

과학기술분야 연구자들의 데이터 출판경험 및 인식 연구*

A Study on Awareness and Experience of Data Publishing by Scientists

황 혜 경(Hyekyong Hwang)**, 정 영 임(Youngim Jung)***
조 성 남(Sung-Nam Cho)****, 서 태 설(Tae-Sul Seo)*****
김 지 현(Jihyun Kim)*****

< 목 차 >

I. 서 론	IV. 분석 결과
II. 선행연구	V. 결론 및 시사점
III. 연구설계 및 방법	

요 약: 본 연구는 오픈사이언스 연구환경 변화가 가속화되면서 데이터 공유의 새로운 채널로 인정받고 있는 데이터 출판에 대한 국내 연구자의 경험과 인식을 조사하고자 수행되었다. 문헌조사와 국내 과학기술분야 5개 정부출연연구기관 연구자와 GeoAI 데이터 학회 회원을 대상으로 설문조사를 실시하여, 데이터 출판에 대한 인식 정도를 측정하였다. 연구결과, 국내 연구자들은 데이터에 대한 설명제공, 데이터로의 안정적인 접근, 인용, 동료심사를 통한 품질확인을 데이터 저널의 장점으로 인식하였으나, 연구실적으로써의 데이터 논문 인정에는 낮은 인식수준을 보였다. 또한 데이터 상세설명, 메타데이터 기술, 영구식별기호 부여 순으로 데이터 출판 속성을 인식하고 있었으나 리포지터리에 데이터 제출과 동료심사에 대해서는 상대적으로 낮았다. 데이터 출판 촉진 요인으로 데이터 논문작성과 리포지터리 연계를 지원하는 투고심사시스템 개발과 데이터 인용을 지원하는 리포지터리 개발의 필요성이 확인되었다. 국내 연구자들의 데이터 출판 경험과 인식은 향후 데이터 출판 서비스와 인프라 구축을 위한 중요한 기초가 될 것이다.

주제어: 데이터 출판, 데이터 저널, 데이터 논문, 데이터 출판경험, 정부출연 연구기관, GeoAI 데이터학회

ABSTRACT: This study aims to investigate the awareness and experiences of domestic researchers regarding data publishing, which has been recognized as a new channel of data sharing as scholarly communication evolves in the open science environment. A survey is conducted among researchers from five government-funded research institutes in the field of science and technology and members of the GeoAI Data Society to confirm the awareness of data publishing. As a result of the study, domestic researchers recognized providing explanations for data, stable access to data, citation, and quality assurance through peer review as the advantages of data journals. On the contrary, a low level of recognition for data paper as one of the research outputs was presented. With regard to the properties of data publication, the respondents answered that the data description, metadata description, and permanent identifiers are highly related, however, their recognition of the relation between the properties of data publication and the data submission to a repository and data peer review was relatively low. Finally, to expand the data publication, the need for the development of an editorial system that supports data paper peer review and cross-linking to a data repository as well as the development of a repository that supports data citation was identified. This study on the domestic researchers' experience and awareness of data publishing can provide insights for the implementation of data publishing services and infrastructure in the future.

KEYWORDS: Data Publishing, Data Journal, Data Paper, Data Publishing Experience, Government-funded Research Institutes, GeoAI Data Academic Society

- * 이 연구는 국가과학기술연구회 글로벌 공동연구 촉진사업 '연구데이터 동료심사 방안 및 데이터 영향력 분석 연구(Grant No. NST-02)' 과제의 지원을 받았음.
- ** 한국과학기술정보연구원 책임연구원(hkhwang@kisti.re.kr / ISNI 0000 0004 6461 3367) (제1저자)
- *** 한국과학기술정보연구원 책임연구원, 과학기술연합대학원대학교 전임교원 (acorn@kisti.re.kr / ISNI 0000 0004 5935 5842) (공동저자)
- **** 한국과학기술정보연구원 책임연구원(chosn@kisti.re.kr / ISNI 0000 0004 6453 1601) (공동저자)
- ***** 한국과학기술정보연구원 책임연구원(tsseo@kisti.re.kr / ISNI 0000 0004 6501 522X) (공동저자)
- ***** 이화여자대학교 문헌정보학과 교수(kim.jh@ewha.ac.kr / ISNI 0000 0004 6336 969X) (교신저자)

- 논문접수: 2023년 2월 20일 • 최초심사: 2023년 3월 6일 • 게재확정: 2023년 3월 15일
- 한국도서관·정보학회지, 54(1), 45-68, 2023. <http://dx.doi.org/10.16981/kliss.54.1.202303.45>

© 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. 서론

공적자금이 투입된 대표적인 연구성과인 논문 등의 출판물과 데이터 공유 및 개방형 연구협력과 소통을 강조하는 오픈 사이언스가 국제적인 정책 의제로 제시되고 있다. 최근 미국 백악관 과학기술정책실(OSTP)은 2022년 8월 연방정부 및 유관기관을 대상으로 완전하고 즉각적인 공공접근 정책 지침을 발표한 바 있다(Office of Science and Technology Policy, 2022). 본 지침은 공적자금으로 생산된 모든 논문의 즉각적이고, 전면적인 공개뿐만 아니라 관련된 데이터의 공개도 의무화하는 정책으로 오픈사이언스와 오픈엑세스 흐름에 시사하는 바가 크다. 그러나 데이터 공유에 드는 노력과 시간, 데이터 오류 노출 가능성과 논문 출판 기회를 선점하지 못할 수 있다는 우려와 데이터 공유에 대한 인센티브 부족이 주요 방해요인이 된다는 선행조사가 있다(Gorgolewski, Margulies, & Milham, 2013). 이러한 상황에서 연구데이터 공유 촉진을 위한 인센티브로 연구자들이 생산한 데이터를 기존의 학술논문과 같이 동료심사 등 품질 확인 절차를 거쳐 누구나 이용할 수 있게 공개하는 데이터 출판이 새로운 데이터 공유 방식으로 주목받고 있다(Kratz & Strasser, 2014). 그러나 국내에서의 데이터 출판에 관한 연구자 인식을 반영한 연구는 국내 생태학분야에서의 데이터 저널 창간 전략을 제안하기 위한 포커스 그룹 인터뷰를 통한 질적연구(정영임 외, 2020)와 데이터 논문을 출판한 국내 저자들을 대상으로 데이터 논문 출판에 대한 인식을 조사한 연구(Lee & Kim, 2021), 그리고 해외 데이터 논문 저자 대상으로 데이터 논문 출판 동기요인을 조사한 연구(이승은, 2020) 등으로 소수이다. 이에 본 연구의 목적은 과학기술분야 연구자의 데이터 출판 경험과 인식을 파악하고 향후 데이터 출판 서비스 및 지원 방안을 제안하는 것이다. 이 목적을 달성하기 위한 구체적인 연구 질문은 다음과 같다.

- 연구질문 1: 과학기술분야 연구자들은 어떤 유형의 연구데이터를 생산하며 이를 어디에 보관하는가?
- 연구질문 2: 연구자들의 데이터 저널 출판 경험의 정도와 선택 기준 및 장단점에 대한 인식은 어떠한가?
- 연구질문 3: 연구자들의 향후 데이터 저널 투고 의향 여부에 따라 데이터 출판 속성의 중요도 인식에 차이가 존재하는가?
- 연구질문 4: 연구자들이 중요하다고 인식하는 데이터 출판 촉진 요인은 무엇인가?

본 연구에서는 데이터 출판의 대표적인 유형인 데이터 저널이 대다수 과학기술분야에서 발행되고 있음(Adamczak, 2021)을 고려하여 과학기술분야 연구자들을 대상으로 연구를 수행하였다. 본 연구 결과는 기존 연구의 결락을 보완하고 연구자들의 요구에 맞는 서비스와 인프라 구축을 위한 기초자료가 될 것으로 기대한다.

II. 선행연구

1. 데이터 출판

데이터 출판을 논의한 초기 연구로서 Lawrence et al.(2011)는 데이터 출판을 “인터넷을 통해 데이터를 가능한 한 영구적으로 이용할 수 있게 하는 것”으로 정의하고 있다. 출판된 데이터는 영구식별기호를 부여받아 안정적으로 인용 및 활용될 수 있어야 하며 데이터에 대한 장기적인 큐레이션이 필요함을 제안하였다. 더불어 데이터의 맥락과 출처를 나타내는 상세한 메타데이터 기술 및 동료심사를 기반으로 하는 데이터와 메타데이터 품질 관리 필요성을 강조하였다. Lawrence et al.(2011)는 데이터 출판 개념을 논의한 초기 연구인 만큼 학술지를 통한 데이터 공유의 방식들을 5가지 유형으로 구분하고 있다. 첫째, 리포지터리를 통한 독립적인 데이터 출판, 둘째, 학술지 연구논문 출판 시 데이터에 대한 설명자료와 연구논문을 연계하는 방식의 데이터 출판, 셋째, 학술지 연구논문 출판 시 데이터를 보충 자료로 출판, 넷째, 학술지 연구논문 출판 시 데이터를 리포지터리에 기탁하게 하는 방식으로 출판, 다섯째, 데이터 논문을 출판하는 오버레이(overlay) 저널을 통한 데이터를 출판하는 방식이다. Kratz와 Strasser(2014)는 데이터 출판을 위한 주요 속성을 제안하였다. 데이터에 영구식별기호를 부여하여 공식적으로 인용될 수 있도록 하는 인용가능성, 데이터에 대한 상세 설명을 제공하는 데이터 기록화, 데이터에 대한 품질확인과 데이터 재이용이 가능하도록 공개하여야 한다는 활용 가능성의 네 가지 속성이다. 이 중에서도 데이터 동료심사 등 품질 확인의 영역은 다른 영역에 비해 뚜렷하게 정립되지 못한 상태이며 이에 대한 지속적인 논의가 필요함을 제안하였다. Austin et al.(2017)는 데이터 출판의 구성요소를 필수요소와 추가 구성요소로 구분하였다는 점에서 차별점이 있다. 필수요소로 영구식별기호 부여, 데이터에 대한 메타데이터 등 상세한 설명 제공, 데이터 활용을 가능하도록 하기 위한 인용 형식 준수와 데이터 세트에 관한 기술과 데이터 리포지터리 기록과의 일치성을 두었다. 그리고 추가 구성요소로는 맥락, 품질, 접근가능성 및 가시성을 고려할 수 있다고 제안하였다. Austin et al.(2017)는 리포지터리 및 데이터 저널을 통해 데이터 출판이 이루어지며 향후 이용을 위해 출판된 연구데이터를 장기간에 걸쳐 기록화하고 데이터 출판을 위한 큐레이션을 통해 상호운용성, 인용, 품질 및 검색 가능성을 보장해야 한다고 제안하였다. 이러한 데이터 출판의 궁극적인 목적은 데이터의 재이용과 재현 가능성을 지원하는 것이다(Candela et al, 2015; Kratz & Strasser, 2014). Austin et al.(2017)는 Kratz와 Strasser(2014)가 제시한 데이터 출판의 속성 중에서 영구식별기호 부여, 데이터에 대한 메타데이터 등 상세한 설명 제공, 데이터를 활용 가능하도록 공개하는 부분은 데이터 출판의 필수 요소로 인정하였으나, 동료심사 등을 통한 데이터 품질확인 구성요소는 추가적인 구성요소로 제시하고 있음을 알 수 있다.

데이터 출판에서 강조되는 속성이 연구마다 다소 차이가 있기는 하지만 공통적으로 데이터의 활용 가능성 확보, 데이터에 대한 상세한 설명 제공, 데이터의 인용을 지원하는 영구 식별기호 부여 및 데이터 품질 확인 수행이 데이터 출판의 중요한 속성으로 일관되게 제시되고 있다.

2. 데이터 저널

데이터 저널은 데이터 논문을 출판하는 학술지를 지칭하며 데이터 논문을 주로 출판하는 순수형(pure) 저널과 데이터 논문뿐만 아니라 여러 유형의 논문도 출판하는 혼합형(mixed) 저널로 구분된다(Candela et al., 2015). 데이터 저널과 관련해서는 품질 관리에 대한 연구가 다수 존재하고 있다. 데이터 저널의 중요한 특징 중 하나는 동료심사를 통해 데이터 및 데이터 논문의 품질 관리를 수행한다는 것이다. 그러나 데이터 동료심사의 경우 데이터 품질 평가 기준과 방법에 대한 합의가 부족한 상황이며 데이터 동료심사에 대한 비판적인 의견들도 존재한다(Huang, Hawkins, & Qiao, 2013). 데이터 동료심사를 수행해 줄 수 있는 적절한 심사위원을 구하기 어려운 점과 데이터 용량이 너무 크거나 데이터에 대한 설명 부족 등으로 인한 데이터 논문에 대한 이해의 어려움을 지적하였다. Candela et al.(2015)는 혼합형 데이터 저널 109종과 순수형 데이터 저널 7종을 대상으로 데이터 저널에서의 동료심사 방법을 조사하였다. 대부분의 데이터 저널은 출판 전 심사 및 익명 심사(blind review)로 대표되는 비공개방식의 동료심사를 수행하고 있었으며 이는 데이터 저널이 일반적인 학술지의 동료심사 방법을 따르고 있다는 것을 나타내고 있었다. 또한, 매우 소수의 저널에서만 공개방식의 동료심사를 수행하고 있었으며 일부 저널은 공개와 비공개 방식을 절충한 방식을 사용하고 있었다. 그 외 연구자들을 대상으로 데이터 출판에 대한 기대를 조사한 연구(Kratz & Strasser, 2015)를 살펴보면 데이터가 동료심사를 거쳐 출판될 것이라 기대하는 비율은 전체 응답자의 29%에 불과하였으나 동료심사를 통해 출판된 데이터는 더 높은 품질을 가지고 신뢰할 수 있을 것이라 인식하고 있었다. 이승은(2020)은 Web of Science에 등재된 데이터 저널 중 데이터 논문의 비율이 20% 이상인 8개 저널의 교신저자 453명의 설문조사 응답을 바탕으로 데이터 투고 단계, 데이터 동료심사 단계, 그리고 게재 확정 및 출판 이후 단계에서의 저자들의 경험이 재투고 의향에 미치는 영향을 조사하였다. 이를 통해 데이터 저널 익숙도, 데이터 리포지터리 가이드라인의 유용도, 데이터 템플릿 및 가이드라인 유용도, 데이터 투고시스템 만족도, 동료심사 과정 만족도, 게재 확정 및 출판 후 만족도와 재투고 의향 간에는 통계적으로 유의한 관계가 있음을 확인하였다. Lee와 Kim(2021)은 국내 연구기관 소속 교신저자를 포함하고 있는 4종의 데이터 저널을 선정하여 이들 교신저자 중 인터뷰 참여에 동의한 7명을 대상으로 데이터 논문 투고 경험에 대한 질적 연구를 수행하였다. 연구결과, 인터뷰 참여자들은 데이터 논문 출판으로 연구성과를 인정받는다라는 점에서 데이터 저널의 가치를 높이 평가하고

있었으며, 데이터 저널은 데이터에 관한 상세한 기술과 더불어 리포지터리를 통해 데이터를 공유하고 인용을 통해 데이터 기반의 후속 연구가 이루어질 수 있는 기반을 마련한다는 점에서 중요한 학술 커뮤니케이션 채널로 인식되고 있었다. 이상과 같이 국내 연구자 중 데이터 저널에 데이터 논문을 투고한 경험이 있는 연구자들을 대상으로 데이터 저널에서의 출판 경험과 인식을 조사한 연구가 소수 진행되었으나 국내 연구자를 대상으로 설문조사를 통해 과학기술분야 전반을 조사한 연구는 아직 이루어지지 않은 실정이다. 본 연구에서는 데이터 논문 출판이 비교적 활성화되어 있는 과학기술분야 연구자들을 대상으로 연구자들의 데이터 출판 경험과 인식을 조사하고 시사점을 논의하고자 한다.

III. 연구설계 및 방법

본 연구는 데이터 출판 경험과 데이터 저널에 대한 인식을 조사하기 위해 과학기술분야에 종사하는 연구자를 대상으로 설문조사를 실시하였다. 자세한 연구방법과 내용은 다음과 같다. 설문조사 설계를 위해 연구자들을 대상으로 데이터 공유에 대한 인식을 조사한 연구(한국사회과학자료원, 2021; Tenopir et al., 2015)와 데이터 출판에 관한 해외 연구자 인식조사 연구(이승은, 2020; Kratz & Stasser, 2015)의 설문지를 주로 참고하였다. 이들 연구는 비교적 최근에 수행되었으며 연구자들 간 데이터 공유 및 출판 경험과 인식 수준의 다양성을 고려하여 설문 문항을 설계하였다는 점에서 본 연구의 설문조사 설계에 적절한 참고자료라고 판단하였다.

설문지는 연구데이터 생산/관리(4문항), 연구데이터 공유/출판(13문항), 연구데이터 출판에 대한 인식(6문항), 인구통계정보(4문항)의 네 영역으로 구성하였다. 그리고 후속 연구 인터뷰 참여 동의 여부를 묻는 1문항을 포함하여 총 28문항으로 작성되었다(〈표 1〉 참조).

〈표 1〉 설문지 구성도

구분	설문 내용
연구데이터 생산과 관리 (4문항)	데이터 수집 및 생산 유형, 데이터 보관 장소
연구데이터 공유와 출판 (13문항)	데이터 공유 범위, 데이터 공유 방법, 데이터 저널 인식여부, 데이터 논문 출판 여부/건수, 데이터 저널 선택 기준, 데이터 저널의 장점에 대한 동의정도, 데이터 저널의 우려사항에 대한 동의정도, 데이터 리포지터리 선택 기준, 데이터 공유를 위한 리포지터리 명, 데이터 공유 장애이유
연구데이터 출판에 대한 인식 (6문항)	데이터 출판 속성별 인식정도, 데이터 논문의 동료심사 평가 측면, 데이터 출판 촉진 요인, 데이터 논문 출판 의향, 데이터 논문 투고 시기
인구통계정보, 후속 인터뷰참여 (5문항)	연령, 성별, 최종 학위, 연구분야, 후속 인터뷰 참여 여부

설문 참여율을 높이기 위해 응답에 관한 사례로 음료쿠폰을 발송하였으며 후속 인터뷰 참여에 동의한 응답자 연락을 위해 설문조사 후 개인정보 수집에 동의하는 응답자들을 대상으로 핸드폰 번호를 수집하였다. 후속 인터뷰 참여에 동의하지 않는 경우 음료 쿠폰 발송 후 즉시 파기하고 후속 인터뷰 참여에 동의한 경우 인터뷰 수행 후 파기하는 것으로 안내하였다.

온라인 설문조사(서베이몽키, <https://ko.surveymonkey.com>)기반으로 설문지를 구축하여 2022년 6월 7일부터 10일에 걸쳐 파일럿 테스트를 통해 문항 내용 및 설문지 기능 오류를 확인하고 수정작업을 거쳐 최종 온라인 설문지를 완성하였다. 온라인 설문지 배포 및 응답 수집은 2차에 걸쳐 수행되었다. 1차 수집은 5개 정부출연연구기관(국립생태원, 한국과학기술정보연구원, 한국지질자원연구원, 한국표준과학연구원, 한국한의학연구원)의 홈페이지 게시판에 설문지 링크를 올리는 방식으로 6월 15일부터 29일까지 2주 동안 진행되었다. 2차 수집은 국내 데이터 저널인 Geo Data를 발행하는 학회인 GeoAI 데이터학회 회원들을 대상으로 6월 29일부터 7월 5일까지 1주 동안 진행되었다. 1차 수집을 통해 275명의 설문 응답을 받았고 2차 수집에서는 55명으로부터 응답받았다. 총 330개의 응답 중 불완전한 응답(거의 응답하지 않았거나 절반 이하로 응답)은 62개(1차: 51개, 2차: 11개)이었으며 이를 제외한 268개 응답을 최종 분석 대상으로 정하였으며, 268개 응답 중 9개는 인구통계 질문에 미응답하였으나 나머지 주어진 모든 문항에 대해 답변을 제공하였으므로 분석대상에 포함하였다.

설문 데이터는 본 연구 주제 범위인 데이터 출판과 인식에 중점을 두어 분석하였으며, 연구주제 범위에 해당하지 않는 연구데이터 공유 부분은 본 고에서 제외하였다. 4장의 분석 결과는 주로 기술 통계 분석에 기반하여 서술하였다. 또한 두 집단의 차이를 비교하는 비모수 검정 기법인 Mann-Whitney 검정을 활용하여 데이터 출판 경험 유무에 따른 데이터 저널의 한계에 대한 인식 차이와 향후 데이터 저널 투고 의향 여부에 따른 데이터 출판 중요도 인식 차이를 조사하였다.

IV. 분석 결과

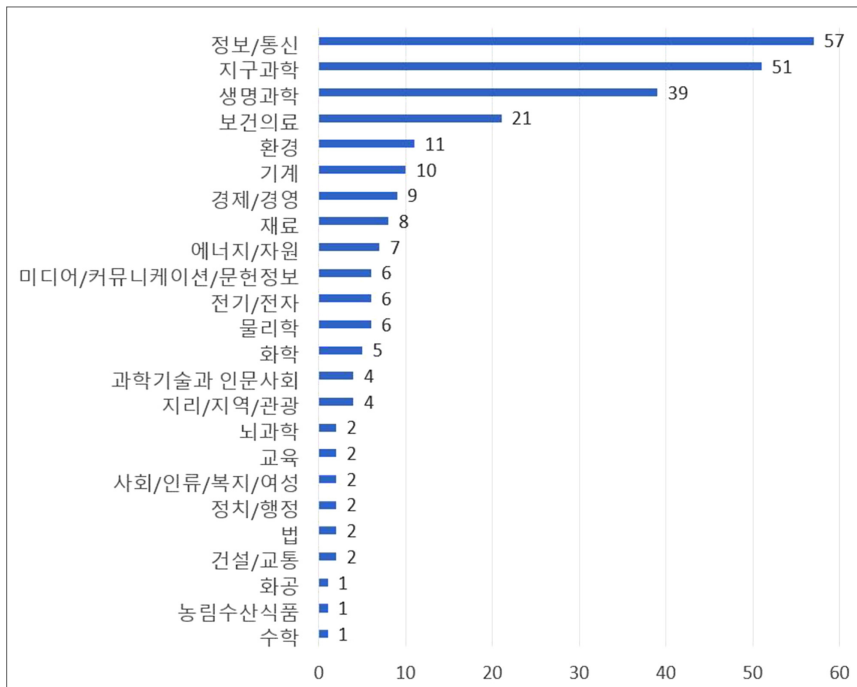
1. 인구통계학적 특징

설문 응답자 268명 중 인구통계 질문에 답한 259명의 응답을 바탕으로 연령, 성별, 학위 및 연구경력을 <표 2>에 제시하였다. 연령별로는 30대(37.1%)와 40대(34.0%)가 약 71%를 차지하였으며 남성의 비율이 64.9%이었다. 응답자 중 151명(58.3%)이 박사학위 소지자인 것으로 나타났다.

〈표 2〉 응답자 인구통계 특성

인구통계정보	범주	빈도	비율(%)
연령	20대	30	11.6
	30대	96	37.1
	40대	88	34.0
	50대	39	15.1
	60대 이상	6	2.3
성별	여성	91	35.1
	남성	168	64.9
학위	학사	36	13.9
	석사	72	27.8
	박사	151	58.3
합계		259	100.00

응답자들의 연구 분야를 응답 빈도가 높은 순서대로 나타내면 〈그림 1〉과 같다. 설문 문항에서는 한국연구재단 국가과학기술표준분류의 대분류 33개 범주로 제시하였으며 이 중에서 24개 분야에 응답이 존재하였다. 대부분 과학기술 분야이며 경제/경영, 미디어/커뮤니케이션/문헌정보 등 사회과학 분야도 일부 포함되어 있다. 응답자 비율이 높은 연구 분야는 정보/통신(57, 22.0%), 지구과학(51, 19.7%), 생명과학(39, 15.1%) 순으로 나타났다.



〈그림 1〉 응답자 연구분야

2. 연구데이터 생산 및 보관

본 연구의 첫 번째 연구 문제로 과학기술분야 연구자들이 생산하는 연구데이터 유형과 이를 보관하는 장소를 살펴보는 것을 제시하였다. 이를 위해 우선 응답자들에게 연구와 관련하여 데이터를 수집 또는 생산한 경험 유무를 질문하였을 때 경험이 있다는 응답자가 268명 중 218명(81.3%)인 것으로 나타났다. 이들을 대상으로 생산·수집한 연구데이터의 유형, 데이터 보관 정도 및 장소를 조사하였다. 연구데이터 유형 중 실험 데이터를 생산·수집한다는 응답 비율이 가장 높았으며 다음으로 모델 및 시뮬레이션 데이터, 설문 데이터, 관찰 데이터의 순으로 응답 비율이 높았다(〈표 3〉 참조). 실험 데이터를 생산·수집한다는 응답자 중 지구과학(18명), 생명과학(16명), 보건의료(8명), 재료(6명), 정보/통신(17명) 분야 연구자들이 다수를 차지하였다. 모델 및 시뮬레이션 데이터의 경우 정보/통신(27명) 분야 연구자들이 가장 많았으며 지구과학(10명) 및 기계(8명) 분야 연구자들의 응답이 다음으로 많았다. 설문 데이터를 생산·수집한다는 응답자 중에는 보건의료(14명), 정보/통신(10명) 외에 경제/경영(3명), 지리/지역/관광(3명), 미디어/커뮤니케이션/문헌 정보(4명) 등 사회과학 분야 연구자들의 응답이 포함되었다. 관찰 데이터의 경우 생명과학(12명), 정보/통신(8명) 외에 보건의료(6명), 환경(6명), 지구과학(5명), 과학기술과 인문사회(3명) 분야 연구자들이 응답하였다. 기타에 응답한 22명 중 18명이 데이터 유형을 기술하였으며 이 중 12명이 텍스트 데이터 또는 이에 상응하는 데이터 유형을 언급하였다. 텍스트 데이터라는 응답 외에 논문 등 참고문헌, 특허, 뉴스 댓글, 학술지나 논문의 메타데이터, 문헌 조사 데이터 등 텍스트 데이터의 상세 유형을 명시한 응답도 있었다. 그 외 교통 정보 및 지리정보, 식물 표본정보, 기업 데이터, 기타 원격탐사 데이터, 통계 데이터를 기타 유형으로 제시하였다.

〈표 3〉 연구데이터 유형

구분	빈도*	비율(%)
실험 데이터(실험실 환경, 실험 장비 혹은 현장실험에서 생산되는 데이터)	91	20.5
모델 및 시뮬레이션 데이터	67	15.1
설문 데이터	57	12.9
관찰 데이터	47	10.6
생물(biotic) 연구데이터	41	9.3
비생물 원격탐사 데이터(기후 데이터 등)	40	9.0
비생물(abiotic) 연구데이터(토양, 수문[hydrology] 등)	36	8.1
인터뷰 데이터	26	5.9
생물 원격탐사 데이터	16	3.6
기타	22	5.0
합계	443	100.0

(*복수응답 허용)

연구데이터 보관 장소의 경우 개인 PC, USB 또는 하드 드라이브를 활용한다는 응답(196, 50.4%)과 개인용 클라우드에 보관한다는 응답(45, 11.6%) 등 연구자 개인의 온오프라인 저장 공간이 가장 많이 활용되는 것으로 나타났다. 연구팀 공용 드라이브나 공동 클라우드에 보관(68, 17.5%) 한다는 응답과 소속기관의 리포지터리에 보관(54, 13.9%) 한다는 응답이 각각 두 번째와 세 번째로 비율이 높은 것으로 나타나 팀 단위 또는 소속기관 단위의 데이터 보관이 어느 정도 이루어지고 있다는 것을 알 수 있었다(〈표 4〉 참조).

〈표 4〉 연구데이터 보관 장소

구분	빈도(*)	비율(%)
개인 PC나 USB, 또는 하드 드라이브에 보관	196	50.4
연구팀 공용 드라이브나 공용 클라우드에 보관	68	17.5
소속기관의 리포지터리에 보관	54	13.9
개인용 클라우드에 보관	45	11.6
내 연구 분야의 데이터 아카이브나 리포지터리에 보관	9	2.3
학술지 발행기관이나 출판과 연계된 리포지터리에 보관	6	1.5
연구비지원기관의 리포지터리에 보관	6	1.5
기타	5	1.3
합계	389	100.0

(*복수응답 허용)

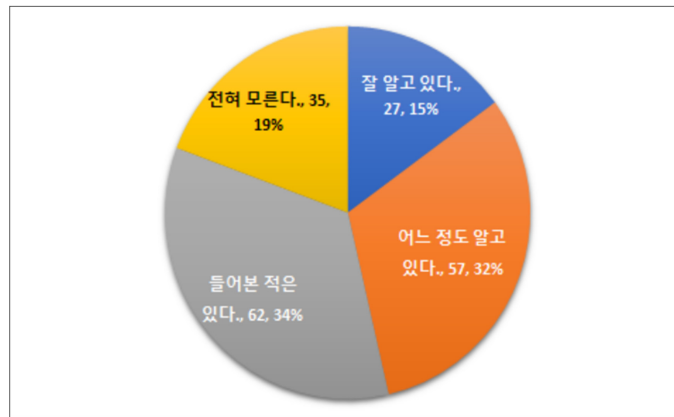
그러나 연구 분야 내의 리포지터리나 학술지 및 연구비 지원기관에서 운영하는 리포지터리에 데이터를 보관한다는 응답은 소수인 것으로 나타났다. 기타 응답으로 한국과학기술정보연구원에서 운영하는 국가연구데이터플랫폼인 DataON, 데이터를 생산한 가속기 연구소, 서비스 관리 시스템, 소속기관 홈페이지 등이 제시되었다.

3. 데이터 논문 출판 경험 정도와 선택 기준 및 장단점 인식

본 연구에서 제시한 두 번째 연구 문제는 과학기술분야 연구자들의 데이터 논문 출판 경험의 정도와 데이터 저널의 선택 기준 및 장단점을 어떻게 인식하고 있는지에 관한 것이다. 이를 위해 먼저 연구데이터 출판의 대표적인 유형인 데이터 저널에 대한 인식 여부를 확인하고 이를 인식하고 있는 응답자들을 대상으로 데이터 저널에서의 데이터 논문 게재 경험 유무를 확인하였다. 이후 데이터 저널 선택 시 고려하는 기준과 장단점에 관하여 질의하였다.

데이터를 생산·수집한 경험은 있으나 데이터를 공유해 본 적은 없는 37명을 제외한 181명을 대상으로 데이터 논문을 출판하는 데이터 저널에 대해 어느 정도 알고 있는지를 질문하였고 그 결과는 〈그림 2〉와 같다. 181명 중 62명(34%)이 데이터 저널에 대해 들어본 적이 있다고 응답하

였고 어느 정도 알고 있다는 응답자는 57명(32%)인 것으로 나타났다. 데이터 저널에 대해 잘 알고 있다는 응답자는 27명(15%)이었다. 데이터 저널에 대해 전혀 모른다고 응답한 35명을 제외하고 데이터 저널을 인식하고 있는 146명을 대상으로 데이터 저널에 데이터 논문을 출판한 경험이 있는지를 질문하였고 이 중 36명(24.7%)이 경험이 있다고 응답하였다(〈그림 2〉 참조).



〈그림 2〉 데이터 저널 인식 정도

데이터 저널에 논문을 출판한 적이 있다고 응답한 36명에게 데이터 논문 게재 편수를 기재하도록 했고 모두 응답하였다. 이 중에서 100편이라고 기재한 응답이 있어 오류라고 판단하여 제외하였고 나머지 35명의 데이터 논문 게재 편수를 〈표 5〉에 제시하였다.

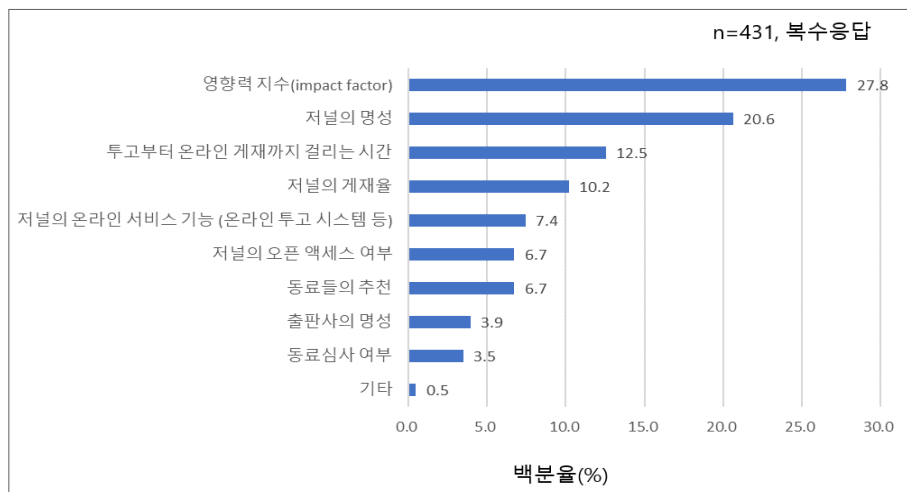
응답자 35명 중 데이터 논문 1편을 출판하였다는 응답이 13명(37.1%)으로 가장 많았고 2편과 3편이 각각 5명(14.3%)과 7명(20.0%)로 나타나 대다수가 1편~3편의 데이터 논문을 데이터 저널에 게재하였음을 알 수 있다. 데이터 논문을 10편 게재하였다는 응답자도 3명이었고 가장 많은 게재 편수는 12편(1명)인 것으로 나타났다.

〈표 5〉 데이터 논문 게재 편수

구분	빈도	비율(%)
1편	13	37.1
2편	5	14.3
3편	7	20.0
4편	4	11.4
5편	2	5.7
10편	3	8.6
12편	1	2.9
합계	35	100.0

가. 데이터 저널 선택 기준

데이터 저널을 인식하고 있는 146명을 대상으로 데이터 저널에 투고를 고려할 때의 선택 기준과 데이터 저널에서의 데이터 논문 출판의 장점과 한계에 어느 정도 동의하는지를 질문하였고 응답 결과는 다음과 같다. <그림 3>에 따르면 데이터 저널의 영향력 지수(impact factor)를 선택 기준으로 고려한다는 응답(120, 27.8%)이 가장 많았으며, 저널의 명성(89, 20.6%)은 두 번째로 응답 비율이 높은 기준이었다. 투고부터 온라인 게재까지 걸리는 시간(54, 12.5%)과 저널의 게재율(44, 10.2%)도 상당수의 응답자가 데이터 저널을 선택할 때 고려한다는 것을 알 수 있었다. 그러나 데이터 저널에서의 동료심사 여부를 선택 기준으로 응답한 경우는 15명(3.5%)에 불과하여 제시된 기준 가운데 가장 응답 비율이 낮은 것으로 나타났다. 기타 응답으로 가장 관련성이 높은 주제의 데이터 저널을 선택한다는 의견이 제시되었다.



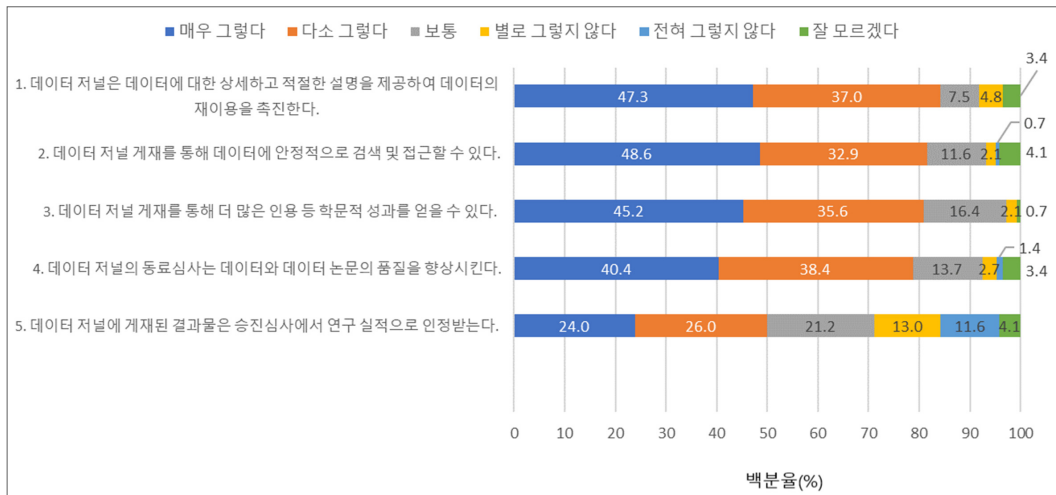
<그림 3> 데이터 저널 선택 기준

나. 데이터 저널의 장점에 대한 인식

데이터 저널을 알고 있는 146명에게 데이터 저널의 장점에 대해 어느 정도 동의하는지를 5점 척도로 응답하게 하였다. 그리고 각각의 장점에 대해 '매우 그렇다'와 '다소 그렇다'의 응답 비율이 높은 순서대로 <그림 4>에 정리하였다.

장점 1번 문항(데이터에 대한 상세한 설명을 제공하여 데이터 재이용 촉진)에 동의한다는 응답 비율이 가장 높았으며 장점 2번 문항(데이터에 안정적으로 검색 및 접근)에 동의하는 비율이 두 번째로 높았다. <표 6>에 따르면 데이터 저널의 장점을 묻는 1번 문항과 2번 문항의 평균은 5점 만점을 기준으로 거의 같은 것으로 나타났다. 그 외 장점 3번 문항(더 많은 인용 등 학문적

성과 획득)과 장점 4번 문항(동료심사를 통해 데이터와 데이터 논문의 품질 향상)에 대해서도 동의하는 비율이 80% 내외인 것으로 나타났다(〈그림 4〉 참조). 이들 문항의 평균 역시 4점을 상회하는 것으로 나타나 해당 문항들에 대해서도 응답자들이 대체로 동의함을 알 수 있었다.



〈그림 4〉 데이터 저널의 장점에 대한 응답자 인식 정도

〈표 6〉 데이터 저널 장점별 평균 및 표준편차

구분	응답자 수*	평균	표준편차
1. 데이터 저널은 데이터에 대한 상세하고 적절한 설명을 제공하여 데이터의 재이용을 촉진한다.	141	4.31	0.82
2. 데이터 저널 게재를 통해 데이터에 안정적으로 검색 및 접근할 수 있다.	140	4.32	0.82
3. 데이터 저널 게재를 통해 더 많은 인용 등 학문적 성과를 얻을 수 있다.	145	4.25	0.80
4. 데이터 저널의 동료심사는 데이터와 데이터 논문의 품질을 향상시킨다.	141	4.18	0.88
5. 데이터 저널에 게재된 결과물은 승진심사에서 연구 실적으로 인정받는다.	140	3.39	1.32

(* '잘 모르겠다'를 선택한 응답자를 전체 응답자(n=146)에서 제외)

그러나 장점 5번 문항(데이터 저널에 게재된 결과물은 승진심사에서 연구 실적으로 인정)에 대해 동의하는 응답 비율은 50%에 불과하였으며 이에 동의하지 않는 응답 비율도 24.6%로 나타났다. 따라서 5번 문항은 다른 문항들에 비해 동의/비동의 수준에 있어 많은 차이를 보이고 있다(〈그림 4〉 참조). 5번 문항 응답 평균도 5점 만점에 3.39로 동의 정도가 보통 수준인 것으로 나타났다(〈표 6〉 참조). 데이터 출판의 경험 유무에 따른 5번 문항 응답의 차이를 보기 위해 데이터 출판 경험이 있는 36명의 5번 문항에 대한 응답 분포를 살펴보았다. 그 결과 '매우 그렇다' 11명, '다소 그렇다' 3명으로 나타나 36명 중 14명(38.9%)만이 데이터 논문이 승진심사에서 실적으로

인정된다는 것에 동의하였다. 그 외 '보통' 9명(25.0%), '별로 그렇지 않다' 7명(19.4%), '전혀 그렇지 않다' 5명(13.9%), '잘 모르겠다' 1명(2.8%)이었다. 이처럼 5번 문항에 동의하지 않는 응답자도 33.3%로 나타나 동의하지 않는 비율도 상당함을 알 수 있었다. 또한 장점 5번 문항에 대해 데이터 저널에 데이터 논문을 출판한 응답자와 그렇지 않은 응답자 집단으로 나누어 비모수 검정 방법인 Mann-Whitney 검정을 통해 살펴본 결과 두 집단 간 5번 문항 응답에서의 유의한 차이는 없는 것으로 나타났다(<표 7> 참조). 장점 5번 문항은 각 표본이 정규 분포를 따르고 있지 않았으므로 독립 표본 t 검정 대신 독립 표본 2개에 대한 서열 척도 값에 순위를 매겨 비교하는 비모수 검정 방식인 Mann-Whitney 검정을 사용하였다.

<표 7> 데이터 논문 출판 경험 유무에 따른 데이터 저널 장점 인식 차이 (Mann-Whitney 검정 결과)

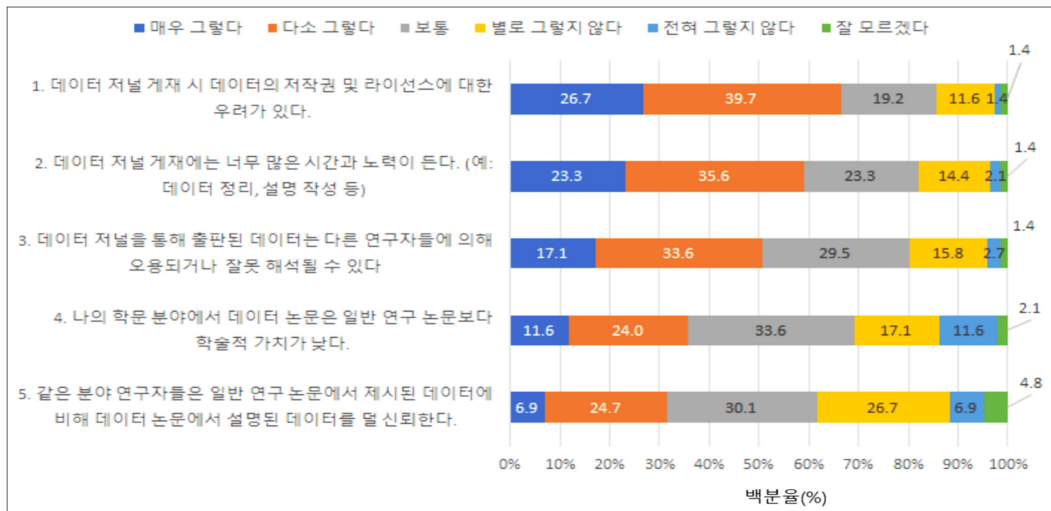
	데이터 논문 출판 경험	응답자 수	순위 합	평균 순위 합	z	p
1. 데이터 저널은 데이터에 대한 상세하고 적절한 설명을 제공하여 데이터의 재이용을 촉진한다.	있음	35	2651	75.7	-0.87	0.383
	없음	106	7360	69.4		
2. 데이터 저널 게재를 통해 데이터에 안정적으로 검색 및 접근할 수 있다.	있음	35	2269	64.8	1.05	0.294
	없음	105	7601	72.4		
3. 데이터 저널 게재를 통해 더 많은 인용 등 학문적 성과를 얻을 수 있다.	있음	36	2577	71.6	0.25	0.800
	없음	109	8008	73.5		
4. 데이터 저널의 동료심사는 데이터와 데이터 논문의 품질을 향상시킨다.	있음	36	2501	69.5	0.28	0.779
	없음	105	7510	71.5		
5. 데이터 저널에 게재된 결과물은 승진심사에서 연구 실적으로 인정받는다.	있음	35	2315	66.1	0.76	0.450
	없음	105	7556	72.0		

따라서 다수의 선행연구에서 데이터 논문 출판을 통해 데이터 공유 활동을 연구 실적으로 인정 받을 수 있다는 것을 장점으로 강조한 것과는 대조적으로 응답자들이 이에 동의하는 정도는 높지 않음을 알 수 있었다. 또한 데이터 논문을 출판한 적이 있는 응답자 중에서도 데이터 논문이 연구 실적으로 인정받는다는데 동의하지 않는 비율이 상당히 높았고, 이들의 경험상 데이터 논문이 연구 성과물로 인정되지 못했을 가능성도 생각해볼 수 있었다. 이 밖에 장점 1번~4번 문항에 대해서도 데이터 논문 출판 경험 유무에 따른 인식 차이가 있는지를 알아보기 위해 Mann-Whitney 검정을 수행하였고 모두 유의한 차이가 없는 것으로 나타났다(<표 7> 참조).

다. 데이터 저널의 한계에 대한 인식

데이터 저널의 한계에 대한 응답자들의 동의 정도를 각 문항에 '매우 그렇다'와 '다소 그렇다'의 5점 척도로 응답한 비율이 높은 순서대로 정리하였다(<그림 5> 참조). 한계 1번 문항(데이터의 저작권 및 라이선스에 대한 우려)에 동의하는 응답 비율이 66.4%로 가장 높았으며 한계 2번 문항

(데이터 저널 게재에는 많은 시간과 노력 소요)에 동의하는 응답 비율이 58.9%로 두 번째로 높았다. 그러나 데이터 저널의 장점에 대해 응답자들이 동의한 정도에 비하면 그 비율이 낮음을 알 수 있다.



〈그림 5〉 데이터 저널의 한계에 대한 응답자 인식 정도

〈표 8〉에 따르면 데이터 저널의 한계에 대한 문항별 평균 역시 4점 미만으로 나타나 응답자들은 보통 수준에서 데이터 저널의 한계에 동의하고 있음을 보여주었다. 특히 한계 5번 문항(같은 분야 연구자들이 일반 연구 논문의 데이터에 비해 데이터 논문에서 설명된 데이터를 덜 신뢰함)에 대해서는 동의하는 응답자 비율(31.6%)보다 동의하지 않는 응답자 비율(33.6%)이 더 높았다(〈그림 5〉 참조). 〈표 8〉에서도 한계 5번 문항 응답 평균은 2.98로 3점 미만으로 나타나 응답자들이 그다지 동의하지 않는다는 것을 알 수 있다. 한계 5번 문항에 대해 ‘잘 모르겠다’는 응답도 7명(4.8%)으로 가장 많았다(〈그림 5〉 참조). 또한 한계 4번 문항(데이터 논문은 일반 연구 논문보다 학술적 가치가 낮음)에 대해서도 동의하는 비율이 35.6%로 동의하지 않는다는 비율(28.7%)과 큰 차이가 없었으며 응답 평균도 3.07로 낮은 수준이었다(〈그림 5〉, 〈표 8〉 참조).

〈표 8〉 데이터 저널 한계별 평균 및 표준편차

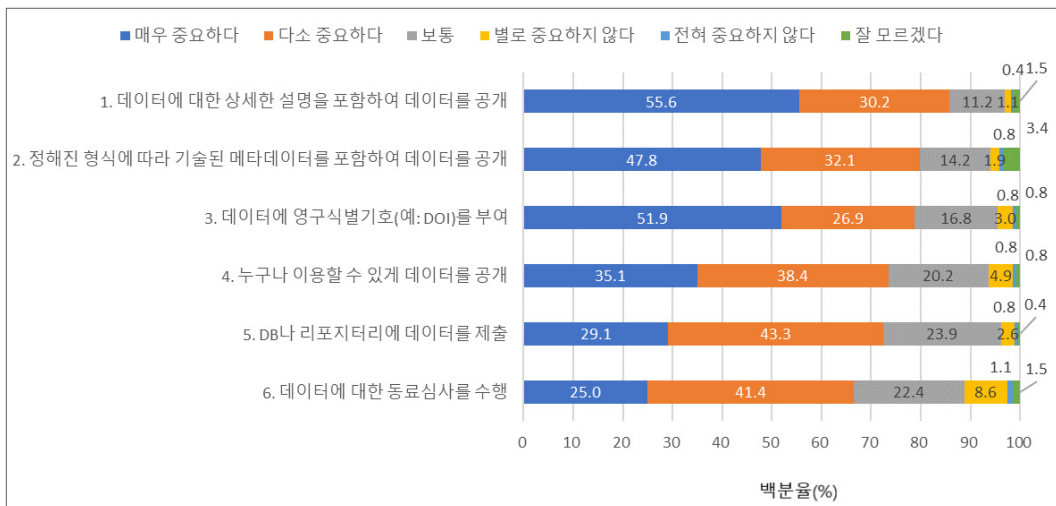
문항	응답자 수*	평균	표준편차
1. 데이터 저널 게재 시 데이터의 저작권 및 라이선스에 대한 우려가 있다.	144	3.80	1.01
2. 데이터 저널 게재에는 너무 많은 시간과 노력이 든다. (예: 데이터 정리, 설명 작성 등)	144	3.65	1.06
3. 데이터 저널을 통해 출판된 데이터는 다른 연구자들에 의해 오용되거나 잘못 해석될 수 있다	144	3.47	1.04
4. 나의 학문 분야에서 데이터 논문은 일반 연구논문보다 학술적 가치가 낮다.	143	3.07	1.17
5. 같은 분야 연구자들은 일반 연구 논문에서 제시된 데이터에 비해 데이터 논문에서 설명된 데이터를 덜 신뢰한다.	139	2.98	1.06

(*‘잘 모르겠다’를 선택한 응답자를 전체 응답자(n=146)에서 제외)

이처럼 데이터 저널의 한계에 대해 응답자들은 데이터 저널의 장점만큼 높은 수준으로 동의하지는 않음을 알 수 있다. 특히 일반 학술 논문과 비교하여 데이터 논문의 학술적 가치를 낮게 평가하거나 데이터 논문을 통해 공유된 데이터를 일반 학술 논문의 기반이 되는 데이터보다 덜 신뢰하는 경향은 낮은 것으로 보인다. 데이터 논문 출판 경험 유무에 따른 데이터 저널 한계에 대한 인식 차이를 조사하기 위해 문항별로 각 표본의 정규성 및 등분산 여부를 조사하였고 한계 1번 문항을 제외하고는 모두 정규 분포와 등분산 가정을 만족하였다. 이에 따라 한계 1번 문항에 대해서는 Mann-Whitney 검정을, 나머지 문항에는 독립 표본 t 검정을 수행하였다. 그 결과 독립 표본 t 검정을 수행한 한계 2번 문항(데이터 저널 게재에는 너무 많은 시간과 노력이 든다)에 대해 데이터 논문 출판 경험이 있는 응답자 집단(M=3.28, SD=0.198)과 경험이 없는 응답자 집단(M=3.77, SD=0.095) 간 응답의 평균($t(142)=2.45, p=0.016$)에서 통계적으로 유의한 차이가 있는 것으로 나타났다. 이로 인해 데이터 논문 게재 경험이 없는 응답자들이 데이터 논문 게재에 드는 시간과 노력을 한계로 인식하는 정도가 경험이 있는 응답자에 비해 높다는 것을 알 수 있었다.

4. 향후 데이터 저널 투고 의향에 따른 데이터 출판 속성 중요도 인식

본 연구의 세 번째 연구 문제는 향후 데이터 저널 투고 의향 여부에 따라 데이터 출판 속성의 중요도 인식에 차이가 있는지에 관한 것이다. 우선 설문에 참여한 모든 응답자 268명을 대상으로 데이터 출판의 속성에 대하여 어느 정도 중요하게 생각하는지를 5점 척도로 질문한 결과를 '매우 중요하다'와 '다소 중요하다'의 응답 비율이 높은 순서대로 정리하였다(〈그림 6〉 참조).



〈그림 6〉 데이터 출판의 속성에 대한 인식

〈그림 6〉에 따르면 속성 1번 문항(데이터에 대한 상세한 설명을 포함하여 데이터 공개)에 대해 중요하다고 생각하는 응답 비율이 85.8%로 가장 높았고 평균 역시 가장 높은 값인 4.42로 나타났다(〈표 9〉 참조). 속성 2번 문항(정해진 형식에 따라 기술된 메타데이터를 포함하여 데이터를 공개)에 대해 중요하다고 인식한 비율이 79.9%로 두 번째로 높았으며 평균도 2위인 4.29이었다(〈표 9〉 참조). 속성 3번 문항(데이터에 영구식별기호 부여) 역시 중요하다고 인식한 비율이 78.8%이었고 평균 순위는 3위(4.27)로 나타났다. 이를 통해 응답자들은 데이터에 대한 설명과 메타데이터 및 영구식별기호 제공을 데이터 출판의 중요한 속성으로 인식함을 알 수 있었다. 속성 4번 문항(누구나 이용할 수 있게 데이터 공개)도 상위 3개 문항에 비해 평균(4.03)이 다소 낮았으나 대체로 중요하다고 인식하는 것으로 나타났다. 그러나 속성 5번 문항(DB나 리포지터리에 데이터 제출)과 속성 6번 문항(데이터에 대한 동료심사 수행)에 대한 중요도 인식 비율은 각각 72.4%와 66.4%로 나타났으며 해당 문항들의 평균 역시 4점 미만으로 나타났다. 따라서 리포지터리에 데이터를 기탁하는 것과 데이터 동료심사를 수행하는 것을 데이터 출판의 중요한 속성으로 인식하는 정도는 다소 낮음을 알 수 있었다.

〈표 9〉 데이터 출판의 속성 문항별 평균 및 표준편차

문항 구분	응답자 수*	평균	표준편차
1. 데이터에 대한 상세한 설명을 포함하여 데이터를 공개	264	4.42	0.77
2. 정해진 형식에 따라 기술된 메타데이터를 포함하여 데이터를 공개	259	4.29	0.84
3. 데이터에 영구식별기호(예: DOI)를 부여	266	4.27	0.90
4. 누구나 이용할 수 있게 데이터를 공개	266	4.03	0.91
5. DB나 리포지터리에 데이터를 제출	267	3.98	0.84
6. 데이터에 대한 동료심사를 수행	264	3.82	0.95

(* '잘 모르겠다'를 선택한 응답자를 전체 응답자(n=268)에서 제외)

그리고 데이터 출판의 속성별 중요도 인식 수준이 향후 데이터 저널에 투고할 의향에 따라 차이가 있는지를 조사하였다. 향후 데이터 저널에 투고할 의향이 있다는 응답자는 268명 중 153명(67.1%)이었고 투고 의향이 없다는 응답자가 26명(9.7%), 모르겠다는 응답자가 89명(33.2%)으로 나타났다. 이를 다시 투고 의향이 있는 응답자 집단(153명, 67.1%)과 투고 의향이 없거나 불확실한 응답자 집단(117명, 42.9%)으로 구분하여 데이터 출판 속성별 인식 차이를 조사하였다. 데이터 출판 속성 문항별로 정규 분포를 따르는지를 살펴본 결과 모든 문항이 정규성을 만족하지 못하는 것으로 나타나 비모수 검정 방법인 Mann-Whitney 검정을 수행하였다. 그 결과 향후 데이터 저널 투고 의향이 있는 응답자 집단과 투고 의향이 없거나 불확실한 집단 간 모든 데이터 출판 속성 문항별 중요도 인식에서 통계적으로 유의한 차이가 존재하는 것으로 나타났다(〈표 10〉 참조).

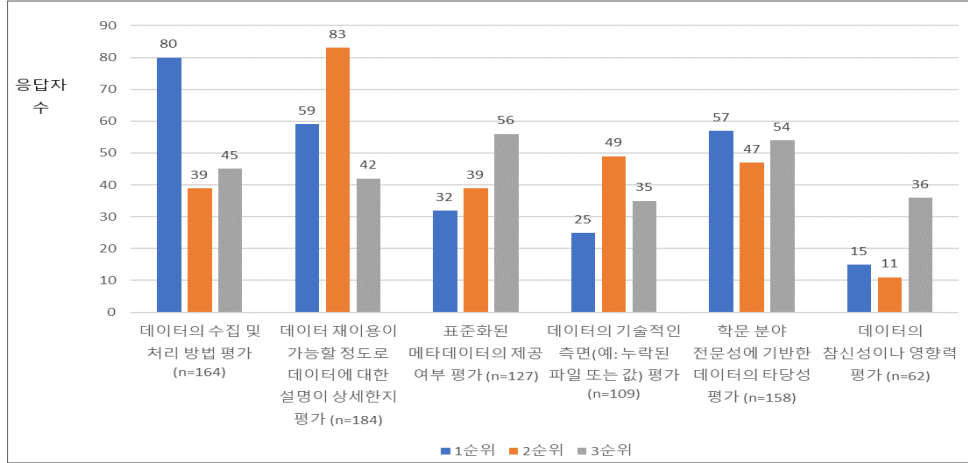
〈표 10〉에 따르면 모든 문항에서 향후 데이터 저널 투고 의향이 있는 집단의 평균 순위 합이 투고 의향이 없거나 불확실한 집단의 평균 순위 합보다 높은 것으로 나타났다. 이는 향후 데이터 저널 투고 의향이 있는 집단에서 데이터 출판 속성들을 더 중요하게 인식하고 있다는 것을 의미한다.

〈표 10〉 향후 데이터 저널 투고 의향에 따른 데이터 출판 속성의 중요도 인식 차이 (Mann-Whitney 검정 결과)

문항 구분	향후 데이터 저널 투고 의향	응답자수*	순위 합	평균 순위 합	z	p
1. 데이터에 대한 상세한 설명을 포함하여 데이터를 공개	있음	149	21711	145.7	-3.60	0.000
	없거나 불확실함	115	13269	115.4		
2. 정해진 형식에 따라 기술된 메타데이터를 포함하여 데이터를 공개	있음	147	21093	143.5	-3.62	0.000
	없거나 불확실함	112	12577	112.3		
3. 데이터에 영구식별기호(예: DOI)를 부여	있음	152	22178	145.9	-3.33	0.001
	없거나 불확실함	114	13334	117.0		
4. 누구나 이용할 수 있게 데이터를 공개	있음	152	22803	150.0	-4.29	0.000
	없거나 불확실함	114	12708	111.5		
5. DB나 리포지터리에 데이터를 제출	있음	152	22817	150.1	-4.18	0.000
	없거나 불확실함	115	12961	112.7		
6. 데이터에 대한 동료심사를 수행	있음	150	21251	141.7	-2.36	0.018
	없거나 불확실함	114	13730	120.4		

(* '잘 모르겠다'를 선택한 응답자를 전체 응답자(n=268)에서 제외)

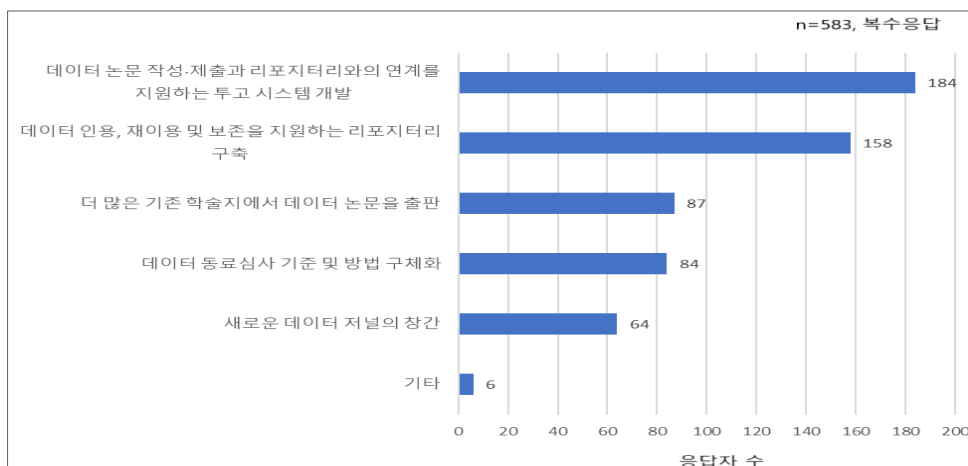
데이터 동료심사 기준에 대해 1순위에서 3순위까지 응답하게 한 질문에서 〈그림 7〉에 제시된 바와 같이 '데이터의 수집 및 방법 평가'를 1순위로 선택한 응답자가 80명으로 가장 많았다. 1순위를 선택한 응답자 수를 비교하였을 때 두 번째로 많은 59명이 '데이터의 재이용이 가능할 정도로 데이터에 대한 설명이 상세한지 평가'하는 것을 1순위로 선택하였다. '학문 분야 전문성에 기반한 데이터의 타당성 평가'를 1순위로 선택한 응답자는 57명으로 1순위 응답 중 세 번째로 많은 것으로 나타났다. 2순위 응답자 중에 가장 많은 83명이 '데이터 재이용이 가능할 정도로 데이터에 대한 설명이 상세한지 평가'하는 것을 선택하였다. 해당 기준은 순위를 매긴 응답자가 184명으로 가장 많았으며 1순위와 2순위로 선택한 빈도 역시 높았다. '학문 분야 전문성에 기반한 데이터의 타당성 평가' 역시 순위를 매긴 응답자 총수가 158명으로 '데이터의 수집 및 처리 방법 평가'에 이어 세 번째로 많았으며 1순위에서 3순위까지 비교적 고른 응답 분포를 보였다. 따라서 '데이터의 수집 및 처리 방법 평가'와 '데이터 재이용이 가능할 정도로 데이터에 대한 설명이 상세한지 평가', '학문 분야 전문성에 기반한 데이터의 타당성 평가'의 기준들이 데이터 동료심사에서 활용될 수 있다. '데이터의 참신성이나 영향력 평가'의 경우 순위를 매긴 응답자 총수는 62명으로 가장 적었으며 데이터 동료심사 기준으로 크게 고려되지 않는 것을 알 수 있었다.



〈그림 7〉 데이터 동료심사 기준

5. 데이터 출판 촉진 요인

본 연구의 모든 설문 응답자 268명을 대상으로 데이터 출판을 촉진하는 데 중요하다고 생각하는 요인을 질문하였으며 응답 결과는 〈그림 8〉과 같다. ‘데이터 논문 작성 제출과 리포지터리와의 연계를 지원하는 투고 시스템 개발’(184, 31.6%)이 가장 높은 응답 비율을 보인 요인이었다. 다음으로 높은 응답 비율을 나타낸 요인은 ‘데이터 인용, 재이용 및 보존을 지원하는 리포지터리 구축’(158, 27.1%)이었다. 이를 통해 데이터 출판을 효과적으로 지원하는 투고 시스템과 리포지터리 개발이 데이터 출판 활성화를 위한 과제로 수행될 필요가 있음을 확인할 수 있었다.



〈그림 8〉 데이터 출판 촉진 요인

그 외 기타 의견으로 데이터 출판 결과를 인사고과 등 연구 실적 평가에 반영하는 체계나 보상 체계 마련, 저작권 보호 방안 마련 및 데이터의 수정, 가공 및 공개 여부 추적이 가능한 버전 관리 시스템 개발이 제시되었다.

한편, 데이터 저널에 투고할 의향이 있는 153명을 대상으로 투고 예정 시기를 질문하였고 1년 초과~3년 이내에 투고하겠다는 응답자가 68명(44.4%)으로 가장 많았다(〈표 11〉 참조). 1년 이내가 36명(23.5%)이었고 투고 계획은 있으나 시기는 미정이라는 응답도 30명(19.6%)인 것으로 나타났다. 응답자 다수가 3년 이내 투고할 계획이라고 응답하였으므로 이들 잠재적 수요자의 요구를 지원하는 서비스와 시스템 개발을 고려할 필요가 있음을 시사하고 있다.

〈표 11〉 데이터 저널 투고 예정 시기

문항구분	빈도	비율(%)
지금 당장	6	3.9
1년 이내	36	23.5
1년 초과~3년 이내	68	44.4
3년 초과~6년 이내	12	7.8
언젠가 투고할 계획이나 시기는 미정	30	19.6
기타	1	0.7
합계	153	100.0

마지막으로 데이터 출판에 대한 의견을 자유롭게 기술하는 질문에서 180명(67.2%)이 유효한 응답을 제시하였고 이를 통해 나타난 주제 및 응답 예시를 〈표 12〉에 제시하였다. 응답자 다수가 데이터 출판의 결과물이 연구 실적으로 반영되어야 데이터 출판이 활성화될 수 있다는 것을 강조하였다. 이는 선행연구에서도 데이터 출판의 동기로 강조되는 측면(Huang & Jeng, 2022; Lee & Kim, 2021)이며 이를 지원하는 방안에 대한 논의가 필요함을 알 수 있다. 이와 더불어 데이터 품질을 보장하는 것이 중요하다는 의견이 다수 제시되었으며 동료심사를 언급한 의견도 있었으나 기존의 동료심사 외에 자동화된 검증 방식 등 새로운 방식에 대한 의견도 존재하였다. 데이터의 활용 가능성 역시 데이터 출판 결과물을 연구 실적에 반영하는 것만큼 중요하다고 언급한 응답자들이 있었다. 이 외에도 데이터 논문 출판 플랫폼 개발, 효과적인 리포지터리 서비스 제공 및 데이터 출판을 위한 제도 기반 마련에 대한 의견이 제시되었다. 데이터 출판과 관련된 우려 사항도 자주 언급되었는데 저작권, 소유권 및 개인정보 보호 문제, 데이터 오용 및 데이터 공유로 인해 논문 출판의 기회를 잃을 수 있다는 인식, 데이터 출판이 중복출판이 아닌지에 대한 우려 및 대량의 데이터로 인한 데이터 출판의 어려움 등이 제시되었다(〈표 12〉 참조).

〈표 12〉 자유응답 예시

구분	응답 예시
데이터 출판에 대한 교육과 안내 필요성	“데이터 출판의 중요성은 인지하지만 정확한 개념과 시스템을 확인하기 어렵다. 이러한 교육이 있으면 좋겠음”(S4, 40대, 생명과학)
	“데이터 출판에 대한 기본적인 정보 안내가 필요한 거 같습니다.”(S97, 30대, 생명과학)
데이터 출판의 결과물이 연구 실적으로 반영 필요	“연구성과가 개인의 인사고과나 인센티브에 적극적으로 반영되지 않는다면, 굳이 출판을 해야되나에 대한 의문을 해결하기 어려울 것 같습니다. 적극적으로 반영이 된다면, 기존에 연구직에 계신 분들이 평소에 두고하시던 저널에 플러스알파의 개념으로 데이터 저널 투고도 긍정적으로 바라보실 수 있을 것 같습니다.”(S143, 30대, 기계)
	“데이터 출판 또는 데이터 논문 작성을 통한 데이터 저널 게재 시, 연구자들이 실제 유효하다고 느낄 수 있는 보상책이 필요합니다. DOI 부여를 통해 인용을 가능하게 하는 것 외에도, 연구자들의 소속기관에서 논문/특허 성과와 동등한 실적으로 인정해주는 제도 개선도 꼭 필요합니다.”(S158, 30대, 화공)
	“데이터 저널이 정식 학술 논문으로 평가받아야 더 많은 사람의 투고 및 이용이 가능할 것 같습니다.”(S237, 40대, 지구과학)
데이터 품질 관리 방안 필요	“동료심사를 통한 데이터의 신뢰성 확보 필요”(S261, 40대, 생명과학)
	“기존의 동료평가 방식이 아닌 자동화된 검증 방식이 필요”(S117, 40대, 기계)
	“출판된 다른 데이터를 이용하고자 할 때 가장 중요한 것은 이용 가능한, 타당한, 신뢰가 있는 데이터 인지의 여부이다. 이게 보증되어야 출판된 데이터를 다른 연구자가 이용할 수 있을 것이다.”(S145, 40대, 경제/경영)
데이터의 활용 가능성 보장	“데이터 학술지에 대한 논문이 충분한 실적으로 인정되는지, 공유한 데이터가 적절하게 사용되는지, 두 부분이 중요한 것 같습니다.”(S246, 20대, 지구과학)
	“데이터 생산자에 대한 보상을 우선적으로 고려해야 하며, 데이터 자체의 신뢰도 보다는 데이터의 활용성을 높이기 위해서 필요한 방법에 대한 고민이 필요.”(S254, 50대, 지구과학)
데이터 논문 출판 플랫폼 필요	“활용이 용이한 데이터 논문 출판 플랫폼을 구축하여 제공해 주시기 바랍니다.”(S139, 50대, 정보/통신)
	“데이터 출판은 접근성이 좋은 플랫폼이 선제”(S225, 40대, 기계)
효과적인 리포지터리 서비스 필요	“리포지터리 데이터 등록은 매우 장려할만한 일이나 수고스러움을 최대한 덜 수 있어야 활발히 이루어질 것으로 보임”(S6, 30대, 지구과학)
	“국내 학술지도 데이터 논문을 주요한 논문 유형으로 채택하도록 노력함과 아울러 연계할 공동의 데이터 리포지터리 서비스가 제공되어야 할 것이다.”(S224, 60대 이상, 미디어/커뮤니케이션/문헌정보)
데이터 출판을 위한 제도/규정 필요	“데이터 출판에 대한 관련 규정이 필요함”(S78, 50대, 보건의료)
	“데이터 출판이 활성화될 수 있는 제도 기반 마련 필요”(S268, 50대, 지구과학)
데이터 저작권 및 소유권 문제	“연구데이터의 공공 활용을 위해 데이터 출판은 필수적이라고 생각되나, 저작권 보호 방안 확립이 필요.”(S40, 40대, 지구과학)
	“재분석 자료 소유권 관련 이슈 문제 해결 방안 모색 필요”(S179, 40대, 건설/교통)
개인정보 보호 문제	“연구를 위한 데이터 활용과 공개할 때의 저작권이나 개인정보 등 책임져야 할 부담이 큼니다. 개개인에게 이런 책임을 묻는 구조라면 연구자 대상으로도 데이터 공유가 쉽지 않을 것 같습니다.”(S171, 40대, 정보/통신)
	“공유의 의미는 좋으나, 의미가 변질되거나 오용되는 것에 근심이 된다.”(S138, 20대, 생명과학)
데이터 오용 및 논문 출판 기회 상실 우려	“죽은 데이터를 살릴 수 있는 기회이지만, 내가 쓸 수 있는 논문을 뺏긴다는 기분도 들 수 있음”(S140, 30대, 기계)
	“중복출판에 대한 우려는 없을지 두 번 일하게 하는 거 아닌지 모르겠어요.”(S219, 30대, 보건의료)
대량의 데이터로 인한 출판의 어려움	“어느 정도의 데이터를 가지고 출판을 해야 할지 잘 모르겠다 (데이터의 양)”(S25, 30대, 지리/지역/관광)
	“데이터 출판은 연구자 입장에서 불가능함. - 사이즈 (연간 100PB)가 문제 - 실험에 참여한 연구자만 실험 데이터를 활용할 수 있다는 규정”(S45, 50대, 물리학)

V. 결론 및 시사점

본 연구의 설문조사 응답자들은 주로 정부출연연구기관 소속의 연구자들로 남성 및 박사학위 소지자이면서 연령별로는 30대와 40대의 젊은 연구자가 높았다. 연구 분야로는 정보/통신, 지구과학, 생명과학, 보건의료 분야에 속한 응답자들의 비율이 높았으며 총 24개 연구 분야에 응답자들이 분포되었다.

첫 번째 연구 문제인 연구데이터 유형 및 보관 장소와 관련하여 총 268명의 응답자 중 데이터 생산·수집 경험이 있는 218명이 응답을 제시하였다. 실험 데이터, 모델 및 시뮬레이션 데이터, 설문 데이터를 생산한다는 응답 비율이 높았으며 약 73%의 응답자들이 생산·수집한 데이터의 대부분 혹은 전부를 보관하고 있었다. 연구자 개인의 온오프라인 저장 공간이 데이터 보관 장소로 가장 많이 활용되고 있었으며 연구팀 공용클라우드 혹은 소속기관의 리포지터리도 활용되고 있었다.

두 번째 연구 문제인 데이터 논문 출판 경험의 정도와 선택 기준 및 장단점에 대한 인식을 살펴보았을 때 데이터 논문을 출판하는 데이터 저널에 대해 들어보았거나 알고 있다는 응답자는 146명이었으며 이들 중 데이터 저널에 데이터 논문을 게재한 경험이 있는 응답자는 36명이었다. 이들을 대상으로 데이터 저널의 선택 기준을 조사한 결과 영향력 지수와 저널의 명성을 선택한 응답자의 비율이 높았으며 이는 일반적인 학술지 선택 기준과 유사함을 알 수 있었다. 또한 데이터 저널의 장점으로 데이터에 대한 상세한 설명 제공, 데이터로의 안정적인 접근, 인용 등 학문적 성과 획득 가능, 동료심사를 통한 품질 확인에 대해서는 동의하는 비율이 높았다. 그러나 데이터 논문이 승진심사에서 연구 실적으로 인정받는다든가 문항에서는 동의 비율이 절반 정도로 낮게 제시되었다. 데이터 논문을 게재한 경험이 있는 36명 중에서도 이에 동의하지 않는 응답자 비율이 약 33%, 동의하는 응답자 비율이 약 39%로 큰 차이가 없었다. 따라서 데이터 논문이 연구 실적으로 반영될 수 있도록 하는 제도 및 인식 개선이 필요함을 알 수 있었다. 데이터 저널의 한계에 대한 동의 비율은 장점에 대한 동의 비율보다는 낮았으며 보통 수준의 동의 수준을 보였다. 특히 학술지 연구논문과 비교하여 데이터 논문의 학술적 가치나 데이터의 신뢰성이 낮다고 인식하는 경향은 적은 것으로 나타났다. 데이터 저널의 한계 중 데이터 논문 게재에 많은 시간과 노력이 든다는 문항에 대해 데이터 논문 게재 경험이 있는 응답자와 그렇지 않은 응답자 간 인식 수준에 통계적으로 유의한 차이가 있음을 확인하였다. 데이터 논문 게재 경험이 있는 응답자는 해당 문항을 한계로 인식하는 정도가 경험이 없는 응답자보다는 낮은 것으로 나타났다.

데이터 출판의 속성에 대한 중요도 인식을 조사한 결과 데이터에 대한 상세한 설명 제공, 정해진 형식으로 메타데이터 기술, 영구식별기호 부여 및 데이터를 활용 가능하도록 공개하는 것에 동의하는 응답 비율이 높았다. 그러나 리포지터리에 데이터를 제출하거나 데이터의 동료심사 수행이 데이터 출판의 속성으로 중요하다는 인식은 상대적으로 낮았다. 이를 바탕으로 세 번째 연구

문제인 향후 데이터 저널 투고 의향 여부에 따른 데이터 출판 속성의 중요도 인식 차이를 분석한 결과 모든 속성에 대해 통계적으로 유의한 차이를 확인하였다. 향후 데이터 저널에 투고 의향이 있는 응답자 집단이 투고 의향이 없거나 불확실한 응답자 집단보다 데이터 출판 속성을 중요하게 인식하는 정도가 높은 것으로 나타났다.

이 밖에도 데이터 동료심사 기준으로 데이터의 수집 및 처리 방법 및 재이용이 가능할 정도로 상세한 데이터에 대한 설명이 제공되는지를 1순위로 선택한 응답자 비율이 높게 나타났다. 그러나 데이터의 참신성과 영향력을 동료심사 기준으로 고려하는 비율은 낮았다.

마지막으로 데이터 출판을 촉진하는 요인으로 응답자 다수가 데이터 논문 작성과 리포지터리 연계를 지원하는 투고 시스템 개발과 데이터 인용 및 재이용을 지원하는 리포지터리 개발을 제시하였다. 응답자 268명 중 153명이 향후 데이터 저널에 투고할 의향이 있다고 응답하였고 이 중 3년 이내에 데이터 저널에 투고를 고려한다는 응답자 비율이 153명 중 70% 이상으로 나타났다. 따라서 데이터 출판을 효율적으로 지원하는 논문투고 관리시스템과 리포지터리 개발이 주요 과제로 수행될 필요가 있었다. 그러나 데이터 논문을 연구 실적으로 인정받는다든 문항에 동의하지 않는 비율이 상당히 높았고 자유 응답 문항에서도 데이터 출판의 결과물이 연구실적으로 인정받을 수 있게 하는 체계 마련이 필요하다는 의견이 다수 제시되었다. 데이터 출판과 연구 평가 체계와의 연계를 위한 제도 및 인식 개선 노력이 지속적으로 이루어져야 하며 데이터의 저작권 및 소유권 문제에 대한 응답자의 우려 정도가 높았으므로 이에 대한 컨설팅 및 지원 서비스 제공도 고려되어야 할 것이다.

참 고 문 헌

- 이승은 (2020). 데이터 출판의 동기요인과 활성화 방안 연구. 석사학위논문, 이화여자대학교 일반대학원 문헌정보학과.
- 정영임, 권오석, 김기동, 김소형, 서태설, 김선대 (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>
- 한국사회과학자료원 (2021). 연구데이터 관리와 이용에 대한 연구자 인식조사. [미간행].
- Adamczak, B. (2021). Data journals and data papers in various research areas and scientific disciplines-bibliometric analysis based on incites. TASK Quarterly, 25, 433-471.
<https://doi.org/10.34808/tq2021/25.4/p>
- Austin, C. C., Bloom, T., Dallmeier-Tiessen, S., Khodiyar, V. K., Murphy, F., Nurnberger, A.,

- Raymond, L., Stockhause, M., Tedds, J., & Vardigan, M. (2017). Key components of data publishing: using current best practices to develop a reference model for data publishing. *International Journal on Digital Libraries*, 18(2), 77-92.
<https://doi.org/10.1007/s00799-016-0178-2>
- 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>
- Gorgolewski, K., Margulies, D. S., & Milham, M. P. (2013). Making data sharing count: a publication-based solution. *Frontiers in Neuroscience*, 7.
<https://doi.org/10.3389/fnins.2013.00009>
- Huang, P. P. & Jeng, W. (2022). Data paper as a reward? motivation, consideration, and perspective behind data paper submission. *Proceedings of the Association for Information Science and Technology*, 59(1), 437-441.
<https://doi.org/10.1002/pra2.648>
- Huang, X., Hawkins, B. A., & Qiao, G. (2013). Biodiversity data sharing: will peer-reviewed data papers work?. *BioScience*, 63(1), 5-6. <https://doi.org/10.1525/bio.2013.63.1.2>
- Kratz, J. E. & Strasser, C. (2014). Data publication consensus and controversies. *F1000Research*, 3(94). <https://doi.org/10.12688/f1000research.3979.3>
- Kratz, J. E. & Strasser, C. (2015). Researcher perspectives on publication and peer review of data. *PLoS One*, 10(2), e0117619. <https://doi.org/10.1371/journal.pone.0117619>
- 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>
- Lee, J. & Kim, J. (2021). Korean researchers' motivations for publishing in data journals and the usefulness of their data: a qualitative study. *Science Editing*, 8(2), 145-152.
<https://doi.org/10.6087/kcse.246>
- Office of Science and Technology Policy (2022). Memorandum for the heads of executive departments and agencies, August 25, 2022. Available:
<https://www.whitehouse.gov/wp-content/uploads/2022/08/08-2022-OSTP-Public-Access-Memo.pdf>
- Tenopir, C., Dalton, E. D., Allard, S., Frame, M., Pjesivac, I., Birch, B., Pollock, D., & Dorsett, K. (2015). Changes in data sharing and data reuse practices and perceptions

among scientists worldwide. PLOS ONE, 10(8), e0134826.
<https://doi.org/10.1371/journal.pone.0134826>

• 국한문 참고문헌의 영문 표기

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

- Jung, Youngim, Kwon, Ohseok, Kim, Kidong, Kim, Sohyeong, Seo, Tae-Sul, & Kim, Suntae (2020). 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>
- Korea Social Science Data Archive (2021). Survey on the management of the research data and use data. [Unpublished].
- Lee, Seungeun (2020). A Study on Motivations of Data Publishing and a Plan for its Promotion. Master's thesis, Ewha womans university. South Korea.