

임산공학 연구데이터 관리 및 공유를 위한 메타데이터¹

김주섭 전북대학교 문헌정보학과 박사

1. 머리말

최근 데이터 주도형 연구의 중요성이 대두됨에 따라 다양한 방법으로 생산된 방대한 데이터를 통한 과학적 발견이 시도되고 있다. 데이터의 수집, 관리, 유통 등 데이터 개념 및 데이터 관리 방법론에 대한 정립도 이뤄지기 시작하였다. 해외의 경우 연구결과의 활용 및 재사용성 제고를 위하여 데이터 관리를 단일화하고 공유체계를 마련하는 등 데이터 관리체계를 개선하고 있다. 이에 따라 연구개발 과정을 개방화하고 연구데이터를 공유하는 오픈 사이언스 운동이 확산되기 시작했다. 국내에서도 오픈 사이언스의 영향에 따라 데이터 공유를 위한 연구데이터 정의 및 데이터관리계획 도입 등 데이터를 국가 자산으로 인식하고, 공유 및 활용하기 위한 정책을 마련하였다.

2021년 1월 1일부터 시행된 ‘국가연구개발정보

처리기준’에서는 국가연구개발과제를 선정할 때 데이터관리계획에 따른 연구데이터 생산·보존·관리의 충실성 및 공동활용 가능성 등을 검토해야 한다고 명시하고 있다. 연구데이터 생산, 관리, 보존 및 재사용(공동활용)은 데이터 주도 연구에서 필수불가결한 요건이 되었다. 특히, 분야별로 다양하게 생산되는 연구데이터를 관리하기 위해서는 학문 특성에 맞게 연구데이터가 관리되어야 한다. 임산공학 분야에서 생산되는 데이터는 다른 분야보다 훨씬 복잡하여 다양한 형태로 생산되고 있고 그 범위와 데이터의 양이 상당하다. 따라서 오픈 사이언스 시대의 핵심 테마인 데이터 공유, 보존 및 재사용을 위해서는 체계적 데이터 관리가 필요하며 그 핵심은 바로 메타데이터이다.

일반적으로 메타데이터는 ‘데이터에 대한 데이터’ 그리고 ‘어떤 목적을 가지고 만들어진 데이터’를 뜻한다. 메타데이터는 데이터를 검색할 수

¹ 본고는 2020년도 국립산림과학원 연구비지원에 의해 수행된 ‘목재·목재제품 관련 국제 표준에 따른 메타데이터 스키마 설계’ 보고서의 일부를 수정·보완한 것이다

있고, 접근할 수 있고, 상호운용성이 있으며, 재 사용할 수 있도록 해준다. 메타데이터는 해당 학문이나 분야의 특성을 반영하여 개발되어 활용되고 있으며, 실제로 인문학, 사회과학, 물리학, 지구과학, 생물학 등 다양한 학문 분야에서 메타데이터가 개발되었다.

이러한 배경 및 필요성에 따라 본고에서는 임산공학 분야의 연구데이터 공유 및 관리를 위한 메타데이터 요소 도출 및 스키마 개발에 관하여 논의하고자 한다.

2. 메타데이터 표준 동향

국내외를 종합하여 임산공학 분야의 메타데이터는 아직까지 개발된 사례가 없다. 유사한 메타데이터로서 NEFIS가 있지만 이 메타데이터도 산림분야 정보자원을 기술하기 위한 메타데이터이다. 따라서 본 절에서는 임산공학 메타데이터를 개발하기 위해 지리분야의 메타데이터인 ISO 19115, NEFIS 및 INSPIRE, 그리고 연구데이터 출판 및 인용을 위한 메타데이터 스키마인 DataCite를 살펴보고자 한다.

2.1 ISO 19115 지리정보 메타데이터

ISO 19115:Geographic information — Metadata(이하 ISO 19115)는 지리 정보 및 연관 서비스를 기술하기 위한 국제 메타데이터 표준이다. 디지털 지리 데이터의 배포, 공간참조, 공간적 및 시간적 스키마, 데이터 품질, 범위, 식별에 대한 정보를 제공한다. ISO 산하 지리 정보/지질학 기술 위원회인 ISO/TC211은 2003년 ISO 19115:2003 Geographic information — Metadata 초판을 발표하였다. 이후 ISO 19115 지리정보 메타데이터는 세 파트로 분할되었다.

분할된 ISO 19115는 파트 1 기본원칙, 파트 2 수집 및 처리를 위한 확장, 파트 3 XML 스키마 구현을 위한 기본 개념으로 구성된다. 세 파트로 구성된 ISO 19115에서 ISO 19115-1 기본원칙은 지리정보 메타데이터의 개념적 모델을 제시하고 있으며, ISO 19115-3은 XML을 통한 실질적 구현에 대한 내용을 담고 있다[3].

ISO 19115의 특징은 다음과 같다. 첫째, 지리 정보에 특화된 메타데이터로 지리적 규모를 가지는 정보자원을 기술할 수 있다. 둘째, 범용 메타데이터의 성격을 지녀 지리 정보 메타데이터 표준이지만 지리 데이터뿐만 아니라 지리적 특징을 포함하지 않는 비지리 데이터를 비롯한 모든 유형의 정보자원을 기술할 수 있다. 셋째, 지리정보 표준 ISO 19100 시리즈의 일부로서 특정 측면의 정보를 확장하여 지리정보 메타데이터를 기술할 수 있는 연관 표준이 존재한다. 넷째, 해당 표준을 기반으로 확장한 다양한 메타데이터 어플리케이션 프로파일이 존재한다.

2.2 NEFIS 메타데이터

NEFIS(Network for a European Forest Information Service) 메타데이터 스키마는 유럽 내 산림 분야 관련 정보자원을 기술하기 위해 개발된 메타데이터이다. 2003년부터 2005년까지 유럽산림연구소(European Forest Institute)는 유럽위원회의 Quality of Life and Management of Living Resources Programme의 동반 조치로 NEFIS 프로젝트를 수행하였다. NEFIS 프로젝트는 유럽 내 다양한 유형의 산림 정보의 처리 및 유통을 제고하기 위해 네트워크를 구축하는 프로젝트로서, 산림 정보 시스템 구축, 통일된 표준 및 절차 개발, 메타데이터 표준과 통제어휘/키워드 리스트와 같은

상이한 인터넷 기반 산림 정보 서비스 간 의사소통을 위한 통일된 운영 방법 개발 등을 목표로 하였다. 이 과정에서 산림 분야 정보자원을 식별하고 목록 작성을 위해 NEFIS 메타데이터 스키마가 개발되었다[4].

NEFIS 메타데이터 스키마는 NEFIS 프로젝트의 선행 프로젝트인 EFIS 프로젝트에서 개발된 메타데이터 초안과 더블린 코어 그리고 ISO 19115를 기반으로 설계되었다. 전반적으로 더블린 코어에 중점을 두어 개발되었으므로 더블린 코어의 어플리케이션 프로파일의 성격을 지닌다고 볼 수 있다.

2.3 INSPIRE

2007년 유럽에서 공간 데이터 인프라 구축을 위해 INSPIRE (Infrastructure for Spatial Information in the European Community, 이하 INSPIRE) 지침이 발효되었다. 이 지침은 환경 관련 34개 주제의 공간 데이터를 중심으로 공공기관 간 환경 공간정보 공유 및 공간정보에 대한 공공 이용 제고, 정책 수립 지원을 목적으로 하고 있다.² INSPIRE 메타데이터 규약(regulation)은 2008년 처음 제정되었으며, 2014년 INSPIRE 메타데이터 요소의 정의에 따른 규약 Commission Regulation (EU) No 1311/2014으로 개정되었다. 지침에 따라 EU 회원국은 공간 데이터셋 및 서비스의 메타데이터를 작성하고 최신 상태를 유지해야 한다. INSPIRE 메타데이터 규약은 법적 구속력을 갖지만 메타데이터 요소만 제시하고 있다. INSPIRE는 실질적인 INSPIRE 메타데이터 규약의 구현을 지원하

기 위해 기술 가이드라인(Technical Guideline)을 제공하고 있으며, 각국은 이 가이드라인을 참고하여 지역적 요구를 반영하여 필요에 따라 확장하여 사용할 수 있다[2].

본고는 INSPIRE가 제공하는 ISO 19139 기반 기술 가이드라인을 참고하였다. ISO 19139 기반 기술 가이드라인은 공간정보 데이터 서비스와 공간정보 데이터셋을 위한 메타데이터와 관련된 요구사항의 구현과 관련되어 있으며, INSPIRE 서비스의 상호운용성(interoperability)을 최대화하기 위해 설계되었다.

2.4 DataCite

DataCite 메타데이터 스키마는 연구데이터 및 연구 산출물을 위한 영구 식별자(DOI)를 제공하는 국제 비영리 조직 DataCite 컨소시엄 산하 메타데이터 워킹 그룹에서 개발한 메타데이터이다. DataCite 컨소시엄은 데이터의 인용 및 공유를 통해 국제 학술 활동을 장려하고 있으며, 영구식별자를 통해 연구 산출물 검색 및 메타데이터 이용을 지원한다. 이러한 맥락에서 DataCite 컨소시엄의 메타데이터 워킹 그룹은 데이터의 인용과 검색, 편리한 접근을 지원하기 위해 DataCite 메타데이터 스키마를 개발하였다.

DataCite 메타데이터 스키마는 명확하고 일관되게 자원을 식별하기 위해 선정한 핵심 메타데이터 속성들로 이루어져 있으며, 데이터의 탐색과 인용을 지원한다. 식별 대상은 모든 종류의 자원을 포괄하지만 일반적으로 데이터셋(dataset)이다. DataCite 메타데이터 스키마에서는 데이터셋이란 용어를 주로 이루어

² INSPIRE 지침 아래 메타데이터, 데이터 사양, 네트워크 서비스, 데이터 및 서비스 공유, 공간 데이터 서비스, 모니터링 및 보고 등 6개의 의무적인 실행 규칙(implementing rules)이 존재한다.

〈표 1〉 임산공학 분야 메타데이터 요소 도출을 위한 매핑 작업

DataCite	ISO 19115	NEFIS	INSPIRE
Identifier		Identifier	Unique resource identifier Coupled resource
Title	Dataset title	Title	Resource title
Creator	Dataset responsible party	Creator	
Publisher		Publisher	
Publication Year			Date of publication
Date	Dataset reference date	Date Coverage	• Date of creation • Date of last revision
Language	• Dataset language • Metadata character set	Language	Resource language
Description	Abstract describing the dataset	Description Relation	Resource abstract
Subject	Dataset topic category	Subject	• Topic category • Spatial data service type
Contributor		Contributor	
Rights		Rights	
Resource Type		Type	Resource type
AlternateIdentifier			
RelatedIdentifier			
Funding Reference			
Format		Format	
Version			
Size			
Geographic Location	• Geographic location of the dataset • Additional extent information for the dataset • Spatial resolution of the dataset • Reference system • Spatial representation type	Reference system Coverage	• Geographic bounding box • Spatial resolution • Geographic bounding box • Spatial resolution
	Lineage	Source	Lineage
	Distribution format		
	On-line resource		
	• Metadata file identifier • Metadata language • Metadata character set • Metadata point of contact • Metadata date stamp • Metadata standard name • Metadata standard version		• Metadata point of contact • Metadata date • Metadata language
			Keyword value
			Originating controlled vocabulary
		Audience	
		Access Rights	• Limitations on public access • Conditions applying to access and use
			• Responsible party • Responsible party role
			Resource locator
			Temporal extent
			Specification
			Degree

진 데이터로 한정하지 않고 연구 객체(research object)를 포함하는 광의의 의미로 사용하고 있다. DataCite 메타데이터 스키마는 데이터의 탐색 및 인용을 지원하며, 특정 영역에 국한되지 않고 다양한 분야의 연구데이터를 다룰 수 있는 일반적인 성격을 지닌다. 유럽의 OpenAIRE (Open Access Infrastructure for Research in Europe)는 데이터 아카이브 내 데이터셋을 기술하기 위해 DataCite 메타데이터 스키마를 채택하고 있다. 최근 DataCite 메타데이터 워킹그룹은 더블린 코어 메타데이터 이니셔티브(DCMI, Dublin Core Metadata Initiative) 과학 및 메타데이터 공동체(SAM, Science and Metadata Community)와 협력, 스키마의 개방성과 미래의 확장성 지원 그리고 더블린 코어와의 상호운용성 유지를 위해 DataCite 더블린 코어 어플리케이션 프로파일(DC2AP)를 개발하였다[1].

3. 메타데이터 요소 도출

본 장에서는 2장에서 논의한 ISO 19115와 NEFIS, INSPIRE, DataCite를 매핑하여 임산공학 분야의 메타데이터 요소를 도출하였다. 다음의 <표 1>은 임산공학 메타데이터 요소 도출을 위해 매핑한 내용을 정리한 것이다. 매핑 기준은 연구데이터 출판 및 인용을 위한 메타데이터 스키마인 DataCite로 설정하였다.

매핑한 결과, DataCite에서 19개, NEFIS에서 2개 그리고 INSPIRE에서는 11개의 요소가 도출되었다. 다음은 각 메타데이터에서 도출한 요소를 정리한 것이다.

- **DataCite** : Identifier, Title, Creator, Publisher, Publication Year, Date, Language, Description, Subject,

Contributor, Rights, Resource Type, AlternateIdentifier, RelatedIdentifier, Funding Reference, Format, Version, Size, Geographic Location

- **NEFIS** : Audience, Access rights
- **ISO 19115** : Metadata on metadata, Keyword value, Originating controlled vocabulary, Responsible party, Limitations on public access, Specification, Degree, Temporal extent, Resource Locator

4. 임산공학 분야 메타데이터 스키마

본 장에서는 3장에서 도출한 메타데이터 요소를 중심으로 임산공학 분야 메타데이터 스키마를 설계하였다. 각 요소의 우선순위의 상세 내용은 다음과 같다.

- **필수(M)** 속성: 반드시 제공
- **권장(R)** 속성: 선택 사항이지만 상호운용성을 위해 강력히 권장됨
- **선택(O)** 속성: 선택 사항이며 더 자세한 설명을 제공함

우선순위를 적용한 결과 필수요소 6개, 권고요소 17개 그리고 선택요소 7개로 구성되었다. 다음의 <표 2>는 설계된 임산공학 분야 메타데이터 스키마를 정리한 것이다.

5. 맺음말

본 연구는 임산공학 분야의 연구데이터 공유 및 관리를 위한 메타데이터 요소 도출 및 스키마 개발이 목적이다. 목적을 달성하기 위해 관련된 메타데이터인 ISO 19115, NEFIS, INSPIRE 그리고 DataCite 등에 대한 메타데이터를 분석하였다. 또한 임산공학 분야 메타데이터 요소를 도출하기 위하여 각 메타데이터를 매핑하였다. 다음은 본 연구결과인 우선순위를 적용한 메타데이터 요소를 나타낸 것이다.

<표 2> 임산공학 메타데이터 요소


ID	요소명(영/한)		정의	우선순위
1	Title	제목	데이터셋의 이름 또는 명칭	M
2	Identifier	식별자	데이터셋의 식별자 또는 식별값	M
3	Creator	저자	데이터셋을 생성하거나 묶은 사람, 조직 등	M
4	Publication Year	출판년도	데이터가 공개되었거나 공개될 연도	M
5	Publisher	출판사	데이터셋을 웹에 공개하거나 출판하는 주체	M
6	Resource Type	자원 유형	데이터셋의 유형	M
7	Related Identifier	관련 자원 식별자	관련 자원의 식별자	R
8	Contributor	기여자	자원의 내용에 기여한 책임이 있는 사람 또는 기관과 같은 개체	R
9	Date	관련 날짜	업무와 관련된 다른 날짜	R
10	Language	언어	메타데이터(제목, 설명 등) 또는 데이터셋 배포 시 값에 쓰이는 자연어	R
11	Description	추가 설명	데이터셋의 내용에 대한 설명	R
12	Subject	주제	자원을 설명하는 주제, 키워드, 분류 코드 또는 주요 문구	R
13	Keyword value	키워드값	데이터셋의 주제를 설명하는 키워드/태그	R
14	Originating controlled vocabulary	통제어휘	유사한 공신력 있는 키워드 원천(source) 또는 공식적으로 등록된 시소러스의 이름	R
15	Audience	이용 대상	데이터에 대한 메타데이터 작성 시 해당 정보 자원의 대상	R
16	Access rights	접근 권한	데이터셋의 접근 제한 또는 이용 제한 유형	R
17	Rights	권한	데이터셋의 저작권 정보	R
18	GeoLocation	공간정보	데이터가 수집되었거나 데이터가 초점을 둔 공간 지역 또는 명명된 장소	R
19	Size	크기	데이터셋의 크기 또는 기간	R
20	Format	유형(포맷)	자원의 기술적 포맷(구성 방식)	R
21	Version	버전	데이터셋의 버전 번호	R
22	Funding Reference	펀딩 참조 정보	수록된 자원의 재정 지원(펀딩)에 대한 정보	R
23	Metadata on metadata	메타데이터의 메타데이터	메타데이터의 메타데이터	R
24	Alternate Identifier	대체 식별자	해당 엔티티의 추가적인 식별자	O
25	Responsible party	연락처	데이터셋에 대한 문의 사항에 대응할 수 있는 담당자의 연락처 정보	O
26	Limitations on public access	라이선스	데이터셋에 제공될 수 있는 라이선스와 권리 정보	O
27	Specification	규격	기술된 자원이 준수하는 표준과 실행 규칙의 인용, 제목 및 참조일	O
28	Degree	적합성 정도	실 표준과 실행 규칙에 대한 자원의 적합성 정도	O
29	Temporal extent	시간 범위	데이터셋이 다루는 기간	O
30	Resource Locator	온라인 자원	데이터셋을 배포하는 온라인상 위치에 접근할 수 있는 리소스의 URL	O

- **필수요소** : Title, Identifier, Creator, Publication Year, Publisher, Resource Type
- **권고요소** : Related Identifier, Contributor, Date, Language, Description, Subject, Keyword value, Originating controlled vocabulary, Audience, Access rights, Rights, GeoLocation, Size, Format, Version, Funding Reference, Metadata on metadata

- **선택요소** : Alternate Identifier, Responsible party, Limitations on public access, Specification, Degree, Temporal extent, Resource Locator

본 연구를 통해 설계된 메타데이터 스키마는
임산공학 분야 연구데이터의 체계적인 관리와

데이터 공유 및 재사용이 가능해져 연구효율성 및 생산성 증대로 이어질 것으로 기대된다. 또한 임산공학 분야 거점 연구기관인 국립산림과학원

의 위상이 높아지고 영향력이 증대될 것으로 판단된다. 

참고문헌

- [1] DataCite Metadata Working Group. (2019). DataCite Metadata Schema Documentation for the Publication and Citation of Research Data. Version 4.3. DataCite e.V. <https://doi.org/10.14454/7xq3-zf69>
- [2] European Commission. 2008. Commission Regulation (EC) No 1205/2008 of 3 December 2008 implementing Directive 2007/2/EC of the European Parliament and of the Council as regards metadata (Text with EEA relevance) Metadata Implementing Rules Part B. [online] <<https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:32008R1205>>
- [3] ISO. [n.d.]. International Organization for Standardization (1 April 2014). "ISO 19115-1:2014(en)". [online] <<https://www.iso.org/standard/53798.html>>
- [4] NEFIS - Network for a European Forest Information Service. 2019. [online] <<https://www.efi.int/projects/nefis-network-european-forest-information-service>>