I M F S T A F F D I S C U S S I O N N O T E

빅데이터 : 잠재적 과제 및 통계적 의미

Cornelia L. Hammer, Diane C. Kostroch, Gabriel Quirós 및 STA내부 그룹

2017. 9

통계학과

빅데이터 : 잠재적 과제 및 통계적 의미

Cornelia L. Hammer, Diane C. Kostroch, Gabriel Quirós 및 STA[1]가 준비함 외부 그룹[2]

Louis Marc Ducharme의 배포 승인

JEL Classification Numbers: E0, Y2, Z0
키워드: 빅데이터, 거시 경제 및 금융 통계, 공식 통계, 데이터 품질, 감시
저자의 이메일 주소: CHammer@imf. org, DKostroch@imf. org, GQuiros@imf. org

1 통계청 (STA) 빅데이터에 관한 내부 그룹 은 거시 경제 및 재무 통계에 대한 빅데이터의 기회와 도전을 조사하기 위해 2016 년 8 월에 설립되었습니다. 이 그룹은 STA 부국장 가브리엘 퀴 로스가 이끌고 세르칸 아르 슬라 나프, 호세 마리아 리아 카르 타스, 아미 청 카이 수 에트, 다니엘라 코 미니, 사우라 부타, 안드레아스 헤이 크, 코넬리아 L. 해머, 게리 스티븐 존스, 벤 카테 스와 루 요 실라, 다이앤 C. Kostroch , Stephanie Medina Cas, Edgardo Ruggiero 및 Patrizia Tumbarello

2 저자들은 특히 이전의 초안 Louis Marc Ducharme, El Bachir Boukherouaa 및 ITD 동료들에게 뛰어난 IT 관련 기여에 대해 귀중한 의 견을 표명했습니다 . 검토 및 편집을 위한 Claudia Dziobek 및 Mark Van Wersch; 그녀의 귀중한 의 견에 대한 Manuela Goretti. 우리는 그래픽을위한 IMF 멀티미디어 서비스 및 COM 동료 Jim Beardow, Linda Long 및 Lucy Scott Morales의 자서전 지원 에 감사드립니다 . 이 보고서는 또한 IMF 지역 부서 (AFR, APD, EUR, MCD, WHD)와 기능 부서 (COM, FAD, ICD, ITD, LEG, MCM, RES, SPR) 및 검토 부서로부터받은 검토 및 의 견으로부터 큰 혜택을 얻었습니다. James Chan, Liliya Nigmatullina 및 Zula Oimandakh (모든 STA)가 제공 한 분석 입력 및 관리 지원

목차

개요

본 스태프 토론 노트는 거시경제 및 재무 통계에 대한 빅데이터의 잠재력, 과제 및 시사점을 반영한다. 그것은 "공식"데이터와 통계의 광범위한 이해관계자를 다루고 관심 있는 사용자와 생산자를 다룬다. 모든 사회와 경제를 위한 전략적 요소인 좋은 데이터와 통계는 민간과 공공 부문에서 건전한 정책 결정을 하는데 필수적이다. 지금쯤은 국내외 기관뿐 아니라 민간기업들도 '빅데이터'는 단순한 유행어가 아니라 장기적인 비전을 요구하는 중기 개념으로 보는 시각이 많다.

빅데이터는 진화적이며 경제 및 재무 분석을 위한 혁신적이고 실시간이며 보다 세분화된 통찰력을 제공할 수 있다. 그러나 **개별 국가의 빅데이터 기회는 비대칭적일 것이며 국가의 특성과 빅데이터를 생성하는 시스템과 네트워크의 가용성에 따라 달라질 것이다**. 빅데이터는 통계 작성자와 사용자가 빅데이터를 작업 계획에 적절히 통합하기 시작할 때를 알아야 하는 공식 통계에 대한 기회, 과제 및 시사점을 제공한다.

빅데이터의 수많은 개별 애플리케이션은 이미 사용자나 데이터 및 통계 작성자에 의해 수행되고 있다. 그러나 체계적이고 체계적인 논의는 부족하다. 이 SDN은 거시경제와 금융 통계에 대한 함의를 강조하면서 그러한 제안을 하려고 한다. 빅데이터가 IMF 감시 업무를 직간접적으로 지원할 수 있는지 여부를 이해하기 위해서는 추가적인 연구와 상세한 분석이 필수적이다.

빅데이터의 의미는 무엇인가? 비록 합의 된 정의는 없지만, 용어는 종종 3Vs—대량, 속도, 다양성 등으로 특징지어진다. 시간이 지남에 따라 진실성과 변동성 등 더 많은 V가 추가되었다. 빅데이터는 특정 목적을 위해 수집된 통계자료와 달리 기업 및 행정시스템, 소셜네트워크, 사물인터넷 등에서 발견되는 부산물로 구성된다. 토론을 구조화하기 위해, 이 논문은 거시경제와 재정 통계와 관련된 빅데이터 분류를 제시한다.

빅데이터의 잠재력은 무엇인가? 빅데이터는 거시경제와 금융 통계에 이익을 줄 수 있으며, 궁극적으로 다음과 같은 세 가지 특성을 통해 정책을 수립할 수 있다.

1. 새로운 질문에 답하고 새로운 지표를 만들어냄
2. 공식 통계의 가용성에 있어 시간을 지연시킴으로써 기존 지표의 적시성 예측을 지원한다.
3. 공식 통계 제작의 혁신적 자료 출처로서

빅데이터에 어떤 문제가 발생하나? 데이터 품질 문제, 액세스에 대한 어려움, 새로운 필수 기술과 기술은 빅데이터의 주요 과제다. 그리고 빅데이터는 주로 통찰력, 상관관계, 동향 및 정서를 측정하는 반면, 국제적으로 합의 된 표준에 따라 상세한 국가별 시간 시리즈는 시간이 지남에 따라 국가의 경제성과와 정책을 측정하고 감시하는데 매우 중요하다.

통계적 의미: 나아가 국제통계협력은 빅데이터 당면과제를 극복하고 국가 및 국제 통계기관, 사용자 및 데이터 소유주 간의 지속적인 파트너십 구축의 핵심이다. 그 의미는 이해관계자들이 그들의 조직에서 필요한 기술과 기술을 구축해야 한다는 것을 포함한다. 국가 통계기관에 있어서 기회, 과제 및 잠재적 함축성이 특히 높다. **즉, 빅데이터를 새로운 데이터 소스로 통합하는 것은 전통적인 데이터 소스를 보완하거나 대체하는 것으로 방법론, 조직적 및 예산적 과제를 면제하지 않을 것이다.**

빅데이터 프로젝트의 성공은 특정 기술을 구현하는데 있는 것이 아니라 빅데이터 혁신을 앞당기고 이를 작동시키는 사람과 프로세스의 환경을 구축하는데 있다. **빅데이터를 처리하는데 필요한 다양한 기술을 감안할 때, 조직들이 데이터 및 통계 사용자와 생산자 사이의 내부 사일로를 해소할 수 있는 기회도 제공한다.**

빅데이터의 개별적 적용에서부터 그 통계의 체계적, 규칙적, 대규모 생산에 통합되는 것까지 (비용과 시간이 많이 소요되는 투자에 참여하기 전에) 조직은 개념 증명부터 시작해야 하며, 조직에서 발견이 가치 있고 실현 가능한 것으로 입증된 후에만 프로젝트를 운영해야 한다. 비틀림적 관점 통계기관은 사례별로 결정하고 기존 통계를 보완할 수 있는 가장 유망한 빅데이터 프로젝트를 선정해야 한다. 게다가, 발전을 따라잡기 위해, 기관들은 가장 시급한 연구 요구를 해결하기 위해 빅데이터 소스를 적극적으로 검색해야 한다. 또한 선정된 빅데이터 프로젝트는 가용 빅데이터 소스의 혜택을 받을 수 있는 역량 구축에 대한 구성원 자격을 지원하기 위해 역량 개발 활동에 통합될 수 있다. 앞으로, 모범 사례의 연구 및 편집 - 진실성과 변동성을 다루는 통계 기법 및 방법론에 대해, 특히 통계 커뮤니티의 최우선 과제가 되어야 한다.

빅데이터가 정적이 아니라 동적이라는 점을 감안할 때 빅데이터를 생성하는 시스템과 네트워크는 지속적으로 진화하고 있으며, 이와 함께 통계용 빅데이터의 가능성, 과제 및 한계도 함께 진화하고 있다. 따라서, **빅데이터와 공식 통계의 세계가 진화함에 따라 본 논문에서 이루어진 전반적인 평가는 재검토될 필요가 있을 것이다.**

I. 소개

1. 빅데이터는 단순한 유행어가 아니다. 빅데이터는 계속 유지될 것이다. 초기 과대광고를 회피했던 조직들은 이제 빅데이터와 미래의 조직 문화를 서로 얽을 것인지에 대한 결정을 내려야 할 필요성을 인식하게 되었다. 다른 사람들은 첫 번째 운영자들로부터 압력을 받고 있다. 조직은 모범 사례에서 배우고 의미 있는 통찰력을 가질 수 있는 잠재력이 가장 큰데이터를 조정함으로써 빅데이터를 구현하고데이터 과학의 혜택을 누릴 수 있다. 또한 가장 시급한 연구 요구에 부응하는데 도움이 될 수 있는 빅데이터 소스를 사전 예방적이고 혁신적으로 검색할 수 있다. IMF 팀들은 이미 이 관행을 성공적으로 실행하고 있다.

2. 민간 부문의 많은 기업이 혁신적인 툴을 구현하여 분석(Davenport 2006)에 경쟁하고 있다. 정교한 정보 시스템, 엄격한 분석, 뛰어난 데이터 활용은 차별화의 마지막 포인트 중 하나이다. 기업들은 축적된 데이터를 캐내어 마케팅 분석으로 전환하고 있다. 동적 콘텐츠는 시청자의 이익이나 과거의 행동에 근거하여 광고, 웹사이트 또는 이메일 본문을 조정하며, 마케팅 세분화는 과거의 구매를 바탕으로 기업의 판촉을 최적화하며, 채널간 개인화는 기업이 고객 타깃을 변경할 수 있도록 지원한다. 그들의 소셜 미디어 공급에 광고를 심음으로써. 시장(Chain Store Age 2015)은 고객의 e-쇼핑 경험을 개별화함으로써 참여율이 높아지고 소비자 충성도가 높아지고 브랜드 인지도가 향상되었다고 믿는다. 빅데이터 애플리케이션은 소매 및 전자 상거래에서 고객을 상대하는 기업에서 은행, 의 료 및 제조업에 이르기까지 많은 조직에 대한 관심을 활용했다. 동시에, 스토리지와 컴퓨팅 비용의 감소와 클라우드의 유연성, 확장성, 신뢰성 및 사용자 친화성 때문에 경쟁 환경이 근본적으로 이 공간을 민주화하게 되었다. 데이터 분석은 모든 규모의 다양한 비즈니스에서 경쟁 우위가 되고 있다.

3. 공공부문의 기관들도 그들의 권한을 보다 효과적이고 효율적으로 전달하기 위해 빅데이터와 새로운 기술을 사용하는 것에 관심이 있다. 국가 및 국제 기관에서는 생활 환경을 개선하기 위해 점점 더 비전통적인 방법이 사용되고 있다. 빅데이터는 가뭄, 기후 조건, 이주, 가격, 이전 생산 수준(Data Floq)에 대한데이터 조합을 통해 식량 부족을 예측하는데 도움이 될 수 있다. 자동차 사고와 교통 체증에 대한 정보를 담은 실시간 GPS데이터를 건너면 세계 곳곳에서 대중교통의 흐름과 자동차 속도를 높일 수 있다(Data Revolution Group 2014). 빅데이터가 용어화되기 훨씬 전에 위성 이미지는 날씨 예측과 지리적 위치 확인데이터에 사용되어 왔다. 오늘날에는 산림가, 도시계획가, 농업관리자 등이 있지만 연방 및 개발기관도 위성사진이나 "대량의 공간데이터"를 사용하여 다양한 유형의 데이터를 결합하고 해석하여 현장에서 물리적 자산의 효과를 극대화한다. 빅데이터를 이용하기 위해서는 정부와 국제기구가 관심데이터를 생성하는 기업 및 기타 기관과 제휴를 맺을 필요가 있을 것이다.

4. 데이터 혁명이 개발 과제를 해결하는데 도움을 줄 수 있는 잠재력을 널리 인정받고 있다. 휴대전화 기술, 인터넷 트래픽, 소셜 네트워킹의 광범위한 사용은 개발도상국의 개인들이 은행 서비스, 고용 정보,

의 료 서비스 및 시장에 접근할 수 있게 해준다. 동시에, 그러한 혁신의 부산물로 생성된 대량의 데이터는 인간의 행동과 사회적 패턴에 대한 더 넓고 깊은 통찰력을 얻을 수 있는 새로운 기회를 제공한다. 이러한 통찰력은 사회 발전 지표를 포함하여 "전통적으로" 수집되고 있는 지표를 보완하고 확대할 수 있다(UNGP 2012). 세계은행(World Bank 2016)은 빅데이터 분석 프로그램의 혁신을 통해 개발도상국들이 빅데이터의 기능을 운용할 수 있도록 지원하는 이니셔티브를 수립했다. 그러나 세계은행은 빅데이터가 개발도상국에서 가치를 더할 수 있는 큰 잠재력을 보고 있으며, 용량 개발과 공공 부문과 민간 부문의 협력 의 지가 핵심이 될 것임을 인정한다(ODI 2015). 프랑스 통신회사 오렌지(Orange)가 아프리카에서 개발데이터(D4D) 챌린지(Challenge 4 Development 2013) 연구원들과 자발적인데이터 공유는 민관 부문 간 필요한 협력관계를 어떻게 키울 것인가를 보여주는 좋은 사례다.

5. 중앙은행들이 빅데이터와 신기술 활용에 강한 관심을 보이고 있다(BIS 2015). 중앙은행들은 빅데이터 이용의 장점을 거시경제와 금융 안정성 분석을 지원하는 잠재적으로 효과적인 예측 도구로 보고 있다. 통화정책 분석(2016년 중앙은행)은 거시경제 변수의 더 좋고 더 시기 적절하게 예측되는 혜택을 받을 수 있다. 거시정책과 미시적 정책도 혜택을 받을 수 있다.

II. 빅데이터가 의미하는 바는 무엇인가?

8. 빅데이터라는 용어가 완전히 새로운 것은 아니다. 2006년 하버드 비즈니스 리뷰에서 톰데이븐포트는 아마존, 캐피털 원, 보스턴 레드삭스와 같은 조직들이 그들의 분야에서 지배하기 위해 사용하는 한 가지 방법 즉 경쟁적 차별화 요소로서의 분석"기업들은 데이터와 데이터 관리자들에게 넘쳐났다"고 지적했다. 2010년에 Hal Varian은 컴퓨터 매개 거래에 대해 논의 했는데, 여기에는 판매점 단말기, 현금 레지스터, 그리고 좀 더 최근의 전자 상거래와 같은 컴퓨터가 포함된다. 비록 저자들이 "빅데이터"라는 용어를 명시적으로 사용하지 않지만, 그들이 언급하는 현상과 정보는 이후에 빅데이터 토론에 포함될 것이다.

9. 현재 빅데이터에 대한 합의 된 정의 가 존재하지 않지만, 용어는 종종 3Vs—대량, 속도, 다양성으로 특징지어진다. [3] 대용량(High volume)은 기계, 네트워크 및 인간 상호작용에 의해 생성되는 엑사바이트(Exabyte)데이터의 증가, 고속(High Speed)은 데이터가 생성, 처리 및 저장되는 속도를 의미하며, 다양성(High Various)는 데이터 유형과 소스의 범위와 복잡성과 관련이 있다. 데이터 세트는 매우 크고 복잡하여 기존의 데이터 처리 애플리케이션은 데이터를 캡처, 저장 및 분석하기에 불충분해진다. 대신 빅데이터를 처리하려면 인적 기술, 첨단 기술, 데이터 액세스 인프라 등의 네트워크가 필수적이다. 이는 빅데이터를 툴킷에 통합하고자 하는 통계학자 및 정책 입안 기관의 핵심

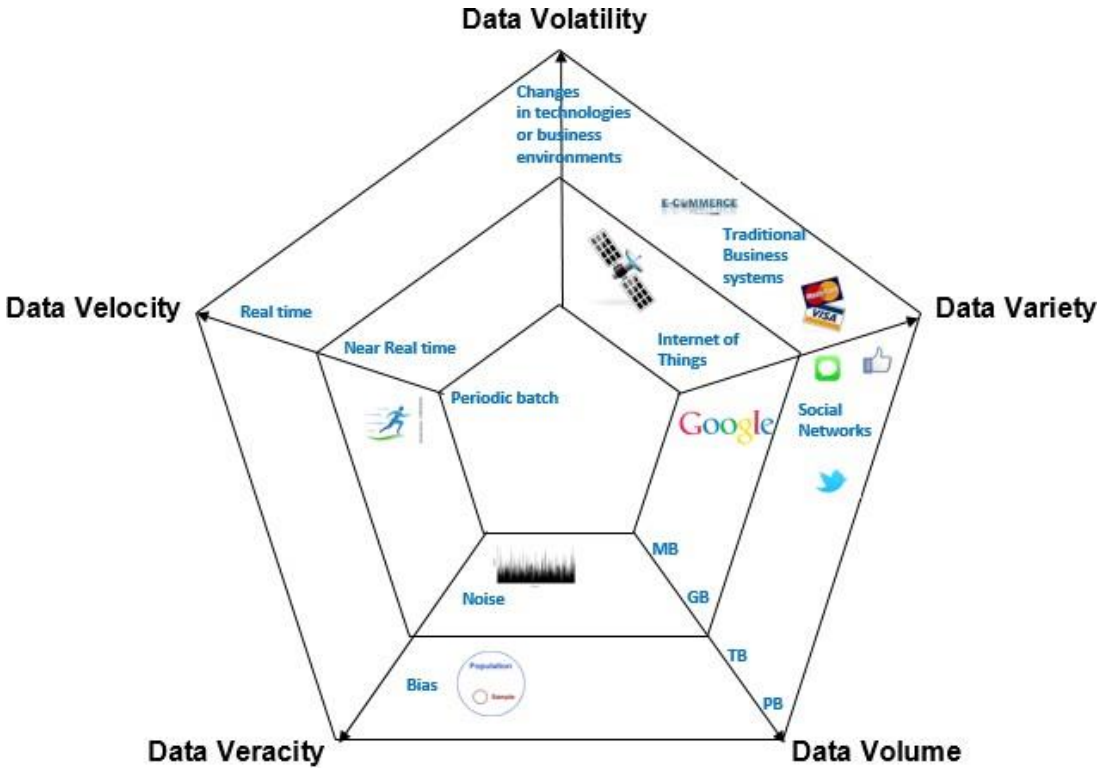
3 Gartner analyst Doug Laney came up with the famous three Vs back in 2001

과제다.

10. 특정 목적을 위해 수집된 통계("만들기") 자료와 달리, 빅데이터는 비즈니스 및 행정 시스템, 소셜 네트워크, 사물 인터넷 등에서 "발견되는" 부산물이다. 소셜 네트워크는 사람들이 비슷한 관심사를 가진 다른 사람들과의 사회적 관계를 형성하도록 돕는 온라인 플랫폼이다(Facebook, Twitter, LinkedIn. 사용자들은 블로그와 프로필을 만들고, 사진을 공유하고, 메시지를 교환함으로써 디지털화되고 저장된 인간 소스의 정보를 제공한다. 소셜 네트워크의 데이터는 종종 통제되지 않고 구조화되지 않는다. 빅데이터 분류에 있어서, 유엔 유럽 경제 위원회(UNECE)는 또한 소셜 네트워크 인터넷 검색과 인간 정보로 더 널리 이해될 수 있는 모바일 데이터를 포함한다. 전통적인 비즈니스 시스템은 기업이 고객에게 가치를 제공하고 관리 기록을 포함한 프로세스 매개데이터를 생성하기 위해 정의 한 프로세스와 절차다. 비즈니스 시스템은 관계형 데이터베이스 시스템에 저장된 비즈니스 이벤트(고객 등록이나 주문 수신과 같은 상업적 거래)와 관련된 거래, 위치 및 메타데이터에 대한 통제가 잘되고 구조화된 정보를 기록한다. 사물의 인터넷은 물리적인 세계의 사건 및 상황을 측정하고 기록하는 인터넷 연결과 내장된 센서와 내장된 상호관련 컴퓨팅 장치를 데이터로 생산하는 시스템이다. 이들의 출력은 기계로 생성된 데이터(센서 기록, 컴퓨터 로그, 웹캠, 휴대전화 위치/GPS)로 구성된다.

11. Vs 목록은 시간이 지남에 따라 커져서 빅데이터를 기존 비즈니스 운영에 통합 할 때 회사와 조직이 직면한 기회와 과제를 모두 강조합니다 (그림 1). Veracity는 빅데이터의 가치와 타당성을 가져오는데 이었어 가장 큰 과제 중 하나인 데이터의 오류(noise)와 편향(bias)를 말합니다. 변동성은 빅데이터가 생성되는 기술 또는 비즈니스 환경의 변화를 의미하며, 이는 잘못된 분석 및 결과로 이어질 수 있으며 빅데이터의 데이터 소스로서의 취약성으로 이어질 수 있습니다.

그림 1. 빅데이터의 5V - 다양성, 다양성, 속도, 신뢰성 및 볼륨



Based on Doug Laney, 2001

12. 거시경제 및 재무 통계와 점차 관련이 있는 빅데이터 분류는 박스 1에 제시되어있다. 대상별로 빅데이터를 생성하는 네트워크, 시스템, 기계의 범주는 UNECE 빅데이터 분류(부록 I 참조)에 기초하고 있지만, 본 논문은 행정데이터, 비즈니스 웹 사이트, 온라인 뉴스 등 (이탈릭체) 추가적인 하위 범주를 포함한다. 사용 사례의 결과에 기초하여, 이러한 추가 하위 범주는 거시경제 및 재무 통계와 마지막으로 감시를 제공할 잠재성도 있는 것으로 보인다. 그러나 비디오, 의 료 기록, 사진 등과 같은 범주는 비록 나중에 그렇게 될 수 있지만 현재 거시경제와 재정 통계에는 적용되지 않기 때문에 제외된다.

Box 1. Adapted UNECE Big Data Classification

1. Social Networks (human-sourced information)

1100. Social Networks: Facebook, Twitter, *LinkedIn*

1200. Blogs and comments

1600. Internet searches on search engines (*Google*)

1700. Mobile data content: text messages, *Call Detail Record, Data Detail Record, Location update, Radio coverage updates*

Online news

2. Traditional Business systems (process-mediated data)

21. Data produced by public agencies

Administrative data

22. Data produced by businesses

2210. Commercial transactions

2220. Banking/stock records

2230. E-commerce

2240. Credit cards, *Business websites Scanner data*

3. Internet of Things (machine-generated data)

31. Data from sensors

311. Fixed sensors

3111. Home automation

3112. Weather/pollution sensors

3113. Traffic sensors/webcam

3114. Scientific sensors

312. Mobile sensors (tracking)

3121. Mobile phone location *(GPS)*

3122. Cars

3123. Satellite images

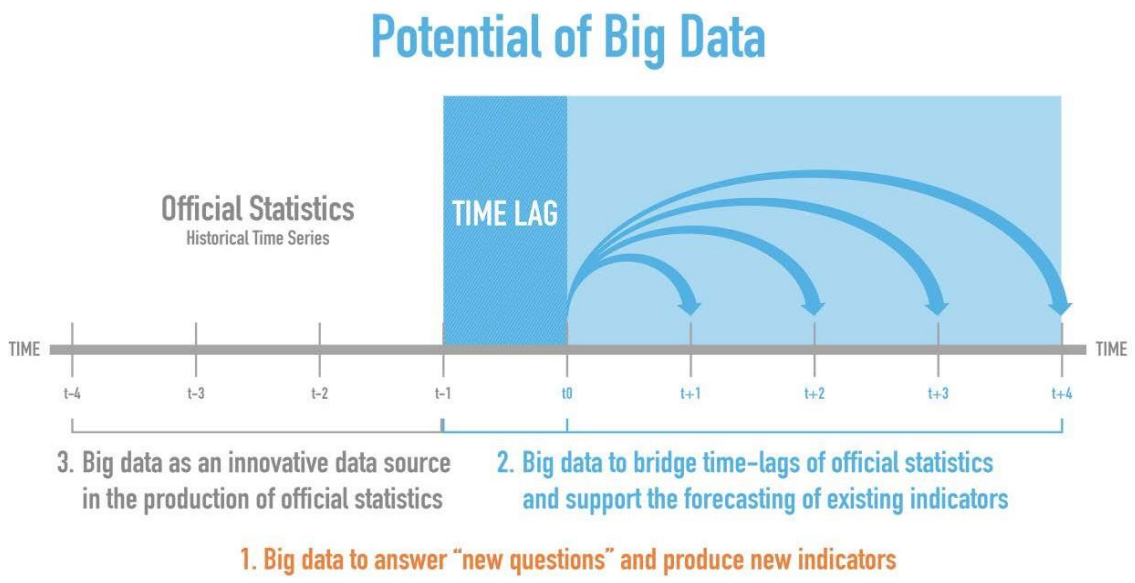
13. 빅데이터의 잠재력은 크지만 각 국가의 빅데이터 기회는 국가의 특성에 달려 있습니다. 소셜 네트워크, 기존 비즈니스, 관리 시스템 및 빅데이터를 생성하는 사물 인터넷의 가용성은 다양합니다. 결과적으로 빅데이터 기회와 정책 결정을 위한 잠재적 인 응용 프로그램을 평가하려면 데이터 및 관련 도구의 가용성, 사용자의 기능, 개인 정보 및 보안 문제, 법률 및 기술 시스템을 고려해야 합니다.

III. 빅데이터의미래

14. 섹션 III의 논의 는 빅데이터가 거시경제 및 금융 통계에 직간접적으로 이익을 줄 수 있는 세 가지 특징을 중심으로 구성되며, 마지막으로 정책 수립(그림 2):

1. 새로운 질문에 답하고 새로운 지표를 만들어냄
2. 공식 통계의 가용성에 있어 시간을 지연시킴으로써 기존 지표의 적시성 예측을 지원한다.
3. 공식통계 생산에 혁신적인 자료 출처를 제공함.

Figure 2. The Potential of Big Data



이 접근법은 거시경제와 재무 분석을 위해 가장 중요한 요소를 증식할 뿐만 아니라 빅데이터에 대한 광범위한 논의를 구조화할 수 있다.

15. 세 가지 특성은 직접(1과 2) 또는 간접(3)으로 정책 분석을 위한 귀중한 데이터를 제공할 수 있지만, 서로 연결되어 있으므로 완전히 분리할 수 없다. 본 문서는 예제를 제공하지만 포괄적 인 빅데이터 프로젝트 인벤토리를 제공하지는 않습니다.[4] 오히려 빅데이터 프로젝트가 어떻게 정책 분석을 향상시킬 수 있는지 설명하려고 합니다. 그러나 그 가능성을 보장하기 위해서는 앞으로 더 많은 탐구와 개념 증명이 필요합니다.

4 A big data inventory can be found here: https://unstats. un. org/bigdata/inventory/.

A. 새로운 질문에 답변하고 새로운 지표를 작성하는 빅데이터

16. 빅데이터는 인과관계를 검색하는 전통적인 방법에서 깨진다. 빅데이터를 다루는 것은 우리에게 왜 어떤 일이 일어나고 있는지 말해주는 것이 아니라 그것이 일어나고 있다는 것을 우리에게 경고하는 패턴과 상관관계를 찾는 것을 의미한다(Mayer-Schönberger and Cukier 2014). 이런 맥락에서 실시간 상관관계를 확보하고, 실질, 외부, 재정, 금융 분야의 시스템 리스크뿐만 아니라 국가별 구축도 감시할 수 있는 보다 포괄적인 조기경보시스템(Kitchin 2015)을 구축하기 위한 새로운 지표를 개발할 수 있다. 구글 웹 검색과 페이스북 게시물은 이미 주식시장 유동성(아리우리 등 2014년)을 예측하고 주식시장 활동을 예측하는 심리지수를 구축하는데 사용되고 있다(카라불럿 2013).

17. 빅데이터의 효과적이고 스마트한 활용이 진화하고 있지만, 우리가 정보를 바라보는 방식을 변화시킬 수 있는 가능성은 논란의 여지가 없다. 빅데이터 탐사는 새롭고 더 나은 질문을 하는 연습일 뿐만 아니라 통계의 수집과 생산에 대한 우리의 전통적인 생각에 도전하는 기회다. 이러한 견해는 국제통화기금(IMF)의 빅데이터 및 분석 심포지엄이나 유럽 빅데이터 해커톤과 같은 국제적인 이니셔티브에 반영되어 참가자들(현장의 전문가, 학계, 민간 부문)이 정책 및 의사결정에 도움이 되는 빅데이터의 혁신적인 활용 방안을 모색하도록 도전한다. 의 심할 여지 없이 더 많은 아이디어가 생겨날 것이며, 그 중 일부는 개념증명을 통과한 후 전통적인 통계의 범위를 넓히고 연구 요구에 대응할 수 있을 것이다.

18. 어떤 경우에는 빅데이터를 통해 정책 분석이 집계를 넘어 정책 대응을 보다 효과적으로 알릴 수 있다. 휴대전화, 통화내역기록, 위성사진,[5] 및 특별히 개발된 앱은 사회경제적 패턴, 인구이동, 신용위험 프로파일, 기후스마트 농업 등을 예측하고 위기와 관련된 스트레스의 증가를 감지할 수 있는 잠재력을 가지고 있다. 빅데이터의 보다 세분화된 정보는 정책 상호연계를 명확히 하고, 따라서 정책 권고사항이 기업과 가정에 미치는 영향을 보다 신속하게 식별하여 정치적 긴장 및/또는 불평등 역학을 강조할 수 있다. 빅데이터는 개발 과제를 해결하고 성평등과 같은 SDG(Sustainable Development Goals) 지표의 컴파일 요구를 충족할 수 있는 잠재력을 가지고 있다. [6] 민간기업 LinkedIn은 이미 세분화된데이터를 사용하여 성별 다양성 통계를 발표하고 성 통계 관련 교육을 실시하고 있다(Karani 2017).

19. IMF 빅데이터 및 분석 심포지엄에 이어, IMF 사내 빅데이터 혁신 챌린지는 IMF의 향후 업무에서 빅데이터를 활용할 수 있는 혁신적인 방법의 길을 닦았다. 상위 6개 아이디어는 개념 개발 증명용으로 승인되었다. (1) SWIFT데이터를 사용하여 글로벌 재무 흐름을 모니터링하고, (2) 감성 기반 조기 경보 시스템, (3) 현재 Google 추세데이터를 사용하여 GDP 예측, (4) 해변 지수 @ 주간 자동화 및 확장, (5)

5 Australia, China, Colombia, and Mexico have discovered the potential of satellite data for agricultural statistics.

6 The IMF Statistics Department is involved in the following SDGs: (1) indicator 8. 10. 1. a, Number of commercial bank branches per 100,000 adults, and indicator 8. 10. 1. b, Number of ATMs per 100,000 adults (Tier I);

감시 및 정책 분석을 강화하기 위한 정부 현금 흐름데이터 풀링, (6) 세금 및 관세 관리를 개선하기 위한 분석 적용이온. 다른 예로는 국제통화기금(IMF) 재정부 세입관리국 격차 분석 프로그램에 대한 빅데이터를 사용하여 부가가치세 준수 격차를 결정하고, 국제통화기금(IMF) 연구부가 저소득 국가의 경제활동에 미치는 영향을 측정하는 것이다. 그러한 생각은 시간이 지남에 따라 기존의 관행을 증가시키거나 보완하기에 충분한 견고성과 정확성을 보여줄 수 있다.

20. 빅데이터는 금융 포함, 금융 서비스 액세스 및 경제성장의 영향을 더 잘 측정할 수 있다. M-Pesa (박스 2)와 같은 전자화폐 시스템은 개발도상국에서 급성장하고 있다(Donovan 2012).[7] 이들과 함께 재정포용이 빈곤감소, 성별 불평등, 경제성장에 미치는 영향을 측정할 수 있는 기회를 제공한다. 이동형 전송 시스템의 데이터를 통해 피어투피어 트랜잭션을 더 자세히 추적할 수 있으며, 이는 다시 송금, 지역 가처분 소득, 소비 패턴 및 재무 포함에 대한 더 정확한 추정치를 산출하는데 도움이 될 수 있다.

Box 2. 경제 정책 수립을 위해 모바일 전송 시스템에 저장된데이터를 사용하는 M-Pesa

M-Pesa는 케냐에서 2007년에 설립된 소규모 휴대전화 기반 송금 시스템이다. 이후 아프리카, 동유럽, 중동, 남아시아 등 개발도상국으로 확대됐다. 이용자들은 휴대전화 앱에 저장된 계좌에 있는 대리점을 통해 돈을 입금한다. 그들은 개인 식별이 보장된 문자메시지를 이용하여 어떤 휴대폰 사용자에게도 돈을 송금할 수 있다. 송금된 돈은 수혜자가 어떤 M-Pesa대리점에서 현금으로 수령할 수 있다. 이용자들은 송금과 인출 수수료를 받는다. 원래 돈 이전을 용이하게 하기 위해 고안된, 봉급, 소액 금융, 상품과 서비스의 구매를 다루기 위해 서비스가 확대되었다. 케냐에서는 85% 이상의 가정이 M-Pesa를 사용한다. 전국 각지에 7만8000여명의 요원이 분산돼 거의 모든 마을에 이른다.

빅데이터의 잠재적 사용

M-Pesa 사용자 간의 현재 전송 측정: 가계는 청구서와 월 보험료 등을 내거나 연금이나 사회복지대금을 받는다. 많은 시골 가정들은 도시 중심지로부터의 송금에 의 존하여 생존한다. M-Pesa는 송금 지불에 사용되었던 많은 비효율적이거나 제한된 옵션들(돈 송금 사업자, 외환국, 버스 회사, 친구 및 가족)을 대체하고 시골 지역의 가구들에게 금융 서비스에 접근할 수 있게 해준다. M-Pesa데이터는 송금, 가처분소득, 금융포함 등과 같은 지표의 더 정확한 추정치를 산출할 수 있다. 즉, 실제 금융 외부 부문 지표에 필요한 모든 관련 정보를 제공한다.

소비 패턴 추정: 소비자들은 소매점에서 현금 유입/현금 인출에 M-Pesa를 사용하는데, 그 중 다수는 도심 외곽에 위치해 있다. M-Pesa데이터는 산업별 지역적 수요를 보다 명확하게 파악할 수 있으며 수요 패턴의 변화의 초기 신호로 작용할 수 있다.

7 More than 110 money mobile systems servicing more than 40 million customers.

21. 빅데이터는데이터 문제가 더 심각한 저소득 국가(LIC)의 감시를 지원할 수 있다. M-Pesa의 사례에서 알 수 있듯이 빅데이터는 데이터가 부족하고 시대에 뒤떨어진 LIC의 거시경제 및 재무 동향을 모니터링하는데 특히 도움이 될 수 있다. 좋은 예로는 대개 수집하기 힘든 가계데이터가 있다. 기업이 생산하는 소셜 네트워크, 모바일 데이터 컨텐츠, 거래상대방데이터에서 인간이 제공하는 정보는 가계의 데이터 가용성 문제를 극복하는 좋은 출발점이 될 수 있다. 빅데이터는 또한 데이터 수집 관행을 개선하고, 새로운 통계를 편집하며, 중복데이터 수집(ADB 2013)과 같은 가능한 비효율성을 줄일 수 있는 기회를 제공할 수 있다. 국제통화기금(IMF) 주간 @해수욕지수(박스 3)는 관광이 GDP의 상당 부분을 차지하고 정기적인 비용 및 물가지수가 잘 개발되지 않는 카리브해 지역의 귀중한 도구다.

Box 3. 일주일 @ 해변 지수

The Economist's Big Mac Index에서 영감을 받아, The Week @ the Beach Index는 전 세계 해변 휴가의 비용을 비교한 가격 지수를 구성한다. 이 지수는 해변 휴가에 일반적으로 소비되는 상품 바구니, 즉 호텔 요금, 택시 요금, 식사 및 음료 가격(물, 커피, 맥주)에 대한 국내 비용을 포함한다. 지수의 데이터는 Expedia 및 TripAdvisor(호텔 요금), Worldcabfare(택시 요금), Numbeo(식료 및 음료 가격)에서 도출한다).

응용 프로그램 사례

이 지수는 분기별로 기록되고 있으며 (1) 경쟁력 지표로서 외부 부문 평가를 보완하는 지표로서, 특히 관광이 경제의 큰 몫을 차지하는 일부 국가에서는 유용하고, (2) 평형 실질 환율 측정을 위한 대체 수단으로서, (3) 설명으로서 활용되고 있다. 특히 긴 시계열이 컴파일됨에 따라 경험적 작업에서의 아토리 변수. 예를 들어 관광흐름의 결정요인을 추정할 때 가격탄력성 및 공급요인으로 포함되는 호텔의 수 등 개별변수를 추정하는 변수로 지수를 활용했다. 또 공식 통계에서 포착되지 않는 공유경제(예: Airbnb)를 편입시켜 시장과 가격데이터를 개선할 수 있었다

Source: Laframboise and others 2014.

22. 국제통화기금(IMF)과 다른 기관들은 빅데이터를 사용하여 현재 제3자가 생산하고 있는 감시에 사용되는 지표를 상호 확인하고 방법, 자료 출처, 지표의 품질 및 특성을 평가하는데 도움을 줄 수 있다. 빅데이터는 민간기업과 기관이 IMF 공식 직원 보고서에 자주 사용되는 제3자 지표(TPI)를 생산하기 위해 사용한다. 대체 지표는 빅데이터 이니셔티브를 통해 확보하여 생산할 수 있으며, TPI의 품질, 특성, 방법론적 건전성을 평가하는 벤치마크 역할을 할 수 있다.

23. 단순한 모델과 더 많은데이터가 더 적은데이터를 기반으로 더 정교한 모델을 능가할 경우 빅데이터 분석이 논의 의 빛을 발할 수 있다(Mayer-Schönberger 및 Cukier 2014). 빅데이터에서 얻을 수 있는 몇 가지 잠재적 이득은 더 빠른 처리나 더 나은 알고리즘에서 오는 것이 아니라 단순히 더 많은 다양한 데이터가 있기 때문이다. "모든 것의 데이터화"는 더 넓은 범위의 정보의 잠재 가치를 밝혀낸다. 따라서 빅데이터는 경제적 모델링이 어떻게 수행될 수 있는지에 대해 재고하고 되돌아볼 수 있는 기회가 될 수 있다. 경우에 따라 가용한 (새로운)데이터에 대한 분석의 적응을 고려할 수 있는가, 아니면 경제 분석에 대한데이터 적응이라는 기존의 접근방식을 고수하는 것을 선호하는가?[8]

8 빅데이터는 IMF에서 기존의 분석기법을 폭넓게 채택하도록 유도한다. 수십 년 동안 존재해 온 알고리즘을 이용한 기계학습 방법은 더 많은데이터를 이용할 수 있다는 점에서 엄청난 혜택을 받았다. 국제통화기금(IMF) 실무보고서인 '어둠 속의 목격: 레바논의 현재 방송의 기계학습 접근법'은 이 방법들이 펀드에서 어떻게 사용되고 있는지를 보여주는 한 예다. 게다가, "자연어 처리"는 많은 관심을 불러일으켰다. IMF는 최근 세미나를 개최했다. "텍스트 마이닝, 정보 및 연결"에서, 외부 및 내부 연사들이 자신의 아이디어를 제시하고 비정형 콘텐츠로부터 의미, 컨텍스트 및 정서를 추론하기 위해 진행 중인 작업

B. Big Data to Bridge Time Lags of Official Statistics 및 기존 지표의 지원 예측

24. 예를 들어 금융 및 가격데이터 같은 주요 변수들이 거의 즉각적으로 관찰될 수 있기 때문에 빠른 통찰력은 빅데이터의 가장 큰 약속 중 하나이다. 경제 및 재무 개발을 모니터링하고 안정성 위험의 조기 경고 신호를 제공하기 위해 정책 사용에 적합한 통계에 적시에 필요한 데이터가 필요하다. 승리한 IMF 빅데이터 혁신 프로젝트는 SWIFT데이터를 거래량 및 금융 시장 가격에 사용하여 전 세계 금융 흐름을 감시할 것을 주창했다(박스 4). [9] SWIFT데이터는 가능한 전염과 유출 효과에 대한 적절한 통찰력을 얻을 수 있는 유망한 사례다. 다른 연구에서는 SWIFT데이터를 사용하여 네트워크 집중도와 국가간 트랜잭션을 평가하고 있다. [10]

상자 4. SWIFT를 사용하여 글로벌 재무 흐름 모니터링

SWIFT(Society for Worldwide Bank Financial Telecommunications)는 200여 개국의 약 1만1000개 금융기관에서 사용하는 표준화된 보안 메시징 시스템이다. 매일 2500만개 이상의 SWIFT 메시지가 전송된다. SWIFT는 기본 거래의 성격에 대한 정보를 제공하는 100개 이상의 서로 다른 메시지 유형을 가지고 있다. 각 메시지 유형에 대해, 전송된 총 메시지 수와 이러한 지불의 총 값을 사용할 수 있다.

응용 프로그램 예제

글로벌 금융 흐름 모니터링: SWIFT데이터는 지역과 통화별로 분류된 전 세계 금융 흐름을 포착하는 지표를 작성하는데 사용된다. SWIFT지수는 SWIFT데이터를 사용하여 경제 활동의 거울로 SWIFT 트래픽(볼륨)을 사용하여 초기 GDP 성장 추정치를 제공하는 한 예다. 2015년 SDR을 검토할 때 IMF는 SWIFT의 데이터를 국제거래에서 중국 위안화의 역할이 커지는 지표로 활용했다. [11] 2015년 IMF 혁신도전의 맥락에서 SWIFT데이터를 사용하여 글로벌 금융 흐름을 감시하자는 제안이 나왔다. IMF 직원 팀은 2016년 동안 SWIFT데이터 사용에 대한 개념 증명 훈련을 실시했다. 이러한 맥락에서 SWIFT데이터는 시기적절성을 높이기 위해 국제 무역 통계를 현재 방송하는데 사용되어 왔다.

혜택들

통화구성: SWIFT데이터는 글로벌 및 지역적으로 금융 거래에서 서로 다른 통화의 사용 추세를 감시하는데 사용될 수 있다. 수출/수입 지표: 거래와 금융 거래를 구분하기 위해 다른 SWIFT 메시지 유형을 사용할 수 있으며, 공식데이터가 이용되기 전에 현재 거래 흐름에 정보를 제공할 수 있다. 인출 특파원 은행 관계: SWIFT데이터는 통신원 은행 관계에 대한 위험 평가 및 모니터링을 지원할 수 있다. 또 민간 금융기관의 SWIFT데이터에 대한 접근은 자금세탁 방지 원칙 등 은행의 실사 대책을 뒷받침할 수 있다.

9 특히 SWIFT데이터 사용과 관련된 제한사항은 접근 제한사항이다. 기관은 SWIFT데이터에 대한 구독을 구입해야 한다. 인터페이스는 사용자에게 친숙하지만, 단일 다운로드의 크기에 제한이 있다. 게시 제어(데이터 기밀): SWIFT는 기밀 유지를 위해데이터를 사용하는 외부 게시를 승인해야 한다. 데이터 처리: 대용량데이터 볼륨을 처리하려면 처리 용량이 축적되어야 한다. 직원 비용: 어떤 훈련은 인터페이스와 메시지 유형에 익숙해지는데 유용하다. 엑셀과 같은 익숙한 도구는데이터 작업을 하기에 충분하지 않을 것이다.

10 Cook and Soraméki 2014와 Sy와 Wang 2016은 SWIFT 메시지 유형 MT103을 사용하여 전 세계의 지역 네트워크 집중도를 보여주며, 미국은 네트워크 코어가 된다. 2009년 국제결제은행(BIS)은 SWIFT데이터를 사용하여 국경간 거래를 보다 투명하게 할 수 있는 기회를 강조하였는데, 특히 발신자 이외의 경제에 위치한 중개은행이 관여하는 경우 더욱 그러하다. 2015년에 BIS는 SWIFT데이터의 유용성을 반복하여 통신사 은행 관계를 더 잘 이해하고, 통신사 은행 관계에 대한 인출 및 압력의 관련 위험을 모니터링할 수 있도록 하였다.

11 IMF 2015년 11월 정책서 "SDR 평가방법 검토."

25. 빅데이터를 사용하면 거의 실시간으로 경제 신호를 추출하고 공식 수치가 발표되기 전에 경제 흐름을 전망할 수 있는 기회를 창출한다. 빅데이터에서 도출된 시기 적절한 정보는 개별 경제와 세계 경제의 재무 및 경제 상태를 평가하는데 도움이 될 것이다. 그것은 빠르게 움직이는 금융, 상품 및 기타 시장에 대한 거의 실시간 정보를 제공함으로써 안정성을 유지함에 있어 감시를 지원할 수 있으며, 따라서 국제 통화 시스템의 위기를 예방하거나 최소한 완화시키는데 도움이 될 수 있다. 현재 방송 연습을 통해 얻은 정보는 금융 시장, 공공 금융 개발 및 지역 경제 전망과 같은 분야에서 세계적인 전망을 도출함으로써 정책 입안을 더욱 강화할 것이다.

26. Nowcasting(또는 실시간 예측)은 이미 많은 지표에 대해 민간 및 공공 부문에서 널리 사용되고 있다(FRBNY 2017; Banbura 등 2013). [12] MIT의 10억 달러 가격 프로젝트는 웹 스크래핑을 사용하여 온라인 소매업체로부터 가격변동과 인플레이션을 일일 빈도로 모니터링하고 인플레이션을 전환한다. 이온 경향 트위터는 가격 정보가 담긴 트윗을 필터링하고 모델링함으로써 인도네시아의 식품 가격을 현재 예측하는데 사용되었다(UNGP 2014). 민간기업 프레다타(2016년)는 오픈소스 소셜 미디어와 협업 매체에 걸친 디지털 대화를 감시함으로써 웹 주변의 데이터 소스를 지정학적 위험 신호로 압축한다. 그 후, 이러한 신호에 기계학습 알고리즘을 적용하여 자산 가격 변화에서 시민 시위, 노동 파업, 선거 결과 및 국가 안보 결과에 이르기까지 다양한 발전을 예측한다.

27. 현재 전망의 다른 두드러진 예로는 관광, 실업, 소매업, 무역 흐름에 관한 경제 흐름이 있다. [13] 통계 작성자들이 소셜 네트워크 세계와 위성 이미지의 새로운 응용 분야에 과감히 진출하고 있다. [14] 네덜란드는 소비심리를 추정하기 위해 페이스북과 트위터데이터를 사용하고 있고 중국과 이탈리아는 대략 조에 가깝다. 웹 스크랩을 통한 공실률 다른 잘 알려진 예로는 자동차와 부동산 판매뿐 아니라 실업률과 소비자 심리를 파악하기 위해 공개적으로 이용 가능한 구글 트렌드 데이터를 사용하는 것이다(Pavlicek and Kristoufem 2015).

28. 선도 지표(OECD 1987)[15]는 전통적으로 재무 및 경제 발전을 예측하는데 사용되었지만, 빅데이터는 가용데이터의 양 덕분에 더 나은 예측 변수를 제공하고 예측 정확도를 향상시킬 수 있다. 첫 번째 단계로서, 예측 연습은 최근의 과거가 어떻게 예상된 것과 비교하여 발전했는지를 평가하기 위해 이용할 수 있는 정보의 범위를 살펴본다. 여러 연구에서는 빅데이터가 제공하는 보다 상세하고 세분화된 정보를 통해 추정 경제 시리즈의 품질을 개선할 수 있다는 사실을 밝혀냈다(Galbraith and Tkacz 2013). 예측 연습은 기존의 통계 시리즈를 보다 세분화, 고주파수 및 사회경제데이터로 보완함으로써 빅데이터의 이점을 얻을 수 있다. 빅데이터 소스를 활용하고 관찰 횟수를 늘림으로써

12 기업은 공급망 관리를 효율화하기 위해 예측 기법을 사용한다.

13 콜롬비아 재무부는 구글 트렌드에서 파생된 경제활동을 감시하기 위해 단기 동향을 이용한다.

14 국제 연합 통계 부서의 빅데이터 이니셔티브

15 선행지표는 경제활동의 전환점 조기경보(예: 경제협력개발기구 종합선도지표)를 제공하는데 사용되었다.

빅데이터는 예측 정확도를 향상시킬 수 있다.

29. 새로운데이터 소스가 예측에 대한 입력의 역할을 하더라도, 일관되고 조화를 이룬 과거 시계열이 여전히 필요하다. 공식 통계와 관련된 시차 및 빈도 문제는 빅데이터에서 지표의 개발에 박차를 가하는 반면, 이 둘은 상호 보완적인 것으로 보아야 한다. 두 출력 모두 벤치마크 역할을 할 수 있는 기존 시계열에서 새로운 지표의 견고성을 보장하기 위해 비교해야 한다. 빅데이터는 주로 "실제 정보"(예: 이동, 동향, 정서)와 대조적으로 "실제 정보"(예: 연말 미불 부채의 위치)와 상관관계를 측정한다.

C. 데이터 출처로서의 빅데이터 및 공식 통계 작성의 혁신

30. 빅데이터 시대에 공식 통계가 어떻게 진화해야 하는지에 대한 전 세계의 논의는 통계기관이 상당한 변화를 겪기 시작하고 있음을 보여준다. 유엔 통계위원회는 2014년 45차 회의 에서 "빅데이터가 무시할 수 없는 정보의 원천을 구성한다"(UNGWG 2017)고 공식 인정했다. 유럽에서는 2013년에 모든 유럽 국가 통계청장이 "빅데이터 및 공식 통계에 관한 회의 록"(EC 2014)에 서명하고, 공식 통계 빅데이터 전략 실행 계획 및 로드맵을 보다 광범위한 지배 전략과 통합하는데 전념하였다. Eurostat는 EU 회원국들의 전형적인 빅데이터 사용은 고립된 것이 아니라 이미 존재하는 데이터 출처(EC 2017)와 결합된 것이라고 강조했다. 실제로, 세계 통계는 빅데이터를 입증하고 이 새로운 데이터 소스를 앞으로의 도전과 기회로 받아들이기 위해 노력하고 있다.

31. 빅데이터를 공식 통계로 통합하는 전략은 부분적으로 기존 통계 출처를 대체하는 것에서부터 보완적이거나 완전히 새로운 통계 산출물을 제공하는 것까지 다양할 수 있다. (Flowscu 및 기타 2014) 예산이 한정되어 있고 설문 조사에 대한 응답이 감소하는 시기 (예: Meyer, Mok 및 Sullivan 2015)에서 공식 통계는 빅데이터와 같이 관련성이 높지만 시기 적절하고 더 많은 가능성이 있는 새로운 데이터 소스를 탐색해야 합니다. 기존데이터 수집 방법보다 비용 효율적입니다. [16] 국가 당국은 응답 부담을 줄이고 통계적 생산 프로세스를 현대화하는 동시에 더 빠르고 빈번한 통계를 생성하는데 유망한 장점이 있습니다.

32. 작성자들은 주로 전통적인데이터 소스를 다듬고 보완하기 위해 빅데이터 프로젝트를 시범적으로 진행하고 있다. 공식 작성자에 의 한 빅데이터 이용의 일부 잘 문서화된 예는[17] (1) 관광, 교통 및 도시 통계(예: Eurostat, 벨기에, 브라질, 인도네시아, 이스라엘, 이탈리아, 나이지리아의 세계은행, 폴란드), (2) 가격지표, 노동시장 지표, 기업 프로파일링(Eurostat, 중국, Eurostat, 에콰도르, 핀란드, 독일, 헝가리, 일본), (3) 에너지 및 환경 통계용 스마트 미터(Eurostat, 벨기에, 캐나다), (4) 가격 및 기타 경제 통계용 신용카드, 현금 레지스터 및 스캐너데이터. 섹션 V의 표 1은 공식 통계 영역에 공급될 가능성이 있는 빅데이터 소스를 제시한다. 그것은 유엔 분류에 근거하고 있으며(박스 1 참조), 감시를 위한 IMF 통계

16 UNSD GWG Survey and Project Inventory, https://unstats. un. org/bigdata/inventory/.

17 UNSD Big Data Project Inventory https://unstats. un. org/bigdata/inventory/.

요구에 맞춘 것이다.

33. 많은 국가에서 빅데이터는 공식 통계를 작성하는 저비용 고품질데이터 소스 대안이 될 수 있다. 유망한 예는 에스토니아의 모바일 포지셔닝을 국경 조사의 대안으로 지급 잔액 통계를 위한 여행 서비스데이터를 수집하는 것이다(박스 5). 이 경우 빅데이터를 사용하면 비용을 크게 절감할 뿐만 아니라 기존의 조사 기반 추정치에 비해 추정치의 품질, 정확도, 데이터 가용성 및 포괄성도 향상되었다. 관광에 크게 의존하는 나라들에게 이것은 고려될 수 있는 기회다.

Box 5. 국제여행서비스 통계자료 출처로서의 데이터 이동 배치

배경: 2010년 통계 에스토니아는 국경 조사를 중단했고, 이로 인해 에스티 판크(에스토니아 중앙은행)는 여행 서비스 추정치를 수집하는 다른 방법을 모색해야 했다. 조사된 다양한 대안(도로 센서, 신용카드 정보, 숙박 기반데이터) 중에서 모바일 위치 확인데이터는 가장 간단하고 비용이 적게 들며 가장 시의 적절하다는 것이 입증되었다. 더욱이 에스토니아에서는 이러한 데이터에 접근하고 이용하는데 있어서 법적인 장애물이 없다.

여행 서비스 수출/이미지 견적 이동 위치데이터는 여행 기간, 인바운드 및 아웃바운드 해외 여행자, 월별 여행 서비스 수출/이메이트를 추정하는데 사용할 수 있다.

휴대폰 사용 및 위치데이터는 SIM 카드 ID, 날짜 및 시간, 위치, 국가 코드 등의 정보를 제공한다. 고객의 거주지는 SIM 카드 거주지로 추정된다. 방법론은 익명데이터의 SIM 카드의 로밍 패턴 분석에 기초한다. (1) 거주자 이동 통신사의 네트워크에 있는 비거주자는 인바운드 여행자 및 국가의 추정치를 제공한다. (2) 거주자 이동 통신사가 비거주 모바일 파트너 회사로부터 수신한 로밍 활동 보고서를 사용하여 아웃바운드 여행을 추정한다. rs와 시골 알고리즘은 짧고 긴 방문을 고려하고 운송 여행(항로, 공항, 환승 도로)을 조정하며, 다른 국가의 정규직 근로자와 국경 소음(선박 교통 및 무작위 전환)을 살펴본다. 이러한 데이터는 또한 여행자의 체류 기간을 제공한다: 첫 번째와 마지막 로밍 활동 사이의 인바운드 여행자의 경우, 인바운드 여행자의 경우 인바운드 여행자의 경우와 동일하며, 한 번의 여행으로 둘 이상의 국가를 지도화한다.

이점: 기존 조사 및 행정 자료의 데이터와 비교하여 이러한 데이터를 사용함으로써 상당한 이득을 얻을 수 있다. 향상된 품질: 기존 조사 기반 추정치에 비해 추정치의 정확성, 품질 및 포괄성이 향상되었다. 포획된 비급여 호텔(친지 또는 친구와 함께 체류)과 미등록 호텔 및 기타 숙박 시설을 갖춘 여행자들은 견적을 더욱 포괄적으로 만들었다. 여행 서비스는 지역별 또는 거주 국가별로 보다 상세하게 분류될 수 있다. 시기: 데이터는 거의 실시간으로 제공되므로 월별 및 분기별 추정치를 실질적으로 시간 지연 없이 생성할 수 있다. 비용: 데이터를 쉽게 사용할 수 있으므로, 조사, 관리 및 기타 출처의 데이터 수집 및 처리 비용보다 훨씬 저렴하다.

Source: http://www. oecd. org/trade/its/46287481. ppt.

34. 통계기관은 변화에서 앞서기 위해 사용자들에게 부가적인 서비스를 제공함으로써 빅데이터와 함께 제공되는 새로운 기회를 포착해야 한다. 여기에는 조사 결과의 적시성을 위한 플래시 추정치의 작성과 경제적, 사회적 관심사에 대한 혁신적인 단기 지표의 생산이 포함될 수 있다. 또한 통계기관은 제3자 또는 공공 또는 민간 부문에 의해 생성된 데이터 세트의 인가나 인증과 같은 새로운 업무를 고려할 수 있다. 그 권한을 확대함으로써, 그것은 품질을 통제하고 투명성, 적절한 품질 및 건전한 방법론의 테스트에 실패한 데이터 세트를 조작하는 개인 빅데이터 생산자와 사용자의 위험을 제한하는데 도움이 될 것이다(MacFee 2016).

Box 6. 관리데이터 및 빅데이터

행정데이터는 (일반적으로) 대규모 행정 시스템의 부산물로 발생하며 일반적으로 공식 통계 이외의 목적으로 생성되는 빅데이터의 독특한 형태로 간주될 수 있다. 세계의 대부분의 통계기관은 행정자료를 이용한다. 조사("기본데이터") 대신 잠재적으로 광범위한 커버리지가 있는 비통계적 목적("2차데이터")을 위해 수집된 데이터에 의 존하는 것은 제한된 자원과 감소하는 응답률에도 불구하고 점점 더 많은 더 나은 통계에 대한 수요를 해결하는데 도움이 되었다. 세금 자료, 공적 자금 거래 자료, 사회보장 기록 등이 대표적인 예다.

행정데이터에 대한 수년간의 경험은 통계 기관이 빅데이터를 사용할 수 있는 길을 닦은 것일 수 있다. 통계기관은 데이터 품질, 방법론적 건전성, 프라이버시 보호 및 기밀성 보호를 보장해야 혁신적인 데이터 출처를 활용할 수 있다는 점을 염두에 두고 있다. 이러한 전제조건은 1차 및 2차데이터 출처를 적절하게 관리할 수 있는 컴파일 당국의 능력과 배포된 공식 통계에서 대중의 신뢰를 유지하는데 중요하다. 대부분의 통계기관은 유럽 북유럽 국가(덴마크, 핀란드, 아이슬란드, 노르웨이, 스웨덴)와 같은 예외를 제외하고 행정 자료에서 데이터 수집 프로세스의 설계 및 운영에 대한 통제 또는 영향력이 제한적이다. 그러나 대부분의 빅데이터와 달리 행정 기록은 공식 통계청과 동일한 공공 부문에서 비롯된다. 그렇기는 하지만 빅데이터의 사용은 행정데이터의 통합보다 더 어려운 것으로 판명될 수 있다.

모든데이터 수집 프로그램과 마찬가지로, 소스데이터의 사용에 대한 고려는 비용과 편익의 균형을 맞추는 일이다. 행정 출처의 2차데이터를 사용하면 수집 비용이 절감되고 자주 비난 받는 응답자의 부담이 감소하여 통계 기관은 새로운 데이터 수요를 협상할 때 여유를 갖게 된다. 행정기록은 다음과 같은 다양한 통계적 용도를 가지고 있다. (1) 일부 지표의 유일한 출처로서(예: 상품거래에 대한 관세자료, 국제거래보고시스템(International Transactions Reporting Systems for Payment Statistics)), (2) 기타 출처(예: 생산을 측정하기 위한 세무기록, 소형버스용 과세자료 사용). 필수사항), (3) 검증 및 교정을 위한 (예를 들어, 관련 행정 프로그램의 추정치와 조사 추정치의 비교), (4) 간접 추정(벤치마킹), (5) 조사 설계에 대한 (세부 내용의 풍부함을 이용)
북유럽 국가에서는 통계에 대한 레지스터의 사용이 잘 확립되어 있다. 벤치마크 예로는 세계 최초로 덴마크에 등록한 인구 및 주택 검열을 들 수 있다(1980—핀란드(1990), 노르웨이와 스웨덴(2011). 인구 구조는 1970년대 모든 북유럽 국가에서 등록부를 기반으로 작성되었다.

등록부는 다양한 방법으로 관리데이터를 사용하는데 중심적이다.

- 하나의 레지스터에 기초한 통계(인구구조, 중요 통계)
- 여러 레지스터를 함께 결합(인구 및 주택 조사)

- 레지스터와 측량데이터 결합
- 샘플링 프레임
- 품질 관리

행정데이터의 효과적이고 효율적인 사용을 위한 전제조건은 다음과 같다.

- 국가 통계청이 공식 통계를 작성하기 위한 행정 출처에 대한 접근 권한
- 모든 개인과 기업에 대한 고유한 식별 정보, 모든 관리 레지스터에서 널리 사용되어 서로 다른 소스의 데이터를 결합함
- 기밀 유지; 즉, 행정 자료의 생성 및 수집에 있어 통계 기관의 영향을 집계한 형태에서만 사용됨: UNWDF 2017.

IV. 빅데이터로 인해 발생하는 것은 무엇인가?

A. 데이터 품질

35. 정책 수립을 위해서는 빅데이터에서 도출된 지표의 품질평가가 지배구조, 정치, 평판 리스크를 최소화하는데 중요할 것이다. 시기적절하고 보다 세분화된 데이터에 대한 요구가 강하지만 지표의 품질과 기초적인 데이터 출처를 평가해야 한다. 실험 단계에서는 새로 생성된 지표를 기존 지표를 벤치마킹해 평가할 필요가 있다. 이는 새로운 지표들이 현실, 외부, 재정, 통화 및 재무 통계에 대해 오랫동안 확립된 품질 프레임워크에 정의 된 최소데이터 품질 표준을 충족하도록 보장하기 위한 것이다.

36. 어떤 새로운 데이터 소스의 적합성은 정확성, 지속성, 방법론적인 건전성 등 다수의 핵심 특징에 대해 평가할 필요가 있을 것이며, 반면에 메타데이터[18]은 새로운 데이터 소스를 해석하고 평가하는 열쇠가 될 것이다. [19] 빅데이터를 생성하는 프레임워크와 상호작용이 f에 있을 것이라는 보장은 없다. 미래 빅데이터는 대부분 민간에서 비즈니스 모델과 기술의 부산물로 생산되기 때문이다. 그것은 경쟁 시장의 맥락에서 변화될 수 있다. 따라서 시계열의 가용성, 비교 가능성 및 일관성은 위험하다(Kitchin 2015).

37. 많은 유형의 빅데이터는 모집단의 무작위 샘플을 나타내지 않는다. 예를 들어, 소셜 미디어 서비스의 경우, 기술을 사용하지 않는 모집단 하위 그룹은 달리 캡처하거나 조정하지 않더라도 충분히 대표되지 않을 것이다. 빅데이터의 사용자는 특정 빅데이터 소스가 방법론적으로 건전한지 여부를 판단해야 하며 분석 대상 인구를 나타내야 한다. 인구, 단위 및 이벤트, 적용된 방법과 프로세스, 그리고 이들의 적합성과 완전성에 대한 메타데이터는 매우 중요하다(UNECE 2014).

38. 빅데이터를 기반으로 한 지표는 시간적 범위가 짧고 특이치가 들어 있어 연속성을 보장할 수 없다. 이 부산물에 접근하는 것은 최근에야 시작되었다. 이것은 시간의 경과에 따른 비교 가능성을 제한한다. 빅데이터는 구조화되지 않은 경우가 많으므로 특이치를 대체하고 관측치를 추정치로 누락하는 등 시계열 관측치 및 세척 변수로 적절히 변환해야 한다(Eurostat 2016). 데이터 제공의 지속성은 규제 프레임워크에 의해 보장될 수 없지만 공식 통계에서 빅데이터를 사용하는 것과 관련이 있을 것이다. 전반적으로, 불안정성은 제도적 변화와 데이터 제공의 불연속성 때문이기도 하지만, 기술 향상과 그에 따른 소비자 행동 변화 때문이기도 하다.

18 "메타데이터"라는 용어는데이터 액세스, 통계 개념, 컴파일 관행, 방법론 등데이터 생산 주기의 모든 측면에 대한 정보를 제공하는 메타데이터를 말한다.

19 이러한 요인들은 각국의 데이터 품질에 대한 종합적인 평가에 사용되는 IMF의 데이터 품질 평가 프레임워크의 중심에 있다(2012년). 관련성, 정확성, 신뢰성, 적시성, 접근성, 해석성 및 일관성의 7가지 특징에 걸쳐데이터 품질을 측정한다.

39. 품질은 공식 통계기관에 의 한 자료의 생산에 있어서 중심적인 관심사다. 공식 통계는 공익이며, 그 전문 표준과 규범은 UN의 기본 공식 통계 원칙(UNSD 2014)에 반영되어 있다. 혁신적 출처나 방법은 공식 통계를 생산하기에 적합하다는 평가를 받을 필요가 있다. 5V 특성에 의해 표현된 바와 같이 빅데이터는 복잡하고 불완전하며 소음이 심하며 특이치와 극단적인 이벤트를 포함할 수 있다. 품질 표준에 대한 경험, 모범 사례, 종합적인 분석, 방법론 및 연구의 부족은 통계 생산자가 준수하는 품질 표준을 훼손할 수 있다. 빅데이터를 실험하고 있는 경제협력개발기구(OECD) 국가들이 주저하는 것은 주로 사용되는 데이터 소스와 관련된 방법론적 프레임워크가 없기 때문이다(UNSD 2015). 앞으로, 진실성과 변동성을 다루는 통계 기법 및 방법론 모범 사례의 연구와 편찬, 구체적으로는 통계 커뮤니티의 최우선 과제가 될 필요가 있다.

40. 빅데이터의 장기적인 이용은 제쳐두고, "현재"이 가능한 " 현재" 정보는 구체적인 행동에 대해 " 현재"으로 사용할 수 있다(Meyer 등 2013). 빅데이터는 포괄적인 데이터 품질 표준을 준수하지 않을 수 있지만, 여전히 우리에게 어떤 일이 일어나고 있다는 것을 경고함으로써 의미 있는 통찰력을 발견할 수 있다. 한 가지 예는 의 견이나 추세를 위해 구조화되지 않은 데이터의 다양한 출처를 채굴하는데 사용되는 정서 분석이다.

B. 빅데이터 액세스

41. 빅데이터의 출처와 생성은 행정자료의 현저한 예외를 제외하고는 대부분 국가나 국제기관의 통제를 벗어난다. 빅데이터에 대한 접근은 일반적으로 민간 부문에서 보유하고 있는 독점데이터에 대한 접근을 의미한다. 빅데이터에 효과적으로 접근, 실험, 활용하기 위해서는 이용자가 독립성을 유지하면서 개인데이터 소유자와 합의해야 하며, 프라이버시와 기밀성을 모두 다루는 법적 환경을 보장해야 한다. 액세스가 허가되면 빅데이터 소스 사용자는 데이터 소스, 합법성 및데이터 사용 목적에 대해 투명해야 한다. 기밀을 보장하기 위해 사용되는 전통적인 기법은 한계에 이를 것이며, 사용자들은 그들의 독립성과 명성과 대중의 신뢰를 보호하는 기술과 방법을 찾아야 할 것이다. 예를 들어, 정부가 그들의 편성 기관을 빅데이터 혁명에 빠뜨릴 수 없다는데 동의 할 때 입법부의 중요한 역할은 빅데이터에 대한 접근을 협상하는 것이다.

42. 빅데이터를 이용할 때 프라이버시, 기밀성, 사이버보안 리스크가 주요 관심사다. 빅데이터는 사생활, 기밀성 및 사이버 보안 위험에 노출될 수 있는 많은 양의 민감한 개인 정보를 포함한다. 충분히 보호되지 않을 경우 이 정보는 사이버 공격에 취약할 수 있으며, 개인 프로필에 사용되며, 제3자에게 판매될 수 있다. 정보는 국가의 법적 체계에 따라 개인의 사전 지식 없이는 합법적으로 사용될 수 없다. 데이터의 잠재적 손실은 평판 훼손과 소비자의 신뢰 상실로 이어질 수 있다. 따라서, 국제기관과 공공기관은 개인정보 보호나 기밀성 제도를 위반하지 않고 사용된 자료 출처와 지표를 입수했는지 확인할 필요가 있을 것이다.

43. 세분화된 빅데이터 소스를 사용할 때 개인정보 보호 절차와 정보 기술 기법은 프라이버시, 기밀성 및 사이버 보안 위험을 최소화하는 열쇠가 될 것이다. 기업, 공공기관 및 제3자데이터 사용자는 철저한 개인정보 보호 절차를 수립해야 할 것이다. 그들은 보안 계층에 투자하고 암호, 익명화, 사용자 접근 제어와 같은 전통적인 정보 기술(IT) 기법을 빅데이터 특성(Moreno, Serrano, Fernández-Medina 2016)에 적용하여 프라이버시를 보호하고 데이터가 재구성되고 개인에게 추적되지 않도록 해야 한다. 더욱이, 기업이 수집된 정보로 무엇을 할 수 있는지를 상세하게 설명하면서, 규제 개입을 통해 소비자 권리를 강화해야 할 수도 있다(Smith 등 2012). 전자는 프라이버시, 기밀성 및 사이버 보안 위험을 최소화하기 위해 필수적일 것이다.

44. 여러 나라(UNSD 2015)가 민관 협력관계 구축에 착수했지만, 접속권은 때로는 어드레스가 없고 불명확한 경우도 있다. 이러한 문제를 극복하기 위해 UN 글로벌 워킹 그룹은 현재 공식 통계 생산자와 데이터 소유자 간의 지속적인 파트너십 구축을 위한 "접근 권고사항"과 모범 사례를 준비하고 있다. 필요한 것은 표준화, 윤리적, 안정된 데이터 공유 툴과 민간 기업과의 프로토콜(Letouzé 및 Jütting 2014)이다. 그러나 변동성의 위험은 계속된다. 분석 및 감시 애플리케이션이 연속성 측면에서 위험에 처할 가능성이 있는 미래에는 회사와 데이터가 존재할 것이라는 확신이 없다.

45. 빅데이터의 가격이 반드시 낮거나 무시할 수 있는 것은 아니다. 수요가 있는 곳에는, 빅데이터 생산 비용이 기업에 미미할 수 있는 반면, 이 정보에 접근하기 위한 허가나 적정 비용은 상당할 수 있다. 이러한 비용은 외부 공급업체를 통해 사내에서 사용하거나 액세스하기 위한 처리 및 저장 기술을 획득하는 것 외에도 발생할 것이다.

46. 빅데이터는 단순한 부산물이 아니라 기업의 주요 자산이 되고 있다. "빅데이터 혁명"에서 발돋움한 신기업은 데이터 브로커, 컨설턴트[20], 현장 기술 플랫폼을 설치하는 기업에서부터 공공 또는 독점 클라우드에서 유료로 빅데이터를 서비스로 제공하는 이른바 클라우드 벤더에 이르기까지 다양하다. 점점 더 많은 민간 기업들은 이윤을 위해 그들의 데이터를 판매한다; 시간이 지남에 따라 빅데이터의 생성은 단순히 부산물이 아니라 그들의 활동의 주요 목표가 될 수 있다. 국가 및 국제 정책 입안자들뿐만 아니라 통계계와 공식 사용자들은 복잡한 협상 과정을 포함한 많은 결정에 직면할 것이다.

47. 선택할 수 있는 수많은 기술 옵션과 플랫폼에도 불구하고, 선택 과정은 상당히 복잡하다. 빅데이터 관행 설정에 기업이 따르는 대표적인 접근방식은 기술 옵션을 중심으로 한 역량 중심 구축이다. 그들은 또한 전문가들과 실무자들이 접근할 수 있는 탐구적인 환경을 구축했다. 첫해에는 (1) 다양한 프로젝트에 종사하는 팀이 프로비저닝할 수 있는 표준 기술 플랫폼을 식별하고, (2) 환경과 관련된 사용, 모니터링 및 비용을 중심으로 거버넌스를 구축하며, (3) 특정 기술을 가진 전문가를 활용할 수

20 www. bigdatapartnership. com.

있는 네트워크를 구축하고, (4) 사업 및 예산 규정을 만들어 향후 비용을 조달한다.

48. 클라우드 기술은 필요에 따라 기술 인프라를 확장하거나 축소할 수 있는 유연성을 제공한다. 이것은 업계에서 선호되는 접근방식이며 일반적으로 첫 해에 약 25만 달러의 예산 배분을 요구할 수 있다. 이후의 연도에 대한 비용의 취급은 "사용하는 것에 대한 지불"에 근거하고 있으며, 자금의 효율적인 사용을 보장하기 위해 강력한 거버넌스를 필요로 한다. 클라우드 기반 시스템은 비교기구의 채택이 확대되고 스토리지 및 컴퓨팅 전력 비용은 계속 하락할 것으로 예상된다. IMF는 클라우드 서비스 프로비저닝을 시작했고 향후 몇 년 동안 빅데이터를 분석 관행에 통합하면서 관련 비용이 점차 증가할 것으로 예상한다. 데이터 비용은 더 예측하기 어려울 것이다. 유용할 수 있는 일부 소스는 아직 별개의 제품으로 존재하지 않거나 심지어 사용할 수 없을 수도 있기 때문이다. 향후, 일부 데이터는 공용 도메인에 속할 수 있는 반면, 일부는 상업적 소스의 면허를 요구할 수 있다. 클라우드 기반 컴퓨팅 서비스와 상용데이터는 일반적으로 반복적으로 라이센스가 부여되며, 이는 행정 예산에 영향을 미친다.

C. 새로운 스킬 프로필 및 기술

49. 빅데이터를 발표하려면 다원적 팀이 필요할 것이다. 빅데이터 프로젝트에 종사하는 통계기관은 풍부한 인적, 기술적 자원과 협력하여 빅데이터를 능숙하게 활용해야 하는 중요성을 알고 있다. UN 글로벌 워킹 그룹은 신중하게 설계된 설문지와 같은 전통적인 수집 수단에 의 존하기 보다는 다른 전문적 배경을 가진 다분야 프로젝트팀이 데이터에 대한 분석을 적응시킬 필요가 있다고 결론지었다. 통계기관, 중앙은행, 공공기관, 국제기구는 빅데이터를 처리하기 위해 기존 인력을 양성하고 개발해야 할 뿐만 아니라 민간과 경쟁해 신규 인력을 채용하거나 감축해야 할 것이다. 이러한 새로운 직원은 빅데이터(예: 데이터 과학자(그림 3 참조), IT 아키텍처 전문가 및 데이터 시각화 전문가와 함께 작업하는 데이터 시각화 전문가)에 익숙해야 한다. 평균적으로 초기 단계에서 빅데이터 실행은 기술 및 비즈니스 기술이 혼합된 3~4명의 핵심 팀으로 시작한다.

그림 3. 데이터 사이언티스트

Box 7. 정보기술과 IT 거버넌스의 재고

피터 라이먼과 할 R. 바리안 2000의 2000년 연구 "얼마나 많은 정보?" (라이먼과 바리안 2000)는 "세계는 매년 1에서 2엑사바이트의 독특한 정보를 생산하는데, 이는 지구상의 모든 남자, 여자, 아이에 대해 약 250메가바이트의 고유한 정보를 생산한다"고 기술한 물리 매체에 저장된 정보의 총량을 계량화한 최초의 것이었다. 또한 2000년에는 프랜시스 X. 디볼드는 "빅데이터는 이용가능하고 잠재적으로 관련성이 있는 데이터의 양에서 폭발적으로 증가하는 것을 의미하며, 주로 최근의 그리고 전례 없는 진보주의 자들의 결과"라고 기술하는 "빅데이터" 동역학적 요소 모델"이라는 논문을 에코메트릭 협회의 제8차 세계회의 에 제출했다. 데이터 기록 및 스토리지 기술의 ts. " 2001년에 Doug Laney는 현재 일반적으로 받아들여지는 빅데이터의 특징인 3V를 설명했다.

빅데이터는 단순히 큰 것이 아니라 확장성이므로 정보 기술과 IT 거버넌스를 재고해야 한다. 새로운 데이터 출처와 분석 기법이 도입됨에 따라 오랜 기간 지속된 사업 관행과 레거시 기술을 재고할 필요가 있을 것이다.

정보 보안 및 개인 정보: 트랜잭션 레벨 또는 관찰 레벨데이터가 있을 때 복수의 소스에서 얻은 데이터의 입증도를 이해하고, 데이터 프라이버시를 위한 국가 및 국제 표준을 준수하는 것이 공공 부문 이니셔티브의 신뢰성에 필수적이다.

연구 프로젝트의 거버넌스: 국제통화기금(IMF)의 데이터 관리 관행은 제도적 가치가 있는 데이터를 보존하고 문서화하기 위해 이사회 논문 발표와 세계경제전망 생산을 중심으로 구조화된 프로세스에 의존해 왔다. 빅데이터 프로젝트는 그러한 프로세스와 관련될 가능성이 적기 때문에 데이터 보존, 데이터 파괴, 대형 모델 실행 비용, 소프트웨어 코드와 알고리즘 공유와 같은 문제를 다루는 지침이 필요할 것이다.

Life Cycle을 통한 연구 프로젝트 관리: 기술 그룹은 소프트웨어와 시스템을 생산 환경에서 지원할 수 있도록 조직되고 최적화된다. 빅데이터 연구 프로젝트는 반복성이 높고 시스템 구축이 아니라 새로운 데이터 세트, 모델 및 시각화를 초래할 수 있다. 이러한 노력은 기술이 보조적인 역할을 하는 특정 프로젝트에 의해 주도될 가능성이 높다.

오픈 소스 소프트웨어가 제시하는 과제 및 기회: 빅데이터 연구 및 분석에 사용되는 기술 중 상당 부분은 오픈 소스 소프트웨어로 빠르게 진화하며 느슨하게 조정된 실무자 커뮤니티에 의해 관리되며 군중 소싱 협업 공간을 통해 지원된다.
IMF에서 이와 유사한 지원 체계를 채택하는 것은 처음에는 반문화적일 것이다.

클라우드 컴퓨팅: 많은 조직들이 클라우드를 정보기술의 플랫폼으로 채택함에 따라 느리고 꺼려하는 모습을 보이고 있다. 빅데이터의 경우, 낮은 비용, 새로운 서버와 소프트웨어의 구축 속도, 그리고 시장에서 사용할 수 있는 급속한 성숙 서비스 및 보안 모델 때문에 이러한 거부감을 해소해야 할 것이다.
This box was prepared by El Bachir Boukherouaa, IMF IT Department.

V. STATISTICAL IMPLICATIONS

50. 통계기관은 빅데이터 프로젝트에 대한 참여를 더욱 강화하고 방법론 개발을 주도하는 국제기구와 긴밀히 협력하여 이미 달성한 것이 무엇인지를 파악해야 한다. 통계기관은 빅데이터 커뮤니티에서 이미 진행 중인 프로젝트와 작업으로부터 기여, 학습 및 이익을 얻을 수 있다. 빅데이터에 관한 유엔 글로벌 워킹 그룹과 같은 기존의 빅데이터 네트워크를 활용하는 것뿐만 아니라 기여하는 것도 필수적일 것이다. 국제 및 국가 협력을 촉진하는 빅데이터를 중심으로 한 강력한 거버넌스 프레임워크는 중복을 피하고 효과적인 협업을 보장할 필요가 있을 것이다.

51. 통계기관은 공식 거시경제와 금융통계를 풍부하게 할 잠재력이 있는 사업들을 재고할 필요가 있다. 데이터 품질 문제와 방법론의 판단에 대한 전문지식이 필요하지만, 제도 변경 관리, 전략적 제휴 구축, 사용자와 의 커뮤니케이션 등의 분야에서도 필요하다. 베스트 프랙티스(best practice)의 재고를 위해서는 학계, 민간, 국제 통계계의 빅데이터 전문가 네트워크와 의 강력한 연계를 도모하고, 네트워크를 활용해 회원국의 교육, 기술, 역량 개발에 대한 국제적 노력을 효율화할 필요가 있다. 빅데이터 애플리케이션은 10년 전 행정데이터의 통합에 버금가는, 그러나 보다 혁명적인 방법으로 통계 시스템에 패러다임 전환을 도입할 수 있다. 그러나 관리데이터의 경우, 데이터 가용성, 접근성, 그리고 따라서 국가 및 조직 전반에 걸쳐 진보가 균일하지 않을 것이다.

52. UNECE High-Level Group for the Modernization of Offic Statistics (HLG- MOS)은 값비싼 모험을 하기 전에 샌드박스[21]를 만들기로 현명한 결정을 내렸다. 샌드박스(Sandbox)는 통계기관과 다른 사용자가 도구, 기술, 워크플로우 및 방법론을 테스트 및 평가하고 향후 동료들과 협업하기 위해 활용할 수 있는 공유 플랫폼 및데이터 저장소(비밀 유지 제약에 따름)이다. 빅데이터를 사용하는 것의 결론은 물론 편익이 비용을 초과할 수 있는지, 그리고 일정 수준의 품질과 국제적 비교 가능성을 보장할 수 있는지 여부다.[22] UNECE 샌드박스 프로젝트에 참여하는 것은 무엇보다도 공식 통계 작성에 사용될 수 있는 접근방식에 대한 논의를 촉진하기 위한 적절한 환경으로 권장된다(UNSD 2015). 특히, 통계 시스템이 덜 발달된 국가는 샌드박스의 혜택을 받을 수 있다.

53. 거시경제와 금융 통계에 대한 가장 유망한 빅데이터 출처는 시간이 지남에 따라 개별 국가에 대한 비대칭적 기회를 통해 확인될 것이다. 소셜 네트워크, 전통적인 비즈니스 시스템 및 사물 인터넷은 표준 통계 영역(국가 계정, 외부 부문, 금융, 정부 재정 및 가격 통계)과 추가 분야로 공급될 수 있는 데이터 유형을 생성한다. 국가의 특성을 고려하고 개념증명을 통과한 후, 통계학자들은 이해관계자들과 공동으로 거시경제와 금융 통계에 가장 유망한 빅데이터 프로그램을 수립할 것이다. 새로운 필수 기술,

21 http://www1. unece. org/stat/platform/display/bigdata/Sandbox. 현재 연간 수수료는 10,000유로다. 플랫폼은 최대 56테라바이트의 데이터를 안정적으로 저장하고 80개의 CPU 코어에 걸쳐 이러한데이터를 고성능 동시 처리할 수 있는 5개의 서버로 구성되어 있다. 사용자는 고대역폭 네트워크 연결을 통해 인터넷에 연결된 통합 로그인 포털에 액세스한다.

22 통계 네덜란드(CBS)는 빅데이터 샌드박스를 프로토타입으로 사용해 자체 내부 플랫폼을 만들었다

빅데이터 사용자의 기술, 개별 국가의 가용 소스데이터의 비용과 복잡성을 고려할 때, 개별 국가의 실제 기회는 비대칭적일 수 있다. 즉, 모든 국가가 같은 방식으로 빅데이터로 부터 이익을 얻는 것은 아니다.

54. 표 1(부록 II)은 다른 유형의 빅데이터와 표준 거시경제, 금융 및 기타 통계 간의 지도를 제시한다. 4개의 열로 구성된 표 1은 박스 1의 빅데이터 분류를 사용하며 이를 기존의 통계 영역과 연계하고 새로운 빅데이터 출처 및 유형으로부터 잠재적인 이익을 얻을 수 있는 추가적인 통계에도 연결한다.

55. 거시경제와 금융 통계와 관련된 유망한 빅데이터 애플리케이션이 많은데 어디서부터 시작해야 할까? 거시경제와 금융 통계에 대한 가장 유망한 빅데이터 출처는 시간이 흐르면서 파악되겠지만, 국가의 특성을 고려할 때 이미 중기의 기반이 될 수 있는 빅데이터 애플리케이션이 많이 진행되고 있다. 표 2는 출발점으로 작용할 수 있다. [23]

56. 일부 빅데이터 소스에는 선호도가 있지만 파생 지표는 기존의 모든 통계 영역에 걸쳐 있습니다. 실제적인 설명을 위해 표 2 (부록 III)는 빅데이터 애플리케이션과 거시 경제, 재무 및 기타 통계에 대한 잠재력에 대한 광범위한 개요를 제공합니다. 4 개의 열로 구성된 표 2는 이러한 응용 프로그램을 보여주고 사용 된 빅데이터 소스, 파생 된 지표 및 잠재적으로 이익이 되는 통계적 도메인 사이의 링크를 만듭니다. 빅데이터 사용 동기는 섹션 III에 정의 된 세 가지 빅데이터 영역을 따라 구성됩니다. (1) 새로운 질문에 답하고 새로운 지표를 생성하고, (2) 공식 통계의 가용성에 있어 시간 지연을 메우고 더 많은 것을 지원합니다. 기존 지표의 적시 예측 및 (3) 공식 통계 생성의 혁신적인 데이터 소스. Google 트렌드, 트위터, 페이스 북, 공식 레지스트리, 스캐너 및 가격데이터와 같은 일부 빅데이터 소스에 대한 선호도를 확인할 수 있지만 추론 된 지표는 모든 기존 통계 도메인을 포괄하고 충분한 새로운 애플리케이션 및 통계를 암시합니다.

57. 공식 통계는 새로운 데이터 품질 개념을 개발하고 기존 프레임워크를 확장하여 빅데이터와 함께 오는 기회와 과제를 통합할 필요가 있다. 비구조적 데이터(Piatetsky-Shapiro 2012)에서 통찰력을 추출하면(예: 기업의 대차대조표 사용) 컴파일 프로세스와 통계 결과물을 다시 방문해야 하는 공식 통계에 새로운 지평을 만든다. 데이터 품질 프레임워크는 새로운 데이터 환경에 맞춰 조정되어야 한다. 동시에, 시간과 국가 간의 데이터 비교 가능성은 가능한 범위까지 보존되어야 하는데, 이는 빅데이터의 본질에 의해 도전 받을 목표다. 통계학자와 사용자 모두 결과적인 절충을 염두에 둘 필요가 있다.

58. 빅데이터 출처 및 방법의 데이터 품질을 평가하고 확인하기 위해 공식 통계 커뮤니티는 통계 및 기타 전문가 그룹에 관한 상임 위원회를 통해 협력하고 노력을 조정해야 한다. [24] 국제통화기금 BPM6 컴파일 가이드 및 UNECE 글로벌 생산 측정 가이드와 같은 국제 통계

23 빅데이터 인벤토리는 https://unstats. un. org/bigdata/inventory/에서도 확인할 수 있다.

24 경제금융통계 국제기구그룹, 국제통화기금수지통계위원회(BOPCOM), 정부재정통계자문위원회(GFSC), 국가계정에 관한 사무국간 워킹그룹(ISWGNA) 등

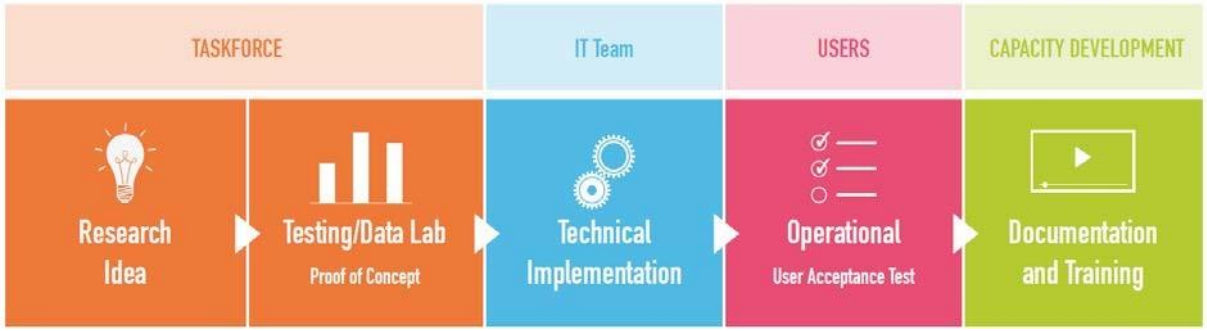
매뉴얼과 가이드는 통계 관행의 개발 및 적용에 대한 견실한 근거를 지속적으로 제공하고 빅데이터에 의해 제기된 새로운 기회와 과제를 통합하기 위해 업데이트되어야 할 수 있다. 공식 통계에 빅데이터를 통합하는 적절한 방법에 대한 상임 위원회의 견해, 조언 및 검증은 데이터 세트의 지속적인 신뢰성, 정확성 및 방법론적 건전성에 대한 확신을 제공할 것이다. 이는 국가 및 국제 기구가 이미 수행한 작업을 토대로 하여 실질적인 지침 노트와 메타데이터 개요를 작성하는 것으로 이어질 수 있다.

59. 중앙은행과 IMF 등 내부 조직과 통계 및 사용자 부서가 힘을 합쳐 정책 수립에 활용할 수 있는 핵심 조기경고 및 스냅샷 지표 세트를 파악해야 한다. 통계 지원은 빅데이터 지표의 적합성과 방법론적 건전성에 대한 평가와 검증의 열쇠가 될 것이다. 새로운 지표에 대한 빅데이터의 사용은 적용 방법론과 데이터 출처 측면에서 투명하게 이루어져야 한다. 그렇지 않으면 정책 조언과 예측의 가치가 심각하게 약화될 수 있다. 통계부서는 모범 사례에 대한 지침 노트 작성과 이러한 지표를 의식적으로 사용하는 직원을 지원하기 위한 견고한 개념의 입안에 주도적인 역할을 할 것이다. 정보 기술 부서는 기술 지원을 제공하기 위해 중요한 이해관계자가 될 것이다.

60. 사일로를 파괴하고 선택된 빅데이터 프로젝트에 집중하기 위해 특화된 태스크 포스가 설립 될 수 있다. 부서 간 그룹 외에도 전문 주제 관련 및 IT 전문가로 구성된 전문 태스크 포스팀을 설립 할 수 있습니다. 태스크 포스는 파일럿 프로젝트에 전념 할 수 있으며, 고려중인 파일럿 프로젝트의 진행 상황, 과제 및 기회에 대해 부서 간 그룹에 보고합니다. 태스크 포스에 여러 분야의 전문 지식을 활용함으로써 빅데이터는 국가 및 국제 기관이 사일로를 깰 수 있는 기회가 될 것입니다.

61. 이들 전문 태스크 포스는 원하는 결과를 식별하기 위한 브레인스토밍 단계(그림 4)로 시작할 수 있다. 1단계와 2단계는 부서간 그룹이 설정한 시간 목표 내에서 작업하면서 확립된 태스크 포스에 배치될 수 있다. 3단계는 태스크포스 전문가와 협력하여 IT 팀이 주도하는 기술적 구현이 될 것이다. 동시에, 신제품은 감시 작업에 대한 "목적 적합성" (적절한) 사용자들에 의해 평가되어야 할 것이다. 통계부서는 방법론적 지침을 제공해야 한다. "목적 적합성"으로 간주될 경우, 신제품은 시범적으로 제작되고, 감시 업무에 통합되며, 더 큰 규모로 확대될 수 있다. 마지막 단계에서, 기관은 신제품을 역량 개발 활동(훈련, 기술 지원)에 포함시킬지를 결정할 수 있었다.

그림 4. 특별 태스크포스별 파일럿 연구



62. 전문화된 태스크 포스의 결과는 동적 및 대화형 지침 노트일 수 있다. 지침 참고사항에는 방법론적 건전성, 데이터 출처의 신뢰성 및 국제 및 국가 조직 모두에 대한 정책 분석에 대한 입력 자료 역할을 할 수 있는 잠재력에 대한 파일럿 프로젝트의 평가가 수반된다. 지침 노트는 동적(시간별 및 고주파 업데이트)과 대화형(대화형 링크 및 검색 기능이 있는 온라인 기반 문서)이어야 한다. 일부 프로젝트의 경우, 새로운 방법론을 개발해야 할 수도 있다.

63. 용량 개발 및 기술 지원. 개발도상국에 대한 기술적 지원에 강력하게 관여하고 있는 IMF와 같은 기관의 경우, 통계와 양자 및 다자간 감시에 직접 사용할 수 있는 지표를 위해, 여기에서 언급된 그들의 잠재력과 과제를 더 탐구한 후에 역량 개발 프로그램에 통합될 수 있다. 종이. 역량 개발은 데이터 품질 문제와 방법론의 판단에 대한 전문지식의 제공뿐만 아니라, 기관 변경 관리, 전략적 파트너십 구축, 사용자와 의 커뮤니케이션에 관한 모범 사례의 전달을 포함할 수 있다. 베스트 프랙티스(best practice)의 재고를 위해 국제기관은 학계, 민간, 국제통계의 빅데이터 전문가 네트워크와 강력한 연계를 구축하고, 네트워크를 활용해 회원국의 교육, 기술, 역량 개발에 대한 국제적 노력을 효율화할 필요가 있다. 9개의 다른 국제 및 지역 조직과 22개 국가로 구성된 UNSD 글로벌 워킹 그룹은 공식 통계 분야의 빅데이터에 대해 이미 확립된 전문가 그룹이다.

VI. 결론 및 작업 머리

64. 지금까지 많은 국내외 통계기관들은 빅데이터가 단순한 유행어가 아니라 비전과 계획을 필요로 하는 잠재력 있는 전략자산이라고 인식했다. 빅데이터는 공식 통계를 보완할 수 있는 가장 유망한 애플리케이션을 선정하고, 시의 성 향상, 기존데이터 세트의 예측 지원, 새로운 지표 제작 등 부가가치를 가져오는 전략이 필요하다.

65. 조직은 기존의 개인 및 분산된 빅데이터의 애플리케이션을 뛰어넘어야 한다. 빅데이터 분석에 착수하는 조직은 측정 가능한 높은 수준의 결과를 제공하기 위한 전략적 조직 계획을 필요로 한다. 특히 저소득 국가에서는 비용이 많이 들고 시간이 많이 소요되는 투자에 착수하기 전에 국제 및 국가 기구는 개념 증명이나 시범 사업으로 시작해야 하며, 그 결과가 조직의 관점에서 가치 있고 실현 가능한 것으로 입증된 후에야 사업을 운영해야 한다. 건전한 파트너십, 법적 문제, 올바른 기술과

기술은 통계적 전문지식, 자료의 대표성과 방법론적 정확성, 그리고 데이터 과학자와 주제 경제학자들 사이의 효과적인 협업만큼 중요하다. 특히 공식 통계의 영역 안에서 국제 조정 노력이 관건이다.

66. 공식 통계기관과 데이터 소유주 간의 지속적인 파트너십 구축을 위한 모범사례는 개발 및 시험되고, 법적 문제가 명확해지고, 모범사례는 현장에서 검증되고 있다. 조직들은 빅데이터 성공이 하나의 기술을 구현하는 것이 아니라 빅데이터 혁신을 추진하고 이를 실현하는 사람과 프로세스의 환경을 통합하는 것이라는 사실을 알게 된다. 인프라스트럭처 공간과 빅데이터 소스는 지속적으로 발전하고 있으며, 이를 통해 가능성, 과제 및 제한 사항이 대두되고 있다. 필요한 다양한 기술과 협업을 감안할 때 빅데이터 프로젝트도 제도적 사일로를 타개할 수 있는 기회.

67. 거시경제와 금융 통계에 빅데이터가 제공하는 실제 기회는 통계 영역에 따라 상당히 다르다. 흐름과 거래, 통찰력, 상관관계, 동향 및 정서에 대한 통계에 대한 유망한 기회가 있지만, 현재는 주식에 대한 통계나 거래로 유입되는 흐름의 분류, 재평가 및 기타 볼륨 변경에 대해서는 그렇지 않다. 국제적으로 합의 된 표준에 따른 세부 국가별 시간 시리즈는 시간 경과에 따른 각국의 경제 성과와 정책을 측정하고 모니터링하는데 여전히 중요하다.

68. 공식 통계를 담당하는 국제기구는 사용자 부서와 긴밀히 협력하여 작업해야 한다. 역량 개발 활동에 통합되는 것을 포함하여 국제 통계 조정과 협력을 위한 새로운 차원을 고려해야 한다. 국제기구의 통계부서와 더 넓은 통계공동체의 통계부서는 각각의 전문분야에서 빅데이터 논의 에 기여할 수 있다. IMF의 통계부는 "건전한 통계 관행의 개발과 적용을 위한 강력한 리더십을 제공하는 임무"를 계속하고, 거시경제와 금융 통계에 대한 빅데이터의 활용을 정책 입안에 대한 투입으로 육성하기 위해 IMF와 공동체의 나머지 부분에 손을 뻗칠 것이다.

69. 빅데이터는 정적이 아니라 동적인 현상이므로 빅데이터를 생성하는 시스템과 네트워크는 빅데이터가 제공하는 기회, 빅데이터가 제기하는 도전, 그리고 그 통계적 함의 와 함께 계속 진화할 것이다.

VII. REFERENCES

Arouri M. , A. Aouadi, P. Foulquier, and F. Teulon. 2014. Can Information Demand Help to Predict Stock Market Liquidity? Google It! https://www. ecb. europa. eu/events/pdf/conferences/140407/Aouadi_CanInformationDemandHel pToPredictStockMarketLiquidityGoogleIt. pdf?7dd64c397041aaf1086faf73b3eac25b.

Asian Development Bank (ADB). 2013. Big Data: Vital Statistics for Development. https://www. adb. org/features/big-data-vital-statistics-development.

Banbura, M. , D. Giannone, M. Modugno, and L. Reichlin. 2013. "Now-Casting and the Real-Time Data Flow. " Working paper series 1564, European Central Bank, Frankfurt. https://www. ecb. europa. eu/pub/pdf/scpwps/ecbwp1564. pdf.

Bank for International Settlements (BIS). 2015. Central Bank Use of and Interest in Big Data. Irving Fisher Committee report. http://www. bis. org/ifc/publ/ifc-report-bigdata. pdf.

Central Banking. 2016. Big Data in Central Banking: 2016 Survey. www. centralbanking. com.

Chain Store Age. 2015. Survey: Marketers Value Personalization. http://www. chainstoreage. com/article/survey-marketers-value-personalization#.

Challenge 4 Development. http://www. d4d. orange. com.

Cook, S. , and K. Soramäki. 2014. "The Global Network of Payment Flows. " SWIFT Institute Working Paper No. 2012-006. https://www. swiftinstitute. org/wp-content/uploads/2014/09/SWIFT-Institute-Working-Paper- No. -2012-006-Network-Analysis-of-Global-Payment-Flows_v5-FINAL. pdf.

Data Floq. How Big Data Can Help the Developing World Poverty. https://datafloq. com/read/big- data-developing-world-beat-poverty/168.

Data Revolution Group. 2014. Data Innovation: Big Data and New Technologies. http://www. undatarevolution. org/data-innovation.

Davenport, T. 2006. "Competing on Analytics. " Harvard Business Review. https://hbr.

org/2006/01/competing-on-analytics.

Donovan, K. 2012. "Mobile Money for Financial Inclusion. " In Information and Communications for Development 2012: Maximizing Mobile. Edited by the World Bank Group, Washington, DC. http://siteresources. worldbank. org/extinformationandcommunicationandtechnologies/resources/ic4d-2012-chapter-4. pdf.

European Commission (EC). 2014. Scheveningen Memorandum.
https://ec. europa. eu/eurostat/cros/content/scheveningen-memorandum_en.

———. 2017. Big Data. https://ec. europa. eu/eurostat/cros/content/big-data. en.

European Commission, Eurostat. 2016. Big Data and Macroeconomic Nowcasting: From Data Access to Modelling. http://ec. europa. eu/eurostat/en/web/products-statistical-working-papers/-/KS-TC-16-024.
Federal Reserve Bank of New York (FRBNY). 2017. "Nowcasting Report. " https://www. newyorkfed. org/research/policy/nowcast.
Florescu, D. , M. Karlberg, F. Reis, P. R. D. Castillo, M. Scaliotis, and A. Wirthmann. 2014. Will 'Big Data' Transform Official Statistics? Eurostat. http://www. q2014. at/fileadmin/user_upload/ESTAT-Q2014-BigDataOS-v1a. pdf.

Galbraith, J. W. , and G. Tkacz. 2013. "Nowcasting GDP: Electronic Payments, Data Vintages and the Timing of Data Releases. "
Karabulut, Y. 2013. Can Facebook Predict Stock Market Activity? https://www. ecb. europa. eu/events/pdf/conferences/140407/Karabulut_CanFacebookPredictStock MarketActivitiy. pdf?902eb04ceaa17187b7353be87992b83a.

Karani, K. 2017. Training Course on Gender Statistics and Gender Budgeting. https://www. linkedin. com/pulse/training-course-gender-statistics-budgeting-kennedy-karani-3.

Kitchin, R. 2015. "Big Data and Official Statistics: Opportunities, Challenges and Risks. " Statistical Journal of the International Association of Official Statistics 31 (3) (9).

Laframboise, N. , N. Mwase, J. Park, and Y. Zhou. 2014. "Revisiting Tourism Flows to the Caribbean: What Is Driving Arrivals?" IMF Working Paper 14/229, International Monetary Fund,

Washington, DC.

Letouzé E. , and J. Jütting. 2014. Official Statistics, Big Data and Human Development: Towards a New Conceptual and Operational Approach, Paris 21. https://www. odi. org/sites/odi. org. uk/files/odi-assets/events-documents/5161. pdf.

Lyman, P. , and H. R. Varian. 2000. "How Much Information?" http://www2sims. berkeley. edu/research/projects/how-much-info/summary. html.

MacFeely, S. 2016. "The Continuing Evolution of Official Statistics: Some Challenges and Opportunities," Journal of Official Statistics 32 (4).

Mayer-Schönberger, V. , and K. Cukier. 2014. A Revolution That Will Transform How We Live, Work and Think: Big Data. John Murray Publishers, Great Britain.

Meyer, C. , T. McGuire, M. Masri, and A. Wahab Shaikh. 2013. "Four Steps to Turn Big Data into Action. "
https://www. forbes. com/sites/mckinsey/2013/10/22/four-steps-to-turn-big-data-into-action/#5dcac9094380.

Meyer, B. D. , W. K. C. Mok, and J. X. Sullivan. 2015. "Household Surveys in Crisis. " Journal of Economic Perspectives 29 (4): 199–226.

Moreno, J. , M. A. , Serrano, and E. Fernández-Medina. 2016. "Main Issue in Big Data Security. " http://www. mdpi. com/1999-5903/8/3/44/pdf.

Nathan, M. , and A. Rosso. 2013. "Measuring the UK's Digital Economy with Big Data. " NIESR, July. http://www. niesr. ac. uk/publications/measuring-uk%E2%80%99s-digital-economy-big-data#. WKI5xGeQyos.

Oostrom, L. , A. Walker, B. Staats, M. Slootbeek-Van Laar, S. Ortega Azurdy, and B. Rooijakkers. 2016. "Measuring the Internet Economy in the Netherlands: A Big Data Analysis. " CBS Discussion Paper 2016/14. https://www. cbs. nl/-/media/_pdf/2016/40/measuring-the-internet-economy. pdf.

Organisation for Economic Cooperation and Development (OECD). 1987. https://search. oecd. org/eco/outlook/35252065. pdf.

———. 2015. OECD Digital Economy Outlook 2015. Paris. http://dx. doi. org/10. 1787/9789264232440-en.

Overseas Development Institute (ODI). 2015. "What Is the Future of Official Statistics in the Big Data Era?" Public panel discussion. https://www. odi. org/events/4068-what-future-official-statistics-big-data-era.

Pavlicek, J. , and L. Kristoufem. 2015. "Nowcasting Unemployment Rates with Google Searches: Evidence from the Visegrad Group Countries. " PLOS One, May 22.

Piatetsky-Shapiro, G. 2012. "Managing Uncertainty: Big Data Hype (and Reality). " Harvard Business Review, October 18. https://hbr. org/2012/10/big-data-hype-and-reality.

Predata. 2016. http://www. predata. com.

Smith, M. , C. Szongott, B. Henne, and G. Von Voight. 2012. "Big Data Privacy Issues in Public Social Media. " In Digital Ecosystems Technologies (DEST), 6th IEEE International Conference, 1–6.

Sy, A. , and T. Wang. 2016. "De-risking, renminbi, internationalization, and regional integration. " Africa Growth Initiative at the Brookings Institution.

United Nations Economic Commission for Europe (UNECE). 2013. Classification of Types of Big Data. http://www1. unece. org/stat/platform/display/bigdata/Classification+of+Types+of+Big+Data.

———. 2014. A Suggested Framework for the Quality of Big Data. http://www1. unece. org/stat/platform/download/attachments/108102944/Big%20Data%20Quality%20Framework%20-%20final-%20Jan08- 2015. pdf?version=1&modificationDate=1420725063663&api=v2.

United Nations Global Pulse (UNGP). 2012. "Big Data for Development: Challenges and Opportunities. "

———. 2014. Nowcasting Food Prices in Indonesia Using Social Media Signals (2014) http://www. unglobalpulse. org/nowcasting-food-prices.
United Nations Global Working Group (UNGWG). 2017. Big Data. http://unstats. un. org/bigdata.
———. Survey and Project Inventory, https://unstats. un. org/bigdata/inventory.

United Nations Statistics Division (UNSD). 2014. Fundamental Principles of Official Statistics. https://unstats. un. org/unsd/dnss/gp/fundprinciples. aspx.
———. 2015. Report of the 2015 Big Data Survey. https://unstats. un. org/unsd/statcom/47th-session/documents/BG-2016-6-Report-of-the-2015-Big-Data-Survey-E. pdf.

United Nations World Data Forum (UNWDF). 2017. A series of presentations by Statistics Norway, Statistics Denmark, Statistics Sweden and Statistics Finland. www. undataforum. org.
World Bank. Big Data in Action for Development. Washington, DC. http://live. worldbank. org/sites/default/files/Big%20Data%20for%20Development%20Report_final%20version. pdf.

———. 2016. Delivering on Big Data. Washington, DC.
.

Appendix I. UNECE 태스크 팀이 빅데이터에 대해 개발한 분류,

(UNECE Wiki 2013년 6월)[25]

1. Social Networks (human-sourced information): This information is the record of human experiences, previously recorded in books and works of art and later in photographs, audio, and video. Human-sourced information is now almost entirely digitized and stored everywhere from personal computers to social networks. Data are loosely structured and often ungoverned.

1100. Social Networks: Facebook, Twitter, Tumblr etc.
1200. Blogs and comments
1300. Personal documents
1400. Pictures: Instagram, Flickr, Picasa, etc.
1500. Videos: YouTube etc.
1600. Internet searches
1700. Mobile data content: text messages
1800. User-generated maps
1900. E-mail

2. Traditional Business Systems (process-mediated data): These processes record and monitor business events of interest, such as registering a customer, manufacturing a product, taking an order, etc. The process-mediated data thus collected are highly structured and include transactions, reference tables, and relationships, as well as the metadata that set its context. Traditional business data are the vast majority of what IT managed and processed, in both operational and business intelligence systems—usually structured and stored in relational database systems. (Some sources belonging to this class may fall into the category of "Administrative data. ")

21. Data Produced by Public Agencies
    2110. Medical records
22. Data Produced by Businesses
    2210. Commercial transactions
    2220. Banking/stock records
    2230. E-commerce

[25] UNECE 2013

2240. Credit cards

3. Internet of Things (machine-generated data): This information is derived from the phenomenal growth in the number of sensors and machines used to measure and record the events and situations in the physical world. The output of these sensors is machine-generated data, and from simple sensor records to complex computer logs, it is well structured. As sensors proliferate and data volumes grow, it is becoming an increasingly important component of the information stored and processed by many businesses. Its well-structured nature is suitable for computer processing, but its size and speed are beyond traditional approaches.

31. Data from Sensors

311. Fixed sensors

3111. Home automation

3112. Weather/pollution sensors

3113. Traffic sensors/webcam

3114. Scientific sensors

3115. Security/surveillance videos/images

312. Mobile sensors (tracking)

3121. Mobile phone location

3122. Cars

3123. Satellite images

32. Data from Computer Systems

3210. Logs

3220. Web logs

Appendix II. Table 1: 빅데이터 및 통계 영역 연결

Data Source+	Data Type	Statistical Domains	Additional Statistical Domains*
Social Networks	1100. Social Networks: Facebook, Twitter, LinkedIn 1200. Blogs and comments 1600. Internet searches on search engines (Google) 1700. Mobile data content: text messages, Call Detail Record, Data Detail Record, Location update, Radio coverage updates Online news	National accounts External sector statistics Financial statistics Price statistics Government finance statistics (public debt statistics)	Sentiment indices (investor, consumer) Social statistics Labor statistics Migration statistics Tourism statistics Population statistics Household consumption statistics Sustainable Development Goals indicators Early-warning indicators Transportation statistics Urban statistics
Traditional Business Systems	Data produced by public agencies Administrative data	Government finance statistics National accounts Price statistics External sector statistics	Sustainable Development Goals indicators
	Data produced by businesses 2210. Commercial transactions 2220. Banking/stock records 2230. E-commerce 2240. Credit cards Business websites Scanner data	National accounts Price statistics External sector statistics Financial statistics	Social statistics Business registers Employment statistics Household consumption statistics Transport statistics Sustainable Development Goals indicators
Internet of Things (machine-generated data)	Data from sensors **311.** Fixed sensors 3111. Home automation 3112. Weather/pollution sensors 3113. Traffic sensors/webcam 3114. Scientific sensors **312.** Mobile sensors (tracking) 3121. Mobile phone location 3122. Cars	National accounts Satellite accounts External sector statistics Government finance statistics Price statistics	Traffic/transport statistics Energy statistics Land use statistics Agricultural statistics Environment statistics Transport and emission statistics Air emission statistics Sustainable Development

	3123. Satellite images		Goals indicators
Based on adapted UN big data classification+			*Based on European Statistical System Committee (2014)

Appendix III. Table 2: 거시경제 및 금융통계에서의 빅데이터의 현재적 활용

Data Origin+	Data Type	Data Source and Techniques	Potential Indicators Derived	Statistical Domains	What May Be the Potential?*
Social Networks	Social networks, blogs and comments 1100. Social Networks: Facebook, Twitter, LinkedIn 1200. Blogs and comments 1600. Internet searches on search engines (Google) 1700. Mobile data content: text messages, Call Detail Record, Data Detail Record, Location update, Radio coverage updates Online news	Google trends and search data	nowcast GDP nowcast unemployment consumer sentiment car and property sales	National accounts External sector statistics Financial statistics price statistics	2
		Mobile phone system data (electronic money schemes, e. g. , M-Pesa) Peer-to-peer transactions	financial inclusion indicators remittances, regional disposable income, consumption patterns poverty reduction SDG "Gender Equality" economic growth	National accounts External sector statistics Financial statistics Price statistics	1,3
		Twitter tweets	consumer confidence index border mobility, tourism, transitioning of migrants nowcast food prices sentiment and topic trend analysis	Mobility and urban statistics Price statistics Demographic and social statistics	1,2,3
		Web scraping of Facebook posts, Wikipedia articles	geopolitical risk indicators price changes civil protests/labor strikes and national security events consumer sentiment inclusive infrastructure for sustainable development	Price statistics National accounts Demographic and social statistics Labor statistics	1,2

		Call Detail Record data	SDG indicators, travel/tourism, transport, migration	Mobility and urban statistics	1,3
Traditional Business Systems	Data produced by public agencies Administrative data	Taxation registers	consumer spending small business income nonresident businesses controlled by resident parent corporations business profiling flight reservation system	National accounts Price statistics External sector statistics Labor statistics Tourism statistics Transportation statistics	2, 3
	Data produced by businesses 2210. Commercial transactions	Population/business registers	multisourcing to derive population and housing census population structure	National accounts Demographic and social statistics	3
		SWIFT data on transaction quantities and financial market prices	global financial flows network concentration cross-border transactions export/import indicators	National accounts Price statistics External sector statistics Financial statistics	2,3

Data Origin+	Data Type	Data Source and Techniques	Potential Indicators Derived	Statistical Domains	What May Be the Potential?*
	2220. Banking/stock records 2230. E-commerce 2240. Credit cards Business websites Scanner data		withdrawal of correspondent banking relationships trade financing		
		Web scraping to collect price data from online retailers Web scraping business websites	daily inflation turning points in inflation trends e-commerce index enterprise profiling job vacancies	Price statistics Financial statistics National accounts Financial statistics Labor statistics	2,3 2,3
		Scanner data Prices and quantities	national and regional consumer prices household income and expenditure	Price statistics National accounts Financial statistics	2,3
		Credit card data	consumer spending growth trends of retail sales	National accounts External sector statistics	
Internet of Things (machine-generated data)	Data from sensors **311.** Fixed sensors 3111. Home automation 3112. Weather/pollution sensors 3113. Traffic sensors/webcam 3114. Scientific sensors	GPS positioning/tracking data	travel services exports/imports trip duration inbound/outbound international travelers remoteness index traffic intensity	National accounts External sector statistics Demographic statistics Transport statistics Urban statistics Tourism statistics Population	1,2,3

312. Mobile sensors (tracking) 3121. Mobile phone location 3122. Cars 3123. Satellite images			statistics	
	Traffic/road sensors	proxy of economic growth/health commuting time traffic intensity incoming/outgoing traffic travel/tourism	National accounts External sector statistics Transport statistics Tourism statistics Mobility statistics	1,2,3
	Satellite imagery Research and mapping of weather and climate data	improved geographical localization of statistical units and assets spatial sampling frame for output measurement land use and geostatistical cartography crop planting area, land use, and agricultural output population and asset location as proxy for SDG "Gender Equality"	National accounts Price statistics External sector statistics Demographic and social statistics Transport statistics Agricultural statistics Demographic and urban statistics	1,3
	Smart meters (energy consumption measures)	nonoccupancy rates household consumption electricity supply and consumption price differentials household structure and size	Environmental and energy statistics National accounts Price statistics Demographic and social statistics Transportation statistics Geospatial statistics Agricultural statistics Rural and population statistics	1,2,3
Based on adapted UN big data classification+	*1. Big data to answer "new questions" and produce new indicators 2. Big data to bridge time lags in the availability of official statistics and supporting the more timely forecasting of			

	existing indicators Big data as an innovative data source in the production of official statistics

Appendix IV. 빅데이터와 디지털 경제

빅데이터는 종종 디지털 경제의 연료로 묘사된다. 데이터는 인적, 재정적 자원에 버금가는 중요한 경제 자산이 되었다. 대부분의 경제활동은 몇 년 안에 데이터에 의존할 것이며, 이것은 많은 경제분야에 기회를 제공할 것이다. 디지털 경제는 인터넷 플랫폼과 스마트폰과 정보통신 기술 하드웨어와 소프트웨어와 같은 디지털 기술을 통해 제공되거나 중개되는 서비스를 포함한다. 전문 서비스 회사는 국경 내외에서 데이터 저장, 전송 및 광업 서비스를 제공하므로 거래 비용이 절감된다(OECD 2015).

디지털화는 가격과 그에 따른 GDP의 부피적 측정에 어려움을 제기한다. 디지털화는 품질 차이를 통제하는 가격 비교를 어렵게 만든다. 특히 인터넷과 모바일 서비스를 위한 가격 모델도 다양하다. 게다가, 디지털화는 새로운 비즈니스 모델이 만들어지는 것을 목격했는데, 이것은 제공되는 서비스의 질을 높이고 제품들 간의 전환을 초래할 수 있다. 이는 품질 대 가격 변화의 개념이 빠르고 구별하기 어려웠던 소프트웨어 및 정보통신 기술 제품에도 공통적이다.

빅데이터는 디지털 경제를 측정하는데 도움이 될 수 있다. 빅데이터는 디지털 거래에 대한 새로운 통찰력을 제공할 수 있다. 통계기관은 웹 스크랩을 통해 인터넷의 거래 및 디지털 제품에 관한 새로운 디지털 중개자(이베이 등)로부터 직접 정보를 수집할 수 있다. 이러한 방식으로 제공되는 새로운 데이터는 더 높은 빈도의 수집과 품질 및 가격 변경에 대한 더 나은 제어를 가능하게 하다.

이 두 가지 예는 빅데이터가 어떻게 디지털 분야를 더 잘 정량화할 수 있는지를 보여준다.

성장 인텔리전스(GI)데이터를 이용한 국립경제사회연구소(NIESR) 논문에 따르면 영국의 디지털 경제는 이전 추정치보다 상당히 컸다(Nathan and Rosso 2013). GI는 웹, 소셜 미디어, 뉴스 피드, 특허 및 기타 출처의 데이터를 사용하고 기업, 기업 하우스(Companies House)의 영국 등록부의 공개데이터 위에 이러한 데이터를 계층화했다.

또 다른 연구는 빅 데이터와 정규 통계를 결합하여 네덜란드 인터넷 경제의 규모를 연구했습니다 (Oostrom and others 2016). Statistics Netherlands, Google 및 Dataprovider는 복잡한 매칭 알고리즘을 통해 비즈니스 웹 사이트 데이터와 비즈니스 통계 정보를 연결하는 연구에 협력했습니다. 인터넷 경제의 핵심은 건설 부문과 거의 같은 규모라는 것을 발견했다.