

혀낸 의사 존 스노우의 수도 지도(그림1)를 데이터 저널리즘의 효시로 보기도 한다. 존 스노우는 당시 오염된 템즈강의 하류에서 물을 끌어오는 회사의 수도라인을 이용하는 구역에서 콜레라가 창궐했다는 것을 밝혀냈고, 당시 그가 지도에 상수도회사의 파이프라인과 해당 주소에서 사망한 사람들의 숫자를 그래프로 나타낸 그림은 2개의 서로 다른 데이터(수원이 다른 상수도 회사의 공급처 - 콜레라로 인한 사망자)를 겹쳐놓음으로써 감춰진 진실을 발견해낼 수 있었다.

이렇듯 데이터저널리즘 보도는 데이터가 핵심이 되는 기사를 가리킨다. 기사를 뒷받침하기 위해 부수적으로 통계를 사용하는 것이 아니라 - 뒷받침하기 위해 입맛대로 골라내는 통계가 적지 않은 왜곡을 불러온 사례를 우리는 익히 알고 있다 - 2개 이상의 데이터들을 겹쳐놓았을 때 나타나는 진실을 보여주는 것이 데이터저널리즘이다. 때문에 미국의 저명한 데이터저널리스트들은 대부분 탐사 기자다. 폴리처상을 받은 뉴욕타임즈의 데이터 기사사라 코헨은 “차고에서 은밀하게 속삭이며 나쁜 짓을 하는 사람들은 이제 찾아보기 힘들다”며 “이제 부조리와 기만적 행태는 대단히 시스템 안에서 보

다 경험에 익숙한 형태로 이뤄지고 있다”고 지적한다. “사람을 쫓아다니며 파헤치는 탐사보다 시스템 안에서 수집되고 있는 데이터를 분석하는 일에 더 집중해야 한다”는 것도 같은 맥락이다. 실제로 그렇다. 현실에서 사과상자를 차 트렁크에 넣어주는 현장을 적발하는 것은 하늘에 별 따기만큼 힘들다. 그러나 정치자금 거래 내역을 분석한다면 얘기는 달라진다는 뜻이다.

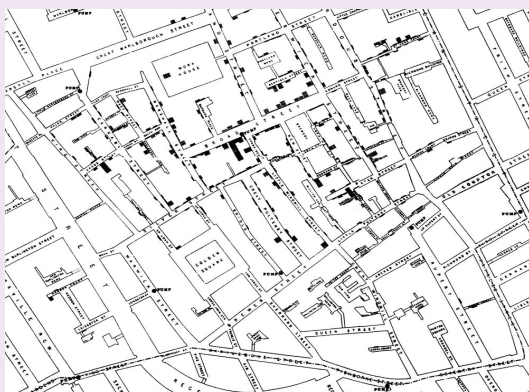
2. 관건은 데이터 확보... 어디 좋은 방법 없나요?

결국 핵심은 데이터다. 구하고자 하는 데이터가 무엇인지, 어디에 있는지, 어떻게 얻을 수 있는지를 빨리 파악할수록 결승선까지 가는 시간은 단축된다. 해마다 IRE(전미탐사보도협회) 등 기자들이 여는 탐사보도, 데이터보도 컨퍼런스에서 가장 인기를 끄는 강좌는 “어떻게 데이터를 수집할 것인가”인 것을 보면 국가를 막론하고 데이터 기자들의 숙명은 비슷하다는 걸 발견한다. 서로 효율적인 방법을 공유하고 머리를 맞댄다.

데이터를 수집하는 방법으로 근래에 많이 쓰이는 것은 알(R)이라는 프로그램 언어다. 데이터셋이 있는 장소에서 원하는 데이터만 골라서 가져오도록 프로그래밍을 하는 것이다. 그러나 R은 숫자로 된 정보는 강력하게 작용하지만 문자로 된 정보에는 약하다. 그래서 등장한 게 파이썬(Python)이라는 프로그램 언어다. 트위터나 텍스트로 된 문서의 크롤링과 분석에 쓰기 적합한 언어다. 다행히도(?) 프로그래밍에 능한 기자는 미국이나 유럽에도 많지 않다.

다시 미세먼지로 돌아가서 생각해본다. 정부의

〈그림 1〉 존스노우의 수도 지도





저감정보발령이 과연 효과가 있는지 의문이 든다면, 어떤 데이터를 수집해야 할까? 우선 1)저감정보 발령이 이루어진 날짜와 2)저감조치가 이루어진 날의 미세먼지 수치를 수집해야 할 것이다. 미세먼지 저감조치정보를 내리는 곳은 환경부, 미세먼지 수치를 수집하고 있는 곳은 환경공단과 기상청이다. 미세먼지 저감정보는 예측치를 기반으로 한다는 점에서 3)예측치도 수집하면 좋을 것이다. 효과를 파악하기 위해서는 저감조치로 인해 4)얼마나 많은 미세먼지 발생이 억제됐는지도 함께 봐야 할 것이다. 해당 데이터들은 다행히 모두 정부에서 생산되는 것들이다. 보도자료에 적힌 데이터를 하나씩 볼 수도 있고, 환경공단 산하 에어코리아에서 실시간으로 집계되는 미세먼지 수치들을 일일이 받아 내릴 수도 있지만 이빨 빠진 데이터(중간 중간 누락되는 데이터를 말한다)를 방지하기 위해선 정보공개청구를 시도하게 된다.

3. 정보공개청구의 세계에 오신 걸 환영합니다

대한민국 국민이라면 누구나 공공기관에 정보의 공개를 청구할 수 있다. 청구 방법은 간단하다. 정보공개포털(open.go.kr 단, 국회, 대법원 등은 별도 홈페이지를 이용)을 통해 신청하면 된다. 정보공개포털에 가입해 로그인만 하면 원하는 정보를 청구할 수 있다. 다만 청구 방법이 간단한 데 비해, 경우에 따라서는 청구한 자료를 받기까지 순탄치 않은 과정이 기다리기도 한다. 청구가 접수된 다음부터 청구 내용을 두고 공공기관과의 줄다리기가 시작되기 때문이다. (윤지희, 2018, 10, 16).

정보공개법에 의하면 공공기관이 보유·관리하는 정보는 공개 대상이고, 다만 그 정보가 정보공개법 제9조 제1항 각 호에 해당할 때만 비공개할 수 있다고 정하고 있을 뿐이다. 그렇지만 해당 기관에서는 심심치 않게 “누구십니까”를 묻는다. “어디에 쓰실

건가요”를 묻는 경우도 허다하다. 결재권자가 궁금해 한다고 묻기도 하고, 사용 목적을 알아야 정보를 줄 수 있다고 압박하기도 한다.

공공기관이 정보공개청구를 접수받고 처리해야 하는 시일은 10일. 10일 이내에 한 번 연장할 수 있어서 사실상 20일이다. 정보를 보유하고 있는 기관은 청와대, 국회부터 서울시 강남구 같은 지방자치단체는 물론 정부에서 예산을 받는 각종 위원회까지 모두 포함된다. 그러나 ‘공개’ 결정 통지를 받았다고 해서 완료된 것은 아니다. ‘공개’ 제목만 달고 실제 공개된 자료는 청구했던 정보의 일부만 공개한 ‘부분공개’ 수준이거나 실제로는 빈 파일을 주는 경우도 적지 않기 때문이다. 엑셀 파일로 달라고 했는데 굳이 PDF 파일로 만들어서 보내주기도 한다. PDF에서 중요한 항목은 지워서 보내면서 ‘공개’ 딱지를 달고 있는 자료도 허다하다.

요즘은 그래도 엑셀, CSV로 정보를 달라는 요청을 잘 받아들여주는 편이다. 그러나 정보공개 담당자들도 모르는 허점이 있으니...바로 정보 자체의 관리가 부실하다는 점이다. 똑같은 자료가 2번 3번씩 중복되어 집계되어 있기도 하고, 누락되어 있기도 하다. 정보공개청구로 받은 자료라고 맹신할 수 없는 셈이다. 이럴 땐 담당자와 한 땀 한 땀 자료를 비교 대조해봐야 하는데 공개에만 걸린 시간보다 더 길 때도 있다. (정보공개 청구로 20일이 걸렸다면 밀고 당기기하는 시간은 또 별도다.)

4. 내 것인 듯 내 것 아닌... 정보공개 담당공무원과 밀당

유행처럼 ‘블랙리스트’와 ‘화이트리스트’를 만든다면 정보공개청구자들 사이에서도 우스개처럼 리스트

가 존재한다. 데이터저널리스트들이 모이는 컨퍼런스에서는 서로의 막노동(속칭 노가다)에 대해 이야기하다가 “앗, KBS도 그 분과?”라며 슬픈 웃음꽃이 터지기도 한다. 경찰청의 000 님, 환경부의 000 님처럼 정보공개담당 주무관들의 이름은 전설처럼 언론사에 기록되고 있다는 걸 알고 계실지 모르겠다.

정보공개청구에서 정부가 제공하는 조언 중 하나가 담당 공무원과 대화를 나눠보라는 것인데 사실 대화를 너무 오래 나누다 못해 대화의 희열을 느끼게 되는 일이 생기면 좋을 텐데 어찌된 일인지 희열은 느끼지 못하고 대화의 절망이 쌓이게 된다. 해당 기관에서 수집하고 보유한 정보는 분명히 그 분(주로 담당 주무관)의 것이 아님에도 불구하고 왜 자신의 것이라고 생각해서 주기를 싫어하시는지... 남자친구보다 공무원과 밀당(밀고 당기기)을 더 많이 해본 것 같다.

경찰청의 교통사고 통계를 받기 위해 정보공개청구를 해 놓고 KBS 데이터분석가가 담당 주무관과 두 달 이상 통화를 하던 시기가 있었다. 교통사고를 당한 사람들의 성별, 연령, 사고 시간대, 위치, 횡단 보도였는지, 인도였는지, 신호는 초록불이였는지, 사고원인을 청구한 것이었는데, 경찰청에서는 해당 정보를 갖고 있으면서도 고집스럽게 “없다”고 우기던 중이었다. 법령에 따라 해당 정보를 수집하지 않으면 경찰청에 귀책사유가 있다는 점을 여러 차례 어필하고, 결국 상급자와 통화하고 나서야(그때가 정보공개 시한이 지난 지 한 달가량 지났을 때였다) 정보공개 담당 주무관은 “다음주 월요일에 정리해서 주겠다”고 했다.

밀당 끝에 오는 정보가 반가운 법. 주말을 편한 마음으로 지내고 출근해 월요일에 경찰청에 전

〈그림 2〉 국정감사 증인 명부(KBS 데이터저널리즘팀)



화하니 황당함이 폭포처럼 밀려온다.

“000 주무관 오늘부터 2주일 간 휴가 가셨는데요.”

프로이트 심리학에 나오는 ‘회피’ 기제를 이렇게도 사용하는구나... 하는 생각과 함께 오기가 발동해 주무관이 휴가 다녀올 때까지 끈질기게 기다렸다 복귀하는 날 전화를 했던 기억이 난다. 지금 생각해보면 그 분도 000 기자라면 이를 같고 계셨을 것 같다. 결론은 반쪽짜리 자료를 받았다. 차일피일 미루면서 반쪽 정보를 공개할 거였으면 내가 왜 그분과 그렇게 밀당을 해야 했는지 이해하기 힘들다. 교통사고 데이터는 분명 그분의 것이 아니었는데 말이다.

5. “야, 공무원도 힘들어!”

힘들면 데이터베이스화 합시다, 우리.

경찰청과 비슷한 사례는 공공기관마다 쉽게 찾을 수 있다. 국회사무처에 일하는 서기관(아는 분)은 이렇게 말한다.

“야 진짜 공무원도 힘들어. 그거 너희가 요구하는 거 다 만들어야 하는 거야. 엑셀 열어서 그거 자료 모아서 합치고 하는 게 얼마나 막노동인 줄 알아?”

그래서 진심으로 전달하고 싶은 말이 있다. 자료를 청구한 사람도 미안하고 안타깝다. 늘 막노동을 하고 있기 때문에 그 마음 이심전심이다. 그런데 만듦 때 제대로 만들었으면 막노동을 할 필요가 없는 자료인데 왜 수집을 그렇게 하는지 이해가 안 간다. 처음부터 데이터베이스를 만들었으면 막노동을 할 필요는 없었을 텐데 말이다.

지난해 KBS 디지털뉴스팀에서 대한민국 국회의 가장 큰 권리자 책무인 국정감사에 대해 자료를 청구했다. 지난 10년 간 국회에서 국정감사에 출석 하라고 요구한 증인들의 명단, 횟수, 상임위, 안건, 출석 여부, 답변 내용, 답변 시간을 분석해 증인이 정말 필요해서 부른 건지, 아니면 보여주기 식으로 부른다는 세간의 비판이 맞는 건지 확인해보기 위함이었다.(국정감사 증인 2,179명 연속보도)

그런데 자료의 질은 어땠을까? 자료가 있다는 것만으로는 설명이 안 되는 수준이었다. 국회사무처에서는 당연히 해마다 상임위에서 요청하는 증인들을 차곡차곡 데이터베이스에 쌓아놨을 것이라고 여겼는데 현실은 각각의 상임위에서 의결한 직후 그 문서를 PDF로 눌러서 첨부해놓은 게 다였다. 심지어 주요 증인이 출석했는지 여부는 상임위 담당 주무관의 능력과 비례하는 것처럼 보일 정도로 들쭉날쭉이었다. 회의록과 일일이 대조를 해보니 출석하지도 않은 사람이 출석한 걸로 표기가 되어 있고, 그 반대이기도 했다. 국정감사 기간의 정신없는 상황은 이해하지만, 국회법상 국정감사 자료는 감사를 마치고 60일 이내에 반드시 정리해 보

고하도록 하고 있는데도, 아무도 점검하지 않았다는 흔적이 분명했다. 결국 국회사무처와 밀당(밀고 당기기)을 통해 크고 작은 오류를 바로잡아 가면서 데이터를 정리하는 데 두 달 가까이 걸렸다.

이 과정에서 담당 공무원은 지난 10년치 자료에서 오류가 수도 없이 발견됐는데도 전혀 문책을 당하지 않았다. 그러나 국회(다른 공공기관도 마찬가지)가 갖고 있는 정보는 국민의 세금으로 수집하는 정보다. 국회가 자발적으로 모으는 것이 아니라 법에 규정되어 있기 때문에 정리해야 하는 정보다. 그럼에도 불구하고 국정감사 증인 명단과 출석 여부 같은 것들을 PDF 파일로 돌아다니게 만들고 있다는 사실은 국회가 국민의 세금을 낭비하고 있는 것과 마찬가지로.

PDF는 공공기관에서 널리 사용되는 파일 형태이지만, 데이터를 다루는 기자 입장에서는 정말 피하고 싶은 파일이다. PDF는 이미지 파일로 오직 읽기만 가능하고 수정이 불가할뿐더러, 그 안에 들어 있는 문자와 숫자는 화석이라고 봐야 한다. 2차 분석을 위해서는 전자화된 형태로 생명을 불어넣어야 한다. 영화 주라기 공원에서 화석을 살려내기가 얼마나 힘들었는지를 떠올려보면 알 것이다.

12월 말 멸절한 인도 벽돌을 뜯어내는 것만이 세금 낭비가 아니다. 세금으로 수집하고 있는 공공기관의 데이터들을 수집만 하고 체계적으로 관리하지 않고 있는 행태는 보도블록을 바꾸는 것보다 훨씬 심한 세금 낭비다. 공개했다고 생색만 내서는 곤란하다. 정제하지 않고, 관리하지 않고 있는 데이터들 '공개했다'고 하는 순간 의심이 덜해져 결국 국민의 눈을 가리는 시스템으로 활용될 가능성이 높기 때문이다.

6. 대한민국 초중고등학교 석면 지도 제작

쓸 만한 가치가 있는 데이터를 얻기 위해서는 정보 공개청구 → 공공기관과의 줄다리기 → 지저분한 데이터들을 구조화(데이터 클리닝) → 분석의 과정을 거쳐야 한다. 공공기관과의 지난하면서도 힘겨운 밀당을 겪고, 잘못된 데이터들을 일일이 수정하고, 또 PDF나 프린트 문서들을 일일이 CSV나 엑셀로 넣어 전자화한 뒤에 나온 보도 중 하나가 “최초제작, 대한민국 초중고등학교 석면 지도”다.

모든 학교건물과 공공건물에 대해 석면 검사를 실시하고, 석면건물일 경우 어디에, 얼마나 석면이 있는지, 그리고 위해등급은 어느 정도인지 알려야 하는 석면관리법이 지난 2012년부터 시행됐었지만, 정작 아는 사람도, 알리는 사람도 없었다. 법으로 조사하도록 돼 있는데 아는 사람이 없으니!

전국 교육청에 정보공개를 청구해 자료를 모으는 데만 닷 달 넘게 걸렸다. 내부 시스템에 석면정보를 올리지 않거나 대충 올린 학교가 적지 않았다. 교육청에선 각 지원청에 연락했고, 지자체에선 귀찮아했다. 광주광역시도 석면 위해등급별 점수가 전산화 되어있지 않아 데이터 리서처들이 3백 개가 넘는 학교의 석면정보를 일일이 문서와 대조해가며 엑셀파일에 입력해야 했다.

그런데 이상했다. 데이터를 모아놓고 보니 석면 위해등급 차이가 말도 안 되게 컸다. 전국에서 광주는 석면 위험 학교들만 있나 싶을 정도로 예민하게 측정된 반면, 다른 곳들은 평가 기준, 세부 항목이 지나치게 허술했다. 천장에 에어컨이 있는데도 기류와 진동으로 인한 석면 손상 가능성은 거의 없다고 한 곳들이 대부분이었다. 인체에 치명적인 석

면을 조사하도록 한 이유가 분명히 있는데도, 각각 자신의 학교만 확인하니 전체적으로 뭐가 다른지 전혀 몰랐던 걸 석면지도로 모아놓고 보니 나타난 셈이다.

7. 시민 참여로 데이터저널리즘 완성

KBS 데이터팀의 전국 초중고 석면지도가 만들어진 직후 가장 먼저 반응한 건 엄마들이 모인 맘카페였다. “아이들보다 교사가 더 문제죠”, “여기는 이렇습니다”는 댓글들이 나오면 기사에 미처 활용하지 못했던 학교의 석면 보고서를 다시 뜯어보기도 하고, 석면제거공사가 끝난 학교의 교무실에서 연락이 와 정보를 업데이트 해달라는 경우도 왕왕 있었다. KBS 9시 뉴스에 석면지도가 방송되면서 시청자들의 반응과 기사가 서로 상승작용을 이뤄냈다. 2016년에 내놓은 석면지도는 현재까지 KBS 데이터팀이 구축한 플랫폼 가운데 가장 많은 사용자들이 유입되는 플랫폼이다. 전국의 각급 학교들은 방학 때마다 석면 철거 공사를 하고, 서울 반포 등 대규모 재건축이 진행되는 곳에서는 오래된 건물에서의 석면 자재 방출이 심각한 문제로 떠오른다. 때문에 기자의 메일함에는 “우리 아이 학교 석면이 걱정된다”는 메일이 끊이지 않는다. 교사와 학부모들에게 경각심을 일깨우고, 그들이 적극적으로 나서 보도와 상호작용을 할 때 데이터저널리즘 기사는 비로소 완성된다.

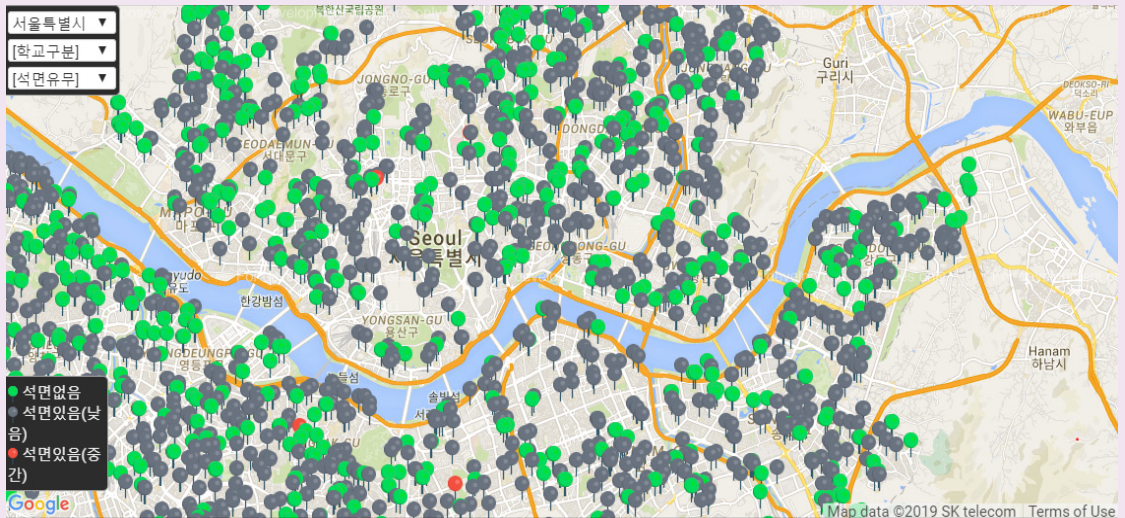
데이터저널리즘의 아버지이자 가디언(The Guardian) 기자 출신인 사이먼 로저스(Simon Rogers)는 데이터를 투명하게 개방하는 것이 데이터 저널리즘이며 뉴스룸에 사용자의 참여를 유도해야 한다고 강조한다. 기획보도를 톱 던져놓는 것이

끝이 아니라 데이터를 공유하는 사람들이 상호작용하면서 그 너머의 새로운 가치를 창출해야 한다는 거다.

중앙일보의 ‘탈탈 털어보자’ 시리즈를 보면 답을 알 수 있다. 중앙일보의 ‘탈탈 털어보자’ 시리즈는 사람들에게 잘 알려지지 않은 기초의회 데이터를 탈탈 털어본다는 데서 시작했다. 대표적으로 지난해 열린 제1회 데이터저널리즘 대상을 수상한 ‘우리 동네 의회살림’이 있다. 데이터저널리즘 뉴스콘텐츠로는 이례적으로 100만 조회 수를 기록했는데, 그 이유가 자못 재미있다. 지역 의회를 탈탈 털어보자는 취지로 기획된 이 기사는 지방의회의 영수증 사용 내역을 수집하고 분석해, 이용자가 자신의(혹은 궁금한) 지역구를 검색하면 해당 지역 의원들이 자주 가는 식당은 어디인지, 활동비를 어디에 썼는지, 조례는 무엇을 만들었는지를 한눈에 볼 수 있도록 했다. 아이러니하게도 해당 인터랙티브 데이터 기사가 폭발적인 반응을 이끌어낸 것은 지역 의회 의원들이 주로 가는 식당이 ‘맛집’이라고 소문이 나면서부터다. 이른바 ‘우리 지역구 맛집 리스트’로 소문이 나기 시작한 이 인터랙티브 데이터는 사용자들이 자신이 사는 지역 혹은 여행지의 맛집 정보를 검색하고 퍼나르며 100만 조회를 이끌어냈다.

사용자의 퍼나르기와 적극적 공유는 기사의 품질에도 당연히 기여했다. 모 지역구 주민들은 해당 의원들이 자주 가는 음식점이 그 지역 의회 의원이 운영하는 것이라는 걸 제보하기도 했고, 활동비가 의원들의 이익과 연결된 곳에서 주로 집행됐다는 것도 매의 눈으로 발견했다. 기초의회에 대한 감시는 국회의원에 대한 감시보다 미미하고 정보 자체가 많지 않다 보니 음식점 주인이 이권을 쥔 사람이라는 걸 기사가 파악하기란 불가능하다. 그러나

〈그림 3〉 전국 석면 지도(KBS 데이터저널리즘팀)



해당 데이터를 공유한 시민이 자신의 커뮤니티에서 그러한 사실을 알아내고, 정보를 제공함으로써, 데이터저널리즘 기사가 진가를 발휘하게 된다.

중앙일보 사례는 또한 그 내용뿐만 아니라 방법론에 있어서도 참여를 이끌어내는 것이 눈에 띈다. 이들은 ‘탈탈 털어보자’ 시리즈의 일환으로 최근 정보공개센터와 함께 받은 공공기관 산하단체들의 이 사들 명부를 소셜 미디어에 올려 어떻게 하면 데이터를 좋게 만들 수 있을지 시민들의 조언을 구하고 있다. 규모는 작지만 ICIJ(국제 탐사보도 언론인 협회)가 파나마 페이퍼스(Panama Papers)를 분석할 때 각국의 참여를 유도했던 방법과 같은 맥락이다.

8. 파나마 페이퍼와 조세도피처의 한국인들

2016년 풀리처상 탐사보도상을 수상한 ICIJ의 파나마 페이퍼스는 조세도피처를 추적하는 국제 공조프로젝트로 한국에서는 뉴스타파가 참여했다.

파일만 1천1백만 건이 넘고 크기는 2.6TB인 역사상 최대 데이터셋을 분석하기 위해 세계 각국에서 기자들이 머리를 맞댔다. 자료의 80%는 이미지. 이미지를 전자화하고, 숫자를 입력하고, 역외 탈세를 한 인물이 한국인인지, 중국인인지 걸러내고, 번역하고, 분석했다. 그 결과 시진핑 중국 국가주석의 매형, 푸틴 러시아 대통령의 최측근이 엄청난 규모의 탈세를 한 것이 나타났고, 아이슬란드 총리는 사임했다.

미국의 비영리 탐사전문매체 프로퍼블리카의 데이터 보도들도 배울 점이 많다. 미국 전역의 의료인 160만 명이 작성한 처방전 12억 건을 분석한 보도가 그 대표적인 사례로, 의사들이 대형 제약회사로부터 큰 후원금을 받고 대가로 해당 브랜드의 값비싼 약을 처방해주고 있다고 연속 보도했다. 프로퍼블리카는 이후 자료를 데이터베이스로 구축해 개방했고, 사용자들은 자신의 주치이나, 관심 지역을 검색하면 문제 있는 의사와 처방된 약품명, 가격까지 확인할 수 있게 되며, 기사의 업데이트를 주도했다.



9. 오늘도 네이버로 뉴스를 보셨습니까? 대한민국에서 뉴스유통의 한계

옥스퍼드대 산하 로이터저널리즘연구소는 해마다 세계 주요국을 대상으로 미디어 이용자 연구를 하는데, 한국에서 나오는 숫자를 보면 좌절할 때가 있다. 2017년 조사에서 대표적인 사례가 뉴스소비에 있어 포털 의존성에 대한 답변이다. 언론사 페이지로 직접 들어가서 본다는 답변이 4%로 전체 대상국 평균인 32%보다 현격히 낮았다. 96%의 경로 중 포털이 차지하는 비중은 77%. 그런데 그 네이버 뉴스에 선택이 되어야 국민의 간택을 받는다는 거다.

네이버 모바일의 초록 창에 들어가는 모바일 메인 뉴스는 정확히 5쪽지, 사진 이미지 뉴스를 포함하면 7쪽지다. 네이버 편집이 2~3시간 간격으로 바뀌는 걸 산술화하면 하루 평균 50건 가량이 네이버 모바일에 선택되는 수혜를 받는다. 그 중에 질 좋은 데이터저널리즘 기사가 들어가기란 하늘에 별 따기다. 들어간다고 해도 이용자들은 데이터 기사의

진면목을 감상할 수 없다. 국내 포털들은 인링크(네이버나 카카오 웹 테두리 안에서 콘텐츠를 볼 수 있도록 하는 형식) 방식으로 콘텐츠를 가두리 양식 장치럼 가둬놓은 데다 사용자와의 상호 작용성을 최대한 살린 인터랙티브(웹상에서 다른 페이지로 연결되는 프로그래밍이나 하이퍼링크)의 작동을 막아놓고 있기 때문이다.

KBS 데이터저널리즘팀은 2014년 국내에서는 처음으로 데이터팀을 꾸려 가동하고 있다. 석면지도를 비롯해 전국 소녀상 지도, 3.1운동 만세 특별 페이지, 우리 동네 재벌상권 등 질 좋은 기사들을 탄생시켰지만 많은 사람들이 알지 못한다. 포털에서 볼 수 없기 때문이다.

네이버는 지난해 3분기부터는 뉴스편집을 하지 않겠다고 밝혔다. 벌써 2019년 1분기가 지났지만, 여전히 약속은 지켜지지 않고 있다. 네이버가 뉴스편집을 하는 게 나쁜 게 아니다. 문제는 많은 사람들이 봐야하는 뉴스를 편집하지 않는 것이다.

10. 데이터저널리즘에 대한 오해와 진실

사정이 이렇다보니 언론사에서는 데이터저널리즘의 필요성에는 공감하지만, 현실적으로는 부정적이다. 그러나 ▲데이터저널리즘은 메이저 언론사에서만 가능한 모델이 아니다.

협업과 데이터를 보는 눈, 막노동에 대한 참을성만 있으면 어디서든 할 수 있다. 현재 미국 로컬 언론사들이 가장 선호하는 모델이 데이터저널리즘이기도 하다.

▲빅데이터만 분석해야 한다?

그렇지 않다. KBS 데이터팀이 만든 소녀상 지도는 전형적인 스톨데이터 모델이다. 소녀상의 위치를 좌표로 바꾸고, 상권과 겹치는지 보면, 사람들이 많이 볼 수 있는 곳에 있는지, 방치되어 있는지 알 수 있다.

▲인터랙티브 요소를 살리기 위해선 디자이너와 개발자가 필수다?

개발자와 디자이너가 별도로 있으면 좋다. 그러나 엑셀을 할 줄 아는 기자와 구글, 태블로 등에서 무료로 제공하는 디자인만 활용해도 데이터저널리즘을 시작할 수 있다.

▲데이터저널리즘은 어렵다?

하고 싶은 이야기가 무엇인지 생각해보자. 주장에 필요한 근거가 필요한지, 근거를 기반으로 이야기를 하고 싶은 건지... 후자라면 데이터저널리즘을 이미 시작한 셈이다.

▲정보공개청구를 꼭 해야만 하나?

KBS 데이터저널리즘팀 윤지희 분석가의 말을 빌리면, 정보공개청구를 통해 데이터를 수집하는 방법은 시간적 측면에서는 효율적이지 않을 수도 있다. 모든 청구 내용에 대해 공개 결정이 내려지는 것도 아니다. 하지만 정보공개청구는 공공기관 내부에 있던 정보를 기관 밖으로 이끌어낼 수 있는 효과적인 방법이다. 이렇게 공개된 정보는 때로는 정보 그 자체만으로 새로운 사실이 되고, 데이터에 대한 분석을 거치면 더 많은 정보와 의미를 드러내기도 한다. 정보 비공개 또는 부존재 통지 역시 그 자체로 의미를 가질 때도 있다. 그래서 시도할 이유는 충분하다. 비록 정보공개청구는 시간도 많이 걸리고 세련된 방식의 데이터 수집 방법도 아니지만, 이를 통해서만 구할 수 있는 정보가 있기 때문이다. 데이터저널리즘의 한편에서 정보공개청구가 앞으로도 계속 활용될 것이라고 생각하는 이유다. 📌



참고문헌

윤지희 (2018, 10, 16). 공공기관과의 정보 줄다리기. <커먼스랩>. URL: <http://bitly.kr/yALwPi>