

키워드 네트워크 분석을 활용한 연구데이터 분야 동향 분석*

- SCOPUS DB를 중심으로 -

Analyzing Trends in Research Data Using Keyword Network Analysis: Focusig on SCOPUS DB

금 효 진 (Hyojin Geum)**

김 선 태 (Suntae Kim)***

초 록

본 연구는 최근 15년간의 연구데이터 관련 연구 현황을 파악하기 위하여 2010년부터 2024년까지의 연구데이터 학술논문의 연구 동향을 분석하고자 하였다. 목적을 달성하고자 Scopus DB에 게재된 학술논문 14,921편을 대상으로 키워드 빈도 분석 및 네트워크 중심성 분석을 수행하였다. 학술지 게재 시기에 따라 1기(2010-2014년), 2기(2015-2019년), 3기(2020-2024년)로 구분하여 UCINET을 활용한 키워드 네트워크 분석을 수행한 결과, 시기에 상관없이 연구되는 주요 키워드와 기간별로 주목받는 키워드, 시간이 지나면서 관심이 줄어드는 키워드를 도출하였다. 최근 15년간 연구데이터 관련 연구가 가장 활발히 이루어진 주제는 데이터 공유인 것으로 확인되었으며, 연결 중심성이 높은 키워드들이 대부분 매개 중심성 또한 높은 것으로 나타났다. 본 연구의 결과는 향후 국내 연구데이터 분야의 연구 방향성을 제시하는 기초자료로서 활용될 수 있을 것으로 판단된다.

ABSTRACT

This study aimed to analyze the research trends of research data academic papers from 2010 to 2024 to understand the research status of research data over the past 15 years. To achieve this goal, keyword frequency analysis and network centrality analysis were conducted on 14,921 academic articles published in Scopus DB. The keyword network analysis using UCINET, which was divided into the first period (2010-2014), second period (2015-2019), and third period (2020-2024) according to the period of publication of academic journals, revealed the main keywords studied regardless of the period, the keywords that attracted attention by period, and the keywords that decreased in attention over time. It was found that the most active topic of research data-related research in the last 15 years is data sharing, and most of the keywords with high Degree Centrality also have high Betweenness Centrality. The results of this study can be utilized as a basis for suggesting future research directions in the field of research data in Korea.

키워드: 연구데이터, 연구 동향, 키워드 네트워크 분석, UCINET, 중심성

Research Data, Research Trend, Keyword Network Analysis, UCINET, Centrality

* This research was supported by National University Development Project at Jeonbuk National University in 2023. 이 논문은 2023년 전북대학교 국립대학 육성사업의 지원을 받아 수행되었음.

** 전북대학교 기록관리학과 석사과정(0133sky@jbnu.ac.kr) (제1저자)

*** 전북대학교 문헌정보학과 부교수, 연구데이터융복합연구소장(kim.suntae@jbnu.ac.kr) (교신저자)

논문접수일자 : 2024년 5월 15일 논문심사일자 : 2024년 5월 19일 게재확정일자 : 2024년 6월 13일
한국비블리아학회지, 35(2): 85-108, 2024. <http://dx.doi.org/10.14699/kbiblia.2024.35.2.085>

※ Copyright © 2024 Korean Biblia 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

전 세계적으로 연구성과 및 과정을 개방하는 오픈 사이언스, 오픈 데이터 정책이 주목받고 있다. 이는 새로운 학술적 발견에 도움이 될 뿐만 아니라 연구결과에 대한 교차 검증을 통한 신뢰도 상승, 중복 연구 진행으로 인한 연구비 지출 방지에도 도움이 될 수 있다.

연구 경쟁의 가속화, 연구 인력의 증가, IT의 눈부신 발전 등으로 인하여 연구데이터의 양이 기하급수적으로 증가하고 있으며(신은자, 2015), 이에 따라 개방된 연구데이터를 체계적으로 축적 및 관리하고 공개·공유할 수 있도록 관련 법과 제도, 시스템의 마련이 요구되고 있다.

이러한 요구에 따라 주요 선진국들은 공적자금을 투입한 연구과제의 연구데이터를 공유 및 활용할 수 있도록 정책을 추진하고 관련 연구를 진행하고 있다. 미국은 경쟁력 강화법(America Competes Reauthorization Act), 호주는 책임 있는 연구수행을 위한 호주 규약(Australian Code for the Responsible Conduct of Research)을 시행 중에 있고, 중국은 과학데이터제출입법과 과학데이터공유조례를 입법하였다(윤종민, 김규빈, 2013). 그 밖에도 영국은 디지털 큐레이션 센터(Digital Curation Center, DCC)를 중심으로 데이터 정책을 분석하는 연구들을 수행하는 등(김지현, 2013) 국가 수준에서 연구데이터 관리 중장기 계획을 수립하고 제도적 측면을 강화해 나가고 있는 것으로 확인된다.

국내 역시 연구데이터를 중요한 국가자산으로 인식하기 시작하여 2018년 국가연구데이터 공유활용전략 수립, 2019년 연구데이터관리계획(Data Management Plan, DMP) 시범 적용,

2022년 ‘국가연구데이터 관리 및 활용 촉진에 관한 법률안’을 제안하는 등 연구데이터 관리에 대한 관심을 기울이고 있다.

국내에 발표된 연구데이터 논문은 2010년 이전에 1편, 2010년부터 2019년까지 39편이 발표되었는데(한상우, 2023), 이를 통해 연구데이터 관련 국내 연구는 2010년대에 들어 본격적으로 시작되었다는 것을 알 수 있다. 이로 인해 연구데이터 관련 국내 연구는 다른 분야에 비해 그 수가 적은 편이며 연구 동향 파악 진행에도 어려움이 있다. 반면, 국외의 경우 국내보다 연구데이터의 중요성에 일찍 주목하여 관련 연구를 많이 수행하고 있다.

이러한 사회적 배경에 따라 본 연구는 키워드 네트워크 분석을 통해 연구데이터 관련 연구 동향을 기간별로 비교·분석하고, 이를 바탕으로 연구데이터 분야의 연구 현황 파악과 시사점 도출을 목적으로 한다. 이를 위한 연구 질문은 다음과 같다.

- 연구 질문 1: 연구데이터 관리에 관한 키워드 빈도와 연결중심성 및 매개중심성을 구한 결과는 어떠한가?
- 연구 질문 2: 연구데이터 관련 연구의 시기별 주요 연구주제는 무엇이고 어떻게 달라졌는가?
- 연구 질문 3: 연구데이터 관리에 관한 중심성 분석 결과를 통하여 도출되는 시사점은 무엇인가?

2. 선행연구

연구데이터 관리 연구 동향 분석과 관련한

국내 연구는 한상우(2023)의 논문이 있다. NetMiner 4를 이용한 빈도 분석, 연결 중심성 및 매개 중심성 분석, 응집구조 분석을 통해 국내 연구데이터 분야 연구 동향을 파악하였다. 연구 결과로 ‘연구데이터관리’, ‘연구데이터공유’, ‘데이터리포지터리’와 같은 주제로 연구데이터 관련 국내 연구가 지속적으로 증가하고 있으나 아직은 그 수가 다소 부족함이 지적되었다. 더불어 ‘연구데이터품질관리’, ‘데이터참조모델’, ‘데이터논문(출판)’ 주제 등 연구데이터 관리 관련 분야에서 더욱 다양한 주제의 연구가 진행되어야 한다고 주장하였다. 또한, 연구 동향 분석 대상을 국내 학술지로 한정했기 때문에 분석 대상에 해외 학술지를 포함한 후속 연구를 제안하였다.

이혜경, 이용구(2023)는 오픈 데이터 관련 연구의 최근 동향을 파악하고자 하였다. 오픈 데이터와 관련한 개념인 오픈 데이터, 링크드 오픈 데이터, 공공 데이터와 같은 다양한 개념을 정리하고 그들 간의 관계를 정리한 후, Scopus에서 1999년부터 2023년까지의 6,543건의 논문을 수집하였고 수집한 논문은 저자키워드를 대상으로 중심성 도출 및 네트워크 분석을 수행하였다. 분석 결과 오픈 데이터 관련 연구는 2009년부터 지속적으로 증가하였으나 2019년 정점 이후 정체 시기에 돌입하였다. 오픈 데이터 관련 연구에서는 ‘big data’가 가장 높은 중심성을 보였으며, 링크드 오픈 데이터 관련 연구는 삼각매개중심성 및 최근접이웃중심성에서 모두 ‘semantic web’이 가장 높은 것으로 나타났다. 주로 상위권에 등장하는 키워드는 정부와 관련한 키워드로 나타났는데, 정부 정책의 공공 데이터보다는 데이터 연계 및 관계의 형성을 중

심으로 연구가 수행된 것으로 확인되었다.

연구데이터와 관련한 주제가 아니지만 국내의 동향 분석을 수행한 연구로는 구본진, 장덕현(2023), 이진우, 박동명(2023), 그리고 정선경(2022)의 논문을 살펴볼 수 있다. 구본진, 장덕현(2023)은 동시출현단어분석, 주성분분석 및 상관관계분석을 활용하여 추후 문헌정보학의 유망 연구 주제를 도출하였다. 이진우, 박동명(2023)은 상관관계분석을 반복적으로 수행하여 적절한 수준의 유사성 집단을 찾아내는 CONCOR 분석을 통해 박물관 마케팅에 관한 국내외 연구 동향을 분석하였고 재규정된 박물관 역할에 따라 변화할 향후 관련 연구의 학술적 방향을 제시하였다. 정선경(2022)은 네트워크 분석 및 토픽모델링을 통해 데이터 거버넌스의 연구 경향을 파악함과 동시에 관련 연구 분야에서 적용하기 위한 시사점을 제시하였다.

국외에서는 Sheriff, Sevukan(2023)이 810개의 채용광고에서 수집한 연구데이터 관리(Research Data Management, RDM) 관련 용어에 대해 텍스트 마이닝을 실시하였다. 그 결과 RDM 교육에 필요한 세 가지 주요 컨텍스트인 ‘도서관 연구 서비스’, ‘연구데이터’, ‘직장 환경’을 파악할 수 있었으며, 데이터 관리자를 시장에 신속하게 공급할 수 있도록 기술 향상 프로그램 및 데이터 관리 교육의 도입을 주장하였다. 해당 연구는 RDM 업무를 장려하기 위하여 필요한 요소와 이에 대한 상세한 시사점을 도출하였다. 또한 Sheriff, Sevukan(2022)은 R 패키지 Biblioshiny, ScientoPy 및 VOSviewer를 이용하여 전세계를 범위로 1926년부터 2020년까지의 연구데이터 관리 연구의 일반적인 분석을 진행하였다. 해당 연구에 따르면, 주요 주

제는 도서관 연구 서비스, 연구데이터, 직장 환경이고 미국, 영국, 캐나다, 독일과 같은 국가가 연구데이터 관리 업무 및 서비스의 주요 참여자였다. 주요 연구 기관은 Harvard, Stanford, Massachusetts Institute of Technology와 같은 공신력 있는 기관이었다. 연구데이터 관련 연구가 이루어지는 주 학술영역과 주로 연구하는 국가, 각 국가에서의 데이터 관리 역량을 파악하는 기초자료로서 그 의미가 있다. Chunlai Yan et al.(2022)은 CiteSpace와 VOSviewer를 이용하여 중국 내 연구데이터 분야 연구의 현재 동향을 파악하였고, 이는 동향 파악에 그치지 않고 중국 내에서 이루어질 향후 연구가 고려해야 할 구체적인 측면을 제시하였다는 점에서 의의를 가진다. 제시한 사안으로는 연구데이터의 조직 간, 단계 간 통합에 대한 도서관의 참여와 연구데이터 공유 플랫폼의 기능 개선, 데이터 리터러시 교육 방법 및 모델의 실천, 그리고 연구데이터 서비스 개선이 있다. Kwanya (2021)는 Harzing의 “Publish or Perish” 소프트웨어와 VOSviewer를 이용하여 사하라 이남 지역의 연구데이터 관련 학술 출판물을 분석하였고, 그 결과 저자 협력이 부족하고 출판물의 가시성이 떨어져 그 양과 품질의 수준이 다소 부족함을 지적하였다. Zang, Eichmann-Kalwara (2019)는 CiteSpace를 이용하여 연구데이터 관리의 네트워크 구조와 중요 주제 및 개발 추세를 도출하였고, 학제 간 분야에서 과학 협력 연구의 클러스터가 활발함을 언급하였다.

본 연구는 2010년부터 2024년 사이 Scopus DB에 등재된 연구데이터 관련 연구 동향을 분석하여 시사점을 도출하고, 연구 현황을 파악하는 것이 목적이다. 기존에 진행된 연구데이

터 관련 선행연구 중 키워드 네트워크 분석을 활용한 연구는 다수 존재하나 국내 연구는 소수에 불과하며, 국외 연구는 다소 한정된 키워드로 데이터를 수집하였다는 점에서 한계를 지닌다. 따라서 보다 더 다양한 키워드로 수집할 수 있는 연구데이터 관리와 관련된 논문이 분석될 필요가 있다는 점에서 본 연구는 의의가 있다. 더불어 연구데이터와 관련된 연구의 흐름을 파악하기 위해서는 비교적 최근 자료를 수집하여 이를 시기별로 분석할 필요가 있다. 이에 본 연구에서는 기존의 동향 분석 논문을 보완하기 위해 연구데이터와 관련한 다양한 키워드를 최대한 반영하여 학술논문을 수집하고, 수집한 논문을 5년 주기로 나누어 시기별 연구 동향을 확인할 수 있도록 진행하고자 한다.

다음 <표 1>은 국내외의 동향 분석 선행연구를 최신 발행연도순으로 정리한 것이다.

3. 연구 설계

본 연구는 2010년부터 2024년까지 Scopus DB에 게재된 연구데이터 관련 학술논문의 연구현황을 파악하고, 저자 키워드를 바탕으로 빈도 분석과 연결 중심성 분석 및 매개 중심성 분석을 실시하였다.

3.1 키워드 네트워크 분석

키워드 네트워크 분석(Keyword Network Analysis)은 텍스트 내에서 의미있는 단어를 추출하고, 추출한 단어 중 핵심 역할을 하는 키워드를 선정하여 텍스트 내에서 구성되는 키워

〈표 1〉 국내외 선행연구

저자(발행연도)	연구 주제	분석 방법
Sheriff, Sevukan(2023)	- RDM 연구의 핵심 요구사항 도출	토픽 모델링(LDA)
구분진, 장덕현(2023)	- 문헌정보학 분야의 유망 연구주제 파악 및 향후 주요 연구주제 선정	키워드 네트워크 분석
이진우, 박동명(2023)	- 키워드 확인을 통한 박물관 마케팅의 연구 동향 파악과 연구 확장을 위한 연계방안 제시	키워드 네트워크 분석
이혜경, 이용구(2023)	- 오픈 데이터 관련 연구의 최근 동향과 지적 구조 고찰	키워드 네트워크 분석
한상우(2023)	- 연구데이터 관리 분야의 주요 키워드와 연구 현황 파악	키워드 네트워크 분석
Sheriff, Sevukan(2022)	- 1926년-2020 문헌을 대상으로 전세계 범위의 RDM 연구 환경 및 추세 분석	키워드 네트워크 분석
Chunlai Yan et al.(2022)	- 중국의 연구데이터 관리 분야 연구주제 분석 및 추후 연구 개선사항 제시	키워드 네트워크 분석
정선경(2022)	- 데이터 거버넌스 플랫폼 연구의 주요 연구주제 도출 및 연구 확대 양상 파악	키워드 네트워크 분석, 토픽 모델링(LDA)
Kwanya(2021)	- 사하라 이남 아프리카 지역의 연구데이터 관리 학술 출판물 현황 분석 및 연구 환경 개선방안 제안	키워드 네트워크 분석
Zang & Eichmann-Kalwara (2019)	- RDM 학술 출판물 분석을 통한 주요 연구주제 확인	키워드 네트워크 분석

드간 연결관계를 파악하고 다양한 특성을 분석해내는 것이다(이수상, 2014). 해당 용어는 네트워크 분석, 언어 네트워크 분석으로도 표현되며 본 연구에서는 키워드 네트워크 분석으로 통일하여 기술한다. 노드는 분석 대상 개체를, 링크는 개체 간의 연결 관계를 의미한다. 본 연구에서는 ‘저자 키워드’를 노드로, ‘키워드 간의 관계’를 링크로 정의하여 진행하였다. 키워드 네트워크 분석의 핵심은 두 노드 사이의 상호관계성을 파악하는 것으로, 연결 관계 데이터를 표현한 네트워크는 노드와 링크의 모양, 색상, 크기 등을 다르게 설정함으로써 다양한 표현의 시각화가 가능하다. 키워드 네트워크 분석은 시각화된 네트워크의 형태, 밀도, 중심성을 파악하는 분석기법으로서 연구의 주제 및 핵심 키워드 간 연결 관계와 그 의미, 개념구조 등의 특징을 시각적으로 파악하고 해석할 수 있어 특정 분야의 연구 동향을 분석하는데 효과

적인 방법론으로 간주되며, 계량정보분석의 중요한 분석기법으로 활용되고 있다(이수상, 2014; Popping, 2000). 이러한 특징으로 인하여 연구 동향을 분석한 국내외의 논문 다수가 프로그램을 이용한 키워드 네트워크 분석 방법을 사용한다. 연구자는 키워드 네트워크 분석을 통하여 주제 분야의 전체적인 연구 동향 파악과 동시에 어떠한 주제 영역에 연구가 집중되는지, 어떠한 키워드가 핵심적 키워드로 작용하는지를 면밀히 살펴볼 수 있다.

3.2 중심성 분석

중심성(centrality)은 사회 네트워크의 분석에서 가장 많이 사용되는 지표로, 하나의 노드가 전체 네트워크에서 중심에 위치하는 정도를 표현한다. 중심성 분석을 통해 한 네트워크에서 중요한 역할을 하는 노드가 무엇인지, 각 노

드는 네트워크 내의 핵심에 어느 정도 접근하고 있는지를 확인할 수 있다.

중심성의 유형은 여러 가지로 나눌 수 있는데, Freeman이 제안한 가장 기본적인 지표인 연결 중심성(degree centrality), 근접 중심성(closeness centrality), 매개 중심성(betweenness centrality), 위세 중심성(prestige centrality) 4 가지를 들 수 있다(Freeman, 1978). 본 연구에서 다룰 중심성 분석은 연결 중심성과 매개 중심성이다.

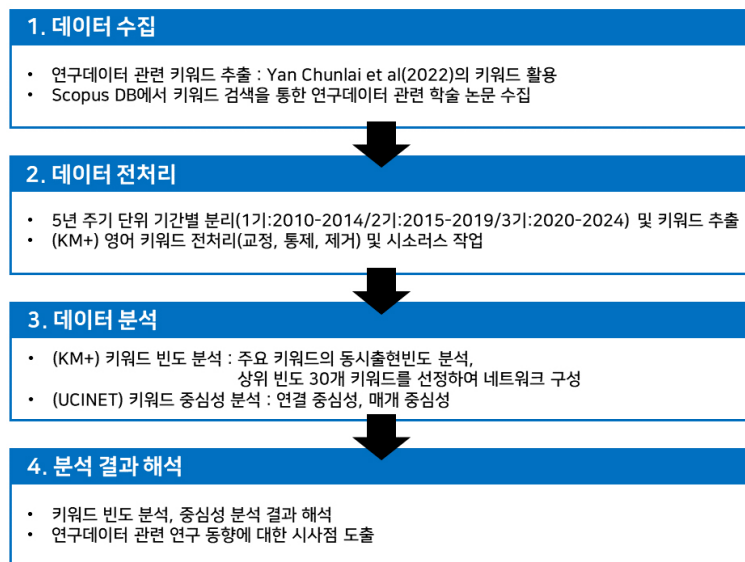
연결 중심성은 네트워크의 노드들이 다른 노드들과 얼마나 많은 연결 정도를 갖고 있는지로 파악된다. 즉, 어느 한 노드에 직접 연결된 다른 노드의 수를 나타낸다(김용학, 김영진, 2011). 따라서 연구 동향 분석 연구에서 연결 중심성은 네트워크 내의 노드가 가지는 직접적인 영향력을 나타내는 지표라고 할 수 있으며(김우주, 2015), 한 네트워크 내에서 큰 연결 중심성

을 갖는 노드는 해당 분야에서 핵심 역할을 하는 키워드로 간주할 수 있다.

매개 중심성은 네트워크 상의 두 노드 사이의 최단 경로를 이용하는 개념으로, 한 노드가 네트워크를 구축함에 있어서 다른 노드들과의 중개 역할을 얼마나 많이 수행하고 있는지에 따라 높은 값을 가진다. 즉, 높은 매개 중심성 값을 가지는 노드는 다른 노드들이 많이 거치는 노드이며 정보에 대한 영향력이 큰 노드이다(김병선 외, 2015). 따라서 연구 동향 분석 연구에서 매개 중심성이 큰 노드는 다양한 연구 키워드의 매개 역할을 하는 것으로 이해할 수 있다.

3.3 연구 방법 및 절차

다음 <그림 1>은 본 연구의 진행 방법 및 절차를 간단히 정리하고 이를 도식화하여 나타낸 것이다.



<그림 1> 연구 프로세스 도식화

먼저, 문헌분석을 통해 학술논문 수집 시에 사용할 검색식에 필요한 연구데이터 관련 키워드를 도출한다. 기존 선행연구에서 확인할 수 있는 연구데이터 관련 개념을 바탕으로 논문 수집에 필요한 검색 키워드를 선정하였다.

선정한 키워드는 학술논문 DB에서 OR 연산자를 통해 연결하여 키워드 검색에 활용한다. 주로 학술논문만을 분석 대상으로 선정한 이유는 학술지가 학계의 연구 동향을 가장 함축적으로 잘 보여주는 연구물이며, 학술적 연구 성과를 공유할 수 있는 주된 교류 경로가 되기 때문이다(김소연, 이은주, 2022). 본 연구에서는 학술논문 검색을 위한 데이터베이스로 Scopus를 선정하였다. Scopus는 국제적으로 그 권위를 인정받고 있는 인용색인 데이터베이스로서 전 세계적으로 다양한 분야의 연구 성과를 게재하고 있어 학술지 검색에 적합하며, 네트워크 분석에 필요한 서지사항 추출이 편리하여 자료 수집이 용이하다는 장점을 지니고 있다. 검색 기간의 범위는 2010년부터 2024년까지로

설정하였고, 검색을 통해 수집한 논문은 발행연도 기준 2010-2014년 구간을 1기, 2015-2019년 구간을 2기, 2020-2024년 구간을 3기로 구분하여 서지사항을 추출하였다.

서지사항 데이터를 대상으로 유의미한 결과를 도출하기 위해 분석대상에 대한 주제적 특성을 잘 나타내는 키워드만을 선택할 필요성이 있다고 판단하여 서지사항 분석에 앞서 데이터 전처리를 수행하였다. 데이터 전처리 과정은 다음과 같다.

키워드가 주제적 특성을 얼마나 잘 나타내는지 판단하는 대표적 기준인 출현빈도(occurrence)를 활용하여 분석 대상 키워드를 추출한다. 다음으로 네트워크 분석의 오류를 최소화하기 위해 키워드를 정제하였다. 본 연구에서는 자체적으로 시소러스를 제작 및 적용함으로써 키워드를 선별하였다. 키워드 정제 작업은 교정작업, 통제작업, 그리고 제거작업으로 구분한다(이수상, 2014). 다음 <표 2>는 키워드 정제 작업에 대한 설명과 그 예시를 정리한 것이다.

<표 2> 키워드 정제 작업 설명 및 예시

정제	설명	예시
교정작업	• 단/복수, 약어, 띄어쓰기, 품사 통일	<ul style="list-style-type: none"> • 단/복수: academic library(단수), academic libraries(복수) → 'academic library' • 약어: application programming interfaces → API • 띄어쓰기: research-data management, researchdata management → 'research data management' • 품사: reproducibility(명사), reproducible(형용사) → 'reproducibility'
통제작업	• 동의어, 유사어, 광의어 등을 하나의 적절한 대표어로 통일	• FAIRness, FAIR priciples, FAIR data → 'FAIR data principle'
제거작업	<ul style="list-style-type: none"> • 고빈도 단어 중 주제적 특성과 관계없는 일반적 키워드 삭제 • 타 키워드와 과도하게 연결되어 분석 결과에 지장을 줄 수 있는 핵심 키워드 삭제 	<ul style="list-style-type: none"> • 국가명 삭제: 'Canada', 'China' 등 • 일반 어휘 삭제: 'object', 'analysis' 등 • 핵심 키워드 삭제: 'research', 'data', 'research data' 등

데이터 추출과 <표 2>의 키워드 정제 작업을 거쳐 최종 분석 대상 키워드를 선정했다면 네트워크 구성을 위하여 키워드 간 관계를 파악할 필요가 있다. 키워드 간의 관계는 대체로 동시출현빈도(Co-occurrence)를 통해 파악할 수 있으며, 이를 위한 분석도구로 KM+1)를 사용하였다. 동시출현빈도는 대상 키워드가 특정 시기 범위 내에서 동시에 출현하는 빈도이다.

마지막으로, 구성된 네트워크에 대한 중심성 분석을 통해 연구 동향을 파악하고 시사점을 도출한다. 본 연구에서 분석한 사항은 동시출현빈도별 키워드 빈도 분석 결과, 그리고 기간별 키워드 중심성 분석 결과이다.

4. 데이터 수집 및 분석

연구데이터 관련 연구 동향 분석을 위하여 Scopus에서 학술논문의 발행 일자를 지난 15년간인 2010년부터 2024년까지로 제한하고, 검색 문헌의 유형은 Article, Conference paper, Review로 제한하여 검색하였다. 검색어 필드는 키워드(keyword)로 한정하였으며, 검색식은 “research data”, “research data manag*”, “research data shar*”, “research data preserv*”, “research data reposit*”, “data reposit*”, “data curat*”, “digital curat*”, “data management plan”, “data lifecycle”, “data literacy”, “data reus*”, “open science”, “data paper”, “data article”, “scientific data”, “re3data”, “fair data”의 18개 키워드를 사용하였다. 18개의 키워드

는 선행연구 중 Sheriff, Sevukan(2023), Chunlai Yan et al.(2022), 그리고 Zang, Eichmann-Kalwara(2019)의 연구에서 사용된 키워드 및 검색식을 참고하여 선정되었다. 해당 검색어를 사용하여 2024년 4월 3일 Scopus에서 수집한 연구데이터 관리 관련 국외 학술논문은 14,921편으로, 최종 선정된 학술논문은 2010-2014, 2015-2019, 2020-2024 기간별로 나누어 서지사항을 추출하였다. 이후 KM+ 프로그램을 통해 키워드 정제 및 키워드별 출현빈도를 확인하고, UCINET 프로그램을 활용한 키워드 네트워크 중심성 분석과 시각화를 진행하여 기간별 연구 동향을 비교하였다.

4.1 학술논문 출판현황

기간별 학술논문의 수는 1기(2010년-2014년)에 3,136편, 2기(2015년-2019년)에 4,809편, 3기(2020년-2023년)에 6,976편이었다. Scopus 학술논문에 대한 자세한 통계는 <표 3>에서 확인할 수 있다.

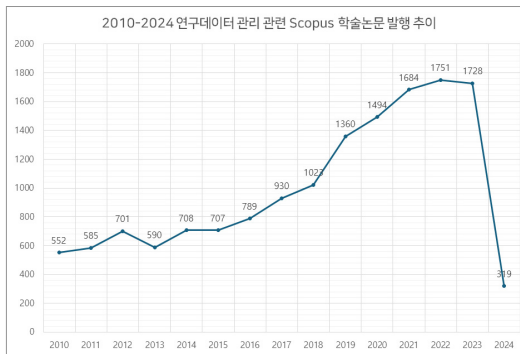
연구데이터 관련 논문의 연도별 편수 추이는 <그림 2>와 같으며 2012년~2013년 구간을 제외한 모든 기간에서 발행 편수가 꾸준히 증가하였음을 확인할 수 있다.

2010년~2019년 사이는 매년 100~200편씩 편수가 증가하였으나, 2018년에서 2019년으로 넘어가는 시기에 약 300편 가량으로 발행 편수가 크게 증가하였다. 2022년에는 1,751편의 논문이 투고되어 가장 많은 논문이 발행된 것으로 확인되었다.

1) 정식 명칭은 KnowledgeMatrixPlus로, KISTI가 2007년에 개발하여 보급한 데이터 전처리 소프트웨어이다.

〈표 3〉 2010-2024 Scopus 학술논문 검색결과

연도	논문 수	연도	논문 수	연도	논문 수
2010	552	2015	707	2020	1,494
2011	585	2016	789	2021	1,684
2012	701	2017	930	2022	1,751
2013	590	2018	1,023	2023	1,728
2014	708	2019	1,360	2024	319
2010-2014	3,136	2015-2019	4,809	2020-2024	6,976



〈그림 2〉 2010-2024 연구데이터 관련 Scopus 학술논문 발행 추이

〈표 4〉 2010-2024 논문 투고 상위 10개국

순위	국가	편수
1	United States	5,002
2	Germany	1,785
3	United Kingdom	1,553
4	China	1,380
5	Nether-lands	767
6	Italy	736
7	France	694
8	Canada	675
9	Spain	646
10	Australia	621

〈표 4〉는 가장 많은 논문을 투고한 10개 국가의 논문 발표 현황을 정리한 것이다. 가장 많은 논문을 게재한 국가는 미국으로, 전체 논문의 33.5%를 차지하였으며 두 번째로 많은 논문을 투고한 독일이 12%로 그 뒤를 이었다. 이어서 영국, 중국, 네덜란드, 이탈리아 등의 국가가 순위권에 들었으며, 한국은 순위권에 들지는 못하였지만 15년간 265편의 논문을 발표(1.8%)하여 20번째로 많은 논문을 게재하였다. 이는 아시아에서 중국, 인도, 일본 다음으로 많은 논문 수이다.

다음의 〈표 5〉는 가장 많이 투고된 상위 10개의 학술분야를 나타낸 것이다.

컴퓨터과학 분야에서 26.3%로 가장 많은 논

문을 게재했음을 확인할 수 있으며, 그 뒤로는 사회과학, 공학, 의학 등의 학술분야가 뒤를 이었다. 키워드 네트워크 분석의 정확도 향상을 위하여 키워드 정제 시에 해당 분야들과 직접적인 관련이 있는 키워드는 분석 대상에서 제외처리하였다.

4.2 키워드 빈도 분석

전체 기간(2010-2024년) 동안 검색할 수 있는 학술논문 14,921편에서는 총 32,191개의 키워드를 확인할 수 있었다. 키워드의 교정, 통제, 제거 작업을 통해 자체적으로 시소러스를 적용한 후 키워드는 각 기간별로 1기(2010-2014년) 6,525

〈표 5〉 2010-2024 학술논문 투고 상위 10개 분야

순위	분야	논문 수	백분율(%)
1	Computer Science	6,940	26.3
2	Social Sciences	3,071	11.7
3	Engineering	2,640	10.0
4	Medicine	2,224	8.4
5	Mathematics	1,619	6.1
6	Biochemistry, Genetics and Molecular Biology	1,409	5.3
7	Physics and Astronomy	803	3.0
8	Decision Sciences	738	2.8
9	Agricultural and Biological Sciences	717	2.7
10	Environmental Science	679	2.6

개, 2기(2015-2019년) 9,575개, 3기(2020-2024년) 15,087개로 확인되었다.

다음 〈표 6〉은 전처리 작업을 통해서 얻은 정제된 데이터를 이용하여 키워드의 빈도를 구한 결과를 정리한 것이다. 키워드 빈도는 1기(2010-2014년), 2기(2015-2019년), 3기(2020-2024년) 기간별로 값을 구하였으며 기간 내에서 빈도가 가장 높은 10개 키워드를 선별하였다.

〈표 6〉에서 상위 5개 키워드를 기간별로 보면, 1기(2010-2014년)의 경우 'Data repository',

'Data curation', 'Data reuse', 'Data sharing', 'Scientific data'이며, 2기(2015-2019년)에는 'Open science', 'Data repository', 'Data curation', 'Open data', 'Data sharing'이었다. 마지막으로 3기(2020-2024년)의 경우에는 'Open science', 'FAIR data management', 'Data literacy', 'Data curation', 'Open data'이다. 상위 5개 키워드 중에서 전 기간에서 공통으로 나타난 키워드는 'Data curation'이었으며 시간이 흐를수록 빈도 순위는 낮아졌으나 빈도는 115, 208,

〈표 6〉 2010-2024 기간별 키워드 빈도 분석 결과 상위 10개 키워드

2010-2014			2015-2019			2020-2024		
순위	키워드	빈도	순위	키워드	빈도	순위	키워드	빈도
1	Data repository	129	1	Open science	623	1	Open science	1,526
2	Data curation	115	2	Data repository	221	2	FAIR data principles	463
3	Data reuse	103	3	Data curation	208	3	Data literacy	363
4	Data sharing	81	4	Open data	194	4	Data curation	262
5	Scientific data	78	5	Data sharing	188	5	Open data	258
6	Open science	76	6	Data reuse	185	6	Data reuse	246
7	Digital curation	71	7	Open access	148	7	Reproducibility	240
8	Metadata	64	8	Metadata	119	8	Data sharing	236
9	E-science	35	9	Data literacy	116	9	Data repository	234
10	Open data	35	10	Digital curation	112	10	Open access	212

262로 꾸준히 증가하였다.

상위 5개 키워드에 항상 해당하지는 않았지만 'Data repository', 'Data reuse', 'Data sharing', 'Open science', 'Open data' 키워드도 전 기간에 걸쳐 공통으로 등장하였다. 'Data repository', 'Data reuse', 'Data sharing' 키워드의 경우, 'Data curation' 키워드와 마찬가지로 빈도 순위가 낮아지는 경향을 보이거나 빈도는 지속해서 증가하고 있다. 반면 'Open science' 키워드의 경우 다른 키워드와는 달리 빈도가 증가하며 동시에 빈도 순위도 꾸준히 증가한 유일한 키워드로 확인되었다. 또한, 'Open science' 키워드는 3기(2020-2024년)에 들어 빈도 2위의 키워드인 'FAIR data management'보다 약 4배 많은 빈도를 보이며 주목받는 중요 키워드임을 입증하였다. 한편, 전 기간에 걸쳐 꾸준히 언급된 'Open science' 키워드와 달리 3기(2020-2024년) 이전에 상위 10개 키워드에 들지 못하던 'FAIR data management' 키워드는 3기(2020-2024년)에 들어 빈도 2위에 위치하여 새로운 주요 연구 키워드로서 주목받고 있다.

4.3 중심성 분석

본 연구에서는 중심적인 연구 키워드를 탐색하기 위해 연결 중심성 및 매개 중심성 분석을 실행하였다. 중심성 분석의 대상이 되는 키워드는 2010-2024 기간별 키워드 빈도 분석 결과에서 추출하였으며, 최종적으로 각 1기(2010-2014년), 2기(2015-2019년), 3기(2020-2024년)의 키워드 빈도 분석 결과에서 가장 빈도가 높은 30개의 키워드를 선정하였다.

다음 <표 7>은 1기(2010-2014년)의 상위 빈

도 30개 키워드에 대한 연결 중심성 및 매개 중심성 값을 정리한 표이다.

1기(2010-2014년)에는 'Data sharing' 키워드가 가장 높은 연결 중심성 값을 보이는 것을 확인할 수 있으며, 'Data repository', 'Data curation', 'Metadata' 등이 그 뒤를 잇고 있다. 이와 함께, 상위 10개 키워드에는 'Data reuse', 'Digital curation', 'Open data' 등 데이터 공유와 관련한 키워드들이 주로 포함되어 있다. 'Data sharing'과 'Data repository' 키워드는 연결 중심성과 매개 중심성 모두에서 높은 순위를 보임으로써, 주요 연구 주제임과 동시에 다른 주제 간의 매개 역할도 다수 수행하고 있는 키워드임을 알 수 있다. 또한, 시각화를 통해 1기(2010-2014년)에는 연결 중심성이 높은 키워드가 매개 중심성에서도 대부분 높게 측정되는 것으로 파악되었다. <표 7>의 키워드 연결 중심성과 매개 중심성을 시각화한 결과는 다음의 <그림 3>, <그림 4>에서 확인할 수 있다. 네트워크 분석 시각화 결과에서 키워드와 그 연결 관계는 노드와 링크로 나타난다. 노드의 크기가 클수록 해당 키워드의 중심성 분석값이 크다는 것을, 링크의 굵기가 굵을수록 각 노드가 강한 연결 관계가 있다는 것을 의미한다. <그림 3>에선 1기(2010-2014년) 기간 내 높은 연결 중심성을 보인 'Data repository', 'Data curation', 'Metadata' 키워드의 노드가 가장 크게 나타난 것을 알 수 있으며 동시에 다른 다수의 키워드와 연결 관계를 맺고 있음을 확인할 수 있다. <그림 4>는 1기(2010-2014년)의 키워드 매개 중심성을 시각화한 것으로, 매개 중심성 상위 5개 키워드 중에서도 'Data repository' 키워드가 네트워크 내에서 큰 비중을 차지한다는 것을 노드의 크

〈표 7〉 1기(2010-2014년) 키워드 연결 중심성 및 매개 중심성 상세

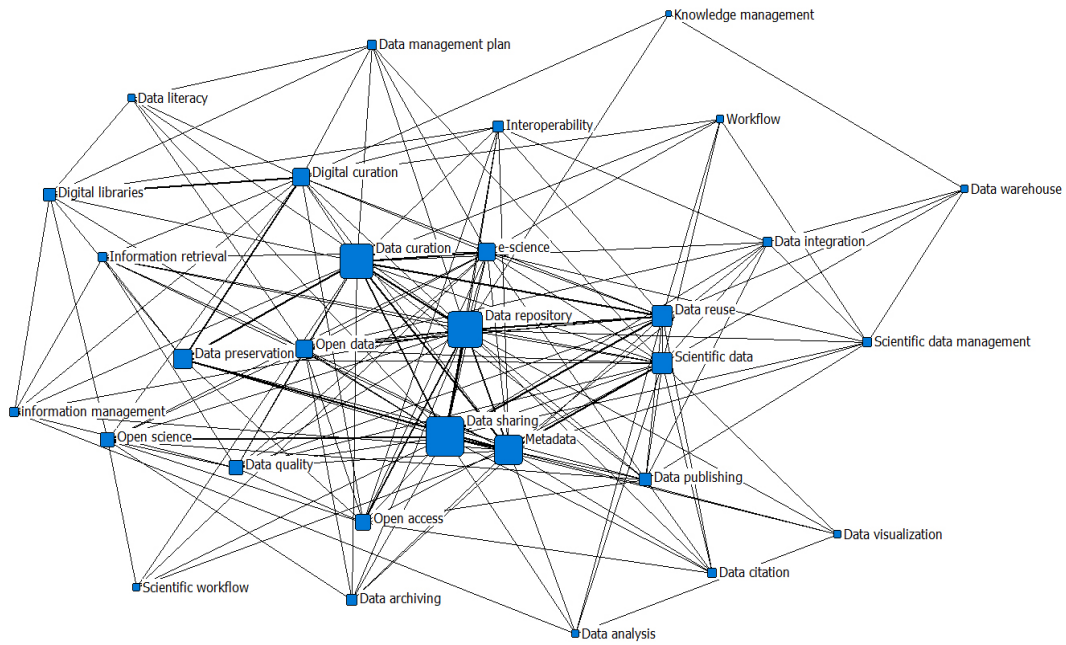
연결 중심성			매개 중심성		
순위	키워드	값	순위	키워드	값
1	Data sharing	96	1	Data repository	51.852
2	Data repository	88	2	Data sharing	37.786
3	Data curation	81	3	Scientific data	29.156
4	Metadata	69	4	Digital curation	22.298
5	Data reuse	44	5	Data curation	22.038
6	Scientific data	43	6	Metadata	21.482
7	Data preservation	39	7	Data reuse	17.948
8	Digital curation	37	8	E-science	14.334
9	E-science	34	9	Open data	8.654
10	Open data	34	10	Open science	5.389
11	Open access	30	11	Data publishing	4.457
12	Data quality	25	12	Data integration	4.212
13	Open science	24	13	Scientific data management	3.958
14	Digital libraries	22	14	Digital libraries	3.709
15	Data publishing	21	15	Open access	3.705
16	Data archiving	16	16	Data preservation	3.663
17	Interoperability	16	17	Data quality	3.291
18	Data citation	13	18	information management	2.85
19	Data integration	13	19	Data warehouse	2.768
20	Scientific data management	13	20	Interoperability	2.42
21	Data management plan	11	21	Data management plan	1.421
22	Information retrieval	11	22	Information retrieval	1.272
23	information management	10	23	Data analysis	1.234
24	Data analysis	8	24	Knowledge management	1.165
25	Data literacy	8	25	Data literacy	0.869
26	Data warehouse	7	26	Workflow	0.785
27	Scientific workflow	7	27	Data archiving	0.475
28	Workflow	7	28	Data citation	0.385
29	Data visualization	6	29	Scientific workflow	0.225
30	Knowledge management	3	30	Data visualization	0.2

기를 통해 직관적으로 확인할 수 있다.

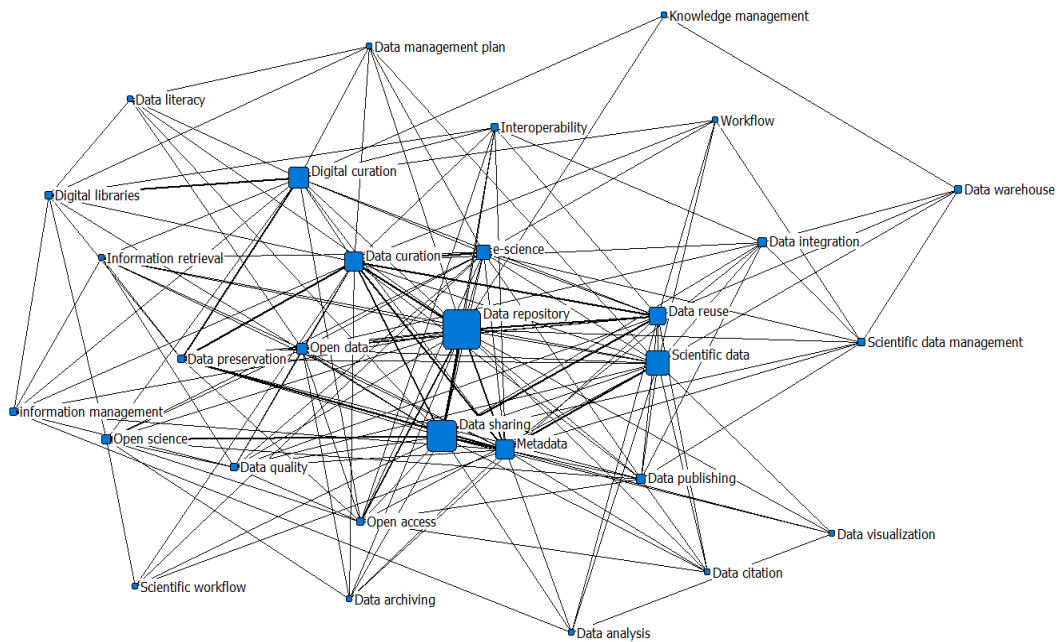
다음 〈표 8〉은 2기(2015-2019년)의 상위 빈도 순위 30개 키워드에 대한 연결 중심성과 매개 중심성 값을 정리한 것이다.

2기(2015-2019년)에서 가장 높은 연결 중심성 값을 가진 키워드는 'Open science'로, 그 뒤

로 'Open data', 'Open access', 'Data sharing' 키워드가 이어져 1기(2010-2014년)에 비해 오픈 사이언스와 관련한 키워드가 많은 관심을 받고 있는 것으로 나타났다. 반면, 매개 중심성에서는 'Metadata'가 첫 번째로 높은 값을 가진 키워드로 확인되었다. 그 뒤로는 'Data curation',



〈그림 3〉 1기(2010-2014년) 키워드 연결 중심성 시각화 결과



〈그림 4〉 1기(2010-2014년) 키워드 매개 중심성 시각화 결과

〈표 8〉 2기(2015-2019년) 연결 중심성 및 매개 중심성 상세

연결 중심성			매개 중심성		
순위	키워드	값	순위	키워드	값
1	Open science	494	1	Metadata	20,646
2	Open data	275	2	Data curation	16,552
3	Open access	224	3	Data sharing	15,734
4	Data sharing	218	4	Open science	12,677
5	Data repository	194	5	Data repository	12,367
6	Metadata	142	6	Open data	11,31
7	Data curation	134	7	Data preservation	9,993
8	Reproducibility	129	8	Open access	9,485
9	Data reuse	109	9	Data quality	7,65
10	FAIR data principles	85	10	Data reuse	6,631
11	Scholarly communication	83	11	Reproducibility	5,82
12	Data quality	64	12	FAIR data principles	5,375
13	Open source	61	13	Scientific data	5,33
14	Interoperability	53	14	Scholarly communication	4,953
15	Data preservation	50	15	Digital curation	4,692
16	Data publishing	49	16	Interoperability	4,514
17	Data science	49	17	Data integration	3,942
18	Digital curation	48	18	Data analysis	3,63
19	Scientific data	47	19	Data science	3,479
20	Peer review	46	20	Open source	2,727
21	Data integration	44	21	Data publishing	2,495
22	Data analysis	43	22	Data collection	1,863
23	Data literacy	43	23	Data lifecycle	1,55
24	Data visualization	36	24	Data literacy	1,521
25	Data access	23	25	Data visualization	1,407
26	Data management plan	22	26	Data management plan	0,518
27	Data collection	18	27	Peer review	0,431
28	Data standards	16	28	Data standards	0,298
29	Data lifecycle	12	29	Scientific data management	0,272
30	Scientific data management	9	30	Data access	0,138

‘Data sharing’ 키워드가 이어져, 1기(2010-2014년)에서 가장 높은 연결 중심성을 보이던 키워드들이 2기(2015-2019년)에서 주로 연구 주제들 간의 매개 역할을 수행하고 있음을 알 수 있다. 흥미로운 사항으로는 연결 중심성에서는 다소 낮은 값을 보인 ‘Data preservation’, ‘Data

quality’ 키워드가 매개 중심성에서는 10위 안의 높은 값을 가졌다는 점이 있다. 이는 데이터 공유와 관련한 연구가 주로 진행되면서 그 매개로서 공유 데이터의 품질에 대한 사항이 함께 언급되었기 때문으로 판단된다. 다음의 〈그림 5〉, 〈그림 6〉은 〈표 8〉의 키워드 연결 중심

다음 <표 9>는 3기(2020-2024년)의 상위 빈도 30개 키워드에 대한 연결 중심성과 매개 중심성 값을 나타낸 것이다.

‘Open science’가 2기(2015-2019년)에 이어 3기(2020-2024년)에서도 가장 높은 연결 중심성 값을 가진 키워드로 확인되었으며, 동시에

매개 중심성에서도 가장 높은 값을 지닌 키워드로 밝혀졌다. 즉, ‘Open science’는 가장 많이 연구자들의 관심을 받는 연구 주제임과 동시에 다른 주제와의 융합에서도 중요한 매개로서 작용하는 키워드로 나타났다. 연결 중심성에서는 ‘FAIR data principles’, ‘Open data’, ‘Open

<표 9> 3기(2020-2024년) 연결 중심성 및 매개 중심성 상세

연결 중심성			매개 중심성		
순위	키워드	값	순위	키워드	값
1	Open science	905	1	Open science	12,398
2	FAIR data principles	394	2	Metadata	12,398
3	Open data	344	3	Data sharing	9,764
4	Open access	303	4	FAIR data principles	9,339
5	Data sharing	296	5	Open data	8,109
6	Reproducibility	290	6	Data repository	7,913
7	Data repository	206	7	Open access	7,361
8	Metadata	180	8	Data curation	6,518
9	Data reuse	177	9	Scientific data	5,702
10	Data curation	146	10	Data science	5,017
11	Interoperability	111	11	Data reuse	4,62
12	Data science	103	12	Data analysis	4,556
13	Scholarly communication	101	13	Data integration	4,31
14	Data quality	82	14	Knowledge graph	3,981
15	Data analysis	81	15	Data collection	3,401
16	Data literacy	81	16	Interoperability	3,259
17	Scientific data	77	17	Data quality	2,751
18	Data integration	62	18	Reproducibility	2,512
19	Data visualization	62	19	Data paper	2,512
20	Dataset	60	20	Dataset	2,06
21	Data Management Plan	58	21	Workflow	1,5
22	Accessibility	45	22	Data literacy	1,365
23	Data collection	43	23	Data visualization	1,266
24	Knowledge graph	42	24	Accessibility	1,21
25	Data paper	40	25	Data Management Plan	1,051
26	Workflow	40	26	Data governance	1,051
27	Data governance	36	27	Data lifecycle	0,848
28	Data lifecycle	36	28	Digital curation	0,713
29	Digital curation	23	29	Scholarly communication	0,322
30	Scientific data management	14	30	Scientific data management	0,195

치한 단어는 'Data repository', 'Data curation', 'Data sharing', 'Metadata'로 나타났으며 이는 전체 기간에 걸쳐 데이터 공유, 재활용과 관련한 연구가 활발히 이루어졌음을 의미한다. 특히 'Data sharing'의 경우 연결 중심성과 매개 중심성 모두에서 높은 값을 가진 중요 키워드로 나타났다. 최근 들어서는 'Open science' 키워드가 연결 중심성과 매개 중심성에서 가장 높게 측정되었으며, 'FAIR data principles'도 주요 키워드로 함께 주목받고 있다. 'Metadata' 키워드는 연결 중심성과 비교하여 매개 중심성에서 높게 측정되어 주요 연구 주제로 언급되기보다 타 연구 주제와의 매개 역할로서 주목받고 있음이 드러났다. 반면, 연결 중심성에 비해 매개 중심성에서 낮게 측정된 키워드는 'Data reuse'로 확인되었다.

5. 결론

본 연구의 목적은 학술논문을 분석하여 연구데이터 관련 연구 현황을 파악하는 것이다. 연구 목적을 달성하기 위해 2010년부터 2024년까지 발표된 연구데이터 관련 학술논문을 Scopus에서 수집하였다. 18개의 연구데이터 관련 키워드를 이용한 검색으로 14,921편의 문헌을 수집할 수 있었으며, 이를 대상으로 키워드 네트워크 빈도 분석 및 중심성 분석 방법을 활용하여 연구데이터 관련 연구의 연구 동향을 살펴보았다. 본 연구에서 수행한 연구데이터 관련 연구 동향 분석 결과를 요약하자면 다음과 같다.

첫째, 1기(2010-2014년)에서 가장 큰 빈도를 보인 키워드는 'Data repository', 'Data curation',

'Data reuse', 'Data sharing', 'Scientific data'였다. 이 중 'Scientific data'를 제외한 4개의 키워드는 연결 중심성 분석에서도 상위 순위를 차지하였고, 나머지 1개의 키워드는 'Metadata'로 확인되었다. 빈도 분석 상위 5개의 키워드 중 'Data reuse' 키워드만이 매개 중심성 상위 5위에 들지 못하였으며 'Digital curation' 키워드가 그 자리를 대체하였다. 이를 통해 1기(2010-2014년)는 'Data repository', 'Data curation', 'Data sharing' 키워드를 중심으로 데이터 공유 관련 연구가 주로 수행되었음을 알 수 있다.

둘째, 2기(2015-2019년)에는 기존의 상위 5개 빈도 키워드였던 'Data repository', 'Data curation', 'Data sharing' 외에도 'Open science', 'Open data' 키워드가 새롭게 등장하였다. 연결 중심성 분석에서는 'Data curation'이 제외되고 'Open access' 키워드가 추가되었다. 연결 중심성에서 가장 높게 측정된 키워드는 'Open science', 매개 중심성에서 가장 높은 값을 가진 키워드는 'Metadata'로 파악되었다. 2016년에는 연구데이터의 저장 및 공유에 활용할 수 있는 프레임워크인 FAIR 데이터 원칙이 등장하였는데, FAIR 원칙은 DOI와 같은 식별자, 데이터와 메타데이터 표준, 기계가 읽을 수 있고 실행 가능한 정보의 중요성을 강조하고 있다. 2기(2015-2019년) 시기에 'Metadata' 키워드가 높은 매개 중심성 값을 보인 것은 이와 유의미한 관계가 있을 것으로 추정할 수 있다. 한편, 1기(2010-2019년)에서 빈도, 중심성 분석의 상위 그룹에 존재했던 'Scientific data' 키워드는 2기(2015-2019년)에서 순위가 상당 수준 하락했다. 2기(2015-2019년)는 1기(2010-2014년)와 비교하여 오픈 사이언스, 오픈 데이터와 관

런 있는 키워드가 주로 상승세를 보였다. 이혜경, 이용구(2023)의 연구에 따르면 Scopus DB에서 확인할 수 있는 오픈 데이터 관련 학술논문은 2010-2014 기간에 870편, 2015-2019 기간에 2,759편이었으며 특히 2019년은 총 645편의 논문이 출판되어 해당 연구의 수집 데이터 기간(1999-2023) 중 오픈 데이터가 가장 많이 연구된 해로 나타났다.

셋째, 가장 최근 기간인 3기(2020-2024)는 'Open science', 'FAIR data principles', 'Data literacy', 'Data curation', 'Open data' 키워드의 빈도가 가장 큰 것으로 나타났다. 이를 통해 'Data curation' 키워드는 전체 기간(2010-2024년) 동안 꾸준히 큰 관심을 받은 키워드임이 확인되었다. 연결 중심성 분석에서의 5개 상위 그룹 키워드는 'Open science', 'FAIR data principles', 'Open data', 'Open access', 'Data sharing'이었으며, 'Data repository' 키워드의 순위가 하락하고 'FAIR data principles' 키워드의 순위가 크게 상승한 것을 알 수 있다. 매개 중심성 분석에서는 'Open science', 'Metadata', 'Data sharing', 'FAIR data principles', 'Open data' 키워드가 상위 5위를 이루었다. 'Metadata' 키워드는 연결 중심성보다 매개 중심성에서 더 높게 측정되어 연구데이터 관련 연구에서 중개 역할로서 주로 활용되고 있는 것으로 드러났다.

이상의 분석 결과에서 'Data sharing' 키워드가 전체 기간(2010-2024년)의 연결 중심성 및 매개 중심성 분석 상위 5위 그룹에서 공통으로 등장하였음을 확인할 수 있었다. 즉, 15년간 '데이터 공유'가 연구데이터 관련 연구의 주요 주제이자 매개체로서 안정적으로 활용되고 있는 것으로 나타났다.

'Metadata'는 1기(2010-2014년)에서 비하여 2기(2015-2019년)에서 매개 중심성 순위가 급증하여 매개 중심성 1위를 차지하였는데, 3기(2020-2024년)에서도 매개 중심성 상위 2위로 확인되면서 2015년부터 약 10년간 연구데이터 관련 키워드 간 연결에 중요한 역할을 하는 것으로 드러났다. 높은 매개 중심성과는 달리 빈도 분석에서는 순위권에 들지 못하였는데, 이는 메타데이터가 연구데이터 연구에서 중요한 연결 키워드임에도 불구하고 메타데이터 그 자체에 대한 연구는 활발하게 진행되고 있지 않다는 것을 의미한다. 따라서 연구데이터 연구와 관련하여 메타데이터를 주제로 한 연구가 진행될 필요가 있다고 제안하는 바이다.

반대로 'Data literacy' 키워드는 최근 10년간 높은 빈도 수를 보이지만 연결 중심성과 매개 중심성에서 순위권에 들지 못하는 것으로 드러났다. 이는 데이터 리터러시에 대한 높은 관심도와는 달리 아직 연구데이터 분야에서 그 영향력이 낮은 것으로 해석할 수 있다. Web of Science DB의 학술논문을 통해 데이터 리터러시 연구의 주경로와 지적구조를 분석한 이재윤(2023)의 연구가 이를 뒷받침할 수 있다. 도서관의 연구데이터 지원 서비스와 관련한 데이터 리터러시 연구의 성장지수는 증가추세이지만 이와 관련한 키워드 중 가장 유망한 키워드는 'Digital Competence'로, 'Data sharing', 'Research Data', 'Research Data Management' 키워드의 출현 빈도는 낮게 나타났다.

한편, 'Open science'와 'Open data', 'FAIR data principles' 키워드는 3기(2020-2024)에서 빈도, 연결 중심성, 매개 중심성 모두에서 높은 순위를 보여 연구데이터 관련 최근 연구에서

주요 주제로 부상하였음을 알 수 있다.

본 연구는 연구데이터 관련 연구 동향을 파악함에 있어서 제한된 기간 내의 특정한 학술 데이터베이스만을 대상으로 하였다. 이러한 특징으로 인해 본 연구의 분석 결과를 해당 분야의 전반적 동향으로서 일반화하기에는 다소 어려움이 존재한다는 점에서 한계점을 갖는다. 또한, 수집한 문헌의 실내용을 상세하게 살펴볼 수 없었던 상태로 키워드 정제 및 동향 분석을 진행하였기 때문에 일련의 과정에서 연구자의 주관성이 개입될 여지가 있어 논의를 진행하는데 제한점이 존재할 수 있다. 따라서 후속 연구에서는 다양한 데이터베이스를 이용한 문헌 수

집으로 신뢰성 있는 분석 표본을 확보하고, 객관성을 보장하는 분석 방법을 통해 동향 분석을 진행하는 것을 제안한다.

위와 같은 제한점에도 불구하고, 본 연구는 다음과 같은 의의를 갖는다. 첫째, 2010년부터 2024년까지 15년간의 연구데이터 관련 분야의 흐름 및 경향을 파악하였다. 둘째, 연구데이터 관련 분야의 동향 비교를 통해 주요점을 발견하고 추후 연구과제를 제안하였다. 본 연구가 향후 연구데이터 관리 분야의 연구 방향성을 제시하는 기초자료로서 활용될 수 있기를 기대한다.

참 고 문 헌

- 구본진, 장덕현 (2023). 문헌정보학분야 해외 연구 동향 및 유망 주제 분석 연구. 한국문헌정보학회지, 57(3), 71-96. <https://dx.doi.org/10.4275/KSLIS.2023.57.3.071>
- 김병선, 정민우, 전상은, 신동빈 (2015). 키워드 네트워크 분석을 이용한 지리공간정보의 글로벌 연구 동향 분석. Spatial Information Research, 23(1), 69-77. <https://doi.org/10.12672/ksis.2015.23.1.069>
- 김소연, 이은주 (2022). 언어 네트워크 분석을 활용한 기록정보서비스 연구동향 분석. 한국기록관리학회지, 22(4), 87-107. <https://dx.doi.org/10.14404/JKSARM.2022.22.4.087>
- 김용학, 김영진 (2011). 사회 연결망 분석(제3판). 서울: 박영사.
- 김우주 (2015). 네트워크 중심성이론: Gephi&R을 사용한 네트워크 빅데이터 분석. 서울: 카오스북.
- 김지현 (2013). 국외 정부연구비지원기관의 연구데이터 관리정책 분석: 미국, 영국, 캐나다, 호주를 중심으로. 한국문헌정보학회지, 47(3), 251-274. <https://dx.doi.org/10.4275/KSLIS.2013.47.3.251>
- 신은자 (2015). 디지털 시대 오픈 데이터 정책의 현황과 과제. 정보관리학회지, 32(3), 49-68. <https://dx.doi.org/10.3743/KOSIM.2015.32.3.049>
- 윤종민, 김규빈 (2013). 과학데이터에 관한 입법례와 관리정책 그리고 대응방안: 호주, 미국, 중국을

- 중심으로. 기술혁신학회지, 16(1), 63-100.
- 이수상 (2014). 언어 네트워크 분석 방법을 활용한 학술논문의 내용분석. 정보관리학회지, 31(4), 49-68.
<https://dx.doi.org/10.3743/KOSIM.2014.31.4.049>
- 이재윤 (2023). 데이터 리터러시 연구 분야의 주경로와 지적구조 분석. 정보관리학회지, 40(4), 403-428.
<http://dx.doi.org/10.3743/KOSIM.2023.40.4.403>
- 이진우, 박동명 (2023). 언어네트워크 분석을 통한 박물관 마케팅의 국내외 연구동향 분석. 박물관학
보, 46, 57-79. <https://dx.doi.org/10.22884/joksms..46.202312.003>
- 이혜경, 이용구 (2023). 동시출현단어 분석을 이용한 오픈 데이터 분야의 지적 구조 분석. 정보관리학
회지, 40(4), 429-450. <http://doi.org/10.3743/KOSIM.2023.40.4.429>
- 정선경 (2022). 텍스트 마이닝을 활용한 데이터 거버넌스 연구 동향 분석: 2009년~2021년 국내
학술지 논문을 중심으로. 디지털융복합연구, 20(4), 133-145.
<https://dx.doi.org/10.14400/JDC.2022.20.4.133>
- 한상우 (2023). 키워드 네트워크 분석을 이용한 연구데이터 관련 국내 연구 동향 분석. 한국도서관·
정보학회지, 54(4), 393-414. <https://dx.doi.org/10.16981/kliss.54.4.202312.393>
- Chunlai Yan, Hongxia Li, Ruihui Pu, Jirawan Deeprasert, & Nuttapong Jotikasthira (2022).
Knowledge mapping of research data in China: a bibliometric study using visual analysis.
Library Hi Tech, 42(1), 331-349. <https://doi.org/10.1108/LHT-11-2020-0285>
- Freeman, L. C. (1978). Centrality in social networks conceptual classification. Social Networks,
1(3), 215-239. [https://doi.org/10.1016/0378-8733\(78\)90021-7](https://doi.org/10.1016/0378-8733(78)90021-7)
- Kwanya, T. (2021). Publishing trends on research data management in Sub-Saharan Africa:
A bibliometrics analysis. IASSIST Quarterly, 45(3-4). <https://doi.org/10.29173/iq996>
- Popping, R. (2000). Computer-assisted text analysis. SAGE Publications, Ltd.
<https://doi.org/10.4135/9781849208741>
- Sheriff, N. & Sevukan, R. (2022). Global research trends in research data management (rdm):
a scientometric view. International Journal of Information Science and Management
(IJISM), 20(4), 117-135. <https://dorl.net/dor/20.1001.1.20088302.2022.20.4.8.7>
- Sheriff, N. & Sevukan, R. (2023). Discovering research data management trends from job
advertisements using a text-mining approach. Journal of Information Science, 0(0).
<https://doi.org/10.1177/01655515231193845>
- Zhang, L. & Eichmann-Kalwara, N. (2019). Mapping the scholarly literature found in scopus
on "research data management": a bibliometric and data visualization approach. Journal
of Librarianship and Scholarly Communication, 7(1).
<https://doi.org/10.7710/2162-3309.2266>

• 국문 참고자료의 영어 표기

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

- Han, Sang Woo (2023). An analysis of domestic research trend on research data using keyword network analysis. *Korean Library And Information Science Society*, 54(4), 393-414. <https://dx.doi.org/10.16981/kliss.54.4.202312.393>
- Jeong, Sun-Kyeong (2022). The study on data governance research trends based on text Mining: based on the publication of Korean academic journals from 2009 to 2021. *Journal of Digital Convergence*, 20(4), 133-145. <https://dx.doi.org/10.14400/JDC.2022.20.4.133>
- Kim, Byeong Sun, Jeong, Min-Woo, Jeon, Sang Eun, & Shin, Dong Bin (2015). Global research trends on geospatial information by keyword network analysis. *Spatial Information Research*, 23(1), 69-77. <https://doi.org/10.12672/ksis.2015.23.1.069>
- Kim, Ji Hyun (2013). An analysis of data management policies of governmental funding agencies in the U.S., the U.K., Canada and Australia. *Journal of the Korean Society for Library and Information Science*, 47(3), 251-274. <https://dx.doi.org/10.4275/KSLIS.2013.47.3.251>
- Kim, So-Yeon & Lee, Eun Ju (2022). Research trends of archival information services using language network analysis. *Journal of Korean Society of Archives and Records Management*, 22(4), 87-107. <https://dx.doi.org/10.14404/JKSARM.2022.22.4.087>
- Kim, Woo Ju (2015). *Network Centrality Theory*. Seoul: Chaosbook.
- Kim, Yong Hak & Kim, Yeong Jin (2011). *Social Network Analysis*. Seoul: Pakyoungsa.
- Koo, Bon Jin & Chang, Durk Hyun (2023). Research on overseas trends and emerging topics in field of library and information science. *Journal of the Korean Society for Library and Information Science*, 57(3), 71-96. <https://dx.doi.org/10.4275/KSLIS.2023.57.3.071>
- Lee, Hyeekyung & Lee, Yong-Gu (2023). Intellectual structure analysis on the field of open data using co-word analysis. *Journal of the Korean Society for Information Management*, 40(4), 429-450. <http://doi.org/10.3743/KOSIM.2023.40.4.429>
- Lee, Jae Yun (2023). Analyzing the main paths and intellectual structure of the data Literacy Research Domain. *Journal of the Korean Society for Information Management*, 40(4), 403-428. <http://dx.doi.org/10.3743/KOSIM.2023.40.4.403>
- Lee, Jin Woo & Park, Dongmyung (2023). Semantic network analysis of domestic and international research trends related to museum marketing. *Journal of Museum Studies*, 46, 57-79. <https://dx.doi.org/10.22884/joksms..46.202312.003>
- Lee, Soosang (2014). A content analysis of journal articles using the language network analysis

methods. *Journal of the Korean Society for Information Management*, 31(4), 49-68.
<https://dx.doi.org/10.3743/KOSIM.2014.31.4.049>

Shin, Eun-Ja (2015). An analysis on trends and tasks of open data policy in the digital era. *Journal of the Korean Society for Information Management*, 32(3), 49-68.
<https://dx.doi.org/10.3743/KOSIM.2015.32.3.049>

Yoon, Chong-min & Kim, Kyubin (2013). Legislation cases, management policies and countermeasures on scientific data: focusing Australia, the United States and China. *Journal of Korea Technology Innovation Society*, 16(1), 63-100.