



### 빅데이터 혁명과 미디어 정책 이슈

정 용 찬 / 정보통신정책연구원 연구위원

요약문 .....	1
1. 서론 .....	2
2. 빅데이터의 특징과 의미 .....	4
3. 빅데이터 활용 사례 .....	9
4. 정책 제언: 데이터 생태계의 재정립 .....	13

## 빅데이터 혁명과 미디어 정책 이슈

### 정 용 찬

정보통신정책연구원 연구위원

\*ycjung@kisd.re.kr, 02-570-4160

\*고려대학교 응용통계학 박사

\*DNI컨설팅 이사(CRM 담당)

\*방송위원회 연구센터 연구위원

\*현 정보통신정책연구원

방송미디어연구실

### 요약문

디지털 경제의 확산으로 우리 주변에는 규모를 가늠할 수 없을 정도로 많은 정보와 데이터가 생산되는 ‘빅데이터(Big Data)’ 환경이 도래하고 있다. 이러한 방대한 규모의 데이터는 미래 경쟁력의 우위를 좌우하는 중요한 ‘자원’이므로 이를 효과적으로 활용할 수 있는 전략의 수립이 시급하다.

‘빅데이터’ 환경은 민간 기업의 경영활동뿐 아니라, 정부를 포함한 공공 부문의 혁신을 수반하는 대변화를 의미한다. 빅데이터 혁명에 효과적으로 대응하기 위해서는 데이터의 생산, 유통, 소비 전 영역에서 패러다임의 전환이 필요하다.

특히 ‘인터넷(네트워크)’은 스마트 생태계에서 ‘자원(데이터)’이 생산되는 원천으로 새로운 데이터 생태계의 발전은 미디어 정책과 밀접한 관련이 있다. 빅데이터의 확산이 미디어와 IT 부문에 미치는 파급효과를 분석하고, 대응전략을 마련하기 위한 마스터플랜의 수립이 시급하다. 실행전략으로 스마트 환경에 부합하는 미디어 데이터 생산 체계의 재정비, 부가가치 창출이 가능한 빅데이터 유통체계의 혁신, 데이터 생태계(data ecosystem) 활성화를 위한 인프라·제도의 정비를 들 수 있다.

정책 기관은 빅데이터 생태계 환경의 선순환을 돕는 ‘촉매자’이자 ‘중재자’ 역할을 담당해야한다. 공공과 민간 영역의 데이터 공유와 활용을 촉진시키기 위해서는 정부데이터의 공개가 우선되어야 한다. 이는 생산데이터의 품질을 높이고 투명하고 효율적인 정책 집행은 물론 데이터공유와 협업을 통한 새로운 부의 창출에 기여할 뿐 아니라 궁극적으로는 소통과 신뢰에 바탕을 둔 지식기반 사회로 진입하기 위한 출발점이다.

## 1. 서론

- 디지털 경제의 확산으로 우리 주변에는 규모를 가늠할 수 없을 정도로 많은 정보와 데이터가 생산되는 ‘빅데이터(Big Data)’ 환경이 도래
  - PC와 인터넷, 모바일 기기의 보급으로 사람들이 도처에 남긴 발자국(데이터)은 기하급수적으로 증가하고 있음
  - 사용자가 직접 제작하는 UCC를 비롯한 동영상 콘텐츠, 휴대전화와 SNS(Social Network Service)에서 생성되는 문자 등은 데이터의 증가 속도뿐 아니라, 형태와 질에서도 기존과는 다른 양상을 보임
  - 특히 블로그나 SNS에서 유통되는 텍스트 정보는 내용을 통해 글을 쓴 사람의 성향뿐 아니라, 소통하는 상대방과의 연결 관계까지도 분석 가능함
- 이러한 방대한 규모의 데이터는 미래 경쟁력의 우위를 좌우하는 중요한 ‘자원’이며, 이를 활용할 수 있는 전략의 수립이 필요함
  - 산업혁명과 IT혁명에 버금가는 스마트 혁명의 시기에 빅데이터는 혁신과 경쟁력 강화, 생산성 향상을 위한 중요한 원천임(McKinsey, 2011)
    - ※ 미래학자 앨빈 토플러는 ‘부의 미래’에서 다가올 ‘제4 물결’의 시대에 부를 창출하기 위한 핵심 요소는 ‘지식’이며, 거대하게 유통되는 정보에서 유용한 지식을 구별해 내기 위한 과학적 방법의 중요성을 강조
  - 민간 기업은 1990년대부터 통합고객 데이터 분석을 통해 마케팅 활동을 추진하는 고객관계관리(CRM, Customer Relationship Management)를 실행하고 있으며, 이를 위한 데이터웨어하우스(Datawarehouse)를 구축·운영함
  - 최근에는 분산처리 등 빅데이터 기술을 활용해 대용량 고객정보를 실시간으로 분석하는 시스템을 구축하고, 트위터와 인터넷에 유통되는 방대한 규모의 기업 관련 검색어와 댓글을 실시간으로 분석해 자사의 이미지와 제품 및 서비스에 대한 고객의 반응을 파악하여 대응전략을 세우고 있음

- ‘빅데이터’ 환경은 민간 기업의 경영활동뿐 아니라, 정부를 포함한 공공 부문의 혁신을 수반하는 패러다임의 변화를 의미
  - 미 대통령 과학자문위원회는 ‘디지털 미래 전략(Designing a Digital Future)’ 보고서에서 ‘모든 연방 정부 기관은 빅데이터 전략이 필요함’을 강조(2010. 12)
  - 미국과 싱가포르 정부는 재난방재와 테러감지, 전염병 확산 등 국가 위협을 예측하고, 분석하는 빅데이터 기반 위협관리 계획을 추진
  - OECD, Eurostat 등 국제기구는 인터넷 공간에 쌓이는 방대한 규모의 데이터와 접속정보를 활용해 정보사회를 대변할 수 있는 지표를 작성하는 방안을 비용과 기술 관점에서뿐 아니라, 사회정치적 관점에서도 검토 중임
  - 2012년에 열린 다보스 포럼에서는 위기에 처한 자본주의를 구하기 위해 제시한 4가지 모델 중의 하나인 ‘사회 기술 모델(Social and Technological Models)’ 영역에서 ‘빅데이터’가 사회현안 해결에 강력한 도구가 될 것으로 예측함
  - 우리나라 국가정보화전략위원회는 ‘빅데이터를 활용한 스마트 정부 구현(안)’을 보고(2011. 10)
- 본 리포트에서는 ‘빅데이터’ 환경에서 민간과 공공부문의 활용 사례를 살펴보고, 미디어 분야에서의 정책 이슈를 중심으로 데이터 생태계의 재정립을 위한 정책 방향을 모색함

## 2. 빅데이터의 특징과 의미

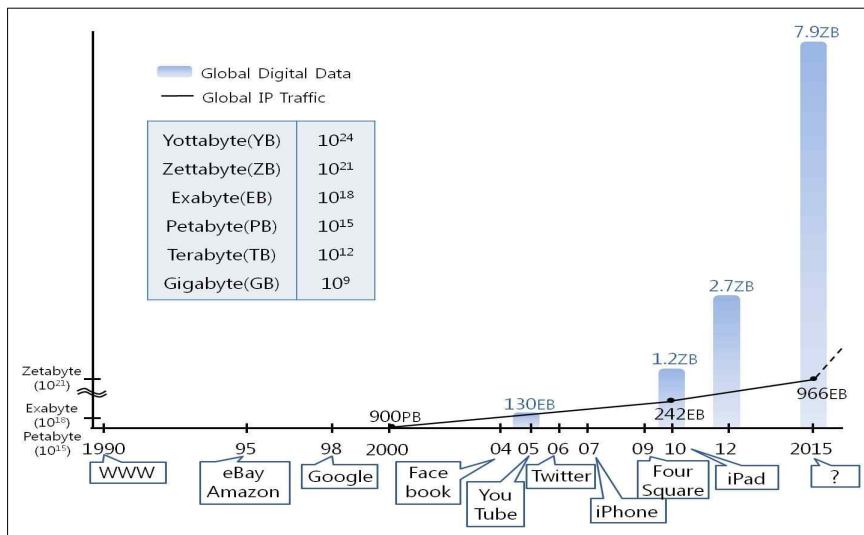
### ◆ 빅데이터의 정의와 특징

- 빅데이터(Big Data)란 데이터의 양, 생성 주기(실시간 생산), 형식(수치 데이터뿐 아니라 문자와 같은 비정형 데이터 포괄) 등에서 과거 데이터에 비해 규모가 크고, 형태가 다양하여 기존의 방법으로는 수집, 저장, 검색, 분석이 어려운 방대한 크기의 데이터를 의미함

- 컴퓨터와 인터넷 이용의 확산, 스마트폰을 포함한 디지털 기기의 보급, 신용카드와 온라인 상거래의 증가, 동영상 콘텐츠의 유통, SNS 이용의 확대 등으로 생성되는 데이터의 규모는 기하급수적으로 증가하고 있음
- 글로벌 데이터 규모는 2012년에 2.7제타바이트(zettabyte,  $10^{21}$  byte), 2015년에는 7.9제타바이트로 증가할 것으로 예측(IDC, 2011)

※ 1엑사바이트(exabyte,  $10^{18}$  byte)는 미 의회도서관 인쇄물의 10만 배에 해당하는 정보량(Lynman, P., & Varian, H., 2003).

[그림 1] 글로벌 디지털 데이터 규모 예측



자료: Cisco(2011)와 IDC(2011)의 예측을 인용하여 작성.

- 현재의 빅데이터 환경은 인터넷을 기반으로 한 디지털 경제 확산의 결과이며, 민간, 공공부문 전 영역으로 확산될 전망
  - 스마트TV를 포함한 스마트 기기와 SNS 외에도 CCTV, 사물통신(M2M)의 확산 등은 유통, 금융, 제조, 의료, 교통 등 사회 전 분야의 데이터 폭증을 의미
  - 센서스(Census), 국세자료, 의료보험, 연금 등 공공 부문에서도 데이터가 양산 중이며, 스마트워크(smart work)의 본격화도 데이터 증가를 가속화함
- ※ 정부의 '스마트워크 활성화 전략'에 따르면 2015년까지 스마트워크 근무율은 전체 노동인구의 30%까지 확대될 예정

〈표 1〉 빅데이터 환경에 영향을 미친 주요 사건

구분	주요 사건
1990년대	<ul style="list-style-type: none"> <li>- 최초의 상업용 웹사이트(Digital Equipment Corporation) 등장(1993)</li> <li>- 온라인쇼핑(eBay, Amazon) 출범(1995)</li> <li>- AOL 메신저 등장(1996)</li> <li>- 구글의 검색기반 광고 서비스 등장(1998)</li> </ul>
2000년대	<ul style="list-style-type: none"> <li>- 블랙베리 이메일 가능 휴대전화(2001)</li> <li>- Facebook 서비스 시작(2004)</li> <li>- YouTube 온라인 서비스(2005)</li> <li>- Twitter 서비스 개시(2006)</li> <li>- 애플 iPhone 등장(2007)</li> <li>- 소셜 커머스 Groupon 등장(2008)</li> <li>- 위치기반 소셜 네트워크서비스 foursquare 등장(2009)</li> </ul>
2010 이후	<ul style="list-style-type: none"> <li>- iPad 출시(2010)</li> <li>- 구글 고객센터정보통합 관리 발표(2012)</li> <li>- 통합커뮤니케이션 서비스(Rich Communication Suite) 등장(2012)</li> </ul>

자료: Craig, T and Ludloff, M. E.(2011)의 내용을 참고하여 작성

### ◆ 빅데이터의 의미와 중요성

- 대용량의 데이터를 분석하는 기업의 마케팅 활동은 이전에도 존재했으나, 현재의 빅데이터 환경은 과거와 비교해 데이터의 양은 물론 질과 다양

### 성 측면에서 패러다임의 전환을 의미

- 기업 차원에서는 보유하고 있는 고객 데이터를 활용하여 마케팅 활동을 활성화하는 고객관계관리(CRM, Customer Relationship Management) 활동을 1990년대부터 시작

※ CRM은 기업이 보유하고 있는 데이터를 통합하는 데이터웨어하우스(Datawarehouse), 고객데이터 분석(Data Mining)을 통한 다양한 마케팅 활동(교차판매-cross-selling, 상향판매up-selling, 고객유지-retention, 이탈방지 등)을 진행하는 것을 뜻하며, 기업의 CRM 활동은 자사 고객의 데이터뿐 아니라 제휴회사의 데이터를 활용한 제휴 마케팅도 포함

- 특히 구매 이력 정보와 웹로그(web-log) 분석, 위치 기반 서비스(GPS) 결합을 통해 소비자가 원하는 서비스를 적기에 적절한 장소에서 제안할 수 있는 기술적 기반을 갖추

- 현재의 빅데이터 환경에서는 과거에 비해 데이터의 양이 폭증하였다는 점과 함께 데이터의 종류도 다양해져 단순히 구매 이력뿐 아니라, 위치정보와 SNS를 통해 고객의 생각과 의견까지도 분석 가능

※ 이용자가 8억 명에 달하며, 매일 30억 회 이상 동영상 검색되는 '유튜브는 선호하는 동영상 채널'을 구성할 수 있는 개별 홈페이지를 제공. 이는 영상 이용 정보, SNS 정보, 인적 네트워크 정보의 결합 분석을 통해 개인 맞춤형 서비스의 제공이 가능함을 의미

- 관련 장비와 소프트웨어 측면에서도 기존의 비싼 스토리지와 데이터베이스, 고비용의 데이터웨어하우스 구축에서 탈피하여 분산병렬처리 기술, 클라우드 컴퓨팅 등을 활용해 효율적인 시스템 구축이 가능

〈표 2〉 빅데이터 환경의 특징

구분	기존	빅데이터 환경
데이터	- 정형화된 수치자료 중심	- 비정형의 다양한 데이터 - 문자데이터(SMS, 검색어) - 영상데이터(CCTV, 동영상) - 위치데이터



구분	기존	빅데이터 환경
하드웨어	- 고가의 저장장치 - 데이터베이스 - 대규모 데이터웨어하우스(Datawarehouse)	- 클라우드 컴퓨팅 등 비용효율적인 장비 활용 가능
소프트웨어/분석 방법	- 관계형 데이터베이스(RDBMS) - 통계패키지(SAS, SPSS) - 데이터 마이닝(data mining) - machine learning, knowledge discovery	- 오픈소스 형태의 무료 소프트웨어 - NoSQL, Hadoop, Cassandra, MapReduce - 오픈 소스 통계솔루션(R) - 텍스트 마이닝(text mining) - 온라인 버즈 분석(opinion mining) - 감성 분석(sentiment analysis)

- 특히 빅데이터에 기반한 분석방법론은 과거에 불가능했던 일을 가능하게 만들
  - 구글의 독감 동향 서비스([google.org/flutrends](http://google.org/flutrends))는 관련 단어 검색어 빈도를 분석해 독감 환자 수, 유행 지역 등을 예측하는 것으로, 이는 미 질병통제본부(CDC)보다 예측력이 뛰어남. 구글은 EC 문서 자료와 수천만 권의 도서 정보를 활용하여 50개 언어 간의 자동번역 시스템 개발에 성공
    - ※ IBM도 캐나다 의회의 '수백만 건'의 문서를 활용, 영어-불어 자동번역 시스템 개발을 시도했으나 실패. 이는 사용 데이터의 규모의 차이에 의한 결과임
  - 서울시장 보궐선거에서도 전통적인 여론조사 결과는 선거 당일까지 '박빙'의 승부를 예상했지만, 트위터 분석은 당선자 측의 우위를 예측
- 빅데이터의 활용은 기업의 입장에서는 고객의 행동을 미리 예측, 대처 방안을 마련하여 기업경쟁력을 강화시키고, 생산성 향상과 비즈니스 혁신을 가능하게 함
  - 미국 기업의 대부분은 가치를 창출할 정도의 정보량인 100terabyte 이

상의 정보를 보유 중(McKinsey, 2011)

- 공공기관의 경우에도 시민이 요구하는 서비스를 제공할 수 있는 ‘기회’로 작용하며, 이를 통해 ‘사회적 비용 감소와 공공 서비스 품질 향상’이 가능

### 3. 빅데이터 활용 사례

#### ◆ 해외 활용 사례

- Economist가 전 세계 약 600개 기업을 대상으로 한 조사에서 대상자의 10%는 빅데이터가 기존 비즈니스 모델을 완전히 바꿀 것이며, 46%는 기업 의사결정의 중요한 요소로 작용할 것으로 응답
  - 그러나 응답자의 25%는 기업 내부에 사용 가능한 데이터는 충분하지만 대부분의 데이터를 방치하고 있으며, 53%는 일부만 활용하고 있다고 응답해 부가가치 창출을 위해서는 더 많은 노력이 필요함을 시사(Economist Intelligence Unit, 2011)
- McKinsey는 의료, 소매, 제조, 개인 위치정보, 공공의 5개 분야별로 빅데이터 활용 사례와 효과를 소개(McKinsey, 2011)
  - 특히 EU의 공공행정 부문에서는 15~20% 행정비용(최대 3,000억 유로)이 절감 가능하며, 생산성은 향후 10년간 0.5% 향상 가능
  - 영국과 스페인은 정부 데이터베이스를 일반에 공개하는 ‘open data’ 정책 추진, 정부데이터와 소셜미디어 데이터를 결합하여 유용한 정보 생산
  - 독일 연방노동부는 대규모 실업데이터 분석을 통해 실직자 개개인에게 맞춤형 일자리를 제공하여 100억 유로 비용을 절감

〈표 3〉 빅데이터 활용 해외 사례

분야	활용 분야	성과
의료	의료 기관별 진료방법, 효능, 비용데이터를 분석하여 효과적인 진료방법 파악 가능 투약과 효능데이터 분석, 저비용 고효율 약 구매 환자 치료데이터 공유, 치료 업무 효율화 전국 의료데이터 연계, 전염병 발생 등 대처	연간 \$3,000억 이상의 비용 절감
소매업	마케팅 영역에서 교차판매(cross-selling), 위치기반 마케팅, 고객세분화, 감성 분석(sentiment analysis) 상품화 계획에서는 가격과 디자인 최적화	최소 0.5%의 생산성 향상 이윤 60% 이상 증가 예상

분야	활용 분야	성과
제조업	제품 개발 공급망 관리(supply chain management) 분야 수요예측	제품개발비 20~50% 절감 경상비 10~25% 절감
개인 위치 정보	통신, 소매, 미디어가 결합된 영역으로 새로운 가치 창출이 가능한 분야 스마트 내비게이션(smart navigation) 위치기반 광고(2020년까지 글로벌 광고시장의 5% 규모로 성장할 것으로 예상)	향후 10년간 새로운 서비스 공급자 매출은 \$1,000억 이상, 최종 소비자는 \$7,000억 이상 혜택 예상
공공	정부 데이터베이스 공개 맞춤형 구직 추천 시스템 SMS를 활용한 간편한 세금신고	행정비용 15~20% 절감 생산성은 향후 10년간 0.5% 향상

- 美 미시간 주는 관련 정부기관 통합 데이터웨어하우스(Integrated Data Warehouse, IDW) 구축으로 시민에 대한 보다 나은 서비스를 제공하고 비용 절감(<http://www.youtube.com/watch?v=ilRWnIJDnRE>)
  - 미시간 주의 21개 정부기관 데이터 통합으로 공공의료보험(Medicaid) 부정행위 발생 감지, 개인 건강관리 개선, 최적의 입양가정 선택 등 서비스 품질 개선
- 美 오하이오 주와 오클라호마 주 정부는 국세청(IRS) 데이터와 고용데이터로 마이닝(mining) 분석을 통해 새로운 세원과 미납세금 확인(The Wall Street Journal, Mar. 12, 2010)

### ◆ 국내 활용 사례

- 국내 공공부문의 빅데이터 활용은 시작 단계에 불과
  - 공공부문도 민간 기업의 CRM 사례를 벤치마킹하여 PCRM(Public CRM)을 도입·시행

- 고객 만족을 최우선으로 하는 서비스 정신의 확산에는 기여하였지만, CRM이 추구하는 고객데이터 분석을 기반으로 한 맞춤형 서비스 제공에는 한계가 있었음
- 이는 공공부문 CRM의 실패라기보다는 고객과의 상호작용(데이터)을 기반으로 고객의 요구를 파악하여 새로운 서비스를 창출하는 CRM의 기본 사상을 공공부문에 적용하는 데에 한계(데이터의 빈곤, 새로운 서비스 제약)가 있었기 때문임
- 최근 인터넷에 산재한 다양한 웹문서, 댓글 등을 통해 특정 이슈에 대한 시민의 의견을 분석하여 대응책을 마련하는 오피니언 마이닝을 도입하는 사례가 늘고 있음
- 국민권익위원회는 연간 약 300만 건에 달하는 민원을 분석하는 ‘민원 동향분석시스템’을 구축
  - 교육, 복지 등 주요 정책과 사회적 이슈 관련 민원의 동향과 원인을 분석하고, 관련 정보를 각 부처에 주간단위로 제공할 예정(2012년 업무보고)
  - 민원 발생 이전에 원인을 차단, 국민의 국정 신뢰도와 행정 서비스 만족도를 높일 계획
- 국민연금공단은 고객관계관리(CRM) 시스템을 구축(2010), 연금제도에 대한 고객의 불신을 해소하고, 소통 확대를 통한 업무효율을 높이고 고객의 요구 변화에 능동적으로 대처하기 위해 노력
  - 고객정보의 통합으로 고객맞춤형 서비스를 적기에 실행
  - ‘여론정보수집분석시스템’을 통해 주요 이슈에 대해 선제적 대응
- 국가정보화위원회에서는 향후 공공부문의 빅데이터 활용 시나리오를 다음과 같이 제시함
  - 과거 재해기록과 사전 피해 예측, 유관기관의 실시간 정보 분석, 소셜 데이터 분석 등을 기반으로 한 실시간 전조감지를 통해 대형재난으로 확산되는 것을 사전에 차단

- 축산 관련 기관별로 산재한 데이터를 융합, 분석하여 가축질병 예방 및 통제를 위한 선제적 의사결정
- 다양한 복지정보 수집을 통한 사회복지통합관리망의 구축으로 맞춤형 복지 서비스 제공
- 생산, 소비, 물류정보를 실시간 모니터링하여 물가 관리와 물가 안정 대책 수립
- DNA, 의료데이터 공유와 활용 촉진을 통해 개인맞춤형 의료 시스템 구축

## 4. 정책 제언: 데이터 생태계의 재정립

- 빅데이터 혁명에 효과적으로 대응하기 위해서는 데이터의 생산, 유통, 소비 전 영역에서 패러다임의 전환이 필요함
  - ‘데이터’는 산업혁명기의 석탄과 같이 스마트 시대의 중요한 ‘자원’임을 인식하고, 이를 효과적으로 관장할 수 있는 정책적, 조직적 대책 마련이 필요
  - 특히 ‘인터넷(네트워크)’은 스마트 생태계에서 ‘자원(데이터)’이 생산되는 원천이므로 새로운 데이터 생태계의 발전은 미디어 정책과 매우 밀접한 관련이 있음
  - 따라서 ‘빅데이터 확산에 따른 미디어와 IT 부문 영향 분석과 대응전략’의 수립이 필요

### ◆ 스마트 환경에 부합하는 미디어 데이터 생산 패러다임 전환

- 미디어 관련 통계생산 체계의 재검토와 설계 필요
  - 인터넷 공간에 쌓이는 데이터는 정부승인통계(official statistics)로 활용할 가치가 충분함
  - OECD는 새로운 기술의 발전으로 통계의 원천이 인터넷 영역으로 전환되는 것에 주목하고, 인터넷 기반 통계에 대한 타당성 연구를 추진(OECD, 2010)
  - Eurostat의 경우 인터넷 접속정보를 활용한 정보사회 지표 작성에 대한 가능성을 검토 중이며, 승인통계로 활용하기 위한 품질 보증 방안 등 관련 연구를 확대(Wirthmann, 2011)
  - 네덜란드 정부(Ministry of Economic Affairs)도 인터넷TV 이용자의 유형별로 이용시간 데이터 분석 관련 연구를 진행(Dialogic, 2008)

※ 우리나라 통계청은 고용보험, 국민연금, 건강보험, 산재보험 등 기존 정보를 활용한 임금근로일자리통계 개발을 계획 중이며, 센서스, 경제총조사도 행정 자료로 대체할 계획임

● 전통적인 통계조사를 보완하기 위한 통계생산 체계 재정립

- 미디어 소비행태는 ‘개인미디어를 통한 소비’에서 참여와 공유를 의미하는 ‘소셜 소비’로 진화 중에 있으며, 단순히 방송 프로그램을 시청하는 데 머물지 않고 시청 중에도 의견을 나누고 시청 전후에 프로그램 관련 정보와 느낌을 공유하는 ‘소셜TV’가 확산되고 있음
- 따라서 SNS나 블로그 분석을 통해 일반적인 통계조사에서 파악하기 어려운 트렌드를 실시간으로 파악하여 기존 통계조사 데이터와 보완 관계를 형성할 필요가 있음

● 미디어 이용행태 변화와 디지털 환경에 부합하는 미디어 측정 자료의 생산이 시급

- 표본 가구 내에 설치된 피플미터(people meter)를 이용한 TV 시청률 측정 방법도 디지털 환경에 부합하는 새로운 측정방식으로서의 변화 모색이 필요
- 특히 방송 콘텐츠 이용이 TV수상기에서 PC, 모바일 기기로 옮겨감에 따라 시간과 공간의 제약을 받지 않는 시청행태로 변화하고 있는 상황에서 기존의 측정방식을 보완하기 위한 다양한 측정 방법과 기술 개발이 진행되어야 함

※ 미국 케이블업계의 조인트벤처인 Canue Venture는 미 광고주협회(ANA)와 공동으로 양방향TV 시청자의 시청률 측정뿐 아니라, 광고효과까지 측정하는 ‘CEE MCC 프로젝트’를 진행. 미국 뉴욕의 TRA는 170만 가정의 초당 시청습관을 담은 데이터와 250만 개의 DVR 및 케이블 셋톱박스 정보와 식료품 구매 기록을 분석해 TV광고 가치를 측정(IDG Korea, 2011)

● 미디어 융합환경에서의 통계 측정 기준 표준화, 미디어 분야의 국가 기간 통계와 부문 통계 간의 연계 강화 등이 필수적임(정용찬, 2012. 12)



- ‘분산형 통계제도(해당 부처가 필요통계를 생산)’를 채택하고 있는 우리나라의 경우 통계조정을 위한 control tower의 역할이 필요
- 부처 내에 관련 통계 생산을 포함하여 빅데이터 환경을 대비하기 위한 전담 조직이 필요함

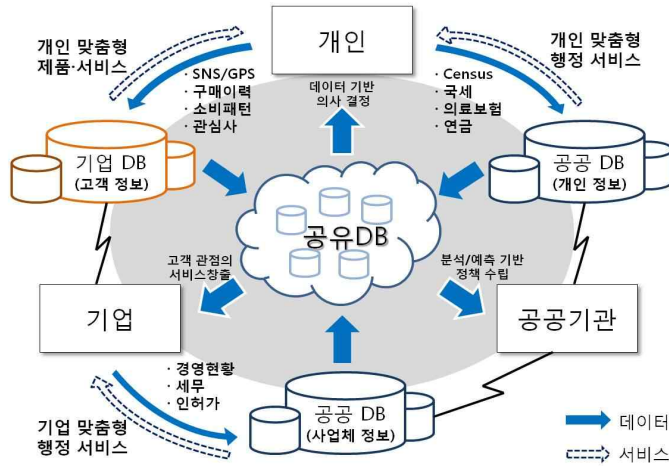
※ 미국 대통령행정집무실 관리예산처(OMB) 소속의 통계수석(Chief Statistician)은 80여 통계작성 기관의 통계 예산안 검토, 정보수집 승인, 통계기준 수립, 기타 통계정책과 조정 업무 기능을 담당

#### ◆ 부가가치 창출이 가능한 빅데이터 유통체계 혁신

- 통계데이터는 태생적으로 통계소비자가 생산과정에 참여하는 사용자제작 콘텐츠(UCC)임(정용찬, 2008)
  - 정부와 기업이 보유하고 있는 데이터의 원천은 ‘개인’임에도 불구하고 이용에 있어서는 제약이 심한 ‘정보불평등’이 존재함
  - 사회 공동의 자산인 데이터의 부가가치를 높이기 위한 활용방안을 적극적으로 모색해야 함
- 정부가 생산하는 데이터를 우선 개방하고, 민간 부문에서 생산하는 데이터의 공동 활용 방안 모색
  - 데이터 공개(공유)는 중복생산 제거를 통한 비용절감뿐 아니라, 공개를 통한 데이터 품질 제고와 투명하고 효율적인 행정이 가능하며 궁극적으로는 새로운 지식 창출에 기여함
  - ※ 정보공개와 청구권이 헌법에 보장된 스웨덴을 비롯해 핀란드 등 정보공개 강국은 국가경쟁력 순위에서도 상위권임
  - 빅데이터 환경에서 정부와 기업이 생산하는 데이터의 소유권과 재사용 등의 활용방안에 대해 개인, 기업, 정부 등 이해당사자 간의 소통과 합의가 필요하므로 단계적인 공유 전략이 필요

- 데이터 공유 환경 조성을 위해서는 어디에 어떤 정보가 있는지를 확인하기 위한 ‘데이터 지도’ 작성이 우선되어야 함
  - 방대한 데이터를 효과적으로 활용하기 위한 메타데이터(metadata)를 포함하여 데이터에 관한 정보가 집약된 ‘데이터웨어(dataware)’의 구축이 필요
  - 데이터웨어는 기존 데이터의 상태뿐 아니라 향후 생산(조정, 개선, 대체, 중지 등을 포함)과 유통 계획을 망라한 종합정보서가 되어야 함
- 소통과 협업에 기초한 지식 창출 기반 구축
  - 데이터 공개와 공유로 촉발된 빅데이터 유통체계의 성공여부는 이를 활용한 지식 창출에 있음
  - 지식 창출 과정은 정부와 기업, 전문가 집단뿐 아니라 다수의 일반 시민이 참여하는 크라우드소싱(crowdsourcing) 전략을 지향
    - ※ 기업의 제품생산이나 서비스 과정에 소비자를 참여시키는 크라우드소싱 전략을 빅데이터 환경에서는 정부정책 수립 과정으로 확대 적용
- 데이터 공유를 통한 다양한 서비스 제공은 소비자의 편익을 증대시키는 기회인 동시에 프라이버시의 침해라는 개인정보 보호 이슈를 제기함
  - 프라이버시를 과도하게 침해하지 않는 범위 내에서 데이터의 활용을 활성화할 수 있는 제도의 마련이 필요
  - 프라이버시의 상품화와 기본 인권 보호에 대한 사회적 합의를 기반으로 개인 정보의 거래와 활용, 정보보호의 범위와 수준 논의
    - ※ 프라이버시를 사고 팔 수 있는 상품으로 보는 미국과, 기본 인권으로 간주하고 있는 유럽의 사례 벤치마킹 필요(Craig, 2011)

〔그림 2〕 데이터 공유 개념도



#### ◆ 데이터 생태계(Data ecosystem) 활성화를 위한 인프라·제도 정비

- 빅데이터 전략은 기업의 경쟁력 강화 관점이 아닌 정부와 사회 구성원 전체의 생산성 혁신과 경쟁력 강화 차원에서 접근할 필요가 있음
- 정책기관은 빅데이터 생태계 환경의 선순환을 돕는 ‘촉매자’이자 ‘중재자’로 다음과 같은 분야의 범정부 차원의 대책과 역할 분담이 필요
  - 데이터 생산 원천이며 유통망인 인터넷 인프라의 구축 전략을 빅데이터 관점에서 재검토
  - 공공과 민간 부문의 데이터 공유와 활용을 촉진시키기 위한 제도 개선
  - 빅데이터 기반 제품과 서비스 산업 활성화를 위한 제도와 법제 정비
  - 빅데이터 관련 기술 개발과 R&D 정책 추진
  - 글로벌 기업의 데이터 공유를 위한 제도적 기반 구축

※ ‘한국 드라마(Korean Drama)’가 별도 장르로 구분된 ‘홀루(hulu.com), ‘K-Pop’을 음악 장르로 신설한 유튜브의 이용자 데이터 공동 활용방안이 필요함

- 빅데이터 관련 인력 양성 정책의 수립과 추진

※ 미국에서만 데이터 분석 전문가가 14만 명에서 19만 명, 데이터에 익숙한 관리자가 150만 명 부족한 것으로 추정(McKinsey, 2011)

- 빅데이터 환경은 데이터 자원을 보유하고 이를 활용할 능력을 가진 개인과 기업에게는 기회로 작용하지만, 이 과정에서 소외되는 정보 약자가 양산될 가능성이 존재

- 새로운 사회자본인 데이터 인프라의 이용 격차를 해소하고, 동반성장 가능한 지원책을 마련할 필요가 있음

- 기업의 입장에서 마케팅을 위한 고객데이터의 분석에서 공익을 위한 목적으로 활용 범위를 넓히는 노력이 필요하며, 이러한 활동은 궁극적으로 기업의 신뢰를 높여 시장 규모 확대에도 도움이 됨

※ 통신분야의 빅데이터 분석은 사회 공헌(Data Philanthropy)으로 활용 가능함. MIT 연구팀은 휴대전화 통화 패턴의 변화를 통해 독감 발생을 예측했고, 텔레포니카(Telefonica) 연구팀은 통화패턴 분석으로 주택, 교육, 의료 등에 대한 수요를 추정함. 콜롬비아 대학 연구팀은 Digicel의 통화데이터를 통해 아이티 지진 시의 난민 이동 분석(Forbes, Sep. 20, 2011).

- 빅데이터 환경에서는 데이터에 근거하여 불확실한 상황에서 합리적인 판단을 내릴 수 있는 능력인 데이터 리터러시(data literacy)에 대한 교육이 필요함. 또한 데이터 시각화(data visualization) 등 통계정보를 쉽게 이해하기 위한 방법론과 관련 기술 개발이 필요

※ '통계에 기초한 정책'을 의제로 열리는 OECD 세계포럼(OECD World Forum on Statistics, Knowledge and Policy)도 '이해하기 쉬운 통계정보 전달'을 주요 의제로 설정함

- 빅데이터 대응 전략의 최종수혜자는 '시민'임. 따라서 관련 정책의 성공여부는 '시민 고객' 관점의 전략 수립과 실행에 달려 있음

## 참 고 문 헌

### <국내 문헌>

- 국가정보화전략위원회 (2011. 11), “빅데이터를 활용한 스마트 정부 구현 (안)”.
- 정용찬 (2008), 『미디어통계의 이해』, 커뮤니케이션북스
- \_\_\_\_\_ (2009), “미디어의 개인화와 관계의 소비”, 『현대사회와 미디어 비판』, 백산서당.
- \_\_\_\_\_ (2011. 12. 8), “빅데이터 시대의 방송통계 개선”, 《2011년 방송통신 통계 개선 및 수요조사 보고회》.
- \_\_\_\_\_ (2011. 5. 26), “성과와 고객 관점의 통계품질진단 개선”, 《한국통계 학회 창립 40주년 기념 국가통계포럼: 국가통계의 현안과 대응》.
- 개인정보 보호법 (법률 제10465호, 2011. 3. 29, 제정).
- IDG Korea (2011. 10. 4), “DW 선택 가이드2”, 《CIO》.

### <국외 문헌>

- Cisco (2011). *Visual Networking Index 2010-2015*.
- Craig, T and Ludloff, M. E. (2011). *Privacy and Big Data*. O'Reilly.
- Dialogic (2008). “Go with the dataflow! Analysing the Internet as a data source”. Ministry of Economic Affairs, Netherlands.
- Economist Intelligence Unit (2011). “Big Data Harnessing a game-changing asset”.
- IDC (2011). “Big Data Is Here, Now What?” 2011 IDC Digital Universe Study.
- Kirkpatrick, Robert (Sep. 20, 2011). “Data Philanthropy is Good for Business”. *Forbes*.

- Lynman, P., & Varian, H. (2003). "How Much Information?". School of Information Management and Systems, UC Berkeley.
- McKinsey Global Institute (2011). *Big data: the next frontier for innovation, competition, and productivity*. McKinsey & Company.
- Roebuck, Kevin (2011). *Storing and managing big data NoSQL, Hadoop and more*. Tebbo.
- Spiezia, Vincenzo (May. 10, 2010). "Official statistics in the era of ubiquitous connectivity and pervasive technologies", OECD.
- Vital Wave Consulting (2012). "Big Data, Big Impact: New Possibilities for International Development". World Economic Forum.
- The Wall Street Journal* (Mar. 12, 2010). "Cash-Starved States Put Tax Scofflaws in Crosshairs-Budget Woes Push Officials to Mine IRS Data, Try Web Infamy to Get Late Cash".
- Wirthmann, Albrecht (2011). "The feasibility of producing relevant indicators about the Information Society using Internet traffic flows as a data source". 58th ISI World Statistics Congress.