

Bruno J. Strasser, *Collecting Experiments: Making Big Data Biology*. Chicago: University of Chicago Press, 2019. xv + 404 pp.*

최석현†

브뤼노 스트라세(Bruno J. Strasser)의 책 『실험의 수집: 빅데이터 생물학의 탄생(*Collecting Experiments: Making Big Data Biology*)』은 최근 수십 년 동안 생명과학이 겪고 있는 이른바 ‘빅데이터 혁명’의 역사적 맥락에 관한 설명을 시도한다. “통설에 따르자면, 이러한 전환은 주로 기술 변화의 차원에서 설명되어 왔다.” 책의 저자는 이런 설명이 충분하지 않다고 지적하며, 생명과학이 현재 겪고 있는 일을 조금 더 큰 맥락, 특히 ‘장기지속(longue durée)’의 차원에서 이해해야 한다고 말한다. 생명과학의 데이터화는 오랫동안 이루어진 생명과학 인식론과 방법론의 변화라는 “보다 심원한 역사적 힘”의 일시적 작용일 뿐이기에 그렇다. 따라서 그는 질문을 바꾸어 이렇게 묻는다. “에당초 생명과학자들은 왜 실험 데이터를 수집하고, 비교하고, 분류하기 시작했던 것일까”(8)?

실험주의(experimentalism)의 확산과 그로 인한 분자생물학과 생명공학의 부상을 강조하는 것은 20세기 생명과학사의 가장 전형적인 서술 방식 중 하나이다. 이 같은 전형적인 서사에 따르면, 실험주의는 “클로드 베르나르(Claude Bernard)가 잘 알려진 바와 같이 실험과학과 관찰과학을 구별하고 전자만이 인과적 설명을 제공할 수 있다고”(11) 천명한 이래 20세기 초반을 즈음해 유전학, 생화학, 세균학 등 다양한 분야에서 ‘실험실 혁명’을 일으키며 관찰과학을 밀어내고 생명과학 연구의 주된 방법으로 자리잡았다. 이에 따라 다양한 동식물의 외양을 수집하고 분류하는 데에만 천착해 왔던 생명과학 연구 풍토에 일대 혁신이 일어났다. ‘실험실 혁명’은 화학과 물리학의 발전과 실험 기법 및 기기의 발달, 종전 이후 정부에서 투입한 막대한 연구비와 연구 인력의 유입에 힘입어 보다 강력한 힘을 갖추게 되었고, 이 모든 흐름이 마침내 분자생물학과 생명공

† 서울대학교, sokionchoi@snu.ac.kr

* 서평을 작성하는 과정에 도움을 주신 배상희, 이두갑 선생님께 감사드립니다.

학이라는 결실을 맺음으로써 생명과학은 오늘날의 모습으로 거듭난 것이다.

하지만 만일 20세기 생명과학사의 큰 흐름이 박물관으로 대표되는 소박한 수집 문화로서의 생물학에 대한 실험실 과학자들의 도전이었다면, 이들은 왜 다시 ‘빅데이터’ 과학에, 그러니까 예전의 수집가들처럼 수많은 자료를 수집하고, 분류하고, 비교하는 작업에 뛰어들었을까? 이 같은 질문을 던지면서 스트라세는 기존 현대 생물학사 내러티브에 의혹을 제기한다. 외견상으로는 두 ‘앎의 방식(ways of knowing)’ 사이에 넘지 못할 벽이 있었던 것 같지만, 사실 “수집하기, 비교하기, 분류하기 등의 앎의 방식들은 …… 생명과학의 역사를 통틀어, 실험적 앎의 방식에 의해 밀려난 게 아닌” 것일지도 모른다는 것이다. 그는 오히려 반대로 생각해보자고 제안한다. 요컨대 20세기 내내 수많은 생명과학자가 “실험적 지식의 데이터베이스 활용에 중점을 두고, 실험과 수집의 동반 상승(synergy)을 일으키는 하이브리드를 만들어왔다”는 것이다(18). 만일 20세기 생명과학사를 이와 같이 이해할 수 있다면, 방대한 실험 데이터를 수집, 분류, 비교하는 ‘빅데이터 생물학’의 출현은 갑작스러운 단절보다는 연속적인 변화의 결과에 가까울 것이다. 두 앎의 방식 사이의 ‘하이브리드화’는 실험주의 또는 실험적 앎의 방식이 출현한 19세기 말 이후 계속해서 진행되어 온 일이고, 빅데이터 생물학은 이러한 장기지속의 일부에 불과한 일이기 때문이다.

이러한 ‘하이브리드화’는 한편으로는 인식론적 차원에서 이루어졌지만, 다른 한편으로는 사회적인 차원에서 이루어졌다. 저자는 이에 대해 “수집은 …… 무엇보다도 공간적 실행(spatial practice)”(24)이라는 점을 강조한다. 그런데 이는 두 가지 의미에서이다. “첫째로, 과학자들은 공간적으로 분산된 이질적인 사물들을 한 장소에 저장된 통약가능한 데이터로 변환시키는 중대한 과제에 부딪힌다. …… 둘째로, 전문가와 아마추어를 포함하는 다양한 사람들 또한 ‘수집될 필요가 있다’(20). 이것은 “어떤 과학적 수집 활동의 성패 여부는 연구 대상(objects) 혹은 데이터를 수집할 수 있는 능력과 이를 유지할 수 있게 해주는 커뮤니티를 발전시키는 능력 둘 모두에 달려 있기”(114) 때문이다. 요컨대 ‘하이브리드화’는 실험과 수집이라는 두 ‘앎의 방식’ 사이에서뿐만 아니라 두 앎의 방식이 수반하는 서로 다른 사회적 조직 형성 방식 사이에서도 만들어진 것이다.

책의 각 장은 모두 이와 같은 ‘하이브리드’의 구성에 관한 탐구에 할애되었다. 우선 책의 첫 두 장에서 저자는 20세기 전반의 생명과학사를 실험과 수집의 두 가지 앎의 방식 간의 투쟁으로 규정한 기존의 생명과학사 서술을 문제삼으

며, ‘하이브리드화’는 이 시기에도 이미 진행 중이었음을 보여준다. 먼저, 제1장 “살아있는 박물관(Live Museums)”에서는 1960년대 이전 구축된 ‘실형체 컬렉션(stock collections)’의 여러 초기 사례를 검토한다. ‘비교학적(comparative)’ 얹의 방식과 ‘실험적(experimental)’ 얹의 방식 사이의 중요한 차이 중 하나는, 전자가 다양한 생물종을 수집해 비교함으로써 생명 현상의 보편적 특질을 찾으려 하는 반면, 후자의 경우 실험 목표에 더 유리한 특질을 가진 하나의 생물종에 집중한다는 점이다. 그렇게 박테리아, 쥐, 옥수수, 초파리, 바이러스 등 다양한 생물종이 ‘모델 생물(model organism)’의 지위를 얻게 되었다. 1장에서 저자는 새로운 모델 생물의 발굴이 언제나 해당 생물에 대한 실험체 컬렉션의 구축을 동반했음을 보여준다. 일반적인 서사와는 반대로, 실험실의 팽창은 오히려 컬렉션의 역할 또한 강화시켰다. 실험 연구자가 많아질수록 이들의 연구 대상은 지속적으로 표준화되어야 했고, 컬렉션은 이러한 표준, 혹은 “두 번째 자연(second nature)”⁽²⁵⁾을 제공해 주었다. 마찬가지로 실험적 얹의 방식이 많은 지역에서 자리를 잡을수록 이들 간의 교류는 어려워졌고, 컬렉션 운영 주체들은 소식지(newsletter) 발간을 통해 같은 대상을 연구하는 사람들 사이의 커뮤니티 형성을 도모했다.

제2장 “혈액 은행(Blood Banks)”에서는 다양한 생물종의 특질을 수집, 비교, 분류함으로써 이들 간의 관계를 규명하는 계통분류학(systematics) 혹은 계통발생학(phylogeny) 분야에서 실험적 얹의 방식을 받아들여 스스로의 연구를 쇄신하려 했던 이들, 특히 생명체의 해부학적 구조나 겉보기 특질이 아닌 혈액의 특질을 바탕으로 생물 진화를 이해할 수 있다고 주장했던 이들의 작업을 다룬다. 이들은 자연사에 실험적 방법을 도입해 “그 분야를 보다 정량적인, 정밀한, 그리고 객관적인 것으로 만듦으로써 자연사를 개혁”(74)하려 했다. 하지만 그들의 정체성은 계속 자연사가로 남아 있었는데, 그들이 어디까지나 다양한 생물종 간의 관계를 규명하는 데에 관심을 가졌을 뿐 아니라 계속해서 분류학 커뮤니티 안에 있으면서 그들과 교류했다는 점 등이 이를 뒷받침한다. 1장과 2장은 대칭적 구조를 이루는데, 요컨대 1장이 실험실 생물학의 탄생 과정에서 실험체 컬렉션이라는 일종의 ‘박물관’이 중요한 역할을 수행했음을 보여준다면, 2장은 비교학적 생물학이 현대화되는 과정에서 실험 방법론의 도입이 중요한 역할을 수행했음을 보여준다. 박물관의 실천 없이는 실험실의 실천이 만들어질 수 없었고, 실험주의 방법론 없이는 비교학적 생물학 또한 앞으로 나아가기 힘든 상

황이었다. 이렇게 책의 첫 두 장은 20세기 초반에 일어났던 변화가 “하나의 문화에 의한 다른 문화의 정복이라기보다는 ‘하이브리드 문화’의 출현”(107)으로 개념화되어야 함을 잘 보여준다.

책의 중반부는 서로 상이한 삶의 방식 사이의 하이브리드가 사물들을 모으는 방식에 대해서뿐만 아니라 사람들을 모으는 방식에 대해서도 나름의 대안을 구성해 왔음을 보여준다. 먼저, 제3장 “데이터 아틀라스(Data Atlases)”는 최초의 단백질 서열 데이터베이스인 마거릿 데이호프(Margaret O. Dayhoff)의 『단백질 서열 및 구조 아틀라스(*Atlas of Protein Sequence and Structure*)』의 사례에 집중한다. 1965년 출판되어 오늘날 빅데이터 생물학에서 활용하는 DNA 혹은 단백질의 “서열 데이터에 대한 최초의 공식적인 컬렉션”인 이 책은, “생명과학 [역사상] 최초로 대량의 실험 데이터에 대한 컴퓨터화된 분석의 가능성을 열어주었다”(113). 데이터의 수집과 저장 그리고 인쇄에 이르기까지 많은 과정에서 컴퓨터의 도움을 받았지만, 『아틀라스』는 어찌 되었건 2장에서 다루었던 비교생화학 전통의 영향을 받아 다양한 생물종의 데이터를 비교할 수 있도록 구조화되었으며, 무엇보다도 전통적 매체인 책의 형태로, 그것도 자연사가들이 오랫동안 활용해 온 방식인 아틀라스 형태로 배포되었는데, 이런 특징을 들어 스트라세는 ‘생물정보학(bioinformatics)’의 주요 기원 중 하나인 데이호프의 작업이 “예전의 연구 실행들과 깊은 연속성”(153)을 갖고 있었음을 설득한다.

그런데 『아틀라스』의 관건이 오직 방대한 실험 데이터를 효과적으로 구조화하는 데에만 있었던 것은 아니다. 스트라세는 “도덕 경제(moral economy)” 개념을 활용하여 데이호프가 그의 작업을 위해 다양한 단백질의 서열과 구조 데이터뿐만 아니라 수많은 사람들을 또한 수집해야 했음을 보여준다. 저자는 사회사가 에드워드 파머 톰슨(Edward Palmer Thompson)과 생물학자 연구자 로버트 콜러(Robert E. Kohler)의 논의를 참조하면서, 도덕 경제를 “특히 소유권(property), 사유권(privacy), 우선권(priority) 등의 쟁점에 결부된 과학 지식 교환의 기저를 이루는 가치 체계”(327, n. 102)를 지칭하는 용어로 정의한다. 그리고 책의 3장에서 스트라세는 이러한 도덕 경제의 문제야말로 데이호프가 『아틀라스』를 구축하는 과정에서 겪은 가장 큰 문제였다고 설명한다. “데이호프가 자신의 데이터베이스에 대한 판권을 취득했고, 재배포를 제한했으며, 그로부터 수익을 추구했고, 『아틀라스』에 수록된 데이터를 스스로의 연구에 활용했다는 사실은, 일부 실험가에게는 실험과학의 도덕 경제에 대한 위협으로 여겨졌

다”(145-146). 직접 산출해낸 데이터를 통해 과학자로서의 ‘인정(credit)’을 얻는 것을 당연하게 여긴 실험가들에게, 자신의 손으로 실험을 수행하지 않는 데이호프의 “이론적 접근법”은 그저 “사무적인” 작업에 불과한 것으로, 과학으로서의 가치가 떨어지는 것이었다 (141-142). 하지만 데이호프의 입장에서 보면 사정은 그렇지 않았다. 아틀라스 구축에는 실험가들이 짐작했던 것보다 훨씬 높은 전문성이 요구되었으며, 보통은 아마추어인 데이터 생산자보다 이를 종합한 사람이 더 많은 인정을 가져가는 것은 자연사에서는 오랜 관습이기도 했다. 3장에서 논의한, 『아틀라스』 구축 과정에서 데이호프가 겪은 어려움은 데이터 생산자와 수집가가 상상하는 서로 다른 도덕 경제 사이의 충돌에 기인한 것이었고, 이러한 충돌의 한복판에서 “데이호프는, 다른 한편, 실험가도 이론가도 아닌 실험적 지식의 큐레이터라는 새로운 역할을 개척해 나가고 있었다”(142).

다음 장인 “가상 컬렉션(Virtual Collections)”에서도 스트라세는 수집의 문제가 사물들에 관한 것이면서 동시에 사람들에 관한 것이기도 하다는 점을 보여준다. 4장은 단백질 구조에 대한 디지털 컬렉션인 ‘단백질 데이터 은행(the Protein Data Bank, PDB)’의 사례에 초점을 맞춘다. 단백질 결정학(crystallography) 연구자들은 흔히 스스로의 작업을 ‘따분한(tedious)’ 것으로 묘사하곤 했다. 이 따분함, 그리고 그 원인인 단백질 “구조 결정의 여러 단계에서 나타나는 뻔뻔한 계산의 필요성은 왜 결정학자들이 처음으로 컴퓨터를 활용한 연구자 집단 중 하나가 되었는지 설명해준다” (158-159). 컴퓨터의 도입과 함께 20세기 이전까지 자연사의 한 분야로 여겨져 온 결정학은 다시 한번 비교연구의 얇은 방식을 받아들였고, 실험주의적 결정학의 과제인 단백질의 분자 구조를 조직화하는 원리 규명은 수많은 단백질 구조 데이터를 비교하고 분류하는 작업과 합류하게 되었다. 이처럼 PDB는 과학 공동체가 단백질 결정과 컴퓨터라는 상이한 행위자들과 상호작용한 결과 출현한 ‘하이브리드’였다.

PDB의 경우에도 도덕 경제의 문제는 중요한 화두였다. 데이호프에게 그랬듯, 에드거 마이어(Edgar F. Meyer, Jr.)와 헬렌 버먼(Helen M. Berman)에게도 과학자로서의 인정을 위한 핵심 자원인 데이터를 선뜻 내어주지 않는 실험가들을 ‘수집하는’ 것은 쉽지 않은 일이었다. 그럼에도 그들은 서서히 그러나 끈질기게 이 문제를 해결해 나갔다. 버먼의 “공동체주의적(communitarian) 이상향”(168)에 감명받은 월터 해밀턴(Walter Hamilton)은 가장 먼저 ‘수집된’ 이들 중 한 사람이었다. 스트라세는 해밀턴의 합류가 PDB에 대한 결정학자들의 우

력을 잠재우는 데에 중요한 역할을 했다고 평가한다. 즉 “해밀턴이 단백질 결정학자가 아니라는 사실이, 마거릿 데이호프와 『단백질 서열 및 구조 아틀라스』 이후 반복된 우려, 즉 데이터 뱅크 관리자들이 자신의 연구를 위해 데이터를 이용할 것이라는 우려를 가진 이들을 안심시켰을 수 있다”(170). 하지만 반감을 줄이는 것만으로는 충분하지 않았다. “명백하게도 공동체주의적 에토스가 결정학자들로 하여금 데이터를 공유하게 만들 정도로 충분한 유인책은 아니었다.” (182) 십여 년에 달하는 PDB 관리자들의 노력이 꽃을 피운 것은 1981년 국제결정학연맹(International Union of Crystallography, IUCr)에서 연맹이 운영하는 학술지에 단백질 구조 관련 논문을 투고하려면 PDB에 데이터를 등록해야 한다는 규정을 입안하면서였다. 특히, 이 규정을 입안하고 얼마 후 IUCr은 데이터 공개를 4년까지 유예할 수 있도록 기준을 완화했는데, 이는 직접 만든 데이터를 충분히 사용해보기 전에 다른 연구자에게 이를 넘겨주는 것에 대해 거부감을 갖고 있던 다수의 실험실 결정학자들을 동참하게 만들기에 충분한 조치였다. 이후 다른 학술지들도 이런 흐름에 합류하면서, PDB는 과학 연구의 공동체주의적 이상향을 현실화하는 데에 성공한 사례로서 역사의 이정표로 남게 되었다.

오늘날 ‘젠뱅크(GenBank)’로 불리는 유전자 서열 데이터베이스 구축의 주도권을 둘러싼 마거릿 데이호프와 월터 고드(Walter B. Goad) 사이의 경쟁을 다룬 5장에서도 저자는 비슷한 방식의 설명을 이어간다. 이미 『단백질 서열 및 구조 아틀라스』를 운영한 경험이 있는 데이호프는 전문성 면에서 고드보다 더 유리한 위치에 있었다. 하지만 결국 이 프로젝트는 고드에게 돌아갔다. 스트라세는 여기에서도 도덕 경제의 문제가 중요한 요인이었다고 분석한다. 고드의 승리는 “비교학적 얇은 방식, 즉 자연적 대상들을 수집하고 비교하는 일의 필요조건을 기존 실험 과학의 도덕 경제에 성공적으로 적응시킨”(225) 덕분이었다는 것이다. 데이호프는 『아틀라스』 때와 마찬가지로 데이터에 대한 권리가 데이터베이스를 구축한 큐레이터에게 주어져야 한다고 생각했다. 반면 고드는 수집가만큼이나 생산자에게도 충분한 권리가 있으며, 따라서 데이터베이스에 대한 권리는 큐레이터 개인이 아니라 그것을 생산하고, 관리하고, 사용하는 학계 전체에 귀속되어야 한다고 이야기했다.

이런 차이는 두 사람의 제안이 채택한 물리적인 모델의 차이로도 이어졌는데, 데이호프가 여전히 “자기 테이프와 책 형태의 인쇄된 서열을 보내줌으로써”

데이터를 배포하려 했던 것과 달리 고드의 팀은 온라인 네트워크를 사용하겠다고 선언했다 (220-221). 또 데이호프는 부족한 정부 예산을 보완하기 위해 기업의 후원이나 유료 구독 제도의 도입을 고려했지만, 고드는 그럼에도 모든 서비스를 무료로 제공해야 한다고 설명했다. 고드의 이 같은 사고방식은 데이터 생산자로서의 권리를 빼앗기고 싶지 않았던 실험실 과학자들에게나 바이돌 법안 (Baye-Dole Act)의 통과 이후 과학계가 상업에 종속될 것을 우려했던 과학 커뮤니티 일반 및 국립보건원(NIH, National Institutes of Health) 등 연구행정 기관들에게나 더 매력적인 선택지로 여겨졌다. 고드의 승리는 그가 더 유능한 데이터베이스 관리자였기 때문이라기보다는 변화하는 시대적 감수성을 읽어내고 더 많은 이들을 동맹으로 끌어들이는 데 성공한 덕분이었다.

이상 3장에서 5장까지의 논의를 통해 스트라세는 사물들을 수집하는 것만큼이나 사람들을 수집하는 것이 데이터 컬렉션 구축의 성패에 있어 중요한 관건이었음을 보여준다. 특히, 저자는 여기에서 사물들의 수집만큼이나 사람들의 수집 또한 충분한 ‘기술(technique)’을 요하는 일임을 드러낸다. 각자의 이해관계를 가진 사람들을 한 곳에 모으기 위해서는 단지 이념이나 규범에 호소하는 것만으로 불충분하다. 사물들을 모으려면 그것들을 질서 있게 배열할 수 있게 해주는 개념적, 물질적 구조가 뒷받침되어야 하듯이, 사람들을 모으기 위해서는 그들 사이에 질서를 부여해 줄 수 있도록 소유권, 사유권, 우선권 등을 분배하는 구조를 마련할 필요가 있다. 이 지점에서 저자가 도덕 경제의 개념이 머튼 규범(Mertonian norms)에 대해 갖는 우위를 명시적으로 정리하지 않은 점은 아쉽다. 스트라세가 설명하듯이, 도덕 경제는 “국지적으로 또 역사적으로 위치 지어져 있으며 따라서 과학 공동체들 사이에서 차이를 보일 수 있다는”(327, n. 102) 점에서 머튼의 ‘보편적’ 규범 개념과 다르다. 그런데 두 개념 사이의 또다른 중요한 차이로 머튼 규범이 정신적인 차원에 있는 것인 데에 비해 도덕 경제는 ‘경제’라는 표현이 시사하듯이 일종의 하부구조의 층위에 자리한다는 점을 지적할 수 있다. 저자가 잘 보여주는 것처럼, 도덕 경제는 소유권, 사유권, 우선권 등을 적절히 분배함으로써 행위자들 사이의 이해관계 정렬을 이루어내는 일종의 인터페이스다. 요컨대 새로운 도덕 경제는 일종의 “관심끌기 장치(device of intersement)”, 즉 “동맹자들을 한 장소에 묶어두는 방법”으로, 확립된 도덕 경제는 “등록(enrolment)” 장치, 즉 동맹자들의 “역할을 규정하고 조직화하는 방법”으로 기능한다는 점에서 머튼 규범과는 상이한 함의를 갖는다.¹ 머튼

규범은 일종의 정신이기 때문에, 이 개념에 의존하는 경우 올바른 규범 질서를 확립하는 것이야말로 과학을 더 좋게 만드는 방법일 테다. 반면, 도덕 경제는 충돌하는 이해관계들을 정렬하는 시스템에 더 가까우며, 이 개념을 받아들이는 경우에는 어떤 규범을 확립하는 것을 넘어서 상이한 이해관계를 갖는 행위자들의 관심을 끌고 이들을 등록시킬 수 있는 ‘기술’을 정교하게 갈고닦는 것이야말로 과학적 실천의 관건이 될 것이다.

책의 마지막 장인 6장에서 저자는 지금까지 논의한, 20세기 초반에서 1980년대에 이르는 장기지속의 역사가 어떻게 오늘날의 세계와 연결되는지를 논의한다. 이 장에서 스트라세는 미국의 젠뱅크와 이에 상응하는 타국의 데이터베이스들인 유럽 분자생물학 연구소(European Molecular Biology Laboratory, EMBL)의 뉴클레오티드 서열 데이터 라이브러리(Nucleotide Sequence Data Library) 그리고 일본 DNA 데이터 은행(DNA Data Bank of Japan, DDBJ) 등이 밀어붙인 오픈 액세스 방침이 어떻게 오늘날의 과학 정보 배포 플랫폼까지 이어졌는지 살핀다.

물론 젠뱅크를 비롯한 새로운 세대의 데이터베이스는 출범 초기 『아틀라스』와 PDB 등이 겪은 것과 같은 어려움을 피할 수 없었다. 하지만 이들에게는 참조할 역사가 있었다. “단백질 데이터 은행이나 『아틀라스』 같은 과거의 경험은 (저자권authorship 등) 적절한 보상이 없다면, 연구자들에게 데이터를 공공화하지 않을 많은 유인이 있다는 점을 명확히 알려주었다”(252). 젠뱅크와 EMBL 라이브러리의 운영자들은 따라서 학술지 편집자들의 관심을 이끌어냈다. 학술지에 논문을 게재하기 위해서 젠뱅크 등에 데이터를 제출하도록 규정하자고 제안한 것이다. 하지만 과학 지식의 ‘문지기(gatekeeper)’ 역할을 독점하고 있던 학술지들의 입장에서 권력을 나누자는 말과 마찬가지로 이러한 제안을 흔쾌히 받아들일 리는 없었다. 학술지의 닫힌 문을 여는 데에 먼저 성공한 쪽은 EMBL이었다. 하이텔베르크 EMBL 연구소 소속이었던 실험생물학자 퍼트리샤 칸(Patricia Kahn)은 먼저 학술지가 저자들에게 수탁 번호를 제공하면, 저자들이 다시 이 번호를 통해 데이터베이스 측에 자료를 제출할 수 있게 하는 방안을 제시했는데, “이는 어떤 제약조건도 포함하지 않았으므로”(233) 학술지의 편의를

1 Michel Callon, “Some Elements of a Sociology of Translation: Domestication of the Scallops and the Fishermen of St. Brieuc Bay,” *The Sociological Review* 84:S1 (1984), 196-233 중 특히 206, 211.

해지지 않았고, 따라서 학술지 측에서도 받아들이 수 있는 것이었다. 변화의 물결은 이후 편집진 일부가 오픈 액세스 이념에 동조적이었던 『핵산 연구(Nucleic Acid Research)』가 EMBL 자료 제출을 의무화하고, 뒤이어 젠뱅크가 『국립과학원회보(Proceedings of the National Academy of Sciences, PNAS)』 측을 설득해 데이터 제출을 의무화하도록 규정하는 데에 성공하면서 보다 거세졌다. 물론 이들 학술지가 데이터 제출 규정을 위반한 논문에 어떠한 제재도 가하지 않았기 때문에 이를 강제 규정으로 보기는 어려웠지만, 그럼에도 불구하고 데이터 제출 의무화는 큰 파급효과를 가져왔다. 젠뱅크의 경우를 보면, 1985년에는 출판된 데이터의 19%만을 수집하는 데에 그쳤지만, 불과 5년 뒤인 1990년에는 데이터의 75%를 확보하기에 이르렀던 것을 알 수 있다(229, 235).

스트라세는 젠뱅크, EMBL, DDBJ의 성공이 이후 과학의 오픈 액세스화를 촉진하는 데에 영향을 끼쳤다고 주장한다. 젠뱅크를 비롯한 여러 성공 사례의 존재가 인간 유전체 프로젝트(Human Genome Project)를 필두로 한 여러 공공 과학 프로젝트의 데이터 공개에서 ‘퍼브메드 센트럴(PubMed Central)’과 ‘퍼블릭 라이브러리 오브 사이언스(PLoS, Public Library of Science)’같은 학술 자료 데이터베이스 및 오픈 액세스 플랫폼의 탄생에 이르기까지 이후의 여러 사건에 중요한 참조점을 제공해 주었다는 것이다. 이처럼 젠뱅크를 위시한 생명과학 데이터베이스의 공동체주의적 이상향에 주목하면서, 스트라세는 지식에 대한 자유로운 접근과 활용이라는 유토피아적 전망의 또다른 계보를 발굴해낸다. 결론에서 저자가 언급하는 “1950년대의 아방가르드 상황주의자들과 1970년대의 컴퓨터 해커들”(263)의 사례는 보다 널리 알려져 있지만, 생명과학 데이터 컬렉션의 도덕 경제가 오늘날 떠오르는 새로운 지식 배포 패러다임에 끼친 영향에 대해서는 그간 많은 이목이 집중되지 못했던 것이 사실이다.² 스트라세는 이처럼 오픈 액세스 패러다임이 상황주의자와 컴퓨터 해커의 미래주의뿐만 아니라 과학자들의 다소 전통적인 이상주의에, 더 나아가 실험체 컬렉션과 혈액 은행을 구축했던 100여 년 전의 실험실 문화와 박물관 문화에 뿌리를 내리고 있음

2 대표적으로 해커 문화에 관한 고전적인 연구인 Christopher M. Kelty, *Two Bits: The Cultural Significance of Free Software* (Durham: Duke University Press, 2008)와 해커 문화와 대항문화(counterculture) 간의 관계를 다룬 Fred Turner, *From Counterculture to Cyberculture: Stewart Brand, the Whole Earth Network, and the Rise of Digital Utopianism* (Chicago: University of Chicago Press, 2006) 등을 언급할 수 있을 것이다.

을 보여줌으로써 데이터 ‘혁명’의 장기지속적 기원을 잘 보여주었다.

방금 이야기한 것처럼, 데이터 혁명 또는 ‘데이터 홍수(data deluge)’의 환상에 사로잡히지 않고 장기지속의 관점에서 생명과학의 데이터화를 설명하는 것은 이 책의 가장 중요한 의의 중 하나이다. 빅데이터 과학에 관한 세간의 담론은 오늘날 인간이 지식을 습득하는 과정 자체가 전례 없는 혁명을 겪고 있는 양 이야기한다. 하지만 스트라세는 이러한 변화조차 사실 더 큰 역사적 흐름의 일부일 뿐임을 강조한다. “최근 생명과학의 전환을 전례 없는 것처럼 여겨서는 안 되며, 장기지속의 관점에서 보아야 한다”(20). 저자가 인용하는 로버트 단턴(Robert Darnton)의 문장이 특히 이러한 관점을 잘 요약해준다. “모든 시대는 정보의 시대였다. 각기 다른 방식으로 말이다”(9).

물론 역사 서술에 있어 이런 관점을 채택하는 것이 스트라세 혼자만은 아니다. 『데이터사들(Data Histories)』이라는 제목으로 출간된 『오시리스(Osiris)』의 2017년 특집호 서문에서, 해당 특집호의 편집자들은 “데이터사(data history)”라는 새로운 역사 서술의 범주를 개발해 보자고 제안한 바 있다. “데이터가 어떻게 수집되는지, 어디에 저장되는지, 누가 그것을 관리하며, 데이터가 어떻게 처리되고 재조합되는지, 데이터 교환의 실행 경제와 과학적, 정치적, 문화적 맥락에서 이를 사용하는 바에 관한 도덕 경제는 어떤지” 등을 질문하는 데이터사는 한편으로는 “다종다양한 실행들, 기술들, 문화들을 장기지속의 맥락에서 고려할 수 있게 하는 새로운 개념적 공간”을 만들어주면서, 다른 한편으로는 비슷한 문제에 대해 고민하면서도 서로 고립되어 있었던 역사 서술의 여러 분과 사이의 “학제적 대화를 촉구”하는 데에 그 의의가 있다.³ 스트라세의 책은 현대 생명과학사 담론에서 흔히 접할 수 있는 기술결정론에 반대해 역사적 장기지속의 맥락에 주목하는 한편, 20세기 내내 생명과학 연구 데이터의 수집, 저장, 관리, 처리, 재조합 등에 관한 실행과 규범이 어떤 방식으로 조직되어 왔는가를 상세히 분석한다는 점에서 데이터사의 흐름에 합류하는 것으로 보인다.

이처럼 스트라세의 연구가 다른 데이터사 서술과 방법론 차원에서 공명하는 점과 비교해 볼 때, 그가 데이터의 다른 ‘역사들’과 생명과학 데이터의 역사 간의 연결고리를 충분히 강조하지 않은 것은 아쉬운 점이다. 예를 들어 저자가 언급한 바와 같이 오픈 액세스 패러다임의 형성에는 생물학의 도덕 경제뿐만 아

3 Elena Aronova, Christine von Oertzen, and David Sepkoski, “Introduction: Historicizing Big Data,” *Osiris* 32 (2017), 1-17 중 특히 15, 12.

나라 60년대 아방가르드 문화와 컴퓨터 해커의 도덕 경제가 모두 중요한 영향을 행사했는데, 책에서 이들과 생명과학 사이의 교차점을 조금 더 강조해 주었다면 어땠을까 하는 아쉬움이 있다. 마찬가지로 생물학과 언제나 다양한 영향을 주고받으며 변모해 왔을 물리학이나 화학 등 인접 분야와의 상호작용 또한 충분히 깊이 있게 탐구되지 못한 듯하다. 특히 책의 곳곳에서, 예컨대 스스로를 “히피”라고 칭하는 헬렌 버먼과 그의 친구들이 내세운 “공동체주의적 이상향”에 월터 해밀턴이 감화되었다는 서술 (168), PDB 출범 초기 “사용자의 절반만이 결정학자였고, 나머지 절반은 직접 실험을 수행하지 않으면서 단백질 구조를 연구하는 컴퓨터 과학자, 물리학자, 화학자, 생물학자 등이었다”(173)는 서술, 그리고 퍼브메드가 “로스앨러모스가 1991년에 물리학 연구를 위해 창설한 arXiv를 모형으로 하여”(250) 제안되었다거나 화학, 물리학, 생명과학에서 오픈 액세스 방침이 확산되는 과정이 달랐을 것이라는 언급(254) 등 생명과학과 다른 분야의 역사 간의 흥미로운 교차점을 엿볼 수 있음을 고려해 보면, 스트라세의 책이 생물학의 도덕 경제가 갖는 자율성을 필요 이상으로 부각시키고 있는 것은 아닐까 싶다.

앞에서 인용한 『오시리스』 데이터사 특집호의 마지막 글에서, 공동 저자로 참여한 스트라세는 글의 제목을 통해 도발적인 질문을 던진 바 있다. “빅데이터가 답이다. …… 그래서 질문은 무엇인가?”⁴ 브뤼노 스트라세의 책 『실험의 수집』은 이 ‘질문에 대한 질문’에 대한 답이다. 책은 빅데이터라는 ‘답’을 도출하기까지 생명과학자들이 던져온 ‘질문’들을 보여주면서, 오늘날 우리가 어떤 질문을 던져야 하는가에 대해 숙고해 보기를 제안한다. 책의 마지막 문장이 다음과 같이 스트라세가 이 책을 통해 하고 싶었던 이야기를 요약해준다. “따라서, 만일 수집과 실험을 결합한 생의학 지식의 새로운 체제가 다시 한번 과학 지식의 생산에 참여할 수 있는 사람들의 범위를 넓히는 데에 성공한다면, 바야흐로 더욱 심원한 역사적 전환이 시작될 것이다”(272). 요컨대 빅데이터 과학이 진정으로 ‘열린’ 과학으로 거듭나려면 더 많은 데이터, 더 많은 연구 대상을 수집해야 할 뿐만 아니라 더 많은 사람들을 끌어들어야만 한다. 빅데이터 과학의 미래는 ‘하이브리드’를 어떻게 구축할 것인가에 달려 있는 것이다.

4 Bruno J. Strasser and Paul N. Edwards, “Big Data Is the Answer … But What Is the Question?,” *Osiris* 32 (2017), 328-45.