발 간 등 록 번 호
11-1240162-000025-01

2016

빅데이터와 행정자료의 이해

통계청
통계교육원

목 차

1. 빅데이터와 행정자료의 정의 ····· 1

1.1 빅데이터의 정의 ····· 2

1.1.1 학문적 의미 ····· 2

1.1.2 현실적 의미 ····· 10

1.1.3 빅데이터의 특징 ····· 21

1.2 행정자료의 정의 ····· 31

1.2.1 학문적 정의 ····· 31

1.2.2 현실적 정의 ····· 33

1.2.3 해외 사례 ····· 38

1.2.4 행정자료의 특징 ····· 39

1.2.5 행정자료의 유형 ····· 43

1.3 빅데이터와 행정자료의 관계 ····· 44

1.3.1 이론적 측면에서의 관계 ····· 44

1.3.2 존재 형태면에서의 관계 ····· 44

1.3.3 현실적 측면에서의 특성 비교 ····· 45

2. 빅데이터의 연구 및 활용 추이 ····· 47

2.1 세계적 추이 ····· 51

2.1.1 선진국 사례 ····· 52

2.1.2 국제기구의 추이 ····· 65

2.2 우리나라의 추이 ····· 77

2.2.1 정부 및 공공기관 ····· 78

2.2.2 민간 및 기업 ····· 85

2.2.3 학계 ····· 87

3. 우리나라 빅데이터 보유 현황 ····· 97
3.1 통계청 ····· 102
3.1.1 국가통계포털 ····· 102
3.1.2 국가지표체계 ····· 105
3.1.3 통계지리정보 ····· 107
3.1.4 통계분류포털 ····· 108
3.1.5 마이크로데이터 ····· 110
3.2 중앙행정기관 ····· 112
3.3 공공기관 및 중앙행정부처 산하기관 ····· 131
3.4 민간(기업, 금융기관, 협회) 등 ····· 143
3.5 지방자치단체 ····· 151
3.6 정부 3.0 ····· 157

4. 우리나라 빅데이터 활용현황과 과제 ····· 159
4.1 통계청의 활용현황 ····· 165
4.1.1 빅데이터 이용 통계작성 ····· 167
4.1.2 빅데이터와 조사통계의 결합 ····· 168
4.1.3 빅데이터를 이용한 통계 벤치마킹 ····· 173
4.2 중앙행정기관별 활용현황 ····· 175
4.3 공공기관별 활용현황 ····· 199
4.4 민간 및 기업의 활용현황 ····· 204
4.5 지방자치단체의 활용현황 ····· 213
4.6 학계 및 연구기관의 활용현황 ····· 226

5. 빅데이터 통계활용에 대한 미래 전략 ····· 231
5.1 현황에서 나타난 한계와 문제점 ····· 233
5.1.1 과잉 기대의 정점에 위치한 빅데이터 ····· 234

5.1.2 빅데이터를 둘러싼 쟁점들 ····· 237
5.2 개선과 미래 발전을 위한 제언 ····· 242
5.2.1 제도적 구조적 개선 방안 ····· 242
5.2.2 현실적 이용확산 방안 ····· 245

6. 행정자료의 통계적 활용방법과 국제적 추이 ····· 249
6.1 행정자료의 통계적 활용방법 ····· 250
6.1.1 행정자료만을 이용한 통계작성 ····· 250
6.1.2 행정자료와 통계조사의 연계 ····· 251
6.1.3 행정자료와 통계조사 간의 벤치마킹 ····· 252
6.2 행정자료 활용의 국제적 추이 ····· 254

7. 우리나라 행정자료 통계활용 현황과 과제 ····· 255
7.1 통계청의 활용현황 ····· 255
7.1.1 행정자료 통계작성 ····· 256
7.1.2 통계조사 항목 대체 ····· 257
7.1.3 통계 벤치마킹 ····· 259
7.2 행정자료 DB와 종합통계등록부 구축 ····· 260
7.2.1 구축배경 ····· 260
7.2.2 종합통계등록부의 구조와 이론적 배경 ····· 263
7.2.3 DB구축 및 시스템 개발 현황 ····· 271
7.2.4 향후 추가개발필요 사업의 내용 ····· 275
7.2.5 시스템운영 및 유지관리 ····· 277
7.2.6 종합통계등록부 연계통합 ····· 286
7.3 중앙행정기관과 지방자치단체의 활용현황 ····· 291
7.4 공공기관의 활용현황 ····· 296
7.5 기타 민간 및 연구기관의 활용현황 ····· 297

8. 행정자료 통계활용에 대한 미래전략 ····· 299
8.1 우리나라 행정자료 통계활용의 한계와 문제점 ····· 299
8.2 미래발전을 위한 제안 ····· 304
8.2.1 제도적 구조적 개선방안 ····· 304
8.2.2 현실적 활용확산방안 ····· 305

9. 행정자료를 이용한 통계개발 사례 ····· 307
9.1 개발 배경 및 환경 ····· 307
9.1.1 개발 배경 ····· 307
9.1.2 개발 환경 ····· 308
9.2 선행사례 고찰 ····· 310
9.2.1 해외 사례 ····· 310
9.3 개발 가능성 검토 ····· 314
9.3.1 행정자료의 일자리 설명력 검토 ····· 314
9.4 임금근로일자리 행정통계 개발 ····· 317
9.4.1 일자리 개념 ····· 317
9.4.2 기업체와 사업체 개념 ····· 318
9.4.3 통계지표 ····· 319
9.5 항목 개발 ····· 321
9.5.1 활용 자료 ····· 321
9.5.2 생산가능 항목 검토 ····· 322
9.5.3 항목별 생산가능성 검토 ····· 322
9.6 지표 개발 ····· 327
9.6.1 통계지표 개발 검토 ····· 327
9.6.2 통계지표 작성(안) ····· 328
9.7. 임금근로일자리 행정통계 작성 방법 ····· 329
9.7.1 작성 방법론 ····· 329

9.7.2 포괄범위 확대 ····· 329
9.7.3 시계열 안정 ····· 330
9.7.4 정확성 제고 ····· 331
9.7.5 통계 처리 ····· 332
9.7.6 통계지표 작성 ····· 332
9.7.7 작성 과정 ····· 333
9.8 향후 개선 과제 ····· 340

10. 행정 빅데이터를 위한 오픈소스 소개 ····· 343
10.1 분석 소프트웨어 ····· 343
10.1.1 SAS 소개 ····· 343
10.1.2 R 소개 ····· 346
10.2 지리정보 시스템 ····· 348
10.2.1 QGIS 소개 ····· 348
10.2.1.1 GIS 개요 ····· 348
10.2.1.2 QGIS 소개 ····· 349
10.2.2 QGIS 설치하기 ····· 350
10.2.3 간단한 QGIS 실행법 ····· 354
10.3 범용소프트웨어(Excel add-in 프로그램) ····· 360
10.3.1 Power Pivot ····· 360
10.3.2 MS Power View ····· 375
10.3.3 Power Map ····· 401

참고문헌 ····· 416
부록 1: 행정자료별 활용 현황(172종, ‘15년말 기준) ····· 421
부록 2: 통계조사별 대체 및 검증·보완 현황(40종, ‘15년말 기준) ····· 426
부록 3: 통계조사별 항목대체 현황(28종, ‘15년말 기준) ····· 427

부록 4: ’16년 행정자료 신규수요 현황 ····· 428
부록 5: 완전대체 항목 ····· 430
부록 6: 일부대체 항목 ····· 431
부록 7: 기존 통계 검증·보완 ····· 436
부록 8: 노르웨이 국가통계와 통계청에 관한 법 ····· 442
부록 9: 노르웨이 국가통계와 통계법 보충 법률 (통계법 시행령) ····· 448
부록 10: 데이터 소재 정보 ····· 454

빅데이터와 행정자료의 이해 개요

학습목표

본 교재의 학습목표는 빅데이터와 행정자료의 존재 현황, 통계적 활용 방법 및 활용현황, 활용가치를 이해하는 것이며, 주요 학습목표는 다음과 같다.

- 빅데이터와 행정자료의 학문적, 현실적 정의와 특징 및 두 자료의 관계를 이론적 현실적 관점에서 학습한다.
- 빅데이터의 연구 및 활용에 관하여 국제적 상황과 우리나라 상황으로 나누어 학습한다.
- 우리나라의 데이터 소재 정보를 공적영역과 민간영역으로 나누어 파악해 보고 활용에 대하여 학습한다.
- 우리나라의 빅데이터 활용현황 및 사례에 대하여 분야별로 나누어 학습하고, 빅데이터 통계활용관련 문제점과 개선을 위한 방법에 대하여 학습한다.
- 행정자료의 연구 및 활용에 관하여 국제적 상황과 우리나라 상황으로 나누어 학습한다.
- 우리나라 행정자료의 보유현황에 대하여 학습한다.
- 우리나라 행장자료의 활용현황 및 사례에 대하여 분야별로 나누어 학습하고, 행정자료 통계활용관련 문제점과 개선을 위한 방법에 대하여 학습한다.
- 행정자료의 실제 구축과정에 대하여 학습한다.
- 빅데이터 분석 소프트웨어(오픈소스)에 대하여 학습한다.

주요용어

빅데이터, 행정자료, 빅데이터 마스터플랜, 정부3.0, 데이터 소재정보, 빅데이터 공통기반 플랫폼, 혜안, 행장자료DB, 종합통계등록부, 개인정보호법, 개인정보 비식별화, 오픈소스, R, Excel, Quantum-GIS

내용요약

- 빅데이터와 행정자료에 관해 현재의 정부기관과 통계청에서 실무적으로 사용되고 있는 개념과 정의, 분류 및 활용방법 등을 중심으로 이론체계를 정리
- 국가에 존재하는 빅데이터와 행정자료의 현황을 기관, 내용, 활용 및 공개 정도별 등으로 분류하여 전체 윤곽을 파악하고 이해할 수 있도록 내용을 정리
- 국가전반의 빅데이터와 행정자료의 존재현황과 내용 및 활용도에 대한 기본지식을 배양하고, 통계적 활용현황, 방법, 가치와 한계 발전방향에 관한 소양을 쌓을 수 있도록 내용을 구성하여 정리
- 빅데이터 분석을 위한 오픈소스를 소개하고 기본적인 내용을 실습해 볼 수 있도록 컴팩트하게 사용방법과 실습내용을 정리
- 이후의 교육과정인 빅데이터 실무기술, 빅데이터 통계작성, 행정자료 통계작성 등의 교육과정을 이수하는데 필요한 선수과정으로서, 개론서 역할을 담당할 수 있도록 내용을 구성함

CHAPTER 1

빅데이터와 행정자료의 정의

학습목표

- 빅데이터를 학문적 의미, 현실적 의미 관점에서 정의하고 특징을 학습한다.
- 행정자료를 학문적 의미, 현실적 의미 관점에서 정의하고 특징을 학습한다.
- 빅데이터와 행정자료의 관계를 이론적 측면, 존재형태 측면에서 학습한다.
- 빅데이터와 행정자료의 공통점과 차이점에 대하여 학습한다.

최근 정보통신 기술의 일상화가 이루어지면서 하드웨어, 소프트웨어에 이어 데이터가 정보화 사회의 새로운 핵심기술로 등장하고 있다. 이는 과거 하드웨어에서 소프트웨어, 그리고 데이터로 관심영역이 점차 확대되고 있다는 것을 의미한다. 이러한 관심영역의 중심에는 빅데이터가 포함되어 있어서 많은 이들과 영역에서 이슈가 되고 있다. 스마트 단말기 확산, SNS 활성화, 클라우드, M2M(machine to machine) 네트워크의 확산으로 데이터 폭발이 더욱 가속화되어 빅데이터의 기반이 확대되고 있다. 따라서 향후 5년 이내 빅데이터를 위한 고급분석 등 관련기술이 성숙될 것으로 전망하고 있다(한국정보화진흥원, 2011). 인터넷이 일상화된 지 10여년이 지난 지금 디지털 데이터가 양적으로 증대하고 있어 전 세계 데이터에 생성될 디지털 정보량은 2011년 1.8 제타바이트(zetta bytes)에서 2015년에는 7.9 제타바이트로 증가하고, 2020년에는 50배 급증하여 10배 많은 서버가 필요할 것이라 전망된다(한국정보화진흥원, 2010; 정보통신산업진흥원, 2011).

1.1 빅데이터의 정의

위키피디어(http://ko.wkipedia.org/)에서는 빅데이터란 “기존 데이터베이스 관리도구로 데이터를 수집, 저장, 관리, 분석할 수 있는 역량을 넘어서는 대량의 정형 또는 비정형 데이터 집합 및 이러한 데이터로부터 가치를 추출하고 결과를 분석하는 기술을 의미한다”라고 설명하고 있다. 이러한 설명의 의미는 그림 1.1의 워드클라우드를 보면 직관적으로 이해할 수 있다.



그림 1.1 빅데이터 워드클라우드
출처:http://biz.heraldcorp.com/view.php?ud=20130821000356

1.1.1 학문적 의미[1)]

빅데이터라는 개념은 1997년 ACM의 VIS(Proceeding of the 8th conference on Visualization)학회에서 Cox와 Ellsworth에 의해 처음 발표되었다. Cox는 그의 논문 ‘Application-controlled demand paging for out-of-core visualization’에서 ‘더 이상 데이터 저장소에 담을 수 없는 사태를 맞이할 것이다’라고 문제를 제기하면서 빅데이터 개념의 출발을 알렸다. 이후 2001년 메타그룹의 Doug Laney는 학문적인 관점에서 3V라고 하는 빅데이터의 세 가지 특징(Velocity, Volume, Variety)을 그림 1.2와 같이 정의하였다.

1) 본 절의 내용은 백두권의 빅데이터란 무엇인가?, 박원준(2013)의 빅데이터 활용에 대한 기대와 우려의 내용을 옮겨오거나 발췌하여 정리한 것임

빅데이터의 첫 번째 특징인 속도(Velocity)는 데이터들이 거의 실시간에 가깝게 생산된다는 것을 의미한다. 두 번째 특징은 대용량의 데이터(Volume)이다. 이는 데이터의 크기와 기록의 양 등이 현저하게 증가함을 의미한다. 마지막은 다양성(Variety)이다(그림 1.2). 과거에는 정형화된 데이터가 중심이 되어 각 정보들이 물건, 사건, 인물에 대한 정보 위주로 가공되었으나, 빅데이터 시대에서는 비정형화된 데이터가 중심이 되어 오락, 일상생활 등 우리가 생활하는 거의 모든 분야에 관한 데이터를 이용한다. 추가적으로, 2012년 가트너는 데이터 관리의 복잡성(Complexity) 속성을 더하여, 빅데이터의 특징을 3VC로 새롭게 정의하였다.[2)]

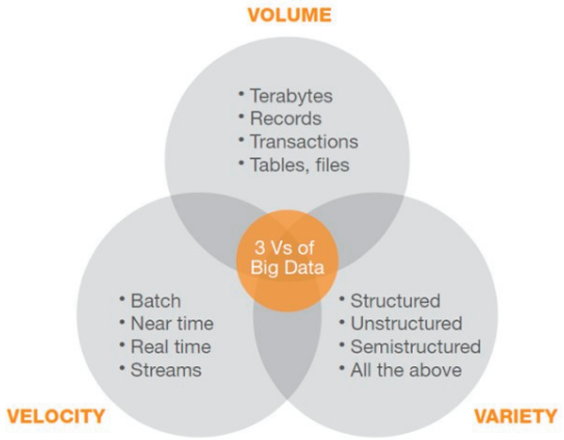


그림 1.2 빅데이터의 세 가지 특징

출처: http://api.ning.com/files/

다음의 표 1.1은 주요 기관에서 정의한 빅데이터의 정의를 정리한 것이다.

빅데이터에 대한 분석은 2008년 미국의 인디애나 주립대학의 조한 볼렌(Johan Bollen)교수에 의해서 시작되었다고 볼 수 있다. 그는 트위터에 올라온 글에서 개인과 관련된 정보가 포함되어 있음을 인지하고, 이 모든 데이터를 종합하여 분석하면 어떤 거대한 흐름을 발견할 수 있을 것으로 기대하였다. 그는 2008년 상반기 트위터에 올라온 모든 데이터를 분석해 이용자들의 집단적 기분 변화가 전국적인 행사와 일치한다는 사실을 발견하게 된다. 또한 추수감사절이 다가올수록 트위터에 올라온 기분변화는 행복지수를 나타내고, 반대로 사람들이 불안감을 느끼면 며칠 뒤 주가지수가 하락한다는 사실도 발견하였다(유지연, 2012). 이러한 분석은 과학적 데이터 즉, 통계나 실험을 통해 얻은 것이

2) http://www.diabetes.or.kr/new_workshop/201403/ab3.html

아니라 인간의 정서나 심리 정보에 해당하는 기분이나 감정을 분석한 것이다. 이와 같은 비정형화된 데이터는 인터넷을 비롯해 스마트 기기 이용을 통해서 생산되는데, 전자상거래 서비스에서 발생하는 거래내역이나 소비자 의견은 물론 게시판, 소셜 네트워크 서비스 이용자들의 모든 메시지가 여기에 해당된다.

표 1.1 기관 별 빅데이터의 정의

기관	빅데이터 정의
포레스트	- 가치를 얻기 위한 데이터와 무엇을 할 것인지 아는 사람이 기업에게 필요하다는 것을 의미하는 기술 - 볼륨, 속도, 다양함, 다양성으로 현재의 기술로 감당 어려운 규모의 데이터 - 경제적 가치를 창출하는 데이터
SERI	- 거대한 데이터 집합으로 대규모 데이터와 관련된 기술 및 도구포함
가트너	- 3V1C 정의 : Velocity, Volume, Variety, Complexity
SAS	- 4V로 정의 : Volume, Variety, Velocity, Value - Volume, Variety는 가트너 정의와 동일 - Velocity : 센서나 모니터링 등 사물정보, 스트리밍 정보 등 실시간성 정보가 증가하고 있고, 이러한 실시간성으로 인한 데이터 생성, 이동과 유통의 속도가 증가하고 있으며 대규모 데이터처리 및 가치 있는 실시간성 정보활용을 위해 데이터 처리 및 분석 속도가 매우 중요하게 되었음을 의미 - Value : 새로운 가치를 창출하는 것을 의미
노무라 연구소	- 빅데이터를 처리할 수 있는 인재· 조직, 데이터 처리·축적·분석기술, 데이터 자원 등을 빅데이터의 3요소로 정의 - 3요소의 조화로운 발전이 데이터의 특성과 컴퓨팅 파워의 발달에 따라 실생활 적용이 빠르게 확산될 것으로 전망

다음의 표 1.2는 데이터의 유형에 따른 구분을 정리해 놓은 것으로 앞으로 언급되는 데이터의 유형 구분에 사용되므로 참조하기 바란다.

표 1.2 데이터 유형에 따른 빅데이터 분류

빅데이터 유형	종류
정형 데이터	고정된 필드에 저장된 데이터 - 데이터베이스, 스프레드 시트 등
반정형 데이터	고정된 필드에 저장되어 있지는 않지만 메타 데이터나 스키마를 포함하는 데이터 - XML, HTML
비정형 데이터	고정된 필드에 저장되어 있지 않은 데이터 - 텍스트 문서, 이미지/동영상 등

최근 비정형화된 데이터는 숫자에 해당하는 기존 데이터의 양적인 측면과 질적인 측면을 넘어 비교할 수 없을 정도로 많은 데이터가 생산되고 있다. 전통적 개념의 데이터는 Google이나 Apple 등과 같은 대기업이나 NASA의 연구과학 프로젝트에서 분석하는 대용량의 숫자로 표현된 정형화된 데이터를 일컫는다. 하지만 빅데이터는 전통적인 데이터의 개념으로 모두 설명할 수는 없다. 아울러 빅데이터는 상대적인 개념으로 받아들이고 있는데, 이유는 기업 규모에 따라 데이터 저장과 분석 용량이 다르기 때문이다.

최근 논의되는 빅데이터는 '대용량의 데이터 외에 앱, SNS 등에서 생산되는 데이터'를 포함하는 개념이다(채승병, 2011). 이러한 광범위한 데이터는 기존 민간 기업이나 정부의 데이터 분석 범위를 넘어 선 것이다. 또한 정용찬(2012)은 '데이터의 양, 생성주기, 형식(수치데이터 뿐만 아니라 문자와 같은 비정형 데이터 포함) 등에서 과거 데이터에 비해 규모가 크고, 형태가 다양하여 기존의 방법으로는 수집, 저장, 검색, 분석이 어려운 방대한 크기의 데이터'라고 빅데이터를 정의하고 있다. 이러한 빅데이터의 정의는 기존 분석 시스템의 한계를 기반으로 빅데이터를 바라보고 있다. 그렇다보니 빅데이터의 정의에서도 조차 빅데이터를 매우 모호하게 규정하고 있다.

여기서는 빅데이터를 '숫자, 그림, 문자, 언어 등을 포함하는 다양한 형태의 대용량 데이터를 S/W를 통해 분석하여 새로운 가치해석이 가능한 데이터'로 정의하여 사용하고자 한다. 이러한 정의는 유무선 네트워크 사용에 따라 생산 또는 저장되는 정형화된 데이터와 SNS, 블로그, 홈페이지, 전자상거래 등에 의해서 생산되고 있는 비정형화 된 데이터를 포함한다. 따라서 빅데이터는 정형(Structured) 데이터, 반정형(Semi Structured) 데이터, 비정형(Unstructured) 데이터로 구분할 수 있다(표 1.2 참조). 정형화된 데이터는 일정한 규칙을 갖고 체계적으로 정리된 데이터를 의미한다. 예를 들어 매년 통계청에서 발표하는 통계자료, 방송통신 실태조사, 각종 과학적 데이터 등이 이에 해당한다. 정형화된 데이터는 그 자체로 의미 해석이 가능하며, 바로 활용할 수 있는 정보를 내포하고 있다. 반정형화된 데이터는 한글이나 MS Word 등으로 작성된 데이터를 의미한다. 대표적인 예는 인쇄매체의 텍스트라 할 수 있다. 반정형화된 데이터는 표나 그림이 될 수도 있지만 일반적으로 문자로 서술된 정보를 담지하고 있다. 비정형화된 데이터는 스마트 기기 등을 통해서 형성되는 데이터로 페이스북, 트위터, 카카오톡 등으로 상호 교류되는 정보가 이에 해당한다.

비정형화된 데이터는 개인, 집단, 사회, 국가 등과 관련된 주제를 스마트 미디어 이용자들이 상호 의견을 교류함으로써 생산되는 정보들이다. 특히 오늘날 빅데이터는 비정형화된 데이터에 관심을 두고 있다.

이러한 빅데이터는 해당 데이터를 분석하고 처리함으로써 기존의 데이터에서 볼 수 없었던 새로운 의미를 산출할 수 있다. 즉, 수많은 데이터를 분석하여 사용자에게 유용한 정보를 제공할 수 있어야 빅데이터는 효용성을 갖는다. 따라서 빅데이터에서 중요한 것은 형식적인 데이터 소스 내에서 외부로 새로운 가치를 창출할 수 있느냐 하는 것이다. 결국 새로운 가치와 의미를 산출하기 위해서는 축적된 데이터를 갖고 무엇을 분석할 것인가에 대한 고민이 필요하다. 이러한 고민은 데이터 마이닝과 연결되는데 빅데이터에서 데이터 마이닝은 텍스트 마이닝(Text Mining)과 웹마이닝(Web Mining) 그리고 소셜 마이닝(Social Mining)을 통해서 현실 마이닝(Reality Mining)에 도달해야 한다. 빅데이터의 현실 마이닝은 우리가 영화에서나 볼 수 있었던 미래를 예측할 수 있는 데이터들이 산출되어 사후대책이 아니라 사전 방지 시스템을 작동시킨다.

아울러 빅데이터를 활용하기 위해서는 기존의 데이터 분석방법이 아니라 새로운 기법의 분석방법을 도입해야 한다. 빅데이터 활용에 따른 분석방법의 변화는 활용의 변화와 분석방법의 변화로 구분할 수 있다. 빅데이터에 대한 활용의 변화는 앞서 언급한 바와 같이 비정형화된 데이터를 기반으로 실시간 분석이 가능해야 하며, 특정기관이 아니라 민간, 기업, 정부 등 다양한 주체들이 활용할 수 있도록 유도해야 한다. 분석의 변화는 데이터의 크기에 대한 변화를 기반으로 다중의 저장 공간과 분석틀로 데이터의 흐름을 분석할 수 있어야 한다. 이를 위하여 빅 테이블(Big Table), 카산드라(Cassandra), 데이터웨어하우스(Data ware house) 및 분석 어플라이언스, 분산시스템(Decentralized system), 구글 파일시스템(Google file system), 하둡(Hadoop), H베이스(Hbase), 맵리듀스 (MapReduce), 비관계형 데이터베이스(non-relational database) 등과 같은 빅데이터 관련 기술이 등장하고 있다(표 1.3 참조).

앞의 설명에서 알 수 있듯이 빅데이터에 관하여 내용을 정리하기 위해서는 자료에 대한 기본적인 의미에 더하여 '데이터의 수집', '데이터의 저장·관리', '데이터의 분석' 기술을 모두 포괄하는 것으로 정리해야 한다. 이와 관련된 자세한 내용은 빅데이터의 분석을 다루는 교육과정에서 더 자세히 다루도록 한다. 여기서는 표 1.4에서와 같이 빅데이터 기술과 전통적인 통계처리 간의 비

교를 통해 개괄적인 내용을 소개하는 것으로 지나가도록 한다.

표 1.3 빅데이터와 연결된 기술

용어	설명
Big Table	• 구글 파일 시스템 상에 구축된 상용 분산 데이터베이스 시스템 • H베이스에 영향을 미침
Cassandra	• 분산 시스템에 방대한 분량의 데이터를 처리할 수 있도록 디자인된 오픈소스 데이터베이스 관리시스템 • 이 시스템은 페이스북에서 개발했으며, 지금은 아파치 소프트웨어 재단의 한 프로젝트로 관리
Data ware house & Appliance	• 데이터웨어하우징을 위해 서버, 스토리지, 운영체제, 데이터베이스, BI, 데이터마이닝 등 기타 여러 가지 소프트웨어가 최적화되어 설치된 통합제품
Decentralized system	• 동시에 일을 처리하기 위해 네트워크로 연결된 컴퓨터들의 집합 • 단일 또는 다수의 컴퓨터 리소스를 부분적으로 활용함으로써 시스템의 가격 대비 성능비, 안정성, 확장성을 향상시킬 수 있음
Google file system	• 구글에서 개발한 분산파일 시스템으로 하둡(Hadoop)과 관련
Hadoop	• 분산시스템 상에서 대용량 데이터 처리분석을 지원하는 오픈 소스의 소프트웨어 프레임워크 • 구글이 개발한 맵리듀스를 오픈소스로 구현한 결과물 • 야후에서 최초 개발되었으며, 지금은 아파치 소프트웨어 재단의 한 프로젝트로 관리되고 있음
Hbase	• 구글의 빅테이블을 참고로 개발된 오픈소스 분산 비관계형 데이터베이스 • 파워셋에서 개발했으며 현재는 아파치소프트웨어 재단에서 하둡의 일환인 프로젝트로 관리되고 있음
MapReduce	• 분산시스템 상에서 대용량 데이터 세트를 처리하기 위해서 구글이 제안한 소프트웨어 프레임워크. 하둡에도 구현 가능
non-relational database	• 비관계형 데이터베이스는 데이터를 테이블에 저장하지 않는 데이터베이스이며 관계형 데이터베이스와는 대조적인 개념 • 이를 사용하면 스키마 없는 엔티티(NoSQL)를 관리할 수 있음

출처: 김성태 (2012).

빅데이터가 이슈화되는 이유는 크게 3가지 요인으로 볼 수 있다. 첫째 요인은 스마트 폰을 비롯하여 모바일 스마트 기기 보급의 활성화이다. 모바일 스마트 기기에 탑재된 센서, 원격감지기술, 소프트웨어, 카메라, RFID 리더 등을 통해서 비정형화된 데이터를 수집할 수 있게 된 것이 데이터 증가의 원인이라고 할 수 있다. 둘째, 클라우드 서비스이다. 클라우드 서비스를 통해서 개인과

조직의 데이터가 한곳으로 축적되고, 저장된 데이터를 분석하여 활용하고자 하는 요구가 증가하고 있다. 셋째, 소셜 미디어의 활용이 일상화되면서 정보유통 구조가 새롭게 재편되고 있다. 소셜 미디어의 특성상 쌍방향 커뮤니케이션이 활발하게 이루어지면서 상호작용 데이터의 증가를 가져왔다. 이러한 상호 작용 데이터를 비즈니스 측면에 활용하려는 관심이 높아졌기 때문이다(McKinsey, 2011). 빅데이터의 활용측면에서 빅데이터의 분석에 관한 것은 매우 중요하다. 빅데이터의 분석은 데이터의 절대적인 크기 기준에서 기존의 분석방법의 적용이 어려울 수 있으므로 기존 방법의 진화 또는 새로운 방법들의 개발 등으로 변화되었다. 이에 관하여 '빅데이터 분석기술'과 '빅데이터 분석기법'으로 구분하여 알아보도록 하자.

표 1.4 빅데이터 및 전통적 통계생산 방식의 기술 비교

구분	빅데이터 기술	전통적인 통계생산 방식
데이터 수집	- 데이터 크롤링 - 각종 센서 집계(모바일 기지국, 교통결제, 비콘, 와이파이 기기 등)	- 수작업을 통한 데이터 수집 (설문기법 활용) - 수작업을 통한 데이터 기입
데이터 저장·관리	- 분산저장 - 병렬식 서버 구조 활용 - 정보시스템을 통한 자동 데이터 관리 - 비관계형 데이터베이스 활용	- 관계형 데이터베이스 활용 - 단일 서버 활용 - 종이 자료
데이터 처리·분석	- 분산처리를 통한 복수의 컴퓨터 자원 활용 - 그리드컴퓨터 기술 활용 - 고급 분석기술 적극 활용 - 데이터를 통한 예측/최적화/기계학습 등 다양한 범위에 적용 - 대용량데이터 처리에 적합한 패키지 활용	- 단일 컴퓨터 자원 활용 - 전통적인 통계분석 방식 사용

□ 빅데이터 분석기술[3)]

3) 이 부분의 내용은 통계개발원의 '빅데이터의 국가통계 활용을 위한 기초연구'의 내용을 내용을 옮겨오거나 발췌하여 정리한 것임

① **기술분석**(Descriptive Analytics) 과거에서부터 현재까지 주어진 데이터로부터 현재의 상황을 설명할 수 있는 패턴(Pattern)을 찾아 사용자의 이해를 돕기 위해 표현하거나 설명하는 것으로 주어진 기간 동안에 어떤 사건이 얼마나 많이 혹은 얼마나 자주, 어디에서 발생했는지에 대한 정보를 제공(빈도분석 등)

② **예측분석**(Predictive Analytics) 과거의 데이터나 사건으로부터 미래에 발생 가능한 상황이나 사건을 예측하여 선제적인 의사결정을 지원(은닉 마르코프 모델* , 베이지안 네트워크 모델*)

③ **최적화**(Optimization) 주어진 가능한 결과들에 대한 평가를 수행하여 최적의 결과를 도출하는 것으로, 제시된 전략을 평가하고 최적의 대안을 선택하도록 지원. 확률적 기법 및 통계적 기법과 함께 기술모형과 예측모형을 병합하며 빅데이터 서비스 및 기계학습* 모델에서 다양하게 활용

* **은닉 마르코프 모델(Hidden Markov Model)** 모델링하는 시스템이 미지의 모수를 가진 Markov Process일 것이라고 가정하여, 그 가정에 기초해서 관측된 모수로부터 숨겨진 모수를 결정하려 하는 하나의 통계모델

* **베이지안 네트워크 모델(Bayseian Network Model)** 랜덤 변수의 집합과 그래프를 통하여 그 집합을 조건부 독립으로 표현하는 확률의 그래픽 모델

* **기계학습(Machine Learning)** 임의의 데이터가 있을 때 이를 컴퓨터가 인지하는 형태로 표현하고 컴퓨터에게 적용하는 방법으로 Deep Neural Networks, Convolutional Deep Neural Networks, Deep Believe Networks와 같은 다양한 딥러닝 기법들이 컴퓨터비전, 음성인식, 자연어 처리, 음성/신호 처리 등의 분야에 적용

□ 빅데이터 분석기법[4)]

① **분류**(Classification) 미리 알려진 클래스들로 구분되는 훈련 데이터 군을 학

4) 이 부분의 내용은 통계개발원의 '빅데이터의 국가통계 활용을 위한 기초연구'의 내용을 옮겨오거나 발췌하여 정리한 것임

습시켜 새롭게 추가되는 데이터가 속할 만한 데이터 군을 찾는 방법

② **군집화**(Clustering) 비슷한 특성이 있는 데이터들을 합쳐가면서 유사 특성 군으로 분류하는 방법

③ **텍스트 마이닝**(Text Mining) 인간의 언어로 쓰인 비정형 텍스트에서 자연어 처리 기술을 이용하여 유용한 정보를 추출하거나, 연계성 파악, 분류 혹은 군집화, 요약 등 빅데이터에 숨겨진 의미 있는 정보를 발견하는 방법

④ **오피니언 마이닝**(Opinion Mining) 소셜 미디어의 텍스트 문장을 대상으로 자연어 처리 기반 감성분석기술을 적용하여 사용자의 의견을 분석하는 방법

⑤ **감성분석**(Sentiment Analysis) 자연어 처리 기법을 이용하여 인간의 언어로 쓰인 텍스트 문장을 분석할 때, 문장에서 주관적인 감정을 나타내는 정보를 찾아내어 긍정, 부정, 중립의 성향을 분석하는 방법

1.1.2 현실적 의미[5)]

빅데이터에 관한 설명은 여러 가지가 있지만, 여전히 빅데이터에 대한 혼란은 대부분 빅데이터의 정의와 관련이 있다. 현실적으로 빅데이터라는 용어를 어떻게 정의하는지 알아보기 위해, 2012년 중반 IBM이 전 세계 95개국 26개 산업 분야의 전문가 1,144명을 대상으로 실시한 연구인 "빅데이터@워크" 연구를 참조하였다. 이 연구는 다양한 분야의 비즈니스 전문가(전체 표본의 54%)와 IT 전문가(46%)로 구성되었으며, 모든 응답자는 자발적으로 인터넷 설문조사에 참여하였다.

빅데이터의 특징을 두 가지씩 선택해 달라고 요청하였고, 빅데이터를 가장 잘 설명할 수 있는 문구로 응답자들이 선택한 대답은 하나의 특징에 집중된 것이 아니라, 데이터의 규모 확대, 새로운 유형의 데이터 및 데이터 분석, 실시간 분석에 대한 요구 증가 등 여러 가지로 분산되어 나타났다(그림 1.3 참조).

5) 본 벌이 내용은 IBM 글로벌 비즈니스 서비스의 요약보고서인 빅데이터의 현실적인 활용(기업이 불확실한 데이터에서 가치를 창출해 내는 방법)의 내용을 옮겨오거나 발췌하여 정리한 것임

빅데이터의 성공적인 활용을 위해서는 규모와 다양성, 속도 측면에서 급증을 거듭하는 데이터를 처리할 수 있는 효율적인 빅데이터 인프라가 구축되어있어야 한다. 이에 대한 의견수렴을 위해 현재 자사의 빅데이터 관련 인프라를 명시해 달라고 요청한 결과, 응답자의 3분의 2 정도가 빅데이터 활용과 관련해 확장이 가능하고 보안 기능이 있는 통합된 정보 기반을 구축하고 있다고 대답하였다. 응답자들은 빅데이터 활용 계획과 관련해 '정보통합', '확장 가능한 스토리지 인프라', '대용량 웨어하우스', '보안 및 관리'요소를 가장 자주 언급하였다(그림 1.4 참조).

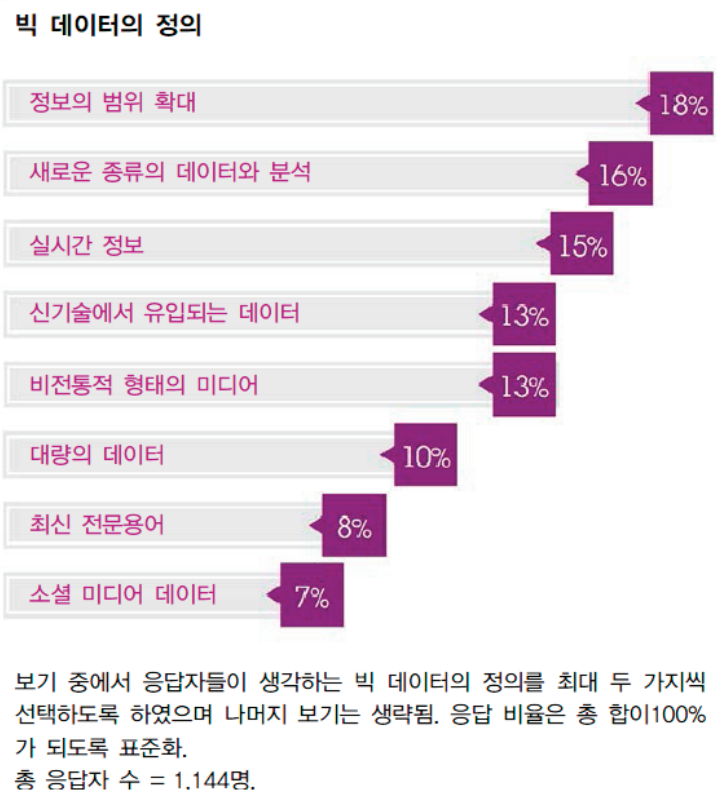


그림 1.3 빅데이터의 정의

정보 통합(integrated information)은 모든 분석 작업의 핵심 요소이며, 빅데이터 관련 작업에서는 더욱 중요한 요소이며, 데이터는 데이터를 필요로 하는 사람이나 시스템이 언제든 쉽게 접근하여 이용할 수 있어야 한다. 조직 간 혹은 부서 간 데이터 전달 및 공유가 이루어져야 하며, 빅데이터 활용을 위해서는 이런 관리와 통합이 복잡하면서도 중요한 문제가 된다. 빅데이터를 활용하고 있다고 대답한 응답자들 중 65%가 자사의 정보 통합 능력이 빅데이터를 지원하기에 충분하다고 응답하였다.

빅데이터 활용을 위한 정보 관리 기반에 있어 두 번째로 중요한 요소는 '확장이 가능한 스토리지 인프라와 대용량 데이터 웨어하우스'로 나타났다. 두 가지 모두 유입되는 현재 및 미래의 데이터 급증을 대비하기 위한 장치이다. 외견상으로는 스토리지를 추가하고 대용량 서버를 한두 대 더 설치하면 정보 관리 기반의 성장을 뒷받침할 수 있을 것처럼 보이지만 인프라 예측과 설계야말로 의도하는 빅데이터의 가치를 끌어낼 수 있는 최대 관건이다.

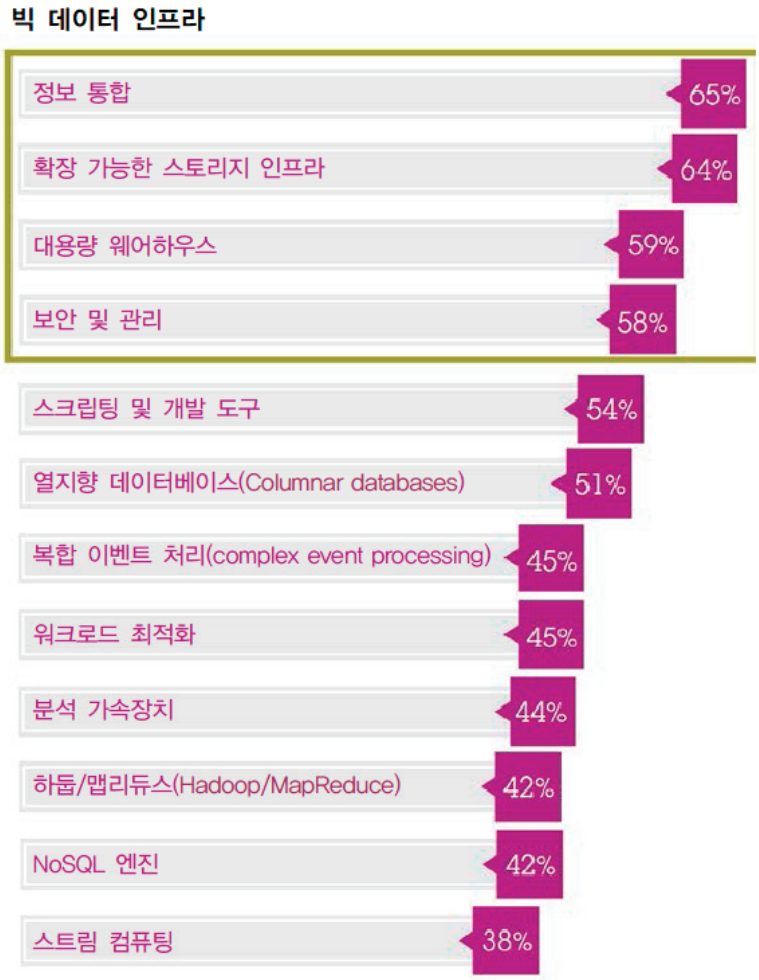


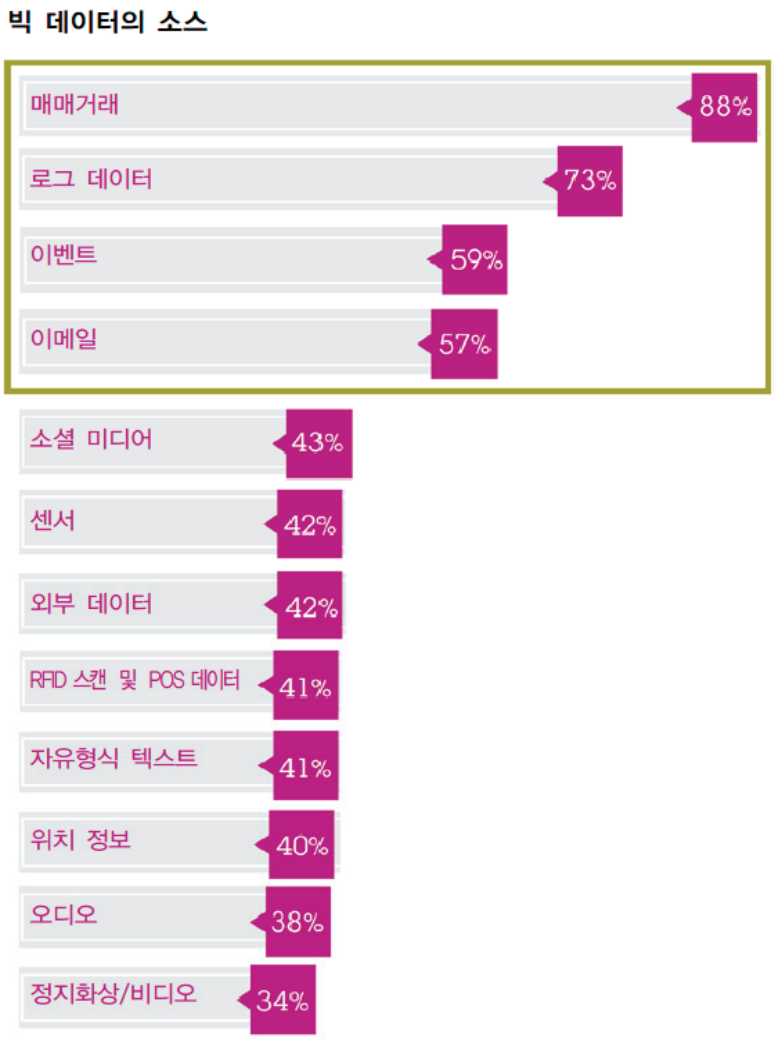
그림 1.4 빅데이터 인프라

사용자들이 필요할 때 데이터에 접근할 수 있도록 급변하는 데이터 흐름을 관리하는 방법을 고려해야 하며, 이와 같은 서버와 스토리지의 균형 잡힌 설정

과 배치는 결국 보다 효율적인 인프라 최적화로 이어지게 된다.

현재 적극적으로 빅데이터를 활용하고 있다고 응답한 경우의 58%가 강력한 보안 및 관리 프로세스를 구비하고 있었다. 보안과 관리는 오랫동안 BI(Business Intelligence)의 본질적인 부분이었지만, 빅데이터에 대한 법률적, 윤리적인 제약들은 새로운 위험 요인이 되고 있다.

인터뷰에 참여한 관련 분야 전문가 및 경영진에 따르면, 보안과 관리는 앞으로 새로운 소스, 특히 소셜 미디어에서 데이터를 수집하게 되면서 더욱 중요한 문제가 될 것이며 프라이버시 관련 규제가 계속 강화되고 국가 별로 다양해지는 것이 앞으로 빅데이터 활용을 어렵게 만드는 문제로 예상하였다.



빅 데이터 활용을 위해 현재 수집 및 분석 중인 데이터 소스. 각 데이터 포인트는 독립적으로 수집됨. 각 데이터 포인트에 대한 총 응답자 수는 557명부터 867명 사이.

그림 1.5 빅데이터 원천

빅데이터 도입 초기에는 주로 내부 데이터를 수집하여 분석하는 데 집중한다. 조사 결과, 응답자의 절반 이상이 조직의 내부 데이터가 빅데이터의 주요 원천이라고 밝혔으며, 이런 현상은 조직들이 빅데이터 도입에 실용적으로 접근하고 있다는 사실과, 각 조직의 내부 시스템 안에 아직 발견하지 못한 굉장한 가치가 내재되어 있다는 사실을 의미한다. 예측 가능한 일이지만, 내부 데이터야 말로 가장 잘 파악하고 있는 익숙한 데이터이다. 그러나 빅데이터는 전통적인 시스템의 저장 및 분석능력을 초과하는 데이터이며, 이런 데이터의 상당 부분은 수집만 될 뿐 분석은 이루어지지 않고 있다. 그림 5는 대상자들이 응답한 빅데이터의 원천을 정리한 것이다.

내부데이터 - 자체적으로 보유한 내부 파일시스템이나 데이터베이스 관리 시스템 등에 접근하여 데이터를 수집

외부데이터 - 웹크롤링 엔진을 통해 인터넷 링크의 모든 페이지 복사본을 생성함으로써 데이터를 수집

빅데이터가 가치를 창출하기 위해서는 더 다양한 데이터에 접근할 수 있어야 하며 소프트웨어 도구와 그 도구의 활용 기술을 두루 갖춘 강력한 분석 능력이 필요하다(그림 1.6 참조).

빅데이터를 활용 중이라는 응답자의 75% 이상이 빅데이터 분석을 위해 쿼리와 리포팅, 데이터 마이닝과 같은 핵심 분석 기능을 이용한다고 대답하였고 67% 이상이 예측 모델링을 이용한다고 대답하였다. 빅데이터를 도입하면 데이터베이스가 방대해져서 비즈니스 분석가나 데이터 분석가가 전통적인 리포팅 및 데이터 마이닝 툴을 이용해 관찰하고 분석하기 곤란한 경우가 많아지므로 고급 데이터 시각화 기능에 대한 요구가 늘어나게 된다. 조사 결과 응답자들은 빅데이터 활용의 71%를 데이터 시각화 기술에 의존하고 있다고 대답하였다.

빅데이터를 활용할 때는 복잡한 데이터 속에서 패턴을 발견하기 위해 점점 더 고급 능력을 원하게 된다. 응답자들은 핵심 비즈니스 프로세스의 개선 방안을 파악하기 위하여 최적화 모델 및 고급 분석을 적용하고 있었다. 또한 빅데

이터 안에서 만나게 되는 수많은 변수들을 분석하기 위해 시뮬레이션 기능도 사용하고 있었다. 조사 결과 빅데이터를 활용 중인 기업들의 50% 정도는 이런 고급 모델링 기능을 사용하고 있는 것으로 밝혀졌다.

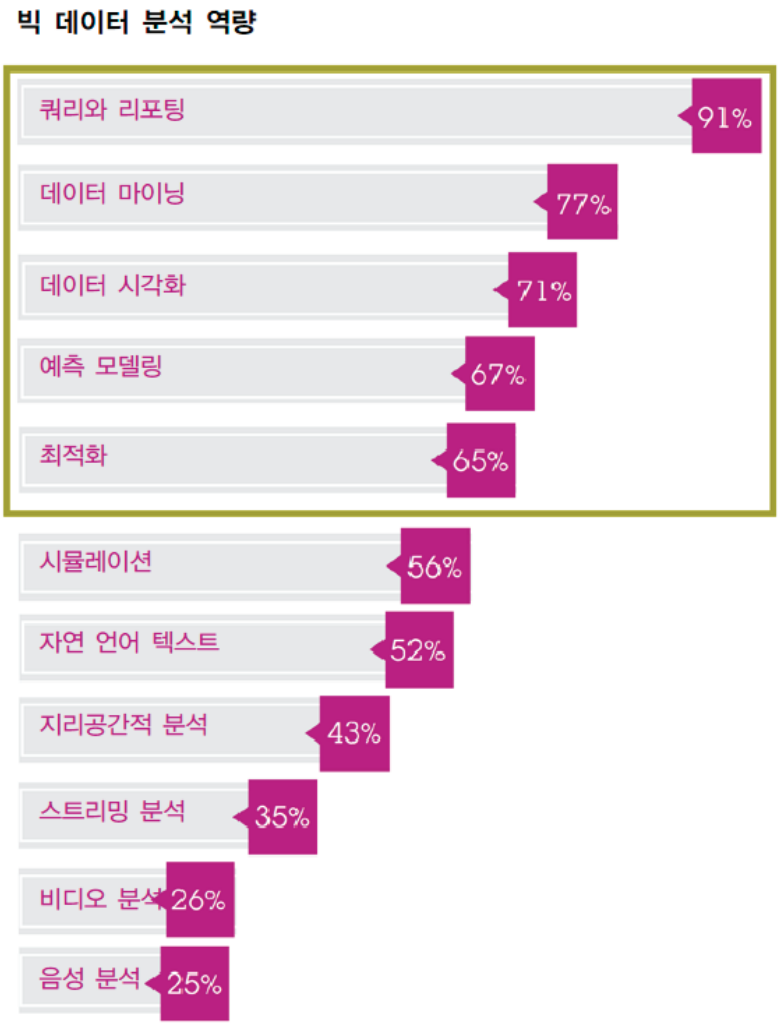


그림 1.6 빅데이터 분석기능

빅데이터는 분명 새로운 가능성을 제시하고 있으며, 사회전반에 걸쳐 혁신적인 변화를 가져올 것이다. 하지만 빅데이터라는 개념정의에서 알 수 있듯이 데이터 규모에 대한 규정과 데이터 종류, 데이터 보관과 활용 차원에서 사회적 문제를 유발할 수 있다. 따라서 빅데이터 효용에 대한 기대 못지않게 가외적인 문제를 어떻게 처리하는가에 따라 빅데이터의 평가가 달라질 수 있다. 이러한

관점에서 현실적으로 빅데이터에 관한 우려에 대한 내용들을 정리해 보았다.

□ 새로운 빅브라더(Big Brother)의 탄생

데이터 집중화 현상에 따라서 데이터 관리 및 데이터 분석 서비스 제공 업체가 나타나게 된다. 이는 곧 빅브라더의 탄생을 의미한다. 빅데이터를 분석하기 위해서는 데이터 보관 및 관리에 대한 책임 있는 기관이나 기업이 등장할 수밖에 없으며, 책임 있는 기관은 자신들의 역량을 강화하기 위해 더 많은 데이터와 정보를 원하게 된다. 결국 빅데이터는 빅브라더라는 일종의 사회적 괴물을 만들어낼 수밖에 없다. 특히 빅데이터는 개인과 관련된 정보가 축적되기 때문에 개인정보의 유출과 사생활 침해라는 문제가 더욱 확산될 것이다.

제러미 벤담(Jeremy Bentham)은 원형 감옥을 예로 들어 감시사회를 언급하였다. 우리는 현재 우리가 살고 있는 사회를 '판옵티콘'의 사회라 부르고 있다. 판옵티콘(Panopticon)은 원형감옥의 형태로 감시자의 시선을 볼 수 없는 죄수들은 감옥에서 제시하는 규율을 철저하게 따른다. 감시자들이 보이지 않는 어두운 곳에서 끊임없이 죄수들을 감시하기 때문이다. 국가인권위원회는 일반인들이 하루에 평균 80회 이상 카메라에 촬영된다고 발표하였다. 즉, 언제, 어디서나, 어디에 머물렀는지 이동경로와 머무는 장소를 확인할 수 있다. 최근 범죄자들을 구속하는데 감시카메라의 역할이 절대적인 기여를 하고 있다.

이와 같은 감시시대를 판옵티콘의 시대라 할 수 있다. 우리가 느끼지 못하는 사이에 우리를 감시하는 시대적 환경이 조성된 것이다. 빅데이터는 판옵티콘 시대를 넘어 슈퍼 판옵티콘을 조성할 우려가 있다. 빅데이터로 쌓이는 대량의 데이터는 상당수가 개인과 관련된 것이기 때문에 '프라이버시' 침해문제가 유발될 것이다. 따라서 사생활 노출은 물론 감시와 처벌이 끊임없이 재생되는 시대의 서막을 알리고 있다.

일례로 2011년 10월 국가인권위원회는 카카오톡의 강제적이고 고압적인 개인정보수집 행위가 개인정보자기결정권을 침해한다고 지적하였고 이는 헌법재판소의 판결[6]에 근거한 것이다. 또한 헌법재판소는 정보통신망법에 대한 위반의 소지가 있다고 판단하여 방송통신위원회에게 모바일 개인정보에 대한 적절

6) 헌법재판소 2005.05.26. 99 헌마 513

한 조치를 취하고 개인정보 처리에 대한 가이드라인을 제정할 것을 권고 하였다. 특히 카카오톡은 '플러스 친구'서비스를 통하여 수집된 개인정보를 모바일 광고를 위한 제3자에게 제공하면서 '선 동의 후 거부방식'을 적용하였다. 이는 개인정보 보호법 제22조 제2항에 규정된 선택적 동의의 획득과정에서 옵트인(opt-in) 방식이 아닌 옵트 아웃(opt-out) 방식을 적용하여 개인정보자기결정권을 침해할 소지를 야기하였다.

또한 2011년 4월 모바일 광고를 목적으로 위치 정보를 무단으로 수집하여 2억 건이 넘는 개인정보를 빼돌린 개인정보 유출사례가 발생하였다. 이 사건은 인기 앱인 '대리운전'등에 위치추적 프로그램을 탑재하여 배포하는 수법을 이용하였다. 유출된 개인정보는 모바일 광고에 이용되었는데 사용자가 심야에 유흥가에 있으면 대리운전이나 술집을 광고하고, 점심시간에는 사무실 밀집지역에 있는 식당 등의 위치와 연락처를 앱 화면 하단에 표출하였다. 이를 통해서 모바일 광고업체는 6억 원 이상의 부당이익을 얻었다. 이러한 불법 광고가 가능했던 이유는 휴대전화 번호만 알면 날짜와 시간대별로 단말기의 좌표가 나타나기 때문에 이 좌표를 구글, 네이버, 다음 등에서 무료로 제공하는 API(Application Programming Interface)에 입력하면 오차범위 1m 이내의 위치를 지도상에 표시해 준다. 이는 기존 휴대전화 위치정보의 오차 범위(50m)와 비교할 수 없을 정도로 정밀해 사용자가 방문한 상점이나 숙박업소 등의 이름까지 추적이 가능하다. 게다가 GPS 기능을 꺼도 스마트폰 전원을 차단하지 않으면 위치가 실시간으로 수집된다.

개인정보나 사생활 침해와 관련된 문제를 해결하기 위해서 '빅데이터 관리 및 데이터 활용에 관한 법률'을 제정하여 기업이나 기관이 저장하고 있는 다양한 데이터를 활용할 때 적정한 기준을 마련해야 한다. 개인정보와 사생활의 존중은 민주주의의 핵심 가치이기 때문에 법률적 보호가 필요하다. 빅데이터와 관련된 법률을 제정하여 규제함으로써 빅브라더의 탄생을 방지하고, 부정적 결과를 최소화해야 한다. 이러한 역할은 국가적 차원의 규제가 필요한 부분이며, 국가 역시 빅데이터 활용에서 나타나는 사회적 문제를 최소화하기 위해서 데이터 저장, 활용, 교류에 대한 원칙이 필요하다.

빅데이터의 대부분은 SNS를 통해서 제공되는 정보로 지식정보와 함께 정서

적인 공감에 바탕을 둔 감성적인 정보가 큰 비중을 차지하고 있다. 감성적인 정보에 대한 축적을 통해 소비자들의 라이프스타일을 파악할 수 있다. 인간의 감성적인 정보와 이성적인 정보는 사람들의 생각을 의미하며, 빅데이터 분석은 인간의 생각을 읽어내는 작업이 된다. 따라서 인간의 생각을 읽는다는 것은 행동을 규제할 수 있음을 의미하고, 행동에 대한 규제는 결국 자유에 대한 억압 가능성을 나타낸다. 이와 같은 개인정보 유출과 사생활 침해로 인한 판옵티콘의 확장은 시대의 산물이지만, 인간의 자유를 어떻게 보장할 것인지 보다 구체적이고 적극적인 고민이 필요하다.

□ 데이터 저작권의 문제

저작권법 제2조 19항과 20항은 데이터베이스에 대해서 저작권법 적용을 인정하고 있다. 19항에서 데이터베이스는 소재를 체계적으로 배열 또는 구성한 편집물로서 개별적으로 그 소재에 접근하거나 그 소재를 검색할 수 있도록 한 것을 말한다. 아울러 20항은 '데이터베이스 제작자'는 데이터베이스의 제작 또는 그 소재의 갱신·검증 또는 보충(이하 '갱신 등'이라 한다)에 인적 또는 물적으로 상당한 투자를 한 자를 말한다고 명시하고 있다. 저작권은 크게 저작인격권과 저작재산권으로 분류되는데, 데이터의 경우 저작재산권으로 볼 때 복제권, 공연권, 공중송신권, 전시권, 배포권, 대여권, 2차적저작물작성권 등에 대한 권리가 원 저작자에게 있다. 이러한 저작권 이용은 제35조 3에 의해서 저작자의 정당한 이익을 부당하게 해치지 않는 보도·비평·교육·연구 등을 위하여 저작물을 무료로 이용할 수 있다. 하지만 그렇지 않을 경우 영리성 또는 비영리성 등 이용의 목적 및 성격에 따라서, 저작물의 종류 및 용도에 따라서, 이용된 부분이 저작물 전체에서 차지하는 비중과 그 중요성에 따라서, 저작물의 이용이 그 저작물의 현재의 시장가치나 잠재적인 시장 또는 가치에 미치는 영향에 따라서 저작권자의 경제적 권리를 인정해야 한다.

미국, 독일 등을 중심으로 온라인 데이터베이스 (DB) 저작권 침해 논쟁이 활발하게 전개되고 있는 가운데 국내에서도 인터넷 등 온라인 서비스를 통해 제공되는 데이터를 저작권 대상에 포함 시켜야 한다는 주장이 제기되고 있다. 저작권심의 조정위원회(2012)는 온라인 서비스상의 데이터도 저작권 보호를 받아야 하며, 이를 위해서 입법화 한다는 방침이다.

빅데이터는 저작권법 적용 대상이 될 수밖에 없으며, 현재 방송이나 책은 물론 모든 창작물은 저작권이나 특허권에서 자유롭지 않다. 빅데이터 역시 최초 데이터 생산자, 유통사, 2차 정보 생산자 등과 관련해 저작권 논란에 휩싸일 수 있다. 최초 데이터 생산자는 고유의 데이터를 비상업적으로 제공하였지만, 데이터를 분석하여 시각적인 정보를 제공하는 2차 데이터 생산자가 상업적으로 판매할 경우 저작권 분쟁이 발생할 수 있다. 따라서 학술적 용도나 개인적 용도가 아니라, 상업적 목적을 갖고 데이터를 이용할 경우를 대비하여 저작권 문제를 해결해야 한다.

빅데이터와 관련된 저작권 문제는 사업의 활성화, 빅데이터 공유, 글로벌 경제력 확보, 사회기여도 등을 명분으로 저작권 집중관리 시스템을 고려해볼 필요가 있다. 저작권 집중관리 시스템의 사례로 음원을 들 수 있다. 음원의 경우 한국음악저작권협회, 한국음악실연자협회 등에서 저작권을 집중관리하여 저작권료를 징수하고 분배한다. 음원사례에서 보듯이 데이터 집중관리 기구를 만들어 유통되는 데이터의 저작권을 관리하고, 계약에 의해서 수익의 일부를 사회에 환원할 수 있는 공익적 성격의 기구를 설립해야 한다. 따라서 데이터 활용과 이용에 관한 자율 규정 및 동의서를 작성하고, 2차 저작물에 따른 저작권 침해에 대한 예외 사항을 최대한 보장해야 한다. 아울러 기업이 모든 정보를 저장할 수 없기 때문에 필요한 정보에 대해서 상호 교환이나 이용에 대한 협조가 필요하다. 특정기업 이나 집단이 정보폐쇄(Information Silo) 방침을 갖고 있을 경우 빅데이터의 활용영역은 매우 위축될 수밖에 없다. 이러한 경우 역시 교류의 조건은 저작권료나 이용에 따른 비용을 지출해야 한다.

□ 네트워크 블랙아웃

Gartner(2011)는 데이터가 21세기 원유와 같은 역할을 할 것이며, 빅데이터는 향후 주목해야 할 이머징 기술이라고 하였다. 또한 PWC(2010)는 빅데이터가 새로운 비즈니스 가치창출의 핵심 키워드가 될 것으로 예견하였다. 빅데이터에 대한 이러한 청사진은 데이터 저장기술, 데이터 분석 소프트웨어, 데이터 해석인력, 적용분야 개발이라는 단계적 발전이 이루어져야 한다. 그렇지 않을 경우 빅데이터는 일종의 사회적 유행에 머물 수도 있다. 뿐만 아니라 기술적

차원에서 최근 나타나는 데이터 트래픽 현상과 DDoS나 해킹에 철저히 대비해야 한다. 특히 DDoS나 해킹으로 '네트워크 블랙아웃'현상이 발생할 가능성이 있다. 네트워크 블랙아웃 현상이 발생하면 네트워크 기반으로 연결돼있는 금융·교통·의료·국방 등에서 심각한 문제가 발생할 뿐만 아니라 의료, 산업, 국방, 경제 등 전 분야에서 혼란이 발생하게 된다.

2016 국가정보보호백서에 의하면 DDoS 공격과 이에 대한 대응·탐지 건수는 2009년 53건에서 2015년 502건으로 빠른 증가세를 보이고 있다(표 1.7 참조).

표 1.7 DDoS 대응 탐지현황

구분	2009	2010	2011	2012	2013	2014	2015
탐지·대응	53	239	290	318	415	457	502

출처 : 2016 국가정보보호백서

얼마전 KT 이용자 870만 명의 개인정보가 해킹으로 유출되었다. 이와 같이 DDoS나 해킹으로 네트워크 블랙아웃현상이 발생할 가능성은 점점 높아지고 있다. 발생한 디도스 공격의 35%가 정치적 혹은 이념적 동기, '허무주의 혹은 파괴주의에 근거(31%)', '온라인 게임 타깃(29%)', '실력과시(25%)' 등이 DDoS 공격 동기로 나타났다(호애진, 2012). 즉, 누구나 쉽게 적은 비용으로 개인 PC는 물론 서버에 DDoS 공격이 가능하다.

트래픽으로 인해 네트워크 블랙아웃 현상이 발생할 가능성은 매우 낮지만, 지속적인 망관리가 이루어지지 않는다면 영국과 일본의 사례에서 볼 수 있듯이 사회적 혼란을 경험하게 될 것이다. 트래픽으로 인해 2012년 7월 14일 영국 통신사 O2의 네트워크가 24시간 동안 다운되었고, 프랑스 오렌지가 9시간 동안 블랙아웃 되는 현상이 발생하였다. 일본에서도 올해 초 NTT 도코모가 수차례 다운되는 등 스마트폰 사용률이 높은 세계 곳곳의 통신사 네트워크가 멈추는 사고가 이어지고 있다(심나영, 2012). 이러한 현상은 결국 스마트폰 사용 급증으로 인해 문자, 동영상, 그림 등과 같은 비정형화된 데이터 사용이 급증하면서 나타나는 현상이라 할 수 있다.

빅데이터 시대에 기업이나 정부가 축적한 데이터를 분석할 때 DDoS 공격으로 인해 네트워크 혼란이나 마비가 초래될 수 있으며, 해킹을 통해서 잘못된 데이터를 제공하거나 데이터를 변경할 경우 엄청난 경제적 손실이 나타날 수 있음을 염두에 두어야 한다.

1.1.3 빅데이터의 특징[7)]

빅데이터를 특징짓는 차원은, 학문적 의미에서 소개한 규모(volume), 다양성(variety), 속도(velocity)를 의미하는 3V에서 새로운 V의 추가로 인한 개념의 확장으로 볼 수 있다. 3V는 빅데이터의 핵심적인 특징을 포함하고 있지만, 새로운 특징들이 계속해서 등장하고 있다. 1.1.1절에서 3VC와 중복되는 면이 있으나 지속적인 빅데이터의 특징을 개념적으로 추가해나가는 과정의 설명이므로 새로운 빅데이터의 속성들을 소개하도록 한다.

표 1.8 빅데이터의 특징

구분	내용
Volume (크기)	세상의 90%의 데이터가 최근 2년 동안 만들어졌다고 한다.
Velocity (속도)	데이터의 생성, 저장, 분석, 이동 등의 속도가 매우 빨라지고 다양한 단말장치에서 발생하는 데이터 또한 매우 빠르게 생성되고 있다.
Variety (다양성)	text 뿐 아니라 pdf, photo, audio, video, email, web, sms, sns, transaction 데이터 등장. 더 나아가 IoT 세상의 다양한 센서 데이터.
Veracity (정확성)	데이터가 깨져있는 것은 일상적이다. 따라서 빅데이터 분석을 위해 데이터를 감사하고 보정하는 작업은 매우 중요하다.
Variability (가변성)	동일한 데이터가 서로 다른 컨텍스트(context)에서 다른 의미를 가질 수 있다.
Volatility (휘발성)	현재 생성되는 모든 데이터를 저장하기 위한 비용이 매우 크기 때문에 대부분의 데이터는 버려진다.
Visualization (시각화)	데이터의 의미를 보여줄 수 있는 시각화 기술
Value (가치)	크기만 한 것이 아니라 가치가 내제되어있다
Viscosity (연계성)	데이터 간의 상관관계를 의미
Virality (확산성)	데이터의 본제, 전달, 확산을 의미
Viability (실행가능성)	데이터가 활용목적에 맞는 가치를 갖고 있어야 하며, 이를 찾을 수 있어야 한다.
Vicinity (인접성)	데이터가 획득 가능하여야 한다.
Virtue (미덕)	빅데이터 취득 및 활용에 있어서 도덕적인 방법이 적용되어야 한다

출처: BDT insights, 디지털타임즈

7) 본 절의 내용은 IBM 글로벌 비즈니스 서비스의 요약보고서인 빅데이터의 현실적인 활용(기업이 불확실한 데이터에서 가치를 창출해 내는 방법)의 내용을 옮겨오거나 발췌하여 정리한 것임

새롭게 등장한 V들을 나열해보면, 특정 유형의 데이터에 내재되어 있는 불확실성을 인식하고 관리하는 일의 중요성을 강조하기 위한 질적인 측면의 정확성(Veracity), 데이터의 분석을 통해 얻을 수 있는 궁극적 가치를 의미하는 가치성(Value), 사용자 친화적인 시각적 기능을 통해 빅데이터의 모든 잠재력 활용을 의미하는 시각화(Visualization), 끊임없이 변화하는 데이터를 의미하는 가변성(Variability) 등이 있고 이들은 비교적 일찍 등장하였다. 최근에 언급되는 것들로는 주요 데이터로써의 가능성을 의미하는 생존성(Viability), 대규모 자료일수록 저장의 부담으로 인해 발생하는 휘발성(Volatility), 데이터 간의 관계를 의미하는 연계성을 의미하는 (Viscosity), 데이터의 확산성(Virality) 등이 있으며 이들을 표 1.8에 정리하였다.

표 1.8의 빅데이터 특성 들은 최근의 소개 된 것들을 모아놓은 것들 이므로 모두 검증된 것이라고 볼 수는 없다. 일반적으로 잘 알려진 것들을 중심으로 다시 자세히 정리하였다.[8)]

① Volume : 지구상 총 데이터 중 90%는 지난 2년 간 생산

해당 분야에 전문가가 아닌 이상 대부분의 사람들에게 빅데이터는 'BIG'이라는 접두어 때문에 대용량의 데이터를 다루는 무언가로 알고 있다. 즉 데이터의 크기가 빅데이터의 정의로 널리 사용되고 있는 것이다. 최근 '기가바이트'나 '테라바이트', '페타바이트' 등의 대용량을 나타내는 용어들이 점차 일반화되는 것은 빅데이터 정의에 있어 용량에 주목하고 있다는 것을 잘 나타내고 있다고 볼 수 있다.

현재 지구상에 존재하고 있는 모든 데이터의 90%가 지난 2년간 생성됐고, 앞으로 2020년까지 2년마다 그 데이터량이 두 배씩 커질 것이라는 보고가 있는 것을 보면 데이터량이 더욱 중요해진다는 것은 어렵지 않게 짐작할 수 있다. 미국의 IT 전문기관인 TDWI가 실시한 2012년 설문조사에 의하면 IT 관련 종사자의 많은 부분은 3~10 테라바이트의 데이터를 분석한다는 결과를 보였다.

8) 이후의 내용은 BigData Transfer Insight의 관련내용을 옮겨오거나 발췌하여 정리한 것임

이렇게 엄청난 데이터는 최근 모바일 기기의 발전과 사물인터넷 등이 급격히 발전하면서 기하급수적으로 늘고 있는 추세다. 따라서 과거에 주로 수집했던 정형 데이터는 빅데이터 분석 범주의 일부분으로 밀려나고 있다. 이제는 존재하는 모든 데이터의 약 90% 이상이 소셜미디어나 웹상의 클릭스트림, 센서 등을 통해 수집되는 비가공 데이터, 즉 비정형 데이터다. 비정형 데이터의 전형적인 예로 SNS 등의 소셜미디어를 들 수 있다. 세계 최대의 가입자를 보유하고 있는 페이스북의 경우 사용자들은 글은 물론, 사진과 음악 등을 페이스북에 업로드 하는데, 그 저장량은 하루에 500 테라바이트에 이른다.

센서 등으로부터 얻어지는 데이터를 분석하여 활용하는 사물인터넷 역시 각광을 받고 있다. 일례로 미국의 경우 비행기 엔진에 센서를 부착해 정보를 받아 분석하여 고장 예측 등으로 활용하고 있는데, 이때 엔진으로부터 만들어지는 데이터는 연 약 25억 테라바이트에 이른다. 또한 자동으로 운전할 수 있는 자동차의 경우 센서를 통해 연 약 2 페타바이트의 데이터가 생산된다.

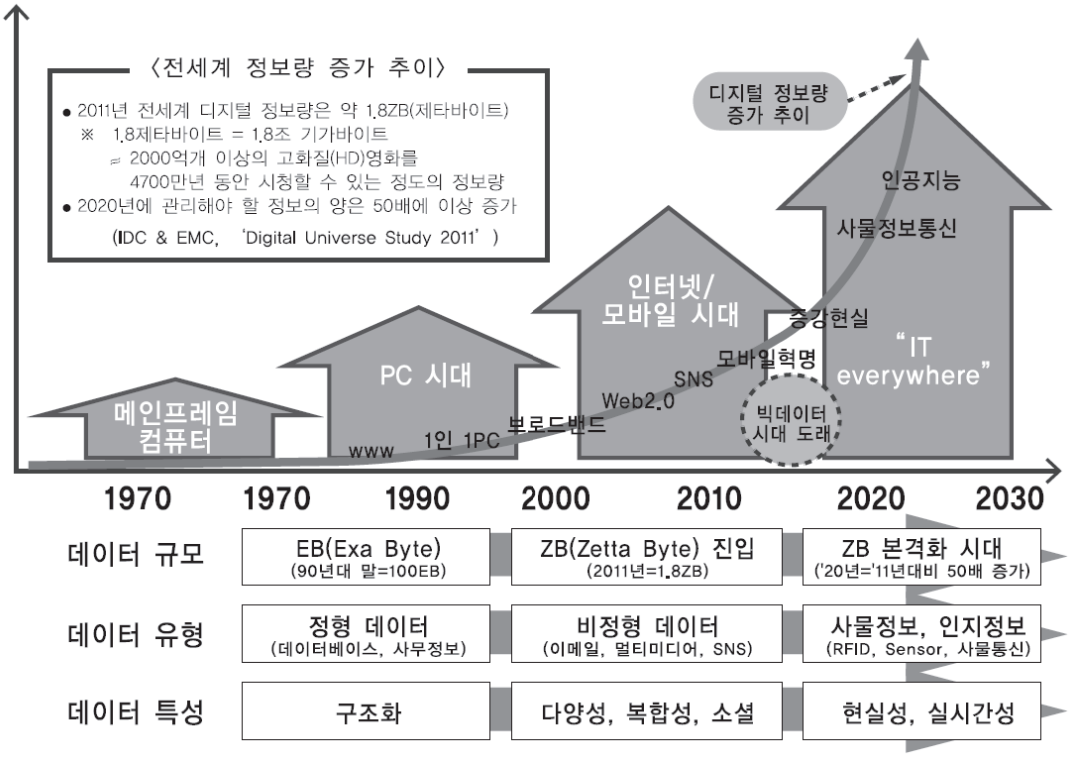


그림 1.7 빅데이터 규모의 변화

출처 : 새로운 미래를 여는 빅데이터 시대

② Variety : 데이터 수집 경로는 웹 접근 가능한 모든 곳

이렇게 방대한 데이터는 앞서 설명했다시피 너무나 다양한 경로를 통해 수집된다. 인터넷이 보급된 후 한참이 지난 2000년대 중반까지만 해도 기업이나 기관에서 활용하기 위해 분석하는 데이터는 전문가 또는 관계자 등이 제공하는 체계적인 정형 데이터가 대부분을 차지했다. 이러한 데이터는 표나 그래프 등을 통해 보기 좋게 가공되기 때문에 분석이 편한 반면, 정보의 범위가 지극히 제한적이었다. 하지만 2000년대 후반 들어 태블릿 PC와 스마트폰의 급격한 보급, 인터넷 활용성 확장 등은 웹상에서 수집할 수 있는 데이터의 경로를 급격히 확장시키는 도구가 되었다. 또한 센서 기술의 발달로 인해 앞서 봤던 비행기 엔진은 물론, 채굴 장비나 자동 운전 자동차, 물류회사의 트럭 및 비행기에 부착된 센서 등으로부터도 데이터 수집이 가능해졌다.

이러한 도구를 통해 수집되는 데이터는 이제 정형 데이터를 넘어, 반정형 데이터, 비정형 데이터로 세분화되어 분석에 활용되기에 이른다. 이 정보는 IBM이 정의했던 바와 같이 기존의 디바이스로는 수집과 분석이 불가능하기 때문에 빅데이터로 정의할 수 있는 것이다. 또한 수집 경로 역시 군수, 기상, CCTV 등 웹으로 통합이 가능한 어떤 곳도 가리지 않고 확장되고 있는 추세다.

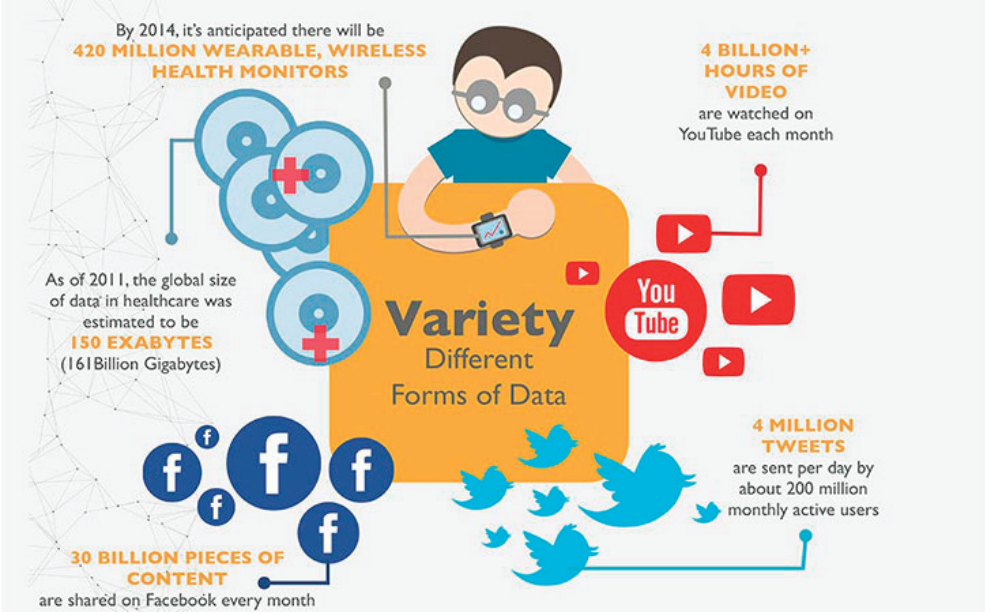


그림 1.8 빅데이터의 다양성

출처: http://adec-it.com/assets/img/technology/big-data/variety.jpg

③ Velocity : 빅데이터 분석의 성공을 좌우하는 '속도'

수집되는 데이터는 다양한 경로로부터 분석을 수행하는 기업이나 기관에게 얼마나 빠르게 전송되는지도 중요하다고 할 수 있다. 예를 들어 얼마 전 발생한 보스톤 마라톤 대회 테러 사건의 범인을 검거를 위해 10 테라바이트의 CCTV 자료 분석이 이뤄졌다. 이 때 만약 그 데이터가 신속하게 전달되지 못했다면 분석 전에 범인은 추적 반경 밖으로 도주했을지도 모른다.

빅데이터를 이러한 속도 측면으로 접근하는 것을 'Velocity'라고 한다. 이는 데이터가 생성되고, 저장되며, 시각화 되는 과정이 얼마나 빠르게 이뤄져야 하는 지에 대한 정의다. 과거에는 이러한 데이터를 일괄로 처리하는 방식이 일반적이었다. 즉 수집 데이터를 하루 또는 일주 단위로 업데이트하는 식이었다. 따라서 컴퓨터나 서버는 새로운 데이터베이스로 갱신하는데 많은 시간이 걸렸다. 하지만 최근 인터넷의 발달로 데이터의 전송은 상상을 초월한다. 예를 들어 유튜브로 전송되는 동영상은 분당 약 100시간 분량이다. 또한 매 분마다 2억 통의 메일이 보내지고 있고, 2천만 개의 사진이 포스팅 되며, Flicker에 업로드 되는 사진은 3만 개에 달한다. 또한 매 분마다 거의 30만 개의 트위터 메시지가 전송되고 있다. 따라서 빅데이터의 정의에 있어 이러한 실시간 정보를 얼마나 빠르게 분석하여 의사결정에 반영하는지에 대한 사항은 너무나 중

요한 측면이라 할 수 있다.

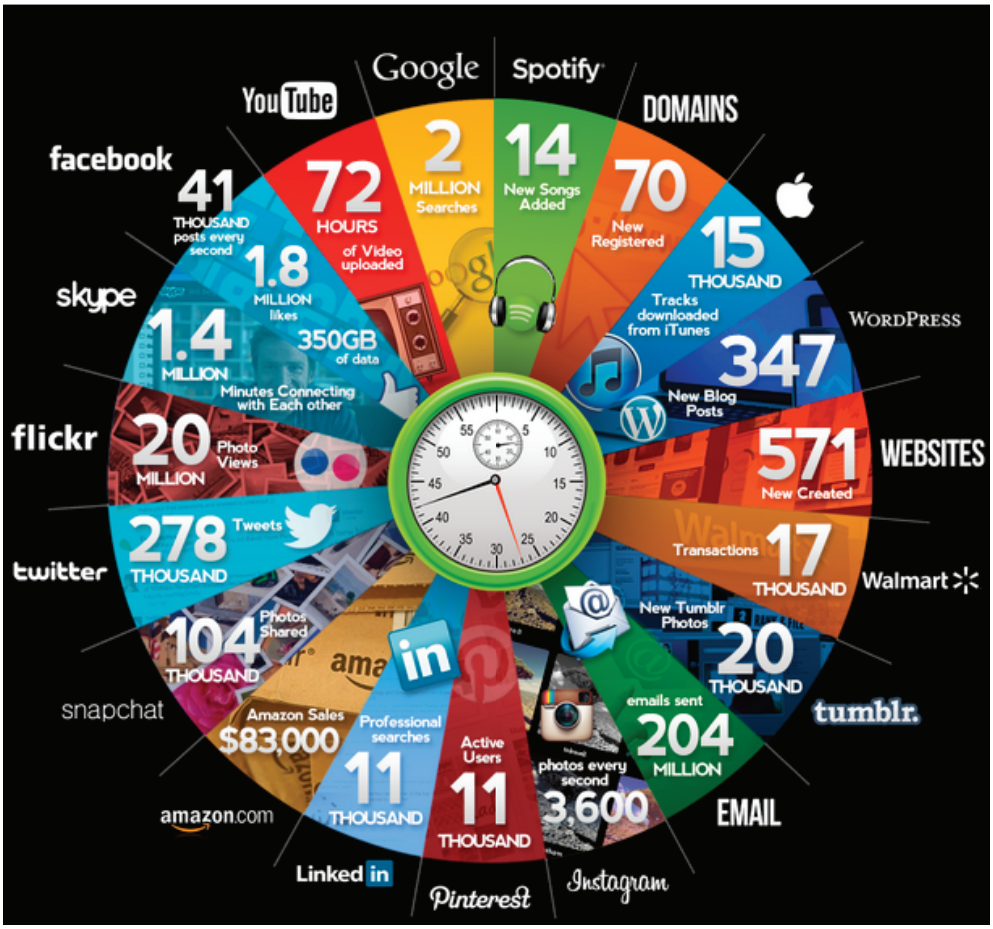


그림 1.9 빅데이터의 생성 속도

출처 : http://cfile3.uf.tistory.com/image/

④ Value : 빅데이터 분석, '과연 돈 될까?'

컨설팅 전문 업체인 맥킨지에 따르면 미국의료보장에 적용되는 빅데이터의 잠재가치는 연간 3조 달러에 이르는 것으로 조사됐다. 또한 유럽 정부 부문에 적용된 빅데이터 시스템으로 인해 연간 약 2억 5천억 유로의 잠재가치를 창출할 수 있을 것이라고 밝혔다. 물론 이는 일반인들에게도 적용된다. 맥킨지 리포트는 전 세계적으로 수집 가능한 지역 정보를 이용해 발생하는 잠재적인 소비자 잉여는 금액으로 2020년까지 6조 달러에 이를 것이라고 전망했다.

이러한 사항들은 빅데이터가 충분한 가치를 낼 수 있는가에 대한 것들이다.

특히 기업의 입장에서는 엄청난 자금을 투자해 빅데이터 관련 인프라를 갖추기 때문에 그 이상의 가치 창출이 무엇보다 중요한 요소라 할 수 있다. 하지만 이러한 성공적인 예가 존재함에도 불구하고 실제 기업에서 가시적인 성과가 있는지에 대해서는 아직 논란이 분분하다.

기업이 확인할 수 있는 가치 창출의 효과는 금전적인 이익이나 기업의 평판, 대상 고객의 증가 등이라고 할 수 있다. 따라서 최근 이뤄지고 있는 소셜미디어 데이터를 이용한 감정분석이나 사용기기에 적용된 센서로 부터의 데이터 분석 등이 어떻게 기업이 원하는 가치로 전환해야 할지에 대한 전략이 중요한 이유가 여기에 있다.



그림 1.10 빅데이터의 가치 창출
출처 : https://media.licdn.com/mpr/mpr/shrinknp_800_800

⑤ Veracity : '대수의 법칙'의 딜레마

흔히 통용되는 '대수의 법칙'에 따르면 표본이 많을 수록 일정한 경향성은 강해진다. 이 법칙을 확인하는데 있어 빅데이터만큼 좋은 경로도 없을 것이다. 일정한 경향을 분석하기 위해 분석해야 할 데이터가 넘치는 것을 넘어, 한 기업이나 기관이 다루기 불가능할 정도이기 때문이다. 문제는 이러한 엄청난 비정형 데이터데이터가 과연 일정 패턴을 설명할 수 있을 만큼 신뢰성이 있는가이다. 역시 대수의 법칙에 따라 데이터가 많을 수록 신뢰하지 못하는 엉터리

데이터도 많아질 가능성이 높아지기 때문이다. 따라서 빅데이터를 분석하는데 있어 기업이나 기관에게 수집한 데이터가 정확한 것인지, 분석할 만한 가치가 있는지 등을 살펴야 하는 과정이 추가되는 것이다. 하지만 기술 발달은 데이터의 수집과 분석을 최대한 자동으로 수행 가능하게 만들었다. 이는 빅데이터 분석의 활용에 있어 정확성은 간과되기 쉽다는 것을 의미한다.

이 지점을 간파해서 일까. 한 조사에 의하면 비즈니스 리더 3명 중 한 명은 의사결정을 위해 사용하는 비정형 데이터를 신뢰하지 않는다고 답했다. 과거에 비해 인터넷에 글을 쓰고, 자신의 의사를 밝히며, 다양한 자료를 제시할 수 있는 방식이 너무나 수월해졌기 때문에, 비전문가가 제공하는 정보 역시 급격히 늘어나게 된 것이 큰 이유라 할 수 있다. 이에 따라 정보의 '진실성'은 빅데이터를 정의하는데 빠져서는 안 될 개념으로 자리 잡았다.

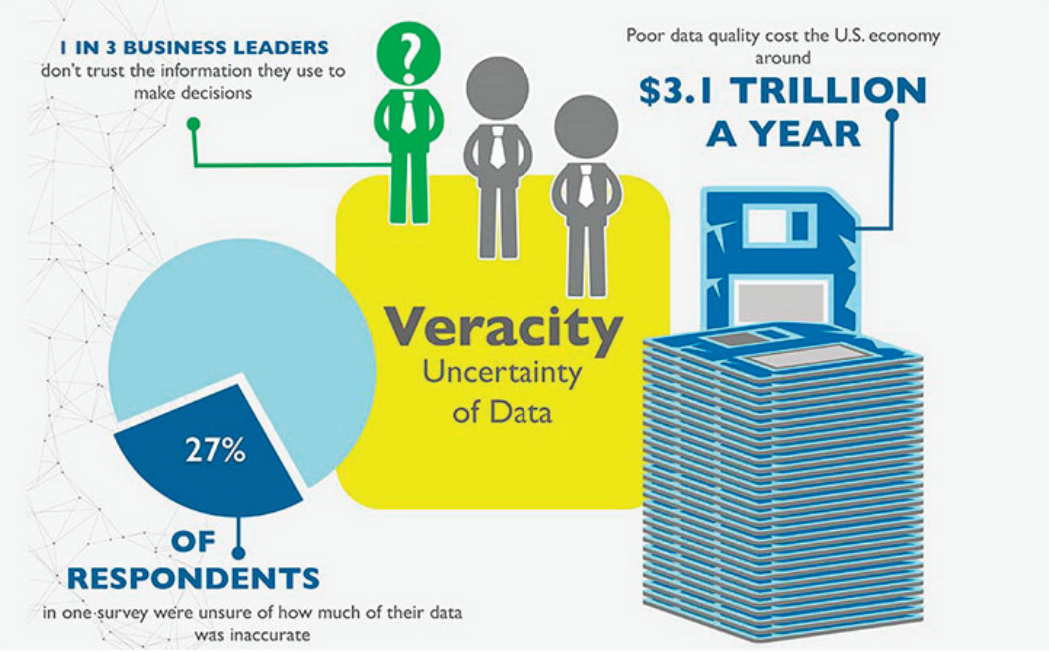


그림 1.11 빅데이터의 진실성
출처 : http://adec-it.com/assets/img/technology/big-data/veracity.jpg

⑥ Variability : 데이터에는 '맥락'에 따라 의미가 달라진다.

분석해야 할 데이터의 진실성과 관련하여 또 하나의 중요한 문제는 그 데이터의 의미가 수시로 변한다는데 있다. 최근 소셜미디어의 확산으로 자신의 의

견을 웹을 통해 자유롭게 게시하는 것이 쉬워졌다. 하지만 실제로 만나 대화하는 것과는 달리, 자신의 생각을 글로 표현하게 되면 상황 맥락이 생략되기 때문에 의도를 오해하기 쉽다. 예를 들어 '풍경이 너무 좋다'는 표현이 실제로는 풍경에 대한 실망을 반어법으로 표현할 것일 수도 있다.

일반적으로 빅데이터 분석은 시스템을 구축하고 프로그램에 따라 움직이기 때문에 인터넷 상의 개인적인 글들을 대상으로 감정분석 등을 수행할 때 상황적 맥락을 놓치는 경우가 많다. 그렇다고 엄청나게 많은 각 글들을 일일이 사람이 살피는 것도 불가능하다. 또한 아무리 기술이 발전하더라도 감정에 따른 맥락을 알고리즘화해 시스템에 적용하는 것은 너무나 어려운 작업일 것이다. 따라서 가변성의 확보는 빅데이터의 중요한 일부로 추가할 수 있게 된 것이다.

⑦ Visualization : 빅데이터 분석 결과는 보기 쉬워야 한다.

빅데이터 분석과 활용 등의 흐름을 넓게 보면, 정형 및 비정형 데이터의 수집으로 시작한다. 이 데이터를 바탕으로 복잡한 분석을 수행한 후 각각의 용도에 맞는 정보로 가공하게 된다. 이때 중요한 것은 정보의 사용 대상자가 쉽게 이해할 수 있어야 한다는 것이다. 그렇지 않으면 정보의 가공을 위해 소모된 시간적, 경제적 비용이 무용지물이 될 수 있기 때문이다.

즉, 빅데이터의 시각화가 중요하다고 할 수 있다. 최근 인포그래픽 관련 서비스를 제공하는 업체가 생겨나는 현상은 이러한 인식을 잘 반영한다고 볼 수 있다. 하지만 빅데이터 분석이 너무 복잡하기 때문에 단순 시각화하는 것만으로는 부족하다. 다행이 그래픽 관련 기술도 함께 발전이 이뤄지고 있기 때문에, 시간이 갈수록 복잡한 분석 결과를 이해하기 쉽게 시각화할 수 있는 방법은 점차 많아지고 있다.

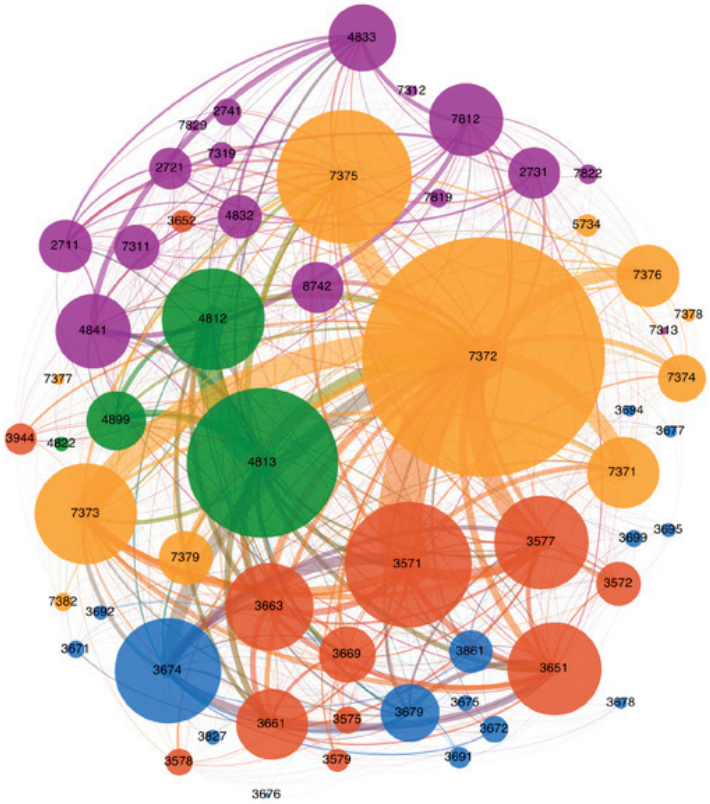


그림 1.12 빅데이터의 시각화

출처: https://static01.nyt.com/images/

살펴본 바와 같이 빅데이터는 어떠한 한 분야를 나타내는 개념으로 사용하기 어려운 측면이 있다. 분석과 활용 과정에 필요한 사항을 설명할 수는 있겠지만 사용 분야마다 중요하게 생각하는 부분은 다를 수 있고, 심지어는 어떤 분야는 위에 살펴본 사항 중 필요 없는 것도 있을 수 있다. 이러한 점이 빅데이터의 정의를 시원하게 답하기 어렵게 만든다. 그럼에도 불구하고 '인간'에 대한 포괄적인 이해가 없다면 특정 인물에 대한 설명도 불가능 하듯이 빅데이터의 포괄적인 개념의 정립이 특정 빅데이터 프로젝트가 성공할 수 있는 초석이라 할 수 있다.

1.2 행정자료의 정의

행정자료의 통계적 활용에 대해서 논의하기 위해서는 우선 행정자료가 무엇인가를 정의할 필요가 있다.

1.2.1 학문적 정의

행정자료를 문자 그대로 해석하면 '행정과정에서 생산된 자료'를 의미하는 것으로 볼 수 있지만, 우리나라의 공식적인 국어사전인 국립국어원의 표준국어대사전에도 고유명사로써의 '행정자료'라는 단어를 별도로 정의하고 있지 않다. 따라서 '행정'과 '자료' 각각에 대해 살펴봄으로써 우리가 마치 복합명사처럼 관용적으로 쓰고 있는 '행정자료'의 의미를 정의할 수 있을 것이다.

먼저 '행정'의 사전적 의미[9]를 살펴보면, 일반적으로 '정치나 사무를 행함'을 의미하지만, 법률적으로는 "국가 통치 작용 가운데 입법 작용과 사법 작용을 제외한 국가 작용" 또는 "법 아래에서 법의 규제를 받으면서 국가 목적 또는 공익을 실현하기 위하여 행하는 능동적이고 적극적인 국가 작용"이라고 정의하고 있다. 전자는 행정법학적 행정개념 중 삼권분립적 공제설(소극설)에 후자는 국가목적실현설(적극설)에 각각 근거를 둔 정의이다. 따라서 행정의 정의를 보다 심도 있게 파악하기 위해서 행정법학, 행정법, 행정학 분야에서 행정을 어떻게 에서 정의하고 있는가를 살펴볼 필요가 있다.

행정법학에서 행정의 개념은 근대 법치국가의 권력분립이론을 배경으로 성립되었다. 초기에는 행정을 입법·행정·사법의 권력구조에서 입법기관에서 성립된 법령을 구체화하는 수단으로 인식했으나, 현대국가에서 종래의 삼권분립 구조에 중대한 변화가 생기고 행정의 역할이 확대됨에 따라 행정의 개념 또한 많이 변화하게 되었다. 현재 행정의 개념을 정치권력을 배경으로 공공정책의 형성 및 구체화를 이룩하려는 공공성을 띤 인간의 협동적인 집단행동으로 정의하는 견해[10]가 타당해 보인다.

9) 여기에서는 각종 용어의 정의는 국립국어원의 표준국어대사전을 기준으로 함
10) 박동서, 한국행정론, 법문사, 1997, p68; 김규정, 행정학원론, 법문사, 1998, p12

행정법에서 행정은 크게 형식적 의미의 행정과 실질적 의미의 행정으로 구분된다.[11] 형식적 의미의 행정은 국가기관을 기준으로 한 개념으로서, 행정부에 속하는 기관에 의해 이루어지는 모든 작용을 의미한다. 실질적 의미의 행정은 국가작용의 성질상 차이를 전제로 행정내용의 본질적 특성을 중심으로 구성되는 개념으로, 실질적 의미의 행정개념이 정립될 수 있느냐에 대해 긍정설과 부정설로 나뉜다(표 1.9참조).

긍정설 중 소극설은 행정을 "입법과 사법을 제외한 국가작용의 전부"라고 정의한다. 적극설은 행정에 대해 "국가의 목표와 법질서의 달성을 위한 국가의 활동"등 다양한 정의를 하고 있으며, 이중에서 "법아래서 법의 규제를 받으면서 현실적·구체적으로 국가목적의 적극적 실현을 위하여 행해지는 전체로서 통일성을 가진 계속적인 형성적 국가활동"이라는 정의[12]가 많은 지지를 받고 있는 것으로 보인다. 국립국어원 표준국어대사전의 행정의 의미도 이를 따르고 있다.

행정학에서 수많은 학자들이 행정에 대하여 행정관리설, 통치기능설, 행정행태설, 발전기능설, 정책화기능설 등 다양한 정의를 내리고 있다. 이는 행정의 본질적 개념에서의 차이라기보다는 행정의 어떤 측면을 강조하는가에 따른 차이이다. 여러 정의 중에서 나이그로(F. Nigro)의 "행정이란 공공성을 지닌 협동적인 집단행동이며, 이는 고도의 합리성을 지닌 정책과정"이라는 정의가 타당하다고 생각한다.[13]

다음으로 자료의 사전적 의미를 보면 "연구나 조사 따위의 바탕이 되는 재료" 또는, "만들거나 이루는 데 바탕이 되는 물자나 재료"라고 정의하고 있다. 또한 재료의 의미를 "어떤 일을 하기 위한 거리"라고 정의하고 있다. 따라서 행정자료를 정의하는데 있어서의 자료의 의미는 행정 또는 행정행위와 관련하여 생성된 유무형의 결과물을 의미한다고 하겠다. 이는 행정의 결과물만을 의미하는 것이 아니라, 행정의 전체 과정에서 생성된 것을 의미한다고 보는 것이 타당할 것이다. 대부분의 행정행위는 문서로 행해지거나 문서를 근거로 하기 때문에 행정자료의 대부분은 문서의 형태로 남게 된다. 오늘날 각종 정보통신

11) 홍정선, 행정법원론(상), 박영사, 2015, p4~6
12) 田中二郎, 行政法總論, 1992, p22
13) 유종해, 이덕로, 현대행정학, 2015, p7

기술이 행정에 적용됨으로써 문서는 '글이나 기호 따위로 일정한 의사나 관념 또는 사상을 나타낸 것'이라는 전통적인 의미와 형식에서 벗어나 매우 다양한 형식으로 변화하고 있다. 이에 따라 전자문서와 데이터 파일 등 새로운 형태의 행정자료가 차지하는 비중이 점차 확대되고 있다.

표 1.9 행정의 의미 구분

구분		내용
긍정설 (입법, 사법과 본질적 차이를 긍정)	소극설 (공제설)	행정이란 '국가작용 가운데서 입법·사법을 제외한 것'이라고 한다 (Jellinek, Hatschek)
	적극설 (목적설)	'국가가 그의 목적을 실현하기 위하여 그 법질서 아래에서 행하는 사법 이외의 작용', 즉 행정은 '국가가 국민의 이익을 도모하는 작용' 등으로 정의한다. 다시 말하면, 행정은 국가목적적 작용·공익목적적 작용인 점에서 입법·사법과 구별된다(Otto Mayer).
	적극설 (양태설)	'법 아래에서 법의 규제를 받으면서 현실적·구체적으로 국가목적의 적극적 실현을 위하여 행해지는 전체로서의 통일성을 가진 계속적인 형성적 국가활동'이라든가, '법 아래에서 현실적으로 국가목적을 실현하여 구체적 결과를 가져오는 작용' 등으로 정의한다 (Fleiner, Sarway, Peters)
부정설 (입법, 사법과 본질적 차이를 부정)	법단계설	헌법의 직접적 집행인 입법과 헌법의 간접적 집행으로 입법보다 한 단계가 낮은 집행으로 나누며, 그 집행은 다시 법관에 의한 집행행위인 사법과 행정청에 의한 집행행위인 행정으로 구분된다 (A. Merkl).
	기관 양태설	헌법의 집행작용인 입법과 법률의 집행작용인 행정, 사법은 각 작용 간의 성질상의 차이는 없기 때문에 그 작용을 담당하고 있는 기관의 조직형태에서 구별기준을 찾아야 한다고 한다

1.2.2 현실적 정의

여기에서 논의하고자 하는 행정자료는 새로운 통계를 작성하거나 기존의 통계를 보완하는데 사용될 뿐만 아니라, 연관된 통계치의 수준을 비교 점검하는 데에도 폭넓게 활용되는 자료를 의미한다. 따라서 통계적 활용이란 관점에서 행정자료의 의미를 현실적으로 정의할 필요가 있다. 우선 통계법을 중심으로 한 국가통계제도를 살펴봄으로써 통계와 연관된 행정자료에 대한 논의를 전개하기로 한다. 우리나라의 통계법은 통계와 그 작성기관 등에 관하여 상세히 규

정하고 있다. 이를 정리하면 표 1.10과 같다.

표 1.10 통계법 제3조의 통계용어 정의

용어	정의
통계	통계작성기관이 정부정책의 수립·평가 또는 경제·사회현상의 연구·분석 등에 활용할 목적으로 산업·물가·인구·주택·문화·환경 등 특정의 집단이나 대상 등에 관하여 직접 또는 다른 기관이나 법인 또는 단체 등에 위임·위탁하여 작성하는 수량적 정보로써 작성기관이 내부적으로 사용할 목적 등 대통령으로 정하는 것은 제외
통계작성기관	중앙행정기관·지방자치단체·통계작성지정기관
통계작성 지정기관	법인 중 정부정책의 수립·평가 또는 경제·사회현상의 연구·분석 등에 이용되는 수량적 정보를 작성하고 있거나 작성하고자 하는 기관 중에서 지정요건을 갖춘 기관의 신청을 받아 통계청장이 지정한 기관 (통계법 제15조, 통계법시행령 제18조)
통계자료	통계작성기관이 통계의 작성을 위해 수집·취득 또는 사용한 자료(데이터베이스 등 전산자료를 포함)
행정자료	통계자료를 제외한 것으로 공공기관이 직무상 작성·취득하여 관리하고 있는 문서·대장 및 도면과 데이터베이스 등 전산자료를 의미
공공기관	중앙행정기관, 지방자치단체 및 다음 각목의 기관을 의미 가.「공공기관의 운영에 관한 법률」에 따른 공공기관 나.「지방공기업법」에 따른 지방공사 및 지방공단 다.「유아교육법」,「초·중등교육법」,「고등교육법」, 그 밖의 다른 법률에 따라 설립된 각급 학교 라. 특별법에 따라 설립된 특수법인 마.「신용정보의 이용 및 보호에 관한 법률」 제25조에 따라 금융위원회의 허가를 받은 종합신용정보집중기관

통계법은 중앙행정기관, 지방자치단체, 그리고 통계작성지정기관이 작성하는 통계만을 국가의 공식통계(official statistics)로 한정하고 있다. 따라서 이들 통계작성기관이 삼권분립의 관점에서 행정부에 속하는가하는 문제는 통계법의 적용범위와 한계를 판단하는데 중요한 기준이 된다. 통계작성기관 중에서 중앙행정기관과 지방자치단체는 각각 정부조직법과 지방자치법을 근거로 한 중앙정부와 지방정부로서 삼권분립의 관점에서 보면 행정부에 속하는 것이 명확하다.

통계작성지정기관은 법인 중 정부정책의 수립·평가 또는 경제·사회현상의 연구·분석 등에 이용되는 수량적 정보를 작성하고 있거나 작성하고자 하는 기관 중에서 지정요건을 갖춘 기관의 신청을 받아 통계청장이 지정하게 되어 있다. 이러한 통계작성지정기관은 다음과 같은 이유로 넓은 의미의 행정부에 속한다고 볼 수 있다.

먼저, 통계작성지정기관의 지정권자가 중앙행정기관인 통계청장이라는 점이다. 통계청장은 통계법에 따라 정부의 국가통계 작성 및 활용 전반에 관련된 사무를 총괄한다. 따라서 통계청장의 지정을 받아 통계작성기관이 되고, 통계법에 따라 통계와 관련된 사안에 대해 지휘감독을 받는 다는 점에서 통계작성지정기관이 행정부에 속하는 성격을 가진다는 점은 분명하다. 실제로 통계작성지정기관을 보면 금융기관, 공사와 공단, 연구기관, 협회와 조합, 기타 기관 등으로 분류할 수 있는데, 대부분 행정부의 관리감독을 받는 산하기관이거나, 행정부의 업무와 관련된 단체 등으로 구성되어 있다. 표 1.11은 통계작성지정기관의 현황이다.

통계법은 통계, 통계자료, 행정자료에 대하여 각각 정의하고 있다. 통계에 대하여는 '통계작성기관이 정부정책의 수립·평가 또는 경제·사회현상의 연구·분석 등에 활용할 목적으로 산업·물가·인구·주택·문화·환경 등 특정의 집단이나 대상 등에 관하여 직접 또는 다른 기관이나 법인 또는 단체 등에 위임·위탁하여 작성하는 수량적 정보'라고 정의하고 있다.

통계자료는 '통계작성기관이 통계의 작성을 위하여 수집·취득 또는 사용한 자료(데이터베이스 등 전산자료를 포함한다)를 말한다.'라고 정의하고 있다.

행정자료에 대하여는 '통계자료를 제외한 것으로 공공기관이 직무상 작성·취득하여 관리하고 있는 문서·대장 및 도면과 데이터베이스 등 전산자료를 말한다.'라고 정의하고 있다.

표 1.11 통계작성지정기관 현황(2016.7.1.현재)

금융기관	한국철도공사	**협회와 조합**
금융감독원	한국토지주택공사	한국기계산업진흥회
보험개발원	한국환경공단	한국로봇산업협회
수산업협동조합중앙회	**연구기관**	한국무역협회
중소기업은행	과학기술정책연구원	한국소프트웨어산업협회
한국거래소	국토연구원	한국엔지니어링협회
한국산업은행	에너지경제연구원	한국여성경제인협회
한국예탁결제원	정보통신정책연구원	한국전기공사협회
한국은행	중소기업기술정보진흥원	한국전자정보통신산업진흥회
공사와 공단	한국건설기술연구원	한국정보통신공사협회
교통안전공단	한국교육개발원	한국정보통신진흥협회
국립공원관리공단	한국노동연구원	한국지방재정공제회
국민건강보험공단	한국보건사회연구원	한국철강협회
국민연금공단	한국보건산업진흥원	해외건설협회
소상공인시장진흥공단	한국여성정책연구원	**기타 기관**
인천국제공항공사	한국전기산업연구원	(재)한국정보통신산업연구원
주택도시보증공사	한국조세재정연구원	건강보험심사평가원
축산물품질평가원	한국지질자원연구원	국가평생교육진흥원
한국공항공사	한국직업능력개발원	국립중앙의료원
한국관광공사	한국청소년정책연구원	농림수산식품교육문화정보원
한국광물자원공사	한국한의학연구원	한국청년기업가정신재단
한국농어촌공사	한국행정연구원	정보통신산업진흥원
한국도로공사	한국형사정책연구원	한국감정원
한국방송광고진흥공사	한국환경산업기술원	한국고용정보원
한국산업단지공단	**협회와 조합**	한국공정거래조정원
한국산업안전보건공단	대한건설협회	한국기상산업진흥원
한국산업인력공단	대한설비건설협회	한국방위산업진흥회
한국석유공사	대한시설물유지관리협회	한국사학진흥재단
한국시설안전공단	대한전문건설협회	한국생산성본부
한국에너지공단	대한측량협회	한국언론진흥재단
한국장애인고용공단	생명보험협회	한국인터넷진흥원
한국전기안전공사	중소기업중앙회	한국전력거래소
한국전력공사	한국건설기술관리협회	
한국주택금융공사	한국금융투자협회	

여기에서 말하는 공공기관은 중앙행정기관, 지방자치단체, 공공기관의 운영에 관한 법률[14]에 따른 공공기관, 지방공기업법에 따른 지방공사 및 지방공단, 유

아교육법·초중등교육법·고등교육법, 그 밖의 법률에 따라 설립된 각급 학교, 특별법에 따라 설립된 특수법인을 말한다. 이렇듯 통계법이 별도로 공공기관에 관한 정의를 도입해서 공공기관의 운영에 관한 법률에서 정의하고 있는 공공기관에다 다른 기관들을 추가한 것은 통계작성에 필요한 행정자료의 범위를 확대하고자 한 것이다.

통계법은 중앙행정기관의 장 또는 지방자치단체의 장이 통계작성의 목적을 위해 공공기관으로부터 행정자료를 수집할 수 있도록 하고 있다.[15)]

앞에서 살펴 본 것과 같이 행정자료에 관한 법적 제도적 접근은 통계법을 중심으로 진행되고 있다. 또한 본 교재에서 논의하고자 하는 것도 행정자료의 통계적 활용에 그 초점이 맞추어져 있으므로, 통계적 측면에서 행정자료의 의미나 정의는 통계법을 중심으로 이해하는 것이 타당할 것으로 보인다. 그리고 앞

14) 제4조(공공기관) ① 기획재정부장관은 국가·지방자치단체가 아닌 법인·단체 또는 기관(이하 "기관"이라 한다)으로서 다음 각 호의 어느 하나에 해당하는 기관을 공공기관으로 지정할 수 있다. 〈개정 2008.2.29.〉
1. 다른 법률에 따라 직접 설립되고 정부가 출연한 기관
2. 정부지원액(법령에 따라 직접 정부의 업무를 위탁받거나 독점적 사업권을 부여받은 기관의 경우에는 그 위탁업무나 독점적 사업으로 인한 수입액을 포함한다. 이하 같다)이 총수입액의 2분의 1을 초과하는 기관
3. 정부가 100분의 50 이상의 지분을 가지고 있거나 100분의 30 이상의 지분을 가지고 임원 임명권한 행사 등을 통하여 당해 기관의 정책 결정에 사실상 지배력을 확보하고 있는 기관
4. 정부와 제1호 내지 제3호의 어느 하나에 해당하는 기관이 합하여 100분의 50 이상의 지분을 가지고 있거나 100분의 30 이상의 지분을 가지고 임원 임명권한 행사 등을 통하여 당해 기관의 정책 결정에 사실상 지배력을 확보하고 있는 기관
5. 제1호 내지 제4호의 어느 하나에 해당하는 기관이 단독으로 또는 두개 이상의 기관이 합하여 100분의 50 이상의 지분을 가지고 있거나 100분의 30 이상의 지분을 가지고 임원 임명권한 행사 등을 통하여 당해 기관의 정책 결정에 사실상 지배력을 확보하고 있는 기관
6. 제1호 내지 제4호의 어느 하나에 해당하는 기관이 설립하고, 정부 또는 설립 기관이 출연한 기관
②제1항의 규정에 불구하고 기획재정부장관은 다음 각 호의 어느 하나에 해당하는 기관을 공공기관으로 지정할 수 없다. 〈개정 2007.12.14., 2008.2.29.〉
1. 구성원 상호 간의 상호부조·복리증진·권익향상 또는 영업질서 유지 등을 목적으로 설립된 기관
2. 지방자치단체가 설립하고, 그 운영에 관여하는 기관
3. 「방송법」에 따른 한국방송공사와 「한국교육방송공사법」에 따른 한국교육방송공사
③제1항제2호의 규정에 따른 정부지원액과 총수입액의 산정 기준·방법 및 동항제3호 내지 제5호의 규정에 따른 사실상 지배력 확보의 기준에 관하여 필요한 사항은 대통령령으로 정한다.

15) 제24조(행정자료의 제공) ① 중앙행정기관의 장 또는 지방자치단체의 장은 통계의 작성을 위하여 필요한 경우에는 공공기관의 장에게 행정자료의 제공을 요청할 수 있다. 〈개정 2009.4.1., 2012.12.18.〉
②공공기관의 장은 제1항에 따라 행정자료의 제공을 요청받은 때에는 국가기밀, 개인과 기업의 중대한 비밀의 침해 등 대통령령으로 정하는 정당한 사유가 없는 한 이에 응하여야 한다.
③제2항에 따라 행정자료를 제공하는 경우 그 제공범위 및 방법 등에 관하여는 요청기관의 장과 제공기관의 장이 협의하여 결정하되, 제공기관의 장은 요청기관의 장에게 행정자료에 포함되어 있는 개인이나 법인 또는 단체 등의 정보를 보호하기 위하여 사용방법·사용부서나 그밖에 필요한 사항에 대하여 제한을 하거나 행정자료의 안전성 확보를 위하여 필요한 조치(이하 "정보보호조치"라 한다)를 강구하도록 요청할 수 있다.
④제2항에 따라 공공기관으로부터 제공받은 행정자료는 이를 통계작성 외의 목적으로 사용하거나 다른 자에게 제공하여서는 아니 된다.
⑤행정자료 제공기관의 장은 요청기관의 장이 제3항에 따라 요청한 정보보호조치를 하지 아니하거나 제4항을 위반하는 경우에는 행정자료의 제공을 중지 또는 제한할 수 있다.

으로 이 책에서 언급하는 행정자료는 통계적 활용을 전제로 하는 것임을 밝혀 둔다.

1.2.3 해외 사례[16)]

UN의 『통계메타데이터용어집』은 행정자료에 대해 '통계청 외부의 원천들로 부터 수집한 자료로, 수집한 자료의 원천이 사람, 장소나 물건인 것' 이라고 정의하고 있다.[17)]

캐나다 통계청의 『품질지침』은 '행정기록은 다양한 비통계적 사업을 수행하기 위해 수집된 자료'라고 정의하고 있으며,[18)] 『통계자료와 메타데이터 교환지침』은 'Administrative source는 하나 이상의 행정규정을 시행하기 위한 목적으로 수집되고 관리되는, 정보를 담고 있는 자료보유'라고 정의하고 있다.[19)]

OECD와 EUROSTAT의 『통계용어집』[20)]은 행정원천(Administrative source), 행정자료(Administrative data), 행정자료 수집(Administrative data collection)을 각각 다음과 같이 정의하고 있다.

행정원천은 통계자료의 기초자료가 될 것으로 보이는 상응하는 등록단위와 거래에 있어서 하나 또는 여러 개의 규정을 집행할 책임이 있는 조직단위를 의미한다.[21)] 행정자료는 행정원천으로부터 가져온 자료나 자료의 집합을 의미한다.[22)] 행정자료 수집은 하나 또는 그 이상의 행정원천으로부터 자료를 수집,

16) 송성헌, 행정자료를 이용한 통계작성 방법론, 통계교육원, 2015
17) 「Terminology of Statistical Metadata」, UN, 2000, 'The data collected by sources external to statistical office. The source of something is the person, place or thing which you get it from'
18) 「Quality Guideline」, 'Administrative records are data collected for the purpose of carrying out various non-statistical program
19) 「Statistical Data and Metadata eXchange, SDMX」, Administrative source: A data holding containing information collected and maintained for the purpose of implementing one or more administrative regulations
20) OECD, IMF, ILO, Interstate Statistical Committee of the Commonwealth of Independent States, "Measuring the Non-Observed Economy: A Handbook", Second Draft, Annex 2, Glossary, Paris, 2002.
21) Administrative source is the organizational unit responsible for implementing an administrative regulation (or group of regulations), for which the corresponding register of units and the transactions are viewed as a source of statistical data.
22) Administrative data is the set of units and data derived from an administrative source.

처리, 저장, 통계자료의 평가 등과 관련된 일체의 행위를 의미한다.[23]

1.2.4 행정자료의 특징

통계에 행정자료를 활용하게 된 이유를 살펴봄으로써 행정자료의 특징의 많은 부분을 이해할 수 있다. UN의 『공식통계를 위한 행정원천과 2차 원천의 사용』[24] 핸드북은 이를 다음과 같이 설명하고 있다.

"*세계의 통계기관들은 통계작성과정의 효율성을 개선하라는, 특히 비용과 투입인력을 절감하라는 압력을 받고 있다. 동시에 통계응답자의 응답부담을 경감하라는 정치적 요구가 커지고 있다. 이것은 경영체가 응답자인 경우 특히 심해, 많은 정부에서는 경영체의 발전을 지원하고 촉진하는 핵심 조치로 관료조직을 감축하는 방안을 찾고 있다.*"

이런 압력 하에서 통계책임자들은 자료를 수집하는 방법으로써 현장조사에 의한 전통적인 방법을 변경할 것을 고려해야 하는 상황에 점점 직면해가고 있다. 어디엔가 사용 가능한 자료가 존재하고 있는지를 찾는 것이 아마도 가장 명확한 해답일 것이다. 통계와는 관련이 없는 많은 기관에서 다양한 형태의 자료를 수집하고 있고, 이런 자료가 통계조사를 통해 얻어야 하는 자료를 바로 대체하는 것이 비록 드물다고 해도, 때로는 여러 경로의 자료를 조합하거나, 전적 혹은 부분적으로 직접적인 통계자료 수집을 대체할 가능성을 제공한다.

통계작성과정에서의 행정자료의 이용정도는 국가에 따라 상당히 다른데, 이는 국가가 충분히 기능하는 발전된 등록부 기반(register-based) 통계작성체계를 가진 경우부터 이제 겨우 이런 접근법을 고려하는 경우까지 큰 차이가 있다.

행정원천은 다른 정부기관에서 조세, 편익 또는 서비스 행정의 목적으로 수집되고 사용된 후 보유하고 있는 자료를 수집한 것으로 전통적으로 정의되

23) Administrative data collection is the set of activities involved in the collection, processing, storage and dissemination of statistical data from one or more administrative sources.
24) 「Using Administrative and Secondary Sources for Official Statistics: A Handbook of Principles and Practices」, 1~4page, UN, 2011.

었다. 아마도 행정자료의 특징을 가장 포괄적으로 정의한 것은 캐나다 통계청의 Gordon Brackstone일 것이다.[25)]

그는 행정자료가 전통적인 조사자료 등과 구별되는 4가지 특징을 다음과 같이 설명하고 있다.

1) 통계기관에 행정자료를 제공하는 제공자(예:국세청)가 그 자료의 당사자(예:기업체)와 다르다(이와 대조적으로 대부분의 조사통계는 당사자가 제공자가 된다).
2) 행정자료는 원래 비통계적인 목적으로 수집되었으며, 이점은 원 자료의 처리에 영향을 미칠 수 있다.
3) 행정자료는 목표집단의 완전한 파악을 목적으로 한다.
4) 행정자료를 수집하고 처리하는 방법의 결정권은 해당 행정기관에 있다.

1996~7년에 유럽연합 내 연구팀은 다른 영역의 통계 간의 행정원천의 사용과 관련된 효과적인 조정방법에 관해 실험하였다. 이 연구팀은 어떻게 행정원천이 정의되어야 하는가를 판단하는데 간단한 원천자료 유형분류체계를 사용하였다. 먼저 모든 원천자료는 1차 원천(통계목적으로 수집된 자료)과 2차 원천(1차 원천 이외의 모든 자료)으로 나누어진다.

전통적인 또는 협의의 행정원천의 정의에는 오직 공공부분의 비 통계적 원천만으로 구성되는 반면, 광의의 정의에 따르면 민간부문 원천도 포함된다. 광의의 접근법은 『통계메타데이터전문용어집』에서 유럽통계기관장회의가 적용한 행정자료의 정의와 일치한다(그림 1.13).

협의의 정의에 따르면 행정원천은 2차 원천의 부분집합이지만, 광의의 정의에 따르면 행정원천과 2차 원천은 동일한 의미가 된다. 행정원천과 2차 원천에 대한 광의의 정의는 새로운 유형의 데이터 원천의 잠재가치 창의적인 평가를 강조한다. 이런 이유로 행정원천과 2차 원천의 정의는 통계기관에게 어떤 인공적인 제한을 두지 않으며, 가능한 폭넓게 해석되어야 한다. 따라서 행정원

25) Brackstone G J: "Statistical Issues of Administrative Data: Issues and Challenges", in "Statistical Uses of Administrative Data -An International Symposium", organised by Statistics Canada, 23-25 November 1987 (Proceedings published by Statistics Canada, Ottawa, December 1988)

천과 2차 원천이란 용어가 동의어로 취급되며, UN의 핸드북에서는 두 개념을 포괄하는 의미로 행정원천을 사용하고 있다.

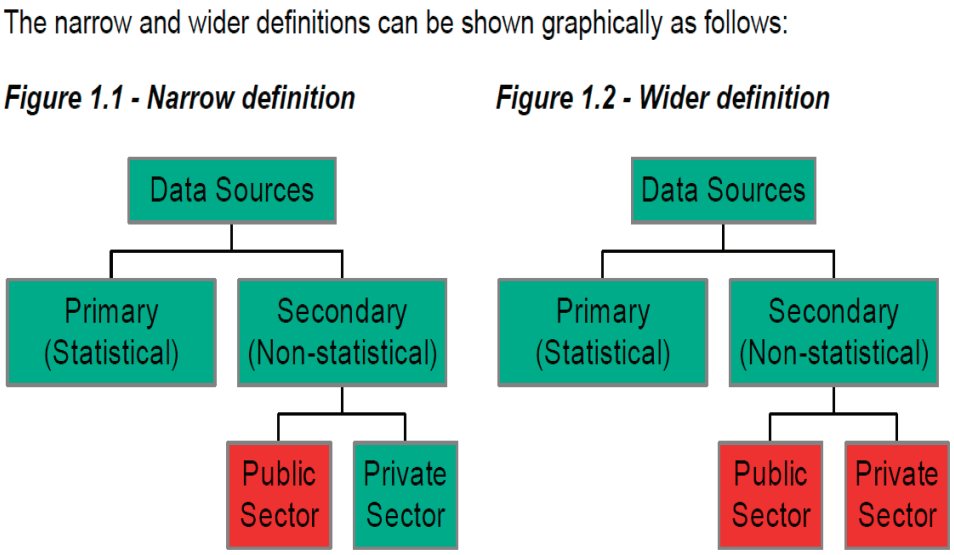


그림 1.13 광의와 협의의 행정자료

광의의 정의를 선호하는 이유가 증가하는 현황을 살펴보면 다음과 같다.

- **정부기능의 민영화 증가**

여러 국가에서 과거 정부부처나 기관에서 수행하던 규제기능이 민간기구 또는 준 민간기구로 이양되고 있다. 전형적인 사례로서 대개 보건, 교육, 복지, 공공사업분야에서 이전의 정부독점이 점진적으로 민간회사 또는 비영리기구로 이양되고 있다. 몇몇 국가에서는 정부부처를 대신한 행정등록의 운영을 포함한 등록기능의 민영화도 또한 고려중에 있다. 이는 공공부문과 민간부문 간의 전통적인 구별이 점점 희미해지고 있음을 의미하며, 전통적 혹은 협의의 행정원천의 정의가 너무 제한적인 것이 되고 있음을 의미한다.

- **민간부문 자료의 성장과 '가치부가 재판매업'**

세계의 디지털 정보의 양은 기하학적으로 성장하고 있는데, 매 5년마다 거의 10배로 증가하고 있다. 동시에 데이터의 상업적 가치는 분명해지기 시작했고, 민간부문에서의 데이터 시장은 빠르게 증가하고 있다. 이것은 마케팅 목적의 주소명

부의 발전 및 판매와 함께 시작했고, 기업경영정보와 신용등급자료의 제공으로 범위를 확장하여, 이제는 거의 모든 형식의 데이터로 범위가 확장되었다. 이 시장의 규모가 증가함에 따라, 많은 업체들이 그것으로부터 이익을 얻고 있다. 민간부문은 데이터가 매우 가치 있는 상품이라는 것을 인지하고 있다.

비교적 최근의 발전은 데이터 시장에서 민간부문의 '가치부가 재판매업'을 출현시켰다. 이들 사업은 다양한 공공 및 민간 부문의 원천으로부터 현존하는 데이터를 가져다가, 결합·정제·인증한 후, 다른 기관에 되판다. 이런 유형의 데이터 원천은 공식통계 공급자에게도 이익이 될 수 있는데, 이는 이들 민간부문 데이터 공급자들이 통계기관보다 싼 가격으로 데이터를 실제로 가공하여 공급할 수 있기 때문으로, 단순하게 보면 그들이 많은 수의 구매자들에게 비용을 분산시킬 수 있기 때문이다. 기업집단에 대한 유럽통계 등록부의 발전을 위한 'Eurogroups' 프로젝트는 정확히 이와 같은 이유 때문에 그런 원천을 사용한다.

- **새로운 형식의 데이터에 대한 사용자의 관심**

공식통계의 이용자들은 끊임없이 새로운 형식의 데이터를 요구하고 있다. 통계작성비용과 통계조사에 대한 응답부담을 줄이라는 압력은, 통계수요를 충족시키기 위한 새로운 통계조사를 도입하는 것을 어렵게 했고, 통계기관은 점차 이를 대체할 수 있는 해결책을 찾아야 할 필요가 있었다. 민간부문 원천의 범위, 내용, 규모가 성장함에 따라 통계조사의 대체수단으로서의 매력을 발산하고 있다.

1.2.5 행정자료의 유형

앞에서 통계목적으로 사용될 수 있는 행정원천의 잠재적인 범위가 크고 계속 성장하고 있음을 밝혔다. 다음 표 1.12에 예시되는 행정자료의 유형목록은 완전한 것임을 의미하는 것이 아니라 잠정적인 데이터 원천의 유형과 범위를 보여주기 위한 것이며, 행정원천의 사용 가능한 정의에 도달하는 방향으로 마지막 걸음을 내딛는 의미이다.

표 1.12 행정자료의 유형

• Tax data	-Personal income tax
	-Value Added Tax (VAT)
	-Business I profits tax
	-Property taxes
	-Import I export duties
• Social security data	-Contributions
	-Benefits
	-Pensions
• Health I education records	
• Registration systems for persons I businesses I property I vehicles	
• Identity cards I passports I driving licenses	
• Electoral registers Register of farms	
• Local council registers Building permits	
• Licensing systems e.g. television.sale of restricted goods	
• Published business accounts	
• Internal accounting data held by businesses	
• Private businesses with data holdings	- Credit agencies
	- Business analysts
	- Utility companies
	- Telephone directories
	- Retailers with store cards etc.

1.3 빅데이터와 행정자료의 관계

최근 들어 인터넷 등 정보화의 진전에 따라 다양한 분야에서 빅데이터가 생산·축적됨에 따라 이를 통계적으로 활용하려는 경향이 세계적으로 확산되고 있다. 그러나 현재 빅데이터가 무엇이며 어떻게 활용할 것인가에 관한 연구가 국제기구를 중심으로 진행 중에 있기 때문에 확정된 정의나 분류를 찾기 어렵다. 지금까지 논의 되어온 결과를 중심으로 빅데이터와 행정자료의 관계를 살펴보면 다음과 같다.

1.3.1 이론적 측면에서의 관계

빅데이터와 행정자료를 공통의 기준을 가지고 비교할 수는 없다. 그것은 빅데이터가 자료의 형태를 나타내는 용어이고, 행정자료는 자료의 생성원천을 나타내는 용어이기 때문이다. 즉 빅데이터는 초대용량인 자료의 크기, 빠른 생성속도, 내용의 다양성이란 특징을 가진 데이터를 의미하고, 행정자료는 데이터의 생성원천이 행정기관(협의 또는 광의)인 데이터를 의미한다. 따라서 빅데이터는 그 내용이 행정자료인 자료일 수도 행정자료가 아닌 자료일 수도 있으며, 마찬가지로 행정자료 또한 빅데이터 형태로 존재할 수도 빅데이터가 아닌 자료의 형태로 존재할 수도 있는 것이다. 다만, 행정자료가 규정에 따라서 전체의 행정대상에 대한 장기간에 걸친 행정원천으로부터 생성되는 경우가 많아 빅데이터의 형태를 띨 가능성이 많다. 물론 빅데이터가 얼마 이상의 크기를 가진 자료라는 확정된 정의가 없는 것이, 행정자료와 빅데이터의 관계를 명확히 정리하게 하지 못하게 하는 가장 큰 원인임은 분명하다.

1.3.2 존재 형태면에서의 관계

행정자료는 일반적인 데이터 형태로 구성된 행정자료와 빅데이터 형태로 구성된 행정자료일 수 있고, 빅데이터 형태로 구성된 행정자료는 여러 종류의 빅데이터 중의 하나라고 이해할 수 있다. 빅데이터는 생성된 신용카드 등 거래자료, 통신사의 모바일 위치 등 각종 추적자료, 위성사진 등 센서자료, 각종 온라인 거래자료, 트위터 등의 각종 의견자료 등 여러 종류의 데이터 생성원천으로부터

생성된다. 마찬가지로 행정분야에서 생성된 빅데이터 형태의 데이터가 행정자료 빅데이터라고 규정하는 것이 타당할 것이다. 사실 ICT 기술이 쉼 없이 발전하고 데이터의 활용이 고도화 되면서 크기만으로 빅데이터를 규정하는 것은 점점 의미를 잃고 있다. 빅데이터에 대한 관심과 활용정도가 높아지면서 처음 3개로 출발하였던 V는 표 1.8에서처럼 여덟 가지로 늘어났으며, 이는 빅데이터를 바라보는 관점이 다양화 되었다는 것을 나타낸다. 최근에는 너무나도 대규모로 빠르게 생산되는 데이터를 모두 저장하고 분석해야 하는가? 이들을 어떻게 처리해야 하는가? 에 대한 관심이 고조되고 있다. 따라서 절대적인 자료의 규모 관점에서 빅데이터를 보는 것 보다는 자료가 포함하는 정보의 크기 관점에서 빅데이터를 논의하는 것도 필요할 것으로 보인다. 이런 관점이라면, 행정자료 또는 결합된 행정자료는 빅데이터로 분류하기에 충분할 것이다.

1.3.3 현실적 측면에서의 특성 비교

빅데이터는 정형적 또는 비정형적 형식의 데이터를 포함할 수 있다. 행정자료 또한 정형적 비정형적 데이터를 포함할 수 있지만 법령의 규정에 따른 정형적인 행정행위를 주축으로 하는 경향이 많아 주로 정형적 데이터 형태를 띠고 있다. 빅데이터와 행정자료, 공식통계의 특징을 비교해보면 표 1.13과 같다.

표 1.13 빅데이터, 행정자료 및 공식통계 특징 비교

특 성	빅데이터(행정자료 제외)	행정자료	공식통계
생산이유	특정 비즈니스 목적	행정관리 목적	통계생산 용도
분석·처리목적	비즈니스 목표에 부합 (모델링/최적화)	행정관리 지원	모집단 추론 정책 효율성 지원
모집단 대표성	약함	강함(대상 집단)	강함
정보 추가	통계생산 용도는 추가적인 보완조치 필요	좌동	추가 가능
주 분석기법	데이터마이닝 기계학습 최적화	빈도분석 (빅데이터 기법 일부 활용)	확률 기반 표본이론
수집비용 (자료단위 기준)	일반적으로 낮음	중간	높음
수집간격	실시간	거의 실시간	주기적

CHAPTER 2

빅데이터의 연구 및 활용 추이[26)]

학습목표

- 빅데이터의 연구 및 활용에 관하여 선진국, 국제기구의 현황과 사례를 학습한다.
- 빅데이터의 연구 및 활용에 관하여 우리나라 정부 및 공공기관의 현황과 사례를 학습한다.
- 빅데이터의 연구 및 활용에 관하여 민간, 기업 및 학계의 현황과 사례를 학습한다.

빅데이터는 '21세기 원유'라 불리며, IT·금융·유통 등 다양한 업무 분야의 새로운 패러다임이자 신성장동력으로 부상하고 있으며, 용어의 사용 및 기술의 발전 등은 민간에서 시작되었지만, 정보화 사회에서 빅데이터에 대한 연구나 활용이 국가경쟁력을 의미할 정도로 급격하게 성장하고 있는 분야이다. 민간주도의 빅데이터 사업과 국가주도의 빅데이터 인프라 구축이 모두 중요하며, 특히 국가의 역할이 강조되고 있는 분야이기도 하다. 다음의 내용은 국가주도의 빅데이터의 활용에 의해 어떤 가치들이 창출되는지를 정리한 것이다[27)].

빅데이터 활용을 통해 가치가 창출되는 요소는 첫째, 투명성(transparency), 둘째, 혁신성(innovation), 셋째, 맞춤성(customization), 넷째, 의사결정의 시의적절성(better and timelier decision) 등으로 요약할 수 있다(Trends E-Magazine, 2011: 6-7; Yiu, 2012: 13-16).

26) 이장의 내용은 '빅데이터 동향 및 정책적 시사점'의 내용을 그대로 옮겨오거나 발췌하여 정리한 것임
27) 다음의 내용은 '공공부분 빅데이터의 활용성과 위험성'을 그대로 옮겨오거나 발췌하여 정리한 것임

투명성은 특히 공공부문의 경우 조직 내의 의사결정과정 및 업무절차에 대한 대중의 접근성을 제고하는 것을 가리킨다. 여기서의 투명성 제고는 기존의 전자정부의 가장 큰 편익으로 꼽았던 전통적인 투명성 개념을 확장시키는 것이다. 빅데이터에 의한 투명성 제고는 대중의 공공기관에 '대한' 투명성(예를 들어 업무처리절차의 공개) 뿐 아니라 공공기관에 '의한' 대중에 '대한' 자세한 정보를 제공해 주기 때문이다. 예를 들어 빅데이터는 조세당국이 납세자에 대한 자세한 정보를 바탕으로 과세하고 국민들이 납세의무를 성실하게 이행하고 있는지 판단한다. 또한 빅데이터는 사회복지기관이 적절한 계층에게 그에 맞는 복지서비스를 제공하고, 부당한 복지서비스의 수혜자를 적발하여 서비스요구를 거부하고 재원낭비를 방지하는 근거가 되는 자료를 효과적으로 제공한다. 투명성은 특히 데이터의 공유와 학습을 통해 실현될 수 있는데, 수많은 정부부처와 조직으로 구성된 공공부문은 조직간 주요 정보를 공유하고 공개함으로써 정부업무를 투명하게 수행하는 데에 기여할 뿐 아니라 빅데이터와 같은 엄청난 규모의 정보력을 바탕으로 정책의 대상집단에 대한 정부서비스 제공상황, 서비스로 인한 효과를 정확하고 신속하게 측정함으로써 정부업무의 효과성과 효율성, 그리고 투명성을 실현한다.

혁신성은 과거와는 차별되는 재화나 서비스를 제공하거나 업무처리 방식을 변화시키는 것을 의미한다. 즉, 다양하고 풍부한 데이터 분석을 통해 마케팅전략이나 제품생산 과정을 혁신하거나(민간부문), 공공서비스의 종류와 제공절차를 바꾸기도 한다(공공부문). 따라서 빅데이터는 비용을 절감하고 조직의 효율성을 증대시켜 줌으로써 조직의 생산성 증대는 물론이고 국가 전체 경제성장에도 기여한다. 또한 빅데이터는 성과가 떨어지는 분야를 찾아내거나 보다 생산적인 분야에 자원을 재배분하여 조직이나 기관의 생산성을 제고한다. 공공부문에서의 조직의 혁신을 통한 생산성 증대는 궁극적으로 납세자나 공공서비스의 사용자 등 일반대중에게 편익을 제공하는 데에 기여한다.

맞춤형 접근은 빅데이터 분석으로 인해 다양한 고객을 최대한 세분화함으로써 조직으로 하여금 고객의 요구에 맞는 재화나 서비스를 제공하고 궁극적으로 정부서비스에 대한 국민들의 만족도를 높이는 데에 기여한다. 다양한 고객의 요구에 대응하는 전략은 아마존, 구글 등 이미 주요 기업들에 의해 활발하게 적용되고 있는 민간부문 뿐 아니라, 획일적인 서비스가 정형화된 공공부문

에서도 최근 그 필요성이 증대되고 있는데, 빅데이터는 이러한 요구를 충족시키는 데에 중요한 역할을 하게 된다. 빅데이터의 맞춤성은 사물이나 현상에 대한 거시적 시각을 미시적 시각으로 전환시켜 주는데, 예를 들어 기존의 데이터가 전국민의 건강정도 또는 특정상품에 대한 선호도의 평균치를 정확하게 밝혀내 준다면, 빅데이터는 각 개인의 건강정도와 그에 따른 건강유지방법이나 상품구매의 선호사항 등을 개별적으로 밝혀 준다. 예를 들어 공공복지분야에서의 개별성은 복지서비스제공의 기준이 사업에서 개인으로 전환되는 것을 의미한다. 정부는 복지서비스 수혜 자격이 있고 수혜가 필요한 시민들에게 그들이 요청하기 전에 필요한 서비스를 제공해 주며, 시민들은 자신이 이용하고 싶은 서비스를 선택하게 된다. 또한 이 단계에서는 개인의 이력관리와 복지서비스 제공 기준에 대한 변경관리가 실시간으로 이루어져, 개인 상황의 변화를 실시간으로 감지하여 복지서비스 제공이 그에 맞게 이루어지게 된다.

의사결정의 시의 적절성은 공공과 민간 부문 전체에 걸쳐 조직의 생존과 운영, 공공의 요구 파악 및 문제해결을 위한 대응에 매우 중요하다. 문제해결을 위해서 빅데이터 분석은 기존에 적은 규모의 데이터로서는 파악할 수 없었던 숨겨진 패턴이나 상관관계를 찾아내거나 신뢰성 있는 예측을 도출해 낸다. 즉, 몇 가지 정책대안들이 어떤 결과를 낳을 것인지에 대한 정확한 예측을 통해 문제해결에 가장 우수한 정책대안을 선택할 수 있다. 전염성이 강한 질병이 빠른 속도로 확산되는 경우 최대한 빨리 원인을 규명하고 피해자를 치료하며, 확산을 방지하는 데에 빅데이터가 유용하게 활용되는 것이 좋은 예이다.

이와 같이 빅데이터 활용을 통한 가치창출은 빅데이터를 분석함으로써 사회문제 해결을 위한 처방을 제공한다는 것이다. 이들 빅데이터 유용성 요소들은 상호배타적이기 보다는 연계적이다. 예를 들어 복지 프로그램의 부정지출을 방지함으로써 효율적인 재정을 운영하여 결국 적절한 계층에게 적합한 서비스가 제공되고, 납세자들로 하여금 납세의 의무를 성실히 다하게 하여 조세정의를 실현하게 된다면, 이는 위에서 언급한 혁신성과 투명성, 맞춤성 및 의사결정의 시의적절성 모두와 연관된다.

국가는 다양하고 방대한 주민정보를 소유하고 있기 때문에 빅데이터의 수집, 보관, 활용에서 가장 큰 주체 가 될 수밖에 없는 상황이므로 빅데이터 활용 제

고를 위한 주도적이고 적극적인 역할이 필요하다. 정부의 역할 이외에도 인프라, 소프트웨어 등 빅데이터 관련 시장은 빅데이터를 활용한 서비스 모델의 활성화 여부에 따라 수요가 결정되므로 빅데이터 활용 모델의 개발이 매우 중요하다. 향후 서비스 사업자들이 혁신적인 비즈니스 모델을 개발할 수 있도록 다음과 같은 정책 이슈들이 논의될 필요가 있다(Australian Government, 2013. 3).

첫째, 개인정보 보호 및 프라이버시 관련 이슈이다. 실제적으로, 빅데이터의 잠재적 가치는 데이터 셋으로부터 새로운 패턴, 동향, 시사점을 제시하는 것에 있다. 이러한 분석을 위해서는 개인의 사생활을 권리를 침해하지 않는 범위에서 정보의 수입, 축적, 분석, 활용을 용인하는 사회적 분위기와 이를 위한 제도적 기반이 조성되어야 한다.

둘째, 데이터의 공유 및 가공하는 과정에서 발생하는 이슈이다. 익명화된 집단 데이터를 사용하는 과정에 있어서 모자이크 효과가 발생할 수 있는 위험성이 있다. 모자이크 효과(Mosaic Effect)란 단편적인 정보 자체로는 개인을 식별할 수 없지만, 다른 정보와 결합하면 프라이버시 혹은 보안 위험이 발생할 수 있는 상황을 의미한다(The White House, 2013. 5). 최근 미국 백악관 또한 이러한 데이터 이용 시, 개인정보 가능한 정보를 추출할 수 있는 가능성에 대해 경고 하였는데, 이러한 데이터 공유 및 처리 시 관련 규정 정립도 해결되어야 한다.

셋째, 기술적인 측면에도 이슈가 있다. 빅데이터를 활용하기 위해서는 큰 데이터 세트의 복잡한 분석이 필수적이어서, 이를 위한 대용량 데이터를 저장하고 분석하기 위한 ICT 시스템과 솔루션이 필요하게 된다. 또한 도입 시 비용 절감 문제도 함께 제기된다. 따라서 보다 저렴한 스토리지, 메모리 처리 및 클라우드 기반 스토리지 그리고 고성능 프로세싱 플랫폼에 의한 새로운 소프트웨어 개발 등이 필수적일 것이다.

넷째, 빅데이터 활용을 위한 전문 인력이 필요할 것이다. 빅데이터 기술 및 서비스를 개발하려면 과학 기술, 통계, 분석력, 해석 스킬, 시사점과 창의력 또한 비즈니스 프로세스와 기획 의도의 근본적인 본질을 이해하는 능력 등 다양한 역량이 필요하다. 따라서 관련 전문가 집단의 확보, 육성이 절대적으로 요구된다.

2.1 세계적 추이[28)]

최근 10년 사이 디지털 데이터가 폭증함에 따라 대규모 데이터가 중대 이슈로 떠오르면서 전 세계적으로 빅데이터를 막대한 가치를 창출할 수 있는 핵심 자원으로 인식하고 있다. 즉, 방대한 양의 빅데이터를 처리할 수 있는 기술이 발전하면서 과거에는 발견하기 어려웠던 가치를 창출할 수 있게 된 것이다.

이러한 빅데이터 환경은 민간 기업의 경영활동 뿐 아니라, 정부를 포함한 공공부문의 혁신을 수반하는 패러다임의 변화를 의미한다. 향후 빅데이터는 기업의 성패 뿐만 아니라 미래 국가경쟁력에도 큰 영향을 미칠 것이며, 다양한 정보미디어의 등장과 진화의 과정에서 새롭게 생성되는 정보의 효율적인 이용이 국가나 기업의 경쟁력을 좌우할 수 있다.

민간 기업들이 빅데이터와 관련된 이슈에 대한 준비를 가장 먼저 시작했으며, 이후 민간 글로벌 경제전문지, 컨설팅 그룹 등이 빅데이터를 비중 있게 보도·분석하기 시작하였다. 다수의 글로벌 IT 컨설팅 기업 및 시장조사 기관은 빅데이터 시장 규모가 지속적으로 확대될 것으로 전망하는 예측치를 제시하였다. 빅데이터 시장의 발전 및 적극적인 도입과 확산에 따라 맥킨지를 비롯한 여러 컨설팅 기관들 역시 빅데이터가 국가경쟁력은 물론 경제활성화에 기여할 수 있다고 강조하고 있다.

맥킨지는 '게임 체인저: 미국 성장과 부흥을 위한 5가지 기회'라는 보고서를 통해 빅데이터가 미국의 생산성을 향상시키고 비용 절감 효과를 가져옴으로써 유통·제조업 등이 2020년까지 연간 3,250억 달러 이상 GDP에 기여할 수 있을 것이며, 특히 비효율성을 감소시켜 정부 서비스 비용과 헬스케어 부문에서 연간 2,850억 달러의 비용을 절감할 수 있을 것이라고 예측하고 있다. 이는 고령화 등 인구 특성의 변화, 투자 감소, 생산성 개선 둔화 문제 등을 빅데이터로 해결할 수 있을 것이라는 기대감이 크게 작용하고 있기 때문이라고 할 수 있다.

영국의 경제경영연구소(Center for Economics and Business Research)는 영

28) 이절의 내용은 '주요국의 빅데이터 추진전략 및 시사점'의 내용을 옮겨오거나 발췌하여 정리한 것임

국에 빅데이터를 도입할 경우 공공 및 민간영역에서 괄목할 만한 경제적 효과가 발생할 것이며, 2012년부터 2017년까지 영국 산업 전체에서 약 2,160억 파운드의 경제적 효과가 발생할 것으로 전망하고 있다. 또한 영국 대외정책연구원(Policy Exchange)은 빅데이터로 공공부문에서 연간 160억에서 330억 파운드를 절감할 것이며, 이는 영국 정부 총예산인 7,000억 파운드의 약 2.5%에서 4.5%에 해당하는 것이라고 보고한 바 있다.

일본 총무성 역시 빅데이터 도입 및 활용에 따라 농업, 도시, 환경, 유통, 의료 등 다양한 분야에서 생산성과 효율성을 높일 수 있을 것이며, 10조 엔의 부가가치 창출과 약 12조 엔에서 15조 엔의 사회비용 절감이 가능할 것으로 예측하고 있다.

이처럼 빅데이터는 사회·경제적 비용 절감은 물론 데이터를 기반으로 한 새로운 국가정보화 전략추진으로 국정운영 및 사회 현안 등을 해결하여 더 나은 미래사회로의 방향을 모색할 수 있다. 빅데이터를 활용할 경우 복지, 안전·위험 등 국가사회 현안을 발굴하여 국민이 행복한 생활을 영위할 수 있는 맞춤형 서비스를 제공할 수 있고, 이머징 신기술을 기반으로 신성장동력 발굴 및 신시장을 개척하여 양질의 일자리도 창출할 수 있을 것이다. 이외에도 급변하는 국내외 환경 변화의 동인을 발견하고 발생 가능한 미래를 예측하여 선제적인 대응전략을 마련할 수 있다.

이에 우리나라보다 앞서서 빅데이터를 활용하기 위해 적극적으로 정책을 추진하고 있는 미국, 영국, 일본 등 주요국의 빅데이터 활용전략은 무엇이며, 빅데이터 활성화를 위한 정책과 노력에 대해 살펴본다.

2.1.1 선진국 사례

1. 미국: 민간은 글로벌 빅데이터 시장주도, 정부는 빅데이터를 통한 공공개혁추진

미국은 국가 당면 과제의 해결방안을 찾을 수 있는 원천으로 빅데이터의 가능성을 파악하고, 빅데이터로의 새로운 지식과 인사이트를 분석, 추출, 획득할 수 있는 역량 확보가 시급하다고 판단하였다. 따라서 미 대통령 과학기술자문

위원회(PCAST)가 연방 정부 수준에서의 빅데이터 관련 기술 투자 필요성을 대통령에게 건의하였다. 이후 2012년 3월에는 대통령실 내 과학기술정책실(OSTP)은 국가차원의 다양한 부처가 참여하는 2억 달러 규모의 '빅데이터 연구개발 이니셔티브(Big Data R&D Initiative)'를 발표하였다.

빅데이터 연구개발 이니셔티브는 빅데이터 핵심 기술 확보, 사회 각 영역에 활용, 인력 양성의 3가지 측면을 중점 추진한다. 기술 측면에서는 방대한 데이터 수집·저장·분석·공유를 위한 핵심 기술의 최첨단화를 추진할 것이다. 활용 측면에서는 과학기술의 가속화, 국가안보 강화, 교육의 변화를 위한 기술로 활용할 예정이다. 마지막으로 빅데이터 기술의 개발 및 활용에 필요한 전문 인력을 양성할 계획이다.

또한, 과학기술정책실은 NITRD 프로그램의 일환으로 빅데이터 연구개발 조정, 이니셔티브 목표 확인 등을 위해 빅데이터 협의체인 '빅데이터 고위운영그룹(BDSSG)'을 구성하였다. 이 그룹은 빅데이터 기반의 과학기술 발전 및 관련 기관 협조·서비스 발굴, 연방정부 데이터 관리, 인력 및 인프라 개발을 추진하고, 데이터 수집·저장·보존·관리·분석·공유와 관련된 핵심 기술의 최신성을 유지하기 위해 노력할 것이다. 이외에도 빅데이터 관련 부처 간 연계 프로젝트로 얻을 수 있는 편익을 분석하고 실현 가능한 협업 프로젝트를 지속적으로 개발하고 제안할 예정이다.

빅데이터 고위운영그룹은 정부기관 및 소속기관으로 구성되어 빅데이터 계획을 주도적으로 추진할 예정인데, 유전자 연구 및 의료, 교육, 지구과학 등 빅데이터 활용 효과가 뛰어난 분야의 기관들이 우선적으로 참여하여 추진하고 있다. 현재 8개 연방부처 및 기관이 빅데이터 고위운영그룹과 함께 프로젝트를 진행하고 있으며, 지속적으로 확대할 예정이다(그림 2.1 참고).

미국의 주요 부처에서 추진하고 있는 빅데이터 관련 정책 및 프로젝트에 대한 자세한 내용은 표 14에서 살펴볼 수 있다.

미국은 연방정부 외에 각 주정부에서도 지역에서 발생하는 각종 문제해결을 위해 빅데이터를 활용하고 있다. 범죄 방지 및 치안 유지 등의 사회 안전, 금

융사기를 비롯한 탈세방지, 행정서비스 개선 등을 위해 데이터를 분석하여 활용하는 사례들이 점차 증가하고 있다.

미국 국세청(IRS)과 고용분야 빅데이터를 이용한 환급신청서 검토 등으로 새로운 세원과 미납세금을 확인한 결과, 오하이오와 오클라호마, 텍사스 등 7개 주에서 16억 달러의 세수 증가 효과를 가져왔으며, 이러한 결과를 기반으로 미국 국세청은 지속적으로 자금세탁 및 금융사기 방지, 신용위험관리, 고객행동분석 등에 활용하고 있다.

정부기관 내 데이터를 통합·공유하여 주민 서비스를 개선하고 예산의 효율적 배분 등 정책업무 지원에 활용하는 경우도 있다. 미시간 주 정부는 21개 기관의 데이터를 통합하고 의료보장 부정수급 탐지, 개인별 범죄 관리 등 분석업무에 활용하여 1일 1백만 달러의 예산절감 효과를 발생시켰다.

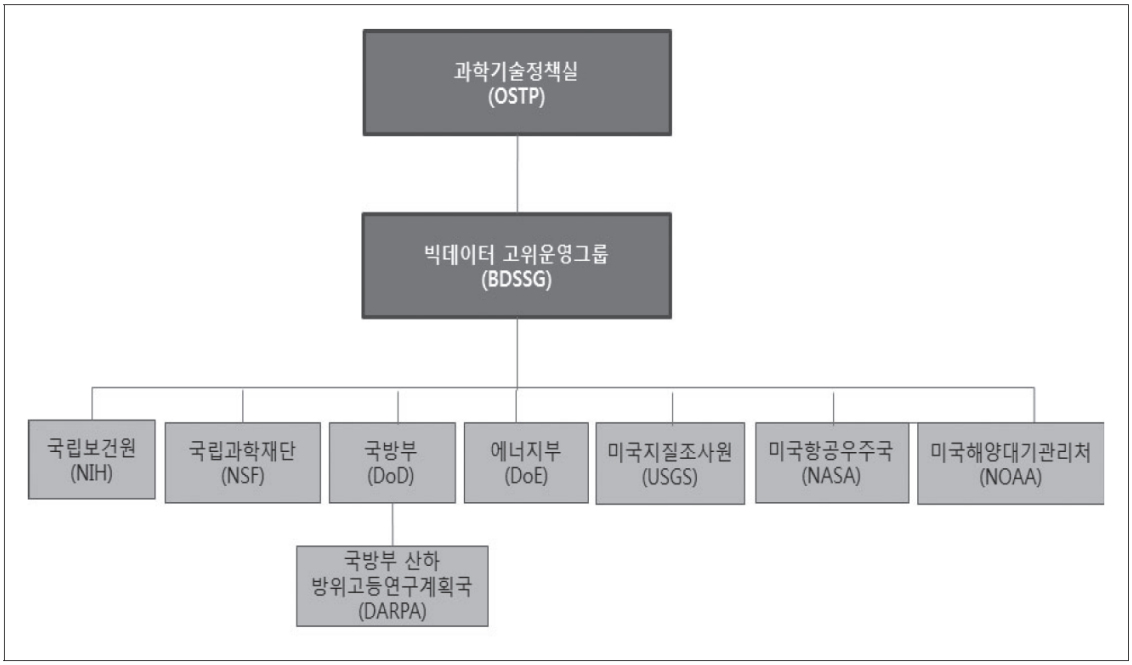


그림 2.1 미국의 빅데이터 추진체계

표 2.1 미국의 부처 별 빅데이터 프로젝트

부처명	주요내용
국방부	- 데이터로부터 의사결정을 위해 군사관련 빅데이터 프로젝트에 연간 2억 5,000만 달러를 투입하였으며, 향후 신규 프로그램에 6,000만 달러 추가 지원 예정. - 대용량 데이터를 활용한 자율 의사결정시스템 구축 및 상황인식 능력 개선을 위해 '의사결정을 위한 데이터' 프로젝트 수행. - 빅데이터 관련 기술 혁신을 촉진시키기 위해 연구 경진대회 개최.
방위고등 연구계획국	- 대용량 데이터를 분석할 수 있는 컴퓨팅 기술과 소프트웨어 개발. - 영상 데이터 처리기술, 데이터 암호화와 관련된 프로그래밍 언어 개발, 국가 위협 및 안보 요소를 감시할 수 있는 시스템 개발 등.
지질 조사원	- 지구시스템 과학 분야에 빅데이터 활용방안 연구 '존 웰시 파월 분석종합센터'를 통해 보조금을 지급하여 지구과학 시스템에 빅데이터를 활용하여 지구과학의 혁신 도모. - 향후 기후변화, 지진발생률, 차세대 생태계 지표 등에 대한 이해 증진방안 모색.
국립 보건원	- 게놈 프로젝트 데이터의 아마존 무료 공개 등을 통해 신경과학 연구 및 생리학 등 다양한 빅데이터 핵심 기술 개발. - 영상, 세포, 분자, 전기생리학, 화학, 행태, 전염병학, 임상, 기타 건강과 질병 등에 관한 데이터를 수집하여 연구개발.
국립 과학재단	- 국립보건원과 공동으로 빅데이터 과학 및 공학을 위한 핵심 기술 개발 및 데이터 사이언티스트 육성. - 대규모의 분산된 데이터를 분석하여 유용한 정보를 추출·시각화하기 위해 요구되는 기술개발 추진.
에너지부	- 진보된 컴퓨팅 파워 이용 과학적 발견을 위한 빅데이터 연구개발 추진. - 2,500만 달러의 기금을 조성해 확장형 데이터 관리, 분석 및 시각화 기관(Scalable Data Management, Analysis and Visualization Institute) 설립을 추진하고 있으며, 6개 국립 연구소와 7개 대학의 전문가들이 협력하여 점차 복잡해지고 대용량 데이터를 분석할 수 있는 툴 개발 추진. - 미생물학·식물학 등과 관련된 연구 데이터를 제공해 줄 뿐만 아니라 연구 설계에 따른 향후 결과 예측치까지 제시할 수 있는 개방형 데이터베이스 'KBase' 운영.
미국항공 우주국	- 우주 개발을 통해 얻은 다양한 자료의 공유 및 사용자 맞춤형 상품 검색을 위한 자동화 알고리즘 개발. - 주요 빅데이터 프로그램: 첨단정보시스템 기술(Advanced Information System Technology: AIST), 지구과학 데이터 및 정보시스템(Earth Science Data and Information System: ESDIS), 전 세계 지구관찰 복합시스템(Global Earth Observation System of Systems: GEOSS), 행성 데이터 시스템(Planetary Data System: PDS), 우주망원경 과학연구소의 다중미션 기록보관소(Multimission Archive at the Space Telescope Science Institute: MAST)
미국 행양대기 관리처	- 미국해양대기관리처와 국립기상청은 50년 전부터 대규모 데이터를 수집하고 분석하여 예측·예보 시스템 구현. - 매년 30페타바이트의 신규 데이터를 관리하고 위성, 선박, 항공기, 부표 등에 장착된 센서에서 매일 35억 개 이상의 데이터를 수집하여 복잡하고 정확도가 높은 예측 모델링과 함께 국립기상청에 제공

이외에도 샌프란시스코, 로스앤젤레스, 샌타바버라 등 캘리포니아 주정부와 뉴욕시를 포함하고 있는 뉴욕 주정부 등에서 가장 활발히 적용하고 있는 분야는 범죄 예방, 치안 유지 등 사회 안전 분야이다. 이를 위해 정보공유와 상황 인지 등을 통해 정부와 민간의 정보 활용을 극대화하여 국가 안보 및 국민생활을 위협하는 요인을 사전에 감지하고 대응할 수 있도록 노력 중이다.

2. 영국: 정보공유·활용을 위한 데이터 개방 중심의 정책

영국 정부는 빅데이터 활용의 기반이 되는 공공부문의 정보공유 및 활용에 따른 가치창출을 위한 데이터 공개·공유 중심의 정책을 추진하고 있다. 데이터를 '사회와 경제 성장을 위한 21세기 새로운 원자재 및 연료'로 정의하고, '영국 역사상 가장 투명한 정부'를 목표로 오픈 데이터 전략을 추진 중이다.

이를 위해 우선 기업혁신기술부(BIS)는 공공정보 공개 및 데이터의 가치창출을 위해 2012년 3월에 '데이터 전략위원회(Data Strategy Board)'를 설립하였다. 데이터 전략위원회는 내각사무처를 비롯한 각 부처의 '오픈 데이터 전략(Open Data Strategy)'에 대한 의견 제시는 물론 전략의 수정·검토를 할 수 있다. 또한 데이터 공개여부 판단, 데이터 활용을 통한 비즈니스 서비스 발굴 가능성 여부 등을 검토하여 공개·활용 등에 대한 제언을 하고 있으며, 공공 데이터의 접근 개선과 활용을 위해 일관성 있는 데이터의 제공 및 접근 방식을 고려한다. 데이터 전략위원회는 오픈 데이터 사용자 그룹, 기상 및 지리정보 사용자 그룹 등을 구성하고, 공공 데이터 그룹(Public Data Group)과의 협력 체계로 구축되어 있으며, 데이터 전략위원회 의장 및 데이터 재사용자, 기상 및 지리정보 사용자 그룹 대표, 지역정보협회 등 총 12명의 전문가들로 구성되어 있다(그림 2.2 참조).

데이터 전략위원회가 2013년 6월에 발표한 공공 데이터와 서비스에 대한 평가를 한 '정부 대응에 대한 세익스피어 검토(Government response to the Shakespeare review)' 보고서에 따르면, 데이터의 공개 및 활용은 영국 정부는 물론 기업을 위한 새로운 기회를 창출할 수 있을 것이라고 강조하고 있음을 알 수 있다.

데이터 전략위원회의 구성과 함께 내각사무처(Cabinet Office)는 데이터 접근성 강화 및 데이터 개방 지침, 향후 개방·공개 데이터 목록 등에 관한 '오픈

데이터 백서(Open Data White Paper)'를 발표하였다. 이에 2012년 6월에 기업혁신기술부(BIS)를 비롯한 총 16개 부처는 부처별 특성에 맞는 '오픈 데이터 전략(Open Data Strategy)'을 발표함에 따라 영국의 데이터 전략이 활성화되었음을 알 수 있다. 각 부처는 데이터 공유 플랫폼(data.gov.uk)의 재정비를 통한 데이터 접근성 강화 및 서비스 활성화 방안을 모색하고, 오픈 데이터 평가 방법을 도입한다. 또한 검색 기능 개선, 정보이용 방법의 단순화, GIS 데이터의 시각화, 보유 목록에 대한 접근성을 확대하고, 팀 버너스리(Tim Berners-Lee)가 데이터 품질과 재이용성을 평가하기 위해 개발한 평가방법을 기준으로 정부가 공개하는 오픈 데이터 평가를 실시할 예정이다. 각 부처는 수집된 데이터를 빅데이터, 개인정보와 관련된 My Data, 서비스에 대한 이용자 경험 및 만족도와 관련된 데이터로 구분하는 작업을 추진하고 있으며, 2015년까지 의료, 교육, 세금, 고용, 기상 및 지리 데이터 등에 대해 순차적으로 확대하여 개방할 예정이다.

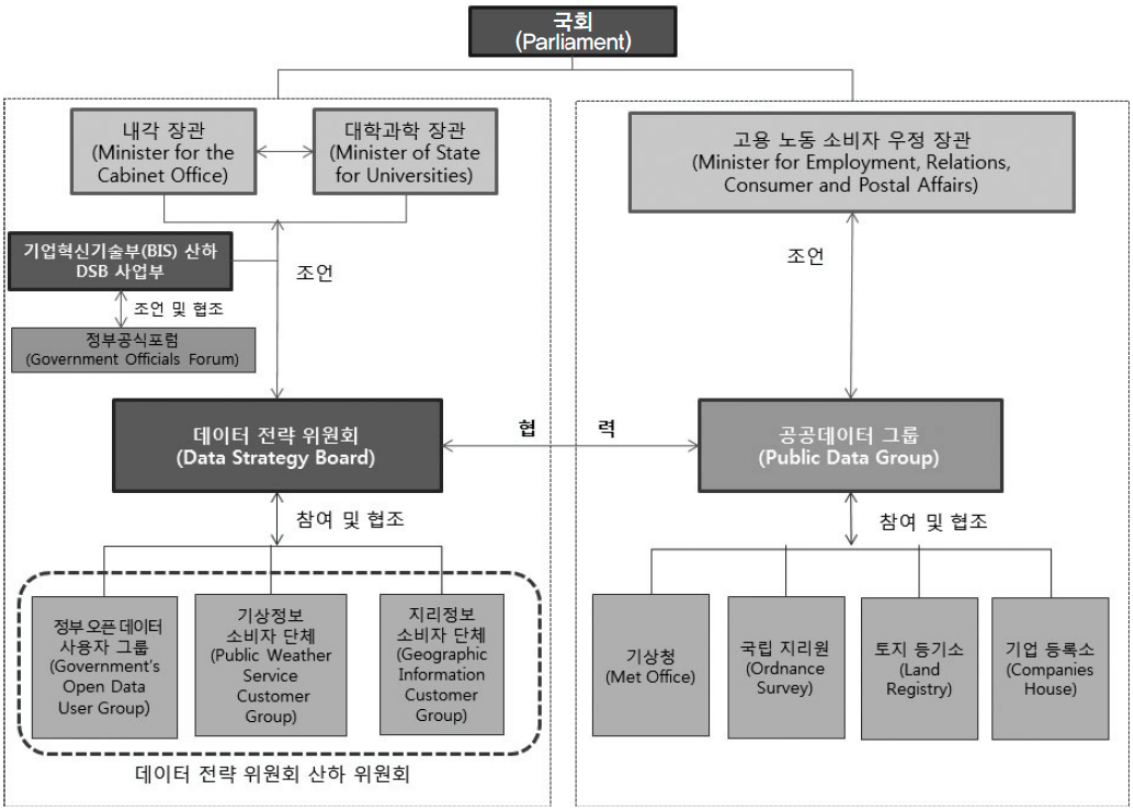


그림 2.2 영국 데이터 전략위원회 조직 구성 체계

표 2.2 영국 부처별 데이터 공개 현황

부처	데이터구분	공개 내용
교통부	철도 데이터	철도운영, 성능, 철도망 지리정보, 요금 등
	버스 데이터	버스 시간표, 버스 출도착 알림 등
	도로 데이터	(고속)도로 정보, 도로보수 정보, 도로상태 등
	항공 데이터	승객 경험데이터 수집을 통한 항공프로그램
	운전 데이터	운전시험 불합격 데이터, 운전면허 정보, 교통과 운전자 정보에 관한 요청사항 등
국세청	국세청 국가공식통계	100가지 이상의 국가공식통계 데이터(세금, 혜택, 거래 데이터 등)
	조달 데이터	조달 거래 관련 데이터
	세금 데이터	납부해야 할 세금과 건강보험 계산데이터
	수입자 데이터	EU로부터 영국으로 상품을 수입하는 13만 명 이상의 사업자 이름과 주소가 제공되는 수입자 세부 정보
환경식품 농촌부	환경 데이터	공기, 물, 생물 등 다양한 환경 통계 데이터
	농업 데이터	농장 조사, 농업 생산량 및 매출 등 데이터
문화매체 체육부	올림픽 정보	자원봉사자, 올림픽으로 인한 영향력 통계, 훈련 캠프 통계 등
	복권	복권 당첨자 세부 데이터
	브로드밴드	브로드밴드 펀드에 참가하는 프로젝트의 총 부가가치 및 고용데이터
	지속가능사업 모델	문화유산, 스포츠 등 지속가능한 사업 재정데이터
	정부예술품	예술작품 검색가능 이미지 데이터
기업혁신 기술부	고등교육 데이터	학생들의 고등교육 진학 시 고려사항
	기업등록소 데이터	기업 명, 기업 수, 등록 주소, 분류 상태 및 원문코드(SIC) 무료 다운로드
	국립지리원 데이터	국립지리원의 오픈 데이터, 국립자연탐방로 데이터 세트
노동 연금부	워크 프로그램	고용유지 지원금, 일자리 창출 결과 등에 대한 통계
	건강정보기록	근로자 건강정보
	일반 신용거래	근로자 신용관련 데이터
보건부	의사업무 관련 일반 데이터	의료진의 정보를 공개, 환자들이 병원간 치료 및 생존률 비교, 치료 관련 정보(치료시간, 대기시간, 상담자료 등)
	암 치료 데이터	1차 암 치료와 관련된 과정정보 및 결과 정보
	국민보건 서비스	연도별, 트렌드별로 성과비교가 가능한 국민보건서비스(NHS) 데이터
외무부	영사관·대사관 데이터	영국 국민에 대한 통계(외국 여행, 외국에서 사망, 외국에서 체포 여부 등

이외에도 공공 데이터의 공개를 통해 데이터에 대한 용이한 접근과 데이터의 활용 기회 증대를 위해 내각사무처는 2012년 9월에 '오픈 데이터 연구소(Open Data Institute)'를 설립하였다. 오픈 데이터 연구소는 기술전략위원회를

중심으로 오픈 데이터의 활용 기회를 높여 국가 전반의 혁신을 촉진하고, 이를 위해 오픈 데이터 활용 기업의 배양 및 육성에 주력할 것이다. 또한 오픈 데이터의 잠재력을 활용하여 새로운 비즈니스 모델을 개발하고, 신생기업을 중점적으로 지원하여 정부의 오픈 데이터를 이용한 혁신과 가치창출 등 생태계구축을 지원할 예정이다. 현재 19명이 상주 근무하고 있으며, 5개의 자체 벤처기업을 양성 중이다. 2013년 6월 현재, 오픈 데이터 연구소는 '오픈정부 파트너십을 위한 국가액션플랜 2013(Open Government Partnership: UK draft National Action Plan 2013)'에 대한 초안을 내각사무처에 제출했으며, 본 보고서에는 데이터 공개 및 투명성에 대한 내용과 공개 데이터에 대한 정의, 데이터에 대한 접근 및 활용 등에 대해 검토한 내용이 담겨있다(표 2.2 참조).

3. 일본: 사회 현안 해결을 위한 데이터 추진전략

일본 정부는 동일본 대지진을 계기로 데이터의 중요성을 재인식하고, 빅데이터를 국가경쟁력 강화에 기여할 수 있는 전략적 자원으로 평가하기 시작하였다. 수많은 데이터를 실시간으로 수집·전송·분석 등 활용하여 과제해결에 연계함과 동시에 수십조 엔의 데이터 활용 시장을 창출할 수 있을 것이라고 평가함에 따라 일본의 사회 현안 해결을 위한 데이터에 대한 가치를 재평가하였다.

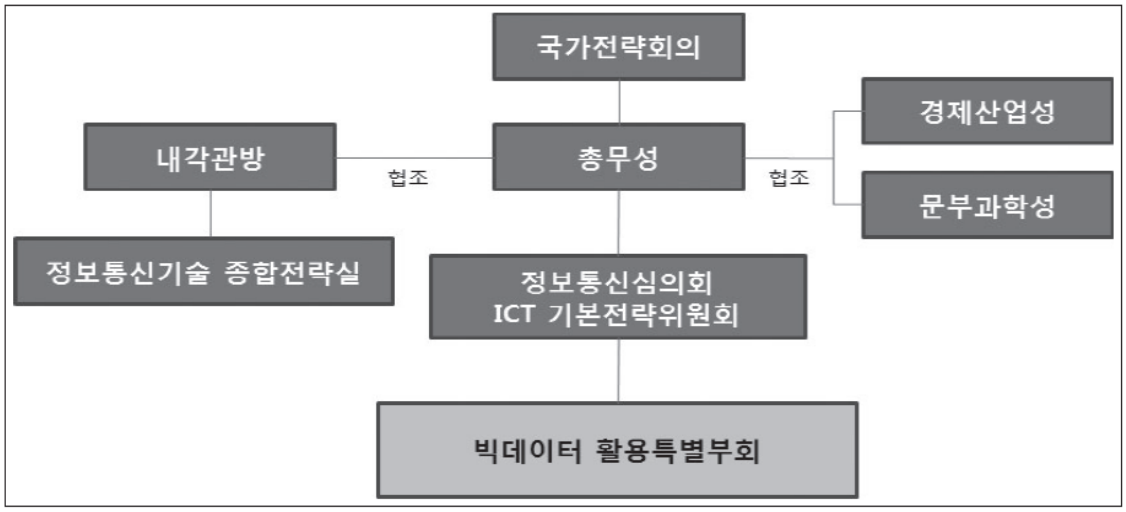


그림 2.3 일본의 빅데이터 추진 체계

그동안 경제산업성과 문부과학성에서 독립적으로 빅데이터 관련 R&D를 추진해왔으나, 최근에 총무성을 중심으로 산·학·연 참여를 통한 빅데이터 추진체

계를 재구축하였다. 이를 위해 민간위원으로 구성되는 총무성 산하 정보통신심의회 ICT 기본전략위원회에서 빅데이터 활용특별부회를 운영하여 빅데이터 활용을 위한 전략을 수립하기 위한 다양한 활동들을 추진하였다.

이처럼 빅데이터를 포함한 일본의 강력한 정보화 정책에 대한 박차를 가하기 위해 일본 정부는 2013년 6월에 새로운 조직을 신설하였다. 신설된 IT 전담조직은 '정보통신기술 종합전략실'이며, 이 조직은 내각관방 내에 설치되어 향후 일본의 효율적인 IT 정책과 국민의 편의성을 향상시킬 것이다. 또한, 중앙부처의 정보시스템을 변화시키기 위한 업무를 가장 적극적으로 추진할 예정이며, 향후 5년간 현재 1,500개가량의 일본 중앙부처 정보시스템을 절반 이하로 감소시켜 운영비용을 30%가량 절감하는 것을 목표로 하고 있다(그림 2.3 참조).

이외에도 클라우드와 모바일, 빅데이터 등으로 급변하고 있는 IT 환경에 맞는 전략을 준비하는 역할을 수행할 예정이며, 특히 공공 데이터를 민간에 개방하는 등 국민의 편의성 향상을 중점적으로 추진할 예정이다. 이를 위해 통계나 방재정보 등의 공공 데이터를 민간에도 공유시켜 국가의 IT 정책이 국민과 기업의 편의성을 향상시킬 수 있도록 하고자 한다. 이와 함께 개인정보 보호와 데이터 활용이 양립할 수 있는 방안을 마련하기 위한 전략을 2013년 말까지 수립할 예정이다.

2012년 5월, 빅데이터 활용특별부회는 빅데이터 활용을 위한 정책의 기본방향을 검토하여 '빅데이터 활용 방안'을 발표하였으며, 여기서 발표한 내용은 '액티브 재팬 ICT(Active Japan ICT) 전략'에 최종 반영하였다. 2012년 7월에 총무성이 발표한 차기 ICT 전략인 액티브 재팬 ICT(Active Japan ICT) 전략은 5개의 추진전략으로 이루어져 있으며, 빅데이터의 이용과 활용에 의한 사회·경제 성장을 위해 '액티브 데이터 전략'을 중점 목표로 하고 있다. 액티브 데이터 전략은 민간 분야를 고려하면서 빅데이터 활용 활성화 정책을 추진할 예정이며, 이를 위해 데이터 개방, 기반기술 연구개발, 표준화, 활용인재 확보, 사물 간 통신 촉진, 규제 개선, 산·학·관이 제휴하여 추진, 성과 평가 방법 등을 마련하고 있다(그림 2.4 참조).

액티브 재팬 ICT(Active Japan ICT) 전략 외에도 IT 종합전략본부(구, IT 전략본부)는 2012년 7월에 공표한 일본 재생 전략의 과학기술 이노베이션·정보

통신 전략을 통해 빅데이터 환경을 위한 공공 데이터 전략을 추진하고 있다. 본 전략에서는 2015년까지 빅데이터의 활용을 통해 새롭게 약 2조 엔 규모의 시장을 창출하고, 2020년까지 약 10조 엔 규모의 관련 시장을 육성할 계획이다. 빅데이터 활성화를 위해 공공 데이터를 적극적으로 활용할 것이며, 공공 데이터를 활용할 수 있는 인재를 육성하고 새로운 비즈니스와 서비스를 창출할 수 있도록 지원할 예정이다.

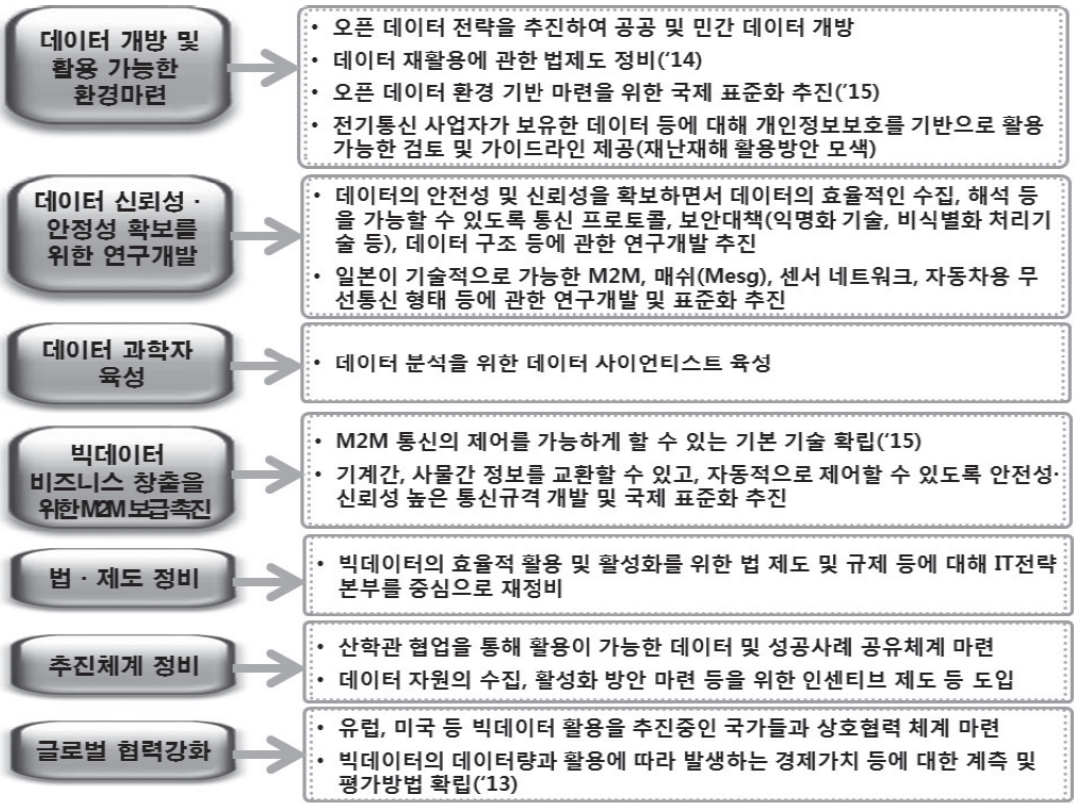


그림 2.4 엑티브 데이터 전략을 위한 7대 추진과제

이에 IT 종합전략본부는 '세계 최첨단 IT 국가 창조선언'을 통해 혁신적인 신산업 및 서비스 창출과 전(全) 산업의 성장을 촉진하는 사회의 실현을 위해 공공 데이터의 민간 개방 추진과 시장에 존재하는 빅데이터를 유기적으로 활용하여 기업 활동과 소비자 행동에 혁신을 일으킬 수 있도록 정책을 추진할 예정이다. 이를 위해 '전자행정 오픈 데이터 전략'에 의거한 로드맵을 책정·공표하고, 2013년부터 공공 데이터의 자유로운 2차 이용을 인정하는 이용방침을 재검토하여 기계판독에 적합한 국제표준 데이터 형식에서의 공개를 확대해나갈 예정이다. 또한, 각 부처가 공개하는 공공 데이터의 안내와 수평적 검색을 가능하게 하는 데이터 카탈로그 사이트에 대해서는 2013년 중에 베타판을 공개

하고, 국민의 의견을 수렴하여 2014년부터 본격적으로 운용할 예정이다(표 2.3 참조).

이렇게 빅데이터의 활성화를 위해 정부가 주도적으로 추진하고 있는 일본은 2013년 7월에 총무성이 발표한 '2013년 정보통신백서'에 따르면, 데이터 유통량 증가율과 노동 생산성 증가율은 부의 상관관계가 있음을 증명하고 있다. 그 결과로 개인의 구매 이력 등 빅데이터 활용 분야에서 연간 7조 7,700억 엔의 경제효과가 나타났다고 추정하고 있으며, 특히 소매업, 제조업, 농업, 인프라 등 4개 분야에서 빅데이터를 활용한 매출 확대와 비용이 절감되었다고 발표하였다. 즉, 소매업 분야에서는 판매 촉진 및 발주 최적화 등을 통해 약 1조 1,500억 엔, 제조업에서는 인건비 효율화 및 전기요금 절약을 통해 약 4조 7,899억 엔, 농업 분야에서는 생산비 감소와 판매 단가 향상에 따라 3,972억 엔, 인프라 부문에서는 인프라 수명 연장을 통한 신규 건설비 절감과 교통 연비 절감으로 약 1조 4,300억 엔의 경제적 효과가 발생하였다.

표 2.3 일본의 빅데이터 관련 구체적 시책 방향

구분	주요 내용
일본판 오픈 데이터	행정기관, 민간의 미활용 데이터의 활용 기반 정비, '14년까지 데이터 2차 이용 규칙 정비
센서 네트워크 안전 표준 주도	일본이 우위에 있는 센서 네트워크의 안전적 국제 통신 규격 주도
빅데이터 활용 인재 육성	산학관 프로젝트 등을 통한 인재 육성
빅데이터 활용사업 창출 촉진을 위한 규제 완화	빅데이터의 활용을 저해하는 규제, 이를 활용한 비즈니스에 대한 규제를 완화
빅데이터 활용 보급책	성공사례의 공유 및 홍보, 활용촉진 인센티브, 산학관 프로젝트 추진

미국, 영국, 일본 등 주요국은 빅데이터를 공공서비스 효율화 및 국가경쟁력 강화를 위한 중요 수단으로 인식하고 빅데이터 활성화를 위한 다양한 정책을 추진하고 있다. 국가위기관리 등 특정한 부처에서 추진하기 어려운 범정부 차원의 분석과제를 발굴하여 데이터 기반의 분석·활용 방안을 마련하고, 범정부

빅데이터 추진체계 마련을 통해 각 부처가 보유한 데이터의 연계·활용을 촉진하고 있다(표 2.4 참조, 표 2.5에 우리나라 포함 오픈데이터 사이트 제시).

특히 빅데이터는 민간 기업의 경영활동뿐 아니라, 정부를 포함한 공공부문의 혁신을 수반하는 패러다임의 변화를 의미하고 있으며, 재난재해·사회 안전, 복지·의료 등 수많은 국가사회 현안을 해결하기 위해 활용되고 있음을 알 수 있다. 즉, 데이터에 기반을 둔 객관적이고 과학적인 접근 방식을 통해 국가사회 현안해결과 미래사회 변화에 대응할 수 있으며, 데이터 분석을 기반으로 하는 창조적 접근 방식으로 더 나은 미래, 창조경제, 국민행복을 달성하여 더 나은 미래사회로의 방향을 모색할 수 있다.

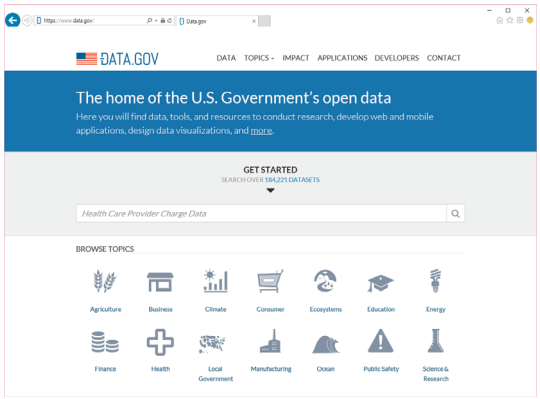
표 2.4 주요국의 빅데이터 추진전략 및 체계

구분	미국	영국	일본
의사결정기관	과학기술정책실	내각사무처	총무성
전담기구	국가과학기술위원회	기업혁신기술부	정보통신심의회
추진기구	빅데이터 협의체	데이터전략위원회 오픈 데이터 연구소	빅데이터 활용특별부회 정보통신기술 종합전략실
참여부처	국방부, 방위고등연구계획국, 에너지부, 국립보건원, 국립과학재단, 지질조사원, 항공우주국, 국립해양대기관리처	기업혁신기술부, 지역사회지방정부부, 노동연금부, 교육부, 보건부, 국제개발부, 에너지기후변화부, 외무부, 국세청, 국방부, 법무부, 재무부, 교통부, 내무부, 문화매체체육부, 환경식품농촌부	내각관방, 총무성, 경제산업성, 문무과학성
정책 및 전략	빅데이터 연구개발 이니셔티브	오픈 데이터 전략	Active Japan ICT 전략과 새로운 정보통신기술전략 內 일부포함 전자행정 오픈 데이터 전략
데이터 플랫폼	data.gov	data.gov.uk	openlabs.go.jp

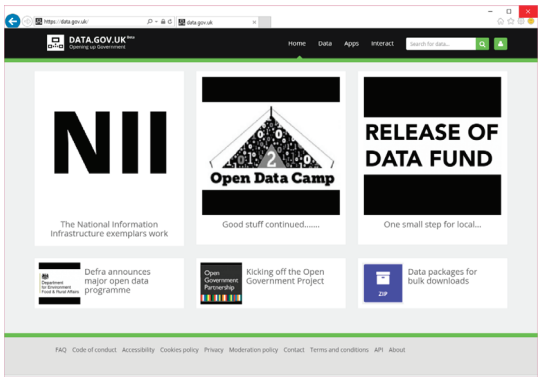
이에 우리나라도 창조경제를 달성하기 위한 핵심 엔진으로 데이터의 창조적 활용을 강조하고 있으며, 정보사회에서 데이터를 기반으로 한 새로운 가치창출의 창조사회로 변화하기 위해서는 다양한 영역 간 협력과 창의적 아이디어의 결합을 필요로 하고 있다.

이렇게 창조경제 및 국민행복을 위한 국가로의 변화를 위해서는 미래사회를 예측하고 변화발생 요인에 대한 발굴과 진단을 위한 사회, 기술, 산업, 글로벌 트렌드와 관련한 다양한 데이터를 확보하고 이를 활용하는 것이 중요하다. 이를 위해 데이터를 공유하고 활용할 수 있는 민관 협력 모델을 발굴하고, 개인정보 등 민감한 정보가 노출되지 않도록 하는 기술적인 기반을 구축해야 할 것이며, 데이터의 활용으로 사회 현안과 국민의 니즈를 파악하여 미래전략 수립, 선제적 공공서비스 제공 등 정부혁신을 지원할 수 있을 것이다.

표 2.5 각국의 오픈데이터 플랫폼



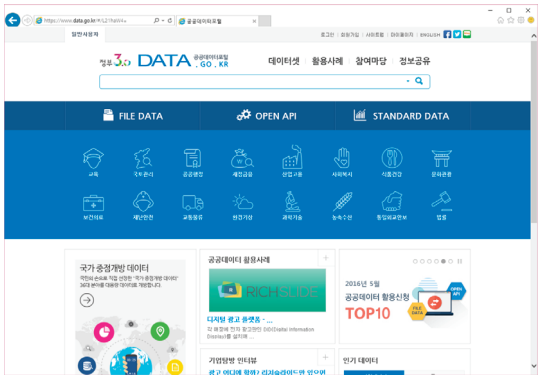
(a) 미국(http://data.gov)



(b) 영국(http://data.gov.uk)



(c) 일본(http://openlabs.go.jp)



(d) 한국(http://data.go.kr)

2.1.2 국제기구의 추이

1. OECD[29)]

(1) 배경 및 경과

OECD에서의 빅데이터 논의는 '새로운 성장요인(New Sources of Growth, 이하 NSG): 지식기반자본(Knowledge Based Capital, 이하 KBC)에 관한 수평적 업무'라는 프로젝트 차원에서 빅데이터 이슈를 포괄적으로 포함시켜, 전문분과위원회 중 정보통신정책위원회(Committee for Information, Computer and Communications Policy, 이하 ICCP) 및 산업혁신기업가위원회(Committee on Industry, Innovation, and Entrepre −neurship, 이하 CIIE)에서 주로 진행되고 있다. 동 프로젝트는 2010년 OECD 혁신전략에서 기인한 것으로, 2011년부터 신성장동력으로서 지식기반자본의 경제적 가치의 증거 제공, 정책 도전과제에 대한 이해 개선이라는 목표아래 시작되었다. 2012년까지 1단계 작업은 빅데이터, 경쟁, 기업보고, 자원배분의 효율성, 글로벌 가치사슬, 혁신, 측정 및 조세 등 분야에 대한 연구가 CIIE위원회 주도아래 이뤄졌으며, 2013년부터의 2단계 작업은 ICCP위원회 책임하에 정책결정의 개선 및 성장에 긍정적인 영향을 미치는 지식기반자본의 특수 형태에 대한 연구에 초점을 맞추고 진행 중에 있다. 이와 관련 ICCP위원회는 2단계 프로젝트 시작 이전에 2012. 10월 OECD 기술전망 포럼에서 '새로운 성장요인으로서의 데이터 활용: 빅데이터 분석 및 정책'을 주제로 빅데이터 분석의 의미, 기회, 문제점 등을 논의하기도 하였다.

(2) 1단계 작업

CIIE위원회에서 진행한 1단계 작업은 지식기반자본의 유형 및 국가들의 활용사례, 지식기반자본과 연관된 정책관련 핵심 주제 영역을 분석 및 연구하였다. '지식기반자본'은 R&D, 데이터, 소프트웨어, 특허권, 새로운 조직 프로세스, 기업맞춤형의 기술과 디자인 등 비물리적 자산에 대한 사업 투자의 산물로, 컴퓨팅 정보(소프트웨어, 데이터베이스), 혁신 자본(R&D, 저작권, 디자인 등), 경제적 역량(브랜드자산, 인적자본, 조직 노하우, 광고마케팅 등) 등 3가지 유형의 무형자산으로 구성되어 있다. 핀란드, 스웨덴, 영국, 미국 등 주요 선진국들

29) 이 절의 내용은 'OECD의 빅데이터 관련 논의 동향(2013)의 내용을 옮겨오거나 발췌하여 정리한 것임

은 KBC 투자가 성장 및 생산성을 증가시키고, 무형자산에의 기업투자와 비즈니스 주요성과가 밀접히 연관되어있다고 보고, 이미 설비 등 유형자산 보다 무형자산인 KBC의 활용 및 투자에 주력하고 있다. 이에 OECD는 지식기반자본 프로젝트의 정책관련 핵심 주제 및 영역으로 '경쟁 정책, 조세 정책, 글로벌 가치사슬에 있어 지식기반자본의 역할, 무형자산에 대한 기업보고, 지식기반자본에의 비즈니스 투자에 대한 측정 개선' 등을 선정하여 연구를 진행하였고, 각 핵심 영역별 정책적 시사점을 제시하였다(표 2.6 참조). 한편 CIIE위원회는 2013년 2월 동 프로젝트의 최종 컨퍼런스를 개최하고 프로젝트 종합보고서를 발표하였고, 최종 본은 OECD 이사회에 회부 및 2013년 5월 각료이사회에 제출함으로써 완료되어 발간될 예정이다.

표 2.6 1단계 작업 핵심 영역의 정책적 시사점

정책분야	논의배경 및 정책적 시사점
혁신	- 정책결정자들은 R&D가 절대적인 비중을 차지했던 전통적 시각을 넘어, 혁신의 확대된 개념에 적응할 필요 - 1) 혁신을 위한 협력 활성화 방안, 2) 혁신기업을 위한 금융지원 등 필요한 지원 조치, 3) 장기간 지속가능한 혁신 프로그램의 재설계, 4) 불확실성을 최소화하는 정책안정성 등이 요구됨
기업가 정신 및 비즈니스 개발	- KBC의 축적과 최적 사용에는 신규 기업들의 아이디어 및 실험이 필요하며, 이러한 기업들의 노력을 지원하는 정부 정책이 필요 ※ 2001~'11년 OECD 10개국에서 KBC를 집중 사용하는 신규 기업들이 전체 고용의 18%에 불과했지만 신규 일자리는 47%를 창출 - 이를 위해 1) 합리적인 상품, 노동시장 및 주식 금융 시스템, 2) 실패를 과도하게 처벌하지 않는 파산법 등이 필수적임 ※ 실제로, KBC 투자 규모가 큰 국가들이 혁신기업에의 자원배분이 더욱 효과적(미국, 스웨덴은 이탈리아, 스페인보다 KBC 투자가 GDP 비중상 2배 많고, 특허 기업의 자본 유치 규모는 4배 많음)
조세	- R&D 조세 인센티브는 정부의 혁신 장려 정책의 핵심적 요소이나, 다국적 기업의 R&D 조세 경감이 지나치게 크고 지사를 활용한 국경 간 조세 이득 및 KBC의 전환 문제가 발생 가능하다는 점에서, R&D 조세 인센티브의 효과에 대한 재평가가 바람직 - 1) R&D 조세 인센티브 신규 중소기업에 초점, 2) 다국적기업의 조세경감을 줄이는 국제협력, 3) 국가간 경쟁심화가 조세수입 감소를 초래할 수 있다는 위험인식 등이 필요
경쟁정책	- KBC를 기반으로 하는 산업에 있어 경쟁정책이 새로운 과제에 직면하고 있는 바, 플랫폼 간의 경쟁이 플랫폼 내의 경쟁보다 혁신과 소비자 복지에 있어 더욱 중요 - 경쟁 정책은 1) 플랫폼 간 경쟁에 대한 적절한 고려, 2) 불필요한 반경쟁적 시장 규제의 철폐, 3) 혁신을 보호 및 장려할 효과적인 경쟁법의 강화

	가 포함되어야 함
지적 재산권	- KBC의 부상과 함께 지적재산권은 몇몇 분야에서만 중요한 기술적 이슈에서 경제 전반에 걸친 중요성을 갖는 이슈로 변화 - 지식자산기반 사회에서 1) 경쟁친화적인 정책과 효율적인 사법시스템이 뒷받침되는 지재권 시스템, 2) 특허권의 질적 하락을 제고하는 조치, 3) 국제적으로 조화된 지재권 시스템 등이 필요
글로벌 가치사슬에서의 가치 확보	- KBC 투자는 글로벌 가치사슬과 국제경쟁력에서 중요한 역할 담당 - 1) 글로벌 가치사슬에 있어 고부가가치의 고용 창출 및 유지를 위해서는 올바른 정책틀이 중요 ※ 2006년 iPod의 일자리 창출은 41,000개(국내 14,000, 국외 27,000), 그러나 국내 근로자 수입이 해외 근로자 수입의 2배 이상
금융시장	- KBC를 활용한 부채의 증권화 과정에서 혁신이 일어나면서 이러한 혁신을 촉진시킬 정책적 과제들이 대두됨 - 1) KBC 집약적인 신규 기업들에 영향을 미치는 초기 단계의 위험자본 문제를 해결하는 시장개선 정책, 2) 기업들이 비즈니스 모델 내에서 KBC의 가치에 대하여 상호 인식할 수 있는 개선된 방법이 필요
기업 보고	- 성공한 기업의 가치는 전적으로 KBC에 달려있고 기업의 시장 가치와 KBC 투자 간에는 양의 상관관계가 존재하나, 기업 재무보고는 KBC 투자에 관한 제한된 정보만 제공하며 이로 인해 기업재원 마련과 관리가 어려워짐 - 1) 자발적인 권고 및 가이드라인 수립 또는 민간 분야의 기존 기업보고 이니셔티브 지원을 통한 기업 공개의 개선, 2) KBC 투자에 대한 기업 보고 촉진메커니즘 구축, 3) 회계 감사를 위한 체계 도입, 4) 기업들이 제공한 정보의 국가 간 비교가능성 제고위한 국제협력 동참 등이 필요
측정	- 1) KBC 투자에 대한 적절한 측정, 비교가능하고 보편적인 측정 가이드라인 개발에 정부의 역할이 요구되며, 2) 중단기적으로 추가적인 수단을 개발하여 GDP의 국가 간 비교 가능성을 제고 및 유지해야 함
경제적 자산으로 데이터의 활용	- 빅데이터로부터의 경제적 가치 창출이 바로 비즈니스 혁신을 선도하는 것임 - 급속한 환경변화 및 불확실한 최적정책의 여건하에서, 정부는 1) 프라이버시 보호, 2) 개방적인 데이터 접근, 3) 정보통신기술 인프라 및 관련 기술 분야에서 일관성 있는 정책을 수행하기 위해 노력해야 함
교육 및 훈련	- 인적자본은 KBC의 기반이며, 따라서 KBC 투자 증가는 인적자본 정책의 올바른 정립이 중요함을 시사 - 1) 고급기술의 수요와 공급을 균형 있게 조정할 수 있는 효과적인 정책, 2) 교육의 질 개선 노력이 중요

출처 : OECD(2013a)

(3) 2단계 작업

CIIE위원회가 수행 완료한 1단계 작업에 이어, ICCP위원회는 2013~14년 말까지 다른 위원회와의 수평적 업무 및 ICCP위원회 주도의 수직적 업무를 동시에 추진하는 2단계 작업을 진행하고 있다.

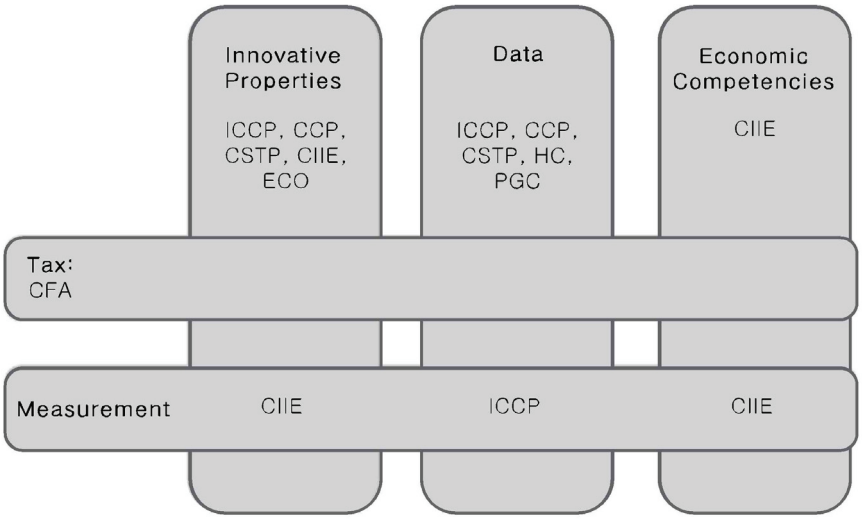


그림 2.5 2단계 작업 구조(수직 및 수평적 작업)

출처 : OECD, 2013b

2단계 작업의 목표는 정책결정의 개선을 위한 잠재적 기반을 제공해주고 결과적으로 성장에 긍정적인 영향을 미치는 지식기반자본의 특수 형태에 대한 연구 분석이며, 수평적 프로젝트인 만큼 다양한 위원회와 협업으로 진행된다. ICCP위원회는 수직적 업무로 '혁신 자본 및 지재권정책에의 시사점, 데이터 견인 지식 창출 및 정책결정, 경제적 역량' 등의 정책영역을 주도하여 작업할 예정이며, 수평적 업무로는 '지식기반 자본 측정 및 생산성 분석, 지식기반자본의 조세정책' 등의 정책 분야를 연구할 예정이다.

ICCP위원회는 동 작업을 통해 정책결정자들 및 분석가들의 지식기반자본에 대한 이해도 개선, 혁신 증진에 있어 지재권 관련 정책의 효율성 개선, 민간분야가 창출하는 지식기반자본을 지원할 수 있는 회계정책의 비용효율성, 지식기반자본에 기반을 두고 성장과 관련된 정책 분석을 지지하는 측정, 통계, 지표의 개선 등의 기대효과를 달성할 것으로 전망하고 있다.

OECD의 빅데이터 논의는 회원국들의 논의 동향 및 사례를 고려하여 핵심

주제 및 정책 영역을 선별하여 정부의 공공정책의 개선에 도움이 되도록 한다는 목적지향적인 특성을 나타낸다. 우리나라 역시 방송통신위원회에서 2011년 '스마트워크 활성화 추진계획', 2012년 '빅데이터 서비스 활성화 방안', 국가정보화전략위원회에서 2012년 '빅데이터를 활용한 스마트 정부 구현(안)'을 추진하는 등 공공부문의 빅데이터 정책을 구축해가고 있다. 그러나 아직 민간부문에 비해 활용 수준은 시작 단계라고 할 수 있으며, 좀 더 능동적인 추진 노력이 필요하다. 이에 OECD의 빅데이터 논의 및 연구는 우리정부의 향후 본격적인 정책방향 설정에 있어 참고할 만한 좋은 사례이자 근거이다. 보통 선진국간 국제기구인 OECD에서의 논의가 향후 글로벌 범위로 확대되거나 국제논의를 선도하는 측면이 많은 만큼, 우리 정부도 OECD의 연구내용을 주시하고 결과를 참고하여 국내 정책 수립 및 조정에 적극 활용할 필요가 있다.

2. UN 통계위원회(United Nations Statistical Commission, UNSC)[30]

유엔통계위원회는 160여개 국가 및 국제기구의 통계수장들이 모여 경제, 환경, 세계무역, 보건 등 글로벌 이슈에 대한 측정 기준과 통계적 지원을 결정하는 국제통계사회 최대 규모의 최고위급 위원회로

- o 통계 및 통계기법 개선, 새로운 방법론 개발 및 보급
- o 각 국가 간 통계 비교성 향상을 위한 통계배포 인프라 강화
- o 각종 정책현안(i.e. MDG, SDG 등)에 대한 통계적 지원
- o 주요 통계적 이슈제기 및 공론화
- o 통계의 수집, 분석, 작성등과 관련 개선 권고

등을 주요 기능으로 갖는다. 통계위원회에서 빅데이터 관련 활동을 정리하면 다음과 같다. 상시기구가 아니므로 활동을 중심으로 개조식으로 정리하였다.

(1) UNSD 빅데이터 관련 세미나 개최

- o 제44차 통계위원회 중 'Big Data for Policy, Development and Official Statistics'를 주제로 세미나 개최하였으며, 빅데이터 통계작성 사례 공유

30) 이 절은 "빅데이터의 국가통계 활용을 위한 기초연구"의 내용을 옮겨오거나 발췌하여 정리한 것임

및 빅데이터 문제점 등 논의하였음('13.2월)

- o 제46차 통계위원회 중 'Big Data : How do we meet the expectations?'를 주제로 빅데이터 세미나를 개최하였으며, 국가통계기관에서의 빅데이터 전략 및 프로젝트 경험을 공유하였음('15.3월)

(2) 빅데이터 활용 설문조사 실시

- o UNECE(United Nations Economic Commission for Europe)와 공동으로 국가통계기관을 대상으로 '빅데이터 활용 설문조사'를 실시하여 빅데이터 원천, 문제점 및 활용분야 파악('14.9~10월), 조사결과는 '빅데이터의 국가통계 활용을 위한 기초연구(통계청)'참조하기 바람.

(3) 빅데이터 국제컨퍼런스 개최

- o 빅데이터 관련 문제에 대한 이해를 돕고, 경험공유를 위하여 UNSD(United Nations Statistics Division)와 중국 국가통계국 공동 주관으로 국제컨퍼런스를 개최, 40여 개국 120여 명 참가('14.10.27~30, 북경)
  - 이동통신, GPS 등 추적장치, 위성영상, 소셜미디어로 대표되는 빅데이터 자료에 초점을 두고 빅데이터의 효용과 해결과제 등에 관하여 논의를 진행함.

(4) 글로벌 워킹그룹(GWG) 발족 및 회의 개최

- o 제45차 UN통계위원회 중 빅데이터의 효용성에 대해 인지하고 글로벌 워킹그룹 발족에 합의함('14.3월)
  - 빅데이터의 활용과 관련 기술, 인력, 품질 및 정보보호 등에 관한 활동을 지원하기 위하여 글로벌 워킹그룹 발족('14.5월)
  - 북경 빅데이터 국제컨퍼런스에 연하여 최초의 글로벌 워킹그룹 회의를 개최, 글로벌 워킹그룹의 위임사항을 제정하며, 실무 작업팀과 그에 따른 산출물을 합의('14.10.31, 북경)

- o 빅데이터 공식통계를 위한 글로벌 워킹그룹 위임사항을 간략히 정리하면

다음과 같다.

- 포스트 2015 의제의 지표를 포함하여 공식통계에 대한 빅데이터 글로벌 프로그램의 전략적 비전, 방향과 조정을 제공
- 빅데이터 소스의 실질적인 사용을 촉진
- 역량 강화, 훈련, 경험공유를 촉진
- 빅데이터 사용의 의사소통과 지지기반을 조성
- 빅데이터의 공식통계를 위한 대중의 신뢰를 구축

〈글로벌 워킹그룹 회원〉

▸국가(18) : 호주, 방글라데시, 카메론, 중국, 콜롬비아, 덴마크, 이집트, 인도네시아, 이태리, 멕시코, 모로코, 네덜란드, 오만, 파키스탄, 필리핀, 탄자니아, 아랍에미리트, 미국

▸기구(10) : Eurostat, GCC-Stat(걸프협력회의, Gulf Cooperation Council), ITU(국제전기통신연합, International Telecommunication Union), OECD, UN Global Pulse, UNECE, UNESCAP/SIAP, UNSD, UPU(만국우편연합, Universal Postal Union), World Bank

○ 작업팀(8개) 및 조정팀 구성, 2015년 작업팀별 사업프로그램 및 산출물을 정리하면 다음과 같다.

- 공식통계를 위한 빅데이터 소스 탐색을 위한 훈련, 기술·역량 강화 제안서 및 활동
- 빅데이터와 지속발전목표 간의 연계성에 관한 보고서
- 빅데이터 공식통계를 위한 지지기반 및 의사소통 전략
- 방법론, IT, 개인정보 문제에 관한 해결책을 제시하고, 포스트 2015 의제 모니터링에 사용가능 여부 등을 포함한 공식통계를 위한 모바일 데이터 사용에 관한 보고서
- 방법론, IT, 개인정보 문제에 관한 해결책을 제시하고, 포스트 2015 의제 모니터링에 사용가능 여부 등을 포함한 공식통계를 위한 소셜미디어 데이터 사용에 관한 보고서
- 방법론, IT, 개인정보 문제에 관한 해결책을 제시하고, 포스트 2015 의제 모니터링에 사용가능 여부 등을 포함한 공식통계를 위한 위성 영상 사용

에 관한 보고서
- 빅데이터 소스 접근 강화 및 민간부문과의 협업을 위한 제안서
- 분류, 프레임워크 등 이슈에 관한 보고서

3. 유럽 통계 생산 및 서비스 현대화를 위한 고위 전문가 단체(HLG)

HLG는 High-Level Group for the Modernization of statistical Production and Services의 약자로 통계의 현대화와 관련된 업무를 감독하고 조정하기 위해 2010년 유럽 통계인 회담(CES, Conference of European Statistician)을 통해 설립된 단체이다. 이들의 업무는 다음과 같다.

- 통계 생산과 서비스에 대한 표준 기반 현대화(standards-based modernisation) 촉진 업무를 담당
- 업무는 유럽 통계인 회담에 직접 보고하고 업무를 직접 지시받는 형태로 이루어짐

HLG의 목표와 역할 및 이와 관련된 업무수행 내용을 정리하면 다음과 같다.

- 공식통계 생산물과 조직화 측면에 있어서의 새로운 통계 생산 개발 및 통계기관 내부를 포함한 대내외 관련 기관과의 정보공유 유도 · 촉진
- 공식통계 현대화와 관련된 전략적 개발 방향에 대하여 CES에 조언하고, 통계산업 내 집중(convergence)과 협력(coordination)의 극대화를 보장

위와 같은 목표 달성을 위하여 전문가 그룹 워크샵을 매년 개최하여 유사중복 업무를 조율하는 한편, 업무의 우선순위를 선정하고 전략을 수립하고 있다.

- 워크샵에서 우선순위가 높게 선정된 한두 개의 업무를 중심으로 통계의 현대화와 관련된 도전과제 해결을 위한 연간 단위의 프로젝트를 수행
- 제3차 워크샵에서 빅데이터 프로젝트가 우선순위 업무로 선정되고, 이에 관한 제안서 제출('13.11.25~27, 제네바)
- 제4차 워크샵에서 빅데이터 프로젝트에 관한 세션이 마련되어 빅데이터 프로젝트 현황, 프로젝트 세부팀별 보고 등('14.11.19~20, 제네바)

4. EUROSTAT 및 ESS(European Statistical System)

'12년 HLG 세미나(High-level Seminar on Modernization of Statistical Production and Services, 12.10월, St Petersburg), '13년 UN통계위원회 제44차 세미나(Big data for Policy, Development and Official Statistics)에서 빅데이터의 공식통계 활용에 관련된 의제가 대두됨에 따라 유럽통계의 현대화를 위하여 ESS(European Statistical System)는 빅데이터와 관련하여 민간과의 협력, 빅데이터 인증, 새로운 데이터 소스, 기술 습득 등 빅데이터 관련 문제에 대한 접근을 시도하게 되었다.

2013년 9월 EU 국가통계기관 국장급 연간회의(Directors General of the national statistical Institutes, DGINS)에서 ESS의 빅데이터 타당성과 실행계획 채택의 필요성을 인식하고 스케브닝겐 제안서(Scheveningen Memorandum)를 통해 8개 과제를 명시하였고, 명시된 과제는 다음과 같다.

1 빅데이터가 새로운 기회임을 인식, ESS가 그 잠재성을 검토
2 공식통계와 빅데이터 전략 마련
3 데이터 활용과 개인정보에 관한 법안의 채택
4 경험공유 및 빅데이터 프로젝트 협력
5 훈련과정 등 과학기술 구축
6 정부·학계·민간 간의 시너지와 협력
7 품질평가, 정보기술 관련 이슈를 포함하여 새로운 방법론의 개발
8 제안서 후속조치로서 2014년 중반까지 실행계획 및 로드맵 채택

- Eurostat은 TF를 구성하여 ESS와 EU 내에 빅데이터 공식통계 개발을 선도·조직하고, 스케브닝겐 제안서에 따라 공공-민간을 포함한 네트워크 구축, UNECE 등 국제 협력, 상황점검, 전문지식·기술의 이전, 방법론 연구 탐색 등을 목적으로 하는 빅데이터 활용 및 발전 전략을 개발함('14~'16년)

- ESS 프로젝트(Facilitation of methodological cooperation in the ESS)의 협약에 따라 스케브닝겐 제안서 및 TF 목적달성을 위하여 ESS Big Data

Event 개최하고('14.3.31~4.1, 로마), 다음의 사항을 논의하였음.

- 27개국 107명이 참가, 빅데이터의 공식통계와 관련하여 빅데이터의 방법론 및 품질이슈, 전문인력양성, 공식통계를 위한 전략 및 협력 등을 논의
- 빅데이터 공식통계에 관한 ESS 로드맵을 위한 제안에서는 방법론적 이슈(9), 교육훈련(8), 기획(10), IT측면(11)을 다룸

5. 국제 빅데이터 프로젝트 조사(Big Data Project Survey)

2014년 9월~10월 UNSD와 UNECE는 빅데이터의 공식통계 활용을 위한 조사를 실시하고, 조사결과를 2015년 2월에 발표하였다. 조사대상은 UNECE 빅데이터 프로젝트 참가자와 UN 빅데이터 글로벌 워킹그룹 중 78개국 통계작성기관, 28개 국제기구로 하였으며, 설문지를 발송하고 응답을 받은 방법으로 자료를 수집하였다. 설문의 내용은 빅데이터 프로젝트에 관한 사항 및 이를 수행하는 조직에 관한 사항 두 부분으로 구성하였다. 발송된 설문에 대하여 24개국, 3개 기구에서 총 57개 프로젝트에 관하여 응답하였고, 32개국, 3개 기구에서 조직에 관한 설문지에 응답하였다. 빅데이터 수행조직에 관한 조사결과와 빅데이터 프로젝트에 관한 조사결과를 간략히 정리하면 다음과 같다.

(1) 빅데이터 수행 조직 조사결과

- IT, 기술, 법제화, 방법론 등에 우선해 빅데이터 프로젝트에 있어 가장 큰 문제점은 데이터 접근에 대한 제약이며, 온라인업체, 이동통신, 은행 등 민간과의 긴밀한 협력 관계를 매우 중요하게 인식함
- (빅데이터 전략) 일부 국가만이 빅데이터 활용에 관한 장기 비전을 가지고 있으며, 대부분은 빅데이터 전략수립의 초기단계 또는 미수립 단계임
- (빅데이터 거버넌스) 빅데이터 프로젝트 수행 인력이 있는 경우가 반 정도이며, 그 중 전담조직이 있는 경우는 소수, IT/방법론 담당부서의 전문가로 구성된 임시조직인 경우가 일반적임
- (빅데이터 품질) 2/3 이상이 빅데이터 품질평가체계를 정의하고 있지 않으며, 일부만이 그 체계를 구축하고 있음

- (빅데이터 정보보호) 대부분의 경우 전통적인 통계에서 활용하는 정보보호체계를 차용하고, 빅데이터 보유주체로부터 합계자료를 제공받아 활용하는 등 개인정보 보호 문제를 극복하고 있음
- (빅데이터 기술) 대부분 신규 인력을 채용하기보다는 기존 인력을 활용하고 있으며, 데이터 관리, 대용량 자료접근뿐만 아니라 Java, SAS, SQL, R같은 IT/통계기술에 관한 훈련을 마련함

(2) 빅데이터 프로젝트 조사결과

- 57개 프로젝트 중 12%는 완료, 39%는 수행 중, 49%는 제안·승인·예산확보 단계이며, 통계생산을 고려하는 경우는 48%, 연구수준의 프로젝트는 49%임
- 대부분의 프로젝트는 경제·재정, 인구·사회, 가격분야를 다루고 있으며, 인구동향, 교통, 환경, 관광, 노동, 이동통신 등도 포괄함
- 프로젝트 수행을 위해서 타기관과 협력하는 경우는 58%이며, 협력기관은 산업체 51.7%, 정부 44.8%, 학계 13.8%, NGO단체 10.3% 순이고, 협력유형은 데이터 제공 65.5%, 분석 44.8%, 데이터 통합·디자인·기술협력은 각각 17.2%임
- 대부분의 경우 프로젝트에 적합한 데이터 소스를 규정하고 있으나, 데이터를 자체 보유하는 경우는 18%에 불과하며 외부로부터 제공받은 경우는 29%, 외부와 협의 중이거나 협의가 필요한 경우가 46%에 달함
- 데이터의 지역범위는 국가전체 59%, 지역 32%, 국제 9%이며, 데이터 전송을 위한 자동연계가 구축된 경우(API, 파일내려받기)는 27%임
- 데이터 단위는 초(1000분의 1초)당 23%, 분당 4%, 주간 4%, 월간 4%, 연간 5%를 차지하며, 현행화주기는 수시 23%, 시간당 4%, 일간 4%, 주간 5%, 월간 9%, 분기 5%, 연간 9%임
- 빅데이터와 기존의 전통적 개념의 데이터를 통합하는 경우는 58%에 달하며, 데이터 소스의 41%가 일반에 공개하고, 그 중 18%는 CSV, XML, Exce 등의 형식으로 손쉽게 이용 가능함
- 기술적 측면에서 빅데이터 기반의 호스트, 저장시스템, 분석도구에 대해서는 무응답이 많은 편이나, 응답된 경우, 정보보호를 이유로 내부시스템(internal hosting solution)을 선호하며, 자료분석은 주로 R과 SAS를 이

용하며, 데이터의 저장은 하둡, RDBMS를 활용하는 경우가 많음

2.2 우리나라의 추이

한국은 세계최고의 ICT기술력과 인프라를 보유하고 많은 데이터를 생산하고 있으나, 아직은 빅데이터 이용에 대한 관심이 형성되는 수준이다. 다만, 다수의 기업이 관심을 가지고 사업을 시도하고 있다는 점에서 성장 가능성은 높게 평가되고 있다. 현재까지 국내 빅데이터 시장을 예측한 기관은 한국과학기술정보연구원(KISTI)이 유일하다. KISTI 전망치에 따르면, 한국의 전체 시장규모가 2015년 2억 6천3백만 달러, 2020년 9억 달러로 성장하여 국내 IT시장에서 빅데이터 분야가 차지하는 비중은 2013년 0.6%에서 2020년에는 2.6%까지 증가할 것으로 전망하였다(그림 2.6참조).

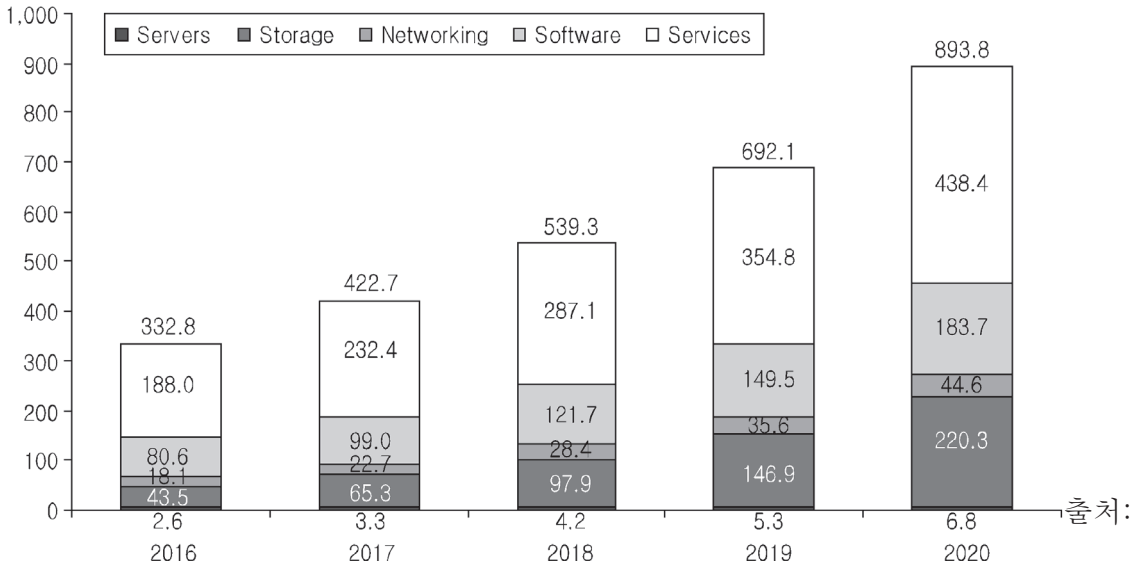


그림 2.6 국내 빅데이터 시장규모 단기전망 (단위: 백만 $)

KISTI(Market Report 2013. 4)

KISTI(2013)의 빅데이터 시장전망은 IDC 세부 시장 구분을 그대로 적용하고 있어 세부 시장별 비중을 비교해 볼 수 있다. KISTI의 세부 시장별 성장 전망은 전망 기간에 따라 차이가 있는데, 2015년까지의 단기 전망에서는 서비스 → 스토리지 → 서버 → 네트워킹 → 소프트웨어 순으로 시장규모가 형성되다가, 2016~2020년까지의 중장기에는 서비스 → 스토리지 → 소프트웨어 → 네트워킹 → 서버 순으로 소프트웨어 비중은 급증하고, 서버 비중은 급감할 것으로 전망하였다. 이러한 세부시장 전망은 소프트웨어 시장규모를 높은 순위에

두었던 IDC, Gartner 등의 전 세계 시장 전망과 차이가 있다. 이러한 결과는 상대적으로 소프트웨어 부문이 취약한 우리나라 현실을 반영한 결과로 볼 수 있다.

2.2.1 정부 및 공공기관[31)]

해외 주요국보다는 늦었지만 우리나라도 2011년 말 부터 정부 차원에서 빅데이터에 관한 논의가 시작되었다. 이후 관련 정부부처에서 정책과제로 빅데이터를 채택하였으며, 연구소 및 포럼 등이 활발히 설립되었다. 그러던 중, 2012년 11월 28일, 국가정보화전략위원회가 관계부처 합동으로 '빅데이터 마스터플랜'안을 발표하였다(표 2.8 참조).

표 2.8 빅데이터 관련 정책 추진 연혁

구분	정책	주무부서
2011년 11월	빅데이터를 활용한 스마트 정부 구현	정보화전략위원회
2011년 12월	2012년 방송통신 핵심과제로 빅데이터 포함	방송통신위원회
2012년 1월	빅데이터 관련 과제 선정(신융합 원천기술 개발 신규 과제)	지식경제부
2012년 1월	빅데이터 소프트웨어 연구소 개설	ETRI
2012년 2월	방송통신 연구 개발 시행 계획에 빅데이터 포함	방송통신위원회
2012년 4월	빅데이터 국가전략 포럼 설립	-
2012년 7월	빅데이터 마스터플랜 추진현황 및 향후 계획	행정안전부
2012년 11월	빅데이터 마스터플랜	관계부처 합동

빅데이터 마스터플랜은 행정안전부, 국가과학기술위원회, 방송통신위원회, 지식경제부, 교육과학기술부 등 관련 부처 국장들과 민간전문가 등으로 빅데이터 추진협의회를 구성하여 계획을 수립하였다. 빅데이터 마스터플랜은 선제적 사회현안 해결, 맞춤형 대국민 서비스, 데이터 기반정책수립 지원, 빅데이터 기반 신성장 동력 창출 등 4개의 목표를 수립하였으며, 또한 추진 전략으로 범부처 및 민간 협력체계 구축, 선택과 집중을 통한 단계적 확산, 전략적 기술연구 개

31) 이 절의 내용은 '정부3.0 실현을 위한 빅데이터 활용방안'의 내용을 옮겨오거나 발췌하여 정리한 것임

발지원, 수평적 데이터 활용 등을 선정하였다(그림 2.7 참조).

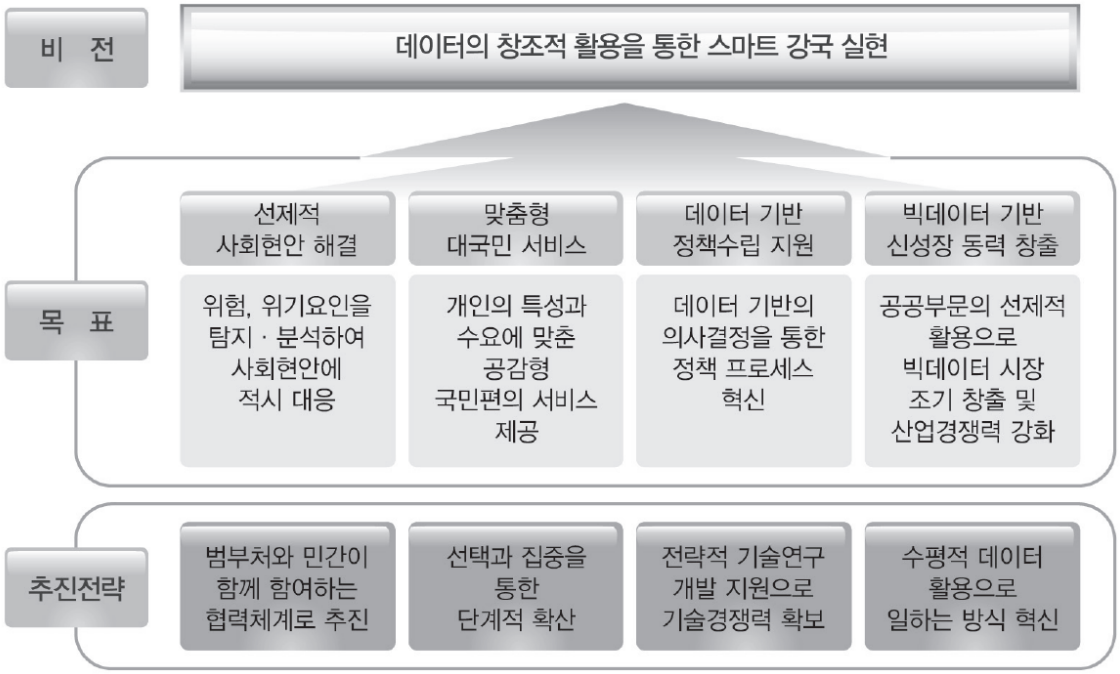


그림 2.7 빅데이터 마스터플랜

비전은 '데이터의 창조적 활용을 통한 스마트 강국 실현'으로 내세웠으며 목표는 크게 4가지로 설정했다.

- 선제적 사회현안 해결: 빅데이터 활용을 통한 미래예측
- 맞춤형 대국민 서비스: 데이터 활용을 통해 더욱 더 세분화되고 개인에 맞춰진 서비스를 제공
- 데이터 기반 정책수립 지원: 수많은 데이터를 통해 의사결정에 있어서도 합리적 결정을 하고자 함
- 빅데이터 기반 신성장 동력 창출: 빅데이터의 선제적 개발과 활용으로 시장에서 경쟁력을 갖추고 선점하고자 함

선택과 집중을 통한 단계적 확산을 빅데이터 마스터플랜의 로드맵은 다음과 같다.

- 1단계, 2012-14년: 빅데이터 활용의 타당성 확인, 우선추진 활용과제 추진, 활용기반 조성

- 2단계, 2015-16년: 빅데이터 활용 확산, 적용가능한 과제로 확산, 기술연구개발과 전문인력 양성 등 활용기반 조성
- 3단계, 2017년 이후: 국가 전반에 빅데이터 활용, 기술수준 고도화 지속 추진

현 정부 출범 이후, 현재 행정자치부의 창조정부조정실에서 정부3.0을 담당하고 있으며, 공공정보의 개방과 공유과 주요 추진방향으로 하고 있다. 행정자료, 공공자료는 다른 것으로 구분할 수 있지만 빅데이터의 활용 관점에서 자료의 연계, 정보의 활용을 생각한다면 정부3.0 계획을 살펴볼 필요가 있다.

1. 정부3.0 계획

현 정부 출범 후 정부는 저성장 구조 속에서 경제부흥의 새로운 모멘텀 필요, 기존의 방식으로 풀기 어려운 복잡다기한 사회문제의 대두, 지식정보사회로의 전환에 따른 정부-국민 간 관계 변화, 지식과 기술의 융복합 혁명이 새로운 기회요인으로 등장 등 행정환경의 변화와 새로운 도전에 직면하여, 정부운영의 패러다임 전환을 위해 정부3.0계획을 수립하고, 현재 행정자치부 중심으로 추진하도록 하였다.

정부3.0에서 제시한 비전, 즉 나아가고자 하는 방향은 '국민 모두가 행복한 대한민국'으로 설정하였다. 그리고 정부3.0의 궁극적인 목표는 '수요자 맞춤형 서비스 제공'과 '일자리·신성장동력 창출'이다. 이러한 목표를 달성하기 위해서 정부는 '서비스', '투명함', '유능함'이란 키워드를 가지고 왔으며, 그 세부에는 각각의 추진과제들이 있다. 추진과제들은 아래에서 설명하도록 하겠다. 그리고 이러한 전략이 효율적으로 이루어지려면 정부에서는 개방, 공유, 소통, 협력이 중요하고 필요하다고 언급하고 있다(그림 2.8 참조).

정부3.0의 전체적인 추진 방향은 (1) 공공정보의 개방과 공유, (2) 정부-국민 간의 소통과 협력을 확대, (3) 국가보다 국민 개개인의 행복에 초점을 추어 맞춤형 서비스 제공, (4) 민간의 창의와 활력이 증진되는 혁신 생태계 조성, (5) 부처 간 칸막이를 뛰어넘는 통합형 정부운영 지향, (6) 정부가 직접 개입하지 않고, 민간의 능동적 참여를 유도하는 플랫폼 정부로 설정하고 있다. 전체적으

로 정부1.0부터 정부3.0까지 단계적 발전을 염두 해 두고 있다.

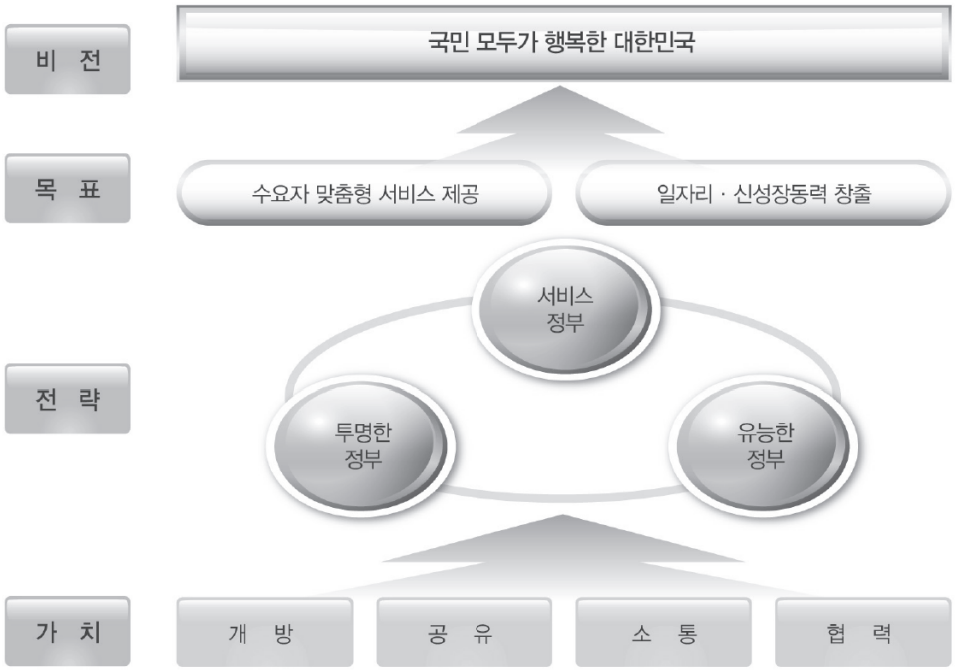


그림 2.8 정부3.0 비전 및 전략

목표를 달성하기 위해 정부는 크게 3가지의 핵심과제를 제시했으며 그 세부 추진 과제로 10가지 과제를 명시했다.

1) 소통하는 투명한 정부

- 공공정보 적극 공개로 국민의 알권리 충족
- 공공데이터의 민간 활용 활성화
- 민·관 협치 강화

2) 일 잘하는 유능한 정부

- 정부 내 칸막이 해소
- 협업·소통 지원을 위한 정부운영 시스템 개선
- 빅데이터를 활용한 과학적 행정 구현

3) 국민 중심의 서비스 정부

- 수요자 맞춤형 서비스 통합 제공
- 창업 및 기업활동 원스텝 지원 강화
- 정보 취약계층의 서비스 접근성 제고
- 새로운 정보기술을 활용한 맞춤형 서비스 창출

한국의 빅데이터 종합정책은 상대적으로 미성숙한 시장을 직접적으로 진흥하면서도 정부기관의 데이터 개방 및 활용을 동시에 지향하고 있어 일본의 정책과 유사한 측면이 있다. 다만, 해외와 비교할 때 다음과 같은 차별성이 있다.

첫째, 한국은 종합적인 청사진을 담고 있어, 공공정보 공개, 공공서비스 품질 개선 기반 형성에 주력하는 해외와는 차이가 있다. 우리 정부의 정책은 공공에 더불어 민간 활성화를 위한 방안도 동시에 담고 있다는 점에서 보다 큰 미래 비전을 제시하고 있다. 둘째, 시장활성화를 위한 법제도 개선을 계획안에 포함시키고 있다. 그동안 개인 정보 보호를 비교적 강하게 추진해 왔고, 상대적으로 빅데이터 활성화를 위한 관련 법제도 개선 논의가 부족한 현실을 고려할 때 시의적절하다고 할 수 있다. 다만, 현재는 계획 단계라 구체적인 제도 개선까지는 상당한 논의 과정이 필요할 것이다. 셋째, 아직 해외 대비 서비스 초기라 계획수립 단계로서 세부 정책 방안을 조기에 구체화할 필요성이 있다. 특히, 해외는 정부의 빅데이터 활용을 위해 기초기술투자 및 관련 기술개발 및 투자 계획이 시행되고 있으나, 국내는 아직 방향성만 제시되어 있다.

2. 개인정보 보호[32)]

각 국과 국제기구의 빅데이터 활용현황에서 알 수 있듯이 빅데이터의 잠재적 가치는 데이터 셋으로부터 새로운 패턴, 동향, 시사점을 제시하는 것에 있다. 이러한 분석을 위해서는 개인의 사생활을 권리를 침해하지 않는 범위에서 정보의 수입, 축적, 분석, 활용을 용인하는 사회적 분위기와 이를 위한 제도적 기반이 조성되어야 한다. 따라서 개인정보 보호와 데이터의 활용이 양립할 수 있는 방법을 찾는 것이 매우 중요한 관건이며, 이는 정책적으로 해결할 수 있는 문제이다. 이와 관련하여 정부부처를 중심으로 여러 가지 노력이 있었으며 그 간의 진행현황을 정리한 것인 표 2.9이다.

32) 이 내용은 '개인정보 비식별 조치 가이드라인(2016)의 내용을 옮겨오거나 발췌하여 정리한 것임

표 2.9 개인정보 보호를 위한 활동

구분	공공정보 개방·공유에 따른 개인정보 보호지침	개인정보 비식별화에 대한 적정성 자율평가 안내서	빅데이터 개인정보보호 가이드라인	빅데이터 활용을 위한 개인정보 비식별화 기술활용 안내서
발간	2013.9	2014.12	2014.12	2015.6
주관부처	행저자치부	행정자치부	방송통신위원회	미래창조과학부
대상	공공	공공, 민간	민간	공공, 민간
목적	공공정보 개방·공유 등에 따른 개인정보 보호조치 요령 안내	개인정보 비식별화 적정성 평가, 재식별 위험 관리 및 조치요령 안내	빅데이터 관련 개인정보 보호법령 적용 방안 설명	빅데이터 활용을 위한 비식별화 기술 활용방법 및 관련법규 안내
주요 내용	(1단계) 공공정보 개방·공유 근거 확인 (2단계) 개인식별요소 제거 (3단계) 개인식별 가능성 검토 (4단계) 비식별 기법 및 재식별 가능성에 관한 주기적 모니터링 실시	개인정보가 포함된 정보는 개인정보 식별요소 제거를 통해 개인을 식별할 수 없는 형태로 정보를 변경 후 이용 재식별 위험에 대한 관리적 조치 및 재식별시 대응 조치 공공정보 개방·공유 등에 따른 개인정보 보호지침과 유사한 조치 및 프라이버시 보호모델에 따른 검토를 하도록 함	공개된 정보, 이용내역 정보, 생성된 정보는 비식별화 조치 후 처리 비식별화 조치한 경우 이용자의 동의 없이 수집·이용 가능 비식별 조치 후 저장 시 보호조치 필요 비식별화 조치된 개인정보는 이용자 동의 없이 제3자 제공 가능	개인식별이 가능한 정보는 삭제 다른 정보와 결합으로 재식별될 위험 최소화 및 사후관리 철저 가명처리, 총계처리, 데이터 값 삭제, 범주화, 데이터 마스킹 등 18가지 비식별화 기술 및 적용사례 제시

출처 : 개인정보 비식별 조치 가이드라인(2016)

개인정보 보호를 위한 가장 최근의 조치로 2016년 6월 30일 행정자치부, 방송통신위원회, 금융위원회, 미래창조과학부, 보건복지부, 국무조정실 등 정부부처가 합동으로 마련한 `개인정보 비식별화 조치 가이드라인`을 발표하였다. 이는 빅데이터 활용에 필요한 비식별 조치 기준·절차·방법 등을 구체적으로 안내하여 안전한 빅데이터 활용기반 마련과 개인정보 보호 강화를 도모하기 위한 것으로 개인정보를 비식별 조치하여 이용 또는 제공하려는 사업자 등이 준

수하여야 할 조치 기준을 상세히 제시한 것이다.

비식별화는 개인 정보 일부 또는 전부를 삭제하거나 개인을 식별할 수 없도록 다른 값으로 변경하는 것을 말한다. 가이드라인에는 금융, 의료, 물류, 유통, 교통, 제조, 안전, 환경 등 정보통신기술(ICT) 융합 환경에서 빅데이터 산업 활성화로 신규 일자리를 창출하고 새로운 성장 동력을 확보하자는 정부의 의지가 담겨있다. 유관 부처가 모두 한자리에 모여 학계와 산업계 전문가 의견을 수렴해 마련됐다는 데 그 의의를 크게 둘 수 있다. 가이드라인의 근거가 되는 법은 개인정보보호법으로 이 법에서는 데이터 이용과 제공이 “통계 작성 및 학술 연구 등의 목적을 위해 필요한 경우로서 특정 개인을 알아볼 수 없는 형태로 개인 정보를 제공하는 경우”에서 목적 외로 가능하다고 규정하고 있다.

표 2.10 개인 식별요소 삭제 방법 - 일반적 기법

처리기법	예시	세부기술
가명처리 (Pseudonymization)	• 홍길동, 35세, 서울 거주, 한국대 재학 → 임꺽정, 30대, 서울 거주, 국제대 재학	① 휴리스틱 가명화 ② 암호화 ③ 교환 방법
총계처리 (Aggregation)	• 임꺽정 180cm, 홍길동 170cm, 이콩쥐 160cm, 김팥쥐 150cm → 물리학과 학생 키 합 : 660cm, 평균키 165cm	④ 총계처리 ⑤ 부분총계 ⑥ 라운딩 ⑦ 재배열
데이터 삭제 (Data Reduction)	• 주민등록번호 901206-1234567 → 90년대 생, 남자 • 개인과 관련된 날짜정보(합격일 등)는 연단위로 처리	⑧ 식별자 삭제 ⑨ 식별자 부분삭제 ⑩ 레코드 삭제 ⑪ 식별요소 전부삭제
데이터 범주화 (Data Suppression)	• 홍길동, 35세 → 홍씨, 30~40세	⑫ 감추기 ⑬ 랜덤 라운딩 ⑭ 범위 방법 ⑮ 제어 라운딩
데이터 마스킹 (Data Masking)	• 홍길동, 35세, 서울 거주, 한국대 재학 → 홍○○, 35세, 서울 거주, ○○대학 재학	⑯ 임의 잡음 추가 ⑰ 공백과 대체

출처 : 개인정보 비식별 조치 가이드라인(2016)

가이드라인의 기본 방향은 기업에 의한 빅데이터 처리를 허용하되 오·남용을 막기 위한 기업의 책임도 동시에 강화하는 것이다. 비식별화된 데이터는 개인정보로 보지 않았으며, 비식별화된 데이터는 기업이 이용할 수 있게 하였고, 제3자로의 제공도 가능하도록 하였다. 비식별화 조치의 적정성은 기업 평가단을 구성해 평가하게 하였다. 비식별화된 데이터의 외부 유출을 방지하기 위한 기술 및 관리 차원의 보호 조치를 이행하도록 하였으며, 산업 내 다른 기업 간 데이터 결합도 고려하여 산업별 비식별화 전문 기관도 허용되었다(전자신문 전문가기고, '비식별화조치 가이이드라인, 실천이 더 중요', 2016,7.17) 표 2.10과 2-11은 비식별화를 위한 일반적 방법과 프라이버시 보호모델에 대한 간략한 정리로 참고하기 바란다.

표 2.11 재식별 가능성 검토 기법 - 프라이버시 보호 모델

기법	의미	적용례
k-익명성	• 특정인임을 추론할 수 있는지 여부를 검토, 일정 확률수준 이상 비식별 되도록 함	• 동일한 값을 가진 레코드를 k개 이상으로 함. 이 경우 특정 개인을 식별할 확률은 1/k임
l-다양성	• 특정인 추론이 안된다고 해도 민감한 정보의 다양성을 높여 추론 가능성을 낮추는 기법	• 각 레코드는 최소 l개 이상의 다양성을 가지도록 하여 동질성 또는 배경지식 등에 의한 추론 방지
t-접근성	• ℓ-다양성 뿐만 아니라, 민감한 정보의 분포를 낮추어 추론 가능성을 더욱 낮추는 기법	• 전체 데이터 집합의 정보 분포와 특정 정보의 분포 차이를 t이하로 하여 추론 방지

출처 : 개인정보 비식별 조치 가이드라인(2016)

2.2.2 민간 및 기업

주요 해외 기업들이 빅데이터 시장을 점령하고 있는 가운데, 우리 기업들은 경쟁력을 갖추기 위해 2012년 하반기부터 본격적으로 빅데이터 사업에 뛰어들고 있다. 현재 국내 빅데이터 시장에서 네이버, 다음, SK텔레콤, kt등 주요 기업들의 비즈니스는 무엇이며, 어떤 형태로 이루어지는지 살펴보고자 한다.

우선, 네이버는 국내 기업 중에서는 가장 많은 데이터를 가지고 가장 높은 수

준의 빅데이터 분석 및 활용 기술을 가지고 있다. 네이버는 빅데이터라는 용어가 나오기 전인 2006년부터 빅데이터의 주요 기술 중에 하나인 하둡 기술을 도입하여 매일 생산되는 방대한 데이터를 관리하고 있다. 현재 네이버는 빅데이터를 활용하여, 실제 서비스에 적극 활용하여 음악추천기능인 서비스, 자동완성, 연관검색어, 실시간 급상승 검색어 등의 서비스를 제공하고 있으며 그 활용 범위는 계속 확대될 전망이다. 통신사들도 빅데이터를 활용한 새로운 비즈니스를 선보이고 있다. SK텔레콤은 2011년 2월 지도와 위치기반서비스(LBS) 기술을 결합해 상권분석, 고객관리 등을 지원하는 '지오비전' 솔루션을 출시하였다. 이 서비스는 자사 이용자의 위치정보를 활용한 서비스로, 특정 지역의 유동인구, 지역분석 정보 등을 활용하여 최적화된 비즈니스 솔루션을 제공하는 서비스이다. 또한 빅데이터 인프라와 시스템을 강화, 다양한 빅데이터 영역이 결합된 비즈니스를 창출하기 위해 지난해 11월 네이버와 업무협력도 체결하였다. 2013년 5월초 2,700만 이용자 정보를 분석 및 빅데이터 신산업 분야 발굴을 위한 1조 3천 억원의 투자 발표를 계기로 빅데이터 사업계획이 구체화되고 있다.

표 2.12 국내 민간분야 빅데이터 주요 사업 현황

네이버	국내 최대의 검색엔진 포털 사이트로 매일 수많은 데이터가 생성됨 2006년부터 대용량의 데이터를 관리하기 위하여 HADOOP 기술을 도입하여 현재 실시간, 연관 검색 및 음악추천 기능 등의 서비스를 제공 중
다음 소프트	'소셜메트릭스'를 활용하여 소셜미디어 분석 전문기업으로 성공적으로 변신 SNS상의 검색어들을 동태 소셜 모니터링이나 사회관계망 분석 등의 서비스를 제공 중
SKT	2011년 지오비전을 출시하여 이용자 정보를 활용한 빅데이터 사업 중이고 지속적인 투자 예정 상권분석, 맞춤형 지도서비스, 스마트 세일즈 등의 서비스를 제공 중
kt	2012년 후반부터 KCB, 서울시와 제휴를 맺는 등의 적극적인 빅데이터 사업 추진 중이며 자체적인 빅데이터 분석 플랫폼을 보유하려 노력 중

출처: Bloter.net(2012), SK 지오비전 홈페이지, 경향신문(2013), 중앙일보(2013) 재구성

한편 kt는 지난해 9월 코리아크레딧 뷰로(KCB)와 업무협조를 체결하고, 본격적으로 빅데이터 사업을 추진하고 있다. KCB의 지역기반 분석서비스 '알지

오’에 올레맵을 결합한 서비스를 제공 중이다. 이는 지역별 소비자들의 연소득 등의 구매능력을 분석하여 기업 및 자영업자들이 마케팅 등에 활용하도록 지원하였다. 지난 4월에는 서울시와 MOU를 체결하였는데, kt는 서울시의 교통, 복지, 경제, 문화, 도시 등 여러 분야의 현안 해결을 위한 빅데이터 활용 및 분석 · 검증을 위한 기술 지원 및 빅데이터 통계정보 제공 등의 역할을 할 예정이다. 더 나아가, kt는 앞으로 빅데이터의 활용도가 높아질 것에 대비하여, 자체적인 빅데이터 플랫폼 구축을 시도하고 있다(표 2.12 참조).

2.2.3 학계

1. 대학, 학회, 연구소 현황

빅데이터의 연구 및 활용추이를 하계의 입장에서 본다면, 교육과 연구가 주된 내용이 될 것이다. 이 절에서는 갈수록 그 중요성을 더하고 있는 빅데이터를 효율적으로 교육하기 위한 대학의 노력을 나타내는 빅데이터 전공, 협동과정, 학과 개설 사항에 대하여 알아보고, 연구소나 협회 설립에 관한 내용을 정리한다. 빅데이터 교육을 위한 커리큘럼에 대한 연구결과와 향후 빅데이터의 발전을 이끌 10대 기술에 대하여 소개하였다.

많은 대학들이 빅데이터 교육에 관심을 표명하고 있으나 2016년 7월 기준(대학알리미 검색결과)으로 ‘빅데이터’ 교육과정을 개설하고 있는 대학은 표 2.13에서 처럼 모두 4개이다(그 밖에 통계 교육을 하는 학과, ‘데이터’를 학과명에 포함하고 있는 많은 학과에서 빅데이터 분석 교육을 실시하겠으나, 학과명에 직접적으로 언급되지 않는 경우는 제외하였다).

이 밖에도 적지 않은 대학들이 2017학년 또는 그 이후부터 빅데이터 교육과정(학사, 석사)을 개설할 것으로 예상된다(서울대학교, 고려대학교, 연세대학교, 서강대학교. 경희대학교, 부산대학교, 국립순천대학교, 강원대학교, 단국대학교, UNIST 등).

학계의 중요한 축인 학회(학회 이름에 빅데이터가 명시된 곳)는 모두 두 개가 활동 중에 있다. 2013년 한국빅데이터 학회(KBS, Korea Bigdata Society,

http://www.kbigdata.kr)가 설립되어 활발하게 활동 중이다. 이 학회의 설립취지는 다음과 같다.

표 2.13 빅데이터 교육과정 개설 대학

학과명	학교명	교육목표
빅데이터 협동과정	부산대학교 일반대학원	다수의 정보단말들이 생성하는 방대한 규모의 빅데이터를 분석/처리하는 기술과 이를 기반으로 사회현상 변화의 합리적 해석과 효율적 의사결정기법의 개발을 목표로 한다.
빅데이터 경영통계 전공	국민대학교	최근 업계의 화두로 급부상하고 있는 분석경영학(Business Analytics)에 중점을 두어 특화된 전공이다. 분석경영학은 경영학, 통계학, 빅데이터분석학 등 다양한 분야의 지식을 포괄하는 융합분야로서 데이터 기반의 경영과 의사결정을 위해 필요한 제반 도구와 프로세스를 다루는 학문이다. 경영학과 통계학, 빅데이터분석학의 융합을 통하여 비즈니스 마인드와 분석적 마인드를 바탕으로 현대 정보사회의 방대한 정보를 보다 효과적·효율적으로 분석·평가하고 미래를 예측할 수 있는 분석경영전문가(Business Analyst), 데이터과학자(Data Scientist) 등의 전문인력을 양성하는 것을 목표로 한다.
빅데이터 학과간 협동과정	충북대학교 대학원	IT, BT 분야에 사회과학, 비즈니즈, 의학, 인문, 예술 등의 분야를 추가하여 융합 교과과정을 신설함으로써 창의적인 융합인재 양성을 목적으로 함.
빅데이터 분석학 협동과정	이화여자대학교 대학원	시대가 요구하는 빅데이터 분석 전문가 양성을 목적으로 한다. STEM(과학,기술,공학,수학)영역에서 전통적으로 수적인 열세에 있는 여성인력을 양성하고, 도입 초기의 학문인 빅데이터 관련 이론과 연구 및 실제 활용을 활성화하고, 산업계에서 필요로 하는 이론과 실습을 겸비한 융합형 전문 고급 인력(데이터과학자)를 양성한다

“미래 성장동력으로서 빅데이터의 생태계를 조망하고, 개인정보에서 공공정보까지 데이터의 가치와 데이터 가치 활용도를 높이는 분석의 중요성을 토대로 학술적 성과를 도모하는 것이다”

또 다른 학회로 한국빅데이터서비스학회(Korean Society for Big Data Service, http://kbigdata.or.kr/)가 있다. 2013년 설립되었으며, 설립취지는 다음과 같다.

“IT 및 초대규모 데이터 관리를 기반으로 하는 IT 기술과 수학 및 통계영역

을 기반으로 하는 고도 데이터 분석기술 및 빅데이터를 공공 및 산업 분야에 다양하게 적용하여 효과를 극대화하는 응용 기술 분야의 산업과 학문 발전을 목적으로 설립되었다"

대학알리미에서 빅데이터를 명시한 대학연구소를 검색한 결과(2015년 공시자료) 모두 6개의 대학에서 연구소 또는 연구원을 설립한 것으로 나타났다(표 2.14참조).

이 밖에도 여러 대학에서 2014년 이후 빅데이터 관련연구소를 설립하여 활발히 운영하고 있으나, 2015년 기준 대학알리미 자료가 2014년 상황까지 정리되어있어, 추가로 소개하는 것을 생략하였다.

표 2.14 대학 연구소 리스트

연구소명	대학	학문분야
빅데이터 연구소	건국대학교	공학
빅데이터연구센터	경희대학교	사회과학
빅데이터문화창조연구소	동의대학교	공학
빅데이터분석연구소	명지대학교	사회과학
빅데이터연구원	서울대학교	공학
빅데이터산업기술연구소	세종대학교	공학

2. 빅데이터 교육을 위한 커리큘럼[33)]

2014년 미래창조과학부와 한국정보화진흥원은 '빅데이터 커리큘럼 참조모델 1.0'을 전국의 빅데이터 관련 대학(원)과 협회, 학회 등에 보급한다고 발표하였다. 참조모델 1.0은 국내외 85개 대학/대학원 및 글로벌 전문기업의 빅데이터 관련 커리큘럼 분석 내용을 기반으로 작성됐으며 학계 및 업계의 검토를 거쳐 전문성을 높였다. 이 모델은 4대 핵심역량인 기반역량(Foundation), 기술역량(Platform Technique), 분석역량(Analysis Technique), 비즈니스업무역량(Business Analytics)을 토대로 전문가 수준을 파악할 수 있도록 구성됐다(그림 2.9참조).

33) 이 내용은 '빅데이터 커리큘럼 참조 모델 Ver 1.0'의 내용을 옮겨오거나 발췌하여 정리한 것임

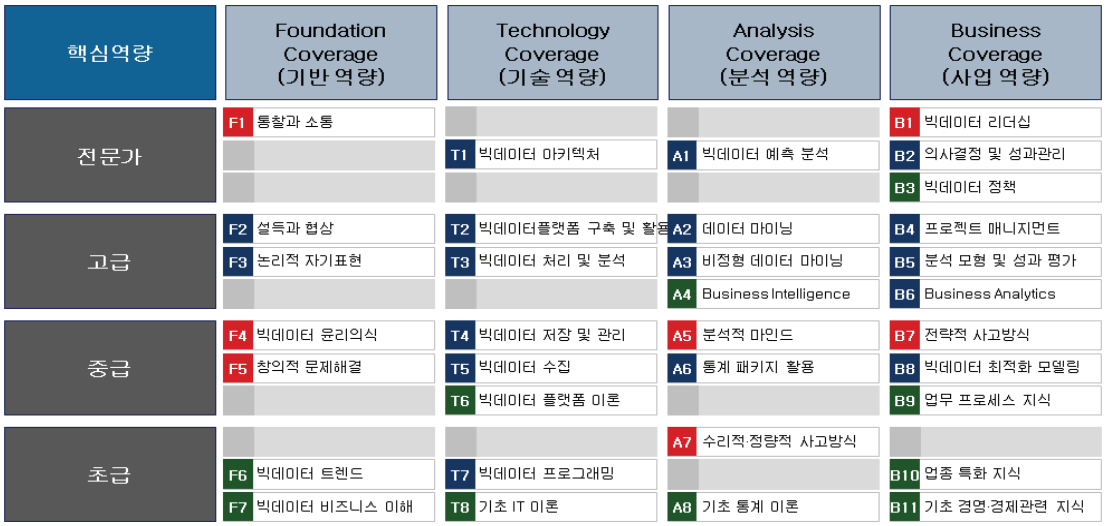


그림 2.9 빅데이터 커리큘럼 참조모델 1.0

좀 더 세부적인 사항은 표 2.15를 참고하면 알 수 있다.

표 2.15 4개 핵심역량별 34개 역량

역량 영역	핵심 역량	핵심 역량 설명
기반 역량 (Foundation)	빅데이터 비즈니스 이해	빅데이터와 비즈니스의 상관관계를 이해하는 역량
	빅데이터 트렌드	빅데이터 최신 동향 및 향후 발전 방향을 이해하는 역량
	빅데이터 윤리의식	빅데이터 보안 및 개인정보 보호를 실현할 수 있는 역량
	창의적 문제해결	창의적으로 빅데이터 활용 방안을 도출할 수 있는 역량
	통찰과 소통	사물이나 현상을 통찰하여 빅데이터 활용 방안을 제시할 수 있는 역량
	논리적 자기표현	빅데이터 활용 결과를 논리적으로 표현 할 수 있는 역량
	설득과 협상	빅데이터 사업 기회를 공유하고 실현시킬 수 있는 역량
기술 역량 (Platform Technique)	기초 IT 이론	IT에 대한 기본적인 이론과 지식을 이해하는 역량
	빅데이터 프로그래밍	빅데이터에 사용되는 프로그램 언어를 활용할 수 있는 역량
	빅데이터 플랫폼 이론	빅데이터 플랫폼의 개념과 주요 기능을 이해하는 역량
	빅데이터 수집	빅데이터 수집 유형, 방법 및 핵심 기술을 이해, 사용할 수 있는 역량
	빅데이터 저장 및 관리	빅데이터 저장 유형, 방법 및 핵심 기술을 이해, 사용할 수 있는 역량
	빅데이터 처리 및	빅데이터 분석을 위한 처리 유형, 방법 및 핵심 기술을

역량 영역	핵심 역량	핵심 역량 설명
	분석	이해, 사용할 수 있는 역량
	빅데이터 플랫폼 구축 및 활용	빅데이터 플랫폼 응용 시스템을 사용, 신규 개발 할 수 있는 역량
	빅데이터 아키텍처	빅데이터 IT 환경을 설계하고, 운영을 총괄할 수 있는 역량
분석 역량 (Analysis Technique)	수리적/정량적 사고방식	숫자(정량적) 중심으로 문제를 해석하고 결과를 도출할 수 있는 역량
	분석적 마인드	다양한 대량의 수와 통계치에서 새로운 Insight를 도출할 수 있는 역량
	기초 통계 이론	기본적인 확률//통계 이론, 분석기법 등을 이해하는 역량
	통계 패키지 활용	사용 목적에 따라 다양한 통계 패키지를 사용 및 응용할 수 있는 역량
	Business Intelligence	경영의사결정을 지원하는 분석 결과물을 기획, 산출할 수 있는 역량
	데이터 마이닝	데이터 마이닝의 지식을 바탕으로 데이터 처리/분석할 수 있는 역량
	비정형 데이터 마이닝	비정형 데이터의 이해 및 비정형 성격 별 데이터 처리/분석할 수 있는 역량
	빅데이터 예측 분석	빅데이터 분석을 통해 미래를 예측할 수 있는 역량
사업 역량 (Business Analytics)	기초 경영/경제 관련 지식	기본적인 경영/경제 이론적 지식을 이해할 수 있는 역량
	업종 특화 지식	산업(제조/유통/통신 등)의 핵심 업무 및 산업 빅데이터 전략을 이해할 수 있는 역량
	업무 프로세스 지식	업무 프로세스(영업/마케팅/생산 등) 및 빅데이터 적용 전략을 이해/수립할 수 있는 역량
	빅데이터 최적화 모델링	산업/업무 별 빅데이터 사업전략을 수립할 수 있는 역량
	Business Analytics	빅데이터 사업전략에 따라 분석 방향성 및 적용 방안을 수립할 수 있는 역량
	분석 모형 및 성과 평가	분석 전략 및 기획 결과를 기반으로 분석을 수행하고, 그 결과를 평가할 수 있는 역량
	프로젝트 매니지먼트	목표 성과를 창출할 수 있도록 빅데이터 프로젝트를 관리할 수 있는 역량
	빅데이터 리더십	빅데이터 사업의 이해관계자 관리 및 총괄 책임자로서 사업을 추진할 수 있는 역량
	의사결정 및 성과관리	빅데이터 사업 중 발생하는 다양한 의사결정사항을 효과적으로 수행할 수 있는 역량

역량 영역	핵심 역량	핵심 역량 설명
	전략적 사고방식	기업 비전/전략과 빅데이터 사업 전략의 상관관계를 이해/향상 시킬 수 있는 역량
	빅데이터 정책	국내외 ICT 및 빅데이터 정책을 이해하고 이를 빅데이터 사업에 적용할 수 있는 역량

3. 향후 10년을 이끌 Top 10 빅데이터 기술[34)]

향후 Top10 빅데이터 기술은 현재 높은 가치를 보유하고 있더라도 3년 내에 새로운 기술 가치로 전환될 것으로 예측되는 기술은 상위 순위로 배치하지 않고, 향후 5년까지 성장세를 지속할 것으로 예측되는 기술을 중심으로 선정하였다.

(1) Predictive analytics

'Predictive analytics'는 향후 10년간 안정적 성장세를 보이고, 비즈니스적 가치가 높은 기술로 평가되었으며 기업이 자사 및 자사 제품의 시장에서의 평가, 마케팅의 최적화를 통해 비즈니스적 성과를 향상 시키고, Risk를 줄이는 방법을 제공하는 가장 우수한 솔루션으로 선정하였다(그림 2.10 참조).

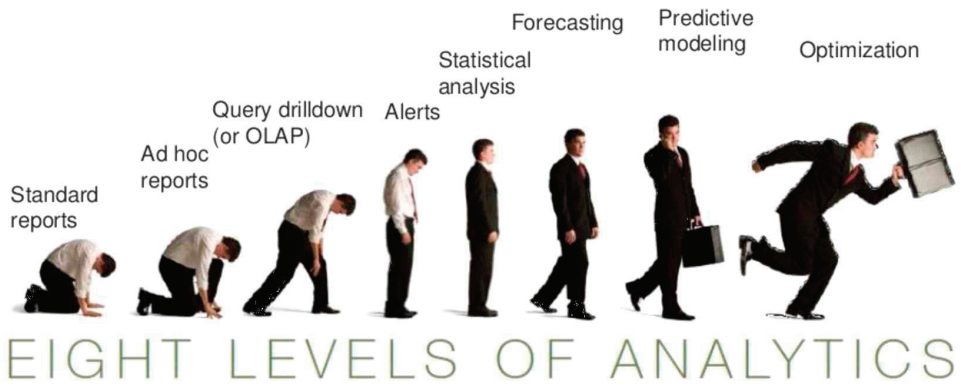


그림 2.10 자료의 분석단계

출처: SAS

34) 이 내용은 한국정보화진흥원 '빅데이터 동향과 이슈 제21호'의 내용을 옮겨오거나 발췌하여 정리한 것임

(2) NoSQL databases

데이터베이스는 그동안 계층형, 네트워크형, 관계형으로 발전해 왔으나 빅데이터 처리가 일반화되면서 NoSQL에 대한 관심과 가치가 크게 높아지는 추세로 향후 5년까지는 NoSQL의 기술 가치는 지속될 것으로 분석하였다. NoSQL 기술은 복수의 저가 서버들을 클러스터링, 샤팅 등의 방법을 통해 데이터를 분리하여 처리하고 있어 보다 대량의 데이터를 빠르게 처리할 수 있는 기술로 평가받고 필요에 따라 Key-value, Documents, Graph Database 등으로 활용할 수 있다는 점이 특징이다.

(3) Search and knowledge discovery

정형 또는 비정형 데이터의 검색과 빅데이터 분석을 위해 제공되는 툴과 기술은 향후 5년간 지속적인 성장세를 보일 것으로 예측하였다. 이는 단순한 검색 도구가 아니라 파일 시스템, 데이터베이스, 스트림, API 및 기타 플랫폼, 애플리케이션에 있는 다양한 유형의 데이터를 분석하여 새로운 Insight를 제공하는 툴로서 매우 중요한 가치를 지니는 것으로 평가된다.

(4) Stream analytics

네트워크 또는 시스템에서의 서로 다른 다양한 데이터 소스, 포맷이 다른 데이터 소스를 모두 취합하고 분석하여 필요한 데이터를 선택 후 처리하는 소프트웨어 분야가 향후 높은 성장세를 보일 것으로 평가하였다. 데이터 분석 및 추론 등을 통해 얻는 지능은 데이터의 규모가 커질수록 다방면의 최신 데이터를 통합하여 얻을수록 더 내재 가치가 높은 지능을 얻을 수 있으며 분석 결과의 정확도도 높아진다는 측면에서 대량의 이종 데이터를 실시간으로 분석할 수 있는 환경 구축 가능성이 점차 커지고 있어 이를 지원하기 위한 기술도 발전할 것으로 예상된다.

(5) In-memory data fabric

In-memory database는 디스크가 아닌 주 메모리에 모든 데이터를 보유하고

있는 데이터베이스로 디스크 검색보다 자료 접근이 훨씬 빠른 것이 가장 큰 장점이며 데이터양의 빠른 증가로 데이터베이스 응답 속도가 떨어지는 문제를 해결할 수 있는 대안이 되고 있다. 인메모리 방식은 메모리상에 색인을 넣어 필요한 모든 정보를 메모리상의 색인을 통해 빠른 검색이 가능하다는 점에서 높은 성장세가 예측되는 분야이다.

(6) Distributed file stores

저장해야 하는 데이터양이 크게 증가하면서 필요 저장 장치의 수가 증가하게 되었으며 다수의 저장 장치 이용 시 일부 저장장치가 사용 불가능하여도 파일의 가용성을 보장할 뿐만 아니라 대용량의 데이터를 빠르게 처리할 수 있다는 점에서 파일 분산 저장 기술은 필수적이다.

(7) Data virtualization

데이터 가상화는 이종으로 구성된 인프라스트럭처를 하나의 논리적으로 통합된 리소스인 것처럼 접근, 관리, 최적화할 수 있게 해주는 방법으로 이를 통해 일부 서비스, 기능 또는 기타 리소스의 내부적 구현에서 외부 인터페이스를 추상화할 수 있다는 장점을 가지고 있어 빅데이터 기술로 밝은 전망을 갖고 있다.

(8) Data integration

산재된 데이터 소스와 시스템, 애플리케이션 등으로부터 데이터를 취합하고 가치를 지니는 정보가 되도록 만들어주는 과정을 의미하며 데이터를 복제하거나 애플리케이션의 기존 소스를 변경하지 않고도 즉시로 이용할 수 있게 한다는 점에서 우월하며 정기적으로 반복되지 않는 물음에 신속한 대답을 얻을 수 있다는 것이 강점이다.

(9) Data preparation

데이터 준비 프로세스의 자동화로 복잡하고 시간 소비적인 수동 데이터 준비 시간을 절약할 수 있으며 데이터의 유효성 검증을 더욱 빠르고 정확하게 수행

할 수 있는 방법론을 제공한다.

(10) Data quality

빅데이터에 대한 자동 데이터 프로파일링, 데이터 품질 분석, 데이터 품질 분석 결과 보고 등을 제공하는 품질 관리 솔루션의 중요성이 크게 증가할 것으로 예상된다.

CHAPTER 3

우리나라 빅데이터 보유 현황

학습목표

- 우리나라 (행정자료)빅데이터 보유현황을 통계청을 중심으로 정리한다.
- 중앙행정기과 제공 자료들의 데이터 소재정보와 내용을 파악한다.
- 공공기관, 지방자치단체, 민간 및 기업 제공 데이터 소재정보를 파악한다.

정부 주도의 빅데이터 정책과 정부 3.0에 의한 공공데이터 개방으로 우리나라의 빅데이터 또는 활용 가능한 공개된 데이터는 주로 행정자료이거나 이와 관련된 자료들로 구성되어 있는 것이 현실이다. 그림 3.1은 우리나라 데이터의 현황을 나타내는 여러 지표를 나타낸 것이지만, 이는 규모는 추정할 수 있을 뿐이지 실제 사용가능한 데이터에 관한 것은 아니다. 공개되고 있는 데이터의 규모를 알아보기 위하여 2015년 행정자치연감의 공공데이터 개방현황과 공공데이터 민간 활용 실적을 알아보았다.

표 3.1 공공데이터 개방현황

구분	기관 수		데이터셋 수		오픈API 수		합계	
연도	2013	2014	2013	2014	2013	2014	2013	2014
계	556	736	4,718	11,796	554	1,612	5,272	13,408
중앙행정기관	76	101	950	1,795	178	471	1,128	2,266
지방자치단체	209	244	2,846	7,630	112	295	2,958	7,925
교육행정기관	91	125	164	226	1	17	165	243
기타공공기관	180	266	758	2,145	263	829	1,021	2,974

출처 : 2015년 행정자치통계연보

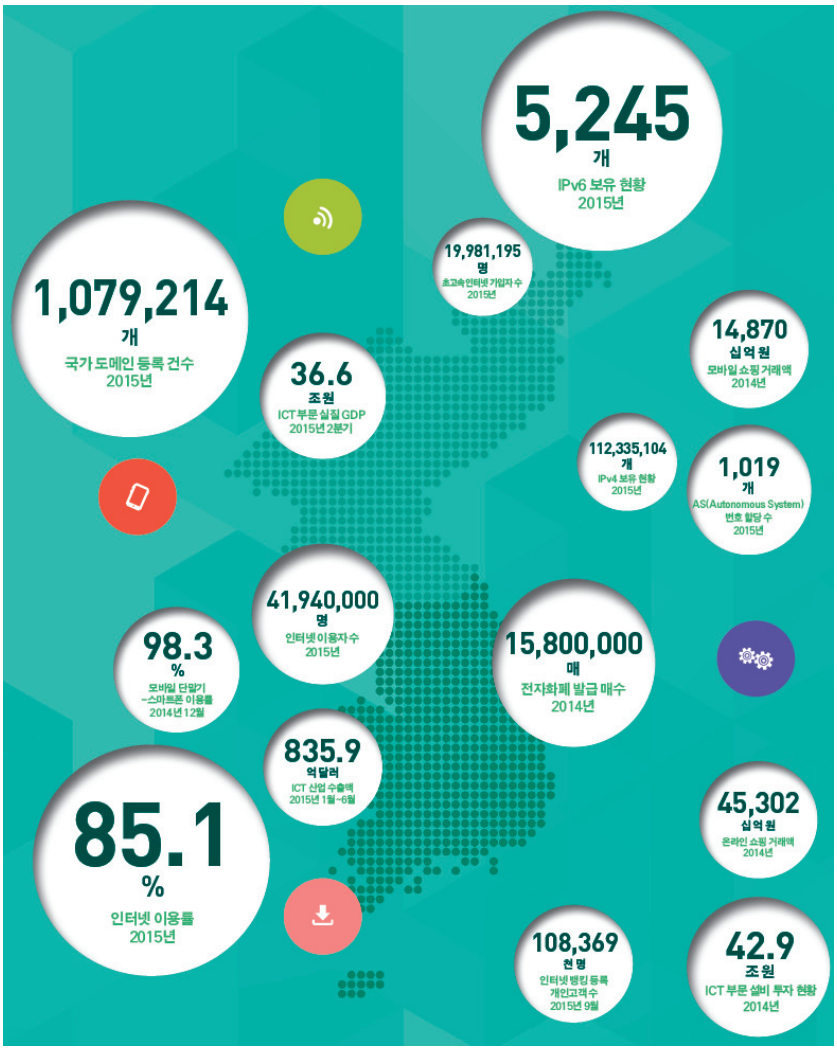


그림 3.1 국내현황 데이터

출처 : 2015 국가정보화백서

표 3.2 공공데이터 민간 활용 실적

구분	데이터셋 다운로드건 수		오픈API활용 신청건 수		합계	
연도	2013	2014	2013	2014	2013	2014
계	4,108	115,525	9,815	23,872	13,923	139,397
중앙행정기관	2,205	31,663	4,743	11,579	6,948	43,242
지방자치단체	1,233	56,770	2,681	4,209	3,914	60,979
교육행정기관	45	1,671	0	30	45	1,701
기타공공기관	625	25,421	2,391	8,054	3,016	33,475

출처 : 2015년 행정자치통계연보

정부 3.0 정책으로 인하여 2013년과 2014년 비교에서 공공데이터 개방과 민간 활용 실적이 급증한 것을 알 수 있다. 많은 정보가 공개되고 활용되고 있는 것을 알 수 있지만, 표 3.1과 표 3.2로는 내가 원하는 정보를 어디서 찾을 수 있는지는 알 수 없다. 물론 대표적인 정보포털사이트를 알고 있다면, 이를 통해서 자료를 구할 수 있겠지만, 더 세부적인 정보는 원 자료를 제공했던 사이트에 있을 수도 있다. 이와 같이 내가 찾고자 하는 사이트들이 어디에 있는지를 정리한 것을 '데이터 소재정보' 또는 '데이터 맵'이라고 한다.

아직까지 우리나라 데이터 소재정보(분석용 자료에 관한 분야)는 제공되지 않고 있어서 통계분석을 위한 자료제공을 중심으로 소재정보를 파악해 보았다. 이 과정에서 정리할 수 있었던 구조적인 형태는 자료의 제공이 주로 행정기관 및 관련기관들 위주로 이루어지고 있다는 것이며, 그 중심에 통계청의 국가통계정보포털(KOSIS) 있다는 것이다. 민간분야에서도 제공이 되는 것이 있지만 상대적으로 매우 적은 규모인 것으로 나타났다. 이는 자료의 공개여부로 판단해야 할 부분이다. 데이터가 자산으로 여겨지는 상황에서 민간분야의 자료공개는 쉽지 않은 문제이다. 또한 정보를 제공하는 사이트들은 제공방식에 따라 다음의 7가지 유형으로 구분할 수 있었다.

유형 1) 자료의 검색, 다운로드, 분석까지 가능한 정보포털 사이트
유형 2) 자료의 검색, 다운로드 가능한 정보포털 사이트
유형 3) 화면에서 데이터 확인이 가능하며, 다운로드 가능
유형 4) 화면에서 데이터 확인만 가능
유형 5) 연감자료, 보고서 등 비정형자료 제공, 데이터 파일 제공
유형 6) 데이터를 파일형태로 제공
유형 7) 연감자료, 보고서 등 비정형 자료 제공

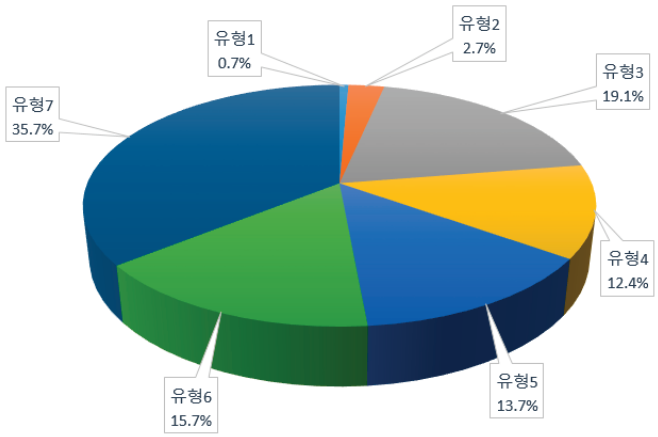


그림 3.2 데이터 제공 사이트들의 유형 별 분류

전체 445개의 사이트들 중에서 편리하게 사용할 수 있는 유형 1과 유형 2에 속하는 경우는 모두 15곳 3.4%로 그 수가 적음을 알 수 있고, 연감자료, 보고서 등을 파일 형태(PDF, Excel)로 제공하는 경우인 유형 5,6,7이 290곳 65.2%로 전체의 반 이상을 차지하고 있었다(그림 3.2참조).

445개의 자료제공 사이트 구분을 보면, 중앙행정기관 관련, 산하기관, 공공기관, 민간기관(금융, 협회, 기타 등), 지방자치단체 등으로 구분할 수 있었다. 이들을 정리한 결과는 다음과 같다.

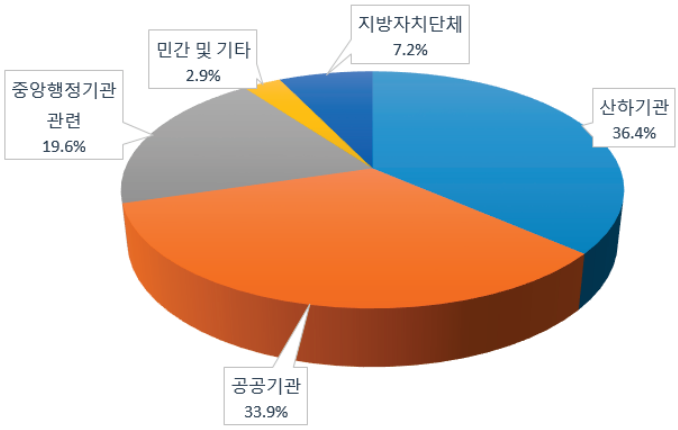


그림 3.3 자료제공 기관별 데이터 제공 사이트 구분

자료를 제공하는 기관들의 97%정도가 공공기관의 성격을 갖고 있어 국가의 역할이 중요한 것을 확인할 수 있다(그림 3.3참조).

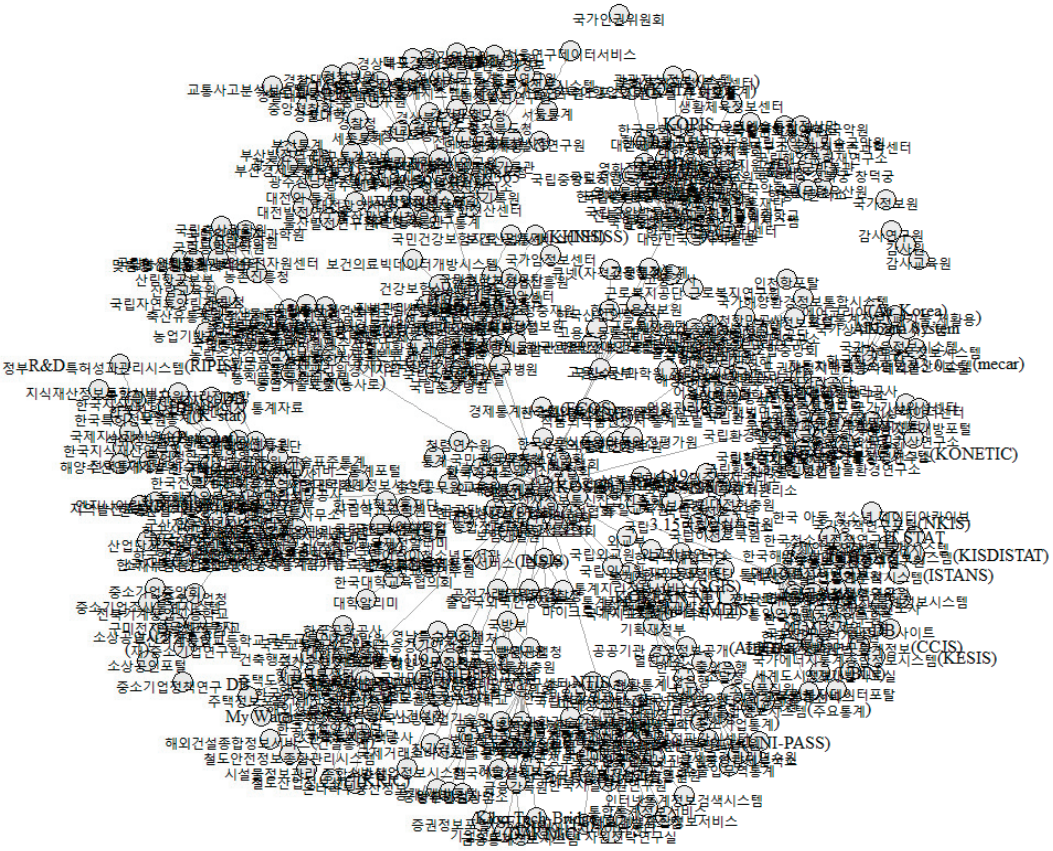


그림 3.4 데이터 소재 cytoscape

데이터 소재를 정리해보면 공공정보 중심이 될 수밖에 없으므로 KOSIS에서 제공하는 정보와 KOSIS에서 제공하지 않는 정보로 구분할 수 있다. KOSIS가 데이터 소재정보를 파악하는 중심이 될 수밖에 없는 이유이다.

그림 3.4는 조사한 445개 정보사이트의 연결관계를 정리해서 그려본 cytoscape 이다. 이 그림의 중심에는 KOSIS가 있지만 너무 복잡해서 연결관계와 각 사이트의 이름은 확인이 어렵다. 이 그림의 가독성은 없지만 전체를 파악하기 위해 제시하는 것이고, 이제 이렇게 많은 데이터들이 어디에 어떤 형식으로 존재하는지를 알아보도록 하자. 특히 대부분의 정보가 모이고 있는 KOSIS의 운영부처인 통계청에 대해서 먼저 알아보도록 하자. 이 장에서 제시하는 모든 정보제공 사이트의 주소 및 기본 정보는 부록 10에 정리되어있으며, 본문에서는 표로 정리되어있는 사이트의 주소는 가독성과 편집을 위해 제외하였다.

3.1 통계청

통계청(http://kostat.go.kr/portal/korea/index.action)을 방문해 보면 통계청 핵심서비스로 모두 5개의 정보서비스를 각각의 웹사이트를 통해서 제공하고 있다(그림 3.5 참조).



그림 3.5 통계청 핵심서비스

이들 각 서비스는 각각의 특징을 갖고 있으며, 활용성이 매우 높은 정보들을 제공하고 있으므로 각각에 대한 소개를 자세히 하도록 한다.

3.1.1 국가통계포털

우리나라 행정자료 또는 이와 관련된 대부분의 정보가 통합되는 국가정보포털인 KOSIS(Korean Statistical Information System, http://kosis.kr/)는 국내, 국제, 북한 주요통계를 한 곳에 모아 이용자가 원하는 통계자료를 한 번에 찾을 수 있도록 통계청이 제공하는 서비스로. 335개 기관 883종의 승인통계가 정리되어 있으며 125,205개의 통계표가 존재한다(2015년 말 기준). 방대한 자료를 무료로 제공하며, 이용자 편의성을 지속적으로 개선하고 있다. 2015년 628만 명이 방문하였으며 하루 평균 17,200명이 방문하는 통계청의 대표적인 서비스 사이트이다.

KOSIS에서 제공되는 자료는 '주제별통계', '기관별통계', '과거·중지통계'로 구분되어있다. 제공분류를 2015년 6월 기준으로 알아보면 다음과 같다(그림 3.6 참조).

- 주제별통계: 제공 통계를 내용에 따라 16개 주제로 분류하여 제공
- 기관별통계: 제공 통계를 중앙행정기관(34개), 지방자치단체(17개), 금융기관(8개), 공사/공단(24개), 연구기관(19개), 협회/조합(20개), 기타기관(15개)로 분류하여 제공

- 과거·중지통계: 이 부분은 ‘광복이전통계’를 14개 주제로 제공하며, ‘대한민국통계연감’을 16개 주제로 제공한다. 통계연감을 문서형식으로 제공하는 것이 아니라 검색 가능한 자료형식으로 제공한다. 마지막으로 ‘작성중지통계’를 61종 별로 제공한다.



그림 3.6 KOSIS 메인화면

KOSIS에서 제공하는 추가적인 자료로 온라인간행물이 있다. 모두 14개 주제 95종을 제공하며, 엑셀(Excel) 또는 PDF 형식으로 자료가 제공된다.



그림 3.7 KOSIS 통계설명 자료

KOSIS의 장점은 자료의 검색, 분석 기능과 같은 것도 있지만 가장 큰 것은 '통계설명자료'이다. 여기서는 KOSIS에서 제공하는 통계에 대한 설명을 볼 수 있어 사용하는 자료에 대한 이해를 높일 수 있다. 추가적으로 조사통계 작성을 위하 사용된 조사표들도 조회할 수 있다(그림 3.7 참조).

KOSIS는 자료를 생산하는 모든 행정·공공기관의 자료가 정리되어있다고 보아도 무방할 정도의 방대한 자료를 제공하고 있다. 그렇다면 모든 자료들이 KOSIS에서 제공되는 것인가? 이 물음에 대한 답변은 '그렇지는 않은 것 같다'이다. 왜냐하면 KOSIS에 자료를 제공하는 행정·공공기관들 중에는 별도의 정보포털을 운영하는 곳들이 적지 않게 있기 때문이다. 모든 자료를 제공한다면 굳이 그럴 필요가 없을 것이기 때문이다. KOSIS에 자료를 제공하는 기관들을 중심으로 데이터 소재정보(정보포털 포함)를 부록 10에 정리하였다. 너무 많은 곳이 존재하여 본문에는 제시하지 못하였으며, 이후 각 기관별로 정보포털 중심으로 설명을 하도록 한다.

3.1.2 국가지표체계

국가지표체계(http://www.index.go.kr/potal/main/PotalMain.do)에서는 국정에 관한 주요통계지표를 분석자료와 함께 제공하며 제공지표를 내용, 생산부처 등의 기준으로 정리된 체계도도 함께 제공한다. 정보의 홍수 속에서 국민들이 우리나라를 보다 쉽고 체계적으로 파악할 수 있는 것을 목적으로 한다. 제공하는 지표는 크게

○ 국가주요지표(국가 발전 상황을 균형 있게 보여주는 분야별 핵심지표)
○ e-나라지표(다양한 방면에서 우리나라 현 위치를 보여주는 지표)
○ 국민 삶의 질 지표(국민의 삶의 현 위치를 보여주는 지표)
○ 녹생성장지표 (녹생성장의 현 위치를 보여주는 지표)

로 구성되며, 화면에서 선택 후 내용을 확인할 수 있다(그림 3.8 참조).

1. 국가주요지표

국민의 웰빙과 국가 발전(national progress)의 주요 상황을 체계적으로 파악할 수 있도록 경제・사회・환경 3개 부문 – 15개 영역 – 60개 하위영역으로 구성되어있다. 제공되는 지표의 수는 국민의 관심이 크고 주요 정책 수립에 긴요하게 활용될 수 있는 성과(outcome) 중심 지표로 주요지표 139개, 주요지표를 보완할 수 있는 보조지표 186개, 다른 국가와의 비교를 위해 국제비교지표 125개로 총 450개의 지표가 제공된다. 사용자가 쉽게 이해할 수 있도록 그래프 형태로 제공되며, 지표의 정의와 측정방법, 해설, 상세통계표, 용어해설을 함께 제공하고 있다.

2. e-나라지표

e-나라지표 시스템은 국정 전 분야에 걸친 각종 국정통계에 대해 정확한 통계와 추이, 생생한 통계의미 분석 등을 제공함으로써 국민, 전문가, 정책결정자들 모두에게 국정을 알기 쉽게 설명하고, 모니터링할 수 있도록 할 뿐만 아니라 필요한 국정통계정보를 손쉽게 접근할 수 있도록 하는 것을 목표로 한다.

제공하는 지표들은 국가 공식 승인 통계자료뿐만 아니라, 현황이나 각종 행정 자료들을 가공한 내용들도 포함하며 시계열 자료를 통하여 정책 결과의 변동을 볼 수 있고, 이를 그래프로 도식화하여 쉽게 그 추이를 알 수 있도록 하였다.



그림 3.8 국가지표체계

3. 국민 삶의 질 지표

삶의 질은 삶을 가치 있게 만드는 모든 요소를 포괄하는 개념으로 "객관적인 생활조건과 이에 대한 시민들의 주관적 인지 및 평가"로 구성된다. 여기서는 개인의 '삶의 질'과 함께 전체 사회와 관련된 '사회의 질'도 포함하는 개념으로 활용된다. 삶의 질 지표는 물질과 비물질 부분 12배 부문 81개 지표를 제공한다.

4. 녹색성장지표

국가의 저탄소녹색성장을 위한 국가전략에 따른 '녹색성장 5개년 개획'의 정책지원 도구로서 녹색성장통계 작성을 실시하였다. 녹색성장정책지표로 22개를 제공하고 있으며, 국민의 녹색생활 실천수준을 파악하고 표준적 녹색생활 양식을 홍보하기 위해 녹색가정, 녹색교통, 녹색커뮤니티의 3대 분야와 7대 관심영역에 대하여 46개 지표를 제공한다. 녹색생활 실천에 따른 성과를 파악하기 위한 지표로, 녹색생활실천지표와 같은 '녹색가정','녹색교통','녹색커뮤니티' 3대 분야와 7대 관심영역에 대하여 13개의 지표를 제공한다.

또한 우리나라의 지속가능한 발전수준을 점검하고 바람직한 국가의 발전방향을 제시하기 위하여 사회/경제/ 환경 3개 분야에 대하여 77개 지표를 작성하여 제공하고 있다. 또한 리오회의 이후 각국의 환경개선 이행 정도를 측정하기 위한, 환경보건 및 생태계 건전성의 2대 정책목표와 9개 정책부문에 대하여 20개 지표로 구성하고 작성한다. 녹색성장지표에서는 총 178개의 지표를 제공한다.

3.1.3 통계지리정보

SGIS 플러스(http://sgis.kostat.go.kr/view/index)는 Statistical Geographic Information Service의 약자로서 통계청이 보유한 인구·가구·주택·사업체 등 공간통계정보와 각종 외부자료를 융·복합하여 소지역 공간의사결정을 지원하는 개방형 공간통계정보 플랫폼이다.

또한 사용자에게 통계정보와 지리정보를 융·복합하여 새로운 서비스를 만들 수 있는 기반을 지원한다. 포털 서비스를 통해 사용자가 직접 플랫폼에서 제공하는 다양한 인터랙티브맵, 통계주제도 등의 다양한 서비스를 이용할 수 있다(그림 3.9 참조).

SGIS에서는 통계지도 이외에도 정부기관 및 민간에서 활용하여 더 큰 부가가치를 창출할 수 있도록 자료를 제공하는 것으로 현재 센서스 공간통계 자료를 파일형태로 서비스하고 있다. 제공하는 데이터는 집계구별 인구, 가구, 주택, 사업체 통계자료와 지도 작성을 위한 센서스용 행정구역경계(전체, 시도,

시군구, 읍면동)를 shape 파일 형태로 제공한다.



그림 3.9 통계지리정보 서비스

3.1.4 통계분류포털

우리나라와 같은 분산형 통계제도에서는 각종 경제, 사회, 보건 분류의 제정·개정, 통계용어 정비 등 통계 기준을 명확히 설정하는 것이 통계의 정확성과 비교성 제고를 위해 매우 중요하다.

이에 통계청장은 통계작성기관이 동일한 기준에 따라 통계를 작성할 수 있도

록 국제표준분류를 기준으로 산업, 직업, 질병·사인, 무역 등에 관한 표준분류를 작성·고시하도록 통계법 제22조에 규정하고 있다.



그림 3.10 통계분류포털

통계청에서는 통계분류포털(https://kssc.kostat.go.kr:8443/ksscNew_web/index.jsp)을 통해 이러한 표준분류 이외에도 14종의 특수목적분류를 제정·운영하고 있으며, 그 밖에도 행정구역, 상품용도, 가계수지항목, 생활시간, 교육, 종사상지위, 한국재화 및 서비스 분류 등 7종의 기타분류를 운영하고 있다. 특히, 특정산업분야의 통계작성을 위해 제정된 11종의 특수산업분류는 표준산업분류에서 관련 분야의 항목을 선정하여 그 목적에 맞게 만든 것으로, 관련 산업을 일관성 있고 체계적으로 조사·분석할 수 있는 매우 유용한 분류이다. 최근 들어 통계에 기반하여 정책을 수립·추진하려는 각 부처로부터 지원대상 산업의 범위와 구조 등 실태 파악을 위한 특수 분류 제정 요구가 증가하고 있는 실정이다. 통계분류포

털에서는 표준분류 5종, 특수목적 산업분류 11종, 특수목적 직업분류 3종, 기타분류 7종으로 모두 26종의 분류자료를 제공하고 있다(그림 39 참조).

3.1.5 마이크로데이터

MDIS(Microdata Integrated Service, https://mdis.kostat.go.kr/index.do)에서는 통계청 40종 및 통계작성기관 21종의 마이크로데이터에 대하여 온라인 통합 서비스를 제공하고 있다. 공표자료(KOSIS에서 제공)이외의 통계자료는 MDIS를 통해 제공한다. 서비스 제공은 다음의 5가지 유형이 있으며, 수수료가 발생한다(그림 3.11 참조).



그림 3.11 마이크로데이터 통합 서비스

- 추출다운로드 : 사용자가 마이크로데이터를 직접 다운 받아 사용(무료)
- 원격접근서비스 : 사용자가 집이나 사무실에서 인터넷을 통해 승인용 마이크로데이터를 분석할 수 있는 서비스. 데이터는 다운되지 않고 분석 결과표만을 승인 받아 반출
- 이용센터 서비스 : 지정된 장소에서 승인용 또는 특수목적용 마이크로데이터를 이용하여 분석하는 서비스. 데이터는 다운되지 않고 분석 결과표만을 승인 받아 반출
- 주문형서비스 : 공용 마이크로데이터를 이용하여 이용자가 원하는 분석 결과표를 작성해주는 서비스
- 마이크로데이터 CD : 공공용 마이크로데이터를 CD로 제공

3.2 중앙행정기관

정부조직은 2015년 기준 17부 5처 16청/2원 5실 6위원회로 구성되어 있으며 각 조직에서 운영하는 홈페이지에서 대부분 자료를 제공하고 있다(그림 3.12 참조).

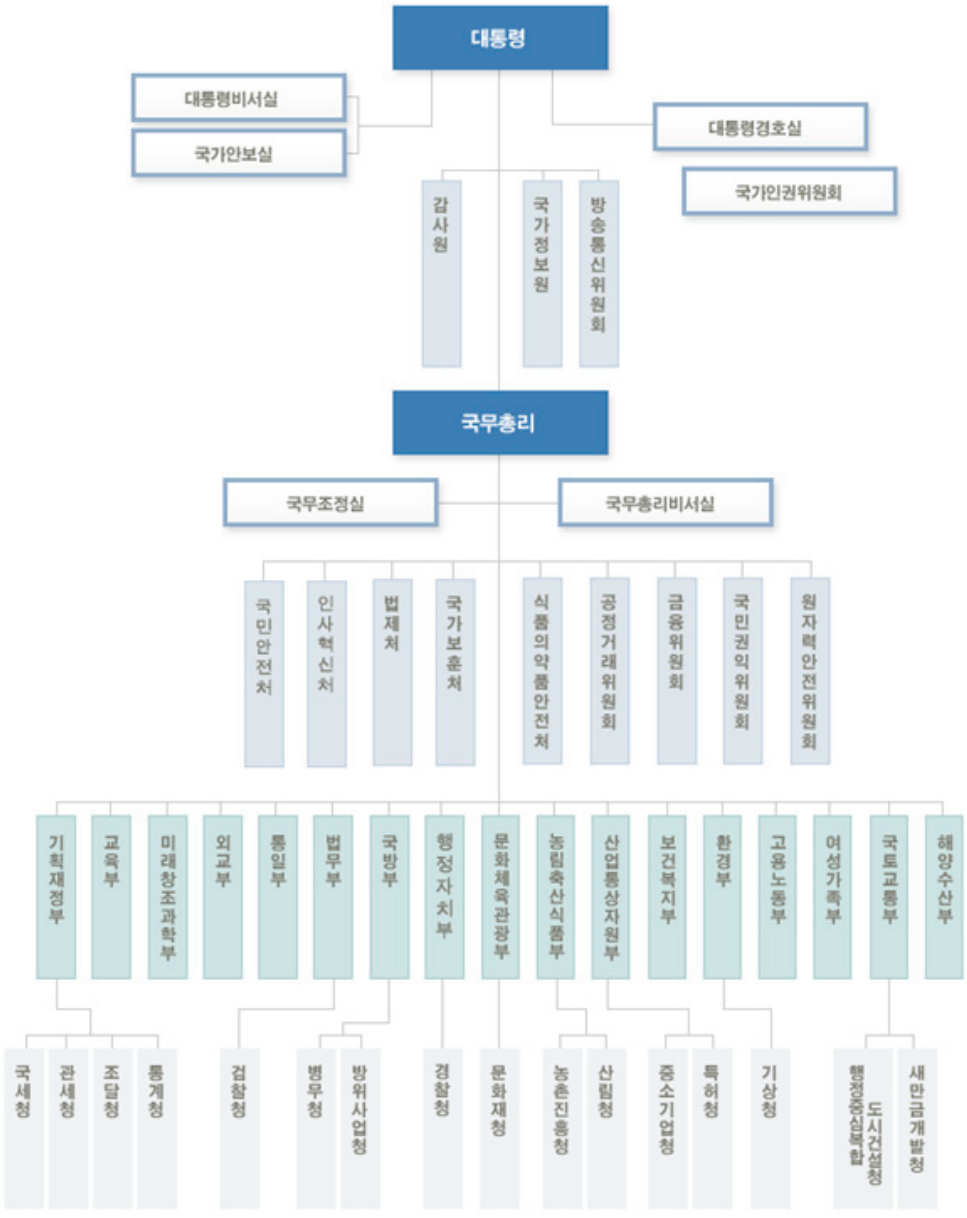


그림 3.12 중앙행정기관 조직도

출처 : 2015 행정자치통계연보

각 중앙행정기관에서 제공하는 자료는 기본적으로 국가통계포털(KOSIS)에서 확인할 수 있다. 그러나 각 행정기관에서는 각자의 홈페이지에서 정형, 비정형

자료를 제공하고 있으며, 경우에 따라서는 직접 또는 위탁기관을 통해 간접적으로 정보포털을 운영하고 있다. 그러므로 중앙행정기관에 연관된 사이트까지 정리하면 매우 복잡해지므로 이 절에서는 중앙행정기관의 홈페이지에서 직접 제공하는 자료만으로 정리하였다. 정리한 결과는 그림 3.13과 같다.



그림 3.13 중앙행정기관 제공정보 cytoscape

그림 3.13은 기관 중심으로 KOSIS에 연결되는 것을 정리한 것이므로 사실 그림 3.12와 차이가 없다. 대부분의 데이터들이 중앙행정기관의 관리 하에 생성되는 것이므로 그림 3.13만으로는 관계를 표현하기에 부족하고, 모든 관계를 그림에 넣으면 그림 3.4와 같이 복잡해진다. 따라서 표 3.3과 표 3.4에서 중앙행정기관 홈페이지에서 직접적으로 접근할 수 있는 정보사이트들을 정리하였다. 표 3.3과 3-4의 내용은 추후 공공기관, 산하기관 정리에서 중복이 발생할 수 있다.

표 3.3 중앙행정기관 통계자료 제공 사이트(KOSIS 연계)

중앙행정기관	사이트 이름
검찰청	1. 통계자료
경찰청	2. 경찰통계연보
고용노동부	3. 고용노동통계*
공정거래위원회	4. 통계연보

관세청	5. 수출입무역통계**
	6. 통관포털(UNI-PASS)*
교육부	7. 교육통계
	8. 지방교육재정 알리미*
	9. 교육통계서비스(한국교육개발원)**
국가보훈처	10. 통계자료
국민안전처	11. 주요통계
	12. 국민재난안전포털*
국세청	13. 국세통계
국토교통부	14. 국토교통 통계누리**
	15. 온나라부동산정보 통합포털**
	16. 국가공간정보포털*
	17. 실거래가 공개시스템
	18. 건축행정시스템(세움터)
기상청	19. 기상재해통계
	20. 기상자료개방포털**
기획재정부	21. 기획재정부 통계
	22. 열린재정*
	23. 공공기관 경영정보 공개시스템(ALIO)*
농림축산식품부	24. 통계정보, 물가정보
농촌진흥청	25. 농업기술포털(농사로)
문화재청	26. 통계정보
문화체육관광부	27. 국가도서관통계시스템*
	28. 문화통계포털(문화셈터)**
	29. 관광지식정보시스템*
미래창조과학부	30. NTIS 현황통계*
	31. ITSTAT**
방송통신위원회	32. 자료마당(파일자료)
법무부	33. 출입국 외국인정책본부 통계자료실
병무청	34. 공개/개방포털
보건복지부	35. 통계포털**
산림청	36. 산림통계
	37. 맞춤형 산림통계서비스*
산업통상자원부	38. 통계정보
	39. 국가기술표준통계
	40. 국가에너지통계 종합정보시스템(KESIS)**
	41. 산업연구원(KIET) 서비스통계포털**
	42. 소재부품 종합정보망
	43. 한국자원정보서비스*
	44. 정보통신통계서비스**(ITSTAT, 25와 동일)
	45. 전력통계*
	46. Petronet*
	47. 한국전자산업진흥회(전자산업정보)*
	48. 한국자동차산업협회(자동차통계유)*-유료회원
	49. 한국섬유산업연합회(섬유정보포털)*
	50. 한국정밀화학산업진흥회(정밀화학통계)*
	51. 한국산업단지공단(산업단지통계)*
	52. 지역발전포털(REDIS)**

식품의약품안전처	53. 통계자료
	54. 식품의약품안전처 통계포털**
여성가족부	55. 통계자료
인사혁신처	56. 통계정보(인사통계)
조달청	57. 나라장터 조달통합정보시스템(주요통계)
중소기업청	58. 중소기업조사통계시스템**
통일부	59. 통계자료
특허청	60. 지식재산권통계
해양수산부	61. 수산정보포털**
	62. 어업자원포털*
	63. 해양안전종합정보시스템*
행정자치부	64. 통계
	65. 내고장알리미
	66. 지방재정365*
	67. 주민등록 인구통계*
환경부	68. 환경통계포털**
	69. 환경디지털도서관

표 3.4 중앙행정기관 통계자료(KOSIS 비연계)

중앙행정기관	사이트 이름
감사원*	1. 감사연보
국가인권위원회*	2. 인권통계자료
국가정보원*	3. 산업스파이적발통계
금융위원회*	4. 금융시장동향(일일상황보고)
외교부*	5. 국가 및 지역정보
여성가족부	6. 통계자료
인사혁신처	7. 통계정보(인사통계)
조달청	8. 나라장터 조달통합정보시스템(주요통계)
중소기업청	9. 중소기업조사통계시스템**
통일부	10. 통계자료

중앙행정부처 홈페이지에서 바로 연결이 가능한 통계자료 제공사이트는 79개였으며, 이 중 KOSIS와 연계되어있는 곳이 69곳(표 28), 연계되지 않은 곳이 10곳(표 29) 었다.

특히 이 중 '**'표시가 된 곳(15개)은 관련분야의 다양한 정형자료를 검색·다운로드 할 수 있는 정보포털의 형식을 갖추고 있는 것을 표시한 것이다. '*'는 정보포털의 형식은 갖고 있으나 제공되는 자료의 다양성이나 검색기능이 상대적으로 부족한 경우에 해당한다. 아무 표시도 없는 경우는 홈페이지의 한 항목으로 통계자료를 제공하는 형식인 경우에 해당된다. '**' 표시된 곳은 자료분석을 위해 편리하게 참조할 수 있는 사이트들이므로 간략하게 소개한다.

(1) 수출입무역통계(관세청)



항목	내용
① 명칭	수출입무역통계
② 유형분류	화면에서 데이터 확인이 가능하며, 다운로드 가능
③ 주소	https://unipass.customs.go.kr:38030/ets/
④ 운영기관	관세청
⑤ 상위기관	기획재정부
⑥ 제공자료	수출입통계, 남북교역통계, 수출입 화물통계, 지역별 무역통계, 테마별 무역통계, 월간동향분석, 경제현안분석, 수입가격공개 등을 제공
⑦ 제공자료형식	정형 데이터(숫자), 비정형 데이터(월보)
⑧ 소개	- 일국의 경제영역으로 반입(수입)되거나 반출(수출)됨으로써 그 나라의 물적 자원을 증가 또는 감소시키는 모든 상품이 기록. - 한 나라를 통과하는 물품이나 일시적으로 반입, 반출되는 물품은 무역통계에 포함되지 않음.

(2) 교육통계서비스



항목	내용
① 명칭	교육통계서비스
② 유형분류	화면에서 데이터 확인이 가능하며, 다운로드 가능
③ 주소	http://kess.kedi.re.kr/index
④ 운영기관	한국교육개발원
⑤ 상위기관	교육부
⑥ 제공자료	유초중등교육통계, 대학통계, 평생통계, 국제통계, 주요지표, 시계열통계, 이슈통계, 시도별 간추린 통계, 교육통계연보, 간추린 교육통계, 취업통계연보 등 제공
⑦ 제공자료형식	정형 데이터(숫자), 비정형 데이터(월보, 연보)
⑧ 소개	- 교육통계는 교육기본통계조사, 고등교육기관 졸업자 건강보험 DB연계 취업통계조사, 평생교육통계조사, 평생학습개인실태조사 등 4종의 국가승인통계 조사를 의미 - 각각의 조사를 통해 산출된 통계 결과들은 교육정책 수립과 교육관련 연구에 기초가 되는 데이터로 활용되고 있으며, 국민의 알 권리를 충족을 위해 다양한 방법으로 자료가 제공

(3) 국토교통통계누리



항목	내용
① 명칭	국토교통 통계누리
② 유형분류	화면에서 데이터 확인이 가능하며, 다운로드 가능
③ 주소	https://stat.molit.go.kr/portal/event/eventIntroQuiz.do
④ 운영기관	국토교통부
⑤ 상위기관	없음
⑥ 제공자료	교통문화실태조사, 국가교통조사, 대중교통현황조사, 건축물통계, 건축허가 및 착공통계, 도시계획현황, 통계연보 등 제공
⑦ 제공자료형식	정형 데이터(숫자), 비정형 데이터(연보)
⑧ 소개	- 국토교통통계누리는 국토교통분야 주요 통계를 한 곳에 모아 이용자가 원하는 통계를 쉽게 찾을 수 있도록 국토교통부에서 제공하고 있는 통계 서비스 - 국토, 주택, 토지, 수자원, 도로, 교통, 항공, 철도 등 국민생활과 밀접한 분야의 통계자료를 승인통계, e-나라지표, 행정자료로 구분하여 제공

(4) 온나라부동산정보 통합포털



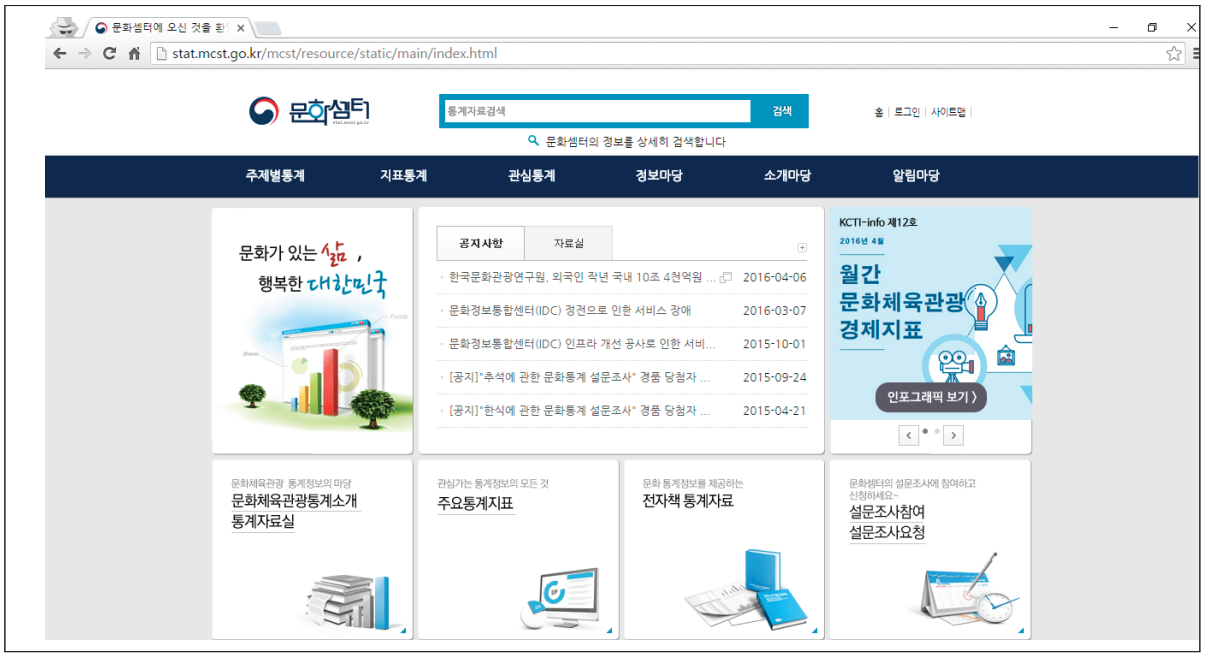
항목	내용
① 명칭	온나라부동산정보 통합포털
② 유형분류	자료의 검색, 다운로드, 분석까지 가능한 정보 포털 사이트
③ 주소	http://www.onnara.go.kr/
④ 운영기관	한국토지주택공사
⑤ 상위기관	국토교통부
⑥ 제공자료	토지 거래현황, 순수토지 거래현황, 건축물 거래현황, 주택 거래현황, 아파트 거래현황, 지가변동률, 토지현황, 도시현황, 주택보급률, 주택건설실적, 미분양 주택현황 등 자료제공
⑦ 제공자료형식	정형 데이터(숫자), 비정형 데이터(월보)
⑧ 소개	- 전국의 토지, 주택 등 부동산가격(실거래가/공시지가/주택공시가격), 분양정보, 토지이용규제정보를 필지별로 제공하는 부동산 정보 포털 서비스 - 2007년 온나라포털 오픈하여 GIS기반의 부동산 정보 통합 제공, 2011년 온나라포털 3.0 구축 - 부동산 통계 메뉴에서 통계현황을 확인 할 수 있음 - 통계자료들을 비교하는 서비스 제공 - 부동산통계 월보 제공

(5) 기상자료개방포털



항목	내용
① 명칭	기상자료개방포털
② 유형분류	자료의 검색, 다운로드 가능한 정보포털사이트
③ 주소	https://data.kma.go.kr/cmmn/main.do
④ 운영기관	기상청
⑤ 상위기관	환경부
⑥ 제공자료	종관기상관측, 방재기상관측, 농업기상관측, 북한기상관측, 연직바람관측, 해양기상부이, 등표기상관측, 파고부이, 부유분진측정, 지역예보모델, 국지예보모델, 초단기예보모델, 지역파랑모델 등 제공
⑦ 제공자료형식	정형 데이터(숫자), 비정형 데이터(월보, 연보)
⑧ 소개	- 기상청은 국가기후자료를 보다 정확, 신속, 편리하게 사용할 수 있도록 기후자료 관리시스템을 구축하여 다양한 정보를 제공하도록 함 - 통계자료 메뉴는 관측분야별통계, 요소별 분석, 다중지점 통계, 연속발생일수 분석, 위험기상추적, 한반도 기후통계로 나뉘어져 있음 - 응용정보메뉴는 응용기후, 지역맞춤형기상정보, 기후지수 통계, 기상자원지도, 사용자 입력자료 분석으로 나뉘어져 있음 - 기후특성을 분석한 간행물을 연보와 월보로 제공

(6) 문화통계포털(문화셈터)



항목	내용
① 명칭	문화셈터
② 유형분류	자료의 검색, 다운로드 가능한 정보포털사이트
③ 주소	http://stat.mcst.go.kr
④ 운영기관	한국문화관광연구원
⑤ 상위기관	문화체육관광부
⑥ 제공자료	국민여가활동조사, 국민여가생활 실태, 공연예술실태조사, 문화예술인실태조사, 전국도서관통계조사, 콘텐츠산업통계, 광고산업통계조사, 국민독서실태조사, 정기간행물등록현황 등 제공
⑦ 제공자료형식	정형 데이터(숫자), 비정형 데이터(연보)
⑧ 소개	- 주제별통계, 지표통계, 관심통계, 정보마당, 소개마당, 알림마당을 제공 - 주제별통계에서는 문화예술, 문화산업, 관광, 체육으로 나누어 제공 - 지표통계에서는 문화지표통계, 문화동향통계, e-나라지표로 나누어 제공 - 관심통계에서는 문화통계, 주요통계지표로 나누어 제공 - 정보마당에서는 자료실, 시각화 서비스, 지도서비스, 해외통계 제공

(7) ITSTAT



항목	내용
① 명칭	ITSTAT
② 유형분류	화면에서 데이터 확인이 가능하며, 다운로드 가능
③ 주소	http://www.itstat.go.kr
④ 운영기관	정보통신정책연구원
⑤ 상위기관	미래창조과학부
⑥ 제공자료	ICT산업총괄, 대표사업체수, 부문별 사업체수, ICT산업 총 종사자수, 유선통신서비스, 태블릿 PC 가입자 등 제공
⑦ 제공자료형식	정형 데이터(숫자), 비정형 데이터(보고서, 월보)
⑧ 소개	- 거시경제, ICT산업, SW산업, 통신/방송, 정보화/ICT활용, R&D, 우편/금융, 국제ICT지수로 나누어 자료 제공 - 시계열 통계데이터, 통계간행물, 세계 속 우리ICT, 수출 상위10대 품목, ICT 생산, 우리경제 속 ICT, 주요 거시지표, 서비스 가입자 현황 등을 제공

(8) 보건복지부 통계포털



항목	내용
① 명칭	보건복지부 통계포털
② 유형분류	연감자료, 보고서 등 비정형자료 제공, 데이터 부분은 정형자료 제공
③ 주소	http://stat.mohw.go.kr/
④ 운영기관	보건복지부
⑤ 상위기관	없음
⑥ 제공자료	HIV/AIDS신고현황, 가계금융조사, 가계동향조사, 병원경영실태조사, 보육실태조사, 보건소 및 보건지소운영현황, 아동종합실태조사, 학대피해아동보호현황 등 제공
⑦ 제공자료형식	정형 데이터(숫자), 비정형 데이터(연보, 월보, 백서)
⑧ 소개	- 국가통계통합DB 관리시스템에서 작성하여, 보건복지부 통계포털 및 국가통계포털에서 통계 데이터베이스 서비스 제공 - 주요지표 100선을 제공함 - 보건복지 통계는 보건복지 주제별, 기념일별, 색인별로 통계를 찾을 수 있도록 편리하게 메뉴를 구성해 놓음 - 국제통계는 주제별, 국제기구별로 나누어 제공하며, 각 통계표를 KOSIS와 연동하여 제공 - e-나라지표는 국기지표체계와 연동하여 제공

(9) 국가에너지통계 종합정보시스템(KESIS)



항목	내용
① 명칭	국가에너지통계 종합정보시스템
② 유형분류	화면에서 데이터 확인이 가능
③ 주소	http://www.kesis.net/
④ 운영기관	에너지경제연구원
⑤ 상위기관	산업통상자원부
⑥ 제공자료	주요에너지지표, 에너지 수출입, 지역별·국가별요약통계, 에너지 순위통계, 에너지 자원통계, 에너지수요전망, 에너지통계월보, 에너지통계연보, 지역에너지통계연보, 에너지총조사보고서, Energy info. Korea, Country Profile 등 제공
⑦ 제공자료형식	정형 데이터(숫자), 비정형 데이터(연보, 보고서)
⑧ 소개	- 국가에너지 통계종합정보시스템(KESIS)은 국가에너지정보통계를 종합적으로 관리, 지원하고 에너지정보통계를 통합 포털사이트를 통하여 서비스하는 시스템 - 이용자에게 편리성 제공 및 만족도를 증대시키고 정책입안·결정 관계자 및 대국민 서비스 기반을 마련하는 등 에너지 정보·통계의 국가 지식 기반화를 목적으로 구축

(10) 산업연구원(KIET) 서비스통계포털



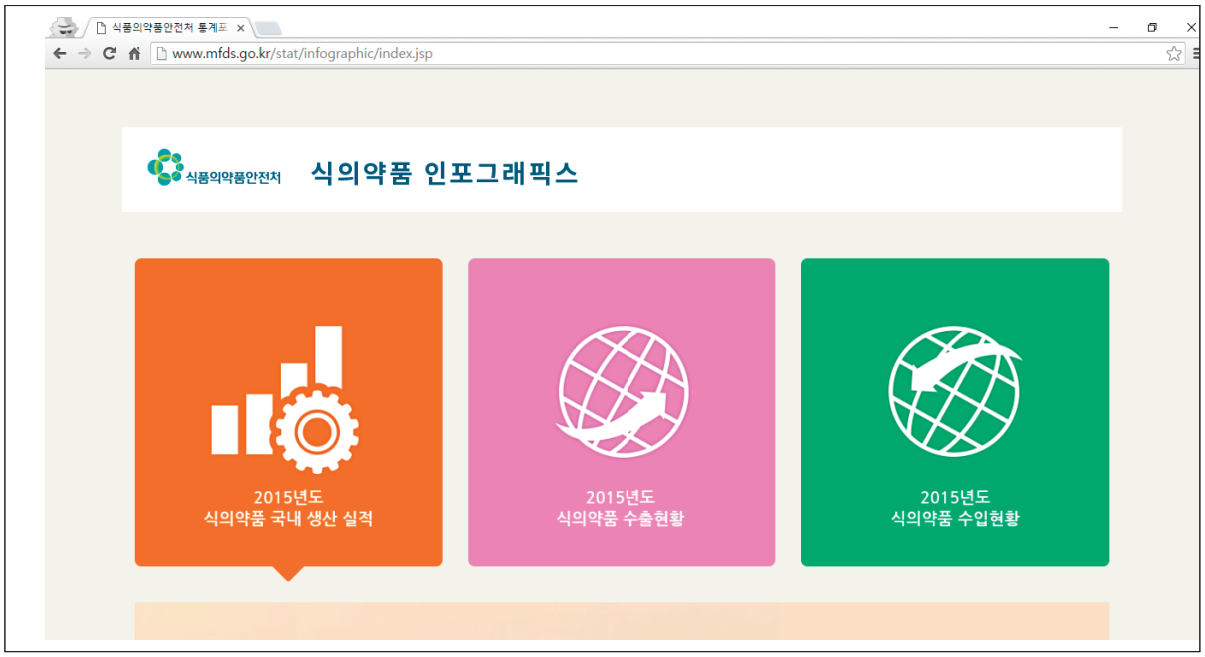
항목	내용
① 명칭	KIET 서비스통계포털
② 유형분류	화면에서 데이터 확인이 가능하며, 다운로드 가능
③ 주소	http://service.kiet.re.kr/html/
④ 운영기관	산업연구원
⑤ 상위기관	산업통상자원부
⑥ 제공자료	매출액, 매출액 비중, 부가가치, 부가가치 비중, 서비스업생산지수, 설비투자, 해외직접투자, 외국인직접투자, 사업체수, 종사자 비중, 산업별 GDP, KEIT 분석자료 등 제공
⑦ 제공자료형식	정형 데이터(숫자), 비정형 데이터(보고서)
⑧ 소개	- 산업연구원 서비스통계포털은 주제별통계에서는 생산, 투자, 고용, 무역, 자산구조, R&D/기술혁신 등에 대한 통계를 서비스산업 업종별로 제공 - 기관별 통계에서는 서비스산업 통계를 통계청, 한국은행, 수출입은행 등 통계 생산기관별로 분류하여 제공 - 글로벌 통계에서는 GDP 대비 서비스산업 비중, 서비스산업의 노동생산성 등의 지표를 주요 국가별로 제공한다. 산업별 분석에서는 산업연구원과 국내/외 연구소 등에서 생산된 서비스산업 관련 자료를 요약, 정리, 분류하여 제공

(11) 지역발전포털(REDIS)



항목	내용
① 명칭	지역발전포털(REDIS)
② 유형분류	화면에서 데이터 확인이 가능
③ 주소	http://www.redis.go.kr/route/main_home.do
④ 운영기관	한국산업기술평가관리원
⑤ 상위기관	산업통상자원부
⑥ 제공자료	지역기반통계, 지역산업통계, 지역경제통계, 지역생활통계, 지역역량통계, 지역투자통계 등 제공
⑦ 제공자료형식	정형 데이터(숫자), 비정형 데이터(보고서)
⑧ 소개	- 지역발전포털은 지역발전사업평가를 효율적으로 수행하기 위하여 부처 및 지자체의 지역발전사업에 대한 정보시스템을 구축·운영 - 지역행복생활권 정보 확충 및 강화, 지자체 정책 역량강화, 주민참여 소통증진, 지역 간 협력과 상생 지원 강화를 목표로 함 - 맞춤형 지역정보 제공 - 통계보고서로 분기별 지역경제동향과 연차별통계보고서 제공 - 지역행복생활권통계는 지도에 데이터를 표시하여 보여줌

(12) 식품의약품안전처 통계자료



항목	내용
① 명칭	식품의약품안전처 통계자료
② 유형분류	화면에서 데이터 확인이 가능하며, 다운로드 가능
③ 주소	http://www.mfds.go.kr/stat/infographic/index.jsp
④ 운영기관	식품의약품안전처
⑤ 상위기관	국무총리
⑥ 제공자료	식의약품국내 생산 실적, 식의약품 수출현황, 식의약품 수입현황
⑦ 제공자료형식	정형 데이터(숫자)
⑧ 소개	- 국내 생산실적, 수출현황, 수입현황 등 데이터를 그래프와 지도에 표시하여 제공 - 2013년, 2014년 데이터 제공

(13) 중소기업조사통계시스템



항목	내용
① 명칭	중소기업 조사통계시스템
② 유형분류	화면에서 데이터 확인이 가능하며, 다운로드 가능
③ 주소	http://stat2.smba.go.kr/index.jsp
④ 운영기관	중소기업청
⑤ 상위기관	산업통상자원부
⑥ 제공자료	기본통계, 종합실태, 인력실태, 경영지표, 기술실태, 임금실태, 경기전망지수, 생산 및 가동률, 설비투자, 수출부분, 창업 및 부도 등 제공
⑦ 제공자료형식	정형 데이터(숫자), 비정형 데이터(월보, 연보, 보고서)
⑧ 소개	- 중소기업청, 중소기업중앙회를 비롯한 여러 기관에서 작성한 중소기업관련 조사·통계를 DB화하여 제공하는 시스템으로 다양한 방법에 의한 검색과 다운로드가 가능 - 중소기업 관련 조사의 조사개요 및 보고서 원문을 다양한 형식(hwp, pdf 등)으로 볼 수 있으며, 분야별, 기관별, 주기별로 조회 가능

(14) 수산정보포털



항목	내용
① 명칭	수산정보포털
② 유형분류	자료의 검색, 다운로드 가능한 정보포털사이트
③ 주소	http://www.fips.go.kr/
④ 운영기관	해양수산부
⑤ 상위기관	없음
⑥ 제공자료	연도별 등록어선통계, 시도별 톤급별 등록어선통계, 연도별 선령별 등록어선통계, 업종별 선종별 등록어선통계, 원양어업통계, 수산물 수출입 통계 등 제공
⑦ 제공자료형식	정형 데이터(숫자)
⑧ 소개	- 어선통계, 어업생산통계, 원양어업통계, 수산물 수출입 통계, 유통정보, 수산물 가공, 자율관리어업, 인공어초, EEZ 조업정보로 나누어 자료 제공 - 수산물가격정보로는 산지위판정보, 소비자 가격정보를 제공 - 연근해어업에서는 어선조업정보, EEZ 조업정보를 제공 - 양식어업에서는 해수면어업 총괄현황, 내수면 총괄현황, 낚시산업, 어류양식통계 제공 - 원양 • 무역에서는 원양어업통계, 수산물 수출입 통계 제공 - 유통정보에서는 유통정보, 산지위판정보, 소비자가격정보, 유통단계별 통계정보, 수산물재고 정보 제공 - 수산통계에는 어업생산통계, 계통판매통계, 수산물가공 제공

(15) 환경통계포털



항목	내용
① 명칭	환경통계포털
② 유형분류	자료의 검색, 다운로드 가능한 정보포털사이트
③ 주소	http://stat.me.go.kr
④ 운영기관	환경부
⑤ 상위기관	없음
⑥ 제공자료	공공기관녹색제품구매실적, 소음도현황, 실내공기질관리, 환경산업통계조사, 국립공원기본통계, 멸종위기 야생동식물 현황, 생물표본 확인수, 수질오염 실태 보고 등 제공
⑦ 제공자료형식	정형 데이터(숫자), 비정형 데이터(연감)
⑧ 소개	- 환경통계 고객만족도 조사결과 별도의 통계전용 홈페이지 구축 등 통계정보를 상호 공유하고 편리하게 활용할 수 있는 통계정보시스템 구축 - 정책과 제도의 기반이 되는 통계지표를 구비함으로써 통계와 정책 간 연계를 강화, 정책 및 제도의 효율성 제공 - 통계조회에서는 분야별통계, 환경통계연감, 명칭별 통계, 지역별통계, 타기관통계 제공 - 테마별 통계에서는 주요환경지표, e-나라지표, 그래프로 보는 주요통계, OECD 주요환경지표, Open API, 내 지역 환경통계 제공 - 통계연구/간행물에서는 환경통계 Brief Guide, 생활 속의 통계이야기, 환경통계상식, 통계간행물 제공

3.3 공공기관 및 중앙행정부처 산하기관

공공기관이란 정부의 투자·출자 또는 정부의 재정지원 등으로 설립·운영되는 기관으로 2016년 현재 기준 공기업 30개, 준정부기관 90개 기타공공기관 203개로 총 323개가 있다(ALIO). 정부산하기관은 정부로부터 출연금·보조금 등을 받는 기관 또는 단체, 법령에 의하여 직접 정부로부터 업무를 위탁 받거나 독점적 사업권을 부여 받은 기관 또는 단체로서 이 법을 적용 받는 기관으로 정의되며, 2015년 행정자치연감에 의하면 시험연구기관 61곳, 교육훈련기관 36곳, 문화기관 37곳, 의료기관 10곳, 기타행정기관 99곳으로 243곳이 있다. 이들의 정리는 책임기관(중앙행정기관)의 분류에 따라 진행하였다. 공공기간과 산하기간은 그 수가 너무 많고, 유형분류에서 1, 2에 해당하는 경우가 거의 없어 연계구조를 나타내는 cytoscape와 리스트만 본문에 제시하였다.

(1) 위원회, 처 통합

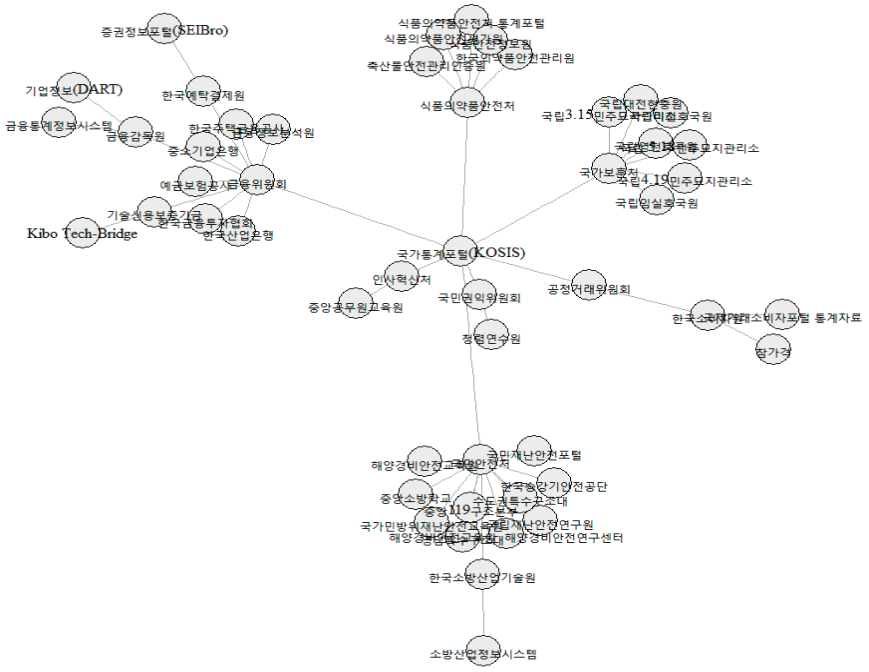


그림 3.14 위원회, 처 통합 cytoscape

표 3.5 위원회, 처 소속 공공기관과 산하기관

공공기관	산하기관
Kibo Tech-Bridge	국립 5.18민주묘지관리소
국제거래소비자포털 통계자료	국립3.15민주묘지관리소
기술신용보증기금	국립4.19민주묘지관리소
식품안전정보원	국립대전현충원
예금보험공사	국립영천호국원
중소기업은행	국립이천호국원
증권정보포털(SEIBro)	국립임실호국원
참가격	국립재난안전연구원
축산물안전관리인증원	금융정보분석원
한국산업은행	수도권특수구조대
한국소방산업기술원	식품의약품안전평가원
한국소비자원	영남특수구조대
한국승강기안전공단	중앙119구조본부
한국예탁결제원	중앙공무원교육원
한국의약품안전관리원	중앙소방학교
한국주택금융공사	청렴연수원
국가민방위재난안전교육원	해양경비안전교육원
	해양경비안전연구센터

(2) 고용노동부, 교육부

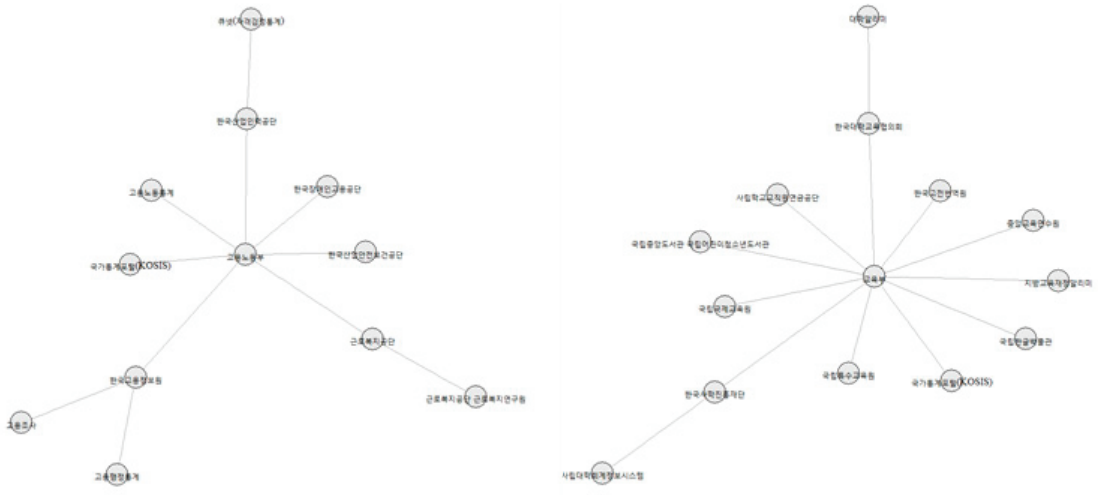


그림 3.15 고용노동부(좌)와 교육부(우) cytoscape

표 3.6 고용노동부와 교육부 소속 공공기관과 산하기관

고용노동부		교육부	
고용조사	공공기관	사립대학회계정보시스템	공공기관
고용행정통계	공공기관	사립학교교직원연금공단	공공기관
근로복지공단	공공기관	한국고전번역원	공공기관
근로복지공단 근로복지연구원	공공기관	한국사학진흥재단	공공기관
큐넷(자격검정통계)	공공기관	국립국제교육원	산하조직
한국고용정보원	공공기관	국립중앙도서관 국립어린이청소년도서관	산하조직
한국산업안전보건공단	공공기관	국립특수교육원	산하조직
한국산업인력공단	공공기관	국립한글박물관	산하조직
한국장애인고용공단	공공기관	중앙교육연수원	산하조직

3. 국무조정실, 국방부(병무청)

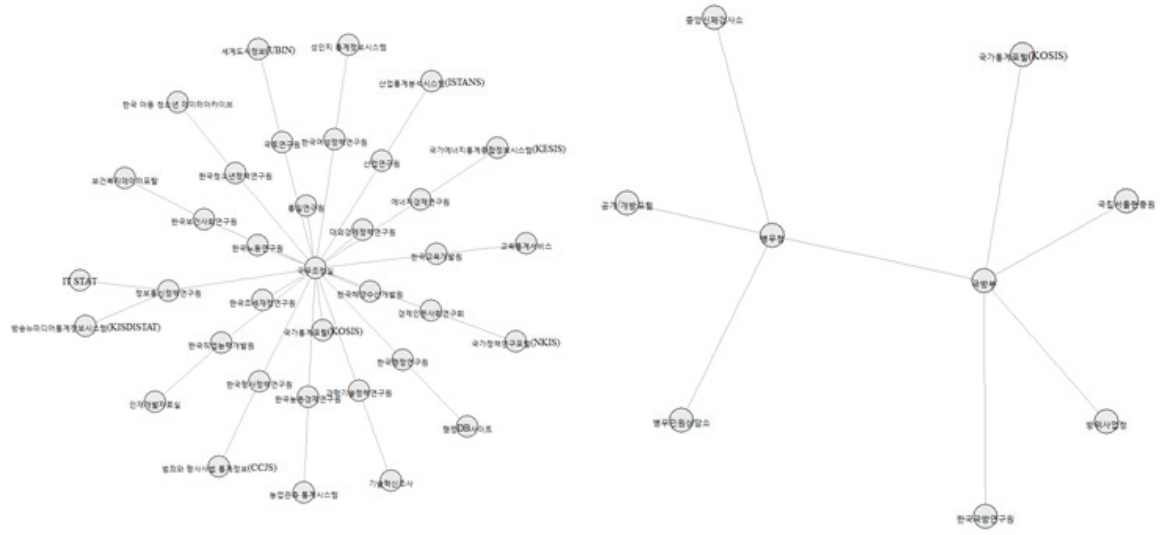


그림 3.16 국무조정실(좌)와 국방부(우)의 cttoscape

표 3.7 국무조정실과 국방부 소속 공공기관과 산하기관

국무조정실 공공기관	국무조정실 공공기관	
한국형사정책연구원	행정DB사이트	
한국행정연구원	한국 아동 청소년 데이터아카이브	
한국해양수산개발원	인재개발자료실	
한국청소년정책연구원	성인지 통계정보시스템	
한국직업능력개발원	보건복지데이터포탈	
한국조세재정연구원	농업관측 통계시스템	
한국여성정책연구원	교육통계서비스	
한국보건사회연구원	방송뉴미디어통계정보시스템(KISDISTAT)	
한국농촌경제연구원	IT STAT	
한국노동연구원	국가에너지통계종합정보시스템(KESIS)	
한국교육개발원	산업통계분석시스템(ISTANS)	
통일연구원	세계도시정보(UBIN)	
정보통신정책연구원	기술혁신조사	
에너지경제연구원	국가정책연구포털(NKIS)	
산업연구원	국방부	
대외경제정책연구원		
국토연구원	한국국방연구원	공공기관
과학기술정책연구원	국립서울현충원	산하기관
경제인문사회연구회	병무민원상담소	산하기관
범죄와 형사사법 통계정보(CCJS)	중앙신체검사소	산하기관

4. 국토교통부, 기획재정부(관세청, 통계청, 국세청, 조달청)

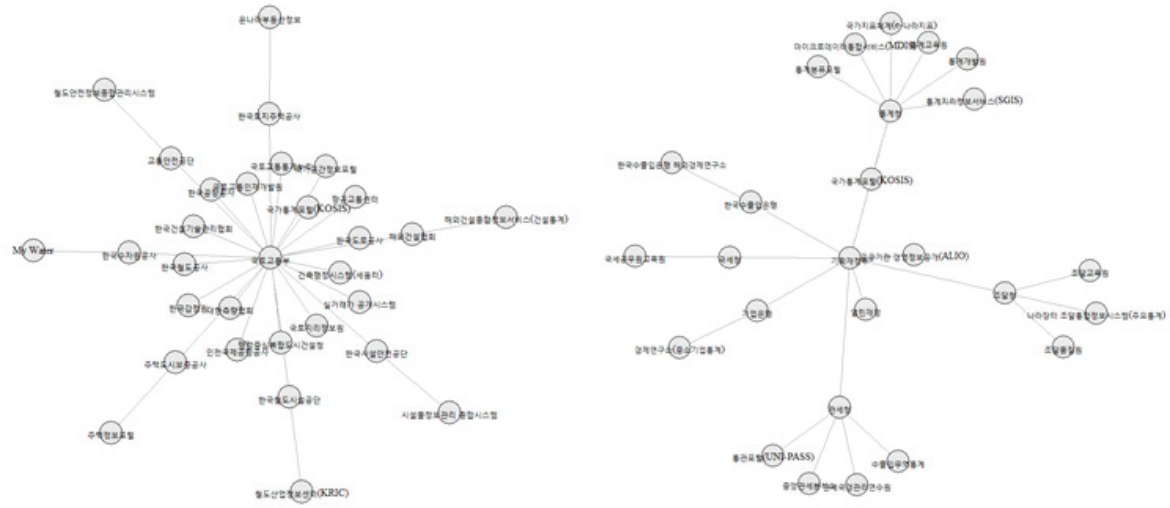


그림 3.17 국토교통부(좌)와 기획재정부 cytoscape

표 3.8 국토교통부와 기획재정부 소속 공공기관과 산하기관

국통교통부		국토교통부	
My Water	공공기관	한국토지주택공사	공공기관
교통안전공단	공공기관	국토교통인재개발원	산하조직
시설물정보관리 종합시스템	공공기관	국토지리정보원	산하조직
온나라부동산정보	공공기관	항공교통센터	산하조직
인천국제공항공사	공공기관	기획재정부	
주택도시보증공사	공공기관		
주택정보포털	공공기관	한국수출입은행	공공기관
철도산업정보센터(KRIC)	공공기관	한국수출입은행 해외경제연구소	공공기관
철도안전정보종합관리시스템	공공기관	관세국경관리연수원	산하조직
한국공항공사	공공기관	국세공무원교육원	산하조직
한국도로공사	공공기관	조달교육원	산하조직
한국수자원공사	공공기관	조달품질원	산하조직
한국시설안전공단	공공기관	중앙관세분석소	산하조직
한국철도공사	공공기관	통계개발원	산하조직
한국철도시설공단	공공기관	통계교육원	산하조직

5. 농림축산식품부(농촌진흥청, 산림청), 문화체육관광부(문화재청)

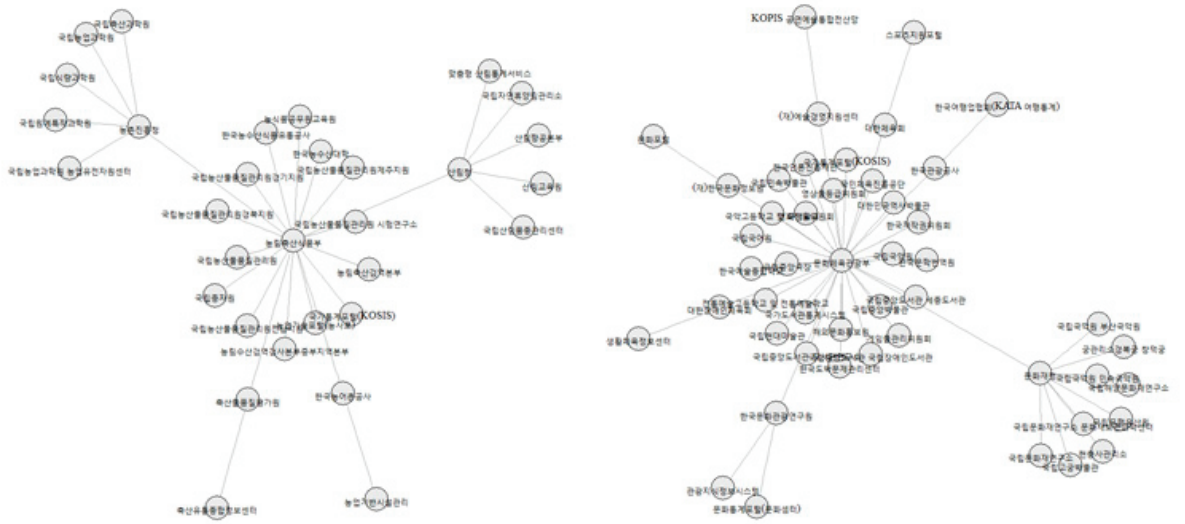


그림 3.18 농림축산식품부(좌)와 문화체육관광부(우)의 cytoscape

표 3.9 농림축산식품부와 문화체육관광부 소속 공공기관과 산하기관

농림축산식품부		문화체육관광부	
농업기반시설관리	공공기관	문화포털	공공기관
축산물품질평가원	공공기관	생활체육정보센터	공공기관
축산유통종합정보센터	공공기관	스포츠지원포털	공공기관
한국농수산식품유통공사	공공기관	영상물등급위원회	공공기관
한국농어촌공사	공공기관	영화진흥위원회	공공기관
국립농산물품질관리원	산하조직	한국관광공사	공공기관
국립농산물품질관리원 시험연구소	산하조직	한국도박문제관리센터	공공기관
국립농산물품질관리원 경기지원	산하조직	한국문학번역원	공공기관
국립농산물품질관리원 경북지원	산하조직	한국문화관광연구원	공공기관
국립농산물품질관리원 전남지원	산하조직	한국언론진흥재단	공공기관
국립농산물품질관리원 제주지원	산하조직	한국여행업협회 (KATA 여행통계)	공공기관
국립농업과학원	산하조직	한국저작권위원회	공공기관
국립농업과학원 농업유전자원센터	산하조직	국립고궁박물관	산하조직
국립산림품종관리센터	산하조직	국립국악원	산하조직
국립식량과학원	산하조직	국립국악원 민속국악원	산하조직
국립원예특작과학원	산하조직	국립국악원 부산국악원	산하조직
국립자연휴양림관리소	산하조직	국립국어원	산하조직
국립종자원	산하조직	국립무형유산원	산하조직
국립축산과학원	산하조직	국립문화재연구소	산하조직
농림수산검역검사본부 중부지역본부	산하조직	국립문화재연구소 문화재보존과학센터	산하조직
농림축산검역본부	산하조직	국립민속박물관	산하조직
농식품공무원교육원	산하조직	국립중앙극장	산하조직
산림교육원	산하조직	국립중앙도서관 국립장애인도서관	산하조직
산림항공본부	산하조직	국립중앙도서관 도서관연구소	산하조직
한국농수산대학	산하조직	국립중앙도서관 세종도서관	산하조직
문화체육관광부		국립중앙박물관	산하조직
(재)예술경영지원센터	공공기관	국립해양문화재연구소	산하조직
(재)한국문화정보원	공공기관	국립현대미술관	산하조직
KOPIS 공연예술통합전산망	공공기관	국악고등학교 및 국악학교	산하조직
게임물관리위원회	공공기관	궁관리소경복궁 창덕궁	산하조직
관광지식정보시스템	공공기관	대한민국역사박물관	산하조직
국민체육진흥공단	공공기관	전통예술고등학교 및 전통예술학교	산하조직
대한장애인체육회	공공기관	한국예술종합학교	산하조직
대한체육회	공공기관	해외문화홍보원	산하조직
문화통계포털(문화셈터)	공공기관	현충사관리소	산하조직

6. 미래창조과학부, 법무부

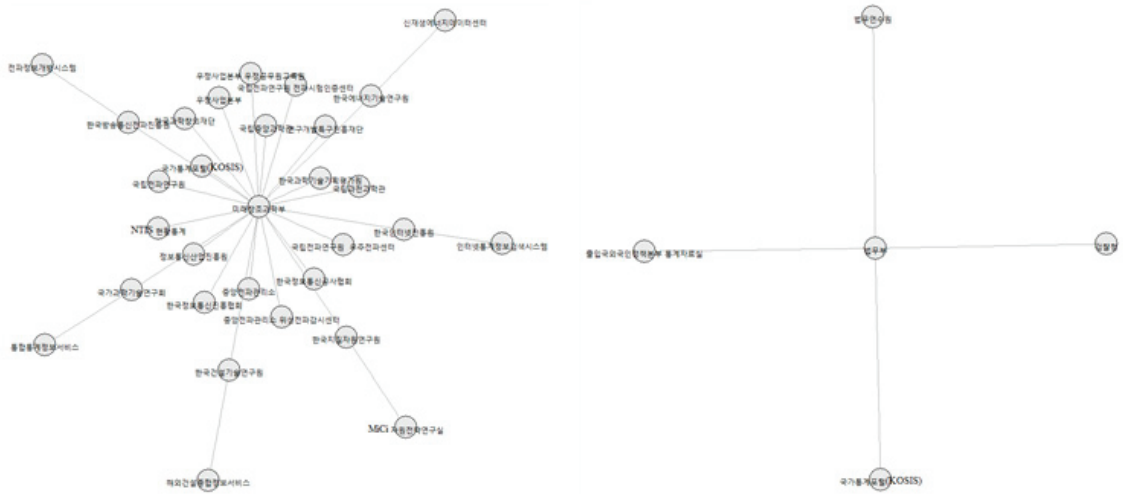


그림 3.19 미래창조과학부(좌)와 법무부(우)의 cytoscape

표 3.10 미래창조과학부와 법무부 소속 공공기관과 산하기관

미래창조과학부		미래창조과학부	
MiCi 자원전략연구실	공공기관	한국지질자원연구원	공공기관
국가과학기술연구회	공공기관	해외건설종합정보서비스	공공기관
신재생에너지데이터센터	공공기관	국립과천과학관	산하기관
연구개발특구진흥재단	공공기관	국립전파연구원	산하기관
인터넷통계정보검색시스템	공공기관	국립전파연구원 우주전파센터	산하기관
전파정보개방시스템	공공기관	국립전파연구원 전파시험인증센터	산하기관
통합통계정보서비스	공공기관	국립중앙과학관	산하기관
한국건설기술연구원	공공기관	우정사업본부	산하기관
한국과학기술기획평가원	공공기관	우정사업본부 우정공무원교육원	산하기관
한국과학창의재단	공공기관	중앙전파관리소	산하기관
한국방송통신전파진흥원	공공기관	중앙전파관리소 위성전파감시센터	산하기관
한국에너지기술연구원	공공기관	법무부	
한국인터넷진흥원	공공기관	법무연수원	산하기관

7. 보건복지부, 산업통상자원부(중소기업청, 특허청)

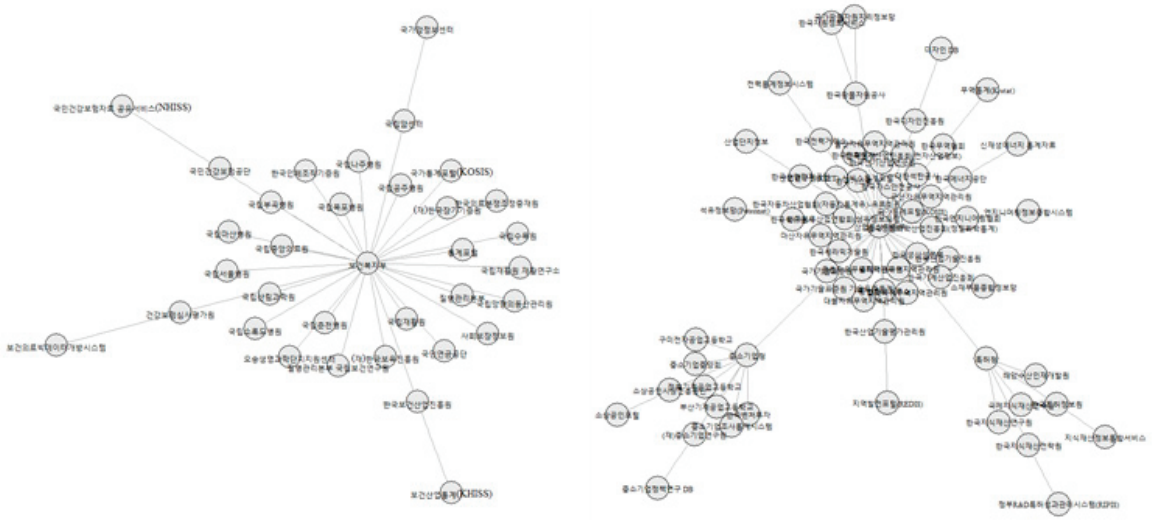


그림 20 보건복지부(좌)와 산업통상자원부(우)의 cytoscape

표 11 보건복지부와 산업통상자원부 소속 공공기관과 산하기관

보건복지부		산업통상자원부	
(재)한국보육진흥원	공공기관	신재생에너지 통계자료	공공기관
국가암정보센터	공공기관	전력통계정보시스템	공공기관
국립암센터	공공기관	정부R&D특허성과관리 시스템(RIPIS)	공공기관
국민건강보험공단	공공기관	중소기업정책연구 DB	공공기관
국민건강보험자료 공유서비스(NHISS)	공공기관	지식재산정보통합서비스	공공기관
국민연금공단	공공기관	지역발전포털(REDIS)	공공기관
보건산업통계(KHISS)	공공기관	한국가스공사	공공기관
보건의료빅데이터개방시스템	공공기관	한국가스안전공사	공공기관
사회보장정보원	공공기관	한국광물자원공사	공공기관
한국보건산업진흥원	공공기관	한국디자인진흥원	공공기관
한국의료분쟁조정중재원	공공기관	한국벤처투자	공공기관
한국인체조직기증원	공공기관	한국산업기술진흥원	공공기관
국립공주병원	산하조직	한국산업기술평가관리원	공공기관

국립나주병원	산하조직	한국석유공사	공공기관
국립마산병원	산하조직	한국세라믹기술원	공공기관
국립망향의동산관리원	산하조직	한국에너지공단	공공기관
국립목포병원	산하조직	한국자원정보서비스	공공기관
국립부곡병원	산하조직	한국전력공사	공공기관
국립산림과학원	산하조직	한국지식재산연구원	공공기관
국립서울병원	산하조직	한국지식재산전략원	공공기관
국립소록도병원	산하조직	한국특허정보원	공공기관
국립수목원	산하조직	광업등록사무소	산하조직
국립재활원	산하조직	구미전자공업고등학교	산하조직
국립재활원 재활연구소	산하조직	국가기술표준원	산하조직
국립춘천병원	산하조직	국가기술표준원 기술표준통계	산하조직
오송생명과학단지지원센터	산하조직	국제지식재산연수원	산하조직
질병관리본부	산하조직	군산자유무역지역관리원	산하조직
질병관리본부 국립보건연구원	산하조직	김제자유무역지역관리원	산하조직
산업통상자원부		대불자유무역지역관리원	산하조직
(재)중소기업연구원	공공기관	동해자유무역지역관리원	산하조직
국가광물자원지리정보망	공공기관	마산자유무역지역관리원	산하조직
대한석탄공사	공공기관	부산기계공업고등학교	산하조직
디자인 DB	공공기관	울산자유무역지역관리원	산하조직
산업단지정보	공공기관	율촌자유무역지역관리원	산하조직
석유정보망(Petronet)	공공기관	전북기계공업고등학교	산하조직
소상공인시장진흥공단	공공기관	해양수산인재개발원	산하조직
소상공인포털	공공기관		

8. 여성가족부, 외교부

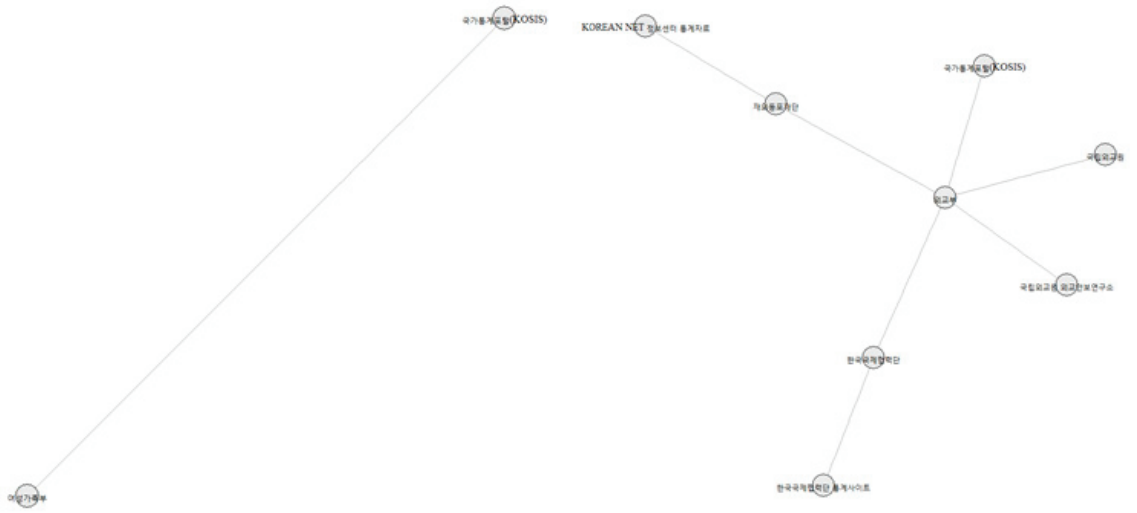


그림 3.21 여성가족부(좌)와 외교부(우)의 cytoscape

9. 통일부, 해양수산부

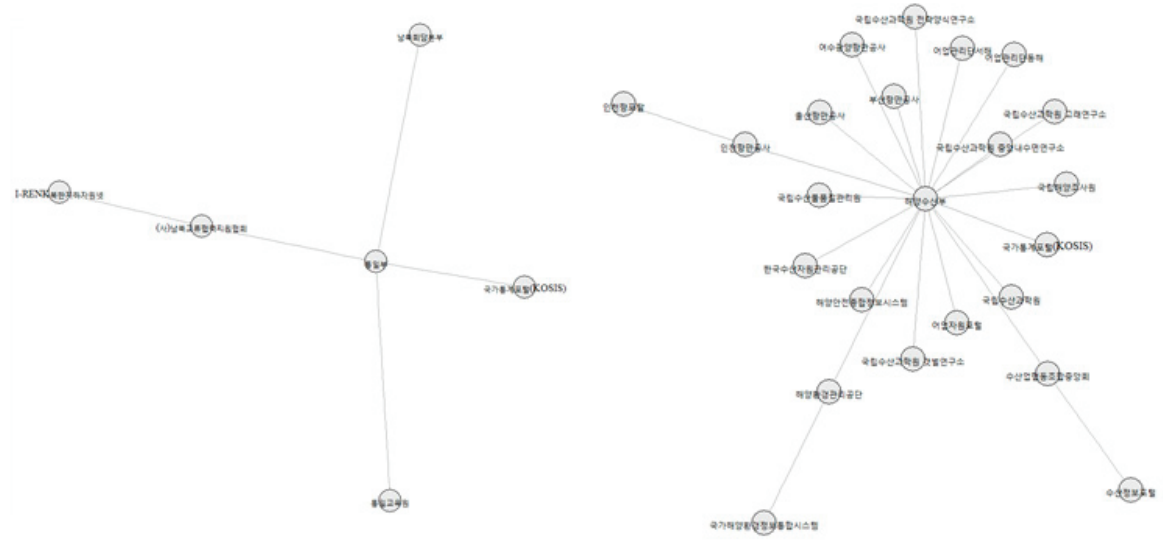


그림 22 통일부(좌)와 해양수산부(우)의 cytoscape

표 3.12 외무부, 통일부, 해양수산부와 산업통상자원부 소속 공공기관과 산하기관

외무부		해양수산부	
한국국제협력단	공공기관	울산항만공사	공공기관
재외동포재단	공공기관	인천항만공사	공공기관
한국국제협력단 통계사이트	공공기관	인천항포탈	공공기관
KOREAN NET 정보센터 통계자료	공공기관	한국수산자원관리공단	공공기관
국립외교원	산하조직	해양환경관리공단	공공기관
국립외교원 외교안보연구소	산하조직	국립수산과학원	산하조직
통일부		국립수산과학원 갯벌연구소	산하조직
(사)남북교류협력지원협회	공공기관	국립수산과학원 고래연구소	산하조직
I-RENK북한지하자원넷	공공기관	국립수산과학원 전략양식연구소	산하조직
통일교육원	산하조직	국립수산과학원 중앙내수면연구소	산하조직
남북회담본부	산하조직	국립수산물품질관리원	산하조직
해양수산부		국립해양조사원	산하조직
국가해양환경 정보통합시스템	공공기관	어업관리단동해	산하조직
부산항만공사	공공기관	어업관리단서해	산하조직
여수광양항만공사	공공기관		

⑩ 행정자치부(경찰청, 지방자치단체 제외), 환경부(기상청)

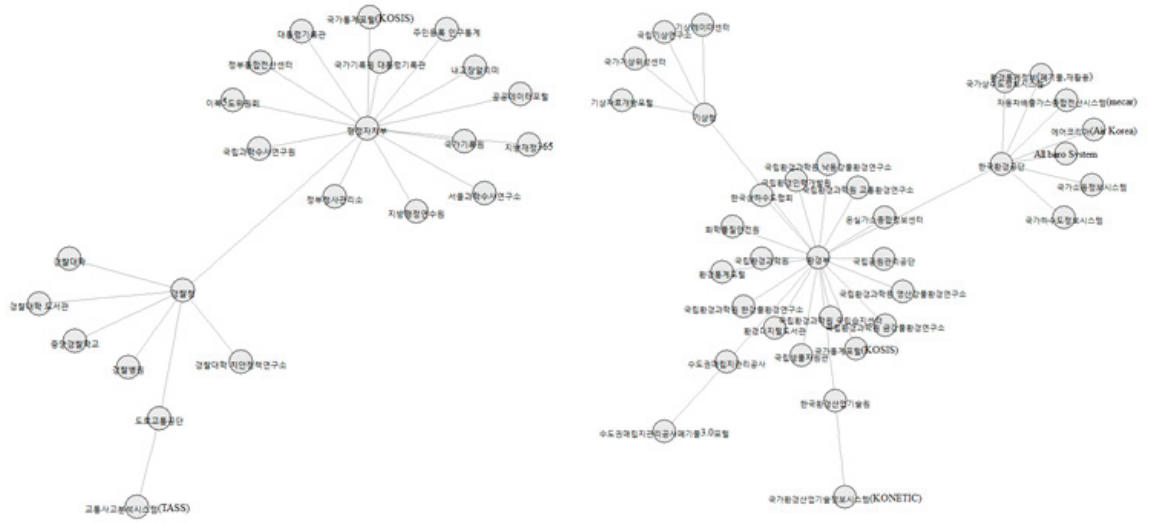


그림 3.23 행정자치부(좌)와 환경부(우)의 cytoscape

표 3.13 행정자치부와 환경부 소속 공공기관과 산하기관

행정자치부		환경부	
공공데이터포털	공공기관	수도권매립지관리공사	공공기관
교통사고분석시스템(TASS)	공공기관	수도권매립지관리공사폐기물3.0포털	공공기관
경찰대학	산하조직	에어코리아(Air Korea)	공공기관
경찰대학 도서관	산하조직	자동차배출가스종합전산시스템(mecar)	공공기관
경찰대학 치안정책연구소	산하조직	한국상하수도협회	공공기관
경찰병원	산하조직	한국환경공단	공공기관
국가기록원	산하조직	한국환경산업기술원	공공기관
국가기록원 대통령기록관	산하조직	환경통계정보(폐기물,재활용)	공공기관
국립과학수사연구원	산하조직	국가기상위성센터	산하조직
서울과학수사연구소	산하조직	국립기상연구소	산하조직
이북5도위원회	산하조직	국립생물자원관	산하조직
정부청사관리소	산하조직	국립환경과학원	산하조직
정부통합전산센터	산하조직	국립환경과학원 교통환경연구소	산하조직
중앙경찰학교	산하조직	국립환경과학원 국립습지센터	산하조직
지방행정연수원	산하조직	국립환경과학원 금강물환경연구소	산하조직
환경부		국립환경과학원 낙동강물환경연구소	산하조직
All baro System	공공기관	국립환경과학원 영산강물환경연구소	산하조직
국가상수도정보시스템	공공기관	국립환경과학원 한강물환경연구소	산하조직
국가소음정보시스템	공공기관	국립환경인력개발원	산하조직
국가하수도정보시스템	공공기관	기상레이더센터	산하조직
국가환경산업기술정보시스템(KONETIC)	공공기관	온실가스종합정보센터	산하조직
국립공원관리공단	공공기관	화학물질안전원	산하조직

3.4 민간(기업, 금융기관, 협회) 등

민간기관이나 기업에서 지속적으로 관리되는 빅데이터를 공개하는 경우는 거의 찾을 수 없었으며, 금융기관의 경우 공공기관은 아니지만, 다소 공공성을 갖고 있는 기관에서 자료를 개방하고 있었다. 먼저 민간의 경우는 아래의 세 경우를 찾을 수 있었으며, 그나마 다음소프트나 구글의 경우는 검색어의 변화에 대한 결과만을 제시해주고 있어 원자료를 이용한 분석은 불가능하다(Google Trends는 도표작성에 사용된 자료는 제공).

표 3.14 데이터 제공 민간 사이트

민간기관	사이트 이름
네이버	1. 네이버 데이터랩**
다음소프트	2. SOCIAL metrics
구글	3. Google Trends

반면에 네이버의 '데이터랩(DATA LAB)'은 다양한 민간/공공 데이터를 한눈에 볼 수 있고, 더 나아가 이를 자신이 보유한 데이터에 융합해 다방면으로 활용할 수 있도록 구성되어있다.

데이터랩은 크게 네이버 검색 데이터를 개인 이용자가 보유한 데이터와 융합분석할 수 있는 〈데이터 융합 분석〉, 위치정보를 기반으로 지역별 특정 분야의 데이터를 시각적으로 보여주는 〈지역 통계〉, 현재 인기 검색어들의 기간별 트렌드를 파악할 수 있는 〈검색어로 알아보는 대한민국〉섹션 등으로 구성되어 있으며, 공공데이터포털(data.go.kr), 서울열린데이터(data.seoul.go.kr), 통계청, 한국고용정보원, 한국 국제협력단 통계사이트에서 제공하는 데이터 목록확인 및 연결이 가능하도록 되어있다.

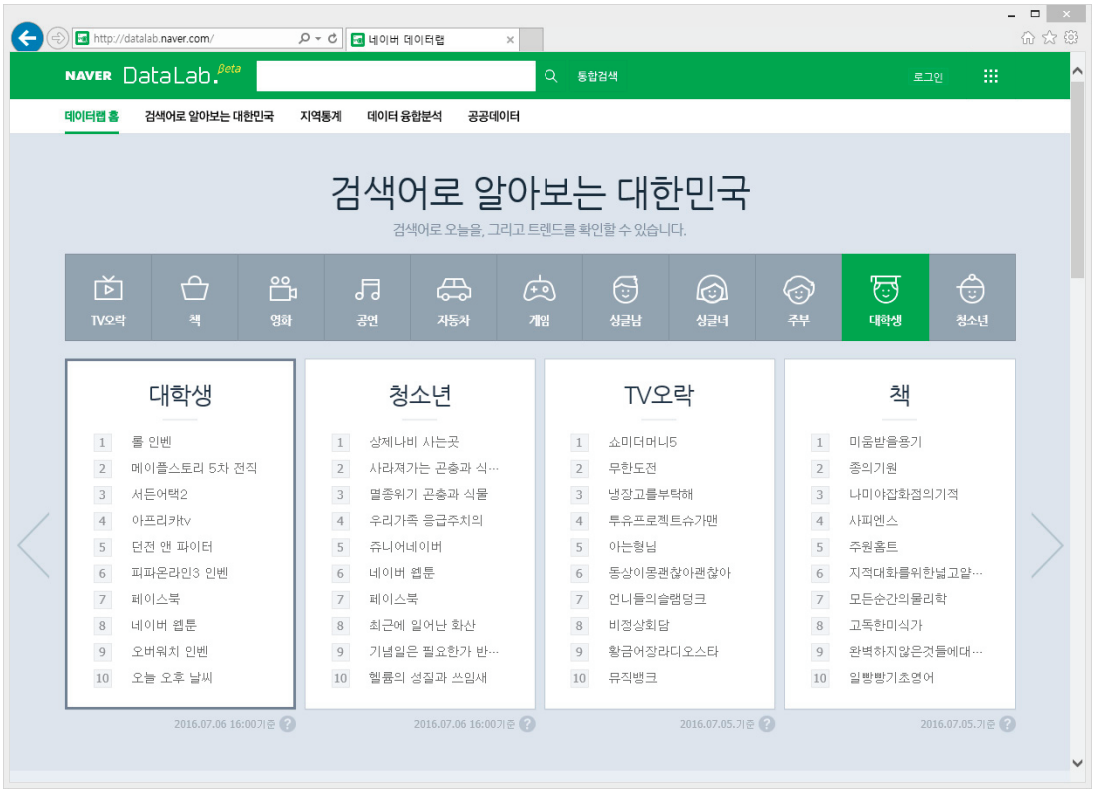


그림 3.24 네이버 데이터랩

금융기관의 경우 10곳에서 개방된 자료를 찾았으며, 모두 자료분석에 활용할 수 있을 정도의 충분한 자료를 제공하고 있었다. 금융·경제지식이 수반된다면 매우 높은 활용 가치가 있을 것으로 예상된다.

표 3.15 금융기관 데이터제공 사이트

금융기관	사이트 이름
금융감독원	4. 금융통계정보시스템**
	5. 기업정보(DART)*
보험개발원	6. 보험통계포털서비스(INSIS)**
수산업협동조합중앙회	7. 수산정보포털**
	8. 수산경제연구원 조사통계
기업은행	9. 경제연구소(중소기업통계)
한국거래소	10. Marketdata*
한국은행	11. 경제통계시스템(ECOS)**
한국예탁결재원	12. 증권정보포털(SEIBro)*

특히 이 중 '**'를 표시한 세 곳은 정보포털 형식과 아울러 다양한 자료를 제공하고 있어 자세히 살펴 볼만한 가치가 있다.

(1) 금융통계정보시스템



항목	내용
① 명칭	금융통계정보시스템
② 유형분류	화면에서 데이터 확인이 가능하며, 다운로드 가능
③ 주소	http://fisis.fss.or.kr
④ 운영기관	금융감독원
⑤ 상위기관	금융위원회
⑥ 제공자료	단수통계표조회, 복수통계표조회, 금융통계월보, 금융권역관 주요통계비교, 국내외 금융동향 및 금융지표, 금융산업과의 과거와 오늘 등 제공
⑦ 제공자료형식	정형 데이터(숫자), 비정형 데이터(월보)
⑧ 소개	- 금융정보시스템 설치 후 통계표 조회 가능 - 금융통계월보는 e-book으로 제공 - 금융권역간 주요통계비교에서는 일반현황, 재무현황, 수익성, 여신건전성, 유동성, 자산건전성 제공 - 국내외금융동향 및 금융지표에서는 국내외금융시장, 외국인투자, CP 및 전단체 제공

(2) 보험통계포털서비스



항목	내용
① 명칭	보험통계포털서비스(INsis)
② 유형분류	자료의 검색, 다운로드 가능한 정보포털사이트
③ 주소	http://www.insis.or.kr/insis/insisweb/index.jsp
④ 운영기관	보험개발원
⑤ 상위기관	없음
⑥ 제공자료	보험현황통계, 자율통계, 월간보험통계, 종목별 실적대비표, 손해보험종목별 원수보험료 등 제공
⑦ 제공자료형식	정형 데이터(숫자), 비정형 데이터(월보, 연보)
⑧ 소개	- INsis는 INsurance Statistics Information Services의 약자로, 보험개발원에서 보험통계정보를 대외에 서비스하는 보험통계 포털 시스템 - 보험개발원은 42개 보험회사로 부터 자동으로 접수한 기초통계자료를 집적하고, 통계산출을 위하여 약150억건의 자료를 가공하여 매년 2000여종의 정기통계와 수시통계를 생산 - INsis는 이러한 통계자료를 보험회사에서 분석, 출력, 다운로드 등 편리하게 활용할 수 있도록 하여 상품개발 및 언더라이팅 업무 등에 효율성을 높이도록 구축 - 보험통계연감 30년간의 자료 등 4종의 일반인을 위한 통계자료도 제공

(3) 경제통계시스템(ECOS)



항목	내용
① 명칭	한국은행 경제통계시스템(ECOS)
② 유형분류	화면에서 데이터 확인이 가능하며, 다운로드 가능
③ 주소	http://ecos.bok.or.kr/
④ 운영기관	한국은행
⑤ 상위기관	없음
⑥ 제공자료	통화 및 유동성지표, 지급결제, 물가, 기업경영분석, 대출행태서베이, 거시경제분석 지표, 100대 통계지표, 시각화통계 등 제공
⑦ 제공자료형식	정형 데이터(숫자), 비정형 데이터(월보, 연보, 보고서)
⑧ 소개	- 시각화통계에서는 세계 속의 한국경제, 국민소득, 국제수지, 물가, 기업경영분석지표 제공 - 통계월보로는 조사통계월보, 주요경제지표, 통화금융, 물가 등 제공

협회 및 기타기관에 대해서는 23개의 사이트를 찾아서 정리할 수 있었다. 많은 경우 협회에서 작성하는 연보를 PDF나 Excel로 제공하고 있었으며, 정제된 데이터를 정보포털의 형태로 제공하는 것도 있었다. 특히 한국금융투자협회의 '자본시장통계'와 '한국무역협회'의 무역통계는 살펴볼 만하다.

표 3.16 협회 및 기타기관 데이터 제공 사이트

협회 및 기타	사이트 이름
대한건설협회	13. 통계*
대한기계설비건설협회	14. 정보자료, 통계연보
대한시설물유지관리협회	15. 시설물관리유지업 통계연보
대한전문건설협회	16. 통계연보
생명보험협회	17. 생명보험통계
중소기업중앙회	18. 중소기업통계*
한국건설기술관리협회	19. 통계
한국금융투자협회	20. 자본시장통계**
한국기계산업진흥회	21. 정보센터(기계산업무역통계)
	22. 소재부품종합정보망*
한국로봇산업협회	23. 조사/통계사업
한국무역협회	24. 무역통계(K-stat)**
한국소프트웨어산업협회	25. SW기술자 평균임금
한국엔지니어링협회	26. 엔지니어링정보종합시스템
한국여성경제인협회	27. 여성기업 종합정보포털
한국전기공사협회	28. 전기공사종합정보시스템(ECIC)
한국전자정보통신산업진흥회	29. 통계자료
한국정보통신공사협회	30. 법령자료(공사업관련통계)
한국정보통신진흥협회	31. ITSTAT**
한국철강협회	32. 통계자료*
해외건설협회	33. 해외건설종합정보서비스(건설통계)*
한국방위산업진흥회	34. 방산통계
한국생산성본부	35. 생산성통계DB*

(1) 금융투자협회 자본시장통계



항목	내용
① 명칭	금융투자협회 자본시장통계
② 유형분류	화면에서 데이터 확인이 가능하며, 다운로드 가능
③ 주소	http://freesis.kofia.or.kr/?
④ 운영기관	금융투자협회
⑤ 상위기관	금융위원회
⑥ 제공자료	한 눈에 보는 자본시장 통계, 주식, 채권, 파생상품, 자산관리, 펀드, 자문일임, 신탁, 단기자금, 금융투자업자, 시장지수, 교차통계 등 제공
⑦ 제공자료형식	정형 데이터(숫자)
⑧ 소개	- 교차통계는 금융투자협회와 및 타기관에서 제공하는 통계 중 자본시장과 금융투자업 동향 분석 시 주로 활용되는 통계 항목 간 시계열 추세분석 또는 상관관계 분석 등 다양한 통계 항목 선택시 시계열 데이터, 다양한 차트 기능, 상세통계 연결 기능, 다운로드 기능, 부가정보 등을 제공 - 상세통계는 증권, 파생상품, 펀드, 채권, 프리보드, 금융투자업자 총 6개의 카테고리로 구분되어 있으며, 각 카테고리 메인화면에서는 카테고리별 주요 통계를 요약하여 제공

(2) 무역통계(K-stat)



항목	내용
① 명칭	무역통계(K-stat)
② 유형분류	화면에서 데이터 확인이 가능하며, 다운로드 가능
③ 주소	http://stat.kita.net/
④ 운영기관	한국무역협회
⑤ 상위기관	산업통상자원부
⑥ 제공자료	국내통계, 해외통계, 세계통계, 시계열통계, 매트릭스 통계, 수출의탑 시뮬레이션, 수출진단통계, 무역통계로 보는 주요국 등 제공
⑦ 제공자료형식	정형 데이터(숫자)
⑧ 소개	- 연월별, 국가별, 품목별, 대륙경제권별 수출입 통계수록국가 : 한국, 남북, 미국, 일본, 중국, EU(28개국), 홍콩, 인도, 대만, ASEAN(9개국),브라질, 러시아, 터키, 호주, 캐나다, 칠레, 페루, 멕시코, 사우디아라비아, 우즈베키스탄, 뉴질랜드 ,콜롬비아 ,남아프리카공화국 ,우크라이나 - 전 세계 국가 간 교역현황 및 관련통계 - 240여 개국을 대상으로 통화, 금융증권, 세정, 물가지수, 국내수지, 무역, 외환, 국내 총생산 등에 관한 각종통계

3.5 지방자치단체

지방자치단체로서 통계를 제공하는 사이트는 243개가 존재하지만 시도 단위의 지방자치단체에서 정보포털과 발전연구원 등을 통해 다양한 정보를 제공하고 있어 중복을 피하는 입장에서 17개 시도와 발전연구원을 중심으로 데이터 소재정보를 정리하였다. 각 지방자치단체 대부분은 각 지역의 특징을 나타내는 통계정보를 잘 가공하여 제공하고 있었으며, KOSIS에서 제공하는 지역통계들을 검색할 수 있는 항목을 함께 제공하고 있었다. 대부분의 경우 정보 활용이 원활하도록 높은 수준의 정보포털의 형식을 갖추고 있다.

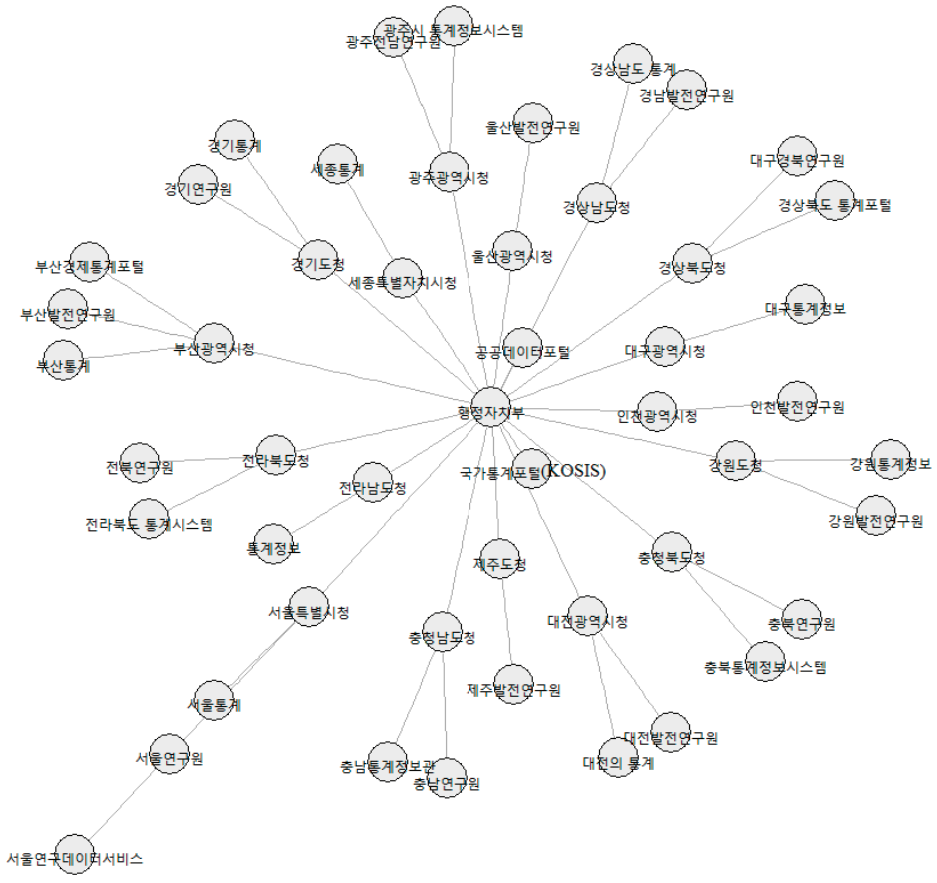


그림 3.25 지방자치단체 cytoscape

그림 3.25는 행정자치부를 중심으로 지방자치단체와 발전연구원이 연계되어 있음을 나타낸 것이다.

표 3.17 지방자치단체 정보사이트 정리

지방자치기관	사이트이름
강원도	1. 강원통계정보**
경기도	2. 경기통계**
	3. 경기데이터드림**
경상남도	4. 경상남도 통계*
경상북도	5. 경상북도 통계포털*
광주광역시	6. 광주시 통계정보시스템*
대구광역시	7. 대구통계정보*
대전광역시	8. 대전의 통계**
부산광역시	9. 부산통계
	10. 부산도시서비스분석시스템(SSA-GIS)**
	11. 부산경제통계포털*
서울특별시	12. 서울통계**
	13. 서울열린데이터광장**
세종특별자치시	14. 세종통계*
울산광역시	15. 통계*
인천광역시	16. 인천통계
전라남도	17. 통계정보
전라북도	18. 전라북도 통계시스템**
제주특별자치도	19. 통계정보
충청남도	20. 충남통계정보관*
충청북도	21. 충북통계정보시스템**

대부분의 지방자치단체들의 정보제공 사이트(표 91)들은 살펴볼 가치가 있을 만큼 잘 정비되어있었지만, 특히 서울의 '열린데이터광장'(경기도의 경기데이터드림도 유사한 성격임)과 부산의 '도시서비스분석 시스템'은 다른 지역에 비하여 특화된 모습을 보여준다(그림 3.26).

이름에서 알 수 있듯이 이 사이트(http://sgis.busan.go.kr)는 자료제공을 목적으로 하기 보다는 분석서비스를 제공하는 것에 초점이 맞추어져 있는 것이 특징이다. 지방정부 3.0에 부합하는 과학적 행정구현을 위한 수요분석에 세분화된 인구 제공을 위하여, 인구산정을 위한 기초자료는 모바일 관련 빅데이터(1시간 정보량 4G), 소지역(50m×50m) 인구의 시각화를 위한 지도정보시스템을 결합하여 현주인구를 추정하여 제공(그림 39)한다.



그림 3.26 부산도시서비스분석시스템(SSA-GIS)



그림 3.27 부산 현주인구 추정 구조

'서울 열린데이터 광장'(http://data.seoul.go.kr)에서는 2600여종의 공공데이터를 개방해 놓은 사이트로 오픈데이터, 데이터서비스(시각화, 25개 자치구 열린데이터광장 연결)를 제공한다. 사실 이 사이트 정부 3.0 공공데이터 개방의 일환으로 구축된 것으로 분석용 자료를 다운로드하기에는 그리 편리한 것은 아

니지만, 다른 지역에서 참고해볼 만큼 내용들이 잘 구성되어있다(그림 3.28).



그림 3.28 서울열린데이터광장

지방자치단체의 경우 통계자료의 확보는 정리한 자치단체별 통계사이트에서 충분히 확보할 수 있다. 지방자치단체의 통계자료를 찾을 수 있는 또 다른 곳은 지역별 발전연구원 들이다. 정형자료들의 경우 앞서 소개한 20개의 사이트와 중복되는 부분들을 피할 수 없지만, 연감이나 연구보고서 등과 같은 비정형 자료들을 체계적으로 정리하여 제공하는 점들을 특징으로 볼 수 있다.

표 3.18 시도 데이터 제공 발전연구원 사이트

지방자치기관의 발전연구원	사이트 이름
서울연구원	22. 서울연구데이터서비스*
인천발전연구원	23. 월별경제동향(문서자료)
대구경북연구원	24. 경제동향분석(문서자료)
대전발전연구원	25. 월간대전경제(문서자료)
광주전남연구원	26. 광주전남경제동향(문서자료)
울산발전연구원	27. 통계 DB*
부산발전연구원	28. BDI 간행물
경기연구원	29. 통계로보는 경기도*
강원발전연구원	30. 연구보고서(문서자료)
충북연구원	31. CRI 발간물(문서자료)
충남연구원	32. 연구원 간행물
전북연구원	33. 보고서
경남발전연구원	34. 연구보고서(문서자료)
제주발전연구원	35. 제주통계

각 지자체의 연구원은 14곳을 찾을 수 있었다(표 3.18). 통계데이터를 제공하는 형식은 앞서 소개되었던 사이트들에 비하여 다소 부족한 점이 있지만, 발전연구원의 연구보고서나 조사결과 등을 비교적 자세히 정리하여 제공하고 있다. 서울연구원의 '서울연구데이터서비스'는 다른 지역 대비 상대적으로 잘 정리되어 있는 것을 그림 3.29에서 알 수 있다.

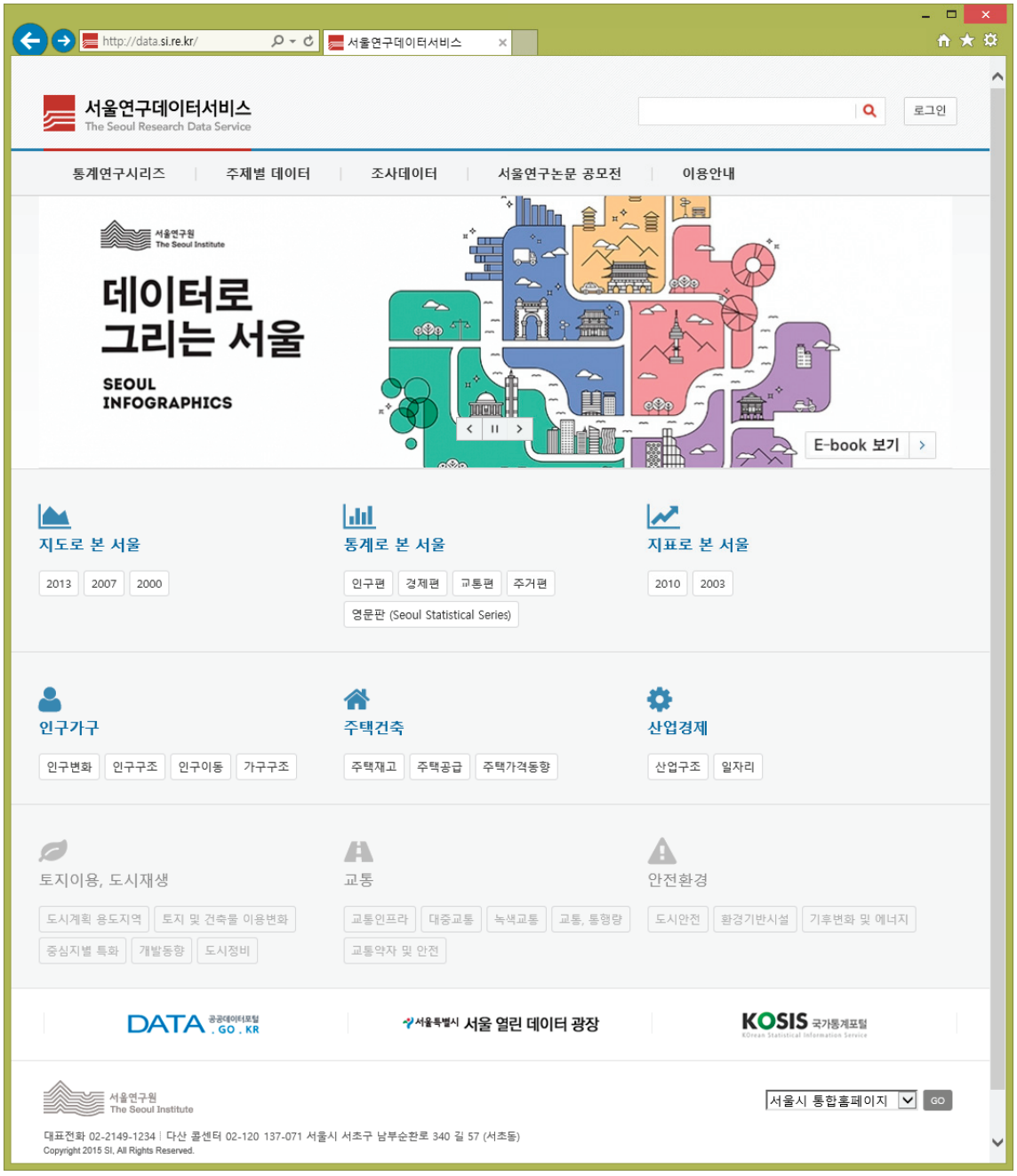


그림 3.29 서울연구데이터서비스

3.6 정부 3.0

현재 정부가 정의한 정부3.0의 개념(표 3.19 참조)은 다음과 같다.

“공공정보를 적극 개방·공유하고, 부처 간 칸막이를 없애고 소통·협력함으로써, 국정과제에 대한 추진동력을 확보하고, 국민 맞춤형 서비스를 제공함과 동시에 일자리 창출과 창조경제를 지원하는 새로운 정부운영 패러다임”

표 3.19 전자정부 발전모형 특징비교

구분	전자정부1.0	전자정부2.0	전자정부3.0
가치	정부중심	시민참여	시민, 정부협업
정부서비스	단방향 임시처방	쌍방향 사후대응	맞춤형/개인화 사전예측
정보기술	웹1.0(텍스트와 링크가 주된 형태, 하이퍼텍스트 기반의 월드와이드 웹)	웹2.0(정보의 개방을 통해 인터넷 사용자들 간의 정보공유와 참여를 이끌어냄, 이를 통해 정보의 가치를 지속적으 로 증대시킴	웹3.0(지능화된 웹이 시맨틱 기술을 이용해서 상황인식을 통해 이용자에게 맞춤형 콘텐츠 및 서비스를 제공)

출처 : 류광택 외(2012)

정부 3.0의 3대전략과 10대 중점과제에 의하면, ‘공공정보 적극 공개로 국민의 알권리 충족’, ‘공공데이터의 민간 활용 활성화’, ‘빅데이터를 활용한 미래지향적 행정구현’ 등이 있다. 이의 일환으로 개발되어 제공되는 것이 ‘공공데이터 포털’ 이다.

공공데이터포털(Data Portal)은 한국정보화진흥원에서 운영하는 공공데이터 통합제공 시스템이다. 대한민국 정부가 보유한 다양한 공공데이터를 개방하여 누구나 편리하고 손쉽게 활용할 수 있게 하는 것을 목적으로 한다(그림 3.30 참조).

2016년 7월 초기준, 16,554개(네이버 데이터랩에서도 조회돠 다운로드 가능, 17,943개의 자료가 제공되는 것으로 나옴)의 공공데이터가 공개되어 있다. 정

부부처 및 산하기관에서 발행하는 백서, 주제별 전문가가 선정한 최신 이슈와 관련된 지식, 지역/분류체계/제공기관/활용방법 별 공공데이터 개방현황 등이 제공된다. 파일 다운로드 또는 Open API를 통해 이용할 수 있다.



그림 3.30 공공데이터포털

CHAPTER 4

우리나라 빅데이터 활용현황과 과제[35)]

학습목표

- 통계청의 빅데이터 활용현황을 통계작성, 조사통계와의 결합 관점에서 학습한다.
- 중앙행정기관과, 공공기관, 지방자치단체의 빅데이터 활용현황에 관하여 학습한다.
- 민간 및 기업, 학계 및 연구기관의 빅데이터 활용현황에 관하여 학습한다.

2014년 안전행정부는 국가 정책수립, 대국민 서비스(복지·창업지원) 및 사회현안 대응(치안·사이버 공격) 등 국정운영 전반에 빅데이터 활용이 본격화되고, 공공·민간의 공개된 데이터를 수집·분석하는 '빅데이터 분석 플랫폼'을 관계부처 공동으로 이용할 수 있도록 하는 '정부 3.0'의 핵심과제인 빅데이터 활용 확대 방안의 추진을 발표하였다.

이를 위하여 먼저, 2017년까지 97개 빅데이터 활용사업을 추진하여, 데이터 기반의 과학적 정책수립과 맞춤형 서비스 제공을 확대하기로 하였다(표 4.1 참조). 이중 국민생활·안전, 일자리 창출 및 국정과제 관련 23개 사업은 안행부·미래부가 중점지원 과제로 선정, 우선적으로 예산을 지원한다. 향후 계획에 따라 신규 사업을 추진하고, 2013년에 완료된 여가부 '위기 청소년 조기경보', 경찰청 '범죄확률 사전예측', 통계청 '일일 물가분석' 등 주요 과제들의 서비스

35) 이 장의 내용들을 통계청의 빅데이터 관련 보도자료 내용을 옮겨오거나 발췌하여 정리한 것임

제공도 본격화 하였다.

표 4.1 빅데이터과제 추진계획

구 분	계	’13년	’14년	’15년	’16년	’17년
합 계 (과제 수)	97	25	25	26	15	6
중앙부처	61	19	15	12	12	3
지자체	36	6	10	14	3	3
(중점과제)	(23)	(8)	(4)	(4)	(3)	(4)

* 빅데이터 예산 ’13년 360억(중앙 345억, 지자체 15억), ’14년 460억(중앙400억, 지방60억)
출처 : 통계청 2014년 1월 7일 보도자료

또한, ‘빅데이터 공통기반 플랫폼’(’13.12구축)을 통해 행정기관들이 공개된 공공과 민간의 데이터를 자유롭게 수집·분석·공유할 수 있게 지원하고, 공공데이터전략위원회(국무총리 소속)를 활용하여 공공부문의 빅데이터 추진전략과 국가적 우선순위를 심의·조정 할 계획이다.

아울러, 미래부 등 관계부처와 함께, 빅데이터 분석을 통한 미래전략 수립과 현안해결을 지원할 ‘국가미래전략센터’ 구축방안을 마련하고, 산·학·연 협력을 통한 빅데이터 처리·저장 등 기초기술 개발과 전문인력 양성도 지원해 나갈 계획이다. 이로써 빅데이터 활용으로 정부 3.0의 핵심인 국민 맞춤형·지능형 서비스 제공을 효과적으로 추진할 수 있는 기반을 갖추게 되었다고 볼 수 있다.

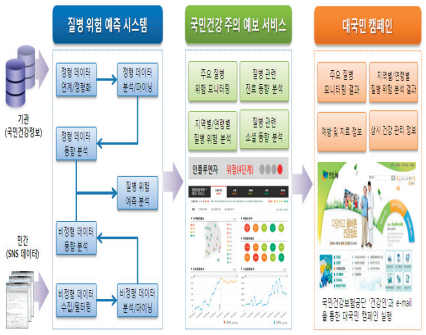
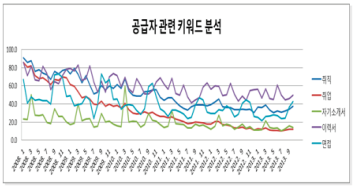
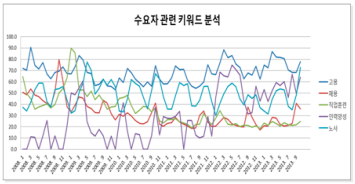
다음은 빅데이터 활용을 통한 정부운영 및 국민생활 변화 사례들을 정리한 것이다(표 4.2~표 4.5).

표 4.2 빅데이터 활용을 통한 정부운영 및 국민생활 변화 사례

여성부 : 위기청소년 지도(맵) 분석으로 위기 청소년 조기 발견·상담
청소년쉼터에서 근무하는 김씨(남, 33)는 가출 청소년이 비행, 범죄, 성매매 등에 노출되기 전에 조기 발견하여 안전한 청소년쉼터로 인도하는 거리상담활동(아웃리치)을 한다. 그동안 주로 유동인구가 많은 곳에 나가서 홍보하는 것에 그쳤으나, 빅데이터 분석으로 개발된 위기청소년 지도에 따라 밀집지역, 활동시간, 이동경로를 중심으로 거리상담 활동을 한 결과 더 많은 가출청소년을 보다 효율적으로 발견하게 되었다. 그 결과 00역 주변의 거리·노숙 청소년과 가출팸은 현저히 줄어들었으며, 청소년 쉼터 이용률도 크게 증가하여 보다 많은 청소년들이 희망을 충전하게 되었다.
식약처 : 인터넷 수집 데이터를 분석으로 의약품 부작용 조기경보 가능
식약처에 근무하는 임모 사무관은 한국의약품안전관리원 등을 통해 의약품 부작용 정보를 수집하여 안전정보를 제공하는 업무를 담당하고 있다. 정보부족으로 의약품 안전성 이슈를 조기에 판단하는데 어려움을 겪어왔다. 하지만, 빅데이터 분석을 통해 SNS상의 의약품 부작용 정보와 위해요인을 검출할 수 있게 되어 이슈가 될 만한 부작용 정보를 신속히 파악할 수 있었다. 또한, 병원의 환자진료기록정보 데이터 및 의약품 부작용 보고관리 DB와 비교 검증하여 의약품 안전성 여부를 조기에 판단하여 신속히 알릴 수 있게 되었다.
중기청 : 소상공인 창업지원 서비스를 통해 창업성공률 제고
퇴직금 3억원을 종자돈으로 호프집을 창업할 계획이었던 홍고민씨는 상권정보시스템에서 제공하는 ‘점포평가 서비스’ 도움을 받고 분식점 창업으로 마음을 돌렸다. 홍고민씨가 선택한 점포에 대한 최근 5년간의 개업 이력을 조회한 결과, 호프집은 1년 남짓 영업하다가 폐업한 경우가 많음을 알게 되었기 때문이다. ‘점포 평가 서비스’를 통해 점포 창업시에 필요한 과거 개·폐업 이력, 추정매출, 입지 분석 등을 통해 해당 점포에 가장 적합한 업종을 찾을 수 있어 창업에 성공할 수 있었다.
서울시 : 통화량 유동인구 빅데이터를 통한 심야버스 노선수립 지원
회사원 K씨, 밤 11시가 넘었는데도 아직도 할 일은 산더미처럼 쌓여있다. “오늘도 이 일 다 끝내고 가려면 택시를 타고 가야겠군. 회사에서 택시비라도 지원이 되면 좋으련만..” 안 그래도 박봉에 시달리는 K씨는 고민이 이만저만이 아니다. 또 다른 회사원 L씨, 간만에 친구들과 홍대입구에서 만나 시간이 가는 줄 모르고 회포를 풀고 있다. “앗, 벌써 한시야? 지하철 막차가 끊겼네. 집까지 가려면 택시비 만만치 않은데..” 라며 걱정하던 K씨, L씨. 심야시간 통화량 분석을 통한 버스노선 조정으로 집 앞까지 심야버스가 운행하는 것을 알고 만족해하였다.

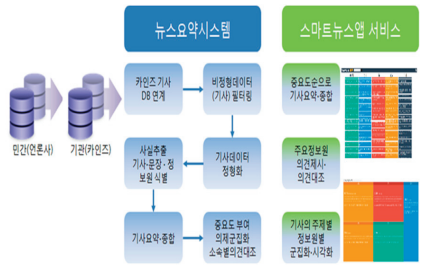
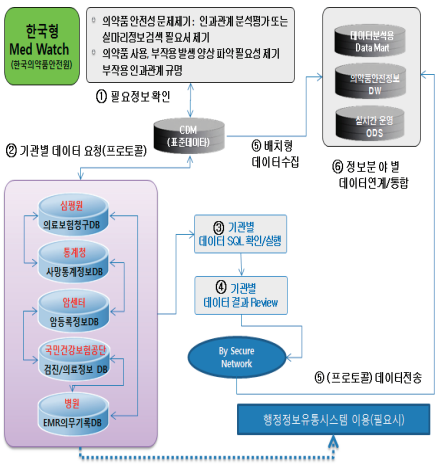
출처 : 통계청 2014년 1월 7일 보도자료

표 4.3 정부 부처별 빅데이터 과제 수행 사례 1

구 분		내 용	주요 이미지
빅데이터 로그분석으로 보안사고 및 장애 예방 강화 안행부	배경	정보통합전산센터가 운영하는 장비의 급격한 증가에 따라 기존 기술로는 실시간 운영 상태 점검에 한계 도달	
	내용	**(분석 데이터)** 2만여 정보시스템의 Log 및 작업승인 내역 ※ 6개월 동안 Log 데이터(2PB) **(분석 내용)** 악성위협IP의 접근기록 탐지 및 활동흔적 추적 및 장애의 패턴을 인지하여 동일 패턴의 Log 발생 시 담당자 경보	
	효과	**(분석시간 단축)** 종합분석시간 12시간→10분 이내로 단축함으로써 대응능력 강화 **(장애식별 용이)** 장비상태를 한 곳에서 보여줌으로써 장애 원인 식별 용이 **(사이버공격감시 용이)** 해킹, DDoS 공격징후 분석·탐지, 악의적 이용자의 접근 실시간 감시	
국민건강정보 DB를 활용한 맞춤형 건강서비스 개발 복지부	배경	예방 중심의 건강관리를 위해 국민건강정보 DB를 활용한 맞춤형 통합 건강관리 서비스 제공 필요	
	내용	**(국민건강 주의예보)** 질병관리 본부와 협조체계를 구축('14년)하여 공단의 질병정보와 민간 SNS 자료분석을 통해 특정 질병에 대해 위험 예보 **(맞춤형 건강서비스)** 건강in 홈페이지 고도화 사업을 통하여 개인 건강기록 및 건강나이, 뇌졸중 위험도 예측 등 서비스 제공('14년)	
	효과	사전예방적 건강증진 체계 구축 및 지역별 건강통계를 지역보건의료 정책 수립의 기반으로 활용 가능	
일자리 현황분석을 통한 고용 수급 예측 고용부	배경	미래 일자리 현황에 선제적으로 대응하고 인력수급전망 결과 활용도를 제고하기 위해, SNS, 전화상담 내역 등을 추가로 활용할 필요성 제기	 
	내용	**(분석 데이터)** 경제활동인구조사, 국민계정, 고용노동통계조사, 패널조사데이터, 고용보험, 산재보험, 워크넷, SNS상의 직업 관련 검색 및 키워드, 전화상담기록 등 **(분석 상세내용)** 정형적비정형* 데이터 융합을 통해 미래 주요 산업(직종) 트렌드, 이슈 등 발굴 *직업 관련 검색 키워드 정보 등	
	효과	다양한 사회·문화적 요인을 반영한 인력수급 전망 가능	

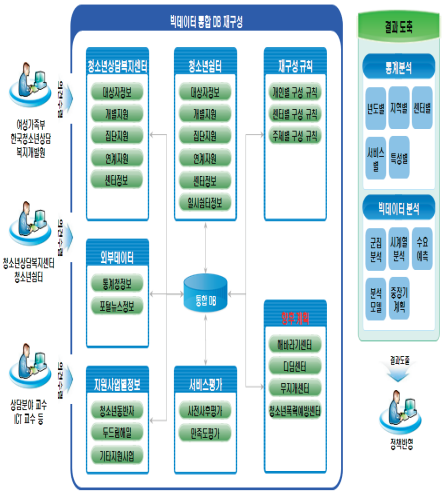
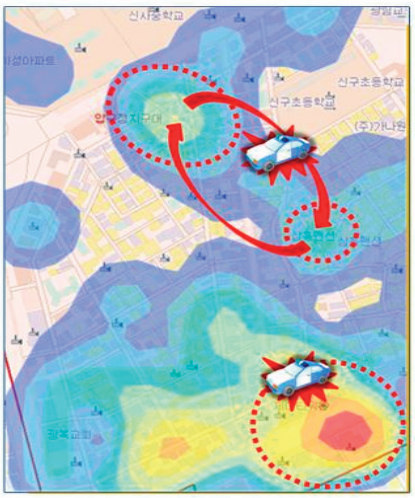
출처 : 통계청 2014년 1월 7일 보도자료

표 4.4 정부 부처별 빅데이터 과제 수행 사례 2

구 분	내 용		주요 이미지
빅데이터를 활용한 물가통계 생산 통계청	배경	국민이 체감할 수 있는 시의성 있는 통계를 생산하기 위하여 온라인가격데이터 등을 활용한 통계 생산 필요	
	내용	**(분석 데이터)** 온오프라인 병행 사이트 등 데이터 수집에 적합한 사이트로부터 수집한 가격정보 및 비정형 데이터 **(분석 상세내용)** BPP*를 벤치마킹하여 우리 실정에 맞는 일일물가지수를 개발 * BPP(Billion Price Project) : 온라인에서 수집하는 가격정보에 기반을 둔 물가통계 생성 프로젝트 -일일물가지수, 감성지수(물가변동에 대한 국민정서 영향지수) 추세 분석 및 물가 예측	
	효과	다양한 물가정보를 적시에 제공함으로써 신속한 정책수립에 활용 가능	
빅데이터 활용 스마트 뉴스 모바일 앱 개발 교육부	배경	빅데이터를 활용한 지능형 뉴스 제공	
	내용	**(분석 데이터)** 대량의 기사DB에 대한 중요도, 관계도 등 **(분석 상세내용)** -뉴스기사 시간별 매체별 요약, 중요성의 관계 시각화 -뉴스정보원 시각화로 전문가 식별 및 관계 파악	
	효과	신속한 지능형 뉴스 검색 서비스 제공 가능	
빅데이터 기반의 의약품 안전성 조기경보 서비스 식약처	배경	소비자(환자)의 자발적 부작용 신고 비율이 낮아 수집되지 않은 잠재적인 부작용 정보가 있을 수 있어 부작용 수집 경로의 다양화 필요	
	내용	**(분석 데이터)** 포털·SNS상의 의약품 부작용 정보, 병원 환자진료기록정보 등 **(분석 상세내용)** -포털, SNS상의 비정형화된 의약품 부작용 정보 분석을 통해 부작용/오남용 관련 이슈 도출의 가능성 검토 -향후 제약사, 의약전문가 등으로부터 보고된 부작용 정보의 분석·평가 시 활용하여 의약품의 안전성 이슈를 사전에 예측하고 소비자 등에 알림서비스 제공	
	효과	촘촘한 부작용 정보 수집으로 의약품 위해요인 조기발견 및 선제적인 안전조치에 따른 의약품 부작용 피해 최소화	

출처 : 통계청 2014년 1월 7일 보도자료

표 4.5 정부 부처별 빅데이터 과제 수행 사례 3

구 분	내 용		주요 이미지
위기청소년 빅데이터 분석으로 적시 대응체계 구축 여가부	배경	청소년자살, 학업중단, 가출 등 청소년 위기문제를 사전 예측하여 선제적으로 대응할 수 있는 시스템 필요	빅데이터 분석을 위한 통합DB 재구성 및 결과 도출 수행 
	내용	**(분석 데이터)** CYS-Net정보망(개인상담, 전화·문자·사이버상담 정보 등), 청소년쉼터정보시스템(가출원인, 성별, 연령, 상담정보 등), 각종 소셜데이터 등 **(분석 상세내용)** -1388청소년전화 발신정보 및 상담내용 분석을 통해 위기청소년의 지역별분포, 위험시간, 이동경로 파악 -자살 원인, 원인별 위기 패턴을 분석으로, 향후 온라인, SNS 등에 남긴 메시지를 분석하여 부모, 교사에게 위기경보를 조기에 알리는 안심서비스 제공	
	효과	청소년 위기 문제에 선제적으로 대응하여 비행범죄 등 고위기로의 이행 차단 및 충동적 자살 예방	
지리적 프로파일링 시스템 고도화로 범죄예측력 강화 경찰청	배경	지리적 프로파일링을 통해 범죄에 선제적으로 대응하고 있으나, 빅데이터를 활용하여 분석 정확도 향상 필요	
	내용	**(분석 데이터)** 범죄기록, 유동인구, 주민신고정보, CCTV위치정보 등 **(분석 상세내용)** -인구통계학적 변인을 가미한 범죄위험지역 분석 -치안활동을 기존 '관서 및 행정동'에서 범죄위험 예측·분석결과를 반영한 '치안블록' 단위로 개편 -현장 활용성 제고를 위한 경찰 內 타부서 정보 연계를 통해 장소·시간대별 범죄확률 향상	
	효과	범죄 예측을 통한 순찰차 사전 배치 등 선제적 치안 활동 전개로 범죄 발생감소 및 신속한 수사 가능	

출처 : 통계청 2014년 1월 7일 보도자료

4.1 통계청의 활용현황

통계청의 빅데이터 활용은 '새로운 정보 제공', '민간의 빅데이터 통계 생산 지원', 빅데이터 활용기반 구축'으로 요약할 수 있다. 이들에 관한 내용을 정리하면 다음과 같다.

1. 빅데이터를 활용한 새로운 정보 제공

공공데이터와 민간 빅데이터를 연계·분석한 정보 제공을 말하며, 통계청 이 보유한 원시 공공데이터(행정자료 89종+조사통계 42종)와 민간 보유 빅데이터 간 연계·분석을 통하여 유용한 정보를 제공하게 된다. 이의 사례로 통계청 통계 + 통신사 로밍통신 자료 + 신용카드 매출액 + 외국인 환율정보 4가지를 연계 활용한 외국인 관광객 소비동향 통계가 있다. 이종의 자료를 연결할 식별정보가 없는 경우는 통계적 연계(Statistical matching) 방법을, 식별정보가 포함된 자료의 경우 현행법상 비식별화 후 정확 연계(Exact matching) 방법 등 사용한다. 이 밖에도 민생관련 정책수립을 지원하기 위해 민간 빅데이터를 분석하여, 기존 민생 통계를 보완하는 보조적인 통계를 생산하는 분야를 확대해 나갈 예정이다.

통계청에서 빅데이터를 활용한 통계생산의 예로 다음과 같은 것들이 있다.

- 대형마트, 인터넷 쇼핑몰 등 온라인 가격정보를 활용한 온라인물가지수
- 카드사용 데이터를 활용하여 실시간 업종별(음식점업, 숙박업 등) 경기 지표 개발
- 웹상 공개된 소셜 미디어(블로그, 페이스북 등) 데이터(메시지의 빈도수 및 긍부정 감정 등)를 활용한 새로운 소비자심리지수 개발

2. 민간의 빅데이터 통계 생산 지원

민간의 빅데이터 활용을 지원하기 위하여 정부가 보유한 빅데이터를 민간에 제공하며 이때 비식별화(Masking) 처리 지원 및 기법 관련한 컨설팅을 제공한

다. Masking 방법이란 익명화, 개별 자료를 구간 자료로 대체, 개별 자료를 집단 평균자료로 대체, 자료 중 일부를 추출 제공함으로써 데이터 제공을 통한 개인정보의 유출을 데이터 단계에서 차단하는 기법을 말한다.

또한 민간에서 발표하는 빅데이터 분석결과 가운데 검증이 가능하고 중요한 사안에 대해, 이용자 혼선 방지를 위해 신뢰성 검증을 추진한다. 신뢰성 검증이란 통계청이 보유한 공공데이터(행정자료, 조사자료 등)와의 추세 및 시계열 분석 등을 통해 비교·사후 검증하는 것을 말한다.

3. 빅데이터 활용기반 구축

통계청에서는 빅데이터 활용을 위한 기반구축을 위해 국내외 네트워크 구축, 통합 DB 구축, 제도적 기반구축을 추진하고 있다. 먼저 국내외 네트워크 구축을 위하여 네이버 MOU 체결, 강원창조혁신센터 MOU 체결 등 지속적으로 기간 간 협력체계를 확대해 나가고 있으며 빅데이터 관련 업무 추진 시 발생되는 문제의 실질적 해결을 위해 학계·업계·전문가·정부부처 관계자 등이 참여하는 빅데이터 활용전략의 수립과 문제 해결방안을 논의하는 포럼을 구성·운영하고 있다. 또한 UN의 빅데이터 전문가작업반(GWG) 회원 가입 및 국제 전문가들과의 네트워크 구축을 통해 국제동향 파악을 꾸준히 진행하고 있다.

중장기적 계획을 빅데이터 분석 기반 구축을 위한 행정자료와 민간자료를 통합한 공동 활용 DB 구축을 추진하고 있으며, 민간 빅데이터 수집, 활용 근거 등 법적 근거 마련과 행정자료 활용확대를 위한『행정자료를 이용한 통계작성 법률(안)』제정 추진 및 민간 빅데이터 활용 확대를 위한『금융실명법』등 개정 검토를 추진하고 있다.

4.1.1 빅데이터 이용 통계작성

빅데이터를 활용한 농가판매가격조사[36)]

통계청은 오는 2016년 7월 1일부터 농가에서 생산된 농축산물 등의 농가수취가격을 조사하여 농가판매가격지수를 작성하는 농가판매가격조사의 자료수집 방식을 표본 대상처를 직접 방문하는 현장조사 방식에서 농협중앙회가 보유한 빅데이터를 활용하는 방식으로 전면 전환한다. 현재 농가판매가격조사는 69개 대상품목 중 마늘, 양파를 포함한 48개 품목에 대해 농협 및 농가를 방문하여 매순기 38개 품목 및 매월10개 품목 조사를 실시하고 있다(축산물 등 21개 품목은 관련 협회 등을 통해 입수하고 있음).

조사에 의한 통계작성을 빅데이터를 활용하여 작성하기 위하여 통계청은 지난해 농협중앙회와 자료공유에 대한 협의를 완료하고 그동안 농협 빅데이터 기반 통계작성을 위한 시험연구를 진행하였다. 품목별로 품목매칭 및 품질을 표준화하고 최신 통계이론을 바탕으로 이상치를 탐색·제거함으로써 농협 빅데이터 기반 농가판매가격의 대표성과 안정성을 확보하기 위해 노력하였다. 아울러, 통계청은 농협중앙회 빅데이터의 활용에 따라 국가통계의 정확성 제고는 물론 체감물가와의 차이를 크게 해소할 수 있을 것으로 기대하고 있다. 그동안 농가판매가격 현장조사가 주산지를 중심으로 구성되었음에도 불구하고 일부 품목의 경우 거래빈도가 낮아 체감물가를 제때에 반영하기 어려운 한계가 있었다. 이에 반해 농협중앙회의 빅데이터는 전국의 일일자료 형태로 축적되어 있어 농가판매가격의 정확성 제고에 크게 기여할 것으로 보인다.

통계청은 이번 농가판매가격조사에서의 성공적 경험을 토대로 민간 및 공공부문이 보유한 빅데이터를 통계작성에 전면 활용하는 국가통계생산 패러다임 혁신을 점차적으로 확대해 나갈 계획이다. 이를 통해 현재 현장조사에 의존하는 고비용-저효율의 통계생산방식을 탈피하고 과학적인 통계방법론에 기반을 둔 통계작성으로 국가통계의 정확성 제고 및 신뢰 확보를 위해 노력해 나갈 예정이다.

36) 통계청 2016년 6월 27일 보도자료의 내용을 옮겨오거나 발췌하여 정리한 것임

4.1.2 빅데이터와 조사통계의 결합[37)]

통계청의 빅데이터와 조사통계의 결합사례를 찾는 것은 어려움이 따른다. 1장에서 연계된 행정자료를 행정자료 빅데이터로 볼 수 있음을 언급하였으나 통계청은 이를 행정자료로 구분하고 있기 때문이다. 통계청의 빅데이터통계과(2015년 9월 신설)를 중심으로 조사통계와 빅데이터를 결합하려는 연구는 계속되고 있겠지만, 정확하게 일치하는 사례가 발표된 것은 찾을 수 없었다. 통계청의 2016년 업무계획을 보면, '국가통계 생산 패러다임 전환'에 대한 내용을 다루고 있으며, '행정자료 활용을 위한 협업체계 강화'와 '빅데이터 활용 가반 구축 추진'에 대한 내용이 있다.

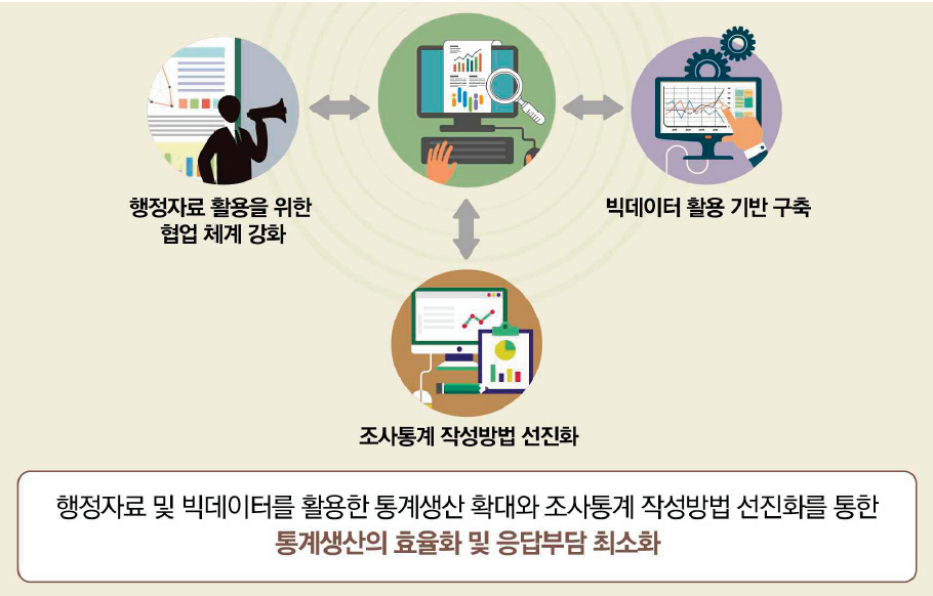


그림 4.1 국가통계 생산 패러다임의 변화

특히 '빅데이터 활용 기반 구축 추진'에 대해서는 다음과 같이 설명하고 있다.

> 빅데이터*와 통계데이터**간 융복합으로 의미 있는 통계를 작성할 수 있도록 제도적 기반을 마련하고, 모범 사례를 발굴하여 빅데이터 활용 확산
>
> * 통신자료, 신용자료 등 민간이 보유한 대용량 자료
> ** 조사자료, 행정자료 등 통계작성을 위하여 수집한 자료

37) 이 절의 내용은 통계청의 2016년 주요업무 추진계획의 내용을 옮겨오거나 발췌하여 정리한 것임

‘빅데이터 활용 기반 구축 추진’에 관하여 4가지 구체적인 내용을 설명하고 있으며 다음과 같다.

(1) 제도적 기반 마련

통계작성을 위해 통계자료와 빅데이터를 연계·융합할 수 있도록 제도적 기반 구축하기 위하여 개인정보를 보호하면서 데이터간 연계로 빅데이터의 잠재가치를 활용하는 방안에 대한 범국민적 공감대 형성 추진한다(개인정보보호법상 통계 목적으로는 자료 공유가 가능하지만, 개별법에 의하여 자료 공유가 금지되어 있어서 연계·분석에 한계가 있음).

각계의 의견을 수렴하여 통계 목적으로 빅데이터를 활용하기 위한 통계법 등 관련 법 개정 검토 추진(외국의 사례를 토대로 통계청이 각종 데이터를 수집, 연계, 분석하여 그 결과를 공유하는 방식을 검토)

> * 네덜란드, 덴마크 등은 행정자료는 물론 금융자료 등 빅데이터도 통계청이 수집, 연계, 분석하는 데이터 허브 기능 수행

(2) 개인정보 보호 강화

빅데이터 활용시 개인정보를 완전하게 보호할 수 있도록 제도적 장치를 마련하며, 외국 사례를 참조하여 비식별화 등 개인정보를 보호하기 위한 장치를 규정하는 통계청장 훈령 제정 추진한다.

(3) 빅데이터 통계교육 활성화

빅데이터를 이용한 신규통계 개발 및 빅데이터 활용기반 구축을 위하여 전문 통계인력 양성한다. 기존 빅데이터 교육과정 내용 개편 및 수준별 교육과정 운영을 통한 선진통계기법 적용 능력 배양하고, 신규 교육과정 개발 및 관련기관과의 협업을 통해 빅데이터 통계교육 실시한다.

※ 통계청-한국데이터베이스진흥원 통계교육 MOU(보도자료 2015.12.22.)
양 기관은 빅데이터 관련 통계교육에 있어서 통계부문과 데이터부문에서 각각 특화된 교

육과정을 분담함으로써 교육효과를 배가시키고, 교육프로그램의 개선과 개발, 보급확산에도 공동의 노력을 기울임으로서 빅데이터 통계교육의 발전에 기여하기로 하였다.

(4) 연계사례 발굴

통계청이 보유한 통계데이터(행정자료 93종, 조사통계 40종)와 빅데이터간 연계를 통한 통계작성 활용 사례 발굴 및 DB 시범구축을 진행한다. 통계 목적의 빅데이터 활용을 확산하고, 이를 제도화 할 수 있는 계기로 활용하도록 한다.

구축 DB 유형
ㅇ (부채·신용 DB) 통계청의 가구 모집단 통계와 민간 부채·신용 빅데이터를 연계(개인정보 마스킹, 자료에디팅 등 통계기법 적용으로 개인정보 문제해결)하여 부채·신용 DB 구축 및 공동연구 추진 ⇒ 가구 유형별, 가구주 특성별 부채 현황 파악, 취약계층의 부채 정보 다원화 및 가계금융복지조사의 통계조사 보완 등 활용 ㅇ (창업지원 DB) 네이버 등 민간과 공동으로 국민들의 창업 활동을 지원하기 위해 빅데이터를 활용한 다양한 창업 정보 제공(소비자들의 창업관련 검색기반 정보, 업체의 신용정보, 기업체의 업종별·소지역단위별 창·폐업 등 경기관련 정보 등을 제공)

※ 빅데이터 기반의 가계부채 연구를 위한 통계청-KCB 업무협력 체결(보도자료, 2016.1.22.)
공공데이터와 민간 빅데이터간 연계·분석하는 최초의 빅데이터 활용 민관협력 사례로서 의미가 있음.

통계청에서는 행정자료와 조사통계 빅데이터의 결합을 위하여 다각도로 노력을 하고 있어서 추후 정보력이 높은 유용한 통계가 새롭게 나타날 것으로 기대된다.

최근 유경준 통계청장의 '빅데이터 시대에 즈음한 경제총조사의 진화(서울신문, 2016년 7월 25일)'라는 기고에서 "인구주택총조사와 경제총조사가 모두 현장조사가 아닌 행정자료를 기반으로 하는 등록센서스로 전환된 것은 빅데이터 시대에 즈음한 통계생산 방식의 패러다임 변화의 출반선이라고 하였다" 경제총조사의 경우 개인의 주민등록번호처럼 기업과 사업체에 고유한 번호를 부여하고, 국세자료와 현장조사 결과를 최신 상태로 통합하여 원스톱으로 처리할 수

있는 기업등록부(BR: Business Register) 구축을 가장 큰 진전으로 언급하였다. 인구총조사의 경우는 11개 행정기관의 자료연계와 20%의 표본조사의 결합이 이와 같은 진보에 해당될 것이다.

행정자료빅데이터라는 관점에서 경제총조사와 인구추택종조사(등록센서스)를 간략히 소개하였다.

경제총조사는 우리나라 전체 산업에 대한 고용, 생산, 투입(비용) 등에 관한 구조를 파악하기 위하여 동일시점에 통일된 조사기준으로 실시하는 대규모 전수조사를 말하며(그림 참조), 국내에서 산업활동을 수행하고 있는 종사자 1인 이상인 모든 사업체가 경제총조사의 대상이 된다. 이번 경제총조사에서 통계청은 앞서 언급하였던 것처럼 이번 조사를 바탕으로 기업등록부를 구축하였으며 이를 통해 차기 경제총조사를 행정자료 기반의 '등록센서스' 방식으로 추진해 현장조사를 최소화함으로써 보다 정확하고 효율적으로 통계를 제공할 계획이다(보도자료, 2016.7.24.)

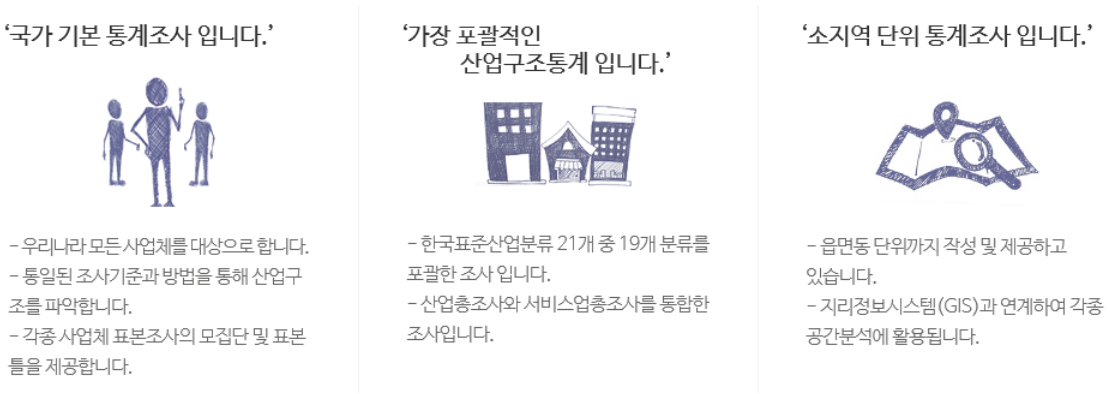


그림 4.2 경제총조사

새로운 방식의 인구주택총조사 방법인 등록센서스는 다시 완전등록센서스(fully Register-based Census)와 부분등록센서스(partly Register-based Census)로 구분할 수 있다. 완전등록센서스는 100% 행정자료만을 연계·활용하여 인구주택총조사를 대체하는 방식으로, 덴마크, 스웨덴, 오스트리아 등에서 도입하고 있고, 부분등록센서스는 행정자료와 함께 표본조사 등 현장조사를 병행하는 방식으로 독일, 스페인, 터키 등이 도입하고 있다(표 4.6 참조, 그림 4.3 참조).

표 4.6 센서스 방식의 구분

구분	설명
전통적인 방식 (현장조사 기반)	인구 · 가구 · 주택의 규모, 구조, 모든 특성에 대한 정보를 조사표와 함께 현장 면접조사, 우편조사 등을 통해서 수집하여 제공
부분 등록센서스 (행정자료 + 현장조사)	인구 · 가구 · 주택의 규모 및 구조와 행정자료만으로 파악이 가능한 기본적인 특성 정보는 행정자료를 이용하여 생산하고, 기타 상세한 특성에 대한 정보는 표본조사를 통해 수집하여 제공
완전 등록센서스 (행정자료 기반)	인구 · 가구 · 주택의 규모와 구조 뿐 아니라, 모든 특성에 대한 정보를 행정자료만을 이용하여 생산하고 이용자에게 제공



그림 4.3 등록센서스에 행정자료 제공 관계기관

4.1.3 빅데이터를 이용한 통계 벤치마킹

온라인 물가작성 시스템(매일 50만건의 온라인 가격데이터 개방)[38]

통계청에서는 그동안 소비자물가와 체감물가 간 차이 완화를 위하여 많은 노력을 해왔다. 국민들이 물가에 대한 이해를 쉽게 할 수 있도록 하는 온라인 서비스(우리집 물가 알아보기 프로그램, CPI 소비자 물가지수) 뿐 아니라 물가작성 과정 개선 등이 이에 해당된다. 통계청은 이에 대하여 '빅데이터를 활용한 온라인물가 작성시스템'을 개발하였다. 이 시스템은 온라인상의 가격자료를 일일 단위로 수집하여 물가통계를 작성하는 프로젝트로, 미국 MIT에서 개발된 사례(BPP: Billion Prices Project)를 벤치마킹하여 개발한 것이다. 인터넷사이트에 공개된 가격정보를 수집 분석하여(매일 50만건의 온라인 가격데이터) 일일 물가 통계를 작성하며, 안전행정부의 '빅데이터 공통기반 및 시범과제 구축 사업'의 일환으로 통계청, 안전행정부가 공동 개발한 시스템이다(그림 4.4 참조).

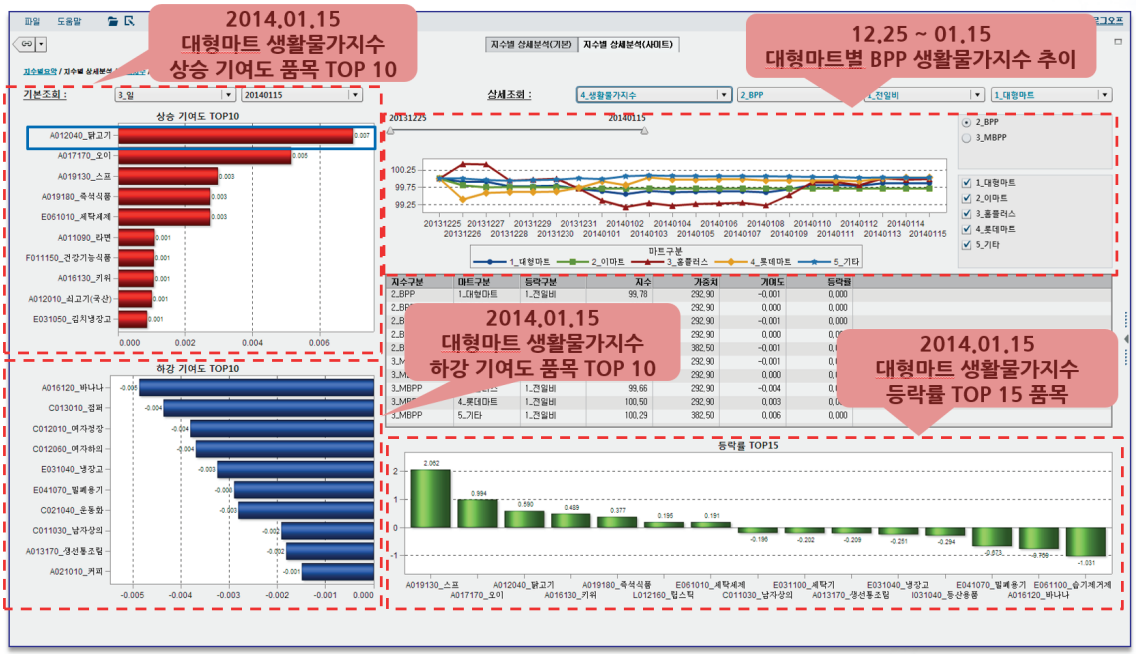


그림 4.4 빅데이터를 활용한 온라인물가 작성시스템

통계청은 본 시스템을 내부적으로 시험 운영하며 일일 단위로 수집되는 물가자료를 지속적으로 모니터링하고, 향후 일일물가지수를 개발하여 업무에 활용

38) 이 내용은 통계청 2014년 2월 17일 보도자료의 내용을 옮겨오거나 발췌하여 정리한 것임

할 예정이다. 본 시스템은 온라인에서 수집 가능한 품목을 대상으로 방대한 가격정보를 매일 수집하여 소비자물가에 대한 신속한 모니터링과 예측을 수행할 수 있으며, 향후 축적된 수집 정보를 활용하여 일일물가지수를 개발하고, 시의성 있는 물가관련 보조자료를 생산·제공함으로써 신속한 정책수립 지원이 가능할 것으로 기대된다.

4.2 중앙행정기관별 활용현황[39)]

2011년 11월 대통령 직속의 국가정보화전략위원회는 '빅데이터를 활용한 스마트 정부 구현(안)'을 마련하였다. 특히 교육과학기술부, 행정자치부, 지식경제부, 방송통신위원회, 국가과학기술위원회와 협의하여 스마트 국가 구현을 위한 빅데이터 마스터플랜을 수립하였다. 6개 분야에서 대상 과제 16개를 제시하였고, 유용성을 고려하여 3개 과제를 2013년도에 우선 추진한 후 2017년 까지 수행할 단계적 활용 계획을 수립하였다. 6개 분야는 사회 안전, 국민 복지, 국가 경제, 국가 인프라, 산업 지원, 과학기술이며 대상 과제별로 관련 주관 부처를 선정하였다.

2013년 말에는 관계 부처 합동으로 '빅데이터 산업 발전 전략'을 발표하였다. 이 전략에서는 ICT 인프라 강국에 이은 빅데이터 활용 강국으로의 도약이라는 중장기 비전 아래 수요 측면에서 초기 시장 창출 및 확대, 공급 측면에서 산업 육성 기반 확충, 인프라 측면에서 지속 발전 가능한 데이터 생태계의 조성을 중점 추진 과제에 포함시켰다.

향후 빅데이터 관련 추진 전략은 제조, 건강, 기상, 스포츠, 재난, 유통, 금융 등 유망 업종에서 빅데이터 선도 프로젝트를 추진 (3년간 350억 원, 민간 매칭)하고, 빅데이터로 시민 체감형 도시 문제 (교통, 주차 등) 해결 시범 사업을 추진(2017년까지 6곳)한다. 또한 빅데이터 산업 활성화와 개인정보 보호가 조화되도록 개인정보 보호법 등 관련 법제 개정을 추진할 것이다(실제 2016년 6월 30일 행정자치부, 방송통신위원회, 금융위원회, 미래창조과학부, 보건복지부, 국무조정실 등 정부 부처가 합동으로 마련한 `개인정보 비식별화 조치 가이드라인`을 발표하였다. 이에 대한 내용은 추가적으로 다루기로 한다).

1. 범정부 빅데이터 추진체계

박근혜 정부 출범 이후 행정안전부가 안전행정부(현행 행정자치부)로 바뀌면서 정보화전략실에서 담당하던 빅데이터는 직제 개편으로 창조정부조직실 공

39) 2015년도 국가정보화 백서, 빅데이터의 국가통계 활용을 위한 기초연구(통계청)의 내용을 옮겨오거나 발췌하여 정리한 것임

공정보정책과에서 담당하면서 정부3.0 과제로 전환하였다. 빅데이터 업무는 크게 행정자치부와 미래창조과학부에서 진행해왔다. 두 부처는 각각 정부차원의 빅데이터 활용에 관한 업무와 민간차원의 빅데이터 활용에 관한 업무를 나누어 진행하는 추진체계를 갖는다.

정부차원의 빅데이터 활용에 관한 전반적인 업무는 행정자치부에서 추진하며, 대용량의 정형 또는 비정형 데이터 셋에 관한 제도·정책 및 기본계획의 수립·시행, 빅데이터 기반 구축·운영 및 활용에 관한 업무를 총괄한다.

민간차원의 빅데이터 활용에 관한 업무는 미래창조과학부에서 추진하며, 빅데이터 활성화 정책 수립·추진, 관련 정책·제도의 연구·기획 및 조정, 상호운용성·표준화·품질관리·유통에 관한 사항, 정부·공공·민간 협력에 관한 업무를 담당한다.

2. 빅데이터 공통기반 플랫폼[40)]

부처별 별도의 시스템 구축 없이 빅데이터 인프라를 통해 공공데이터와 민간데이터를 융합하여 분석 서비스를 제공하는 것을 빅데이터 공통기반 플랫폼이라 한다. 범정부적으로 빅데이터를 공유·분석할 수 있는 체계와 이를 활용한 표준모델을 제공하고, 빅데이터 활용 수요 증가로 인한 각 부처별 중복 투자문제를 해결하기 위하여 범정부 차원의 빅데이터 분석 플랫폼을 구축('13.12월, 행정자치부) 하였다.

공통기반 플랫폼은 연계·수집 시스템으로써 행정정보공동이용센터를 통한 공공데이터 수집(ESB Agent 활용), 공공데이터포털 자료 연계(Open API 활용)와 SNS, 블로그, 카페, 뉴스 등 Web 상의 민간 공개데이터 수집(민간데이터 수집기 활용 및 구매)등 민간분야의 데이터를 수집·연계하는 역할을 한다.

40) 혜안에 관한 내용은 빅데이터 공통기반 가이드북 1권을 내용을 옮겨오거나 발췌하여 정리한 것임

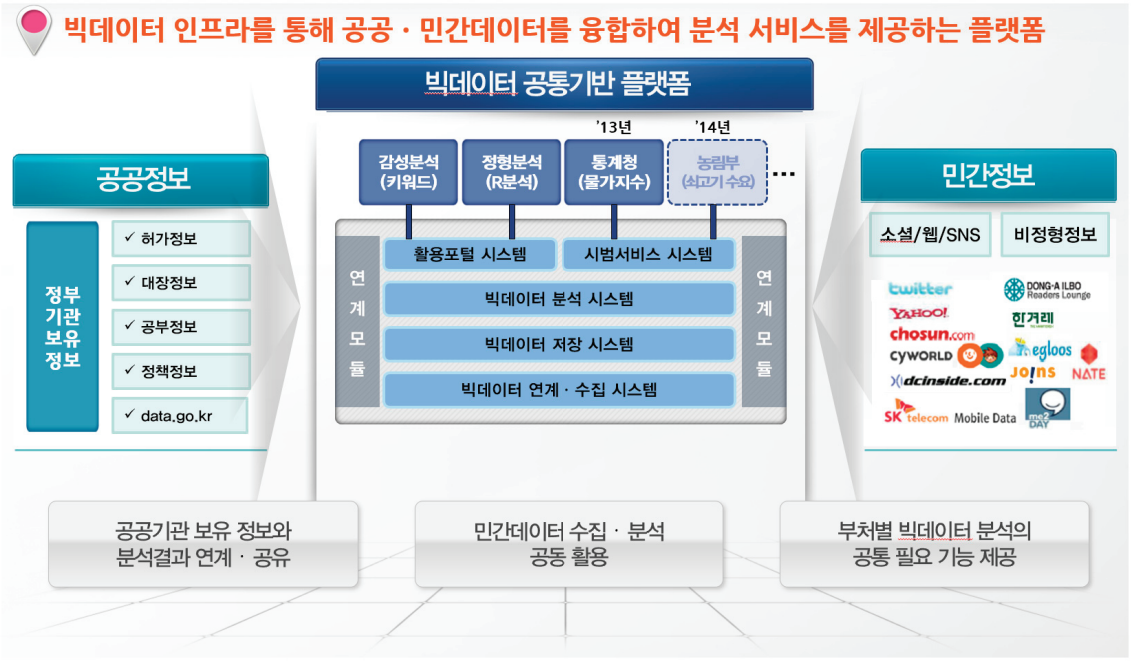


그림 4.5 빅데이터 공통기반 플랫폼 구성

공통기반 플랫폼은 저장시스템, 분석시스템, 활용포털시스템인 혜안(http://www.insight.go.kr)으로 구성되며(그림 4.6 참조) 행정업무망 내에서 빅데이터 활용포털사이트에 접속, GPKI 인증서를 통해 회원가입 후 활용한다. 공통기반 플랫폼의 구성은 그림 4.7과 표 4.7을 참조하기 바란다.

표 4.7 빅데이터 공통기반 플랫폼의 구성과 역할

구성	역할
저장시스템	Hadoop Ecosystem(R, Mahout, Hive, RHIVE, Hbase)을 통하여 빅데이터 저장환경을 구성하고, 수집된 데이터 및 분석결과를 분산 파일시스템에 저장·관리하는 역할
분석시스템	수집된 정보를 활용 텍스트마이닝 등 빅데이터 분석기반을 제공하는 역할
활용포털시스템(혜안)	이용자가 빅데이터 수집·분석 요청 및 결과 등을 이용할 수 있는 단일창구 역할



그림 4.6 빅데이터 공통기반 활용포털 혜안의 초기화면

출처 : 빅데이터 공통기반 가이드북 1권

빅데이터 공통기반에 적용된 전문분석 도구가 크롬(Chrome) 브라우저에 최적화되어있어 혜안 포털에 접속하기 위해서는 크롬브라우저를 기본으로 사용해야 한다.

혜안 포털은 공회원관리, 커뮤니티 등의 포털 기본 기능과 서비스 소개, 테마분석, 서비스 신청, 데이터 검색, 커뮤니티, 전문분석 6개의 메인 메뉴와 15개의 서비스로 구성된다(그림 4.7 참조). 각 메뉴의 내용은 다음과 같다.

○ 서비스 소개

혜안 포털에서 제공하는 서비스들에 대한 간략한 소개를 제공하는 화면으로 빅데이터 공통기반과 혜안 포털 서비스를 제공하기 위한 범정부 추진 체계에 대한 소개를 한다.

○ 테마분석

사용자의 관심 키워드에 대한 기본적인 소셜 분석과 실시간 언론스크랩 내용을 확인할 수 있다.

- 키워드 분석 : 실시간으로 수집되는 웹소셜데이터(뉴스, 블로그, 트위터)에 대한 기본적인 키워드 분석 기능을 제공하는 서비스, 관심 키워드를 입력 후 검색하여 감성분석, 연관검색어 분석 등의 결과 제공
- 언론스크랩 : 행정자치부관련 키워드들을 기반으로 실시간으로 수집한 언론 기사들에 대한 기사 출처, 기사 제목, 수집시간, 감성분석 결과 등을 가장 최근 기사부터 게시판 형태로 보여 준다
- 이동인구분석 : 이동인구분석은 빅데이터 공통기반 사업 수행 과정에서 확보한 유동인구 데이터를 기반으로 지역별, 시간대별, 연령별, 성별 등으로 유동인구 분포를 확인할 수 있다

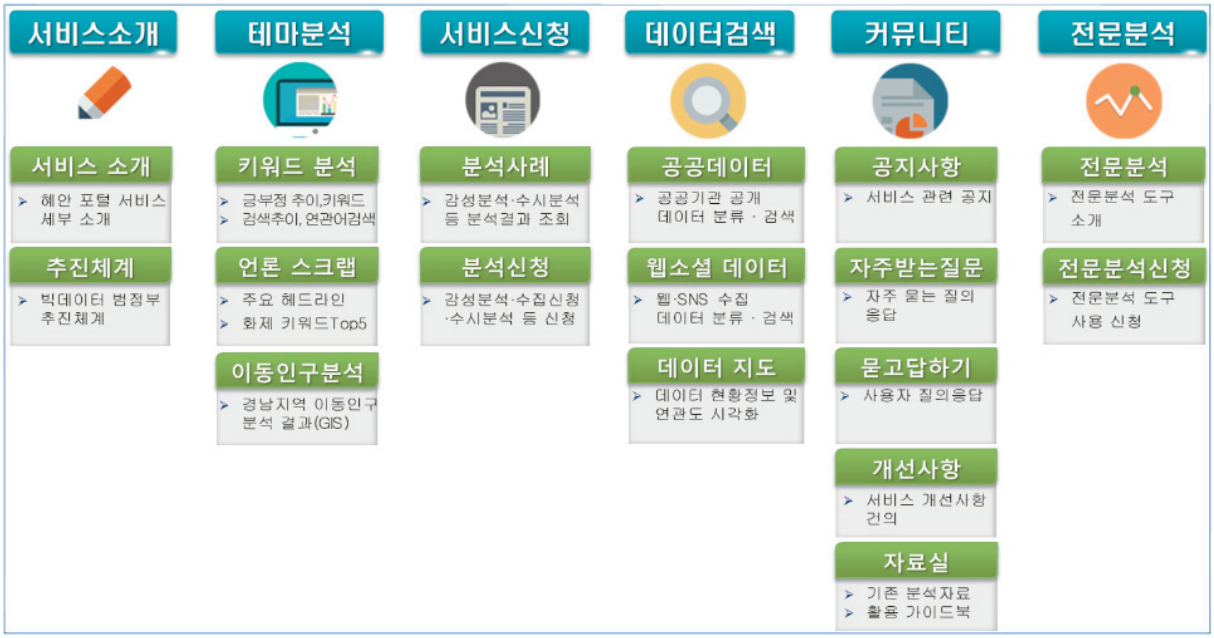


그림 4.7 혜안 포털의 메인 메뉴와 서비스

○ 서비스 신청

공통기반을 이용한 분석 사례 결과 조회와 공통기반에서 제공하는 분석 서비스에 대한 신청을 수행한다.

- 분석사례 : 빅데이터 공통기반을 활용하여 기존에 진행했던 수시분석 및 시범과제의 내용과 결과물을 조회할 수 있다.
- 분석신청 : 분석 신청 서비스에서는 감성분석, 수집요청, 수시분석, 분석과제 4가지 유형의 분석 서비스가 있다(표 4.8 참조). 수시분석 및 분석과제를 신청할 때에는 신청정보, 분석개요, 활용계획, 분석데이터(자체 보유, 타기관 보유, 웹소셜데이터), 기대효과 등을 입력하여야 한다.

표 4.8 분석신청 유형별 정리

신청유형	내용 설명
감성분석	- 혜안 포털의 테마분석 기능보다 심도 있는 감성분석이 필요할 경우에 신청 - 웹소셜데이터 분석에 한함 - 신청 내용이 승인될 경우, 플랫폼운영계에서 데이터 수집 및 감성 분석 진행 후 결과물을 제공
수집요청	- 현재 혜안 포털에서 수집하고 있지 않은 특정한 사이트의 데이터를 수집할 필요가 있을 경우에 신청 - 혜안 포털이 수집하고 있지 않은 과거 특정 시점의 웹소셜데이터가 필요한 경우에 신청 - 신청 내용이 승인될 경우, 해당 사이트 또는 기간에 대한 데이터를 수집하여 제공
수시분석	- 신청기관 보유 데이터를 기반으로 타기관 데이터와 웹소셜데이터 등을 결합한 전문분석이 필요할 경우에 신청 - 신청 내용이 승인될 경우, 데이터분석계 전문 분석인력이 해당 주제에 대한 분석 작업 수행 후 결과물을 제공
분석과제	- 사용할 데이터의 범위와 분석 영역이 수시분석보다 더 넓고 전문적이며, 별도의 서비스 시스템 구축 등이 필요한 경우에 신청 - 신청 내용이 승인될 경우, 행정자치부 시범사업 주관부서인 공공정보정책과에 이관하여 시범사업의 분석 주제로 포함할지 여부를 검토

출처 : 빅데이터 공통기반 가이드북 1권

○ 데이터 검색

공공데이터 및 소셜 데이터에 대한 검색 및 조회, 다운로드 기능을 제공한다.

- 데이터 분류 체계 및 수집 원천 : 공공데이터(공공데이터 포털의 16개 분류를 1레벨로 정부기능 표준분류체계 67개를 2레벨로 매핑하여 분류)는 공공데이터포털과 연계하여 제공, 웹소셜데이터(자동분류기술에 의한 16개

분류)는 뉴스, 블로그, 트위터 및 각 분야별 주요 정보 사이트의 데이터를 수집하여 제공

- 공공데이터 : 공공데이터는 16개의 주제별 분류기준을 선택하여 데이터를 조회하거나, 검색어 기반으로 데이터를 조회 가능하며, 검색어 기반 조회 시에는 분류기준이나 검색기간들을 설정할 수 있으며, 동의어와 배제어를 입력하여 보다 상세하고 정확한 검색을 수행하는 것이 가능하다.

- 웹소셜데이터 : 공공데이터 검색과 같은 방식으로 분류기준을 클릭하거나 검색어를 입력하여 데이터 검색이 가능하다

- 데이터지도 : 데이터 지도 서비스는 공공기관에서 생성하여 공개하였거나 보유하고 있는 데이터의 현황 정보를 분류 체계별로 시각화하여 보여주는 서비스로 검색 과정에 유의어 사전을 적용하여 원하는 검색어와 유사한 의미를 가지는 모든 데이터를 한 번에 찾을 수 있도록 지원한다.(그림 4.8 참조)

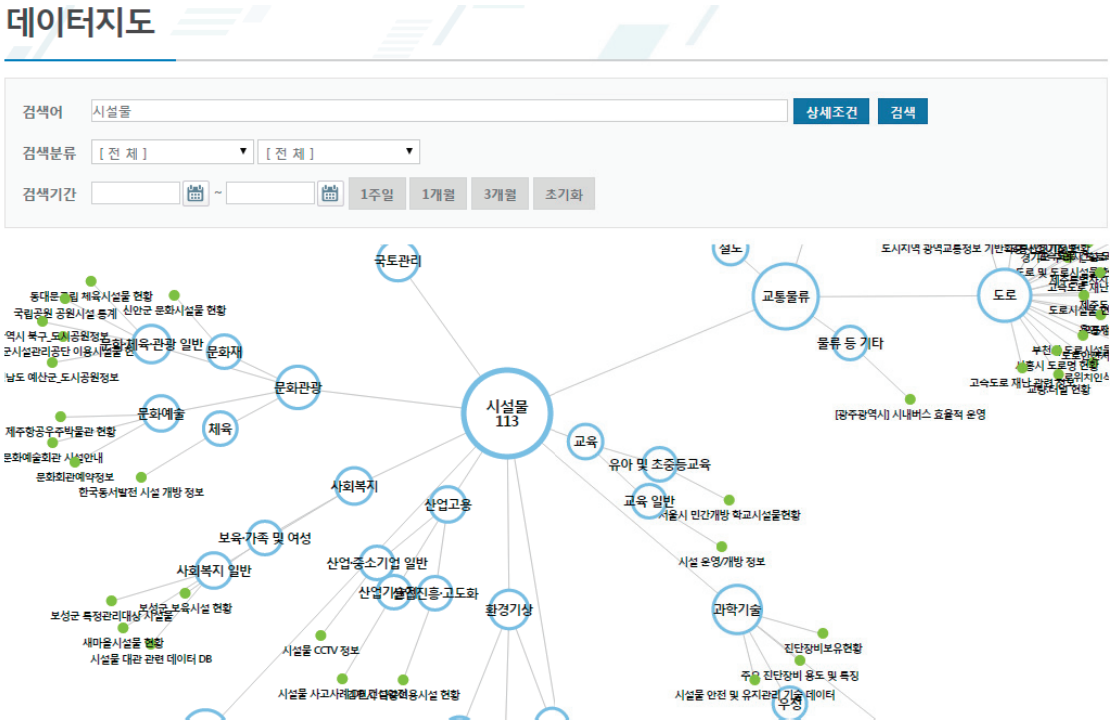


그림 4.8 데이터지도 검색 결과 화면

출처 : 빅데이터 공통기반 가이드북 1권

○ 커뮤니티

공지사항, Q&A, FAQ, 개선사항, 자료실 등 5개의 화면으로 구성되어있으며, 혜안 포털 사용자들에 대한 정보 제공, 불편 해결, 서비스 개선 등의 목적으로 하고 있다.

○ 전문분석

사용자가 빅데이터 공통기반 내에서 데이터 처리/분석/시각화 등을 직접 수행할 수 있는 전문분석 도구 소개 및 사용 신청 서비스 제공한다. 모두 3가지의 전문분석 도구를 제공한다(표 4.9 참조). 전문분석 도구를 활용하기 위해서는 별도의 신청 과정을 거쳐야 한다.

표 4.9 전문분석 도구 설명

도구명	사용분야	설명
Hadoop 2.X 플랫폼 (SBP)	데이터 저장/처리	- Hadoop 2.0 기반의 Big Data 플랫폼 - GUI기반 Big Query, Workflow 등 지원 - Hadoop 인프라 관리 및 모니터링 기능
R-GUI 솔루션 (SRA)	데이터 분석	- R기반 데이터 분석을 GUI 환경으로 제공 - 분석 프로세스에 대한 Workflow화 가능 - Hadoop 데이터 연계 분석 가능
시각화 솔루션 (Tableau)	분석 시각화	- 다양한 형태의 차트를 통한 분석 시각화 - Hadoop 데이터 및 R분석 모듈 연동

출처 : 빅데이터 공통기반 가이드북 1권

1) 공통기반 플랫폼 활용 사례(시범사업 지원과제)

〈 2013년, 1건 〉

1 물가지수 분석을 위한 시범 사업(통계청)

- (분석데이터) 20개 온라인사이트의 295개 품목에 대한 가격정보
- (분석 내용) 물가 변동요인 파악 등 물가관련 통계 분석·검증 및 트렌드 분석 ➡ 기존 월단위 물가정보에 대한 분석결과를 일단위로 단축

〈 2014년, 5건 〉

1 개인정보 보호체계 운영기반 구축(안행부)

- (분석데이터) 인터넷진흥원(개인정보침해 신고정보), 통신사·은행 등 기관 자료(취급데이터·제공기관 등), 안행부(개인정보 노출 점검업체·노출내용) 등
- (분석 내용) 개인정보 노출 위험도가 높은 취약업체를 선정 ➡ 취약분야 기획점검, 침해·유출사고 예방교육 등 사전조치

2 도로위험상황 예보 시스템 구축(경찰청, 도로교통공단)

- (분석데이터) 경찰청(사고·집회·공사 등 돌발정보), 도로교통공단(급경사, 결빙구간·안개구간, 교통사고·소통), 기상청(기상상황 정보) 등
- (분석내용) 지역별·시간별 사고 패턴 등 실시간 위험정보를 도로상황판 및 네비게이션으로 운전자에게 제공 ➡ 위험지역 운전 회피 · 주의로 교통사고율 감소

3 의약품 부작용 분석(식약처)

- (분석데이터) 식약처(SNS, 의약품 부작용 신고정보, 의약품 성분·주의사항), 건강보험공단(요양급여청구자료, 병원진료기록)
- (분석내용) 각종 의약품 부작용 정보 간 상관관계 분석 ➡ 의약품 표시정보 개선, 처방·복용시 활용하여 의약품 부작용 최소화

4 빅데이터 기반 산불 위험예보(산림청)

- (분석데이터) 산림청(임상도 종류·밀도), 기상청(산악습도·강수확률·풍속·풍향)
- (분석 내용) 산불 피해 예상지역 및 위험도를 실시간으로 파악 ➡ 순찰 우선순위 선정 등 예방활동 수행을 통한 산불피해 최소화

5 스마트 재난상황실(스마트 빅보드) 구축(재난안전연구원)

- (분석데이터) 기상청(기상정보), 재난연구원(침수예상도, 태풍정보), 경찰청(CCTV), 방재청(방재자원도, 급경사지 위험정보), 민간(SNS, NEWS, 방송, 휴대폰 현장 영상) 등
- (분석 내용) 침수지역 등 위험상황 사전 예측 ➡ 재난 위험정보의 실시간 공유·분석·전파를 통한 신속 대처

2) 공통기반 플랫폼 활용 사례(수시분석 과제)

<전략과제>

1 시내버스 효율적 운영(광주광역시)

- (분석데이터) 교통카드 이력데이터(173,388,169건), 버스운행이력데이터(2,755,106건), 버스정류장(2,632건), 교통약자시설(590건), 주거인구데이터(4,717,776건)
- (분석내용) 대중교통 사각지대 발굴(26건), 환승시설물 위치선정(50개), 탄력배차 우선순위 선정(23개), 저상버스 우선도입 선정(30개)

<일반과제>

1 법령정보센터 웹로그 분석(법제처)

- (분석방법) 시차연관성 분석, SNA 분석(법령에 대한 주요 조회 패턴 및 법령 간의 연관성 분석 목적)
- (분석내용) 사용자 그룹별 SNA 그래프 생성 후, 법령 조회 패턴들의 연결망 군집 추출(Light 그룹 7개, heavy 그룹 15개)

3. 빅데이터 분석활용센터

2013년 미래창조과학부와 한국정보화진흥원(https://www.nia.or.kr/)은 공공·민간의 빅데이터 서비스 도입을 촉진하고, 데이터 기반 창조경제 및 정부3.0 실현을 위하여 빅데이터 분석활용센터를 개소하였다.

빅데이터 분석활용센터에서는 빅데이터 분석활용 및 교육실습을 위한 분석 인프라(서버, 솔루션, 테스트 데이터 등)를 구축, 개인, 회사, 연구소 등에 실증·분석환경 제공한다. 원격에서 이용자가 센터에 접속하여 솔루션 개발 및 교육 컨텐츠를 활용할 수 있도록 웹기반으로 분석환경 제공(http://www.kbig.kr)하고, 이용자는 회원가입 후 사전 협의된 자원을 스케줄에 따라 활용할 수 있다(그림 4.9 참조). 분석인프라 사용절차는 그림 4.10을 참고하기 바란다.



그림 4.9 빅데이터 센터 초기화면



그림 4.10 빅데이터 분석인프라 사용절차

분석인프라를 이용하여 데이터 분석, 워크플로우 관리, 시각화 등을 위해 하둡 관련 오픈소스 S/W(하둡(Hadoop), 하이브(Hive), 피그(Pig), 머하웃(Mahout), R, 플룸(Flume), 스쿱(Sqoop) 등)를 사용하는 것이 가능하다.

분석기능 이외에서 빅데이터 분석활용센터에서는 국내외 빅데이터 관련 사

업, 사이트와의 협조 및 자체수집을 통해 활용도가 높은 데이터를 수집하여 공개하고 있다. 제공 정형 데이터는 저작권 보호를 위해 빅데이터 분석활용센터 내의 인프라에서만 사용 가능하며, 국토·교통(12종), IT서비스(9종), 산업·경제(8종), 방송·통신(4종), 의료·보건(4종), 일반행정(3종), 교육(2종), 통계(2종), 기상·환경(1종) 등 총 45종임(2015.5월말 기준). 비정형 데이터로 미래전략라이브러리를 구축, 정책기술 동향, 시장동향, 수출정보, R&D 정보, 우수기술 정보 등 미래전략 수립에 활용할 수 있는 정보를 각종 보고서, 논문, 정책·연구보고서 형태의 자료로 수집하고 공유·보관하여 사용자에게 파일로 제공하고 있다.

그 밖에도 빅데이터 분석활용센터에서는 빅데이터 관련 노하우를 공공·민간에 기술전수하고, 애로사항에 대한 기술자문 등을 빅데이터 전문인력 풀을 구성하여 지원하고 있다. 전문인력 풀은 대학교, 기업, 공공기관 등 각 분야의 전문 인력이 분석/통계/마이닝(11명), 플랫폼/관리기술(7명), 아키텍처/클라우드 설계(5명), 데이터 시각화(2명)으로 구성되어있다(2015.5월말 기준). 이 서비스를 활용하기위한 절차는 그림 4.11을 참고하기 바란다.

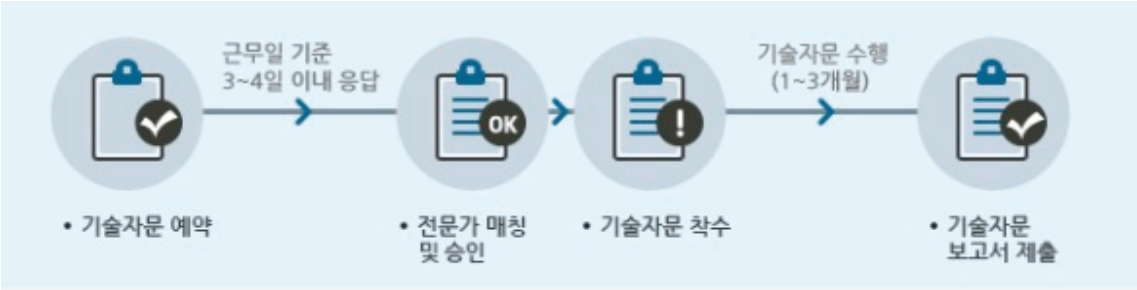


그림 4.11 빅데이터 기술자문 진행절차

지금까지 중앙행정기관이 빅데이터 활용을 위하여 어떤 일들을 해왔는지를 알아보았다. 이제 사례들을 통해 구체적인 결과들을 살펴보도록 하자.

사례 1 - 특수질병 및 질환 예방 및 관리(보건복지부, 국립보건연구원)[41)]

가. 추진 배경

○ 보건 의료 산업의 핵심 재료로 인체 자원이 급부상
- 질병 지표 발굴 및 조기 진단을 위해 대규모 인체 자원이 활용됨에 따라 질병예방 및 맞춤 치료 등 삶의 질 향상에 기여하기 위해 인체 자원 정보를 공유

나. 추진 내용

○ 수요자 맞춤형 분양시스템 구축
- 1기 사업에서 수집한 50 만 명 분의 병원별로 특성화된 질환 집단에 속하는 인체 자원을 관리, 이를 활용할 연구자에게 적극 분양하기 위해 One-Stop 분양이 가능한 온라인 분양 포털 마련(그림 4.12)
- 2013년에 2기 사업으로 진행됨

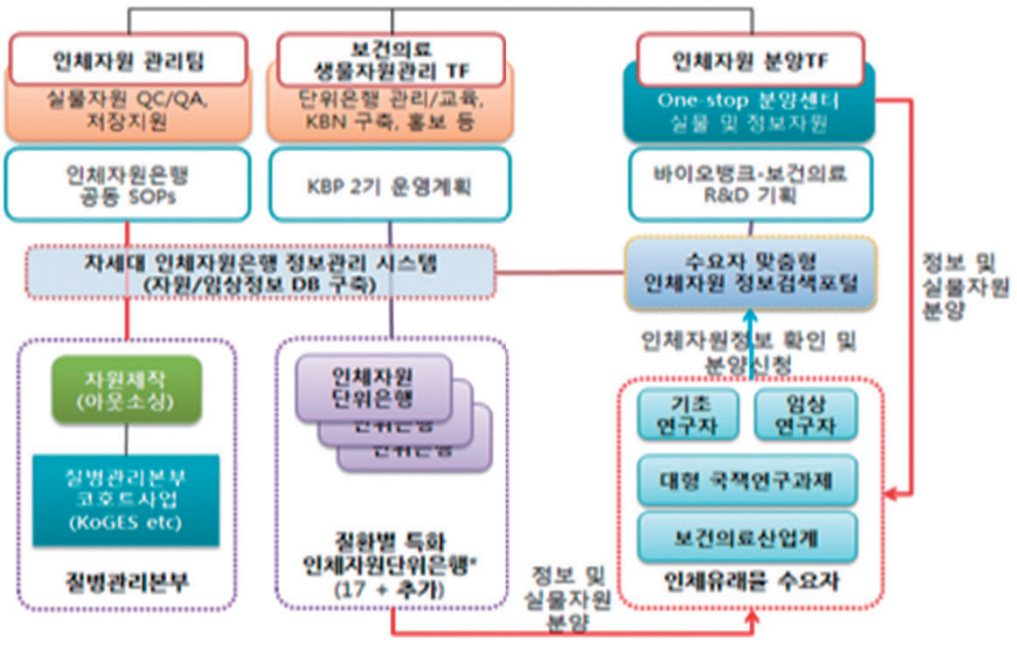


그림 4.12 One-Stop 분양 포털 추진 체계

출처 : 보건복지부 2013

41) 이 후 사례들에 관한 내용은 '창조경제 실현을 위한 2013 빅데이터 국내 사례집'의 내용을 옮겨오거나 발췌하여 정리한 것임

○ 관계 부처별 유전체 자원 보유, 활용현황을‘ 국가 생명연구자원 정보센터’(KOBIC) 에서 중점관리
- 2021년까지 10만 명의 한국인 유전체 자원을 확보, 인체자원연구지원센터를 통해 맞춤의료, 예방치료 등 보건의료 신 성장 분야 지원

○ 인체자원의 품질 표준화를 위한 운영 프로토콜 기반 관리
- 단위은행은 인체 자원의 품질 표준화를 위해 표준화된 은행 운영 프로토콜(규약)을 기반으로 인체 자원을 수집, 보관 및 분양
- 수집된 자원을 기반으로 다양한 질병연구, 바이오마커 개발, 정도관리지표물질 개발 등 인체 자원 활용화 연구 수행



그림 4.13 인체자원은행 전국현황(1개 중앙은행, 17개 단위은행)

출처 : 보건복지부 2013

다. 성과 및 시사점

○ 국가 주도의 보건 의료 분야 생명 연구 자원의 글로벌 경쟁력 기반 확보
- 생명 연구 자원의 체계적 수집과 정보 표준화, 통합 네트워크 구축 가능 (그림 4.13)

- 질병관리본부, 단위은행 뿐만 아니라 민간 인체자원은행의 보유자원까지 표준화하여 보관, 활용하는 표준화 관리 정책 개발 도입 가능

○ 미래 바이오산업의 신성장 동력으로 자리매김 가능
- 연구자가 요구하는 자원을 분양받을 수 있도록 파생 시료를 공급하고, 고부가가치 제품 개발에 필요한 인체 자원 확보와 연계 확대

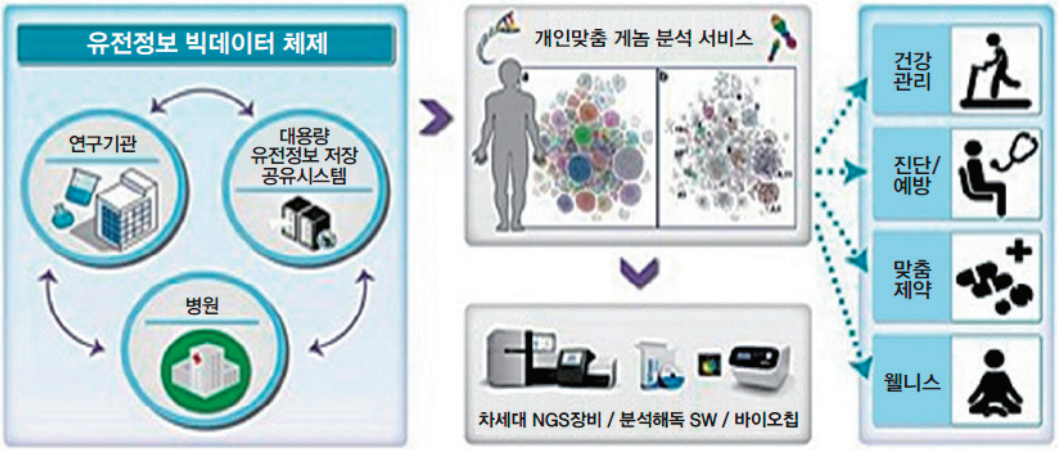


그림 4.14 유전정보 빅데이터 시스템 흐름도

출처 : 쿠키뉴스, 2013

사례2 - 통합형, 재난안전 정보 네트워크, 스마트 빅 보드(안전행정부, 재난안전연구원)

가. 추진 배경

○ 재난 상황에 대한 통합 데이터가 부족
- 그간 재난·안전사고 발생 시 주로 문서나 유선을 통한 상황 보고에 그친 현장 정보와 언론사나 지자체에서 제공된 영상 정보를 수동적으로 이용함
- 재난상황을 실시간으로 파악하고 능동적으로 분석하여 국민들의 신속한 안전 확보 에 도움을 주기 위한 체계적인 서비스에 대한 필요가 꾸준히 제기됨

나. 추진 내용

○ 스마트 빅 보드(재난 및 안전사고 시 현장 중심의 모든 가용한 정보 네트워크를 가동하여 위험상황을 분석하는 최첨단 스마트 재난 상황실)를 통해

기상, 재난이력 및 국내외 재난정보의 정보 공유를 통해 재난현장 상황 정보 취득 가능(그림 4.15, 4-16)

- CCTV·무인항공기·스마트폰·인공위성영상 등 다양한 첨단 기술을 통해 지상 은 물론 항공에서 제공되는 생생하고 입체적인 재난현장 상황 조회 가능
- 종합적인 원스톱 재난상황 모니터링 : 위치기반 SNS 정보 및 트위터 등 재난현장 에서 제공되는 다양한 정보를 이용하여 신속한 대응이 가능
- 빅데이터 정보 : 재난관리에 활용성이 높은 실시간 트윗정보, 과거 재난이력 및 원인분석결과, 위성영상, 시뮬레이션, 관련 웹사이트 등을 빅데이터 메뉴로 분류하여 예방, 대비, 대응, 복구 재난관리의 전체 프로세스에 과학적이면서도 직관적인 재난관리 수행 지원

○ 상황실 개선 방향

- 기존 재난관리 관련 기관에서 개별적으로 관리·서비스하고 있는 정보를 연계하여 하나의 프레임 안에서 위치 기반 서비스 (LBS : Location Based Service) 형태 로 표출하여 다양한 유형의 재난에 신속하게 대응(표 4.10)

표 4.10 상황실 개선 방향

구분	현재	개선
정보형태	기상, CCTV 등	지능형 CCTV, SNS, 재난지도 등
정보분석	경험적 위험성 판단	시뮬레이션 등 과학적 분석
위험경보	기상청 등 단순정보 의존	지역특성을 고려한 스마트 재난경보
현장경보	TV, 뉴스 등 언론정보	위성, UAV, MMS(이동차량) 등
정보통합	각 부처별 재난정보 관리	부처간 정보통합 구축
정보표출	보고서 방식의 단순 정보표출	위치기반 전자지도 표출
민관협력	정부중심의 일방적 대응	SNS 등 국민 ↔ 정부 간 쌍방향 대응

출처: 안전행정부, 2013

다. 성과 및 시사점

○ 다양한 유형의 재난에 신속 정확하게 대응

- 상황 판단에 필요한 정보를 신속하게 제공하고 최적의 상황대응 의사결정을 하는데 도움을 줌

  예) 2013년 4월 산대저수지 붕괴사고와 7월 아시아나 항공기 사고 시 언론보다 먼저 위험상황 전파가 이루어졌던 사례는 스마트폰 3천 5백만 대

시대의 미래형 재난관리의 청사진을 제시(국립재난안전연구원, 2013)

○ 국민생활안전지도의 제작·보급에 도움

- 재난 상황을 전자 지도로 표출하고, 재난 예측 및 피해 시뮬레이션 결과를 바탕으로 과학적 재난 의사 결정 지원 가능

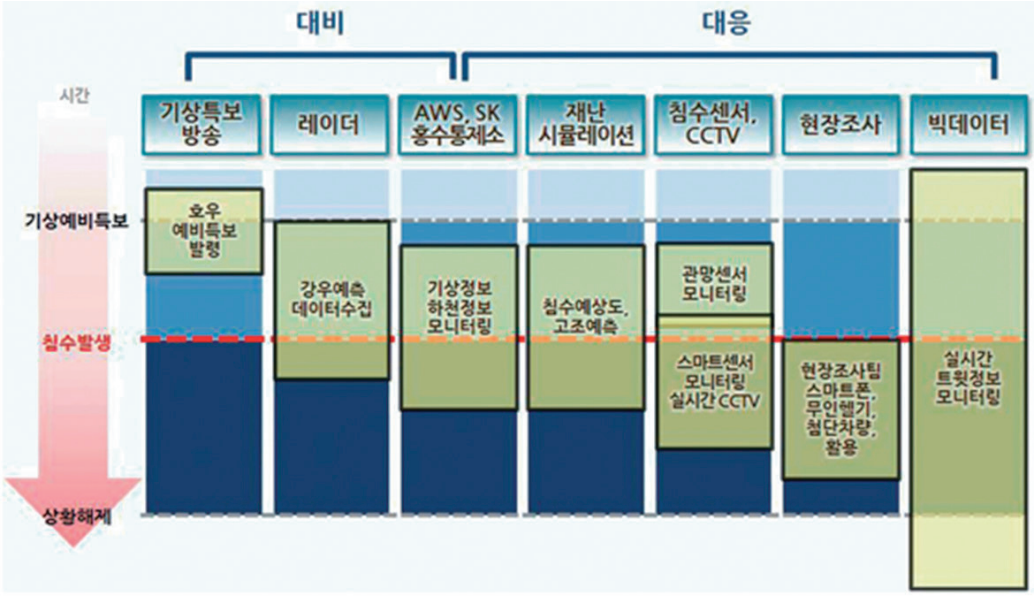


그림 4.15 스마트 빅 보드 운영 시나리오

출처 : 최우정 외, 2013



그림 4.16 스마트 빅 보드 메인화면

출처 : 최우정 외, 2013

사례3- 지리정보 통합한 지리적 프로파일링 시스템 구축(GeoPros, 경찰청)

가. 추진 배경

○ 지리적 프로파일링 시스템(다양한 공간통계 분석기법을 경찰의 범죄수사데이터에 적용, 범죄위험지역 예측을 통한 방법 전략 수립 및 연쇄 범죄자 거주지 예측을 통한 수사 활동 전개가 가능한 시스템(경찰청, 2013)) 개발의 필요성 대두

- 방대한 범죄수사 자료가 축적되고 수사기법이 발전되어 왔으나, 한국 실정에 맞는 지형공식이 없어 범죄수사의 범위를 줄여줄 지리적 프로파일링 시스템 개발의 필요성이 생김

나. 추진 내용

○ 범죄 유형에 따른 범죄 위험 지역을 한 눈에 알아보게끔 시각화

- 각종 범죄의 유형에 따라 지역 및 시간대별로 범죄다발지역과 위험도를 지도에 각기 다른 색으로 표시(그림 4.17, 범죄 밀집 정도에 따라 색상변화 → 붉은 곳이 범죄가 가장 많이 발생하는 지역)

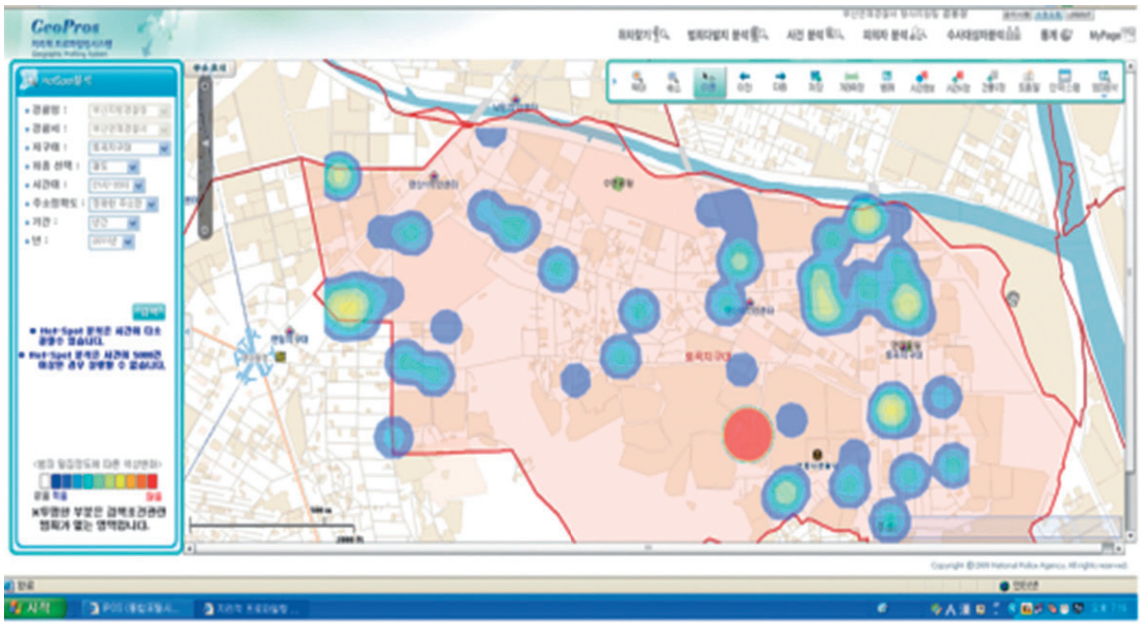


그림 4.17 지리적 프로파일링 시스템 활용 범죄 다발지역 분석 예시

출처 : 부산연제경찰서, 2012

○ 추가 범죄 가능성을 예측하여 우선 수사 및 예방
- 기존 발생한 범죄의 데이터를 통합 수집 및 분석하여 추가 범죄의 가능성과 방향성을 찾고 우선 수사 대상자 및 대상 지역 선별
- 연쇄 범죄의 발생 가능성이 높은 장소와 시간을 예측하고 용의자가 거주할 만한 지역을 예측
- 범죄위험도에 따라 순찰차를 배치하고 순찰 경로를 선정
- 성폭력 등 범죄에 대한 지역별, 시기별 맞춤형 치안대책 수립

다. 성과 및 시사점

○ 강력 범죄 발생의 감소 및 예방
- 지리적 프로파일링 시스템을 통해 분석한 자료를 토대로 2012년 8월 1개월간 범죄 다발지(5대 범죄) 중심 형사기동차량순찰 및 필요지점에 대한 거점근무를 실시한 결과, 해당 범죄가 전년 동기간보다 10.5% 감소(전년 동기간 296건, 2012년 8월 265건) (부산연제경찰서, 2012)

○ 순찰 현장의 대응력 강화로 주민 안전 고취
- 향후 112 신고출동시스템 및 순찰차 내비게이션과 연계하여 시간별, 지역별 범죄 위험지역을 순찰차에서 바로 확인하여 전략적 순찰활동 전개 가능 (안전행정부, 2013)

○ 주민의 관심도를 높이고 주민 스스로의 범죄 예방 유도
- 범죄예방지도 등 생활치안 정보를 생산하여 국민에게 공개함으로써, 지역 치안에 대한 관심을 높이고 범죄를 예방하는 효과 기대(안전행정부, 2013)

사례4 - 국가공간정보 통합데이터베이스 구축(국토교통부)

가. 추진 배경

○ 유비쿼터스, 모바일, 가상공간 등 정보의 융-복합 기술 발전에 따른 공간정보 수요증가

○ 산재해 있던 다양한 분야의 GIS (지리 정보 체계) 정보를 연계 통합 필요성 제기

- 중앙부처, 지방자치단체, 공공기관 등이 토지, 도로, 지하시설물, 통계 등의 정보를 각각 구축, 운영함에 따라 상호 호환성이 떨어지고, 중복으로 개발하여 예산이 낭비되고 시스템 간 연계체계의 미비로 활용도가 떨어짐
- 다수의 기관에서 개별적으로 구축, 활용되던 공간정보를 연계, 통합하여 행정업무에서 공동으로 활용할 수 있게끔 공간정보통합체계를 구축

나. 추진 내용

○ 중앙부처 공간정보시스템 연계 지속 추진
- 국가공간정보 공동 활용의 통합 관리 기관 역할을 강화하기 위해서 25개 기관 76개 시스템 연계 완료
- 246개(광역 17, 기초 229) 지자체 확산을 통해 국가공간정보 통합체계 정착 완료(국토교통부, 2013)

○ 민간 활용수요가 높은 공간정보의 공개
- 국가공간정보센터를 통해 행정정보, 수자원/해양, 환경, 산림, 보전지역, 토지 정보, 지형, 지질, 관광/문화 등 9개 분야 31종의 정보 공개에 이어 2013년 7월 택지정보, 도시계획정보, 등산로정보, 사업지구정보, 국가지명, 해안선정보, 국가지명, 교통CCTV, 국가교통정보 등 16종의 공간정보 추가 개방(국토교통부, 2013)
- 포털, 통신사, 내비게이션 회사 등이 영업점 설치, 부동산개발, 길안내 및 지도 서비스 갱신이나 최신 정보 구축에 활용 가능(국토교통부, 2013)

○ 공간 빅데이터의 융합된 데이터베이스를 바탕으로 새로운 가치 창출 가능한 분석모형 개발(그림 4.18, 4-19)
- 3차원 분석, 시계열 분석, 기반시설 입지분석 등
- 부동산, 교통, 안전, 복지 등 구체적인 행정 분야에 공간 빅데이터 플랫폼을 적용 하여 다양한 서비스 개발 및 제공(국토교통부, 2013)

다. 성과 및 시사점

○ 분산 관리 되던 국가공간정보를 통합하여 필요한 국민에게 제공

- 공공기관의 행정·공간정보와 민간 정보 간 통합을 통한 생활 공감형 서비스 마련

○ 보다 정확한 국토계획수립이 가능하게끔 지원

- 국가공간정보 플랫폼에 공간정보와 행정정보를 융합하여, 주택 수급이나 교통 등 인프라 구축 등의 행정서비스의 효율성과 질 향상 가능



그림 4.18 국가공간정보 통합데이터베이스 구축

출처 : 국토교통부 홈페이지



그림 4.19 공간 빅데이터를 활용한 '부동산 임차시장 수요패턴 분석 및 대응방안'

출처 : 국토교통부 홈페이지

사례5 - 역외탈세 정보공유(관세청, 국세청)

가. 추진 배경

○ 탈세 적발의 강화 가능성 발견
- 역외탈세와 불법외환거래는 동시에 발생하는 특징을 보임
- 해외발생 소득, 신고누락에 따른 역외탈세를 적발하는 국세청과 불법외환거래를 파악하는 관세청 간에 일시적, 소극적 정보공유가 아닌 지속적, 적극적 정보 공유를 통해 불법외환거래 및 역외탈세에 대한 단속 효과를 동시에 높이는 것이 가능

나. 추진 내용

○ 두 기관 간 지속적, 적극적 과세 정보 공유
- 관세청과 국세청이 각자 보유 중인 신고자료 등의 과세 정보 교환을 통해 정보 협력 체제 강화
- 국세·관세 탈루 등 혐의 정보를 '외환 거래 감독기관 협의회'를 통해 정기적으로 교환

예) 2013년 9월 현재 관세청은 불법외환거래 조사 중 발견한 역외탈세 등 내국세 탈루 혐의 정보를 국세청에 제공, 국세청은 관세탈루 및 채권미회수 등 불법외환거래 혐의 정보를 제공 해 정밀 분석

다. 성과 및 시사점

○ 행정 효율성과 조세형평 증대
- 인력이나 예산의 추가 투입 없이 정보 공유만으로 행정효율성과 조세의 형평을 증대시키는 정부 부처 간 대표 협업 사례로 떠오름
- 지능적 역외탈세에 효과적으로 대처

○ 세수 증대 및 국고 손실 방지(안전행정부, 2013)
- 탈세 혐의정보 공유의 획기적 증가로 연간 약 900억원 세수증대

- 과세 기초정보 공유의 확대로 연간 약 1,000억원 세수 증대
- 복지부 등 다양한 기관과의 정보 공유 확대를 통해 약 485억 원 상당의 국고 손실 방지

○ 2013년도 〈정부 3.0 우수사례 대상〉으로 선정
○ 해외 협력으로 국제 공조 네트워크 구축으로 발전
- 미국, 영국, 호주와 함께 공동 조사하는 역외탈세정보 공유 네트워크 구축

사례6 - 빅데이터 분석을 통한 2013 국민인식 변화 분석(문화체육관광부)

가. 추진 배경

○ 문화, 체육, 관광 관련 정책 수립에는 간접적인 수요와 연관된 국민들의 여가에 대한 인식 현황 파악이 중요
○ 인터넷 및 스마트폰의 보급과 동시에 최근 국민들의 생활에 변화가 있었을 것으로 예측 가능하나 실제 데이터로 분석한 바가 없어 그 필요성 대두

나. 추진 내용

○ 온라인 뉴스 및 SNS(소셜 네트워크 서비스) 등의 정보를 통해 국민의 라이프 스타일 조사
- 다음소프트와 협력하여 2011년 1월 ~ 2013년 5월까지 29개월 간 트위터, 블로그, 온라인 뉴스 등을 통해 36억여 건의 메시지를 분석

○ 해당 기간 중 언급이 증가한 키워드와 연관어 분석
- 키워드 : 현재, 일상, 퇴근 후, 소소하다, 지르다, 혼자 등 소소한 일상을 소중하고 행복하게 여기고, 작은 물건들을 사며 뿌듯해하며 영화 감상, 여행, 커피 등을 혼자 즐기는 추세도 증가하는 것으로 나타남
- 정책 관련 : 일자리, 사회복지, 무상교육, 영어 조기교육 등 일상생활에 밀접한 연관이 있는 분야에 관심
- 박근혜 정부 4대 국정기조 관련 : 경제부흥, 국민행복, 문화융성, 평화통일 기반구축 중 가장 많이 언급된 키워드는 '국민행복'(48% 차지)(문화체육관광부, 2013)

다. 성과 및 시사점

○ 국민이 정부에 기대하는 요소를 파악함으로써 그에 부응하는 정책 수립 지원 가능

- 국민안전에 대한 관심, 튼튼한 안보에 대한 언급이 많았던 것으로 나타나 정부가 생활안전, 안보 등 최소요건을 담보해주기를 기대하는 것으로 나타남
- 창조경제를 통해 경제에 활력을 주고 일자리를 만들어주기를 기대하며, 일자리 및 부동산 등 삶에 직접 영향을 주는 정책에 대한 관심이 높았음
- 건강 관련 먹거리 안전에 대한 요구, 문화 관련 취약 계층에 대한 문화 대상 확대 등도 관심이 높아 관련 정책의 우선순위 설정에 참고 가능

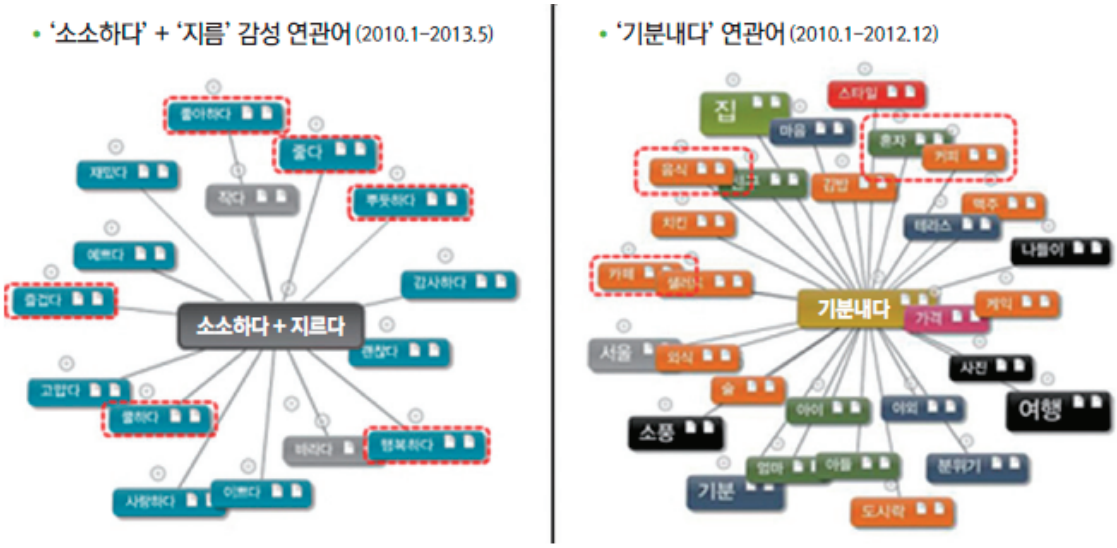


그림 4.20 키워드 연관어 분석

4.3 공공기관별 활용현황

사례1 - 국민건강 주의 예보 시범 서비스 구축

- 주관기관 : 국민건강보험공단
- 참여기관 : 다음소프트

> 어린이집을 운영하는 C씨는 아이들 중 한명이라도 열이 오르면 가슴이 덜컥 내려앉는다. 지난 여름에는 눈병이 유행하여 수많은 엄마들의 원망과 지탄에 시달려야 했다. 결국 비위생적인 어린이집이라는 오명 속에 몇 명의 아이까지 떠나갔다. 그러나 감염병 주의예보 서비스를 알게 되면서 걱정을 덜었다. 유행이 한차례 쓸고 지나간 뒤 대처할 수밖에 없던 전과 달리, 2주전 미리 예보를 접함으로써 미리미리 손씻기, 소독 등으로 대비할 수 있게 되었기 때문이다. 엄마들 한테도 미리 예보내용을 공지하여 주의토록 할 생각이다. 어린이집 이미지도 좋아질 것이다.
>
> 국민건강보험공단 '건강인' 웹사이트(http://hi.nhis.or.kr)에서 제공 예정

가. 추진배경

○ 전염병 유행에 대해 기존 체계는 사후 대처만 가능하여 사회경제적 비용이 발생할 수밖에 없음

○ 빅데이터 분석을 활용하여 예방과 예측 중심의 의료서비스 제공을 시도 함

○ 기상청의 일기예보와 같이 각종 전염성 질환의 위험도를 사전예보할 수 있는 체계를 구축하고자 함

나. 추진 내용

○ 활용데이터

기관명	데이터 명
국민건강보험공단	국민건강정보(진료내역 데이터)
다음소프트	SNS(트위터, 다음 카페) 데이터

○ 주요 개발 내용
- 현재 시점의 주요 유행성 질병에 대한 위험도 예보 및 지역별·연령별 위험도 예보 등 감염병 유행 징후 시 주의 예보 제공
- 독감을 예로 들면, 어떤 지역을 중심으로 환자가 늘고 있는지, 환자들의 연령대는 어떤지, 진료 환자 숫자가 얼마나 늘어나는지, 소셜미디어에서 검색하는 사람들의 추이는 어떤지 등을 파악해 미리 알려주는 방식
- 주요 유행성 질병에 대한 건강보험정보와 소셜미디어 정보를 융합하여 질병 예측 모델을 개발
- 주요 유행성 질병의 위험도와 동향을 한눈에 파악할 수 있는 대시보드 서비스를 제공하고, 지역별 주요 유행성 질병 위험도 정보, 지역 내 질병 관련 진료 동향 및 연령별 진료 현황정보 제공

○ 주요 유행성 질병 진료현황, 과거 진료통계, 질병 동향 및 고위험 지역 정보 제공(그림 4.21)

○ 자연어처리기술을 이용한 텍스트마이닝 기법, 통계분석/기계학습 기술을 이용한 데이터마이닝 기법 등을 활용함(그림 4.22)



그림 4.21 국민건강 주의 예보를 위한 빅데이터 분석 도입 효과
출처 : 한국정보화진흥원 2013

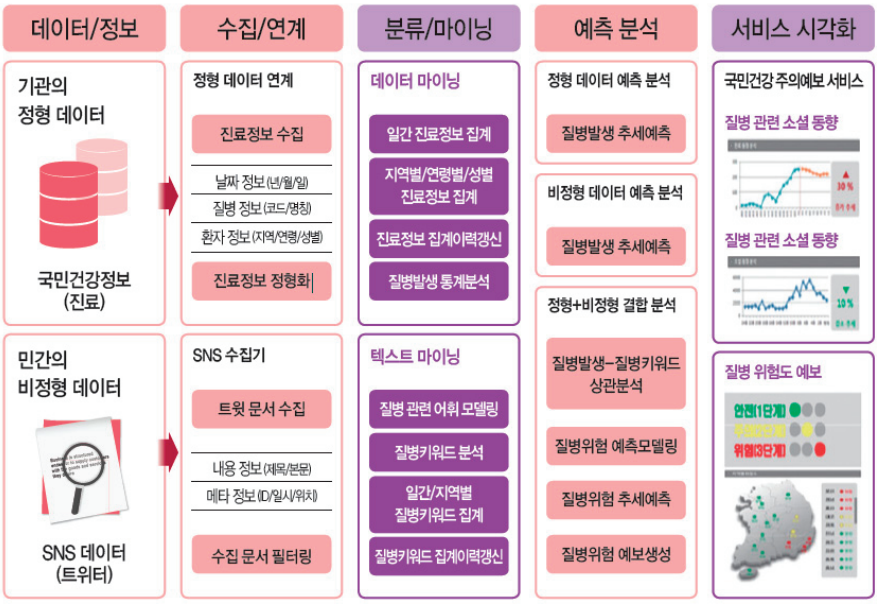


그림 4.22 국민건강 주의 예보 시범 서비스 과정

출처 : 국민건강보험공단, 2013

다. 성과 및 시사점

○ 현재까지의 건강보험 진료서비스는 질병발생 후의 치료 중심이었으나, 국민건강 주의예보 서비스는 급성기(전염병) 질병, 계절적·주기적 발생 질병, 다수 및 다빈도 질병 등 건강이상 징후에 대하여 사전에 건강주의 예보를 발동하는 것으로 앞으로 건강보험이 예방 중심의 서비스로 전환하는 커다란 계기가 될 것

○ 주요 감염병 유행을 예측하여 관련 기관과 국민 개개인이 유행에 대비하고 예방 할 수 있도록 지원함으로써 국민건강 증진 및 사회적 편익 극대화

○ 공단은 건강주의 예보가 발동되면 지역별·집단별(가족, 학교, 직장 등) 맞춤형 건강관리 안내사업을 실시하고, 향후에는 희망자에게 개인별 맞춤형 모바일 건강서비스를 제공할 예정이어서 국민건강증진 및 사회적 편익이 극대화될 것으로 기대

○ “공단이 보유한 국민건강정보 빅데이터를 활용해 질병예방과 생애주기별 맞춤형 건강서비스를 제공할 수 있도록 노력할 계획” (건강보험공단, 2013)으

로 이는 공단 빅데이터의 경제적 사회적 가치를 높이는 주요 기반이 될 것

사례2 - 고용정보통합분석시스템 구축(한국고용정보원)

가. 추진 배경

○ 고용정책을 위한 현황정보의 적시 제공 필요
- 과거 고용 관련 이슈 분석에 지체되는 시간 때문에 적시에 효과적인 연계 분석 어려움
- 개별 운영되는 시스템에 자료를 별도로 요청할 때 소요되는 시간이 지나치게 길어 시기적절한 분석 결과 활용이 어려웠음

○ 데이터의 품질을 높여 대국민 서비스 개선 필요

나. 추진 내용

○ 고용정보 전산망 통합 관리(그림 4.23)
- 고용보험전산망, 워크넷, 직업훈련전산망, 외국인고용관리전산망 등 11개 전산망 보유 및 관리

○ 신속하고 다양한 분석을 통해 고용정보의 적시 제공(한국고용정보원, 2013)
- 연계분석 : 고용보험, 취업알선(워크넷), 직업훈련 등의 다양한 고용정보를 연계 분석한 자료 파악
- 고용동향 : 워크넷, 외국인, 자활, 사회적 일자리, 구인업체개척사업, 경제활동 인구조사 등 관련 통계 자료 추출 및 분석
- 고용보험 : 사업장, 피보험자, 모성보호, 지원금, 실업급여 관련 통계 자료 활용
- 직업훈련 : 실업자, 재직자, 고용촉진, 지도점검, 훈련교사 관련 통계 자료 활용
- 부서현안자료 : 고용노동부 각 팀 별 해당 통계 자료 제공

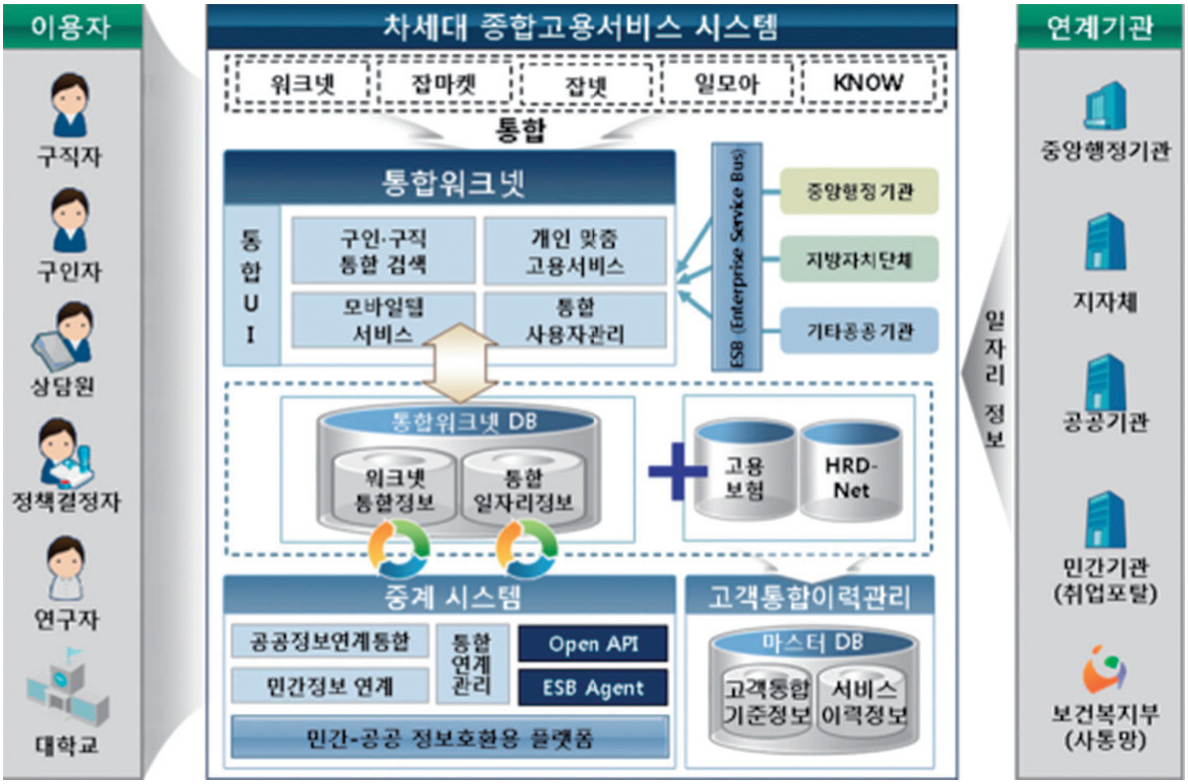


그림 4.23 차세대 종합 고용 서비스 시스템 구성도

출처: 공개SW포털

다. 성과 및 시사점

○ 민간의 노동시장 관련 정보 직접 추출 이용 가능

○ 정부 고용정책의 실효성 증대

- 축적된 데이터베이스를 활용해 신뢰성 있는 정보를 제공, 다양한 연계분석으로 고용정책 실효성 높임

  예) 고용정보시스템을 통해 개인 특성, 구직자 고용 변경 추이 및 현재 상태 파악부터 사업장 특성, 고용 유지 기간 등의 통합 정보 확인 가능

  예) 직업훈련시스템을 통해 개인특성이나 사업장 특성 등의 특성별 정보 취득 : 취업자의 재직자 훈련 이수 여부, 실직자의 실업자훈련 프로그램 참여 여부 확인 및 사업자의 고용 유지 추이 파악 가능

4.4 민간 및 기업의 활용현황

사례1 - 소상공인 창업지원을 위한 점포평가 서비스

- 주관기관 : ㈜오픈메이트
- 참여기관 : 비씨카드(주), 한국감정원

> 최근 명예퇴직한 B씨는 카페를 창업할 계획이다. 퇴직금을 모두 투자해야 하기 때문에 신중할 수 밖에 없으나, 마땅히 창업정보를 얻을 길이 없다. 프랜차이즈 설명회에도 가보고 부동산도 찾아가 봤지만 신뢰해도 되는지 확신이 서질 않는다. 그러나 스토어스토리* 서비스를 알고 나서 눈이 번쩍 띄었다. 서울 시내 어디든 건물내 점포의 개폐업 이력, 매출정보와 데이터 기반의 과학적 상권분석, 적정업종, 추정매출 등 정말 알고 싶었던 정보를 얻을 수 있었기 때문이다.
>
> http://partners.bccard.com에서 시범 서비스 제공

가. 추진 배경

○ 현재 국내 자영업 비율은 OECD 평균의 두 배가 넘는 반면, 3년 내 폐업율은 50%가 넘어 매우 치열한 경쟁 속에 놓여있음

○ 창업 준비과정에서 개인 및 영세 자영업자는 프랜차이즈 설명회, 부동산 사업자의 컨설팅 등에 의존할 수밖에 없는 등 창업정보 인프라가 매우 열악한 실정

○ 민간, 공공의 데이터를 연계 활용하여 소상공인의 창업 의사결정에 도움이 되는 정보 인프라를 제공하고자 함(그림 4.24, 4-25)

나. 추진 내용

○ 활용데이터

기관명	데이터 명
비씨카드	월 2억건씩 누적되는 카드거래데이터 중 서울지역 2억건 (3년치)
한국감정원	부동산 임대시세 및 건축물 대장 데이터
소상공인진흥원	소상공인상가벙보(점포명, 업종, 주소, 전화번호) 총 1억건(3년치)
공공오픈데이터	인구, 가구세대, 36만 블럭

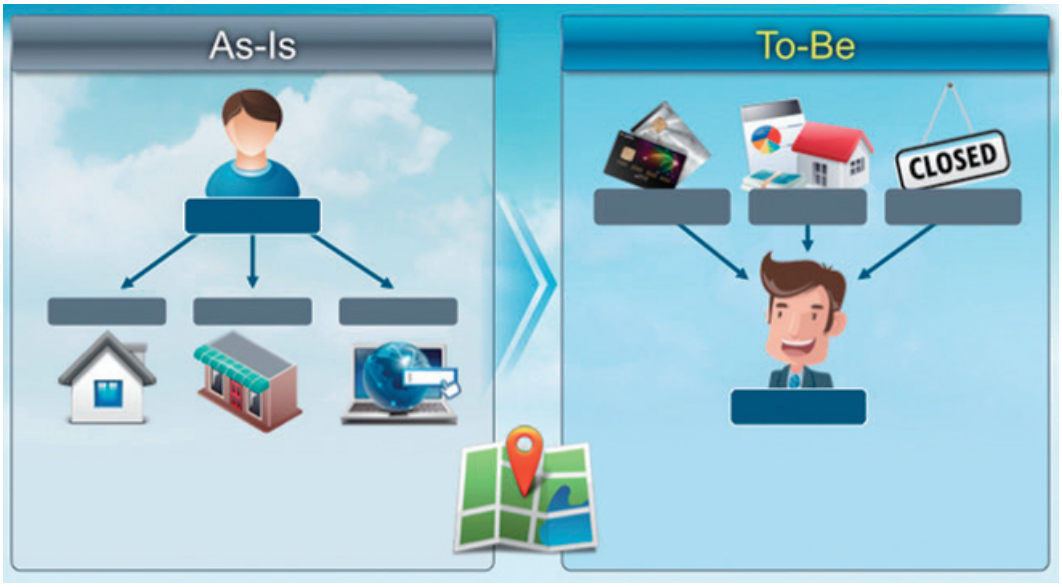


그림 4.24 점포평가 서비스를 위한 빅데이터 분석 도입 효과

출처 : 한국정보화진흥원 2013

○ 주요 개발 내용(그림 4.26)

- 약 1억 건의 상가업소데이터, 6억건 이상의 카드사용 트래픽 데이터 기반 분석
- 동일 지점의 최근 3년간 점포 개폐업 이력 추적
- 점포별 수년간의 개폐업이력을 추적하고, 업종변화별로 해당 점포의 매출 추정
- 창업자가 어떤 업종으로 창업하면 가장 높은 매출을 낼 수 있을지, 어떤 업종일 때 영업기간이 짧고 폐업율이 높았는지, 점포의 입지는 어떤 수준인지 등 창업결정을 위한 지표정보 제공
- 임대시세, 추정매출, 점포진단평점 등 입지상권분석에 필요한 기초정보 제공
- 점포의 과거 개폐업 이력과 더불어 다양한 경쟁지표를 중심으로 상권을 평가하여 특정업종에 대한 평가와 더불어 업종 추천도 가능하게 함
- 행정동 단위의 통계데이터를 보다 활용도 높은 36만개의 블록으로 재가공하고 이에 대해 다양한 컨텐츠를 만들어 상권과 입지분석에 활용할 수 있도록 제공

	사 례	필요성
창업자의 고민	– 내가 창업하고자하는 업종이 현 위치에서 과거 실패한 경우가 있는가? – 이 자리에서 가장 높은 매출을 일으킬 만한 업종은 무엇인가? – 임대료, 권리금은 적정한가? – 예상 매출에 비해 비용이 높지 않은가?	부동산 중개업자 제공 정보 프랜차이즈컨설턴트 분석 의견, 창업전문가 분석 의견에 의존한 창업 → 점포거래, 가맹로열티의 이해관계로 인하여 객관적이고 정확한 정보제공이 되지 않음
정부의 지표	– 1년에 약 36%의 손바꿈이 일어나는 시장 (통계청 생멸율 발표 2011년, 국세청 2011년 829,669개 개인사업자 폐업) – 국내 소상공인수 (265만, 소상공인진흥원) (통계청 생멸율 발표 2011년, 국세청 2011년 829,669개 개인사업자 폐업) – 베이비부머 세대의 본격적인 은퇴시기 도래 (1955년 ~ 63년생 695만 명 : 통계청)	다산, 다사 구조의 창업시장 악순환 신중한 진입을 위한 정보서비스 미비 → 창업시장 경쟁구조 변화 필요
관련 서비스	– 소상공인진흥원 상권정보서비스 상가업소의 현황만을 제시 : 예측 서비스 미비 – 민간서비스 (SKT지오비전, 나이스비즈앱, 창업포탈) 현상파악 위주 insight 제시 부족, 유료모형, 점포중개 등 수익 추구 – 국가 데이터 공개 정책 확대 데이터 공개 후 활용에 대한 컨텐츠 확보 필요	현황정보만을 제공하여 의사결정 지원이 어려움 → 빅데이터 분석을 통한 insight 발굴 미래예측 정보 제공이 필요

그림 4.25 창업자의 고민과 현 수준

출처 : ㈜오픈메이트, 2013

다. 성과 및 시사점

○ 유동인구가 많은 지역을 찾아내고 먹거리·의류 등 상권별 특성을 과학적으로 파악할 수 있는 지도가 마련되면 자영업자들의 실패 확률을 줄일 것으로 기대

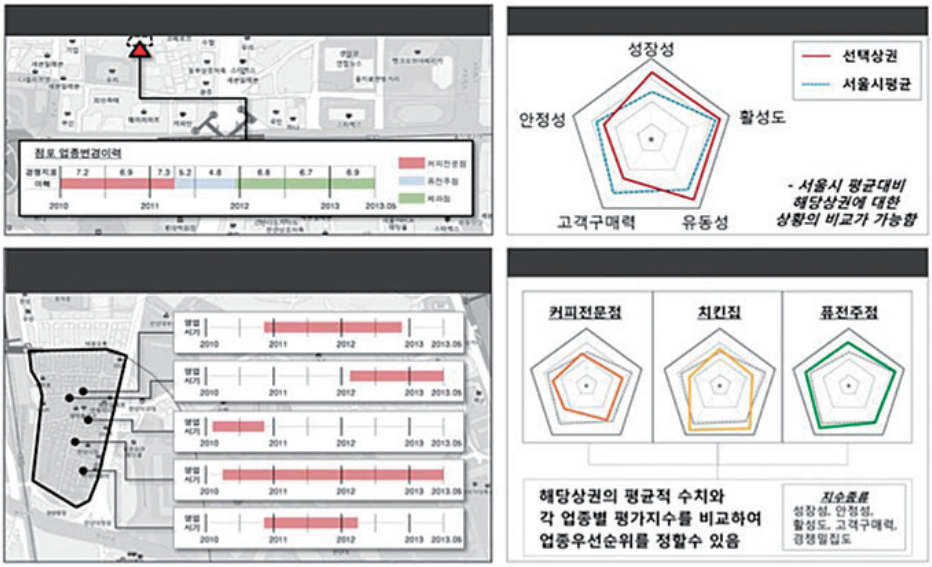


그림 4.26 상권 평가 및 업종 추천

출처 : ㈜오픈메이트, 2013

○ 성공률 높은 업종, 상권 조건이 우수한 점포를 기반으로 선별창업을 할 수 있고, 부정확한 컨설팅으로 인한 피해나 비용지출 감소

○ 창업 점포의 부동산 가치를 객관적으로 판단함으로써 창업 과정에서 나타나는 과잉지출 또는 피해를 최소화

사례2 - 고객 맞춤형 서비스를 위한 빅데이터 센터(신한카드)

가. 추진 배경

○ 국민의 신용카드 사용이 상용화됨에 따라 카드업계의 경쟁이 과열되는 상황
- 신한카드와 국민카드가 올해 3,4분기 전체 카드 이용액이 각각 96조8000억원, 68조원으로 업계 1, 2위이지만, 같은 기간 체크카드 이용액은 KB국민카드가 14조1000억원, 신한이 11조3000억원을 기록하면서 국민카드가 신한을 다소 앞선 상태 (파이낸셜뉴스, 2013.12. 27)
- 카드사들은 새로운 서비스를 제공하는 신카드 출시를 통해 시장점유율 향상에 총력
- 각 카드사는 마케팅전략의 핵심인 매출액 증대를 위해 다각적인 노력을 기울이고 있는데, 신한카드가 빅데이터 이용을 통한 소비패턴 분석에 주력을 둔다면, 국민 카드는 다양한 회원층을 확보하는 데 적극적임

나. 추진 내용

○ 빅데이터 센터 구축
- 2200만 고객의 정보를 모아 새로운 가치를 재창출해 고객에게 되돌려 주는 게 신한카드가 지향하는 빅데이터
- 기존에 다수를 대상으로 하던 매스마케팅을 벗어나 개별적인 특성을 강조한 마케팅에 관심을 기울이고 빅데이터센터를 새로운 성장동력으로서 추진
- 빅데이터센터를 영업 부문이 아니라 경영기획 부문에 배치해 단순한 영업이 아니라 상품개발, 영업, 상담 등 전사적 관점에서 활용하는 '빅데이터 경영'을 추진
- 빅데이터 센터는 빅데이터 기반 '마케팅 파트', 내부 효율화를 위한 '비즈

니스 인사이트 파트', '빅데이터 플랫폼 파트' 등 3개 부문으로 구성

○ 빅데이터를 활용한 소비 패턴 분석
- '빅투그레이트(Big to Great)'라는 광고 문구에 '시장점유율 1위 카드사'라는 멘트를 전면에 내세운 광고를 시작한 가운데 최근 빅데이터를 활용한 '콤보(Combo) 서비스'를 확대(그림 124)
- 콤보서비스는 빅데이터를 통해 카드 회원들의 소비 성향을 분석한 결과를 바탕으로 선호도가 가장 높은 품목에 대해 할인 혜택을 집중 제공하는 서비스
- 2013년 12월부터 콤보 서비스 적용 카드가 하이포인트 카드, 빅플러스 카드 등 7종의 신용카드에서 러브 카드와 에스초이스 카드 등 2종류의 체크카드에도 추가로 적용돼 활용 폭이 확대 (대상 카드 수가 30만장에서 1700만장으로 획기적으로 증가)



그림 4.27 콤보 서비스

자료: 신한카드 홈페이지

다. 성과 및 시사점

○ 개개인의 카드소비 성향패턴을 분석함으로써 고객의 만족도를 높임과 동시에 카드사의 매출액 증대에 기여

○ 과거의 단순 매스마케팅(다수에 대한 마케팅)을 넘어 개개인의 니즈를 위한 고객 과 기업의 소통을 활성화 시킬 수 있음

○ 고객 개인의 소비패턴 분석은 단순한 소비 성향분석을 넘어 관련 분야 마케팅 전략에 활용 가능 (예 : 소비자에 인기 있는 음식점에 대한 정보 제공)

사례3 - 스마트 통합물류 시스템(CJ 대한통운)

가. 추진 배경

○ 물류기업에 있어 전문가들은 빅데이터를 '미래 물류시장의 신호을 읽어내는 안테나'로 요약(그림 125)

○ 빅데이터 분석 시스템 구축을 위해 세계적인 물류기업들이 발빠르게 대응

- 최근 DHL이 물류산업에서의 빅데이터 활용방법을 소개한 '빅데이터 인 로지스틱스 (Big Data in Logistics)'를 발표
- 빅데이터가 물류 운송의 최적화 커다란 변화를 가져오고 있음
- 많은 물류 기업들은 이미 빅데이터의 활용능력이 추후 기업의 성패를 가늠할 핵심 경쟁력으로 인식



그림 28 빅데이터 인 로지스틱

출처 : DHL 홈페이지

나. 추진 내용

○ CJ 대한통운의 스마트 통합 단말기

- 스마트 통합 단말기에 디지털 운행기록계를 결합해 차량의 위치와 경로, 화물의 상태, 연료 소모량, 속도 등을 실시간으로 파악하고 관제하여 물류차의 전반적인 관리와 통제를 효과적으로 할 수 있도록 함
- 최적의 운송경로를 찾아 택배 배송기사들에게 지급된 스마트폰과 태블릿 등에 네비게이션 기능을 통해 전달
- 운전자가 화물을 내린 장소에서 가장 가까운 곳에 있는 다른 화물의 정보

를 실시간으로 확인할 수 있어 공차율을 낮출 수 있을 것으로 전망
- 디지털 운행기록정보를 분석해 급출발, 급정지, 급가속 등의 10대 안전지표를 관리하여 각종 안전사고를 줄일 계획
- 전체 시스템을 새롭게 하여 정보 수집 및 저장 시스템도 단일화 함으로써 빅데이터를 활용하여 화물비용 관련 데이터를 한 곳에서 처리할 수 있도록 함
- 각종 차량의 수집된 정보는 빅데이터 분석을 통해 연료절감을 위한 노력을 꾀할 수 있고, 물류 효율성을 높일 수 있음
- 빅데이터 분석으로 불필요한 지출을 억제시켜 인력 및 비용절감 효과를 누릴 수 있음

○ 통합물류 시스템 확장
- 최근 들어 증가한 택배 물량을 통합적이고 효율적으로 관리하기 위해 파워7+ 프로세서로 보다 강화된 성능을 제공하는 IBM 파워 시스템즈를 추가적으로 도입
- 통합물류시스템 확장을 통해 국내 택배 물동량을 사전에 예측하고 선제적으로 관리함으로써 운영 효율성과 안정성을 증대
- 추석, 설날 등 택배 물량이 급증하는 성수기에 대비한 처리 용량을 확보해 갑작스러운 시장 변화에도 새로운 추가 업무에 유연하게 대응

다. 성과 및 시사점

○ 내적 외적 경쟁력 강화
- 통합 물류 시스템 확장을 통해 안정적이고 최적화된 IT 인프라를 마련하고 고객 서비스 경쟁력을 한층 더 강화
- 서비스의 속도 향상과 연속성을 보장하는 안정된 기반을 확대해 물류 시스템의 전반적인 퍼포먼스 상승은 물론 운영 및 관리 비용의 절감

사례4 - 맞춤형 고객 상품추천 서비스(GS샵)

가. 추진 배경

○ 온라인 커머스 시장에서의 경쟁력 강화를 위한 상품 추천 서비스 플랫폼

구축 요구

○ 내부 역량을 활용해 다양한 서비스에도 확산시킬 수 있도록 자체 역량 확보가 우선시

○ 아마존닷컴의 경우 빅데이터 기반 추천 시스템을 통해 매출의 35%가 증가

나. 추진 내용

○ 아마존닷컴의 추천시스템을 벤치마킹

- 아마존의 추천시스템은 빅데이터 분석을 기반으로 하여 아마존 사이트에 접속한 사용자들이 남긴 로그데이터나 상품과 상품의 유사도 분석 등의 대량의 데이터를 분석해 상품을 추천
- 누군가 책을 사려고 아마존에 접속하면 그의 이전 구매 내역 및 행동 패턴을 분석해 관심을 가질만한 책을 추천
- 고객이 셔츠를 사면 그 셔츠와 어울릴만한 바지를 보여주며 함께 구매하도록 유도
- 추천시스템의 성과를 확인한 아마존은 매년 이익의 10%를 추천 시스템 성능 향상에 투자
- 2012년 7월부터 하둡 기반 빅데이터 플랫폼으로 사용자의 구매 내역과 온라인 게시판에 남긴 텍스트 데이터를 이관하는 동시에 CEP(Continuous Event Processing)나 오피니언 마이닝(Opinion Mining) 등과 같은 빅데이터 관련 내부 연구 과제를 마련해 국내외 파트너사들과 함께 수행
- 국내 유통업체 최초로 하둡기반 빅데이터 플랫폼을 구축하고, 이를 기반으로 추천 시스템(사용자가 온라인 쇼핑 사이트에 접속하면 그가 관심가질 만한 상품을 추천)을 구축 (하둡에코 시스템 도입)
- 사용자들의 구매 내역을 바탕으로 상품 선택 패턴을 분석한 다음 그 결과를 토대로 사용자가 관심을 가질 만한 신제품이나 관련 상품들을 맞춤식으로 추천해주는 새로운 상품 추천 서비스를 구축
- GS샵은 하둡 기반 시스템이 현재는 상품 추천 서비스에만 도입됐지만 보다 다양한 서비스로까지 그 적용범위를 확대해 나갈 방침

다. 성과 및 시사점

- 다양한 몰에서도 사용 가능한 빅데이터 플랫폼 마련
- 내부 인력의 연구개발과 실제 적용을 병행할 수 있는 수준의 내부 개발자들 역량 확보
- GS샵이 구축한 추천 서비스를 위한 데이터 플랫폼은 좀더 거시적인 관점에서 해석하면 상품의 기획 및 준비 단계에 도움을 줄 수 있고, 단기적으로는 보다 정교한 실시간 상품 추천이 가능

4.5 지방자치단체의 활용현황[42)]

지방자치단체의 빅데이터 활용현황을 광역시와 도지역 순서로 구분하여 정리하였으며 구체적인 활용사례는 최근 것을 중심으로 별도로 정리하였다.

1. 서울특별시

2011년부터 현재까지는 '시민과 함께하는 스마트 서울'을 비전으로 개방·공유·소통의 열린 전자정부를 표방하며 공공 정보 개방, 시민 맞춤형 서비스, 모바일 시정 추진 등 양방향 시민 소통 및 참여를 확대해 오고 있다. 최근 활용사례는 다음과 같다.

- KT와 협력해 유동 인구, 교통 관련 빅데이터를 분석해 심야버스 노선을 개선
- 교통·복지·경제·문화 등 다양한 현안 문제 해결을 위해 빅데이터를 활용한 과제 발굴을 추진

2. 부산광역시

시민 중심의 공감형 생활 정보화, ICT를 활용한 도시 기반 정보화, 삶의 질 향상을 위한 스마트 라이프, 건전한 정보문화 정착 및 정보 복지 구현, 정보화를 통한 과학적 행정 구현을 추진 전략으로 삼아 지속적인 부산 지역 정보화를 추진하고 있다. 최근 활용 사례는 다음과 같다.

- 부산시는 시민들이 편리하게 통계를 활용할 수 있도록 부산 도시 서비스 현황 정보를 지도 위에 제공하는 '부산 서비스 인구 통계 시스템'을 개발하였다.
- 2014년부터 2016년까지 연차적으로 국립재난안전연구원과 협력해 스마트 빅보드를 도입키로 하고, 2014년에 부산시와 기장군에 시범 사업을 추진하였다.

42) 이 절의 내용은 2015년도 국가정보화 백서의 내용을 옮겨오거나 발췌하여 정리한 것임

3. 대구광역시

참여와 소통을 위한 행정정보화 실현, 대 시민 정보 서비스 제공을 위한 공공데이터 개방을 위한 시스템 구축을 추진하고 있다. 최근 활용 사례는 다음과 같다.

- 수요자 맞춤형 가치 있는 공공데이터 제공을 위하여 시스템 구축 및 데이터 개방 활성화

4. 인천광역시

'녹색 창조도시 실현을 위한 열린 융복합 정보화'란 비전을 세우고, '새로운 인천, 행복한 시민'을 위한 정보화 서비스에 총력을 기울이고 있다. 이를 실천하기 위해 소통과 공유 기반의 맞춤형 행정 체계 마련, 생활공간 기반의 맞춤형 정보 제공 체계 마련, 산업공간 기반의 맞춤형 지원 체계 마련 등을 추진하고 있다. 최근 활용 사례는 다음과 같다.

- 전국 최초로 전·월세, 매매 등 부동산 실거래가 모바일 지도 서비스와 공간 Big Data 시민 공개 서비스인 '시민행복 오감(五感) 지도 서비스'를 시작
- 빅데이터 활용·분석을 위한 빅데이터 정보화 전략 계획을 마련해 다가올 미래에 필요한 시민 서비스, 예측 민원 서비스 제공, 정책 입안 시 정확한 데이터를 제공 하는 체계 구축 예정

5. 광주광역시

스마트 기기의 보급 대중화 및 ICT 융·복합 기술의 발전으로 시민 중심의 지역 정보화 발전을 이룩하기 위한 새로운 정보화 전략과 정책 수단이 요구되고 있어, 2015년도에는 5개년(2015~2019년) 정보화 중장기 발전 계획을 수립해 추진하고 있다. 최근 활용 사례는 다음과 같다.

- 도로 및 교통수단에 전자통신 등 첨단 ICT 기술을 접목한 효율적인 지능형 교통 시스템(ITS : Intelligent Transportation Systems)을 구축

- 호남 지역 중소기업을 대상으로 정보 보호 지원 사업 및 정보 보호 교육 운영 등 정보 보호 종합 지원 체계를 마련하고자 호남권 정보보호지원센터 구축

6. 대전광역시

2015년 다양한 정보통신기술(ICT)을 활용해 시민의 삶의 질을 향상시키는 데 중점을 두고, '지역정보통합센터'의 이전·구축을 통한 예산 절감 및 정보 자원의 효율성 제고를 목표로 두고 있다. 최근 활용 사례는 다음과 같다.

- '시민이 안전한 도시, 대전'을 만들기 위해, 기존에 운영 중인 U-City통합센터의 CCTV 영상 정보를 전국 최초로 112·119 상황실에 제공하는 사업으로까지 확 장해 시범 운영할 계획

7. 울산광역시

안전한 시민 생활을 보장하고 미래형 첨단 도시 건설을 위해 ICT (Information and Communications Technology) 기술을 이용해 도시의 복잡한 상황을 관리할 수 있는 U-CITY 통합관리센터 구축을 추진하고 있다. 최근 활용 사례는 다음과 같다.

- '시민이 안전한 도시, 대전'을 만들기 위해, 기존에 운영 중인 U-City통합센터의 CCTV 영상 정보를 전국 최초로 112·119 상황실에 제공하는 사업으로까지 확 장해 시범 운영할 계획

8. 세종특별자치시

안전하고 편리한 삶의 질을 향유할 수 있는 정보 소통 도시 구현 및 유비쿼터스 사회를 선도하는 명품 U-City 구현을 목표로 U-정보 서비스, U-도시 인프라, U-비즈니스 등 3개 분야에 대한 추진 전략을 수립하여 추진. 최근 활용 사례는 다음과 같다.

- U-방범 서비스는 읍면 및 건설 지역 496개소 819대 방범용 CCTV를 통

합 운영이 가능하도록 하여 범죄 발생 시 방범용 CCTV 카메라가 현장을 관제하고 도시 방범센터 관제 요원이 사건·사고에 유연하게 대처하는 등 실질적인 범죄로부터 안전한 도시가 되도록 고도화 된 서비스를 제공
- U-교통 서비스는 시민 생활의 이동 편의성 제공과 생활 편의 교통 체계의 효율적인 관리를 위하여 실시간 데이터를 신뢰성 있고 예측력 높은 정보로 가공하여 효과적인 교통 관리가 가능하도록 하고 교통 관리 전략을 수립하는 모니터링 체계를 구축
- U-시설물 관리 서비스는 U-City의 각종 시설물 도면 조회와 VMS (Variable Message Sign), BIT(Bus Information Terminal), CCTV 등의 상태를 모니터링 하고 도시통합정보센터 및 현장 시설물의 제원을 통합적으로 관리하여 시설물 관리 업무의 효율성을 제고.

9. 경기도

빅파이 프로젝트를 추진하여 경쟁력 있는 민간 참여를 최대한 유도하고 경기도는 이를 지원하는 방식으로 사업을 추진, 공공데이터 뿐 아니라 민간이 보유한 데이터와 서비스를 받아 수요자 중심의 빅데이터 서비스 모델을 만드는 것이 핵심 내용이다. 빅파이 프로젝트는 빅데이터(Big-data)와 프리 인포메이션(Free Information)의 합성어로 최근 활용 사례는 다음과 같다.

- 빅데이터에 기반을 둔 과학적 행정 서비스 제공과 데이터 산업 생태계 조성을 통한 일자리 창출 등 빅데이터 활용에 대한 규정을 담은 조례를 국내에서 경기도가 처음으로 공포
- 빅데이터의 활용과 관련된 사항을 심의하기 위해 빅데이터위원회 설치
- 재난안전에 관한 모든 정보를 한 곳에 모아 지도를 기반으로 과거 재난이 발생했던 지역과 사고발생 위험이 있는 지역이나 시설물을 지도에 표시해 재난을 사전에 감지·예측하는 서비스인‘ 전 대동여지도’구축
- ‘빅데이터 전문 인력 양성 사업’ 실시

10. 강원도

산학협동으로 비즈니스 데이터 분석 전문 인력을 양성하고 기업 문제, 공공의 현안 문제, 사회 문제를 공개하여 해결할 수 있는 전국 규모의 오픈 플랫폼 구축·운영 등의 사업을 추진. 최근 활용 사례는 다음과 같다.

- 대학의 SW 전공 운영 체계 개선을 통해 우 수 SW 융합 인재를 양성하여 전문 데이터 과학자를 양성

11. 충청북도

도시의 대부분이 농촌과 혼재된 도농 복합 도시의 형태로 도시 기반 정보화가 활발히 진행되고 있지는 않다. 다만, 점진적으로 이뤄지는 도시화가 도심을 중심으로 이루어지고 있으며, 개별적으로는 부분적인 도시 개발 및 발전의 요인으로 공간 정보화 사업이 추진되고 있다. 최근 활용 사례는 다음과 같다.

- 지하 시설물 DB 구축으로 웹기반 시설물 관리 시스템, 도로 점용 인허가 시스템, 공간 정보 자 동 갱신 시스템 등 표준 시스템 등을 구축 추진

12. 충청남도

모바일 전자정부 지원센터의 공통 기반을 토대로 하는 ‘충남 모바일 오피스’는 모바일 네트워크를 위한 망 연계, 문서 변환 서비스, 바이러스 차단 시스템 도입 등 스마트워크 환경 조성과 스마트 충남 구현을 위한 노력에 역량을 집중하고 있다. 최근 활용 사례는 다음과 같다.

- 충청남도 모바일 행정 업무시스템 ‘바로일터’ 구축 및 개선 추진

13. 전라북도

빅데이터 기반의 스마트 도정 시스템을 도입하는 걸 목표로 농업 · 관광 · 재난 등 다양한 분야의 사회 현안 해결과 미래 변화 예측을 통한 선제적 대응을

추진하고 있다. 최근 활용 사례는 다음과 같다.

- ‘기상 요소와 농작물 생육 간 상관관계’를 빅데이터로 분석하기 위한 기상 관측 센서 및 분석 시스템 구축 사업을 진행
- 빅데이터를 활용한 한옥마을 관광객들의 성향 분석을 통해 맞춤형 관광 콘텐츠를 개발 및 지원
- 전주시 시내버스 노선 체계 개편을 위해 이용자 현황에 대한 빅데이터를 분석, 2016년 전주시 시내버스 노선 체계 개편을 위한 의사결정 지원

14. 전라남도

매년 발생하는 정책들을 가시적으로 분석하고 전라남도가 보유하고 있는 데이터를 가공·분석하여 정책 결정의 기반 자료로 사용할 수 있는 토대를 마련할 계획이다. 또한 민간 빅데이터를 활용하여 도정 정책 결정의 공신력 있는 증빙 자료로 활용 가능하도록 유동 인구 및 매출 현황 관련 빅데이터 를 활용한 공간 정보 시스템을 구축을 추진하고 있다. 최근 활용 사례는 다음과 같다.

- 기본적인 공간 정보 데이터를 기반으로 한 정책주제도에 빅데이터 분석 결과를 결합하여 표기하는 공간 정보와 빅데이터 융복합 시스템의 구현을 통해 누구나 손쉽게 이용할 수 있는 사용자 중심의 플랫폼 구현

15. 경상북도

신청사 이전에 따른 정보통신 시스템의 재구축, 활성화 등을 추진하고 있다.
- 이전에 따른 국가정보통신망의 개선 추진

16. 경상남도

‘당당한 경남시대’비전 달성을 지원하기 위해 지역 정보화 사업을 적극적으로 추진하고 있다. 또한 5년마다 수립하는‘지역 정보화 기본 계획’을 토대로 매년 시행 계획을 마련하여 다양한 분야의 정보화 사업을 추진하고 있다. 최근 활용 사례는 다음과 같다.

- 18개 행정업무와 인사, 재·세정, 건축행정 등의 공통 행정업무를 전산화·표준화.
- 전자문서 처리 시스템인 온-나라 시스템을 안정적으로 운영하기 위하여 저장 공간 용량 증설, 통합 백업 체계 구축 등 체계적으로 사업을 추진
- 개인정보 보호를 강화하기 위해 교육 대상을 세분화(일반/전문/담당자/수탁자)하여 집중 교육을 실시

17. 제주특별자치도

도에서 보유하고 있는 각종 데이터 DB를 개방형·표준형(Open-API)으로 전환하여 민간 활용을 증대하고 SNS, 모바일, 웹 표준화·접근성 등 최신 환경에 대응하기 위하여 대표 홈페이지를 개편을 추진하였다. 최근 활용 사례는 다음과 같다.

- 2015년도에 부서별로 운영 중인 24개의 홈페이지를 도 대표 홈페이지(www.jeju.go.kr)에 통합하고, 보건·복지 분야, 경제·투자 분야, 역사·문화 분야 등 4개의 통합 홈페이지와 이를 관리하기 위한 관리자 시스템을 구축
- 도 대표 홈페이지에 자동 번역 시스템을 적용함으로써 통합되는 서브 페이지의 모든 외국어 홈페이지에 대하여 실시간 번역 서비스를 제공

사례1 - 빅데이터 분석을 통한 심야버스 노선 정책 지원

- 주관기관 : ㈜KT
- 참여기관 : 서울특별시

> 야근이 잦은 직장인 A씨는 요 근래 한시름 덜었다. 12시 넘어 끝나는 일이 많다 보니 몸도 피곤하지만 택시비가 늘 부담이었고, 무엇보다 택시를 잡는 것 자체가 큰일이었다. 그러나 심야버스로 모든 문제가 해결됐다. 새벽 5시까지 다니기 때문에 대중교통의 공백이 사실상 없어졌다. 사람 많은 동네를 두고 먼 곳으로만 다니던 버스가 신기하게도 이번엔 동네로 들어왔다. 통신사의 유동인구 빅데이터를 분석한 과학적 노선 수립 덕분이다.
>
> 서울시 심야버스 7개 노성(2013. 9월) 수립 지원

가. 추진 배경

○ 심야시간대 택시 승차거부로 인한 불편, 심야 근로자의 교통수단 부재, 교통비 부담 등으로 인해 서울시는 심야시간대 대중교통의 편의 마련이 필요

○ 수익성이 낮은 만큼 최소의 비용으로 서울시내 전역을 지원할 수 있는 합리적 노선 수립방법이 필요한 상황이었고, 이를 위해 통신사업자인 KT와 협력하여 데이터 기반의 객관적 자료를 근거로 한 심야버스 노선 수립을 추진하게 됨(그림 4.29)

나. 추진 내용

○ 활용데이터

기관명	데이터명
㈜KT	CDR(Call Detail Record), 고객통계 정보(30억건)
서울특별시	시내버스 현황, 정류소 현황 등

○ 주요 개발 내용

- 지역별 유동인구 파악을 위해 서울시를 블록 단위로 분할하고, 목적지, 통행량 등 을 산출하기 위해 거리 기반 알고리즘 제작
- 심야시간대 유동인구의 밀집도를 분석하여, 시민들의 이동경로에 있어 버스노선이 가장 필요한 곳을 중심으로 구역을 설정하고, 이용객의 최적화된 수준을 효율성의 관점에서 고려하여 배차간격 조정
- 빅데이터에 기반을 둔 유동인구 분석은 시간대별 이용객 수뿐만 아니라 이용승객의 특성에 이르기까지 복합적이고, 구체적인 분석을 가능하게 함
- 2013년 4월부터 실시된 ‘심야전용버스’ 정책지원 서비스는 시민들의 큰 호응에 힘입어, 심야시간대 유동 인구가 많은 지역을 중심으로 7개 노선이 추가되었고, 9월부터는 총 9개의 올빼미 버스노선이 운용 중에 있음



그림 4.29 버스노선 수립을 위한 빅데이터 분석 도입 효과

출처 : 한국정보화진흥원 2013

다. 성과 및 시사점

○ 민원, 버스사업자 의견 등에 의해 결정되던 버스노선을 데이터 분석을 통해 결정함으로써, 데이터에 기반을 둔 신뢰성 있고 과학적인 행정구현 및 시민 편익 증진(그림 4.30)

○ 버스 운송량 증가 및 버스 노선에 대한 지역 간 분쟁 소지를 차단하였고, 기존 시내버스에 비해 만족도 향상(기존 시내버스 만족도 74.30점, 본 심야 버스 만족도 80.15점)

○ 빅데이터가 앞으로 정책결정에 상당한 영향을 미칠 수 있음을 증명하였으며, 지방행정정보화 연찬회에서 본 서비스로 서울시가 대통령상 수상(2013.9.12.)

○ 빅데이터 융합, 분석 결과를 정책에 반영하는 획기적인 사례로 창조적인 행정 실현을 위한 민관 협력의 계기가 됨

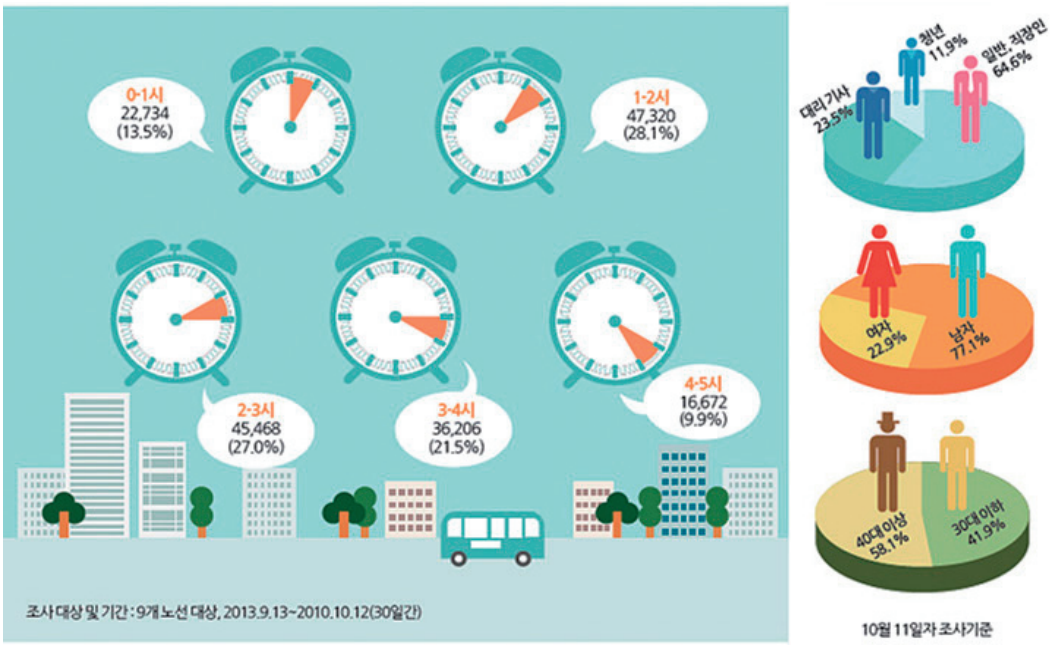


그림 4.30 심야 전용 버스 이용 승객수 및 특성

출처 : 서울시 홈페이지

사례2 - 빅데이터 기반 영주사과작황정보 서비스(경상북도, 미래창조과학부, 한국정보화진흥원)

가. 추진배경

○ 경상북도의 지역적 특성에 기반을 둔 사과재배의 중요성
- 경상북도의 농업인구는 도민의 18%이고, 농업면적 또한 28만ha로 전체 가용면적의 49%에 달함(경상북도 공공분야 빅데이터 활용방안 연구, 2013)
- 전국 사과 생산량의 63%가 경상북도에 의존하고 있는 상황이며(영남일보, 2012), 그 중에서도 영주는 전국 최대 사과 생산의 주산지로서 생산량이 7만여 톤에 이름(매일신문, 2012)

○ 기존 정부 사업의 결과로 광대역망이 구축되어 있음
- 2010년 4월, 방송통신위원회와 범부처, 지자체, 통신사업자 등이 참여하여 농어촌 스마트 인프라 구축 시작(한국정보화진흥원, 2010)

○ 생산량 제고를 위한 기존 u-IT 시스템의 문제점 고찰
- 2012년부터 온도, 습도, 토양 내 수분, 강우 등과 같은 환경적 특성이

u-IT 기술을 통해 측정되는 생장환경관리 시스템이 영주시내 사과 재배 농가 중 13가구에서 실시·운용되고 있음
- 하지만, u-IT 센서 설치에도 불구하고, 기기관련 유지·보수가 적절하게 이루어지지 않거나, 일부 기능이 미비하여 시스템 활용에 애로점이 상존함

○ 사과 생산성 제고 및 적절한 사과 판매가격 책정에 기여
- 농업 관련 공공정보와 시민의 민간정보를 통합한 데이터를 분석하여 병충해 징후 발견, 생산량 추정, 가격동향 등의 사과작황정보를 제공하여 궁극적으로 사과 생산량을 증가시키고 적절한 사과가격을 책정하는데 참고자료로 활용 가능하도록 함

○ 경북도 영주사과 빅데이터를 수집하고 정제한 후 텍스트마이닝과 데이터마이닝을 실시하고 빅데이터 모형을 개발해 검증해 봄으로써 우리나라 빅데이터 연구 와 산업의 발전에 기여하고자 함

나. 추진내용

○ 빅데이터 셋 구축

정형 데이터

항목	세부항목	수집기간	데이터량(건)
기후	기온, 습도, 일조, 풍속, 강수량, 기타	2000~2013년	90,608
u-IT	기온, 습도, 토양수분	2011년 11월~2013년	약 200만
병해충	복숭아순나방, 복숭아심식나방, 애모무늬잎말이나방, 기타	2005~2012년	2,058
생산량	전국 및 영주 사과생산량, 재배면적	2000~2012년	39
가격	전국 및 영주산 가격, 대체과일가격, 물가지수, 사과지수	2003~2013년	약 185만

비정형 데이터

항목	수집기간	데이터량(건)
영주, 사과, 가격, 병충해, 생산	2000~2013년	약 20만
관련 핵심어		약 70만

* 핵심어는 웹&SNS 기반으로 데이터를 수집한 데이터 다섯 가지 분석단어(영주, 사과, 가격, 병해충, 생산)로 수집된 자료를 분석하여 생성한 단어

○ 빅데이터 모형 개발

- 사과 작황에 영향을 미치는 병충해 징후 발견과 사과 생산 추정 및 가격 동향 정보 의 제공을 위하여 관련 데이터를 수집하고 정확한 서비스 모형 개발
- u-IT 센서와 기상청으로부터 수집한 기후 데이터와 페르몬 트랩을 통해 수집된 병해충 유찰 수 등 정형 데이터와 웹과 SNS의 병해충 데이터 등 비정형 데이터를 수집, 분석해 병해충의 발생 경위와 정보를 공유, 농민들에게 농약 살포 시기 등 유용한 정보를 제공(그림 128)

○ 빅데이터 분석 결과 제공

- 2014년 1월부터는 스마트폰이나 컴퓨터, IPTV 등을 통해 관련 서비스를 제공할 계획이며, 농촌 광대역망 구축 및 활용을 위하여 마을회관 5곳에 IPTV, 사과재배 농가 25곳에 와이파이를 설치하여 결과 제공

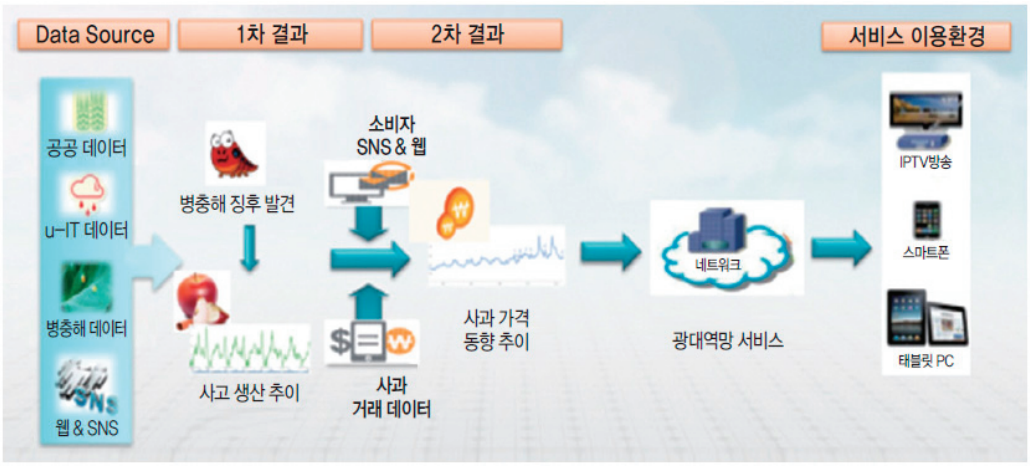


그림 4.31 사과 작황정보 서비스 개념도

출처 : 경상북도, 미채창조과학부, 한국정보화진흥원 2013

- 웹, 블로그, 카페, 트위터, 페이스북, 유투브 등의 비정형 빅데이터(Unstructured Big Data)는 The SCRM을 통해 수집하고 처리한 후 패턴 분석과 상관성을 통한 예측분석을 실시하여 정보를 제공함(그림 4.31)

다. 성과 및 시사점

○ 병충해 징후 발견

- u-IT 센서 데이터와 기상청의 기후데이터를 분석하여, 병충해의 징후를 발견함으로써 사전에 예방할 수 있도록 하여, 궁극적으로는 사과 생산 증대에 기여

o 생산 추정

- 사과생산 정형 데이터와 웹 & SNS상의 사과 생산량과 관련된 비정형 데이터를 수집 및 분석하여, 앞으로의 생산량을 추정함으로써 사과 마케팅과 관련 정책 수립에 참고자료로 활용

o 가격동향

- 사과가격 정형 데이터와 웹 & SNS상의 사과 가격과 관련된 비정형 데이터를 수집 및 분석하여, 앞으로의 가격 동향을 파악함으로써 사과 출하시기 결정과 판매가 책정에 참고자료로 활용 가능



그림 4.32 영주 사과 작황 정보 플랫폼 메인 화면

출처 : 경상북도, 미채창조과학부, 한국정보화진흥원 2013

4.6 학계 및 연구기관의 활용현황[43)]

데이터의 증가는 일반 기업의 문제가 아니라 과학계에도 영향을 미치고 있다. 마이크로소프트의 짐 그레이는 과학의 방법론 자체를 바꾸어야 한다는 제4 패러다임(The Fourth Paradigm)을 주장하였다(Hey et al., 2009). 그의 주장에 따르면 최초의 과학은 자연현상을 관찰함으로 시작되었고, 다음으로 자연을 이해하기 위한 이론적 모델을 만드는 방식으로 과학 문제를 해결하였으며 뉴턴의 법칙이 좋은 예이다. 이후 컴퓨터가 등장한 지난 수십 년 동안 과학은 복잡한 현상을 시뮬레이션하는 방식을 사용하였다. 제4 패러다임은 이제는 측정 장치에 의해 얻어진 데이터를 소프트웨어로 분석하여 새롭게 얻은 지식을 발표하는 방식으로 바뀌었음을 주장한다. 실제 우주과학 분야에 있어 연구자는 천체망원경을 조작하며 데이터를 얻는 것이 아니라 이미 얻어진 수천 개 이상의 천체 관측 데이터를 소프트웨어를 이용하여 분석하고 이로부터 새로운 이론을 추출하는 방식을 사용하고 있다(그림 4.33).

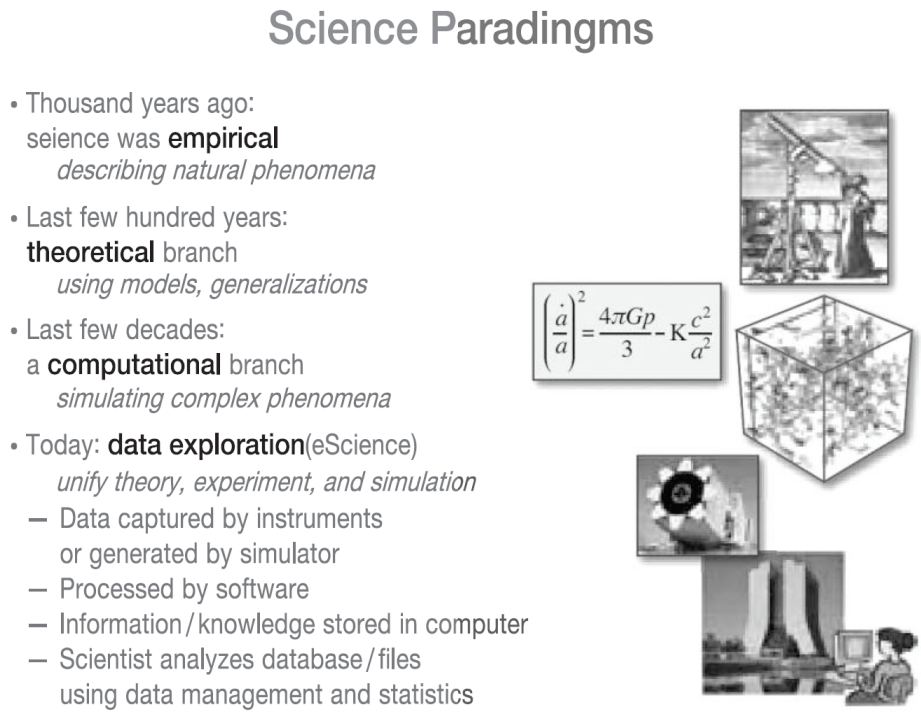


그림 4. 33 제4 패러다임 개념 요약

출처 Hey et al.(2009)

43) 이 절의 내용은 빅데이터와 공공데이터 활용의 내용을 옮겨오거나 발췌하여 정리한 것임

짐 그레이는 미래의 연구방식은 이러한 데이터 분석기법이 다양한 과학연구의 일반적인 현상이며 과학자는 데이터를 다루는 능력을 가져야 하고 이를 위해서는 데이터 분석 소프트웨어 기술을 더욱 발전시켜야 함을 강조한다. 또한 인터넷은 단순히 연구결과를 배포하며 필요한 자료를 검색하여 사용하는 수단으로 그치지 말고, 검색된 연구 논문에 포함된 실험결과뿐 아니라 실험에 사용된 데이터를 함께 제공하는 방식으로 바뀌어야 함을 역설하고 있다. 연구결과의 검증 내지는 추가적인 연구는 이러한 데이터를 활용함으로써 발전할 수 있으며 이러한 방법을 사용할 경우 과학기술을 한 단계 격상할 수 있을 것이라고 주장한다. 빅데이터는 많은 실험을 거쳐야 하는 생명과학 분야에서 우선적으로 주목받고 있으나 모든 과학 연구에서 도입되고 있다.

학계 및 연구기관 활용사례

학계와 연구기관의 활용사례를 표 4.11에 정리하였다. 다양한 활용사례들이 존재하고 있어 공공분야의 활용사례와 차별성이 있는 것들을 중심으로 간략하게 정리하였다.

표 4.11 학계 및 연구기관 활용사례 정리(2015~2016년)

기관	활용사례
빅데이터, 보험 전 부문에 영향	보험연구원의 '보험산업의 빅데이터 활용 가능성'에 따르면 보험산업에서 요율 및 상품개발, 판매 및 마케팅, 언더라이팅, 보상에 이르기까지 전 부문에 걸쳐 빅데이터가 영향을 미칠 것으로 전망
서울대, 만화영화 빅데이터 기반 인공지능기술 개발	서울대 공대, 컴퓨터공학부와 인지과학연구소 장병탁 교수 연구팀(하정우 연구원)은 만화영화를 보고 그림과 언어가 연합된 개념을 습득하며 스스로 지능을 향상시키는 '상상력 기계"를 개발. 향후 외국어 개인교사나, 피교육자의 학습능률 향상에 기여 할 것으로 기대
소셜 빅데이터 분석으로 청소년 자살 막을 수 있다?	한국보건사회연구원 통계정보연구실은 국내 소셜 빅데이터를 활용한 청소년 자살사례에 대한 연구결과를 발표, 온라인 등의 소셜 빅데이터를 분석해 활용하면 청소년의 자살도 예방할 수 있다고 주장. 향후 소셜 빅데이터를 활용한 보건복지 정책 동향과 수요를 분석하고 예측할 계획

기관	활용사례
국립암센터 연구팀, 빅데이터 기반 항암신약개발 프로세스 구축	국립암센터가 빅데이터 기반 항암신약개발 프로세스를 구축. PATHOME 네트워크 분석법을 개발하여, 작용기전의 검증이 가능한 약물작용점 확보 가능. 특히 위암의 바이오 빅데이터에 적용해 HNF4α - WNT5A로 이뤄진 새로운 약물작용점 발굴.
KAIST, 빅데이터 통해 고전음악 창작원리 규명	KAIST문화기술대학원 박주용 교수 연구팀이 빅데이터를 이용해 서양 고전음악의 창작, 협력, 확산의 원리를 밝힘. 연구팀은 ArkivMusic과 세계 최대 음반 정보 사이트인 올뮤직 가이드(All Music Guide)를 첨단 데이터와 모델링 방법을 사용해 분석, 고전음악 작곡가들의 시대와 스타일이 어떤 패턴을 이루는지 탐구하여, 수 백 년의 차이가 있는 음악가들 사이에서도 긴밀한 네트워크가 존재함을 발견. 또한 소비자들의 음악적 취향이 고전음악 성장에 어떤 영향을 끼쳤는지 규명
수산업 미래, 빅데이터 활용에 달려	수산업의 각종 자료를 망라한 빅데이터와 첨단기술 활용이 국내 수산양식 산업의 미래를 좌우할 것이라고 전망. 또한 "IT(정보기술)와 BT(생명공학), CT(문화 기술), ET(환경기술), NT(나노기술)를 적용한 새로운 양식 기술을 창조해 수산업의 생산력을 높여야 한다"고 주문(2015 세계양식학회)
빅데이터 기반 초고성능 유전자 진단기술 개발	김민수 대구경북과학기술원(DGIST) 정보통신융합공학전공 교수와 구재형 뇌인지과학전공 교수 공동연구팀이 4년간의 연구 끝에 빅데이터 분석 기반 초고성능 유전자 진단기술(MRPrimer)을 개발. 연구팀이 개발한 유전자 진단용 프라이머(Primer) 디자인 기술은 사람이나 동식물의 전체 유전자 데이터에 대해 맵리듀스(MapReduce)기반 복잡 알고리즘을 적용, 프라이머를 자동으로 완벽하게 모두 찾아낼 수 있는 세계 최초의 초고성능 기술로, 기존 유전자 진단 프라이머 디자인 기술(Primer3)의 문제점들을 해결. MRPrimer 기술은 유전자가 알려진 생물 종에 대해 프라이머들을 빠짐없이 모두 찾아줌과 동시에 진단율이 우수한 순서대로 결과를 보여주는 뛰어난 성능과 함께, 유전자 기반 신종 바이러스 진단, 암 진단, 유전자 변형농산물(GMO) 탐지 등 광범위하게 사용 가능할 것으로 기대.
강원대, 한국학자료 빅데이터 구축 계획	강원대학교 고문서연구센터가 교육부와 한국학중앙연구원이 추진하는 '한국학기초자료사업'에 선정. 한국학자료센터 사업은 국내외에 산재한 한국학 고전 자료를 체계적으로 수집·정리·분석해 학술적·교육적 기반을 구축하는 사업으로 한국학자료 부분의 빅데이터 구축사업으로 불리며, 강원대 고문서연구센터는 앞으로 6년간 12억원을 지원 받아 '강원권 한국학자료센터'를 운영하게 될 계획

기관	활용사례
경희대학교, 한글 텍스트 마이닝 도구 RHINO 개발	한글 텍스트 마이닝 도구인 RHINO를 개발. 기존 한글로 기록된 대량의 데이터에서 비정형이거나 반정형인 단어들을 추출해 내기가 쉽지 않다는 단점에 기반하여, BK21 플러스 사업팀에서는 국어학 전문가인 최석재 교수를 영입, 새로운 분석기를 개발. RHINO의 가장 큰 장점은 맥락을 참조하지 않는 기존의 분석기와는 달리 앞과 뒤의 단어 배열을 참조하여 단하나의 최종, 최적 결과만을 제시 가능.
한양대 김상욱 교수팀, SNS 빅데이터 분석속도 향상 기술 개발	SNS 분석을 위한 핵심 연산의 하나인 두 희소행렬 간의 곱셈(sparse matrix multiplication)을 기존 방법보다 수백 배 빠르게 처리할 수 있는 방법을 개발, 그래픽처리장치(GPU)를 활용해 SNS 분석의 핵심 연산인 두 희소행렬 곱셈을 빠르게 처리하는 획기적 방법을 제안.
KISTI, 70배 빠른 초고속 빅데이터 분석기술 개발	동일한 컴퓨팅 환경에서 전통적인 데이터베이스 관리시스템 등을 적용한 시스템과 비교한 결과 70배 빠른 빅데이터 분석기술을 개발. 이번에 개발한 빅데이터 분석 시스템인 투픽스(Turning Pixels into Knowledge and Science)는 데이터 저장·계산 방식의 변화를 통해 대용량 데이터를 병렬분산처리할 때 발생하는 원형 데이터의 전처리·불러오기·재구성 과정 없이 바로 원형 데이터에 접근하는 방식을 취하며, 아울러 환경설정을 통해 손쉽게 필요한 만큼 컴퓨팅 노드를 추가·연동할 수 있다는 것이 특징.

CHAPTER 5

빅데이터 통계활용에 대한 미래 전략[44)]

학습목표

- 빅데이터 통계활용 현황에서 나타난 한계와 문제점에 관하여 학습한다.
- 개선과 발전방향에 대하여 제도적 구조개선, 현실적 이용확산 측면에서 학습한다.

빅데이터 정책주체로서 공공부문의 역할은 빅데이터 관련 정책이 개별 부처의 범위를 넘어서는 다부처 정책통합으로 추진되도록 하는 것이다. 신생기술을 둘러싼 새로운 기술혁신체제의 발전과정에는 유인기제와 저해기제가 동시에 존재하는데 이는 전통적인 기술 혹은 산업과는 전혀 다른 정책적 시사점을 주고 있다. 유인기제가 해당 기술에 대한 사회적 기대를 반영하는 것이라면 저해기제는 해당 기술이 가지는 잠재적 갈등이 상존함을 의미한다.

빅데이터 시대에서 공공부문의 역할이 주목받는 이유는 공공부문이 다양한 데이터를 생산하는 주체이자 동시에 사회 현안 문제의 효율적 해결을 위해 다양한 데이터를 필요로 하는 수요의 주체라는 점, 또한 적절한 규제 혹은 진흥정책을 통해 빅데이터 생태계에 영향을 미치는 정책의 주체이기 때문이다. 따라서 빅데이터 정책에는 데이터의 생산자와 수요자로서 정부의 바람직한 역할과 함께 동시에 빅데이터 유인기제를 강화하는 진흥정책과 저해기제를 해소하기 위한 규제정책이 적절히 혼합되어야 한다. 신기술이 가지는 특성 중의 하나

44) 이 장의 내용은 '새로운 기술패러다임으로서 빅데이터: 쟁점과 과제'의 내용을 옮겨오거나 발췌하여 정리한 것임

는 해당 기술의 발전방향이 아직 열려있다는 것이다. 따라서 공공부문의 R&D 정책은 단지 해당 기술의 유인기제를 강화하는 역할 뿐만 아니라 저해기제를 해소하기 위한 다양한 기술적 가능성까지 염두에 두어야 한다.

그림 5.1에서와 같이 공공부문에서 데이터를 생산하는 주체, 데이터를 필요로 하는 주체, 규제 부처, 진흥 기관, R&D 정책의 주체는 모두 다르다. 정부 내에서조차 빅데이터에 대한 이해관계가 다를 수 있기 때문에, 빅데이터를 둘러싼 다양한 쟁점과 과제들을 해결하기 위해서는 무엇보다 공공부문과 함께 민간부문의 이해관계자들까지 포함하여 공통의 비전을 만들어 나가기 위한 다양한 합의과정이 필요하다. 따라서 빅데이터 분야는 다른 정책보다 정책통합 혹은 통합적 혁신정책(성지은·송위진, 2010)이 더욱 절실한 영역이다. 빅데이터가 만병통치약이 아님에도 불구하고 빅데이터 자체의 활용 효과만을 일방적으로 과장하는 행태를 정부는 지양해야 하며, 빅데이터를 통해 해결하려는 사회문제가 더 중심에 자리매김 되어야 할 것이다. 빅데이터로 모든 '데이터 관련 정책'을 포괄하는 것은 불가능하다. 수단이 목표가 되어서는 안 된다.

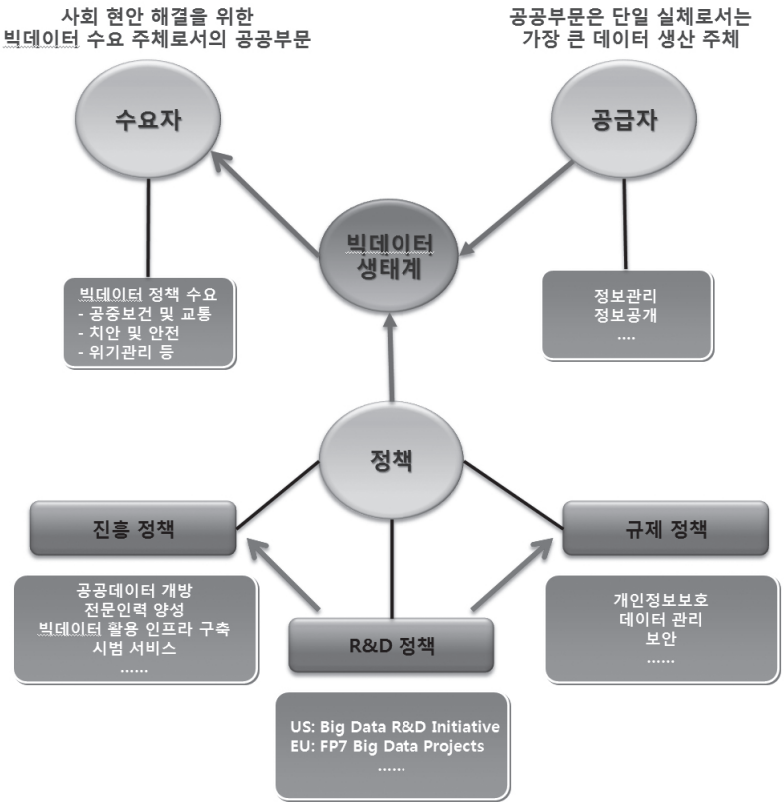


그림 5.1 빅데이터 시대에서 정부의 역할

5.1 현황에서 나타난 한계와 문제점

빅데이터에 대한 관심이 달아오르면서 국내외 유수의 시장분석기관들은 빅데이터가 주도할 거대한 시장규모를 예측·발표하고 있으며, 대중매체에서는 연일 빅데이터 활용 사례를 언급하며 빅데이터 관련 시장의 급성장 과정에서 한국이 뒤쳐지지 않기 위해서 보다 적극적인 관심과 투자의 필요성을 강조하고 있다. 정부 또한 빅데이터 활용의 잠재력을 극대화하기 위해 다양한 정책을 수립·시행하고 있다. 관계 부처 합동으로 '정부 3.0 추진 기본계획'을 발표하였고 현재 60개의 정보공유 및 시스템 연계·통합 과제와 21개의 빅데이터 시범과제를 추진하고 있다. 제3차 과학기술기본계획에서 빅데이터는 30대 중점기술로 선정되었으며, 미래창조과학부는 최근 빅데이터 활용을 통해 사회문제를 해결하고자 일자리, 청소년 복지, 소상공인 지원, 보건의료 분야에서 4개의 빅데이터 사업화 컨설팅 과제를 선정한 바 있다.

한편 빅데이터에 대한 열광과 기대의 반대편에는 여전히 우려와 냉소의 시선이 존재한다. 최근 빅데이터에 대한 기술영향평가에서는 빅데이터 활용능력에 따라 경제적 양극화가 심화될 수 있고, 국내 빅데이터 분석이 외국계 기업에 의존하고 있기 때문에 국가정보의 대외 유출 위험이 초래될 수 있으며, 데이터 독점에 따른 빅브라더 정부의 출현 가능성과 프라이버시 침해 등 다양한 우려의 시선을 보내고 있다(미래창조과학부, 2013). 전사적 자원관리(ERP), 고객관계관리(CRM), 웹서비스 (Web Service) 등 한 시대를 풍미했던 엔터프라이즈 IT 기술처럼 빅데이터도 그냥 유행어에 그칠 것이라는 냉소적인 시각도 제기되고 있다(전승우, 2012).

열광과 기대 그리고 우려와 냉소가 공존한다는 것은 신기술 혹은 미래기술에 대한 자연스러운 사회적 반응이다. 신생기술(emerging technologies)의 발전과정에는 해당 기술의 발전을 촉진하는 유인기제(inducement mechanism)와 저해기제(blocking mechanism)가 동시에 존재하기 때문에 기술을 둘러싼 사회적 갈등은 기술발전과정에 있어서 필연적으로 내재하게 된다(Bergek et al., 2010). 따라서 빅데이터를 새로운 기술패러다임 혹은 새로운 기술혁신체제로 파악하고자 한다면 무엇보다 빅데이터를 둘러싼 기술생태계의 발전을 가로막는 저해기제와 잠재적인 사회적 갈등을 어떻게 해소할지가 중요한 정책적 고려대

상이 되어야 한다. 여기서는 이러한 측면에서 빅데이터를 둘러싼 쟁점을 정리하고 정책과제를 제시하고자 한다.

5.1.1 과잉 기대의 정점에 위치한 빅데이터

빅데이터에 대한 관심이 비교적 최근에 급부상하였다는 데에는 이론의 여지가 없을 것이다. 그림 5.2에서와 같이 2011년 하반기 이후 빅데이터(big data)라는 검색용어의 사용이 급격히 증가하였고, 구글과 네이버의 검색 빈도 증가 패턴이 매우 유사함을 알 수 있다. 반면 미국 톰슨로이터가 운영하는 WoS(Web of Science)에서 'big data'를 키워드로 주제검색을 한 결과는 209편에 불과하였다. 이는 '빅데이터'가 현재까지는 학술적으로 핵심적인 키워드 혹은 독자적인 학문분과로 자리매김 되지 않은 상황을 반영하는 것이다. 즉 빅데이터는 지식의 전통적인 확산과정인 '학계→산업계→사회'라는 경로가 아니라 그 반대인 사회적 관심과 기대로부터 촉발되었다는 특징을 지니고 있다.

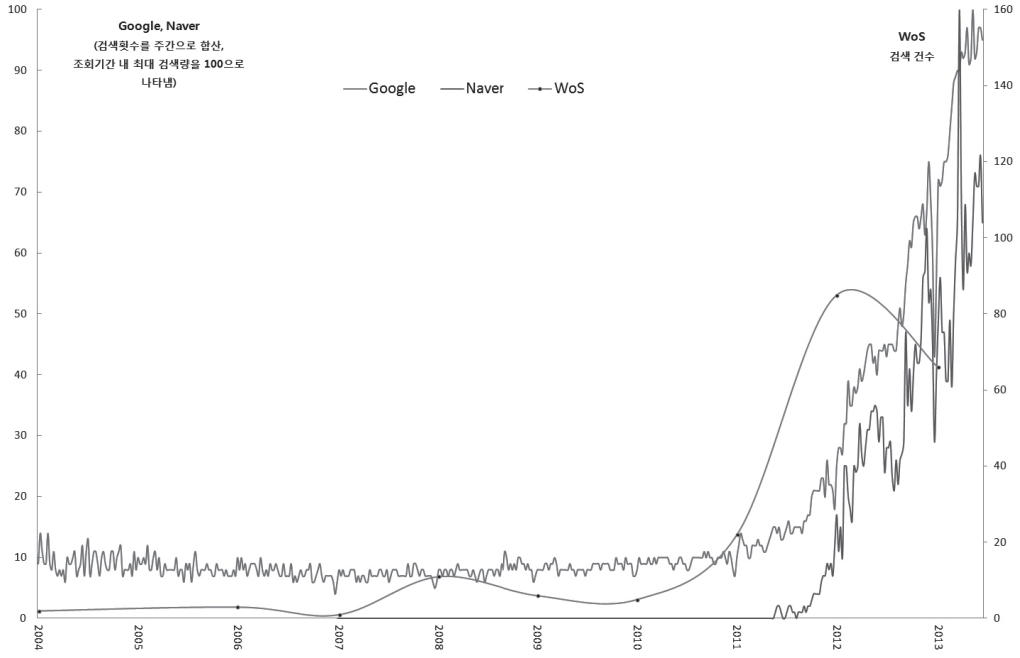


그림 5.2 구글과 네이버의 '빅데이터(big data)' 검색 빈도 및 WoS 학술논문의 시계열 변화

가트너의 기대곡선(hype cycle)2)은 현재 상황에서 빅데이터가 대중매체에서

집중적으로 홍보되고 있는 일종의 과잉 기대의 정점에 서 있음을 보여주고 있다(그림 5.3 참조). 가트너의 기대곡선에 신생기술(emerging technologies)로써 빅데이터가 처음 포함된 해는 불과 2012년이었다. 미국 대선에서 오바마의 빅데이터 활용 전략, 트위터 등과 같은 소셜서비스 분석에 대한 수요 증가, 구글의 플루(flu) 트렌드를 통한 질병의 예측 가능성, 빅데이터를 활용한 일부 기업의 소비자 마케팅 성공사례 등 다양한 데이터 활용사례들이 언론에 집중 조명되면서 이 모든 것을 아우르는 상징으로서 '빅데이터'가 자리매김된 것이다.

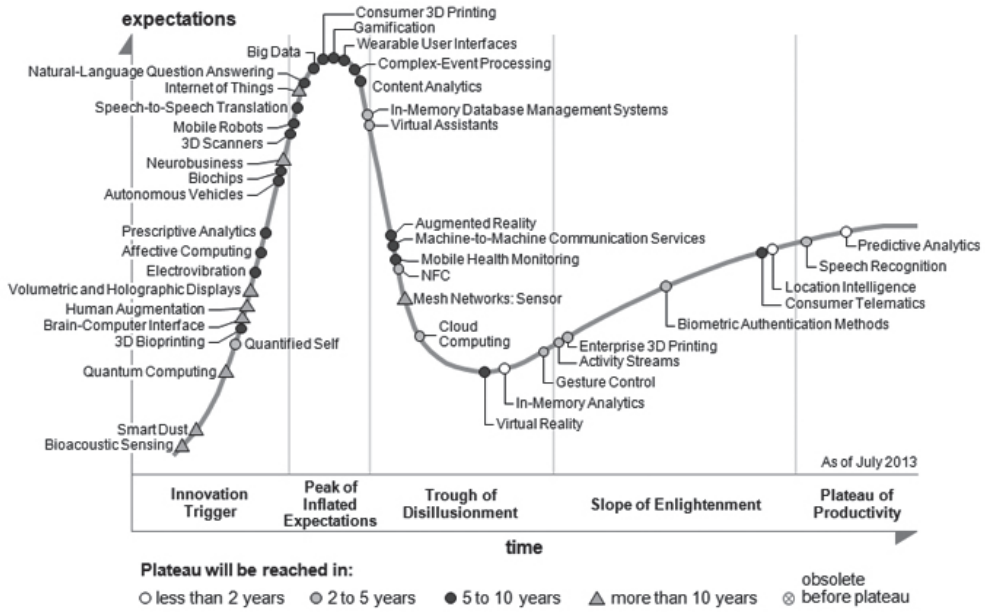


그림 5.3 가트너의 2013 기대곡선

가트너의 기대곡선이 시사하는 바는 특정 기술이 과잉 기대라는 일종의 버블 상태를 거치면서 해당 기술이 과도한 사회적 기대를 충족시킬 수 없어지면서 급격히 매력을 잃게 되고, 이전의 기대가 과장이었다는 것을 인식하는 동시에 부정적인 측면이 강조되어 대중의 관심에서 점차 멀어지게 된다는 점이다. 이러한 환멸기(Trough of Disillus-ionment)를 거쳐 비록 언론의 주목을 받지는 못하지만 해당 기술에 대한 현실적 기대가 충족되면서 시장에서 생존하는 기술로 정착하게 된다. 물론 이 과정에서 도태되는 기술도 존재한다. 빅데이터의 경우 과잉 기대의 정점을 지나 환멸기에 진입했다는 전망도 나오고 있다(Sicular, 2013). 미국의 350개 기업 IT 임원중에 빅데이터 프로젝트로 효과를 보았다는 응답이 7%에 불과하였다는 조사결과는 서서히 거품이 꺼져가는 하나의 징조로 보인다. 이와 같은 급격한 등장에 대한 우려는 현실로 나타나게 된

다. 가트너의 2014, 2015년 기대곡선을 보면, 빅데이터는 2014년은 정점을 지난 하강 국면에서 나타나며, 2015년에는 기대곡선에서 사라지게 된다(그림 5.4와 그림 5.5 참조).

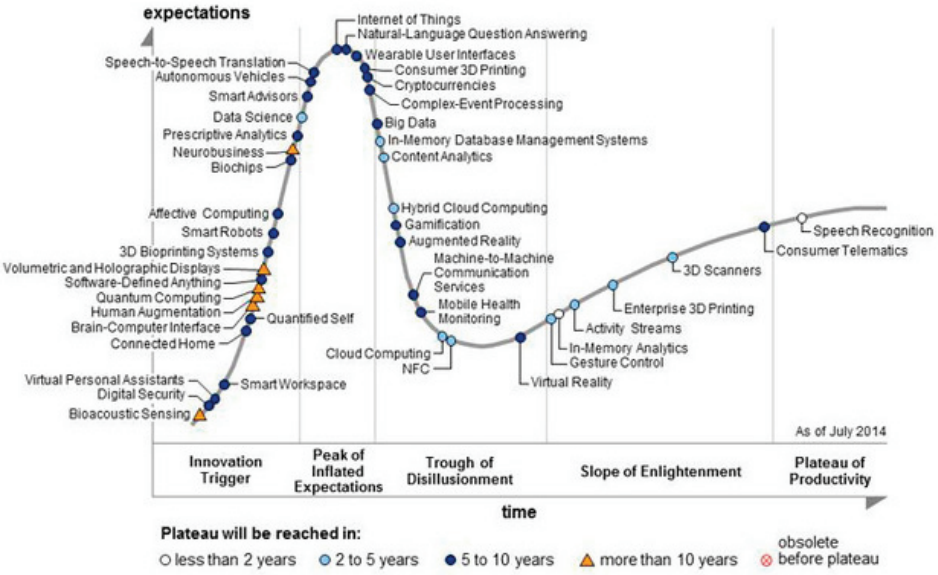


그림 5.4 가트너의 2014 기대곡선

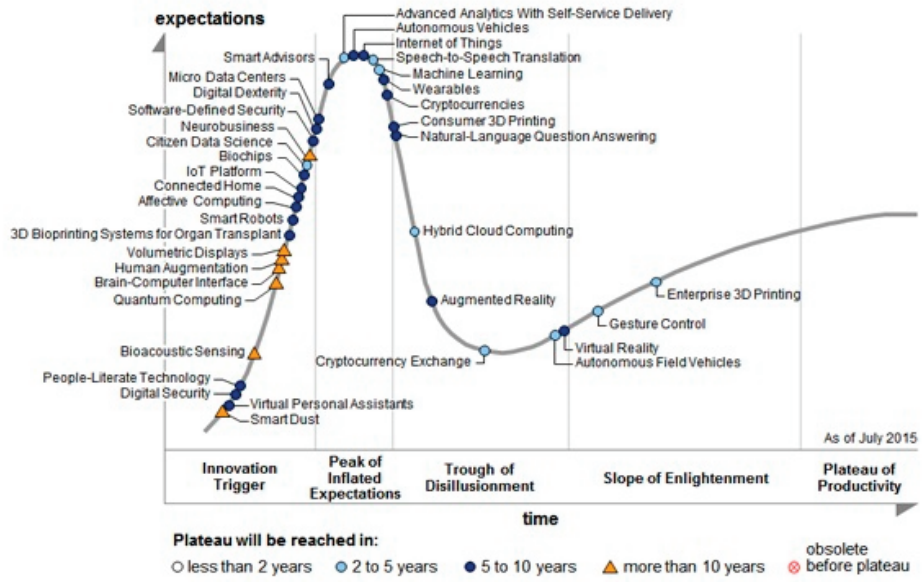


그림 5.5 가트너의 2015 기대곡선

5.1.2 빅데이터를 둘러싼 쟁점들

1. 비정형 데이터의 분석가능성과 한계

일반적으로 정형 혹은 비정형 데이터에 대한 구분은 데이터를 관리하는 시스템(DBMS)에 스키마(schema)가 존재하느냐 여부로 결정하는데, 이런 비정형 데이터에는 텍스트 형식으로 저장된 데이터, 이미지, 음성, 멀티미디어 등이 포함된다. 앞에서 언급한 바와 같이 빅데이터의 급부상 배경에는 무엇보다 SNS로 대변되는 비정형 데이터와 관련이 있으나, 여전히 비정형 데이터의 근본적인 속성들과 관련된 문제가 제기되고 있다. SNS, 스마트폰, 태블릿 PC 등 모바일 환경이 급속히 확장되면서 비정형 데이터에 대한 분석의 중요성 또한 더욱 부각되고 있음에도 비정형 데이터에 대한 분석이 과연 유용한지에 대해서는 회의적인 시각이 팽배하다.

우선 블로그나 트위터 정보 그 자체가 왜곡될 가능성이 존재한다. 블로그나 트위터에 대한 감성분석(sentiment analysis) 혹은 오피니언 마이닝(opinion mining)에는 근본적인 한계가 존재하는데, 예컨대 은어, 사투리, 빈정대거나 과장하는 어투, 역설, 약어 등이 일상적으로 담긴 진술의 의미를 과연 현재의 기술 수준에서 제대로 분석할 수 있을지, 앞으로 가능할지, 텍스트마이닝을 통해 특정 진술에서 그 진술을 대표하는 키워드를 얼마나 정확하게 추출할 수 있는지는 여전히 회의적이다(Global Pulse, 2012). 특히 대규모 데이터를 대상으로 하는 분석에서는 수리모형에 맞게 어떤 형태로든 데이터는 축소되기 마련인데, 편향되고 왜곡된 정보 혹은 주관적인 정보가 많은 비정형 데이터를 수집하고 정량화하는 과정에서 데이터에 내포된 의미가 왜곡되어 분석결과의 신뢰성이 하락될 수 있다는 지적도 있다(Boyd & Crawford, 2012; 전승우, 2012). 이처럼 비정형 데이터를 분석하고 그 결과를 올바로 해석하기까지는 현재의 기술 수준이 따라잡지 못하고 있는 게 현실이다. 가트너는 2012년 빅데이터 요소기술에 대한 기대곡선을 통해 시맨틱웹(Semantic Web) 기술이 안정기에 다다르기까지는 10년 이상이 소요될 것으로 예측하였다. 현재 비정형 데이터에 대한 분석은 맥락을 이해하는 것보다는 단어의 출현빈도, 단어와 단어 간의 관계를 보여주는 것에 그치는 실정이다. 특히, 비정형 데이터가 영어가 아니라 한글일 경우 비정형 데이터의 분석가능성과 한계는 더 명백해진다.

2. 빅데이터를 활용한 문제해결 가능성: 상관관계와 인과관계의 혼동

빅데이터 그 자체가 중요한 것이 아니라 빅데이터를 어떻게 활용하고 분석할지가 더 중요하다는 것에는 대부분 동의를 할 것이다. 그러나 여전히 빅데이터를 통해 도출된 결과가 과연 얼마만큼 신뢰할 수 있을 가에는 의구심이 존재한다. 빅데이터를 통해 도출된 결과는 많은 경우 통계적인 상관관계를 보여 주는 경우가 대부분이다. 그러나 인과관계를 밝히는 것은 전혀 다른 문제이다. 그럼에도 인과관계와 상관관계를 혼동할 경우, 특정 정책적 선택이 전혀 엉뚱한 결과를 보여줄 가능성이 매우 높아진다(Global Pulse, 2012; NESSI, 2012). 빅데이터 분석의 바람직한 사례로 자주 언급되는 구글의 플루(flu. 독감) 트렌드를 보자. 구글의 플루 트렌드가 플루를 포함한 일반적인 호흡기 질환을 예측하는데 어느 정도 도움이 될지도 모르나 플루 자체를 예측하는 데에는 한계가 있다는 지적(Global Pulse, 2012)이 있으며, 구글의 플루 트렌드가 미국 질병관리센터의 예측보다 25% 정도 낮다는 연구결과도 발표되었다(전승우, 2012). 빅데이터는 문제해결을 위한 수단이자 도구이지 그 자체가 해결책이 아니다. 예컨대 공공부문에서 빅데이터를 활용한다고 해서 McKinsey Global Institute(2011) 등이 추정한 것과 같은 비용 절감이나 세수 증대 등의 효과가 정말 나타날 것인가? 공공부문의 도덕적 해이와 부패, 정경유착, 담합구조 등이 단순히 빅데이터로 환원하여 분석한들 그 해결책을 찾을 수 있을까?

3. 빅데이터의 인식론적 한계

Boyd와 Crawford(2012)는 보다 근본적인 차원에서 빅데이터에 대해 인식론적 문제를 제기하고 있다. 데이터 분석의 핵심은 '해석'인데 데이터의 규모와 상관없이 해석에는 연구자의 주관적인 한계와 편견이 내재할 수밖에 없다. 따라서 이들은 마치 빅데이터를 통해 '객관적이고 정확한' 분석이 가능하다고 말하는 것은 실재를 호도하는 것이라고 비판한다. 아울러 규모가 더 큰 데이터가 항상 더 좋은 데이터가 아니라고 주장한다. 이들은 대표적인 비정형 데이터인 트위터를 통해 다양한 패턴을 조사한 학술연구에서조차 스스로 트위터 데이터의 한계를 지적하고 있다며, 무엇보다 트위터가 애초에 분석하고자 했던 모집단을 대표하는 것이 아니라고 지적한다. 실제 분석에 사용되는 빅데이터와 전체 (모집단) 데이터는 동일한 것이 아니기 때문에 샘플 데이터의 특성을 명확

히 설명하지 못한다면 데이터의 크기 자체는 무의미하다고 주장한다. 오히려 Boyd와 Crawford(2012)는 빅데이터보다 'small data'의 가치를 인식하는 것이 점차 더 중요하게 될 것이라고 전망한다.

4. 빅데이터와 기술종속의 문제

빅데이터라는 용어 자체는 구글, 아마존, 페이스북, 트위터 등 글로벌 기업에서 자사의 서비스 개선을 위해 적극적으로 도입된 것이며, 특히 이러한 서비스가 국경을 가로지르는 인터넷에 기반하고 있다는 것에 주목해야 한다(NESSI, 2012). 국내에서도 이미 빅데이터 시장에서 글로벌 기업의 독과점 현상과 향후 기술종속의 문제가 빈번히 지적되고 있다. 현재 IBM, 오라클 등 글로벌 기업들이 국내 비즈니스 분야와 공공부문 빅데이터 처리 분야에서 초기 시장을 선점하고 있는 상황이고(교육과학기술부 외, 2012), 외국계 글로벌 기업에 의존한 빅데이터 분석으로 인해 정보의 대외 유출 등과 같은 위험을 초래할 수 있는 가능성도 매우 높다(미래창조과학부, 2013). 따라서 국내 빅데이터 시장이 주로 글로벌 기업의 기술에 의존하고 있기 때문에 빅데이터 활성화를 위해서는 무엇보다 국내의 데이터 분석 인프라 구축이 시급하다(김한나, 2012). 이러한 상황에서 적절한 대비책이 없다면 정부의 성급한 빅데이터 산업 진흥정책이 오히려 글로벌 기업의 국내 시장 독과점을 고착화하는 바람직하지 않은 정책 결과를 초래할 수도 있다. 아울러 빅데이터 산업이 형성되었을 때 시장에서 가장 높은 부가가치를 창출할 수 있는 정보원천인 이용자(소비자) 데이터는 구글, 페이스북, 애플, 아마존, 야후 등 글로벌 기업들이 이미 만들어 놓은 진입장벽을 극복하기 어려운 게 현실이다. 자칫 국내 시장 활성화라는 작은 욕심으로 인하여 데이터는 글로벌 인터넷 기업으로부터 구매하고, 데이터 처리와 분석은 글로벌 IT 기업에 의존하는 돌이킬 수 없는 소탐대실의 과오를 범할 수 있다.

5. 공공부문의 비효율성과 빅데이터

많은 보고서들이 공공부문의 비효율성을 빅데이터를 통해 어느 정도 개선할 수 있다고 주장한다. 미국과 독일에서는 빅데이터를 활용하여 예산을 절감하였다는 사례도 보고되고 있다. 실제 공공부문에서 낭비되는 예산에 대해 다양한

추정치가 존재하며, 빅데이터를 활용하여 공공부문에서 직접적으로 효과를 얻을 수 있는 영역에서 이러한 비효율성을 제거할 수 있는 가능성은 매우 높다. 그러나 빅데이터를 활용한 공공부문의 비용절감 효과의 규모에 대해서는 논란의 여지가 있다. 예컨대, 유럽의 23개 정부에서 빅데이터로 인한 행정관리 비용 절감 추정액은 15~20%로 추정되며, 이는 1,500억 유로~3천억 유로에 해당된다(McKinsey Global Institute, 2011). 이 추정액에는 공공정보(PSI)에 대한 광범위한 접근과 효과적 이용으로부터 발생하는 편익은 포함되지 않았다고 한다(OECD, 2013). Yiu(2012)는 MGI(2011)의 방법론을 적용하여 영국 공공부문의 빅데이터 활용 가치를 추산한 결과 연간 약 160억 파운드에서 330억 파운드로 나타났다. 그러나 Cebr(2012)은 영국 공공부문의 빅데이터 활용가치를 부정적발에서 20억 파운드, 성과관리에서 36억 파운드 규모의 비용을 절감할 것으로 추정하였는데 이는 2012~2017년까지의 누적액이다. Yiu(2012)와 Cebr(2012)의 전망치는 많은 차이를 보여주고 있다. McKinsey Global Institute(2011)는 운영효율화를 위한 비용 절감의 경우, 전체 예산의 20~25%가 비용 절감이 가능한 영역이고 이중 빅데이터를 활용하여 절감 가능한 비중을 15~20%로 추정하고 있다. 이러한 전망치가 과연 타당한지에 대해서는 여전히 의문시된다.

6. 데이터 개방전략과 빅데이터 정책

빅데이터 관련 정책에 있어서 데이터 개방전략을 마치 빅데이터 활성화를 목적으로 하는 정책으로 오인하는 경우가 종종 있다. 물론 데이터 개방전략이 빅데이터 시장 활성화에 어느 정도 도움이 될 수 있겠지만, 유럽을 중심으로 추진되고 있는 데이터 개방전략은 빅데이터 시장 활성화 그 자체 보다는 기본권으로서의 정보접근권 확대에서 출발했다고 보아야 할 것이다. 주요 선진국의 오픈데이터 정책은 정보접근권을 하나의 인권으로서 보장하기 위해 시행되고 있다. 정보접근권은 “정보에 대해 자유롭게 접근할 수 있는 권리로서 정보의 자유 또는 알 권리의 기초가 되는 개인의 기본적 인권”(안석모, 2013)을 의미하며, 유럽의회의 “행정정보에 대한 접근과 정보의 자유에 관한 권고” 혹은 인터넷 접속권을 표현의 자유와 동등한 기본권으로 규정하거나(EU) 국민의 기본권으로 법제화하는 노력(핀란드) 등을 통해 정보공개 및 정보격차 해소를 위한 목적으로 추진되고 있는 것이다(안석모, 2013). 과연 우리나라 빅데이터에 대

한 논의에서 많은 주목을 받고있는 사례인 영국의 오픈데이터 정책과 EU의 데이터 개방전략(Open Data Strategy: ODS)이 빅데이터 진흥을 위한 적절한 사례인가에 대해서는 더 많은 논의가 필요할 것으로 보인다. 물론, EU의 데이터 개방전략은 단순히 데이터에 대한 접근뿐만 아니라 정당한 목적의 '재사용 권한'도 함께 보장하고 있기 때문에 데이터를 활용한 상업용 서비스를 개발하고 판매하는 것도 허용한다는 측면에서 빅데이터 시장 활성화와 관련이 있을 수도 있다. 그러나 EU의 데이터 개방전략은 '빅데이터' 혹은 '시장활성화' 그 자체보다는 공공정보 및 데이터에 대한 보편적 접근과 개방의 원칙아래 진행되어 왔다고 보는 것이 더 적절하다. 최근 우리나라도 공공데이터 이용권을 폭넓게 보장하는 내용의 「공공데이터의 제공 및 이용 활성화에 관한 법률」이 국회를 통과, 2014년 11월부터 법이 시행되고 있으나, 데이터 개방 이전에 데이터의 생성, 데이터 품질관리, 데이터 연계가 더 중요한 문제로 지적되고 있다.

5.2 개선과 미래 발전을 위한 제언

5.2.1 제도적 구조적 개선 방안

현재 추진 중인 정부의 빅데이터 마스터 플랜은 해외와 마찬가지로 공공 데이터 개방 및 활용에 상당한 비중을 두고 있다. 물론 공공부문 데이터 개방을 통하여 빅데이터 시장 형성이나 사회적 관심도를 높이는 등 서비스 활성화에 전기를 제공할 수 있다. 그러나 향후에는 아직까지 상대적으로 비활성화된 민간 부문 데이터 활용 및 서비스 활성화를 위한 방안도 함께 고려될 필요가 있다. 공공부문 뿐만 아니라 민간 영역에도 빅데이터로 활용할 수 있는 광대하고 다양한 데이터들이 산재해 있기 때문이다.

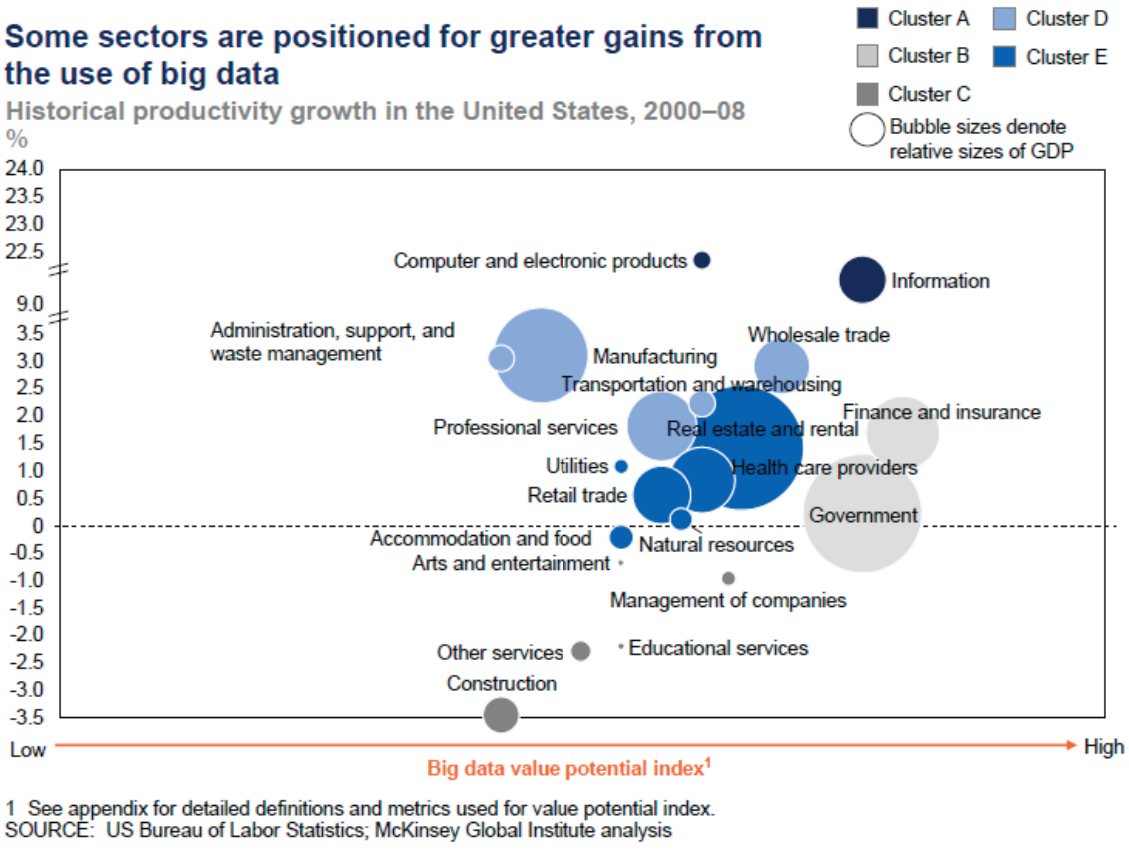


그림 5.6 빅데이터 잠재가치 지수

출처: McKinsey(2011. 6)

McKinsey(2011. 6)에 따르면 정부의 공공 정보는 잠재적 가치는 높으나, 수

집에 어려움이 있는 반면, 인터넷 · 통신 등의 '정보 데이터'는 높은 가치와 수집의 용이성을 가진다고 평가하였다. 이러한 관점에서 정부 보유 데이터 이외에 금융, 유통, 의료 정보 등 다양한 민간정보도 합리적 기준에 따라 공개되고 활용할 수 있는 환경조성이 필요하다. 특히, 공공부문의 데이터와 민간부문의 데이터가 융합되어 활용될 때 정보의 가치와 활용성은 더욱 향상될 수 있다는 점도 고려할 필요가 있다(그림 5.6).

민간부문의 데이터 활성화를 위해서 우선적으로 관련 법제도 개선 논의도 필요하다. 빅데이터 개념이 도입되면서 데이터가 수집/이용/파기 단계에서 새로운 유형의 정보형태가 생성되고 있는데, 관련 법제도 논의는 아직까지 활발하지 않은 측면이 있기 때문이다.

표 5.14 빅데이터 라이프 사이클별 개인정보 보호 이슈

단계	기존 규정	예상 이슈
수집 및 이용	- 필요한 범위 내 수집 - 수집하는 개인정보에 대한 동의 - 이용에 관한 사항에 대해서도 동의	- 가공정보에 대하여 정보주체의 동의 또는 이용사실에 대한 고지 여부
	- 영상정보처리 시 과도한 영상 수집 방지	- 개인정보와 지리정보가 결합된 신규 서비스 적용 여부
활용	- 클라우드 기반 집적화/가공정보는 정보 주체에 고지, 동의	- 개인정보를 세부적인 단위로 구분하여 절차에 따라 정보의 중요도에 관한 규정을 두고 동의를 받는데 한계
	- 개인정보의 국외이전 시 정보주체의 동의 의무	- 국외 이전의 목적, 형태 등을 고려하지 않고 획일적인 동의 절차 - 글로벌 경제활동에 지장 초래
	- 개인정보 제공 목적 이외 용도로 이용 또는 제3자 제공 금지	- 빅데이터를 구성하는 정보를 결합하여 신규 서비스를 창출에 제약
파기	- 개인정보처리자가 보유기간이 경과, 개인정보의 처리목적 달성 등으로 개인정보가 불필요할 때, 지체 없이 파기 의무	- 개인정보가 가공되어 민감정보, 고유식별정보 등을 또 다른 정보의 요소로 사용 가능
저작권	- 창작물을 만든이(저작자)가 자기 저작물에 대해 가지는 배타적인 법적 권리	- 분석되는 가공정보에 대하여 새로운 저작권 개념 도입 및 법적 보호

출처 : 신영진(2012, 9) 재구성

예를 들어, 빅데이터 사업을 추진하는 민간사업자들은 개인정보의 정의가 애매하고 포괄적이어서 법 위반 여부를 판단하고 사업방향을 결정하는데 애로를 겪고 있다. 또한 사업자의 인센티브를 독려하기 위해서 새로운 유형의 정보에 대한 소유권(저작권)에 대한 법적 보호도 수반되어야 한다. 사회적 합의하에 개인정보 보호와 민간 시장의 성장잠재력 간의 적절한 균형점을 모색할 필요가 있다. 이러한 시점에서 2016년 6월30일 국무조정실, 행정자치부, 방송통신위원회, 금융위원회, 미래창조과학부, 보건복지부 등 관계부처가 합동으로 현행 개인정보 보호 법령의 틀 내에서 빅데이터가 안전하게 활용될 수 있도록 하는데 필요한 개인정보의 비식별 조치 기준과 비식별 정보의 활용 범위 등을 명확히 제시한 것은 큰 의의가 있다고 하겠다.

빅데이터 이용이 가장 활발한 미국에서는 학계를 중심으로 빅데이터 관련 법제도 이슈가 활발히 논의되고 있다. 여기서 논의되는 이슈들에는 무차별 복제를 방지하고 가공 · 유통 · 활용을 위한 지적재산권(Copy right) 이슈, 개인정보의 활용, 처리, 삭제 등 정보 보호 이슈, 해킹 대비 보안 시스템을 구축하고 정보 손실 시 관련자에게 고지하는 사이버 보안 및 정보유출 고지 이슈(Cyber Security & Data Breach Notification), 데이터 신뢰성을 보장하기 위해 부정확한 정보활용으로 인한 피해의 책임범위를 규정하는 이슈(Liability) 등이 포함된다. 향후 국내에서 이러한 이슈에 대한 논의가 필요할 것이다.

빅데이터 활성화를 위해서 데이터를 저장, 처리, 전송을 위한 클라우드 서비스 활성화와 네트워크 고도화가 필수적이다. 클라우드 서비스는 효율적 데이터 처리를, 네트워크 인프라는 데이터 전송에 필요한 속도를 제공하여 실시간으로 분석기반을 구축하는데 필수적인 요소들이다. 특히 클라우드 서비스는 빅데이터를 수집 · 관리 · 운송비용을 현격히 낮출 수 있어 빅데이터 상용화에 크게 기여할 수 있다. 한편 빅데이터 활성화를 위해서는 대용량 데이터가 끊김없이 전달될 수 있는 안정적인 네트워크의 구축도 고려해야 한다. 그간 한국의 네트워크 인프라는 세계 최고 수준이라고 인식되어져 왔다. 그러나, 최근 동향을 보면 한국의 우수한 네트워크 기반 경쟁력을 상실되고 있는 상태이다(kt경제연구소, 2013년 1월 한국이 6년여 동안 7%p 보급률 증가에 그친 반면, 영국, 미국, 프랑스는 15%p 내외의 성장, 미국도 100Mbps급 초고속의 홈패스율이 80% 이상이며 Verizon, Comcast, BT 등이 300Mbps급 상품을 출시(또는 예

정)하여 국가 간 격차는 더욱 줄어들 것이라 전망). 앞서 언급한 것처럼 해외 주요국은 네트워크 인프라 진흥정책을 병행함으로써 빅데이터 정책의 효율성을 제고하고 있다. 따라서 향후 우리나라의 빅데이터 정책 논의 시, 네트워크 활성화를 위한 정책방안도 반드시 병행하여 논의되어야 할 것이다.

5.2.2 현실적 이용확산 방안[45)]

1. 빅데이터에 대한 인식 조정

고가의 시스템 하에서 처리되는 절대적으로 큰 데이터라는 생각 보다는 과거에 비하여 상대적으로 규모가 커지고, 연결을 통해 자세한 의미파악이 가능해진 데이터로 시각 교정 필요. 여기에 다양하게 개발된 통계분석기법, 데이터마이닝 기법, 빅데이터 시각화 방법론 등을 효과적으로 사용하는 것으로 인식할 필요가 있다.

2. 방향

빅데이터의 활성화는 다양한 분야에서 생성되는 다양한 데이터를 언제 어디서나 활용할 수 있을 때 가능하다. 이를 위해서는 사회 전체가 지식과 정보의 공유가 가능하도록 해야 한다. 그러나 단순한 공개가 아니라 데이터의 유통을 촉진하고, 타 분야와의 연계될 수 있도록 해야 한다.

특히 개인정보 데이터의 취급을 둘러싼 문제 등 실생활에의 적용에서 발생하는 제도적 과제를 해결해야 하며, 기술 개발에 있어서도 해외 규제의 특성을 반영할 수 있어야 한다. 예를 들어, 미국의 구글과 같은 플랫폼 사업자에 의한 이용자의 행동 이력 등 각종 정보의 활용은 주요 국가에서 개인 정보 보호와 관련하여 다양한 문제를 발생시키고 있다. 이외에도 빅데이터 활용에 필요한 통계 등의 지식을 갖춘 인재 부족 문제를 해결하고, 빅데이터 활용의 경제적 효과와 사회적 문제해결 효과가 가시화될 수 있도록 함으로써 각 산업, 지역, 국가 단위의 이용자에게 빅데이터의 의의를 이해하고 활용할 수 있는 환경을 조성하고 정비해야 하는 과제들이 존재하고 있다. 따라서 본 장에서는 법제도,

45) 이 절의 내용은 '정부3.0 실현을 위한 빅데이터 활용방안(제5장)'의 내용을 옮겨오거나 발췌하여 정리한 것임

운영, 조직측면에서 어떤 정책적 제도 보완이 필요한지 살펴보고자 한다.

3. 원칙

빅데이터의 활성화를 위해서는 세 가지 원칙이 중요하다.

(1) 원칙1: 부처 간 협업

빅데이터의 가치를 얻기 위해서는 분석을 하는 주체가 중요하다. 이는 결국 조직 구조와 직결되어 있다. 빅데이터 분석을 위해서는 현장의 집행부서와 이를 뒷받침하는 연구조직, IT 부서가 결합되어야 한다. 이들 부서 간의 원활한 커뮤니케이션은 정부의 맞춤형 서비스 제공을 위한 필수 조직이라고 할 수 있다. 그런데 이들 조직이 분리되어 있어 국민에게 맞춤형 서비스를 제공하기 위한 분석결과를 얻을 수 없다. 왜냐하면 집행부서는 데이터를 분석할 능력이 없고, 지원부서는 무엇을, 왜 분석해야 하는지 맥락을 모르는 기계적 분석활동이 주가 된다. 따라서 좀 더 효율적인 빅데이터 활용을 위해 조직을 융합하기도 하나 이미 융합이 반드시 효과적이지 않다는 것이 알려져 있다. 이를 위해서는 빅데이터 활용을 명확히 해야 한다.

또 분석할 빅데이터는 정부 및 민간자료를 포함해서 활용해야 한다. 특히 비정형 자료를 어떻게 활용하느냐가 매우 중요하다. 정보를 공유하는 것은 생각보다 쉽지 않다. 특히 국민의 정보를 부처 간 공유할 때 장애요인으로 개인정보를 부처 내 이용하는 것에 대한 저항과 규정이 지나치게 엄격하기 때문이다. 결국 이들 부서간의 협업이 매우 중요하다. 본질적으로 빅데이터는 다양한 부서가 분석 작업에 협업해야 함을 요구한다.

(2) 원칙2: 빅데이터 성과의 객관화

빅데이터란 시스템을 통합해서 정책대상자에 대한 데이터가 모두 합쳐질 때 비로써 정확한 데이터 분석이 가능하다. 즉 데이터 수집이 기본전제 조건이다. 그런데 부처 간 데이터 공유가 되지 않는 상황에서 정책대상자에 대한 분석은 어려울 것이다. 그래서 각 부처의 데이터를 통합하는 작업에 대해 진지한 고민

을 하게 된다. 그런데 이러한 데이터 통합작업은 상당한 비용이 들어간다. 그런데 아직까지 빅데이터 활용이 비용대비 효과가 명확하게 나타나지 않았다. 과거에도 정부는 다양한 시스템 구축을 위해 엄청난 IT투자를 했다. 그런데 최근에 와서는 이러한 투자에 대한 회의적 시각이 있는 것도 사실이다. 즉 빅데이터 활용을 위해서는 상응하는 시스템 투자 및 체제의 확립이 필요한데, 효과가 명확하지 않은 상황에서 최고 책임자는 투자를 주저하는 것이 당연하다. 빅데이터 활용의 투자 효과를 산정하는 것은 쉽지 않다. 따라서 정부가 대규모 투자를 통해 빅데이터에 투자를 하는 것보다는 시범운영 등을 통해 효과를 검증하는 방식이 요구된다.

(3) 원칙3: 시중의 통계분석 소프트웨어를 적극 활용하라

데이터를 분석하는 소프트웨어 분야에서는 통계나 IT지식이 없는 담당자들도 분석 작업을 할 수 있는 시대가 도래 하였다. 통계프로그램을 활용하는 것은 어려운 일이 아니며 그에 대한 해석 또한 충분히 할 수 있는 역량을 갖추고 있기 때문이다.

특히 최근에는 통계프로그램이 사용자 중심으로 바뀌면서 매뉴얼을 활용한다면 좀 더 쉽게 데이터를 활용할 수 있다. 빅데이터 분석 도구의 발전을 활용하면 빅데이터 활성화의 선결과제인 분석인력 확보 문제도 해결할 수 있을 것이다.

CHAPTER 6

행정자료의 통계적 활용방법과 국제적 추이

학습목표

- 우리나라를 중심으로 행정자료의 통계적 활용방법을 주요 특징별로 나누어 학습한다.
- 행정자료 통계적 활용방법에 관한 국제적 추이를 학습한다.

행정자료를 어떻게 통계적으로 활용할 것인가는 하는 것은 사용되는 행정자료의 종류나 내용에 따라 매우 다양할 것이다. 갈수록 행정의 내용이 복잡해지고 행정의 민영화 추세와 함께 행정자료 또한 다양한 기관에서 다양한 내용으로 생성되고 있기 때문이다. 이 장에서는 우리나라를 중심으로 행정자료의 통계적 활용방법을 주요한 특징 별로 유형화하여 설명하고 국제적인 추이를 살펴보고자 한다.

6.1 행정자료의 통계적 활용방법

행정자료의 통계적 활용방법은 통계작성의 목적과 내용에 따라 달라 질 수 있다. 그러나 무엇보다도 이용하는 행정자료의 현황에 따라 활용방법이 달라진다. 아무리 작성하고 싶은 통계의 내용을 잘 설계한다 해도 결국은 현실적으로 존재하는 행정자료의 내용에 따라 결과가 좌우될 것이기 때문이다. 그렇기 때문에 행정자료의 존재형태에 따라 달라지는 다양한 활용방법을 모두 언급할 수는 없으며 여기에서는 행정자료의 현황에 따른 활용방법을 몇 가지로 유형화하여 설명하도록 한다.

6.1.1 행정자료만을 이용한 통계작성

행정자료를 통계적으로 활용하는 방법 중 가장 효과적이고 경제적인 방법은 해당 행정자료만을 이용하여 목표하는 통계를 작성하는 것이다. 그러나 행정자료가 통계작성을 목적으로 생성된 것이 아니라 특정목적을 위한 행정행위의 결과로 수집된 자료이기 때문에, 두 목적이 일치하는 경우가 아니라면 행정자료만으로 통계를 작성하는 것은 흔하지 않은 경우일 것이다. 행정자료만으로 통계를 작성하는 경우는 크게 두 가지로 나누어 볼 수 있다.

1. 단일 행정자료 이용

특정 행정과 관련하여 생성된 하나의 행정자료를 이용하여 통계를 작성하는 경우이다. 대표적인 것으로 주민등록통계가 있다. 행정자치부에서 전국의 읍면동에 설치된 주민등록시스템에 등록된 등록사항을 집계하여 통계로 만든 것이다. 이러한 유형은 정부부처나 지방자치단체, 또는 공공기관에서 고유 행정업무를 수행함에 따라 생성된 행정자료를 이용하여 작성하는 통계로 일종의 행정업무 수행내용 집계결과의 성격을 띠고 있다. 국토교통부의 건축물통계, 환경부의 상수도통계, 국민건강보험공단의 건강보험통계 등이 그 대표적인 사례이다.

여기에 사용되는 행정자료의 형태를 보면, 등록부형식의 데이터베이스 형식, 단

독기관의 행정업무처리내역을 집계하는 형식, 중앙과 지방 또는 본부와 지사의 보고체계를 통해 행정업무처리내역을 집계하는 형식 등 다양하다. 정보통신의 발전에 따라 각 기관에서 고유한 형태의 업무처리용 데이터베이스를 운용하고 있는 경우가 많아 향후, 이러한 유형의 통계작성은 계속 늘어날 것으로 보인다.

2. 2개 이상의 행정자료 이용

단일한 행정자료만으로는 원하는 목적의 통계를 작성할 수 없어 다른 행정자료를 추가로 결합하여 통계를 작성하는 경우이다. 국토교통부의 토지소유현황통계는 토지대장과 주민등록을 연계하여 작성하고, 통계청의 개인별주택소유통계는 건축물대장과 공시가격 자료를 연계하여 작성한다. 여러 개의 행정자료를 연계함으로서 단일 행정자료만으로 파악할 수 없는 다양한 특성변수를 찾아낼 수 있어 행정자료의 통계적 활용을 촉진하고 발전시킬 중요한 요인이 될 것이다. 보다 다양한 행정자료를 통계적으로 활용할 수 있도록 자료공유가 확대될수록 행정자료 간의 연계활용도 증가하게 될 것이다.

문제는 행정자료를 서로 연결할 수 있는 공통의 변수가 있느냐 여부와 행정자료 간의 용어정의, 포괄범위, 분류 등 자료의 특성을 서로 비교 가능한 작성기준의 통일성이 얼마만큼 확보되느냐에 달려있다. 국가의 여러 행정행위와 그것을 기록하는 방법이 체계화되고 표준화 될수록 행정자료의 통계적 활용과 그 가치는 향상되게 될 것이다.

6.1.2 행정자료와 통계조사의 연계

행정자료와 통계조사를 연계하는 가장 큰 전제는 두 자료를 연계하는 동일변수(key)가 존재해야한다는 것이다. 주민등록번호, 이름, 주소, 사업체 명 등이 그것이다. 그 다음으로 중요한 것은 두 자료의 특성(대상, 포괄범위, 시점, 용어의 정의, 분류기준 등)의 일치성 또는 유사성이다. 자료의 특성이 서로 일치하는 정도가 높을수록 연계자료의 품질 또한 높아지는 것은 당연하다. 행정자료와 통계조사를 동시에 통계작성에 활용하는 방법은 크게 두 가지로 나누어 볼 수 있다.

1. 통계조사를 행정자료로 보완

기존 통계조사를 행정자료를 이용해 보완하는 방법 중 가장 전형적인 방법은 통계조사의 특정 조사항목을 행정자료에 있는 항목으로 대체함으로써 응답부담을 경감하고 통계조사 비용 등을 절감하고자 하는 것이다. 이 경우 별다른 문제가 없이 조사되던 특정 조사항목을 단순히 통계조사의 경제적 효율성을 추구하는 차원에서 행정자료로 대체하는 경우와 기존 통계조사의 조사항목 중 응답결과의 질에 문제가 있어 대신 행정자료에 있는 항목을 대신 사용하는 경우이다. 후자의 예로 응답자의 성실한 답변을 기대하기 어려운 소득관련 조사항목을 관련 세무자료 상의 항목으로 대체하는 경우이다.

다음으로는 통계조사로는 파악하기 어렵거나 불가능한 항목의 통계 값을 행정자료의 통계 값으로 대체하는 경우이다. 가령 개인이 정부에 대한 전체 지출내역과 개인이 정부로부터 받는 각종 수입내역을 조사하는 경우, 금전과 물품 등 유형의 지출과 수입은 조사할 수 있지만, 정부로부터 받은 무형의 집합서비스를 개인별로 조사할 수는 없을 것이다. 이 경우 정부세출자료를 이용하여 집합서비스의 개인별 평균수입액을 추산하여 적용하는 대안을 생각할 수 있다.

2. 행정자료를 통계조사로 보완

행정자료를 이용하여 통계를 작성하고자 할 때, 원하는 통계항목이 기존의 행정자료에 없는 경우이다. 해당 행정자료가 통계작성을 염두에 두고 만들어진 것이 아니라 고유의 행정을 수행한 결과물이기 때문에 이런 경우가 발생할 수 있다. 이 때 목표로 하는 통계를 작성하기 위하여 행정자료에 없는 항목을 추가로 조사하는 경우가 있는데, 흔히 해당 행정행위를 수행하는 과정에서 필요한 항목에 대한 부가조사를 하는 방식으로 진행된다. 예를 들면 통계청의 인구동향조사에 있어서 출생, 사망, 혼인, 이혼 등 신고서식을 토대로 작성된 행정자료 이외에 기아발견조서, 실종선고, 혼인취소신고내용을 포함한 인구동향조사사항을 별도로 조사하여 통계를 작성하는 것이 그것이다.

6.1.3 행정자료와 통계조사 간의 벤치마킹

행정자료를 집계하여 얻어진 통계 값과 통계조사 결과 얻어진 통계 값을 상호

비교하여 통계 값을 보완하는 것이다. 이는 행정자료 또는 통계조사의 품질수준을 점검하는 방법으로 사용된다. 두 자료가 각기 가지고 있는 강점을 이용하여 단점을 보완하고자 하는 목적이다. 행정자료는 소관 행정대상 모집단 전체를 포괄하는 정보를 가지고 있다는 장점이 있다. 행정자료로부터 계산된 모수와 표본조사 결과를 가지고 추정한 모수를 서로 수준을 점검해 볼 수 있을 것이다.

또 두 자료가 공통적으로 가지고 있는 각각의 변수도 같은 방법으로 비교 활용할 수도 있을 것이다. 그러나 유의할 점은 행정자료가 가진 각종 특성과 통계조사 자료가 가진 특성에 차이가 존재할 수 있다는 점이다. 그러한 차이를 감안하지 않을 경우 양 자료를 비교하여 얻을 수 있는 이점의 대부분이 상실될 수 있기 때문이다.

6.2 행정자료 활용의 국제적 추이

행정자료를 활용한 통계생산은 이미 국제적 추세이다. 미국, 유럽 등 선진국은 행정자료를 활용한 통계작성이 보편화되어 있다. 대부분의 OECD 국가는 기업체 통계작성에 과세자료를 활용하고, 덴마크, 핀란드 등 북유럽 국가는 인구주택총조사에 행정자료를 활용한 등록센서스를 실시하고 있다(표 6.1 참조).

표 6.1 주요국의 행정자료 활용 사례

국가	행정자료로 생산하는 통계	활용하는 행정자료
미 국	사업체모집단명부, 소규모사업체조사	국세청의 사업체명부, 경영실적자료 등
프랑스	기업통계, 인구통계, 가계소득통계	부가세, 기업수익, 개인소득 신고자료 등
호 주	모집단 구축, 사업체생멸통계 등	사업자등록번호, 기업소득세, 기업활동신고서 등
이탈리아	행정자료기반 센서스	조세대장, 전국인구지표, 학생DB, 고용자명부 등
독 일	행정자료기반 센서스	인구등록부 등
핀란드	등록센서스	주민정보, 부동산 정보 등
스페인	행정자료 기반 센서스	인구등록부, 사회보장파일, 국세청자료, 토지대장 등
스웨덴	등록부 기반의 인구·사업체·노동통계	인구등록부, 사업체등록부, 직업등록부, 토지등록부
덴마크	등록센서스, 등록부 기반 노동력조사	주민등록부, 사업체등록부 등

UN, EU 등 국제기구에서도 응답부담 감소를 위해 행정자료 활용을 적극 권장하는 등 행정자료 활용은 세계적인 트렌드로 자리잡아가고 있다. 유엔 공식 통계 기본원칙(제5원칙 : 비용-효율성의 원칙)은 '통계자료는 통계조사나 행정자료 등 모든 형태의 자료 출처로부터 수집될 수 있으며, 이 때 통계기관은 통계품질, 시의성, 비용 및 응답자 부담을 고려하여 수집방법을 선택해야 한다.' 라고 권고하고 있다.

CHAPTER 7

우리나라 행정자료 통계활용 현황과 과제

학습목표

- 통계청의 행정자료 활용을 통계작성, 행정DB 및 종합통계등록부 구축과저에 관하여 단계별로 학습한다.
- 중앙행정기관과 지방자치 단체, 공공기관의 행정자료 활용 통계작성에 관하여 학습한다.
- 민간 및 연구기관의 행정자료 활용 통계작성, 공공데이터 활용현황에 관하여 학습한다.

우리나라의 행정자료의 통계적 활용은 통계청을 중심으로 시작되어 점차 그 영역이 확대·발전되고 있다. 이를 살펴보면 다음과 같다.

7.1 통계청의 활용현황

정부와 공공기관 중에서 행정자료의 통계적 활용이 가장 필요한 기관은 바로 통계청이다. 통계청은 국가의 중앙통계기관으로서 주요 국가통계를 작성해오고 있으며, 인구·사회·경제 등 국가 전체를 대상으로 하는 대규모 통계조사는 많은 예산과 인력을 필요로 한다. 따라서 통계작성에 투입되는 자원과 통계조사에 따른 응답부담을 경감하라는 압력이 세계의 다른 국가와 유사하게 증가되어온 것은 당연한 일일 것이다. 통계청은 이러한 요구에 부응하고 나아가 통계작성

의 효율성과 품질을 개선하기 위한 노력을 계속해오고 있다. 통계청은 2016년 7월 1일 현재, 57종의 통계를 작성하고 있으며, 통계의 작성방법별로 조사통계(40종), 보고통계(2종), 가공통계(15종)로 구분하고 있다. 행정자료를 통계작성에 이용하는 유형별로 분류해 살펴보면 다음과 같다.

7.1.1 행정자료 통계작성

UN의 행정자료 정의에 의하면 통계작성을 목적으로 수집된 자료는 행정원천이 아니다. 따라서 여기서 행정자료를 이용한다고 하는 경우에서의 행정자료는 행정원천만을 의미한다. 통계청은 이를 다른 행정업무에 수반하여 수집된 자료라고 정의하고 있으며, 이런 행정자료를 이용하여 통계를 작성하는 대표적인 사례가 보고통계(Reporting Statistics)[46]이다. 그 내용을 살펴보면 표 7.1과 같다.

표 7.1 행정자료 이용 보고통계 작성 사례

통계명	주요 내용
국제인구 이동통계	작성방법 : 법무부 출입국 기록을 이용 주요내용 : 출입국 구분, 일자, 항구, 성, 연령, 국적, 체류자격 등 작성주기 : 매년 6~7월(월별 구분 자료)
국내인구 이동통계	작성방법 : 행정자치부 주민등록자료 이용 주요내용 : 주민등록인구, 인구변동요인(전입, 전출, 출생, 사망 등), 전입지, 전입일자, 전출지, 주민등록번호(생년월일, 성) 작성주기 : 매월

이러한 방식은 행정자료가 통계작성에 이용되는 전형적인 형태라 할 수 있을 것이며, 통계작성을 위한 별도의 자료 수집이나 통계조사가 필요하지 않다는 점에서 가장 효율적인 통계작성방법 중의 하나라고 볼 수 있다. 문제는 이러한 행정자료의 보유기관이 관련 자료를 통계작성기관에 제공하느냐 여부에 달려있다. 통계청의 입장에서는 행정자료의 공유가 활성화되어야 이러한 유형의 통계작성이 활발해 질 수 있을 것이다.

46) 법령에 따른 개인이나 단체의 신고, 보고, 신청, 인허가 등과 같이 다른 행정업무에 수반하여 수집된 자료를 이용하여 작성한 통계를 말한다. 제2의 통계라고도 하며, 주민등록인구현황, 공중위생관계업소 실태보고 등이 이에 속한다. 통계청 통계행정편람(2011)- 통계표준용어.

7.1.2 통계조사 항목 대체

기존의 통계조사를 통해 수집하던 자료를 모두 다 행정자료를 이용해 대체할 수 있는 경우란 많지 않다. 따라서 가능한 일부 항목만이라도 행정자료로 대체하는 방식이 차선책으로 강구되어 왔다. 이러한 통계작성기관의 노력은 조사환경의 악화(응답회피)에 따른 응답부담 완화 요구와 통계조사에 투입되는 예산과 인력을 절감하라는 압력 하에서 더욱 가속되어 왔다.

국가의 중앙통계기관으로서 가장 많은 자원을 통계조사에 사용하고 있는 통계청이 이러한 연구와 노력의 선두에 서온 것은 당연하다 할 것이다. 이를 위해 통계청은 가능한 많은 행정자료를 확보하기 위해 다각적인 방안을 추진해 왔다. 행정자료를 입수하기 위해 통계법 개정 등 제도적 기반을 구축함은 물론, 관련기관과의 협의와 설득을 통해 행정자료를 공유하기 위한 다양한 노력을 지속해 왔다. 그 결과 행정자치부, 국토교통부, 국세청 등 주요 행정원천 보유기관으로부터 행정자료를 입수하여 통계에 활용할 수 있게 되었다. 주민등록자료, 토지주택대장, 사업자등록자료, 각종 과세자료 등이 그 대표적인 행정자료이다. 본 장에서는 행정자료를 통계조사와 융합하여 활용한 주요 내용을 살펴보고, 본 절의 활용단계에서 발견된 문제점과 한계와 함께 상세한 행정자료 입수 및 활용현황, 미래발전 전략을 다루게 될 것이다.

실제 많은 통계조사에 있어서 다양한 형태나 비중으로 행정자료가 사용되고 있지만, 이를 명확하게 파악하기는 쉬운 일이 아니다. 따라서 대표적인 사례들을 살펴봄으로서 이를 파악하는 안목을 키우는데 도움이 되도록 한다(표 7.2 참조).

표 7.2 조사통계와 행정자료의 융합 사례

통계 명	주요 내용
인구총조사 주택총조사	작성방법 : 전수조사구(20%)는 통계조사(인터넷조사 병행) (행정자료로 대체 가능한 항목 제외) 표본조사구(80%)는 모든 항목 행정자료로 대체 행정자료 : 주민등록자료, 건축물대장, 가족관계등록부 등 작성주기 : 5년

통계 명	주요 내용
경제총조사	작성방법 : 전수조사 실시(인터넷조사 병행) 행정자료 : 법인세, 사업자등록자료 등 대체항목 : 소규모사업체의 매출액, 자산. 자본, 업종 등 작성주기 : 5년
건설업조사	작성방법 : 협회를 통한 통계조사와 공사기본정보 병용 행정자료 : 법인세, 부가가치세, 공사기본정보(협회) 등 대체항목 : 매출액, 자본금, 건설공사매출액 등 작성주기 : 1년
광업제조업 동향조사	작성방법 : 통계조사와 협회행정자료 병용 행정자료 : 시도별 가스, 대두, 레미콘 생산량 등 대체항목 : 생산량, 출하량, 재고량 등 작성주기 : 월
광업 제조업조사	작성방법 : 시도를 통한 통계조사와 행정자료 병용 행정자료 : 법인세, 부가가치세 등 대체항목 : 매출액(출하액), 영업비용 등 작성주기 : 연
국가자산 통계	작성방법 : 각종 통계자료를 이용한 가공통계 행정자료 : 법인세, 토지(임야)대장, 지적전산자료 등 대체항목 : 재고자산, 유형자산 등 작성주기 : 연
기업활동 조사	작성방법 : 통계조사와 행정자료 병용 행정자료 : 법인세, 부가가치세 등 대체항목 : 매출액, 사업비용 등 작성주기 : 연
도소매업 조사	작성방법 : 통계조사와 행정자료 병용 행정자료 : 법인세, 사업소득신고자료 등 대체항목 : 매출액, 비용, 종사자 등 작성주기 : 연
사망원인 통계	작성방법 : 통계조사(부가)와 행정자료를 이용한 가공통계 행정자료 : 사망신고 자료 대체항목 : 사망원인통계의 기본항목 작성주기 : 연
서비스업 동향조사	작성방법 : 통계조사와 행정자료 병용 행정자료 : 연료판매액, 건강보험료징수액 등 대체항목 : 가정용 및 차량용 연료소매, 자동차판매 등 작성주기 : 월
서비스업 조사	작성방법 : 통계조사와 행정자료 병용 행정자료 : 법인세, 사업소득신고자료 등 대체항목 : 매출액, 비용, 종사자 등 작성주기 : 연
설비투자 지수	작성방법 : 각종 통계자료를 이용한 가공통계 행정자료 : 부가가치세, 수출입자료 등 대체항목 : 설비투자 기준생산액 산정 작성주기 : 월
운수업 조사	작성방법 : 통계조사와 행정자료 병용 행정자료 : 법인세, 사업자등록자료 등 대체항목 : 매출액, 영업비용 등 작성주기 : 연

통계 명	주요 내용
전국사업체 조사	작성방법 : 통계조사와 행정자료 병용 행정자료 : 사업자등록자료, 부가가치세 등 대체항목 : 매출액, 사업장정보 등 작성주기 : 연
전문과학기술 서비스조사	작성방법 : 통계조사와 행정자료 병용 행정자료 : 법인세, 부가가치세 등 대체항목 : 매출액, 사업비용, 종사자 등 작성주기 : 연
지역소득	작성방법 : 각종 통계자료를 이용한 가공통계 행정자료 : 간편장부대상자소득신고자료, 건강보험자료 등 대체항목 : 사업비용, 자산, 부채 등 작성주기 : 연

7.1.3 통계 벤치마킹

행정자료를 통계적으로 활용하는 방법 중 가장 효과적이고 경제적인 방법은 해당 행정자료만을 이용하여 목표하는 통계를 작성하는 것이다. 그러나 행정자료가 통계작성을 목적으로 생성된 것이 아니라 특정 목적을 위한 행정행위의 결과로 수집된 자료이기 때문에, 두 목적이 일치하는 경우가 아닌 경우 일반적인 행정자료에 들어 있는 특정 항목을 통계조사 내의 동일 항목으로 대체함으로써 통계조사의 응답부담과 비용을 줄이는 것이다. 가능한 내용의 통계조사 항목에 의한 대체는 근거 규정에서 기존의 통계조사를 통해 수집하던 자료 모두 다 행정자료를 이용해 대체할 때와 조사통계를 융합하는 것은 구체적인 방법이나 활용된 행정자료가 차지하는 비중 등에 따라 다양한 형태를 지닐 수 있다. 최근 들어 이러한 방식이 인구주택총조사와 경제총조사 등 통계청의 대표적 대규모 통계조사에 적용됨으로써, 조사의 정확성 제고나 예산절감 측면에서 커다란 효과를 거두고 있다.

통계청에서는 172종의 행정자료를 입수하여 이 가운데 40종을 통계를 작성하고 있다. 40종의 통계 가운데 완전대체는 귀농귀촌인통계, 기업생멸행정통계, 영리법인기업체행정통계 등 7종, 일부대체는 전국사업체조사, 건설업조사, 기업활동조사 등 21종, 검증보완은 외국인고용조사, 어가경제조사, 농림어업조사 등 12종이다. 자세한 통계 현황은 부록5~부록7의 완전대체 항목, 일부대체 항목, 그리고 검증보완 항목 표를 참고하기 바란다.

7.2 행정자료 DB와 종합통계등록부 구축[47)]

통계청에서는 국민의 응답부담 경감, 저비용·고효율의 통계생산과 행정자료 활용을 위해 통계단위별로 행정자료DB를 구축을 추진해왔으며, 행정자료 활용 확대기반 마련과 체계적인 행정자료 관리 및 통계생산용 서비스를 위해 구축 중인 행정자료DB 연계·통합을 통한 종합적인 통계등록부 구축을 추진하고 있다.

종합통계등록부는 기 구축된 행정자료 DB와 모집단DB 등 활용부문별 DB에 주소ID, 건물ID, 사업체고유번호를 부여하여 연계하는 것을 말한다(그림 7.1 참조).



그림 7.1 행정자료 기반 분야별 통계등록부간 연계·통합 체계 개념도

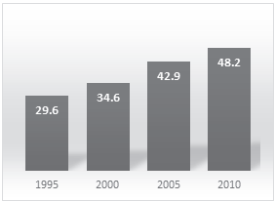
7.2.1 구축배경

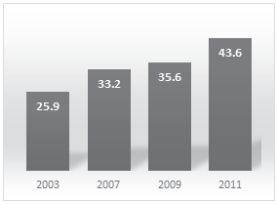
통계 조사환경 악화 및 과중한 응답부담으로 통계조사를 대체할 수 있는 새로운 대안을 강구할 필요성이 대두되었다. 최근 조사환경 변화 내용 가운데 통계등록부의 구축배경이 된 6가지 중요한 항목을 정리하였다.

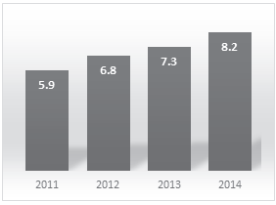
47) 통계청 종합통계기록부 보고서를 바탕으로 재구성함.

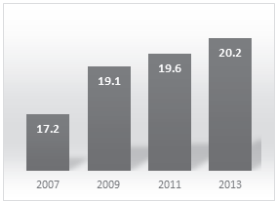
1) 부재·맞벌이 가구, 응답거부 증가 등으로 현장조사가 점점 곤란

1.2인 가구 비율(%)		맞벌이가구 비율(%)		경제활동 불응률(%)		가계동향조사 불응률(%)	
연도	응답비율	연도	응답비율	연도	불응률	연도	불응률
1995	29.6	2003	25.9	2011	5.9	2007	17.2
2000	34.6	2007	33.2	2012	6.8	2009	19.1
2005	42.9	2009	35.6	2013	7.3	2011	19.6
1010	48.2	2011	43.6	2014	8.2	2013	20.2



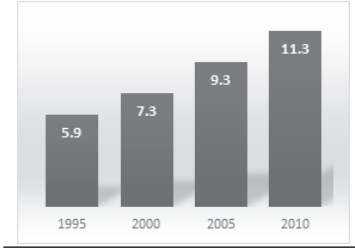


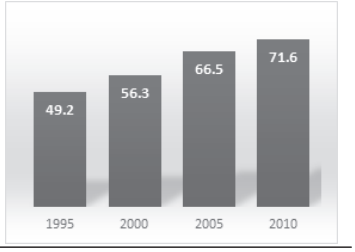




2) 고령화, 외부인 출입통제 주택 증가로 응답자 접근성 약화

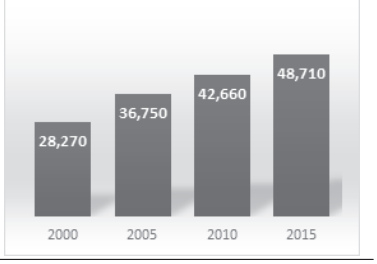
65세 이상 인구 비율(%)		공동주택 비율(%)	
연도	응답비율	연도	응답비율
1995	5.9	1995	49.2
2000	7.3	2000	56.3
2005	9.3	2005	66.5
2010	11.3	2010	71.6





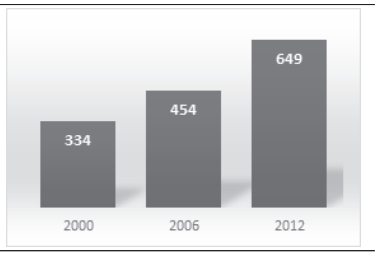
3) 인건비 상승 등으로 통계조사비용은 갈수록 증가 추세

조사인건비(단가)	
연도	인건비
2000	28,270
2005	36,750
2010	42,660
2015	48,710



4) 통계조사 종수 증가 및 항목 확대로 국민·기업의 응답부담 가중

기업대상 통계종수	
연도	종수
2000	334종
2006	454종
2012	649종



아울러 공공데이터의 제공 및 이용 활성화에 관한 법률이 제정 및 시행되어('14.11.19) 공공데이터의 개방과 활용 활성화 및 통계작성 여건이 성숙되었다. 이러한 법적근거 마련으로 통계청은 60여개 기관으로부터 172종의 행정자료를 입수하여 각종 통계작성에 활용하고 있다.

5) 행정자료를 활용한 통계생산 요구 증대

통계청이외의 많은 기관에서 행정자료 활용을 위한 다양한 요구가 진행되어 왔다. 2011년 6월 국회 재정위원회에서는 국세청 과세자료를 활용하여 보다 정확한 통계를 생산할 것을 요구하였으며, 2011년 6월 청와대에서는 국가통계 인프라 강화 차원에서 행정자료 이용에 적극적인 관심을 표명 하였다. 2011년 5월 재정부에서는 비용절감을 위해 행정자료 활용을 적극적으로 권장하였고, 2011년 3월 감사원에서는 성과감사나 정기감사 등을 통해 통계청에 대한 지원 의사를 표명하였다. 또한 국내의 많은 언론기관에서 사회적 이슈에 관한 통계의 부재, 정확도 등에 대한 문제를 수시로 제기 하였다.

6) 통계의 중요성과 행정자료 활용 필요성에 대한 사회적 공감대 형성

행정자료를 활용한 통계생산으로 응답부담 감소, 비용절감, 열악한 조사환경 극복 등에 대한 사회적 공감대가 형성되어 가고 있다. 또한 2007년 4월 통계법에 행정자료(개체 식별자료 포함) 활용에 대한 조항이 신설되었으며, 2009년 2월에는 국세기본법 개정을 통한 과세자료 요구 규정이 마련되었다. 이에 따라 통계청장이 국가통계 작성을 목적으로 과세정보를 요구하는 것이 가능하게 되었다.

아울러 고비용 구조의 총조사를 저비용 구조의 행정자료 기반 센서스로 전환

할 필요성이 제기되었다. 이에 따라 열악한 조사환경에 대처, 응답부담 경감 및 비용 절감 등을 위해 행정자료 기반 센서스를 추진 중이고, 이를 위해 행정자료 DB구축과 종합통계등록부 구축, 자료분석, 표준화·연계·보정 등을 위한 (관리)시스템적 기반 마련을 위해 사업을 추진하고 있다(그림 7.2 참조)



그림 7.2 종합통계등록부 구축 및 관리시스템 개발로 인한 효과

7.2.2 종합통계등록부의 구조와 이론적 배경

1 등록부(Register)

□ 등록부의 정의

등록부는 다음 두 가지의 조건을 갖출 것이 요구된다. 첫째, 목표로 하는 대상에 대한 완전한 목록을 구비할 것이 요구된다. 물론 현실적인 품질문제로 인하여 전체 목록을 다 포괄하지 못할 수도 있다. 둘째, 목록에 있는 각각의 객체의 식별이 가능해야 한다. 이로써 각 객체에 관한 자료의 갱신이 가능해진다. 여기에서 레지스터(Register), 레지스트리(Registry), 카탈로그(Catalog), 디렉토리(Directory), 리스트(List) 등은 거의 동일한 의미로 사용된다.

□ Register Survey

조사대상에 대한 전수조사나 표본조사 등 통계조사를 통한 데이터 수집과정 대신에 기존 행정활동의 결과로 축적된 행정등록부에 존재하는 데이터를 통계적 목적에 맞도록 가공해 가는 과정을 말한다. Register Survey를 수행한다는 의미는 기존의 행정자료를 이용하여 새로운 통계등록부(Statistical Register)를 만드는 것을 의미한다.

<참고> Statistics Canada, Quality Guidelines
'Survey'라는 용어는 일반적으로 통계데이터를 수집하거나 확보하는 모든 활동을 포괄하는 의미로, 표본/전수 조사뿐만 아니라 행정자료를 이용한 데이터의 수집, 가공 통계를 만드는 과정을 포함함

□ 행정등록부를 통계적 목적으로 사용하기 위한 4가지 원칙

① 변형 원칙(Transformation Principle)
행정등록부를 이용해 통계등록부를 만드는 과정에서 필요한 자료들이 통계목적에 적합하게 변형되어야 한다는 원칙

② 시스템 원칙(System Principle)
모든 통계등록부는 하나의 등록부 시스템에 포함되고 통합되어져 모든 데이터가 일관성 있고 효과적으로 사용되어져야 한다는 원칙

③ 일관성 원칙(Consistency Principle)
각각의 register survey의 통계값들이 일관성(coherence)을 갖기 위해서는 모집단과 변수 또한 일관성을 유지할 필요가 있다는 원칙

④ 품질 원칙(Quality Principle)
표본조사 등 통계조사의 경우에는 확률이론과 추정방법론 등 이론적 배경이 튼튼하지만, 행정통계의 품질에 관한 이론적 토대가 부족하여 임시방편으로 등록부 시스템에 대한 품질평가가 필요하다는 원칙

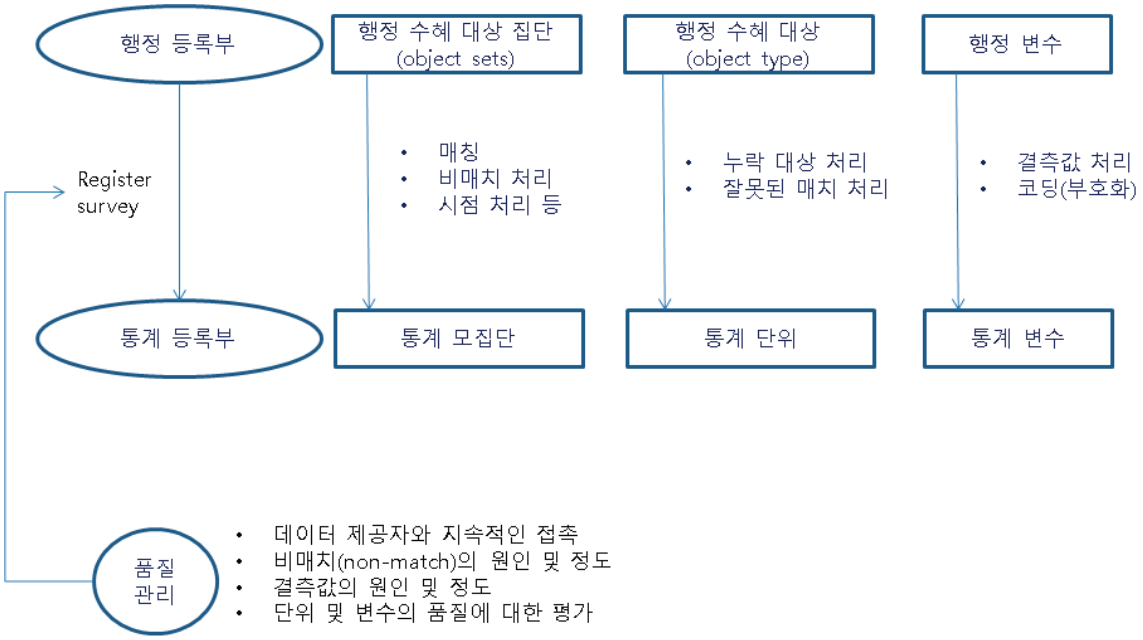


그림 7.3 행정등록부로부터 통계등록부를 만드는 과정

2 등록부 시스템(Register System)

□ 등록부 시스템에서의 객체의 종류와 객체간의 관계

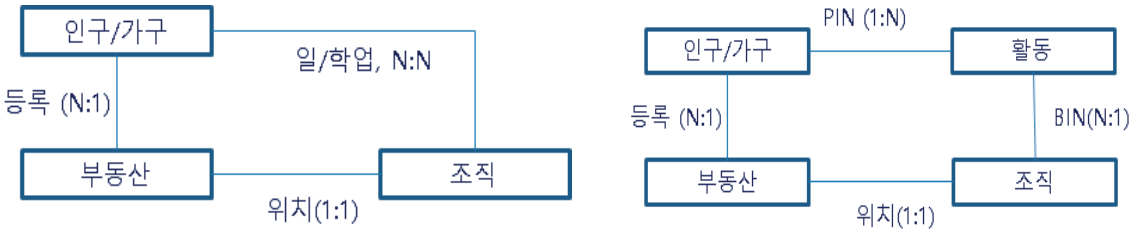
등록부 시스템에서 통계적으로 의미 있는 주요 객체의 종류와 관계를 살펴보면 다음과 같다.

〈객체의 종류〉
- 인구 및 가구
- 조직(기업체, 사업체)
- 부동산(빌딩과 주거지)

〈객체간의 관계〉
- 사람과 조직 간의 관계 : 고용
- 사람과 부동산 간의 관계 : 거주
- 조직과 부동산 : 위치

객체간의 관계가 N(다수):N(다수)일 경우에는 관계를 나타내는 새로운 객체

를 도입하는 것이 좋으며, 이 경우 새 객체(1)와의 관계는 1:N의 관계를 형성하게 된다.



□ 스웨덴 통계청의 등록부 시스템

스웨덴 통계청의 등록부 시스템은 4개의 기본등록부(base register)와 50여개의 등록부로 구성되어 있으며, 50여개의 등록부는 식별변수를 통하여 4개의 기본등록부와 연계되어 있다.

스웨덴은 분산형 통계제도지만 다도 중앙집중적인 요소를 보유하고 있어서 등록부 중 일부는 타 통계기관에서 관리하고 있다. 분산형 통계제도 하에서 일관되게 등록부를 관리하기 위하여 최소한의 등록센서스 및 국민계정과 관련된 등록부를 중앙통계기관인 통계청에서 관리할 필요가 있다.

□ 등록부 시스템의 용도

모집단의 성격을 가진 등록부 시스템은 표본추출틀로 활용된다. 등록부의 변수들은 모집단을 층화할 때 사용되거나 통계의 모수를 추정할 때 정밀도를 높이기 위한 보조변수로 활용이 가능하다. 또한 등록부상의 변수들을 활용하면 통계조사 시에 불필요하게 변수와 중복된 조사항목을 조사하지 않을 수 있다.

□ 등록부 시스템의 비밀보호

등록부에 있는 모든 텍스트 정보는 코드화를 통해 비밀을 보호한다. 공식 식별키는 통계청 내부의 식별키로 전환하고 원 행정자료에 대한 접근권은 최소한의 직원에게만 부여함으로써 비밀유출의 가능성을 감소시킨다.

□ 데이터 입수부서와 데이터의 주제와 관련된 부서의 역할 분담

데이터 입수 부서	주제 관련 부서
입수 시 전송받은 레코드 수, 각 레코드 별 변수 개수, 전송 형식 점검 전송받은 데이터를 통계청 내부 형태로 변환 비밀보호를 위하여 공식 식별키를 내부키로 변환 최초 내검(오류값을 갖는 변수 찾기)	데이터 관련 기관 접촉 ·행정자료에 대한 메타데이터 작성 최종 내검(결측값 또는 오류값에 대한 임퓨테이션) 통계등록부 생성 및 메타데이터 작성

□ 등록부와 시점의 관계

① 현시점 등록부(current stock register): 가장 최신 상태로 현행화된 등록부
② 특정시점 등록부(register referring to a specific point of time)
③ calendar year register: 특정 연도에 존재했던 모든 객체를 포함한 등록부

Calendar year register for 2012

Object identity	Existed 1/1	Added	Ceased to exist	Existed 31/12	Other variables
PIN1	Yes	-	20120517	No	...
PIN2	Yes	-	-	Yes	...
PIN3	No	20120315	20120925	No	...
PIN4	No	20120606	-	Yes	...

④ 사건 등록부(event register): 특정 기간 동안 일어났던 사건들을 기록

Events register for 2012 regarding change of address

Object identity	Address 1/1	Date of change of address	New address
PIN11	Address 11	20120517	Address 21
PIN12	Address 12	20120606	Address 22
PIN13	Address 13	20120911	Address 23

⑤ 히스토리 등록부(historical register): 각 객체의 생애주기 변화를 모두 기록

Historical register regarding change of address

Object identity	From address	Date of change of address	To address
PIN21	Born	19670517	Address 1
PIN21	Address 1	19810606	Address 2
PIN21	Address 2	20120911	Address 3

⑥ 종단면 등록부(longitudinal register): 특정기간 동안의 정보를 기록

Longitudinal register for 2010-2012

Object identity	Existed 1/1/2010	Added	Ceased to exist	Income 2010	Income 2011	Income 2012
PIN31	Yes	-	20110517	183 450	97 600	-
PIN32	Yes	-	-	273 500	281 360	258 340
PIN33	No	20110315	20120925	-	193 570	204 520
PIN34	No	20120911	-	-	-	56 300

□ 변수와 시점의 관계

① 유량변수(flow)는 calenda year register에서 정의되어져야 함

② 저량변수(stock)는 특정시점 기록부에서 정의되어야 함

□ 등록부 시스템에서 표준화 변수

연계키(link key)와 분류(classification)

□ 파생변수(Derived Variable)

① 동일 등록부 내에서 변수의 값을 묶어 class 구간을 만드는 경우

PIN	AGE	AGE CLASS	COB	Geographical category
1	76	70-79	스웨덴	북유럽
2	49	40-49	프랑스	유럽
3	32	30-39	노르웨이	북유럽
4	11	10-19	칠레	비유럽

② 동일 등록부 내에서 다른 변수들을 산술적으로 연산해서 만드는 경우

예) 개인의 가처분 소득= 벌어들인 소득+자본이익+이전받은 소득-세금-이전해준 소득

③ 두 개의 등록부가 N:1의 관계에 있을 때, N측 등록부로 1측 등록부의 변수를 가져오는 경우(by adjoining)

예) 인구등록부(N)와 부동산등록부(1)의 관계를 '실거주지' 관전에서 볼 때 N:1의 관계가 성립하며, 이때 부동산 등록부의 부동산 주소를 인구등록부로 가져오는 경우

④ 두 개의 등록부가 1:N의 관계에 있을 때, 1측 등록부로 N측 등록부 변수의 aggregation 값을 가져오는 경우(by aggregation)

□ 변수의 종류

〈원천에 따른 구분〉
- local primary variable: 행정등록부에서 직접 취한 변수
- locally derived variable: 한 등록부에서 파생된 변수
- imported primary variable: 다른 등록부에서 가져온 primary variable
- imported derived variable: 다른 등록부에서 가져온 derived variable

〈원천에 따른 구분〉
- 식별키
- 소통을 위한 변수(객체의 이름, 주소, 연락처 등)
- 외래키(다른 객체와의 관계를 기술하는데 사용되는 변수)
- 시점
- meta data variable(출처, 주석 등)
- 통계변수(spanning variable, response variable)
  * spanning variable : 분류의 관점을 제시하는 변수
  * response variable : 자극에 대한 반응이나 결과를 나타내는 변수, 영향을 받는 변수로서 결과변수라고도 함. 독립변수의 변화에 따라 변함.

□ 등록부 시스템의 3가지 역할

- 데이터의 원천으로서 통계생산비용과 응답부담을 경감
- 모집단과 변수의 원천으로서 데이터의 일관성을 향상
- survey간의 마이크로데이터 비교를 가능하게 하여 품질향상에 기여

□ 3가지 종류의 통계등록부

	변수 종류	역할
기본등록부 (Base Register)	Local primary 변수: 식별키, 통신변수, 참조변수(외래키), 시점	행정자료 입수 객체집단 생성 객체 정의 분류 변수 생성
1차등록부 (Primary Register)	Local primary 변수: 식별키, 통계변수	행정자료 입수 실제 통계변수 생성
통합등록부 (Integrated Register)	Imported 변수: 식별키, 통계변수, Locally 파생 변수	등록부 시스템에 있는 데이터를 이용하여 새 정보 생성 다른 필드로부터 정보 컴파일 다른 시점으로부터 정보 컴파일

① 1차등록부

- 1개 이상의 행정등록부와 직접적으로 연관된 통계등록부로 많은 통계변수가 행정등록부에서 오거나 이를 이용하여 생성됨

② 통합등록부

- 이미 생성된 다른 통계등록부들을 이용하여 만드는 등록부

  * 통합등록부의 한 형태로 종단면 등록부(longitudinal register)가 있음

③ 기본등록부

· 등록부시스템에서 기본등록부는 백본과 같은 역할을 수행
· 기본등록부의 역할
  - 모든 조사의 객체 또는 통계 단위를 정의
· 기본등록부의 특징
  - 중요한 객체 형식을 정의
  - 중요한 객체 그룹과 표준화된 모집단 정의
  - 다른 기본등록부와의 연계(links) 포함
  - 동일한 객체(통계단위)를 포함하는 다른 등록부와의 연계 포함
  - 표본추출틀의 역할
  - 객체들(사람, 활동, 부동산 등)에 대한 demographic[48] statistics를 산출하는데 사용

48) relating to the dynamic balance of a population especially with regard to density and capacity for expansion or decline

· 기본등록부의 포괄범위 및 분류 변수
  - 기본등록부는 포괄범위가 좋아야 하며(good coverage),
  - 분류를 위한 변수는 높은 품질을 보유하여야 하며,
  - 기본등록부는 시점 정보를 포함하여야 함
· 분류변수에서의 분류에러 처리

〈분류에러가 없을 경우〉

산업	매출	이익	비율
제조업	300	30	10%
서비스업	250	10	4%

〈분류에러가 10%인 경우〉

매출	이익	비율
300-30+25=295	30-3+1=28	9.5%
250-25+30=255	10-1+3=12	4.7%

· 시점과의 관계
  - 3가지 버전의 등록부
  * 사건발생일시, 등록일시, 변화일시 등

RIN	주소 코드	From (사건발생)	등록	To (사건종료)	탈퇴
2141	214	19890301	19890420	20121128	20130201
3244	310	20051101	20060108	NULL	NULL
3523	353	20121121	20130201	NULL	NULL
6229	134	19991201	20000115	20121130	20130201
6229	263	20121201	20130201	NULL	NULL
7048	339	20101001	20101105	20121105	20130201
9566	310	20121220	20130201	NULL	NULL

current stock 20121231	at 20121231	Calendar 2012
1	0	1
1	1	1
0	1	1
1	0	1/12
0	1	11/12
1	0	1
0	1	1

7.2.3 DB구축 및 시스템 개발 현황

다음으로 종합통계등록부 DB 구축 현황에 대하여 살펴보자.

1. 연도별 신규 행정자료 DB구축 현황(2009~2015, 총28개 기관 93종)

구 분	계	'09	'10	'11	'12	'13	'14	'15
입수기관 수	28개*	5개	9개	15개	3개	4개	4개	4개
자료 수	93종	7종	13종	49종	4종	6종	6종**	8종

* 연도별 입수기관수 중복제거 / ** 항목만 추가한 자료(농어업경영체등록자료)는 제외

구 분		계	'09	'10	'11	'12	'13	'14	'15
행정DB	경제	35	6	8	3	3	3	5	7
	농림어업	21	-	2	15	1	1	1	1
	등록센서스	10	1	3	4	-	2	-	-
	건설업	27	-	-	27	-	-	-	-
	합계	93	7	13	49	4	6	6	8
통계용 행정DB *	경제	3	1		-	-	1	1	
	농림어업	14	2		-	8	4	-	
	등록센서스	10	3		2	5	-	-	
	합계	27	6		2	13	5	1	
분석용 행정DB **	경제	13	12		-	1	-	-	
	농림어업	6	2		-	4	-	-	
	등록센서스	6	3		-	3	-	-	
	합계	25	17		0	8	0	0	

* 통계용 행정DB : 행정DB의 자료 중 두 개 이상의 자료를 연계하고 정제를 통해 통계적인 기준으로 처리하여 통계작성을 위해 기초로 구축한 DB
** 분석용 행정DB : 행정DB의 자료 중 두 개 이상의 자료를 연계하고 정제를 통해 조사통계의 검증, 대체, 보완용으로 분석하기 위해 구축한 DB

2. 연도별 행정자료DB 통합관리시스템 개발 현황

년도	주요개발 내용
2009년	- 국세청 사업자등록자료 등 5개 기관 7종 행정자료 DB구축 - 경제부문 자료연계, 조회분석시스템 개발 등
2010년	- 행안부 주민등록자료 등 9개 기관 13종 행정자료 DB구축

년도	주요개발 내용
	- 등록센서스 관련 자료 표준화, 조회분석시스템 구축 등
2011년	- 농식품부 농지원부 등 51종 행정자료 DB구축 - 행정자료를 이용 등록센서스DB, 통계용DB 정확성 제고 - 인구·가구·주택수 추·보정 기능 개발 - 농림어업 자료 표준화, 건설업조사시스템 구축 - 행정자료 포털, 자료관리시스템 구축 등
2012년	- 국세청 소득신고자료 등 25종 행정자료 DB구축 - 국세청 업종코드에 대한 산업분류코드 부여 정확성 제고 - 인구·가구·주택수 추·보정 기능 고도화 - 행정, 조사자료 통합한 농림어업통계등록부 작성 등
2013년	- 국세청 법인합병신고서자료 등 11종 행정DB구축 - 종사자정보 생성, 사업자등록자료 조사구 부여 기능 개발 - 인구 가구 주택 DB 내검 기능 개발 - 농림어업 통계등록명부 생성 및 연계자료 상세항목 조회기능 개발 - 건설업 수집자료 협회등록 시스템 개선 - 외부기관 자료수집시스템 개발 등
2014년	- 국세청 수입배당금명세서 등 7종 행정DB구축 - 학적부 등을 이용한 실거주지 보완, 건축물 생멸 이력관리 기능 개발 - 명부 미연계자료의 수작업 연계기능 개발, 농림어업 통합가구명부 등 3종 명부 작성 및 항목 연계 - 종사자DB 구축, 분석용DB 현황 집계표 개발, 원천 행정자료 입수 내검 기능 개발 - 임금근로일자리 행정통계생산시스템 개발 등
2015년	- 조사자료와 행정자료 간 연계조건 설정기능 개선 - 산업분류 내검 기능 개발 - 과세자료통합 기능 개발, 주소분리 및 표준화관리 기능 개발, 지역·종사자 규모 등 시계열 조회 기능 개발 - 개인별 주택소유 통계 및 귀농인통계 생산시스템 개발 등

3. 행정자료DB 통합관리시스템 주요 기능

구분	주요 기능
등록센서스부문	○ 주소모집단관리 기능 ○ 행정자료 간, 행정자료와 조사자료간 자료 연계 기능 ○ 인구기초 보완, 인구·가구·주택DB 생성 및 관리기능

구분		주요 기능
		○ 인구·가구·주택수 추정 및 보정 기능(SAS) ○ 다문화가구 관리 기능 ○ 집단시설 관리 기능 ○ 건축물 생멸관리 기능 ○ 개인의 상태변동 이력관리 기능 등
경제부문		○ 행정자료와 조사자료간 간 고유번호 연계 기능 ○ 행정자료에 산업분류코드 및 행정구역분류코드 부여 등 기능 ○ 행정자료 사업자·기업체 모집단 생성 및 관리기능 ○ 본지사 연계관리 기능 ○ 3종 이상자료 통합검색 ○ 분석용 자료 생성 기능 ○ 종사자 DB 생성 기능 ○ 행정자료에 조사구부여 기능 ○ 전국사업체조사 자료와 동일한 항목 생성 기능 ○ 과세자료 통합체계 관리 등
농림어업부문		○ 행정자료 간, 행정자료와 조사자료간 자료 연계 기능 ○ 통합농림어업행정등록부 생성 기능 ○ 농림어업통계등록부(명부·항목) 생성 기능 ○ 농림어업총조사(명부·항목) 생성 기능 ○ 행정자료로 보완한 농어가경제조사 명부 및 항목, 농림어업 조사 명부, 세종시 특별조사(농림업) 명부 생성 기능 등
행정통계생산부문	임금근로 일자리통계	○ 임금근로일자리 인적정보 기초자료 추출 및 항목추가 ○ 기업체단위 기초자료 생성 및 보완 ○ 인적정보와 기업체단위자료의 결합 ○ 임금근로일자리 지표 생성 ○ 무등록대체 자료 반영 ○ 통계표 생성 및 조회
	개인별 주택소유 통계	○ 기초자료 추출 및 점검 ○ 전년도 기초자료와 연계 및 비교 ○ 기초자료 생성 및 연계 ○ 통계표 생성 및 조회
	귀농인통계	○ 기초자료 추출 및 점검 ○ 기초DB 구축 및 조회 ○ 귀농가구DB 구축 및 조회 ○ 통계표 생성 및 조회
공통부문		○ 각 업무부문별 주소분리, 주소ID 부여 등 주소표준화 기능 ○ 사용자등록 등 관리자 기능 ○ 행정자료 품질점검(내검) 기능 ○ 행정자료 수집시스템(행정정보중계, 외부기관자료수집시스템) ○ 행정자료 DB 수록 기능(ETL) ○ 원자료에디팅 기능 ○ 행정자료 메타관리 - 행정자료 레이아웃, DB테이블 이력관리 등 ○ 행정자료 DB관리 - 테이블관리, 테이블용량관리, 암호화항목관리 등 ○ 분석용DB OLAP기반 조회분석 기능 ○ 행정자료 홈페이지 관리 ○ 자료추출요청, 승인 및 원격분석시스템 탑재 기능 등

4. 종합통계등록부 개발 현황 및 관리시스템 주요 기능

년도	주요개발 내용
2015년	- 분야별 통계등록부 DB 시범구축('13년기준) - 분야별 통계등록부 간 연계 - 신혼부부 등에 대한 통계용DB 시범구축 및 시산

구분	주요 기능
데이터수록	○ 자동 데이터 수록(SQL관리, 배치관리) ○ 수작업 데이터 수록 ○ 데이터 수록 내검(대상관리, 항목관리, 규칙관리, 실행 배치관리, 결과조회)
통계등록부 연계	○ 통계등록부 자동 연계(대상관리, 마스터관리, 조건관리, 배치작업 관리, 결과조회) ○ 통계등록부 수작업 연계(1:N비교연계, 검색연계, 중복점검연계) ○ 통계등록부 연계결과 내검(규칙관리, 배치작업 관리, 내역조회)
품질점검	○ 품질점검 등록(대상관리, 항목관리, 규칙관리) ○ 품질점검 실행(배치관리) ○ 품질점검 결과조회(결과조회, 이력조회, 현황조회)
메타데이터 관리	○ 메타데이터 관리(테이블 메타관리, DB수록작업 메타관리, 코드관리, DB구축절차 기술서 관리) ○ 메타데이터 조회(테이블 메타조회, 코드조회, DB구축절차 기술서 조회, DB용량 조회)
데이터 조회/분석	○ 데이터 조회(일반검색) ○ 데이터 분석(수준분석)
시스템 관리	○ 사용자 지원(질의응답) ○ 작업 관리(작업자관리, DB수록작업 진척현황 관리) ○ 시스템 관리(사용자관리, 메뉴관리, 권한관리, 소속관리) ○ 이용현황(시스템 이용현황, 조회분석 이용현황)

7.2.4 향후 추가개발필요 사업의 내용

행정자료 DB·종합통계등록부 구축 및 관리시스템 개발을 위해 필요한 주요 내용은 다음과 같다.

1. 행정자료 DB 구축 및 관리시스템

- (DB구축) 행정통계 생산을 위한 주민등록초본 등 원행정DB 구축 및 조사통계 정합성 검증을 위한 경제등록DB(가칭) 등 구축
  - 대출·연체자료 등 민간 입수자료 구축
- (경제·농림어업) 전국사업체조사 등 경제통계 활용성을 높이기 위한 산업분류코드 부여화면 기능 개선, ’16년 농림어업총조사 항목에 대한 ‘16년 입수 행정자료와의 연계 기능 개선 등
- (등록센서스) 등록센서스 결과 공표를 위한 결과표 및 마이크로데이터 작성·KOSIS, MDSS 이관 기능 개발 등
- (행정통계생산시스템) 귀농인통계 정의 변경 및 통계표 추가에 따른 귀농인통계 생산시스템 확대 개발
- (원격분석시스템) 원격SAS 서버 용량 증설 및 자원의 효율적 관리를 위한 개인별 사용량 모니터링 기능 개발 등

2. 종합통계등록부 DB 구축 및 관리시스템

- (사업체/기업체 통계등록부) 사업자·법인 등록번호 미부여 자료의 원인 구분 및 자료 보완
- (인구/가구 통계등록부) 개인식별번호 미부여 자료의 원인분석 및 추가 연계 key 개선
- (주택/건물 통계등록부) 주소ID이외의 건축물관리번호, 행정구역 등 주소 표준화 작업 구축 및 보완
- (경제활동 통계등록부) 경제활동을 일용근로 및 비임금근로 일자리 등으로 확대하여 구축
- (종합통계등록부관리시스템) ‘15년 시범 구축을 보완하여 다년도 통계등록부 본구축 및 등록부간 연계 Key 정비 등
  - 민간자료와 사업체/기업체, 인구/가구 통계등록부 연계 등
- (행정자료포털 개편) 행정자료 홈페이지를 이용자 편의성 및 종합통계등록부 소개 및 구축현황 등을 포함한 컨텐츠를 추가하여 개편
- (생산용 통계DB 구축) 통계등록부간 연계를 통해 신혼부부통계 등과 같은 생산용 통계DB 구축

7.2.5 시스템운영 및 유지관리

1. 목적

시스템운영 및 유지관리의 목적은 공공기관 행정자료의 지속적인 입수 및 DB구축을 통하여 통계조사 항목 또는 조사자체를 대체하고 새로운 통계를 작성에 적극 활용할 수 있도록 지원하고 시스템의 안정적 유지하고 시스템의 성능 향상을 도모하는데 있다.

2. 주요 내용

- 행정자료 DB 관리·운영
  - 2014년까지 구축된 행정자료의 변경관리 및 최신자료 수록

- 행정자료통합관리시스템 등 운영지원
  - 행정자료 추출 및 제공을 위한 시스템 운영 지원
  - 원격분석시스템 및 서버용 SAS 운영 지원
  - 행정통계생산시스템 통계작성 운영 지원
  - 외부자료수집시스템 운영 지원

- 행정자료통합관리시스템의 기능 보완
  - 2014년까지 개발되어 운영되고 있는 시스템의 기능 개선

3. DB 현황

DB현황에 관해서는 표 7.3~표 7.7에 정리하였다.

표 7.3 시스템 주요 기능

구분	주요기능
공통부문 시스템 관리	○사용자 권한 관리 및 시스템·데이터 접근 이력 로깅 기능 ○행정자료 수집시스템(행정정보 중계시스템) ○학적부 등 외부기관자료 수집시스템 ○행정자료 DB 수록 기능(ETL)

구분	주요기능
	○원자료에디팅 기능 ○행정자료 메타관리 ○원자료조회 기능 ○주소표준화 기능
경제부문 행정자료 관리	○행정자료 품질점검(내검) 기능 ○행정자료와 조사자료간 간 고유번호 연계 기능 ○국세청 과세자료에 대해 산업분류코드 부여 기능 ○행정자료에 행정구역분류코드 부여 등 기능 ○행정자료 사업자·기업체 모집단 생성 및 관리기능 ○본지사 연계관리 기능 ○분석용 자료 생성 기능 ○종사자DB 생성 기능 ○행정자료에 조사구부여 기능 ○전국사업체조사 자료와 동일한 항목 생성 기능
등록센서스부문 행정자료관리	○행정자료 품질점검(내검) 기능 ○자료 연계·활용 위한 주소표준화관리 기능 ○주소모집단관리 기능 ○행정자료간, 행정자료와 조사자료간 자료 연계 기능 ○인구기초 보완, 인구·가구·주택DB 생성 및 관리기능 ○인구·가구·주택수 추정 및 보정 기능(SAS)
농림어업부문 행정자료관리	○행정자료 품질점검(내검) 기능 ○자료 연계·활용 위한 주소표준화관리 기능 ○행정자료 간, 행정자료와 조사자료간 자료 연계 기능 ○통합농림어업행정등록부 생성 기능 ○농림어업통계등록부(명부·항목) 생성 기능 ○행정자료로 보완한 농어가경제조사 명부 및 항목, 농림어업 조사 명부, 세종시 특별조사(농림업) 명부 생성 기능 등
행정통계생산부문 시스템 관리	○임금근로일자리 통계 실행 프로세스 편집 및 순서관리 기능 ○임금근로일자리 통계표조회 기능
OLAP을 이용한 조회· 분석 기능	○경제부문 및 등록센서스부문 등 행정자료 조회분석 ○일반검색 : 행정자료/조사자료의 원시자료 조회 ○연계검색 : 행정자료 간, 행정자료와 조사자료간 연계 및 비교 검색 ○수준분석 : 행정자료 및 조사자료의 집계분석 ○연계수준분석 : 연계자료의 집계결과 비교분석
행정자료 홈페이지	○행정자료 DB구축부서 업무 소개(경제, 등록센서스, 농림어업, 건설업부문) ○행정자료 DB구축 절차 및 분야별 DB구축 현황, 행정자료통합관리시스템 개요 등 ○게시판(공지사항, DB구축 및 시스템 개발 상황, 질의응답, 개선의견) ○사용자 등록 및 자료추출 요청관리 ○조회· 분석시스템 사용법 동영상 ○행정자료 레이아웃 조회

표 7.4 행정자료 DB구축 현황(~2014년)

구 분	계	2009	2010	2011	2012	2013	2014
입수기관 수	27개	5개	9개	15개	3개	4개	4개
자료 수	85종	7종	13종	49종	4종	6종	6종

표 7.5 행정자료 DB구축 현황(~2014년)-상세

입수기관	자료명	입수년도	이용부문	비고
국세청 (17종)	사업자등록자료(법인, 개인)	2009	경제	
	부가가치세자료(개인, 법인, 간이)	2009	“	
	법인세자료	2009	“	
	사업장현황신고서	2010	“	
	원천징수이행사항신고서	2010	“	
	일용근로소득지급명세서	2010	“	
	근로소득지급명세서	2010	“	
	사업자단위사업장자료	2011	“	
	사업장별 부가가치세자료	2011	“	
	사업소득지급명세서(원천징수,연말정산)	2012	“	
	사업소득신고자료_복식기장	2012	“	
	사업소득신고자료_간편장부	2013	“	
	법인합병신고서	2013	“	
	개인소득신고자료	2013	“	
	수입배당금명세서	2014	“	
	이익잉여금처분계산서	2014	“	
	특수관계자간 거래명세서	2014	“	
법원행정처	법인등기자료	2009	“	
대법원	가족관계등록부	2010	등록센서스	
국민연금공단	국민연금	2010	경제, 등록센서스	
국민건강보험공단	건강보험	2010		
한국고용원	고용보험	2010		
근로복지공단	산재보험	2010		
농림축산식품부 (3종)	농업경영체 등록자료(농업인)	2010	농림어업	
	농업경영체 등록자료(농업법인)	2010	“	
	축산업등록명부	2011	“	
한국농어촌공사	농지원부	2011	“	
해양수산부 (3종)	어선원부	2011	“	
	어업인허가대장	2011	“	
	어업경영체 등록자료	2014		
축산물품질평가원	쇠고기이력제자료	2012	“	
산림청 (9종)	임업후계자명부	2011	농림어업	
	입목별 벌채허가 및 신고대장	2011	“	
	임산물굴취채취허가(신고)대장	2011	“	
	조경수협회명부	2011	“	
	산림조합원명부	2011	“	
	국유림(대부, 사용허가)신청서	2011	“	
	독림가명부	2011	“	
	관상수시업상황개인별조사서	2011	“	

입수기관	자료명	입수년도	이용부문	비고
	산림사업법인	2011	“	
수협중앙회 (2종)	수협조합원명부	2011	“	
	수협위탁판매자료	2013	“	
각 지자체	내수면어가대장	2011	“	
안전행정부 (2종)	주민등록자료	2010	등록센서스	
	재산세(주택부문)자료	2014	경제	
법무부 (2종)	외국인등록자료	2010	“	
	출입국자료	2011	“	
지식경제부	공장등록대장	2011	경제	
한국신용평가	기업자료	2009	“	
대한상공회의소	기업자료	2009	“	
교육부	대학생기숙사명부	2011	등록센서스	
한국전력	전기사용자명부(주거용)	2011	“	
보건복지부	사회복지시설명부	2013	“	
대학교	학적부	2013	“	
국토교통부 (7종)	건축물대장	2009	“	
	공동주택 공시가격자료	2011	“	
	공사기본정보	2011	건설업	
	수급인정보	2011	“	
	하수급인정보	2011	“	
	임대사업자등록현황자료	2012	경제	
	물류창고업등록자료	2014	“	
조달청 (17종)	공사계약서	2011	건설업	
	공사공동도급	2011	“	
	경영실적여유율	2011	“	
	경영실적여유율내역	2011	“	
	시공능력평가금액	2011	“	
	실적금액	2011	“	
	실적_신인도등급기준	2011	“	
	건설부정	2011	“	
	건설부정업종	2011	“	
	실적_건설안전	2011	“	
	건설재해	2011	“	
	건설환경	2011	“	
	실적_부실처벌	2011	“	
	실적_산재보고	2011	“	
	하도우수	2011	“	
	실적_하도처벌	2011	“	
	실적_행정처분	2011	“	
건설공제조합	재무제표(대차,손익,원가)	2011	“	
전문건설공제조합	재무제표(대차,손익,원가)	2011	“	

입수기관	자료명	입수년도	이용부문	비고
설비공제조합	재무제표(대차,손익,원가)	2011	“	
건설관련협회 (4종)	명부	2011	“	
	일반	2011	“	
	실적	2011	“	
	결산	2011	“	

표 7.6 조사자료 DB구축 현황(~2014년)

구 분	계	2009	2010	2011	2012	2013	2014
자료 수	32종	1종	7종	5종	9종	3종	7종

표 7.7 조사자료 DB구축 현황(~2014년)-상세

자료명	세부자료명	구축년도	이용부문	비고
사업체모집단	사업체모집단	2009	경제	
전국사업체조사	전국사업체조사	2010	“	
2010 경제총조사	2010 경제총조사	2011	“	
기업체모집단	기업체모집단	2010	“	
광업제조업조사	광업제조업조사	2010	“	
도소매업및서비스업조사	도소매업및 서비스업조사	2010	“	
전과기업조사	전문과학기술서비스업조사	2010	“	
기업활동조사	기업활동조사	2010	“	
운수업조사	운수업조사	2010	“	
2010 농림어업총조사	농림어가가구명부	2011	농림어업	
	농가			
	임가			
	해수면어가			
	내수면어가			
농림업조사	농림조사가구명부	2011	“	
	농업가구	2012	“	
	농업가구원			
	임업가구			
	임업가구원			
어업조사	어업조사가구명부	2011	“	
	어업가구원	2012	“	
	어업가구			
	어선			
	양식			
	응답자			
농작물생산조사	경지면적조사명부	2012	“	
	농작물생산조사명부			
어류양식동향조사 대상처명부	농가개체	2012	“	
	소 조사표			
	돼지 조사표			
	닭 조사표			

자료명	세부자료명	구축년도	이용부문	비고
	오리 조사표			
농가경제 조사	명부	2012	“	
	가구원현황			
	출타가구원			
	토지			
	건물			
	부채			
어가경제 조사	어가명부	2012	“	
	표본명부			
	어장현황			
	가구원현황			
	같이살고있지않는가구원			
	토지			
	건물			
	선박기계기구비품			
	무형자산			
	부채			
세종시 특별센서스	(농림업)명부	2013	“	
	(농림업)결과자료	2014	“	
‘14년 가구주택기초조사 (시험조사)	명부	2014	“	
가구주택기초조사 시험예행조사	명부	2014	“	
어업생산동향조사	명부	2014	“	
농림어업총조사	3차 시험조사명부	2014	“	
가축동향조사	가축동향조사 (항목추가)	2014	“	
인구주택총조사 전수자료	전수가구	2011	등록센서스	
	전수가구원			
	전수주택			
인구주택총조사 표본자료	표본자료 가구	2012	“	
	표본자료 가구원			
	표본자료 주택			
인구주택총조사 사후조사	사후조사 가구	2012	“	
	사후조사 가구원			
인구동향조사	사망자료	2012	“	
	이혼자료			
	출생자료			
	혼인자료			
인구주택총조사 시험조사	인구주택총조사 시험조사	2013	“	
지역별 고용조사	지역별 고용조사	2013	“	
가구주택기초조사자료	가구주택기초조사자료 (본조사)	2014	“	

4. 행정자료 운영 업무 흐름도

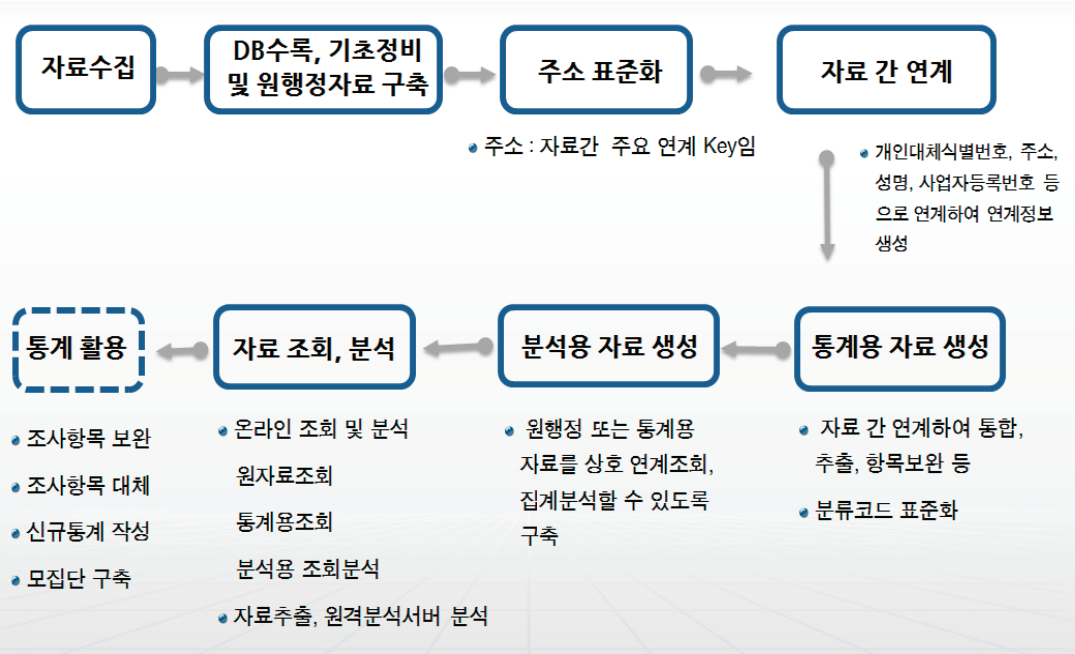


그림 7.4 행정자료 운영 업무 흐름도

5. 행정DB 구축 절차



그림 7.5 전체 흐름도

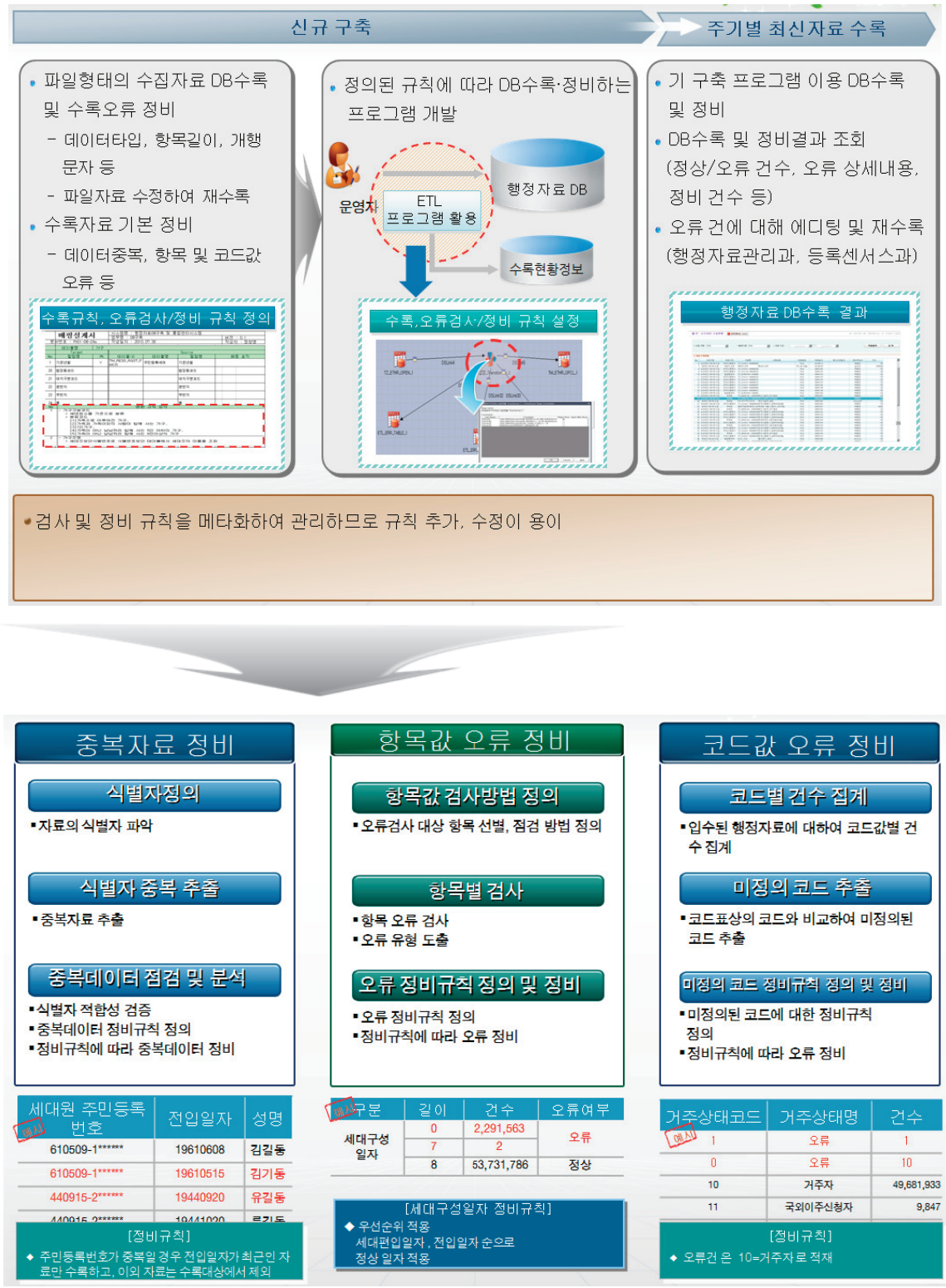


그림 7.6 원행정 DB 구축

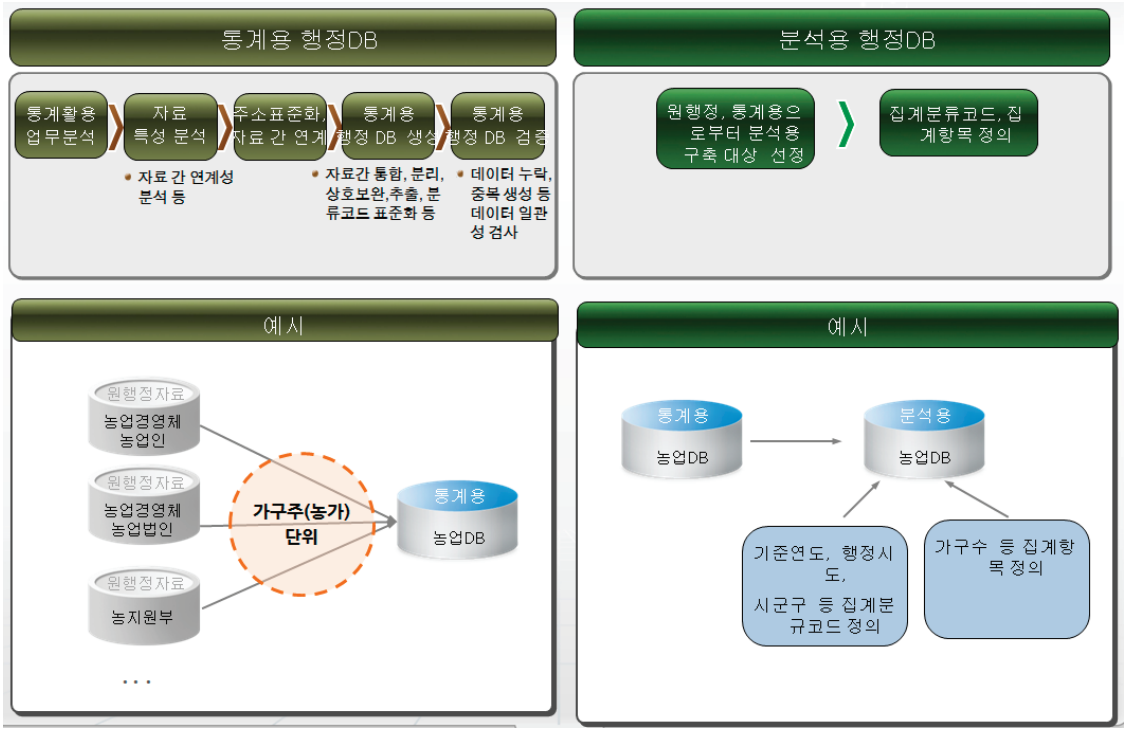


그림 7.7 통계용 및 분석용DB 구축

6. DB 보안조치 및 암호화 절차

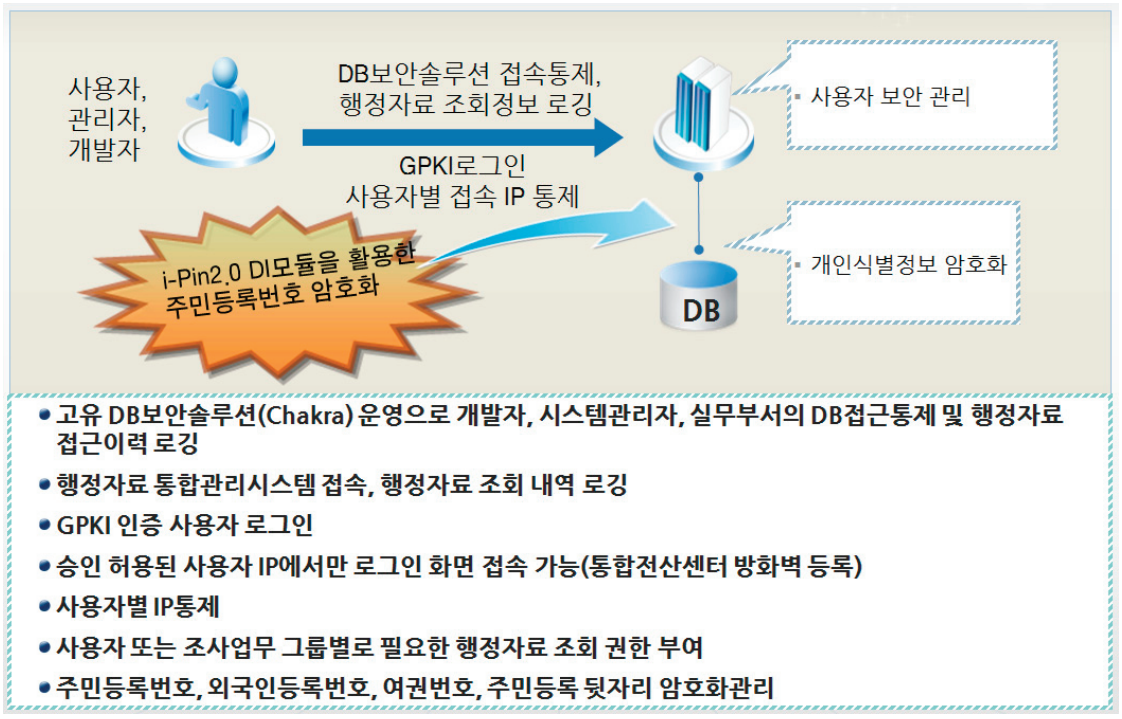


그림 7.8 기술적 보안 조치

7.2.6 종합통계등록부 연계통합

본 절에서는 종합통계등록부 연계통합 절차에 대하여 설명한다.

1) 종합통계등록부 구축 방안

가. 구축배경

구축배경의 첫 번째로는 국가통계 생산을 위한 패러다임의 변화라 할 수 있다. 최근 들어와서 일반 국민들의 통계조사에 대한 응답거부가 증가하고 있다. 이에 따라 국민의 응답부담을 감소시키고 또한 증가하고 있는 조사비용을 절감하기 위하여 행정자료 기반의 통계 생산 방식의 선진화 전략을 수립하고 추진하여야 한다. 또한 2009년부터 입수되는 172종의 행정자료의 효율적인 활용을 위하여 유기적인 연계 및 통합을 통한 통계등록부 구축이 필요하게 되었다.

다음으로는 국민 맞춤형 통계 개발 및 개선이 필요하게 되었다. 경제의 활성화 및 국민 생활과 밀접한 통계 수용에 대응하기 위하여 다양한 신규 통계의 개발 및 개선이 요구되면 또한 기존의 통계를 최근 추세를 따르고 정확성을 제고하여야 한다.

나. 연혁

종합통계등록부 연계통합을 위한 추진 경과는 다음과 같다.

- 종합모집단 구축을 위한 정보화전략계획(ISP) 수립('14년) - 내·외부 환경분석, 선진사례 분석, 목표모델 정의, 종합모집단 DB의 구성·관리 방안 및 단계별('15 ~ '17) 추진전략 수립

- 종합통계등록부 3개년 계획 추진

- ('15년) 분야별 통계등록부('13년기준) 구축, 연계시스템 개발 및 신혼부부 통계 시범구축

- ('16~'17년) 통계등록부 구축 확대('10~'15년), 연계·변동관리 시스템 고도화, 통계 생산용DB 시범구축 및 생산지원 확대

다. 향후 구축 방향

종합통계등록부 연계통합을 위한 향후 구축 방향에 대하여 알아보자.

□ 2015년 시범사업 결과를 반영한 통계등록부 연계·통합 개선

2015년 시범사업의 결과를 반영한 연계 및 통계 개선 방향은 구축확대와 표준화로 진행한다. 구축확대는 시계열 분석 및 통계 작성이 가능하도록 2010년부터 2014년 까지 작성된 분야별 통계 등록부를 구축 활용 한다. 또한 효율적인 통계등록부 연계를 위하여 주요 연계항목에 대한 표준화 및 일치화 그리고 연계 로직 고도화를 추진한다.

다음 표는 기존 행정자료 DB와 종합 통계 등록부의 장단점을 비교한 표이다.

표 7.8 기존 행정자료DB와 종합통계등록부 장·단점 비교

구분	행정통계DB	종합통계등록부
장점	·입수 즉시 신속한 DB구축 가능 ·조사자료의 항목대체 및 보완시 간편하게 활용 가능	·연계된 통계등록부를 이용하여 다양한 통계 개발 수요에 대응 가능 ·통계등록부간의 연계를 통한 대용량 데이터의 신속한 처리 향상
단점	·행정자료간 공통된 연계키 값 부재 및 정합성 결여로 개별 행정자료 및 조사자료와의 연계 곤란 ·행정DB간, 행정DB-조사자료간 기준시점과 입수시기의 불일치 ·연도별 입수된 자료를 추가하여 DB적재하므로 항목의 시계열 관리가 부재	·초기데이터 구축으로 인한 구축비용 소요 ·이상적인 개념이므로 단기간에 완성하는 것은 불가능함 ·통계등록부간 연계 매칭률 확장 한계

다음 표 7.9는 통계등록부 활용에 대한 해외 사례를 정리한 표이다.

표 7.9 통계등록부 활용 해외 사례

국 가	활용자료	활용분야	비고
노르웨이	• 인구·교육·소득 등록자료 • 토지재산, 주소 및 건물등록자료 • 소득·사회보장 등록자료	• 인구센서스 • 단기경제통계	분야별 통계등록부를 구축하여 통계생산에 활용
덴마크	• 인구, 건물/거처 등록자료 • 사업자, 세부사업, 사회보장 등록자료 • 사회보장등록자료	• 인구센서스 • 단기경제지표	
핀란드	• 주민정보, 부동산, 기업정보 시스템 • 고용, 구직자, 조세, 연금 등록자료	• 인구센서스 • 건물, 주거, 주택현황통계 • 소득분포통계, 임금통계	

□ 통계작성 기반 마련을 위한 생산용(주제별)DB 시범구축

두 번째 구축 방향은 통계작성 기반 마련을 위한 생산용(주제별) 데이터베이스를 시범 구축하는 것이다. 구축된 종합통계등록부를 활용하여 작성대상 선정, 필요항목 추가, 내용검토, 통계표 작성 등의 통계작성 시스템을 개발을 위한 기반을 구축하고 제2차 중기 행정자료 활용전략의 신규통계 개발 대상인 중·장년층통계, 자영업자통계에 대한 생산용DB를 시범구축 하고자 한다.

생산용 DB 구축의 하나의 사례로 신혼부부통계 구축을 살펴보자. 구축과정에서 연도별 변동내역을 확인하기 위해 해당 자료와 연계작업(혼인·이혼)을 반복 실시하였고 자료 간 대상 중복 발생하였다. 다음 그림 7.9는 신혼부부통계 생산용 DB 구축과정을 정리한 그림이다.

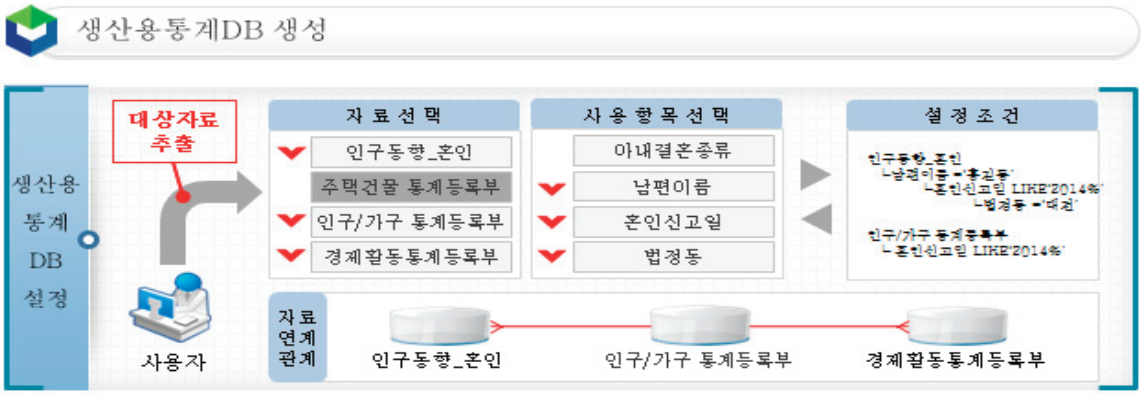


그림 7.9 신혼부부통계 생산용 DB 구축과정

□ 정책 맞춤형 통계생산 지원 가능

세 번째 구축방향은 정책수립에 필요한 대상을 파악하고 기초현황에 대한 분석을 실시하고 통계간 불일치 축소 등을 위한 정책 지원을 가능하도록 한다.

□ 통계작성기관 생산통계 지원 가능

네 번째 구축방향은 외부 통계작성기관 및 다양한 분야의 통계활용을 위하여 One-Stop으로 통계작성에 활용할 수 있도록 구축함으로써 통계의 활용을 높이고자 한다. 이를 위하여 조사통계의 모집단 구성, 다양한 기준의 표본추출, 관련 통계조사의 보완 등에 활용할 수 있도록 구축한다.

라. 기대효과

종합통계등록부 구축에 따른 기대효과는 다음과 같이 정리할 수 있다.

- 사회·경제 융합형 통계 등 신규통계 생산 및 분석 기틀 마련 - 고용 관련 인구/가구 부문과 사업체/기업체 부문 간 연계 등 융복합(Socio-Economic) 통계의 생산이 가능해진다.

- 시계열·종단면 분석, 패널통계 등 장기 데이터 분석 - 연계키(Key)에 기반을 둔 종합통계등록부 구축은 연도별 DB를 활용한 시계열 분석, 패널통계 등 데이터 분석 역량을 제고시킨다.

- 타 통계작성기관의 통계생산 지원 - 통계작성기관의 통계조사 표본추출, 조사항목 대체 등을 통해 응답부담 경감 및 통계생산 지원 가능하게 한다.

2) 종합통계등록부 시범사업 실적

□ 분야별 통계등록부 시범 구축('13년 기준)

- (사업체/기업체) 사업자등록자료와 전국사업체조사자료를 기반으로 자료간 연

계, 기업단위 생성 및 중요항목(산업분류, 종사자, 매출액 등)을 추가하여 구축
- (인구/가구) 등록센서스의 인구DB와 가구DB를 기반으로 주소, 개인대체번호로 연계하고 중요항목(가구구분, 가구관계, 거처 등)을 추가하여 구축
- (주택/건물) 등록센서스의 주택DB를 기반으로 건축물대장자료, 건축물기초자료, 주택공시자료 등을 연계하여 중요항목(주택형태, 면적, 건축년도 등)을 추가하여 구축
- (경제활동) 임금근로일자리통계 기초자료를 기반으로 임금근로, 비임금근로 등을 반영하고 중요항목(근무기간, 임금, 사업장업종 등)을 추가하여 구축

□ 신혼부부 생산용DB 시범 구축

- (개요) 분야별 통계등록부의 기본항목들과 신혼부부 대상선정, 경제활동(임금근로, 대표자 등), 주택(소유, 임차) 관련 행정자료 연계를 통해 신혼부부에 관한 통계표를 작성

- (작성절차) 종합통계등록부를 기준으로 주제에 따른 행정자료 DB(인구동향자료, 주택임대차확정일, 재산세)를 연계하여 구축
  ① 신혼부부 기초자료: 인구/가구통계등록부와 인구동향자료(혼인, 이혼, 출생)를 연계하여 신혼부부 기초데이터 구축
  ② 신혼부부 경제활동, 주택소유: 신혼부부 기초자료와 종합통계등록부를 연계하고 주택임대차확정일자, 재산세 자료를 추가 연계하여 경제활동(임금근로, 사업체경영 등), 주택소유(주택소유, 전·월세)데이터 구축 및 통계표 작성

- (최종결과) 혼인건수 1,614,347건 중 재혼, 사망, 해외, 이혼 등이 확인된 경우를 제외한 1,453,876건을 신혼부부 기초자료로 구축하였고, 신혼부부 경제활동, 주택소유 관련 통계 시범작성

7.3 중앙행정기관과 지방자치단체의 활용현황

우리나라의 국가기관에서 작성되는 행정통계는 2016년 7월 26일 기준 300여 기관에서 452종이 작성되고 있으나, 이 절의 내용작성을 위한 자료취합 기준일인 2015년 6월 30일에는 310개 작성기관에서 445종의 통계를 작성하고 있다. 이 가운데 중앙행정기관에서는 139종, 지방자체단체에서는 261종의 행정통계가 작성되고 있었다.

시도·시군구 단위에서 단위에서는 행정자료를 활용하여 작성하는 기본통계가 24종, 보건·복지·사회 부분이 45종, 인구와 교통·정보통신 부분이 22종으로 작성되고 있다. 다음 표 7.10은 중앙행정기관과 지방자치단체에서 작성하는 통계의 수를 정리한 표이다.

표 7.10 기관별 작성현황(2015년 6월 30일 기준)

기 관	작성통계수	작성방법		
		조사	보고	가공
계	935	415	445	75
정부기관	767	311	400	56
· 중앙행정기관	340	172	139	29
· 지방자치단체	427	139	261	27
민간지정기관	168	104	45	19

1) 행정자료를 활용한 통계작성

기본통계는 총 224종으로 전체의 55%를 차지한다. 인구, 노동, 사업체, 보건 및 사회보장, 환경, 교육 및 문화, 소득 및 지출 등 19개 부문 413개 통계표가 작성된다. 각 지자체의 업무수행 결과로 생산되는 행정자료 중 대외적으로 활용이 필요한 공통적인 지표로 구성한다. 다음 표는 1995년부터 2014년 까지 연도별로 작성된 통계표의 숫자를 정리한 것이다. 1995년도에 339건 이었던 통계표가 매년 꾸준히 증가하여 2014년에는 413개로 증가한 것을 확인할 수 있다.

표 7.11 연도별로 작성된 통계표의 수

연도	부문	통계표	연도	부문	통계표
1995	18	339	2005	19	389
1996	18	352	2006	19	391
1997	19	366	2007	19	391
1998	19	372	2008	19	386
1999	19	384	2009	19	383
2000	19	384	2010	19	379
2001	19	387	2011	19	376
2002	19	387	2012	19	409
2003	19	388	2013	19	408
2004	19	389	2014	19	413

□ 서울서베이

서울서베이는 조사와 행정자료가 결합된 도시사회정책지표 체계로 작성되는 통계이다. 총 190여개의 지표 중 120여개는 행정자료에서 추출하고, 나머지는 사회조사성격의 조사자료에서 산출한다. 이러한 조사자료와 행정자료의 결합은 기존 행정자료의 데이터 유지관리에도 유용한 자극을 주며 새로운 현상을 조사자료와 행정자료를 결합하여 다양한 형태의 정책발굴이 가능하게 하는 주요 원천으로 이용될 수 있다. 다음 표 7.12는 서울서베이의 영역별 주요 발굴 주제를 정리한 것이고, 다음 그림 7.10은 서울서베이의 주요 조사영역을 정리한 것이다.

표 7.12 영역별 주요 발굴 주제

영 역	주 제
인구	• 인구구성의 변화와 싱글가구 → 인구센서스 데이터와 부동산 자료 결합 • 2인가구와 가족구성의 변화 • 신고령세대 거주특성 집중분석 → 센서스데이터와 국토부 주택실태조사 자료의 결합
주거	• 아파트 문화와 중산층 주거 취향: 지속가능한 도시주거 탐색
복지	• 사회서비스에 대한 복지 우선영역에 관한 컨센서스
가치와 의식	• 세대갈등과 조정 • 감성중심 소비 • 기술융합과 안전 • 다원주의



그림 7.10 서울서베이 조사영역 개괄

2) 행정자료 정책활용 현황

다음으로 행정자료를 이용하여 주요 정책에 활용한 사례 현황에 대하여 살펴보자. 광주광역시 광산구에서는 'GIS로 그리는 광산구 행정지도'를 통하여 GIS 정보를 가지고 주요 정책을 결정하는데 있어서 과학적 근거를 제시하고자 하였다.

먼저 내부행정자료, 유관기관, 경찰서의 협조를 통하여 데이터베이스를 구축하고 지역의 주민들이 살아가는 인문·사회·공간적 특성을 GIS지도에 중첩시켜 도시가 지니고 있는 가능성과 장점, 기회적 측면을 시각적으로 구현 하고자 하였다. 이를 통하여 주요 정책결정시 합리적 의사결정의 근거자료로 활용하고 및 부서별 주요 행정업무의 예측가능성을 높이고자 하였고 또한 정확성을 제고하고자 하였다.

도시관리 측면에서는 공간 특성을 구체적으로 진단해 광산구가 가지고 있는 문제점에 대한 해결 방안을 모색하고자 하였다. 다음 표 7.13은 광산구내 부서별 GIS 정보에 대한 활용 계획을 정리한 표이다.

표 7.13 부서별 GIS 정보 활용 계획

부서명	활용계획	부서명	활용계획
기획관리실	·지역현안 발표 및 사회공헌 사업 유치	보건소	·신규 보건지소 입지분석
복지정책과	·수급자, 차상위자, 외국인 관련시설 입지 분석	교통과	·버스노선 증설 및 대중교통 편의성 제고방안 검토
노인장애인 복지과	·장애인 복지예산 지원	도시개발과	·친자연 걷기 좋은 길 조성 방안 마련
여성아동보육과	·국공립보육시설 확충방안	방재관리과	·자연재해 피해 예방지도 제작

다음 세 그림(그림 7.11~7-13)은 GIS 정보의 활용에 대한 예로 각 지역의 자살 증가율, 10대 인구분포, 범죄지도이다.


광주일보 2012년 7월 25일 수요일

광주 자살 증가율 최고

2010년 442명 3년새 45% 급증-노인 자살도 늘어

전문 센터 확대·관련기관 협조체계 시급

'09년~'11년 변사사건

변사사건 | 543회 중 정위치 확인된 366건 맵핑

시청 및 구청

밀집가구 밀집지역

APT호당 5권역이하

변사사건 분포

높음

낮음

북구

광산구

서구

변사사건 제작자료 : 광주광역시 생공

그림 7.11 GIS 정보 활용 - 자살증가율

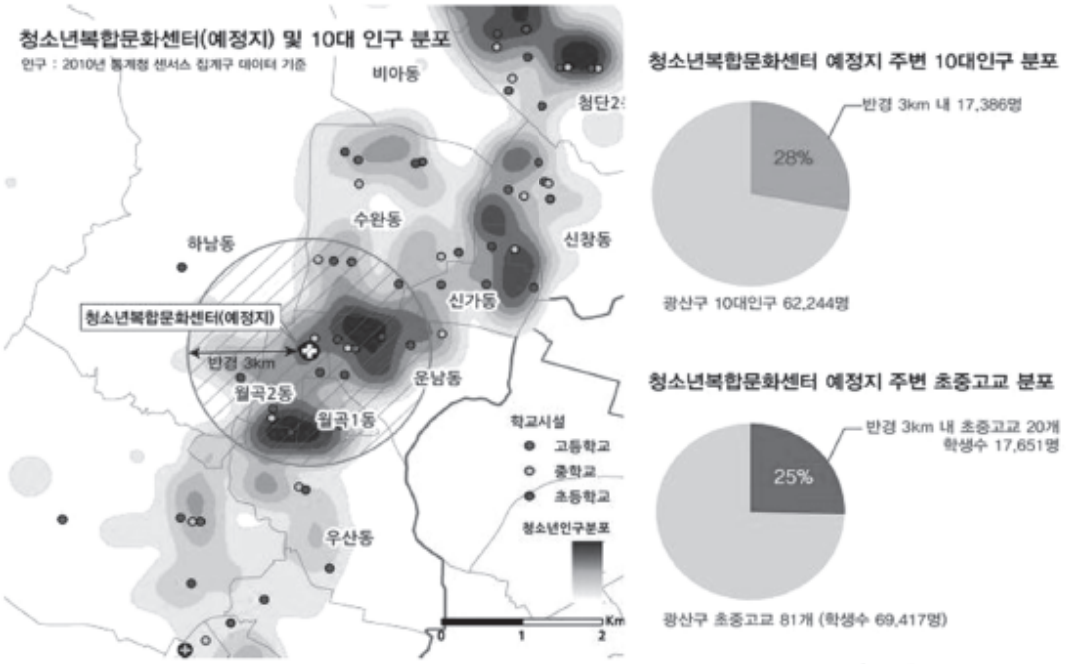


그림 7.12 GIS 정보 활용 - 청소년 인구 분포

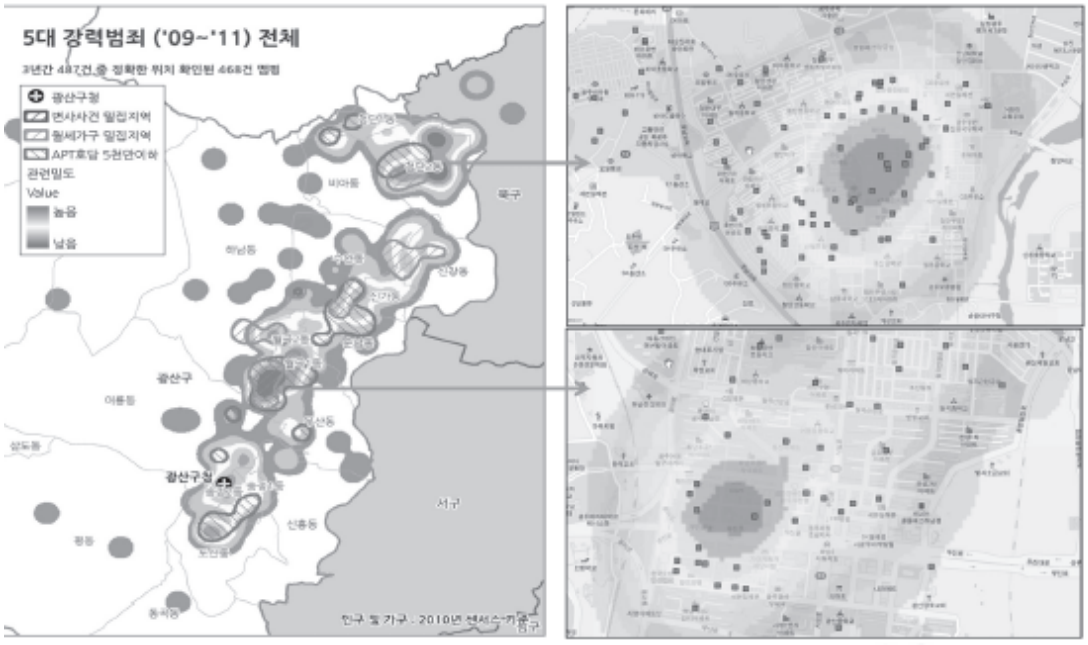


그림 7.13 GIS 정보 활용 - 범죄지도

이러한 정보를 이용하여 새로운 청소년복합문화센터를 건립하는데 건립 장소의 타당성을 평가하고 활용도를 증대 시키는데 도움이 될 수 있을 것이다. 또한 범죄지도를 통하여 적절한 대응방안을 마련하여 범죄를 예방하고 차단하는데 큰 도움이 될 수 있을 것이다.

7.4 공공기관의 활용현황

공공기관에서의 행정자료 활용은 그동안 대부분 기관의 업무실적을 집계하는 선에 머물러 왔다. 공사와 공단 등 공기업들의 업무가 소관 행정기관에서 법령 등에 근거에 위임·위탁된 업무가 많아, 이들 업무를 수행하는 과정에서 생성된 행정원천(광의의)을 업무실적보고의 기초자료로 활용하는 것이 대부분이었다.

공공기관의 통계작성 현황을 보면 94개 기관에서 작성하는 171종의 통계 중에서 조사통계가 102종으로 60%를 차지하고 있으며, 보고통계는 50종으로 29%를 차지하고 있다. 대부분의 보고통계가 행정원천을 이용해 연간 실적을 집계하는 방식으로 작성되고 있다. 다음 표 7.14는 2016년 7월 23일 현재 공공기관의 통계작성 현황을 정리한 것이다.

표 7.14 공공기관의 통계작성 현황 (2016.7.23. 현재)

구분	작성기관수	작성통계수	작성방법별		
			조사	보고	가공
지정기관	94	171	102	50	19
- 금융기관	8	23	9	6	8
- 공사/공단	27	50	21	27	2
- 연구기관	20	37	31	4	2
- 협회/조합	22	32	27	3	2
- 기타기관	17	29	14	10	5

7.5 기타 민간 및 연구기관의 활용현황

통계청을 중심으로 정부부문에서의 행정자료의 통계적 활용이 확산되고, 공공데이터의 개방이 진척됨에 따라 민간 및 연구기관의 행정자료 활용도 증가하게 될 것이다. 정부 3.0에 따른 정보공개의 확대로 많은 행정정보가 공유되어 활용되고 있다.

최근의 추세를 보면 2013년부터 2015년 사이에 개방건수는 3배, 활용건수는 52배로 급증하고 있음을 알 수 있다. 정부와 공공기관에서 제공하는 행정자료의 내용이 보다 상세해지고 그 종류가 다양해질수록 자료의 활용은 폭발적으로 증가하게 될 것이다. 결국 이는 사회의 다양한 부문에서 통계를 활용한 합리적인 의사결정을 가능하게 하여 국가경쟁력의 제고에도 도움을 줄 수 있을 것이다. 다음 표는 우리나라 공공데이터 개방 및 활용현황을 정리한 표이다.

표 7.15 〈공공데이터 개방 및 활용현황〉 단위 : 건(누적)

	2013년		2014년		2015년	
	개방건수	활용건수	개방건수	활용건수	개방건수	활용건수
소계	5,272	13,923	13,157	153,320	15,912	783,773
오픈API	554	9,815	1,361	33,687	1,790	73,846
파일데이터	4,718	4,108	11,796	119,633	14,122	709,927

CHAPTER 8

행정자료 통계활용에 대한 미래전략

학습목표

- 우리나라 행정자료 통계활용 시 한계와 문제점을 행정자료에 대한 인식, 공유시스템 관점에서 학습한다.
- 행정자료 통계활용을 위한 제도적 구조적 개선방안과 활용확산을 위한 방안에 대하여 학습한다.

8.1 우리나라 행정자료 통계활용의 한계와 문제점

행정자료를 통계적으로 활용하는 것은 이미 세계적인 추세이다. 특히 유럽의 경우 잘 발달된 레지스터 시스템과 정부 내 부처 간 데이터 공유체계, 유럽연합 내 국가 간 통일된 자료 공유체계 등 제도적 기술적 기반을 배경으로 국가의 공적통계의 대부분을 행정자료를 활용하여 작성하고 있다. 이런 측면에서 우리나라의 현실을 보면 이제 막 시작단계 정도라고 볼 수 있다. 통계청에서는 국가가 보유한 막대한 행정자료를 통계적으로 활용하기 위해 다양한 노력을 경주해오고 있지만, 근본적이고 구조적인 시스템의 변화 없이는 그 성과는 제한적일 수밖에 없을 것이다. 그 문제점을 살펴보면 다음과 같다.

1) 행정자료에 대한 인식

정부 각 부처나 지방자치단체, 공공기관 등이 보유한 행정원천을 국가의 자

원 내지 공공의 이익을 위해 사용해야할 공공재로 인식하지 않고, 보유기관의 전유물로 인식하는 경향이 일반화되어 있다. 모든 기관들이 국익을 위해 존재하고 있고 그들이 보유하는 자원 또한 국익을 위해 쓰여야 하는 것은 당연한 이치이다. 따라서 각 기관이 보유한 각종 행정원천이 국익을 위해 필요한 곳에 원활하게 쓰일 수 있어야 한다.

특히 통계는 국가와 국민의 공공재로써 합리적인 의사결정을 위한 필수적인 요소이다. 이러한 통계가 다양한 행정원천을 이용해 작성되고 국가의 모든 기관들이 공동으로 활용한다면, 동일한 통계를 근거로 정책이 수립됨에 따라 기관 간 정책의 일관성이 유지되고, 정책의 효과에 대한 합리적인 측정이나 평가도 가능해 질 수 있을 것이다.

2) 행정자료 공유시스템 미흡

각 기관이 보유한 행정원천으로부터 필요한 자료를 원활히 추출하여 통계작성에 이용할 수 있으려면, 이를 위한 체계적인 공유시스템이 구축되어 있어야 한다. 현재 각 기관이 소관 업무를 수행하기 위해 운영하고 있는 행정원천으로써의 업무관리 또는 정보관리 시스템은 그 물리적 구조나 운영방식, 관리되는 정보의 형태 등에서 매우 다양하다. 이렇게 수많은 시스템에 들어있는 다양한 정보를 통계활용을 위한 시스템으로 통합하고 관리한다는 것은 매우 어려운 일임에 분명하다. 이를 위해서는 행정원천의 공유를 위한 법적 제도적 기반과 이를 실현하기 위한 기술과 인력, 장비와 물리적 공간 등 막대한 자원이 필요할 수 있다. 하나의 통합된 시스템이든, 각각의 시스템을 클라우딩 방식으로 연결한 것이든 어떤 형식이든 간에 공동 활용체계를 구축하지 않는다면, 국가 행정원천의 효율적인 활용은 어려울 수밖에 없을 것이다. 행정자료의 활용도가 낮아 기존의 통계조사를 계속해야 하고, 이용 가능한 정보가 사장된다면 국가적으로 크나큰 손실이 아닐 수 없다.

행정원천 공유를 위한 법적 기반은 ‘통계법’, ‘개인정보 보호법’, 통계청의 ‘내부 운영 규정’ 등에서 찾아볼 수 있다. 이와 유관한 내용을 2.2.1절에서 다룬 바 있으므로 여기서는 관련 법령과 규정을 간략한 내용과 함께 정리하였다(출처: 2016 제1기 빅데이터와 행정자료). 참고하기 바란다.

관련 법령 및 규정

□ 통계법

○ 제4조의③(국가 등의 책무), 제24조(행정자료의 제공), 제24조의2(사법기관 등의 자료 제공), 제33조(비밀의 보호) 및 제34조(통계종사자 등의 의무)에 의하여 개인이나 단체 등의 비밀에 속하는 사항에 대한 비밀보호 의무를 지고, 위반 시 제39조(벌칙), 제41조(과태료) 등에 의하여 처벌을 받도록 규정되어 있음

제4조의③(국가 등의 책무)통계작성기관의 장은 통계의 작성을 위하여 질문을 받거나 자료제출 등의 요청을 받고 답변을 하거나 자료제출 등을 하는 개인이나 법인 또는 단체 등(이하 "통계응답자"라 한다)의 부담을 최소화하고, 비밀이 보호되도록 노력하여야 한다.

제24조(행정자료의 제공) ① 중앙행정기관의 장 또는 지방자치단체의 장은 통계의 작성을 위하여 필요한 경우에는 공공기관의 장에게 행정자료의 제공을 요청할 수 있다. 〈개정 2012.12.18〉

제24조의2(사법기관 등의 자료 제공) ① 통계청장은 통계의 작성을 위하여 필요한 경우에는 가족관계 등록전산자료의 제공을 법원행정처장에게 요청할 수 있다. [본조신설 2014.5.14.]

제33조(비밀의 보호) ①통계의 작성과정에서 알려진 사항으로서 개인이나 법인 또는 단체 등의 비밀에 속하는 사항은 보호되어야 한다. ②통계의 작성을 위하여 수집된 개인이나 법인 또는 단체 등의 비밀에 속하는 자료는 통계작성 외의 목적으로 사용되어서는 아니 된다.

제34조(통계종사자 등의 의무) 통계종사자, 통계종사자이었던 자 또는 통계작성기관으로부터 통계 작성업무의 전부 또는 일부를 위탁받아 그 업무에 종사하거나 종사하였던 자는 직무상 알게 된 사항을 업무 외의 목적으로 사용하거나 다른 자에게 제공하여서는 아니 된다.

제39조(벌칙) 다음 각 호의 어느 하나에 해당하는 자는 3년 이하의 징역 또는 1천만원 이하의 벌금에 처한다.

1. 통계의 작성을 목적으로 수집되거나 제공을 받은 개인이나 법인 또는 단체 등에 속하는 사항을 그 목적 외의 용도로 사용하거나 이를 다른 자에게 제공한 자

제41조(과태료) ③ 다음 각 호의 어느 하나에 해당하는 자에게는100만원 이하의 과태료를 부과한다.

1. 제24조의 제4항을 위반하여 공공기관으로부터 제공받은 행정자료(비밀에 속하는 사항을 제외한다)를 제공받은 목적 외의 목적으로 사용하거나 다른 자에게 제공한 자

□ 개인정보 보호법

○ 제1장(총칙)~제9장(벌칙)에 개인정보 보호 원칙, 국가 등의 책무, 개인정보 보호지침, 벌칙 등이 규정되어 있음

- 동법 제58조(적용의 일부 제외)에 공공기관이 처리하는 개인정보 중 「통계법」에 따라 수집되는 개인정보의 경우 제3장부터 제7장의 적용(*)에서 제외됨

* 제3장 개인정보의 처리 , 제4장 개인정보의 안전한 관리 , 제5장 정보주체의 권리 보장 , 제6장 개인정보 분쟁조정위원회 , 제7장 개인정보 단체소송

□ 내부 운영 규정

○ 「행정자료의 정보보호에 관한 규정」(예규 제182호, '16.2.29.일부개정)

- 통계법 제24조 및 제24조의2에 따라 접수한 행정자료의 안전한 관리를 위한 정보 보호 강화 및 접수된 행정자료의 활용을 위한 이용범위, 이용절차에 필요한 사항을 정하고 있음

- 주요 내용
  - ◦ 행정자료 DB를 인터넷망과 물리적으로 분리된 보안망에 구축하도록 명시(제6조제3항)
  - ◦ 주민등록번호,여권번호,운전면허번호,외국인등록번호등고유식별정보의 정보보호를 위하여 암호화 규정 마련(제6조제4항)
  - ◦ 접수한 행정자료의 이용목적을 달성한 경우에는 해당 행정자료를 복구할 수 없는 방법으로 파기하도록 함(제7조의2)
  - ◦ 청내 행정자료 활용 범위 규정 신설(제7조의3)
    - ◦◦ 행정자료 DB 및 개별파일 형태로 보관중인 개체식별 자료의 이용 시에는「통계법 시행령」제38조 제1항 제3호에 따른 사용목적대로 이용하되, 다른 목적으로 이용 시에는 자료제공기관의 승인을 득하여 사용
    - ◦◦ 행정자료에 통계적 처리를 거친 통계자료는 청내 공동 활용하되, 통계자료제공심의회의 심의를 거쳐 이용

○ 「정보보안업무 규정 시행세칙」(훈령 제251호 2012. 3.30.)

- 국가 정보보안 기본지침, 국가 사이버안전 매뉴얼, 개인정보보호법에 의하여 정보보안활동에 필요한 세부사항을 규정

◦ 통계청 전체의 개인정보보호 업무 총괄책임자(CPO)는 통계데이터허브국장이 담당하고 정보시스템 접근 시 반드시 사용자 인증과정을 거쳐야 접근이 가능

- **제1조(목적)** 국가 정보보안 기본지침, 국가 사이버안전 매뉴얼, 개인정보보호법에 따라서 통계청 정보보안활동에 필요한 세부사항 규정
- **제14조(개인정보보호관리체계)** 통계청 전체의 개인정보보호업무총괄책임자(CPO)는 통계데이터허브국장이 담당, 영상정보관리자는 운영지원과장이 담당
- **제21조(접근 통제)** ① 통계청의 모든 정보시스템은 반드시 사용자 인증과정을 거쳐야 접근이 가능하도록 함. ② 정보보안담당관은 각 정보시스템의 사용자가 관리자 권한으로 접근하고자 할 경우 IP 주소를 이용하여 특정 단말에서만 접근이 허용하도록 조치하여야 한다. ③ 시스템 관리자는 동일한 계정으로 여러 명이 동시에 정보시스템에 로그인하지 못하도록 차단하여야 하며, 업무상 필요한 경우에 정보보안담당관의 승인을 받아 제한적으로 그룹 계정을 사용할 수 있다.

※ “정보보안” 또는 “정보보호” 란 정보통신수단으로 수집・가공・저장 ・검색・송수신되는 정보의 유출・위변조・훼손 등을 방지하거나 정보통신망을 보호하기 위하여 관리적・물리적・기술적 수단을 마련하는 일체의 행위를 말한다.

○ 「통계청 개인정보보호지침」(통계청예규 제160호 201511.09. 일부발췌)

- 제1조(목적) 이 개인정보 보호지침 (이하 "지침”이라 한다)은 「개인정보보호법」(이하 "법”이라 한다) 에 따라 개인정보보호에 필요한 개인정보의 처리에 관한 기준, 개인정보 침해의 유형 및 예방조치 등에 관한 세부적인 사항을 규정함을 목적으로 한다.
- 제3조(적용범위) 이 지침은 전자적 처리 여부를 불문하고 수기문서를 포함한 모든 형태의 개인정보파일을 운용하는 통계청과 통계교육원, 통계개발원, 지방통계청(이하 '소속기관’이라 한다)의 직원 및 계약관계에 있는 외주 직원에거i 적용된다.
- 제5조(다른 지침과의 관계) 통계청이 보유한 개인정보 중 통계법에 따라 수집된 행정자료 및 통계조사자료에 관해서는 본 지침이 적용되지 않으며, 「행정자료의 정보보호에 관한 규정」 및 「정보보안업무규정시행세칙」이 적용된다.

8.2 미래발전을 위한 제안

행정자료의 공동 활용을 활성화하기 위한 방안은 여러 가지가 있을 수 있다. 이를 크게 나누어 볼 때, 기존의 문제점을 일시에 근본적으로 개선함으로써 획기적인 효과를 기대할 수 있는 제도적 구조적 개선방안과, 그러한 근본적 개선 없이 현상을 유지하면서 제한된 성과를 기대하는 미시적 개선방안을 살펴보고자 한다.

8.2.1 제도적 구조적 개선방안

우리가 행정자료 이용활성화를 위한 제도적 구조적 개선방안을 논의함에 있어 전제되어야 할 것이 있다. 그것은 행정자료 공유에 대한 국민 또는 각 기관의 인식과 행정자료 공유체계를 구축하는데 소요되는 예산과 인력의 조달 가능성 등 개선에 필요한 환경적 제약조건을 배제하고, 오직 행정자료의 이용을 가장 활성화 할 수 있는 방안을 모색한다는 가정이다. 만일 이러한 전제 없이 행정자료 이용활성화를 제약하는 현실적 여건 – 행정원천 보유기관의 호응도, 관련법령 등 제도적 뒷받침, 관련 시설과 장비, 기술과 인력, 운영예산의 확보 등 –을 고려한다면, 가장 합리적인 방안을 발견하기 어려울 것이다. 따라서 마치 백지에 그림을 그리듯, 행정자료의 통계적 이용을 가장 활성화 할 수 있는 최적의 대안을 아무런 제약 없이 구상해 봄으로서 궁극적으로 우리가 추구할 목표가 무엇인가를 발견할 수 있을 것이다.

이런 전제하에서 행정자료 공동 활용의 효율성에 최상의 가치를 두었을 때, 우리가 이상적으로 생각해 볼 수 있는 방안이 물리적으로 단일한 데이터베이스시스템이다. 국가 모든 기관의 행정원천이 하나로 통합된다면 그 활용성은 더 이상의 설명이 필요 없을 정도로 엄청나게 증가될 것이다. 각 기관이 국민으로부터 수집한 모든 정보는 국가의 통일된 데이터베이스에 구축되고 공동 활용된다. 각 기관은 동일한 정보를 기초로 정책을 수립하고 집행하게 됨으로써 정책간의 조화와 일관성이 유지될 수 있을 것이다. 국민이 특정 행정행위의 객체로서 국가기관에 제공한 정보는 국가데이터베이스에서 안전하게 관리 활용하게 된다. 따라서 국민은 동일한 정보를 중복하여 국가에 제공할 필요가 없어

져 국민의 부담이 경감됨은 물론, 행정의 효율성 또한 향상될 것이다. 논리적으로도 하나의 국가 하나 된 정부의 이념과 가장 어울리는 제도라고 하지 않을 수 없다.

8.2.2 현실적 활용확산방안

앞에서 언급한 구조적 개선방안은 궁극적으로 추구해야 할 만한 대안이기는 하지만 그것을 현실화하기 위해서는 행정구조의 일대 혁신에 가까운 대변화를 수반하게 될 것이다. 궁극적으로 모든 정부기관의 행정원천이 하나의 데이터베이스로 통합되는 것이기 때문이다. 그러나 이러한 방향으로 진전하는데 필요한 많은 기반이 구축되어가고 있다. 전자정부시스템이 그것이다. 현재 우리나라는 거의 대부분의 정부기관이 통합된 업무처리시스템인 전자정부시스템에 의해 연결되어 있다. 결재시스템을 통한 의사결정은 물론, 각종 지식정보를 공유하고 있다. 이 시스템이 기본적으로는 단위 기관 단위별 시스템을 연계하고 필요한 항목만을 공유하게 되어 있지만, 필요에 따라서는 언제든지 전체를 하나의 시스템으로 변경시킬 수 있는 기반이 구축된 것이다.

문제는 이러한 범용적인 업무처리시스템이 아니라 조세, 건축물, 환경 등 각 기관이 독립적으로 운영하고 있는 행정원천 시스템이다. 이러한 각 기관의 행정원천을 통계적으로 활용하려면, 통계작성기관이 해당 행정원천에 접속하거나 해당 원천의 보유기관이 그 사본을 통계작성기관에 제공하는 방법이 있을 수 있다. 전자의 대표적인 사례가 노르웨이의 행정자료 활용체계이다.[49] 대표적인 통계작성기관인 통계청이 각 부처가 운영하고 있는 행정원천에 실시간으로 접속하여 통계를 작성할 수 있도록 하고 있다.

우리나라는 후자의 경우이다. 통계청이 원하는 자료의 사본을 통계청에 제공하고 통계청을 이를 가공하여 통계를 작성하는 방식이다. 통계청은 제공받은 행정자료를 데이터베이스로 구축하고 이를 바탕으로 종합통계등록부 형태로 발전시켜 각종 통계작성에 활용하는 사업을 진행 중에 있다. 이러한 방법은 현재 단계에서 현실적으로 가능한 최적의 대안이라고 생각되며 이를 발전시켜 향후 통합시스템으로 가는 발판으로 삼을 수 있을 것이다.

49)노르웨이의 행정자료 활용을 이해할 수 있는 관련 법령 등 자료는 부록8~9 참조

행정자료의 통계적 활용 가능성과 효과를 좌우하는 것이 바로 항목의 정의, 분류 기준, 포괄 범위 등 작성기준의 통일성이다. 행정자료와 통계자료가 동일한 기준을 가지고 작성될 때, 양 자료 간 공동 활용도가 극대화될 수 있을 것이다. 이러한 측면에서 통계청이 행정자료를 통계적으로 활용하면서 각종 작성기준을 통일하는데 주도적인 역할을 하는 것이 중요하다. 어차피 각 행정기관에서 행정원천을 작성하는 것이 통계적 활용을 목적으로 하지 않기 때문에, 통계청이 이러한 상황과 통계적 요구를 여하히 조화시켜 작성기준을 일치시켜 나가는 가가 중요한 과제가 될 것이다.

CHAPTER 9

행정자료를 이용한 통계개발 사례[50)]

학습목표

- 행정자료를 이용한 통계개발 환경에 대하여 학습한다.
- 행정자료를 이용한 통계개발 해외사례에 대하여 학습한다.
- 우리나라의 사례로 임금근로일자리 행정통계 개발과정 및 개선과제에 대하여 학습한다.

9.1 개발 배경 및 환경

9.1.1 개발 배경

우리나라는 저출산·고령화의 급속한 진전으로 노동인력은 계속 줄어들고 있고, 최근 민간소비와 기업의 투자 위축으로 저성장 기조가 고착화되면서 경제상황도 어려움에 직면하고 있다. 청년층의 취업은 갈수록 어려워지고 있고, 베이비부머(1955~1963년생) 세대의 은퇴, 노년층의 재취업이 늘어나면서 일자리에 대한 사회적인 관심은 커져가고 있다. 정부에서는 경기회복과 일자리 창출을 위한 각종 경제정책을 추진하고 있는데, 이를 뒷받침하는 일자리 관련 통계는 충분치 못한 실정이다.

통계청의 경제활동인구조사는 가구를 대상으로 조사하고 있어, 산업·직업, 종사

50) 본 장은 통계교육원 행정자료 활용 교육과정(2015) 교재에 수록된 이정수의 '행정통계 사례분석'을 전재한 것임

자 규모 등 기업체와 연계한 정보를 파악하는 데에는 한계가 있다. 전국사업체조사와 기업활동조사는 사업체(기업체) 중심으로 작성하기 때문에 종사자의 특성(학력, 연령, 근속기간 등)에 대한 자료가 취약하다. 이와 같이 현재 작성하고 있는 조사통계는 종사자와 기업체의 특성을 동시에 파악하는데 어려움이 있다.

언론에서도 기업체의 '일자리'를 파악할 수 있는 통계가 부족하다는 보도가 있었다. 그렇다고 해서 기업체와 종사자의 특성을 동시에 파악하는 일자리 통계를 현장조사를 통해 새로 작성하기에는 제약이 있다. 갈수록 조사환경은 어려워지고 있고 조사비용은 늘어가고 있기 때문이다.

어려운 통계작성 환경에 대처하기 위한 방안으로 행정자료 활용의 필요성이 부각되고 있다. 국회나 감사원 등 외부기관에도 행정자료를 활용한 통계작성을 강조하고 있으며, 사회적인 공감대가 형성되어 가고 있다.

이상 살펴본 바와 같이 통계조사환경 악화와 일자리통계 수요에 적극 대처하기 위해 공공기관이 보유하고 있는 행정자료를 활용하여 임금근로일자리 통계를 개발하게 되었다.

9.1.2 개발 환경

행정자료를 통계작성에 활용하기 위해서는 몇 가지 전제조건이 필요하다. 우선, 전국단위의 품질이 좋은 모집단을 확보할 수 있어야 하고, 행정자료를 서로 연계할 수 있는 고유키가 존재해야 한다. 그리고 행정자료를 이용할 수 있는 법적근거가 마련되어 있어야 하며, 통계작성 목적으로 전환이 가능해야 한다. 아래 내용은 임금근로일자리 통계를 작성하는데 필요한 전제조건을 충족하고 있는지에 대해 검토한 내용이다.

1) 품질이 우수한 행정자료 존재

임금근로일자리 통계를 작성하기 위해 활용 가능한 행정자료를 살펴보면, 건강보험, 국민연금, 고용보험, 산재보험 등 4대 사회보험 자료와 국세청이 보유하고 있는 과세자료(법인세, 부가가치세, 근로소득지급명세서)가 있다. 사회보험은 최근에 가입대상을 근로자 1인 이상 사업장으로 확대하면서 포괄 범위가 넓어졌고, 가입률도 꾸준히 증가하고 있기 때문에 행정자료가 모집단을 포

괄할 수 있는 범위는 확대되어 가고 있는 상황이다. 국세청 과세자료의 경우에도 신용카드와 현금영수증 발행 증가로 품질이 향상되어 가고 있다.

2) 행정자료 간 연계키 존재

고용보험, 국민연금, 건강보험 자료에는 주민등록번호가 공통적으로 등록되어 있고, 산재보험, 과세자료 등에는 사업자등록번호가 존재하므로 가입자(종사자)와 사업체의 연계가 가능하다. 주민등록번호를 통계목적의 연계키로 활용할 때에는 개인식별번호(PIN)로 암호화하여 활용할 수 있다.

3) 행정자료를 활용한 법적 근거

행정자료를 통계목적에 활용하기 위해서는 법적근거가 반드시 마련되어 있어야 한다. 통계법에 행정자료 활용 조항(제24조)을 신설('07.4)하였고, 2009년 2월에는 국세기본법에 과세자료 요구 규정을 개정하였으며, 2014년 2월에는 지방세기본법을 개정하였다. 관련법의 개정으로 통계청에서는 행정자료를 통계작성 목적에 활용할 수 있게 되었다.

4) 통계작성 목적으로 전환

행정자료에 연계키가 있다고 하여 무조건 통계를 작성할 수 있는 것은 아니다. 활용하는 행정자료 간에 중복자료를 걸러 낼 수 있어야 하고, 통계작성 기준에 맞도록 전환이 가능해야 한다. 임금근로일자리 통계작성에 필요한 행정자료에는 개인식별번호가 공통적으로 존재하기 때문에 통계기준 시점에 맞추어 중복자료의 처리가 가능하고 가입자와 사업주의 동일인을 식별할 수 있으므로 미대상자(비임금근로자 등)를 제외할 수 있다. 또한, 법인등록번호를 이용하여 사업자 단위로 등록된 행정자료를 기업체 단위로의 전환이 가능하다.

9.2 선행사례 고찰

9.2.1 해외 사례

1) 행정자료 기반의 일자리 통계 작성 사례

뉴질랜드, 덴마크, 핀란드, 스웨덴은 행정자료를 활용하여 다양한 일자리 통계를 작성하고 있다. 이들 나라는 행정자료에 기반을 둔 기본등록부를 구축하여 각종 통계지표를 생산하고 있다. 다음 표 9.1은 외국의 일자리 통계 작성 현황을 정리한 것이다.

표 9.1 외국의 일자리통계 작성 현황

국 가	통 계 명	주요지표	행정자료-조사자료	
			행정자료	조사자료
뉴질랜드	Linked Employer - Employee Data (LEED)	임금근로일자리, 비임금근로일자리	국세자료	사업체모집단(NZ's Business Frame)
덴마크	Register-based labour force statistics	일자리, 통근자, 소지역	인구등록부, 사업체 과세자료	
핀란드	Register-based employment statistics	일자리, 산업세세 일자리	중앙인구등록부, 과세등록부	
스웨덴	The Integration of Immigrants into the Labour Market	일자리, 통근근로자, 산업분류, 소지역	직업등록부, 취업등록부	

2) 스웨덴의 기본통계등록부

스웨덴 통계청은 행정자료 기반으로 4가지의 기본통계등록부를 구축하여 각종 통계작성에 활용하고 있다. 기본통계등록부는 인구, 토지, 직업, 사업체로 구분된다. 직업통계등록부(Job register)는 이용자의 새로운 통계요구에 부응하고 등록통계의 품질을 높이기 위해 “2001년 스웨덴의 등록부 미래 발전 계획”에 활동기본통계등록부[51]로 개념을 확장하였다. 활동기본등록부는 경제활동을 중심으로 학교활동, 노동의 수요와 공급 연계, 거주지와 근무지 연계, 기타 자료와 연계한 통계생산 기반 마련하기 위해 경제활동 이외의 학업활동과 기타

51) 활동기본통계등록부는 Job register(직업등록부)를 경제활동에서 학업활동, 기타활동을 추가한 개념이다.

활동(가사 등)까지 포함하여 구축되어 있다. 다음 그림 9.1은 스웨덴의 기본통계등록부 모형을 정리한 것이다.

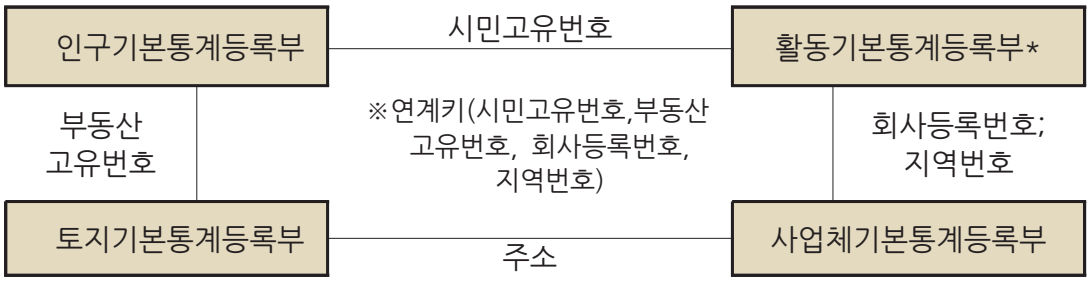


그림 9.1 스웨덴의 기본통계등록부 모형

스웨덴은 활동기본통계등록부를 이용하여 취업자를 직업분류(세분류 3-digit, 세세분류 4-digit)에 따라 성, 연령, 학력, 일자리, 거주지별로 통계를 작성하고 소지역 단위(County/Municipality)의 주야간 인구, 출퇴근/통학 등 이동인구, 산업별 취업자 통계를 작성하고 있다.

3) 뉴질랜드의 LEED[52)]

뉴질랜드 통계청은 국세청으로부터 15세 이상의 소득원천 징수 대상자의 과세자료를 전송 받아 사업체명부(BF)[53)]와 결합하여 LEED를 구축하고 노동·고용 통계를 생산하고 있다. 국세청의 과세자료와 사업체명부를 연결해 주는 주된 식별자는 IRD[54)]번호이며, 고용인의 성명, 주소 등을 보조 식별자로 활용하고 있다.

사업체명부는 동일 시점에서 사업체의 3각 관계성(기업-사업체-지리적 단위 수준의 사업체)을 표시하며, 사업체의 고용인과 피고용인 간의 연계성은 과세자료를 통해 파악이 가능하다. 요컨대, 사업체등록부와 과세자료의 통합으로 피고용인-고용인-사업체와 지리적 단위의 사업체-기업체 정보를 모두 연계하여 다양한 통계를 생산할 수 있다. 그림 9.2는 뉴질랜드의 LEED 자료 간 연계방법을 정리한 것이다.

52) LEED : Linked Employer-Employee Data(고용인과 피고용인의 정보가 결합된 행정데이터 집합)
53) 사업체명부(BF) : Business Frame(뉴질랜드 내의 제조·서비스업 사업체 목록)
54) IRD(Inland Revenue Department) : 한국의 세무서에 해당하는 세무관서이며, IRD번호는 과세목적으로 부여하는 개인식별번호

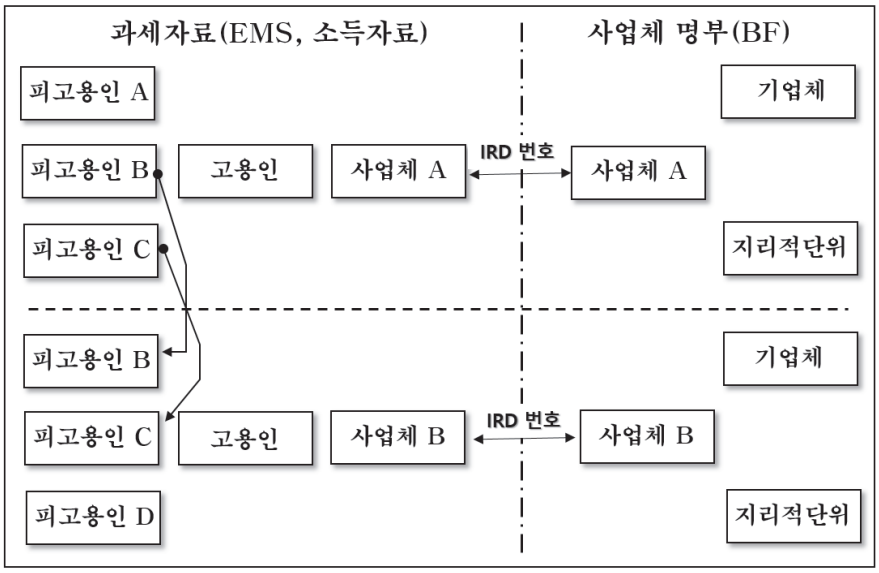


그림 9.2 뉴질랜드의 LEED 자료 간 연계방법

표 9.2 LEED를 활용하여 생산하는 주요통계

주요 통계	주요 작성 항목	활용 분야
일자리 현황	조사기준시점 기준, 일자리 현황	경기활동지표
신규 채용 수	전 조사대상분기 이후 신규로 채용된 피고용인 수	산업별·지역별·연령별· 성별 근로자 현황
퇴직자 현황	전 조사대상분기 이후 퇴직한 피고용인 수	
이직률	총 신규채용과 퇴직자 평균에 대한 조사대상분기와 직전 분기내 총 고용평균의 비율 *(신규채용자+퇴직자)/2:(조사대상분기 일자리+직전분기 일자리)	고용 안정성 측정, 연령 및 성별 이직률 현황 * 이직률은 기업단위가 아닌 지리적 단위 사업체별로 산출
일자리 창출력	직전 조사기준시점 이래로 사업확장 또는 창업으로 인해 창출된 일자리 수 *예: 100인 사업체가 10명의 신규채용과 5명의 퇴직자가 있었다면 이 사업체의 일자리 창출력은 '5'임	산업별·부문별·지역별 일자리 창출 및 감소현황
일자리 감소율	직전 조사기준시점 이래로 사업 축소 및 폐업으로 인해 감소한 일자리 수 *예: 100인 사업체가 5명을 신규채용하고 15명의 퇴직자가 발생한 경우, 이 사업체의 일자리 감소율은 '10'	사업체 규모별 또는 기타 요소별로 일자리 창출력/감소율 파악
평균 소득	조사대상분기 내 현존하는 모든 일자리 소득의 평균 및 중위값 * 단, 일부 파트타임 소득은 제외	최상 및 최저 평균소득·중앙값을 가지는 산업이나 지역 분석
중위 소득		

뉴질랜드 통계청은 LEED를 활용해 일자리 창출현황, 이직률, 평균소득 및 중위소득 등의 통계를 생산하여 산업 및 부문별 일자리와 고용자 흐름 파악 등에 활용하고 있다. 다음 표 9.2는 뉴질랜드에서 LEED를 활용하여 생산하는 주요통계를 정리한 것이다.

9.3 개발 가능성 검토

행정자료를 활용한 임금근로일자리 통계의 개발가능성을 검토하기 위해 경제활동인구조사, 사업체조사 등 유사통계와 비교분석을 실시하였다.

9.3.1 행정자료의 일자리 설명력 검토

1) 조사자료와의 상관성 분석

2003년~2010년까지 행정자료(건강보험 가입자)와 조사자료 간의 임금근로자 상관관계는 매우 높고, 사업체조사와의 임금근로자 규모는 유사한 수준으로 나타나 행정자료의 일자리 규모 변화에 대한 설명력은 충분하다고 판단되었다(표 9.3, 표 9.4 참조).

표 9.3 행정자료와 조사자료와의 임금근로자 상관계수(r)

	경제활동인구조사	전국사업체조사	사업체노동실태
행정자료 (건강보험가입자)	0.997794	0.966764	0.982869

* 상관계수(r) : $-1 \leq r \leq 1$ 사이의 값, −1 또는 1에 가까울수록 연관성이 높음

표 9.4 행정자료와 조사자료의 임금근로자 현황

연도	행정자료 (건강보험 가입자)	임금근로자		
		경제활동인구조사	전국사업체조사	사업체노동실태
2003	8,805	14,149	10,462	9,355
2004	9,283	14,584	10,594	9,486
2005	9,746	14,968	10,908	9,749
2006	10,415	15,351	11,188	10,021
2007	11,175	15,882	11,352	10,431
2008	11,617	16,104	11,980	10,746
2009	12,146	16,479	12,464	11,144
2010	12,764	17,048	13,229	11,849

* 전국사업체조사 및 사업체노동실태는 상용근로자와 일시·일용근로자를 합친 수치임

* 경제활동인구조사(부가조사)는 8월 기준으로 작성한 임금근로자 수치임

2) 경제활동인구조사와 비교

2010년 말 기준의 행정자료를 활용한 임금근로일자리는 14,065,000개로, 경제활동인구조사(2010.8월, 부가조사)의 17,048,000명보다 2,983,000개 적게 나타났다.

표 9.5 행정자료와 경제활동인구조사 임금근로자 (단위: 천명)

연도	행정통계	경제활동인구조사	차 이
2009년	13,486	16,479	-2,993
2010년	14,065	17,048	-2,983

두 자료 간의 차이가 발생하는 주요 원인으로는 ①조사기준에 따른 차이, ② 미가입·미등록 일자리 등에서 찾아 볼 수 있다.

경제활동인구조사는 약 30,000가구를 대상으로 조사대상 주간에 1시간 이상 수입을 목적으로 일한 사람을 임금근로자로 조사하여 작성하고 있으나, 행정통계는 행정등록부의 가입요건* 기준으로 신고한 근로자에 한해 작성할 수밖에 없으므로 두 자료 간의 차이가 발생한 것으로 예상된다.

표 9.6 경활 부가조사(2011.8) 임금근로자 및 사회보험 가입률

구 분	임금근로자(천명)	사회보험 가입률(%)		
		국민연금	건강보험	고용보험
합 계	17,510	65.1	68.3	64.6
○ 정규직	11,515	79.1	80.9	77.4
○ 비정규직	5,995	38.2	44.1	42.3
- 한시적	3,442	58.2	66.9	62.7
- 시간제	1,702	11.0	13.0	13.5
- 비전형	2,427	21.1	29.6	28.0

* (가입요건) 고용·산재보험: 월 60시간, 주15시간이상 근무한 근로자, 국민연금: 1개월 이상, 월60시간 이상 근무한 근로자

또한, 임금근로자 중에서 비정규직 등 취약계층(시간제, 비전형근로*)의 사회보험 가입률이 낮아 조사 자료와의 차이가 발생하는 주요 원인으로 추정된다(표 9.6 참조).

* 비전형근로자(2,427천명) : 파견(197천명), 용역(673천명), 특수형태근로(614천명), 일일근로(962천), 가정 내 근로(75천명)[55]

3) 전국사업체조사와 비교

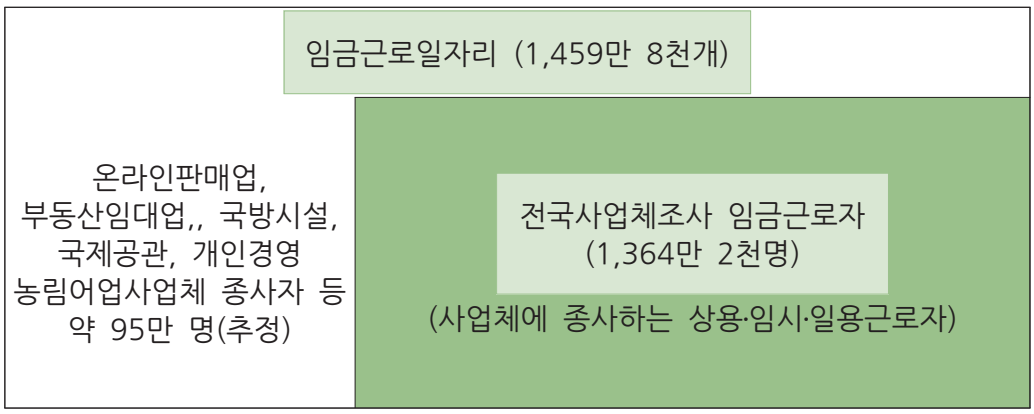
2011년 말 기준의 행정통계의 임금근로일자리는 총 14,598천개로, 동년 전국사업체조사 임금근로자 13,642천명(상용 11,345천명, 임시·일용 2,297천명)보다 956천개 많게 나타났다. 두 자료간의 차이가 발생하는 주요 원인으로는 ① 사업체 정의·포괄범위, ②조사대상 차이 등으로 볼 수 있다(표 9.8 참조).

온라인판매업, 부동산임대업 등 사업장이 없는 사업체에 종사하는 근로자는 전국사업체조사에서 제외되어 차이가 발생한 것으로 보인다. 그리고, 국방시설, 국제공관, 개인경영 농림어업사업체는 전국사업체조사에서 제외되기 때문에 종사자 차이가 발생한 것으로 추정된다.

* 사업체(establishment) : 일정한 물리적 장소 또는 지역 내에서 하나의 단일 또는 주된 경제활동에 독립적으로 종사하는 기업체 또는 기업체를 구성하는 기본 단위

※ 전국사업체 조사대상에서 제외되는 사업체
① 국방시설(군부대, 군사학교, 군병원 등)
② 국제기구·공관(한국주재 대사관, 영사관 등)
③ 가정내 고용활동(가정부, 가정교사 등)
④ 개인이 경영하는 농림어업 사업체(농장, 목장, 양어장, 양식장 등)
⑤ 고정시설이 없는 노점, 행상, 상근종사자 없는 어촌계, 이동차량 판매업자 등

표 9.8 전국사업체조사 임금근로자 ↔ 임금근로일자리 행정통계 비교



55) 자료출처 : 2011년 8월 경제활동인구조사 근로형태별 및 비임금 근로 부가조사 결과(2011.10.28)

9.4 임금근로일자리 행정통계 개발

9.4.1 일자리 개념

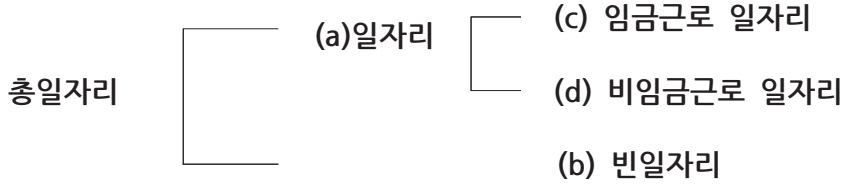
일자리란 기업체에서 통계작성 기준 일에 임금근로·비임금근로 활동을 하는 근로자(인적단위)가 '점유한 고용위치(filled employment position)'를 말한다.

※ 일자리 정의 : 사업체(또는 기업)에서 근로자에 의해 채워진 고용지위
"A job is an employment position filled by a worker."
(「Job Creation and Destruction」, David, Haltiwanger and Schuh, 1996)

※ 일자리 변동의 예시 : 1년 동안 일자리 수 10개가 감소
• 2010년 100명의 근로자 일자리 → 2011년 90명의 근로자 일자리

'고용위치'라는 측면에서 일자리는 취업자와는 개념에 차이가 있다. 예를 들면, 어떤 개인이 A회사에 다니면서 B학원의 시간강사를 하고 야간에는 C대리운전회사에서 운전기사로 일하는 경우, 취업자는 1명이나 일자리는 3개(A회사 직원, B학원 강사, C회사 직원)이다. 1일 3교대 근무하는 경우에는 일자리와 취업자는 3개로 동일하다. 이와 같이 취업자는 사람 중심이고, 일자리는 사람이 점유하고 있는 고용위치를 의미하기 때문에 개념 차이가 있다.

※ **취업자 (경제활동인구조사)**
• 조사대상 주간 중 수입을 목적으로 1시간 이상 일한 자
• 자기가구에서 경영하는 농장이나 사업체의 수입을 높이는 데 도운 가족종사자로서 주당 18시간이상 일한 자(무급가족종사자)
• 일시적인 병, 일기불순, 노동쟁의 등의 이유로 일하지 못한 일시휴직자



* 임금근로 일자리 행정통계는 '(c) 임금근로 일자리'를 중심으로 작성

일자리는 작성기준시점에 ‘사람이 점유하고 있는 일자리’ 와 ‘점유하고 있지 않은 빈일자리‘로 구분된다. 점유하고 있는 일자리는 현물 또는 현금을 댓가로 수입이 있는 경우에는 임금근로일자리로, 그렇지 않은 경우에는 비임금근로일자리로 구분한다.

9.4.2 기업체와 사업체 개념

1) 기업체

기업체란 재화 및 서비스를 생산하는 법적 또는 제도적인 최소 경영단위로 자원배분에 관한 의사결정의 자율성이 있고, 수입·지출 및 자금관리에 관한 기록을 독립적으로 유지·관리하는 단위를 말한다. 기업체는 조직형태에 따라 회사법인, 회사이외법인, 국가·지방자치단체, 비법인단체, 개인사업체로 구분한다.

표 9.9 조직형태에 따른 기업체 분류

회사법인	회사이외 법인	국가·지방자치단체	비법인 단체	개인 사업체
상법에 의하여 영리추구를 목적으로 설립하여 법인격을 갖춘 조직체(주식회사, 합명회사, 합자회사, 유한회사, 외국회사)	민법에 의한 재단·사단법인과 특별법에 의한 특별법인(재단법인, 사단법인, 정부투자기관 및 재투자기관 등)	중앙행정기관, 지방자치단체, 국회, 법원, 국·공립학교	법인격이 없는 각종단체나 모임(문화단체, 경기후원단체, 후원회 등)	법인격을 갖추지 않고, 1인 또는 공동으로 경영하는 기업

2) 사업체

영리·비영리를 불문하고 일정한 물리적 장소에서 재화의 생산·판매, 서비스 제공 등 경제활동을 독립적으로 수행하고 있는 모든 경영단위를 말한다. 어떤 회사가 사업장을 달리하는 본사, 지사, 지점, 공장, 영업소 등이 있을 경우 기업체는 1개이고, 사업체는 여러 개로 구성된다.

9.4.3 통계지표

1) 통계지표 체계

임금근로일자리 통계지표는 해외 사례, 국내선행 연구사례에 기초하여 노동 전문가 및 관계부처 의견을 반영하여 아래와 같이 지표체계를 구성하였다. 다음 그림 9.3은 임금근로 일자리에 대한 통계지표 체계도이다.



그림 9.3 임금근로일자리 통계지표 체계

2) 통계지표 개념

- 임금근로총일자리: 지속일자리와 변동일자리의 합으로 구성됨

- 지속일자리: 동일한 기업체에서 두 기간(당해년도와 전년도)에 걸쳐 지속적으로 동일근로자에 의해 일자리가 점유된 경우(예 : 전보)

- 대체일자리: 동일한 기업체에서 두 기간(당해년도와 전년도)에 걸쳐 입직 또는 퇴직에 의해 일자리의 변화 없이 근로자만 바뀐 경우

- 신규일자리: 동일한 기업체에서 두 기간(당해년도와 전년도)에 걸쳐 기업체의 생성 또는 기업체의 사업확장에 의해 늘어난 일자리

- 소멸일자리: 동일한 기업체에서 두 기간(당해년도와 전년도)에 걸쳐 기업체의 소멸 또는 기업체의 사업축소에 의해 사라진 일자리

- 일자리 순증: 신규일자리와 소멸일자리의 합. 다만, 신규와 감소의 차이인 일자리순증 총계는 전년도와 비교한 증감과 정확히 일치하나 그룹 내에서 그룹 간 일자리 대체 등이 있어 하위그룹에서는 일치하지 않을 수 있음

- 변동일자리: 대체일자리와 일자리순증을 합한 것

- 임금근로일자리 변동률: 두 기간(전년도와 당해연도)의 평균일자리 중에서 기업체의 사업의 확장(창업) 또는 축소(폐업)에 따른 일자리의 변동 정도

$$\text{* 산식} = \frac{(\text{신규일자리}+\text{소멸일자리})/2}{(\text{당해년도 일자리}+\text{전년도 일자리})/2} \times 100$$

9.5 항목 개발

9.5.1 활용 자료

임금근로일자리 행정통계를 작성하기 위해 우선, 원자료의 활용 가능성을 검토하였다. 종사자 정보는 국민연금, 건강보험, 고용보험, 근로소득지급명세서를 자료원으로 활용할 수 있고, 기업체 정보는 부가가치세, 법인세, 법인등기자료, 산재보험, 사업자등록자료, 사업장현황신고, 전국사업체조사, 기타소득자료의 활용이 가능하다. 행정자료로부터 임금근로일자리 통계생산에 필요한 가입자, 사업장, 기업체 자료원에서 추출할 수 있는 항목은 아래 표 9.10과 같다.

표 9.10 항목추출 행정자료

항목 \ 자료원	가입자						사업장			기업체					
	국민연금	고용보험	건강보험	일용근로소득	기타소득	근로소득지급명세서	국민연금	산재보험	건강보험	부가가치세	사업장현황신고	법인세	법인등기자료	사업자등록번호	전국사업체
개인식별번호	○	○	○	○	○	○									
(보안식별번호)	○	○	○	○	○	○									
자격취득일자		○	○	○	○	○									
학력		△													
사업장관리번호	○	○	○	○	○	○									
사업장관리번호							○	○	○						
사업자등록번호							○	○	○						
사업주개인식별번호							○	○	○						
사업자등록번호										○	○	○	○	○	○
행정동										○	○	○	○	○	○
산업분류								○		○	○			○	○
창설연월												○	○	○	○
종사자														○	○
매출액										○	○	○	○	○	○
자본금												○	○		○

9.5.2 생산가능 항목 검토

현재 활용할 수 있는 행정자료의 가입자, 기업체 정보에서 작성 가능한 항목을 12개로 정리하였다. 종사자 항목은 성별, 연령, 근속기간, 교육정도, 소득 등 5개 항목의 작성이 가능하나, 교육정도는 2015년 등록센서스 이후로, 소득 항목은 타 자료와의 비교분석 등 품질점검을 충분히 할 필요성이 있다는 의견이 많아 공표항목 검토에서 일단 유보하였다.

기업체 항목은 기업형태, 산업분류, 종사자수, 존속기간, 매출액, 자본금, 공공분야 등 7개 항목에 대해 작성이 가능한데 자본금은 자료 검증·보완이 필요할 것으로 판단하였다(표 9.11 참조).

표 9.11 기업형태별 주요 통계항목

구 분	행정자료	통계항목	기업 형태							
			회사법인		회사이외법인		정부	비법인단체		개인기업
종사자 (5개)	개인식별자료	**성별**	O	O	O	O	O	O	O	O
		연령	O	O	O	O	O	O	O	O
	학력	**교육정도**	O	O	O	O	O	O	O	O
	자격취득일 (입사일)	**근속기간**	O	O	O	O	O	O	O	O
	소득	**소득**								
기업체 (7개)	사업자등록	**기업형태**	O	O	O	O	O	O	O	O
	산업분류	**산업분류**	O	O	O	O	O	O	O	O
	법인분류등	**공공/비공공**								
	종사자	**종사자수**	O	O	O	O	O	O	O	O
	존속기간	**존속기간**	O	O	O	O	O	O	O	O
	매출액	**매출액**	O	O						
	자본금	**자본금**	O	O						

9.5.3 항목별 생산가능성 검토

항목별 생산가능성은 통계품질 관점에서 볼 때,

- 첫째, 포괄범위가 어느 정도인가?

- 둘째, 시계열 유지가 가능한가?
- 셋째, 유사통계와 정합성이 맞는가?
- 넷째, 통계 내에서 유사항목 간에 일관성이 있는가?

등을 사전에 충분히 검토해야 한다.

항목의 생산가능성은 경제활동인구조사와 비교하여 검토하였다.

1) 성별

성별은 포괄범위가 81%이며, 경제활동인구조사와 성별분포가 유사하고, 시계열의 일관성이 있는 것으로 나타났다.

성 별	임금근로일자리(12월)			경활 임금근로자(12월)			차이 (A-B)
	2009	2010	증감(A)	2009	2010	증감(B)	
전 체	13,486	14,065	579	16,555	17,154	599	△20
남 자	8,646	8,908	263	9,591	9,889	298	△35
여 자	4,840	5,157	317	6,964	7,265	301	16

(단위 : 천명, 천개)

2) 연령

행정자료의 연령별 분포는 경제활동인구조사와 유사하며 시계열의 일관성도 있는 것으로 나타났다.

연 령	임금근로일자리(12월)			경활 임금근로자(12월)			차이 (A-B)
	2009	2010	증감(A)	2009	2010	증감(B)	
전 체	13,486	14,065	579	16,555	17,154	599	△20
15~19세	78	90	13	180	201	21	△8
20~29세	2,861	2,740	△116	3,476	3,414	△62	△54
30~39세	4,189	4,332	144	4,749	4,847	98	46
40~49세	3,597	3,805	206	4,410	4,608	197	9
50~59세	2,107	2,371	261	2,667	2,909	242	19
60세이상	653	727	72	1,072	1,174	103	△31

(단위 : 천명, 천개)

3) 종사자규모

종사자 규모별 분포의 경우에도 경제활동인구조사와 유사하며, 시계열상으로도 일관성이 있는 것으로 확인되었다.

종사자규모	임금근로일자리(12월)			경활 임금근로자(12월)			차이 (A-B)	고용DB (2010)
	2009	2010	증감(A)	2009	2010	증감(B)		
전 체	13,486	14,065	579	16,555	17,154	599	△20	10,185
100인 미만	5,982	6,285	303	12,962	13,402	440	△137	(300인미만) 6,837
100~300인 미만	1,595	1,673	77	1,646	1,711	64	13	
300인 이상	5,909	6,107	198	1,946	2,041	95	103	3,348

(단위 : 천명, 천개)

4) 산업분류

산업분류별 분포는 경제활동인구조사와 유사하며, 시계열의 일관성을 유지하고 있는 것으로 나타났다.

산업(대분류)	임금근로일자리(12월)			경활 임금근로자(12월)			차이 (A-B)
	2009	2010	증감(A)	2009	2010	증감(B)	
전 체	13,486	14,065	579	16,555	17,154	599	△20
농림어업(A)	44	47	3	127	148	21	△18
광업(B)	19	19	-	21	25	4	△4
제조업(C)	3,238	3,442	204	3,278	3,577	299	△95
건설업(F)	902	903	1	1,322	1,333	10	△9
도소매·음식숙박업(G,I)	1,518	1,613	95	3,135	3,227	91	4
사업·개인·공공·서비스업(E, L~U)	5,903	6,208	305	6,476	6,655	179	126
전기·운수·통신·금융업(D, H, J, K)	1,861	1,834	△28	2,196	2,190	△6	△22

(단위 : 천명, 천개)

5) 근속기간별

근속기간별의 분포가 경제활동인구조사와 유사하게 나타났다.

근속기간	임금근로일자리(12월)		경활 임금근로자(12월)		차이 (A-B)
	2010	구성비(A)	2010	구성비	
전 체	14,065	100.0	17,154	100.0	-
1년 미만	4,409	31.3	5,899	34.4	△3.1
1~3년 미만	3,209	22.8	3,890	22.7	0.1
3~5년 미만	1,755	12.5	1,866	10.9	1.6
5~10년 미만	2,068	14.7	2,357	13.7	1.0
10~20년 미만	1,627	11.6	2,093	12.2	△0.6
20년 이상	997	7.1	1,048	6.1	1.0

(단위 : 천명, 천개)

6) 공공부문 일자리통계 생산가능성 검토

다양한 행정자료(사업자등록자료, 중앙·지방정부 소속현황자료, 내부문서)를 활용하여 공공부문 일자리통계의 생산가능성을 사전에 검토한 결과, 행정자료(사업자등록자료)와 경제활동인구조사의 공공부문 일자리 차이는 2만 8천개로 유사한 것으로 확인되었다.

다른 행정자료(중앙·지방정부 산하기관에 대한 재정부와 안행부의 내부자료)와 경제활동인구조사와의 차이는 10만 8천개로 다소 차이가 있었다.

경제활동인구조사 공공부문 ('11.11.15일, 내부자료)	행정자료 공공부문 (2011 임금근로일자리통계를 재분류)	
	사업자등록자료 기준	내부문서 기준
2,113천명 (정부, 산하기관)	2,085 천개 (가집계) (정부 1,664천, 공기업 등 산하기관 421천)	2,005 천개 (가집계) (정부 1,664천, 중앙산하기관 291천*, 지방산하기관 50천**)

* 기획재정부 내부자료 (중앙산하기관 정규직 248천, 비정규직 43천, '12년 기준)

** 안행부(공기업과) 내부자료 (지방산하기관 정규직 50천명, 비정규직은 자료없음, '11년 기준)

이상과 같이 생산가능한 항목을 검토한 결과 공표 가능한 항목을 아래와 같이 도출하였다(표 9.12 참조). 단, 매출액과 자본금은 자료의 품질을 심층적으로 충분히 검증하여 공표할 필요가 있다고 판단하였다.

표 9.12 임금근로일자리 행정통계 공표 가능 항목

구분	행정자료	통계항목	기업 형태				
			회사법인	회사이외법인	국가지방자치단체	비법인단체	개인기업
근로자 (3개)	개인식별자료	성별	O	O	O	O	O
		연령	O	O	O	O	O
	자격취득일자(입사일)	근속기간	O	O	O	O	O
기업체 (6개)	사업자등록	기업형태	O	O	O	O	O
		공공/비공공부문	O	O	O	O	O
	산업분류	산업분류	O	O	O	O	O
	종사자	종사자수	O	O	O	O	O
	존속기간	존속기간	O	O	O	O	O
	매출액	매출액	O				
	자본금	자본금	O				

9.6 지표 개발

9.6.1 통계지표 개발 검토

1) 지속일자리

지속일자리는 동일한 기업체에서 당해년도와 전년도에 걸쳐 지속적으로 동일 근로자에 의해 일자리가 점유된 경우로 2년간(t년도, t-1년도)의 종사자 정보를 이용하면 동일인임을 확인할 수 있으므로 지표 작성에 문제가 없다는 의견에 전문가 및 관계부처 담당자가 공감하였다.

2) 신규일자리 및 대체일자리

전년도에는 존재하지 않았으나 작성기준년도에 신규로 등록된 종사자는 신규일자리이거나 대체일자리에 해당된다. 즉, 어떤 기업체에서 몇 명의 직원을 새로 채용하였을 경우에 순수하게 새로 생긴 일자리인지 아니면 퇴직자를 보충하기 위해 채용한 대체일자리인지 구분을 해야 한다.

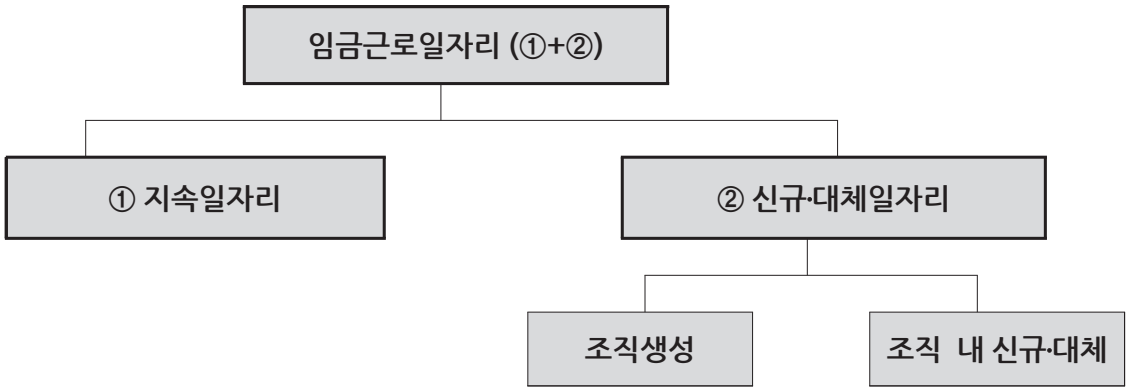
새로 입사한 직원의 신규일자리 또는 대체일자리를 구분하기 위해서는 기업체의 구인정보, 채용정보, 조직 확장 등과 같은 상세 정보와 신규자의 종사상 지위, 직종, 입직 시기 등의 정보가 필요하다. 그러나 현 행정자료에는 입직시기와 고용보험의 직종자료만 존재하기 때문에 신규/대체일자리 구분하여 공표하는 것은 충분한 연구가 필요하다는 의견이 많았다.

특히, 신규일자리는 일자리 창출 등 정부정책과 국민의 최대 관심사이므로 자칫 객관적인 설명이 충분히 뒷받침 하지 못하는 통계지표를 작성하기에는 문제가 발생할 수 있으므로 신규일자리와 대체일자리를 구분하는 방법은 좀 더 연구할 필요가 있다고 판단하였다.

따라서, 신규/대체일자리는 당분간 구분하지 않고 통합하여 공표하되, 향후 심층적인 연구를 통해 도출한 결과를 검토하여 공표여부를 결정하기로 방향을 설정하였다.

9.6.2 통계지표 작성(안)

임금근로일자리 통계를 작성하기 위해 전문가 및 관계부처와의 수차례에 걸친 논의를 하였고, 그 결과를 아래와 같이 정리하였다.



임금근로일자리 행정통계 지표 체계

○ **통계지표 정의**

- **(임금근로일자리)** 기업체에서 현물 또는 현금을 대가로 상품생산이나 서비스 활동을 하는 임금근로자가 '점유한 고용위치'(Filled employment position)

- **(지속일자리)** 동일한 기업체에서 당해년도와 전년도에 걸쳐 지속적으로 동일근로자에 의해 일자리가 점유된 경우

- **(신규.대체일자리)** 통계기준시점 현재 당해년도와 전년도 사이에 조직생성이나 조직 내 신규.대체로 인해 변화된 일자리

  ■ **(조직생성)** : 법인 설립 등 새로운 조직 생성에 따른 신규일자리
  ■ **(조직 내 신규.대체)** : 동일 조직에서 조직 확장이나 근로자의 입.퇴직 등으로 인한 일자리 변화

- **(일자리 증감)** 전년 대비 일자리 수의 증가 혹은 감소한 정도

  * 일자리증감 = 2010.12.31현재 일자리수 - 2009.12.31현재 일자리 수

9.7. 임금근로일자리 행정통계 작성 방법

이제까지는 행정자료를 활용한 외국의 임금근로일자리 통계작성 사례와 우리나라에서의 작성여건과 개발 가능성에 대해 살펴보았다. 이번 장에서는 작성방법론과 과정에 대해 살펴보기로 한다.

9.7.1 작성 방법론

행정자료를 활용하여 통계를 작성하기 위해서는 몇 가지 관점에 주목할 필요가 있다. 우선, 모집단을 포괄할 수 있는 범위를 최대한 확대해야 하고, 시계열 유지 및 정확성을 확보해야 한다. 그리고 중복을 제거하거나 무등록 자료의 통계적 처리 과정을 거쳐야 한다.

9.7.2 포괄범위 확대

행정자료는 이용목적에 따라 그 대상을 달리하므로 통계작성 대상인 추정 모집단 규모에 근접하기 위해서는 필요한 행정자료를 모두 활용하여 포괄범위를 확대해야 한다. 즉, 개인식별번호를 활용하여 행정자료 간의 신고자 중복과 누락을 점검하고, 작성대상자를 최대한 확보해야 한다.

예를 들어, 건강보험은 가입하였으나 고용보험은 가입하지 않은 경우 고용보험만을 활용하여 통계를 작성한다면 건강보험에 가입한 근로자가 누락될 것이다. 반대의 경우에도 마찬가지이다.

표 9.13은 사회보험 3종과 근로소득 신고자의 중복여부를 체크하여 집계한 것이다. 2012년 말 기준으로 근로소득 신고 건수가 1,341만건으로 가장 많다. 이중에서 사회보험에는 가입하지 않고 근로소득만 신고한 자가 156만건이나 된다. 이들도 임금근로일자리 대상에 모두 포함해야 하므로 임금근로일자리는 총 1,591만 3천개로 포괄범위를 최대화하여 작성함이 바람직하다.

표 9.13 사회보험 가입자 및 근로소득 신고자 수(2012.12.31. 기준) (단위 : 천명)

구 분	계(A+B)	중 복 여 부					
		없음(A)	있음 (B : B1+B2+B3)				
			소계(B)	건강/국민(B1)	건강/고용(B2)	고용/국민(B3)	건강/국민/고용(B3)
계*	15,913,000	4,317,646	11,595,354	943,661	845,668	356,756	9,449,269
건강보험	13,315,075	2,076,477	11,238,598	943,661	845,668	-	9,449,269
고용보험	11,184,500	532,807	10,651,693	-	845,668	356,756	9,449,269
국민연금	10,900,152	150,466	10,749,686	943,661	-	356,756	9,449,269
근로소득	13,410,365	1,557,896	-	-	-	-	-

* 중복이 있는 경우의 합계는 세로 합계가 아니며, 비임금근로자를 제외한 결과임

9.7.3 시계열 안정

행정자료는 정부정책에 종속되므로 정부정책이 바뀌는 경우에 단일자료를 활용한다면 시계열이 지나치게 변경될 수 있으므로 시계열 안정화를 위해서는 필요한 행정자료를 통합하여 통계를 작성해야 한다.

최근 사회보험의 적용대상 확대, 근로소득신고 강화 정책에 따라 사회보험 가입률이 증가하고 있는 추세이므로 실제 증가한 임금근로일자리와 차이가 발생할 수 있으므로 작성기준시점별로 재산정한 시계열 안정화가 필요하다(표 9.14 참고).

표 9.14 연도별 임금근로자 증감 현황 (단위 : 천명, %)

	2010년	2011년			2012년		
			증감	증감률		증감	증감률
임금근로	14,974	15,505	531	3.5	15,913	408	2.6
건강보험	12,218	12,806	587	4.8	13,323	517	4.0
고용보험	9,775	10,348	573	5.9	10,835	487	4.7
국민연금	10,232	10,855	623	6.1	11,397	541	5.0
근로소득	12,276	12,897	621	5.1	13,410	513	4.0

* 비임금근로일자리는 제외한 결과임

9.7.4 정확성 제고

사회보험 가입률 증가가 경제활동인구조사의 취업자 증가 보다 많기 때문에 실질적인 노동시장의 변화를 과대 계상할 소지가 있으므로 행정자료의 통합은 정확성 제고 측면에서 반드시 필요하다. 표 9.15는 정확성 제고를 위하여 파악하여야 할 사회보험 적용 제외자를 정리한 것이다.

표 9.15 사회보험 적용 제외자

	의료급여자 또는 유공자	외국인근로자	60세 또는 65세이상	기타
건강보험	○생활유지 능력이 없거나 일정수준 이하의 저소득층을 대상으로 국가재정에 의하여 기본적인 의료혜택을 제공하는 의료급여자 ○독립유공자 예우에 관한 법률·국가유공자등 예우 및 지원에 관한 법률에 의하여 의료보호를 받는 자			
국민연금			60세 이상(다만, 가입기간이 부족하여 연금을 받지 못하거나 가입기간을 연장하여 더 많은 연금을 받고자 하는 경우는 임의 가입)	○ 타 공적연금 가입자인 공무원, 군인, 사립학교 교직원 및 별정우체국직원 ○ 조기노령연금 수급권자 등
고용보험		외국인 근로자는 적용제외, 다만 다음의 근로자에 대하여는 적용대상 : 출입국관리법 시행령 제12조의 규정에 의한 외국인 체류자격 중 ①거주(F-2), 영주(F-5)의 체류 자격, ② 주재(D-7), ③ 기업투자(D-8), ④무역경영(D-9)의 체류자격을 가진자(상호주의에 따라 판단)	65세 이상인 자 (실업급여만 제외)	○ 공무원, 교사 등 특정직종 근로자 ○ 소정근로시간이 대통령이 정하는 시간 미만인 자 ○ 월간소정근로시간이 60시간(주간 소정근로시간 15시간) 미만인 근로자

9.7.5 통계 처리

1) 중복 처리

동일한 일자리를 점유하고 있는 근로자가 사회보험(3종 : 국민연금, 고용보험, 건강보험) 가입여부 및 근로소득 신고에 따라 행정자료 간의 개인식별번호 중복이 발생하게 된다.

개인식별번호가 중복되는 경우에는 한 개의 내용만 선택하여야 하는데, 행정자료의 정확성과 경제활동인구조사 등 유사통계와의 정합성을 고려하여 선택해야 한다.

2) 무등록 처리

행정자료의 일부 항목은 무등록자료가 존재하거나 자료연계 과정에서 행정자료 간의 등록 항목 차이로 결측치가 발생할 수 있다. 이렇게 발생한 결측항목은 통계적 방법으로 처리해야 하는데, 확률분포를 아는 경우와 모르는 경우에 따라 달리 적용하였다.

확률분포를 아는 경우에는 확률에 의해 대체(Probabillty Imputaion)하고 모르는 경우에는 '대체군을 이용한 대체 방법(hotdeck)'을 사용하였다.

9.7.6 통계지표 작성

행정자료는 통계목적으로 작성되지 않아 자료의 정의 및 필요한 지표가 없으므로, 통계목적에 맞게 자료를 재분류하고 조합하여 가공하는 절차가 필요하다. 임금근로일자리 통계지표(지속일자리와 신규·대체일자리)를 작성은 두 개년도의 자료 비교를 통해 가능하다.

지속일자리는 전년도와 당해 연도에 동일한 근로자에 의해 점유된 일자리로 2개년도의 종사자 정보를 비교하여 동일인인 경우에 해당되고, 신규·대체일자리는 당해 연도에 신규로 등록된 종사자가 해당된다.

9.7.7 작성 과정

임금근로일자리 통계과정은 통계작성 모집단 및 항목 생성과 통계적 처리 및 지표 작성 과정으로 구분할 수 있다.

일자리 모집단 및 항목 생성을 위해서는 우선, 연계키를 사용하여 행정자료를 통합해야 한다. 통합한 자료를 활용하여 종사자 및 기업체 DB를 구축하고, 인적 항목과 기업체 항목을 각각 생성한 후에 기업체 항목과 인적속성을 결합한 통합 DB를 생성한다.

다음으로는 중복 및 비임금근로일자리를 제거하고 무등록자료를 대체한 후에 마지막으로 일자리 통계지표를 작성한다. 전체 과정을 그림 9.4로 정리하였다.

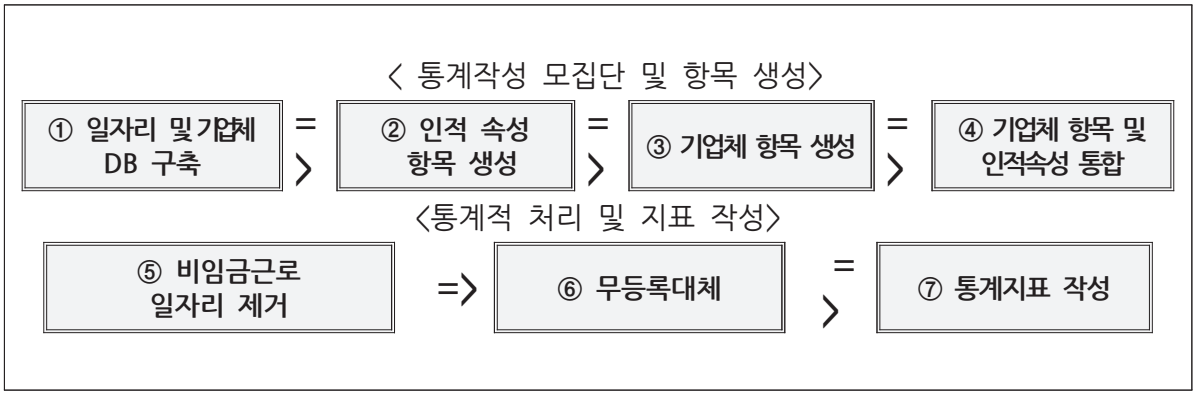


그림 9.4 임금근로일자리 행정통계 작성 과정

1) DB 구축 및 항목 생성

① 1단계 : 일자리 및 기업체 DB 구축

일자리 DB를 생성하기 위해서는 우선, 개인식별번호를 연계키로 사회보험 3종(국민연금, 고용보험, 건강보험)의 종사자 정보를 통합한다. 중복이 있는 경우는 먼저 가입한 날짜를 기준으로 통합하고, 중복되지 않는 경우에는 모두 포함되도록 한다(그림 9.5 참조).

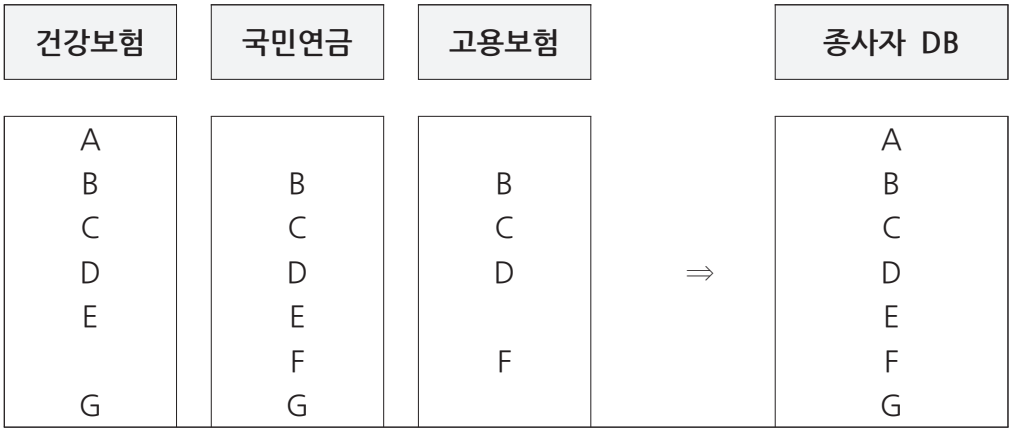


그림 9.5 사회보험 종사자 통합

다음에는 사업장관리번호가 있는 국민연금, 건강보험, 산재보험자료를 활용하여 사업자 정보를 통합한다(그림 9.6 참조).

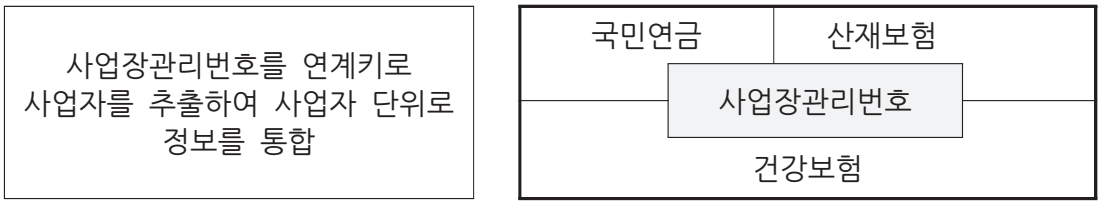


그림 9.6 사업자 정보 통합

“종사자 통합자료”와 “사업장 통합자료”에 공통으로 존재하는 ‘사업장관리번호’를 활용하여 통합한다(그림 9.7 참조).

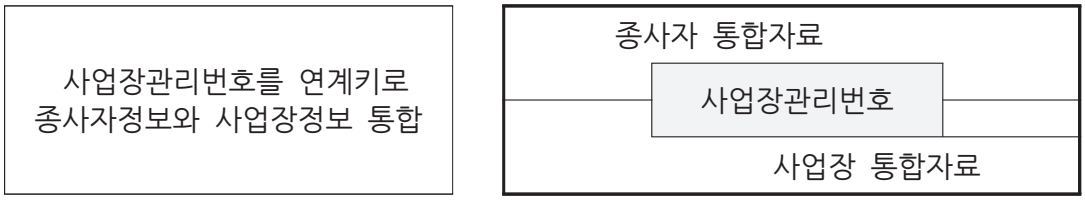


그림 9.7 종사자와 사업장 정보 통합

사회보험 통합자료와 근로소득원천징수명세서를 통합하여 일자리DB를 구축한다. 이때, 연계키는 사회보험DB와 근로소득원천징수명세서에 동일하게 존재하는 생년월일, 성별, 사업자등록번호를 이용한다(그림 9.8 참조).

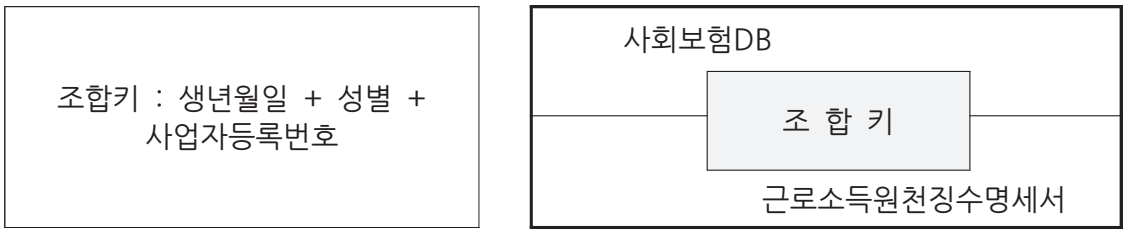


그림 9.8 일자리 DB 구축

일자리DB는 사업자등록번호 단위에서 통계작성기준인 기업체 단위로 전환한다. 기업체는 회사법인, 회사이외의 법인, 정부, 비법인단체, 개인기업체로 구분한다.

② 2단계 : 인적속성 항목 생성

인적항목은 표 9-16과 같이 자료원 및 작성규칙에 따라 추출한다.

표 9.16 인적항목 작성규칙

구 분	자료원 및 작성규칙	활용분야	
		항목생성	무등록대체 요소
개인식별번호	고용보험,국민연금,건강보험,근로소득원천징수* * 생년월일 + 성별		
성별	고용보험,국민연금,건강보험,근로소득원천징수	성별	내/외국인
생년월일	고용보험,국민연금,건강보험,근로소득원천징수	연령	연령
근무시작일자	고용보험>국민연금>건강보험>근로소득원천징수 * 고용보험 기업체단위	근속기간	입·이직 차이
근무종료일자	고용보험>국민연금>건강보험>근로소득원천징수 * 고용보험 기업체단위		
기준소득월액	건강보험>국민연금		1년차 월평균 소득
직종코드	고용보험		직종

③ 3단계 : 기업체 항목 생성

기업체 항목은 표 9.17과 같이 자료원 및 작성규칙에 따라 추출한다.

표 9.17 기업체 항목 작성규칙

구 분	자료원 및 작성규칙	활용분야	
		항목생성	무등록대체 요소
사업자등록번호	사업자등록자료, 건강보험, 국민연금, 산업재해보험, 근로소득원천징수 *무순위	기업형태, 기업체작성	기업형태
법인등록번호	사업자등록자료, 부가가치세, 근로소득원천징수 *무순위	기업체 작성, 공공분야	
공공분야	사업자등록자료	공공/비공공 분야별	공공/비공공분야
산업분류	사업자등록자료, 전국사업체, 산업재해보험	산업분류별	산업분류
매출액	법인세> 부가가치세 *자료 내 동일사업자등록번호의 경우 최대 매출액	매출액규모별	매출액
자본금	법인세> 법인등기자료		자본금
회사설립일 (사업자등록일 포함)	법인등기자료>법인세>사업자등록자료	존속기간별	존속기간별

④ 4단계 : 일자리 DB와 기업체 DB 통합

사업자등록번호를 활용하여 일자리DB와 기업체 DB를 통합한다(표 9.18).

표 9.18 인적항목과 기업체 항목 통합

ID	인적속성							기업체 모집단							
일자리번호 (개인식별번호+기업체)	개인식별번호	성별	생년월일	기준소득월액	직종코드	근무시작일	근무종료일	사업자등록번호	법인번호	기업체	산업분류	매출액	자본금	존속기간	일자리수

2) 통계처리 및 지표 작성

① 5단계 : 비임금근로일자리 제거

통합된 DB는 대부분 임금근로자이나 자영업자가 일부 포함되어 있으므로 제거해야 한다. 국민연금과 고용보험 자료 중에서 영세 자영업자 또는 일반자영업자는 비임금근로일자리로 처리한다(표 9.19).

표 9.19 비임금근로 일자리 제거

구 분		국민연금*	고용보험	일자리행정통계
임금근로자				
자영업자	고용인이 없는 자영업자			비임금근로자
	50인 미만 고용인이 있는 자영업자			
	50인 이상 고용인이 있는 자영업자			

② 6단계 : 무등록자료 처리

행정목적과 통계목적이 달라 통계기준에 맞게 작성되지 않은 경우, 정책 대상에서 제외되어 등록되지 않은 경우와 집행과정에서 미등록된 경우 등 무등록 자료가 존재한다.

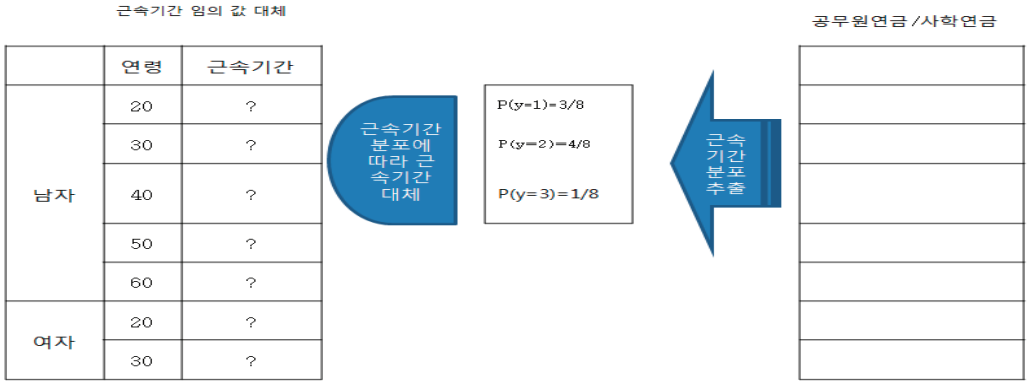


그림 9.9 확률에 근거한 대체 방법

자료 입수과정에서 특정항목에 대한 정보를 미입수 하였으나 미입수 자료에

대한 확률분포를 아는 경우에는 확률에 근거한 대체(Probabillty Imputaion) 방법을 적용하였다. 즉, 각 무응답 값을 관측된 값들의 구성 비율에 비례하는 확률로 대체하는 방법을 사용하였다. 예를 들어, 공무원 및 교사의 연령분포는 모르나 공무원연금 및 사학연금 자료에서 공무원 및 교사에 대한 확률분포를 추출하여 무등록자료를 대체한다(그림 9.9).

확률분포를 모르는 경우의 무등록자료 대체는 '대체군을 이용한 대체 방법(hotdeck)'을 사용한다. 행정자료는 가입자 등록이 연령, 국적 등에 따라 제한되므로 비슷한 유형의 대체방법을 적용한다(그림 9.10 참고).

고용보험	○	×	○	×	○	○	
(연령)	15~17세		18~59세		60~65세		(연령)
(국적)	내국인	외국인	내국인	외국인	내국인	외국인	(국적)
	×	○	○	○	○	×	국민연금

그림 9.10 대체군을 이용한 대체방법

* ○ 등록자료, ×무등록자료

③ 7단계 : 통계지표 작성

행정자료를 통합한 DB에는 작성하고자 하는 임금근로일자리 통계지표가 없으므로 규칙에 따라 2개년(t년, t-1년) 자료를 비교하여 통계지표를 구분한다.

예를 들어, t-1년도에 6명이었던 A기업체의 임금근로자가 t년도에 들어 2명은 정년퇴직하고 3명이 신규로 입사하였다고 가정하자. 이 회사의 근로자는 2년간에 5명의 변동이 발생하였다. 그렇다면 지속일자리와 신규·대체일자리는 몇 개로 작성해야 하는가?

우선, 이 회사의 임금근로일자리는 6개에서 7개로 1개 늘어났다. 변동이 없는 지속일자리는 4개이고, 변동이 있는 신규·대체일자리도 3개이다. 그런데, 퇴직자를 대체한 신규직원이 누구인지 행정자료로서는 확인하기가 어렵다. 그러므로 신규·대체일자리는 구분하지 않고 통합하여 3개로 작성한다(그림 9.11 참조).

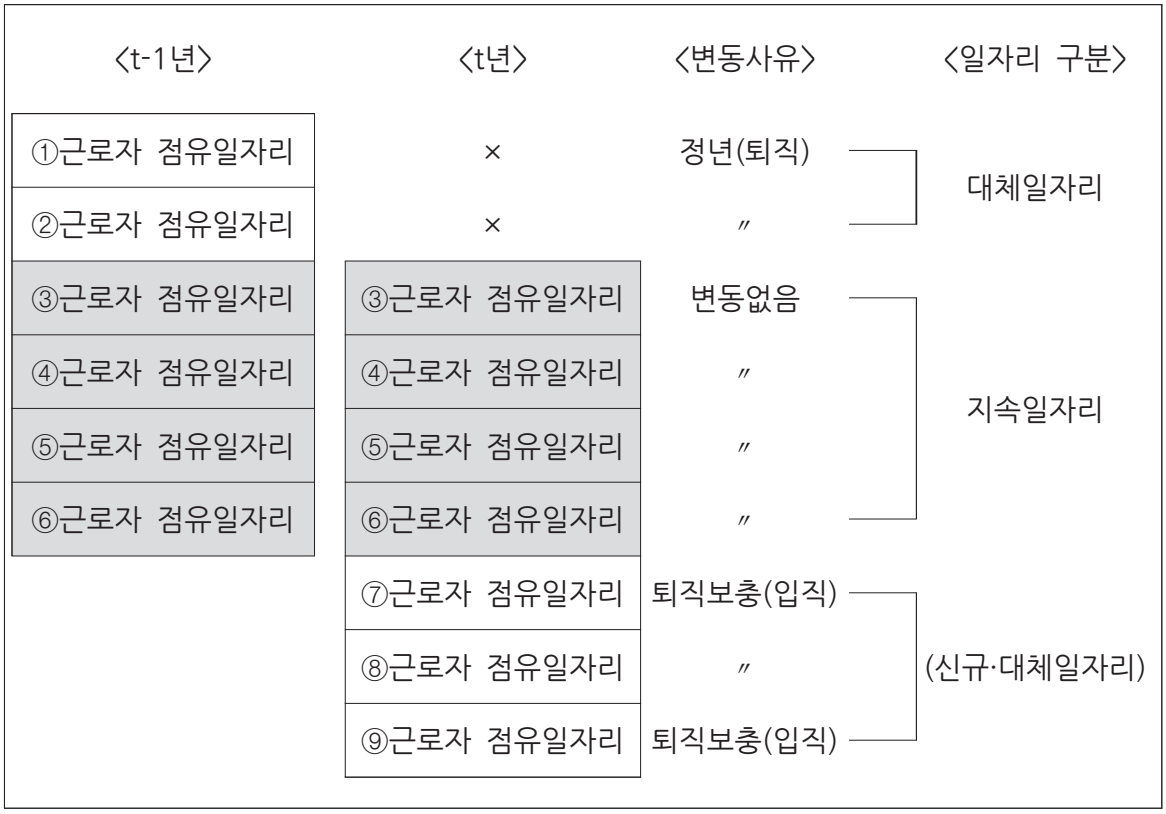
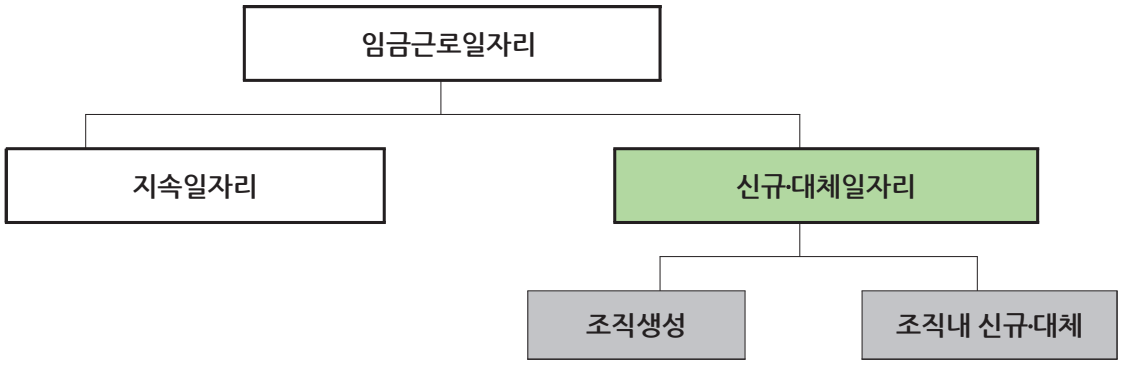


그림 9.11 예시 : A기업체의 지속일자리, 신규·대체 일자리 구분

9.8 향후 개선 과제

2012년부터 행정자료를 활용한 임금근로일자리 통계를 매년 작성하여 제공하고 있으나 일자리 변동 현황과 지역별 일자리를 파악할 수 있는 세분화된 통계는 여전히 부족하다. 현재는 임금근로일자리에 대해서만 통계지표를 작성하고 있는데, 전체적인 일자리 규모를 파악하기 위해서는 비임금근로일자리까지 작성 범위를 확대할 필요가 있다.

〈기 존〉



〈개 선〉

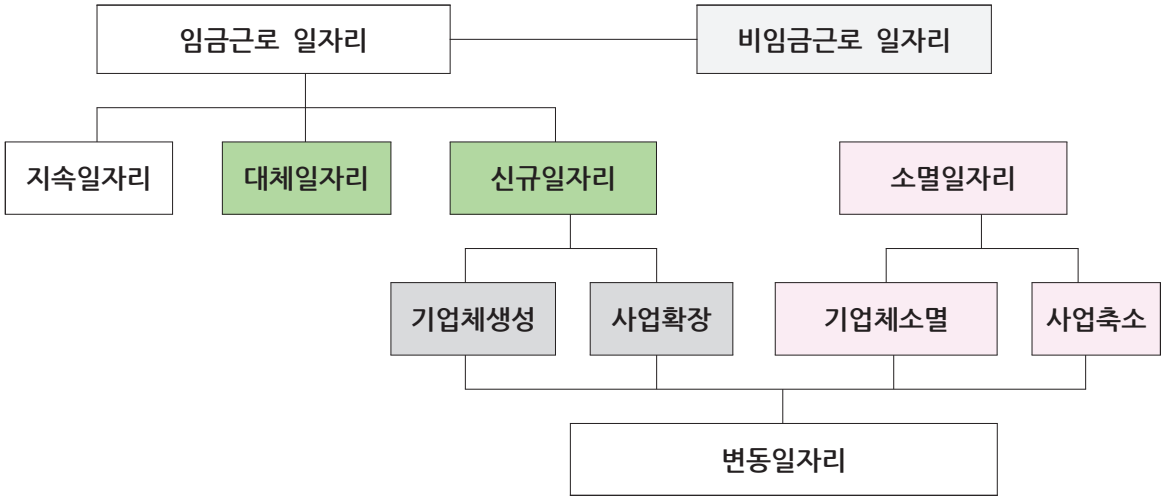


그림 9.12 일자리 통계지표 개선 방향

그리고, 일자리 증감 현황을 파악할 수 있는 통계지표의 세분화가 필요하다. 일자리 변동은 기업체의 생성과 소멸, 사업확장 또는 축소, 퇴직자를 대체하는

신규자 채용에 의해 발생한다. 이러한 변동상태를 파악할 수 있는 일자리 지표의 개선이 요구된다(그림 9.12 참조).

또한, 지역단위의 고용·노동정책을 뒷받침하는 소지역별 일자리 통계가 요구되고 있으나 전국단위의 통계작성에 국한될 수밖에 없는 것이 현재 행정자료의 한계이다. 왜냐하면 사회보험이나 과세자료에는 기업체의 사업장 주소가 본사중심으로 등록되어 있기 때문이다. 즉, 공장, 지사, 지점, 영업소 등이 사업장이 위치한 주소로 등록되어 있지 않고 본사 주소로 등록되어 있어 지역별 일자리통계를 작성하기가 곤란한 실정이다.

그럼에도 불구하고 사업장 주소를 지역별로 현실화하는 방안을 모색하여 지역별 일자리 통계를 생산해야 할 필요성은 커지고 있다. 사업장 주소를 현실화하기 위해서는 두 가지 방안을 검토할 수 있다.

첫 번째는 행정자료를 보유하고 있는 기관에서 신고서식의 사업장 주소를 실제 위치하고 있는 장소로 수정하는 것이다. 이 방법은 중앙행정기관의 의지와 신고서를 취급하는 일선기관의 절대적인 협조가 필요하다. 수천만 건에 해당하는 행정자료를 전면적으로 수정하는 작업은 쉽지 않기 때문이다.

두 번째는 다른 데이터를 이용하는 방법이다. 대부분의 직장인은 휴대폰을 이용하고 있기 때문에 통신사의 위치정보를 활용하면 근무지 추적이 가능할 것으로 보인다. 통신사 자료는 민간기업이 보유하고 있으므로 유상입수가 불가피하고, 개인정보 보호법에 의해 개인별 위치정보 제공은 불가능하다는 입장이다. 따라서, 이용 가능한 최소한의 구역까지 개인특성별 집계한 자료를 입수하여 근로자의 사업장 주소를 현실화하는 방법론 연구를 진행해 나갈 계획이다.

한편, 현행 행정자료 DB는 개별적으로 구축하여 활용하여 왔기 때문에 상호 연계·통합된 구조라고 볼 수 없었다. 이에 통계청에서는 2014년도에 행정자료 기반의 효율적 통계생산을 지원하기 위해 종합모집단 구축 청사진을 제시하는 ISP(Information strategy plan)를 수립하였다. 종합모집단은 인구·가구, 사업체·기업체, 주택·건물, 고용 등 4개 분야로 나누어 데이터를 상호 연계하여 활용할 수 있도록 구축하는 것이다. 이 사업은 2015년부터 2017년까지 구축할 예정에 있다.

고용분야는 일자리통계등록부를 구축하여 고용모집단 및 신규통계 개발 등에 활용할 계획이다. 일자리통계등록부는 주민등록과 외국인등록부에 등재된 모든 사람의 근로활용, 구직활동, 학업활동, 기타활동(가사 등)을 포괄할 수 있도록 구축할 예정이다. 그러다 보니 구축내용과 용량이 방대하고, 행정자료의 품질이 다양하여 몇 년에 걸친 작업을 진행될 것으로 보인다. 일자리통계등록부를 구축하면 전국 및 지역별 일자리 구조 및 동향통계, 일자리 패널통계 등 다양한 신규통계 생산과 모집단으로 활용이 가능하다. 또한, 지리정보와 연계한 시각화 서비스로 활용범위는 한층 넓어질 것으로 기대된다.

행정자료 활용은 어려워지고 있는 조사환경을 극복하고 저비용·고효율 통계작성의 중요한 수단이므로 행정자료 활용은 계속 확대해 나가야 할 방향임에 틀림이 없다. 그러기 위해서는 통계생산자의 고민과 열정이 무엇보다 요구되며 관계기관과의 협업이 필요하다. 정책수립에 필요한 통계를 선제적으로 발굴하고, 통계생산에 필요한 행정자료 활용을 더욱 활성화하여야 할 것이다.

CHAPTER 10

행정 빅데이터를 위한 오픈소스 소개

학습목표

- 빅데이터 분석에 사용할 수 있는 오픈소스에 대하여 학습한다.
- SAS(Strategy Analysys System)은 오픈소스는 아니며, 관심있는 사용자를 위하여 소개에 포함하였다.
- 오픈소스인 R과 Quantum-GIS를 학습한다.
- Excel을 활용할 수 있는 Power Pivot, Power View, Power Map을 학습한다.

10.1 분석 소프트웨어

10.1.1 SAS 소개[56)]

SAS는 원래 컴퓨터를 이용한 통계분석의 필요에 의해 미국 North Carolina 주의 SAS Institute Inc.에서 개발되어, 다양한 유형의 자료를 SAS 시스템 내에서 포괄적으로 처리, 분석할수 있도록 제반 수단을 제공하는 통계분석 시스템이다.

SAS소프트웨어는 개발 이후 지속적으로 발전되어 통계분석 이외에도 품질관리 및 계량경제모형 분석 등으로 데이터 웨어하우징, 데이터마이닝에 이르기까

56) 김기영외 (2012), 예제로 배우는 SAS 프로그래밍 입문, 자유아카데미 내용을 옮겨오거나 발췌하여 정리한 것임

지 다양한 목적의 작업을 수행할 수 있는 매우 강력한 소프트웨어로 변모하였다. 따라서 현재의 SAS는 단순한 통계패키지의 범위를 넘어서 하나의 고급화된 컴퓨터언어 또는 전략적 응용 소프트웨어로 자리 잡고 있는 실정이다. 더 나아가 SAS는 자금세탁방지, 사기검출 시스템, 예지보전, 전사적 리스크 관리 시스템, 소셜미디어 분석 등의 비즈니스 솔루션을 제공하고 있다.

SAS는 다른 소프트웨어에 비해서 시스템의 접근과 이용절차가 비교적 편리하고, 데이터 파일을 효율적으로 관리할 수 있으며, 다른 소프트웨어에 의해 만들어진 시스템 파일들을 직접 이용하는 것이 가능하고, 기본적으로 SAS 제품 이외에 다양한 수요에 호응하여 공급되는 여러 가지 제품들을 이용할 수 있다는 장점을 가지고 있다. 이와 같은 SAS 시스템은 각각의 기능을 담당하는 다양한 제품들로 구성되어 있는데 그중 중요한 제품들의 기능을 간단히 나열하면 다음과 같다.

- **Base SAS**

SAS/CORE와 함께 SAS 시스템의 근간이 되는 제품으로 데이터셋의 입,출력과 저장 SAS파일의 편집과 관리, 기본적인 통계분석 및 보고서 작성 등의 기능을 가지고 있다. 책에서 서술하고 있는 대부분의 내용은 Base SAS의 기능과 사용법에 대해서 다루고 있다.

- **SAS Enterprise Guide**

데이터 접근 및 가공, 다양한 리포트 작성, 통계적인 분석 등의 과정이 프로그래밍 없이 메뉴 방식으로 가능하도록 지원한다. 따라서 특히 초보적인 SAS 사용자들에게 유용하다.

- **SAS Enterprise Miner**

데이터의 생성 및 추출, 탐색 및 변형, 모형화, 평가 등으로 이루어지는 일련의 데이터마이닝 작업을 하나의 에플리케이션 작업 공간에서 쉽게 수행할 수 있도록 구성되어있다.

- **SAS / ETS**

계량경제 및 시계열 데이터를 이용한 분석 및 예측, 재무분석 등을 지원한다.

- **SAS / IML**

행렬 연산 및 처리를 위한 대화식 기능을 가지고 ldtek. 또한 SAS/IML Studio는 대화식 인터페이스를 제공한다.

- **SAS / GRAPH**

다양한 형태의 플롯과 도표를 작성하고 편집할 수 있도록 지원한다.

- **SAS/ STAT**

전통적으로 SAS에서 가장많이 이용되는 제품 중 하나로서 각종 통계분석의 처리를 포괄적으로 수용하고 있다.

SAS의 주요 기능

1) 데이터 입력과 출력 : 실제적으로 거의 모든 형태의 데이터를 읽어 이를 SAS의 고유 형식을 통해 저장할 수 있고, 다른 형식의 외부 파일로 데이터를 출력할 수 있다.

2) 데이터의 수정과 프로그래밍 : 데이터 값의 수정 및 변환, 새로운 변수의 생성, 누적연산, 오류검정 등을 위해 사용할 수 있는 다양한 프로그램 명령문을 가지고 있다.

3) 정보의 출력 및 보고서 작성 : SAS고유의 형식에 따른 보고서 뿐 아니라 사용자기 원하는 형태로 정보를 출력하고 보고서를 작성 할 수 있다.

4) 기본적인 통계자료 분석: 기초통계분석이나 일변량자료 탐색등 기본적인 통계분석을 수행할 수 있다.

5) 파일의 관리 및 조작 : SAS데이터 파일이 가지는 특별한 구조를 통해 여러 SAS 데이터셋들을 사용자의 필요에 따라 조작하여 원하는 형태의 데이터셋을 창조 할 수 있다.

10.1.2 R 소개[57)]

R프로그래밍 언어(줄여서 R)는 통계 계산과 그래픽을 위해 특화된 프로그래밍언어이자 소프트웨어 환경이다. 뉴질랜드 오클랜드 대학의 로스 이하카와 로버트 젠틀맨에 의해 시작된 R은 GPL 하에 배포되는 S프로그래밍 언어의 구현으로 GNU S라고도 하는데 R은 통계 소프트웨어 개발과 자료 분석에 널리 사용되고 있으며, 패키지 개발이 용이하여 통계학자들 사이에서 통계 소프트웨어 개발에 많이 쓰이고 있다(출처: 위키피디아).

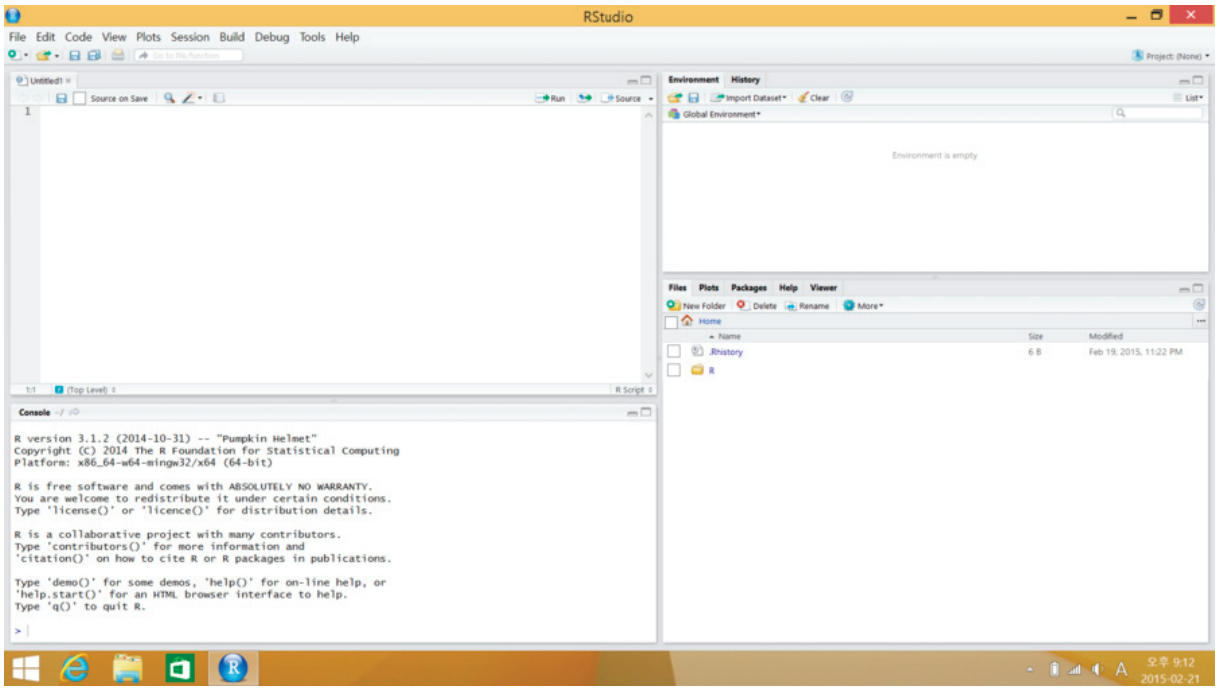


그림 10.1 R프로그램 구성

R의 장점

- R은 다양한 통계 기법과 수치 해석 기법을 지원한다.
- R은 사용자가 제작한 패키지를 추가하여 기능을 확장할 수 있다. (즉, 패키지 개발이 용이하다.)→ 기본 패키지는 R과 함께 설치가 되고, 추가 패키지는 CRAN을 통해 내려 받을 수 있다.
- 그래픽 기능으로 수학 기호를 포함하며, 출판물 수준의 그래프를 제공한다.

57) http://blog.naver.com/kkokkoddu/220279194246

- 행렬 계산을 위한 도구로서 사용될 수 있다.
- 윈도우, 맥 그리고 리눅스를 포함한 UNIX 플랫폼에서 이용 가능하다.
- 데이터 분석에 특화된 프로그래밍 언어로 오픈소스 패키지의 집합체이다.
- 오픈소스 패키지의 집합이기 때문에 통계, 기계학습, 금융, 생물정보학, 그래픽스에 이르는 다양한 패키지가 무료이다.
- 설계를 포함하는 자유로운 데이터 분석이 가능하다.

10.2 지리정보 시스템

10.2.1 QGIS 소개[58)]

10.2.1.1 GIS 개요

지리정보시스템(GIS)은 공간데이터를 생성, 시각화, 검색 및 분석할 수 있는 소프트웨어이다. 공간데이터(Geospatial data)란 객체의 지리적 위치에 관한 정보를 의미한다. 공간데이터는 경위도와 같은 지리적 좌표를 사용하는 경우가 많다. 좀 더 일반적으로는 Spatial data라고 표현하며 지리데이터(geographic data), GIS데이터, 지도 데이터, 위치데이터, 좌표 데이터, 공간 도형 데이터 등으로 불린다. 다음은 공간 분석의 예이다.

GIS 정보에 사용되는 주요 데이터 형태는 다음과 같다.

- **래스터 데이터**

래스터 데이터 중 가장 널리 사용되는 것은 위성 사진이나 항공사진이다. 높이 값 혹은 수치표고모델(DEM: digital elevation model)또한 래스터로 표현됩니다. 지도 객체도 모두 래스터 데이터로 표현될 수 있다. 래스터는 셀(cell) 혹은 사진의 경우 픽셀(pixel)로 구성된 격자형 자료이다. 래스터의 경우 행과 열의 숫자가 고정되어 있습니다. 각 셀에는 숫자가 들어 있으며, 크기가 일정(예: 30m x 30m)하다. 여러 가지 색을 지정해서 표현해야 하는 영상의 경우(예: 적색/녹색/청색을 조합하여 칼라 영상을 만드는 래스터) 래스터를 여러 개 중첩하여 사용한다. 위성영상의 경우 여러 개의 밴드(band)로 데이터를 표현한다. 각각의 밴드는 기본적으로 별도로 분리되었지만, 공간적으로는 위치가 동일한 래스터로서, 각 밴드별로 특정한 파장의 빛의 반사값을 담고 있다. 셀의 크기가 작아질수록 더 자세한 형태를 표현할 수 있지만, 파일의 크기가 커진다. 따라서 셀의 크기는 파일 저장목적과 분석 혹은 지도제작 목적 간에 적절

58) QGIS 시연 - 민경주

히 균형을 맞춰 결정해야 한다.

● **벡터 데이터**

벡터데이터도 공간정보 응용프로그램에서 사용되는 데이터 포맷이다. 벡터는 위치를 좌표로 표현 방법으로 x, y 값을 이용하여 지리적 위치를 표현한다. 이는 일종의 x 축, y 축으로 표현되는 직교좌표계로 고려할 수 있다. 지리적 좌표를 표현하는 방법은 목적에 따라 여러 가지 방법이 있다. 벡터 데이터의 좌표 정의 방법은 다음과 같다.

1. 점 하나의 좌표(x,y)로 구성되며 분리된 지리적 위치를 표현한다.
2. 선 여러 개의 좌표 (x1 y1, x2 y2, x3 y4, ... xn yn)로 구성되는데, 점(x1 y1) 에서 점(x2 y2) 로 선을 그리는 것처럼 일정한 순서로 서로 연결된다. 각 점사이의 부분을 선분(line segment) 이라 한다. 선은 길이가 있으며, 점의 순서에 따라 방향이 있을 수도 있다. 기술적으로는 선은 서로 연결되는 좌표 쌍이며, 선열(line string)은 서로 연결된 여러 개의 선이다.
3. 폴리곤 선이 2개 이상의 점으로 서로 연결되고, 마지막점이 최초의 점과 연결되면 이를 폴리곤이라 한다. 삼각형, 원, 직사각형 등이 모두 폴리곤이다.

10.2.1.2 QGIS 소개

Quantum GIS(QGIS)는 오픈소스 지리정보시스템(GIS :Geographic Information System)이다. QGIS는 2002년 5월 서비스를 시작하였고, 같은 해 6월 SourceForge 프로젝트로 등록되었다. QGIS는 현재 Unix 플랫폼, Windows, OS X 등에서 모두 작동된다. QGIS의 최초 목적은 GIS 데이터 뷰어(data viewer)로 시작하여 현재 많은 분야에서 일상적인 GIS 데이터 검색에 사용되고 있다. QGIS는 여러 가지 래스터 및 벡터 포맷을 지원하고 있으며, 새로운 포맷도 플러그인 구조를 사용하면 쉽게 추가할 수 있다. QGIS의 주요 기능은 다음과 같다.

● 데이터 보기

QGIS는 내부포맷이나 공통포맷으로 변환하지 않더라도 다양한 포맷(벡터, 레스터)을 확인하고 중첩하여 작성할 수 있다.

- 데이터 조사 및 지도 제작
  QGIS를 사용하여 GUI 한경에서 공간데이터를 조회하고 지도를 작성할 수 있다.

- 데이터 생성, 편집, 관리, 내보내기
  QGIS를 사용하여 여러가지 포맷의 벡터지도를 생성, 편집, 관리, 출력할 수 있다. 또한 다른 포맷으로 생성하여 내보내기가 가능하다.

- 데이터 분석
  PostgreSQL/PostGIS 에서 공간데이터를 분석할 수 있으며, fTools Python 플러그인을 이용하면 OGR 지원 포맷에 대해 분석할 수 있다. QGIS는 현재 벡터분석, 샘플링, 자료처리, 기하학적 계산 및 데이터베이스 관리도구를 제공한다.

- 인터넷에 지도 발행하기
  QGIS는 데이터를 mapfile로 내보내어, UMN MapServer 웹서버를 사용하여 인터넷에 업로드 하여 사용할 수 있다.

- 플러그인을 통한 QGIS 기능의 확장
  QGIS는 확장 가능한 플러그인 구조를 사용하여 자신만의 특별한 목적에 맞도록 기능을 추가할 수 있다. QGIS는 플러그인을 제작하는데 사용되는 라이브러리를 제공하고 있다. 아울러 C++이나 Python을 이용해 새로운 응용 프로그램을 개발할 수 있다.

10.2.2 QGIS 설치하기

1) QGIS 공식 홈페이지인 http://qgis.org/ko/site/forusers/download.html 에 접속하여 QGIS 최신 버전의 독립 설치관리자 버전을 내려 받는다.

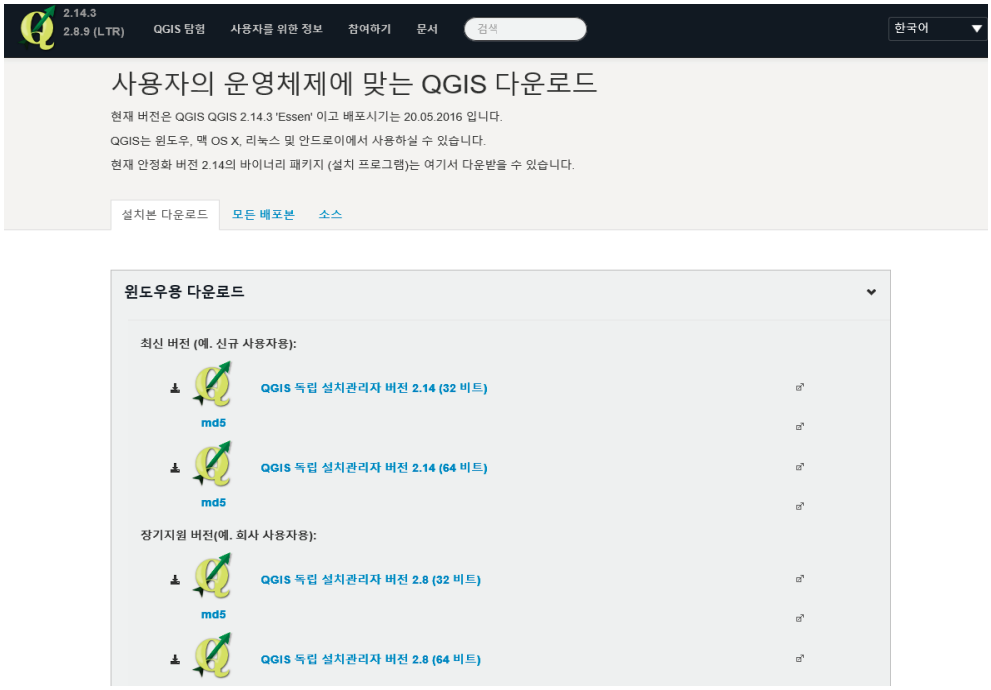


그림 10.16 QGIS 다운로드 페이지

2) 설치파일을 더블클릭하여 설치를 시작한다. QGIS외의 다른 GIS프로그램은 설치하지 않는다.



그림 10.17 설치 그림 1

3) 라이센스 등록 후 기본 옵션만 선택하고 진행한다.

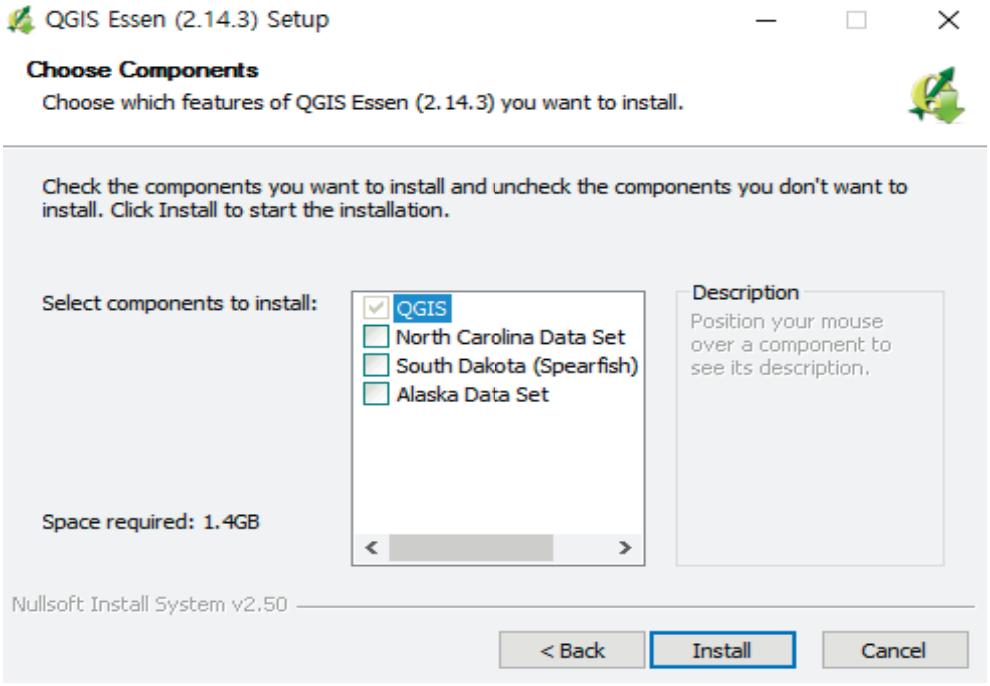


그림 10.18 설치 그림 2

4) 다음을 누르면 QGIS 설치가 진행이 된다.

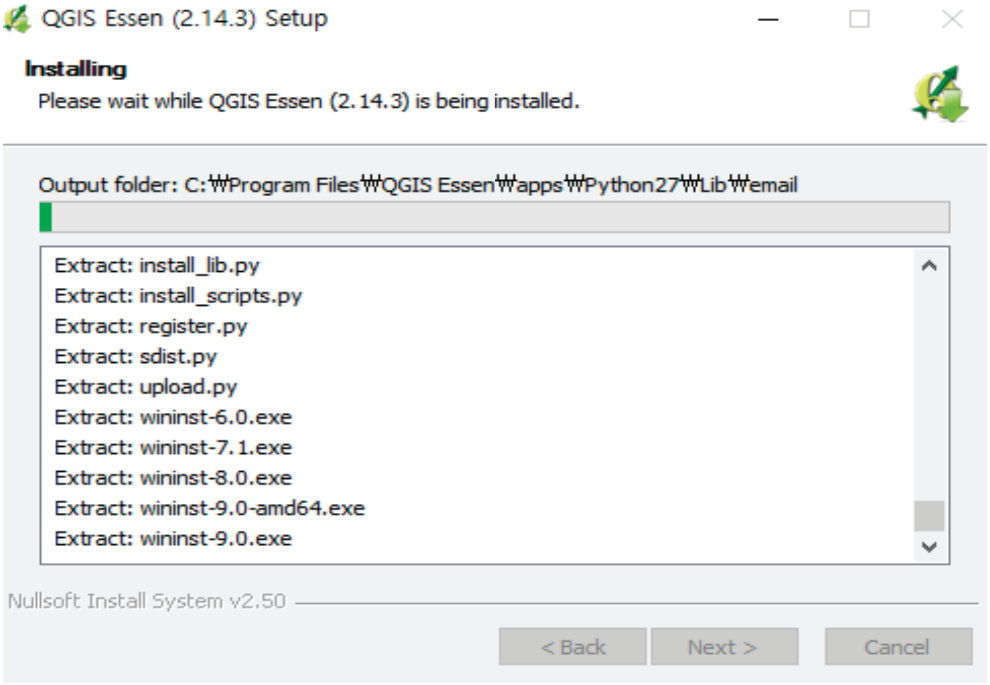


그림 10.19 설치 그림 3

5) 설치를 완료한다.

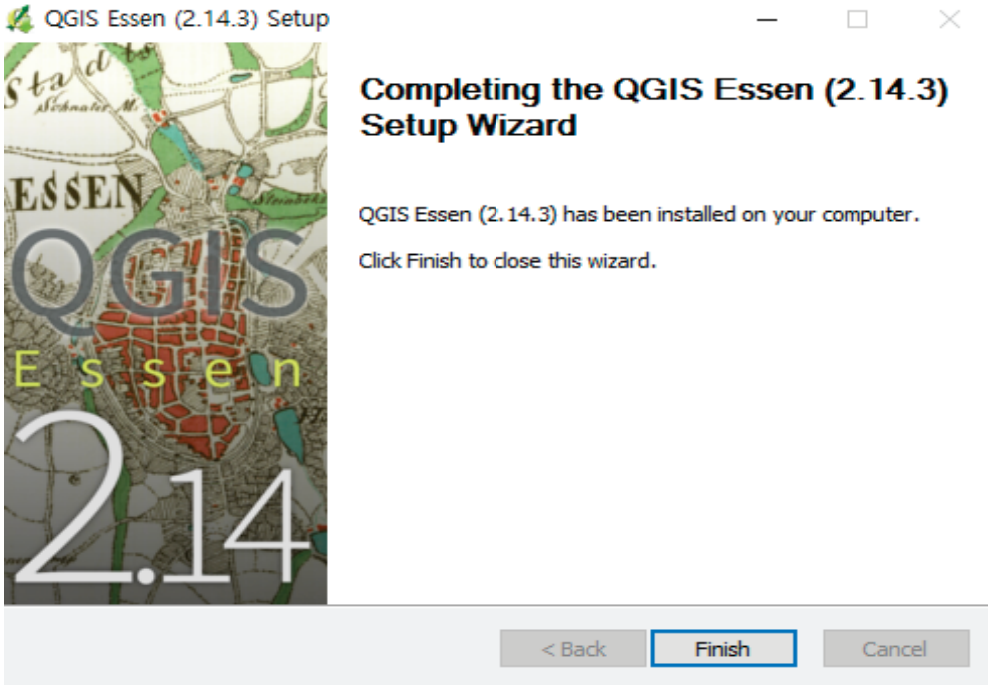


그림 10.20 설치 그림 4

6) 정상적으로 설치되면 QGIS를 실행한다.



그림 10.21 QGIS 실행 화면

7) QGIS의 기본창이 나오게 된다.

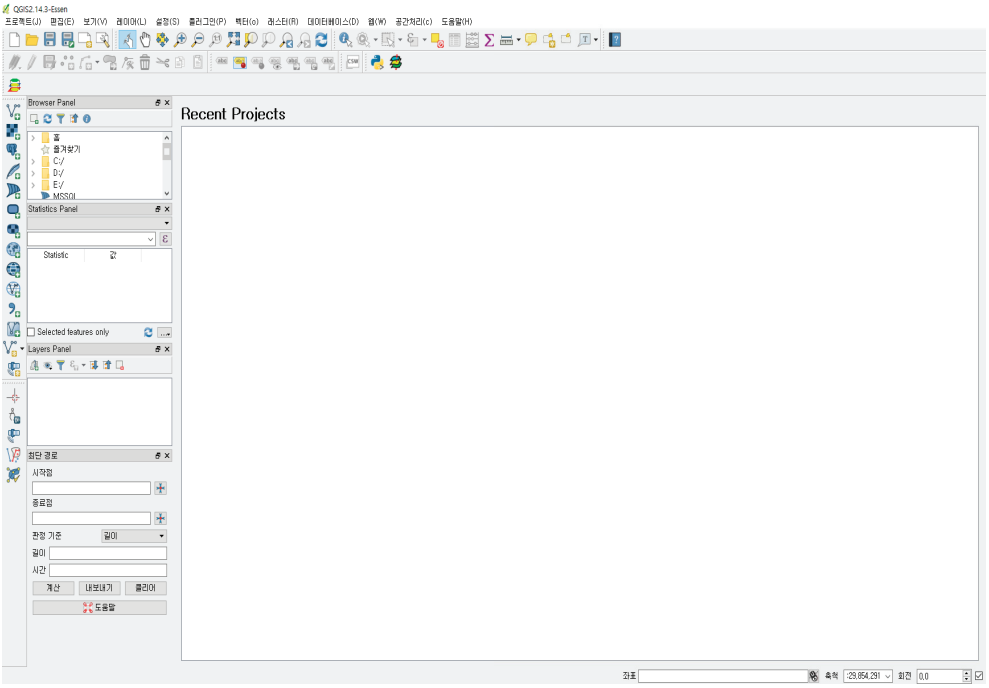


그림 10.22 QGIS 실행 화면

10.2.3 간단한 QGIS 실행법

1) QGIS의 사용자 인터페이스

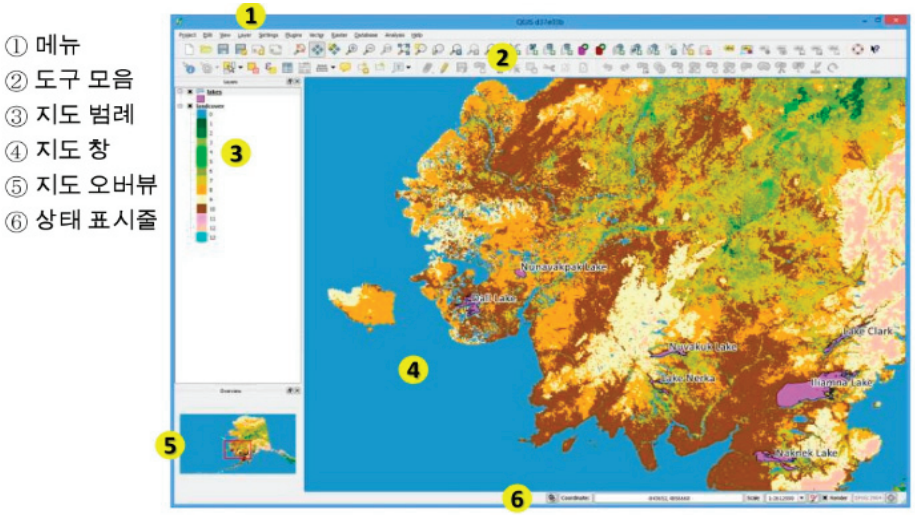


그림 10.23 QGIS 인터페이스 화면

2) 사용자 정의 좌표계

QGIS에서는 세계적으로 사용되는 1000개 이상의 좌표계를 지원하지만 우리나라에서 주로 사용하는 bessel.katac이나 TM중부 같은 좌표계를 모두 포함되지 않아 사용자가 필요한 좌표계를 직접 설정해야 한다.

□ QGIS에서 좌표계 설정하는 방법

1) 설정 > 옵션 > 좌표계 메뉴에서 QGIS에서 지원하는 모든 타원체/좌표계 목록을 확인 할 수 있다.



그림 10.24 QGIS 옵션 창 열기

2) 우측 상단 “선택” 버튼을 누르면 다음 화면 “좌표계 선택”이 나타난다.
3) “좌표계 선택”창 두 번째 항목 “세계의 좌표계”에서 QGIS에서 지원하는 모든 좌표계 목록과 사용자가 추가로 정의한 좌표계를 확인할 수 있다.

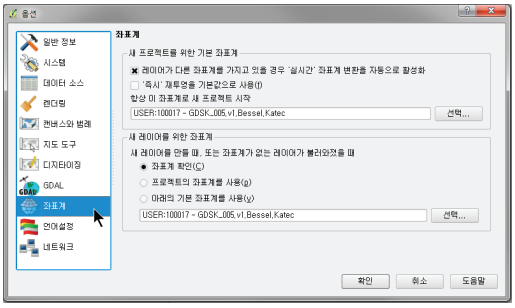


그림 10.25 QGIS 옵션 창 좌표계

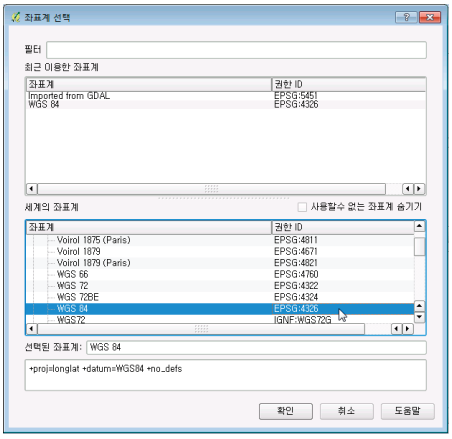


그림 10.26 좌표계 선택 창

□ 국내에서 사용하는 좌표계

국내에서 지도정보에서 주로 사용하는 좌표계는 WGS84 타원체 이다. 다음 표는 주로 사용되는 좌표계를 정리한 표이다.

표 10.1 국내(GDSK)에서 사용하는 좌표계

순번	타원체	좌표계	설명
000	Bessel	경위도	경위도 좌표
001		TM서부	평면좌표계 특성상 원점에서 멀어질수록 오차가 커짐으로 경도원점을 달리한 4개의 TM좌표계가 사용됨.
002		TM중부	
003		TM동부	
004		TM동해	
005		KATEC(TM128)	TM좌표가 여러 개의 원점을 가지고 있고, 제주도 일부 지역이 음수로 표시되는 관계로 이를 대체하기 위해 고안된 좌표계
006		UTM-K	UTM 좌표계의 원점을 우리나라에 맞춘 좌표계
007		UTM-K(수정원점)	UTM-K 경도원점에 10.04초 보정 값을 사용
008		UTM Zone 51	범 세계적으로 같은 기준의 평면좌표를 사용하기 위해 고안된 좌표계
009		UTM Zone 52	
010		TM서부(구원점)	경도 원점에 10.04초 보정하지 않은 원점 좌표 일부 관공서에서 제공받은 데이터가 구원점을 사용함.
011		TM중부(구원점)	
012		TM동부(구원점)	
013		TM동해(구원점)	
014		EKSYS 좌표계	경위도 값, 1/100초 단위
000	WGS84/ GRS80	경위도	
001		TM서부	
002		TM중부	
003		TM동부	
004		TM동해	
005		KATEC(TM128)	
006		UTM-K	
007		UTM-K(수정원점)	
008		UTM Zone 51	

□ 사용자 정의 좌표계 설정방법

1) 설정메뉴 > 사용자 정의 좌표계… 메뉴를 실행한다.

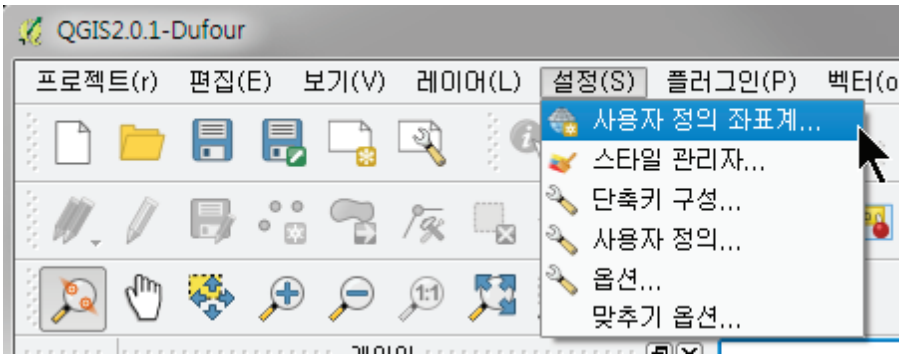


그림 10.27 사용자 정의 좌표계 메뉴

2)“사용자 정의 좌표계 정의” 창에서 “새로운 좌표계 추가” 버튼을 클릭한다.

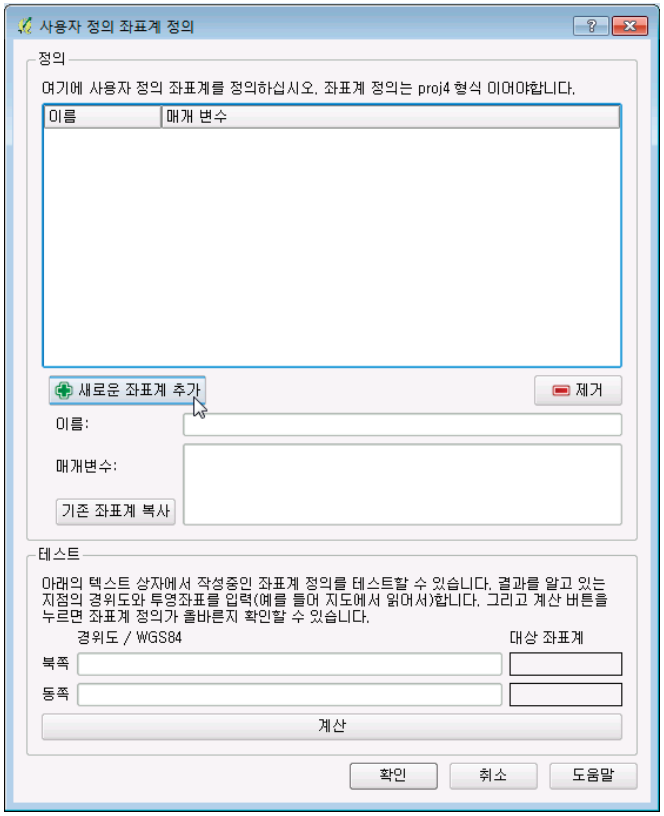


그림 10.28 사용자 정의 좌표계 정의

3) Bessel/카텍 좌표의 재원을 입력한다. “이름”, “매개변수”란에 값을 입력하고 확인을 클릭한다.

이름 : GDSK_005.v1.Bessel.Katec

매개변수 : +proj=tmerc +lat_0=38 +lon_0=128 +k=0.9999 +x_0=400000 +y_0=600000 +ellps=bessel +towgs84=-115.8,474.99,674.11,1.16,-2.31,-1.63,6.43 +units=m +no_defs

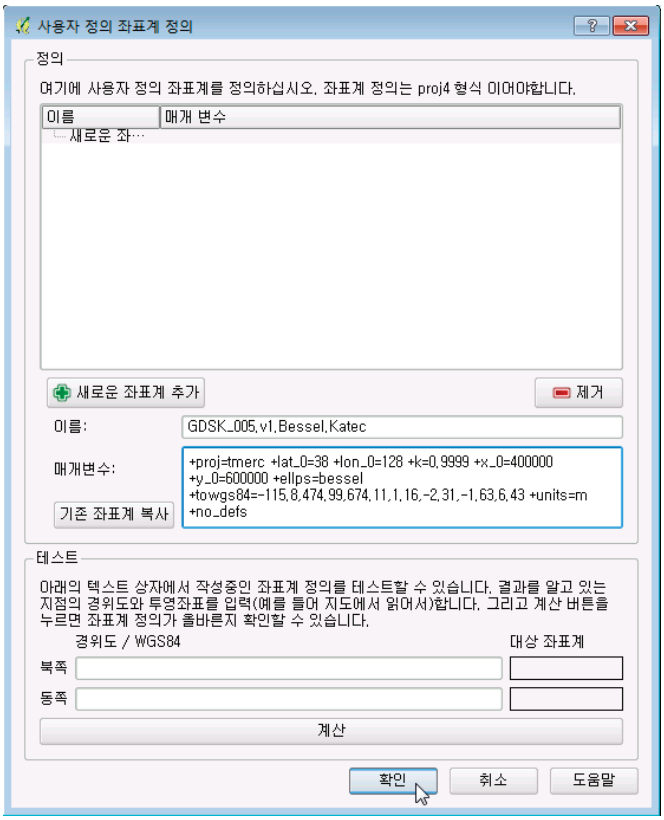


그림 10.29 사용자 정의 좌표계

4) Bessel/Katec 좌표계 확인

“QGIS에서 지원하는 좌표계” 항목에서 설명한 방법으로 추가한 Bessel/Katec 좌표계를 확인할 수 있다. 설정메뉴〉 옵션〉 좌표계〉 좌표계 선택창〉 “세계의 좌표계”〉 “사용자 정의 좌표계” 에서 위에서 정의한 “GDSK_005.v1.Bessel.Katec” 항목을 찾으면 된다.

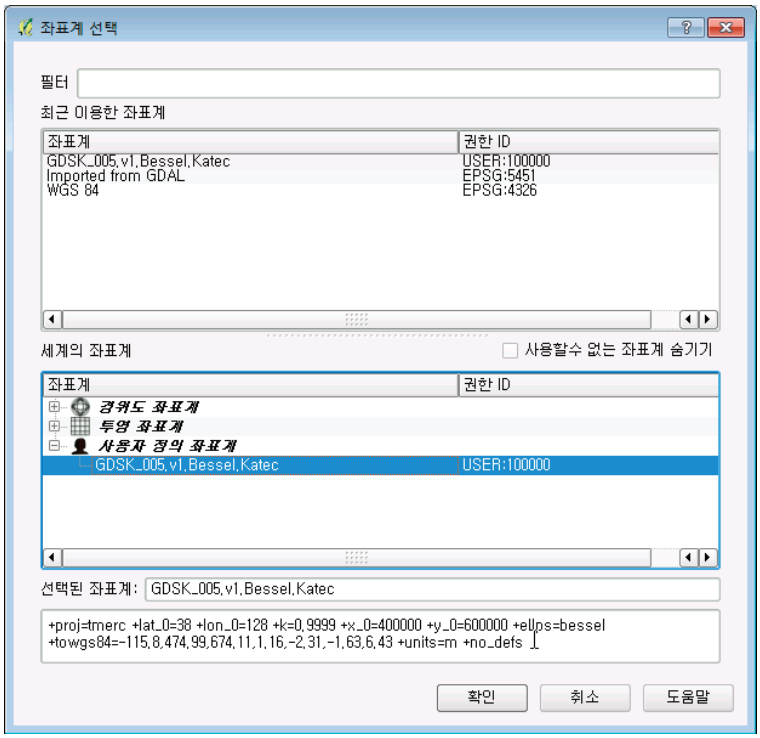


그림 10.30 좌표계 선택

10.3 범용소프트웨어(Excel add-in 프로그램)[59)]

10.3.1 Power Pivot

1. 파워 피벗 추가기능 설치하기

파워 피벗 추가 기능은 엑셀 2010부터 사용할 수 있었지만 별도의 설치 과정을 거쳐야 했다. 하지만 엑셀 2013부터는 파워 피벗 추가 기능이 내장되어 있어 별도로 설치할 필요 없이 편리하게 사용할 수 있다. 파워 피벗 추가 기능을 효율적으로 사용할 수 있도록 활성화하는 작업에 대해서 알아보자.

파워 피벗 추가 기능의 장점
피벗 테이블에 관계 기능이 추가되면서 파워 피벗 추가 기능을 사용해야 하는 이유가 줄었다. 하지만 워크시트의 데이터 관리 능력(104만 행)을 넘어서는 빅 데이터(Big Data)를 분석하려면 여전히 파워 피벗 추가 기능이 필하다. • 관계를 좀 더 직관적으로 설정할 수 있다. • 총 1억 개의 데이터를 관리하고 분석할 수 있다. • 2G 크기의 파일과 4G 메모리를 사용할 수 있다.

파워 피벗 추가 기능을 사용할 때 리본 메뉴의 [개발 도구] 탭이 표시되어 있으면 편리하다. [개발 도구] 탭을 표시하는 방법은 리본메뉴의 [파일] – [옵션] – [리본 사용자 지정] – ☑개발도구 순서대로 클릭하여 설정해주면 된다.

(1) 리본 메뉴의 [개발 도구] 탭 - [추가 기능] 그룹 - [COM 추가 기능]을 클릭한다. [COM 추가 기능] 대화상자가 표시되면[Microsoft Office PowerPivot for Excel 2013]에 체크표시하고 [확인]을 클릭한다.

59) 최준선, (2013), 엑셀 2013 바이블 회사에서 필요한 기능은 모두 있다의 내용을 옮겨오거나 발췌하여 정리한 것임



그림 10.31 Power Pivot 설치 1

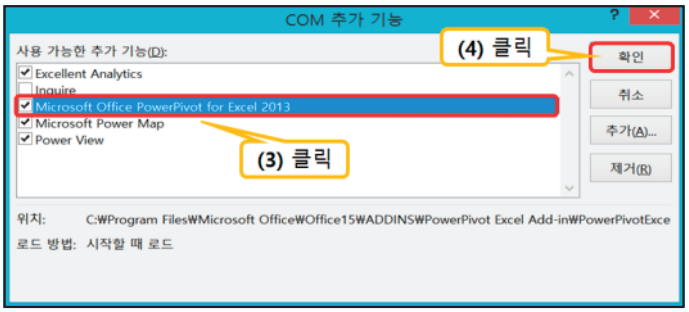


그림 10.32 Power Pivot 설치 2

(2) 리본 메뉴 오른쪽에 [PowerPivot] 탭이 생성된다.



그림 10.33 Power Pivot 설치 3

2. 파워 피벗 기능을 활용해 액세스 데이터를 피벗으로 분석하기

※ 예제파일 – [dbSales.accdb]

파워 피벗을 이용하면 SQL server, 액세스, 엑셀, 텍스트 파일 등 다양한 데이터 형식을 피벗 테이블 보고서로 분석할 수 있다. 또한 파워 피벗은 기존의 피벗 테이블 보고서와 달리 다양한 형태로 피벗 테이블 보고서를 만들 수 있다. 파워 피벗을 이용해 액세스 데이터베이스 데이터를 피벗으로 분석하는 방법에 대해서 알아보자.

(1) 빈 엑셀 파일을 하나 연다. 리본 메뉴의 [POWERP1VOT] 탭 - [데이터 모델] 그룹 - [관리]를 클릭 한다.



그림 10.34 Power Pivot으로 분석 1

(2) 파워 피벗 창이 열린다. 파워 피벗은 엑셀 워크시트의 104만 행의 제약에서 벗어나기 위해 별도의 데이터 관리 창을 사용한다. 예제 파일인 'dbSales.accdb' 데이터를 가져와 분석해보자. 리본 메뉴의 [홈] 탭 - [외부 데이터 가져오기] 그룹 - [데이터베이스] - [Access에서]를 선택한다.

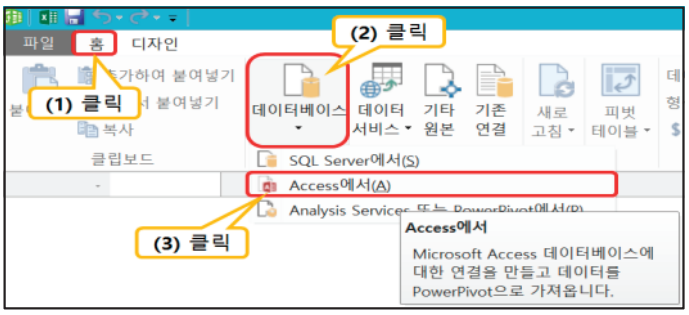


그림 10.35 Power Pivot으로 분석 2

(3) [테이블 가져오기 마법사] 대화상자가 나타나면 [찾아보기]를 클릭한다. 예

제파일이 들어있는 폴더로 이동한 후 'dbSales.accdb' 파일을 선택하고 [열기]를 클릭한다.

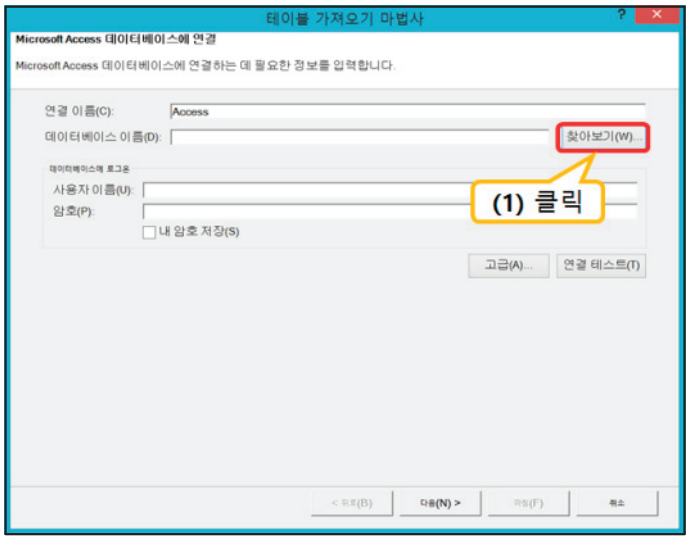


그림 10.36 Power Pivot으로 분석 3

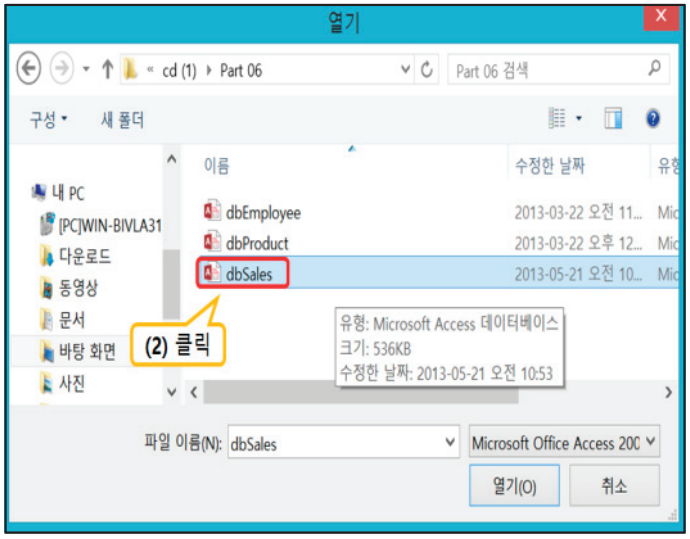


그림 10.37 Power Pivot으로 분석 4

(4) [데이터베이스 이름]에 선택한 경로가 표시되며, 확인 후 [다음]을 클릭한다. 여기에서는 데이터를 가져올 방법을 선택한다. [테이블 및 뷰 목록에서 가져올 데이터 선택]을 클릭하고 [다음]을 클릭한다.

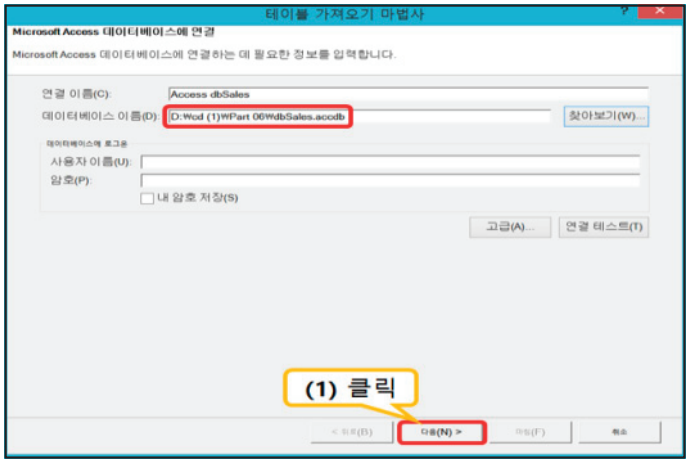


그림 10.38 Power Pivot으로 분석 5

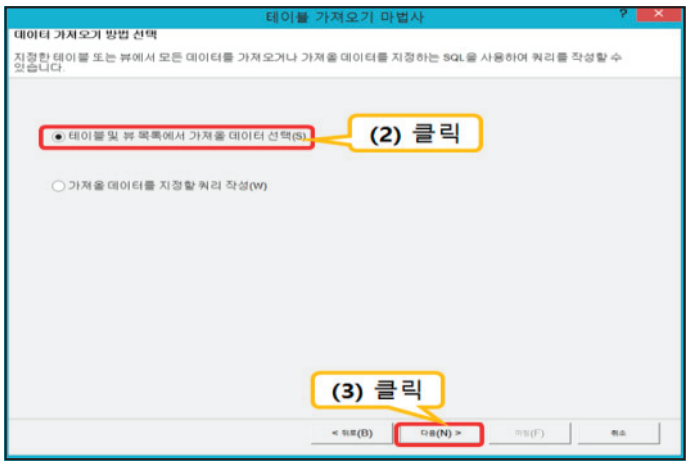


그림 10.39 Power Pivot으로 분석 6

뷰(View)란 무엇인가

뷰(View)는 데이터베이스의 가상 테이블로 액세스에서 쿼리(Query)라고 부른다. 뷰(View) 표현은 SQL server에서 쿼리(Query)를 부를 때 사용하는 명칭이므로 쿼리와 같은 명칭 정도로 이해한다.

(5) dbsales 데이터베이스에 있는 테이블 및 쿼리 목록이 표시된다. [판매대장] 테이블에 체크 표시하고 [마침]을 클릭한다. 데이터 연결에 성공했다는 메시지가 나타나면 [닫기]를 클릭해 창을 닫는다.

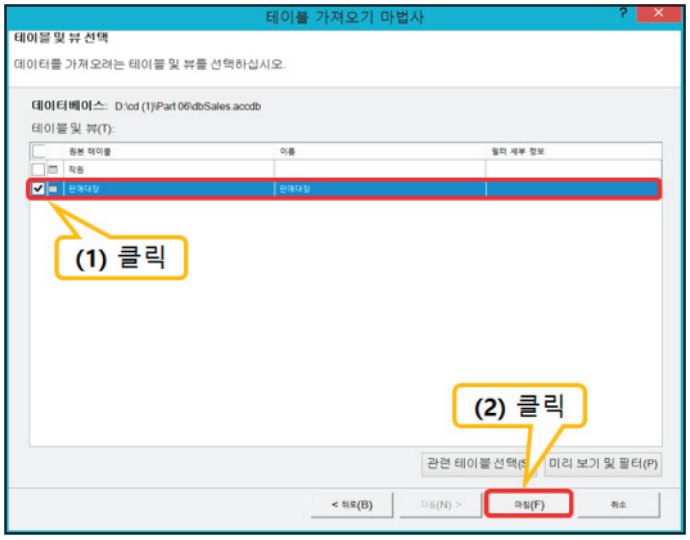


그림 10.40 Power Pivot으로 분석 7



그림 10.41 Power Pivot으로 분석 8

(6) [파워 피벗] 창에 dbSales 데이터베이스의 [판매대장] 테이블의 데이터가 가져와 표시된다. 피벗 테이블 보고서를 만들기 위해 리본 메뉴의 [홈] 탭 - [피벗 테이블] 콤보 상자의 [▽더 보기]를 클릭한 후 [피벗 테이블]을 선택한다.

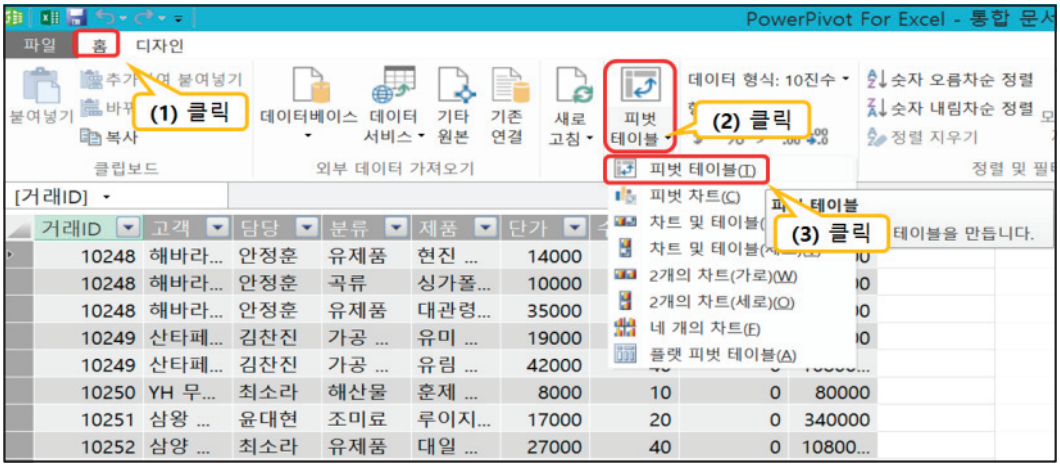


그림 10.42 Power Pivot으로 분석 9

Plus) 파워 피벗에서 만들 수 있는 피벗 테이블 종류

파워 피벗에서 만들 수 있는 피벗 테이블의 종류는 다음 설명을 참고한다.

- 피벗 테이블 : 엑셀의 피벗 테이블과 동일한 형식으로 피벗 테이블 보고서를 구성한다.
- 피벗 차트 : 피벗 차트만 구성한다.
- 차트 및 테이블(가로) : 피벗 테이블과 피벗 차트를 구성한다. 피벗 차트가 왼쪽에 표시되고 피벗 테이블이 오른쪽에 표시되도록 구성한다.
- 차트 및 테이블(세로) : 피벗 차트가 상단에 피벗 테이블이 하단에 표시되도록 구성한다.
- 두 개의 차트(가로) : 두 개의 피벗 차트를 왼쪽과 오른쪽에 각각 구성한다.
- 두 개의 차트(세로) : 두 개의 피벗 차트를 상단과 하단에 각각 구성한다.
- 네 개의 차트 : 네 개의 차트를 2열 2행으로 구성한다.
- 플랫 피벗 테이블 : 엑셀 2003 방식의 피벗 테이블 보고서를 구성한다.

(7) [피벗 삽입] 대화상자가 표시되면 [새 워크시트] 옵션을 선택하고 [확인]을 클릭한다.

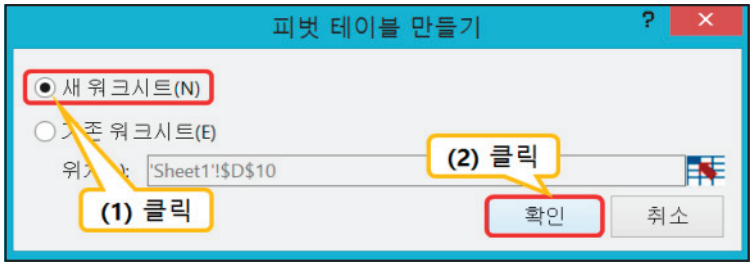


그림 10.43 Power Pivot으로 분석 10

(8) 피벗 테이블 보고서 레이아웃이 나타난다. 엑셀의 피벗 테이블 보고서 구성과 동일하다.

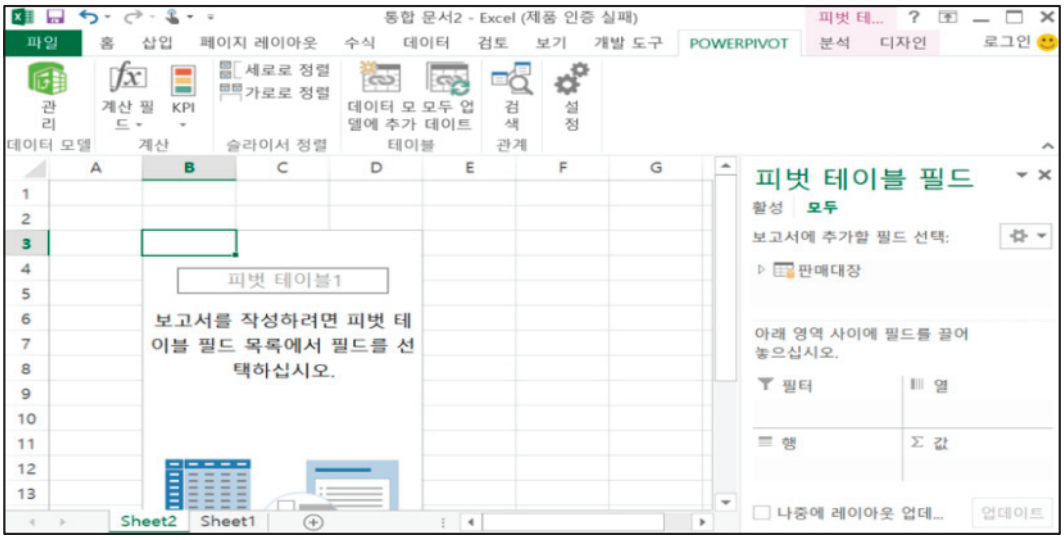


그림 10.44 Power Pivot으로 분석 11

(9) [피벗 테이블 필드] 작업 창에서 [담당] 필드를 [행] 영역에, [수량]과 [판매] 필드를 [값] 영역에 각각 추가한다. 다음과 같은 피벗 테이블 보고서를 완성한다.

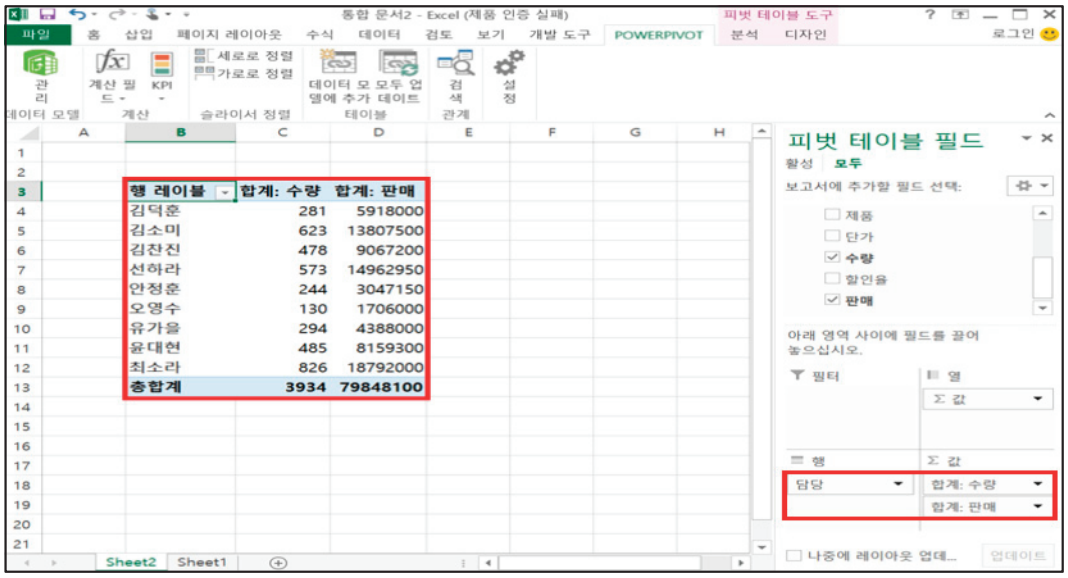


그림 10.45 Power Pivot으로 분석 12

(10) 슬라이서 창을 하나 삽입한다. 리본 메뉴의 [분석] 탭 - [필터] 그룹 - [슬라이서 삽입]을 클릭한다. [분류] 필드의 확인란에 체크 표시하고 [확인]을 클릭하면 다음과 같은 슬라이서 창이 삽입된다.

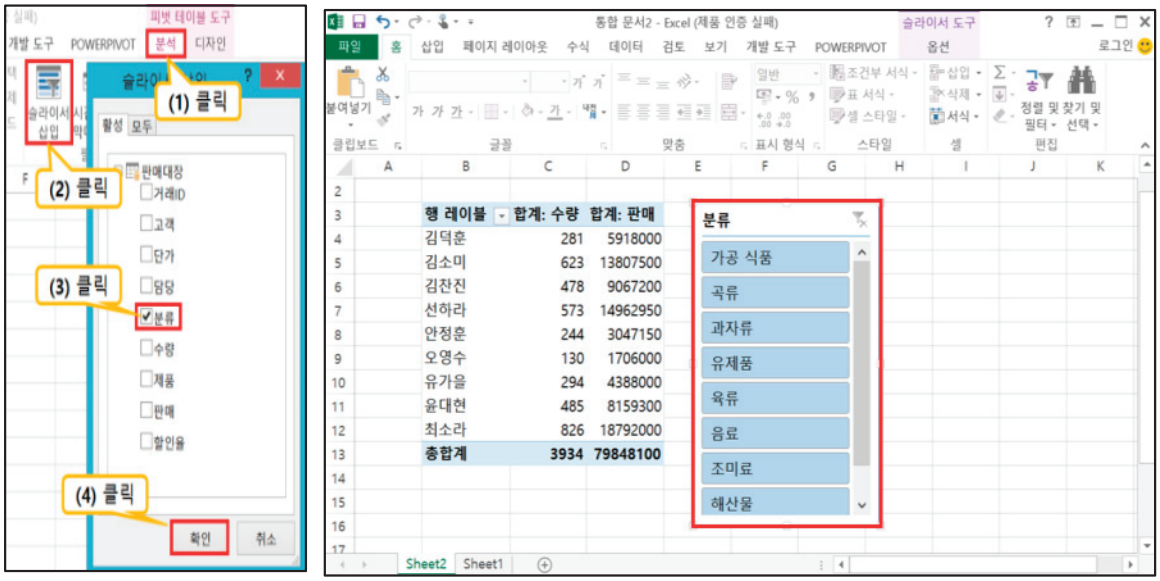


그림 10.39 Power Pivot으로 분석 13

(11) 리본 메뉴에서 슬라이서를 제어할 수 있다. 리본 메뉴의 [POWERPIVOT] 탭 - [슬라이서 정렬] 그룹 - [세로로 정렬]을 클릭한다. 슬라이서 창이 피벗 테이블 보고서의 왼쪽에 맞춰 정리된다.

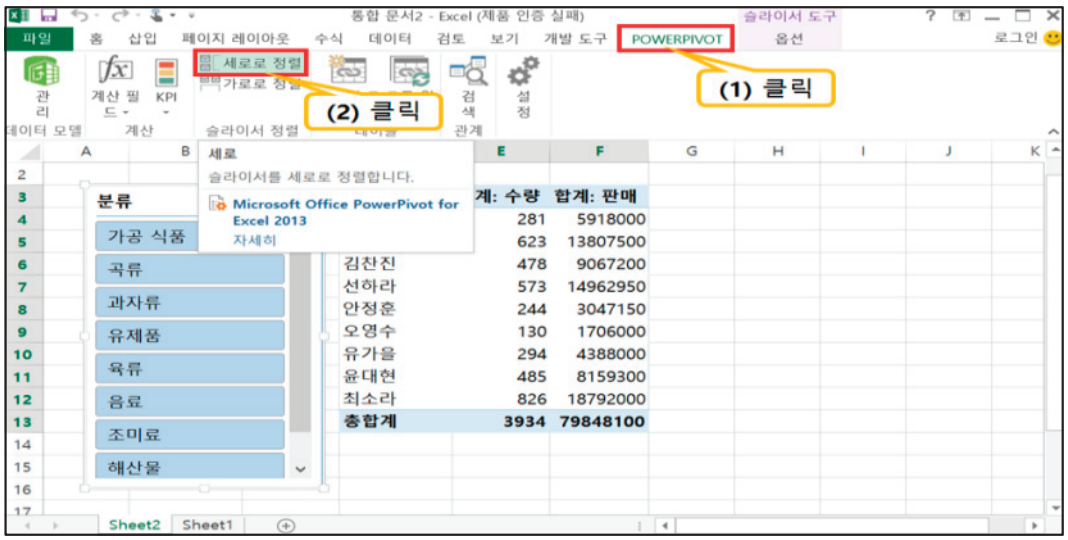


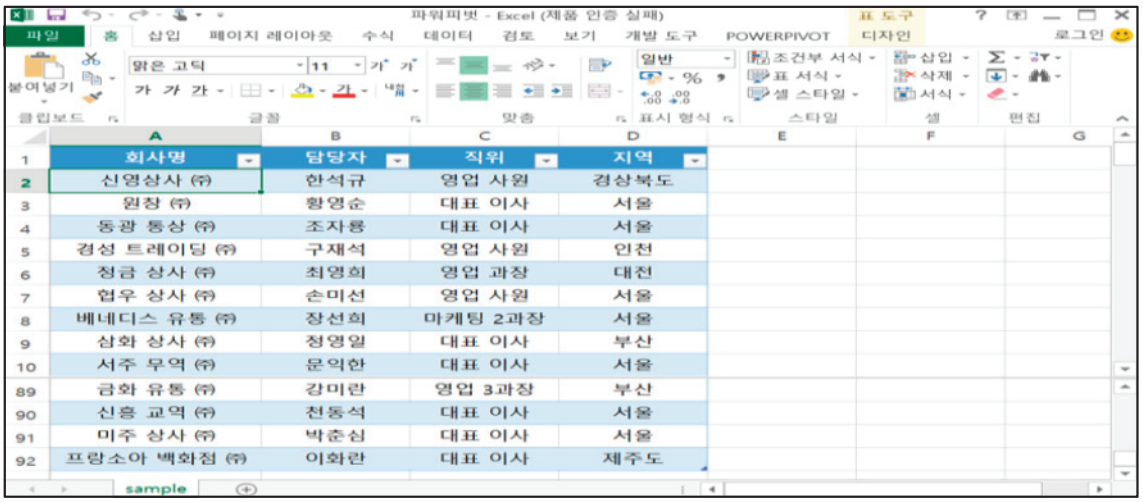
그림 10.40 Power Pivot으로 분석 14

3. 액세스와 엑셀 데이터 연동해 피벗으로 분석하기

※ 예제파일 – [파워피벗.xlsx, dbSales.accdb]

파워 피벗의 장점은 서로 다른 데이터 형식을 연결해 데이터를 분석할 수 있다는 점이다. 엑셀 2013부터는 관계 기능을 이용해 피벗 테이블 만으로 이 작업을 수행할 수 있다. 엑셀 워크시트에 담을 수 없는 대량의 데이터를 다룰 수 있다는 것은 실무에서 유용하게 활용 할 수 있는 부분이므로, 이 내용을 잘 이해해두는 것이 좋다.

(1) 예제 '파워피벗.xlsx' 파일을 열면 다음과 같은 엑셀 표를 확인할 수 있다. 이 표에는 거래하고 있는 회사의 담당자 및 지역 데이터가 입력되어 있다. 이 표와 'dbSales.accdb' 데이터베이스 의 [판매대장] 테이블을 연결해 지역별 매출 실적 보고서를 파워 피벗으로 만들어 보자.



	A	B	C	D
1	회사명	담당자	직위	지역
2	신영상사 ㈜	한석규	영업 사원	경상북도
3	원창 ㈜	황영순	대표 이사	서울
4	동광 통상 ㈜	조자룡	대표 이사	서울
5	경성 트레이딩 ㈜	구재석	영업 사원	인천
6	청금 상사 ㈜	최영희	영업 과장	대전
7	협우 상사 ㈜	손미선	영업 사원	서울
8	베네디스 유통 ㈜	장선희	마케팅 2과장	서울
9	삼화 상사 ㈜	정영일	대표 이사	부산
10	서주 무역 ㈜	문익한	대표 이사	서울
89	금화 유통 ㈜	강미란	영업 3과장	부산
90	신흥 교역 ㈜	천동석	대표 이사	서울
91	미주 상사 ㈜	박준심	대표 이사	서울
92	프랑소아 백화점 ㈜	이화란	대표 이사	제주도

그림 10.41 Power Pivot으로 분석 15

엑셀 표의 '이름'과 dbSales.accdb 데이터베이스

외부 데이터와 연결해 작업하려면 엑셀에서는 표를 '엑셀 표'로 변환해야 한다. 이때 엑셀 표의 이름이 중요하므로, 본인이 직접 변환해놓은 표가 아니라면 리본 메뉴의 [디자인] 탭 - [속성] 그룹 - [표 이름]에서 엑셀 표 이름을 확인해야 한다.

(2) 엑셀 표 데이터를 파워 피벗 창으로 전송한다. 리본 메뉴의 [POWERPIVOT] 탭 - [테이블] 그룹 - [데이터 모델에 추가]를 클릭한다.

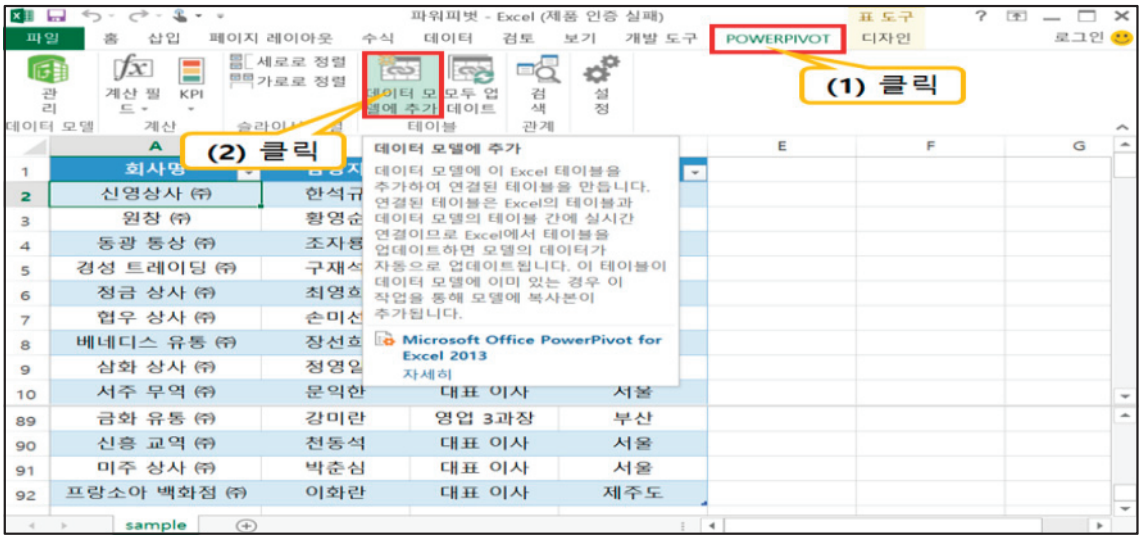


그림 10.42 Power Pivot으로 분석 16

(3) 자동으로 [파워 피벗] 창이 열리면서 전송된 결과를 확인할 수 있다.

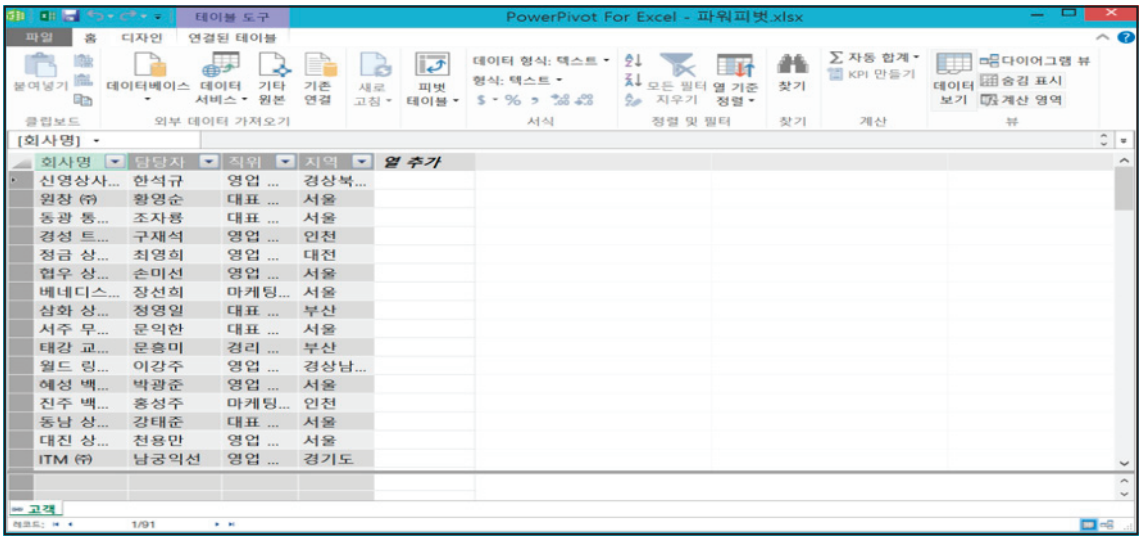


그림 10.43 Power Pivot으로 분석 17

(4) 'dbSales.accdb' 데이터베이스의 [판매대장] 테이블을 추가로 가져온다. [파워 피벗] 창의 리본 메뉴에서 [홈] 탭 - [외부 데이터 가져오기] 그룹 - [데이터베이스] - [Access에서]를 선택한다.

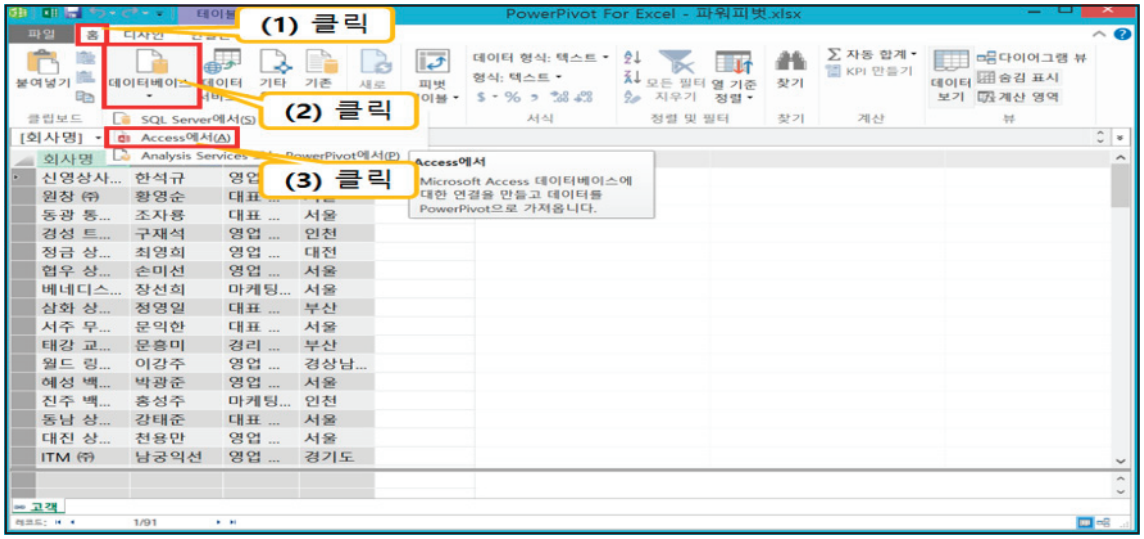


그림 10.44 Power Pivot으로 분석 18

(5) '파워 피벗 기능을 활용해 액세스 데이터를 피벗으로 분석하기'의 03. ~ 08. 과정을 참고해 [판매대장] 테이블을 가져온다. 그러면 [파워 피벗] 창 하단에서 [고객] 탭과 [판매대장] 탭을 확인할 수 있다.

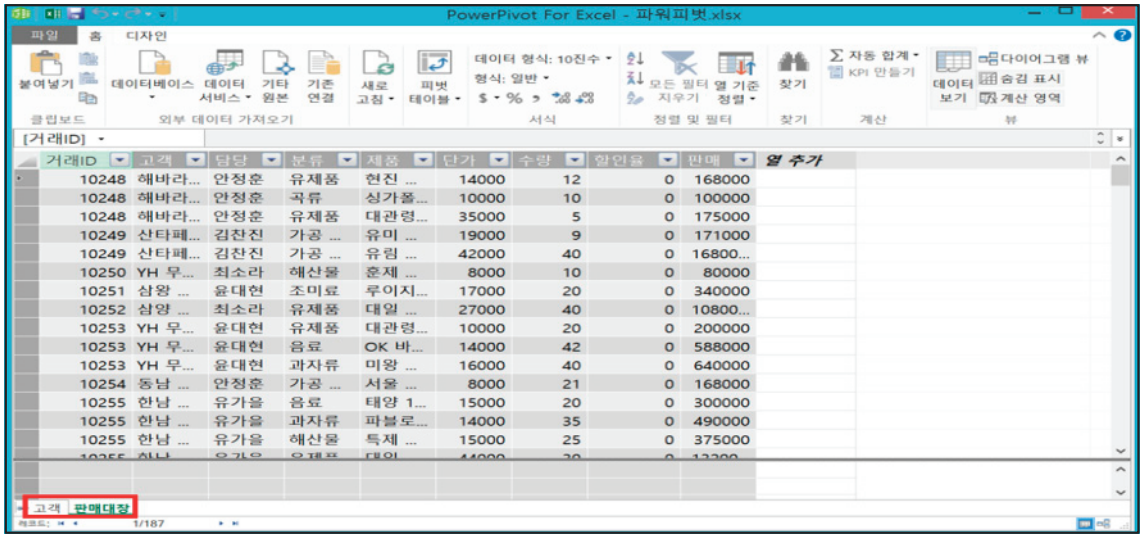


그림 10.45 Power Pivot으로 분석 19

[파워 피벗] 창에서 가져온 데이터 확인하기

[파워 피벗] 창의 하단 탭을 선택하면 가져온 데이터를 확인할 수 있다. [고객] 표는 엑셀 워크시트의 [고객] 엑셀 표와 연결된 표이고, [판매대장] 표는 'dbSales.accdb 데이터베이스에서 [판매대장] 테이블의 데이터를 가져온 것 이다.

(6) [파워 피벗] 창으로 가져온 두 표의 관계를 설정해 하나의 표처럼 사용한다. [파워 피벗] 창에서 리본 메뉴의 [디자인] 탭 - [관계] 그룹 - [관계 만들기]를 클릭한다.

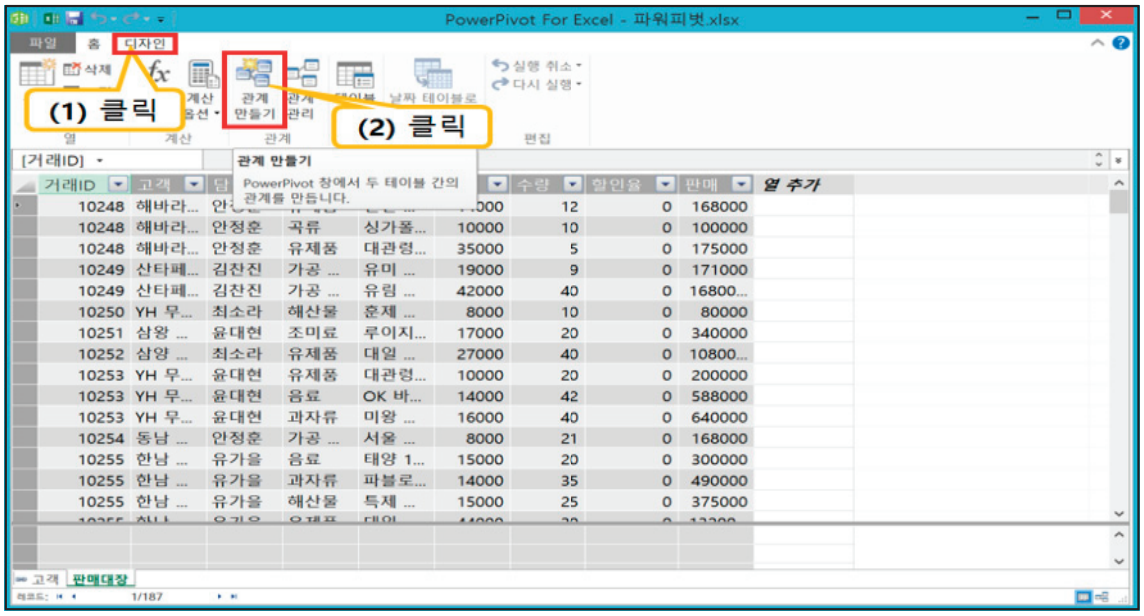


그림 10.46 Power Pivot으로 분석 20

(7) [관계 만들기] 대화상자가 나타난다. 다음과 같이 설정하고 [만들기]를 클릭한다.

테이블 : 판매대장
열 : 고객
관련 조회 테이블 : 고객
관련 조회 열 : 회사명



그림 10.30 Power Pivot으로 분석 21

[관계 만들기] 대화상자 설정 이해하기

엑셀에서 '관계'를 설정하는 방법과 '파워 피벗'의 관계를 설정하는 방법은 동일하다. 따라서 표의 주종 관계를 잘 이해해야 한다. 하나의 고객(기업)에서 여러 건의 거래를 할 수 있으므로, [고객] 표가 주이고 [판매대장] 표가 종이다. 판매대장 테이블의 필드가 먼저 선택되어야 하며, 고객 표에서 연결한 열을 나중에 설정해야 한다. 반대로 정보를 입력해도 자동으로 위치가 변경되어 관계가 만들어진다.

(8) 관계를 설정했으면 피벗 테이블 보고서를 생성한다. [파워 피벗] 창에서 리본 메뉴의 [홈] 탭 - [피벗 테이블]을 클릭한다.

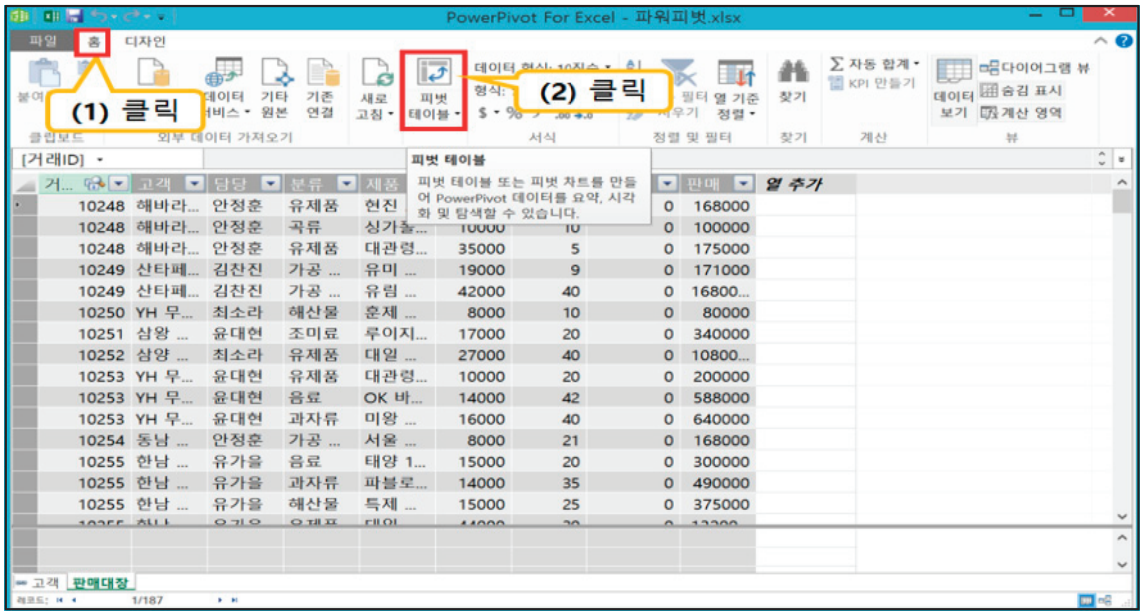


그림 10.48 Power Pivot으로 분석 22

(9) [피벗 삽입] 대화상자가 나타나면 [새 워크시트] 옵션이 선택된 상태에서 [확인]을 클릭한다.

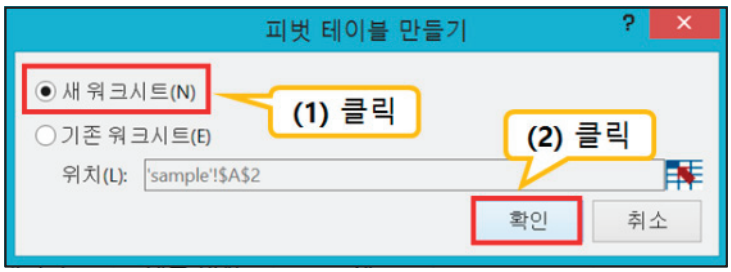


그림 10.49 Power Pivot으로 분석 23

(10) 피벗 테이블 보고서가 나타난다. [피벗 테이블 필드] 작업 창에서 [고객]과 [판매대장] 두 개의 표를 확인할 수 있다.

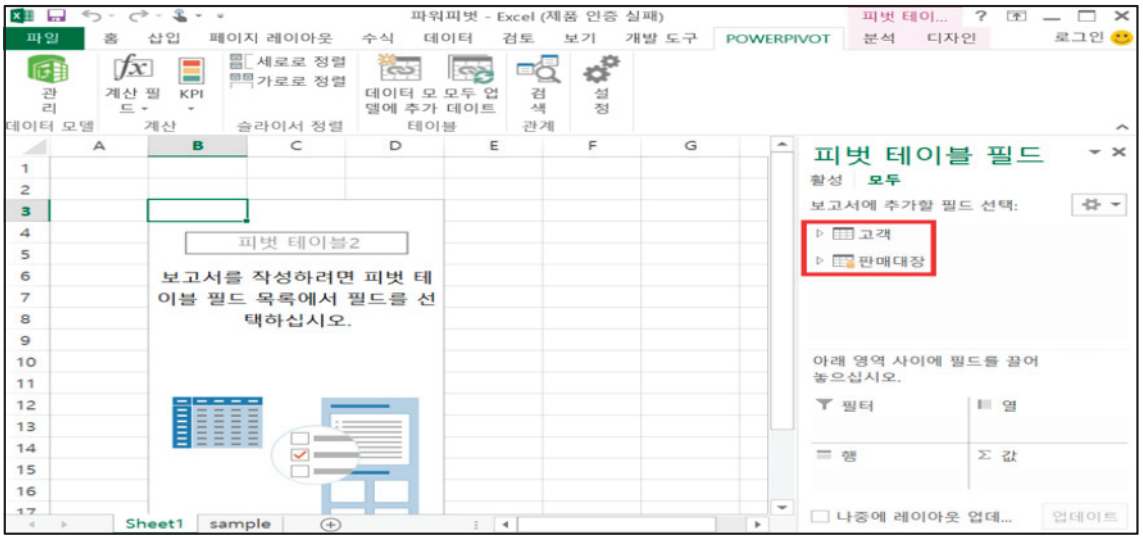


그림 10.50 Power Pivot으로 분석 24

(11) [고객] 표의 [지역] 필드를 [행] 영역에 삽입하고, [판매대장] 테이블의 [판매] 필드를 [값] 영역에 삽입한다. 그러 면 다음과 같은 지역별 매출 실적 보고서를 만들 수 있다.

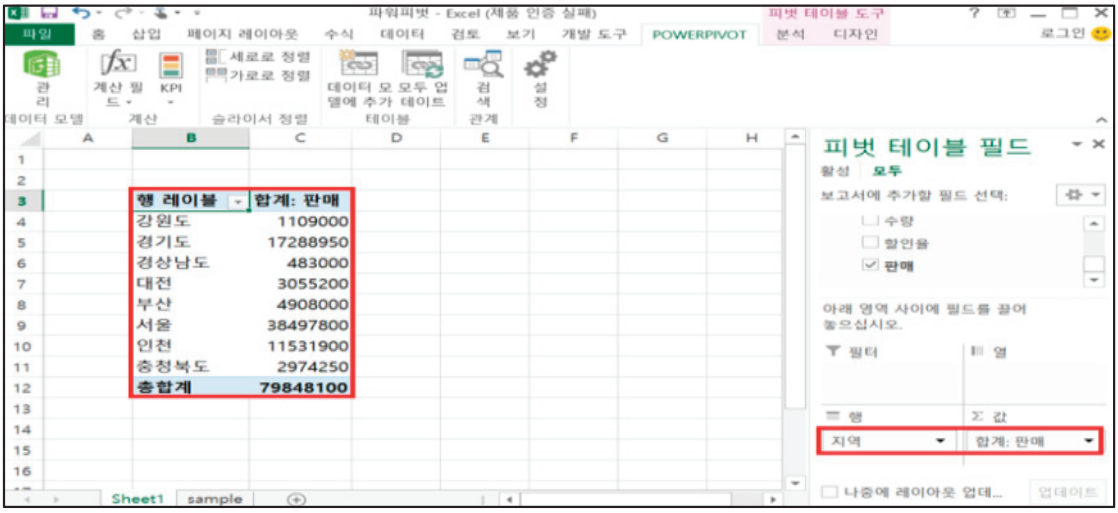


그림 10.51 Power Pivot으로 분석 25

10.3.2 MS Power View

1. 파워뷰 추가기능 설치

엑셀 2013에서 새로 추가된 파워뷰 기능은 시각적인 보고서를 빠르고 효율적으로 구성할 수 있도록 도와준다. 피벗 테이블 보고서와 유사한 점이 많은데 파워뷰는 추가 기능으로 우선 설치한 후 사용할 수 있다. 사용자 환경에 따라 마이크로소프트사에서 배포하는 silverlight를 먼저 설치해야 하는 경우도 있다.

(1) 엑셀을 실행하고 빈 파일을 하나 연다. 리본 메뉴의 [삽입] 탭 - [보고서] 그룹 - [power view]를 클릭한다. 다음과 같은 안내 메시지 창이 나타나는데 메시지 내용을 잘 읽어보고 [사용]을 클릭한다.

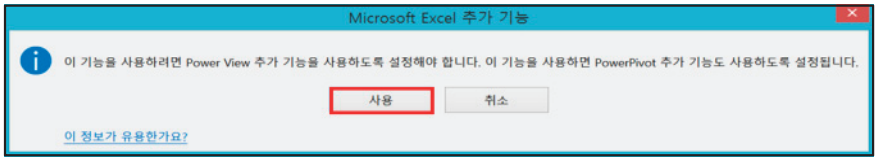


그림 10.52 Power View 설치 1

(2) sliverlight가 설치되지 않은 PC에서는 다음의 경고 메시지 줄이 표시된다. [sliverlight설치] 하이퍼링크를 클릭해 설치한다. 자동으로 다운로드 과정이 진행되며 끝나면 화면과 같은 창이 나타난다. [지금 설치]를 클릭해 설치한다.



그림 10.53 Power View 설치 2



그림 10.54 Power View 설치 3

(3) 설치가 완료되면 엑셀 화면으로 돌아오는데 02. 과정의 경고 메시지 줄이 계속 표시된다. 경고 메시지 줄의 [다시 로드]를 클릭하면 경고 메시지 줄이 사라지고 시트 탭에는 [power view1]시트가, 리본 메뉴 상단에는 [POWER VIEW] 탭과 [POWERPIVOT] 탭이 추가된 것을 확인할 수 있다.

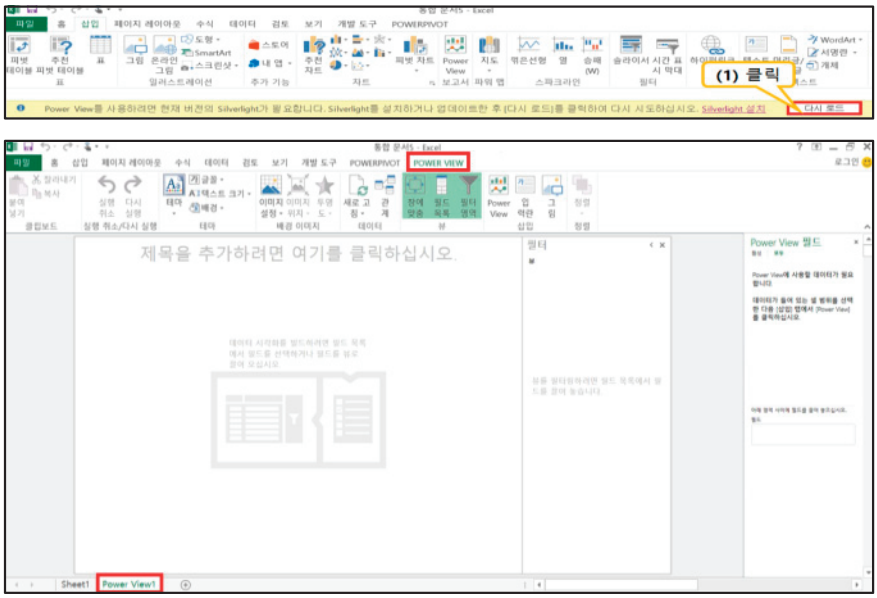


그림 10.55 Power View 설치 4

2. 데이터 시각화하기[60)]

※ 예제파일 – [Special_PowerView_예제.xlsx]

우선 Special_PowerView_예제.xlsx라는 엑셀 파일을 불러온다. 이 파일은 한 영업부서에서 작성한 영업자별, 매출액과 주문받은 날짜 등이 들어있는 데이터베이스이다.

(1) 불러온 예제 파일에서 데이터가 있는 부분 중 아무 셀이나 클릭한다. 그 상태에서 [삽입] 탭 - [보고서] 그룹 - [Power View]를 클릭한다. 이 메뉴가 안 보인다면 옵션을 수정해야 한다.

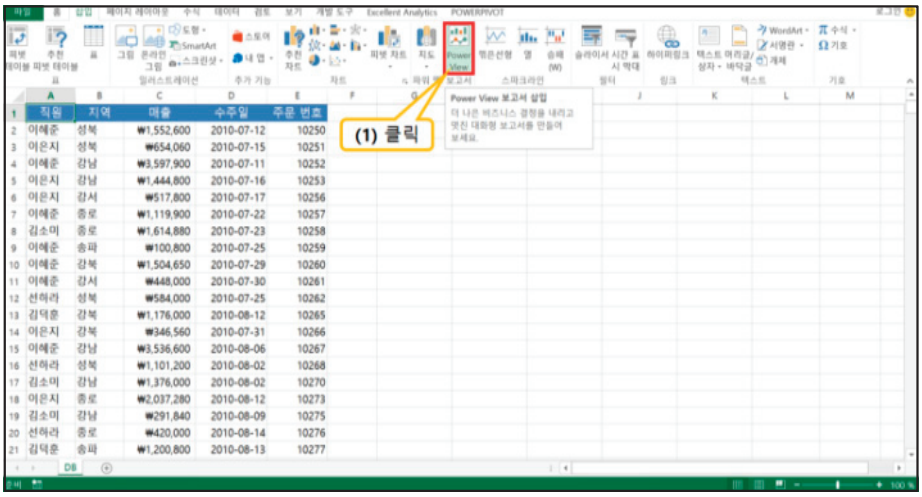
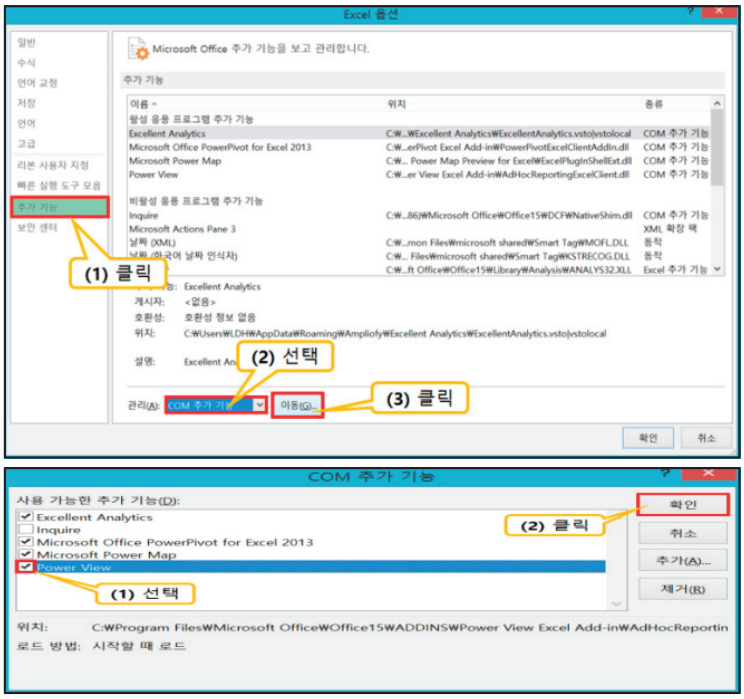


그림 10.56 Power View를 이용한 데이터 시각화 1

60) 김철, (이지스퍼블리싱), 된다! 엑셀 능력자 데이터 시대의 엑셀 기본서의 내용을 옮겨오거나 발췌하여 정리한 것임

Power Vew 기능을 사용하기 위한 [Excel 옵션] 수정

Power View 는 마이크로소프트의 silverLight 기술을 이용한 엑셀 2013의 고유 기능이므로 [COM 추가 기능]을 사용하도록 설정해야 한다. 리본 메뉴의 [파일] - [옵션] - [추가 기능]의 하단에 있는 [관리(A)]에서 [COM 추가 기능]을 선택하고 [이동] 버튼을 클릭한다.



나타난 [COM 추가 기능] 대화 상자에서 [Power View]를 선택하고 [확인] 버튼을 클릭한다.

(2) 하단에 새로운 시트가 생성된 것을 볼 수 있다. 우측의 [power view 필드] 목록에서 나타내고자 하는 필드만 클릭한다. 여기에서는 [매출] 필드와 [직원] 필드를 선택한다.

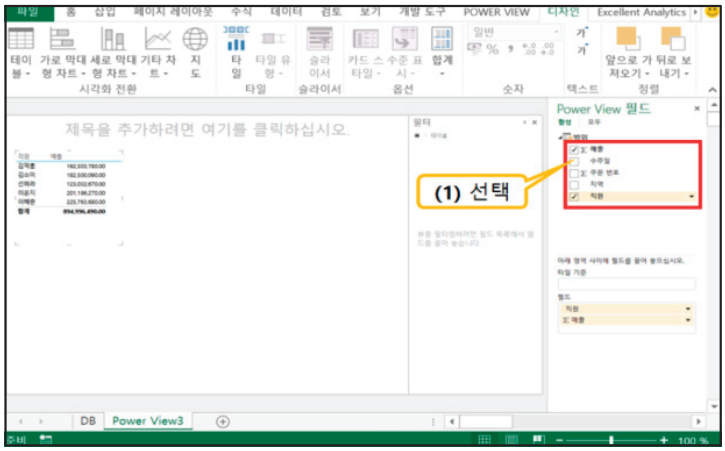


그림 10.57 Power View를 이용한 데이터 시각화 2

(3) 직원별 매출 금액을 차트로 작성하기 위해 [디자인] 탭 - [시각화 전환] 그룹 - [세로 막대형 차트]의 확장버튼을 클릭해서 [묶은 세로 막대형]을 선택한다.

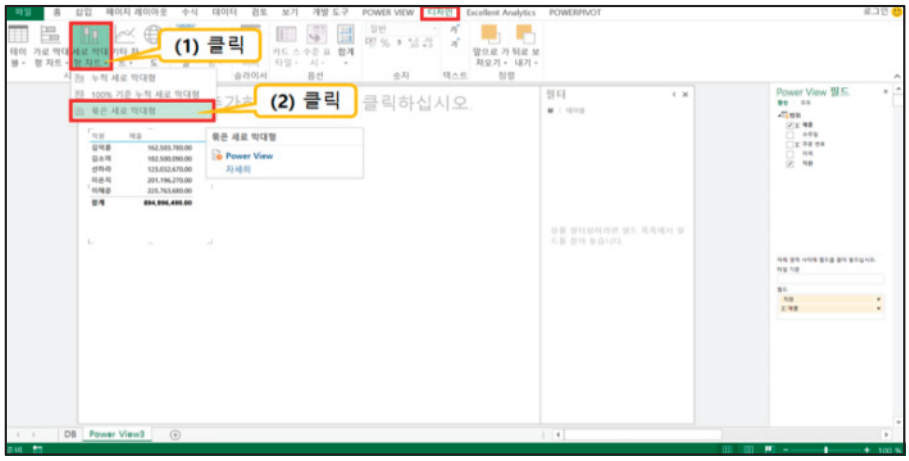


그림 10.58 Power View를 이용한 시각화 3

(4) 생성된 [묶은 세로 막대형] 차트에서 우측 하단의 조절점을 드래그하여 적당한 크기로 차트를 조절한다.

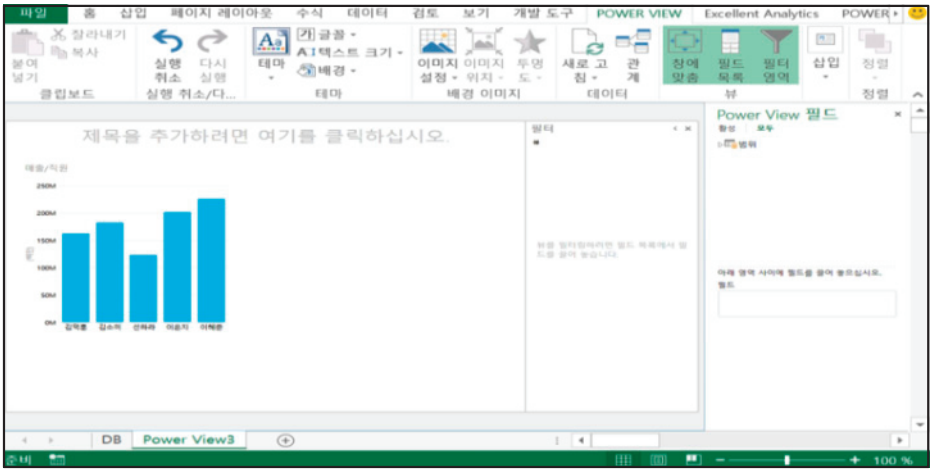


그림 10.59 Power View를 이용한 데이터 시각화 4

(5) [묶은 세로 막대형] 차트 우측은 지역별 매출을 나타내고자 한다. [묶은 세로 막대형] 우측의 빈 공간을 클릭하고 좌측의 [Power View 필드] 목록에서 [범위]를 확장해서 [지역], [매출]을 선택한다.

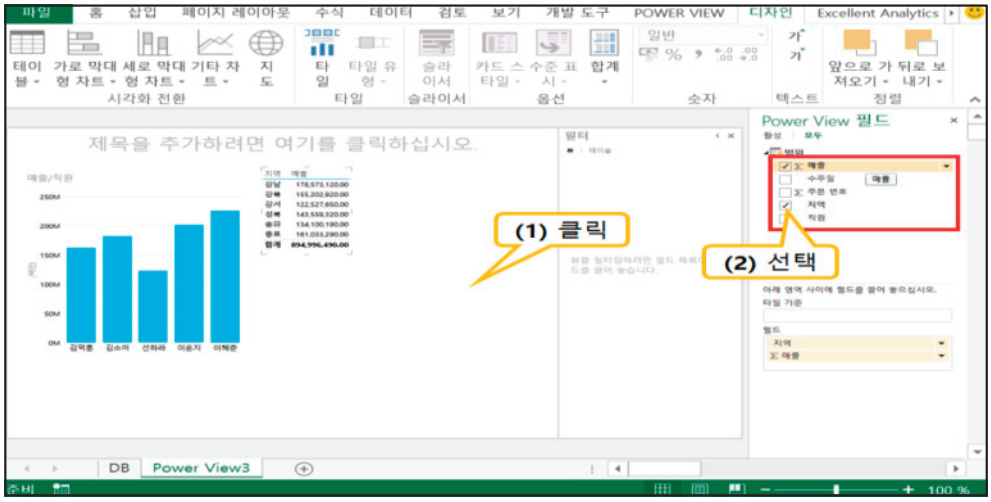


그림 10.60 Power View를 이용한 데이터 시각화 5

(6) [디자인] 탭 - [시각화 전환] 그룹 - [기타 차트]의 확장 버튼을 클릭해서 [원형]을 선택한다.

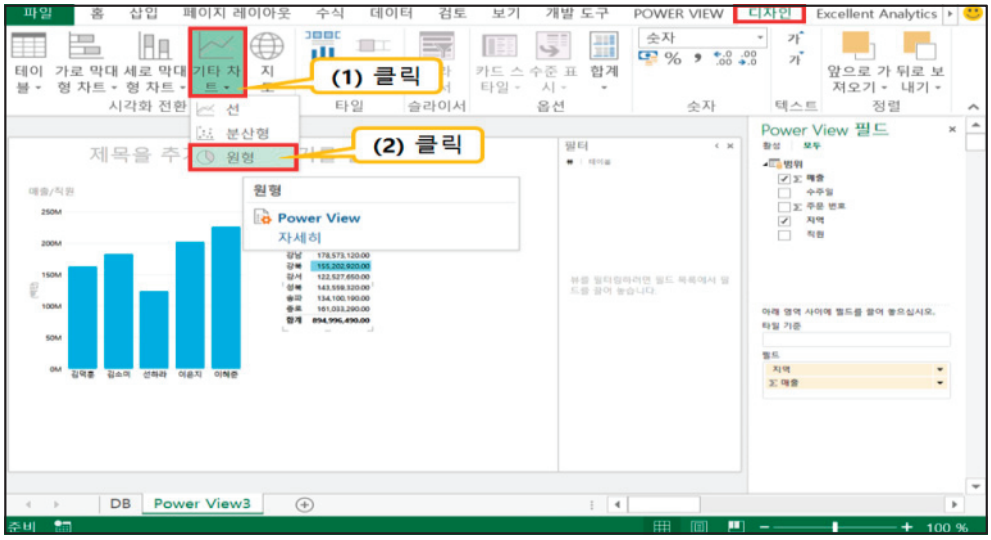


그림 10.61 Power View를 이용한 데이터 시각화 6

(7) 이번에는 만들어 놓은 차트의 제목을 입력하겠다. [제목을 추가하려면 이곳을 클릭하십시오.] 부분을 클릭해서 '개인, 지역별 매출현황'이라고 입력한다.

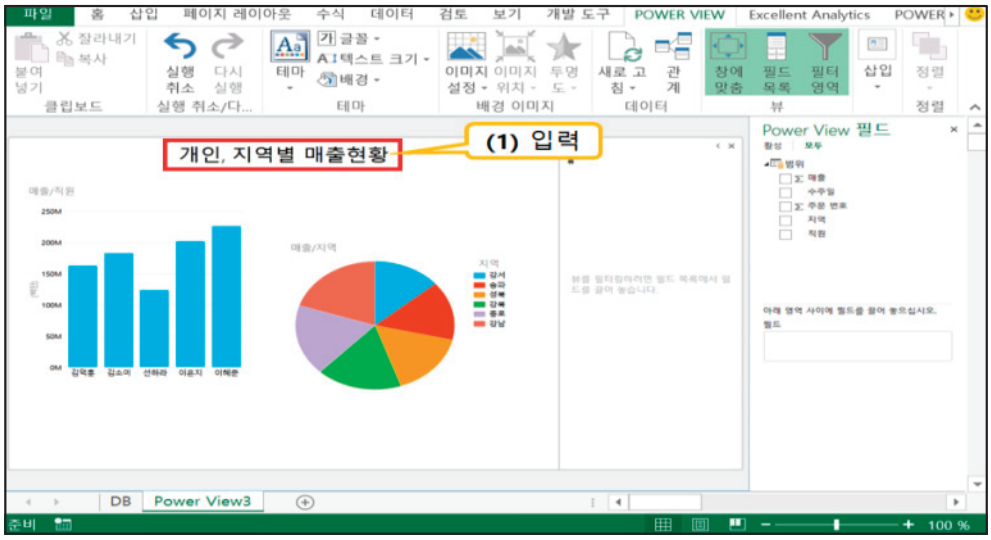


그림 10.62 Power View를 이용한 데이터 시각화 7

(8) 시각화한 차트를 크게 보기 위해서 [Power View 필드] 목록의 [필드 목록 닫기] 버튼(X)을 클릭해서 닫는다.

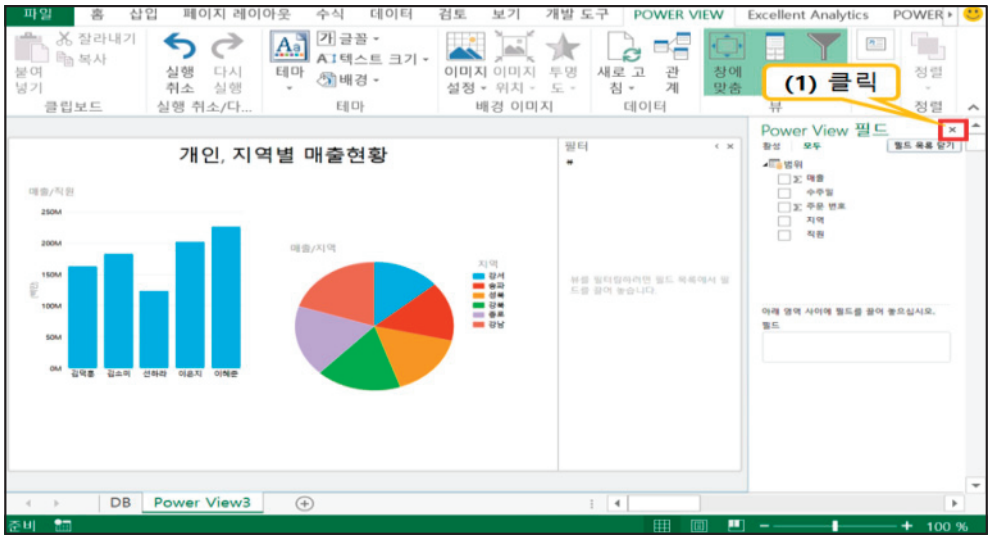


그림 10.63 Power View를 이용한 데이터 시각화 8

(9) [필터]의 [필터 영역 닫기] 버튼(X)을 클릭해서 [필터] 영역도 닫는다.

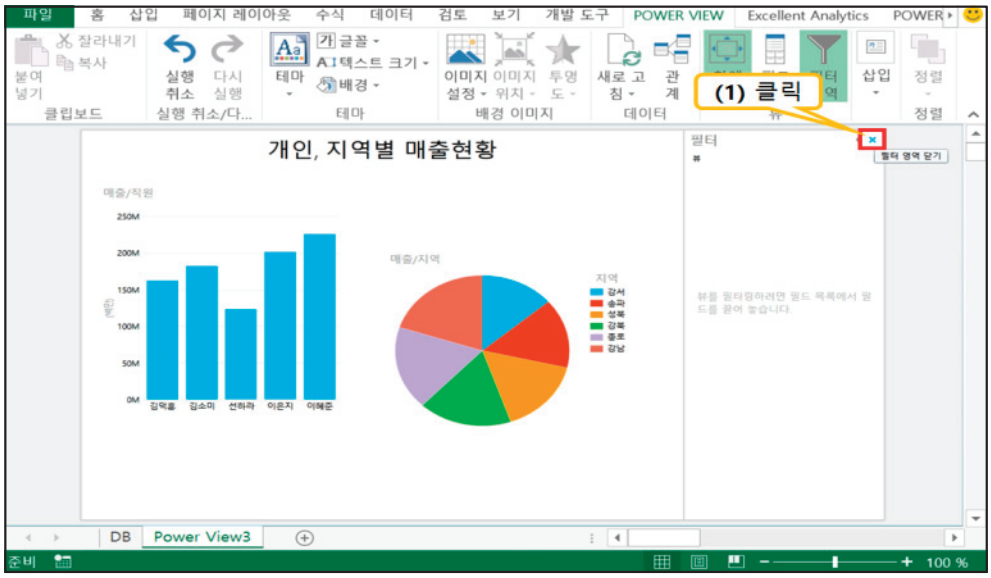


그림 10.64 Power View를 이용한 데이터 시각화 9

(10) 마지막으로 시각화된 차트를 이용해서 매출이 가장 많은 '이혜준'의 지역별 매출 분포를 확인해 보자. [묶은 세로 막대형] 차트의 '이혜준' 막대를 클릭한다.

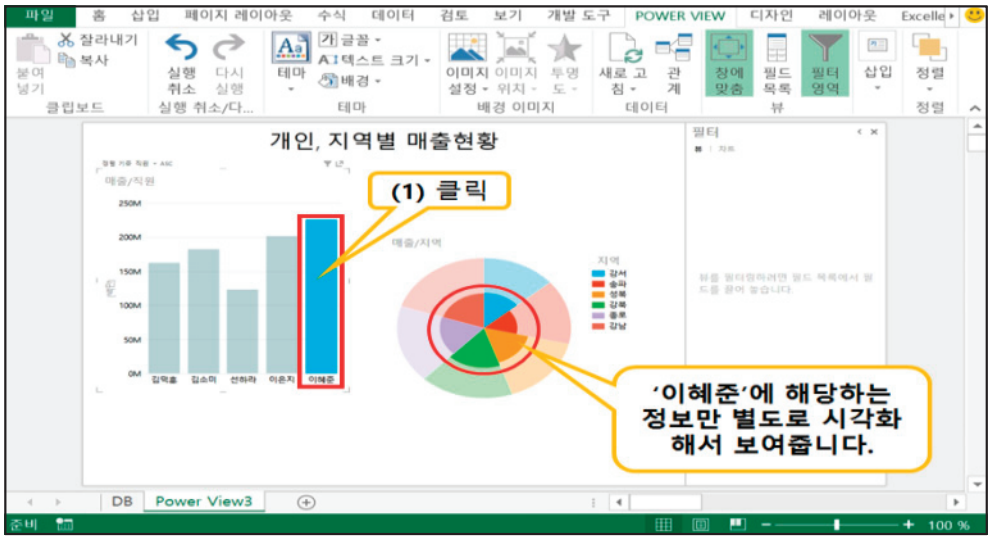


그림 10.65 Power View를 이용한 데이터 시각화 10

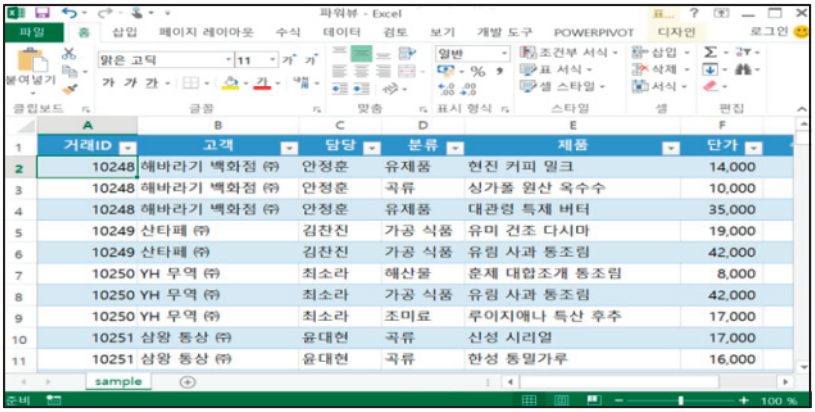
다른 사람의 매출 분포를 확인하려면 다른 막대를 클릭하면 되고, 원상태로 환원하려면 [묶은 세로 막대형] 차트의 [그림 영역]을 클릭하면 된다. 또 특정 지역의 개인별 매출 분포를 보려면 [원형 차트]에서 원하는 지역을 클릭해서 분포를 확인할 수 있다.

3. 시각적인 보고서 만들기

※ 예제파일 – [파워뷰.xlsx]

파워뷰 추가 기능을 사용하면 피벗 테이블 보고서를 구성하듯 시각적인 보고서를 구성할 수 있어 편리하다. 파워뷰 기능을 이용해 월간 보고서를 작성하면서 파워뷰 추가 기능을 이용하는 방법에 대해서 알아보자.

(1) 예제를 열면 화면과 같은 표를 확인할 수 있다. 이 표의 데이터가 1월 데이터라고 가정하고 월 실적을 집계하는 보고서를 파워뷰 기능을 이용해 만들어보자.



	A	B	C	D	E	F
1	거래ID	고객	담당	분류	제품	단가
2	10248	해바라기 백화점 ㈜	안정훈	유제품	현진 커피 밀크	14,000
3	10248	해바라기 백화점 ㈜	안정훈	곡류	싱가폴 원산 옥수수	10,000
4	10248	해바라기 백화점 ㈜	안정훈	유제품	대관령 특제 버터	35,000
5	10249	산타페 ㈜	김찬진	가공 식품	유미 건조 다시마	19,000
6	10249	산타페 ㈜	김찬진	가공 식품	유림 사과 통조림	42,000
7	10250	YH 무역 ㈜	최소라	해산물	훈제 대합조개 통조림	8,000
8	10250	YH 무역 ㈜	최소라	가공 식품	유림 사과 통조림	42,000
9	10250	YH 무역 ㈜	최소라	조미료	루이지애나 특산 후추	17,000
10	10251	삼왕 통상 ㈜	윤대현	곡류	신성 시리얼	17,000
11	10251	삼왕 통상 ㈜	윤대현	곡류	한성 통밀가루	16,000

그림 10.66 Power View를 이용한 시각적 보고서 만들기 1

(2) 엑셀 표에서 리본메뉴의 [삽입] 탭 - [보고서 그룹] - [Power view]를 클릭한다. 화면과 같은 [pover View1] 시트가 추가되면서 엑셀 표 데이터를 요약한 결과를 표시한다.

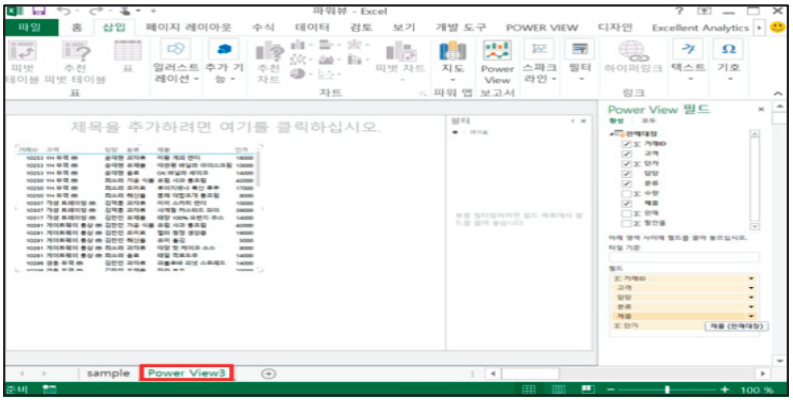


그림 10.67 Power View를 이용한 시각적 보고서 만들기 2

(3) 두 번째 과정에서 표시된 결과는 배제하고 영업사원별 실적을 요약하는 표를 집계한다. 오른쪽 [Power View 필드] 작업 창에서 [판매대장] 필드의 [담당], [수량], [판매], [할인율]에만 체크 표시하고 나머지는 모두 체크 표시를 해제한다.

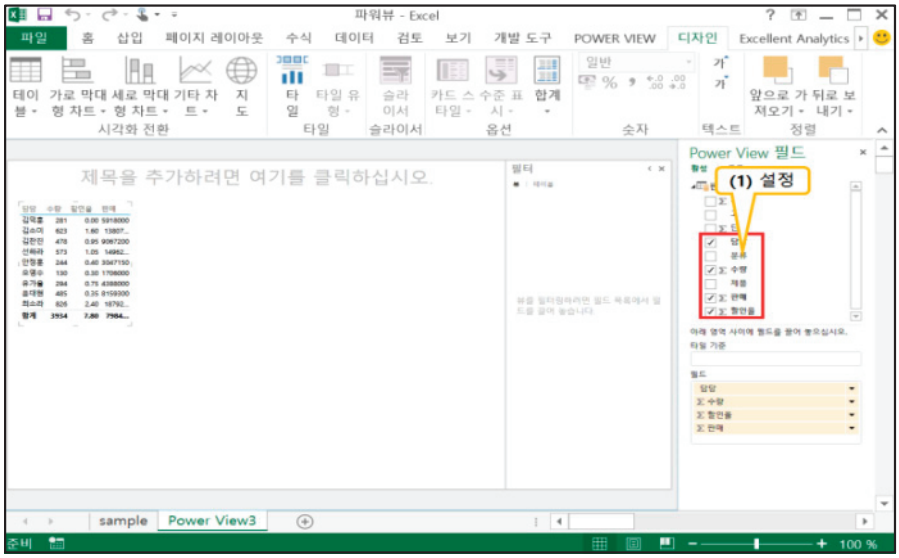


그림 10.68 Power View를 이용한 시각적 보고서 만들기 3

(4) 영업사원별 수량, 판매, 할인율의 합계가 구해진다. 할인율은 합계 대신 평균으로 계산한다. [Power View 필드]의 작업 창의 [필드] 리스트에서 [할인율] 필드를 클릭하면 메뉴가 표시되는데, 요약 방법을 선택할 수 있다. 여기서는 [평균]을 선택한다.

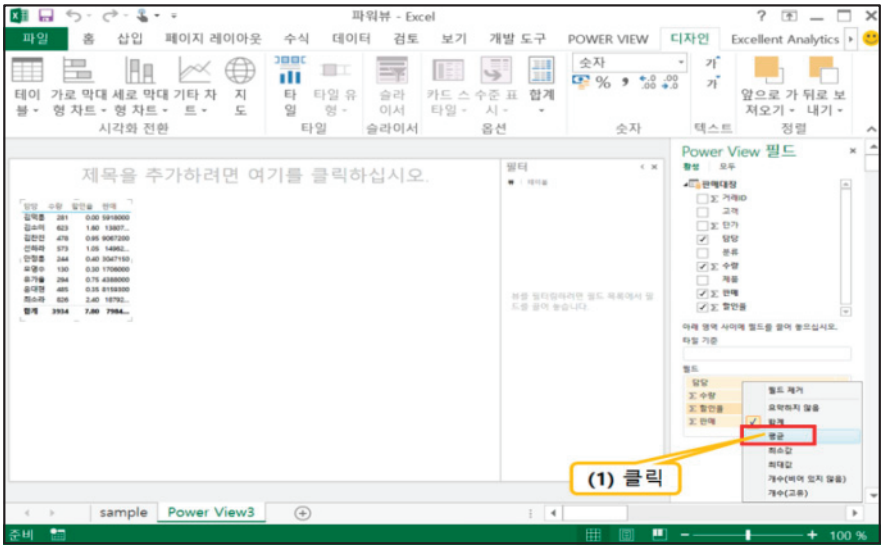


그림 10.69 Power View를 이용한 시각적 보고서 만들기 4

(5) 집계된 결과를 보기 좋게 구성하기 위해 천 단위 구분 기호와 백분율 기호를 추가한다. [Power view 1] 시트의 [수량] 필드 내 아무 값이나 선택하고 리본 메뉴의 [디자인] 탭 - [숫자] 그룹 - [쉼표 스타일]을 클릭한다. 같은 방법으로 [판매] 필드를 선택한다. [쉼표 스타일]을 클릭한 후 [할인율의 평균] 필드를 선택한다. [백분율 스타일]을 클릭한다.



그림 10.70 Power View를 이용한 시각적 보고서 만들기 5

(6) [판매] 필드의 열 너비가 부족해 전체 실적이 제대로 표시되지 않는다. [판매] 필드의 열 너비를 넓히기 위해 [판매]와 [할인율의 평균] 사이로 마우스 커서를 옮긴 후 오른쪽으로 드래그 한다.

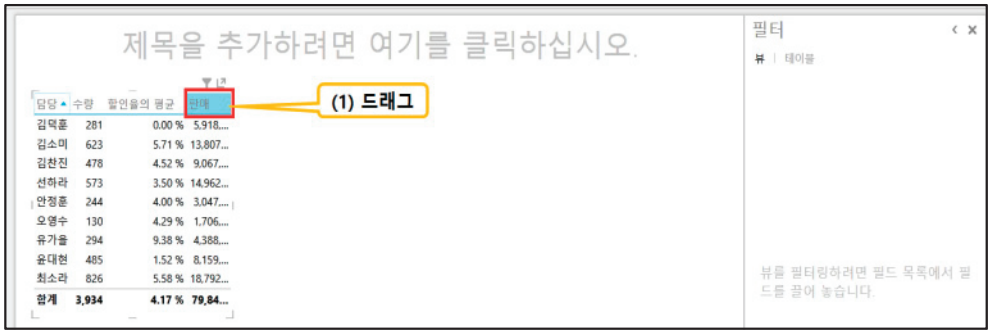


그림 10.71 Power View를 이용한 시각적 보고서 만들기 6

(7) 고객별 실적을 세로 막대형차트로 보고서에 추가한다. 새로운 집계 값을 추가해야 한다. [Power View] 시트의 빈 영역을 클릭하면 [Power View 필드] 작업 창의 필드가 모두 해제된 상태로 표시된다.

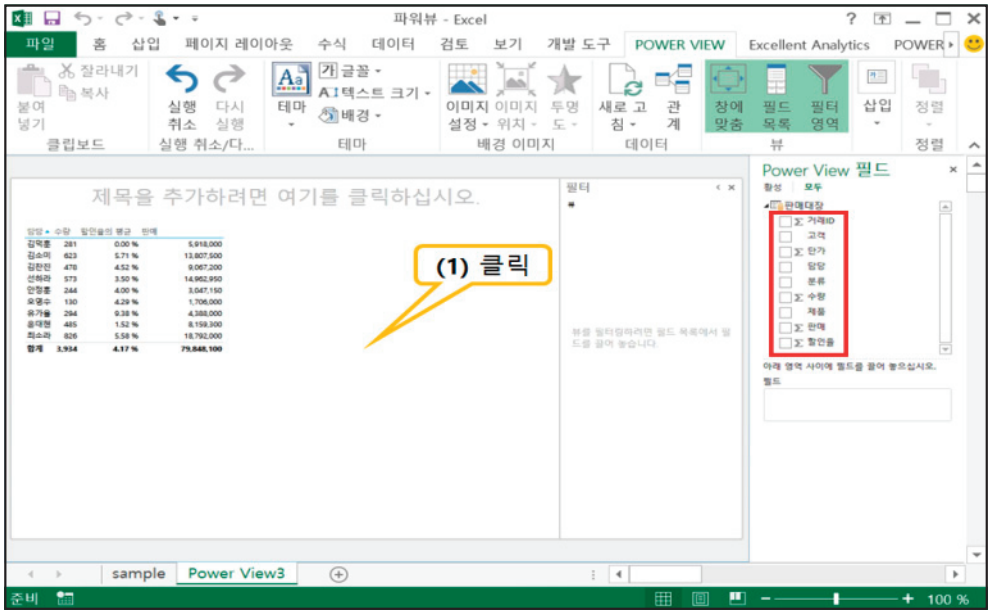


그림 10.72 Power View를 이용한 시각적 보고서 만들기 7

(8) [Power View 필드] 작업 창에서 [고객], [판매] 필드에 체크 표시하면 [Power View] 시트에 새로운 집계표가 나타난다. [판매] 필드를 선택하고 리본 메뉴의 [디자인] 탭 - [숫자] 그룹 - [쉼표 스타일]을 클릭해 천 단위 구분 기호를 삽입한다.

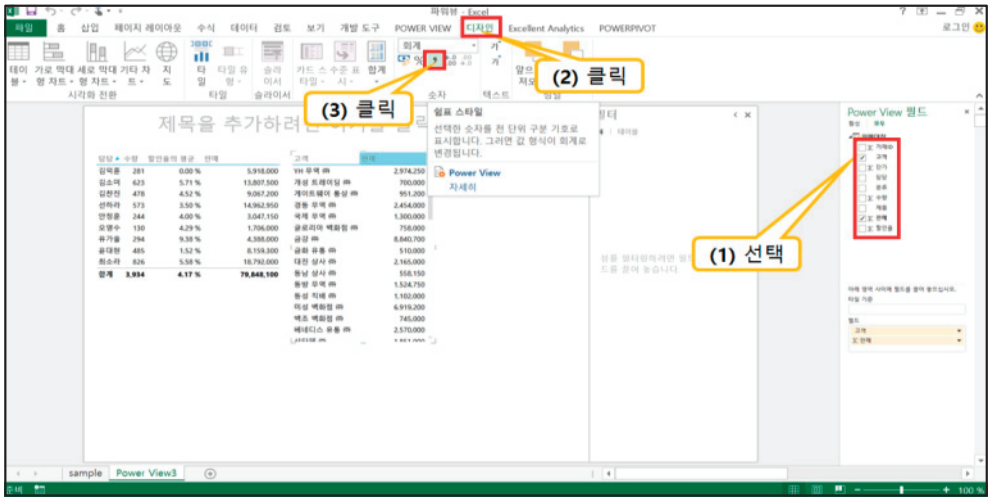


그림 10.73 Power View를 이용한 시각적 보고서 만들기 8

(9) 고객별 실적은 매출이 높은 순으로 표시한다. [판매] 필드의 머리글 부분을 두 번 클릭하면 매출 실적이 높은 순으로 정렬되어 나타난다.

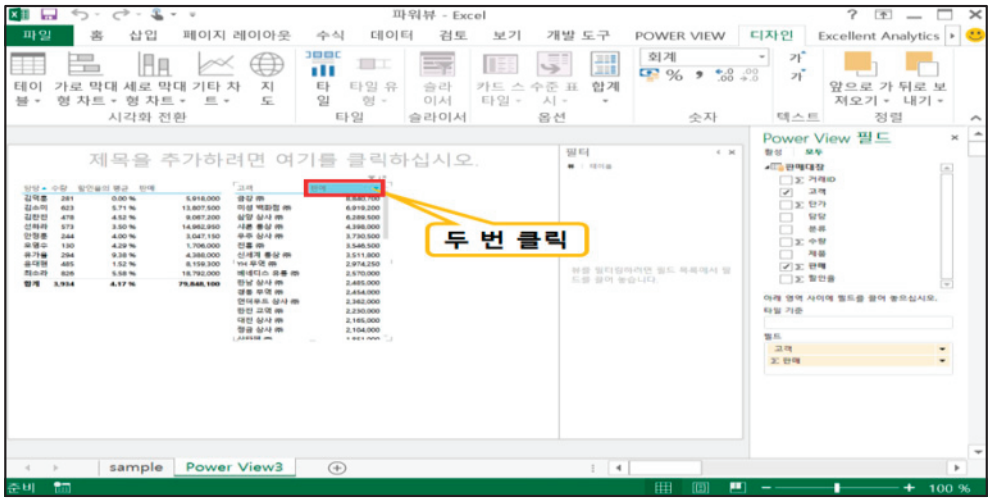


그림 10.74 Power View를 이용한 시각적 보고서 만들기 9

(10) 고객별 실적과 같이 항목이 많은 경우는 데이터를 제한할 수 있다. [필터] 영역의 [테이블]을 선택한다. [판매] 필드를 클릭하고 파란색 막대를 오른쪽으로 드래그하면 [Power View] 시트의 고객 데이터 개수가 줄어든다. 화면을 참고해 적당한 위치까지 드래그 한다.

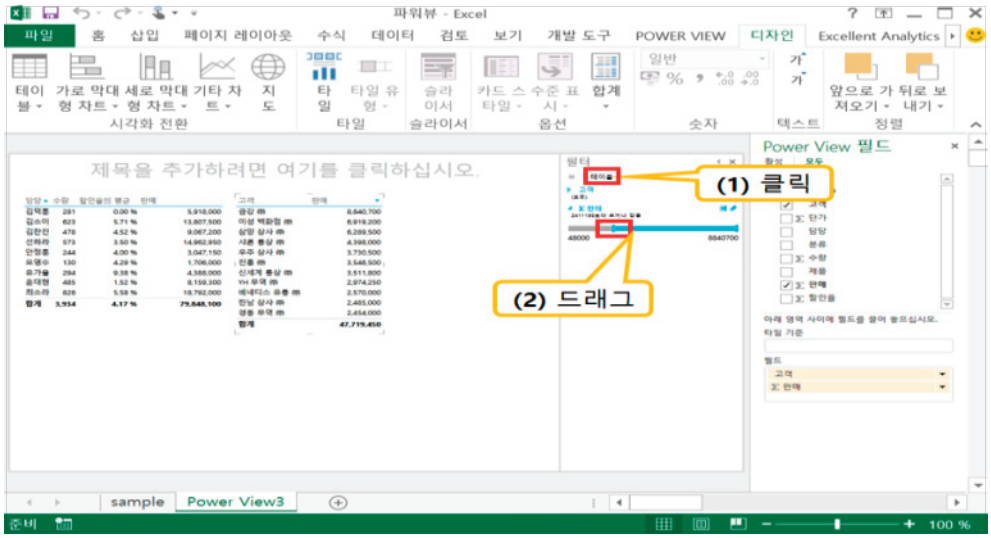


그림 10.75 Power View를 이용한 시각적 보고서 만들기 10

(11) 차트로 변경한다. 리본 메뉴의 [디자인] 탭 - [시각화 전환] 그룹 - [세로 막대형 차트] - [묶은 세로 막대형]을 선택한다.

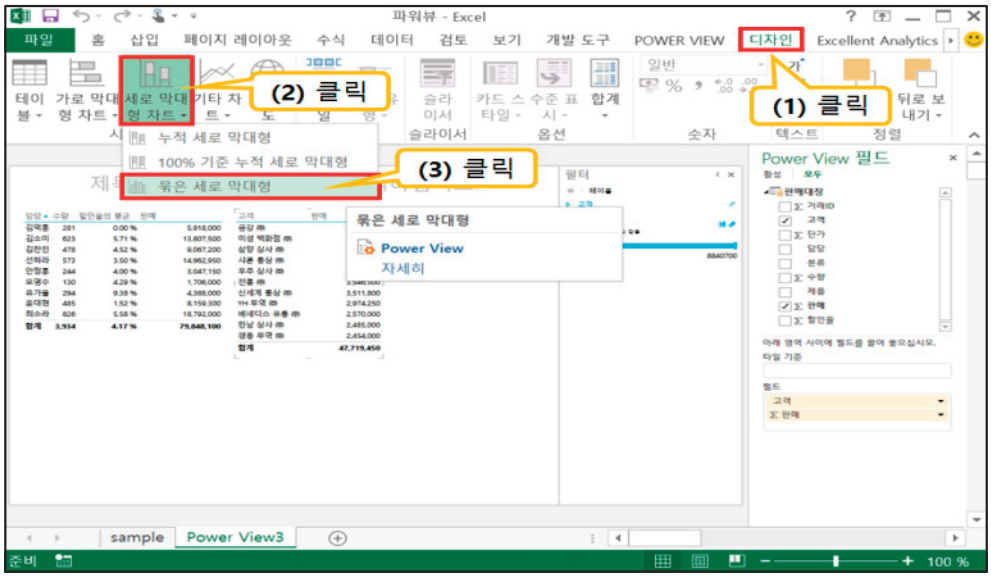


그림 10.76 Power View를 이용한 시각적 보고서 만들기 11

(12) [Power View] 시트 내 집계표가 세로 막대형 차트로 변경된다. 우측 하단의 크기 조정 핸들을 드래그해서 적당한 크기로 조정한다.

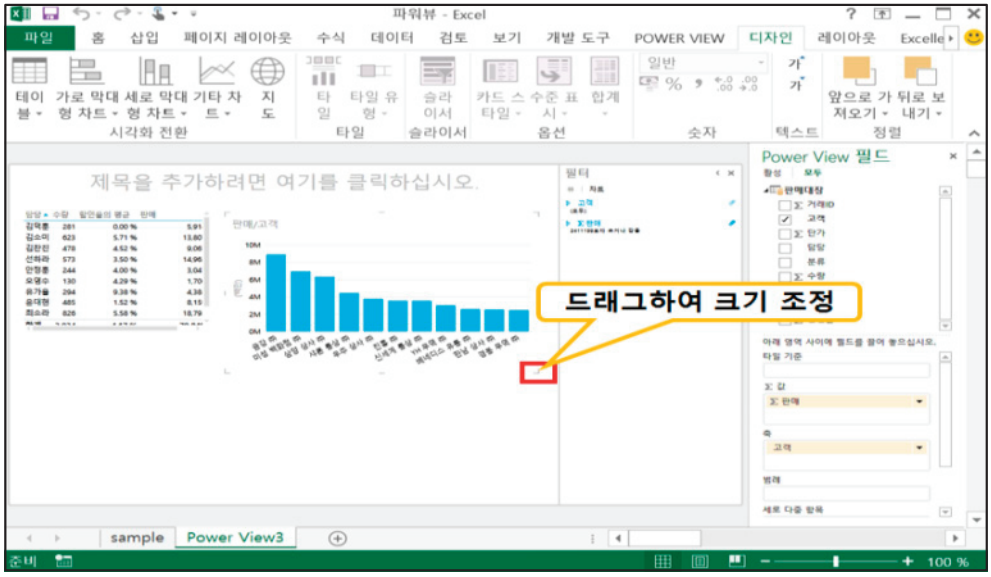


그림 10.77 Power View를 이용한 시각적 보고서 만들기 12

(13) 이런 방식으로 원하는 집계표와 차트를 모두 추가한 후 [Power View] 시트 상단의 제목을 '월간 보고서'로 입력한다.



그림 10.78 Power View를 이용한 시각적 보고서 만들기 13

(14) 마지막으로 [필터] 영역은 Power View 보고서를 조정할 때는 편리하지만, 인쇄하면 해당 부분까지 인쇄된다. 보고서 부분만 나오도록 하려면 이 부분을 축소해야한다. 13. 과정에서 [필터] 영역 닫기 버튼 왼쪽에 [최소화] 버튼이 있는데, 클릭하면 화면과 같이 [필터] 영역이 축소된다. 이 상태에서 인쇄하면 [필터] 영역이 나타나지 않는다.

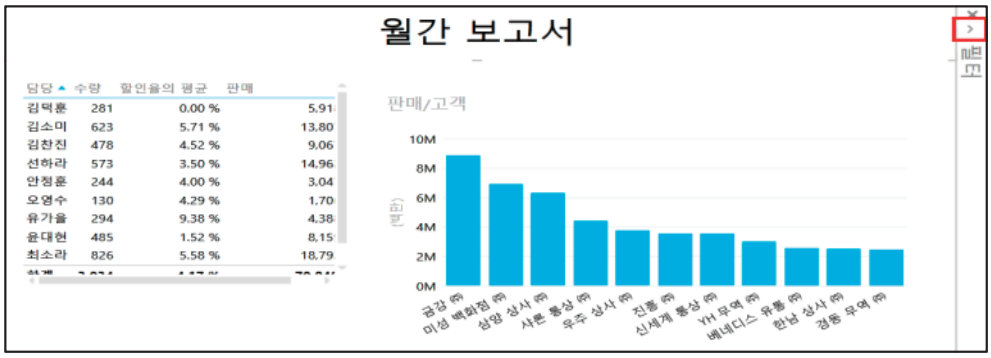


그림 10.79 Power View를 이용한 시각적 보고서 만들기 14

4. 고객 이름별 판매분석을 위한 타일 시각화[61)]

※ 예제파일 – [Sample – Superstore Sales (Excel)]

(1) Superstore Sales (Excel) 파일을 열면 화면과 같은 표를 확인할 수 있다. 이 표의 데이터가 Superstore의 판매샘플 데이터라고 가정하고 고객별 정보를 시각화하는 연습을 하기 위해 파워뷰 기능을 사용해 보자. [Order] 시트의 데이터를 모두 선택한 뒤 엑셀 표에서 리본메뉴의 [삽입] 탭 - [보고서] 그룹 - [Power view]를 클릭한다.

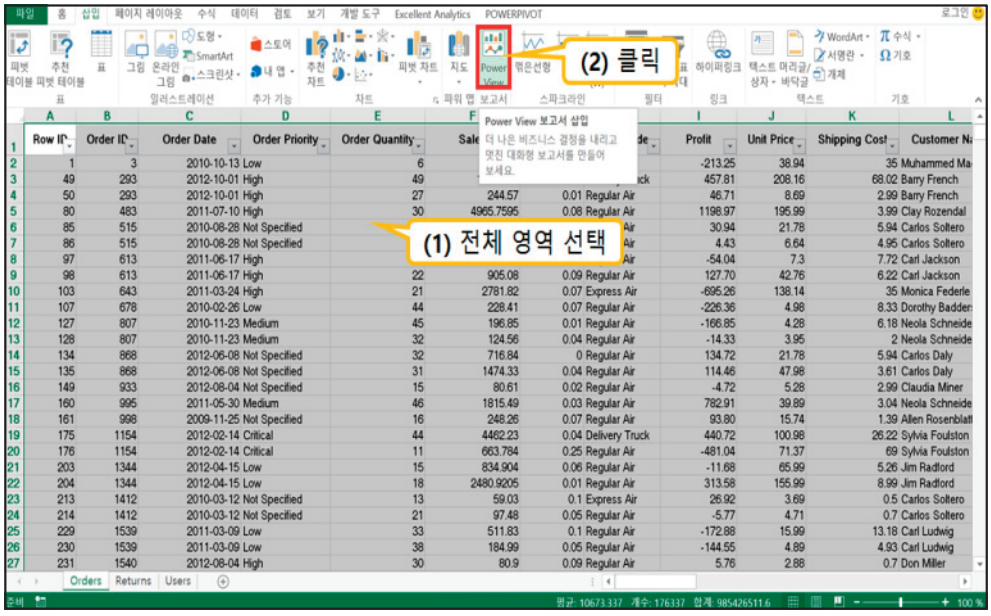


그림 10.80 Power View를 이용한 타일 시각화 1

61) http://seunghoon-chae.com/ch01-%EC%97%91%EC%85%80%EB%A1%9C-%ED%95%98%EB%8A%94-%EB%8D%B0%EC%9D%B4%ED%84%B0-%EC%8B%9C%EA%B0%81%ED%99%94-power-view-1/

(2) 화면과 같이 [Power View1] 시트가 추가되면서 엑셀 표 데이터를 요약한 결과를 표시한다.

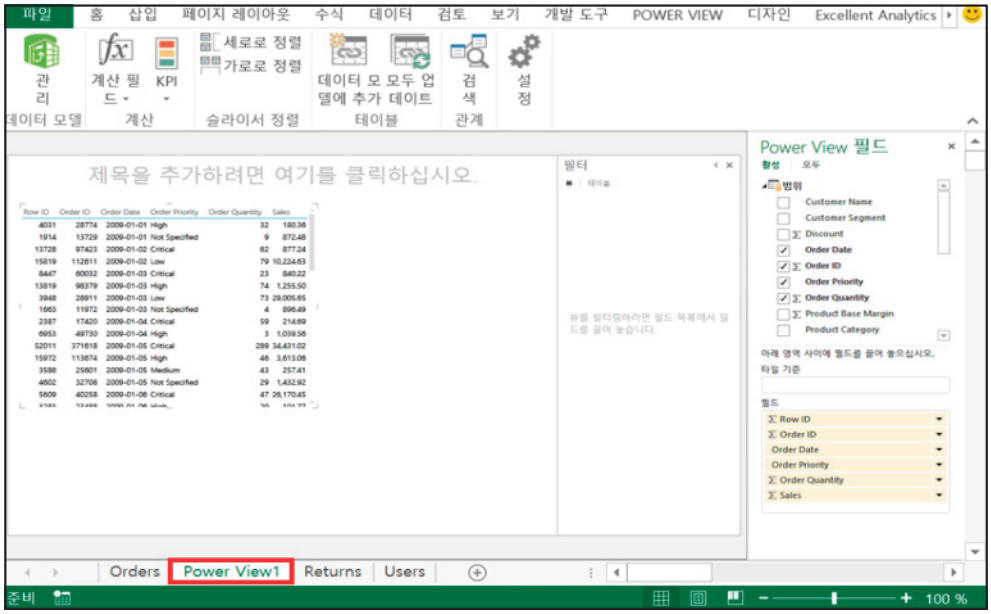


그림 10.81 Power View를 이용한 타일 시각화 2

Power View가 제공하는 다양한 시각화 기능
Power view는 막대 및 차트, 타일, 지도 시각화, 대시보드 까지 다양한 시각화 기능을 지원한다. 사용자가 원하는 시각화 방법에 따라 [Power View 필드] 작업 창에서 원하는 필드를 선택하면 된다.

(3) 각 고객별로 구매물품과 판매액, 수익을 보고자 한다. 이에 따라 [Power view 필드]에서 Customer Name, Product Sub-Category, Profit, Sales를 선택한다. (추가로 원하는 데이터 항목들을 추가할 수 있다.)

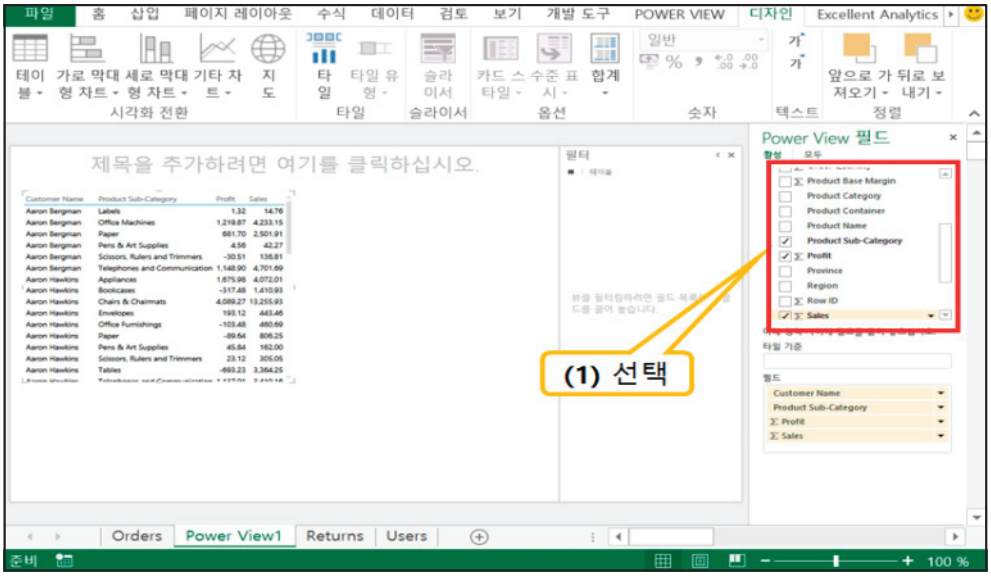


그림 10.82 Power View를 이용한 타일 시각화 3

(4) 리본 메뉴의 [디자인] 탭 - [타일] 그룹 - [타일]을 클릭한다.

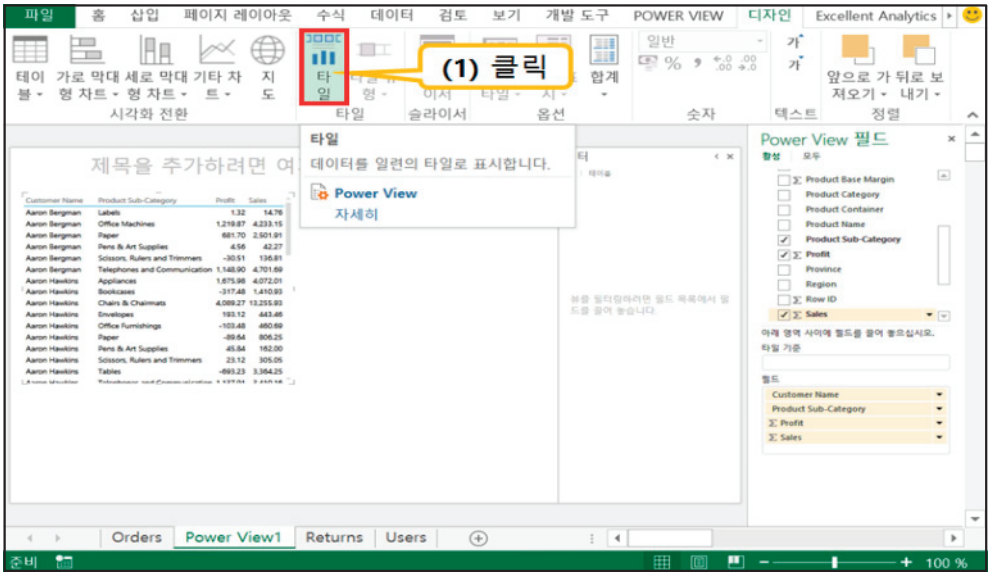


그림 10.83 Power View를 이용한 타일 시각화 4

(5) 아래와 같이 고객 이름을 중심으로 한 타일이 생성된다. 그리고 각 고객명을 선택하면 해당 고객이 구매한 금액과, 물품, 그리고 각 물품별 수익율을 볼 수 있다.

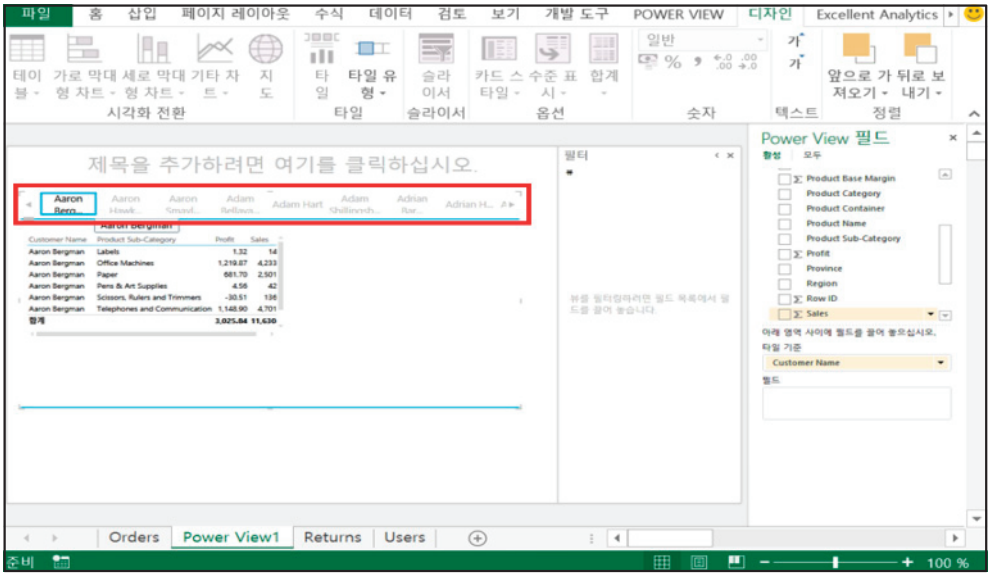


그림 10.84 Power View를 이용한 타일 시각화 5

5. 지역별 매출분석을 위한 지도 시각화[62)]

※ 예제파일 – [Sample – Superstore Sales (Excel)]

(1) Superstore Sales (Excel) 파일을 열어서 [Power View1] 시트를 생성한다. [Power view 필드]에서 Province (지역명 데이터)와 Sales를 선택한다. 이를 통해 지역별 판매현황 지도를 만들어보고자 한다.

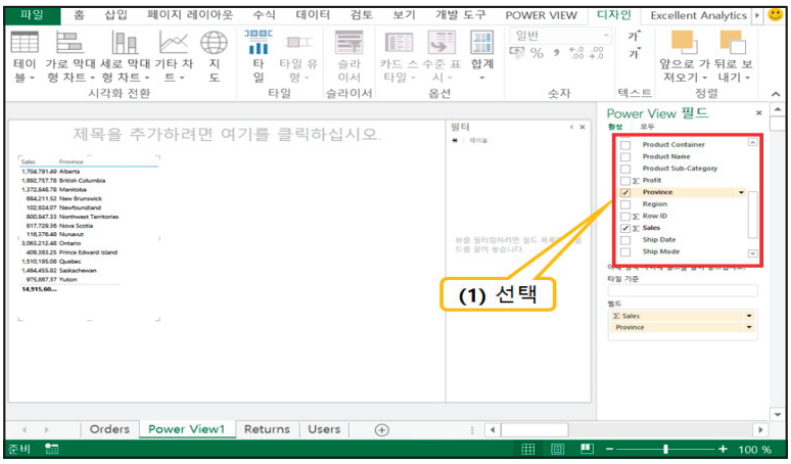


그림 10.85 Power View를 이용한 지도 시각화 1

(2) [디자인] 탭 - [지도]를 클릭한다.

62) http://seunghoon-chae.com/ch02-%EC%97%91%EC%85%80%EB%A1%9C-%ED%95%98%EB%8A%94-%EB%8D%B0%EC%9D%B4%ED%84%B0-%EC%8B%9C%EA%B0%81%ED%99%94-power-view-2/

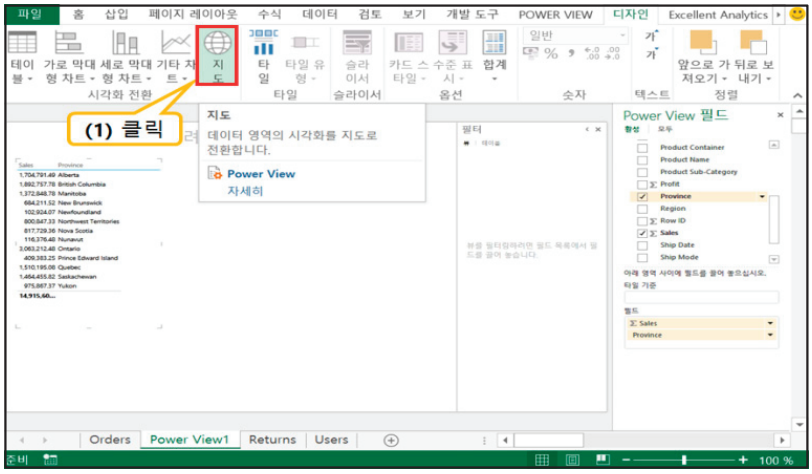


그림 10.86 Power View를 이용한 지도 시각화 2

(3) Power view의 지도는 Microsoft Bing의 지도 API를 받아서 사용하고 있다. 따라서 다음과 같은 경고 메시지 줄이 표시된다. 콘텐츠 사용을 클릭한다.

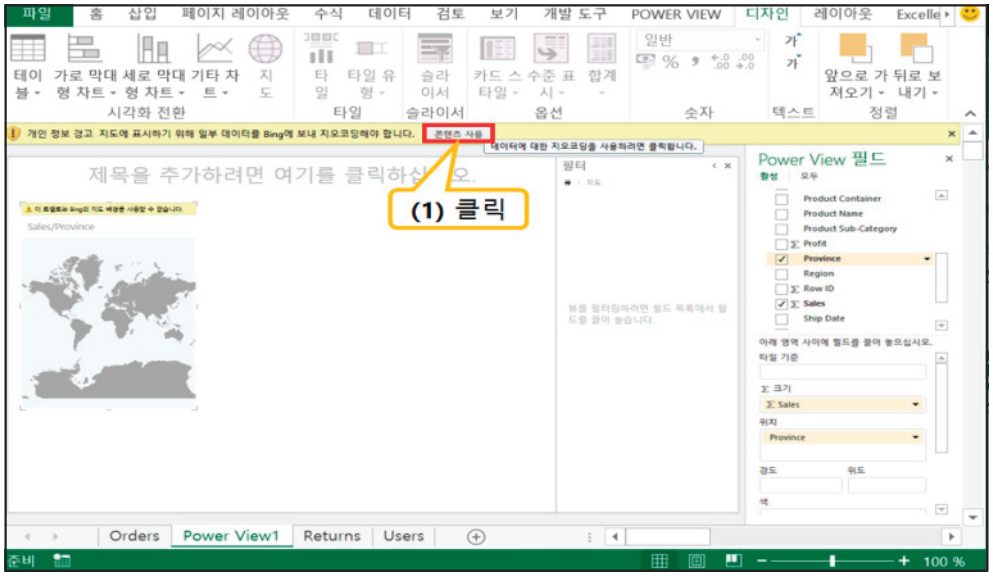


그림 10.87 Power View를 이용한 지도 시각화 3

(4) 아래 화면에서는 지도가 제대로 세부적으로 표현되지 않으며 경고표시와 함께 '이 로캘로는 Bing의 지도 배경을 사용할 수 없습니다.'라고 나온다. 이는 Power view 와 Bing의 호환문제이므로 일단 무시하고 지도 시각화를 이어서 진행하도록 한다.

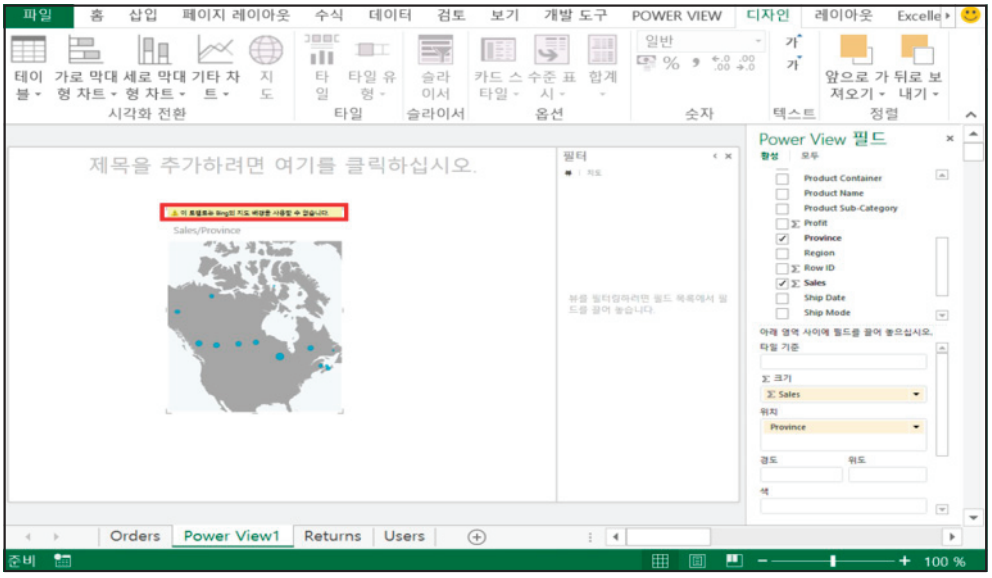


그림 10.88 Power View를 이용한 지도 시각화 4

(5) [Power view 필드]에서 Product category을 선택하면 상품 카테고리별로 지역별 매출을 볼 수 있다. 이처럼 필드에서 데이터 항목을 추가시키면 바로 지도나 차트, 그래프에 반영되는 것이 Power view 의 가장 큰 장점이다.

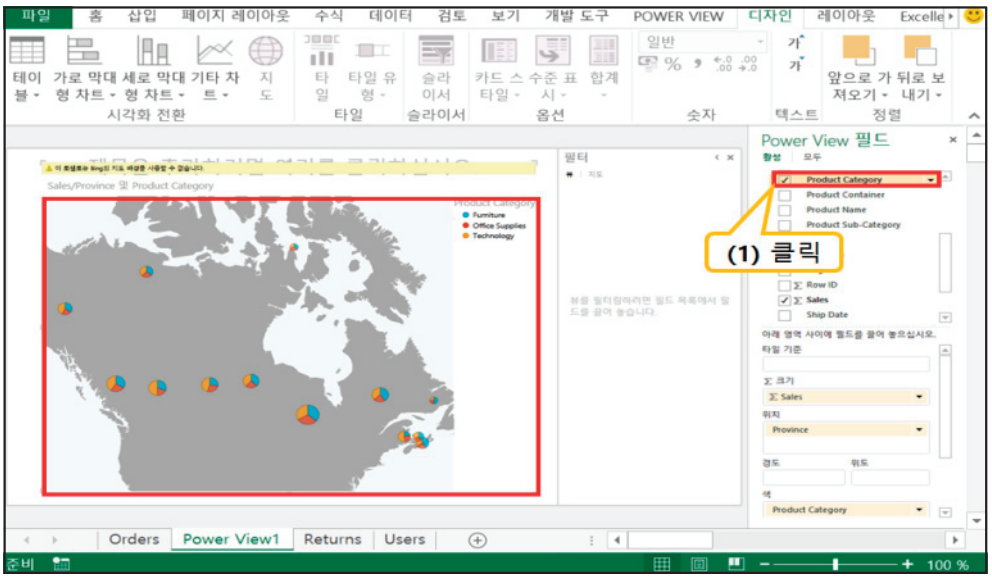


그림 10.89 Power View를 이용한 지도 시각화 5

(6) 이번에는 필터 기능을 사용해보자. 필터를 통해 특정 상품이나 지역에 대해 선택적으로 시각화가 가능하다. 자신이 원하는 상품 및 지역을 선택해서 지도상에서 효과적으로 데이터를 탐색할 수 있다.

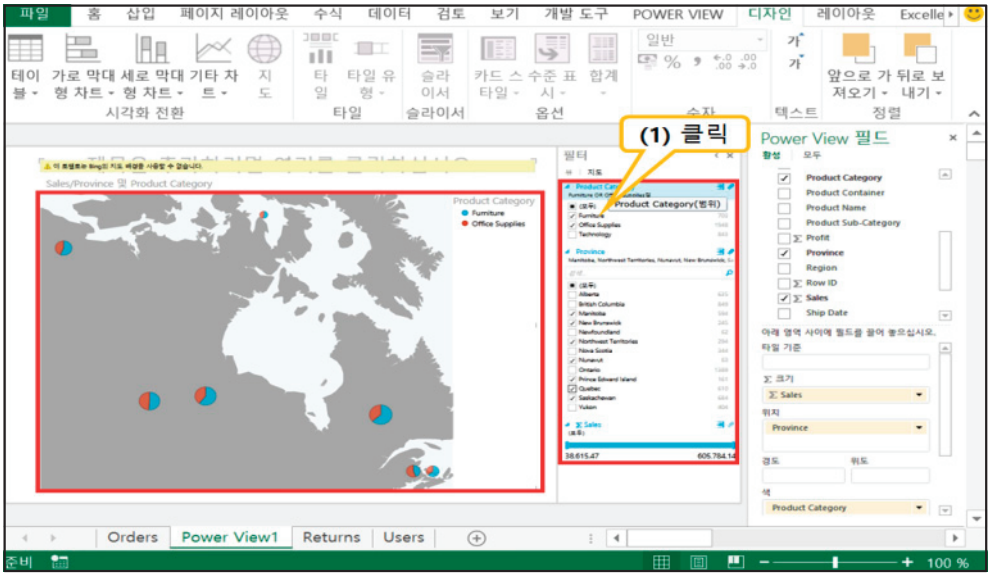


그림 10.90 Power View를 이용한 지도 시각화 6

6. 대시보드를 활용한 매출보고서 작성

이제 Power View 에서 대시보드를 활용해서 매출보고서를 만들어보자. Power View 대시보드의 경우 해당 전용 기능이 있는 것이 아니라 Power View 필드의 데이터들을 체크하면 한 화면 안에 여러 가지 차트 및 그래프를 생성할 수 있다.

(1) 먼저 '고객 이름별 판매분석을 위한 타일 시각화'과정에서 만들었던 고객 매출현황을 볼 수 있는 타일형 테이블을 생성한다.

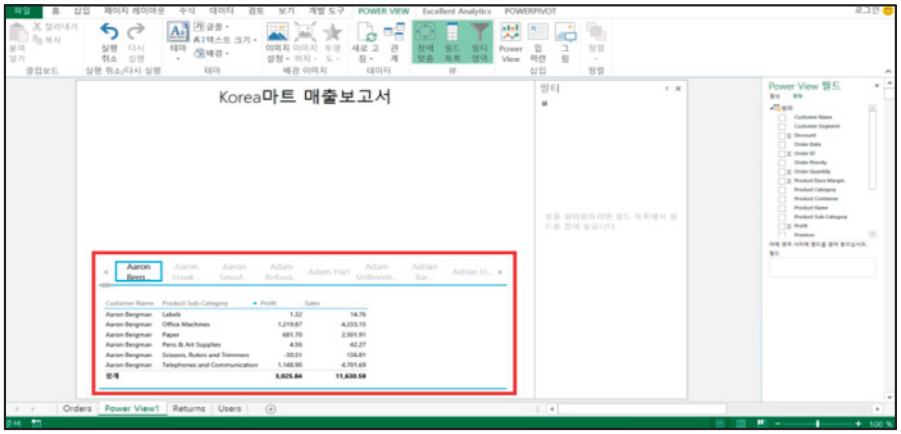


그림 10.91 Power View를 이용한 매출보고서 작성 1

(2) [Power View] 시트의 빈 영역을 클릭하면 [Power View 필드]작업창의 필드가 모두 해제된 상태로 표시된다. [Power View 필드] 작업창에서 [Sales], [Order data] 필드에 체크 표시하면 [Power View]시트에 새로운 집계표가 나타난다.

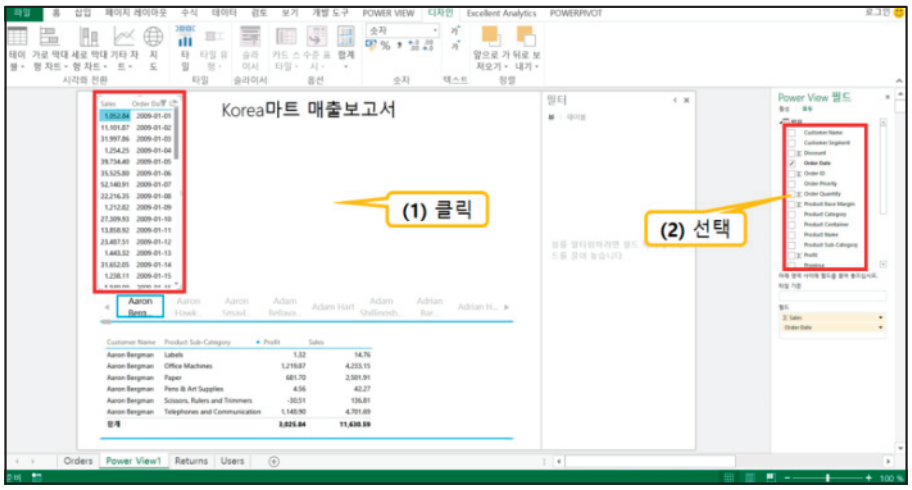


그림 10.92 Power View를 이용한 매출보고서 작성 2

(3) 새로운 집계표가 선택된 상태에서 리본 메뉴의 [디자인] 탭 - [시각화 전환] 그룹 - [기타 차트] - [선]을 선택한다.

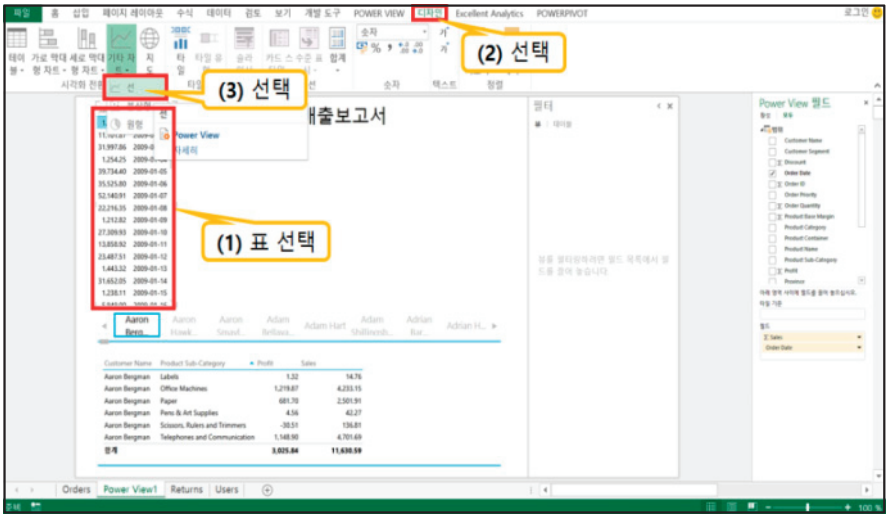


그림 10.93 Power View를 이용한 매출보고서 작성 3

(4) 만들어진 선 그래프 역시, 필터 기능을 통해 원하는 일자만 선택해서 데이터를 탐색 할 수 있다.

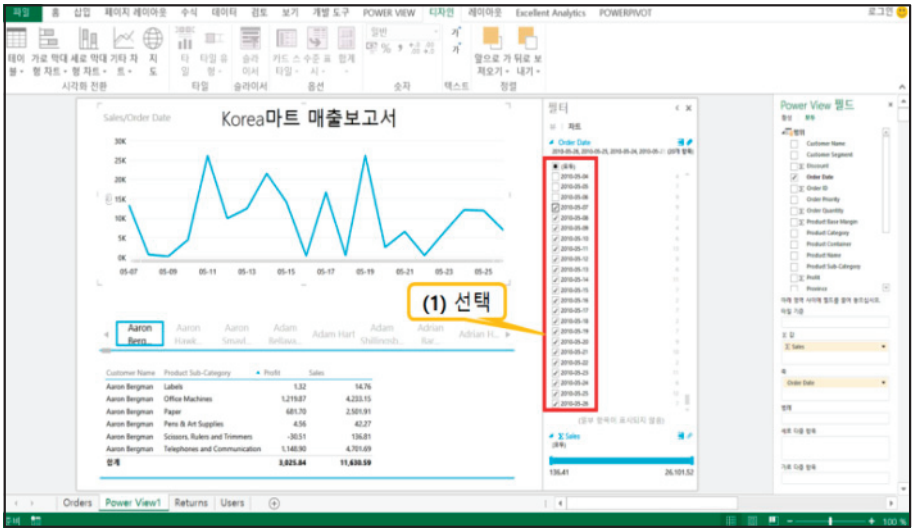


그림 10.94 Power View를 이용한 매출보고서 작성 4

(5) 추가로 상품수량과 매출간의 연관성을 보여주는 산포도를 만들어 보도록 하겠다. [Power View] 시트의 빈 영역을 클릭한 뒤에, [Power View 필드] 작업창에서 [Order Quantity], [Product Sub-Category], [Sales] 필드에 체크

표시한다. 새롭게 만들어진 표가 선택된 상태에서 리본 메뉴의 [디자인] 탭 - [시각화 전환] 그룹 - [기타 차트] - [분산형]을 클릭한다.

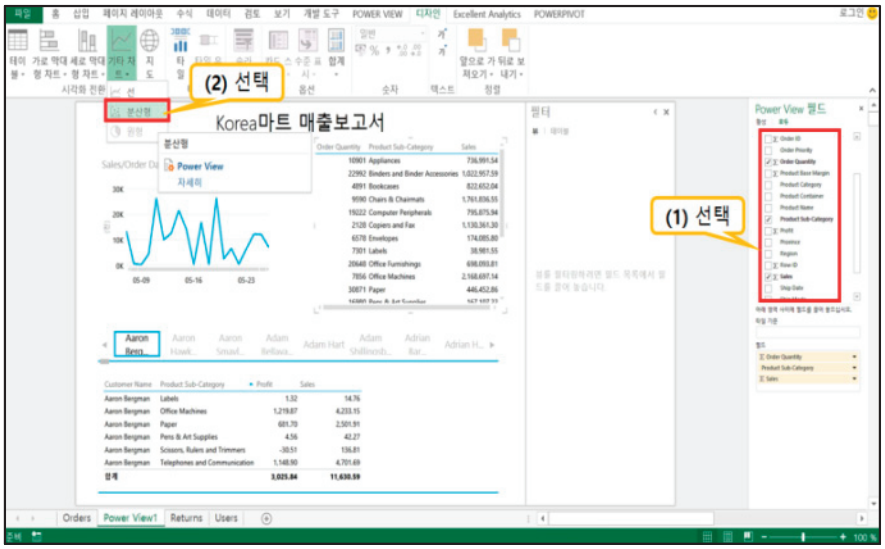


그림 10.95 Power View를 이용한 매출보고서 작성 5

(6) 만들어진 분산형 그래프 역시, 필터 기능을 통해 원하는 일자만 선택해서 데이터를 탐색 할 수 있다.

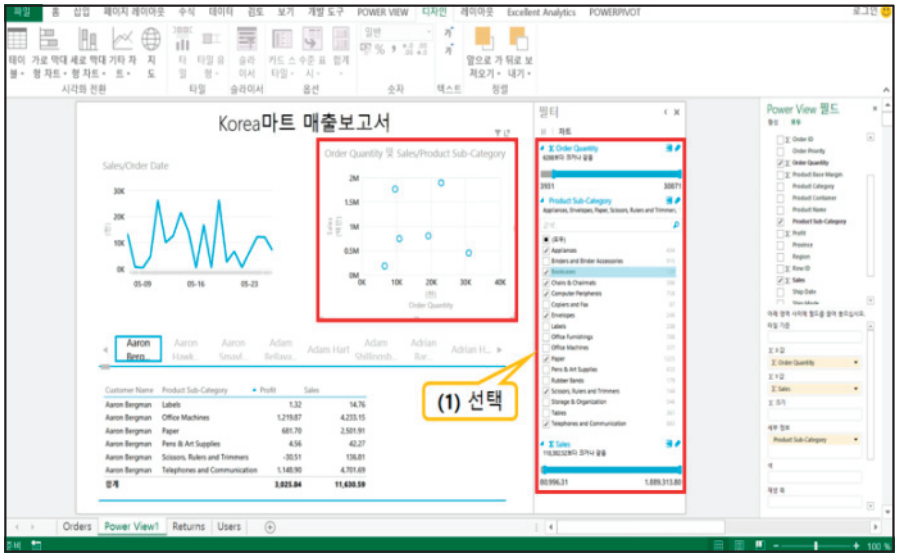


그림 10.96 Power View를 이용한 매출보고서 작성 6

10.3.3 Power Map

1. 파워 맵 설치하기[63]

Power Pivot이나 Power View와는 달리 Power map은 별도의 설치파일을 다운받아서 설치해야 한다. 엑셀 2013 버전부터 지원된다.

(1) 구글에서 “power map excel 2013 download”로 검색한 뒤 첫 번째 페이지에 접속한다.

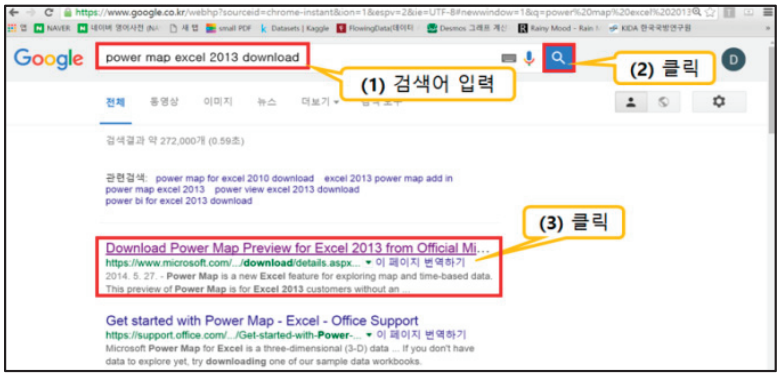


그림 10.97 파워맵 설치하기 1

(2) Download 버튼을 클릭한다.

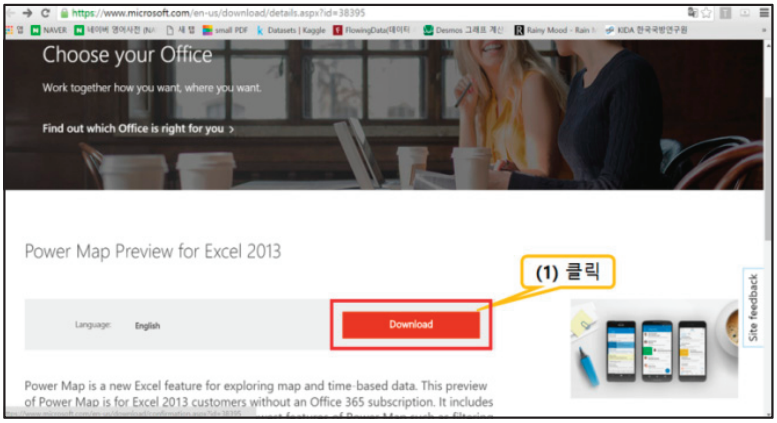


그림 10.98 파워맵 설치하기 2

63) http://seunghoon-chae.com/ch03-%EC%97%91%EC%85%80%EB%A1%9C-%ED%95%98%EB%8A%94-%EB%8D%B0%EC%9D%B4%ED%84%B0-%EC%8B%9C%EA%B0%81%ED%99%94-power-map-1/

(3) Download Center 홈페이지에 접속하면 자동으로 프로그램 다운로드가 시작된다.

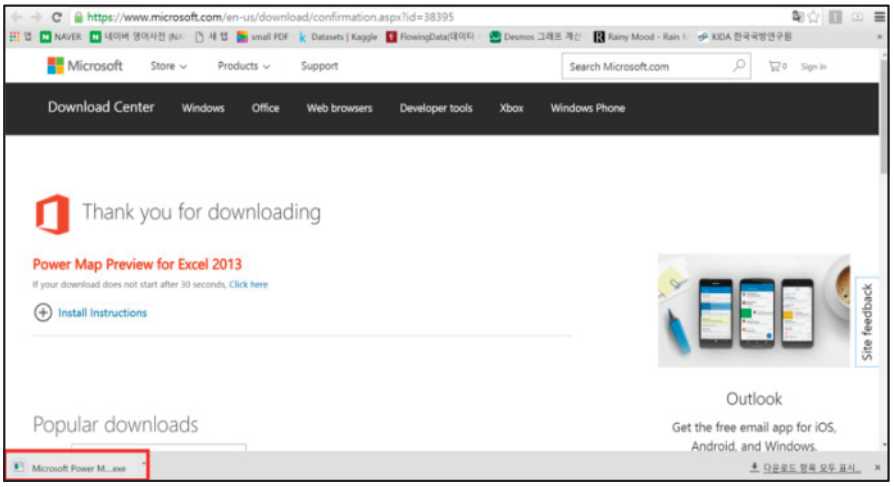


그림 10.99 파워맵 설치하기 3

(4) 다운로드받은 설치파일을 열고, Agree와 Next를 반복하며 설치를 진행한다.

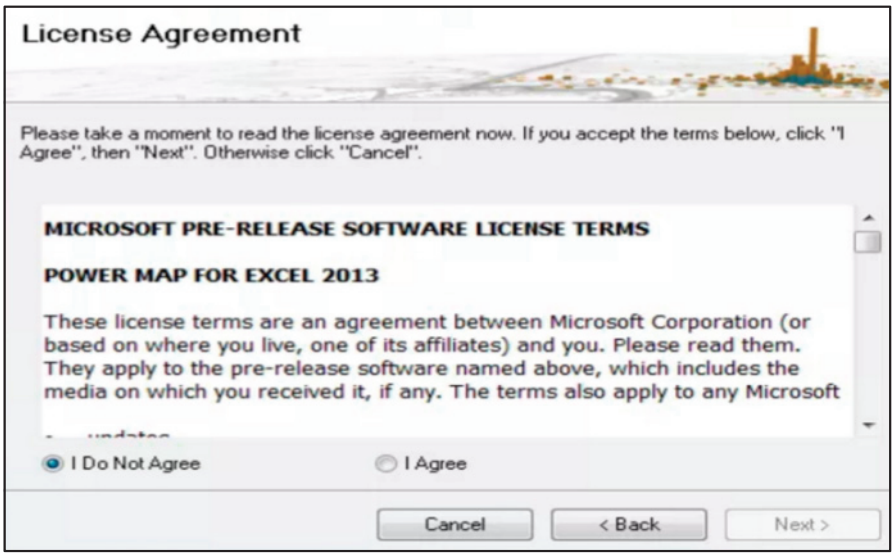


그림 10.100 파워맵 설치하기 4

(5) 설치가 끝나면 리본메뉴의 [파일] - [옵션] - [추가기능] - [COM 추가 기능]을 선택해서 우측에 있는 [이동]을 클릭한다.

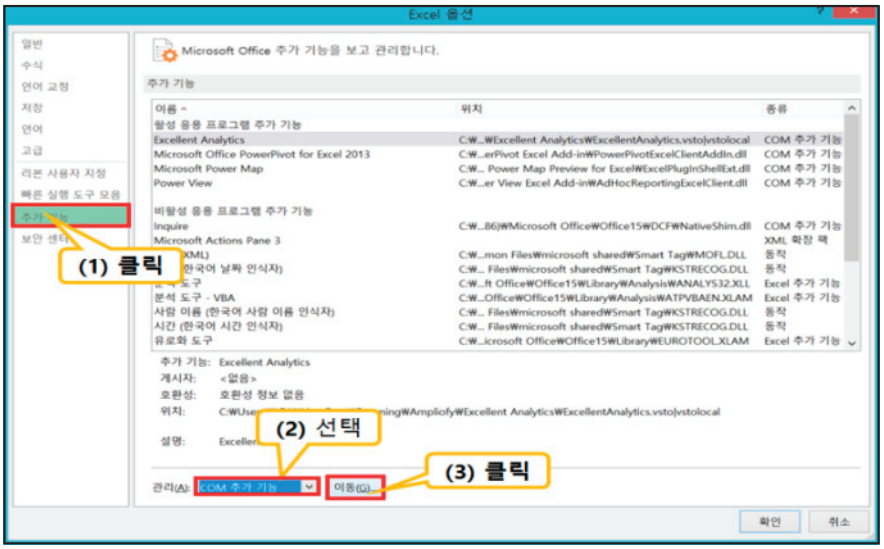


그림 10.101 파워맵 설치하기 5

(6) COM 추가 기능 창이 나타나면 Microsoft Power Map을 체크하고 확인을 눌러준다.

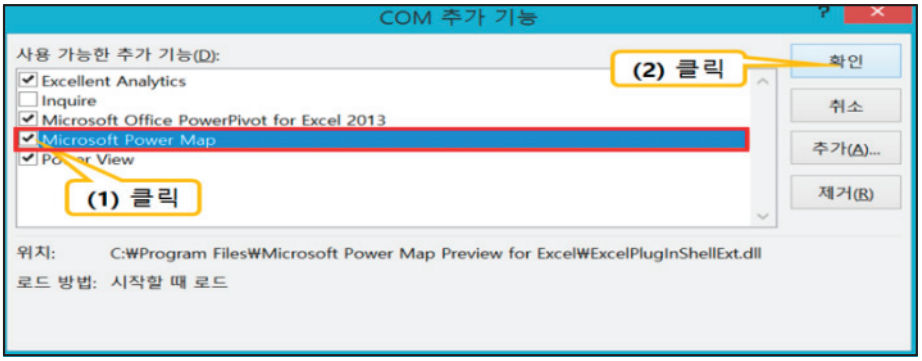


그림 10.102 파워맵 설치하기 6

(7) 리본 메뉴의 [삽입]탭의 중간에 [파워 맵]그룹의 [지도]가 생성되었다.

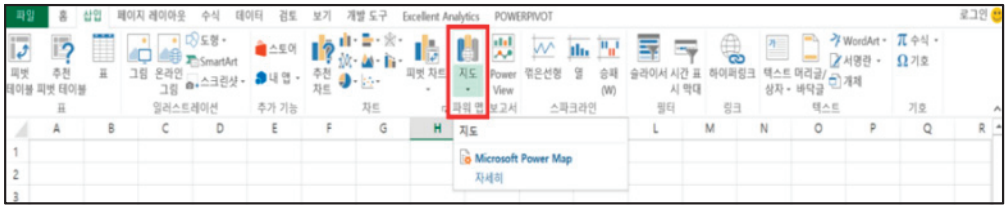


그림 10.103 파워맵 설치하기 7

2. 파워맵 인터페이스 및 기본 사용법[64)]

※ 예제파일 – [Sample – Superstore Sales (Excel).xlsx]

Sample – Superstore Sales (Excel).xlsx 라는 엑셀 파일을 불러온다. 이 파일을 이용해서 Power map의 인터페이스 및 기본 사용법에 대해 알아보도록 하겠다.

(1) 불러온 예제 파일에서 데이터가 있는 부분 중 아무 셀이나 선택한다. 그 상태에서 [삽입] 탭 - [지도] - [파워 맵 시작]을 클릭한다. 이 메뉴가 안보인다면 옵션을 수정해야 한다.

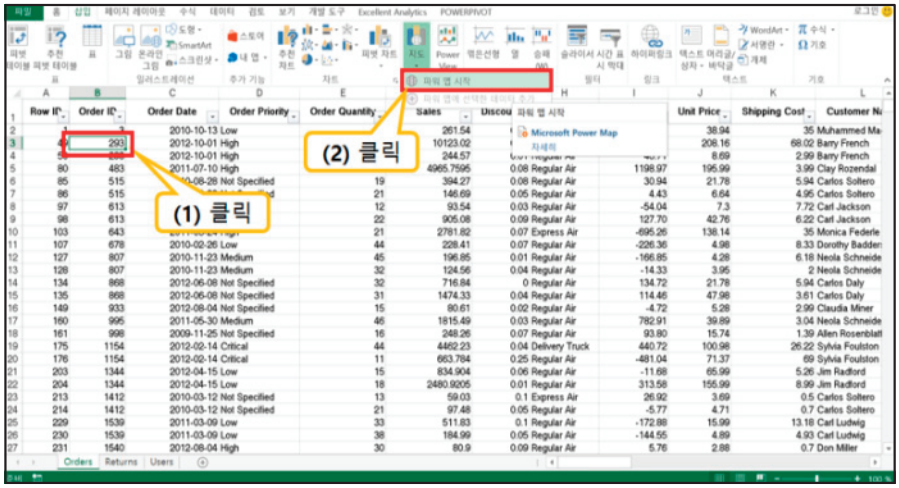


그림 10.104 파워맵 인터페이스 및 기본 사용법 1

64) http://seunghoon-chae.com/ch04-%EC%97%91%EC%85%80%EB%A1%9C-%ED%95%98%EB%8A%94-%EB%8D%B0%EC%9D%B4%ED%84%B0-%EC%8B%9C%EA%B0%81%ED%99%94-power-map-2/

Power map 을 실행하면 Power map 전용 시트가 생성된다. 먼저 메뉴를 한번 살펴보면 [홈] 탭에 [투어] 그룹, [지도] 그룹, [레이어] 그룹, [시간] 그룹, [보기] 그룹이 있고 이를 통해 다양한 결과물을 만들어볼 수 있다.

먼저 [투어] 그룹 내의 [비디오 만들기]의 경우 Power map 에서 작업한 결과물을 자동으로 동영상 파일로 만들어준다. 예를 들어 연도별 매출액 변화추이를 시각화 했다면, 연도별로 해당 각 지역별로 매출액이 어떻게 변화하는지를 보여주는 (ex 막대그래프의 높낮이 변화) 동영상이 생성된다.

또한 Power map 은 Microsoft의 Bing 지도에 기반하고 있기 때문에 구글 맵 못지 않은 자세한 위치정보를 가지고 있다. Power map 의 시각화 방식은 레이어를 바탕으로 이루어진다. 즉 한 레이어에서 한 가지 그래프나 차트를 만들고, 레이어를 추가하여 다른 차트를 그 위에 추가하는 방식이다. 이를 통해 지도상에 원하는 차트 및 그래프를 나타낼 수 있다. (지오코딩: 위치나 지도 데이터를 지도상에 매핑 시키는 것)

이를 위해서는 Power map 역시 지오코딩을 먼저 해주어야 한다. 즉 위치나 지역명 데이터를 Power map 에게 넘겨주어야 한다.

(2) 화면 오른쪽 레이어1에서 'Provice'를 체크한다. 아래쪽에 지리 및 지도 수준에 'Provice'(시/도)가 추가되었다. 다음 버튼을 누르게 되면 자동으로 지오코딩이 완료된다.

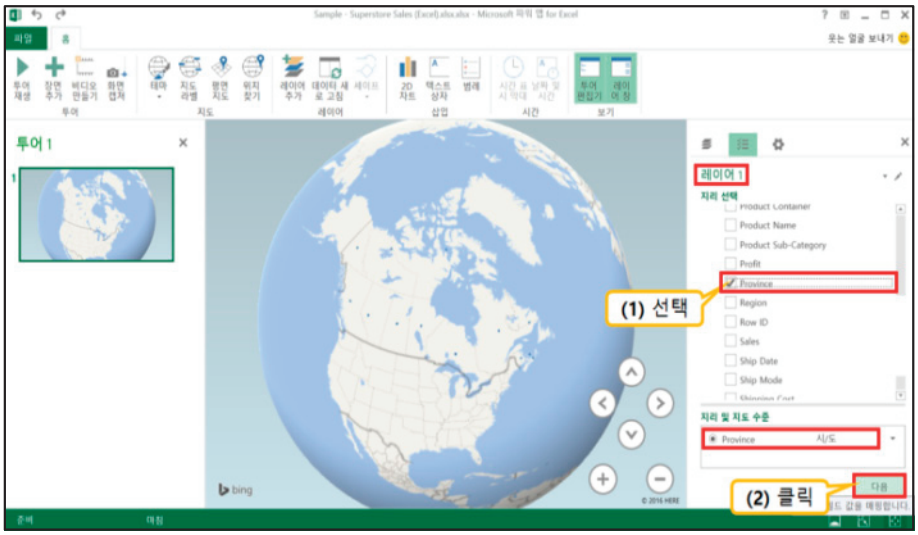


그림 10.105 파워맵 인터페이스 및 기본 사용법 2

(3) Province의 위치명을 기반으로 지도가 설정되었다. 이제 화면 오른쪽 [레이어1] 필드에서 데이터를 체크하는 방식으로 시각화를 해보자. 먼저 'Sales'를 체크하니 지도상에 3D 막대그래프가 생성된다. 지구본 형태가 신기하기는 하나 제대로 보기가 어렵다면 [홈] 탭 - [지도] 그룹 - [평면지도]를 선택해준다. 그러면 지구본이 평면지도로 바뀌게 된다.

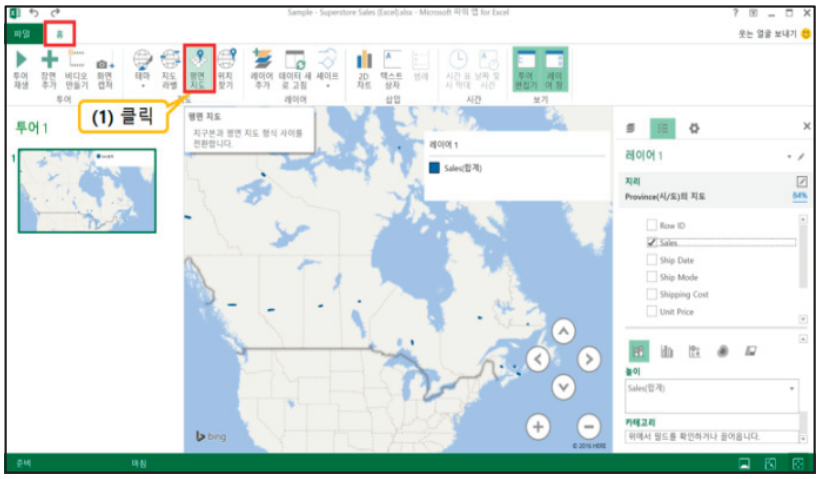


그림 10.106 파워맵 인터페이스 및 기본 사용법 3

(4) 단순한 지역별 매출보다는 상품 카테고리별 매출현황을 알고 싶다. [레이어1] 필드에서 'Product category'를 선택한다. 그러면 각 상품 카테고리별로 매출액 막대그래프가 생성된다. 추가로 [레이어1] 필드를 살펴보면 중간에 그래프 모양의 아이콘들을 살펴볼 수 있다. 본인이 원하는 그래프 형태를 선택해 준다. 여기서는 [시각화를 묶은 세로 막대형으로 변경]을 선택해 보자.

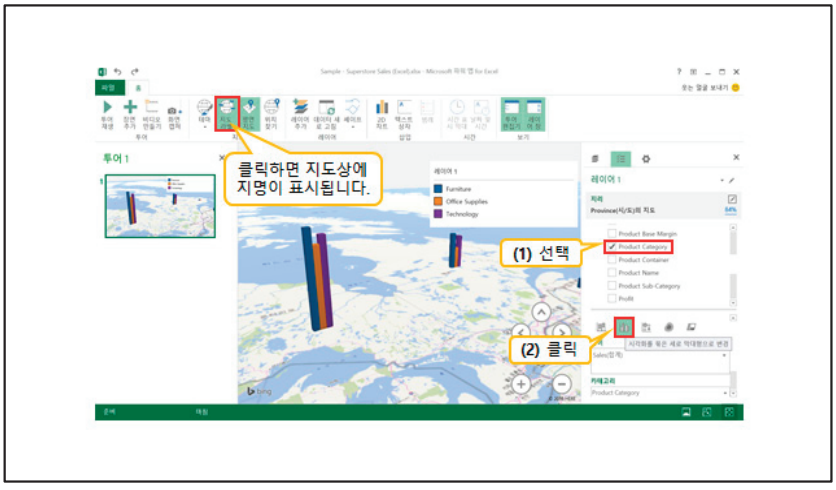


그림 10.107 파워맵 인터페이스 및 기본 사용법 4

(5) 같은 방식으로 기존에 [레이어1]에서 선택되어 있던 'Product category'를 선택 해제하고 'Product sub category'를 선택하면 상품 하위 카테고리별 매출액 막대그래프가 생성된다. 모든 데이터 시각화가 그러하듯이 Power map 의 경우도 Raw data가 상당히 중요하다. 데이터의 양이 크고 디테일 할수록 Power map 이 빛을 발하게 된다.

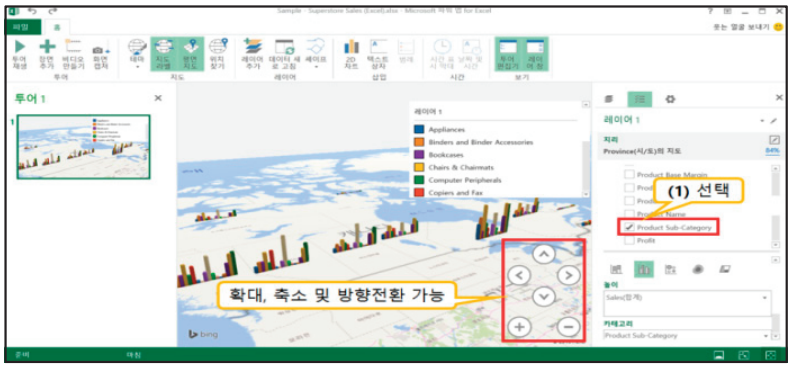


그림 10.108 파워맵 인터페이스 및 기본 사용법 5

마우스 스크롤로 지도를 확대하면 자세한 지도를 볼 수 있다. 국내의 경우 동, 읍, 면 단위까지 지원이 되기 때문에 각 지역별로 세부적인 데이터들만 확보할 수 있다면 상권분석이나, 도시개발 등 다양한 분야에서 유용하게 사용될 수 있겠다.

(6) [홈] 탭 - [지도] 그룹 - [테마]를 선택해서 다음과 같이 테마를 변경할 수 있다.

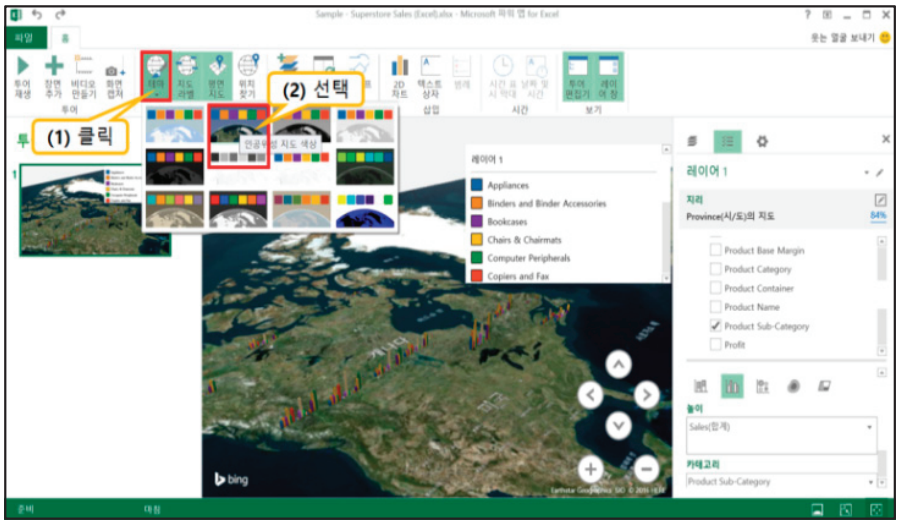


그림 10.109 파워맵 인터페이스 및 기본 사용법 6

(7) 이제 비디오를 만들어 보자. 비디오를 만들기 위해서는 데이터 수치의 변화를 보여 줄 수 있는 '시간 데이터'가 필요하다. 예를 들어 Sales 데이터 같은 경우에는 Oder date 데이터가 있어야 한다. [레이어1] 필드에서 'Oder date'를 체크하거나 마우스 드래그 한 뒤 하단에 시간 창에 넣어 준다. 그러면 플레이 버튼이 생기는데 재생시켜보면 Oder date 별로 매출액이 어떻게 변화했는지가 막대그래프의 높낮이로 나타난다. 이 상태에서 [홈] 탭 - [투어] 그룹 - [비디오 만들기]를 클릭하면 동영상 파일을 만들 수 있다.

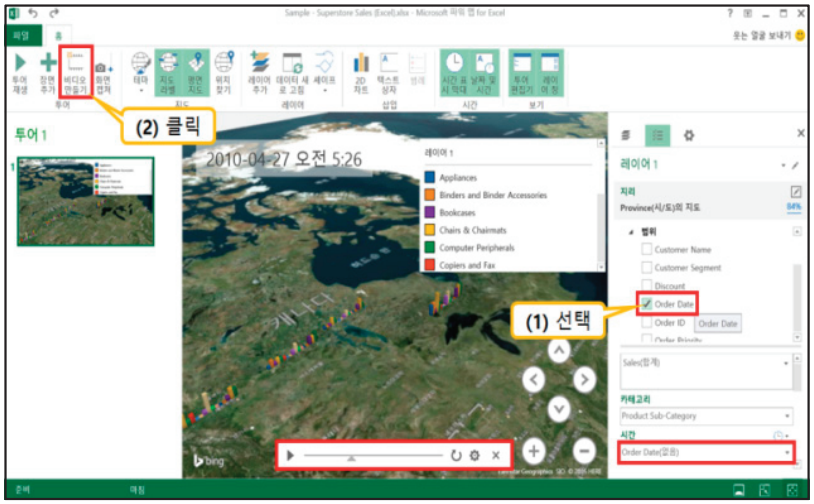


그림 10.110 파워맵 인터페이스 및 기본 사용법 7

(8) [홈] 탭 - [투어] 그룹 - [비디오 만들기]를 클릭 후에 다음과 같은 창이 뜨면, 비디오 품질을 선택하고 [만들기]를 클릭한다.

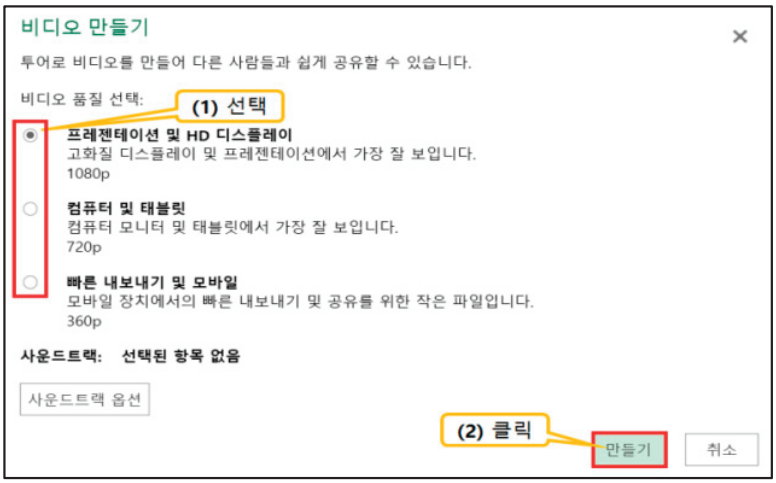


그림 10.111 파워맵 인터페이스 및 기본 사용법 8

(9) 원하는 파일의 저장위치를 선택하고, 파일명을 입력한 뒤 [저장]을 클릭한다.



그림 10.112 파워맵 인터페이스 및 기본 사용법 9

(10) 다음의 과정을 거쳐 파일이 생성되면, 동영상 파일이 생성되고 이를 재생해서 확인 한다.

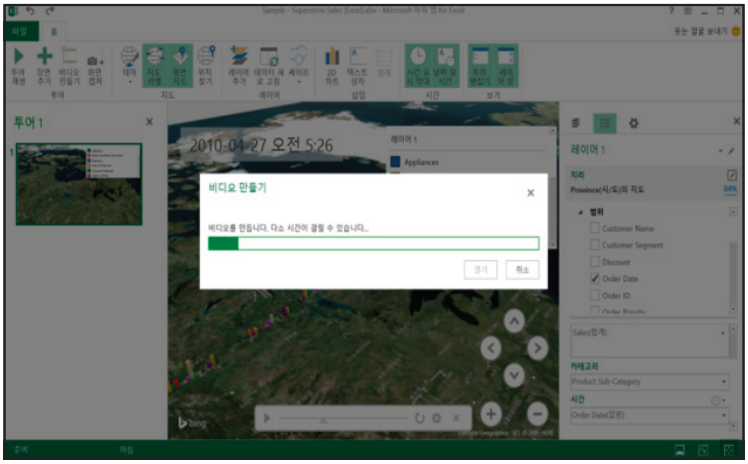


그림 10.113 파워맵 인터페이스 및 기본 사용법 10



그림 10.114 파워맵 결과물

3. 레이어 기능[65)]

※ 예제파일 – [나라별 종합지표.xlsx]

아래 이미지가 이번에 Power map으로 만들 최종 콘텐츠이다. 히트맵은 공기오염도를 의미한다. 히트맵이 진한색일 수록 공기오염도가 심한 국가이다. 히트맵 위에 파란그래프는 출산율을, 빨간 그래프는 유아사망률을 나타낸니다. 시각적으로 히트맵이 진한 국가 일수록 빨간 막대그래프의 높이가 높다. 즉 해당 데이터만 두고 본다면 공기오염도와 유아사망률은 서로 상관관계를 지니고 있다. 이처럼 Power map 은 레이어 기능을 통해 여러 시각화 유형들을 지도상에 나타낼 수 있다.



그림 10.115 파워맵 레이어 예제

65) http://seunghoon-chae.com/ch04-%EC%97%91%EC%85%80%EB%A1%9C-%ED%95%98%EB%8A%94-%EB%8D%B0%EC%9D%B4%ED%84%B0-%EC%8B%9C%EA%B0%81%ED%99%94-power-map-3/

(1) 불러온 예제 파일에서 데이터가 있는 부분 중 아무 셀이나 선택한다. 그 상태에서 [삽입] 탭 - [지도] - [파워 맵 시작]을 클릭한다. 이 메뉴가 안보인다면 옵션을 수정해야 한다.

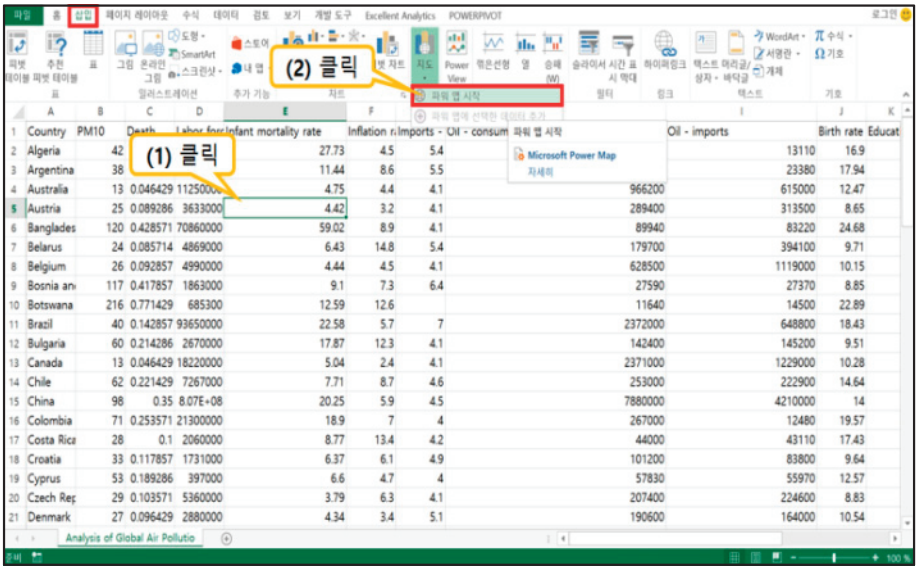


그림 10.116 파워맵 레이어 기능 1

(2) Power map 전용시트가 생성된다. 지오코딩을 위해 [레이어1] 필드에서 'Country'(국가명 데이터)를 선택한 뒤 다음을 눌러 진행한다.

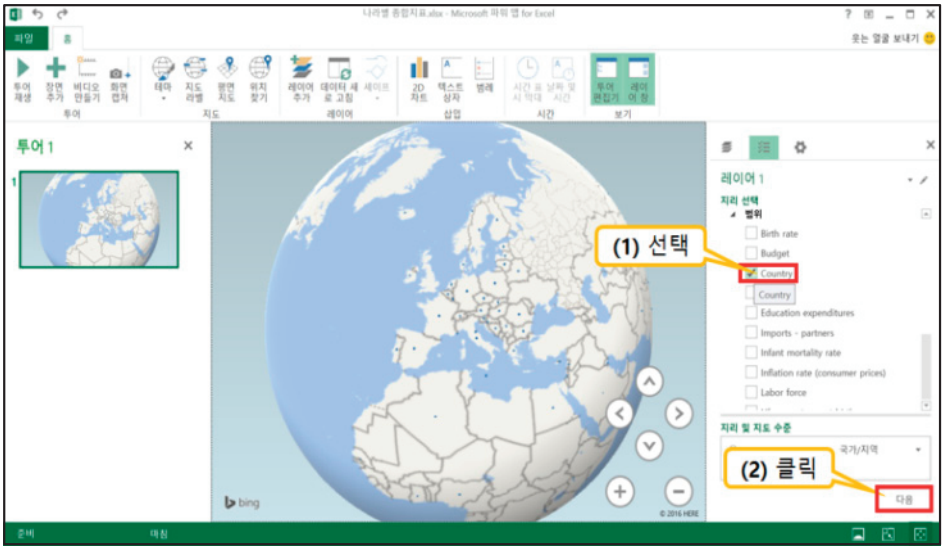


그림 10.117 파워맵 레이어 기능 2

(3) [레이어1] 필드에서 'PM10'(공기오염도)를 선택한 뒤 시각화 유형으로 히트맵을 선택한다.

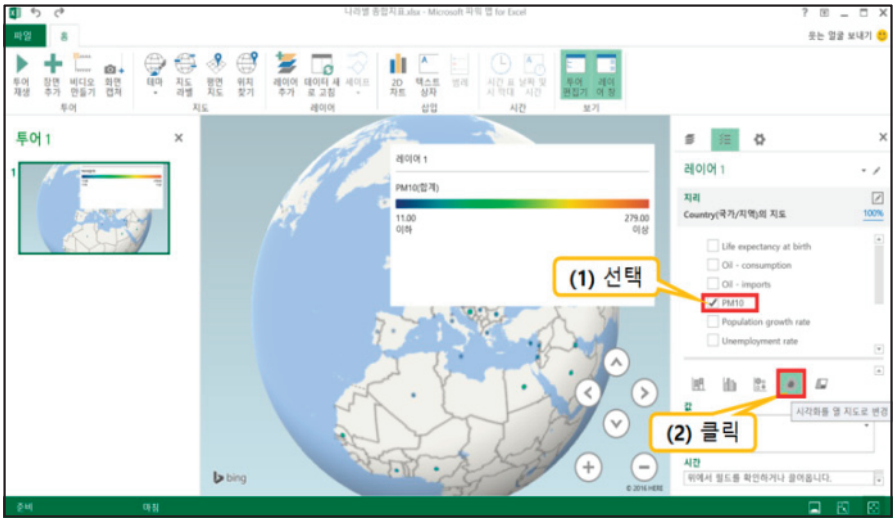


그림 10.118 파워맵 레이어 기능 3

(4) [레이어 옵션]에서 히트맵의 불투명도 및 영향권 반경을 조정할 수 있다. 공기오염도가 높을수록 히트맵이 주황색으로 표시된다. (색상 조절도 가능하다.)

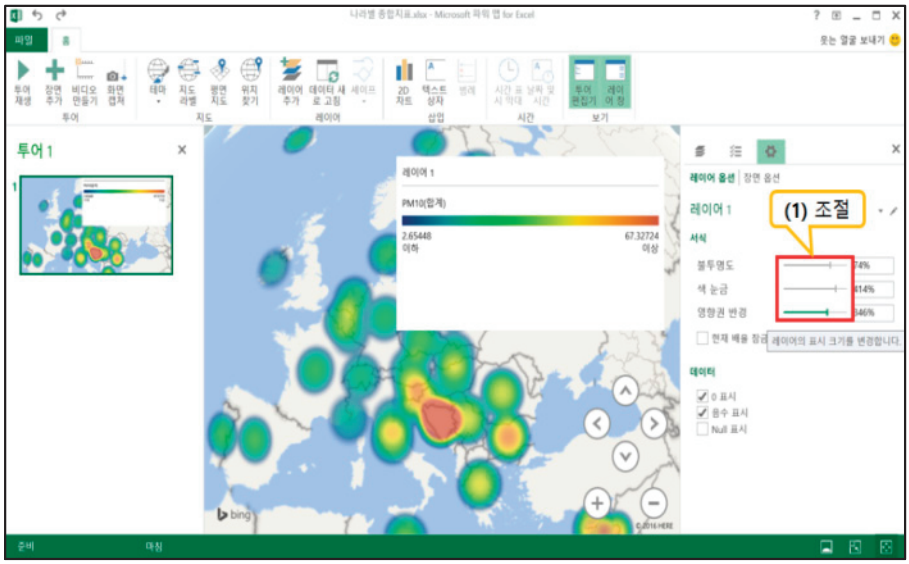


그림 10.119 파워맵 레이어 기능 4

(5) 화면 우측에서 [레이어 관리자] 버튼을 클릭하고, 레이어 추가 버튼을 클릭한다. [레이어2] 역시 [레이어1]과 동일하게 'Country'를 기준으로 지오코딩을 해준다. [레이어1] 위에 [레이어2]가 놓인다고 생각하면 이해하기가 쉽다.

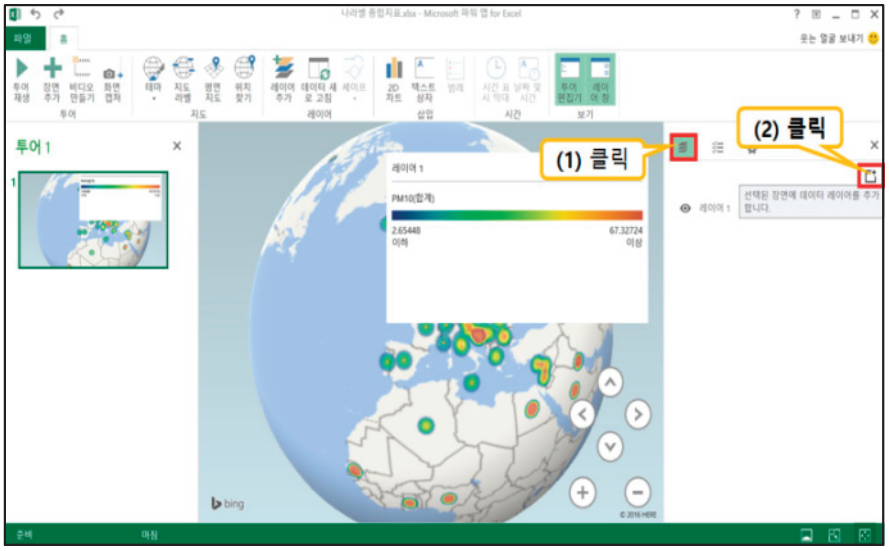


그림 10.120 파워맵 레이어 기능 5-1

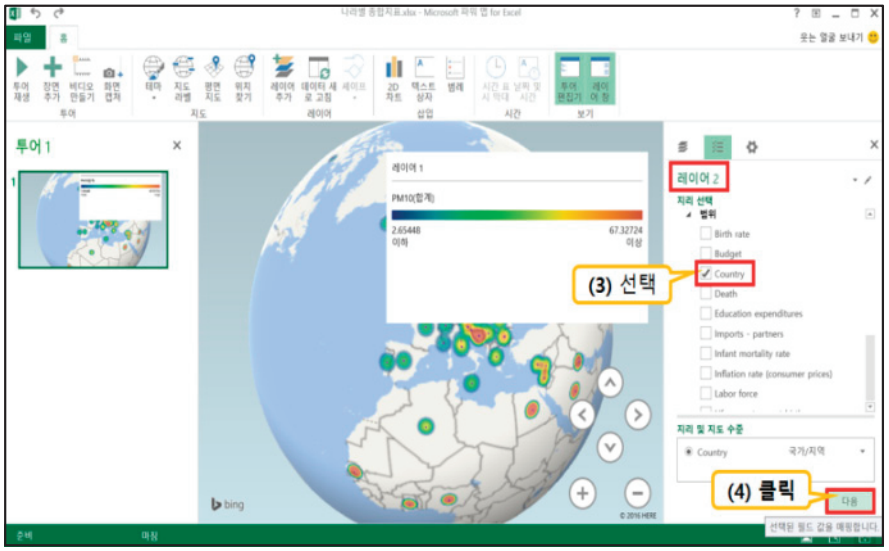


그림 10.121 파워맵 레이어 기능 5-2

(6) [레이어2] 필드에서는 'Infant mortality rate'을 선택하고 시각화 유형은 막대그래프로 하겠다. 그럼 히트맵[레이어1]위에 막대그래프[레이어2]가 생성된다.

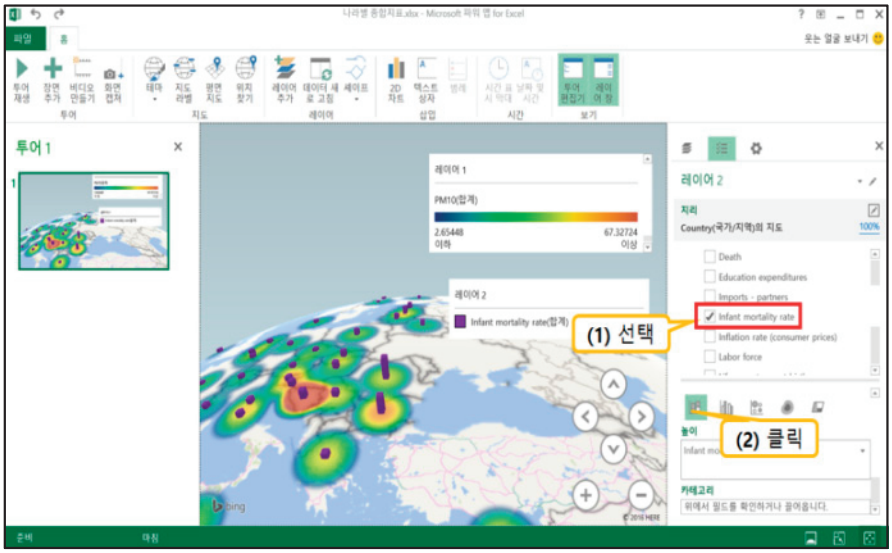


그림 10.122 파워맵 레이어 기능 6

(7) [홈] 탭 - [지도] 그룹 - [지도라벨]을 클릭하여 지구본에 지역명을 표시한다. 그리고 [레이어2] 필드에서 'Birth rate'을 추가로 선택하여 막대그래프를 만든다. 레이어를 통해 지구본위에 국가별 공기오염도, 유아사망률, 출생률이 함께 시각화되었다.

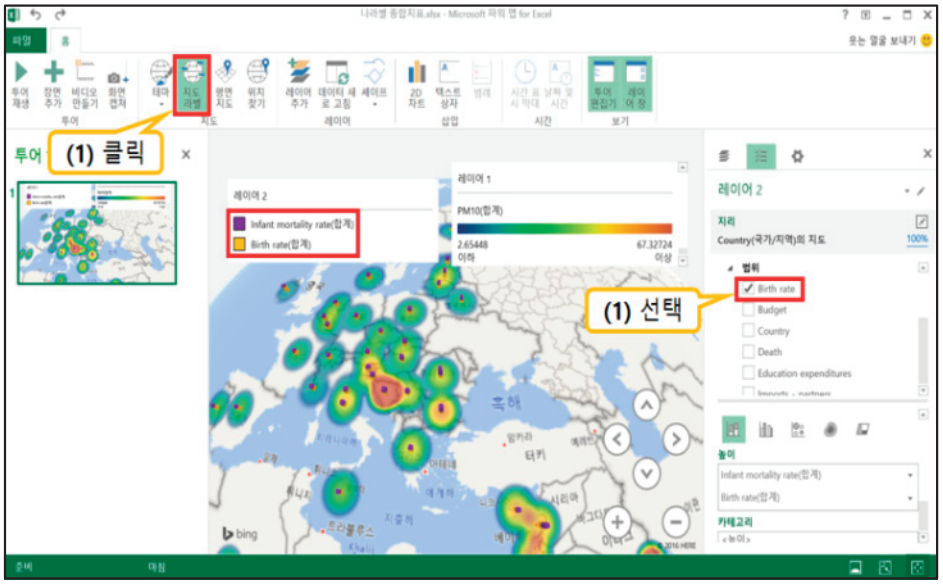


그림 10.123 파워맵 레이어 기능 7

(8) 추가로 [홈] 탭 - [지도] 그룹 - [평면지도]를 클릭해서 평면지도로 만들어보자.

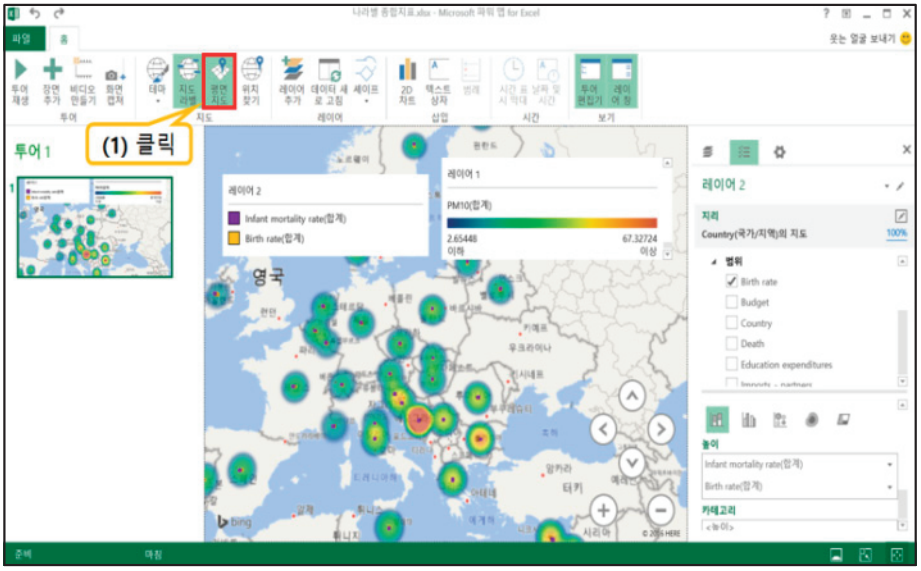


그림 10.124 파워맵 레이어 기능 8

참고문헌

Anders and Rritt Wallgren(2007), Register-based Statistics Administrative Data for statistical Purposes, Statistics Sweden, SwedenBergek, A., Jacobsson, S., Hekkert, M. and Smith, K.(2010), Functionality of Innovation Systems as a Rationale for and Guide to Innovation Policy, in Smits, R. E., Kuhlmann, S. and Shapira, P. (eds.), The Theory and Practice of Innovation Policy - An International Research Handbook, pp. 115~144, Edward Elgar.

Boyd, D. and Crawford, K.(2012), Critical Questions for Big Data - Provocations for a cultural, technological, and scholarly phenomenon, Information, Communication & Society, Vol. 15, No. 5, pp. 662~679.

Brackstone G. J.(1988), Statistical Issues of Administrative Data: Issues and Challenges, in Statistical Uses of Administrative Data -An International Symposium, organised by Statistics Canada, 23-25 November 1987 (Proceedings published by Statistics Canada, Ottawa, December 1988).

Cebr(2012), Data equity: Unlocking the value of big data, Report for SAS, SAS.

David, Haltiwanger and Schuh(1996), Job Creation and Destruction.

Global Pulse(2012), Big Data for Development: Challenges & Opportunities, Global Pulse.

Hey, T., Tansley, S., & Tolle, K. (2009), Jim Gray on eScience: A transformed scientific method, the fourth paradigm, Redmond, WA: Microsoft Research.

IBM(2012), 빅데이터의 현실적인 활용(기업이 불확실한 데이터에서 가치를 창출해 내는 방법), IBM 글로벌 비즈니스 서비스 요약보고서.

KISTI(2013), Market Report.

McKinsey Global Institute (2011), Big data: the next frontier for innovation, competition and productivity, McKinsey & Company.

NESSI(2012), Big Data - A New World of Opportunities, NESSI White Paper, NESSI.

OECD(2013), Exploring Data-Driven Innovation as a New Source of Growth: Mapping the Policy Issues Raised by 'Big Data, OECD Digital Economy

Papers, No. 222, OECD Publishing.
OECD(2013a), New Sources of Growth: Knowledge-based Capital - key anlaysis and policy conclusions, [DSTI/IND(2013)7].
OECD(2013b), New Sources of Growth: Knowledge-based Capital Outline of the Work in 2013~14, [DSTI/ICCP(2013)1].
Statistics Canada(2013), Quality Guideline.
Statistics Canada(2013), Statistical Data and Metadata eXchange, SDMX.
Statistics New Zealand(2010), Linked Employer-Employee Data.
UN(2000), Terminology of Statistical Metadata.
UN(2011), A Handbook of Principles and Practices.
Yiu, C.(2012), The Big Data Opportunity - Making government faster, smarter and more personal, Policy Exchange.
경상북도(2013), 경상북도 공공분야 빅데이터 활용방안 연구.
공개SW포털(2013), 한국고용정보원 - 공개SW 기반으로 고용정보 시스템 서비스 개선.
교육과학기술부, 행정안전부, 지식경제부, 방송통신위원회, 국가과학기술위원회(2012), 스마트 국가 구현을 위한 빅데이터 마스터플랜.
국무조정실, 행정자치부, 방송통신위원회, 금융위원회, 미래창조과학부, 보건복지부(2016), 개인정보 비식별 조치 가이드라인.
국민건강보험공단(2013), 건강보험 빅데이터 현황 및 활용계획.
국토교통부(2013), 국가공간정보통합체계.
김기환(2012), 공공부분 빅데이터의 활용성과 위험성, 정책분석평가학회보, 23권 2호, pp. 1-27.
김규정(1997), 행정학원론, 법문사, pp. 12.
김성웅(2013), OECD의 빅데이터 관련 논의 동향, 정보통신정책연구원.
김성태 (2012), 新가치창출 엔진, 빅데이터의 새로운 가능성과 대응 전략, IT & Future Strategy, 한국정보화진흥원.
김한나(2012), 빅데이터의 동향 및 시사점, 방송통신정책, 제24권 제19호.
류광택, 이용건, 조성배(2012), 미래 전자정부 모델에 대한 탐색적 연구: 지속가능한 발전을 위한 전자벙부 3.0의 가치와 개념을 주심으로, 한국정책학회 하계학술대회 발표논문집, pp. 319-340.
문화체육관광부(2013), 혼자, 일상에서 소소한 행복을 느끼다 - 36억여 건의

빅데이터 분석을 통해 본 2013 국민 인식 변화.
미래창조과학부(2013), 2012년도 기술영향평가 결과(안) - 빅데이터(Big Data) 분석기술과 활용, 국가과학기술심의회.
박동서(1988), 한국행정론, 법문사, pp. 68.
박원준(2013), 빅데이터에 활용에 대한 기대와 우려, Journal of Communications & Radio Spectrum, pp. 28-45.
박진서, 박경석, 이용호(2013), 새로운 기술패러다임으로서 빅데이터: 쟁점과 과제. 과학기술정책 통권 192호, pp. 17-30.
배동민, 박현수, 오기환(2013), 빅데이터 동향 및 시사점, kt경제경영연구소.
보건복지부(2013), 제2차 한국인체자원은행사업 추진계획 2013.
부산연제경찰서(2012), 지리적 프로파일링 시스템 활용한 범죄다발지(hot-spot) 형사활동.
성지은, 송위진(2010), 탈추격형 혁신과 통합적 혁신정책, STEPI Working Paper, 3호, 과학기술정책연구원.
송성헌(2015), 행정자료를 이용한 통계작성 방법론, 통계교육원.
안석모(2013), 정보인권 보고서 발간 배경 및 의의, 정보인권 보고서 발간 기념 공동학술대회 자료집, pp. 3-13.
안전행정부(2013), 관세청-국세청, 정부 3.0 우수사례 대상 공동 수상. 보도자료.
안전행정부(2013), 안전한 사회, 유능한 정부, 성숙한 자치로 행복한 대한민국을 만들겠습니다. 2013년 업무보고.
안전행정부(2013), 모든 재난 정보, 한눈에 확인하고 즉시 대응.
유경준(2016), 빅데이터 시대에 즈음한 경제총조사의 진화, 서울신문, 2016년 7월 25일.
유종해, 이덕로(2015), 현대행정학, pp. 7.
유지연(2012), 세계경제포럼(WEF)을 통해 본 빅데이터 논의 동향과 함의, 정보통신정책 동향, 24권 4호, 통권 526, pp. 55-62.
윤미영(2013), 주요국의 빅데이터 추진전략 분석 및 시사점, 한국정보화진흥원.
윤윤규, 고영우(2009), 일자리 창출과 소멸에 관한 연구.
이만재(2011), 빅데이터와 공공 데이터 활용, Internet and Information Security, 제2권, 제2호, pp. 47-64.
이재호(2013), 정부3.0 실현을 위한 빅데이터 활용방안, 한국행정연구원.

이지영(2015), 빅데이터의 국가통계 활용을 위한 기초연구, 2015년 상반기 연구보고서, 통계개발원, pp. 49-123.
전승우(2012), 빅데이터에 대한 기대와 현실, LG Business Insight, LG경제연구원.
田中二郎(1992), 行政法總論, pp. 22.
정부통합전산센터(2016), 빅데이터 공통기반 가이드북 1권.
채승병(2011). 정보홍수 속에서 金脈 찾기: 빅데이터(Big Data) 분석과 활용, SERI 경영노트.
최우정, 이종국, 여운광(2013), 국립재난안전연구원 재난안전 스마트 빅 보드 개발 착수, 재난관리 3.0을 선도하는 새로운 플랫폼. 물과 미래. 46권 9호, pp. 73-78.
쿠키뉴스(2013), 개인유전정보로 맞춤형 의료서비스-정부, 유전정보 빅데이터 체제 구축 추진 한국.
통계청(2013), 국가통계 선진화를 위한 중기(2011~15) 행정데이터 활용전략(개정).
통계청(2014), 보도자료 2014년 1월 7일.
통계청(2014), 보도자료 2014년 2월 17일.
통계청(2015), 보도자료 2015년 12월 22일.
통계청(2016), 2016년 주요업무 추진계획.
통계청(2016), 보도자료 2015년 1월 22일.
통계청(2016), 보도자료 2016년 6월 27일.
통계청(2016), 보도자료, 2016년 7월 24일.
통계청(2014), 임금근로일자리 행정통계 개발보고서.
통계청(2010), 행정자료 활용 외국사례집.
한국정보화진흥원(2013), 새로운 미래를 여는 빅데이터 시대.
한국정보화진흥원(2013), 빅데이터 성과물 한자리에, 빅데이터 페어 2013.
한국정보화진흥원(2014), 빅데이터 커리큘럼 참조 모델 Ver 1.0.
한국정보화진흥원(2014), 창조경제실현을 위한 2013 빅데이터 국내 사례집.
한국정보화진흥원(2015), 2015 국가정보화백서.
한국정보화진흥원(2016), 국가정보보호백서.
한국정보화진흥원(2016), 빅데이터 동향과 이슈, 제21호.
행정자치부(2015), 2015 행정자치통계연보.

홍정선(2015), 행정법원론(상), 박영사, pp. 4-6.
BigData Transfer Insight, http://www.bdtinsights.com/kr/
K-ICT 빅데이터센터, https://kbig.kr
KOSIS, http://www.kosis.kr
백두권, http://www.diabetes.or.kr/new_workshop/201403/ab3.html
통계청, http://www.kostat.go.kr
한국정보화진흥원, http://www.nia.or.kr

부록 1: 행정자료별 활용 현황(172종, ‘15년말 기준)

번호	행정자료명	소관기관	대체, 검증/보완
1	AIDS 등록자료	질병관리본부	검증/보완
2	TAC 업서버 자료(참홍어 및 붉은대게)	수산자원관리사업단	검증/보완
3	가정용 및 차량용 연료판매액	한국석유공사	대체
4	가족관계등록부	대법원	대체
5	가족등록신고(전산시스템)	대법원	대체
6	간접상품설정잔액	금융투자협회	대체
7	간편장부 대상자 소득금액	국세청	대체
8	감사보고서/사업보고서 작성대상 기업	금융감독원	검증/보완
9	개별공시지가	국토교통부	대체
10	개인소득신고자료	국세청	검증/보완
11	건강보험 징수액(장기요양급여지급액 포함)	국민건강보험공단	대체
12	건강보험 총진료비	건강보험심사평가원	대체
13	건강보험급여비지출	국민건강보험공단	대체
14	건강보험자료	국민건강보험공단	대체
15	건설폐기물수집운반 및 처리업	한국환경공단	대체
16	건축물대장	국토교통부	대체
17	결산_건설관련협회	건설관련협회	대체
18	결핵 등록자료	질병관리본부	검증/보완
19	고용보험 지급액	한국고용정보원	대체
20	고용보험(회사부담금,개인부담금)	고용노동부	대체
21	고용보험자료	고용노동부	대체
22	골프장 내장객수	한국골프장경영협회	대체
23	공동사업자자료	국세청	대체
24	공동주택 공시지가 자료	국토교통부	대체
25	공립학교 회계(초중고)	교육부	대체
26	공사계약서	조달청	대체
27	공사공동도급	조달청	대체
28	공사기본정보	국토교통부	대체
29	공장등록자료	산업통상자원부	검증/보완
30	관광호텔 카지노수입	한국카지노업관광협회	대체
31	관상수시업상황개인별조사서	산림청	검증/보완
32	광물탐사비용	한국광물자원공사	대체
33	교육비 특별회계(초중고)	교육부	대체
34	교정시설 수용자 및 소년원원생명부	법무부	대체
35	교통사고 자료	경찰청	검증/보완
36	국공립대학 결산자료	기획재정부	대체
37	국립대 기성회계	교육부	대체

번호	행정자료명	소관기관	대체, 검증/보완
38	국립대 일반회계, 기타특별회계	기획재정부	대체
39	국민연금자료	국민연금공단	대체
40	국유림(대부, 사용허가)신청서	산림청	검증/보완
41	국적취득자 명부	법무부	대체
42	군내사망자료	국방부 육해공본부	검증/보완
43	군인(사병)명부	국방부	대체
44	궁중입장수입	문화재청	대체
45	근로소득지급명세서	국세청	대체
46	금속원료재생처리량	한국환경공단	대체
47	내수면어가명부	해양수산부	검증/보완
48	농업경영체등록자료(농업법인)	농림축산식품부	검증/보완
49	농업경영체등록자료(농업인)	농림축산식품부	대체
50	농지원부	농림축산식품부	대체
51	대구획 경지정리내역	농림축산식품부	검증/보완
52	대학생기숙사명부	대학교	대체
53	독림가명부	산림청	검증/보완
54	등록외국인 명부 및 국내거소신고 외국국적동포 명부 자료	법무부	대체
55	면세점 판매액	관세청	대체
56	명부_건설관련협회	건설관련협회	대체
57	물류창고업등록자료	국토교통부	대체
58	미곡 정부수매량 및 수매액	농림축산식품부	대체
59	배구경기장관중수	한구배구연맹	대체
60	벌과금 자료	법무부	대체
61	법인등기자료	대법원	대체
62	법인세자료	국세청	대체
63	법인합병신고서	국세청	대체
64	법정전염병 등록자료	질병관리본부	검증/보완
65	변사자 부검자료	국립과학수사연구소	검증/보완
66	변사자 자료	경찰청/해양경비안전본부	검증/보완
67	보험사별 영업수익(경과보험료)	금융감독원	대체
68	복권판매액	복권위원회	대체
69	부가가치세자료	국세청	대체
70	부동산거래액	국토교통부	대체
71	불법체류자(단기체류자격)명부	법무부	대체
72	사립학교 법인회계(초중고)	교육부	대체
73	사망신고자료	대법원/지자체	대체
74	사망자 수진기록	국민건강보험공단	검증/보완
75	사업소득신고자료_간편장부	국세청	대체

번호	행정자료명	소관기관	대체, 검증/보완
76	사업소득신고자료_복식기장	국세청	대체
77	사업소득지급명세서	국세청	검증/보완
78	사업자등록자료	국세청	대체
79	사업장단위 과세적용사업장	국세청	대체
80	사업장별 부가가치세	국세청	대체
81	사업장현황신고서	국세청	대체
82	사회시설명부	보건복지부	대체
83	산림사업법인	산림청	검증/보완
84	산림조합원명부	산림청	검증/보완
85	산재보험 지급액(진료비지급액 포함)	근로복지공단	대체
86	산재보험자료	고용노동부	검증/보완
87	산재사망자자료(근로복지공단)	근로복지공단	검증/보완
88	산재사망자자료(산업안전보건공단)	한국산업안전보건공단	검증/보완
89	새주소모집단	행정자치부	대체
90	석유가스탐사비용	한국석유공사	대체
91	소년원 원생 명부	법무부	대체
92	손해보험료 및 보험금	보험개발원	대체
93	쇠고기 이력제 자료	농림축산식품부	검증/보완
94	수급인정보	국토교통부	검증/보완
95	수입배당금액명세서	국세청	대체
96	수입차 판매액	한국수입차협회	대체
97	수출입 통계자료	관세청	대체
98	수출자료	관세청	검증/보완
99	수협위탁판매자료	수협중앙회	검증/보완
100	수협조합원명부	수협중앙회	검증/보완
101	시도별 가스 공급량	한국도시가스협회	대체
102	시도별 대두 공급현황	한국연식품공업협동조합	대체
103	시도별 레미콘 생산, 출하량	한국레미콘공업협동조합연합회	대체
104	실적_건설관련협회	건설관련협회	대체
105	아이사랑카드 결제금액	보건복지부	대체
106	암 등록 자료	국립암센터	검증/보완
107	어선원부	해양수산부	검증/보완
108	어업경영체등록자료	해양수산부	검증/보완
109	어업인허가대장	해양수산부	검증/보완
110	영화관람수입	영화진흥위원회	대체
111	예산 집행자료	국민연금공단	대체
112	예산 집행자료	한국농어촌공사	대체
113	외국인등록부	법무부	대체
114	외국적동포거소신고자 명부	법무부	대체

번호	행정자료명	소관기관	대체, 검증/보완
115	운수수입(철도, 화물 , 전철)	한국철도공사	대체
116	운전면허등록자수	경찰청	대체
117	원양어업생산량	해양수산부	검증/보완
118	위탁판매량	수산업협동조합	검증/보완
119	은행 및 대출기관 대출금 잔액	한국은행	대체
120	은행,신협,투신사, 금고 결산서	금융감독원	대체
121	응급의료센터내원정보	중앙응급의료센터	검증/보완
122	의료보장(회사부담금,개인부담금)	국민건강보험공단	대체
123	의무해양경찰명부	해양경찰청	대체
124	의무경찰 명부	경찰청	대체
125	의무소방대원명부	중앙소방본부	대체
126	이익잉여금처분계산서	국세청	대체
127	이자배당소득자료(가구)	국세청	대체
128	인삼 생산량 및 생산액	농림축산식품부	대체
129	일반_건설관련협회	건설관련협회	대체
130	일용근로소득지급명세서	국세청	검증/보완
131	임목별벌채허가및신고대장	산림청	검증/보완
132	임산물굴취채취허가대장	산림청	검증/보완
133	임업후계자명부	산림청	검증/보완
134	자동차세자료	행정자치부	검증/보완
135	재산세자료(건축물)	행정자치부	검증/보완
136	재산세자료(주택)	행정자치부	대체
137	재산세자료(토지)	행정자치부	검증/보완
138	재외국민거소 신고자 명부	법무부	대체
139	재판별 최종판결 건수	대법원	대체
140	전기시설명부	한국전력	대체
141	전월세거래금액	국토교통부	대체
142	전자상거래 물품반입자료(목록통관,간이,일반신고자료)	관세청	대체
143	정당의 활동개황 및 회계보고서	중앙선거관리위원회	대체
144	조경수협회명부	산림청	검증/보완
145	주거용부동산관리비	국토교통부	대체
146	주민등록 전입자료	행정자치부	대체
147	주민등록자료	행정자치부	대체
148	주택공시가격자료	국토교통부	대체
149	주택임대차확정일자자료	한국감정원	대체
150	중앙정부 결산서	기획재정부	대체
151	중앙정부 예산 집행자료	기획재정부	대체
152	증권거래대금	한국거래소	대체
153	지방정부 결산서	행정자치부	대체

번호	행정자료명	소관기관	대체, 검증/보완
154	지방정부 예산 집행자료	행정자치부	대체
155	지적재산권 보유건수	특허청	대체
156	지정외 폐기물수집운반 및 처리량	한국환경공단	대체
157	지정폐기물 수집운반 및 처리량	한국환경공단	대체
158	직장가입자 보험료(공무원, 사립교원)	국민건강보험공단	대체
159	직장가입자(근무처, 근무내역) 변동신고서	국민건강보험공단	대체
160	초중고 교직원수	한국교원학술정보원	대체
161	축구경기장관중수	한국프로축구	대체
162	축산업등록명부	농림축산식품부	대체
163	출입국심사자료	법무부	대체
164	취득세자료(차량)	행정자치부	검증/보완
165	토지(임야)대장/지적전산자료	국토교통부	대체
166	통관기준 수출입	관세청	검증/보완
167	특수관계자간거래명세서	국세청	대체
168	하수급인정보	국토교통부	대체
169	학적부	대학교	대체
170	해외주재 공관원명부	외교부	대체
171	행복e음자료	보건복지부	검증/보완
172	휘발유, 경유, 등유, 자동차용 LPG 품목 가격	한국석유공사	대체

부록 2: 통계조사별 대체 및 검증·보완 현황(40종, ‘15년말 기준)

통계명	전부대체	일부대체	검증/보완
가계금융복지조사			○
가계동향조사			○
가축동향조사			○
개인별주택소유통계	○		
건설업조사		○	○
경제총조사		○	○
광업제조업동향조사		○	○
광업제조업조사		○	○
국내인구이동통계	○		
국민대차대조표		○	
국제인구이동통계	○		
귀농귀촌인통계	○		
기업생멸행정통계	○		○
기업활동조사		○	○
농가경제조사			○
농림어업조사			○
농림어업총조사			○
농어업법인조사		○	○
농업면적조사			○
대기배출계정			○
도소매업조사		○	○
사망원인통계		○	○
서비스업동향조사		○	
서비스업조사		○	○
설비투자지수		○	
소비자물가조사		○	
어가경제조사			○
어업생산동향조사			○
영리법인기업체행정통계	○		
온라인쇼핑동향조사		○	
외국인고용조사			○
운수업조사		○	○
인구동향조사			○
인구주택총조사(2)		○	○
임금근로일자리행정통계	○		
전국사업체조사		○	○
전문과학기술서비스업조사		○	○
전산업생산지수		○	
지역소득		○	○
합 계	7	21	12

부록 3: 통계조사별 항목대체 현황(28종, ‘15년말 기준)

번호	통계명	대체항목	행정자료명
1	귀농귀촌인통계	귀농인의 가족관계, 영농경험, 영농품목 등	농지원부, 농업경영체등록자료 등
2	기업생멸행정통계	법인기본사항, 합병 현황 등	법인세, 법인합병신고서 등
3	영리법인기업체행정통계	산업별, 기업체규모별, 사업장별 현황 및 재무현황 등	사업자등록, 법인세 자료 등
4	임금근로일자리행정통계	기업체형태별, 산업분류별 일자리 등	4대보험, 근로소득명세서, 부가세자료 등
5	개인별주택소유통계	주택유형별 주택소유현황 등	재산세자료, 건축물 대장 등
6	건설업조사	매출액, 자본금, 건설공사매출액 등	법인세, 부가세, 공사기본정보 등
7	경제총조사	매출액, 자산, 자본, 업종 등	법인세, 사업자등록자료 등
8	광업제조업동향조사	생산량, 출하량, 재고량 등	시도별 가스, 대두, 레미콘 생산량 등
9	광업제조업조사	매출액(출하액), 영업비용 등	법인세, 부가가치세 등
10	국민대차대조표	재고자산 및 유형자산 등	법인세자료, 토지(임야)대장 / 지적전산자료 등
11	국내인구이동통계	시도, 시군구별 전입, 전출	주민등록 전입자료
12	국제인구이동통계	장기이동자	출입국신고 전산자료
13	기업활동조사	매출액, 사업비용 등	법인세, 부가가치세 자료 등
14	농어업법인조사	조직형태, 경영실태 등	법인등기, 법인세 자료 등
15	도소매업조사	매출액, 비용, 종사자 등	법인세, 사업소득신고자료 등
16	사망원인통계	사망원인통계 기본자료	사망신고자료
17	서비스업동향조사	가정용 및 차량용 연료 소매, 자동차 판매 등	연료판매액, 건강보험 징수액 등
18	서비스업조사	매출액, 비용, 종사자 등	법인세, 사업소득신고자료 등
19	설비투자지수	설비투자 기준생산액 산정	부가가치세, 수출입통계 자료
20	소비자물가조사	소비자 판매 가격	휘발유, 경유, 등유, 자동차용 LPG 품목 가격
21	운수업조사	매출액, 영업비용 등	법인세, 사업자등록자료 등
22	온라인쇼핑동향조사	수입통계 부분	전자상거래 물품반입자료(목록통관, 간이, 일반신고자료)
23	인구총조사	인구, 가구, 주택 항목보완	가족관계등록부, 건축물대장, 주민등록자료 등
24	주택총조사		
25	전국사업체조사	매출액, 사업장 정보 등	사업자등록, 부가가치세 자료 등
26	전문과학기술서비스업조사	매출액, 사업비용, 종사자 등	법인세, 부가가치세 자료 등
27	지역소득	사업비용, 자산 및 부채 등	간편장부대상자 소득금액, 건강보험급여비지출 등
28	전산업생산지수	공공행정부문 지수작성	중앙·지방정부 예산집행자료 등

부록 4: ’16년 행정자료 신규수요 현황

보유기관	행정자료	활용통계
국세청	근로소득원천징수명세서*/근로소득지급명세서*	퇴직연금가입통계
	법인세*	
	부가가치세*	
	사업자등록자료*	
	연금가입자 퇴직소득 지급내역 명세서	
	연금계좌원천징수영수증 및 퇴직소득세액 계산부표	
	퇴직소득원천징수영수증/지급명세서	
	퇴직연금 부담금 납입확인서/조정명세서	
	간편장부소득금액계산서 서식자료*	지역소득
	추계소득금액계산서 서식자료	
	단순경비율 적용대상자용 과세표준확정신고 및 납부계산서 서식자료	
	복식부기의무자 대차대조표, 손익계산서, 원가명세서	
	이자, 배당 지급명세서 서식자료	
	일용근로*·근로*·연금·퇴직소득 지급명세서 서식자료	
	사업자등록자료*	임금근로자일자리 행정통계
	사업자소득지급명세서_연말정산/원천징수*	
	일용근로소득지급명세서*	
고용노동부	육아휴직자료(한국고용정보원)	일·가정 양립지표
	고용보험_사업자등록번호*	퇴직연금가입통계
	퇴직연금규약신고서	
	퇴직연금가입사업장자료(금융감독원 공동)	
국토교통부	주택공시가격자료**	가계금융복지조사
	자동차등록 및 이륜차 신고현황	대기배출계정
	임대사업자자료	개인별주택소유통계
	주택공시가격자료**	
보건복지부	기초연금	농가경제조사
		임금근로자일자리 행정통계
행정자치부	주민등록(초본)자료**	가계금융복지조사
		가계동향조사
		농·어가경제조사
		귀농귀촌인통계
		임금근로자일자리 행정통계
법무부	가족관계등록부**	일·가정 양립지표
	외국인등록자료**	임금근로자일자리 행정통계

보유기관	행정자료	활용통계
해양수산부	어업인허가대장(면허, 허가, 신고, 근해*/연안*/구획*)	귀어인통계
교육부	학적부**	임금근로자일자리 행정통계
		대졸자 취업 및 직장변동 통계
	(국공·사립)유치원·초·중·고·대학 알리미 등록자료(한국교육학술정보원	전국사업체조사
		경제총조사
공무원연금 관리공단	현직/퇴직자 공무원연금자료	임금근로자일자리 행정통계
국민건강보험공단	건강보험 사업장 자료*_사업자등록번호	임금근로자일자리 행정통계
국민연금관리공단	국민연금자료*_종사자·사업자부문	농가경제조사
		퇴직연금가입통계
	국민연금 상실자 자료	임금근로자일자리 행정통계
사학연금관리공단	현직/퇴직자 사학연금자료	임금근로자일자리 행정통계
농협중앙회	거래자별 매출 상세내역	농가경제조사
	출하처별 출하주 실적	
	하나로마트 회원 영수증 상세내역	
수협중앙회	수협조합원/어촌계원 명부	귀어인통계
한국공정거래 조정원	프랜차이즈 가맹본부 등록자료	전국사업체조사
		경제총조사
한국관광호텔업 협회	관광호텔업 등록자료	전국사업체조사
		경제총조사

* 기존 입수자료 중 제공범위 확대 요청　　** 등록센서스과 기 입수자료

부록 5: 완전대체 항목

통계조사명	행정자료명	보유기관	활용분야
귀농 귀촌인 통계	농업경영체등록자료	농림축산 식품부	성명, 성별, 연령, 배우자동반, 영농기간, 경영형태, 재배면적, 영농품목, 임차비율, 축종
	농지원부		성명, 성별, 연령, 배우자동반, 경지소유, 임대차여부, 영농기간, 영농품목
	축산업등록명부		지역별 이동, 축종 및 마리수
	직장가입자 변동신고서	건강보험공단	근무처, 근무내역 등
	국민연금자료	국민연금공단	
	근로소득지급명세서	국세청	
기업생멸 행정통계	법인세자료	국세청	매출액, 사업비용, 자산 및 부채, 재고정보
	부가가치세자료		산업분류별 매출액 비율 파악
	사업자등록자료		법인기본사항, 사업장 정보
	법인합병신고서		합병 현황
	근로소득지급명세서		종사자수 파악
	사업장단위 과세적용사업장		업종, 사업장 정보
	사업장별 부가가치세		산업분류별 매출액 비율
영리법인 기업체 행정통계	건강보험자료	건강보험공단	종사자수 파악
	고용보험자료	고용노동부	종사자수 파악
	국민연금자료	국민연금공단	종사자수 파악
	법인세자료	국세청	기업체 수, 경영실적
	부가가치세자료		매출액
	사업자등록자료		기업체 수
	근로소득지급명세서		종사자수 파악
	산재보험자료	고용노동부	종사자수 파악
임금근로 일자리 행정통계	건강보험자료	건강보험공단	근속기간, 성별, 연령별
	고용보험자료	고용노동부	근속기간, 성별, 연령별
	국민연금자료	국민연금공단	근속기간, 성별, 연령별
	근로소득지급명세서	국세청	근속기간, 성별, 연령별
	법인등기자료	대법원	기업형태, 산업분류, 자본금
	법인세자료	국세청	기업형태, 산업분류, 매출액
	부가가치세자료		기업형태, 산업분류, 매출액
	사업자등록자료		존속기간, 기업형태, 산업분류
	산재보험자료	고용노동부	근속기간, 성별, 연령별
개인별 주택소유통계	건축물대장	국토교통부	주택소유현황, 주택유형
	주택공시가격자료		주택소유현황, 주택유형
	재산세자료	행정자치부	주택소유현황, 주택유형
국내인구 이동통계	주민등록 전입자료	행정자치부	시도 시군구별 전입·전출자 파악
국제인구 이동통계	출입국신고 전산자료	법무부	장기이동자 파악

부록 6: 일부대체 항목

통계조사명	행정자료명	보유기관	활용분야
건설업조사	건설관련협회 일반, 명부, 실적, 결산	건설관련협회	명부보완 및 일반·실적·결산 항목
경제총조사	국공립대학 결산자료	기획재정부	교육서비스업 조사활용
	국립대 일반회계, 기타특별회계		
	중앙정부 결산서		정부부문 조사활용
	지방정부 결산서	행정자치부	정부부문 조사활용
	은행, 신협, 투신사, 금고 결산서	금융감독원	금융 및 보험업 조사활용
	공립학교 회계(초중고)	교육부	교육서비스업 조사활용
	교육비 특별회계(초중고)		
	국립대 기성회계		
	사립학교 법인회계(초중고)		
	법인세자료	국세청	사업장정보, 사업실적, 자산, 자본
	부가가치세자료		사업실적
	사업자등록자료		자산, 자본, 업종, 사업장 정보
	사업소득신고자료		개인사업자 매출액, 영업비용 등
광업제조업 동향조사	시도별 가스 공급량	한국도시가스협회	생산량, 출하량 및 지수 작성
	시도별 대두 공급현황	한국연식품공업협동조합	두부 생산량, 출하량 추정 및 지수 작성
	시도별 레미콘 생산, 출하량	한국레미콘공업협동조합연합회	생산량, 출하량 및 지수 작성
광업제조업조사	법인세자료	국세청	출하액(매출액), 사업실적(원재료비, 연료비 등 항목별 영업비용), 연간 급여액, 재고내역
국민대차대조표	개별공시지가	국토교통부	토지자산 추계
	광물탐사비용	한국광물자원공사	광물탐사 추계
	석유가스탐사비용	한국석유공사	광물탐사 추계
	토지(임야)대장/지적전산자료	국토교통부	토지추계
기업활동조사	법인세자료	국세청	매출액, 사업비용, 자산, 부채, 재고
	부가가치세자료		산업분류별 매출액 비율 파악
	지적재산권 보유 건수	특허청	지적재산권 보유 건수 관련
농어업법인조사	법인등기자료	대법원	모집단 작성, 조직형태, 법인설립연월
	법인세자료	국세청	명부보완, 매출액, 자산, 부채, 자본 등
도소매업조사	법인세자료	국세청	수입금액, 사업비용
	사업소득신고자료_복식		개인사업체 사업실적

통계조사명	행정자료명	보유기관	활용분야
	사업소득신고자료_간편		개인사업체 사업실적
사망원인통계	사망신고자료	대법원/지자체	사망원인통계 기본자료
서비스업 동향조사	가정용 및 차량용 연료판매액	한국석유공사	가정용 및 차량용 연료 소매업
	간접상품설정잔액	금융투자협회	증권 및 선물중개업
	건강보험 징수액 (장기요양급여지급액 포함)	건강보험공단	건강보험업, 사회복지서비스업
	건강보험 총진료비	건강보험심사평가원	보건업
	건설폐기물수집운반 및 처리업	한국환경공단	폐기물 수집운반 및 처리업
	고용보험 지급액	한국고용정보원	고용보험업
	골프장 내장객수	한국골프장경영협회	골프장운영업
	관광호텔 카지노수입	한국카지노업관광협회	갬블링 및 베팅업
	국공립대학 결산자료	기획재정부	고등교육기관(대학교)
	궁중입장수입	문화재청	사적지관리업
	금속원료재생처리량	한국환경공단	금속원료재생업
	배구경기장관중수	한구배구연맹	스포츠클럽, 경기장운영업
	보험사별 영업수익(경과보험료)	금융감독원	생명 및 손해보험업, 재보험업
	복권판매액	복권위원회	복권발행 및 판매업
	부동산거래액	국토교통부	부동산중개업
	산재보험 지급액 (진료비지급액 포함)	근로복지공단	산재보험업, 보건업
	수입차 판매액	한국수입차협회	자동차 판매업
	아이사랑카드 결제금액	보건복지부	보육시설운영업
	영화관람수입	영화진흥위원회	영화 및 비디오물 상영업
	운수수입(철도, 화물 , 전철)	한국철도공사	철도운송업 , 도시철도운송업
	운전면허등록자수	경찰청	운전학원
	은행 및 대출기관 대출금 잔액	한국은행	금융업 (일반은행, 저축기관, 투자기관)
	재판별 최종판결 건수	대법원	법무관련서비스업
	전월세거래금액	국토교통부	주거용부동산임대업
	주거용부동산관리비		주거용부동산관리업
	증권거래대금	한국거래소	증권 및 선물중개업
	지정외 폐기물수집운반 및 처리량	한국환경공단	폐기물 수집운반 및 처리업

통계조사명	행정자료명	보유기관	활용분야
	지정폐기물 수집운반 및 처리량		
	초중고 교직원수	한국교원 학술정보원	초등 및 중등 교육기관(초중고교)
	축구경기장관중수	한국프로축구	스포츠클럽, 경기장운영업
	면세점 판매액	관세청	대형마트 소매판매 항목
서비스업조사	법인세자료	국세청	수입금액, 사업비용
	사업소득신고자료_복식		개인사업체 사업실적
	사업소득신고자료_간편		개인사업체 사업실적
설비투자지수	수출입 통계자료	관세청	설비투자 추계시 대외거래부문
운수업조사	법인세자료	국세청	영업비용
	물류창고업등록자료	국토교통부	모집단 보완, 사업장 주소 등
	부가가치세자료	국세청	매출액
	사업자등록자료		사업형태, 사업장 변동
	사업장현황신고서		면세사업자 매출액 파악
인구주택총조사	가족관계등록부	대법원	인구 및 가구 항목
	국민연금 사업장가입자 자료	국민연금공단	가구 보완
	의무경찰 명부	경찰청	가구 보완
	건축물대장	국토교통부	주택 항목
	주택공시가격자료		주택 모집단 생성
	군인 명부	국방부	가구 보완
	기숙시설(대학) 이용자 명부	대학교	인구 보완(기숙시설거주)
	학적부		거주지 보완
	단기체류자(90일 이상) 명부	법무부	인구(외국인) 항목
	외국인등록부		인구(외국인) 항목
	외국적동포거소 신고자 명부		인구(외국인)보완
	재외국민거소 신고자 명부		인구보완
	출입국자료		인구 보완(해외거주자 분석)
	교정시설 수용자 명부		인구, 가구 보완
	소년원 원생 명부		인구, 가구 보완
	국적취득자 명부		인구, 가구 보완
	사회시설 및 이용자 명부	보건복지부	인구 보완(사회시설거주)
	새주소모집단	행정자치부	거처 모집단 생성
	주민등록자료		인구 및 가구 모집단 생성

통계조사명	행정자료명	보유기관	활용분야
	의무해양경찰명부	국민안전처	가구 항목 보완
	의무소방대원명부		가구 항목 보완
	고용보험 피보험자 자료	한국고용정보원	가구 보완
	전기시설(주택용) 자료	한국전력공사	가구 항목 보완
	해외주재 공관원 명부	외교부	인구 및 가구 항목 보완
전국사업체조사	법인세자료	국세청	사업장정보, 매출액, 자본금
	부가가치세자료		매출액
	사업자등록자료		업종, 사업장 정보
	사업장현황신고서		면세사업자 매출액 파악
전문과학기술 서비스업조사	법인세자료	국세청	사업장정보, 매출액, 사업비용, 유형자산
	사업소득신고자료_복식		개인사업체 사업실적
	사업소득신고자료_간편		개인사업체 사업실적
전산업생산지수	예산 집행자료	국민연금공단 한국농어촌공사	공공행정부문 지수작성
	중앙정부 예산 집행자료	기획재정부	
	지방정부 예산 집행자료	행정자치부	
지역소득	간편장부 대상자 소득금액	국세청	개인 영업잉여 추계
	사업소득신고자료_복식		준법인기업 소득추계
	법인세자료		산업부문 추계, 사업비용, 자산 및 부채
	수입배당금액명세서		배당금 추계
	이익잉여금처분계산서		중간배당액, 배당금 추계
	건강보험급여비지출	국민건강보험공단	정부부문추계
	의료보장(회사부담금, 개인부담금)		피용자 보수
	직장가입자 보험료(공무원, 사립교원)		사회부담금, 사회수혜금 추계
	고용보험(회사부담금, 개인부담금)	고용노동부	피용자보수파악
	공립학교 회계(초중고)	교육부	교육서비스업 부문 추계
	교육비 특별회계(초중고)		
	국립대 기성회계		
	사립학교 법인회계(초중고)		
	국립대 일반회계, 기타특별회계	기획재정부	교육서비스업 부문 추계
	중앙정부 결산서	기획재정부	공공행정, 국방 및 사회보장 부문 추계
	미곡 정부수매량 및 수매액	농림축산식품부	농림어업 생산액 추계
	인삼 생산량 및 생산액	농림축산식품부	농림어업 생산액 추계

통계조사명	행정자료명	보유기관	활용분야
	벌과금 자료	법무부	가계 및 법인 부분 벌과금 구성비 파악
	손해보험료 및 보험금	보험개발원	경상이전 추계
	은행, 신협, 투신사, 금고 결산서	금융감독원	금융 및 보험업 부문 추계
	정당의 활동개황 및 회계보고서	중앙선거 관리위원회	기타 서비스업 부문 추계
	지방정부결산서	행정자치부	정부부문추계
	이익잉여금처분계산서	국세청	중간배당액, 배당금 추계
소비자물가조사	휘발유, 경유, 등유, 자동차용 LPG 품목 가격	한국석유공사	소비자 판매 가격
온라인쇼핑동향조사	전자상거래 물품반입자료	관세청	수입통계 부분

부록 7: 기존 통계 검증·보완

통계조사명	행정자료명	보유기관	활용분야
가계금융 복지조사	건축물대장	국토교통부	가구별 부동산 소유현황
	전월세 확정일자 자료		주택 임대차 현황
	등기부등본	대법원	주택 소유현황 및 담보대출 현황
	행복e음 자료	보건복지부	기초생활수급금, 기초연금, 장애수당
	개인소득신고자료	국세청	가구별 유형별 소득금액 추정
	재산세자료	행정자치부	소유현황, 과표, 면적 등
	국민연금자료	국민연금공단	연금불입 여부, 소득 비교
	건강보험자료	국민건강 보험공단	비소비 지출 내 건강보험료
가계동향조사	개인소득신고자료	국세청	가구별 소득현황
가축동향조사	쇠고기 이력제 자료	농림축산식품부	연령별, 성별 마리수, 생산마리수
건설업조사*	공사계약서	조달청	발주자명, 도급종류, 계약연월, 착공연월 등
	공사공동도급		
	공사기본정보	국토교통부	발주자명, 도급종류, 계약연월, 착공연월 등
	수급인정보		
	하수급인정보		
	법인세자료	국세청	자본금, 매출액, 건설공사매출액, 건설비용
	부가가치세자료		납부부가가치세
경제총조사*	건강보험자료	국민건강 보험공단	업종, 종사자 파악
	고용보험자료	고용노동부	업종, 종사자 파악
	산재보험자료		업종, 종사자수, 급여액
	공장등록자료	산업통상자원부	사업장 정보
	국민연금자료	국민연금공단	종사자수 파악
	근로소득지급명세서	국세청	종사자수 파악
	사업소득지급명세서		
	사업장현황신고서		면세사업자 매출액
	일용근로소득지급명세서		일용근로자수 파악
	법인등기자료	대법원	업종, 조직형태
광업제조업 동향조사*	수출자료	관세청	제조업생산속보지수 작성
광업제조업조사*	사업자등록자료	국세청	사업장 정보(사업체명, 소재지 등), 자산, 자본, 업종
	사업소득신고자료_복식기장	국세청	개인사업자 출하액(매출액), 사업실적(원재료비, 연료비 등 항목별 영업비용), 연간 급여액 등

통계조사명	행정자료명	보유기관	활용분야
	부가가치세자료	국세청	출하액(매출액), 재고액
기업활동조사 *	근로소득지급명세서	국세청	종사자수 파악
	특수관계자간거래명세서		관계회사간 거래
	일용근로소득지급명세서		일용근로자수 파악
	사업자등록자료		법인기본사항, 사업장 정보
	법인등록자료	대법원	기업변동여부(합병, 존속, 파산 등)
	통관기준 수출입	관세청	국내외 기업 간 거래 현황 (직접 수출 및 수입액)
	감사보고서/사업보고서 작성대상 기업	금융감독원	감사보고서/사업보고서 작성대상 기업
기업생멸 행정통계*	일용근로소득지급명세서	국세청	일용근로자수 파악
	합병과세특례 신청서		합병현황
	분할과세특례 신청서		분할현황
	물적분할과세특례 신청서		물적분할현황
농가경제조사	농업경영체등록자료(농업인)	농림축산식품부	농가현황, 농작물, 가축사육시설·규모, 융자보조금, 친환경농산물
	농지원부		농가 일반현황, 소유농지현황, 임차지농지현황, 농지경작현황
	쇠고기이력제자료		가축종류, 성별 마리수, 모집단 보완
	축산업등록명부		가축종류, 사육두수, 모집단 보완
	기초연금	보건복지부	공적보조수입
농림어업조사	관상수시업상황 개인별조사서	산림청	경영주 성명, 나이, 재배면적, 모집단 보완
	국유림(대부, 사용허가) 신청서	산림청	경영주 성명, 나이, 성별, 산림면적, 모집단 보완
	농업경영체등록자료	농림축산식품부	모집단 보완
	농지원부	농림축산식품부	모집단 보완
	독림가명부	산림청	모집단 보완, 성명, 육림구분 등
	산림조합원명부	산림청	경영주 성명, 나이, 성별, 산림면적, 모집단 보완
	쇠고기이력제자료	농림축산식품부	모집단 보완
	수협위탁판매자료	수협중앙회	모집단 보완
	수협조합원명부	수협중앙회	경영주 성명, 경영형태, 어법종류, 모집단 보완
	어선원부	해양수산부	경영주 성명, 어선현왕, 모집단 보완
	어업인허가대장	해양수산부	어업종류, 산물, 모집단 보완
	임목별벌체허가 및 신고대장	산림청	경영주 성명, 산림면적, 경영형태, 모집단 보완
	임산물굴취채취허가대장	산림청	경영주 성명, 산림면적, 임산물종류, 모집단 보완

통계조사명	행정자료명	보유기관	활용분야
	임업후계자명부	산림청	모집단 보완
	조경수협회명부	산림청	경영주 성명, 나이, 모집단 보완
	축산업등록명부	농림축산식품부	모집단 보완
농림어업총조사	관상수시업상황 개인별조사서	산림청	경영주 성명, 나이, 성별, 재배면적
	국유림(대부, 사용허가)신청서	산림청	경영주 성명, 나이, 성별, 산림면적
	내수면어가명부	해양수산부	경영주 성명, 양식장면적, 주요어법
	농업경영체등록자료	농림축산식품부	모집단 보완
	농지원부	농림축산식품부	모집단 보완
	독림가명부	산림청	모집단 보완, 성명, 육림구분 등
	산림조합원명부	산림청	경영주 성명, 나이, 성별, 산림면적, 모집단 보완
	쇠고기이력제자료	농림축산식품부	모집단 보완
	수협위탁판매자료	수협중앙회	모집단 보완
	수협조합원명부	수협중앙회	경영주 성명, 경영형태, 어법종류, 모집단 보완
	어선원부	해양수산부	경영주 성명, 어선현왕, 모집단 보완
	어업인허가대장	해양수산부	어업종류, 산물, 모집단 보완
	임목별벌체허가 및 신고대장	산림청	경영주 성명, 산림면적, 경영형태, 모집단 보완
	임산물굴취채취허가대장	산림청	경영주 성명, 산림면적, 임산물종류, 모집단 보완
	임업후계자명부	산림청	모집단 보완
	조경수협회명부	산림청	경영주 성명, 나이, 모집단 보완
	어업경영체	해양수산부	농림어업총조사 가구명부 작성
	축산업등록명부	농림축산식품부	모집단 보완
농어업법인조사*	농업경영체등록자료 (농업법인)	농림축산식품부	모집단 보완
	산림사업법인	산림청	산림사업법인(개발)
	부가가치세자료	국세청	매출액
	근로소득지급명세서		종사자수 파악
	사업자등록자료		기본사항 및 조직형태, 사업장 변동
	일용근로소득지급명세서		일용근로자수 파악
농업면적조사	대구획 경지정리내역	농림축산식품부	증감사유별 경지면적
도소매업조사*	부가가치세자료	국세청	본지사 임퓨테이션 지사별 매출액
	근로소득지급명세서		종사자수 파악
	사업소득지급명세서		종사자수 파악
	사업자등록자료		사업체명, 소재지, 사업자등록번호 등

통계조사명	행정자료명	보유기관	활용분야
	사업장현황신고서		면세사업자 매출액
	일용근로소득지급명세서		일용근로자수 파악
사망원인통계 *	의료기관정보	건강보험 심사평가원	사망원인 보완
	교통사고 사망 자료	경찰청	사망원인 보완
	변사자 자료		
	공군내 사망자료	국방부(공군)	사망원인 보완, 사망자직업 보완
	변사자 부검자료	국립과학 수사연구소	사망원인 보완
	암 등록 자료	국립암센터	사망원인 보완
	건강보험자료	국민건강보험공단	사망원인 보완
	영아사망 수진자료		
	의료급여자료		
	군내사망자료	국방부	사망원인 보완, 사망자직업 보완
	월간 산재보험자료	근로복지공단	사망원인 보완
	AIDS 사망 자료	보건복지부 (질병관리본부)	사망원인 보완
	감염병 사망 자료		사망원인 보완
	결핵 사망 자료		사망원인 보완
	변사자 자료	국민안전처 (해양경비안전본부)	사망원인 보완
	육군내 사망자료	국방부(육군)	사망원인 보완, 사망자직업 보완
	응급의료센터내원정보	중앙응급의료센터	사망원인 보완
	무연고 사망 공고	지방자치단체	사망원인 보완
	화장신고자료	지방자치단체 (시설관리공단)	사망원인 보완, 영아사망 보완
	연간 산업재해자료	한국산업안전 보건공단	사망원인 보완
	해군내 사망자료	국방부(해군)	사망원인 보완, 사망자직업 보완
	사망자 주민등록번호자료	행정자치부	사망원인 보완
서비스업조사 *	부가가치세자료	국세청	본지사 임퓨테이션 지사별 매출액
	근로소득지급명세서		종사자수 파악
	사업소득지급명세서		종사자수 파악
	사업자등록자료		사업체명, 소재지, 사업자등록번호, 법인등록번호, 사업의 종류
	사업장현황신고서		면세사업자 매출액
	일용근로소득지급명세서		일용근로자수 파악

통계조사명	행정자료명	보유기관	활용분야
어가경제조사	개별공시지가	국토교통부	토지 자산 평가
	건축물대장	국토교통부	건축물 소유현황
	국민연금자료	국민연금공단	연금불입 여부, 소득 비교
	법인등기자료	대법원	기본사항 및 조직형태
	수협위탁판매자료	수협중앙회	위탁판매 현황, 모집단 보완
	수협조합원명부	수협중앙회	경영주 성명, 경영형태, 어법종류
	어선원부	해양수산부	어선종류, 총톤수, 모집단 보완
	어업인허가대장	해양수산부	총톤수, 어선재질, 모집단 보완
어업생산 동향조사	원양어업생산량	해양수산부	생산량, 생산금액, 어법 등
	위탁판매량	수산업협동조합	생산량, 생산금액, 어법 등
	TAC 업서버 자료 (참홍어 및 붉은대게)	수산자원 관리사업단	생산량, 생산금액 등
외국인고용조사	등록외국인 명부 및 국내거소신고 외국국적동포 명부 자료	법무부	입국시기, 거주지주소, 체류자격, 국적 등
운수업조사*	근로소득지급명세서	국세청	종사자수 파악
	사업소득지급명세서	국세청	종사자수 파악
	일용근로소득지급명세서	국세청	일용근로자수 파악
인구동향조사	가족등록신고(전산시스템)	대법원	출생, 혼인 등 파악
인구주택총조사*	건강보험자료	건강보험공단	거주지 보완
	고용보험자료	고용노동부	거주지 보완
	국민연금자료	국민연금공단	거주지 보완
	산재보험자료	고용노동부	거주지 보완
전국사업체조사*	고용보험자료	고용노동부	업종, 종사자수
	산재보험자료	고용노동부	업종, 사업장 정보
	건강보험자료	국민건강보험공단	업종, 종사자수
	국민연금자료	국민연금공단	종사자수
	근로소득지급명세서	국세청	종사자수
	사업소득지급명세서	국세청	종사자수
	사업자등록자료	국세청	업종, 사업장 정보
	사업장단위 과세적용사업장	국세청	업종, 사업장 정보
	사업장현황신고서	국세청	면세사업자매출액
	일용근로소득지급명세서	국세청	일용근로자수
	법인등기자료	대법원	업종, 조직형태
	공장등록자료	산업통상자원부	사업장정보

통계조사명	행정자료명	보유기관	활용분야
전문과학기술서비스업조사*	부가가치세자료	국세청	본지사 임퓨테이션 지사별 매출액
	근로소득지급명세서		종사자수 파악
	사업소득지급명세서		종사자수 파악
	사업자등록자료		사업체명, 소재지, 사업자등록번호, 법인등록번호, 사업의 종류
	사업장현황신고서		면세사업자 매출액
	일용근로소득지급명세서		일용근로자수 파악
지역소득*	사업장현황신고서	국세청	면세사업자 매출액 파악
대기배출계정	자동차등록 및 이륜차 신고현황	국토교통부	일반대기오염, 온실가스 대기배출량 등

부록 8: 노르웨이 국가통계와 통계청에 관한 법

제1장 목적, 정의, 범위

1-1. 목적

이 법은 정보의 수집과 통계목적의 정보이용, 통계청의 조직과 기능 등을 규정함으로써 적합한 통계를 효율적으로 생산하는 것을 목적으로 한다.

1-2. 정의

(1) 통계는 집단내의 개별적인 개체나 이러한 개체의 집합에 대한 정보를 비교가공하거나, 현상에 대한 체계적인 관찰을 통하여 나타난 집단이나 현상에 관한 수량적 자료를 말한다.

(2) 국가통계는 공공의 활용을 목적으로 통계청이나 또 다른 지방청이 생산한 통계를 말한다.

1-3. 범위

(1) 국왕은 이 법이 만들어지고 적용될 노르웨이 경제구역, Svalbard, Jan Mayen 등과 그 부속물 범위를 결정해야 한다.

제2장 국가통계

2-1. 국가통계의 생산에 관한 결정

국가통계의 생산에 관한 결정은 국왕에 의해 행해져야 한다.
(재무부는 칙령 제387호(1989.6.16.)에 따라 이를 결정한다)

2-2. 정보제공의 의무

(1) 국왕은 다른 비밀보호에 관한 의무와 상충되지 않는 한 국가통계의 생산에 필요한 정보를 제공할 의무를 누구에게든 부과할 수 있다.

(재무부는 칙령 제387호(1989.6.16.)에 따라 이를 결정한다)

(2) 정보의 공급에는 정보제공의 기한이 설정되고 정보가 제공될 양식에 관한 조건이 설정되어야 한다. 정보를 기한만료 이전에 제공하지 않을 경우 정보제공의무 위반이 된다.

2-3. 강제벌금

정보제공의무의 대상자인 자가 그 의무를 위반할 경우 국가에 납부해야 할 강제벌금을 부과할 수 있다. 강제벌금의 부과는 납부를 강압하는 것에 기초하여야 한다. 이러한 강제벌금은 압류를 통하여 징수될 수 있다. 특별한 경우에 강제벌금이 의의제기를 당한 경우 그 일부나 전액을 감하여 줄 수 있다.

국왕은 그러한 강제벌금에 관한 보다 상세한 규정을 마련해야 한다.
(재무부는 칙령 제387호(1989.6.16.)에 따라 이를 결정한다)

벌금의 징수를 담당하는 국가기관은 첫 절에서 언급한 강제벌금을 징수하기 위해서 Creditors Security Act 의 2-7조의 규정에 따라 임금이나 이와 유사한 급여에 대한 압류를 지시할 수 있다.

징수담당기관은 청구권에 대한 압류를 통하여 벌금의 납부를 강제할 수 있는데 그 청구권은 레지스터에 등록된 것이거나 Mortgage Act의 제5장 3절에 의거 공시된 저당권을 말한다.

그리고 압류의 실행은 Act relating to the Enforcement of Claims 제6-9조 제1절에 따라 담당기관이 정한 바에 따라 집행되어야 한다.

2-4. 비밀보호의 의무

(1) 국가통계의 생산과 그 준비과정에 종사하는 자는 그 과정에서 알게 된 개인, 정부나 사업, 통계생산과 그 준비과정의 기술적인 수단이나 방법에 관한 정보에 관련한 지식에 접근하려는 것을 방지할 의무를 지고 있다. 비밀보호의 의무는 국가통계의 생산을 위해 수집된 정보에만 적용된다.

(2) 비밀보호의 의무는 해당업무를 마친 후에도 적용된다. 나아가 그러한 자는 이

조에서 언급된 정보를 그 자신의 사업이나 취업, 혹은 타인의 서비스 등에 이용하여서는 아니 된다.

(3) Public Administration Act 의 제13조에서 제13조의e 는 여기에 적용되지 아니 한다.

2-5. 정보의 이용

(1) 어떤 정보제공에 대한 의무규정에 따른 정보이건 자발적으로 제공된 정보이건 국가통계의 생산이나 자료검사관에 의한 승인을 받은 다른 목적을 위해서만 이용되어야 하고, 비밀보호의 범주를 침해해서는 안된다. 만일 정보가 이관되면 제2-4조에 의한 비밀보호의 의무가 또한 적용된다. 자료검사관은 특정장소를 지정하고 특정한 형태의 정보에 대하여 그러한 비밀보호 의무의 예외를 지정할 수 있다.

(2) 그러한 정보를 이관 받은 기관은 정보의 이용, 정보에의 접근과 관리책임자, 정보의 보관 및 이관된 자재의 반환, 복사본의 파괴 등에 관한 내부처리조건을 정한 규정을 마련해야 한다.

2-6. 정보의 공표

정보의 제공에 대한 의무를 규정한 것에 의해 수집된 정보이건 자발적으로 제공된 정보이건 간에 그것이 그것을 제공한 자에 관한 어떤 자료이라는 것, 혹은 그 개체를 식별하려고 추적하여 관련된 사람에게 해를 끼치게 할 수 있는 방법으로 공표하는 것이나, 또는 자료의 제공자나 개체에 대하여 이유없이 해를 입히는 것은 여하한 사정이 있다하여도 아니 되며, 이 경우 제5-1조 제3절에 규정에 의하거나 공공단체가 정한 같은 종류의 벌에 처한다.

2-7. 비밀보호의무의 중지

이 법에 따른 개인에 관한 비밀보호의 의무는 100년 후에 중지된다. 경영과 사업, 기술적인 장치나 방법에 관한 비밀보호의 의무는 60년 후에 정지된다.

제3장 통계청의 임무와 역할

3-1. 통계청의 임무

통계청은 국가통계의 생산과 보급의 중추기관이며, 제1조에 규정된 이 법의 목적이 달성되도록 할 주요 책임을 지고 있다. 이에 관하여 통계청은 다음 각호의 임무를 수행한다.

a) 국가통계 수요의 우선순위를 확정
b) 정부기관들이 생산한 통계들을 포괄한 통계조정
c) 통계기법의 개발과 분석과 연구에 대한 통계지원
d) 제2-5조의 범위 내에서 연구목적의 통계이용과 정부의 계획을 위한 정보의 공급
e) 국제통계협력을 위한 주요한 책무

3-2. 정부자료처리시스템

(1) 통계청은 국가기관과 전국적인 자치조직의 정부자료처리시스템을 국가통계의 기초자료로 이용할 권리를 가지고 있다.

(2) 국가기관이나 전국적인 자치조직이 주요한 정부자료처리시스템을 설치하거나 이를 변경한 때에는 미리 이러한 사항들에 관한 통지를 통계청에 보내야 한다. 통계청은 부가적인 정보를 요구할 수 있다. 통계청은 자료처리시스템이 통계적 안정성을 고려하여 설계되도록 하는 개선방안을 제안할 수 있다.

(3) 국왕은 위 (1), (2)을 실행하기 위하여 보다 상세한 사항을 규정할 수 있다.
(재무부는 칙령 제387호(1989.6.16.)에 따라 이를 결정한다)

3-3. 통계조정

(1) 중앙행정기관이 주요한 조사를 실시하기 전에 이에 관하여 통계청에 통보해야 한다. 통계청은 부가적인 정보를 요구할 수 있다. 통계청은 자료의 수집방법과 통계와 조정의 안정성을 고려하여 통계가 생산되도록 하는 방법에 대한 개선방안을 제안할 수 있다.

(2) 국왕은 공공조사기구가 본 조의 규정에 의한 중앙행정기관으로 간주될 것인가를 결정해야 한다.
(재무부는 칙령 제387호(1989.6.16.)에 따라 이를 결정한다)

제4장 통계청의 조직과 행정구조

4-1. 통계청의 행정구조

(1) 통계청은 국왕이 설립한 직업적으로 독립적인 기관이다.
(재무부는 칙령 제387호(1989.6.16.)에 따라 이를 결정한다)

(2) 통계청은 통계집행위원회와 청장에 의해 운영된다. 통계청장은 통계집행위원회가 협조해야 하는 몇 가지 사항을 제외하고는 통계행정을 지휘 감독한다. 국왕은 통계집행위원회와 통계청장의 임무에 관한 규정을 정한다.
(재무부는 칙령 제387호(1989.6.16.)에 따라 이를 결정한다)

4-2. 통계집행위원회의 임무와 구성

(1) 통계집행위원회는 통계청장이 제출한 장기계획, 예산안과 연간업무계획을 심의의결하고 이것들과 연차보고서를 내각에 제출한다. 다른 한편으로 통계집행위원회는 국가통계의 발전과 통계청의 활동에 관하여 일반적인 감독자로서의 역할을 수행한다.

(2) 통계집행위원회는 7인의 위원으로 구성된다. 국왕은 의장과 부의장을 지명한다. 일반적으로 위원들과 그 대리인들의 임기는 4년이다. 위원 1인과 대리인 1인은 일반 직원들의 추천으로 지명된다.
(재무부는 칙령 제387호(1989.6.16.)에 따라 이를 결정한다)

제5장 기타 규정

5-1. 벌칙

(1) 제2-2조, 정보제공의무와 관련된 제2-2조의 규정에 따른 신고서제출명령, 제2-5조와 관련된 정보이용, 제2-6조와 관련된 정보의 공표에 관한 규정을 고의적으로 위반할 경우 그 행위가 보다 엄중한 범죄규정에 해당하지 않는다면 벌금을 부과한다.

(2) 비밀유지의무 위반, 제2-5조의 규정에 따른 내부처리조건을 정한 규정의 위반은 형법 제121조에 의해 처벌된다.

5-2. 법의 완성 등

국왕은 이법을 완성하고 보충하는 것에 관련된 규정을 제정해야 한다.
(재무부는 칙령 제387호(1989.6.16.)에 따라 이를 결정한다)

제6장 법의 효력. 경과규정. 다른 법률에 대한 효력 등

6-1.

(1) 이법은 국왕이 결재한 날짜로부터 발효된다.

(2) 국왕은 이법이 제정되기 전에 국가통계를 목적으로 수집된 정보에 관한 경우에도 이법의 기간과 정지에 관한 규정이 또한 적용된다는 규칙을 정하여야 한다.
(재무부는 칙령 제387호(1989.6.16.)에 따라 이를 결정한다)

부록 9: 노르웨이 국가통계와 통계법 보충 법률 (통계법 시행령)

이 법은 1989. 6. 16 법률 제54호로 제정된 국가통계와 통계청에 관한 법 제 2-1, 2-2, 2-3, 3-2, 3-3, 4-1(2), 5-2, 6-1(4)조의 규정을 국왕의 칙령(제387)호에 의거 보충한 법이다.

제1장 정부자료처리시스템

1-1.

정부자료처리시스템은 국가기관이나 전국적인 자치조직이 각 기관들의 활동과 관련해서 검색하여 이용할 수 있도록 수집하여 저장하는 정보로서 이해되어야 한다.

“메인정부자료처리시스템”이란 개별적인 자료처리시스템을 조회할 수 있거나 전국이나 국가의 대부분을 포괄하는 것이다.

1-2.

통계청은 정부가 활용할 목적의 자료수집에 보다 효과적인 것으로 판단될 때 정부자료처리시스템에 접속할 수 있다. 만일 통계청의 판단할 때 정보제공의무에 의해 만들어진 정부자료처리시스템의 정보가 적절한 국가통계의 작성에 알맞은 데이터베이스로 확충되어야 한다면, 통계법 제2-2조의 규정에 의거 정보제공의무를 추가로 부과할 수 있다.

1-3.

국가기관이나 전국적인 자치조직은 중요한 새로운 정부자료처리시스템을 계획 중이거나 그러한 시스템을 바꾸려는 계획이 있을 때 통계청에 이에 관한 사항을 보고하여야 한다. 통계청은 보다 자세한 보고양식을 규정할 수 있다.

통계청은 정부자료처리시스템이 어떻게 설계되고 통계적 목적으로 이용되어야 할 것인가에 관하여 제안할 수 있다. 제안에 포함되는 것으로는

- 어떤 정보가 포함되어야 하는 가

- 단위, 변수, 분류 등의 정의와 시스템의 구조.
- 데이터 컨트롤
- 통계청으로 전송해야 할 정보의 종류와 시기

정부자료처리시스템이 설계될 때에는 비용측면이 중요하게 검토되어져야 한다.

제2장 통계조정

2-1.

주요통계조사라 함은 전국적인 조사거나 만일 정보의 개수나 정보의 총량이 크다면 국가 대부분을 조사하는 것을 의미한다.

2-2.

행정기관이 주요 통계조사의 실시를 위한 계획을 수립할 때에는 이에 관한 사항을 통계청에 보고하여야 한다. 통계청은 보다 자세한 보고양식을 규정할 수 있다.

2-3.

통계청이 통계와 조정에 관한 제약을 하고자 할 때는 정보가 어떻게 수집되고 어떤 방법으로 통계가 작성될 것인가에 관한 제안을 할 수 있다. 제안에 포함되는 것으로는

- 통계단위, 변수, 분류 등의 정의와 시스템의 구조.
- 자료의 수집과 양식, 조사표의 내용.
- 데이터의 처리
- 결과의 공표

정보의 수집과 통계의 작성에는 비용측면이 중요하게 검토되어져야 한다.

2-4.

통계청은 통계결과가 널리 공표되어야 할 것인지 통계청의 내부자료로만 활용될 것인가를 결정해야 한다.

제3장 강제벌금

3-1.

통계청은 통계법 제2-2조의 규정에 따른 정보제공의 기한을 어긴 자에 대하여 강제벌금을 부과할 수 있다. 강제벌금은 새로운 정보제공기한의 설정과 동시에 혹은 사전에 규정되어야 한다.

3-2.

강제벌금은 단순벌금액 부과방식인가 의무이행의 지체에 따라 벌칙이자가 합산되는 방식인가가 규정되어야 한다.

단순벌금액 부과방식은 정보제공이 완료되지 않는 한 새로운 제출기한이 정해지고 단순벌금액이 반복하여 부과된다. 벌칙이자가 합산되는 방식은 정보가 제공될 때까지 벌금액과 벌칙이자가 합산된다.

3-3.

통계청은 벌금의 규모를 정하는데 있어서 정보제공자들이 정보제공의무를 완수하는데 소요되는 비용을 고려하여 벌금이 의무불이행을 충분히 방지할 수 있는 벌금액을 정하여야 한다. 다시 말하면 강제벌금은 비용보다는 높아야 한다.

3-4.

강제벌금이 사전에 정하여 지는 경우에, 단순벌금액은 의무불이행이 발생한 날부터 납부가능 하여야 한다. 강제벌금이 사후에 정하여 지는 경우에는, 새로 정한 정보제공의무규정상의 마감일 종료 이후의 첫 번째 근무일부터 납부가능 하여야 한다. 지체된 강제벌금액의 과징금은 원금과 합산하여 납부하여야 한다.

3-5.

통계청은 합당한 사유가 있는 경우 강제벌금을 일부 혹은 전부를 경감시킬 수 있다.

3-6.

강제벌금부과에 대한 이의제기는 재무성에 대하여 할 수 있다.

제4장 통계집행위원회

4-1.

집행위원회는 통계청장이 상정하여 처리할 안건이 있는 경우 또는 3인 이상의 위원의 요구에 의해서 소집된다.

회의소집공고는 서면으로 하되 14일 이전에 고지되어야 한다. 회의소집공고문의 사본과 관련 서류는 대리인들에게도 송부되어야 한다.

통계청장은 통계집행위원회에 출석하여야 한다.

통계집행위원회의 의장은 회의를 주관한다. 의장이 부재시에는 부의장이 회의를 주관한다.

통계집행위원회의 의사정족수는 의장이나 부의장과 3인 이상의 위원이나 대리인, 그리고 통계청장이나 그를 대신한 자가 출석해야 한다.

4-2.

모든 통계집행위원이나 그의 대리인은 위원회의 활동영역 내에서 의안과 관련된 모든 문건에 대하여 접근할 권한을 갖는다. 관리자와의 협상이나 토론의 주제가 의안이 되는 경우에는 일반협정에 의거, 노동자의 경영참여가 의안이 되는 경우에는 특별협정에 의거 노동조합의 지지를 받은 후에 통계집행위원회에 상정되어야 한다.

4-3.

통계집행위원회는 참석자와 통계청장이 서명한 의사록을 유지하여야 한다. 통계집행위원이나 통계청장은 반대의견에 대한 속기록 기록을 요구할 수 있다.

의사록의 사본은 재무부와 감사원장에게 지체 없이 송달되어야 한다.

제5장 통계청장

5-1.

통계청장은 통계집행위원회의 심의를 거쳐야 할 사항을 제외한 통계행정을 총괄한다.

5-2.

통계청장은 통계법의 규정에 의거 통계집행위원회의 요구가 있기 전에 통계청의 업무에 관한 다음 사항을 준비하여 제출하여야 한다.

- 장기계획
- 연간예산안
- 연간업무계획
- 연차보고서

5-3.

통계청장은 통계청의 활동에 대한 책임을 진다. 그리고 사전에 계획된 업무가 현재의 법률과 규칙 하에서 가장 적절하고 경제적으로 실행되도록 할 책임이 있다. 그렇지 않은 통계청의 활동은 현행 법규와 국회를 통과한 의결에 따른 연간 배정이나 주어진 예산범위 그리고 현재 재정과 행정사무를 규율하는 규정에 따라 집행되어야 한다.

5-4.

통계청장은 통계집행위원회에 안건을 제출하기 전에 그것이 매우 중요한 것인지, 국가통계나 통계청의 이익의 원칙에 관한 것인지를 그 안건에 부속된 중요한 예산부담의 문제까지를 포함하여 제시하여야 한다.

5-5.

통계청장은 통계집행위원회와 내각으로 넘길 안건에 대하여 충분히 연구검토하고 적기에 제출할 의무가 있다.

제6장 발효

통계법(법률 제54호, 1989.6.16.)과 본 법규는 1990.6.16.부터 시행한다.

비밀보호의무의 시작과 종료 시기에 관한 통계법 제2-7조의 규정은 1907년 통계법의 규정에 의하여 수집된 정보, 특별한 법률에 의해 수집된 정보, 공공단체에 의해 수집된 정보, 현행 법 이전에 자발적인 제출을 통해 수집된 정보에도 적용된다.

부록 10: 데이터 소재 정보

운영기관	사이트이름	주소	유형
감사교육원	감사교육원 정보자료	http://www.bai-edu.go.kr/edu/index.do	7
감사연구원	감사연구원 연구발간물	http://www.bai-eri.go.kr/eri/index.do	7
감사원	감사원 감사연보	http://www.bai.go.kr/bai/cop/bbs/listBoardArticles.do?mdex=bai28&bbsId=BBSMSTR_100000000018	7
강원도청	강원통계정보	http://stat.gwd.go.kr/	3
강원발전연구원	강원발전연구원 연구보고서	http://www.rig.re.kr/hb/main	7
건강보험 심사평가원	건강보험 심사평가원 통계자료실	http://www.hira.or.kr/dummy.do?pgmid=HIRAA020045010000	5
건강보험 심사평가원	보건의료 빅데이터 개방시스템	http://opendata.hira.or.kr/home.do#none	3
검찰청	검찰청 통계자료	http://www.spo.go.kr/spo/info/stats/stats02.jsp	5
게임물관리위원회	게임물 관리위원회 관련통계자료	http://www.grac.or.kr/Statistics/Statistics.aspx	4
경기도청	경기통계	http://stat.gg.go.kr	3
경기연구원	통계로 보는 경기도	http://www.gri.re.kr/통계로-보는-경기도/	7
경남발전연구원	경남발전연구원 연구보고서	http://www.gndi.re.kr/report/res_report/list.php?topMenu=01&leftMenu=01&cate=0	7
경상남도청	경상남도 통계	http://stat.gsnd.net/jsp/main/main.jsp	3
경상북도청	경상북도 통계포털	http://stat.gb.go.kr/	3
경제인문 사회연구회	국가정책 연구포털(NKIS)	https://www.nkis.re.kr:4445/main.do#none	2
경찰대학 도서관	경찰대학도서관 경찰자료실	https://library.police.ac.kr/#/	7
경찰대학 치안정책연구소	경찰대학치안정책연구소 발간자료	http://www.psi.go.kr/police/psi/index.do	7
경찰대학	경찰대학 정보공개	http://www.police.ac.kr/police/index.do	6
경찰병원	경찰병원 정보공개	http://www.nph.go.kr/index.jsp	7
경찰청	경찰통계연보	http://www.police.go.kr/portal/main/contents.do?menuNo=200141	5
고용노동부	고용노동통계	http://laborstat.molab.go.kr/	3
공정거래위원회	공정거래위원회 통계연보	http://www.ftc.go.kr/policy/case/caseStaticList.jsp	7
과학기술 정책연구원	과학기술 정책연구원 기술혁신조사	http://kis.stepi.re.kr/kis_portal/index.do	5
관세국경 관리연수원	관세국경관리 연수원통계자료	https://ctc.customs.go.kr/	7
관세청	수출입무역통계	https://unipass.customs.go.kr:38030/ets/	3
관세청	통관포털 (UNI-PASS)	https://unipass.customs.go.kr/csp/index.do	2

광업등록사무소	광업등록사무소 주요통계	https://branch.motie.go.kr/	4
광주광역시청	광주시 통계정보시스템	https://stat.gjcity.go.kr/statdb/statdb02.asp	3
광주전남연구원	광주전남 경제동향	http://www.gjri.re.kr/cms/bbs/cms.php?dk_cms=report_03	7
교육부	교육부 교육통계	http://www.moe.go.kr/web/100085/site/contents/ko/ko_0118.jsp?selectId=1085	4
교육부	지방교육 재정알리미	http://www.eduinfo.go.kr/	3
교통안전공단	철도안전정보 종합관리시스템	https://www.railsafety.or.kr/web/index.jsp?section=null&edmui=null	4
구미전자 공업고등학교	구미전자 공업고등학교 정보공개	http://gnet.hs.kr/?main	6
국가과학 기술연구회	통합통계 정보서비스	http://stat.nst.re.kr/index.do	3
국가기록원 대통령기록관	대통령기록관 소장기록물통계	http://www.pa.go.kr/index.jsp	4
국가기록원	국가기록원 뉴스/정보공개	http://www.archives.go.kr/next/viewMain.do	7
국가기상위성센터	국가기상위성센터 정보마당	http://nmsc.kma.go.kr/html/homepage/ko/main.do	7
국가기술표준원	기술표준통계	http://www.kats.go.kr/main.do	7
국가민방위 재난안전교육원	국기민방위 재난안전교육원 정보공개	http://www.ndti.go.kr/	7
국가보훈처	국가보훈처 현황통계	http://www.mpva.go.kr/info/info600_list.asp	5
국가인권위원회	국가인권위원회 인권통계자료	http://www.humanrights.go.kr/03_sub/body03.jsp	6
국가정보원	산업스파이 적발통계	http://www.nis.go.kr/AF/1_5_1_1.do	4
국립5.18 민주묘지관리소	국립5.18 민주묘지관리소 묘역현황	http://518.mpva.go.kr/	6
국립3.15 민주묘지관리소	국립3.15 민주묘지관리소 묘역현황	http://315.mpva.go.kr/	6
국립4.19 민주묘지관리소	국립4.19 민주묘지관리소 묘역현황	http://419.mpva.go.kr/	6
국립고궁박물관	국립고궁박물관 정보공개	http://www.gogung.go.kr/openInfo.do	5
국립공원관리공단	국립공원관리공단 통계자료	http://knps.or.kr/front/portal/stats/statsDtl.do?menuNo=7070020&refId=statistics010967	6
국립공주병원	국립공주병원 정보공개	http://www.knmh.go.kr/	5
국립과천과학관	국립과천과학관 공개정보자료실	http://www.sciencecenter.go.kr/gnsm_web/main/	7
국립과학 수사연구원	국립과학수사 연구원 통계자료	http://www.nfs.go.kr/main.action	7
국립국악원 민속국악원	국립국악원 민속국악원 통계자료	http://namwon.gugak.go.kr/	7

국립국악원 부산국악원	국립국악원 부산국악원 학술연구	http://busan.gugak.go.kr/sub/03_01_03.jsp	7
국립국악원	국립국악원 정보공개	http://www.gugak.go.kr/site/main/index001	5
국립국어원	국립국어원 통계연감	http://www.korean.go.kr/	7
국립국제교육원	국립국제교육원 정보공개	http://www.niied.go.kr/index.do	6
국립기상연구소	국립기상연구소 연구마당	http://www.nimr.go.kr/MA/main.jsp	7
국립나주병원	국립나주병원 정보공개	http://www.najumh.go.kr/	5
국립농산물품질 관리원시험연구소	시험연구소 소통광장	http://www.naqs.go.kr/enr/	7
국립농산물 품질관리원	국립농산물품질 관리원 정보공개	http://www.naqs.go.kr/	7
국립농산물품질관 리원경기지원	경기지원 정보공개	http://www.naqs.go.kr/gg/	6
국립농산물품질관 리원경북지원	경북지원 정보공개	http://www.naqs.go.kr/gb/	6
국립농산물품질관 리원전남지원	전남지원 정보공개	http://www.naqs.go.kr/jn/	6
국립농산물품질관 리원제주지원	제주지원 정보공개	http://www.naqs.go.kr/jeju/	6
국립농업과학원 농업유전자원센터	농업유전자원 정보센터 자료실	http://www.genebank.go.kr/	7
국립농업과학원	국립농업과학원 연구정보	http://www.naas.go.kr/	7
국립대전현충원	국립대전현충원 통계정보	http://www.dnc.go.kr/html/kr/index.html	4
국립마산병원	국립마산병원 진료통계	http://www.mnth.go.kr/	5
국립망향의 동산관리원	국립망향의동산 정보공개	http://www.nmhc.go.kr/	5
국립목포병원	국립목표병원 정보공개	http://www.tbmokpo.go.kr/	5
국립무형유산원	국립무형유산원 자료마당	http://www.nihc.go.kr/nihchome/main/Index.do?mn=KO_01	7
국립문화재연구소 문화재보존 과학센터	문화재보존 과학센터 연구자료실	http://www.conservation.go.kr/	7
국립문화재연구소	국립문화재 연구소 자료마당	http://www.nrich.go.kr/	6
국립민속박물관	국립민속박물관 자료마당	http://www.nfm.go.kr/index.nfm	5
국립부곡병원	국립부곡병원 정보공개	http://www.bgnmh.go.kr/	7
국립산림과학원	국립산림과학원 통합자료실	http://www.forest.go.kr/newkfsweb/kfs/idx/SubIndex.do?orgId=kfri&mn=KFS_14	7
국립산림 품종관리센터	산림생명자원 보존및관리	http://www.forest.go.kr/newkfsweb/kfs/idx/SubIndex.do?orgId=kfsv&mn=KFS_18	7
국립생물자원관	국립생물자원관 정보마당	www.nibr.go.kr	4
국립서울병원	국립정신건강 센터 자료실	http://www.ncmh.go.kr/index.jsp	6

국립서울현충원	국립서울현충원 알림마당	http://www.snmb.mil.kr/mbshome/mbs/snmb/index.jsp	7
국립소록도병원	국립소록도병원 정보자료실	http://www.sorokdo.go.kr/index.jsp	7
국립수목원	국립수목원 연구	http://www.forest.go.kr/newkfsweb/kfs/idx/SubIndex.do?orgId=kna&mn=KFS_15	7
국립수산과학원 갯벌연구소	갯벌연구소 예보속보	http://www.nifs.go.kr/page?id=wf_intro_gb	7
국립수산과학원 고래연구소	고래연구소 멀티미디어 자료	http://www.nifs.go.kr/page?id=wa_intro	7
국립수산과학원 전략양식연구소	전략양식연구소 예보속보	http://www.nifs.go.kr/page?id=kr_index	7
국립수산과학원 중앙내수면연구소	중앙내수면연구소 예보속보	http://www.nifs.go.kr/page?id=fs_intro_01	7
국립수산과학원	국립수산과학원 예보속보	http://www.nifs.go.kr/page?id=kr_index	7
국립수산물 품질관리원	국립수산물품질 관리원 통계	http://www.nfqs.go.kr/2013/index.asp	6
국립식량과학원	국립식량과학원 연구활동	http://www.nics.go.kr/index.do	7
국립암센터	국가암정보센터	http://www.cancer.go.kr/mbs/cancer/index.jsp	7
국립영천호국원	국립영천호국원 묘역시설현황	http://www.ycnc.go.kr/	6
국립외교원 외교안보연구소	외교안보연구소 발간자료	http://www.ifans.go.kr/knda/ifans/kor/main/IfansMain.do	7
국립외교원	국립외교원 주요연구활동	http://www.knda.go.kr/knda/hmpg/kor/main/HmpgMain.do	7
국립원예 특작과학원	주요원예 특작산업통계	http://www.nihhs.go.kr/	7
국립이천호국원	국립이천호국원 묘역시설현황	http://www.icnc.go.kr/	4
국립임실호국원	국립임실호국원 묘역시설현황	http://www.isnc.go.kr/	6
국립자연 휴양림관리소	산림휴양 3.0	http://www.huyang.go.kr/main.action	6
국립재난 안전연구원	국립재난안전 연구원 연구실적	http://www.ndmi.go.kr/index.jsp	7
국립재활원 재활연구소	재활연구소 연구자료	http://www.nrc.go.kr/nrc/data/data01List.jsp?menu_cd=M_04_08&fno=37	4
국립재활원	국립재활원 건강통계	http://www.nrc.go.kr/	4
국립전파연구원 우주전파센터	우주전파센터 실시관관측정보	http://spaceweather.rra.go.kr/	4
국립전파연구원	국립전파연구원 전파자료실	http://www.rra.go.kr/ko/index.jsp	7
국립종자원	국립종자원 종자검정연구	http://www.seed.go.kr/index.jsp	6
국립중앙과학관	국립중앙과학관 자료실	http://www.science.go.kr/	7
국립중앙극장	국립중앙극장 정보공개	https://www.ntok.go.kr/	5
국립중앙도서관 국립어린이 청소년도서관	국립어린이 청소년도서관 발간자료	http://www.nlcy.go.kr/index.jsp	5
국립중앙도서관 국립장애인도서관	국립장애인 도서관 통계자료	http://nlid.nl.go.kr/able?act=searchDetail03	5

국립중앙도서관 도서관연구소	도서관연구소 통계자료	http://www.nl.go.kr/krili/	6
국립중앙도서관 세종도서관	세종도서관 이용소개	http://sejong.nl.go.kr/	7
국립중앙박물관	국립중앙박물관 정보공개	http://www.museum.go.kr/site/main/home	5
국립중앙의료원	국립중앙의료원 통계연보	http://www.e-gen.or.kr/nemc/statistics_annual_report.do?brdclscd=01	5
국립축산과학원	국립축산과학원 통계자료	http://www.nias.go.kr/front/main.do	7
국립춘천병원	국립춘천병원 정보공개	http://www.cnmh.go.kr/	7
국립특수교육원	특수교육 운영계획 및 관련통계	http://www.nise.go.kr/index_http.jsp	7
국립한글박물관	국립한글박물관 정보공개	http://www.hangeul.go.kr/main.do	7
국립해양 문화재연구소	국립해양문화재 연구소연보	http://www.seamuse.go.kr/seamuseweb/main/Index.do?mn=KO_01	7
국립해양조사원	국립해양조사원 해양자료센터	http://www.khoa.go.kr/	7
국립현대미술관	국립현대미술관 정보공개	http://www.mmca.go.kr/	7
국립환경과학원 교통환경연구소	교통환경연구소 연구활동	http://www.nier.go.kr/NIER/egovTprcIndex.jsp	7
국립환경과학원 국립습지센터	국립습지센터 습지자료실	http://www.wetland.go.kr/cmm/main/mainPage.do;jsessionid=h5sJX2kLBQKctZZNLpvtdJzf5TChm4mpyySMgVWMHZShhHhWzVVd!-2142916999	7
국립환경과학원 금강물환경연구소	금강물환경 연구소 주요사업	http://www.nier.go.kr/NIER/egovGeumIndex.jsp	7
국립환경과학원 낙동강물환경 연구소	낙동강물환경 연구소 주요사업	http://www.nier.go.kr/NIER/egovNakDongIndex.jsp	7
국립환경과학원 영산강물환경 연구소	영산강물환경 연구소 주요사업	http://www.nier.go.kr/NIER/egovYeongSanIndex.jsp	7
국립환경과학원 한강물환경연구소	한강물환경 연구소 주요사업	http://www.nier.go.kr/NIER/han/index.do	7
국립환경과학원	환경데이터	http://www.nier.go.kr/NIER/index.jsp	5
국립환경 인력개발원	국립환경인력 개발원 정보공개	http://ehrd.me.go.kr/member/index.do	7
국민건강보험공단	국민건강보험 공단 통계정보	http://www.nhis.or.kr/menu/boardRetriveMenuSet.xx?menuId=F3321	6
국민건강보험공단	국민건강 보험자료 공유서비스(NHISS)	https://nhiss.nhis.or.kr/bd/ay/bdaya001iv.do	3
국민안전처	국민안전처 주요통계	http://www.mpss.go.kr/home/policy/statistics/statisticsData/	3
국민안전처	국민재난 안전포털	http://www.safekorea.go.kr/idsiSFK/index_user.jsp	4
국민연금공단	국민연금공단 통계자료	http://www.nps.or.kr/jsppage/info/resources/info_resources_03_01.jsp?cmsId=statistics_month	5
국민체육진흥공단	체육통계자료	http://www.sports.re.kr/front/main/main.do?menu_seq=0	5
국민안전처	국민재난 안전포털	http://www.safekorea.go.kr/idsiSFK/index_user.jsp	4
국민연금공단	국민연금공단 통계자료	http://www.nps.or.kr/jsppage/info/resources/info_resources_03_01.jsp?cmsId=statistics_month	5

국민체육진흥공단	체육통계자료	http://www.sports.re.kr/front/main/main.do?menu_seq=0	5
국세공무원교육원	국세공무원 교육원정보마당	http://taxstudy.nts.go.kr/ntcs.main.dev	7
국세청	국세청 국세통계	http://www.nts.go.kr/info/info_03_02.asp?minfoKey=MINF49200802112100 12&top_code=&sub_code=&sleft_code=&ciphertext=	6
국악고등학교 및 국악학교	국악고등학교 및 국악학교 정보 공개	http://www.gukak.hs.kr/index.htm	7
국제지식 재산연수원	국제지식 재산연수원 정보마당	http://iipti.kipo.go.kr/main.do	7
국토교통부	국토교통 통계누리	https://stat.molit.go.kr/portal/event/eventIntroQuiz.do	2
국토교통부	건축행정시스템 (세움터)	http://www.eais.go.kr/	6
국토교통부	국가공간 정보포털	http://www.nsdi.go.kr/?menuno=2679	5
국토교통부	실거래가 공개시스템	http://rt.molit.go.kr/	4
국토교통 인재개발원	국토교통인재 개발원 정보공개	http://tilit.molit.go.kr/intro.do	7
국토연구원	세계도시정보 (UBIN)	https://ubin.krihs.re.kr/ubin/index.php	4
국토연구원	부동산 소비심리지수	http://www.krihs.re.kr/issue/consumeList.do	6
국토지리정보원	국토지리정보원 통계자료	http://www.ngii.go.kr/kor/main/main.do?rbsIdx=1	7
군산자유무역 지역관리원	군산자유무역 지역관리원 주요통계	http://www.ftz.go.kr/gunsan/	5
궁관리소 경복궁	문화재청 경복궁 간행물	http://www.royalpalace.go.kr:8080/content/data/data_05.asp	7
근로복지공단	근로복지공단 근로복지연구원	https://www.kcomwel.or.kr/Researchinstitute/lay1/program/S1T13C15/report_pg/list.do	5
금융감독원	금융통계 정보시스템	http://fisis.fss.or.kr	3
금융감독원	기업정보(DART)	http://dart.fss.or.kr/	6
금융위원회	금융시장동향	http://www.fsc.go.kr/info/trd_list.jsp?menu=7230000&bbsid=BBS0069	7
금융정보분석원	금융정보분석원 알림마당	http://www.kofiu.go.kr/index.jsp	7
기상레이더센터	기상레이더센터 알림마당	http://radar.kma.go.kr/index.do	4
기상청	기상재해통계	http://www.kma.go.kr/weather/lifenindustry/life_jisu.jsp	4
기상청	기상자료 개방포털	https://data.kma.go.kr/cmmn/main.do	2
기술신용보증기금	Kibo Tech-Bridge	https://tb.kibo.or.kr/ktms/board/statsView.do?rbsIdx=333	4
기업은행	경제연구소 (중소기업통계)	http://research.ibk.co.kr/research/#	5
기획재정부	기획재정부 통계	http://www.mosf.go.kr/st/submain.do?menuNo=6000000	6
기획재정부	열린재정	http://www.openfiscaldata.go.kr/portal/main.do	5
기획재정부	공공기관 경영정보 공개 (ALIO)	http://www.alio.go.kr/home.do	4

김제자유무역 지역관리원	김제자유무역 지역관리원 주요통계	http://www.ftz.go.kr/gimje/	4
남북회담본부	남북회담본부 회담통계	http://dialogue.unikorea.go.kr/ukd/main/userMain/main.do	4
농림축산검역검사본부중부지역본부	농림축산검역본부 중부지역본부 알림/참여공간	http://www.qia.go.kr/Jungbu/html/main.jsp	6
농림축산검역본부	농림축산 검역본부 식물검역통계	http://www.qia.go.kr/listindexWebAction.do	7
농림축산식품부	통계정보 물가정보	http://www.mafra.go.kr/list.jsp?group_id=1&menu_id=1&menu_name=국민소통&link_menu_id=10&division=H&board_kind=&board_skin_id=&parent_code=&depth=0&tab_yn=	5
농식품 공무원교육원	통계자료	http://www.ati.go.kr/	6
농촌진흥청	농업기술포털 (농사로)	http://www.nongsaro.go.kr	5
대구경북연구원	경제동향분석	http://www.dgi.re.kr/EJournal	7
대구광역시청	대구통계정보	http://www.daegu.go.kr/Statistics/	5
대불자유무역 지역관리원	대불자유무역 지역관리원 주요통계	http://www.ftz.go.kr/daebul/	4
대외경제 정책연구원	세계경제통계	http://www.kiep.go.kr/sub/list.do?bbsId=worldEcoStat	6
대전광역시청	대전의 통계	http://www.daejeon.go.kr/sta/index.do	3
대전발전연구원	월간대전경제	http://www.djdi.re.kr/sub01/boardList.do?hbbsid=economyboard	7
대통령기록관	대통령 기록연구실	http://www.pa.go.kr/research/contents/speech/index.jsp	7
대한건설협회	대한건설협회 통계	http://www.cak.or.kr/stat/statisticsMain.do?menuId=7	5
대한기계설비 건설협회	대한기계설비 건설협회 통계연보	http://www.kmcca.or.kr/main.do#none	7
대한민국 역사박물관	대한민국 역사박물관 열린마당	http://www.much.go.kr/	7
대한석탄공사	유연탄 가격동향	https://www.kocoal.or.kr/board/list.php?table=1038	7
대한시설물유지 관리협회	시설물관리 유지업 통계연보	http://www.fma.or.kr/daehan1/overview/2-8.asp	7
대한장애인체육회	생활체육 정보센터	http://sports.koreanpc.kr/club/club_stats.asp	3
대한전문건설협회	대한전문 건설협회 통계연보	http://www.kosca.or.kr/infoplaza/infoplaza_statistics_view.asp?area=00&GBN=A	7
대한체육회	스포츠지원포털	http://www.sportsg1.or.kr/	3
대한측량협회	대한측량협회 통계자료	http://www.kasm.or.kr/information/listETCurrentstatus.do	3
도로교통공단	교통사고분석 시스템(TAAS)	http://taas.koroad.or.kr/	1
도서관정보 정책위원회	국가도서관 통계시스템	http://www.libsta.go.kr	4
동해자유무역 지역관리원	동해자유무역 지원관리원 주요통계	http://www.ftz.go.kr/donghae/	4

마산자유무역 지역관리원	마산자유무역 지원관리원 주요통계	http://www.ftz.go.kr/kor/main.jsp	4
문화재청	문화재청 통계정보	http://www.cha.go.kr/html/HtmlPage.do?pg=/seek/search/statistics.jsp&mn=NS_03_06	4
미래창조과학부	미래창조과학부 NTIS 현황통계	http://sts.ntis.go.kr	3
방송통신위원회	방송통신위원회 자료마당	http://www.kcc.go.kr/user.do?boardId=1027&page=A02060100&dc=K02060100	5
법무연수원	법무연수원 정보마당	http://www.lrti.go.kr/homepage/index/MainAction.do?method=index	7
병무민원상담소	병무민원상담소 통계자료	http://www.mma.go.kr/minwon/index.do	7
병무청	병무청 공개/개방포털	http://open.mma.go.kr/caisGGGS/	5
보건복지부	보건복지부 통계포털	http://stat.mohw.go.kr/	5
보험개발원	보험통계포털 서비스(INSIS)	http://www.insis.or.kr/insis/insisweb/index.jsp	2
부산광역시청	부산통계	http://www.busan.go.kr/Page.bs?parcode=MNU_00000000089&prgcode=CMS_00000000394	3
부산광역시청	부산경제 통계포털	http://www.becos.kr/	2
부산기계공업 고등학교	부산기계공업 고등학교 정보공개	http://bmt.hs.kr/	4
부산발전연구원	부산발전연구원 BDI 간행물	http://www.bdi.re.kr/program/researchreport/lstResearchReport.asp	7
부산항만공사	부산항만공사 통계자료	http://www.busanpa.com/kor/Contents.do?mCode=MN0931	3
사립학교교직원 연금공단	통계로 보는 사학연금	http://www.tp.or.kr:8088/tp/news/news_st_tab1.jsp	7
사회보장정보원	사회보장통계	http://www.ssis.or.kr/lay1/bbs/S1T206C227/A/62/list.do	7
산림교육원	산림교육원 통계자료	http://www.forest.go.kr/newkfsweb/kfs/idx/SubIndex.do?orgId=fhi&mn=KFS_16	7
산림청	산림통계	http://www.forest.go.kr/newkfsweb/kfi/kfs/stats/localStats.do?mn=KFS_02_03_01_01	5
산림청	맞춤형 산림통계서비스	http://116.67.44.22/forest/#/stats	4
산림항공본부	산림항공본부 통계자료	http://www.forest.go.kr/newkfsweb/kfs/idx/SubIndex.do?orgId=fao&mn=KFS_17	7
산업연구원	산업통계분석 시스템(ISTANS)	http://www.istans.or.kr/	3
산업연구원	산업연구원(KIET) 서비스통계포털	http://service.kiet.re.kr/html/	3
산업통상자원부	산업통상자원부 통계정보	http://www.motie.go.kr/motie/py/sa/industrytrende/inderstryTrendUser/inderstryTrendList.do?frst_yn=Y	4
생명보험협회	생명보험통계	http://www.klia.or.kr/board/boardList.do?bCd=19	5
서울과학수사 연구소	서울과학수사 연구소 통계자료	http://www.nfs.go.kr/branchSeoul.action	4
서울연구원	서울연구데이터 서비스	http://data.si.re.kr/	6
서울특별시청	서울통계	http://stat.seoul.go.kr/jsp3/index.jsp	3
세종특별자치시청	세종통계	http://www.sejong.go.kr/stat.do	5
소상공인 시장진흥공단	소상공인포털	http://www.sbiz.or.kr/sup/main.do	4

수도권매립지 관리공사	수도권매립지관리공 사폐기물3.0포털	http://open.slc.or.kr/main.do	6
수도권특수구조대	수도권특수구조대 통계정보	http://www.rescue.go.kr/menu6_sub4	4
수산경제연구원	수산경제연구원 조사통계	http://fei.suhyup.co.kr/total/total_10.jsp	7
식품안전정보원	식품안전정보원 통계정보	http://www.foodinfo.or.kr/portal/bank/statisticsInfo_14_01.jsp	7
식품의약품안전처	식품의약품안전처 통계포털	http://www.mfds.go.kr/stat/infographic/index.jsp	3
식품의약품 안전평가원	식품의약품 안전평가원 정보마당	http://www.nifds.go.kr/nifds/	7
어업관리단동해	동해어업관리단 알림마당	http://eastship.mof.go.kr/index.do	7
어업관리단서해	서해어업관리단 알림마당	http://westship.mof.go.kr/index.do	6
에너지경제연구원	국가에너지통계 종합정보시스템 (KESIS)	http://www.kesis.net	4
여성가족부	여성가족부 통계자료	http://www.mogef.go.kr/korea/view/policy/policy02_05a.jsp	6
여수광양항만공사	여수광양항만공사 주요통계	http://www.ygpa.or.kr/kr/release_information/public_data_open/major_result_statistics/	6
연구개발특구 진흥재단	연구개발특구 통계자료	https://www.innopolis.or.kr/sub0405	7
영남특수구조대	영남특수구조대 통계정보	http://www.rescue.go.kr/menu6_sub4	4
영상물등급위원회	등급분류통계	http://www.kmrb.or.kr/data/statsMovie1.do	3
영화진흥위원회	영화진흥위원회 조사연구통계	http://www.kofic.or.kr/kofic/business/rsch/findPolicyList.do	7
예금보험공사	예금보험공사 주요통계	http://www.kdic.or.kr/pds/pds_static_list.jsp	6
오송생명과학 단지지원센터	오송생명과학 단지지원센터 정보공개	http://osong.mohw.go.kr/user/index.do	6
온실가스 종합정보센터	온실가스통계	http://www.gir.go.kr/home/main.do	5
외교부	국가 및 지역정보	http://www.mofa.go.kr/countries/index.jsp?menu=m_40	7
우정사업본부 우정 공무원교육원	우정공무원 교육원 알림마당	http://www.kpoti.go.kr/	7
우정사업본부	우정사업본부 통계정보 연차보고서	http://www.koreapost.go.kr/kpost/main/index.jsp	6
울산광역시청	울산광역시 통계	http://www.ulsan.go.kr/rep/ghstat1	5
울산발전연구원	울산발전연구원 통계 DB	http://www.udi.re.kr/ulsan/ulsan02.php	3
울산자유무역 지역관리원	울산자유무역 지역관리원 주요통계	http://ftz.go.kr/ulsan/	4
울산항만공사	항만운영통계	https://www.upa.or.kr/cms/CmsAction.do?cmd=CmsView&menu_key=14&menu_id=72&HOME_SEQ=2&next=/cargo/CargoAction.do?cmd=CargoV&TYPE=1	5
율촌자유 무역지역관리원	율촌자유무역 지역관리원주요통계	http://www.ftz.go.kr/yulchon/	4

이북5도위원회	이북5도위원회 알림마당	http://www.ibuk5do.go.kr/5do/frt/main.do	6
인사혁신처	인사혁신처 인사통계	http://www.mpm.go.kr/mpm/info/infoStatistics/hrStatistics/hrStatistics01/	6
인천광역시청	인천통계	http://www.incheon.go.kr/articles/180	6
인천국제공항공사	항공통계	http://www.airport.kr/co/ko/3/6/3/index.jsp	3
인천발전연구원	월별경제동향	http://www.idi.re.kr/report/publish/list	7
인천항만공사	인천항포탈	http://www.ipus.co.kr/ipa/ipapt/information/statics/selectTonInExportStaticList.do	3
재외동포재단	KOREAN NET 정보센터 통계자료	http://www.korean.net/portal/PortalView.do	3
전라남도청	전라남도청 통계정보	http://www.jeonnam.go.kr/contentsView.do?menuId=jeonnam0508020000	6
전라북도청	전라북도 통계시스템	http://stat.jeonbuk.go.kr/	5
전력거래소	전력통계 정보시스템	http://epsis.kpx.or.kr/epsis/	3
전북기계공업 고등학교	전북기계공업 고등학교 정보공개	http://www.cmt.or.kr/index.do	6
전북연구원	전북연구원 보고서	http://www.jthink.kr/bbs/board.php?bo_table=sub_04_01	7
전통예술고등학교 및 전통예술학교	전통예술고등학교 및 전통예술 학교 정보공개	https://www.kugak-am.hs.kr/high/	7
정보통신 산업진흥원	정보통신 산업진흥원 정책/통계자료	http://www.nipa.kr/know/policyList.it?menuNo=27&page=1	5
정보통신 정책연구원	방송뉴미디어통계정 보시스템 (KISDISTAT)	http://stat.kisdi.re.kr/	3
정보통신 정책연구원	정보통신정책 연구원 (IT STAT)	http://www.itstat.go.kr/home.it	3
정부청사관리소	정부청사관리소 정보공개	http://www.chungsa.go.kr/chungsa/frt/main.do	7
정부통합전산센터	정부통합전산센터 알림마당	http://www.ncis.go.kr/index.jsp	5
제주발전연구원	제주통계	https://www.jdi.re.kr/contents/index.php?mid=1207	5
제주특별자치도청	제주도청 통계정보	http://www.jeju.go.kr/open/governmen/stats/stats.htm	5
조달교육원	조달교육원 참여마당	https://hrd.pps.go.kr:8443/index.do	4
조달청	나라장터 조달통합 정보시스템	http://www.g2b.go.kr:8060/jsp/out/index.jsp	4
조달품질원	조달품질원 정보제공	https://www.pps.go.kr/kor/jibang/center/indexCenter.do	7
주택도시보증공사	주택정보포털	http://housta.khgc.co.kr	4
중소기업은행	중소기업은행(IR)	http://www.ibk.co.kr/common/navigation.ibk?linkUrl=/intro/ir/ir_stock_ann.jsp&pageId=IR05030100#	7
중소기업중앙회	중소기업통계	http://www.kbiz.or.kr/home/homeIndex.do?menuCode=stat	5
중소기업청	중소기업 조사통계시스템	http://stat2.smba.go.kr/index.jsp	3
중앙119구조본부	중앙119 구조본부 자료실	http://www.rescue.go.kr/	6

중앙경찰학교	중앙경찰학교 정보공개	https://www.cpa.go.kr/cpa2013/	6
중앙공무원교육원	중앙공무원교육원 정보공개	http://www.nhi.go.kr/	7
중앙관세분석소	중앙관세분석소관세분석	http://www.customs.go.kr/kcshome/site/index.do?layoutSiteId=cclss	7
중앙교육연수원	중앙교육연수원 자료실	http://www.neti.go.kr/	6
중앙소방학교	중앙소방학교 자료실	http://www.nfsa.go.kr/	7
중앙신체검사소	중앙신체검사소 소식정보	http://www.mma.go.kr/jungang/index.do	7
중앙전파관리소 위성전파감시센터	위성전파감시 센터동향자료	http://www.srmc.go.kr:8080/trend/trendMoonBBSList.do	7
중앙전파관리소	전파관리통계	http://www.crms.go.kr/	6
지방행정연수원	지방행정연수원 정보공개	http://www.logodi.go.kr/	5
질병관리본부 국립보건연구원	국립보건연구원 통계자료	http://www.nih.go.kr/NIH_NEW/main.jsp	7
질병관리본부	질병관리본부 통계자료	http://www.cdc.go.kr/CDC/main.jsp	7
청렴연수원	청렴연수원 정보공개자료실	https://www.acrc.go.kr/edu/index.do	7
축산물안전관리인증원	HACCP관련통계	http://www.ihaccp.or.kr/site/haccp/sub.do?key=221	7
축산물품질평가원	축산물등급 판정통계	http://www.ekape.or.kr/view/micro/statistics/sta_01_01.asp	3
축산물품질평가원	축산유통 종합정보센터	http://www.ekapepia.com/home/homeIndex.do	4
출입국외국인정책본부	출입국 외국인정책본부 통계자료실	http://www.immigration.go.kr/HP/COM/bbs_003/BoardList.do?strNbodCd=noti0097&strOrgGbnCd=104000&strFilePath=imm/&strRtnURL=IMM_6070&strNbodCdGbn=&strType=&strAllOrgYn=N	5
충남연구원	연구원 간행물	http://www.cdi.re.kr/report/list.asp?id=E1&sid=E101&code=040100	7
충북연구원	CRI 발간물	http://www.cri.re.kr/sub.php?menukey=17	7
충청남도청	충남통계정보관	http://www.chungnam.net/statsMain.do	5
충청북도청	충북통계 정보시스템	http://stat.cb21.net/stat/index.do	5
통계개발원	통계개발원 통계자료	http://kostat.go.kr/edu/sri_kor_new/index.action	7
통계교육원	통계교육원 통계자료	http://sti.kostat.go.kr/	6
통계청	국가통계포털 (KOSIS)	http://kosis.kr/	1
통계청	통계지리정보 서비스(SGIS)	http://sgis.kostat.go.kr/view/index	4
통계청	마이크로데이터 통합서비스(MDIS)	https://mdis.kostat.go.kr/index.do	3
통계청	국가지표체계 (e-나라지표)	http://www.index.go.kr/potal/main/PotalMain.do	3
통계청	통계분류포털	https://kssc.kostat.go.kr:8443/ksscNew_web/index.jsp	7
통일교육원	통일교육원 통계자료	http://www.uniedu.go.kr/uniedu/home/cms/page/uniedu/main.do?mid=UNIEDU&main=true	7
통일부	통일부 통계자료	http://www.unikorea.go.kr/content.do?cmsid=3099	6

통일연구원	통일연구원 통계자료	http://www.kinu.or.kr/info/info_04_01.jsp	7
특허청	지식재산권통계	http://www.kipo.go.kr/kpo/user.tdf?a=user.html.HtmlApp&c=3041&catmenu=m04_05_02	6
한국가스공사	KOGAS	http://www.kogas.or.kr/	7
한국가스안전공사	가스통계	http://www.kgs.or.kr/publish/Board.do?method=list&board_id=main_13	7
한국감정원	통계정보	http://www.kab.co.kr/kab/home/trend/fluctuation01.jsp	3
한국거래소	한국거래소 통계	http://marketdata.krx.co.kr/main/main.jsp	3
한국건설 기술관리협회	한국건설기술 관리협회 통계	http://www.gamri.or.kr/storehouse/record.asp	6
한국건설 기술연구원	해외건설종합 정보서비스	http://www.icak.or.kr/sta/sta_0101.php	7
한국고용정보원	고용행정통계	http://statistics.keis.or.kr/stats/index.do	3
한국고용정보원	한국고용정보원 고용조사	http://survey.keis.or.kr/	6
한국고전번역원	국가고전변역 통계정보	http://www.itkc.or.kr/itkc/contents/ContentsView.jsp?menuId=M0522	3
한국공항공사	항공통계	http://www.airport.co.kr/www/extra/stats/airportStats/layOut.do?cid=20151029175015422 53&menuId=397	3
한국과학기술 기획평가원	한국과학 기술기획평가원 통계자료	http://www.kistep.re.kr/c3/sub6.jsp	7
한국과학창의재단	과학창의연례통계	https://www.kofac.re.kr/?page_id=15583	7
한국관광공사	한국관광통계	http://kto.visitkorea.or.kr/kor/notice/data/statis/profit.kto	7
한국관광공사	한국여행업협회 (KATA 여행통계)	https://www.kata.or.kr/search/03_result_1.asp?boardtype=A	7
한국광물자원공사	한국자원 정보서비스	https://www.kores.net/komis/main/userMain/main.do	3
한국광물자원공사	국가광물 자원지리정보망	http://www.kmrgis.net/	4
한국교육개발원	교육통계서비스	http://cesi.kedi.re.kr/index	3
한국국방연구원	국방예산통계	http://www.kida.re.kr/ja_statistic/index.htm	3
한국국제협력단	한국국제협력단 통계사이트	http://stat.koica.go.kr:8077/komis/jsptemp/ps/stat_index.jsp	3
한국금융투자협회	자본시장통계	http://freesis.kofia.or.kr/?	3
한국기계 산업진흥회	정보센터 (기계산업무역통계)	https://www.koami.or.kr/	4
한국기계 산업진흥회	소재부품 종합정보망	http://www.mctnet.org/index.jsp	4
한국노동연구원	고용노동정보	https://www.kli.re.kr/kli/selectBbsNttList.do?bbsNo=6&key=40	5
한국농수산대학	한국농수산대학 정보공개	https://www.af.ac.kr/index.9is	7
한국농수산 식품유통공사	해외시장동향	http://www.at.or.kr/article/apko336800/list.action	7
한국농어촌공사	농업기반 시설관리	http://rims.ekr.or.kr/index.aspx	7
한국농촌 경제연구원	농업관측 통계시스템	http://oasis.krei.re.kr/index.jsp	3
한국대학 교육협의회	대학알리미	http://academyinfo.go.kr/	3
한국도로공사	한국도로공사 교통통계	http://www.ex.co.kr/site/com/pageProcess.do?oneDpMnNo=100300000&selDpPgNm=menuno=100305000&selDpMnNo=100305000&paramStr=&detailTp=&dirNm=portal&kwd=	6

한국도박 문제관리센터	한국도박 문제관리센터 통계	https://www.kcgp.or.kr/np/stats/2/stats.do	4
한국디자인진흥원	디자인 DB	http://www.designdb.com/dtrend/trend.l.asp?menupkid=377	7
한국로봇산업협회	한국로봇 산업협회 조사/통계사업	http://www.korearobot.or.kr/business/business4.htm	7
한국무역협회	무역통계(K-stat)	http://stat.kita.net/	3
한국문학번역원	LTI KOREA DB	http://www.klti.or.kr/ku_01_05_011.do?start_year=2001	4
한국문화 관광연구원	문화통계포털 (문화셈터)	http://stat.mcst.go.kr	2
한국문화 관광연구원	관광지식 정보시스템	http://www.tour.go.kr	3
한국방송 광고진흥공사	방송통신광고 통계시스템	https://adstat.kobaco.co.kr/default.asp	3
한국방송통신 전파진흥원	전파정보 개방시스템	http://spectrummap.kr/index.do	3
한국방위 산업진흥회	방산통계	http://www.kdia.or.kr/content/3/2/1/view.do	4
한국벤처투자	한국벤처투자 정보공개	http://www.k-vic.co.kr/board.do?boardConfigNo=81&menuNo=506	7
한국보건 사회연구원	보건복지 데이터포탈	https://data.kihasa.re.kr/index.jsp	2
한국보건 산업진흥원	보건산업통계 (KHISS)	https://www.khiss.go.kr/	3
한국사학진흥재단	사립대학 회계정보시스템	http://infor.sahak.or.kr/portal_main.asp	3
한국산업 기술진흥원	소재부품 종합정보망	http://www.mctnet.org/index.jsp	7
한국산업 기술평가관리원	지역발전포털 (REDIS)	http://www.redis.go.kr/route/main_home.do	4
한국산업단지공단	산업단지정보	http://www.e-cluster.net/new_app/lab/brief/brief_list.jsp	6
한국산업 안전보건공단	산업재해통계	http://www.kosha.or.kr/board.do?menuId=554	6
한국산업은행	공공데이터개방	https://www.kdb.co.kr/ih/simpleJsp.do	6
한국산업인력공단	큐넷 (자격검정통계)	http://www.q-net.or.kr/crf011.do?id=crf01101&gSite=Q&gId=	6
한국상하수도협회	상하수도통계	http://www.kwwa.or.kr/reference/stats01_list.php	5
한국생산성본부	생산성통계 DB	http://www.kpc.or.kr/Productivity/StatisticDB.asp	3
한국석유공사	석유정보망 (Petronet)	http://www.petronet.co.kr/v3/index.jsp	3
한국섬유 개발연구원	한국섬유산업 연합회 (섬유정보포털)	http://textopia.or.kr/new/index.do	3
한국세라믹기술원	세라믹 생산통계	http://www.kicet.re.kr/Front/Default.aspx?MenuNo=1050201	3
한국소방 산업기술원	소방산업정보시스템	http://www.sobangin.or.kr/sub4/board_lst.asp?inja=tong_b&Lnm=20501	5
한국소비자원	국제거래소비자 포털 통계자료	http://crossborder.kca.go.kr/stats/statsAreaList.do	4
한국소비자원	한국소비자원참가격	http://www.price.go.kr/tprice/portal/main/main.do	3
한국소프트웨어 산업협회	SW기술자 평균임금	https://www.sw.or.kr/biz/b_noi.jhtml	3
한국수산자원 관리공단	한국수산자원 관리공단정기간행물	http://www.fira.or.kr/fira/fira_040401.jsp	7

한국수자원공사	한국수자원공사 My Water	http://www.water.or.kr/wstatic/main/index.do	3
한국수출입은행	한국수출입은행 해외경제연구소	http://keri.koreaexim.go.kr/site/main/index007	3
한국수출입은행	수은리소스센터 통계자료	https://www.koreaexim.go.kr/site/program/resource/resource?menuid=001003004002	3
한국승강기 안전공단	승강기안전공단 사고통계	http://www.koelsa.or.kr/wpge/m_140/info/info020201.do	4
한국시설안전공단	시설물정보관리 종합시스템	http://www.fms.or.kr/	7
한국언론진흥재단	한국언론 진흥재단 미디어통계	http://www.kpf.or.kr/board/Z/detail.do?menu_link=pds/MediaStatsInfomation/MediaStatsInfomation_01/summary&p_menu_id=null&menu_id=197&menu_orderby=5&tabYn=N	5
한국에너지공단	에너지통계	http://www.kemco.or.kr/web/kem_home_new/info/statistics/overview.asp	7
한국에너지공단	신재생에너지 통계자료	http://www.knrec.or.kr/knrec/14/KNREC140300.asp	7
한국에너지 기술연구원	신재생에너지 데이터센터	http://kredc.kier.re.kr/kier/	7
한국엔지니어링 협회	엔지니어링정보 종합시스템	http://www.etis.or.kr/webs/main.jsp	7
한국여성 경제인협회	여성기업 종합정보포털	http://www.wbiz.or.kr/web/board/BD_board.list.do?bbsCd=1011	7
한국여성 정책연구원	성인지 통계정보시스템	https://gsis.kwdi.re.kr/gsis/kr/main.html	3
한국예술종합학교	입시통계자료	http://www.karts.ac.kr/	7
한국예탁결제원	증권정보포털 (SEIBro)	http://www.seibro.or.kr	4
한국은행	경제통계시스템 (ECOS)	http://ecos.bok.or.kr/	3
한국의료분쟁 조정중재원	한국의료분쟁 조정중재원자료실	https://www.k-medi.or.kr/cop/bbs/selectBoardListUser.do?bbsId=BBSMSTR_000000000134	7
한국의약품 안전관리원	의약품안전정보 지식자료실	https://www.drugsafe.or.kr/iwt/ds/ko/bbs/EgovBbs.do?bbsId=BBSMSTR_000000000021&nttId=1557&pageIndex=1&searchCnd=&searchWrd=	7
한국인체 조직기증원	조직기증 통계정보	http://kftd.or.kr/info/statistic.asp	7
한국인터넷진흥원	인터넷통계 정보검색시스템	http://isis.kisa.or.kr/	3
한국임업진흥원	임산물가격 정보시스템	https://fps.kofpi.or.kr/fnt/price/wood/market/list.do	3
한국자동차 산업협회	한국자동차산업 협회 자동차통계	http://www.kama.or.kr	6
한국장애인 고용공단	한국장애인 고용공단통계자료실	https://www.kead.or.kr/common/comm_board.jsp?main=4&sub1=4	7
한국저작권위원회	저작권통계	http://www.copyright.or.kr/information-materials/statistics/registration/index.do	7
한국전기공사협회	전기공사종합 정보시스템 (ECIC)	http://www.ecic.kr/	4
한국전기 산업연구원	표준시장단가	http://erik.re.kr/index.php?mid=data_1	7
한국전력거래소	전력통계 정보시스템	http://epsis.kpx.or.kr/epsis/	3

한국전력공사	한국전력공사 전력통계	http://cyber.kepco.co.kr/kepco/KO/ntcob/list.do?boardCd=BRD_000097&menuCd=FN050301	6
한국전자정보통신산업진흥회	전자산업정보 통계자료	http://www.gokea.org/bbs/list.php?board_id=kor_statistics&idx=34&m=2	7
한국정밀화학산업진흥회	한국정밀화학산업진흥회 (정밀화학통계)	http://www.kscia.or.kr/ver01/html/b_01.php	4
한국정보통신공사협회	법령자료 (공사업관련통계)	http://www.kica.or.kr/index.jsp	4
한국조세재정연구원	조세·재정 통계	http://www.kipf.re.kr/TaxFiscalPubInfo/TaxFiscaltax/A01/1	6
한국주택금융공사	주택금융공사 통계자료	http://www.hf.go.kr/cmspubl/template8/MN00000376.jsp?on=23	5
한국지식재산연구원	한국지식재산연구원 간행물	https://www.kiip.re.kr/board/data/list.do?bd_gb=data&bd_cd=2&bd_item=0&po_item_gb=1	7
한국지식재산전략원	정부R&D특허성과관리시스템(RIPIS)	http://www.rndip.or.kr/main.do?method=main	3
한국지질자원연구원	MiCi 자원전략연구실	https://mici.kigam.re.kr/Portal_335/	5
한국직업능력개발원	인재개발자료실	http://www.nhrd.net/board/list.do?boardId=BBS_0000009&menuCd=DOM_000000103001000000&contentsSid=31	7
한국철강협회	한국철강협회 통계자료	http://www.kosa.or.kr/statistics/production_2011.jsp	3
한국철도공사	한국철도공사 자료실	http://info.korail.com/mbs/www/jsp/board/list.jsp?boardId=9863289&id=www_060700000000	7
한국철도시설공단	철도산업정보센터(KRIC)	http://www.kric.or.kr/jsp/industry/rss/operatingReportList.jsp	3
한국청소년정책연구원	한국아동청소년데이터아카이브	http://archive.nypi.re.kr/	6
한국출판문화산업진흥원	한국출판문화산업진흥원 통합자료실	http://www.kpipa.or.kr/info/totaldataList.do?board_id=50	7
한국토지주택공사	온나라 부동산정보	http://www.onnara.go.kr/index.jsp	1
한국특허정보원	지식재산정보통합서비스	http://www.ipis.or.kr/infoPlace/patentInfo/standardPaState.do	7
한국해양수산개발원	해양수산통계	http://www.kmi.re.kr/web/contents/contentsView.do?rbsIdx=221	6
한국행정연구원	행정DB사이트	http://survey.kipa.re.kr/database/frt/main.do	5
한국형사정책연구원	범죄와 형사사법 통계정보(CCJS)	http://www.crimestats.or.kr/main/index.k2?cmd=main	3
한국환경공단	에어코리아 (Air Korea)	http://www.airkorea.or.kr/index	3
한국환경공단	All baro System	https://www.allbaro.or.kr/index.jsp	3
한국환경공단	국가소음정보시스템	http://www.noiseinfo.or.kr/index.jsp	3
한국환경공단	국가하수도정보시스템	https://www.hasudoinfo.or.kr/stat/selectDataList.do	6
한국환경공단	자동차배출가스종합전산시스템 (mecar)	http://www.mecar.or.kr/veims/support/statProductCarAuthTest.do	3
한국환경공단	환경통계정보 (폐기물, 재활용)	http://stat.recycling-info.or.kr/keco_web/main/	7
한국환경공단	국가상수도정보시스템	http://www.waternow.go.kr/openPage.do?OPENCODE=PG_HOME	6

한국환경 산업기술원	국가환경산업기술 정보시스템 (KONETIC)	http://www.konetic.or.kr/main.asp	7
항공교통센터	항공교통센터 항공통계	http://acc.molit.go.kr/intro.do	6
해양경비 안전교육원	해양경비 안전교육원 정보공개	http://edu.kcga.go.kr/main.jsp	7
해양수산부	어업자원포털	https://www.greensea.go.kr/main.jsp	2
해양수산부	해양안전종합 정보시스템	https://www.gicoms.go.kr/	3
해양수산부	수산정보포털	http://www.fips.go.kr/	2
해양수산 인재개발원	해양수산 인재개발원 통계자료	http://www.ofhi.go.kr/	6
해양환경관리공단	국가해양환경 정보통합시스템	http://www.meis.go.kr/rest/main	6
해외건설협회	해외건설종합 정보서비스 건설통계	http://www.icak.or.kr/sta/sta_0101.php	3
해외문화홍보원	해외문화홍보원 통계자료	http://www.kocis.go.kr/main.do	6
행정자치부	공공데이터포털	https://www.data.go.kr	2
행정자치부	행정자치부 통계	http://www.moi.go.kr/frt/sub/a05/statistics/screen.do	5
행정자치부	내고장알리미	https://www.laiis.go.kr	3
행정자치부	지방재정365	http://lofin.moi.go.kr	3
행정자치부	주민등록 인구통계	http://rcps.egov.go.kr:8081/jsp/stat/ppl_stat_jf.jsp	3
현충사관리소	현충사관리소 충무공자료실	http://hcs.cha.go.kr/cha/idx/SubIndex.do?mn=HCS	7
화학물질안전원	화학물질안전원 정보마당	http://nics.me.go.kr/	7
환경부	환경통계포털	http://stat.me.go.kr	2
환경부	환경디지털도서 관	http://library.me.go.kr	7
(사)남북교류 협력지원협회	I-RENK 북한지하자원넷	https://www.irenk.net:50015/	3
(재)APEC기후센터	APCC Data Service System	http://adss.apcc21.org/	6
(재)예술경영 지원센터	KOPIS 공연예술 통합전산망	http://www.kopis.or.kr/por/stats/perfo/perfoStats.do?menuId=MNU_00027	3
(재)중소기업연구원	중소기업 정책연구 DB	http://db.kosbi.re.kr/kosbiDB/main	7
(재)한국문화정보원	문화포털	http://www.culture.go.kr/cultureplan/statistic.do	4
(재)한국보육진흥원	보육통계	https://www.kcpi.or.kr/	5

빅데이터와 행정자료의 이해

발　행 | 2016년 8월 10일
인　쇄 | 2016년 8월 31일
발행인 | 통계교육원장 송복철
발행처 | 통계교육원
기　획 | 김상희 · 황양주 · 김나리
집　필 | 김기환 · 최보승

주　소 | (35220) 대전광역시 서구 한밭대로 713(월평동) 통계센터 5층 통계교육원
전　화 | 042 – 366 – 6252
홈페이지 | http://sti.kostat.go.kr/

발간등록번호　11–1240162–000025–01
ISBN　978–89–962019–6–0 93310

* 본서의 무단복재행위를 금합니다.

비매품

9 788996 201960 93310
ISBN 978-89-962019-6-0

통계청
통계교육원
(35220) 대전광역시 서구 한밭대로 713(월평동)
통계센터 5층 통계교육원