

빅데이터, 오픈데이터, 마이데이터의 비교 연구*

A Comparative Study of Big Data, Open Data, and My Data

박주석

경희대학교 경영대학

요약

지금은 데이터혁명 시대라고 한다. 데이터혁명 시대는 빅데이터로 시작하였고 오픈데이터를 거쳐서 마이데이터로 완성될 것이라 얘기한다. 본 논문에서는 빅데이터, 오픈데이터, 마이데이터를 비교 분석하고, 디지털자원으로서 마이데이터의 역할과 효과를 제시하고자 한다.

- 중심어 : 데이터혁명, 빅데이터, 오픈데이터, 마이데이터, 데이터 경제, 데이터 생태계

Abstract

With the advent of the fourth industrial revolution, data becomes very important resource. Now is called as 'Data Revolution Age.' It is said that Data Revolution Age started with Big Data, then accelerated with Open Data, finally completed with My Data. In this paper, we compared Big Data, Open Data, and suggested roles and effects of My Data as a digital resource.

- Keyword : Data Revolution, Big Data, Open Data, My Data, Data Economy, Data Ecosystems

I. 서론

지난 60년 동안 3번의 정보 혁명이 있었다고 얘기한다. 즉, 컴퓨터 혁명, 인터넷 혁명, 그리고 디지털 혁명이 일어났다고 한다. 그렇다면 각 정보 혁명을 뒷받침해 준 인프라는 무엇일까? 많은 전문가들이 각 정보 혁명의 인프라는 컴퓨터 인프라, 통신 인프라, 데이터 인프라라고 얘기한다.

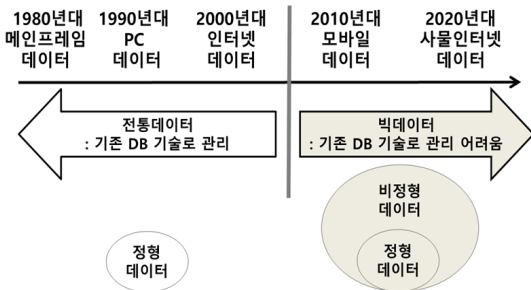
과거 아날로그 혁명에서 핵심자산이 자본, 설비, 인력, 원료 등이었다면, 디지털 혁명의 핵심 자산은 데이터이다. 데이터가 '자산'으로 부각되는 이유는 모든 것이 연결되어 언제나 데이터

를 주고 받을 수 있으며 우리의 모든 활동이 데이터로 기록될 수 있기 때문이다. 이러한 데이터를 매개로 한 생태계가 급속하게 성장하면서 '데이터산업'이 급부상하고 있다. 향후 초연결 사회와 초지능사회로 발전되면서 데이터는 더욱 중요한 자산이 될 것이다[3, 5, 14].

최근에 가장 많이 얘기되는 데이터 분야는 빅데이터(Big Data), 오픈데이터(Open Data), 마이데이터(My Data)이다. 본 논문에서는 빅데이터, 오픈데이터, 마이데이터를 비교 분석하고, 디지털자원으로서 마이데이터의 역할과 효과를 제시하고자 한다.

II. 빅데이터와 데이터 생태계

정보기술이 발전하면서 데이터 규모는 폭발적으로 증가하였다. 데이터 규모뿐만 아니라, 종류도 다양해졌고, 데이터 주기도 빨라졌다. 이러한 빅데이터의 특성을 일반적으로 3V(Volume, Variety, Velocity)로 지칭한다. 하지만 데이터생태계(data ecosystem) 측면에서 보면 다양성(Variety)이 중요하다. 즉, 정형데이터에서 비정형데이터로 확장됨에 따라 데이터생태계가 생성되기 시작했다[4, 5, 15].



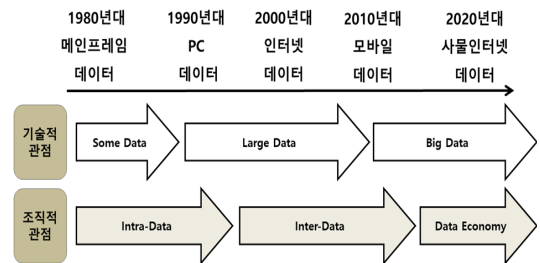
〈그림 1〉 데이터의 변천사 그리고 빅데이터

예를 들어 제조업의 경우 빅데이터를 활용한 공급사슬관리(SCM) 데이터생태계를 목표로 하고 있다. 단순히 스마트공장을 구축하는 것이 아니라 스마트공장, 스마트유통, 스마트제품, 그리고 스마트서비스를 융합하는 것이다. 대표적인 사례가 제너럴일렉트릭(GE) 회사의 지능형 항공운행 사례이다. GE사는 스마트항공기를 생산하여 장비 및 부품 정비는 물론이고 항공 운영상의 다양한 프로세스를 개선하고 있다. 또한 고객과 협력업체를 위한 지능화된 스마트서비스를 준비하고 있다. 궁극적으로 부품업체부터 항공사에 이르는 공급사슬관리의 데이터생태계 구현을 목표로 하고 있다[11]. 이러한 데이터생태계 구현을 위해서는 사물인터넷 데이터가 수집되고 관리되고 분석되어야 한다.

앞에서 언급된 사례는 제조업 데이터생태계

사례이지만, 모든 산업에서 데이터생태계 혁신이 일어나고 있다. 금융산업에서는 핀테크(FinTech) 데이터생태계가 구축되고 있고 의료산업에서는 웰니스(Wellness) 데이터 생태계가 형성되고 있으며, 교통산업에서는 자율주행차를 위한 데이터생태계가 논의되고 있으며 공공분야에서는 스마트 시티(Smart City) 데이터생태계가 나타나고 있다 [11].

최근에 데이터가 새로운 형태의 자산으로 주목받으면서 데이터 유통에 기반한 새로운 생태계인 ‘데이터 경제(Data Economy)’ 개념이 화두가 되고 있다. 데이터 경제란 ‘데이터에 접근하고 활용할 수 있도록 협업하는 과정에서 데이터 생산, 인프라 제공, 연구조사, 데이터 소비 등 서로 다른 역할을 담당하는 구성원으로 이루어진 생태계’를 의미한다[4, 7, 13].



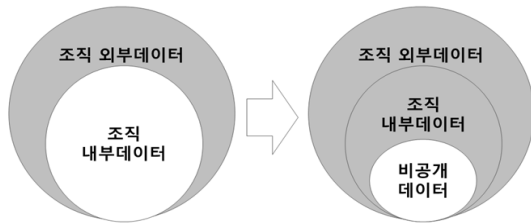
〈그림 2〉 데이터경제 시대의 도래

데이터 경제 규모는 유럽연합(EU)의 경우 2014년 기준으로 2,570억 유로(EU GDP의 1.89%)이었고 2020년에는 6,430억 유로(EU GDP의 3.17%)로 증가할 것으로 전망하고 있다. 유럽연합에서는 데이터를 경제성장뿐 아니라 일자리 창출과 사회 발전을 위한 필수 자원으로 인식하고 데이터 경제 활성화를 위한 다양한 노력을 기울이고 있다[13].

III. 오픈데이터와 데이터 민주주의

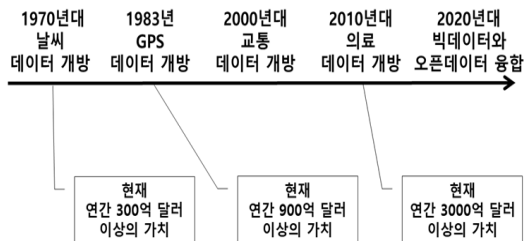
오픈데이터는 저작권 및 특허권의 제약 없이 혹은 합법적 통제를 통해 데이터자원을 모두에

게 원하는 형태로 무료로 활용 및 재배포하는 것이다[5, 11]. 이런 점에서 오픈데이터의 철학은 오픈소스(Open Source)나 오픈이노베이션(Open Innovation) 철학과 같다[5].



〈그림 3〉 오픈데이터의 개념

오픈데이터 개념은 최근에 추진된 ‘정부 3.0’ 정책과 ‘www.data.go.kr’ 포털에 의하여 많이 알려졌다. 하지만 전세계적으로 보면 오픈데이터 사례는 빅데이터 사례처럼 이미 오래 전부터 시작되었다. 많은 전문가들이 오픈데이터는 세계 최대의 무료자원이라고 하면서 웹만큼 거대한 기회가 올 것이라고 주장하였다[5].



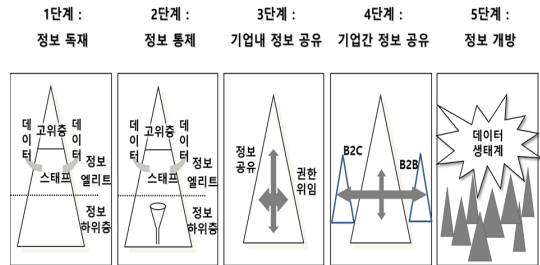
자료원 : Open Data Now(2013), 박주석(2017).

〈그림 4〉 오픈데이터의 역사

오픈데이터의 발전을 ‘데이터 민주주의’ 개념으로 바라볼 수 있다. 역사적으로 크게 보면, 정보독재 시대에서 정보통제 시대를 거쳐서 정보개방 시대가 도래하였다[2].

작게 보면, 행정정보공개 정책부터 데이터 민주주의가 시작되었고 오픈거버넌트(Open Government) 정책부터 본격화 되었다. 또 다른 측면에서 보면 인터넷이 포화화되면서 데이터 민주주의가

시작되었고 개인정보보호 정책에 의해서 데이터 민주주의가 성숙되었다. 따라서 데이터 민주주의는 그동안 데이터 개방과 데이터 보호라는 2가지 정책에 의하여 발전되었다[2, 12].

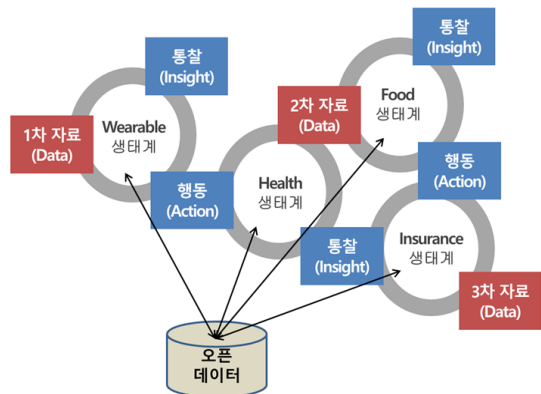


자료원 : e비즈니스인텔리전스(2001), 박주석 개정(2017).

〈그림 5〉 ‘데이터민주주의’의 발전

이런 관점에서 정리하면 데이터 민주주의는 ‘특정한 개인에 대한 정보가 아니면, 누구나 그 데이터를 사용, 재사용 및 재배포 할 수 있어야 한다’는 의미로 정의할 수 있다[2].

데이터민주주의가 확산되면 데이터생태계는 더욱 진화할 것이다. 특정 비즈니스 영역의 데이터 생태계는 오픈데이터와 융합되면서 새로운 관점의 데이터생태계를 만들어낼 것이다. 예를 들어 공공데이터가 개방되면서 공공데이터와 기존 민간데이터를 융합하여 새로운 데이터 서비스나 새로운 비즈니스 모델을 제공하고 있다[4, 5].



〈그림 6〉 데이터생태계와 오픈데이터의 융합

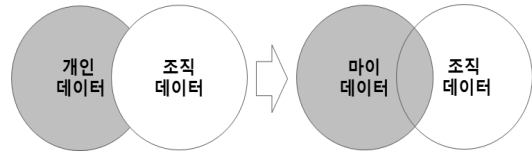
우리나라 데이터 민주주의 수준을 살펴보면, 지난 10년 동안에 괄목할 성장을 하였다. 우리나라는 2011년 농협사태 등을 겪으면서 전 세계에서 가장 강력한 개인정보보호 제도를 갖고 있다. 공공 데이터 개방도 선진국에 비하면 늦게 시작하였지만 적극적으로 추진하여 2015년에는 OECD 공공데이터 개방 지수 1위를 차지하였다. 이러한 성과에도 불구하고 향후 4차 산업혁명이 발전하면서 데이터 보호 정책과 데이터 개방 정책이 충돌되는 문제점이 나타날 것이다[2, 13].

IV. 새로운 디지털 자원, 마이데이터

앞에서 데이터 경제와 데이터 민주주의를 언급하였는데 결국 새로운 관점의 신뢰성 있는 데이터 생태계를 창출하는 것이 매우 중요하다. 특히 개방된 공공데이터와 민간 데이터의 융합이 중요하다. 하지만 공공데이터 개방만으로 가치 있는 데이터 생태계를 완성시키기에는 한계가 있다[2, 12].

조직데이터와 개인데이터가 융합됨으로써 훨씬 더 가치 있는 데이터 생태계를 만들어 낼 수 있다. 이런 점에서 개인정보보호법은 융합 데이터 생태계를 만드는 데 걸림돌이 될 수 있다. 물론 하나의 대안으로 개인정보의 비식별화 정책이 추진되고 있으나 섬세한 데이터 생태계를 만들어낼 수 없기 때문에 근본적인 대책이 될 수 없다.

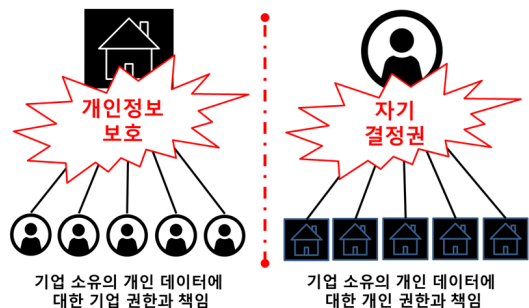
4차 산업혁명을 위하여 데이터 생태계를 자연스럽게 융합해야 한다. 데이터 생태계를 융합하기 위해서는 산업별 마스터데이터가 중요한 연결고리가 되며 특히 개인정보는 가장 중요한 연결고리이다. 예를 들어 개인별로 제조, 금융, 유통, 의료 데이터를 융합하여 새로운 가치를 창출하는 것이 가장 의미있는 데이터 생태계이다. 이런 점에서 마이데이터(My Data) 개념은 융합데이터 생태계를 구현하는데 핵심적인 역할을 할 것이다[12].



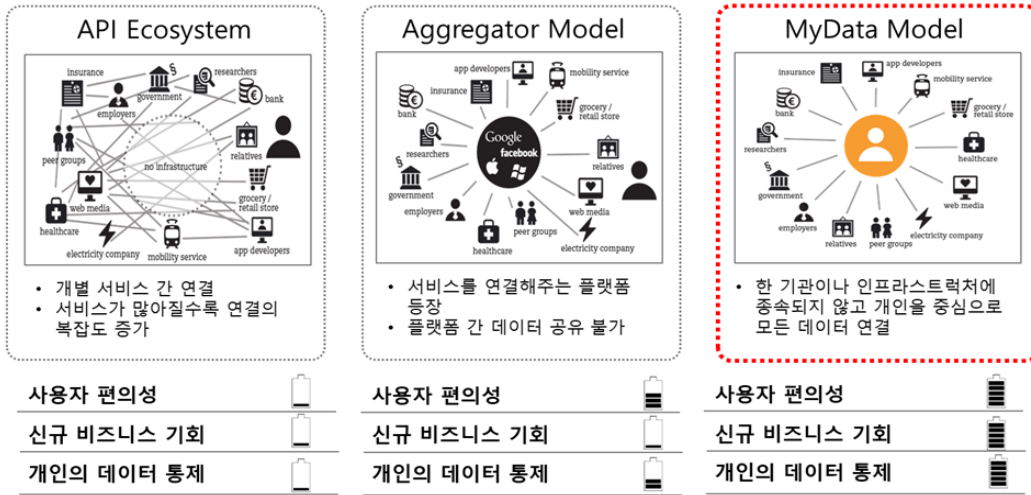
〈그림 7〉 마이데이터의 개념

2011년 IDC 보고서에 의하면 디지털데이터 중 개인데이터 비중은 약 75%이며, 사물인터넷과 인공지능 기술로 인해 개인데이터의 비중과 가치가 더 증가할 것으로 예상된다. 실제로 구글, 페이스북, 아마존 등 인터넷기업은 개인데이터를 기반으로 성장하여 지금은 글로벌 시가총액 TOP 5 기업이 되었다. EU에서 2020년까지 개인데이터로 창출되는 경제적 가치는 약 1조 유로, 정부와 기업이 얻는 연간 이익은 3,300억 유로로 예상하고 있다.

많은 사람들이 개인정보보호와 마이데이터를 상반된 개념으로 오해하고 있다. 개인정보보호는 기업 관점에서 기업소유의 개인데이터에 대한 기업 권한과 책임을 얘기한다. 반면에 마이데이터는 개인 관점에서 기업소유의 개인데이터에 대한 개인 권한과 책임을 얘기한다. 즉, 마이데이터 개념은 기업 중심의 개인데이터 생태계에서 개인 데이터의 주인인 개인에게 자신의 데이터를 통제 및 관리할 수 있는 권한을 돌려주는 개념이다. 따라서 개인정보보호 정책은 유지되면서, 마이데이터 정책을 적극 추진할 수 있다[12].



〈그림 8〉 ‘개인정보보호’ 대 ‘마이데이터’



자료원: Open Knowledge Finland (2015).

〈그림 9〉 융합 데이터생태계 모델의 비교

마이데이터 정책의 핵심은 개인정보의 소유권(ownership)을 당사자에게 돌려주는 것이다. 실제로 많은 개인정보는 개인이 아니라 기업이나 기관이 갖고 있다.

개인의 금융데이터는 그 개인이 이용하는 금융기관이 대부분 갖고 있으며, 개인의 의료데이터는 진료를 받았던 병원이 대부분 갖고 있고, 개인의 공공데이터는 서비스 받았던 공공기관이 대부분 갖고 있다. 하지만 이러한 개인정보의 소유권자는 그 데이터를 관리하는 기관이나 기업이 아니라 그 개인이다. 따라서 그 당사자만이 금융데이터, 의료데이터, 통신데이터, 구매데이터, 공공데이터 등을 활용하고 융합할 수 있다[12].

결국 마이데이터는 개인의 융합 데이터 생태계를 만들 수 있는 기회를 제공하고 있다. 특히 마이데이터의 개인데이터 자기 이동권은 데이터 생태계 융합을 촉진하면서 새로운 비즈니스모델을 창출하고 지능화된 사회를 촉진시킬 것이다[12].

V. 마이데이터 중심의 데이터경제

데이터경제는 새로운 융합 데이터생태계를 의

미한다. 융합 데이터생태계는 3가지 모델로 분류된다. 즉, 인터페이스(API : application program interface) 기반의 생태계, 집합자(Aggregator) 기반의 생태계, 마이데이터(My Data) 기반의 생태계이다[9].

인터페이스(Open API) 기반의 생태계는 개별 서비스 간에 쉽게 연결할 수 있으나, 서비스가 많아질수록 연결의 복잡도는 증가한다. 집합자(Aggregator) 기반의 생태계는 서비스를 연결하는 플랫폼이 중심이 되나 플랫폼 간 데이터의 공유는 어렵다. 마이데이터 기반의 생태계는 특정 기관이나 인프라에 종속되지 않고 개인을 중심으로 모든 데이터를 연결하는 모델이다. 하지만 이제 제도가 수립되고 있으며 솔루션도 아직 실험 단계이다 [10].

마이데이터 기반의 생태계는 개인 사용자 중심의 동기 기반 통제(Consent-Based Control)이며 개인 데이터 인프라의 상호운용성과 이동성을 강조한다. 최근에 블록체인 기반의 마이데이터 생태계가 구축되고 있으며 이러한 모델이 마이데이터 기반의 데이터생태계 표준이 될 것이다.

VI. 결 론

지금은 데이터혁명 시대라고 한다. 데이터혁명 시대는 빅데이터로 시작하였고 오픈데이터를 거쳐서 마이데이터로 완성될 것이라고 얘기한다. 빅데이터, 오픈데이터, 마이데이터는 서로 연관되어 있지만, 각 개념은 분명히 다르다.

본 논문에서는 빅데이터, 오픈데이터, 마이데이터를 데이터생태계 관점에서 비교 분석하고 마이데이터를 위한 데이터생태계 모형을 제시하고자 한다.

참 고 문 헌

- [1] 김승현, 박주석, 박재홍, 김인현, “빅데이터 환경에서 분석자원이 기업성과에 미치는 영향”, 한국빅데이터학회, 제1권, 제1호, 2016.
- [2] 박주석, “데이터 민주주의와 데이터 생태계 융합,” 칼럼 기고, 전자신문, 2018.
- [3] 박주석, “데이터 중심의 공공 정보자원관리”, 한국정보화진흥원, 연구보고서, 2016.
- [4] 박주석, 김인현, “전통적 환경과 빅데이터 환경의 데이터자원관리 비교연구”, 한국빅데이터학회, 제1권, 제2호, 2016.
- [5] 박주석 외 9인, “비즈니스 분석”, 한경사, 2018.
- [6] 이연우, 장현미, 홍승필, “빅데이터 환경 내 개인정보보호를 위한 대용량 개인정보 관리 모델 설계방안”, 한국인터넷정보학회 추계학술발표대회논문집, 제1권, 제2호, pp.29-30, 2012.
- [7] 정용찬, “4차 산업혁명 시대의 데이터 경제 활성화 전략”, 정보통신정책연구원(KISDI), 2017.
- [8] 정준화, “4차 산업혁명 시대의 빅데이터 정책 과제”, 이슈와 논점, 국회입법조사처, 2018.
- [9] 투이컨설팅, “국내 개인데이터 이용 환경 분석”, 연구보고서, 한국데이터진흥원, 2017.
- [10] 투이컨설팅, “MyData 시물레이션 및 시스템 설계 연구”, 연구보고서 한국데이터진흥원, 2017년 11월.
- [11] 한국데이터진흥원, “2017 데이터산업백서”, 한국데이터진흥원, 2017.
- [12] 한국데이터진흥원, “2018 데이터산업백서”, 한국데이터진흥원, 출간 예정.
- [13] European Commission, “Building a European Data Economy”, *Data Policy and Innovation*, EU, January 2017.
- [14] Ibarra, I. A., L. Goff, D. J. Hernández, J. Lanier, and E. G. Weyl, “Should we treat data as Labor?”, *American Economic Association Papers & Proceedings*, Vol.1, No.1, 2017.
- [15] Gartner, “Establish a Data Quality Program to Support Digital Business”, April 2015.

저 자 소 개



박 주 석 (Jooseok Park)

- 1981년 : 서울대학교 산업공학 (학사)
- 1983년 : KAIST 산업공학 (석사)
- 1990년 : University of California, Berkeley,

경영대학 MIS 전공 (박사)

- 현재 : 경희대학교 경영대학 교수
- 관심분야 : 데이터베이스, 모델링, 아키텍처, 정보화전략 등