

# 원문DB 구축예산 150억 돌파의 의미

## - 새로운 국면 맞은 국회도서관 자료의 데이터화

국회도서관장 허용범

yb22032@nnet.go.kr



저작자표시(BY)

한 달여 전쯤, 나는 서울대학교로 오세정 총장을 찾아갔다. 서울대가 대대적으로 추진 중인 인공지능(AI) 연구에 당연히 필수적인 데이터를 어디에서 얻을 것인가, 우리 국회도서관의 것을 이용하라고 먼저 제의했다. 국회도서관이 구축하고 있는 방대한 지식정보 데이터는 국가적 자산이므로, 서울대학교가 AI 연구에 선도적으로 이 데이터를 이용할 수 있도록 필요한 데이터를 전면적으로 공유하겠다는 뜻이었다.

물리학자 출신으로, 내년엔 서울대 인공지능 대학원을 설립하는 등 인공지능 연구의 중요성을 역설해 온 오 총장은 그 뜻을 흔쾌히 받아들였다. 오 총장과의 대화에 이어 9월 중순에는 김명환 서울대 중앙도서관장이 국회도서관을 방문하며, 국립 서울대학교와 국가도서관인 국회도서관 간의 교류협력 강화방안에 대해 보다 구체적인 논의를 이어갔다. 서울대와 국회도서관의 AI-데이터 협력추진은 4차 산업혁명이라는 시대적 과제를 극복해 가는 데 국가 기관 간의 전면적 협력이라는 상징적이고 실질적 의미가 있을 뿐만 아니라, 소위 '미래도서관'이 나아가야 할 방향과 관련해 시사하는 점도 대단히 크다고 나는 믿고 있다.

### “양적 확장에서 질적 도약으로”

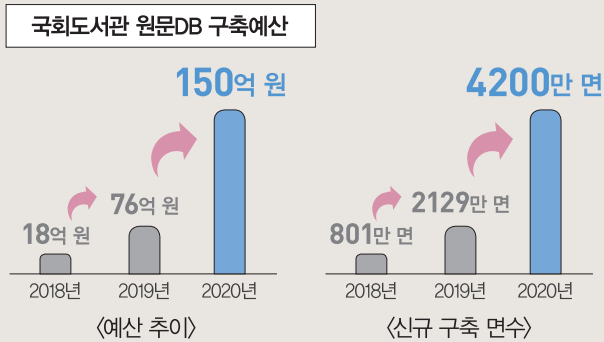
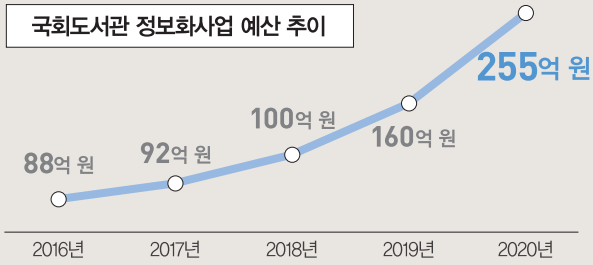
국회도서관이 서울대에 데이터 제공의사를 먼저 적극적으로 제의할 수 있는 것은 우리 도서관의 디지털 능력 확충 덕분이다.

모든 도서관자료가 디지털화되고, 그 이용 또한 디지털 방식으로 이뤄져야 하는 시대에 도서관 자료의 원문DB 구축은 그 모든 것의 기반이자 출발점이라고 할 수 있다. 국회도서관은 이 원문DB 구축예산을 올해(2019년) 76억 원을 확보해, 지난 3년간 내내 18억 원에 그쳤던 것에 비해 320%를 순증시켰다. 이로써 원문DB를 올해에만 2100만 면을 새롭게 구축할 수 있게 되었고, 9월 현재 60% 이상의 순조로운 진척을 보이고 있다.

내년(2020년)은 이 원문DB 구축예산이 정부안으로 올해에서 다시 100% 증액된 150억 원이 책정, 상정되어 있다. 2018년까지의 18억 원과 비교할 때 2년 만에 730%가 증가하는 것이다. 이 예산을 확보하면 우리는 내년엔 4200만 면에 이르는 원문DB를 다시 신규로 구축할 수 있게 되고, 이로써 국회도서관이 구축·보유한 원문DB 총량은 마침내 3억 면을 돌파하게 된다. 단일 도서관의 원문DB 구축예산이 150억 원에 이르고 총 구축량이 3억 면을 넘어선다는 것은 '도서관 자료의 데이터화'가 전혀 새로운 국면으로 진입했음을 의미한다. 단순한 양적 증가가 아니라 질적 도약의 단계로 들어섰다는 뜻이다. 지금까지의 원문DB 구축예산은 어떤 질적 논의의 하기는 턱없이 적은 액수였다. 그러나 150억 원 정도 되는 예산이므로, 이제는 지식정보의 보고로서 도서관이 생산해낼 수 있는 최고 품질의 데이터, 즉 '검증된 양질의 데이터'를 계획적으로 생산해 낼 수 있게 되었다.

디지털 도서관으로서의 서비스를 충실히 해 나가려면 우선 그 서비스의 기반으로서 원문DB 구축량이 절대적으로 많아야 함은 물론이다. 국회도서관도 지금까지는 그런 수준을 벗어나지 못했다. 그러나 이제는 일단 양을 늘리는 식의 단계를 뛰어넘어, 서울대와 같은 특정한 수요자의 요구나 국가적 필요성, 미래 도서관 서비스의 방향 등에 맞춘 '맞춤형 데이터'로서 원문DB를 구축할 수 있게 된 것이다. 이것이 원문DB 구축예산 150억 원 확보의 가장 중요한 의미이다.

특히 지금까지는 예산상 엄두를 내기 어려웠던 고품질의 데이터셋 구축도 본격적으로 할 수 있게 되었다. 국회도서관은 이미 2년 전부터 한국전자통신연구원(ETRI)의 인공지능 프로젝트인 엑소브레인 개발에 법률데이터



를 데이터셋 형태로 제공해오고 있다. 그러나 앞으로 그 영역은 법률을 넘어 지식정보 전 영역으로 확대될 것이다. 우리는 국회도서관이 망라적으로 수집하는 입법·학술·정책 자료를 수요자와 시장의 필요에 맞춰 체계적으로 DB화하고, 이 양질의 디지털 데이터를 사회 곳곳에 개방하고 제공해 나갈 계획을 수립해 놓고 있다.

### ‘맞춤형 데이터셋’ 생산 가능

국회도서관이 보유한 이 입법·학술·정책자료는 현재만 175만 책에 이르는 방대한 양이 쌓여 디지털화를 기다리고 있고, 매년 신규로 엄청난 양이 새롭게 수집되고 있다. 국회도서관이 보유한 이 자료가 중요한 것은, 명확한 저작권을 갖고 상업적 판매물이 대다수인 일반 단행본 책과는 본질적으로 다른, 일종의 공공저작물이기 때문이다. 더욱이 이들 세 부류의 자료는 국가정책 개발이나 인문 사회과학적 연구, 국회의 입법에 필수적인 3종 세트로서, 우리나라에서 국회도서관만이 망라적이고 체계적으로 수집하고 있다. 따라서 이를 원문DB화해 서비스한다는 것은, 오래된 고서(古書)나 희귀도서, 그림 등을 디지털화해 보존하는 것과는 원문DB 구축이라는 이름만 같았지 전혀 다른 목적과 용도인 것이다. 말하자면 국회도서관의 원문DB는 과거 자료의 단순 보존용이 아니라, 살아 있고 활용되는 현재의 지식정보인 셈이다. 나는 국회도서관의 이 입법·학술·정책자료를 100% 디

지탈화해 도서관 이용자 누구나 언제 어디서나 활용할 수 있도록 해야 한다고 줄곧 역설해 왔다. 일본은 국립국회도서관의 원문DB를 인공지능 개발에 활용할 경우 저작권에서 자유롭도록 법까지 2018년에 개정해 놓았다. 현재 우리 도서관이 보유한 175만 책과 신규 수집물량을 전면적으로 원문DB로 구축해 어느 정도 규모를 갖춘 국가DB 자산으로 만들 경우, 연간 200억 원씩 우선 5년간 1000억 원 정도가 필요한 것으로 추산되고 있다. 국회도서관은 그동안 기획재정부 예산실을 비롯한 많은 곳에 이 입법·학술·정책자료의 전면적 DB화 필요성을 호소해 왔고, 그 성과로 올해 150억 원의 예산을 편성 받게 된 것이다.

지금까지 여러 차례 공언해 왔듯이, 우리는 이 원문DB가 국가적 자산이므로 적극적으로 개방되고 공유되어 실질적으로 활용되어야만 한다고 믿고 있다. 그것이 국가도서관으로서 국회도서관이 해야 할 당연한 책무라고 여기고 있다. 우리가 서울대를 먼저 찾아가 데이터 제공의사를 제시한 것은 이런 의미였다. 우리는 조만간 OCLC(Online Computer Library Center)와 같은 세계적 도서관 자료 공유서비스에도 우리가 구축한 원문DB를 체계적으로 올려, 세계인들에게 열려 있는 ‘대한민국 지식정보의 창’이 되도록 그 개방과 공유의 지평을 넓혀 나갈 계획이다.

원문DB 구축은 이처럼 국회도서관이 ‘대한민국 최고의 디지털 도서관’을 지향하고 세계적 수준의 도서관으로 도약하려는 데 있어서 그 기반이자 출발점이다. 또한 이 DB를 대한민국 모든 도서관과 연구소, 대학, 민간 기업까지 개방하고 공유하는 그 자체가 ‘4차 산업혁명 선도 국가중심도서관’으로서의 사명을 실천하는 길이다. 원문DB 구축은 현대 도서관이 제공하는 여러 서비스 중 단지 한 분야를 의미하지 않으며, 혁명적 변화를 겪고 있는 도서관 서비스의 생명력과 같은 것이다.

나는 이 같은 원문DB 구축 역량이 획기적으로 밀받침이 되고, 여기에 첨단 IT기술을 활용한 도서관 서비스 시스템이 지속적으로 개발·적용된다면, 도서관의 혁명기로 불리는 현대에 있어서 국회도서관은 소위 ‘미래도서관의 표준’을 열어가는 선도자로 자연스럽게 자리매김할 것이라고 믿는다. 🌍