

KISDI

Premium Report



빅데이터 산업과 데이터 브로커



정용찬



정보통신정책연구원 ICT통계분석센터장



정보통신정책연구원
KOREA INFORMATION SOCIETY DEVELOPMENT INSTITUTE



KISDI

15 - 04

2015. 05. 08

Premium Report

빅데이터 산업과 데이터 브로커

정 용 찬

정보통신정책연구원 ICT통계분석센터장

요약문	1
1. 머리말	2
2. 데이터 브로커의 정의와 특징	4
3. 데이터 취득과 가공	9
4. 서비스 유형과 구매자	13
5. 시사점	18

정 용 찬

정보통신정책연구원 ICT통계분석센터장

*ycjung@kisdi.re.kr, 043-531-4160

*고려대학교 응용통계학 박사

*전) DN컨설팅 이사(CRM 담당)

*전) 방송위원회 연구센터 연구위원

빅데이터 산업과 데이터 브로커

요약문

디지털 경제의 확산으로 형성된 빅데이터 환경에서 데이터는 부를 창출하는 새로운 ‘자원’으로 주목받고 있다. 미국의 경우 데이터 산업 발전 과정에서 소비자의 개인 정보를 수집해서 제3자와 공유하거나 재판매하는 기업을 의미하는 데이터 브로커(Data Broker)가 등장하게 되었다.

미국 데이터 브로커 산업의 특징은 광범위한 데이터 원천으로부터 다양한 소비자 데이터를 수집, 유통하고 있지만 대부분의 경우 소비자가 이를 정확하게 인지하지 못하고 있다는 점에 있다. 미국 공정거래위원회(FTC)는 데이터 브로커산업의 ‘투명성 부족’에 주목하고 의회에 소비자가 데이터 브로커의 존재와 활동을 인지하고, 이들이 보유하고 있는 개인 정보에 합리적으로 접근할 수 있도록 돋는 법률 제정을 권고했다.

우리나라는 미국처럼 데이터 브로커 산업이 활성화되지 않았지만 ‘데이터 기반(data-driven) 패러다임’이 확산될수록 사회 전 분야에서 데이터의 수집과 활용에 대한 요구는 급증할 것으로 전망된다.

‘정보 약자’인 소비자의 자기정보에 대한 권리 강화는 ‘규제 강화’가 아닌 ‘정보 유통의 투명성 제고’이고, 데이터 유통 활성화와 산업 발전도 소비자 신뢰에 기반 할 때 가능하므로 빅데이터 시대에 부합하는 개인정보 보호의 범위와 수준에 대한 사회적 합의를 위한 논의가 필요하다.

데이터 유통 활성화를 위해서는 데이터 보안, 암호화, 비식별화 등 개인정보보호를 위한 기술 개발이 필수적이며, ‘일상이 곧 데이터’를 의미하는 데이터 기반 사회를 대비하기 위한 국가 차원의 ‘데이터 역량 강화 전략’ 수립이 시급하다.

1. 머리말

- 사람들의 행동은 물론 생각까지도 데이터로 축적되는 ‘빅데이터(Big Data)’ 시대에 효과적으로 대응하기 위해 주요국 정부는 공공 정보의 전면적인 개방과 데이터 활용을 통한 가치 창출을 국가 전략으로 삼고 새로운 혁신을 도모
 - 영국 정부의 ‘오픈데이터 정책’과 ‘데이터 역량 강화 전략’은 데이터 활용을 통해 새로운 성장 동력을 찾기 위한 노력임(BIS 2014; HM Government 2013)
 - 미국 정부도 ‘빅데이터: 기회 포착과 가치 보호(Big Data: Seizing Opportunities, Preserving Values)’ 보고서에서 삶의 질을 높이면서 동시에 개인의 프라이버시를 보호하기 위한 빅데이터 정책을 천명(White House 2014)
 - 디지털 경제의 확산으로 형성된 빅데이터 환경에서 데이터는 부를 창출하는 새로운 ‘자원’으로 주목받고 있음
 - 정부가 공개한 오픈데이터를 활용한 벤처 기업이 빠른 성장으로 큰 부가가치를 창출하는 기업으로 성공하는 사례도 나타남
 - 1억 천만 건이 넘는 미국 주택 데이터베이스를 기반으로 온라인 부동산 장터 서비스를 개설한 벤처기업(Zillow)의 시가총액이 30억 달러에 달하며, 미국 정부의 작물재배 현황과 날씨 정보, 토양 데이터를 활용하여 농부들에게 정보를 제공한 벤처기업(Climate Corp)을 몬산토가 2013년 9억 3천만 달러에 인수한 사례가 있음(The World Bank 2014, 11-12)
- ※ 디지털경제는 인터넷 경제(Internet Economy), 신 경제(the New Economy), 웹 경제(Web Economy)로도 불리는데 전자상거래, 인터넷 쇼핑몰, 검색 서비스 등과

같이 인터넷을 비롯한 정보통신기술(ICT)을 기반으로 이루어지는 경제활동을 의미. 70개국을 대상으로 한 디지털경제 순위 조사에서 우리나라는 13위를 차지 (Economist Intelligence Unit 2010, 4).

- 민간 기업은 1990년대부터 데이터의 중요성을 인식하고, 고객 데이터를 축적, 분석하여 다양한 마케팅 활동을 추진하는 고객관계관리(CRM, Customer Relationship Management)를 실행(정용찬 2012, 6)
 - 미국의 경우 이 과정에서 데이터를 수집, 분석, 가공하는 서비스를 제공하는 데이터 브로커가 등장하게 되었으며 빅데이터 환경에서 급성장하고 있는 이들을 주목하기 시작함
 - 미국 회계감사원(GAO)은 상원의 요청에 따라 데이터 브로커의 성장에 따른 소비자 프라이버시 보호법 개선 방향에 대한 보고서를 작성 (U. S. GAO 2013). 미 의회도 급성장하고 있는 데이터 브로커 산업의 잠재적 위험과 혜택을 평가하고 적절한 소비자 보호 조치를 위한 정책입안자의 지속적인 관리감독의 중요성을 강조하는 보고서를 발표 (Majority Staff of Senate Committee 2013, 36)
 - 이러한 사회 분위기를 반영하여 미국 공정거래위원회는 9개 주요 데이터 브로커에 대한 조사를 통해 산업 실태를 파악한 보고서를 발표 (FTC 2014)
- 본 리포트는 미국 데이터 산업 생태계의 핵심 역할을 담당하고 있는 데이터 브로커의 실태 파악을 통해 빅데이터 산업 활성화와 프라이버시 보호가 조화를 이루기 위한 정책 방향을 모색해 보고자 함

2. 데이터 브로커의 정의와 특징

가. 빅데이터 산업과 데이터 브로커

- 빅데이터 산업은 부가가치 창출을 위해 빅데이터의 생산, 유통, 활용, 관리 등과 관련된 서비스를 제공하는 산업을 의미(정용찬 외 2014, 31)
 - 빅데이터 산업은 데이터를 매개로 하는 산업의 특성을 감안할 때 기반 산업, 지원산업, 활용산업으로 분류

〈표 1〉 빅데이터산업 분류

대분류	중분류	주요 기업
기반산업	빅데이터 관련 장비 제조업	IBM, HP, EMC 등
	빅데이터 관련 소프트웨어 개발업	오라클, SAS, SPSS 등
	빅데이터설비 임대/운영대행업	구글, 아마존, MS 등
지원산업	시스템구축 및 솔루션 제공업	
	컨설팅 및 사업서비스업	IBM, 오라클, SAS 등
	기타 지원업	
활용산업	데이터유통업	
	데이터가공업	Axiom, Corelogic, Datalogix 등
	기타 활용업	

자료: 정용찬 외(2014)의 31-32쪽에서 인용

- 디지털 경제에서 빅데이터(Big Data)는 거대 산업이며, 데이터 브로커 (Data Broker)는 빅데이터 생태계에서 중요한 역할을 담당(FTC 2014, i)
 - 정보 재판매업자(information resellers)로도 불리는 데이터 브로커란 소비자의 개인 정보를 수집해서 그 정보를 제3자와 공유하거나 재판매하는 기업을 의미(FTC 2014, i)
 - 빅데이터산업 분류의 관점에서 데이터 브로커는 데이터유통업과 데이터 가공업을 의미하는 활용산업에 해당

- 미국의 경우 인터넷이나 이메일이 등장하기 전부터 마케팅을 목적으로 소비자 데이터를 수집하고 판매하는 비즈니스 영역이 존재

- 우편 발송을 통한 직접 판매(direct mail)의 적중률을 높이기 위해 우편 번호(ZIP code)의 지역 구분과 잡지 구독, 식품 소비 유형, 정치성향 등 개인 정보를 연결하려는 다양한 시도가 진행(Majority Staff of Senate Committee 2013, 1)

※ 대표적인 사례로는 1974년 Jonathan Robbins이 개발한 PRIZM(Potential Rating Index for Zip Markets)으로 3만 6천개의 우편번호를 센서스 데이터, 소비자 조사 데이터와 결합하여 40개의 그룹으로 분류(Weiss 1988, 3–5). 현재 PRISM은 미국 인구를 생활양식에 따라 62개의 그룹으로, 사회경제적 요인별로 15개의 그룹으로 분류(Weiss 2000, 11–13)

- 이후 우편물을 통한 마케팅뿐 아니라 전화 마케팅 등의 목적을 위해 공공 정보, 사회 조사 결과 등을 활용한 소비자 정보의 축적이 진행되었으며 인터넷의 등장과 전자 결제, 스마트폰의 확산으로 수집되는 소비자 정보의 양이 급속도로 확장

- 이러한 역사적 배경을 지닌 미국의 데이터 유통 시장의 존재는 그동안 일정 부분 베일에 가려져 있어 투명성에 대해 논란이 있었음(Singerjune 2012)

- 미국 공정거래위원회(FTC)는 대표적인 데이터 브로커인 데이터로직스와 페이스북과의 개인정보 관련 협업과정에서 프라이버시 이슈를 위반했는지에 대해 조사할 것을 요구 받기도 함(Mills 2012)

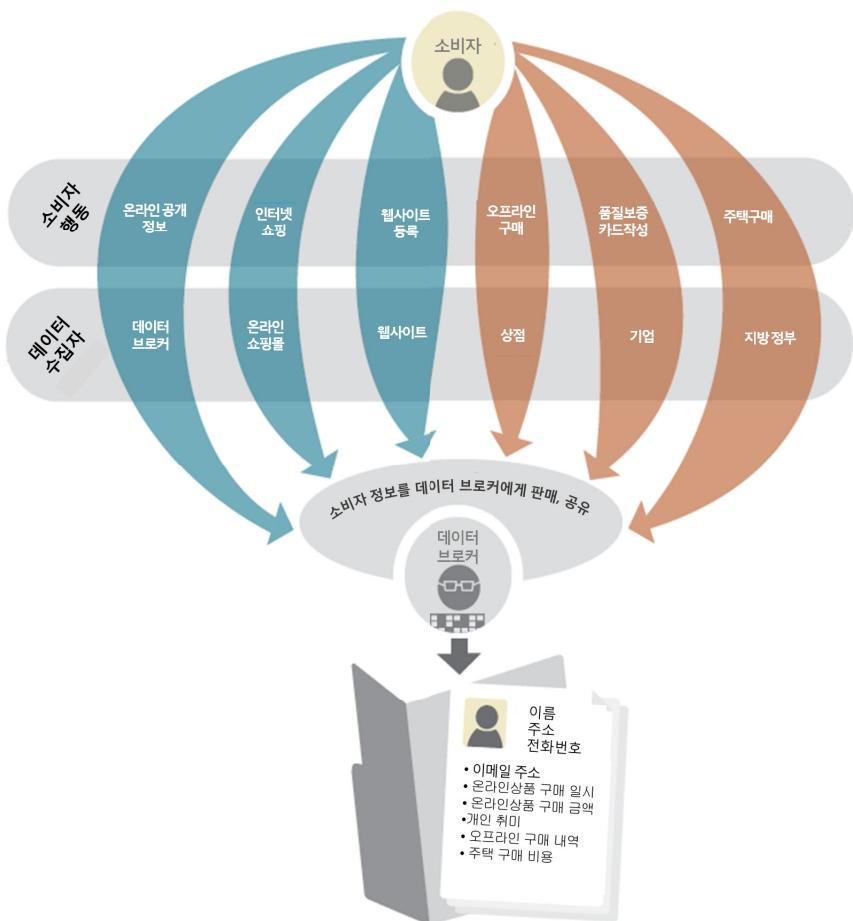
- 회계감사원(GAO)은 현재의 프라이버시 관련 법안이 마케팅 목적으로 소비자 데이터를 수집, 이용, 판매하는 행위에 대해 제한적으로 접근하고 있기 때문에, 민간 기업이 취급하고 있는 소비자 정보의 수집과 판매에 관한 포괄적인 프라이버시 법안의 필요성을 제기(U. S. GAO 2013, 4)

- 의회를 비롯해서 데이터 브로커 산업에 대한 투명성 요구가 확산되자 공정거래위원회(FTC)는 데이터 브로커 기업을 대상으로 조사를 벌여 비즈니스 실체를 공개하는 등 투명성 강화에 노력을 기울임

나. 주요 데이터 브로커와 데이터 수집

- 데이터 브로커가 수집하는 소비자 데이터는 온라인과 오프라인 전 영역에 걸쳐 다양한 유형을 포함하고 있음
 - 오프라인 상점은 물론 온라인 쇼핑몰에서의 구매 내역, 웹사이트의 등록 정보, 블로그, 홈페이지의 공개 정보, 주택 구매와 같은 공공정보에 이르기까지 소비자와 관련된 다양한 행동 정보가 데이터 브로커에 의해 수집되고 있음([그림 1])
- 미국 공정거래위원회(FTC)가 조사한 데이터 브로커는 빅데이터 분석을 통해 오바마 미 대통령의 당선을 도운 것으로 알려진 Axiom을 비롯해 Corelogic, Datalogix, eBureau, ID Analytics, Intelius, PeekYou, Rapleaf, Recorded Future의 9개 회사임
 - FTC조사의 의하면 데이터 브로커는 미국은 물론 전 세계 소비자와 관련된 다양한 데이터를 수집, 분석, 가공하여 마케팅과 광고, 부정거래 탐지 등을 지원하기 위한 다양한 서비스를 제공 중(<표 2>)

[그림 1] 데이터 브로커의 소비자 데이터 수집 과정



자료: FTC(2014)의 p. 2.

<표 2> 데이터 브로커의 소비자 데이터 수집 과정(FTC의 조사 대상 업체)

회사명	특징
Acxiom	- 마케팅캠페인, 부정사용 탐지를 위한 고객데이터 분석 서비스 제공 - 전 세계 7억 명의 소비자 정보가 담긴 데이터베이스 보유
Corelogic	- 산업계와 정부에 재무정보와 부동산정보에 기초한 분석서비스 제공 - 약 8억 건의 부동산 거래정보, 약 1억 건의 담보 데이터베이스 보유
Datalogix	- 거의 모든 미국 소비자의 마케팅 데이터를 제공 - 2012년 페이스북은 페이스북 이용자의 소셜사이트 상품광고 조회와 오프라인 상점의 구매 관련성 측정 위해 데이터로직스와 협력 발표
eBureau	- 마케터와 재무관련 회사, 온라인유통업체에 수익성이 높은 잠재 고객과 부정 거래 예측 서비스 제공 - 매달 평균 30억 건이 넘는 새로운 정보 추가 축적
ID Analytics	- 특정인 확인, 부정 거래 확인 서비스 제공 - 7천억 건의 데이터와 14억 건의 소비자 거래 데이터 보유
Intelius	- 신원 조회와 공문서 정보 제공 - 20억 건이 넘는 데이터베이스 보유
PeekYou	- 소셜미디어사이트, 홈페이지, 블로그의 콘텐츠를 분석 작성자 확인 서비스 제공
Rapleaf	- 이메일 주소와 함께 이메일 주소 소유자의 연령, 성, 우편번호, 소득, 결혼 여부, 자녀 여부와 취미, 구매 유형 등 정보 제공
Recorded Future	- 소비자와 기업의 과거 이력 데이터 분석을 통해 미래 행동 예측 정보 제공

자료: FTC(2014)의 pp. 8-9와 각 사별 홈페이지를 참고하여 작성

3. 데이터 취득과 가공

가. 데이터 취득

- 조사 대상 9개 기업 모두 소비자로부터 직접 데이터를 수집하지 않는 특징을 보임
 - 주요 데이터 취득 원천은 (1) 정부 데이터, (2) 그 외 공개된 정보, (3) 민간 데이터로 구분 가능(FTC 2014, 11)
- 정부 데이터는 연방정부와 주정부, 지방정부로부터 수집
 - 연방정부 데이터로는 센서스국(U.S. Census Bureau)의 인구통계정보와 주소, 선거구 등 행정 정보, 사회보장국(Social Security Administration)의 사망자 명부 정보, 우편서비스(U.S. Postal Service)의 주소 변경 정보, 연방법원의 파산 정보를 활용
 - 주정부와 지방정부 데이터로는 전문직 면허 정보, 부동산, 유권자, 자동차 등록, 법원 정보 등이 대표적
 - 민간 상업 정보로는 소매업과 카탈로그 회사의 거래 내역, 잡지사의 구독자 정보, 전자상거래, 뉴스, 여행 사이트, 금융회사의 거래 정보, 다른 데이터 브로커의 보유 정보 등이 해당

〈표 3〉 데이터 브로커의 분야별 수집 데이터

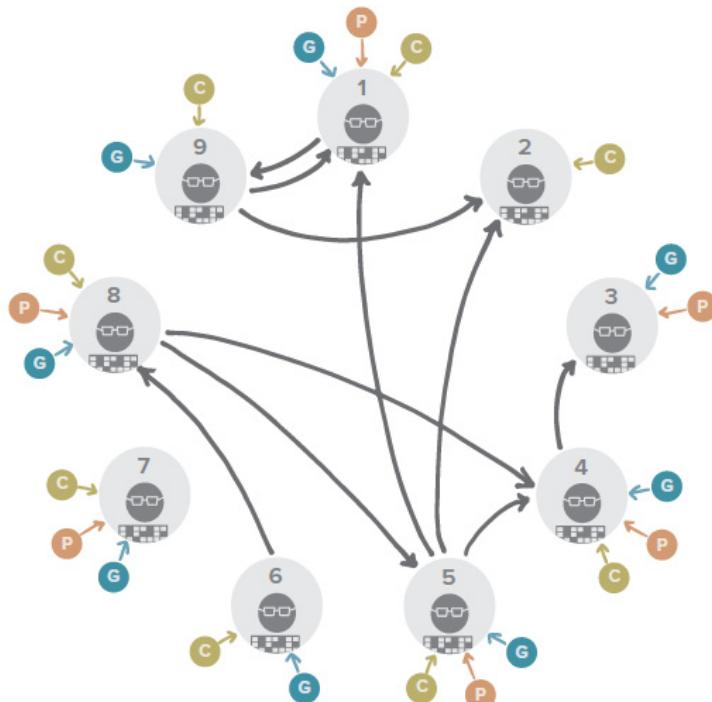
분야	데이터
정부	- 센서스국(U.S. Census Bureau) : 특정 도시나 구역의 인종, 연령, 교육수준, 가족 구성, 수입, 직업, 통근 시간, 도로, 주소, 행정 구역, 선거구, 학군 정보 - 사회보장국(Social Security Administration) : 사망자 이름, 사회 보장번호, 사망일 - 우편서비스(U.S. Postal Service)의 주소 변경 이력 정보 - 연방법원 : 파산 정보
	- 전문직 면허(의사, 변호사, 조종사, 건축사) 정보와 사냥, 낚시 면허 정보
	- 부동산 : 세무, 주택소유권과 담보, 소유자 정보, 부동산 세부 내역 정보(욕실 수, 침실 수, 수영장 유무 등)
	- 유권자 등록정보 : 이름, 주소, 생일, 가입 정당 - 자동차등록정보와 운전 기록 - 법원 : 범죄 기록, 민사 소송, 출생, 결혼, 이혼, 사망 기록
공개정보	- 전화번호부, 보도자료, 블로그와 소셜미디어사이트(링크드인과 같은 공개형) 등 인터넷에 올린 개인정보
상업정보	- 소매업, 카탈로그회사의 거래내역, 품목(예를 들어 고가의 구두, 친환경식품 등), 구매일자, 결재 방법 데이터 - 잡지사의 구독자 데이터 - 전자상거래, 뉴스, 여행 사이트의 고객 이름, 이메일, 우편번호 정보 - 금융서비스회사의 거래 데이터 - 다른 데이터 브로커가 보유한 데이터
	-
	-
	-
	-

자료 : FTC(2014)의 pp. 11-14의 내용 요약

- 대부분의 데이터 브로커는 정부, 공개정보, 상업정보로부터 데이터를 확보하고 있으며 데이터 브로커 회사 간의 데이터 유통도 활발함([그림 2])
 - 다른 데이터 브로커와 데이터를 거래하지 않는 회사는 조사 대상 9개사 중 1개사에 불과
- 데이터 수집 방법으로는 웹으로부터 크롤링(web crawling), 인쇄정보 구매 등을 활용

- 인쇄정보로는 전화번호부, 지방정부 기록 등이 해당
 - 데이터 취득 과정에서 데이터 브로커가 필요한 정보만을 선별적으로 구매하지 못하는 과잉 취득 상황도 발생
- ※ 웹 크롤링(web crawling)은 봇(bots), 웜(worms), 웹 스파이더(web spider), 웹 로봇(web robot) 등으로 불리는 웹 크롤러(web crawler)가 자동화 방법으로 월드 와이드 웹을 탐색하여 필요 정보를 찾아내는 것을 의미(위키피디아)
- 수집된 데이터는 데이터 특성에 따라 실시간, 일간, 주간, 월간 등 기간 별로 현행화(update)

(그림 2) 데이터 브로커의 데이터 획득 경로



자료: FTC(2014)의 p. 15. 여기서 G는 정부, P는 공공, C는 상업 정보 원천을 의미.
1~9는 FTC가 조사한 9개 데이터 브로커 회사를, 화살표는 데이터 브로커 간의 데이터 이전 경로를 의미.

나. 데이터 가공

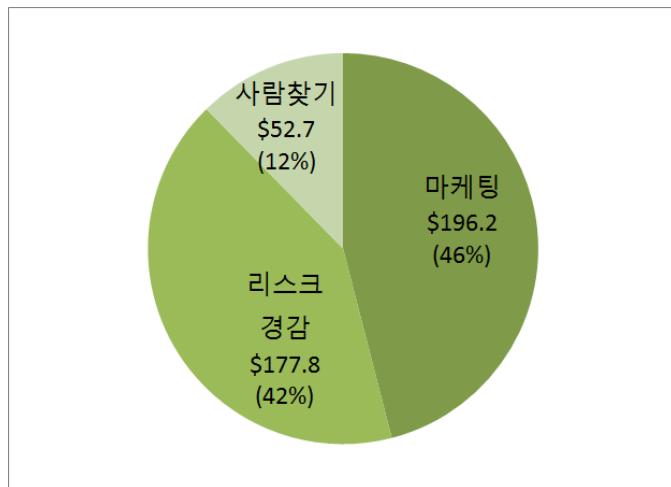
- 수집된 데이터는 기존 데이터와 연결, 가공, 재분류 등 판매를 위한 상품화의 과정을 거침(FTC 2014, 19)
 - 데이터 브로커는 고객에게 원데이터를 판매하기도 하지만 고객 특성을 분석하여 새로운 고객군(segment) 정보를 생성 판매함
- 고객군 정보를 생성하는 방법으로는 유사한 특성을 가진 고객군으로 분류하기 위해 데이터를 요소별로 결합하거나, 고객의 행동을 예측하기 위한 모형을 개발함(FTC 2014, 19-21)
 - 예를 들어 ‘사커맘(soccer mom)’ 고객군은 자녀가 있고 최근 2년 이내에 스포츠 용품을 구매한 경험이 있는 21세에서 45세 사이의 여성 의미
 - 또한 최근 1년 이내에 캠핑 장비를 구매한 경험이 있는 고객의 정보를 분석하여 공통 특징을 발견하여 ‘캠핑 장비를 구매할 가능성이 있는 고객’을 발굴
 - 이들이 개발한 고객군은 ‘부유한 베이비부머(baby Boomer)’, ‘알러지로 고통 받는 사람’, ‘친구가 250명이 넘는 트위터 사용자’ 등과 함께 소수 인종, 고연령층, 저소득층도 분류
- ※ ‘segment’는 마케팅 용어로 특성별로 분류된 집단을 의미하며 이렇게 분류하는 과정을 ‘세분화(segmentation)’라고 표현
- 고객데이터는 일반적으로 인구통계정보, 취미, 구매 습관 등 개인 정보(individual consumer profile) 단위로 저장되나, 경우에 따라 ‘은행 계좌 개설’ 등과 같은 특정 이벤트 단위로 저장하기도 함

4. 서비스 유형과 구매자

가. 제공 서비스

- 데이터 브로커가 제공하는 상품은 (1) 마케팅, (2) 리스크 경감(risk mitigation), (3) 사람 찾기의 세 범주로 구분(FTC 2014, 23)
 - 9개 데이터 브로커의 총매출은 2012년 기준 4억2천6백만 달러(한화로 약 4570억원) 규모로 이 중에서 마케팅 분야가 약 1억9천6백만 달러에 달함(FTC 2014, 23)

[그림 3] 9개 데이터 브로커의 상품별 매출(2012년 기준, 단위: 백만\$)



◆ 마케팅 상품

- 마케팅 상품은 소비자에게 맞춤형 마케팅 메시지를 전달하는 과정에 도움을 주기 위한 목적으로 직접마케팅(direct marketing), 온라인 마케팅, 마케팅 분석으로 구분(FTC 2014, 23)

- 직접마케팅(direct marketing) 상품은 데이터 추가 제공(data append)과 마케팅 명부(marketing list) 작성으로 구성(FTC 2014, 23)
 - 데이터 추가 제공이란 데이터 브로커가 고객의 의뢰를 받아 고객이 가지고 있는 명단에 필요한 항목을 추가로 제공하는 것을 의미
 - 예를 들어 이름과 주소정보만 가지고 있는 고객 명부에 전화번호와 구매 습관 정보, 이메일정보 등을 추가하여 우편 발송을 통한 직접 판매 (direct mail), 텔레마케팅, 이메일 마케팅 캠페인 등에 활용할 수 있도록 제공. 경우에 따라서 집전화나 휴대전화 이메일 주소 정보를 가지고 있을 경우 소유자의 이름과 주소 정보를 결합하여 제공하기도 함
 - 데이터 브로커가 제공하는 정보로는 연령, 성, 직업과 같은 인구통계 정보뿐 아니라 키, 몸무게, 흡연 여부, 자녀 유무, 신용카드 사용, 부동산 담보 유형 등 개인과 가족 정보는 물론 인종, 종교, 정치성향과 같은 민감한 정보도 포함되어 있음
 - 마케팅 명부 작성 상품은 최소한 두 명 이상의 자녀가 있는 가정, 특정 종류의 차를 소유하고 있는 여성, 당뇨병에 관심이 있는 사람 등과 같이 특정한 조건에 해당되는 소비자의 명부를 작성하는 것을 의미. 명부는 마케팅 목적에 따라 이름과 주소(direct mail 용도), 이름과 전화 번호(텔레마케팅 용도), 이메일 주소(이메일 마케팅 용도)정보를 제공
- 온라인 마케팅(online marketing) 지원 상품은 인터넷, 모바일 기기, 케이블과 위성TV를 통해 효과가 가장 높은 고객군을 선별, 마케팅이나 광고 활동을 지원하기 위한 데이터 상품을 의미. 온라인 마케팅 지원 상품은 등록 고객 타겟팅(registration targeting)과 협업 타겟팅(collaborative targeting)으로 구분(FTC 2014, 26)
 - 등록 고객 타겟팅은 특정 목적의 마케팅을 위해 인터넷 업체의 고객 명단을 데이터 브로커에게 보내면 이 중에서 가능성이 있는 고객을

선별해주는 서비스

- 협업 타게팅은 인터넷에 기반한 두 사업체가 협업을 원하지만 각자의 고객 정보를 공유하기를 원치 않을 경우 데이터 브로커가 중간에서 두 사업체의 고객 정보를 결합, 분석하여 각 기업이 필요로 하는 정보를 제공

- 마케팅 분석(marketing analytics)

- 마케팅 분석이란 고객 데이터 분석을 통해 행동을 예측하는 것을 목적으로 한 서비스를 의미. 이를 통해 특정 상품이나 브랜드를 판매하거나 홍보하기 위해서는 어떤 미디어 채널(온라인, 신문, TV 등)이 적합한지, 어떤 지역이 최적인지 조언을 하며, 광고 효과에 대한 분석 서비스를 제공하기도 함(FTC 2014, 31)

◆ 리스크 경감

- 리스크 경감 서비스는 고객으로부터 발생할 수 있는 위험(risk)을 감소시키기 위한 서비스로 신원 식별(identity verification)과 사기 탐지(fraud detection)로 구분(FTC 2014, 32)
- 데이터 브로커의 신원 식별 서비스는 은행과 같은 금융기관이 고객이 거래하는 시점에서 자사 고객의 신원을 인증하는데 도움을 주기 위해 거래 건별로 위험 점수를 산출, 제공하는 등의 방법으로 신원 확인을 지원
- 사기 탐지는 소비자가 회사에 제출한 이메일 주소 등의 정보가 정확한지를 확인하는 것과 같은 민간 영역뿐 아니라 정부보조금 지원을 위해 공공 기관에 제출한 소득 정보의 진위 여부도 판단하는 등 공공 부문에도 활용(FTC 2014, 33-34)

◆ 사람 찾기

- 사람 찾기 서비스는 기업이 보유하고 있는 고객에 대한 세부 정보를 제공하는 서비스로 주로 정부나 소셜미디어 사이트와 같이 공개된 정보를 기초로 작성된 다양한 정보로 구성되는데 여기에는 이름, 주소, 전화번호, 이메일 주소와 같은 정보 이외에 결혼 기록, 생년월일, 범죄 전과, 이혼 경력, 주소 이력, 취미/관심사, 부동산 소유와 거래 내역, 친지, 고용 기록, 재산 소유 내역과 거래 정보 등이 포함(FTC 2014, 34)

나. 데이터 구매자

- 데이터 브로커의 서비스를 이용하는 구매자는 자동차, 에너지, 교육, 의료, 보험, 통신, 미디어, 금융 등 다양한 산업을 포괄
 - 민간 영역뿐 아니라 지방정부와 주정부, 연방정부와 같은 공공 기관, 자선단체와 대학교와 같은 비영리조직도 데이터 구매자에 포함되어 있는 점이 특징(Majority Staff of Senate Committee 2013, 29)

<표 4> 데이터 브로커의 고객군별 서비스 유형

구분	다이렉트 마케팅	온라인 마케팅	마케팅 분석	개인 식별	사기 탐지	사람 찾기
변호사/수사관	✓					
자동차산업	✓	✓	✓			
데이터 브로커	✓	✓	✓	✓	✓	
교육기관	✓		✓	✓		
에너지	✓					
정부	✓		✓	✓	✓	✓
관광여가	✓	✓	✓			
개인고객						✓
보험	✓		✓	✓	✓	✓
대출기관/금융	✓	✓	✓	✓	✓	✓
마케팅/광고	✓	✓	✓	✓	✓	✓
미디어	✓		✓			✓
비영리/정치캠페인	✓	✓		✓	✓	
제약	✓		✓			✓
부동산	✓				✓	✓
소매	✓	✓	✓	✓	✓	✓
기술기업	✓	✓	✓			✓
통신	✓		✓	✓	✓	

자료: FTC(2014)의 pp. 39–40의 표 일부

5. 시사점 : 데이터 생태계와 데이터 브로커

- 미국 데이터 브로커 산업의 특징을 요약하면 광범위한 데이터 원천으로부터 다양한 소비자 데이터를 수집, 유통하고 있지만 대부분의 경우 소비자가 이를 정확하게 인지하지 못하고 있다는 점
 - 데이터 브로커는 거의 모든 미국 소비자와 관련된 데이터를 수집하고 있으며 데이터 브로커 간에도 이를 유통시키는 등 데이터 거래관계가 복잡하여 소비자의 입장에서는 자신의 데이터가 어떤 경로를 통해 유통되는지 파악하기 곤란한 상황
 - 특히 다양한 소비자 정보를 연결하여 분석하는 과정에서 민감한 프라이버시 문제를 야기할 가능성이 상존. 예를 들어 소비자의 행동 데이터 분석을 통해 다양한 고객군(segment)으로 분류하는데 이 중에는 ‘출산 예정’, ‘당뇨병 관심’과 같이 민감한 추론 정보도 존재
- 사기 방지, 개인 기호에 부합하는 상품 추천, 맞춤형 광고 제공 등 데이터 브로커의 활동으로부터 소비자가 얻는 이득도 동시에 존재함
 - 분실된 신용카드를 제3자가 부정 사용할 경우 이상 거래임을 감지하거나, 기존의 광고 시스템에서는 홍보하기 어려운 혁신적인 소기업의 제품을 필요로 하는 사람에게 광고로 전달하거나, 오랫동안 연락이 끊겼던 지인 찾기 등은 소비자의 입장에서 유용한 서비스임
 - 물론 이 과정에서 소비자가 피해를 입을 가능성도 존재함. 예를 들어 ‘오토바이 매니아’나 ‘당뇨병 관심’ 고객군으로 분류될 경우 맞춤형 광고를 통해 필요한 제품 정보와 할인 기회를 얻을 수 있지만 보험사는 이러한 고객을 고위험군으로 분류할 가능성이 있음
 - 특히 이러한 분석 정보는 수집된 다양한 데이터를 기초로 분석과정을 통해 얻은 추론 결과이므로 원천 데이터에 오류가 있거나 추론 결과가

틀릴 경우 해당 소비자가 불이익을 입을 수 있으며, 경우에 따라 불이익을 당하는 소비자가 그 이유를 알지 못할 수 있음

- 미국 공정거래위원회는 데이터 브로커산업 조사 과정에서 ‘투명성 부족’에 주목하고 이를 강화하기 위해 의회에 소비자가 데이터 브로커의 존재와 활동을 인지하고, 이들이 보유하고 있는 개인 정보에 합리적으로 접근할 수 있도록 하는 법률을 제정할 것을 권고함
 - 즉 데이터 브로커가 보유하고 있는 개인 정보에 대해 소비자가 접근할 수 있도록 하고, 마케팅 목적으로 개인정보를 공유할 경우 소비자가 이를 거부할 수 있도록 제안
- 프라이버시를 기본 인권으로 인식하고 있는 유럽과 달리 미국은 사고 팔 수 있는 상품으로 간주하고 있으며 이러한 차이는 각 국가의 역사적 배경과 연관되어 있음(Craig & Ludloff 2011, 78–79)
 - 이러한 인식의 차이로 미국의 경우 프라이버시는 일종의 자산으로 프라이버시의 침해에 대한 판단은 그로 인한 손실의 존재 여부에 달려 있음
 - 최근 미국 공정거래위원회의 데이터 브로커 산업에 대한 투명성 강화 방침은 공정거래위원회가 데이터 브로커 산업에 대해 조사를 시작한 20년 전과 비교할 때 비약적으로 발전한 데이터 산업의 위상에 부합하는 투명성과 소비자 통제권의 강화 필요성을 인식했기 때문임
- 우리나라의 경우 미국처럼 데이터 브로커 산업이 활성화되지 않았지만 ‘데이터 기반(data-driven) 패러다임’이 확산될수록 사회 전 분야에서 데이터의 수집과 활용에 대한 요구는 급증할 것으로 전망
 - 다양한 소비자 정보를 보유하고 있는 기업에 비해 ‘정보 약자’인 소비자의 자기정보에 대한 권리 강화는 ‘규제 강화’가 아닌 ‘정보 유통의 투명성 제고’이고, 소비자 신뢰 확보를 기반으로 할 때 데이터 유통 활성화와 빅데이터 산업 발전이 가능(정용찬 외 2014, 133)

- 방송통신위원회는 개인정보의 오·남용을 방지하면서, 빅데이터 산업의 활성화를 목적으로 ‘빅데이터 개인정보보호 가이드라인’ 발표. 이를 계기로 빅데이터 시대에 부합하는 개인 정보의 유통과 활용, 개인정보 보호의 범위와 수준에 대한 사회적 합의를 위한 논의가 지속되어야 함
※ 개인정보 보호의 범위는 사회의 역사적 배경과 제도에 따라 가변적임. 예를 들어 부패인식지수 조사에서 청렴도 최상위 국가군에 속하는 핀란드의 경우 개인 소득과 납세 정보는 누구나 확인 가능한 공개 정보임
- 데이터 유통 활성화를 위해서는 데이터 보안, 암호화, 비식별화 등 개인정보보호를 위한 기술 개발이 중요함
 - 미국 오바마 행정부는 프라이버시 기술 관련 연구개발을 위해 국립과학재단, 국립보건원, 국토안보부, 국가안보국 등 관련 기관을 중심으로 매년 7천7백만 달러의 예산을 집행(White House 2014, 55)
※ 주요 사업으로는 익명화 기술, 프라이버시 보호 데이터 통합, 프라이버시를 위한 데이터 세분화, 보건데이터 익명화 보호, 클라우드 컴퓨팅·데이터 통합·데이터 마이닝을 위한 프라이버시 솔루션 등이 있음
 - 우리 정부도 ‘빅데이터 산업 발전 전략’에서 빅데이터 기술 핵심 7대 분야 중 하나로 제시한 익명화 처리 등 ‘개인정보보호 기술’ 개발의 실행을 통해 프라이버시 보호와 관련된 원천 기술 확보가 시급(관계부처 합동 2013)
- 온라인 거래의 확산과 디지털 콘텐츠 이용의 확대, CCTV와 사물인터넷의 증가로 우리 환경은 ‘일상이 곧 데이터’를 의미하는 데이터 기반 사회로 진화하고 있으며 이를 효과적으로 대처하기 위한 국가 차원의 전략 수립이 필요
 - 빅데이터 환경에서는 대용량 데이터의 처리 속도와 함께 다양한 데이터의 조합을 통한 정확한 예측 능력이 곧 글로벌 경쟁력이며 이러한 원천 기술의 유무가 국가 경쟁력의 척도로 작용

- 데이터 기반 사회에서도 인적 자원의 중요성은 여전하며, ‘데이터 과학자(Data Scientist)’와 같은 전문가 양성과 함께 국민 개개인의 데이터 역량 강화를 위한 국가 전략이 필요

※ 영국은 ‘데이터 역량 강화 전략’(A strategy for UK data capability)의 핵심을 초·중등 교육 과정 개혁과 빅데이터 활용 선도 역할을 수행할 공무원 교육 강화로 설정(HM Government 2013, 20–28)

참 고 문 헌

〈국내 문헌〉

관계부처 합동(2013). 『창조경제 및 정부3.0 지원을 위한 빅데이터 산업 발전 전략』.

정용찬 (2012). 『빅데이터 혁명과 미디어 정책 이슈』. KISDI Premium Report 2012-02. 정보통신정책연구원.

_____ (2013) 『빅데이터』 커뮤니케이션북스.

정용찬·한은영 (2014). 『빅데이터 산업촉진 전략 연구: 해외 주요국 정부 사례를 중심으로』. 정보통신정책연구원.

〈국외 문헌〉

BIS (2014). "Open Data Strategy: 2014–2016."

Craig, T and Ludloff, M. E. (2011). *Privacy and Big Data*. O'Reilly.

Economist Intelligence Unit (2010). "Digital Economy ranking 2010: Beyond e-readiness".

Federal Trade Commission (2014). "Data Brokers: A Call for Transparency and Accountability."

HM Government (2013). "Seizing the data opportunity: A Strategy for UK data capability."

Majority Staff of Senate Committee on Commerce, Science, & Transportation, Office of Oversight & Investigations (2013). "A Review of the Data Broker Industry: Collection, Use, and Sale of Consumer Data for Marketing Purposes."

Mills, Elinor(2012). "Privacy experts to ask FTC to probe Facebook-Datalogix deal", (Sep. 24). <http://www.cnet.com/news/privacy-experts-to-ask-ftc-to-probe-facebook-datalogix-deal/>

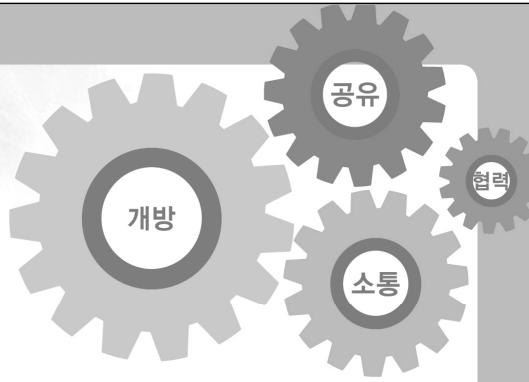
- SingerJune, Natasha (2012). "Mapping, and Sharing, the Consumer Genome." New York Times (June 16). <http://www.nytimes.com/2012/06/17/technology/acxiom-the-quiet-giant-of-consumer-database-marketing.html>
- United States Government Accountability Office (2013). "Information Resellers: Consumer Privacy Framework Needs to Reflect Changes in Technology and the Marketplace."
- Weiss, J. Michael (1988). *The Clustering of America*. Harper & Row Publishers.
- _____ (2000). *The Clusteed World*. Little, Brown and Company.
- White House (2014). "Big Data: Seizing Opportunities, Preserving Values."
- The World Bank (2014). "Open Data for Economic Growth."

행복한
대한민국을 여는

정부 3.0

[개방 · 공유 · 소통 · 협력]

국민의 기대와 희망을 모아 새로운 변화를 시작합니다.
국민 한 분 한 분을 위해 특별한 내일을 준비합니다.
개인의 행복이 커질수록 함께 강해지는 새로운 대한민국
그 희망의 새 시대를 정부3.0이 함께 열어가겠습니다.



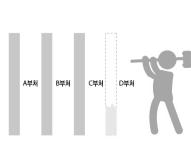
국민 모두가 행복해지는 정부3.0



공공정보 공개확대로
「국민의 알 권리」총족



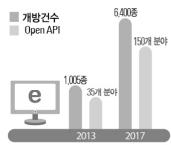
국민의
정부정책 참여확대



정부 내 칸막이 제거로
통합적 행정서비스 제공



정보 취약계층
서비스 접근 제고



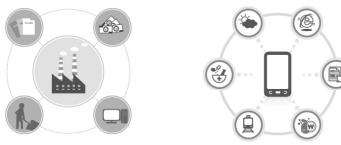
공공데이터 민간활용으로
새로운 일자리창출



정보공유와 디지털협업으로
더 나은 행정서비스 제공



데이터에 기반한
과학적 행정구현



창업과 기업활동
지원 강화

새로운 정보기술을 활용한
맞춤형 서비스 제공