



독재자의 권한 침해와 관료제의 합리주의

상상을 초월한 박근혜와 최순실의 엽기 행각이 사람들을 놀래키고, 분노케 하고, 좌절시키고, 분열시키고 있다. 뜬금없이 “창조”를 말했지만 비선과 비정상으로 꼼꼼하게 챙겨낸 변태 추문이였다. 현재로서는 그

추문의 끝이 어딘지 가늠하기도 어렵다. 국민 행복 시대를 열겠다고 기염궐을 토했지만 단지 그들만의 행복 시대였다. 밀도 끝도 없이 “미래”를 말했지만 박씨의 아버지가 좋맛아 죽었던 유신 독재 시절로 시계 바늘을 거꾸로 돌렸다. 이성과 상식과 엄치를 기대할 수 없는 자들이다. 사리사욕을 채우기 위해 정부 관료제를 멋대로 부러먹은 자들이다. 장난감처럼 갖고 놀다가 망가뜨린 자들이다. 공직사회를 제멋대로 흔들기 기강을 무너뜨린 자들이다. 어찌하여 관료제와 공직자들이 그리도 속절없이 휘둘렸단 말인가?

권한 침해와 행정의 합리주의 현상

이문영(1980)은 정부의 행정과정에서 “합리주의 현상”이 발생하거나 계속되지 못하는 것은 “권한 침해” 때문이라고 단언했다(6쪽). 권한 침해는 “아랫사람이 고유하게 갖는 권한을 윗사람이 부당하게 빼앗는 관행”을 말하는데(이문영 2001: 111), 권위주의나 관료주의 문화에서 흔히 볼 수 있는 현상이다. 예컨대, “정치적 행정에 대한 극도의 간섭증, 업관주의, 상급기관이나 상급자의 하급기관이나 하급자에 대한 권한 남용, 예산의 편성과 집행에[서] 업무담당기관이 아닌 ... 예산기관의 횡포” 등이다(이문영 1980: 6).

권한 침해는 웨버(Max Weber)가 합리성을 가장 잘 갖춘 형태로 묘사한 관료제 이상형(ideal type bureaucracy)과 거리가 멀다. 업무가 기능에 따라 분명하게 나뉘어 지고, 자의성이 배제된 합당한 규칙에 따라 조직이 통제되고, 전문 지식을 가진 사람이 계약에 의해 자리에 임명되고, 업무 성과에 따라서 승진이 이루어지는 관료제의 특성을 해체한다. 단지 아랫사람이라는 이유만으로 그 사람의 직책과 권한을 무시한다면 업무가 분화되지 않은 것이고, 합의된 규칙은 적용되지 않으며, 전문 지식은 인정되지 않고 (웃사람이 “내가 해봐서 아는데”라며 허세를 부리면서 만든 전문가 행세를 하고), 성과가 아닌 위계질서에 의한 무조건 상명하복으로 승진이 결정된다는 뜻이다. 누가 웃사람이 되느냐, 누가 권력을 장악하느냐에 따라서 관료제가 좌우된다는 뜻이다.

박근혜와 최순실의 권한 침해는 정부 관료제를 넘어서 민간기업과 예술인에게까지 미친 엽기와 변태였다. 공과 나를 넘나든 이른바 “갑질”의 최고봉이였다. 공무원 인사를 마음대로 주무르는 것은 물론이고 최씨 존재를 언급하는 것조차

도 금기시하고 보복을 가했다. 무자비한 폭력 그 자체였다. 이런 관료제에서 행정은 예측가능성과 연속성과 안정성을 가질 수가 없다. 언제 어떤 이유로 어떻게 자리에서 쫓겨나갈 지 알 수 없다. 힘센 자들의 “갑질” 때문에 이성과 상식을 따라 일하기 어렵다. 공직자들이 갈피를 잡지 못하고 우왕좌왕하게 된다. 공직자들이 전문성과 자율성을 가진 관료가 아닌 그저 “네 네”하는 심부름꾼이 될 것을 강요받는다. 독재자 편에 줄설 것을 강요하는 관료제는 좋은 경험을 축적시키지 못하고 합리주의를 구현할 수 없게 된다.

합리주의는 행정이 가야할 길

그러면 왜 합리주의인가? 그것이 행정이 지향해야 할 방향이자 걸어가야 할 “제 길”이기 때문이다(이문영 1980: viii). 官의 권력 남용과 民의 무책임한 행동을 극복해야 하기 때문이다. 그래서 이문영(1980)은 벌거벗은 힘(naked power)과 “난동을 불사하는 힘”이 부딪히는 원색적인 대결을 피하고 “관의 정책과 민의 평화”가 호응하는 합리적 관계로 발전해야 한다고 말했다 (vii쪽). 官이 “옷을 입[은] 힘”을 행사하고 民이 자율성이 있는 시민사회를 형성하고, 백성의 정당한 의견과 요구를 정부가 존중하고 받아들임으로써 양자 간 피를 보는 사태를 피할 수 있다(vii쪽).

결국은 官의 권력 남용과 民의 난동은 모두 폭력인데, 이런 극단의 폭력이 난무하는 데서 합리성을 기대하기 어렵다. 이러한 폭력을 극복하기 위하여 비폭력을 대안으로 내세운다. 통치자와 공무원이 권력 남용을 억제하고 백성들이 난동을 자제하는 것이다. 주먹질과 발길질을 멈추고 합당한 말로 사실과 진리를 다투는 것이 바로 비폭력이다. 그런데 官의 비폭력보다도 民의 비폭력이 훨씬 중요하다. “민이 좋은 행동을 官이 배우며 官의 나쁜 행동을 民이 배운다”(이문영 1990: 29). 원래 官은 좋은 행동을 할 능력을 갖고 있지 않다(29쪽). 주인이 부지런하고 모범을 보여야 머슴들의 못된 버릇장머리를 고쳐서 제대로 부릴 수 있다는 뜻이다. “주인인 국민이 만들어내는 감동, ... 국민의 合理的의 抵抗, 祝祭분위기의 편재가 국민의 종인 통치자를 변하게 만든다”(이문영 1991: 30).

물론 비폭력이 궁극의 목표는 아니다. 정부 관료제에서 합리주의나 합리성을 만들고 가꾸어 나가는 것이 우리가 추구하는 것이다. “폭력과 비폭력의 중간가치는 합리성 rationality이다”(1991: 118). 헤겔의 역사발전논리로 치자면 폭력은 정정(thesis), 비폭력은 반反(antithesis), 제 3의 가치인 합리성은 합성(synthesis)에 해당한다(이문영 1991: 30). 모순이 있는 상태를 부정하여 모순을 극복하려는 노력이다. 모순을 부정하는 반이 없이는 승에 이를 수가 없기 때문에 승에 도달하기 위한 과정으로 반을 내세운다(30쪽). 따라서 비폭력, 개인윤리, 사회윤리, 자기회생으로 이어지는 이문영의 초월윤리는 승인 합리성에 도달하는데 필요한 반이지 그 자체가 목적이 아니다(30쪽).

공무원 이상형의 행동강령

이문영은 행정개혁은 (1) 인간계발을 위한 환경 조성 (권한 보장 작업), (2) 과학적 관리와 인간관계 향상, 그리고 (3) 적극적인 인간계발 단계를 거친다고 보았다(이문영 1980: 4). 그런데 처음 두 단계에서 성공을 하는 것은 소극적인 행정조치일 뿐이어서 행정개혁을 완수完遂하기에는 부족하다(7쪽). 궁극적인 힘은 사람에게서만 나오는데, “무엇에나 ‘네, 네’ 하는 사람이나 책임을 회피하는 사람이 아니라 좀더 독립적이며 자신을 갖고 있으며 근면하며 새로운 아이디어를 열심히 탐색하는 사람”이 필요하다(7쪽). 이문영

책임기

김동환의 <빅데이터는 거품이다>

은 이러한 공무원 이상형이 따라야 할 행동강령을 다음과 같이 제시하였다 (2001: 277, 465; 2008: 561).

(1) 공무원이 일을 하는 '방법' 은 전문 지식(능률과 민주주의 이념 포함)을 매일매일 닦아서, 이 학문을 갖고서 조직 내에서는 상사를 존중하며 조직 밖에서는 국민에게 봉사하는 것이지, 전문지식을 연마하지 않고 다만 상사에게 굴종하며 조직 밖의 국민들에게는 교만하는 것이 아니다.

(2) 공무원이 할 '일' 이란 민주·복지국가와 정의를 추구하는 것이지, 상사에게 상납하는 대가로 '이' (利)를 추구하는 것이 아니다.

(3) 공무원인 '사람' 이란 전문 지식(능률과 민주주의 이념 포함) 앞에서는 존귀한 존재이지, 상사와 조직 앞에 세워지는 존재가 아니다.

박근혜 최순실 추문에서 우리가 뼈아프게 반성해야 하는 것은 그들의 "권한 침해"나 과학적 관리 실패가 아니다. 물론 공사公私를 구분 못하고 제멋대로 관료제를 휘젓은 것은 명백한 죄악이다. 하지만 "공무원 이상형"에 가깝게 행동했던 공직자가 드물었다는 사실이 비수치수가 되어 행정학의 가슴을 찌른다.

전문 지식을 닦아서 상사를 존중하고 국민에게 봉사한 것이 아니라, 공직자 대부분은 전문 지식을 내팽개치거나 악용하였고 부당한 지시를 내리는 상사에게 굴종하고 국민들에게는 교만했다. 국민의 기본권과 복지를 신장하고 정의를 실현한 것이 아니라 상사의 부당한 명령까지도 복종하는 대가로 이득을 챙겼다. 자율성을 가지고 전문성 앞에 당당하게 서는 공직자가 아니라 그저 상사와 조직에게 쓰여지고 버려지는 소모품이나 도구를 자처했다. 안중범 전 경제수석비서관과 유진룡 전 문화체육관광부 장관의 모습이 대조되는 이유다.

비루한 고위 공직자의 자화상

특별히 고등고시나 사법고시에 합격하거나 박사학위까지 취득하고 득의양양得意揚揚했던 고위 공직자들을 생각한다. 그들의 비굴한 처신은 사람들을 참담하게 한다. 그들의 전문 지식이란 것이 사

김동환(2016)은 이명박 정권 말기부터 나라를 휩쓸고 있는 "빅데이터" 광풍을 유행이라고 표현했다. 1980년대를 풍미했던 "데이터"라는 유행이 1990년대의 "정보," 2000년대의 "콘텐츠"와 "사이버 공간," 2010년대의 "스마트"를 거쳐서, 30년 만에 다시 "빅"이라는 접두사를 달고 돌아왔다는 것이다(41-42쪽). 빅데이터는 아무리 커도 데이터일 뿐인데 호사가들이 "그럴듯한 신화"로 둔갑시켜 사람들을 미혹迷惑시키고 있다고 했다(155-158쪽).

빅데이터라는 유행의 구조

이러한 "지적 유행"은 그 바닥에 여유자원이 넉넉하고, 그 떡고물을 얻어먹을 수 있는 관(기회)이 벌어져야 하고, 진리를 모르거나 회피하는 열악한 지식 풍토가 있어야 가능하다(113-114쪽). 도박으로 치면 몽치돈을 대주는(잃어주는) 호구, 도박판을 벌여주는 하우스 운영자와 돈을 챙기는 타짜, 타짜와 호구를 엮어주는 바람잡이에 비유된다(115쪽). 빅데이터 광풍으로 치면 각각 눈먼 정부예산, 빅데이터 관련 업체와 정부, 그리고 빅데이터 옹호자와 침묵하는 지식인에 해당한다(115-116쪽). 빅데이터를 모르면서 아는 척하거나 알면서도 모르는 척하고 침묵하는 타락한 지식인이 바람

회에 빛과 소금이 되기는 커녕 출세와 사리사욕을 채우는 데 쓰였을 뿐이다. 그들은 특혜와 특권을 누리는 데에는 "엄치불구"였고, 전문성에 걸맞는 사회 윤리와 책임은 나몰라라 했다. 공학도라면서 산소가스 이산화가스라는 박씨와 "심신이 회폐"한 "최선생님" 앞에만 서면 바짝 쪼그라들어 오금도 펴지 못하는 전문성이라니... 한때 하늘을 찌를듯한 자존심과 자부심은 다 어디에 내팽개쳤단 말인가?

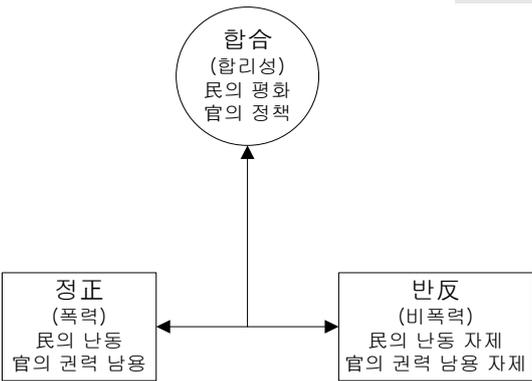
청문회에 나와서도 모른다거나 기억이 안난다거나 누가 봐도 뻔한 거짓말을 늘어놓은 고위 공직자와 자칭 사회 지도층을 기억한다. 걸핏하면 나랏일을 한다며, 혹은 공인이라며, 혹은 사회 발전에 기여해왔다고 온갖 공치사란 공치사는 다 늘어놓았던 자들 아닌가. 이런 자들의 비루한 모습을 보면서 사람들은 과연 무엇을 느꼈을까?

잡이다(15-17쪽). 이와같이 정부, 빅데이터 업체, 타락한 지식인이 쇠같이 단단한 삼각관계를 이루어 광풍을 주도한다고 저자는 분석했다(115-119쪽).

빅데이터에 대한 환상과 홍보에도 불구하고 사업 성과가 지지부진하자 관련 업체와 바람잡이 지식인들은 분석할 빅데이터가 부족하다며 정부에게 빅데이터 포털을 통해 공공 정보를 공개하라고 요구한다(77쪽). 개인정보보호 규제가 빅데이터의 발목을 잡고 있다고 목소리를 높였다. 어마어마하게 생산되는 빅데이터를 이용하려고 시작한 사업이 빅데이터가 없어서 실패한다는 모순은 그리스 신화에 나오는 Ouroboros를 보는 듯하다(146-147쪽). 또한 빅데이터 전문가가 부족한 탓이라며 전문가 양성 교육을 강화하라고 정부에 요구한다(75-77쪽). 결국 빅데이터 호사가들은 반성은 커녕 책임을 떠넘기며 끊임없이 전문가 교육, 공공 정보 공개 등으로 화제를 돌리면서 자가발전을 꾀한다. "정부는 빅데이터에 돈을 대는 것도 모자라, 또 빅데이터 실패의 원인으로 지목되어 욕을 먹는다"(81쪽).

흔히 빅데이터 유행의 근거로 Nature에 발표된 Ginsberg et al. (2009)이 거론되는데, 사실 이 논문은 구글의 검색엔진을 사용하여 독감 의심 환자 비율을 추정(estimation)한 것이지 예측(prediction)한 것이 아니고 저자는 지적한다(89-90쪽). 몇 년 후 Butler (2013)는 구글의 독감환자 추정치는 질병예방통제국(CDC)의 결과와 차이가 큰 경우가 있었다고 보고했다(94쪽). 이에 비해 소수 참여자들의 정보를 분석한 Brownstein의 독감추적 프로그램은 질병예방통제국의 추정치와 가까웠다(96쪽). 말하자면 빅데이터로 미래를 예측한다는 주장이 허구이며 환상이며 망상이다(158-159쪽). "과거에 대한 측정과 추론을 미래에 대한 예측으로 오해"하기 때문이다(159쪽). 미국의 빅데이터 유행도 좋은 물건을 만들거나 서비스를 제공하기보다는 단순히 주가를 올려 이익을 챙기려는 "주주 자본주의(shareholder capitalism)"와 정부의 투자에 편승했다고 보았다(129-135쪽).

말이 빅데이터지 사실은 Facebook, Twitter, 카카오톡 등에 담겨진 문자정보에 의존하게 되는데, 이런 잡담을 굶어모아 분석한다 한들 사회 현상을 이해하고 미래를 예측할 수는 없다는 얘기다(40쪽). 저자는 빅데이터를 분석하여 재난이나 범죄를 예



방한다는 대목에서 특히 절망스러워했다 (56쪽). 주민이 경찰에 신고를 하면 데이터를 분석할 것이 아니라 즉시 출동해야 할 일 아닌가?(48쪽) 빅데이터를 분석한다 한들 산사태와 눈사태가 언제 어디서 일어날지 어떻게 예측한다 말인가?(49-50쪽). 이 책에서 가장 인상깊은 문장은 이것이다. “주민들이 위험하다고 신고할 때 적극적으로 반응하는 행정체제만 유지되더라도 대단한 것이다. ... 그저 평소에 침수 위험 지역을 꾸준히 관리하는 것만으로도 대단히 훌륭한 재난 행정이라고 할 수 있을 것이다”(50쪽). 말하자면 기본에 충실한 최소주의 행정학이다.

한국에 쓸만한 자료가 있던가?

이 책을 처음 소개받고 나는 바로 반응을 했다. 한국에 (쓸만한) 공공 자료가 있던가? 잘 측정되고 정리된 보통 자료도 구경하기도 힘들 지경인데 무슨 빅데이터인가라고 반문했다. 백성들이 원하는 정보를 알차게 담고 있는 공공 자료를 기대하는 것이 얼마나 부질없는 일인가? 공공기관의 문서와 자료가 적절한 형식으로 작성되어 있기를 바라는 것은 희망사항에 가깝지 않았던가? 투박한 자료조각을 찢고 채고 오리고 붙이고 한바탕 난리를 쳐야 그나마 좀 쓸모 있게 보이지 않았던가? 경험에서 얻은 상식이다. 그런 자료일망정 정부에서 문서와 자료를 얻는 일이 그 자체로 얼마나 어려운가? 애초부터 분석을 하지 못하게끔 작정을 하고 교묘한 방법으로 자료를 가공하고 감질나게 찢끔찢끔 민간에 흘리면서 생색이나 낸다고 보는 것이 차라리 속편하다. 그럴 때마다 연구자는 절벽 앞에 선 심정인 것은 나만의 경험인가? 그런데 빅데이터라니... 생똥맞다고 해야 할까? 그냥 하던 일이나, 해야 할 일이나 제대로 하세요...

미국 공공기관의 자료에 관해 충격을 받았던 적이 있다. 어느날 미국인 대학생이 CD-ROM 하나를 들고 내게 찾아와서 어떻게 이 자료를 SAS로 읽어서 분석할지를 물어 왔다. 병원에서 진료를 받은 환자의 질환에 관한 자세한 자료가 고정길이 형식(fixed format)의 문자파일로 저장되어 있었다. 관측치 수(N)는 물론이려니와 변수도 많아서 컴퓨터가 힘들어 할 만큼 SAS 자료 파일이 매우 컸다. 다른 자료분석 프로그램은 아마도 읽기도 어려웠을 것이다. 그런데 놀라운 것은 그 자료가 19세기 중반부터 시작하고 있었다는 점이다. “아... 도대체 이나라는...” 절로 탄식이 나왔다.

미국에서 어지간한 정부 문서와 자료는 어렵지 않게 구할 수 있고, 내용이 적절하고 알차며, 형식도 손질이 잘 되어 있다. 한국 정부의 웹집(Web site)은 휘황찬란해도

자료를 얻기 불편하고 마땅히 얻어갈 것이 없다. 반면 미국 정부의 웹집은 화려하지는 않아도 기능에 충실하고 내용이 풍부하다. 당장 미국 감사원(Government Accountability Office) 웹집에 가보라. 또한 General Social Survey (GSS)와 Current Population Survey (CPS)같은 공공 자료가 많은 연구자들에게 무료로 제공된다. 슈퍼컴퓨터 네트워크를 통해 전국에 있는 자료(예컨대, CDC 자료)를 공유할 뿐만 아니라 연구자에게 유용한 분석도구를 제공한다. 이른바 grid computing으로 data grid 를 실현하고 있다. 주요 기관에 분산되어 저장된 대용량 자료를 효율성있게 활용하는 체제다.

미국은 “빅데이터”가 유행하기 전에 이미 어마어마한 데이터를 쌓아놓고 있었다는 점이 중요하다. 저자가 비판하는 행태주의와 거대주의의 소산물이 아니라 합리성을 추구하는 관료제에서 한 축이 되는 문서주의가 이뤄낸 결과물이다. 거대한 자료보다도 정보관리 체계가 부럽다. 어쩌면 20세기 이후 한국의 정보관리 수준은 15세기 조선왕조를 따라가지 못하는지도 모른다.

책에서 묘사된 “빅데이터 중독”은 생각한 것보다 정도가 심했다. “묻지도 따지지도 않고” 나랏돈을 뿌려대며 “그들만의 잔치”를 벌이는 형국이다. 소문은 무성하고 정부 사업은 요란한데 지금이나 예전이나 쓸모있는 자료는 빈약하기 그지없는 것이 “대한민국 빅데이터”의 현주소다. 한국의 빅데이터 현상은 반성과 비판없이 미국에서 베껴온 것이며, 실제 국민이 필요해서 시작한 것이 아니라 정부가 주도하여 판을 벌였다는 김교수 지적 그대로다(111쪽). 끝장을 보기 전에는 절대 멈추지 않고 질주하는 폭주기관차라고나 할까? 한번쯤은 “왜 우리가 이런 일을 하고 있을까?” 자문해볼 법도 한데 말이다.

빅데이터는 자료인가, 기술인가?

가장 눈에 들어오는 한국과 미국의 차이는 “빅데이터”를 어떻게 정의하고 있는가이다. 2011년 국가정보화전략위원회에서 발행한 <빅데이터를 활용한 스마트 정부 구현(안)>에 의하면 빅데이터는 “대용량 데이터를 활용, 분석하여 가치있는 정보를 추출하고, 생성된 지식을 바탕으로 능동적으로 대응하거나 변화를 예측하기 위한 정보화 기술”이다(38쪽). 한글 위키피디아에는 “기존 데이터베이스 관리도구의 능력을 넘어서는 대량(수십 테라바이트)의 정형 또는 심지어 데이터베이스형태가 아닌 비정형의 데이터 집합조차 포함한 데이터로부터 가치를 추출하고 결과를 분석하는 기술”이라고 되어 있다. 빅데이터가 데이터가 아니라 “기술”이라는 얘기다. 데이터가

커지면 더이상 데이터가 아니라 “기술”로 환골탈태換骨奪胎를 한다는 말인가? 납득하기 어려운 정의다. 빅데이터 사업의 지지부진이 큰 데이터가 없어서라는 변명이 분석 “기술”은 있는데 분석할 “자료”가 없기 때문이라는 사술邪術로 들리는 까닭이다.

미국에서 빅데이터는 어쨌거나 “자료”다. Laney (2001)는 전자상거래에서 자료관리 문제가 세가지 차원에서 폭증하고 있다고 지적했다. 이런 기준에서 보면 빅데이터는 깊이와 폭이 큰 자료(high volume)이고, 자료가 빠르게 생성되고(high velocity), 형식과 구조가 다양한 자료(high diversity)다. Wikipedia에는 너무 크고 복잡해서 기존의 자료처리방법으로는 요리하기 어려운 자료집(data sets)이라고 적혀있다. 빅데이터 업체라 할 수 있는 SAS와 IBM도 규모가 크고, 대량으로 발생하고, 다양한 형식을 가진 자료라고 정의하고 있다. 상식에 맞고 현실성이 있는 정의이다.

어쩌면 비약으로 들릴지는 모르겠으나 전자정부에 관한 정의도 비슷하다. 2001년에 제정된 한국 전자정부법 제 2조는 “정보기술을 활용하여 행정기관 및 공공기관...의 업무를 전자화하여 행정기관의 상호 간의 행정업무 및 국민에 대한 행정업무를 효율적으로 수행하는 정부”라고 정의하고 있다. 2002년 제정된 미국 전자정부법 Section 3601에는 정부가 웹기반의 인터넷과 정보기술을 이용하는 것(“the use of the Government of [W]eb-based Internet applications and other information technologies”)이라고 적고 있다. 한국의 전자정부가 당위나 신화에 가깝다면 미국의 전자정부는 상식과 현실에 가깝다. 어느 정부나 한국에서 정의한 전자정부를 꿈꾸고 있지만 그 누구도 그런 정의에 꼭맞는 전자정부를 가지고 있지 않다. 애초부터 한국의 전자정부는 현실과 다른 차원에서 시작되었고 현재 3.0으로 진화하면서 “그들만의 신화”를 써가고 있다. 전자정부 2.0이든 3.0이든 “이명박근혜 정부”가 일을 잘한다고 생각하는 시민들이 많아보이지 않으니 말이다.

무엇이 빅데이터인가?

빅데이터는 주로 (1) 디지털 거래자료 (예컨대, 신용카드 사용내역), (2) 감시카메라, 인공위성 등에서 얻은 사진과 영상 자료(예컨대, CCTV와 remote sensing), (3) 이동통신 자료(예컨대, 교통카드, 무선결제, 무선전화 사용 자료), (4) 사회매체(social media)에서 발생된 자료 (예컨대, Facebook, Blog, Youtube,)에 집중되어 있다. 그런데 주요 관심은 특별히 사회매체의 한 종류인 사회망서비스(social network service)에서 일어나는 대화나 인터넷 정보(뉴스 기

사)다. (1)과 (3)은 기존의 방법대로 일정한 형식으로 잘 정리된 자료인데 관측치 수만 매우 크고 계속 생산된다는 특징이 있다. (2)는 그림이나 영상 자료라는 특징이 있다. Facebook, Twitter, Skype, 카카오톡 등에는 문자정보 뿐만 아니라 사진, 동영상, 음성, 이진파일과 같은 비문자 정보가 담겨있다. 하지만 비문자정보는 분석하기가 어렵기 때문에 빅데이터 분석은 사실상 문자정보에 치중하고 있다. 물론 사진과 동영상을 분석하여 지하철의 흔적도를 측정하고 특정한 사람들의 행위(에컨대, 테러)를 판별하는 연구가 진행되고 있는 것도 사실이다.

빅데이터의 대표성은 있는가?

빅데이터의 가장 큰 문제는 사회 전체 모습을 반영하지 않는다는 점이다. 예컨대, 신용카드나 교통카드를 사용하지 않는 사람들은 빅데이터에서 빠진다. 지하철을 이용하지 않거나 시골에 머물러 사는 사람들 역시 제외된다. 인터넷이나 사회망서비스를 사용하지 않는 사람들의 대화와 행동도 빅데이터는 담고 있지 않다. 그러면 과연 얼마나 많은 사람들이 사회망서비스에서 글을 남기고 대화에 참여하는가?

한국 시민들의 참여도는 높은 편이지만 그래도 웹마실(Web surfing)을 다니면서 다른 사람이 쓴 글을 눈요기하는 사람들이 다수다. 아무나 블로그를 시작할 수 있지만 좋은 블로거가 되기는 쉽지 않다. 소위 파워 블로거는 교육을 많이 받고 지식과 능력을 가진 교수, 언론인, 변호사들이다(Hindman 2009). 결국 빅데이터는 이런 소수 엘리트나 적극 참여자의 말을 주로 담고 있기 때문에 대표성을 가지기 어렵다.

또한 이름을 가리고(익명으로), 돈들이지 않고, 어렵지 않게 자신의 의사를 표현할 수 있다는 점은 칼의 양날과 같다. 웹이 등장하기 전에는 의사표시를 하기 어려웠던 사람들도 참여할 수 있는 장점이 있는 반면 대화의 품격과 신뢰성이 떨어진다는 단점이 있다. 이른바 “구글세대(Generation Google)”에서 피하기 어려운 가상공간의 시궁창(cyber-cesspool)은 사람들이 얼마나 쉽게 사회매체를 남용하며 그 폐해弊害를 관리하기가 어려운지를 보여준다(Levmore and Nussbaum 2010). 물론 많은 사람들이 사회매체를 선하게 사용하여 유용성과 즐거움을 얻고 있다는 것도 사실이다.

이렇게 사회의 일부 조각을 반영하는 빅데이터와 과장, 왜곡, 조작 등으로 신뢰성이 떨어지는 빅데이터를 분석한다면 그 결과는 기대와 정반대가 될 것이다. 아무리 슈퍼컴퓨터를 동원하고 엄청난 기법을 적용한다고 해도 분석결과와 현실과 거리가 멀 수밖에 없다. 사회현상을 이해하는데 도

움이 아니라 장애가 될 수 있다. 일단 쓰레기가 들어가면 슈퍼컴퓨터라 한들 쓰레기를 토해낼 수밖에 없다(garbage in, garbage out). 자료를 기반으로(data-driven) 하는 방법론의 숙명이다. 이런 빅데이터라면 아무리 규모가 크다 한들 미래를 예측하는 일은 불가능하거나 확실하게 틀린 결과를 내놓을 뿐이다. 시민들이 사고가 난 현장의 모습을 휴대전화로 찍어서 Facebook에 올린 사진을 보고 사태를 수습하는 것과 가상공간의 시궁창에서 벌어지는 싸구려 잡담을 분석해서 미래를 예측하는 것과는 전혀 다른 일이다. 아무리 사례나 관측치 수가 크다 해도 말이다.

빅데이터는 얼마나 커야 하나?

빅데이터는 덩치가 크고 다양한 형태를 가졌기 때문에 기존에 자료를 저장하고, 손질하고, 분석하는 방법으로는 한계가 있다고 한다. 그래서 빅데이터라는 식이다. 참 이해하기 어렵다. 기존의 방법으로 빅데이터를 적절하게 처리할 수 없다면 새로운 방법은 무엇인가? 기존 방법은 계속 발전하고 진화하는데 도대체 무엇을 말하는가? 웹이 출현하던 90년대 이전의 방법인가?

먼저 얼마나 커야 빅데이터인가? 관측치 수가 몇 개면 되는가? 1억 개면 족한가? 변수가 몇 개쯤 되어야 하나? 저장공간은 어떠한가? 1 Peta (1,024 Tera) 바이트면 되는가? 올해는 Peta 바이트면 되고 내년에는 Zeta (1,024 Peta) 바이트면 만족하겠는가? 이런 식이면 20년 전에도 빅데이터는 존재했을 것이고 그 전에도 그랬을 것이다. 당시에 수십만 개 관측치와 변수는 상상하기 어려웠을 것이고, Giga (1,024 Mega) 바이트는 꿈같은 크기였을 것이기 때문이다. 1988년 당시 20 Mega 바이트 하드디스크를 보고 나는 얼마나 감동을 했는가?

누구한테 큰 데이터인가?

사실 데이터가 큰지 아닌지, 복잡하지 아닌지는 연구자의 처리 능력과 기술에 따라 다르다. 누구에게 크고 복잡한 데이터냐가 중요하다. 관측치가 10개라면 유치원생에게는 버거운 크기일 테지만 중학생에게는 식은 죽 먹기일 것이다. 몇 백 개는 대학생에게 만만하겠지만 몇 십만 개는 부담스러울 것이다. 만 개는 유치원생에게 상상을 초월하는 크기다. 몇년 전에 한 학생이 천만 개도 넘는 어마어마한 거래정보라며 들고 왔다. (1)번 빅데이터인 디지털 거래자료였다. 자료처리 능력이 부족한 그 학생에게 몇 Tera 바이트는 “슈퍼 빅데이터”였을 테지만, 내게는 “껌값”은 아니어도 랩탑에서 가볍게 처리할 수 있는 그냥 그렇고 그런 자료일 뿐이었다. 아는 사람은 딱 보면 “견적”이 나오는데 모르는 사람들이 보면

불가능이자 기적이자 충격이다. 어쩌면 빅데이터라며 호들갑을 떠는 인간들은 정작 자료가 무엇인지, 어찌 분석해야 하는지를 잘 모르기 때문에 그러는지 모른다. 실제 자료를 손질하거나 요리해 본 경험이 없거나 능력이 없기 때문에 칼로 무를 써는 것만 봐도 황홀해하며 그 자리에서 자지러지는 위인들이 있는지 모른다. 그래서 빅데이터 드라마를 쓰고 신화를 만드는지 모른다. 자료에 대해서는 쥐뿔도 모르면서 빅데이터라는 “연기”를 그럴듯하게 해서 사람들을 호러먹는 부류는 아닐는지... 천만 개가 넘는 자료를 분석한 결과를 감탄하며 경이롭게 바라보면 그 학생의 얼굴과 빅데이터를 팔고 다니는 바람잡이나 장사꾼들의 얼굴이 겹치는 것은 무슨 까닭일까?

그럼 슈퍼컴퓨터면 되겠니?

자료처리에 관한 지식과는 별개로 실제 자료를 처리할 수 있는 도구와 기술을 따져 보자. 컴퓨팅 파워를 생각해 보자. 빅데이터가 감당할 수 없을 만큼 크고 복잡하다니 컴퓨팅 파워가 가장 빠르다는 슈퍼컴퓨터 수준은 되어야 하지 않을까?

현재 슈퍼컴퓨터는 어떤 절대 컴퓨팅 파워 이상을 가진 컴퓨터가 아니라 <https://top500.org/>에 6개월에 한 번씩 발표되는 목록에 나와 있는 컴퓨터를 말한다. 작년 11월 기준 세계 최고 컴퓨터인 중국의 Sunway TaihuLight는 CPU라 할 수 있는 core가 1천만 개, 주메모리로 1.3 Peta 바이트를 가지고 있다. 이런 컴퓨터로 처리해야 하는 자료라면 충분히 빅데이터라고 할 수 있을 것이다. 일반 랩탑으로도 충분히 분석할 수 있는 자료라면 빅데이터라고 하기에는 좀 거시기하다. 그렇다면 빅데이터를 제대로 만끽할 수 있는 사람은 엄청난 컴퓨팅 파워를 가진 슈퍼컴퓨터 사용자란 말인가? 과연 몇 명이나 슈퍼컴퓨터를 사용해서 빅데이터를 분석하고 있는가?

그러면 대한민국은 현재 얼마나 컴퓨팅 파워를 가지고 있는가? 현재 한국 기상청에서 가지고 있는 Cray사의 Nuri와 Miri 각각 46등과 47등(1년 전에는 28, 29등)을 차지하고 있는데 core가 똑같이 7만 개다. 날씨 예측을 잘못한다고 온갖 비난을 받고 있는 그 컴퓨터다(사실 컴퓨터가 무슨 죄가 있나?). 그리고 대구경북과학기술원(DGIST)의 iREMB가 351등 (만 4천 core), 어느 제조업에서 사용하고 있는 컴퓨터가 404등 (만 6천 core)이다. 고작 이 네 대 뿐이다. 서울대학교의 천둥(2012년 기준 277등, 8천 core)은 벌써 순위 밖으로 밀려났다. 더 이상 슈퍼컴퓨터가 아니라는 소리다.

현재 미국, 중국, 일본이 슈퍼컴퓨팅을 주도하고 있다. 중국은 1등과 2등을, 일본

은 6등과 7등을 차지하고 있는데, 10위 밖에도 두 나라의 슈퍼컴퓨터는 준비하다. 물론 미국은 1, 2등을 놓쳤지만 누가 뭐래도 슈퍼컴퓨팅을 주도하고 있다. 빅데이터가 정말 크고 복잡한 데이터라면 한국은 미국은 물론 중국이나 일본 앞에 명함을 내밀기 어렵다. 닷넷과 골리앗 차이 그 이상이다. 슈퍼컴퓨터로 치면 한국은 아시아에서 사우디 아라비아보다도 못하다. 한국에 아무리 큰 빅데이터가 있다 해도 현재 답아서 처리할 만한 컴퓨팅 파워는 허망하리만치 초라하다. 무역 10대 강국에 전혀 걸맞지 않은 허접한 수준이니 하는 말이다.

그런데도 빅데이터 옹호자와 정부는 데이터가 엄청 크다고만 하고 슈퍼컴퓨터를 새로 사거나 만들자는 얘기는 하지 않는다. 알 수 없는 일이다. 3살박이 아이에게 하늘이 얼마나 큰지, 눈이 얼마나 오는지, 기차가 얼마나 긴지, 아버지를 얼마나 사랑하느냐, 우동을 얼마나 먹는지 차이가 없다. 똑같이 두 팔을 찢어지도록 벌리고 입으로 “이—만큼 많—이”라고 답할 뿐이다.

빅데이터 전용 분석법이 있나?

그러면 빅데이터를 분석하는 별도 방법이 있는가? 내 보기에는 그렇지 않다. 기존의 방법론을 적용하되 일반 사용자들도 쉽게 사용할 수 있도록 사용방법을 개선하고 시각화를 강조하는 듯하다. 10여년 전 미국에서 열린 SAS사용자 학회에 참석했을 때 SAS가 내세웠던 JMP (<http://jmp.com>)가 그러했다. 마우스로 푹푹 눌러서 원하는 분석 결과를 바로 알려주고 그래프로 표시해주는 것이 기억에 남는다. 발표자는 이제 더 이상 통계분석을 공부하지 않아도 되는 시대가 온다고 너스레를 떨었다. 그때 벌써 지금 빅데이터 옹호자들이 하고 있는 얘기를 했다는 점이 흥미롭다. 하지만 현란한 화면 뒤편에서 실제 이루어지는 분석은 기존의 통계기법일 뿐이다. 회귀분석이든 분산분석이든 T-test든 방법론은 그대로이고 어떻게 보여줄 것인가가 다를 뿐이다.

현재 미국에서 Data Science 혹은 (Data) Analytics라는 이름으로 학위 프로그램이나 학과가 생기고 있다. 모두 Big Data를 염두念頭에 두고 시류를 쫓고 있다. 예컨대, 인디애나 대학교는 몇 년 전부터 School of Informatics and Computing (<https://www.soic.indiana.edu>)에서 Master of Science in Data Science 프로그램을 운영하고 있다.

교과과정은 기존의 Computer Science, Informaiton and Library Science, Informatics 과목에 더하여 통계학으로 구성되어 있다. 통계학은 Introduction to Statistics, Exploratory Data Analysis, Statistical Learning and High-Deminsional Data Analysis, Bay-

sian Theory and Data Analysis, Applied Linear Models, Reproducible Results in Stats, Topics in Applied Statistics 등이다. 과목 이름을 자세히 살펴 보라. 컴퓨터과학, 사회정보학(informatics), 통계학 등이 잘 결합되어 있는 것은 사실이지만 빅데이터 학위 프로그램에서 가르치는 과목이 전혀 새로운 것이 아님을 알 수 있다. 예컨대, “빅데이터용 회귀분석”같은 과목이 없다. 기존 과목을 기존 교수진이 가르치고 있다. 같은 포도주를 예쁘장한 새 잔에 담은 셈이다.

요즘 방송 출연자들이 빅데이터를 분석한 결과라며 그래프를 가져와 설명하는 것을 종종 보게 된다. 기억컨대, 사회망서비스에서 얻은 자료를 바탕으로 시간단위로 빈도분석(frequency analysis)하여 추세를 따져보는 것이 전부였다. 방법론으로 치면 기초 수준의 분석방법이다. 관측치 수가 크고 멋진 그래프로 결과를 보여준다는 것 외에 뭐가 새롭다는 것인지 모르겠다.

인과관계냐 상관관계냐?

저자는 빅데이터가 상관관계를 말해줄 뿐이며 기존의 “스몰데이터”는 인과관계를 말해준다고 했다(99-100쪽). 그래서 상관관계를 말해주는 빅데이터가 “스몰데이터”에 비해 나올 것이 없다고 비판했다. 이러한 비판은 사실 공정하지 못하다. 빅데이터이든 스몰데이터든 분석방법이 다르지 않다면 인과관계를 말하지 않는다. 예컨대, 회귀분석(혹은 이런 류의 계량분석)은 독립변수와 종속변수 사이에 어떠한 인과관계가 있는지를 밝혀주지 않는다. 엄밀하게 말해서 회귀분석은 처음부터 인과관계를 가정하고 있지 않다. 결국 상관관계를 말해줄 뿐인데 마치 인과관계가 있는 것처럼 사람들이 해석할 뿐이다. 빅데이터든 스몰데이터든 마찬가지다.

변수 사이의 인과관계는 분석방법이 결정해주는 것이 아니라 연구자가 어떻게 자료가 발생되는가(data generation process, DGP) 혹은 변수가 어떤 속성을 가지고 있는가 등을 따져서 설정해주는 것이다(김수영 2016: 24). 연구자가 머리 속에 그리고 있어야 하는 큰 그림(분석틀)에 관한 문제다. 사물과 현상을 이해하는 이론에 관한 문제다. 결코 빅데이터냐 스몰데이터냐 문제가 아니다. 데이터는 데이터일 뿐이다. 이론을 가지고 있어야 데이터를 설명할 수 있고 그 가설이 맞는지 맞지 않는지를 검증할 수 있다. 이론이 없이 데이터를 이해할 수는 없는 일이다(154쪽).

요즘 통계분석기법과 컴퓨팅 성능이 고도로 발달했다 해도 컴퓨터가 알아서 데이터를 분석해서 원하는 결과를 내주지 못한다. 이렇게 데이터에서 그럴듯한 변수관계

나 모델을 끌어내는 낚시질(data fishing)은 유희遊戱일지는 몰라도 과학은 아니다. 하물며 분석결과를 해석하여 현실에 적용하는 일임에다.

항상 큰 것이 좋은가?

N(관측치 수)으로 치자면 빅데이터가 꼭 좋은 것도 아니다. 물론 자료를 설명할 때는(descriptive statistics) N이 큰 것이 유리하다. 하지만 분석모형을 통하여 통계추론을 할 때는(inferential statistics) 얘기가 달라진다. 어떤 분석모형에서 일정한 효과크기(effect size)와 통계증거력(statistical power)이 주어지면 적절한 표본크기(sample size)가 결정된다. N이 지나치게 크면 통계증거력이 너무 커서 분석할 필요성이 사라지고(의미없는 분석이 되고), N이 지나치게 적으면 신뢰성이 떨어지게 된다. N이 크면 클수록 좋다는 빅데이터를 찬양하는 것은 단지 무지거나 착각이거나 미신일 뿐이다.

중요한 것은 어떻게 대표성을 가진 랜덤 샘플을 뽑을 것인가와 분석모형에 따라 적절한 샘플수를 결정하는 일이다. 빅데이터는 이런 랜덤샘플링에 제대로 답하지 못한다. 예컨대, 빅데이터가 사회매체에서 얻은 대화나 반응에 의존하는 한 자기선택(self-selection)문제에서 벗어나기 어렵다. 따라서 빅데이터가 대표성을 갖기 어렵다. 빅데이터의 정신줄은 한마디로 N이 크면 클수록 좋다는 것이다. 덩치 큰 N으로 문외한門外漢인 청중을 오락질러(압도하여) 자신이 원하는(진리와 상관없는) 주장을 강요하기에 딱 알맞는 주술呪術이다.

빅데이터와 개인정보보호

“빅데이터가 너무 많아서 빅데이터 분석이 필요하다고 해놓고, 이제 와서 빅데이터가 부족하니 정부가 관리하고 있는 국민들의 개인정보를 공개하라고 한다. 그리고 또 말한다. 빅데이터 분석으로 개인을 식별할 수 없다고 말이다”(80쪽). 저자의 비판은 세 가지다. 하나는 빅데이터가 없어서 빅데이터 산업 부흥이 안되니 정부가 관리하고 있는 국민 정보를 공개하라는 논리 모순이다. 빅데이터 산업의 실패를 빅데이터가 아니라 공공 정보를 공개하지 않는 정부의 탓으로 돌리려는 술수라는 것이다(81쪽). 또 한 공공데이터 포털을 만들어 국민들의 비밀스러운 개인정보를 무방비 상태로 만천하에 공개하고 있다고 우려한다(77-78쪽). 마지막으로 “개인정보 비식별 조치”라는 것도 미봉책彌縫策이어서 “비식별 조치된” 개인정보 조각 조각을 끼워맞추면 재식별이 가능하다는 것이다. 이런 재식별도 못하는 빅데이터 분석 기술이라면 모든 개인 정보를 다 공개한다 한들 쓸모있는 분석을 할 수 없을테니(취도 못먹을 테니) 공공

정보를 공개하라고 목소리를 높이는 것이 앞뒤가 안맞는 주장이다(80-81쪽).

이런 비판을 곱씹으며 몇 가지 생각이 든다. 먼저 빅데이터를 그때 그때 편리한 대로 둘러댄다는 느낌이다. 보통은 사회매체에서 벌어지는 대화처럼 손질되지 않은 자료를 말하다가 정부의 공공 정보를 요구할 때는 잘 손질되거나 (1)과 (3)과 같이 덩치가 큰 자료를 지칭한다. 어쩌면 사람들이 생각했던 것보다 사회매체에서 쓸만한 것을 건지지 못했기 때문에 국민의 정보를 담은 공공데이터베이스를 공개하라고 다른 과녁으로 화살을 돌렸을는지 모른다.

저자의 우려에도 불구하고 한국 정부의 공공데이터 포털은 다행(?) 그냥 시늉내기에 머물고 있다. 국민 개인의 은밀한 정보도, 의미있는 정보를 뽑아낼 수 있는 자료도, 분석에 유용한 형태로 손질된 자료도 찾기 어렵다. 생색을 낼 뿐이다. 쉽게 말해 <http://data.go.kr>은 <http://data.gov>를 겹테기만 베꼈다고 말할 수 있다. 어찌어찌해서 그럴듯한 웹집을 만들기는 했는데, 수십 년 치 자료를 한꺼번에 만들 수는 없었던 것이다. 결국 지금까지 해오던 자료축적 수준 그대로를 보여줄 뿐이다. 의미없는 국가단위 통계, 제공 주체에 따라 띄엄띄엄 올려진 자료... 자료를 사용하라고 공개한 것인지, 사용하든 말든 내가 알 바 아니라는 소리인지 알쏭달쏭하다. 게다가 웹표준과 웹접근성(Web accessibility)과도 거리가 있어 한글마이크로소프트 제품을 사용하지 않는 사용자에게는 매우 불친절하다. 표준도 지키지 않고 여기 저기에 자바스크립트(javascript)을 덕지덕지 발라놨으니 말이다. 설령 국가기밀을 수두룩하게 올려놓았다 해도 멀쩡한 외국인이 빼가기 힘든 상황이다. 전자정부 3.0이 다른 것은 몰라도 자료보안에 각별히 심혈을 쏟은 것 같다. 누가 되든지 간에 쓸모있는 정보를 캐내기 어렵게 해놨으니 입에 침이 마르도록 칭찬해야 할는지... 그러니 김교수는 안심하고 두 다리 쪽 뺏고 주무시기 바란다.

한편 개인정보는 자료수집부터 저장, 수정, 사용(접근, 전달, 분석 등), 폐기 전 과정에서 보호되어야 한다. “감출 게 없으면 상관없다”는 논리(nothing-to-hide argument)는 자료를 수집하는데만 시선을 돌림으로써 자료처리 전 과정에서 일어날 수 있는 문제를 보지 못하게 한다(Solove 2011). 예컨대, 저자가 지적한 대로 “개인정보 비식별 조치”는 위협할 것 같지 않은 자료 조각을 붙이면 “재식별”이 가능하다. Solove (2011)는 이런 과정을 aggregation이라고 불렀다. 주민등록번호와 같은 식별정보를 지우거나 일부 자료만 공개하는 자료보안 기법은 빠르게 진화하는 자료 분석 기술에 무

력하다.

마지막으로 개인정보를 공개하는 자체가 위험하고, 그래서 해서는 안되는 일이라는 주장은 적절하지 않다. 법에서 정한 방법과 절차에 따라 개인 정보가 수집되고 저장되고 수정(공개)되고 사용되고 폐기되느냐가 중요하다. 개개인 자신이 정보가 어찌 처리되는지 전 과정을 지켜보고 의사 표시를 할 수 있는가를 물어야 한다. 개인정보를 공개하느냐 마느냐가 아니라 개인 정보 문제에 관하여 국회와 사법부가 어떻게 행정부를 감시하고 통제하는지를 따지는 문제이다.

행태주의와 거대주의는 왜?

저자는 미국의 빅데이터 유행은 행태주의 (behavioralism)와 거대주의 (gigantism)라는 두 신화에 기초하고 있다고 보았다(123-128쪽). 특히 행태주의는 관찰가능한 자료를 계량방법으로 분석하는데, 객관성으로 측정된 행태 자료 간의 상관관계를 발견하는 것이 행태주의의 근본정신이라고 주장했다(151쪽). 그런데 빅데이터가 밝혀내는 것이 인과관계가 아니라 상관관계니까 빅데이터는 행태주의 원리와 그 근본이 같다는 것이다(151쪽). 또한 빅데이터 사업자들은 계량분석을 위주로 한 행태주의 방법론을 이끌었던 통계소프트웨어 회사와 여론조사 회사라고 했다(11쪽). 말하자면 행태주의 = 객관성있는 경험자료 = 계량분석 = 상관관계 = 빅데이터라는 연관성이다. 또한 거대주의는 언제나 큰 것이 좋거나 끊임없이 탐욕을 멈추지 않는 자본주의로 이어진다. 하지만 이러한 상상과 비판은 지나쳐 보인다. 행태주의나 계량분석 자체가 문제가 아니라 그 근본을 잊고 엉터리로 받아들여 사용하는 것이 문제이지 않을까?

문제는 빅데이터가 아니다

사실 문제는 빅데이터가 아니다. 상식에서 가까운 자료조차 챙기지 못하는 것이 현실이다. 마땅히 있어야 할 자료를 만들지 않을 뿐더러 있던 자료도 치우고 지우고 고치는 세상이다. 마땅히 알아야 할 일을 모른다고 하고 기억나지 않는다고 하는 세상이다. 청문회, 검찰 조사, 법원의 재판에서조차 거짓이 난무하는 세상이다. 어쩌면 사실과 진실과 이성과 상식을 빅데이터라는 미신과 허구와 환영으로 덮어버리려는 세상일는지 모른다. 기본이 되어 있지 않은 자신이 한없이 부끄럽고 초라하하다가 영겁 곁에 지푸라기라도 잡는 심정으로 빅데이터를 끌어다가 뭐라도 있는 것처럼 허세를 부리거나 자기최면을 거는 것은 아닐까?

결국 빅데이터가 아니라 한 울, 한 조각이라도 멀쩡한 데이터가 아쉬운 세상이다. 당연히 있어야 하고 필요한 자료를 챙기는

것이 먼저다. 빅데이터를 논하기 전에 사회의 기본 정보를 알차게 축적하여 효율성있게 사용할 수 있는 제도를 말해야 한다. 자료를 모으고, 저장하고, 분석하고, 폐기하는 전 과정에서 개인정보가 보호될 수 있도록 제도를 다듬어야 한다. 상상을 초월하는 컴퓨팅 파워를 자랑하는 21세기의 정보관리 제도가 500년을 버텨온 조선왕조의 체제보다도 못한대서야 어디...

그냥 기본기에나 충실하라

중요한 것은 한마디로 기본기라 할 수 있다. 빅데이터까지는 바라지도 않지만 시민들이 원하는 일상의 정보를 제대로 구축하여, 알맞은 형식과 방법으로 필요한 때에 제공하는 것이 필요하다. 우리도 GSS와 CPS같은 자료를 마음껏 즐길 수 있는 날이 오기를 바란다. 또한 거창하게 빅데이터 분석기술을 따지기 전에 컴퓨팅 파워를 늘리고 기본에 해당하는 자료분석 방법부터 철저하게 익혀야 한다.

저자는 과거를 기록한 자료를 통하여 미래를 예측한다는 터무니없는 말에 귀기울이지 말 것을 권한다. 물론 사람들의 감정이나 의견이 아닌 사람들의 행동이나 물리현상을 분석하는 것은 타당하다. 정부가 나서서 빅데이터를 분석하여 재난과 범죄를 예방하려 애쓰기보다는 법과 절차에 따라서 공정하게 일을 처리해주었으면 한다. 규정대로 건물과 구조물을 살피고 주민들의 요구에 성심성의껏 대응해 주는 정부가 되었으면 한다. 그것만으로도 족하다. 그래서 최대주의 행정이 아니라 최소주의 행정을 구현했으면 한다. 

참고문헌

김동환. 2016. <빅데이터는 거품이다: 대한민국 빅데이터 유행을 말한다>. 서울: 페이퍼로드.
김수영. 2016. <구조방정식 모형의 기본과 확장: Mplus 예제와 함께>. 서울: 학지사.
Ginsberg, Jeremy, Matthew H. Mohebbi, Rajan S. Patel, Lynnette Brammer, Mark S. Smolinski, and Larry Brilliant. 2009. Detecting Influenza Epidemics Using Search Engine Query Data. *Nature* 457: 1012-1014.
Hindman, Matthew Scott. 2009. Blogs: The New Elite Media. In *The Myth of Digital Democracy*, 102-128, Princeton, NJ: Princeton University Press.
Laney, Doug. 2001. 3D Data management: Controlling Data Volume, Velocity, and Variety. *Application Delivery Strategies* 949 (6 February).
Levmore, Saul, and Martha C. Nussbaum, eds. 2010. *The Offensive Internet: Speech, Privacy, and Reputation*. Cambridge, MA: Harvard University Press.
Solove, Daniel J. 2011. *Nothing to Hide: The False Tradeoff Between Privacy and Security*. New Heaven, CT: Yale University Press.