

연구데이터 메타데이터의 품질과 연구데이터플랫폼의 활성화의 관계에서 동기부여 요인의 매개효과 연구*

A Study on the Mediating Effect of Motivation Factors between the Quality of Research Data Metadata and the Activation of Research Data Platform

박 성 은 (Seong-Eun Park)**

목 차

- | | |
|------------------|----------|
| 1. 서 론 | 4. 분석 결과 |
| 2. 이론적 배경 및 선행연구 | 5. 결 론 |
| 3. 연구설계 | |

초 록

본 연구는 바이오 분야 연구데이터플랫폼인 K-BDS를 대상으로, 연구데이터 메타데이터의 품질이 연구데이터플랫폼의 활성화에 미치는 영향 및 이 관계에서 연구데이터플랫폼 이용에 관한 동기부여 요인의 매개효과를 밝히고자 하였다. 먼저 세 변인 간 구조적 관계를 구조방정식모형, 부트스트랩을 통해 분석하였으며 분석 결과, 연구자가 메타데이터의 품질에 대해 중요하다고 생각할수록 연구데이터플랫폼 이용의 동기부여 정도, 그리고 플랫폼의 활성화 의도가 높아지는 것으로 나타났다. 또한 동기부여 요인의 매개효과도 확인되었다. 추가적으로 각 변인의 하위요인간의 세부적인 구조를 회귀분석과 Sobel test를 통해 파악하였다. 그 결과 바이오 분야의 연구데이터 공유의 활성화를 위해서는 검색가능성을, 연구데이터 재이용의 활성화를 위해서는 발견가능성을 높이는 것이 가장 효과적이며, 인용가능성은 플랫폼의 활성화에 영향을 미치지 않는 것으로 나타났다. 따라서 플랫폼을 활성화하기 위해서는 우선적으로 메타데이터 품질을 향상시킴으로써 시스템적인 지원을 충분히 하는 것이 중요하며, 이를 통해 플랫폼에 대한 신뢰를 높이고 인용에 대한 혜택을 제도적으로 정착시켜 갈 필요가 있다는 시사점을 얻을 수 있다.

ABSTRACT

This study focuses on the impact of research data metadata quality evaluation index on the revitalization of K-BDS, a research data platform in the bio field, and examines the mediating effect of motivation factors for utilizing the platform. The investigation employs a structural equation model analysis and bootstrap analysis to explore the interrelationships among the three variables. The findings demonstrate that researchers who prioritize the quality of metadata display higher motivation to use the research data platform, leading to an intention to activate the platform. The study also confirms the mediating effect of motivation factors. Moreover, a comprehensive understanding of the sub-factors within each variable is attained through regression analysis and Sobel test. The results highlight that enhancing searchability is crucial to activate research data sharing in the bio field, while improving discoverability is vital for research data reuse. Interestingly, the study reveals that citationability does not significantly impact platform activation. As a conclusion, to foster platform activation, it is imperative to provide systematic support by enhancing metadata quality. This improvement can not only increase trust in the platform but also institutionally solidify the benefits of citation.

키워드: 연구데이터, 메타데이터, 바이오 연구데이터플랫폼, 국가 연구데이터플랫폼, 플랫폼 활성화

Research Data, Metadata, Bio Research Data Platform, National Research Data Platform, Platform Activation

* 본 논문은 저자의 2023년 성균관대학교 문헌정보학과 박사학위논문(지도교수 고영만) '연구데이터 메타데이터의 품질평가 지표가 연구데이터플랫폼의 활성화에 미치는 영향 및 동기부여 요인의 매개효과 연구'의 일부 내용을 축약·정리한 것임.

** 한국과학기술정보연구원(KISTI) 연구데이터공유센터 선임기술원

(pse3598@kisti.re.kr / ISNI 0000 0005 0591 0799)

논문접수일자: 2023년 7월 20일 최초심사일자: 2023년 8월 1일 게재확정일자: 2023년 8월 11일

한국문헌정보학회지, 57(3): 325-350, 2023. <http://dx.doi.org/10.4275/KSLIS.2023.57.3.325>

© Copyright © 2023 Korean Society for Library and Information Science

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.

1. 서론

1.1 연구의 필요성

데이터 중심 R&D 활성화로 데이터의 공유와 재이용이 학술 커뮤니케이션의 중요한 과제가 됨에 따라(Cragin & Shankar, 2006), 국가 R&D를 통해 축적되는 연구데이터를 다양한 연구자가 공유·활용할 수 있는 체계 구축의 필요성이 제기되어왔다. 이에 정부 차원에서 연구데이터의 관리를 위한 법과 제도를 개선하고, 연구데이터플랫폼 등의 인프라를 구축하는 정책을 추진하고 있다.

한국과학기술정보연구원(Korea Institute of Science and Technology Information, KISTI)에서 구축하여 2020년 1월부터 운영하고 있는 ‘국가연구데이터플랫폼(이하 DataON)’ 또한 이러한 인프라의 하나이다. ‘DataON’은 국가센터로서 분야별 전문센터들이 확보하고 연계한 연구데이터를 취합하는 연결망 체계를 구축하고 있다. 분야별 전문센터는 지질자원, 표준, 화학, 바이오, 소재 등이 있으며, 그 중 빠른 속도로 구축되는 분야는 바이오와 소재 분야이다. 특히 바이오 분야는 COVID-19와 같은 재난 극복과 관련하여 데이터의 품질이 특히 중요한 분야로써 2020년 정부가 발표한 ‘생명연구자원 빅데이터 구축전략’을 근거로, 국가생명연구자원정보센터(Korea Bioinformation Center, 이하 KOBIC)에서 ‘국가바이오데이터스테이션(Korea BioData Station, 이하 K-BDS)’이라는 이름으로 전문센터인 ‘바이오연구데이터플랫폼’을 구축하였으며, 2021년 10월 시범 운영을 시작하였다.

이렇게 국가적 차원에서 구축된 연구데이터 플랫폼이 활성화되기 위해서는 연구데이터를 생산한 연구자가 플랫폼에 적극적으로 연구데이터를 등록 즉, 공유하고, 공유된 데이터가 잘 검색되며 어려움 없이 접근할 수 있어야 한다. 또한 공유된 데이터가 타 연구자들로 인해 많이 재이용될 수 있도록 연구데이터를 인용할 수 있는 충분한 정보가 제공되어야 하며, 이는 궁극적으로 연구데이터 메타데이터의 높은 품질을 통해 보장될 수 있다.

연구데이터플랫폼의 활성화와 관련한 선행연구들은 대체로 플랫폼에 데이터를 공유함으로써 받을 수 있을 것으로 기대되는 보상(Sterling & Weinkam, 1990; Wiig, 1997; Andersen, 2004; Kankanhalli, Tan, & Wei, 2005; Kim, 2007; Piwowar, Day, & Fridsma, 2007; Hsu & Lin, 2008; Kim & Han, 2009; 김은정, 남태우, 2012; 김문정, 김성희, 2015), 플랫폼에 대한 신뢰(김지현, 2012; Kim & Stanton, 2016; Faniel, Kriesberg, & Yakel, 2016; Yoon, 2017; Yoon & Lee, 2019; 김나연, 정은경, 2020), 메타데이터의 품질(Tenopir et al., 2018; Faniel, Kriesberg, & Yakel, 2016) 등을 주요 요인으로 들고 있다. 선행연구에서 다뤄진 연구데이터플랫폼의 활성화에 영향을 미치는 각각의 여러 요인들은 실제로는 복합적으로 작용할 것이다. 하지만 대부분의 연구들은 개별적인 상관관계를 다루고 있으며, 연구의 성격과 범위 또는 연구의 한계적 특성으로 인해 각 요인들의 효과나 영향을 복합적으로 다루지 못하고 있다. 따라서 데이터의 품질과 플랫폼 활성화와의 관계 및 이 관계에서 신뢰와 보상과 같은 동기부여 요인이 갖는 효과 등을 복합적으로 파악할 필요가 있

을 것이다.

이에 본 연구에서는 전문센터 중 비교적 빠르게 구축되고 있는 바이오 분야의 K-BDS를 대상으로, 플랫폼에 공유된 데이터의 품질이 플랫폼의 활성화에 미치는 영향을 살펴보고자 하였다. 나아가 데이터의 품질과 플랫폼의 활성화의 관계에서 동기부여 요인인 보상과 신뢰가 미치는 매개효과를 검증하여 데이터 품질, 동기부여 요인과 플랫폼 활성화의 관계를 복합적으로 밝히고자 하였다. 플랫폼에 공유된 데이터의 품질의 경우 직접 데이터의 품질을 측정하는 방법이 주로 사용되나, K-BDS의 경우 시스템 운영 초기 단계로써 플랫폼에 공유된 데이터가 충분치 않아 그 품질을 정량적으로 측정하기에는 한계가 있다. 때문에 메타데이터 요소의 중요도를 평가하여 플랫폼에 공유된 데이터의 중요도를 측정하고자 하였으며, 메타데이터 요소 각각의 중요도를 측정하기보다 메타데이터의 품질을 평가할 수 있는 지표를 개발하여 그 지표에 대한 중요도를 측정하고자 하였다. 이에 본 연구에서는 메타데이터가 지녀야 할 기능의 특성에 따라 품질을 평가하는 지표를 개발한 DataON의 품질평가 모델을 적용하여(박성은, 고영만, 2022), K-BDS 메타데이터 품질평가 지표를 도출하였다.

본 연구는 데이터 품질과 플랫폼 활성화의 관계뿐만 아니라 동기부여 요인인 보상과 신뢰의 매개효과까지 복합적으로 살펴본다는 점에서 이전의 연구와 차별성을 가진다고 할 수 있다.

1.2 연구의 목적 및 연구문제

본 연구의 목적은 바이오 분야에 있어서 연

구데이터 메타데이터의 품질이 연구데이터플랫폼의 활성화에 미치는 영향 및 이 관계에서 동기부여 요인의 매개효과를 밝히는 것이다. 이에 다음과 같은 연구문제를 설정하고 연구를 진행하였다.

- 연구문제 1. 연구데이터 메타데이터의 품질은 연구데이터플랫폼의 활성화에 영향을 미치는가?
- 연구문제 2. 연구데이터 메타데이터의 품질은 연구데이터플랫폼 이용의 동기부여 요인에 영향을 미치는가?
- 연구문제 3. 연구데이터플랫폼 이용의 동기부여 요인은 연구데이터플랫폼의 활성화에 영향을 미치는가?
- 연구문제 4. 연구데이터 메타데이터의 품질이 연구데이터플랫폼의 활성화에 미치는 영향에 연구데이터플랫폼 이용의 동기부여 요인이 매개효과를 가지는가?

이러한 연구문제를 바탕으로 가설을 설정하였으며, 네 가지의 가설을 세분하여 전체 연구가설을 구성하였다.

2. 이론적 배경 및 선행연구

2.1 연구데이터 메타데이터의 품질

아직까지 연구데이터 메타데이터 품질평가에 대한 표준화된 방법은 없으나, 이와 관련한 실험적인 연구는 계속 시도되고 있으며 특히 국외에서 많이 이루어지고 있다. Rousidis et al.(2014)

는 연구데이터 메타데이터의 품질평가의 필요성을 제기하였는데, 지금까지의 연구가 연구데이터 자체의 품질 문제에 중점을 두고 있을 뿐, 연구데이터의 메타데이터의 품질 문제에 대한 연구가 수행되고 있지 못함을 지적하였다. Dryad 저장소의 메타데이터 요소 분석을 통해 주요 품질 문제를 파악하였으며, 분석 결과, 저자, 기여자 등의 이름 작성 방법과 날짜 입력 형식이 일관적이지 못함을 지적하고, 메타데이터 입력 과정에서 이러한 문제들을 예방할 수 있는 장치를 마련해야 한다고 주장하였다.

그리고 메타데이터 품질평가를 실제로 수행한 연구도 이루어졌는데, Marc et al.(2017)은 2012년부터 2014년까지 HealthData.gov에 등록된 메타데이터를 대상으로 품질평가를 실시하여 완전성, 정확성, 일관성이 낮음을 확인하였고, 이러한 원인을 HealthData.gov가 Dublin Core Metadata Initiative를 균일하게 적용하지 않은 결과로 분석하면서 메타데이터를 표준화하기 위한 정책 수립이 필요하다는 점을 지적한 바 있다.

그리고 Azeroual & Schöpfel(2019)은 연구 활동 및 결과에 대한 정보를 수집, 관리, 제공하기 위해 설립된 CRIS(Current Research Information System)의 12개국의 전문가 17명을 대상으로 CRIS 데이터의 품질 관련 이슈에 대한 설문조사를 진행한 결과, 맞춤법 오류, 중복 및 부정확한 데이터 입력, 데이터 간의 관계 누락, 결측값 등의 순으로 문제점이 관찰된다고 응답하였으며, 품질평가 기준으로 정확성, 완전성, 일관성 등이 중요한 요인으로 측정되어야 한다고 조사하였다. 또한 발생된 데이터 품질 문제를 가능한 빨리 해결하기 위해서는 데이터 정제 및 데

이터 모니터링 기술 등이 필요하며, 데이터 품질을 지속적으로 개선시키기 위해서는 예방적 접근 방법이 요구된다고 주장하였다.

또한 Gonçalves & Musen(2019)는 NCBI에서 관리하는 BioSample과 EBI에서 관리하는 BioSample의 두 가지 온라인 데이터베이스를 대상으로 메타데이터 품질에 대한 분석적 연구를 시행하였는데, 메타데이터의 속성과 그 값을 표준화하기 위해 FAIR 원칙을 따르는 온톨로지를 사용하는 것과, 메타데이터 표준화가 중요하다는 결론을 도출하였다. 이처럼 국외에서는 다양한 리포지토리를 대상으로 메타데이터의 품질을 평가한 연구들이 수행되었다.

연구데이터 메타데이터의 품질평가와 관련한 국내연구로는 고영만 외(2019)의 연구와 박성은, 고영만(2022)의 연구가 있다. 고영만 외(2019)는 'DataON'의 연구데이터 메타데이터의 품질평가 모델을 개발하였으며, 크게 '연구데이터 메타데이터의 품질평가 모델'과 '연구데이터의 유용성평가 모델'로 나누어 구현하였다. 박성은, 고영만(2022)의 연구에서는 고영만 외(2019)의 연구에서 도출된 '연구데이터 메타데이터의 품질평가 모델'의 도출 과정을 참조하여 DataON과 K-BDS의 두 플랫폼에 독립적으로 구축된 메타데이터 요소에 대한 크로스워킹 작업 후, 적합성 검증을 통해 두 플랫폼의 메타데이터 요소를 모두 매핑하였다.

연구데이터의 메타데이터 품질과 관련한 선행연구는 실제적으로 특정 플랫폼의 메타데이터 자체의 품질을 평가한 것에 그쳤으며, 품질평가 모델을 개발한 연구도 있었으나, 그 지표들이 연구데이터를 활용하는 어떠한 행위에 영향을 미치는지에 대한 연구는 찾아보기 어렵다.

2.2 동기부여 요인

2.2.1 연구데이터플랫폼을 통한 보상

연구자는 연구 결과를 공개함으로써 인용, 승진, 명성 등에 직·간접적으로 긍정적인 영향을 미치는 다양한 인센티브를 받게 되는 반면, 아직까지 연구 과정에서 생산한 연구데이터를 공개할 때 받는 인센티브는 거의 없다. 이상환, 심원식(2009)은 데이터를 제출함에 따른 구체적인 인센티브가 없고, 데이터 제출에 대한 정책과 관리 체계가 부재함이 연구데이터의 공유에 장애가 되는 요인이 된다고 하였으며, Borgman(2010)은 기존의 출판 방식과 달리 데이터 공유에는 표준 또는 공식 인용 메커니즘이 없기 때문에 데이터를 수집한 과학자에게 적절한 보상을 제공할 수 없다고 하였다. 따라서 연구자들에게 연구데이터의 제출을 독려하기 위해서는 적절한 인센티브를 제공할 수 있는 매커니즘이 필요하며, 이는 연구데이터플랫폼을 통해 보상을 할 수 있는 체계를 구축함으로써 가능할 수 있을 것이다.

한편 보상체계가 연구데이터의 공유에 주요한 영향 요인임을 밝히는 연구 결과는 많은 선행연구자들에 의해 제시되었다. Sterling & Weinkam(1990)은 보상이 낮거나 없다고 인식하면 다른 사람과 데이터를 공유하지 않을 것이라고 하였고, 반면 Wiig(1997)는 평가와 보상을 통한 동기부여의 유발이 지식공유 행위를 촉진할 수 있는 요인임을 강조하였다. 또한 데이터 공유에 영향을 미치는 요인으로서 Andersen(2004)은 학문적 보상을, Kankanhalli, Tan, & Wei(2005)은 제도적 인정을, Kim(2007)은 전문적 인정을 제시한 바 있다. 그리고 Piwowar,

Day, & Fridsma(2007)는 데이터 저장소를 통해 관련 데이터 세트를 제공한 논문이 데이터 세트를 제공하지 않은 논문보다 더 많은 인용을 받았다는 결과를 제시하였으며, Hsu & Lin(2008), Kim & Han(2009)은 지식공유 행동에서 기대되는 사회적 보상이 지식공유에 대한 개인의 태도와 지식공유 의도에 긍정적인 영향을 미친다는 것을 발견하기도 했다.

국내 연구로써 김은정, 남태우(2012)의 연구에서 데이터 수집에 영향을 미치는 요인에 대한 연구에서 데이터 제출자의 동기부여 요인을 살펴 보았는데, 보상과 관련하여 데이터와 도큐멘테이션이 출판물로 실적 인정이 될 경우(36.5%), 논문출판처럼 데이터 제출이 실적으로 인정될 경우(32.7%), 데이터에 대한 금전적 보상이 주어질 경우(15.4%) 등의 경우에 데이터를 제출하겠다는 응답이 도출되었다. 또한 김문정, 김성희(2015)의 연구에서도 보상이 연구데이터 공유에 유의한 영향을 미치는 것으로 나타난 바 있다. 이처럼 선행연구에서는 연구데이터의 공유 등을 통한 플랫폼의 활성화를 위해서는 보상 체계가 중요함을 알 수 있다.

2.2.2 연구데이터플랫폼에 대한 신뢰

신뢰는 주로 연구데이터를 공유하거나 재이용하는 행위를 이해하기 위한 관점에서 공유 혹은 재이용 의도에 영향을 미치는 요인으로써 연구되어 왔다. 이는 본 연구에서 밝히고자 하는 연구데이터플랫폼에 대한 신뢰가 연구데이터플랫폼의 활성화에 영향을 미치는지에 대한 것과 같은 맥락이라고 볼 수 있다. 리포지터리에 있는 연구데이터의 품질에 대한 신뢰가 리포지터리에 대한 신뢰로 이어지고, 궁극적으로

연구데이터의 재이용 등에도 영향을 미친다는 관련 선행연구들을 살펴보면 다음과 같다.

Zimmerman(2008)에 의하면 데이터의 신뢰성은 '누가 데이터를 생산했느냐'와 밀접하게 연관되며, 데이터를 생산한 생산자 혹은 기관에 대한 신뢰도와 데이터가 기탁된 리포지터리의 신뢰도로 판단할 수 있다. 그리고 Faniel, Kriesberg, & Yakel(2016)은 양적 연구를 하는 사회과학 분야 연구자들 사이에서는 데이터의 품질이 데이터 재이용 만족도에 가장 크게 기여하는 것을 밝힌 바 있으며, 결국 리포지터리가 데이터에 대한 이용자의 신뢰도를 높여 데이터 재이용을 촉진할 수 있다는 사실을 보고하였다.

김지현(2012)은 데이터 신뢰에 대한 문제가 데이터 재이용에 있어 핵심임을 강조하였으며, 김나연, 정은경(2020)은 신뢰성이 높은 데이터는 일정 수준의 데이터 품질이 담보된 것으로 판단할 수 있으며, 신뢰도 높은 기관에서 제공하는 데이터를 사용한 경우에는 이미 신뢰성을 담보 받아 논문을 게재할 때 불필요하게 데이터의 신뢰성과 타당성을 입증할 필요가 없어 재이용이 편리하다고 하였다.

Kim & Stanton(2016)은 데이터 공유 행동에 영향을 미치는 제도적 요인과 개별적 요인의 상호 보완적인 성격을 보여주는 연구 모델을 개발하기 위해, 43개 분야의 1,317명의 과학자를 대상으로 설문조사를 실시하였다. 이 연구에서는 데이터 리포지터리의 가용성이 반드시 과학자들의 데이터 공유를 증가시키는 것은 아닌 것으로 나타났으나, 데이터 리포지터리에 연구데이터를 제출 즉, 공유하고자 할 때 좀 더 편리한 방법으로 연구데이터를 제출할 수 있도록

시스템적인 지원이 반드시 필요하다는 시사점을 도출하기도 했다. 그리고 Yoon & Lee(2019)는 문헌 연구를 통해 데이터를 재이용하고자 하는 연구자, 즉 재이용자의 신뢰 형성에 기여할 수 있는 다섯 가지 요소를 제시하였다. 그 중 마지막 요소는 '데이터 저장소'로써 본 연구에서 다루는 연구데이터플랫폼과도 밀접한 관련이 있는 것으로 볼 수 있다.

한편 신뢰가 공동체에 대한 신뢰, 또는 다른 연구자에 대한 믿음을 의미하는 연구도 있었는데, 김선(2022)은 국내 생명공학 분야 연구자를 대상으로 연구데이터의 공유 의도에 영향을 미치는 요인을 행위론적인 관점에서 접근하였으며, 다른 연구자에 대한 믿음이 연구자들 사이에 강력한 학술적 상호주의가 형성되었을 때 데이터 공유 의도로 이어질 가능성이 높다는 것을 도출한 바 있다.

2.3 연구데이터플랫폼 활성화

2.3.1 연구데이터의 공유

연구데이터의 공유는 연구데이터의 제출, 수집 등 다양하게 표현되는데, 본 연구에서는 연구데이터의 '공유'라는 표현으로 사용한다. 연구데이터의 공유와 관련한 연구는 다양한 관점에서 이루어져 왔다. 먼저 연구데이터 공유와 관련한 과학기술 분야 연구자들의 인식에 대한 연구로, 송백화(2018)는 화학 분야의 연구자들을 대상으로 연구데이터 공유에 대한 인식을 파악하고, 인지성, 개방성, 보상 및 평가의 요인이 연구데이터 공유에 영향을 미치고 있음을 밝힌 바 있다.

한편 연구데이터 공유의 영향 요인 즉, 동기

부여 요인과 장애요인을 밝힌 연구는 국내·외에서 다수 이루어졌는데, 이 중 국내연구(이상환, 심원식, 2009; 김은정, 남태우, 2012; 김문정, 김성희, 2015; 김선, 2022)는 앞서 다루었기 때문에 생략하였다. Hedstrom & Niu(2008)는 연구비 지원을 받아 연구데이터 제출의무가 있는 연구자를 대상으로 데이터 생산자가 데이터를 제출할 때 영향을 미치는 동기부여 요인을 파악하기 위해 연구자의 태도와 행위에 대한 설문조사를 진행하였다. 조사 결과 연구자들은 데이터를 반드시 제출해야 하는 것으로 인식하고 있지는 않지만, 제출한 데이터를 다른 연구자가 이용함으로써 더 많은 공공혜택을 얻을 수 있다면 데이터를 제출할 의향이 있다고 응답하였다. 그러나 데이터 제출자는 제출할 데이터를 준비할 때 어려움을 느끼기 때문에 데이터 제출을 독려하기 위한 동기부여가 필요하다고 제안하였다. 연구데이터를 공유하는데 있어서의 장애요인에 대해 연구한 Committee on Ensuring the Utility and Integrity of Research Data in a Digital Age(2009)는 연구자들은 연구를 수행할 때 많은 시간과 노력을 투입하기 때문에, 대부분의 경우 본인의 연구 결과가 즉시 공개되어 공유되는 것을 원하지 않는다는 장애요인을 도출하였다. 이에 대한 해결책으로 연구데이터를 제출하고 공유하기 위해서는 연구자가 데이터의 독점적 사용기한을 정해 어느 정도의 시간이 경과한 후 데이터센터 혹은 아카이브에 저장되어야 한다고 하였다.

또한 연구데이터 공유에 있어서 메타데이터 품질의 중요성을 제시한 연구들이 수행되기도 했다. Tenopir et al.(2018)은 American Geophysical Union의 116개국 1,372명의 연구자들을 대상

으로 연구데이터 공유 인식 및 태도에 관한 설문조사를 수행하였다. 조사 결과, 메타데이터의 품질은 다른 연구자의 연구데이터를 사용할 때 데이터에 대한 신뢰를 판단하는 중요한 기준이므로, 메타데이터 표준, 품질 보증에 대한 세부적인 명세, 출처 정보가 포함되는 것이 필요하다고 하였다. 그리고 높은 품질의 메타데이터가 포함되고, 활용된 데이터에 대한 적절한 인용이 보장되며, 데이터 사용에 대한 저작권이 보장되면, 연구자들의 우려를 완화할 수 있다고 주장하였으며, 메타데이터 생성을 지원하고, 데이터의 출처 정보를 추가함으로써 연구자의 연구데이터가 검색 및 인용 가능성을 보장해주는 것이 중요하다고 강조하였다. 그리고 연구자의 보상체계로서 인용의 중요성을 강조하며, 데이터 리포지토리에서 인용 표준을 준수하는 등의 데이터 관리 정책 수립은 물론, 연구자의 승진, 평가에 있어서 데이터 인용 결과를 활용하도록 하여 데이터 공유인식을 확산시켜야 한다고 주장하였다. 그리고 Yoon(2017)은 데이터 품질 보증에 대한 표준이 부족하고 품질 불량 데이터 사용으로 인한 잠재적 위험성이 있으므로, 데이터의 신뢰성 평가가 데이터의 생성 및 재사용에 있어서 점차 중요해지고 있음을 주장하였다. 또한, 연구데이터 등록의 편리함은 연구자들이 데이터셋을 언제 어디서 공유할지에 관한 의사결정에 있어 중요한 요인으로 작용하고 있으며, 연구데이터의 메타데이터 표준화가 부족할 경우 연구자들이 자신의 연구데이터를 공유할 수 없음을 지적하였다.

이처럼 연구데이터의 공유에 대해서는 다양한 관점에서 연구가 이루어져 왔으나, 대부분 공유에 대한 연구자의 인식이나 태도와 관련한 동기

부여 혹은 장애요인을 밝히는 연구가 진행되었다.

2.3.2 연구데이터의 접근

연구데이터 공개원칙인 'FAIR'에 따르면 접근가능성(Accessibility)은 “데이터는 신뢰할 수 있는 저장소에 저장하여야 하며, 메타데이터와 데이터는 인간과 기계가 모두 이해할 수 있어야 한다.”는 의미를 가진다. 연구데이터의 접근 자체를 연구 주제로 다룬 연구는 찾아보기 어려웠으나, 연구데이터의 재이용이나 유용성에 대한 연구에서 연구데이터 접근에 대한 측정이 이루어졌다. Faniel, Kriesberg, & Yakel (2016)은 Inter-University Consortium for Political and Social Research(이하 ICPSR)을 통해 데이터 재이용 경험이 있는 연구자들을 선정하고 이들을 대상으로 데이터 재이용에 대한 사회과학자의 만족도 요인을 파악하기 위해, ICPSR에서 2008년부터 2012년까지 출판된 논문 중 데이터를 인용한 1,480명의 논문 저자에 대해 설문조사를 실시하였다. 그리고 수집된 총 249건의 설문지를 분석하여 데이터 완전성, 접근가능성, 조작 용이성, 신뢰성, 도큐멘테이션 품질을 데이터 재이용에 만족을 주는 주요 요인으로 도출하였다. 해당 연구에서는 ‘데이터 접근성은 데이터 재이용자의 만족도와 긍정적인 관련이 있을 것이다.’라는 가설을 도출하였으며, 검정 결과 데이터 접근성은 데이터 재이용자의 만족도와 가장 강한 관계를 나타냈다.

다음으로 연구데이터 서비스의 유용성 평가 모형을 개발한 박진호, 고영만, 김현수(2019)의 연구에서도 연구데이터의 접근성과 관련한 척도를 발견할 수 있다. 해당 연구는 다양한 사례를 통해 도출한 데이터 서비스의 유용성 평

가 요소로부터 연구데이터에 내재된 평가척도인 검색성, 접근성, 상호운용성, 재활용성 4개와 각각의 측정 지표 총 20개를 도출하였으며, 연구데이터의 잠재적 이용자 164명을 대상으로 설문조사를 실시하여 평가척도와 측정 지표에 대한 타당성과 신뢰성을 검정하였다.

2.3.3 연구데이터의 재이용

연구데이터의 재이용은 연구데이터의 이용, 재사용, 재활용 등으로 표현되며, 특히 사회과학 분야에서 데이터 재이용에 대한 인식이나 영향요인을 도출한 연구가 이루어졌다.

박지원, 장우권(2021)은 연구데이터 재이용 영향에 대해 분석하고 활성화 방안을 제시하기 위해 연구데이터를 재이용하여 새로운 연구 결과물을 산출한 경험이 있는 국내 사회과학 분야 연구자를 대상으로 설문조사를 실시하였다. 178부의 설문분석을 통해 도출한 10가지 요인(사회적 조건, 학문적 유용성, 연구의 효율성, 데이터 취약성, 연구자의 우려사항, 직접적 노력, 간접적 노력, 재이용 적합성, 데이터 완전성, 데이터 유용성)을 상관분석한 결과, 사회적 조건 및 연구의 효율성은 ‘연구데이터 재이용 의도’와 ‘양적 상관관계’가 있고, 데이터 취약성 및 연구자의 우려사항, 직접적 노력은 ‘연구데이터 재이용 의도’와 ‘음적 상관관계’가 있음이 나타났다. 회귀 분석 결과 10가지 요인들은 모두 연구데이터 재이용 의도에 유의미한 영향을 미치는 것으로 나타났으며, 연구의 효율성, 사회적 조건, 직접적 노력, 연구자의 우려 사항, 데이터 취약성 순서로 연구데이터 재이용 의도에 영향을 미치는 것으로 나타났고, 이를 토대로 연구데이터 재이용 활성화 방안을 제시하였다.

한편 데이터 재이용에 대한 인식이나 영향 요인을 자연과학·의학 및 공학 분야 과학자들을 대상으로 도출한 연구도 이루어졌다. Joo, Kim, & Kim(2017)은 건강과학자의 태도, 사회적, 자원 요인이 데이터 재사용 행동에 어떠한 영향을 미치는지 설문조사를 통해 분석하였다. 분석 결과, 건강과학자의 데이터 재사용 의도는 데이터 재사용에 대한 태도, 데이터 재사용에 대한 커뮤니티 규범, 학문적 연구 분위기, 조직 지원 요인에 의해 좌우되는 것으로 나타났다. 이 연구는 또한 데이터 재사용의 인지된 유용성과 데이터 재사용과 관련된 인지된 우려 모두가 데이터 재사용에 대한 건강과학자의 태도에 상당한 영향을 미친다는 것을 발견했다.

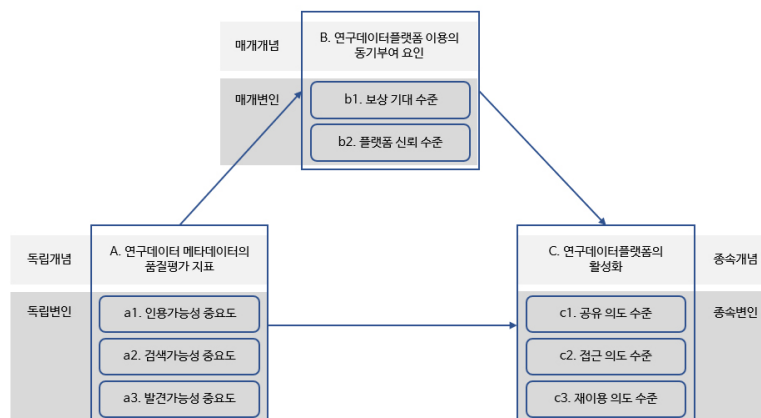
3. 연구설계

3.1 연구모형 및 변인

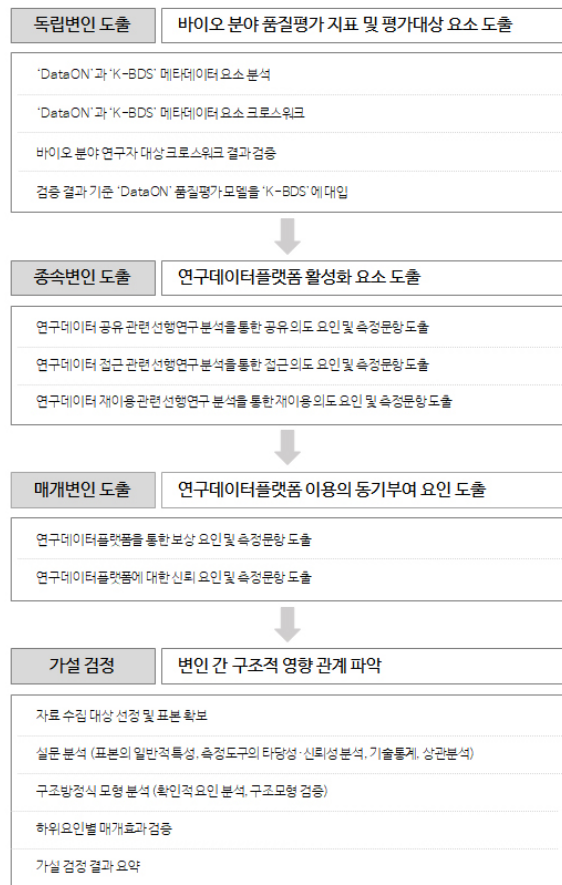
국내·외 선행연구를 통해 ‘연구데이터 메타

데이터의 품질’, ‘연구데이터플랫폼 이용의 동기부여 요인’ 그리고 ‘연구데이터플랫폼의 활성화’ 간 영향 관계를 확인하였다. ‘보상’과 ‘신뢰’가 각각 ‘연구데이터플랫폼의 활성화’에 영향을 미치는 것을 선행연구를 통해 확인하였으므로, 본 연구에서는 ‘연구데이터플랫폼 이용의 동기부여 요인’이 ‘연구데이터 메타데이터의 품질’과 ‘연구데이터플랫폼의 활성화’를 매개하는지 살펴보았다(〈그림 1〉 참조).

연구절차는 다음과 같다(〈그림 2〉 참조). 먼저 독립변인 도출 단계는 선행연구를 통해 이루어졌다(박성은, 고영만, 2022). 독립개념인 바이오 분야 ‘연구데이터 메타데이터의 품질평가 지표’를 측정하기 위한 변인은 국가연구데이터플랫폼인 DataON을 대상으로 2019년 개발된 메타데이터 품질평가 지표(고영만 외, 2019)를 토대로 K-BDS 메타데이터의 품질평가 지표를 개발하기 위해 두 플랫폼에 독립적으로 구축된 메타데이터 요소에 대한 크로스워킹 작업 후, 적합성 검증을 통해 두 플랫폼의 메타데이터 요소를 모두 매핑하였으며(박성은, 고영만,



〈그림 1〉 연구모형



〈그림 2〉 연구절차 도식화

2022), 매핑 결과를 토대로 DataON에 사용된 메타데이터 품질평가 지표의 수정 작업을 거쳐 인용가능성, 검색가능성, 발견가능성을 K-BDS의 메타데이터 품질평가 지표로 설정하였다.

다음으로 선행연구 분석을 통해 중속변인과 매개변인의 요소들을 도출하였다. 그리고 설문 조사를 통해 K-BDS의 품질평가 지표의 변인인 인용가능성, 접근가능성, 발견가능성의 중요도와 플랫폼 활성화의 변인인 연구데이터의 공유, 접근, 재이용 의도 수준 및 동기부여 요인의 변인인 보상 기대 수준과 플랫폼 신뢰 수준에

관한 데이터를 수집하고, 수집된 데이터를 분석하여 가설을 검정함으로써 변인 간의 구조적 영향 관계를 파악하였다.

본 연구의 변인은 독립개념인 '연구데이터 메타데이터의 품질평가 지표'를 측정하기 위한 '인용가능성 중요도', '검색가능성 중요도', '발견가능성 중요도'로 설정하였다. 매개개념인 '연구데이터플랫폼 이용의 동기부여 요인'을 측정하기 위한 변인은 선행연구 분석을 통해 '보상 기대 수준', '플랫폼 신뢰 수준'으로 설정하였으며, 중속개념인 '연구데이터플랫폼의 활성화'를 측

정하기 위한 변인 역시 선행연구 분석을 통해 ‘공유 의도 수준’, ‘접근 의도 수준’, ‘재이용 의도 수준’으로 설정하였다. 각 변인의 조작적 정의는 <표 1>과 같다.

3.2 연구 방법

3.2.1 연구 대상 선정 및 가설 설정

바이오 분야를 대상으로 하는 연구문제를 해결하기 위해, 2020년 8월 과학기술정보통신부에서 발표한 ‘바이오 연구데이터 등록 양식 표준화 계획(안)’에 제시된 15대 바이오 연구 분야의 주요 학회·협회 목록을 참고하여 각 학회에서 발간하는 학술지에 게재된 논문의 교신저

자 이메일 주소를 수집하였다. 총 36개의 학회 및 협회 리스트 중 학술지를 발행하지 않는 학회 및 협회를 제외하고, 학회 회원에게만 학술지가 공개된 경우를 제외하여, 최종적으로 26개 학회의 학술지를 대상으로 하였으며, 총 2,344개의 이메일 주소를 수집하여 연구 대상으로 삼아 설문조사를 실시하였다.

3.2.2 설문지 배포 및 회수

설문지는 인구통계학적 특성을 묻는 7개의 문항을 비롯하여 ‘연구데이터 메타데이터의 품질 평가 지표’ 20문항, ‘연구데이터플랫폼의 활성화’ 15문항, ‘연구데이터플랫폼 이용의 동기부여 요인’ 12문항 등 총 54개 문항으로 이루어져 있다.

<표 1> 변인의 조작적 정의

구분	개념	변인	조작적 정의	문항 수	출처
독립	연구데이터 메타데이터의 품질평가 지표	인용가능성 중요도	연구데이터 인용을 위해 필수적인 메타데이터 요소들이 포함되어 있는지에 대한 중요도	8	고영만 외(2019) 박성은, 고영만(2022)
		검색가능성 중요도	연구데이터 검색을 위해 필수적인 메타데이터 요소들이 포함되어 있는지에 대한 중요도	3	
		발견가능성 중요도	연구데이터 발견을 위해 필수적인 메타데이터 요소들이 포함되어 있는지에 대한 중요도	9	
매개	연구데이터플랫폼 이용의 동기부여 요인	보상 기대 수준	연구데이터플랫폼에서의 연구데이터의 가치 및 그에 대한 성과에 대해 공정하게 평가하여 인정함으로써 조직적, 사회적 보상을 유발할 수 있는 시스템에 대한 기대 정도	4	김은정, 남태우(2012) 김문정, 김성희(2015) Kim & Stanton(2016)
		플랫폼 신뢰 수준	연구데이터플랫폼에 있는 연구데이터의 품질 및 관리 체계에 대한 믿음의 정도	8	Kim & Stanton(2016) Yoon & Lee(2019) 박지원, 장우권(2021)
종속	연구데이터플랫폼의 활성화	공유 의도 수준	연구데이터플랫폼에 본인의 연구데이터를 공유하고자 하는 의도의 정도	6	김은정, 남태우(2012) 김문정, 김성희(2015) 송백화(2018) 김선(2022)
		접근 의도 수준	연구데이터플랫폼에 있는 연구데이터에 접근하고자 하는 의도의 정도	4	Faniel, Kriesberg, & Yael(2016) 박진호, 고영만, 김현수(2019)
		재이용 의도 수준	연구데이터플랫폼에 있는 연구데이터를 재이용하고자 하는 의도의 정도	5	김문정, 김성희(2015) Joo, Kim, & Kim(2017) 박지원, 장우권(2021)

설문조사는 2022년 9월 16일부터 10월 5일 까지 약 3주간 온라인으로 진행되었다. 설문지는 Survey Monkey를 활용하여 이메일을 통해 앞서 도출한 연구 대상 2,344명에게 배포하였으며, 373부를 회수하여 15.9%의 응답률을 나타냈다. 이 중 응답자가 모든 설문조사 항목을 완료하지 않은 43부를 제외한 330부를 최종 데이터 분석에 사용하였다.

3.2.3 분석 방법

본 연구에서는 설문조사를 통해 수집된 자료를 분석하고 가설을 검증하기 위해 SPSS version 25와 AMOS 25 프로그램을 사용하여 <표 2>와 같이 통계 분석을 실시하였다.

4. 분석 결과

4.1 설문 분석

4.1.1 연구대상자의 특성

총 330명의 설문 응답자의 특성을 파악하기

위해 소속기관, 신분, 전공분야, 연구데이터 생산 및 이용 경험, K-BDS 인지여부에 대해 조사하였다(<표 3> 참조).

4.1.2 측정 도구의 타당성 분석

본 연구에서 활용한 측정 도구의 타당성을 검증하기 위해 탐색적 요인분석(Exploratory factor analysis)을 실시하였다. 요인분석 방법 중 원래의 변인들의 분산 중 가급적 많은 부분을 설명하는 요인을 추출하면서 정보손실을 최소화하는 주성분분석(Principal component analysis)을 사용하였고, 요인의 독립성을 유지하면서 요인구조가 가장 뚜렷할 때까지 요인을 회전시키는 배리맥스 회전(Varimax rotation)을 사용하여 분석하였다. 요인 분류는 고유값(Eigen value)이 1 이상일 때 하나의 요인으로 구성하였으며, 요인 적재량(Factor loading)이 .50을 초과하면 해당 요인으로 분류했다.

타당성을 저해하는 14개 항목(인용1, 인용2, 인용6, 인용7, 발견1 발견2, 발견3, 발견4, 발견8 발견9, 신뢰6, 신뢰7, 신뢰8, 공유5)이 제외되어 최종적으로 33개의 항목으로 요인분석을 실시

<표 2> 분석 방법

구분	분석 목적	분석 방법
응답자 특성	인구통계학적 특성 분석	빈도분석(Frequency Analysis)
측정 도구 분석	변인의 구성 타당성	탐색적 요인분석(Exploratory Factor Analysis)
	측정 도구의 신뢰성	크론바흐 알파 계수 산출(Cronbach's alpha)
기술 통계	연구 변인의 수준 파악	평균, 표준편차 산출
	변인들의 정규성 가정 충족 여부 파악	왜도, 첨도 산출(Skewness, Kurtosis)
변인 간의 관련성	변인 간의 상관관계 분석	Pearson 상관분석(Pearson correlation coefficient)
가설의 검증	변인 간의 구조적 관계 검증	구조방정식 모형 분석(Structural Equation Model)
	간접효과의 유의성 검증	부트스트랩 분석(Bootstrap)
	매개효과 검증	Baron과 Kenny가 제안한 3단계에 걸친 회귀분석 Sobel test를 통한 재검증

〈표 3〉 연구대상자의 일반적 특성

항목	구분	N	%
소속기관	대학	207	62.7
	연구기관	94	28.5
	기타	29	8.8
신분	대학원생 - 박사과정	4	1.2
	교수	198	60.0
	연구원	103	31.2
	기타	25	7.6
전공 분야	신약	36	10.9
	의료기기	7	2.1
	뇌과학	24	7.3
	바이오제제	14	4.2
	혁신형 융복합	40	12.1
	보건	25	7.6
	중자/가축	15	4.5
	지능형 농업	32	9.7
	해양수산	21	6.4
	식품	40	12.1
	동식물 치료제	10	3.0
	바이오화학/에너지	12	3.6
	공정	16	4.8
	바이오소재	13	3.9
	환경바이오	25	7.6
연구데이터 생산 경험	있음	312	94.5
	없음	18	5.5
연구데이터 이용 경험	있음	279	84.5
	없음	51	15.5
K-BDS 인지 여부	있음	22	6.7
	없음	308	93.3
전체		330	100.0

하였다. 분석 결과, KMO 측도는 .858로 나타났다. Bartlett의 구형성 검증 결과 또한 유의하게 나타나($p < .001$), 요인분석 모형은 적합한 것으로 판단되었다.

분석 결과 여덟 개의 요인으로 분류되었으며, 여덟 개의 요인은 68.714%의 요인 설명력을 보였다. 첫 번째 요인은 4개 항목으로 '인용가능성 중요도', 두 번째 요인은 3개 항목으로 '검

색가능성 중요도', 세 번째 요인은 3개 항목으로 '발견가능성 중요도', 네 번째 요인은 5개 항목으로 '공유 의도 수준', 다섯 번째 요인은 4개 항목으로 '접근 의도 수준', 여섯 번째 요인은 5개 항목으로 '재이용 의도 수준', 일곱 번째 요인은 4개 항목으로 '보상 기대 수준', 여덟 번째 요인은 5개 항목으로 '플랫폼 신뢰 수준'으로 구성되었다.

4.1.3 측정 도구의 신뢰성 분석

본 연구를 위해 측정한 변인들의 척도에 대한 신뢰성을 판단하기 위해 크론바흐 알파 계수(Cronbach's alpha)를 산출하였다(〈표 4〉 참조). 일반적으로 크론바흐 알파 계수가 .60 이상

이면 척도를 구성하는 항목들 간에 내적 일관성이 양호한 것으로 판단한다(Hair et al., 2006). 본 연구의 모든 변인은 크론바흐 알파 계수가 .60 이상으로 나타나 사용된 측정 도구들의 신뢰성이 확보되었다고 할 수 있다.

〈표 4〉 신뢰성 분석 결과

구분	개념	변인	문항 수	측정 문항		Cronbach's Alpha
독립	연구데이터 메타데이터의 품질평가 지표	인용 가능성 중요도	4	인용3	생성자(Creator)	.722
				인용4	출판자(Publisher)	
				인용5	출판연도(Publication Year)	
				인용8	출판자 소재지(Location of publisher)	
		검색 가능성 중요도	3	검색1	키워드(Keyword)	.668
				검색2	설명(Description)	
				검색3	주제분류(Subject)	
		발견 가능성 중요도	3	발견5	파일설명(File Description)	.845
				발견6	파일포맷(File Format)	
				발견7	파일유형(File Type)	
		전체	10			.784
매개	연구데이터 플랫폼 이용의 동기부여 요인	보상 기대 수준	4	보상1	연구데이터플랫폼 활용시, 공정한 보상체계가 필요하다.	.939
				보상2	연구데이터의 가치는 공정하게 측정되어야 한다.	
				보상3	연구업무실적으로 인정한다면 연구데이터를 제출할 것이다.	
				보상4	연구데이터를 공유할수록 전문가로서 인정받을 수 있다.	
		플랫폼 신뢰 수준	5	신뢰1	나는 연구데이터플랫폼을 신뢰한다.	.778
				신뢰2	연구데이터플랫폼은 데이터 전문가가 관리한다.	
				신뢰3	데이터 전문가는 데이터에 대한 전문지식이 있다.	
				신뢰4	전문가가 선별한 데이터는 잘 관리된다.	
				신뢰5	연구데이터는 신뢰할 수 있는 연구데이터플랫폼에 기탁되어 있다.	
		전체	9			.825
종속	연구데이터 플랫폼의 활성화	공유 의도 수준	5	공유1	나는 동료들의 업무 수행에 나의 연구데이터가 필요하다고 요청할 경우, 나의 연구데이터를 제공할 것이다.	.945
				공유2	나는 향후 연구를 위해 다른 연구자들과 내 데이터를 공유할 것 같다.	
				공유3	나는 나의 연구데이터를 기꺼이 제공할 것이다.	
				공유4	나는 향후 연구를 위해 다른 연구자들과 내 데이터를 공유할 계획이다.	
				공유5	나는 향후 연구를 위해 다른 연구자들과 내 데이터를 공유하려고 노력할 것이다.	
				공유6	나는 향후 연구를 위해 다른 연구자들과 내 데이터를 공유하려고 노력할 것이다.	
		접근 의도 수준	4	접근1	나는 내가 원하는 연구데이터를 쉽게 검색할 수 있다.	.865
				접근2	나는 내가 원하는 연구데이터에 제한사항(IP제한, 인가자 접근 등) 없이 자유롭게 접근할 수 있다.	
				접근3	내가 원하는 연구데이터는 별도 추가처리 과정 없이 접근 즉시 처리할 수 있는 형식이다.	
				접근4	내가 원하는 연구데이터는 개방형 포맷으로 제한없이 쉽게 접근할 수 있다.	
		제이용 의도 수준	5	제이용1	나는 미래의 연구를 위해 다른 연구자의 데이터를 제이용할 가능성이 있다.	.692
				제이용2	향후 연구를 위해 다른 연구자의 데이터를 제이용할 계획이다.	
				제이용3	나는 나의 미래 연구를 위해 다른 연구자의 데이터를 제이용하려고 노력할 것이다.	
				제이용4	나는 내 연구를 위해 다른 연구자의 데이터를 제이용할 의사가 있다.	
				제이용5	나는 나의 연구를 위해 다른 연구자들의 데이터를 제이용할 것이다.	
		전체	14			.851

4.1.4 기술통계

본 연구에서 측정한 연구 변인의 수준을 파악하기 위해 평균과 표준편차를 산출하였으며, 변인들의 정규성 가정 충족 여부를 판단하기 위해 왜도(Skewness)와 첨도(Kurtosis)를 산출하였다. 왜도는 절댓값 3 미만, 첨도는 절댓값 10 미만이면 정규분포에 근사하는 것으로 판단하는데(Kline, 2016), 본 연구에서는 <표 5>와 같이 모든 변인이 정규성 가정을 충족하는 것으로 나타났다.

4.1.5 상관분석

변인 간 상관관계를 파악하기 위해 Pearson의 상관분석을 실시하였으며(<표 6> 참조), 상관분석 결과를 연구의 주요 변인을 중심으로 정리하면 다음과 같다.

첫째, 연구데이터 메타데이터의 품질평가 지표는 연구데이터플랫폼 이용의 동기부여 요인($r=.359, p<.001$)과 그 하위요인인 보상 기대 수준($r=.345, p<.001$), 플랫폼 신뢰 수준($r=.264, p<.001$)과 유의한 정(+)적 상관관계를 보였으며,

연구데이터플랫폼의 활성화($r=.280, p<.001$)와 그 하위요인인 공유 의도 수준($r=.263, p<.001$), 재이용 의도 수준($r=.266, p<.001$)과 유의한 정(+)적 상관관계를 보였다.

둘째, 연구데이터플랫폼 이용의 동기부여 요인은 연구데이터 메타데이터의 품질평가 지표($r=.359, p<.001$)와 그 하위요인인 인용가능성 중요도($r=.127, p<.05$), 검색가능성 중요도($r=.348, p<.001$), 발견가능성 중요도($r=.337, p<.001$), 연구데이터플랫폼의 활성화($r=.333, p<.001$)와 그 하위요인인 공유 의도 수준($r=.358, p<.001$), 접근 의도 수준($r=.139, p<.05$), 재이용 의도 수준($r=.199, p<.001$)과 유의한 정(+)적 상관관계를 보였다.

셋째, 연구데이터플랫폼의 활성화는 연구데이터 메타데이터의 품질평가 지표($r=.280, p<.001$)와 그 하위요인인 검색가능성 중요도($r=.251, p<.001$), 발견가능성 중요도($r=.276, p<.001$), 연구데이터플랫폼 이용의 동기부여 요인($r=.333, p<.001$)과 그 하위요인인 보상 기대 수준($r=.221, p<.001$), 플랫폼 신뢰 수준($r=.338, p<.001$)과

<표 5> 기술통계 분석 결과

개념	변인	최소값	최대값	평균	표준편차	왜도	첨도
연구데이터 메타데이터의 품질평가 지표	인용가능성 중요도	1.75	5.00	3.66	0.65	-0.02	-0.01
	검색가능성 중요도	2.67	5.00	4.36	0.51	-0.72	0.45
	발견가능성 중요도	1.00	5.00	3.85	0.70	-0.41	0.59
	전체	2.33	5.00	3.96	0.46	-0.19	0.49
연구데이터 플랫폼 이용의 동기부여 요인	보상 기대 수준	1.00	5.00	3.99	0.59	-0.54	1.53
	플랫폼 신뢰 수준	1.00	5.00	3.54	0.63	-0.39	0.89
	전체	1.00	5.00	3.76	0.52	-0.51	1.99
연구데이터 플랫폼의 활성화	공유 의도 수준	2.00	5.00	3.95	0.64	-0.45	0.52
	접근 의도 수준	1.00	5.00	3.13	0.65	-0.03	-0.08
	재이용 의도 수준	2.00	5.00	3.79	0.67	-0.19	-0.01
	전체	2.00	5.00	3.62	0.45	0.02	0.49

〈표 6〉 상관관계 분석 결과

	1	2	3	4	5	6	7	8	9	10	11
1. 품질평가지표	1										
2. 인용가능성	.700***	1									
3. 검색가능성	.725***	.267***	1								
4. 발견가능성	.800***	.264***	.456***	1							
5. 동기부여요인	.359***	.127*	.348***	.337***	1						
6. 보상	.345***	.120*	.339***	.324***	.835***	1					
7. 신뢰	.264***	.095	.254***	.248***	.856***	.431***	1				
8. 플랫폼활성화	.280***	.101	.251***	.276***	.333***	.221***	.338***	1			
9. 공유	.263***	.090	.272***	.238***	.358***	.281***	.323***	.751***	1		
10. 접근	.052	.050	.002	.055	.139*	.031	.199***	.602***	.168**	1	
11. 재이용	.266***	.070	.249***	.280***	.199***	.150**	.185**	.727***	.406***	.090	1

* p<.05, ** p<.01, *** p<.001

유의한 정(+)적 상관관계를 보였다.

준으로 나타나 측정모형이 적합한 것으로 판단되었다.

4.2 구조방정식 모형 분석

4.2.1 확인적 요인분석

(1) 측정모형의 적합도

구조방정식에서 연구모형 분석을 진행하기 전에 측정변인들이 잠재변인을 잘 설명하고 있는지 확인하기 위해 확인적 요인분석(Confirmatory Factor Analysis, CFA)을 통해 측정모형의 적합도를 검증하였다. 측정모형의 모수 추정을 위해 최대우도법(Maximum likelihood)을 사용하였으며, 하위요인을 바탕으로 묶음지표(Item parceling)로 측정변인을 구성하였다(Sass & Smith, 2006).

측정모형의 적합도 결과를 살펴보면, $\chi^2(df=17)$ 값은 28.271($p<.001$)로 나타났으나, χ^2 값은 표본의 크기와 모형의 복잡성에 많은 영향을 받기 때문에 다른 적합도 지수를 확인하였다(Bollen & Long, 1993). 그 결과, RMR=.016, GFI=.979, AGFI=.956, RMSEA=.045로 적합했으며, IFI=.970, TLI=.949, CFI=.969로 모두 양호한 수

(2) 확인적 요인분석 결과

확인적 요인분석에서 각 측정변인이 잠재변인을 잘 반영하는지 파악하기 위해 측정변인들의 요인 부하량을 확인한 결과, 〈표 7〉과 같이 모든 요인 부하량이 유의하게 나타나 대체로 측정변인들이 해당 잠재변인을 잘 반영하는 것으로 판단되었다.

(3) 수렴타당도 검증

수렴타당도(Convergent validity)란 동일한 개념을 측정하기 위하여 서로 다른 방법으로 측정한 값 사이에 높은 상관관계가 있어야 한다는 것이다. 즉, 수렴타당도는 동일개념을 측정하는 복수의 문항들이 어느 정도 일치하는가를 검증하는 것이다. 이러한 수렴타당도를 검증하기 위해 각 변인에 대한 개념신뢰도(CR)와 평균분산추출값(AVE)을 산출하였다.

개념신뢰도는 .70 이상, 평균분산추출값은 .50

〈표 7〉 확인적 요인분석 결과

요인		B	S.E.	β	C.R.	p
품질평가 지표	→ 인용	1		.360		
품질평가 지표	→ 검색	0.748	0.101	.696	7.389***	<.001
품질평가 지표	→ 발견	0.491	0.402	.673	5.098***	<.001
동기부여 요인	→ 보상	1		.682		
동기부여 요인	→ 신뢰	0.987	0.148	.632	0.677***	<.001
플랫폼 활성화	→ 공유	1		.750		
플랫폼 활성화	→ 접근	0.280	0.096	.207	2.929**	0.003
플랫폼 활성화	→ 재이용	0.756	0.135	.542	5.583***	<.001

* p<.05, ** p<.01, *** p<.001

이상이면 수렴타당도가 양호한 것으로 판단하는데(우종필, 2012), 개념신뢰도의 경우 ‘연구데이터플랫폼 이용의 동기부여 요인’은 기준치에 근접한 값(.603)이며 이외의 값은 기준치를 상회(.795, .714)하였고, 평균분산추출값의 경우 ‘연구데이터플랫폼 이용의 동기부여 요인(.432)’과 ‘연구데이터플랫폼의 활성화(.497)’는 기준치를 상회하지 않았으나 기준치에 근접한 값으로 본 모형의 수렴타당도는 전반적으로 양호한 것으로 판단되었다.

(4) 판별타당도 검증

판별타당도(Discriminant validity)란 서로 다른 변인들 간에는 그 측정치에도 분명한 차이가 있어야 한다는 것을 의미한다. 이러한 판별타당도를 검증하기 위해 각 변인에 대한 평

균분산추출값(AVE)을 측정하였고, 변인들 간의 상관계수를 산출하였다. 판별타당도를 평가하기 위한 기준은 각 변인 간 상관계수 제곱값이 평균분산추출값을 상회하지 않아야 하는데, 〈표 8〉을 보면 모든 상관계수 제곱값이 평균분산추출값을 상회하지 않았으므로 판별타당도는 확보되었다고 볼 수 있다.

이상의 확인적 요인분석 결과를 도식화하면 〈그림 3〉과 같으며, 그림에서는 본 연구의 개념과 변인에 해당하는 용어를 축약해서 표현하였다.

4.2.2 구조모형 검증

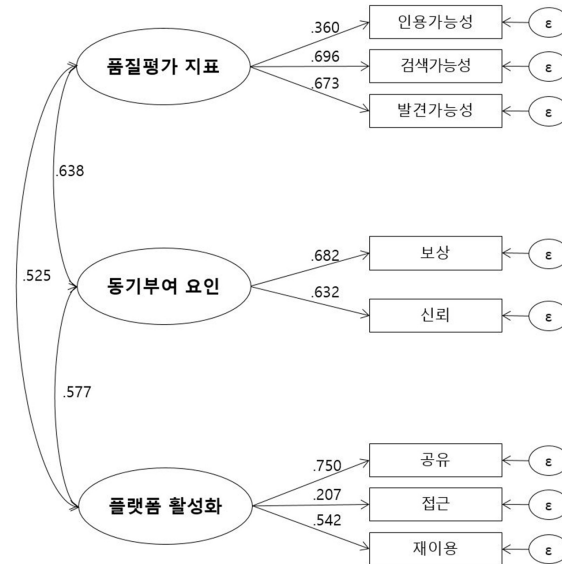
(1) 연구모형의 적합도

연구모형의 적합도 결과는, $\chi^2(df=17)$ 값의 경우 28.271(p<.001)로 나타났으나, χ^2 값은 표

〈표 8〉 판별타당도 검증 결과

	품질평가 지표	동기부여 요인	플랫폼 활성화
품질평가 지표	(.581)	.407	.276
동기부여 요인	.638	(.432)	.333
플랫폼 활성화	.525	.577	(.497)

대각선: AVE, 대각선 아래: 상관계수, 대각선 위: 상관계수 제곱



〈그림 3〉 확인적 요인분석 모형

본의 크기와 모형의 복잡성에 많은 영향을 받기 때문에 다른 적합도 지수를 확인하였다(Bollen & Long, 1993). 그 결과, RMR=.016, GFI=.979, AGFI=.956, RMSEA=.045로 적합했으며, IFI=.970, TLI=.949, CFI=.969로 모두 양호한 수준으로 나타나 연구모형이 적합한 것으로 판단되었다.

(2) 연구모형의 경로분석

구조모형에서 잠재변인 간 영향 관계를 검증하기 위해 경로분석을 실시하였으며, 이를 통해

연구가설을 검정하였다(〈표 9〉 참조).

첫째, 본 연구는 바이오 분야 ‘연구데이터 메타데이터의 품질평가 지표’가 ‘연구데이터플랫폼의 활성화’에 영향을 미칠 것이라는 가설을 설정하였다(가설 1). 이 변인 간의 영향력을 분석한 결과 연구데이터 메타데이터의 품질평가 지표는 연구데이터플랫폼의 활성화($\beta=.264$, $p<.05$)에 통계적으로 유의한 정(+)의 영향을 미치는 것으로 나타나 가설 1은 채택되었다.

둘째, 본 연구는 바이오 분야 ‘연구데이터 메타데이터의 품질평가 지표’가 ‘연구데이터플랫폼

〈표 9〉 연구모형의 경로분석 결과

경로	B	S.E.	β	C.R.	p
품질평가 지표 → 동기부여 요인	0.542	0.090	0.542	5.990***	<.001
품질평가 지표 → 플랫폼 활성화	0.268	0.134	0.264	2.006*	<.05
동기부여 요인 → 플랫폼 활성화	0.488	0.170	0.409	2.874**	<.01

* $p<.05$, ** $p<.01$, *** $p<.001$

품 이용의 동기부여 요인'에도 긍정적 영향을 미칠 것이라는 가설을 설정하였다(가설 2). 이를 분석한 결과, 연구데이터 메타데이터의 품질평가 지표는 연구데이터플랫폼 이용의 동기부여 요인($\beta=.542$, $p<.001$)에도 통계적으로 유의한 정(+)의 영향을 미치는 것으로 확인되어, 가설 2는 채택되었다.

셋째, '연구데이터플랫폼 이용의 동기부여 요인'이 '연구데이터플랫폼의 활성화'에 영향을 미칠 것이라는 가설 3에 대한 분석 결과, 연구데이터플랫폼 이용의 동기부여 요인이 연구데이터플랫폼의 활성화($\beta=.409$, $p<.01$)에 통계적으로 유의한 정(+)의 영향을 미치는 것으로 나타나, 가설 3도 채택되었다.

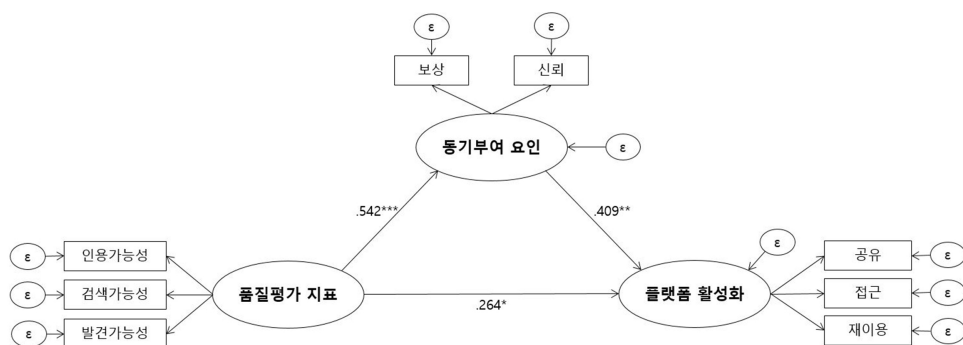
즉 바이오 분야 '연구데이터 메타데이터의 품질평가 지표'의 중요도가 높을수록 '연구데이

터플랫폼 이용의 동기부여 요인', '연구데이터플랫폼의 활성화'가 높아지고, '연구데이터플랫폼 이용의 동기부여 요인'이 높을수록 '연구데이터플랫폼의 활성화'가 높아진다고 할 수 있다(〈그림 4〉 참조).

(3) 매개효과 검증

구조모형에서 각 변인들 간의 직접효과, 간접효과 및 전체효과를 더 자세히 살펴보기 위해 모형의 효과분해를 실시하였다(〈표 10〉 참조).

다음으로 부트스트랩(Bootstrap)을 통해 직접효과, 간접효과, 총효과의 크기를 산출하고, 그 유의성을 검증하였다. 그 결과, 〈표 11〉과 같이 바이오 분야 '연구데이터 메타데이터의 품질평가 지표'가 '연구데이터플랫폼 이용의 동기부여 요인'을 매개하여 '연구데이터플랫폼의 활성화



〈그림 4〉 연구모형의 경로분석 결과

〈표 10〉 구조모형의 효과분해

경로	직접효과(β)	간접효과(β)	총효과(β)
품질평가 지표 → 동기부여 요인	0.638**		0.638**
품질평가 지표 → 플랫폼 활성화	0.264**	0.261	0.525**
동기부여 요인 → 플랫폼 활성화	0.409**		0.409**

* $p<.05$, ** $p<.01$, *** $p<.001$

〈표 11〉 매개효과 검증 결과

경로	effect	95% CI		p
		LLCI	ULCI	
품질평가 지표 → 동기부여 요인 → 플랫폼 활성화	0.525**	0.317	0.791	<.01

* p<.05, ** p<.01, *** p<.001

화’에 미치는 간접효과($B=0.525$, $p<.01$)가 통계적으로 유의한 것으로 검정되어 가설 4도 채택되었다.

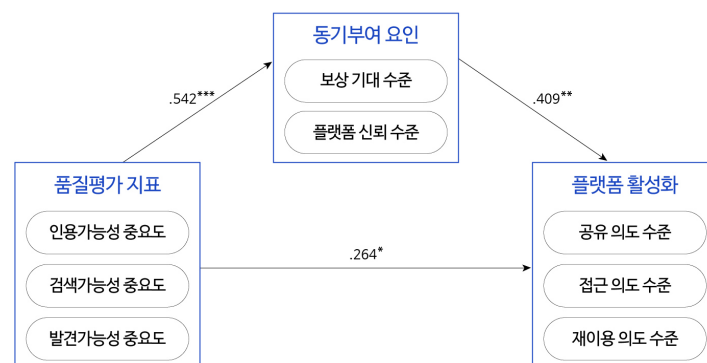
따라서 바이오 분야 ‘연구데이터 메타데이터의 품질평가 지표’는 ‘연구데이터플랫폼의 활성화’에 직접적인 영향을 미치며, ‘연구데이터플랫폼이용의 동기부여 요인’을 통하여 간접적으로도 ‘연구데이터플랫폼의 활성화’에 유의한 정(+)의 영향을 미친다고 할 수 있다.

바이오 분야 ‘연구데이터 메타데이터의 품질평가 지표’와 ‘연구데이터플랫폼 이용의 동기부여 요인’, ‘연구데이터플랫폼의 활성화’ 3개 변인의 구조적 관계를 파악하기 위해 경로분석 및 변인간 효과분석을 통한 가설 검정 결과를 정리하면 〈그림 5〉와 같다. ‘연구데이터 메타데이터의 품질평가 지표’, ‘연구데이터플랫폼의 활

성화’, ‘연구데이터플랫폼 이용의 동기부여 요인’ 3개 변인 간의 구조적 관계를 살펴보았을 때, 연구자가 품질평가 지표에 대해 중요하다고 생각할수록 ‘연구데이터플랫폼 이용의 동기부여’ 정도, 그리고 ‘연구데이터플랫폼의 활성화’ 의도가 높아지는 것을 확인하였다. 또한 품질평가 지표가 ‘연구데이터플랫폼의 활성화’에 영향을 미칠 때, ‘연구데이터플랫폼 이용의 동기부여 요인’의 매개효과도 확인되었다.

4.3 품질평가 지표의 하위요인별 매개효과 검증

바이오 분야 ‘연구데이터 메타데이터의 품질평가 지표’ 각각의 요소와 ‘연구데이터플랫폼의 활성화’ 각각의 요소의 관계에서 ‘연구데이터플



〈그림 5〉 가설 검정 결과 1

랫폼 이용의 동기부여 요인' 각각의 요소의 매개효과를 검증함으로써 좀 더 세부적인 구조를 파악하기 위해 Baron과 Kenny(1986)가 제안한 3단계에 걸친 위계적 회귀분석과 Sobel test를 통한 재검증을 수행하였으며, 그 결과를 요약하면 다음과 같다(〈그림 6〉 참조).

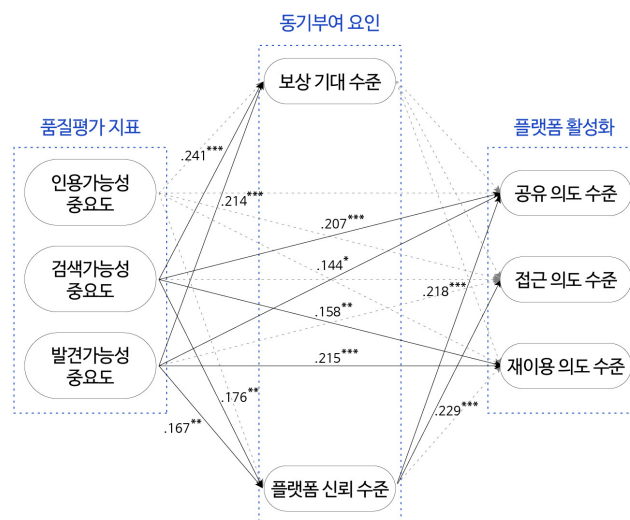
연구데이터 공유 의도 수준에 대해서는 검색가능성과 발견가능성 중요도가 공유 의도 수준에 미치는 영향에서 플랫폼 신뢰 수준의 매개효과가 검증되었고, 연구데이터 접근 의도 수준에 대해서는 검색가능성과 발견가능성 중요도는 접근 의도 수준에 직접적인 영향을 미치는 않지만, 플랫폼 신뢰 수준을 매개로 연구데이터에 대한 접근 의도에 간접적으로 영향을 미치는 것으로 나타났으며, 연구데이터 재이용 의도 수준에 대해서는 검색가능성과 발견가능성 중요도는 재이용 의도 수준에 직접적인 영향을 미치는 것으로 나타났지만, '연구데이터 메타데이터의 품질평가 지표'(인용가능성, 검색

가능성, 발견가능성 중요도)가 재이용 의도 수준에 미치는 영향에서 보상 기대 수준과 플랫폼 신뢰 수준의 매개효과는 확인되지 않았다.

이를 통해 바이오 분야의 연구데이터플랫폼이 활성화되기 위해서는 다음의 사항을 고려할 필요가 있는 것으로 분석되었다.

첫째, 연구데이터플랫폼에서의 공유를 활성화 시키기 위해서는 연구데이터의 메타데이터 품질 중 검색가능성을 높이는 것이 가장 효과적이다. 독립변인인 검색가능성 중요도($\beta=.207$, $p<.001$)와 발견가능성 중요도($\beta=.144$, $p<.05$)가 종속변인인 공유 의도 수준에 유의한 정(+)의 영향을 미치는 것을 확인하였으며, 검색가능성, 발견가능성 중요도 순으로 공유 의도 수준에 영향을 미치는 것으로 나타났기 때문이다.

둘째, 연구데이터플랫폼에서의 재이용을 활성화시키기 위해서는 연구데이터의 메타데이터 품질 중 발견가능성을 높이는 것이 가장 효과적이다. 독립변인인 검색가능성 중요도($\beta=.158$,



〈그림 6〉 가설 검정 결과 2

$p<.01$)와 발견가능성 중요도($\beta=.215$, $p<.001$)가 종속변인인 재이용 의도 수준에 유의한 정(+)의 영향을 미치는 것을 확인하였으며, 발견가능성, 검색가능성 중요도 순으로 재이용 의도 수준에 영향을 미치는 것으로 나타났기 때문이다.

5. 결 론

본 연구는 COVID-19 이후 국내 연구데이터 인프라 역량 강화의 필요성이 빠른 속도로 제기되는 상황에서 바이오 분야 연구데이터플랫폼인 K-BDS를 대상으로 연구데이터 메타데이터의 품질과 연구데이터플랫폼의 활성화, 그리고 연구데이터플랫폼 이용에 관한 동기부여 요인의 영향 관계를 살펴보았다. 나아가 연구데이터플랫폼 이용의 동기부여 요인이 연구데이터 메타데이터의 품질과 연구데이터플랫폼의 활성화 간의 관계를 매개하는지를 분석하였다. 연구는 품질평가 지표와 플랫폼 활성화에 관한 변인 및 매개변인들에 대한 설문조사를 실시하여 각 변인간의 영향 관계를 분석하는 형태로 진행되었다. 연구 결과를 통해 도출할 수 있는 시사점은 다음과 같다.

첫째, 활성화에 영향을 미치는 요인과는 달리 연구데이터 메타데이터의 품질평가 지표 중 인용가능성 중요도는 어느 변인에도 유의한 영향을 미치지 않는 것으로 나타났다. 연구데이터를 연구데이터플랫폼에 공유하는 연구자는 본인의 연구데이터가 많이 인용되기를 바랄 것으로 가정하였으나, 인용가능성의 중요도는 검색가능성이나 발견가능성의 중요도보다 상대

적으로 낮게 나타났다. 이는 아직 연구데이터 인용에 대한 제도가 정착되어 있지 않고, 연구데이터 인용으로 인한 혜택이 크지 않기 때문일 수 있다. 따라서 연구데이터 인용에 대한 혜택이 제도적으로 정착될 필요가 있을 것이다.

둘째, 보상과 관련한 가설이 모두 통계적으로 유의미하지 않은 결과를 나타내었는데, 이는 보상에 대한 기대가 연구데이터플랫폼의 활성화에 유의한 영향을 미치지 않은 것을 의미한다. 그러나 선행연구에서는 보상체계가 연구데이터의 공유에 주요한 영향 요인임을 밝히는 결과가 많이 도출된 바 있다. 이는 앞서 인용가능성에 대한 고찰에서 언급한 것과 같은 맥락에서 연구데이터의 공유 혹은 재이용에 대한 보상체계가 정규화되지 않아 보상에 대한 인식이 확립되지 않았기 때문일 수 있다. 한편으로는 시스템적으로 연구데이터 메타데이터의 품질이 좋고, 이를 통해 연구데이터플랫폼에 대한 신뢰가 충족될 경우 연구데이터플랫폼을 통한 보상이 없더라도 연구데이터플랫폼이 활성화될 수 있을 것으로도 해석할 수 있다. 따라서 연구데이터플랫폼을 활성화하기 위해서는 보상보다 우선적으로 연구데이터플랫폼의 메타데이터 품질을 향상시킴으로써 시스템적인 지원을 충분히 하고, 이를 통해 연구데이터플랫폼에 대한 신뢰를 높이는 것이 중요하다는 시사점을 얻을 수 있다.

마지막으로 바이오 분야를 대상으로 수행된 본 연구 결과를 전체 학문 분야로 확대하여 해석하는 것에는 한계가 있으며, 후속 연구를 통해 기각된 가설에 대하여 인터뷰 등을 통해 심층적인 추가 분석을 진행하여 그 원인을 밝혀보는 작업이 수행될 수 있을 것이다.

참 고 문 헌

- 고영만, 안혜연, 김현수, 심지우 (2019). 국가연구데이터플랫폼의 메타데이터 품질 및 유용성 평가 모델 연구, 한국과학기술정보연구원.
- 김나연, 정은경 (2020). 사회과학 분야 연구자의 데이터요구와 데이터 재이용 행위에 관한 연구. 정보관리학회지, 37(4), 1-26.
- 김문정, 김성희 (2015). 과학기술분야 연구자의 연구데이터 공유의 영향요인에 대한 연구. 한국문헌정보학회지, 49(2), 313-334.
- 김선 (2022). 생명공학 분야 연구자의 연구데이터 공유 의도에 영향을 미치는 요인에 관한 연구: 학술적 평판의 조절효과를 중심으로. 정보관리학회지, 39(1), 45-68.
- 김은정, 남태우 (2012). 연구데이터 수집에 영향을 미치는 요인 분석. 정보관리학회지, 29(2), 27-44.
- 김지현 (2012). 대학내 연구자들의 연구데이터 관리에 대한 연구. 한국도서관·정보학회지, 43(3), 33-45.
- 박성은, 고영만 (2022). 국가 연구데이터플랫폼과 바이오 연구데이터플랫폼의 메타데이터 상호운용성에 관한 연구. 정보관리학회지, 39(2), 159-202.
- 박지원, 장우권 (2021). 사회과학분야 학술 연구자의 연구데이터 재이용 영향요인 연구. 정보관리학회지, 38(4), 199-230.
- 박진호, 고영만, 김현수 (2019). 연구데이터 서비스의 유용성 평가 모형 연구. 정보관리학회지, 36(4), 129-159.
- 송백화 (2018). 화학분야의 연구데이터 공유에 관한 연구. 석사학위논문, 중앙대학교 대학원 문헌정보학과.
- 우종필 (2012). 구조방정식모델 개념과 이해. 서울: 한나래.
- 이상환, 심원식 (2009). 과학데이터의 공유와 활용 (KISTI 지식리포트 제5호), 한국과학기술정보연구원. 출처: <https://repository.kisti.re.kr/handle/10580/7608>
- Andersen, D. L. ed. (2004). Digital scholarship in the tenure, promotion, and review process. NY: M.E. Sharpe, Inc.
- Azeroual, O. & Schöpfel, J. (2019). Quality issues of CRIS data: an exploratory investigation with universities from twelve countries. Publications, 7(1), 14.
- Baron, R. M. & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. Journal of Personality and Social Psychology, 51(6), 1173-1182.
- Bollen, K. A. & Long, J. S. (1993). Testing Structural Equation Models, 154. CA: Sage.
- Borgman, C. L. (2010). Research data: who will share what, with whom, when, and why? In Paper Presented at the Fifth China-North America Library Conference, 1-21. Beijing,

- National Library of China.
- Committee on Ensuring the Utility and Integrity of Research Data in a Digital Age (2009). Ensuring the Integrity, Accessibility and Stewardship of Research Data in the Digital Age. Washington, D.C.: National Academy Press.
- Cragin, M. H. & Shankar, K. (2006). Scientific data collections and distributed collective practice. *Computer Supported Cooperative Work*, 15(2-3), 185-204.
- Faniel, I. M., Kriesberg, A., & Yakel, E. (2016). Social scientists' satisfaction with data reuse. *Journal of the Association for Information Science and Technology*, 67(6), 1404-1416. <https://doi.org/10.1002/asi.23730>
- Gonçalves, R. S. & Musen, M. A. (2019). The variable quality of metadata about biological samples used in biomedical experiments. *Scientific Data*, 6(1), 1-15. <https://doi.org/10.1038/sdata.2019.21>
- Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2006). *Multivariate Data Analysis*. Upper Saddle River, NJ: Pearson Prentice Hall.
- Hedstrom, M. & Niu, J. (2008). Incentives for Data Producers to Create "Archive-Ready" Data: Implications for Archives and Records Management. *Society of american archivists - 2008 Research Forum*.
- Hsu, C. L. & Lin, J. C. C. (2008). Acceptance of blog usage: the roles of technology acceptance, social influence and knowledge sharing motivation. *Information & Management*, 45(1), 65-74.
- Joo, S., Kim, S., & Kim, Y. (2017). An exploratory study of health scientists' data reuse behaviors: examining attitudinal, social, and resource factors. *Aslib Journal of Information Management*, 69(4), 389-407.
- Kankanhalli, A., Tan, B. C. Y., & Wei, K. K. (2005). Contributing knowledge to electronic knowledge repositories: an empirical investigation. *MIS Quarterly*, 29(1), 113-143.
- Kim, B. & Han, I. (2009). The role of trust belief and its antecedents in a community-driven knowledge environment. *Journal of the American Society for Information Science and Technology*, 60(5), 1012-1026.
- Kim, J. (2007). Motivating and impeding factors affecting faculty contribution to institutional repositories. *Journal of Digital Information*, 8(2), 1-11.
- Kim, Y. & Stanton, J. M. (2016). Institutional and individual factors affecting scientists' data-sharing behaviors: a multilevel analysis. *Journal of the Association for Information Science and Technology*, 67(4), 776-799. <https://doi.org/10.1002/asi.23424>

- Kline, R. B. (2016). Principles and Practice of Structural Equation Modeling. New York: Guilford publications.
- Marc, D. T., Beattie, J., Herasevich, V., Gatewood, L., & Zhang, R. (2017). Assessing Metadata Quality of a Federally Sponsored Health Data Repository. AMIA ... Annual Symposium proceedings. AMIA Symposium, 2016, 864-873.
- Piwowar, H. A., Day, R. S., & Fridsma, D. B. (2007). Sharing detailed research data is associated with increased citation rate. PLoS ONE, 2(3), 1-5.
- Rousidis, D., Garoufallou, E., Balatsoukas, P., & Sicilia, M. A. (2014). Metadata for Big Data: a preliminary investigation of metadata quality issues in research data repositories. Information Services and Use, 34(3-4), 279-286.
- Sass, D. A. & Smith, P. L. (2006). The effects of parceling unidimensional scales on structural parameter estimates in structural equation modeling. Structural Equation Modeling, 13(4), 566-586.
- Sterling, T. D. & Weinkam, J. J. (1990). Sharing scientific data. Communications of the ACM, 33(8), 112-119.
- Tenopir, C., Christian, L., Allard, S., & Borycz, J. (2018). Research data sharing: practices and attitudes of geophysicists. Earth and Space Science, 5, 891-902.
- Wiig, K. M. (1997). Knowledge management: where did it come from and where will it go?. Expert Systems with Applications, 13(1), 1-14.
- Yoon, A. (2017). Data reusers' trust development. Journal of the Association for Information Science and Technology, 64(8), 946-956.
- Yoon, A. & Lee, Y. (2019). Factors of trust in data reuse. Online Information Review, 43(7), 1245-1262. <https://doi.org/10.1108/OIR-01-2019-0014>
- Zimmerman, A. (2008). New knowledge from old data: the role of standards in the sharing and reuse of ecological data. Science, Technology, & Human Values, 33(5), 631-652.

• 국문 참고자료의 영어 표기

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

- Bak, ji won & Chang, Woo kwon (2022). A study on factors affecting the reuse of research data by academic researchers in the social sciences. Journal of the Korean Society for Information Management, 38(4), 199-230.
- Kim, Eun-Jeong & Nam, TaeWoo (2012). Factor analysis of effects on research data collection.

- Journal of the Korean Society for Information Management, 29(2), 27-44.
- Kim, Jihyun (2012). A study on university researchers' data management practices. Journal of Korean Library and Information Science Society, 43(3), 433-455.
- Kim, Moon Jung & Kim, Seonghee (2015). A study on the factors affecting sharing of research data of science and technology researchers. Journal of the Korean Society for Library and Information Science, 49(2), 313-334.
- Kim, Nayon & Chung, Eunkyung (2020). An investigation on data needs and data reuse behavior in the field of social sciences. Journal of the Korean Society for Information Management, 37(4), 1-26.
- Kim, Sun (2022). An exploratory study of biotechnology scientists' research data sharing intention: the moderating effects of academic reputation. Journal of the Korean Society for Information Management, 39(1), 45-68.
- Ko, Young-Man, Ahn, Hye-yeon, Kim, Hyun Soo, & Shim, Ji-Woo (2019). Study on a model for metadata quality and usability assessment of the national research data platform. Korea Institute of Science and Technology Information.
- Lee, Sang-Hwan & Shim, Wonsik (2009). Sharing and utilization of scientific data (KISTI knowledge report vol.5), KISTI. Available: <https://repository.kisti.re.kr/handle/10580/7608>
- Park, Jin Ho, Ko, Young-Man, & Kim, Hyun Soo (2019). A study on evaluation model for usability of research data service. Journal of the Korean Society for Information Management, 36(4), 129-159.
- Park, Seong-Eun & Ko, Young Man (2022). A study on metadata interoperability between the national research data platform and the bio research data platform. Journal of the Korean Society for Information Management, 39(2), 159-202.
- Song, Baek-hwa (2018). Research on the Sharing of Research Data in the Field of Chemistry. MA. thesis, Dept. of LIS., Graduate School of Chung-Ang University.
- Yu, jongpil (2012). Structural Equation Model Concepts and Understanding. Seoul: Hannarae.