

NIE Issue Brief

NIE IB 22-02(통권 16호) 2022년 11월 30일 | 발행처 : 국립생태원 | 발행인 : 조도순 | www.nie.re.kr

EcoBank 생태 빅데이터의 공유 및 활용

저자 | 생태정보팀 선임연구원 권용수



기존의 빅데이터는 규모(Volume), 속도(Velocity), 다양성(Variety)으로 특징지을 수 있는 정보 자원으로 구분하며, 단순히 데이터 처리 응용 소프트웨어로는 수집·저장·분석·처리하기 어려울 정도로 방대한 양의 데이터를 의미하였다. 그러나 최근 빅데이터를 통한 가치 창출의 중요성이 증대함에 따라 가치(Value), 진실성(Veracity), 정확성(Validity), 휘발성(Volatility)을 더해 7V로 확대되는 등 필요에 따라 계속 변화되고 있다. 이러한 특성으로 인해 Frank(2012)는 그의 저서에서 빅데이터를 "Today's big data is not Tomorrow's big data"라고 표현하며 빅데이터의 유동성을 강조하였다. 이에 따라 최근의 '빅데이터'는 대용량 데이터를 활용하고 분석하여 가치 있는 정보를 추출하고, 생성된 지식을 바탕

으로 능동적으로 대응하거나 변화를 예측하기 위한 전 과정을 의미하는 광의의 의미로 사용되고 있다(조 등, 2012). 최근 많은 언론매체를 비롯해 일상생활 전반에서 쉽게 접하고 있는 '빅데이터'는 4차 산업혁명을 견인하는 핵심 동력으로써 정보통신기술(CT)의 급속한 발전과 함께 그 중요성이 대두되고 있다. 특히 정보통신기술(CT)의 발전과 함께 사회가 복잡한 구조를 가짐에 따라 정부는 지금까지의 방식으로는 풀 수 없는 새로운 사회문제에 직면하고 있으며, 미래사회의 불확실성에 대응하여 새로운 가치를 창출할 핵심 자원으로써 빅데이터에 주목하고 있다. 이를 위하여 전 세계적으로 빅데이터 산업 육성을 위한 다양한 정책 수립과 플랫폼 구축을 위한 노력이 정부 차원에서 이루어지고 있다.

빅데이터 공유를 위한 국내·외 노력

디지털 경제로의 전환이 가속화됨에 따라, 주요국들은 디지털 경제를 구성하는 기본 요소인 데이터 산업에 대한 지원을 강화해나가고 있다(K-data, 2019; NIA, 2020). 미국은 오픈 데이터 정책을 통해 누구나 데이터를 활용하여 새로운 가치를 창출할 수 있도록 지원하고 있으며, 대표적으로 2009년 오바마 정부에서 시작된 Data.gov(<https://data.gov/>)를 서비스하고 있다. 더불어 국가 데이터 저장 네트워크 구축(Data Storage Network)과 데이터 관련 법 제정 및 시행으로 데이터 개방 활성화를 위한 제도적·기술적 토대를 마련하고 있다. 유럽연합(EU)은 유럽 위원회(European Commission)를 중심으로 공공데이터를 취합·공유·활용하도록 하고 있으며, 공공데이터 개발 전략(Open Data Strategy)을 수립하여 모든 공공데이터에 대한 온라인 개방을 의무화하고 있다. 또한 일본의 경우 2021년 '디지털청'을 발족하며 사회 전반의 디지털화를 위해 민관 협력을 통한 데이터 산업 발전 및 디지털 전환을 추진하고 있다. 더불어 중국은 기업 성장 지원을 통한 데이터 산업을 육성하고자 '빅데이터 발전 촉진 행동 요강(2015년)'을 발표하는 등 국가의 전략적 자원이자 핵심적 혁신 요소로 빅데이터를 인식하고 데이터의 공유·개방 확대 및 안전보장을 목표로 하고 있다.

국내에서는 지난 2011년 공공데이터포털(<https://www.data.go.kr/>) 서비스를 시작으로 공공정보를 적극적으로 공개하면서, 공공의 혁신과 함께 국민의 알 권리 보장과 공공 빅데이터의 민간 활용 활성화에 힘쓰고 있다. 특히 정부는 '정부3.0(박근혜 정부)', '한국판 뉴딜(문재인 정부)', '디지털 플랫폼 정부(윤석열 정부)' 구현과 같은 국가 정보화 및 전자정부 정책을 통해 국가 주도의 공공 빅데이터 개방 활성화를 추진하고 있다. 공공 빅데이터 개방은 국가의 미래 전략을 수립하고 사회 현안을 해결하는 역할 뿐 아니라, 민간에게 제공하는 서비스의 질 향상과 기업의 경쟁력과 생산성 향상 등에 큰 역할을 하고 있다. 더 나아가 정부는 공공 빅데이터를 활용 또는 다른 분야와 융합하여 새로운 재화나 서비스를 창출하거나, 이를 통해 국민의 삶의 질 제고와 신사업 및 일자리 창출로 인한 경제 활성화에 기여하는 것을 목표로 하고 있다(전과 김, 2017).

특히 포스트 코로나 시대 정보통신기술(ICT)을 중심으로 한 디지털 전환의 중요성이 높아지며, 국내 기업들을 위한 수요 및 일자리 창출, 국가 혁신역량 강화로 이어질 수 있는 디지털 경제로의 전환이 가속화되고 있다. 더불어 빅데이터 활용 요구 증가로 공공기관의 데이터 개방 확대 필요성이 점차 증가하고 있다. 그러나 국내에서는 일부 기관들에서 생물종 정보 등 일부 단편적인 생물자료만을 제공하고 있을 뿐 생태계 전반에 대한 정보의 공유·활용은 부족한 실정이다. 이에 국립생태원에서는 국내·외 생태 빅데이터의 체계적인 구축 및 통합정보 제공을 통해 4차 산업혁명시대의 초석을 마련하고, 국민의 알권리 및 생태복지 증진에 기여하고자 국내 최초로 ICT 기반의 생태정보

종합은행인 EcoBank(에코뱅크; <https://nie-ecobank.kr/>) 시스템을 구축하고 서비스를 실시하였다.



[그림] EcoBank 시스템 구축 배경

EcoBank를 통한 생태 빅데이터 공유

1) EcoBank 시스템 구축

기후변화 및 환경오염과 같은 전 지구적인 환경변화가 더욱 다양하고 복잡해짐에 따라 생태계 관련 문제들을 효과적으로 해결하기 위해서는 다양한 생태 빅데이터의 종합적 수집을 통해 생태계 현황에 대한 정확한 이해가 필요하다(최와 김, 2002). 이를 위해서는 다양한 정보 주체(시민, 학자, 정부 등)가 보유한 광범위한 생태 빅데이터들을 체계적으로 관리하고 공유·활용할 수 있도록 하는 통합 플랫폼의 필요성이 높아졌다(성 등, 2018). 이에 국립생태원은 생태 빅데이터 활용 플랫폼 구축을 위해 2015년 정보화전략계획(ISP)을 수립하고 5년간의 개발을 거쳐 2019년 12월 EcoBank의 정식서비스를 개시하였다. EcoBank는 사용자 계층별(시민, 정책입안자, 연구자) 맞춤형 서비스를 통해 국립생태원에서 조사·연구한 생태 빅데이터와 그 외의 국내·외 주요 생태 빅데이터들을 하나의 온라인 공간에서 공유 및 활용할 수 있도록 하고 있다. 또한 환경부 소속·산하기관(국립생태원, 국립생물자원관, 국립공원공단, 국립낙동강생물자원관)의 자연환경관련 정보를 한곳에서 보여 줄 수 있도록 수요자 제공 중심 자연환경조사 정보 통합서비스 기반을 마련하였다.



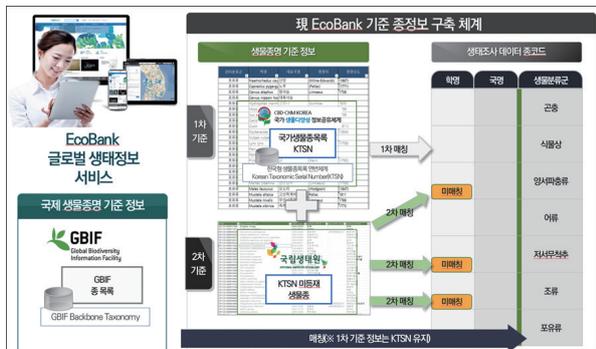
[그림2] ICT 기반 생태정보 플랫폼 서비스

특히 EcoBank에서는 국내 최대 생태 조사 데이터인 전국자연환경조사 자료를 비롯해 생태계 정밀조사, 습지조사 현황자료, 장기생태, 로드킬 현황 및 분석자료를 개방하고 있으며, 생태·자연도를 지도 및 도엽별 공간분석 보고서와 함께 제공하고 있다. 또한 시민참여형 데이터인 조류 유리창 충돌사고 자료, 수원시 하천생물상, 전국자연환경조사 시민참여 자료를 구축하여 제공하고 있으며, EcoBank의 국제적 활용성 증대를 위해 국외(베트남) 맹그로브숲 생태정보도 서비스하고 있다.

2) EcoBank 생태 빅데이터 표준화

데이터 표준화는 정보시스템별로 산재해 있는 데이터 정보 요소에 대한 명칭, 정의, 형식, 규칙에 대한 원칙을 수립·적용하는 것으로, 데이터의 정확한 의미를 파악할 수 있게 할 뿐만 아니라 데이터에 대한 상반된 시각을 조정하는 역할을 수행한다(권 등, 2020). 더불어 ① 데이터에 대한 의사소통 증대, ② 데이터의 소재 파악 용이, ③ 데이터 품질 향상, ④ 데이터 변환 및 정제 비용 감소와 같은 긍정적 효과를 기대할 수 있다.

EcoBank에서는 국내 생태 데이터의 품질을 확보하고 국내 생물·생태·자연환경 관련 정보시스템을 비롯해 국외 데이터 정보시스템과의 정보공유 활성화에 기여하고자 데이터 표준화 방안을 마련하였다. 데이터 표준화는 제2차 ~ 제4차 전국자연환경조사 자료를 대상으로 ① 용어 표준화, ② 테이블 통합, ③ 불필요한 속성 및 입력 오류 제거, ④ 조사자별 상이한 입력항목 단일화, ⑤ 데이터 코드화, ⑥ 입력항목별 코드 매핑을 수행함으로써 데이터를 정제하고 국내 최초의 생태데이터 표준안 마련의 기반을 수립하였다. 더불어 생물종명에 대한 기준정보로 국립생물자원관의 KTSN(Korean Taxonomic Serial Number)을 사용하고 있으며, KTSN에 미등재된 종명에 대해서는 국립생태원 자체 종목록(전국자연환경조사 기준)을 적용하고 있다. 또한 국외 생태정보 수집 및 서비스의 확대를 위하여 GBIF(Global Biodiversity Information Facility)의 생물종 목록을 함께 구축하여 적용하고 있다.



[그림3] 생태 빅데이터 생물 중정보(종목록) 일치

3) EcoBank 생태 빅데이터 품질관리

일반적으로 빅데이터가 많은 양의 데이터를 지칭하는 개념으로 받아들여지고 있으나, 단순히 데이터의 양이 많다고 해서 활용도가 높아지는 것은 아니다. 최근에는 빅데이터를 통한 가치 창출의 중요성이 증대하면서 데이터의 접근성과 신뢰성 확보가 해당 빅데이터의 활용성을 결정하는 가장 중요한 요소로 평가받고 있다.

데이터의 접근성은 무수히 많은 데이터를 중에서 사용자가 필요로 하는 데이터를 일정한 기준에 따라 빠르고 정확하게 식별할 수 있도록 해주는 메타데이터가 필수적이다. 메타데이터는 데이터의 데이터로서 데이터의 출처, 종류, 구조 등 해당 데이터를 정의하는 속성의 집합을 지칭한다. 즉, 메타데이터는 책상 위에 흩어진 데이터라는 책을 속성에 따라 정리해주는 책장이라고 할 수 있다. EcoBank 또한 이러한 메타데이터의 중요성을 인식하여 한국정보통신기술협회(Telecommunications Technology Association, TTA)로부터 전국자연환경조사 등 생태 연구 데이터에 대한 메타데이터 구성표준¹⁾을 획득하여 생태조사·연구에서 획득되는 다양한 형태의 생태 빅데이터를 수집·저장·분석·관리하는데 활용하고 있다.

한편 데이터를 개방하기 위해서는 데이터의 오류 추출 및 정제를 통하여 데이터 품질이 확보되어야 하며, 이를 위해서 정확한 품질진단과 개선작업이 병행되어야 한다. EcoBank의 데이터 품질진단은 6개 분야(값, 구조, 성능, 표준화, 연계, 관리체계)를 대상으로 진행하게 되며, 고품질의 데이터를 제공하고 향후 주기적인 데이터 품질관리를 체계적으로 진행할 수 있도록 하고 있다. 또한 연구데이터의 생산 단계에서 개방까지의 원 수준의 관리 기준 및 체계 확립을 위해 연구데이터 관리 지침을 제정(22.12월 예정)하고 '23년부터는 연구데이터의 일관된 품질관리를 위한 데이터 관리계획(Data Management Plan, DMP)을 작성 및 이행하도록 계획하고 있다.

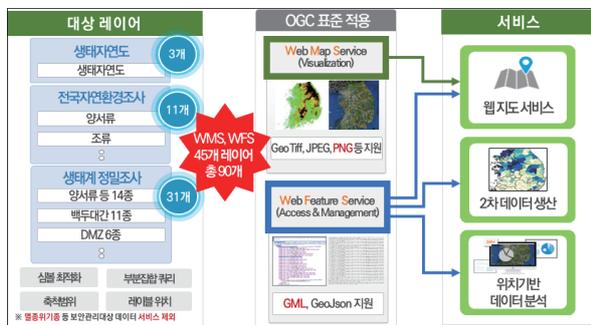


[그림4] 생태 빅데이터 품질관리 체계

1) 생태 연구 데이터 관리 및 공유를 위한 메타데이터(TTAK.KO-10.1249)

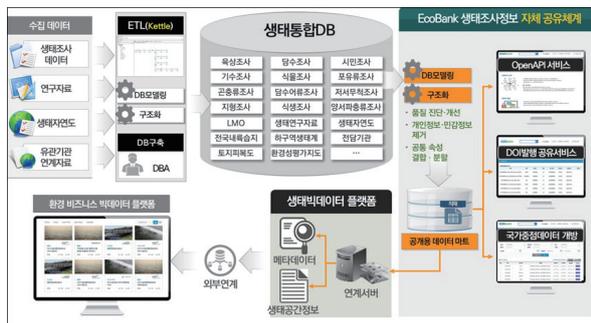
4) EcoBank 생태 빅데이터 개방 · 공유

인터넷상에서 등록된 연구자료는 웹페이지의 폐쇄와 같은 이유로 손실되어 데이터 복구가 어려워지기도 한다. EcoBank는 이러한 데이터 손실을 방지하고 영구적인 추적이 가능하도록 데이터에 DOI²⁾(Digital Object Identifier)를 부여하고 개방하고 있다. DOI는 사람의 주민등록번호와 같아서 언제나 데이터의 소재를 파악할 수 있게 해주며, 새로운 연구자료를 생성할 때 참고할 수 있는 정보로도 유용하게 쓰인다. 또한 EcoBank에서는 다양한 방식(OpenAPI³⁾, ETL⁴⁾, off-line 등)을 통해 국내 · 외 생태 빅데이터를 구축 · 연계하여 고품질 공공 빅데이터를 민간으로 확대 개방하고 있다.



[그림5] 생태공간정보의 OpenAPI 연계 서비스

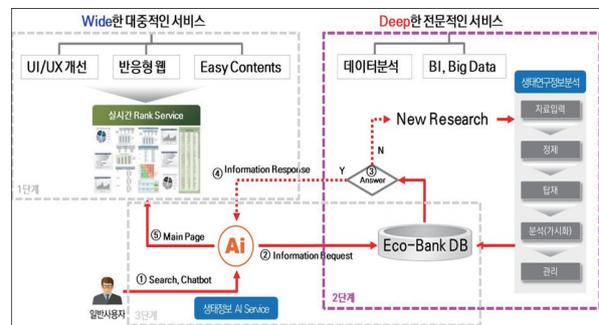
한편 생태통합DB에 적재 관리되고 있는 이들 생태 빅데이터는 비공개 민감정보 등을 포함하고 있어 외부에 공유하기 위해서는 별도의 가공처리가 필요하다. 따라서 서비스 목적 및 데이터 특성에 따라 외부 공개 · 공유를 위해 임시데이터 또는 파일을 데이터마트의 형태로 별도의 가공처리를 거쳐 지속적인 활용과 체계적인 관리가 될 수 있도록 하고 있다.



[그림6] 생태 빅데이터 공개 · 공유용 데이터마트

5) EcoBank 생태 빅데이터 활용

연구데이터는 생산하고 쌓아놓는 것만으로는 죽은 정보가 되어 버린다. 따라서 살아있는 데이터는 구축, 갱신 및 지속적인 활용을 요구한다. 최근의 빅데이터 분석 기법의 발달은 이러한 연구데이터에 새로운 가치를 부여하고 창출하도록 하여 사회문제의 해결 및 미래사회의 불확실성에 대응할 수 있는 해법으로 제시되고 있다. 특히 디지털 융복합 기술(Deep Learning, AI 등)을 이용한 EcoBank 생태 빅데이터 분석은 생태관련 정책 마련을 위한 과학적 근거를 제시하고 생태 빅데이터 활용성 증대를 위한 데이터에 기반한 정책자원을 강화하는 수단으로 사용될 수 있을 것이다.



[그림7] EcoBank 발전 방향

<참고문헌>

권용수, 송교홍, 김목영, 김기동. 2020. 전국자연환경조사 데이터 정제와 표준화 방안 연구. 생태와환경. 53(4): 380-389.
 성선용, 권용수, 김기동. 2018. 생태정보의 공유를 위한 생태정보 포털서비스(EcoBank) 구축 및 활용 방안. 생태와환경. 51(3): 212-220.
 전병진, 김희웅. 2017. 공공 빅데이터 개방 및 활용 활성화 방안에 대한 연구. 정보화정책. 24(3): 27-41.
 조지연, 김예진, 박건철, 이봉규. 2012. 생태계 관점에서의 빅데이터 활성화 위한 구조 연구. 한국IT서비스학회지. 11(4): 277-294.
 최병길, 김성수. 2002. GIS를 이용한 강화지역 자연생태정보관리시스템. 대한공간정보학회지. 10(3): 123-130.
 한국데이터산업진흥원(K-data). 2019. 2019 데이터산업 백서. 한국데이터산업진흥원.
 한국정보화진흥원(NIA). 2000. 2019 빅데이터 플랫폼 및 센터 데이터 표준화 가이드 : 빅데이터 연계 활용 조사 분석서. 한국정보화진흥원.
 Franks, B. 2012. Taming the big data tidal wave. Wiley.

2) DOI(Digital Object Identifier; 객체의 디지털식별자): 인터넷 주소가 바뀌어도 사용자가 그 문서의 새 주소로 찾아가도록 웹 파일이나 인터넷 문서에 영구적으로 부여된 식별자
 3) OpenAPI: 인터넷 이용자가 일반적으로 웹 검색 결과 및 사용자인터페이스(UI) 등을 제공 받는데 그치지 않고, 누구나 접근하여 직접 응용 프로그램과 서비스(지도서비스 등)를 개발할 수 있도록 하는 연계방식
 4) 시스템에서 시스템으로 데이터를 이동시키는 방식, 추출(Extract), 변환(Transform), 적재(Load)