

연구 데이터 공유 활성화를 위한 더블린 코어 및 데이터 관리 계획 통합 메타데이터 설계

한성근, 이정철

한국과학기술정보연구원

sghan@kisti.re.kr, jcleee@kisti.re.kr

Design of Integrated Metadata Combining Dublin Core and Data Management Plans to Promote Research Data Sharing

Sunggeun Han, Jeongcheol Lee

Korea Institute of Science and Technology Information

요 약

본 연구는 연구 데이터 공유 활성화를 위해 더블린 코어(Dublin Core) 메타데이터와 데이터 관리 계획(Data Management Plan)을 결합하여 새로운 통합 메타데이터 스키마를 제안하였다. 현재 사용되고 있는 데이터 관리 계획 문서의 경우 대부분 개방형 질문 형태로 구성되어 연구자마다 작성 방식과 내용에 대한 일관성이 부족한 단점이 있다. 이를 해결하기 위해 본 연구에서는 한국과학기술정보연구원(KISTI), 한국연구재단(NRF), Horizon Europe에서 제공하고 있는 데이터 관리 계획 문서를 분석하여 주요 메타데이터 요소를 추출하고, 더블린 코어 메타데이터와 결합하여 66개의 통합 메타데이터 요소를 도출하였다. 본 연구를 통해 연구 데이터 관리 및 공유를 효율적으로 수행할 수 있으며, 데이터 품질 관리 또는 메타데이터 자동 추출 기능 개발에 기여할 것으로 기대된다.

I. 서 론

최근 데이터 기반 연구에 대한 중요성이 강조되면서 연구 과정에서 생산된 연구 데이터에 대한 체계적인 관리의 필요성과 함께 데이터 관리 계획(Data Management Plan, DMP)에 대한 요구가 증가하고 있다. 데이터 관리 계획이란, 연구 수행 전이나 수행 중에 작성하는 공식적인 문서로, 연구 과정 전반에 걸쳐서 연구 데이터에 대한 지침을 제공함으로써 데이터에 대한 활용 가치를 높이 위한 것이다[4]. 그러나, 대부분의 데이터 관리 계획 문서를 보면 개방형 질문으로 구성되어 있다. 이는 연구자들이 데이터에 대해 일관된 형식이나 내용으로 문서를 작성하기에는 한계가 있음을 보여준다[5]. 더블린 코어(Dublin Core) 메타데이터는 디지털 자원의 효율적인 검색과 관리를 위한 국제 표준 메타데이터 요소 집합이다. 기본 메타데이터는 15가지 요소로 구성되며, 기본 요소를 확장하여 다양한 분야를 지원할 수 있도록 유연한 구조를 갖추고 있다[3]. 그러나, 더블린 코어 메타데이터의 확장성에도 불구하고 연구 데이터의 특성을 충분히 반영하지 못하는 한계점도 가지고 있다[2].

따라서, 본 연구에서는 개방형 질문 형태의 데이터 관리 계획 문서의 한계를 극복하고, 더블린 코어 메타데이터의 확장성과 표준화된 구조를 활용하여 연구 데이터의 공유 활성화를 위한 통합 메타데이터 스키마를 제안한다.

II. 통합 메타데이터 설계

2.1. 통합 메타데이터 설계 방법

본 연구에서는 데이터 관리 계획과 더블린 코어 메타데이터의 효과적인 통합을 위해 다음과 같은 과정을 수행한다.

- 데이터 관리 계획 주요 메타데이터 요소 추출

한국과학기술정보연구원, 한국연구재단, Horizon Europe에서 요구하는 데이터 관리 계획 문서를 수집하고 분석하여 공통적으로 요구되는

주요 정보 요소를 파악한다. 그런 다음 각 데이터 관리 계획 문서에 포함된 질문들을 분석하여 메타데이터 요소로 변환할 수 있는 항목들을 추출한다.

- 메타데이터 요소 통합

추출한 데이터 관리 계획의 주요 메타데이터 요소와 더블린 코어의 15개 기본 요소를 비교하여 중복되는 요소와 고유한 요소를 분류한다. 이 과정에서 유사한 의미를 가진 요소들은 통합하고, 연구 데이터의 특성을 반영하는 새로운 요소들을 추가한다.

- 메타데이터 스키마 설계

추출한 메타데이터 요소들에 대해 카테고리, 명칭, 변수명, 속성을 정의한 다음, 메타데이터 요소 간의 관계를 설정한다.

2.2. 통합 메타데이터 정의

더블린 코어의 15개 기본 요소와 한국과학기술정보연구원, 한국연구재단, Horizon Europe에서 제공하고 있는 데이터 관리 계획 문서로부터 추출한 메타데이터 요소를 결합하여 12개의 주요 카테고리와 66개의 통합 메타데이터 요소를 도출하였으며, 그 내용은 표 1과 같다.

〈표 1〉 통합 메타데이터 정의

카테고리	설명	메타데이터
프로젝트 기본 정보	프로젝트의 기본 개요와 배경 정보	프로젝트 제목
		프로젝트 ID
		데이터 생성일
		프로젝트 설명
		프로젝트 기간
		연구 분야
책임자 정보	데이터의 생성과	프로젝트 자금 정보 (예산)
		연구 책임자 정보

	관리에 관련된 모든 책임 주체에 관한 정보	데이터 생성자 정보
		데이터 관리자 정보
		기여자 정보
		발행 기관 정보
데이터 설명	데이터 자체의 내용과 특성을 설명	데이터 컬렉션 제목
		데이터 컬렉션 ID
		버전 정보
		데이터 등록일
		데이터 최종 수정일
		데이터 설명
		주제 키워드
		데이터 유형
		데이터 수집 범위
		원본 데이터 출처
		데이터 활용 분야
		사용 언어
		데이터 수집 방법
		파일 정보
파일 형식		
파일 크기		
파일 수		
품질 관리	데이터의 신뢰성과 정확성을 보장하기 위한 프로세스에 관한 정보	체크섬
		품질 관리 방법
		데이터 검증 결과
		데이터 표준 규격
데이터 형식 및 표준	데이터의 구조와 형식에 관한 표준 및 문서화 정보	데이터 표준 규격
		스키마 파일
		용어 사전
		문서화 계획
데이터 공유 및 접근성	데이터 공유 및 재사용을 위한 접근 조건과 방법에 관한 정보	DOI
		접근 권한
		엠바고 제한 사유
		데이터 접근 방법
		라이선스
		공개 일정
		인용 방법
		관련 자료 링크
		데이터 저장소
		장기 보존 계획
		API 접근 정보
데이터 보존 및 아카이빙	데이터의 장기적인 보존과 관리에 관한 계획 정보	공유 대상
		보존 기간
		보존 책임자
		데이터 폐기 절차
저장 및 백업	데이터의 안전한 저장과 손실 방지를 위한 기술적 방안에 관한 정보	아카이빙 전략
		백업 전략
		백업 주기
		보안 조치
윤리 및 법적 측면	데이터 사용에 관한 윤리적, 법적 고려사항	개인정보 포함 여부
		익명화 처리 여부
		동의 획득 여부
		윤리 심의 정보
		법적 제한 사항
시스템 정보	데이터 관리 시스템 운영에 대한 기술 정보	지적 재산권 정보
		등록자 ID
		등록자 IP
기타 정보	추가적인 정보	메타데이터 변경 이력
		기타 정보

III. 결론

본 연구에서는 연구 데이터의 공유 활성화를 위해 더블린 코어 메타데이터와 데이터 관리 계획 문서의 요소를 결합한 통합 메타데이터 스키마를 제안하였다. 이를 위해 한국과학기술정보연구원, 한국연구재단, Horizon Europe의 데이터 관리 계획 문서를 분석하여 주요 메타데이터 요소를 추출하였으며 더블린 코어의 15개 기본 요소와 결합하여 66개의 메타데이터 요소로 통합하였고, 이를 12개의 카테고리로 분류하였다. 제안한 통합 메타데이터 스키마를 통해 연구자들은 구조화된 메타데이터 형식의 일관된 문서 작성 방식을 사용하여 연구 데이터를 기술할 수 있으며, 다양한 기관의 데이터 관리 계획 문서 작성 요구에 대해서도 효과적으로 대응할 수 있게 된다. 향후에는 제안한 통합 메타데이터를 다양한 연구 분야에 적용할 것이다. 또한, 통합 메타데이터를 활용한 데이터 품질 관리 체계를 구현하고, 메타데이터 자동 추출 기능 개발을 수행할 예정이다.

ACKNOWLEDGMENT

본 논문은 정부(과학기술정보통신부)의 재원으로 한국연구재단의 디지털융합 R&D 플랫폼 연구개발 및 서비스 사업 (No. NRF-2022M3C1A6090416), 국가과학기술연구회 글로벌 TOP 전략연구단 지원사업(No. GTL24031-700)의 지원을 받아 수행되었습니다.

참 고 문 헌

[1] European Commission, "Horizon Europe: Data Management Plan Template," 2021. [Online]. Available: https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/temp-form/report/data-management-plan_he_en.docx. [Accessed: May 9, 2025].

[2] S. Farnel and A. Shiri, "Metadata for research data: Current practices and trends," in Proc. Int. Conf. Dublin Core Metadata Appl., 2014.

[3] DCMI, "DCMI Metadata Terms," Dublin Core Metadata Initiative. [Online]. Available: <https://www.dublincore.org/specifications/dublin-core/dcmi-terms/>. [Accessed: May 9, 2025].

[4] W. K. Michener, "Ten simple rules for creating a good data management plan," PLOS Comput. Biol., vol. 11, no. 10, e1004525, 2015. [Online]. Available: <https://doi.org/10.1371/journal.pcbi.1004525>

[5] N.-M. Pham, H. Moulaison-Sandy, B. W. Bishop, and H. Gunderman, "Data management plans: Implications for automated analyses," Data Sci. J., vol. 22, p. 2, 2023. [Online]. Available: <https://doi.org/10.5334/dsj-2023-002>

[6] 한국과학기술정보연구원, "연구데이터 관리 계획(DMP) 가이드라인." [Online]. Available: https://dataon.kisti.re.kr/data_mgmt_guideline_03.do. [Accessed: May 9, 2025].

[7] 한국연구재단, "연구데이터 및 데이터관리계획 설명회 개최 안내," 2019. [Online]. Available: https://www.nrf.re.kr/cms/board/general/view?menu_no=53&page=1&nts_no=124731. [Accessed: May 9, 2025].