

Original article

전국자연환경조사 데이터 정제와 표준화 방안 연구

권용수 · 송교홍¹ · 김목영 · 김기동^{2,*}

국립생태원 에코뱅크팀, ¹국립생태원 자연환경조사팀, ²국립생태원 생태정보연구실

A Study on the Data Cleaning and Standardization of National Ecosystem Survey in Korea. Yong-Su Kwon (0000-0002-2573-5922), Kyohong Song¹ (0000-0002-1751-4302), Mokyoung Kim (0000-0001-6690-2486) and Kidong Kim^{2,*} (0000-0001-8948-6999) (EcoBank Team, National Institute of Ecology, Seocheon 33657, Republic of Korea; ¹Team of National Ecosystem Survey, National Institute of Ecology, Seocheon 33657, Republic of Korea; ²Division of Ecological Information, National Institute of Ecology, Seocheon 33657, Republic of Korea)

Abstract Research on diagnosing and predicting the response of ecosystems caused by environmental changes such as artificial disturbance and climate change is emerging as the most important issue of biodiversity and ecosystem researches. This study aims to clean, standardize, and provide the results of National Ecosystem Survey which should be considered fundamentally in diagnosing and predicting ecosystem changes in the form of dataset. To refine and clean the dataset we developed a simple verification program based on the fifth National Ecosystem Survey Guideline and applied that program to the data from the second (1997~2005), third (2006~2013) and fourth (2014~2018) National Ecosystem Survey. Data quality control processes were implemented including (1) standardization of terminology, (2) similar data table integration, (3) unnecessary attribute and error elimination, (4) unification of different input items, (5) data arrangement in codes, and (6) code mapping for input items. These approaches and methods are the first attempt propose an option for ecological data standardization in Korea. The standardized dataset of National Ecosystem Survey in Korea will be easily accessible, reusable for both researchers and public. In addition, we expect it will contribute to the establishment of diverse environmental policies concerning environmental assessments, habitat conservation, prediction of endangered species distribution and ecological risks due to climate change. The dataset through this study is open freely online via EcoBank (nie-ecobank.kr) which is the first ecological information portal system in Korea developed by National Institute of Ecology.

Key words: data quality control, ecological information, EcoBank, data management

서론

최근 생태계, 생물다양성, 자연환경 분야 연구와 관련하여 생물과 생태에 대한 데이터와 정보가 국가적, 지역

적, 국제적 계획들을 통해 매우 광범위하게 공유되고 있다 (Global Biodiversity Outlook 4, 2014). 또한, 자연환경 체계에 관한 데이터가 점점 더 복잡해지고 방대해짐에 따라, 이러한 자료들을 수집, 관리, 분석, 시각화, 공유하는 방안에 대한 관심도 점차 커지고 있다 (Bingham *et al.*, 2017). 그러나, 데이터를 자유롭게 공유하고 사용할 수 있다 해도 기존의 데이터 네트워크에 통합할 때 공통의 표준

Manuscript received 11 December 2020, revised 16 December 2020, revision accepted 17 December 2020
* Corresponding author: Tel: +82-41-950-5620, Fax: +82-41-950-6143
E-mail: kidong@nie.re.kr

© The Korean Society of Limnology. All rights reserved.

This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provide the original work is properly cited.

(standards)과 관례 (practices)를 사용하지 않는 경우가 있으며, 이러한 표준의 부재로 이후의 데이터 분석을 더 어렵게 하고 있다(Hobern *et al.*, 2019).

생물다양성과 생태계 연구관련 조사 데이터를 축적하는 것은 대부분의 연구자에게 일반적이고 보편적인 과정이다. 그러나, 연구자들이 데이터를 수집하거나 분석한 형식 그대로의 원시데이터로 저장하는 것을 선호하거나 규격화된 자료 형식이 차후 데이터 재사용에 유리하다는 것을 인식하지 못하기 때문에, 축적된 데이터들의 재사용에는 한계가 있는 것이 현실이다. 그러므로 데이터의 폭넓고 다양한 재사용과 체계화된 저장을 허용하는 데이터 표준을 설계하기 위해 데이터 생산자, 데이터 재 사용자 및 데이터 관리자 간의 상호협력체계를 수립하고 유지하는 것이 매우 중요하다(Poisot *et al.*, 2019). 또한, Zimmerman (2008)에 의해 데이터 표준화는 과학적 지식을 공공의 영역으로 전달하는 하나의 수단이라고 제안되었으며, 생물다양성 협약(CBD: Convention on Biological Diversity)에서도 공통의 데이터 표준과 관례를 통해 향후 데이터의 이동과 접근을 강화하고 촉진할 것을 권장하고 있다(Secretariat of the Convention on Biological Diversity, 2014). 데이터 표준화 규칙에 따라 정제과정을 거쳐 구조화된 데이터는 다른 데이터와 연계하여 통합하기 쉬우며, 새롭고 독창적인 해석과 분석을 위해 기존 데이터의 1차 목적을 변경하여 더욱 유용하게 활용할 수 있다. 또한, 데이터 처리 코드를 쉽게 재사용 할 수 있어 소요되는 시간과 노력을 절약하게 되고, 원시데이터에 대한 추적도 용이하여 디지털 객체 인식자(DOI: Digital Object Identifier)와 같은 인식표를 달기 쉽다는 이점이 있다(Poisot *et al.*, 2019).

온라인상에서 적용이 가능한 개방형 자료 표준 형식의 채택은 데이터간의 이질성을 완화하고 접근성을 향상시키는데 도움이 되며, 메타데이터 및 데이터 표준화의 중요성 강화(예: 공통 모델 및 인코딩에 대한 정의)는 연구자, 대학, 도서관 및 일반 대중과 같은 이해관계자의 데이터 저장 및 검색 용이성을 촉진하는 것으로 알려졌다(Bigagli and Nativi, 2017). 또한, 데이터를 수집하고 표현하며 저장하는데 있어 데이터 표준을 사용하는 방식은, 다양한 연구에서 수집된 데이터를 더 쉽게 집계하고 비교할 수 있게 해주며, 과학자들 본인이 직접 수집하지 않은 데이터를 사용하려고 할 때 느낄 수 있는 거리감을 극복하는 하나의 수단이 될 수 있다(Zimmerman, 2008). Poisot *et al.* (2019)에 의하면 현재의 생태데이터 구조 형식은 연구자들이 데이터를 간단하게 복제하거나 재사용 할 수 있을 만큼 충분히 표준화되어 있지 않다는 점을 지적하면서도, 복수의 출처로부터 생산된 데이터의 통합 구조화 정도가 향상된다

면 새로운 연구가 급속도로 향상되고 대규모 가설의 시험도 가능할 것이라고 한다.

이와 같이 전 세계적으로 생물다양성, 생태계, 자연환경 관련 데이터에 대한 표준 수립과 개방에 의한 데이터 공유의 필요성과 당위성을 피력한 연구가 지속적으로 수행되고 있으며, 생물다양성 협약의 지구생물다양성전망 보고서에서도 중요한 사항으로 제기되며 이에 대한 이행을 촉구하고 있다. 그러나, 아쉽게도 국내의 경우 생태조사자료와 생태 데이터의 구축과 분석에 관한 몇몇 연구(Choi and Kim, 2002; Song *et al.*, 2014; Youn, 2014; Youn *et al.*, 2017)가 진행은 되었으나 다양한 계층의 이해관계자들에게 만족스러운 결과를 제시하지는 못하고 있다. 개별 연구자와 대학 및 각 연구기관에서는 고유의 저장 방식에 의한 데이터와 생태 정보를 보유하고 있으나, 이에 대한 국가차원의 생태 정보 표준화가 마련되어 있지 않아 데이터 간 상호 연계 및 통합 정보를 활용한 자연환경보전 정책 수립 및 생태계 연구 전반에 걸친 다학제 간 연구 활동이 제한적으로 진행되고 있는 상황이다.

데이터 표준화는 정보시스템별로 산재해 있는 데이터 정보 요소에 대한 명칭, 정의, 형식, 규칙에 대한 원칙을 수립하여 이를 적용하는 것을 의미하며, 데이터 표준화를 통해 다음과 같은 효과를 기대할 수 있다. 첫째, 명칭 통일로 명확한 의사소통을 증대시킬 수 있다. 즉, 동일한 데이터에 대해서는 동일한 명칭을 사용함으로써 조사자, 개발자, 연구자, 운영자 등 다양한 계층 간에 명확하고 신속한 의사소통이 가능해진다. 둘째, 필요한 데이터의 소재 파악에 소요되는 시간과 노력을 최소화할 수 있다. 표준 데이터를 관리함으로써 새로운 정보 요건 사항 발생 시 정보 사용자는 데이터의 의미와 데이터의 위치 등을 신속하게 파악할 수 있어 적시에 정확한 정보를 활용할 수 있게 된다. 셋째, 일관된 데이터 형식 및 규칙을 적용하여 데이터의 품질이 향상될 수 있다. 취득한 원시데이터를 데이터 표준에 맞게 적용함으로써 데이터의 입력 오류를 예방하고 데이터의 품질을 향상시킬 수 있게 된다. 넷째, 정보시스템 간 데이터 인터페이스 연결시 데이터 변환 및 정제 비용을 감소할 수 있다. 데이터 표준에 의해 데이터를 관리함으로써 데이터의 전송, 공유, 가공을 위해 정보시스템 간 또는 정보시스템 내에서 별도의 포맷 변환이나 정제 작업 없이 원천 그대로의 데이터를 활용할 수 있으므로 향후 추가적인 비용이 발생하지 않게 된다(NIE, 2019-1).

본 연구의 목적은 국내의 생태 조사, 연구데이터를 데이터베이스로 구축하고 운영 및 관리할 시에 적용할 데이터의 표준을 정함으로써 국내 생태 데이터의 품질을 확보하고 국내 생물다양성, 생태, 자연환경 정보시스템 간은 물

론 국외 데이터 정보시스템과의 정보공유 활성화에 기여하는데 있다. 국내 생태 데이터의 표준화를 수행하기에 앞서 국내 최대의 생태 조사 데이터인 제2차~제4차 전국자연환경조사 자료의 비표준화 현황과 표준화 필요성에 대하여 분석한 후 데이터 오류 정제를 거쳐 표준화를 수행하였다.

재료 및 방법

본 연구에서는 1986년부터 국내 자연환경과 생물다양성 구성요소의 분포 및 현황을 파악하여 국토의 효율적 관리를 하기 위해 매 10년을 주기로 시작한 전국자연환경조사 자료를 참고하였으며, 자연환경조사 결과에 공간적인 좌표를 기재하기 시작한 제2차 (1997~2005), 제3차 (2006~2013) 및 제4차 전국자연환경조사 (2014~2018) 결과를 활용하였다. 제4차 전국자연환경조사 시점인 2014년부터는 5년 주기로 전국에 대한 조사가 수행되었으며 현

재 제5차 조사 (2019~2023)가 수행되고 있다 (NIE, 2019-2). 1997년 시작된 제2차 조사부터 육상생태계에 대해서는 지형, 식생, 식물상, 포유류, 조류, 양서·파충류, 육상곤충, 어류, 저서성대형무척추동물 등 9개 분야를 대상으로 하고 있으며, 이들 조사 결과에 대한 데이터를 정제하고 표준화하였다.

결과 및 고찰

1. 데이터 오류 개선

전국자연환경조사 데이터 표준화를 수행하기에 앞서 전국자연환경조사 데이터베이스가 보유한 데이터 오류를 정제 가능여부, 수치와 값, 명칭, 날짜, 시간 순서, 업무규칙 등 6개 항목으로 분류하고 문제점을 분석하였다 (Table 1). 분석된 문제점을 토대로 제5차 전국자연환경조사 지침을 기반으로 제작한 표준 오류 검출프로그램을 이용해 오류

Table 1. Data error classification & Improvement.

Class	Main issue	Improvement
Y/N	· Invalidate letters	· Modified into Y/N · Deleted invalidate letters in consultant with data producers
Numerical value	· Invalidate data in different unit or scope	· Deleted the letters after confirming unit or scope in the column · Some of data with letters excluded
Name	· Special character and multiple data	· Some of the data with special characters modified · Others (multiple data) left unimproved
Date	· Invalidate data found in Date domain	· Modified in standard format (YYYY-MM-DD)
Chronological sequence	· Invalidate data out of chronological sequence	· Modified into standard format (YYYY-MM-DD) · Some of invalidate data modified with date in the other column
Data rules	· Invalidated data out of data rules	· Modified in consultant with data producers

Table 2. The result of data improvement through standard error validation program.

Class	Number of data					Error rate (%)	
	Total	Errors	Improvements	Exceptions	Unimproved	Before	After
Y/N	2,433,743	134,856	134,856	0	0	5.5411	0.0000
Numerical value	7,128,970	35,384	34,971	413	0	0.4963	0.0000
Name	904,407	23	9	13	0	0.0025	0.0000
Date	3,814,769	4,486	4,469	0	17	0.1176	0.0004
Chronological sequence	1,219,490	3,066	3,060	0	6	0.2514	0.0005
Data rules	918,858	10,899	0	0	10,899	1.1861	1.1861
Total	16,420,237	188,714	177,365	426	10,922	1.1493	0.0665

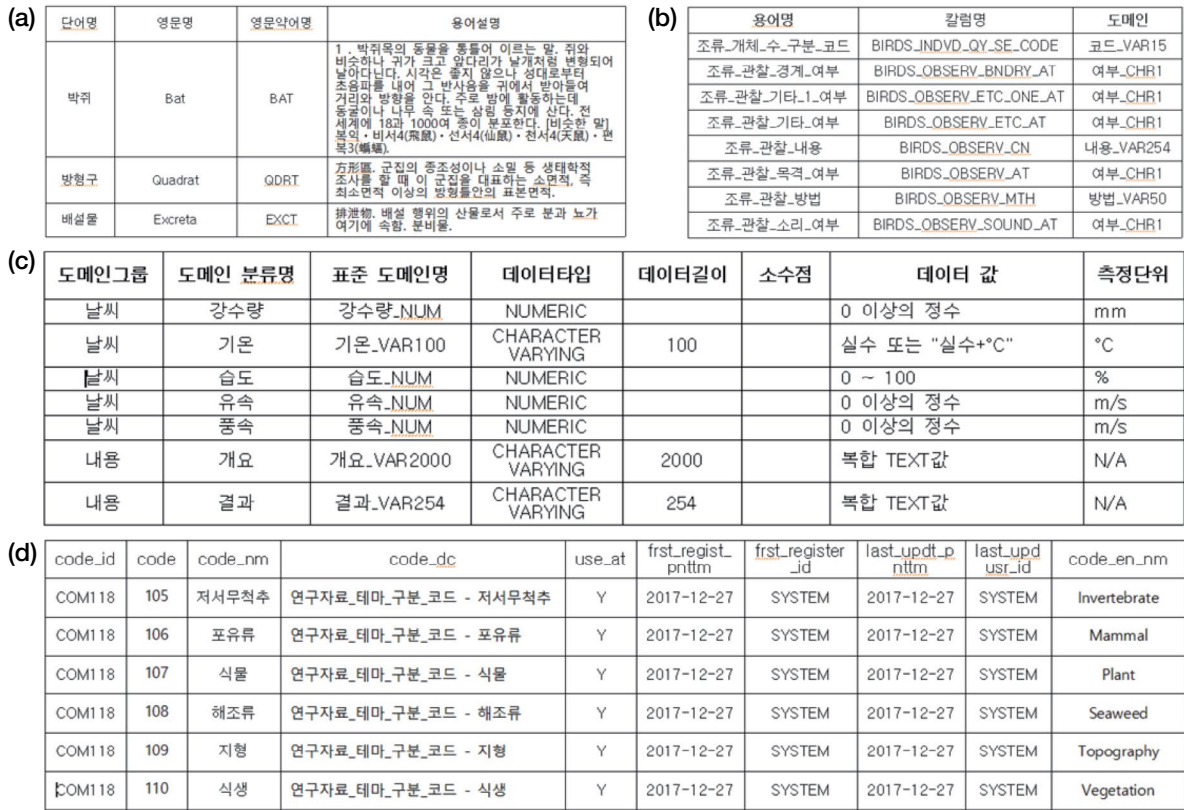


Fig. 1. Example of standardized national ecosystem survey data, (a) Standard word, (b) standard term, (c) standard domain, (d) standard code.

를 추출한 결과 총 16,420,237건의 전국자연환경조사 데이터 값 중 188,714건이 오류를 포함하는 것으로 나타났다 (Table 2). 이들 데이터 (오류 건수) 중 전문가 검수에 의해 오류로 판단되지 않거나 보관이 필요한 데이터 값 426건 (오류제외 건수)을 제외한 177,365건에 대한 오류 수정 (개선 건수)을 통해 오류율을 1.1493%에서 0.0665%로 개선하였다. 개선된 데이터는 각 조사 분류군별 전문가 검수를 통해 최종적으로 표준화 대상으로 선정하였다.

2. 표준 데이터

표준 데이터란 데이터 표준 관리를 효과적으로 수행하기 위한 구성요소로 본 연구에서는 표준 단어, 표준 용어, 표준 도메인 그리고 표준 코드로 구성하였다 (Fig. 1).

1) 표준 단어

전국자연환경조사 결과를 데이터베이스로 구축한 자연환경종합 GIS-DB의 데이터 모델 및 용어집을 통해 생태 관련 모든 단어를 추출하였다. 추출된 단어는 단어의 종류와 유형을 분류하고 업무 정의 및 용도를 고려하여 표준

단어를 정의하였다. 표준 단어 사전을 정의할 경우 이음동어, 동음이의어 처리에 주의하였다. 또한 정의된 표준 단어는 정의된 표준화 원칙을 참고하여 영문명과 영문 약어를 정의하였다. 용어 표준의 기반은 공공기관의 데이터베이스 표준화 지침 (행정안전부고시 제2019-20호)에 따른 행정표준용어를 참고하여 생태 용어들을 반영한 용어 사전을 작성하였다 (Fig. 1a).

2) 표준 용어

단어, 도메인, 코드 표준을 바탕으로 표준 용어를 구성하고, 단어의 조합, 도메인 분류, 데이터 타입 길이, 코드 값을 기준으로 하였다. 표준 단어 사전에서 적절한 의미를 나타내는 단어를 선택하고, 표준 도메인 사전에서 해당 데이터 (숫자, 문자, 날짜, 코드 등)를 정의할 수 있는 도메인명을 선택하였다. 코드성 도메인의 경우 표준 코드 사전에서 해당되는 코드를 선택하였다. 표준 단어 사전, 표준 도메인 사전, 표준 코드 사전에서 적절한 항목을 찾지 못한 경우 신규로 해당 사전에 추가하였다. 작성된 표준도메인사전 및 표준단어 사전에서 선택한 용어를 적용하였고, 총 1,806개 생태데이터 용어로 정의하였다 (Fig. 1b).

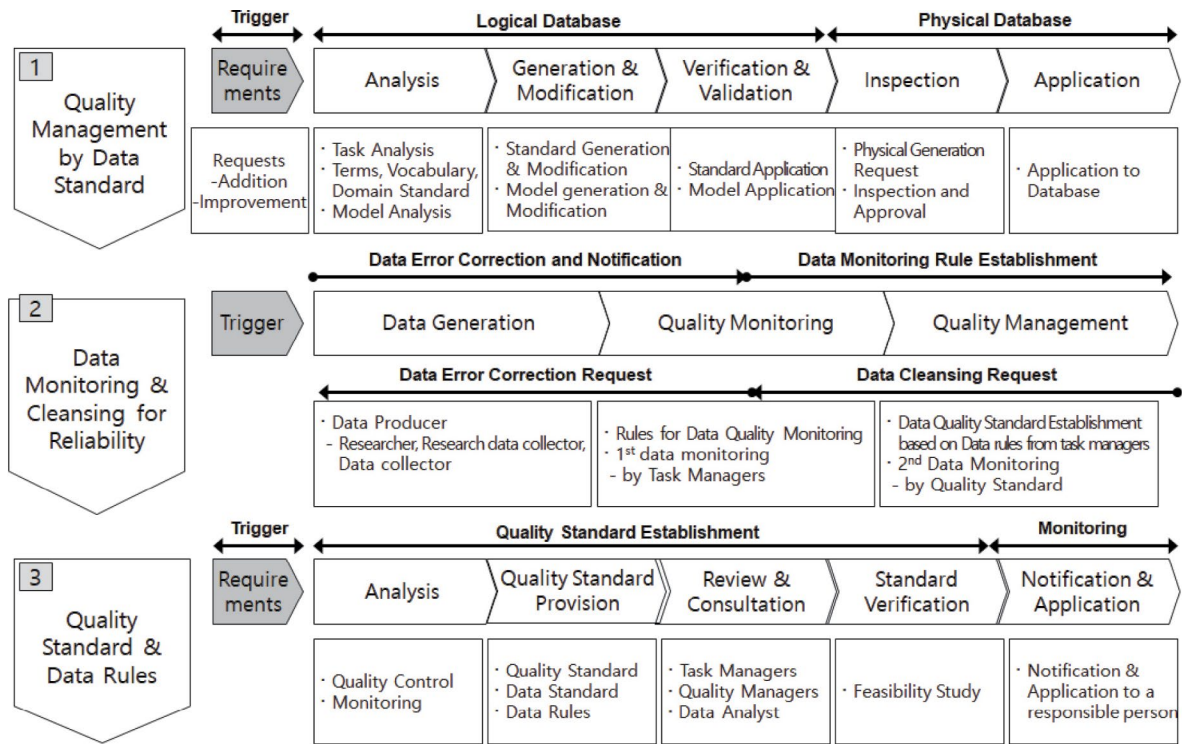


Fig. 3. Management guideline and process for integrated data management of national ecosystem survey.

3) 표준 도메인

표준 용어 조합의 마지막에 위치하는 도메인(분류어)은 데이터의 성격을 분류한 것으로 표준 용어가 가질 수 있는 데이터 형식(데이터 타입, 길이 등)을 제한한다(NIE, 2019-1). 또한 업무적인 용도와 사용 빈도, 데이터의 논리적·물리적인 특성 등을 고려하여 도메인을 추출하고 데이터 타입을 부여한다. 본 연구에서는 총 11개 도메인 그룹, 169개 도메인에 대하여 정의하였고, 중복 도메인 제거 및 인포타입(Information type: 데이터 타입, 길이)을 정비하였다(Fig. 1c).

4) 표준 코드

표준 도메인의 유형 중 하나로서, 코드 도메인에 허용된 데이터 값을 제한된 범위 내에서 구체적으로 열거하여 정의한 것을 지칭한다(NIE, 2019-1). 데이터 값을 코드값 또는 유효값이라 하며 각각의 코드 값에는 의미를 부여한다. 이 의미를 ‘코드값명’ 또는 ‘유효값명’이라 한다. 메타데이터 관리 시스템 상에서는 유효값, 유효값명으로 명명하였다. 표준 코드는 다른 표준과는 달리 데이터값, 즉 유효값에 대해서도 정의가 관리되어야 하고, 코드명에 대한 표준화를 진행할 경우 유사한 코드가 중복되어 관리되지 않도록

유의해야 한다(NIE, 2019-1). 본 연구에서는 표준화 대상 데이터의 표준 코드를 정의하였고, 연계 및 수집 데이터 관련 코드를 추가하였다(Fig. 1d).

3. 구조 데이터

논리 데이터 모델과 물리 데이터 모델을 구성하는 구조적 특징을 갖는 데이터를 의미하며, 주제영역, 엔티티, 속성, 테이블 및 컬럼으로 구성하였다(Fig. 2).

1) 주제영역

데이터 모델의 개념을 보다 관리하기 쉽도록 특정 주제를 중심으로 분류한 체계로 본 연구에서는 전국자연환경 조사를 하나의 주제영역으로 묶어 에코뱅크(<https://nie-ecobank.kr>)에서 서비스하는 다른 데이터베이스들과 구분하였다.

2) 엔티티

업무에서 정보로 관리할 필요가 있는 데이터 집합의 실체로, 서로 밀접한 관련성을 가지고 있는 속성별로 데이터를 분류하였다. 전국자연환경조사에 해당되는 10개의 조사분류군과 6개의 공간 범위로 분류하였다.

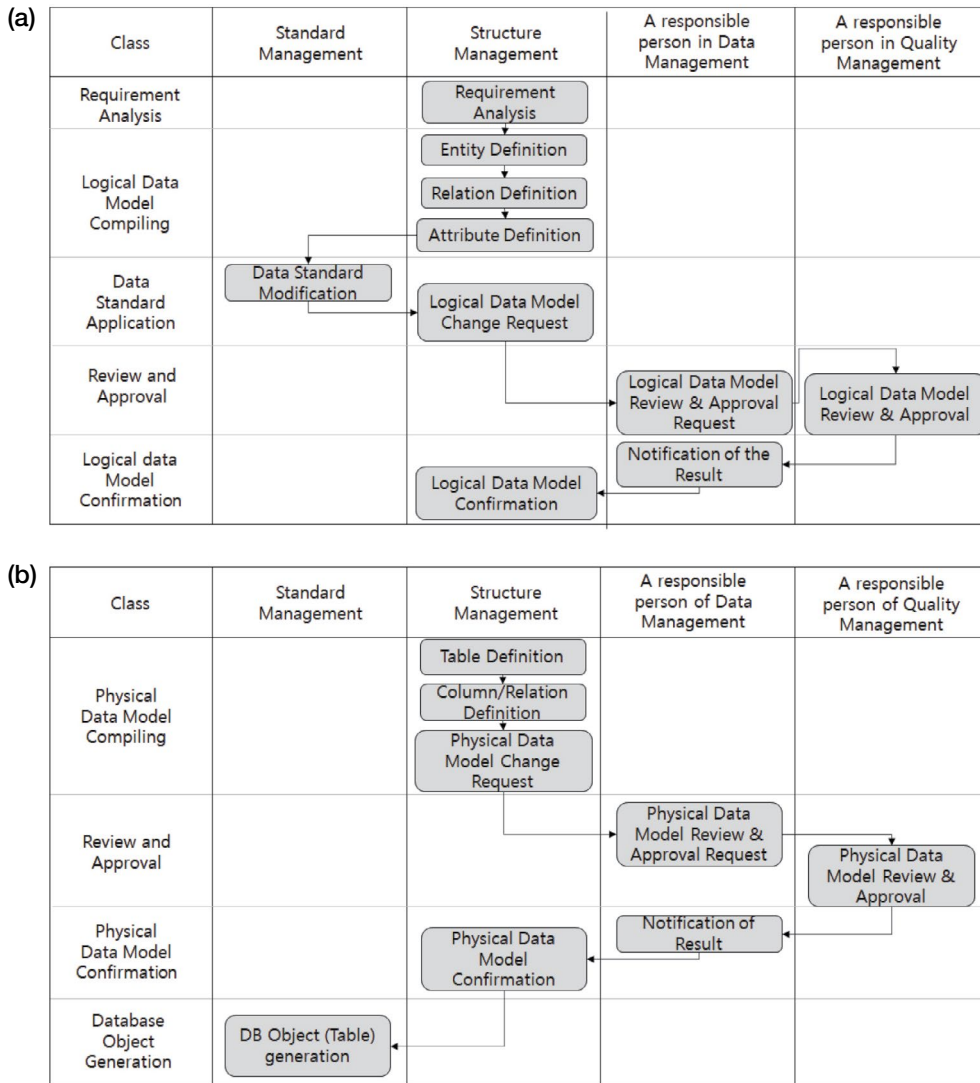


Fig. 4. Data model transforming process (a) logical data model management (b) physical data model management.

3) 테이블

자료를 하나의 표처럼 구조화해 놓은 형태로써 관계형 데이터 베이스에서 자료를 관리하는 기본 단위를 뜻한다. Fig. 2에서 제시된 개체관계도(ERD)의 10개 분류군별 엔티티와 6개의 공간범위 엔티티에 포함된 속성의 조합으로 구성되며, 엔티티별 특성에 따라 서로 다른 구성을 가진다.

4) 컬럼

논리적 데이터 요소인 속성과 대응되는 물리적 개념으로, 데이터의 가장 기본적인 물리적 단위로 사용하였다. 즉, 성격이 유사한 데이터 속성들을 하나로 묶는 범주를 정의하는 데이터 테이블의 구성요소인 열을 의미한다. 전국자연환경조사 개체관계도(ERD)에서 각 엔티티의 특성

에 따라 다양한 컬럼으로 구성되었다.

4. 데이터 표준 관리

1) 데이터 표준 관리 프로세스

데이터 표준 관리를 위한 프로세스는 데이터 표준화 원칙에 따라 정의된 표준 단어 사전 및 도메인 사전, 표준 용어 사전, 표준 코드, 데이터 관련 요소 표준 등을 적합한 형태로 정의하고 관리하기 위한 프로세스를 말한다 (NIE, 2019-1). 이는 데이터베이스 설계와 개발을 지원하고 데이터 표준의 사용 및 재사용을 통해 시스템간 상호 운용성, 데이터 공유, 시스템 통합, 비즈니스 프로세스 개선 등을 지원하며, 데이터의 일관성과 정확성을 유지할 수 있다. 이

를 위해 업무에 의한 소스, 테이블 및 모델 생성 시점에 표준을 적용하였다. 또한 지속적인 표준화에 대한 교육과 개선 모니터링 활동으로 표준 프로세스가 조직과 관련 담당자에게 구체화되도록 하여야 한다. 데이터 표준은 조사자의 의견이 반영되어야 하겠지만, 관습적으로 잘못 사용되어 온 용어를 모두 수용할 수는 없으므로 조정이 필요하다. 이를 위해 원천 데이터 오류를 모니터링 하였고, 데이

터 모니터링에 대한 역할을 규정하였다. 일관된 데이터 형식 및 규칙의 적용으로 데이터 품질을 향상시키고, 데이터 표준에 대한 관리 프로세스를 제대로 정의함으로써 데이터 표준을 지속적으로 유지할 수 있도록 해야 한다. 이를 위해 품질 관리 및 모니터링 관점의 기준을 마련하였고, 기준, 표준, 규칙 정의 후 협의 및 타당성을 검토하여 적용하였다 (Fig. 3).

Table 3. Establishment data management policy for integrated database management of national ecosystem survey in Korea.

Class	Title	Contents
Data management policy	· Ecological data management principles	· Ecological data management principles
Data management process	· Data naming rule · Data standard rule · Metadata management rule · Data structure management rule · Data quality management rule · Data publish management rule · Database management	· Naming standard · Vocabulary standard · Domain standard · Data modeling guide · Data flow management guide · Data publish management guide · DB backups and restoration · Data security · Performance management
Data management organization	· Data management organization	· Composition of organization · Duties, roles, and responsibilities

Table 4. Role of data management organization.

Class	Characteristics of Organization
Data quality performance	· Define quality requirements, standard management, data verification for reliability and monitoring · Analysis of quality problems (error analysis, etc.) and support and operate data improvement measures (cleaning, etc.), and problem solving · Development of indicators for data quality measurement · Support and operate the measurement for quality level · Collaboration with data quality managers · Composition: Quality manager by tasks, data owner by tasks, DB manager
Data quality management	· Establishing and implementing a standard management system for ecological information (guideline and process) · Support system development plan and decision-making · Data modeling, database design, and monitoring · Define quality requirements, data verification for reliability, monitoring, and correction and improvement of errors · Reviewing legal and institutional requirements for data operation and management and education · Composition: Data quality manager (EcoBank Team)
Data management committee	· Adjusting, deliberating, decision making for data quality · Arrange strategic advice, and implementation policies for data quality policies · Acquisition of Quality Management Sponsorship · Composition: Data Quality Review Committee, Chairman of the Data Quality Management Committee

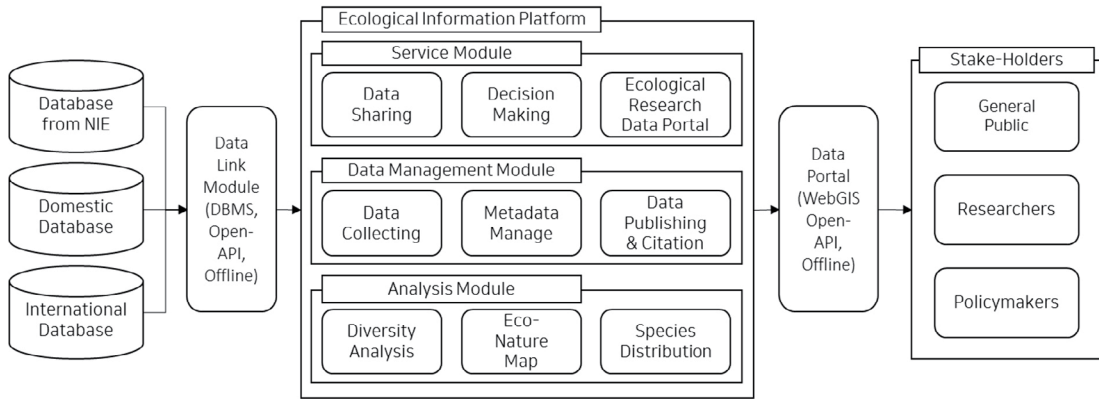


Fig. 5. The architecture and service flow of EcoBank system (Sung et al., 2018).

2) 데이터 모델 변경 프로세스

데이터 소스, 테이블, 모델 변경 시 수립된 표준을 확인하여 검증을 하였고, 표준 적용 시 추가, 변경 표준 관리를 마련하였다. 데이터 모델 변경을 위한 프로세스는 Fig. 4에 제시된 바와 같다.

3) 데이터 관리 정책 수립

효율적이고 체계적인 전국자연환경조사에 대한 통합 DB운영을 위하여 데이터 관리 프로세스와 데이터 관리 조직으로 구성된 데이터 관리 정책을 수립하였다 (Table 3). 데이터 관리 프로세스는 데이터 명명규칙 등 7가지 원칙을 통해서 명명표준, 용어표준, 도메인 표준, 데이터 모델링 가이드, 데이터 흐름 관리 가이드, 데이터 출판 관리 가이드, DB 백업 및 복구, 데이터 보안 원칙, 성능관리 원칙에 대한 사항을 정의하였으며, 데이터 관리 조직은 조직 구성도, 업무정의 및 역할과 책임에 대한 내용으로 구성하였다. 또한 데이터 품질관리의 지속성을 확보를 위한 데이터 관리 위원회, 데이터 품질 관리팀, 데이터 품질 수행팀의 데이터 관리 조직 구성 및 관리조직별 특성을 제시하였다 (Table 4).

5. 표준화된 전국자연환경조사 데이터 개방

본 연구를 통해 정제되고 표준화된 전국자연환경조사 데이터는 국립생태원이 2015년부터 개발한 생태정보포털 서비스 시스템인 에코뱅크 (<https://nie-ecobank.kr>)를 통해 누구나 자유롭게 사용가능하도록 제공하였다 (Fig. 5)

본 연구에서는 제2차~제4차전국자연환경조사 데이터의 정제 및 표준화를 통해 객관적이고 신뢰할 수 있는 데이터의 공유와 활용 방안을 제시하고자 하였다. 본 연구 결과를 활용하여 관련기관과 연구자들의 의견을 반영한 국내

생태데이터 표준안을 수립한다면 국내 생물·생태 관련 기관별의 데이터베이스 시스템 간 상호 연계가 편리해질 것으로 판단된다. 이를 통해 하나의 시스템에서 생태와 생물 다양성 관련 자료를 일목요연하게 조회, 검색하는 것이 가능할 것이며, 생태계 현황파악과 변화 예측 시 시공간적인 연속성과 상호 요소, 인간생활과의 상관성 분석 등에 효율적으로 활용이 가능해질 것으로 보인다. 일반적으로 데이터 표준화는 시스템별로 산재해 있는 데이터 정보 요소에 대한 명칭, 정의, 형식, 규칙에 대한 원칙을 수립하여 이를 공통적으로 적용하는 것을 의미한다 (NIE, 2019-1). 이러한 데이터 표준화 작업은 데이터의 정확한 의미를 파악할 수 있게 할 뿐만 아니라 데이터에 대한 상반된 시각을 조정하는 역할을 수행한다. 본 연구결과를 토대로 국내 생태조사 데이터의 공유와 활용 활성화는 물론, 향후 지속적인 보안을 통해 국내 표준안을 제정하여 생물다양성 협약에서도 요구하고 있는 공통의 데이터 표준과 관례를 통해 향후 데이터의 이동과 접근을 강화하고 촉진할 것이라는 권장 사항의 적극 이행에도 기여할 것으로 기대된다.

적 요

본 연구에서는 최근, 자연환경의 변화에 따른 생태계 예측과 지속가능한 개발과 보전의 정량적 근거 제시를 위해 전세계적으로 연구중요도와 활용도가 높아지고 있는 생태조사연구 데이터의 정제와 표준화 및 서비스 방안에 대하여 제시하였다. 1997년부터 매 10년 단위로 전국을 대상으로 수행된 제2차, 제3차 전국자연환경조사 결과와 2014년부터 매 5년 단위로 수행된 제4차 전국자연환경조사 결과를 대상으로 전국자연환경조사 지침을 따라 별도 개발된 입력 오류 검증 프로그램을 통해 1차 정제 후 분야별 전문

가의 검수를 거쳤다. 또한, 국내외 데이터 표준 사례들을 분석하여 국내의 자연환경과 생태조사 자료인 생태정보를 표준화하는 방안을 제시하였다. 전국자연환경조사 결과에 대해 (1) 조사자료인 데이터 용어의 표준화, (2) 유사한 유형의 데이터 테이블 통합 및 (3) 불필요한 속성과 입력오류를 제거하였고 (4) 현장 조사자 별로 상이한 입력항목의 단일화, (5) 데이터를 일련번호 코드로 정리하고 (6) 입력항목 별 코드 매핑을 수행하였다. 연구결과, 연구자와 일반인들이 활용하기 쉽도록 전국자연환경조사 데이터를 정제하고 국내 최초의 생태데이터 표준안 마련의 기반을 수립하였다. 또한, 연구결과로 산출된 전국자연환경조사 데이터는 국내의 환경영향평가, 서식지 보전지역 설정, 멸종위기종 분포 예측, 기후변화에 따른 생태계 위험 예측은 물론, 국외 생태정보 연계 등에 손쉽게 공유 및 활용가능하여 다양한 환경정책수립과 관련 연구활성화에 기여할 것으로 기대된다. 본 연구를 통해 정제된 전국자연환경조사 데이터는 국립생태원에서 국내 최초로 개발중인 생태정보포털서비스 시스템인 에코뱅크 (<http://nie-ecobank.kr>)를 통해 개방하여 누구나 온라인상에서 쉽게 접근하고 활용할 수 있도록 하였다.

저자정보 권용수 (국립생태원 전임연구원), 송교홍 (국립생태원 연구원), 김목영 (국립생태원 연구원), 김기동 (국립생태원 실장)

저자기여도 연구설계: 권용수, 송교홍, 김기동, 그림 및 표 작성: 권용수, 송교홍, 문헌정리: 김목영, 원고작성: 권용수, 송교홍, 김목영, 김기동, 원고 수정 및 검토: 권용수, 김기동

이해관계 본 연구는 이해관계의 충돌 여지가 없습니다.

연구비 본 연구는 환경부의 재원으로 국립생태원 에코뱅크 구축 (NIE-전략연구-2020-01)과 전국자연환경조사 (NIE-법정연구-2020-01) 결과를 지원 받아 수행하였습니다. 본 연구에서 제시하고 있는 전국자연환경조사 생태정보는 조사자료의 확인을 위한 것이므로 법적 효력이 없음을 밝힙니다.

REFERENCES

Bigagli, L. and S. Nativi. 2017. Mobilising data: Environmental data, technical and governance issues. p. 121-140. In: Open data and the knowledge society (Wessels, B., R. Finn, K. Wadhwa and T. Sveinsdottir, eds.). Amsterdam

university press. Amsterdam.

Bingham, C.H., M. Doudin, L.V. Weatherdon, K.D. Belmonte, F.T. Wetzel, Q. Groom, E. Lewis, E. Regan, W. Appeltans, A. Güntsch, P. Mergen, D. Agosti, L. Penev, A. Hoffmann, H. Saarenmaa, G. Geller, K.D. Kim, H.J. Kim, A.S. Archambeau, C. Häuser, D.S. Schmeller, I. Geizendorffer, A.G. Camacho, C. Guerra, T. Robertson, V. Runnel, N. Valland and C.S. Martin. 2017. The Biodiversity Informatics Landscape: Elements, Connections and Opportunities. *Research Ideas and Outcomes* 3: e14059. <https://doi.org/10.3897/rio.3.e14059>.

Choi, B.G. and S.S. Kim. 2002. Natural eco-information management system using GIS in Kwang-Wha. *Journal of Korean Society for Geospatial Information System* 10(3): 123-130.

Hobern, D., B. Baptiste, K. Copas, R. Guralnick, A. Hahn, E.V. Huis, E.S. Kim, M. McGeoch, I. Naicker, L. Navarro, D. Noesgaard, M. Price, A. Rodrigues, D. Schigel, C.A. Sheffield and J. Wieczorek. 2019. Connecting data and expertise: a new alliance for biodiversity knowledge. *Biodiversity Data Journal* 7: e33679. 1-20.

National Institute of Ecology (NIE). 2019-1. Guideline for data standard in construction of EcoBank-phase I (4th year). NIE, Seocheon, Korea.

National Institute of Ecology (NIE). 2019-2. Guideline for the 5th national ecosystem survey. NIE, Seocheon, Korea.

Poisot, T., A. Bruneau, A. Gonzalez, D. Gravel and P.P. Neto. 2019. Ecological data should not be so hard to find and reuse. *Trends in Ecology & Evolution* 34(6): 494-496.

Secretariat of the Convention on Biological Diversity. 2014. Global Biodiversity Outlook 4. Montréal, 155 pages. www.cbd.int/GBO4

Song, J.H., I.J. Kang, S.H. Hong and D.H. Park. 2014. Construction of vegetation information management system using GIS. *Journal of the Korean Society for Geospatial Information System* 22(4):99-106.

Sung, S., Y.S. Kwon and K.D. Kim. 2018. Development and applications of ecological data portal service (EcoBank) for sharing ecological information of Korea. *Korean Journal of Ecology and Environment* 51(3): 212-220.

Youn, C., C. Kim and H. Moon. 2017. The establishment for technology development plan for national spatial information infrastructure cloud service. *Journal of the Korean Academia-Industrial Cooperation Society* 18(3): 469-477.

Youn, J.H. 2014. The establishment of BPR for national spatial data infrastructure quality management system. *Journal of Korean Society for Geospatial Information System* 22(4): 81-89.

Zimmerman, A.S. 2008. New Knowledge from Old Data: The Role of Standards in the Sharing and Reuse of Ecological Data. *Science, Technology, & Human Values* 33(5): 631-652.