

1- COVER STORY —

기획·구성

박선주 한국정보화진흥원 책임연구원

데이터, 그리고 미래정부

이제는 데이터 기반 정부다!

국내외 데이터 전문가들이 주목한 데이터 기반 정부의 기회와 도전

- ① 데이터와 공공행정
- ② 데이터와 정치·선거
- ③ 데이터와 경제·산업
- ④ 데이터와 사회·문화

데이터, 그리고 미래정부 이제는 데이터 기반 정부다!

박주석

경희대학교 교수

서울대 산업공학과에서 학사, KAIST 산업공학과에서 석사, 그리고 미국 UC Berkeley 경영대에서 박사학위를 받았고, 현재 경희대학교 MIS 전공 교수이다. 한국EA학회 회장, 한국경영정보학회 부회장, 한국경영학회 부회장을 역임하였고 현재 한국빅데이터학회 회장이다. 지난 25년 동안 국내 여러 금융기관의 대규모 시스템 구현에 깊이 관여하였고 공공 데이터 정책수립에 적극적으로 참여하였다. 주요 관심분야는 정보화전략, 모델링, 아키텍처, 비즈니스분석이다.

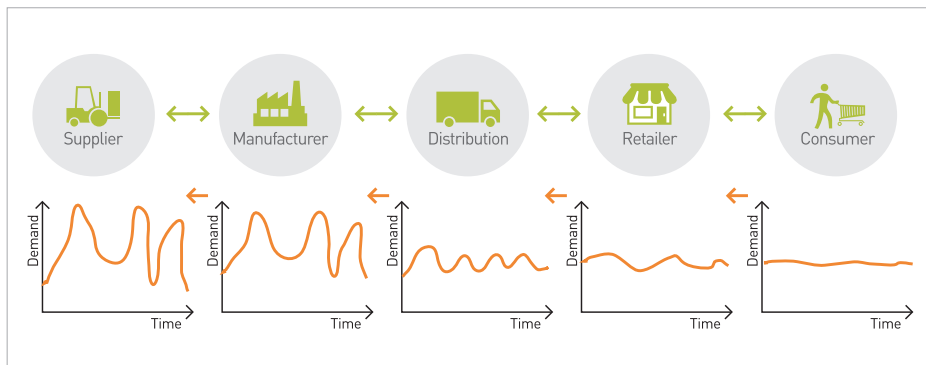
데이터, 미래를 보장하는 차세대 정보자원

양(Quantity)과 질(Quality)의 보완이 관건이다

정보자원(Information Resource)은 하드웨어 자원, 소프트웨어 자원, 데이터 자원 등으로 분류할 수 있다. 수많은 ICT 관계자들은 초기에 가격이 비쌌던 하드웨어 자원의 효율적 활용을 고려했으나, 시간이 지나면서 실제로 업무를 자동화시키는 소프트웨어 자원에 더 큰 관심을 가지게 되었다. 특히, 소프트웨어 자원에 따라 조직 생산성에 확실한 차이가 나타나면서부터 민간, 공공의 많은 조직들이 소프트웨어 자원에 집중하기 시작했다. 조직업무의 대부분이 정보화된 이후에는 정보시스템을 통해 생성되는 데이터 자원으로 관심이 옮겨갔다. 즉, 조직 차원에서 데이터 자원 연계와 통합으로 새로운 가치를 창출하고, 데이터 분석에 기반한 빠르고 정확한 의사결정이 가능하게 된 것이다.

데이터 자원은 한 부처나 기관의 정보화가 진행되고 난 이후, 시스템이나 프로세스가 혁신으로 변화하더라도 살아 남는다. 결국, 축적되는 데이터 자원의 품질이 조직의 내부행정과 서비스 등을 제공하는 정보자원의 품질을 좌우하게 된다. 개별 부처와 공공기관, 하나의 정부(Whole of government)에도 마찬가지다. 이와 같이, 현재와 미래의 정보자원은 그 시작과 끝이 데이터로 통한다고 해도 과언이 아니다. 이미 수년 전 가트너가 “데이터는 21세기 원유”라고 주장했던 외침을 부정할 이는 아무도 없다. 따라서, 미래사회의 정부 또한 데이터 기반의 정부(Data-Driven Government)로 그 패러다임이 전환될 것이다.

| 소채찍 효과(Bullwhip Effect) |



- 공급망을 거슬러 올라갈수록 변동폭이 커지는 현상을 소채찍 효과라고 한다. 소를 물 때 쓰는 긴 채찍의 경우, 손잡이 부분에서는 작은 힘이 가해져도 끝 부분에서는 큰 파동이 생기는데 착안하여 붙여진 용어이다.

미래정부의 새로운 원유로 자리잡을 데이터 자원의 가치와 특성에 대해 조금 더 고민해보자. 2008년 미국 소매업체의 재고데이터를 분석한 한 연구사례에 따르면, 전국 37개 지점에서 획득한 약 37만 여 개의 재고데이터 중 부정확한 65%가 소매업체의 운영성과에 부정적 영향을 초래한 것으로 밝혀졌다.¹⁾ 품질이 떨어지는 재고데이터는 제품의 수요예측과 창고관리 등을 위한 의사결정시스템에 투입해 온 소매업체의 비용(연 매출액의 1%에 해당하는 약 1조원) 대비 효과를 떨어뜨렸기 때문이다. 또한, 공급망(Supply chain) 상에서의 소채찍 효과(Bullwhip effect)도 초래했다. 이에 따라, 소매업체는 물론, 물류업체, 제조업체, 협력업체 등의 성과에까지 커다란 악영향을 미치게 되었다. 이 연구결과는 단순히 한 소매업체 내부의 데이터 오류를 넘어, 조직(소매업체) 간 정확한 데이터 연계에 대한 중요성을 일깨워 준다.

1) Nicole DeHoratius and Ananth Raman. 2008. Inventory Record Inaccuracy: An Empirical Analysis, Management Science, Vol. 54, No. 4. pp 627-641.

데이터 자원 관리에 대한 경각심을 전 세계에 인식시킨 다른 사례도 있다. 바로 1929년 경제대공황에 버금가는 수준의 경제적 혼란을 가져온 2008년 세계 금융위기의 원인에 관한 이야기다. 금융위기 이전, 앨런 그린스펀 전 미연방준비제도이사회(FRB) 의장은 “금융기관의 발전된 슈퍼컴퓨팅 파워와 위험 관리시스템을 통해 서브프라임 모기지 채무자들의 신용을 충분히 평가할 수 있을 것이다” 라고 자신하였다. 하지만 2008년 금융위기 이후 앨런 그린스펀 전 의장은 “문제는 위험관리시스템에 입력된 데이터가 저 품질이었다”라고 후회하면서 고개를 숙였다. 물론, 금융위기는 다양한 원인에 의해 비롯되었겠지만, 이 일화는 하드웨어나 소프트웨어 자원이 우수해도 데이터 자원이 잘못되면 아무런 의미가 없음을 여실히 보여줬다 하겠다.

2008년 금융위기 전(前)

“

금융기관의 발전된 컴퓨팅 파워와
리스크 관리 모델을 통해
서브프라임 모기지 채무자들의 신용을
충분히 평가할 수 있을 것이다.

”



2008년 금융위기 후(後)

“

문제는 위험관리시스템에 입력된
데이터가 저 품질이었다는 점이다.
결국 아무리
최첨단 컴퓨터시스템이라고 해도
입력되는 데이터가 잘못되면
의미가 없다.

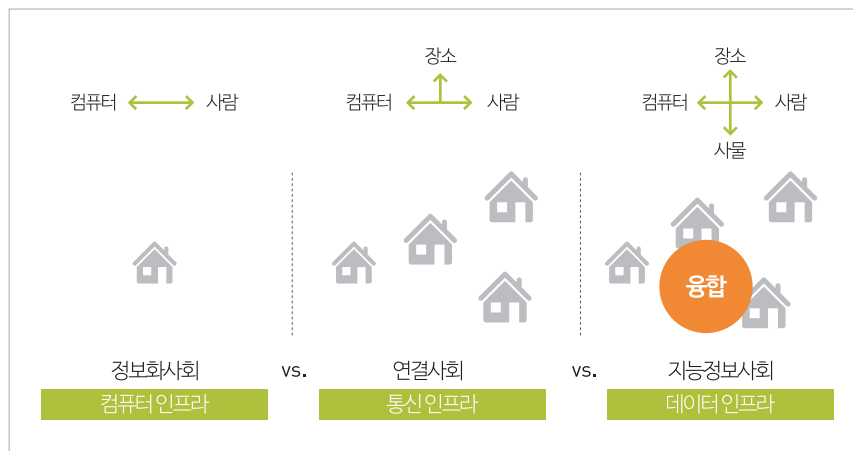
”

이렇듯 데이터 자원은 양과 질을 함께 고려해야 한다. 혹자는 양보다 질을 우선시한다. 물론, 방대한 양을 가졌더라도 데이터의 품질을 제대로 관리하고 어떻게 이를 활용하느냐 하는 것은 중요하다. 그렇지만, 분석대상에 관한 많은 케이스(Cases)가 모여야 보다 유용한 결과를 일반화시켜 현실에 적용할 수 있기 때문에, 데이터의 양도 중요하다. 미래를 보장하는 차세대 정보자원으로서 데이터를 다루는 정부는 양과 질의 균형 잡힌 시각으로 정책을 결정하고, 현안 종류나 사회적 특성 등을 고려해서 적절히 데이터 양과 질을 구분·활용해야 할 것이다.

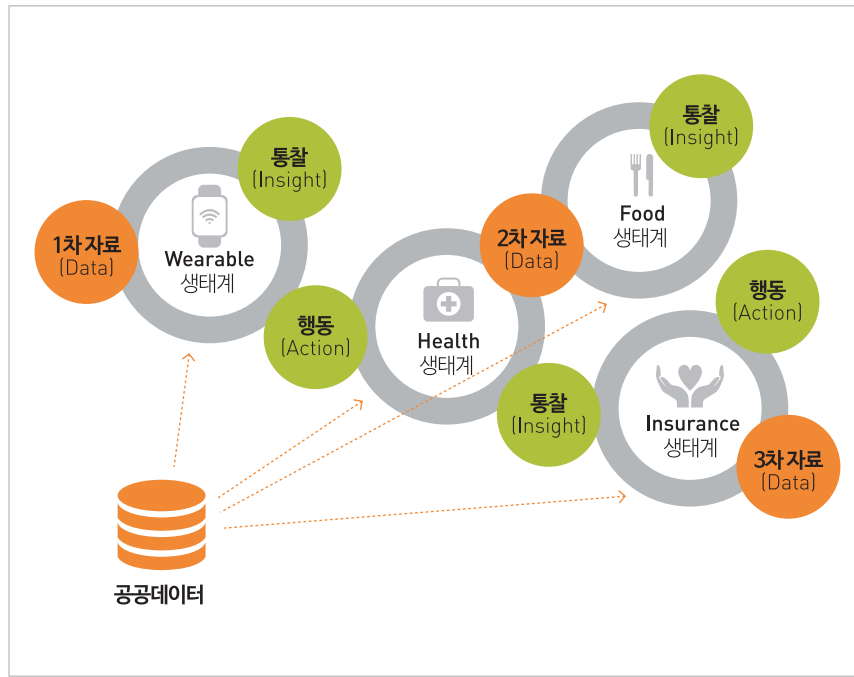
데이터 기반 정부의 “D.I.A.”

스스로 개방하고 지능화되며 유연해져라!

시스코(CISCO) 회사의 챔버스 회장은 “2020년에는 25억 명의 사람과 370억 개의 사물이 인터넷에 연결되고, 2030년에는 500억 개의 사물이 인터넷에 연결되어 혜택을 줄 것이다.”라고 말했다. 또한, 저명한 미래학자들은 ICT 발전으로 사람, 컴퓨터, 장소, 사물 등 모든 것이 상호작용하는 초연결사회(Hyper-connection Society)가 도래할 것이라고 예측한다. 최근에는 여기에 인공지능(AI: Artificial Intelligence) 기술까지 강조되면서 미래사회를 지능정보사회로 지칭하기도 한다. 그렇다면, 지능정보사회를 위한 국가사회의 핵심 인프라는 무엇일까? 바로 앞서 그 중요성을 강조했던 데이터 자원으로 이루어진 인프라이다. 과거 정보화사회의 인프라가 컴퓨터였고 인터넷 기반의 연결사회가 인프라가 통신이었다면, 지능정보사회의 인프라는 데이터가 될 것이고 이러한 데이터는 정부의 정책·서비스를 제공하는 소스이자 정보자원 플랫폼으로서 기능할 것이다.



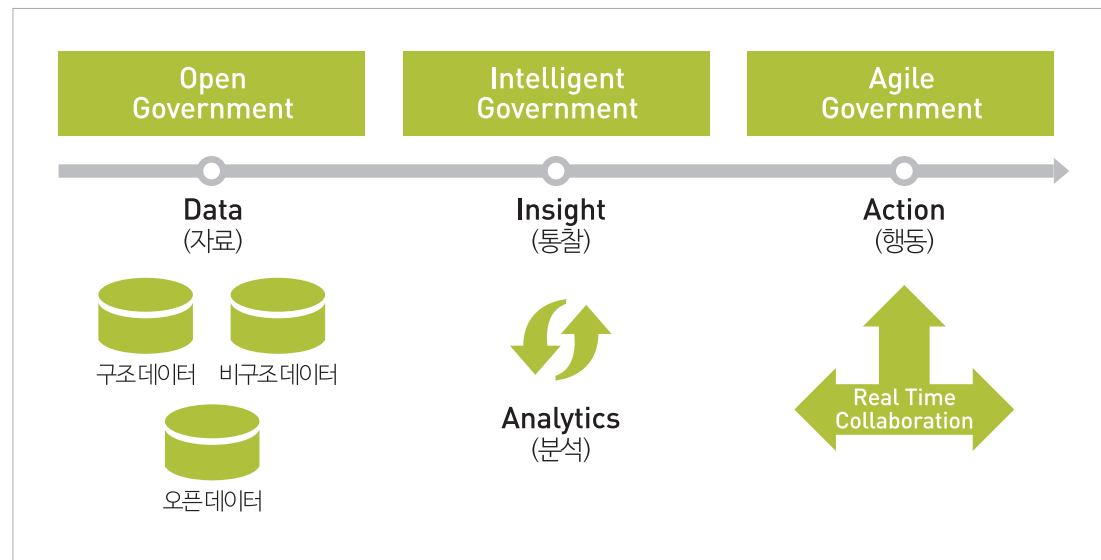
데이터 기반 정부의 데이터인프라는 사물인터넷, 빅데이터 기술을 중심으로 발전해나갈 것이다. 사물인터넷은 센서 등을 비롯한 다양한 디바이스를 통해서 실시간으로 엄청난 데이터를 생성해 낼 것이고, 빅데이터는 그렇게 생성된 데이터를 가공하고 분석하여 데이터인프라가 지능화될 수 있도록 도울 것이다. 예를 들어 웰니스(Wellness) 데이터생태계는 웨어러블 생태계, 헬스 생태계, 푸드 생태계, 보험 생태계가 융합되어진다. 웨어러블생태계는 다양한 센서를 통해서 건강 데이터를 체크하고 헬스생태계는 건강기록과 의료기록을 관리하며 푸드 생태계는 개인 맞춤형 식품을 공급하며 보험생태계는 개인에 맞는 건강보험을 제공한다. 즉, 개인의 웰니스를 위하여 다양한 관점의 데이터가 융합되어 서비스 된다. 이러한 서비스는 축적된 데이터를 활용하면서 지능화된다.



이와 같이, 향후 데이터 기반 정부는 기본적으로 “D(Data), I(Insight), A(Action)”를 따라 데이터 인프라를 정비하고, 지능화된 정책결정과 유연한 서비스를 제공할 수 있을 것이다. 이를 위해서는 우선, 공공과 민간의 다양한 경로를 통해 축적된 방대한 데이터를 개방하고 공유하는 열린정부가 되어야 제대로 된 ‘양+질의 데이터’를 거를 수 있다. 특히, 현재 높은 활용성이 기대되에도 불구하고 이용이 저조한 공공데이터는 미로에 빠져 있다. 미로에 있기 때문에 활용할 수도 없고 연계할 수도 없고 분석할 수도 없다. 데이터 미로를 없애고 데이터지도 를 만들어야 한다. 비유하자면 자동차를 운전하기 위해 지도 네비게이션이 필요하듯이, 융합된 서비스를 위해서 데이터 네비게이션이 필요하다.

두 번째, 통찰(Insight)은 양질의 데이터를 제대로 분석하고 결과값에 대한 정책적 통찰력을 도출하는 것이다. 여기서는 데이터 분석을 정책 프로세스에 내재화시키고, 통찰에 따라 필요한 분석방법을 자동으로 실행할 수 있도록 지능화된 정부업무 처리와 일하는 방식의 개선이 요구된다. 또한, 이 단계에서는 데이터의 공공성과 시장가치를 정확하게 읽고 미래 트렌드에 접목시켜 효율적인 서비스를 디자인할 수 있는 데이터 전문인력이 필요하다. 데이터 사이언티스트를 배양하거나, 공공인력의 데이터 전문성을 함양시킬 수 있는 전략이 보완되어야 할 것이다.

이제, 마지막 남은 것은 행동(Action)하는 것이다. 이때 데이터 기반 정부는 통찰과정을 즉각적으로 반영할 수 있는 유연성을 갖추어야 한다. 이를 위해서는 데이터 분석에 따라 실시간으로 대응이 가능한 기관 간 협업과 보유한 데이터 연계 등이 정비되어야 한다. 특히, 시시각각 변하는 다양한 수요에 따라 순간 빅데이터를 종합 분석한 실시간 결과는 정부의 현장행정 대응력과 위기관리역량, 국민반응을 고려하는 감성형 서비스 제공 등으로 정부만족과 신뢰성 제고에 기여할 것으로 기대된다. 나아가, 데이터 기반 정부의 정책역량을 키우고 정책영역을 넓히기 위해 고려해야 할 법제도, 규제 등의 수정 및 보완을 추진함으로써 제대로 된 정부 행동이 보장될 것이다.



전통적인 정보이론은 DIKW(Data, Information, Knowledge, Wisdom) 사이클이다. 데이터를 가공하면 정보가 되고, 정보가 축적되면 지식이 되며, 지식이 일반화되면 지혜가 된다는 것이다. 최근에 새로운 정보이론은 DIA(Data, Insight, Action) 사이클이다. 데이터를 분석하여 통찰(Insight)을 얻고 바로 기업은 필요한 행동을 실행해야 한다. 이러한 DIA 사이클은 공공조직에도 적용될 수 있다.

① 데이터와 공공행정

정보에서 지식으로, 지식에서 지혜로!

지능화된 공공행정의 시작

이회정

인포와이즈 대표

SK케미칼 전략기획실에서 IT기획을 담당했고 2000년 벤처기업으로 옮겨 사업기획을 담당하였다. 2년 후 인포와이즈를 창업하여 조달청 나라장터 구축과 국내 다수 아이마켓코리아, 바츠 등 전자상거래 업체들의 품목 표준화 컨설팅을 수행하였으며 LG전자, 삼성전자 등 국내 유수 기업들의 MDM 전략 수립 및 구축 작업에 참여하였다.

데이터, 시스템이라는 동굴 벽의 그림자들

홍수다. 데이터의 홍수. 플라톤의 비유처럼 우리는 시스템이라는 동굴 속에서 벽을 바라보고 살고 있다. 벽면에 비춰지는 수많은 그림자들을 바라보며 그것들을 해석하느라 분주하다. 그림자로 보여지는 데이터의 실체가 무엇인지 많은 고민을 하며 그것을 통해 지혜를 얻기를 원하고 있다. ICT 없는 생활을 상상조차 할 수 없는 현대사회 속에서 우리가 접하게 되는 데이터는 기하급수적으로 증가(Volume)하고 있으며, 점점 더 다양한 원천들을 통해 다양한 형태로(Variety) 생겨나고 있을 뿐만 아니라 순식간에(Velocity) 온 세상을 흘러 다닌다. 이러한 데이터와 관련된 일련의 상황을 우리는 '빅데이터'로 이야기한다.



우리가 원하던 원하지 않던 우리는 데이터와 무관하게 살아갈 수 없다. 이를 예견했을까? 1934년 시인 엘리엇(T.S. Eliot)은 자신의 시 한 구절에서 이렇게 노래했다. “삶은 무엇인가? 우리는 생활 속에서 그것을 잊어버렸네. 지혜는 무엇일까? 우리는 지식 속에서 그것을 잊어버렸네. 지식은 무엇일까? 우리는 정보 속에서 그것을 잊어버렸네.” 이 시가 지금 우리의 현실은 아닐까? 우리는 거대하게 쏟아져 들어오는 데이터 속에서 어떻게 지혜를 얻을 수 있을까? 우리 모두의 고민이다. 지능정보사회의 데이터 기반 정부 역시 데이터에 묻혀 해안을 잃어버리지 않기 위해 노력해야 한다. 그러나 그림자를 보여주는 실체를 찬찬히 파악하기에는 그림자가 너무 자주, 그리고 너무 많이 비춰지고 있다. 진실과 지혜를 얻기 힘든 이유가 여기 있다. 우리가 동굴 밖을 나가 직접 모든 것을 확인할 능력이 없다면 우리는 이 그림자들을 통해 의미 있는 무엇인가를 얻어야만 한다. 빅데이터는 정보화된 현대사회의 핵심적인 현상일 뿐이다. 진정한 문제는 정부가 ‘이 빅데이터를 어떻게 이해하고 분석하여 현상 이면의 실체를 파악하고 이를 진실과 지혜로 바꾸어 낼 수 있는가’이다.

더 나은 행정과 손에 잡히는 공익 창출을 위해

첫째, 결국은 분석! 그 중에서도 분석의 축이 먼저

최근 빅데이터를 주도하고 있는 미국에서의 화두는 역시 분석이다. 효율을 이끌어내고 의사결정을 개선할 도구로써 빅데이터 분석에 대해 관심이 매우 높다. 빅데이터 분석은 의사결정을 지원하기 위해 정형화된 데이터를 기반으로 하는 통상적인 BI(Business Intelligence)와는 다르게, 정형화된 시스템 정보뿐만 아니라 방대한 비정형적(Unstructured) 정보를 포함하여 정교한 알고리즘의 활용을 통해 수행하는 분석을 말한다. 특히 공공부문에서 이러한 분석 알고리즘은 정부의 정책 적용 이전에 사기, 예산 낭비나 오용 등과 같은 문제를 파악하고 빠른 주기로 분석을 실행하는 분석 프로그램의 정확성을 높이고 정책의 실효성을 극대화하는데 적극적으로 활용되고 있다. 또한 보건·복지 부문이나 범죄 예방과 안전 부문, 그리고 최근에는 교통 정책 분야에서도 활발하게 적용되고 있다. 또한 시민들의 필요와 요구사항을 좀 더 빠르고 정확하게 이해하고 예견하기 위해서도 빅데이터 분석을 적극적으로 활용하고 있다. 따라서 산재되어 있는 방대한 데이터를 유의미한 자원으로 전환하기 위한 핵심은 데이터에 대한 분석역량의 확보라 할 수 있다.

분석은 주어진 값, 즉 데이터를 내가 파악하려고 하는 기본적 틀 위에 올려놓고 이에 대한 해석 작업을 통해 유의미한 판단을 수행하는 것이다. 따라서 빅데이터 활용의 초석은 결국 분석의 축을 얼마나 정교하고 안정적으로 만드느냐에 달려 있다고 해도 과언이 아니다. 2012년 오바마 대선캠프가 빅데이터를 활용한 선거 운동을 진행할 때 가장 먼저 확보했던 정보는 유권자 정보였고 이 유권자 정보의 정확도를 높이기 위해 많은 노력을 기울였다. 이 과정에서 미네소타의 할머니들이 주도적으로 참여하여 지역을 순회하며 수소문을 통해 가장 정확한 유권자 정보를 제공했던 일화는 매우 유명하다. 이렇게 정제된 유권자 정보를 기반으로 다양한 경로를 통해 수집한 데이터를 여러 관점에서 분석하여 실제 선거운동을 집중적으로 추진해야 할 대상을 찾아내고(마이크로타겟팅), 타겟별 유권자들의 관심사와 성향에 맞는 선거 운동의 진행을 통하여 대선에서 승리했던 일은 빅데이터 활용의 교과서와 같은 사례다.

이런 상황 속에서 데이터 기반 정부가 가장 시급히 준비해야 할 일은 바로 분석의 축이 될 수 있는 핵심 정보를 정의하고 이에 대한 데이터 품질을 확보하는 것이다. 이를 위해 가장 우선적으로 준비해야 하는 것이 기준정보 관리(MDM)이다. 기준정보는 정보시스템에서 다양한 업무를 야기하는 실체에 대한 정보를 의미하며, 정형적인 정보이든 비정형적인 정보이든 모든 데이터는 이 기준정보를 기반으로 쌓여지게 된다. 그리고 기준정보가 견실하지 않으면 데이터를 해석하기 위해 더 복잡한 알고리즘이 필요하게 되고 더 많은 노력과 시간을 소모하게 된다. 따라서 업무에 빅데이터 분석을 적용하기 이전에, 무엇이 핵심이 되는 기준정보인지 정의하고 이를 어떻게 관리하여야 할지에 대한 대안을 수립해야 한다. 기준정보는 그 특성상 다수의 사용자와 시스템에 의해 사용되고 있기 때문에 그 중요성에 비하여 관리가 쉽지 않다. 이의 개선을 위해서는 단순히 시스템 보완 측면의 접근보다는 업무적 관점에서 관리체계를 확실하게 만들어가는 노력이 중요하며, 특히 각 기관에서 공통으로 사용되고 있는 핵심 정보의 경우 단위 기관이나 부처가 아닌 범정부 차원에서의 통합적 관리가 필요하다.

둘째, 데이터 전문가가 필요하다

사과는 항상 떨어져왔다. 그러나 떨어지는 사과를 통해 인류인력을 발견한 사람은 뉴턴 밖에 없었다. 데이터는 늘 존재한다. 하지만 그 데이터 속에서 의미를 찾고 지혜를 찾아 내는 사람은 흔하지 않다. 데이터는 앞서 언급한 것처럼 현실세계 시스템으로의 투사이기 때문에 빅데이터를 활용하려면 현실세계, 즉 업무의 관점에서 데이터를 바라볼 수 있는 해안이 필요하다. 이것은 기술 이전의 문제이며 근본적인 문제이다. 이는 뉴턴과 같이 현실(업무)에 대한 많은 고민을 가지고 근본적인 문제를 탐구하며 시스템을 이해하고 있는 역량을 가진 사람만이 해결할 수 있는 문제이다. 이러한 인력은 일순간에 만들어지지 않기 때문에 지속적인 관심과 체계적인 인력 양성 프로그램의 도입이 필요하다. 준비된 전문가의 '적절하고 올바른 질문을 시스템에 던질 수 있는 역량'이야말로 빅데이터와 시스템 인프라의 가치를 극대화하는 필수적인 요소라 할 수 있다.

셋째, 궁극의 숙제, 빅데이터 생태계의 활성화

‘빅데이터 정부’라는 별칭으로도 불리는 미국의 오바마 정부는 빅데이터 생태계를 혁신하고 더욱 확신 있는 의사결정을 하기 위해 범정부 조직인 ‘NITRD(네트워킹 및 IT R&D)’ 프로그램의 빅데이터 시니어 스티어링 그룹을 통해 「연방 빅데이터 연구개발 전략계획」을 수립했다. (2016.5)에 발표하였다. 7가지로 요약된 이 전략계획의 핵심은 빅데이터 활용 기반인 기술요소들에 대한 연구개발을 수행하고, 데이터에 대한 공유·관리에 관한 정책 수립을 통해 데이터의 가치를 증가시키며, 빅데이터를 혁신할 수 있는 생태계를 지속적으로 발전시켜 나가야 한다는 것이다.



— 오바마 정부는 NITRD를 통해 국가 빅데이터 전략계획을 수립하였다.
<http://www.nitrd.gov>

최근 정부에서도 데이터 정부를 표방하며 빅데이터 생태계 구축을 위해 많은 노력을 기울이고 있다. 이러한 노력이 국가적 결실을 맺기 위해서는 미국처럼 범정부적 기구의 구성이 필요할 것으로 생각된다. 이를 통해 빅데이터에 대한 청사진을 마련하고 전반적 국가 데이터 생태계가 더욱 활성화될 수 있도록 종합적이고 지속적인 노력을 경주해 가는 것이 중요할 것이다. 이러한 노력으로 조성된 건강한 데이터 생태계를 통하여 데이터가 정보로, 정보가 지식으로, 지식이 지혜로 발전하여 갈 때 국가 행정 전반이 국민들에게 최상의 서비스와 행복한 경험들을 안겨줄 수 있을 것으로 기대한다.

② 데이터와 정치·선거

데이터 정치의 서막은 열렸다

주먹구구 선거에서 현미경 선거로

윤희웅

오피니언라이브 여론분석센터장

리서치와 빅데이터분석의 융합을 시도하는 오피니언라이브에서 여론분석센터장을 맡고 있다. 정치와 선거를 비롯한 공공영역의 데이터 분석 및 컨설팅을 실시하고 있다. 주요 일간지와 월간지에 '윤희웅의 SNS민심', '윤희웅의 선거와 경제', '윤희웅의 활을 쏘다' 등의 고정 칼럼을 기고하고 있으며 특강 강사로도 활동 중이다.

데이터와 선거는 불가분의 관계

현대 민주주의 국가에서 데이터 없는 선거란 상상하기 어렵다. 모든 유권자를 개별적으로 만날 수 없기 때문에 사회와 유권자에 대한 각종 데이터를 참조해 접근해야 한다. 유권자를 이해할 수 있는 양질의 데이터가 많을수록 효과적인 선거캠페인을 실시할 수 있다. 또 투표율과 득표율 등 선거결과 자체가 계량적 특성을 지니고 있다는 점에서도 양자는 불가분의 관계에 있다.

과거에는 데이터보다 자금과 조직이 중시되었다. 대규모 군중동원을 통해 세를 과시하는 캠페인이 대세였다. 선거법이 엄격해지고, 유권자 의식수준이 제고되면서 인위적이고 불도저식의 캠페인은 효력이 사라지고 있다. 1987년만 하더라도 100만 명 이상이 모인 유세가 여럿 존재했다. 하지만 지금은 어렵다. 뜨거운 대선이라 하더라도 여쭙다 1만 명이 넘으면 큰 뉴스거리가 될 정도로 흔치 않다. 돈과 조직원을 이용하는 대신 지금은 미디어를 통해 국민에게 메시지를 전달하는 루트에 집중하고 있다. 당연히 국민에 대한 연구가 필수적이게 되었다. 국민이 요구하는 바와 사회경제적 환경 등 특성을 알아야 한다. 또 지역별, 세대별, 경제수준별, 원적지별, 종교별, 거주형태별, 주거 점유형태별 등으로 유권자를 세분화(Segmentation)해 살펴보아야 한다.

미국에서는 국민들의 소셜미디어 데이터를 바탕으로 유권자의 취향까지 읽어내 온라인 기반의 일대일 접근방식을 선호하는 이른바 ‘마이크로타겟팅(Micro-targeting)’ 기법까지 나아가고 있다. 감에 의존하던 주먹구구식 캠페인 시대를 벗어나 데이터 기반의 과학적 정치시대로 진입한 것이다.

따라서, 누가 양질의 데이터를 많이 확보하고 있느냐, 그리고 이를 잘 분석해 활용할 수 있는지에 따라 선거캠페인의 질은 차이가 난다. 이 때문에 최근 후보자들의 캠프에는 데이터분석팀을 별도로 두는 경향이 있다. 2012년 미 대선의 오바마 캠프에서는 다음과 같은 구인광고를 냈다. “2012년 11월까지 일할 예측모델 과학자, 데이터 마이닝 분석가 구함. 목표는 오바마 대통령의 재선”. 이 팀은 성공적 평가를 받았다. 알려졌다시피 당시 데이터분석팀 인력들은 민간기업에 고액에 스카웃되거나 당시 경험을 바탕으로 데이터 전략회사를 차렸다. 정치와 선거영역의 전문인력이 민간영역의 인정을 받아 옮겨가는 일은 과거에는 거의 없었다. 새로운 데이터 패러다임이 정부와 시장, 정치와 기업, 공적 영역과 사적 영역 간의 핵심기술 공유로 단절되었던 양측을 연결하고 있다. 즉, 데이터 기반 정부가 다루어야 할 정치와 선거 영역, 이슈와 쟁점, 국민들과 이해관계자 집단 등 그 대상과 범주가 넓어지면서 기술 기반의 다양한 변화요소를 포함하게 되었다.



- CNN LIVE

| 미 대선 패러다임 변화 |



- <http://blog.dt.co.kr/blog>

선거활용 데이터:

여론조사, 소셜 빅데이터, 공공데이터

선거캠페인의 요체는 ①유권자의 요구 ②(유권자가 생각하는) 후보 본인의 강점 ③(유권자가 생각하는) 경쟁후보의 단점을 각각 밝혀 이 세 요소의 교집합을 발견하는 작업이다. 가령, 유권자들이 복지확대를 원하고 후보가 복지 분야의 전문성이 있는 반면, 경쟁후보가 복지 확대에 반대해왔다면 복지 확대를 전면에 내세운다. 아무리 안보가 중요해졌다고 해도 경쟁후보는 군장성 출신인데 무작정 안보를 전면에 내세우면 효과가 떨어지는 원리와의도 같다.

이것들을 알기 위한 데이터는 어디서 얻는가. 첫째, 여론조사이다. 전화조사 등의 정량조사(Quantitative research)와 표적집단조사(FGI) 등의 정성조사(Qualitative research)를 통해 유권자의 요구사항과 후보자들에 대한 인식을 파악한다. 둘째, 소셜 빅데이터다. 온라인에서의 소셜미디어를 통해 유권자들의 보다 심층적이면서도 즉각적인 반응을 찾아낸다. 특히, 소셜미디어는 후보자의 메시지가 제대로 효과를 얻고 있는지 파악이 용이하고, 유권자들에게 ‘먹히고 있는지’도 쉽게 점검할 수 있어 최근 중시되고 있다. 이번 미 대선에서도 소셜미디어에서 영향력을 높이려는 두 후보의 싸움이 치열하다. 셋째, 공공데이터다. 과거의 공공데이터는 지난 선거결과에 대한 통계 수치만을 의미했다. 하지만 중앙정부와 지방정부가 보유한 인구센서스, 지리정보, 통계연보, 지역별 전월세자료 등 각종 데이터들이 최근 새로운 보고로 부상하고 있다.



홍보관계자, 자원봉사자, 일반 국민 지지자 등이 오바마 대통령 캠프 사무실에서 선거 후, 투표결과에 기뻐하고 있다.

<http://www.larryelder.com/b/An-Obama-Campaign-Photo-That-Looks-Like-a-Young-Republican-Rally/888192416915360101.html>

| 오바마대통령의 재선을 가능케 했던 선거전략 |

짐 메시나	레이드 가니	버락 오바마
 시카고 선거캠프 본부장 데이터마인딩 분석가 구인 목표는 오바마 대통령 재선 선거는 2년 전 데이터 분석팀 “우리는 선거운동의 일거수일투족을 수치화했다”	 빅데이터 분석 전문가 소비자 구매 패턴 분석가 슈퍼마켓 매출 상승 기여 “유권자를 철저히 분석해 특정그룹에 어떤 방식으로 접근해야 하는지 세부 전략만들어 보고”	 데이터분석팀 보고서 백악관 오바마에게 보고 캠프 내부 비판보다 데이터분석팀 전략 수용 “정치 역사상 최고의 선거팀”

— <http://sens.tistory.com/243>

8번의 낙선경험이 있는 아브라함 링컨은 “나무를 자르는데 6시간 주어진다면 4시간을 도끼를 가는데 사용하겠다(Give me six hours to chop down a tree and I will spend the first four sharpening the axe)”며 준비를 강조한 바 있다. 최근 신뢰할만한 데이터의 출현은 선거준비를 이전과 다른 차원에서 할 수 있게 해주고 있다. **선거캠페인을 위한 데이터 중 공공데이터는 그간 경시되어왔다. 공개된 데이터가 부족하기도 했고 ‘살아있는 데이터’라는 신뢰가 없었기 때문이다. 하지만 각종 통계자료 구축 및 공개 흐름은 데이터 기반 정부에 게 정치와 선거영역에서의 지능적인 접근 기회를 제공하고 있다.**

| 소셜미디어 팔로워 규모(Hillary Clinton vs Donald J. Trump) |

	Hillary Clinton	Donald J. Trump
페이스북	5,928,410	10,713,645
트위터	8,847,608	11,588,375
인스타그램	2,130,885	2,462,769
유튜브 채널	71,847	54,638

2016년 9월 18일 기준

공공데이터 개방 확대는 선거의 공익기능 강화

데이터는 단지 선거과정에서만 의미 있는 게 아니다. 선거에서 뽑히는 대표들의 정책 우선순위를 결정하는 과정이기도 하다. 후보자와 캠프 그리고 정당은 선거에 심혈을 기울인다. 온 에너지를 동원하여 공약을 정리해 내놓는다. 이때 이른바 정책마켓(Policy market)이 열리게 된다. 이런 상황에서 유권자의 요구, 사회의 문제를 부실한 데이터로 인해 파악하지 못하면 이때 내놓는 공약도 시의성과 적실성이 떨어질 수밖에 없다. 이런 의미에서 최근 정부3.0으로 양질의 공공데이터가 더욱 폭넓고 깊이 있게 공개되면 선거의 공익실현 기능도 더욱 강화될 것이다.

| 선거 활용 데이터 분류 |

no	구분	주요 내용
1	여론조사	전화조사 등 정량조사, 표적집단조사 등 정성조사
2	소셜빅데이터	후보자에 대한 소셜미디어 분석자료, 후보자별 소셜미디어 팔로워 규모 및 호응도 등
3	공공데이터	역대선거결과, 인구센서스, 통계지리정보서비스자료, 국가통계포털 자료, 지역별 전월세 비율 자료 등

③ 데이터와 경제·산업

미래정부가 만드는 기술복합적 생태계 “데이터 기반 경제”

Mark Koh

프로스트 앤 설리번 수석 컨설턴트

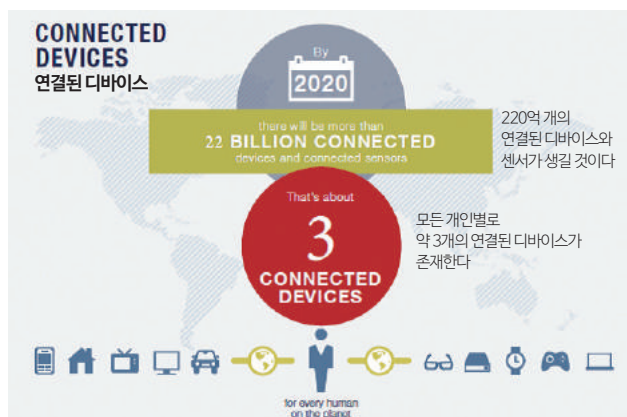
빅데이터 분석 및 AI, 사물 인터넷, 모바일, IP 네트워크 설계 및 아키텍처 등 이동 통신 산업에 17년간 종사하였으며, 루센트 테크놀로지(Lucent Technologies)와 알카텔(Alcatel-Lucent)을 거쳐 프로스트 앤 설리번에 근무하고 있다. 이동통신 분야의 기술 및 운영 경험 조사와 빅데이터 분석 및 사물 인터넷 기술 & 시장 이해, 계량 경제학 및 통계학 지식, 엔드 유저 연구 방법론 등을 전문으로 한다.

디지털 전환과 데이터의 증가

데이터는 디지털 경제의 통화로써 빠르게 자리매김하고 있다. 기업들은 전통적으로 해오던 신체적 활동 중심의 투입 대신, 데이터를 투입함으로써 그들의 서비스와 비즈니스 모델을 새롭게 혁신하고 있다. 오늘날, 거대한 규모를 가지고 성공가도를 달리고 있는 기업들은 데이터, 정보 기반의 기업들이며, 이들은 데이터를 통해 새로운 상품과 서비스를 만들거나 경쟁자보다 우위를 점하기 위한 경쟁력을 확보하고 있다. 실제로, 구글(Google), 우버(Uber), 페이스북(Facebook) 등과 같은 기업들의 수익은 아시아 태평양 지역의 일부 국가들의 국내총생산(GDP)을 능가한다.

Digital transformation and the growth of data

Data is fast becoming the new currency in the digital economy. Enterprises are increasingly reinventing their services and business models using data as inputs, rather than traditional raw physical inputs. Many of the largest and most successful enterprises today are data and information driven companies, either leveraging data as a competitive advantage to outperform their competitors or creating new products and services from data. The combined revenue of some of these enterprises like Google, Uber, and Facebook are larger than the Gross Domestic Product of several countries in Asia Pacific.



새로운 수익의 원천, 혹은 경쟁력 있는 자원으로써 데이터 활용은 스마트 시티에서 사물인터넷 현상에 따라 디지털화되고 연결된 물리적 세상에서 더욱 중요해질 것으로 기대된다.

2020년까지 다양한 산업부문 간 220억 개의 센서와 디바이스들이 네트워크로 연결될 것이다. 이러한 연결된 센서와 디바이스의 확산은 소셜미디어 데이터를 대신할 머신 데이터의 급속한 성장을 가져올 것으로 예측된다.

The importance of using data as a competitive advantage or as a new source of revenue is expected to grow as more of our physical world gets connected and digitalized in the Internet of Things phenomenon and in smart cities. By 2020, it is expected that there will be over 22 billion sensors and devices that will be connected to the network across different industries. This proliferation of connected sensors and devices is expected to drive exponential growth in machine data which will supersede social media data.

데이터가 만드는 비즈니스 모델

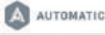
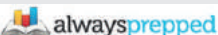
기업들은 새로운 디지털 경제체제에서 고객을 위한 가치창출의 혁신적인 방법을 찾고 있다. 많은 기업들이 데이터를 활용해 열린 경제에 적합하거나 기존에 없던 새로운 상품과 서비스를 만들고 있다. 예를 들어, 우버는 각 개인들이 모바일 기반 통신 데이터를 활용해 원하는 택시 서비스를 이용할 수 있도록 해준다. 즉, 수요와 공급에 따라 다르게 적용되는 비용과 택시 승차 경험을 연결시킴으로써 고객들에게 편리한 서비스를 제공하는 것이다. 이러한 트렌드는 열린 경제에서 활발히 진행 중이며, 특히, 의료, 물류 산업 부문에서 활성화되고 있다.

사물인터넷 또한, 이와 유사하게 기업들의 비즈니스와 상품, 서비스 생산방식을 바꾸고 있다. 본질적으로 사물인터넷은 기업을 위한 빅데이터 시나리오와 같다. 수천억 개의 디바이스와 센서 간 연결은 방대한 데이터를 만들고, 이런 데이터가 분석되어 적절히 활용되지 못한다면 무의미해질 수 있다. 사물인터넷에서 생성된 방대한 데이터를 인지하면서부터, IBM, 마이크로소프트(Microsoft), 보쉬(Bosch), GE 등 많은 선진 기업들은 이미 대규모 투자를 진행하고 있다.

Data enabled business models

In the new digital economy, enterprises are finding innovative ways to create value for their customers. Many of these enterprises make use of data to enable them to create new services and product offerings that previously did not exist or to tap into the open economy. For example Uber provides a platform that enables individuals to offer taxi services to others by leveraging mobile and telematics data. Uber uses this data to offer convenience services to its customers by enabling them to control their taxi ride experience, to match demand and supply, and to provide differentiated pricing. Similar trends to tap on the open economy are happening in the hospitality industries and logistics industries.

The Internet of Things is similarly transforming the way enterprises conduct their businesses and offer their products and services. In essence the Internet of Things is a Big Data scenario for enterprises. The connecting of thousands of devices and sensors will generate large volume of data, which in itself is a useless exercise unless this data is analyzed and properly harnessed. Recognizing the value of data from the Internet of Things, many leading enterprises have already made large investment in their respective domains, including IBM, Microsoft, Bosch and GE.

데이터 소스 Data Sources	주요 활동 Key Activity	데이터 종합 Aggregation	데이터 분석 Analytics	데이터 생성 Data generation
추적 및 생성된 데이터 Tracked and generated			데이터 생성 및 분석 이 카테고리 분류된 기업들은 외부 데이터 분석에 집중하기 보다 데이터 생성에 관심을 갖는다. 기업들은 클라우드 소싱, 웹 분석 기업, 연결된 디바이스 등에서 만들어진 데이터로 구분된다. data generation and analysis Companies in this category generate data rather than relying on external data and analyze this data. Companies are segregated by data generated from crowdsourcing, web analytics, connected devices, connected devices generated data.	  
고객이 제공한 데이터 Customer provided data	서비스로써 데이터 종합 기업들은 고객을 위한 다양한 내부 소스로부터 얻은 정보를 종합, 수집한다. Data aggregation as a Service Companies collect & aggregate information from multiple internal sources for their customers.		서비스로써 분석 기업들은 고객에 의해 제공되는 데이터에 기반하여 분석결과 및 인사이트를 제공한다. Analytics as a service Companies provide analytics and insights based on data provided by the customers.	매쉬업 및 분석 기업들은 무료 이용 및 내부 고객의 데이터 등 다양한 자원으로부터 얻은 데이터를 종합하고, 결합된 데이터셋에 대한 분석을 수행한다. Mash up and analysis Companies that aggregate data from a variety of sources, both freely available analytics and internal customer data, and perform analytics on the combined data set.
공공데이터 Public Data	종합자료를 무료 데이터 기업들은 이용가능한 데이터로부터 정보를 수집, 종합하고 데이터 배분을 제공한다. Free data aggregator Companies collect & aggregate information from available sources, & provide data distribution.	 	무료 데이터 지식 발견 기업들은 비정형의 포맷에서 이용가능한 무료 데이터 분석을 돕는다. Free data knowledge discovery Companies help analyze freely available data, which might be in a non standard format.	      

Adapted from Cambridge University

사물인터넷 데이터는 기업들이 서비스를 제공하는 방식을 바꾼다. 미쉐린(Michelin)은 영업부터 서비스 제공까지 사물인터넷을 통해 비즈니스 전환을 진화시키고 있다. 스마트 센서에 따른 데이터는 운전자 모니터링 서비스나 자동차를 위한 서비스 부가가치를 높여 미쉐린을 위한 새로운 기회를 제공한다. 또한, 웨어러블 형태의 사물인터넷 데이터는 이전에는 이용이 어려웠던 행태와 습관 등을 수량화할 수 있는 기회를 제공한다. 보험회사들은 운전습관부터 운동습관까지의 데이터 분석을 통해 보험료 책정에 활용할 수 있다.

특히, 개인화된 데이터 등을 현금화하는 것은 새로운 일입니다. 과거에도 온라인 기업, 광고, 보험, 금융 등의 산업 부문에서는 데이터를 통해 수익을 창출해왔다. 데이터가 순조롭게 이용이 가능해짐에 따라, 더 많은 산업 부문에서 데이터가 수익창출의 보고가 될 것이다. 이는 위 그림과 같이 기업들이 어떻게 그들의 고객과 조직을 위한 가치창출에 데이터를 활용하는지를 잘 보여준다.

The data from the Internet of Things has enabled companies to change the way they offer their services. As an illustration, by leveraging the Internet of Things, Michelin is able to evolve their business from selling of tyres to offering of tyres services. The data from smart sensors opens up a new universe of opportunity for Michelin, enabling them to provide value added services beyond pay per use tyres, to providing safety services for vehicle fleet or driver monitoring services. Similarly, data from the Internet of Things, in the form of wearables, are now being used to quantify behavior and habits that were previously not available. From driving habits to exercising habits, many insurance companies are now leveraging such data to offer premium adjusted usage based insurance.

The monetizing of data, in particular personal data, is not new. The online, advertising, insurance, and finance industries have been particularly active in the use of data to generate revenue in the past. As data become more readily available, more and more industries will look to data as a source of revenue. This is best illustrated in the diagram, which depicts how enterprises are leveraging data to create value for their customers and their organization.

스마트 시티와 정부에 관한 데이터

지방의 도시기획자들은 투자를 유치하고 혁신을 추진하기 위해 더 효율적이고 안전하며, 지속가능한 스마트 시티로의 여정에 박차를 가하고 있다. 스마트 시티 발전의 일환으로, 도시기획자들은 도시 인프라에 센서를 내재화시켜 데이터를 수집, 관리, 분석함으로써 다른 정부기관 및 일반대중이 이용하도록 하는 등 도시를 가상화시키고 있다.

Data in the context of smart cities and government

Many city planners in the region are embarking on their smart cities journey, with the goals of making their cities more efficient, safer, sustainable, to attract investments, and foster innovation. As part of their smart city development, city planners are virtualizing the city by embedding sensors into the infrastructure, collecting, managing and analyzing the data, and making it available to different government agencies and to the public. This is illustrated in the following diagram.

데이터 활용을 촉진하는 매쉬업과 어플리케이션(오픈 데이터)

Mash ups and applications leveraging data (open data)



Mobility
apps



Portal
apps



Citizen
apps



Developer
platform



Industry
apps



데이터 종합 및 조정을 제공하는 총괄 단계의 플랫폼과 미들웨어

Platform and middleware for orchestration layer providing data aggregation & control



Homes



Public Infra



Comms



Utilities



Social media



Manufacturing



센서를 통한 도시의 가상화 실현

Virtualizing the city through sensors



도시기획자들은 데이터에 대한 유의미한 분석 및 협업 촉진을 위해 수집된 데이터 관리와 공유를 고민해야 한다. 현재 축적된 데이터는 여전히 개별 조직과 정부기관 내의 사일로에서 만들어지며, 기존 프로세스에 통합되지 않고 있다. 이는 관련 정보가 생태계에서 주요 이해관계자들이 수월하게 접근·이용할 수 있는 스마트 시티를 만드는 데 장애 요인이 된다. 스마트 시티 기획자는 다른 정부기관 간, 그리고 국민에게 오픈 데이터 개방을 촉진할 견고하고 안전한 플랫폼/미들웨어를 구축해야 한다. 이러한 플랫폼/미들웨어는 개별 국민과 기업들이 데이터를 가지고 혁신할 수 있는 생태계를 실현하고, 정부 조직 간 향상된 효율성을 위한 기반을 만들 것으로 기대된다.

The challenge for many city planners is the managing and sharing of the data collected to facilitate collaboration and meaningful analysis of the data. Today data collected currently still reside in silos within individual organizations and government agencies, and not well integrated into existing processes. This is a barrier to implementing a smart city where relevant information is made readily available to key stakeholders in the ecosystem. The challenge for smart city planner is to create a robust and secure platform/middleware that will facilitate the sharing of information across different government agencies and to the public in the form of open data. This platform/ middleware will lay the foundations for improved efficiencies across government entities and to foster an ecosystem where individuals and businesses can innovate on data.

정책적 제언

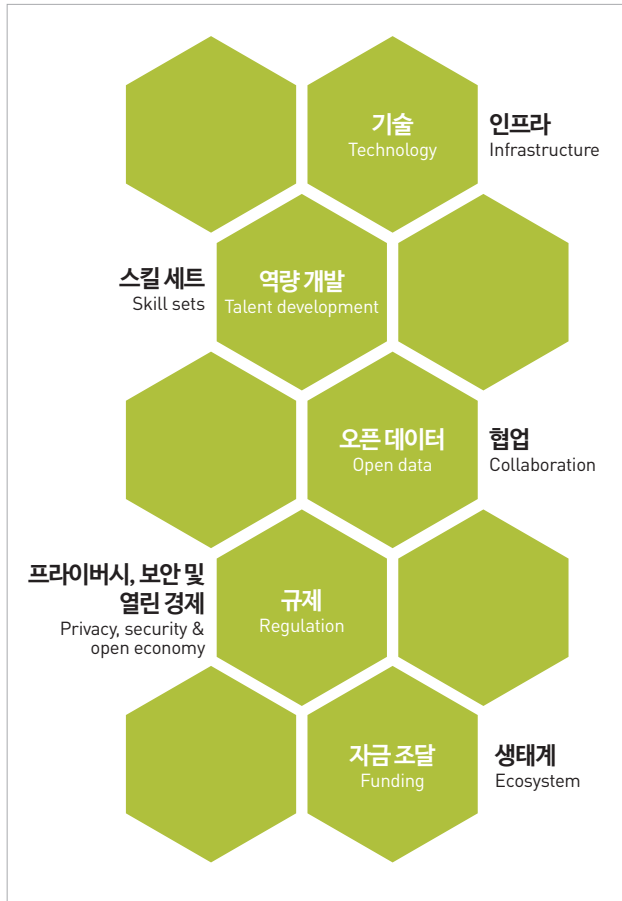
데이터 기반의 경제를 발전시키기 위해서는 혁신을 조정하고 새로운 상품과 서비스 창출을 촉진하는데 포괄적인 생태계 및 관련 기반이 요구된다. 정부와 규제자들의 역할은 이러한 기반을 형성하고 생태계 발전을 실현하는데 중요하다.

디지털화된 사물과 디바이스가 더 많이 연결될수록 보안은 우리의 일상을 위협하는 잠재적 요인이 된다. 과거의 보안 문제는 금융자산과 지적재산권을 보호하고 커뮤니케이션을 방해하는 일 등과 관련되었지만, 새로운 디지털 경제와 사물인터넷 시대에는 사이버 범죄가 현실적인 위해 원인이 된다. 또한, 도시가 연결됨에 따라 중요한 인프라와 디지털 서비스는 사이버 공격에 취약해진다. 안정적이고 복원력이 우수한 보안 솔루션, 정책을 만드는 것은 디지털 서비스 이용을 촉진하고 서비스 장애요인의 영향을 줄일 수 있다.

Final word

To develop a data driven economy a comprehensive ecosystem and foundation is required to foster innovation and drive the creation of new products and services. The role of the government and regulators is important to lay the foundation and to nurture the development of such an ecosystem.

As more physical objects get digitized and devices get connected, security becomes a paramount concern as the potential to disrupt our daily lives increases. Security concerns in the past predominantly revolve around protecting financial assets, intellectual property and disrupting communications, however in the new digital economy and the Internet of Things cyber criminals have the ability to cause physical harm. As cities become hyper connected, critical infrastructure and digital services become increasingly vulnerable to cyber-attacks. Ensuring robust and resilient security solution and policies is important to encourage adoption of digital services and reducing the impact of disruption of services.



데이터 프라이버시 역시 중요한 문제로, 데이터와 디지털 서비스 수요와 관련된다. 많은 지방정부들은 새로운 디지털 경제를 준비하는데 그들의 데이터 보호 법제를 간간하게 정비하고 수정하기 시작했다. 이런 규제들은 개인정보에서 데이터 활용과 관리에 대한 명확한 가이드라인을 제공하고 데이터 타협에 관한 책임과 의무를 다룬다. 시너지 효과를 내려면, 빅데이터 전략, 스마트 시티 전략, 보안 규제, 열린 경제 관련 규제 등과 같은 정부기관 간 다른 전략과 규제에 대해 조정하는 것이 중요하다.

Data privacy is becoming a key concern and needs to be addressed to encourage the adoption of big data and digital services. Many governments in the regions have started amending and tightening their personal data protection laws in preparation of the new digital economy. These regulations provide explicit guidelines on the use and manage-

ment of data, in particular personal data, and the responsibilities and liability associated with data compromise. It is also important ensure coordination across different initiatives and regulation across government agencies, such as big data initiatives, smart cities initiatives, security regulation, and open economy regulations, to ensure synergies are achieved.

관련 역량도 데이터 기반 경제를 견인하는 것의 장애요소 중 하나로 남아 있다. 즉, 이런 경제를 이끌기 위해 요구되는 스킬 (Skill)을 가진 지역 인재가 부족하다. 따라서, 민간 및 공공부문은 시장의 수요를 충족시킬 수 있는 전문가 인증과 학문적 프로그램 등과 같이 데이터 과학 및 데이터 엔지니어링을 함께 도모해야 한다. 중요한 빅데이터 인프라 발전을 견인하고, 빅데이터, AI기술 등을 활용하는데 요구되는 생태계 발전을 위한 투자형태로서의 정부 참여 또한 중요한 부문의 데이터 기반 경제를 구축하는데 필요하다. 공급과 수요 측면의 인센티브는 인력의 재개발, 세금 인센티브나 R&D 보조금, 새로운 기술 활용 등을 위한 지원이 될 수 있다.

Relevant talent remains one of the largest hurdles to the adoption of data led economy. There is an acute shortage with candidates in the region with the necessary skill set to fuel this economy. Private sector and public sector collaboration is needed to jointly create necessary programs, such as data science and data engineering professional certification and academic programs, to meets the need of the market.

Government involvement in the form of funding to develop the necessary ecosystem, in driving the development of necessary big data infrastructure, and to encourage the adoption big data and AI technologies, is also key to developing and shaping a data driven economy in key sectors. Supply and demand side incentives given can be in the form of subsidies for redevelopment of manpower, and tax incentives or subsidies for R&D and adoption of newer technologies.

Lastly, the development of a data ecosystem can be led by the government by its promotion and implementation of a coherent open data initiative. It will be important that the government embraces a data driven culture and promotes the sharing of data amongst its agencies, and also to share data and communicate with the public via its open data initiative.

④ 데이터와 사회·문화

디지털과 아날로그가 결합된 사회,

데이터는 올바르게 사회현상을 다룰 수 있는가?

황철현

데이터마루 이사

데이터 전문기업인 데이터마루에서 데이터 아키텍처와 빅데이터 기획에 대한 컨설팅, 정부 연구과제를 수행하고 있다. 현재 국민대학교 BIT전문대학원에서 겸임교수로 재직 중이며 한국데이터베이스진흥원의 빅데이터 아카데미 등에서 빅데이터 기획과 데이터 분석에 대한 강사로도 활동 중이다.

데이터는 주관적인가, 객관적인가?

‘데이터(Data)’는 라틴어인 dare(주다, to give)의 과거분사형으로서 ‘주어지는 것’이란 의미로 처음 사용되었다. ‘주어지는 것’이란 관찰자의 의도에 의해 주관적으로 파악되는 것이 아니라 관찰 대상이 스스로 내어주는 것으로 사실이나 객관적인 의미로 주로 사용된다. 이로 인해 우리는 데이터가 그 자체로서 객관적이라는 전제를 가정한다. 하지만 데이터는 과연 객관적인가?

취재현장에서 사진기자가 어느 위치와 방향에서 어느 정도의 노출을 주고 사진을 찍는가에 따라 사진이 주는 메시지가 완전히 바뀌는 것처럼 데이터도 관찰자(데이터 생산자)나 모델러(구조 설계자)의 의도에서 자유로울 수 없다. 이것은 데이터 생산자나 설계자의 의도와 무관하게 어떤 사물이나 현상의 모든 정보를 파악하거나 기록할 수 없고, 이 가운데 일부를 선택하는 과정에서 본질적으로 주관적인 특성을 배제할 수 없음을 의미한다. 나아가, 데이터의 양이 아무리 많아지고 다양하게 수집된다 하더라도 그 자체만으로는 객관적일 수 없다는 사실도 시사한다. 이러한 데이터의 주관적인 특성은 사회현상에 대한 연구방법이나 이론이 정당하더라도 결과를 보장할 수 없는 문제를 야기시킨다.

빅데이터의 신화를 다룬 여러 논문에 의하면, 대부분의 사람들은 막연한 기대로 데이터가 커지면 그 자체로서 객관적이고 정확해진다고, 창의적인 아이디어가 도출되고 새로운 지식으로 발전할 것이라 생각한다.¹⁾ 이러한 기대는 특히, 복잡하고 해결이 어려운 문제들이 뒤섞여 있는 미래사회에서 데이터 기반 정부가 사회현상을 파악하고 문제를 해결할 때, 매우 위험할 수 있다. 우리사회에서 데이터는 단순히 현상을 나타내는 언어가 아니라 사회현상을 결정하는 사회 구성원이기 때문이다.



— 빅데이터 분석을 통해 2016년 사회 분야 이슈를 도출한 결과이다.
한국정보화진흥원, NEAR & Future Monthly 1월호, 2016

또 다른 사회구성원, 데이터

어떤 사실과 그 이면의 의미까지 아울러야 사회를 이해한다

데이터가 사회를 구성하는 일원이 된 것은 데이터의 중요성이 증가되면서부터다. 연구자나 정책입안자, 의사결정자에게 있어 자신이 속한 사회(연구나 비즈니스 대상) 자체를 이해한다는 것은 조직의 가치이자 필수적인 생존능력이 되었다. 따라서 우리가 속해 있는 사회를 이해하고 설명하기 위해 오랫동안 각종 이론들을 발전시켜왔고 더 많은 자료(데이터)를 축적하기 위해 노력해 왔다. 또한 사회구조가 복잡해질수록 이해관계자나 환경 요인들이 증가하고 이를 해석하는 이론들 역시 많은 변수를 추가함으로써 사회 현상을 설명하는 방향으로 진행되어 왔다. 이러한 복잡한 구조는 디지털화된 데이터를 기반으로 하지 않으면 사회 현상을 이해하고 해석할 수 없고 이는 곧 생존과 직결된다는 인식으로 이어졌다.

1) 한신갑, 「빅데이터와 사회과학하기: 자료 기반의 변화와 분석 전략의 재구성」 (한국사회학49(2), 2015), pp. 161~192.
Boyd, Danah, and Kate Crawford. [2012]. Critical Questions for Big Data (Information, Communication & Society. vol.15. No. 5)

데이터가 의사결정의 중요한 변수로 자리매김하면서 데이터가 존재하는지, 어떤 방법으로 생산되었는지, 누가 생산했는지, 얼마나 유통되고 있는지가 사회현상을 결정짓는 한 요인으로 작용하게 되었다. 즉, 데이터 자체가 이미 사회 구성원인 셈이다. 이는 특정한 사회문제에 대한 의사결정을 논하기 전에 데이터가 생산되기 이전과 이후의 사회변화 가능성을 고려해야 함을 의미한다. 특히, 다른 생산자의 데이터를 활용하는 빅데이터 시대에서는 이 문제를 더 깊게 고려해야 한다. 따라서 과거의 사례를 잘 설명한다고 해서 미래를 예측하는 수단으로 충분할 수 없다는 점, 데이터의 주관적인 특성이나 생산자의 의도를 배제할 수 없다는 점, 분석·예측된 데이터에 기반해 내려진 의사결정과 실제 사회문제 간 구조적 관계의 불명확성 등을 충분히 고려하지 않고서는 데이터를 기반으로 한 사회현상 분석과 참여는 매우 위험하게 전개될 수 있다.

데이터를 더 이상 사물이나 사건에 대한 객관적인 사실로만 보지 않고 주관적 견해를 담은 주장으로 본다면 ‘데이터가 어떤 대상과 사실을 기록하고 있는가?’ 보다 ‘누가 어떤 환경에서 기록했는지?’에 더 관심을 가져야 한다. 이것은 특정사건을 기록한 데이터를 그대로 해석하는 것이 아니라 사건의 참여자나 이해당사자들의 주장임을 전제로 하고 데이터에서 지칭하는 사물이나 현상을 뛰어넘어 존재하는 실체를 발견하고 이와 관련된 모든 데이터를 종합함으로써 사회 자체를 이해하는 과정이다.

이 과정에서 주관적인 것과 객관적인 것을 분리하는 ‘해체’, 객관적인 사실들을 하나의 주제로 다시 모으는 ‘종합’, 처음과는 다른 주관적인 영역으로 ‘확산’이 이루어진다. ‘해체-종합-확산’의 과정을 거치는 이유는 해체와 종합, 확산을 이어주는 것은 시간이나 현상을 가로지르면서 존재하는 실체(또는 객관적인 사실)를 파악하기 위해서인데, 이것이 우리가 이해하고자 하는 사회이고 세계이다. 세상(주관)에서 수집된 데이터를 통해 세계(객관)를 이해하고 다른 세상(주관, 시간적이든 공간적이든 개념적이든)으로 인식을 확장할 수 있을 때 데이터를 통한 사회 변혁과 참여를 정당화할 수 있는 근거를 가진다. 이것이 바로 빅데이터에서 가장 중요한 요소인 ‘통찰력(Insight)’을 얻는 방법이다. 이런 과정 없이는 ‘과거의 현상을 잘 설명한다고 해서 미래를 예측하는 능력이 있는 것이 아니란 것’과 ‘한 현상에 대한 설명력이 다른 유사한 현상을 설명하는 것의 충분한 근거가 되지 못한다’란 문제를 넘어서지 못한다.

데이터로 문제해결을 원하는가?

인문학에서 배운 참된 질문을 먼저 해라

데이터 생산자가 모든 데이터를 생산할 때 기본적으로 작용하는 주관성(가치관)과 특정 데이터를 생산할 때에만 영향을 미치는 주관성(의도)을 구분하고 이를 종합할 수 있을 때만이 데이터를 통해 사회를 이해하는 것이 설득력을 가질 수 있다. 이것은 보다 보편적이고 메타적인 학문의 영역이다. 데이터 분석의 대상이 사회이며 이것은 인문학의 영역이라는 것만으로 이해의 대상이 되지만, 데이터와 관련된 각종 기술적 측면에서도 인문학을 기반으로 한 구체적인 내용들을 접할 수 있다. 빅데이터 또는 데이터베이스 등에서 사용하는 핵심적인 용어가 인문학(특히 철학)에서 매우 중요히 다루는 개념어와 동일하게 사용하고 있고, 데이터의 모델 구조가 인식론, 구조주의, 포스트모더니즘과 같은 사상구조와 일치하는 사례들을 찾을 수 있다. 데이터는 현실세계의 무엇인가를 바라보는 관점과 기술(記述)의 집합체이기 때문에 현실세계를 연구하는 인문학과 연계되는 것은 어쩌면 당연한 것이다. 따라서 데이터 기반 정부(조직, 공무원 등)가 데이터를 이용해서 사회를 이해하고 사회참여의 변혁에 대한 정당성을 확보하는데 인문학은 구체적인 방법과 정당성을 제시한다. 이것은 선택이 아닌 필수적으로 가져야 할 능력이다.

결론적으로 데이터 양의 증가와 처리 기술의 발전이 사회 발전에 기여하고 새로운 가치를 창출할 것이라는 막연한 기대는 매우 위험한 발생이다. 오히려 특정 주장을 뒷받침하는 잘못된 근거로 활용되거나 원하지 않는 방향으로 사회현상을 유도하는 범인이 될 가능성이 높다. 따라서 데이터를 통해 사회를 이해하고 사회의 변혁을 이끌 수 있는 힘은 데이터를 처리하는 기술이나 방법에서만 도출되는 것이 아니며 기존의 사회현상을 분석하는 방법과 이론에 국한되어서도 안 된다. 두 영역의 결합은 결국 분석 대상이 되는 사회 자체에 대한 이해를 증진시킬 수 있는 근거 학문에서만 찾을 수 있음을 다시 한 번 되새겨야 한다.



- 2010년 스티브 잡스가 아이패드를 소개할 때 인용한 발표영상의 한 장면으로, 그는 기술과 일반 교양의 교차점에 서 있으려고 노력한 덕분에 창의적인 제품을 만들 수 있었다고 설명했다. 기술에 앞서 사람과 사회를 전반적으로 이해하고 포용하는 시각이야말로 세상에 없었던 새로운 질문과 답을 찾을 수 있음을 시사한다.

<https://brunch.co.kr/@neo3xdh/20>

국내외 데이터 전문가들이 주목한 데이터 기반 정부의 기회와 도전 |

[D.gov 인물포커스]

김진형 지능정보기술연구원 원장 / 공공데이터전략위원회 공동위원장

‘感’ 대신, 데이터로 합리적 의사결정을 해야죠!

“10%의 부작용에 두려워 말고, 90%의 부가가치 창출에 몰입해야 합니다.
이 때 정부는 너무 많은 일을 하려하지 말고, 민간에게 데이터를 깔아줘야 합니다.”



김진형 원장은 국내 최초로 인공지능 박사학위를 받고, IBM Watson 연구소, KAIST 등을 거쳐 40년 이상 교육과 연구개발에 힘써 왔다. 지난 2월부터 공공데이터전략위원회 공동위원장을 맡고 있는 동시에, 소프트웨어정책연구소 소장을 거쳐 현재 인공지능 등 지능정보기술을 개발 확산하기 위한 지능정보기술연구원 소장으로써 미래 핵심기술 발전 전략을 구상하고 있다.

미래 신기술을 연구 개발하는 연구소 원장이자, 공공데이터 개방과 활용을 촉진하는 위원회장으로서 동시에 두 가지 역할을 하고 있는 김진형 원장은 데이터와 인공지능 기술이 같은 맥락에 있다고 운을 뚫었다. 미래의 지능정보사회에서 데이터와 기술은 서로 상호작용하면서 발전해 나가야 실질적인 시너지를 낼 수 있다는 말이다.

“데이터 기반 정부요? ‘데이터’와 ‘정부’라는 단어가 붙어 있을 때 키워드는 데이터를 이용한 의사결정의 합리화죠! 단순히 데이터의 경제적 가치를 논하는 것을 넘어, 정부가 양질의 의사결정을 하는 것에 중점을 두어야 합니다. 즉, 정부는 감(感)에 의해 의사결정을 내리는 시대가 아니고, 빅데이터 분석에 의해 보다 합리적인 의사결정을 해야 한다는 의미죠. 이때 합리적 의사결정까지 가는 과정에는 데이터 수집과 분석 기술이 필요합니다. 결국, 데이터나 인공지능은 같은 맥락의 이야기입니다.”

Q 데이터 패러다임의 변화와 정부의 역할이 궁금하다

“과거에는 ‘컴퓨터화’라고 해서, 컴퓨터 안에 들어 있는 문서에 관한 데이터 이야기를 많이 했죠. 그런데 지금은 현장에서 센서들을 통해 직접 수집하는 살아 있는 데이터가 중요해졌습니다. 이전 데이터 패러다임에서는 정제되었던 정형 데이터가 중심이 되었다면, 미래에는 방대한 로우 데이터(Raw data), 즉, 비정형 데이터가 주를 이룰 것입니다. 이에 따라, 더 방대한 비정형 데이터를 다루기 위한 분석기술도 함께 발전하게 될 것입니다.”

김진형 원장은 새로운 데이터 패러다임을 고려한다면, 정부의 역할도 변해야 한다고 강조했다. 즉, 정부가 나서서 정책을 만들고 추진해야 한다는 과거 방식은 지속적인 효과를 내기가 어렵다는 것이다. 방대한 데이터 홍수 속에서 정부가 취급하고 다룰 수 있는 데이터에는 한계가 있기 때문이다.



— 김진형 원장은 2016년 2월부터 공공데이터전략위원회 2기 공동위원장으로 활동하고 있다. 2기 위원회는 더욱 활발한 공공데이터 개방을 통한 유의미한 서비스 확산은 물론, 개인 데이터에 대한 활용과 보호에 관한 균형을 찾고자 노력하고 있다.

<http://www.odsc.go.kr/home/homeIndex.do>;

<http://blog.daum.net/jkim.kaist/124>

<http://www.datanet.co.kr/news/articleView.html?idxno=104632>

“미래 데이터 패러다임을 고려한다면, 정부가 주도적으로 무엇을 해야 한다면 시각은 바람직하지 않습니다. 향후에는 ‘바텀-업(Bottom-up)’ 방식의 데이터 정책이 활성화되어야 합니다. 정부는 이를 통해 국민이 리드하는 데이터 소비사회 기반을 만들 수 있어야 합니다. 정부가 공공데이터 개방, 공유, 활용을 더욱 촉진해야 함은 당연하며, 이를 통해 국민의 창의성 발현과 산업발전, 나아가 모두가 정보를 누리고 소통하는 민주주의 정신에도 기여하게 될 것입니다.”

“또한, 공무원이 아닌, 국민들이 데이터로 필요한 서비스를 직접 만들 수 있는 환경이 구현되어야 합니다. 즉, 고객 입장에서 가려운 데를 긁어주는 서비스가 되려면 공무원이 고민해서는 안됩니다. 정부가 서비스를 기획하는 게 아니라, 데이터라는 판을 깔아주면 국민과 민간기업의 아이디어로 서비스가 만들어지고 이를 구매하는 것이 정부역할이 아닐까 생각합니다. 영국은 이미 정부가 여러 가지 앱을 거래하는 앱 스토어를 만들어서 활용하고 있으며, 일본은 신기술을 정부가 나서 제 값에 구매하고 문제점을 찾아내기도 합니다.”



**국민이 리드하는
데이터 소비사회 기반을 만들어야 합니다.
정부역할은 국민과 기업이 데이터로 만든
서비스를 구매하는 것입니다.**



데이터 활용에 따른 부작용보다 성과에 집중할 수 있는 공공의 조직문화를 법적으로 보장하려는 노력이 필요합니다.

Q 데이터 기반 정부로의 전환에 있어 문제점과 해결방안은?

김진형 원장에게 보다 구체적으로 정부의 역할 변화에 대해 물었다. 김진형 원장의 대답 속에서, 데이터 기반 정부의 걸림돌과 이에 대한 정부의 해결정책은 자연스럽게 전자정부와 연결되었다.

“우리나라 전자정부는 세계 최고 수준입니다. 현재 추진 중인 정부 3.0 역시 우수한 서비스를 제공하고 있다고 알고 있습니다. 데이터 기반 정부 성과를 가장 잘 보여줄 수 있는 건 역시 대국민 서비스 효과입니다. 우리는 이미 데이터를 통해 민원24나 연말정산과 같은 편리한 서비스를 만든 바 있습니다. 그런데 아직까지 상당히 번거로운 서류절차가 필요한 서비스도 남아 있습니다. 이는 아직까지 정부기관 간에 정부가 가진 공공데이터가 제대로 연계, 통합되지 않았음을 보여줍니다. 국민이 자신의 공공데이터를 제시할 필요 없는 서비스 제공을 위한 고민이 필요합니다. 예를 들어, 어떤 집에 남자 아이가 새로 태어난 후에 출생신고부터 입학, 나중에 군대 입대까지 알아서 찾아오는 서비스가 구현되려면 그 아이에 대한 공공데이터 관리와 활용이 관건이 될 것입니다.”

덧붙여, 김진형 원장은 이와 같은 데이터 기반의 전자정부 서비스 고도화를 위한 방안의 하나로 부처 간 칸막이 없애기를 강조했다. 실효성 높은 데이터 연계와 통합을 위해서는 사람, 조직 간 장애가 없어야 한다는 것이 그의 생각이었다. 또한, 김진형 원장은 공무원의 의식구조와 조직문화에 따른 걸림돌도 지적했다.

“공무원들이 데이터나 기술을 더 잘 쓸 수 있기 위해서는 더 많이 공부해야 합니다. 데이터는 전 부처의 모든 정책영역에 걸쳐 있는 업무자원이자 의사결정의 소스로서 학창시절의 필수과목과 같이 여겨야 합니다. 데이터 활용의 장점과 한계를 제대로 모르고서는 어떤 정책을 과감히 추진하고 장기적인 관점을 견지할 수 있겠습니까?”

“또한, 공공문화 차원에서 공무원들은 부작용이 100% 중에 10%라도 있으면 추진하기 어려운 환경에 놓여 있습니다. 데이터 활용 결과의 90%가 가져올 부가가치가 크다는 건 잊은 채, 10%의 정보유출이나 프라이버시 침해 등의 잘못된 따지기 때문이죠. 특히, 공공부문에서는 더 부작용에 민감합니다. 이런 문화에서 누가 나서서 정책을 자신 있게 이끌겠습니까?”

공무원이 업무상 열심히 공공데이터를 개방하고 활용을 촉진하면서 어쩔 수 없이 나타난 실수나 부작용에 대해 수용하고 더 큰 실익을 인정할 수 있는 인식변화가 필요합니다. 일례로, 우리 위원회에서는 관련 법에 공무원의 실수나 업무상 부작용에 대한 면책조항을 만들었습니다¹⁾. 데이터 활용에 따른 성과를 활성화할 수 있는 조직문화를 법적으로 보장해 주자는 노력이 필요하죠.”

김진형 원장은 우리나라가 세계적으로도 데이터 보호에 관한 법제가 강력한 수준에 속한다고 이야기하면서 이러한 규제는 역설적으로 데이터를 활용한 성과 역시 저조하게 만드는 원인이라고 강조했다. 마지막으로 범정부 차원의 데이터 컨트롤타워의 필요성도 역설했다.

“정부부처와 기관은 각자의 역할에 충실합니다. 예산의 적정한 사용을 감시하는 기관과 기술을 더 저렴하게 구입하고 장비를 조달하려는 기관, 신기술 개발 확산을 빠르게 촉진하려는 기관들이 제각기 소임을 다 잘하고 있습니다. 그런데, 이들을 한 데 모으면? 더 원활하게 정책이 추진되어야 하지만, 그렇지 않습니다. 데이터 정책도 마찬가지입니다. 무엇 때문일까요? 컨트롤타워의 부재가 문제입니다. 데이터가 국가사회 전반에 관여되는 만큼, 총괄적인 조정과 적정한 이용활성화 촉진을 위한 범정부 차원의 역량 있는 컨트롤타워 조직이 필요합니다.”

1) 관련 내용은 「공공데이터의 제공 및 이용 활성화에 관한 법률」의 제5장 보칙의 제36조 면책조항에 해당한다.

Q 추천할만한 데이터 활용 정책사례를 제안해 주신다면?

“얼마 전부터 우리 역사를 바탕으로 만들어진 재미 있는 영화나 드라마들이 인기가 많은 것 같습니다. 저도 봤지만, 왕의 남자나 관상, 역린 같은 대중적인 영상 콘텐츠는 흥행몰이와 함께 한류에 이르기까지 무한한 경제적 부가가치를 만들고 있죠. 그런데, 이런 것들이 가능해진 큰 이유는 일반인이 접근하지 못하던 어려운 한문으로 쓰여진 조선사 콘텐츠가 전자화되었기 때문입니다. 대표적으로 조선왕조실록을 한글번역 하고 온라인으로 데이터를 개방한 사업은 본받을 만한 일입니다. 이제 누구라도 온라인으로 언제 어디서나 저렴하게 역사적 데이터를 이해하고, 이를 통해서 다양한 상상력을 입힐 수 있게 되었습니다. 조금 지나면 어떤 내용의 제2, 제3의 왕의 남자가 탄생할지 모릅니다. 이것이 데이터의 힘이고, 정부가 해야 하는 정책사업입니다.”

“세종대왕께서 한글을 창제한 지 600여 년이 지난 지금, 과거부터 지금까지 한글로 만들어진 모든 문서의 전자화를 추진하는 것도 좋은 방법입니다. 지금까지는 주로 행정에 쓰였던 문서가 기록물로 우선 분류되어 전자화되었지만, 이는 부가가치 창출 효과가 별로 없죠. 비즈니스 경쟁력과 문화적 세계화를 추구하기 위해서는 우리만의 고유한 역사적 콘텐츠에 대한 데이터 개방사업이 더 중요합니다. 우리나라 공영방송국인 KBS가 가지고 있는 수많은 영상자료도 마찬가지입니다. 이를 테면, 이승만 대통령의 취임 연설부터 시작해서 도롱뇽이 알을 낳은 장면 같은 영상들이 개방된다면, 톡톡 튀는 아이디어를 가진 젊은이들이 생각지 못했던 스토리로 엮은 콘텐츠가 세계적인 소셜 네트워크 등을 통해 그 가치를 더 확산시킬 수도 있을 거라 기대됩니다.”

Q 유망기술로 부상한 인공지능과 데이터에 대한 견해가 궁금하다

김진형 원장은 인공지능도 결국 데이터를 모아서 머신러닝으로 가능해진다고 말하면서, 데이터가 없거나 잘못된 데이터 셋을 활용하면 인공지능도 소용없다고 설명했다.

“인공지능은 고깃국의 양념과 같습니다. 데이터가 섞여 있는 맹물에 양념이 들어가면 국이 더 맛있어지는 것처럼 인공지능이 어우러져 더 편리하고 유용한 서비스로 거듭나는 것입니다.

현재 지능정보기술연구원에서 개발 중인 전자상거래용 인공지능 프로그램을 잠깐 소개할까 합니다. 오늘 내가 ‘버버리 코트’를 사고 싶다면, 이제 디바이스에 말만 하면 됩니다. 인공지능은 사용자의 언어를 이해해서 묻습니다. ‘버버리’라는 특정 브랜드의 코트를 원하는 것인지, 일반적인 레인코트를 사고자 하는 것인지요. 내가 레인코트라고 답하면, 취향에 맞는 색깔, 브랜드, 가격, 디자인 등을 모두 분석해 추천하고 집에 보내주기까지 할 겁니다. 즉, 대화형 쇼핑로봇은 기존 비교검색 사이트를 능가할 것으로 기대됩니다. 데이터에 인공지능이 결합된 고깃국의 완성이죠.”

“전자정부 서비스는 이러한 인공지능의 효과가 크게 발현될 수 있는 분야입니다. 우리는 이미 축적된 데이터로 만들어진 서비스를 가지고 있기 때문입니다. 여기에 어떤 양념을 쳐서 맛을 낼 것인지는 정부의 선택이자, 이제 모두의 고민이죠.”



**인공지능은 데이터라는 국의 양념입니다.
전자정부 서비스가 인공지능으로
어떤 맛을 더할까요?**