

미국과 유럽 매체의 '데이터저널리즘' 활용과 트렌드

SBS 신승이 기자/UC Berkeley 연수=

서론

데이터저널리즘은 데이터, 특히 빅데이터를 분석해 새로운 정보를 발굴해 보도하는 것으로, 단편적인 자료의 수집이나 일부의 증언만으로는 잘 드러나지 않는 장기적인 트렌드 또는 숨겨진 진실을 데이터 분석으로 찾아내는 과정이다. 또 이렇게 찾아낸 기사, 즉 스토리를 직관적으로 이해할 수 있는 표나 그래프, 인포그래픽(Infographic), 맵핑(mapping)기술, 때로는 독자들이 직접 참여할 수 있는 상호작용적(interactive) 요소의 시각적 방법으로 전달한다. 데이터저널리즘의 기본 과정은 결국 '데이터 획득과 수량화', '분석', 그리고 '시각화 기법을 통한 전달'로 요약할 수 있다. 데이터저널리즘은 과거 컴퓨터활용보도(CAR)나 과학적 정밀 보도와 유사하지만 데이터가 기사의 보완재가 아닌, 그 자체로서 기사의 원천이 된다는 점에서 결정적인 차이가 있다. 또한 누군가 배포한 정보에 단순히 반응하거나 일어난 뉴스를 다루는 것에서 벗어나 트렌드를 규정하고 중요한 사안에 새로운 정보를 찾아내 드러내는 보다 능동적인 과정을 데이터저널리즘이라고 규정할 수 있다(Dunham, 2019).

데이터저널리즘은 기존에 존재하지 않았던 정보를 새롭게 찾아내 드러내는 과정인 만큼 가장 독창적이고 독점적인 기사를 생산할 수 있게 하고 이를 통해 언론 매체의 경쟁력을 높여준다. 또한 익명의 취재원이나 자료에 기대던 관례와 달리 데이터와 분석 방법을 독자와 시청자에게 투명하게 공개함으로써 기사의 신뢰도를 높이고 기사의 폭발력을 키울 수 있다. 정파주의와 정치적 양극화가 어느 때보다 심하고 저널리스트들에 대한 공격이 거칠어지는 미디어 환경에서, 반박하기 어려운 객관적 분석 보도로서의 데이터저널리즘은 저널리스트들과 독자 양측으로부터 더욱 주목을 받고 있다.

2천년대 중반부터 데이터저널리즘이 본격적으로 시작된 지 약 15년이 지난 지금 데이터저널리즘은 어느 수준까지 확장되었고 최근 데이터저널리즘의 트렌드는 무엇인지, 또 데이터저널리즘 활성화를 위한 과제는 없는지 등을 미국과 유럽 매체들을 중심으로 살펴본다.

본론

1. 데이터저널리즘의 확대

데이터저널리즘은 미국과 영국 정부의 공공 정보 공개가 시작된 2천년대 중반부터 본격적으로 나타나기 시작했다. 현재 미국 정부의 공공 기록 공개 웹사이트인 'data.gov'에만 해도 거의 25만 개의 자료가 모여있고(The NYT/2019년 6월 12일 기사) 각종 소셜미디어와 기업들이 생산하는 기

록 등 현대 사회의 데이터 용량은 상상을 초월한다. 여기에 디지털 미디어 스토리지와 대역폭 가격의 하락, 스마트폰과 태블릿을 포함한 디바이스, 디지털 미디어의 급격한 증가 등이 빅데이터 전략과 구현을 급증하게 했다(김위근 외, 2019).

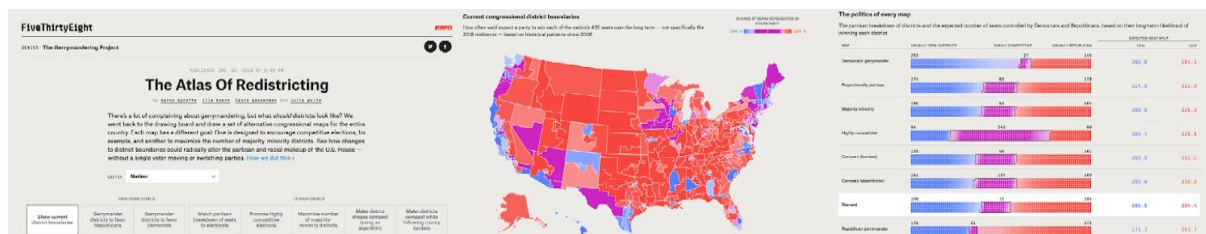
새로운 수입원과 독자 창출. 영향력 회복이 필요한 미국과 유럽의 기성 언론매체들은 발 빠르게 데이터저널리즘을 도입했다. 영국 '가디언(The Guardian)'의 '데이터블로그(Data blog)', 미국 '뉴욕 타임스(The New York Times)'의 '업샷(Upshot)', '월스트리트저널(The Wall Street Journal)', '영국 BBC' 등이 일찌감치 데이터 전문 기사 섹션을 갖추고 매일 관련 기사를 생산하고 있다. 이제 데이터저널리즘의 기술은 숫자로 이뤄진 통계만이 아니라 그림, 비디오, 오디오, 책 등에서도 텍스트마이닝(text mining)을 이용해 새로운 정보를 찾아내는 수준으로 발전했고, 더욱 혁신적인 시각화 기술을 도입, 전달력을 높이는 한편 젊은 독자들을 끌어들이고 있다.

구글(Google)의 최근 조사에 따르면 이미 조사대상의 42% 기자들이 정기적으로 기사에 데이터를 쓰고 있고 미국과 유럽 뉴스 매체의 51%는 데이터저널리즘 담당자가 있으며 디지털 뉴스 회사의 경우 이 비중이 60%에 이르는 것으로 나타났다(Rogers, 2017). 비영리단체인 저널리스트국제연맹 ICFJ(International Center for Journalists)가 지난해 전세계 149개국 저널리스트 4,100여명을 대상으로 한 또다른 설문조사에서는 응답자 61%가 적어도 일주일에 한 번 이상 데이터로 기사를 쓰고 있는 것으로 나타났다. 일반 기자의 65%, 데스크 등 보도국 간부의 55%가 '데이터저널리즘이 독자 또는 시청자를 더 유입하는 수단이 되는 동시에 기사의 질과 생산성을 향상시킨다'고 믿고 있었다. 데이터저널리즘의 효용성과 중요성에 대한 저널리스트들의 인식 확대, 다양한 분석 및 시각화 툴(tool)의 개발과 보급, 빅데이터 환경 등이 더해지면서 데이터저널리즘 시장은 점점 커지고 있다. 재정적 어려움에 직면해 있는 미디어 산업에서 거의 유일하게 끊임없이 투자되고 성장하는 분야(Weber, Kosterich, 2018; Zamith, 2019)라는 진단이 나오는 이유다.

2. 데이터저널리즘의 활용예

(1)선거

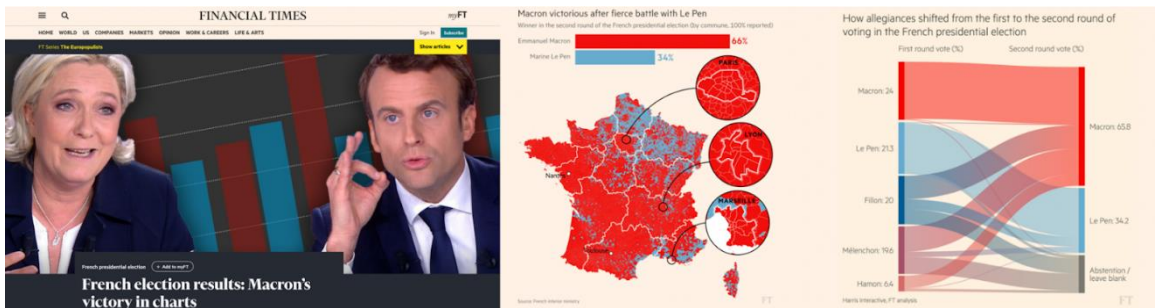
#'파이브서티에이트'의 <게리맨더링 분석>... '과학적, 실증적 접근'



FiveThirtyEight, 2018년 1월 25일/ <지역구 재설정 지도 (The Atlas Redistricting)>

2018년 데이터저널리즘상에 출품된 미국 데이터 분석 전문 웹사이트 '파이브서티에이트(FiveThirtyEight)'의 <지역구 재설정 지도(The Atlas of Redistricting)>는 '게리맨더링(특정 정당이나 후보에게 유리하게 지역구 범위를 바꾸는 것)'이 실제로 선거 결과에 어떤 영향을 주는지를 데이터저널리즘 방식을 통해 독자들이 직관적으로 파악할 수 있도록 한 보도다. 민주당과 공화당에 각각 유리하게 선거구를 나눌 경우, 경합 지역을 많이 배치했을 경우 등 다양한 시나리오에 따라 독자가 직접 선거구를 나눠볼 수 있게 하고, 그 결과 어떤 정당이 더 많은 의석을 차지하는지 확률에 따라 색깔별로 표시했다. 약간의 선거구 획정 변화가 특정 정당의 과대 대표 또는 과소 대표를 가져올 수 있다는 것을 독자가 쉽게 확인할 수 있다. 이 작업에는 과거 두차례 미국 대통령 선거의 각 선거구별 개표 결과와 인종 비율 등의 정보가 포함된 인구 센서스 자료 데이터가 사용됐다. 소수 선거구의 결과 또는 전문가의 지적만을 앞세워 게리맨더링 문제를 지적해 온 일반적인 보도가 자칫 추상적인 개념 비판에 머물 수 있는 것과는 달리 훨씬 더 정밀하면서도 독자들에게 강력한 메시지를 전달할 수 있었다.

#'파이낸셜타임스'의 <프랑스 선거 결과 분석> ... 속도에 깊이를 더하다'

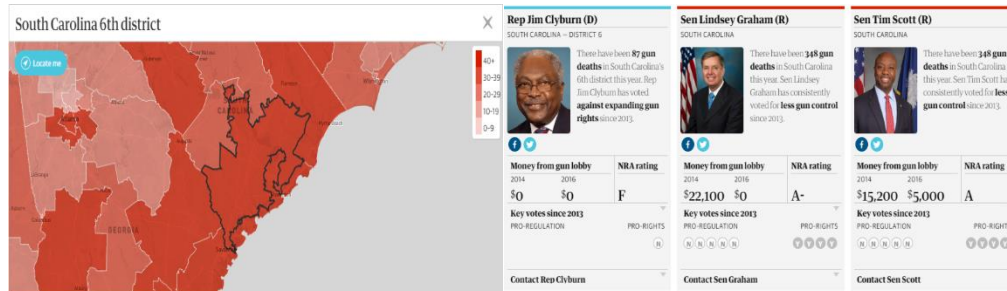


FT 2017년 5월 9일/ <프랑스 선거결과: 표로 보는 마크롱의 승리(French election results: Macron's victory in charts)>

영국 파이낸셜 타임스(The Financial Times)의 <프랑스 선거 결과: 표로 보는 마크롱의 승리(French election results: Macron's victory in charts)>는 빅데이터 분석 기술을 통해 깊이 있는 선거 보도를 가능하게 했다는 평가를 받는다. 지역적 특성과 함께 연령, 교육 수준, 기대수명 등 여러 지표를 대입해 투표자 특성과 지지 후보와의 상관관계를 분석해 내고 1차 투표와 2차 투표 때 두 후보에 대한 지지가 어떻게 이동했는지 등을 단순하면서도 직관적인 그래프를 통해 표현했다. 분석 결과를 토대로 '소득과 교육 수준, 기대 수명이 높고 진보 성향을 가진 그룹들의 폭넓은 지지로 마크롱 후보가 당선될 수 있었다'는 결과를 도출했다. 선거 결과가 여러 변수에 의해 복잡적으로 좌우된다는 점을 고려할 때, 다양한 데이터를 접목해 결과를 분석하는 데이터저널리즘 기법은 선거분야에서 특히 빛을 발할 수 있다. 데이터저널리즘에 있어 데이터의 발굴 못지 않게 불필요한 데이터를 정리해 내는 작업이 관건인데, '글로벌 에디터스 네트워크(GEN)'은 "정확한 데이터 수집(scraping)과 불필요한 데이터를 정리하는 데이터 청소(cleaning)에 최적화된 앱 기술을 사용한 점, 해당 지역의 학계 전문가, 지역 특파원과 협업해 데이터 분석에 오류가 없도록 노력한 점 등이 눈에 띈다"며 위 보도에 데이터저널리즘 상을 수여했다. 차별화된 데이터저널리즘 시도 덕분에 파이낸셜타임스는 유럽 각종 선거 결과를 보도하면서 매체의 지역 기반인 영국과 미국에서보다 유럽 다른 국가들로부터 더 많은 독자를 유입할 수 있었다.

(2)정치인, 정책

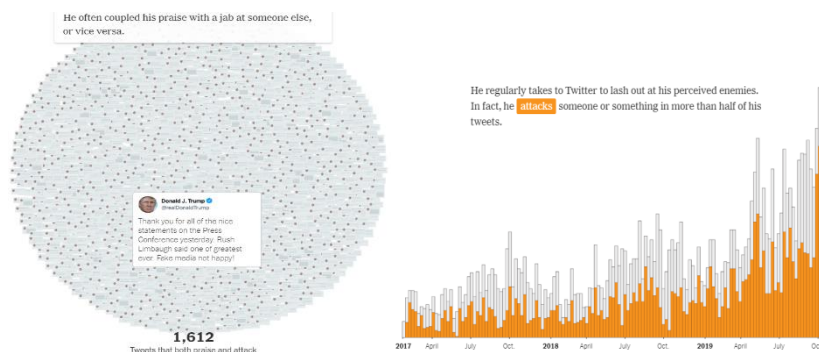
#'가디언' <총기사고.. 의회는 뭐했나?>...'입법활동과 후원금 해부'



The Guardian, 2015년 12월 14일 / <여러분 지역의 총기 사망: 당신이 뽑은 대표는 무엇을 했습니까(Gun deaths in your district: what have your elected representatives done?)>

가디언 미국 본부(Guardian USA)가 2015년 발표한 기사 앱은, 미국 연방의회 의원들이 대형 총기 사고 예방을 위해 어떤 입법 노력을 벌였는지, 그리고 그 결과가 총기협회의 로비와는 어떤 연관이 있는지를 파헤친 보도다. 독자들은 자신의 지역구 의원이 총기 관련 입법을 몇 건이나 했는지, 관련 법안에 어떤 표를 행사했는지, 또 총기협회로부터 어떤 평점을 받고 얼마만큼의 후원금을 받았는지 등을 한 명 한 명씩 확인할 수 있다. 이 매체는 이를 위해 비영리단체인 '총기폭력 자료(Gun Violence)'의 총기 관련 사상자 자료와 함께 후원금, 투표 기록 등을 다각도로 찾아 매칭시켰다. 앱(App)을 통해 독자가 즉석에서 해당 의원에게 전화를 걸거나 트위터로 글을 보낼 수 있게 해 독자들의 참여를 독려하는 한편 기사의 영향력을 확대했다. 미국의 경우 대부분의 의회 활동과 후원금 모금 내역이 투명하게 공개되고 디지털 자료로 보관되어 있음에도 유권자들이 이를 하나하나 직접 감시하기는 어렵다. 이런 측면에서 의회의 정치활동을 모니터하는 것 역시 데이터저널리즘이 큰 가능성을 보이는 영역이다.

#'뉴욕타임스'의 <트럼프의 트위터 정치>..공인에 대한 새로운 검증 방식

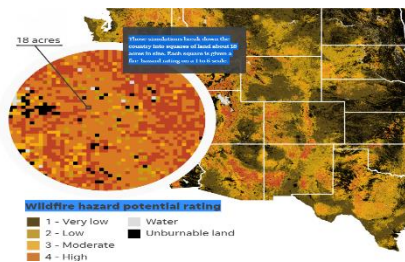


The New York Times 2019년 11월 2일/ <트럼프는 11,000개 넘는 트윗으로 대통령이란 자리를 어떻게 바꿨나(How Trump reshaped the presidency over 11,000 Tweets)>

'뉴욕타임스(The New York Times)'의 2019년 11월 2일자 데이터 기사는 트럼프 미국 대통령이 정치 무기로 삼고 있는 트위터 내용을 해부한 기사이다. 취임 후 2년 9개월 동안 트럼프 대통령이 올린 11,000개 이상의 트윗과 리트윗(retweet)을 빅데이터 분석 기술로 분석했다. 트위터의 절

반 이상이 민주당, 언론, 사법부 등을 공격하는 내용이었고, 선거를 앞두고 민주당의 탄핵 절차가 시작될 무렵 가장 많은 트위터 글을 올렸으며 '폭스뉴스(Fox news)'등 보수 매체의 내용을 그대로 옮긴 트윗이 전체의 15%에 달한다는 내용 등 다양한 각도의 문제 의식이 기사에 담겨 있다. 데이터 저널리즘은 더이상 단순히 숫자 데이터에만 머물지 않고 단어, 문장 등 글의 내용을 분석해 경향성을 파악하고 있다. 특히 많은 정치인들이 소셜미디어를 이용해 사실을 왜곡하거나 무리한 정치 홍보 수단으로 사용하는 경우가 늘어나면서 내용의 오류, 가짜 정보, 또는 정치인으로서의 성향, 자질 등을 검증하기 위한 방법으로 데이터 저널리즘이 유용하게 쓰일 수 있다.

(3) 환경, 재난 사고



The Arizona Republic 2019년 7월 22일/ <다음 대 화재는 어디서 일어날까? 위험은 도처에 있다 (Where will the West's next deadly wildfire strike? The risks are everywhere)>

'아리조나 리퍼블릭(The Arizona Republic)'과 'USA투데이(the USA TODAY)'가 협업한 "화재의 전조(Ahead of the Fire)" 보도는 재난 수준의 화재 문제를 '우리 동네'와 연결시켜 독자들에게 경각심을 심어준 데이터저널리즘의 예이다. 캘리포니아는 매년 대형 산불이 점점 더 빈번하게 발생하는데, 화재 점검 시스템, 구조 인력 현황, 노약자 비율 등에 따라 피해 규모에 큰 차이가 난다. 이 신문은 이 지점에 착안해 각 지역의 지리적 특성과 함께 대피 경로, 인구 구성 등의 정보를 담은 인구센서스 데이터, 그리고 미국 산림청 화재 데이터를 접목해 내가 살고 있는 지역의 대화재 피해 위험이 얼마나 되는지를 수치로 계산, 인포그래픽 등으로 실증적으로 보여줬다. 분석 결과 지난 2018년 대화재로 엄청난 인명과 재산피해를 낸 캘리포니아 파라다이스(Paradise) 지역보다 잠재적 화재 피해 가능성이 더 큰 지역이 서부지역 11개주(州) 526개 마을에 이른다는 것을 찾아냈다. 이 보도는 지난해 우수한 사회과학적 방법 보도에 수여하는 필립마이어 상(Philip Meyer Award)을 수상했는데 당시 정밀한 분석뿐 아니라 '뛰어난 그래픽과 생생한 현장 사진 등도 겸비해 환경 분야 보도에 본보기가 될 수 있다'는 평가를 받았다.

3. 새로운 트렌드

(1) 자동화(automation) 기술

비영리단체인 국제저널리스트네트워크(IJN)는 2019년 데이터저널리즘의 두드러진 트렌드 중 하나로 자동화(automation)기술과 기계학습(machine learning) 기술의 확산을 들었다. 예를 들어 독일 '슈피겔(Der Spiegel)'의 <외국인을 위한 집은 없다(Wir müssen draußen bleiben)> 기사는 자동

화 기술을 이용해 독일 주택시장에 만연해 있는 '차별' 문제를 짚어냈다. 전형적인 독일인 이름과 외국 이민자의 이름, 그리고 단기 일자리와 안정적인 일자리를 가진 사람 등 각각의 경우에 해당하는 가상의 인물을 만들고, 전국 10개 도시에서 나오는 임대 광고 7천여건에 대해 2만 통 이상의 지원서를 작성해 보낸 뒤 그 결과를 분석했다. 컴퓨터 자동화 프로그램을 이용해 30분에 한번씩 새로운 주택 임대 광고를 찾아내고 각각의 지원서를 적절한 비율로 작성, 발송하도록 했다. 답변을 분석한 결과 외국인은 유창한 독일어를 사용하더라도 독일인에 비해 집주인으로부터 긍정적인 피드백을 받지 못하는 비율이 훨씬 높게 나타나는 등 출신 지역에 따른 차별이 광범위하게 존재하는 것으로 드러났다. 저널리스트들은 답신 자료에 오류나 스팸 메일이 없는지 등을 걸러내고 자료 해석에 집중하는 반면, 대량의 광고 수집과 자료 발송 과정은 자동화 기술에 맡김으로써 독일 전역에 걸친 대규모 프로젝트가 짧은 시간 안에 가능했다.

(2) 글로벌 협업 프로젝트

국경과 매체를 뛰어넘는 협업 프로젝트도 최근의 트렌드로 꼽힌다. 가장 대표적인 예로 탐사기자 자국제연맹(ICIJ)이 주도한 <임플란트 플라이즈(Implant Flies)> 보도를 들 수 있다. 전 세계 36개국 250여명의 저널리스트들이 참여한 프로젝트로, 제대로 검증되지 않았거나 부실 검사 받은 각종 의료 장치 이식 때문에 얼마나 많은 피해가 전세계적으로 발생했는지를 파헤친 보도다.

또 올해 데이터저널리즘 상인 'SIGMA AWARDS' 선정작 중 하나로 뽑힌 '조직적범죄와 부패 범죄 보도 프로젝트(OCCPR)'의 <트로이카 돈 세탁기(The Troika Laundromat)> 역시 국경을 뛰어넘은 대규모 협업 프로젝트로 거론된다. 굴지의 국제 기업들의 탈세, 비자금 의혹을 파헤치기 위해 15개 나라에서 발생한 약 3천개 기업과 은행들의 거래를 조사했는데 그 결과 과거 7년 동안 26조 유로 이상의 수상한 거래가 있었음을 파악해 폭로했다. 이 프로젝트에는 유럽과 아프리카, 아시아, 중동 라틴아메리카, 미국에 이르기까지 거대한 네트워크 협업이 이뤄졌다.

디지털 시대 이전에는 모든 뉴스 매체와 각 기자들이 개별적으로 경쟁하는 환경이 당연했지만 이제는 각 매체와 플랫폼, 그리고 국경까지 뛰어넘는 협업 보도가 점점 늘고 있다. 이런 협업 프로젝트는 재정적 어려움에 처해 있으면서 어느 때보다도 높은 수준의 탐사보도 기능을 요구 받는 언론 매체들에 훌륭한 도구로 활용될 수 있다. 각 매체가 치러야 하는 비용을 절감할 뿐 아니라 훨씬 더 풍부한 기사를 생산하고, 더 강력한 어젠다 설정 기능을 수행할 수 있으며, 무엇보다 전 지구적 차원의 복잡하고 어려운 문제의 실마리를 풀어나가는데 큰 도움이 될 수 있기 때문이다 (Carson, Farhall, 2018)

(3) 일상적 기사에 데이터 적용

데이터저널리즘은 규모가 크고 오랜 시간이 필요한 추적보도, 탐사보도에 주로 접목돼 왔지만, 보다 편리하게 데이터를 분석하고 시각화할 수 있는 툴(tool)이 보급되고, 무엇보다 저널리스트들의 데이터 친숙도가 높아지면서 정부 발표나 사건사고 같은 일상적인 기사에도 데이터 적용이 점점 확대되고 있는 추세다. 미국 미네소타주의 지역 신문인 '스타 트리뷴(The Star Tribune)'은 데이터저널리즘 웹사이트인 'datajournalism.com'에 관련 예를 소개하고 있다.

“2019년 미네소타 교육당국은 고교졸업성적 결과를 발표하면서 학생들의 전반적인 성적 이 향상됐다고 발표했다. 대부분의 언론이 당국의 발표를 받아 기사를 썼지만 이 신문의 한 교육전문 기자는, 장기적인 관점에서 성적이 향상됐다는 발표는 사실과 다르다고 반박 보도했다. 매년 공개 되는 성적 데이터를 모아 꾸준히 트렌드를 분석해 오고 있었기 때문에 가능한 일이었다. 또 이 데이터를 통해 점점 더 많은 지역 학생들이 정부가 주관하는 표준 시험을 아예 치르지 않고 있다는 사실도 파악할 수 있었다. ”

엑셀(Excel)이나 스프레드시트(Spreadsheet)같은 기초적인 프로그램을 활용해 꾸준히 관심 분야의 데이터를 축적하고 확인해 데이터베이스를 확보하면 데일리 뉴스의 디테일을 보완하고 생명력을 불어넣을 수 있으며 무엇보다 시청자와 독자가 더 큰 그림을 볼 수 있게 할 수 있다(Webster, 2019). 또 대형 사건사고가 일어난 지 일정 기간 후 정부의 문제 개선 약속이 얼마나 이행되었는지 등도 통계로 점검할 수 있다. 기사의 주제에 맞는 적절한 취재 장소나 대상을 물색할 때도 이와 같은 데이터분석을 이용해 가장 적절한 곳을 찾을 수 있다.

4. 데이터저널리즘 본격 도입을 위한 과제

(1) 뉴스 매체의 편차, 고급 기술의 활용은 아직

2019년 전 세계 데이터 저널리스트들을 대상으로 한 조사에서도 데이터저널리즘 활용 도구로 스프레드시트를 사용한다는 응답이 78.7%였던 반면 프로그래밍 언어(39.4%)나 GIS 또는 네트워크 분석(23.7%), 통계 분석이나 기계학습(8.2%)을 사용한다는 사람은 매우 적었다(Camaj, 2019). 즉, 상당수 뉴스 매체의 데이터저널리즘이 아직은 기본적인 단계에 머물러 있고 최신의 데이터저널리즘을 구사하지 못하고 있다는 것을 알 수 있다(Heravi, 2017). 특정 온라인 전문 매체와 대형 미디어 기업들을 중심으로 데이터저널리즘이 확대되고 있다 보니 전체적인 데이터저널리즘 기사 비율도 아직은 비교적 낮은 수준이다. 데이터저널리즘이 가장 활발히 시행되고 있는 영국조차도 가디언, 데일리메일, 인디펜던트 등 15개 신문 기사 가운데 약 6~22%만이 데이터 관련 뉴스 콘텐츠인 것으로 파악된다(Knight, 2015; 임종섭, 2019). 국내언론사의 사정도 다르지 않다. 데이터저널리즘을 아직 뉴스 제작에 활발하게 응용하지 못하고 있는데, 무엇보다 기자들과 IT 기술 인력의 협업이 부족하고 언론사 차원에서 투자가 어려우며 데이터를 정리하는데 큰 작업이 필요하기 때문으로 파악된다(임종섭, 2015; 신동희, 김장현, 최명군, 2013)

(2) 교육의 부재 - 전 보도국 교육에 앞장선 NYT

한 조사에 따르면 (Whibey, 2019) 미국의 저널리스트들 가운데 데이터분석이 매우 중요하다고 느끼는 저널리스트들이 80%에 달했지만 스스로 분석 능력을 잘 갖췄다고 생각하는 비율은 25%에 불과한 것으로 나타났다. 지난해 ICFJ(International Center for Journalists)의 조사에서도 79%의 저널리스트들이 그 어떤 교육보다 데이터분석 교육을 받고 싶어했지만 해당 교육을 제공하는 매체는 35%에 불과했다. 심지어 데이터저널리스트라는 직함을 가진 기자들 중에서도 관련 지식의

부족을 토로하는 경우가 적지 않아, 전세계 43개국 기자들과 데이터저널리스트들을 대상으로 한 데이터저널리즘 설문 조사 결과, 이들이 가장 어려움을 겪는 것은 자료의 부족(52%)과 함께 전문 지식의 부족(44%)인 것으로 조사됐다. 86%가 자신이 데이터 저널리스트라고 생각했지만, 정작 데이터저널리즘의 숙련도 측면에서는 18%만이 자신감을 보였다. 그럼에도 데이터저널리즘 관련 교육을 받은 경우는 절반에 그쳤고 특히 데이터 분석, 통계, 코딩, 기계학습, 데이터 시각화 같은 정규 교육 비율은 매우 낮았다.

이런 환경에서 '뉴욕타임스'가 최근 전 뉴스룸을 대상으로 데이터저널리즘 교육에 나선 것은 큰 시사점을 갖는다. 뉴욕타임스는 2018년부터 뉴스룸 기자들과 데스크들을 대상으로 본격적인 데이터저널리즘 교육을 실시해 오고 있다고 밝혔다. 석 달 과정의 훈련프로그램과 집중적인 부트캠프를 통해 데이터 분류, 검색, 필터링 같은 기초과정에서부터 데이터 청소 같은 고급 기술에 이르기까지, 또 데이터저널리즘 윤리와 데이터에 적합한 기사 구성 방법 등도 커리큘럼에 포함해 교육을 진행했다. 약 1년 반 동안 이 교육을 이수해 데이터 관련 기사를 쓰기 시작한 일반 취재기자가 60명 이상이라고 한다. '업샷(Upspot)'이라는 데이터 전문 웹 페이지를 운영하며 이미 세계 최고 수준의 데이터 저널리스트들과 그래픽 전문가, 관련 팀을 갖추고 있음에도 불구하고 뉴욕타임스가 일반 기자들을 대상으로 광범위한 교육 프로젝트를 시작한 것은, 더이상 데이터저널리즘이 특정 영역, 특정 전문가에게만 국한되지 않는다는 것을 의미한다. 사회 모든 분야에서 정보가 더 복잡해지고 맥락을 더욱 빨리 알고 싶어하는 뉴스 이용자의 수요가 높아지면서 이제 제너럴리스트(generalist)로서의 기자들을 위한 공간은 거의 없어졌으며(Whibey 2017; Anderson, Bell, Shirky, 2014) 데이터에 대한 비판적 사고와 분석, 전달 능력 없이는 수준 높은 기사 생산이 어렵다고 판단한 것으로 볼 수 있다.

(3) 투명성과 뉴스 소비자의 참여의 향상

데이터저널리즘을 일반 저널리즘과 구분 짓는 중요한 요소 가운데 하나가 바로 '참여를 전제로한 공개성(participatory openness)'과 '오픈 데이터, 오픈소스의 일상화'라고 할 수 있다(Zamith, 2019). 일방적으로 내용을 전달하는게 기존의 일반 뉴스 방식이라면, 데이터저널리즘은 이용자들의 반응을 토대로 수정되는 과정을 거친다. 이용자가 인터랙티브 그래픽을 직접 확인하고 언론사가 제공하는 데이터를 언제든지 내려받아 분석하는 비선형의 방식이 가능하다. 즉 이용자들이 사용자라는 객체로 머물지 않고 뉴스 생산자의 역할까지 한다는 것이 중요하다(임종섭, 2015). 이 때문에 자료의 출처와 원자료 전체에 대한 접근 링크, 그리고 조사방법론, 협업 기관 등을 명시하는 것도 데이터저널리즘이 지향하는 투명성의 중요한 요소가 된다. 그러나 아직 현실은 데이터저널리즘의 중요 요소인 투명성, 공개, 오픈소스 등이 충분히 적용되지 못하고 있다. 미국의 '뉴욕타임스'와 '워싱턴포스트'의 2017년 상반기 데이터저널리즘 기사(159건)를 분석한 결과 87.4%의 기사에서 원 데이터(raw data)가 드러나지 않았고 기사 중 16.4%에서만 데이터 출처나 데이터 수집 방법 등이 제대로 명시된 것으로 나타났다(Zamith, 2019). 한국의 주요 방송뉴스(SBS, KBS, MBC, JTBC)와 미국의 주요 방송뉴스(ABC, NBC, CBS, CNN)를 비교한 또다른 연구에서도 양쪽 미디어 모두 주로 정부 자료에 의존해 데이터 기사를 작성했고, 원 데이터를 함께 공개하는 일 역시 거의

없었다(임종섭, 2019). 마찬가지로 2019년 22개국 데이터 저널리스트 61명을 대상으로 한 설문 조사에서는 37.9%의 응답자만이 데이터 클라우드소싱을 위해 온라인 플랫폼을 이용해 봤다고 답하는 등 데이터의 원천이 대부분 정부기관으로 나타났다. 데이터저널리즘이 지향하는 투명성과 참여 또는 더 많은 상호작용 등은 언론매체에 대한 신뢰 하락 문제를 해결할 수 있는 방안으로 여겨진다(Zamith, 2019; Usher 2018). 따라서 데이터저널리즘이 지향하는 순기능을 적극적으로 활용하기 위해서는 데이터에 대한 접근을 확대하고 이용자와의 상호 작용을 더 활발히 하며 다양한 출처의 데이터를 입체적으로 활용하려는 노력이 필요해 보인다.

결론

2018년 갤럽이 미국인들을 대상으로 실시한 여론조사에서 매스 미디어에 대한 신뢰도는 의회(Congress) 다음으로 가장 낮은 것으로 조사됐다(Gallup, 2018년 6월 1-10일 조사). 응답자들은 미디어를 신뢰하지 않는 이유로 '부정확성', '편견에 치우친 보도', '당파성' 등을 꼽았다. 소셜미디어의 확산과 에코챔버(echo chamber) 현상, 정치적 양극화(polarization) 심화 등과 맞물리면서 기성 뉴스 미디어에 대한 신뢰 추락 속도는 점점 더 빨라지고 있으며 우리나라 역시 상황이 크게 다르지 않다.

데이터에서 입각한 과학적 보도, 검증 가능한 열린 보도로서 데이터저널리즘은 독창적, 독점적 기사 생산 측면뿐 아니라 언론의 신뢰 회복을 위한 중요한 가능성을 보여준다. 실제로 많은 언론인들이 데이터저널리즘을 '좀 더 체계적이고 정확하며 신뢰할 수 있는 저널리즘'이라고 여기고 있으며(Borges-Rey 2017) 데이터저널리즘을 표방한 기사들의 양도 최근 10여년 사이 급격히 늘고 있다.

이제 데이터저널리즘은 정치와 경제, 사회, 사건사고, 스포츠에 이르기까지 데이터가 축적된 거의 모든 영역에서 활용되고 있으며 대규모 탐사 보도 프로젝트뿐 아니라 일상적인 발표, 사건사고 기사에도 다양하게 적용되고 있다. 갈수록 데이터의 양이 증가하고 인공지능(AI) 기술과 데이터 수집·분석 앱, 그리고 각종 시각화 기술 등이 발전하면서 데이터저널리즘 기사의 생산 속도는 점점 빨라지고 분석 수준은 더 향상되고 있다.

그러나 한편으로는 아직 데이터저널리즘에 대한 언론 현장과 학계의 이해, 연구는 제대로 정착되지 못했다. 뉴스 매체의 규모에 따라 기사의 질적 편차가 심하고 방향성이 모호한 경우도 적지 않으며 무엇보다 데이터저널리즘이 추구하는 '투명성'과 '참여'의 가치도 충분히 구현되지 못하고 있다. 특히 뉴스 사용자와 보도국 내부에서 데이터저널리즘 기사에 대한 수요가 급격히 커진 반면, 전문 인력의 확보나 저널리스트들의 협업 시스템 구축, 보도국 교육 같은 투자가 미비한 것도 해결해야 할 과제이다. 덧붙여 데이터저널리즘 역시 저널리즘을 구현하는 한 방식이라는 점에서 자칫 현장과 동떨어진 기교적 기사에 치우치지 않도록 다양한 취재원의 이야기를 직접 듣고 확인한 뒤 이를 함께 반영하려는 다각적인 노력이 필요하다는 지적도 있다.

참고 문헌 및 웹사이트

1. Megan Knight(2015) Data Journalism in the UK: a preliminary analysis of form and content
2. Stephen Cushion, Justin Lewis, Robert Callaghan (2017) Data Journalism, Impartiality and Statistical Claims
3. Linda Camaj, Jason Martin, Gerry Lanosga (2019), Global Initiatives in Data Journalism in the Age of Big Data: Challenges and Opportunities
4. Richard Scott Dunham (2019) 'Multimedia Reporting: How Digital Tools Can Improve Journalism Storytelling)
5. 뉴욕타임스 2020년 6월 12일자 기사 'How 5 Data Dynamos Do Their Jobs'
<https://www.nytimes.com/2019/06/12/reader-center/data-reporting-spreadsheets.html>
6. 김위근, 류민호, 권오성 (2019) 컴퓨터이셔널 저널리즘과 뉴스룸 트랜스포메이션
7. Simon Rogers (2017) A new data journalism handbook
8. Matthew S. Weber &Allie Kosterich (2017) Coding the News: The role of computer code in filtering and distributing news
9. Rodrigo Zamith (2019) Transparency, Interactivity, Diversity, and Information Provenance in Everyday Data Journalism
10. 파이브서티에이트(FiveThirtyEight) 2018년 1월 25일 기사 'The Atlas Of Redistricting'
<https://projects.fivethirtyeight.com/redistricting-maps/>
11. 2018년 데이터저널리즘 수상작 <https://medium.com/data-journalism-awards>
12. 파이낸셜타임즈(The Financial Times) 2017년 5월 9일 기사 'French election results: Macron's victory in charts' <https://www.ft.com/content/62d782d6-31a7-11e7-9555-23ef563ecf9a>
13. 가디언(The Guardian) 2015년 12월 14일 기사 'Gun deaths in your district: what have your elected representatives done?' <https://www.theguardian.com/us-news/ng-interactive/2015/dec/14/gun-control-laws-congress-shooting-deaths-nra-lobby-campaign-donations>
14. 뉴욕타임스(The New York Times) 2019년 11월 2일 기사 'How Trump reshaped the presidency over 11,000 Treests'
<https://www.nytimes.com/interactive/2019/11/02/us/politics/trump-twitter-presidency.html>
15. 아리조나 리퍼블릭(The Arizona Republic) 2019년 7월 22일 기사 'Where will the West's next deadly wildfire strike? The risks are everywhere' <https://www.azcentral.com/in-depth/news/local/arizona-wildfires/2019/07/22/wildfire-risks-more-than-500-spots-have-greater-hazard-than-paradise/1434502001>
16. 2019년 필립마이어(Phillip Meyer Awards)상 수상작 <https://www.ire.org/awards/philip-meyer-awards/2019-philip-meyer-award-winners>
17. Marianne Buechart (2019) Key trends for data journalism in 2019: machine learning, collaborations, and code art (<https://datajournalismawards.org/2019/01/17/key-trends-for-data-journalism-in-2019-machine-learning-collaborations-and-code-art/>)
18. 슈피겔(Der Spiegel) 2017년 6월 22일 기사 'No Place for Foreigners'
<https://www.hanna-und-ismail.de/english/methodology.html>
19. OCCRP 2019년 연속 기사 'THE TROIKA LAUNDROMAT' <https://www.occrp.org/en/troikalaundromat/>
20. Maryjo Webster (2020) Bringing the power of data to deadline stories <https://datajournalism.com/read/longreads/how-to-bring-the-power-of-data>
21. Bahareh R. Heravi (2017) Teaching Data Journalism
22. 임종섭 (2019) Representation of data journalism practices in the South Korean and US television news
23. 임종섭 (2015) 데이터저널리즘
24. John Wihbey, Mark Coddington (2017) Knowing the Numbers: Assessing Attitudes among Journalists and Educators about Using and Interpreting Data, Statistics, and Research
25. Gallup poll(2018.6) <https://news.gallup.com/poll/236243/military-small-business-police-stir-confidence.aspx>
26. Eddy Borges-Rey (2017) Data Journalism Sustainability/ An outlook on the future of data-driven reporting