

생태정보를 위한 메타데이터 카탈로그 시스템 설계 기술 보고서

한국과학기술정보연구원



저자소개

박근철

한국과학기술정보연구원/선임연구원

Email : gcpark@kisti.re.kr

김상완

한국과학기술정보연구원/선임연구원

Email : sangwan@kisti.re.kr

남덕윤

한국과학기술정보연구원/선임연구원

Email : dynam@kisti.re.kr

황순욱

한국과학기술정보연구원/책임연구원

Email : hwang@kisti.re.kr

목 차

I. 생태계 정보를 위한 메타데이터 카탈로그 기술 개요	3
1. 생태관측 데이터	3
2. 메타데이터 카탈로그 기술	3
II. 연구동향 분석	4
1. ECN(Environmental Change Network, 영국)	5
2. TERN(Terrestrial Ecosystem Research Network, 호주)	7
3. NEON(National Ecological Observatory Network, 미국)	14
4. 장기생태연구(한국)	17
5. 소결론 - 비교/분석	19
6. 기타 참고 사이트	23
III. 생태계정보를 위한 메타데이터 카탈로그 구조 설계	24
1. 메타데이터 카탈로그 시스템 구축 방법 비교 분석	25
2. 해외 생태정보 메타데이터 카탈로그 구조 분석	25
3. 한국형 메타데이터 카탈로그 구조 설계	29
IV. 생태정보를 위한 메타데이터 카탈로그 시스템 설계	32
1. 생태계정보 메타데이터 카탈로그 시스템 설계	33
2. 생태계정보 메타데이터 등록	33
3. e-EcoHub 포털 연계	36
4. 생태계정보 분석 연계	37
5. 국내외 생태계정보 공유	38
V. 향후계획 및 기대효과	42
1. 향후계획	43
2. 기대효과	43
VI. 참고 문헌	44

I. 생태계 정보를 위한
메타데이터 카탈로그 기술 개요

I. 생태계정보를 위한 메타데이터 카탈로그 기술 개요

1. 생태관측 데이터

생태관측 인프라는 국가차원의 인프라에 해당되는 생태분야의 관측데이터를 장기적으로 보관 가능하고, 발견되는 생태종의 증가에 따라 유연한 플랫폼이 요구된다. 또한, 다양한 생태관련 분석 툴과의 접근성과 글로벌 생태데이터와의 비교 및 분석 연구를 위해 다수의 국가에서 추구하는 표준 메타데이터 스키마 및 전송 프로토콜을 수용해야 한다.

생태관측 자료는 관측분야, 관측요소가 다양하고 관측 기관별로 다양한 종류의 원시 데이터를 수집하고, 생태 관측 자료에 대한 메타데이터 스키마를 자체적으로 정의하여 오라클과 같은 관계형 데이터베이스관리시스템(RDBMS)에 메타데이터시스템을 구축한다. 앞서 언급한 다수의 국가에서 공동 활용하기 위한 전송 프로토콜을 위해 생태분야에서도 EML(Ecological Metadata Language)이라는 메타데이터 표준 언어가 정의되어 있어서 생태학자들간에 서로의 데이터를 공유할 수 있는 기반은 마련되어 있다. 특히, NEON은 기본적으로 EML을 기반으로 하여서 NEON의 요구사항을 반영한 EML을 확장한 메타데이터 시스템을 구축함으로써 NSF에서 구축하는 또 다른 생태계정보 인프라인 LTER(Long Term Ecological Research Network)과의 생태관측데이터 공동 연계 활용을 모색하고 있다. 이에 우리나라도 EML 기반의 메타데이터 시스템을 구축하여서 NEON, LTER 등과 연계한 글로벌 생태관측 데이터 공동 활용을 위한 전송 프로토콜을 고려한 데이터 관리가 필요하다.

2. 메타데이터 카탈로그 기술

생태관측 서비스는 생태관측 정보를 통합관리 및 공유를 위한 웹기반의 생태/환경분야 e-사이언스 연구환경 서비스라 말할 수 있다. 표준화된 메타데이터를 기반으로 연구자들을 포함한 사용자에게 생태계 정보 접근을 위한 웹기반의 통합 인터페이스를 제공하고, HTML5, 메쉬업, GIS 등, 최신의 웹기술을 접목하여 사용자 친화적이고 유연한 사용자 연구환경을 웹으로 제공한다.

메타데이터(metadata)는 다른 데이터(원시 데이터)에 관한 구조화된 데이터로, 다른 데이터를 기술해 주는 데이터(data about data)로 속성정보(attribute)라고도 하며, 대량의 정보 가운데서 찾고자하는 정보를 효율적으로 찾아내기 위해 일정 규칙에 따라 콘텐츠에 대하여 부여되는 일종의 요약 정보라고 할 수 있다. 천문우주나 지구관측 분야에서 흔히 보는 것처럼 생태관측분야에서도 다양한 장소와 다양한 수집 방법에 의해서 수백만의 원시 데이터가 다양한 파일(text, image, sound 등) 형태로 여러 장소에 분산·저장되므로 연구자들이 이들 수백만 원시데이터 파일 이름을 일일이 기억할 수도 없을 뿐만 아니라 필요한 원시데이터를 탐색하고자 할 때, 단지 이들 이름만으로 하는 것은 비효율적인 일이다. 따라서 이들 수백만의 원시데이터들을 효율적으로 탐색, 관리 및 공유를 위한 요약정보 시스템인 메타데이터 시스템 구축은 필수적이라 할 수 있다.

II. 연구동향 분석

II. 연구동향 분석

본 챕터에서는 장기적인 생태조사 및 연구를 지원하는 국내외 사이트에 대한 동향 분석을 데이터 포털 중심으로 설명하고자 한다. 세계 여러나라에서는 일찍부터 환경변화에 따른 생태계의 움직임에 관심을 보여왔다. 그리고, 그러한 데이터를 바탕으로 환경지표종을 개발하고, 환경변화에 대응하는 정책에 활용하고 있다.

1. ECN(Environmental Change Network, 영국)

영국 ECN(Environmental Change Network)는 생태관측 서비스에서 가장 오랫동안 구축해오는 서비스로 영국의 장기적인 환경 관측 및 연구를 위한 프로그램으로서 1992년부터 정기적으로 동식물군과 그들의 물리화학적 생태 환경에 대한 관측을 해옴. 이렇게 관측된 데이터들은 기후변화, 공기오염 등을 포함한 환경적인 변화가 영국의 생태계에 미치는 영향을 분석하는데 활용되고 있다. 관측된 데이터들은 ECN Data Centre(<http://data.ecn.ac.uk/>)로 보내져서 검증 절차를 거친 후 데이터베이스에 저장되어 연구, 교육 및 비상업적인 목적을 위해 활용되어진다.

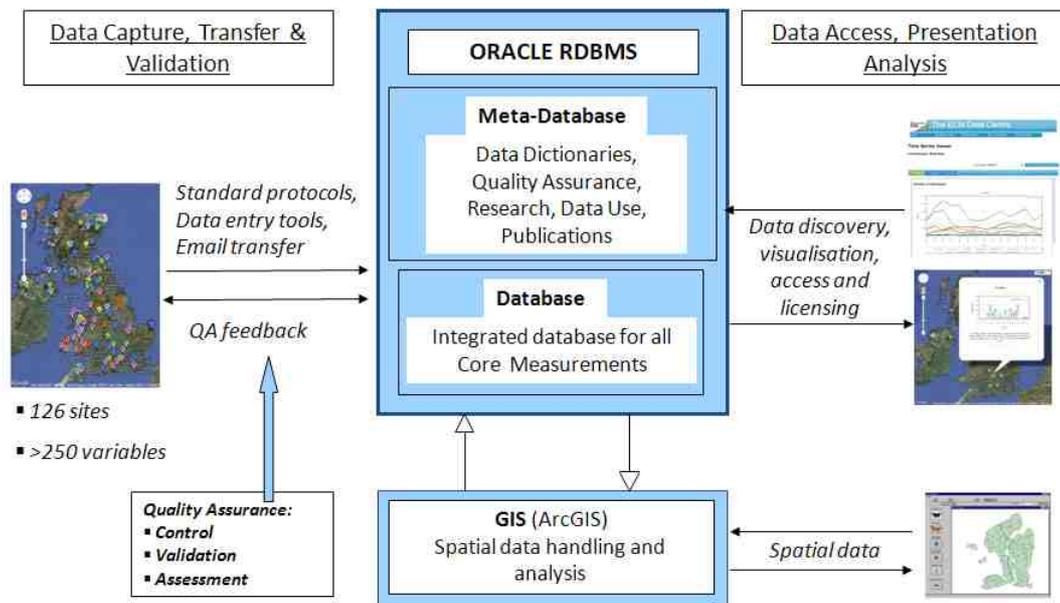


그림 3 ECN Data Management Architecture

ECN은 위의 구조도에서 보는 것과 같이 수집된 데이터(Raw Data)와 이에 관련된 메타 데이터(Meta-data)를 전통적인 Oracle RDBMS에 저장하고 있으며, 공간 데이터 처리를 위해 Arc GIS를 활용하고 있음. 모든 데이터는 중앙 데이터베이스에서 관리되며 고속 네트워크 망을 통해서 접근이 가능하다. ECN의 데이터베이스는 논리적으로 데이터 테이블과 메타 데이터 테이블로 분리되어 있으며, 데이터는 매달, 매분기, 년별로 요약된 버전(summary database)을 제공하고 있음. 이러한 요약된 데이터에 대해서는 무료로 열람이 가능하며, 원 데이터에 대해서는 특정 절차를 거쳐야 검색 및 열람이 가능하게 되어 있다. ECN에서는 데이터(요약된 통계본)의 가시화를 위해 ECN 데이터의

변수간의 상관관계를 표시할 수 있는 ECN Data Explorer와 시간의 흐름에 따른 데이터 변동 추이를 나타낼 수 있는 Time Series Viewer를 제공하고 있다.

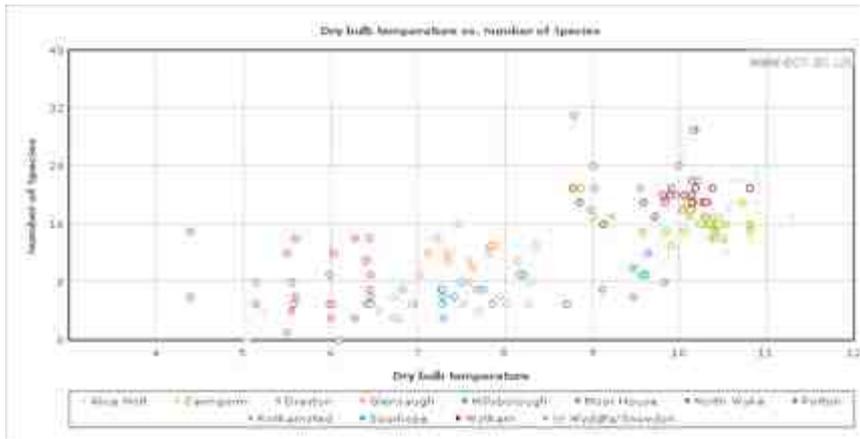


그림 4 ECN Data Explorer

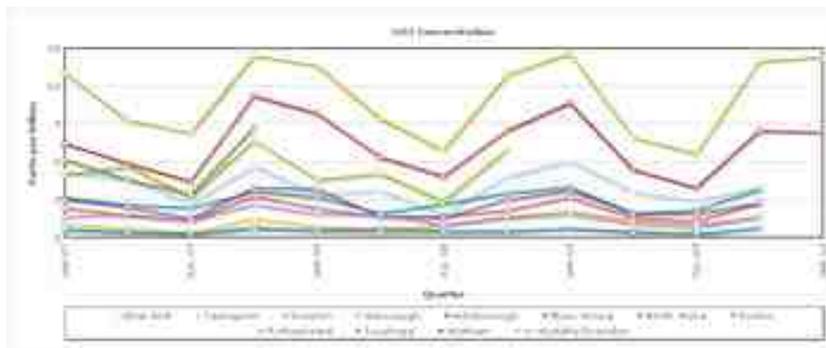


그림 5 Time Series Viewer

영국 ECN은 가장 엄격한 기준으로 데이터와 메타데이터를 수집하고 서비스 한다. 모든 관측자료는 정해진 엄격한 프로토콜에 따라 조사되어야 하며, 두 단계의 인증 절차를 거쳐서 메타데이터 카탈로그 및 데이터 저장소에 탑재된다.

ECN의 각 사이트는 표준 프로토콜의 절차를 지켜 데이터를 수집하고 데이터의 질을 유지하기 자체적인 Quality Control을 수행해야 하며 데이터의 Quality에 대한 Reporting을 의무적으로 수행해야 한다.

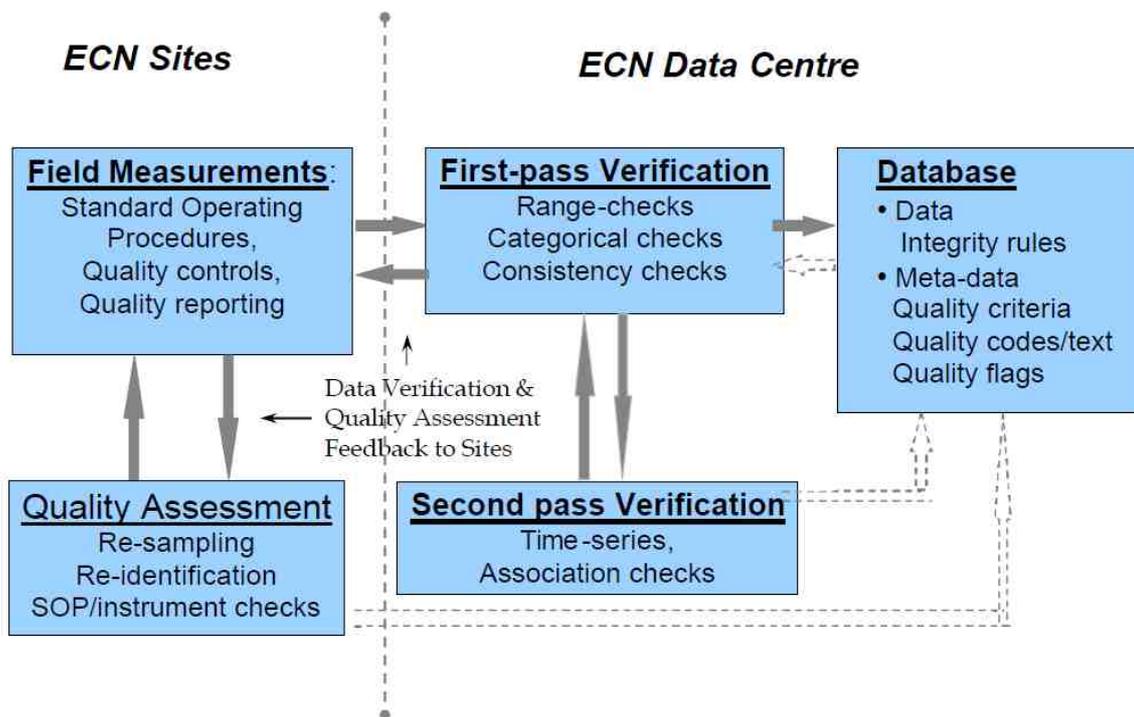
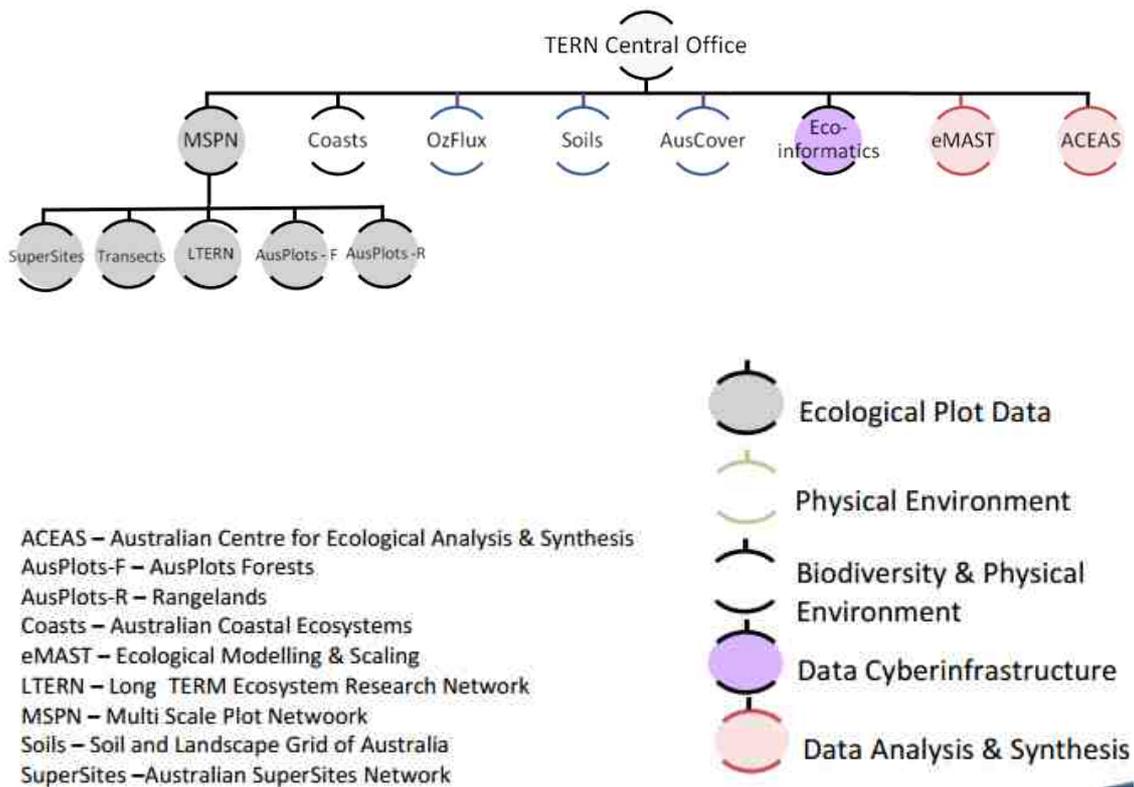


그림 6 ECN Quality Assurance 체계

ECN 데이터 센터에서는 수집 데이터를 2단계를 거쳐서 인증 하고 인증이 완료된 데이터와 메타데이터를 데이터베이스에 탑재하여 서비스 한다. 영국 ECN은 엄격한 관리를 거친 데이터를 탑재하여 데이터의 질이 높고 데이터 간의 비교 분석이 가능한 장점을 가진다.

2. TERN(Terrestrial Ecosystem Research Network, 호주)

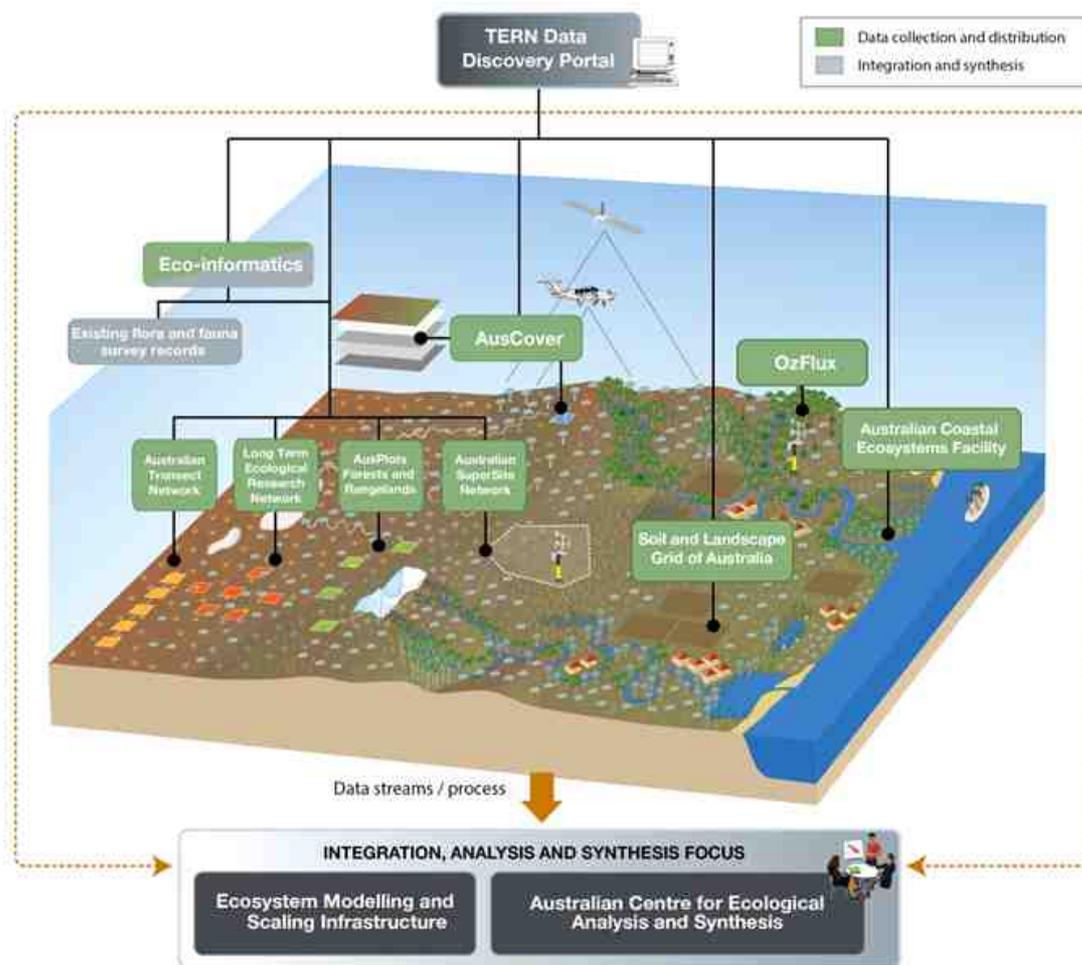
호주의 TERN은 협력 생태계 과학 커뮤니티를 지원하기 위한 인프라스트럭처이자 네트워크이다. 이는 장기 관측과 표준 측정 및 프로토콜을 따르며, 데이터 품질을 증진하고, 데이터 접근을 용이하도록 인터페이스를 제공한다. 다양한 사이트 네트워크로 구성된 데이터를 기반으로 연구 플랫폼을 구축하여 호주 전체 생태계 연구를 지원한다. 이는 생태 커뮤니티가 다양한 특징을 가지는 사이트에 접근하여 데이터를 조작할 수 있도록 한다. TERN은 2010년부터 호주 정부의 예산으로 진행되는 프로젝트로 AEKOS, ACEF, Soil and Land scape Grid, AusCover, OxFlux와 같은 주요 생태 사이트로 구성된다.



대중엔 AEKOS Facility의 Eco-Informatic와 같이 관측장비가 없는 사이버 인프라 Facility도 있고, eMAST, ACEAS와 같이 데이터 분석 및 통합을 하는 Facility도 있다. TERN의 관측 Facility의 성격에 따라 사이트가 나뉘어져 있다. 관측 데이터는 생태, 환경, 기후, 인간 및 에너지에 걸쳐 다양한 데이터로 구성이 되며, 세부 데이터 셋은 아래 표와 같다.

생태	Plants/Animals, Vegetation, Terrestrial ecosystem, Ecological dynamics
환경	Freshwater, Land surface and Soils, Oceans & Coasts, Water & gas
기후	Climate
인간	Agriculture, Human-nature
에너지	Energy

각 관측 Facility에서 수집된 데이터는 각각의 QA를 통해, 메타데이터와 함께 Data Portal에 모여지고, 사용자에게 검색 서비스를 통해 제공된다.



The screenshot displays the TERN Data Discovery Portal interface. On the left, there are search filters including 'Current Search' (with terms like ecosystem, ecology, dynamics, fire, species, interaction, invasive), 'Refine Search' (with an advanced Boolean search box), 'Dates' (1950-2020), 'Region' (Australia), and 'Field of Research' (Biological Sciences, Environmental Sciences, Earth Sciences, Agricultural And Veterinary Sciences). Below these are 'Facilities' such as Australian Ecological Knowledge and Observation System, Australian Supersite Network, OzFlux, Auscover, ACEAS, Australian Coastal Ecosystems Facility, TERN, Soils, and LTERN.

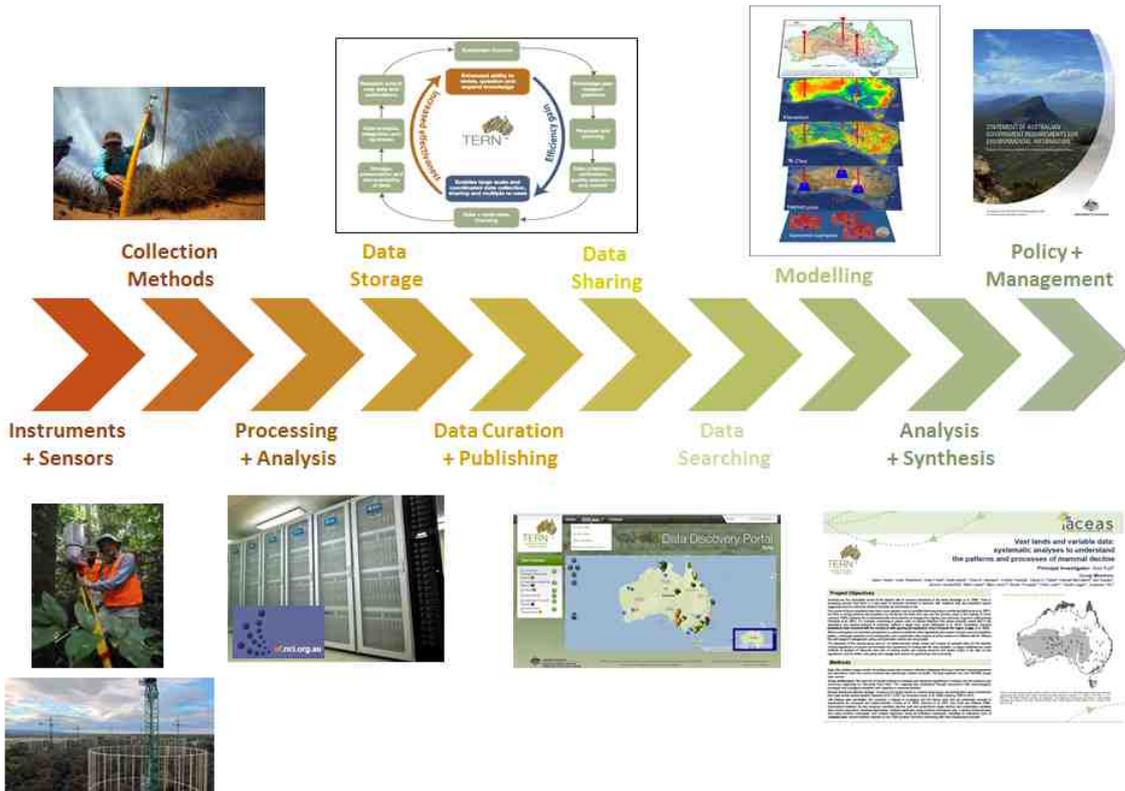
The main area features a map of Australia with numbered markers (1-10) and a table of data items:

Map ref#	Title	Date released
1	Jervis Bay Booderee National Park Plot Network: site-level plot data collected during an assessment of the ecological and cost effectiveness of invasive plant management in achieving conservation goals: a long-term, experimental study, 2014.	20-05-2014
2	Land Cover Dynamics - MODIS, LPDAAC MCD12Q2 mosaic, Australia coverage	16-04-2014
3	Conserving Koalas in the 21st Century: Synthesising the dynamics of Australia's Koala populations	20-05-2014
4	Acoustic Sensor Data - Samford Ecological Facility, 2013	06-01-2014
5	Samford Ecological Research Facility (SERF) Weather Monitor	18-10-2012
6	Prominent Hill (Oxiana) Survey by Ecological Horizons	25-09-2013
7	Samford Ecological Research Facility OzFlux tower site	16-04-2014
8	Avian Survey Data - Samford Ecological Facility October 2013	09-01-2014
9	Desert Ecology Plot Details LTERN Sites (P56T182)	13-06-2014
10	A Vegetation Survey of the Samford Ecological Research Facility (SERF)	15-10-2012

TERN에서 정의되는 데이터셋은 저널 논문과 같이 학문적으로 인용 가능한 출판되어 데이터를 공유하거나 재사용하는데 활용된다. 이를 위해 DOI(Digital Object Identifier), 국제적 ISO 표준을 도입하고 있다. 호주는 국제 DataCite(데이터 인용) 컨소시엄의 회원으로, 호주 ANDS(The Australian National Data Service)에서 DOI 핸들링 서비스를 운영하고 있다. TERN의 DOI-minting 서비스는 이를 활용하여 생성되

어지는 데이터셋에 DOI를 부여하여 관리한다.

TERN에서 제시하는 메타데이터의 항목은 다소 제한적인 편이 있는데, 이는 ANDS에서 공동활용하기 위해 관리되는 메타데이터 항목을 일부 따르고 있으며, 이에 대한 확장된 메타데이터를 TERN에서 운용한다. 메타데이터 항목의 예로, Data set Citation, Data set Owner(s), Associated Party, Abstract, Keywords, Geographic Coverage, Taxonomic Coverage, Contact(s), Methods Info, Access, Project Info, Additional Metadata 등으로 메타데이터가 구성된다.



TERN은 아주 간략한 메타데이터와 링크 정보만을 가지고 있으며 실제 데이터는 각 사이트에 존재하고 관리된다. 각 사이트는 수집한 데이터 셋에 라이선스를 명기하고 고유번호를 위하여 DOI를 발급하여 할당하고 TERN에서 공유하기 위한 간단한 메타데이터를 생성하여 TERN에 관련정보 등록 요청을 하면 TERN Data Discovery Portal의 관리도구에서 해당 메타데이터를 승인하고 등록하게 된다.

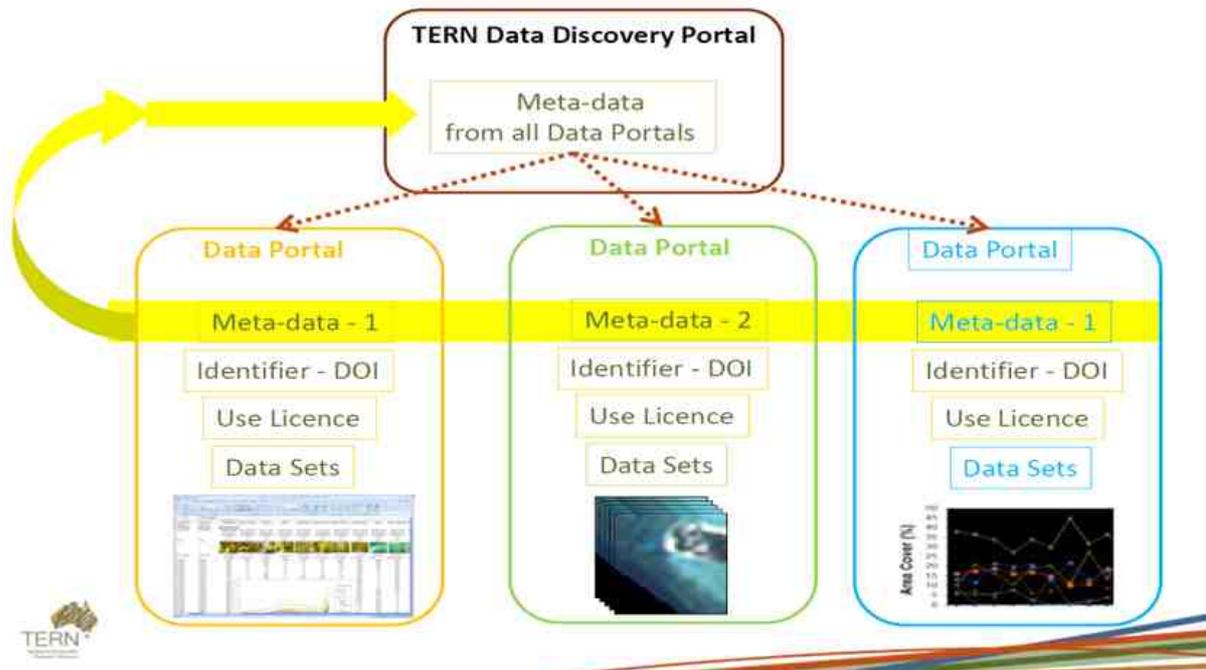


그림 12 호주 TERN 데이터 구성

TERN의 데이터 인프라스트럭처는 다양한 형태의 데이터를 수집할 수 있는 구조로 되어 있으며 데이터의 외부 연계는 Data 카탈로그 포털에서 제공하는 Search API를 통하여 이루어진다.

TERN Data Infrastructure: Key Elements

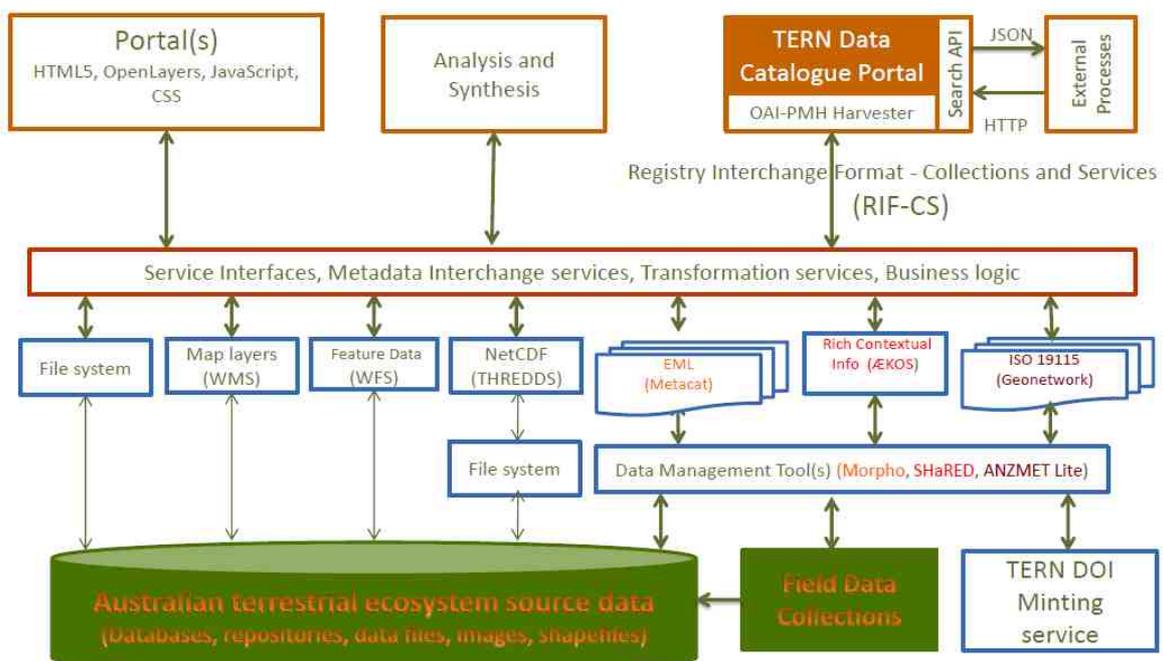


그림 13 호주 TERN 데이터 인프라스트럭처

호주 생태정보 지식시스템으로 TERN에 참가하고 있는 기관중 하나인 Eco-informatics에서 운영하는 AEKOS는 호주의 생태관련 수집데이터를 모두 담아두는 리파지토리 역할을 수행하고 있다.

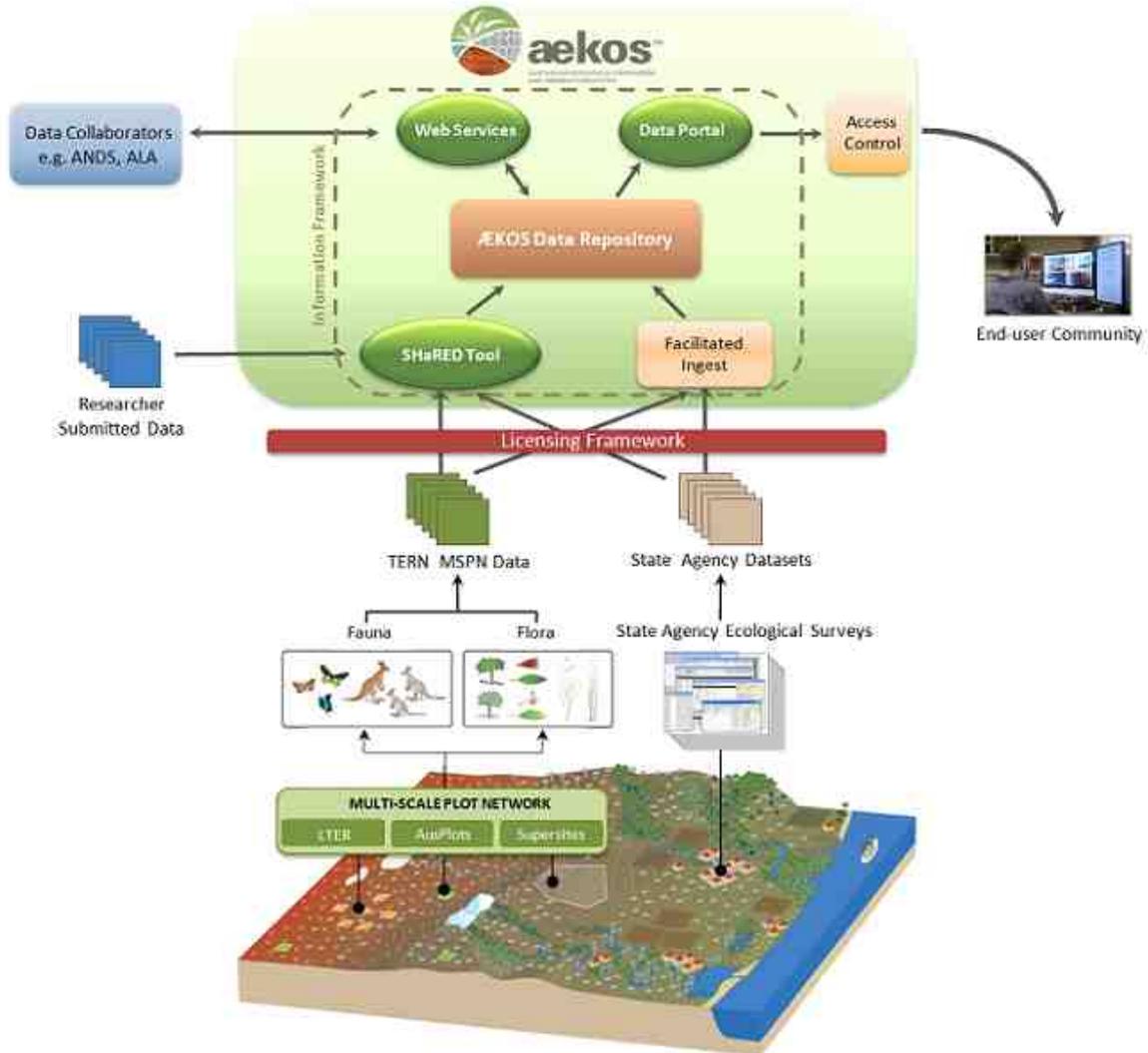
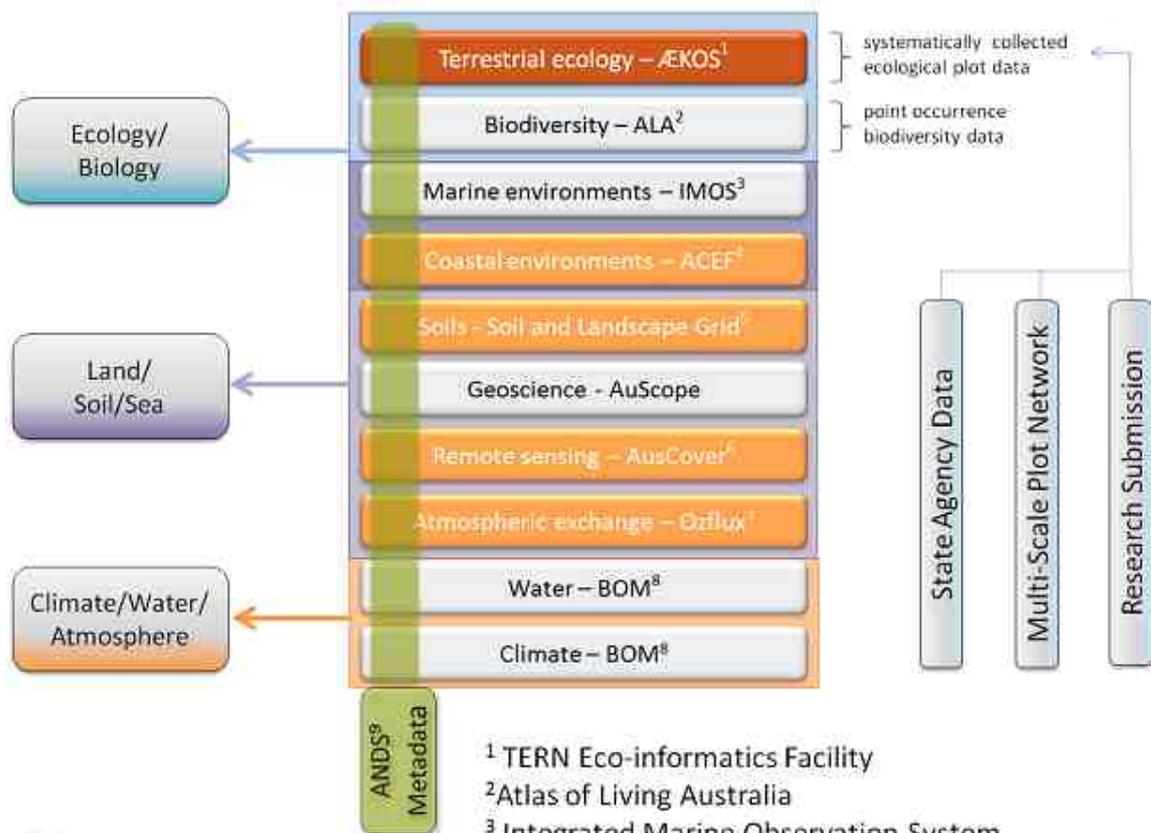


그림 14 AEKOS 시스템

TERN과 AEKOS의 데이터 측면에서의 차이점은 TERN의 경우 산하 기관의 사이트에서 데이터를 관리하고 정해진 규격에 따라 아주 간단한 메타데이터만 TERN에서 관리하고 서비스를 제공하는 반면 AEKOS는 메타데이터와 실제 관측 데이터를 모두 AEKOS에서 수집하여 저장하고 서비스를 제공하고 있는 점이다.



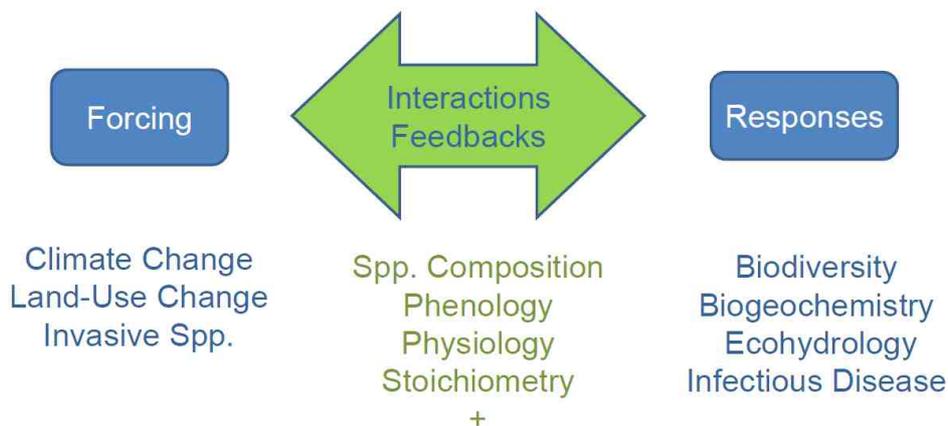
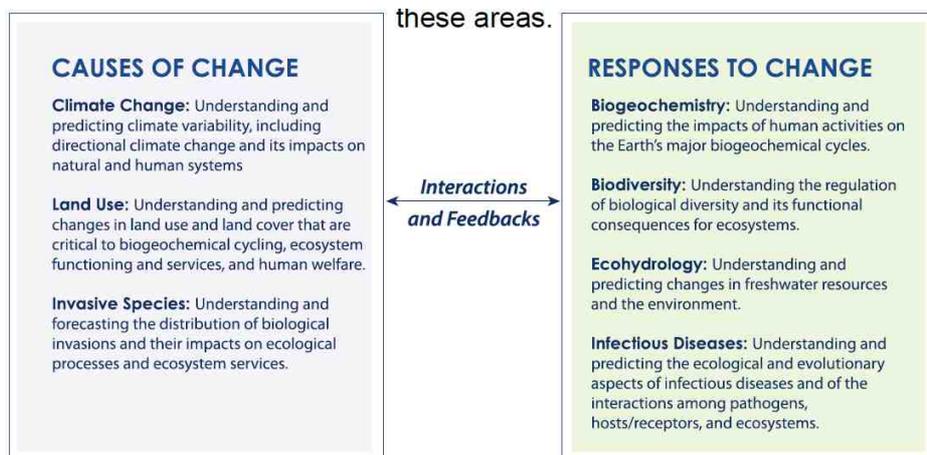
- ¹ TERN Eco-informatics Facility
- ² Atlas of Living Australia
- ³ Integrated Marine Observation System
- ⁴ TERN Australian Coastal Ecosystems Facility
- ⁵ TERN Soil and Landscape Grid
- ⁶ TERN AusCover Facility
- ⁷ TERN OzFlux (Carbon, Water, Energy) Facility
- ⁸ Australian Government Bureau of Meteorology
- ⁹ Australian National Data Service

Orange boxes represent TERN facilities

그림 15 AEKOS 시스템 구조

3. NEON(National Ecological Observatory Network, 미국)

미국 NEON(National Ecological Observatory Network)은 미국 대륙을 대상으로 기후변화, 토지이용변화, 외래종의 침입 등이 생태계에 미치는 영향을 파악하고 예측하기 위한 생태계 정보 네트워크 인프라(Information & Physical Infrastructure)를 제공하는 것이 목적으로 한다.



생태학적 데이터와 정보에 대한 오픈 액세스를 통해 과학자, 교육자, 및 정책 결정자들에게 생태계에 대한 인류의 영향을 파악하고 예측하고 매핑함으로써 생태학적으로 중요한 이슈들을 해결할 수 있는 토대를 마련하고자 한다. 또한 NEON Cyber Infrastructure는 이러한 NEON의 목적을 달성하기 위한 생태 데이터 수집 (Production), 가공(Analysis) 및 가시화(Visualization)를 위한 플랫폼을 제공하는 것이 목표이다. 전체적인 구조는 사람에 의해 수집되고 관측된 데이터를 처리하기 위한 Observation and Sample Management(OSM), 수집된 데이터를 가공하기 위한 모듈 간의 연계 및 통신을 위한 Message Handling System(CMES), 데이터에 대한 QA(Quality Assurance) 및 운영을 위한 Operations Support System(OSS), 다양한 알고리즘을 활용하여 데이터를 가공하는 Data Processing Management System(DPMS), 가공된 데이터 및 이에 대한 메타데이터를 관리하는 Common Data Services(CDS), 과학자/교육자들에게 데이터를 공유하고 검색할 수 있도록 제공하는 Science & Education Portal로 구성되어 있다. 이 중 70%이상을 COTS/Open Source S/W로 구현하려는 계획을 가지고 있다.

NEON Framework

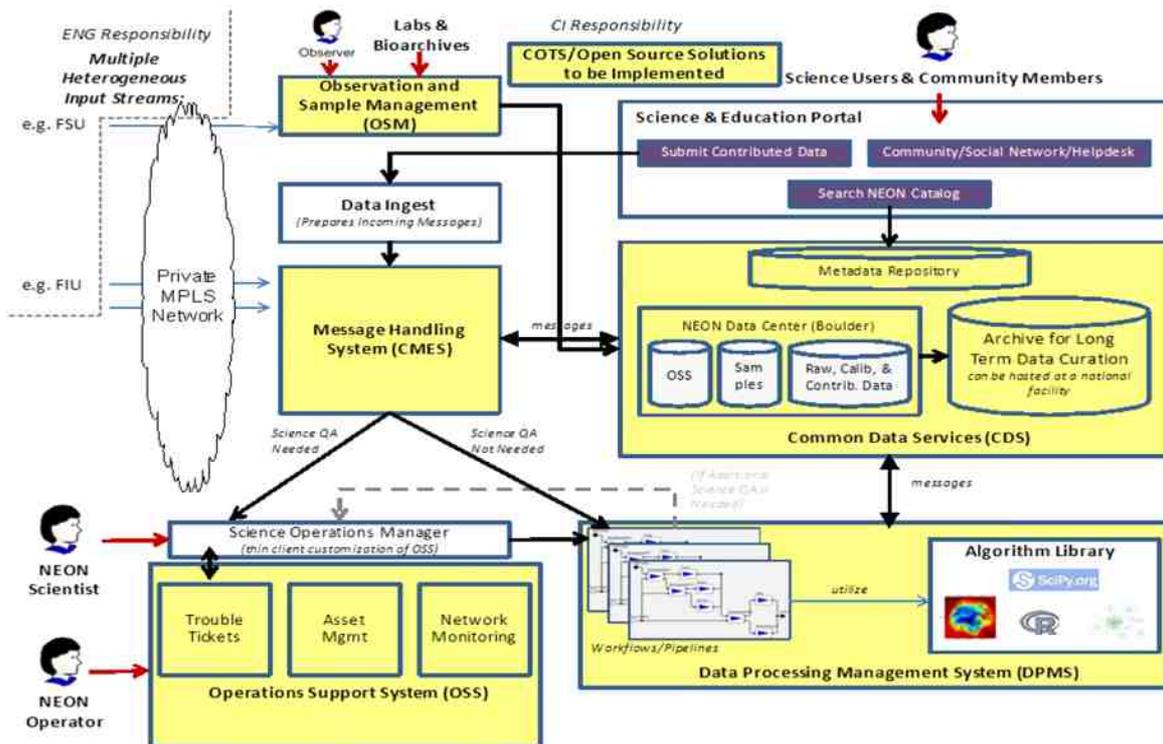
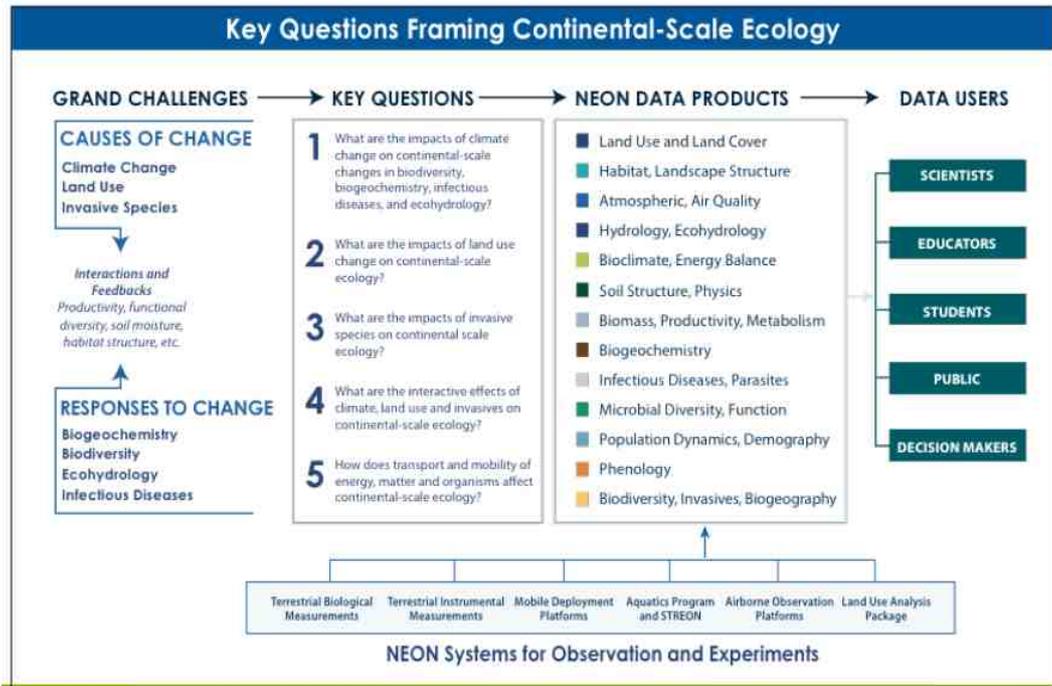


그림 19 NEON Cyber Infrastructure

NEON 프로젝트는 2004년에서 2011년에 걸치는 개념 정립 및 설계를 거쳐 2012년부터 2017년까지 미국 전역에 주요 관측소들을 설립할 예정임. 이후 2017년부터 2046년

까지 생태 데이터를 수집 및 가공할 예정인데, 이러한 일정에 맞추어서 Cyber Infrastructure도 같이 준비해나가는 것으로 보이나, 아직까지 구체적인 설계와 사용될 기술 등이 정해지지 않은 것으로 보인다. NEON의 데이터 수집 시스템은 방대한 넓이의 미국 대륙을 대상으로 생태 관측 정보를 수집해야한다. 종전의 관측원이 직접 관측하는 방법과 또는 관측장비를 활용하는 방법을 모두 수용하는 FSU(Fundamental Sentinel Unit) 뿐만 아니라, 타워 형태의 대형 관측시설을 도메인마다 설치하여 장기적인 관측을 시도하고 있다. 고정된 관측시설을 보완하기 위해 MRP(Mobile Relocatable Platform)인 RTS(Relocatable Tower System)도 활용하여 추가적 관측 지역을 커버하기도 한다.

4. 장기생태연구(한국)

장기생태연구는 2004년부터 연구기반 및 시설 설치를 시작으로 조사지점 기초조사를 시작했다. 2007년부터는 사업의 정착 및 지소를 확립하고 생태분류군에 대한 심화연구를 수행하였고, 2010년부터 사업 종료, 2013년까지 생태계 장기변화 예측과 한반도 생태계 보전 대책을 수립함으로써 장기생태연구에 기여하였다.

기후 변화는 한반도 환경변화의 주요 원인 중 하나로, 자연생태계 교란 및 생물다양성의 변화가 우려되어진다. 상세하고, 과학적인 미래 기후변화 시나리오를 바탕으로 미래 기후를 전망하는 것이 생태계변화 관리체계를 확립하는 큰 요소가 될 수 있다.

기후변화에 따른 환경변화를 모니터링하고 그 결과를 추적하여 한반도에서 현재와 미래에 일어날 수 있는 환경문제에 지속적인 대응 방안을 마련하는 것이 장기생태연구의 가장 큰 목적이라 할 수 있다. 이를 위해 국가장기생태 연구에서는 생태계 변화에 대한 모니터링 자료를 지속적으로 수집할 수 있는 시스템 구축을 했다. 데이터베이스 구축은 크게 육상생태계, 담수생태계, 연안생태계, 동물생태계로 4개 대분야 80여개 이상의 지소별로 이뤄졌고, 지소별 데이터베이스 구축은 같은 분야에 대해 같은 항목으로 수집을 하여, 지역별 동일한 장소와 내용에 대해 자료의 비교분석을 가능하게 하였다.

장기생태연구는 GIS-DB를 통해 연구성과의 홍보, 자료축적을 통한 DB구축을 하였고, 연구자들 뿐만 아니라 대국민도 시각적으로 쉽게 이해할 수 있도록 개발되었다.

ILTER와 차별점은 ILTER는 전문가 집단을 위주로 운영되어, 전문가들이 개발한 Morpho, Kepler, EML을 등을 이용하기에 프로그램 다운로드, 설치 및 사용방법에 대한 전문성이 요구된다. 반면에 장기생태연구 사이트는 DB를 기반으로 PHP를 이용한 개발로 이뤄졌으며, 생태정보와 함께 공간정보의 활용으로 웹상에서 쉽게 테이블 리스트, 그래프, 통계, 메타데이터를 지도와 병행하여 결과를 볼 수 있도록 되어 있다.



국가장기생태연구 메인 페이지

GIS-DB 조사지역구분

조사지역구분	조사지역구분	조사지역구분	조사지역구분	조사지역구분	조사지역구분	조사지역구분	조사지역구분	조사지역구분	조사지역구분	조사지역구분	조사지역구분	조사지역구분	조사지역구분	조사지역구분	조사지역구분	조사지역구분	조사지역구분	조사지역구분	조사지역구분	조사지역구분
조사지역구분																				

조사지역 메타데이터



조사지역 연도별 통계

○ 문제점

장기생태연구 사이트는 국내에서 처음으로 생태정보시스템 구축의 장기적 모델을 제시했다는 점에서 그 의미가 크다. 하지만, 10년동안 제한적 DB 구축의 문제점으로 장기생태연구가 국내에 국한 되는 문제점이 있었다. 장기생태연구는 ILTER라는 국제 기구를

중심으로 각국의 데이터를 서로 교환하고 비교 분석할 수 있는 체제를 갖추고 있지만, 장기생태연구 사이트는 국제표준 메타데이터인 EML의 한국어 버전이 존재치 않아 국제적 공유의 단절이 있다. 또한 개별 연구거점들에서 생산된 정보를 관리하는 책임자의 부재와 국가차원에서 종합 관리하는 관리자가 없는 것도 그들의 보고서에서 문제점으로 언급되었다.

5. 소결론 - 비교/분석

	ECN	호주 TERN	미국 NEON	장기생태연구
(Overview)	-UK's long-term environmental monitoring and research programme -collect, analyse and interpret long-term data from a network of sites	-Infrastructure and network to support coordinated, collaborative ecosystem science community -operate as a network of various facilities	-understanding and forecasting of the impacts of climate change, land use change, and invasive species on continental-scale ecology -Planning/design phase (2012), Full operation (by approximately 2017)	-10 Year Plan of National Long-term Ecological Research * 기후변화와 환경오염에 따른 생태계 변화 및 조성에 관한 연구 * 모니터링을 통해 이들 사이의 상관관계를 규명 * 미래의 환경변화에 능동적으로 대처 및 국가차원의 대응 전략을 수립 * 장기적으로 국내 생태학 및 환경관련 연구 인력을 양성
주요 특징 (Key Features)	-Long-term observations science -Standard measurements and protocols -Quality assured, accessible data -Site Network -Whole ecosystem approach -Research platforms -Community access to site and data -Partnership based	-Long-term observations science -Standard measurements and protocols -Quality assured, accessible data -Site Network -Whole ecosystem approach -Research platforms -Community access to site and data -Partnership based	-World's first ecological observatory-Long-term observations science -Solely funded by the NSF -Free and open access NEO data and data products -Timeframe * A decade of community input and design * Beginning of a 5-year construction period * 30 year period of observations -Experiment infrastructure backbone for experiment s.e.g. STREON exp. -Standard measurements and protocols -Quality assured, accessible data -Site Network -Whole ecosystem approach -Research platforms -Community access to site and data -Partnership based	-연구기반 및 시설 설치 -조사지점 기초조사 -지소 확립 -생태계 장기 변화 예측 능력 배양 -한반도 생태계 보전 대책 수립 -국제 협력:ILTER
기간 (Period)	- 1992 ~	- 2010 ~	- 2012 ~ (omitting community input and design)	- 2004 ~ 2013
ECN/TERN sites	-45 ECN Freshwater Sites -12 ECN Terrestrial Sites	-AEKOS -ACEF -Soil and Landscape Grid -AusCover -OxFlux	-106 sites * 60 terrestrial * 36 aquatic * 10 aquatic experimental	
Extended Networks	-UK Environmental Change Network-12+45 sites -UK Environmental Change	-AusCover (data accessible via TDDP or AusCover portal)	-20 Domains * 20 Core sites (wildland)	-21 Coresites * 9 Terrestrial core sites

	ECN	호주 TERN	미국 NEON	장기생태연구
(ECN)or Facilities(TERN)	eBiodiversityNetwork*-4 1sites -UKUplandWatersMonito ringNetwork-22sites -UKLakeEcologicalObser vatoryNetwork-11sites -HabitatsMonitoringNet work-7sites * ECN terrestrial sites are part of ECBN	-AusPlots(viaAEKOSdata portalorSoilstoSatellitesP ortal,inthefuture,willbese archableviaTDDP) -TheAustralianCentreforE cologicalAnalysisandSynt hesis(ACEAS)(viaTDDPor ACEASportal) -TheAustralianCoastalEc osystemsFacility(ACEF)(vi aTDDPorACEFportal) -TheAustralianSupersite Network(ASN)(viaTDDPo rASNportal) -TheAustralianTransectN etwork(ATN)(viaAEKOSor SoilstoSatellitesportal,int hefuture,willbesearchabl eviaTDDP) -Eco-Informatics(viaAEK OSportal,inprogressofsu bmittingmetadatatotDD P)-TERN 장비가없는CIFacility -TheEcosystemModelling andScalingInfrastructure(e-MAST)(partiallyavailabl eviaeMASTportal,inprogr essofsubmittingmetadat atotoTDDP) -TheLongTermEcological ResearchNetwork(LTERN) (viaTDDPorLTERNportal) -OxFlux(viaTDDPorOzFlu xportal) -SoilandLandscapeGridof Australia(underdevelopm ent,Bestavailabledatapro ductsviaTDDP)	* 40 Relocatable sites (land-use sites) * 10 Mobile laboratories * 3 Airborne Observation Platforms -DataAcquisitionGroupso f20Domains * FIU: Fundamental Instrument Unit * FSU: Fundamental Sentinel Unit * MRP: Mobile Relocatable Platform > RTS: Relocatable Tower System > RDS: Rapid Deployment System * AOP: Airborne Observation Platform * LUAP: Land-use Analysis Package	* 5 Freshwater core sites * 3 Coastal core sites * 4 Animal Ecology core sites
예 산 (Budget)				
주 관 기 관 (Funding Agency)			NSF	환경부/국립환경과학원
생 태 관 측 변 수 (Variables)	~260개 (physical, chemical and biological variables)		100 summary variables	
Datasets	-525개 * 137개 사이트, 5 networks, variables >250 (physical, chemical and biological variables)		-539Level1and118Level4 data -Newdataproducsdevel opedbasedoncommunity inputs	-539Level1and118Level4 data -Newdataproducsdevel opedbasedoncommunity inputs
C o r e datasets		-Plants/Animals -Vegetation -Terrestrialecosystem -Ecologicaldynamics -Freshwater -Landsurface&Soils -Agriculture -Oceans&Coasts -Climate -Human-nature -Energy	-BioclimateSuite * Temperature, p r e c i p i t a t i o n , humidity, radiation -BiodiversitySuite(include sinvasivespp.) * Abundance and diversity (mosquitoes, aquatic invertebrates, beetles, fish, birds, plants, ...)	-기 후 -육 상 -담 수 -연 구 -생 태 -물

	ECN	호주 TERN	미국 NEON	장기생태연구
		-water&gas	<ul style="list-style-type: none"> * Phenology (mosquitoes, beetles, plants) * Microbial function and diversity (functional genes, metagenomes) * Bioarchive (all taxa, substrates) -BiogeochemistrySuite <ul style="list-style-type: none"> * Carbon stocks, fluxes, isotopes * Nutrient stocks, fluxes, isotopes * Chemical climate (N-deposition, Ozone) -EcohydrologySuite <ul style="list-style-type: none"> * Water balance components (storage and flows) -InfectiousDiseaseSuite <ul style="list-style-type: none"> * Disease prevalence (Dengue, Hanta virus, Lyme disease, WestNileVirus) -LandUseandLandCoverSuite <ul style="list-style-type: none"> * Remote sensing data (vegetation performance and structure) * Geographic data (topography, historical climate, etc.) * Statistical data (human geography) 	
구축	<p>-ECN홈페이지 * http://www.ecn.ac.uk/ * 사업소개(최신뉴스, 이벤트, 홍보 등) * ECN Sites 소개 (12 terrestrial, 45 freshwater) -ECNDataCenter * http://data.ecn.ac.uk/ * Data Discovery Portal (http://data.ecn.ac.uk/Data_discovery/searchresult.s.asp?t=1&rec=ALL)</p>	<p>-TERN홈페이지 * http://www.tern.org.au/ * 사업소개(최신뉴스, 이벤트, 홍보 등) * TERN Theme & Facility별 분류 - T E R N 데 이 터 DiscoveryPortal(TDDP) * http://portal.tern.org.au/ * Dataset별로 분류 * 메타데이터 탐색 기능 제공</p>	<p>-NEON홈페이지 * http://www.neoninc.org/ * 사업소개(최신뉴스, 이벤트, 홍보 등) * 진행사항 * NEON Domain</p>	<p>-국가장기생태연구홈페이지 * http://www.knlter.net * 사업소개(최신뉴스, 이벤트, 홍보 등) * 사업소개(최신뉴스, 이벤트, 홍보, 연구성과 등) * 연구진현황 -GISDB(데이터검색) * 동일 사이트 / GIS-DB 메뉴 * Database별, Attribute별로 분류 및 검색 * 기초 분석툴 제공 * GIS 기반의 검색 기능 제공</p>
DataPortal - ECN Data Center Portal - TERN Data Discovery	<p>-manageanintegrationinformation,whichstoresalldataandmetadatacollectedbythenetworks -OracleRDBMSisusedwithlinkstoArcGISforspatialdatahandling -Dataareregularlysentinfromsitesandarequalityassuredbeforebeingloggedintothedatabase * TDDP와 는 달 리 ECNDataCenterPortal에 실제데이터를오라클DB에 저장및관리</p>	<p>-provideaccesstoallTERNdataanddatacontributedbyotherresearchers *TDDP에서는각데이터포탈의메타데이터만보관하고실제데이터는각데이터(facility)포탈에저장되어있음 *TDDP는 단어 그대로 datadiscovery역할만담당하고실제데이터접근각데이터(facility)포탈을통해접근함) *주요기능:Discovery</p>	<p>- NEON Core CI에서 Data,Metadata관리 -RestrictedPortal:AOP,FSU,LUAP,NEONHQoperationstaff -DataCollection(2017~2046) * 현재 Data available 도메인 없음.</p>	<p>-GISDB * MySQL에서 Geo-spatial 데이터, 데이터, 메타데이터를 통합 관리 * DB 구성: site별 DB 이름, 조사내용별 테이블 이름 관리 * PhpMyadmin를 사이트에서 직접활용하여 데이터 입력(or SQL Gate 활용) * 주요기능:</p>

	ECN	호주 TERN	미국 NEON	장기생태연구
Portal	*ECNDataCenterPortal *ECNDataDiscoveryPortal 페이지 이동 하 여 서 ECNDataset탐색 *주요 기 능:Discover,View,Access			
Org . hierarchy	0.Networks:5ea 1.Sites:137ea	0. Facilities: 20ea	0.Domains:20ea 1.sites:106ea	- 연구지소: 21ea

6. 기타 참고 사이트

"<http://www.ecosystem.nier.go.kr>" 국가생태계정보네트워크

"<http://www.alienplant.nier.go.kr>" 한국의외래식물종합검색시스템

"<http://www.naris.go.kr>" 국가자연사연구종합정보시스템

"<http://www.nature.go.kr>" 국가생물종지식정보시스템

"<http://www.kbif.re.kr/chm2>" 한국생물다양성정보기구

"<http://www.biodiv.ccbp.re.kr>" 생물자원정보

"<http://www.bioall.org>" 국가생물자원정보포털사이트

"<http://www.bric.postech.ac.kr>" 생물학연구정보센터

III. 생태정보를 위한 메타데이터 카탈로그 구조설계

III. 생태계정보를 위한 메타데이터 카탈로그 구조 설계

1. 메타데이터 카탈로그 시스템 구축 방법 비교 분석

메타데이터 카탈로그 시스템의 구축 방법은 크게 2가지로 분류 할 수 있다. 한 가지 방법은 메타데이터를 단순히 RDBMS에 저장하고 특별한 메타데이터 카탈로그 소프트웨어를 사용하지 않고 직접 접근하는 방법이고, 다른 방법은 메타데이터를 관리하는 소프트웨어를 사용하는 방법이다. RDBMS 기반의 자체 시스템을 구축하는 경우는 메타데이터 관리, 보안 및 공유와 관련된 모든 기능을 구현해야 한다. 오픈소스 기반의 메타데이터 카탈로그 소프트웨어를 사용하여 구축하는 경우 기본적으로 구현되어 있는 기능들을 활용할 수 있고 추가적으로 필요한 부분과 데이터 포털과 연계하는 기능 부분만 구현하여 사용할 수 있는 장점이 있다.

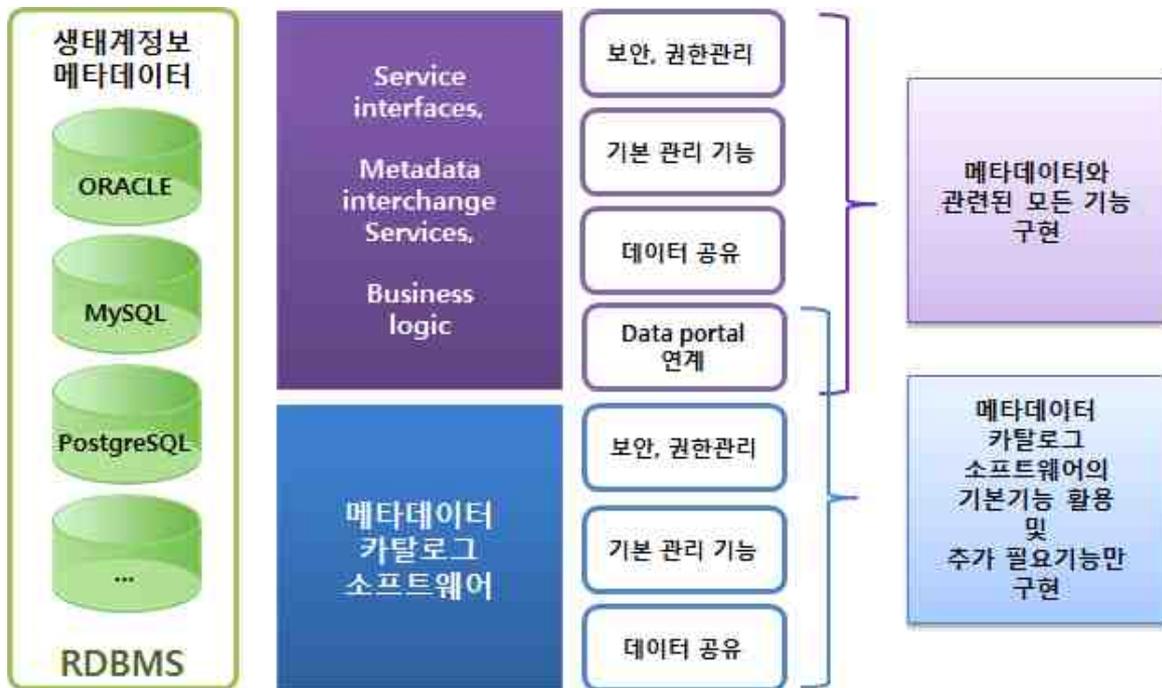


그림 23 메타데이터 카탈로그 서비스 구축 방식

영국의 ECN의 경우 메타데이터 카탈로그 소프트웨어를 사용하지 않고 RDBMS 기반의 자체 시스템을 구축하여 서비스를 하고 있다. 호주의 TERN과 AEKOS의 경우 대부분을 자체 시스템으로 구축하여 서비스 하고 있으나 외부 연계를 위하여 Metacat을 사용하고 있다.

국제 생태 연구 네트워크인 ILTER(International Long-Term Ecological Research Network)에서는 Metacat을 활용하여 생태정보를 공유 하는 것을 권장하고 있다.

2. 해외 생태정보 메타데이터 카탈로그 구조 분석

대부분의 생태계 수집 정보에 대하여 관리, 활용 및 공유의 목적으로 자세한 메타데이터를 생성하여 관리한다. 생태계 정보를 수집하여 관리하는 포털의 정책에 따라 관리되

는 메타데이터가 차이점을 보인다.

영국 ECN의 메타데이터는 엄격한 프로토콜에 의해 관리되고 모든 정보를 수집하므로 매우 자세한 정보를 표준 프로토콜에 의하여 수집하여 관리한다.

Title	ECN Freshwater Flow data from Lathkill
Alternative title	N/A
Dataset language	English
Abstract	Freshwater Discharge data from the UK Environmental Change Network's (ECN) Lathkill site. ...
Topic category	inlandWaters <i>definition: inland water features, drainage systems and their characteristics</i> <i>Examples: rivers, salt lakes, dams, floods, water quality, hydrographic charts</i> environment <i>definition: environmental resources, protection and conservation</i> <i>Examples: environmental pollution, waste storage and treatment, environmental impact assessment, monitoring environmental risk, nature reserves, landscape</i>
Keyword	flow, flux, stage. stream, water discharge
Temporal extent	These data span the period from 07/09/1998 to 31/12/2012
Dataset reference date	07/09/1998
Lineage	Data are collected according to the standard ECN protocol. ...
West bounding longitude	-1.67228269
East bounding longitude	-1.67228269
North bounding latitude	53.17893982
South bounding latitude	53.17893982
Extent	United Kingdom - River Site - Derbyshire, England
Vertical information extent	N/A
Spatial reference system	WGS 84
Spatial resolution	10 metres
Resource locator	http://www.data.ac.uk/Data_discovery/search.asp
Data format	Text files
Responsible organisation	Owner: Environment Agency Distributor: Centre for Ecology and Hydrology Point of Contact: Environmental Change Network
Frequency of update	Continual
Limitation on public access	Licence

Use constraints	Data are available for research and commercial use providing a licence is obtained and ECN and the originators are acknowledged in any publications produced.
A d d i t i o n a l information source	Environmental Change Network website - http://www.ecn.ac.uk ECN Data Centre website - http://data.ecn.ac.uk Protocol document - http://www.ecn.ac.uk/measurements/freshwater/fwd
Metadata date	12 June 2014
Metadata language	English
Metadata point of contact	Susannah Rennie - ECN Data Centre Manager
Unique resource identifier	ECN-ECN-FWD-R06 Code space: http://data.ecn.ac.uk
Resource type	Dataset

표 3 영국 ECN 메타데이터 샘플

호주의 TERN의 경우에는 TERN의 Data Discovery Portal에서 관리하고 제공되는 아주 간단한 메타데이터와 해당 데이터를 수집한 기관에서 제공하는 상세한 메타데이터로 분리되어 관리되고 있다.

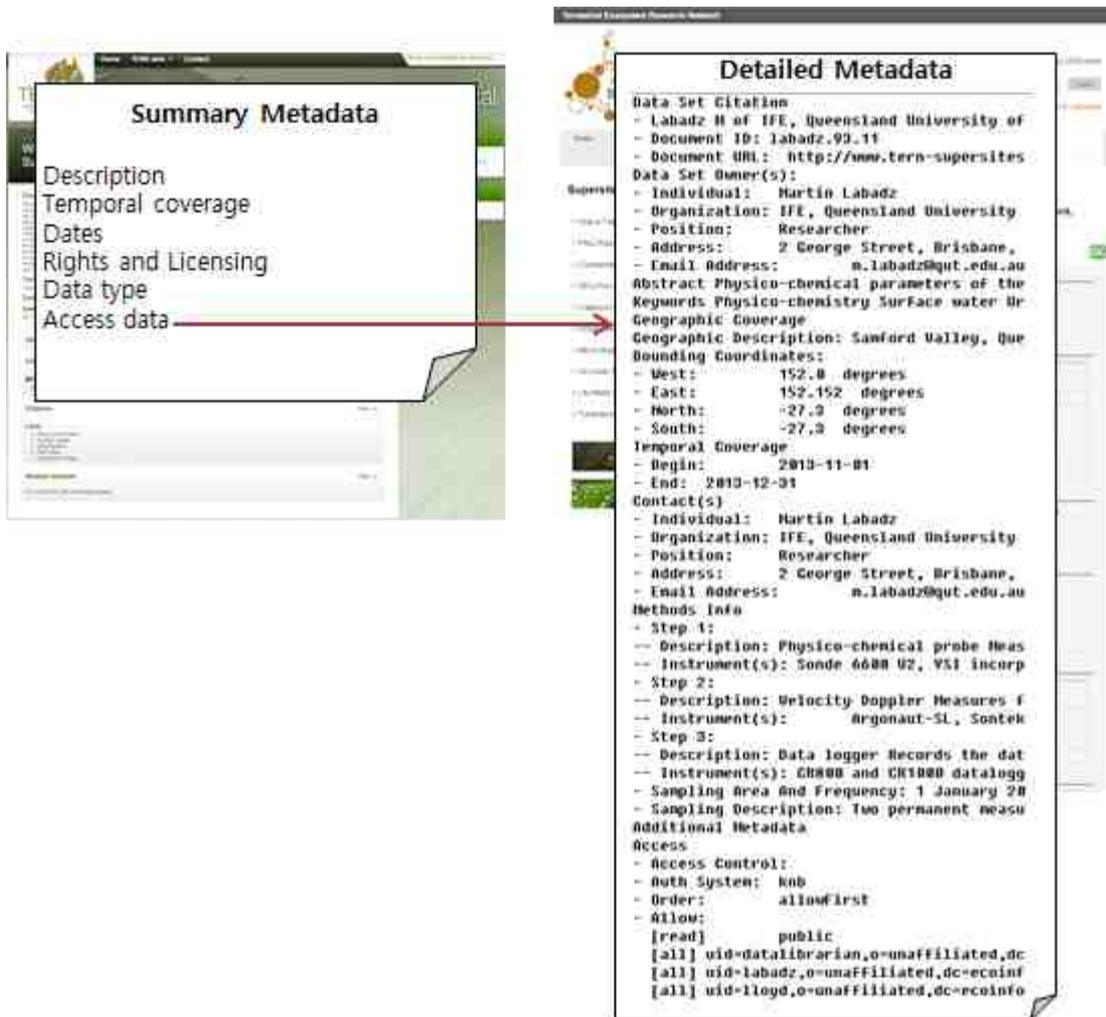


그림 24 호주 TERN의 메타데이터

호주 AEKO의 메타데이터는 데이터를 직접 수집하여 관리하는 사이트이기 때문에 상세한 메타데이터의 성격을 지닌다.

Overview		Scope	
	Title		Temporal
	Program		Spatial
	Custodian		Thematic
	Scope		Taxonomic
	Accessibility		Latest Update
	Abstract		Curation Status
	Objectives		
Scope		Methods	

	Temporal	Conditions of Use	
	Spatial		Citation
	Thematic		License
	Taxonomic		Rights Statement
	Latest Update		Access
	Curation Status		Contact
	Completeness		Resources

표 4 AEKOS 메타데이터 구성

미국의 NEON 또한 데이터를 일괄 수집하기 때문에 상세한 메타데이터를 관리하고 있다. 현재 메타데이터는 데이터와 함께 제공되고 있으며, NEON Data Portal에서 다양한 메타데이터와 관련한 문서를 확인 할 수 있다. 메타데이터에 대한 상세한 검색기능은 제공하지 않고 있다.

3. 한국형 메타데이터 카탈로그 구조 설계

한국형 생태계정보의 효율적인 관리, 활용 및 공유를 지원하기 위하여 본 연구에서는 메타데이터를 기본 메타데이터, 표준 메타데이터, 확장 메타데이터의 3단계의 구조로 설계한다.

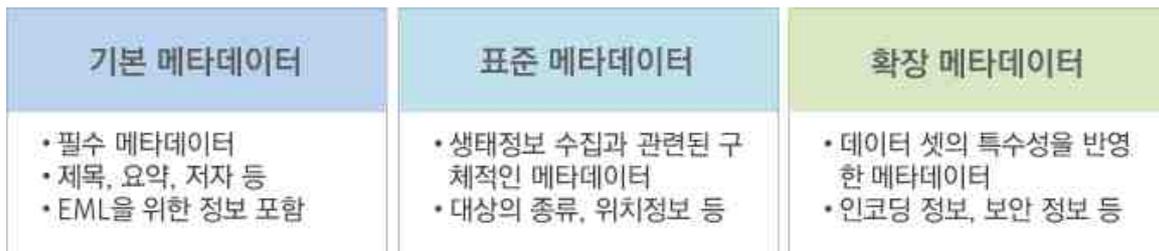


그림 25 메타데이터 3단계 구조 설계

국내에는 현재까지 생태계정보를 통합 관리하는 시스템이 없으며 데이터 수집 및 메타데이터와 관련한 표준 프로토콜이 없는 실정이다. 이러한 상황에서 그동안 수집된 데이터들은 통일성을 찾아보기 힘들다. 이러한 기존의 데이터를 최대한 수용하고 호주 TERN에서와 같이 링크를 통해서 서비스를 제공하기 위해서는 아주 기본적이고 최소화 된 메타데이터가 필요하다. 또한, ILTER의 표준 EML 지원을 위해서도 기본 메타데이터는 반드시 필요하다. 기본 메타데이터는 생태계 정보가 아니어도 모든 정보가 가지는 기본적인 메타데이터라고 생각 할 수 있다.

데이터 항목	설명
제목	데이터 셋의 이름
저자	데이터 셋의 생성자
키워드	데이터 셋의 검색을 위한 키워드
생성일	데이터 셋의 생성일
최종수정일	데이터 셋의 최종 수정일
조사기간	데이터 조사의 기간
라이선스 정보	데이터의 라이선스 정보
고유번호(ID)	데이터를 식별할 수 있는 고유번호
데이터형식	데이터의 형식(ex:텍스트파일, CSV 파일 등)
접근 경로	데이터의 셋의 접근을 위한 URL 정보

표 5 기본 메타데이터

표준 메타데이터는 생태계정보와 관련한 필수적인 요소를 포함하고 관측 대상에 따라 일부 변화가 있는 형태로 설계한다. 관측 대상에 대한 정보와 관측지역, 관측주기, 방법 등 생태계정보 조사의 표준 프로토콜과 관련된 각종 정보가 포함된다.



그림 26 표준 메타데이터 구성요소

기본 메타데이터와의 차이점은 특정 생태정보에 특화된 정보가 기록된다는 점이다. 물론 대부분의 생태정보가 공통적으로 가지는 위치 등의 정보도 있지만 식물의 경우 지름

정보나 나비의 경우 날개길이 정보 등의 특화된 정보가 포함된다. 이러한 특화된 정보는 여러 관측대상을 비교 분석하는데 사용이 가능하다.

확장메타데이터는 특별한 정보를 기록하기 위한 부분이다. 이는 관측 데이터의 포맷과 연관 될 수도 있고, 데이터의 활용, 공유 및 보안 등의 부분과 관련될 수 있다. 또한 데이터의 2차 가공 및 분석 정보를 포함 한다.

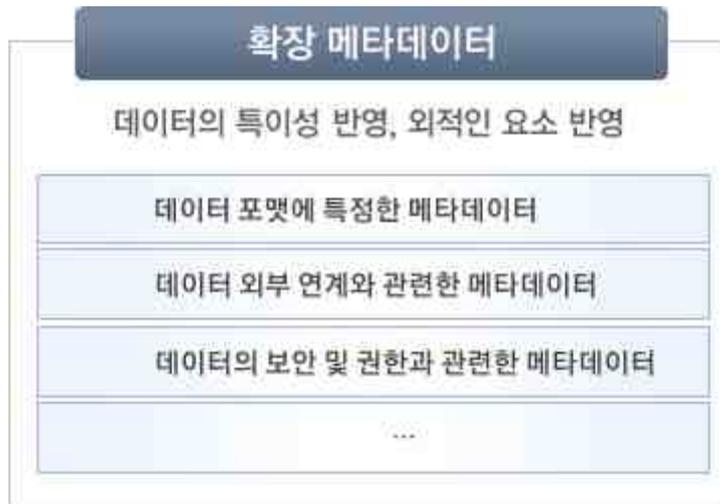


그림 27 확장 메타데이터 특성

그 외의 추가적인 메타데이터로 외부 메타데이터를 생각해 볼 수 있다. 외부 메타데이터는 대부분의 생태관측과 밀접하게 관련을 가지는 기상, 기후, 지형 및 토양정보 등을 말한다. 이러한 외부 데이터는 생태정보 관측 시 센서를 통하여 동시에 수집되기도 하고 외부의 기관과 연계하여 관련 데이터를 활용하기도 한다.



그림 28 외부 메타데이터

IV. 생태정보를 위한 메타데이터 카탈로그 시스템 설계

IV. 생태정보를 위한 메타데이터 카탈로그 시스템 설계

1. 생태계정보 메타데이터 카탈로그 시스템 설계

생태계정보를 위한 메타데이터 시스템 설계를 위해서는 어떠한 데이터가 어떻게 등록되고 저장되며 사용자에게 서비스로 제공되는 일련의 프로세스를 예상하고 이에 필요한 기능들을 설계해야 한다. 또한 구축된 생태정보를 외부에 서비스하기 위해 필요한 기능들을 포함하여 시스템을 설계한다.

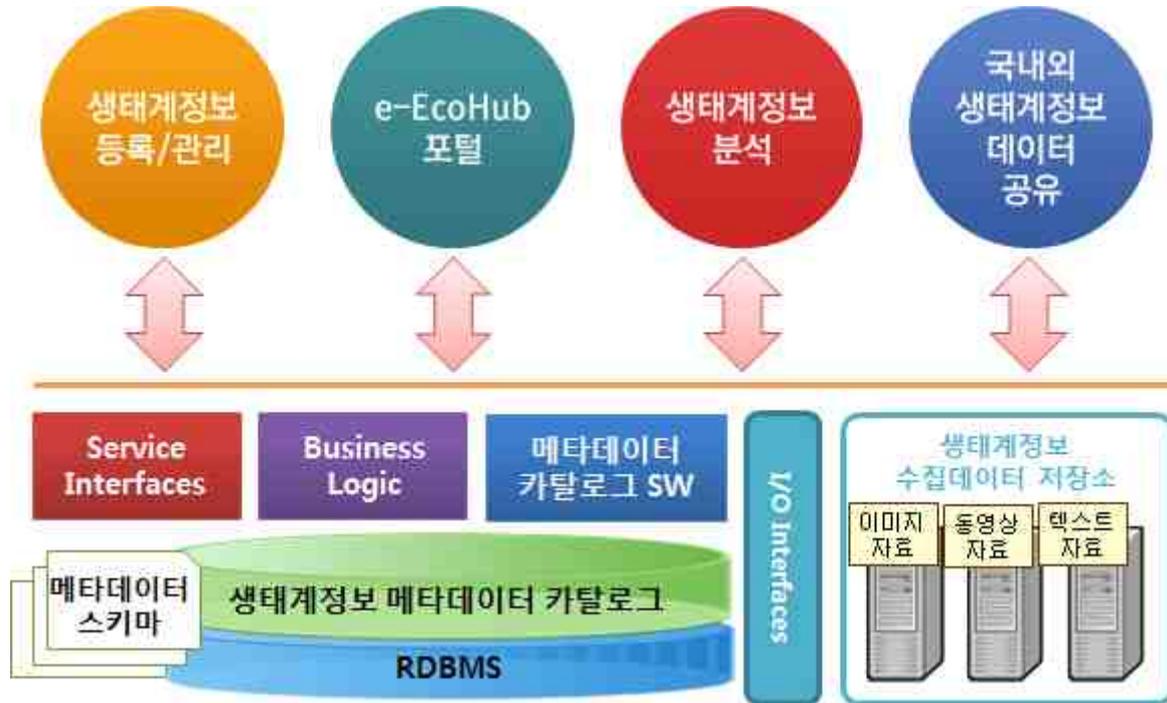


그림 29 생태계정보 메타데이터 카탈로그 시스템 구성도(안)

2. 생태계정보 메타데이터 등록

메타데이터 카탈로그 시스템에 등록되는 생태계정보의 데이터는 크게 기본메타데이터만 등록되는 경우, 표준메타데이터 까지 등록되는 경우와 기존의 데이터를 마이그레이션하는 경우의 3가지 형태로 분류할 수 있다.

기본메타데이터만 등록되는 경우는 비교적 간단한 절차를 통하여 처리 된다. 등록을 요청하는 요구자는 수집한 데이터 셋을 웹 사이트에 등록하고 링크 정보를 포함한 기본 메타데이터를 생성하여 생태계정보 메타데이터 카탈로그 시스템 관리자에게 등록을 요청하게 된다. 통합 생태정보 관리자는 등록 요청을 검토하여 기본 메타데이터 등록 도구를 활용하여 해당 기본 메타데이터를 등록하고 데이터 포털과 연계하여 서비스를 제공한다. 호주의 TERN의 경우도 유사한 절차로 이루어지고 있다.

CASE A. 기본 메타데이터 등록

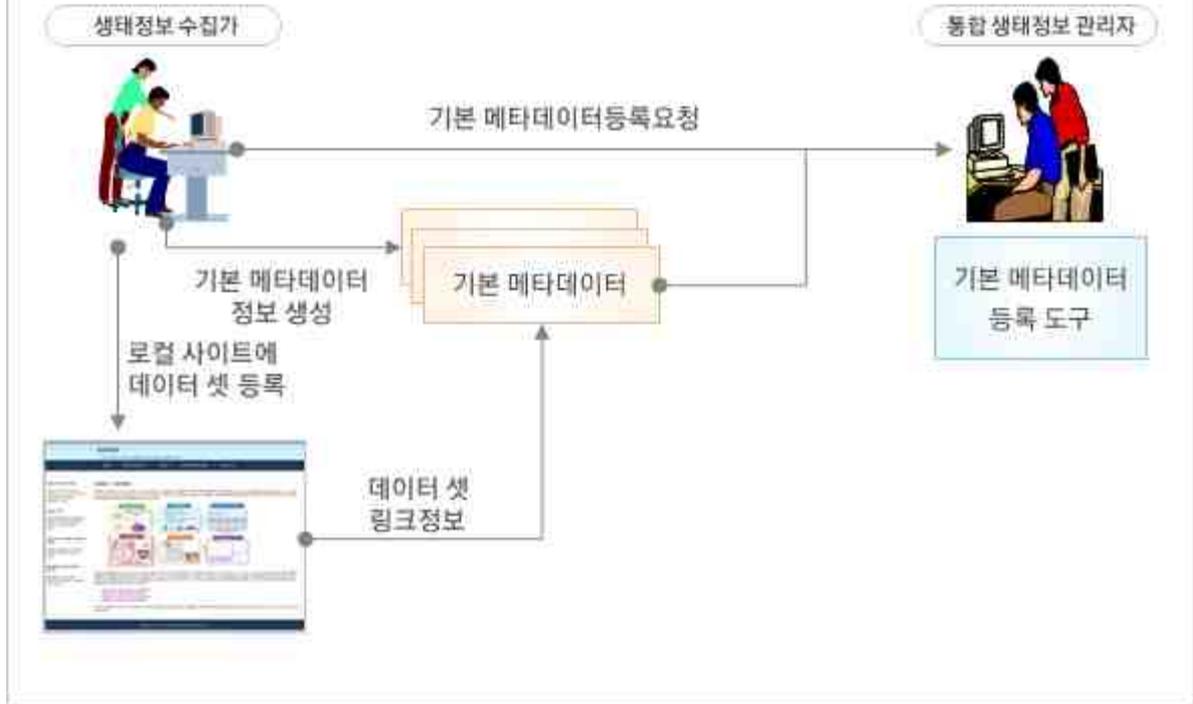


그림 30 기본 메타데이터 등록 절차

수집된 생태정보를 서비스 하거나 저장할 시스템이 없는 경우 또는 필요에 의하여 수집데이터와 메타데이터를 모두 통합 관리해야 하는 경우에는 관련 데이터를 모두 수집하여 저장해야 한다. 이 경우에는 기본 메타데이터와 표준 메타데이터를 모두 수집해야 하고 수집된 메타데이터를 에코-Hub 포털에서 서비스해야 하므로 데이터의 품질관리 절차를 수행해야 한다.

데이터의 등록을 위해서는 생태정보 수집가가 수집한 생태관측 데이터 셋을 기반으로 기본 및 표준 메타데이터를 생성하고 생성한 데이터를 통합생태정보 관리자에게 전송하여 등록을 요청한다. 통합생태정보 관리자는 전송받은 데이터에 대한 QA를 수행하고 등록여부를 결정한다.

수집가의 메타데이터 및 관측 데이터 셋을 전송하는 방법은 메일 등을 통한 직접 전송과 에코-Hub 포털을 통한 등록 방법의 두 가지 방법을 제공한다. 에코-Hub 포털을 통해 등록하는 방법을 사용하는 경우 기본 및 표준 메타데이터 입력 페이지를 통하여 적절한 양식으로 작성되었는지 일차적인 검사가 가능하고 생태관측 데이터 셋에서 기본 및 표준 메타데이터의 일부를 자동 생성하는 기능을 통하여 보다 편리하게 메타데이터를 생성하고 등록 할 수 있다.

통합생태정보 관리자는 등록된 또는 전송되어온 메타데이터 및 데이터 셋에 대하여 정해진 QA 프로세스 프로토콜에 따라 검증하고 이를 등록 도구를 사용하여 통합 생태정

보 메타데이터 저장소와 생태관측 데이터 저장소에 등록한다. 이를 위해서 메타데이터와 수집 데이터 셋의 QA 프로세스를 지원하기 위한 도구가 필요하고 생태관측 저장소와 연계된 데이터 셋 등록 및 저장도구가 필요하다.



그림 31 데이터 셋 및 표준 메타데이터의 등록

기존 생태 데이터를 마이그레이션 하기 위해서는 기존 생태데이터 셋으로부터 메타데이터를 자동으로 생성하는 기능과 새로운 데이터 셋의 형식에 적합하게 데이터 셋을 변환하는 기능이 필요하다. 이 기능을 통하여 새로운 데이터 셋과 기본, 표준 메타데이터를 생성하여 메타데이터 등록 도구를 사용하여 메타데이터와 데이터 셋을 등록한다.

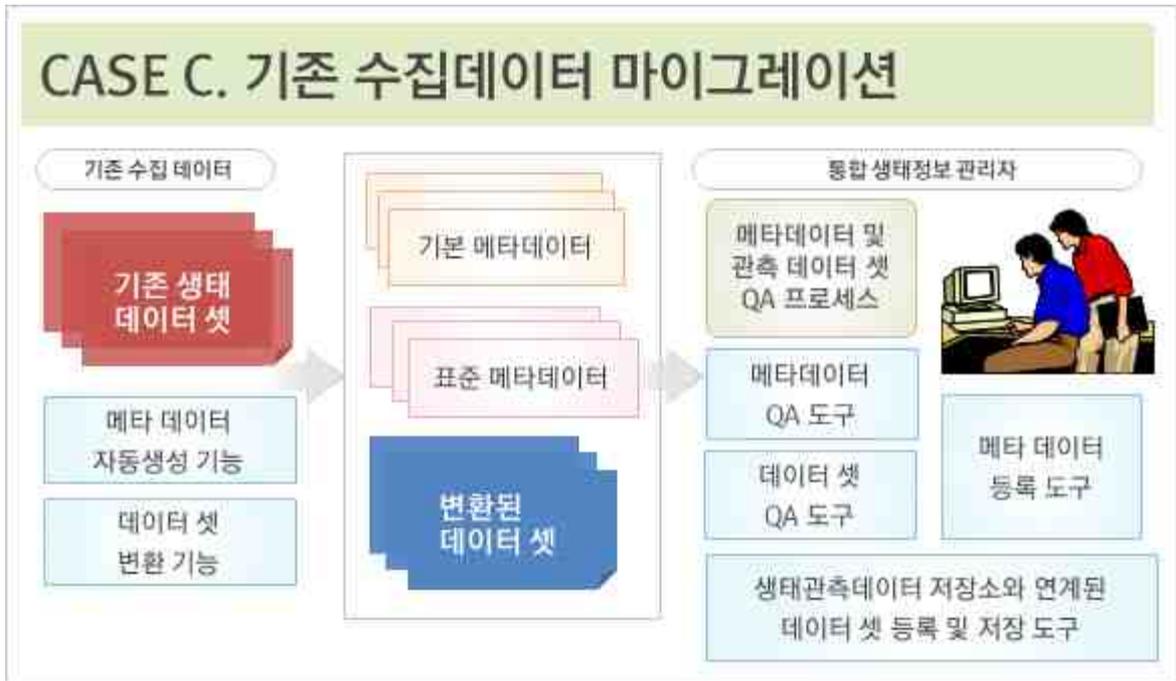


그림 32 기존 수집데이터 마이그레이션

3. e-EcoHub 포털 연계

메타데이터를 사용자에게 제공하는 포털을 위하여 연계 API 및 비즈니스 로직을 구현해야 한다. e-EcoHub에서 메타데이터를 검색하고 검색 결과 및 메타데이터를 화면에 출력하기 위한 API를 제공해야 하며, 외부 포털과 연계를 위한 검색 API를 제공해야 한다.

웹과 데이터베이스를 연동하기 위한 기술에는 JDBC, CGI, 웹서버 API, 데이터베이스 접속을 위한 미들웨어(middleware) 등을 이용하는 방법들이 있다. CGI는 가장 일반적인 방법으로 모든 웹 브라우저를 지원하며 빠른 프로토타이핑과 표준화된 인터페이스를 제공하지만 대화형 프로그램의 작성이 어렵고 서버시스템의 오버헤드가 크며 보안이 취약하다는 단점을 가지고 있다. JDBC는 Java 클라이언트와 관계형 데이터베이스 서버와의 연동을 지원한다. Java 클래스 형태로 제공되는 JDBC API는 ODBC(Open DataBase Connectivity)나 기존의 데이터베이스 접근을 위한 인터페이스의 상위계층을 구현할 목적으로 고안되었다. JDBC의 가장 큰 특징은 통합적인 SQL 데이터베이스를 이용할 수 있는 프레임워크로서 데이터베이스에 비종속적인 분산 응용프로그램을 구현할 수 있도록 해 준다는 것이다. 또한 다양한 데이터베이스 접근방식을 제공함으로써 CGI(Common Gateway Interface)를 이용한 수동적인 방식을 벗어나 보다 능동적인 데이터베이스의 이용이 가능하다. 웹서버 API는 웹서버를 개발하는 업체나 데이터베이스를 개발하는 업체에서 특정한 혹은 자사의 데이터베이스 접근을 위해 제공하는 전용 API 이다. 이 전용 API를 이용하면 개발이 용이하며 CGI보다 훨씬 서버의 부담을 줄일 수 있으나, 특정 서버나 데이터베이스에 종속적이라는 문제점을 안고 있다. Java API나 CGI 스크립트 형태로 제공되는 미들웨어는 HTML(HyperText Markup

Language)에서 바로 이용할 수 있으며 미들웨어가 제공하는 OLTP(OnLine Transaction Processing)서비스 등을 받을 수 있다는 장점을 가지고 있다. 이와 비교해 JDBC의 경우 다양한 시스템 구성이 가능하며 서버의 부담을 대폭 줄이고 경량 클라이언트의 구성이 가능하다는 장점을 가지고 있다. 또한 RMI나 JavaIDL을 이용하여 3-계층분산시스템을 구성한다면 보다 효과적인 분산 환경을 구축할 수 있다.

최근 인터랙티브한 웹 환경의 구현을 위해서는 HTML5가 많이 사용된다. HTML5는 웹 표준 기관인 월드와이드웹 컨소시엄(W3C)이 만들고 있는 차세대 웹 언어 규격이다. HTML5는 문서 작성 중심으로 구성된 기존 표준에 그림, 동영상, 음악 등을 실행하는 기능까지 포함시켰다.

HTML5를 이용해 웹사이트를 만들면 국내 전자상거래에서 많이 쓰이는 액티브X, 동영상이나 음악재생에 필요한 어도비 플래시와 같은 플러그인 기반의 각종 프로그램을 별도로 설치할 필요가 없어진다. HTML5는 모바일환경에서 아이폰이나 안드로이드 등의 운영체제를 가리지 않고 모두 호환된다.

HTML5는 W3C에서 만든 차세대 웹 표준으로써 2004년 WHATWG의 초안으로부터 시작되었다. 이는 시맨틱 마크업, 편리한 웹 폼기능, 리치 웹 애플리케이션 API를 담고 있다. 이는 또한 웹 문서 기능을 그대로 유지하고 웹 브라우저 간의 상호 운용성을 위한 세부적인 규칙을 담고 있다. HTML5는 웹 브라우저 구현에 대한 상세스펙을 제공함으로써 과거 스펙의 호환성을 그대로 유지하면서도 추가되는 버전에 기능사용이 가능하다. 기존의 HTML4 보다 확장된 태그들을 지원하고 문서 구조에 적합한 header, footer, nav, section과 같은 구조화 마크업을 사용할 수 있다. 또한 HTML5는 다양한 API를 제공 한다. 이는 더욱 더 풍부하고 방대한 데이터를 처리하는 것을 가능하게 함으로써 더 좋은 애플리케이션 개발에 사용할 수 있다. 다음은 HTML5에서 제공하는 API 기능이다.

4. 생태계정보 분석 연계

정보의 분석을 위해서는 메타데이터와 원시 데이터 셋이 모두 필요하다. 또한, 대용량의 데이터 중 필요로 하는 데이터를 필터링 하는 기능도 필요하다. 이러한 필터링은 메타데이터와 원시 데이터 모두에 적용되어야 하며 이를 위한 별도의 비즈니스 로직과 서비스 인터페이스가 필요하다.

대부분의 응용과학 분야에서 생산되는 데이터가 다량이거나 그 크기가 큰 경우에 데이터의 효율적인 활용과 관리를 위해서 메타데이터를 사용하고 메타데이터는 분석 대상에 포함될 데이터 셋을 검색하고 데이터 셋의 위치를 알려주는 용도로 주로 사용된다.

이러한 구조가 원활하게 동작하기 위해서는 생태정보 분석 프레임 워크에서 생태정보 메타데이터를 쉽게 접근 할 수 있어야 하며 이를 위해서 생태정보 메타데이터는 관련 라이브러리, 인터페이스 또는 API 등의 접근 방법을 제공해야 한다. 분석 프레임 워크에서는 다양한 응용 소프트웨어가 사용될 수 있으므로 보다 많은 종류의 API가 제공될수록 쉽게 활용 될 수 있다.

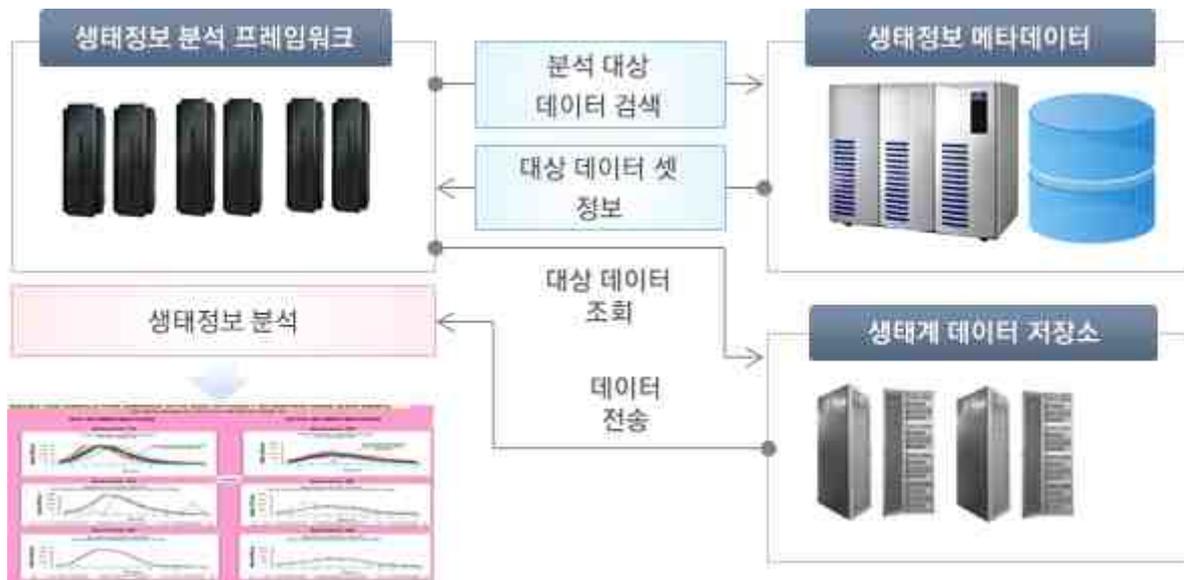


그림 33 생태정보 분석 프레임워크 연계

5. 국내외 생태계정보 공유

생태계 정보를 공유하는 의미는 수집한 관측 데이터 셋의 공유가 아닌 메타데이터의 공유를 의미한다. 어떤 대상에 대하여 언제 어디서 어떻게 관측했는지에 대한 기본 메타데이터 및 표준 메타데이터를 공유하고 여기서 필요로 하는 데이터가 있는 경우 요청하여 공동 활용 할 수 있다. 메타데이터의 국내외 공유를 위한 모델은 Tightly-Coupled Model과 Loosely-Coupled Model로 나눌 수 있다.

Tightly-Coupled Model은 동일한 구조의 메타데이터를 사용하고 이를 메타데이터 카탈로그 소프트웨어를 통하여 실시간으로 공유하는 모델이다. 이를 위해서는 기본조회 기능, 권한 관리기능, 보안기능, 데이터 복제기능 등 다양한 기능을 지원하는 메타데이터 카탈로그 소프트웨어가 필요하다.

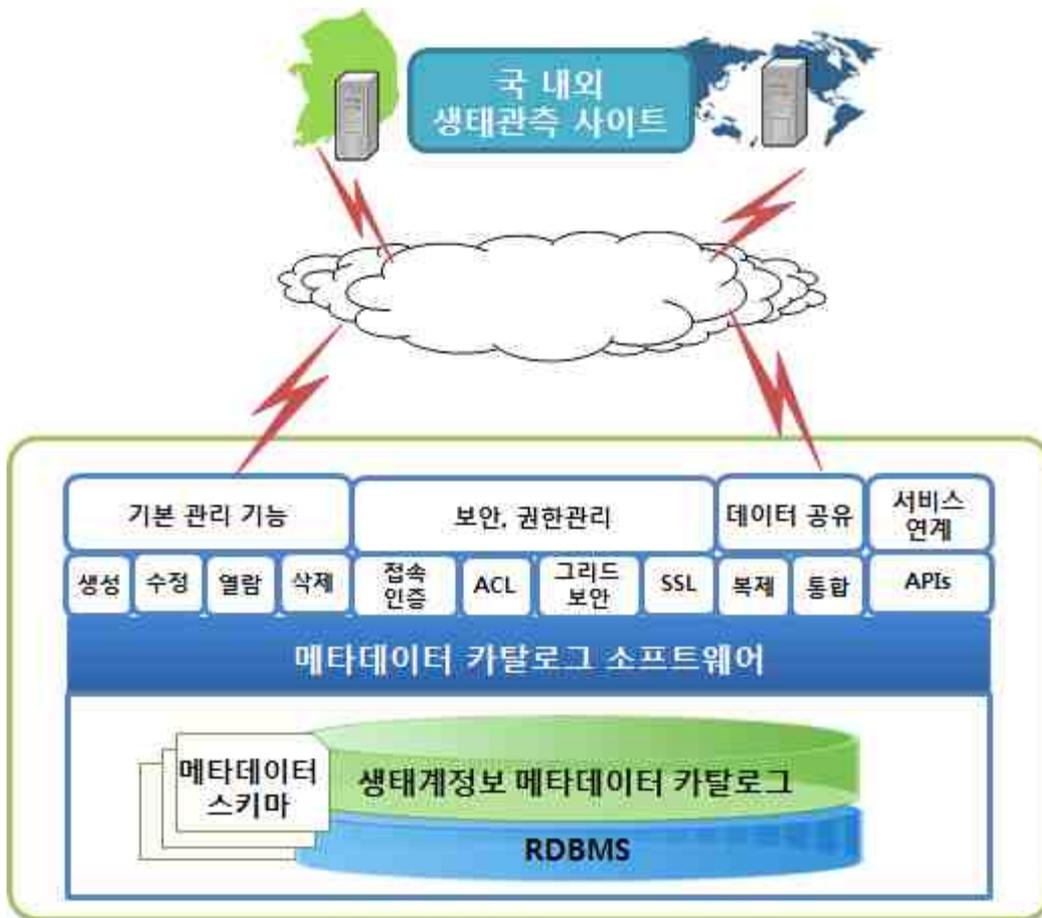


그림 34 Tightly-Coupled 생태계 데이터 공유 모델

이러한 용도로 많이 사용되는 대표적인 메타데이터 카탈로그 소프트웨어로 AMGA가 있다. AMGA는 다양한 RDBMS를 지원하는 메타데이터 카탈로그 시스템이다.

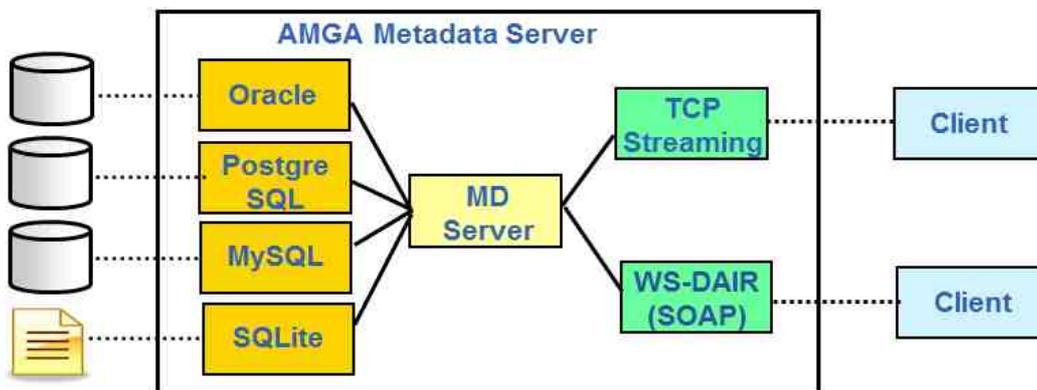


그림 35 메타데이터 카탈로그 시스템 AMGA의 구조도

AMGA의 기술적 특성을 살펴보면 아래와 같다.



- 다양한 인증 방법 지원(ID/Password, GSI, VOMS, X.509)
- 이기종 데이터베이스 지원(Modular back-end: Oracle, MySQL, Postgres, SQLite)
- 표준화된 접근 방식(Modular front-end: TCP Streaming, SOAP WS-DAIR, SSL)
- 멀티-쓰레드/멀티-프로세스 하이브리드 DB
- ACL(Access Control List) 기반 정교한 데이터 권한 설정
- 데이터베이스 접근 표준 인터페이스 지원(SQL)
- 기존 데이터베이스 가져오기(import) 지원
- 계층적 데이터 컬렉션 구조지원(Directory-like structure)
- 언어: AMGA 메타데이터 언어, 네이티브 SQL
- 신뢰성과 확장성을 위한 데이터 복제(Replication) 및 병합(Federation) 제공
- OGF(Open Grid Forum) 표준 WS-DAIR 공식 참조 구현
- AMGA Manager(범용적 사용자 GUI 클라이언트) 제공

Grid metadata catalogue service인 AMGA는 EGEE 프로젝트를 통하여 다양한 분야의 연구 커뮤니티가 원하는 기능을 개발하여 왔다. AMGA는 두 개의 주요 커뮤니티 그룹인 HEP 그룹과 Biomed 그룹의 요구사항을 토대로 개발되었다. HEP 그룹은 수백만 또는 수천만의 많은 수의 파일 및 각 파일과 연계된 메타데이터를 효과적으로 관리하기를 원하였다. Biomed 그룹은 HEP 그룹에 비해 보다 작은 수의 파일과 메타데이터를 사용한다. 하지만 높은 수준의 보안 및 접근제어(ACL:Access Control List)를 요구한다. AMGA는 두 응용의 요구에 의해 유연하고(flexible) 동적이며(dynamic) 계층적인(hierarchy) 스키마(schema)를 지원하는 메타데이터 인터페이스를 구현하였다. 메타데이터의 저장을 위해서는 Oracle, PostgreSQL, MySQL, SQLite의 다양한 RDBMS를 사용한다. AMGA는 X.509기반의 grid authentication과 VOMS의 지원, 세밀한 ACL 등의 보안 기능을 제공한다. 이러한 AMGA의 데이터 공유를 위한 권한 관리 기능은 생태계 정보의 공유에도 적용이 가능하다.

또한 AMGA는 데이터의 공유를 위하여 Replication과 Federation을 지원한다. AMGA의 Replication 기능은 동일한 데이터를 여러 서버에 복제하여 저장하는 기술이고 Federation은 여러 서버에 분산되어 있는 데이터를 하나의 서버에 존재하는 데이터처럼 접근할 수 있게 하는 기술이다.

비슷한 기능을 지원하는 메타데이터 카탈로그 소프트웨어로 Metacat이 있다. AMGA 기능이 간단한 소프트웨어로 ILTER에서는 Metacat을 사용한 메타데이터 공유를 권장하고 있다.

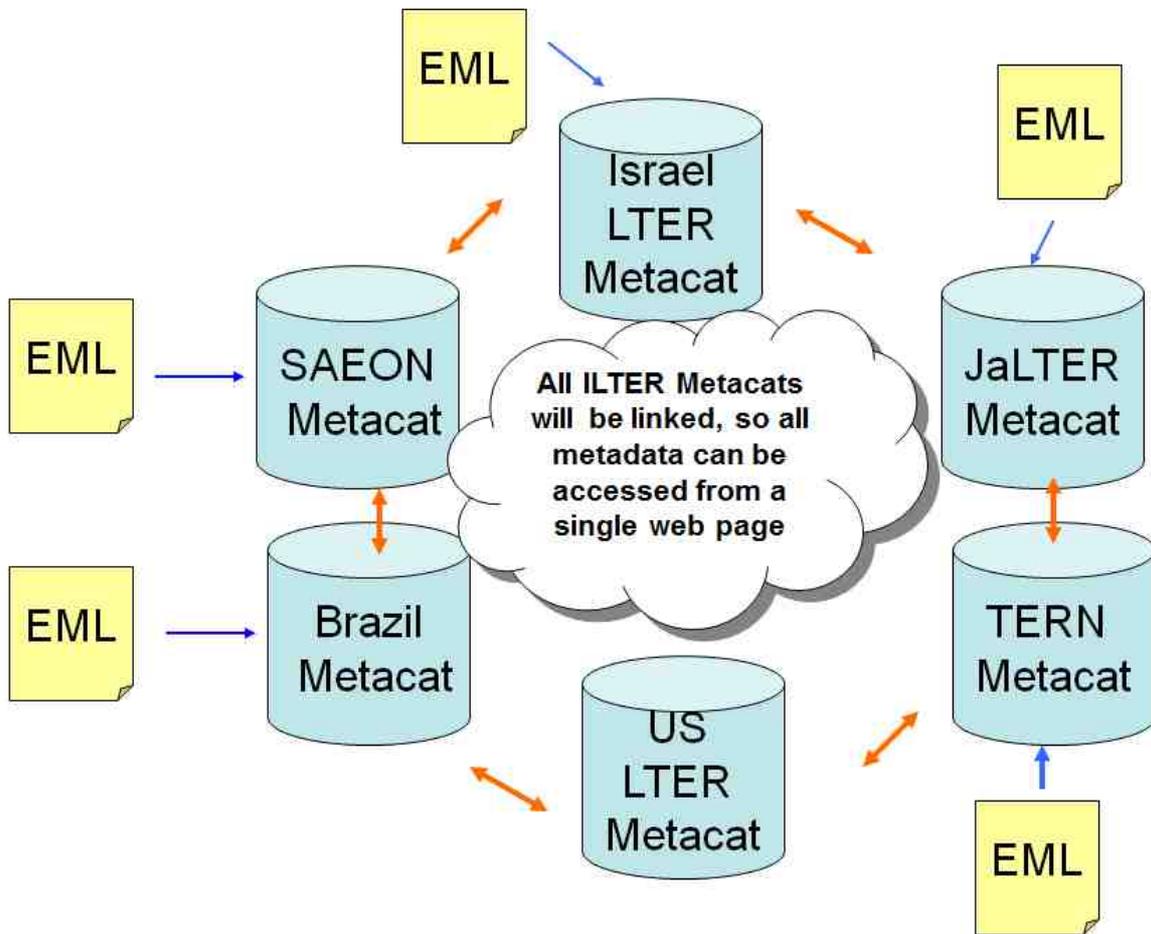


그림 36 Metacat을 활용한 EML 공유

Loosely-Coupled Model은 서로 다른 구조의 메타데이터를 사용하고 있더라도 표준으로 정한 항목을 정해진 포맷으로 공유하는 방식을 말한다. 가장 대표적인 방법으로 ILTER에서 권유하고 있는 Morpho를 사용한 공유가 있다.

Morpho는 XML 에디터로 EML의 데이터를 정해진 양식대로 XML 포맷으로 생성해 준다. 반드시 포함되어야 하는 항목은 다음과 같다.

ILTER EML 패키지에 반드시 포함되어야 하는 항목

- Title
- Owner
- Abstract
- Keywords
- Package-ID(EML 패키지를 구분하는 유니크한 값)

V. 향후계획 및 기대효과

V. 향후계획 및 기대효과

1. 향후계획

- 생태계정보를 위한 메타데이터 구조 상세 설계
 - 실제 관측 데이터를 기반으로 메타데이터 구조 상세 설계
 - 생태계 정보 관측자 및 연구자의 요구사항을 수렴하여 설계에 반영
 - 총괄 과제에서 제시되는 프로토콜 반영
 - 총괄과제의 선정 생태종에 대한 상세 구조 우선 설계
- 생태계정보를 위한 메타데이터 카탈로그 시스템 기능 구현
 - 생태계정보 메타데이터 등록/관리 기능 구현
 - 에코-Hub와 연계 기능 구현
 - 생태계정보 메타데이터 카탈로그 시스템 사용자를 대상으로 인터뷰 실시
 - 요구사항을 반영한 기능 구현
- 생태계정보 메타데이터 국내외 공유 기능 구현
 - 국내 생태계 정보 관련 사이트와의 업무 협의를 통한 메타데이터 공유 방안 마련
 - 메타데이터 공유 정책 수립 및 시스템 구축에 반영
 - ILTER의 EML 공유 관련 기능 반영
 - 메타데이터 공유를 위한 AMGA 등의 메타데이터 카탈로그 시스템 도입 여부 결정

2. 기대효과

- 국내 생태계정보를 효과적으로 관리 활용할 수 있는 메타데이터 구조를 설계하여 생태계정보 활용성 극대화에 기여
- 국내외 생태계정보 메타데이터 공유를 통하여 데이터 활용도 향상 및 중복 조사의 비용 절감
- 생태계정보의 수집 현황 파악 및 데이터의 쉬운 검색과 열람을 지원함으로써 정책 결정에 기여
- 표준화된 메타데이터 관리 시스템 구축 및 활용을 통한 생태데이터의 체계적 관리체계 구축

VI. 참고문헌

II. 참고 문헌

- [1] ECN : <http://www.ecn.ac.uk/>
- [2] TERN : <http://www.tern.org.au>
- [3] NEON : <http://www.neoninc.org/>
- [4] ANDS : ands.org.au
- [5] TERN Eco-informatics : <http://www.ecoinformatics.org.au>
- [6] ISO 19115 :
http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=53798
- [7] EML :
<https://knb.ecoinformatics.org/#external//emlparser/docs/eml-2.1.1/index.html>
- [8] AEKOS : <http://www.aekos.org.au/home>
- [9] ILTER : <http://www.ilternet.edu/>
- [10] UNEP : <http://www.unep.org/>
- [11] Lane, AMJ. (1997). The UK Environmental Change Network Database. An Integrated Information Resource for Long-term Monitoring and Research. *Journal of Environmental Management*, 51 (1), 87-105.
- [12] Lane, A.M.J., Rennie, S.C. and Watkins, J.M. (2004). Integrated Data Management for Environmental Monitoring Programmes. In: *Environmental monitoring*, edited by G.B. Wiersma, 37-62. Boca Raton: CRC Press.
- [13] Dick, J., et al. "A comparison of ecosystem services delivered by 11 long term monitoring sites in the UK environmental change network." *Environmetrics* 22.5 (2011): 639-648.
- [14] Utz, Ryan M., et al. "The National Ecological Observatory Network: An Observatory Poised to Expand Spatiotemporal Scales of Inquiry in Aquatic and Fisheries Science." *Fisheries* 38.1 (2013): 26-35.
- [15] TAYLOR, JR, et al. "THE NATIONAL ECOLOGICAL OBSERVATORY NETWORK: OVERVIEW AND STRATEGIES FOR MANAGING THOUSANDS OF SIMULTANEOUS MEASUREMENTS ACROSS THE CONTINENT." *Data Science Symposium*. 2014.
- [16] Schaaf, Crystal, et al. "Canopy Biomass Lidar (CBL) Acquisitions at NEON and TERN Forest Sites." *AGU Fall Meeting Abstracts*. Vol. 1. 2013.
- [17] Youngentob, K. N., et al. "A survey of long term terrestrial ecology studies in Australia." *Austral Ecology* 38.4 (2013): 365-373.

- [18] Guru, Siddeswara, et al. "Sharing Australia's Nationally Significant Terrestrial Ecosystem Data: A Collaboration between TERN and ANDS." *eScience (eScience)*, 2013 IEEE 9th International Conference on. IEEE, 2013.
- [19] van Dijk, Albert, et al. "Environmental reporting and accounting in Australia: Progress, prospects and research priorities." *Science of The Total Environment* 473 (2014): 338-349.
- [20] Phinn, S. R., R. Christensen, and S. Guru. "Australia's TERN: Advancing Ecosystem Data Management in Australia." *AGU Fall Meeting Abstracts*. Vol. 1. 2013.
- [21] Wieczorek, John, et al. "Darwin Core: An evolving community-developed biodiversity data standard." *PLoS One* 7.1 (2012): e29715.
- [22] Endresen, Dag Terje Filip, and Helmut Knüpffer. "The Darwin Core extension for genebanks opens up new opportunities for sharing genebank datasets." *Biodiversity Informatics* 8.1 (2012).
- [23] Constable, Heather, et al. "VertNet: a new model for biodiversity data sharing." *PLoS biology* 8.2 (2010): e1000309.
- [24] Samy, Gaiji, et al. "Content assessment of the primary biodiversity data published through GBIF network: Status, challenges and potentials." *Biodiversity Informatics* 8.2 (2013).
- [25] Michener, William K., and Matthew B. Jones. "Ecoinformatics: supporting ecology as a data-intensive science." *Trends in ecology & evolution* 27.2 (2012): 85-93.
- [26] Willis, Craig, Jane Greenberg, and Hollie White. "Analysis and synthesis of metadata goals for scientific data." *Journal of the American Society for Information Science and Technology* 63.8 (2012): 1505-1520.
- [27] Michener, William K., et al. "Long term ecological research and information management." *Ecological Informatics* 6.1 (2011): 13-24.
- [28] Skidmore, Andrew K., et al. "Geospatial tools address emerging issues in spatial ecology: a review and commentary on the Special Issue." *International Journal of Geographical Information Science* 25.3 (2011): 337-365.
- [29] Franklin, Erik C., et al. "GeoSymbio: a hybrid, cloud based web

application of global geospatial bioinformatics and ecoinformatics for Symbiodinium - host symbioses." *Molecular ecology resources* 12.2 (2012): 369-373.

[30] Boyd, Doreen S., and Giles M. Foody. "An overview of recent remote sensing and GIS based research in ecological informatics." *Ecological Informatics* 6.1 (2011): 25-36.

[31] Weiss, Anne Elise. "Domain Engineering in GIS: Metadata Standard Interoperability." (2012).

[32] LEI, Zhang, et al. "Metadata Framework of Large-scale GIS Data Center." *Water Resources and Power* 11 (2011): 017.

[33] Steiniger, Stefan, and Andrew JS Hunter. "Free and open source GIS software for building a spatial data infrastructure." *Geospatial free and open source software in the 21st century*. Springer Berlin Heidelberg, 2012. 247-261.

[34] 안성수, et al. "생물다양성데이터 검색포탈 구축." *한국콘텐츠학회 종합학술대회 논문집* 3.1 (2005): 124-128.

[35] 안부영, et al. "생물다양성 데이터교환을 위한 메타데이터 스키마 설계." *한국정보과학회 학술발표논문집* 32.2 (2005): 91-93.

[36] 이슬기, et al. "GIS 를 이용한 산림성 조류의 서식지 예측 모형 및 지도구축." *한국지리정보학회지* 13.1 (2010): 62-73.

[37] 안성수, et al. "KBIF 데이터 노드 구축." *한국정보과학회 학술발표논문집* 31.2II (2004): 280-282.

Acknowledgement:

This work is supported by both Korea Ministry of Environment as "Public Technology Program based on Environmental Policy (Grant No.: 2014000210004)" and the KISTI (Grant No. : K-16-L01-C03).

도서명: 생태정보를 위한 메타데이터 카탈로그 기술 보고서

발행일: 2016년 7월 31일

발행처: 한국과학기술정보연구원
