

01

총론

제1장 빅데이터와 데이터베이스 산업

제1절 데이터의 활용·분석

제2절 빅데이터의 이전과 이후

제2장 데이터를 이용한 창조경제 활성화

제1절 창조경제의 동력

제2절 빅데이터 통찰과 창조경제

제3절 향후 전망

1

CHAPTER

빅데이터와 데이터베이스 산업

제1절 데이터의 활용·분석

1. 데이터의 활용·분석

과거에는 유용한 정보를 얻기 위해 자료를 수작업으로 분석하였지만 근래에는 자료의 양이 커지고 복잡해짐에 따라 자동화된 분석이 필요하게 되었다. 최근 정보산업의 발전으로 대용량 자료(large data set)의 수집, 저장, 분석에 대한 연구와 기술이 발달함에 따라 다양한 분석 방법들을 적용한 데이터 분석이 활발해지고 있다. 1950년대에 신경망(neural networks), 군집분석(cluster analysis), 유전자 알고리즘(genetic algorithm), 1960년대에 의사결정나무(decision trees), 1990년대에 서포트 벡터 머신(support vector machines) 등의 방법론들이 개발됨으로써 대용량 자료를 분석하는 것이 가능하게 되었다.

데이터 분석은 데이터마이닝이라는 기법을 적용하여 숨겨진 패턴을 찾는 과정으로 볼 수 있으며 학계, 산업, 정부 등의 여러 분야에 널리 활용되고 있다. 데이터 분석은 대용량의 자료로부터 정보의 요약과 미래에 대한 예측을 목표로 자료에 존재하는 관계, 패턴, 규칙 등을 탐색하고 DB 마케팅, 신용평가, 생물정보학, 텍스트 마이닝, 사기 방지 등의 다양한 분야에 적용하여 유용한 지식을 추출하는 일련의 과정으로 정의할 수 있다. DB에서의 지식탐색(knowledge discovery in database), 기계 학습(machine learning), 패턴 인식(pattern recognition), 통계학(statistics)등의 많은 분야에서 관련 연구들이 진행되고 있다.

빅데이터 시대에는 데이터가 정보통신에서 분리된 독립적인 주체로 발전되었으며 정보통신의 주도권이 인프라, 기술, 소프트웨어에서 데이터로 전이되고 있다. 스마트 시대에는 데이터의 저장-검색-관리-공유-분석-추론의 전체적인 과정이 업그레이드되고 있으며 정보화 시대와 차별적으로 성장하고 있다. '빅데이터'의 등장으로 데이터 분석의 중요성에 대한 인식과 다양한 데이터마이닝 기법들이 새롭게 재조명되고 있다. 수

집된 데이터에서 유용한 정보를 찾아내는 데이터마이닝은 혁명적인 신기술로 평가되고 있으며 데이터의 절대적인 양이 많아지자 의미 있고 가치 있는 정보를 발견할 가능성을 높여 나가고 있다.

(그림 1-1-1) 데이터의 과거·현재·미래



※ 출처 : 정지선, 新가치 창출 엔진, 빅데이터의 새로운 가능성과 대응 전략, 2011

데이터마이닝은 기업의 비즈니스 전략 외에도 컴퓨터 과학이나 통계적 관점으로 활용되어 부가가치가 높은 정보를 창출하는데 활용되고 있다. 국방 분야(테러위협 분석 등), 의료 분야(질병 진단 및 유전자 분석 등), 에너지 분야(전력수요 예측 및 자원 탐사 등)에서 활용되고 있으며, IBM의 왓슨(Watson), 애플(Apple)의 시리(Siri)처럼 사람의 말을 이해하고, 대안을 제시하는 인공지능 분야에서도 필수 요소로 각인되고 있다. 또한 기업들은 수많은 고객·업무·거래 데이터를 활용하여 발 빠른 의사결정과 미래 수요 예측, 경영 성과관리에 기여하는 BI(Business Intelligence)전략을 도입하여 활용하고 있다. BI는 의사결정에 필요한 데이터를 수집, 저장, 처리, 분석하는 일련의 기술과 응용 시스템으로 대용량 데이터 저장소인 데이터웨어하우스(DW), 데이터 질의 및 보고 도구, 데이터마이닝, 비즈니스 성과관리 등의 요소로 구성된다.

고객관계관리(CRM)도 고객의 행동이나 평가를 분석하여 예측에 기반한 맞춤형 마케팅, 고객 수요 대응, 신뢰도 향상에 기여하고 있다. 데이터 자원이 축적과 공유를 통해 엄청난 규모로 쌓이면서 데이터의 역할은 '분석과 추론'의 방향으로 진화하고 있다. 대규모 데이터를 기반으로 한 자연언어 처리, 기계 학습(machine learning), 인공지능 기술 등을 토대로 맥락을 이해하고 추론서비스에 적용되어 상용화되고 있다. 지능형 서비스들은 분석 데이터가 늘어나고 기계 학습이 진행될수록 자연어에 대한 이해도가 상승하게 되어 맥락 이해와 추론이 기반이 되는 인공지능형 서비스, 상황인식 서비스, 개인 맞춤 서비스의 발전으로 이어지게 되고, 사용자의 개인정보나 상황정보, 환경정보, 위치정보 등과 융합하여 특정 개인에 특화된 서비스의 제공이 가능해진다. 또한 대용량 데이터 분석을 활용한 인지, 지각, 결정을 제공하는 자율시스템이 점차 확대될 것으로 예상된다. 스마트 시대의 주요 패러다임을 선도하기 위해서는 빅데이터의 활용이 필수 불가결한 요소이며 그 수준이 경쟁력과 성패를 좌우할 척도가 될 것이다.

2. 데이터 분석 패러다임

전문가들이 설명하는 창조는 4단계로 진행된다. ①Data+System=Information; 데이터를 가공하여 가치 있는 정보(Information)를 만든다. ②Information+Experience=Knowledge; 정보에 경험을 더해 지식(Knowledge)을 쌓는다. ③Knowledge+Intuition=Wisdom; 지식에 직관(융합)을 담아 지혜(Wisdom)를 터득한다. ④Wisdom+Imagination=Creation; 지혜에 상상의 나래를 달아 창조(Creation)를 일으킨다.

종래에는 데이터를 가공하여 정보를 만들고 정보가 축적되면 지식이 되고 지식이 결합되고 성숙되면 지혜가 된다고 생각했다. 이를 D-I-K-W(Data-Information-Knowledge-Wisdom) 사이클이라고 부른다. 지금까지는 데이터를 축적하고 정보화하는 것이 DB산업이었고, 지식과 지혜는 사람의 영역으로 생각했다. 그러나 데이터가 광범위하게 축적되고 활용할 수 있게 됨으로써 D-I-K-W 사이클은 D-I-A(Data-Insight-Action) 사이클로 대체되고 있다. DB산업은 데이터를 분석하여 의미를 찾아내고 이를 실행으로 연결시키는 산업이 되고 있는 것이다. 이러한 데이터 패러다임의 변화로 인해 DB산업은 창조경제의 주역으로서 충분한 의미를 갖게 되었다.

빅데이터의 등장과 더불어 데이터 분석에 대한 패러다임은 큰 변화를 맞이하고 있다. 빅데이터 시대에 요구되는 새로운 경영 패러다임에 초점을 맞춘 것이다. 제한된 데이터를 바탕으로 의사 결정이 이뤄진 과거의 경영 방식에서는 경험과 상식이 운영의 핵심이었다면, 빅데이터 시대의 경영은 의사 결정에 필요한 모든 데이터를 얻을 수 있다는 전제하에 이뤄지고, 데이터에 근거하지 않은 통념이나 경험은 오히려 걸림돌이 될 수 있다. 따라서 빅데이터 패러다임에서는 종래 기업들의 시장 및 문제점 분석과 예측 활동의 방향성이 바뀌어야 한다.

우선 데이터 분석의 시제를 과거에서 현재로 옮겨오게 된다. '과거에 무엇이 일어났는가'를 분석하는 것에서 '현재 무엇이 일어나고 있는가'를 분석하고 대응하는 것으로 변화해야 하며, 현상 파악에 그치지 않고 구체적인 이윤 창출이나 미래 예측 등의 운영방안 설정을 데이터 분석 방식으로 선택하는 것도 빅데이터 패러다임 속에서 요구되는 변화다. 반도체를 예로 들면 이전에는 반도체를 생산하는 설비 기계 각각의 생산율이 전체 반도체 생산율로 직결된다는 통념이 있었다. 하지만 반도체 공정을 모아 데이터 분석을 실시한 결과, 각각 기계의 생산성이 높다고 해서 전체 반도체 생산율이 증가하는 것은 아니라는 사실이 발견됐다. 그 이유는 공정에 따라 생산 시간 및 생산율에 차이가 있었던 것이다. 이러한 분석을 바탕으로 수십억 원의 예산을 들여 설비를 확충하는 대신 전체 공정을 대폭 수정하는 과정을 거쳤다고 한다. 그 결과 생산성 증가와 비용 절감의 효과를 이끌어 낼 수 있었다는 것이다. 결국 단순히 통계치에 의존한 예측보다는 구체적이고 결과 중심적인 액션을 추구하는 기업 경영이 필요하다는 것이고, 패러다임과 자원의 변화를 이해하는 것이 기업 경쟁력을 높이고 생존하는 조건이 될 것임을 알 수 있다.

과거 데이터 분석에서 활용했던 경험과 주관적인 판단은 데이터-근거 기반의 의사결정으로, 트렌드 분석에서 현재 상황의 실시간 분석 형태로, 분석·리포트를 통한 현황파악에서 직접적인 비용 절감·매출 확대 등

의 결과 도출로, 통계에 근거한 예측에서 구체적인 운영과 결과를 중심으로 변화하고 있다. 마케팅 분야에서는 전통적인 CRM에서 이벤트 기반의 마케팅으로, 판매 분야에서는 제품별·고객별 판매 분석에서 제품의 가격경쟁력과 고객에게 필요한 제품 추천 등으로 변화하고 있다. 과거 IT나 SoC(System on Chip) 분야에서는 시스템의 변화나 운영 현황 등의 통합 모니터링에 초점을 맞추고 있었으나, 현재는 사전 장애에 대한 예측과 IT 자원의 증설계획, 재해에 대한 예측과 다양한 시민의견 분석 등으로 관심이 확대되었다. 또한 정치 분야에서는 선거 전략 도출 및 득표율 예측에 초점을 맞추어 활용하고 있다.

3. 빅데이터 분석·활용

빅데이터에 대한 정의는 다양하지만 다음과 같이 요약할 수 있다. ①기존 방식으로 저장·관리·분석하기 어려운 정도의 큰 규모의 자료, ②대량의 정형 및 비정형 데이터로부터 가치를 추출하고 결과를 분석하는 기술로 요약될 수 있으며, 최근 스마트폰 등 모바일 기기의 확산과 소셜 미디어의 성장으로 다양한 종류의 대규모 데이터가 생산되는 빅데이터 시대가 도래되고 있다.

빅데이터는 단순히 큰 데이터가 아니라 부피가 크고, 변화의 속도가 빠르며, 속성이 매우 다양한 데이터라는 세 가지 특징을 가진 데이터로 정의할 수 있다. 원시(raw) 데이터는 가공되지 않은 수집 상태의 데이터를 의미하며, 빅데이터에서는 분석하는 분야 및 대상의 특징에 따라 원시 데이터를 가공하기 때문에 가공되지 않은 그대로의 상태로 저장하는 경우가 많다. 빅데이터는 원시 데이터에서 사용자가 원하는 것이 생기면 그것에 해당하는 데이터를 찾아서 분석해 제공하는 것이며 규모(volume), 변화 속도(velocity), 다양성(variety)이라는 3가지 특징, 즉 3V를 가지고 있는 데이터에서 효과적인 결과물을 창출해 낼 수 있다.

최근에는 기존의 3V 요소에 데이터의 진실성(veracity), 시각화(visualization), 가치(value) 등을 포함해 빅데이터 기술 요소를 6V로 확대해 규명하고 있다. 결국 빅데이터의 핵심적인 특징은 '가공되지 않은 많은 양의 데이터가 빠르고 다양하게 생산·소비되는 분야에서 효과적으로 데이터를 분석·가공하여 가치 있는 정보를 만들어낼 수 있는 데이터'라 할 수 있다. 빠르게 변화하는 ICT 기술에 적극적으로 대응하고 새로운 가치를 찾는 기술을 빅데이터에 적용하여 의미 있는 정보, 가치 있는 정보를 찾아 활용하는 것이 중요하다는 것이다.

빅데이터는 민간, 공공, 과학기술계별로 각각 다양하게 활용되고 있으며 그 외 다른 분야에서도 활용이 가능하다.

첫째, 민간부문에서는 고객 관련 다양한 빅데이터 분석을 수행하여 소비자의 구매 패턴, 제품 만족도 등을 반영한 소비자 맞춤형 서비스를 제공하기도 하며, 빅데이터 분석을 기반으로 제품의 결함이나 개선 방안 등을 마련하여 제품 생산 비용을 절감하거나 사업의 전반적인 효율성을 높이고 있다. 뿐만 아니라 과거 또는 실시간 빅데이터 분석을 통해 미래 트렌드를 분석·예측하여 유망기술을 발굴하고 신규 사업 개발에 활용하고 있다. 연 매출 2조 원이 넘는 국내 미용 분야 1위 회사인 아모레퍼시픽은 2007년 이래 연평균 71%(B2C)

로 증가하는 고객 빅데이터를 분석하여 고객 유형별로 맞춤서비스를 제공하고 있다. 신세계물은 기존 2~5일 걸렸던 고객 데이터 처리·분석 작업을 개선하기 위해 고객과 고객의 거래 관련 빅데이터 분석 프로젝트(C.S.I 프로젝트)를 수행하여 실시간 분석을 통한 신속한 소비자 대응이 가능해졌다.

한국석유공사는 국제유가의 변동이 소비자 가격에 반영되는 시기와 폭을 예측하기 위해 국제유가와 환율, 금융거래정보 등의 빅데이터 분석을 기반으로 한 국내 유가 예보시스템을 개발하여 서비스를 제공하고 있다. 미국 텍사스주의 환경기업인 웨이스트 매니지먼트(Waste Management)는 트럭 2만 대에 탑재된 센서로부터 과적 차량, 차량 경로, 고객과 매립지에서 머문 시간 등의 빅데이터를 수집·분석하여 좀 더 경제적인 제품을 제안하고 효율적인 경로정보 등을 제공하는 등 쓰레기 물류회사에서 물류 및 정보 제공자로 완전히 탈바꿈하는데 성공하였다.

둘째, 공공부문에서는 국가 차원의 공공 빅데이터 개방 확산을 통해 정부의 투명성을 제고하고 국민의 권리를 향상시키고 있으며, 데이터 공개를 통한 경제적·사회적 가치 증대를 위해 단계적으로 공공데이터 공개를 추진하고 있다. 또한 빅데이터 기술을 활용한 공공데이터 분석·운영으로 정부 운영비용 절감, 대내외 변화에 대한 신속한 대처, 삶의 질과 정부 신뢰도 향상에 초점을 맞추고 있다. 뿐만 아니라 소셜 네트워크, 신문, 잡지, 기사 등으로부터 수집된 정보를 분석하여 국가 안전을 위협하는 질병 및 위기 등의 글로벌 위험요소를 사전 예측함으로써 글로벌 위험재난을 관리하고 선제적 대응마련에 활용하고 있다.

미국과 영국 정부는 각각 data.gov와 data.gov.uk를 통해 공공데이터를 공개하고 있으며, 일부는 RDF로 변환·제공하여 정부의 투명성 향상 및 경제적·사회적 고부가가치 창출을 지원하고 있다. 국내에서도 data.go.kr을 통해 공공데이터를 개방하고 있다. 또한 미국 국세청은 빅데이터 분석 기반의 통합형 탈세 및 정부 사기 방지 시스템을 구축해 연간 3,450억 달러에 달하는 세금 누락 및 불필요한 세금 환급을 절감했다. 미국 국립보건원은 'Phillbox 프로젝트'를 통해 약(藥) 검색 서비스를 시행하여 연간 5,000만 달러의 비용을 절감하고 있다. 싱가포르르는 빈번히 발생하는 테러, 전염병 등으로 인한 불확실한 미래에 대비해 질병, 금융 위기 등 잠재적인 위험요소와 불확실성 요소들에 대한 빅데이터를 수집·분석하여 국가위험관리 계획 수립 시스템(RAHS)을 구축하고 있다. 미국 연방수사국(FBI)은 유전자정보은행 CODIS(Combined DNA Index System)를 구축해 1시간 내에 범인 DNA를 분석하고 주정부 데이터베이스와의 연계 및 빅데이터 실시간 분석으로 글로벌 위험이나 재난 관리에 활용하고 있다.

유럽연합(EU)은 비용 절감, 부정 및 오류에 따른 손실 감소, 세수 증대에 따른 비용 효과가 1,500~3,000억 유로 규모일 것으로 예측하며, 국내의 경우 최소 비율 적용 시 10.7조 원 규모로 추산하고 있다. 맥킨지 글로벌 연구소(McKinsey Global Institute)의 2010년 5월 빅데이터에 관한 보고서에 따르면, 빅데이터를 활용함으로써 유럽 공공기관 분야는 약 2,500억 유로의 경제적 이익을 얻을 수 있을 것으로 예측했다.

셋째, 과학기술 분야에서는 연구 장비의 첨단화 및 고성능화로 인해 다양한 분야의 고품질 데이터가 대량 생산되고, 1·2세대인 실험 및 이론 중심 연구에서 최근 3·4세대의 컴퓨팅 자원 및 데이터 중심 연구로 과학

연구의 패러다임이 변화하면서 데이터 중심의 과학이 등장하고 있으며, 의료·에너지·기후 등의 과학기술 빅데이터를 활용하여 국가·사회적 이슈를 발굴하고 문제를 해결하고 있다.

미국 항공우주국(NASA)은 2000년에서 2005년까지 데이터양이 10배 성장했으며, NASA의 지구관측 시스템 데이터 및 정보시스템(The Earth Observing System Data and Information System: EOSDIS)의 경우 PB급¹⁾의 데이터 크기에 도달하였다. 유럽원자핵공동연구소(CERN)의 LHC(Large Hadron Collider)는 16PB의 데이터를 매년 생산하고 있는데, 미립자 연구자들의 예측에 따르면 2015년에는 엑사바이트(Exabyte = 1024×1 Petabyte)급의 저장소와 페타플롭스급²⁾의 컴퓨팅 파워가 요구되고 있다. 또한 센서 장비 및 네트워크 기술의 발달로 인해 다양한 센서 데이터가 실시간 수집되고, 기후 데이터 또한 2030년에는 350PB에 달할 것으로 예측하고 있다.

미국 캘리포니아의 의료진단회사인 카디오DX(CardioDX)는 1억 개 이상의 유전자 분석을 통해 샘플 모델링 및 관상동맥 질환을 예측하는 데 빅데이터 분석 기술을 활용하고 있으며, 소행성 및 혜성연구 프로젝트인 판-스타스(Pan-STARRS)는 연간 2.5PB가 넘는 천체 관측 데이터를 수집·분석함으로써 블랙홀 등의 과학적 현상을 규명하는 데 활용하고 있다.

제2절 빅데이터의 이전과 이후

1. DB산업 동향

최근 빅데이터가 주요 화두로 떠오르면서 전 산업 활동의 근간을 이루고 있는 데이터베이스(이하 DB)의 활용과 분석, 유통 등 DB산업 활동에 국가적인 관심이 주목되고 있다. 또한 OECD에서는 DB를 무형자산으로 포함하여 국가 자산으로서의 DB가치 측정을 위한 노력도 기울이고 있다. 이렇듯 자산으로서의 DB는 전 세계적으로 산업 경쟁력 제고를 위한 핵심 인프라로 자리 잡아가고 있으며 DB산업은 국가 산업 경쟁력 전반에 지대한 영향력을 미치는 핵심 산업으로 인정받고 있다.

2013년 새 정부와 ICT 전담 부서의 출범으로 DB 및 데이터의 가치가 본격적으로 재조명되면서, 정부의 주도 아래 DB산업 활성화를 위한 준비에 박차를 가하고 있다. 이에 DB산업의 정확한 현황을 제시하기 위한 산업 동향 파악에 대한 요청이 끊이지 않고 있으며, 향후 DB산업 시장에 대한 전망을 제시할 수 있는 자료의 필요성이 증대되고 있다.

1) 1 Petabyte: 5MB 음악 약 2억 1475만 곡 또는 700MB 동영상 약 153만 편에 달하는 공간

2) 1 PetaFLOPS: 1초당 1 Petabyte를 처리할 수 있는 컴퓨터 성능

2. 빅데이터의 산업동향

정보통신 분야가 유무선 통신에 기반한 유비쿼터스 환경으로 빠르게 발전하고 있는 상황에서 무엇보다도 주목해야할 점은 다양하게 축적된 데이터의 활용 방안이다. 현재 빅데이터 시장은 미디어, 통신, 금융, 서비스 등 전 산업영역에 걸쳐 데이터 분석 수요가 급증함에 따라 빅데이터 분석 및 솔루션에 대한 수요가 증가하고 있는 추세이다. 특히 개방형 네트워크를 통한 스마트 혁명으로 데이터의 양과 질이 급격하게 증대·향상되면서 빅데이터 분석을 통한 조직의 데이터 분석 수요가 증가하고 있으며, 관련 데이터를 활용한 신규 시장 창출 및 비즈니스가 급성장하고 있다.

위키본(Wikibon)에 따르면 전 세계 글로벌 빅데이터 시장은 184억 달러로 추산되는데, 빅데이터를 통한 서비스 및 상품의 개선, 전략 방향의 수립, CRM 향상 등이 빅데이터 시장 성장의 주요 요인으로 나타났다. 특히 빅데이터 관련 인프라 기술 컨설팅, 시스템 통합 등 IT 서비스 확대가 빅데이터 시장 성장을 주도하면서 2017년에는 485억 달러에 이를 것으로 전망되고 있다.

그러나 빅데이터의 중요성이 부각되고 빠른 시장성장이 예상되고 있는 가운데 관련 시장의 확산은 늦어지는 모습이다. 전 세계 대기업들은 빅데이터 관련 투자 계획만 세워놓고 실제 프로젝트를 주저하고 있는 것이다. 또한 대규모 투자를 단행했던 기업들도 효과에 대해서는 여전히 확신을 하지 못하고 있다. IT 시장조사 기관인 가트너(Gartner)에 따르면 2013년 전 세계 기업의 64%가 빅데이터에 대한 투자 계획을 세웠거나 투자를 시작했다고 한다. 그러나 빅데이터 투자 계획을 세웠던 기업의 8%만 실제 프로젝트를 시작하는데 그쳤으며, 계획과 예산만 잡아놓고 구체적인 활용에 나선 기업은 드물다고 발표했다. 또한 2014년 투자에 나서겠다고 답한 비율은 15%였다. 가트너가 조사한 지역별 빅데이터 투자 의향을 살펴보면 북미지역 기업의 37.8%가 빅데이터 투자 의향을 보이며 가장 높은 것으로 나타났고, 이어서 유럽, 아시아, 남미 지역순으로 조사됐다. 이는 빅데이터 투자에 대한 의지는 높으나 관련 기술을 어떻게 적용해야 할지에 대한 정확한 전략 수립이 미진하기 때문인 것으로 풀이된다.

빅데이터는 하드웨어, 소프트웨어, 서비스가 모두 절묘하게 융합되어야만 성과를 낼 수 있을 정도로 고난도의 역량을 필요로 한다. 그런 이유로 가트너는 2015년까지 포춘(Fortune) 500대 기업의 85% 이상이 빅데이터를 활용해 경쟁 우위를 확보하는데 실패할 것이라고 예측하기도 했다. 이에 대해 부연하자면, 빅데이터를 도입하는 기업이 많을지라도 이를 제대로 활용해 유의미한 성과를 내는 기업은 적을 것이라는 뜻이다.

3. 빅데이터의 이슈와 전망

빅데이터는 우리가 살아가는 세상의 엄청나게 많은 분야에서 사용될 것이며 그 영향력도 매우 클 것이다. 우리는 적절한 대처 방안을 마련하여 새로운 빅데이터 시대를 맞이해야 한다.

빅데이터의 3대 요소는 자원, 기술, 인력이라고 할 수 있다. 미래에는 빅데이터를 핵심 자원으로 인식하고, 필요한 정보를 뽑아낼 수 있도록 리소스를 키우는 것이 중요한 성공 전략이 될 것이다. 주어진 빅데이터를 관리·처리하는 측면과 함께, 활용할 수 있는 외부 빅데이터 자원을 발견하고 확보하는 전략이 필요하고, 데이터의 품질(quality)은 데이터 활용 결과에 중대한 영향을 미치므로 데이터 관리 체계 및 데이터의 신뢰성 확보가 매우 중요하다.

빅데이터는 데이터 자체 뿐만 아니라 관련 도구(tool)나 플랫폼, 분석 기법까지 포괄하는 용어로 의미가 확장되며 IT 패러다임의 변화를 견인하고 있다. 성장을 촉진하는 신기술에 대한 이해가 부족하면 미래 경쟁력 강화를 위한 기회 포착에도 어려움을 겪을 수 있으며, 조직과 기업의 혁신 전략으로 적용할 수 있도록 빅데이터 플랫폼, 빅데이터 분석 기술 및 데이터 분석 기법에 대한 이해가 필요하다. 신기술과 도구(tool)가 뛰어나도 실제 성과를 낼 수 있는 것은 이를 활용하여 적용하는 사람의 역량에 좌우되는 것은 지극히 당연한 일이다. 데이터 처리와 분석능력을 갖춘 인력은 IT 분야 뿐만 아니라 대부분의 기업과 조직에서 필수적으로 확보해야 할 핵심 인력이다. 조직의 차원에서 인재를 확보하기 위해 내부 역량 강화 및 외부 협력이 중요하다.

빅데이터 시장은 데이터의 가치가 증대됨에 따라 업체들 간 데이터를 거래하는 데이터 마켓플레이스가 주목받게 될 것이며, 물리적 현상을 나타내는 데이터를 수집함에 따라 온라인과 오프라인의 연계가 더욱 중요해질 것이다. 그리고 빅데이터 시장에서 탁월한 성과를 내기 위해서는 무엇보다도 빅데이터를 활용할 수 있는 성숙된 IT 조직 문화를 확립해야 한다. 더불어 IT 및 비즈니스 도메인 지식을 갖춘 인력, 그리고 통계학, 수학적 지식을 갖춘 데이터 사이언티스트(data scientist)를 확보하는 것이 중요한 선결 과제라고 볼 수 있다.

빅데이터의 특성상 네트워크상에서 지인과의 일상 대화부터 공유한 이미지, 영상, 텍스트와 같은 모든 것들이 빅데이터로서 수집되고 분석된다. 즉, 무심코 말한 개인정보가 유출될 수 있고, 산업정보 유출, 나아가 군사정보 유출과 같은 심각한 사태까지 발생할 가능성이 있다. 따라서 정부 차원에서 빅데이터 관리에 유념하여야 하며 유출 사태에 대한 대응 방안을 준비해야 할 것이다.

2 CHAPTER

데이터를 이용한 창조경제 활성화

제1절 창조경제의 동력

1. 창조경제 추진현황

지난 2013년은 ‘창조경제 실현계획’이 수립되었고 신성장 동력, 일자리 창출 등 다양한 정책을 쏟아내며 창조경제의 추진 기반을 조성한 해이자 창업 활성화 정책에 주력한 해이다.

창조경제 생태계 측면에서는 미래창조펀드(6천억 원), 성장사다리펀드(2조 원) 등 벤처·창업 지원을 위한 자금을 조성하고, 재창업 자금 확대 등으로 창업 안건망을 강화하여 고부가가치 지식재산 창출 활성화 기반을 마련하였다.

신산업·신시장 창출 측면에서는 소프트웨어 인력양성 및 시장구조 개선 등을 내용으로 하는 ‘소프트웨어 혁신전략’을 수립하는 등 신산업·신시장 창출을 위한 전략적 틀을 마련하였고, ‘창조비타민 프로젝트 추진계획’, ‘농식품 ICT 융복합 확산 대책’ 등을 통해 기존 산업과 과학기술·ICT의 융합 방향을 제시하였으며, 「정보통신 진흥 및 융합 활성화 등에 관한 특별법」을 통해 ICT 융합의 제도적 기반을 마련하였다.

창조문화·인프라 측면에서는 창업교육센터, 창업선도대학 등을 통해 창업교육을 확대하고, 9월에는 국민 누구나 아이디어를 사업화하도록 멘토링하여 실질적으로 창조경제를 실현하는 공간인 창조경제타운을 온라인에 개설했으며, 12월에는 창조경제 성과물을 전시하는 창조경제 박람회를 개최하였다. 또한, 무한상상 국민창업 프로젝트를 추진함으로써 국민의 창의적 아이디어가 마음껏 발현될 수 있는 토대를 닦았다. 그 결과 지난해 벤처기업 수가 전년보다 3.3% 많은 29,135개로 증가하였고, 대학생 창업동아리 수는 1,833개로 전년보다 50% 증가하는 등 괄목할 만한 성과를 창출하였다.

2014년에는 2013년 실적을 바탕으로 창조경제 확산을 본격화하기 위해 창조경제 정책목표를 ‘기반 정착’과 ‘성과 창출 가속화’에 두고, 추진방향을 ‘①정부 주도→민간 주도’, ‘②중앙정부 중심→지역 확산’, ‘③국내 중심→글로벌 확장’으로 설정하였으며, 이러한 목표를 실현하기 위해 정부는 ‘①창조경제 생태계 조성 가속화’, ‘②신산업·신시장 개척을 위한 성장 동력 창출’, ‘③창조경제 기반강화 및 창조문화 확산에 중점을 두었다.

2. 창조경제 구현

지속적인 신성장 동력 발굴과 국민행복기술 기반의 디지털 혁신을 통해 창조경제가 구현될 수 있다. 이를 위해서는 기초과학, 융합과학 연구를 통해 미래를 선도할 중장기적인 신기술을 개발하여 미래 지속가능한 신성장 동력을 발굴하고, 기존 산업 및 비즈니스를 창조적인 방식, 즉 디지털 경제로 전환시킴으로써 새로운 시장을 창출할 필요가 있다.

지금 우리에게 시급한 것은 중장기적으로 신성장 동력을 발굴해 내는 것보다 국민행복기술을 이용하여 새로운 일자리와 새로운 시장을 창조하는 것이다. 국민행복기술은 국민행복을 실현하고 창의성에 기반한 새로운 성장과 혁신을 열어가는 디지털 기술 즉 정보화 기술이고, 전통산업에 기반하여 소자본으로 소규모 사업을 영위하는 국민들이 새로운 시장에서 경쟁할 수 있는 시장을 창출 가능하게 하며, 소규모 투자를 통해 새로운 서비스를 창출하고 이를 통해 더 많은 일자리를 창출하는 원동력이 될 것이다.

이러한 국민행복기술의 핵심은 데이터이다. 아날로그 및 디지털 산업에서 발생하는 데이터를 통합·축적해 가치 있는 디지털 데이터로 변화시킴으로써 새로운 서비스를 창조하고 새로운 시장을 개척할 수 있는 기회의 창출이 가능하다. 또한 축적된 데이터 자체를 분석·활용함으로써 새로운 서비스, 새로운 비즈니스를 창출할 수 있다. 다시 말해 데이터는 데이터 자체가 창조경제의 자원으로 활용 가능하며, 데이터를 통해 새로운 디지털 경제로의 혁신, 즉 창조경제의 구현이 가능해 질 것이다.

3. 창조경제를 이끄는 힘 : 데이터

데이터의 시장 규모 및 경제적 효과는 앞으로 더욱 커질 전망이다. 무엇보다 공공데이터의 개방이 가져올 효과는 국가 기관의 예산 투명성이 확보되며 정부에 대한 신뢰가 형성되어 소모적인 논쟁이 사라져 사회적 간접비용이 큰 폭으로 감소될 수 있을 것이다. 공공데이터의 개방은 이러한 경제적 효과 뿐만 아니라 국가적 차원에서 정책 및 예산의 투명성을 확보하고 보다 객관적인 데이터 및 의견 수렴이 가능하며 정책의 효율성을 높이는 계기가 될 것이다. 아울러 산업적 차원에서는 새로운 사업이 등장하고 고용과 새로운 부가가치 창출이 가능할 것으로 예상된다.

공공데이터는 그 활용 범위가 매우 넓기 때문에 농업, 어업을 포함한 1차 산업에서부터 CT, IT, BT 등 첨

단기술 분야까지 활용할 수 있다. 개인 및 중소기업은 공공데이터를 활용해 창의력을 기반으로 하는 새로운 사업에 도전할 수 있는 기회가 될 것이다. 특히 소자본으로 새로운 사업에 도전하는 벤처 중심의 사회로 도약하는 기회를 마련할 수 있을 것이다.

공공데이터 개방 시대와 더불어, 누구나 적은 비용으로 데이터를 분석할 수 있는 시대를 맞이하였다. 금융, 생산관리, 마케팅 등 산업 분야를 시작으로 데이터 분석의 효용성과 가치는 점차 확대되고 있으며, 많은 기업들이 유사한 서비스 제품 기술을 이용해 경쟁함에 따라 데이터 분석을 통한 혁신과 가치 창출은 중요한 경쟁력으로 자리매김하고 있다. 이제 우리 사회가 지식, 즉 데이터가 힘이 되는 지식기반 사회로 진화하는 것처럼 정부 역시 데이터 분석을 통한 지식기반 정부로 진화해야 하며, 이를 통해 창조경제 구현을 위한 체계를 완성할 수 있을 것이다.

데이터의 개방과 공유는 데이터 분석·활용을 통한 가치 창출이 핵심이며, 반대로 데이터의 분석과 창의적 활용을 통해서 새로운 비즈니스 창출, 다양한 맞춤형 서비스 개발이 가능해질 것이고, 이를 통해 데이터 개방과 공유 효과를 극대화할 수 있을 것이다. 데이터는 창조경제의 핵심자원이며, 공개된 데이터를 활용함으로써 새로운 비즈니스와 일자리 창출이 가능할 것이다.

데이터가 개방되고 저비용 분석이 가능해진 지식정보시대에서 데이터는 '21세기의 원유'라 불리며 세계적으로 그 가치를 주목받고 있다. 특히 창조경제를 부르는 요즘, 데이터는 창조의 소재로서 자유롭게 연계, 융합되며 전 산업 분야에 새로운 가치를 창출하고 있다. 미래창조의 시대, 데이터로 준비해야 한다. 창조경제 시대에서 데이터가 어떻게 새로운 가치를 창출하는지 조명하고 다가오는 미래에 걸맞게 데이터를 바라보아야 한다.

제2절 빅데이터 통찰과 창조경제

1. 데이터 통찰과 실행

정보화 시대에서 데이터는 D-I-K-W(Data-Information-Knowledge-Wisdom) 사이클에 맞추어 데이터를 가공하고, 지식화하며, 의사결정을 위한 지혜를 얻는 데에 활용되었다. 하지만 데이터 홍수 속에서 맞이한 빅데이터 시대에서의 데이터는 D-I-A(Data-Insight-Action) 사이클을 따라 엄청난 양의 데이터로부터 데이터를 통찰하고, 이를 기반으로 정책적, 그리고 비즈니스적 의사결정을 위한 핵심 요소로 활용되고 있다.

기존에는 경험 및 주관적 판단에 기반하여 실행(의사결정)을 수행하였으며, 데이터 분석은 과거 트렌드 분석, 현황 파악, 통계에 근거한 예측을 수행하기 위한 기초 단계일 뿐이었다. 하지만 분석 대상이 빅데이터로 변화하면서 데이터 분석을 통한 패러다임과 그 위상 또한 변화를 맞이하였다. 근거 기반의 의사결정을 수행

하고, 현재 상황에 대한 실시간 분석, 직접적인 비용 절감 및 매출 확대와 같은 결과 도출 목적에 활용되며, 미래에 대한 예측, 구체적인 운영 및 결과 중심의 실행을 수행하게 된다.

데이터 분석은 데이터 통찰로 진화하고 있다. 이미 발생한 정확한 사실, 즉 가까운 미래에 반복될 가능성이 있으나 미래의 상황에 따라 다른 결과가 나타날 가능성도 높은 서술적 분석(descriptive analytics)에서 가까운 미래에 대한 한두 가지를 예측할 수 있을 정도의 예측적 분석(predictive analytics)을 넘어, 빅데이터 시대에서는 다수의 값에 대한 예측을 결합해 미래의 상태에 대한 다차원적, 종합적, 입체적 그림을 제시하는 지시적 분석(prescriptive analytics), 즉 통찰로 발전하고 있다. 더욱이 통찰에서 끝나지 않고 그 결과를 실행에 옮기기 위한 분석으로 진화하고 있다.

이제는 빅데이터 활용을 통해 통찰(insight)과 실행(action) 사이의 갭을 채우는 것이 매우 중요하며, 이는 실행 중심의 빅데이터 활용이 중요하다는 것이다. 다시 말해 목적의식 또는 목표가 분명해야 함을 의미하기도 한다.

가트너 그룹이 발표한 '2014년 10대 전략기술'에서 빅데이터가 빠졌다. 가트너 하이프 그래프(Gartner Hype Graph)를 보면 2014년에는 빅데이터가 환멸의 굴곡(trough of disillusionment) 시기를 지날 것으로 예상되기 때문이다. 하지만 스토리지 비용이 점점 더 내려가고 분석해야 할 데이터양이 증가하는 것을 고려할 때 빅데이터 기술은 점차 산업 필수 인프라로 자리 잡을 것으로 생각되기 때문에, 이제는 빅데이터 시스템을 만드는 것보다 그것을 어떻게 활용하고 어떻게 가치 있게 만들 것인가에 대해 더 관심을 기울여야 한다. 물론 빅데이터 통찰을 통해 알게 된 사실들이나 결과들은 매우 소중하다. 그러나 그것을 어떤 종류의 가치, 즉 실행으로 바꾸지 않는다면 사실 그다지 쓸모가 없을 수 있다.

빅데이터 분석을 통해 얻은 통찰과 실시간 정보를 결합하면 상황 변화를 빠르게 이해할 수 있고 충분한 정보를 바탕으로 시의적절한 실행(의사결정)을 내리며 빠르게 대응할 수 있다.

2016년에는 70% 이상의 영리기업이 실시간 예측 분석을 이용하여 비즈니스 프로세스를 운영할 것으로 전망되고 있으며, 더 많은 지도자 및 경영인들이 전략적·정책적 의사결정을 위해 빅데이터 분석-통찰-실행이라는 새로운 사이클을 활용하게 될 것이다.

2. 빅데이터 실행을 통한 창조경제

모바일, 클라우드, 소셜 네트워크 서비스(SNS)의 등장으로 근본적인 인프라와 데이터의 변화가 일어나며 데이터를 바라보는 시각이 변화하고 있다. 즉, IT의 주도권이 인프라, 기술, 소프트웨어 등에서 데이터로 전이되고 있고 데이터가 IT에서 분리된 독립적인 주체로 발전하고 있으며, 데이터 폭증에 대한 대응과 데이터 분석 및 실행이 ICT의 중요 이슈로 부각되고 빅데이터가 ICT 시장과 기술발전의 핵심 주제로 인식된다.

축적과 공유를 통해 유의미하게 분석할 수 있는 데이터 자원이 쌓이자 데이터의 역할은 통찰과 실행의 방

향으로 진화하고 있다. 다시 말해, 빅데이터는 사회 환경변화를 신속하게 감지하고 대응하는 역량, 스마트한 지능형·개인화 서비스를 창출하는 창조경제의 원천 요소로 작용한다. 스마트 시대, 즉 창조경제 시대에는 데이터의 저장-검색-관리-공유-분석-추론의 전체적인 과정이 업그레이드되고 정보화 시대와 차별적으로 발전하고 있으며, 통찰과 실행을 지원하기 위해 대규모 데이터의 분석 결과를 구체적·표면적·실용적으로 전달하는 시각화 기술 또한 부각되고 있다.

빅데이터 환경에서는 의미 없이 단절되어 있는 정형·비정형·반정형 데이터들이 상관성에 의해 연계되고 실시간으로 분석되어 정책적 의사결정을 지원할 수 있을 때 그 가치가 극대화된다. 그러나 세상의 모든 데이터를 연계시키는 것은 데이터 규모와 저장 비용을 고려하면 매우 비효율적이므로 패턴으로부터 의미 있는 정보(knowledge)를 찾는 노력이 필요하다. 따라서 빅데이터 자체보다는 데이터 분석 및 실행에 초점을 맞춰야 한다.

사회 발전의 속도가 빨라지고, 위험요인과 복잡성이 증가할수록 시스템적으로 신속하게 환경변화를 감지할 필요성이 증가하고 있다. 빅데이터는 이러한 불확실성, 리스크, 스마트, 융합 등으로 대변되는 미래사회의 특성에 대응하는 역할을 수행하며, 빅데이터의 통찰 및 실행을 통해 미래의 통찰력, 대응력, 경쟁력, 창조력을 통한 기회 요인을 창출하는 미래사회의 가치 창출 엔진이자 창조경제의 핵심 엔진이다.

제3절 향후 전망

1. 공공데이터 개방전략 수립

정부가 공공데이터 개방을 적극 추진하는 이유는 정부 운영의 투명성을 높이는 것보다 새로운 ICT 기술을 이용한 대국민 서비스 질적 향상을 추구하고 혁신적이며 창의적인 연구 및 사업화가 진행될 수 있기를 기대하고 있기 때문이다.

따라서 개방적 차원을 넘어 지원과 제도 개선이 뒤따라야 하는데, 지원 정책은 적절한 보상 체계를 마련하는 것이며 제도 개선은 사업화 및 서비스 도입의 장벽을 제거하는 것이다. 특히 지원정책 수립에 있어 공공데이터를 활용해 국가 지출 비용을 절감하는 효과가 나타날 경우 비용 절감의 일정 비율을 일정기간 사업자 및 개인에게 지원하는 방안을 고려할 수 있을 것이다.

기업 및 개인은 공공데이터를 어떻게 활용할지 적극적인 고민이 필요하다. 공공데이터와 기업의 내부 정보를 결합하여 업무 효율성을 강화하고 새로운 상품 및 서비스 개발에 필요한 자료로 활용할 수 있다. 공공데이터는 향후 기업에 새로운 가능성을 제공하는 동시에 또 다른 위협 요소로 작용하게 될 것이다. 새로운 가능성은 마케팅, 소비자 맞춤 전략 수립, 시장 반응, 새로운 전략 수립 등에 활용이 가능한 한편 소비자들이

더 많은 정보를 갖고 기업 및 서비스를 비교할 수 있게 됨에 따라 소비자의 평가가 더욱 세심하게 이루어질 것이다. 특히 공공분야인 보건의료, 안전, 교육, 교통 등에서의 공공데이터 활용 정도에 따라 공공데이터 활용 확산에 결정적인 영향을 줄 수 있을 것이다.

이러한 공공데이터는 빅데이터의 범주에 포함될 수 있는데, 아직 빅데이터에 대한 활용도가 높지 않다는 점에서 데이터 이용 활성화 및 분석 방법의 다변화가 필요하다.

그리고 미국의 헬스케어 사례에서 알 수 있듯이 관련 데이터 개방이 공동으로 추진되어야 창조적 서비스 및 비즈니스 확산에 효과가 있으며, 아젠다(agenda)별 공공데이터 개방은 분야별 개방 경쟁을 유도하여 개방 속도를 더욱 가속화할 것이다. 아젠다별 개방과 동시에 공공분야 전체를 대상으로 단계별 공공데이터 개방 범위 목표를 수립하고 추진해야 한다. 뿐만 아니라 민간 영역과의 협력을 통해 다양한 비즈니스를 개발함으로써 창조적 비즈니스의 방향성을 제시하는 혁신적 목표도 함께 수립해야 한다.

이에 발맞추어 데이터를 생산·관리하는 오픈 플랫폼 기반의 표준화도 병행하여 추진해야 한다. 이러한 아젠다별 데이터 개방 목표 관리, 공동 전략 추진, 다양한 창조경제 활용 비즈니스 전략 개발을 위한 담당 전문기구를 설립하고 철저한 성과 관리와 혁신적 아이디어 발굴이 필요하다.

2. 창조경제 인프라 구축

교통정보를 기반으로 비즈니스를 확장하고 있는 미국의 교통정보 제공기업 인릭스(INRIX) 사례에서 알 수 있듯이 데이터 인프라를 기반으로 다양한 데이터를 생산하고 이를 통해 새로운 비즈니스로 확장할 수 있다. 창조경제를 위해 공공데이터 자체를 개방하는 것 뿐만 아니라 데이터 생산 및 수집을 위한 인프라를 제공해 민간에서 필요로 하는 데이터를 생산할 수 있도록 지원하는 것도 중요한 방향이다. 데이터 개방과 인프라 구축 뿐만 아니라 데이터 분석을 통한 비즈니스의 새로운 방향성 제시와 구체적 사업모델 개발 지원도 필요하다.

데이터 분석을 통해 정부 스스로 과학적인 의사결정에 적용하고 새로운 일자리 창출에 활용할 수 있을 뿐만 아니라 예산, 인력, 기술 등의 문제로 인해 민간에서 불가능한 대용량 데이터 분석을 지원해 줌으로써 새로운 창조적 비즈니스의 방향성을 수립하는 데 도움을 줄 수 있다. 그러나 이러한 데이터 기반 창조경제 구현의 밑바탕에는 데이터를 활용한 스타트업(start-up)이 활성화될 수 있는 바람직한 스타트업 생태계 조성이 급선무이다.

공공데이터 개방과 빅데이터에 기반한 생태계 구축, 관련 산업 활성화는 창조경제 활성화의 필수불가결한 요소이다. 빅데이터는 데이터를 기반으로 한 새로운 혁신 동력이자 일자리 창출의 원동력이며, 창조경제 실현의 핵심이다. 이러한 데이터는 민간과 공공분야에 산재되어 있지만 민간데이터 개방 및 활용은 어려운 문제이기에 우선 공공분야의 데이터 공개를 통해 관련 산업 활성화를 유도해야 한다. 그리고 이를 기반으로

공공·민간데이터 개방을 통한 생태계 구축과 창조경제 활성화가 필요하다. 이러한 데이터 개방을 촉진하고 효과적으로 추진하기 위해 기술적 측면에서는 개방 플랫폼을 개발, 확산, 보급할 필요가 있다. 개방 플랫폼은 공공데이터를 자유롭게 공개하고 활용할 수 있는 기술적인 장을 마련해 주는 것이며 이는 법과 제도적인 작업에 앞서 적극적인 공공데이터 개방을 물리적으로 촉진할 수 있을 것이다. 또한 플랫폼을 통해 사용된 데이터로 발생된 서비스를 공유하고 이용자들의 자발적인 데이터 업데이트가 가능해짐으로써 창조경제 실현을 유도할 수 있을 것이다. 그리고 무엇보다 중요한 점은 데이터 개방에 필요한 정부 및 공공기관의 시간·인력·예산의 부담을 덜어줄 수 있다는 것이다.

창조경제 인프라 구축을 위한 핵심 엔진인 공공데이터 개방 플랫폼은 데이터 저장, 관리, 개방, 활용의 모든 단계에 적용되는 서비스로 개발되어야 할 것이다. 다시 말해 데이터는 플랫폼에서 지원하는 서비스에 맞춰 표준화된 형태로 자동으로 생성되며 데이터 생산자가 개방 가능한 데이터를 단순히 업로드 처리하면 관련된 모든 작업이 완료될 수 있도록 개발되어야 하며, 개발자는 검색과 목록 확인을 통해 데이터를 활용하고 일반 이용자는 데이터 포털을 통해 시각화된 서비스를 받을 수 있어야 한다. 이러한 플랫폼은 공공 및 민간에서 별도 비용 없이 이용 가능하도록 오픈소스로 구현하고 이를 개발·보급·지원하기 위한 전담조직의 구성이 필요하다.

3. 빅데이터 시대 : 데이터 자원 확보와 품질관리 방안

현재 빅데이터에 대한 논의는 데이터 자체의 증가 국면과 기존 데이터 처리 기술의 한계에 따른 신기술 등장 이슈에 집중되고 있다. 하지만 주어진 빅데이터를 관리하거나 처리하는 측면과 함께, 활용 측면의 빅데이터 확보에 빅데이터를 확보하는 것에 관심을 갖는 것도 중요하다. 활용할 수 있는 외부 데이터를 발굴하여 적극적으로 확보하고, 내부 자원은 결과의 신뢰성을 위해 품질을 제고해야 한다.

이러한 빅데이터 자원의 활용 촉진 및 가치 극대화를 위해 빅데이터 환경에서 자원을 확보하는 방안은 생성, 수집, 공유와 연계, 협력, 참여, 공개 등에 따라 단계적으로 확장 가능하며 이를 위한 원활한 데이터 확보 환경 구축이 필요하다.

원활한 빅데이터 자원 확보 환경을 구축하기 위해서는 행정 및 공공데이터를 공개하고, 데이터가 필요한 조직 및 개인이 접근하여 데이터를 수집할 수 있는 공공데이터 허브와 같은 체계 구축이 필요하다. 또한 국가기밀, 개인정보 등 데이터 성격에 맞는 데이터 등급 제도 및 공익적, 개인적, 사업적 활용에 대한 지침 마련을 포함한 데이터 공개 가이드라인 제정이 시급하다. 빅데이터 활용 결과의 정확성 및 신뢰성 향상을 위해 빅데이터 품질관리 체계 구축도 필요하다. 기업의 정형 데이터, 공공기관이 보유한 공공데이터 등은 개별 데이터에 대한 품질관리를 통해 데이터의 중복성, 불일치성 등을 관리해야 하며 다양한 데이터 소스를 활용하는 빅데이터의 특성을 고려한 종합적인 빅데이터 품질관리 가이드라인 마련이 필요하다. 아울러 빅데이터

품질관리 가이드라인을 준수한 데이터에 부여하는 ‘빅데이터 품질 라이선스’와 데이터 공개 가이드라인을 준수한 데이터에 부여하는 ‘데이터 공유 라이선스’ 등 품질을 보장하고 활용을 극대화하기 위한 빅데이터 자원 품질 인증 방안 연구가 필요하다.

끝으로 데이터 자원 확보와 품질관리를 위해서는 국가, 개인, 기업 등 데이터 생성에 참여하는 주체들 간 지식·자원의 결합과 협력 촉진을 위한 신뢰 기반 형성이 무엇보다 중요하며, 데이터의 안전한 사용을 보장하고 데이터의 신뢰성을 높이기 위한 법적·사회적 측면의 환경 조성이 필수적이다.