2013-OR-62 Working Paper

빅데이터 활용과 정보공개의 경제적 효과 분석

An Analysis of Economic Effect on Utilization of Big Data and Information Disclosure on Public Sector

최 봉

서울연구원
The Seoul Institute

2013-OR-62

빅데이터 활용과 정보공개의 경제적 효과 분석

An Analysis of Economic Effect on Utilization of Big Data and Information Disclosure on Public Sector

연구진

연구책임 최 봉 시민경제연구실 연구위원
연구원 정현철 시민경제연구실 연구원

이 보고서의 내용은 연구진의 견해로서
서울특별시의 정책과는 다를 수도 있습니다.

요약 및 정책건의

1 연구개요

1 1 연구배경, 목적, 내용

1 1 1 연구배경

- 대규모 데이터를 통칭하는 빅데이터(Big Data)라는 용어가 등장하면서 이에 대한 관심이 증가하고 있으며 공공부문에서도 빅데이터 활용에 대한 관심도가 높아지고 있음
- 빅데이터에 대한 관심 고조와 함께 활용가치가 높은 공공데이터의 이용 활성화에 대해 주목

1 1 2 연구목적

- 서울시가 시정에 빅데이터를 이용하였을 경우 절감할 수 있는 잠재적 비용 및 공공데이터 정보공개에 따른 경제적 효과 추정
- 공공데이터 활용 및 정보공개 효과 제고 방안 마련과 서울시 대용량 데이터베이스의 효율적 활용방안 모색

1 1 3 연구내용

- 서울시를 포함한 공공 및 민간부문에서의 빅데이터 활용 현황 파악
- 서울시가 시정에 빅데이터 및 공공데이터를 이용하였을 경우 얻을 수 있는 경제적 효과 추정
- 서울시 공공데이터 정보공개에 따른 경제적 효과 분석

2 주요 연구결과

2.1 서울시 공공행정 비용절감 추정

- 서울시의 공공행정이 빅데이터 활용으로 절감할 수 있는 비용은 연간 최소 7,358억원에서 최대 1조 4,474억원에 이를 것으로 추정(2014년 서울시 예산 대비 3.4%~6.7%)
- 운영효율 개선 분야에서의 절감비용은 최소 6,411억원에서 최대 1조 684억원으로 추정
  - 부정 및 오류 감소 분야에서의 절감비용은 최소 119억원에서 최대 477억원으로 추정
  - 조세제도 개선에 따른 세입 증가 분야에서의 절감비용은 최소 828억원에서 최대 3,312억원으로 추정

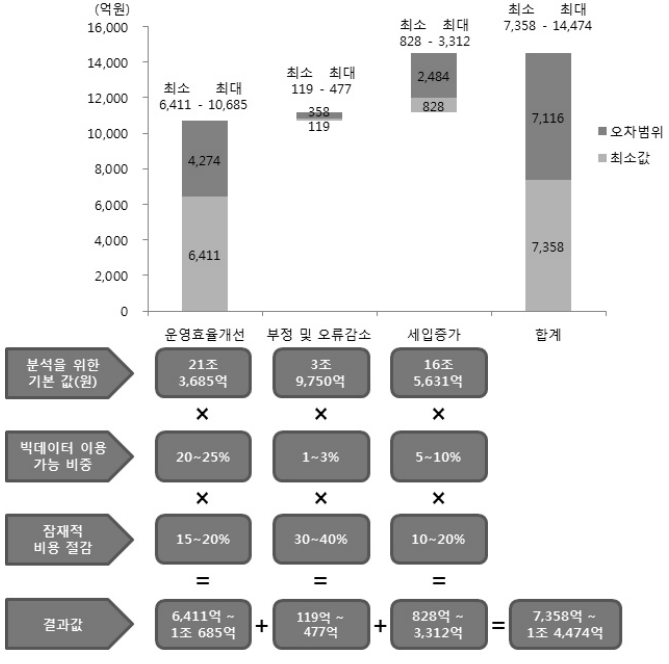


그림 1 서울시 공공부문 빅데이터 잠재력

2.2 정보공개에 따른 민간부문 경제적 파급 효과

- ㅇ 공개된 공공정보는 DB서비스 기업들에 의해 지식정보 서비스와 광고서비스의 형태로 이용자들에게 제공되어 매출을 발생(투입값)
- ㅇ 산업연관표의 부문분류표를 이용해 통합소분류를 추출
  - 산업연관표 상의 소분류에서 공공정보를 이용한 민간활용 서비스는 부가통신 및 정보서비스(142), 컴퓨터관련서비스(153), 광고(151)와 밀접한 관련이 있음
- ㅇ 투입값을 산출하여 산업연관분석에 이용
  - 투입값은 부가통신 및 정보서비스(142)에 3,094억원, 컴퓨터관련서비스(153)에 1,397억원, 광고(151)에 3,570억원 등 총 8,061억원

표 1 산업연관분석을 위한 투입값 산출 결과

연도	DB서비스	산업분류	투입(억원)
2013	정보서비스	142(부가통신 및 정보서비스)	3,094
		153(컴퓨터관련서비스)	1,397
	광고	151(광고)	3,570
	계	-	8,061

- ㅇ 2011년 산업연관표를 이용하여 서울시 공공정보 민간활용에 따른 경제적 파급효과를 계산
  - 생산유발효과 1조 7,775억원, 부가가치유발효과 1조 6,122억원, 취업유발효과 8,920명

표 2 서울시 공공정보 민간활용에 따른 경제적 파급효과

(단위 : 억원)

연도	DB서비스	산업분류	투입	생산유발효과	부가가치유발효과	취업유발효과
2013	정보서비스	142*	3,094	5,520	6,188	1,997명
		153**	1,397	2,388	2,794	1,744명
	광고	151***	3,570	9,866	7,140	5,179명
	계	-	8,061	17,775	16,122	8,920명

*142 부가통신 및 정보서비스 **153 컴퓨터관련서비스 *** 151 광고

2.3 서울시 역점 추진사업 효과분석

2.3.1 서울시 정보기획단 추진 사업

- ㅇ N버스(심야버스, 일명 올빼미버스) 도입으로 인한 서울시민 교통비 절감[1]
  - – 야간통행 이동비용 절감
    - • 심야시간에 택시를 이용해야만 했던 시민들이 대체수단인 야간버스를 이용
    - • 1인당 약 7,050원, 1일 4,440만원, 연간 162억원의 절약효과
  - – 유류비 절감효과
    - • 심야버스 이용으로 승용차 운행 수요를 감소시켜 유류비용을 절감
    - • 1일 665만원, 연간 24억원의 유류비 절감 효과
- ㅇ N버스(올빼미버스) 도입으로 인한 환경개선 효과
  - – 이산화탄소 배출 감소
    - • 승용차 운행을 버스 운행으로 대체함으로써 이산화탄소 배출량 감소
    - • 1일 1,365kg, 연간 498톤의 이산화탄소 배출 억제 효과

2.3.2 정보기획단 추진 예정 사업

1) 상권분석

- ㅇ 예비창업자에게 상권에 대한 정보를 제공함으로써 창업성공률을 높임
- ㅇ 기존점포에 대한 평가를 통해 객관적인 정보를 얻을 수 있고 이를 토대로 창업성공률을 제고

2) 택시 매치메이킹

- ㅇ 수요자인 승객과 공급자인 택시기사들의 정보를 매칭해서 수급불균형을 해소하여 택시 이용의 편리성을 높이고 탑승률을 높임

---

1 부록 참고

2.3.3 시정 집중·핵심사업

1) **공공의료 강화**

- 의료보장 적용인구, 환자 현황, 의료기관 현황 등을 이용하여 공공의료 강화에 이용

2) **여성안전 종합대책, 도시안전마스터플랜**

- 범죄발생현황, 가정폭력 현황, 화재현황, 건축물 현황 등을 이용하여 여성안전 종합대책, 도시안전마스터플랜 수립에 활용

3) **지역맞춤형 수방대책, 태양광 발전 보급**

- 기상관측정보(강수량, 일조량 등)를 이용하여 지역맞춤형 수방대책, 태양광 발전 보급에 이용

4) **채무관리 강화**

- 예산자료, 회계자료, 세금자료, 지방채 등을 이용하여 채무관리 강화

3 정책제언

3.1 공공데이터 활용 및 정보공개 효과 제고 방안

3.1.1 양질의 데이터 제공을 위해 노력

- 서울시는 가치 있는 공공정보의 공개를 위해 노력
  - 단순 문서가 아닌 가치 있고 유의미한 공공정보를 공개
  - 가치 있는 공공정보 발굴을 위해 노력
- 공공데이터의 품질 향상을 위한 오류율 저감 노력이 필요
  - 데이터 개방을 기피하는 이유는 심각한 오류율 때문이므로 이를 보완할 수 있는 방안이 필요

- 서울시는 공공데이터 품질의 업그레이드를 위해 공공정보 개방 건수의 확대보다 데이터의 품질 향상에 관심을 기울여야 함
  - 서울시 부서 간 또는 서울시와 외부 공공기관 간 협의를 통해 융합공공데이터를 제공할 수 있도록 노력해야 함
- 서울시 조직 간(예 : 도시계획국과 교통국 등) 융합데이터를 생성
- 서울시와 국가공공기관(예 : 경찰청, 국세청, 안행부 등)이 공동으로 융합데이터를 생성

312 **공공데이터 활용의 활성화 유도**

- 공공데이터 이용활성화를 유도하기 위해서는 관련 법령의 정비가 필요하며 서울시는 관련 법률 개정을 적극적으로 건의
- 국내에는 데이터를 다룰 수 있는 전문가는 많지만 빅데이터를 다룰 수 있는 전문인력은 부족한 실정으로 이들의 양성이 시급
- 빅데이터 큐레이터 양성 프로젝트[2]의 참여자를 늘리고 이들에게 적극적으로 일자리를 매칭
- 산·관·학 협력 활성화를 통해 공공과 민간의 빅데이터 활용현장의 실무경험을 쌓을 수 있는 기회를 제공
- 개방부터 활용까지 모든 절차에서 수요자인 시민이 편하게 활용할 수 있도록 정보이용 방법을 개편
- 사용하기가 불편하면 공개된 정보의 이용률이 떨어짐
- 서울 열린데이터 광장의 사용자인터페이스(User Interface)에 대해 지속적인 관심과 개선이 필요

2 서울시는 청년 미취업자를 대상으로 전문가의 지도로 전문지식과 직무경험을 축적하고 서울시 전략수립 및 신규사업에 활용 가능한 빅데이터 기반 과제 발굴을 위해 서울형 뉴딜 일자리사업의 일환으로 '빅데이터 큐레이터 양성 프로젝트'를 시행

3.1.3 정보보안 강화

- ㅇ 카드정보 유출로 인한 개인정보 보호에 관심이 증가하고 있는 상황에서 정보보안은 공공정보공개와 밀접한 연관이 있음
- ㅇ 우리나라의 프라이버시에 대한 규제는 매우 엄격한 편으로 데이터를 활용한 상당수의 서비스가 위법에 속하므로 공공정보를 공개할 경우 주의를 요함
  - 민간에서 필요한 상당수 공공정보가 개인정보보호법을 적용받아 개방이 어렵기 때문에 민감한 정보는 제거하고 개방할 수 있도록 데이터를 가공
  - 민간에서 활용할 것을 고려하여 공공데이터 공개 범위를 결정
- ㅇ 공공정보에 개인정보 항목을 삭제 또는 최소화하여도 의미 있는 데이터가 될 수 있도록 하는 방안 모색
  - 예를 들어, 개인의 특성이 담긴 자료를 지도와 융합한다면 민감한 개인정보(주민등록번호 등)를 제거해도 의미 있는 분석이 가능함

3.2 서울시 대용량 데이터베이스의 효율적 활용방안

3.2.1 공공정보 공개에 따른 경제적 효과 극대화

- ㅇ 공공정보를 활용한 광고, 정보서비스 등의 관련 산업 발전을 위한 DB산업 지원책을 마련
- ㅇ 연구조직(연구원 등) 및 정책부서(서울시 등)와 긴밀한 업무협력을 통해 공공정보의 활용방안을 다각도로 모색하고 정책연구에 활용
  - 다양한 빅데이터 및 공공데이터를 융합하여 활용방안을 마련하기 위한 연구가 필요
  - 연구조직과 정책부서의 협력을 통해 빅데이터 및 공공정보 이용의 시너지효과 제고

3 2 2 **시정 집중·핵심사업에 공공정보 및 빅데이터 이용**

- ㅇ 시정사업에 빅데이터를 활용하면 운영 효율성이 증대되어 비용절감 효과가 있을 것으로 판단
  - 기존 정책결정방식의 효율성을 증대시키고 취약점을 보완할 수 있음
- ㅇ 면밀한 분석을 통해 핵심사업과 빅데이터를 결합하여 정책수립 방향을 설정
  - 서울시 주요 핵심사업에 빅데이터를 적용할 수 있는 분야를 지속적으로 탐색 후 사업에 적용
  - 빅데이터 큐레이터와 협업하여 새로운 정책사업을 발굴

3 2 3 **민간부문 데이터 확보**

- ㅇ 다양하고 긴밀한 분석을 위해서는 공공부문뿐만 아니라 민간부문의 데이터가 필요함
  - 공공부문 데이터와 민간부문 데이터 융합으로 시너지효과 기대
  - 민간과 공공이 협력하여 플랫폼을 구축하고 업무를 공유
- ㅇ 민간기업과의 교류 및 협력으로 민간데이터를 확보하여 이용할 수 있도록 해야 함
  - 공공부문은 민간부문에 의미 있는 사업을 제안하여 이들이 보유하고 있는 빅데이터(민간데이터) 확보를 위해 노력

3 2 4 **빅데이터 자문위원회 상설화**

- ㅇ 빅데이터 자문위원회의 상설화 방안 모색
  - 17개 시·도에 창조경제혁신센터가 개설될 예정
  - 창조경제혁신센터 산하에 빅데이터 자문위원회를 설치하여 지역 실정에 맞는 신규 정책발굴 등의 업무를 수행
  - 공공데이터 개방 전담팀인 정보기획단과 긴밀한 협업이 가능하도록 교류 활성화를 추진

- 빅데이터 자문위원회는 빅데이터 전문가 위주로 구성되어야 하며, 위원장은 반드시 빅데이터 전문가가 맡아야 함

3.3 빅데이터를 활용한 정책 제안

3.3.1 사람을 찾습니다

- 서울시의 '사람을 찾습니다' 도입 방향
  - 사진, 인적사항 등의 정보를 매칭하여 가출, 미아, 치매노인 등 실종자를 찾기 위한 시스템 구축 방안을 모색
  - 서울시, 안전행정부, 경찰청, 법무부, 어린이재단 등과 실종자 정보를 공유하여 실종자 DB를 구축
- 실종자 찾기를 통해 사회적 불안 및 비용 감소 등의 효과를 얻을 수 있을 것으로 기대

3.3.2 내가 낸 세금은 어디에 쓰이나

- 서울시의 '내가 낸 세금은 어디에 쓰이나' 도입 방향
  - 서울시는 예산 등의 정보를 공개하고 있으며 이를 이용하여 시민들이 납부한 세금이 어디에 쓰이는지 알 수 있도록 웹페이지 구축
  - 제공하는 정보를 시민들이 알기 쉽게 하기 위해 시각화함
- 시민들이 낸 세금에 대한 사용출처를 명확히 알 수 있으므로 시정에 대한 신뢰도 향상을 기대할 수 있음
  - 개인의 수입을 입력하면 세금납부액과 세금이 사용되는 항목을 알려줌으로써 납세에 대한 이해를 높여주고 세금의 사용처를 명확하게 인지할 수 있음

차례

Ⅰ	**연구개요**	**18**
1	**연구배경 및 목적**	18
1 1	연구배경	18
1 2	연구목적	21
2	**연구내용 및 방법**	22
2 1	연구내용	22
2 2	연구방법	22
3	**빅데이터란 무엇인가**	23
3 1	빅데이터의 정의	23
3 2	빅데이터의 특징	23
3 3	빅데이터와 공공데이터의 관계	25
Ⅱ	**빅데이터 활용과 정보공개**	**28**
1	**국내 공공부문 빅데이터 활용 현황**	28
1 1	공공부문 정보공개 현황	28
1 2	공공데이터 포털	28
1 3	국내 공공부문 빅데이터 활용 사례	30
1 4	부처별 공공정보 활용사례	32
2	**국내 민간부문 빅데이터 활용 현황**	34
2 1	민간에서 이용되는 공공정보	34
2 2	국내 기업의 빅데이터 활용 사례	35
3	**해외사례**	36
3 1	해외 공공부문 빅데이터 활용 현황	36
3 2	해외 민간부문 빅데이터 활용 현황	38

III	**서울시 공공데이터 정보관리 및 정보공개 현황**	**44**
1	**서울시 공공데이터 현황**	44
1 1	서울 열린데이터 광장	44
1 2	서울시 공공정보 공개 현황	48
2	**공개된 정보의 활용 현황**	50
2 1	서울시 공공정보 활용 현황	50
2 2	추후 활용방안	53
3	**정보공개의 경제적 효과 분석**	53
3 1	개요	53
3 2	서울시 공공부문 빅데이터 잠재력	54
3 3	정보공개에 따른 민간부문 경제적 파급 효과	57
3 4	서울시 역점 추진사업 효과분석	59
IV	**정책제언**	**64**
1	**공공데이터 활용 및 정보공개 효과 제고 방안**	64
1 1	양질의 데이터 제공을 위해 노력	64
1 2	공공데이터 활용의 활성화 유도	65
1 3	정보보안 강화	66
2	**서울시 대용량 데이터베이스의 효율적 활용방안**	66
2 1	공공정보 공개에 따른 경제적 효과 극대화	66
2 2	시정 집중·핵심사업에 공공정보 및 빅데이터 이용	67
2 3	민간부문 데이터 확보	67
2 4	빅데이터 자문위원회 상설화	68
3	**빅데이터를 활용한 정책 제안**	68
3 1	사람을 찾습니다	68
3 2	내가 낸 세금은 어디에 쓰이나	69
	참고문헌	**72**
	부 록	**76**

표차례

표 1-1	주요 기관별 빅데이터 정의	23
표 1-2	기존데이터와 빅데이터의 비교	24
표 2-1	공공데이터 신청건수	29
표 2-2	공공데이터 활용건수	30
표 2-3	국가정보화전략위원회 선정 빅데이터 대상과제	31
표 2-4	2013년 미래창조과학부 빅데이터 시범사업 선정 과제	32
표 3-1	서울시 공공데이터 연도별 개방목표	48
표 3-2	서울시 열린데이터광장 분류체계	49
표 3-3	서울시 공공데이터 개방 현황(2014년 3월 기준)	49
표 3-4	산업연관분석을 위한 투입값 산출 결과	59
표 3-5	서울시 공공정보 민간활용에 따른 경제적 파급효과	59
표 4-1	공공데이터 개방 관련 정비대상 주요법령	65

그림차례

그림 1-1 국내 빅데이터 시장 규모 전망(KISTI) 19
그림 1-2 국내 공공정보 민간활용 시장 전망 20
그림 1-3 데이터 관계도 25
그림 3-1 공공데이터 활용 45
그림 3-2 서울 열린데이터 광장 월별 활용 현황 47
그림 3-3 서울시 공공정보를 활용하여 만들어진 앱 51
그림 3-4 서울시 공공정보를 활용하여 만들어진 인포그래픽 52
그림 3-5 서울시 공공부문 빅데이터 잠재력 56
그림 3-6 공공정보 민간활용 산업의 가치사슬 57
그림 4-1 '내가 낸 세금은 어디에 쓰이나' 이미지 70

Ⅰ 연구개요

1 연구배경 및 목적
2 연구내용 및 방법
3 빅데이터란 무엇인가

Ⅰ 연구개요

1 연구배경 및 목적

1.1 연구배경

1.1.1 빅데이터 시대의 도래

- ㅇ 최근 디지털 정보량이 기하급수적으로 증가하면서 무한정에 가까운 정보가 실시간으로 만들어지고 있음
  - – 하루에도 수천 개에서 수억 개의 데이터가 쏟아지고 있으며 인터넷이 일상화된 최근 10년 사이 디지털 데이터가 폭증하고 있음
  - – 스마트폰의 이용이 보편화되면서 개인의 활동, 위치정보 등 새로운 유형의 정보가 실시간으로 생성되고 있음
- ㅇ 대규모 데이터를 통칭하는 빅데이터(Big Data)라는 용어가 등장하면서 이에 대한 관심이 증가하고 있음
  - – 머브 에이드리언(2011)은 데이터 홍수(Data Deluge)의 시대에 빅데이터를 이용한 사업기회 증가에 주목해야 한다고 언급
  - – 창조경제 시대에 적합한 아이템으로 빅데이터를 이용한 사업에 관심이 모아지고 있음
- ㅇ 세계 빅데이터 시장뿐만 아니라 국내시장의 규모도 점차 확대되는 추세(KISTI, 2013)
  - – 빅데이터의 세계시장 규모는 기관마다 차이를 보이고 있으나, 매년 약 39~60% 성장하여 2015년에는 169~321억달러에 이를 것으로 전망
  - – 국내시장 규모도 2013년 1억 6,300만달러에서 지속적으로 성장하여 2020년에는 8억 9,400만달러에 이를 것으로 예측됨

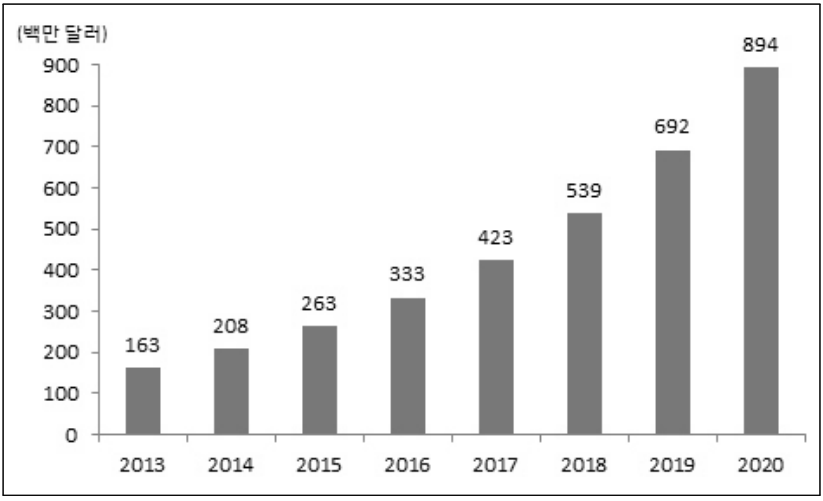


그림 1-1 국내 빅데이터 시장 규모 전망(KISTI)

1.1.2 공공부문과 빅데이터

- 민간부문뿐만 아니라 공공부문에서도 빅데이터 활용에 대한 관심이 높아지고 있음
  - 2013년 11월 제28차 경제관계장관회의에서 미래창조과학부는 빅데이터를 창조경제 시대의 핵심동력으로 선정하고 관련산업을 육성하기 위한 '빅데이터 산업 발전전략'을 관계부처와 합동으로 발표
- 빅데이터에 대한 관심 고조와 함께 활용가치가 높은 공공데이터의 이용 활성화에 대해 주목
  - 공공정보는 공공영역을 넘어 사적 영역의 다양한 사회·경제적 부가가치 창출에 활용될 수 있는 잠재력을 보유(한국데이터베이스진흥원, 2013)
  - 정부 및 지자체는 웹페이지 구축 등을 통해 공공데이터의 일반 공개를 적극 추진
  - 미국과 영국 정부는 공공데이터를 마이닝(Mining)하고 시각화해서 국민(시민)들에게 대용량의 데이터를 쉽게 이해할 수 있도록 제공

- 공공정보의 민간활용 시장도 지속적인 성장이 예측됨
  - 공공정보의 민간 활용시장 규모는 2013년 1조 348억원에서 연평균 11% 성장하여 2017년에는 1조 5,711억원이 될 것으로 추정

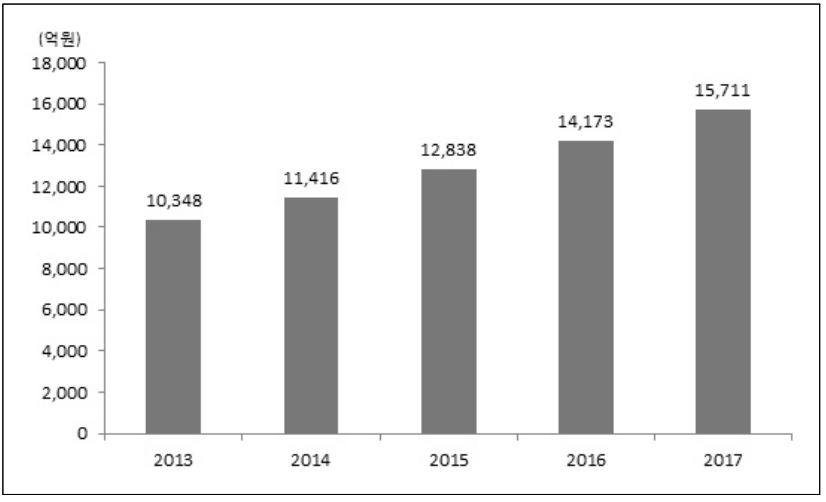


그림 1-2 국내 공공정보 민간활용 시장 전망
자료 : 허필선 외(2013)

- 공개된 공공데이터를 이용하면 효율적인 정책수립뿐만 아니라 새로운 사업기회를 발굴할 수 있음
  - 공공부문은 조직 운영 효율성 증대, 제도 개선 등을 통해 비용을 절감할 수 있음
  - 민간부문은 공공데이터를 활용하여 사업기회를 포착

113 **서울시의 빅데이터 활용 움직임**

- 최근 N버스(심야버스, 일명 올빼미버스)의 사례를 통해 서울시의 빅데이터를 활용한 정책수립이 주목받고 있음
  - 서울시는 빅데이터를 활용하여 심야버스 노선을 개발

- N버스 사례는 2013년 9월 12일 지방행정정보화 연찬회에서 빅데이터를 사용한 우수정책 사례로 선정되어 대통령상을 수상
  - 서울시는 빅데이터 활용을 적극 모색
- 교통, 산업경제, 도시관리, 안전, 문화관광 등의 분야에서 서울시 정보기획단이 데이터를 수집하고 이를 이용한 사업을 지속적으로 발굴하려고 노력하고 있음
- 서울시는 빅데이터 활용도를 높이기 위하여 '빅데이터 큐레이터 양성 프로젝트'를 통해 관련 인력을 양성
  - 서울시는 공공정보의 공개를 통해 정보접근성을 높임
- 서울 열린데이터광장(data.seoul.go.kr)을 통해 공공데이터를 공개하여 (프로그램)개발자뿐만 아니라 시민들이 공공정보에 수월하게 접근할 수 있도록 함
  - 서울시가 보유하고 있는 공공데이터의 공개에 따른 영향 분석이 필요한 시점
- 빅데이터를 활용한 우수사례를 널리 홍보할 필요가 있음
- 공공데이터 정보공개의 경제적 파급효과 분석을 통해 공개 당위성을 알리고 공공데이터의 활용을 촉진해야 함

1.2 연구목적

- 서울시가 시정에 빅데이터를 이용하였을 경우 잠재적으로 절감할 수 있는 비용 추정
- 서울시 공공데이터 정보공개에 따른 경제적 효과 분석
- 서울시가 보유하고 있는 공공정보를 공개하고, 공공정보를 민간에서 활용하였을 경우 발생하는 경제적 파급효과를 추정
- 공공데이터 활용 및 정보공개 효과 제고 방안 마련
- 서울시 대용량 데이터베이스의 효율적 활용방안 모색

2 연구내용 및 방법

2.1 연구내용

- 국내외 공공 및 민간부문의 사례분석을 통한 빅데이터 활용 현황 파악
- 서울시 공공데이터 정보공개 및 활용 현황 검토
- 서울시가 빅데이터 및 공공데이터를 이용해서 얻을 수 있는 경제적 효과 분석
  - 서울시가 빅데이터를 이용해서 잠재적으로 절감 가능한 비용 추정
  - 서울시 역점 추진사업 중 일부사업을 선별하여 공공데이터를 이용한 시민 편익에 대한 진단
- 서울시의 공공데이터 정보공개에 따른 경제적 효과 분석
  - 생산유발, 부가가치유발, 취업유발 창출 등 경제적 효과 파악
- 정책제언
  - 공공데이터 활용을 활성화하고 정보공개 효과를 극대화할 수 있는 방안 제안
  - 서울시가 보유하고 있는 대용량 공공데이터의 효율적인 활용을 위한 방향을 제시

2.2 연구방법

- 문헌조사
- 실 데이터 조사 및 분석
- 전문가 자문 및 인터뷰

3 빅데이터란 무엇인가

3.1 빅데이터의 정의

- 빅데이터는 초기에 단순히 방대한 데이터의 집합을 의미하였으나 최근에는 데이터 자체뿐만 아니라 데이터를 관리하기 위한 하드웨어, 이를 기반으로 하는 응용프로그램과 이용방법을 포함하는 개념으로 사용
  - 데이터를 구성하고 있는 물리적 하드웨어로부터 시작해 이를 기반으로 하는 애플리케이션과 소프트웨어를 포괄하는 거대한 플랫폼(미래창조과학부, 2013)

표 1-1 주요 기관별 빅데이터 정의

기관	정의
국가정보화 전략위원회	대량으로 수집한 데이터를 활용·분석하여 가치 있는 정보를 추출하고 생성된 지식을 바탕으로 능동적으로 대응하거나 변화를 예측하기 위한 정보화 기술
삼성경제연구소	기존의 관리 및 분석 체계로는 감당할 수 없을 정도의 거대한 데이터의 집합을 지칭하며, 대규모 데이터와 관계된 기술 및 도구(수집·저장·검색·공유·분석·시각화 등)도 빅데이터의 범주에 포함됨
맥킨지	일반적인 데이터베이스 체계가 저장, 관리, 분석할 수 있는 범위를 초과하는 규모의 데이터

자료 : 2012 국가정보화백서(한국정보화진흥원), 한국데이터베이스진흥원

3.2 빅데이터의 특징

- 기존데이터와 빅데이터는 데이터의 양, 데이터의 유형, 프로세스 및 기술에서 차이가 있음
  - 기존데이터는 테라바이트(TB)[3] 수준에 불구하지만 빅데이터는 최소 페타바이트(PB)[4] 수준의 데이터 양이 전제되며 그 크기가 끊임없이 변함

3 1 테라바이트(TB) = 10^{12} bytes = 1,000,000,000,000 bytes

4 1 페타바이트(PB) = 10^3테라바이트(TB) = 10^{15} bytes = 1,000,000,000,000,000 bytes

- 기존데이터는 정형화된 구조 위주인 반면, 빅데이터는 비정형화된 동영상, 사진데이터 등의 비중이 높음
- 기존데이터는 처리 및 분석과정이 정형화되어 있으며 원인과 결과 규명 중심이었으나 빅데이터는 상관관계를 규명하는 것으로 비정형화되어 있으며 처리 복잡도가 매우 높음

표 1-2 기존데이터와 빅데이터의 비교

구분	기존 데이터	빅데이터
데이터 양	테라바이트 수준	페타바이트 수준(최소 100테라바이트 이상)
데이터 유형	정형데이터 중심	정형데이터 외에도 소셜 미디어 데이터의 동영상, 사진, 대화 내용, 검색어 통계 등 비정형 데이터 비중이 높음
프로세스 및 기술	처리/분석 과정이 정형화 원인/결과 규명 중심	다양한 데이터 소스, 복잡한 로직 처리 등 처리 복잡도가 매우 높음 상관관계 규명 중심

자료 : 배동민 외(2013)

- 빅데이터는 거대한 데이터 양(volume), 빠른 데이터 이용 속도(velocity), 다양한 데이터 종류(variety) 등 3V로 규정됨
  - 과거의 데이터가 테라바이트(TB) 수준이라면 빅데이터는 페타바이트(PB) 이상의 규모로 이전에 비해 양이 많아짐(Volume)
  - 실시간 정보로 활용하기 위해서는 대량 데이터의 처리 및 분석 속도가 중요(Velocity)
  - 과거 정형화된 데이터를 중심으로 분석이 진행되었다면 빅데이터는 숫자, 문자, 영상데이터 등 비정형화된 것들도 포함(Variety)
- 최근 여기에 새로운 빅데이터 요소를 추가하여 새롭게 정의하려는 움직임이 나타남
  - IBM은 진실성(Veracity)을, 브라이언 홉킨스(Brian Hopkins)는 가변성(Variability)을 추가하여 새롭게 "4V"를 정의(미래창조과학부, 2013)

- 진실성(Veracity), 시각화(Visualization), 가치(Value)의 3V가 새로운 빅데이터 요소로 추가되면서 6V로 규정(범지인 외, 2013)
  - 빅데이터 분석을 통해 비즈니스 의사 결정이나 활동의 배경을 고려하여 이용되므로 신뢰가 중요함(Veracity)
  - 사용자가 이용하기 쉽도록 시각적인 기능을 통해 빅데이터의 활용성을 높여 모든 잠재력이 활용될 수 있도록 해야 함(Visualization)
  - 비즈니스에 사용되기 때문에 분석된 자료는 궁극적으로 가치의 창출로 이어져야 함(Value)

3.3 빅데이터와 공공데이터의 관계

- 빅데이터는 대용량의 다양한 데이터 세트로 지속 생성되고 있으며 일부 정보는 공개되지 않고 있음
- 공개된 데이터 중에서 공공부문이 만들어낸 정보는 열린 정부 데이터
- 공개된 데이터의 생성 주체는 정부, 기관, 회사, 개인 등
- 최근 공공에 의해 만들어진 열린 데이터에 대한 관심이 높아지고 있음

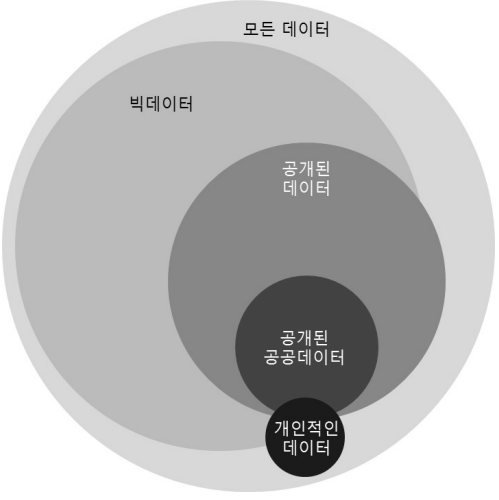


그림 1-3 데이터 관계도
출처 : MGI(2011)

II 빅데이터 활용과 정보공개

1 국내 공공부문 빅데이터 활용 현황
2 국내 민간부문 빅데이터 활용 현황
3 해외사례

II 빅데이터 활용과 정보공개

1 국내 공공부문 빅데이터 활용 현황

1 1 공공부문 정보공개 현황

- ㅇ 공공기관 정보공개에 관한 법률에 따라 일부를 제외하고는 모두 공개하도록 되어 있음
  - – 비공개대상은 공공기관 정보공개에 관한 법률에 적용되지 않는 정보
    - • 국가안전보장에 관련되는 정보
    - • 보안 업무를 관장하는 기관이 국가안전보장과 관련된 정보의 분석을 목적으로 수집하거나 작성한 정보
  - – 공개대상은 문서, 도면, 사진 등 다양한 매체를 포함하지만 원 데이터에 대한 언급은 없음
- ㅇ 공공데이터포털(data.go.kr)을 통하여 정부 및 산하기관의 데이터를 공개하고 있음

1 2 공공데이터 포털

- ㅇ 국가가 보유하고 있는 다양한 공공정보를 국민에게 개방하여 이를 편리하고 손쉽게 활용할 수 있도록 구축한 시스템

1 2 1 운영주체

- ㅇ 안전행정부의 공공정보정책과가 운영하면서 관련 업무를 주관
- ㅇ 한국정보화진흥원이 공공데이터포털과 공공정보활용지원센터를 위탁운영

1 2 2 주요 추진과제

- ㅇ 정보공개의 패러다임 전환 주도
  - – 모든 정보는 공개를 원칙으로 하고 비공개 범위를 최소화
  - – 가능한 한 모든 원천데이터를 민간에 개방
- ㅇ 공공정보 개방 및 공유를 위한 공통기반 조성
  - – 공공정보의 민간활용이 용이하도록 DB를 개방형·표준형으로 전환
  - – 범정부 단일 정보 공유 플랫폼(data.go.kr) 구축
- ㅇ 공공정보의 민간 활용 지원을 통한 창조적 일자리 창출
  - – 1인 창조기업, 사회적기업 대상 공공정보 활용 지원
  - – 공공데이터 융합에 따른 새로운 가치 창출 등을 통해 비즈니스 모델 발굴

1 2 3 개방 및 활용현황

- ㅇ 현재 781개 기관이 8,197개의 데이터를 제공(2014년 3월 31일 현재)
  - – 8,197개의 데이터 중 데이터셋 7,670개, Open API[5] 527개
- ㅇ 프로그램 개발자, 일반시민 등이 공공데이터를 신청하는 건수가 급속도로 증가
  - – 2011년 1,987건에 불과하던 신청건수는 2013년 1만 3,938건으로 급증

표 2-1 공공데이터 신청건수

구분	2011년	2012년	2013년
오픈 API	1,650	2,602	9,826
데이터	337	2,020	4,112
소계	1,987	4,622	13,938

자료 : 공공데이터포털(data.go.kr)

5 Open Application Programmer Interface의 약자로 직접 응용프로그램을 개발할 수 있도록 데이터를 공개한 것이며 누구나 접근하여 사용할 수 있다는 장점이 있음

- 공공데이터 활용건수도 신청건수와 마찬가지로 급속도로 증가
  - 2011년 1억 2천만건에서 2013년에는 16억건으로 14배에 육박하는 증가세를 보임

표 2-2 공공데이터 활용건수

구분	2011년	2012년	2013년
오픈 API	120,433,377	1,092,573,370	1,597,109,616
데이터	327	2,020	4,040
소계	120,433,704	1,092,575,390	1,597,113,656

자료 : 공공데이터포털(data.go.kr)

1.3 국내 공공부문 빅데이터 활용 사례

1.3.1 국가정보화전략위원회(현 정보통신전략위원회)

- 국가정보화전략위원회는 빅데이터를 활용한 스마트 정부구현(안)과 관련하여 '12년 11월 빅데이터 마스터플랜 수립
  - 교육과학기술부, 행정안전부, 지식경제부, 방송통신위원회, 국가과학기술위원회와 협의하여 계획 수립
  - 6개 분야에서 16개 빅데이터 대상과제를 제시

표 2-3 국가정보화전략위원회 선정 빅데이터 대상과제

분야	빅데이터 대상 과제	주관부처
사회 안전	범죄발생 장소·시간 예측을 통한 범죄발생 최소화	경찰청, 행안부, 법무부, 검찰청
	예측기반의 자연재해 조기 감지 대응	방재청, 경찰청, 기상청, 행안부
	음란물 유통차단을 통한 건강한 인터넷문화 조성	방통위, 행안부, 경찰청, 여가부
국민 복지	민원 데이터 분석을 통한 정책의 환류 시스템 마련	권익위, 각 부처, 지자체
	복지 수요·공급 매칭을 통한 맞춤형 서비스 제공	복지부, 노동부, 지자체, 국세청
	일자리 현황 분석·예측으로 고용정책 수립	노동부, 중기청, 교과부
국가 경제	과세 데이터 분석으로 탈세방지 및 국가 재정 확충 지원	국세청, 기재부
	다양한 경제 관련 데이터 분석 기반의 경제정책 수립 지원	기재부, 지경부
국가 인프라	주민참여형 교통사고 감소체계 구축	경찰청, 국토부, 지자체
	실시간 네트워크시스템 재난 관리·대응체계 마련	방통위, 행안부
산업 지원	자영업자 창업 실패 예방 지원	중기청, 노동부, 지경부, 지자체
	제조공정 실시간 장애 예측을 통한 생산효율 고도화 수급 전망에 기반을 둔 농수산물 생산 관리	지경부, 교과부 농림부, 기상청
과학 기술	국가 기후 위험요소에 대한 선제적 대응 체계 구축	기상청, 국토부, 산림청
	유전자·의료 데이터 분석을 통한 국민 건강 증진	교과부, 복지부, 지경부, 식약청
	위성영상 데이터 분석·활용을 통한 재난 대응	교과부, 기상청, 국토부

자료 : 국가정보화전략위원회(2012) ; 빅데이터 기반 융합 서비스 창출 주요 정책 및 시사점(2013)

132 **미래창조과학부**

- 미래창조과학부는 2013년 총 14억원의 예산으로 빅데이터 활용 스마트 서비스 시범사업 6개를 시행
  - 빅데이터 기반 서비스를 발굴하고 확산시켜 공공 및 민간서비스의 경쟁력을 높이고 이를 통한 일자리 및 비즈니스 창출을 지원하기 위함

표 2-4 2013년 미래창조과학부 빅데이터 시범사업 선정 과제

컨소시엄	서비스명	내용
㈜KT (서울특별시)	심야버스 노선 정책 지원	서울시의 교통데이터와 KT의 유동인구 데이터를 연계, 분석하여 최적의 심야버스 노선 정책 수립 지원
국민건강보험공단 (다음소프트)	국민건강 주의 예보 서비스	국민건강보험공단의 건강보험 DB와 SNS 정보를 연계하여 홍역, 조류독감, SAS 등 감염병 발생 예측 모델을 개발하고 주의예보
에스지에이㈜ (한국의약품안전관리원, ㈜와이즈넛)	의약품 안전성 조기경보 서비스	유해사례 DB, 진료기록, SNS 등을 분석하여, 유의 의약품을 추출하고, 병의원, 제약회사 및 유관기관 등에 위험도 예측 서비스 제공
서울아산병원(ETRI, 한국MS, ㈜테크아이, 켐아이넷㈜, 한국쌔스소프트웨어)	심실부정맥 예측 등 보건의료 서비스	포털(다음), 질병관리본부 데이터와 병원 자체 데이터를 활용하여 독감 유행 예측, 심실부정맥 예측, 입원 병상 배정 최적화 등 제공
㈜오픈메이트 (비씨카드㈜, 한국감정원)	소상공인 창업성공률 제고를 위한 점포이력 분석 서비스	카드거래, 부동산, 상가이력 정보 등의 연계 분석을 통해 창업관련 과거/현황 분석 및 예측 정보를 제공하여 소상공인 창업지원
차세대융합 기술연구원 (서울대 융합과학 기술대학원, 이스플러스)	모바일을 통한 지능형 뉴스검색 서비스	대량의 기사 DB에 대해 중요도, 관계도 등 다각도의 고급분석을 적용하여 지능형 뉴스 검색 서비스 제공

자료 : 미래창조과학부(2013)

1.4 부처별 공공정보 활용사례

1.4.1 국토교통부 국가공간정보 통합데이터베이스 구축

- ㅇ 공간정보의 수요가 증가하고 있는 상황에서 다양한 분야의 GIS(지리정보체계)정보를 연계하여 통합할 필요성이 제기
  - – 중앙부처, 지자체, 공공기관 등이 토지, 도로, 지하시설물, 통계 등의 정보를 따로 관리하였음
  - – 개별적으로 구축된 정보를 공동으로 활용할 수 있게 통합데이터베이스를 구축
- ㅇ 주요 추진 내용
  - – 흩어져 있는 공간정보에 대해 지속적으로 연계를 추진하여 통합체계를 구축

- 택지정보, 등산로정보, 교통CCTV 등 민간 활용수요가 높은 공간정보의 공개
- 공간 빅데이터의 융합 데이터베이스를 바탕으로 새로운 가치 창출이 가능한 분석모형 개발

○ 성과 및 시사점

- 국민들이 국가공간정보를 편리하게 이용할 수 있게 함
- 공간정보와 행정정보의 융합으로 교통이나 주택문제 관련 행정서비스에서 효율성과 질의 향상이 가능

142 **경찰청 지리적 프로파일링 시스템(GeoPros)**

○ 한국 실정에 맞는 지형공식이 없어 범죄수사의 범위를 줄여줄 지리적 프로파일링 시스템 개발의 필요성 대두

- 미국의 지리 프로파일링 시스템(CrimeStat)에 국내 지역 특성과 사건 데이터를 적용

○ 주요 추진내용

- 범죄다발지역과 위험도를 지도에 각기 다른 색으로 표시하여 범죄 위험 지역을 알아보기 쉽게 시각화
- 기존에 발생한 범죄데이터를 분석하여 추가 범죄의 가능성과 방향성을 예측하여 수사 및 범죄예방에 우선 활용

○ 성과 및 시사점

- 강력 범죄 발생의 감소 및 예방

• 부산연제경찰서(2012)에 따르면 2012년 8월 1개월간 5대 범죄가 전년 동기 대비 10.5% 감소(2011년 296건에서 2012년 265건으로 31건 감소)

- 순찰 현장의 대응력 강화로 주민 안전 고취
- 주민의 관심도를 높이고 주민 스스로 범죄 예방 유도

• 범죄예방지도를 작성하여 공개

1.4.3 국세청 역외탈세 정보공유

- ㅇ 역외탈세와 불법외환거래가 동시에 발생하는 특징을 보인다는 점에 착안하여 탈세 적발의 가능성을 높이는 방안을 발견
  - 불법외환거래 및 역외탈세에 대한 단속효과를 동시에 높임
- ㅇ 주요 추진내용
  - 관세청과 국세청이 각자 보유 중인 신고자료 등의 과세정보 교환
- ㅇ 성과 및 시사점
  - 지능적 역외탈세에 효과적으로 대처
  - 탈세 혐의정보 공유의 획기적 증가로 연간 약 900억원 세수 증대
  - 과세 기초정보 공유의 확대로 약 1,000억원 세수 증대
  - 복지부 등 다양한 기관과의 정보 공유 확대를 통해 485억원 상당의 국고 손실 방지

2 국내 민간부문 빅데이터 활용 현황

2.1 민간에서 이용되는 공공정보

- ㅇ 민간에서 가장 활발하게 활용되고 있는 공공정보는 기상정보, 지리정보, 법률정보, 특허정보 등
- ㅇ 기상청 기상정보
  - 생활밀착형 기상정보 서비스를 만들어 국민이 무료로 이용 가능
  - 맞춤형 날씨정보, 항해를 위한 해상기상 등 특정 수요자를 대상으로는 유료로 제공
- ㅇ 국토지리정보원 지리정보
  - 특정사업자(지도제작, 온라인 맵, 내비게이션 등 제작)가 일정한 가공을 거쳐 책자, 웹서비스, 스마트폰 등으로 변환하여 활용
- ㅇ 법제처 법률정보

- 일반 국민을 위한 생활법령서비스로 활용
- 민간 사업자에 의해 고급 법률정보서비스나 법학분야 논문정보 등 유사 DB서비스사와 결합되어 신 서비스 창출

○ 특허청 특허정보

- 한국특허정보원의 특허정보를 제공받아 민간사업자들이 변리사 등 특수 수요자들에게 고급 특허정보 서비스를 개발하여 판매

2.2 국내 기업의 빅데이터 활용 사례

2.2.1 네이버 모바일 빅데이터 맞춤형 서비스

○ 검색엔진인 네이버 자체가 빅데이터이며 이를 이용하여 데이터를 수집하고 분석

○ 주요 추진내용

- PC와는 차별화되는 모바일의 맞춤형 검색 서비스 시작
  - 지역명소 : 네이버 검색창에 지역명과 가볼만한 곳을 입력(예 : 서울시 가볼만한 곳)하면 검색결과의 최상단에 지역명소가 나옴

○ 성과 및 시사점

- 원하는 정보를 더욱 손쉽게 찾을 수 있게 검색 기능을 고도화
- 모바일 기기의 특성에 따른 고객 요구사항 반영 가능

2.2.2 CJ대한통운 스마트 통합물류 시스템

○ 세계적인 물류기업들은 빅데이터 분석 시스템 구축을 위해 발빠르게 대응하고 있으며 물류시장에서 빅데이터는 '미래를 읽어내는 안테나'로 인식되는 등 중요성이 강조

○ 주요 추진내용

- 최적의 운송경로를 찾아주는 스마트 통합 단말기
- 국내 택배 물동량 사전예측 및 선제적 대응 등 통합물류시스템 확장을 통해 운영 효율성과 안정성 증대

- 성과 및 시사점
  - 통합 물류 시스템의 확장을 통해 서비스 경쟁력을 한층 더 강화
  - 물류시스템의 전반적인 효율 향상뿐만 아니라 운영 및 관리 비용의 절감

223 **신한카드 고객 맞춤형 서비스를 위한 빅데이터 센터**

- 경쟁 과열 상태인 카드업계가 고객을 늘리기 위한 소비패턴 분석을 통하여 다양한 회원층 확보에 주력
- 주요 추진내용
  - 빅데이터 센터를 구축하여 수집된 데이터를 관리
  - 소비패턴을 분석하여 선호도가 가장 높은 품목에 대한 할인혜택을 집중 제공하는 서비스 개발
- 성과 및 시사점
  - 개인의 카드소비패턴을 분석하여 상품을 개발하였기 때문에 고객의 만족도가 높아지는 동시에 매출이 증대
  - 개개인의 니즈를 고려한 타깃 마케팅이 가능

3 해외사례

31 해외 공공부문 빅데이터 활용 현황

311 미국

- 국토보안
  - 국토안보부를 중심으로 테러·범죄 방지를 위한 범정부적 빅데이터 수집, 분석 및 예측체계 도입
  - 국내외 금융시스템의 감시로 테러자금 조달 색출 기능 강화
- 치안 : FBI의 종합 DNA 색인 시스템 CODIS
  - 범인의 DNA와 관련된 빅데이터를 구축

- 범인의 DNA가 확보되면 1시간 안에 분석이 가능하도록 주정부 데이터와 연계 및 실시간 분석 솔루션을 확보
  - 의료 : 필박스 프로젝트(Pillbox)
    - 국립보건원(National Library of Medicine)의 사이트로 약(Medicine) 검색이 가능
    - 사용자가 사용 중인 약에 대해 정보가 불분명할 때 검색을 통해 정확한 정보를 획득
    - 검색을 통해 수집된 빅데이터를 이용하여 주요 질병 분포, 연도별 증가 등에 대한 통계치 확보 가능

312 영국

- 정보공개
  - 정부 사이트(data.go.uk)를 통해 공공부문의 정보 공유 및 활용을 위한 데이터 원스톱 서비스 제공
  - 일반인들의 참여를 장려하는 커뮤니티 제공
    - 아이디어 수렴, 앱 개발, 데이터 공개 등의 주제에 대한 토론 등

313 싱가포르

- 국가위험관리
  - 테러 및 전염병으로 인한 불확실한 미래 대비
  - RAHS(Risk Assessment & Horizon Scanning) 시스템을 통해 모든 국가적 위험을 수집 및 분석하여 위험을 선제적으로 관리
  - 수집된 정보는 분석되어 사전에 위험을 예측

314 호주

- 정보공개
  - 정보관리청을 통한 정보개방

- 자동화된 툴을 활용하여 방대한 양의 정보검색, 분석, 재사용 시간 등이 이전에 비해 단축됨
- 정부 데이터에 대해 포털사이트를 운영하여 공개(data.gov.au)

3.2 해외 민간부문 빅데이터 활용 현황

3.2.1 새로운 비즈니스 창출에 활용

- ㅇ 기후보험을 개발하여 판매
  - 미국 Climate Cops.는 미국정부가 개방한 데이터를 기반으로 기후 및 곡물 수확량을 예측하여 이상기후 발생 시 해당 농가에 보험을 지급하는 기후보험을 개발
    - Climate Cops.가 이용한 정보는 과거 60년간 지역별 작물수확량, 2제곱마일 단위로 제공되는 토양성분, 미국 전역 100만 개 지점의 기후 정보 등
- ㅇ 개인 맞춤형 보험상품 안내 서비스
  - 보험사(도쿄해상화재보험)와 통신사(NTT도코모)가 제휴
  - 통신사의 GPS 위치정보를 통해 고객이 스키장이나 골프장에 도착하였을 때 보험사는 특성에 맞는 맞춤형 보험상품을 실시간으로 안내
- ㅇ 소비자 행동정보 분석을 통한 정보 제공
  - 버라이즌은 2012년 10월부터 “Precision Market Insight”라는 서비스를 제공
  - 이용자의 위치, 선호하는 콘텐츠, 전화사용 패턴 등의 데이터를 수집한 후 이를 분석하여 도출한 시사점을 수요자에게 공급하는 서비스
    - 예상되는 소비자의 행동이 아닌 실제 관찰된 소비자의 행동정보를 분석하기 때문에 정확성이 높음

3.2.2 업무 효율성 향상 및 서비스 개선에 활용

- ㅇ 작업시간 단축
  - 미국 금융지주사 'BB&T'는 자금 세탁 추적에 빅데이터를 활용
  - 기존 데이터를 분석한 후 약 25개의 시나리오를 만들어 다양한 거래 내역을 실시간으로 추적한 결과 수개월씩 걸리던 자금 추적 작업시간이 하루 정도로 단축됨
  - 빅데이터를 자금 세탁 추적에 활용하여 세탁 추적 작업 시간을 단축
- ㅇ 모델 개발을 통한 수익성 향상
  - 미국의 '제스트파이낸스'는 대출 신청자의 통화습관, 소셜미디어 메시지 등 수천 개의 변수를 반영한 비정형데이터를 분석하여 고객 신용평가 모델을 개발하고 이를 활용
  - 단기 대출 연체율은 20%까지 감소하고 20%의 수익 향상
- ㅇ 전사적 서비스 개선에 활용
  - 자동차 렌탈업체 '허츠(Hertz)'는 세계 8,000여 지점의 이메일, 온라인(고객의 목소리), 콜센터에서 불만과 관련된 키워드 자동 추출
  - '지점-서비스-고객 불만' 연관성을 파악하여 전사적 서비스 개선에 활용

3.2.3 개발 및 마케팅 전략 수립에 활용

- ㅇ 소비지출에 대한 예측
  - 미국에서는 소비자태도지수 설문조사 결과보다 구글트렌드를 활용한 결과가 더 정확한 것으로 나타남
- ㅇ 소비자 니즈를 신제품 개발에 활용
  - 미국 포드(Ford) 자동차는 차량에 설치된 센서로 고객의 운전습관, 주행정보 등의 데이터를 획득하여 분석
  - 분석한 정보를 토대로 소비자의 니즈를 파악하여 신차 개발에 적용
- ㅇ 맞춤형 상품추천
  - '이베이(Ebay)'는 명절이나 기념일처럼 선물 구입이 증가하는 시점에

맞춰 고객의 소셜미디어 활동 내역과 과거 구매이력을 분석하여 고객이 선물할 만한 지인의 프로파일을 추정하고 적합한 선물을 추천

- ‘아마존(Amazon)’은 특정 물품을 구매할 경우 이용자들의 소비성향을 분석하여 연관되는 제품을 추천하고 이를 구매하도록 유도

ㅇ 고객맞춤형 마케팅

- 신용카드회사인 ‘아멕스(AMEX)’는 소셜 네트워킹 웹사이트인 ‘포스퀘어(Foursquare)’와 제휴하여 고객의 거래 정보 및 위치 정보, 소셜 미디어 활동 정보 등을 마케팅에 활용

• 고객이 선호하거나 자주 가는 인근 레스토랑, 카페의 할인쿠폰을 실시간으로 전송하는 등 고객맞춤형 마케팅을 수행

- 호텔 체인인 ‘하라스(Harrah’s) 그룹’은 카지노에 대해 주민 타깃 마케팅을 실시하여 만족도 1위를 달성

• 카지노 고객의 행동 패턴을 분석한 결과 핵심고객은 관광객이 아니라 지역주민이라는 것을 인지함

• 지역주민을 대상으로 하는 맞춤형 마케팅 전략을 수행하여 성공

324 **영업 전략에 활용**

ㅇ 상품 라인업 조정

- 미국 유통업체인 ‘월마트(Walmart)’는 지역 특성을 고려하여 상품 라인업을 조정

• 소셜미디어에서 수집한 빅데이터의 분석결과를 특정 점포의 상품진열에 반영

• 캘리포니아 마운틴뷰 지역에 자전거에 관심을 갖는 거주자가 많다는 사실을 파악하여 해당 점포의 상품 라인업을 조정

ㅇ 고객 이탈 방지

- 미국의 ‘T-Mobile’은 번호이동[6] 위험을 감지하는 시스템을 운영하여 고객 이탈을 방지

- 3,000만 명이 넘는 가입자로부터 매일 170억 건 이상의 통화 및 송수신 내역을 담은 빅데이터 및 고객 간 소셜네트워크를 분석
- 다른 통신사로 회선을 옮긴 고객이 사전에 보였던 특유의 이용 패턴을 발견하고 이를 실시간으로 포착하는 시스템을 구축하여 이탈 징후를 보이는 고객에게 맞춤형 추가 혜택을 제공

325 **기타**

- ㅇ 펀드 운용에 활용
  - 런던의 투자기관 '더웬트 캐피털 마켓츠'는 시장 투자심리를 이용한 빅데이터를 활용하여 헤지펀드를 운용
    - 매일 1억 개의 '트윗'(트위터)을 분석하여 시장 분위기를 측정한 뒤 이를 포트폴리오에 반영하는 '트위터 펀드'를 개발·운용해 높은 수익률을 유지
    - S&P500 지수가 2.2% 하락한 2011년 7월에 헤지펀드 평균운용수익률이 0.76%에 그친 데 비해, '트위터 펀드'는 1.86%를 기록(최근 청산)
- ㅇ 재고 관리
  - 미국 유통업체 '시어스(Sears)'는 빅데이터를 분석하여 최적의 가격을 책정하고 재고를 관리
    - 점포별 재고, 경쟁업체 가격, 날씨예보, 고객정보 등의 다양한 빅데이터를 분석하여 상품별·시간별 최적의 가격을 책정
    - 실시간 주변상황과 고객의 취향을 동시에 고려한 개인별 특화 쿠폰을 적시에 발급하여 매출 증가와 적정 재고 관리를 달성

6 통신사를 변경하는 것을 의미함

III 서울시 공공데이터 정보관리 및 정보공개 현황

1 서울시 공공데이터 현황
2 공개된 정보의 활용 현황
3 정보공개의 경제적 효과 분석

III 서울시 공공데이터 정보관리 및 정보공개 현황

1 서울시 공공데이터 현황

1 1 서울 열린데이터 광장

1 1 1 열린데이터 광장 개설 목적

- ㅇ 서울시가 보유하고 있는 공공정보를 민간에 공개하고 시민과 소통하기 위한 목적으로 만들어진 웹사이트
  - – 시정의 공익성, 업무효율성, 투명성을 높임
  - – 시민의 자발적인 참여로 새로운 서비스와 공공가치를 창출할 수 있도록 유도함
- ㅇ 서울시 공공데이터 서비스 통합창구의 역할을 함
  - – 공공데이터의 존재여부 및 소재정보 등을 검색을 통해 확인할 수 있음
  - – 공공정보 원천데이터의 다운로드 가능
  - – Open API 서비스를 제공
- ㅇ 사회경제적으로 가치가 높은 보유 공공데이터를 개방하여 민간에게 제공함
  - – 다양한 비즈니스 창출 기회 제공
  - – IT콘텐츠 산업육성에 기여

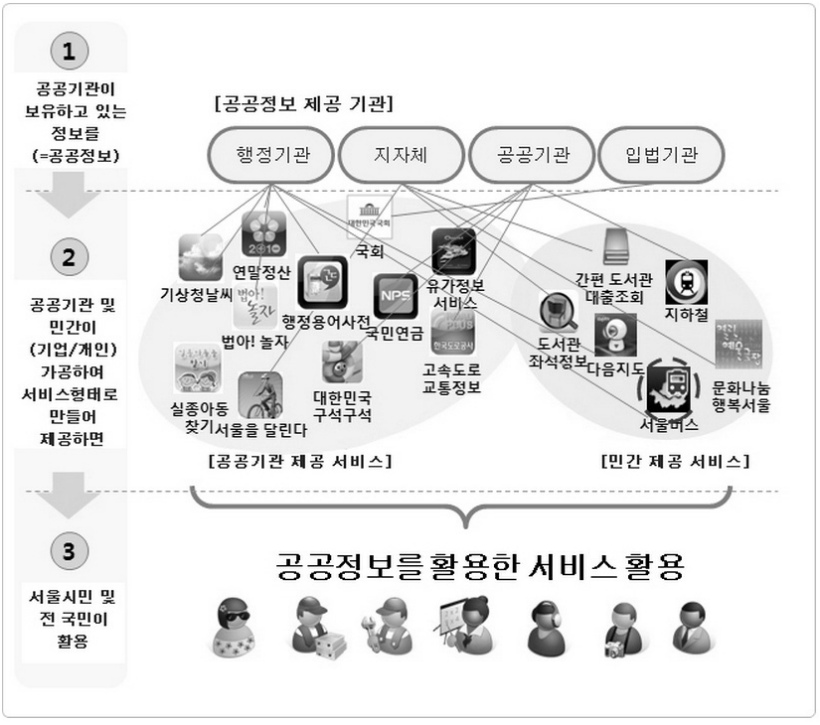


그림 3-1 공공데이터 활용
출처 : 서울열린데이터광장(data.seoul.go.kr)

112 **열린데이터 광장 추진과정**

- ㅇ 공공데이터 개방 프로세스 수립
  - – 공공데이터 분류체계 수립, 발굴, 수요조사, 품질관리, 플랫폼 구축 등의 세부과정에 대한 계획을 수립
- ㅇ 공공데이터 분류체계 수립
  - – 수요자 관점에서 서비스 활용성을 극대화시키기 위해 쉽고 편리하며 직관적으로 이해할 수 있는 분류체계 마련
- ㅇ 공공데이터 발굴
  - – 서울시 내부 400여개 시스템의 전수조사를 통해 개방 가능한 데이터 목록을 작성

- ㅇ 공공데이터 수요 조사
  - 공공데이터 수요자, 제공자, 전문가 그룹의 의견 수렴을 통해 활용성이 높은 공공데이터 발굴을 시도
- ㅇ 데이터 품질관리
  - 데이터 품질 진단 기법인 '프로파일링'을 통해 공공데이터 품질을 측정하고 결과를 바탕으로 지속적으로 수정 및 보완
- ㅇ 플랫폼 구축
  - 공공데이터 개방지침과 프로세스를 기반으로 공공데이터 공유와 활용을 위한 통합 플랫폼을 구축

113 **연혁**

- ㅇ 2012년 5월 21일 시범서비스를 위한 서울 열린데이터 광장 오픈
  - 시스템 37종 데이터셋 905개로 서비스 시작(2012년 5월 16일 기준)
- ㅇ 2013년 2월 구축된 홈페이지의 문제점을 개선하여 리뉴얼
  - 시민의 접근성과 활용성을 극대화
- ㅇ 2014년 2월 17일 서울 열린데이터 광장이 새로운 기능을 추가하여 재오픈
  - 데이터 시각화 시범 서비스 개시
  - 예산현황, 대기오염현황, 통학/통근수단 현황, 외국인 관광객 현황, 구청별 인구현황 등을 직관적으로 인식할 수 있도록 비쥬얼 차트를 활용한 시각화 서비스 시작
  - 우리동네 열린데이터 광장 포털 서비스 시범 개방
  - 구로구와 서대문구의 열린데이터 광장을 오픈하였으며 향후 25개 전 자치구로 확장할 예정
  - 공공데이터 활용 기능 강화
  - Sheet : 필터 서비스 강화
  - Chart : 다양한 유형의 차트 제공

- Map : 지도 메시업(Mashup)[7] 기능 제공
- Open API : 인증키 다중발급 기능 및 인증키 중지, 삭제 기능 제공

– 시민 공유의 공간 확대

- 활용갤러리 확대 : 앱, 웹, 프로그램, 인포그래픽, 행사, 문서 등 다양한 정보 공유의 장을 확대

1 1 4 이용 현황

ㅇ 이용객 수는 꾸준히 증가하고 있는 추세

– 2012년 5월 개설한 첫 한 달간 이용건수는 2만여 건에 불과

– 2013년 4월 기준 누적 이용건수는 267배가 증가한 540만여 건으로 시민이나 기관에서 하루 평균 15만여 건의 데이터를 활용

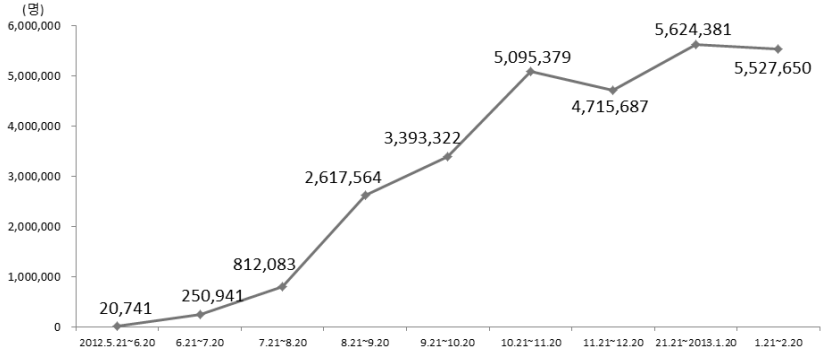


그림 3-2 서울 열린데이터 광장 월별 활용 현황
출처 : 서울 열린데이터 광장 백서

7 다양한 콘텐츠와 서비스를 융합하여 새로운 웹서비스를 만들어내는 것

1.2 서울시 공공정보 공개 현황

1.2.1 정보공개 목표

- ㅇ 서울시는 2014년 150개의 공공정보 시스템 개방을 목표로 하고 있음
  - 2011년 20개에서 2014년 150개로 단계적으로 공공정보 개방을 확대 (2014년 3월31일 현재 145개 개방)

표 3-1 서울시 공공데이터 연도별 개방목표

연도	2011	2012	2013	2014
개방목표	20	60	100	150

1.2.2 정보공개 현황

- ㅇ 서울시는 공공정보를 시민들이 효과적으로 활용할 수 있도록 주요 데이터를 공개
  - 민간에서 생성하기 힘든 유동인구, 버스정보, 지하철정보 등 공공정보를 개방
- ㅇ 10개 분야 주요 데이터를 개방
  - 교통, 안전, 환경, 일반행정, 도시관리, 보건, 복지, 산업경제, 문화관광, 교육 등 대분류
  - 대분류 안에 49개의 중분류가 포함됨

표 3-2 서울시 열린데이터광장 분류체계

대분류(10개)	중분류(49개)
교통	교통운영, 교통시설, 대중교통, 자가교통, 물류수송
안전	시설물 안전, 소방, 재난
환경	대기환경, 토지환경, 수질환경, 생활환경, 기상, 공원녹지, 폐기물
일반행정	정책, 행정, 재정/예산/세무, 통계, 정보통신
도시관리	도시계획, 도로, 상수도, 하수도, 하천, 건축, 주택, 부동산
보건	식품, 의약품, 보건의료
복지	일반복지, 여성복지, 장애인복지, 아동·청소년복지, 노인복지, 가족·다문화
산업경제	산업, 경제, 고용노동, 창업, 벤처, 국제교류, 에너지 자원
문화관광	문화, 체육, 관광
교육	학교교육, 평생교육, 학술

자료 : 서울 열린데이터 광장 백서

- 2014년 3월 현재 13개 기관이 146종의 시스템, 1,376개 데이터셋, 2,509개 서비스를 제공
  - 분야별 서비스는 문화관광분야가 627개로 가장 많고, 그다음은 일반행정, 환경의 순

표 3-3 서울시 공공데이터 개방 현황(2014년 3일 기준)

분류	Sheet	Chart	Map	Link	File	Open API	합계
교통	29	5	4	47	8	44	137
안전	16	1	8	39	9	15	88
환경	93	43	54	53	57	104	404
일반행정	100	8	40	147	68	104	467
도시관리	28	0	10	66	10	45	159
보건	121	4	2	35	5	120	287
복지	4	0	0	93	1	4	102
산업경제	32	0	0	103	0	32	167
문화관광	182	1	102	49	104	189	627
교육	14	0	4	35	6	12	71
합계	619	62	224	667	268	669	2509

자료 : 서울열린데이터광장(data.seoul.go.kr)

2 공개된 정보의 활용 현황

2.1 서울시 공공정보 활용 현황

- ㅇ 열린데이터 광장을 통해 공개된 데이터를 활용하여 스마트폰 앱, 홈페이지 등이 만들어져 시민들에게 다양한 편의를 제공
- ㅇ 서울 열린데이터 광장에서는 시민들이 직접 만든 앱과 홈페이지 정보를 공유할 수 있도록 활용갤러리를 운영
  - 공공데이터를 활용하여 만들어진 앱, 웹사이트, 인포그래픽[8], 행사 등의 정보를 공유

2.1.1 스마트앱

- ㅇ 공공데이터를 이용하여 스마트폰 앱을 개발하고 이를 촉진하기 위해 서울시 스마트앱 공모전을 매년 개최
  - 교육, 교통, 문화관광, 환경 등 다양한 분야의 앱이 만들어짐
  - 2013년 대회에는 총 118건의 앱이 접수
- ㅇ '서울시 야간도서관 정보안내' 앱
  - Kate & Lu Awesome 프로젝트팀이 서울시 야간도서관 정보가 필요한 사람을 위해 서울 열린데이터 광장의 Open API를 이용하여 개발
  - 구별로 야간도서관 운영시간과 위치 등의 정보를 제공
- ㅇ '서울교통정보' 앱
  - 유료로 존재하는 앱을 무료로 배포하기 위한 목적으로 제작
  - 서울도시고속도로 교통정보(smartway.seoul.go.kr)의 데이터를 이용하여 주요 도로 및 남산터널 등 교통정보를 실시간으로 확인할 수 있음
- ㅇ '서울시 문화재' 앱
  - 서울시에 있는 문화재에 대한 소개와 위치 정보 등을 제공

---

8 공공정보를 그림, 표, 그래프, 사진 등을 활용하여 한눈에 보기 쉽게 표현하는 것

- 구별로 소재한 문화재를 찾아볼 수 있음
- ‘서울 대기환경’ 앱
- 최근 관심이 부각되고 있는 대기환경정보를 제공
- 구별로 실시간 대기환경 정보 확인, 통합 대기환경 정보에 대한 간단한 설명을 제공

그림 3-3 서울시 공공정보를 활용하여 만들어진 앱
출처 : 서울 열린데이터 광장(data.seoul.go.kr)

2.1.2 웹사이트

- 공공데이터를 이용하여 웹페이지를 통해 정보를 제공
- 통계정보, 위치(지리)정보, 행사정보 등을 주로 제공
- 공공정보를 활용한 주요 웹사이트
- 위치정보 제공
  - 서울시 장난감 도서관 위치
  - 서울시 문화재 지도
  - 서울지도 홈페이지 추천 지리정보
- 환경정보 제공
  - 서울의 일별 평균 대기오염도
- 일정 안내
  - 서울의 축제 및 행사 타임맵

213 **인포그래픽**

- ㅇ 공공정보를 활용하여 인포그래픽을 제작
  - 인포그래픽을 활용하여 정보를 보다 손쉽게 이해할 수 있게 함
  - 시민들에게 서울시의 정책 및 정보를 효율적으로 전달
- ㅇ 주요 인포그래픽 목록
  - 서울시 고령자들은 어떤 공공자원을 필요로 할까?
  - 서울시 단독주택 재활용 정거장 도입, 무엇이 달라지나?
  - 선행학습 금지법, 어떤 내용이 담겨져 있을까?
  - 2012 서울 월평균 임금은 금융 및 보험업이 349만원으로 최고, 숙박 및 음식점업이 141만원으로 최저
  - 2013년 1/4분기 서울신용보증재단 소기업 체감경기지표(BSI)

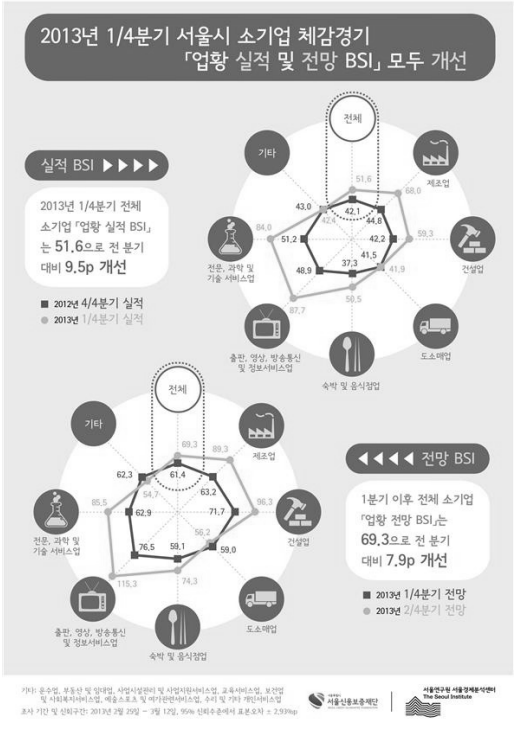


그림 3-4 서울시 공공정보를 활용하여 만들어진 인포그래픽
출처 : 서울 열린데이터 광장(data.seoul.go.kr)

2.2 추후 활용방안

2.2.1 투자출연기관 및 자치구 개방 확대

- 17개 투자출연기관 및 25개 자치구 협력 체계 TF를 구성하여 공공정보 개방 추진 실행력 강화
- 투자출연기관 및 자치구는 서울시 열린데이터 광장을 기반으로 서비스를 구현할 계획
  - 별도의 플랫폼 구축이 필요 없기 때문에 데이터 개방에 전념할 수 있음

2.2.2 공공정보 공유체계 구축

- 카탈로그 서비스 제공
  - 서울시의 공공정보와 안전행정부의 국가자원공유포털의 Open API 서비스 목록을 제공하는 서비스
- 서울시와 민간이 협업할 수 있는 환경 마련
  - 서울시 자체데이터뿐만 아니라 민간이 보유하고 있는 데이터를 공유할 수 있는 환경을 구축

2.2.3 오픈데이터에서 링크드오픈데이터(Linked Open Data)로 진화

- 공공정보를 외부 데이터와 연결하여 활용성을 증진
  - 링크드오픈데이터를 확장하여 공공정보를 풍부하게 할 계획

3 정보공개의 경제적 효과 분석

3.1 개요

- 공공데이터 정보공개의 경제적 효과 또는 비용절감 효과를 세 가지 관점에서 분석

- 빅데이터를 이용하였을 때 서울시 공공부문의 비용절감 효과
- 공공정보를 민간에 공개하였을 때 경제적 파급 효과
- 빅데이터 및 공공데이터를 활용한 서울시 역점 추진(정책)사업에 대한 효과 분석

3.2 서울시 공공부문 빅데이터 잠재력

3.2.1 분석틀

- ㅇ 공공부문 빅데이터 이용에 따른 공공행정의 잠재적 비용절감 효과를 측정하기 위해 '효율성 측면의 비용절감', '부정 및 오류 감소', '조세제도의 개선' 등 3가지 관점에서 접근
- 이 같은 접근방식은 McKinsey Global Institute(2011)의 분석틀을 적용한 것
- 동일한 방식으로 영국 공공부문의 비용절감 효과를 추정(Chris(2012), 부록 참조)
- ㅇ 공공부문에서 빅데이터를 사용하였을 경우 기대할 수 있는 효과
- 효율성 측면의 비용절감
  - 부처의 조직을 효율화하고 업무의 우선순위 선정에 대한 객관성을 증대시킴으로써 정부운영의 직접비용이 감소
  - 분석을 위한 기본 값으로 공공부문의 지출액을 활용
- 부정 및 오류 감소
  - 사회복지제도의 허점을 이용한 부정수급, 계산착오에 따른 보조금 과다지급 등의 예산 낭비를 방지
  - 복지제도 내에서 부정수급 또는 과다지급을 찾아낼 수 있으며, 기본 값으로 사회복지 지출액을 활용
- 조세제도의 개선
  - 빅데이터를 활용해서 부과된 세금과 실제 징수된 세금 간 차이를 줄일 수 있는 방안을 모색할 수 있음

- 분석을 위한 기본 값으로 공공부문의 세입금액을 활용
  - 앞서 언급된 3가지 부문(효율성 측면의 비용절감, 부정 및 오류 감소, 조세제도의 개선)에서 빅데이터를 적용할 수 있는 분야의 비중을 파악
  - 빅데이터를 적용할 수 있는 분야에서 발생하는 효과의 비중을 파악하여 최종적으로 공공부문의 비용절감 효과를 산정

3.2.2 서울시 공공행정 비용절감 추정

- 분석의 기반이 되는 금액을 설정(분석을 위한 기본 값)
  - 빅데이터를 이용하면 시정 운영에 따른 비용감소를 기대할 수 있으며,기본 값으로 서울시의 1년 총 지출금액을 활용
    - 2012년 서울시의 총 지출액은 21조 3,685억 4,051만 897원
  - 빅데이터를 활용하여 복지제도 내에서의 부정수급 및 보조금 과다지급 등을 파악하기 위하여 서울시 사회복지 지출액을 기본 값으로 설정
    - 2012년 서울시 사회복지 지출액은 3조 9,750억 4,553만 7,010원
  - 빅데이터를 조세와 관련된 분야에 적용한다면 업무효율화 및 징수과정에서 오류를 줄여줄 수 있으며, 기본 값으로 서울시의 일반회계 세입금액을 설정
    - 2012년 서울시의 일반회계 세입금액은 16조 5,631억 386만 298원
- 각 분야(효율성 측면의 비용절감, 부정 및 오류 감소, 조세제도의 개선)에서 빅데이터를 적용할 수 있는 비중을 산출[9]
  - 서울시 지출금액 중 빅데이터 이용 가능 비율은 20~25%로 추정
  - 서울시 사회복지 지출금액 중 빅데이터 이용 가능 비율은 1~3%
  - 서울시 일반회계 세입금액 중 빅데이터 이용 가능 비율은 5~10%
- 빅데이터를 이용할 수 있는 분야 중에서 잠재적으로 비용절감을 기대할 수 있는 부분에 대한 비율을 산출

---

9 McKinsey Global Institute(2011)에서 적용한 비중을 참고

- 운영효율 개선 분야의 비용절감 효과는 15~20%
- 부정 및 오류 감소 분야의 비용절감 효과는 30~40%
- 조세제도 개선으로 인한 세입증가 분야의 비용 절감효과는 10~20%

○ 서울시의 공공행정이 빅데이터 활용으로 절감할 수 있는 비용은 연간 최소 7,358억원에서 최대 1조 4,474억원에 이를 것으로 추정(2014년 서울시 예산 대비 3.4%~6.7%)

- 운영효율 개선 분야에서의 절감비용은 최소 6,411억원에서 최대 1조 684억원으로 추정
- 부정 및 오류 감소 분야에서의 절감비용은 최소 119억원에서 최대 477억원으로 추정
- 조세제도 개선에 따른 세입 증가 분야에서의 절감비용은 최소 828억원에서 최대 3,312억원으로 추정

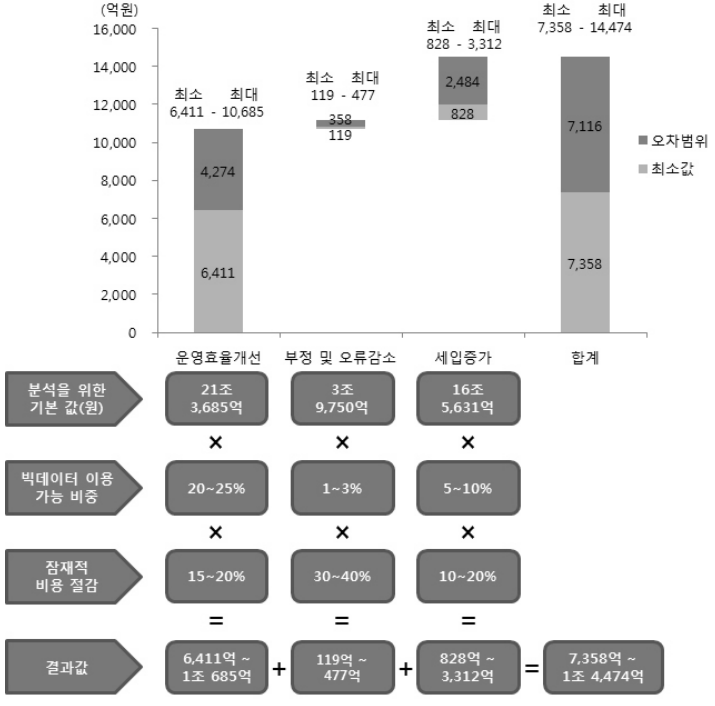


그림 3-5 서울시 공공부문 빅데이터 잠재력

3.3 정보공개에 따른 민간부문 경제적 파급 효과

3.3.1 분석틀

- ㅇ 공공정보가 민간에서 활용되어 DB서비스 시장을 형성
  - 기업들은 구축된 공공정보 데이터베이스를 이용하여 서비스를 만들어 이용자에게 제공
  - 공공정보가 민간에서 이용되었을 때 효과를 분석하기 위해서는 DB서비스 시장의 분석이 필요
- ㅇ 공개된 공공정보는 DB서비스 기업들에 의해 지식정보서비스와 광고서비스의 형태로 이용자들에게 제공되어 매출을 발생(투입값)
  - DB서비스 기업이 정보서비스를 만들어 이를 판매할 때 발생하는 수익은 직접 매출(지식정보서비스)에 해당
  - DB서비스 기업이 공공정보를 소비자의 니즈가 반영된 다양한 형태의 서비스로 재가공할 때 발생하는 수익 및 연관된 파생 분야는 간접 매출(광고서비스)에 해당

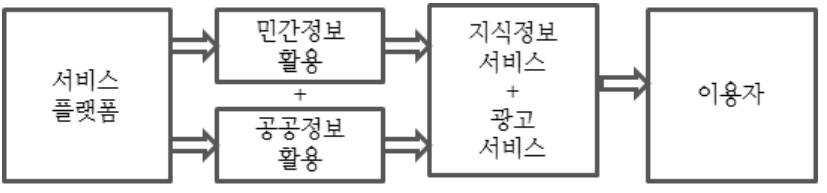


그림 3-6 공공정보 민간활용 산업의 가치사슬
출처 : 허필선 외(2013)

3.3.2 DB서비스 시장

- ㅇ DB산업의 구성
  - DB산업은 DB구축, 컨설팅, 솔루션, 서비스 분야로 구성되어 있음
- ㅇ DB서비스 시장
  - DB서비스 기업이 공공정보를 이용하여 사업을 시행할 때 발생하는 경제적 파급효과를 분석

- DB서비스 기업의 주요 매출은 정보서비스 및 광고 등을 통해 발생

333 **공공데이터 공개에 따른 민간부문 파급효과 산정**

- ㅇ 전체 DB산업의 시장 규모를 파악
- 2013년 전체 DB산업의 규모는 11조 6,517억원(한국데이터베이스진흥원, 2013)
- ㅇ DB서비스 산업의 시장규모를 파악
- 2013년 DB서비스 산업의 규모는 5조 1,369억원(한국데이터베이스진흥원, 2013)
- ㅇ 서울의 DB서비스 시장 규모를 추정
- 서울의 DB서비스 시장 규모를 추정한 데이터가 없으므로 추론과정을 거쳐야 함
- DB서비스에 포함되는 세세분류(표준산업분류)는 포털 및 기타 인터넷 정보매개서비스(63120), 뉴스제공업(63910), 데이터베이스 및 온라인 정보 제공업(63991) 등
- DB서비스에 포함되는 세세분류 기업들의 매출에서 서울시 소재 기업이 차지하는 비중은 70.1%
- ㅇ DB서비스 중 공공부문이 차지하는 비중을 파악
- 전체 DB기반 지식정보서비스 매출 중 공공정보의 매출기여도는 2013년 22.4%로 가정(허필선 외, 2013)
- ㅇ 산업연관표의 부문분류표를 이용해 통합소분류를 추출
- 산업연관표 상의 소분류에서 공공정보를 이용한 민간활용 서비스는 부가통신 및 정보서비스(142), 컴퓨터관련서비스(153), 광고(151)와 밀접한 관련이 있음
- ㅇ 투입값을 산출하여 산업연관분석에 이용
- 투입값은 부가통신 및 정보서비스(142)에 3,094억원, 컴퓨터관련서비스(153)에 1,397억원, 광고(151)에 3,570억원 등 총 8,061억원

표 3-4 산업연관분석을 위한 투입값 산출 결과

연도	DB서비스	산업분류	투입(억원)
2013	정보서비스	142(부가통신 및 정보서비스)	3,094
		153(컴퓨터관련서비스)	1,397
	광고	151(광고)	3,570
	계	-	8,061

- o 2011년 산업연관표를 이용하여 서울시 공공정보 민간활용에 따른 경제적 파급효과를 계산
  - 생산유발효과 1조 7,775억원
  - 부가가치유발효과 1조 6,122억원
  - 취업유발효과 8,920명

표 3-5 서울시 공공정보 민간활용에 따른 경제적 파급효과

(단위 : 억원)

연도	DB서비스	산업분류	투입	생산유발효과	부가가치유발효과	취업유발효과
2013	정보서비스	142*	3,094	5,520	6,188	1,997명
		153**	1,397	2,388	2,794	1,744명
	광고	151***	3,570	9,866	7,140	5,179명
	계	-	8,061	17,775	16,122	8,920명

*142 부가통신 및 정보서비스 **153 컴퓨터관련 서비스 *** 151 광고

3.4 서울시 역점 추진사업 효과분석

3.4.1 서울시 정보기획단 추진 사업

- o N버스(올빼미버스) 도입으로 인한 서울시민 교통비 절감[10]
  - 야간통행 이동비용 절감
    - 심야시간에 택시를 이용해야만 했던 시민들이 대체수단인 야간버스를 이용

10 부록 참고

- 1인당 약 7,050원, 1일 4,440만원, 연간 162억원의 절약효과

– 유류비 절감효과

- 심야버스 이용으로 승용차 운행 수요를 감소시켜 유류비용을 절감
- 1일 665만원, 연간 24억원의 유류비 절감 효과

ㅇ N버스(올빼미버스) 도입으로 인한 환경개선 효과

– 이산화탄소 배출 감소

- 승용차 운행을 버스 운행으로 대체함으로써 이산화탄소 배출량 감소
- 1일 1,365kg, 연간 498톤의 이산화탄소 배출 억제 효과

342 정보기획단 추진 예정 사업

ㅇ N버스에 이어 현재 서울시 정보기획단이 추진할 예정인 사업으로는 상권분석과 택시매치메이킹 사업이 있음

1) 상권분석

ㅇ 예비창업자에게 상권에 대한 정보를 제공함으로써 창업성공률을 높임

– 유동인구, 소득, 소비/매출, 부동산 시세 등의 자료를 통해 업소 과밀정보, 생존가능성 추정, 상권분석 서비스 제공 등

ㅇ 기존점포에 대한 평가를 통해 객관적인 정보를 얻을 수 있고 이를 토대로 창업성공률을 제고

2) 택시 매치메이킹

ㅇ 수요자인 승객과 공급자인 택시기사들의 정보를 매칭해서 수급불균형을 해소하여 택시 이용의 편리성을 높이고 탑승률을 높임

– 유동인구, 택시 승하차 데이터, 기상 데이터 등을 이용하여 요일별, 시간별, 날씨별 택시매치메이킹 최적 위치 선정 모형을 개발

– 택시이용객들에게 택시를 쉽게 탈 수 있는 지역의 정보를 제공하고 택시기사들에게는 손님을 쉽게 태울 수 있는 지역의 정보 제공

3.4.3 시정 집중·핵심사업

- 시정 집중·핵심사업에 빅데이터 및 공공데이터를 이용한다면 공공행정의 운영효율을 높일 수 있으며 정책적 효과를 극대화할 수 있음

1) 공공의료 강화

- 의료보장 적용인구, 환자 현황, 의료기관 현황 등을 이용하여 공공의료 강화에 이용
  - 의료보장인구가 많고 의료기관이 상대적으로 적은 의료취약지를 선정하여 공공의료원을 우선적으로 입지
  - 환자 현황 및 치료내역을 관리하여 분석한 후 맞춤형 예방접종, 백신확보 등을 수행
- 해외사례
  - 미국 국립보건원은 유전자데이터를 확보하여 유전적 다양성을 분석하여 이를 이용해 질병을 치료
  - EU 메디시스(Medisys)는 의학전문 사이트로 뉴스포털 등에서 수집한 뉴스를 분류하고 추적한 후 공중보건에 영향을 미치는 사건을 탐지하여 위험 경고

2) 여성안전 종합대책, 도시안전마스터플랜

- 범죄발생현황, 가정폭력 현황, 화재현황, 건축물 현황 등을 이용하여 여성안전 종합대책, 도시안전마스터플랜 수립에 활용
  - 범죄예방 가능성이 높은 지역에 대한 순찰을 강화
  - 여성들의 안전한 귀가를 도울 수 있는 안심귀가도우미의 효율적인 배치에 이용
  - 가정폭력이 일어나는 유형에 대해 진단하여 적극적인 조치를 취해야할 대상을 선정
  - 화재 위험성이 높은 건물을 찾아내 소방안전점검을 실시

- ㅇ 해외사례
  - 뉴욕시는 위치정보, 사업정보, 교통위반, 범죄, 구급차전화, 화재 등 다양한 종류의 데이터와 결합시켜 위험도가 높은 건물을 찾아내 예방
  - LA시는 고급수학과 컴퓨터 학습기술을 적용하여 범죄발생을 예측하고 집중 순찰

3) **지역맞춤형 수방대책, 태양광 발전 보급**

- ㅇ 기상관측정보(강수량, 일조량 등)를 이용하여 지역맞춤형 수방대책, 태양광 발전 보급에 이용
  - 강수량 패턴을 지역별로 파악하여 지형정보가 결합한 후 수해취약지역을 선정하고 맞춤형 수방대책을 수립
  - 일조량이 풍부한 지역에 태양광발전이 많이 이루어질 수 있도록 정보 제공
- ㅇ 해외사례
  - 미국 국립해양대기청은 기상 데이터를 기반으로 다양한 예측시스템을 개발

4) **채무관리 강화**

- ㅇ 예산자료, 회계자료, 세금자료, 지방채 등을 이용하여 채무관리 강화
  - 국세청의 역외탈세정보를 서울시와 공유하여 포탈된 세금 현황을 파악하고 이를 징수
  - 불필요한 예산이 쓰이고 있지는 않은가에 대한 감시를 통해 채무감축에 노력
- ㅇ 해외사례
  - 미국 국세청은 납세와 관련된 정부데이터, 소셜 네트워크에서 드러난 세금 관련 범죄 데이터를 종합적으로 분석해서 세금 범죄를 패턴화하여 파악 후 사전 예방

Ⅳ 정책제언

1 공공데이터 활용 및 정보공개 효과 제고 방안
2 서울시 대용량 데이터베이스의 효율적 활용방안
3 빅데이터를 활용한 정책 제안

Ⅳ 정책제언

1 공공데이터 활용 및 정보공개 효과 제고 방안

1 1 양질의 데이터 제공을 위해 노력

- ㅇ 서울시는 가치 있는 공공정보의 공개를 위해 노력
  - – 단순 문서가 아닌 가치 있고 유의미한 공공정보를 공개
    - • 2013년 6월 현재 중앙정부의 단위업무 중 유의미한 정보는 2.5%에 불과(전자신문, 2013-12-10, 정부3.0 공공정보 개방을 넘어 활용으로 진화해야)
  - – 가치 있는 공공정보 발굴을 위해 노력
- ㅇ 공공데이터의 품질 향상을 위한 오류율 저감 노력이 필요
  - – 데이터 개방을 기피하는 이유는 심각한 오류율 때문이므로 이를 보완할 수 있는 방안이 필요
    - • 데이터 검수를 위한 전담직원을 배치하여 오류율을 줄여나가는 방안을 마련
    - • 전문가의 자문을 통해 오류를 줄일 수 있는 방안을 강구
  - – 서울시는 공공데이터 품질의 업그레이드를 위해 공공정보 개방 건수의 확대보다 데이터의 품질 향상에 관심을 기울여야 함
- ㅇ 서울시 부서 간 또는 서울시와 외부 공공기관 간 협의를 통해 융합공공데이터를 제공할 수 있도록 노력해야 함
  - – 서울시 조직 간(예 : 도시계획국과 교통국 등) 융합데이터를 생성
  - – 서울시와 국가공공기관(예 : 경찰청, 국세청, 안행부 등)이 공동으로 융합데이터를 생성

12 공공데이터 활용의 활성화 유도

- 공공데이터 이용활성화를 유도하기 위해서는 관련 법령의 정비가 필요하며 서울시는 관련 법률 개정을 적극적으로 건의

표 4-1 공공데이터 개방 관련 정비대상 주요법령

개정법령	법령상 정비사항
공간산업진흥법 및 하위법령	이용자 범위 확대 : 이용하고자 하는 모든 사람 이용절차 간소화 : 신청절차, 서식 등 삭제
기상법/기상산업진흥법 및 하위법령	이용자 범위 확대 : 이용하고자 하는 모든 사람 이용절차 간소화 : 신청절차, 서식 등 삭제 제공창구 : 단일화된 제공시스템 규정
통계법 및 하위법령	이용자 범위 확대 : 이용하고자 하는 모든 사람 제공창구 및 방법 규정 : 공공데이터법 준용
발명진흥법 및 하위법령	이용자 범위 확대 : 이용하고자 하는 모든 사람 이용절차 간소화 : 신청절차, 서식 등 삭제
저작권법 및 하위법령	공공저작물의 자유이용 규정 신설

자료 : 부처별 데이터개방 제약법령 정비계획

- 국내에는 데이터를 다룰 수 있는 전문가는 많지만 빅데이터를 다룰 수 있는 전문인력은 부족한 실정으로 이들의 양성이 시급
  - 빅데이터 큐레이터 양성 프로젝트[11]의 참여자를 늘리고 이들에게 적극적으로 일자리를 매칭
  - 산·관·학 협력 활성화를 통해 공공과 민간의 빅데이터 활용현장의 실무경험을 쌓을 수 있는 기회를 제공
- 개방부터 활용까지 모든 절차에서 수요자인 시민이 편하게 활용할 수 있도록 정보이용 방법을 개편
  - 사용하기가 불편하면 공개된 정보의 이용률이 떨어짐

11 서울시는 청년 미취업자를 대상으로 전문가의 지도로 전문지식과 직무경험을 축적하고 서울시 전략수립 및 신규사업에 활용 가능한 빅데이터 기반 과제 발굴을 위해 서울형 뉴딜 일자리사업의 일환으로 '빅데이터 큐레이터 양성 프로젝트'를 시행

- 서울 열린데이터 광장의 사용자인터페이스(User Interface)에 대해 지속적인 관심과 개선이 필요

1 3 정보보안 강화

- ㅇ 카드정보 유출로 인한 개인정보 보호에 관심도가 증가하고 있는 상황에서 정보보안은 공공정보공개와 밀접한 연관이 있음
- ㅇ 우리나라의 프라이버시에 대한 규제는 매우 엄격한 편으로 데이터를 활용한 상당수의 서비스가 위법에 속하므로 공공정보를 공개할 경우 주의를 요함
- 민간에서 필요한 상당수 공공정보가 개인정보보호법을 적용받아 개방이 어렵기 때문에 민감한 정보는 제거하고 개방할 수 있도록 데이터를 가공
- 민간에서 활용할 것을 고려하여 공공데이터 공개 범위를 결정
- ㅇ 공공정보에 개인정보 항목을 삭제 또는 최소화하여도 의미 있는 데이터가 될 수 있도록 하는 방안 모색
- 예를 들어, 개인의 특성이 담긴 자료를 지도와 융합한다면 민감한 개인정보(주민등록번호 등)를 제거해도 의미 있는 분석이 가능함

2 서울시 대용량 데이터베이스의 효율적 활용방안

2 1 공공정보 공개에 따른 경제적 효과 극대화

- ㅇ 공공정보를 활용한 광고, 정보서비스 등의 관련 산업 발전을 위한 DB산업 지원책을 마련
- ㅇ 연구조직(연구원 등) 및 정책부서(서울시 등)와 긴밀한 업무협력을 통해 공공정보의 활용방안을 다각도로 모색하고 정책연구에 활용

- 다양한 빅데이터 및 공공데이터를 융합하여 활용방안을 마련하기 위한 연구가 필요
- 효율적인 융합데이터 생성을 위한 TF팀 가동을 고려
- 연구조직과 정책부서의 협력을 통해 빅데이터 및 공공정보 이용의 시너지효과 제고

22 시정 집중·핵심사업에 공공정보 및 빅데이터 이용

- 시정사업에 빅데이터를 활용하면 운영 효율성이 증대되어 비용절감 효과가 있을 것으로 판단
- 기존 정책결정방식의 효율성을 증대시키고 취약점을 보완할 수 있음
- N버스 사례는 빅데이터를 활용하여 버스노선을 효율적으로 선정하였으며 노선결정을 위한 논리적 근거(유동인구 등)도 뒷받침
- 면밀한 분석을 통해 핵심사업과 빅데이터를 결합하여 정책수립 방향을 설정
- 서울시 주요 핵심사업에 빅데이터를 적용할 수 있는 분야를 지속적으로 탐색 후 사업에 적용
- 빅데이터 큐레이터와 협업하여 새로운 정책사업을 발굴

23 민간부문 데이터 확보

- 다양하고 긴밀한 분석을 위해서는 공공부문뿐만 아니라 민간부문의 데이터가 필요함
- 공공부문 데이터와 민간부문 데이터 융합으로 시너지효과 기대
- 최신트렌드를 반영한 실시간 정보 등은 민간부분이 우위에 있음
- 민간과 공공이 협력하여 플랫폼을 구축하고 업무를 공유
- 민간과 공공부문에서 보유하고 있는 기존의 빅데이터를 활용하기 위한 알고리즘을 개발

- ㅇ 민간기업과의 교류 및 협력으로 민간데이터를 확보하여 이용할 수 있도록 해야 함
  - 공공부문은 민간부문에 의미 있는 사업을 제안하여 이들이 보유하고 있는 빅데이터(민간데이터) 확보를 위해 노력
- 민간부문의 빅데이터 사업은 여전히 낮은 수익성을 보이고 있는 것이 현실

2.4 빅데이터 자문위원회 상설화

- ㅇ 빅데이터 자문위원회의 상설화 방안 모색
  - 17개 시·도에 창조경제혁신센터가 개설될 예정
  - 창조경제혁신센터 산하에 빅데이터 자문위원회를 설치하여 지역 실정에 맞는 신규 정책발굴 등의 업무를 수행
  - 공공데이터 개방 전담팀인 정보기획단과 긴밀한 협업이 가능하도록 교류 활성화를 추진
  - 빅데이터 자문위원회는 빅데이터 전문가 위주로 구성되어야 하며, 위원장은 반드시 빅데이터 전문가가 맡아야 함

3 빅데이터를 활용한 정책 제안

3.1 사람을 찾습니다

- ㅇ 서울시의 '사람을 찾습니다' 도입 방향
  - 사진, 인적사항 등의 정보를 매칭하여 가출, 미아, 치매노인 등 실종자를 찾기 위한 시스템 구축 방안을 모색
  - 서울시, 안전행정부, 경찰청, 법무부, 어린이재단 등과 실종자 정보를 공유하여 실종자 DB를 구축

- 실종자 찾기를 통해 사회적 불안 및 비용 감소 등의 효과를 얻을 수 있을 것으로 기대

> ※ 부산 금정구 CCTV를 이용한 실종자 찾기 시스템
>
> 실종자가 발생하면 경찰에서 금정구 CCTV통합관제센터에 실종자 사진을 전달해 관내 설치된 CCTV가 실종자를 찾아내는 제도
>
> 실종자가 안면 인식 프로그램이 추가된 CCTV를 지나가게 되면 곧바로 관제센터에서 인식하여 실종자를 발견
>
> 실종자가 발견되면 안면 인식 프로그램이 기존 사진과 대비해 유사율을 분석
>
> 실종자로 확인되면 근처 CCTV 8개 화면이 분할해서 뜨고 관계자에게 문자 전송
>
> 경찰의 보다 신속한 수사가 가능해질 것으로 기대

3.2 내가 낸 세금은 어디에 쓰이나

- 서울시의 '내가 낸 세금은 어디에 쓰이나' 도입 방향
  - 서울시는 예산 등의 정보를 공개하고 있으며 이를 이용하여 시민들이 납부한 세금이 어디에 쓰이는지 알 수 있도록 웹페이지 구축
  - 제공하는 정보를 시민들이 알기 쉽게 하기 위해 시각화함
- 시민들이 낸 세금에 대한 사용출처를 명확히 알 수 있으므로 시정에 대한 신뢰도 향상을 기대할 수 있음
  - 개인의 수입을 입력하면 세금납부액과 세금이 사용되는 항목을 알려줌으로써 납세에 대한 이해를 높여주고 세금의 사용처를 명확하게 인지할 수 있음

> ※ 영국정부의 공공데이터를 활용한 납세정보 확인
>
> 공개지식재단(Open Knowledge Foundation)은 영국 정부에서 제공하는 공공데이터를 바탕으로 세금의 사용처에 대한 정보를 시각화를 통해 보여주고 있음
>
> 개인의 수입액을 지정하면 세금 납부액과 세금이 어떻게 사용되는지에 대한 정보를 보여줌으로써 납세에 대한 이해를 높임(그림 4-1 참조)

그림 4-1 '내가 낸 세금은 어디에 쓰이나' 이미지
출처 : Open Knowledge Foundation(2011)

참고문헌

참고문헌

국가정보화전략위원회, 2011,「빅데이터를 활용한 스마트 정부 구현(안)」

_____, 2012,「스마트 국가 구현을 위한 빅데이터 마스터 플랜」

김종호, 2013, "빅데이터 시대의 정부의 역할, 주간기술동향", 정보통신산업진흥원

머브 에이드리언, 2011, "이제는 빅 데이터이다", Teradata Magazine

미래창조과학부, 2013, 창조경제 실현을 위한 2013 빅데이터 국내 사례집

배동민·박현수·오기환, 2013, "빅데이터 동향 및 정책 시사점", 방송통신정책 제25권 10호, 통권 555호, 정보통신정책연구원

범지인·최성종·송두한, 2013, "빅데이터(Big Data) 활용 사례와 시사점", 농협경제연구소 CEO Focus 312호

서울특별시, 2013, 서울 열린데이터 광장 백서

윤미영, 2013, "공공부문의 빅데이터 활용사례 및 정책적 활성화 방안", 부동산포커스 Vol.61

이규철·원희선, 2012, "新기술(빅데이터) 등장에 따른 경제적 파급효과 및 법(규제) 연구", 정보와 통신

장병열·김영돈, 2013, "빅데이터 기반 융합 서비스 창출 주요 정책 및 시사점", 과학기술정책 제23권 3호

한국데이터베이스진흥원, 2013,「2013 데이터베이스백서」

한국정보화진흥원, 2013,「새로운 미래를 여는 빅데이터 시대(증보판)」

허필선·박광만·박원주·조기성·류원, 2013, "공공정보 민간활용 시장 및 파급효과", 한국전자통신연구원 전자통신동향분석 제28권 제4호

IBM, 2012, "신세계몰, DW에서 빅데이터 분석까지 효과만점의 시스템 구축"

IT world(2013.07.22.), CJ대한통운, 물류 시스템 확장 위해 IBM 파워 7+도입

KISTI(한국과학기술정보연구원), 2013,「빅데이터 산업의 현황과 전망」

McKinsey Global Institute(MGI), 2011, **Big data : The next frontier for innovation, competition, and productivity**

Gartner, 2012, **The Importance of 'Big Data' : A Definition**

Chris Yiu, 2012, **The Big Data Opportunity, Policy Exchange**

Open Knowledge Foundation, 2011, **Where does my money go?**

http://data.seoul.go.kr/ 서울특별시 열린데이터광장

https://data.go.kr 공공데이터 포털

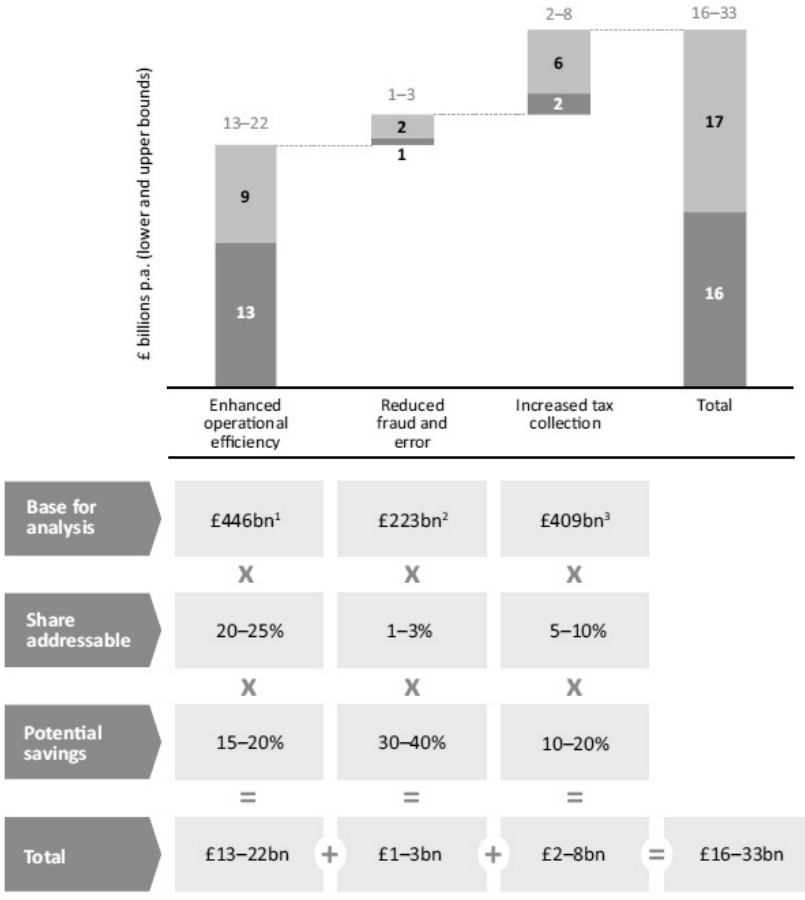
부록

부록

1 영국 공공부문이 빅데이터를 이용하여 절감할 수 있는 비용 추산[12]

- ㅇ 빅데이터를 활용하면 공공부문의 서비스 수준을 높이고 효율을 향상시키며 국가 재정의 건전성을 향상시킬 것으로 기대
  - – 영국의 공공부문 생산성은 십 년 이상 큰 변화가 없었음
  - – 빅데이터를 잘 이용한다면 공공부문의 생산성 향상이라는 변화를 가져올 수 있을 것으로 기대
- ㅇ Chris(2012)는 영국 공공부문 빅데이터 이용에 따른 잠재적 비용 절감 효과를 160억파운드에서 330억파운드로 추정
  - – 이는 1인당 연간 약 250~500파운드를 절감하는 것과 같음
  - – 정부 총예산인 7,000억파운드의 2.5%~4.5%에 해당
  - – 부문별 비용절감 효과
    - • 정부운영의 효율성 증대 부문의 비용절감효과는 130억파운드에서 220억파운드로 추정
    - • 복지제도 내 부정 및 오류의 근원을 찾아내 이를 감소시켜 재정부담을 완화하는 금액은 10억파운드에서 30억파운드로 추정
    - • 조세제도 개선에 따른 세입 증가 분야의 비용절감 효과는 20억파운드에서 80억파운드로 추정

12 Chris Yiu(2012)의 "The Big Data Opportunity"의 내용을 요약



부록 그림 1 영국 공공부문 빅데이터의 잠재력

출처 : Chris(2012)

2 N버스 이용에 따른 서울시민 교통비 절감 계산

2.1 야간통행 이동비용 절감

- ㅇ 일 평균 이용승객 6,298명(A)
- ㅇ N버스 이용 승객 평균 승차거리 8.2㎞(B)
- ㅇ 택시 요금 계산(B=8.2㎞[13]) 결과 8,900원(야간할증 고려) → (a)
- ㅇ 심야버스 이용요금 1,850원 → (b)
- ㅇ 1인당 약 7,050원, 1일평균 4,440만원의 절약효과
- – (A) × {(a)-(b) } = 44,400,900원(1인 1택시 이용)
- – 참고 : (A) ÷ 1.75 × {(a)-(b)} = 25,371,943원(1.75인 1택시 이용)

2.2 이산화탄소 배출 감소

- ㅇ 심야버스 1일 총 운행거리 : 3,591㎞ → (c)
- ㅇ CNG버스 1㎞당 이산화탄소 배출량 : 721g → (d)
- ㅇ 심야버스 이산화탄소 배출량 : 2,589,111g → {(c) × (d)} = (e)
- ㅇ 심야버스 미이용 시 1일 승용차 운행거리 : 29,511㎞ → (B) × (C) = (f)
- ㅇ 승용차 이산화탄소 배출량(아반떼 기준 134g/㎞) : 3,954,474g → (f) × 134g = (g)
- ㅇ 1일 1,365㎏[{(g) - (e)} × (A)], 연간 498톤의 이산화탄소 배출 억제 효과
- – 30년생 소나무 75,506그루가 흡수하는 양

---

13 - 8.2㎞ 산출근거
① N버스가 확대시행된 시점(2013.09.13)부터 연말(2013.12.31)까지의 이용객에 대한 교통카드 자료를 분석
② 승객 1인당 이동거리를 분석
- 8.2㎞ 산출의 주요 데이터
① N버스 전체 노선의 정류장 간 거리
② N버스 이용건별 출발정류장 및 도착정류장, 승객수 정보
- 출발 및 도착정류장 정보와 정류장 거리 정보를 이용
① 건별 이동거리를 산출
② 이용객 수를 반영한 후 평균 산출

23 유류비 절감 효과

- 승용차 운행 시 유류(휘발유 1,850원/ℓ) 1㎞당 0.1ℓ 사용 → (i) = 0.1 ℓ/㎞
- 유류절감량 2,951ℓ/일→ (f) × (i) = (k)
- 1일 546만원{(k) × 1,850원}, 연간 20억원의 유류비 절감 효과

서울연 2013-OR-62

빅데이터 활용과 정보공개의 경제적 효과 분석

발행인 이창현
발행일 2014년 2월 28일
발행처 서울연구원
137-071
서울특별시 서초구 남부순환로 340길 57
전화 (02)2149-1234 팩스 (02)2149-1025

비매품 ISBN 979-11-5700-011-1 93320

본 출판물의 판권은 서울연구원에 속합니다.