

데이터논문 동료심사를 위한 핵심 개념 분석과 프로세스 모델링*

Analysis and Modeling of Essential Concepts and Process
for Peer-Reviewing Data Paper

안 성 수 (Sungsoo Ahn)**

조 성 남 (Sung-Nam Cho)***

정 영 임 (Youngim Jung)****

〈 목 차 〉

I. 서 론	V. 데이터논문 동료심사 개념 및 프로세스 모델링
II. 관련 연구	VI. 논의 및 시사점
III. 연구 방법	VII. 결론 및 향후 과제
IV. 데이터논문의 구성요소 및 동료심사 프로세스 분석	

요약: 연구데이터를 서술하는 데이터논문은 데이터를 생산한 연구자에게 연구논문과 같은 학술적 성과로 인정받을 수 있게 하고 다른 연구자가 데이터논문에서 공유한 데이터를 활용하여 기존 연구를 재현하거나 새로운 연구를 시작할 수 있도록 도움을 준다. 이러한 장점으로 데이터논문의 동료심사, 출판, 인용은 지속적으로 증가하고 있다. 데이터논문을 출판하고자 하는 학술단체는 연구논문과 다른 데이터논문의 핵심 개념, 데이터논문 심사 절차, 데이터논문 출판과 관계된 정보시스템 등을 이해하고 데이터논문에 포함할 구성요소 정의, 동료심사 절차 수립, 그리고 데이터저장소와 연계 등 세부 사항을 결정하는 것이 필요하다. 하지만, 데이터논문의 동료심사와 관련하여 다루어야 할 정보가 막대하고, 데이터논문 출판에 관한 연구 및 체계적인 정보가 부족하여 국내 학술단체는 데이터논문의 동료심사 업무에 어려움을 겪고 있다. 이러한 문제를 해결하는 데 도움이 될 수 있도록, 본 논문은 국내의 다섯 종 데이터학술지의 데이터논문 양식, 동료심사 사례를 조사·분석하여 데이터논문의 핵심 구성요소를 포함한 데이터논문 동료심사의 핵심 개념을 제안하고 프로세스를 모델링하였다. 데이터논문을 신규로 출판하거나, 기존의 연구논문과 함께 데이터논문을 심사하여 출판하고자 하는 학술단체는 본 논문에서 제시한 데이터논문의 핵심 개념과 동료심사 프로세스를 활용하여 데이터논문의 양식 개발, 체계적인 동료심사 프로세스 정립, 출판 정책 개발 등의 학술지 업무에 신속하고 효과적으로 적용할 수 있을 것으로 기대된다.

주제어: 데이터논문, 구성요소, 핵심 개념, 동료심사, 프로세스, 모델링

ABSTRACT: A data paper describing research data helps credit researchers producing the data while helping other researchers verify previous research and start new research by reusing the data. Publishing a data paper and depositing data to a public data repository are increasing with these benefits. A domestic academic society that plans to publish data papers faces challenges, including timely acquiring tremendous knowledge concerning data paper structures and templates, peer review policy and process, and trustworthy data repositories, as a data paper has different characteristics, unlike a research paper. However, the need for more research and information concerning the critical elements of data paper and the peer-review process makes it difficult to operate for data paper review and publication. To address these issues, we propose essential concepts of the data paper and the data paper peer-review, including the process model of the peer-review with in-depth analysis of five data journals' data paper templates, articles, and other guides worldwide. Academic societies intending to publish or add data papers as a new type of paper may establish policies and define a peer-review process by adopting the proposed conceptual models, effectively streamlining the preparation of data paper publication.

KEYWORDS: Data Paper, Elements, Essential Concepts, Peer Review, Process, Modeling

* 이 연구는 2023년도 경상국립대학교 교내연구비와 국가과학기술연구회 글로벌 공동연구 촉진사업 '연구데이터 동료심사 방안 및 데이터 영향력 분석 연구(Grant No. NST-02)' 과제의 지원을 받아 수행되었음.

** 경상국립대학교 항공우주및소프트웨어공학부 조교수(sungsoo.ahn@gnu.ac.kr / ISNI 0000 0004 7373 7812) (제1저자)

*** 한국과학기술정보연구원 책임연구원(chosn@kisti.re.kr / ISNI 0000 0004 6453 1601) (공동저자)

**** 한국과학기술정보연구원 책임연구원, 과학기술연합대학원대학교 전임교원
(acorn@kisti.re.kr / ISNI 0000 0004 5935 5842) (교신저자)

• 논문접수: 2023년 8월 22일 • 최초심사: 2023년 8월 29일 • 게재확정: 2023년 9월 20일

• 한국도서관·정보학회지, 54(3), 321-346, 2023. <http://dx.doi.org/10.16981/kliss.54.3.202309.321>

※ Copyright © 2023 Korean Library and Information Science Society

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 (<https://creativecommons.org/licenses/by-nc-nd/4.0/>) which permits use, distribution and reproduction in any medium, provided that the article is properly cited, the use is non-commercial and no modifications or adaptations are made.

I. 서 론

데이터논문은 연구 활동을 통해 생성된 연구데이터를 체계적으로 서술하는 논문으로 데이터와 관계된 연구의 재현, 데이터 재사용, 또는 데이터를 기반으로 새로운 연구 활동을 촉진하는 데 도움이 되고 있다(Chavan & Penev, 2011; Lawrence et al., 2011). 데이터논문을 심사하고 출판하는 데이터학술지는 데이터논문을 통해 연구데이터의 공유와 재사용을 증진하고 데이터제공자의 데이터 공유 활동과 노력을 연구성과로 인정하는 것을 목적으로 하고 있다. 오픈액세스 정책을 따르는 많은 데이터학술지는 심사위원과 편집위원의 전문지식을 바탕으로 데이터논문의 심사를 진행하고, 학술지의 목적과 범위에 적합한 데이터논문을 출판하고 있으며, 데이터논문과 데이터의 검색, 접근, 상호운용성, 재사용성을 보장하려고 노력하고 있다(서태설, 2022; Wilkinson et al., 2016).

하지만 학술지 관계자가 데이터논문을 중심으로 하는 데이터학술지를 새롭게 창간하거나 연구 논문과 함께 새로운 논문의 유형으로 데이터논문을 출판하고자 할 때 데이터논문에 포함할 구성 요소의 정의, 데이터논문 동료심사 기준과 절차 수립, 데이터를 저장하고 접근하는 데이터저장소 연계 등 데이터논문의 출판에 참고할 수 있는 연구와 세부 정보가 부족하여 데이터논문의 심사 및 출판 업무에 어려움을 겪고 있다(이현정, 정영임, 황혜경, 2023; 황혜경 외, 2023). 데이터논문의 핵심 구성요소를 구체적으로 이해하고 정의하지 않은 상황에서 기존 연구논문의 구성요소를 준용하여 데이터논문의 심사를 진행할 때 해당 학술지에 제출되는 데이터 논문은 연구논문과 다른 데이터논문의 특성, 예를 들면, 데이터의 생성 배경, 생성 방법과 재사용성 등을 제대로 반영하지 못하여 데이터논문의 형태와 내용 품질에 편차가 클 수 있다. 연구데이터의 중요 특성을 체계적으로 서술하지 않은 데이터논문의 심사는 심사위원에게 학술지에 제출되는 논문에 대해 좋지 않은 인상을 주고 잠재적인 저자들은 고품질의 데이터논문을 이 학술지에 제출하는 것을 꺼릴 수 있으며 심사위원도 이러한 학술지의 데이터논문 심사요청에 적극적으로 응답하지 않을 수 있다. 또한 기존의 연구논문 심사기준과 절차를 따르게 되면 해당 학술지의 심사는 향후 재사용성이 높거나 가치가 높은 연구데이터를 서술하는 데이터논문의 목적과 필요성을 간과하고 데이터논문의 특징을 반영한 심사가 이루어지지 않을 수 있다. 이러한 데이터논문의 심사는 결과적으로 학술지의 권위, 신뢰 및 명성을 점점 저해하는 원인이 될 수 있다(Gomes et al., 2022; Tedersoo et al., 2021).

본 연구는 데이터논문을 심사하고 출판하고자 하는 학술단체가 당면하고 있는 위의 문제를 해결하는 데 도움이 될 수 있도록, 데이터논문의 구성요소, 데이터논문 심사의 핵심개념과 프로세스 모델을 제안한다. 이를 위해 본 논문은 국내외 다섯 종의 데이터학술지 데이터논문 동료심사 및 출판 사례를 조사하고 분석하였다. 이후 연구논문과 차별되는 데이터논문의 핵심 구성요소,

데이터논문 동료심사와 관계된 주요 개념, 동료심사 프로세스를 모델링하였다.

본 논문은 제2장에서 국내외 데이터논문과 관련된 연구를 서술하고, 제3장에서는 본 연구의 주요 연구방법을 설명한다. 제4장에서는 국내외 데이터학술지가 제공하는 데이터논문의 구성요소와 심사 절차를 분석한다. 제5장은 앞장에서 분석한 결과를 바탕으로 데이터논문의 핵심 구성요소, 동료심사를 위한 개념 모델과 프로세스 모델을 제시한다. 제6장은 본 연구를 통해 얻은 국내외 데이터학술지 동료심사의 특징과 시사점에 대해 논의하고, 학술단체가 데이터논문의 심사를 진행할 때 유의점 등을 설명한다. 제7장은 본 논문의 연구 결과를 요약하고 향후 연구를 서술한다.

II. 관련 연구

1. 국내 데이터논문 동료심사 및 출판에 관한 연구

데이터논문 동료심사 및 출판에 관한 국내 연구는 주로 학술지 편집인 또는 관계자를 대상으로 설문조사를 통해 데이터 공유정책, 학술지 발행전략, 데이터논문 출판에 관한 인식조사에 관한 연구와 지구과학 분야의 데이터학술지 출판을 위한 연구가 있다.

Kim, Yi, Huh(2019)는 국내 학술지 편집위원을 대상으로 학술지의 데이터 공유 정책 현황과 계획을 웹 기반 설문조사를 통해 분석하였다. 설문에 응답한 100개의 학술지 중 데이터 공유정책을 가지고 있다고 응답한 13개 학술지는 의학 또는 보건학 분야가 10개로 많았고, 공유정책의 시행 구분은 의무사항(1개), 선택사항(2개), 권고사항(10개)으로 조사되었다. 데이터 공유 정책을 가진 학술지는 재현성 촉진, 과학적 확신에 기여를 주요 이유로 응답하였다. 87개 학술지는 데이터 공유 정책이 없었는데, 이것의 주된 요인은 '저자들이 데이터를 공유하지 않을 것', '데이터 공유로 인한 문제 발생 가능성', '데이터 공유의 지식 부족' 등으로 조사되었다. Kim, Yi, Huh(2019)의 연구는 연구논문에 사용된 데이터의 공유에 관한 정책과 계획 등을 조사했지만, 본 연구는 새로운 논문의 유형으로 데이터논문을 출판할 때 제시해야 할 데이터논문의 구성요소, 연구논문의 동료 심사와 다를 수 있는 심사 단계를 분석하고 제시한 점이 차이가 있다.

정영임 외(2020)는 데이터논문을 심사하는 국내외 데이터학술지 출판 현황을 분석하고, 전문가 표적 집단 인터뷰를 통해 국내 상황을 고려한 생태학 분야의 데이터학술지 발행 전략을 제시하였다. 구체적으로, 이 연구를 통해 생태학 분야 데이터학술지의 발행 방향, 데이터논문 투고 양식에 포함되어야 할 10개의 공통 항목, 데이터논문의 안정적인 수급 방법, 학술지 구성 및 발행주기, 학술지 편집위원회 구성 등의 전략을 제시하였다. 정영임과 조성남(2022)은 또한 국내외 데이터출판 현황을 조사하여 데이터학술지가 발행하는 데이터논문의 수가 2015년 이후 Data in Brief, Scientific Data

등을 중심으로 급격하게 증가하고 있음을 보고하였다. 정영임 외(2020) 연구는 생태학 분야의 데이터논문 출판을 위한 전반적인 방향과 정책을 다루었지만, 본 연구는 과학, 의학, 사회과학, 인문학 등의 데이터논문 출판에 사용될 수 있는 데이터논문의 구성요소, 동료심사 프로세스의 세부 단계를 연구한 점이 차이가 있다. 그리고, 정영임 외(2020) 연구는 생태학 데이터논문의 구성요소로 10개 공통 항목을 제시하였지만, 논문의 다른 구성요소(예: 논문의 제목, 저자 데이터와 같은 메타데이터)와 관계, 배치 등의 자세한 설명이 부족하였다. 본 논문은 데이터논문 작성에 필요한 공통 구성요소를 그룹화하고 배치하였고 학문 분야별 선택 요소 등을 제안하여 전체적으로 논문의 구성요소를 쉽게 이해할 수 있도록 하였다.

황혜경 외(2023)는 국내 과학기술 분야에 종사하는 5개 정부출연연구기관 연구자, GeoAI데이터학회 회원을 대상으로 연구데이터 생산과 보관 장소, 연구데이터 공유, 데이터저널, 및 데이터리포지터리에 대한 인식, 연구데이터의 동료심사, 출판 촉진 요인 및 의향 등을 중심으로 설문조사를 진행하였다. 분석 결과, 데이터저널을 알고 있는 연구자들은 데이터 동료심사가 데이터와 데이터논문의 품질을 향상시킨다는 것에 동의(80% 정도)하였지만, 데이터논문의 출판이 승진심사에서 인정받는다고 응답한 비율은 50% 정도이고, 그렇지 않다고 응답한 비율은 24.6%로 나타났다. 연구기관별로 데이터논문에 대한 인식 차이가 있음을 알 수 있었다. 이현정, 정영임, 황혜경(2023)은 국내 학술지 편집인을 대상으로 데이터 공유 정책과 데이터논문의 인식 정도를 조사하고 분석하였다. 이 조사에서 72개 학술지의 응답 중 20개 학술지가 데이터 공유 정책이 있었으며, 이 정책의 시행은 의무사항(1개), 선택사항(7개), 권고사항(12개)으로 나타났다. 이 분석에서 학술지 편집인은 데이터논문 인식 정도에 관한 문항에서 ‘데이터논문을 (어느 정도) 알고 있다’(42개 학술지)와 ‘전혀 알지 못한다’(28개 학술지)로 분석되었다. 황혜경 외(2023), 이현정, 정영임, 황혜경(2023)의 연구는 설문조사를 통해 연구데이터의 보존, 데이터학술지, 데이터 리포지터리 등 광범위한 내용에 관한 학술지 관계자의 인식 등을 조사하고 분석했지만, 본 연구는 현재 데이터논문을 출판하는 학술지의 논문 양식, 동료심사 가이드, 문헌 조사 등을 통해 데이터논문의 구성요소와 심사 프로세스를 구체적으로 제안하고 모델링한 점이 차이가 있다.

김기동(2023)은 개방형 지구과학 데이터 오픈 플랫폼인 Geo Data의 동료심사 방식과 심사기준을 소개하고 Geo Data의 향후 계획을 소개하였다. 국내 유일한 데이터학술지인 Geo Data의 동료심사 방식은 전문 분야의 편집위원이 심사위원을 초청하며 이중암맹 동료심사 방식으로 제출된 데이터논문을 심사하고 있다. 일반 학술지와 다르게 투고된 데이터논문에 대해 Geo Data에서는 데이터 제공 여부, 획득된 데이터의 정합성, 오류 여부, 중복성, 활용 가능성 등 데이터논문이 기술하고 있는 데이터 관련한 항목을 중심으로 심사함을 밝혔다.

한종규(2023)는 지구과학 분야에서 연구데이터 출판의 중요성을 강조하며, 대표적인 데이터리포지터리인 팬케아(PANGAEA) 및 한국지질자원연구원과 국립생태원에서 구축한 지오빅데이터

오픈플랫폼과 에코뱅크의 현황을 소개하였다. 지구과학 분야 대표 데이터학술지인 Earth System Science Data와 Geoscience Data Journal을 통한 데이터논문 출판 현황을 소개하고, 신뢰성을 갖춘 데이터저장소의 특징, 그리고 연구데이터 공유에 대한 장벽과 다양한 도전 사항을 논의하였다.

2. 국외 데이터논문 동료심사 및 출판에 관한 연구

Lawrence et al.(2011)은 학술분야에서 데이터를 공식적으로 출판할 수 있도록 데이터 동료심사에 필요한 기반 구조, 데이터 출판 모델, 출판된 데이터 인용 방법에 관한 연구를 진행하였다. 데이터 출판은 데이터를 생산하는 연구자에게 학술적 성과를 인정해주고, 과학적인 결론을 분석하고 검증할 수 있는 장치를 제공하고 있어 과학계에서 점점 그 중요성이 높아지는 상황이다. 또한, 데이터 동료심사는 데이터셋의 품질을 보장하는 장치를 제공할 수 있는데, 이러한 동료심사를 진행할 때 기대되는 여러 종류의 품질 관련한 활동을 제안하였다. 특히, 데이터 동료심사 관련해서 심사 검토 기준을 크게 세 가지(데이터 품질, 메타데이터 품질, 일반사항) 그룹으로 나누고, 데이터 심사를 위한 세부 기준을 제안하였다.

Chavan과 Penev(2011)는 1차 생물다양성데이터를 쉽게 접근하거나 발견할 수 없는 원인 중의 하나로 생물다양성데이터를 공유하고 출판하는 연구자에 대한 인센티브 부족 문제를 다루었다. 이러한 인센티브 중 하나가 생물다양성데이터에 관한 상세 메타데이터의 학술출판을 통한 데이터 생산자의 인정이다. 이 문제해결을 위해 생물다양성데이터에 관한 상세 메타데이터를 출판하는 연구자의 기여에 대한 보상으로서 학술적 성과 및 인용할 수 있는 학술논문으로 인정받을 수 있는 생물다양성 데이터논문을 제안하였다. 이 데이터논문과 관련하여 세계생물다양성정보기구(GBIF, Global Biodiversity Information Facility), 펜소프트 출판사와 협업한 결과인 프로젝트의 목적, 워크플로우, 성과 등을 소개하였다.

Candela et al.(2015)은 데이터 출판을 방해하는 장애 요소(인정과 보상 부족, 모호한 인용 행태, 품질 이슈, 일반적인 데이터 공유 문화 부족)에 대해 주목하고, 이러한 문제를 데이터젠틸이 어떻게 해결하고 있는지 조사하였다. 이를 위해, 106개의 데이터젠틸의 데이터셋 서술, 데이터셋 가용성, 인용, 품질, 그리고 오픈액세스의 진흥을 위한 접근방식을 조사하고 분석하였다.

학술출판과 상업출판을 구분하는 중요 요소는 동료심사이다. 2010년 이후 데이터 공유가 지속적으로 증가추세인 상황에서 Carpenter(2017)는 연구데이터 동료심사의 구성요소, 그리고 연구 데이터 동료심사와 관계된 실제 학술지의 심사 사례를 이해하기 위해, 데이터논문의 동료심사와 출판을 수행하는 다양한 분야에 걸친 39개 학술지의 데이터 동료심사 정책을 조사하였다. 이를 위해 데이터 동료심사 정책을 다섯 가지 범주(데이터논문의 적합성 심사, 메타데이터 심사, 데이터 품질 심사, 방법론 심사, 및 기타 심사)로 나누고, 각 범주의 세부 항목에 조사 학술지의 동료심사

정책이 얼마나 포함되는지 정략적으로 분석하였다.

3. 개념 모델링에 관한 연구

컴퓨터과학, 특히 데이터베이스와 소프트웨어공학 분야에서는 현실 세계의 핵심 개념과 프로세스를 이해하고 이러한 개념을 데이터베이스와 소프트웨어 시스템에 적용할 수 있도록 고수준의 개념 모델링에 관한 연구가 진행되었다.

기업 또는 조직의 업무에서 발생하는 데이터를 모델링할 수 있는 다양한 개념 데이터 모델 중에 대표적인 것은 ER(Entity-Relationship) 모델이다. ER 모델은 엔티티(entity), 속성(attributes), 관계성(relationship)의 핵심 개념을 사용하여 데이터를 모델링한다(Elmasri & Navathe, 2015). 소프트웨어공학 분야의 개념 모델링 방법의 하나인 객체지향 모델링은 UML(Unified Modelling Languages)의 클래스 다이어그램에서 현실 세계의 객체를 모델링하고 있다. UML 클래스 모델은 ER 모델과 비슷한 클래스, 속성을 핵심 개념으로 가지고 있고, 객체에 데이터를 처리하는 오퍼레이션(operation)이 추가로 있는 점이 ER 모델과 다르다. 이것 이외에 객체간 관계성으로 일반화, 특별화를 모델링할 수 있다(Booch, Rumbaugh, & Jacobson, 2005). 개념 모델링에 ER 모델과 UML 모델 이외에 기업 또는 사용자의 목표(goals)를 중심으로 핵심개념을 모델링하고 소프트웨어 요구사항을 도출하는 목표지향 모델링 기법이 있다(Ahn et al., 2021; Mylopoulos, Chung, & Yu, 1999).

비즈니스 프로세스를 모델링하는 기법은 여러 가지가 있으나 BPMN(Business Process Model and Notation)을 활용한 모델링 기법이 널리 사용되고 있다. BPMN은 비즈니스 이해당사자가 비즈니스 프로세스를 설계, 관리, 및 실현하는 데 사용되며, 소프트웨어 프로세스 컴포넌트로 변환될 수 있을 만큼 자세하게 모델링할 수 있는 특징이 있다(White & Bock, 2011). 본 연구에서는 데이터논문의 구성요소, 데이터논문의 심사과정 및 심사와 관련된 개념을 명확하게 식별하고 모델링하기 위해 UML, BPMN 모델링 기술을 채택하고 활용하였다. 위 모델링 기법을 통해 본 연구는 학술지 관계자가 데이터논문의 구성요소와 심사 프로세스를 더욱 명확하게 이해하고 다른 이해당사자와 원활한 소통을 지원하려고 의도하였다.

기존의 국내외 연구와 본 연구의 가장 큰 차이점은 데이터학술지를 발행하는 학술지 관계자 관점에서, 데이터논문의 작성 및 동료심사에 필요한 데이터논문의 구성요소를 체계적으로 조사하고 분류하여 핵심 요소를 제시하였고, 데이터논문의 동료심사 과정에 필요한 핵심개념과 프로세스를 소프트웨어공학에서 널리 사용되는 모델링 기법을 활용하여 모델링한 것이다. 이러한 동료심사의 핵심개념과 프로세스는 데이터논문 동료심사 및 출판 관계자의 소통을 쉽게 하고 소프트웨어 개발을 통한 동료심사시스템에서 활용될 수 있는 점이 차별점이라 할 수 있다.

III. 연구 방법

본 논문의 연구 절차와 연구방법을 <그림 1>에 나타내었다. 첫 번째 단계로, 학술지 관계자가 데이터논문의 심사와 출판 업무에 당면하고 있는 문제와 도전 사항에 대해 아래와 같은 연구 질문(Research Question, RQ)을 구성하고, 이 연구 질문에 답할 수 있도록 데이터학술지의 데이터논문 사례, 학술지 데이터논문 제출 및 심사 가이드 등을 분석하고 핵심 개념을 모델링 하였다.

RQ1: 데이터논문을 통해 연구데이터의 공유 및 재사용을 촉진하고자 할 때, 연구데이터를 서술하는 데이터논문의 핵심 구성요소는 무엇인가?

RQ2: 연구데이터를 서술하는 데이터논문의 심사 프로세스는 전통적인 연구논문의 심사 프로세스와 어떻게 다른가?

RQ3: 데이터논문을 심사, 출판하기 위해 핵심적으로 필요한 개념은 무엇이고 이 개념간 관계는 어떠한 의미를 갖는가?

본 연구의 두 번째 단계로, 위의 연구 질문에 대한 답을 찾기 위해, 데이터논문을 출판하는 국외의 대표적인 출판사(Elsevier, Springer Nature, Pensoft, Ubiquity Press)의 학술지 네 종과 국내 학술지 한 종을 선정하였다. 다음으로, 이 학술지들에서 제공하는 데이터논문 저자, 심사위원 가이드, 데이터논문 양식, 데이터논문의 질문 모음집(Frequently Asked Questions, FAQs), 출판된 데이터논문 등을 수집하고 분석하여 개별 학술지의 데이터논문의 제출, 심사, 출판 과정을 분석하였다.

Candela et al.(2015)은 데이터논문을 출판하는 116종의 학술지를 선정하고 데이터논문의 다양한 특징을 조사하였다. 여러 출판사가 학문 분야별 다양한 데이터학술지를 출판하고 있지만, 출판사별로 데이터논문의 심사 및 출판 정책을 각 출판사의 데이터학술지에 공유하고 적용하는 것에 주목하여, 출판사의 대표적인 학술지를 선정하고 데이터논문의 세부 특징을 조사하였다. 이 연구에서 언급된 주요 출판사는 BioMed Central, Chemistry Central, Pensoft Publishers, SpringerOpen, Ubiquity Press이며 각 출판사는 85종, 3종, 7종, 8종, 3종의 데이터학술지를 출판하였다.

본 연구는 Candela et al.(2015)가 데이터학술지를 선정하여 연구한 방법을 일부 채택하고 적용하여 다섯 종의 데이터학술지를 선정하고, 데이터논문과 동료심사를 심도 있게 분석하여 연구 질문에 대한 해결방안을 탐색하였다. 현재 BioMed Central과 SpringerOpen은 Springer Nature 출판사의 자회사라는 점을 고려하여 본 연구에서는 Springer Nature에서 학제간 데이터논문을 출판하는 Scientific Data 학술지를 선정하였다. 또한 Pensoft 출판사의 대표적인 데이터학술지인

Biodiversity Data Journal, Ubiquity 출판사의 Journal of Open Humanities Data를 선정하였다. 그리고, Elsevier 출판사는 Data in Brief 학술지를 통해 최근 많은 데이터논문을 출판하며 관련 시장을 선도하고 있어 본 연구에 이것을 포함하였다. 국내외에서 출판되는 데이터논문에 관한 분석 내용을 상세히 비교할 수 있도록 국내 데이터학술지인 Geo Data도 본 연구의 조사 대상에 포함하였다.

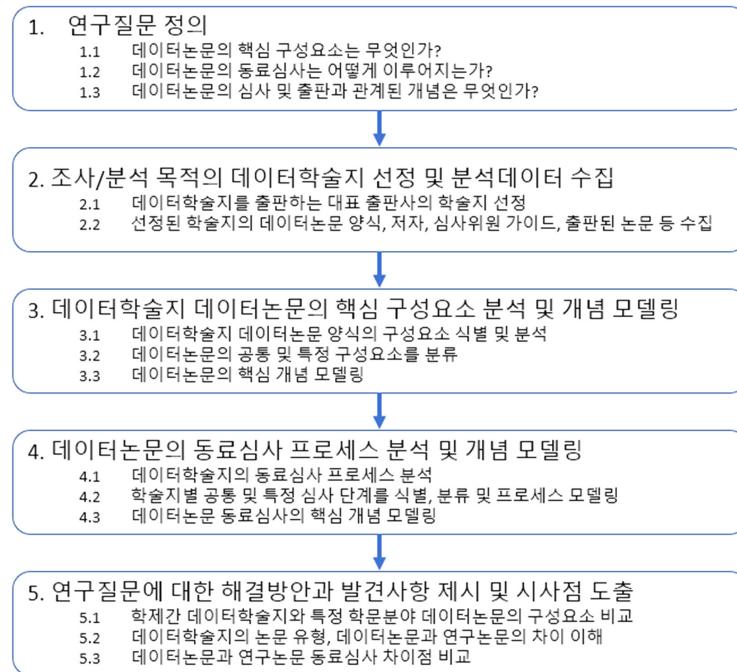
본 연구에서 선정한 국내외 다섯 개 출판사의 대표적인 데이터학술지는 의학, 공학, 농학, 사회 과학을 포함하는 범용적 학문 분야의 학술지(Data in Brief, Scientific Data), 생물다양성 분야의 학술지(Biodiversity Data Journal), 지질, 생태, 항공 분야를 포함하는 지구과학 분야의 학술지(Geo Data), 인문학 분야의 학술지(Journal of Open Humanities Data)로, 다양한 학문 분야에서 채택되어 사용되고 있는 데이터논문의 핵심 구성요소, 심사과정 및 심사와 관련된 핵심 개념을 조사, 분석하고 개념적인 모델링을 위해 사용되었다.

세 번째 단계는 다섯 종의 데이터학술지가 제공하는 데이터논문의 양식을 분석하여 데이터논문의 핵심 구성요소를 이해하기 위해, 데이터논문 양식의 구성요소를 분석하고, 이것을 공통요소와 개별요소로 구분한 뒤 세부 내용을 모델링하였다. 먼저 데이터학술지의 데이터논문 양식을 바탕으로 데이터논문의 핵심 구성요소를 조사하였다. 이후, 다섯 종의 데이터학술지의 데이터논문 양식에서 나타나는 공통 구성요소와 개별 학술지에서만 나타나는 특정 요소를 비교하고 분류하였다. 이후, 데이터논문에 사용될 수 있는 핵심 구성요소를 도출하고 개념 모델링을 통해 이러한 핵심 구성요소를 클래스 다이어그램으로 표현하였다.

네 번째 단계는 다섯 종의 학술지 심사위원 가이드를 기반으로 데이터논문 동료심사 절차를 분석하여 각 학술지 동료심사 과정의 유사점과 차이점을 분류하였다. 이러한 분석 결과를 알기 쉽게 BPMN 프로세스 다이어그램을 활용하여, 데이터논문의 핵심 동료심사 과정을 모델링 하였다. 먼저, 개별 학술지의 데이터논문 동료심사 절차를 조사, 분석하여 학술지별 동료심사의 핵심 개념과 절차를 이해하고 연구논문과 다른 프로세스를 식별하였다. 다음으로 각 동료심사의 유사점과 차이점을 분석하였고, 이후 이러한 분석을 바탕으로 데이터논문의 핵심 프로세스를 모델링하고, 개념 모델링을 통해 동료심사의 핵심 개념간 관계를 UML 클래스 다이어그램으로 표현하였다.

다섯 번째 단계는, 본 연구의 결과와 연구 수행 중 관찰했던 내용을 바탕으로 연구 질문에 대한 해결방안과 발견사항을 분석하여 제시하고 데이터논문 동료심사와 관련한 시사점을 도출하였다.

데이터논문 동료심사를 위한 핵심 개념 분석과 프로세스 모델링



〈그림 1〉 연구 절차

IV. 데이터논문의 구성요소 및 동료심사 프로세스 분석

본 장에서는 앞의 3장에서 언급한 세 가지 연구 질문에 답하기 위한 본 논문의 연구 단계별 연구내용을 상세히 설명한다. 데이터논문의 구성요소를 이해하기 위해 국외 네 종, 국내 한 종의 데이터학술지가 제공하는 데이터논문 양식을 분석하여 세부 구성요소를 조사하였다. 또한 데이터학술지의 데이터논문 심사 프로세스를 조사하여 기존 연구논문 심사 프로세스와 차이점을 알아보고, 데이터논문 심사에 공통으로 사용되는 핵심개념과 프로세스를 분석하였다.

1. 생물다양성데이터저널(Biodiversity Data Journal)¹⁾

생물다양성데이터저널(Biodiversity Data Journal, 이하 BDJ)은 커뮤니티 기반 동료심사를 수행하는 오픈액세스 데이터학술지로 포괄적인 온라인 플랫폼을 제공한다. BDJ는 펜소프트 출판사(Pensoft Publishers)가 2013년부터 데이터논문을 출판하고 있으며 생물다양성 관련한 모든 데이

1) <https://bdj.pensoft.net/about>

터의 출판, 확산, 및 공유를 촉진하고 있다. BDJ는 데이터논문 이외에 오믹스 데이터논문(Omics Data Paper), 연구논문(Research Article), 방법(Methods), R 패키지, 소프트웨어 설명(Software Description) 등 다양한 유형의 논문을 출판하고 있다.

BDJ에 사용되는 데이터논문 양식의 주요 구성요소를 <표 1>에 나타내었다. 1번에서 7번까지는 데이터논문의 메타데이터로 분류할 수 있는데, 논문제목, 저자, 초록, 키워드를 포함한다. 8번에서 17번까지는 데이터 또는 데이터셋을 설명하는 요소로, 데이터 소개, 목적, (수집/생성된 데이터의) 표본 추출 방법, 지리적 범위, 분류학적 범위, 사용 라이선스, 데이터 자원 또는 데이터셋의 정보가 포함된다. 18번부터 20번까지는 저자기여, 참고문헌을 나타내고 있다.

<표 1> BDJ 데이터논문의 구성요소

* 필수사항, 그 외 선택사항

번호	구성요소	번호	구성요소
1	Title*	11	Sampling methods*
2	Authors*	12	Geographic coverage*
3	Abstract* - Background, New Information	13	Taxonomic coverage*
4	Keywords*	14	Collection data
5	Classifications	15	Usage license* - Usage license, IP rights notes
6	Funder	16	Data resources* - Data package title, Resource link, Alternative identifiers - Number of data sets: Data set name, Download URL, Data format
7	Nanpublications	17	Additional information
8	Introduction*	18	Acknowledgements
9	General description* - Purpose	19	Author contributions*
10	Project description	20	References*

BDJ의 데이터논문 심사 프로세스는 데이터논문 제출, 데이터논문 접수, 심사위원 초청, 데이터논문 심사, 데이터논문 판정과 같이 기존 연구논문의 동료심사 프로세스와 거의 유사하지만, BDJ의 데이터논문의 제출, 심사과정에서 기존 연구논문의 동료심사 프로세스와 몇 가지 차이점이 있다. 저자의 관점에서, 연구논문의 동료심사와 다른 점은 첫째, BDJ 데이터논문 원고를 제출하기 전에 관련 데이터셋을 신뢰할 수 있는 데이터저장소, 예를 들면 제노도(Zenodo), 드라이어드(Dryad), 피그쉐어(Figshare), 하버드 데이터버스(Harvard Dataverse) 등에 기탁하고 접근할 수 있게 공개하여야 한다. 둘째, 데이터 기탁 후, 데이터논문의 데이터 자원 항목에 데이터 패키지(데이터셋) 제목, 영구 ID를 통한 접근 정보, 데이터셋에 관한 상세 정보를 서술해야 한다. 셋째, 데이

터논문 원고 제출 때 수행하는 활동으로 FAIR(Findable, Accessible, Interoperable, Reusable) 원칙에 상응하는 데이터셋에 관한 자체 검사를 수행하고 그 결과를 저장하는 과정이 있다. 심사위원 관점에서, 기존 연구논문의 심사 대비 주요 차이점으로 심사위원은 데이터논문에 대한 심사뿐만 아니라, 데이터셋, 그리고 데이터논문과 데이터셋 간의 일관성에 대해 추가로 심사를 진행해야 한다.

BDJ 동료심사방식은 기본적으로 단일암맹(Single blind peer-review)으로 진행된다. 하지만, BDJ는 단일암맹 심사방식 이외에 패널 또는 (회원을 대상으로) 공공 심사위원(Panel or Public Reviewers)을 초청하여 심사를 진행할 수 있는 방식을 지원한다. 이것은 비공식적 논문심사 방식으로 간주되지만, BDJ 회원은 데이터논문에 대해 의견(comments)을 제시할 수 있는 점이 특징이다.

2. 사이언티픽데이터(Scientific Data)²⁾

사이언티픽데이터(Scientific Data, 이하 SciData) 학술지는 데이터논문과 연구논문을 출판하는 동료심사, 오픈액세스 저널이며, 스프링거 네이처(Springer Nature)가 2014년부터 이 저널을 출판하고 있다. SciData는 자연과학, 의학, 공학 및 사회과학의 모든 분야에서 제출되는 데이터논문(Data Descriptors)을 출판하고 있다. 이 데이터논문 이외에 분석(Analysis), 연구(Article), 분석 및 연구(Analysis & Article), 의견(Comments)과 같은 논문 유형이 있다.

〈표 2〉는 SciData 학술지 데이터논문의 구성요소를 보여주고 있다. 1번에서 3번은 제목, 저자, 요약을 나타낸다. 다른 데이터학술지와 비교하여 SciData 데이터논문은 키워드를 서술하는 요소가 없다. 4번에서 9번까지는 데이터셋과 관련한 배경 및 요약, (수집/생성) 방법, 데이터 레코드, 기술 검증, 코드 가용성 등을 서술한다. 10번부터 14번까지는 참고문헌, 사사 정보, 저자 기여사항, 이해 충돌, 추가정보를 작성한다. SciData 데이터논문은 데이터 처리에 사용된 소프트웨어를 명시적으로 서술할 수 있는 ‘코드 가용성’ 항목을 별도로 제공하고 있다.

〈표 2〉 SciData 데이터논문의 구성요소

번호	구성요소	번호	구성요소
1	Title*	8	Usage Notes
2	Authors*	9	Code availability*
3	Abstract*	10	References*
4	Background & Summary*	11	Acknowledgements*
5	Methods*	12	Author contributions*
6	Data Records*	13	Competing interests*
7	Technical Validation*	14	Additional information*

2) <https://www.nature.com/sdata/>

SciData 데이터논문에 관한 동료심사는 단일암맹 심사방식으로 진행되지만, 중요한 차이점은 데이터 품질, 재사용에 중점을 두고 동료심사를 진행한다는 점이다. 저자의 경우, 데이터 리포지터리에 데이터논문이 서술하는 데이터셋을 기탁하지 않으면 데이터논문의 동료심사가 진행되지 않는다. SciData는 학제적 논문을 접수하기 때문에, 논문접수 과정에서 편집팀에서 적극적으로 참여하여 논문을 관련 분야의 편집위원에게 전달하는 과정을 거친다. 심사위원은 기술적 측면에서 데이터의 품질, 데이터논문 구성요소에 관한 서술의 완전성, 데이터파일과 리포지터리 레코드의 무결성을 중점으로 심사를 진행한다.

3. 데이터인브리프(Data in Brief)³⁾

데이터인브리프(Data in Brief, 이하 DIB) 학술지는 다분야, 오픈액세스, 동료심사 학술지로 엘스비어(Elsevier) 출판사가 2014년부터 출판하고 있다. DIB 데이터논문은 이해하기 쉽고 짧은 형태이며 데이터와 관련 연구를 더 쉽게 이해하고 세부 내용을 발견할 수 있게 구성되어 있다. DIB는 연구 협업의 문을 열고 중복 노력을 감소시켜 오픈사이언스에 기여하고, 연구 재현의 향상을 목표로 한다. DIB의 논문 유형은 데이터논문(data articles)이 큰 부분을 차지하며, 지원논문(supporting articles,) 갱신논문(update articles), 관점 논문(perspective articles)이 있다.

DIB의 데이터논문 구성요소를 <표 3>에 나타내었다. 1번에서 4번까지는 논문제목, 저자, 초록, 키워드이다. 5번에서 9번까지는 데이터셋과 관련된 요소이다. 특히, 5번의 명세표는 데이터셋의 요약 정보를 데이터논문의 앞부분에 보여주고 있어, 독자들이 데이터셋을 빠르게 이해하는 데 도움이 된다. 이 명세표는 주제, 세부 주제영역, 데이터 타입, 데이터 획득 방법, 데이터 형식, 데이터 컬렉션 설명, 데이터 소스 위치, 데이터 접근성, 관련 연구논문 항목으로 구성되어 있다. 6번-9번 요소는 데이터 가치, 목적, 데이터 설명, 실험 디자인/재료/방법을 상세히 서술할 수 있다. 10번에서 14번은 윤리선언, 이해총돌, CRedit에 기반한 저자 기여, 참고문헌을 서술할 수 있다.

DIB의 데이터논문 심사방식은 단일암맹 방식이며, 2명 이상의 심사위원이 참여하고 있다. 저자는 데이터논문 제출 전에 데이터논문이 서술하는 데이터를 공공 데이터저장소에 필수로 기탁하여야 심사가 진행된다. 소프트웨어 코드의 경우 깃허브(GitHub)에 기탁한다. 심사위원은 데이터논문이 데이터셋의 획득 방법과 재사용, 데이터 가치 등을 상세히 서술하고 있는지, 그리고 데이터가 공공의 데이터저장소에 기탁되고 접근할 수 있는지를 중심으로 심사한다.

3) <https://www.sciencedirect.com/journal/data-in-brief>

〈표 3〉 DIB 데이터논문의 구성요소

번호	구성요소	번호	구성요소
1	Title*	9	Experimental Design, Materials and Methods*
2	Authors*	10	Ethics Statements*
3	Abstract*	11	Competing Interest*
4	Keywords*	12	CRedit Author Statement*
5	Specification Table* - Subject*, Specific subject area*, Type of data*, How the data were acquired*, Data format*, Description of data collection*, Data source location*, Data accessibility*, Related research article*	13	Acknowledgments
6	Value of Data*	14	References*
7	Objective*		
8	Data Description*		

4. 오픈인문학데이터저널(Journal of Open Humanities Data)⁴⁾

오픈인문학데이터저널(Journal of Open Humanities Data, 이하 JOHD)은 재사용성이 높은 인문학 데이터 또는 기법을 서술한 데이터논문을 동료심사하고 있으며, 유비쿼터 출판사(Ubiquity Press)가 2015년부터 JOHD의 논문을 출판하고 있다. 과거 수 세기 동안 인문학 데이터는 고문서, 인쇄된 책에 기록됐는데, 이러한 데이터의 사용은 주관적, 개인적 해석에 의존하고 문맥이 중요한 역할을 하는 특징이 있다. JOHD의 논문 유형은 데이터를 서술하는 단편 데이터논문과 데이터와 관계된 정규 연구논문으로 나뉜다.

JOHD 데이터논문의 구성항목을 〈표 4〉에 정리하였다. 1번에서 4번까지의 구성요소는 논문제목, 저자, 초록, 키워드를 나타낸다. 5번에서 8번까지는 데이터셋의 개요(데이터저장소 링크, 문맥), 방법(세부 단계, 샘플링 전략, 품질관리), 데이터 상세설명, 재사용 가능성을 서술한다. 6번 ‘방법’ 요소는 독자가 데이터셋이 어떻게 생성되었는지 이해하고 데이터를 재생성할 수 있도록 서술해야 한다. 7번 ‘데이터 상세설명’ 항목은 데이터 이름, 형식과 버전, 생성 날짜, 생성자, 언어, 라이선스, 저장소 이름, 출판날짜를 포함한다. 8번 ‘재사용 가능성’ 항목은 데이터 재사용을 위한 구체적이고 유용한 설명을 제공해야 한다. 9번에서 13번은 사사, 기금정보, 이해충돌, 저자기여, 참고문헌 항목에 대해 서술할 수 있다.

JOHD의 동료심사는 단일암맹 방식으로 진행되며, 두 명 이상의 심사위원이 JOHD가 제공하는 심사정책과 심사양식을 사용하여 심사를 진행한다. 동료심사 프로세스 중 저자의 논문 제출 프로세스에서 주의할 점은 저자는 데이터셋을 데이터 리포지터리에 기탁한 후 얻게 되는 데이터셋의 DOI(Digital Object Identifier)를 데이터논문에 추가 후 데이터논문을 제출해야 한다. 이후, 데이터

4) <https://openhumanitiesdata.metajnl.com/about>

논문의 게재 승인이 이루어지면, 데이터논문 DOI를 데이터 리포지터리에 추가해야 한다.

〈표 4〉 JOHD 데이터논문의 구성요소

번호	구성요소	번호	구성요소
1	Title*	8	Reuse Potential
2	Authors*	9	Acknowledgments
3	Abstract*	10	Funding information
4	Keywords*	11	Competing interests
5	Overview* - Repository location, Context	12	Author Contributions
6	Method* - Steps, Sampling strategy, Quality control	13	References*
7	Data Descriptions* - Object name, Format names and versions, Creation dates Dataset creators, Language, License, Repository name, Publication date		

5. 지오데이터(Geo Data)⁵⁾

지오데이터(Geo Data) 학술지는 지리, 생태, 해양 및 극지, 원격 센싱 분야의 연구데이터 공유, 출판 및 연구를 위한 국내 최초 데이터학술지로 2019년 12월 창간되었다. Geo Data는 지구과학 분야의 데이터논문을 동료심사하고 오픈액세스로 출판하고 있다. 이 학술지의 데이터논문은 데이터셋을 활용하여 심화 및 응용 연구를 촉진할 수 있도록, 과학적 방법으로 획득되거나 수집된 데이터셋에 관한 설명을 제공하는 것을 목표로 하고 있다. Geo Data 학술지의 논문 유형은 데이터논문(Data paper), 리뷰논문(Review paper), 단편논문(Shorts)이 있다.

Geo Data는 누리집에서 데이터논문의 구성항목을 안내하고는 있지만 별도로 데이터논문 작성 을 위한 양식은 제공하지 않고 있다 Geo Data 누리집의 데이터논문 작성 안내 사항과 김기동 (2023)의 자료를 기반으로 분석한 데이터논문의 주요 구성항목을 〈표 5〉와 같이 나타낼 수 있다. 1번에서 4번까지의 구성요소는 논문제목, 저자, 초록, 키워드이다. 논문제목, 저자, 초록은 국문과 영문으로 작성하고 키워드는 영문으로만 작성하고 있다. 5번 본문 항목의 경우, 세부 구성요소는 명시하고 있지 않지만, 출판된 일부 데이터논문 사례를 조사한 결과, 암묵적으로 서론, 본론(소개, 방법, 시간적/공간적 범위, 결과 및 고찰 등), 결론으로 구성되고 있다. 6번에서 9번은 이해충돌, 기금정보, 데이터가용성, 참고문헌을 서술한다. 10번 항목은 표 형태로 데이터셋에 대한 메타데이터(제목, DOI 이름, 범주, 초록, 시간 범위, 공간 범위, 생성자, 라이선스, 프로젝트, 도구)를 서술하고 있으며, 이 표는 데이터논문의 끝에 배치된다.

Geo Data는 앞에서 서술된 네 종의 국외 학술지가 채택한 단일암맹 방식이 아닌 이중암맹

5) <https://geodata.kr/about/>

데이터논문 동료심사를 위한 핵심 개념 분석과 프로세스 모델링

(Double-blind) 동료심사 방식을 활용하고 있다. 저자는 데이터논문을 제출하기 전에 데이터를 공공 데이터저장소에 기탁하여야 한다. 편집위원장은 데이터논문을 접수한 후 특정 분야의 편집 위원에게 논문을 배정하고, 담당 편집위원은 두 명의 심사위원을 초청하여 심사를 진행한다. 심사 위원은 데이터논문과 데이터가 기술적으로 정확한지, 데이터가 데이터저장소에 기탁되어 접근 가능한지, 데이터가 용성 문장의 서술 등을 심사하고 심사 의견을 제출한다. 편집위원은 이러한 심사 결과를 참고하여 논문의 계재를 판정한다.

〈표 5〉 Geo Data 데이터논문의 구성요소

번호	구성요소	번호	구성요소
1	Title* (제목)	6	Conflict of Interest*
2	Authors* (저자)	7	Funding Information*
3	Abstract* (초록)	8	Data Availability Statement*
4	Keywords*	9	References*
5	Main Body* (본문) - 주요 구성요소는 서론, 본론 (개요, 방법, 시간적/공간적 범위, 결과, 고찰), 결론임	10	Metadata for Dataset* - Title*, DOI name*, Category*, Abstract*, Temporal Coverage*, Spatial Coverage*, Personnel*, CC License* - Project, Instrument

V. 데이터논문 동료심사 개념 및 프로세스 모델링

본 장에서는 앞 장에서 조사, 분석한 개별 학술지 데이터논문의 구성요소를 상위 수준에서 비슷한 개념의 구성요소를 모아 목적별로 분류하고, 다섯 종의 데이터학술지 데이터논문의 세부 구성요소를 이 분류항목에 매핑하였다. 이후, 분류된 데이터논문의 구성요소를 기반으로 객체지향 모델링 방법 중 클래스 모델링 기법을 적용하여 데이터논문의 핵심 개념 모델을 설계하였다.

1. 데이터논문의 구성요소 분류 및 모델링

다섯 종 학술지의 데이터 논문 구성요소를 분석한 후 본 논문은 〈표 6〉과 같이, 데이터논문을 메타데이터, 데이터셋을 서술하는 요소, 참고문헌, 그리고 사사, 윤리, 이해충돌 등 선언 부분과 같이, 상위 수준에서, 네 가지로 그룹으로 분류하고, 이 그룹에 해당되는 논문의 구성요소를 배치하였다. 〈표 6〉에서 식별된 데이터논문의 구성요소는 연구 질문 1에서 제기한 학술지가 관계된 학문분야의 연구데이터 공유 및 재사용, 기존 연구의 재현에 도움이 될 수 있다.

첫 번째 분류항목인 데이터논문의 메타데이터는 논문제목, 저자 정보, 초록, 키워드를 포함한다. 두 번째는 데이터셋을 서술하는 부분으로, 학술분야에 독립적인 요소와 의존적인 요소로 나눌 수 있다. 학술분야에 독립적인 요소는 소개, 목적, (데이터셋을 생성 단계, 관찰, 추출, 품질관리) 방법, (데이터셋 이름, 형식, 버전, 접근 링크 등을 포함한) 데이터셋 정보, 라이선스, (재)사용 가능성성이 포함할 수 있다. 학술분야에 의존적인 요소는 생물다양성데이터 분야의 경우, 예를 들면, 분류학적 범위, 지리적 범위 데이터가 포함될 수 있고, 인문학 분야의 경우, 데이터를 작성한 ‘언어’ 요소가 될 수 있다. 세 번째는 참고문헌으로, 한 개 이상의 인용할 수 있는 연구논문, 데이터논문을 포함한다. 네 번째는 선언 부분으로 사사 (또는 감사), 윤리 선언, 이해충돌, 저자기여, 기금정보 등을 포함한다. 이러한 상위 수준의 분류체계를 바탕으로, 앞의 다섯 종 학술지 데이터논문 구성요소의 핵심 요소를 선별하고 제안된 분류체계에 매핑하였다.

〈표 6〉의 데이터논문 분류체계와 세부 요소를 바탕으로, 데이터논문의 핵심 구성요소를 〈그림 2〉와 같이 UML(Unified Modelling Language) 클래스 다이어그램을 활용하여 개념적으로 모델링할 수 있다. 데이터논문은 네 가지 클래스로 구분할 수 있고, 각 클래스는 세부 핵심 구성요소를 포함하고, 이 구성요소는 동료심사 및 출판 목적에 따라 더 세분화한 클래스로 정제될 수 있다. 예를 들면, 제시된 클래스 다이어그램에서, 참고문헌은 한 개 이상의 연구논문 또는 데이터논문을 포함할 수 있는데, 이러한 부분은 개념간 관계의 복잡성을 피하고자 본 논문에서는 표시하지 않았다.

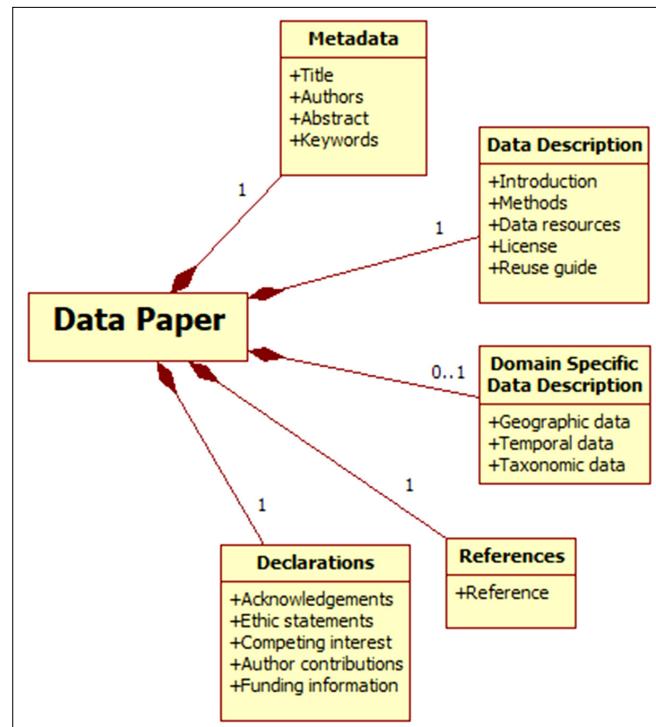
〈표 6〉 본 연구에서 분석한 데이터논문의 핵심 구성요소

* 필수사항, ✓ 필수 적용, △ 암묵적 적용, 그 외 선택사항

번호	구분	구성요소	BDJ	SciData	DIB	JOHD	Geo Data
1	1. Metadata of Data Paper	Title*	✓	✓	✓	✓	✓
2		Authors*	✓	✓	✓	✓	✓
3		Abstract*	✓	✓	✓	✓	✓
4		Keywords*	✓		✓	✓	✓
5	2. Data Descriptions (Domain independent elements)	Introduction*	✓	✓	△	✓	△
6		Methods*	✓	✓	✓	✓	△
7		Data resources*	✓	✓	✓	✓	✓
8		License*	✓		✓	✓	✓
9		Reuse guide*	△	✓	✓	✓	△
10		Taxonomic coverage	✓				
11	Data Descriptions of Dataset (Domain specific elements)	Geographic coverage	✓				✓
12		Language				✓	
13		References*	✓	✓	✓	✓	✓
14	4. Declarations	Acknowledgements	✓	△	✓	✓	△
15		Ethic statements			✓		
16		Competing interest		✓	✓	✓	✓
17		Author contributions	✓	✓	✓	✓	
18		Funding information				✓	✓

데이터논문 동료심사를 위한 핵심 개념 분석과 프로세스 모델링

이러한 데이터논문의 개념 모델은 데이터논문의 핵심 구성요소를 가시화하고 이해에 도움을 주며, 학술지 관계자간 의사소통을 원활하게 하는 것을 지원한다. 또한, 이러한 개념 모델은 추후 온라인 동료심사시스템의 요구사항 개발 및 정제, 서비스에 기술적 기반으로 활용될 수 있다.



〈그림 2〉 데이터논문 핵심 구성요소 모델링

2. 데이터논문 동료심사를 위한 프로세스 모델링

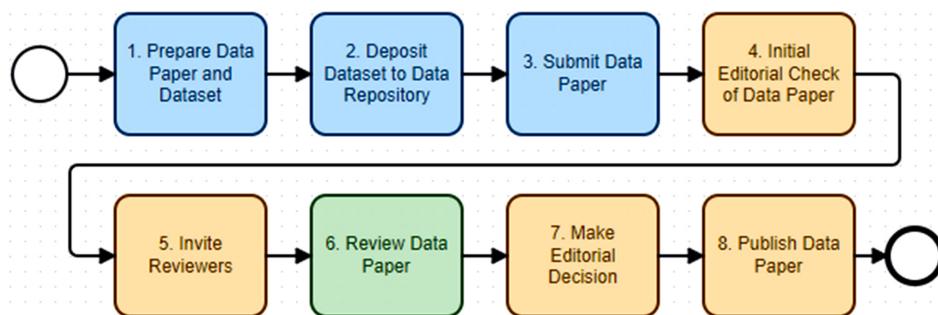
본 연구는 데이터논문의 동료심사 프로세스에 적용되는 공통 단계를 도출하기 위해 제1 수준의 프로세스를 여덟 단계로 분류하고, 다섯 종 학술지가 적용하는 동료심사의 세부 단계를 이 분류체계에 매핑하였다. 연구논문 동료심사와 비교하여 가장 많은 차이가 있는 저자, 심사위원 중심으로 프로세스를 모델링하였다. 이 프로세스는 연구 질문 2에서 제기한 연구데이터를 서술하는 데이터논문의 심사 프로세스는 연구논문의 심사 프로세스와 무엇이 다른가에 대한 설명을 제공한다.

〈표 7〉 본 연구에서 분석한 데이터논문 동료심사 단계

✓ 필수 적용, △ 암묵적 적용, 그 외 미적용

동료심사 단계	동료심사 세부단계	BDJ	SciData	DIB	JOHD	Geo Data
1. 데이터논문 및 데이터셋 준비	- 데이터논문 양식을 활용한 데이터논문 준비	✓	✓	✓	✓	△
	- 데이터셋 안내서를 참고하여 데이터셋 생성/수집	✓	✓	✓	✓	✓
2. 데이터셋을 데이터저장소에 기탁	- 데이터셋을 데이터저장소에 기탁	✓	✓	✓	✓	✓
	- 데이터논문 내용에 FAIR와 관련한 항목 체크	✓	△	△	△	△
3. 데이터논문 제출	- 데이터논문의 메타데이터 입력 및 논문 파일 첨부	✓	✓	✓	✓	✓
4. 편집위원회 데이터논문 초기 체크	- 편집위원장(또는 편집팀)은 심사할 데이터논문 선별	✓	✓	✓	✓	✓
5. 심사위원초청	- 도메인 전문가 초청	✓	✓	✓	✓	✓
	- 폐널 또는 일반 전문가 초청	✓	△	△	△	△
6. 데이터논문 동료심사	- 메타데이터 품질 심사	✓	✓	✓	✓	✓
	- 데이터셋 품질 심사	✓		✓	✓	✓
	- 메타데이터와 데이터셋간 관계에 대한 품질 심사	✓	✓	△	△	△
7. 데이터논문 게재 판정	- 논문 게재 판정	✓	✓	✓	✓	✓
8. 데이터논문 출판	- 논문 출판	✓	✓	✓	✓	✓

〈표 7〉에 기반하여, 제1 수준의 데이터논문의 동료심사 프로세스를 〈그림 3〉과 같이 모델링할 수 있다. 저자는 데이터논문과 데이터셋을 준비한 후 데이터셋을 공공의 데이터저장소에 기탁하고 데이터논문을 제출한다. 그러면, 데이터학술지의 편집위원장 또는 편집팀은 데이터논문의 품질을 초기에 체크하고, 적합한 논문에 대해 심사위원을 초청하여 심사를 진행한다.

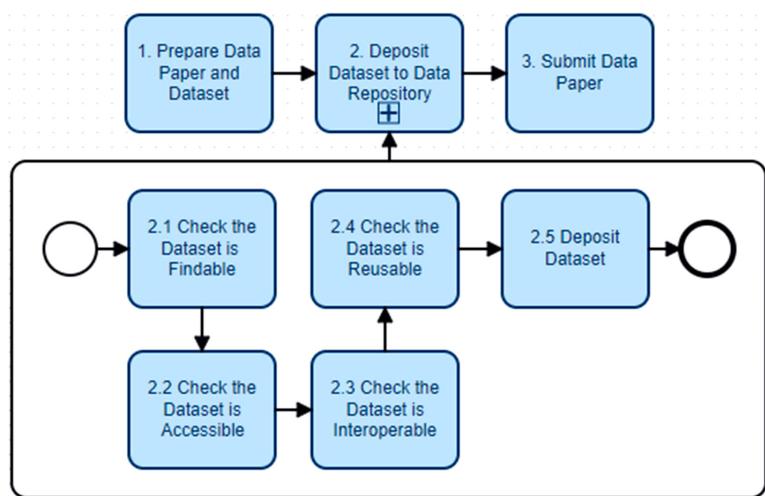


〈그림 3〉 데이터논문 동료심사 프로세스 (제1수준)

심사위원은 전문성과 경험을 바탕으로 데이터논문에 관한 심사 의견을 작성한다. 이러한 심사 의견을 참고하여 데이터학술지의 편집위원(장)은 데이터논문의 게재 및 출판을 진행한다.

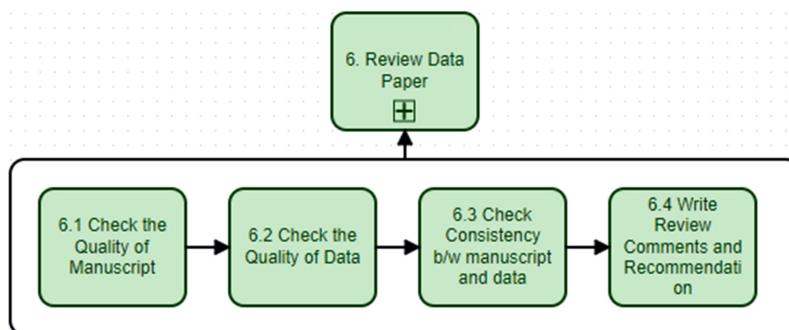
연구논문의 심사과정과 비교하여 데이터논문 제출 과정의 추가 단계는 저자가 데이터셋을 데이터저장소에 기탁하는 것이다. 저자가 데이터셋을 데이터저장소에 기탁하면 이 데이터셋에 대한 DOI와 같은 영구 ID를 발급받을 수 있는데, 이러한 DOI를 데이터논문의 Data Resources 요소에

서술해야 한다. 이때, 데이터논문이 서술하는 데이터셋이 FAIR 원칙을 준수할 수 있도록, 영구 ID에 기반하여 데이터셋의 검색, 접근, 상호운용성, 재사용성을 점검할 수 있는 하위 프로세스를 제2 수준에서 <그림 4>와 같이 모델링 할 수 있다. 이러한 프로세스 모델은 BDJ 동료심사시스템에서 명시적으로 적용하고 있으나 다른 네 종의 학술지는 암묵적으로 저자에게 안내하고 있지만 동료심사시스템의 저자 서비스는 제공하지 않고 있다.



<그림 4> 저자의 데이터셋 기탁 프로세스 (제2수준)

심사위원은 데이터 논문을 심사할 때, 데이터논문에 대한 품질뿐만 아니라, 데이터셋 품질과 데이터논문과 데이터셋의 관계에 대해 심사하는 것이 필요하다. 이러한 과정을 <그림 5>와 같이 명시적으로 모델링하고 동료심사에 적용할 수 있다. 이러한 데이터셋에 대한 심사, 그리고 데이터논문과 데이터셋에 대한 관계에 대한 심사는 일반 연구논문의 심사과정과 차별되는 부분이다.



<그림 5> 심사위원의 데이터논문 심사 프로세스 (제2수준)

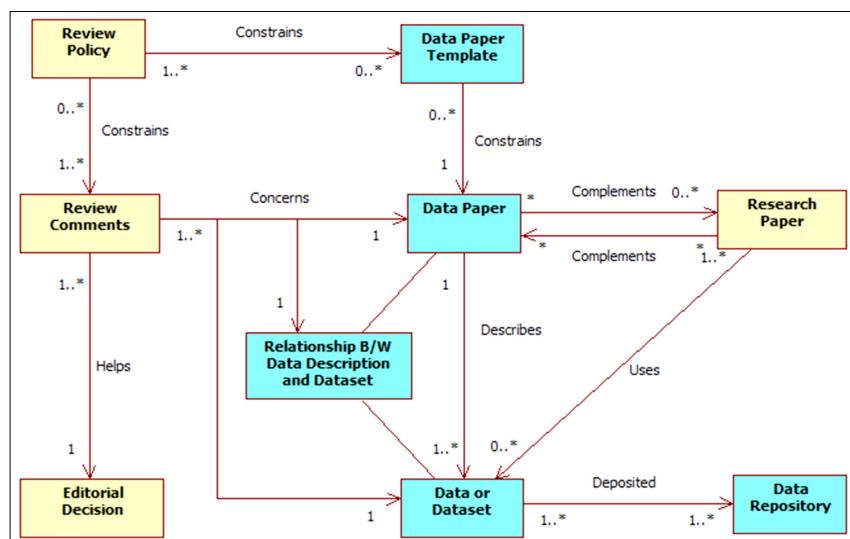
3. 데이터논문 동료심사 개념 모델링

데이터논문 동료심사에 여러 개념이 서로 영향을 미친다. 연구 질문 3에서 제기했던 데이터논문을 심사 출판하기 위해 관계된 핵심 개념을 식별하고, 이러한 개념간 관계를 이해하기 위하여 다섯 종 학술지의 동료심사 프로세스 분석과 모델링을 기반으로 동료심사에 사용되는 핵심 개념을 <표 8>에서 정의하였다.

<표 8> 데이터논문 동료심사의 핵심개념에 대한 정의

개념	정의
데이터논문(Data Paper)	연구 활동을 통해 생산된 데이터셋을 서술한 논문
데이터논문 양식(Data Paper Template)	데이터논문의 세부 형식과 구성요소에 대한 세부 설명을 제공하는 문서
연구논문(Research Paper)	자연현상의 이해, 문제에 대한 해결책 등을 제안한 논문
데이터셋(Dataset)	연구 활동을 통해 생산된 데이터 또는 데이터의 집합
데이터논문과 데이터셋의 관계성 (Relationship between Data Paper and Dataset)	데이터논문과 데이터셋의 관계성 (예: 일관성, 무결성)
데이터저장소(Data Repository)	연구자가 생산한 데이터셋을 저장하여 타 연구자가 접근하고 활용할 수 있도록 서비스를 제공한 인터넷 공간
심사의견서(Review Comment)	심사위원이 논문을 심사하여 작성한 의견서
심사정책(Review Policy)	학술지의 데이터논문 심사를 정책, 방향 또는 기준
논문판정(Editorial Decision)	편집위원(장)의 데이터논문에 대한 최종 출판 판정

<그림 6>은 <표 8>의 핵심개념과 핵심개념간 관계를 클래스 모델링 기법을 사용하여 나타내었다. 데이터논문 구성요소의 개념을 모델링 한 것과 같이, 데이터논문 동료심사와 출판 프로세스와 관련한 개념을 정제하고 정의하여 상세 개념과 개념간 관계를 도출하고 모델링할 수 있다.



<그림 6> 데이터논문 심사의 핵심개념과 개념간 관계

〈그림 6〉의 개념간 관계를 간략히 설명하면, 학술지의 데이터논문 심사정책은 데이터논문 양식과 세부 구성요소에 영향을 미치고, 심사의견서의 세부 구성요소와 심사권고 결정에 영향을 미친다. 심사의견서는 데이터논문, 데이터셋, 데이터논문과 데이터셋에 관한 심사 의견과 심사권고 결정을 포함하며 심사위원의 심사권고 결정은 편집위원장의 최종 판정에 영향을 미치게 된다. 데이터논문에서 서술하는 데이터셋은 Zenodo, Dryad, Figshare, Harvard Dataverse와 같은 공공 데이터저장소에 기탁되고, 데이터논문과 연구논문은 상호 보완적인 관계로 모델링하였다. 연구논문은 공유된 데이터셋을 활용하여 그 분석 결과 또는 새롭게 발견한 사항을 포함할 수 있을 것이다.

VII. 논의 및 시사점

본 연구에서는 학술지 관계자가 새로운 논문의 유형으로 데이터논문을 접수, 심사, 출판하는 과정에서 데이터논문이 핵심적으로 포함해야 할 구성요소 선정, 데이터논문의 동료심사 절차 수립, 데이터논문 동료심사와 관련된 필요한 개념 및 정보시스템의 명확한 이해 등의 문제를 해결할 수 있도록 세 가지 연구 질문을 제기하고 이에 관한 결과를 제시하였다.

첫 번째 질문인 데이터논문에 포함해야 할 핵심 구성요소의 경우, 본 연구는 학술지에서 공통으로 발견되는 요소와 학술지별 특성을 반영한 요소를 분석, 식별하고 제안하였다. 즉, 학문 분야별 데이터학술지 데이터논문의 구성요소는 공통점과 차이점이 존재한다. 데이터논문 구성요소의 공통점은 데이터논문의 메타데이터, 데이터셋에 대한 요소(데이터 생성 배경과 목적, 방법, 데이터 자원의 정보), 참고문헌, 데이터논문의 선언 요소이다. 반면, 학문 분야에 따라 특정 분야에 의존적인 구성요소(예: 생물분류, 위치, 시간, 언어)가 있으므로, 이러한 의존적인 요소는 별도로 추가 정의하고 데이터논문 양식에 반영해야 할 것이다.

두 번째 질문인 연구논문과 차별화되는 데이터논문의 추가적인 동료심사 프로세스의 경우, 본 연구는 데이터셋과 관련한 심사 단계가 필요한 것을 발견하였다. 데이터논문을 제출하는 저자는 데이터논문이 서술하는 데이터셋을 데이터저장소에 기탁하고 공개한 후, 데이터셋에 관한 정보, 즉 DOI와 같은 영구 ID를 데이터논문의 정보자원 부분에 포함한 후 데이터논문을 학술지에 제출하는 단계가 필요하다. 심사위원이 데이터논문을 심사할 때 관련 데이터셋에 접근할 수 없을 때, 데이터논문에 대한 심사는 보류되거나 거절하는 단계도 필요하다. 둘째, 연구논문의 동료심사가 주로 논문의 주장과 품질을 심사하는 것에 비해, 데이터논문의 동료심사는 데이터논문의 품질, 데이터셋의 품질, 데이터논문과 데이터셋의 관계에 대한 품질에 대해 주로 심사를 진행하기 때문에, 심사위원은 연구논문의 심사과정보다 보통 두 단계 심사과정을 추가로 진행하게 된다.

데이터논문의 동료심사와 관련하여 동료심사 방식은 네 종의 학술지가 단일암맹 심사방식을 적용하고 있고 한 종의 학술지는 이중암맹 심사방식을 채택하고 있다. 이 심사방식 이외에 심사위원이 참여하는 패널심사(Panel review), 학술지 회원 중심으로 진행되는 공공 심사(Public review) 방식, 일반 독자가 참여할 수 있는 공개 심사(Open Review) 방식이 있으므로 학술지 정책을 고려하여 학술지에 적합한 심사방식을 적용할 수 있을 것이다(정용일 외, 2022).

세 번째 질문인 데이터논문 동료심사와 관계된 핵심 개념으로는 연구논문, 데이터셋, 데이터저장소, 데이터논문 심사정책, 데이터논문 양식을 식별하고 그 관계성을 설명하였다. 데이터논문은 데이터셋을 생성한 방법과 데이터셋에 접근하여 활용할 수 있는 정보를 중점적으로 제공하는 것에 비해, 연구논문은 주로 자연현상의 이해, 모델링, 또는 새로운 해결책에 관한 주장과 검증을 주로 서술한다. 이러한 특징을 고려하여 데이터논문과 연구논문이 상호 보완될 수 있도록, 예를 들면, 학술지 관계자는 연구논문과 데이터논문을 연계하여 검색 또는 접근 서비스를 제공할 수 있을 것이다. 데이터학술지는 이러한 데이터논문의 제출과 심사를 체계적이고 원활하게 진행할 수 있도록, 데이터논문 정책(예: 심사기준)을 수립한 후, 이와 관련한 데이터논문 양식, 데이터논문 제출 체크리스트, 데이터논문 심사 체크리스트, 데이터저장소 가이드 등을 준비하고 제공하여 저자와 심사 위원이 활용할 수 있도록 안내하는 것이 필요하다.

그 외, 데이터학술지의 분석 및 관찰 내용은 다음과 같다. 데이터논문을 출판하는 학술지는 특정 학문 분야의 학술지(예: BDJ, JOHD, Geo Data)와 다양한 학문 분야를 포괄하는 학술지(예: SciData, DIB)로 구분할 수 있다. 따라서, 데이터논문을 출판하고자 하는 저자는 저자의 학문 분야와 관계되는 데이터학술지를 탐색하고 평가하여, 적합한 학술지를 선택할 필요가 있다. 예를 들면, 데이터셋이 다학제에서 재사용되거나 도움이 될 경우, 저자는 범용적인 데이터학술지에 데이터논문을 제출할 수 있을 것이다. 본 논문에서 분석한 데이터학술지 다섯 종에서 동료 심사하고 출판하는 논문의 유형은 데이터논문이 많은 부분을 차지하지만, 그 외의 논문 유형(예: 데이터와 관련된 연구논문 또는 분석논문)도 심사하고 출판하고 있다. 따라서, 기존 연구논문만을 출판하는 학술지가 새롭게 데이터논문을 출판하고자 할 경우, 이러한 점을 참고하여 데이터논문을 새로운 유형으로 추가로 제출받아 심사하고 연구논문과 데이터논문을 함께 출판할 수 있을 것이다.

본 논문은 네 종의 국외 데이터학술지와 한 종의 국내 데이터학술지를 기반으로 데이터논문의 주요 구성요소, 데이터논문의 동료심사 개념과 프로세스를 분석하였기 때문에, 본 논문의 분석 결과와 모델링 산출물이 모든 데이터논문의 동료심사 개념과 특징을 반영한다고 일반화할 수 없다. 개별 학술단체가 데이터논문의 심사와 출판을 준비할 때 학문 분야별 데이터논문의 구성요소, 동료심사 정책 및 프로세스가 다름을 이해하고 본 논문의 연구내용을 참고하여 학문 분야, 학술지의 목적과 범위 등을 고려하여 응용하여야 할 것이다.

VII. 결론 및 향후 과제

본 논문은 데이터학술지 다섯 종을 조사, 분석하여 데이터논문의 핵심 구성요소를 식별 및 분류하고 클래스 다이어그램을 활용하여 개념 모델링하였다. 다음 단계로 연구논문 동료심사 프로세스와 차이가 있는, 데이터논문 동료심사에 활용되는 상세 프로세스를 식별하고, 저자, 심사위원 중심으로 데이터논문 동료심사 프로세스를 모델링하였다. 또한, 데이터논문 동료심사에 사용되는 핵심 개념들을 도출하고 이러한 개념간 관계를 모델링하고 제안하였다. 본 논문의 기여 사항은 다섯 종의 데이터학술지를 분석하여 데이터논문의 특성을 반영한 데이터논문의 핵심 구성요소를 식별하고, 동료심사에 관한 핵심 개념 및 프로세스를 모델링하여 그 결과를, 저자의 인지하는 범위에서, 선도적으로 제안한 것이다. 데이터논문 중심의 데이터학술지를 신규로 출판하거나 데이터논문을 기존의 연구논문과 함께 추가로 심사, 출판하고자 하는 학술단체는 본 논문에서 제시한 데이터논문의 핵심개념, 데이터논문 동료심사 프로세스 및 핵심개념 모델을 활용하여 데이터논문의 양식 개발, 동료심사 프로세스 및 정책개발 등의 학술지 업무에 적용할 수 있을 것이다. 이를 통해, 데이터학술지 관계자는 데이터논문의 핵심 개념이 반영된 양질의 데이터논문을 접수하고, 효율적인 심사 프로세스를 통해 품질 높은 데이터논문을 출판하는 데 도움이 될 것으로 기대된다.

본 연구에서 데이터학술지의 데이터 동료심사 정책에 관해서는 연구범위와 지면상 상세히 분석하지 않았는데 이러한 심사정책과 데이터논문 구성요소의 관계, 심사정책과 동료심사의 여러 개념 간 관계가 데이터논문의 동료심사에 어떠한 영향을 주는지에 관한 연구가 필요하다. 또한, 데이터학술지가 신뢰할 수 있는 데이터저장소의 특징, 데이터 동료심사 시스템에서 데이터저장소 시스템으로의 원활한 데이터 연계 등에 관해 추가 연구도 필요하다.

참 고 문 헌

- 김기동 (2023). GEO DATA 저널 데이터 논문 동료심사 현황. 2023년 GeoAI데이터학회 춘계컨퍼런스 논문집, 89-89.
- 서태설 (2022). 데이터 공유에서 데이터 출판으로. KCSE Newsletter, 41, 9-12.
- 이현정, 정영임, 황혜경 (2023). 학술지 대상 데이터 공유/데이터 논문 정책 및 인식조사. 2023년 GeoAI데이터학회 춘계컨퍼런스 논문집, 82-82.
- 정영임, 권오석, 김기동, 김소형, 서태설, 김선태 (2020). 생태학 분야 데이터 저널 발행 전략 연구: K기관을 중심으로. 한국도서관·정보학회지, 51(4), 83-100.
<https://doi.org/10.16981/kliss.51.4.202012.83>

- 정영임, 조성남 (2022). 데이터 논문 출판 현황 분석, 데이터출판 세미나, 서울, 국립중앙도서관.
- 정용일, 노지윤, 조성남, 안성수 (2022). 과학기술 학회의 개방형동료심사에 관한 인식 연구: 한국과학기술총연합회 산하 학회를 중심으로. *한국콘텐츠학회논문지*, 22(10), 59-73.
<https://doi.org/10.5392/JKCA.2022.22.10.059>
- 한종규 (2023). 지구과학분야 연구데이터 출판 현황. 2023년 GeoAI데이터학회 춘계컨퍼런스 논문집, 88-88.
- 횡혜경, 정영임, 조성남, 서태설, 김지현 (2023). 과학기술분야 연구자들의 데이터 출판경험 및 인식 연구. *한국도서관·정보학회지*, 54(1), 45-68. <https://doi.org/10.16981/kliss.54.1.202303.45>
- Ahn, R., Supakkul, S., Zhao, L., Kolluri, K., Hill, T., & Chung, L. (2021). Validating business problem hypotheses: a goal-oriented and machine learning-based approach. In International Conference on Big Data (pp. 17-33). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-96282-1_2
- Booch, G., Rumbaugh, J., & Jacobson, I. (2005). Unified Modeling Language User Guide, The 2nd Edition, Addison-Wesley Object Technology Series.
- Candela, L., Castelli, D., Manghi, P., & Tani, A. (2015). Data journals: a survey. *Journal of the Association for Information Science and Technology*, 66(9), 1747-1762.
<https://doi.org/10.1002/asi.23358>
- Carpenter, T. (2017). What constitutes peer review of data? A survey of peer review guidelines. *The Scholarly Kitchen*.
- Chavan, V. & Penev, L. (2011). The data paper: a mechanism to incentivize data publishing in biodiversity science. *BMC bioinformatics*, 12, 1-12.
<https://doi.org/10.1186/1471-2105-12-S15-S2>
- Elmasri, R. & Navathe, S. B. (2015). Fundamentals of Database Systems (Seventh edition). Pearson.
- Gomes, D. G., Pottier, P., Crystal-Ornelas, R., Hudgins, E. J., Foroughirad, V., Sánchez-Reyes, L. L., ... & Gaynor, K. M. (2022). Why don't we share data and code? perceived barriers and benefits to public archiving practices. *Proceedings of the Royal Society B*, 289(1987), <https://doi.org/10.1098/rspb.2022.1113>
- Kim, S. Y., Yi, H. J., & Huh, S. (2019). Current and planned adoption of data sharing policies by editors of Korean scholarly journals. *Science Editing*, 6(1), 19-24.
<https://doi.org/10.6087/kcse.151>
- Lawrence, B., Jones, C., Matthews, B., Pepler, S., & Callaghan, S. (2011). Citation and

- peer review of data: moving towards formal data publication. International Journal of Digital Curation, 6(2), 4-37. <https://doi.org/10.2218/ijdc.v6i2.205>
- Mylopoulos, J., Chung, L., & Yu, E. (1999). From object-oriented to goal-oriented requirements analysis. Communications of the ACM, 42(1), 31-37. <https://doi.org/10.1145/291469.293165>
- Tedersoo, L., Küngas, R., Oras, E., Köster, K., Eenmaa, H., Leijen, Ä., ... & Sepp, T. (2021). Data sharing practices and data availability upon request differ across scientific disciplines. Scientific Data, 8(1), 192. <https://doi.org/10.1038/s41597-021-00981-0>
- White, S. A. & Bock, C. (2011). BPMN 2.0 Handbook: Methods, Concepts, Case Studies and Standards in Business Process Management Notation (Second Edition). Future Strategies Inc., Book Division.
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., ... & Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data, 3(1), 1-9. <https://doi.org/10.1038/sdata.2016.18>

• 국한문 참고문헌의 영문 표기

(English translation / Romanization of references originally written in Korean)

- Han, Jong-Gyu (2023). GEO current state of data publishing in earth system science. The Spring Conference of GeoAIDData Society, 88-88.
- Hwang, Hyekyong, Jung, Youngim, Cho, Sung-Nam, Seo, Tae-Sul, & Kim, Jihyun (2023). A study on awareness and experience of data publishing by scientists. Journal of Korean Library and Information Science Society, 54(1), 45-68. <https://doi.org/10.16981/kliss.54.1.202303.45>
- Jeong, Yong-il, Ro, Ji-Yoon, Cho, Sung-Nam, & Ahn, Sungsoo (2022). A study on science and technology scholarly societies' understanding on open peer review. The Journal of the Korea Contents Association, 22(10), 59-73. <https://doi.org/10.5392/JKCA.2022.22.10.059>
- Jung, Youngim & Cho, Sung-Nam (2022). Analysis of Current Data Paper Publication, Data Publication Seminar, Seoul, National Library of Korea.
- Jung, Youngim, Kwon, Ohseok, Kim, Kidong, Kim, Sohyeong, Seo, Tae-Sul, & Kim, Suntae (2020). A study on the strategies for publishing data journals in the field of ecology:

- focused on K institution. Journal of Korean Library and Information Science Society, 51(4), 83-100. <https://doi.org/10.16981/kliss.51.4.202012.83>
- Kim, Kidong (2023). GEO DATA: current state of data peer review. The Spring Conference of GeoAIData Society, 89-89.
- Seo, Tae-Sul(2022). Towards data publishing from data sharing. KCSE Newsletter, 41, 9-12.
- Yi, Hyun Jung, Jung, Youngim, & Hwang, Hyekyong (2023). Survey: data sharing/data paper policies by editors of Korean scholarly journals. The Spring Conference of GeoAIData Society, 82-82.