

## 빅데이터의 정책적 활용과 과제\*

— 빅데이터 추진 현황과 정책적 적용 쟁점을 중심으로 —

### Policy Implementation and Alternative Policy Options of Big Data: Issues of Present Condition and Policy Application with Big Data

김 성 우 (부경대학교 공공정책연구소: 주저자)

정 건 섭 (부경대학교 행정학과 교수: 교신저자)

#### *Abstract*

Sung-Woo Kim / Kyoung-Sup Chung

The study is concerned with policy implementation and alternative policy options of big data which has characteristics such as volume, velocity, and variability. Nowadays, what is so called, the issue of big data is ubiquitous in recent years, such as openness public data, solving alternatives of social problems using big data, etc. Especially, the role and issue of government will be focused on the application of big data. It is evident that scientific policy implementation and citizens' administrative convenience should be improved by way of using big data technology in government. Unfortunately, there is no rule out that the gathering and analysis of unreliable big data will cause the error or noise of the policy. In order to reduce the error or noise of the big data related policy, it is necessary to have the expert or specialized person of big data more than anything else. There is no doubt that the number of experts of big data are equal to the regional competitive power in the near future. Notwithstanding, the infrastructure of Information Technology is inferior to the national capital regions; accordingly, the regional governments have to dependent on the central government in terms of the structure in system establishment, analysis, application of big data. This kind of disparity will cause the regional inequality in the era of big data. Therefore, the regional government should foster and support the IT or big data experts with the relation of the regional universities.

**주제어:** 빅데이터, 소셜 네트워킹, 데이터 계급, 정책적 활용

**Key Words:** Big Data, Social Networking Service, Data-Classes, Policy Implementation

\* 이 논문은 2012학년도 부경대학교 연구년교수 지원사업에 의하여 연구되었음(PS-2012-0705).

## I. 서론

과거와 달리 최근의 대중교통 이용자의 모습에 많은 변화를 가져오고 있다. 신문과 같은 인쇄 매체보다는 스마트폰을 이용하는 이용자의 모습이 늘어났다. 출·퇴근길에 주요 뉴스를 검색하거나, 개인 간의 문자 정보를 주고받거나 SNS의 글을 올리거나 댓글을 확인하는 모습이 그것이다. 한편, 가전제품도 스마트 시대를 맞이하고 있다. 집 밖에서 가전제품들의 상태를 모니터링 하거나, 조작 명령을 통해 상태를 변화시킬 수 있다. 이렇듯 최근 IT기술의 발전과 함께 스마트폰의 확산으로 소셜 네트워킹의 활성화되고, 사물간의 정보통신이 늘어나면서 이들이 생성하는 데이터의 양은 기하급수적으로 증가하고 있다.

IT 기술의 발달과 함께 증가하고 있는 데이터의 처리 및 관리 기술도 함께 개발되고 있다. 최근 구글은 빅데이터를 처리할 수 있는 대표적인 프레임워크 기술인 하둡(Hadoop)을 오픈소스로 공개하였다. 하둡(Hadoop)은 비용으로 빅데이터가 처리 가능할 뿐만 아니라 높은 사용편의성으로 제공해 주고 있다. 한편, 아마존, 야후, 페이스북, 트위터 등의 대형 글로벌 인터넷 서비스 업체들도 자사의 기술과 플랫폼을 공개하여 빅데이터 처리기술 발전에 공조하고 있다.

스마트폰과 SNS(social networking service)의 보급으로 실시간 소통이 이루어지면서 광대한 규모(volume), 속도(velocity), 다양성(variability)의 특성을 지닌 ‘빅데이터(big data)’가 발생되고 있다. 더불어 빅데이터에 대한 처리기술에 대한 공개와 함께 비용이 감소하면서 빅데이터는 새로운 일자리 창출과 새로운 경제 성장 동력으로 기대를 모으고 있다. 맥켄지(McKinsey)는 빅데이터를 향후 비즈니스 지형을 변화시킬 10대 기술트렌드 중 하나로 선정하고, 빅데이터를 수집, 저장하여 이를 토대로 정보를 마이닝 하는 것이 중요한 가치창출효과를 가져 올 것이라고 전망하고 있다(김성태, 2013: 16).

최근 정부는 향후 빅데이터가 미래국가경쟁력의 주요한 요소가 될 것으로 인식하고 공공정책 분야에도 빅데이터를 활용한 사회현안 및 행정수요를 파악해 공공서비스의 대응성을 높일 수 있을 것으로 예상하고 있다. 정부3.0에 따른 공공정보와 민간정보의 융합 및 공유체계의 구축과 함께 사회 환경의 급변성과 불확실성을 대비하기 위한 「데이터기반의 정책결정」 과정이 재조명되고 있다. 불확실성의 고위험 사회를 대비하는 것은 시급성이 높은 국가의 중요한 정책적 의제이므로 데이터기반의 정책과정과 의사결정에 대한 고려의 필요성이 높아지고 있다(한국정보화진흥원, 2012b: 34). 지방정부 역시 빅데이터 실현과 관련된 사례를 제시하고 있다.

빅데이터 기술에 대한 밝은 청사진에만 초점이 집중되면서, 빅데이터 기술 활용과

관련된 문제 및 쟁점에 대한 논의는 상대적으로 부족하다. 빅데이터 활용과 관련하여 허성욱(2014), 최혜민(2014) 등은 된 개인정보보호의 문제점을 지적했고, 윤상오(2013)는 빅데이터와 관련된 위험을 유형별로 분류하여 제시하였다. 이재현(2013)은 데이터를 통한 사회현상에 대한 인식 오류를 지적 하였다. 빅데이터 활용과 관련된 문제점을 지적한 국내연구는 소수에 불과하다. 한편, 이들 연구의 경우 대부분이 빅데이터의 활용과 관련된 문제점을 지적하고 있을 뿐 논의를 정책적 활용측면으로 확장시킨 연구는 부족하다. 따라서 본 연구는 최근 이슈가 되고 있는 빅데이터 활용과 관련된 사업 및 이슈를 살펴보고 정책 활용과 관련된 쟁점과 과제에 대하여 살펴보고자 한다.

## II. 빅데이터의 및 활용

### 1. 빅데이터의 개요

빅데이터에 대한 정의는 연구자들마다 다양하게 정의하고 있다. 일반적으로 빅데이터란 기존의 데이터베이스 시스템으로 처리할 수 없는 용량을 넘어선 데이터를 말하는데(오라일리미디어 2013: 14), Newman(2011)는 이러한 빅데이터에 대한 특징을 규모(volume), 다양성(variety), 속도(velocity) 3가지로 규정하였다. IBM은 이러한 3V 요소 중 두 가지 이상을 충족시킬 경우 빅데이터 기술이라고 정의하고 있다. 빅데이터의 가장 큰 특징은 텍스트와 이미지뿐만 아니라 동영상 등 비정형성을 갖고 있다는 것이다. 규모뿐만 아니라 빠르게 전파되기 때문에 중요한 패턴을 찾기가 쉽지 않고, 유용한 정보의 증가만큼 불필요한 정보도 급증하고 있다(최진원·김이연 2012: 168).

데이터의 처리측면에서 Gantz와 Reinsel(2011)은 빅데이터 기술을 아주 방대한 양의 다양한 데이터를 매우 빨리 모니터링 함으로써 그냥 지나칠 수 있던 것을 발견 또는 분석함으로써 그 안에 존재하는 가치를 경제적으로 추출할 수 있게끔 디자인 되어 있는 차세대 기술이며 구조라고 묘사하고 있다(이정미 2013: 58). 안창원·황승규(2012)는 빅데이터는 데이터의 규모가 방대하고(volume), 데이터의 종류가 다양하며(variety), 데이터 처리 및 분석을 적시에 해결해야 하는(velocity) 특성을 가지고 있으며, 그 결과로 새로운 가치를 창출해 낼 수 있어야 하고, 일반적인 데이터베이스로 저장, 관리, 분석 할 수 있는 한계를 넘어서며, 기업정보, 웹, 이미지/동영상, SNS, 센서 스트림 등 정형/비정형 데이터를 모두 포함하고, 분석과 예측에 있어서 실시간 처리 등 적시성을 요구한다고 정의하고 있다.

김현영 외(2014)는 빅데이터를 기존 데이터베이스 S/W를 이용한 데이터 수집·저장·관리·분석의 한계를 넘어서는 큰 규모의 정형 또는 비정형 데이터를 이용한 실시간 분석을 통하여 데이터로부터 숨겨진 가치를 추출해내는 미래예측 기술 및 아키텍처를 의미한다고 정의하였다. 한편 Gatner(2011)은 규모, 속도, 다양성의 특성에 복잡성(Complexity)를 추가하여 4가지 특성을 제시하였다. 데이터의 복잡성이란 구조나 데이터의 획득과 처리에 드는 속도, 도메인이나 규칙, 저장타입 등 데이터의 발생과 처리, 정제 등 모든 과정이 복잡해지는 것을 말한다(최규현 2012: 17).

이상의 정리하면, 빅데이터는 단순한 양이 많은 것을 의미하는 개념을 넘어 규모(volume), 다양성(variety), 속도(velocity), 복잡성(complexity)의 4가지 특성을 가지고 있는 정형 또는 비정형화된 것으로 이것으로 처리·분석하여 그 결과 새로운 가치를 창출해낼 수 있는 기술로 초고속 데이터의 수집, 처리, 분석, 활용 포함된 기술로 정의될 수 있다.

## 2. 빅데이터 시장규모

빅데이터 기술은 2012년 세계경제 포럼에서 10대 기술 중 첫 번째로 선정될 만큼 과학기술적 측면뿐만 아니라 정치, 경제, 사회, 문화 등 다양한 분야에 가치 있는 서비스를 제공할 것으로 기대를 모으고 있다. 오늘날 11억 인구가 SNS를 이용하고, 이 중 2억5천만명이 매일 페이스북에 사진을 업로드하고 있다(김동완, 2013: 40). 지난 10년간의 데이터의 생산량보다 앞으로 2년간 생성될 데이터량은 더 많을 것으로 예측하고 있다. 2011년 한 해 동안 생성된 전 세계 디지털 양은 약 1.8ZB(1.8조 GB)에 이르는데 두 시간짜리 HD급 영화 2,000편과 맞먹는 수치이다(김상락·강만모, 2014: 8). 2020년의 경우 약 50배가량이 증가할 것으로 예상 된다(김민수, 2014: 30). 데이터의 형식면에서도 과거 정형 또는 텍스트 중심의 자료였다면 최근에는 그림, 동영상, 음성 위주의 비정형 데이터가 급속도로 증가하고 있다.

Manyika et al.(2011)은 빅데이터를 활용한 주요산업의 가치는 22.3조 달러에 달하고, 미국의 경우 2018년까지 14~19만 명의 전문 인력과 150만 명의 데이터관리 인력이 필요한 것으로 추정하고 있다. 한국정보화진흥원(2012)은 한국의 경우 향후 5년간 약 52만 개의 한국의 빅데이터 기반 일자리가 창출될 것으로 전망하고 있다. 과거 우리나라의 경우 IMF 극복과정에서 IT 관련 산업의 발전으로 창업과 청년일자리 증가를 경험한 적이 있다. 이처럼 데이터를 효과적으로 관리·활용함으로써 발생하는 경제적 가치뿐만 아니라 인프라 형성과정에서 파급되는 사회·경제적 영향은 매우 클 것으로 전망된다.

## 2. 빅데이터와 정책 활용

빅데이터의 경우 상당부분 위치와 시간이라는 정보를 갖고 있다. 예를 들어 독감증세에 대하여 검색하거나, 부동산 거래를 위해 부동산 정보를 검색할 경우 해당 정보의 위치와 검색시간 등이 여기에 포함된다. 빅데이터는 시간과 장소 혹은 위치정보를 기반으로 다양한 빅데이터와 융합되면서 활용 가치가 증가한다. 실세계에 존재하는 모든 정보의 80% 이상이 위치 또는 공간과 관련되어 있는 상황에서 공간정보는 빅데이터 분석에서 결코 빠뜨릴 수 없는 정보라고 할 수 있다(김민수, 20104: 31). 특히 위치와 같은 정보는 시간이 지남에 따라 매우 빈번하고 다양하게 발생하는 대용량의 특성을 갖고 있다

시간과 공간 정보를 포함한 빅데이터는 언제 어느 지역에서 어떤 현상이 일어나는지를 파악할 수 있기 때문에 효과적으로 사회환경에 대응할 수 있다. 따라서 물리적 환경에 대한 분석뿐만 아니라 사회·경제적 환경·인간의 행태 분석에까지 공간을 융·복합함으로써 복잡한 사회문제 해결에 도움이 될 수 것이다(이영주 외, 2014: 12-13).

<표 1> 정책과정별 빅데이터 활용 방안

정책과정	기존 방법	빅데이터 활용 방법	차이점
현과과제 진단	· 통계자료 분석 · 설문 등 조사	· SNS 등 실시간 자료 수집 및 분석	· 대용량의 자료 · 비용과 시간 절약
대안마련	· 전문가 집단	· 계획지원체계 활용 · 실시간 의견수렴	· 계획과정 투명화 · 다양한 의견수렴
의사결정	· 전문가 집단	· 대안별 과급효과 시뮬레이션	· 과학적이고 투명한 의사결정
집행	· 집행과정에 대한 설문조사 등	· 실시간 의견수렴 · 실시간 효과분석	· 집행에 따른 반응과 효과· 실시간 확인
평가	· 통계자료로 확인 · 설문 등 조사	· 실시간 의견수렴 · 실시간 효과분석	· 집행에 따른 반응과 효과 실시간 확인

자료 : 김대중(2014: 45)

사회문제 해결을 위한 정책결정과별로 빅데이터는 <표 1>과 같이 활용될 수 있다. SNS와 같은 소셜빅데이터의 경우 현실 문제를 해결하는 아이디어가 담겨있거나, 정책 집행과정에서 실시간 의견수렴을 통해 정책집행과정에서 효율성을 높일 수 있을 것이다. 나아가 정책의 효과분석에도 활용될 수 있다. 이들 빅데이터에 대한 축적과 분석기법의 개발은 정책과급효과에 대한 시뮬레이션 기능을 강화해 의사결정과정에서 정책대안을 분석하는데 도움을 줄 수 있을 것이다. 미국, 일본, 영국 등은 국가차원에

서 빅데이터 전략을 수립하고 막대한 R&D투자를 추진하고 있다. 우리나라에서도 2011년 빅데이터를 활용한 스마트 정부를 구현하기 위한 방안을 모색하고 있다(김희수, 2014: 7). 한편, 「공공데이터 제공 및 이용 활성화에 관한 법률」이 2013년 7월에 제정되면서 빅데이터 지원 및 활용과 관련된 제도적 장치가 마련되었다.

### 3. 빅데이터 사업의 추진 현황

빅데이터 관련 분야는 빅데이터의 사용과 화보를 위한 스마트폰, 인터넷 등의 시장 부문에서 치열한 경쟁을 벌이고 있다. 이들 분야 대부분에서 글로벌 기업이 우위를 점하고 있다. 대표적인 기업으로 구글, 애플, 아마존과 같은 서비스 기업뿐만 아니라 IBM, SAS 등 IT기업들도 빅데이터 인프라 및 플랫폼 구축 사업을 확대하고 있다(한국정보화진흥원, 2013: 182). 한국의 경우 스마트폰 개발과 같은 하드웨어적인 부분은 경쟁우위를 나타내고 있지만, 소프트웨어부문의 인프라가 부족해 데이터 활용도가 낮고, 빅데이터 관련 사업 환경 또한 미성숙한 상황이다.

<표 2> 글로벌 기업의 빅데이터 기반 산업 추진

	빅데이터 기반 사업
구글	검색 및 위치 정보, 안드로이드OS 기반의 모바일 콘텐츠 이용정보, 구글+를 통한 소셜정보, 구글박스를 통한 클라우드 서비스 등
애플	iOS 기반의 모바일 콘텐츠 이용정보, 아이튠즈를 통한 콘텐츠 이용정보, iCloud를 통한 클라우드 서비스 등
아마존	인터넷 비즈니스를 통한 사용자 구매 정보 및 콘텐츠 이용정보, 클라우드 서비스, 킨들을 통한 모바일 이용정보 등
페이스북	세계 최대 소셜정보 보유, 콘텐츠 기업과의 제휴/협력을 통한 이용정보 등

자료 : 한국정보화진흥원(2013: 182)

국내 공공부문의 빅데이터 사업은 2012년 빅데이터 분석 기반 정책결정, 업무혁신, 맞춤형 대국민 서비스 제공 등을 위하여 행정자치부, 지식경제부, 교육과학기술부, 방송통신위원회, 국가과학기술위원회 등이 참여하여 ‘스마트국가 구현을 위한 마스터 플랜’을 수립(김민수, 2014: 35) 진행 중에 있다. 행정자치부는 공간정보를 활용한 시도 행정정보시스템이 구축사업이 추진 중이고, 국교통부는 공간빅데이터 구축사업을 진행하고 있다.

&lt;표 3&gt; 빅데이터 추진계획

구분	계	13년	14년	15년	16년	17년
합계(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)

자료: 통계청(2014.1.7.)

정부 각 부처는 빅데이터 지원 시범사업들을 발굴하여 성과를 발표하고 있다. 향후 국정운영 전반에 빅데이터 활용을 본격화 하고, 정부 부처별 공동으로 빅데이터 플랫폼을 활용할 계획이다. 빅데이터 관련 정부지원 규모는 2017년까지 97개 빅데이터 활용사업을 추진하고 빅데이터 관련 2013년 360억원에서 2014년에는 460억원의 예산이 지원된다. 이중 지방정부가 추진하는 빅데이터 사업은 36개로 2014년에는 10개 사업이 추진될 계획이다. 정부 예산규모는 2013년 15억에서 2014년에는 60억원이 지원되어 중앙정부와 지방자치단체 모두로 빅데이터 관련 사업의 활발히 추진 중이다.

국토교통부(2014)에 따르면 국가공간정보정책에 따라 2014년에는 385개의 공간정보사업이 시행되었다. 세부적으로 공간정보 융·복합사업, 실내공간정보사업, 부동산 행정정보 일원화 및 공간정보활용 확산사업, 공간정보기반조성사업 등에 약 2,946억원이 투자되었다. 한편, 공간빅데이터 체계를 구축하기 위하여 2016년까지 공간정보인 지형 및 지적과 함께 행정정보인 인구, 소득, 기상정보, 부동산 가격 및 민간 정보인 SNS 및 검색정보 등에 대한 DB구축 및 분석모델 개발을 위한 사업이 단계별로 추진될 계획이다.

열린정부(open government)의 흐름 속에 공공정보의 개방이 확대되고 있다. Chadwick(2003)는 IT 기술이 정부와 시민사회 그리고 시민사회 내부에 다자간 더 많은 소동과 참여를 촉진시키고, 결과적으로 민주주의를 증대시키는 효과를 가져 온다고 하였다(최창우 외, 2012: 5). 정부는 웹 3.0과 대비되는 정부 3.0으로 개방, 공유, 소통, 협력의 가치를 국정운영의 혁신체계로 삼고 있다. 정부 3.0은 기본적으로 국가운영과정에 발생하는 다양한 데이터를 공개함으로써 새로운 가치를 창출하고, 운영의 투명성과 효율성을 향상시키고자 하는 전략이 진행 중이다(김우영, 2014: 24)

### Ⅲ. 빅데이터의 활용과 정책추진

#### 1. 민간부문 빅데이터 활용 사례

빅데이터의 활용 노력은 민간영역에서는 구글이 가장 발 빠르게 현실화 하고 있다. 구글은 사용자가 입력한 독감 관련 검색어(콧물, 고열 등)의 빈도 등을 분석 하여 지역별, 국가별로 5단계의 독감 유행 수준 예측치를 제공하고 있다(김희수, 2014). 한편, 구글은 사람이 미리 번역한 문서를 기초로 비교하여 실시간 자동번역시스템을 마련하고 있다. 온라인 종합 쇼핑몰 아마존닷컴(amazon.com)은 모든 고객들의 구매 내역을 빅데이터로서 수집, 관리 및 분석함으로써 고객 맞춤형 상품을 메일과 홈페이지 상에서 제공하고 있다(신신애 외, 2014: 7). IBM은 자체 채용시스템인 “Blue Pages”을 통해 전 세계에서 60만 명의 정보를 관리하고 있지만 동료를 찾거나 특정 업무의 책임자를 찾는 데 어려움이 있어 빅데이터의 “in-memory” 기술과 “하둡(hadoop)”을 적용하여 실시간적으로 인력의 정보를 관리하고 있다(최진명, 2012: 6).

SK텔레콤은 소셜 네트워크에서 여론분석을 위한 스마트 인사이트 시스템을 개발하여 기업이 원하는 키워드를 중심으로 온라인 여론을 분석하여 실시간으로 제공하는 서비스를 제공 중에 있다(한국정보화진흥원, 2012a: 62). 기타 개인정보의 많이 보유하고 있는 신용카드 기업들도 빅데이터 기반 소비패턴을 분석하여 개인 맞춤형 마케팅활동에 활용하고 있다. 한편 삼성전자의 경우 빅데이터 시스템 구축의 필요성이 인식 되면서 본격적인 빅데이터 전담조직을 구성하여 빅데이터 분석체계를 구축하고 있다. 현대자동차 역시 차량의 품질 및 서비스 개선을 위한 빅데이터 사업을 추진 중에 있어 제조업 중시의 기업들도 최근 빅데이터 중요성을 인식하고 해당 분야에 대한 투자를 늘리고 있다. 이처럼 빅데이터 기술의 공공 활용은 주민 개개인의 맞춤형 정책을 가능하게 할 것 이다.

#### 2. 공공분야 빅데이터 사업

공공분야는 민간 기업에 비해 많은 데이터를 보유하고 있다. 각국에서는 빅데이터는 새로운 혁신요소로 인정받으면서 전략적으로 국가차원의 사업이 추진 중에 있다. 미국의 경우 2012년 ‘빅데이터 연구개발 이니셔티브(big data R&D initiative)’ 발표 이후 다양한 형태의 빅데이터 관련 지원 프로그램을 추진 중에 있다. 한편 사회문제 해결을 위해 탈세방지와 보건의료, 범죄 및 테러대응에 빅데이터를 활용하고 있다.

영국 빅데이터 활용의 기반이 되는 공공부문의 정보공유 및 활용에 따른 가치창출을 위한 데이터 공개·공유 중심의 정책을 추진하고 있다(윤미영, 2013: 35). 전염병



과 비만과 같은 보건 복지 분야와 자연재해 및 잠재적 재해와 관련된 위험대응 프로그램에 빅데이터를 활용하고 있다. 아시아 국가에서도 싱가포르와 일본 등도 보건, 교통, 화재, 지진 등 다양한 주요 정책현안에 대한 대응력을 높이기 위해 빅데이터를 활용하고 있다.

<표 4> 주요국의 빅데이터 활용 정책 및 사회문제 해결 사례

주요국	분야	정책활용
미국	탈세 방지 프로그램	탈세 행위를 미리 예측하고 방지하여 국가 재정 확충 및 사회문제 해결
	국립보건원 의료 체제 개혁	유전자 데이터 공유를 통한 질병치료체계 구축, Pillbox 프로그램 기반 의료 개혁 실시
	테러 대응 프로그램 ADVICE	9.11 테러 이후 급증한 테러 위협 예측 및 대응을 위한 통합적 대 테러 프로그램 ADVICE 구축
	범죄 예측 프로그램	범죄 예방 및 범죄율 감소를 위해 범죄 발생 빈도와 위치를 분석하고 범죄자들을 지속 관찰하여 범죄 징후를 예측
영국	비만 대책 프로그램	비만을 사회 문제로 규정하고 비만 문제에 대한 효과적인 대응책 마련
	잠재적 위험 관리 프로젝트	영국을 위협할 수 있는 잠재적 위험, 자연 재해 등에 대한 대책 마련
	전염병 대응 프로그램	글로벌 차원의 전염병 발생 현황과 동물 전염병 동향 파악을 통한 전염병 예측 및 대응 방안 마련
싱가포르	해안 안전 시뮬레이션	해상 테러 및 해안 침투 등을 방지하여 해안 영역 안전 확보
	조류 인플루엔자 시뮬레이션	조류 인플루엔자 미래 시나리오 구축 및 대응 방안 마련
일본	지능형 교통시스템	교통 정보 분석을 통해 최적의 교통 상황 및 경로 안내
	재난 대응 프로그램	지진, 화재, 방사능 등 재난 발생을 예측하고 대응 능력을 강화

자료 : 성지은·박기량(2014)

한국은 안전행정부, 미래창조과학부, 국토교통부 등을 중심으로 시범사업을 추진 중에 있다. 한국정보화진흥원(2014)은 <표 5>와 같이 국내 빅데이터 국내사례를 미래창조과학부 선도사업 6종과 공공사업 11종, 민간사례 13종을 소개하고 있다.

지방정부의 대표적인 빅데이터 활용사례는 서울시 심야버스 노선수립과 관련된 사례이다. 서울시는 합리적인 버스노선을 설정하기 위하여 빅데이터 분석을 이용하였다. 분석 자료는 서울시의 시내버스 및 정류소 현황과 KT의 CDR(Call Detail Record)<sup>1)</sup>

를 이용하여 심야시간대 통화가 가장 많이 발생하는 지역을 기초로 유동인구를 파악하고, 이를 서울시 시내버스 현황과 정류소 현황 정보 등과 비교 분석하였다. 이러한 빅데이터 분석 기법을 바탕으로 9개로 확대하였다.

<표 5> 2013년 국내 빅데이터 사례

부서	사업명	활용데이터	배경	성과
미래 창조 과학부	빅데이터 분석을 통한 심야버스 노선정책지원	· (주) KT, 서울시	합리적인 심야버스 노선 수립	데이터 기반 신뢰성 있는 행정 구현 및 시민편익증진
	국민건강 주의 예보 서비스 구축	· 국민건강보험공단, 다음소프트	전염성 질환의 위험도 사전예보 체계구축	지역별 연령별 질병에 대한 위험도 예보제공
	의약품 안전성 조기경보시스템	한국 의약품안전관리원, 와이즈넷, 아주대학교병원	의약품 부작용 및 오남용사례 조기 검출	우리나라에 적합한 약물부작용 조기경보시스템 개발을 통한 기술의 유용성 검증
	심실부정맥 예측, 입원병상 최적화 등 보건의료서비스	서울아산병원 외	환자모니터 정보의 활용성 증대 인플루엔자조기경부 마약류 정보	중환자실 환자 생존률 증가 의료장비 로그데이터 분석 기반기술 확보
	소상공인 창업지원 서비스를 위한 점포평가 서비스	BC카드, 한국감정원, 소상공인지원센터	민간·공공 데이터를 연계 활용하여 소상공인 창업의사결정도움 인프로 제공	과학적 상권분석을 통한 자영업자의 실패확률 감소
	스마트 뉴스제공 서비스	한국언론진흥재단	스마트 뉴스서비스 제공	디지털뉴스의 새로운 모형 제안
공공 사례	빅데이터 기반 영주사과작황정보서비스	경상북도, 미래창조과학부	빅데이터 모형을 개발해 검증해 봄으로써 우리나라 빅데이터 연구와 산업의 발전에 기여	병충해 징후 발견 생산 추정 가격동향
	기상 빅데이터 분석을 통한 효율적인 위험기상 예측	기상청	기상 예측의 정확성을 높이기 위한 모델구성	단일속성모델 분석을 통한 조합속성 모델을 생성하고, 생성된 모델 중 최적의 모델을 선택
	안전한 금융서비스를 위한 여신상시감시시스템	금융감독원	불법 및 부실 혐의 여신을 사전에 탐지하고 사전에 방지	
	특수질병 및 질환 예방 및 관리	국립보건연구원	인체 자원 정보를 공유	생명 연구 자원의 글로벌 경쟁력 기반 확보
	통합형 재난안전정보 네트워크, 스마트 빅 모드	재난안전연구원	재난상황을 실시간으로 파악하고 능동적으로 분석하여 국민들의 신속한 안전 확보에 도움을 주기 위한 체계적인 서비스에 대한 필요가 꾸준히 제기	다양한 유형의 재난에 신속대응

자료 : 한국정보화진흥원(2014)

- 1) 고객들이 유선전화나 휴대폰으로 통화할 때마다 통화위치와 대상, 시간 등의 로그 데이터를 의미.

기타 부산시 해운대구도 빅데이터를 이용하여 구정활동에 개선의 노력을 하고 있다. 해운대구는 2013년 빅데이터 분석팀을 신설하여 부산대 “빅데이터처리플랫폼연구센터(ITRC)”와 협약을 통해 해운대 관광객의 트위터, 블로그, 페이스북 등 SNS의 어휘 분석을 통해 관광정책에 수립에 활용하고 있다. 기타 일자리, 민원 및 도시정책 수립 등의 활용도를 점차 높이고 있다.

## IV. 빅데이터 활용 쟁점 및 과제

### 1. 개인정보 보호의 문제

정보통신기술이 발달하면서 생활의 편리성을 누리려 한다면, 개인정보 유출 등과 같은 부작용도 발생하고 있다. 최근 스마트폰과 사물통신 및 SNS와 같은 새로운 미디어 환경의 발전되면서 빅데이터에 대한 이슈가 확대되고 있다. 밝은 미래전망과 함께 개인의 정보보호문제가 지속적으로 제기되고 있다. 모바일 어플리케이션이나 소셜 네트워크 서비스 등의 정보를 이용하면서 생성되는 위치 및 이동정보, 검색패턴에 대한 로그 등의 그림자 데이터(digital shadow)는 개인 사생활을 침해할 가능성을 높인다(이환수 외, 2013: 126).

외견상으로는 무의미하지만, 이들 데이터가 결합될 경우 자신이 누구인지에 대하여 파악할 수 있을 것이다. 이는 추가적으로 개인의 사생활 침해와 더불어 감시의 문제가 발생된다. 빅데이터 시대의 도래로 인해 기업이 핵심노조원이나 정부 주요인사의 행동과 활동을 감시할 수 있고, 국가가 테러나 범죄예방을 명목으로 불특정 다수의 개인들을 감시하고나, 개인이 경쟁 상대나 관심대상에 대한 감시가 활동이 얼마든지 가능하다(윤상오, 2013: 108-109). 특히 이러한 정보들이 해킹에 의해 제3자에 의해 악용될 경우 빅데이터 시대에 개인정보 과잉에 대한 문제점이 발생할 수 있다.

이렇듯 명백한 범죄행위 이외에도 다양한 형태의 문제점이 발생할 수 있다. 기업의 관점에서는 빅데이터를 이용하여 개인의 성향을 파악하고, 각 개인에게 맞는 맞춤형 마케팅 전략을 세우려고 할 것이다. 이런 상황에서 빅데이터의 과잉 활용이 문제될 수 있다. 예를 들어 특정 어휘를 사용하거나, 또는 특정 음원을 선호할 사람의 경우 신용카드 연체률이 높고, 개인파산자가 될 가능성이 높다는 분석결과를 바탕으로 특정인의 대출을 제한하거나 높은 이자율을 적용한다면, 서비스 개선이 아닌 개인에 대한 차별로 이어질 가능성이 있다. 이러한 관점에 대해서 사회 내에서 충분히 논의되기 위한 여건이 아직 조성되어 있지 않다. 따라서 빅데이터로 분석 및 활용에 대한 사회적 합의 및 개인사생활의 보호와 범죄 등에 악용되지 않도록 법과제도 마련에 대

한 다양한 논의가 요구된다.

## 2. 빅데이터 수집 및 관리 방법론

앞서 빅데이터의 정의에서 언급하였듯이 빅데이터는 단순한 양이 많은 것을 의미하는 개념을 넘어 규모(volume), 다양성(variety), 속도(velocity), 복잡성(complexity)의 4가지 특징을 가지고 있는 정형 또는 비정형화된 것으로 이것으로 처리 분석하여 그 결과 새로운 가치를 창출 해 낼 수 있는 기술로 초고속 데이터의 수집, 처리, 분석, 활용 포함된 기술로 정의될 수 있다.

전통적 연구방법은 수집할 데이터의 대상을 먼저 설정한 이후 표본조사 등을 통한 데이터의 수집 및 분석이 이뤄지는 과정을 거친다. 하지만 빅데이터 시대에서는 우선 수집된 방대한 빅데이터 속에서 의미 있는 데이터를 마이닝하여 의미를 찾게 되는데 이러한 과정에 다양한 형태의 오류가 발생할 가능성이 있다.

첫째, 빅데이터로부터 추출한 데이터가 과연 대표성을 가지고 있느냐에 대한 문제가 된다. 전통적인 표본추출 방법에서는 표본의 대표성 확보를 위한 다양한 노력이 강구되지만, 빅데이터의 마이닝 과정에서 이들 데이터가 특정 집단 또는 특정 의식을 갖은 계층에 편향될 가능성이 발생한다.

둘째, 데이터를 이용하여 분석으로 활용하기 위해서는 표준화 작업이 필요한데, 이러한 표준화 과정에서 표준화 또는 디지털화가 불가능한 데이터는 모두 하찮고 쓸모 없어진다. 빅데이터 분석방법에 따른 차이는 있지만, 빅데이터 분석을 위해 크롤링(crawling)<sup>2)</sup>을 통해 수집된 데이터를 정제하는 과정을 거치게 되는데, 이 과정에서 중요한 데이터의 누수 현상이 발생하게 한다. 이와 더불어 표준화 과정에 분석자의 개인적 주관이 개입할 수 있는 가능성도 함께 높아진다.

셋째, 문서, 그림, 영상처럼 정형화 되지 못한 비정형화 데이터에 대한 분석의 경우 형식이 정해져 있지 않기 때문에 분석을 위해서는 정형화 하는 과정이 필요하다. 이영주 외(2013)는 대전지역 국공립어린이집 부족문제를 해결하기 위해 공간빅데이터를 통한 분석모형에 관한 연구에서 공간빅데이터 사회분야의 적용 방법을 살펴보면 SNS와 뉴스, 카페, 트윗을 통해 키워드를 추출하고, 이들 간의 관계를 키워드 트리로 정의한 이후 감성적 표현까지 구분하여 분석에 활용하였다. 하지만, 단순한 텍스트가 갖는 의미만을 분석하고 있다. 과거 행태주의가 비판받았던 것처럼 “개인의 활동 및 행동결과에 대한 해석은 상황적 맥락 속에서 이해되어야 한다”는 점에서 한계가 있다. 즉, 상황적 맥락을 파악하지 못한 기계적 의미해석은 오히려 왜곡된 결과를 초래하고,

2) 무수히 많은 컴퓨터에 분산 저장되어 있는 문서를 수집하여 검색 대상의 색인으로 포함시키는 기술(IT용어사전, 한국정보통신기술협회).

자칫 정책결정에 노이즈로 작용할 기능을 배제할 수 없다. 따라서 빅데이터 분석을 위한 다양한 기법과 이들 자료의 신뢰성을 높이기 위한 방안이 동시에 강구되어야 할 것이다.

### 3. 정보의 격차와 정보의 역설

정보격차와 디지털 격차와 같이 빅데이터 격차도 심각한 사회문제가 될 수 있다. Manovich(2011)는 빅데이터 시대에 “데이터 계급”(data-classes)을 세 집단으로 분류하였다. 데이터를 만들어내는 계급, 데이터를 수집할 수 있는 계급, 데이터를 분석할 전문성을 가진 계급이 그것이다. 특히 세 번째 계급의 숫자는 극히 작을 뿐만 아니라 계급 간의 격차는 출신과 데이터에 대한 접근성과 컴퓨터 과학과 고급통계에 대한 지식이 결정된다고 보고 있다(이재현, 2013: 134). 이렇듯 빅데이터 시대의 경우 분석전문가의 역할이 크다.

하지만 국내기업은 아직 빅데이터 처리에 대한 기술적 이해나 경험이 부족한 편이다. 빅데이터 처리 및 분석에 필요성에 대한 시각차이가 문제일 수 있지만, 향후 빅데이터 처리 및 활용에 필요한 전문기술의 양성이 요구된다. 공공기관 역시 대부분의 예산을 빅데이터의 활용 사례발굴에 초점이 맞춰져 있을 뿐 전문 인력 양성에 대한 노력은 많지 않다. 지방정부 또한 빅데이터 분석을 위한 다양한 사례를 내놓고 있지만, 빅데이터를 활용 및 분석은 대기업이거나 대부분 수도권에 위치한 IT기업에 의존하고 있어 빅데이터 사업에 대한 본래의 취지를 살리지 못하고 있다. 이처럼 지역 간의 IT 산업의 격차는 빅데이터 전문 인력의 확보 측면에서도 수도권과 지방간의 격차가 발생 할 수 있다. 따라서 지방정부의 경우 지역 간 빅데이터 격차를 줄이기 위한 지역 대학 및 연구기관과 연계된 빅데이터 교육프로그램의 개발과 전문가 양성을 위한 노력이 병행되어야 한다.

빅데이터 시대에 또 다른 문제점은 정보의 역설 현상이 발생할 수 있는 것이다. 우리는 일반적으로 자기의 주장을 뒷받침하기 위해 통계자료를 활용하게 되는데, 데이터가 많고, 통계적 방법이 늘어날수록 데이터의 결과가 진실을 반영한다고 사람들 믿게 된다. 한편, 우리는 너무 많은 것을 알고 있음에도 사실 그것들이 무엇인지 모르는 이른바 “정보의 역설”에 주의해야 한다(오라일리 미디어, 2013: 122). 데이터 뒤에 숨어 있는 진실에 대하여 관심을 가져야 하지 수학적 방법에 현혹되어서는 안 된다. 동일한 데이터라도 분석 모델과 알고리즘이 어떻게 설계하느냐에 따라 결과가 전혀 달라질 수 있다(윤상오, 2013: 107). 따라서 빅데이터를 이용한 정책결정 과정에서 몇 가지 검증 절차를 확보하는 것이 바람직하다. 잘못된 분석결과는 정책결정의 오류를 범할 가능성이 높기 때문이다.

## V. 결 론

빅데이터는 우리의 미래의 삶에 크고 작은 영향을 줄 것이다. IT 기술의 개발에 따른 지능형과학의 발전과 생활의 편리성과 새로운 경제발전의 혁신적인 요소가 될 수 있을 것이다. 정부는 이러한 과학기술의 청사진에 기초하여 빅데이터 관련 분야에 대하여 육성·발전시키려고 있다. 공공의 데이터를 개방하고, 빅데이터의 관련 성공사례를 발굴하는 등 사회문제 해결을 위한 다양한 형태의 시도를 하고 있다.

빅데이터 기술의 정부 정책과정에 도입은 과학적 정책결정 및 집행을 통해 행정적 행정과 주민편의를 높일 수 있을 것이다. 하지만, 분석결과에 사로잡혀 오히려 주민과의 대면접촉이 감소되어 행정의 대응성을 감소시키는 결과를 낳을 수 있다. 신뢰성이 낮은 빅데이터 수집 및 분석은 정책결정의 오류 또는 노이즈로 작용할 가능성이 높다. 이러한 오류를 감소시키고 빅데이터 기술의 행정과정에 성공적으로 도입 활용되기 위해서는 무엇보다 빅데이터 활용과 관련된 전문 인력 양성이 절대적으로 요구되고 있다. 향후에는 얼마나 많은 빅데이터 전문 인력을 확보하고 있느냐가 핵심적 경쟁력이 될 것이다.

하지만 지방정부의 경우 IT 인프라가 매우 부족하다. 대부분 IT관련 산업이 수도권에 집중되어 있어 지방정부는 빅데이터 관련 시스템 구축과 분석 및 활용측면에도 수도권 지역에 의존할 수밖에 없는 구조이다. 이러한 격차는 빅데이터 시대에 지역불균형을 초래하고 지방의 중앙의 의존도를 높이는 주요한 요소가 될 것이다. 따라서 지방정부는 지역대학과의 연계를 통한 빅데이터 전문 인력 양성 및 IT 인프라가 갖춰질 수 있는 생태계조성을 위한 지원책 마련에 초점을 두어야 할 것이다.

## <참고문헌>

- 김대중. (2014). 정부3.0을 위한 공간빅데이터 구축 및 활용방안. 「월간국토」, 387: 42-51.
- 김동완. (2013). 빅데이터의 분야별 활용사례. 「경영논총」, 34: 39-52.
- 김민수. (2014). 국내외 공간빅데이터 정책 및 기술동향. 「월간국토」, 389: 30-39.
- 김상락·강만모. (2014). 빅데이터 분석기술의 오늘과 미래. 「정보과학회지」, 32(1): 8-17.
- 김성태. (2013). 「새로운 미래를 여는 빅데이터 시대」. 한국정보화진흥원
- 김우영. (2014). 공공데이터 개방: 정부3.0 로드맵과 빅데이터. 「CERIK Journal」, 2014.2: 23-26.

- 김현영 외. (2014). 빅데이터에 따른 지적정보의 효율화 방안 연구. 「한국지적정보학회지」, 16(1): 29-48.
- 김희수. (2014). 공간빅데이터 체계구축·활용 정책방향. 「월간국토」, 389: 6-11.
- 성지은·박기량. (2014). 빅데이터를 활용한 정책 사례 분석과 시사점. 「과학기술정책」, 24(2): 94-106.
- 신신애 외. (2014). 「창조경제 실현을 위한 2013 빅데이터 국내사례집」. 한국정보화진흥원.
- 안창원·향승규. (2012). 빅데이터 기술과 주요이슈. 「정보과학회지」, 30(6): 10-17.
- 오라일리미디어. (2013). 「오라일리 심층보고서 빅데이터 어떻게 활용할 것인가?」. 제이펍.
- 윤미영. (2013). 주요국의 빅데이터 추진전략 분석 및 시사점. 「과학기술정책」, 192: 31-43.
- 윤상오. (2013). 빅데이터의 위험유형분류에 관한연구. 「한국지역정보화학회지」, 16(2): 93-122.
- 이영주 외. (2014). 복지·안전국토 실현과 공간빅데이터 활용. 「월간국토」, 389: 12-21.
- 이영주 외. (2013). 「공간빅데이터기반의 복지·공간 융합모델 시범연구: 영·유아 복지 공간 진단을 통한 복지정책 지원방안」. 국토교통부.
- 이완수 외. (2013). 빅데이터 시대의 개인정보 과잉이 사용자 저항에 미치는 영향. 「지능정보연구」, 19(1): 125-139.
- 이재현. (2013). 빅데이터와 사회과학: 인식론적·방법론적 문제들. 「커뮤니케이션이론」, 9(3): 127-165.
- 이정미. (2013). 빅데이터의 이해와 도서관 정보시스템의 활용. 「한국비블리아학회지」, 24(4): 53-73.
- 최진명. (2012). 「빅데이터시대의 데이터 활용과 전략」. 한국지역정보개발원.
- 최진원·김이연. (2012). 빅데이터시대에 효과적인 시각커뮤니케이션을 위한 인포그래픽 연구: 정부부처 및 공공기관을 중심으로. 「한국과학예술포럼」, 11(1): 165-175.
- 최창우 외. (2012). 「지리정보와 공공정보의 결합을 통한 지역 공공서비스 활성화 방안 연구: 공공정보를 이용한 플랫폼 구현 및 주민참여 지도 구축방안을 중심으로」. 한국지역정보개발원.
- 최혜민. (2014). 빅데이터 시대와 현행 개인정보보호법제의 부정합 문제 및 그 해결 방안에 대한 연구. 「IT와 법연구」, 8: 357-382.
- 한국정보화진흥원. (2012a). 「빅데이터로 진화하는 세상 : Big Data 글로벌 선진사례」. 한국정보화진흥원.
- 한국정보화진흥원. (2012b). 「선진국의 데이터기반 국가미래전략추진 현황과 시사점」. 한국정보화진흥원.

- 한국정보화진흥원. (2013). 「새로운 미래를 여는 빅데이터 시대」. 한국정보화진흥원
- 한국정보화진흥원. (2014). 「창조경제 실현을 위한 2013 빅데이터 국내 사례집」. 한국정보화진흥원
- 허성욱. (2014). 한국에서 빅데이터를 둘러싼 법적 쟁점과 제도적 과제. 「정보통신 정책학회 학술발표논문집」, 2014(3): 173-193.
- IT용어사전, 한국정보통신기술협회. www.naver.com 검색결과
- 국토교통부(2014). “정부, 올해 385개 공간정보사업 추진, 2,946억원 규모”국토교통부 보도자료. 2014.1.23.
- 전자신문(2014). “공공빅데이터 컨트롤타원 기능 불명확 중복투자와 효과 미흡 우려”(http://www.etnews.com). 2014.4.23.일자 기사.
- 통계청(2014). “빅데이터 활용 본격화, 복지·치안·물가등해결사”. 통계청 보도자료. 2014.1.07
- 미디어잇(2014). 심야버스부터 상품배치까지 생활 속 빅데이터 사례 5선”(http://www.it.co.kr) .2014.05.07.일자 기사.
- Chadwick, A. (2003). Bringing E-Democracy Back In : Why it Matters for Future Research on E-Governance. *Social Science Computer Review*, 21(4): 443-455.
- Gantz, J. and Reinsel D. (2011). Extracting value from chaos. *IDC I-View*. June 2011. http://www.emc.com
- Gartner. (2011). Pattern-Based Strategy : Getting Value from Big Data. *Gartner Special Report(Jun 2011)*.
- Manovich, L. (2011). Trending: The Promises and the Challenges of Big Social Data. In M. K. Gold (ed.), *Debates in the Digital Humanities*. Minneapolis, MN: The University of Minnesota Press: 460-475.
- Manyika et al. (2011). *Big Data: The Next Frontier for Innovation, Competition, and Productivity*. Technical Report, McKinsey Global Institute.
- Newman, D. (2011). *How to Plan, Participate and Prosper in the Data Economy*. Gartner, Stamford, CT.
- Sang-Hyun Ju · Sang-Yup Lee. (2014). Analysis on Value Creative Economic System of Japan and Policy Implications on Korea. *Korean Comparative Government*, 18(2): 289-266.

접수일(2000년 11월 14일)

수정일자(2011년 11월 30일)

게재확정일(2011년 12월 09일)