

사회과학분야 학술 연구자의 연구데이터 재이용 영향요인 연구*

A Study on Factors Affecting the Reuse of Research Data by Academic Researchers in the Social Sciences

박지원 (Ji Won Bak)**

장우권 (Woo Kwon Chang)***

초 록

이 연구는 연구데이터 재이용에 대한 연구자 및 재이용 데이터를 조사를 통해 연구데이터 재이용 영향에 대한 분석 및 활성화 방안을 제시하는데 있다. 이를 위해 연구데이터를 재이용하여 뉴 연구결과물을 산출한 경험이 있는 국내 사회과학분야 학술 연구자를 대상으로 설문배포와 회수를 토대로 178부를 분석하였다. 그 결과, 1) 대부분의 연구자들은 데이터 리포지터리, 데이터관리시스템, 연구데이터 DB와 같은 시스템을 통해 재이용 데이터를 획득하고, 실험 및 관찰을 통해 생산된 분석데이터를 주로 재이용하는 것으로 나타났다. 또한 성공적으로 연구데이터를 재이용한 연구자임에도 불구하고 연구데이터 공유에 대한 인식이 낮을뿐더러 다양한 문제에 직면하여 공유를 하지 않았다. 2) 문헌 검토와 요인 분석을 통해 도출한 10가지 요인(학문적 유용성, 연구의 효율성, 연구자의 우려사항, 데이터 취약성, 직접적 노력, 간접적 노력, 재이용 적합성, 데이터 완전성, 데이터 유용성, 사회적 조건)의 신뢰성과 타당성을 검증하였다. 3) 상관 분석결과 연구 효율성, 사회적 조건은 연구데이터 재이용 의도와 양적 상관관계가, 연구자의 우려사항, 데이터 취약성, 직접적 노력은 연구데이터 재이용 의도와 음적 상관관계가 나타났다. 회귀 분석 결과 이 요인들은 모두 연구데이터 재이용 의도에 유의미한 영향을 미치는 것으로 나타났으며, 연구의 효율성, 사회적 조건, 직접적 노력, 연구자의 우려사항, 데이터 취약성 순서로 연구데이터 재이용 의도에 영향을 미치는 것으로 나타났다. 이를 토대로 연구데이터 재이용 활성화 방안을 제시하였다.

ABSTRACT

This study is to present an analysis and activation plan for the effect of reuse of research data through investigation of researchers and reuse data on reuse of research data. To this end, 178 copies were analyzed based on the distribution and collection of surveys targeting academic researchers in the field of social science in Korea who have experience in calculating new research results by reusing research data. As a result, 1) Most researchers acquire reuse data through systems such as data repositories, data management systems, and research data DBs, and mainly reuse analysis data produced through experiments and observations. In addition, despite being a researcher who successfully reused research data, the awareness of research data sharing was low and did not share it in the face of various problems. 2) The reliability and validity of 10 factors derived through literature review and factor analysis (academic usefulness, research efficiency, researcher concerns, data vulnerability, direct effort, indirect effort, suitability for reuse, data completeness, data usefulness, and social conditions) were verified. 3) As a result of correlation analysis, research efficiency, social conditions showed a quantitative correlation with research data reuse intention, researcher concerns, data vulnerability, and direct effort showed a negative correlation with research data reuse intention. As a result of regression analysis, all of these factors had a significant effect on the intention to reuse research data, and in the order of research efficiency, social conditions, direct efforts, researchers' concerns, and data vulnerability. Based on this, a plan to revitalize the reuse of research data was proposed.

키워드: 사회과학 학술연구자, 연구데이터, 재이용데이터, 재이용 의도, 재이용 영향요인

social science academic researcher, research data, reuse data, reuse intent, reuse influencing factors

* 이 연구는 석사학위논문을 수정·보완하였음.

** 전남대학교 대학원 기록관리학과 석사수료(7won12@naver.com) (제1저자)

*** 전남대학교 문헌정보학과 교수(wk1961@jnu.ac.kr) (교신저자)

■ 논문접수일자: 2021년 11월 22일 ■ 최초심사일자: 2021년 12월 3일 ■ 게재확정일자: 2021년 12월 15일
■ 정보관리학회지, 38(4), 199-230, 2021. <http://dx.doi.org/10.3743/KOSIM.2021.38.4.199>

© Copyright © 2021 Korean Society for Information Management

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

지능정보사회 환경의 핵심은 '데이터'이다. 제6차 국가정보화 기본계획(2018~2022) (과학기술정보통신부 지능정보사회추진단, 2018.12)에 서는 데이터가 지능정보사회를 견인하는 핵심 요인임을 제시하여 '데이터 경제 활성화'를 과제 삼아 데이터를 적극 개방하여 활용률을 높여야 한다고 하였다.

이러한 변화는 학술 분야 및 연구 활동에서도 마찬가지이다. 지식의 발전은 앞선 연구와 데이터를 기반으로 이루어진다. 정보 환경과 학문이 발전함에 따라 연구 과정 중 수 많은 데이터가 생산되고 있으며, 이때 생산되는 연구 데이터가 바로 지식을 이루는 기본 단위이다(Mooney & Newton, 2012). 또한 2015년 OECD의 "Making Open Science a Reality" 보고서(OECD, 2015, 55)에서는 모든 이용자에게 과학적 연구 결과물에 대한 접근이 가능하도록 하는 오픈 사이언스(Open Science)에 대한 논의를 구체화하여, 그 중 "모든 사람이 자유롭게 사용, 재사용 및 재배포 할 수 있는 데이터"인 오픈 데이터(Open Data)가 핵심이라고 하였다. 이는 데이터를 공유함으로써 연구 결과에 대한 검증뿐만 아니라 다른 데이터와 함께 관련 가설 혹은 새로운 가설을 탐구할 수 있기 때문이다(김지현, 2012, 433-455; Piwowar & Todd, 2013, 1-25).

이와 같이 연구의 최종 결과물뿐만 아니라 연구 과정에서 생산되는 연구 데이터도 연구의 성과물로써 학술 커뮤니티 내 중요한 기반으로 자리를 잡아가고 있다. 따라서 연구 데이터는 연구의 효율성 및 연속성을 위해 적절히 공유 및 관리되어야 한다. 우리나라에서는 데이터를

체계적으로 관리할 수 있고, 공유 및 재이용이 활성화 될 수 있는 환경을 마련하기 위해 노력하고 있다.

이에 '공공데이터의 제공 및 이용 활성화에 관한 법률'(시행 2020.12.10. 법률제17344호), '공공데이터포털'(www.data.go.kr/)도 구축하여 운영하고 있다. 또한 연구데이터 체계적인 관리의 중요성에 따라 2019년 9월 1일부터 국가 연구개발 사업 과제 참여시 데이터관리계획(Data Management Plan, DMP, 미국, 유럽 등에서 활성화되어 있는 연구데이터 관리제도로 연구데이터의 수집, 저장, 활용, 공동 활용 등에 대한 계획을 담는 1-2페이지의 짧은 계획서) 작성이 의무화되었다. 이처럼 연구 데이터의 공유와 개방을 확산시키기 위해 법적, 제도적, 기술적 환경을 마련하기 위해 노력하고 있다. 뿐만 아니라 많은 대학 및 연구기관 내에서는 데이터 리포지터리 구축 및 관련 서비스를 지원하기 위해 노력하고 있으며, 학술적인 측면에서는 연구데이터 활용 및 관리 실태 및 인식, 공유 활성화 방안 등과 관련된 연구들이 다수 수행되었다.

그러나 연구데이터에 대한 접근 경로 확대 및 인프라의 마련만으로는 데이터 재이용을 촉진하는 데 한계가 있다. 또한 연구데이터의 재이용을 위해서 공유가 선행되어야 하는 것이 맞지만, 연구데이터가 공유되었다고 재이용이 활성화 된다고 볼 수는 없으며, 재이용에 대한 인식과 경험은 학문적, 사회적, 인프라, 요구사항, 인식 등의 여러 요인에 의해 모두 다르게 나타난다(Faniel & Jacobsen, 2010, 356; Faniel & Zimmerman, 2011, 60). 따라서 연구데이터 재이용 활성화를 위해서는 재이용 연구자와

재이용 관행, 그리고 데이터에 대한 심도 있는 이해가 필요하다.

이 연구는 사회과학분야 학술연구자들을 대상으로 연구데이터 재이용 관행을 조사하였다. 여기에서 어떤 요인들이 연구데이터 재이용에 영향을 미치는지를 분석하고, 그 결과를 토대로 연구데이터 재이용과 연결하여 연구의 기초를 마련, 연구데이터 재이용을 위한 활성화 방안을 제안하고자 한다.

이를 위해 다음 연구문제를 설정하였다.

[RQ. 1] 연구데이터 재이용에 영향을 미치는 요인은 무엇인가?

[RQ. 2] 각 요인들은 연구데이터 재이용에 어떠한 영향을 미치는가?

[RQ. 3] 각 요인들은 연구데이터 재이용에 어느 정도 영향을 미치는가?

이 연구에서의 연구대상은 2018.1.1. ~ 2021.7.26. 사이에 연구데이터를 재이용하여 연구 결과물을 산출한 경험이 있는 국내 사회과학분야 학술 연구자와 재이용하여 산출한 연구 결과물은 학위 논문과 학술지에 게재된 학술 논문으로 한정하였다. 따라서 연구데이터 재이용 경험이 있는 사회과학분야 학술 연구자들을 확인하기 위해 국내 대표적인 사회과학분야 데이터 아카이브인 한국사회과학자료원(KOSSDA)의 ‘자료이용문헌DB’를 이용하였다. 최근까지 생산된 자료이용문헌 2,066건 중 공동 연구자를 포함하여 총 3,432명의 연구자를 파악하였다. 여기에서 연구대상 범위에서 벗어나는 자료이용문헌(연구보고서, 통계자료 등)과 국외연구자를 제외하여 총 1,402명의 연구자를 파악하였다.

설문은 서베이몽키(surveymonkey) 이메일

을 사용하여 실시하였으며, 이메일 수신을 거부한 연구자를 제외하여 최종적으로 1,342명의 연구자를 연구대상으로 확정하였다. 설문조사(2021년 9월 6일~9월 12일)가 진행되었으며, 회수된 응답 220부 중 178부(결측치 제외)를 분석대상으로 하였다.

2. 이론적 배경

2.1 연구데이터

정보기술이 발달하고 지능정보사회로 도래하며 데이터 집약형 연구 패러다임이 확산되고, 데이터 중심의 협업 연구가 활발하게 진행되고 있다(최희윤, 서태설, 2020, 17). 또한 연구 결과의 검증, 새로운 연구의 토대 마련 등을 위해 데이터를 보존하고 재이용하는 것이 중요해지고 있다. 이에 따라 연구 결과물인 학술지 논문뿐만 아니라 연구과정 중 생산된 데이터 또한 개방 및 공유하여 누구나 자유롭게 이용할 수 있도록 하는 것이 중요해졌다. 이렇게 데이터 중심의 연구 패러다임의 변화와 함께 오픈 액세스, 오픈 데이터 운동이 활발해지며 등장한 개념이 바로 오픈 사이언스(Open Science)이다.

OECD의 “Making Open Science a Reality” 보고서에서는 오픈 사이언스에 관한 논의의 장을 확장하며, 이를 오픈 액세스, 오픈 데이터, 오픈 리포지토리 등으로 구체화하였다(OECD, 2015, 18-65). 이 중 핵심이 되고 있는 오픈 데이터는 “모든 사람이 자유롭게 사용, 재사용 및 재배포할 수 있는 데이터”로, 김주섭, 김선태, 최상기(2019, 318)는 이를 학술 연구 영역에서의 연구

데이터라고 하였다. 이는 많은 시간과 비용을 투자한 중요한 정보자원으로 연구의 연속성과 효율성 등을 위해 적절히 관리되어야 함을 강조하였다. OECD(2007, 13-14)에서는 연구데이터를 수치, 텍스트, 이미지, 음성데이터 등과 같이 사실에 기반한 디지털 형식의 데이터로, 주로 과학분야 연구를 위한 기반으로 활용되며 연구결과를 검증하는 역할을 한다고 하였다. 연구데이터는 과학분야 연구의 기반이 되는 사실에 기반한 데이터로 모든 사람에게 공개하여 자유로운 재이용이 가능한 데이터이다.

김선태 외(2010, 84)는 과학분야에서는 연구데이터 보다 과학데이터라는 용어가 더 세부적인 설명이 가능하다고 언급하였다. 최현옥, 이해영(2009, 121-123)은 데이터가 아닌 기록학적 관점에서 연구데이터를 바라보며 공과대학 연구실에서 생산되는 과학분야 연구로 범위를 한정하였다. 반면 신영란, 정연경(2012, 97)은 인문사회분야의 관점에서 연구데이터를 정의하였다.

이렇게 특정 학문분야를 대상으로 연구데이터를 정의한 연구가 있는 반면, 모든 학문분야를 아우르는 관점에서 연구데이터를 정의한 연구와 지침도 있다. 서울대학교 연구윤리지침(2010, 제3장)에서는 연구데이터를 “연구실 또는 실험실에서 연구 수행 결과나 설문조사 등에 의한 통계적 처리 결과로 생성, 관찰된 1차 자료(원자료)와, 1차 자료를 분석, 처리한 2차 자료를 통칭”하는 것, 김은정, 남태우(2012, 29)는 연구데이터를 연구자들이 연구 또는 연구 수행과정에서 생산하는 연구의 부산물이자 결과물이라고 이야기하며, “연구수행과정 중에 산출되는 데이터로서 원자료와 분석된 2차 자료를 포함한

관찰, 조사, 실험, 경험에 기반하는 사실”로 정의하고 있다.

연구데이터는 연구데이터, 연구기록물, 과학데이터 등으로 혼용되어 사용하고 있으나 연구데이터라는 용어가 보편적이며, 서울대학교 연구윤리지침(2010), 김은정, 남태우(2012)의 정의가 특정 학문분야가 아닌 전체 학문분야에서 생산되는 모든 연구데이터를 포괄하고 있음을 알 수 있다.

이 연구에서는 연구데이터를 ‘연구의 계획부터 종료까지 전 과정에서 생산되는 1차 혹은 이를 분석·처리한 2차 자료로, 누구에게나 공개할 수 있으며, 후속 연구 진행 혹은 연구결과의 검증과 재현 등 재이용 및 보존가치가 지속적인 데이터’로 정의한다.

연구데이터는 생산된 학문분야 범위, 연구주제, 방법, 단계, 데이터의 형식, 연구 단계 등에 따라 그 유형이 매우 다양함을 알 수 있다(<표 1> 참조).

따라서 연구데이터는 학문분야의 범위 및 데이터 형식과 관계없이 연구자의 연구 및 분석 방법과 같은 연구 행위에 따라 구분되며, 그에 따라 실험데이터, 관찰데이터, 분석데이터로 구분하였다. 실험데이터는 실험 방법을 통해 생산되는 데이터로 주로 과학분야에서 많이 생산되며, 모델 및 시뮬레이션 데이터, 참조데이터 등이 있다. 관찰데이터는 특정 사물이나 현상을 관측함으로써 생산되는 데이터로 설문조사 자료, 센서스, 표본조사 자료 등이 있다. 분석데이터는 관찰과 실험을 통해 생산된 데이터를 분석 및 적절하게 처리한 2차적으로 생산된 데이터이다(통계자료, 관찰·면접·구술 채록, 오디오 테이프, 비디오 테이프 등).

〈표 1〉 선행연구 및 지침을 통해 살펴 본 연구데이터의 유형

출처	유형
OECD(2007)	사실에 기반한 디지털 형식의 데이터: 수치, 텍스트, 이미지, 음성파일 등
서울대(2010)	1차 자료 및 2차 자료: 연구 수행 결과, 설문조사, 설문조사 통계 처리 자료 등
김선태 외 (2010)	- 관측, 조사, 실험, 분석 등의 과정을 통해 생산된 아날로그 및 디지털 형식의 데이터: 이미지, 오디오, 동 영상 등 / - 지구관측 및 환경분야: 공간정보, 수치정보, 이미지정보 등 관측데이터 / - 컴퓨터과학 분야: 도표정보, 수치정보 / - 사회과학분야: 설문조사 관련 데이터, 통계데이터
김은정, 남태우 (2012)	- 실험데이터: 모델 및 시뮬레이션 데이터, 참조데이터
박옥남(2018)	- 관찰데이터: 조사 데이터, 채수집이 불가능한 역사 기록
신영란, 정연경 (2012)	- 추출데이터: 설문 분석 데이터, 오디오 테이프, 비디오 테이프, 모형, 테스트 응답
기초학문 자료센터	- 센서스, 표본조사 자료, 통계 자료, 사료, 문화·민속 자료, 정치·경제·사회문제 자료, 문화·예술자료, 어문학 자료
기초학문 자료센터	- 센서스, 표본조사 자료, 경제 및 행정 통계 자료, 정부기록문서, 관찰·면접·구술 채록, 사진 및 동영상, 연구결과물, 연구 회의기록, 발간보고서, 출판물

2.2 연구데이터 재이용

연구데이터 재이용 과정은 데이터 생애주기를 살펴봄으로써 이해할 수 있다. 영국 디지털 큐레이션 센터(Digital Curation Center, DCC) (www.dcc.ac.uk)에서 개발한 데이터 라이프 사이클 모델을 통해 이를 확인하고자 한다. 그 이유는 DCC는 연구데이터의 큐레이션을 위한 목적으로 설립된 기관이며, DCC의 데이터 라이프 사이클 모델 또한 연구데이터에 적합하게 개발된 모델로 이 연구에서 다루고자 하는 연구데이터 재이용을 설명하기에 적합하기 때문이다.

연구데이터는 연구와 학문을 뒷받침하는 증거이자, 이를 통해 새로운 연구를 다시 시작함으로써 계속적으로 지식을 창출하고 학문이 발전할 수 있도록 하는 기반이기 때문에 데이터 생애주기 전반에 걸쳐서 적극적으로 관리해야 한다. 따라서 데이터 생산부터 1차 이용을 통한 연구결과물 산출, 보존, 그 후 재이용까지 연구의 생애주기 전 과정에서 데이터를 보존하고, 새로운 가치를 부가할 수 있도록 적극적으로 관리하는 데이터 큐레이션은 매우 중요하다.

이에 DCC는 〈그림 1〉과 같이 데이터 생애주기에 걸친 디지털 큐레이션 모델을 제안하였다. DCC의 디지털 큐레이션 모델은 크게 3가지 활동(전체 라이프사이클 활동, 순차적인 활동, 비정기적인 활동)과 15개의 하위요인으로 구성되어 있다.



〈그림 1〉 DCC Digital Curation Lifecycle Model(Higgins, 2008, 136)

먼저 전체 라이프사이클 활동은 정보의 기술 및 표현, 보존 계획, 커뮤니티 참여, 보존 및 관리로 이루어지며, 데이터에 메타데이터를 부여

하여 데이터를 적절하게 설명하고, 관련 커뮤니티 참여를 통해 데이터 관리와 관련된 표준이나 기술 등을 개발 및 공유함으로써 데이터의 생애 주기 전반에 걸친 보존 계획을 세우는 것이다.

데이터 재이용이라는 용어는 복잡한 개념이며, 아직 공식적인 정의가 없다(김지현, 2012, 4-5). 연구자마다 다른 기준으로 연구데이터 재이용의 개념을 설명하고 있다. 크게 연구자들의 연구 경험을 바탕으로 살펴본 연구(Faniel, Kriesberg, & Yakel, 2012; Curty, 2015; Kim & Yoon, 2017; Yoon & Kim, 2017)와 데이터 재이용을 하나의 연구 방법(Heaton, 2008; Zimmerman, 2008; Faniel & Jacobsen, 2010; Yoon & Lee, 2019)으로써 정의한 연구가 있다. 연구자들의 경험을 조사한 연구에서는 데이터 재이용을 데이터 공유 사이클 및 연구과정에 따라 설명하였다. 연구데이터 재이용은 이미 생산된 데이터를 연구 주기의 효율성, 연구 잠재력 증가, 학문분야 발전 등의 이유로 다시 새로운 연구 기반으로 사용하게 되는 순환하는 자연스러운 연구과정 중 하나이다. 특히 Faniel & Jacobsen(2010, 367)은 데이터 재이용은 원래 생산 목적을 뛰어넘어 새로운 목적을 달성하고 연구 가치를 창출하는 것이 중요하다고 언급, 그 범위에 대한 구체적이고 명확한 정의가 필요하다고 하였다.

한편 데이터 재이용을 연구 방법으로써 정의한 대부분의 연구는 원래 생산된 목적이 아닌 새로운 연구 문제를 해결하기 위해 새로운 방법으로 이용하는 것으로 정의하였다. Yoon & Lee(2019, 2)는 데이터를 생산하지 않은 사람이 다른 사람이 생산한 데이터를 이용하는 것과 새로운 연구문제를 해결하는 것뿐만 아니라 기존 연구결과의 재현까지도 데이터 재이용 행

위로 보다 광범위하게 정의하였다. 이는 데이터 생산자를 기준으로 데이터 재이용을 정의하였으며, 기존 연구 재현도 데이터 재이용으로 보아 재이용 범위를 확장하였다는 의미가 있다.

이러한 데이터 재이용은 2차 분석과 동일한 의미를 지녔다고 볼 수 있다(Zimmerman, 2008, 30; Yoon & Lee, 2019, 1246). 2차 분석은 이미 생산된 데이터 및 자료를 분석 자료로 활용하는 방법(Heaton, 2008, 506)으로, 주로 양적 자료를 대상으로 한다. 하지만 양적 자료뿐만 아닌 질적 자료의 2차 분석에 대한 논의 또한 지속적으로 이루어지고 있으며, 구체적으로는 질적 자료의 2차 분석 필요성과 방법, 윤리적 문제와 같이 고려해야 할 사항 등에 대한 연구(Heaton, 2008; Bishop, 2015)가 진행되어왔다. 따라서 2차 분석 자료의 범위에는 통계 자료, 설문 자료, 관찰·면접·구술 채록 등 연구데이터 뿐만 아니라 개인기록인 일기, 편지, 사진과 같은 문서나 다양한 연구기록물도 포함된다(Heaton, 2008, 506). 이는 앞서 정의한 연구데이터(2.1 연구데이터 정의)와 그 유형(〈표 1〉참조)에 부합하지 않는 것으로 보아, 연구데이터 재이용은 보다 한정적으로 2차 분석(양적, 질적 분석자료 등)을 설명하는 용어임을 알 수 있다.

이와 같이 연구데이터 재이용을 바라보는 관점에 따라 연구자마다 다르게 정의하고 있으며, 아직 공통된 정의가 없음을 알 수 있다. 그럼에도 공통된 확인된 내용은 원래의 생산 목적과 다른 새로운 연구 목적을 달성하기 위해 사용한다는 것이며, 광범위하게는 이용 목적이 같더라도 다른 사람이 생산한 데이터이거나 기존 연구를 재현하는 경우까지도 연구데이터 재이용 행위로

볼 수 있다. 또한 앞서 살펴본 데이터 생애주기를 통해 알 수 있었듯이, 연구데이터 재이용은 적절한 처리과정을 거치거나 다른 데이터와 결합하여 새로운 연구의 기반으로 이용하는 연구의 본질적인 과정이다. 이를 종합하여 본 연구에서는 연구데이터 재이용을 '새로운 연구 문제를 해결하거나, 기존 연구 결과의 재현을 위해 다른 사람이 생산한 데이터를 이용하는 행위'로 정의한다. 이는 연구의 본질적인 과정 중 하나로 새로운 연구를 위해 적절하게 처리하거나, 새로운 데이터와 결합하는 등의 방법을 통한다.

2.3 연구데이터 재이용에 영향을 미치는 요인

연구자는 선행연구를 종합적으로 살펴보고, 연구자들의 연구데이터 재이용에 영향을 주는 6개의 요인(재이용 유용성, 재이용 위험성, 재이용하기 위한 노력, 재이용 가능성, 재이용 용이성, 사회적 조건)과 20개 하위 요인을 다음과 같이 도출하였다.

2.3.1 재이용 유용성

① 지식의 확장: 데이터를 재이용하는 과정을 통해 연구자의 지식의 지평을 넓히고, 학문분야의 발전을 도모하게 되는 것(김지현, 2012; Curty, 2015; Niu, 2009; Piwowar & vison, 2013). ② 효율성: 데이터를 수집·생산하는데 필요한 시간과 노력을 줄이고, 다른 연구자와 동일한 데이터를 수집하게 되는 불필요한 중복작업을 최소화함으로써 연구를 효율적으로 수행할 수 있게 하는 것(김나연, 2019; Curty, 2015; Kiecolt & Nathan, 1985).

2.3.2 재이용 위험성

① 저평가에 대한 두려움: 데이터를 재이용하는 것이 가치가 낮은 것으로 인식되어, 나와 나의 연구가 저평가 될 것이라는 걱정(Curty, 2015; Goodwin, 2012; Kim & Yoon, 2017; Martin, 1995). ② 윤리적 문제 위반: 데이터에 연구대상자의 개인정보가 포함되어 있거나 추가적인 동의 없이 연구가 진행됨으로써 윤리적 문제가 발생하거나 데이터 보호 규범을 위반할 수 있다는 걱정(Curty, 2015; Goodwin, 2012; Kim & Yoon, 2017; Martin, 1995). ③ 재이용 과정에서 실수: 데이터를 잘못 해석하여 오용하게 될 가능성과 관련된 걱정(김나연, 2019; Curty, 2015; Tenopir et al., 2011; Yoon, 2014; Yoon & Kim, 2017). ④ 숨겨진 오류에 대한 취약성: 데이터 자체에 숨겨진 오류를 쉽게 식별하지 못해, 부정확한 데이터를 사용하게 될 수 있다는 걱정(Curty, 2015; Faniel & Zimmerman, 2011; Zimmerman, 2008).

2.3.3 재이용하기 위한 노력

① 새로운 발견: 기존 데이터를 기반으로 데이터를 생산한 기존 연구자, 혹은 다른 재이용자가 아직 다루지 못했거나 간과한 다양한 연구문제를 탐색하려는 노력(Curty, 2015; Zimmerman, 2008). ② 데이터 접근: 나의 연구 목적 및 주제와 관련되며 재이용이 가능한 연구데이터를 찾아내고, 이용 권한을 얻기 위한 노력(Curty, 2015; Tenopir et al., 2011). ③ 불일치 처리: 나의 연구에 부합하게 데이터를 조작, 처리하거나 혹은 데이터를 재이용할 수 있도록 연구 설계를 조정하는 등 재이용하려는 데이터와 내가 필요한 데이터 사이의 불일치를 처리하기 위한 노력

(김지현, 2012; Curty, 2015; Kiecolt & Nathan, 1985). ④ 기존 연구 이해: 데이터 자체와 데이터가 생산된 연구 맥락을 충분히 이해하기 위한 노력(김지현, 2012; Faniel & Jacobsen, 2010; Zimmerman, 2008).

2.3.4 재이용 가능성

① 데이터 설명 자료 제공: 데이터에 대한 맥락 정보를 충분히 파악할 수 있도록 하는 다양한 형태의 자료들(김지현, 2012; Curty, 2015; Faniel & Jacobsen, 2010; Zimmerman, 2003; 2007; 2008). ② 데이터 적합성: 데이터의 주제, 유형, 분석 수준 등이 연구 목적과 연구 질문을 해결하기에 알맞는 상태(Curty, 2015; Niu, 2009). ③ 데이터 신뢰성: 데이터의 생산자, 데이터가 기탁된 기관에 대한 신뢰로 판단되는 데이터에 대한 믿음(김나연, 2019; 김지현, 2012; 전리나, 2017; Curty, 2015; Faniel & Jacobsen, 2010; Faniel, Kriesberg, & Yakel, 2012; Zimmerman, 2008). ④ 데이터 품질: 데이터의 일관성과 완전성의 보장(Curty, 2015; Faniel, Kriesberg, & Yakel, 2012). ⑤ 연구의 엄격성: 데이터가 생산된 기존 연구의 설계와 실행의 엄격함(전리나, 2017; Curty, 2015).

2.3.5 재이용 용이성

① 데이터 리포지터리 이용: 데이터에 쉽게 접근할 수 있도록 수집, 관리하고 있는 시스템, 인프라에 접근 및 이용(김지현, 2012; Curty, 2015; Faniel, Kriesberg, & Yakel, 2012; Faniel & Zimmerman, 2011; Piwowar & Todd, 2013; Yoon & Kim, 2017). ② 데이터 생산자와 소통: 데이터에 대한 추가 정보를 데이터 생산자

가 재이용자에게 제공할 수 있고, 재이용자가 데이터에 대한 궁금점을 생산자에게 문의할 수 있는 것(Curty, 2015; Faniel & Zimmerman, 2011). ③ 재이용 지원 및 서비스: 데이터를 재이용과 관련된 공식적인 교육 및 재이용을 돕는 전담 직원의 서비스와 지원(심윤희, 김지현, 2019; 심원식 외, 2015; Curty, 2015; Faniel, Kriesberg, & Yakel, 2012; Kriesberg et al., 2013).

2.3.6 사회적 조건

① 재이용 수용성: 학문분야에서의 데이터 재이용의 보편화, 일반화 정도에 따른 데이터 재이용 수용 정도(김나연, 2019; Curty, 2015). ② 동료의 권장: 연구를 함께 하는 동료들이 다른 연구자가 생산한 데이터를 자주 재이용하거나, 재이용을 권장, 지원하는 것(Curty, 2015; Kriesberg et al., 2013).

2.4 선행연구

연구데이터 재이용에 영향을 미치는 요인 분석을 위한 토대를 마련하기 위해, 연구데이터 활용과 관련된 연구자들의 인식과 관행 전반을 파악한 연구(김지현, 2012; 2015; 신영란, 정연경, 2012; 심원식, 2015; Faniel & Jacobsen, 2011; Faniel & Zimmerman, 2011; Faniel et al., 2012; Kriesberg et al., 2013; Zimmerman, 2008)와, 두 번째로 연구자의 경험을 바탕으로 연구데이터의 공유 및 재이용에 영향을 미치는 요인을 도출한 연구(김은정, 남태우, 2012; 김문정, 김성희, 2015; 전리나, 2017; 김나연, 2019; Curty, 2015; Faniel, Kriesberg, & Yakel, 2016; Kim & Yoon, 2017)를 국내외로 나누어 살펴보았다.

〈표 2〉 연구데이터 활용과 연구자들 인식

연구자	연구내용
김지현(2012:2015)	연구데이터를 체계적으로 관리하기 위한 노력을 하고 있으나 아직 미흡한 점을 지적하며, 효과적인 연구데이터 관리 지원 서비스를 구축하기 위해서는 연구자들의 연구데이터 활용에 대한 관행과 인식에 대한 이해가 선행되어야 함
신영란, 정연경(2012)	서구에서는 가치 있는 인문사회분야 연구데이터를 아카이빙하고 활용할 수 있도록 하는 데이터 아카이브가 발달해 왔으나, 국내에서는 인식 부족과 기술 및 제도적 문제로 인해 제대로 관리 및 활용되지 못하고 소실되는 경우가 빈번하다는 점을 문제로 제기
심원식, 안혜연, 변제연(2015)	기초학문자료센터를 중심으로 인문학 분야 연구데이터의 체계적인 수집, 관리 및 활용을 위한 과정과 전략을 제시

〈표 3〉 연구데이터 공유와 재이용

연구자	연구내용
김은정, 남태우(2012)	연구데이터 활용을 위해서는 우선 수집이 되어야 함을 주장하며, 연구데이터 공유 활성화를 위해 2010년 SCI에서 확인된 논문 중 국내 저자들을 대상으로 연구자들의 연구데이터 제출 의도에 영향을 미치는 요인을 분석
김문정, 김성희(2015)	과학기술분야 연구자를 대상으로 연구데이터 공유에 영향을 미치는 요인간의 인과관계 분석
김나연(2019)	연구데이터의 활용 활성화가 필요하다는 점은 동의하나, 단순히 데이터 공급의 확대와 접근가능성 확보만으로는 연구데이터 재이용을 촉진하는데 한계가 있음을 지적
전리나(2016)	연구데이터 재이용에 영향을 미치는 리포지터리의 환경 요인에 대해 분석
김문정, 김성희(2015)	구글애널리틱스를 활용하여 의학도서관 웹 사이트 이용행태를 분석
Zimmerman(2008)	연구데이터 재이용 경험이 있는 생태학자 인터뷰를 통해 다른 사람의 연구데이터를 재이용한 경험을 분석
Faniel & Jacobsen(2010)	활발한 연구데이터 활용을 위해 관련 인프라 및 정책 등을 개발하는데 많은 노력을 기울이고 있지만 연구데이터 공유와 재이용이 여전히 활성화되지 않았음을 지적하며, 지진공학 연구자 14명을 대상으로 다른 사람의 연구 데이터 재이용 가능성을 어떻게 평가하는지 조사
Faniel & Zimmerman(2011)	연구데이터를 공유하고 재이용하는 것이 중요해지고, 이에 관한 논의가 확장되고 있으나, 아직 대부분의 연구는 연구데이터를 공유 하려는 동기 요인에 초점이 맞춰져 있음을 지적. 또한 데이터를 널리 공유하는 것이 재이용을 보장할 수 있다고 많은 선행연구에서 암묵적으로 가정하고 있지만, 그것만으로는 연구데이터 재이용을 보장하기 어렵다고 함
Faniel, Kriesberg, & Yakell(2012:2016)	재이용에 대한 지원과 서비스, 데이터의 품질(완전성, 신뢰성), 데이터 재이용 가능성(조작 용이성, 설명 자료의 품질, 접근성)이 데이터 재이용 의도에도 영향을 미칠 것으로 예측
Kriesberg et al(2013)	2011년 Interuniversity Consortium for Political and Social Research(ICPSR)에서 주최하는 하계 프로그램에 참가한 고고학, 양적 연구를 하는 사회학, 동물학 총 3개 학문분야 92명의 인터뷰를 통해 연구 경험이 많지 않은 연구자들의 데이터 재이용이 어떻게 이뤄지는지 살펴보고, 데이터를 재이용 하여 지식을 창출하는 과정을 확인
Curty(2015)	ICPSR과 Havard Database Network의 이용자 중 사회과학분야 연구자 13명을 선정하여 인터뷰를 실시하였으며, 인터뷰 결과와 선행연구 고찰을 통해 총 6가지 영역(인지된 이익, 인지된 어려움, 인지된 노력, 재이용성 판단, 재이용 활성화 요인, 사회적 요인)과 25개의 세부 영향요인이 도출
Kim & Yoon(2017)	연구자들의 데이터 재이용 행동에 영향을 미치는 요인을 확인하고 다양한 분야의 데이터 재이용에서 발생하는 일반화된 행동 양상을 파악

선행연구를 살펴본 결과 국외의 경우 연구 데이터 재이용에 대한 논의가 다양한 관점에서 활발하게 이루어지고 있으나, 국내의 경우 연구데이터 활용과 관련된 전반적인 관행을 확인하고 있을 뿐 연구데이터 재이용을 다각적인 관점에서 분석한 연구는 거의 없는 실정이다. 이에 따라 이 연구에서는 선행연구를 바탕으로 연구데이터 재이용에 영향을 미치는 요인을 도출하고, 연구데이터 재이용 경험이 있는 국내 사회과학분야 학술 연구자를 대상으로 연구데이터 재이용 영향 요인을 분석하고자 한다.

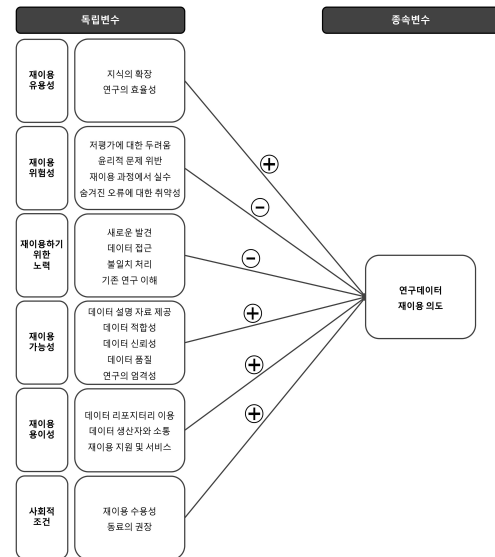
3. 연구설계 및 분석

3.1 연구설계

3.1.1 연구모형

선행연구를 통해 도출한 6가지 요인(재이용 유용성, 재이용 위험성, 재이용하기 위한 노력, 재이용 가능성, 재이용 용이성, 사회적 조건)이 연구자들의 연구데이터 재이용 의도에 어떠한 영향을 미치고 있는지 분석하기 위해 각 요인을 독립변수로, 연구데이터 재이용 의도를 종

속변수로 설정하였다. 독립변수와 종속변수의 관계를 나타낸 연구 모형은 <그림 2>와 같다.



<그림 2> 연구 모형

3.1.2 가설설정

이 연구의 연구 모형을 통해 재이용 유용성, 재이용 위험성, 재이용하기 위한 노력, 재이용 가능성, 재이용 용이성, 사회적 조건이 연구데이터 재이용 의도에 영향을 미친다고 가정하였다. 이를 토대로 다음의 <표 4>와 같은 가설을 설정하였다.

<표 4> 연구 가설 설정

가설 1	연구데이터 재이용으로 인한 이점은 연구자의 연구데이터 재이용 의도에 긍정적 영향을 미칠 것이다.
가설 2	연구데이터 재이용으로 인한 우려사항과 잠재된 위험은 연구자의 연구데이터 재이용 의도에 부정적 영향을 미칠 것이다.
가설 3	연구데이터 재이용을 위한 노력은 연구데이터 재이용 의도에 부정적 영향을 미칠 것이다.
가설 4	연구데이터 재이용 가능성은 연구데이터 재이용 의도에 긍정적 영향을 미칠 것이다.
가설 5	연구데이터 재이용에 용이한 환경적 조건은 연구데이터 재이용 의도에 긍정적 영향을 미칠 것이다.
가설 6	연구자의 사회적 조건은 연구데이터 재이용 의도에 긍정적 영향을 미칠 것이다.

3.1.3 측정항목과 설문 구성

위의 가설을 검증하기 위하여 연구데이터 재이용 의도에 영향을 주는 요인을 설명하는 문항을 작성하여 연구자들의 연구데이터 재이용 의도를 조사하였다. 또한 연구데이터 재이용 경험 및 공유 경험에 대한 질문을 통해 연구자들의 연구데이터 활용 관행을 파악하였다. 설문조사 응답 데이터를 바탕으로 응답평균을 조사하여 연구자들의 연구데이터 재이용과 관련된 인식을 파악하였고, 이후 각 요인들이 연구데이터 재이용 의도에 어떤 영향을 얼마나 미치는지에 관한 영향도를 분석하였다.

설문 문항은 연구데이터 재이용 의도에 영향을 주는 요인에 관한 질문과 연구데이터 재이용 및 공유 경험에 관한 질문, 일반사항 및 인구통계학적 기초 자료와 관련된 질문으로 총 50문항을 구성하였으며, 5점 리커트 척도를 사용하였다. 문항 별 측정 항목은 다음의 <표 5>와 같다.

3.1.4 자료 현황분석

1) 응답자 정보

① 성별 및 연구 경력: 분석한 설문 응답자의 성별과 연구 경력에 관한 정보는 <표 6>과 같다.

<표 5> 설문 측정항목

	구분	측정항목
1	연구데이터 재이용 유용성(A)	연구데이터 재이용 이점과 동기에 대한 인지 정도
2	연구데이터 재이용 위험성(B)	연구데이터 재이용을 함으로써 처할 수 있는 위험사항 및 걱정에 대한 인지 정도
3	연구데이터 재이용하기 위한 노력(C)	연구데이터 재이용을 하기 위한 개인의 노력 필요성에 대한 인지 정도
4	연구데이터 재이용 가능성(D)	연구데이터 재이용을 가능하게 하는 데이터의 특성에 대한 중요성 인지 정도
5	연구데이터 재이용 용이성(E)	연구데이터 재이용을 용이하게 돕는 환경적 조건의 중요성에 대한 인지 정도
6	연구자의 사회적 조건(F)	연구데이터를 재이용함에 있어 연구자의 사회적 환경의 중요성에 대한 인지 정도
7	연구데이터 재이용 의도	연구데이터를 재이용하고자 하는 의지
8	연구데이터 재이용 경험	재이용 연구데이터를 획득한 경로
9	연구데이터 공유 경험	연구데이터의 공유 빈도
10	인구통계학적 질문 및 일반사항	연구자의 연구 경력 / 연구자의 주 전공 및 연구 분야 연구자의 주된 연구 방법 / 연구자가 주로 이용하는 데이터 유형 연구자가 주로 생산하는 데이터 유형 / 성별

<표 6> 응답자의 일반정보

	구분	빈도	퍼센트	누적 퍼센트
성별	여성	93명	52.2%	
	남성	85명	47.8%	
연구 경력	5년 미만	28명	15.7%	15.7%
	5~10년	54명	30.3%	46.1%
	11~15년	44명	24.7%	70.8%
	16~20년	27명	15.2%	86.0%
	20년 이상	25명	14.0%	100.0%
전체		178명	100%	

응답자는 여성 93명(52.2), 남성 85명(47.8)으로, 여성이 조금 더 많으나 큰 차이 없이 성별의 균형이 적절하게 구성되었다. 연구 경력은 석사 과정 시작 시점을 기준으로 정하였다. 그 이유는 본 연구의 연구대상은 '연구데이터 재이용을 통해 학술적 연구 결과물을 산출한 연구자'로, 학술지에 게재된 논문뿐만 아니라 학위논문이라는 결과물을 산출한 연구자까지 포함하기 위해서이다.

② 연구분야 및 방법: 연구데이터 활용에 관한 선행연구를 통해, 학문 분야에 따라 혹은 연구 방법 및 경력에 따라 그 인식이나 관행이 다르게 나타나는 것을 확인할 수 있다. 이 연구에서는 사회과학분야 학술 연구자로 대상의 범위를

한정하였으나, 보다 세부적으로 연구데이터 재이용 관행을 파악하고 결과에 대한 이해를 돕기 위해 응답자의 연구 분야와 주된 연구 방법을 조사하였다(〈표 7〉 참조).

연구 분야는 한국사회과학자료원에서의 학문 분류표를 기본으로 사용하였으며, 한국연구재단의 학술연구분야분류표를 참고하여 일부 세분화하였다. 응답자의 연구 분야는 사회복지학이 25.3%로 가장 많았으며, 행정학 > 사회과학일반 > 심리과학 > 교육학 > 사회학 > 경제학 > 경영학 > 정책학 > 정치외교학 > 국제/지역개발 > 관광학 > 군사학 > 농업경제학 > 무역학 > 신문방송학 > 지역학 순으로 분포되어 있다.

〈표 7〉 연구분야 및 연구 방법

구분	빈도	퍼센트
연구 분야	경영학	7명 3.9%
	경제학	8명 4.5%
	관광학	1명 0.6%
	교육학	17명 9.6%
	국제/지역개발	2명 1.1%
	군사학	1명 0.6%
	농업경제학	1명 0.6%
	무역학	1명 0.6%
	사회학	16명 9.0%
	사회복지학	45명 25.3%
	신문방송학	1명 0.6%
	심리과학	19명 10.7%
	정책학	7명 3.9%
	정치외교학	3명 1.7%
	지역학	1명 0.6%
	행정학	27명 15.2%
	사회과학일반	21명 11.8%
연구 방법	양적 연구	141명 79.2%
	질적 연구	1명 0.6%
	혼합 연구	36명 20.2%
전체	178명	100.0%

연구 방법은 양적 연구가 79.2, 질적 연구가 0.6)으로 양적 연구를 주로 하는 연구자가 절대적으로 많은 것을 알 수 있다. Martin(1995)과 Niu(2009)의 연구에서 같은 사회과학분야에서도 양적 연구자와 질적 연구자의 연구데이터 재이용에 앞서 걱정하고 있는 요인이 다른 것으로 나타났음을 확인할 수 있다. 따라서 본 연구의 재이용 영향요인의 분석 결과를 해석함에 있어 이 점을 유의할 필요가 있다.

③ 재이용 데이터 획득 및 경로: 국내의 경우 연구자들이 연구데이터의 공유 및 재이용에 대한 중요성을 어느 정도 인지하고 있으나 실제로는 체계의 부족, 시스템 및 인프라 등의 환경 조건이 미흡함을 알 수 있다(김지현, 2012, 453; 박옥남, 2018, 104). 따라서 이미 연구데이터를 재이용하여 연구 결과물까지 성공적으로 산출해 낸 연구자들은 재이용 데이터를 어떤 경로를 통해 주로 획득하는지 질문하였다. 재이용 데이터 획득 경로는 <표 8>과 같다.

거의 모든 연구자들은 재이용 데이터를 주로 데이터 리포지터리, 데이터관리시스템, 연구데이터 DB와 같이 기탁되어 공개된 것을 주로 이용하는 것으로 나타났다. 케이스 중 퍼센트(전체 응답자수에 기초하여 계산한 항목별 상대 비율로, 전체 응답자 중 해당 항목을 선택한 비

율)를 통해 확인한 응답 분포는 데이터 리포지터리, 데이터 관리 시스템, 연구데이터DB의 이용이 91.0, 개인적으로 연구자를 통한 연락이 16.9, 출판된 논문에서 데이터를 추출하는 경우가 12.9, 기타가 5.1로 나타났다.

기타 응답으로는 구매를 통해 수집할 수 있는 데이터였으나, 가격이 너무 비싸서 다른 연구자에게 공유를 요청, 데이터를 수집한 연구자가 공동연구를 요청하여 그 데이터를 이용하여 공동연구를 진행함, 소속 연구실에서 수집이 된 데이터를 이용, 공공기관에서 조사하여 공개된 자료 이용, 기존 연구 및 연구보고서의 데이터를 적절하게 처리하여 이용 등의 답변이 확인되었다. 이러한 기타 응답 또한 데이터 리포지터리 등을 통해 공개되었거나, 공동연구자 혹은 연구실의 다른 연구자를 통해 공유 받거나, 출판된 연구 결과물에서 확인 가능한 데이터를 적절하게 처리하여 이용하는 것으로 구분할 수 있다.

④ 이용 생산하는 데이터 유형

연구자들이 주로 이용하거나 생산하는 데이터를 확인해보았다. 데이터 유형은 본 연구에서 살펴본 연구데이터의 유형을 토대로 응답자들이 보다 쉽게 유형을 이해하고 응답할 수 있도록 Curty(2015)의 설문 문항 항목을 참고하

<표 8> 재이용 데이터 획득 경로

구분	빈도	퍼센트	케이스 퍼센트
데이터 리포지터리, 데이터 관리 시스템, 연구 데이터 DB 이용	162명	72.3%	91.0%
개인적으로 연구자에게 연락	30명	13.4%	16.9%
출판된 논문에서 데이터 추출	23명	10.3%	12.95%
기타	9명	4.0%	5.1%
전체	224명	100%	125.8%

여 구성하였다. 복수응답을 허용, 응답 결과는 <표 9>와 같다.

주로 이용하는 데이터는 공식 통계자료가 79.8(복응답 허용, 케이스 퍼센트)로 가장 많았다. 다음으로는 설문 결과 데이터> 설문지> 인터뷰데이터> 실험데이터> 관찰에 의한 데이터> 임상실험 자료> 기타> 오디오 파일> 비디오 파일 순이었다.

공식 통계자료가 굉장히 큰 비율을 차지했으며, 이는 공식 적인 통계자료의 경우 데이터에 대한 메타데이터를 충분히 확인할 수 있으며, 보다 쉽게 접근할 수 있기 때문이라고 예측할 수 있다. 기타 응답으로는 '수집한 데이터를 바탕으로 시뮬레이션으로 재생성한 데이터', '패널자료', '외부 기관에서 제공하는 2차 자료' 등을 확인할 수 있었다.

반면 주로 생산하는 데이터는 설문 결과 데이터가 65.2로 가장 많았다. 그 다음으로 설문지> 공식 통계자료> 인터뷰데이터> 관찰에 의한 데이터> 실험데이터> 임상실험 자료> 기타> 비디오파일> 오디오파일로 나타났다. 이는 주로 이용하는 데이터에서는 공식 통계자료가 가장 많은 부분을 차지하였으나, 주로 생산하는 데이터에서는 세 번째에 위치한다는 점을 제외하고는, 순위와 비율에 큰 차이가 없음을 알 수 있다. 기타 응답으로는 '영상분석 데이터', '설문지 데이터를 재구조화 한 설문지' 등으로 나타났다.

따라서 이러한 데이터 유형을 생산측면에서 실험, 관찰, 추출로 나누어 이 연구에서 정의한 데이터 유형에 맞춰 구분하면 <표 10>과 같다.

<표 9> 주요 이용 및 생산 데이터

유형	주요 이용 데이터			주요 생산 데이터		
	명	%	케이스 퍼센트	명	%	케이스 퍼센트
설문지	105	23.4	59.0	97	26.7	54.5
설문결과 데이터	129	28.8	72.5	116	32.0	65.2
인터뷰 데이터	38	8.5	21.3	50	13.8	28.1
실험 데이터	11	2.5	6.2	13	3.6	7.3
오디오파일	2	0.4	1.1	1	0.3	0.6
비디오파일	1	0.2	0.6	2	0.6	1.1
임상실험 자료	6	1.3	3.4	5	1.4	2.8
관찰에 의한 데이터	10	2.2	5.6	17	4.7	9.6
공식 통계 자료	142	31.7	79.8	58	16.0	32.6
기타	4	0.9	2.2	4	1.1	2.2
전체		100	251.7		100	203.9

<표 10> 데이터의 유형 구분

유형	응답된 항목
실험데이터	실험데이터, 임상 실험 자료, 시뮬레이션 데이터
관찰데이터	설문지, 관찰에 의한 데이터, 패널데이터
추출데이터	설문 결과 데이터, 인터뷰데이터, 통계자료, 2차 자료, 영상분석 데이터, 설문지를 재구조화하여 생산한 설문지자료

기타 응답은 제외하여 실험데이터, 관찰데이터, 추출데이터로 구분하여 전체 퍼센트를 구하면 <표 11>과 같다.

주로 이용하는 데이터는 추출데이터>관찰데이터>실험데이터 순으로 나타났고, 주로 생산하는 데이터 또한 추출데이터>관찰데이터>실험데이터 순으로 나타났다.

주로 이용하는 데이터와 생산하는 데이터 사이에 순위나 비율에는 큰 차이가 없는 것으로 확인되었으며, 특히 주요 생산 데이터 중 추출데이터가 93.3으로 나타나 연구를 진행함에 있어 실험과 관찰을 통해 생산된 데이터를 분석·처리하는 과정을 거의 필수적으로 거치고 있는 것을 알 수 있다.

⑤ 연구데이터 공유 경험

이 연구는 연구자들의 연구데이터 재이용 관행에 대해 조사를 하는 연구이지만 공유와 재이용은 상호 연결된 과정(Curty, 2015; Faniel

& Jacobsen, 2011)이기 때문에 공유경험 결과는 <표 12>와 같다.

공유 경험이 한 번도 없는 연구자>1번 있는 연구자>2~4번인 연구자>5번 이상>10번 미만인 연구자>10번 이상인 연구자 순으로 나타났다. 공유 경험이 있는 연구자 중 그 빈도가 5번 미만인 연구자는 총 57.9로 절반에 해당하는 것을 알 수 있고, 5번 이상인 연구자는 총 14.7로 급격히 적은 분포를 띄고 있음을 알 수 있다. 응답자는 모두 재이용 경험이 최소 한번 이상 있는 사회과학분야 연구자임을 미루어 보았을 때, 재이용 경험은 있으나 공유 경험이 없는 연구자가 27.5나 되는 것을 알 수 있다. 재이용 경험은 있으나 공유는 하지 않은 이유에 대한 결과를 요약·정리하면 <표 13>과 같다.

첫째, 공유에 대한 인식이 부족하고, 필요성을 인지하지 못하거나 공유하더라도 내가 생산한 데이터의 활용성이 떨어질 것이라고 느낀다는 것이다. 특히 기관에서 생산되어 공유되는

<표 11> 유형별 주요 이용 및 생산데이터

유형	주요 이용 데이터		주요 생산 데이터	
	빈도	케이스 퍼센트	빈도	케이스 퍼센트
실험데이터	17	9.6	18	10.1
관찰데이터	115	64.6	114	64.0
추출데이터	167	93.8	166	93.3

<표 12> 공유 경험

구분	빈도	퍼센트	누적 퍼센트
없음	49명	27.5%	27.5%
1번	39명	21.9%	57.9%
2~4번	64명	36.0%	
5번~9번	17명	9.6%	14.7%
10번 이상	9명	5.1%	
전체	178명	100.0%	100.0%

〈표 13〉 연구데이터 공유 경험이 없는 이유

공유하지 않은 이유	응답
공유에 대한 인식 부족	“공유 필요성에 대한 인식이 없었음” / “별 다른 이유가 없음” / “연구데이터 공유에 대해 잘 모름” / “기관에 공개된 데이터 아닌 개인 공유한 데이터를 재이용해본 적 없어 인식 부족”
공유 필요성 및 활용 가치 부족	“공유의 필요성이 없다고 느낌(활용되지 않을 것 같음)” / “연구 목적에 따라 데이터를 처리하는 방법이 다르기 때문” / “연구 주제 상 데이터 공유가 무의미함” / “연구 관련 변수들이 범용적이지 않음” / “공유 요청을 받은 적 없으며, 실제 연구자가 본인 연구문제와 설계에 부합하지 않으면 공유한다고 하더라도 개인 연구데이터를 이용하여 연구를 진행한다는 것이 쉽지 않음”
공유 요청의 부재	“요청한 사람이 없어서” / “이용하고 싶다고 한 연구자가 없었음” / “요청이 없었음”
공유 기회의 부족	“공유할 루트(경로)가 없었다” / “기관/연구소에서 수집한 데이터가 아닌, 개인이 수집한 데이터를 공유할 수 있는 플랫폼 부재 및 신뢰도 부족” / “공유할 기회가 없었으며, 어디서 어떻게 공유하는지 모름”
데이터에 대한 윤리적 문제	“연구 대상자 동의가 필요하기 때문에”
데이터 소유권 문제	“데이터 권한이 생산한 기관에 있기에” / “책임 소재의 문제” / “개인 연구자료가 아닌 경우가 많아 공동연구자의 의견 수렴 등 공유 절차가 번거로울 것이라고 생각” / “데이터관리, 위탁비용의 부담 및 공동연구자들의 데이터 공유에 대한 동의를 구하기 어려움” / “연구에 이용한 데이터가 국가 db이기에
연구 미완성	“연구 진행 중인 주제들이 있기 때문에” / “연구 종료 후 공유 예정”
공유를 위한 노력	“부가적인 노력이 필요하기 때문에”
데이터 재이용 연구를 진행하여 본인이 생산한 데이터 부재	“대부분 공개되어 있는 데이터를 활용하여 연구를 진행했기 때문에 내가 수집한 1차 데이터가 따로 없음” / “데이터 재이용하는 연구만 진행하였음” / “2차 데이터만 주로 활용해서” / “데이터의 높은 수집 비용으로 직접 수집한 데이터 없음”

데이터가 아닌 개인 연구자가 생산하는 데이터의 공유 및 재이용에 대해 잘 알지 못하거나, 공유를 하더라도 연구 주제나 목적 및 방법에 따라 데이터를 처리하고 활용하는 과정이 복잡하기 때문에 이용이 쉽지 않을 것이다(예상).

둘째, 공유 기회의 문제이다. 나의 연구데이터를 이용하고 싶다고 요청한 연구자가 없었기 때문에 공유하지 않았거나, 공유하고 싶어도 개인이 생산한 연구데이터를 공유할 수 있는 플랫폼이 부재하며 신뢰할 수 없다고 하였다. 연구자가 먼저 적극적으로 연구데이터를 공유해야 할 필요성을 느끼지 못하고 있으며, 또한 공유와 재이용을 활발하게 하기 위한 환경적 요건이 부족함을 알 수 있다. 신뢰할 수 있는 데이터 리포지터리 등 구축 및 관련 서비스를 제공하여 연구데이터 공유에 대한 인식을 높이고 쉽게 공

유할 수 있는 환경을 제공해야 할 것이다.

셋째, 데이터에 대한 윤리적 문제 및 데이터 소유권에 관한 문제 때문이다. 연구대상자에 대한 추가적인 동의를 구할 수 없다는 우려사항은 데이터에 대한 소유권과 연구대상자의 동의 문제가 최근 국내에서 중요한 논의점으로 나타나고 있다는 김지현(2012)의 주장과 일치하였다. 또한 공동연구를 진행한 경우 공유하기 위해 공동 연구자들의 동의를 구해야 한다는 점과 국가나 기관의 지원을 받아 진행한 연구의 경우 데이터에 대한 소유권이 개인에게 있지 않다.

넷째, 연구가 아직 진행 중이기 때문이다. 이는 연구 성과와 관련이 있는 문제이다. 즉 연구가 종료된 후 본인의 연구 실적으로 반영되는 연구 논문이나 특허 등록 등 지적재산권을 확보

하기 전에는 데이터 공개를 꺼리는 것이다(김은정, 2012, 74-75). 실제 응답을 통해서도 연구가 진행 중이기 때문에 공유하지 않았지만, 연구가 종료된 후에는 공유할 의향이 있다.

마지막으로는 공유를 위해 부가적인 노력이 필요하기 때문이다. 김지현(2012:2015)의 연구에서도 연구자들이 연구데이터를 공유하기 위해 메타데이터 및 설명 자료 작성 등의 일이 업무 부담을 가중시키는 일로 인식하고 있다. 또한 앞서 분석데이터가 가장 많이 이용 및 생산되고 있음을 확인한 바와 같이, 이미 생산된 데이터를 이용하여 연구를 진행하였기 때문에 본인이 생산한 데이터가 없다고 인식하고 있었다.

이러한 공유에 대한 인식과 관행 및 재이용 경험과의 관계는 후속 연구를 통해 더욱 심도 있게 연구될 필요가 있으나, 데이터 재이용을 통해 학술적 성과를 얻은 연구자들도 공유에 있어서는 여러 문제에 부딪히고 있다는 것을 확인했다는 점에서 의의가 있다. 다시 말해 연구데이터를 재이용한 경험이 있는 연구자는 비교적 연구데이터 공유에 대한 인식이 높을 것이라고 예상할 수 있으나, 실제로는 연구데이터를 공유에 대한 인식이 높지 않을 뿐더러 다양한 문제에 부딪히며 공유를 하지 않는다는

것이다.

2) 영향 요인

① 재이용 유용성: 연구자들이 인지하고 있는 연구데이터 재이용의 이점 및 재이용하도록 하는 동기이다. 이를 측정하기 위해 하위 요인인 지식의 확장과 효율성에 따라 구성(6개 문항), 결과는 <표 14>와 같다.

지식의 확장과 관련된 문항인 유용1_1부터 유용1_3까지는 응답자들이 평균적으로 4점(그렇다)에 응답했다. 재이용 유용성과 관련된 유용2_1부터 유용2_3까지는 모두 평균이 4.0 이상으로 나타나 4점(그렇다)에 응답했다.

즉 재이용 유용성은 대부분의 응답자들의 재이용 유용성에 대한 인지 정도가 높음을 알 수 있다.

② 재이용 위험성: 재이용을 함으로써 처할 수 있다고 생각되는 걱정과 잠재된 위험 사항을 말한다. 이를 측정하기 위해 문헌 검토를 통해 확인한 저평가에 대한 두려움, 윤리적 문제 위반, 재이용 과정에서 실수, 숨겨진 오류에 대한 취약성 요인을 도출하여 6개의 문항을 구성하였다(<표 15> 참조).

저평가에 대한 두려움과 관련된 위험1_1, 위험1_2은 2(그렇지 않다), 윤리적 문제 위반에

<표 14> 연구데이터 유용성

하위 요인	측정 항목		평균	표준편차
지식의 확장	유용1_1	이전 연구 결과의 재검증	3.84	.794
	유용1_2	이전 연구의 재현가능	3.70	.814
	유용1_3	학문적 발전에 기여	4.06	.722
재이용 유용성	유용2_1	데이터 수집 작업의 중복 방지	4.24	.769
	유용2_2	연구와 관련된 금전적비용 절감	4.51	.731
	유용2_3	연구에 소요되는 시간절약	4.44	.736

〈표 15〉 연구데이터 재이용 위험성

하위 요인	측정 항목		평균	표준편차
저평가에 대한 두려움	위험1_1	연구를 소홀히 하는 것으로 인식되어 좋지 못한 평가 우려	2.39	1.043
	위험1_2	가치가 낮은 연구로 인식되어, 좋지 못한 평가 우려	2.34	1.084
윤리적 문제 위반	위험2_1	윤리강령 및 데이터 보호 규범 위반 우려	2.67	1.083
	위험2_2	연구대상자의 추가적 동의 없이 연구 진행	3.20	1.038
재이용 과정에서 실수	위험3_1	데이터 잘못 해석으로 인한 오류	3.05	1.161
숨겨진 오류에 대한 취약성	위험3_2	데이터 오류 파악 어려움	3.39	1.161

관련된 위험2_1, 위험2_2는 평균적으로 3(보통)에 응답한 것을 알 수 있다. 즉 대부분의 연구자들은 저평가에 대한 걱정을 하지 않고 있는 것이다. 응답자의 79.2%가 주로 양적 연구를 하며, 질적 연구를 주로 하는 연구자가 1%도 되지 않음으로 미루어보아, 저평가 및 윤리적 문제에 대한 우려 사항은 주로 사회과학분야의 질적 연구자들에게 보이는 문제라는 선행연구(Bishop, 2015; Goodwin, 2012)에서 확인한 내용과 일치하였다.

한편 Niu(2009)는 양적 연구자들은 저평가에 대한 걱정은 크게 하지 않으나, 데이터를 잘못 해석하거나 데이터 자체에 오류가 있어 잘못 사용하게 되는 점을 걱정한다고 하였다. 즉 저평가에 대한 걱정보다는 숨겨진 오류에 대한 취약성 등 데이터 자체의 오류를 우려하였다.

③ 재이용하기 위한 노력: 데이터를 재이용하기 위해 연구자가 예측하고 있는 작업 혹은 노력을 말하며, 이를 측정하기 위해 기존 데이터로 새로운 발견, 데이터에 접근, 불일치 처리, 기존 연구의 이해로 구분하여 총 8문항을 구성하였다.

재이용하기 위한 노력과 관련된 문항에 5(매우 필요하다)에 답변한 경우 ‘이러한 노력이 매우 필요하다’, 즉 ‘이러한 노력을 하지 않으면

재이용하기 어렵다’라는 의미이다. 따라서 다른 문항과는 다르게 척도가 커지는 방향이 부정적으로 나아가기 때문에 반대로 역코딩하였다(〈표 16〉 참조).

노력1_1부터 노력4_2까지 연구데이터를 재이용하기 위한 노력과 관련된 항목은 모두 약 1.7점으로, 연구자들이 연구데이터 재이용을 하기 위해서는 많은 노력이 필요하다고 생각하고 있음을 알 수 있다. 즉 이는 그러한 노력 없이는 연구데이터를 재이용하기 어렵다는 것을 의미한다.

④ 재이용 가능성: 데이터의 속성을 말한다. 이는 데이터 설명자료 제공, 데이터 적합성, 데이터 신뢰성, 데이터 품질, 연구의 엄격성의 하위요인으로 이루어지며 총 9개의 응답 결과는 〈표 17〉과 같다.

다음 〈표 18〉을 보면 가능1_1부터 가능5_1까지는 대부분 4(중요하다)에 응답이 된 것을 확인할 수 있다. 가능5_2의 평균은 3.87로 다른 문항에 비해 조금 낮기는 하지만 연구데이터 재이용 가능성과 관련된 문항은 전반적으로 모두 긍정적으로 평가되었다고 볼 수 있다.

따라서 연구자들은 데이터 재이용을 가능하게 하는 데이터의 속성을 중요하게 생각하고 있음을 알 수 있다.

〈표 16〉 연구데이터 재이용하기 위한 노력

하위 요인	측정 항목		평균	표준편차
기존 데이터로 새로운 발견	노력1_1	재이용 연구의 중요성을 정당화하는 노력	4.20	.754
	1_1 역문항		1.80	.754
	노력1_2	재이용 할 수 있는 연구방법 찾는 노력	4.26	.789
	1_2 역문항		1.75	.789
데이터에 접근	노력2_1	재이용이 가능한 데이터를 찾는 노력	4.37	.669
	2_1 역문항		1.63	.669
	노력2_2	데이터 재이용 권한을 얻는 노력	3.87	.826
	2_2 역문항		2.13	.826
불일치 처리	노력3_1	재이용을 위해 연구 설계를 조정하는 노력	4.06	.858
	3_1 역문항		1.93	.858
	노력3_2	연구방법에 맞게 데이터를 적절하게 조작/처리 노력	4.33	.703
	3_2 역문항		1.67	.703
기존 연구의 이해	노력4_1	데이터가 생산된 기존 연구를 이해하는 노력	4.13	.837
	4_1 역문항		1.87	.837
	노력4_2	데이터 자체를 이해하는 노력	4.47	.648
	4_2 역문항		1.53	.648

〈표 17〉 연구데이터 재이용 가능성

하위 요인	측정 항목		평균	표준편차
데이터 설명자료 제공	가능1_1	데이터 설명자료의 제공	4.54	.563
데이터 적합성	가능2_1	연구 주제와 목적에 부합	4.13	.817
	가능2_2	필요한 데이터 형식에 부합	4.22	.666
데이터 신뢰성	가능3_1	신뢰할 수 있는 출처에서 생산	4.65	.501
	가능3_2	신뢰할 수 있는 기관에 기탁되어 있음	4.50	.691
데이터 품질	가능4_1	누락된 데이터 없이 완전	4.06	.925
	가능4_2	데이터가 정확하고 정교	4.42	.741
연구의 엄격성	가능5_1	잘 설계된 연구의 생산물	4.31	.761
	가능5_2	연구 목적을 달성한 연구의 성과물	3.87	.939

〈표 18〉 연구데이터 재이용 용이성

하위 요인	측정 항목		평균	표준편차
데이터 리포지터리 이용	용이1_1	나의 소속기관에 데이터 리포지터리가 존재	3.87	.812
	용이1_2	데이터가 기탁되어 있는 리포지터리 찾을 수 있음	4.07	.741
데이터 생산자와 소통	용이2_1	데이터 1차 생산자에게 쉽게 연락하여 데이터에 대한 정보 획득 가능	3.78	.930
재이용 지원 및 서비스	용이3_1	재이용 데이터를 찾는 방법에 대한 교육을 받을 수 있음	3.59	.930
	용이3_2	재이용에 유용한 기술과 관련된 교육을 받을 수 있음	3.67	.967
	용이3_3	소속기관에 데이터 재이용을 돕는 담당자가 있음	3.61	.993

⑤ 재이용 용이성: 재이용을 보다 쉽게 할 수 있도록 돕는 조건을 의미한다. 이는 데이터 리포지터리, 데이터 생산자와의 소통, 재이용 지원 및 서비스의 하위요인으로 이루어진다(〈표 18〉 참조). 연구자들은 데이터 리포지터리 이용과 관련

된 용이1_2는 평균적으로 4(중요하다)에 응답을 하였다. 그 외 용이1_1과 데이터 생산자와의 소통과 관련된 용이2_1, 재이용 지원 및 서비스와 관련된 용이3_1, 용이3_2, 용이3_3은 평균적으로 3(보통)에 응답을 하였다. 다시 말해, 데이터 재이용을 돕는 환경 요소 중 데이터 리포지터리가 약간의 중요성을 보이니, 이를 제외하고는 중립적으로 평가되었다.

⑥ 사회적 조건: 사회적 조건은 데이터 재이용 연구자가 속한 사회적 분위기를 의미하며, 재이용 수용성과 동료의 권장으로 구분(4개 문항)된 결과는 <표 19>와 같다.

재이용 연구자의 사회적 조건에 관한 응답은 모두 평균적으로 3(보통)에 응답이 되었다. 하지만 재이용 수용성과 관련된 사회1_1과 사회1_2는 평균 3.92로 4에 가까운 것을 알 수 있다. 중립적으로 평가되었으나, 다소 긍정적인 편이라고 볼 수 있다.

⑦ 재이용 의도: 연구자가 재이용하고자 하는 의지를 말하며, 계획행동이론(TPB)에서는 이러한 의도와 실제 행동 사이에 강한 연관성이 있다고 하였다.

재이용 의도(내 연구를 위해 다른 연구자가 이미 수집한 연구데이터를 이용할 의도가 있는 것)에 대해서 연구자들은 평균4.24(그렇다), 표준편차(.744)로 응답하였다. 이는 연구자들이 앞으로 연구를 진행함에 있어 연구데이터 재

용을 고려할 것이라는 의미이며, 이러한 의지는 앞으로 실제 연구데이터 재이용 행동으로 이어질 가능성이 매우 크다. 연구데이터 재이용 경험이 있는 연구자들은 또 다시 연구데이터를 재이용 할 의도가 충분하다는 것이다. 따라서 연구데이터 재이용 의도에 영향을 미치는 요인을 분석하여 연구데이터 재이용 의도에 방해가 되는 요인을 낮추고, 재이용 의도를 높이는 요인을 강화시켜, 연구데이터 재이용이 활성화 될 수 있도록 해야 한다.

3.2 연구분석

3.2.1 신뢰성과 타당성

요인분석 방법을 통해 측정 도구의 신뢰성과 타당성을 검증하였다. 신뢰성은 크론바하 알파 계수(Cronbach's α)를 통해 내적 일관성을 측정 방법으로 사용하였다. 타당성은 탐색적 요인분석을 통해 검증하였다. 요인분석은 변수들 간의 상관을 기초로 분석이 실시되며 상관관계는 KMO(Kaiser-Meyer-Olkin)와 Bartlett의 구형성 검증으로 확인할 수 있다. Bartlett의 구형성 검증은 변수들이 모두 상관관계가 없다는 귀무가설과 상관관계가 있다는 대립가설을 설정하여, 유의 확률 $P < .05$ 로 대립가설이 채택되는지를 통해 확인한다. 하지만 실제 요인분석 결과의 타당성을 확인하기 위해서는 요인 적재 값

<표 19> 연구데이터 재이용 연구자의 사회적 조건

하위 요인	측정 항목		평균	표준편차
재이용 수용성	사회1_1	나의 연구 분야에서 일반적으로 데이터 재이용이 가능한 것	3.98	.781
	사회1_2	나의 연구 분야에서 보편적으로 데이터를 재이용 하고 있는 것	3.86	.862
동료의 권장	사회2_1	동료들이 연구데이터 재이용을 권장, 지원하는 것	3.54	.921
	사회2_2	나의 동료들이 연구데이터 재이용을 자주 하는 것	3.53	.934

과 공통성(요인모델에서 변수가 차지하는 상대적 기여도)도 함께 고려해야 한다(강현철, 2013, 589; 김효창, 2018, 234). 이는 일반적으로 0.40을 기준으로 하지만, 이 연구는 0.50을 기준으로 공통성을 확인하였다. 이러한 과정을 통해 신뢰성과 타당성을 저해하는 측정 항목은 분석에서 제외하여 신뢰성과 타당성을 확보하였다.

요인분석의 추출방법은 타당성을 저해하는 항목을 쉽게 제거할 수 있는 주성분 분석(김홍길, 2016, 322)과 Varimax 회전방식을 사용하였으며, 요인은 고유값 1.0 이상을 기준으로 추출하였다. 그 결과 KMO 값은 0.761, 유의 확률은 0.000으로 확인되었다. 하지만 유용1_3의 공통성은 0.454, 가능1_1은 0.473으로 낮게 확인되어 이를 제거하고 다시 분석하였다. 그 결과 KMO 값은 0.753, 유의 확률은 0.000으로 확인되었으며, 공통성 또한 모든 항목에서 0.50 이상으로 나타났다. 하지만 가능3_2와 가능3_1의 요인 적재 값이 0.50 미만으로 나타나 어떤 요인도 유의미하게 설명하지 못한다고 판단하여 제거하였다. 이와 같은 방법으로 요인분석을 다시 실시하여 요인 적재 값이 0.50 미만으로 나타나는 항목(노력1_1, 위험2_2)을 제거한 후, 최종적으로 타당성을 확보하였다. 따라서 타당성을 저해하는 요소로 제거된 항목은 유용1_3(공통성 0.454), 가능1_1(공통성 0.473), 가능3_2(적재 값 0.458), 가능3_1(적재 값 0.412), 노력1_1(적재 값 0.486), 위험2_2(적재 값 0.478)이다. 요인분석과 신뢰성 검증결과는 <표 20>과 같다.

분석 결과, 고유값을 기준으로 총 10개의 요인이 추출되었으며 전체 설명력은 71.489로 일반적인 기준인 60%를 상회하는 것으로 확인되었다. KMO 값은 0.743, 유의 확률은 0.000으로

나타났으며, 공통성과 요인 적재 값 또한 각 항목 별로 0.50 이상으로 나타나 해당 요인을 측정하기에 타당함을 입증하였다. 크론바하 알파 계수도 각 요인 별 모두 0.60 이상으로 나타나 측정 항목들의 신뢰성이 비교적 높다.

‘재이용 용이성’은 용이1_1, 용이1_2, 용이2_1, 용이3_1, 용이3_2, 용이3_3으로, ‘사회적 조건’은 사회1_1, 사회1_2, 사회2_1, 사회2_2로 구분되며 이 두 요인은 문헌검토를 통해 도출한 요인과 같다. 그 외 문헌검토를 통해 도출한 나머지 4개 요인은 요인분석을 통해 도출된 요인과 약간의 차이가 있다. 하지만 새로운 요인을 형성한 것이 아닌, 4개의 요인에서 내용의 관련도와 주체에 따라 보다 세부적으로 요인이 구분되어 사회과학분야 연구자들의 재이용 영향요인을 보다 면밀하게 분석할 수 있었다.

먼저 문헌검토를 통해 도출된 ‘재이용 유용성’은 재이용을 통해 학문과 지식이 발전하는 지식의 확장과 연구와 관련된 비용과 시간을 절약하며 중복 작업을 방지하는 연구의 효율성의 하위요인으로 이루어졌다. 이는 연구데이터를 재이용의 이점 혹은 재이용하도록 하는 동기를 말한다. 하지만 ‘재이용 유용성’은 요인분석을 통해 이점을 얻는 주체에 따라 구분되었다. 이에 따라 구분된 요인을 ‘학문적 유용성’과 ‘연구 효율성’으로 명명하였다. 유용1_1과 유용1_2로 구분된 ‘학문적 유용성’은 데이터 재이용을 통해 연구자가 학문적 발전에 기여를 하는 것을 말하며, 유용2_1, 유용2_2, 유용2_3으로 구분된 ‘연구 효율성’은 재이용함으로써 연구자 본인이 얻게 되는 이점을 말한다.

‘재이용 위험성’은 데이터를 재이용함으로써 연구자가 우려하는 사항과, 데이터 자체의 결함

〈표 20〉 신뢰성과 타당성

요인	측정항목	요인분석				신뢰도
		적재 값	공통성	고유 값	설명력	
학문적 유효성(A1)	유용1_1	0.892	0.845	1.416	5.866	.856
	유용1_2	0.898	0.848			
	유용1_3	삭제				
연구 효율성(A2)	유용2_1	0.705	0.569	1.754	6.677	.766
	유용2_2	0.802	0.775			
	유용2_3	0.801	0.711			
연구자의 우려사항(B1)	위험1_1	0.868	0.827	2.565	7.145	.826
	위험1_2	0.920	0.872			
	위험2_1	0.667	0.652			
	위험2_2	삭제				
데이터 취약성(B2)	위험3_1	0.804	0.778	1.015	5.223	.748
	위험4_1	0.890	0.816			
직접적 노력(C1)	노력1_1	삭제				
	노력1_2	0.608	0.696	2.774	8.207	.766
	노력2_1	0.744	0.645			
	노력2_2	0.650	0.581			
	노력3_1	0.787	0.705			
	노력3_2	0.625	0.537			
간접적 노력(C2)	노력4_1	0.632	0.668	1.288	5.827	.690
	노력4_2	0.80	0.794			
재이용 적합성(D1)	가능1_1	삭제				
	가능2_1	0.709	0.610	1.168	5.735	.616
	가능2_2	0.709	0.593			
	가능5_2	0.582	0.659			
	가능3_1	삭제				
	가능3_2					
데이터 완전성(D2)	가능4_1	0.772	0.708	1.627	6.006	.788
	가능4_2	0.800	0.792			
	가능5_1	0.618	0.635			
재이용 용이성(E1)	용이1_1	0.628	0.693	6.281	10.531	.841
	용이1_2	0.62	0.743			
	용이2_1	0.594	0.535			
	용이3_1	0.819	0.742			
	용이3_2	0.802	0.741			
	용이3_3	0.728	0.647			
사회적 조건 (F1)	사회1_1	0.85	0.792	3.703	9.678	.901
	사회1_2	0.887	0.807			
	사회2_1	0.851	0.782			
	사회2_2	0.874	0.795			
KMO 0.743		P 0.000		설명력 71.489%		

혹은 정보 부족으로 구분된다. 이에 따라 재이용 위험성 요인은 위험1_1, 위험1_2로 이루어진 '연구자의 우려사항'과 위험3_1과 위험4_1로 이루어진 '데이터 취약성' 요인으로 구분된다.

'재이용하기 위한 노력'은 데이터를 재이용하기 위한 직접적인 노력과 간접적인 노력으로 구분되었다. '재이용하기 위한 직접적 노력'은 노력1_2, 노력2_1, 노력2_2, 노력3_1, 노력3_2로 구분되며, 데이터를 재이용을 하기 위해 데이터를 찾고, 처리하는 등 연구 진행 과정에서 필요한 직접적인 노력을 의미한다. 반면 노력4_1, 노력4_2은 데이터 재이용을 하기 위한 간접적인 노력에 속한다. 이렇게 요인이 구분되었지만 실제 데이터를 재이용하기 위해서는 직접적·간접적 노력이 모두 필요하다.

재이용 가능성은 데이터 재이용 가능 여부를 결정하기 전 평가하게 되는 데이터와 관련된 속성을 의미한다. 이는 데이터 자체에 대한 속성과 연구의 목적과 방법에 부합하는 정도로 구분된다. 가능2_1, 가능2_2, 가능5_2로 구분된 '재이용 적합성'은 데이터가 목적 달성을 위해 가진 내용적, 형식적 조건을 말하며, 가능4_1, 가능4_2, 가능5_1로 구분된 '데이터 완전성'은 데이터를 재이용함에 문제가 없을 만큼 정확하고, 정교한 것을 의미한다. 데이터가 완전하고 이용

하기에 적합할 때 재이용 가능하다고 볼 수 있다. 문헌 검토와 요인분석을 통해 도출해 낸 요인은 <표 21>과 같다.

3.2.2 상관분석

앞서 요인분석을 통해 도출된 10가지 요인(독립변수)간에 서로 어떤 관계가 있는지 상관분석을 실시하여 파악하고, 다중공선성이 존재하는지 확인하였다. 상관분석은 피어슨의 적률상관계수를 사용하였다. 상관관계 분석 결과는 <표 22>와 같다.

학문적 유용성은 연구 효율성($r=.210, p=.005$)과 데이터 완전성($r=.216, p=.004$)과 낮은 양적 상관관계가 나타났으며, 간접적 노력($r=-.219, p=.003$)과는 낮은 음적 상관관계가 나타났다. 그 외로 연구자의 우려사항($r=-.033, p=.662$), 데이터 취약성($r=-.123, p=.101$), 직접적 노력($r=-.068, p=.365$), 재이용 적합성($r=.183, p=.014$), 재이용 용이성($r=.139, p=.064$), 사회적 조건($r=.086, p=.252$)과는 거의 상관관계가 없는 것으로 나타났다.

연구 효율성은 연구자의 우려사항($r=-.243, p=.001$), 데이터 완전성($r=.273, p=.000$), 사회적 조건($r=.304, p=.000$)과 낮은 양적 상관관계를 나타냈다. 그 외 데이터 취약성($r=-.016$,

<표 21> 요인 도출

문헌 검토	요인 분석
재이용 유용성 (A)	학문적 유용성 (A1) / 연구 효율성 (A2)
재이용 위험성 (B)	연구자의 우려사항 (B1) / 데이터 취약성 (B2)
재이용하기 위한 노력 (C)	직접적 노력 (C1) / 간접적 노력 (C2)
재이용 가능성 (D)	재이용 적합성 (D1) / 데이터 완전성 (D2)
재이용 용이성 (E)	재이용 용이성 (E1)
사회적 조건 (F)	사회적 조건 (F1)

〈표 22〉 상관분석¹⁾

	A1	A2	B1	B2	C1	C2	D1	D2	E	F
A1	1.00									
A2	.210**	1.00								
B1	-.033	-.243**	1.00							
B2	-.123	-.016	.328**	1.00						
C1	-.068	-.187*	-.055	-.094	1.00					
C2	-.219**	-.036	-.016	.040	.439**	1.00				
D1	.183*	.179*	-.019	-.077	-.223*	-.260**	1.00			
D2	.216**	.273**	.020	-.168*	-.325**	-.276**	.492**	1.00		
E	.139	.075	.225*	0	-.232**	-.265**	.468**	.451**	1.00	
F	.086	.304**	-.061	-.072	-.105	.001	.166*	.187*	.159*	1.00
의도	.117	.441**	-.310**	-.204**	-.216**	-.116	.127	.152*	.047	.399**

p=.835), 직접적 노력($r=-.187$, $p=.012$), 간접적 노력($r=-.036$, $p=.634$), 재이용 적합성($r=.179$, $p=.017$), 재이용 용이성($r=.075$, $p=.323$)과는 상관관계가 거의 없는 것으로 나타났다.

연구자의 우려사항은 데이터 취약성($r=.328$, $p=.000$)과 재이용 용이성($r=.225$, $p=.003$)과는 낮은 양적 상관관계가 나타난 반면, 그 외 직접적 노력($r=-.055$, $p=.467$), 간접적 노력($r=-.016$, $p=.834$), 데이터 적합성($r=-.019$, $p=.805$), 데이터 완전성($r=.020$, $p=.787$), 사회적 조건($r=-.061$, $p=.421$)과는 상관관계가 거의 없는 것으로 나타났다.

데이터 취약성은 직접적 노력($r=-.094$, $p=.210$), 간접적 노력($r=.040$, $p=.599$), 재이용 적합성($r=-.077$, $p=.304$), 데이터 완전성($r=-.168$, $p=.025$), 재이용 용이성($r=.000$, $p=.995$), 사회적 조건($r=-.072$, $p=.343$)과 상관관계가 거의 나타나지 않았다.

반면 직접적 노력은 사회적 조건($r=-.105$, $p=.163$)과는 상관관계가 거의 나타나지 않았으

나, 간접적 노력($r=.439$, $p=.000$)과는 비교적 높은 양적 상관관계가, 재이용 적합성($r=-.223$, $p=.003$), 데이터 완전성($r=-.325$, $p=.000$), 재이용 용이성($r=-.232$, $p=.002$)과는 낮은 양적 상관관계를 보였다.

간접적 노력 또한 사회적 조건($r=.001$, $p=.993$)과 상관관계가 거의 없었으나, 재이용 적합성($r=-.260$, $p=.000$), 데이터 완전성($r=-.276$, $p=.000$), 재이용 용이성($r=-.265$, $p=.000$)과는 낮은 양적 상관관계가 있는 것으로 확인되었다.

재이용 적합성은 데이터 완전성($r=.492$, $p=.000$)과 재이용 용이성($r=.468$, $p=.000$)간에 비교적 높은 양적 상관관계가 나타났고, 사회적 조건($r=.166$, $p=.027$)과는 상관관계가 거의 없었다.

데이터 완전성은 재이용 용이성($r=.451$, $p=.000$)과는 비교적 높은 양적 상관관계가 나타났으며, 사회적 조건(F)은 데이터 완전성($r=.187$, $p=.012$)과 재이용 용이성(E)($r=.149$, $p=.048$)간에 상관관계가 거의 나타나지 않았다.

1) 표의 간결함을 위해 〈표 22〉에서 구분한 바와 같이, A1=학문적 유용성, A2=연구 효율성, B1=연구자의 우려사항, B2=데이터 취약성, C1=직접적 노력, C2=간접적 노력, D1=재이용 적합성, D2=데이터 완전성, E=재이용 용이성, F=사회적 조건으로 표기하였다.

종속변수인 재이용 의도(의도)와의 상관관계를 살펴보면 학문적 유용성($r=.117$, $p=.121$), 간접적 노력($r=-.116$, $p=.123$), 재이용 적합성($r=.127$, $p=.091$), 데이터 완전성($r=.152$, $p=.043$), 재이용 용이성($r=.047$, $p=.533$) 간에는 상관관계가 거의 나타나지 않았다. 반면 연구 효율성($r=.441$, $p=.000$)과는 비교적 높은 양적 상관관계가 나타났으며, 사회적 조건($r=.399$, $p=.000$) 간에는 낮은 양적 상관관계가 나타났다. 반면 직접적 노력($r=-.216$, $p=.004$)과 연구자의 우려사항($r=-.310$, $p=.000$)과 데이터 취약성($r=-.204$, $p=.006$) 간에는 낮은 음적 상관관계가 나타났다.

이상의 내용을 정리하자면 재이용 의도와 연구 효율성, 사회적 조건, 직접적 노력, 연구자의 우려사항, 데이터 취약성과는 상관관계가 있음을 확인하였다. 또한 독립변수간의 상관정도도 높지 않았기 때문에(.50 이하) 다중공선성의 문제는 없는 것으로 판단하였다.

3.2.3 회귀분석

연구데이터 재이용 의도와 영향을 주는 요인간의 인과관계를 조사하여 보다 설득력 있는 설명을 하고자 다중회귀분석을 실시하였다. 종속변수는 연구데이터 재이용 의도, 독립변수는 요인분석을 통해 도출된 10개의 요인으로 설정하여 회귀분석을 실시한 결과는 <표 23>과 같다.

분석 결과 F 값은 19.313(유의확률은 <.001)로 회귀모형이 통계적으로 유의함을 알 수 있다. 또한 DW는 1.846으로 2에 근접하여 잔차 간의 상관관계가 없는 것으로 확인되었고, Cook의 거리 또한 최대값이 1.0미만으로 극단치 존재 여부도 문제가 없음을 확인하였다. 공차 한계는 모두 .100이상이며, VIF는 10미만으로 다중공선성의 문제도 없음을 확인하였다. 즉 연구데이터 재이용 의도는 독립변수들에 의해 35.8%(수정된 R^2 으로는 24.0%)의 수준에서 유의하게 설명되고 있다고 할 수 있다. 따라서 이 회귀모형을 통해 연구데이터 재이용 의도를 예측하는 것은 적합하다고 보았다.

<표 23> 회귀분석

모형 1	표준화 계수	t	유의확률	공차	VIF
(상수)	3.126	4.919	<.001		
학문적 유용성	-0.001	-0.02	0.984	0.878	1.139
연구 효율성	0.358	4.14	<.001	0.752	1.33
연구자의 우려사항	-0.148	-2.581	0.011	0.765	1.308
데이터 취약성	-0.102	-2.073	0.04	0.818	1.222
직접적 노력	-0.203	-2.07	0.04	0.719	1.392
간접적 노력	-0.063	-0.754	0.452	0.724	1.382
재이용 적합성	-0.006	-0.064	0.949	0.66	1.515
데이터 완전성	-0.071	-0.812	0.418	0.601	1.665
재이용 용이성	0.003	0.039	0.969	0.649	1.541
사회적 조건	0.275	4.285	<.001	0.875	1.143

F = 9.313

P < .001

R = .598

R² = .358

adj R² = .320

DW = 1.846

Cook의 거리 최대값 = .225

각 요인들의 연구데이터 재이용 의도에 대한 영향력 및 통계적 유의성을 확인한 결과 학문적 유용성, 간접적 노력, 재이용 적합성, 데이터 완전성, 재이용 용이성은 t 값이 ± 1.96 보다 작고, 유의확률이 0.05보다 크기 때문에 연구데이터 재이용 의도에 유의미한 영향을 미친다고 판단할 수 없다.

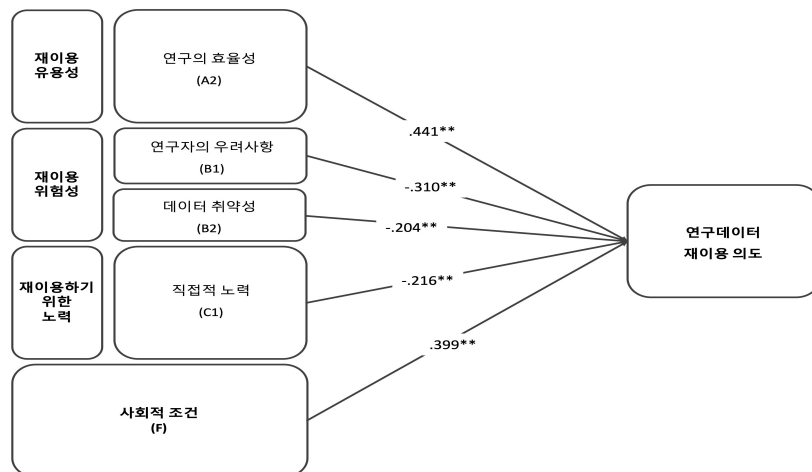
반면 연구효율성 t 값은 4.14($P < .001$), 연구자의 우려사항의 t 값은 -2.581($p = 0.011$), 데이터 취약성의 t 값은 -2.073($p = 0.04$), 직접적 노력의 t 값은 -2.07($p = 0.04$), 사회적 조건의 t 값은 4.285($p < .001$)로 모두 연구데이터 재이용 의도에 유의한 영향을 미치고 있음이 확인되었다.

독립변수들의 상대적 영향정도를 확인할 수 있는 표준화 계수를 확인하면 연구 효율성(0.358), 사회적 조건(0.275), 직접적 노력(-0.203), 연구자의 우려사항(-0.148), 데이터 취약성(-0.102) 순서로 연구데이터 재이용 의도에 영향을 미치고 있음을 알 수 있다.

이러한 결과 이용 의도에 영향을 미치는 요인을 정리하면 <그림 3>과 같다.

3.3 가설 검정

다음 <표 24>는 이 연구의 가설 검정이다.



<그림 3> 연구결과

<표 24> 가설 검정

가설		결과
가설 1	연구데이터 재이용으로 인한 이점은 연구자의 연구데이터 재이용 의도에 긍정적 영향을 미칠 것이다.	부분채택
가설 2	연구데이터 재이용으로 인한 우려사항과 잠재된 위험은 연구자의 연구데이터 재이용 의도에 부정적 영향을 미칠 것이다.	채택
가설 3	연구데이터 재이용을 위한 노력은 연구데이터 재이용 의도에 부정적 영향을 미칠 것이다.	부분채택
가설 4	연구데이터 재이용 가능성은 연구데이터 재이용 의도에 긍정적 영향을 미칠 것이다.	기각
가설 5	연구데이터 재이용에 용이한 환경적 조건은 연구데이터 재이용 의도에 긍정적 영향을 미칠 것이다.	기각
가설 6	연구자의 사회적 조건은 연구데이터 재이용 의도에 긍정적 영향을 미칠 것이다.	채택

4. 결 론

이 연구의 결과는 다음과 같다.

① 연구데이터란 연구의 계획부터 종료까지 전 과정에서 생산되는 1차 혹은 이를 분석·처리한 2차 자료로, 누구에게나 공개할 수 있으며, 후속 연구의 진행 혹은 연구 결과의 검증과 재현 등 재이용 및 보존의 가치가 지속적인 데이터이다. 연구데이터에의 유형은 어떠한 방법을 통해 생산되었는가에 따라 실험데이터, 관찰데이터, 분석데이터로 구분할 수 있다. ② 대부분의 연구자들은 주로 데이터 리포지터리, 데이터 관리 시스템, 연구데이터 DB와 같은 시스템을 이용하여 재이용 연구데이터를 획득하였다. 또한 실험과 관찰을 통해 생산된 데이터를 분석·처리한 분석 데이터를 가장 많이 재이용하는 것으로 나타났다. ③ 연구자들은 연구데이터 재이용을 통해 성공적으로 연구 결과물을 산출하였으나, 반대로 공유에 있어서는 여러 문제에 직면한다는 것을 확인하였다. 공유에 대한 인식 및 기회의 부족, 미흡한 환경 조건, 데이터 재이용 과정의 어려움으로 인한 낮은 활용성, 데이터에 대한 윤리적 문제 및 소유권, 연구 성과에 대한 문제, 데이터 공유를 위한 작업의 부담의 문제 때문에 공유를 하지 않는 것으로 나타났다. 이는 연구데이터를 재이용한 경험이 있는 연구자는 비교적 연구데이터 공유에 대한 인식이 높을 것으로 예상한 것과 달리 실제로는 다양한 문제에 부딪히며 공유하지 않음을 알 수 있었다. 특히 공유하지 않는 이유 중, '재이용 과정의 어려움으로 인한 낮은 활용성'은 실제 재이용 경험에서 연구자가 어려움을 겪었음을 예측할 수 있다. 따라서 이러한 내용을 종합적으로 반영해

본다면, 데이터 리포지터리를 구축하거나 관련 서비스를 제공하여 보다 쉽게 연구데이터를 공유할 수 있는 환경을 제공할 필요가 있음을 알 수 있다. ④ 연구데이터 재이용에 영향을 미치는 요인은 문헌 검토를 통해 재이용 유용성, 재이용 위험성, 재이용하기 위한 노력, 재이용 가능성, 재이용 용이성, 사회적 조건으로 도출하였다. 선행연구를 바탕으로 재이용 유용성, 재이용 가능성, 재이용 용이성, 사회적 조건은 연구데이터 재이용 의도에 긍정적인 영향을 미칠 것이며, 재이용 위험성과 재이용하기 위한 노력은 연구데이터 재이용 의도에 부정적 영향을 미칠 것으로 가설을 검증하였다.

실증분석 결과를 정리하면 다음과 같다.

먼저, 요인분석을 통해 문헌 검토를 통해 도출된 6가지 요인을 보다 구체화 하여 타당성을 확보하였다. ① 재이용 유용성은 학문적 유용성과 연구의 효율성으로, 재이용 위험성은 연구자의 우려사항과 데이터 취약성, 재이용하기 위한 노력은 직접적 노력과 간접적 노력, 재이용 가능성은 데이터 적합성과 데이터 완전성으로 나뉘었다. 따라서 연구데이터 재이용에 영향을 미치는 요인은 학문적 유용성, 연구의 효율성, 연구자의 우려사항, 데이터 취약성, 재이용하기 위한 직접적 노력, 재이용하기 위한 간접적 노력, 데이터 적합성, 데이터 완전성, 재이용 유용성, 사회적 조건으로 총 10가지 요인으로 도출되었다. ② 위의 도출된 요인이 연구데이터 재이용 의도에 미치는 영향정도를 분석한 결과 연구의 효율성, 사회적 조건은 연구데이터 재이용 의도에 긍정적인 영향을 미치는 것으로 확인되었다. 반면 연구자의 우려사항, 데이터 취약성, 재이용하기 위한 직접적 노력은

연구데이터 재이용 의도에 부정적 영향을 미치는 것으로 나타났다. 영향 정도는 연구 효율성, 사회적 조건, 직접적 노력, 연구자의 우려사항, 데이터 취약성 순으로 나타났다. 그 외 학문적 유용성, 간접적 노력, 재이용 적합성, 데이터 완전성, 재이용 용이성은 재이용 의도에 유의한 영향을 미치지 않는 것으로 나타났다. 연구자들은 연구데이터를 재이용하는데 있어 비용과 시간, 노력의 절약과 학술 커뮤니티 내 연구데이터 재이용의 수용 정도에 따른 사회적 분위기를 가장 중시하고 있는 것을 알 수 있다. 다시 말해 데이터를 새로 수집하는 것보다 재이용하는 것이 훨씬 효율적이라고 판단되고, 주위 동료들이 보편적으로 데이터를 재이용하고 있다면 데이터의 완전성과 적합성이 조금 부족하더라도 재이용하게 될 것이다. 또한 연구데이터 재이용이 활발하게 이루어지고 있지 않는 국내의 실정에서는 연구자가 데이터를 적절하게 처

리하고 변환하는 과정에서 많은 노력이 필요함을 인지하고 있었다. 이는 초보 연구자들을 대상으로 한 선행연구의 결과(Faniel, Kriesberg, & Yakel, 2012; Kriesberg et al., 2013)와 일치함을 확인할 수 있다.

따라서 연구데이터 재이용 생산과 공유를 위한 인프라로 연구데이터 리포지터리 구축, 관리, 이용이 필요하다. 또한 연구데이터 재이용 과정에서 연구자들의 부담을 덜어 줄 수 있는 연구데이터 재이용을 활성화 방안을 제시하면 다음과 같다.

① 연구 효율성을 제고하기 위해 연구데이터에 대한 접근성을 높여야 한다. ② 연구데이터 재이용은 지식과 학문이 발전할 수 있도록 하는 연구의 본질적인 과정임을 알려 연구데이터 재이용에 대한 인지를 높여야 한다. ③ 연구데이터를 재이용하는데 필요한 노력을 줄일 수 있는 방안을 고려한다.

참 고 문 헌

- 강현철 (2013). 구성타당도 평가에 있어서 요인분석의 활용. *대한간호학회지*, 43(5), 587-594.
<http://doi.org/10.4040/jkan.2013.43.5.587>
- 김나연 (2019). 국내 사회과학 학술연구자의 데이터 재이용 행위와 데이터요구 연구. 석사학위논문, 이화여자대학교 대학원 문헌정보학과.
- 김문정, 김성희 (2015). 과학기술분야 연구자의 연구데이터 공유의 영향요인에 대한 연구. *한국문헌정보학회지*, 49(2), 313-334. <http://doi.org/10.4275/KSLIS.2015.49.2.313>
- 김선태, 한선화, 이태영, 김용 (2010). 과학데이터 보존 및 활용모델에 관한 연구. *한국비블리아학회지*, 12(4), 81-93. <http://doi.org/10.14699/kbiblia.2010.21.4.081>
- 김은정, 남태우 (2012). 연구데이터 수집에 영향을 미치는 요인 분석. *정보관리학회지*, 29(2), 27-44.
<http://doi.org/10.3743/KOSIM.2012.29.2.027>
- 김주섭, 김선태, 최상기 (2019). 연구 데이터 관리 및 서비스를 위한 핵심요소의 기능적 요건. *한국문헌*

- 정보학회지, 53(3), 317-344. <http://doi.org/10.4275/KSLIS.2019.53.3.317>
- 김지현 (2012). 대학 내 연구자들의 연구데이터 관리에 관한 연구. 한국도서관·정보학회지, 43(3), 433-455. <http://doi.org/10.16981/kliiss.43.3.201209.433>
- 김지현 (2015). 데이터 관리와 공유에 대한 대학 연구자들의 인식에 관한 연구. 한국문헌정보학회지, 49(3), 413-436. <http://doi.org/10.4275/KSLIS.2015.49.3.413>
- 김홍길 (2016). 커피점포의 선택속성이 고객의 이미지 및 행동의도에 미치는 영향에 관한 연구. 관광레저연구, 28(7), 313-328.
- 김효창 (2018). 혼자서 완성하는 통계분석. 서울: 학지사.
- 디지털 큐레이션 센터. <http://www.dcc.ac.uk> [검색일 2021.09.30]
- 박옥남 (2018). 연구데이터 관리를 위한 온톨로지 설계에 대한 연구. 한국기록관리학회지, 18(1), 101-127. <http://doi.org/10.14404/JKSARM.2018.18.1.101>
- 서울대학교 (2010). 연구윤리지침. 서울대학교.
- 신영란, 정연경 (2012). 국내 인문사회 연구데이터 아카이브의 개선방안에 관한 연구. 한국기록관리학회지, 12(3), 93-115.
- 심원식, 안혜연, 변제연 (2015). 인문학 분야 연구데이터의 수집 및 활용성 증진을 위한 전략연구: 기초 학문자료센터를 중심으로. 한국문헌정보학회지, 49(3), 155-183. <http://www.doi.org/10.4275/KSLIS.2015.49.3.155>
- 심윤희, 김지현 (2019). 국내 대학도서관의 연구데이터관리서비스 개발 방안에 관한 연구: 서울대학교 소속 연구자들의 요구분석을 중심으로. 정보관리학회지, 36(3), 61-80. <http://doi.org/10.3743/KOSIM.2019.36.3.061>
- 전리나 (2017). 연구 데이터 재이용 의향에 영향을 미치는 데이터 리파지토리 환경요인에 관한 연구. 석사학위논문, 연세대학교 대학원 문헌정보학과.
- 최현옥, 이혜영 (2009). 우리나라 공과대학 연구실 연구기록의 체계적 관리 방안에 관한 연구. 한국기록관리학회지, 10(1), 119-139.
- 최희운, 서태설 (2020). 글로벌 연대와 상생의 길 오픈사이언스. 대전: 한국과학기술정보연구원.
- Bishop, L. (2015). Ethical sharing and reuse of qualitative data. *Australian Journal of Social Issues*, 44(3), 255-272.
- Curry, R. G. (2015). Beyond "Data Thrifting": An Investigation of Factors Influencing Research Data Reuse In the Social Sciences. Ph.D. Dissertations - ALL, Syracuse University.
- Faniel, I. M. & Jacobsen, T. (2010). Reusing scientific data: how earthquake engineering researchers assess the reusability of colleagues' data. *Computer Supported Cooperative Work*, 19, 355-375.
- Faniel, I. M. & Zimmerman, A. (2011). Beyond the data deluge: a research agenda for large-scale data sharing and reuse. *The International Journal of Digital Curation*, 6(1), 58-69.
- Faniel, I. M., Kriesberg, A., & Yakel, E. (2012). Data reuse and sense making among novice

- social scientists. *Proc. Am. Soc. Info. Sci. Tech.*, 49(1), 1-10, Available:
<http://www.doi.org/10.1002/meet.14504901068> [cited in September 30, 2021]
- Faniel, I. M., Kriesberg, A., & Yakel, E. (2016). Data reusers' trust development. *Journal of the Association for Information Science and Technology*, 67(6), 1404-1416.
- Goodwin, J. (2012). *The Secondary Analysis of Qualitative Data*. n.p.: Thousand Oaks, CA: SAGE Publications.
- Heaton, J. (2008). Secondary analysis of qualitative data: An overview. (n.p.): Historical Social Research/Historische Sozialforschung.
- Higgins, S. (2008). The DCC curation lifecycle model. *International Journal of Digital Curation*, 3(1), 134-140. Available: <http://doaj.org/article/06ec0a3f0db2431a811034013d01427e> [cited in October 10, 2021]
- Kiecolt, K. J. & Nathan, L. E. (1985). *Secondary analysis of survey data*. (n.p.): Beverly Hills: SAGE Publications.
- Kim, Y. S. & Yoon, A. (2017). Scientists' data reuse behaviors: a multilevel analysis. *Journal of the Association for Information Science and Technology*, 68(12), 2709-2719.
- Kriesberg, A., Frank, R. D., Faniel, I. M., & Yakel, E. (2013). The role of data reuse in the apprenticeship process. *Proc. Am. Soc. Info. Sci. Tech.*, 50(1), 1-10. Available:
<http://www.doi.org/10.1002/meet.14505001051> [cited in September 30, 2021]
- Martin, P. A. (1985). More replication studies needed. *Applied Nursing Research : ANR*, 8(2), 102-103. [http://doi.org/10.1016/s0897-1897\(95\)80574-5](http://doi.org/10.1016/s0897-1897(95)80574-5)
- Mooney, H. & Newton, M. P. (2012). The anatomy of a data citation: discovery, reuse, and credit. *Journal of Librarianship & Scholarly Communication*, 1(1), 1-16.
- Niu, Jinfang (2009). Overcoming inadequate documentation. *Proc. Am. Soc. Info. Sci. Tech.*, 46, 1-14. Available: <http://doi.org/10.1002/meet.2009.145046024> [cited in September 30, 2021]
- OECD (2007). *OECD Principles and Guidelines for Access to Research Data from Public Funding*. (n.p.): OECD.
- OECD (2015). *Making Open Science a Reality*. (n.p.): OECD.
- Open Data Guide. *Open Data Handbook*. (n.d.). Available: <http://opendatahandbook.org>. [cited in October 10, 2021]
- Piwowar, H. A. & Vison, T. J. (2013). Data reuse and the open data citation advantage. *PeerJ* e175, 1-25. <http://doi.org/10.7717/peerj.175>.
- Tenopir, C. Allard, Suzie, Douglass, Kimberly, Aydinoglu, Arsev Umur, Wu, Lei, Read, Eleanor, Manoff, Maribeth, & Frame, Mike (2011). Data Sharing by Scientists: Practices and Perceptions. *PLOS ONE*, 6(6), 1-21. Available:

<http://www.doi.org/10.1371/journal.pone.0021101> [cited in October 10, 2021]

- Yoon, A. & Kim, Y. S. (2017). Social scientists' data reuse behaviors: exploring the roles of attitudinal beliefs, attitudes, norms, and data repositories. *Library and Information Science Research*, 39, 224-233.
- Yoon, A. & Lee, Y. Y. (2019). Factors of trust in data reuse. *Online Information Review* 43 no.7, 1245-1262. Available: <http://www.doi.org/10.1108/OIR-01-2019-0014> [cited in October 10, 2021]
- Yoon, A. (2014). making a square fit into a circle: researchers' experiences reusing qualitative data. *Proc. Am. Soc. Info. Sci. Tech*, 51(1), 1-4. Available: <http://doi.org/10.1002/meet.2014.14505101140>.
- 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 참고문헌에 대한 영문 표기
(English translation of references written in Korean)

- Choi, H. O. & Rieh, H. Y. (2009). A study on the methods for systematic management of research records in engineering school laboratories in Korea. *Journal of Korean Society of Archives and Records Management*, 10(1), 119-139.
- Choi, H. Y. & Seo, T.S. (2020). *Open Science, the Path of Global Solidarity and Coexistence*. Daejeon: Korea Institute of Science and Technology.
- Digital Curation Centre. <http://www.dcc.ac.uk> [cited in September 30, 2021]
- Jeon, L. N. (2017). A Study on Environment Factors of Data Repository Affect Intention to Reuse Shared Research Data. MA. thesis, Dept. LIS, Graduate School of Yeonsei University.
- Kang, H. C. (2013). A guide on the use of factor analysis in the assessment of construct validity. *Journal of Korean Academy of Nursing*, 43(5), 587-594.
<http://www.doi.org/10.4040/jkan.2013.43.5.587>
- Kim, E. J. & Nam, T. W. (2012). Factor analysis of effects on research data collection. *Journal of the Korean Society for Information Management*, 29(2), 27-44.
<http://doi.org/10.3743/KOSIM.2012.29.2.027>
- Kim, H. C. (2018). *Statistical analysis done by himself*. Seoul: Hakjisa.
- Kim, H. G. (2016). A study on the effect of selection attributes on store image and behavioral intention of coffee shop. *Journal of Tourism and Leisure Research*, 28(7), 313-328.

- Kim, J. H. (2012). A study on university researchers' data management practices. *Journal of Korean Library and Information Science Society*, 43(3), 433-455.
<http://doi.org/0.16981/kliss.43.3.201209.433>
- Kim, J. H. (2015). A study on the perceptions of university researchers on data management and sharing. *Journal of Korean Library and Information Science Society*, 49(3), 413-436.
<http://doi.org/10.4275/KSLIS.2015.49.3.413>
- Kim, J. S., Kim, S. T., & Choi, S. K. (2019). The functional requirements of core elements for research data management and service. *Journal of the Korean Library and Information Science*, 53(3), 317-344. <http://doi.org/10.4275/KSLIS.2019.53.3.317>
- Kim, N. Y. (2019). A Study on Data Reuse Behaviors and Data Needs of Academic Researchers in the Korean Social Science Field. MA. thesis, Dept. of LIS., Graduate School of Ewha Womans University.
- Kim, S. T., Han, S. H., Lee, T. Y., & Kim, Y. (2010). A study on a model for using and preserving scientific data. *Journal of the Korean Biblia Society for Library and Information Science*, 12(4), 81-93. <http://doi.org/10.14699/kbiblia.2010.21.4.081>
- Kim, M. J. & Kim, S. H. (2015). A study on the factors affecting sharing of research data of science and technology researchers. *Journal of the Korean Library and Information Science*, 49(2), 313-334. <http://doi.org/10.4275/KSLIS.2015.49.2.313>
- Korean Research Memory. <http://www.krm.or.kr> [cited in September 30, 2021]
- Park, O. N. (2018). A study on ontology design for research data management. *Journal of Korean Society of Archives and Records Management*, 18(1), 101-127.
<http://doi.org/10.14404/JKSARM.2018.18.1.101>
- Seoul National University (2010). Research ethics guidelines. Seoul National University.
- Shim, W. S., Ahn, H. Y., & Byun, J. Y. (2015). Strategic for improving the collection and use of research data in the humanities. *Journal of the Korean Library and Information Science*, 49(3), 155-183. <http://doi.org/10.4275/KSLIS.2015.49.3.155>
- Shim, Y. H. & Kim, J. H. (2019). A study on the development of research data management service in a domestic university library: focused on the analysis on the needs of researchers affiliated in Seoul national university. *Journal of the Korean Society for Information Management*, 36(3), 61-80. <http://doi.org/10.3743/KOSIM.2019.36.3.061>
- Shin, Y. R. & Chung, Y. K. (2012). A study on the improvement of the humanities and social sciences research data archives in Korea. *Journal of Korean Society of Archives and Records Management*, 12(3), 93-115.