

# CEO Information

2013. 4. 10. (제891호)

## 기업의 新경쟁력, 빅데이터 큐레이션

### 목차

I. 왜 빅데이터 큐레이션인가

II. 활용 분야와 4大 성공 수칙

III. 시사점

작성 : 박성민 수석연구원(3780-8518)  
sm320@seri.org

박성배, 채승병 수석연구원  
김영도, 김지환 수석연구원



삼성경제연구소

## 《 요약 》

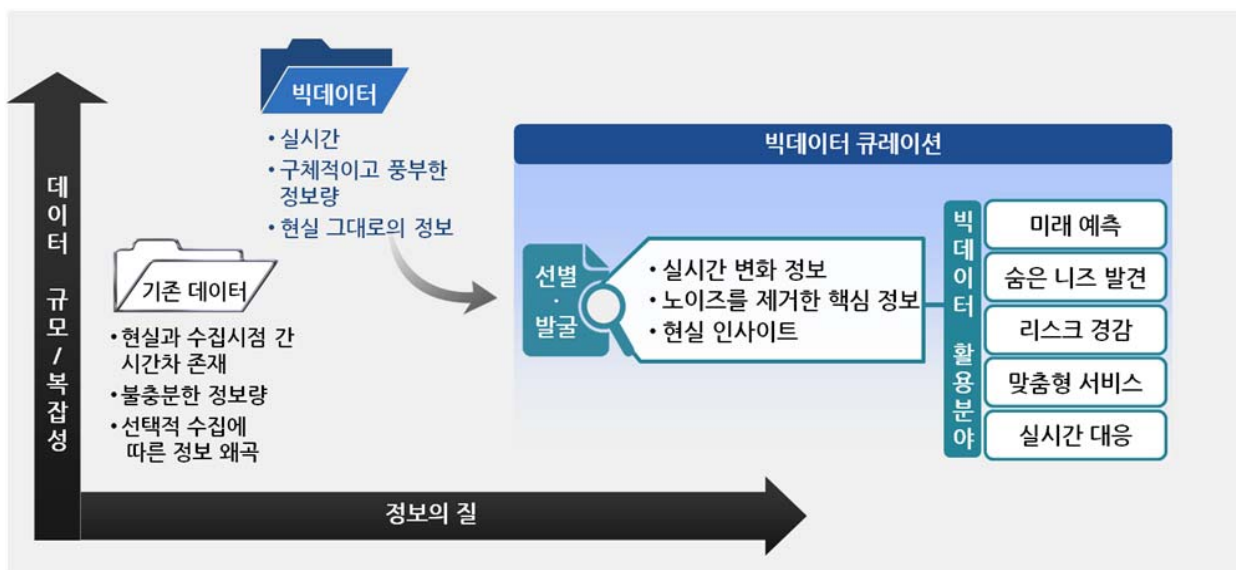
### 빅데이터 質的 활용의 필수 요건, '빅데이터 큐레이션'

전 세계적으로 빅데이터의 활용이 확대되는 가운데, 방대하고 복잡한 데이터에서 '질' 높은 정보를 선별해 발굴하는 것이 핵심역량으로 부상하고 있다. 아무리 현실을 잘 반영하는 빅데이터가 있더라도 경영에 도움이 되는 정보를 찾아내지 못한다면 빅데이터에 대한 하드웨어 투자는 비용에 불과하다. 따라서 빅데이터를 최적으로 구축하고 분석 및 활용까지 전 과정을 지휘하는 활동인 '빅데이터 큐레이션'을 활성화해야 한다.

### 빅데이터 큐레이션의 5大 활용 분야

빅데이터를 활용하여 기존 방식과는 차별화된 성과를 창출하는 분야로 다음 다섯 가지가 있다. ① [미래 예측] 실시간 예측 및 자동 업데이트로 예측의 정확도가 높아진다. 병원은 환자의 증세-질환-입원 패턴을 추출해 재입원 가능성을 예측한다. ② [숨은 니즈 발견] 소비자의 일상 생활 데이터로부터 새로운 패턴을 발견해 숨은 니즈를 발견한다. 자동차업체는 고객의 운전습관 데이터를 분석해 신차의 소비자 니즈를 파악한다. ③ [리스크 경감] 정성적 정보를 대폭 확보함으로써 관리 가능한 리스크의 범위와 정확도가 증대된다. 의료정보시스템은 웹데이터로부터 공중보건에 해로운 이벤트를 탐지해 위험을 경고한다. ④ [맞춤형 서비스] 고객 개인별로 차별화해 스팸이 아닌 유용한 정보를 제공한다. SNS업체는 관계정보 데이터를 분석해 개인 특성에 맞춰 친구를 추천한다. ⑤ [실시간 대응] 고객과 경영환경에 대한 실시간 모니터링과 즉각적인 대응이 가능해진다. 카드사는 거래정보를 실시간으로 분석해 부정거래를 감시하고 고객에게 즉시 고지하여 문제 확산을 최소화한다.

### 빅데이터 큐레이션



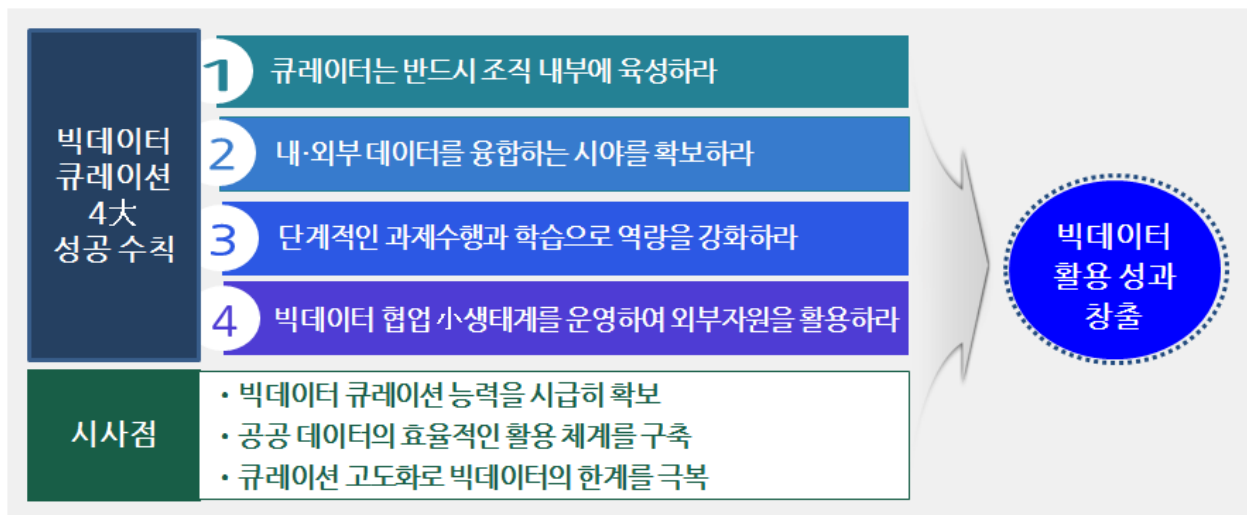
## 빅데이터 큐레이션의 4大 성공 수칙

빅데이터 활용 성과를 지속적으로 창출하기 위해서는 다음 네 가지 성공 수칙을 실천해야 한다. 첫째, 큐레이터는 반드시 조직 내부에 육성한다. 큐레이터는 필요할 때마다 즉각적으로 조달할 수 없는 핵심인재이다. 또한 빅데이터 보안사고가 발생하면 그 피해가 막대하므로 큐레이터를 내재화해야 한다. 둘째, 내·외부 데이터를 융합하는 시야를 확보한다. 내부 데이터는 외부 데이터와 결합될 때 진가를 발휘한다. 셋째, 큐레이션 관련 인력과 시스템 역량을 지속적으로 성장시키기 위해서는 단계적인 과제수행과 자기학습으로 역량을 강화해야 한다. 빅데이터 모범기업인 아마존은 CEO가 높은 기술 이해력과 솔루션에 대한 감각으로 직원들에게 끊임없이 큐레이션 영감을 불어넣고 있다. 넷째, 빅데이터 협업 小생태계를 운영하여 외부 자원을 활용한다. 빅데이터 관리와 분석에는 광범위한 배경과 전문지식을 가진 전문인력이 필요하므로 전부 내재화할 수는 없다.

## 큐레이터 육성과 올바른 빅데이터 활용에 주력할 필요

세계적인 박물관들이 큐레이터의 훌륭한 안목을 바탕으로 기획전을 열며 소장품과 명성을 쌓아가듯 기업도 큐레이터를 활용해 빅데이터 활용 역량을 축적해야 한다. 이를 위해서는 우선적으로 유능한 큐레이터를 육성하는 것이 급선무다. 정부도 개별 기관이 보유한 데이터 간 연계성을 강화하고 적극적으로 개방함으로써 공공 데이터의 활용성을 제고할 필요가 있다. 한편, 빅데이터가 만능은 아니므로 엄연히 한계와 부작용이 존재한다. 하지만 이는 고객의 감성을 이해하고 프라이버시를 보호하는 식의 섬세한 큐레이션으로 극복할 수 있다. 기업은 빅데이터 큐레이션 역량을 심화 발전시켜 '빅데이터 + 큐레이션 = 섬세하고 안전한 기술'로 인식될 수 있도록 적극적인 노력을 기울여야 한다.

## 빅데이터 활용의 성과 창출을 위한 방안



# I . 왜 빅데이터 큐레이션인가

빅데이터, 量에서 質로의 전환이 필요

□ 방대하고 복잡한 데이터에서 '질' 높은 정보를 선별해 발굴하는 것이 빅데이터 시대에 성공할 수 있는 핵심 역량

- 빅데이터는 기존의 데이터 수집, 저장, 분석 도구로는 처리하기 힘든 정형·비정형 데이터<sup>1)</sup>로서 최근 혁신의 新자원으로 부상

- 기존 데이터에 비해 구체적이고 풍부한 정보를 신속하게 제공, 왜곡 없는 현실 반영 등이 장점

- 전 세계 IT 및 현업 담당자 600여 명에게 설문조사한 결과, 67%가 빅데이터 활용이 기업 경쟁력 제고의 기회라고 응답<sup>2)</sup>

- 빅데이터에서 노이즈를 제거하고 필요한 정보를 만들어낼 때 의미 있는 경영 성과를 창출

- 현실 이해를 왜곡하는 콘텐츠를 제거하거나 문맥과 뉘앙스를 파악하는 등 비정형 데이터 처리 기술을 적용하는 기업이 증가

## 빅데이터의 대표적 혁신 사례



<sup>1)</sup> 통상 100TB 이상의 규모(volume), 실시간에 가까운 생성 및 갱신 속도(velocity), 비정형 등의 형태 다양성(variety), 분석의 복잡성(complexity) 등의 특성을 지님

<sup>2)</sup> Informatica (2012). Balancing Opportunity and Risk in Big Data: A Survey of Enterprise Priorities and Strategies for Harnessing Big Data(White Paper).

"가용 데이터의 양과 종류가 급증하는 상황에서 데이터를 수집하고 분석하는 수준으로는 충분하지 않으며 의사결정에 직접 활용할 수 있는 정보를 적시에 제공해야 한다."<sup>3)</sup>

□ 여전히 많은 기업들은 빅데이터를 어떻게 활용할 것인가를 고심 중

- 기업 대상 설문조사 결과, 빅데이터 활용 파일럿 프로젝트 및 구축 사업을 시행(28%)하는 경우보다 활용 로드맵을 작성(47%)하거나 빅데이터에 대한 개념 이해에 집중(24%)한다는 응답이 다수<sup>4)</sup>
- 빅데이터 확보 및 관리, 관련 인프라 구축에 집중하면서 실효성 있는 분석과 비즈니스 인사이트를 도출하지 못하는 경우도 발생

"어떤 문제를 해결할지 고민도 없이 시스템을 구축하고 데이터를 수집해놓으려는 경우도 빈번한데, 이런 경우 대부분 난항을 겪거나 실패" (SAS Korea 이진권 본부장 인터뷰)

### 빅데이터 질적 활용의 필수 활동, '빅데이터 큐레이션'

□ '빅데이터 큐레이션'이란 빅데이터 전략을 제시하고 최적의 빅데이터 구축에서 분석 및 결과 활용까지 전 과정을 지휘하는 활동

※ '큐레이션(curation)'은 기획 의도에 맞게 작품을 선정, 수집, 연구, 구성, 전시, 해설하는 활동을 총칭<sup>5)</sup>

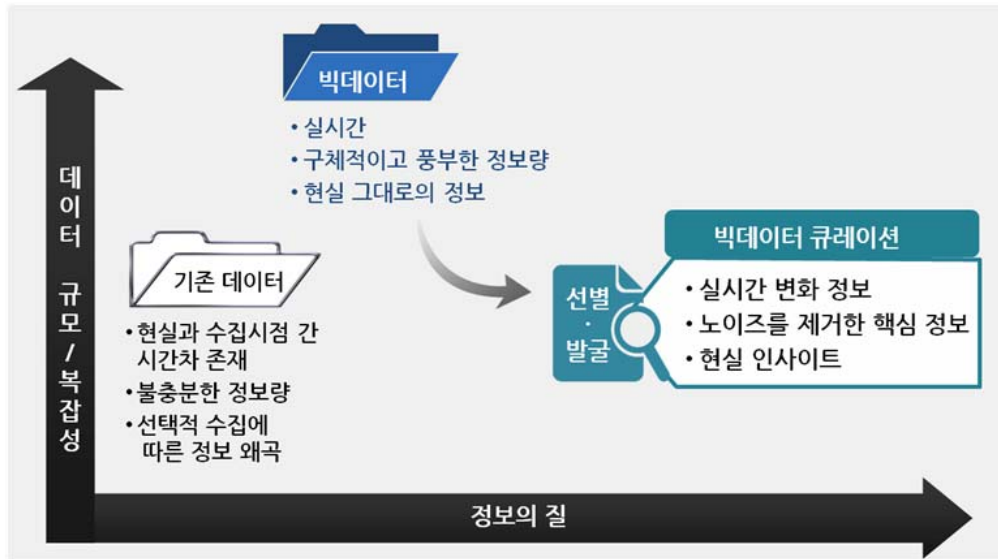
- 아무리 현실을 잘 반영하는 빅데이터가 있더라도 경영에 도움이 되는 정보를 찾아내지 못한다면 빅데이터 투자는 낭비에 불과

<sup>3)</sup> Blechar, M., et al. (2012). Information Infrastructure and Big Data.(Predicts 2012). Gartner.

<sup>4)</sup> 95개국 26개 산업분야 전문가 1,144명(경영진, 현업 부서 담당자, IT 담당자 등) 대상 설문 결과 (IBM Institute for Business Value (2012. 10) Analytics: The real world use of big data.)

<sup>5)</sup> Oxford Dictionaries <<http://oxforddictionaries.com/definition/english/curate--2>>; Blechar, M., et al. (2012). Information Infrastructure and Big Data(Predicts 2012). Gartner.

빅데이터의 질적 활용을 가능케 하는 빅데이터 큐레이션



빅데이터 큐레이션의 개념

- ▷ 미술관이나 박물관에서의 큐레이션은 '작품에 생기(生氣)를 부여'하는 활동
  - 돌보다, 보살피다(take care)라는 뜻의 라틴어 'curare'가 어원
  - 기존에 알려진 작품에 새로운 의미를 부여하고 우수한 작품을 발굴하여 여러 작품들로 하나의 주제를 형상화
  - 주제를 잘 전달하기 위한 작품 배치, 조명 조정, 해설 활동이 모두 큐레이션에 해당
- ▷ 빅데이터 큐레이션은 '데이터의 숨은 가치와 잠재력 발굴'을 추구하는 활동
  - 일반적인 큐레이션의 수혜 대상이 관람객이라면, 빅데이터 큐레이션의 수혜자는 기업의 각 부서와 최고 의사결정자
  - 새로운 빅데이터를 창출하거나 전략, 제품, 서비스 변화의 기폭제로도 기능

(자료: Creative City Network of Canada. et al. (2010). Curatorial Toolkit: A Practical Guide for Curators.)

□ 빅데이터 시대에는 빅데이터 큐레이션이 필수

- 빅데이터를 활용하면 어떤 현안을 해결할 수 있는지, 이를 위해 어떻게 큐레이션을 해야 하는지 살펴보고 시사점을 도출

## II. 활용 분야와 4大 성공 수칙

- 본 장에서는 기업에서 주목할 만한 빅데이터 큐레이션의 활용 분야를 소개한 후, 성공적인 활용 방안을 제시

### 1. 빅데이터 큐레이션 활용 분야

- 빅데이터로 차별화된 성과를 창출할 수 있는 대표적인 5大 분야:

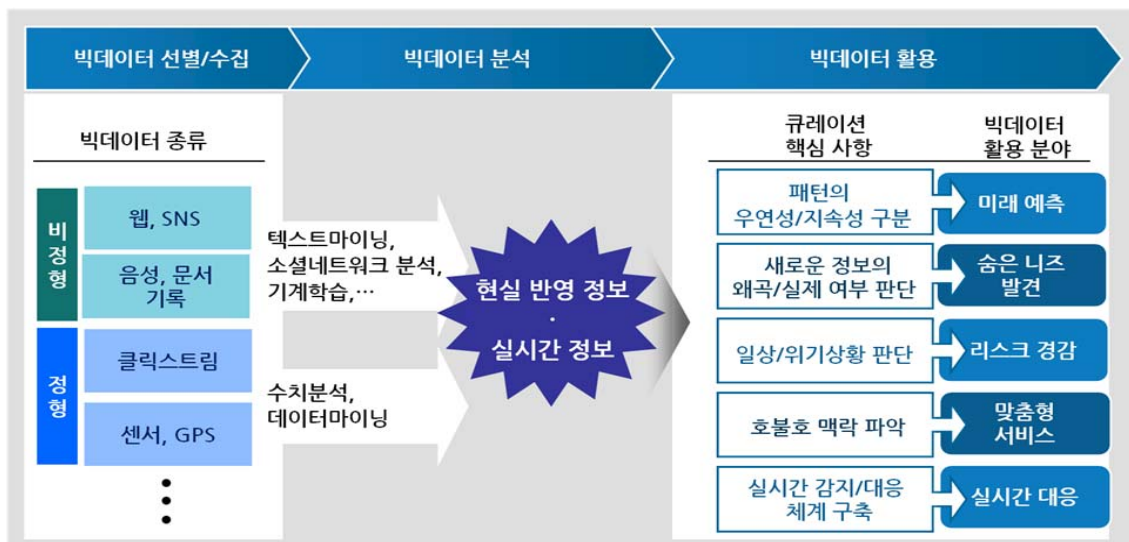
- ① 미래 예측, ② 숨은 니즈 발견, ③ 리스크 경감, ④ 맞춤형 서비스, ⑤ 실시간 대응

- 빅데이터에는 시장, 고객, 거래 관련 기존의 수치데이터에 비해 현실을 보다 구체적이고 신속하게 반영한 정보가 풍부

- 웹, SNS, VOC(Voice of Customers: 고객의 소리), 진료차트 등 비정형 데이터와 클릭스트림<sup>6)</sup>, 센서 등 정형 데이터를 모두 포괄

- 다수의 샘플로부터 왜곡이 적으며 빠르게 갱신된 분석결과를 확보할 수 있는 것이 빅데이터 활용의 최대 장점

#### 빅데이터의 대표적 활용 분야



<sup>6)</sup> 웹사이트에서 사용자가 마우스로 '클릭(click)'한 행위 정보로, 클릭을 통해 이동하는 웹페이지의 경로나 횟수 등이 포함

## ① 미래 예측: 실시간 예측 및 자동 업데이트

- 장기간 축적된 데이터로부터 과거와 현재의 규칙성과 상관관계를 밝히고, 이를 토대로 특정 이벤트의 발생 여부나 수요·판매량을 예측
  - 빅데이터 적용으로 날씨, 교통, SNS, 고객 특성 및 구매이력, 센서 정보 등을 별도의 가공 없이 인과관계 파악이나 패턴 분석에 활용
    - 과거 수리적 통계 모형을 이용한 예측방식에 비해 변수 선택의 폭이 확대되고 최신 정보를 시차 없이 반영할 수 있는 것이 장점
- 빅데이터에서 발견한 인과관계가 일시적 또는 우연에 의한 것인지 반복적으로 지속될 패턴인지 구분하는 것이 큐레이션의 핵심 역량
  - 미국 세튼 헬스케어 패밀리 병원: 200만 명 환자정보와 진료기록 중 입원 관련 데이터를 추출 → '중세-질환-입원' 연관패턴을 발견 → 재입원 가능성이 높은 환자를 예측 → 예방조치·병상 운영에 활용
  - 패스트 패션 기업 자라: 전 세계 매장의 환경 정보, 품목별 특징과 전시위치, 판매실적 등을 실시간 집계 → 의류 품목별 최소 및 최대 판매량을 도출 → 매장·품목별 적정재고 및 주문 가이드로 활용

### 英 에너지 기업 센트리카, 스마트계량기로 전력수요 예측

- ▷ 센트리카는 30분마다 고객의 스마트계량기로부터 받는 검침 데이터와 날씨, 기온, 습도 등 데이터의 패턴을 분석함으로써 전력수요를 예측
    - 1일 48회 발생하는 검침 데이터로 고객별 에너지 소비 패턴을 파악하고, 이를 그룹화하여 그룹별 행동특성을 도출 → 미래 전력소비 예측에 활용
  - ※ 센트리카는 영국 최대 에너지 공급회사로 성장(영국 68% 가정에 전력 공급)
- (자료: 시로타 마코토 (2013). 『빅데이터의 충격』 (김성재 역). 한빛미디어.)



## ② 숨은 니즈 발견: 무의식적으로 표출되는 고객 니즈까지 파악

□ 소비자의 일상이 담긴 데이터에서 발견한 새로운 패턴으로 경쟁사나 고객 스스로가 인지하지 못하는 고객 니즈를 발견

- 웹, SNS, 센서 및 구매 데이터 등의 빅데이터로부터 소비자의 실제 감성과 선호를 추출하여 비즈니스에 가치 있는 패턴을 발견

- 기존 시장설문조사에서 종종 왜곡됐던 소비자 니즈를 파악하여 신제품 개발, 브랜드 포지셔닝, 고객 발굴에 활용

### 美 호텔 체인 하라스(Harrah's), 핵심고객 발견

▷ 하라스 그룹은 카지노 주고객층이 관광객이 아닌 평범한 지역 주민임을 인지, 가족오락공간으로 브랜딩하고 주민 맞춤 마케팅으로 업계 만족도 1위 달성

- 2004년부터 매년 3,000만 명에 달하는 고객의 행동패턴(참여한 게임종류, 베팅액, 손실액, 방문빈도, 거주지까지의 거리, 게임시간 등)을 분석

→ 60개 유형으로 고객을 분류 → 핵심고객 파악 및 유형별 마케팅 도입

- 가족 이벤트, 주변 레스토랑 및 기업과 연계한 마케팅으로 가족고객이 자연스럽게 카지노를 방문하도록 유도

※ 하라스는 업계 3위였지만 2005년 1위인 시저스를 인수하는 대역전극을 연출

(자료: 김지유 (2011). "괜히 카지노 1위가 아니다, 하라스의 분석경영". 『IGM비즈니스 리뷰』, 14.)

□ 빅데이터에서 발견한 패턴이 데이터 오류나 분석 착오가 아닌지 관련 전문가가 해석·검증하여 제품 및 서비스에 활용할 방안을 마련

- 미국 자동차업체 포드: 400만 대의 차량에 설치된 센서로 운전습관, 주행환경, 소음 등의 데이터를 수집 → 상황별 차량 이용, 편의장치 사용 등 운전자 행태를 분석 → 신차에 반영할 소비자 니즈 파악

### ③ 리스크 경감: 다양한 리스크를 더욱 정교하게 관리 가능

□ 정성적 정보의 양과 종류가 과거보다 대폭 증가하여 이상징후 감지, 고위험 이벤트 발생 경고 등 포괄적인 리스크 관리가 가능

- 관계, 정서 등 정성적인 정보를 기존 데이터와 함께 분석하면서 리스크 관리의 폭과 깊이가 확대

· SNS 관계정보를 활용한 위험집단 식별, 경제심리 분석 등으로 부정·불법 모니터링, 위기 감시 및 경고

□ 상황과 이슈별로 분류한 고객불만을 분석하여 트렌드 변화 및 특정 이슈를 관찰하고, 불만의 우선순위를 정해 근원적 불만요소를 식별

- VOC 업무가 수작업과 직감에 의존하던 방식에서 벗어나 자동적이고 체계적으로 분석해 해결책을 도출하는 방식으로 진화

· 노동집약적인 VOC 관리를 분석업무 중심으로 전환

- 자동차 렌탈업체 허츠(Hertz): 세계 8,000여 지점의 이메일, 콜센터 및 온라인 VOC에서 고객불만 키워드를 자동으로 추출 → '지점-서비스-고객불만' 연관성 파악 → 전사적 서비스 개선에 활용

#### 美 레스토랑 P.F.Chang's의 고객불만 파악

▷ P.F.Chang's은 SNS 데이터를 분석하여 일부 매장의 핵심적인 고객불만사항을 파악하고 전사적 서비스 개선에 활용

- 2,000만 개의 SNS 콘텐츠로부터 매장별 이슈 추출 → 일부 매장에서 전화 주문처리와 접객을 직원 한 명이 담당하여 고객만족도가 떨어졌음을 발견

- 주문전화는 콜센터가 처리하고 매장직원은 접객서비스에 전념하도록 개선

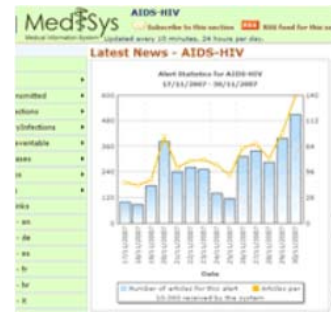
(자료: P.F.Chang's adds Social Media to its Menu. (2012.11.9.). *Wall Street Journal*.)

- 정보가치가 낮은 데이터를 필터링하는 알고리즘을 개발하여 일상적인 불안요소와 긴급하고 영향력이 큰 리스크를 구별·해석하는 것이 중요
  - 정성적 데이터를 수량화하는 방식에 따라 위험의 패턴과 강도가 다르므로 반복적 실험에서 검증된 규칙 사용이 필수
  - 미국 국세청: 납세정보 DB와 SNS에서 납세정보, 탈세이력, 범죄자의 인적관계를 추출 → 사기패턴과 범죄그룹을 파악하여 이상징후 감시 → 불필요한 세금 환급 및 탈세 건수 감소

**유럽연합 의료정보시스템 메디시스(Medisys), 질병 발생 경고**

▷ 메디시스는 웹데이터에서 공중보건에 영향을 미치는 사건을 탐지하여 위험을 경고

- 의학전문사이트 400개, 뉴스포털 3,750개 등에서 수집한 뉴스를 수백 개 그룹으로 분류하여 지속 추적
- 주기적인 사건 등 정보가치가 낮은 데이터를 필터링하는 알고리즘으로 중요한 이벤트만 포착



(자료: Europe Media Monitor <<http://emm.newsbrief.eu/overview.html>>)

**④ 맞춤형 서비스: 스팸이 아닌 유용한 정보를 제공**

- 개별 상황에 대한 정확한 이해를 바탕으로 가장 적합한 방식과 내용으로 효과적인 메시지를 전달함으로써 서비스의 효용을 극대화
  - 다양한 경로를 통해 축적되는 고객 정보로 과거의 연령, 성별 등 인구통계학적인 분류 외에 맥락 기반의 분류가 가능
    - GPS 연동 앱(App)이나 위치기반서비스에 남긴 위치정보, SNS에 올린 감정을 표현한 콘텐츠 등 단편적이지만 구체적인 정보들을 종합하여 상황에 따른 선호 경향을 파악

- 맞춤형 추천과 광고는 소비자를 스팸으로부터 해방시켜주고, 기업의 마케팅 효율을 증가시키는 등 사회적 비용을 경감

□ 상황 및 감정 관련 데이터를 종합적으로 분석하여 호불호를 느끼는 맥락을 짚어내고 획일적이지 않은 콘텐츠 콘셉트를 도출하는 게 중요

- 고객의 기존 관심 분야에서 뻗은 정보를 제공하거나 새로운 분야에서 어울리지 않는 추천을 피하기 위해 소셜네트워크의 관계 정보와 지인들의 선호 경향을 반영
- 아마존, 넷플릭스: 개개인의 정보 외에 지인이나 비슷한 성향을 가진 사람들의 정보를 통합하는 사용자 기반 협업 필터링(collaborative filtering) 기술로 추천의 정확도를 제고

#### 빅데이터를 활용한 이성(異性) 추천, 태그드닷컴

▷ 소셜네트워크서비스 태그드닷컴(Tagged.com)은 관계정보 데이터를 분석하여 실시간 개인맞춤형 네트워크 데이팅 서비스를 제공

- 사용자 1억 명에 대한 빅데이터로 사람 간의 관계를 예측하여 연결 가능성이 높은 사람들을 선별
- 선별된 사람들의 친구 및 대화상대 관계를 파악해 개인맞춤화하여 데이트 성공 가능성을 향상



<태그드닷컴 홈페이지>

- 기존의 남녀만남 주선은 주선자의 개인적 친분 혹은 나이, 직업, 재력, 학벌 등 만남 대상의 프로필 매칭에 주로 의존

(자료: 함유근, 채승병 (2012). 『빅데이터, 경영을 바꾸다』. 삼성경제연구소.)

#### ⑤ 실시간 대응: 경영환경을 실시간 모니터링하고 즉각적으로 대응

□ 시장 상황 및 경쟁 동향을 실시간 파악하여 환경 변화에 신속하고 자동적으로 대응하는 체계를 구축함으로써 업무 성과를 제고

- 실시간 확보 가능한 시장과 소비자 정보로 현 시점에서 유효수요를 정확히 파악하여 효율적인 타겟 마케팅을 진행
- 과거에는 경쟁환경의 이슈 대응이 '사후 뒷수습' 방식이었지만 빅데이터를 활용하면 실시간 감지 및 맞대응이 가능

□ 빅데이터 큐레이션의 핵심은 실시간으로 경영상 기회와 위기요인을 정확히 읽어내고 즉시 대응할 수 있는 체계를 구축하는 것

- 일본 통신사 NTT 도코모: GPS 위치정보를 실시간으로 분석 → 고객이 스키장이나 골프장에 도착할 시점에 맞춤형 보험상품을 안내함으로써 순간의 사업기회를 포착



<NTT 도코모 홈페이지>

- 글로벌 카드사 VISA: 거래정보 실시간 분석시스템으로 부정거래를 감시 → 고객에게 즉시 고지하여 문제 해결 비용을 최소화

**美 유통업체 시어스, 경쟁환경에 대한 실시간 대응체제 구축**

- ▷ 시어스는 실시간으로 점포별 재고, 경쟁업체 가격, 날씨예보, 고객정보 등 다양한 빅데이터를 분석하여 최적의 가격을 책정하고 재고를 관리
  - 과거에는 8주 걸리던 전체 품목의 가격탄력성 분석이 10배 많은 데이터 분석에도 불구하고 1주로 단축되었고, 일부 품목은 실시간으로 가격을 책정
  - 실시간 주변상황과 고객의 취향을 동시에 고려한 개인별 특화 쿠폰을 적시에 발급하여 매출 증가와 적정 재고관리를 동시에 달성
- ▷ "우리 회사는 실시간 대응력을 갖춘 디지털 기업으로 변신하고 있다"  
(필 셸리, 시어스 홀딩스 CEO)

(자료: How Sears Uses Big Data to Get a Handle on Pricing. (2012.6.14.). Wall Street Journal.)

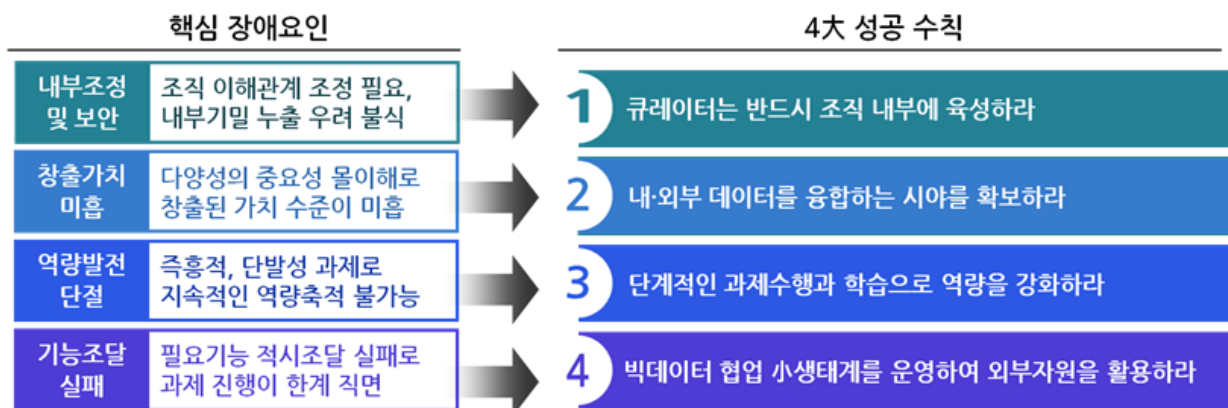
## 2. 빅데이터 큐레이션의 4大 성공 수칙

□ 빅데이터 큐레이션이 실질적인 성과를 이끌어내기 위해서는 주요 장애요인을 보완하는 네 가지 수칙을 준수

- ① 내부조정 및 보안 이슈: 빅데이터 분석 결과 실행 시 조직의 참여한 이해관계와 기밀누출에 대한 우려가 표출되는 경우도 존재  
→ 기업 현황을 정확히 이해하는 내부 인사가 담당
- ② 창출가치 미흡 이슈: 기존 업무에 활용하던 한정된 종류의 데이터만 분석하면 기대 이하의 결과만을 산출  
→ 내·외부 데이터를 융합하는 폭넓은 시야의 아이디어맨을 배치
- ③ 역량발전 단절 이슈: 기술에 대한 깊은 이해와 장기적 비전이 없는 즉흥적, 단발성 과제는 아무리 반복해도 빅데이터 역량이 담보  
→ 단계적인 과제수행과 학습으로 발전해가는 스마트 혁신을 달성
- ④ 기능조달 실패 이슈: 빅데이터의 요소기능은 조직 내부는 물론 외부에서도 매우 부족하므로 필요할 때 적시에 조달하기가 곤란  
→ 사내·외 인재들로 구성된 빅데이터 협업 소(小)생태계 운영

□ 네 가지 수칙을 이끌어가기 위해서는 빅데이터 큐레이션의 책임자인 큐레이터의 역할이 특히 중요

빅데이터 큐레이션의 4大 성공 수칙

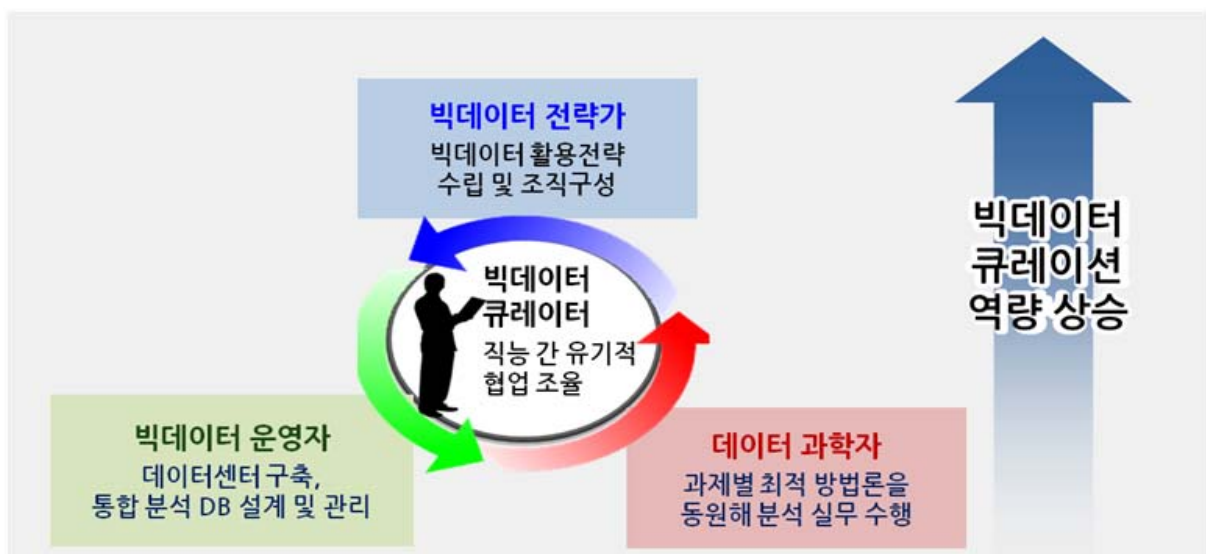


## ① 큐레이터는 반드시 조직 내부에 육성하라

□ 빅데이터 큐레이터는 외부에서 그때그때 조달할 수 없는 핵심 인재

- 빅데이터 큐레이터는 기업의 전략적 지향점과 각 사업부문의 특성을 명확히 이해하고 빅데이터 접목 포인트를 짚어내는 감각이 필요
  - 외부인력은 기업의 현실과 괴리된 결과를 내놓거나 내부조직 간 이해관계 조정에 실패해 실행으로 옮기지 못하는 경우가 빈번
- 특히 빅데이터는 자사의 민감한 경영정보나 고객정보 관리와 직결된 문제이므로 내부 인사조직이 리스크를 통제하는 것이 필수
  - 과도한 IT 아웃소싱이 최근 보안사고 증가의 한 원인이라는 지적  
→ 빅데이터 보안사고가 발생하면 피해가 매우 클 것으로 예상되므로 필히 내재화
- 큐레이터로서의 자질을 갖춘 핵심인력을 우선 내부에서 육성하고, 별도의 큐레이션 조직으로 확장할지 여부는 선별적으로 판단

빅데이터 큐레이터가 조율해야 할 관련 주요 전문직능



## ② 내·외부 데이터를 융합하는 시야를 확보하라

□ 내부 데이터는 의외의 외부 데이터와 결합될 때 진가를 발휘

"데이터의 다양성이 3V(Volume, Velocity, Variety) 가운데 가장 중요하다. 가장 강력하고 영향력 있는 사례는 다양한 종류의 데이터를 결합한 사례들일 것이다... 그러나 많은 조직들이 아직도 이 부분에 취약하다."<sup>7)</sup> (테드 프리드먼, 가트너 애널리스트)

□ 빅데이터 큐레이터는 여러 곳에 있는 데이터 자원을 파악하고 이를 창조적으로 조합하는 방법을 끊임없이 고민

- 경쟁사들이 생각하지 못한 데이터의 활용처를 발견하기 위해서는 다양한 데이터에 대한 관심이 필요
- 선형적인 편견에 사로잡히지 않고 끊임없이 새로운 외부 데이터를 수집하여 상식을 검증하는 자세를 견지

### 기업 내·외부 데이터 결합: 검색과 SNS부터

- ▷ 구글 트렌드 서비스 등을 통해 제공되는 검색통계 데이터와 페이스북, 트위터 등의 SNS 데이터는 가장 접근성이 높은 외부 데이터
- ▷ 기업은 물론 각국의 중앙은행 및 공공기관들도 이들 데이터를 보완적으로 활용
  - 이스라엘 중앙은행은 국민의 각종 소비재 검색통계 빈도로부터 경기수준을 추산하는 보조지표를 개발하여 정책금리 결정에 활용
  - 영국 중앙은행은 '실업수당'과 같은 실업관련 검색통계 빈도로부터 실업률을 빠르게 예측하여 대응

(자료: Ito, A. & Odenheimer, A. (2012.8.15.). Your 119 Billion Google Searches Now a Central Bank Tool. *Bloomberg*.)

<sup>7)</sup> McKenna, B. (2013.2.1.). Don't make big data stand alone, warns Gartner. Computer weekly.com <<http://www.computerweekly.com/news/2240177374/Dont-make-big-data-stand-alone-warns-Gartner>>



### ③ 단계적인 과제수행과 학습으로 역량을 강화하라

- 큐레이터는 기술 학습에 앞장서고 단계별로 실현 가능한 분석과제를 설계하여 역량 강화 경로를 제시하는 스마트 혁신가의 능력이 필요
  - 장기적인 빅데이터 역량 확대 비전을 구상하고 점차 수준 높은 과제를 통해 직원들의 지적 사기(士氣)를 고양
    - 당장 손에 잡히는 제한된 데이터로 단발성 과제만을 반복하면 역량이 축적되지 않고 분석 실무자들의 의욕도 급격히 저하
  - 큐레이터 스스로가 기술에 대한 열의를 갖고 끊임없이 학습하여 항상 경영현안과 함께 살아 숨쉬고 진화하는 빅데이터를 유지
    - 큐레이터와 분석 실무자 간 소통이 단절되고 아이디어 장벽이 생기는 순간부터 빅데이터는 도태

#### 빅데이터 큐레이션의 모범사례, 아마존과 CEO 제프 베조스(Jeff Bezos)

▷ 아마존을 세계 최고의 온라인 유통회사로 키운 CEO 제프 베조스는 빅데이터의 중요성을 인식하고 회사를 빅데이터 큐레이션의 모범기업으로 만듦

- 직접 빅데이터 분석업무를 하지는 않았으나 누구보다 높은 기술 이해력과 솔루션에 대한 감각으로 직원들에게 끊임없이 영감을 부여



<아마존 홈페이지>

▷ "다른 곳에서는 불가능한 것을 온라인에서 할 수 있어야 한다"는 모토를 갖고 업무에서 파생되는 빅데이터를 차별화된 서비스 창출의 지렛대로 활용

- 웹사이트에서 고객활동 데이터를 확보 → 고객의 쇼핑 편의성을 높일 서비스 아이디어 도출 → 효과 검증을 위한 데이터 추가 확보 장치를 마련하도록 반복적으로 독려하면서 아마존의 빅데이터 역량을 고양
- 최고 수준의 관심제품 추천 서비스, 원클릭 구매환경 등이 큐레이션의 산물

(자료: Brandt, R. L. (2011). *One Click*. Portfolio.)

#### ④ 빅데이터 협업 소(小)생태계를 운영하여 외부 자원을 활용하라

□ 빅데이터 큐레이션 역량을 지속 성장시키려면 사내·외의 유능한 주체들로 구성된 작은 생태계를 조성

- 빅데이터 관리와 분석에는 광범위한 배경과 전문지식을 가진 전문인력이 필요하므로 전부 내재화하는 것은 거의 불가능
- 큐레이터는 각 부문에서 탁월한 역량을 지닌 파트너들을 미리 파악하고 과제의 필요에 따라 협조를 받을 수 있도록 관리
- 서로가 '윈-윈'할 수 있는 호혜적인 협업 과제를 계속 창출하고 역량 성장을 주도적으로 이끌어주는 생태계 시각에서 접근

· IT 업체 시스코의 조사에 따르면, 기업은 데이터는 많지만 예산, 학습시간, 솔루션, IT 인력, 전문성 등 내부 자원이 부족<sup>8)</sup>

#### 美 오바마 대통령 재선의 핵심 원동력, 빅데이터 생태계 리더십

- ▷ 2012년 미국 대선에서 오바마 대통령이 재선에 성공한 데는 캠프가 광범위한 선거 관련 빅데이터 분석으로 '디지털 선거전' 전략을 펼친 것이 주요
  - 'The Cave'라고 불리는 50명의 빅데이터 분석팀은 미세 맞춤형 유권자 공략으로 효과적인 모금운동을 벌였으며, 경합주의 여론동향을 정확히 파악하고 결과를 예측하는 시스템을 구축해 대선 승리에 크게 공헌
- ▷ 오바마 캠프가 2008년 초선 당시부터 관리해온 유능한 인재 및 전문업체로 구성된 협력 생태계가 최고의 분석팀을 만드는 데 결정적인 역할
  - 빅데이터 선거전에 대한 관심을 환기시키며 실리콘밸리 및 주요 연구기관의 인재 풀을 관리 → 많은 인재들이 낮은 보수에 개의치 않고 캠프에 참여
  - 액센추어 출신의 마케팅전문가 레이드 가니, 입자물리학자 미켈란젤로 다고스티노, 픽사 출신의 SW엔지니어 캐빈 데어 등으로 드림팀 구성에 성공

(자료: EngageDC.com (2013). Inside the Cave. <<http://www.engagedc.com/inside-the-cave/>>)

<sup>8)</sup> Cisco (2013). Big Data: Big Potential, Big Priority(Cisco Connected World Technology Report).

### Ⅲ. 시사점

#### 기업은 빅데이터 큐레이션 능력을 시급히 확보

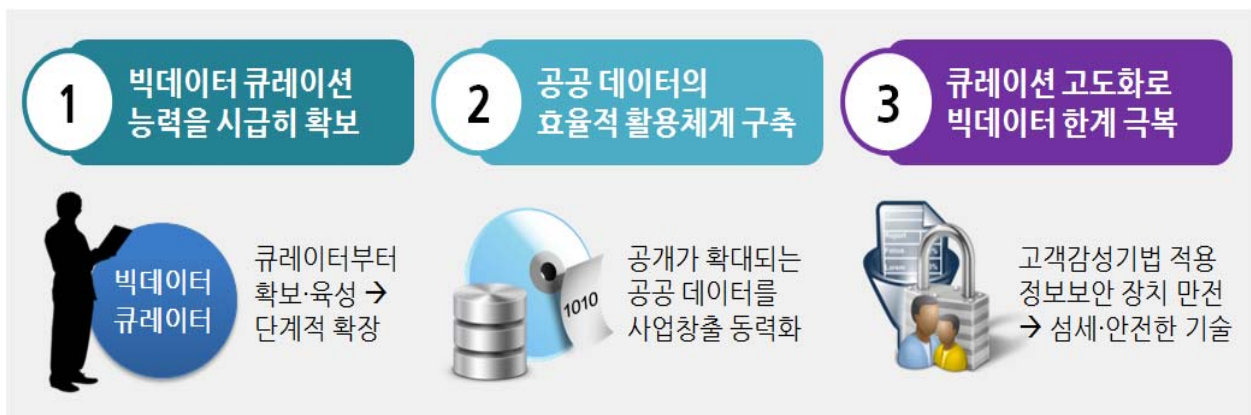
□ '빅' 혁신을 위해서는 '빅' 데이터가 필요

- 창조적 혁신이 강조되는 현재 시점에서 다른 경쟁자들의 사업모델이나 제품스펙에 연연하는 자세로는 더 이상 경쟁이 불가능

"벤치마킹으로는 근본적으로 타사와 다른 제품을 만들 수 없다"  
(제임스 다이슨, 영국 다이슨 설립자)

- 처음부터 빅데이터 운영·분석인력을 모두 모아 완벽한 팀을 만들기 보다는 유능한 빅데이터 큐레이터부터 확보·육성하는 것이 순서
- 세계적인 박물관들이 큐레이터의 훌륭한 안목을 바탕으로 기획전을 반복하며 차근차근 소장품과 명성을 쌓아가듯 기업도 빅데이터 큐레이터를 활용해 빅데이터 역량을 축적
  - 빅데이터를 이용한 가치창출은 데이터 관리-분석 이슈뿐 아니라 경영진과의 소통과 내부 이해관계 조정이 수반된 고난도 과제

#### 빅데이터 활용 성과를 창출하기 위한 제언



## 공공 데이터의 효율적인 활용체계를 구축

- 공공기관의 업무처리 과정에서 생성되는 공공 데이터는 해당 국가의 자연환경 및 인프라, 국민 경제활동 및 일상생활 등을 포괄하고 있어 규모가 크고 활용 가치가 높음
  - 빅데이터 분석기법 적용으로 예산 집행의 효율성 증진, 정확한 복지 대상자 선별, 탈세 방지 등 공공부문에 직접적 효과가 기대
    - 공공부문 빅데이터 잠재효과(McKinsey Global Institute):  
EU 전체 1,500억~3,000억유로, 영국 160억~330억파운드<sup>9)</sup>
  - 2010년 이후 미국, 영국 등 선진국 정부가 개방한 공공 데이터를 개인과 기업이 가공 및 분석하여 새로운 비즈니스 창출에 기여

### 美 기후보험 업체 클라이밋 코퍼레이션, 공공 데이터 기반으로 성장

- ▷ 클라이밋 코퍼레이션은 공공 기후데이터를 이용해 이상 기후 발생 시 해당 농가에 보험금을 지급하는 기후보험(Total Weather Insurance)을 개발·판매
  - 미국정부가 개방한 데이터(과거 60년간 지역별 작물수확량, 2제곱마일 단위로 제공되는 토양성분, 미국 전역 100만 개 지점에 설치된 도플러 레이더로 스캔한 기후 정보 등)를 기반으로 기후 및 곡물 수확량을 예측
  - ※ 클라이밋 코퍼레이션은 구글 출신인 데이빗 프리드버그와 시라지 칼리크가 2006년에 창업한 회사

(자료: 시로타 마코토 (2013). 『빅데이터의 충격』 (김성재 역). 한빛미디어.)

- 개별 기관이 보유한 데이터 간 연계를 강화하고 적극 개방함으로써 공공 데이터의 활용성을 제고할 필요

<sup>9)</sup> Manyika, J., et al. (2011). *Big data: the next frontier for innovation, competition, and productivity*. McKinsey Global Institute.

- 행정서비스의 고도화 추세로 서로 다른 기관과 데이터를 공유해야 하는 사안이 증가하나 매번 별도의 협의와 조정이 필요한 상황<sup>10)</sup>
  - 사회복지통합관리망(2010년 1월 개통)은 42개 기관의 소득·재산, 지원·수혜이력 등 452종의 정보를 연계해 운영
- 공공 데이터 개방에 필요한 제도적 장치와 함께 데이터 접근성을 높일 수 있는 기술 기반을 구축<sup>11)</sup>
  - 사용자가 수작업으로 열람해야 하는 형식이 아니라 자동 업데이트 가능한 방식으로(Linked Data, API<sup>12)</sup> 등) 공유할 필요
  - 익명화 기술 등 데이터 보안 대책과 안정성 확보도 병행

### 큐레이션 고도화로 빅데이터의 한계를 극복

- 빅데이터 처리를 위한 최신 기술과 전문가의 지혜를 더해 보다 세련된 큐레이션이 가능
  - 현실 왜곡 정보를 제거하는 기술과 운영 노하우가 빠르게 발달
    - 특히 SNS와 같은 비정형 데이터에서 사용자 계층의 편중 현상과 의도적인 여론 개입 등을 필터링하는 솔루션이 증가
  - 같은 현상에 대해서도 상이한 맥락을 이해하고 섬세한 대응이 가능
    - 고객이 표출하는 표현의 뉘앙스와 이유를 찾아내는 고객 감성 분석(Customer Sentiment Analysis) 기법이 발달
  - 빅데이터에서 충분히 정보를 얻을 수 없거나 방향을 잡기 힘든 거대 담론은 전문가의 지혜를 반영하여 보완


<sup>10)</sup> 이만재 (2011). "빅 데이터와 공공 데이터 활용". 『Internet and Information Security』, 2(2).

<sup>11)</sup> 윤미영 (2011). "신가치창출을 위한 주요국의 빅데이터 추진전략 분석". 『IT & Future Strategy』, 11.

<sup>12)</sup> 외부에서 웹사이트의 내부 데이터를 자유롭게 접속할 수 있도록 공개된 인터페이스

- 데이터 분석 결과만으로는 경기 활성화 방안이나 차세대 스마트 기기 예측과 같은 이슈를 해결하는 데 한계<sup>13)</sup>

□ 빅데이터의 부작용으로 크게 우려되고 있는 정보 보안 리스크도 빅데이터 기술의 심화·발전으로 해결 가능

- 빅데이터 시대에는 외부로부터 사이버 공격 위협이나 내부 데이터에 대한 비이상적인 행동의 실시간 감지·대응 등으로 정보 보안 관련 기술 수준이 한층 업그레이드될 전망<sup>14)</sup>
  - 미국 500여 은행 지점을 보유한 지온스(Zions Bancorporation)는 빅데이터 기술을 적용하여 수 페타( $10^{18}$ )바이트에 이르는 내외부 DB와 네트워크를 실시간 감시
- 빅데이터 활용 기업은 데이터 운영의 투명성과 안전성을 높일 수 있는 기술 및 제도적 장치 마련을 최우선 과제로 전개
  - 데이터 암호화, 해킹 방지, 식별코드 분산 저장, 데이터 관리자의 엄중한 보안 책임 등 신뢰성 있는 조치가 필요
  - 꼭 필요한 자료인지, 어떻게 활용할 것인지 뚜렷한 방향도 없이 '일단 수집하고 보자'는 식의 접근은 지양
- 빅데이터의 유용성과 안전성을 쉽게 체감할 수 있는 대표 사례를 개발해 홍보함으로써 '빅데이터 + 큐레이션 = 섬세하고 안전한 기술'이라는 인식을 확산 

<sup>13)</sup> 브룩스, D. (『뉴욕타임스』 칼럼리스트) (2013.2.22.). "빅데이터의 한계". 『동아닷컴』 . <<http://news.donga.com/3/all/20130221/53210215/1>>

<sup>14)</sup> Curry, S., et al. (2013.1.). Big Data Fuels Intelligence-driven Security. *RSA Security Brief*.

## 【 지표로 보는 경제 Trend 】

## &lt; 1일 지표 &gt;

		4. 2	4. 3	4. 4	4. 5	4. 8
환율	원/달러(증가기준)	1,118.00	1,117.50	1,123.80	1,131.80	1,140.10
	엔/달러(뉴욕시장)	93.205	93.385	92.965	96.355	98.605
	달러/유로(뉴욕시장)	1.28460	1.28240	1.28430	1.29280	1.29930
금리	회사채(3년AA-, %)	2.84	2.84	2.85	2.81	2.84
	국고채(3년, %)	2.47	2.47	2.48	2.44	2.47
	CD (91일, %)	2.81	2.80	2.80	2.80	2.80
	LIBOR (3개월, %)	0.28210	0.28110	0.28040	0.27940	0.27940
국제 원자재 가격	두바이(S, \$/배럴)	108.41	107.96	105.14	104.19	103.13
	WTI(S, \$/배럴)	97.08	94.44	93.21	92.76	93.37
	CRB 현물가격지수	478.58	478.68	478.08	..	..
주가지수(KOSPI, 종가)		1,986.15	1,983.22	1,959.45	1,927.23	1,918.69
한국 5년만기 CDS 프리미엄		82	85	84	88	88

## &lt; 월별 지표 &gt;

	2011년	2012년	2012.11월	2012.12월	2013.1월	2013.2월	2013.3월
제조업생산 증가율 <sup>1)</sup>	6.0	0.8	1.9	-0.5	8.0	-9.8	..
평균가동률	80.2	78.1	78.0	78.4	78.7	77.8	..
서비스업생산 증가율	3.2	1.6	1.4	0.7	1.3	0.0	..
실업률	3.4	3.2	2.8	2.9	3.4	4.0	..
실업자(만명)	85.5	82.0	71.0	73.7	84.7	99.0	..
총취업자 수(만명)	2,424.4	2,468.1	2,494.1	2,440.2	2,405.4	2,398.4	..
소비자물가 상승률	4.0	2.2	1.6	1.4	1.5	1.4	1.3
생산자물가 상승률 <sup>2)</sup>	6.1	0.7	-0.9	-1.2	-1.6	-1.6	..
생활물가 상승률	4.4	1.6	1.0	0.7	0.8	0.8	0.8
수출(억달러, FOB) <sup>3)</sup>	5,552.1	5,480.8	477.9	450.3	457.0	423.4	..
(증감률)	(19.0)	(-1.3)	(3.9)	(-5.7)	(10.9)	(-8.9)	..
수입(억달러, CIF)	5,244.1	5,195.8	433.8	431.1	452.0	403.2	..
(증감률)	(23.3)	(-0.9)	(0.9)	(-5.2)	(3.9)	(-10.6)	..
외환보유액(억달러)	3,064.0	3,269.7	3,260.9	3,269.7	3,289.1	3,274.0	3,274.1

1) 통계청 (2013. 3. 29.) “2013년 2월 산업활동동향” 에 의거하여 작성

2) 2012년 12월부터 생산자물가 기준년도를 2010년으로 변경(2010=100)

3) 관세청 (2013. 3. 18.) “2013년 2월 수출입 동향(확정치)” 에 의거하여 작성

## &lt; 분기 지표 &gt;

	2011년	2012년	2011년 4분기	2012년 1분기	2012년 2분기	2012년 3분기	2012년 4분기
GDP 성장률 <sup>1)</sup>	3.6	2.0	3.3	2.8	2.3	1.5	1.5
민간소비	2.3	1.8	1.1	1.6	1.1	1.6	2.8
설비투자	3.7	-1.8	-3.3	8.6	-3.5	-6.5	-5.1
건설투자	-5.0	-1.5	-2.1	1.5	-2.1	-0.2	-4.1
SERI 소비자태도지수 (연간은 연말 기준)	45.4	44.9	45.4	44.2	46.6	43.6	44.9
개인금융자산(조원) <sup>2)</sup>	2,303.4	..	2,303.4	2,365.3	2,384.8	2,449.9	2,485.7
개인금융부채(조원)	1,103.5	..	1,103.5	1,106.9	1,121.4	1,135.4	1,158.8
재정수지(조원)	18.6	..	18.6	-11.3	-11.5	-3.7	..
경상수지(억달러) <sup>3)</sup>	260.7	432.5	115.1	25.6	111.9	145.6	149.4
총대외지불부담(억달러) <sup>4)</sup>	3,987.2	4,134.4	3,987.2	4,116.6	4,157.7	4,193.8	4,134.4

1) GDP 성장률은 전년동기 대비

2) 개인금융자산 및 부채는 자금순환계정 기준

3) IMF 국제수지통계매뉴얼(BPM5) 기준

4) IMF, World Bank 등 9개 국제기구가 마련한 새로운 편제기준, 분기별 발표

## &lt; 주요국 통계 &gt;

(전년동기 대비, %)

		2011년	2012년	2011년 4분기	2012년 1분기	2012년 2분기	2012년 3분기	2012년 4분기
미국	GDP 성장률 <sup>1)</sup>	1.8	2.2	4.1	2.0	1.5	2.0	-0.1
	실업률 <sup>2)</sup>	8.9	8.1	8.5	8.2	8.2	7.8	7.8
	소비자물가 <sup>2)</sup>	3.2	2.1	3.0	2.7	1.7	2.0	1.7
일본	GDP 성장률 <sup>1)</sup>	-0.6	1.9	-1.2	5.2	0.3	-3.5	-0.4
	실업률 <sup>2)</sup>	4.6	4.3	4.5	4.5	4.3	4.2	4.2
	소비자물가 <sup>2)</sup>	-0.3	0.0	-0.2	0.5	-0.2	-0.3	-0.1
유로 지역	GDP 성장률 <sup>1)</sup>	1.4	-0.4	-0.3	0.0	-0.2	-0.1	-0.6
	실업률 <sup>2)</sup>	10.2	11.4	10.7	11.0	11.4	11.6	11.7
	소비자물가 <sup>2)</sup>	2.7	2.5	2.7	2.7	2.4	2.6	2.2
중국	GDP 성장률 <sup>1)</sup>	9.3	7.8	8.9	8.1	7.6	7.4	7.9
	실업률 <sup>2)</sup>	4.1	4.1	4.1	4.1	4.1	4.1	4.1
	소비자물가 <sup>2)</sup>	5.4	2.6	4.1	3.6	2.2	1.9	2.5

1) 미국, 일본, 유로지역 GDP 성장률은 전기 대비 연율, 중국 GDP 성장률은 전년동기 대비

2) 실업률 및 소비자물가의 경우 분기 말 기준