

# 일본의 빅 데이터 동향<sup>1)</sup>

## Current Status of Big Data in Japan

송태민 한국보건사회연구원 연구위원

### 1. 서론

최근 빅 데이터가 주목을 받고 그 활용에 대한 기대감이 높아지게 된 것은 종래의 데이터베이스 시스템이 단지 거대해졌기 때문이 아니다. 인터넷과 웹이 사회전반에서 널리 보급되어 일상 생활속에서 활발하게 사용됨에 따라 종래의 데이터베이스시스템에는 포함되지 않았던 비정형적인 구조를 가진 방대한 양의 데이터가 발생하고, 이를 폭넓고 심도있게 그리고 신속히 분석하여 활용하는 기술이 개발 된 것이 계기가 되었다. 사회적 또는 경제적인 문제를 해결하거나 업무의 새로운 가치를 창조하는데 있어 빅 데이터가 담당할 수 있는 역할은 크게 세가지로 꼽을 수 있다. 첫 번째는 시스템 로그 및 축적된 데이터에서 이변을 감지하여 위협의 발생 가능성과 영향도를 측정하여 위협을 방지하거나 최

소화한다는 것이다. 이는 비즈니스의 분야에서 뿐만 아니라 사회적 위협에도 적용할 수 있다. 두 번째는 소셜미디어에서 발생된 방대한 의견을 분석함으로써 최신 고객의 요구와 시장동향을 파악하고 기존 사업에서의 보다 앞선 대응과 신규사업의 개발에 연결하는 것이다. 세 번째는 도시계획에 있어 데이터의 분석과 사회문제에 대한 시뮬레이션을 수행하여 최적의 사회환경을 실현하는 것이다. 예를 들어 센서 데이터를 활용하여 에너지효율을 높일 수 있으며 교통상황의 실시간 분석을 바탕으로 특정지역에 대해 피크 시간대의 교통혼잡료를 부과함으로써 정체완화를 도모하는 것 등이다. 이러한 점에서 빅 데이터 활용의 이점은 보다 빠르고 상세하게 한층 다면적으로 데이터를 분석함으로써 최적의 의사결정에 기여하는데 있다고 할 수 있다. 그리고 빅 데이터의 의미는 지금까지 해결하지

1) 본 내용은 '일본의 빅 데이터 동향 분석'을 위해 '2013. 7. 15~ 7. 19' 기간 동안 일본 동경 출장지(동경공업대학, 노무라연구소, HITACHI 연구소)에서 수집된 자료(번역자: 동경공업대학 이종순교수)를 분석한 것임을 밝힘.

못한 사회적 과제를 극복하고 더 나은 사회를 형성하며, 지금까지 불가능했던 것을 가능하게 함으로써 새로운 비즈니스의 기회를 낳는 것으로 그 가치를 정리할 수 있다. 본 고에서는 우리나라와 경제적, 사회적 현상이 유사한 일본의 빅 데이터 동향에 대해 알아보고 시사점을 논의코자 한다.

## 2. 일본의 빅 데이터 가치

빅 데이터 활용에 의한 효과는 다음의 3가지 면에서 크게 구분할 수 있다. 업무의 효율화와 부가가치창출, 개인정보를 기반으로 한 시장의 창출, 데이터 활용에 따른 새로운 ICT 기술 솔루션의 창출이다. 첫째, 빅 데이터 이용에 의한 업무 효율화와 부가가치 창출에 관해서는 헬스케어 분야에서의 의료비 최적화, 행정분야의 업무효율화 및 사회보장 혜택의 시정에 의한 조세증수, 소매분야에서의 이익 증가, 제조분야에서의 개발비용 절감, 교통분야에서 프로브(Probe) 교통정보 도입에 의한 정체해소 효과 등으로 약 총 10 조엔 이상이 창출되는 것으로 보고하고 있다<sup>2)</sup>(총무성, 2012). 둘째, 개인정보를 기반으로 한 시장의 창출로 시장규모가 전체적으로 약 1.2조엔에 달하며<sup>3)</sup>(일본정보처리개발협회, 2012) 그중에서도 특히 유망한 분야가 금융 및

소매업으로 전망되고 있다. 금융 분야의 대표적인 서비스로는 신용위험을 분산하기 위한 데이터 분석, 금융상품의 익명화 무역정보 활용서비스 등이 있다. 또한 소매분야의 대표적인 서비스로 효율적인 판매촉진을 위한 고객행동패턴이나 구매행동 분석, 실제매장의 POS정보를 가상점포 소매업에 제공 등의 서비스이다. 셋째, 빅 데이터 활용에 따른 새로운 ICT 기술 솔루션 창출에 관해서는 데이터 수집과 관련된 M2M (Machine to Machine), 정보관리와 관련한 클라우드 서비스, 스토리지 소프트웨어, 데이터 분석과 관련한 비즈니스 인텔리전스툴과 같은 사업에서 2020년까지 약 5.4조원의 사업규모가 될 것으로 예상하고 있다(총무성, 2012). 이와 같이 일본의 빅 데이터의 활용이 촉진되면 부가 가치의 창출이나 사회적 비용의 절감에서 총 16조원 이상의 경제적인 효과가 얻어질 것으로 예상되어 기대가 모아지고 있다. 세계에서 일본의 국내 총 생산이 점유하는 비율이 오랜 기간 하락하고 있는 상황에서 빅 데이터의 활용은 일본의 경제성장의 기폭제가 될 수 있다고 해도 과언이 아니다. McKinsey Global Institute(2011)<sup>4)</sup>에 따르면 일본의 정보축적량은 북미의 약 9분의 1, 유럽의 약 5분의 1 수준이다. ICT 기업경영 측면에서의 활용은 일본은 세계 주요국에 비해 뒤처지고 있는 실정이지만, 향후 방대한 데이터를 최대한 활용해 나가는 것은 상기의 추산

2) 総務省2012'平成 24 年度版 年度版 情報通信白書

3) 独立行政法人 情報処理推進機構(2012, 8), くらしと経済の基盤としてのITを考える研究会報告書, つながるITがもたらす豊かなくらしと経済, p.26.

4) McKinsey Global Institute(2011). Big Data: The Next Frontier for Innovation, for Innovation, Competition, and Productivity, McKinsey Inc.

에서 보는 바와 같이 상당한 경제적인 가치를 낳는다는 점에서 유익하다고 할 수 있다. 또한 일본은 현재 세계에 유례가 드문 저출산과 초고령화 사회로 진입하고 있고 동시에 다양한 사회문제들도 가지고 있어 앞으로 변화가 없이는 더 이상 지속가능한 사회를 유지할 수 없는 상황에 몰려있다. 이러한 상황에서 비약적으로 증가하는 정보를 충분히 활용하는 것이야말로 혁신·창출과 그에 따른 사회적 문제해결에 크게 기여하게 될 것이라 보인다. 따라서 사회구조 개혁을 위한 빅 데이터 활용방안을 조속히 검토해 나가고 있다.

### 3. 일본의 빅 데이터 기술 동향

#### 1) 대규모 처리에 있어서의 이론적 기술

빅 데이터의 대규모 처리를 지원하는데 있어서 현재의 WSC(Warehouse Scale Computer)을 가지고 지금까지의 계산 모델과 같이 각 계산의 순간, 순간에서 일체의 일관성을 유지하고 엄밀하게 처리하는 방식으로는 빠른 주기로 시스템의 스케일을 확장시켜 나갈 수가 없다. 따라서 이를 보완하는 이론으로 CAP정리, 에러망각형 컴퓨팅(Failure-oblivious Computing), 결과적 일관성(Eventual Consistency) 등의 기술이 논의되고 있다. CAP정리는 버클리 대학의 Eric Brewer 교수가 Inktomi의 운영에서 얻은 경험과 통찰력을 바탕으로 얻어진 이론으로, 높은 확장성을 가진 분산컴퓨팅 시스템을 만들기 위해 시스템 요건의 균형을 찾는데 적용되

는 정리이다. CAP은 모든 노드에서 동시에 동일한 데이터를 볼 수 있게 하는 데이터 무결성(Consistency), 장애가 발생한 노드가 다른 정상 노드의 기능에 영향을 주지않는 시스템가용성(Availability), 그리고 통신장애 등에 의해 일부 메시지가 손실되더라도 시스템은 정상작동을 해야 하는 네트워크분단허용성(tolerance to network Partitions)을 의미하는 것이다. 상기의 세가지의 모두를 보장하는 것은 불가능하며 이중 어느 두 가지에 중점을 두고 시스템을 설계하느냐가 중요하다는 것이다. 이 법칙의 예로 네트워크 분단에 대한 허용성(P)을 고려하지 않는 시스템으로서의 클러스터 데이터베이스, 가용성(A)을 고려하지 않는 시스템으로서의 분산 데이터베이스, 일관성(C)을 고려하지 않는 시스템으로서의 DNS를 들고 있다. 에러망각형 컴퓨팅(Failure-oblivious Computing)은 MIT의 Martin Rinard 교수 등이 OSDI(Operating Systems Design and Implementation)의 학회에서 2004 년에 발표한 서버의 가용성을 높이는 개념으로, 일반적으로 오류가 발생하면 데이터 처리를 도중에 중단시키지만 에러망각형 컴퓨팅에서는 처리의 불일치가 발생해도 그 로그만 남기고 처리를 계속한다. 이렇게 하면 서버의 가용성을 높일 수 있을 뿐만 아니라 서버가 공격을 받고 있는 상황에서는 공격자들이 예상한 오류의 반응을 서버가 보여주지 않음으로 해서 보안능력도 향상 할 수 있다. 대규모 데이터처리에 있어서 오류처리를 사후에 실시하여 전체 시스템의 효율성을 유지하는데 중요함을 보여주고 있다. 결과적 일관성(Eventual Consistency)은 시스템 전체에 걸쳐 엄밀하게

일관성을 요구하는 것이 아니라 결과적으로 일관성이 유지되면 되는 데이터에 대해 일관성을 유지하는 모델이다. 많은 관계 데이터베이스에서는 데이터의 무결성을 보장하기 위해 중간 무결성을 유지하기 위한 제약조건이 있다. 그러나 대규모 데이터처리의 서비스에서는 모든 상태에서 일관성을 가지도록 하면 시스템의 확장이 어려워진다. 따라서 DNS와 같이 중간 상태에서 일관성이 결여되어도 결과적으로 유용한 상태가 되면 유용하다고 보는 것이다.

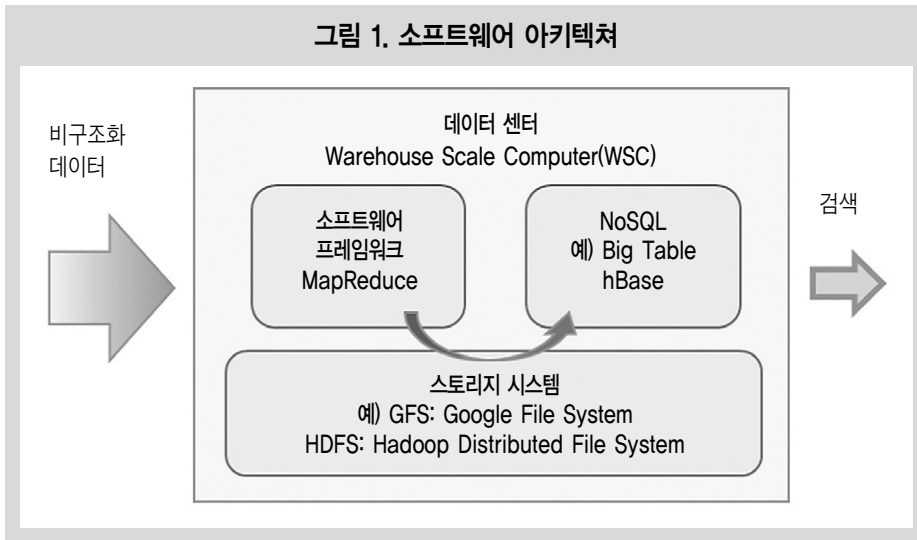
## 2) 하드웨어와 소프트웨어 기술

빅 데이터를 처리하기 위해서는 WSC (Warehouse Scale Computer) 규모의 하드웨어 시스템이 필요하며, CPU 대수와 저장용량이 대규모로 커짐으로 해서 지금까지 드러나지 않았던 전력문제와 신뢰성의 문제가 표면화되게 된다. 네트워크에 대해서도 대용량 전송이 요구되지만, 빅 데이터 수준의 것을 간단히 전송하기는 쉬운일이 아니다. 첫째, CPU 처리 능력으로 2010년 현재 Google의 SOCC(Symposium of Cloud Computing) 데이터처리 용량은 하루 24 페타바이트 정도로 보고 있다(Google, 2010). 현재 생성되는 데이터의 80%가 영상이미지이거나 센서 등에서 수집되는 것이라 보고 있어 이러한 비정형 데이터의 변환이 필요하며, 빅 데이터의 특성에 따른 빠른 컴퓨팅 기술이 필요하다. 둘째, 스토리지 능력으로 빅 데이터를 유지

하기 위해 스토리지는 전체 용량도 중요하지만 견고성과 가용성이 중요하다. 여기에는 하드웨어 장애뿐만이 아니라 네트워크 장애로 인한 지역 분단 등에 대해 글로벌한 대처가 필요하다. 소프트웨어 이중화는 파일시스템 등의 기술로 이루어지고 있지만, 지리적 분산에 의한 견고성과 가용성은 하드웨어(데이터센터) 수준에서 이루어지고 있다. 셋째, 데이터 전송능력으로 빅 데이터를 수집하기 위해 인터넷을 사용하는 것은 필수가 되고 있다. 인터넷 자체는 공유 인프라로서 어느 기업이나 국가가 독점하는 것이 아니지만, 2010년의 시점에 인터넷 트래픽의 최대 12%가 YouTube를 포함한 Google사의 것으로 보고되고 있다. 인터넷에서 대용량의 데이터를 수집하고 있는 것은 사실이지만, 인터넷은 그 구조상 하나의 사이트에서 대량의 데이터를 보내는 데는 적합하지 않다. 넷째, 소프트웨어 아키텍처 기술능력으로 빅 데이터를 처리하는 소프트웨어는 요구되는 처리의 특성을 살리면서, 데이터웨어하우스 규모의 컴퓨터를 충분히 활용할 수 있는 형태로 구성되며 구조화되지 않은 데이터가 검색되도록 처리하는 과정을 나타낸다. 여기에는 대용량 데이터를 지원하는 분산 스토리지 시스템, 그위에서 데이터처리를 담당하는 소프트웨어 프레임워크인 MapReduce, 데이터 검색을 다루기 위한 NoSQL 데이터베이스가 포함되어 있다(그림 1)<sup>5)</sup>.

5) Dean, Jeffrey and Ghemawat, Sanjay. "MapReduce: simplified data processing on large clusters." Communications of the ACM, Volume 51, Issue 1, January 2008

그림 1. 소프트웨어 아키텍처



다섯째, 비 구조화 데이터를 구조화 데이터로 바꾸는 기술 능력으로, 빅 데이터에는 매출정보와 고객정보 등과 같이 정형화된 포맷에 따라 표형식으로 집약된 구조화 데이터와 자연적인 문장의 텍스트, 영상 이미지 및 음향 등 비구조화 데이터가 있다. 일반적으로 빅 데이터의 80%는 비 정형데이터로 추정되고 있으며, 이러한 다양한 형태의 데이터를 구조화된 데이터로 바꾸는데 있어 음성인식, 로그분석기술, 네트워크 기술, 패턴인식기술, 그리고 데이터마이닝의 기술이 주목을 받고 있다.

마지막으로 보안기술 능력으로 악의를 가진 공격자가 공개된 정보에서 개인정보를 추론하는 것을 방지하는 기술인 PPDM(Privacy-Preserving Data Mining) 혹은 PPDP(Privacy-Preserving Data Publishing)이 최근 활발히 연구되고 있다. 여기서는 사용자의 의도하지 않은 암호의 노출 문제와 정보의 유출, 손실, 삭제오

류 등과 같이 시스템의 입장에서의 보안문제를 다룬다. 암호의 노출문제에 있어서는 빅 데이터 환경에서 인터넷에 올려진 데이터는 무엇이든 기본적으로 아무에게나 사용되어질 수 있다고 볼 수 있다. 웹에 올라온 모든 데이터를 저장하는 기술도 개발되고 있어, 일단 공개된 정보는 영원히 어딘가에 남아 있을 위험이 있다. 따라서 중요한 데이터를 은닉하는 기술의 요구는 지속적으로 증가하고 있다. 데이터를 은닉하는 경우에도 여기에 사용되는 암호의 강도에 주의할 필요가 있다. 일본 국내의 전자정부에서 권장하는 암호의 안전성을 평가하는 CRYPTREC는 슈퍼컴퓨터의 성능향상과 암호노출의 관계를 나타낸 그래프를 공개하고, 1024 비트 RSA의 소인수 분해 문제는 2015년에서 2020년경에 풀릴 것으로 예측하고 있다. 따라서 암호키를 2048 비트 방식으로 변경하기를 권장하고 있지만, 컴퓨팅리소스에 제한이 있는 기기에서

는 사용하기 어려운 것이 실상이다. 서버로 부터의 정보유출은 오래된 중요한 문제이지만 여전히 해결되지 않고 있다. 특히 관리자에 의한 악의적인 유출은 방지가 어렵다. 보안에 대해 매우 엄격한 것으로 여겨지고 있는 Google에서도 2010년 간부 엔지니어가 Gmail의 내용을 도청하여 악용한 사건이 발생했다. 또한 악의가 없다하더라도 설정 실수나 버그에 의한 정보유출의 우려도 있다. 정보손실은 본래 저장되어 있어야 할 데이터가 서버에서 사라지는 문제이다. 이 사례로는 개인정보를 기억하는 서비스의 Evemote에서 6,323명의 신규로 편집한 데이터가 손실된 사건이 있었다. 이것은 1대의 서버에서 발생한 장애가 원인인 것으로 보고되었다. 삭제오류는 정보손실의 반대로 사용자가 데이터를 삭제했음에도 불구하고 서버에 남아 있는 문제이다. 예를 들어 Facebook이나 Google+에서 사용자가 이미지 데이터를 지웠음에도 불구하고, 서버에서 사라지지 않아 문제가 지적되었다. 이러한 이미지 데이터는 보통의 URL로 누구나 접근할 수 있다는 더 큰 문제의 소지가 있는 것이다. 이러한 문제로 인하여 정보유출, 손실, 삭제오류는 시스템 장애, 소프트웨어 버그 또는 관리자나 사용자의 악의적인 행위에 의해 여러 형태로 발생하기 때문에 현재로서 근본적인 해결방법이 없어 제 삼자 기관의 감사에 의한 지적과 각종 보안정보를 수집하고 문제가 있으면 대처해 가는 평소의 관리가 매우 중요하다.

## 4. 일본의 빅 데이터 추진전략

일본 정부의 빅 데이터 추진전략은 2013년 6월 고도 정보통신 네트워크 사회추진 전략본부(약칭 IT종합전략본부(본부장: 수상))가 발표한 「세계 최첨단 IT 국가 창조 선언」(이하 ‘창조적 선언’)의 공정표에 잘 나타나 있다. IT종합 전략 본부는 IT 정보자원을 활용하여 미래를 창조하는 국가비전으로서 ‘창조적선언’(2013년6월 14일 내각 결정)을 수립했다. 이 선언에는 향후 2020년까지 세계 최고수준의 IT활용 사회를 실현하는 것을 목표로 ① 혁신적인 신산업·신서비스의 창출과 전체 산업의 성장을 촉진하는 사회의 실현, ② 국민이 건강하고 안심하고 쾌적하게 생활하는 세계에서 가장 안전하고 재해에 강한 사회의 실현, ③ 공공서비스를 누구나 언제 어디서나 원스톱으로 받을 수 있는 사회의 실현의 3항목에 대한 지향해야 할 사회모습을 제시하였다. 이 공정표는 어느 부처가 언제까지 구체적으로 무엇을 실시하는지를 밝히고, 각 부처 간 연계가 필요한 시책에 대해서는 개별 역할분담과 달성해야 할 사항을 명확히 하여 꾸준히 구체적인 성과로 연결시키는 것을 목적으로 책정한 것이다. 공정표에는 ‘창조적 선언’에 나타난 전략과 목표에 대해 단기, 중기, 장기로 나누어 각 부처가 실시할 시책이 명시되어 있다. 오픈데이터·빅 데이터의 활용의 추진 시책은 상기의 첫째 목표인 ‘혁신적인 신산업·신서비스의 창출과 전체 산업의 성장을 촉진하는 사회의 실현’의 실현사항으로 제시되어 있으며, 다음은 단기, 중기, 장기별로 나누어 추진할 계획이다.

## 1) 단기 계획(2013 ~2015년도)

계획	세부계획
제도정비	<ul style="list-style-type: none"> <li>○ 개인 데이터의 취급에 대해 IT종합 전략 본부 아래에 새로운 검토조직을 설치(6월중)하고 검토를 시작한다(내각관방).</li> <li>○ 개인정보 및 프라이버시 보호를 배려한 개인 데이터의 활용규칙을 명확히 한 후, 개인정보보호 가이드라인의 재검토 동의를 취득 절차의 표준화 등의 작업을 연내의 가능한 빠른 시기에 착수하고 2015년말까지의 검토를 바탕으로 필요에 따라 개인정보보호 가이드라인의 검토 및 동의 취득절차의 표준화 등을 실시한다(내각관방).</li> <li>○ 규제개혁 실시계획에 따라 소비자청은 빅 데이터의 이용에 참고를 목적으로 '개인정보보호법에 대해 자주 묻는 질문과 답변(Q &amp; A)'의 개정을 2013년 상반기에 실시한다. 또한 내각관방과 소비자청은 협력하여 합리적인 익명화 조치의 내용을 명확히 한 가이드라인을 2014년 상반기에 수립한다. 개인정보보호법의 사업 등 분야별 가이드라인 등 소관부처는 합리적인 익명화조치의 내용 사업 등 분야별 가이드라인 등을 2014년 중에 명확히 제시한다(내각관방, 소비자청, 사업분야별 가이드라인의 소관부처).</li> <li>○ 제 삼자기관의 설치를 포함하여 새로운 법적 조치도 염두에 둔 제도 재검토의 방침을 연내에 수립하고, 제도 재검토의 방침에 따라 국제적인 협력도 배려하면서 각 시책을 실시한다(내각관방, 관계부처).</li> <li>○ (선행적 조치의 추진) 선행적으로 규정이 책정된 스마트폰의 이용자 정보의 취급분야에 대해 관계부처, 업계단체, 사업자가 함께 규정의 보급을 추진한다. 또한 2국가 간 혹은 다 국가 간의 회합의 기회를 활용하여 국제적인 정보공유와 협력을 추진한다(총무성, 관계부처).</li> </ul>
활용촉진	<ul style="list-style-type: none"> <li>○ 빅 데이터를 활용한 새로운 비즈니스의 창업에 대한 지원을 실시하여 민간 부문에 있어서의 새로운 부가가치 창출을 추진한다(총무성, 경제산업성)</li> <li>○ 각 분야 (지역활성화, 대중교통, 방재, 의료·건강, 에너지 등)의 실증 프로젝트 등을 통하여 빅 데이터의 활용을 검토하고 새로운 부가가치의 창출을 위한 지식을 얻는다(내각관방, 내각부, 총무성, 후생노동성, 농림수산업성, 경제산업성, 국토교통성, 문부과학성)</li> </ul>
인재육성	<ul style="list-style-type: none"> <li>○ 새로운 서비스, 새로운 사업전략 입안과 신기술의 창출에 빅 데이터를 활용할 수 있는 인재(데이터 사이언티스트 등)의 육성에 착수한다(문부과학성).</li> </ul>
기술개발	<ul style="list-style-type: none"> <li>○ 빅 데이터의 활용을 촉진하기 위해 데이터와 네트워크의 안전성·신뢰성 향상 및 상호 운용성 확보, 대규모 데이터의 축적·처리기술의 고도화 등 공통기술의 조기 확립을 도모함과 동시에 새로운 비즈니스 새로운 서비스의 창출로 이어지는 새로운 데이터 활용기술의 연구개발과 그 활용을 추진한다.</li> <li>○ 구체적으로는 빅 데이터의 활용기술(수집, 전송, 처리, 활용·분석 등)에 대한 각 부처의 역할을 명확히 하고 나아가 각 부처가 연계하여 다른 목적으로 수집된 다양한 데이터에서 유용한 정보·지식을 실시간으로 추출할 수 있는 기초 기술, 광통신 기술(400Gbps급) 네트워크 가상화 기술 등의 연구개발을 실시함과 동시에 정보를 유통 순환시켜, 분야의 경계를 넘어서 정보가 활용됨으로써, 신사업·신서비스의 창출을 촉진한다(총무성, 문부과학성, 경제산업성).</li> </ul>

자료: 高度情報通信ネットワーク社会推進戦略本部, 世界最先端 IT 国家創造宣言 工程表(2013. 6.14)

## 2) 중기 계획(2016~2018 년도)

계획	세부계획
제도정비	<ul style="list-style-type: none"> <li>○ 퍼스널 데이터의 활용 규정을 바탕으로 한 검토에 따라 필요에 의해 개인정보보호 가이드라인을 수정하고 등의 취득 절차의 표준화 등을 실시한다(내각관방, 소비자청, 사업 분야별 가이드라인의 소관부처)</li> <li>○ IT종합 전략본부 아래에 설치된 새로운 검토 조직에서 정리된 제도 재검토 방침에 따라 국제협력도 배려하면서 각 시책을 실시한다(내각관방, 관계부처).</li> </ul>
활용촉진	<ul style="list-style-type: none"> <li>○ 빅 데이터를 활용한 새로운 비즈니스의 창업에 대한 지원을 실시하여 민간 부문에 있어서의 새로운 부가가치 창출을 추진한다(총무성, 경제산업성).</li> <li>○ 각 분야(지역활성화, 대중교통, 방재, 의료·건강, 에너지 등)의 실증 프로젝트 등을 통하여 빅 데이터의 활용을 검토하고 새로운 부가가치의 창출을 위한 지식을 얻는다(내각관방, 내각부, 총무성, 후생노동성, 농림수산성, 경제산업성, 국토교통성, 문부과학성).</li> </ul>
인재육성	<ul style="list-style-type: none"> <li>○ 새로운 서비스, 새로운 사업전략 입안과 신기술의 창출에 빅 데이터를 활용할 수 있는 인재(데이터 사이언티스트 등)의 육성에 착수한다(문부과학성).</li> </ul>
기술개발	<ul style="list-style-type: none"> <li>○ 빅 데이터의 활용을 촉진하기 위해 데이터와 네트워크의 안전성·신뢰성 향상 및 상호운용성 확보, 대규모 데이터의 축적·처리기술의 고도화 등 공통기술의 조기 확립을 도모함과 동시에 새로운 비즈니스 새로운 서비스의 창출로 이어지는 새로운 데이터 활용기술의 연구개발과 그 활용을 추진한다.</li> <li>○ 구체적으로는 데이터 활용 기술(수집·전송, 처리, 활용·분석 등)에 대한 각 부처의 역할을 명확히 하고 나아가 각 부처가 연계하여 다른 목적으로 수집된 다양한 데이터에서 유용한 정보·지식을 실시간으로 추출할 수 있는 응용 기술을 확립하고 구축된 기술의 시장전개를 도모하면서 국제 표준화 기구에 제안하여 일본의 개발 기술확립을 추진함과 동시에 정보를 유통 순환시켜, 분야의 경계를 넘어서 정보가 활용됨으로써, 신사업·신서비스의 창출을 촉진한다(총무성, 문부과학성, 경제산업성)</li> </ul>

자료: 高度情報通信ネットワーク社会推進戦略本部, 世界最先端 IT 国家創造宣言 工程表(2013, 6.14)



### 3) 장기 계획( 2019~2021년도)

계획	세부계획
제도정비	○ IT종합 전략본부 아래에 설치된 새로운 검토 조직에서 정리된 제도 재검토의 방침에 따라 국제협력도 배려하면서 각 시책을 실시한다(내각관방, 관계부처).
활용촉진	○ 각 분야(지역 활성화, 대중교통, 방재, 의료·건강, 에너지 등)의 실증 프로젝트 등의 성과를 살려, 새로운 서비스의 창출을 촉진한다(내각관방, 내각부, 총무성, 후생노동성, 농림수산업성, 경제산업성, 국토교통성, 문부과학성).
인재육성	○ 새로운 서비스, 새로운 사업전략 입안과 신기술의 창출에 빅 데이터를 활용할수 있는 인재(데이터 사이언티스트 등)의 육성에 착수한다(문부과학성).
기술개발	○ 빅 데이터의 활용을 촉진하기 위해 데이터 및 네트워크의 안전성·신뢰성 향상 및 상호운용성 확보, 대규모 데이터의 축적·처리기술의 고도화 등 공통기술의 조기 확립을 도모함과 동시에 새로운 비즈니스 새로운 서비스의 창출로 이어지는 새로운 데이터 활용기술의 연구개발 및 그 활용을 추진한다. ○ 구체적으로는 데이터 활용기술(수집·전송, 처리, 활용·분석 등)에 대한 각 부처의 역할을 명확히 하고 나아가 각 부처가 연계하여 다른 목적으로수집된 다양한 데이터에서 유용한 정보·지식을 실시간으로 추출할 수 있는 기술을 실용화하고, 정보를 유통 순환시켜, 분야의 경계를 넘어 정보가 활용됨으로써, 신사업·신서비스의 창출을 촉진한다(총무성, 문부과학성, 경제산업성)

자료: 高度情報通信ネットワーク社会推進戦略本部, 世界最先端 IT 国家創造宣言 工程表(2013. 6.14)

## 5. 일본의 빅 데이터 활용 사례 및 동향

### 1) 일본의 빅 데이터 활용 사례

혼다자동차를 생산하는 혼다기연공업은 드라이버의 쾌적한 카라이프를 실현하기 위한 드라이브 정보 네트워크로서 'intemavi' 서비스를 2002년부터 제공하고 있다. 안전, 안심, 방재, 날씨, 연료절약 루트 등의 정보를 제공하여 2012년 5월 현재 회원수 145만명을 확보하고 있다. 이 서비스는 2003년 회원으로부터 매 5분 간격으로 수집된 장착차량의 주행데이터(Floating Car Data: FCD)를 공유함으로써 교통체증을 피

하고 목적지까지 보다 빠른 경로를 안내하는 'Floating car system'을 도입하고 있다. 이를 통해 주요 간선을 대상으로 하는 VICS(Vehicle Information & Communication System)가 보완되고 있으며 현재 매일 1억km의 데이터가 업로드 되어, 2012년 5월말 현재 누적 주행데이터는 25억km에 달한다.

본 시스템의 효과로는 ① VICS와 FCD를 융합한 정체예측의 효과를 검증한 결과, 동경-오사카 간의 고속도로의 100샘플의 시뮬레이션에서 약 20%의 빠른 경로를 안내하고, CO2로 환산하여 약 16%의 절감효과가 있었다. ② 사이타마현의 도로관리 행정업무에 정보를 제공하여, FCD로부터 급브레이크의 다발지점을 추

출하여 그 지점에서의 가로수 정리와 도로면의 표식을 정비함으로써 급브레이크횟수가 약 70% 감소되었다. ③ 2006년의 NPO법인 방위 추진 기구의 연구에 협력하여 2007년 니가타현의 나카고에서 발생한 지진, 2008년 이와테미야기 내륙의 지진에서 FCD를 활용한 통행실적 맵이 작성되어 공개되었다. 2011년 3월 동일본 대지진에서 동일한 맵을 다음날 아침에 공개함과 동시에 Google과 NPO 법인 ITS-Japan을 비롯한 각종 행정기관이나 연구기관에 제공하여 활용되었다. ④ 목적지의 주행경로에 대해 일본 기상협회에서 제공되는 동결예측 등의 기상정보와 연계하여 노면 동결 발생의 예측시간과 예측시점 등을 카내비게이션 화면에 표시하고 음성경고 등으로 기상재해 방지정보를 제공하고 있다.

동일본 철도회사(JR East Japan)의 자판기 활용사례에는 동일본 철도회사의 자회사인 'JR 동일본 워터 비즈니스' 사가 2010년 8월부터 전개하고 있는 청량 음료의 자동판매기에는 보통 상품 견본을 두는 곳에 대형 디스플레이를 설치하여 제품 정보 광고 등의 다양한 정보를 제공하고 있다. 이 자판기의 흥미로운 점은 디스플레이의 바로 위에 카메라를 갖추고 자판기 앞에서 있는 손님의 외모를 인식하는 기능을 가지고 있다는 점이다. 카메라의 이미지를 분석하여 성별, 연령을 추측하고, 그날의 기온이나 날씨에 따라 상품을 추천한다. 자료에 따르면 10대 이하에서 60대 이상까지 연령대와 성별의 14패턴의 특성을 판별하는 것으로 되어 있다.

기상 정보 서비스의 활용사례에서는 모바일 단말을 가진 사용자가 실제 공간에 일어나는

상황을 위치정보와 함께 발신함으로써 이용자 자신이 '센서'로서의 역할을 할 수 있다. 하드웨어 센서를 여러 곳에 설치한다는 것은 유지보수에 큰 비용이 들며, 그렇다고 필요에 따라 인원을 파견하여 수동으로 상황과 시간장소를 기록하게 하는 것은 비용과 효율면에서 더욱 현실적이지 못하다. 서비스 이용자가 자신의 위치에서 어느 순간의 사진을 시간과 장소 정보와 함께 제공한다면 '센서'에서 획득한 정보 이상으로 매우 유용하게 이용될 수 있을 것이다. 이용자가 발신한 정보를 활용하고 있는 예로는 웨더뉴스(<http://weathernews.jp/gensai/>)가 제공하는 기상 콘텐츠가 있다. 서비스 이용자가 '웨더리포터'가 되어 자신의 위치에서 사진을 찍어 위치 정보와 함께 코멘트를 전송하면 홈페이지의 지도 상에 게시되어 실시간의 기상 정보가 표시되는 것이다. 현재 웨더리포터는 유료회원 중에서 30만명으로 구성되어 있으며, 그 중에서 특히 8만명이 '게릴라 뇌우 방어대원'으로 활동하고 있다. 이용자가 보내 온 정보에는 사진과 함께 '습한 바람이 불기 시작했다'라든가 혹은 '갑자기 비구름이 몰려오고 있다'라든가 혹은 '체감온도'와 같이 인간의 오감으로 느낀 정보가 코멘트로 들어 있어 종래의 관측에서 얻어 내기 어려운 내용까지 실시간으로 제공되고 있다. 특히 예측이 어려운 게릴라성 폭우 정보의 제공에 효과를 발휘하여, 도쿄에서는 90% 이상의 확률로 게릴라성 폭우가 발생되기 42.4분 전에 회원들에게 메일로 전달되고 있다. 이렇게 하여 얻어진 데이터는 방재 데이터베이스에 축적되고 있다. 리포터의 리포터 작성은 1일 3회까지 제한되어 있으며,

리포터가 제공한 과거 정보를 리포터 단위로 정리해서 볼 수 없도록 하여 이용자가 특정 리포터의 이동을 추적할 수 없게 했다. 리포터는 유료회원이며 서비스 공급자 측면에서는 동일 인물로 식별되고 있지만, 이용자 간에는 공개하지 않고 링크가 불가능한 상태로 하여 데이터의 식별과 익명성을 양립하여 프라이버시를 보호하고 있다.

## 2) 일본의 빅 데이터 이용 동향

일본 국내에서의 빅 데이터 이용은 크게 IT 벤더에 의한 서비스, 플랫폼 사업자에 의한 서비스, 자사 내에서의 데이터 이용, 데이터분석

지원, 데이터의 사외에서의 2차 이용으로 대분하고 있다.

### (1) IT 벤더에 의한 서비스

국내외를 막론하고 모든 IT업체가 빅 데이터의 활용과 지원서비스에 주력하고 있다. 미국의 IT벤더는 2009년 이후 빅 데이터 관련 기업을 인수 합병하거나 제품과 서비스를 계속해서 투입하고 있다. 일본 국내에서도 2010년 이후 주요 IT사업자가 Hadoop의 인테그레이션 사업을 전개하고 있고, 복합 이벤트처리 기술과 강화된 탐색 기술의 개발에 경쟁을 벌이고 있다.

그림 2. 웨더 리포트(<http://weathernews.jp/report>)

The screenshot shows the 'Weather Report' page from weathernews.jp. It features a map of Japan with weather icons indicating conditions like '晴れ' (clear), '曇り' (cloudy), '雨' (rain), '雪' (snow), '台風' (typhoon), and '地震' (earthquake). A sidebar on the right lists various regions (prefectures) such as 北海道, 東北, 関東, etc. The page also includes a 'Topics' section with '大地を濡す雨' and '西日本は大雨警戒'.

표 1. IT 벤더에 의한 서비스 이용 동향

회사명	시기	주요사책
히타치제작소	2011년 11월	클라우드 컴퓨팅사업 'HarmoniousCloud' 서비스에 '빅 데이터 활용 서비스'와 '스마트 인프라 서비스'의 서비스 메뉴를 추가했다.
일본IBM	2011년 11월	일본기업의 글로벌 진출과 현지 비즈니스 강화를 지원하기 위해, 고객의 방대한 정보를 경영전략에 활용하기 위한 기반을 제공하는 클라우드 서비스를 개시했다. 분석대상은 중국어, 영어, 스페인어, 불어, 독일어 등 11개 언어에 대응
NEC	2011년 11월	센서 등에서 수집 빅 데이터를 실시간으로 분석하여 사용자에게 제공함과 동시에 분석 시스템에서 사용하는 분산 스토리지의 소비전력을 기존대비 2/3로 줄일 빅 데이터 처리 기반기술을 개발했다.
일본HP	2011년 12월	기업의 빅 데이터 활용을 위한 IT 인프라 구축을 지원하는 솔루션인 'HPHadoop HBase서비스'를 발표했다. 이 솔루션은 'Hadoop'에 각종 서비스를 원스톱으로 제공한다.
후지쯔	2012년 1월	빅 데이터를 활용하기 위한 기반 클라우드서비스인 '데이터활용기반 서비스'를 제공하기 시작했다. 대량의 센싱 데이터를 수집, 축적, 통합하고 실시간으로 처리하고 일괄처리에 의한 장래예측 등을 실시하는 '정보관리 통합서비스', '통신제어 서비스' 등을 제공한다.
NTT데이터	2012년 1월	비 정형데이터를 활용하는 분석솔루션을 제공하기 위하여, 미국 MarkLogic사와의 협력에 합의했다. 이 제휴로, 비 정형데이터 분석에 효과적인 '네이티브형 XML 데이터베이스'를 활용한 빅 데이터 활용 솔루션을 제공할 수 있게 되었다

자료: 独立行政法人 情報処理推進機構(2012, 8), くらしと經濟の基盤としてのITを考える研究会報告書, つながるITがもたらす豊かなくらしと經濟, p.136.

## (2) 플랫폼 사업자에 의한 서비스

모바일 통신 사업자를 대표하는 텔레콤 분야와 인터넷 접속 사업 분야에서는 통신 로그와 웹 액세스 로그와 같은 빅 데이터가 대량으로 발생하고 있다. 특히 모바일 통신사는 스마트폰의 급속한 보급으로 위치 정보를 비롯한 라이프 로그(소비자의 행동 이력을 포함한 데이터)를 분석하는 수요가 급속히 늘어나 이에 대한 분석을 시작하고 있다. 모바일 통신사업자 이외에도 플랫폼 사업을 하는 업체가 빅 데이터를 수집 활용하려는 움직임이 있다. 예를 들어 '츠타야'

를 비롯한 엔터테인먼트 사업과 'T포인트' 사업을 운영하고 있는 '컬처컨비니언스클럽'은 약 4,500만명의 회원 데이터베이스를 기반으로, 'T 회원'의 데이터와 편의점 체인 '로손'과의 공동 포인트 프로그램인 'Ponta'의 회원 데이터를 마케팅에 활용하는 시도를 하고 있다. '리크루트' 사도 전사적으로 100대의 Hadoop서버를 가동하여 자사의 데이터를 분석하려고 하고 있다. 여기에는 자사의 주택정보지 '스모'의 약 200만건의 정적인 데이터뿐만 아니라 트위터와 같은 동적인 데이터에 대해서도 실시간으로 처리 분석하고 있다.

### (3) 자사 내에서의 빅 데이터 이용

축적된 데이터를 자사의 업무에 활용하는 이른바 '유저 기업'의 빅 데이터의 활용에는 ① 제품 개발, 업무 개선, 판매 촉진, 예상 고객의 획득 ② 성능 모니터링, 품질관리, 이상 감지 ③ 유지보수 지원의 강화, 고장이나 피해 예방 등의 목적이 있다. 'JR 동일본 워터비즈니스'는 자사가 전개하는 자동판매기를 지능화하고 고객정보의 분석에서 매출을 올림과 동시에 고객의 행동에서 패트병의 '떨어지지 않는 캡'을 장착한 제품을 판매하기 시작했다. 일식 레스토랑 체인인 '간코우(고집쟁이)푸드 서비스'는 점원에 센서를 장착하고 오랜 전통을 지낸 온천지역인 '기노사키 온센'은 IC카드로 된 '육탕 권'을 발행하는 등 점원과 이용자의 행동을 수집하여 업무개선과 판매 촉진에 활용하고 있다. '히타치 플랜트 테크놀러지'는 크레인의 가동 상황을 모니터링하고 이상을 감지하면 즉시 통보하는 서비스를 제공하고 있고, '후지제록스시즈오카'는 복합 복사기에 장착된 센서를 통해 기기의 문제 발생 데이터를 수집 분석하고 그 결과를 현장의 기술자에게 피드백하고 있다.

### (4) 데이터 분석 지원

기업의 데이터 활용을 위한 분석 업무를 지원하는 사업도 이미 시작되고 있다. '미츠이정보'는 SAP재팬이 제공하는 인메모리 소프트웨어 'SAPHANA'를 이용한 데이터 분석의 실증 실험에 착수했다. 그 첫째 작업으로 암 연구에서 게놈 분석과 약물 개발의 프로세서에 있어서 화

합물의 데이터 해석에 빅 데이터 분석을 이용하는 것이다. 또한 고객이 경영전략을 수립하는 프로세스를 지원하기 위해, SAPHANA를 데이터 기반으로 한 실시간 정보 분석방법을 확립하여 빅 데이터분석에 적용하고 있다. 브레인패드는 빅 데이터 분석 소프트웨어로 급성장한 회사로 대기업을 중심으로 99개사에 실적을 가지고 있다. 웹 사이트 방문자의 성별이나 연령대 등의 특성이나 검색기록 데이터 등을 바탕으로 기업의 사이트가 수익에 기여하고 있는지 등을 검증하고 이를 효과적인 사이트 구축과 인터넷 광고 제작에 활용하고 있다.

### (5) 빅 데이터의 외부 재이용

수집된 데이터를 분석하여 이를 이차적으로 이용하기 위해 제공하는 움직임도 있다. 주택개발 사업의 마케팅 전문회사인 아트랙터스 연구소는 인터넷 상에 공개된 국내 부동산 리스트를 크롤러에서 자동으로 수집 축적하고 고객이 요구하는 정형과 비정형의 분석 요구에 응하여 ASP/클라우드 서비스의 형태로 제공하고 있다. 스포츠 데이터를 분석하여 발신하고 있는 '데이터스타디움'은 야구, 축구, 럭비, 농구의 경기를 데이터화하여 신문사 등의 매스컴에 제공하고 팀의 경기력 강화를 위한 컨설팅 사업도 전개하고 있다.

## 3) 일본의 빅 데이터 활용 조사 결과

일본 IPA(독립행정법인 정보처리 추진기구)는 일본 빅 데이터 이용동향 조사의 일환으로

웹 설문조사의 방식에 의해 IT벤더가 아닌 일반 기업들을 대상으로 빅 데이터를 활용하는데 대한 인식조사를 실시하였다. 조사개요는 다음과 같다.

- ① 조사목적: 기업의 빅 데이터 이용에 관한 인식조사
  - ② 조사대상: 'ITpro' 독자 중 닛케이 BP 컨설팅이 모니터 관리를 실시하고 있는 9만 7000명
  - ③ 조사시기: 2012년 1월
  - ④ 조사방법 : 웹 설문조사 (닛케이 BP 컨설팅웹조사시스템 'AIDA')
  - ⑤ 유효화 응답수: 4,005명
- <표 2>와 같이 비 IT 벤더기업에서의 빅 데이

터 활용이 현 시점에서는 아직 발전단계인 것으로 보인다. 빅 데이터의 이용목적은 대체로 사업성과에 직접적인 향상이나 개선을 목적으로 한 것이 상위를 차지하고 있다. 그리고 이용 목적으로 하위의 항목에 대해서는 아직 충분한 분석방법이 확립되어 있지 않고 연구단계에 머물러 있기 때문에 이 분야에서 획기적인 분석방법이 개발되면 보다 크게 성장 할 가능성이 있다고 보고 있다.

빅 데이터를 이용하는 있어서 문제가 되는 요인으로는 대량 데이터를 분석하는 인재의 확보가 가장 큰 과제이며 또한 그 필요성이 높게 인식되고 있는 것으로 나타났다. 그리고 사내에 대량 데이터를 처리하는 시스템 기반을 강화할

표 2. 일본의 빅 데이터 동향 조사 결과

비 IT벤더 기업의 빅데이터 대처 상황(n=2,485)	빅 데이터 이용 목적(n=570)
<ul style="list-style-type: none"> <li>• 빅 데이터를 이미 보유하고 업무에 활용하고 있음: 22.9 %</li> <li>• 빅 데이터를 이미 보유하고 있지만 아직 활용하지 못하고 있음: 17.7 %</li> <li>• 빅 데이터가 업무상 발생하고 있지만 보유하거나 축적하지 않음: 4.9%</li> <li>• 빅 데이터라고 부를 만한 데이터양이 발생하지 않음: 42.5 %</li> <li>• 기타 혹은 잘 모름: 각각 0.4%, 11.5%</li> </ul>	<ul style="list-style-type: none"> <li>• 경영전략, 사업전략수립: 44.4 %</li> <li>• 고객서비스개선: 44.4 %</li> <li>• 판매촉진, 예상고객의 획득: 39.3 %</li> <li>• 제품개발: 34.6 %</li> <li>• 회계 업무의 강화: 27.7 %</li> <li>• 의사결정의 자동화, 실시간 처리 촉진: 23.9 %</li> <li>• 보수 유지의 강화, 장애발생 예방: 23.2 %</li> <li>• 기초 연구: 19.5 %</li> <li>• 내부 통제, 부정 방지: 19.1 %</li> <li>• 고객 로열티 향상, 회원탈퇴 방지: 16.5 %</li> <li>• 광고 효과의 측정: 9.1 %</li> <li>• EC 사이트의 동선 분석, 행동 마케팅: 7.5 %</li> <li>• EC 사이트 외의 동선 분석, 행동 마케팅: 6.1 %</li> <li>• 블로그, SNS에 있어서의 평판 파악: 3.9 %</li> <li>• 기타 혹은 잘 모름: 각각 5.8%, 3.2%</li> </ul>

자료: 独立行政法人 情報処理推進機構(2012, 8), くらしと經濟の基盤としてのITを考える研究会報告書, つながるITがもたらす豊かなくらしと經濟, p.144.

필요가 있으며, 대량 데이터를 이용·활용하기 위한 권리 관계에 대한 명확한 규정이 없다라는 의견이 있는 것으로 나타났다.

## 6. 일본의 빅 데이터 활용을 위한 프라이버시 보호 제도<sup>6)</sup>

빅 데이터를 활용하는데 있어서는 개인의 각종 기록정보를 수집하여 분석함으로써 개인의 프라이버시를 침해할 우려가 있다. 또한 외부로부터의 침입에 의한 데이터의 유출의 위험성과 데이터 취급자의 부적절한 업무처리로 인한 무의식적인 공개와 같은 잠재적 위험도 있다. 프라이버시가 노출되거나 개인정보가 악용된 경우 경제적 면에서 다음과 같은 3가지 국면에서의 비용 발생이 상정된다. ① 대책비용의 발생 ② 손해배상에 대한 대응 ③ 사업활동의 지속 혹은 신용저하에 따른 매출감소이다. 실제로 2010년에 연간 약 1,700건의 개인정보유출과 1,000억엔 이상의 손해배상이 있다고 보고하고 있다(일본 네트워크 보안 협회, 2012)<sup>7)</sup>. 이러한 프라이버시 보호를 위한 다양한 기술적인 기법이 도입되고 있고 조직적인 방책 그리고 법제도의 대책 등이 검토되고 있다. 기술적으로는 프라이버시 보호 데이터 마이닝(PPDM; Privacy-Preserving Data Mining) 등의 연구가 진행되고 있고 조직적인 방책으로는 사업자의

자주적인 행동 의지를 바탕으로 행동규범의 확립이 무엇보다 중요하다는 인식하에 시스템 운용과 관리상의 각종 기법이 소개되고 있다. 법제도의 측면에서는 미국의 ‘소비자 프라이버시 권리장전 (Consumer Privacy Bill of Rights)’이나 유럽의 ‘EU 데이터보호 규정(General Data Protection Regulation)’이 정비되어 있으며, 일본에서는 ‘개인정보보호법’이 이를 제도화하고 있다. 그러나 개인정보보호에 중점을 두면 빅 데이터의 활용을 저해하게 될 우려가 있다. 이는 소비자와 공급자의 양자가 상호간에 다양한 효용성을 얻을 수 있는 기회를 놓치는 것이 된다. 법률상의 취급규정이 불명확한 경우 모험하지 않는 일본기업의 특성을 감안할 때 법제도 상에서 개인정보의 취급에 방법을 명확하게 할 필요가 있다. 일본의 ‘개인정보보호에 관한 법률’은 ‘국가 및 지방공공단체 일반기업의 개인정보취급에 관련된 법률’(약칭 ‘개인정보보호법’)으로 2003년 5월 23일에 수립하여 2005년 4월 1일 전면 시행되었다. 일본의 개인정보보호법은 민간 사업자의 개인 정보 취급에 관한 규정을 정하고 있지만 각 사업 분야에 있어 공통되는 최소한의 규정에 국한되고 있다. 이런 상황을 감안하여 각 사업 분야를 소관하는 주무부처가 각 분야의 실정에 맞는 ‘개인정보보호에 관한 가이드라인’을 책정하고 있다. 이는 사업 분야마다 취급하는 개인 정보의 내용이나 성질, 용도가 다르기 때문이다. 각 사업자가 사

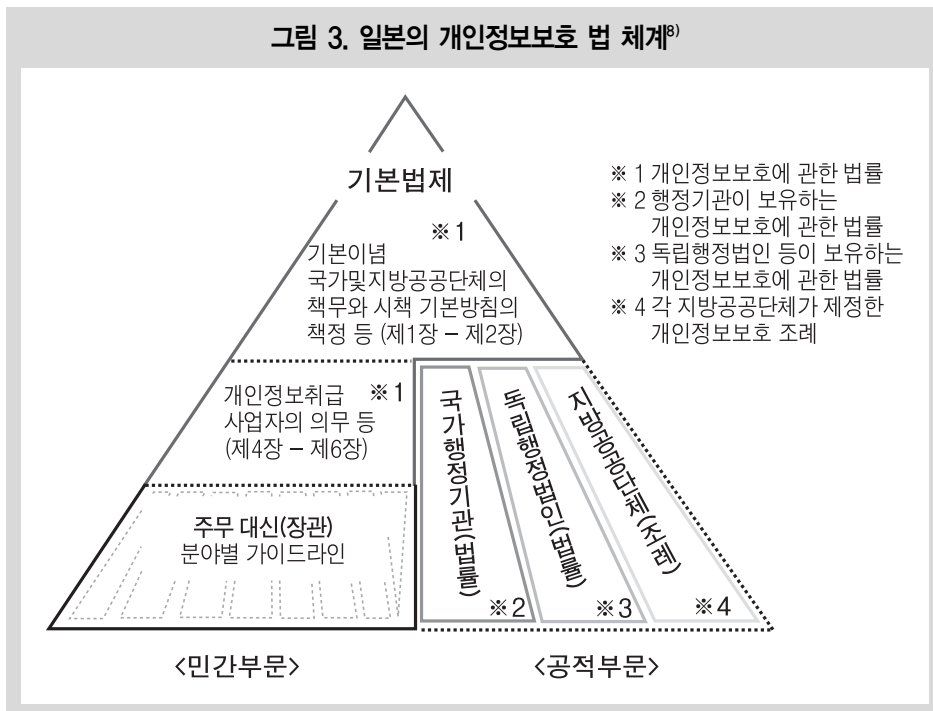
6) 본 내용은 ‘2013 인터넷 건강정보 게이트웨이 시스템 구축 및 운영’ 연구보고서에 게재 예정인 ‘빅 데이터 활용을 위한 프라이버시 보호방안(저자: 동경공업대학 이중순교수)’ 논문에서 발췌한 것임을 밝힌.

7) 日本ネットワークセキュリティ協会(2012. 12. 7) 2012年 情報セキュリティインシデントに関する調査報告書 -個人情報漏えい編-

업 활동을 함에 있어 개인정보를 취급하는 경우에는 개인정보보호법 외에도 그 사업 분야의 가이드라인을 준수해야 한다. 가이드라인이 법률은 아니지만 그 내용에 주무 대신(장관)의 지도, 권고, 명령 등이 포함되어 있어 각 사업자는 자신이 소속된 업종의 가이드라인을 잘 이해하고 바르게 적용해야 한다. 가이드라인은 재판에 있어 법률 해석의 근거가 되기도 한다. 2009년 3월 31일 현재 24분야에 37개의 가이드라인이 각 담당 부처에 의해 책정되어 있다. 일본 정부는 2008년 7월 개인정보보호 관계부처 연락회의를 개최하여 개인정보보호에 관한 가이드라인의 공통화를 추진키로 하였다. 가이드라인의

공통화는 각 부처가 수립한 가이드라인에 차이가 있는 항목을 면밀히 검토하여 각 사업분야의 특성이나 유일성에 의거하는 부분을 제외하고 통일되지 않은 부분에 대해서는 개인정보보호를 정부가 종합적이고 일체적으로 추진하는 관점에서 내각부의 방침에 따라 개정하도록 한 것이다. 분야별 특수한 상황을 감안하면서도 가능한 가이드라인 사이의 차이를 축소하려는 노력의 일환이다. 이러한 노력은 가이드라인 명칭의 공통화 등 형식적인 정리를 촉진하고 민간분야의 개인정보보호 제도를 대외적으로 이해하기 쉽게 함을 목적으로 한 것이다.

그림 3. 일본의 개인정보보호 법 체계<sup>8)</sup>



8) 消費者庁(2012), 個人情報保護法 体系イメージ



## 7. 향후 과제와 시사점

빅 데이터의 기술적인 측면에서는 일단 빅 데이터를 정제하는 기술은 확립되어 있으며, 각종 센서에서 실시간으로 대규모 데이터를 수집하는 방법론도 가닥을 잡아가고 있다. 향후 이러한 데이터를 어떻게 효율적으로 분석하느냐가 관건이다. 기존에 연구된 데이터마이닝과 클러스터링 기술 등의 통계적 분석기술이 필요하지만 이러한 기술은 컴퓨터자원을 다량으로 필요로 하는 것이어서 빅 데이터를 처리하는 데는 개선의 여지가 많다. 빅 데이터의 미래예측 측면에서는 빅 데이터는 현재 상태를 나타내는 데이터로부터 미래의 추세를 판단한다는 점에서 의미가 있지만 장기적인 예측에는 한계가 있을 것으로 보인다. 이는 과거부터 빅 데이터를 분석하여 단기 예보와 장기 예보를 하는 기상 예측의 예에서와 같이 어떤 현상이 어느 일정기간 계속되는 모델이라면 데이터 분석에서 어느 정도의 단기에측은 가능하다. 그러나 장기적으로는 아주 사소한 현상이 큰 현상으로 발달하는 나비효과(북경에서 나비가 날개짓을 하면 뉴욕에서 폭풍이 일어난다는 현상 등으로 설명되는 카오스 이론)와 같은 혼돈요인이 잠재하면 장기 예측을 정확히 하기는 어려운 것이다. 이러한 효과들이 그대로 모든 빅 데이터에 적용되는지는 불분명하지만, 그러한 현상이 가지는 특이성을 염두해 두고 빅 데이터 분석의 한계를 이해할 필요가 있을 것이다. 인적 자원의 측면에서는 앞 절의 인터넷 설문 조사에서도 밝혀진 바

와 같이 빅 데이터를 활용하는 하는데 있어 가장 큰 과제는 대량의 데이터를 취급할 수 있는 인재의 부족이다. 빅 데이터 인재는 빅 데이터 분석에 정통한 '데이터 사이언티스트'라는 분석가로서의 기량을 가진 사람과 그 분석결과를 응용할 수 있는 경영책임자의 양자를 가리킨다. 현재 급성장하는 빅 데이터시장의 수요에 비해 공급이 부족하다. 대학과정에서 빅 데이터분석에 필요한 지식을 가르치게 된 것이 최근 1년 정도이므로 앞으로 몇 년 동안에 걸쳐 개선될 것으로 보인다. 또한 빅 데이터를 활용해야 하는 경영책임자는 교육을 통해 데이터 활용의 장점을 배우고 직감과 경험뿐만이 아니라 데이터에 근거한 합리적인 경영판단을 하는 일의 중요성을 습득함으로써 한층 더 빅 데이터의 활용을 기대할 수 있을 것으로 본다.

일본의 빅 데이터 동향분석을 통하여 한국의 보건복지 분야 빅 데이터의 효율적 활용을 위한 전략<sup>9)</sup>은 다음과 같다.

첫째, 보건복지 빅 데이터를 통합적으로 관리하기 위한 범부처 차원의 (가칭)보건복지 빅 데이터 관리 위원회의 운영이 필요하다. 현재 보건복지 빅 데이터는 보건복지부, 고용노동부, 지식경제부(현 산업통상자원부, 미래창조과학부), 식품의약품안전처, 통계청 등 많은 정부부처와 국민건강보험공단, 건강보험심사평가원, 국책연구기관 등 많은 공공기관에서 관리·운영되고 있어 각 기관에서 운영 중인 정보의 연계와 공유를 위해서는 범정부 차원의 조직이 필요할 것이다.

9) 본 전략은 '송태민(2012. 11). 보건복지 빅데이터의 효율적 활용방안, 보건복지포럼, 한국보건사회연구원'의 내용을 수정·보완함.

둘째, 비정형화된 보건복지 빅 데이터를 관리하고 있는 민간 기관과의 협조체제가 마련되어야 할 것이다. 비정형화된 보건복지 빅 데이터는 민간 기관의 검색포털이나 SNS를 통해서 생산·저장되고 있어 민간기관과의 긴밀한 협조체제(가칭: 보건복지 빅데이터 포럼)가 구축되어야 할 것이다.

셋째, 국가 차원의 오픈 API(Open Application Programming Interface)의 제공이 필요하다. 보건복지의 대부분의 빅 데이터는 공공부문에서 독점하고 있다. 정보를 수집/분석하고 수집과 동시에 정보를 실시간으로 웹상에 공개하는 것도 중요하지만 보건복지 빅 데이터를 효과적이고 효율적으로 활용하기 위해서는 정부차원의 API 공개를 적극적으로 검토할 필요가 있다. 2013년 5월 기준으로 공유자원포털(www.data.go.kr)에서 공개되어 있는 공공정보는 1,721종으로 이중 보건의료 21종, 복지 29종에 불과하다. 따라서 보건복지 빅 데이터의 공개는 관련기관과 빅 데이터 전문가의 참여로 정부와 국민이 필요로 하는 정보를 분류하고 공개대상 정보는 개인정보를 철저히 보안하여 국가지식 플랫폼에 저장할 수 있을 것이다. 넷째, 보건복지 빅 데이터를 분석 처리할 수 있는 관련 기술의 개발이 필요하다. 스마트 시대에는 비 관계형, 비정형 데이터의 저장과 분석, 클라우드 서비스의 확산, 시멘틱 검색 서비스, 추론에 기반한 상황 인식 서비스 등의 기술이 핵심이 될 것이다. 따

라서 관련부처와 협력하여 보건복지 분야 빅 데이터를 ‘수집→저장→분석→추론’ 할 수 있는 기술개발은 물론 기술 표준화가 우선적으로 추진되어야 할 것이다. 다섯째, 구조화되지 않은 대규모 데이터 속에서 숨겨진 정보를 찾아내는 데이터 사이언티스트(Data Scientist)의 인재 양성이 필요할 것이다. 빅 데이터 시대에는 데이터를 관리하고 분석할 수 있는 인력이 매우 중요하다. 이미 글로벌 IT업체에서는 데이터 사이언티스트에 대한 인재 확보와 역량 강화에 많은 노력을 경주하고 있다. 따라서 교육부와 협력하여 보건복지 분야 데이터 사이언티스트를 양성할 수 있는 전략이 마련되어야 할 것이다.

끝으로 보건복지 빅 데이터의 개인정보와 기밀정보에 대한 보안정책이 마련되어야 할 것이다. 보건복지 빅 데이터는 개인에 대한 거의 모든 정보가 저장되어 있지만 아직 법·제도는 미비한 상황이며 논의조차 되지 못하고 있다. 빅 데이터의 활용도 중요하지만 과도한 개인정보의 유출은 프라이버시 침해는 물론 사이버 인권 침해나 범죄에 악용될 수 있다. 빅 데이터로부터 개인을 보호하기 위해 가장 중요한 것은 특정 개인을 식별하지 못하도록 하는 익명화와 정보접근 및 정보처리에 대한 통제이다. 그러나 정보접근 및 정보처리에 대한 통제를 강하게 하면 정보 활용이 활성화되지 않기 때문에 보건복지 빅 데이터 ‘활용과 보호의 균형’에 대한 효과적인 정책이 마련되어야 할 것이다. 