동부속건물수
4	대지구분	20	동구조코드	36	동부속건물면적
5	본번	21	동구조명	37	옥내자주식대수
6	부번	22	동지붕코드	38	옥내자주식면적
7	총괄일련번호	23	동지붕명	39	옥외자주식대수
8	동일련번호	24	동건축면적	40	옥외자주식면적
9	특수지번명	25	동연면적	41	옥내기계식대수
10	총괄표제부구분	26	동용적률산정시연면적	42	옥내기계식면적
11	주부속구분코드	27	동높이	43	옥외기계식대수
12	위반건축물구분	28	동대지면적	44	옥외기계식면적
13	양성화여부	29	동건폐율	45	동허가일자
14	건축물명칭및번호	30	동용적률	46	동착공일자
15	동명칭및번호	31	동지상층수	47	동사용승인일자
16	동주용도코드	32	동지하층수		

출처 : 통계청 내부자료

3) 건축물대장 자료

분석에 앞서 자료의 특성을 살펴보는 것은 자료를 이해하는 데 도움이 된다. 특히 건축물대장은 통계적 목적으로 작성된 것이 아니기 때문에 통계분석에 여러 가지 한계를 갖기 쉽다. 그러한 한계 가운데 하나가 자료의 누락이다. 아래 <표 4-10>은 건축물대장의 주요 필드에 대한 누락률을 지역별로 제시한 것이다. 대전 판암1동, 강화군 화도면, 광주 백운1동 가운데 광주 백운1동의 누락률이 유난히 높다. 가장 누락률이 높은 것은 용도코드로 99.3%이다. 그렇지만 용도명의 누락률은 5.8%로 용도에 대해서는 기록이 없다기보다는 단지 수치화가 되어있지 않을 뿐이다. 실제로 가장 높은 누락률은 준공년도로 93.6%에 이른다. 이와 비슷하게 구조도 구조명은 6.5%만 누락되어 있으나 코드는 43.%가 누락되어 전반적으로 수치화가 잘 되어 있지 않다. 그 다음으로는 용적률, 건폐율, 지상층수는 모두 3,069개의 자

료가 누락되어 있으며, 대지면적과 연면적도 각각 3,063, 3,060 개의 자료가 누락되어 있어 43.6%의 누락률을 보이고 있다. 이렇듯 높은 누락률은 광주시 백운1동의 행정력이 아직 제대로 전산화되어 있지 않기 때문에 나타난 현상인 것으로 보인다.

<표 4-10> 건축물대장의 주요 항목별 누락률

동 항목	대전 판암1동		강화군 화도면		광주 백운1동	
	누락수	비율(%)	누락수	비율(%)	누락수	비율(%)
연면적	0	0.0	0	0.0	3,060	43.5
건축면적	9	0.6	0	0.0	414	5.9
대지면적	26	1.7	55	2.5	3,063	43.5
용적률	26	1.7	52	2.3	3,069	43.6
건폐율	26	1.7	52	2.3	3,069	43.6
구조(코드)	132	8.4	20	0.9	3,066	43.6
구조(명)	2	0.1	6	0.3	459	6.5
용도(코드)	10	0.6	13	0.6	6,988	99.3
용도(명)	4	0.3	5	0.2	407	5.8
준공년도	13	0.8	879	39.6	6,589	93.6
지상층수	3	0.2	2	0.1	3,069	43.6
전체수	1,571		2,221		7,040	

주 : 용도(명)은 '동기타용도', 준공년도는 '동사용승인일자'

한편 대전시 판암1동이나 강화군 화도면의 경우는 대부분 항목들의 누락률은 낮은 편이다. 다만 강화도 화도면의 경우 준공년도의 누락률이 39.6%로 다른 항목에 비해 상당히 높으며, 대전시 판암1동의 구조코드가 8.4%로 상대적으로 높은 편이다. 강화도의 경우는 오래된 건물이나 부속건물 등이 많기 때문으로 보이며, 판암1동의 경우는 수치화가 아직 되지 않기 때문으로 보인다.

참고로 서울시 강남구 역삼동, 종로구 창신동, 서대문구 창천동 건축물대장의 항목별 누락률을 살펴보았다(<표 4-11>). 일단 대전시 판암1동이나 강화도 화도면에 비해 누락률이 높다. 또한 전반적으로는 건축면적, 대지면적, 용적률, 건폐율 등 측정이 필요한 항목에서 누락률이 높은 편이다. 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.

<표 4-11> 서울시 건축물대장의 주요 항목별 누락 비율 (2002년 기준)

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

그렇지만 서울시의 누락률은 지역의 주택 특성과 어느 정도 관련이 있어 보인다. 그러나 여기서 분석할 지역의 건축물대장 누락률에서는 일정한 지역 주택의 특성과의 연관성을 발견하기 어렵다. 그러한 차이의 일부는 지역의 주택 특성에 따라 발생하는 것이지만, 다른 부분은 지역자치단체의 행정적 관리능력 차이에서 발생한 것일 가능성이 크다. 이런 측면에서, 앞으로 행정자료를 통계에 활용하기 위해서는 지역별로 차이가 나는 행정자료의 (수치)관리 능력을 어느 정도 일관된 수준으로 끌어 올린 다음 지역의 특성에 따라 자료 누락의 문제를 해결하는 것이 바람직하다.

3. 자료의 연결

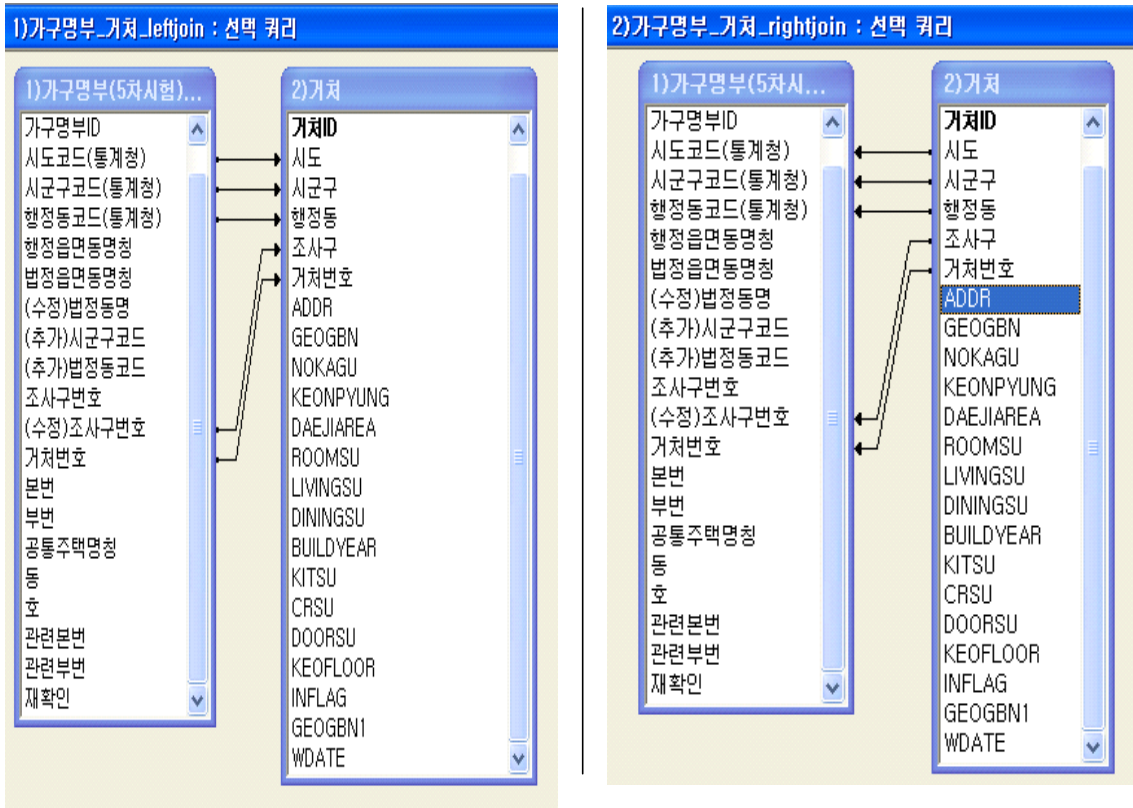
1) 통계청 제공 자료의 연결 방법과 문제점

(1) 가구명부, 거처, 건축물대장의 연결방법

이 연구에서 분석에 사용한 자료는 가구명부, 거처, 건축물대장을 연결(매칭)한 것이다. 통계청에서는 각각의 자료를 데이터베이스의 조인(join)을 사용하여 연결하였다. 먼저 두 자료(테이블)를 연결하기 위해서는 일련번호와 같은 고유의 값이 각각의 자료에 있어야 하며 이들은 같은 형식의 정보이어야 한다. 두 자료는 이 정보를 연결고리로 하여 연결되는데 여기서는 이러한 정보를 매칭값이라고 부른다. 가구명부, 거처, 건축물대장을 연결하는 매칭값은 가구와 거처가 있는 건물의 주소를 중심으로 만들었다. 상세한 설명은 뒤에 개별 자료 연계에서 다룬다.

일반적으로 데이터베이스의 조인에는 세 가지 방식이 사용된다. 첫 번째 방식은 내부조인으로 두 자료에서 매칭값이 일치하는 경우만을 연결한다. 두 번째 방식은 왼쪽 외부조인으로 왼쪽 자료(명함 테이블)의 모든 자료에 오른쪽 자료(메모 테이블) 가운데 왼쪽 자료의 매칭값과 일치하는 자료만을 선별하여 연결한다. 세 번째 방식은 오른쪽 외부조인으로 오른쪽 자료의 모든 자료에 왼쪽 자료 가운데 오른쪽 자료의 매칭값과 일치하는 자료만 선별하여 연결한다.

통계청에서는 가구명부와 거처조사 결과를 연결한 후 이 자료에 건축물대장 자료들을 연결하였다. 자료의 연결에는 데이터베이스의 쿼리(query) 기능을 사용하였다. 가구명부와 거처조사결과자료를 왼쪽 조인하고 다시 오른쪽 조인하였다(<그림 4-1>). 두 자료를 연결할 때 사용한 매칭값은 시도코드, 시군구코드, 행정동코드, 조사구 번호, 거처번호이다. 연결의 과정에서, 일단 왼쪽 조인을 통해 가구명부에 있는 모든 자료에 대해 이와 상응하는 거처자료의 개별 자료들이 모두 연결되었으며, 오른쪽 조인을 통해 나머지 거처자료가 추가되었다. 자료 연계 결과 가구명부와 거처정보는 대체로 잘 매칭 되어 3개 지역의 총 8,856 가구 가운데 11건수만 매칭되지 않고 8,845 건은 매칭 되었다(<표 4-12>).



<그림 4-1> 가구명부와 거처정보 매칭

출처 : 통계청 내부자료

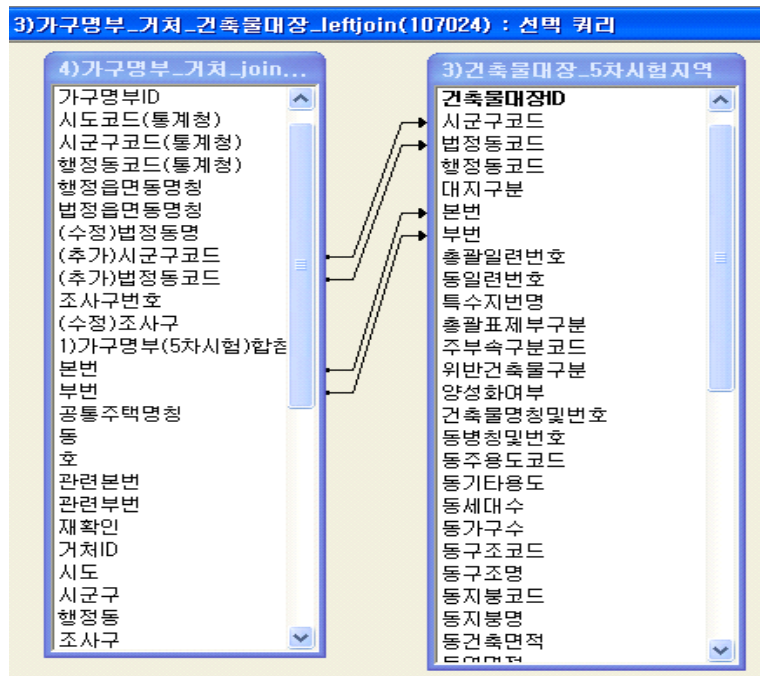
<표 4-12> 가구명부와 거처 매칭 및 비매칭 자료수

구분 \ 지역	대전시 판암1동	강화군 화도면	광주시 백운1동	계
매칭	4,136	1,566	3,143	8,845
가구명부 비매칭	0	3	7	10
거처정보 비매칭	1	0	0	1
계	4,138	1,569	3,151	8,856

이렇게 만든 자료에 다시 건축물대장을 연결하였다. 여기에서는 가구명부/거처정보가 기준이 되고 있다는 점에서 왼쪽 외부조인 방식으로 자료를 매칭 하였다. 매칭값은 시군구코드, 법정동코드, 본번, 부번을 사용하였다(<그림 4-2>).

여기서 한 가지 문제점은 통계청에서 사용한 쿼리 방식의 매칭으로 산출된 자료에는 중복자료가 존재하게 된다. 특히 아파트 단지와 같이 동일한 지번에 여러

체의 건물이 존재할 경우 중복자료가 생성된다. 한 예로, 건물 5동이 있는 아파트 단지의 경우 건축물대장에는 한 지번에 5동의 자료가 개별적으로 있다. 이 아파트의 가구명부-거처 자료와 건축물대장을 지번으로 연결하면 한 거처에 그 지번의 주거용 건물 5개가 모두 연결되어 중복자료가 만들어지면서 자료수가 5배로 증가한다. 이런 이유에서 대전 판암1동의 경우 가구명부에는 거처수(자료수)가 4,136개지만, 건축물대장과 자료의 연결을 통해 생성된 자료(거처)수가 27,678개에 이른다 (<표 4-13>).



<그림 4-2> 가구명부-거처정보와 건축물대장 매칭

출처 : 통계청 내부자료

<표 4-13> 가구명부-거처와 건축물대장 매칭률 - 전체자료와 중복제거후 자료

구분 \ 지역	대전시 판암1동	강화군 화도면	광주시 백운1동	계
전체자료수(중복허용)	27,678	2,870	8,071	38,619
① 중복제거후 자료수	3,828	923	2,910	7,661
② 가구명부-거처 자료수	4,136	1,566	3,143	8,845
매칭률(%) (①/②)	92.6	58.9	92.6	86.6

중복자료는 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 사용되었다. 그 다음으로 원래 테이블에서 기존 테이블의 끝에 쿼리 결과 집합의 레코드를 추가하는 실행 쿼리인 추가쿼리를 만들어 새 테이블로 실행한다. 이런 과정을 통해서 기본 키 필드에는 중복자료가 들어갈 수 없으므로 중복자료가 없는 테이블이 만들어진다. 이런 방식으로 만들어진 자료는 대전시 판암1동이 3,828건, 강화동 화도면인 923건, 광주시 백운1동인 2,910건으로 각각 가구명부-거처 연결 자료의 92.6%, 58.2%, 92.6%에 해당되었다(<표 4-13>).

(2) 데이터베이스 기반 매칭의 문제점

그러나 이러한 방식으로 연결한 자료도 그다지 쓸만한 것은 아니다. 데이터베이스의 JOIN기능으로 중복자료를 제거하면, 중복된 매칭 항목 중 임의의 매칭 항목만이 남게 된다. 중복자료 가운데 어느 하나를 임의로 골라 매칭 하기 때문이다. 또한 매칭 되지 않은 거처 항목들이 배제된다. 중복자료의 경우 매칭 된 자료는 실제 거처와 건축물이 정확히 연결된 것이 아니며, 건축물대장과 매칭 되지 않은 경우는 자료 자체에서 사라지기 때문에 자료의 정확성이나 포괄성이 떨어진다.

가구명부-거처 자료와 건축물대장을 연결하는 데 또 다른 문제가 있다. 이것은 가구명부와 거처 자료는 행정동을 사용하는데 반해 건축물대장은 법정동을 사용한다는 점이다. 아래 <표 4-14>는 행정동으로 표기된 조사지역의 법정동을 기술한 것이다. 이를 보면 하나의 행정동에 여러 법정동이 포함되어 있으며, 하나의 법정동에 다른 행정동이 함께 섞여 있다. 따라서 가구명부와 거처자료에는 있으나 건축물대장에 없는 건물이 있으며, 반대로 건축물대장에는 있으나 가구명부와 거처자료에는 없는 건물이 있기 마련이다. 이 때문에 가구명부와 거처자료를 건축물대장과 연결할 때 연결이 되지 않는 부분이 발생할 수 밖에 없다.

<표 4-14> 조사지역에 해당하는 법정동명

지역(행정동)	행정동에 포함된 법정동	법정동에 포함된 행정동
대전시 동구 판암1동	판암동, 삼정동	
인천시 강화군 화도면	화도면, 내리, 상방리, 문산리, 덕포리, 사기리, 동막리, 흥왕리, 여차리, 장화리	
광주시 남구 백운1동	서동, 백운동	백운2동, 주월2동

2) 수작업을 통한 자료연결

현재 데이터베이스의 기능을 기반으로 통계청 자료를 연결할 때 발생하는 문제를 해결할 수 있는 알려진 방법은 없다. 중복자료의 경우 중복이 나타나는 이유가 각양각색이며, 중복된 정보 가운데 어느 것이 적절한 것인지를 파악할 수 있는 근거가 없기 때문이다. 또한 행정동과 법정동 처럼 실제 혼란으로 말미암아 발생하는 문제도 개별 자료마다 모두 다르다. 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 가능하면 매칭 자료를 만들었다. 수작업 매칭에서는 가구명부와 거처조사결과를 연결한 자료와 건축물대장 자료를 직접 하나 하나씩 대조하고 연결하여 자료를 생성하였다. 수작업을 통해 개별 자료마다 발생하는 문제를 파악하고, 그러한 문제의 원인을 찾아보며, 논리적으로 의미 있는 유형으로 구분하고자 하였다.

먼저 지역별로 매칭률을 비교하기 위해 데이터베이스를 사용한 매칭의 결과와 직접 자료를 대조하였다. 그 결과는 <표 4-15>에 제시하였다. 데이터베이스를 사용한 매칭 자료는 'DB매칭'으로, 직접 수작업으로 대조하여 매칭 한 자료는 '대조매칭'으로 표시하였다. 매칭에서 지번을 매칭값으로 사용하기 때문에 법정동과 지번이 일치하는 건축물대장이 존재하는 거처는 건축물대장 자료와 매칭 할 수 있다. 이렇게 매칭 한 건수를 전체 거처수로 나누어 그 비율을 매칭률로 하였다.

<표 4-15> 지역별 가구명부-거처정보와 건축물대장의 매칭률 비교

구분 \ 지역	대전 판암1동		강화군 화도면		광주 백운1동	
	DB매칭	대조 매칭	DB매칭	대조 매칭	DB매칭	대조 매칭
매칭 된 거처수	3,828	3,841	923	926	2,910	2,902
매칭 안된 거처수	308	295	643	640	233	241
총 거처수*	4,136	4,136	1,566	1,566	3,143	3,143
매칭률(%)**	92.6	92.9	58.9	59.1	92.6	92.3

* 총 거처수 = 가구명부-거처정보의 거처수

** 매칭률 = (매칭된 거처수/총 거처수) x 100

데이터베이스를 사용한 매칭의 결과와 직접 대조를 통한 매칭 결과에 차이가 나타났다. 그 차이는 다음과 같은 원인들에 따른 것이었다. 우선 지번이 '산 번지'로 되어 있는 경우이다. 데이터베이스 매칭에서는 '산'이라는 표기 때문에 해당하는 건축물대장이 없어서 매칭이 되지 않았던 반면, 수작업에서는 '산'을 제외한 번지수를 사용하여 건축물대장 자료를 매칭 하였다. 다음으로는 수작업의 결과 데이터베이스 작업에서 발생하는 상이한 오류들을 교정하였다. 이러한 오류들은 가구명부 작성 과정에서 발생하였거나, 한 지번에 여러 주택유형이 존재함으로써 나타난 오차 때문에 나타난 것이다. 데이터베이스를 사용한 매칭 결과에 비해 직접 대조하여 매칭한 경우 대전 동구 판암1동과 인천 강화군 화도면에서는 다소 매칭률이 높아진 반면, 광주 남구 백운1동에서는 다소 매칭률이 낮아졌다.

매칭률로 보면 대조매칭은 DB매칭과 그다지 큰 차이가 없다. 그렇지만 대조매칭은 DB매칭과 내용면에서 많은 차이를 갖는다. 특히 한 지번에 여러 채의 건물이 있거나 여러 주택유형이 있는 경우 DB매칭을 사용하면 1 단계에서 중복자료가 만들어지고, 2 단계 작업에서 중복을 제거한다. 이때 건축물대장의 여러 자료 가운데 하나가 임의로 선택되어 거처정보와 연결된다. 반면 대조매칭에서는 이를 모두 확인하여 가장 적합한 건축물대장 자료를 선별하여 연결시키기 때문에 DB매칭에 비해 정확한 자료를 확보하게 된다. 가장 대표적인 경우가 아파트로 DB매칭에서는 거처정보에 아파트의 관리실 자료가 연결될 수도 있지만, 대조매칭에서는 그 거처에 해당되는 건축물대장 자료가 연결된다. 아래 이러한 DB매칭과 대조매칭의 질적 차이를 한 아파트 지역의 자료 연결 예시를 통해 보여주고자 한다.

지역별로 매칭률을 살펴보면 큰 차이가 나타났다. 대전 동구 판암1동과 광주

남구 백운1동의 경우에는 모두 90% 이상의 매칭률을 나타낸 반면, 인천 강화군 화도면의 매칭률은 60%에 못 미쳤다. 이러한 지역별 차이에 대해서는 여러 가지 원인이 있을 수 있다. 도시 지역에 비해 농어촌 지역에서 가건물을 짓기가 쉽기 때문에 농어촌 지역의 대장에는 존재하지 않는 건축물이 많을 가능성이 상대적으로 높다. 용이할 수 있다. 또한 농어촌 지역에서 행정자료의 관리가 상대적으로 허술할 가능성이 있다. 앞에서 인구수에 대한 주민등록자료의 총계와 인구주택총조사의 자료를 비교해 봤을 때 강화군 화도면에서 차이가 상대적으로 컸다. 건축물대장이나 주민등록자료 모두 행정자료로 실제에 맞지 않을 가능성이 일반적으로 농어촌지역에서 더욱 클 수 있음을 시사한다.

• DB매칭과 대조매칭의 질적 차이 예시

여기서는 DB매칭과 대조매칭의 질적 차이를 실제 자료를 통해 보여주고자 한다. <표 4-16>은 가구명부-거처정보 자료와 이에 상응하는 건축물대장 자료이다. 이 자료를 보면 판암1동 178-0번지는 판암주공아파트로 101동부터 108동까지 8동의 건물로 구성되어있다. 108동에는 10 거처에 10 가구가 살고 있다.

<표 4-16> 동일 지번의 거처정보 일부와 건축물대장 예시

거처정보							건축물대장					
거처번호	시군구 코드	법정동 코드	본번	부번	공동주택명칭	동 호	시군구 코드	법정동 코드	본번	부번	건축물명칭	동명칭
1	30110	10700	178	0	판암주공아파트	108 101	30110	10700	178	0	주공아파트	101동
2	30110	10700	178	0	판암주공아파트	108 102	30110	10700	178	0	주공아파트	102동
3	30110	10700	178	0	판암주공아파트	108 103	30110	10700	178	0	주공아파트	103동
4	30110	10700	178	0	판암주공아파트	108 104	30110	10700	178	0	주공아파트	104동
5	30110	10700	178	0	판암주공아파트	108 105	30110	10700	178	0	주공아파트	105동
6	30110	10700	178	0	판암주공아파트	108 106	30110	10700	178	0	주공아파트	106동
7	30110	10700	178	0	판암주공아파트	108 107	30110	10700	178	0	주공아파트	107동
8	30110	10700	178	0	판암주공아파트	108 108	30110	10700	178	0	주공아파트	108동
9	30110	10700	178	0	판암주공아파트	108 109	30110	10700	178	0	주공아파트	109동
10	30110	10700	178	0	판암주공아파트	108 110	30110	10700	178	0	주공아파트	110동

가구명부-거처와 건축물대장을 시군구코드, 법정동코드, 본번, 부번을 매칭값으로 하여 왼쪽 외부조인으로 연결한다(<그림 4-2>). <표 4-16>에 있는 자료는 모두 지번이 '대전시 동구 판암1동 178-0'로 같기 때문에 왼쪽에 있는 각각의 거처정보에 대해 101동부터 108동까지 8개의 건축물대장 자료가 존재한다. 따라서 한 거처당 101, 103, 105, 104, 106, 107, 108, 102동의 8개의 건축물 대장이 매칭 된다(<표 4-17>).

<표 4-17> 데이터베이스 매칭 결과 - 중복 허용

거처 번호	거처정보							건축물대장	
	시군구 코드	법정동 코드	본번	부번	공동주택명칭	동	호	건축물명칭 및 번호	동명칭 및 번호
1	30110	10700	178	0	판암주공아파트	108	101	주공아파트	101, 102, 103, 104, 105, 106, 107, 108
2	30110	10700	178	0	판암주공아파트	108	102	주공아파트	101, 102, 103, 104, 105, 106, 107, 108
3	30110	10700	178	0	판암주공아파트	108	103	주공아파트	101, 102, 103, 104, 105, 106, 107, 108
4	30110	10700	178	0	판암주공아파트	108	104	주공아파트	101, 102, 103, 104, 105, 106, 107, 108
5	30110	10700	178	0	판암주공아파트	108	105	주공아파트	101, 102, 103, 104, 105, 106, 107, 108
6	30110	10700	178	0	판암주공아파트	108	106	주공아파트	101, 102, 103, 104, 105, 106, 107, 108
7	30110	10700	178	0	판암주공아파트	108	107	주공아파트	101, 102, 103, 104, 105, 106, 107, 108
8	30110	10700	178	0	판암주공아파트	108	108	주공아파트	101, 102, 103, 104, 105, 106, 107, 108
9	30110	10700	178	0	판암주공아파트	108	109	주공아파트	101, 102, 103, 104, 105, 106, 107, 108
10	30110	10700	178	0	판암주공아파트	108	110	주공아파트	101, 102, 103, 104, 105, 106, 107, 108

데이터베이스의 추가 쿼리를 이용하여 중복자료를 제거하면 건축물대장 자료 가운데 임의로 101동 자료를 선택하여 가구명부-거처정보와 매칭 한다(<표 4-18>). 이와 같은 오류는 추가로 매칭값을 사용한다든지 다른 알고리즘을 찾든지 하지 않으면 해결하기 어렵다. 그러나 어떤 매칭값을 사용할지 어떤 알고리즘이 필요한지 파악하기 위해서는 먼저 자료의 속성을 자세히 이해하고 있어야 한다. 그런 측면에서 수작업으로 각각의 자료를 대조하여 매칭하는 작업을 통해 분석할 자료를 만들었다. 그러한 방식으로 만든 자료는 <표 4-19>에 제시되어 있다. 수작업 결과 아파트의 동 번호를 찾아 가구명부-거처정보와 건축물대장을 적절히 매칭 할 수 있었다.

<표 4-18> 데이터베이스 매칭 결과 - 중복 삭제

거처정보								건축물대장	
거처 번호	시군구 코드	법정동 코드	본번	부번	공동주택명칭	동	호	건축물명칭 및 번호	동명칭 및 번호
1	30110	10700	178	0	관암주공아파트	108	101	주공아파트	101동
2	30110	10700	178	0	관암주공아파트	108	102	주공아파트	101동
3	30110	10700	178	0	관암주공아파트	108	103	주공아파트	101동
4	30110	10700	178	0	관암주공아파트	108	104	주공아파트	101동
5	30110	10700	178	0	관암주공아파트	108	105	주공아파트	101동
6	30110	10700	178	0	관암주공아파트	108	106	주공아파트	101동
7	30110	10700	178	0	관암주공아파트	108	107	주공아파트	101동
8	30110	10700	178	0	관암주공아파트	108	108	주공아파트	101동
9	30110	10700	178	0	관암주공아파트	108	109	주공아파트	101동
10	30110	10700	178	0	관암주공아파트	108	110	주공아파트	101동

<표 4-19> 수작업 매칭 결과

거처정보								건축물대장	
거처 번호	시군구 코드	법정동 코드	본번	부번	공동주택명칭	동	호	건축물명칭 및 번호	동명칭 및 번호
1	30110	10700	178	0	관암주공아파트	108	101	주공아파트	108동
2	30110	10700	178	0	관암주공아파트	108	102	주공아파트	108동
3	30110	10700	178	0	관암주공아파트	108	103	주공아파트	108동
4	30110	10700	178	0	관암주공아파트	108	104	주공아파트	108동
5	30110	10700	178	0	관암주공아파트	108	105	주공아파트	108동
6	30110	10700	178	0	관암주공아파트	108	106	주공아파트	108동
7	30110	10700	178	0	관암주공아파트	108	107	주공아파트	108동
8	30110	10700	178	0	관암주공아파트	108	108	주공아파트	108동
9	30110	10700	178	0	관암주공아파트	108	109	주공아파트	108동
10	30110	10700	178	0	관암주공아파트	108	110	주공아파트	108동

V. 지역별 비교 분석

1. 개요

이 장에서는 수작업으로 가구명부-거처정보와 건축물대장을 연결한 자료를 사용한다. 다루고자 하는 내용은 다음과 같다.

첫째, 자료의 매칭에서 발생하는 다양한 경우들을 유형화한다. 다수의 사례에 대한 일반화를 바탕으로 자료 매칭의 유형화를 제시함으로써 사례의 단편적인 예시를 넘어 다양한 가능성들의 상대적인 분포를 파악할 수 있다. 이를 바탕으로 유형별로 건축물대장 자료의 활용가능성을 평가함으로써 정책 대안을 제시할 수 있는 기초를 마련한다.

둘째, 자료의 매칭 유형들의 상대적인 빈도를 지역별로 파악하여 비교한다. 이를 위해 이 장에서는 먼저 해당 지역의 특성을 살피고, 지역별로 자료 매칭의 유형이 어떻게 나타나는지를 파악한다.

셋째, 건축물대장 정보가 거처에 대한 조사정보와 실제 어느 정도 매칭 되는지를 살펴보고, 활용 가능성을 평가한다.

넷째, 자료를 직접 대조하면서 매칭시키는 과정에서 발견된 문제의 사례들을 제시함으로써 매칭 작업의 정확성을 높이는데 필요한 정보를 제공한다.

끝으로 분석결과를 요약하고 이를 바탕으로 앞으로 건축물대장 자료를 인구주택총조사에 사용하는데 도움이 될 수 있는 정책 대안을 제시한다.

2. 가구명부-거처정보와 건축물대장의 매칭 유형

지번을 매칭값으로 사용하여 가구명부-거처 정보와 건축물대장 정보를 매칭할 때 발생 가능한 여러 유형들을 나누어 볼 수 있다. 우선 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 5-1>과 같이 세분화할 수 있으며 이에 따라 건축물대장 정보의 활용정도와 방법이 크게 달라진다.

<표 5-1> 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 매칭 된 대단위 공동주택	건축물 동수 구별 후 활용
	유형3	대단위 공동주택이 아니면서 거처 다수와 건축물정보 1건이 매칭 된 경우	공동주택 확인 후 활용
	유형4	거처 1곳에 1개의 주택과 다수의 기타용도 건축물정보가 매칭 된 경우	주택 이외의 건축물이 주택의 부속건물인 경우 활용
	유형5	거처 1곳에 다수의 주택이나 시설에 대한 건축물정보가 매칭 된 경우	실제 조사에서 대조자료로서 활용
	유형6	한 지번에 다수의 거처와 다수의 건축물정보가 동시에 매칭 된 경우	실제 조사에서 대조자료로서 활용
비매칭	유형7	거처에 대한 건축물정보가 없는 경우	활용 불가능

유형1은 거처 1곳과 건축물대장 1건이 일대일로 매칭 된 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 대단위 공동주택의 경우이다. 이 경우는 다수의 거처에 건축물대장 1건을 매칭 할 수 있다. 지번을 매칭값으로 사용할 경우 다대일의 매칭인 경우도 있으나 다대다의 매칭으로 나타나는 경우도 있다. 예컨대 한 지번에 여러 동의 아파트와 상가동이 존재하는 경우 한 거처에 여러 동의 건축물대장 자료와 상가동의 건축물대장 자료가 모두 매칭 된다. 데이터베이스 프로그램의 조인(JOIN) 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. 이를 위해서는 현재 사용한 것과는 다른 매칭방법이 필요하다. 거처의 동 명칭과 이에 해당하는 건축물대장 항목을 매칭값으로 사용하여 좀 더 나은 매칭 결과를 얻을 수 있다. 거처의 동 명칭을 사용하여 매칭값으로 사용하기 위해서는 다양한 동 명칭(예: 101동, A동, 가동, 셋별동 등)을 인식할 수 있는 방법이 선행되어야 한다.

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 생활근린 시설 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 다수의 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. 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.

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

유형7은 거처에 해당하는 건축물대장 자료가 존재하지 않는 경우이다. 이 유형은 1) 건축물로 등록이 되지 않은 무허가 건물인 경우, 2) 등록된 건축물대장 정보가 전산입력시 누락된 경우, 3) 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 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. 지역별 자료의 매칭 유형 비교

1) 지역별 유형 비교

지역별로 매칭 유형을 비교한 결과를 <표 5-2>에 제시하였다. 지역별로 매칭 유형의 차이가 두드러지게 나타났다. 거처들 중 매칭이 되지 않은 경우인 유형7은 앞에서 매칭률의 비교에서 보았듯이 강화군 화도면에서 두드러지게 많았다. 지역별로 매칭 유형의 차이가 특히 두드러졌던 것은 대단위 공동주택의 매칭유형인 유형2였다. 대전 동구 판암1동에서는 유형2가 가장 많아 전체 거처의 80% 이상에 달했

으며, 광주 남구 백운1동에서도 절반 이상의 거처에 해당되었다. 반면 인천 강화군 화도면에서는 유형2가 차지하는 비율이 0.6%로 극히 소수에 불과했다.

<표 5-2> 가구명부-거처조사자료와 건축물대장의 지역별 매칭유형 분포

지역 유형	대전 판암1동		강화군 화도면		광주 백운1동	
	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

다른 유형들에서도 지역별로 주목할 만한 차이가 나타났다. 다수의 거처에 한 개의 건축물대장 자료가 매칭 되는 경우인 유형3은 특히 강화군 화도면에서 11.6%나 되어 상대적으로 많이 나타났다. 강화군 화도면에 건축물대장으로 등록되지 않은 건축물이 많았던 사실을 고려해 보면 유형3의 경우 다세대 주택이기보다는 해당 지번에 건축물대장에 있는 건축물과 건축물대장에 없는 건축물이 공존하는 사례일 가능성이 크다.

하나의 거처에 주택과 부속건물로 추정되는 건축물대장 자료가 매칭 된 경우인 유형4는 광주 백운1동에서 두드러지게 많아 전체 거처의 21.0%나 되었다. 실제 광주 백운1동에 별개의 독립된 부속건물을 가지는 주택유형이 많은 것인지, 아니면 건축물대장의 작성 방식에서 생긴 차이인지에 대해서는 실제 조사를 통한 확인이 필요하다.

다수의 거처에 다수의 건축물대장 자료가 매칭 되는 경우인 유형6은 강화군 화도면에서 13.0%로 가장 많았다. 유형6이 강화군 화도면에서 상대적으로 많은 까

닭은 농어촌 지역에서 한 필지에 다수의 건축물이 있는 경우가 많기 때문인 것으로 보인다.

지역별 각 유형의 차이는 매칭이 된 경우를 기준으로 각 유형이 차지하는 비율을 계산했을 때 더욱 두드러지게 나타난다. <그림 5-1>에 제시한 결과를 보면 인천 강화군 화도면의 경우 매칭 된 경우의 절반 가량이 일대일 대응인 유형1에 해당되었다. 대전 동구 판암1동에서는 매칭 된 경우 95% 이상이 일대일 대응(유형1)이거나 대단위 공동주택(유형2)이었으며, 그 외 유형은 극히 일부다.

<그림 5-1> 매칭이 된 경우에서 각 유형이 차지하는 비율

조사구별 유형교차의 결과 역시 조사구별로 유형에 상당한 차이가 있다(<표 5-3, 5-4, 5-5>). 이러한 결과는 일차적으로 아파트 지역의 조사구에서는 모두 공동주택의 형태로 나타나기 때문이다. 대전 판암1동에서는 아파트가 차지하는 비중이 커서 상대적으로 조사구별 차이가 전체적으로 크지 않은 반면, 일부 조사구(0061, 0071 등)에 매칭이 되지 않는 경우인 유형7이 집중되어 있음을 볼 수 있다. 강화군 화도면의 경우 유형7은 조사구 0071, 0021에 상대적으로 집중되어 있으며, 유형3은 조사구 0161과 0171에 상대적으로 집중되어 있다. 광주 백운1동의 경우 유형7이 조사구 028A, 029A에 집중되어 있으며, 유형4가 특히 조사구 0511에서부

터 0591까지 집중되어 있다.

<표 5-3> 조사구별 매칭 유형 : 대전 판암 1동

조사구	유형 1	유형 2	유형 3	유형 4	유형 5	유형 6	유형 7	계	조사구	유형 1	유형 2	유형 3	유형 4	유형 5	유형 6	유형 7	계
0011	23		2		1	6	10	42	038A		60						60
0021	23	10	2		3		20	58	039A		70						70
0031	17					5	21	43	040A		70						70
0041	19		5				27	51	041A		70						70
0054							2	2	042A		60						60
0061	1		1				49	51	043A		60						60
0071	2		4				46	52	044A		60						60
0081	11		3	1	1	7	25	48	045A		60						60
0091	27		5	1	2	15	15	65	0461	16	21	5	1	3		5	51
010A		60						60	0471	12		2			6	8	28
011A		60						60	0481	16		18				10	44
012A		60						60	0491	13	31	2		1	6	8	61
013A		60						60	0501		54						54
014A		60						60	051A		66						66
015A		58						58	052A		67						67
016A		60						60	053A		60						60
017A		60						60	054A		60						60
018A		60						60	055A		60						60
019A		60						60	056A		54						54
020A		60						60	057A		60						60
021A		60						60	058A		60						60
022A		60						60	059A		60						60
023A		58						58	060A		60						60
024A		60						60	061A		54						54
025A		60						60	062A		60						60
026A		60						60	063A		60						60
027A		50						50	064A		60						60
028A		60						60	065A		60						60
029A		40						40	0661	39		4	1		1	5	50
030A		50						50	067A		68						68
031A		50						50	068A		68						68
032A		50						50	069A		68						68
033A		70						70	070A		72						72
034A		70						70	0711							5	5
035A		60						60	0721	27		4			3	22	56
036A		60						60	0731	11		4			20	17	52
037A		60						60	계	257	3,439	61	4	11	69	295	4,136

<표 5-4> 조사구별 매칭 유형 : 강화 화도면

조사구	유형1	유형2	유형3	유형4	유형5	유형6	유형7	계
0011	16		2		1	12	27	58
0021	17		4		3	17	25	66
0031	13		10		1	11	24	59
0041	16		6		4	18	21	65
0051	20					20	18	58
0061	20		12		4		34	70
0071	21		3		2	5	56	87
0081	20		2		2	28	29	81
0091	24		4	1	1	20	10	60
0101	20				2	22	28	72
0111	21		2	1	4	13	20	61
0121	27		1	1	2		37	68
0131	16		4	2	3	2	30	57
0141	19		20	2	6		27	74
0151	25		16	2	4		22	69
0161	14		30	1	2	14	15	76
0171	17		25	1	4	1	21	69
0181	18		14	1	2		14	49
0191	17		7	1	3	4	41	73
0201	16		1	1	2	3	32	55
0211	16	10	4	1	4		25	60
0221	18		1		4	4	40	67
0231	16		8	1	1	7	16	49
0241	22		5		6	2	28	63
계	449	10	181	16	67	203	640	1,566

<표 5-5> 조사구별 매칭 유형 : 광주 백운1동

조사구	유형 1	유형 2	유형 3	유형 4	유형 5	유형 6	유형 7	계	조사구	유형 1	유형 2	유형 3	유형 4	유형 5	유형 6	유형 7	계
0011	17			10			2	29	035A		66						66
0021	10			15	2	4	1	32	036A		60					7	67
0031	9			22			5	36	037A		60						60
0041	7			9	1		3	20	038A		60						60
0051	10			22			3	35	0391	8			32	1		2	43
0061	14			17	1		1	33	040A		53						53
0071	15			25	1		8	49	041A		52						52
0081	28			15			8	51	042A		54						54
0091	39		3		1		11	54	043A		54						54
010A		60						60	044A		71						71
011A		60						60	045A		74						74
012A		60						60	0461	6	37		12			1	56
013A		60						60	0473					1			1
014A		60						60	0481	3			28			2	33
015A		59						59	049A		50						50
016A		68						68	050A		50						50
017A		73						73	0511	5			20				25
018A		72						72	0521	2			24			1	27
0191	18		2	11			7	38	0531	7			24				31
0201	15			28			8	51	0541	12			17				29
0211	8		2	17	1			28	0551	7			26				33
0221	5			26			1	32	0561	6		2	22	1			31
0231	3			27			2	32	0571	1			28			1	30
0241	10			12			2	24	0581	12			12	1		2	27
0251	14			22	1		3	40	0591	5			14	1	4		24
0261	12			36			4	52	0601	24							24
0271	4			21		3	3	31	0611	22			10				32
028A							75	75	0621	24			4	1		2	31
029A							72	72	063A		64						64
0301	16			17			2	35	064A		66						66
0311	8			34			2	44	065A		64						64
032A		54						54	066A		70						70
033A		56						56	067A		62						62
034A		54						54	계	406	1,805	9	659	14	11	241	3,143

4. 건축물대장 정보의 활용에 대한 평가

1) 건축물대장의 매칭률

조사구를 확정하기 위해서는 특정 지역 내에 소재하는 거처수의 파악이 필요하며, 이를 위해 건축물대장의 정보가 활용될 수 있다. 건축물대장 정보의 활용 가능성을 평가하기 위해 건축물대장 자료 가운데 매칭이 안 된 것에 대한 분석을 시도하였다.

먼저 자료의 매칭에서 건축물대장이 활용된 정도를 평가해 보면 <표 5-6>와 같다. 전체 건축물대장 자료 중 일부분만이 가구명부 및 거처조사결과와 매칭 되었다. 판암1동과 백운1동에서는 각각 건축물대장 자료 중 27.3%, 28.1%만이 매칭 되었다. 화도면에서는 매칭률이 다소 높았으나 36.3%에 불과했다.

<표 5-6> 지역별 건축물대장의 매칭률(수작업 매칭)

구분 \ 지역	대전 판암1동	강화군 화도면	광주 백운1동
매칭 된 자료수	430	806	1,975
매칭 안 된 자료수	1,142	1,415	5,065
전체 자료수	1,571	2,221	7,040
매칭률(%)	27.3	36.3	28.1

매칭률=(매칭 된 자료수/전체 자료수) x 100

건축물대장 자료 중 매칭 되지 않는 건축물이 많은 것은 주택용도가 아닌 건축물의 비율이 매우 높기 때문일 가능성이 있다. 비주택용도의 건축물이 많다면 가구 거처 자료와 매칭 되는 건축물이 적을 수밖에 없다. 하지만 매칭 되지 않은 건축물대장 자료들 중 주택용도의 건축물이 차지하는 비율을 살펴보면, <표 5-7>에서 나타나듯이 매칭 되지 않은 건축물대장에서 주택용도의 건축물이 차지하는 비중이

상당히 높다. 판암1동의 경우 주택용도의 건축물이 매칭 되지 않은 건축물대장 자료들 중에서 차지하는 비중이 82.7%에 달하고, 화도면에서도 60.8%에 달하며, 백운 1동에서는 88.1%에 달한다.

다음으로 가구명부와 거처정보가 행정동을 사용하는데 반해 건축물대장은 법정동을 사용하여 주소가 일치하기 않기 때문에 나타나는 결과일 가능성이 있다. 조사는 행정동으로 이루어졌다. 그렇지만 조사지역이 있는 법정동은 조사지역의 경계 밖에 있는 다른 행정동까지 포함할 수 있다. 행정자치부의 주민등록 전산주소 코드 현황에 따르면 조사지역인 행정동 판암1동에는 법정동 판암동과 삼정동에 분포되어 있다. 반면 법정동인 판암동은 판암1동과 판암2동 지역에 분포되어 있다(행정자치부, 2004b). 조사지역인 행정동 백운1동에는 법정동 서동과 백운동이 포함되어 있으며, 반면 법정동 백운동에는 행정동의 백운 1동과 2동 뿐만 아니라 주월2동에 걸쳐 있다. 이러한 경우 건축물대장 자료의 매칭률은 낮을 수밖에 없다.

그렇지만 행정동과 법정동의 불일치 문제가 나타나지 않는 화도면의 경우에도 매칭률이 높지 않다는 사실은 낮은 매칭률에 대한 다른 원인이 있음을 시사한다. 첫째, 화도면의 경우 대도시에 비해 건축물대장이 잘 정비되지 않았을 가능성이 있다. 이는 실제 조사된 거처들에 대한 건축물대장과의 매칭률이 상대적으로 매우 낮았다는 사실과 같은 맥락에서 이해할 수 있다. 둘째, 건축물대장에 등록되어 있는 주택 소재 거처들에 대한 조사가 불충분하게 진행되었을 가능성이 있다. 화도면이 대도시 행정구역에 비해 상대적으로 넓고 외딴 곳에 있는 거처 등 접근이 어려운 곳이 많아 실제 조사가 제대로 이루어지지 못했을 가능성이 있다. 이러한 문제는 강 화도 외에 많은 지방에서 발생할 수 있으며 특히 산간지방에서 심각할 수 있고, 본 조사에서도 발생할 수 있는 어려움이다. 이럴 경우 건축물대장의 자료를 유용한 증거 자료로서 사용할 수 있다. 건축물대장의 주용도 자체가 적지 않은 오차를 가질 수도 있지만 가구명부 작성에서 누락된 거처들을 파악하는데 도움을 줄 것이다.

<표 5-7> 매칭 되지 않은 건축물대장 자료 중 주택용도 건축물

지역 동주용도(주용도코드)		대전 판암1동		강화군 화도면		광주 판암1동	
		자료수	비율(%)	자료수	비율(%)	자료수	비율(%)
주택용도	단독주택(1000)	868	76.0	860	60.8	4,061	80.2
	다중주택(1002)	-	-	-	-	1	0.0
	공동주택(2000)	18	1.6	-	-	44	0.9
	아파트(2001)	25	2.2	-	-	-	-
	연립주택(2002)	2	0.2	-	-	-	-
	다세대주택(2003)	22	1.9	-	-	4	0.1
	생활편익시설(2004)	1	0.1	-	-	-	-
	기숙사(2007)	-	-	-	-	1	0.0
	계	936	82.7	860	60.8	4,111	88.1
비매칭 건축물대장 자료 총수		1,142		1,415		5,065	

비율은 매칭 되지 않은 건축물대장 자료의 총수를 기준으로 계산

2) 건축물대장의 건축물 주용도와 자료의 매칭유형

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 매칭 된 건축물대장 자료들이 세 가지로 구분한 건축물의 주용도에 따라 매칭 유형에 차이가 어떻게 나타나는지를 살펴보았다(<표 5-8>).

여기에서 특히 주목할 것은 기타시설 용도의 건축물대장 자료들이다. 건축물대장에 명기된 주용도를 바탕으로 거쳐 수를 추정할 경우 단독주택이나 공동주택 이외의 목적으로 용도가 분류된 건축물들은 거쳐가 아닌 건축물로 추정된다. 하지만 거쳐와 매칭 된 기타시설 용도의 건축물대장 자료들 중에서 일대일 매칭의 유형 1에 해당하는 경우가 상당히 많았다. 판암1동에서는 매칭 된 기타시설 용도의 건축물대장 자료의 73.8%가 일대일 형태로 거쳐 정보와 매칭 되었다. 화도면에서도 기타시설 용도의 건축물대장 자료 중 유형1에 속하는 비율이 50.0%나 되었다. 백운1

동에서는 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.

<표 5-8> 지역별 건축물 주용도에 따른 자료의 매칭유형

대전시 판암1동								
구분	단독주택		공동주택		기타시설		전체	
	빈도	비율(%)	빈도	비율(%)	빈도	비율(%)	빈도	비율(%)
유형1	226	65.9	-	-	31	73.8	257	59.9
유형2	-	-	39	88.6	-	-	39	9.1
유형3	20	5.8	1	2.3	-	-	22	4.9
유형4	3	0.9	-	-	5	11.9	8	1.9
유형5	24	7.0	-	-	3	7.1	27	6.3
유형6	70	20.4	4	9.1	3	7.1	77	17.9
계	343	100.0	44	100.0	42	100.0	429	100.0
강화군 화도면								
유형1	356	57.4	-	-	91	50.0	447	55.7
유형2	-	-	1	100.0	-	-	1	0.1
유형3	48	7.7	-	-	7	3.8	55	6.8
유형4	16	2.6	-	-	17	9.3	33	4.1
유형5	111	17.9	-	-	44	24.2	155	19.3
유형6	89	14.4	-	-	23	12.6	112	13.9
계	620	100.0	1	100.0	182	100.0	803	100.0
광주시 백운1동								
유형1	343	19.3	4	14.8	54	34.0	401	20.4
유형2	-	-	20	74.1	-	-	20	1.0
유형3	3	0.2	-	-	1	0.6	4	0.2
유형4	1,390	78.0	3	11.1	92	57.9	1,485	75.5
유형5	41	2.3	-	-	12	7.5	53	2.7
유형6	4	0.2	-	-	-	-	4	0.2
계	1,781	100.0	27	100.0	159	100.0	1,967	100.0

이러한 사실은 건축물대장의 주용도를 바탕으로 거쳐수를 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 거쳐와 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.

5. 매칭 오류의 사례 및 가능성

1) 매칭 오류의 사례

방대한 규모의 인구주택총조사에 건축물대장 자료를 사용하기 위해서는 데이터베이스를 이용하는 것이 필요하다. 앞에서는 한 지번에 다수의 거처나 다수의 건축물대장 자료들이 존재할 때 나타나는 매칭유형들을 살펴보았다. 하지만 데이터베이스를 이용하여 자료를 매칭 하는 실제의 과정에서는 이외에도 다양한 유형의 매칭 오류가 생길 수 있다. 본 연구에서 DB를 사용하지 않고 수작업으로 매칭을 한 결과를 데이터베이스를 사용한 매칭 결과와 비교한 결과 다음과 같은 매칭 오류의 사례들이 발견되었다(<표 5-9>).

<표 5-9> 동일 지번에 여러 주택유형이 동시에 존재할 때 발생하는 매칭 오류

가구명부								거처조사 자료			건축물대장			
(수정) 법정 읍면동 명칭	조사 구 번호	거 처 번 호	본 번	부 번	공 통 주 택 명 칭	동	호	거 처 구 분	연 건 평	대 지 면 적	동 일 련 번호	동 명 칭 및 번호	동 기 타 용 도	동 세 대 수
관암동	0021	38	293	1				1	9	270	2	A동	공동주택	6
관암동	0021	39	293	1	관암연립	A	102	4	18		2	A동	공동주택	6
관암동	0021	40	293	1	관암연립	A	101	4	18		2	A동	공동주택	6
관암동	0021	41	293	1	관암연립	A	202	4	18		2	A동	공동주택	6
관암동	0021	42	293	1	관암연립	A	203	4	18		2	A동	공동주택	6
관암동	0021	43	293	1	관암연립	A	103	4	18		2	A동	공동주택	6
관암동	0021	44	293	1	관암연립	A	201	4	18		2	A동	공동주택	6
관암동	0021	45	293	1	관암연립	B	101	4	18		2	A동	공동주택	6
관암동	0021	46	293	1	관암연립	B	102	4	18		2	A동	공동주택	6
관암동	0021	47	293	1	관암연립	B	201	4	18		2	A동	공동주택	6
관암동	0021	48	293	1	관암연립	B	202	4	18	25	2	A동	공동주택	6
관암동	0021	38	293	1				1	9	270	1	B동	공동주택	4
관암동	0021	39	293	1	관암연립	A	102	4	18		1	B동	공동주택	4
관암동	0021	40	293	1	관암연립	A	101	4	18		1	B동	공동주택	4
관암동	0021	41	293	1	관암연립	A	202	4	18		1	B동	공동주택	4
관암동	0021	42	293	1	관암연립	A	203	4	18		1	B동	공동주택	4
관암동	0021	43	293	1	관암연립	A	103	4	18		1	B동	공동주택	4
관암동	0021	44	293	1	관암연립	A	201	4	18		1	B동	공동주택	4
관암동	0021	45	293	1	관암연립	B	101	4	18		1	B동	공동주택	4
관암동	0021	46	293	1	관암연립	B	102	4	18		1	B동	공동주택	4
관암동	0021	47	293	1	관암연립	B	201	4	18		1	B동	공동주택	4
관암동	0021	48	293	1	관암연립	B	202	4	18	25	1	B동	공동주택	4

첫째 경우는 가구명부 및 거처 정보의 오차에서 발생하는 매칭 오류의 사례이다. 대전 동구 관암 1동의 가구명부 및 거처 자료 중에는 법정동, 조사구번호, 본번, 부번, 거처번호가 모두 동일한 두 개의 거처가 입력된 사례가 발견되었다. 두 개의 거처에 대한 조사 내용을 비교해 보면 한 거처는 연건평 45평 규모의 단독주택이었으며 다른 거처는 연건평 40평의 다세대주택이었는데, 빈집으로 조사되었다.

이러한 문제는 가구명부 작성 과정에 따른 단순한 오류로 추정되나, 빈집의 처리 방식에서 나타난 오류일 가능성도 있다.

둘째 경우는 공동주택에서 가구명부의 자료와 건축물대장의 정보가 일치하지 않는 경우이다. 광주 남구 백운1동에 있는 아파트의 경우 가구명부에는 101동부터 111동까지 있는 것으로 나타났으나 건축물대장의 동별 표제부는 101동부터 104동까지 밖에 없다. 이 경우 가구명부 작성의 오차일수도 있고, 건축물대장의 누락일수도 있다. 어느 경우라도 현재 지번을 매칭값으로 사용하는 데이터베이스의 매칭 방식에서는 문제로 발견되지 않는다.

셋째 경우는 한 지번에 단독주택과 공동주택이 동시에 있을 때 발생할 수 있는 오차이다. 다음의 표는 대전 판암1동의 DB매칭에서 나타난 매칭 오류의 경우이다. 이 경우에는 한 지번에 연립주택 형식의 공동주택과 단독주택이 동시에 있는데, 건축물대장 정보는 공동주택에 해당하는 것만 있다. DB매칭에서는 아래의 표에서 음영으로 칠한 칸에 나타나듯이 공동주택의 건축물대장 정보가 건축물대장이 없는 단독주택에도 매칭 되고 있다. 명백히 공동주택과 별개인 단독주택에 공동주택의 건축물정보가 동시에 매칭 됨으로써 실제 건축물정보가 없어서 매칭이 되지 않는 거처가 매칭이 되는 것으로 구분되었다. 이 경우 중복을 제거하고 공동주택의 동명칭에 따라 건축물대장 정보를 매칭 하면 건축물대장 정보가 없는 단독주택을 파악할 가능성이 생긴다.

2) 법정동코드 사용의 문제점

다음으로는 가구명부-거처자료와 건축물대장을 매칭할 때 법정동코드를 매칭값으로 사용할 때 발생할 수 있는 문제점을 살펴본다. 인구주택총조사에서 조사원은 조사표에 조사대상 가구의 주소를 기입할 때 법정동을 사용한다. 해당 가구의 법정동은 조사구 지도에서 확인하거나 가구원에게 직접 물어 봄으로써 확인할 수 있다. 지금까지 조사가구주소의 법정동은 전산자료로 입력되지 않았다. 하지만 건축물대장 자료를 사용할 때 법정동 주소의 중요성은 커진다. 건축물대장 자료에는 법정동을 사용하여 주소가 입력되어 있기 때문에 건축물대장의 정보를 조사 대상 주택에 적용하기 위해서는 정확한 법정동 주소가 필요하다.

조사구 지도나 가구원에 대한 질문을 통해 법정동을 파악하는 것이 대부분

큰 문제가 되지 않지만 일부 지역에서는 문제가 발생할 수 있다. 특히 법정동 사이의 경계 지역에서는 이로 인한 혼란이 발생할 수 있다. 법정동의 지역 구분이 항상 뚜렷한 경계선을 가지는 것이 아니기 때문에 지도상에 위치시킬 때도 항상 명확한 것은 아니다. 또한 조사대상 가구원 역시 항상 자신의 법정동 주소를 정확히 알고 있다고 전제할 수 없다. 많은 일반인들은 행정동과 법정동이 구별해서 존재한다는 사실조차 모른다. 오히려 행정동은 해당 동사무소 혹은 주민자치센터의 이용을 위해 기억할 수 있는 반면 법정동이 낯설 수 있다.

다음의 <표 5-10>은 조사대상 지역 중 판암1동과 백운1동에서 두 법정동 사이에 같은 지번이 있는 경우를 제시한 것이다. 중복수와 가능매칭수는 법정동을 고려 않고 지번만을 기준으로 자료를 매칭할 때, 양쪽 법정동에 같은 지번이 있을 경우 발생할 수 있는 매칭의 횟수를 의미한다. 현재 유형은 법정동을 거쳐 구분의 기준으로 고려한 결과 구분된 해당 지번의 현재 매칭 유형이다.

판암1동의 경우 거처의 지번이 삼정동과 판암동 사이에서 일치하는 경우가 없었다. 반면 백운1동에서는 4개의 지번이 서동과 백운동에 동일한 값을 가지고 있었다. 지번 값이 같은 건축물대장이 양쪽 법정동에 동일하게 존재하는 경우는 더욱 많았다. 판암1동의 경우 판암동과 삼정동 사이에서 동일한 지번값을 가지는 건축물대장의 경우가 10회였고, 백운1동에서는 서동과 백운동 사이에서 지번은 같지만 법정동이 다른 경우가 77회나 되었다.

다음으로 거처와 건축물대장의 지번이 일치하지만 법정동이 서로 달라서 매칭이 되지 않는 경우를 살펴보았다. 예컨대 판암동에 속한 거처가 삼정동에 같은 지번값을 가지는 건축물대장을 가지는 경우가 42회나 되었다. 그렇다면 이러한 경우가 실제 무엇을 의미하는지는 정확치 않다. 실제 이 경우 거처와 건축물대장이 지번값이 같더라도 서로 다른 건축물에 해당할 수 있다. 그러나 이 42개 지번에 있는 거처들의 매칭 유형을 보면 해당 건축물대장이 없는 경우인 유형7에 해당되는 지번의 개수가 35개나 되었다. 이 35개 지번들은 법정동을 고려하지 않고 건축물대장을 매칭하면 매칭 될 수 있으나 법정동 구분 때문에 매칭 되지 않는 경우들이다. 이러한 사실은 가구명부에서 거처의 법정동과 건축물대장의 법정동이 모두 옳게 파악되고 기재되어 있는지를 검토할 필요가 있음을 보여 준다. 실제 판암1동에서는 같은 건축물의 거처와 건축물대장이 속한 법정동이 다르게 되어 있어 매칭이 안 된 경우가 발견되었다5).

물론 이러한 문제 사례들이 일반적이라고 말할 수는 없다. 즉 인접지역에서 건축물의 지번값이 같다면 법정동이 다르더라도 같은 필지에 속한다고 일반화시킬 수 없다. 백운1동의 자료에서는 동일한 지번값을 가지는 건축물대장이 다른 법정동에 있는 거처들 중 유형7에 해당하는 경우가 극히 소수였다. 결국 실제 조사를 통해 실상을 파악할 수 밖에 없다.

이와 관련하여 지적할 수 있는 근본적인 문제는 구 주소 체계에 체계적인 주소 부여 방식이 존재하지 않는다는 사실이다. 지번의 절대값이 가지는 의미가 없기 때문에 지번의 유용성이 매우 낮다. 주소를 체계적이고 규칙적으로 부여하였다면 여기에서 제시한 문제의 경우 현지 답사를 하지 않더라도 일정한 규칙성을 바탕으로 문제해결 방안을 모색할 수 있었을 것이다. 하지만 현재는 주소를 부여할 때 기준으로 삼은 것이 무엇이었으며 계속 유지되어 왔는지를 알 수 없기 때문에 지번값이 갖는 있는 정보의 유용성이 매우 제한되고 더욱 정교한 방식으로 주소의 값들을 활용하는데 도움이 되지 않는다.

5) 이 내용에 대해서는 다음 절의 분석 내용을 참조.

<표 5-10> 대전 동구 판암1동에서 판암동과 삼정동 사이의 중복 지번

일련번호	거처 판암동 삼정동 중복지번		건축물대장 판암동 삼정동 중복지번		판암동 거처 중 삼정동 건축물대장과 일치 지번			삼정동 거처 중 판암동 건축물대장과 일치지번		
	지번	중복수	지번	중복수	지번	가능매칭수	현재유형	지번	가능매칭수	현재유형
1			5-0	2	25-3	1	유형7	1-0	6	유형7
2			29-0	2	33-0	1	유형7	1-6	1	유형7
3			54-1	1	36-1	1	유형7	1-12	1	유형7
4			60-1	7	60-1	14	유형3	2-0	1	유형7
5			95-0	1	60-5	1	유형7	2-1	1	유형7
6			200-0	1	60-11	1	유형7	2-2	1	유형7
7			298-0	2	68-0	1	유형7	54-1	2	유형3
8			322-1	1	69-1	1	유형7	189-14	22	유형7
9			327-0	1	86-9	1	유형7	214-0	2	유형7
10			352-0	1	92-0	1	유형7			
11					95-0	1	유형1			
12					97-1	1	유형7			
13					97-2	1	유형7			
14					97-4	1	유형7			
15					98-1	38	유형7			
16					98-61	1	유형7			
17					153-1	1	유형7			
18					154-0	6	유형7			
19					154-1	9	유형7			
20					159-0	1	유형7			
21					166-0	1	유형7			
22					200-0	1	유형1			
23					201-2	1	유형7			
24					269-0	6	유형7			
25					270-2	1	유형7			
26					271-0	1	유형7			
27					272-0	1	유형7			
28					273-2	1	유형7			
29					298-0	2	유형1			
30					299-0	1	유형7			
31					319-0	1	유형7			
32					321-0	1	유형7			
33					322-1	1	유형1			
34					327-0	19	유형2			
35					335-0	1	유형7			
36					342-2	1	유형7			
37					342-3	1	유형7			
38					346-3	1	유형7			
39					346-4	1	유형7			
40					359-0	2	유형7			
41					365-0	1	유형7			
42					423-2	1	유형7			
합계	0	0	10	19	42	130		8	37	

<표 5-11> 광주 남구 백운1동에서 서동과 백운동 사이의 중복 지번

일련번호	거처 판암동 삼정동 중복지번		건축물대장 판암동 삼정동 중복지번		판암동 거처 중 삼정동 건축물대장과 일치 지번			삼정동 거처 중 판암동 건축물대장과 일치지번		
	지번	중복수	지번	중복수	지번	가능 매칭수	현재 유형	지번	가능 매칭수	현재 유형
1	72-3	1	1-7	2	72-3	2	유형3	1-7	1	유형4
2	72-6	1	3-3	2	72-6	4	유형3	4-0	1	유형4
3	74-6	1	4-0	2	74-8	2	유형3	4-3	7	유형4
4	74-8	1	4-3	14				6-3	1	유형1
5			4-5	1				6-4	1	유형4
6			6-3	1				6-9	1	유형4
7			6-4	2				9-0	1	유형4
8			6-9	2				9-4	3	유형4
9			9-0	1				9-6	2	유형1
10			9-4	6				9-7	2	유형4
11			9-6	2				9-8	1	유형4
12			9-7	4				9-9	1	유형4
13			9-8	2				9-10	1	유형4
14			9-9	6				9-11	1	유형4
15			9-10	2				13-2	1	유형4
16			9-11	2				17-0	1	유형4
17			13-2	2				17-2	1	유형4
18			17-0	2				19-2	1	유형4
19			17-2	2				20-07	1	유형1
20			19-2	2				20-8	1	유형1
21			20-2	2				20-9	1	유형1
22			20-7	1				28-1	2	유형1
23			20-8	1				28-2	2	유형1
24			20-9	1				28-3	2	유형7
25			28-1	2				29-5	1	유형4
26			28-2	2				29-6	1	유형4
27			29-3	2				29-7	2	유형1
28			29-5	2				29-9	2	유형1
29			29-6	2				29-12	1	유형4
30			29-7	2				29-13	1	유형4
31			29-9	2				29-15	1	유형7
32			29-12	2				32-1	147	유형7
33			29-13	2				33-3	2	유형7
34			33-1	2				33-4	2	유형1
35			33-4	2				33-6	1	유형1
36			33-6	1				33-7	1	유형4
37			33-7	2				33-8	2	유형1
38			33-8	2				33-10	1	유형4
39			33-10	2				33-11	1	유형1
40			33-11	1				35-3	1	유형4
41			35-3	2				35-4	2	유형4
42			35-4	6				35-8	2	유형1
43			35-8	2				36-3	2	유형4
44			36-3	4				37-0	1	유형1
45			37-0	3				39-0	1	유형4
46			39-4	2				39-4	1	유형4
47			46-1	5				46-1	253	건축물

일련번호	거처 관암동 삼정동 중복지번		건축물대장 관암동 삼정동 중복지번		관암동 거처 중 삼정동 건축물대장과 일치 지번			삼정동 거처 중 관암동 건축물대장과 일치지번		
	지번	중복수	지번	중복수	지번	가능 매칭수	현재 유형	지번	가능 매칭수	현재 유형
48			55-6	1				55-6	1	유형1
49			59-8	2				59-8	2	유형1
50			61-3	1				61-3	1	유형1
51			62-2	2				62-2	2	유형1
52			62-5	1				62-5	1	유형1
53			62-6	2				62-6	2	유형1
54			63-3	2				63-3	2	유형1
55			64-4	1				64-4	1	유형1
56			65-10	1				65-10	1	유형1
57			65-11	3				65-11	3	유형1
58			65-13	1				65-13	1	유형1
59			65-15	1				65-15	1	유형1
60			66-3	1				66-3	1	유형1
61			66-7	2				66-7	2	유형1
62			66-10	1				66-10	1	유형1
63			68-2	1				68-2	1	유형1
64			68-3	2				68-3	2	유형1
65			68-4	1				68-4	1	유형1
66			72-3	2				72-3	1	유형6
67			72-6	10				72-6	1	유형6
68			74-6	8				74-6	4	유형6
69			74-7	2				74-7	1	유형4
70			74-8	4				74-8	2	유형6
71			88-3	2				88-3	1	유형4
72			197-2	2				197-2	1	유형4
73			204-3	3						
74			206-1	2						
75			209-7	4						
76			236-0	1						
77			241-2	2						
횟수	4	4	77	186	3	8		73	503	

6. 소결 및 정책대안

이 장에서는 가구명부 및 거처의 정보와 건축물대장 정보를 매칭 시킬 때 발생하는 유형들과 지역별 차이를 살펴보았다. 분석 결과 두 정보를 매칭 할 때 나타나는 경우를 7가지 유형으로 구분하였으며, 매칭 유형에 따라 건축물대장 정보를 인구주택총조사에서 활용할 수 있는 가능성이 다르다. 따라서 건축물대장 정보를 효과적으로 활용하기 위해서는 각 매칭 유형별로 활용 전략을 구분하여 개선책을 마련하는 것이 필요하다.

인구주택총조사의 실행 전략과 관련해서 특히 주목할 만한 사실은 다양한 매칭 유형들의 분포가 지역별로 상이하다는 점이다. 이는 인구주택총조사에서 건축물대장 정보를 활용할 때 지역별로 주로 적용되는 활용방식이 달라져야 한다는 것을 시사한다. 그러므로 건축물대장 정보를 인구주택총조사에서 사용한다면 건축물대장 정보의 활용에 대해 조사원들에게 교육을 할 때 지역별로 차별화된 교육 프로그램을 마련하는 것이 효과적일 것이다.

이 연구에서는 가구명부 및 거처 자료와 건축물대장의 자료를 직접 매칭 하는 작업을 수행하는 과정에서 여러 유형의 매칭 오류들을 발견할 수 있었다. 중요한 오류 사례들로는 가구명부 및 거처 정보의 오차에서 발생하는 매칭 오류가 있었고, 가구명부의 자료와 건축물대장의 정보가 일치하지 않는 경우가 있었으며, 한 지번에 공동주택과 건축물대장이 없는 단독주택이 동시에 존재할 때 발생할 수 있는 오류가 있었다. 또 다른 오류의 가능성으로는 법정동이 잘못 조사되거나 건축물대장 정보에 법정동 주소가 잘못 기재되었을 경우를 지적할 수 있다. 이러한 문제들을 해결하기 위해서는 공동주택의 동명칭에 따라 건축물대장 정보를 매칭시킬 수 있는 정교한 매칭 방법의 개발이 필요하고 체계성과 규칙성을 갖춘 주소체계가 요구된다.

인접한 법정동 사이에서 중복지번이 적지 않다. 조사 대상자들이 자신의 지역이 속해 있는 법정동을 알고 있다고 전제하지만 실제 이러한 전제가 얼마나 타당한지 검증할 필요가 있다. 건축물대장에 수록되어 있는 정보를 적합하게 사용하기 위해서는 건축물들에 대한 많은 정보의 축적이 필요하다. 이를 위해서는 우리나라 건축물들이 실제 어떠한 구조로 위치해 있으며 구주소 체계를 효과적으로 활용할 수 있는 방안은 무엇인지에 대한 연구가 선행되어야 한다.

끝으로 이 연구에서 나타난 여러 유형의 오류들을 볼 때 조사원들이 인구주택총조사의 실제 과정에서 많은 다양한 문제들을 접하게 될 것으로 예상할 수 있다. 이러한 문제들의 발견은 앞으로 센서스 방법을 개선하고 조사의 효율성과 자료의 신뢰성을 제고하는 데 크게 기여할 수 있다. 이를 위해 센서스 기간 동안 투입되는 조사원들이 다양한 경험들을 보고할 수 있도록 하고, 이 내용을 DB화 시키는 방안을 제안한다. 이러한 DB의 구축을 통해 실제 활용에 더욱 도움이 되는 예비 질문 및 응답(FAQ)을 발굴할 수 있을 것이다.

VI. 조사구별 매칭률 심층 분석 - 대전시 판암 1동

1. 개요

판암 1동의 조사구별 매칭률은 두 가지 방식으로 조사되었다. 첫 번째 방식은 일종의 조매칭률(crude matching rate)을 구하는 것이다(<표 6-1>). 이 방식은 거처정보와 건축물대장 자료의 중복매칭을 허용한다. 즉 조매칭률 계산은 거처와 건물이 일대일로 매칭 되는 경우뿐만 아니라 한 거처에 다수의 건물이 매칭 되는 일대다로 매칭 되는 경우도 포함한다. 따라서 조매칭률의 최대값은 한 거처에 얼마나 많은 건축물이 중복되어 매칭 되는가에 따라 달라진다.

두 번째 방식의 매칭은 중복을 허용하지 않는 순매칭률을 구하는 것이다. 순매칭률의 최대값은 100%이다. 중복된 사례를 삭제하기 위해서는 database 수준에서 기계적으로 중복된 레코드를 삭제하는 방법이 있다. 그리고 연구자가 거처정보와 건축물대장의 법정동, 본번, 부번을 고려하여 직접 매칭 하는 방법도 있다.

여기에서는 연구자가 직접 수작업으로 매칭 하는 방법을 선택하였다. 무엇보다도 기계식 매칭 방법에서 매칭 오류가 다수 발생하였다. 아울러 상대적으로 기계식 매칭이 연구자의 수작업에 비해 매칭률이 낮다는 점을 고려하였다(92.6% 대 94.4%).

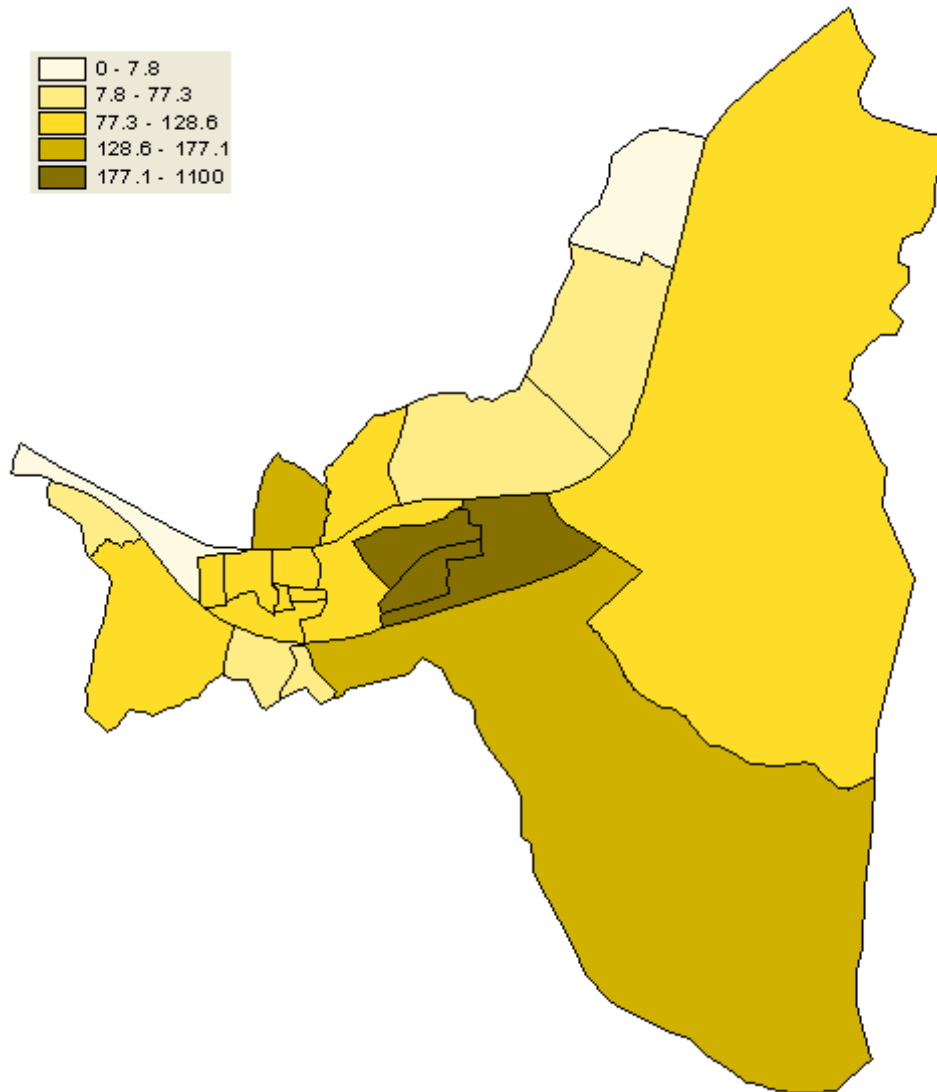
조매칭률을 조사구별로 살펴보면, 공동주택 비율이 높은 지역에서는 건축물대장의 특성상 일대다 중복 매칭으로 인해 매칭비율이 높다. 건축물대장에 기록되어 있는 건물의 수에 따라, 한 거처에 약 8개 혹은 11개 정도의 건축물 정보가 중복되어 매칭 된다. 판암1동의 경우 조사구 9-36이 여기에 해당된다. 반면에 특수한 시설인 정신병원이 위치한 조사구 5와 공장과 교회가 위치한 조사구 71에서는 거처가 건축물대장의 건물정보와 전혀 매칭이 되고 있지 않다. 그 밖에 특별한 매칭률을 보이는 조사구로는 조사구 6이 있다. 이 조사구의 경우는 다른 조사구에 비해 현격하게 낮은 매칭률(7.8%)을 보이고 있다. 이러한 조사구에 대해서는 좀 더 심층적인 자료 분석이 필요하다. <그림 6-1>은 이러한 조매칭률을 그림으로 나타낸 것으로, 색깔이 진한 지역일수록 매칭률이 높다.

<표 6-1> 판암 1동 조매칭률 현황 (중복 포함)

조사구	거처수	매칭 건축물 수	매칭률(%)	조사구	거처수	매칭 건축물 수	매칭률(%)
0011	42	73	173.8	038A	60	60	100.0
0021	58	57	98.3	039A	70	70	100.0
0031	43	23	53.5	040A	70	70	100.0
0041	51	24	47.1	041A	70	70	100.0
0054 [*]	2	0	0.0	042A	60	60	100.0
0061	51	4	7.8	043A	60	60	100.0
0071	52	57	109.6	044A	60	60	100.0
0081	48	85	177.1	045A	60	60	100.0
0091	65	480	738.5	0461	51	46	90.2
010A	60	480	800.0	0471	28	20	71.4
011A	60	480	800.0	0481	44	34	77.3
012A	60	480	800.0	0491	61	53	86.9
013A	60	480	800.0	0501	54	54	100.0
014A	60	480	800.0	051A	66	66	100.0
015A	58	480	827.6	052A	67	67	100.0
016A	60	480	800.0	053A	60	60	100.0
017A	60	480	800.0	054A	60	60	100.0
018A	60	480	800.0	055A	60	60	100.0
019A	60	480	800.0	056A	54	54	100.0
020A	60	480	800.0	057A	60	60	100.0
021A	60	480	800.0	058A	60	60	100.0
022A	60	480	800.0	059A	60	60	100.0
023A	58	464	800.0	060A	60	60	100.0
024A	60	480	800.0	061A	54	54	100.0
025A	60	480	800.0	062A	60	60	100.0
026A	60	480	800.0	063A	60	60	100.0
027A	50	550	1100.0	064A	60	60	100.0
028A	60	660	1100.0	065A	60	60	100.0
029A	40	440	1100.0	0661	50	45	90.0
030A	50	550	1100.0	067A	68	68	100.0
031A	50	550	1100.0	068A	68	68	100.0
032A	50	550	1100.0	069A	68	68	100.0
033A	70	770	1100.0	070A	72	68	94.4
034A	70	770	1100.0	0711	5	0	0.0
035A	60	660	1100.0	0721	56	72	128.6
036A	60	370	616.7	0731	52	34	65.4
037A	60	60	100.0	계	4,136	16,920	409.1

주: 조매칭률 = (매칭 건축물 수 / 거처수) × 100

조사구 5 : 대전정신요양원, 동대전정신병원 / 조사구 71 : 영진조극사, 창성교회



<그림 6-1> 판암1동 조매칭률

중복을 배제한 순매칭률도 대규모 공동주택이 많은 조사구에서 매칭률이 과장된 부분을 제외하면, 대부분은 조매칭률과 유사한 유형을 보여준다(<표 6-2>).⁶⁾ 단독주택이 상대적으로 많은 지역이나 현저하게 매칭률이 떨어지는 조사구는 매칭률에 큰 차이를 보이고 있지 않는다. 다만, 단독주택과 공동주택이 혼합되어 있는 조사구나 주택과 부속건물이 한 지번에 존재하는 경우가 많은 조사구의 경우는 조매칭률과 순매칭률에 많은 차이를 보이고 있다. <그림 6-2>는 순매칭률을 그림으로 나타낸 것이다. 역시 색깔이 진한 지역일수록 매칭률이 높다.

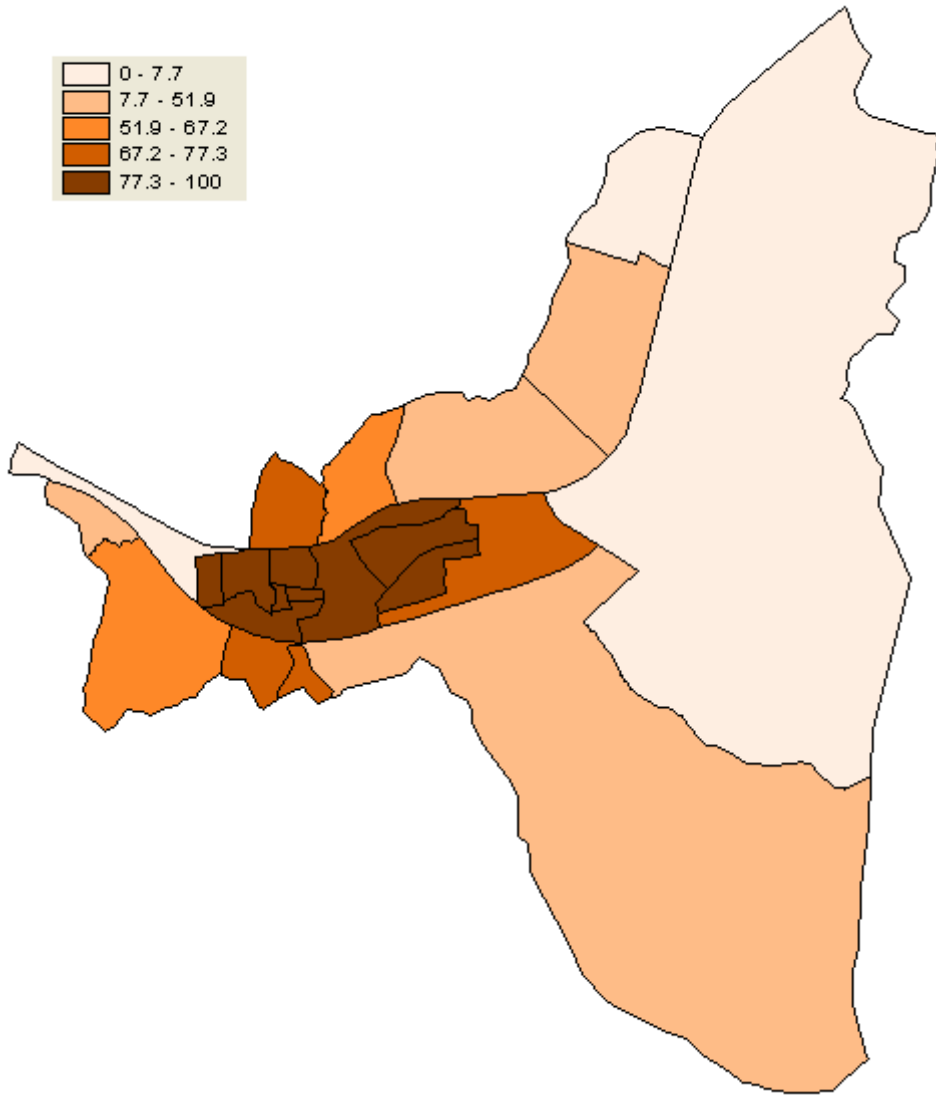
6) 예를 들면, 대규모 공동주택이 많은 조사구의 경우, 자료를 연구자가 직접 대조매칭 작업을 하면 매칭률은 100%로 조정된다.

<표 6-2> 판암 1동 순매칭률 현황 (직접매칭, 중복배제)

조사구	거처수	매칭 건축물수	매칭률(%)	조사구	거처수	매칭 건축물수	매칭률(%)
1	42	32	76.2	38	60	60	100.0
2	58	39	67.2	39	70	70	100.0
3	43	20	46.5	40	70	70	100.0
4	51	24	47.1	41	70	70	100.0
5	2	0	0.0	42	60	60	100.0
6	51	2	3.9	43	60	60	100.0
7	52	4	7.7	44	60	60	100.0
8	48	21	43.8	45	60	60	100.0
9	65	50	76.9	46	51	46	90.2
10	60	60	100.0	47	28	20	71.4
11	60	60	100.0	48	44	34	77.3
12	60	60	100.0	49	61	53	86.9
13	60	60	100.0	50	54	54	100.0
14	60	60	100.0	51	66	66	100.0
15	58	58	100.0	52	67	67	100.0
16	60	60	100.0	53	60	60	100.0
17	60	60	100.0	54	60	60	100.0
18	60	60	100.0	55	60	60	100.0
19	60	60	100.0	56	54	54	100.0
20	60	60	100.0	57	60	60	100.0
21	60	60	100.0	58	60	60	100.0
22	60	60	100.0	59	60	60	100.0
23	58	58	100.0	60	60	60	100.0
24	60	60	100.0	61	54	54	100.0
25	60	60	100.0	62	60	60	100.0
26	60	60	100.0	63	60	60	100.0
27	50	50	100.0	64	60	60	100.0
28	60	60	100.0	65	60	60	100.0
29	40	40	100.0	66	50	45	90.0
30	50	50	100.0	67	68	68	100.0
31	50	50	100.0	68	68	68	100.0
32	50	50	100.0	69	68	68	100.0
33	70	70	100.0	70	72	72	100.0
34	70	70	100.0	71	5	0	0.0
35	60	60	100.0	72	56	34	60.7
36	60	60	100.0	73	52	27	51.9
37	60	60	100.0	계	4,136	3,828	92.6

주 : 순매칭률 = (순매칭 건축물 수 / 거처수) × 100

조사구 5 : 대전정신요양원, 동대전정신병원 / 조사구 71 : 영진조극사, 창성교회



<그림 6-2> 판암1동 순매칭률

2. 주요 조사구별 매칭 현황 분석

조사구별 매칭 현황을 볼 때, 순매칭률이 100%에 이르는 공동 주택의 경우는 큰 문제가 없는 것으로 파악된다. 그러나 다른 조사구에 비해 현저하게 매칭률이 떨어지는 경우는 그 이유를 파악하는 것이 중요하다. 판암 1동에서는 총 73개의 조사구 중에 이러한 목적을 위해, 5개의 조사구에 대해 심층적인 분석을 진행하였다. 여기에는 평균적인 조사구(조사구 1와 조사구 47)⁷⁾, 매칭률이 낮은 조사구(조사구

7) 조사구 1의 경우 순매칭률이 76.2이고, 조사구 47의 순매칭률은 71.4이다.

6), 공동 주택으로 이루어진 조사구(조사구 29), 매칭이 전혀 되지 않은 조사구(조사구 71)가 포함되어 있다.

먼저 조사구 1(0011)의 현황을 살펴보면, 매칭이 안되는 대부분 경우에 건축물대장에서 해당 지번에 대한 정보가 누락되어 있다. 이 경우 거처정보의 정보가 문제가 있는 것으로 추측할 수 있다. 혹은 건축물대장에서 기록이 누락되었거나 잘못 기재되어 있기 때문일 수도 있다. 여기에서는 이러한 경우에서 공통적으로 발견되는 점을 중심으로 그 이유를 생각해 보고자 한다. 첫째, 많은 경우에 1950년 이전에 사용허가가 난 건물이라는 점에서 건축물대장 정보의 정확성과 관련이 깊은 것으로 추측된다. 둘째, 상가 건물에 거처가 존재할 때, 매칭에 문제가 발생하였다. 셋째, 부속건물이 존재할 때, 복수의 건축물이 하나의 거처와 매칭 되었다.

<표 6-3> 판암1동 조사구 1의 정보 매칭 요약

	매칭정보	비매칭정보		계
		거처정보	건축물대장	
거처수	32*	5	5	42
%	75.2	11.9	11.9	100.0

* 일대일 매칭 거처수는 22, 일대다 매칭 거처수는 10개. 지도를 근거로 일대다 매칭을 보면, 건축물대장 문제로 7건, 거처정보의 문제 3건 추정

<표 6-4> 판암1동 조사구 1의 매칭 상세정보

거처 번호	매칭 건물수	주소	거 처 명 부	건 축 물 대 장	지도	비고
1 ↓ 8	1					
9	2	310-12	연건평 33, 대지 69, 방 4, 거실 1, 식당 2, 부엌 2, 화장실 1, 출입구 1	주택, 목조, 세와, 45.62, 1926 주택, 흙벽돌,스레트,48.46,1985	건물 1 310-12	건축물 대장
10	9	488-0	빈집 1	주택, 목조, 스테트 26.44, 1940 주택, 세와, 51.41, 1951 주택, 와즙, 191.76, 1959 주택, 스테트, 26.44, 1945 주택, 스테트, 26.45, 1942 주택, 세와, 35.24, 1953 주택, 세와, 107.48, 1936 주택, 스테트, 26.45, 1919 주택, 스테트, 26.44, 1940	건물 5 웨이리 치킨 2	건축물 대장
11	9	488-0	빈집 1	상동		상동
12	9	488-0	빈집 1	상동		상동

주 : 빈집=매칭 된 거처, 건축물대장정보=지번(다를 경우), 구조명, 지붕구조명, 연면적, 사용승인일자

거처 번호	매칭 건물수	주소	거 처 명 부	건 축 물 대 장	지도	비고
13	0	308-0	연건평 21, 대지 320, 방수 3, 거실 1, 식당 0, 건축연도 2, 부엌 1, 화장실 2, 출입구 1	해당 지번에 건축물대장 없음		건물 부재
14	0	307-4	연건평 20, 대지 45, 방수 3, 거실 1, 식당 1, 건축연도 12, 부엌 1, 화장실 1, 출입구 1	해당 지번에 건축물대장 없음		건물 부재
15	0	308-21	연건평 31, 대지 59, 방수 3, 거실 1, 식당 0, 건축연도 2, 부엌 1, 화장실 2, 출입구 1	해당 지번에 건축물대장 없음		건물 부재
16	0	308-1	연건평 30, 대지 80, 방수 5, 거실 1, 식당 1, 건축연도 10, 부엌 2, 화장실 2, 출입구 1	해당 지번에 건축물대장 없음		건물 부재
17	0	308-4	연건평 18, 대지 72, 방수 3, 거실 1, 식당 2, 건축연도 9, 부엌 2, 화장실 1, 출입구 1	해당 지번에 건축물대장 없음		건물 부재
18 ↓ 22	1					
23	2	487-0	연건평 40, 대지 200, 방 3, 거실 1, 식당 1, 건축연도 12, 부엌 1, 화장실 1, 출입구 3	창고주택, 목조, 도단, 181.8,1901 목조, 스테트, 26.44, 1901	건물 3 487-0	건축물 대장노 후건물
24	1					
25	2	487-0	연건평 15, 대지 70, 방 1, 거실 0, 식당 1, 건축연도 12, 부엌 1, 화장실 1, 출입구 2	창고주택, 목조, 도단, 181.8,1901 목조, 스테트, 26.44, 1901	건물 3 487-0	건축물 대장노 후건물
26	9	488-0	연건평 37, 대지 100, 방 6, 거실 2, 식당 2, 건축연도 12, 부엌 2, 화장실 1, 출입구 1	주택, 목조, 스테트 26.44, 1940 주택, 세와, 51.41, 1951 주택, 외죽, 191.76, 1959 주택, 스테트, 26.44, 1945 주택, 스테트, 26.45, 1942 주택, 세와, 35.24, 1953 주택, 세와, 107.48, 1936 주택, 스테트, 26.45, 1919 주택, 스테트, 26.44, 1940	건물 5 훼미리 치킨 2	건축물 대장노 후건물
27	0	468-10	연건평 48, 대지 136, 방 6, 거실 2, 식당 2, 건축연도7, 부엌 2, 화장실 2, 출입구 4	해당 지번에 건축물대장 없음	건물 1 486-10	거처 정보 사랑의 약국
28 ↓ 33	1					
34	0	488-3	연건평 20, 대지 40, 방 3, 거실 1, 식당 2, 건축연도 5, 부엌 2, 화장실1,출입구1	해당 지번에 건축물대장 없음	건물 2 488-0	거처 정보 판암유리
35	0	492-3	연건평 23, 대지 0, 방 2, 거실 1, 식당 0, 건축연도 6, 부엌 2, 화장실1,출입구1	해당 지번에 건축물대장 없음	건물 2 492-0	거처 정보 보람주유소
36	0	518-0	빈집 1	해당 지번에 건축물대장 없음	건물 2 518-2	거처 정보

주 : 빈줄=매칭 된 거처, 건축물대장정보=지번, 구조명, 지붕구조명, 연면적, 사용승인일자

거처 번호	매칭 건물수	주소	거 처 명 부	건 축 물 대 장	지도	비고
37	3	519-0	연건평 15, 대지 15, 방 2, 거실 1, 식당 1, 건축연도 9, 부엌 1, 화장실1,출입구1	주택, 목조, 스테트 36.66, 1985 주택, 흙벽돌, 스테트 45.36,1985 주택, 흙벽돌, 스테트 33.2, 1985	건물 3 519-1	거처 정보
38	0	518-0	연건평 50, 대지 320, 방 4, 거실 1, 식당 0, 건축연도 4, 부엌 1, 화장실1,출입구2	해당 지번에 건축물대장 없음	건물 2 518-2	거처 정보
39	3	519-0	연건평 28, 대지 60, 방 5, 거실 1, 식당 0, 건축연도 11, 부엌 1, 화장실1,출입구1	주택, 세부, 스테트 36.66, 1985 주택, 흙벽돌, 스테트 45.36,1985 주택, 흙벽돌, 스테트 33.2, 1985	건물 3 519-1	거처 정보
40	3	519-0	연건평 12, 대지 26, 방 2, 거실 1, 식당 0, 건축연도 10, 부엌 1, 화장실1,출입구1	주택, 세부, 스테트 36.66, 1985 주택, 흙벽돌, 스테트 45.36,1985 주택, 흙벽돌, 스테트 33.2, 1985	건물 3 519-1	거처 정보
41	1					
42	1					

주 : 빈줄=매칭 된 거처, 건축물대장정보=지번, 구조명, 지붕구조명, 연면적, 사용승인일자

조사구 6(0061)의 경우는 매칭률이 10%미만으로 다른 조사구에 비해 현저하게 낮다(<표 6-5>). 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. 예를 들면, 거처정보에는 98-27번지의 한솔빌라라고 되어 있지만 건축물대장에는 98-27번지에 대한 건축물 정보가 없다. 대신에 97-28번지의 19가구로 이루어진 다가구주택이 존재하는 것으로 보아서 거처정보에 문제가 있음을 알 수 있다. 물론 이것은 조사자의 오류일 수도 있고, 입력자의 오류일 수도 있다(<표 6-6>).

<표 6-5> 판암1동 조사구 6의 정보 매칭 요약

	매칭정보	비매칭정보			계
		거처정보	건축물대장	매칭오류	
거처수	2	24	0	25	51
%	3.9	47.1	0.0	49.1	100.0

주 : 매칭오류는 거처정보와 건축물대장에 지번 존재하나 매칭 파일에 존재하지 않은 거처

<표 6-6> 판암1동 조사구 6의 매칭 상세정보

거처 번호	매칭 건물수	주소	거 처 명 부	건 축 물 대 장	지도	비고 오류
1	0	423-2	연건평 0, 대지 0, 방 3, 거실 1, 식당 1, 건축연도 6, 부엌 1, 화장실 1, 출입구 1	해당 지번에 건축물 없음	건물부재 423-5 양지포도 423-6	거처정보
2	0	86-4	연건평 26, 대지 100, 방 4, 거실 1, 식당 0, 건축연도 7, 부엌 1, 화장실 2, 출입구 1	해당 지번에 건축물 없음	건물부재	거처정보
3	0	86-9	연건평 38, 대지 100, 방 6, 거실 2, 식당 1, 건축연도 7, 부엌 2, 화장실 2, 출입구 1	주택 조사	86-9 양지포도 원	매칭오류
4	0	112-2	연건평 60, 대지 80, 방 4, 거실 1, 식당 1, 건축연도 9, 부엌 2, 화장실 2, 출입구 1	해당 지번에 건축물 없음	112-0 목조, 스레트, 1941	거처정보
5	1	60-1	-	해당 주소에 건축물대장 8개 건물 존재		매칭오류
6	0	60-0	연건평 33, 대지 60, 방 5, 거실 1, 식당 2, 건축연도 9, 부엌 2, 화장실 1, 출입구 1	건축물대장에 60-0 건물부재, 60-1 목조, 스투트 1959	동구 삼정동 조사구 7	거처정보
7	0	92-0	연건평 25, 대지 240, 방 3, 거실 1, 식당 1, 건축연도 11, 부엌 1, 화장실 1, 출입구 1	세벽, 새와, 77.74, 1980	건물존재 92-0	매칭오류 법정동다름
8	0	97-2	연건평 25, 대지 110, 방 3, 거실 1, 식당 1, 건축연도 2, 부엌 1, 화장실 1, 출입구 1	세벽, 새와, 68.94, 1980	건물존재 97-2	매칭오류 법정동다름
9	1	95-0	빈집	주택, 주유소	-	-
10	0	97-1	연건평 35, 대지 100, 방 4, 거실 1, 식당 1, 건축연도 9, 부엌 2, 화장실 1, 출입구 1	세벽, 새와, 67.28, 1980	건물존재 97-1	매칭오류 법정동다름
11	0	97-24	연건평 25, 대지 125, 방 4, 거실 1, 식당 1, 건축연도 9, 부엌 2, 화장실 2, 출입구 1	97-2 세벽, 새와, 68.94, 1980		매칭오류
12	0	97-4	연건평 35, 대지 100, 방 4, 거실 1, 식당 1, 건축연도 9, 부엌 2, 화장실 1, 출입구 1	조적도, 연와조, 78.4, 1996		매칭오류
13	0	98-1	삼정빌라 A 111	다세대 주택 98-1 삼정빌라A, B		매칭오류
14	0	98-1	삼정빌라 A 112	다세대 주택 98-1 삼정빌라A, B		매칭오류
15	0	98-1	삼정빌라 A 101	다세대 주택 98-1 삼정빌라A, B		매칭오류
16	0	98-1	삼정빌라 A 102	다세대 주택 98-1 삼정빌라A, B		매칭오류
17	0	98-1	삼정빌라 A 201	다세대 주택 98-1 삼정빌라A, B		매칭오류
18	0	98-1	삼정빌라 A 202	다세대 주택 98-1 삼정빌라A, B		매칭오류
19	0	98-1	삼정빌라 A 301	다세대 주택 98-1 삼정빌라A, B		매칭오류
20	0	98-1	삼정빌라 A 302	다세대 주택 98-1 삼정빌라A, B		매칭오류
21	0	98-1	삼정빌라 A 401	다세대 주택 98-1 삼정빌라A, B		매칭오류
22	0	98-1	삼정빌라 A 402	다세대 주택 98-1 삼정빌라A, B		매칭오류
23	0	98-1	삼정빌라 B 111	다세대 주택 98-1 삼정빌라A, B		매칭오류
24	0	98-1	삼정빌라 B 101	다세대 주택 98-1 삼정빌라A, B		매칭오류
25	0	98-1	삼정빌라 B 102	다세대 주택 98-1 삼정빌라A, B		매칭오류

주 : 건축물대장정보=지번, 구조명, 지붕구조명, 연면적, 사용승인일자

거처 번호	매칭 건물수	주소	거 처 명 부	건 축 물 대 장	지도	비고
26	0	98-1	삼정빌라 B 201	다세대 주택 98-1 삼정빌라A, B		매칭오류
27	0	98-1	삼정빌라 B 202	다세대 주택 98-1 삼정빌라A, B		매칭오류
28	0	98-1	삼정빌라 B 301	다세대 주택 98-1 삼정빌라A, B		매칭오류
29	0	98-1	삼정빌라 B 302	다세대 주택 98-1 삼정빌라A, B		매칭오류
30	0	98-1	삼정빌라 B 401	다세대 주택 98-1 삼정빌라A, B		매칭오류
31	0	98-1	삼정빌라 B 402	다세대 주택 98-1 삼정빌라A, B		매칭오류
32	0	98-27	한솔빌라 101	97-28, 단독주택, 철근 콘크리트, 159.07, 1997 (19가구)		거처정보
33	0	98-27	한솔빌라 102	97-28, 단독주택, 철근 콘크리트, 159.07, 1997 (19가구)		거처정보
34	0	98-27	한솔빌라 103	97-28, 단독주택, 철근 콘크리트, 159.07, 1997 (19가구)		거처정보
35	0	98-27	한솔빌라 104	97-28, 단독주택, 철근 콘크리트, 159.07, 1997 (19가구)		거처정보
36	0	98-27	한솔빌라 105	97-28, 단독주택, 철근 콘크리트, 159.07, 1997 (19가구)		거처정보
37	0	98-27	한솔빌라 106	97-28, 단독주택, 철근 콘크리트, 159.07, 1997 (19가구)		거처정보
38	0	98-27	한솔빌라 201	97-28, 단독주택, 철근 콘크리트, 159.07, 1997 (19가구)		거처정보
39	0	98-27	한솔빌라 202	97-28, 단독주택, 철근 콘크리트, 159.07, 1997 (19가구)		거처정보
40	0	98-27	한솔빌라 203	97-28, 단독주택, 철근 콘크리트, 159.07, 1997 (19가구)		거처정보
41	0	98-27	한솔빌라 304 (204오기?)	97-28, 단독주택, 철근 콘크리트, 159.07, 1997 (19가구)		거처정보
42	0	98-27	한솔빌라 205	97-28, 단독주택, 철근 콘크리트, 159.07, 1997 (19가구)		거처정보
43	0	98-27	한솔빌라 206	97-28, 단독주택, 철근 콘크리트, 159.07, 1997 (19가구)		거처정보
44	0	98-27	한솔빌라 301	97-28, 단독주택, 철근 콘크리트, 159.07, 1997 (19가구)		거처정보
45	0	98-27	한솔빌라 302	97-28, 단독주택, 철근 콘크리트, 159.07, 1997 (19가구)		거처정보
46	0	98-27	한솔빌라 303	97-28, 단독주택, 철근 콘크리트, 159.07, 1997 (19가구)		거처정보
47	0	98-27	한솔빌라 304	97-28, 단독주택, 철근 콘크리트, 159.07, 1997 (19가구)		거처정보
48	0	98-27	한솔빌라 305	97-28, 단독주택, 철근 콘크리트, 159.07, 1997 (19가구)		거처정보
49	0	98-27	한솔빌라 306	97-28, 단독주택, 철근 콘크리트, 159.07, 1997 (19가구)		거처정보
50	0	98-27	한솔빌라 401	97-28, 단독주택, 철근 콘크리트, 159.07, 1997 (19가구)		거처정보
51	0	98-61	연건평 25, 방 3, 거실 0, 식당 1, 건축연도 9, 부엌 1, 화장실 1, 출입구1	98-6,세벽,후형스레트,101.6,1980 98-61, 교육연구시설, 철근 콘크리트, 1847.5, 1982	대전 신학교 후생관	거처정보

주 : 건축물대장정보=지번, 구조명, 지붕구조명, 연면적, 사용승인일자

조사구 29(029A) 전체 거처가 11개의 공동 주택으로 이루어져 있다. 수작업으로 매칭하면 모든 거처와 건축물대장을 매칭할 수 있다. 또한 지번으로만 매칭하기보다는 공동주택의 명칭과 동 번호를 지번과 함께 매칭값으로 사용한다면, 데이터베이스로도 모두 매칭시킬 수 있다(<표 6-7>, <표 6-8>).

<표 6-7> 판암1동 조사구 29의 정보 매칭 요약

	매칭정보	비매칭정보			계
		거처정보	건축물대장	매칭오류	
거처수	40	0	0	0	40
%	0.0	0.0	100.0	0.0	100.0

<표 6-8> 판암 1동 조사구 29의 매칭 상세정보

거처 번호	매칭 건물수	주소	거 처 명 부	건 축 물 대 장	지도	비고
1	11	202-0	판암주공아파트 211-101	주공아파트 201동 - 211동	-	건축물대장 (공동주택)
2	11	202-0	판암주공아파트 211-102	주공아파트 201동 - 211동	-	건축물대장 (공동주택)
3	11	202-0	판암주공아파트 211-103	주공아파트 201동 - 211동	-	건축물대장 (공동주택)
4	11	202-0	판암주공아파트 211-104	주공아파트 201동 - 211동	-	건축물대장 (공동주택)
5	11	202-0	판암주공아파트 211-105	주공아파트 201동 - 211동	-	건축물대장 (공동주택)
6	11	202-0	판암주공아파트 211-106	주공아파트 201동 - 211동	-	건축물대장 (공동주택)
7	11	202-0	판암주공아파트 211-107	주공아파트 201동 - 211동	-	건축물대장 (공동주택)
8	11	202-0	판암주공아파트 211-108	주공아파트 201동 - 211동	-	건축물대장 (공동주택)
9	11	202-0	판암주공아파트 211-109	주공아파트 201동 - 211동	-	건축물대장 (공동주택)
10	11	202-0	판암주공아파트 211-110	주공아파트 201동 - 211동	-	건축물대장 (공동주택)
11	11	202-0	판암주공아파트 211-201	주공아파트 201동 - 211동	-	건축물대장 (공동주택)
12	11	202-0	판암주공아파트 211-202	주공아파트 201동 - 211동	-	건축물대장 (공동주택)
13	11	202-0	판암주공아파트 211-203	주공아파트 201동 - 211동	-	건축물대장 (공동주택)
14	11	202-0	판암주공아파트 211-204	주공아파트 201동 - 211동	-	건축물대장 (공동주택)
15	11	202-0	판암주공아파트 211-205	주공아파트 201동 - 211동	-	건축물대장 (공동주택)
16	11	202-0	판암주공아파트 211-206	주공아파트 201동 - 211동	-	건축물대장 (공동주택)
17	11	202-0	판암주공아파트 211-207	주공아파트 201동 - 211동	-	건축물대장 (공동주택)
18	11	202-0	판암주공아파트 211-208	주공아파트 201동 - 211동	-	건축물대장 (공동주택)
19	11	202-0	판암주공아파트 211-209	주공아파트 201동 - 211동	-	건축물대장 (공동주택)
20	11	202-0	판암주공아파트 211-210	주공아파트 201동 - 211동	-	건축물대장 (공동주택)
21	11	202-0	판암주공아파트 211-301	주공아파트 201동 - 211동	-	건축물대장 (공동주택)
22	11	202-0	판암주공아파트 211-302	주공아파트 201동 - 211동	-	건축물대장 (공동주택)
23	11	202-0	판암주공아파트 211-303	주공아파트 201동 - 211동	-	건축물대장 (공동주택)
24	11	202-0	판암주공아파트 211-304	주공아파트 201동 - 211동	-	건축물대장 (공동주택)
25	11	202-0	판암주공아파트 211-305	주공아파트 201동 - 211동	-	건축물대장 (공동주택)
26	11	202-0	판암주공아파트 211-306	주공아파트 201동 - 211동	-	건축물대장 (공동주택)
27	11	202-0	판암주공아파트 211-307	주공아파트 201동 - 211동	-	건축물대장 (공동주택)
28	11	202-0	판암주공아파트 211-308	주공아파트 201동 - 211동	-	건축물대장 (공동주택)
29	11	202-0	판암주공아파트 211-309	주공아파트 201동 - 211동	-	건축물대장 (공동주택)
30	11	202-0	판암주공아파트 211-310	주공아파트 201동 - 211동	-	건축물대장 (공동주택)
31	11	202-0	판암주공아파트 211-401	주공아파트 201동 - 211동	-	건축물대장 (공동주택)
32	11	202-0	판암주공아파트 211-402	주공아파트 201동 - 211동	-	건축물대장 (공동주택)
33	11	202-0	판암주공아파트 211-403	주공아파트 201동 - 211동	-	건축물대장 (공동주택)
34	11	202-0	판암주공아파트 211-404	주공아파트 201동 - 211동	-	건축물대장 (공동주택)
35	11	202-0	판암주공아파트 211-405	주공아파트 201동 - 211동	-	건축물대장 (공동주택)
36	11	202-0	판암주공아파트 211-406	주공아파트 201동 - 211동	-	건축물대장 (공동주택)
37	11	202-0	판암주공아파트 211-407	주공아파트 201동 - 211동	-	건축물대장 (공동주택)
38	11	202-0	판암주공아파트 211-408	주공아파트 201동 - 211동	-	건축물대장 (공동주택)
39	11	202-0	판암주공아파트 211-409	주공아파트 201동 - 211동	-	건축물대장 (공동주택)
40	11	202-0	판암주공아파트 211-410	주공아파트 201동 - 211동	-	건축물대장 (공동주택)

주 : 건축물대장정보=지번.. 구조명, 지붕구조명, 연면적, 사용승인일자

조사구 47(0471)의 경우는 8개 거처에서 해당 지번의 건축물대장에는 건축물이 없었다. 또한 지도에서도 해당하는 건축물이 없기 때문에 거처정보와 매칭 되지 않는다고 판단하였다. 이외의 거처에서는 매칭이 문제가 없었다(<표 6-9>, <표 6-10>).

<표 6-9> 판암1동 조사구 47의 정보 매칭 요약

	매칭정보	비매칭정보			계
		거처정보	건축물대장	매칭오류	
거처수	20	8	0	0	28
%	71.4	28.6	0.0	0.0	100.0

<표 6-10> 판암1동 조사구 47의 매칭 상세정보

거처 번호	매칭 건물수	주소	거 처 명 부	건 축 물 대 장	지도	비고
1 ↓ 8	1	395-22				
7	1	393-0				
8	0	395-17	연건평 29, 대지 120, 방 5, 거실 1, 식당 2, 건축연도 9, 부엌 2, 화장실 2, 출입구 1	해당 지번에 건축물 없음	건축물 부재	거처정보
9	1	395-18				
10	1	395-19				
11	0	395-16	연건평 35, 대지 147, 방 4, 거실 1, 식당 2, 건축연도 9, 부엌 2, 화장실 1, 출입구 1	해당 지번에 건축물 없음	건축물 부재	거처정보
12	1	395-15				
13	0	395-5	연건평 67, 대지 70, 방 10, 거실 3, 식당 2, 건축연도 8, 부엌 4, 화장실 4, 출입구 2	해당 지번에 건축물 없음	건축물 부재	거처정보
14	0	395-5	연건평 26, 대지 50, 방 4, 거실 0, 식당 2, 건축연도 11, 부엌 2, 화장실 2, 출입구 1	해당 지번에 건축물 없음	건축물 부재	거처정보
15	1	395-6				
16	1	395-30				
17	0	395-13	연건평 29, 대지 95, 방 5, 거실 1, 식당 3, 건축연도 9, 부엌 3, 화장실 3, 출입구 1	해당 지번에 건축물 없음	건축물 부재	거처정보
18	0	393-17	연건평 20, 대지 33, 방 3, 거실 1, 식당 1, 건축연도 9, 부엌 1, 화장실 1, 출입구 1	해당 지번에 건축물 없음	건축물 부재	거처정보
19	1	393-18				
20	1	393-15				
21	0	393-4	연건평 20, 대지 108, 방 4, 거실 1, 식당 1, 건축연도 9, 부엌 1, 화장실 1, 출입구 1	해당 지번에 건축물 없음	건축물 부재	거처정보
22	1	393-11				
23	1	393-29				
24	1	393-0				
25	1	391-5				
26	0	394-0	연건평 25, 대지 60, 방 5, 거실 2, 식당 2, 건축연도 9, 부엌 2, 화장실 2, 출입구 1	해당 지번에 건축물 없음	건축물 부재	거처정보
27	1	393-30				
28	1	393-31				

주 : 빈줄=매칭 된 거처, 건축물대장정보=지번, 구조명, 지붕구조명, 연면적, 사용승인일자

조사구 71(0711)의 경우도 모든 거처가 건축물과 매칭이 되지 않았다(<표 6-11>). 다섯 개의 거처가 존재하지만 거처정보에 있는 해당 지번에 어떠한 건축물 정보도 없다. 물론 여기에서 이러한 문제가 거처정보의 문제인지, 건축물대장의 문제인지 판단내리기 쉽지 않다. 다만, 거처정보의 부번이 모두 0인점, 건축물대장에서 유사한 건축물이 존재하는지 찾아본 결과, 본번은 일치하지만 부번이 0이 아닌

다수의 건축물이 존재한다는 점, 그리고 창성교회의 지번이 454가 아니라 사실은 445라는 점을 고려할 때, 거처정보의 지번 정보에 문제가 있다는 점을 추측할 수 있다(<표 6-12>).

<표 6-11> 판암1동 조사구 71의 정보 매칭 요약

	매칭정보	비매칭정보			계
		거처정보	건축물대장	매칭오류	
거처수	0	5	0	0	5
%	0.0	100.0	0.0	0.0	100.0

<표 6-12> 판암 1동 조사구 71의 매칭 상세정보

거처 번호	매칭 건물수	주소	거 처 명 부	건 축 물 대 장	지도	비고
1	0	452-0	영진조극사	해당 지번에 건축물 없음 452-2, 452-3 근린생활시설	452-3 리더 뉘시타운	거처정보
2	0	436-0	연건평 30, 대지 66, 방 2, 거실 1, 식당 0, 건축연도 10, 부엌 1, 화장실 1, 출입구 1	해당 지번에 건축물 없음 436-4, 436-5, 436-6에 각각 주택 존재	436-4 436-6 436-8	거처정보
3	0	434-0	연건평 15, 대지 45, 방 3, 거실 1, 식당 1, 건축연도 10, 부엌 1, 화장실 1, 출입구 1	해당 지번에 건축물 없음		거처정보
4	11	453-0	연건평 43, 대지 800, 방 4, 거실 1, 식당 1, 건축연도 9, 부엌 1, 화장실 1, 출입구 1	해당 지번에 건축물 없음		거처정보
5	11	454-0	창성교회	해당 지번에 건축물 없음	445 창성교회	거처정보

주: 건축물대장정보- (지번), 구조명, 지붕구조명, 연면적, 사용승인일자

3. 거처 특성과 매칭 현황

여기에서는 거처 특성별로 매칭 현황을 살펴보고자 한다. 먼저, 거처의 종류별로 매칭 현황을 살펴본다(<표 6-13>). 단독 주택의 경우는 약 46.3%가, 영업용 건물 내 주택의 경우는 약 66.7%가 건축물대장과 일대일로 매칭 된다. 일대일 매칭

이 되지 않은 경우는 주로 다대일로 매칭 되거나, 다대다로 매칭 된다. 다대일 매칭은 실제로는 여러 가구가 거주하지만, 가구 거주 단위별로 별도 매매가 되지 않는 다가구주택 같은 경우에 발생한다. 물론 거처정보와 건축물대장의 문제로 인해 다대다 매칭과 매칭 되는 건축물대장이 없는 경우도 있다.

아파트의 경우 매칭에 큰 문제는 없는 것으로 보인다. 즉 아파트의 경우는 다대일 매칭이지만, 매칭률이 100%에 이른다. 따라서 아파트의 경우, 건축물대장의 활용 가능성이 가장 높다. 다만 건축물 대장의 공동주택명이나 동 정보를 활용하지 않은 현재의 거처정보 작성방식으로는 거처와 아파트의 동이 서로 정확히 매칭 되지 않는 문제가 발생하고 있다. 이와 유사한 경우가 연립주택이다. 연립주택도 아파트와 마찬가지로 경우는 약 81.2%가 다대일 매칭이 되고 있다. 건축물 부재로 약 18.9%만이 매칭 되고 있지 않다.

한편 그 밖의 형태의 거처는 매칭률이 상대적으로 떨어진다. 다세대주택은 약 31.5%가, 기숙사/특수 사회시설의 경우는 100%가 해당하는 건축물대장 정보를 매칭시킬 수 없었다. 다세대주택의 경우를 자세히 살펴보면, 41.4%만이 공동주택으로 다대일 매칭이 되고 있다. 다세대주택이 비공동주택으로 다대일 매칭이 되는 경우는 15.3%에 이른다. 여러 가지 이유 중에서 무엇보다도 거처정보를 작성할 때 거처구분이 쉽지 않기 때문일 것이다. 현실적으로 다세대주택과 다가구주택이 구분이 쉽지 않다. 따라서 비공동주택 다대일 매칭을 보이는 다세대주택은 실제 건축물대장에는 다가구주택, 즉 단독주택일 경우도 많을 것으로 추측된다.

<표 6-13> 거처 종류별 매칭 현황

매칭유형	거처구분		단독주택		아파트		연립주택		다세대주택		영업용건물내 주택		기숙사/특수 사회시설		계	
	거처 수	%	거처 수	%	거처 수	%	거처 수	%	거처 수	%	거처 수	%	거처 수	%	거처 수	%
일대일	259	46.3	0	0.0	0	0.0	5	4.5	22	66.7	0	0.0	286	6.9		
다대일(공동주택)	4	0.7	3,308	100.0	85	80.2	46	41.4	0	0.0	0	0.0	3,443	83.6		
다대일(비공동주택)	40	7.1	0	0.0	1	0.9	17	15.3	1	3.0	0	0.0	59	1.4		
일대다(주부속건물)	5	0.9	0	0.0	0	0.0	0	0.0	1	3.0	0	0.0	6	0.1		
일대다(복수건물)	14	2.5	0	0.0	0	0.0	0	0.0	1	3.0	0	0.0	15	0.4		
다대다	72	12.9	0	0.0	0	0.0	8	7.2	1	3.0	0	0.0	81	2.0		
건축물대장부재	166	29.6	0	0.0	20	18.9	35	31.5	7	21.2	2	100.0	230	5.6		
계	560	100.0	3,308	100.0	106	100.0	111	100.0	33	100.0	2	100.0	4,120	100.0		

이 밖에 다른 거처 특성별로 매칭 유형을 살펴보면, 거주유무구분(거주/빈집) 과 건축연도에 의한 차이가 가장 두드러지게 나타난다(<표 6-14>). 사람이 실제로 거주하는 경우와 빈집의 경우 모두 공동주택으로 다대일 매칭을 보이는 비율이 가장 높다. 그런데 빈집은 공동주택으로 다대일 매칭을 보이는 비율이 84.7%에 이른다. 사람이 실제로 거주하는 경우 이 비율이 48.4%에 불과하다. 즉 빈집은 단일 건물에 다수의 가구가 거주하는 상태에서 가장 많이 발견된다는 사실을 알 수 있다.

<표 6-14> 거주유무별 매칭 현황

매칭유형	거주유무		거주		빈집		계	
	거처수	%	거처수	%	거처수	%	거처수	%
일대일	19	11.9	267	6.9	286	7.1		
다대일(공동주택)	77	48.4	3,302	84.7	3,379	83.3		
다대일(비공동주택)	5	3.1	54	1.4	59	1.5		
일대다(주·부속건물)	0	0.0	6	0.2	6	0.1		
일대다(복수건물)	0	0.0	15	0.4	15	0.4		
다대다(복수거처, 복수건물)	11	6.9	70	1.8	81	2.0		
건축물대장부재	47	29.6	183	4.7	230	5.7		
계	159	100.0	3,897	100.0	4,056	100.0		

건축연도에 따라서도 매칭 유형이 확연히 다르다(<표 6-15>). 1959년 이전에 건축된 거처의 경우는 단독주택 일대일 매칭이 30.7%를 차지한다. 이러한 결과는 1959년 이전에 건축된 대부분의 건물이 단독주택이기 때문에 나타나는 현상으로 보인다. 그리고 건축물대장 자체가 존재하지 않은 비율도 32.7%에 이른다. 반면에 1990년과 1994년에 건축된 건축물의 경우는 공동주택으로 다대일 매칭이 이루어진 비율이 97.9%에 이른다. 이 시기에는 건축물대장이 존재하지 않거나 다대다 매칭이 이루어진 비율도 현저하게 작다. 이러한 결과는 전체적인 매칭 유형뿐만 아니라 조사구별로 살펴보아도 확인할 수 있다(<표 6-16>, <표 6-17>).

<표 6-15> 건축년도별 매칭 현황

매칭유형		건축년도												
		2004	2003	2002	2001	2000	95-99	90-94	85-89	80-84	70-79	60-69	60년 이전	계
일대일	건	1	4	1	1	9	22	35	15	56	48	43	31	266
	%	25.0	40.0	14.3	25.0	56.3	4.7	1.2	31.9	36.6	49.0	46.7	30.7	6.7
다대일 (공동주택)	건	0	0	0	0	2	432	2,886	19	12	7	0	1	3,359
	%	0.0	0.0	0.0	0.0	12.5	91.7	97.9	40.4	7.8	7.1	0.0	1.0	85.0
다대일 (비공동주택)	건	0	2	0	1	0	1	1	0	24	5	10	9	53
	%	0.0	20.0	0.0	25.0	0.0	0.2	0.0	0.0	15.7	5.1	10.9	8.9	1.3
일대다(주·부 속건물)	건	0	0	1	0	0	1	3	0	1	0	0	0	6
	%	0.0	0.0	14.3	0.0	0.0	0.2	0.1	0.0	0.7	0.0	0.0	0.0	0.2
일대다 (복수건물)	건	0	1	0	0	0	3	2	2	2	1	0	4	15
	%	0.0	10.0	0.0	.0	0.0	.6	0.1	4.3	1.3	1.0	0.0	4.0	.4
다대다 (복수거처, 복수건물)	건	0	0	1	0	1	5	4	8	10	7	11	23	70
	%	0.0	0.0	14.3	0.0	6.3	1.1	0.1	17.0	6.5	7.1	12.0	22.8	1.8
건축물 대장 부재	건	3	3	4	2	4	7	16	3	48	30	28	33	181
	%	75.0	30.0	57.1	50.0	25.0	1.5	0.5	6.4	31.4	30.6	30.4	32.7	4.6
계	건	4	10	7	4	16	471	2,947	47	153	98	92	101	3,950
	%	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

<표 6-16> 건축년도별 매칭 된 거처수 (판암 1동 조사구 1)

거처번호		건축년도								
		2002	2000	95-99	90-94	85-89	80-84	70-79	60-69	60년 이전
1	0	0	0	1	0	0	0	0	0	
2	0	0	1	0	0	0	0	0	0	
4	0	0	0	0	0	1	0	0	0	
5	0	0	0	0	0	0	0	0	1	
7	0	0	0	0	0	0	0	0	1	
8	0	0	0	1	0	0	0	0	0	
9	0	0	0	0	0	0	0	0	2	
18	0	0	0	0	0	0	0	0	1	
19	0	0	0	0	0	0	0	0	1	
20	1	0	0	0	0	0	0	0	0	
21	1	0	0	0	0	0	0	0	0	
22	0	0	0	1	0	0	0	0	0	
23	0	0	0	0	0	0	0	0	2	
24	0	0	0	0	1	0	0	0	0	
25	0	0	0	0	0	0	0	0	2	
26	0	0	0	0	0	0	0	0	9	
28	0	0	0	0	1	0	0	0	0	
29	0	0	0	0	0	0	1	0	0	
30	0	0	0	0	0	0	1	0	0	
31	0	0	0	0	0	0	1	0	0	
32	0	0	0	0	0	1	0	0	0	
33	0	1	0	0	0	0	0	0	0	
37	0	0	0	0	0	3	0	0	0	
39	0	0	0	0	0	0	0	3	0	
40	0	0	0	0	0	0	3	0	0	
41	0	0	0	1	0	0	0	0	0	
42	1	0	0	0	0	0	0	0	0	

<표 6-17> 건축년도별 매칭 된 거처수 (판암 1동 조사구 2)

건축년도 거처번호	2003	2002	2001	2000	95-99	90-94	85-89	80-84	70-79	60-69	60년 이전
1	0	0	0	0	0	0	0	0	0	0	3
2	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	1	0
4	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	1	0	0
8	0	0	0	1	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	1	0	0	0
10	0	0	0	0	0	0	0	2	0	0	0
11	0	0	0	1	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0	1	0	0
14	0	0	0	0	0	0	0	0	0	1	0
15	0	0	0	0	0	0	0	0	0	0	0
16	0	0	0	0	0	0	0	0	0	0	0
17	0	0	0	0	0	0	0	0	0	0	0
18	0	0	0	0	0	0	0	0	0	0	0
19	0	0	0	0	0	0	0	0	0	0	0
20	0	0	0	0	0	0	0	0	0	0	0
21	0	0	0	0	0	0	0	0	0	0	0
22	0	0	0	0	0	0	0	0	0	0	0
23	0	0	0	0	0	0	0	0	0	0	0
24	0	0	0	0	0	0	0	0	0	0	0
25	0	0	0	0	0	0	0	0	0	0	0
26	0	0	0	0	0	0	0	0	0	0	0
27	0	0	0	0	0	0	0	0	0	0	0
28	0	0	0	0	0	0	0	0	0	0	0
29	0	0	0	0	0	0	0	0	0	1	0
30	0	0	0	0	0	1	0	0	0	0	0
31	0	0	0	0	0	0	0	1	0	0	0
32	0	0	0	0	0	0	0	0	0	0	1
33	0	0	0	0	0	0	0	0	0	0	1
34	0	0	0	0	0	0	0	1	0	0	0
35	0	0	0	0	0	1	0	0	0	0	0
36	0	0	0	0	0	0	0	0	0	1	0
37	0	0	0	0	0	0	0	0	0	0	0
38	0	0	0	0	0	2	0	0	0	0	0
39	0	0	0	0	0	0	0	2	0	0	0
40	0	0	0	0	0	0	0	2	0	0	0
41	0	0	0	0	0	0	0	2	0	0	0
42	0	0	0	0	0	0	0	2	0	0	0
43	0	0	0	0	0	0	0	2	0	0	0
44	0	0	0	0	0	0	0	2	0	0	0
45	0	0	0	0	0	0	0	2	0	0	0
46	0	0	0	0	0	0	0	2	0	0	0
47	0	0	0	0	0	0	0	2	0	0	0
48	0	0	0	0	0	0	0	2	0	0	0
49	0	0	0	0	0	0	0	0	1	0	0
50	0	0	0	1	0	0	0	0	0	0	0
51	0	0	0	0	0	0	0	0	0	1	0
52	0	0	0	0	0	1	0	0	0	0	0
53	0	0	0	0	0	0	0	0	0	1	0
54	0	0	0	0	0	0	0	0	0	1	0
55	0	0	0	0	0	0	0	0	1	0	0
56	0	0	0	0	0	0	0	0	5	0	0
57	0	0	0	0	0	0	0	1	0	0	0

마지막으로 연건평과 매칭 유형을 살펴보면, 이들 사이에 뚜렷한 관계가 나타나지는 않는다(<표 6-18>). 다만 두 가지 사실에 주목할 필요가 있다. 첫째, 연건평이 넓을수록 일대일 매칭이 되는 비율이 높다. 반면에 연건평이 작을수록 다대일 매칭이 되는 비율이 높다. 둘째, 특정 연건평대의 거처에서는 다른 평대의 거처에 비해 건축물대장이 없는 경우가 상대적으로 많다. 70평-89평대 사이 거처에서 건축물대장이 없는 비율이 약 44.4%에 이른다. 그리고 40평-49평대의 거처의 경우 이 비율이 38.9%로 나타나고 있다.

<표 6-18> 연건평별 매칭 현황

연건평(평) 매칭유형		≤ 19	20-29	30-39	40-49	50-69	70-89	90-149	150 ≤	계
일대일	건	68	104	52	18	10	4	4	3	263
	%	2.3	12.4	44.8	50.0	58.8	44.4	66.7	100.0	6.7
다대일 (공동주택)	건	2,713	628	17	0	0	0	0	0	3,358
	%	92.9	74.9	14.7	0.0	0.0	0.0	0.0	0.0	85.1
다대일 (비공동주택)	건	25	16	9	1	2	1	0	0	54
	%	0.9	1.9	7.8	2.8	11.8	11.1	0.0	0.0	1.4
일대다 (주·부속건물)	건	0	4	1	1	0	0	0	0	6
	%	0.0	0.5	0.9	2.8	0.0	0.0	0.0	0.0	0.2
일대다 (복수건물)	건	2	5	5	1	2	0	0	0	15
	%	0.1	0.6	4.3	2.8	11.8	0.0	0.0	0.0	.4
다대다 (복수거처/건물)	건	31	27	10	1	1	0	0	0	70
	%	1.1	3.2	8.6	2.8	5.9	0.0	0.0	0.0	1.8
건축물대장 부재	건	80	54	22	14	2	4	2	0	178
	%	2.7	6.4	19.0	38.9	11.8	44.4	33.3	0.0	4.5
계	건	2,919	838	116	36	17	9	6	3	3,944
	%	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

4. 거처정보와 건축물대장 정보 비교

건축물대장은 건축물의 가장 기초적이고 중요한 정보를 포함하고 있는 행정 자료로 비교적 엄격하게 관리되고 있다. 그러나 실제 값과는 여러 가지 이유로 인해 차이가 날 수 있다. 예를 들면, 자료의 수정과 보완에서 문제가 있을 수 있다. 항시 담당 공무원에 의해 건축물 자료가 수정되고 보완되지 않으면 실제 값과는 차이가 날 수 있다. 또 자료를 입력할 때 문제가 생길 수도 있다.

이와 달리 거처정보는 일정한 시점에 실제로 조사한 결과이다. 따라서 가장 현실에 가까운 자료이다. 그러나 이 자료도 조사와 부호화 등 여러 과정에서 오류

가 발생할 수 있다. 예를 들면, 조사자가 주소를 잘 못 기재하였을 수도 있고, 주택과 관련된 정보를 잘 못 기재했을 수도 있다. 또한 조사대상에 대한 정보가 잘 못 되어 있을 수도 있다.

여기서는 거처정보와 건축물대장의 정보를 비교함으로써 앞으로 행정자료를 이용하는데 유용성을 평가하기로 한다. 이를 위해서 판암1동의 경우 일대일 매칭이 된 287개 건물만을 선택하여 평가한다. 이렇게 선택한 이유는 무엇보다도 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.

어떠한 정보를 비교하는가하는 것은 또 다른 차원의 문제이다. 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.

일반적으로 연면적은 내 대지 안에 있는 건물면적의 총 합계를 말하고, 단위로는 제곱미터(m^2)를 사용한다. 연건평은 같은 의미를 가지지만 제곱미터에 0.3025를 곱하여 평으로 환산한 것이다($m^2 \times 0.3025 = \text{평}$). 물론 공동주택의 경우에는, 여러 동이 존재할 수 있기 때문에 거처정보의 연건평과 건축물대장의 동연면적을 그대로 비교하기는 힘들다. 그러나 단독 주택의 경우는 특별한 문제가 없을 것이다. 아래 <그림 6-3>은 건축물대장의 동연면적을 평으로 환산하여, 거처정보의 연건평 자료와의 차이를 계산하여 제시한 그림이다. 여기에서는 모두 233개 거처만을 비교하였는데, 이는 약 50여개 거처와 매칭 된 건축물대장의 건축물에 대한 동연면적 정보가 없기 때문이다.

이를 살펴보면, 약 22.3%의 거처만이 건축물대장과 정보의 차이가 4평 이하로 난다. 9평 이하로 범위를 넓히면 약 40%의 거처가 이에 해당한다. 이러한 사실에 미루어 몇 가지 점을 고려해야 할 것이다. 만일 거처정보가 사실에 가깝다면, 적어도 단독 주택의 경우에 건축물대장의 정보의 유용성을 매우 제한적으로 받아

들여야 할 것이다. 반면에 건축물대장의 정보가 좀 더 사실에 근접한 정보라면, 건축물대장을 적극적으로 활용할 필요가 있다. 조사원들이 건축물대장의 관련 정보를 확인하는 방식을 취한다면 시간이나 비용이라는 측면에서 뿐만 아니라, 정보의 정확성이라는 면에서도 상당한 기여를 할 것이다.⁸⁾ 사실 공동 주택이 아닌 단독 주택에서 거주자가 연건평을 정확하게 알고 있고 조사자가 이를 정확하게 조사하기는 쉽지 않을 것이다.⁹⁾

<그림 6-3> 거처정보의 연건평과 건축물대장의 동연면적 차이

5. 소결

1) 조사구별 매칭률

매칭률은 두 가지 방식으로 계산할 수 있다. 첫째, 한 거처에 대해 복수의 건축물대장 정보가 매칭 되는 것을 인정하는 방식이 있다. 이를 조매칭률이라고 할 수 있다. 한 거처에 여러 건물 정보가 매칭 되는 경우는 대개 거처가 공동주택이거

8) 조사원들이 건축물대장을 활용할 경우 개인정보보호의 문제도 제기될 수 있다. 이에 대한 심층적인 평가가 추후에 이루어져야 할 것이다.

9) 실제 조사하는 사람들의 이야기로는 주변 공인 중개소 등의 도움을 받는다고 한다.

나 주건물에 부속 건물이 동시에 한 지번 내에 존재할 때 발생한다. 혹은 상가 건물이 한 지번 내에 밀집해 있고, 여기에 거처가 있을 때 발생하기도 한다. 물론 건축물대장의 정비가 잘 이루어져 있지 않을 경우에도 발생한다.

둘째, 순매칭률은 중복매칭을 배제하고, 한 거처에 하나의 건물정보만 연결시켜 매칭률을 계산한 것이다. 물론 데이터베이스 내에서 이 과정을 기계적으로 수행하여 중복 매치를 배제시킬 수도 있다. 이 연구에서는 일종의 기본 자료를 확보하기 위하여, 연구자가 거처정보와 건축물대장을 수작업으로도 매칭시켜, 순매칭률을 계산하였다.

중복을 배제한 기계적 방식의 매칭작업과 수작업 방식의 매칭작업의 결과는 전체적으로 볼 때, 매칭률에서 큰 차이를 보이고 있지 않다. 다만, 매칭의 내용이 다르다. 예를 들면, 공동주택의 경우, 다수의 건축물 정보 중에서 어떠한 정보를 선택하는가에 따라 매칭의 내용이 많이 달라질 수 있다. 따라서 기계적 방식의 매칭도 거처의 종류에 따라 주소이외의 정보 (공동주택명, 동번호 등)를 동시에 사용하여야 수작업과 유사한 매칭내용을 얻을 수 있을 것이다.

2) 주요 조사구의 매칭 현황

관암 1동에서 총 73개의 조사구 중에 모두 5개의 조사구에 대해 심층적인 매칭 현황에 대한 분석을 진행하였다. 여기에는 평균적인 조사구(조사구 1와 조사구 47), 매칭률이 낮은 조사구(조사구 6), 공동 주택으로 이루어진 조사구(조사구 29), 매칭이 전혀 되지 않은 조사구(조사구 71)가 포함되어 있다.

전체적인 결과를 살펴보면, 첫째, 거처정보의 오류라고 판단되는 부분이 다수 발견된다. 이러한 이유는 조사자의 오류일 수도 있고, 입력자의 오류일 수도 있다. 둘째, 한 거처에 다수의 건물이 매칭 되는 경우는 매칭 되는 건물의 특성과 관련이 깊다. 즉, 공동주택, 1950년 이전에 사용허가가 난 노후 건물, 상가건물, 또는 주건물과 부속건물이 존재하는 경우, 한 거처에 다수의 건축물이 매칭 된다. 여기에서 공동주택의 경우는 기술적인 처리를 한다면 매칭과 건축물대장 활용에 큰 문제가 없을 것이다. 다른 건물의 경우는 향후 건축물대장이 정비됨에 따라 문제가 해결될 것으로 기대한다. 다만 그 동안은 이들 건물에 대해서는 조사원에 의한 직접 조사는 불가피할 것으로 보인다.

3) 거처 특성과 매칭 현황

거처 특성별로 매칭 현황에 대한 분석도 조사구별 분석과 유사하다. 구체적으로 이를 살펴보면, 단독 주택의 경우는 절반 정도가 건축물대장과 일대일 매칭이 된다. 다세대 주택을 제외한 다른 형태의 거처에서는 적어도 매칭에 있어서 큰 문제는 없는 것으로 보인다. 아파트의 경우는 100% 매칭이 됨으로써, 건축물대장의 활용 가능성이 가장 높다. 연립주택의 경우는 약 80.2%가 공동주택으로 다대일 매칭이 되고 있다.

한편 다세대주택의 경우는 41.4%만이 공동주택으로 다대일 매칭이 된다. 비공동주택으로 다대일 매칭이 되는 경우는 15.3%에 이른다. 이 밖에 빈집 구분과 건축연도가 매칭률과 관련이 있는 것으로 보인다. 즉 대부분의 빈집은 다세대주택에서 가장 많이 발견된다. 그리고 건축연도가 1959년 이전의 거처의 경우는 일대일 매칭률이 상대적으로 낮다.

4) 거처정보와 건축물대장 비교

일대일 매칭이 이루어진 거처(대부분 단독주택)의 연건평(거처정보)과 동연면적(건축물대장)을 비교하였다. 분석 결과, 두 개의 정보의 차이가 많이 난다. 이러한 사실은 두 가지 방식으로 해석할 수 있다. 만일 거처정보가 사실에 가깝다면 적어도 이러한 거처에 대해서는 건축물대장 자료와의 연결을 매우 제한적으로 활용해야 할 것이다. 반면에 건축물대장의 정보가 좀 더 사실에 근접한 정보라면, 인구주택총조사에서 건축물대장을 좀 더 적극적으로 활용할 필요가 있다. 즉 건축물대장을 조사구 설정에 사용하는 것에 그치지 말아야 할 것이다. 조사원들이 건축물대장의 관련 정보를 확인하는 방식을 취한다면 시간이나 비용이라는 측면에서 뿐만 아니라, 정보의 정확성이라는 면에서도 상당한 기여를 할 것이다. 사실 공동 주택이 아닌 단독주택에서 거주자가 연건평을 정확하게 알기도 어렵고, 조사자가 이를 정확하게 조사하기도 쉽지 않을 것이다. 따라서 건축물대장을 조사구 설정과 실제 조사에서 적극적으로 활용하는 방안을 모색해야 할 것이다.

6. 정책적 제안

1) 거처정보의 오류 최소화

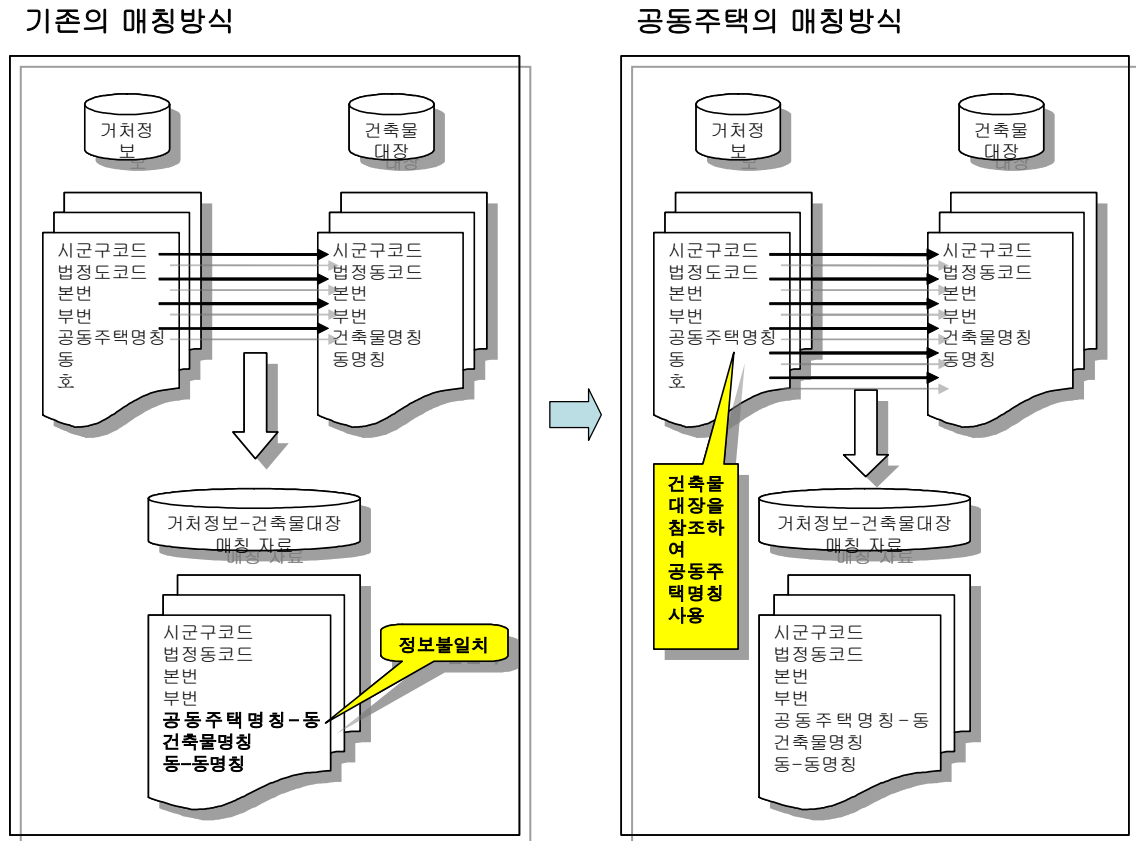
건축물대장 같은 행정정보를 거처정보에 연결시키기 위해서는 좀 더 정확한 거처정보의 정보를 확보할 필요성이 있다. 특히 거처정보와 건축물대장의 연결시키는 주요 정보인 법정동, 본번, 부번에서 발생하는 오류를 최소화할 필요가 있다. 거처정보 자료를 작성할 때, 여러 가지 단계에서 오류가 발생할 수 있다. 조사 단계에서 뿐만 아니라 입력 단계에서도 오류가 발생할 수 있다. 예를 들면, 조사자나 입력자가 해당 조사구에 존재하지 않은 지번을 입력하거나, 부번을 입력할 수도 있다. 혹은 잘못된 법정동 정보를 입력할 수도 있다.

따라서 앞으로 조사에서는 이러한 오류를 최소화하는 방식으로 조사와 입력 과정을 재구조화할 필요가 있다. 예를 들면, 중앙화된 입력 프로그램을 통해 해당 조사구와 지번, 법정동 정보, 공동주택명 등을 입력할 때, 잘못된 정보를 즉시 수정할 수 있도록 하는 체계를 만들 필요가 있다. 즉 해당 조사구에 존재하지 않은 정보를 입력할 때, 오류 발생 등을 알려줄 수 있는 시스템을 구성하여, 개인적인 오류의 발생을 최소화해야 할 것이다.

2) 매칭방법

인구주택총조사에서는 연구자의 수작업으로 모든 거처를 건축물대장과 매칭을 할 수 없을 것이다. 따라서 데이터베이스를 이용한 기계적인 매칭은 필수적이다. 다만 좀 더 효율적인 매칭을 위해서는 매칭방식에서 거처의 종류를 고려할 필요가 있다. 먼저 거처의 종류에 따라 매칭 유형이 많이 다르다. 예를 들면, 관암1동에서는 공동주택의 경우는 일대다 매칭이 많은 반면, 단독주택의 경우 일대일 매칭이 절반에 이른다. 따라서 거처 종류별로 매칭 방식을 다르게 적용한다면 좀 더 효율적인 매칭이 될 수 있을 것이다. 이와 관련하여 공동주택의 경우 두 자료를 매칭할 때 매칭값으로 시군코드, 법정도, 본번, 부번 이외에도 거처의 종류에 따라 다른 매칭값을 추가로 사용하면 좀 더 좋은 결과를 얻을 수 있다. <그림 6-4>에서 보듯이, 거처정보의 공동주택명과 동일련번호 및 건축물대장의 건축물명칭과 동명칭을

각각의 매칭값으로 사용하면 공동주택의 상당한 부분을 일대일로 정확하게 매칭 할 수 있다.



<그림 6-4> 데이터베이스 매칭 방식 개선 방안 - 공동주택의 경우

3) 거처 특성과 건축물대장

비교분석 결과, 거처의 종류와 건축연도 같은 거처 특성에 따라 매칭률과 매칭 유형이 달라진다. 따라서 일률적으로 건축물대장을 적용하는 것은 현재로서는 무리가 있을 수 있다. 따라서 1959년 이전의 노후 건물과 단독주택, 상가건물에 존재하는 거처에 관해서는 조사원에 의한 직접적인 조사가 당분간은 불가피하다.

4) 건축물대장 활용

현재로서는 건축물대장을 조사구 설정 단계에서만 활용하는 것으로 그친다.

앞으로는 건축물대장을 조사구 설정에 사용하는 것에 그치지 말아야 할 것이다. 조사가가 주택에 관한 정보를 정확하게 조사하기는 쉽지 않을 것이다. 예를 들면, 공동 주택이 아닌 단독주택에서 거주자가 연건평이나 건축연도를 정확하게 알기가 힘들 것이다. 따라서 조사원들이 건축물대장의 관련 정보를 확인하는 방식을 취한다면, 시간이나 비용이라는 측면에서 뿐만 아니라, 정보의 정확성이라는 면에서도 상당한 기여를 할 것이다. 물론 이러한 건축물대장의 활용은 건축물대장의 정보가 얼마나 정확한가에 달려있다. 따라서 건축물대장의 정비와 보완이 비교적 잘 되어 있는 대도시부터 적용해야 할 것이다.

VII. 결론 및 제안

1. 현행 조사방법에 대한 평가

조사환경의 변화와 함께 읍면동의 통계기능이 시군구로 이전되면서 인구주택총조사는 여러 가지 어려움에 직면하고 있다. 특히 조사구 설정은 지역주민들에 대한 구체적인 거처 및 가구에 관한 정보를 필요로 하기 때문에 과거에는 읍면동에서 가설정한 다음 통계청에서 최종적으로 승인하였다. 그러나 이제는 읍면동의 통계기능이 없기 때문에 지역에 관한 새로운 정보자원 필요하다. 이에 통계청은 과거 2000년도 인구주택총조사에서 축적된 정보와 2000년 이후 새로이 건축된 건물 및 변경된 건물에 대한 건축인허가대장과 건축물대장 등의 행정자료를 활용하였다. 앞으로 준비조사기간에 실제 확인을 통해 새로운 조사구 설정방식의 적절성과 효율성이 확인이 될 것이다.

이와 함께 수치지도를 이용하여 건물과 도로를 확인하고 이를 바탕으로 조사구 경계를 설정하는 작업을 진행하고 있다. 준비조사기간에는 수치지도를 바탕으로 만든 조사구 지도에 변경이나 오류 사항 등을 직접 수정하여 실제 조사에서는 현상에 가장 가까운 지도를 사용할 수 있는 장점이 있다. 또한 수치지도는 도형자료와 속성자료를 연계하여 사용할 수 있기 때문에 조사구 내부에 있는 개별 건물의 속성을 파악하고 이를 지도에 표시할 수 있는 이점이 있다. 건물이나 거처 및 가구에 대한 속성 자료가 불완전 할 경우 조사를 통해 이를 시각적으로 대조하여 쉽게 파악할 수 있는 장점도 있다.

전체적으로 보면, 건물, 거처, 가구 등 주거에 관한 자료가 불완전한 상황에서 수치지도를 바탕으로 이들 자료를 서로 보완적으로 조사구 설정에 활용하는 것은 효과적인 방법이다. 이러한 방법으로 잘못된 정보를 발견하고 수정하거나 보완하기가 손쉽기 때문이다. 앞으로 수치지도를 지속적으로 최근의 것으로 갱신하고, 조사에 필요한 관련 행정자료들 역시 최신의 정보로 갱신할 수 있는 체계를 갖춘다면 과거 읍면동에서 지원하던 부분을 완전히 대체할 수 있을 것이다. 다만 행정자료를 통계적으로 활용하기 위해서는 자료의 정비가 필요하며, 이들을 서로 연계하여 활용하기 위해서는 자료연결방법에 관한 연구가 필요하다.

2. 가구명부-거처정보와 건축물대장 자료 매칭에 대한 평가

이 연구에서는 2004년 7월에 수행한 인구주택총조사 시험조사 결과 생성된 가구명부와 거처정보 자료를 사용하였다. 조사대상 지역은 대전시 동구 판암1동, 인천시 광화군 화도면, 광주시 남구 백운1동이며, 조사항목은 인구주택총조사의 주택 관련 항목이다. 이와 더불어 주거부문 행정자료로는 건축물대장을 택하였다. 건축물대장은 주거부문 여러 행정자료 가운데 전국을 포괄하는 가장 대표적인 행정자료이기 때문이다. 여기서는 가구명부, 거처정보 자료를 건축물대장과 매칭한 자료를 분석하였다. 분석의 목적은 인구주택총조사에서 건축물대장 자료의 활용가능성을 평가하는 것이었다.

건축물대장의 활용가능성을 평가하는 것 이외에도, 가구명부, 거처정보, 건축물대장의 매칭은 인구주택총조사의 조사방법을 개발하는 데 중요한 역할을 할 것이다. 앞으로 인구주택총조사에서는 지난 인구주택총조사에서 축적된 정보를 활용해야 하는데 이를 수정·보완하기 위해서도 자료 매칭에 대한 방법은 지속적으로 개발되어야 한다. 또한 인구주택총조사에서 획득한 정보를 잘 못된 행정자료를 찾아 수정하기 위해서도 매칭 방법에 대한 연구가 필요하다. 이런 측면을 염두에 두고 자료 매칭에 대한 평가와 함께 개선 방안을 살펴본다.

건축물대장을 가구명부 및 거처정보와 함께 사용하려면 일단 두 자료가 매칭되어야 한다. 이 연구에서는 시험조사자료와 건축물대장 자료의 매칭에 두 가지 방법을 사용하였다. 두 가지 방법은 데이터베이스의 쿼리(query)기능을 사용한 매칭(DB매칭)과 수작업으로 한 매칭(대조매칭)이다. 두 방법 모두에서 시군구코드, 법정동코드, 본번, 부번을 기본 매칭값으로 하여 거처정보와 건축물대장 자료를 매칭하였다.

수작업 자료와 통계청의 매칭자료를 비교한 결과 데이터베이스의 쿼리를 사용하여 매칭시킬 때 다양한 유형의 오류들이 발생한다는 것을 알 수 있었다. 중요한 오류 사례들로는 가구명부 및 거처 정보의 오차에서 발생하는 매칭 오류가 있었고, 가구명부의 자료와 건축물대장의 정보가 일치하지 않는 경우가 있었으며, 한 지번에 공동주택과 건축물대장이 없는 단독주택이 동시에 존재할 때 발생할 수 있는 오류가 있었다. 또 다른 오류로는 법정동이 잘못 조사되었거나 건축물대장 정보에 법정동 주소가 잘못 기재되었을 경우 일수도 있다. 다시 말해, 조사자의 오류일 수

도 있고 건축물대장의 입력 오류일 수도 있다.

인구주택총조사에서는 수작업으로 거처를 건축물대장과 매칭을 할 수 없으며, 데이터베이스의 기능을 사용하는 기계적인 매칭이 필수적이다. 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 고려할 필요가 있다. 거처의 종류에 따라 매칭 유형이 많이 다르기 때문이다. 예를 들면, 관암1동에서는 공동주택의 경우는 일대다 매칭이 많은 반면, 단독주택의 경우 일대일 매칭이 절반에 이른다. 따라서 거처 종류별로 매칭 방식을 다르게 적용한다면 더 효율적으로 매칭 할 수 있을 것이다.

이와 관련하여 공동주택의 경우 두 자료를 매칭 할 때 사용하는 매칭값으로 시군코드, 법정도, 본번, 부번 이외에도 거처의 종류에 따라 다른 매칭값을 사용할 수 있다. 주소와 함께 공동주택명, 동번호, 동명칭을 동시에 사용하면 수작업과 유사한 매칭의 효과를 얻을 수 있을 것이다. 특히 동명칭은 다양한 형태로 있는데, 이를 고려할 수 있는 매칭 방법을 개발할 수 있으면 정교한 매칭을 할 수 있다. 공동주택의 매칭이 정교해지면 대도시 주거지역에서 건축물대장 정보의 활용이 크게 향상될 수 있다.

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. 예를 들면, 조사자나 입력자가 해당 조사구에 존재하지 않은 지번을 입력하거나, 부번을 입력할 수도 있으며, 잘못된 법정동 정보를 입력할 수도 있다. 따라서 앞으로 조사에서는 이러한 오류를 최소화하는 방식으로 조사와 입력 과정을 재구조화할 필요가 있다. 예를 들면, 중앙화된 입력 프로그램을 통해 해당 조사구와 지번, 법정동 정보, 공동주택명 등을 입력할 때, 잘못된 정보를 즉시 수정할 수 있도록 하는 체계를 만들 필요가 있다. 즉 해당 조사구에 존재하지 않은 정보를 입력할 때, 오류 발생 등을 알려줄 수 있는 시스템을 구성하여, 개인적인 오류의 발생을 최소화해야 할 것

이다.

3. 자료 매칭의 시사점

거처정보와 건축물대장을 매칭한 결과 강화군 화도면에서는 매칭률이 60%에 못 미친 반면, 광역시의 주거지역인 판암1동과 백운1동에서는 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.

거처 특성별로 매칭 유형에 차이를 살펴보면, 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 판암1동에서 매칭률이 92.6%, 공동주택이 62.4%인 백운1동에서 92.6%인 반면, 공동주택이 없는 화도면에서는 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 이전의 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

가구명부-거처정보와 건축물대장 정보를 매칭 할 때 다양한 매칭 유형들이

존재하며, 이들 유형에 따라 건축물대장 정보의 활용 가능성이 달랐다. 거처와 건물의 연결이 가장 손쉬운 경우는 거처 1곳과 건축물대장의 자료 1건이 매칭 될 때이다. 단독주택에 1 가구가 거주하는 전형적인 주거의 형태라고 할 수 있다. 그렇지만 건축물대장만으로는 그 건물에 있는 거처수와 가구수를 알 수는 없기 때문에 조사구 설정에서 건축물대장의 정보에만 의존할 수는 없다. 단지 조사구 지도에서 건물의 존재유무와 거처의 존재유무를 확인하는 데는 도움이 된다.

다음으로 다수의 거처와 건축물대장의 정보 1건이 매칭 된 경우를 살펴보았다. 실제 주거가 공동주택인 경우이지만 건축물대장에는 단독주택으로 기재된 경우도 있어 역시 건축물대장의 정보만으로는 조사구 설정을 할 수 없다. 그렇지만 건축물대장의 주용도가 공동주택으로 등록된 경우는 건축물대장의 정보를 조사구 설정에 활용할 수 있다. 일단 건물과 거처의 존재 파악에 활용할 수 있다. 현재 건축물대장 자료로는 가구수를 파악할 수 없으나, 지방자치단체에 있는 건축물의 상세 자료가 전산화되어 통합된다면 가구수 파악도 가능하다. 현재는 주로 공시가격 파악을 위해 한국감정원이 보유한 아파트 관련 자료로 가구수를 파악하고 있다.

아파트와 같은 큰 규모의 공동주택은 건축물대장으로 쉽게 파악할 수 있다. 그러나 그렇지 않으면서 다수의 거처와 건축물 대장 정보 1건이 매칭 된 경우는 상대적으로 파악하기 어렵다. 등록 정보로는 1 건축물이지만 실제로 여러 가구가 사는 경우로 도시에서는 주로 다가구주택이라고 할 수 있다. 그러나 강화군 화도면에서 그러한 거처가 11.6%에 이르러 도시지역보다 월등히 높다는 점은 건축물대장에 등록되지 않은 거처가 많아서 생긴 문제일 가능성을 보여준다. 기실 건축물대장을 중심으로 보면 일대일 매칭과 다대일 매칭을 구별할 수 없다. 한 지번에 건축물대장이 하나 있으며 주용도가 공동주택이 아닌 경우는 실제로 그 지번을 조사하여 거처수와 가구수를 파악할 수 밖에 없다. 건축물대장은 단지 그 지번에 적어도 하나의 건물과 거처가 존재한다 사실을 알려줄 뿐이다.

그 외에도 매칭은 되었으나 건축물대장 정보의 활용이 매우 제한된 경우는 다음과 같은 네 가지 유형을 찾아 볼 수 있다. 1) 거처 1곳에 1개의 주택과 여러 기타용도의 건축물대장이 매칭 된 경우, 2) 거처 1곳에 다수의 주택 또는 시설에 대한 건축물대장이 매칭 된 경우, 3) 한 지번에 거처와 건축물대장 자료가 모두 다수로 매칭 된 경우 등이다. 거처 1곳에 1 주택과 변소나 창고와 같은 다른 용도의 건물이 있는 경우는 흔히 한옥에서 많이 볼 수 있으며 농촌에 많은 주택형태로 짐

작할 수 있다. 그러나 이러한 유형의 거처는 화도면에는 1%에 불과한 반면 광주시 백운1동에서는 21.0%에 이르렀다. 따라서 건축물대장의 1 주택과 다수의 기타용도 건물은 실제로 여러 주택일 수도 있다. 한편 거처 1곳에 해당되는 건축물대장에는 다수의 주택이나 다수의 시설이 매칭 된 경우는 거처가 실제 다수의 주택으로 구성 되었을 수 있기도 하지만 해당 지번에 여러 개의 독립된 주택이나 독립된 시설들이 존재할 수도 있다. 이러한 경우는 화도면에서 4.3%로 가장 높게 나타났다. 또한 한 지번에 거처와 건축물대장 자료가 모두 다수로 매칭 된 경우도 역시 화도면에서 13.0%로 다른 지역에 비해 월등히 높게 나타났다. 이러한 결과는 농촌지역의 건축물대장이 잘 정비되지 않아 나타난 현상일 가능성도 배제할 수 없다. 이들 세 유형을 건축물대장을 중심으로 종합하면, 한 지번에 건축물대장은 하나지만 공동주택이 아니면서 여러 채의 건물이 있거나 여러 개의 건축물대장이 있는 경우 건축물대장은 단지 건물과 거처의 존재만을 알려줄 뿐이다.

끝으로 거처에 해당하는 건축물대장이 존재하지 않는 경우이다. 이 유형은 1) 건축물로 등록이 되지 않은 무허가 건물인 경우, 2) 등록된 건축물대장 정보가 전산 입력시 누락된 경우, 3) 법정동이나 지번이 잘못 입력되어 건축물대장 자료를 매칭 하지 못한 경우 등에 의한 것일 수 있다. 이 경우 거처의 지번에 해당하는 건축물대장 정보가 없어서 현재는 활용할 수 없으나 만약 그 이유가 자료의 전산화 과정이나 조사과정에서 발생했을 때는 오류 수정 후 건축물대장 정보를 활용할 수 있다.

이 모두를 건축물대장을 중심으로 종합하면, 건축물대장에서 공동주택으로 등록된 경우 건물과 거처를 파악할 수 있다. 만약 조금 더 상세한 정보를 활용할 수 있으면 가구수도 파악할 수 있으며 조사구 설정에 직접 활용할 수 있다. 분석 자료에서 판암1동의 아파트 주거비율은 83.1%, 백운1동은 57.4%로 높게 나타났다. 더욱이 아파트의 주거는 계속 증가하는 추세이다. 따라서 앞으로 조사구 설정에서 공동주택의 건축물대장 활용은 적극적으로 추진되어야 할 부분이다.

그러나 그 외의 경우는 지번에 건축물대장이 있거나 없거나에 상관없이 실제 조사를 통해야만 가구수를 파악할 수 있다. 어떤 지번에 건축물대장이 존재하지 않더라도 건물이 있을 수 있기 때문이며, 한 건물만 존재하더라도 그 건물에 여러 가구가 거주할 수 있기 때문이고, 여러 건물이 존재할 경우는 어느 것이 주택인지 부속건물인지를 알 수 없기 때문에 어차피 실제 조사를 해야 한다.

5. 건축물대장의 정보 활용

단독주택의 경우 거처정보와 건축물대장이 잘 매칭 되었기 때문에 두 자료의 정보가 얼마나 일치하는지를 파악할 수 있었다. 대전시 판암1동의 경우 거처정보의 연건평과 건축물대장의 동연면적을 대조해본 결과 차이가 ± 4 평 이내에 드는 거처가 22.3%(52/233), ± 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.

6. 행정자료 활용을 위한 제도적 개선방안

1) 건축물대장

이 연구에서 살펴본 주된 행정자료는 건축물대장으로 여러 가지 문제점을 발견할 수 있었다. 실제로 존재하는 건축물이 등록되지 않은 경우, 반대로 이미 허물

어졌는데 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 작업부터 시작되어야 할 것이다. 이보다 좀 더 구조적인 문제는 입력 내역에 대한 표준화 문제이다. 예를 들면 여러 필지를 묶어 한 건물이 들어선 경우 어떤 지번을 사용할 것인지에 대한 표준 지침이 있어야 할 것이다. 또한 건축물대장을 입력과 갱신에 대한 표준 지침이 있어야 한다. 변동이 발생할 경우 언제 입력할 것인지, 지역자치단체에서 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 정비하는 데 혼란이 있어 건축물대장을 정비하는 효과가 감소할 위험이 있기 때문이다. 이런 측면에서, 연구자들은 건물센서스(building census)를 제안한다.

이와 비슷한 이유에서, 수치지도에 대한 표준화 작업도 있어야 한다. 특히 지방자치단체에서 개별적으로 만드는 수치지도 사업이 여러 민간기업에 외주를 주어 진행되기 때문에 편차가 컸다. 또한 기준점, 지도 작성 방식, 갱신 방법 및 절차에 대해서도 구체적인 지침을 만들어 제시하고 교육하여야 할 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. 대부분의 일반 사람들은 행정동이 무엇이며 법정동이 무엇인지, 또 그 차이는 무엇인지를 알 수 없을뿐더러 알려고도 하지 않는다. 이러한 주소체계는 농업사회의 토지

소유를 근간으로 만들어져 이후 도시화 과정에서 일관된 원칙 없이 주소가 추가되면서 체계를 갖지 못했다. 그 결과 한 본번에 많은 부번, 한 지번에 여러 채의 건물 등이 있어 지번이 혼란스러울 뿐만 아니라 실제 이를 바탕으로 특정 장소를 찾기도 수월하지 않다. 또한 행정동과 법정동명이 서로 일치하지 않아 혼란도 있다.

이런 혼란에 대응하여 정부는 새주소체계를 만들기 시작했다. 새로운 주소체계는 주소를 부여하는 체계가 일관되어 있어 주소로 특정 장소를 찾기 수월하도록 도로와 건물 중심으로 편성되어 있다. 새주소체계는 토지와 건물을 분리하여 지번 중심의 구주소는 토지에만 사용하고, 건물에는 길 이름에 건물별로 번호를 부여하여 사용하는 것을 골자로 한다. 새주소의 일차적 목적은 찾기 쉬운 주소체계이다. 새주소체계에 부가적으로 발생하는 이점들이 몇 가지 있다. 1) 도로를 따라 건물마다 번호를 붙이기 때문에 자연히 건물의 고유번호가 만들어진다. 만약 건축물대장에서 구주소 대신 새주소를 사용한다면 기존의 지번보다 상세하고 정확한 건물에 관한 자료를 관리할 수 있다. 특히 건물별로 식별이 가능한 번호가 매겨지면 건물의 속성, 소유자, 사용자 등에 관한 건물 관련 정보를 건물별로 데이터베이스화하여 행정에 활용하거나 통계를 생산하기에 수월하다. 또한 새주소사업의 하나로 진행된 수치지도는 이러한 건물정보와 함께 사용하면 지리정보시스템(Geographical Information System) 기반의 행정서비스 뿐만 아니라 통계자료를 파악하기 쉽게 가공하고 자료를 시각적으로 확인하는 부수적 효과를 가질 수 있다.

그렇지만 새주소체계에서는 건물 안의 거처에 번호를 줄 수 있게는 되어있지 않다. 이런 측면에서 앞으로 통계적 목적으로 고려한다면 새주소체계는 적어도 거처수준까지는 파악이 될 수 있도록 확인 수준을 확대할 필요가 있다. 이와 함께 현 새주소시스템의 번호체계에서 사용하는 숫자가 너무 많다. 건물까지만 주소를 부여하는데도 27개의 숫자로 이루어져 있다. 여기에 거처 수준, 가구 수준까지 확대된다면 숫자의 단위는 족히 30개를 넘어설 것이다. 외국의 경우 가구 수준까지 파악할 수 있도록 주소체계가 되어 있음에도 덴마크는 18개, 핀란드는 23개의 숫자로 이루어져 있다. 새주소체계를 만들 때 번호체계에 대한 관심을 기울이지 않은 탓으로 보인다. 따라서 새주소체계의 번호체계를 새로 설계할 필요가 있다. 이번에는 통계적 목적을 고려하여 숫자의 단위를 가능하면 적게하는 것이 바람직하다.

참고문헌

- 건설교통부. 1999. 『건축물 대장의 기재 및 관리 등에 관한 규칙』. 건설교통부령 제189호.
_____. 2003. 『건축법 시행령』.
_____. 2004. 『건축물대장 정리 및 정비 요령』.
- 국토지리정보원. 2004. “1/25,000 수치지도/종이지도 및 1/50,000 종이지도 인덱스.”
<http://www.ngi.go.kr>.
- 김호철. 2002. “주택재고 및 공가분석.” 『한국의 주택』 (윤주현 편). 통계청
- 박현주. 2000. “새주소사업의 발전방향.” 『새주소사업의 발전방향 모색 토론회』.
국토개발원.
- 서울특별시. 2001. 『공간데이터웨어 하우스 구축 기본설계』.
_____. 2002. 『새주소 활용방안 연구』.
_____. 2003. 『서울시 정보기술아키텍처 구축』.
- 서울시정개발연구원. 2001. 『서울시 도시계획 정보관리시스템 기초데이터 구축연구』.
_____. 2002a. 『서울시 도시계획정보관리시스템 유지관리방안 연구』.
_____. 2002b. 『공간데이터웨어 하우스 구축 시범사업(조사)』.
- 안경애. 2004. “전국 건축물대장 DB 구축.” 『디지털 타임스』. 8월 30일.
- 양경진. 2003. “행정자료 활용 주택DB 구축 방안.” 『2005년 인구주택총조사 준비 제2차 워크숍』.
- 임명선. 2003. “2005년 총조사 조사구설정방법 개선방안.” 『2005년 인구주택총조사 준비 제2차 워크숍』.
- 통계청. 2002. 「2000 인구주택총조사 CD」.
_____. 2004a. 『2005 인구주택총조사 조사지침서』.
_____. 2004b. 『2005 인구주택총조사 기본계획안(요약)』.
_____. 2000a. 「인구주택총조사」. <http://kosis.nso.go.kr>.
_____. 2000b. 「농업총조사」. <http://kosis.nso.go.kr>.
_____. 2000b. 「어업총조사」. <http://kosis.nso.go.kr>.
- 행정자치부. 2004a. 「읍면동 기능전환 관련 시군구 이관사무 처리 편람」.

- 2004b. 「주민등록주소코드현황」. <http://www.mogaha.go.kr/>
- Ai, Pieter, and Bart Bakker. 2000. *Re-engineering Social Statistics by Micro-integration of Different Sources: an Introduction*. Netherlands Official Statistics.
- Bierau, Dieter. 2000. "Testing a Register-Based Census." Insee-Eurostat Seminar on Census after 2001. Paris.
- Borchsenius, Lars. 2000. "From a Conventional to a Register-based Census of Population." Insee-Eurostat Seminar on Census after 2001. Paris.
- Cassel, Paul and Eddy Craig. 2001. *Sams Teach Yourself MS Access 2002 in 21 Days*. Princeton, Indianapolis: Sams.
- Desplanques, Guy. 2000. "The New Census in France." Insee-Eurostat Seminar on Census after 2001. Paris.
- Dumais, Jean et al. 2000. "An Alternative to Traditional Census Taking: the Continuous Census of the Population." Insee-Eurostat Seminar on Census after 2001. Paris.
- Harala, Riitta and Jari Nieminen. 1998. *Use of the Building and Dwelling Register in the Production of Statistics*. Statistics Finland.
- Harala, Riitta and Reinikainen Anna-Leena. *Confidentiality in the Use of Administrative Data Sources*. Statistics Finland.
- Harder, Christian. 1997. *Arcview Gis Means Business*. Redlands, CA: ESRI Press.
- Hendricks, Coen, Paul Inge, and Harald Utne. 2000. "Establishing a Dwelling Register in Norway: Completing the Register System." Insee-Eurostat Seminar on Census after 2001. Paris.
- Hohl, Pat and Brad Mayo, 1997. *Arcview Gis Exercise Book*. Stamford, CT: OnWord Press.
- Jensen, Lothar. 2000. "Opening speech." Insee-Eurostat Seminar on Census after 2001. Paris.
- Laihonen, Aarno. 1999. "Development of the Use of Administrative Data in Population and Housing Censuses in Europe." Working Paper No. 6. Joint ECE/Eurostat Work Session on Registers and Administrative Records in Social and Demographic Statistics. Geneva.
- 2000. "2001 Round Population Censuses in Europe." Insee-Eurostat Seminar on Census after 2001. Paris.
- Lee, Jay and David W. S. Wong. 2000. *Statistical Analysis With Arcview GIS*. New York: John Wiley & Sons Inc.
- Olexa, Milan, and Alexandra Petrasova. 2002. "Administrative Data Sources - Their Quality and Application to Social Statistics and Demography." Working Paper No. 9. Joint ECE/Eurostat Work Session on Registers and Administrative

- Records in Social and Demographic Statistics. Geneva.
- Myrskylä, Pekka. 1999. "New Statistics Made Possible by the Use of Registers" Working Paper No. 8. Joint ECE/Eurostat Work Session on Registers and Administrative Records in Social and Demographic Statistics. Geneva.
- _____. 2000. "Effects of Changes to Registers: Experiences from Finnish Register-based Censuses." Insee-Eurostat Seminar on Census after 2001. Paris.
- Nordberg, Lief, Irmeli Penttilä, and Susanna Sandström. 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.
- Santa, József. 2002. "Towards the Use of Administrative Records and Registers for Statistical Purposes at the Hungarian Central Statistical Office." Working Paper No. 1. Joint ECE/Eurostat Work Session on Registers and Administrative Records in Social and Demographic Statistics. Geneva.
- Statistics Finland, 1999. "Use of Administrative Data in Population Censuses- Definition of Main Type of Activity as an Example." Working Paper. Joint ECE/Eurostat Work Session on Registers and Administrative Records in Social and Demographic Statistics. Geneva.
- Spieker, Finn. 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.
- Van der Laan. 2000. "The 2001 Census in the Netherlands Integration of Registers and Surveys." Insee-Eurostat Seminar on Census after 2001. Paris.

부록 : 서울시의 행정자료 활용 사례 - 건물·주거 관련

인구센서스에서 조사구 설정 등에 행정자료를 이용하기 위해서는 건축물대장, 토지대장(건설교통부, 지차자체 생산), 새주소시스템(행정자치부, 지자체 생산) 등 이미 생산되고 있는 공간자료(spatial data)를 활용할 수 있어야 한다. 현재 건축물의 공간적 양태를 알려주는 행정자료는 건축물대장, 토지 대장, 과세 대장, 주민등록자료 등이 있다. 이 자료들은 생산부서와 사용용도에 따라 다른 목적으로 만들어져 개별 자료를 연결하여 하나의 통합 데이터베이스(database)로 만드는 작업이 쉽지 않다. 더욱이 개별 자료가 현실을 그대로 완전하게 반영하지 못하기 때문에 자료의 활용은 어려울 수밖에 없다.

서울시는 그러한 어려움을 극복해보고자 몇 가지 시도를 했다. 여기서는 서울시의 그러한 노력의 하나로 만들어진 서울시 공간데이터웨어하우스(spatial data warehouse)를 살펴봄으로써 인구센서스에서 건축물대장을 활용하는 데 시사점을 찾아보고자 한다. 먼저 서울시 공간데이터웨어하우스가 무엇이며, 이 사업이 현재 어떻게 진행되고 있고, 데이터베이스들간 연동은 어떤 방식으로 이뤄지는 지를 중심으로 살펴본 다음, 인구센서스와 직접적 관련이 있는 자료인 서울시 주택자료와 건축물 자료의 현황과 특성을 구체적으로 서술한다. 이후 서울시의 행정자료 데이터베이스 연동현황, 공간자료 데이터웨어하우스 구축 경험이 통계청의 인구주택총조사에 주는 함의를 찾아보고자 한다.

1. 서울시 공간데이터웨어하우스

서울시에서는 1995년 수립된 「서울시 지리정보시스템 구축 기본계획」에 따라 각종 공간자료를 생산해 왔다. 수치지도(digital map), 항공사진 이미지 데이터 등 기본도 및 부서별로 도시계획정보관리시스템, 도로관리시스템, 상수도시설관리시스템, 하수도전산관리시스템, 소방방재시스템 등 단위시스템이 구축되었다.

그러나 실·국별로 부서별 사용을 중심으로 구축된 지리정보시스템은 연결체계가 미흡하며, 최신의 데이터를 취득하기가 쉽지 않고(시스템 표준화, 데이터베이스의 표준화 등), 각 시스템에 구축되어 있는 데이터의 소재와 구체적 내용 파악이

어려웠다. 더욱이 분산적 데이터베이스 개발은 중복구축과 이에 따른 예산낭비의 문제를 일으켰다. 그 결과 서울시의 공간자료시스템은 다양하게 활용되지 못하였다.

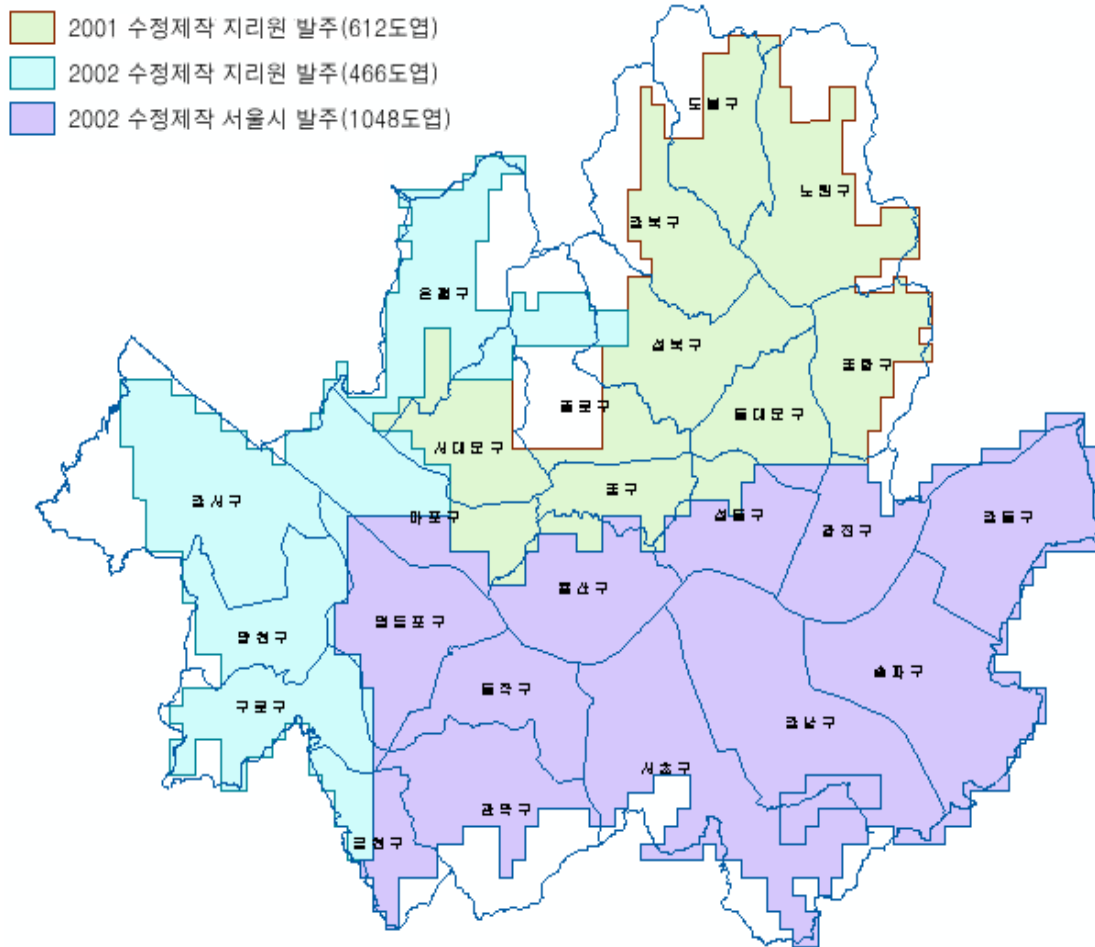
서울시는 이러한 문제를 해결하고 데이터의 공유와 활용성을 증대시키기 위해 모든 부서에서 공통으로 사용하는 데이터를 중앙에서 관리하고, 네트워크로 이 GIS DB 들을 연결하는 방향으로 서울시는 2001년부터 공간데이터웨어하우스 (Spatial Data Warehouse)를 구축하였다. 실·국별 다양한 GIS 데이터의 내부적 공유체계와 부서들 사이의 외부적 유통체계를 확립하기 위해 2001년에 「공간데이터웨어하우스 기본계획 설계」 연구, 이 기본계획을 바탕으로 2002년에는 시범조사 사업을 수행하였다. 2003년부터 본격적인 운영계획을 수립하여, 현재까지 SDW를 운영하고 있다. 이제 SDW 서울시 공간자료의 기반자료로서의 중요한 역할을 수행하고 있다.

SDW는 공간자료로 서울시의 수치지도에 기반을 두고 있다. 수치지도란 지도에 표현된 지형·지물 등 도형자료와 지명·수치 등 속성자료를 컴퓨터에서 처리할 수 있는 디지털 자료(digitalized)로 변환한 지도이다. 그림과 문자를 컴퓨터에서 처리하기 위해서는 점, 선, 면과 같은 좌표와 기하학적 요소의 형태로 표현하여 선추적형으로 변환하여 수치자료 좌표(X, Y, Z)로 입력한 지도이다. 서울시에서 작성한 수치지도는 1/5000 수치지도와 1/1000 수치지도가 있다.

서울시의 1/5000 수치지도는 국가지리정보시스템(National Geographical Information System) 사업의 일환으로 1997년에 제작 완료된 것이다. 내용과 정확도에서 그다지 훌륭한 품질은 아니지만 1/1000 수치지도가 포괄하지 못하는 산림이나 서울시 외곽부를 포괄하고 있는 유일한 수치지도라는 점에서 유용하다. 레이어(layer)는 1/1000 수치지도 정위치 자료와 비슷하나 좀 더 세분화되어 있다. 대략 항공사진 측정은 94년부터 96년 사이에 이루어졌으며, 현지조사는 95년에서 97년 사이에 시행되었다. 한 가지 문제점은 다른 도면과 불일치인데, 1/1000 수치지도의 동일 지역과 중첩시킬 경우 지형/지물이 일치하지 않는 경우가 많다.

서울시의 1/1000 수치지도는 서울시 지리정보센터에서 2003년 말에 완성한 지도로 서울시의 대부분을 포괄하고 있는 지도이다. 386개의 레이어를 갖고 있으며, 1996년 처음 수치지도화 작업이 시작되어 2001년부터 2003년까지 계속 수정·보완되었다. 현 수치지도는 서울시 전 지역을 포괄하지 못하는데, 그러한 지역은 대부분이 산악지역으로 필요할 경우 1/5000 수치지도와 연계하여 작업한다. 수

치지도에 표기된 지적은 도시계획도(1/3000)의 지적을 사용하나 실제와 다르기 때문에 참조용으로만 사용한다. 또한 수치지도의 행정동과 법정동 경계가 잘못 구축된 곳이 많다. 따라서 이럴 경우 다른 행정구역도(구경계도, 행정동경계동 등)를 사용해야 한다. 기본도형은 <그림 부록-1>과 같다.



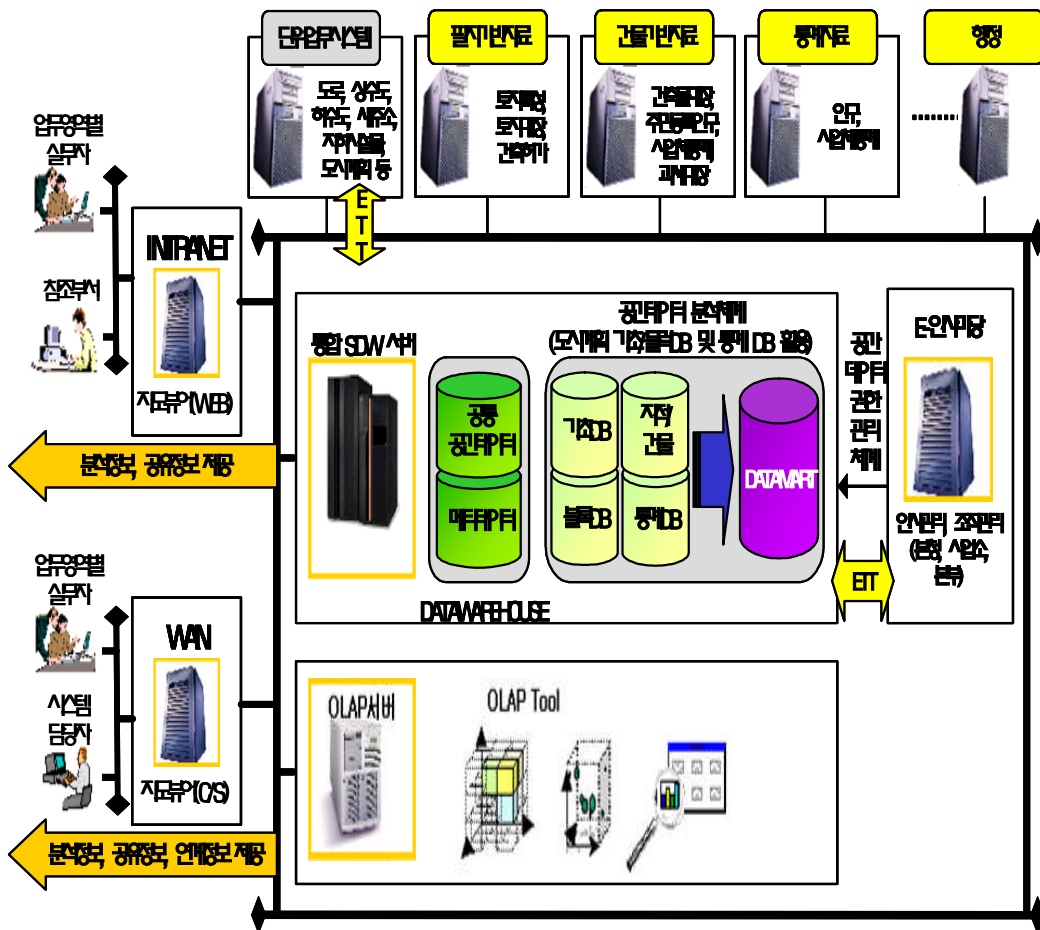
<그림 부록-1> 수치지도 : 서울시 기본지도

출처 : 서울시정개발연구원 연구정보시스템(ris.sdi.re.kr)

SDW을 구축하기 위해서는 기본적인 시스템 구조가 만들어져야 하며, 그 다음 단계는 어떤 데이터를 담을 것인가가 결정해야 한다. 서울시에서는 부서별 업무에 활용도가 높은 핵심·공통 데이터를 저장·관리하여 사용자의 요구 및 업무상 필요시 효율적으로 데이터에 접근할 수 있는 체계를 갖추는 것이 필요하다는 판단을 하였다. 아련 판단에서 서울시는 GIS 메타데이터를 이용하여 핵심이 되는 공통

데이터를 지정하고, 단위 업무시스템에 구축되어있는 모든 GIS 데이터에 대한 검색을 수행할 수 있게 SDW를 구축하였다. 더 나아가, 서울시의 GIS 데이터를 연결할 뿐 아니라 국가수준의 국가지리정보유통기구와도 연결되어 공간데이터의 총괄적인 관문(gateway)의 기능을 수행하도록 SDW 구축의 목적을 설정하였다.

SDW 시스템 구성도는 <그림 부록-2>에 제시하였다. SDW에 포함되어 있는 데이터는 크게 기초/블럭 데이터와 공통데이터로 구분된다(<표 부록-1>). 기초/블럭 데이터는 건물, 토지, 인구, 사업체 등 서울시에서 생산되는 기본적인 데이터와 통계청의 조사구 자료를 포함하고 있으며, 공통데이터는 여러 부서에서 공통으로 참조하는 데이터를 포함하고 있다.



<그림 부록-2> SDW 시스템 구성도

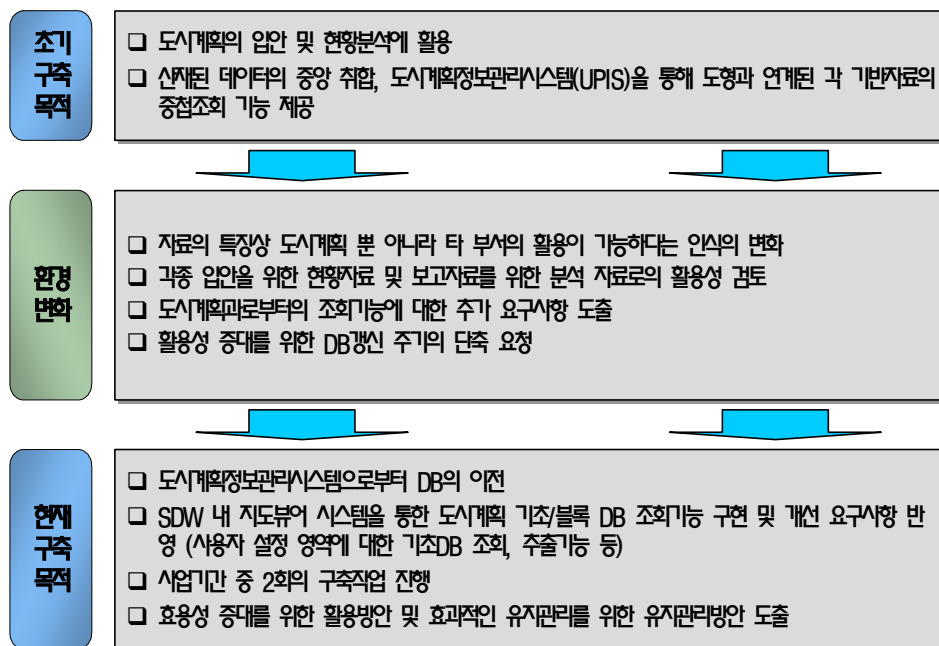
출처 : 서울시 GIS개발팀 내부자료, 2004

SDW에서 공통레이어로 선정된 데이터 목록은 5개 이상 부서에서 참조하는

레이어로 임 구축된 881개 레이어에서 선정된 것이다. 공간데이터 외에 사용빈도와 활용도가 높은 기초블록DB의 경우 필지기반 데이터와 건물기반 데이터를 선정하였다. 이와 함께 이 데이터베이스와 연계될 수 있는 대장이나 조서도 선정하였다.

모든 단위업무시스템의 데이터베이스를 SDW DB에 전체적으로 연결시키지 않았다. 따라서 SDW 안에서 데이터의 연동은 먼저 하나의 단위업무시스템에서 특정 데이터가 변경되면, 이를 추출하여, SDW에 전송함으로써 SDW의 해당 데이터가 변경된다. 이렇게 SDW DB가 바뀐 후 이 데이터가 필요한 다른 단위업무시스템에 데이터를 전송하여 전체적으로 일관된 DB를 생성한다. 이 과정을 통해 데이터 생산 부서에서 단위시스템 갱신과 SDW DB의 갱신이 이뤄진다.

서울시의 SDW 데이터베이스 중 통계청에서 활용과 밀접한 부분이 기초/블럭 DB이다. 기초/블럭DB 구축과정에서 몇 단계의 변화를 겪었다. 초기에는 도시계획 정보관리시스템을 통해 도형과 연관된 기반자료의 중첩조회가 가능한 시스템을 구축하고자 했다. 그러나 이후 SDW의 블럭DB가 도시계획에 국한되는 것이 아니라 타 부서에서도 다양하게 활용될 수 있다는 인식과 함께 데이터베이스의 갱신주기를 단축하게 되었다. 또한 데이터베이스가 SDW로 이전되면서 데이터 조회 기능 뿐 아니라 추출 기능까지를 포함하게 되었다. <그림 부록-3>은 SDW 기초/블럭 DB의 변화과정을 나타낸 것이다.



<그림 부록-3> SDW 시스템 내 기초/블럭 DB 구축의 변화과정

SDW의 기초/블럭DB에 포함된 내용을 살펴보면, 건물기반 자료, 필지기반자료, 행정동 기반자료, 기타 자료 등이 있다. 건물기반자료에는 건축물대장, 과세대장, 사업체기초통계조사자료, 주민등록자료 등이 포함되며, 필지기반 자료는 토지특성과 대장, 건축허가자료 등을 포괄한다. 행정동 기반 자료는 연도별 행정동 도형자료(shape 형태), 인구주택총조사 자료 등이 포함되어 있으며, 공통데이터에는 통계청의 조사구 도형자료, 새주소건물도형자료 등이 포함되어 있다(<표 부록-1>, <표 부록-2>).

<표 부록-1> SDW 데이터 목록

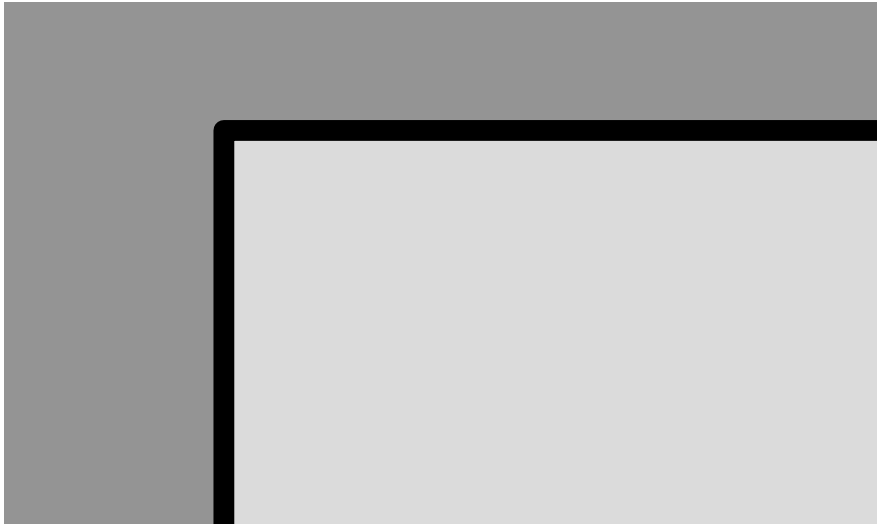
데이터 구분	레이어명 및 대장/조서명	
기초/블럭 데이터	토지대장, 토지특성자료, 건축허가자료, 건물도형, 건축물대장, 주민등록 인구자료, 과세대장, 사업체기초통계조사, 행정동 도형, 인구주택총조사	
공통데이터 (52개 레이어)	수치지도	시경계, 구경계, 행정동경계, 법정동 경계, 전력전신주, 철도, 철도중심선, 지하철역사, 성절토상단, 성절토하단, 석축상단, 석축하단, 옹벽상단, 옹벽하단, 하천경계, 하천 중심선, 복개하천중심선, 호수/저수지, 댐, 보, 제방상단, 제방하단, 수문, 잔교, 선착장, 등고선, 표고점, 색인도 (1/1000), 색인도(1/5000), 통신관로
	새주소	건물
	도로관리 시스템	도로시설물(면), 맨홀, 포장, 도로부속물, 도로중심선, 도로경계선, 도로면
	상수도	상수도관로, 소화전
	하수도	하수관거, 배수구역, 배수분구, 처리구역, 오수토실, 오수 중계펌프장, 하수맨홀
	지하시설 물	가스배관, 난방열배관, 송전-지중관로, 배전-지중관로, 통신관로
	도시계획	용도구역, 용도지구, 용도지역
	토지	편집지적
	항공사진	항공이미지사진
	외부자료	지하철 노선

<표 부록-2> SDW 기초/블록 DB 상세 내용

구분	세부내용
건물기반 자료	<ul style="list-style-type: none"> ▪ 건축물대장 ▪ 과세대장 ▪ 사업체기초통계조사 ▪ 주민등록인구조사
필지기반 자료	<ul style="list-style-type: none"> ▪ 토지특성 자료 ▪ 토지대장자료 ▪ 건축허가자료
행정동 기반 자료	<ul style="list-style-type: none"> ▪ 연도별 행정동 도형 ▪ 연도별 인구 및 주택총조사자료
기타	<ul style="list-style-type: none"> ▪ 블록별 집계 데이터 ▪ 통계청 기초단위구(조사구) 도형 ▪ 편집지적도 ▪ 새주소 건물도형자료

SDW 기초/블록 데이터베이스 구축에서 통계청 조사구와 관련 있는 공간자료는 블록기반 집계자료이다. 여기서 블록이란 통계청의 조사구 3-4개 정도를 더한 영역을 의미한다. 기초/블록 DB의 내역은 <표 부록-2>에 상세히 제시하였다. 기초/블록 DB 유지관리는 건물기반자료 유지관리와 행정동 기반자료 유지관리로 나뉜다. 건물기반자료 구축 유지관리 방안으로는 건축물대장에 ID 부여하여 새주소 도형자료와 연결하고 이를 건물도형과 연결하여 건물기반자료를 구축한다. 행정동 기반자료와 통계자료는 통계청의 조사구 도형자료를 기반으로 블록 데이터베이스를 구축하며, 여기에는 인구와 통계관련 DB를 붙인다(<그림 부록-4>).

SDW에서는 이와 같은 시스템 개발을 통해 DB연동체계를 구축했으며, 현재 서비스되고 있는 응용시스템은 지도뷰어 시스템으로 고급사용자를 위한 C/S 와 일반사용자를 위한 WEB 서비스가 있다(<그림 부록-5>). 고급사용자와 시스템 관리자를 대상으로 한 C/S 지도뷰어(map viewer) 어플리케이션은 기초/블록 DB 조회 기능이 있으며, 데이터베이스를 추출할 수 있다(<그림 부록-6>).

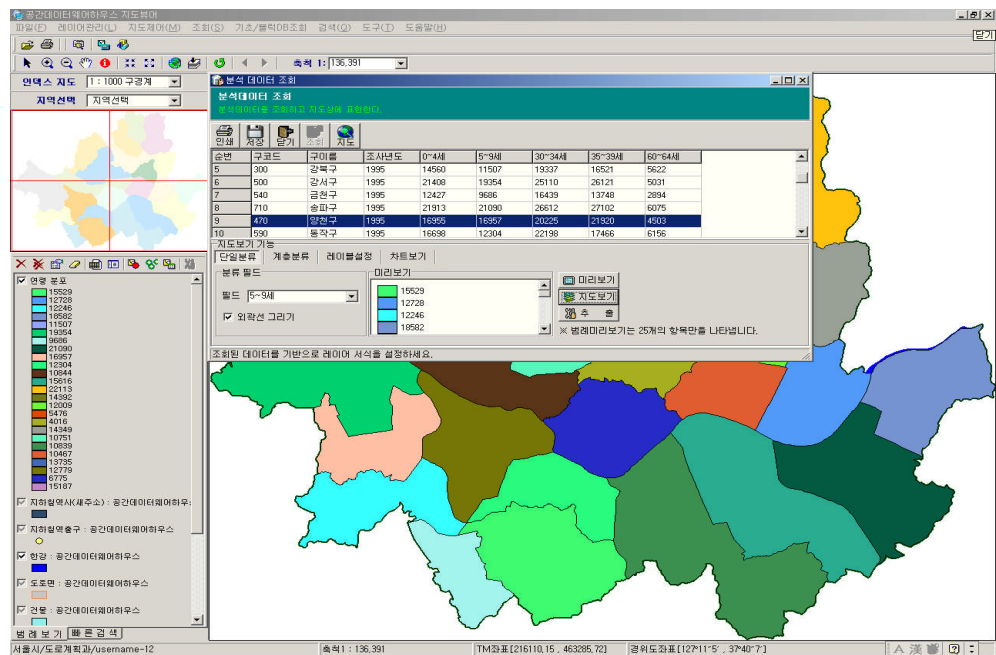
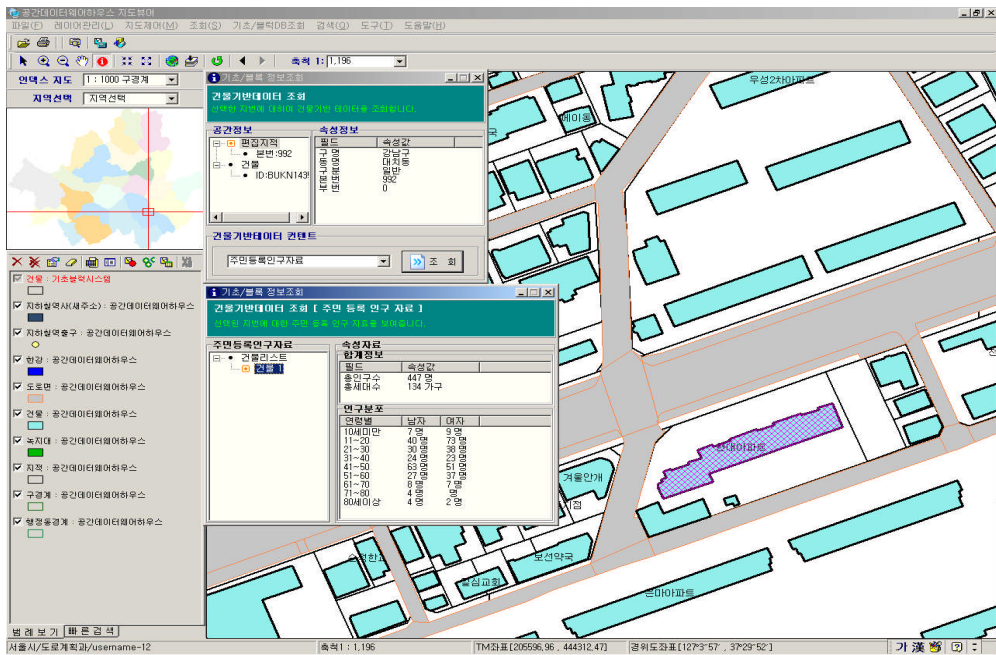


<그림 부록-4> SDW 기초/블록 DB 구축 현황

출처 : 서울시 GIS개발팀 내부자료(2004)

<그림 부록-5> SDW 기초/블록 DB 유지관리 방안

출처 : 서울시 GIS개발팀 내부자료(2004)



<그림 부록-6 > 지도 뷰어 어플리케이션 화면

2. 서울시의 주택 및 건물 관련 자료

서울시의 주택 수는 인구주택총조사에서 파악되며, 나머지 연도에는 추정치를 활용한다(<표 부록-3>). 그러나 통계청의 인구주택총조사의 자료는 제공단위가 구별

이어서 동별 자료는 사용할 수 없다. 서울시 주택기획과와 주택재개발과, 한국주택은행이 다양한 주택관련 표본조사를 수행하고 있으나 조사의 표본규모가 작기 때문에 여전히 동 단위 자료를 확보하기 어렵다. 또한 조사주관 부서와 조사목적에 따라 조사대상이나 집계기준이 달라 사용하기 어렵다. 과세자료로도 주택 수를 추정할 수 있으나 비과세건물이 누락되기 때문에 여전히 문제는 남는다.

<표 부록-3> 주택자료 원천과 특징

조사기관	원시자료	특징
통계청	인구주택총조사	전수조사이나 구별단위로 취합
주택기획과	내부자료	주택건설촉진법에 의해 승인된 건축물(자치구 집계)
주택재개발과	내부자료	도심재개발법에서 승인된 건축물, 자치구 단위 집계
한국주택은행	도시주택가격동향조사	샘플조사, 집계단위가 강북/강남

이렇듯 직접적인 조사통계로 주택 수를 파악하기 어려울 경우, 여러 가지 자료를 함께 사용하여 추정하기도 한다. 일반적으로 주택과 건물에 대한 자료는 인구주택총조사, 재산세 과세자료, 건축물대장을 사용한다(<표 부록-4>). 주택 수는 일단 건물 수를 파악하고, 각 건물별로 층수, 층별 호수를 확인하여 추정할 수 있다. 건물 수는 건축물대장을 이용하여 파악할 수 있다. 건물 층수는 건축물대장과 과세대장에서 파악할 수 있다. 그러나 건축물대장은 해당 자료의 누락률이 높아 실질적인 이용의 어려움이 있으며, 과세대장은 과세 제외 대상건물 및 층수항목 기재 누락이 많아 사용하기 어렵다.

<표 부록-4> 주택 및 건물 분야 분석을 위한 필수데이터

필수데이터		자료원
주택	동별 주택	인구주택총조사 / 건물분 재산세 과세자료
	용적률	건축물대장 / 건물분 재산세 과세자료
건물	동별 건물 수	건축물대장 / 건물분 재산세 과세자료
	용적률	건축물대장 / 건물분 재산세 과세자료
	층수	건축물대장

이럴 경우 수치지도를 함께 활용하여 주택 수를 추정할 수 있다. 수치지도에 서는 자료 구축 시점의 차이로 알고자 하는 시점의 완전한 자료를 얻기 어렵고 해당란의 누락이 많아 자료의 한계가 있다. 그렇지만 공간자료를 시각적으로 얻기가 쉬우며, 지도가 구축된 시점에서는 자료의 신뢰성이 높아 여타 자료에 비해 가장 많이 활용되고 있다. 따라서 건축물대장과 수치지도의 건물도형을 연결시켜 공간자료와 속성자료를 확보하는 것이 현재로는 가장 적절한 방법이다.

서울시 건축물대장은 행정자치부 건축물대장 전산화 기본계획(1997-2001)에 의해 건축물대장 전산화가 2000년도에 완료되었으며, 건축물대장 관리 및 발급을 위한 시스템이 구청 지적과에 설치·운영되어 있다. 또한 시·구청 건축과 및 주택과에 모두 보급되어 있는 건설교통부 건축행정정보시스템은 건축주택관련 행정업무 뿐 아니라 건축사 사무소의 인허가 신청 및 지적과의 건축물대장발급 등 민원업무와 건축허가에서 준공까지의 업무를 처리하고 있다.

행정자치부의 전산화된 건축물대장을 건설교통부 건축행정정보시스템으로 변환하는 작업이 2002년 7월에서 2003년 2월까지 진행되었으며, 2003년 말까지 데이터베이스 변환작업이 완료되었다. 은평구와 강남구에는 건설교통부 건축행정정보시스템이, 나머지 자치구에는 행정자치부 건축물대장 전산화 자료를 기반으로 건축물대장이 구축되었으나 2004년 현재 25개 자치구가 건설교통부 건축행정정보시스템으로 통합되어있다. 건축행정정보시스템은 본청의 별도 통합DB서버가 없이 각 자치구에 서버를 두고 직접 접속·관리하는 체계로 운영된다.

3. 서울시 공간데이터웨어하우스 구축과정의 경험과 시사점

여기서는 서울시 공간데이터웨어하우스의 기초/블록 DB 구축을 중심으로 데이터 오류의 발생원인과 이의 해결방안을 살펴보고자 한다. 먼저 데이터 오류가 발생하는 일반적인 논의를 살펴보고, 데이터오류가 발생하는 몇몇 사례들을 검토해본다. 그런 다음 SDW의 기초/블록 DB의 데이터연결과정을 살펴본다.

3) 데이터 오류

일반적으로 데이터 오류의 발생원인은 정합성(logical consistency) 오류와 기계적 오류로 구분된다. 정합성 오류는 입력 오류, 자료갱신 오류, 법적 기준과 현실이 너무 동떨어져 나타나는 오류 등으로 구분되며, 기계적 오류는 시스템의 오류이다(<표 부록-5>).

<표 부록-5> 데이터 오류발생 원인

구분	오류 원인	오류발생 내용	오류발생 장소
정합성 오류	입력오류	입력 작업자의 입력 오류 데이터 편집과정의 오류	단위업무시스템
	자료갱신 오류	데이터 갱신절차(주기, 방법 등) 차이에 따른 오류	
	법적기준 과 현실 상이	법적기준이 현실에 적합하지 않아 발생하는 오류	단위업무시스템 공간데이터웨어하우스
기계적 오류	시스템 오류	데이터 변환과정에서 데이터 손실, 불량, 데이터 전송과정에서 손실 등 오류	단위업무시스템 데이터웨어하우스

출처 : 서울시정개발연구원(2002)

정합성 오류는 지도에서 도로 위에 건물이 위치하는 등 공통데이터 간 데이터 오류가 발생한다. 여기에는 데이터를 입력하는 작업자에 의해 발생하는 입력오류, 단위업무시스템별로 데이터 갱신절차(주기, 방법 등)의 상이에 따른 오류인 자료갱신 오류, 데이터는 실제 지형지물과 동일하지만 법적기준에 적합하지 않아 발생하는 오류인 법적기준과 현실이 상이한데서 발생하는 오류 등이 있다. 데이터간 정합성 오류 검사를 위해서는 우선 각각의 단위업무시스템에서 데이터 내부에 존재하는 오류(입력오류, 갱신절차 상이 등에 따른 오류 등)들이 수정되어야 하며, 이러한 오류들은 단위업무시스템 구축 시 데이터 유지관리 차원에서 해결되어야 한다.

4) 사례

(1) 건물도형과 건축물대장의 연결사례 (서울시, 2002)

서울시 새주소사업팀은 새주소사업을 하면서 건축물대장과 건물번호를 연계하는 작업을 시도했다. 도로명과 건물번호 부여사업에서 건물번호가 부여된 건물도형과 건축물관리대장을 연결하기 위해서는 건물도형과 지번을 연결해야 한다. 이를 위해 토지관리정보체계에서 구축한 연속지적도를 이용하여 연속지적도의 지번과 건물번호를 연결하고, 이에 다시 건축물대장과 연결하였다. 건물도형과 지번을 전산프로그램으로 자동 연결한 결과 건물도형과 지번(건축물 관리대장의 관리번호)이 1:1로 연결되는 경우는 39.1%에 불과하며, 나머지는 1:N, N:1, N:N로 연결되었다(<표 부록-6>).

<표 부록-6 > 건물도형과 건축물대장 연결 형태

건물 : 대장	1 : 1	1 : N	N : 1	N : N
비율(%)	39.1	1.9	9.7	49.4

연계율은 구별로도 다른데, 이를 파악하기 위해 종로구과 구로구에 대해 현장 조사를 하였다. 지번을 건축물대장과 연결해 본 결과 81.3% 정도 지번이 맞는 것으로 나타났으며, 건물지번이 확인이 안되거나 건축물대장과 개별 연결을 확인할 수 없는 경우가 20% 정도 였다. 또한 서울시 도시계획정보관리시스템 구축사업에서 건축물대장과 건물도형 연계율을 현지조사를 통해 살펴본 자료가 있다. 이때 강남구 역삼동, 종로구 창신동, 서대문구 창천동 등을 시범조사 하였는데, 건축물대장과 건물도형을 연계할 수 없는 비율이 강남구 역삼동 11.9%, 종로구 창신동 24.8%, 서대문구 창천동 17.6%로 나타났다(<표 부록-7>).

<표 부록-7> 건축물대장과 건물도형의 연계율

연계항목 동	건물물대장수	개별연계	지번 ≙ 편집지적	미연계	미연계율*
강남구 역삼동	4,637	4,084	30	523	11.9
종로구 창신동	3,042	2,286	227	529	24.8
서대문구 창천동	1,696	1,396	89	211	17.6

* 미연계율은 지번과 편집지적이 맞지 않는 경우와 미연계 수를 합한 것을 건축물대장수로 나눈 것

(2) 도로와 건물 데이터의 중첩 사례 (서울시정개발연구원, 2002)

공통데이터의 정합성 오류 원인을 분석을 위해 「도로관리시스템」에서 생성된 「도로면 데이터」와 「새주소전산통합시스템」에서 생성된 「건물 데이터」를 중첩하였을 경우의 오류발생 원인을 살펴보았다. 이 두 시스템의 도로 데이터와 건물 데이터는 다음과 같다.

- 도로 및 도로면 데이터 : 도로관리시스템의 도로면 레이어, 1/1,000 수치지도의 도로(AD001) 레이어
- 건물 데이터 : 새주소전산통합시스템의 건물 레이어, 1/1,000 수치지도의 건물(AA001) 레이어

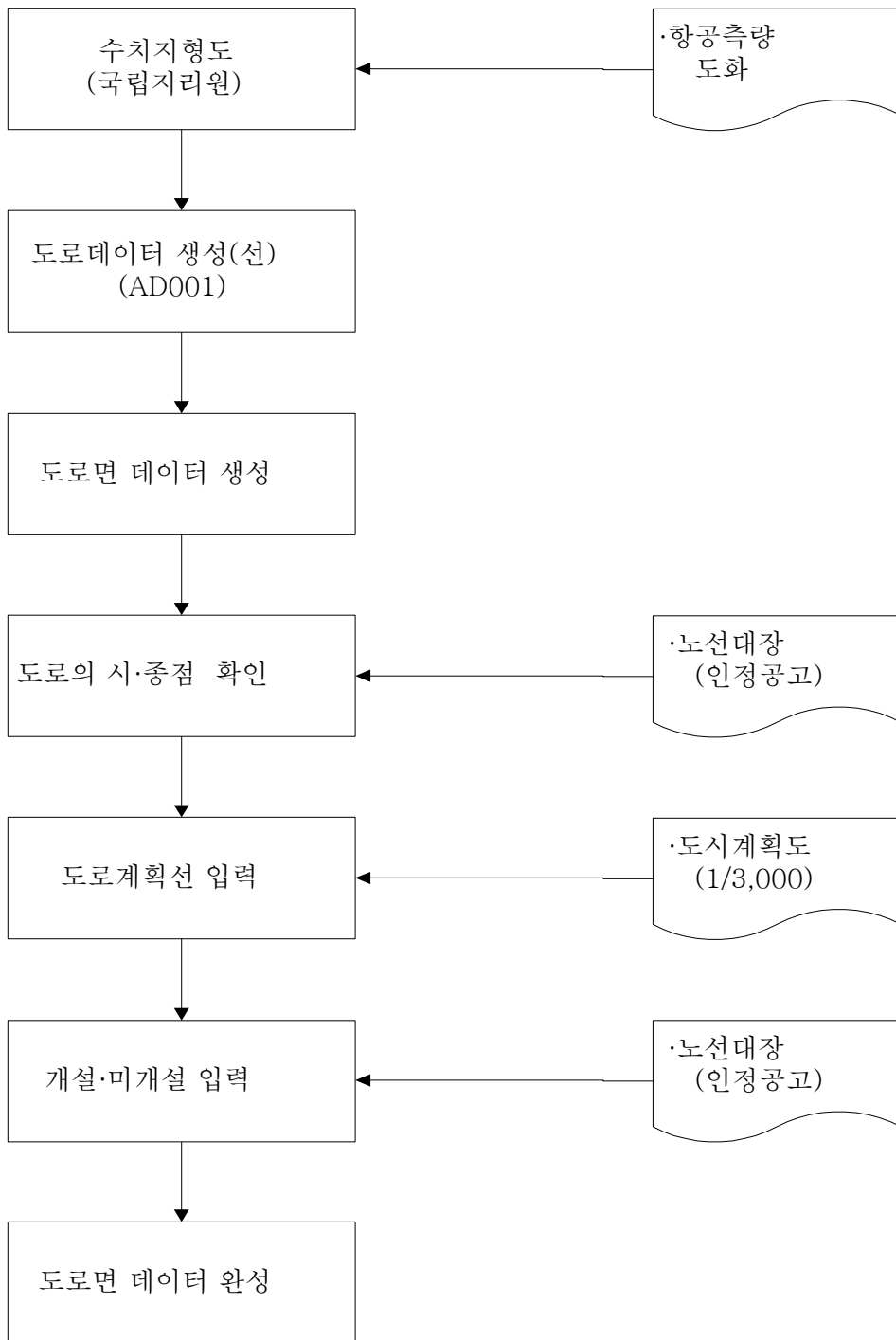
앞에서 논의한 바와 같이 정합성 오류는 입력오류, 갱신오류, 법적 기준과 현실의 괴리에서 발생한다. 그 가운데 입력오류는 자명하며, 법적 기준과 현실의 괴리는 그러한 경우가 발생할 때마다 나타나는 것으로 특별한 설명이 필요 없을 것이다. 이에 반해 갱신오류는 갱신절차와 갱신방법이 얼마나 합리적으로 되어 있는가에 따라 달라질 수 있다. 서울시에서는 이러한 이유에서 갱신방법과 갱신절차를 마련하여 적용하고 있다. 아래 <표 부록-8>는 자료 갱신방법에 관한 것이며, <그림 부록-7>과 <그림 부록-8>은 각각 도로관리시스템의 도로면과 새주소전산시스템의 건물 갱신과정에 대한 흐름도이다.

<표 부록-8> 도로면과 건물 자료 갱신방법

단위업무 시스템	구분	도로면	건물
도로관리 시스템	정비대상	인정 공고가 난 도로	도로 개설 등에 따른 신규 및 변경 데이터
	도형갱신 방법	노선대장의 인정공고에 등재 된 도로의 시종점 및 개통, 미개통을 도시계획도를 이용 하여 확인 후, 도로계획선을 디지털화(<그림 부록-7> 참조)	수치지도상의 건물을 수정
	기타	실측 없이 대장과 도면을 이 용한 실내작업	
새주소 전산통합 시스템	정비대상	건물 신축 및 변경 등에 따 른 신설 및 변경 데이터	1998년 1월1일 이후 건축물 관리대장에 신규 등재된 모 든 건축물(주)
	도형갱신방 법기준점	자치구에서 원시자료 확보가 어려움에 따라 본청에서 도 로의 수정을 통합관리하며, 자치구에서 필요한 원시자료 를 첨부하여 처리요청이 있 는 경우 본청에서 처리	건축물관리대장에 첨부된 건 축물 현황도 전산화(지적경 계선 및 건물), 편집지적 및 연속지적의 지적경계선 좌표 를 참조하여 건물데이터작성 * (<그림 부록-8>참조)
	기타	교량, 고가차도, 지하차도, 터널 등 도로에 부가되는 레이어도 도로의 일부로 관리	

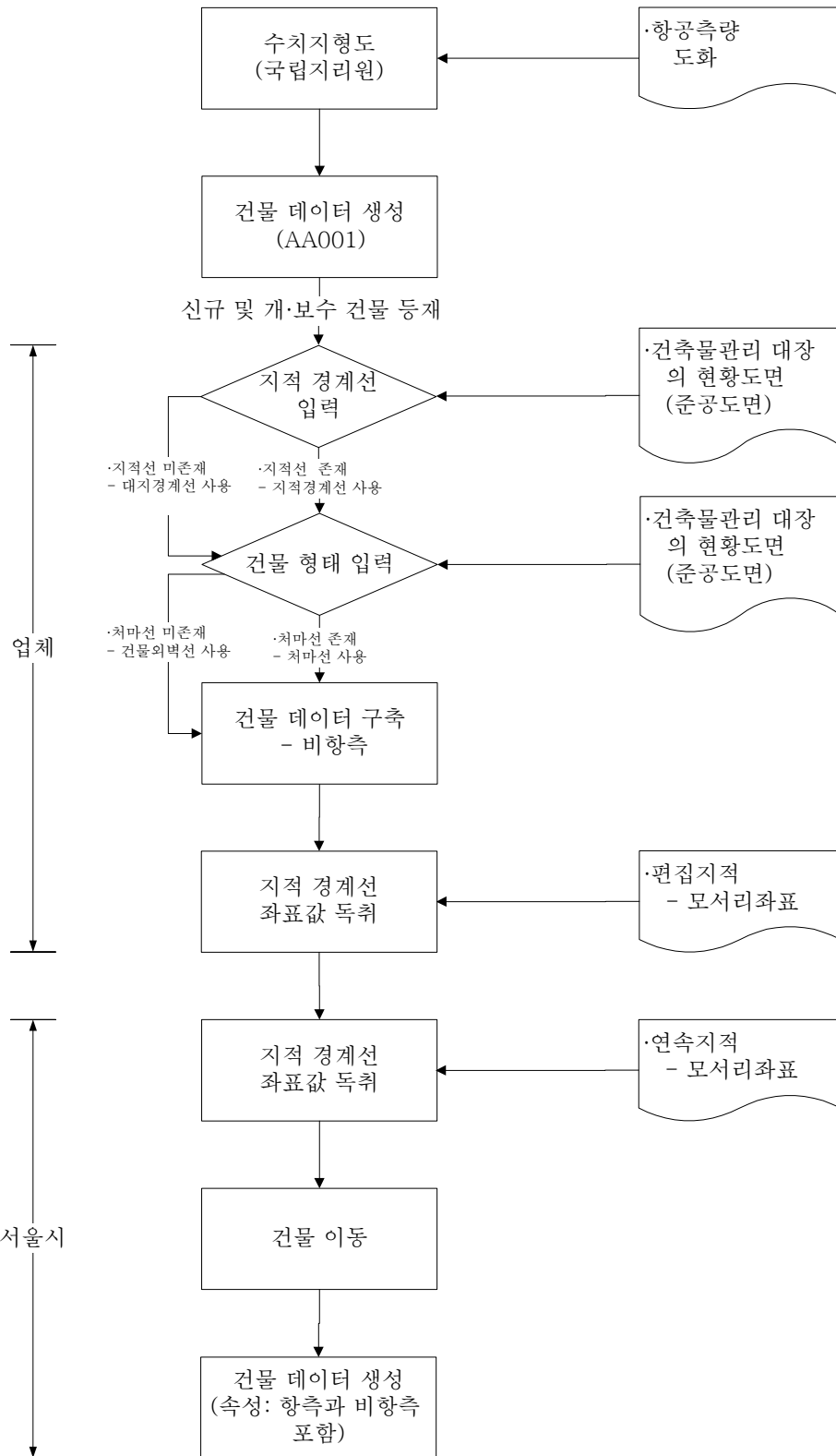
* 1997년 12월 5일 건축물관리대장의 기재 및 관리 등에 관한 규칙 개정 :

건축물관리대장에 건축물 현황도 첨부



<그림 부록-7> 도로관리시스템 도로면 갱신과정

출처 : 서울시정개발연구원(2002)



<그림 부록-8> 새주소전산통합시스템 건물 갱신과정

출처 : 서울시정개발연구원(2002)

중구와 송파구를 선정하여 새주소전산통합시스템의 건물 데이터와 도로관리 시스템의 도로면 자료를 중첩시켜 일반적 공간관계(면대면) 검사하였다. 그 결과 두 데이터 레이어의 중첩에서 데이터의 입력 오류 및 갱신절차 상이 등에 의한 오류가 발생하고 있으며, 이들에 대한 구분이 불가능하였다. <표 부록-9, 10>은 새주소전산 통합시스템의 건물 레이어와 도로관리시스템의 도로면 레이어의 중첩 결과 나타난 오류의 양과 유형이다. 여기서 오류는 입력 오류 및 갱신절차상이 등에 의한 오류의 합을 의미한다.

<표 부록-9> 건물 및 도로면 레이어의 정합성 오류 집계 -중구, 송파구

구청	오류비율(%) (오류건물수 / 총건물수)	총 오류면적(m ²)	호당 오류면적(m ²) (총 오류면적/오류건물수)
중구	2,688/25,047 = 10.73	10,797	4.02
송파구	583/29,283 = 1.99	4,746	8.14

출처 : 서울시정개발연구원(2002)

<표 부록-10> 건물·도로면 레이어의 정합성 오류 유형-중구, 송파구

오류 유형	오류 내용	비고
• 건물은 갱신되었으나 도로면은 갱신되지 않은 경우	• 건물과 도로면이 겹칩 • 건물이 도로면에 의해 분리된 경우	<그림 부록 9-1>
• 도로면은 갱신되었으나 건물은 갱신되지 않은 경우	• 건물과 도로면이 겹칩 • 건물이 도로면에 의해 분리된 경우	<그림 부록 9-2>
• 건물과 도로면이 모두 갱신된 경우	• 건물과 도로면이 겹칩 • 건물이 도로면에 의해 분리된 경우	<그림 부록 9-3>
• 갱신은 되지 않았으나 건물과 도로면의 미세한 차이로 건물의 일부가 도로면을 침범한 경우	• 건물과 도로면이 미세하게 겹치는 경우	






출처 : 서울시정개발연구원(2002)



(a) 1/1,000 수치지도의 건물과 도로면

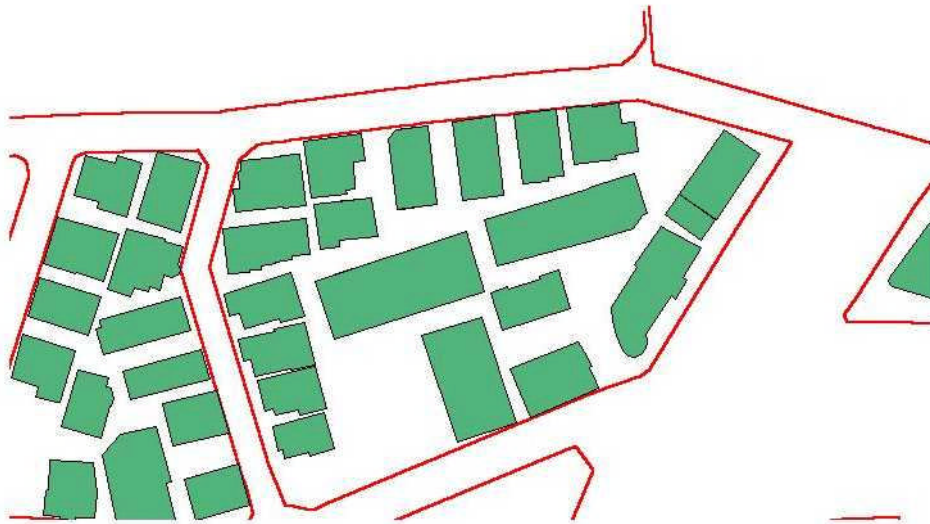


(b) 새주소전산통합시스템의 건물과 도로관리시스템의 도로면

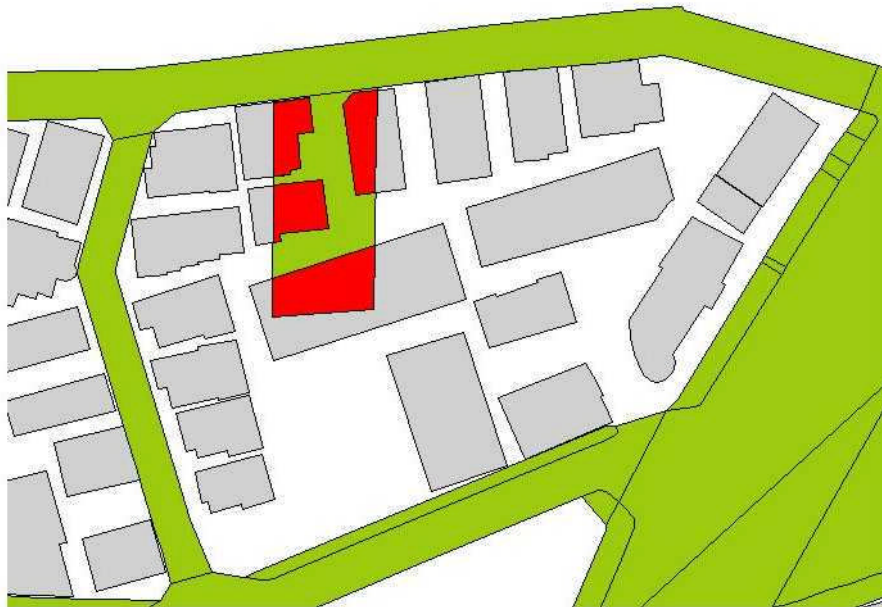
-  건물(1/1,000 수치지도)
-  도로(1/1,000 수치지도)
-  도로건물 중첩
-  건물(도로관리시스템)
-  도로면(새주소전산통합시스템)








<그림 부록-9.1> 건물은 갱신되었으나 도로면은 갱신되지 않은 경우



(a) 1/1,000 수치지도의 건물과 도로면

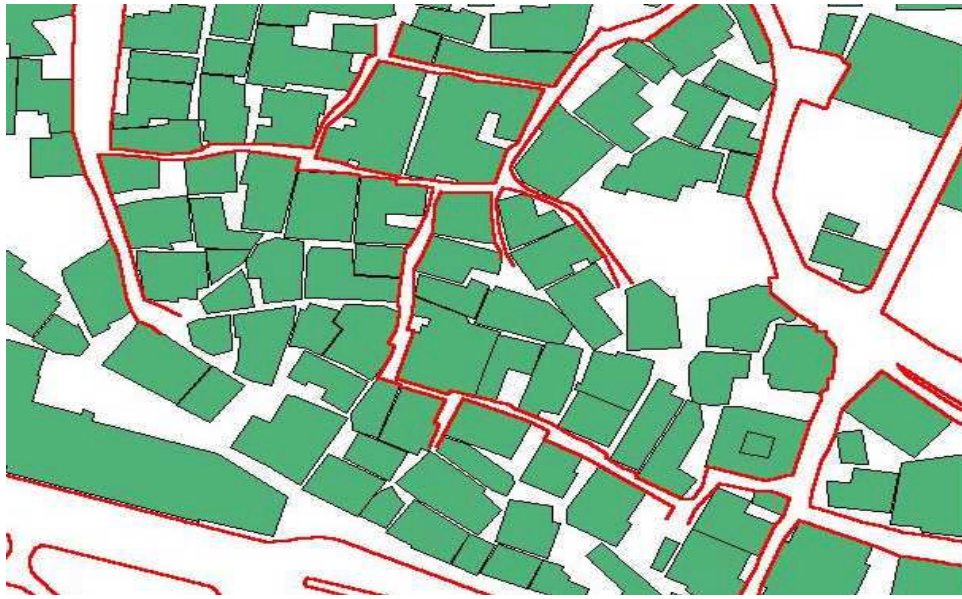


(b) 새주소전산통합시스템의 건물과 도로관리시스템의 도로면

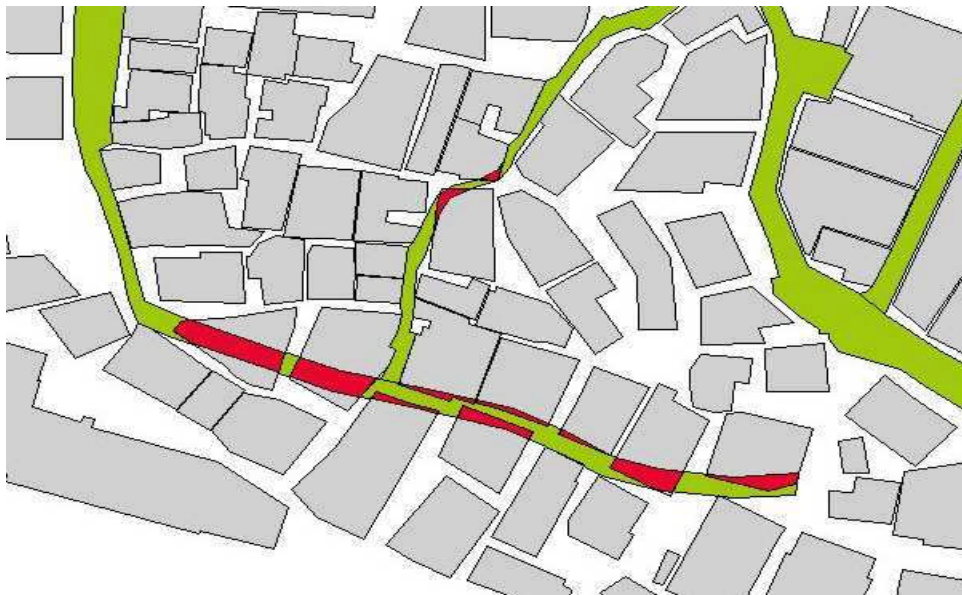
-  건물(1/1,000 수치지도)
-  도로(1/1,000 수치지도)
-  도로건물 중첩
-  건물(도로관리시스템)
-  도로면(새주소전산통합시스템)








<그림 부록-9.2> 도로면은 갱신되었으나 건물은 갱신되지 않은 경우



(a) 1/1,000 수치지도의 건물과 도로면



(b) 새주소전산통합시스템의 건물과 도로관리시스템의 도로면

-  건물(1/1,000 수치지도)
-  도로(1/1,000 수치지도)
-  도로건물 중첩
-  건물(도로관리시스템)
-  도로면(새주소전산통합시스템)



<그림 부록-9-3> 건물과 도로면이 모두 갱신된 경우

건물과 도로에서 정합성 오류의 가장 큰 원인은 단위업무시스템별로 1) 데이터 갱신의 차이, 2) 지형지물 입력방법에서 발생하는 차이이다(<표 부록-11> 참조).

<표 부록-11> 정합성 오류 발생 원인

정합성 오류 발생 원인		사 례	
		도로면 데이터 (도로관리시스템)	건물 데이터 (새주소전산통합시스템)
단위 시스템별 데이터 갱신절차 상이	데이터 갱신주기 상이	인정 공고시 속성갱신 (개통 및 미개통)	건축물관리대장에 등재시(신축 및 개보수)
	실측(측량)의 부재	도로 : 도시계획도 (1/3,000)상의 도로계획선을 디지털화	건물 : 현황도면(준공도면)을 디지털화 - 기준점 : 편집지적 및 연속지적 좌표 활용
단위 시스템별 지형지물 입력방법 상이	갱신대상 지형지물의 기준선 상이	도로 : 도시계획선 건물 : 건물처마선 (정사투영면)	도로 : 준공도면 건물 : 건물 처마선 또는 건물 외벽
	배경도면 지형지물 갱신방법 부재	도로 대한 체계적 갱신방법 부재	건물에 대한 체계적 갱신방법 부재

단위 시스템 별 데이터 갱신절차가 다른 경우는 데이터 갱신 주기가 다르거나 실제 측량이 이뤄지지 않으면 정합성 오류가 발생한다. 먼저 갱신 주기를 보면, 새주소 전산통합시스템의 건물데이터는 건축물관리대장에 등재 할 때 신축건물과 건물의 개·보수 시에 데이터 갱신 시기에서 차이가 나타난다. 이럴 경우 주기의 차이에서 오는 데이터 오류가 발생할 수 있다. 이를 방지하기 위해서는 갱신주기를 표준화 할 필요가 있다. 또 실제측량의 부재로 인한 오류를 들 수 있다. 건물의 현황도면 데이터는 일반적으로 준공도면을 지도화한 것으로 편집지적과 연속지적의 좌표를 활용하여 기준점을 잡는다. 그런데 실제 측량을 하지 않은 상태에서 자료를 통해 데이터의 갱신이 이뤄짐으로써 현실을 정확하게 반영하지 못할 위험이 있다.

다음으로 단위시스템별 지형지물을 입력하는 방법이 달라서 발생하는 정합성

오류가 있다. 이러한 오류는 갱신 대상 지형지물의 기준선이 다르거나 배경도면 지형지물의 갱신방법이 없기 때문에 나타난다. 예를 들어 도로관리시스템의 도로의 기준선은 도시계획선이나 새주소시스템에서 도로의 기준선은 준공도면에 준한다. 건물 역시 도로관리시스템에서의 기준선과 새주소통합시스템의 건물 기준선이 다르다. 이러한 차이로 인해 데이터의 정합성 오류가 발생한다.

이 같은 문제점을 해결하기 위해서는 좀 더 상세한 데이터 갱신주기의 표준절차가 필요하다. 앞서 제시한 현행 표준절차에는 세부 사항들에 관해 일관된 지침이 없기 때문에 이제는 운영에서 나타나는 오류들을 체계적으로 분석하여 이에 대처할 수 있는 표준절차를 강화해 나아가야 한다. 이와 함께 법적으로 모든 시스템에서 동일한 대상에 대해서는 기준을 일치시키는 작업들이 이뤄져야 할 것이다.

단위시스템의 개선에 대해서 덧붙이면, 건축물대장 작성과 관련하여 건축과에서 사용하는 건축시스템과 지적과에서 사용하는 지적시스템의 주소체계가 다르며, 이런 이유로 대장 자체가 부실한 경우도 있다. 예를 들어 토지에 관한 자료 중 편집지적과 토지대장이 일치하지 않는 경우도 찾아볼 수 있다.

(3) 입력오류

<표 부록-12>은 서울시의 3개 동에 대한 2002년 기준 건축물대장의 항목별로 입력오류 가운데 가장 빈번한 누락의 비율을 제시한 것이다. 3개 동의 누락률을 살펴보면 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. 그럼에도 건축면적이나 대지면적 등 측정이 필요한 항목에서는 누락률이 높은 편이다. 이러한 현상은 구청에서 입력하는 가장 기본적인 단계에서부터 건축물대장이 정비되어야 한다는 사실을 보여주고 있다.

<표 부록-12> 건축물대장의 주요 항목별 누락 비율 (2002년 기준)

항목	동	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
준공연도		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

4. 소결

여기서는 서울시가 공간데이터웨어하우스(SDW)을 바탕으로 다양한 부서에서 생산되는 행정자료의 공통 활용과 이미 구축된 데이터베이스의 활용성을 높이고, 자료의 중복구축 문제 해결을 통한 시스템 효율성의 증대 등의 시도를 살펴보았다. 서울시 SDW는 다양한 GIS데이터의 내부적 공유체계와 외부적 유통체계를 확립하면서 공간자료 중복구축 문제의 해결과 함께 공통되는 자료의 집적시스템을 통한 자료검색과 추출의 편의성 증대를 목표로 하였다. 서울시 SDW 데이터베이스는 기초/블록 데이터와 공통데이터(52개 레이어)로 구성되어있다. 기초/블럭 데이터는 통계청의 조사구 도형 자료를 비롯하여 건축물대장, 주민등록자료, 사업체자료 등이 포함되어 있으며, 공통데이터는 수치지도, 새주소의 건물도형, 도시계획 관련 자료, 편집지적 등이 포괄되어 있다. 서울시는 이러한 데이터 웨어하우스를 통해 자료의 유통체계를 공통으로 관리하고 자료 접근 과정에서 사용자 편의를 보장하여 더욱 효율적인 행정자료의 생산, 유통, 유지관리 방안 등을 마련하고자 하였다.

한편, 이러한 공통자료 품질에 대한 유지관리가 잘 이루어지기 위해서 단위시스템에서 생산되는 데이터의 품질이 반드시 전제되어야 한다. 서울시 건축물대장자료와 건물도형 자료의 매치문제, 건축물대장 자체의 데이터 품질 문제 등을 검토하

였으며, 이후 데이터의 정합성이 발생하는 원인, 사례분석을 통한 데이터 정합성 유지를 위한 방안 등을 검토하였다.

서울시 건축물대장은 다른 시도에 비해 상대적으로 정비가 잘 되어 있는 편이다. 그럼에도 단독주택밀집지역의 경우 건축물대장의 누락률이 높게 나타나고 있다. 이러한 자료 누락을 방지하기 위해서는 기초자치단체의 자료 입력에 대한 명확한 지침과 체계가 필요하며, 조직적인 보완도 있어야 할 것이다. 또한 현재 누락된 부분을 보정하기 위해서는 지속적인 현장작업을 통한 데이터 수정이 필요하다.

한편, 행정자료의 연결을 통한 자료 활용성을 높이기 위해서는 데이터의 정합성 오류 문제를 해결해야 한다. 공간자료(도형자료)와 속성자료(특성자료)의 매칭을 통해 이미 생산된 행정자료를 이용하고자 할 때 자료들을 서로 연계하기 어려운 문제에 봉착하게 된다. 단위시스템에서 입력오류가 원인일 수 있으며, 단위업무시스템별로 자료의 갱신주기나 방법, 기준의 상이함 때문에 나타날 수 있다. 이 문제를 해결하기 위해서는 상세한 갱신절차와 시스템의 표준화를 마련해야 한다. 절차적 표준화는 데이터 갱신주기의 일관성, 데이터 기준설정의 표준안, 단위시스템간 데이터 연동을 위한 시스템 표준화 등을 고려하여야 한다. 이 과정이 선행되지 않으면 많은 비용과 시간을 들여 구축된 시스템, 데이터 등의 활용성이 현저히 낮아질 것이다.

5. 정책제언

서울시의 사례에서 보듯이 먼저 내부 데이터 유통체계 확립이 필요하다. 단위시스템의 데이터 품질을 관리하기 위해서는 데이터 갱신에 따른 표준화된 절차마련이 필요하다. 특히 건물관련 데이터는 실시간 데이터 갱신이 가능하도록 하는 절차적 표준화 방안 마련이 요청된다. 이와 관련하여 현재 서울시 새주소시스템의 도형자료는 다른 공간자료에 비해 갱신주기와 빠른 편이며, 또한 건물에 대한 속성 데이터 체계도 잘 갖춰져 있는 편이다. 지금까지 살펴본 바와 같이 지번체계는 토지현황과 변화를 추적하기에는 어느 정도 기여하나 이를 건물자료와 연계시키기에는 오차가 많이 발생한다. 따라서 현재 체계화되어 있는 새주소시스템을 건물을 파악에 사용하는 것이 바람직하다.

데이터의 고품질 확보를 위해서는 기준데이터 선정, 구축 및 갱신에 대한 표

준지침을 마련하여, 각 단위시스템에서 기준데이터 구축뿐 만 아니라 관련 데이터의 구축·갱신에도 활용하여야 한다. 이러한 일의 효과적 추진을 위해서는 우선 부서별로 생산·저장되는 각종 데이터 및 전산환경을 조사한 후, 데이터의 공동이용에 관한 수요조사와 공동활용 데이터 목록을 작성하고, 이를 적극적 이용하기 위한 공동활용체계를 수립하여야 한다. 특히 통계청의 경우 부서 간 협력조정기제를 마련하는 것이 필요하다.

많은 행정부서의 행정에서 많은 데이터들이 생성된다. 또한 전자정부 이후 각 부처와 지방자치단체들은 많은 비용을 들여 정보시스템을 구축하였다. 그러나 이들 단위정보시스템에는 많은 데이터들이 중복구축되고 있다. 서울시 내부만 하더라도 192개의 단위정보시스템이 구축되어 있다. 따라서 많은 비용과 시간이 들어가는 데이터의 중복구축을 방지하기 위해서는 시스템 연동에 따른 데이터의 공유가 필요하다. 다시 말해, 부서 간 활용정도가 높은 데이터들은 외부적 유통체계를 마련하는 것이 필요하다. 중앙부처의 경우 부서 간 장벽이 존재하지만 전자정부 구축과 유비쿼터스 전자정부로의 발전방향을 고려한다면 어느 정도 가능할 것이다.

원활한 시스템 연계 및 데이터 공동 활용을 위해서는 관련되는 각종 요소들의 표준화 작업이 필요하다. 데이터 타입, 속성코드, 테이블 명명방법 및 자료변환 등에 관한 표준안 작성이 요구되며, 데이터의 공동 활용에서 혼란이 예상되는 여러 유사자료들을 평가할 수 있는 기준자료선정과 이를 이용한 다른 자료의 갱신 및 신규자료구축에 관한 표준절차도 마련되어야 할 것이다.