

빅데이터 시대와 데이터 융합

▣ 박 선 우*

빅데이터라는 새로운 형태의 데이터의 등장과 함께 산업 내에서는 이와 같은 빅데이터를 활용하기 위한 다양한 시도들이 이루어지고 있다. ICT기술의 발전으로 인해 급속도로 생성되고 있는 빅데이터를 저장·처리하는 기술 또한 점차 진보해나가고 있다. 이로 인해 빅데이터 시대에는 단순히 대용량의 데이터를 보유하는 것을 넘어서 어떻게 하면 그 안에서 가치있는 정보를 정제해낼 수 있을지에 더 많은 관심이 집중되고 있다. 그 중에서도 이종간 데이터 결합은 서로 다른 종류의 데이터를 결합하여 데이터의 잠재가치를 무궁무진하게 넓힐 수 있다는 점에서 주목받고 있다. 본고에서는 국내에서 이루어지고 있는 다양한 빅데이터 융합분석 사례들을 살펴보고 이를 통해 빅데이터 시대의 새로운 데이터 활용방안에 대한 시사점을 제시하고자 한다.

목 차

I. 서 론 / 2

II. 빅데이터 시장 현황 / 3

1. 세계 빅데이터 시장 규모 / 3
2. 국내 빅데이터 시장 규모 / 6

III. 빅데이터와 데이터 융합 / 8

1. 공공데이터: 공공데이터포털 / 8
2. 민간데이터 / 10
3. 민-관 데이터 융합 13

IV. 결론 및 향후 전망 / 22

* 한국신용정보원 조사역, (02)3705-5787, psw@kcredit.or.kr

I. 서론

오늘날 우리는 빅데이터의 시대를 살아가고 있다. 빅데이터(Big Data)란 기존의 데이터 용량을 넘어서는 거대한 규모의 다양한 정형·비정형 데이터를 가리킨다. 글로벌 시장조사 기관인 IDC에 의하면 2025년 전 세계의 데이터 생산량은 약 163ZB 정도의 크기가 될 것으로 예측되며, 이는 2016년에 생성된 16.1ZB보다 약 10배정도 늘어난 규모이다. ICT기술의 발전과 함께 등장한 스마트 기기의 확산, 소셜네트워크서비스(SNS)의 활성화, 사물인터넷(IoT)의 보급 등으로 인해 빅데이터의 기반은 더욱 확대되고 있으며, 이들로부터 생성되는 방대한 양의 데이터들을 저장하고 처리하는 기술 또한 점점 진보해나가고 있다.

빅데이터의 등장은 개별 산업과 기업의 생존전략을 완전히 바꾸어 놓았다. 앞서 세계적인 IT 리서치 전문회사인 가트너의 수석부사장 피터 손더가드(Peter Sondergaard)는 “빅데이터는 21세기의 원유”라고 정의한 바 있다. 원유를 어떻게 가공하느냐에 따라 만들 어낼 수 있는 제품의 종류가 무궁무진하듯이, 빅데이터 역시 데이터를 어떻게 활용하느냐에 따라 이전에는 미처 발견하지 못했던 수많은 새로운 가치들을 창조해낼 수 있다. 따라서 빅데이터 시대에는 단순히 데이터를 수집하고 저장하는 데에서 더 나아가 보유한 데이터를 어떻게 잘 활용할 수 있는지가 기업의 흥망성쇠를 결정짓게 되었다고 해도 과언이 아니다. 이러한 측면에서 오늘날 개별기업이 당면한 가장 큰 과제는 넘쳐나는 데이터의 홍수 속에서 데이터 정제를 통해 가치 있는 정보만을 선별하고 새로운 부가가치를 창출하는 기술을 개발하는 것이다. 현재 각 산업 분야별로 많은 기업들이 빅데이터를 활용하여 새로운 서비스를 개발하려는 시도를 하고 있지만 실제적으로 수익을 내는 모델로 발전시킨 경우는 극히 드물다. 이는 기업들이 빅데이터라는 새로운 기술을 활용한다고는 하지만 해당 업권에서 기존에 보유하고 있던 정보의 양이 늘어난 것 외에 큰 틀에서 다양한 원천의 정보를 활용해 수익을 창출할 수 있는 빅데이터 분석력은 아직 부족하다는 의미이다.

이런 한계를 극복하고자 빅데이터 시대에 새롭게 부상하고 있는 기술 중 하나가 바로 이종간 데이터 융합이다. 지금까지는 유사한 종류의 데이터 결합 및 분석을 통해 새로운 기회를 발굴해내는 데 초점을 두었다면, 빅데이터 시대에서는 상이한 특성을 가진 데이터

들을 상호 결합하여 데이터의 활용 범위를 넓히고 보다 깊은 함의를 도출해 낼 수 있다. 이런 시대적 흐름에 발맞춰 국내에서도 다양한 데이터 융합 시도가 이루어지고 있다. 민-관 데이터 연계, 업무협약을 통한 서로 다른 업권 간 데이터 교류 등의 사례는 빅데이터라는 새로운 기술을 효과적으로 활용하고자 하는 국가와 산업 내의 다양한 노력의 결과물이라고 할 수 있다. 이에 따라 본고에서는 현재까지 국내에서 이루어지고 있는 민-관, 산업간의 이중 데이터 융합 사례들을 살펴보고 이를 통해 빅데이터 시대의 새로운 데이터 활용 방안을 탐구해보고자 한다.

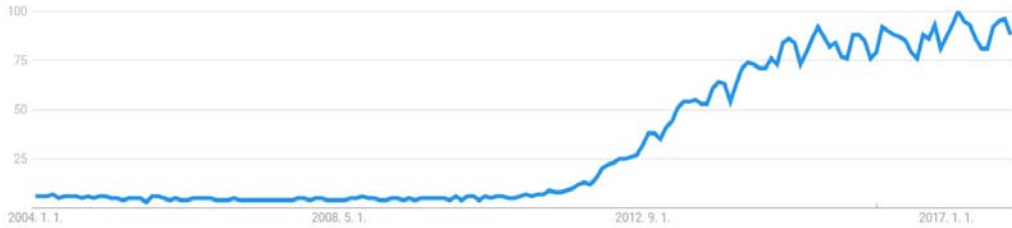
II. 빅데이터 시장 현황

1. 세계 빅데이터 시장 규모

구글트렌드(<https://trends.google.com/>)에 따르면 '빅데이터(Big Data)'의 구글 검색 관심도*는 구글트렌드 서비스가 시작된 2004년 초반에는 거의 0에 가까웠으나, 2012년부터는 점차 가파르게 증가하기 시작하여 2016년에는 100에 가까운 수치를 기록하였다. 또한 세계경제포럼에서 '2012년의 떠오르는 10대 기술(Top 10 Emerging Technologies of 2012)' 중 하나로 빅데이터를 지목한 이후로 전 산업권에 걸쳐 빅데이터에 대한 관심은 꾸준히 증가해왔으며 이에 따라 세계 빅데이터 시장 규모 또한 점차 확대되고 있는 추세이다.

* 특정 지역 및 기간의 차트에서 가장 높은 지점에 대한 검색 관심도를 0~100사이의 수치로 나타낸 것. 100은 검색어의 최고 인기도이며, 검색어의 인기도가 최고 인기도 대비 1% 미만인 경우에는 0으로 표시.

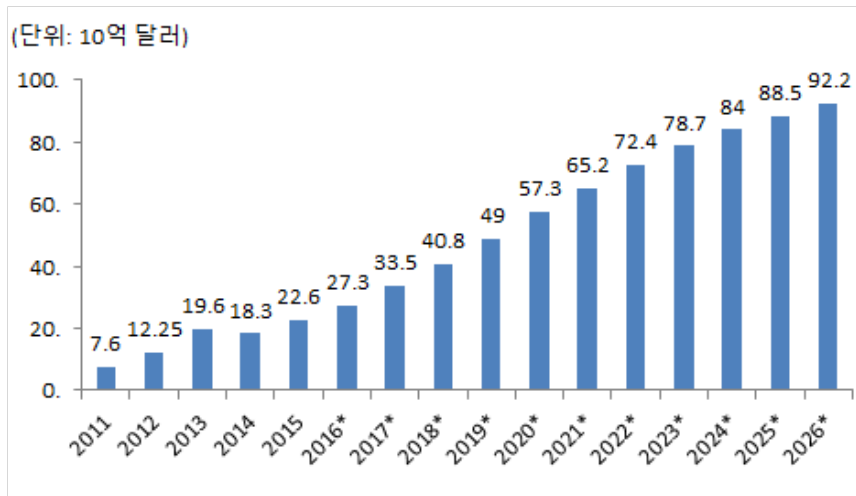
[그림 1] 구글트렌드 ‘빅데이터(Big Data)’ 검색 관심도: 2004~현재



출처: 구글트렌드 (2017.12.13.)

시장조사기관인 Wikibon에 따르면 소프트웨어, 하드웨어, 서비스를 모두 포함한 세계 빅데이터 시장은 앞으로 10년 후인 2026년에 총 922억 달러의 규모로 성장할 것으로 전망된다. 이는 2014년 기록했던 183억 달러에서 약 404% 증가한 수치이며, 2014년부터 2026년까지의 연평균 성장률은 14.4%에 달한다.

[그림 2] 빅데이터 세계시장 규모 예측: 2011~2026



주: *는 예측치, 출처: Statista

원출처: Wikibon Big Data in the Public Cloud Forecast, 2016-2026 (Wikibon)

IDC 또한 빅데이터 및 분석시장(Big Data and Business Analytics)이 2016년부터 매년 11% 이상의 높은 성장률로 성장하여 2020년에는 2천 100억 달러 규모에 다다를 것으로 전망하고 있다. 현재 세계 빅데이터 및 분석시장은 금융과 제조업이 견인하고 있는 구조이다. 2016년도 빅데이터 및 분석시장 산업별 비중은 은행업(13.1%)이 가장 높았으며 뒤이어 조립제조(11.9%), 공정제조(8.4%), 연방/중앙정부(7.6%), 전문서비스(7.4%) 순이었다.

〈표 1〉 빅데이터 및 분석시장 세계 규모

년도	매출액 (단위: 10억 달러)	성장률
2015	122.0	-
2016	130.1	6.6%
2017*	150.8	15.9%
2018*	168.7	11.9%
2019*	188.8	11.9%
2020*	210.0	11.2%

주: *는 예측치, 출처: Statista

원출처: Worldwide Semiannual Big Data and Analytics Spending Guide (IDC)

〈표 2〉 2016년 빅데이터 및 분석시장 산업별 비중

산업	비중
은행업	13.1%
조립제조	11.9%
공정제조	8.4%
연방/중앙정부	7.6%
전문서비스	7.4%
기타	51.7%

출처: Statista

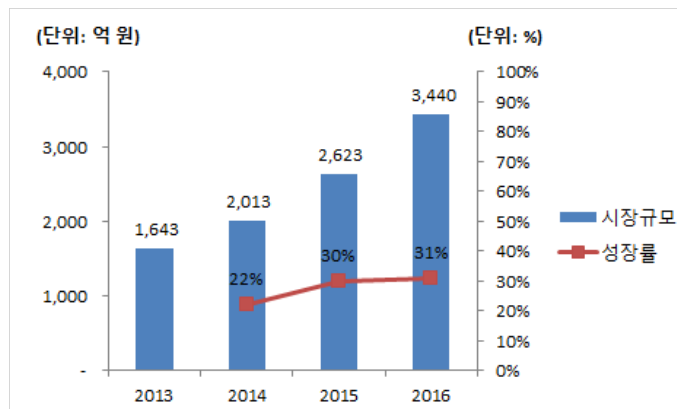
원출처: Worldwide Semiannual Big Data and Analytics Spending Guide (IDC)

이처럼 각국의 정부와 산업계가 빅데이터가 가진 잠재성에 주목하기 시작하면서 세계 빅데이터 시장은 매년 가파르게 성장하고 있다. 빅데이터 기술이 기존 산업의 수익성 뿐만 아니라 미래 산업의 경쟁력을 높이기 위한 필수 자산으로 인식되면서 당분간 세계 빅데이터 시장의 지속적인 성장세는 지속될 전망이다.

2. 국내 빅데이터 시장 규모

국내 빅데이터 시장 역시 기업들의 빅데이터 분야에 대한 투자 증가, 정부의 빅데이터 산업 지원 노력 등에 힘입어 꾸준히 성장하고 있는 추세이다. 한국정보화진흥원의 '2016년 빅데이터 시장현황 조사'에 따르면 2016년 우리나라의 빅데이터 시장 규모는 전년 대비 31.1% 성장한 3,439.6억원이었다. 뿐만 아니라 국내 빅데이터 시장은 최근 3년(2014~2016년) 동안 연평균 27.9%의 성장률을 기록하며 매우 빠른 속도로 그 규모가 확대되고 있다.

[그림 3] 국내 빅데이터 시장규모 추이: 2013-2016



출처: 2016년 빅데이터 시장현황조사 (한국정보화진흥원)

2016년 국내 빅데이터 시장의 영역별 비중을 살펴보면 민간시장 71%, 정부·공공시장 29%로 이루어져 아직까지는 민간영역이 빅데이터 시장을 주도하고 있는 것으로 보인다. 그러나 정부·공공시장의 경우 정부의 적극적인 빅데이터 활용정책에 따라 전년 대비 43.1%의 높은 성장률로 시장비중이 2%p 증가하였으나, 민간시장은 전년 대비 26.8% 성장하는 데 그쳐 시장비중이 오히려 하락하였다.

〈표 3〉 2016년 시장영역별 국내 빅데이터 시장 규모

시장구분		2015년 (단위: 억 원)	2016년 (단위: 억 원)	성장률
정부/공공 시장		698.0	998.6	43.1%
민 간 시 장	기업의 빅데이터 시스템 투자	1,791.0	2,278.4	27.2%
	빅데이터 분석 대행 서비스 시장	134.0	162.6	21.3%
	소계	1,925.0	2,441.0	26.8%
합계		2,623.0	3,439.6	31.1%

출처: 2016년 빅데이터 시장현황조사 (한국정보화진흥원)

한국과학기술정보연구원의 예측에 따르면 2020년 국내 빅데이터 시장 규모는 약 900만 달러에 다다를 것으로 전망된다. 국내 ICT 시장에서 빅데이터 분야가 차지하는 비중 또한 2020년에는 2.6%로 2013년의 0.6%에서 약 4배 가량 증가할 것으로 보인다. 이처럼 국내 빅데이터 시장 또한 세계 빅데이터 시장의 확대에 힘입어 지속적으로 성장해나가고 있는 추세이다.

〈표 4〉 국내 빅데이터 시장 규모 전망

연도	국내 빅데이터 시장 규모 (단위: 만 달러)	국내 ICT 시장 대비 비중
2013	163.4	0.6%
2014	208.1	0.7%
2015	263.2	0.9%
2016	332.8	1.1%
2017	422.7	1.3%
2018	539.3	1.6%
2019	692.1	2.0%
2020	893.8	2.6%

주: 국내 빅데이터 시장 규모 = 국내ICT 시장 규모 × 빅데이터 시장 진입 비율

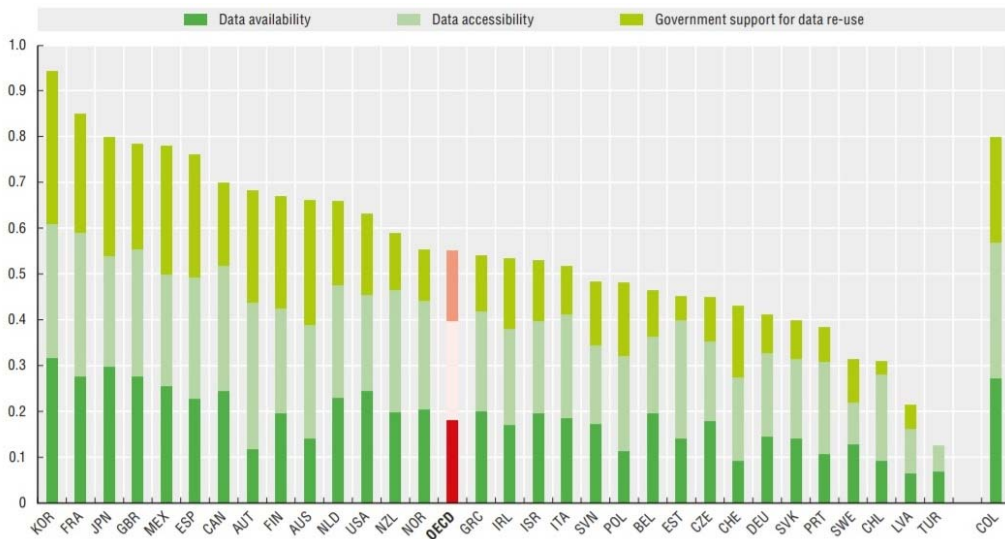
출처: 빅데이터 산업의 현황과 전망 (KISTI MARKET REPORT, 2013.4)

Ⅲ. 빅데이터와 데이터 융합

1. 공공데이터: 공공데이터포털

우리나라는 2013년 「공공데이터 제공 및 이용 활성화에 관한 법률」을 제정한 이후 공공데이터를 적극적으로 개방하려고 노력하고 있다. 정부는 이러한 공공데이터 개방을 통해 정책수립과정에 대한 국민들의 이해력 증대와 행정처리의 투명성 및 책임성 강화 등의 효과를 기대하고 있다. 이런 노력 덕분에 우리나라는 2017년 OECD 공공데이터 개방 평가에서 1위를 차지하였고, 현재 우리나라의 데이터 개방 정도는 세계적으로도 매우 높은 수준이라고 할 수 있다.

[그림 4] 2017년 OECD 공공데이터 개방 평가



출처: Government at a Glance 2017 (OECD)

행정안전부에서는 공공기관이 생성 또는 취득하여 보유하고 있는 공공데이터를 한 곳에서 제공하는 통합창구로 공공데이터포털(<https://www.data.go.kr>)을 운영하고 있다. 2011년 7월 ‘국가공유자원포털’이란 이름으로 처음 포털을 개방한 이후, 공공데이터포털에서 제공하고 있는 데이터 건수는 2013년 5,272건에서 2017년 12월 현재 24,636건으로 약 5배 가량 늘어났다. 공공데이터포털에서는 파일데이터, 오픈API, 표준데이터, 시각화 등의 다양한 유형으로 데이터를 제공하고 있으며, 포털에서 제공하지 않고 있는 공공데이터의 경우 공공데이터 제공신청을 통해 제공여부를 심의받은 후 이용할 수 있다.

〈표 5〉 공공데이터포털 데이터 개방 건수

구 분	2013년	2014년	2015년	2016년	2017년 (12.13 현재)
	누적결산 (‘13년말)	누적결산 (‘14년말)	누적결산 (‘15년말)	누적결산 (‘16년말)	
파일데이터	4,718	11,796	14,115	19,184	22,121
오픈API	554	1,361	1,790	2,143	2,457
표준데이터	-	-	7	31	58
합 계	5,272	13,157	15,912	21,358	24,636

출처: 공공데이터 개방 및 활용 주요 현황 (공공데이터포털, 2017.12.1.)

공공데이터포털을 통한 데이터 활용 또한 점차 증가하고 있다. 2017년 10월말 기준 공공데이터포털의 데이터 다운로드(활용신청) 누적건수는 총 3,505,731건으로 2013년의 13,923건에서 약 251배 증가하였다. 2017년 한 해 동안의 데이터 다운로드 건수도 1,621,481건이나 되어 매년 백만건 이상의 데이터 다운로드가 이루어지는 것으로 나타났다. 공공데이터를 활용하여 웹/앱 서비스를 개발한 사례 또한 2017년 10월 말 누적건수 총 1,401건으로 2013년의 42건에서 약 32배 정도 늘어났다.

〈표 6〉 공공데이터포털 데이터 활용 현황

구 분	2013년	2014년	2015년	2016년	2017년	총 계
파일데이터	4,108	115,525	589,783	978,104	1,497,464	3,184,984
오픈API	9,815	23,872	40,159	108,766	104,635	287,247
표준데이터	-	-	511	13,607	19,382	33,500
합 계	13,923	139,397	630,453	1,100,477	1,621,481	3,505,731

구 분	2013년	2014년	2015년	2016년	2017년 (누적)
웹/앱 서비스 개발사례	42개	395개	695개	1,056개	1,401개

출처: 공공데이터 개방 및 활용 주요 현황 (공공데이터포털, 2017.12.1.)

개인과 기업은 공공데이터포털에 공개된 공공데이터를 다운받아 자신이 보유한 외부 데이터와 결합하여 새로운 서비스나 신기술을 개발하는 데 활용할 수 있다. 이처럼 공공 데이터포털은 일반 국민들의 데이터 접근성을 높이고 공공데이터의 활용 범위를 넓혔다는 점에서 의의가 있다.

2. 민간데이터

(1) 네이버 데이터랩(DataLab)

민간기업이 자사가 보유한 데이터를 공개하고 있는 가장 대표적인 사례는 2016년 1월에 베타버전을 오픈한 네이버의 ‘데이터랩(DataLab)’을 들 수 있다. 네이버는 같은 해 6월에 모바일에서 특정 노래를 검색하면 네이버 뮤직 이용자 데이터를 활용하여 해당 곡의 성별 및 연령별 인기도와 기간별 추이 등을 제공하는 ‘뮤직 데이터랩(DataLab)’ 베타버전의 서비스도 시작하였다. 현재 데이터랩에서는 분야별 인기 검색어, 검색어 트렌드, 지역별 관심도 등의 데이터와 함께 지역·업종별 카드사용통계를 제공하고 있다.

〈표 7〉 네이버 데이터랩 제공 데이터

분류	제공항목	내용
검색어로 알아보는 대한민국	급상승 트래킹	30초 단위로 실시간 급상승 검색어 순위(1~20위) 제공
	분야별 인기 검색어	주제별(TV오락, 쇼핑, 영화, 자동차, 게임), 사용자별(싱글녀, 주부, 대학생, 청소년) 검색어 순위(1~10위) 제공
	검색어 트렌드	기간별/연령별/성별/검색기기(모바일, PC)에 따른 네이버 통합 검색어와 네이버 쇼핑 클릭수 추이를 제공
지역통계	지역별 관심도	지역(시/군/구)과 특정 업종을 선택하면, 선택된 업종에 대한 하위 지역들(읍/면/동/리)의 검색 관심도를 제공
	카드사용통계	비씨카드에서 제공하는 데이터를 기반으로, 지역별/기간별/업종별 카드사용통계를 제공
	맞춤형 트렌드 분석 도구	선택한 지역과 업종의 성별/연령별 카드사용통계를 제공 (복수의 지역과 복수의 업종간의 카드사용통계 비교 가능)
공공데이터	공개된 공공데이터	e-나라지표, 경기도 공공데이터, 공공데이터포털, 국가기록원, 문화재청, 서울열린데이터광장, 통계청, 특허청, 한국고전번역원, 한국문화정보원 등에서 공개한 공공데이터 제공

주 1): 검색 관심도는 특정 검색어의 주간 네이버 검색횟수를 0~100 사이의 상대적 지표로 표기한 것
 주 2): 카드사용금액은 상대적으로 표현한 값이며 0~100 사이의 값으로 표시

데이터랩 사용자들은 “검색어 트렌드 분석”을 통해 여러 개의 주제어에 대한 검색 관심도를 비교해볼 수 있으며, 기간·성별·연령·검색기기(모바일, PC) 등의 세부항목을 설정하여 보다 자세한 분석을 진행할 수 있다. 특히나 BC카드의 카드결제 데이터를 네이버의 검색어 데이터와 융합하여 분석할 수 있다는 점이 데이터랩의 가장 큰 특징이라고 할 수 있다. 이는 민간 사업자의 데이터와 네이버의 검색어 데이터를 융합한 최초의 사례이자 정형데이터와 비정형데이터의 결합으로서 그 의미가 매우 크다. 네이버 검색어 데이터가 소비자들의 구매를 예측하는 데 유용했다면, BC카드의 카드결제 데이터를 통해서도 소비자의

실제 구매 결과를 확인할 수 있다. 따라서 향후에는 데이터랩에서 제공되는 이러한 상이한 두 유형의 빅데이터 융합분석을 통해 사회현상에 대한 보다 정확한 진단과 분석이 가능할 것으로 기대되고 있다.

(2) SKT 빅데이터 허브(Big Data Hub)

자사가 보유한 데이터를 공개하고 있는 또 다른 사례로는 SKT의 ‘빅데이터 허브(Big Data Hub)’가 있다. 2013년 10월에 오픈한 SKT 빅데이터 허브는 민간 소유의 빅데이터를 대중에게 개방한 국내 첫 사례이다. SKT 빅데이터 허브는 SK텔레콤이 보유한 통신 데이터와 공공기관에서 공개한 데이터들을 주제별로 분류하여 제공하고 있다. 이 외에도 개별 사업자가 SKT 빅데이터 허브 사이트에 가입하여 법인 저작권의 데이터를 직접 등록 및 전시할 수 있다.

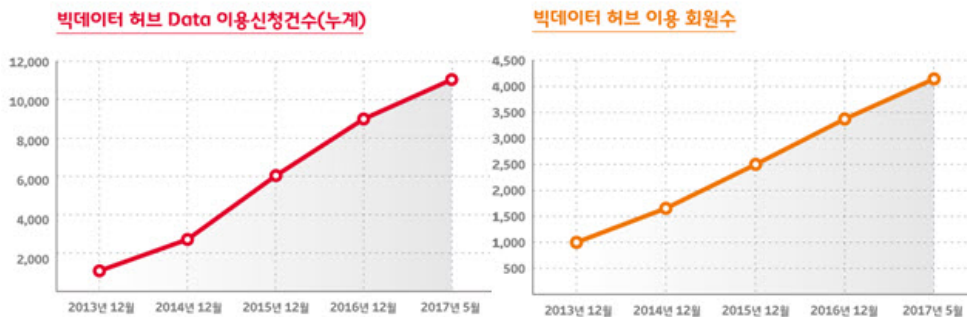
〈표 8〉 SKT 빅데이터 허브 제공 데이터

분류	제공 데이터
산업/경제	특히 실용신안 통합이력정보, 해외특허 서비스 등
문화/예술	외국인 가입자 시간대별 체류지 데이터, 국민관광 실태조사 등
정보통신/IT	월간 가입동향, 광고 집행금액 및 트래픽 리포트 등
사회/교육	지역별 콜택시 이용 분석, 유아교육기관 이용분석 데이터 등
공공/행정	웨더플래닛 기상정보, 전국 시내버스 노선정보, 사회 약자 시간대별 체류지 데이터 등

2017년 5월 말 기준으로 SKT 빅데이터 허브 사이트의 회원수는 4천명을 넘어섰으며, 데이터 이용 신청 건수도 1만 1천건 이상을 기록했다. 2017년 12월 13일 현재 SKT 빅데이터 허브에서 제공되고 있는 데이터는 총 979건으로 다른 데이터포털과 비교하면 상대적으로 적은 수준이지만, SKT 빅데이터 허브가 가진 가장 큰 장점은 “DATA 매쉬업”

이라는 메뉴를 통해 외부 데이터와의 융합분석이 가능하다는 점이다. “DATA 매쉬업”은 CHART, MAP 등의 기능을 활용해 T-데이터와 외부 데이터를 결합하여 다각도로 분석해볼 수 있게 해 주는 서비스이다. “DATA 매쉬업” 기능을 통해 사용자는 단일 데이터를 사용했을 때보다 훨씬 더 통찰력 있는 분석결과를 도출할 수 있다.

[그림 5] SKT 빅데이터 허브 이용 현황



출처: SKT '빅데이터 허브'

3. 민-관 데이터 융합

빅데이터 시대에는 데이터가 바로 경제적 자산이 된다. 매일 매순간 생성되고 있는 엄청난 크기의 데이터들을 어떻게 잘 활용하느냐에 따라 빅데이터가 창출해낼 수 있는 가치 또한 무궁무진하다. 이에 따라 우리나라 정부 또한 빅데이터를 활용하여 정부의 역량을 강화하고 보다 나은 공공서비스를 제공하기 위해 다양한 정부/민간 데이터 융합분석을 추진해오고 있다.

(1) 해외유입 감염병 차단

메르스, 지카바이러스와 같은 해외 감염병의 국내 유입 위험이 높아짐에 따라 과학기술정보통신부(舊 미래창조과학부)와 질병관리본부는 KT가 가진 통신 데이터를 활용하여 해

외감염병 예방을 위한 민·관 빅데이터 융합사업을 추진하였다. 과학기술정보통신부는 질병관리본부와 함께 KT가 제안한 “로밍 빅데이터를 활용한 해외유입 감염병 차단 서비스”를 ‘16년 빅데이터 선도 시범사업’으로 선정하였으며, 같은 해 11월 16일부터 “스마트 검역 정보시스템 고도화 사업”을 시범적으로 운영하였다. 이와 같은 사업시행이 가능했던 이유는 2015년에 「감염병의 예방 및 관리에 관한 법률」 개정으로 제76조의2(정보제공요청 등) 조항이 신설됨에 따라 해당 조항에 근거하여 KT가입자의 해외 로밍데이터를 해외감염병 예방에 활용할 수 있게 되었기 때문이다.

〈감염병의 예방 및 관리에 관한 법률〉

제76조의2(정보 제공 요청 등)

- ① 보건복지부장관 또는 질병관리본부장은 감염병 예방 및 감염 전파의 차단을 위하여 필요한 경우 관계 중앙행정기관(그 소속기관 및 책임운영기관을 포함한다)의 장, 지방자치단체의 장(「지방교육자치에 관한 법률」 제18조에 따른 교육감을 포함한다), 「공공기관의 운영에 관한 법률」 제14조에 따른 공공기관, 의료기관 및 약국, 법인·단체·개인에 대하여 감염병환자등 및 감염이 우려되는 사람에 관한 다음 각 호의 정보 제공을 요청할 수 있으며, 요청을 받은 자는 이에 따라야 한다.
1. 성명, 「주민등록법」 제7조제3항에 따른 주민등록번호, 주소 및 전화번호(휴대전화번호를 포함한다) 등 인적사항

스마트 검역정보시스템 고도화 사업은 해외 로밍데이터를 활용해 감염병 오염국가에서 국내로 입국하는 사람을 확인한 후 감염병 잠복기간 동안 모니터링하는 검역정보화 사업이다. 먼저 통신사 고객이 감염병 오염국가에 방문하여 로밍을 하면, KT에서는 감염병 오염국가 방문정보를 질병관리본부에 제공하고 고객에게 감염병 신고안내 문자를 전송한다. 감염병 오염국가를 방문한 고객이 국내 도착 시 질병관리본부에 입국자 정보가 전송되고 KT에서는 고객에게 입국시 자진신고 안내 문자를 발송한다. 또한 고객이 감염병 증상이 발현하여 의료기관에 방문하게 되면 건강보험심사평가원과 국민건강보험공단에서는

질병관리본부가 보유한 오염국가 방문이력정보를 해당 의료기관으로 전달하여 진료 시에 해당 내용을 조회할 수 있도록 했다. 이를 통해 통신사 고객은 오염국가를 방문하거나 귀국시에 관련 감염병 안내 문자를 받아볼 수 있으며, 질병관리본부에서는 오염국가 방문자 통계를 검역 정책 수립 및 해외유입 감염병 대응방안 마련을 위한 자료로 활용할 수 있다. 또한 의료기관에서는 환자의 감염병 오염국가 방문 이력 정보를 확인하여 진료에 활용할 수 있다.

[그림 6] 스마트 검역정보시스템 고도화 사업 개념도



출처: 민·관 빅데이터 융합을 통한 해외유입 감염병 차단 본격화 (미래창조과학부 보도자료, 16.5.24)

<표 9> 스마트 검역정보시스템 주요 활용데이터

구분	분류	주요변수	설명	보유기관
로밍 데이터	로밍 등록 신호	로밍 휴대전화번호	일 50,000건 내외	(주)케이티
		로밍국가코드		
		로밍사업자코드		
	로밍 취소 신호	휴대전화번호		
감염병 발생국가 데이터	감염병 오염국가 리스트	귀국 여부	메르스/지카바이러스 오염국가 전체	질병관리본부
		감염병명		
		국가명		
		사용여부		

출처: 2016 빅데이터 스마트서비스 시범사업 결과보고서 (한국정보화진흥원)

이와 같은 스마트 검역시스템의 목표 중 하나는 기존에 오염국가 방문여부를 확인할 수 없었던 ‘오염국가 방문 후 청정국가 경유 귀국자’*들을 로밍데이터를 활용하여 구분해냄으로서 적절한 검역서비스를 제공하는 것이었다. 시범사업 기간동안 매일 약 5,000명 정도의 감염병 오염국가 방문자와 귀국자가 발생하였으며, 특히 시범사업의 목표였던 ‘오염국가 방문 후 청정국가 경유 귀국자’를 매일 200명 정도씩 찾아 추가적인 검역서비스를 제공할 수 있었다.

〈표 10〉 스마트 검역정보시스템 시범서비스 결과

조건	11월22일	11월23일	11월24일	11월25일	11월26일	11월27일	11월28일
해외도착SMS 발송자수	5,045	4,623	5,129	5,011	4,489	4,636	4,751
국내입국SMS 발송자수	3,458	3,363	3,336	4,312	4,326	4,927	4,740

청정국가 경유 귀국자	11월21일	11월22일	11월23일	11월24일	11월25일	11월26일	11월27일
메르스	9	8	19	16	11	24	29
지카바이러스	168	144	178	188	209	255	293

출처: 2016 빅데이터 스마트서비스 시범사업 결과보고서 (한국정보화진흥원)

2017년 4월부터는 서비스 대상이 KT 뿐만 아니라 LGU+, SKT로 확대 적용되어 휴대 전화를 사용하는 전 국민을 대상으로 감염병 안내 문자 서비스가 제공하고 있다. 이를 통해 우리나라는 세계 최초로 “통신 로밍 데이터를 활용한 감염병 확산 방지 모델”을 성공적으로 구축한 사례로 평가받고 있다.

* 과거 메르스 1번 확진자 및 지카바이러스 1번 확진자 모두 청정국가 경유 입국으로 인해 감염병 정보 전달이 제대로 이루어지지 못함 (2016 빅데이터 스마트서비스 시범사업 결과보고서, 한국정보화진흥원)

(2) 서울시 심야버스노선 개발

서울시에서는 빅데이터 분석을 통해 실제로 시민이 필요로 하는 심야버스노선을 개발하고 2013년부터 해당 노선을 운행하는 일명 ‘올빼미버스’ 서비스를 시행하고 있다. 우리나라 수도권 서울에서는 낮 시간대 못지 않게 야간 시간의 경제활동 역시 활발하게 이루어지고 있으나 심야에 이용할 수 있는 교통편은 거의 택시로만 한정되어 있는 상황이었다. 야간 교통서비스의 수요 대비 공급 미달 상황과 그로 인한 교통서비스 수준의 악화(택시 승차거부, 비용부담), 안전 문제 등으로 인해 시민들의 야간 경제활동은 많은 제약을 받고 있었다. 이러한 문제를 해결하기 위해 서울시에는 공공 심야전용버스를 개발하려는 계획을 세웠다. 이런 심야전용버스 서비스가 성공적으로 시행되기 위해서는 심야버스의 경로별 수요를 파악하여 최적의 운행노선도를 개발하는 것이 무엇보다 중요했다. 서울시는 심야버스노선 개발을 위해 KT와 업무협약(MOU)을 맺고 휴대전화 통화이력 데이터에서 추출한 개인위치정보를 활용하기로 했다. KT는 공익사업을 위한 민·관 데이터 협력이라는 큰 틀 하에서, 개인정보보호를 위해 개인 신변과 관련된 데이터 항목을 제외한 통화기록 정보를 서울시에 제공하였다.

서울시는 우선 2013년 3월 한 달 동안의 KT 통화데이터 30억 건을 분석하여 심야시간(0~5시)대의 유동인구 밀집지역을 파악하였다. 여기에 500만 건에 달하는 시민들의 심야택시 승하차 데이터를 결합하여 최적의 심야버스노선을 구축하였다. 3개월(4.19~7.31) 동안 시내버스 2개 노선을 시범운행한 결과 총 22만 명(일평균 2,100명)이 심야전용버스를 이용한 것으로 나타났다. 그리하여 같은 해 9월부터는 본격적인 심야전용버스 서비스를 시작하였고 이후 2016년 11월에는 운행지역을 확대하고 운행차량의 증차를 통해 시민들의 편의를 제고하였다. 2016년 12월 한 달 동안의 심야전용버스 일평균 승객은 9,883명으로 2013년 시범운영 때보다 약 5배 정도 증가하였다. 노선별로는 N61(1,865명), N13(1,395명), N26(1,235명) 순으로 가장 많은 승객이 이용한 것으로 집계되었다. 요일별로는 금요일에서 토요일로 넘어가는 새벽 시간대의 이용승객수가 가장 많았으며, 월평균 이용 승객수는 금요일(12,378명), 토요일(10,741명), 목요일(10,342명) 순으로 나타났다.

〈표 11〉 심야전용 시내버스 노선별/요일별 이용승객수

대상노선		일평균 이용승객수(명)	요일	월평균 이용승객수(명)
N13	상계동~장지동	1,395	월	8,857
N15	우이동~사당역	1,152	화	9,713
N16	도봉산~온수동	1,125	수	10,240
N26	방화동~신내동	1,235	목	10,342
N30	강일동~서울역	596	금	12,378
N37	진관동~장지동	1,020	토	10,741
N61	신정동~노원역	1,865	일	5,953
N62	신정동~면목동	1,121		
N65	개화동~시흥동	374		
총계		9,883		

출처: '올빼미버스' 운행 50일, 하루 6천명 이용 (서울특별시청)

뿐만 아니라 심야전용버스 증차후 택시 승차거부 신고건수도 2015년 12월 823건에서 2016년 12월 774건으로 전년 동기 대비 6%(49건)가 줄어들었다. 이와 같은 서울시의 '올빼미버스' 운영 사업은 한정된 자원을 효율적으로 운영하여 시민들에게 직접적인 편익을 제공한 사례로서 2016년에 열린 “제1회 대한민국 지방자치 정책대상”에서 대상을 수상하였다.

(3) 한국교통연구원의 ViewT 플랫폼

교통량과 관련된 정보는 도로정책을 수립하는 데 있어 핵심지표로 활용되고 있으며 해당 데이터의 정확성에 따라 도로정책의 효과도 크게 달라질 수 있다. 그러나 교통량 정보는 고속도로 이외의 도로에서는 매우 제한적으로 수집되고 있으며 일부 지자체에서는 아예 수집하지 않는 경우도 있어 실제로 활용하기에는 많은 어려움이 존재했다. 국토교통부와 한국교통연구원은 이와 같은 교통 데이터의 한계점을 극복하고자 공공부문의 교통량 빅데이터와 민간의 내비게이션 빅데이터를 융합하여 교통량 자료가 수집되지 않는 도로의 교통량까지도 추정할 수 있는 기술을 개발하였다. 민간의 내비게이션 자료는 현대엠엔소

프트에서 제공하는 2015년 1년치 내비게이션 데이터 87억 건*이 활용되었다. 민간의 내비게이션 자료는 내비게이션 이용의 대중화로 방대한 양의 빅데이터가 수집되고 있으나 「개인정보보호법」에 따라 외부 제공이 제한되고 있으므로 주로 교통소통정보 도출에 국한되어 활용되고 있다.

한국교통연구원은 일부 구간에서만 관측되던 공공의 관측교통량 자료와 전국 단위로 수집되는 민간의 내비게이션 자료를 융합하여 전국의 미관측 도로의 교통량을 추정할 수 있는 ‘ViewT 1.0’을 올해 9월부터 서비스하기 시작하였다. 기존 현장조사를 통한 전국도로대비 교통량 수집량은 고작 3%에 불과하였으나 한국교통연구원이 개발한 민·관 빅데이터 융합 기술을 활용하면 전국도로의 95%까지 교통량을 확인할 수 있게 되었다. 공공의 관측교통량 자료와 민간의 내비게이션 자료는 서로 매우 밀접하게 연관되어 있으면서도 각자의 한계를 보완해줄 수 있는 빅데이터라고 할 수 있다. 내비게이션 자료는 전국을 대상으로 수집되기 때문에 데이터 수집범위 측면에서의 장점이 있으며, 관측교통량 자료는 내비게이션 표본 데이터를 전수화하는 데 있어 기준역할을 하므로 두 자료는 교통량 추정에 있어 상호보완의 관계라고 할 수 있다. 그 결과 두 데이터를 융합하여 미관측 도로의 교통량 추정 알고리즘 기술을 개발할 수 있었던 것이다. ‘View T 1.0 (<http://viewt.ktadb.go.kr>)’을 통해 분석한 우리나라에서 교통량이 제일 많은 도로는 평일 일평균 161,741대의 교통량이 발생한 강남대로(한남IC→한남대교)로 나타났다. 이 외에도 특정도시나 특정도로의 일일 차량대수나 시간대별 교통량 등의 분석도 가능하다.

* 한 대의 차량이 10개 링크를 통행하였다면 10건으로 기록

〈표 12〉 ViewT 1.0 제공 서비스

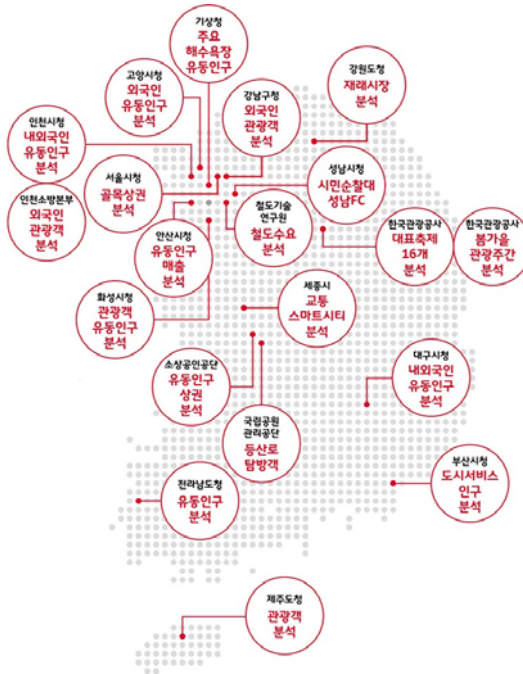
구분	제공 서비스	
주요분석기능	Selected Link Analysis	특정도로를 지나는 차량들의 진입/퇴출 경로 분석
	Time Maps	특정도로에서 30분/1시간 동안 이동 가능한 거리 분석
	사용자 혼잡구간선정	사용자가 설정한 혼잡기준에 따른 혼잡구간을 선정
	주요경로	동일한 출발기와 도착지에서 운전자가 어떤 경로를 주로 선택하는지 분석
	우리동네 교통지표	사용자가 바운더리를 설정하여 교통지표와 교통현상을 분석
	실시간 소통상황	현재 시간에 도로의 소통상황이 어떤지를 분석
지표분석	교통량	관측교통량, 추정교통량, 차량주행거리
	속도	평균속도, 혼잡시 평균속도, 정상시 평균속도
	교통혼잡지표	혼잡시간강도, 혼잡빈도강도, 혼잡기대강도, 교통혼잡비용

‘ViewT 1.0’에서 제공하는 분석기능에는 일반 시민들이 전국의 교통량을 파악하는데에 도움이 될 뿐만 아니라 지금까지 교통량 자료를 수집하지 않았던 지자체들이 해당 지역의 교통현황을 파악하는데 활용할 수 있는 여러 자료들이 포함되어 있다. 이처럼 민·관 교통 빅데이터 융합분석을 통해 도출한 다양한 결과들은 향후 효과적인 국가교통정책 수립을 위한 가치 있는 참고자료로도 활용될 수 있을 것으로 기대된다.

본고에 언급된 사례들 외에도 민·관 데이터 융합을 위한 다양한 시도들은 계속해서 이루어지고 있다. 먼저 서울시에서는 영세 소상공인들의 골목상권 창업에 도움이 되는 ‘우리마을가게’ 라는 상권분석서비스를 제공하고 있다. 해당 서비스를 통해 서울시는 시가 보유한 공공데이터에 BC카드, KB카드, 나이스 지니데이터, KT, 오픈메이트 등의 민간데

이터를 결합하여 서울시내 골목상권의 소비지출 동향을 분석하고 그 결과를 제공한다. 제주시에서는 전국 최초로 ‘지역 거점형 민·관 융합 데이터 서비스 표준모델과 포털’ 구축을 추진하고 있다. 해당 사업은 제주도가 공공데이터포털을 통해 공개하고 있는 교통, 관광 관련 공공데이터에 카카오가 보유한 생활기반 데이터를 융합하여 통합 플랫폼을 구축하는 방식으로 진행된다. 또다른 예로 안산시에서는 ‘민·관 빅데이터 융합을 통한 선제적 범죄 예방 지원 서비스 개발’의 일환으로 오픈메이트와의 협업을 통해 신규 CCTV 설치장소 위치를 도출하는 사업을 진행하였다. 민간기업인 SK텔레콤 역시 지자체 및 공공기관과 총 80여 건에 달하는 데이터 협업 프로젝트를 진행하며 민·관 빅데이터 융합분석 활성화에 기여하고 있다.

[그림 7] SKT-지자체·공공기관 데이터 협업 사례



출처: SKT '빅데이터 허브'

IV. 결론 및 향후 전망

4차 산업혁명 시대를 맞아 지난 2016년 12월 8일 우리나라 정부에서는 공공 및 민간의 빅데이터 활성화를 지원하기 위한 「민관 합동 빅데이터 TF」를 발족하였다. 그동안 각 부처에서는 독립적으로 빅데이터 관련 협의체를 운영해 왔으나 이와같은 빅데이터 TF가 구성됨에 따라 범정부적인 빅데이터 정책 추진이 가능하게 되었다. 빅데이터 TF에는 과학기술정보통신부, 행정안전부, 통계청이 참여하고 있으며 지난 2017년 8월 30일에는 “제 3차 민관 합동 빅데이터 TF 회의”를 개최한 바 있다. 특히 이번 제 3차 회의에서는 통계 기반의 빅데이터 산업 활성화 방안이 논의되었으며, 통계청이 보유한 공공데이터와 민간의 신용·위치정보 빅데이터를 연계해 구축한 가계부채 DB와 같은 빅데이터 연계사례들이 공유되었다.

이처럼 정부에서 직접 나서서 빅데이터의 활용과 민·관 데이터 융합에 주목하고 있는 이유는 다가오는 지능정보사회에서는 데이터 활용 경쟁력이 바로 국가 경쟁력으로 연결되기 때문이다. 정부뿐만 아니라 민간기업들 역시 외부와의 데이터 협력을 적극적으로 고려하고 있다. 데이터 융합을 통해 개별 기업이 보유한 빅데이터의 가치를 더욱 높일 수 있으며, 이 과정에서 얻게 되는 고도의 데이터분석 경험을 활용하여 성공적인 미래사업모델을 구축할 수 있기 때문이다.

이러한 측면에서 빅데이터 융합분석은 향후 다양한 산업분야에 걸쳐 지금보다 더욱 활발하게 이루어질 전망이다. 매일 새롭게 생성되는 데이터들로 넘쳐나는 빅데이터의 시대에는 이종간 데이터 결합이 데이터가 가진 잠재력을 최대한 이끌어낼 수 있는 가장 효과적인 방법 중 하나가 될 수 있기 때문이다. 각종 데이터들의 융합분석을 통해 빅데이터가 가지고 있는 무한대의 가능성을 발굴해내는 사례는 앞으로 점점 더 늘어날 것이며 그 방식 또한 지금보다 훨씬 더 다양해질 것으로 기대된다.

참고문헌

- Big Data Fades to the Algorithm Economy (Forbes, 2015.8.14.,
<https://www.forbes.com/sites/gartnergroup/2015/08/14/big-data-fades-to-the-algorithm-economy/#4b988f6b51a3>)
- Wikibon Big Data in the Public Cloud Forecast, 2016-2026 (Wikibon, 2016.5.31.)
- Data Age 2025: The Evolution of Data to Life-Critical (IDC, 2017.4)
- Google Trends (<https://trends.google.com>)
- Government at a Glance 2017 (OECD, 2017.7.13.)
- Statista (<https://www.statista.com>)
- The World Economic Forum : The top 10 emerging technologies for 2012 (<https://www.weforum.org/agenda/2012/02/the-2012-top-10-emerging-technologies/>)
- Worldwide Semiannual Big Data and Analytics Spending Guide (IDC, 2017.3)
- SKT, “빅데이터 허브’ 공개 4년…공공 활용 선도” (아이티데일리, 2017.6.30.,
<http://www.itdaily.kr/news/articleView.html?idxno=83945>)
- SKT 빅데이터허브(www.bigdatahub.co.kr/)
- 2011 국가정보화백서 (행정안전부, 2011)
- 2016년 빅데이터 시장현황조사 (한국정보화진흥원, 2017.3)
- 2016 빅데이터 스마트서비스 시범사업 결과보고서 (한국정보화진흥원, 2017.8)
- 공공데이터 개방 및 활용 주요 현황(17.10월) (공공데이터포털, 2017.12.1.)
- 공공데이터포털(<https://www.data.go.kr>)
- 네이버 데이터랩(<https://datalab.naver.com/>)
- 민·관 빅데이터 융합을 통한 해외유입 감염병 차단 본격화(미래창조과학부 보도자료, 2016.5.24.)

민관 빅데이터 융합을 통한 선제적 범죄예방 지원 서비스 개발(안산시, 2016.11)
서울시 우리마을가게 상권분석서비스 (<https://golmok.seoul.go.kr/>)
서울시 올빼미버스 제1회 지방자치정책 대상 수상(서울특별시청: 보도자료, 2016.11.25.)
'올빼미버스' 운행 50일, 하루 6천명 이용 (서울특별시청: 보도자료, 2013.11.15.)
'올빼미버스' 증차후, 하루 2천여 명 이용 증가, 혼잡도 완화(서울특별시청: 보도자료,
2017.1.16.)
우리나라에서 교통량 가장 많은 도로가 궁금하다면? (국토교통부 보도자료,
2017.9.4.)
제주도-카카오, 민관 융합 빅데이터 비즈니스 창출 (한라일보, 2017.7.5.,
<http://www.ihalla.com/read.php3?aid=1499232646569781073>)
빅데이터 산업의 현황과 전망 (KISTI MARKET REPORT, 2013.4)
빅데이터를 이용한 교통계획: 심야버스와 사고줄이기 (이신·허유경·김혜미, 서울시
교통정책 패키지, 2017.3)
빅데이터로 해외감염병 대응한다 (미래창조과학부 보도자료, 2016.11.15.)
차량운행경로 빅데이터의 도로정책 활용방안 (박종일, 국토정책 Brief, 2017.7.31.)
통계데이터로 빅데이터 산업 활성화 한다 (과학기술정보통신부·행정안전부·통계청
보도자료, 2017.8.30.)
한국교통연구원 View T 1.0 (<http://viewt.ktdb.go.kr>)