



# 이슈와 논점



이슈와 논점 | 제1482호 | 2018년 7월 4일 | 발행처 국회입법조사처 | 발행인 이내영 | www.nars.go.kr

## 4차 산업혁명 시대의 빅데이터 정책 과제

정 준 화 \*

### 1. 들어가며

우리 사회에서 빅데이터(big data)가 본격적으로 논의되기 시작한 지 어느덧 7년이 흘렀다. 2011년부터 빅데이터는 인터넷 포털사이트의 주요 검색어로 등장했고 국가정보화전략위원회의 「빅데이터를 활용한 스마트 정부 구현」과 같은 정부 정책에 공식 반영되기 시작했다.

빅데이터 도입 초기에는 그 개념과 방법이 생소하여 실체가 모호하고 실용성이 의심된다는 평가가 있었던 것도 사실이다. 그러나 최근 4차 산업혁명 시대가 도래하면서 빅데이터의 실체성과 중요성에 대해 반론을 제기하는 사람은 그리 많지 않다. 모바일·사물인터넷 같은 4차 산업혁명 핵심 기술은 디지털 데이터의 수집·생산 규모를 폭발적으로 증가시켜 빅데이터의 실현 가능성을 높였고, 동시에 4차 산업혁명의 성패를 좌우할 인공지능은 빅데이터 기반의 기계학습 덕분에 더욱 정교해지고 있기 때문이다.

이러한 빅데이터의 가치와 가능성을 충분히 발현하기 위해서는 빅데이터가 건설한 산업으로 성장할 수 있어야 한다. 그러나 국내 빅데이터 산업은 아직까지도 태동 단계를 넘어서지 못하고 있다. 산업 내부의 자율적인 성장을 기다

려 볼 수도 있겠지만 빅데이터의 중요성을 감안한다면 정책적 개입을 통한 해결이 효과적일 수 있다. 따라서 이 글을 통해 국내 빅데이터 산업의 현황과 문제점, 그리고 빅데이터 산업 활성화를 위한 정책적 과제를 모색해 보고자 한다.

### 2. 빅데이터의 개요

#### (1) 빅데이터의 개념

빅데이터는 규모, 종류, 변동성(volume, variety, velocity; 3Vs)이 모두 커서(big) 기존의 방식으로 분석하기 어려운 디지털 데이터를 의미한다. 여기서 기존의 방식이란 숫자 데이터를 정형화된 형태(spreadsheet 등)로 저장하고 모집단의 특성·추세 등을 통계적으로 분석하는 것을 말한다. 이에 반해 빅데이터는 수치화가 어렵고 형태가 다양하고 값이 빠르게 변하는 비계량·비정형 변수들을 포함하며 분석 결과는 모집단의 특성값이 아니라 시각화(visualization)로 나타나는 경우가 많다.

개념의 이해를 돕기 위해 간단한 예를 들어보자. 대형마트가 마케팅 개선 목적으로 소비자의 장바구니 품목 데이터, 실시간 포털사이트 검색어, SNS 데이터 등을 결합하여 만든 데이

터는 그 규모·종류·변동성이 크기 때문에 빅데이터에 해당한다. 그러나 통계청이 5년마다 실시하는 인구주택총조사는 데이터 규모 자체는 아주 크지만 그것이 단일 데이터 세트이고 5년간 값이 고정되기 때문에 빅데이터라기보다는 ‘대형(large) 데이터’라고 보는 것이 타당하다.

## (2) 4차 산업혁명과 빅데이터

4차 산업혁명은 기존 산업의 디지털 전환(digital transformation)이나 태생적인 디지털(born digital) 생태계를 배경으로 한다. 그 결과 대량의 디지털 데이터가 실시간으로 생성되어 빅데이터 분석을 해볼 수 있는 영역이 넓어지고 분석의 정확도가 개선된다. 이를 통해 지금까지는 볼 수 없었던 경향·관계 등을 밝혀내서 사회·경제적으로 더 큰 가치를 창출하게 된다. 구체적으로 기업은 신제품 개발, 제품·서비스 개선, 고객 맞춤형 마케팅 등을 할 수 있다. 창업자들은 혁신적인 창업의 기회를 얻고, 공공 부문은 새로운 정책 수요와 대안을 발굴할 수 있다.

한편 4차 산업혁명은 과거 ‘정보화’가 차지했던 영역을 인공지능 기반의 ‘지능화’로 업그레이드하는 것이다. 따라서 4차 산업혁명의 성공 여부는 인공지능 수준과 직결될 것인데, 최근 인공지능은 빅데이터의 영향을 크게 받는다. 빅데이터가 활발하게 이용되는 환경이 조성된다면 기계학습을 통해 인공지능의 알고리즘이 더욱 정교하게 개선되고, 궁극적으로는 4차 산업혁명의 완성도를 높이게 될 것이다.

결국 빅데이터는 4차 산업혁명의 산출물이자 동시에 추동력이다. 빅데이터 생태계를 어떻게 조성하는지에 따라 4차 산업혁명의 가능성·속도·성과가 달라질 수밖에 없다.

## 3. 국내 빅데이터 산업의 현황과 문제점

### (1) 국내 빅데이터 산업 현황

2017년 기준 국내 빅데이터 시장 규모는 4,547억원으로 전체 ICT 산업 총생산(2016년 기준 428조원)의 약 0.1% 수준에 불과하다.<sup>1)</sup>

빅데이터 시장 규모가 크지 않은 것은 빅데이터를 도입한 기업·기관의 수와도 관련이 있다. 2017년 하반기를 기준으로 종사자 100인 이상 국내 기업·기관의 빅데이터 도입률은 약 7.5%이다. 업종별로 보면 상대적으로 빅데이터 수요가 큰 금융업은 22.7%로 높지만 제조업은 4.2%에 불과하여 업종간 편차가 크다. 기업 규모도 빅데이터 도입에 영향을 미치는데 매출액 1,000억원 이상 기업의 빅데이터 도입률은 13.8%로 전체 평균에 비해 약 1.8배 높다.

### (2) 빅데이터 산업의 문제점

빅데이터 제품·서비스의 수요자(수요기업) 입장에서 본다면 빅데이터 도입에 필요한 자원과 도입 필요성에 대한 인식이 낮고, 빅데이터에 대한 이해가 부족한 것이 큰 문제점이다.

자원 부족은 빅데이터라고 할 만한 데이터가 없고, 빅데이터를 운용할 전문인력과 예산이 부족하다는 것을 의미한다. 특히 기업 입장에서는 내부의 데이터 부족을 해결하기 위해 다른 기관이 보유하고 있는 데이터를 가지고 와서 활용하는 것이 필요한데, 현행 법·제도상 타기관이 보유한 개인정보를 제공받기가 어렵고,<sup>2)</sup> 공

1) 한국정보화진흥원, 「2017 BIG DATA 시장현황 조사」, 2018.

2) 한국은 「개인정보 보호법」의 적용을 받는 ‘개인정보’의 개념 범위가 넓고 정보주체의 동의를 받지 않아도 되는 예외가 명백히 정보주체 등의 이익을 위해 필요한 경우, 통계·학술 목적을 위한 경우 등으로 제한되어 있다. 또한 개인정보에서 식별가능성을 제거하는 조치 및 그 효력에 대한 기준이 법률적으로 명확하지 않아 실제 사용이

공기관이 공개한 정보들도 막상 활용하기에는 품질이 낮아서 부족한 데이터를 보완하는 것이 쉽지 않은 상황이다.

자원이 부족해도 빅데이터에 관한 수요가 크다면 빅데이터 도입이 활발할 수 있는데 국내 기업·기관들은 빅데이터 분석을 하기에는 조직 규모가 작고, 빅데이터 분석이 필요한 업무가 없고, 경영진의 관심도 낮다고 인식하고 있어서 빅데이터 수요 자체가 크지 않다. 그리고 소수이기는 하지만 여전히 빅데이터를 잘 모르거나 그 효과를 불신하는 경우도 있다.

〈표 1〉 빅데이터 미도입 이유

구분	빅데이터 미도입 이유(% , 중복응답)
자원 부족	자체 보유한 데이터 없음 (45.6) 전문 인력 없음 (26.4) 예산 부족 (15.4)
필요성 및 관심 부족	기업 규모가 작음 (23.2) 도입 필요한 업무 없음 (14.9) 경영진의 무관심 (9.2)
이해 부족	도입 효과 불신 (5.4) 빅데이터 잘 모름 (5.1)

자료 : 한국정보화진흥원(2018)

빅데이터 프로젝트를 수행하는 빅데이터 공급기업(사업자)과 수요기업 입장에서 본다면 자금 조달의 곤란, 빅데이터에 대한 낮은 인식, 전문인력 부족이 공통적인 문제점이다. 여기에 공급기업은 협소한 국내시장과 해외시장 진출의 곤란을, 그리고 수요기업은 믿을 만한 성공 사례 부족을 추가적인 문제로 제기한다.

이상의 내용을 종합해 볼 때 국내 빅데이터 산업의 핵심 문제점은 데이터 부족, 전문인력 부족, 낮은 인식 등으로 요약할 수 있다.<sup>3)</sup>

어렵다. 그 결과 정보주체의 개별적인 동의가 없다면 기관간 개인정보를 이전하는 것이 현실적으로 쉽지 않다.  
3) 데이터·인력·인식 등이 개선되면 자연스럽게 빅데이터에 대한 투자와 시장 수요가 확대될 수 있으므로 자금 조달 어려움, 시장 수요 부족 등은 독자적인 문제점으로 분류하지 않기로 한다.

## 4. 개선 과제

### (1) 데이터 확대<sup>4)</sup>

많은 데이터들이 개인정보 보호에 관한 규제에 묶여 이동·결합 등이 제한되고 있는 상황을 감안했을 때, 개인정보 수집·활용·제공 등에 대한 사전동의 원칙의 예외를 현실적인 수준으로 설정하여 빅데이터 분석에 활용할 수 있는 데이터 양을 확대하는 것이 필요하다.

외국은 이미 이러한 입장을 적극적으로 수용하고 있는데, 대표적으로 유럽연합은 2018년 5월부터 「일반개인정보보호법」(GDPR)을 시행하면서 가명처리(pseudonymisation)된 개인정보를 공익·연구·통계의 목적으로 사용할 경우에는 정보주체의 동의 없이도 빅데이터 분석할 수 있도록 했다. 일본은 2017년 5월부터 시행된 「개인정보보호법」 개정 법률을 통해 ‘익명가공정보’는 목적의 제한 없이 정보주체의 동의 없이도 이용할 수 있도록 했다. 미국은 개별 법률에서 개인정보의 보호와 활용을 규정하고 있는데, 예를 들어 「의료정보보호법」(HIPAA)은 개인 의료정보를 엄격하게 보호하면서도 동시에 이름·주소·전화번호 등 18가지의 개인식별자가 제거된 데이터는 자유롭게 이용할 수 있도록 규정하고 있다.

한국도 정보주체의 동의 없이 개인정보를 처리할 수 있는 예외적인 상황을 보다 분명하고 현실적인 수준으로 설정하는 것이 필요하다. 실무에서는 ‘비식별정보’라는 말이 통용되고 있는데 이 표현을 법률상 용어로 수용할 것인지, GDPR의 ‘가명정보’ 표현 등을 활용할 것인지,

4) 데이터 거래 기반(데이터마켓) 조성도 대안이 될 수 있으나 이는 개인정보 이전, 공공데이터 품질 제고 이후에 실현될 수 있는 것이어서 이번 논의에서는 제외한다.

지 판단해야 한다. 그리고 개인정보를 식별 가능성이 낮은<sup>5)</sup> 데이터로 전환하는 기술·절차에 관한 핵심 사항은 법률로 정해야 한다. 이와 관련하여 정부가 2016년 발표한 「개인정보 비식별조치 가이드라인」<sup>6)</sup>의 법률적 근거를 마련하는 것도 방법이다. 또한 정보주체의 동의를 받지 않아도 되는 예외적 상황에 시장조사와 같은 상업적 목적도 포함시킬 것인지에 대한 사회적 합의 도출이 필요하다.

이와 동시에 개인정보 보호의 체계성과 실효성을 높이기 위해 여러 법률<sup>7)</sup>에 산재해 있는 개인정보 조항을 체계적으로 정비하고, 법률 위반에 대한 제재 수준을 현실화해야 한다.

다음으로 공공데이터는 그 자체로도 중요하고 다른 데이터와 결합될 가능성도 높아 빅데이터 활용의 촉매제가 될 수 있기 때문에 공공데이터 개방의 질과 양을 개선해야 한다. 관련 법률을 정비하여 공개할 수 있는 정보의 양을 확대하고, 현재 개방되어 있는 정보의 활용도를 높이는 질적 개선 노력이 병행되어야 한다. 또한 개방된 데이터에 대한 메타데이터(meta data)를 구축하여 어디에 어떤 데이터가 있는지 편리하게 알 수 있도록 해야 한다.

## (2) 인재 양성

빅데이터 프로젝트를 기획·추진·관리할 수 있는 다양한 경험과 지식을 가진 빅데이터 전문가가 필요하다. 이를 위해 대학을 통한 신규인

력 양성과 각 분야별 재직자에 대한 빅데이터 직무교육이 균형을 이루어야 한다. 단기적으로는 소관 부처인 과학기술정보통신부·교육부·고용노동부 등의 정책 협조가 필요하며, 장기적으로 관련 법령에 빅데이터 인재 양성에 관한 규정을 마련하는 방안을 고려해 볼 수 있다.

## (3) 인식 확산

빅데이터에 대한 이해도를 높여야 한다. 일반인들의 인식 개선도 중요하지만 기업·기관의 정보화 담당자나 의사결정자가 빅데이터를 단순한 비용 지출이 아니라 의사결정의 합리성을 개선하는 투자라고 인식할 수 있도록 해야 한다. 동시에 빅데이터 도입의 성공사례를 확산시켜 빅데이터 도입을 망설이는 수요자에게 구체적인 선례를 제공해 주는 것이 필요하다.

이를 위해 「정보통신 진흥 및 융합 활성화 등에 관한 특별법」 제35조(정보통신융합등 문화의 확산 장려)에 빅데이터와 같은 신규 정보통신기술에 대한 홍보, 인식개선, 성공사례 확산 등을 추가하는 방안을 고려해 볼 수 있다.

## 5. 나가며

빅데이터와 같은 신기술은 남들보다 빨리 그 편익을 경험하고, 이를 기반으로 다음 단계로 발전하는 ‘확산의 속도’가 중요하다. 이를 위한 대안은 데이터 확대, 인력 양성, 인식 확산이다. 앞으로 빅데이터에 대한 과감하고 구체적인 정책과 시의 적절한 입법적 지원을 기대해 본다.

□ 「이슈와 논점」은 국회의원의 입법활동을 지원하기 위해 최신 국내외 동향 및 현안에 대해 수시로 발간하는 정보 소식지입니다.

5) 식별 가능성이 없으면 익명정보(anonymous data)가 되기 때문에 「개인정보 보호법」의 적용 대상이 아니다.

6) 이 가이드라인은 개인정보의 비식별조치·적정성평가·결합 등의 방법과 수행기관을 규정하고 있다. 그러나 법률의 근거(위임)가 없다는 비판을 받고 있다.

7) 「개인정보 보호법」, 「정보통신망 이용촉진 및 정보보호 등에 관한 법률」, 「신용정보의 이용 및 보호에 관한 법률」 등이 대표적이다.