

세계경제포럼(WEF)을 통해 본 빅데이터 논의 동향과 함의

유 지 연*

1. 개요

빅데이터(Big Data)가 오늘 IT분야의 중요한 키워드가 되고 있다. 빅데이터는 새로운 개념은 아니다. 1990년대 컴퓨터의 발달과 함께 언급되던 정보 폭발(information explosion), 정보 홍수(information overload) 등과 연장선상에 있다.¹⁾ 다만, 최근의 개념은 진화한 개념으로, 단순히 데이터의 양이 많은 것만을 의미하는 것은 아니며, 기존 데이터에 비해 너무 방대해 일반적으로 사용하는 방법이나 도구로 수집·저장·검색·분석·시각화 등을 하기 어려운 정형 또는 비정형 데이터세트를 의미한다.²⁾ 그리고 빅데이터의 크기는 움직이는 표적으로, 현재는 수십 테라바이트(Terabyte, TB)에서 수 페타바이트(Petabyte, PT)까지 다양하다. 즉, 빅데이터는 데이터 양(volume), 데이터 속도(velocity) 그리고 데이터 다양성(variety) 등 3가지 차원의 데이터 성장 변화를 의미한다.³⁾

* 정보통신정책연구원 미래융합연구실 부연구위원, (02)570-4238, jyy@kisdi.re.kr

1) 《ZDNet Japan》(2011. 8. 30): 《Economist》(2010. 2. 25)

2) Mckinsey Global Institute(2011)

3) Douglas(2001); Beyer(2011)

이러한 빅데이터가 이슈화되고 성장하게 된 데는 크게 3가지 요인이 있다. 첫째 요인은 스마트폰을 비롯하여 모바일기기 보급의 활성화이다. 모바일기기에 탑재된 센서, 원격감지기술, 소프트웨어 로그, 카메라, 마이크, RFID 리더 등을 통해 비정형 데이터를 수집할 수 있게 된 것이 데이터를 증가시키는 요인이 되고 있다.⁴⁾ 둘째 요인은 클라우드 서비스이다. 클라우드 서비스를 통해 개인과 조직의 데이터가 한곳으로 축적되었고 데이터 분석과 활용에 대한 요구가 증가하고 있다.⁵⁾ 셋째 요인은 소셜미디어 활용의 일상화이다. 소셜미디어를 통해 커뮤니케이션과 정보의 유통 구조가 새롭게 재편되고 이들을 통해 상호작용되는 데이터들이 증가하면서 이들 데이터를 비즈니스 측면에서 활용하려는 관심 또한 증대하고 있다.⁶⁾

빅데이터는 공공·행정, 제조업, 소매업, 의료·건강, 개인위치정보 등 사회 각 분야로 확산되면서 사회 전반의 생산성 향상과 가치창출에 기여할 전망이다. EU의 경우 15~20%의 공공관리 비용 감소와 2~4천 달러의 가치 창출, 향후 10년간 0.5% 생산성 증가 효과 등을 기대한다. 제조업 적용시에 상품개발 및 조립비용을 50% 이상 절감 가능하며 운전자본도 7% 이상 절감할 것으로 본다. 또한 소매업은 이윤 60% 이상 증가와 년 0.5~1% 생산성 향상을, 의료·건강은 매년 3,300억 가치와 년 0.7%의 생산성 증가를, 개인 위치 정보 서비스는 1,000억 이상의 공급자 매출과 7,000억의 사용자 혜택을 줄 것으로 전망한다.⁷⁾

빅데이터에 대한 관심은 데이터를 어떻게 유통하고 활용하는가가 디지털시대의 경쟁력이 되어감에 따라 향후 얼마간은 중요한 이슈가 될 전망이다. 2011년에 전 세계에서 새롭게 생성되거나 복제된 정보량은 1.8제타바이트(Zettabyte, ZB)(1.8조 기가바이트(Gigabyte, GB))를 넘어섰으며,⁸⁾ 전 세계의 디지털 정보량은 약 2년마다 2배로

4) 《GigaOM》(2008. 11. 9); Segaran and Hammerbacher(2009), p.257.

5) IDC(2011), p.4.

6) 채승병(2011), p.4.

7) Mckinsey Global Institute(2011)

8) 전 세계 모든 사람이 매일 2억 1,500만 화소의 고해상도 MRI(자기공명영상) 촬영하는 데이터, 한 사람이 쉬지 않고 4,700만년 동안 시청할 수 있는 2,000억개 이상의 고화질(HD) 영화(상영시간 2시간

증가하는 추세다.⁹⁾

우리나라도 빅데이터에 대한 관심이 높아지고 있으며 2012년을 빅데이터 시대 원년으로 보고 있다.¹⁰⁾ 이에 본고에서는 지난 1월 25일부터 29일까지 열린 세계경제포럼(WEF)에서 발표된 “빅데이터 영향: 국제 개발을 위한 새로운 가능성(Big Data, Big Impact: New Possibilities for International Development)” 보고서¹¹⁾를 통해 빅데이터 관련 논의를 정리하고 정책적 차원에서 우선적으로 준비해야 할 점에 대해서 살펴보고자 한다.

2. 빅데이터 관련 논의 동향

(1) 데이터 생태계(Data Ecosystem)

경제개발 도구로서 빅데이터는 새로운 데이터 생태계를 구성하며, 생태계의 여러 주체들에 의해서 데이터를 생산·축적·활용한다. [그림 1]은 새로운 데이터 생태계에서 행위 주체별 데이터 유형과 이익, 그리고 요구사항 등을 정리한 것이다.

행위 주체별 빅데이터의 활용성은 높지만, 데이터 자원을 가장 필요로 하는 공영역에서는 그 활용성이 더욱 기대된다. 공공영역에서는 이전부터 인구, 건강, 소비 및 세금 등에 대한 거대 규모의 데이터세트들이 유지·관리되었다. 인터넷과 모바일 혁명은 이러한 데이터세트에 새로운 데이터원이 되고 있다. 일례로 미국 재난관리 오픈소스 플랫폼 우샤히디(Ushahidi)는 자연재해 및 선거 관련 정보를 실시간으로 취합하여 구호 등 관련 활동을 수행한다. 즉 스마트폰에 접근할 수 있는 세계 다수는 더 민주적이고 투명하게 정보 수집 과정에 기여하고 있다. 또한 특정지역의 휴대전화 시간의 감소는 그 지역 인구의 소득 손실을 의미하기도 한다. 특히 공공 의료는 모바일 및 인터넷

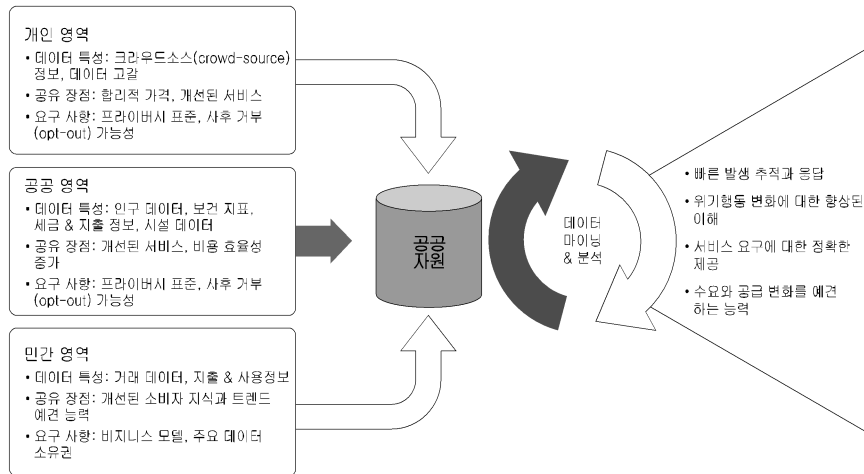
기준), 32GB(기가바이트) 용량의 아이패드 575억개 등의 용량과 같음 《헤럴드경제》(2011. 6. 30)

9) IDC(2011), p.1.

10) 《ETnews》(2012. 1. 26)

11) WEF(2012)

[그림 1] 데이터 생태계의 행위주체와 공공자원



자료: WEF(2012), p.4.

데이터가 공공이익으로 이어질 수 있는 가장 효과적인 분야 중 하나이다. 일례로 샌프란시스코 기반의 글로벌 바이러스성 예측 이니셔티브(Global Viral Forecasting Initiative, GVFI)는 바이러스가 세계적인 전염병이 되기 전에 발생한 지역의 종합적인 위치와 소스 등에 대해서 인터넷을 통해 얻은 정보를 향상된 데이터 분석에 사용한다. 이러한 분석은 전통 기법과 지표에 의존하는 국제기구인 세계보건기구(WHO)의 발생 전망 보다 일주일 앞서 예측할 수 있도록 한다.

(2) 공공자원 활용을 위한 고려사항

데이터 생태계의 공공자원을 활용하기 위해서는 몇가지 생태계 요건들이 마련되어 있어야 한다. 즉 협력적인 데이터 공유를 위해서는 먼저 개인에 의해 생산되는 데이터에 대한 합법적인 활용 방안을 모색해야 한다.

1) 프라이버시와 보안

데이터 생태계의 행위주체들이 데이터를 사용하려면 데이터를 생산한 개인에 대한

신뢰, 표현의 자유 그리고 기밀성 위반 등에 대한 우려에 직면한다. 데이터를 공유하기에 앞서 개인을 위한 메카니즘으로 개인의 프라이버시와 보안이 보장될 수 있어야 한다.

2) 데이터 개인화

모바일을 통한 개인의 일상 데이터 등 데이터 특성이 중요한 가운데, 개인이 여러 개의 모바일 SIM 카드를 가지면 동일한 개인에 대한 SIM 데이터 분석이 불가능하다. 더욱이 이 데이터는 인구 구성별 습성에 대한 이야기를 구성하는 인구통계학적 지표로도 유용함으로써, 모바일 데이터는 가능한 한 개별화 되도록 할 필요가 있다.

3) 데이터 공유에 대한 보상

개인정보 프라이버시 및 보안 문제가 두려워 개인정보 공유를 하지 않으려 한다. 또한 민간영역의 기업들은 데이터 독점에 따른 활용 이익이 크기 때문에 공유하지 않는다. 더욱이 정부는 공공계약에 의해 수집된 데이터를 공유하거나 민간영역에서 이용될 수 있는 정부 데이터를 공유하도록 강제할 수 없다. 이에 데이터 공유에 따른 위험 보다 이익과 보상을 얻을 수 있도록 해야 한다.

4) 인적 자본 확보

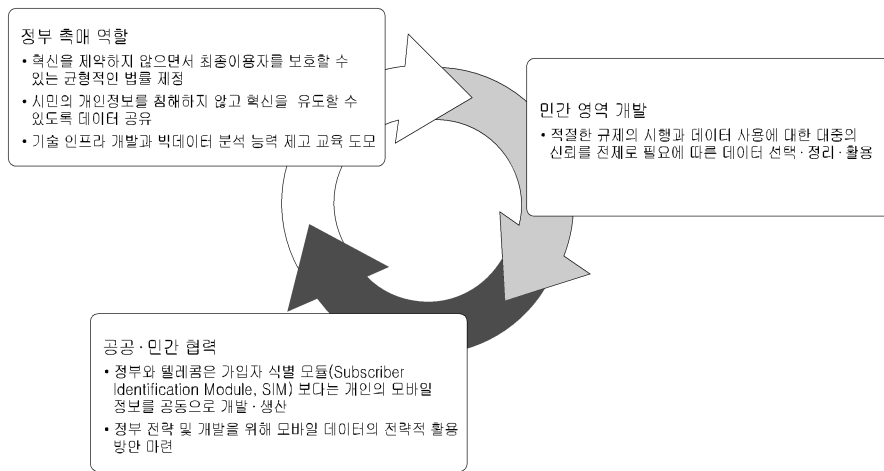
정확하고 실용적인 데이터 마이닝과 분석은 상당한 기술적인 숙련을 요구한다. 이로 인하여 데이터 과학자들은 인력의 공급 부족과 고임금의 문제를 안고 있다. 심지어는 많은 대기업들이 새로운 마이닝 기법을 개발해야 하는 전문지식을 가지고 있지 못하다. 공익 발전을 위한 전문인력의 공헌에 대한 보상과 인력 육성에 대한 장기적인 노력 등이 필요하다.

(3) 데이터 촉매로서의 정부

모바일 데이터를 공유하기 위한 환경을 만들어 상호이익을 실현하려면 모든 생태계 주체들이 모두 적극적이고 개방적으로 참여해야 한다. [그림 2]는 데이터 생태계 행

위 주체들의 역할을 정리한 것이다. 정부는 개인을 보호하고 데이터 공개원칙 및 법적 프레임워크 등에 대해서 설정한다. 민간 영역은 대중이 혜택받을 수 있도록 데이터 공유를 위한 메카니즘을 바르게 작동시켜야 한다. 그리고 개발기구는 데이터 사회 공헌에 대한 정부 지원과 공익 및 비즈니스 가치가 원활하게 공유될 수 있도록 환경을 조성한다.

[그림 2] 데이터 생태계 행위주체별 역할



자료: WEF(2012), p7.

3. 결 어

세계경제포럼의 빅데이터 보고서는 빅데이터를 전체 데이터 생태계 구조로 고찰하고 공공자원으로서 빅데이터의 중요성과 활용성을 강조하고 있다. 그리고 공공자원으로서 빅데이터 활성화·안정화를 위한 정부의 촉매 역할을 기대하고 있다. 구체적으로 빅데이터가 공공자원으로서 활용될 수 있도록 정부 영역, 민간영역, 개인영역의 데이터 수집 및 분석을 지원하고, 그에 앞서 혁신을 제약하지 않으면서 개인의 프라이버시 및 보안을 보호할 수 있는 균형적인 제도 개선의 추진을 제안했다.

우리나라도 빅데이터에 대한 관심이 고조되고 있다. 그러나 빅데이터에 대한 논의는 아직 민간영역 중심이며, 공공영역에서는 공공영역 내 데이터 연계를 중심으로 논의되고 있다. 빅데이터는 데이터를 수집하는 주체에 따라서 정부영역의 데이터, 민간영역의 데이터, 개인영역의 데이터이기도 하지만, 어디에 어떻게 활용하느냐에 따라서 새로운 가치를 창출하며, 공공의 이익에 기여할 수도 있다. 이에 우리나라 전체의 데이터 생태계에 대한 고찰과 공공자원으로서 빅데이터 개발이 필요한 시점이다. 공공자원으로서 공통의 빅데이터를 개발하고 금융·전략·의료·교육 등 새로운 활용 영역을 발굴함으로써 국가자산의 확충과 발전을 도모할 수 있을 것이다. 더불어서 빅데이터 활용 주체로서 이용자 역할을 제고하고 공공 빅데이터 활용을 활성화할 수 있는 방안을 모색해야 한다.

참고문헌

- Beyer, Mark (2011), “Gartner Says Solving ‘Big Data’ Challenge Involves More Than Just Managing Volumes of Data”. Gartner.
- Cisco (2012), “Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2011~2016”.
- Douglas, Laney (2001), “3D Data Management: Controlling Data Volume, Velocity and Variety”. Gartner.
- IDC (2011), “Extracting Value from Chaos”. 《IDC IVIEW》.
- Mckinsey Global Institute (2011), “Big Data: The next frontier for innovation, competition, and productivity”.
- Segaran, Toby and Hammerbacher, Jeff (2009), 『Beautiful Data: The Stories Behind Elegant Data Solutions』. 1st Edition. O'Reilly Media.
- WEF (2012), “Big Data, Big Impact: New Possibilities for International Development”.
- White, Tom (2009). 『Hadoop: The Definitive Guide』. 1st Edition. O'Reilly

Media.

채승병 (2011), “정보홍수 속에서 金脈 찾기: ‘빅 데이터(Big Data)’ 분석과 활용”,

《SERI 경영노트》, 2011. 2. 10.

《헤럴드경제》 (2011. 6. 30), “올 생성 디지털 정보량...서울시 면적의 2.1배”.

《Economist》 (2010. 2. 25), “The data deluge”.

《ETnews》 (2012. 1. 26), “[새해 새설계-공공기관이 함께 된다]〈11〉한응수 한국
데이터베이스진흥원장”.

《GigaOM》 (2008. 11. 9), “Parallel Programming in the Age of Big Data”.

《ZDNet Japan》 (2011. 8. 30), “「ビッグデータ」がもたらす機会と課題--特集
「ビッグデータとは何か」”.