

연구보고서 2020-30

전사적 연구데이터 관리 체계를 마련을 위한 연구

정영철
정소희·이기호·김은주·진재현·안수인

사람을
생각하는
사람들



KOREA INSTITUTE FOR HEALTH AND SOCIAL AFFAIRS



한국보건사회연구원
KOREA INSTITUTE FOR HEALTH AND SOCIAL AFFAIRS

【연구책임자】

정영철 한국보건사회연구원 연구위원

【공동연구진】

정소희 한국보건사회연구원 연구원

이기호 한국보건사회연구원 부연구위원

김은주 한국보건사회연구원 전문원

진재현 한국보건사회연구원 전문연구원

안수인 한국보건사회연구원 연구원

연구보고서 2020-30

전사적 연구데이터 관리 체계 마련을 위한 연구

발행일 2020년 12월

발행인 조흥식

발행처 한국보건사회연구원

주소 [30147] 세종특별자치시 시청대로 370
세종국책연구단지 사회정책동(1~5층)

전화 대표전화: 044)287-8000

홈페이지 <http://www.kihasa.re.kr>

등록 1994년 7월 1일(제8-142호)

인쇄처 (주)삼일기획

발|간|사

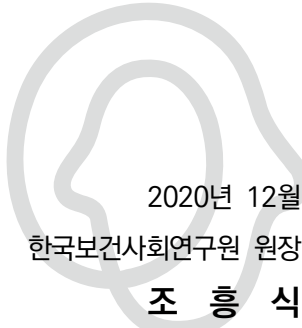
ICT 발달에 따라 엄청난 데이터와 정보를 수집, 처리, 저장할 수 있게 되면서 데이터 가치와 영향은 더욱더 커지고 있다. 정부출연연구기관에서의 객관적·과학적 증거를 활용한 증거 기반 정책(Evidence-based Policy) 연구에서도 연구의 투명성, 재현성, 투자 효율성을 높이기 위해 연구 성과와 과정을 개방·공유하는 연구데이터중심의 오픈 사이언스 정책(Open Science Policies)이 도입되어 추진되고 있다. 특히 경제인문사회계 연구기관에서는 과학기술계 연구기관과는 달리 아직 오픈 사이언스 정책(Open Science Policies)의 필요성과 중요성이 크게 부각되지 않은 상황에서 이번 연구는 한국보건사회연구원을 대상으로 연구데이터 관리 체계와 활용 방안을 마련함으로써 경제인문사회계 연구 부문에서의 오픈 사이언스 정책(Open Science Policies) 필요성을 강조한 점에 그 의의가 있다.

「공공데이터의 제공 및 이용 활성화에 관한 법률」에 의하면 공공기관인 정부출연연구기관이 생산한 전자적 방식의 연구데이터는 공공데이터로 간주할 수 있는 바, 연구데이터를 이용의 편리성과 보편성, 평등성 등을 보장하여 제공하여야 한다. 「공공기관의 정보공개에 관한 법률」에 의하면 정부출연연구기관에서는 연구데이터 공개를 위해 정보 관리 체계 정비에 노력하여야 하며 비공개 대상 정보에 대해 유의할 필요가 있다. 또한 「저작권법」에 의하면 정부출연연구기관의 연구데이터는 업무상 저작물에 해당되어 해당 기관이 저작자가 되며 「생명윤리 및 안전에 관한 법률」에 의하면 정부출연연구기관이 조사데이터를 민감 정보가 포함되지 않게 공개 시 기관생명윤리위원회의 심의를 면제할 수 있는 조건에 해당된다.

이번 연구에서는 이와 같은 법적 근거에 기반하여 정부출연연구기관 조사데이터의 특성을 규정하고, 당원 자체 규정인 「조사자료 관리지침」을 전부 개정하였으며 이를 위한 추진 조직으로 조사데이터TFT와 조사데이터 제공심의위원회를 제안하였다. 한편 조사데이터 관리 인프라로서 보건복지데이터포털은 추가 기능 개발뿐 아니라 연구성과관리시스템(RPMS)과 Research@kihasa 와의 정보시스템 간 정보 연계를 추진하였다.

이번 연구를 통해 연구데이터 재사용이 용이하고, 연구의 진실성을 확보하며, 중복 연구를 걸러주는 등 연구데이터의 가치 향상과 함께 연구 효율성을 제고하고 경제인문사회계 연구 부문에서의 오픈 사이언스 정책 필요성이 부각되기를 기대한다.

이번 보고서는 정영철 연구위원 책임 하에 정소희 연구원, 이기호 부연구위원, 김은주 전문원, 진재현 전문연구원, 안수인 연구원에 의해 작성되었다. 보건복지데이터포털 기능 개발을 위해 많은 고민을 함께 해주신 정보기술팀에게 감사드리며, 연구 수행에 있어 설문지를 통해 많은 협조와 유익한 의견을 주신 경제인문사회계 연구기관의 연구조정실장님들과 연구책임자들, 그리고 연구 내용에 대해 많은 조언을 주신 함영진 연구위원, 이용희 서울시립대 교수에게 감사의 말씀을 전한다. 마지막으로, 이 보고서의 내용은 우리 연구원의 공식적인 의견이 아님을 밝힌다.



2020년 12월
한국보건사회연구원 원장
조 흥 식



Abstract	1
요 약	3
제1장 서 론	17
제1절 연구 배경 및 목적	19
제2절 연구 내용 및 방법	24
제2장 관련 제도 및 정책 분석	27
제1절 오픈 데이터·오픈 사이언스 정책과 연구데이터 관리·활용 정책	29
제2절 우리나라 오픈 사이언스 관련 법·제도 현황	35
제3절 소결	48
제3장 연구데이터 관리 현황 및 문제점: 한국보건사회연구원 사례	51
제1절 관련 지침 분석	53
제2절 관련 사이트 분석	60
제3절 소결	67
제4장 연구데이터 관리 체계 및 인식·실태 조사	69
제1절 조사 개요	71
제2절 조사 결과	75
제3절 소결	121

제5장 연구데이터 관리 체계 개선 및 활용 활성화 방안 125

 제1절 연구데이터 관리 대상 및 절차 127

 제2절 관련 조직 및 위원회 구성 139

 제3절 연구데이터 제공 인프라 개편 방안 142

 제4절 연구데이터 활용 활성화 방안 151

참고문헌 157

표 목차

KOREA INSTITUTE FOR HEALTH AND SOCIAL AFFAIRS



〈표 1〉 연구데이터 관리 현황과 관리, 공유, 활용에 관한 인식 및 실태 조사 개요	5
〈표 2〉 관리 대상 조사데이터의 범위	12
〈표 3〉 한국보건사회연구원 조사데이터 관리를 위한 필요 기능	14
〈표 1-1〉 한국보건사회연구원 최근 3년간 연구에서의 조사 현황	22
〈표 1-2〉 연구데이터의 종류	24
〈표 2-1〉 '제공'과 '공개'의 차이 비교	40
〈표 2-2〉 공공데이터 관리 지침상 기계 판독이 가능한 형태의 단계별 구분·비교	41
〈표 2-3〉 연구데이터에 대한 FAIR 원칙	45
〈표 2-4〉 연구데이터 관리 절차	45
〈표 3-1〉 한국보건사회연구원 「조사자료 처리 및 관리 지침」	54
〈표 3-2〉 한국보건사회연구원 「조사자료 처리 및 관리 지침」 수정 필요 내용(일부)	55
〈표 3-3〉 한국보건사회연구원 조사자료 관리 매뉴얼 구성 내용	56
〈표 3-4〉 한국보건사회연구원 마이크로데이터 제공 업무 가이드라인 내용	58
〈표 3-5〉 마이크로데이터 제공 전 확인 사항	59
〈표 3-6〉 통계청 MDIS에서 제공하는 마이크로데이터 목록 현황	60
〈표 3-7〉 통계청 MDIS에서 제공하는 서비스	62
〈표 3-8〉 KOSSDA에서 제공하는 자료	63
〈표 3-9〉 KOSSDA에서 제공하는 이용자별 서비스	63
〈표 3-10〉 한국보건사회연구원 보건복지데이터포털에서 제공하는 마이크로데이터 목록 현황	64
〈표 4-1〉 연구데이터 관리 현황과 관리, 공유, 활용에 관한 인식 및 실태 조사 개요	72
〈표 4-2〉 연구기관 대상 연구데이터 관리 현황 조사 내용	73
〈표 4-3〉 연구데이터 관리, 공유, 활용에 관한 인식 및 실태 조사 내용	74
〈표 4-4〉 연구기관의 최근 3년간(2017~2019년) 수행 연구 중 조사 수행률	76
〈표 4-5〉 한국보건사회연구원 연구책임자 응답자 현황	85
〈표 4-6〉 원내 타 연구자의 조사데이터 활용 시 어려웠던 점	90
〈표 4-7〉 한국보건사회연구원 연구책임자들이 조사데이터 관리 체계에서 보완이 시급하다고 평가한 부분	95

〈표 4-8〉 한국보건사회연구원 조사데이터 관리 체계에 관한 자유 기재 의견	95
〈표 4-9〉 조사데이터 공유 및 개방 시 어려웠던 점	99
〈표 4-10〉 경제인문사회연구회 소속 기관 연구책임자 응답자 현황	101
〈표 4-11〉 경제인문사회연구회 소속 기관 연구책임자들의 조사데이터 개방을 위해 필요한 부분 평가	112
〈표 4-12〉 타 연구자의 조사데이터 활용 시 획득 경로-원외	113
〈표 4-13〉 타 연구자의 조사데이터 활용 시 어려웠던 점-원외	114
〈표 4-14〉 조사데이터 공유 및 관리를 위해 필요한 부분에 관한 기타 의견-원외	119
〈표 5-1〉 관리 대상 조사데이터의 범위	131
〈표 5-2〉 조사데이터 관리 대상 항목	133
〈표 5-3〉 조사데이터 관리 계획서 구성 내용	137
〈표 5-4〉 한국보건사회연구원 조사데이터 관리를 위한 필요 기능	140
〈표 5-5〉 한국보건사회연구원에서 제공 중인 마이크로데이터 활용 실적	152

그림 목차

KOREA INSTITUTE FOR HEALTH AND SOCIAL AFFAIRS



[그림 1] 한국보건사회연구원 연구책임자들의 오픈 사이언스에 대한 인식	7
[그림 2] 경제인문사회연구회 소속 기관 연구책임자들의 오픈 사이언스에 대한 인식	7
[그림 3] 한국보건사회연구원 연구책임자들의 타 연구 조사데이터 필요 또는 활용 경험 평가	8
[그림 4] 한국보건사회연구원 연구책임자들의 조사데이터 개방 필요성에 대한 인식	9
[그림 5] 경제인문사회연구회 소속 기관 연구책임자들의 조사데이터 개방 필요성에 대한 인식	9
[그림 6] 한국보건사회연구원 연구책임자들이 조사데이터 개방이 필요하다고 생각하는 이유 (복수 응답)	10
[그림 7] 경제인문사회연구회 소속 기관 연구책임자들이 조사데이터 개방이 필요하다고 생각하는 이유(복수 응답)	10
[그림 8] 조사데이터 관리 및 활용 처리 절차	13
[그림 9] 한국보건사회연구원의 조사데이터 관리 체계	15
[그림 10] 관련 정보시스템 간 통합 정보 연계 개념도	16
[그림 1-1] 연구 패러다임의 변화	19
[그림 2-1] 미국의 연구데이터 관리 체계	32
[그림 2-2] 영국의 연구데이터 관리 체계	33
[그림 2-3] 호주의 연구데이터 관리 체계	34
[그림 2-4] 공공데이터 제공·관리 단계	42
[그림 2-5] 저작권 및 기타 권리 이용허락 확인서	43
[그림 2-6] 연구데이터 관리 계획 포함 내용	47
[그림 3-1] 한국보건사회연구원 조사관리·지원 체계	57
[그림 3-2] 보건복지데이터포털 이용 안내	66
[그림 4-1] 연구기관의 조사데이터 관리 체계 및 관련 규정 등 마련 현황	77
[그림 4-2] 연구기관의 조사데이터 관련 정보시스템 구축 현황	78
[그림 4-3] 연구기관의 데이터 공개에 대한 인센티브 제공 현황	79
[그림 4-4] 연구기관의 내부 조사데이터 신청 및 제공 현황	81

[그림 4-5] 연구기관의 조사데이터 외부 공개 현황	82
[그림 4-6] 연구기관의 조사데이터 제공 형식(중복 응답)	83
[그림 4-7] 연구기관의 조사데이터와 함께 제공하는 관련 자료(중복 응답)	83
[그림 4-8] 한국보건사회연구원 연구책임자들의 오픈 사이언스에 대한 인식	86
[그림 4-9] 한국보건사회연구원 연구책임자들의 연구데이터 관리 계획에 대한 인식	87
[그림 4-10] 한국보건사회연구원 연구책임자들의 원내 타 연구자의 조사데이터 필요 또는 활용 경험 및 타 기관의 조사데이터 활용 경험 여부	88
[그림 4-11] 한국보건사회연구원 연구책임자들의 타 연구 조사데이터 필요 또는 활용 경험 평가	89
[그림 4-12] 한국보건사회연구원 연구책임자들의 조사데이터 관리자 실태와 인식	93
[그림 4-13] 한국보건사회연구원 연구책임자들의 당 원 조사데이터 관리 체계 운영 평가	94
[그림 4-14] 한국보건사회연구원 연구책임자들의 조사데이터 개방 필요성에 대한 인식	96
[그림 4-15] 한국보건사회연구원 연구책임자들이 조사데이터 개방이 필요하다고 생각하는 이유(복수)	98
[그림 4-16] 한국보건사회연구원 연구책임자들의 보건복지데이터포털 인지 및 활용 경험 여부	100
[그림 4-17] 경제인문사회연구회 소속 기관 연구책임자들의 오픈 사이언스에 대한 인식	102
[그림 4-18] 경제인문사회연구회 소속 기관 연구책임자들의 연구데이터 관리 계획에 대한 인식	104
[그림 4-19] 경제인문사회연구회 소속 기관 연구책임자들의 타 연구자 조사데이터 활용 경험 평가	105
[그림 4-20] 경제인문사회연구회 소속 기관 연구책임자들의 조사데이터 관리자 실태와 인식	106
[그림 4-21] 경제인문사회연구회 소속 기관 연구책임자들의 소속 기관 조사데이터 관리 체계 운영 평가	107
[그림 4-22] 경제인문사회연구회 소속 기관 연구책임자들의 조사데이터 개방 필요성에 대한 인식	108



[그림 4-23] 경제인문사회연구회 소속 기관 연구책임자들이 조사데이터 개방이 필요하지 않다고 생각하는 이유(복수 응답)	110
[그림 4-24] 경제인문사회연구회 소속 기관 연구책임자들이 조사데이터 개방이 필요하다고 생각하는 이유(복수 응답)	111
[그림 5-1] 발주처 발간확인서	129
[그림 5-2] 조사데이터 공개를 위한 확인 필요 내용	131
[그림 5-3] 조사데이터 관리를 위한 필요 항목 구성	132
[그림 5-4] 조사데이터 관리 및 활용 처리 절차	136
[그림 5-5] 한국보건사회연구원의 조사데이터 관리 체계	141
[그림 5-6] '보고서 발간/배포 요청서' 양식 설명	143
[그림 5-7] 연구관리시스템(RPMS)의 '보고서 발간/배포 요청서' 연구자용 화면(개선 후)	144
[그림 5-8] '조사데이터 이관 신청서' 양식	146
[그림 5-9] 연구관리시스템(RPMS)의 '조사데이터 이관 신청' 연구자용 화면(신설)	147
[그림 5-10] 관련 정보시스템 간 통합 정보 연계 개념도	148
[그림 5-11] 조사데이터 관리 단계별 기능 설명	149
[그림 5-12] 조사데이터 시스템 관리 화면	150



Abstract

A Study on the Establishment of Enterprise-wide Research Data Management System

Project Head: Chung, Youngchul

The core of “Open Science”, which seeks open and shared research findings and processes, is research data. Open science policies based on research data have been introduced to government-funded research institutes, where research is conducted to assist in evidence-based policy making. However, unlike the case is with science and technology research institutes, the necessity and importance of research data management and openness are little discussed in economic and humanities research institutes.

To enhance the value of research data and research efficiency and to highlight the need for open science policies in economic and humanities research institutes, this study attempted to establish a research data management system and a data use plan for the Korea Institute of Health and Social Affairs (KIHASA). To this end, we analyzed policies, laws, and institutions related to open science that are in place in Korea and other countries. Also, some of KIHASA’s guidelines and some online sites were examined to identify the status and problems of its research data management. We also conducted

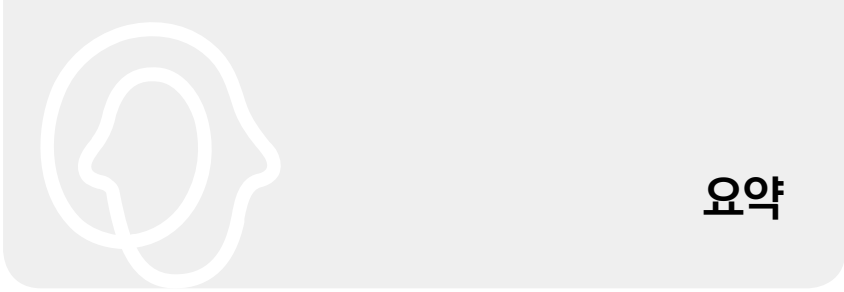
Co-Researchers: Jung, Sohee · Lee, kiho · Kim, Eunjoo · Jin, Jaehyun · An, Suin

2 전사적 연구데이터 관리 체계 마련을 위한 연구

online surveys to examine the status of survey data management at government-funded research institutes including KIHASA, and to understand perceptions of researchers toward the current status of research data management, data sharing, and data utilization.

To activate the utilization of research data, first, it is necessary to benchmark or link with related organizations (online sites) such as MDIS (Microdata Integrated Service) of Statistics Korea and Korea Social Science Data Archive (KOSSDA) of Seoul National University. Second, there is a need to introduce an incentive system and support measures for data transfer and sharing, such as granting research scores and supporting human resources and budgets at the Research Council level. Third, ways need to be sought to encourage the use of survey data, by for example, holding paper contests for survey data users and posting lists of research projects that make use of the data on the website. Fourth, data utilization can be encouraged by providing a list of survey data available internally or externally. Fifth, in the case of commissioned research projects, it is necessary to promote management and provision of data through active discussion or a collective agreement with the ordering organizations. Sixth, it is necessary to form a consensus among researchers through continuous education and promotion. Lastly, data should be translated into English to meet the needs of non-Korean-speaking clients.

* Key words: Open Science, Research Data Management System, Health and Welfare Data Portal



1. 연구의 배경 및 목적

ICT 발달에 따라 엄청난 데이터와 정보의 수집, 처리, 저장이 가능하게 되면서 사회 모든 영역에서 데이터의 가치와 영향이 더욱더 커지고 있다. 특히 연구 성과와 과정을 개방하고 공유하는 ‘오픈 사이언스(Open Science)’, ‘오픈 사이언스 정책(Open Science Policies)’의 핵심은 연구데이터로, 정부출연연구기관에서의 객관적·과학적 증거를 활용한 증거 기반 정책(Evidence-based Policy) 연구에서도 연구데이터를 중심으로 한 오픈 사이언스 정책이 도입되어 추진되고 있다. 그러나 과학기술계 연구기관과는 달리 경제인문사회계 연구기관에서는 연구데이터 관리와 개방의 필요성과 중요성이 공론화도 되지 않은 상황에서(정영철, 2020) 이번 연구는 한국보건사회연구원을 대상으로 연구데이터 관리 체계와 활용 방안을 마련하고자 하였다. 이를 통해 연구데이터의 재사용성, 연구의 진실성 확보, 중복 연구 방지 등과 같은 연구데이터의 가치 향상과 연구 효율성을 제고하며 경제인문사회계 연구 부문에서의 오픈 사이언스 정책 필요성을 부각하고자 한다. 한편 이번 연구에서의 연구데이터는 경제인문사회계 정부출연연구기관들이 가장 많이 생산하고 있는 설문조사 ‘조사데이터’를 중심으로 하였다.

2. 주요 연구 결과

오픈 사이언스 정책과 관련된 국내 관련 법으로 「공공데이터의 제공 및 이용 활성화에 관한 법률」, 「공공기관의 정보공개에 관한 법률」, 「저작권법」, 「생명윤리 및 안전에 관한 법률」에서 시사하는 바는 다음과 같다.

4 전사적 연구데이터 관리 체계 마련을 위한 연구

우선 「공공데이터의 제공 및 이용 활성화에 관한 법률」은 공공기관이 보유하고 관리하는 데이터 제공과 이용 활성화에 관한 사항을 규정한 법이다. 동 법 기준 하에서는 공공기관인 정부출연연구기관이 생산한 전자적 방식의 연구데이터는 공공데이터로 간주할 수 있는 바, 정부출연연구기관에서는 연구데이터를 이용의 편리성과 보편성, 평등성 등을 보장하여 제공하여야 하며 이용자에 대한 의무 사항도 적시할 필요가 있다. 「공공기관의 정보공개에 관한 법률」은 공공기관이 보유하고 관리하는 정보에 대해 국민의 공개 청구와 공공기관의 공개 의무에 관한 사항을 정한 법률로 동 법 기준에 의해 정부출연연구기관에서는 연구데이터 공개를 위해 연구데이터를 적절히 보존하고 보다 신속하고 편리하게 찾을 수 있는 정보 관리 체계 정비에 노력하여야 하며 연구데이터 제공 시 기준과 지침에 있어 특히 비공개 대상 정보에 대해 유의할 필요가 있다. 「저작권법」은 저작자의 권리를 보호하고 저작물의 공정한 이용을 도모하기 위한 법률로 동 법 기준에 의해 정부출연연구기관의 연구데이터의 경우, 직원 등이 업무상 작성하여 기관 등의 명의로 공표되었으므로 원칙적으로 업무상 저작물에 해당되어 특별한 정함이 없는 한 해당 기관이 저작자가 된다. 「생명윤리 및 안전에 관한 법률」은 인간의 존엄과 가치를 침해하거나 인체에 위해를 끼치는 것을 방지하여 생명윤리 및 안전을 확보하기 위한 법률로 동 법 기준에 의해 정부출연연구기관이 조사데이터를 민감 정보가 포함되지 않게 공개 시 기관생명윤리위원회의 심의를 면제할 수 있는 조건에 해당된다.

마이크로데이터 관리 및 제공 기능에 있어 통계청의 마이크로데이터 통합서비스(MDIS) 사이트, 서울대학교 한국사회과학자료원 사이트를 한국보건사회연구원의 보건복지데이터포털(data.kihasa.re.kr)과 비교하여 분석한 결과, 향후 보건복지데이터포털은 전사적 차원의 모든 조사데

이터를 수집하고 보관하기 위한 기능이 미흡하여 이에 대한 개선이 필요하며 질적, 양적으로 더 많은 마이크로데이터 제공과 더불어 다운로드 서비스뿐 아니라 다양한 서비스를 개발할 필요가 있다.

한국보건사회연구원의 연구데이터 관리 체계 마련을 위한 기초 자료로 활용하기 위해 한국보건사회연구원을 포함한 경제인문사회계열 정부출연연구기관들의 조사데이터 관리 현황과 연구책임자들의 연구데이터 관리, 공유, 활용에 관한 인식과 실태를 온라인 조사하였다. 조사 대상 및 조사 내용 등은 <표 1>과 같다.

<표 1> 연구데이터 관리 현황과 관리, 공유, 활용에 관한 인식 및 실태 조사 개요

구분	원내		원외	
	기관	개인	기관	개인
조사 내용	연구데이터 관리 현황	연구데이터 관리, 공유, 활용에 관한 인식 및 실태	연구데이터 관리 현황	연구데이터 관리, 공유, 활용에 관한 인식 및 실태
조사 대상 (응답자)	한국보건사회연구원	한국보건사회연구원 연구책임자(49명)	경제인문사회연구회 소속 연구기관 (19개 기관)	경제인문사회연구회 소속 연구기관 연구책임자(182명)
조사 기간	2020. 09. 10.부터 약 2주간		2020. 10. 14.부터 약 한 달간	2020. 10. 14.부터 약 3주간
조사 방법	담당자 면담	'서베이몽키'를 활용한 온라인 조사	이메일 조사	'서베이몽키'를 활용한 온라인 조사

자료: 저자 작성.

조사 결과, 한국보건사회연구원과 경제인문사회연구회 소속 연구기관들의 최근 3년간 조사 연구 수행률은 각각 39.6%, 37.7%를 나타내어 조사데이터의 중요성을 보여주고 있다. 조사데이터 관리 체계는 한국보건사회연구원과 응답 기관 19개 기관 중 9개 기관이 마련되어 있다고 응답하였으며 관련 지침 또는 매뉴얼은 한국보건사회연구원과 응답 기관 19개

6 전사적 연구데이터 관리 체계 마련을 위한 연구

기관 중 7개 기관이 갖추고 있었다.

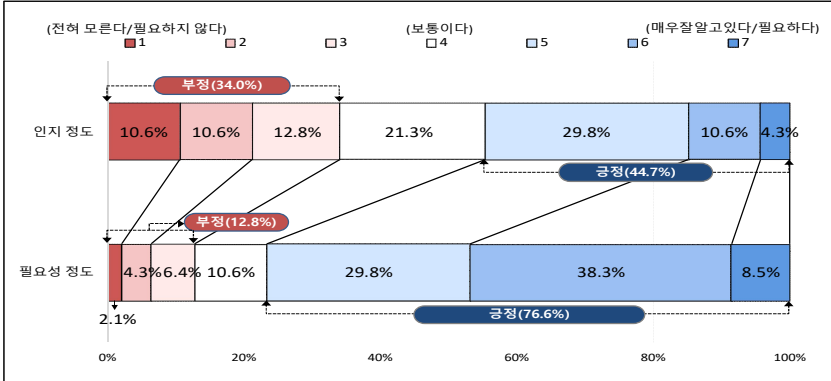
조사데이터 이관 혹은 외부 공개의 경우, 한국보건사회연구원에서는 연구자에게 인센티브를 제공하지 않고 있었으나 연구회 소속 연구기관 중 1개 기관에서는 제공하고 있었다. 조사데이터 목록을 내부에서 공유하고 있는 기관은 14개 기관이었으며 한국보건사회연구원에서는 별도로 제공하고 있지는 않았다. 조사데이터는 한국보건사회연구원과 응답 기관 19개 기관 중 15개 기관이 외부에 공개하고 있었다.

연구기관들이 제시한 조사데이터 제공 우수사례로는 외부 공개 데이터 목록과 활용 성과물을 홈페이지에 제공, 조사데이터 활용 확산을 위한 논문경진대회 추진 등이었다. 이는 향후 활용 활성화 방안에서 도입할 필요가 있다 하겠다.

한국보건사회연구원과 연구회 소속 연구기관들은 조사데이터 관리 체계에 있어 전담 인력, 추가 작업 지원 등에 어려움을 적시하고 있었으며 연구회 차원에서 이를 위한 인력 및 예산 지원 방안도 제시하고 있다.

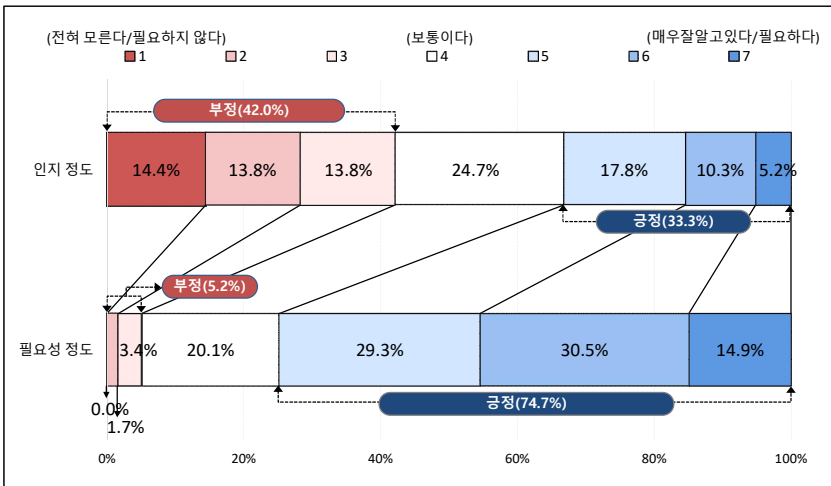
다음으로 한국보건사회연구원 연구책임자들과 연구회 소속 연구기관 연구책임자들의 응답을 비교·분석해 본 결과, 연구책임자들은 오픈 사이언스 개념과 동향, 연구데이터 관리 계획에 대해 현재 알고 있는 것보다 필요성, 중요성에 대해 더 많이 공감하고 있었다([그림 1], [그림 2] 참조).

[그림 1] 한국보건사회연구원 연구책임자들의 오픈 사이언스에 대한 인식



자료: 저자 작성.

[그림 2] 경제인문사회연구회 소속 기관 연구책임자들의 오픈 사이언스에 대한 인식



자료: 저자 작성.

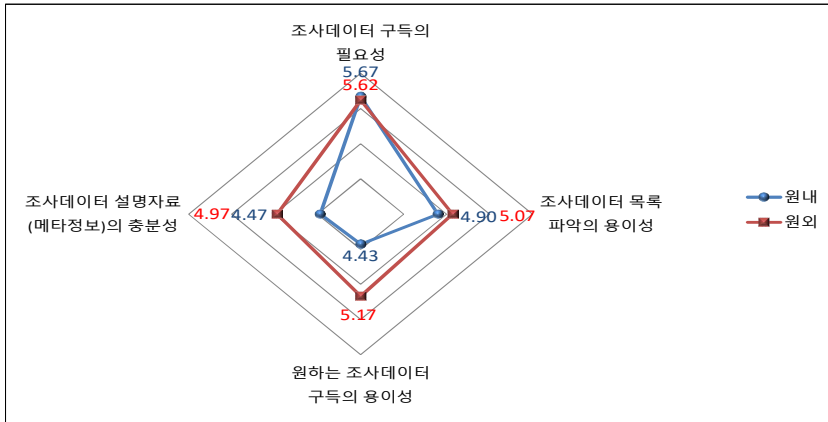
한국보건사회연구원 연구책임자들을 대상으로 한 조사에서 원내 혹은 타 기관 조사데이터 활용 경험자 중 구독의 필요성, 목록 파악의 용이성, 구독의 용이성, 메타정보의 충분성을 평가한 결과 상대적으로 구독의 용

8 전사적 연구데이터 관리 체계 마련을 위한 연구

이성과 메타정보의 충분성에서 원내 활용성이 떨어지는 것으로 나타났다 ([그림 3] 참조).

[그림 3] 한국보건사회연구원 연구책임자들의 타 연구 조사데이터 필요 또는 활용 경험 평가

(단위: 점)

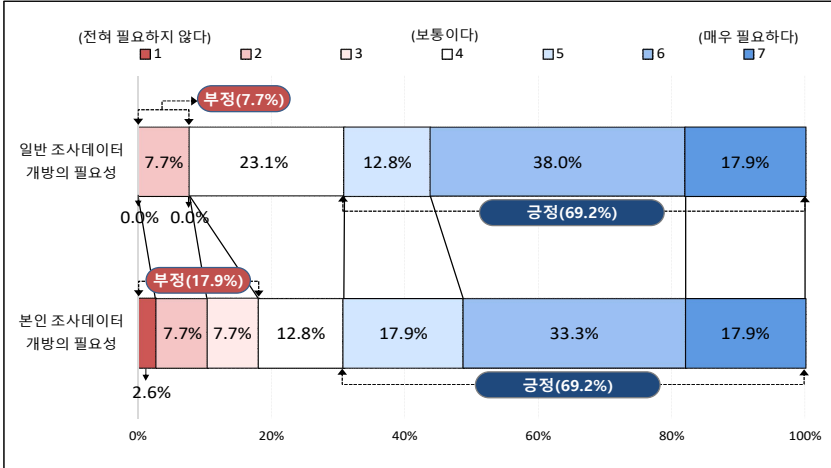


자료: 저자 작성.

최근 3년간 조사데이터를 생산한 경험은 한국보건사회연구원 연구책임자들은 55.8%, 연구회 소속 연구기관 연구책임자들은 64.2%를 나타냈다. 또한 연구 종료 후 조사데이터 보관 및 관리는 현재 연구자(책임자, 연구진)가 주로 하고 있었으나 한국보건사회연구원 연구책임자들은 65.2%, 연구회 소속 연구기관 연구책임자들은 57.1%가 기관 전담팀이 맡는 것이 적절하다고 제시하였다.

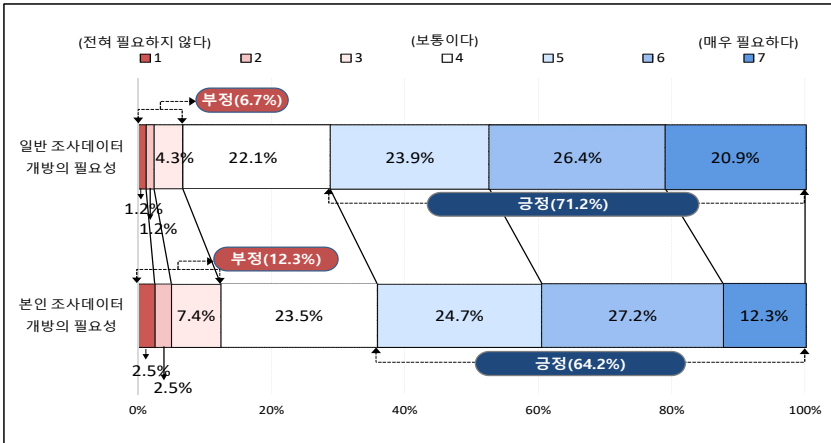
현재 소속 기관 조사데이터 관리 체계에 대해서는 모두 보통으로 생각하고 있었으며 조사데이터 개방에 대해서는 한국보건사회연구원 연구책임자와 연구회 소속 연구기관 연구책임자 모두 본인이 생산한 조사데이터 대비 일반적인 조사데이터 개방의 필요성이 더 많다고 제시하고 있다 ([그림 4], [그림 5] 참조).

[그림 4] 한국보건사회연구원 연구책임자들의 조사데이터 개방 필요성에 대한 인식



자료: 저자 작성.

[그림 5] 경제인문사회연구회 소속 기관 연구책임자들의 조사데이터 개방 필요성에 대한 인식



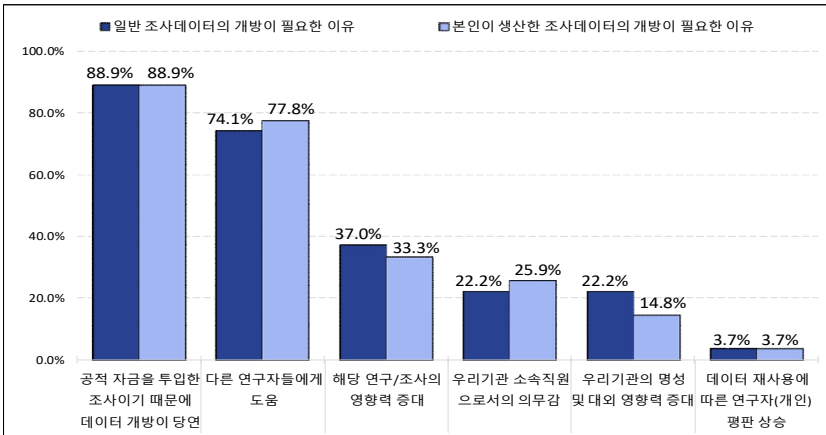
자료: 저자 작성.

한편 그림에도 조사데이터 공유와 개방은 다른 연구자에게 도움이 되기 때문에, 공적 자금을 투입한 조사이기 때문에 당연하다고 생각하고 있

10 전사적 연구데이터 관리 체계 마련을 위한 연구

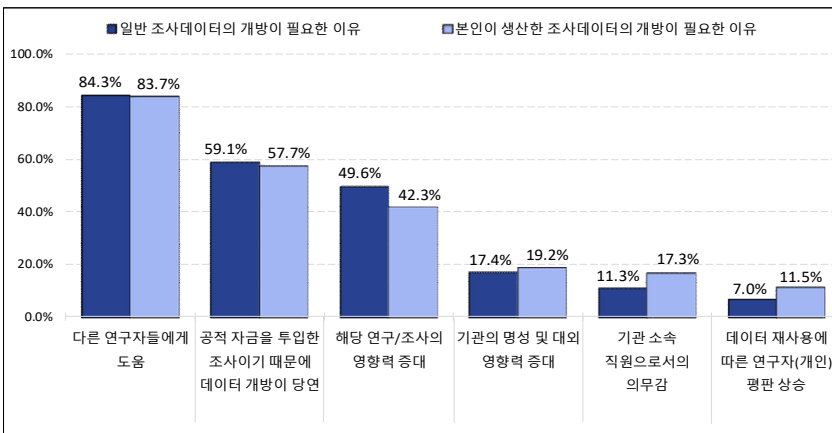
었으며 조사데이터 공유와 개방을 위해서는 인력 및 조직 체계 보완, 관련 인식 개선을 시급한 과제로 제시하였다(그림 6, [그림 7] 참조).

[그림 6] 한국보건사회연구원 연구책임자들이 조사데이터 개방이 필요하다고 생각하는 이유(복수 응답)



자료: 저자 작성.

[그림 7] 경제인문사회연구회 소속 기관 연구책임자들이 조사데이터 개방이 필요하다고 생각하는 이유(복수 응답)



자료: 저자 작성.

그 외 한국보건사회연구원 조사데이터 관리 체계 정립을 위하여 통계청 MDIS, 서울대학교 한국사회과학자료원 KOSSDA와의 연계 및 협력 체계 구축이 필요하며 관리 인프라인 보건복지데이터포털의 콘텐츠와 기능 보강, 그리고 적극적인 조사데이터 홍보 필요성이 제시되었다. 또한 데이터 공개, 활용에 대한 명확한 기준 제시가 필요하며, 전담 인력 및 조직의 필요성, 공감대 형성 및 인식 개선, 인센티브 제공, 연구 윤리에 관한 우려 등을 나타내었다.

3. 결론 및 시사점

조사데이터 관리 대상에 있어 자체 과제의 경우 조사데이터는 공공저작물이므로 보관과 동시에 제공 대상이 된다. 수탁 과제의 경우에는 발주 기관이 조사데이터 관리에 대한 특별한 요구가 없는 한, 보관하며 발주 기관에 저작권이 있으므로 제공은 발주처와의 합의가 필요하다. 조사는 내부에서 수행하든 외부에서 수행하든, 조사 규모가 모집단을 대상으로 하든 표본 집단을 대상으로 하든 조사데이터 보관(이관)과 제공에 영향을 끼치지 않는다. 다음으로 조사데이터 유형에 있어서는 양적 데이터는 보관(이관)과 제공 대상이 되며, 다만 질적 데이터의 경우에는 보관(이관)은 하되, 제공 대상에 대한 판단은 연구자가 판단·결정하는 것이 바람직하다. 또한 마이크로데이터와 이를 가공하여 공표한 공표데이터는 모두 이관 대상이며, 공표 이전의 마이크로데이터에 대한 제공 여부는 연구자 판단에 따르도록 한다(〈표 2〉 참조). 또한 이때 「정보공개법」 제9조 제1항에 따른 비공개 대상 정보와 「저작권법」 등에 따라 이용 허락을 받지 않은 정보를 비공개할 수 있다.

12 전사적 연구데이터 관리 체계 마련을 위한 연구

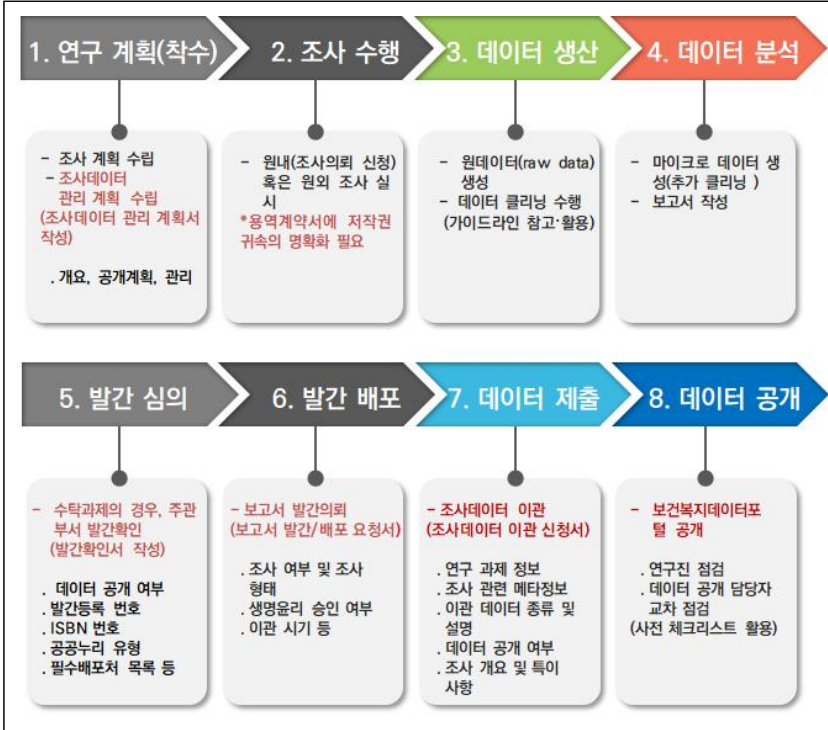
〈표 2〉 관리 대상 조사데이터의 범위

구분		보관	제공
생산 과제 종류	자체 과제	○	○
	수탁 과제	○(다만, 발주처 요청에 의해 이관불가일 수 있음)	△(발주처와 합의 필요)
조사 수행 주체	내부	○	○
	외부	○	○
조사 규모	모집단	○	○
	표본집단	○	○
조사데이터 유형	질적 데이터	○	△(연구자 결정)
	양적 데이터	○	○
	질적+양적	○	△(연구자 결정)
공표 여부	공표 이전 데이터	○	△(연구자 결정)
	공표데이터	○	○

자료: 저자 작성.

조사데이터 관리를 위해서는 연구 계획 단계에서 조사데이터 관리 계획을 수립하게 되며, 조사 수행 후 이를 통해 생산된 원자료(raw data)는 입력오류, 논리오류 등을 수정하는 작업을 거쳐 마이크로데이터를 생성하며 이를 이용하여 데이터를 분석한다. 이후 보고서 작성, 보고서 발간, 보고서 배포 시 배포 요청서상 조사 여부 및 조사 형태 등을 기록하며 보고서 발간 후 6개월 이내에 조사데이터를 이관하여 기관에서 수행한 모든 조사에 대한 데이터는 일괄 관리할 수 있게 한다. 이관 시 데이터 공개 여부 등을 결정짓게 되며 데이터 공개 담당자와 연구진의 점검을 거쳐 보건복지데이터 포털에서 데이터를 공개한다([그림 8] 참조).

[그림 8] 조사데이터 관리 및 활용 처리 절차



자료: 저자 작성.

또한 조사데이터 관리를 위해 필요한 기능으로는 조사데이터 관리 계획 검토 및 사후관리, 조사데이터 이관에 필요한 관련 자료 검토, 조사데이터 공개를 위한 처리, 조사데이터 관리 인프라 운영 및 관리, 조사데이터 활용 성과 측정, 조사데이터 관련 Q&A 응대 등이 있으며(〈표 3〉 참조), 이를 위해 초창기에는 TFT(Task Force Team)를 구성, 운영하여 점차 독립적인 조직으로 나아가갈 필요가 있다.

14 전사적 연구데이터 관리 체계 마련을 위한 연구

〈표 3〉 한국보건사회연구원 조사데이터 관리를 위한 필요 기능

필요 기능
- 조사데이터 관리 계획 검토 및 사후관리
- 조사데이터 이관 절차 담당
- 조사데이터 이관에 필요한 관련 자료 검토
- 조사데이터 품질 관리
- 조사 요도, 조사표, layout 등 관리
- 조사데이터 공개를 위한 처리(개인정보 처리, 저작권 검토, 공개 항목 검토, 설문지 및 layout 검토 등)
- 조사데이터 제공 심의위원회 운영 및 관리
- 조사데이터 관리 및 제공에 대한 법률적 자문 의뢰
- 조사데이터 관리 인프라 운영 및 관리
- 조사데이터 활용 전략 수립 및 실행 계획 추진
- 조사데이터 활용 성과 측정
- 조사데이터 연계 및 융합 방안 제시
- 조사데이터 관리 체계 정착을 위한 유인책 마련
- 조사데이터 관련 Q&A 응대

자료 : 저자 작성.

한편 비공개 조사데이터에 대해 원내외에서 공개 요청이 들어온 경우 공개 여부, 이용 승인에 관한 사항, 승인 없이 제3자 제공으로 인한 연구 윤리 위반 시 처리 등을 심의하고 처리하기 위해 ‘조사데이터 제공 심의 위원회’ 구성이 필요하며 초창기에는 별도 구성하기보다는 연구심의위원회의 심의 안건으로 포함하여 운영하는 것이 효율적일 수 있다.

이상과 같이 연구데이터 관리 대상 및 절차, 관련 조직 및 위원회 구성 등을 포함하는 당 연구원의 조사데이터 관리 체계를 도식화하면 [그림 9]와 같다.

[그림 9] 한국보건사회연구원의 조사데이터 관리 체계

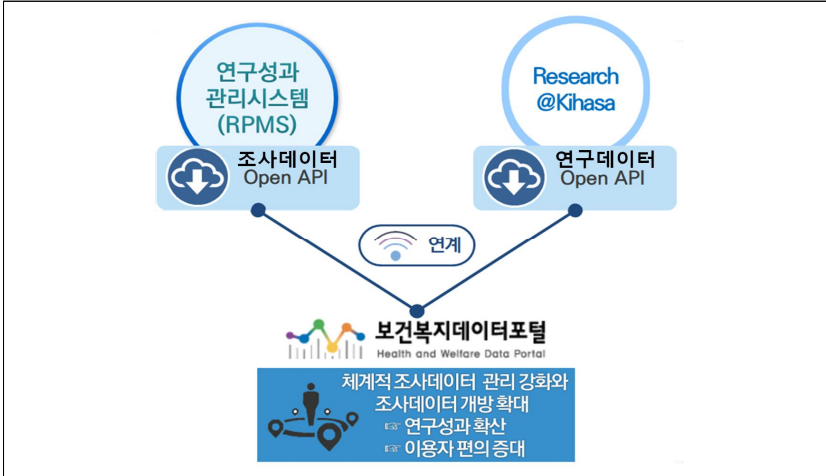


자료: 저자 작성.

한편 조사데이터 관리 인프라로서 보건복지데이터포털은 연구성과관리시스템(RPMS), 보건복지데이터포털, Research@kihasa 등 3개 관련 정보시스템 간 통합 정보 연계를 추진하였으며([그림 10] 참조), 현재 조사데이터 제공 기능 외에 조사데이터 관리시스템을 구축하였다.

16 전사적 연구데이터 관리 체계 마련을 위한 연구

[그림 10] 관련 정보시스템 간 통합 정보 연계 개념도



자료: 저자 작성.

연구데이터 활용을 활성화하기 위해서는 첫째, 통계청의 MDIS, 서울대학교 한국사회과학자료원의 KOSSDA 등과 같은 관련 기관(사이트)과의 벤치마킹 혹은 연계가 필요하며 둘째, 연구 평점 부여 및 기관 혹은 연구회 차원의 인력 및 예산 지원 등과 같은 조사데이터 이관 및 공개에 따른 인센티브 제도 도입과 지원 방안 마련이 필요하다. 셋째, 조사데이터 활용 논문 경진대회 개최 혹은 조사데이터 활용 성과를 홈페이지에서 제공하는 등 조사데이터 활용을 장려하기 위한 방안 도입이 필요하다. 넷째, 내부 혹은 외부에서 사용 가능한 조사데이터 목록 제공을 통한 활용 장려가 필요하다. 다섯째, 수탁 과제의 경우 발주처와의 적극적 협의 혹은 일괄 협약 등을 통해 관리와 제공을 추진해 나갈 필요가 있다. 여섯째, 지속적인 교육 및 홍보를 통해 연구자들의 공감대를 형성할 필요가 있으며 마지막으로 해외 수요에 맞추어 데이터 영문화 작업도 필요하다.

* 주요 용어: 오픈 사이언스, 조사데이터 관리 체계, 보건복지데이터포털

사람을
생각하는
사람들



KOREA INSTITUTE FOR HEALTH AND SOCIAL AFFAIRS



제 1 장

서론

제1절 연구 배경 및 목적

제2절 연구 내용 및 방법

제 1 장 서론


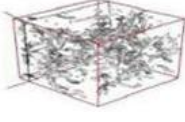

제1장에서는 이번 연구를 수행하게 된 배경과 필요성, 연구 목적, 연구 대상, 연구 내용 및 방법 등을 기술하였다.

제1절 연구 배경 및 목적

정보통신과 IT 분야의 기술 발달로 인해 엄청난 데이터와 정보를 수집, 처리, 저장할 수 있으며 이른바 빅데이터 시대라 일컫게 되었다. 이와 같은 데이터는 사회 모든 영역에 많은 영향을 끼치고 빅데이터 처리 기술을 통해 인간과 사회에 대해 현재를 진단하기도, 미래를 예측하기도 하는 등 그 가치를 더해가고 있다.

연구 영역도 기존의 이론이나 실험 등을 통합한 데이터를 중심으로 하는 제4세대 패러다임으로 변화하면서(Hey, Tansley, & Tolle, 2009, 국가과학기술연구회, 2019b 재인용) 데이터의 중요성이 더욱 커지고 있다(〈그림 1-1〉 참조).

[그림 1-1] 연구 패러다임의 변화

1세대 연구	2세대 연구	3세대 연구	4세대 연구
실험 중심연구	이론 중심연구	컴퓨팅자원 중심연구	데이터 중심연구
	$\left(\frac{a}{a}\right)^2 = \frac{4\pi G\rho}{3} - K\frac{c^2}{a^2}$		

자료: Hey, T., Tansley, S., & Tolle, K. (2009). Jim Gray on eScience: A Transformed Scientific Method. The Fourth Paradigm: Data-Intensive Scientific Discovery. Microsoft Research(pp. xvii-xxxi). Newyork: Microsoft Research; 국가과학기술연구회. (2019b). 출연(연) 연구데이터 관리·활용 방안 연구. p. 1에서 재인용.

과학계에서는 기존 지식에 대한 접근, 학습, 검증, 확산을 새로운 지식과 혁신 창출의 근간으로 보는 오픈 사이언스(Open Science) 개념이 뿌리 깊은 연구 규범으로 자리잡고 있는 가운데(Merton, 1942; 신은정, 안형준, 정원교, 2016 재인용) 디지털 기술이 발전하고 확산되면서 연구를 기획·수행하고 그 성과를 확산·활용하는 방식이 빠르게 변화하고 있다(신은정 외, 2017). 오픈 사이언스(Open Science)는 과학계에서 연구 자료를 공개하고 검증하는 개방적 연구 규범을 지칭하는 개념으로 도입된 이후 인터넷이 보급되고 디지털 연구 자원이 축적되면서 2000년대 초반부터 OECD를 중심으로 이론적 논의를 넘어 글로벌 정책 어젠다로 논의되기 시작하였다(신은정, 2015).

연구의 투명성, 재현성, 투자 효율성을 높이기 위해 연구 성과와 과정을 개방·공유하는 ‘오픈 사이언스(Open Science)’, ‘오픈 사이언스 정책(Open Science Policies)’의 핵심은 연구데이터이다.

이러한 데이터 활용을 통해 정책 수립 영역에 있어서도 과거의 통념·관행에 의존했던 의견 기반 정책(Opinion-based Policy) 수립에서 정책의 효율성과 합리성을 제고시키기 위해 객관적·과학적 증거를 활용한 증거 기반 정책(Evidence-based Policy) 수립으로 점차 변화하고 있다(한국정보화진흥원, 2015). 증거 기반 정책은 정책 결정 과정에서 관련 통계, 연구 및 평가 결과 등과 같은 근거를 고려하여 결정하도록 하는 것으로서 과정이 투명하고 내용을 신뢰할 수 있으며 설득력을 얻을 수 있다(오세영 외, 2017).

증거 기반 사상은 1990년대 초반 영국 의료 분야에서 처음 사용된 이후 임상심리학, 사회복지학을 비롯한 여러 학문 분야에서 폭넓게 사용되었으며 '90년대 후반 이후 사회복지, 교육, 형사 사법 등 다양한 분야에서 정책 수립과 실천 영역에 확대되었다(Sowaki, 2009; 한국정보화진흥원,

2015 재인용).

한편 정부는 모든 영역에서의 공공 문제를 해결하기 위해 ‘정책’을 계획·실행·평가하고 있으며 이를 지원받기 위해 정부출연연구기관을 설립하였다. 정부출연연구기관은 정부가 재원을 출연하여 설립한 싱크탱크로, 경제인문사회계 기관과 과학기술계 기관으로 나눌 수 있다¹⁾.

경제인문사회계 기관에서는 정책 연구를 통해 정책을 개발하는 것에 우선권을 두고 있으며 과학기술계 기관에서는 연구·개발 자체에 우선권을 두고 있다(김용훈, 오영균, 2008). 공공 영역에서의 연구는 국가, 사회 문제가 복잡하고 다양해질수록 역할이 확대되고 있다(이민형, 김태양, 2020).

이러한 연구에서 주안점은 증거에 기반하고 있다는 것이다. 정부출연연구기관에서 정책을 개발하기 위해 수행하는 정책 연구는 증거 기반, 근거에 기반하여 정책을 수립하고 분석·평가하는 것이며 순수 연구는 과학의 이론과 원리에 의해 증거에 기반한 새로운 지식을 탐구하는 것으로 이에 필수 요소는 데이터이다.

이와 같은 데이터의 가치 향상을 위해 세계 각국에서는 오픈 사이언스 정책을 도입하여 추진하고 있는 가운데 우리나라 연구 영역에서도 과학기술계 정부출연연구기관 연구 개발 사업에 오픈 사이언스 정책이 도입되었으나 상대적으로 경제인문사회계 정부출연연구기관에서는 연구데이터 관리와 개방의 필요성, 중요성도 공론화되지 않은 상황이다(정영철, 2020).

한국보건사회연구원은 경제인문사회계 대표적 연구기관으로 2008년부터 한경비즈니스에서 시작한 ‘대한민국 100대 싱크탱크’ 조사 중 정치·사회 부문에서 2018년까지 7년 연속 1위, 그리고 2020년 다시 1위에

1) 우리나라 정부출연연구기관은 「정부출연연구기관 등의 설립·운영 및 육성에 관한 법률」 제8조 제1항에 따른 24개 기관과 「과학기술 분야 정부출연연구기관 등의 설립·운영 및 육성에 관한 법률」 제8조 제1항에 따른 22개 기관 등 총 46개 기관이 있음.

22 전사적 연구데이터 관리 체계 마련을 위한 연구

선정되는(한국보건사회연구원, 2020. 3. 31.) 등 국가적 차원의 사회보장, 사회정책을 개발하고 평가하는 싱크탱크이다. 이러한 한국보건사회연구원에서는 다양한 연구데이터가 생산되고 있으며 그 중에서도 사회조사 결과 생산되는 조사데이터는 대표적인 연구데이터 중 하나로 최근 3년간 수행한 연구 과제에서 조사를 수행한 연구가 평균 약 40%를 차지하였다. 이는 근거 기반 연구를 수행함에 있어 많은 연구에서 조사 사업을 실시하고 있으며, 이에 따라 조사데이터가 중요한 연구 산출물임을 잘 나타내고 있다(〈표 1-1〉 참조).

〈표 1-1〉 한국보건사회연구원 최근 3년간 연구에서의 조사 현황

(단위: %)

구분	조사 사업 포함 연구	조사 방법		
		양적 조사	질적 조사	양적+질적
평균	39.6	65.9	21.1	13.0
2017년	38.8	74.4	16.7	9.0
2018년	40.8	63.4	22.0	14.6
2019년	39.1	60.5	24.4	15.1

자료: 정영철, (2020). 근거기반 정책 지원을 위한 연구데이터 관리와 개방. 보건복지포럼, 286. p. 17. 〈표 5〉를 직접 인용.

한편 명실상부한 우리나라 사회정책 연구의 중심 기관인 한국보건사회연구원에서는 이러한 조사데이터에 대한 관리의 필요성을 일찌감치 수용하여 2012년에 「조사자료 처리 및 관리 지침」을 제정하였으며 최근에 이르기까지 4번의 개정 작업이 이루어졌다. 그럼에도 불구하고 실질적인 내용에 있어서는 표준화된 관리 내용과 관리 절차, 총괄적 관리 주체, 통합적 관리 매체 및 일원화된 제공 체계가 미흡하고 빈번하게 변화하는 현 상황을 제대로 반영하고 있지 못하는 등 연구데이터의 수집·관리·공유·활용에 있어 총체적인 문제점을 나타내고 있다. 더욱이 연구를 수행하면

서 생산된 연구데이터를 연구자 개인 소유로 인식하여 개별적으로 활용(제3자 제공 혹은 목적 외 사용)하거나 관리 시 사장·분실 혹은 기관 이직 시 인수인계가 제대로 되지 않기도 하며²⁾ 일반 과제와 같이 다년도에 걸쳐 수행되고 있는 대형 과제들이 늘어나고 있는 현 상황에서 전국적, 시·계열적 연구데이터 관리의 필요성과 중요성은 더욱더 증가하고 있다. 또한 과거 데이터 중에서도 역사적 가치를 지니고 있는 조사데이터에 대한 수요도 꾸준히 생기고 있어 체계적이면서 지속적인 관리가 이루어져야 할 것이다.

그러므로 이번 연구는 경제인문사회계 대표 기관인 한국보건사회연구원을 대상으로 연구데이터 관리 현황과 문제점을 파악하여 연구데이터의 가치 창출을 위해, 또 연구데이터 관리에 대한 정부 정책에 부합하기 위해 연구데이터의 수집, 관리, 공유, 활용을 위한 체계적이고 통합적인 연구데이터 관리 체계와 활용 방안을 마련하고자 한다. 이를 통해 연구데이터 재사용이 용이하고, 연구의 진실성을 확보하며, 중복 연구를 걸러주는 등 연구데이터의 가치 향상과 함께 연구 효율성을 제고하며 경제인문사회계 연구 부문에서의 오픈 사이언스 정책 필요성이 부각되기를 기대한다.

2) 연구원 내 퇴직 서류 중 퇴직자 업무 정리 사항에 조사데이터에 대한 내용이 적시되어 있지 않음.

제2절 연구 내용 및 방법

「국가연구개발사업의 관리 등에 관한 규정」에 의하면 연구데이터는 “연구개발과제 수행 과정에서 실시하는 각종 실험, 관찰, 조사 및 분석 등을 통하여 산출된 사실 자료로서 연구 결과의 검증에 필수적인 데이터” (제2조 제18항)로 정의하고 있으며 예비 분석 결과, 논문이나 저술의 초안, 연구노트, 보고서 등은 포함하지 않고 있다(국가과학기술연구회, 2019a). 이러한 연구데이터 종류에는 실험데이터, 관측데이터, 시물레이션데이터, 파생데이터, 참조데이터, 조사데이터 등이 있다(최명석, 김현우, 이상환, 2019, <표 1-2> 참조).

<표 1-2> 연구데이터의 종류

구분	내용
실험데이터	실험 장비에서 생산되는 데이터 - 가속기, 화학/바이오 실험데이터
관측데이터	관측 장비를 통해 생산되는 데이터 - 망원경, 전자현미경, 인공위성 등
시물레이션데이터	모델링을 통해 생산되는 데이터 - 기후 모델링, 경제 전망 모델링
파생데이터	원천 데이터로부터 재생산된 데이터 - 텍스트마이닝, 3D 모델링
참조데이터	평가를 거쳐 신뢰성이 공인된 데이터 - 플라즈마 물성표준, 뇌MRI 영상 참조 표준 데이터
조사데이터	설문조사 등을 통해 생산된 데이터 - 시장 조사, 예측 조사

자료: 최명석 외. (2019). 국가 연구데이터 공유의 시작, 데이터 관리 계획, KISTI ISSUE BRIEF 13, p. 4.

이번 연구는 연구데이터 중에서도 경제인문사회계 정부출연연구기관들이 가장 많이 생산하고 있는 설문조사 ‘조사데이터’를 대상으로³⁾, ‘한국보건사회연구원’의 관리 체계를 실 사례로 살펴보았다.

이를 위해 관련 기존 문헌, 관련 법령 및 지침 등을 분석하였으며 한국보건사회연구원을 포함하여 경제인문사회계 연구기관의 연구데이터 관리 현황과 각 소속 연구책임자들에 대한 연구데이터 관리·공유·활용에 대한 인식 및 실태를 파악하기 위해 온라인 설문조사를 실시하였다. 또한 한국보건사회연구원의 관련 지침, 관리 절차, 관련 서식을 분석하여 문제점을 제시하였으며 이러한 결과들을 종합하여 오픈 사이언스 사상에 맞는 연구데이터 관리 체계를 마련하고 활성화 및 활용 방안을 제시하였다.

이번 보고서는 총 5개 장으로 구성하였다.

제1장 ‘서론’에서는 연구 패러다임의 변화, 연구데이터에서의 ‘오픈 사이언스’ 사상, 증거 기반 정책 연구의 필요성, 한국보건사회연구원에서의 연구데이터 관리·활용에서의 문제점, 그리고 이러한 배경 하에서의 연구 목적을 제시하였다.

제2장 ‘관련 제도 및 정책 분석’에서는 오픈 데이터와 오픈 사이언스 정책에 대해 개략적으로 살펴보고 해외 연구데이터 관리 및 활용 정책에 대해 살펴보았다. 또 우리나라에 있어 오픈 사이언스 관련 법제도로 행정안전부의 「공공데이터의 제공 및 이용 활성화에 관한 법률(이하 공공데이터법)」, 행정안전부의 「공공기관의 정보공개에 관한 법률(이하 정보공개법)」, 문화체육관광부의 「저작권법」, 행정안전부의 「공공데이터 관리 지침」, 과학기술정보통신부의 「국가연구개발사업의 관리 등에 관한 규정」, 국가과학기술연구회의 「연구데이터 관리 가이드라인」과 「연구데이터 관리 계획(DMP) 가이드라인」 등을 살펴보았다.

3) 이후 기관의 ‘연구데이터’는 ‘조사데이터’와 동일한 의미로 사용됨.

제3장 ‘연구데이터 관리 현황 및 문제점: 한국보건사회연구원 사례’에서는 한국보건사회연구원의 관련 지침으로 「조사자료 처리 및 관리 지침」과 조사자료 관리 매뉴얼, 마이크로데이터 제공 업무 가이드라인을 분석하고 문제점을 제시하였으며 관련 사이트로는 통계청의 MDIS, 서울대학교 한국사회과학자료원의 KOSSDA와 한국보건사회연구원의 보건복지데이터포털(data.kihasa.re.kr)을 비교·분석하고 문제점 및 시사점을 제시하였다.

제4장 ‘연구데이터 관리 체계 및 인식·실태 조사’에서는 온라인 조사를 통해 한국보건사회연구원과 경제인문사회계열 정부출연연구기관들의 조사데이터 관리 체계 현황과 연구책임자들의 연구데이터 관리, 공유, 활용에 대한 인식 등을 파악하였다.

제5장 ‘연구데이터 관리 체계 개선 및 활용 활성화 방안’에서는 연구데이터 관리 절차와 범위, 관련 조직 및 위원회 구성, 연구데이터 제공 인프라 개편 방안, 연구데이터 활용 활성화 방안을 제시하였다.



제2장

관련 제도 및 정책 분석

- 제1절 오픈 데이터·오픈 사이언스 정책과 연구데이터
관리·활용 정책
- 제2절 우리나라 오픈 사이언스 관련 법·제도 현황
- 제3절 소결

제2장 관련 제도 및 정책 분석

제2장에서는 전사적 차원의 연구데이터 관리 체계를 실증적으로 제안하기 위해 오픈 데이터, 오픈 사이언스, 해외 연구데이터 관리·활용 정책에 대해 기존 문헌을 참고하여 개략적으로 기술하였으며 우리나라의 관련 법·제도들을 살펴보았다.

제1절 오픈 데이터·오픈 사이언스 정책과 연구데이터 관리·활용 정책

1. 오픈 데이터와 오픈 사이언스 정책

학문의 자유를 위하여, 공적 자금 투자에 대한 사회경제적 편익을 제고하기 위하여 2000년 전후 학계를 중심으로 세계 각국에서는 디지털 기술 확산에 따라 연구 성과와 과정을 개방하여 새로운 가치를 창출하고자 하는 오픈 사이언스(Open Science) 운동이 확산되고, 이에 대한 정책이 추진되고 있으며 민간 부문보다는 공공 정책 영역으로 자리잡아가고 있다(신은정 외, 2017). 오픈 사이언스(Open Science)는 초창기에 “과학계의 보편적·집단적 지식 생산 메커니즘의 기저를 이루는 자율적 연구 규범(Merton, 1942; 신은정 외, 2016 재인용)”을 일컫는 개념으로 시작하였으나 정보기술 발달에 따라 인터넷 보급률이 향상되고 디지털 연구 자원 축적이 급증하자 2000년대 초반부터 OECD를 중심으로 글로벌 정책 어젠다로 등장하게 되었다(신은정, 2015).

오픈 사이언스(Open Science)는 “공공 연구의 성과물(출판물 및 데이터)을 디지털 형태로 공개 및 확산시켜 사회경제적 편익을 제고하려는 일련의 노력(OECD, 2015; 생명공학정책연구센터, 2018 재인용)”으로 정의하고 있으며, 오픈 사이언스 정책(Open Science Policies)은 이를 위한 공적 개입을 의미하고 있다. 이러한 오픈 사이언스는 디지털 자료의 보관, 처리, 전송이 용이해짐에 따라 온라인 환경에서의 출판으로 이어지는 오픈 액세스(Open Access), 연구 과정 중에 산출된 데이터 및 소재를 공개하고 공유하는 오픈 데이터(Open Data) 정책 등을 포괄하며(OECD, 2015; 신은정 외, 2016 재인용), 온라인상에서 연구 개발 과정이 이루어지는 오픈 사이언스 랩(Open Science Lab) 등도 등장하고 있다(신은정 외, 2016)

유럽연합, 미국, 영국, 일본 등 각국에서는 공공정책 차원의 다양한 오픈 사이언스 정책이 추진되고 있으며 우리나라에서도 과학기술계에서 추진하고 있다.

오픈 사이언스에서는 자연 현상에 대한 보편적 진리를 탐구하는 보편주의(Universalism), 연구자로서 중립적 자세에서 진리를 추구하는 탈이해주의(Disinterestedness), 연구자 개인이나 조직 차원이 아닌 과학계 전체가 하나의 커뮤니티로서 임하게 하는 공동체주의(Communism), 자의적 해석이나 방법론이 아닌 논리적이고 과학적인 방법론과 절차에 입각하여 기존 지식에 대해 지속적으로 검증해 나가는 조직적 회의주의(Organized skepticism) 등을 규범으로 내세우고 있다(Merton, 1942; 신은정 외, 2016 재인용).

2. 연구데이터 관리·활용 정책⁴⁾

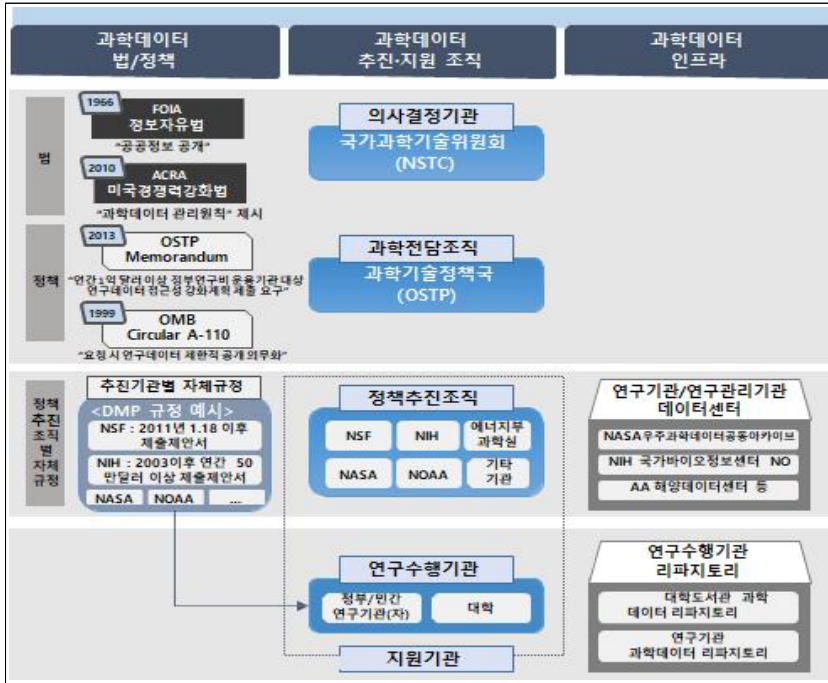
미국은 연방 기관 중심으로 국가 차원에서 연구데이터 공유·활용 정책을 시행하고 인프라를 구축·운영하고 있으며 백악관의 과학기술정책국(OSTP: The Office of Science and Technology Policy)에서 이를 위한 원칙을 제시하고 있다. 법·제도적으로는 정보자유법(FOIA: Freedom of Information Act), 미국 경쟁력 강화법(ACRA: America COMPETES Reauthorization Act)을 통해 법적 근거를 마련하였으며 관련 정책은 2013년 대통령 메모(memorandum)를 통해 지시되었다. 미국항공우주국(NASA: National Aeronautics and Space Administration), 미국 국립보건원(NIH: National Institutes of Health), 국립해양대기국(NOAA: National Oceanic and Atmospheric Administration)과 같은 연구 관리 기관의 자체 규정에 의해 연구 지원을 받는 대학이나 연구기관은 연구 성과를 공개하고 있다.

이러한 미국의 연구데이터 관리 체계를 도식화하면 [그림 2-1]과 같다.

4) 최명석(2017)의 내용을 참고하여 작성하였음.

32 전사적 연구데이터 관리 체계 마련을 위한 연구

[그림 2-1] 미국의 연구데이터 관리 체계



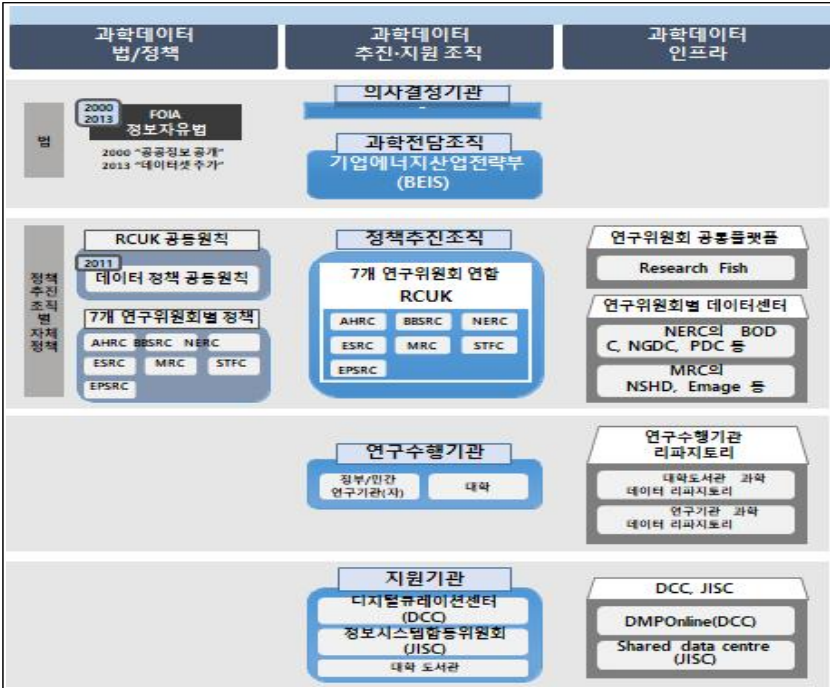
자료: 최명석.(2017). 오픈 연구데이터 정책 동향. ie매거진 24(4). p. 32.

영국은 연구회(Council) 중심으로 국가 차원에서 연구데이터 공유·활용 정책을 시행하고 인프라를 구축 및 운영하고 있으며 영국연구위원회(RCUK: Research Councils UK)가 필요한 데이터 정책에 관한 공동 원칙(Common Principles on Data Policy)을 제시하고 있다. 법·제도적으로는 정보자유법(FOIA: Freedom of Information Act)을 통해 법적 근거를 마련하였으며 관련 정책은 영국연구위원회(RCUK)가 데이터 정책 공동 원칙(Common Principles on Data Policy)과 7개 연구회)별 각각의 정책 규정에 의해 공개 및 공유하고 있다. 이러한 영국의

5) 예술 및 인문학연구회(AHRC: The Arts and Humanities Research Council), 생명공학

연구데이터 관리 체계를 도식화하면 [그림 2-2]와 같다.

[그림 2-2] 영국의 연구데이터 관리 체계



자료: 최명석.(2017). 오픈 연구데이터 정책 동향. ie매거진 24(4). p. 33.

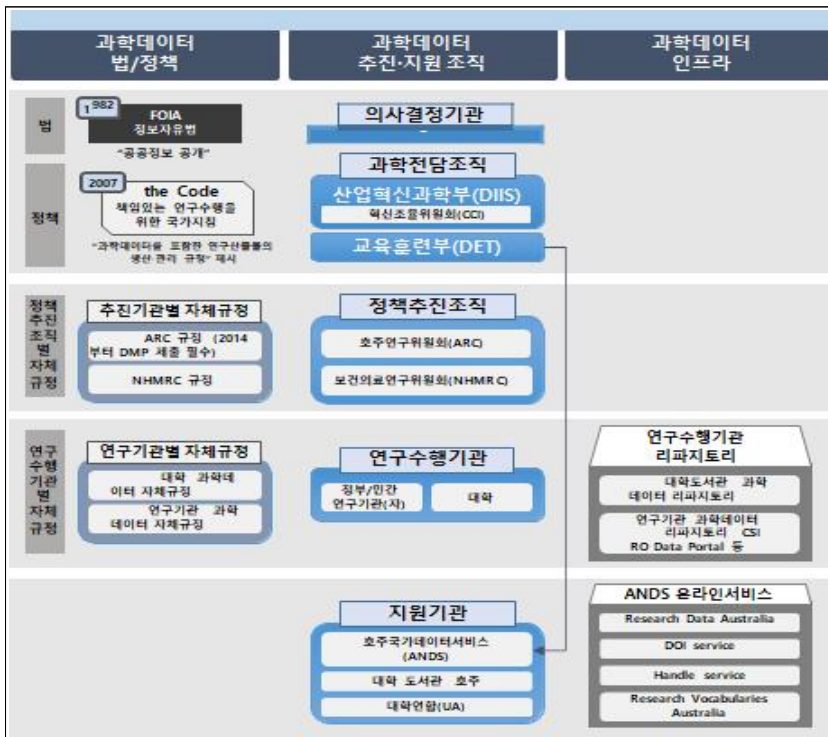
호주 역시 연구회(Council) 중심으로 국가 차원에서 연구데이터 공유·활용 정책을 시행하고 인프라를 구축 및 운영하고 있으며 정보자유법(FOIA: Freedom of Information Act)을 통해 법적 근거를 마련하였

및 생물과학연구회(BBSRC: Biotechnology and Biological Sciences Research Council), 자연환경연구회(NERC: The Natural Environment Research Council), 경제사회연구회(ESRC: The Economic and Social Research Council), 정부의학연구회(MRC: Government Medical Research Council), 과학기술연구회(STEC: The Science and Technology Facilities Council), 공학 및 물리학 연구회(EPSRC: The Engineering and Physical Sciences Research Council).

34 전사적 연구데이터 관리 체계 마련을 위한 연구

다. 연구데이터 등 연구 산출물 생산 및 관리는 ‘책임 있는 연구 수행을 위한 국가 지침(Australian Code for the Responsible Conduct of Research, the Code)’을 통해 제시하고 있다. 호주연구위원회(ARC: Australian Research Council), 보건의료연구위원회(NHMRC: The National Health and Medical Research Council) 과 연구 수행 기관에서는 ‘the Code’에 근거하여 자체 규정을 수립하여 연구 성과를 공개하고 있다. 이러한 호주의 연구데이터 관리 체계를 도식화하면 [그림 2-3]과 같다.

[그림 2-3] 호주의 연구데이터 관리 체계



자료: 최명석.(2017). 오픈 연구데이터 정책 동향. ie매거진 24(4). p. 33.

제2절 우리나라 오픈 사이언스 관련 법·제도 현황

이번 절에서는 우리나라에서의 오픈 사이언스 정책(Open Science Policies) 이행과 관련된 법률 및 관련 규정들을 살펴보았다.

1. 행정안전부의 「공공데이터의 제공 및 이용 활성화에 관한 법률 (이하 공공데이터법)」

「공공데이터법」은 2013년 6월 ‘정부3.0 기본 계획’에 공공데이터 개방이 천명되면서 2013년 7월 30일 제정된 법으로, 공공기관이 보유하고 관리하는 데이터 제공과 이용 활성화에 관한 사항을 규정하여 국민의 이용권을 보장하고 민간 활용 촉진을 추구하고 있다(제1조 목적).

동 법에서 ‘공공데이터’라 함은 데이터베이스, 전자파일과 같이 법령 등에서 정한 목적을 위해 공공기관이 생성 또는 취득하여 관리하고 있는 전자적 방식의 자료 또는 정보이며, ‘제공’이란 전자적 방식의 자료를 공공기관이 이용자에게 전달하는 것을 의미한다(제2조 정의). 이에 의하면 공공기관인 정부출연연구기관이 생산한 전자적 방식의 연구데이터도 공공데이터로 간주할 수 있다.

동 법 기본원칙(제3조)에서는 공공기관에 대해 이용의 편리성과 보편성, 평등성, 영리성 등을 보장토록 하고 있으며 이용저해행위가 불가함을 규정하고 있다. 또 이용자에게는 공익이나 타인의 권리를 침해하지 않도록 의무를 준수하고, 신의에 의한 성실한 이용을 요구하고 있다.

한편 제공하지 않아도 되는 정보로는 「공공기관의 정보공개에 관한 법률」 제9조에 따른 비공개 대상 정보와 「저작권법」 등에 의해 이용 허락을 받지 않은 정보로 규정하고 있으며(제17조 제공 대상 공공데이터의 범위)

제18조(공공데이터 목록의 등록), 제19조(공공데이터 목록 정보의 공표)에서는 공공데이터 목록을 국민 누구나 쉽게 이용할 수 있도록 등록하고 공표토록 하고 있다.

또한 공공데이터 제공 및 이용을 활성화하고 효율적 관리를 위해 공공데이터 제공 형태 및 제공 기술, 분류 체계 등에 대해 표준을 제정하고 시행하여야 함을 규정하고 있으며(제23조 공공데이터의 표준화) 이에 따라 2020년 8월 11일 공통표준용어 338개, 공통표준단어 195개, 공통표준도메인 13개로 「공공데이터 공통표준용어(행정안전부고시 제2020-42호)」가 제정, 시행되었다.

동법 제24조(공공데이터의 제공 기반 구축)에서는 이용자가 편리하게 이용할 수 있게 하기 위한 공공기관과 정부의 역할을 제시하고 있다.

그러므로 공공기관인 정부출연연구기관에서는 「공공데이터법」 입법 취지에 맞게 공공데이터의 일종인 연구데이터를 이용의 편리성과 보편성, 평등성 등을 보장하여 제공하여야 하며 이용자에 대한 의무 사항도 적시할 필요가 있다.

2. 행정안전부의 「공공기관의 정보공개에 관한 법률(이하 정보공개법)」

「정보공개법」은 공공기관이 보유하고 관리하는 정보에 대해 국민의 공개 청구와 공공기관의 공개 의무에 관한 사항을 정한 법률로(제1조 목적), 여기에서 ‘공개’라 함은 공공기관이 정보를 열람하게 하거나 그 사본·복제물을 제공하는 것 또는 정보통신망을 통해 정보를 제공하는 것을 의미하며(제2조 정의), 제3조(정보공개의 원칙)에서는 공공기관이 보유하고 관리하는 정보는 적극적으로 공개하여야 함을 천명하고 있다.

동 법에서는 모든 국민이 정보공개 청구권을 가짐을 명시하고 있고(제

5조 정보공개 청구권자), 공공기관의 의무로서, 정보를 적절히 보존하고 신속한 검색이 가능하도록 정보관리 체계 정비에 노력할 것 등을 규정하고 있다(제6조 공공기관의 의무). 한편 제9조(비공개 대상 정보)에서는 공공기관이 공개하지 않아도 되는 정보로 1) 타 법령 등에서 비밀이나 비공개 사항으로 규정된 정보, 2) 공개 시 국가의 이익을 해칠 우려가 있는 정보, 3) 공개 시 국민의 생명, 신체 및 재산 보호에 지장을 끼칠 우려가 있는 정보, 4) 공개 시 직무 수행이 곤란해지거나 공정한 재판에 영향을 끼칠 우려가 있는 정보, 5) 공개 시 공정한 업무 수행이나 연구·개발에 지장을 초래하는 정보, 6) 공개 시 사생활 비밀 또는 자유를 침해할 우려가 있는 정보, 7) 공개 시 법인·단체 또는 개인의 경영상·영업상 비밀에 관한 정보, 8) 공개 시 특정인에게 이익 또는 불이익을 줄 우려가 있는 정보 등을 제시하고 있으며 공공기관에서는 비공개 대상 정보 범위에 관한 세부 기준을 수립하여야 함을 규정하고 있다.

그러므로 공공기관인 정부출연연구기관에서는 「정보공개법」 입법 취지에 맞게 연구데이터의 공개를 위해 연구데이터를 적절히 보존하고 보다 신속하고 편리하게 찾을 수 있는 정보관리 체계 정비에 노력하여야 하며 연구데이터 제공 시 기준과 지침에 있어 특히 비공개 대상 정보에 대해 유의할 필요가 있다.

3. 문화체육관광부의 「저작권법」

「저작권법」은 저작자의 권리를 보호하고 저작물의 공정한 이용을 도모하기 위한 법률로(제1조 목적), ‘저작물’은 인간의 운동 또는 감정을 표현한 창작물, ‘저작자’는 저작물을 창작한 자, ‘공동저작물’은 2인 이상이 공동으로 창작한 저작물로서 각자의 이바지한 부분을 분리하여 이용할

수 없는 것을 말하며, '업무상저작물'은 법인·단체 그 밖의 사용자의 기획 하에 그 업무에 종사하는 자가 업무상 작성하는 저작물로 정의하고 있다(제2조 정의).

동 법에서는 법인 등의 명의로 공표되는 업무상저작물의 저작자는 계약 또는 근무 규칙 등에 다른 정함이 없는 때에는 그 법인 등이 된다고 규정하고 있으며(제9조 업무상저작물의 저작자), 저작자는 그의 저작물을 공표 혹은 공표하지 않는 것을 결정할 권리를 가진다(제11조 공표권).

국가 또는 지방자치단체가 업무상 작성하여 공표한 저작물이나 계약에 따라 저작재산권의 전부를 보유한 저작물은 몇몇 경우를 제외하고 허락 없이 이용 가능성을 규정하고 있다(제24조의2 공공저작물의 자유 이용).

정부출연연구기관의 연구데이터의 경우, 직원 등이 업무상 작성하여 기관 등의 명의로 공표되었으므로 원칙적으로 업무상 저작물에 해당되어 특별한 정함이 없는 한 해당 기관이 저작자가 되므로 연구 수행 후 연구데이터에 대한 저작 권한은 기관이 가지게 된다. 다만 공동 연구 수행에 있어서의 조사 시 혹은 외부 조사업체에 의한 조사 수행 시에는 저작권 귀속 등에 대해 명확히 할 필요가 있다.

4. 보건복지부의 「생명윤리 및 안전에 관한 법률(이하 생명윤리법)」

「생명윤리법」은 인간과 인체 유래물 등을 연구하거나 배아나 유전자 등을 취급할 때 인간의 존엄과 가치를 침해하거나 인체에 위해를 끼치는 것을 방지하여 생명윤리 및 안전을 확보하기 위한 법률이다(제1조 목적). 동 법에서는 생명윤리 및 안전을 확보하기 위해 인간대상연구를 수행하는 기관에 '기관생명윤리위원회'를 설치하여 연구 계획서의 윤리적·과학적 타당성, 연구 대상자로부터의 적절한 절차에 의한 동의 여부, 연구 대

상자 등의 개인정보 보호 대책 등을 심의토록 하고 있으며(제10조 기관생명윤리위원회의 설치 및 기능), 인간대상연구에서는 기관위원회의 심의를 받도록 규정하고 있다(제15조 인간대상연구의 심의). 그럼에도 불구하고 공개된 정보를 이용하는 연구에서 연구대상자가 특정되지 않고 기존 자료나 문서를 이용하는 연구의 경우에는 기관위원회의 심의를 면제할 수 있다(동법 시행규칙 제13조 기관위원회의 심의를 면제할 수 있는 인간대상연구).

정부출연연구기관이 조사 사업을 수행하는 경우 인간대상연구가 대부분으로 기관생명윤리위원회의 심의를 받고 있다. 이후 **조사데이터가 생산되고 이를 공개 시에는 민감 정보가 포함되어 있지 않아 이용자가 이를 이용 시 기관생명윤리위원회의 심의를 면제할 수 있는** 인간대상연구에 속하게 된다.

5. 행정안전부의 「공공데이터 관리지침(행정안전부고시 제2019-71호)」

동 지침은 「공공데이터의 제공 및 이용 활성화에 관한 법률」 제11조(공공데이터 관리지침)에 근거하여 제정되었으며 공공기관이 보유한 공공데이터의 관리 및 제공의 효율적 시행을 위해 공공기관이 준수해야 할 관리 원칙과 기준을 제시하고 있다(제정 목적). 이는 1) 개요, 2) 공공데이터 제공 관리 단계별 기준, 3) 공공데이터 개방 표준, 4) 공공데이터 품질 관리 단계별 기준 및 5) 공공기관의 중복·유사 서비스 금지에 관한 내용으로 구성되어 있다.

동 지침에서는 ‘제공’과 ‘공개’ 용어에 대해 <표 2-1>과 같이 소관 법률, 목적, 범주, 제공 형태, 제공 창구 등을 비교하여 제시하고 있다.

40 전사적 연구데이터 관리 체계 마련을 위한 연구

〈표 2-1〉 ‘제공’과 ‘공개’의 차이 비교

구분	제공	공개
① 소관 법률	「공공데이터의 제공 및 이용 활성화에 관한 법률」	「공공기관의 정보공개에 관한 법률」
② 목적	민간 활용을 통한 신규 비즈니스와 일자리 창출, 국민편익 향상	국민의 알권리 보장과 국정운영의 투명성 확보
③ 범주	전자적 자료 또는 정보	전자적·비전자적 자료 또는 정보
④ 제공 형태	반드시 기계 판독이 가능한(machine-readable) 형태	특정한 공개 형태가 없음
⑤ 제공 창구	공공데이터포털 data.go.kr	정보공개포털 open.go.kr
대상 예시	실시간 버스운행(노선, 위치, 도착 시간 등) 데이터 등	공문서, 버스운행계획보고, 버스회사 보조금 지급 계획 및 결과 보고 문서 등

자료: 공공데이터 관리지침, 행정안전부고시 제2019-71호. (2019. 9. 4. 시행), p. 8.

또 공공데이터 제공 시 기계 판독이 가능한 형태의 HWP, XLS, JPG, MP3 등과 같이 특정 소프트웨어에서 읽고 수정 및 변환이 가능한 최소 충족 요건을 제시하고 있다(〈표 2-2〉 참조). **그러므로 정부출연연구기관에서 연구데이터 제공시에는 최소한 최소충족포맷인 2단계 이상의 파일 형태로 제공하는 것이 필요하다.**

〈표 2-2〉 공공데이터 관리 지침상 기계 판독이 가능한 형태의 단계별 구분·비교

구분	1단계	2단계	3단계	4단계	5단계
기계 판독이 가능한 형태	미충족 포맷 (공공데이터 포털등록 불가)	최소충족포맷	오픈포맷*		
특징	특정 소프트웨어에서 읽을 수만 있는 데이터로 자유로운 수정, 변환 불가	특정 소프트웨어에서 읽고 수정, 변환 가능	모든 소프트웨어에서 읽고 수정, 변환 가능	URI**를 기반으로 데이터 속성 특성 관계를 기술하고 있는 데이터 구조	웹상의 다른 데이터와 연결, 공유 가능
예시	PDF	HWP, XLS, JPG, PNG, WMV, MPEG, MP3, SWF	CSV, JSON, XML	RDF	LOD

자료: 공공데이터 관리지침, 행정안전부고시 제2019-71호. (2019. 9. 4. 시행). pp. 5-6.

한편 동 지침에서는 공공데이터 제공 관리 단계를 생성·수집, 처리·운영, 등록·관리, 제공, 사후관리 등 5단계로 구분하여 각 단계별로 준수해야 할 관리 기준 및 원칙을 제시하고 있다(그림 2-4) 참조). 이 중에서 특히 저작권이 일부라도 외부인(기관)에 있는 경우에는 저작권자에게 ‘저작물 이용허락 동의서’를 받아야 함을 명시하고 있다.

이를 통해 정부출연연구기관에서도 연구데이터 제공시 특히 제1단계, 제2단계에 대한 준비가 필요하며 특히 외부 위탁을 통해 연구데이터를 생성·구축하는 경우, 위탁 업체와 계약 시 ‘저작권 및 기타 권리 이용허락 확인서(그림 2-5) 참조’를 받아야 한다.

42 전사적 연구데이터 관리 체계 마련을 위한 연구

[그림 2-4] 공공데이터 제공·관리 단계



자료: 공공데이터 관리지침, 행정안전부고시 제2019-71호. (2019. 9. 4. 시행). p. 12의 내용을 도식화하였음.

[그림 2-5] 저작권 및 기타 권리 이용허락 확인서

**○○○○사업(특정사업/용역명)을 통해 발생한
저작권 및 기타 권리 이용허락 확인서**

① ○○○○기관(특정사업/용역수행기관명)은 ○○○○사업(특정사업/용역명)을 수행하는 과정에서 발생한 최종 결과물을 공공데이터로 제공하여 자유롭게 이용(상업적 이용, 변경 등)하는 것에 동의하며, 상기 권리에 대한 세부 내역을 본 사업 종료 시까지 제출하겠습니다.

② 또한 제1항 소정의 최종 결과물에 제3자의 저작물이나 초상 등이 이용되거나 포함되어 있는 경우, 이를 공공데이터로 활용(상업적 이용, 변경 등)하는 데 제약이 없도록 이용허락 등 필요한 조치를 취하고, 상기 권리에 대한 세부 내역을 본 사업 종료 시까지 제출하겠습니다.

○○○○년 ○○월 ○○일

권리자 : (인)

주소 :

법인등록번호 :

연락처 :

※ 공동 수급체 구성원 각각(하도급이 있을 경우 하도급 포함)은 직접 혹은 별도의 고용·용역을 통해 발생한 최종 결과물에 이용되거나 포함된 제3자의 저작권 및 기타 권리에 대해 공공데이터 개방을 위한 이용허락을 확보해야 한다.

자료: 공공데이터 관리지침, 행정안전부고시 제2019-71호. (2019. 9. 4. 시행). p. 74.

6. 과학기술정보통신부의 「국가연구개발사업의 관리 등에 관한 규정」

「국가연구개발사업의 관리 등에 관한 규정」은 과학기술 분야 국가연구개발사업 추진 시의 기본 원칙과 기준을 마련한 규정으로 2019년 개정안(대통령령 제29625호, 2019. 3. 19. 일부 개정)에서 오픈 사이언스 개념 도입을 위한 기반을 마련하였다.

이에 “연구데이터란 연구개발 과제 수행 과정에서 실시하는 각종 실험, 관찰, 조사 및 분석 등을 통하여 산출된 사실 자료로서 연구 결과의 검증에 필수적인 데이터”로, “데이터 관리 계획이란 연구데이터의 생산·보존·관리 및 공동활용 등에 관한 계획”으로 규정하고 있다(제2조 정의).

이러한 국가연구개발사업을 수행 혹은 참여하기 위해서는 연구개발 계획서에 데이터 관리 계획이 포함되어야 하며(제6조 제4항 제9호) 중앙행정기관에서는 데이터 관리 계획에 따른 연구데이터 생산·보존·관리의 충실성 및 공동활용 가능성을 검토하도록 규정하고 있다(제7조 제3항 제10호).

7. 국가과학기술연구회의 「연구데이터 관리 가이드라인」과 「연구데이터 관리 계획(DMP) 가이드라인」

「연구데이터 관리 가이드라인」은 과학기술계 정부출연연구기관들이 국가연구개발사업 수행시 연구데이터 수집, 관리, 공유, 활용에 대한 기준을 제시한 것이다.

동 가이드라인에서는 연구데이터 관리에 있어 검색 가능성, 접근 가능성, 상호운용 가능성, 재사용 가능성의 FAIR(Findable, Accessible, Interoperable, Reusable) 원칙을 요구하고 있다(〈표 2-3〉 참조).

〈표 2-3〉 연구데이터에 대한 FAIR 원칙

구분	내용
Findable (검색성)	연구데이터에 메타데이터와 DOI 등 영구식별자를 할당하고 연구데이터 리포지터리에 등록하여 검색 가능하도록 함.
Accessible (접근성)	표준화된 프로토콜을 사용하여 식별자를 통해 연구데이터(연구데이터가 이용 가능하지 않을 경우 메타데이터)에 접근 가능하도록 함.
Interoperable (상호운용성)	연구데이터는 표준적인 방식과 어휘를 사용하여 상호운용 가능하도록 해야 하며, 관련된 다른 데이터에 대한 인용을 포함함.
Reusable (재사용성)	명확한 데이터 라이선스와 프로비던스 정보를 제공하여 재사용 가능하도록 함.

자료: 정영철. (2020). 근거기반 정책 지원을 위한 연구데이터 관리와 개방. 보건복지포럼, 286. p. 14. 〈표 3〉을 직접 인용.

출처: Wilkinson et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data, 3(1). 1-9. doi: 10.1038/sdata.2016. 18; 국가과학기술연구회. (2019a). 연구데이터 관리 가이드라인. p. 2에서 재인용한 내용을 표로 재구성함.

또한 ① 연구데이터 관리 체계 수립 ② 연구데이터 수집·저장 ③ 연구데이터 출판 ④ 연구데이터 공유·활용 등 4단계의 연구데이터 관리 절차를 제시하고 있다(〈표 2-4〉 참조).

〈표 2-4〉 연구데이터 관리 절차

구분	내용	
관리 체계 수립	정책과 절차	- 관리 기준 정의, 정책 및 절차 수립, 기관 내 공표
	전담 조직 및 인력	- 전문 인력 충원 및 전담 조직 구성
	요구사항 정의	- 기관 내 연구데이터의 크기, 종류, 생산주기, 검색 요구사항 정의 - 연구데이터 및 프로세스의 표준화 방안 마련
	연구 데이터 인프라	- 안정적인 연구데이터 관리와 장기간 보존을 제공하고 다양한 연구데이터 공유·활용을 지원하는 데 필수적인 IT 인프라 필요
	지원 서비스	- 체계적인 연구데이터 관리·활용을 위해 연구자에게 교육, 자문, 지원 서비스 등 제공

46 전사적 연구데이터 관리 체계 마련을 위한 연구

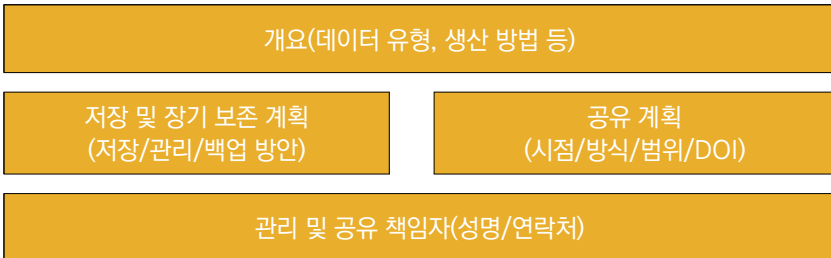
구분		내용
수집 및 저장	연구데이터 정제 및 표준화	- 관리·보존을 위해 연구기관의 데이터 리포지터리에 제출하며, 이를 위해 연구기관에서 정의한 연구데이터 관리 기준에 적합하게 정제되고 표준화된 연구데이터 준비
	메타데이터 작성	연구기관에서 정의한 메타데이터 스키마를 준용하며, 장기간 보존 및 활용을 위해 풍부하게 기술
	제출	연구과제 종료 후 연구기관에서 지정한 기간 이내에 데이터 관리 계획에서 지정한 데이터 리포지터리에 연구데이터 제출
출판	승인 및 검토	- 연구기관의 연구데이터 공개 기준에 따라 비공개, 내부 공개, 대외 공개를 선택하고, 공개의 경우 엠바고 설정을 통해 공개 시한 지정 <ul style="list-style-type: none"> • 연구데이터 주요 검토 사항 : 데이터 형식, 메타데이터 기술 형식 준용 여부, 연구데이터 공개 및 라이선스 적합 여부, 민감 정보 포함 여부 등 - 승인 절차와 성과물 관리를 위해 연구기관의 정보관리 시스템과 연계 가능
	출판 절차	- 다른 연구자들이 (재)활용할 수 있도록 데이터를 공개하기 위해 연구기관 등의 데이터 리포지터리를 통해 온라인으로 접근 가능한 형태로 출판 가능 - 데이터 저널을 통해 연구데이터 자체에 대해 논문의 형태로 출판 가능 - 관련 주제 분야 혹은 데이터 전문가의 피어 리뷰 등 별도의 품질 점검 절차 도입 가능
공유 및 활용	데이터 식별자 부여	- 데이터셋의 식별 및 활용 추적을 위해서 공개된 연구데이터에 대해 DOI(Digital Object Identifier) 등과 같은 식별자 부여 권장 - DOI는 한국 DOI 센터, DataCite 등 DOI 등록 기관으로부터 발급 받을 수 있으며, 해당 연구 분야의 식별자 관리 기관이 존재하면 해당 기관 활용
	라이선스 적용	- 연구기관은 연구데이터 대·내외 공개와 저작권 확보를 위해 라이선스 기준과 지원 서비스를 제공하고, 연구자가 각 연구데이터에 대해 적절한 라이선스를 선택하여 적용
공유 및 활용	접근 및 공유	- 다른 연구자가 연구데이터를 쉽게 활용할 수 있도록 표준 프로토콜 및 기술 형식에 따라 연구데이터 관리 및 배포 - 연구데이터 리포지터리의 검색에 노출되는 연구데이터 및 메타데이터의 공개 여부는 이를 제출하는 연구자에 의해 지정 - 해당 연구 분야의 데이터 검색 사이트 또는 외부 연구데이터 사이트와의 연계를 통해 검색 노출 빈도수를 높임으로써 연구데이터 활용도 증진
	활용	- 연구기관에서는 내외부 연구자에게 보유 중인 연구데이터의 처리·분석을 위한 다양한 연구 환경 제공 - 타 연구자의 연구데이터를 활용한 경우 연구 논문 등에서 데이터 인용을 통해 출처 표기

자료: 정영철. (2020). 근거기반 정책 지원을 위한 연구데이터 관리와 개방. 보건복지포럼, 286, p. 14. <표 4>를 직접 인용.

출처: Wilkinson et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data, 3(1). 1-9. doi: 10.1038/sdata.2016. 18; 국가과학기술연구회. (2019a). 연구데이터 관리 가이드라인. pp. 3-6에서 재인용한 내용을 표로 재구성함.

한편 「연구데이터 관리 계획(DMP) 가이드라인」은 과학기술계 국가연구개발사업 시 생산되는 연구데이터에 대해 관리 계획(DMP: Data Management Plan)을 작성하고 점검하는 기준을 제시한 것이다. 연구데이터 관리 계획에는 개요, 저장 및 장기 보존 계획, 공유 계획, 관리 및 공유 책임자 등을 포함하며([그림 2-6] 참조) “과기정통부에서는 2019년 시범 R&D 과제를 대상으로 데이터 관리 계획 제출 의무화를 우선 시행하고, 2020년부터는 단계적으로 적용 대상을 확대할 예정(한국과학기술정보연구원 기술사업화 플랫폼 홈페이지, n.d.)”으로 머지않아 정부 예산으로 수행하는 모든 R&D 과제에 도입될 것이다.

[그림 2-6] 연구데이터 관리 계획 포함 내용



자료: 국가과학기술연구회. (2019c). 연구데이터 관리 계획(DMP) 가이드라인. p. 2.

제3절 소결

연구 영역, 공공정책 영역에서 공적 자금 투자에 대한 사회경제적 편익 추구를 위한 오픈 사이언스, 오픈 데이터에 대한 전세계적인 흐름은 IT 기술 발달에 따라 더욱더 거세지고 있다.

공공기관인 정부출연연구기관이 생산하고 있는 연구데이터는 「공공데이터법」에 의한 일종의 공공데이터로 간주할 수 있으므로 이를 제공 시에는 이용의 편리성, 보편성, 평등성 등에 대한 보장을 위해 노력하여야 하며, 이용자에게도 공익이나 타인의 권리를 침해하지 않도록 신의에 따라 성실히 이용토록 의무 사항을 적시할 필요가 있다.

또한 「정보공개법」에 의해 연구데이터 공개를 위해 적절히 보존하고 보다 신속하고 편리하게 찾을 수 있도록 정보관리 체계 정비에 노력하여야 하며, 비공개 대상 정보에 대한 명확한 기준과 지침 마련이 필요하다.

한편 정부출연연구기관의 연구데이터는 업무상 저작물로 간주할 수 있어 해당 기관이 저작자가 되며, 공동연구 수행에 있어서 조사 수행 시, 혹은 외부 조사업체에 의한 조사 수행 시 저작권 귀속 등에 대해 명확히 할 필요가 있다.

정부출연연구기관의 조사 사업은 대부분 인간대상연구로 기관생명윤리위원회의 심의를 받게 된다. 이후 조사데이터 공개 시에는 민감 정보가 제외되어 기관생명윤리위원회 심의를 면제할 수 있다.

우리나라 연구 영역에서의 오픈 사이언스 정책은 과학기술계 정부출연연구기관 연구개발사업에 도입되어 연구계획서 작성 시 ‘데이터 관리 계획’을 포함토록 하고 있으며 ① 관리 체계 수립(정책과 절차, 전담 조직 및 인력, 요구 사항 정의, 연구데이터 인프라, 지원 서비스)-② 수집 및 저장(연구데이터 정제 및 표준화, 메타데이터 작성, 제출, 승인 및 검토)-③ 출

판(출판 절차, 데이터 식별자 부여, 라이선스 적용)-④ 공유 및 활용(접근 및 공유, 활용)의 연구데이터 관리 절차를 제시하고 있다.

그러나 경제인문사회계 정부출연연구기관에서는 아직 이러한 가이드 라인을 적용하지 않고 있으며 심지어 연구데이터 관리와 개방의 필요성, 중요성도 공론화되지 않고, 현황조차 파악이 안 되어(정영철, 2020) 있다는 점에서 이번 연구의 의의를 찾을 수 있다 하겠다.



제3장

연구데이터 관리 현황 및 문제점: 한국보건사회연구원 사례

제1절 관련 지침 분석

제2절 관련 사이트 분석

제3절 소결

제 3 장

연구데이터 관리 현황 및 문제점: 한국보건사회연구원 사례

제3장에서는 한국보건사회연구원을 사례로 연구데이터 관리 현황과 문제점을 파악하기 위해 당 원의 연구데이터 관리 관련 지침과 관련 사이트를 분석하였다.

제1절 관련 지침 분석

제1절에서는 한국보건사회연구원의 연구데이터 관리와 관련된 지침으로 「조사자료 처리 및 관리 지침」, 조사자료 관리 매뉴얼, 마이크로데이터 제공 업무 가이드라인에 대해 살펴보았다.

1. 「조사자료 처리 및 관리 지침(2020. 1. 1. 개정)」

한국보건사회연구원 「조사자료 처리 및 관리 지침」은 2012년 1월 31일 제정된 이후 2016년 5월 10일 전문 개정을 거쳐 2020년 1월에 이르기까지 4차에 걸친 개정이 이루어졌다. 당 원 조사자료의 처리 및 관리·제공에 필요한 사항을 규정한 지침으로 사실상 당 원의 자료 관리 체계에 있어 가장 근간을 이루는 지침이라 할 수 있다. 이는 총칙, 원내 자료 처리, 자료의 관리 및 활용 등 3개 장과 부칙으로 구성되어 있으며 8종 서식을 규정하고 있다(〈표 3-1〉 참조).

그러나 조사자료 처리, 조사자료 관리에 대한 개념이 불분명하고 이에 따른 용어 정의도 명확하지 않아 이는 곧 지침 제목에도 영향을 끼쳐 지침명도

54 전사적 연구데이터 관리 체계 마련을 위한 연구

수정이 필요하다 하겠다. 또 전체적인 구성 및 내용도 현실 상황과는 차이를 보이고 있으며 조사데이터 계획부터 시작하여 생산, 이관, 가공, 보관, 제공 및 활용 등 전 주기 내용이 담겨야 할 것으로 전체적인 수정·보완, 그리고 별지 서식에 있어서도 삭제 혹은 내용 수정이 필요하다(〈표 3-2〉 참조).

〈표 3-1〉 한국보건사회연구원 「조사자료 처리 및 관리 지침」

구분	내용	비고
지침 제목	조사자료 처리 및 관리 지침	- 용어 정의에 따라 제목도 수정 가능
제1장 총칙	제1조 목적 제2조 용어의 정의(자료 처리, 조사자료, 조사자료 처리 주관 부서, 조사자료 관리·제공 주관 부서, 마이크로데이터&부가자료)	- 목적과 용어 정의 수정 필요
제2장 원내 자료 처리	제3조 조사 의뢰 제4조 접수 제5조 조사자료의 이관 제6조 종합자료철 비치	- 장 제목 및 내용 수정 필요
제3장 자료의 관리 및 활용	제7조 자료의 관리 제8조 조사자료의 DB 이미지화 제9조 자료의 보존 제10조 자료의 제공 및 활용 제11조 외부 조사자료 제12조 특별 자료 처리	- 용어 정의에 따라 장 제목 및 내용 수정 필요
부칙	제1조 시행일	
별표1. 별지 제1호 서식 별지 제2호 서식 별지 제3호 서식 별지 제4호 서식 별지 제5호 서식 별지 제6호 서식 별지 제7호 서식 별지 제8호 서식	기초 자료 및 부책 보존 기간 조사의뢰 신청서 조사의뢰신청 기록부 조사자료 이관 신청서 조사자료 제출 및 관리 안내 확인서 조사자료의 저작권 이용허락 동의서 조사자료 DB 관리대장 자료 이용 신청서 마이크로데이터 & 부가자료 제공 관리대장	- 서식 삭제 및 수정 필요

자료: 저자 작성.

〈표 3-2〉 한국보건사회연구원 「조사자료 처리 및 관리 지침」 수정 필요 내용(일부)

현안	변경안
지침명: 「조사자료 처리 및 관리 지침」	⇒ 지침명: 조사자료 관리 지침으로 변경 필요
1. “자료처리”라 함은 조사사업을 위하여 자료를 수집하고, 수집된 자료를 활용 및 제공 가능하도록 하는 일련의 작업 과정을 총칭한다.	⇒ 1. “조사자료”라 함은 연구원 내·외부에서 수행한 면접조사, 전화조사, 우편조사, 웹조사 등과 관련하여 생성된 조사데이터, 조사표, 부호화지침서(Coding book), 변수구성표(Layout) 등 조사와 관련된 일체의 자료를 말한다.
2. “조사자료”라 함은 연구원 내·외부에서 수행한 면접조사 및 전화조사, 우편조사, 웹조사 등과 관련하여 조사자료 DB(조사표, 부호화지침서(Coding book), Layout(변수구성표), 조사데이터, 이미지화된 조사자료)를 말한다. <개정 2019. 4. 9.>	⇒ 2. “관리”라 함은 조사자료를 생산, 이관, 가공, 보관, 제공하는 일련의 작업 과정을 총칭한다.
3. “조사자료 처리 주관 부서”라 함은 조사자료의 DB를 처리하고 수집하는 부서를 말한다. <개정 2019. 4. 9.>	⇒ 3. “마이크로데이터”란 조사 원자료(raw data)에서 개인정보, 입력오류, 논리오류 등을 제거한 조사데이터를 말한다.
4. “조사자료 관리·제공 주관 부서”라 함은 조사자료 DB를 관리·보존하고 원내·외 이용자가 자료를 활용할 수 있도록 지원하는 부서를 말한다.	⇒ 삭제 필요
5. “마이크로데이터 & 부가자료”라 함은 입력오류, 조사오류 등을 제거하여 통계자료 작성의 기초 자료로 사용되는 통계 기초자료를 말하며, 개인, 가구, 사업체, 법인 또는 단체 등 개별 자료가 실질적으로 식별되지 않는 조사자료를 말한다.	⇒ 3항으로 변경 필요

자료: 저자 작성.

특히 수행 주체별 역할에 있어서도 조사자료 처리 주관 부서, 조사자료 관리·제공 주관 부서와 같이 조사자료 처리 업무와 조사자료 관리·제공 업무를 이원화하여 규정하고 있으나 이러한 업무 구분과 현 조직 체계와는 맞지 않는 문제점을 지니고 있다. 별지 서식에 있어서도 제3호 서식 ‘조사자료 이관 신청서’의 경우, 관리 항목 등 추가가 필요하며 제4호 서식 ‘조사자료 제출 및 관리 안내 확인서’와 제5호 서식 ‘조사자료의 저작권 이용허락 동의서’의 경우 삭제가 필요하다.

2. 조사자료 관리 매뉴얼(2016. 6. 9. 작성)

한국보건사회연구원 조사자료 관리 매뉴얼은 앞에서 언급한 「조사자료 처리 및 관리 지침」에 의거하여 연구사업에서 생산된 조사자료를 실질적으로 관리하기 위한 상세 설명서로 2016년 6월 9일 작성되었으며 당 원의 데이터 거버넌스 및 관리지원 체계, 조사자료 관리·지원 체계, 조사자료 관련 기안문 및 담당자, 관련 서식으로 구성되어 있다(〈표 3-3〉 참조).

〈표 3-3〉 한국보건사회연구원 조사자료 관리 매뉴얼 구성 내용

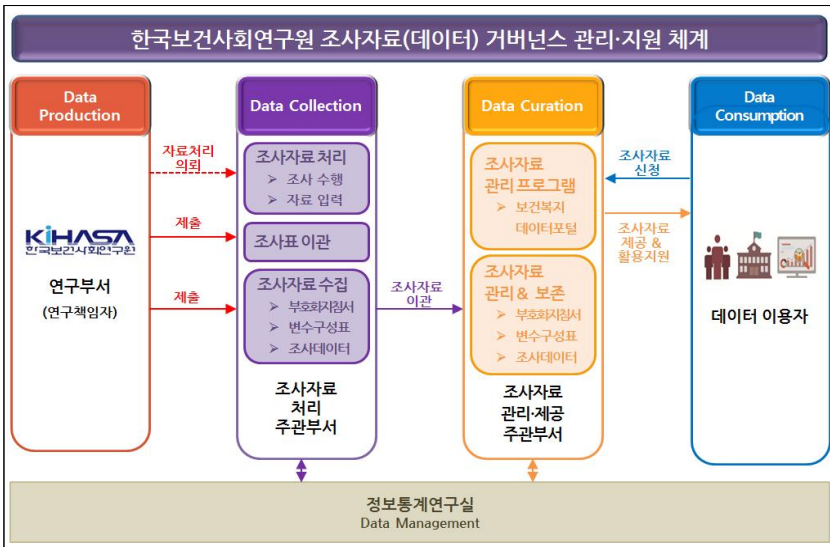
구분	내용	비고
연구원 데이터 거버넌스 및 관리 지원 체계	- 조사자료 관리의 필요성 및 목적 - 연구원 데이터 거버넌스 및 관리지원 체계 Framework	- 필요성 및 목적 수정 필요 - 연구데이터 관리 체계 그림 수정 필요
조사자료 관리·지원 체계	- 조사자료의 정의 및 범위 - 조사자료 수집 - 조사자료 관리 - 외부 구입 조사자료 관리 - 조사자료 제공	- 상세 범위 규정 필요 - 관련 서식 변경 필요 - 내용 수정 필요 - 내용 수정 필요 - 사용자 내용 등 수정 필요
조사자료 관련 기안문 및 담당자	- 조사자료 관련 기안문 위임 전결 기준 - 조사자료 수집 담당자 - 조사자료 관리·제공 담당자	- 위임 전결 내용 수정 필요 - 담당자 변경 필요 - 담당자 변경 필요
관련 서식	- 조사 의뢰 신청서 - 조사자료 이관 신청서 - 조사자료 제출 및 관리 안내 확인서 - 조사자료의 저작권 이용허락 동의서 - 조사자료 이용 신청서 - 서약서 - 보고서 발간/배포 요청서(조사자료 생산 추가)	- 각종 양식 수정·변경 및 삭제 필요

자료: 저자 작성.

우선 '조사자료 관리의 필요성 및 목적'에서는 자료 관리의 필요성을 적시하고 있으나 내용 중 외부 조사자료 구매에 대한 것은 현 상황과 맞

지 않으므로 해당 내용을 삭제할 필요가 있으며 관리 팀 명칭도 변경할 필요가 있다. 다음으로 한국보건사회연구원 데이터 거버넌스상에서 ‘조사자료 처리’와 ‘조사자료 관리·제공’으로 단계를 나누고 있으며 ‘조사자료 처리’는 조사자료를 원내 혹은 원외에서 생성하는 것으로, ‘조사자료 관리·제공’은 조사자료를 보건복지데이터 포털을 통해 보관하고 제공하는 것으로 나타나고 있다([그림 3-1] 참조). 그러나 현재 조사데이터 관리 업무에서는 보건복지데이터 포털을 통해 이관받지 못하고 파일 형태의 자료를 업무 담당자가 전달받아 업무 PC에 관리하는 등 매뉴얼과는 일치하지 않는 업무 체계가 수행되고 있으며 매뉴얼상 업무 흐름도 보완되어야 한다. 관련 서식의 경우, 현재 업무에서 활용되고 있으나 이 또한 향후 업무 절차의 수정 및 변경에 따라 삭제 혹은 수정 및 변경이 필요하다.

[그림 3-1] 한국보건사회연구원 조사관리·지원 체계



자료: 한국보건사회연구원.(2016. 6. 9.) 조사자료 관리 매뉴얼. p. 2,

3. 마이크로데이터 제공 업무 가이드라인(2019. 9. 작성)

한국보건사회연구원 마이크로데이터 제공 업무 가이드라인은 「조사자료 처리 및 관리 지침」에 의거하여 수집한 조사데이터를 가공하여 제공하기 위한 일련의 상세 업무 설명서로 2019년 9월 작성되어 실무 담당자가 활용하고 있다. 내용은 마이크로데이터에 대한 기본 설명, 제공 프로세스, 마이크로데이터 관리, 그리고 기타 업무로 구성되어 있다(〈표 3-4〉 참조). 한편 업무 가이드라인에는 포함되어 있지 않으나 실무 담당자는 마이크로데이터를 제공하기 위해 현재 〈표 3-5〉 내용을 확인하여 업무를 수행하고 있다.

〈표 3-4〉 한국보건사회연구원 마이크로데이터 제공 업무 가이드라인 내용

구분	내용	비고
마이크로데이터 기본 설명	<ul style="list-style-type: none"> - 배경 - 제공 규정 - 보유 현황 - 이용료 - 관련 담당자 	- 보유 현황, 이용료, 관련 담당자 등 수정 필요
마이크로데이터 제공 프로세스	<ul style="list-style-type: none"> - 원내 제공 - 원외 제공 	- 내용 수정 필요
마이크로데이터 관리	<ul style="list-style-type: none"> - 보건복지데이터포털 관리 	- 내용 수정 필요
기타 업무	<ul style="list-style-type: none"> - 보건복지데이터포털 관련 - 마이크로데이터 문의 - 공개용 마이크로데이터 요청 - 보건복지포럼을 이용한 홍보 - 이용자 주요 문의 사항 	- 내용 수정 필요

자료: 저자 작성.

〈표 3-5〉 마이크로데이터 제공 전 확인 사항

<p>데이터셋(제공용 마이크로데이터, 설문지, 코드북 및 파일설계서 등)이 갖추어져 있는가?</p>	<ul style="list-style-type: none"> - 설문지, 코드북 및 파일설계서의 파일 형식이 기계 판독이 가능한 최소충족포맷에 해당하는가?(PDF라면 HWP 또는 XLS 등으로 변환) - 제공용 마이크로데이터의 파일 형식이 기계 판독이 가능한 최소충족포맷에 해당하는가?(SPSS, SAS 및 STATA뿐 아니라 CSV 등의 3단계 오픈 포맷의 형태로도 제공 필요)
<p>개인정보 등 비공개 대상 정보가 포함되었는가?</p>	<ul style="list-style-type: none"> - 비공개 대상 정보가 포함된 경우, 분리 가능한가? - 성명, 생년월일, 주민등록번호, 연락처 및 이메일 등 개인에 관한 정보가 포함되어 있는가? - 조사표 등에 조사를 수행한 연구책임자와 연구진의 성명, 직위 및 연락처 등의 내용이 포함되어 있는가? - 조사표 등에 조사를 수행한 위탁조사업체에 관한 정보 및 위탁 조사업무를 수행한 자의 성명 및 직위 등이 포함되어 있는가?
<p>저작권 등 제3자 권리가 포함되었는가?</p>	<ul style="list-style-type: none"> - 필요 시 저작권자에게 “저작물 이용허락 동의서”를 받았는가? - “저작물 이용허락 동의서”에서의 이용 허락 범위를 확인하였는가? (저작권재산권 7개 항목(복제권, 공연권, 공중송신권, 전시권, 배포권, 대여권, 2차적 저작물작성권)에 대해 모두 동의를 받았는가?) - 「저작권법」제24조의2 및 「공공저작물 저작권 관리 및 이용 지침」 제11조에 근거하여 공공누리를 적용하였는가?
<p>마이크로데이터의 변수명과 코드북의 변수명이 일치하는가?</p>	<ul style="list-style-type: none"> - 비공개 변수가 제거되었는가? - 마이크로데이터에 불필요 변수는 없는가? - 마이크로데이터에 중복된 변수는 없는가? - 코드북 등에서 이전 연도 마이크로데이터에서 삭제되거나 추가된 변수에 대한 표시나 안내가 있는가?
<p>IRB 승인 구득 여부 및 관련 정보가 적절히 제시되어 있는가?</p>	

자료: 저자 작성.

이러한 가이드라인 내용은 대부분 현 상황에 맞추어 변경 및 삭제할 필요가 있으며 제공 전 확인 사항에 대한 실사례들도 체계적으로 정리하여 가이드라인에 포함시킬 필요가 있다.

제2절 관련 사이트 분석

이번 절에서는 마이크로데이터 관리 및 제공 사이트를 분석함에 있어 통계청의 마이크로데이터 통합서비스(MDIS) 사이트와 서울대학교 한국 사회과학자료원 사이트를 한국보건사회연구원의 보건복지데이터포털 (data.kihasa.re.kr)과 비교 분석하였다.

1. 통계청의 마이크로데이터 통합서비스(MDIS: Microdata Intergrated Service)(mdis.kostat.go.kr)

통계청의 마이크로데이터 통합서비스는 사업체 부문에서 경제총조사 등 13종, 인구 부문에서 인구주택총조사 등 5종, 가계 부문에서 가계금융 복지조사 등 10종, 농림어업 부문에서 농가경제조사 등 12종, 행정통계 부문에서 일자리행정통계 등 8종 등 총 48종의 마이크로데이터를 제공하고 있다(〈표 3-6〉 참조).

〈표 3-6〉 통계청 MDIS에서 제공하는 마이크로데이터 목록 현황

영역	조사명	제공 연도
사업체 부문 (13종)	건설업조사	2000~2018 (1년, 계속)
	경제총조사	2010, 2015 (5년, 계속)
	광업제조업조사	1967~2009, 2011~2014, 2016~2018 (1년, 계속)
	국부통계조사	1997 (1년, 계속)
	기업활동조사	2006~2018 (1년, 계속)
	도소매업조사	1997~2009, 2011~2014, 2016 (1년, 중지)
	서비스업조사	2002~2004, 2006~2009, 2011~2014, 2016~2018 (1년, 계속)
	서비스업총조사	1996, 2001, 2005 (중지)
	소상공인실태조사	2018

영역	조사명	제공 연도
	운수업조사	2000~2018 (1년, 계속)
	전국사업체조사	1994~2018 (1년, 계속)
	전문과학기술서비스업조사	2007~2014 (1년, 중지)
	프랜차이즈조사	2018
인구 부문 (5종)	국내인구이동통계	1986~2019(연, 계속)
	사망원인통계조사	1983~2018 (1년, 계속)
	인구동향조사	1981~2019 (연, 계속) (혼인) 1981~2019 (이혼) 1981~2019 (출생) 1981~ 2018 (사망) 1983~2018
	인구주택총조사	1960, 1966, 1970~2019 (전수 1년·표본 5년, 계속)
	주택총조사	1960, 1966, 1970~2019 (전수 1년·표본 5년, 계속)
가구 부문 (10종)	가계금융복지조사	2010~2019 (1년, 계속)
	가계동향조사	1990~2020
	가구소비실태조사	1996, 2000 (4년, 중지)
	경제활동인구조사	1981~2020 (월, 계속)
	녹색생활조사	2011, 2013 (중지)
	사회조사	1990~2019 (1년, 계속)
	생활시간조사	1999~2019 (5년, 계속)
	이민자체류실태및고용조사	2012~2019 (1년, 계속)
	지역별고용조사	2008~2019 (반기, 계속)
	초중고 사교육비조사	2007~2019 (1년, 계속)
농림어업 부문 (12종)	농가경제조사	1998~2019 (1년, 계속)
	농림어업조사	2002~2004, 2006~2009, 2011~2014, 2016~2019 (1년, 계속)
	농림어업총조사	2000, 2005, 2010, 2015 (5년, 계속)
	농어업법인조사	2001~2014 (1년, 중지)
	농업면적조사	2014~2019 (1년, 계속)
	농작물생산조사	2008~2019 (1년, 계속)
	농축산물생산비조사	2003~2019 (1년, 계속)
	세종시특별센서스조사	2013
	양곡소비량조사	2009~2019 (1년, 계속)
	어가경제조사	2003~2019 (1년, 계속)
	어류양식동향조사	2011~2018 (반기, 계속)
	어업생산동향조사	2008~2018 (월, 계속)
행정통계 (8종)	귀농어·귀촌인통계	2013~2017 (1년, 계속)
	기업생멸행정통계	2016~2018(1년, 계속)
	신혼부부통계	2015~2018 (1년, 계속)

62 전사적 연구데이터 관리 체계 마련을 위한 연구

영역	조사명	제공 연도
	영리법인기업채행정통계	2014~2018 (1년, 계속)
	일자리행정통계	2016~2018(1년, 계속)
	주택소유통계	2012~2017 (1년, 계속)
	중장년층행정통계	2016~2018(1년, 계속)
	퇴직연금통계	2015~2018(1년, 계속)

자료: 통계청 MDIS 홈페이지. (n.d.). 제공자료.
https://mdis.kostat.go.kr/infoData/serviceData.do?curMenuNo=UI_POR_P9007
 에서 2020. 10. 25. 인출.

이러한 데이터들은 가입한 회원에 한하여 무료 혹은 유료로 다운로드 서비스, 원격접근 서비스, 이용센터 서비스, 주문형 서비스와 같이 다양한 서비스를 제공하고 있다(〈표 3-7〉 참조).

〈표 3-7〉 통계청 MDIS에서 제공하는 서비스

서비스 유형	설명	제공자료 수준	유/무료
다운로드 서비스	원하는 자료를 선택하고 PC로 다운로드하여 분석하는 서비스	- 공공용 자료 - 개인식별 정보, 민감변수 등을 제외한 일반 공개용 자료	무료
원격접근 서비스 (Remote access service)	이용자가 집, 사무실에서 원격접근 시스템에 접속하여 제공받은 자료를 분석하고 결과만 승인 하에 반출하는 서비스	- 인가용 자료 - 공공용 자료보다 제공항목 추가, 항목별 제공 수준 상세화 등 보다 세부적인 자료 (RAS 전용 자료)	유료
이용센터 서비스(RDC, Research data Center)	물리적 보안 환경이 갖추어진 특정 장소(이용센터)에 방문하여 제공받은 자료를 분석하고 결과만 승인 하에 반출하는 서비스	- RAS전용 자료보다 제공항목 추가, 항목별 제공 수준 상세화 및 연계식별정보(암호화 대체기) 등 가장 세부적인 자료	유료
주문형 서비스 (On-demand Service)	통계 작성용 가구 혹은 사업체 명부 제공 서비스	- 사업체 기본정보, 가구명부 정보	유료
	이용자의 자료와 통계청의 사망 원인통계자료 연계 서비스	- 사망원인 정보	

자료: 통계청 MDIS 홈페이지. (n.d.). MDIS 소개.
https://mdis.kostat.go.kr/pageLink.do?link=content/SISMN091002&curMenuNo=UI_POR_P9002에서 2020. 10. 25. 인출.

2. 서울대학교 한국사회과학자료원의 KOSSDA(Korea Social Science Data Archive)(kossda.snu.ac.kr)

서울대학교 한국사회과학자료원의 KOSSDA는 사회과학 분야 데이터 아카이브로 다양한 양적 자료, 통계 자료, 질적 자료를 메타데이터(조사 방법, 조사 내용, 자료이용 문헌 등)와 함께 제공하고 있다(〈표 3-8〉 참조).

〈표 3-8〉 KOSSDA에서 제공하는 자료

서비스 유형	설명
양적 자료	국내 대표적인 반복횡단조사자료, 패널조사자료, 국제비교조사자료 등 다양한 주제의 조사자료 2,400여 세트
통계 자료	사회 및 경제 동향 통계표 1,900여 건
질적 자료	인구 및 가족, 지역공동체, 민족 및 외국인, 빈곤, 사회 및 노동 운동 등에 관한 텍스트, 녹음, 동영상, 이미지 등 다양한 형태의 질적 자료 200여 세트

자료: 한국사회과학자료원 KOSSDA 홈페이지. (n.d.). 양적, 질적 자료DB.
<https://kossda.snu.ac.kr/data-search/data>에서 2020. 10. 15. 인출.

회원제로 운영하고 있으며 무료로 다운로드서비스, 온라인 통계분석 서비스, 방법론 교육 등을 제공하고 있다(〈표 3-9〉 참조).

〈표 3-9〉 KOSSDA에서 제공하는 이용자별 서비스

유형	검색 및 브라우징	공개자료 다운로드	제한적 공개 자료 신청	NESSTAR 제표/분석	방법론 교육 프로그램 제공	메일링 서비스	기타자료 온라인 퍼블리싱 및 이용통계 제공
비회원	○						
회원	○	○	○	○	○	○	
자료공유회원	○	○	○	○	○	○	○

자료: 한국사회과학자료원 KOSSDA 홈페이지. (n.d.). 이용안내.
<https://kossda.snu.ac.kr/component/about/register>에서 2020. 10. 15. 인출.

3. 한국보건사회연구원 보건복지데이터포털(data.kihhas.re.kr)

한국보건사회연구원 보건복지데이터포털은 보건복지 관련 통계 정보 및 마이크로데이터를 제공하는 사이트이다. 제공하는 마이크로데이터는 한국보건사회연구원에서 실시한 조사 사업 중 2020년 11월 기준 승인 통계 14종, 미승인 통계 15종 등 총 29종(〈표 3-10〉 참조)을 가입한 회원에게 신청, 내부 검토, 승인 절차를 거쳐 사이트에서 무료 다운로드 서비스(파일 형태)를 제공하고 있다(〈그림 3-2〉 참조).

그러나 현재 보건복지데이터포털은 한국보건사회연구원의 조사데이터 관리 체계에 있어 조사데이터 수집, 가공, 보관, 제공의 일련의 작업 중 제공을 위한 인프라로서의 역할은 수행하고 있으나 내부적으로 전사적 차원의 모든 조사데이터를 수집하고 보관하기 위한 기능은 미흡하여 이에 대한 개선이 필요하다.

또 3개 사이트를 비교해 보면 보건복지데이터포털에 비해 상대적으로 MDIS 혹은 KOSSDA의 경우, 제공 데이터와 서비스가 다양하며 회원, 비회원에게 차별화된 서비스를 제공하고 있다. 그러므로 향후 보건복지 데이터포털은 질적, 양적으로 더 많은 마이크로데이터 제공과 더불어 다운로드 서비스뿐 아니라 다양한 서비스를 개발할 필요가 있다.

〈표 3-10〉 한국보건사회연구원 보건복지데이터포털에서 제공하는 마이크로데이터 목록 현황

조사명	승인 여부	제공 연도
건강위험 인지 조사	미승인	2014
국외입양인 실태 조사	미승인	2013
노숙인 등의 실태 조사	승인	2016
노인보호(학대) 실태 조사	승인	2019
노인 실태 조사	승인	1994, 1998, 2004, 2008, 2011, 2014, 2017

조사명	승인 여부	제공 연도
미혼여성의 여성계 건강에 관한 조사	미승인	2014
사회문제와 사회통합 실태 조사	미승인	2017
사회복지시설 종사자 보수수준 및 근로여건 실태 조사	미승인	2014, 2017
사회이동과 사회통합 실태 조사	미승인	2015
사회통합 및 국민행복 인식 조사	미승인	2014
사회통합 실태 및 국민인식 조사	미승인	2016
소셜데이터 통계	미승인	2019
어린이집 이용 경험 및 보육서비스 품질 실태 조사	미승인	2019
영아모성사망 조사	승인	2007
우리나라 치매노인 실태 조사	미승인	1997
웹다잉에 관한 전국민 인식 조사	미승인	2018
의료서비스 경험 조사	승인	2017, 2018
임신 및 출산에 대한 조사	미승인	2013
장애인 실태 조사	승인	2000, 2005, 2008, 2011, 2014, 2017
저소득층 자활사업 실태 조사	승인	2002, 2003, 2004
저출산 고령화 대응 자영업자 복지수요 조사	미승인	2012
전국 결혼 및 출산 동향 조사	승인	2005, 2009
전국 노인 장기요양보호 서비스 욕구 실태 조사	승인	2001
전국 다문화 실태 조사	승인	2009
전국 출산력 및 가족 보건·복지 실태 조사	승인	1991, 1994, 1997, 2000, 2003, 2006, 2009, 2012, 2015, 2018
중년층의 생활 실태 및 복지욕구 조사	승인	2010
차상위계층 실태 조사	승인	2004, 2007
한국사회의 사회심리적 불안 실태 조사	미승인	2015
환자 조사	승인	1996, 1999, 2002, 2005, 2008, 2009년부터 2016년까지 매년

자료: 보건복지데이터포털 홈페이지. (n.d.). 마이크로데이터 전체목록.

<https://data.kihasa.re.kr/microdata/apply/list>에서 2020. 10. 25. 인출.

한국보건사회연구원에서는 보건복지데이터포털 외에도 한국의료패널 홈페이지(khp.re.kr)와 한국복지패널 홈페이지(koweps.re.kr)에서 각각 한국의료패널 조사데이터와 한국복지패널 조사데이터를 이메일, 다운로드서비스로 제공하고 있다.

[그림 3-2] 보건복지데이터포털 이용 안내



자료: 보건복지데이터포털 홈페이지. (n.d.). 마이크로데이터 이용 안내.
(<https://data.khisa.re.kr/microdata/guide>에서 2020. 11. 10. 인출.

제3절 소결

한국보건사회연구원의 연구데이터를 관리하기 위한 지침 등을 살펴본 결과, 용어의 불명확성, 현실을 반영하지 못한 업무 절차 등이 문제점으로 나타났으며 구성 내용에 있어서도 보완 및 수정이 필요하다. 특히 조사데이터를 ‘처리’하는 것과 ‘관리 및 제공’하는 업무를 구분하고, 이에 따른 역할과 담당자가 제시되어 있으나 이렇게 이원화하여 관리함으로써 두 업무 간 연계성이 부족하고 ‘관리 및 제공’ 업무를 해당 과제 소속 연구진이 담당하게 되어 있어 연속성의 문제도 제기되는 등 전반적인 재검토가 필요한 부문이다.

또 연구데이터 관리 절차상 필요한 관리 기준(대상), 관리 항목(메타데이터), 관리 인프라, 이관 절차, 공개 기준 및 절차, 이용 등에 대한 내용 보완도 필요하다.

한편 연구데이터 관리를 위한 인프라로서 한국보건사회연구원에서는 보건복지데이터포털을 운영하고 있다. 그러나 현재는 조사데이터를 전반적으로 관리하기 위한 목적보다는 조사데이터 중 공개 데이터를 웹상에서 제공하기 위한 수단으로 사용하고 있다. 그러므로 향후 조사데이터를 제공하기 위한 수단뿐 아니라 조사데이터 수집, 가공, 보관과 같은 전반적인 관리 수단으로서 역할을 수행하기 위해 내부 업무시스템과의 연계, 조사데이터 관리 기능 개발이 필요하다. 또 웹상에서 조사데이터를 질적, 양적으로 더 많이 제공하는 것과 함께 다운로드 서비스 이외에도 검색 기능 서비스, 맞춤형 신청서비스, 통계 기능 서비스 등과 같은 사용자 요구에 맞는 다양한 서비스 개발이 필요하다.



제4장

연구데이터 관리 체계 및 인식·실태 조사

제1절 조사 개요

제2절 조사 결과

제3절 소결

제4장

연구데이터 관리 체계 및 인식·실태 조사

제4장에서는 한국보건사회연구원의 연구데이터 관리 체계 마련을 위한 기초 자료로 활용하기 위해 한국보건사회연구원을 포함한 경제인문사회계열 정부출연연구기관들의 조사데이터 관리 현황과 연구책임자들의 연구데이터 관리, 공유, 활용에 관한 인식과 실태를 온라인 조사를 통해 파악하였다.

제1절 조사 개요

이번 조사는 1) 한국보건사회연구원의 연구데이터 관리 현황 및 연구책임자들의 연구데이터 관리, 공유, 활용에 대한 인식과 실태 조사 2) 경제인문사회연구회 소속 정부출연연구기관의 연구데이터 관리 현황 조사 3) 경제인문사회연구회 소속 정부출연연구기관 연구책임자들의 연구데이터 관리, 공유, 활용에 대한 인식과 실태 조사 등 3가지로 나누어 실시하였다.

첫 번째 한국보건사회연구원에 대한 조사는 9월 초 기관에 대한 연구데이터 관리 현황 파악과 아울러 당 원 연구책임자들을 대상으로 2020년 9월 10일부터 약 2주간 ‘서베이몽키’ 온라인 설문조사 플랫폼을 통해 실시하였으며, 총 49인이 참여하였다.

두 번째 경제인문사회연구회 소속 정부출연연구기관에 대한 연구데이터 관리 현황 조사는 2020년 10월 기준 경제인문사회연구회 홈페이지에

72 전사적 연구데이터 관리 체계 마련을 위한 연구

공개되어 있는 26개 소관 기관⁶⁾ 중 한국보건사회연구원과 KDI 국제정책대학원을 제외한 24개 연구기관을 대상으로 2020년 10월 14일부터 약 한 달간 이메일 조사를 실시하였으며 이 중 19개 연구기관이 참여하였다.

세 번째 경제인문사회연구회 소속 정부출연연구기관 연구책임자들에게 대한 조사는 해당 기관을 통해 10월 14일부터 약 3주간 ‘서베이몽키’ 온라인 설문조사 플랫폼을 통해 실시하였으며 16개 기관의 총 182명이 참여하였다.

이와 같은 3가지 조사의 대상, 기간, 방법을 정리하면 <표 4-1>과 같다.

<표 4-1> 연구데이터 관리 현황과 관리, 공유, 활용에 관한 인식 및 실태 조사 개요

구분	원내		원외	
	기관	개인	기관	개인
조사 내용	연구데이터 관리 현황	연구데이터 관리, 공유, 활용에 관한 인식 및 실태	연구데이터 관리 현황	연구데이터 관리, 공유, 활용에 관한 인식 및 실태
조사 대상 (응답자)	한국보건사회연구원	한국보건사회연구원 연구책임자(49명)	경제인문사회연구회 소속 연구기관 (19개기관)	경제인문사회연구회 소속 연구기관 연구책임자(182명)
조사 기간	2020. 09. 10.부터 약 2주간		2020. 10. 14.부터 약 한 달간	2020. 10. 14.부터 약 3주간
조사 방법	담당자 면담	‘서베이몽키’를 활용한 온라인 조사	이메일 조사	‘서베이몽키’를 활용한 온라인 조사

자료: 저자 작성.

6) 경제인문사회연구회 소속 기관은 「정부출연연구기관 등의 설립·운영 및 육성에 관한 법률」에 따라 설립되는 24개 연구기관에 육아정책연구소와 KDI국제정책대학원을 포함하여 총26개 기관임.

각 조사에서 기관을 대상으로 한 조사 내용과 개인을 대상으로 한 조사 내용을 살펴보면, 먼저 기관을 대상으로 한 관리 현황 조사 내용은 크게 기관 개요와 조사데이터 관리 체계로 구분하여 각 3개 항목, 13개 항목으로 구성하였다(〈표 4-2〉 참조).

〈표 4-2〉 연구기관 대상 연구데이터 관리 현황 조사 내용

구분	조사 항목
기관 개요	<ul style="list-style-type: none"> - 해당 기관의 연구책임자 수 - 최근 3년간(2017~2019년도) 연구 수행 건수 - 최근 3년간(2017~2019년도) 수행 연구 중 조사 수행 연구 건수
조사데이터 관리 체계	<ul style="list-style-type: none"> - 조사데이터 관리 체계 유무 - 조사데이터 관리 체계 관련 규정/지침 혹은 매뉴얼 유무 및 명칭 - 전담팀 유무 및 규모 - 조사데이터 관리 정보시스템 구축 여부 - 조사데이터 이관 혹은 공개 시 연구자에 대한 인센티브 유무 및 인센티브 내용 - 이관 실적 - 조사데이터 내부 신청 및 제공 유무, 신청 및 제공 방식 - 외부 공개 조사데이터 유무 및 조사데이터명 - 조사데이터 외부 공개 방식 및 경로 - 조사데이터 제공 형식 - 제공 데이터셋 종류 - 조사데이터 제공 우수사례 - 조사데이터 관리 체계 관련 의견

자료: 저자 작성.

다음으로 개인을 대상으로 한 조사 내용은 응답자 특성, 연구데이터 관련 개념에 대한 인식, 타 연구데이터 활용 경험, 연구책임자로서 생산한 연구데이터 관리 경험, 소속 기관 연구데이터 관리 체계에 대한 인식, 연구데이터 개방에 대한 인식 및 경험 등 6가지 영역으로 구성하였다(〈표 4-3〉 참조).

74 전사적 연구데이터 관리 체계 마련을 위한 연구

〈표 4-3〉 연구데이터 관리, 공유, 활용에 관한 인식 및 실태 조사 내용

구분	원내(개인) 조사 항목	원외(개인) 조사 항목
응답자 특성	<ul style="list-style-type: none"> - 성별 - 연령대 - 한국보건사회연구원에서의 연구책임자 경력 기간 - 연구 수행에 있어 연구 방법론 중 조사 연구 방법에 대한 중요성 정도 	<ul style="list-style-type: none"> - 소속 기관 - 성별 - 연령대 - 소속 기관에서의 연구책임자 경력 기간 - 연구 수행에 있어 연구 방법론 중 조사 연구 방법에 대한 중요성 정도
연구데이터 관련 개념에 대한 인식	<ul style="list-style-type: none"> - 오픈 사이언스 개념 및 동향 인식 정도 - 경제인문사회계열 연구에서의 오픈 사이언스 정책 도입 필요성 정도 - 연구데이터 관리 계획 개념 및 동향 인식 정도 - 연구데이터 관리 계획 정책 중요성 정도 - 당 원의 연구데이터 관리 계획 정책 도입 필요성 정도 	<ul style="list-style-type: none"> - 오픈 사이언스 개념 및 동향 인식 정도 - 본인 연구 분야에서의 오픈 사이언스 정책 도입 필요성 정도 - 연구데이터 관리 계획 개념 및 동향 인식 정도 - 연구데이터 관리 계획 정책 중요성 정도 - 본인 연구 분야에서의 연구데이터 관리 계획 정책 도입 필요성 정도
타 연구 조사데이터 활용 경험	<ul style="list-style-type: none"> - 원내 타 연구자의 조사데이터 필요 또는 활용 경험 여부 - 원내 타 연구자의 조사데이터 필요 또는 활용 경험 평가 - 활용한 원내 타 연구자 조사데이터명 - 원내 타 연구자의 조사데이터 활용 시 어려웠던 점 - 타 기관의 조사데이터 활용 경험 여부 - 타 기관의 조사데이터 활용 경험 평가 - 타 기관 조사데이터 획득 경로 - 활용한 타 기관 조사데이터명 - 타 기관 조사데이터 활용 시 어려웠던 점 	<ul style="list-style-type: none"> - 타 연구자의 조사데이터 활용 경험 여부 - 타 연구자의 조사데이터 활용 경험 평가 - 타 연구자의 조사데이터 활용 시 어려웠던 점 - 타 연구자의 조사데이터 획득 경로
연구책임자로서 생산한 조사데이터 관리 경험	<ul style="list-style-type: none"> - 최근 3년간 조사데이터 생산 경험 여부 - 조사데이터의 주된 보관·관리자 - 조사데이터의 적절한 보관·관리자 - 조사데이터 소실 경험 여부 - 조사데이터 소실 원인 	<ul style="list-style-type: none"> - 최근 3년간 조사데이터 생산 경험 여부 - 조사데이터의 주된 보관·관리자 - 조사데이터의 적절한 보관·관리자 - 조사데이터 소실 경험 여부 - 조사데이터 소실 원인
소속 기관 조사데이터 관리 체계에 대한 인식	<ul style="list-style-type: none"> - 당 원 조사데이터 관리 체계 평가 - 당 원 조사데이터 관리 체계 보완 시급한 부분 평가 	<ul style="list-style-type: none"> - 소속 기관 조사데이터 관리 체계 평가
조사데이터 개방에 대한 인식 및 경험	<ul style="list-style-type: none"> - 조사데이터 공유 및 개방 필요성 정도 - 조사데이터 공유 및 개방이 필요한/필요하지 않은 이유 - 본인이 생산한 조사데이터 공유 및 개방 필요성 정도 	<ul style="list-style-type: none"> - 조사데이터 공유 및 개방 필요성 정도 - 조사데이터 공유 및 개방이 필요한/필요하지 않은 이유 - 본인이 생산한 조사데이터 공유 및 개방 필요성 정도

구분	원내(개인) 조사 항목	원외(개인) 조사 항목
	<ul style="list-style-type: none"> - 본인이 생산한 조사데이터 공유 및 개방이 필요한/필요하지 않은 이유 - 본인이 생산한 조사데이터 공유 및 개방 경험 여부 - 본인이 생산한 조사데이터 공유 및 개방 시 어려웠던 점 - 보건복지데이터포털을 통한 원내 조사데이터 개방에 대한 인지 여부 - 보건복지데이터포털을 통한 원내 조사데이터 활용 경험 및 개선점 	<ul style="list-style-type: none"> - 본인이 생산한 조사데이터 공유 및 개방이 필요한/필요하지 않은 이유 - 조사데이터 공유 및 개방을 위해 필요한 부분 평가

자료: 저자 작성.

제2절 조사 결과

이번 절에서는 경제인문사회연구회 소관 연구기관의 연구데이터 관리 현황과 한국보건사회연구원의 조사데이터 관리 현황을 비교·분석하고, 연구책임자들의 관리, 공유, 활용에 관한 인식 및 실태에 대해 원내(한국보건사회연구원), 원외(경제인문사회연구회 소속 정부출연연구기관)로 나누어 살펴보고자 한다.

1. 한국보건사회연구원과 경제인문사회연구회 소속 연구기관의 연구데이터 관리 현황

가. 기관 개요

한국보건사회연구원을 제외한 경제인문사회연구회 소속 연구기관들의 최근 3년간(2017~2019년) 연구(자체, 수탁 모두 포함) 사업을 진행하면서 조

사를 수행한 연구 비율(조사데이터 생산 현황)을 조사한 결과, 평균 37.7% (표준편차 5.4)로 나타났다. **한국보건사회연구원의 조사 수행률은 39.6%로 평균보다 약간 높게 나타났다.** 이러한 조사 수행률은 최대 80.0%, 최소 5.2%로 기관별로 많은 차이를 보이고 있었으며 해당 문항 응답 기관 14개 중 5개 기관이 40% 이상으로 나타났다(〈표 4-4〉 참조). 조사 사업 수행은 조사데이터 관리의 필요성을 나타내는 것으로 조사데이터 관리 체계의 출발점이 된다고 할 수 있겠다.

〈표 4-4〉 연구기관의 최근 3년간(2017~2019년) 수행 연구 중 조사 수행률

구분	기관 수(비율) ^{주)}	평균
계	14개 (100.0%)	37.7%
20% 미만	3개 (21.4%)	
20%~40% 미만	6개 (42.9%)	
40%~60% 미만	3개 (21.4%)	
60%~80% 미만	1개 (7.1%)	
80% 이상	1개 (7.1%)	

주: 1개 기관의 경우 자체 과제만을 대상으로 파악됨.
 자료: 저자 작성.

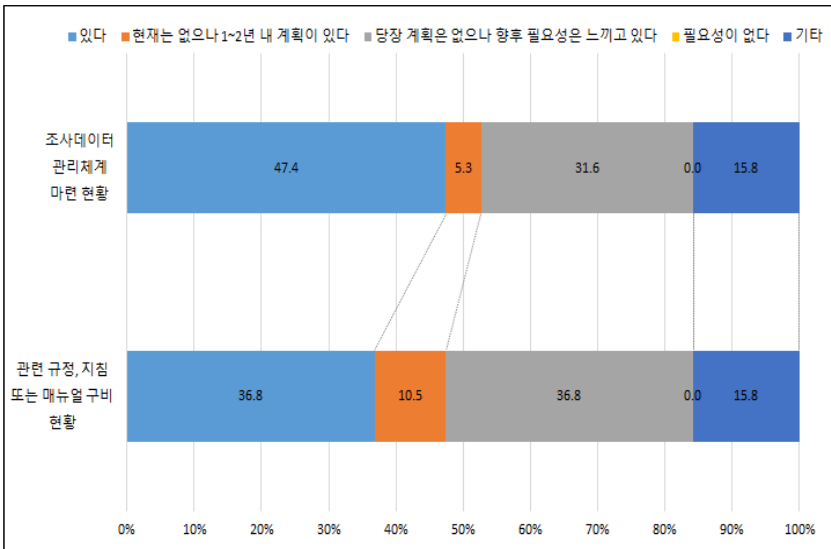
나. 조사데이터 관리 체계

다음으로 각 연구기관에서 연구 종료 후 조사데이터를 관리(데이터 이관, 보관, 공유, 공개 등)하는 체계가 마련되어 있는지에 대한 설문에는 응답 기관 19개 중 47.4%가 마련되어 있다고 응답하였으며, 5.3%가 1~2년 내 계획이 있으며 31.6%는 당장 계획은 없으나 필요성은 느끼고 있는 등 대부분이 관리 체계의 필요성을 인정하고 있었다. 더불어 조사데이터 관리 체계와 관련된 규정, 지침 또는 매뉴얼이 있는지에 대한 설문에는 응답 기관 19개 중 36.8%가 구비되어 있다고 응답하였다. 관리 체

계가 마련되어 있다고 응답한 기관 중에서도 15.8%(3개 기관)는 관련 규정 등을 갖추고 있지 않은 것을 알 수 있다. 관련 규정 등의 명칭은 ‘통계자료 관리 지침’, ‘조사자료 공개 지침’, ‘조사 도구 및 조사자료 관리 지침’, ‘연구 자료 관리 규칙’, ‘자료 관리 규칙’, ‘저작물 관리 지침’ 등 다양하였다. 관련 규정 등에 대한 필요성이 없다고 응답한 기관은 없었으며, 기타 의견 15.8%는 일부 규정만 구비하고 있다는 응답으로서 자료보안 관리 규칙만 구비하고 있거나 데이터 공개에 대한 절차의 일부만 규정하고 있다고 답하였다(그림 4-1) 참조).

한국보건사회연구원의 경우에는 2012년 제정된 ‘조사자료 처리 및 관리지침’을 2016년 전문 개정하여 조사자료 처리 주관 부서와 조사자료 관리·제공 주관 부서로 이원화하여 조사데이터 보존 및 외부 제공을 위한 근거 규정을 마련한 바 있다.

[그림 4-1] 연구기관의 조사데이터 관리 체계 및 관련 규정 등 마련 현황



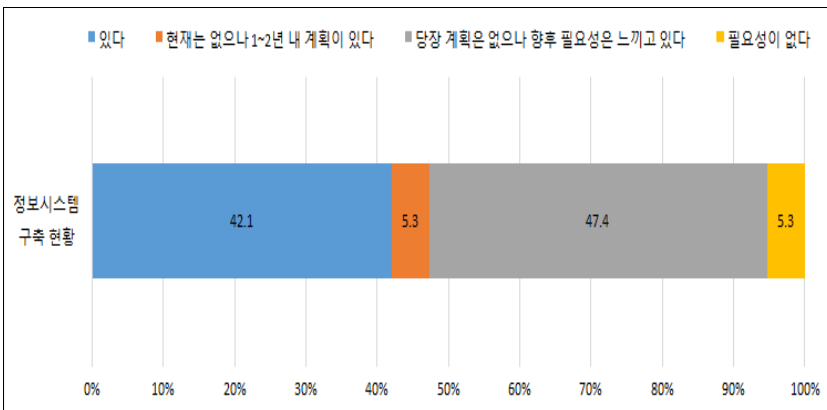
자료: 저자 작성.

조사데이터를 관리하는 전담팀이 있는 기관은 총 19개 기관 중 5개 기관이었으며, 전담팀이 없는 나머지 14개 기관의 경우에는 연구관리팀, 성과홍보실, 여론분석팀, 성과공유팀, 경영기획실 등 기관의 공공데이터 업무를 총괄하는 팀에서 전담하고 있었다.

한국보건사회연구원의 경우에는 조사데이터 수집 및 관리는 사회조사 업무를 수행하는 정보통계연구실 사회조사팀에서, 조사데이터 제공은 보건복지데이터포털 운영 사업을 수행하는 정보통계연구실 해당 연구진에서 수행하고 있으며, 공공데이터 업무 책임 부서는 정보통계연구실로 지정되어 있다.

조사데이터 관리 정보시스템의 경우, 총 19개 기관 중 42.1%가 구축되어 있다고 답하였으며 대부분 필요성은 느끼고 있었다(그림 4-2 참조). 해당 시스템명은 연구사업시스템, 내부 조사DB시스템, 지식자산공유시스템, 업무종합관리시스템과 같은 사내 전산시스템(그룹웨어)이나 외부 인터넷시스템(국가에너지통계종합정보시스템 등)으로 나타났다.

[그림 4-2] 연구기관의 조사데이터 관련 정보시스템 구축 현황

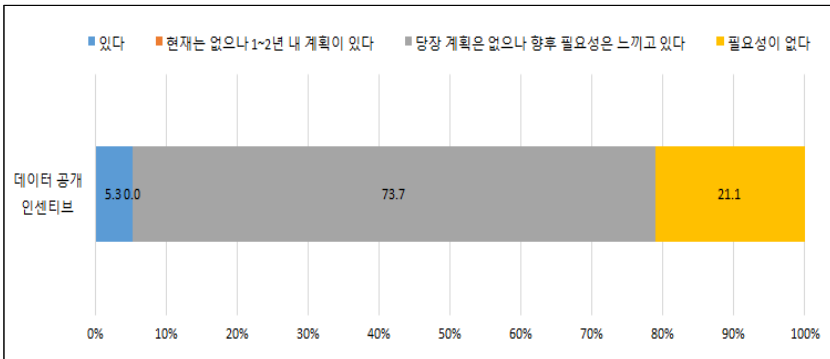


자료: 저자 작성.

한국보건사회연구원의 경우에는 내부 업무시스템에서는 제공하고 있지 않으며 인터넷시스템인 보건복지데이터포털(data.kihasa.re.kr)을 통해 전국출산력 및 가족 보건복지실태조사 등 총 29종의 조사데이터를, 한국복지패널조사와 한국의료패널조사 데이터는 각각의 홈페이지(koweps.re.kr, khp.re.kr)를 통해 제공하고 있다.

다음으로 조사데이터를 기관에 이관하거나 기관을 통해 외부에 공개한 경우 해당 연구자에게 제공하는 인센티브가 있는지에 대한 문항에서는 1개 기관(5.3%) 만이 있다고 응답하였으며, 인센티브 내용은 ‘개인 연구 활동 평가에 대한 평점 부여(1건당 2점, 최대 연간 20점)’였다. 향후 필요성을 느끼고 있는 기관은 73.7%로 많은 기관이 필요성에 공감하고 있는 반면 필요성이 없다고 응답한 기관도 21.1%로 나타났다(그림 4-3) 참조). 이처럼 필요성이 없다고 응답한 기관은 공적 자금을 투입하여 생산한 조사데이터이므로 기관이 이관하거나 외부에 공개하는 것을 당연하게 생각하기 때문이거나 이관하거나 공개하는 것이 이미 연구원 문화로 정착되었기 때문인 것으로 추측할 수 있다.

[그림 4-3] 연구기관의 데이터 공개에 대한 인센티브 제공 현황



자료: 저자 작성.

한국보건사회연구원에서는 아직까지 데이터 이관 및 외부 공개에 대해 별도의 인센티브 제도가 마련되어 있지 않다.

다음으로 각 연구기관에서 이관한 조사데이터의 수는 2종에서 2,840종까지 다양하게 나타났으나 이는 질적 데이터에 대한 숫자, 시계열 자료의 연도별 데이터에 대한 숫자 등 조사데이터의 범위 및 단위를 기관마다 상이하게 적용하여 비교 가능한 결과는 아니라고 판단하여 이번 보고서에서는 다루지 않기로 하였다.

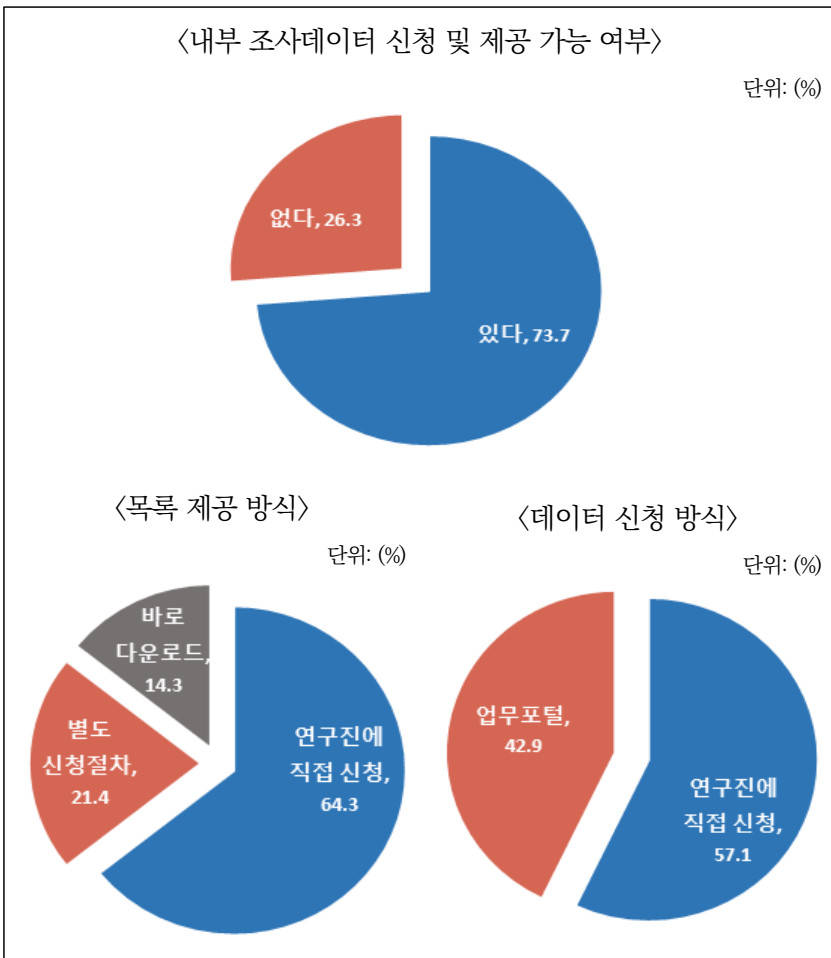
각 기관에서 내부적으로 조사데이터를 신청 및 제공받을 수 있는지에 대한 문항에서는 총 19개 기관 중 73.7%(14개 기관)가 내부적으로 조사데이터를 공유하고 있다고 응답하였다. 공유하고 있다고 응답한 기관들이 조사데이터 목록을 제공하는 방식을 살펴보면, 연구보고서 등을 보고 연구진에게 문의해서 데이터를 신청하는 경우가 57.1%, 업무포털에 원내에서 생산되어 이관된 조사데이터 목록을 제시하는 경우가 42.9%로 나타났다. 조사데이터를 신청하는 방식은 연구진에 직접 신청하여 수령하는 경우가 64.3%로 가장 많은 가운데, 별도의 신청 절차(조사자료 신청서 제출 → 심의)가 있는 경우가 21.4%(3개 기관), 업무포털에서 바로 다운로드하는 경우가 14.3%(2개 기관)였다(그림 4-4) 참조.

반면 한국보건사회연구원은 업무포털에서 이관된 조사데이터 목록을 따로 제공하고 있지는 않으며, 조사데이터가 전면 무료화(2020. 1. 1.)된 이후로는 보건복지데이터포털, 한국의료패널 홈페이지, 한국복지패널 홈페이지를 통해 내부 직원도 외부와 동일하게 조사데이터를 구독하고 있다.

조사데이터를 외부에 공개하고 있는지에 대한 항목에서는 19개 기관 중 15개 기관(78.9%)이 공개를 하고 있다고 응답하였다. 공개하고 있는 조사데이터는 재정패널조사, 에너지총조사, 가구에너지패널조사, 한국교육고용패널조사 등 국가승인통계가 대부분이었다. 조사데이터 공개 경로는 기관에서 운영하는 자체 데이터포털(별도 홈페이지 포함)이 45.0%,

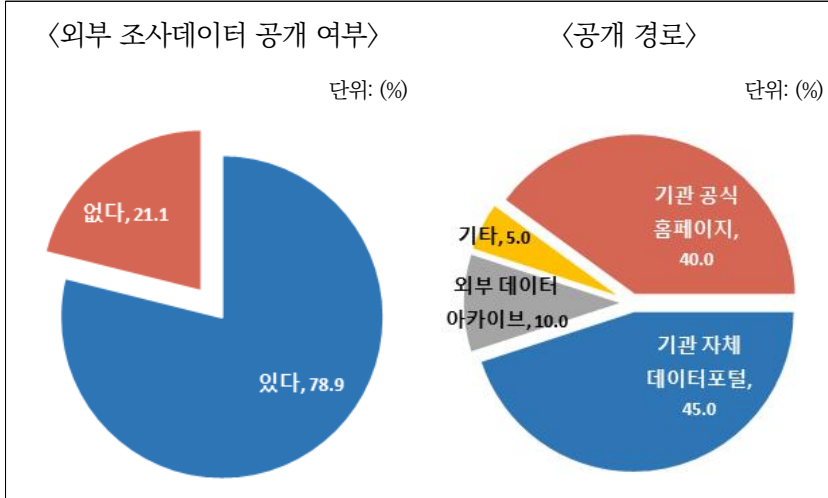
기관 공식 홈페이지를 활용하는 경우가 40.0%, 외부 데이터 아카이브(한국사회과학자료원)를 활용하는 경우가 10.0%로 나타났으며, 일부 기관의 경우에는 USB를 통해 조사데이터를 제공하고 있는 것으로 나타났다 ([그림 4-5] 참조).

[그림 4-4] 연구기관의 내부 조사데이터 신청 및 제공 현황



자료: 저자 작성.

[그림 4-5] 연구기관의 조사데이터 외부 공개 현황

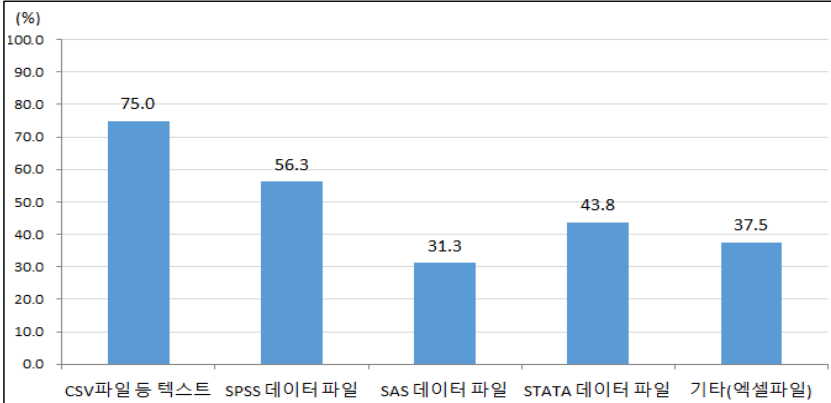


주: 공개 경로는 중복 응답. 공개 경로는 외부 조사데이터 공개를 하고 있는 14개 기관을 대상으로 작성
 자료: 저자 작성.

각 기관에서 내·외부에 제공하는 조사데이터는 CSV파일 등 텍스트 형태가 가장 많았으며(75.0%), 다음으로 SPSS 파일(56.3%), STATA 파일(43.8%), 엑셀파일(37.5%), SAS 파일(31.3) 순으로 나타났다(그림 4-6 참조). 참고로 공공데이터포털에서는 특정 소프트웨어에 종속되지 않는 오픈 포맷 형식(CSV, JSON, XML)으로 공공데이터를 제공할 것을 권장하고 있다.

한국보건사회연구원에서는 한국복지패널조사와 한국의료패널조사의 경우 SAS, SPSS, STATA 형식으로 제공하고 있으며, 보건복지데이터포털에서는 CSV(텍스트 포함), SPSS, SAS, 엑셀파일 중 연구진에서 이관한 형식 그대로 내·외부에 제공하고 있다.

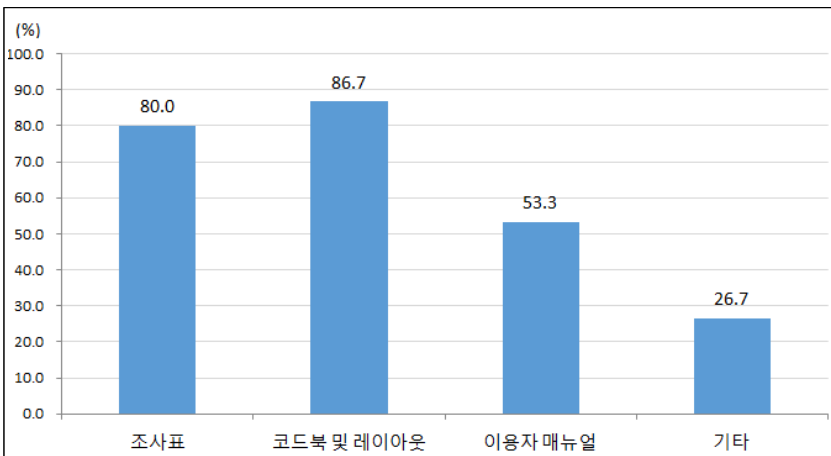
[그림 4-6] 연구기관의 조사데이터 제공 형식(중복 응답)



자료: 저자 작성.

조사데이터와 함께 제공하는 자료로는 코드북 및 레이아웃이 86.7%로 가장 많았으며 기타 자료로 조사지침서, 데이터 설명회 자료 등도 제공하는 것으로 나타났다(그림 4-7 참조).

[그림 4-7] 연구기관의 조사데이터와 함께 제공하는 관련 자료(중복 응답)



자료: 저자 작성.

연구기관의 조사데이터 제공 우수사례로는 외부에 제공하고 있는 조사 데이터를 활용한 성과 목록과 활용 성과물을 홈페이지를 통해 공개하고 있는 경우, 조사데이터 활용 확산을 위해 논문 경진대회를 추진하고 있는 경우 등을 제시하였다. **한국보건사회연구원에서는 한국의료패널, 한국복지패널 데이터를 활용한 학술대회를 매년 개최하고 있다.**

마지막으로 조사데이터 관리 체계에 대한 의견을 자유롭게 기술하라는 문항에서는 현재 각 연구기관에서 느끼는 조사데이터 관리 체계의 필요성에 대한 공감대와 현실적인 관리의 어려움을 적시하였다. 예를 들어 각 연구기관에서는 개별 연구자 입장에서 본인이 구축한 조사데이터를 공유할 유인이 부족하고, 조사데이터를 다른 사람들이 효율적으로 활용하기 위해 데이터를 정제하고 부가적인 작업 등을 수행하는 주체에 대한 이견, 연구 수행 과정에서 일회성에 그치는 설문조사의 경우 시간이 지나면서 그 유효성이 크게 저하됨에도 해당 자료를 관리해야 하는지에 대한 이견 등 다양한 의견을 제시하였다. 또한 데이터 관리 및 보존이 되지 않아 연구자 퇴직이나 이동으로 인해 자료가 유실되는 경우가 많으며, 연구책임자별로 연구 수행을 위해 생산되는 조사연구 및 결과 데이터를 전사적으로 취합하여 관리 및 공유하는 것은 데이터와 정보의 공유 및 확산, 예산 절감, 데이터 검증 및 연구 검토를 위해 반드시 필요한 작업이라는 데는 공감대를 함께 하는 것으로 나타났다. 이를 위해 일부 기관에서는 **연구회 차원에서 인력 및 예산을 적극적으로 지원하는 방안도** 제시하고 있었다.

2. 한국보건사회연구원 연구책임자 대상 데이터 관리, 공유, 활용에 관한 인식 및 실태

가. 응답자 특성

한국보건사회연구원 연구책임자 응답자 49명에 대한 특성을 살펴보면 남성이 30.6%, 여성이 69.4%로 나타났으며, 연령별로는 40~49세가 61.2%로 가장 많았다. 연구책임자로 근무한 경력 기간은 5년 미만이 44.9%, 5년 이상 10년 미만이 26.5%, 10년 이상 15년 미만이 14.3% 순으로 나타났다. 한편 연구 수행에 있어 조사 연구 방법을 얼마나 중요하게 생각하는지에 대해 7점 만점에 6.2점으로 **응답자들이 조사 연구 방법을 매우 중요하게 인식하고 있는 것으로 나타났다**(〈표 4-5〉 참조).

〈표 4-5〉 한국보건사회연구원 연구책임자 응답자 현황

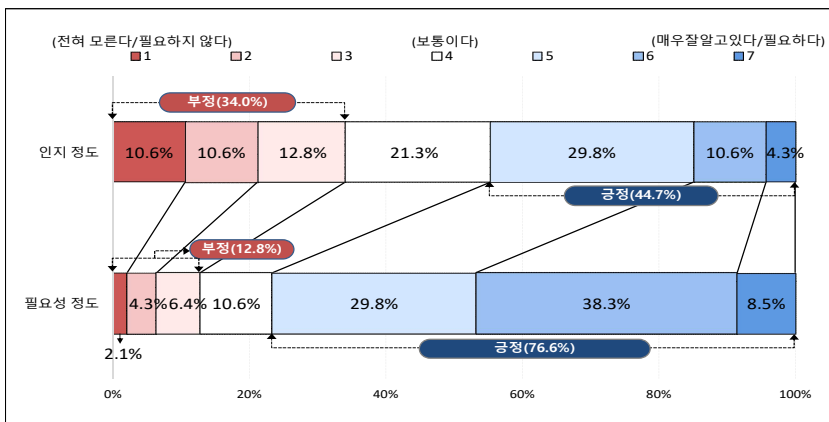
구분		연구책임자 수(명)	비율(%)
계		49	100.0
성별	남자	15	30.6
	여자	34	69.4
연령	30~39세	9	18.4
	40~49세	30	61.2
	50~59세	10	20.4
우리 연구원에서의 연구책임자 경력 기간	5년 미만	22	44.9
	5년~10년 미만	13	26.5
	10년 이상	14	28.6
	- 10년 이상 15년 미만	7	14.3
	- 15년 이상 20년 미만	3	6.1
	- 20년 이상	4	8.2
	조사 연구 방법 중요성 인식 정도	부정(①+②+③)	1
	보통(④)	1	2.0
	긍정(⑤+⑥+⑦)	47	95.9

자료: 저자 작성.

나. 연구데이터 관련 개념에 대한 인식

응답자 47명의 오픈 사이언스 개념과 동향 인지 정도, 오픈 사이언스 정책 도입의 필요성 인식 정도를 7점 척도로 측정하였다. ‘오픈 사이언스의 개념과 동향’에 대해 긍정적 응답(매우 잘 알고 있다+알고 있다+약간 알고 있다)은 44.7%, 부정적 응답(전혀 모른다+모른다+별로 모른다)은 34.0%로 나타나 큰 차이를 나타내지는 않았다. 반면 ‘경제인문사회계열 연구에서 오픈 사이언스 정책 도입이 필요하다’는 문항에서는 긍정적 응답(매우 필요하다+필요하다+약간 필요하다)이 76.6%, 부정적 응답(전혀 필요하지 않다+필요하지 않다+별로 필요하지 않다)은 12.8%로 긍정적 응답이 부정적 응답에 비해 매우 높게 나타났다. 이는 즉 한국보건사회연구원 연구책임자들은 오픈 사이언스 개념과 동향에 대해 현재 알고 있는 것보다 필요성에 대해 더 많이 공감하고 있다는 것을 알 수 있다(그림 4-8 참조).

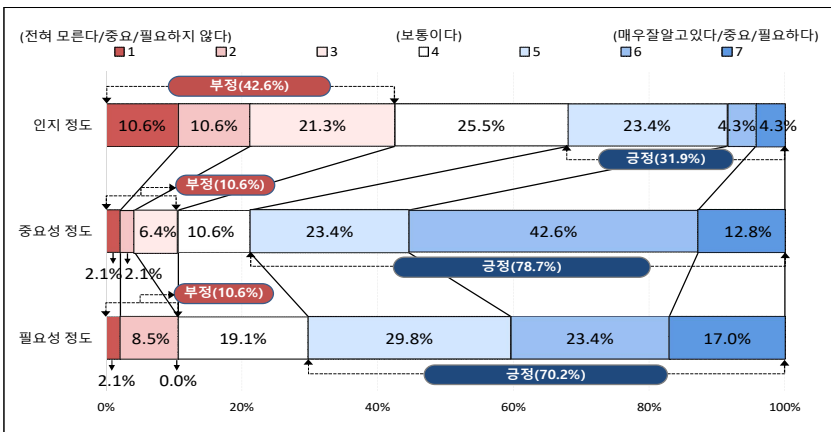
[그림 4-8] 한국보건사회연구원 연구책임자들의 오픈 사이언스에 대한 인식



자료: 저자 작성.

한편 응답자들의 연구데이터 관리 계획에 대한 인식을 알아보기 위해 연구데이터 관리 계획 개념과 동향의 인지 정도와 연구데이터 관리 계획 정책의 중요성, 정책 도입의 필요성을 7점 척도로 측정하였다. ‘연구데이터 관리 계획의 개념과 동향’에 대해 긍정적 응답(⑦매우 잘 알고 있다+⑥알고 있다+⑤약간 알고 있다)은 31.9%로 부정적 응답(①전혀 모른다+②모른다+③별로 모른다) 42.6%에 비해 다소 낮은 수준으로 나타났으나, ‘연구데이터 관리 계획 정책의 중요성’과 ‘연구데이터 관리 계획 정책 도입의 필요성’에 대해서는 긍정적 응답(⑦매우 중요하다+⑥중요하다+⑤약간 중요하다)이 각각 78.7%, 70.2%이고 부정적 응답(①전혀 필요하지 않다+②필요하지 않다+③별로 필요하지 않다)이 각 10.6%인 것에 비해 매우 높게 나타났다. 즉 한국보건사회연구원 연구책임자들은 현재는 연구데이터 관리 계획 개념에 대해 크게 인지하지 않고 있다 하더라도 관련 정책의 중요성과 도입의 필요성은 상대적으로 높게 인식하고 있다는 것을 알 수 있었다(그림 4-9) 참조).

[그림 4-9] 한국보건사회연구원 연구책임자들의 연구데이터 관리 계획에 대한 인식



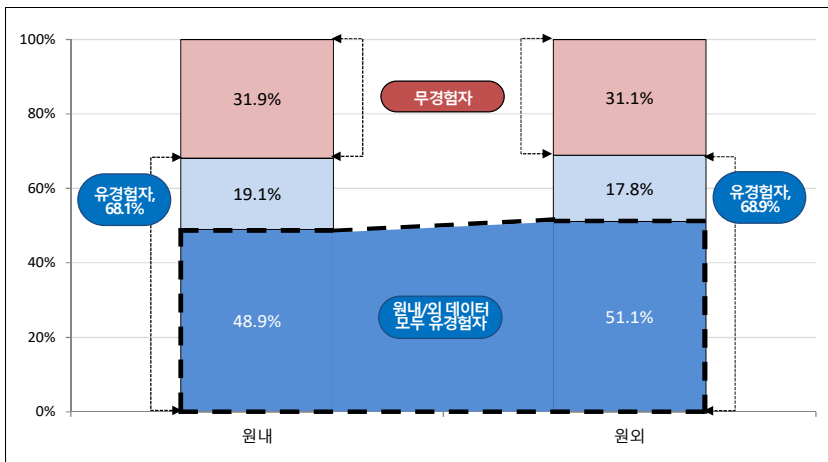
자료: 저자 작성.

다. 타 연구 조사데이터 필요 또는 활용 경험

한국보건사회연구원 연구책임자들의 원내 타 연구자의 조사데이터 필요 또는 활용 경험과 타 기관의 조사데이터 활용 경험에 대한 응답을 살펴보면 다음과 같다.

먼저 ‘원내 타 연구자 조사데이터에 대한 필요 또는 활용 경험’에 대한 문항에서는 응답자 47명 중 32명(68.1%)이 있다고 답하였으며 ‘타 기관 조사데이터 활용 경험’에 대한 문항에서는 응답자 45명 중 31명(68.9%)이 있다고 답하여 원내 타 연구자의 조사데이터와 타 기관의 조사데이터 경험률은 비슷한 수준으로 나타났다. 한편 ‘원내 타 연구자의 조사데이터 필요 또는 활용 경험’과 ‘타 기관의 조사데이터 활용 경험’이 모두 있는 경우는 23명으로 타 연구 조사데이터를 활용하는 경우는 원내외 관계 없이 다수가 활용하는 것을 알 수 있었다(그림 4-10 참조).

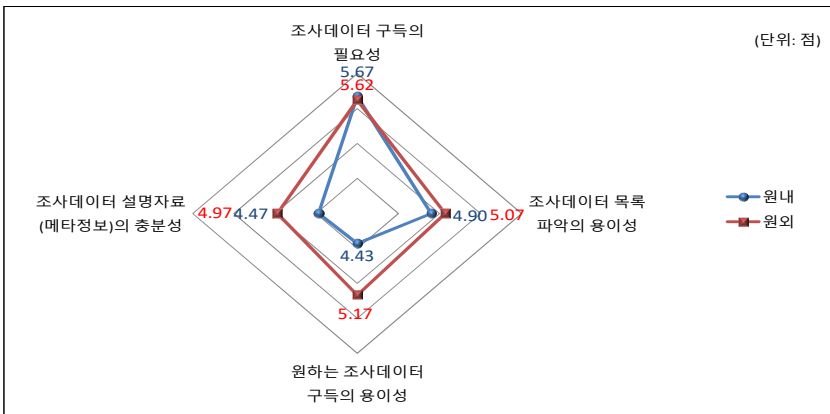
[그림 4-10] 한국보건사회연구원 연구책임자들의 원내 타 연구자의 조사데이터 필요 또는 활용 경험 및 타 기관의 조사데이터 활용 경험 여부



자료: 저자 작성.

다음으로 원내 타 연구자 조사데이터 필요 또는 활용 경험 또는 타 기관 조사데이터 활용 경험이 있다고 응답한 사람을 대상으로 해당 경험에 대한 ‘조사데이터 구득의 필요성’, ‘조사데이터 목록 파악의 용이성’, ‘원하는 조사데이터 구득의 용이성’, ‘조사데이터 설명자료(메타정보)의 충분성’을 7점 척도로 평가하게 하였다. 그 결과 응답자들은 원내·외 조사데이터 활용 경험에 대하여 ‘조사데이터 구득의 필요성’은 5.67점과 5.62점, ‘조사데이터 목록 파악의 용이성’은 4.90점과 5.07점, ‘원하는 조사데이터 구득의 용이성’은 4.43점과 5.17점, ‘조사데이터 설명 자료(메타정보)의 충분성’은 4.47점과 4.97점으로 평가한 것으로 나타났다. ‘조사데이터 구득의 필요성’과 ‘조사데이터 목록 파악의 용이성’에 대해서는 원내와 타 기관이 비슷한 수준이었으나 ‘원하는 조사데이터 구득의 용이성’과 ‘조사데이터 설명 자료(메타정보)의 충분성’에 대해서는 원내 보다는 타 기관에서 좀 더 높은 점수를 나타내, 상대적으로 원내 조사데이터 관리 체계 보완이 필요하다는 것을 알 수 있었다(그림 4-11) 참조).

[그림 4-11] 한국보건사회연구원 연구책임자들의 타 연구 조사데이터 필요 또는 활용 경험 평가



자료: 저자 작성.

한편 원내 타 연구자의 조사데이터를 활용한 경우, 해당 조사데이터명 결과를 분석해 본 결과 원내 보건복지데이터포털, 한국복지패널, 한국의료패널 홈페이지에서 제공하고 있는 공개용 데이터였으나 일부는 공개용 데이터 목록에 있지 않은 데이터로 이는 연구자들 간 데이터를 주고받아 연구에 활용하는 경우가 있는 것으로 파악할 수 있다. 사실상 외부 공개용 데이터 외에 내부 연구진 간에도 제한적 공개가 필요할 수 있다. 그러므로 향후 원외뿐 아니라 원내에서 공개하는 조사데이터의 범위에 대해서도 논의할 필요가 있다 하겠다.

또 원내 타 연구자의 조사데이터 활용 시 어려웠던 점에 대하여 주관식으로 응답한 내용을 살펴보면 본격적인 데이터 활용 이전 단계인 데이터 접근이나 구독 단계에서의 어려움, 조사데이터 활용 시 발주처나 조사데이터 공동 생산 기관 등 관계 기관과의 협의(또는 협조)와 관련된 어려움, 조사데이터의 분석 및 활용 단계에서의 어려움도 피력하고 있다(〈표 4-6〉 참조). 이러한 내용들은 모두 향후 조사데이터 관리 체계에서 반영이 필요할 것으로 생각된다.

〈표 4-6〉 원내 타 연구자의 조사데이터 활용 시 어려웠던 점

구분	응답 내용
데이터 접근 혹은 구독	원하는 자료 존재 여부 파악이 용이하지 않음
	일괄되게 보관하는 장소가 없어 누구에게, 어떻게 문의해야 하는지에 대해 막막함
	조사데이터에 접근하는 것이 어려움
	공개되지 않았으나 조사된 변수 요청 시 자료 구독 어려움
	데이터 이용 프로세스에 대한 정보를 찾기 어려웠고, 데이터를 탐색적으로 이용하는 경우 이용허가가 어려울 것으로 예상됨
	신청절차가 번거로움(회원가입, 연구계획서 검토 등)
	원자료를 제공받는 데 상당한 시간이 소요됨
	상당히 오래전 과거 데이터가 부족함

구분	응답 내용
관계 기관과의 협의(협조)	공동 생산 기관이 제공을 반대함
	발주처의 승인 절차가 비공식적이라 협의에 어려움
	발주처의 승인이 필요함.
데이터 분석 및 활용	시계열 자료, 변수 확인, 변수명 변경 이력 확인 등이 어려움
	데이터 설명이 불충분함(메타데이터 미비)
	실제 데이터 활용 시 정확성이 떨어짐
	비공개 데이터의 경우, 데이터 사용에 있어 향후 윤리적 문제에 대한 우려가 있음.
	원자료를 활용해 변수 구성이 어려움

자료: 저자 작성.

한편 타 기관의 조사데이터 활용 시 획득 경로에 대해서는 응답자 중 65.5%가 통계청 MDIS를, 응답자 중 37.9%가 해당 기관 자체 포털을, 응답자 중 10.3%가 한국사회과학자료원(KOSSDA) 혹은 해당 연구진에 직접 신청하는 것으로 나타났다. 이는 **향후 당 원의 조사데이터 창구로 통계청과 한국사회과학자료원을 적극 활용하는 것이 필요함을 시사하고 있다.** 활용해 본 타 기관의 조사데이터 목록으로는 가계동향조사, 가계 금융복지조사, 고령화연구패널조사, 국민노후보장패널, 한국노동패널조사 등 주로 주기적으로 생산되는 조사데이터, 패널데이터로 이러한 **조사데이터를 한곳에서 쉽게 접근 가능한 경로를 만들어주는 것도 필요하다** 하겠다. 타 기관 조사데이터 활용 시 어려움으로는 원내 타 연구자의 조사데이터 활용 시 어려움에 추가적으로 **데이터 공개 시기가 늦어져 시의성이 떨어지는 점, 원내 땅 분리로 인해 접속이 어려움** 등과 같은 내용이 포함되어 있었다.

이처럼 원내외 조사데이터를 불문하고 연구자들이 조사데이터 활용 시 겪는 어려움을 '5W2H'로 요약해 볼 수 있다.

- What: 이용 가능한 조사데이터는 무엇인가?

- Where: 조사데이터는 어디에서 구독할 수 있는가?
- Who : 조사데이터 문의는 누구에게 할 수 있는가?
- When : 조사데이터 공개 시기는 언제인가?
- Which: 어떤 항목을 이용 가능한가?
- How : 조사데이터 이용 절차는 어떻게 되는가?
- How Long: 조사데이터 구독까지 어느 정도 시간이 소요되는가?

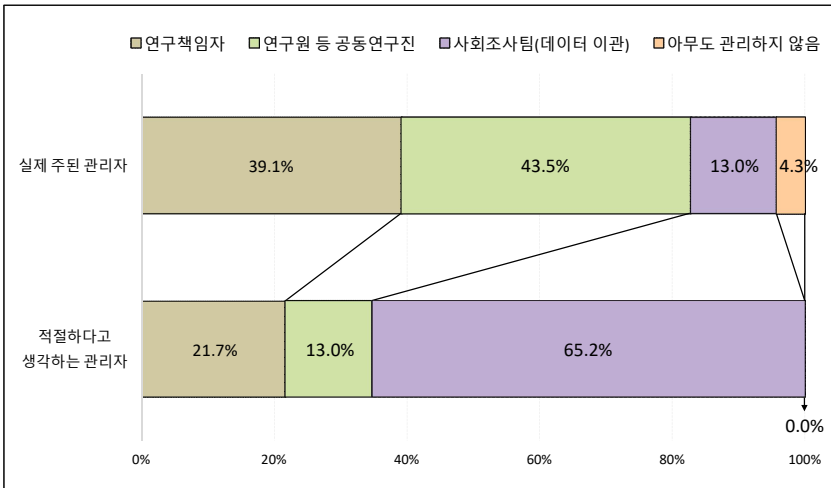
라. 조사데이터 생산 및 관리에 대한 경험과 인식

한국보건사회연구원 연구책임자 중 ‘최근 3년간 우리 연구원에서 연구책임자로 수행한 연구과제에서 조사데이터를 생산한 경험이 있다’고 응답한 연구책임자는 응답자 43명 중 24명(55.8%)으로 나타났다. 이들에게 연구 종료 후 조사데이터 보관 및 관리는 주로 누가 하고 있는지에 대해 물어본 결과 ‘연구원 등 공동연구진’(43.5%), ‘연구책임자’(39.1%), ‘사회조사팀(데이터 이관)’(13.0%) 순으로 나타났고, ‘아무도 관리하지 않음’이라는 응답도 4.3%를 나타냈다. 반면 누가 관리하는 것이 적절하다고 생각하는지에 대해서는 ‘사회조사팀(데이터 이관)’(65.2%), ‘연구책임자’(21.7%), ‘연구원 등 공동연구진’(13.0%) 순으로 나타나 **한국보건사회연구원 연구책임자들은 현재는 분산되어 관리하고 있으나 과반수 이상이 조사데이터를 전담팀에 이관하여 관리하는 필요성에 공감하고 있음을 알 수 있었다**([그림 4-12] 참조).

또 조사데이터 생산 경험이 있다고 응답한 연구자들을 대상으로 조사데이터 소실 경험 유무를 알아본 결과 2명만이 ‘연구 종료 후 조사데이터를 소실한 경험이 있다’고 응답했다. 조사데이터 소실의 원인(복수응답)으로는 ‘데이터 저장 장비의 고장에 따른 소실’, ‘바이러스, 해킹 등 외부

침입에 따른 데이터 소실’, ‘데이터 관리 부주의에 따른 데이터 삭제 또는 소실’을 응답하였다.

[그림 4-12] 한국보건사회연구원 연구책임자들의 조사데이터 관리자 실태와 인식



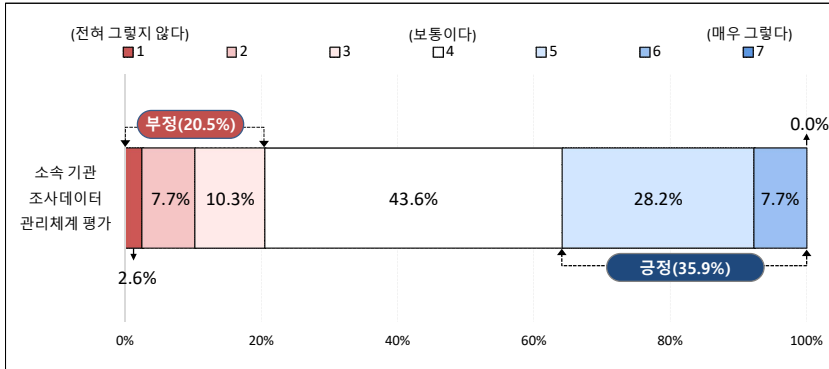
자료: 저자 작성.

마. 한국보건사회연구원 조사데이터 관리 체계에 대한 인식

한국보건사회연구원 연구책임자들의 당 원 조사데이터 관리 체계(조사 자료 관리 매뉴얼에 의해 조사자료를 이관, 관리, 개방) 운영에 대해 인식 정도를 7점 척도로 측정된 결과. 부정적 응답(①전혀 그렇지 않다+②그렇지 않다+③별로 그렇지 않다)이 20.5%, 긍정적 응답(⑦매우 그렇다+⑥그렇다+⑤약간 그렇다)이 35.9%, ‘보통이다’라는 응답은 43.6%로 나타났다. 점수는 4.10점으로, 한국보건사회연구원 연구책임자들은 당 원 조사데이터 관리 체계에 대해 보통의 수준으로 평가하고 있음을 알 수 있었다([그림 4-13] 참조).

94 전사적 연구데이터 관리 체계 마련을 위한 연구

[그림 4-13] 한국보건사회연구원 연구책임자들의 당 원 조사데이터 관리 체계 운영 평가



자료: 저자 작성.

다음으로 한국보건사회연구원 조사데이터 관리 체계에 있어 보완이 시급한 부분에 대해서는 6가지 항목에 대해 1순위부터 6순위까지의 순위형 질문으로 측정하였으며 순위별 가중치(7)를 부여하여 점수를 산출하였다. 응답자 39명에 대한 분석 결과 ‘조사데이터 관리(이관, 보관 등)를 위한 전담 인력 및 전담 조직(410.3점)’, ‘조사데이터 개방을 위한 데이터 품질 검토 인력 지원(407.7점)’, ‘조사데이터 관리에 대한 교육 등 공감대 형성(361.3점)’, ‘효과적인 자료 제공을 위한 정보시스템 개편(356.5점)’, ‘조사데이터 관리 지침 및 규정 개정(341.2점)’, ‘조사데이터 관리 및 제공 심의위원회 설치(223.0점)’ 순으로 나타났다(〈표 4-7〉 참조). 이에 따라 한국보건사회연구원 연구책임자들은 현 조사데이터 관리 체계에서 인력 및 조직 관련 체계 보완에 우선순위를 높게 인지하고 있음을 알 수 있었다.

한편 한국보건사회연구원 조사데이터 관리 체계와 관련하여 자유롭게 의견을 작성한 내용을 살펴보면 기준과 범위의 명확화, 전담 인력 및 조직, 안내·교육 및 홍보, 공감대 형성 등 앞의 제시 내용과 크게 다르지 않

7) 1순위 6점, 2순위 5점, 3순위 4점, 4순위 3점, 5순위 2점, 6순위 1점의 가중치를 부여하였음.

은 것으로 나타났다(〈표 4-8〉 참조). 이러한 내용들은 향후 당 원 관리체계 마련 시 개선안을 마련하여 반영하여야 할 것이다.

〈표 4-7〉 한국보건사회연구원 연구책임자들이 조사데이터 관리 체계에서 보완이 시급하다고 평가한 부분

구분	연구 책임자 수(명)	순위 비율(%)							가중치 (점)
		계	1순위	2순위	3순위	4순위	5순위	6순위	
조사데이터 관리(기관, 보관 등)를 위한 전담 인력 및 전담 조직	39	100.0	25.6	20.5	23.1	10.3	10.3	10.3	410.3
조사데이터 개방을 위한 데이터 품질 검토 인력 지원	39	100.0	20.5	23.1	20.5	20.5	10.3	5.1	407.7
효과적인 자료 제공을 위한 정보시스템 개편	39	100.0	7.7	23.1	28.2	12.8	15.4	12.8	356.5
조사데이터 관리 지침 및 규정 개정	39	100.0	15.4	15.4	10.3	25.6	20.5	12.8	341.2
조사데이터 관리 및 제공 심의위원회 설치	39	100.0	5.1	7.7	5.1	10.3	30.8	41.0	223.0
조사데이터 관리에 대한 교육 등 공감대 형성	39	100.0	25.6	10.3	12.8	20.5	12.8	17.9	361.3

자료: 저자 작성.

〈표 4-8〉 한국보건사회연구원 조사데이터 관리 체계에 관한 자유 기재 의견

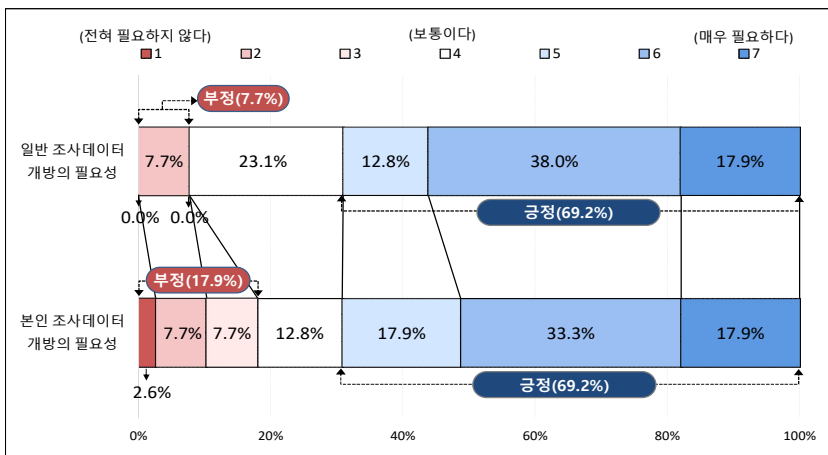
구분	응답 내용
관련 기준 및 범위의 명확화	공개 범위의 지침(기준 및 원칙)이 필요함.
	기관 범위의 지침(기준 및 원칙)이 필요함(질적 자료 포함 필요).
전담 인력 및 조직	연구진의 추가 업무 부담이 최소화되는 방안 강구가 필요함. (예로 작업파일 그대로 이관하여 전담팀에서 재정리하는 방안 등)
	전문, 전담 인력에 의한 일원화가 필요함.
안내·교육 및 홍보	우리 원 조사 관리 체계, 조사데이터 목록에 대한 원내외 적극적 홍보가 필요함.
	조사데이터 개방을 위한 필수 자료 생산에 대한 사전 안내가 필요함.
	연구계획서 작성, 조사표 개발 시부터 조사데이터 관리에 요구되는 사항에 대한 충분한 숙지가 필요함.
공감대 형성	조사데이터 목록 공개가 필요함.
	자발적 동참을 유도함.

자료: 저자 작성.

바. 조사데이터 개방에 대한 인식 및 경험

한국보건사회연구원 연구책임자들의 조사데이터 공유 및 개방 필요성에 대한 인식을 알아보기 위해 일반적인 조사데이터의 공유 및 개방의 필요성과 본인이 생산한 조사데이터의 공유 및 개방의 필요성에 대해 7점 척도로 측정하였다. ‘일반적인 조사데이터 공유 및 개방’과 ‘본인이 생산한 조사데이터 공유 및 개방’에 대해 긍정적인 응답(⑦매우 필요하다+⑥필요하다+⑤약간 필요하다)은 동일하게 69.2%로 나타났다. 그러나 부정적인 응답(①전혀 필요하지 않다+②필요하지 않다+③별로 필요하지 않다)에서는 ‘일반적인 조사데이터 공유 및 개방’이 7.7%, ‘본인이 생산한 조사데이터 공유 및 개방’은 17.9%로 나타나 전반적으로 한국보건사회연구원 연구책임자들은 조사데이터 공유 및 개방에 대해서는 과반수 이상이 공감하고는 있으나 본인이 생산한 데이터의 공유 및 개방에서 좀 더 부정적인 인식이 높게 나타났다(그림 4-14 참조).

[그림 4-14] 한국보건사회연구원 연구책임자들의 조사데이터 개방 필요성에 대한 인식



자료: 저자 작성.

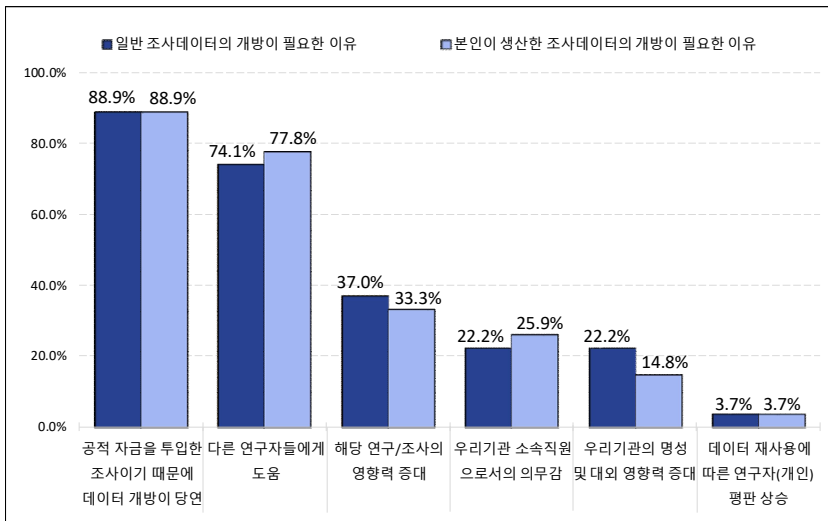
이러한 응답이 나온 이유를 알아보기 위해 부정적 응답자(①전혀 필요하지 않다+②필요하지 않다+③별로 필요하지 않다)들을 대상으로 복수 응답이 가능한 8지선다형으로 질문한 결과, ‘일반 조사데이터’의 경우에는 ‘데이터에 대한 오해석과 오용에 대한 우려(100.0%)’, ‘데이터 개방에 따른 개인정보 침해 우려(66.7%)’, ‘연구의 기밀성(33.3%)’, ‘논문 출판 등을 위한 독점 권한 상실(33.3%)’, ‘데이터 개방을 위한 시간과 노력에 대한 부담감(33.3%)’, ‘기관의 보상 체계 미흡(33.3%)’ 순으로 나타났다. 반면 ‘본인이 생산한 조사데이터’의 경우에는 ‘데이터에 대한 오해석과 오용에 대한 우려(100.0%)’, ‘데이터 개방에 따른 개인정보 침해 우려(80.0%)’, ‘데이터 개방을 위한 시간과 노력에 대한 부담감(80.0%)’, ‘연구의 기밀성(60.0%)’, ‘논문 출판 등을 위한 독점 권한 상실(40.0%)’, ‘표본의 대표성 결여에 대한 우려(40.0%)’, ‘기관의 보상체계 미흡(20.0%)’ 순으로 나타났다. 이에 따라 원내 연구책임자들의 조사데이터 개방을 활성화하기 위해서는 무엇보다도 ‘개인정보 보호’ 조치와 ‘데이터 개방을 위한 제반 노력’을 경감시켜 줄 수 있는 조치가 필요함을 알 수 있었다.

또한 앞선 문항에서 긍정적 응답자(⑦매우 필요하다+⑥필요하다+⑤약간 필요하다)를 대상으로 조사데이터 개방이 필요하다고 생각하는 이유에 대해 복수 응답이 가능한 6지선다형으로 질문하였다. 그 결과 ‘일반 조사데이터’의 경우에는 ‘공적 자금을 투입한 조사이기 때문에 데이터 개방이 당연하다(88.9%)’, ‘다른 연구자들에게 도움(74.1%)’, ‘해당 연구/조사의 영향력 증대(37.0%)’, ‘우리 기관 소속 직원으로서 의무감(22.2%)’, ‘우리 기관의 명성, 대외 영향력 증대(22.2%)’, ‘데이터 재사용에 따른 연구자(개인) 평판 상승(3.7%)’ 순으로 나타났다. ‘본인이 생산한 조사데이터’의 경우에는 ‘공적 자금을 투입한 조사이기 때문에 데이터 개방이 당연하다(88.9%)’, ‘다른 연구자들에게 도움(77.8%)’, ‘해당 연구/조사의 영향력 증대(33.3%)’, ‘우리 기관 소속 직원으로서 의무감

(25.9%)’, ‘우리 기관의 명성 및 대외 영향력 증대(14.8%)’, ‘데이터 재사용에 따른 연구자(개인) 평판 상승(3.7%)’ 등 순서는 동일하게 나타났다. 이를 통해 조사데이터를 개방한다는 것에 대해 원내 연구책임자들은 **공적 자금 투입에 대한 의무와 타 연구자에 대한 기여에 대하여 가장 크게 인식하고 있음을** 알 수 있었다([그림 4-15] 참조).

다음으로 응답자 37명 중 14명(37.8%)이 조사데이터 공유 및 개방 경험이 있다고 답하였으며, 이들의 조사데이터 공유 및 개방 시 어려웠던 점에 대한 주관식 문항에서는 **추가 업무 부담, 기준의 모호, 그리고 발주처와의 협의** 등과 같은 점을 제시하고 있어(〈표 4-9〉 참조) 이는 관리 체계 마련 시 **방안 수립이 필요함**을 알 수 있었다.

[그림 4-15] 한국보건사회연구원 연구책임자들이 조사데이터 개방이 필요하다고 생각하는 이유(복수)



자료: 저자 작성.

〈표 4-9〉 조사데이터 공유 및 개방 시 어려웠던 점

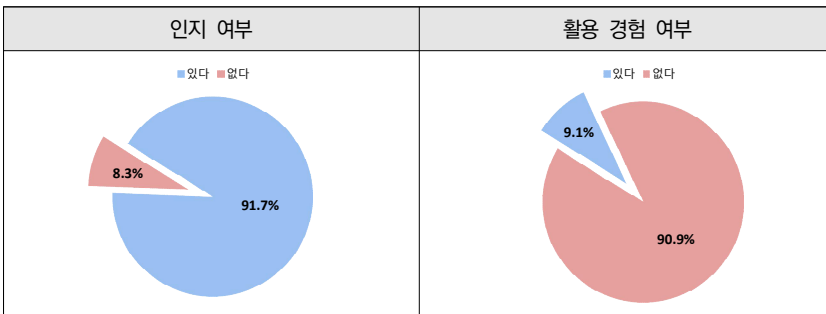
구분	응답 내용
추가 업무 부담	연구 목적의 데이터, 코딩북과 다른 공개용 데이터, 코딩북을 별도 작업해야 하는 부담이 있었음.
	개인정보 삭제 등
	레이아웃 작성
	제3자 사용을 위한 데이터셋 정리
	사후 확인 등 추가 서비스 요구의 부담
	연구 종료 후에 이루어지는 업무이므로 추가 업무 부담으로 느껴짐. 공개는 종료 후가 되더라도 분석 활용 과정에서 조사팀과 공유하는 방안도 검토 필요
기준 (원칙, 범위, 방침 등) 모호	공유·개방을 위한 내부적인 방침/지침에 대한 지식 및 인지 부족,
	개방 범위 설정의 어려움
	개방의 원칙 모호
발주처 협의	발주 부처와의 협의
	저작권이 외부 기관에 있는 수탁 과제에 경우 데이터 공유와 개방 권한이 없으므로 결정할 수 없음.
데이터 신뢰성 문제	데이터 정확성
	신뢰 높은 조사데이터 공유 및 개방을 위해 요구되는 지식 부족, 생산 자료에 대한 개방에 따른 위험성에 관한 상식 부족
데이터 문의 대응 문제	제공 과정에서는 큰 어려움 없었음. 단, 과거 조사 결과(책임 연구자가 바뀐 경우)에 대한 외부 문의 발생 시, 대응에 어려움을 경험함(예: 그 당시 자료 제공 범위에 대한 판단 근거 등).
	개별적인 상세한 조사 문의 답변 작성에 대한 시간 노력 소모가 많음.
연구 윤리 위반	일반 과제에서 수행된 조사로, 차년도 연구에서 이를 활용하였으나 인용표기를 하지 않았음.

자료: 저자 작성.

마지막으로 보건복지데이터포털을 통한 조사데이터 개방에 관한 응답자들의 인지 정도와 활용 경험에 대해서는 ‘알고 있다’고 응답한 연구책임자는 응답자 36명 중 33명(91.7%)으로 대부분 알고 있었으며, 이를 ‘활용’한 연구책임자는 응답자 33명 중 3명(9.1%)으로 매우 낮게 나타났다(그림 4-16) 참조). 이처럼 활용률이 낮은 이유에 대해서는 원하는 조사데이터가 없어서인지, 사용이 불편해서인지 파악하지 못하였으나 앞의

조사에서 응답자의 68.1%가 원내 타 연구자의 조사데이터 필요 또는 활용 경험이 있다고 한 것과 비교해 봐도 매우 낮은 결과로 앞으로 보건복지데이터포털에 대해 원내외 이용자들의 관심을 이끌어내고 활성화를 위한 고민이 필요하다 하겠다.

[그림 4-16] 한국보건사회연구원 연구책임자들의 보건복지데이터포털 인지 및 활용 경험 여부



자료: 저자 작성.

3. 경제인문사회연구회 소속 정부출연연구기관 연구책임자 대상 데이터 관리, 공유, 활용에 관한 인식 및 실태

가. 응답자 특성

경제인문사회연구회 소속 정부출연연구기관 응답자 182명에 대한 특성을 살펴보면 남성이 64.8%, 여성이 35.2%로 나타났으며, 연령별로는 40~49세(48.4%), 30~39세(28.6%), 50~59세(19.2%), 30세 미만(2.2%), 60세 이상(1.6%) 순으로 나타났다. 소속 기관에서 연구책임자로 근무한 경력 기간은 5년 미만(46.7%), 5년 이상 10년 미만(23.6%), 10년 이상 15년 미만(12.6%), 15년 이상 20년 미만(11.0%), 20년 이상(6.0%) 순으로 나타났다. 한편 연구 수행에 있어 조사 연구 방법을 얼마나 중요하게

생각하는지에 대해 7점 만점에 5.84점 나타나 연구회 소속 연구기관 응답자들이 조사 연구 방법을 중요하게 인식하고 있는 것으로 나타났다(〈표 4-10〉 참조). 조사 연구 방법에 대한 중요성 인식 정도는 한국보건사회연구원 연구책임자들은 이보다는 다소 높은 평균 6.2점을 나타내 차이를 보이고 있다.

〈표 4-10〉 경제인문사회연구회 소속 기관 연구책임자 응답자 현황

구분		연구책임자 수(명)	비율(%)
계		182	100.0
성별	남자	118	64.8
	여자	64	35.2
연령	30세 미만	4	2.2
	30~39세	52	28.6
	40~49세	88	48.4
	50~59세	35	19.2
	60세 이상	3	1.6
소속 기관에서의 연구책임자 경력 기간	5년 미만	85	46.7
	5년~10년 미만	43	23.6
	10년 이상	54	29.7
	- 10년~15년 미만	23	12.6
	- 15년~20년 미만	20	11.0
	- 20년 이상	11	6.0
조사 연구 방법 중요성 인식 정도	부정(①+②+③)	4	2.2
	보통(④)	21	11.5
	긍정(⑤+⑥+⑦)	157	86.3

자료: 저자 작성.

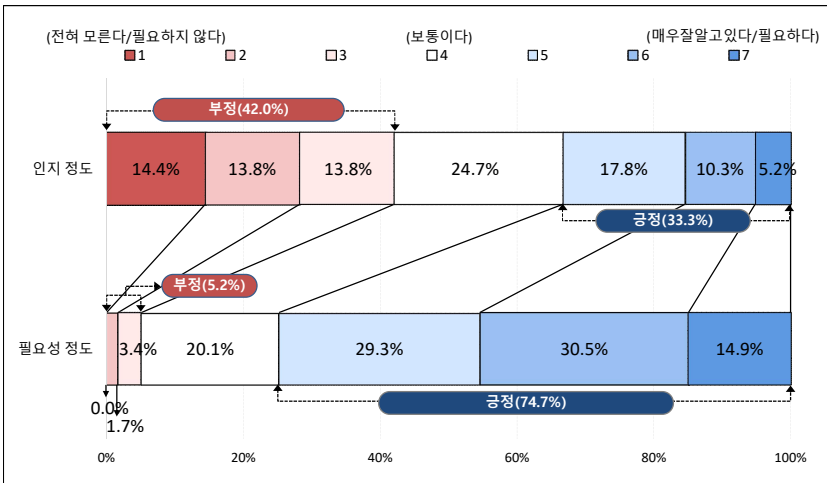
나. 연구데이터 관련 개념에 대한 인식

응답자 174명의 오픈 사이언스에 대한 개념과 동향 인지 정도, 오픈 사이언스 정책 도입의 필요성 인식 정도를 7점 척도로 측정하였다. ‘오픈 사

이언스의 개념과 동향'에 대해 긍정적 응답(⑦매우 잘 알고 있다+⑥알고 있다+⑤약간 알고 있다)은 33.3%, 부정적 응답(①전혀 모른다+②모른다+③별로 모른다)은 42.0%로 긍정보다는 부정적 응답이 높게 나타나 긍정적 응답이 상대적으로 높게 나타난 한국보건사회연구원 연구책임자들을 대상으로 한 결과와는 상이한 결과를 보여주고 있다.

'본인의 연구 분야에서 오픈 사이언스 정책 도입의 필요성'에 대해서는 긍정적 응답(⑦매우 필요하다+⑥필요하다+⑤약간 필요하다)은 74.7%로 부정적 응답(①전혀 필요하지 않다+②필요하지 않다+③별로 필요하지 않다) 5.2%에 비해 매우 높게 나타났다. 즉 연구회 소속 연구기관 연구책임자들은 현재는 오픈 사이언스 개념에 대하여 인지하지 못하고 있더라도 그 필요성에 대해서는 많이 공감하고 있다는 것을 알 수 있었으며([그림 4-17 참조], 이는 한국보건사회연구원 연구책임자들을 대상으로 한 연구 결과와도 동일한 결과를 보여주고 있다.

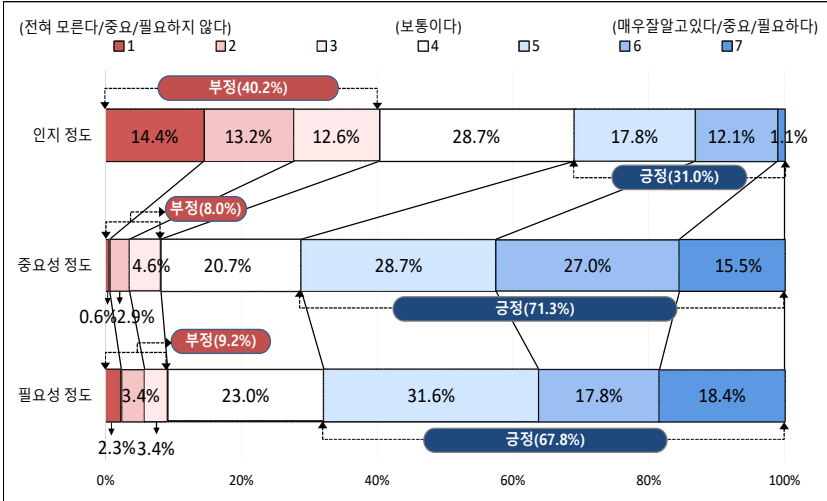
[그림 4-17] 경제인문사회연구회 소속 기관 연구책임자들의 오픈 사이언스에 대한 인식



자료: 저자 작성.

응답자들의 연구데이터 관리 계획에 대한 인식을 알아보기 위해 연구데이터 관리 계획 개념과 동향 인지 정도, 연구데이터 관리 계획 정책의 중요성 정도, 정책 도입의 필요성 정도를 7점 척도로 측정하였다. ‘연구데이터 관리 계획의 개념과 동향’에 대해 긍정적 응답(⑦매우 잘 알고 있다+⑥알고 있다+⑤약간 알고 있다)은 31.0%로 부정적 응답(①전혀 모른다+②모른다+③별로 모른다) 40.2%에 비해 다소 낮은 수준으로 나타났다. ‘연구데이터 관리 계획 정책의 중요성’과 ‘연구데이터 관리 계획 정책 도입의 필요성’에 대해서는 긍정적 응답(⑦매우 중요하다+⑥중요하다+⑤약간 중요하다)이 각각 71.3%, 67.8%이고 부정적 응답(①전혀 모른다+②모른다+③별로 모른다)이 각각 8.0%, 9.2%인 데 비해 매우 높게 나타났다. 이는 경제인문사회연구회 소속 기관 연구책임자들이 현재는 연구데이터 관리 계획 개념에 대해 잘 모르고 있다 하더라도 관련 정책의 중요성과 도입의 필요성은 상대적으로 높게 인식하고 있다는 것을 나타낸다([그림 4-18] 참조). 한국보건사회연구원 연구책임자 대상으로 한 분석 결과에서도 동일한 결과를 보이고 있다.

[그림 4-18] 경제인문사회연구회 소속 기관 연구책임자들의 연구데이터 관리 계획에 대한 인식



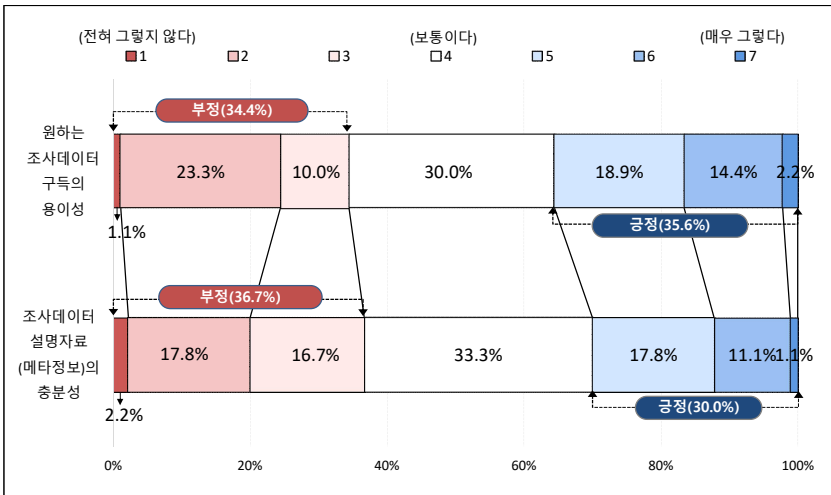
자료: 저자 작성.

다. 타 연구 조사데이터 활용 경험

경제인문사회연구회 소속 기관 연구책임자들에 대해 타 연구자의 조사 데이터 활용 경험을 물어본 결과, 응답자 174명 중 97명(55.7%)이 있다고 답하였다. 타 연구자 조사데이터 활용 경험이 있다고 응답한 사람을 대상으로 ‘원하는 조사데이터 구득의 용이성’, ‘조사데이터 설정자료(메타정보)의 충분성’을 7점 척도로 측정한 결과, 긍정적 응답(⑦매우 그렇다+⑥그렇다+⑤그렇지 않다)이 35.6%와 30.0%, 부정적 응답(①전혀 그렇지 않다+②그렇지 않다+③별로 그렇지 않다)은 각각 34.4%와 36.7%로 거의 유사하게 나타났으며 측정 평균점수는 각각 3.94점과 3.84점으로 나타났다. 따라서 연구회 소속 연구기관들의 연구책임자들도 타 연구자 조사데이터 활용 경험에 대해 ‘조사데이터 구득의 용이성’과 ‘조사데이터

의 충분성'에 대해 긍정도 부정도 아닌 보통 수준으로 평가하고 있다는 것을 알 수 있었다([그림 4-19] 참조). 한국보건사회연구원 연구책임자들을 대상으로 한 분석 결과보다는 다소 낮은 결과를 보이고 있다.

[그림 4-19] 경제인문사회연구회 소속 기관 연구책임자들의 타 연구자 조사데이터 활용 경험 평가



자료: 저자 작성.

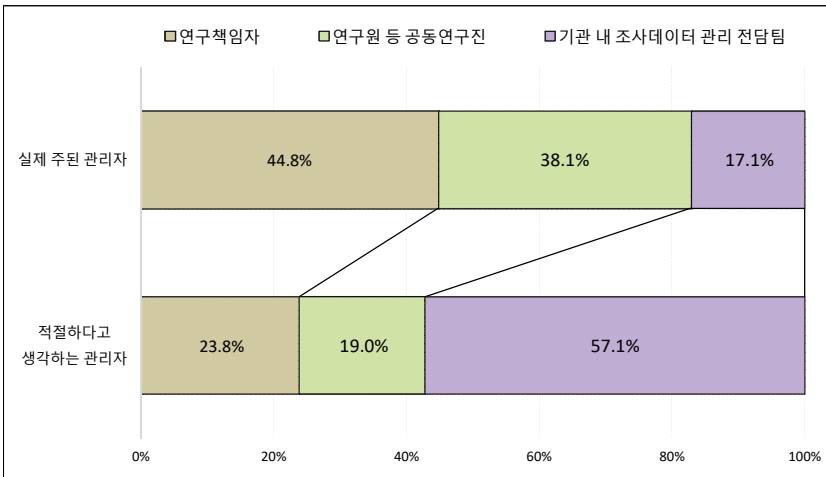
라. 조사데이터 생산 및 관리에 대한 경험과 인식

경제인문사회연구회 소속 기관 연구책임자들 중 ‘최근 3년간 소속 기관에서 연구책임자로 수행한 연구과제에서 조사데이터를 생산한 경험이 있다’고 응답한 연구책임자는 응답자 165명 중 106명(64.2%)으로, 이는 한국보건사회연구원 연구책임자 대상으로 한 분석 결과 55.8%보다는 다소 높은 것이다.

다음으로 이들에게 연구 종료 후 조사데이터 보관 및 관리는 주로 누가

하고 있는지에 대해 물어본 결과, ‘연구책임자’(44.8%), ‘연구원 등 공동 연구진’(38.1%), ‘기관 내 조사데이터 관리 전담팀’(17.1%) 순으로 나타났으며, 이는 한국보건사회연구원 연구책임자 대상으로 한 분석 결과와는 연구책임자와 연구원 등 공동 연구진의 순위가 바뀐 차이를 보이고 있다. 반면 누가 관리하는 것이 적절하다고 생각하는지에 대해서는 ‘기관 내 조사데이터 관리 전담팀’(57.1%), ‘연구책임자’(23.8%), ‘연구원 등 공동연구진’(19.0%) 순으로 나타났으며 한국보건사회연구원 연구책임자를 대상으로 한 분석 결과와 마찬가지로 조사데이터 관리를 위한 전담팀의 필요성에 동의하고 있음을 알 수 있었다([그림 4-20] 참조).

[그림 4-20] 경제인문사회연구회 소속 기관 연구책임자들의 조사데이터 관리자 실태와 인식



자료: 저자 작성.

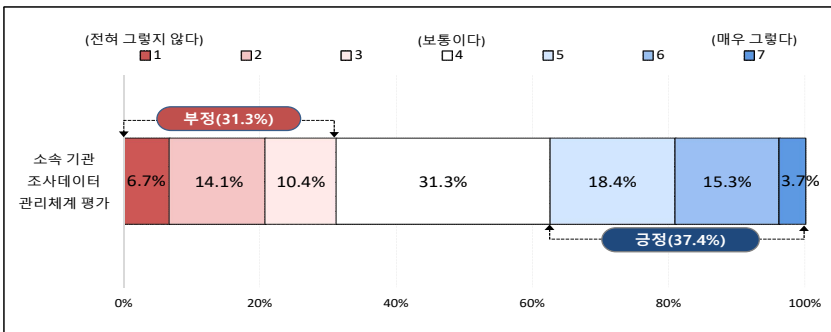
다음으로 최근 3년간 조사데이터 생산 경험자 중 조사데이터 소실 경험자는 응답자 105명 중 15명(14.3%)으로 나타났다. 조사데이터 소실 원인으로서는 ‘데이터 관리 부주의에 따른 데이터 삭제 또는 소실’(66.7%),

‘데이터 저장 장비의 고장에 따른 소실(33.3%)’, ‘바이러스, 해킹 등 외부 침입에 따른 데이터 소실(6.7%)’, ‘데이터 저장 장비 자체의 분실(6.7%) 순으로 나타났다.

마. 소속 기관 조사데이터 관리 체계에 대한 인식

경제인문사회연구회 소속 기관 조사데이터 관리 체계에 대한 해당 기관 연구책임자들의 인식 정도를 7점 척도로 측정된 결과, 긍정적 응답(⑦매우 그렇다+⑥그렇다+⑤약간 그렇다)이 37.4%, 부정적 응답(①전혀 그렇지 않다+②그렇지 않다+③별로 그렇지 않다)이 31.3%이며 평점은 4.01점으로 연구회 소속 연구기관의 연구책임자들은 보통 수준으로 평가하고 있음을 알 수 있었다([그림 4-21] 참조). 한국보건사회연구원 연구책임자들을 대상으로 한 분석 결과에서도 유사한 결과를 보여주고 있다.

[그림 4-21] 경제인문사회연구회 소속 기관 연구책임자들의 소속 기관 조사데이터 관리 체계 운영 평가

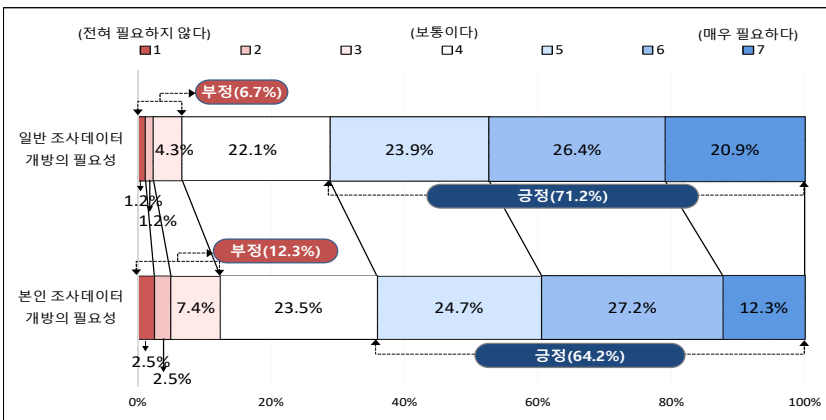


자료: 저자 작성.

바. 조사데이터 개방에 대한 인식

경제인문사회연구회 소속 기관 연구책임자들의 조사데이터 공유 및 개방 필요성에 대한 인식을 알아보기 위해 일반적인 조사데이터와 본인이 생산한 조사데이터로 나누어 각각 7점 척도로 측정하였다. ‘일반적인 조사데이터 공유 및 개방’에 대해서는 긍정적 응답(⑦매우 필요하다+⑥필요하다+⑤약간 필요하다)은 71.2%로 나타났으나 ‘본인이 생산한 조사데이터 공유 및 개방’에 대한 긍정적 응답(⑦매우 필요하다+⑥필요하다+⑤약간 필요하다)은 64.2%로 상대적으로 낮게 나타났다. 즉 연구회 소속 연구기관들의 응답자들은 일반적인 조사데이터 개방의 필요성에 대해서는 다소 높게 평가하는 반면, 본인이 생산한 조사데이터 개방의 필요성에 대해서는 상대적으로 낮게 평가하고 있다는 것을 알 수 있었다(그림 4-22) 참조). 이는 한국보건사회연구원 연구책임자들이 답한 결과와도 동일하다.

[그림 4-22] 경제인문사회연구회 소속 기관 연구책임자들의 조사데이터 개방 필요성에 대한 인식



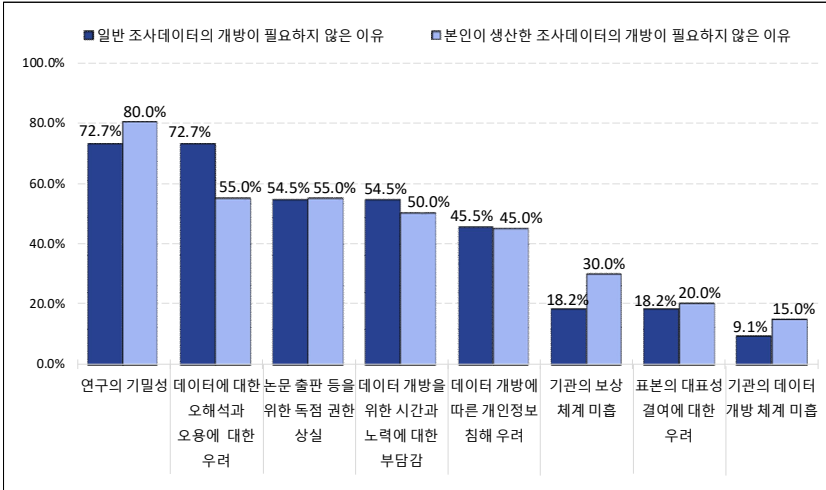
자료: 저자 작성.

앞선 문항에 대한 이유를 알아보기 위해 부정적 응답자((①전혀 필요하지 않다+②필요하지 않다+③별로 필요하지 않다)들을 대상으로 복수 응답이 가능한 8지선다형으로 질문한 결과, 일반 조사데이터의 경우에는 ‘연구의 기밀성’(72.7%)과 ‘데이터에 대한 오해석과 오용에 대한 우려’(72.7%)를 가장 많이 선택하였으며 ‘논문 출판 등을 위한 독점 권한 상실(54.5%)’, ‘데이터 개방을 위한 시간과 노력에 대한 부담감(54.5%)’, ‘데이터 개방에 따른 개인정보 침해 우려(45.5%)’, ‘표본의 대표성 결여에 대한 우려(18.2%)’, ‘기관의 보상체계 미흡(18.2%)’, ‘기관의 데이터 개방 체계 미흡(9.1%)’ 순으로 나타났다.

본인이 생산한 조사데이터의 경우에는 ‘연구의 기밀성(80.0%)’, ‘논문 출판 등을 위한 독점 권한 상실(55.0%)’, ‘데이터에 대한 오해석과 오용에 대한 우려(55.0%)’, ‘데이터 개방을 위한 시간과 노력에 대한 부담감(50.0%)’, ‘데이터 개방에 따른 개인정보 침해 우려(45.0%)’, ‘기관의 보상체계 미흡(30.0%)’, ‘표본의 대표성 결여에 대한 우려(20.0%)’, ‘기관의 데이터 개방 체계 미흡(15.0%)’ 순으로 나타났다([그림 4-23] 참조), 즉 연구회 소속 연구기관들의 연구책임자들은 **연구의 기밀성, 데이터에 대한 오해석과 오용에 대한 우려 등을 이유로 조사데이터 개방을 부정적으로** 생각하고 있었다.

앞선 문항에 대한 응답의 이유를 알아보기 위해 ‘조사데이터의 공유 및 개방이 필요하다’고 응답한 연구책임자와 ‘본인이 생산한 조사데이터의 공유 및 개방이 필요하다’고 응답한 연구책임자를 대상으로 그 이유를 복수 응답이 가능한 6지선다형 질문으로 측정하였다. ‘조사데이터의 공유 및 개방이 필요하다’고 응답한 연구책임자 116명 중 무응답 1명을 제외한 115명과 ‘본인이 생산한 조사데이터의 공유 및 개방이 필요하다’고 응답한 연구책임자 104명은 그 이유를 답하였다.

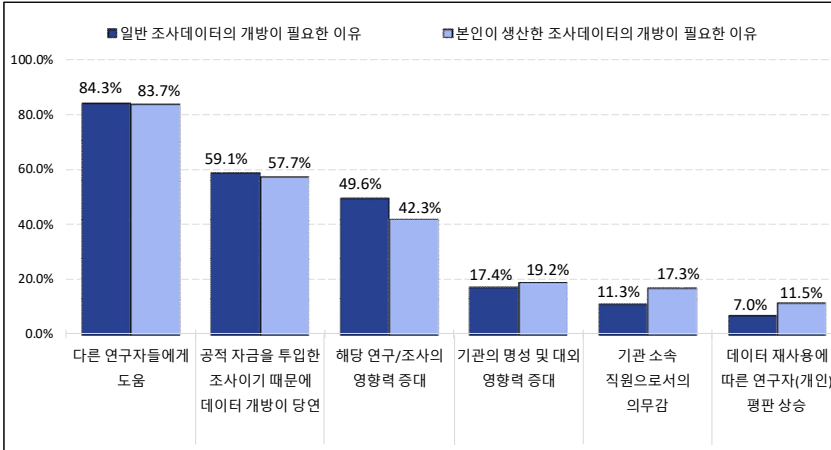
[그림 4-23] 경제인문사회연구회 소속 기관 연구책임자들이 조사데이터 개방이 필요하지 않다고 생각하는 이유(복수 응답)



자료: 저자 작성.

그 결과 ‘조사데이터의 공유 및 개방이 필요하다’고 응답한 연구책임자와 ‘본인이 생산한 조사데이터의 공유 및 개방이 필요하다’고 응답한 연구책임자 모두 ‘다른 연구자들에게 도움’이 되기 때문이라는 응답이 각각 84.3%, 83.7%로 가장 높았고, 그 다음으로는 ‘공적 자금을 투입한 조사이기 때문에 데이터 개방이 당연(59.1%, 57.7%)’, ‘해당 연구/조사의 영향력 증대(49.6%, 42.3%)’, ‘기관의 명성 및 대외 영향력 증대(17.4%, 19.2%)’, ‘기관 소속 직원으로서 의무감(11.3%, 17.3%)’, ‘데이터 재사용에 따른 연구자(개인) 평판 상승(7.0%, 11.5%)’ 순으로 나타났다. 이에 따라 개방의 필요성에 공감하는 연구책임자들은 타 연구자에 대한 기여와 공적 자금의 적절한 사용에 대하여 가장 크게 인식하고 있음을 알 수 있었다(그림 4-24) 참조). 이는 한국보건사회연구원 연구책임자들을 대상으로 한 연구 결과와도 동일한 내용을 보여주고 있다.

[그림 4-24] 경제인문사회연구회 소속 기관 연구책임자들이 조사데이터 개방이 필요하다고 생각하는 이유(복수 응답)



자료: 저자 작성.

다음으로 경제인문사회연구회 소속 기관 연구책임자들의 조사데이터 개방을 위해 필요한 부분에 관한 평가를 알아보기 위해 6가지 항목에 대해 1순위부터 6순위까지의 순위형 질문으로 측정하였다(〈표 4-11〉 참조). 응답에 대해 순위별로 가중치⁸⁾를 부여하여 순위별 점수를 합산한 결과, ‘조사데이터 관리(이관, 보관 등)를 위한 전담 인력 및 전담 조직’의 점수가 424.1점으로 가장 높게 나타났고, 이어서 ‘조사데이터 관리에 대한 교육 등 공감대 형성(416.7점)’, ‘효과적인 자료 제공을 위한 정보시스템 마련(389.5점)’, ‘조사데이터 개방을 위한 데이터 품질 검토 인력 지원(346.9점)’, ‘조사데이터 관리 지침 및 규정 제·개정(339.5점)’, ‘조사데이터 관리 및 제공 심의위원회 설치(183.3점)’ 순으로 나타났다. 이에 따라 연구회 소속 연구기관들의 연구책임자들은 조사데이터 개방을 위하여

8) 1순위 6점, 2순위 5점, 3순위 4점, 4순위 3점, 5순위 2점, 6순위 1점의 가중치를 부여하였음.

가장 우선적으로 인력 및 조직 체계 보완이 필요하다고 생각할 뿐 아니라 관련 인식 개선을 시급하게 생각하는 것을 알 수 있었다.

〈표 4-11〉 경제인문사회연구회 소속 기관 연구책임자들의 조사데이터 개방을 위해 필요한 부분 평가

(단위: 점)

구분	사례수	1순위	2순위	3순위	4순위	5순위	6순위	계
조사데이터 관리(이관, 보관 등)를 위한 전담 인력 및 전담 조직	162	24.7	24.1	21.0	16.0	9.3	4.9	424.1
조사데이터 개방을 위한 데이터 품질 검토 인력 지원	162	6.2	22.8	21.6	21.6	16.7	11.1	346.9
효과적인 자료 제공을 위한 정보시스템 마련	162	13.0	24.7	25.3	19.8	10.5	6.8	389.5
조사데이터 관리 지침 및 규정 제·개정	162	10.5	18.5	15.4	19.8	27.2	8.6	339.5
조사데이터 관리 및 제공 심의위원회 설치	162	1.9	0.6	7.4	11.7	25.9	52.5	183.3
조사데이터 관리에 대한 교육 등 공감대 형성	162	43.8	9.3	9.3	11.1	10.5	16.0	416.7

주: 해당 문항 무응답 수는 20명임.
자료: 저자 작성.

4. 연구책임자 주관식 응답 결과 분석

이상의 객관식 문항 응답 결과와 관련하여 주관식 문항에 대한 응답 결과를 살펴보면 다음과 같다.

먼저 경제인문사회연구회 소속 연구기관(이하 원외) 연구책임자를 대상으로 한 ‘조사데이터 관리, 공유, 활용에 관한 인식 및 실태조사’ 응답자 중 97명이 ‘타 연구자의 조사데이터를 활용했던 경험이 있다’고 응답하였다. 97명의 응답자 중 90명이 해당 조사데이터명에 대한 주관식 문항에 답하였다.

중복 응답을 고려하여 획득 경로별 응답 건수를 살펴보면 해당 연구진(혹은 기관)에 직접 신청(37.4%), 통계청 MDIS(25.2%) 그리고 해당 기관 자체 데이터포털(20.9%) 순으로 높은 것으로 나타났다(〈표 4-12〉 참조).

〈표 4-12〉 타 연구자의 조사데이터 활용 시 획득 경로-원외

(단위: 건, %)

구분	경로명	응답 수*	
		건수	비중
1	해당 연구진(혹은 기관)에 직접 신청	52	37.4
2	통계청 MDIS	35	25.2
3	해당 기관 자체 데이터 포털	29	20.9
4	공공데이터포털 등 통계포털	8	5.8
5	한국사회과학자료원(KOSSDA)	5	3.6
6	유관 기관 또는 협회의 데이터 신청	3	2.2
7	기타(정부기관에 직접 신청, 개방 자료 활용, NKIS 사이트 활용 등)	7	4.9
계		139	100

*: 중복 기재함.
자료: 저자 작성.

응답 수가 높은 상위 3개의 획득 경로를 살펴보면 원내 조사결과와는 확연히 다르게 **해당 연구진(혹은 기관)에 직접 신청하는 비율이 가장 높게 나타났다**. 하지만 이외에는 통계청 MDIS 사이트와 해당 기관의 자체 데이터포털 사이트 이용 빈도가 높게 나타났으며, 이러한 원외 조사 응답 결과는 원내 응답 조사 결과와 마찬가지로 시사하는 바가 동일하다.

외부에서 연구자 개인에게 조사데이터 활용에 관한 요청을 할 경우 연구자 개인이 데이터를 주고받거나 요청을 거절하는 등의 응대를 하는 것은 연구자 개인에게 부담이 될 수 있어 바람직하지 않다. 우리 원내에서 개인 차원이 아닌 전사적 차원의 합리적인 체계(조사데이터 수집, 이관

및 공개에 관한 체계)를 마련하여 신뢰성 있게 조사데이터를 제공하고 우리 원의 연구 결과물을 확산하고 홍보할 수 있는 장을 마련해야 한다.

따라서 우리 원에서 생산하는 조사데이터의 잠재적 이용자를 위해 통계청 MDIS와 보건복지데이터포털 모두를 조사데이터 공개 창구로 적극 활용하는 것이 필요함을 시사한다. 특히 보건복지데이터포털은 데이터 제공의 중심 창구 역할을 부여하고 이에 관한 적극적인 홍보가 필요함을 시사한다.

타 연구자의 조사데이터를 활용해 본 경험이 있는 97명 중 활용 시에 어려움을 겪었다고 답한 응답자 수는 68명(70.1%)으로 원내 조사결과보다는 높게 나타났다. 타 연구자의 조사데이터 활용 시에 어려움을 겪었다고 답한 68명의 응답 내용을 살펴보면 <표 4-13>과 같다.

원의 응답자들은 타 연구자의 조사데이터 활용 시에 다양한 어려움을 겪은 것으로 나타났으며 데이터 접근이나 구득의 어려움, 비공개 변수 등의 활용 제약 및 데이터 비공개, 데이터 활용 그리고 메타데이터 미비에 관한 응답 빈도가 높게 나타났다. 데이터 접근이나 구득의 어려움과 비공개 변수 등의 활용 제약 및 데이터 비공개에 관한 어려움의 내용은 원내·외 응답자들 모두 유사하게 나타났다.

<표 4-13> 타 연구자의 조사데이터 활용 시 어려웠던 점-원외

구분	응답 내용
데이터 접근성 구득 어려움 (20)	자료 자체를 얻기 어려움.
	원자료 구득의 어려움.
	데이터 구득 및 활용
	접근성
	데이터 접근성 한계
	비협조
	원자료 구득 필요
필요한 데이터가 분산되어 있어 누가, 어떤 자료를 가지고 있는지 파악하기 어려움.	

구분	응답 내용
	<p>또한 친분이 어느 정도 있는 관계가 아니라면 제공받기 어려움.</p> <p>데이터 보유 연구자를 알지 못하면, 구하기 어려웠을 것으로 예상됨.</p> <p>원자료 접근이 어려움. 기관 협조가 쉽지 않음.</p> <p>원하는 데이터를 받기가 어려움.</p> <p>비공식 요청에 대한 부담감</p> <p>요청하는 것 자체가 어려움.</p> <p>설문조사의 경우 원자료 취득에 어려움.</p> <p>자의적인 기준에 의한 데이터 공유</p> <p>자료 습득이 어려우며,</p> <p>조사데이터 확보가 어렵고,</p> <p>구득이 어려움.</p> <p>데이터에 대한 접근에 한계가 있을 수밖에 없어 협조가 필요함.</p>
데이터 활용 (15)	<p>연구에 접목시켜 활용하려다 보니 시간이 많이 소요되었음.</p> <p>자료 상세도(공간적 세분 등), 조사 범위 등이 부족하여 제한적인 활용 등 재가공</p> <p>연구 목적의 불일치성, 조사 대상의 상이성 등</p> <p>내가 직접 다루기 어려웠음.</p> <p>원하는 방식으로 데이터 가공, 활용 시 추가적인 노력이 필요했음.</p> <p>조사 목적이 다소 상이함에 따른 활용 등</p> <p>조사 목적과 연구 목적의 상이성으로 인한 적용 가능성</p> <p>원자료가 아닌 집계데이터 사용으로 목적에 맞게 사용하기 곤란한 경우 있음.</p> <p>본인 연구의 맞춤형 데이터 부족</p> <p>특정 연구를 위한 조사자료를 일반화하여 본인의 과제에 적용하기 위해서는 추가 가공이 필요함.</p> <p>조사데이터가 나의 연구 설계에 적합하지 않은 부분이 많아서 활용에 한계가 있었음.</p> <p>조사데이터를 연구 목적에 맞게 재가공하는 데 시간과 노력이 필요</p> <p>가공된 자료라서 내가 원하는 원시자료 사용이 어려움.</p> <p>조사 목적이 애초에 다르므로 충분한 정보는 얻기 힘들.</p>
메타데이터 미비 (9)	<p>메타데이터가 부실했음.</p> <p>설문조사의 경우 조사지 같은 기초적인 조사 원문을 찾기 어려웠음.</p> <p>메타정보가 제공되고는 있으나 구체적인 표본 설계, 조사 방법 등을 파악하기 어려운 점이 있음.</p> <p>데이터 컬럼 등의 정의가 없어 데이터의 의미를 찾기 어려움.</p> <p>데이터 구조에 대한 설명 불충분</p>

116 전사적 연구데이터 관리 체계 마련을 위한 연구

구분	응답 내용
	조사데이터의 설명 자료가 충분하지 않음.
	코딩북 결여
	변수명 불확실
	조사데이터에 대한 자세한 메타데이터 부재
메타데이터 미비 (조사 설계 등 설문조사 관련) (6)	보편적이지 않은 조사 설계 문항 설계 시의 이론적 논의 구조를 알 수 없는 점 조사의 전체적인 전체나 조사 수행의 조건 등을 알 수 없었음. 자료의 형식, 조사 방법 등이 표준화되어 있지 않음. 데이터의 종류와 수집 기간의 한계 설문조사 항목이 현실과 맞지 않음.
변수 그룹핑, 변수 비공개 등 변수 활용 제약 및 데이터 비공개 (13)	공개 범위 설정 내가 원하는 분류로 다시 데이터를 정리하지 못한 점이 아쉬웠음. 데이터를 제3자가 이용하는 권한의 범위 내가 원하는 데이터 항목 부재 구하기가 어렵고, 개인정보 보호 때문에 연구에 정말 필요한 세부 정보들을 구할 수가 없었음. 본 연구에 맞게 재가공을 위해서는 원자료를 제공받아야 하나, 이를 제공하는 연구자는 전무. 가공되어 있는 데이터를 재가공하는 과정에서 자료 누락과 왜곡이 발생함. 연구 대상과 방법이 연구마다 다른데 이미 타 연구자의 조사데이터는 해당 연구의 내용에 맞춰 선별, 코딩된 상태로 내가 보고자 하는 내용에 대한 데이터가 누락되거나 훼손되어 있을 수 있음. 대부분은 활용이 불가능했고, 활용이 가능했던 경우에는 큰 어려움은 없었음. 해당 기관에 조사데이터 활용에 대한 기준이 명확하지 않아 데이터 활용의 일관성이 없음. 보안 등의 이유로 공유를 거절하는 연구자의 비협조 일부만 공개, 데이터 연결이 어려움. 연구자가 가지고 있는 데이터와 타 연구자의 조사데이터의 병합이 필요한 경우가 있는데 개인 식별의 위험이 존재하지 않는 경우에도 타 연구기관에서 데이터 병합에 소극적인 태도를 지님. 나의 연구에 맞는 세분화된 데이터 확보
데이터 신청 절차 및 소요 시간	구득 과정에 시간 소요 자료 요청의 번거로움. 자료를 받기 전까지 구체적인 사항에 대해서 자세히 알기 어려움. 또한 활용을 위해 협의하는 시간이 많이 걸림. 데이터 획득 절차

구분	응답 내용
데이터클리닝	데이터 클리닝이 잘 되어 있지 않음
	원자료의 보정 과정에 대한 습득 및 이해 어려움.
	데이터 클리닝을 다시 해야 했다는 점에서 데이터 관리 필요성 체감
데이터의 정확성	일부 데이터에서 다소 비상식적인 값이 발견되었지만, 그 데이터의 정확성 여부를 확인할 수 없었음.
	샘플사이즈가 충분히 크지 않았음.
	자료의 신뢰성
기타	일단, 둘째는 분석 결과가 가지고 있는 경우가 많았으며, 분석에 활용한 데이터 코드는 공유하지 않는 경우가 많았음.
	마이크로데이터 공개에 대한 개념이 없던 시절이었으므로, 데이터 활용 필요성 설명 및 공유를 요청해야 하는 번거로움 발생
	민간 데이터의 경우, 재사용이 어렵고 데이터 구매 비용이 비쌘.

자료: 저자 작성.

그러나 원내 조사 결과에서는 데이터 가공 등 데이터 활용에 대한 어려움에 관한 응답이 거의 없었던 반면에 외부 조사 결과에서는 눈에 띄게 많은 것으로 나타났다. 또 원외 응답 결과에서는 설문조사와 관련된 조사 설계에 관한 내용 등의 정보가 충분하지 못하는 등 메타데이터가 미비하여 어려움을 겪는 경우가 더 많은 것으로 나타났다. 데이터 활용에 관한 부분은 어디까지나 연구자 본인의 몫이나 연구자가 조사데이터를 이용하는 데에 있어서 활용상의 제약을 받지 않도록 조사데이터 공개 시에 충분한 정보를 제공하는 등 최대한의 노력을 기울일 필요가 있다.

현재 보건복지데이터포털에서는 마이크로데이터를 다운로드 받을 경우 조사표, 코드북 및 레이아웃 등 기본적인 메타데이터 파일들을 함께 제공하고 있다. 뿐만 아니라 연구 DB와 데이터 뱅크(Data bank) 페이지를 서비스하고 있으며 각각의 페이지에서는 연구보고서와 해당 연구에서 생산한 데이터의 메타데이터를 제공하고 있다. 연구보고서 메타데이터에는 보고서 원문 링크가 포함되어 있어 데이터 이용 시에 보고서 내용도 함께 참고할 수 있도록 서비스를 제공함으로써 이용자의 데이터에 대한

이해를 돕고 있다. 그러나 시의적절한 콘텐츠 신규 업데이트와 정확한 정보 제공을 위해 기능 개선 등의 노력이 계속해서 필요하다.

앞서 우리 원 연구책임자들이 타 기관의 조사데이터 활용 시에 비공개 변수 등의 활용 제약 및 데이터 비공개 등과 관련된 어려움을 많이 느낀 것으로 응답하였는데 원외 연구책임자들의 경우에도 동일한 어려움을 겪은 것으로 나타났다. **데이터 활용을 위해 자유로운 데이터 가공을 통해 이용자가 연구 목적에 맞게 조사데이터를 활용할 수 있도록 제공해야 하나 개인정보 침해나 데이터 오용 우려 등의 이유로 데이터 활용 범위가 제한될 수도 있다.** 따라서 데이터 공개·활용 시 제약이 가해진다면 이에 대한 명확한 기준이나 지침을 수립할 필요가 있다.

원외 연구책임자의 경우에는 데이터클리닝 문제와 관련하여 어려움을 겪은 경우도 있었는데 이는 **데이터의 정확성과도 직결되는 문제로 데이터 이관 및 공개 시에 필수적으로 점검이 필요한 항목**이다. 원내 데이터 관리 업무와 관련하여 참고해야 할 필요가 있는 내용이다.

원외 연구책임자의 타 연구자의 조사데이터 활용 시 경험한 어려움의 내용은 상기의 몇 가지 특징적인 점들을 제외하면 원내 조사 결과와 그 내용이 유사하다. 따라서 앞서 정리한 내용들을 함께 고려하여 전사적 차원의 데이터 관리 체계 수립 전략을 세우고 시사점을 도출할 필요가 있다.

마지막으로 원외 연구책임자 응답자 중 27명이 조사데이터 공유 및 관리를 위해 필요한 부분에 관한 기타 의견을 주관식으로 응답한 것으로 나타났다. 그 응답 내용을 살펴보면 <표 4-14>와 같다. 원외 연구책임자들이 조사데이터 공유 및 관리를 위해 필요한 부분에 관한 기타 의견을 주관식 문항으로 응답한 내용을 살펴보면 크게 3가지로 요약할 수 있다. 첫 번째로는 **업무 부담과 전담 인력 또는 조직의 필요성에 관한 부분**(업무부담 완화), 두 번째로는 **공감대 형성 및 인식 개선에 관한 부분**(인식 개선)

그리고 세 번째로는 **인센티브 제공 또는 권리 보호의 문제와 연구 윤리 문제에 관한 우려**에 대한 부분(인센티브와 권리 보호)이다. 이를 참고하여 우리 원 여건에 맞는 적절한 조사데이터 공유 및 관리 활성화 방안을 제시할 필요가 있다.

〈표 4-14〉 조사데이터 공유 및 관리를 위해 필요한 부분에 관한 기타 의견-원의

구분	응답 내용
업무 부담	공개 필요성은 있으나, 실제 연구기관의 연구 수행만으로도 업무 부담이 과하여, 이전에 생산한 조사자료의 클리닝, 보관 등을 연구 인력이 하기가 어려움.
	연구자 업무 부담을 가중시키는 정책이라고 판단됨.
	특히 조사가 주요 연구 방법이 되고 있는 일부 연구기관에는 데이터 공유 및 공개 압력이 들어오면 큰 부담이 될 것임.
전담 인력 조직	관련 예산 확보가 필요함.
	인력과 관련 예산을 우선 배정 후 도입되어야 함.
	조사데이터 공유를 위해서는 별도의 인력과 예산이 투입되어야 함.
	전담 인력이 있어야, 공감대 형성, 시스템 개발 등 해당 정책을 추진 가능
	관리 인력은 필수적인 요소임.
	현재 소속 기관의 경우 조사데이터를 생산하는 부서 인원이 매우 부족한 상황
공감대 형성 및 인식 개선	데이터 공유에 대한 공감대가 필수적이라고 판단됨.
	공공기관에서 생산된 데이터는 공적 자산이라는 공감대 형성 필요
	조사데이터 개방은 국가연구기관으로서 당연함에도 불구하고 개인 전유물이라고 생각하는 인식 개선을 위해 제도화 노력 필요
	결국에는 왜 해야 하는지 목적이 필요할 것 같음.
	제도나 규정보다 문화가 우선
	국가 승인 통계의 경우 위의 설문에 적합하다 판단하지만, 기관의 기본 연구 또는 일반 연구, 협동 연구 등에 활용한 조사데이터는 공적 개념보다 연구자 개인의 아이디어인 사유재산의 성격이 강함. 따라서 상호 협조와 신뢰를 바탕으로 한 기관 간 거래의 개념으로 접근해야 할 것으로 판단됨.
	인식 개선과 제도 정비가 병행되어야 함.

120 전사적 연구데이터 관리 체계 마련을 위한 연구

구분	응답 내용
인센티브 제공 또는 권리 보호	<p>현 시점에서 조사데이터를 공유하고 관리한다는 자체가 연구책임자에게는 또 다른 업무로 느껴질 수 있기 때문에 연구자가 이러한 변화에 동의할 수 있도록 공감대를 형성하고 적절한 인센티브 제공 방안을 마련하는 것이 중요하다고 생각함.</p>
	<p>조사데이터 제공을 위해서는 별도의 시간과 노력이 필요하므로 이에 대한 업무 평가(혹은 인센티브 제공)가 필요함.</p>
	<p>조사데이터 제공자에 대한 보상</p>
	<p>자료 생산자의 권한 보호</p>
연구 윤리	<p>연구책임자가 조사데이터 공유 의식을 가지도록 인센티브가 필요함.</p>
	<p>조사데이터를 공개함에 있어 데이터를 제공받은 기관 혹은 개인이 데이터의 출처를 명확히 기재하는 등 선행 연구자(데이터 생산자)의 기여(Contribution)에 대해 존중하는 문화가 확산되어야 부담 없이 연구데이터를 제공할 수 있을 것임.</p>
기타	<p>데이터 개방에 기본적으로 동의하나, 동일 데이터 사용에 따른 연구 윤리 문제가 발생할 수 있음.</p>
	<p>연구데이터는 연구 목적에 따라 구성된 것으로, 연구 목적이 변경될 경우에도 이를 효과적으로 사용하기 위해서는 원 연구책임자 등과의 의사소통이 필수적이라고 생각함.</p>
	<p>연구기관 사이의 데이터 공유는 필요하다고 생각함. 단, 일반에게 데이터를 공개하는 것에 대해서는 다른 원칙과 절차가 필요함.</p>
	<p>조사데이터 관리의 필요성은 인정되나, 심의위원회 설치나 전담 인력 및 전담 조직 등은 시기상조이며, 자원과 예산의 적절한 사용이 아닌 것으로 판단</p> <p>데이터 개방 전, 개인정보를 필터링할 수 있는 시스템 필요</p> <p>연구를 위한 내 수행하기 위해 급하게 자료를 구축하므로 자료 질이 낮을 것 같음.</p>

자료: 저자 작성.

제3절 소결

한국보건사회연구원과 경제인문사회연구원 소속 연구기관들을 비교해 본 결과, 최근 3년간 조사연구 수행률은 각각 39.6%, 37.7%였다.

조사데이터 관리 체계는 한국보건사회연구원과 응답 기관 19개 중 9개 기관이 마련되어 있다고 응답하였으며 관련 지침 또는 매뉴얼은 한국보건사회연구원과 응답 기관 19개 중 7개 기관이 갖추고 있었다.

조사데이터 전담팀이 있는 기관은 한국보건사회연구원과 응답 기관 중 5개 기관이 있었으며 조사데이터 관리정보시스템의 경우에는 한국보건사회연구원과 응답 기관 중 8개 기관에서 내부 업무 시스템 혹은 외부 인터넷시스템이 구축되어 있었다.

조사데이터 이관 혹은 외부 공개의 경우, 한국보건사회연구원에서는 연구자에게 인센티브를 제공하지 않고 있었으나 연구회 소속 연구기관 중 1개 기관에서는 제공하고 있었다. 조사데이터 목록을 내부에서 공유하고 있는 기관은 14개 기관이었으며 한국보건사회연구원에서는 별도로 제공하고 있지는 않았다. 조사데이터는 한국보건사회연구원과 응답 기관 19개 중 15개 기관이 외부에 공개하고 있었다.

연구기관들이 제시한 조사데이터 제공 우수사례로는 외부 공개 데이터 목록과 활용 성과물을 홈페이지에 제공, 조사데이터 활용 확산을 위한 논문 경진대회 추진 등이었다. 이는 향후 활용 활성화 방안에서 도입할 필요가 있다 하겠다.

한국보건사회연구원과 연구회 소속 연구기관들은 조사데이터 관리 체계에 있어 전담 인력, 추가 작업 지원 등에 어려움을 적시하고 있었으며 연구회 차원에서 이를 위한 인력 및 예산 지원 방안도 제시하고 있다.

다음으로 한국보건사회연구원 연구책임자들과 연구회 소속 연구기관

연구책임자들의 응답을 비교·분석해 본 결과, 연구책임자들은 오픈 사이언스 개념과 동향, 연구데이터 관리 계획에 대해 현재 알고 있는 것보다 필요성, 중요성에 대해 더 많이 공감하고 있었다.

한국보건사회연구원 연구책임자들을 대상으로 한 조사에서 원내 혹은 타 기관 조사데이터 활용 경험자 중 구득의 필요성, 목록 파악의 용이성, 구득의 용이성, 메타정보의 충분성을 평가한 결과 상대적으로 구득의 용이성과 메타정보의 충분성에서 원내 활용성이 떨어지는 것으로 나타났다.

최근 3년간 조사데이터를 생산한 경험은 한국보건사회연구원 연구책임자들은 55.8%, 연구회 소속 연구기관 연구책임자들은 64.2%를 나타냈다. 또 연구 종료 후 조사데이터 보관 및 관리는 현재 연구자(책임자, 연구진)가 주로 하고 있었으나 한국보건사회연구원 연구책임자들은 65.2%, 연구회 소속 연구기관 연구책임자들은 57.1%가 기관 전담팀이 맡는 것이 적절하다고 제시하였다.

현재 소속 기관 조사데이터 관리 체계에 대해서는 모두 보통으로 생각하고 있었으며, 조사데이터 개방에 대해서는 한국보건사회연구원 연구책임자와 연구회 소속 연구기관 연구책임자 모두 본인이 생산한 조사데이터 대비 일반적인 조사데이터 개방의 필요성이 더 많다고 제시하고 있다.

한편 그럼에도 조사데이터 공유와 개방은 다른 연구자에게 도움이 되기 때문에, 공적 자금을 투입한 조사이기 때문에 당연하다고 생각하고 있었으며 조사데이터 공유와 개방을 위해서는 인력 및 조직체계 보완, 관련 인식 개선을 시급한 과제로 제시하였다.

그 외 한국보건사회연구원 조사데이터 관리 체계 정립을 위하여 통계청 MDIS, 서울대학교 한국사회과학자료원 KOSSDA와의 연계 및 협력 체계 구축이 필요하며 관리 인프라인 보건복지데이터포털의 콘텐츠와 기능 보강, 그리고 적극적인 조사데이터 홍보 필요성이 제시되었다. 또한

데이터 공개, 활용에 대한 명확한 기준 제시가 필요하며, 전담 인력 및 조직의 필요성, 공감대 형성 및 인식 개선, 인센티브 제공, 연구윤리에 관한 우려 등을 나타내었다.



제5장

연구데이터 관리 체계 개선 및 활용 활성화 방안

제1절 연구데이터 관리 대상 및 절차

제2절 관련 조직 및 위원회 구성

제3절 연구데이터 제공 인프라 개편 방안

제4절 연구데이터 활용 활성화 방안

제 5 장

연구데이터 관리 체계 개선 및 활용 활성화 방안

제5장에서는 연구데이터 관리 대상 및 절차, 관련 조직 및 위원회 구성, 연구데이터 제공 인프라 개편 방안, 그리고 연구데이터 활용 활성화를 위한 방안을 제시하고자 한다.

제1절 연구데이터 관리 대상 및 절차

우선적으로 연구데이터의 관리 대상에 대해 생각해 보면, 조사 사업은 연구 계획 단계에서 계획을 수립하게 되며, 자체 과제 혹은 수탁 과제 모두 수행할 수 있다. 조사 사업은 내부 혹은 외부 조사업체에서 수행하며 조사 규모로는 모집단을 조사하는 경우와 표본집단을 조사하는 것으로 나누어 볼 수 있다. 또한 조사데이터는 조사 방법에 따라 흔히 FGI, 면담 등을 통해 생산되는 질적 자료와 구조화된 설문조사에서 생산되는 양적 자료, 그리고 혼재된(질적+양적) 자료가 있다. 보통 양적 자료의 경우 조사 사업을 통해 생산된 원자료(raw data)는 입력오류, 논리오류 등을 수정하는 작업을 거쳐 마이크로데이터를 생산하게 되며 연구데이터 관리는 사실상 오류 수정 작업을 거친 마이크로데이터부터 시작하게 된다. 이러한 마이크로데이터는 연구에 사용하면서 가공, 편집되어 공표데이터가 만들어지는 것이다. 그러므로 기관 차원에서 조사데이터를 관리(수집, 보관, 가공, 제공 등)하기 위해서는 연구자로부터 수집(이관)하고 궁극적으로 제공, 활용에 대한 범위를 생각해 보아야 할 것이다.

먼저 조사데이터를 생산하는 연구과제가 자체 과제일 경우, 조사데이

터는 공공 저작물이 되어 저작권은 소속 기관에 있다. 그러므로 자체 과제에서 수행되는 과제는 모두 저작권이 있는 기관이 보관(이관)하며 동시에 제공 대상이 된다. 수탁 과제의 경우, 발주 기관과의 계약 시 조사데이터의 관리(보관)를 특별히 정하지 않는 한, 조사데이터는 연구를 수행한 기관에서 관리하는 것이 바람직하며 저작권자는 발주 기관이므로 제공은 발주처와 합의가 필요하다. 그러나 간혹 연구 수행 이후 발주처 요구에 의해 보관이 불가할 수도 있다. 정부출연연구기관에서 수행하는 수탁 과제의 대부분은 국가 또는 지방자치단체가 위탁 기관으로서, 몇몇 예외적인 경우⁹⁾를 제외하고는 보다 적극적인 데이터 개방과 활용이 필요하다. 그러므로 이를 위해서는 기관 차원에서 위탁 기관과 연구데이터 개방에 대한 포괄적 협약을 추진하거나 연구 종료 후 발주처의 발간 확인 시 데이터 공개에 대한 동의 여부를 받을 수 있도록 한다([그림 5-1] 참조).

조사는 내부 혹은 외부에서 수행하고 있는데 내부에서 수행하든 외부에서 수행하든, 조사 규모가 모집단을 대상으로 하든 표본 집단을 대상으로 하든 조사데이터 보관(이관)과 제공에 영향을 끼치지 않는다. 다음으로 조사데이터 유형에 있어 양적 데이터는 보관(이관)과 제공 대상이 되며, 다만 질적 데이터의 경우에는 보관(이관)은 하되, 제공 대상에 대한 판단은 연구자가 결정하는 것이 바람직하다. 또 마이크로데이터와 이를 가공하여 공표한 공표데이터는 모두 이관 대상이며, 공표 이전의 마이크로데이터에 대한 제공 여부는 연구자 판단에 따르도록 한다(<표 5-1> 참조).

9) 국가안전보장에 관련되는 정보를 포함하는 경우, 개인의 사생활 또는 사업상 비밀에 해당하는 경우, 「국유재산법」에 따른 국유재산 또는 「공유재산 및 물품 관리법」에 따른 공유재산으로 관리되는 경우 등(「저작권법」 제24조의2 제1항).

이와는 별개로 데이터 공개(제공)를 위해서는 추가적으로 개인정보 포함 여부, 제3자 권리 등 내용을 점검하여야 하며([그림 5-2] 참조) 연구자는 「정보공개법」 제9조 제1항에 따른 비공개 대상 정보¹⁰⁾와 「저작권법」 등에 따라 이용 허락을 받지 않은 정보를 비공개 판단 근거로 할 수 있다.

-
- 10) 1. 다른 법률 또는 법률에서 위임한 명령(국회규칙·대법원규칙·헌법재판소규칙·중앙선거관리위원회규칙·대통령령 및 조례로 한정한다)에 따라 비밀이나 비공개 사항으로 규정된 정보
2. 국가안전보장·국방·통일·외교관계 등에 관한 사항으로서 공개될 경우 국가의 중대한 이익을 현저히 해칠 우려가 있다고 인정되는 정보
3. 공개될 경우 국민의 생명·신체 및 재산의 보호에 현저한 지장을 초래할 우려가 있다고 인정되는 정보
4. 진행 중인 재판에 관련된 정보와 범죄의 예방, 수사, 공소의 제기 및 유지, 형의 집행, 교정(矯正), 보안처분에 관한 사항으로서 공개될 경우 그 직무수행을 현저히 곤란하게 하거나 형사피고인의 공정한 재판을 받을 권리를 침해한다고 인정할 만한 상당한 이유가 있는 정보
5. 감사·감독·검사·시험·규제·입찰계약·기술개발·인사관리에 관한 사항이나 의사결정 과정 또는 내부검토 과정에 있는 사항 등으로서 공개될 경우 업무의 공정한 수행이나 연구·개발에 현저한 지장을 초래한다고 인정할 만한 상당한 이유가 있는 정보. 다만, 의사결정 과정 또는 내부검토 과정을 이유로 비공개할 경우에는 의사결정 과정 및 내부검토 과정이 종료되면 제10조에 따른 청구인에게 이를 통지하여야 한다.
6. 해당 정보에 포함되어 있는 성명·주민등록번호 등 개인에 관한 사항으로서 공개될 경우 사생활의 비밀 또는 자유를 침해할 우려가 있다고 인정되는 정보. 다만, 다음 각 목에 열거한 개인에 관한 정보는 제외한다.
- 가. 법령에서 정하는 바에 따라 열람할 수 있는 정보
- 나. 공공기관이 공표를 목적으로 작성하거나 취득한 정보로서 사생활의 비밀 또는 자유를 부당하게 침해하지 아니하는 정보
- 다. 공공기관이 작성하거나 취득한 정보로서 공개하는 것이 공익이나 개인의 권리 구제를 위하여 필요하다고 인정되는 정보
- 라. 직무를 수행한 공무원의 성명·직위
- 마. 공개하는 것이 공익을 위하여 필요한 경우로서 법령에 따라 국가 또는 지방자치단체가 업무의 일부를 위탁 또는 위촉한 개인의 성명·직업
7. 법인·단체 또는 개인(이하 “법인등”이라 한다)의 경영상·영업상 비밀에 관한 사항으로서 공개될 경우 법인등의 정당한 이익을 현저히 해칠 우려가 있다고 인정되는 정보. 다만, 다음 각 목에 열거한 정보는 제외한다.
- 가. 사업활동에 의하여 발생하여 위해(危害)로부터 사람의 생명·신체 또는 건강을 보호하기 위하여 공개할 필요가 있는 정보
- 나. 위법·부당한 사업활동으로부터 국민의 재산 또는 생활을 보호하기 위하여 공개할 필요가 있는 정보
8. 공개될 경우 부동산 투기, 매점매석 등으로 특정인에게 이익 또는 불이익을 줄 우려가 있다고 인정되는 정보

〈표 5-1〉 관리 대상 조사데이터의 범위

	구분	보관	제공
생산 과제 종류	자체 과제	○	○
	수탁 과제	○(다만, 발주처 요청에 의해 이관불가일 수 있음)	△(발주처와 합의 필요)
조사 수행 주체	내부	○	○
	외부	○	○
조사 규모	모집단	○	○
	표본집단	○	○
조사데이터 유형	질적 데이터	○	△(연구자 결정)
	양적 데이터	○	○
	질적+양적	○	△(연구자 결정)
공표 여부	공표 이전 데이터	○	△(연구자 결정)
	공표데이터	○	○

자료: 저자 작성.

[그림 5-2] 조사데이터 공개를 위한 확인 필요 내용

마이크로데이터 제공 전 Check List

- 01.** 데이터 셋이 갖추어져 있는가?

- 제공용 마이크로데이터, 설문지, 코드북 및 파일설계서가 갖추어져 있는가?
 - 설문지, 코드북 및 파일설계서의 파일 형식이 HWP 또는 XLS 등으로 변환되어 있는가?
 - 마이크로데이터의 파일 형식이 SPSS, SAS 및 STATA 뿐 아니라 CSV로도 제공될 수 있는가?
- 02.** 개인정보 등 비공개 대상 정보가 포함되어있는가?

- 개인정보와 관련된 변수를 식재하였거나 식별할 수 없도록 변환하였는가? 성명, 전화번호(휴대전화번호, 집전화, 회사전화, 팩스번호), 생년월일, 주민등록번호, 여권번호, 외국인등록번호, 운전면허번호 등 고유식별정보, 의료기록번호, 건강보험번호, 복지 수급자 번호, 각종 자격증 및 면허 번호, 자격증 취득일, 이메일 주소, 홈페이지 URL, 군번, 개인사업자등록번호 등
- 03.** 저작권 등 제3자 권리가 포함되어있는가?

- 수탁과제의 경우, 마이크로데이터의 저작권에 관해 발주처와 합의가 완료되었는가?
(저작권이 우의 원으로 귀속되었는지 또는 마이크로데이터 이용허락에 관한 동의를 득하였는지 통찰 상담 등의 증빙 자료를 통해 확인)
- 04.** 데이터 제공을 위한 조치(처리)가 되어있는가?

- 중복 변수 또는 불필요 변수가 식재되어 있는가?
 - 제공용 마이크로데이터와 코드북의 변수 목록이 일치하는가?
 - 오픈 문항 변수는 일관되게 (미)식재하여 정리되어 있는가?
- 05.** 데이터 이용 시 유의 사항이 확인·정리되었는가?

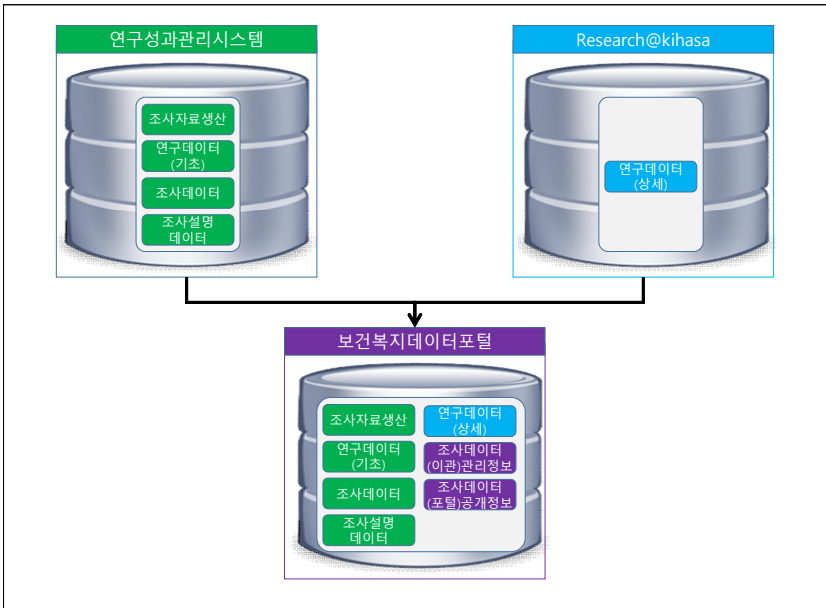
- 코드북이나 파일설계서에 식재되거나 추가된 변수에 관한 정보가 있는가? (시계열적 단절이 있는 변수 또는 지역 변수의 이용 제한 등)
 - IRB 승인 또는 면제 여부에 관한 정보가 적절히 제시되어 있는가?
 - 국가승인통계 또는 법정 통계의 경우 통계 작성 목적과 관련 정보가 적절히 제시 되어 있는가?

자료: 저자 작성.

다음으로 조사데이터를 관리하기 위한 필요 항목은 한국보건사회연구원 내부 연구성과관리시스템(RPMS: Research Project Management System)에서 연계가 필요한 항목, 한국보건사회연구원 연구 레포지터리(Research@kihasa)시스템에서 연계가 필요한 항목, 그리고 보건복지데이터포털에서 생성이 필요한 항목으로 구성할 수 있다([그림 5-3] 참조).

첫째로 연구성과관리시스템에서의 항목은 다시 조사데이터 생산, 조사데이터 이관 영역으로 구분할 수 있으며, 한국보건사회연구원 연구 레포지터리에서의 항목은 관련 연구보고서에 관한 것이다. 보건복지데이터포털에서의 항목은 이관, 공개 영역으로 구분할 수 있다(<표 5-2> 참조).

[그림 5-3] 조사데이터 관리를 위한 필요 항목 구성



자료: 저자 작성.

〈표 5-2〉 조사데이터 관리 대상 항목

영역	구분		설명	
연구 성과 관리 시스템	조사 데이터 생산 정보	연구 과제명	연구성과시스템상의 연구 과제 명칭	
		연구 과제 번호	연구 과제 관리를 위해 부여되는 번호	
		보고서 번호	보고서 발간시 부여되는 번호	
		조사 여부 및 형태	미생산	조사데이터 미생산인 경우
			생산	양적 조사, 질적 조사, 양적+질적 조사 선택
	조사데이터 이관 시기		1개월 이내, 3개월 이내, 6개월 이내 중 택일	
	조사 데이터 이관 기초 정보	연구과제명	연구성과시스템상의 연구과제 명칭	
		연구과제 번호	연구과제 관리를 위해 부여되는 번호	
		보고서 번호	보고서 발간시 부여되는 번호	
		연구책임자	연구책임자 성명	
		자료 문의 담당자	조사데이터 문의 담당자 성명	
		연구 발주 기관	원내	연구 발주 기관이 원내인 경우
	원외		연구 발주 기관이 원외인 경우(발주 기관 및 발주 부서 포함)	
	조사 데이터 이관 부가 정보	데이터명		조사데이터 명칭
		국가 승인 통계 여부	미승인	국가 승인 통계가 아닌 경우
			승인	국가 승인 통계인 경우(국가 승인 통계 번호 포함)
		IRB 승인 여부	미승인	IRB 승인 대상이 아닌 경우
			승인	IRB 승인 대상인 경우(IRB 승인 번호 포함)
		이관 데이터 종류	조사표*	조사표 파일 업로드
			조사데이터*	조사데이터 파일 업로드
			부호화지침서	부호화지침서 파일 업로드
			변수구성표	변수구성표 파일 업로드
			기타	기타 이관 데이터 관련 파일 업로드
		외부 공개 여부	공개	외부 공개 대상인 경우
비공개			외부 비공개 대상인 경우 9개 비공개 사유 중 택일 (상세 사유 포함)	
공개 유보 기간		원내	(외부 공개 대상) 원내 공개 유보 기간	
	원외	(외부 공개 대상) 원외 공개 유보 기간		
보건복지 데이터 포털 공개	데이터 유형	양적, 질적 마이크로데이터		
	연구 분류 코드	연구 분류 코드 최대 3개 선택		
	조사 목적	조사 목적 요약		
	조사 범위	조사 범위 요약		

134 전사적 연구데이터 관리 체계 마련을 위한 연구

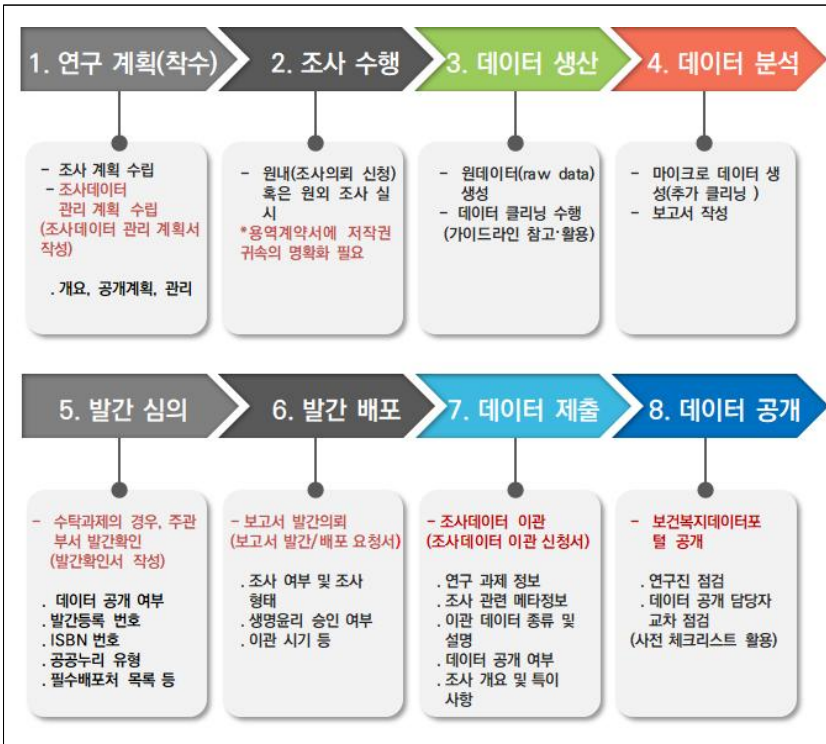
영역	구분	설명	
		조사 방법 요약	
		조사 대상 요약	
		조사 기간	
		표본설계 요약	
		이용 시 유의점	
		관련 근거 법령	조사와 관련한 근거 법령이 있는 경우 작성
		데이터 제공 여부	공개시에만 포털에서 조사데이터 제공
연구원 레포지 터리	연구 보고서 정보	dc.identifier.localId	보고서 발간시 부여되는 번호
		dc.title	발간된 보고서 명칭(국문)
		dc.title.alternative	발간된 보고서 명칭(타언어)
		dc.contributor.author	저자명(연구책임자 및 참여 연구진)
		dc.contributor.affiliatedAuthor	기관 저자
		dc.citation.title	(저널일 경우) 저널 명칭
		dc.citation.volume	(저널일 경우) 권
		dc.citation.number	(저널일 경우) 호
		dc.citation.startPage	(저널일 경우) 시작 페이지
		dc.citation.endPage	(저널일 경우) 끝 페이지
		dc.identifier.bibliographicCitation	표준 서지 인용 정보
		dc.date.issued	발행 연도
		dc.identifier.isbn	ISBN 번호
		dc.identifier.uri	안내 페이지 URL
		dc.description.abstract	요약 정보
		dc.description.tableOfContents	목차
		dc.format	제공되는 정보 형태(jpeg, pdf 등)
		dc.language	언어 형태(kor, eng 등)
		dc.publisher	발행 기관
		dc.rights	저작권 관련 정보
	dc.rights.uri	저작권 관련 안내 페이지	
	dc.subject.keyword	주제어(키워드)	

영역	구분	설명	
보건 복지 데이터 포털	이관 관리 정보	보고서 번호	보고서 발간시 부여되는 번호
		데이터 번호	데이터를 관리하기 위한 번호
		조사데이터 이관 통보 여부	조사데이터 이관 통보 여부
		조사데이터 이관 통보 일자	조사데이터 이관 통보 일자
		조사데이터 이관 신청서 결재 여부	조사데이터 이관 신청서 결재 여부
		조사데이터 이관 신청서 결재 일자	조사데이터 이관 신청서 결재 일자
		메타데이터 작성 여부	조사데이터 관련 메타데이터 작성 여부
		메타데이터 작성 일자	조사데이터 관련 메타데이터 최종 업데이트 일자
		이관 완료 여부	조사데이터 이관 완료 여부(관리자 검토)
		이관 완료 일자	조사데이터 이관 완료 일자(관리자 검토)
	포털 공개 정보	보고서 번호	보고서 발간시 부여되는 번호
		보고서명	발간 보고서명
		데이터 번호	조사데이터를 관리하기 위한 번호
		데이터명	조사데이터 명칭
		데이터 유형	양적, 질적 마이크로데이터
		연구 분류 코드	연구 분류 코드 최대 3개 선택
		조사 목적	조사 목적 요약
		조사 범위	조사 범위 요약
		조사 방법	조사 방법 요약
		조사 대상	조사 대상 요약
이용시 유의점	조사 기간	조사 기간	
	표본설계	표본설계 요약	
	이용시 유의점	이용시 유의사항	
	관련 근거 법령	조사와 관련한 근거 법령이 있는 경우 작성	
	데이터 제공 여부	공개(데이터 제공 가능), 비공개(데이터 제공 불가)	
	데이터 비공개 사유	데이터 비공개시 데이터 비공개 사유	
	데이터 이관 완료 일자	데이터 이관 완료 일자	
	데이터 공개 유보 기간(원내)	데이터 공개 유보 기간(원내)	
	데이터 공개 유보 기간(외부)	데이터 공개 유보 기간(외부)	
	데이터 공개 일자(원내)	데이터 원내 공개 일자	
데이터 공개 일자(외부)	데이터 외부 공개 일자		
데이터 관련 자료	데이터 제공 가능시 데이터 종류 및 제공 경로		

자료: 저자 작성.

관리 절차와 범위를 정하기 위해서 조사데이터가 생산되는 조사 사업에 대해 생각해 보고자 한다. 조사 사업은 연구 계획 단계에서 계획을 수립하게 되며, 자체 과제 혹은 수탁 과제 모두 수행할 수 있다. 조사 사업 또한 내부 혹은 외부 조사 업체에서 수행하게 되며 조사 사업을 통해 생산된 원자료(raw data)는 입력오류, 논리오류 등을 수정하는 작업을 거쳐 마이크로데이터를 생산하게 된다([그림 5-4] 참조).

[그림 5-4] 조사데이터 관리 및 활용 처리 절차



자료: 저자 작성.

우선, ① 연구 계획(착수) 단계에서 연구계획서 작성 시 조사 계획과 더불어 조사데이터 관리 계획을 수립한다. 조사데이터 관리 계획은 국가과학기술연구회에서 제시한 ‘연구데이터 관리 계획’을 참고하여 <표 5-3>과 같이 제시하고자 한다. ‘연구데이터 관리 계획’에서는 연구데이터 개요, 연구데이터의 저장 및 장기적인 보존 계획, 연구데이터의 공유 계획, 데이터 관리 및 공유 책임자로 구성되어 있으나 이번 연구에서는 연구데이터 저장 및 장기적인 보존에 있어 기관에 이관하여 저장 및 보존하므로 이에 대해서는 생략하였다.

<표 5-3> 조사데이터 관리 계획서 구성 내용

1. 연구 과제명	
2. 조사데이터 개요	
조사명	- 조사명
조사데이터 수집·생산 방법	- 조사 방법(대면 설문조사, 우편 설문조사, 전화 조사, 이메일 조사, 인터뷰) - 조사 대상 및 규모
조사데이터 유형	- 데이터 유형(질적, 양적, 질적+양적) - 데이터 파일 유형(텍스트 파일, 음성 파일, 동영상 파일 등)
3. 조사데이터 공개 계획	
데이터 공개 방법	- 데이터 이관 시기 - 데이터 공개 여부(공개 불가시 불가 사유) - 데이터 공개 범위(내부, 외부)와 공개 시기(내부, 외부)
공개 시 유의사항	- 데이터 공개 시 유의항목 및 유의사항
4. 데이터 관리	
데이터 관리 책임	- 연구진 중 데이터의 전반적인 관리 및 세부 사항에 대한 관리 책임자 이름 및 전화번호(내선번호, 핸드폰번호)

자료: 저자 작성.

② 조사 수행 단계에서는 원내 혹은 원외로 조사를 의뢰하여 수행하며 원외로 조사를 의뢰하여 수행할 경우, 외부 조사업체와의 용역계약서 등

에 조사데이터 저작권에 관한 귀속을 명확하게 하는 것이 필요하다.

③ 데이터 생산 단계에서는 조사에서 생산한 원데이터(raw data)의 오류 검증 등 데이터 클리닝 작업을 수행하게 된다.

④ 데이터 분석 단계는 데이터 생산 단계에서 생산한 원데이터에 데이터 클리닝 작업을 거쳐 마이크로데이터가 생성되는 단계로 이를 활용하여 보고서를 작성하게 된다.

⑤ 발간 심의 단계는 보고서 작성 후, 수탁 과제의 경우 발주처에 발간 확인을 받게 되는데 이때 정부 부처와 수탁 관계를 맺은 경우, 계약에 저작권에 관해 별도 규정을 안 했다면 저작권 공동소유 상황이 발생할 수 있어 자유롭게 사용하지 못할 수도 있다. 그러므로 계약시 기관에서 운영하는 사이트(이번 연구에서는 보건복지데이터 포털) 게시 및 활용, 공개 등에 관하여 별도 규율하는 것이 바람직하다.

⑥ 보고서를 발간 시 보고서 발간 배포 요청서에 조사 수행 여부와 형태, 생명윤리 승인 여부, 이관 시기 등을 작성하게 하여 해당 연구 과제에서 조사데이터가 생성되었음을 인지할 수 있게 한다.

⑦ 보고서 발간 후 6개월 이내에 조사데이터를 이관하게 하여 기관에서 수행한 모든 조사에 대한 데이터는 일괄 관리할 수 있게 한다. 조사데이터 이관 신청서에는 연구 과제에 대한 정보와 조사에 대한 메타 정보, 데이터 공개 여부 등을 작성하게 한다.

⑧ 조사데이터 이관 시 공개를 선택한 경우, 데이터 공개 담당자와 연구진의 점검을 거쳐 보건복지데이터 포털에서 데이터를 공개하게 된다.

제2절 관련 조직 및 위원회 구성

현재 한국보건사회연구원 「조사자료 처리 및 관리 지침」에 의하면 ‘조사자료 처리 주관 부서’(현 사회조사팀)가 원내 수행된 모든 조사자료를 이관받아(시스템 처리 부재) 그 중에서 공개하는 데이터를 ‘조사자료 관리·제공 부서’(현 보건복지데이터 포털 운영자)에 재이관토록 하고 있다. 이는 조사데이터 관리 체계상 모든 조사데이터가 정보시스템에서 관리되고 있지 않으며 관리 주체가 이원화, 총괄 관리 주체가 부재하다는 문제점을 지니고 있다. 조사데이터를 수집(이관), 보관, 가공, 제공하는 등의 관리 업무는 사실상 서로 연관되어 있을 뿐 아니라 과정마다 필요 업무들이 존재한다. 그러므로 이러한 일련의 관리 업무를 총체적으로 수행할 조직과 인력이 필요하다.

이러한 조직과 인력은 전담 혹은 담당 업무를 수행하기 위한 것으로 연구 환경에서 공유, 개방의 중요성이 부각되면서 조사표 및 항목 검토, 비공개 대상 정보 검토, 민감 정보에 대한 비식별화 등 전문성과 지속성을 보유하여야 하며, 제4장 연구책임자 조사 결과에서도 제시되었듯이 기관 차원에서는 조사데이터 이관, 개방에 따라 공개용 데이터 검토 및 생성, 관련 자료 작성 등과 같은 연구진의 부가적인 업무를 최소화할 수 있는 방안을 마련할 필요가 있다. 이는 관련 조직 및 인력과 연구진과의 유기적인 협력 관계 속에서 이루어질 수 있으며 동시에 연구진이 점차 데이터 공개를 연구 수행 절차의 일부분으로 받아들일 수 있는 연구 환경 조성도 필요하다.

현재 ‘전자적 조사데이터 관리’ 업무에 맞는 필요 기능들을 나타내면 <표 5-4>와 같다. 한편 이러한 조직과 인력은 초창기에는 TFT(Task Force Team)에서 시작함으로써 구성 및 운영에서의 효율성을 꾀할 수 있다.

〈표 5-4〉 한국보건사회연구원 조사데이터 관리를 위한 필요 기능

필요 기능
- 조사데이터 관리 계획 검토 및 사후관리
- 조사데이터 이관 절차 담당
- 조사데이터 이관에 필요한 관련 자료 검토
- 조사데이터 품질 관리
- 조사 유도, 조사표, layout 등 관리
- 조사데이터 공개를 위한 처리(개인정보 처리, 저작권 검토, 공개 항목 검토, 설문지 및 layout 검토 등)
- 조사데이터 제공 심의위원회 운영 및 관리
- 조사데이터 관리 및 제공에 대한 법률적 자문 의뢰
- 조사데이터 관리 인프라 운영 및 관리
- 조사데이터 활용 전략 수립 및 실행 계획 추진
- 조사데이터 활용 성과 측정
- 조사데이터 연계 및 융합 방안 제시
- 조사데이터 관리 체계 정착을 위한 유인책 마련
- 조사데이터 관련 Q&A 응대

자료 : 저자 작성.

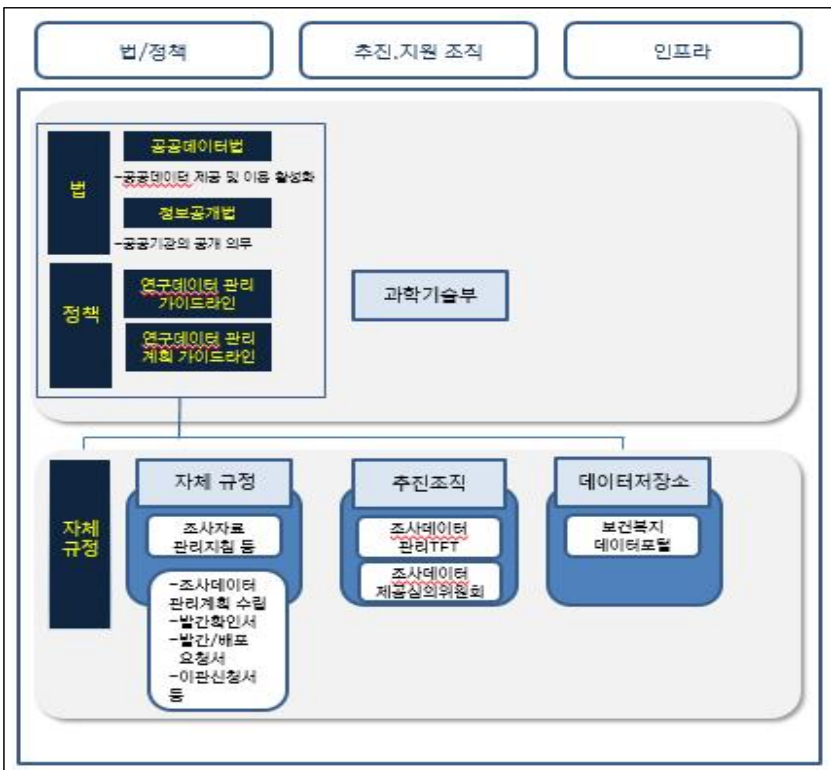
한편 비공개 조사데이터에 대해 원내외에서 공개 요청이 들어온 경우 공개 여부, 이용 승인에 관한 사항, 승인 없이 제3자 제공으로 인한 연구 윤리 위반 시 처리 등 조사데이터 관리상 제반 문제가 발생하였을 때 이를 심의하고 처리하기 위해 ‘조사데이터 제공 심의위원회’ 구성이 필요하다.

조사데이터 제공 심의위원회 역시 초창기에는 별도 구성보다는 연구심의위원회의 심의 안건으로 포함하여 운영하는 것이 효율적일 수 있다.

이상과 같이 연구데이터 관리 대상 및 절차, 관련 조직 및 위원회 구성 등을 포함하는 당 연구원의 조사데이터 관리 체계를 도식화하면 [그림 5-5]와 같다. 조사데이터 관리에 있어 「공공데이터법」과 「정보공개법」을 근거로 하며 조사데이터 관리 정책을 수립함에 있어서는 과학기술부의 연구데이터 관리 가이드라인과 연구데이터 관리 계획 가이드라인을 참고하였다. 다만 과학기술계 정부출연연구기관과는 다르게 경제인문사회계 정부출연연구기관에서는 아직까지 연구데이터 관리 정책에 대한 추진·지원 조직과 인프라가 부재하여 당 원 연구데이터 관리 정책을 수립함에 있

어 그 연결점을 가지지 못하였다. 이러한 배경 하에서 당 원 규정인 「조사 데이터 관리 지침」에 조사데이터 관리 계획 수립 등과 같은 내용을 포함 하고 발간 확인서, 발간/배포 요청서, 이관 신청서 등 관련 서식을 대폭 수정·보완하여 조사데이터 관리 체계는 조사데이터 관리 TF 전담 조직이 추진하기로 한다. 또한 조사데이터 제공 심의 기능을 수행하기 위해 조사 데이터 제공심의위원회를 설치 혹은 타 위원회에서 이와 같은 기능을 수 행하기로 한다. 마지막으로 조사데이터 저장소 및 제공 인프라는 보건복지 데이터포털이 담당하기로 한다.

[그림 5-5] 한국보건사회연구원의 조사데이터 관리 체계



자료: 저자 작성.

제3절 연구데이터 제공 인프라 개편 방안

이번 절에서는 조사데이터 관리를 좀 더 효율적으로 하기 위해 업무시스템인 연구관리시스템(RPMS), 대민용 인터넷 시스템인 보건복지데이터포털(data.kihasa.re.kr)에 조사데이터 관리 기능을 추가하고 조사데이터의 인터넷 공개시 필요한 연구데이터의 중복 입력에 따른 불필요한 예산 낭비와 데이터 입력오류를 최소화하고자 Research@kihasa (repository.kihasa.re.kr)의 연구데이터와의 정보 연계 방안을 제시하였다.

1. 연구관리시스템(RPMS) 기능 개선

연구관리시스템(RPMS) 기능 개선은 연구성과관리시스템(RPMS)의 조사데이터 관리 기능을 개선하고, 이를 통해 축적된 마이크로데이터 및 관련 정보를 보건복지데이터포털과 연계하는 것이다.

연구관리시스템(RPMS)에는 기존의 ‘보고서 발간/배포 요청서’ 기능에 조사데이터 생산 여부를 관리할 수 있도록 개선하고, 조사데이터 이관을 위한 ‘조사데이터 이관 신청서’ 기능을 신설하였다.

가. 조사데이터 생산 여부 관리 기능 개선

조사데이터의 체계적인 관리를 위해서는 연구 진행 과정부터 조사데이터 생산 여부를 파악하고 관리할 필요가 있다. 이를 위해 ‘보고서 발간/배포 요청서’ 서식을 일부 변경하여 ‘조사 여부 및 형태’, ‘생명윤리 승인 여부’, ‘조사자료 이관 시기’ 등 조사자료 생산 여부 등을 관리할 수 있는 항목을 추가하였다(그림 5-6 참조).


[그림 5-6] ‘보고서 발간/배포 요청서’ 양식 설명

보고서 발간/배포 요청서 양식		설명
보고서 발간 / 배포 요청서 기안자 연구책임자 발 시회/시스템장 학술정보 담당자 협 주위 연구책임장 결 의 의 조 제		
발간행사 요청 연구결과물 보안사항 (연구책임자 요청사항) <input type="checkbox"/> 일반 <input type="checkbox"/> 대외비 (구체적 사유:)		
보고서 제목(국문) _____ 보고서 제목(영문) _____ >가변블록 > 수직선배 _____ ISBN _____ 보고서번호 / 영문 _____ <input type="checkbox"/> 발간/배포서 부여하는 경우 있음		
발간/배포번호 _____ 보고서 최종 쪽수 151 pg 표지 구분 발행구분 인쇄일시 (상산일기)		
보고서 표지 종류		
보고서 편성 기재 (나라문란, 4-6배판, A4) 4*6배판 <ul style="list-style-type: none"> * 기본연구: 권리문헌이 한미 (해당)할 경우 불요 * 수탁연구: 수탁기관인 요청 함 * 기관문헌: 요청과기했다 있음 / 4-6배판: 포함과기 / A4: 사용지 과기 		
[책임연구자/공동연구자] * 저자표기 기준 개정안 (2021. 7. 11 문수)		
발 행 기 비밀보기를 지어줄지 여부		
① 조사자료 생산 조사 여부 및 형태 <input type="checkbox"/> 생산 <input type="checkbox"/> 양적조사 <input type="checkbox"/> 질적조사 <input type="checkbox"/> 양적+질적조사 <input checked="" type="checkbox"/> 미생산 생명윤리 승인 여부 <input type="checkbox"/> 승인 <input type="checkbox"/> 미승인 조사자료 이관 시기 <input type="checkbox"/> 1개월 이내 <input type="checkbox"/> 2개월 이내 <input type="checkbox"/> 6개월 이내		
② 홈페이지 공개 여부 * 공개 <input type="checkbox"/> 비공개(구체적 사유:) * 수탁교계의 홈페이지 공개는 표지, 목차, 요약까지입니다. 주위 용어 해당 국문: 사회서비스, 사회서비스원, 성과분석		
배포 요청 총 인쇄 수량 100 부 [내 배출] [출판]		
보고서 발행 사항 발행부수 50부 최종 발행일시 2020-07-15 중기요청사항 _____		
연구책임자 / 공동연구자 연구진 총 12부 or (수령인) _____		
원고접수/발주 참고사항 원고 접수 배부 5부(원정실 1, 부원정실 1, 연구위원 1, 학술정보팀 2) 원내보관용 20부		
외부배포 참고사항 발송소재지 13부		

자료: 저자 작성.

또한 연구관리시스템(RPMS)에 변경된 ‘보고서 발간/배포 요청서’ 양식을 적용할 수 있도록 기능을 개선하였다(그림 5-7) 참조).

[그림 5-7] 연구관리시스템(RPMS)의 '보고서 발간/배포 요청서' 연구자용 화면(개선 후)

'보고서 발간/배포 요청서' 연구자(기안자) 화면	설명
	<p>① 조사자료 생산 정보 관리 기능 추가</p> <ul style="list-style-type: none"> - 조사 여부 및 형태 - 생산(양적, 질적, 양적+질적) - 미생산 - 생명윤리 승인 여부 - 승인 - 미승인 - 조사자료 이관 시기 - 1개월 이내 - 3개월 이내 - 6개월 이내 <p>(홈페이지 공개)</p> <ul style="list-style-type: none"> - 홈페이지 공개 여부 - 공개 - 비공개(사유 포함) - 주요 용어(3개)

자료: 저자 작성.

한편 조사데이터 관리 담당자는 전자결재시스템(smart flow)의 결재 서류를 일일이 확인하여 별도 엑셀파일로 관리하지 않아도 조사데이터를 생산한 연구 과제 목록과 각 연구 과제별 조사데이터 형태, 생명윤리 승인 여부, 조사자료 이관 시기 등 조사데이터 관리를 위해 관련 정보를 조회 및 수정할 수 있는 관리자용 화면을 신설하였다.

나. 조사데이터 생산 여부 관리 기능 개선

한편 2020년 조사데이터 이관을 체계적으로 관리하기 위해 조사데이터 이관 절차를 정립하고 ‘조사데이터 이관 신청서’ 양식을 개발하여 이를 시행한 바 있다([그림 5-8] 참조).

하지만 ‘조사데이터 이관 신청서’는 행정 절차를 위해 전자결재시스템 (smart flow)에만 우선 적용되어, 조사데이터 이관 신청과 관련한 정보가 DB로 관리되고 있지 않은 문제점을 지니고 있어 연구관리시스템 (RPMS)에 ‘조사데이터 이관 신청서’ 양식을 적용할 수 있도록 기능을 신설하였다([그림 5-9] 참조). 한편 조사데이터 관리 담당자용 기능도 새로이 추가 개발하였다.

[그림 5-8] '조사데이터 이관 신청서' 양식

조사데이터 이관 신청서

번호	조사데이터 담당자	기안자	연구책임자	소속부서장	정보통계 연구실장	결과
조사						재
연구과계명			연구과계 번호			
			보고서 번호			
연구책임자	자료문의담당자					
연구발주기관	() 원내 () 원외(기관명: _____ 부서명: _____)					
데이터명 1	조사수행기간					
국가승인통계 여부	() 승인(승인번호: _____) () 미승인		IRB 승인 여부	() 승인(승인번호: _____) () 미승인		
이관 데이터 종류	() 조사표 () 조사데이터(SAS, SPSS, STATA 등) () 부호화지침서(Code book) () 변수구성표(Layout) () 기타 (_____)					
외부 공개 여부	() 공개 * 공개 내용 및 범위에 대해 연구책임자와 협의 과정 거침 [공개시 주의 및 참고 사항]		공개 유보 기간 (임박고)	원내(____ 개월) 원외(____ 개월) * 이관 시점으로부터 12개월 이내 (유보 기간 없이 공개시에는 0으로 기재)		
	() 비공개 (사유: * ㉠ 기타 선택시 상세 사유 기재)					

* 동일 과제에서 수행한 조사 개수에 따라 데이터명 2. 데이터명 3...으로 표를 추가하여 작성(2연의 [이관 신청서 작성 방법] 참조)

• 비공개 사유(공공기관의 정보공개에 관한 법률 제9조)

- ① 다른 법률 또는 법률에서 위임한 명령(국회규칙·대법원규칙·헌법재판소규칙·중앙선거관리위원회규칙·대통령령 및 조례로 한정한다)에 따라 비밀이나 비공개 사항으로 규정된 정보
- ② 국가안전보장·국방·통일·외교관계 등에 관한 사항으로서 공개될 경우 국가의 중대한 이익을 현저히 해칠 우려가 있다고 인정되는 정보
- ③ 공개될 경우 국민의 생명·신체 및 재산의 보호에 현저한 지장을 초래할 우려가 있다고 인정되는 정보
- ④ 진행 중인 재판에 관련된 정보와 범죄의 예방·수사·공소의 제기 및 유지, 형의 집행, 교정(矯正), 보안처분에 관한 사항으로서 공개될 경우 그 직무수행을 현저히 곤란하게 하거나 형사피고인의 공정한 재판을 받을 권리를 침해한다고 인정할 만한 상당한 이유가 있는 정보
- ⑤ 감사·감독·검사·시험·규제·입찰계약·기술개발·인사관리에 관한 사항이나 의사결정 과정 또는 내부검토 과정에 있는 사항 등으로서 공개될 경우 업무의 공정한 수행이나 연구·개발에 현저한 지장을 초래한다고 인정할 만한 상당한 이유가 있는 정보. 다만, 의사결정 과정 또는 내부검토 과정을 이유로 비공개할 경우에는 의사결정 과정 및 내부검토 과정이 종료되면 제10조에 따른 청구인에게 이를 통지하여야 한다.
- ⑥ 해당 정보에 포함되어 있는 성명·주민등록번호 등 개인에 관한 사항으로서 공개될 경우 사생활의 비밀 또는 자유를 침해할 우려가 있다고 인정되는 정보
- ⑦ 법인·단체 또는 개인(이하 "법인등"이라 한다)의 경영상·영업상 비밀에 관한 사항으로서 공개될 경우 법인등의 상당한 이익을 현저히 해칠 우려가 있다고 인정되는 정보
- ⑧ 공개될 경우 부당한 투기, 매점매석 등으로 특정인에게 이익 또는 불이익을 줄 우려가 있다고 인정되는 정보
- ⑨ 기타

자료: 저자 작성.

[그림 5-9] 연구관리시스템(RPMS)의 '조사데이터 이관 신청' 연구자용 화면(신설)

'조사데이터 이관 신청' 연구자(기안자)용 화면	설명
	<p>① 조사데이터 이관 정보 관리 기능 (연구 정보) - 연구 과제명 - 연구 과제 번호 - 연구책임자 - 자료 문의 담당자 - 보고서 번호 - 연구 발주 기관 - 첨부파일</p> <p>(데이터 정보) - 데이터명 - 조사 수행 기간 - 국가 승인 통계 여부 - IRB 승인 여부 - 이관 데이터 종류 - 외부 공개 여부 - 공개 유보 기간</p> <p>② 홈페이지 공개를 위한 조사데이터별 메타데이터 관리 기능(조회, 입력, 수정) - 데이터 유형 - 연구 분류 코드(3개) - 조사 목적 - 조사 범위 - 조사 방법 - 조사 대상 - 조사 기간 - 표본설계 - 통계승인여부 - 통계 승인 여부 - 이용시 유의점 - 관련 근거 법령 - 데이터 제공 여부</p>

자료: 저자 작성.

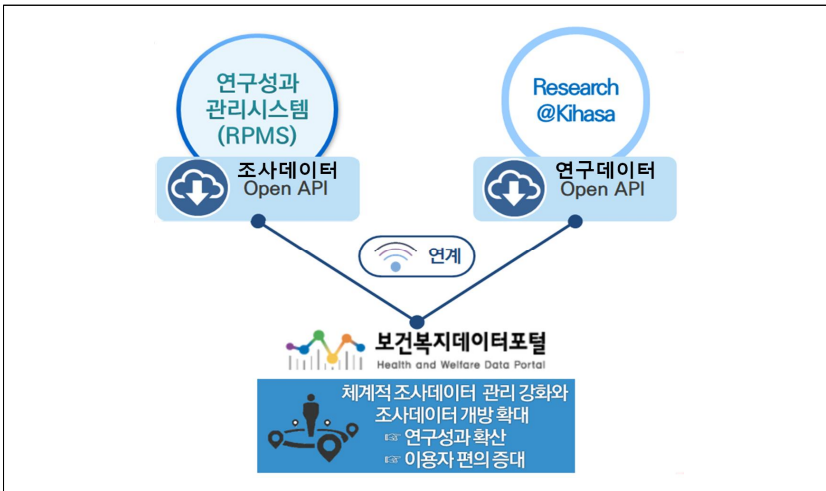
2. 보건복지데이터포털 기능 개선

현재 보건복지데이터포털에서는 조사데이터를 관리하는 기능이 다소 미흡하고 관련 정보가 연계되지 않아 서비스 확대에 한계가 발생하고 있다. 그러므로 관련 정보시스템 간 정보 연계 체계를 구축하고 보건복지데이터포털 내 조사관리시스템을 구축하고자 한다.

가. 관련 정보시스템 간 통합 정보연계 체계 구축

조사데이터 관리 및 연구데이터 연계를 위해 연구성과관리시스템(RPMS), 보건복지데이터포털, Research@kihasa 등 3개 관련 정보시스템 간 통합 정보 연계를 구축하였다(그림 5-10 참조).

[그림 5-10] 관련 정보시스템 간 통합 정보 연계 개념도



자료: 저자 작성.

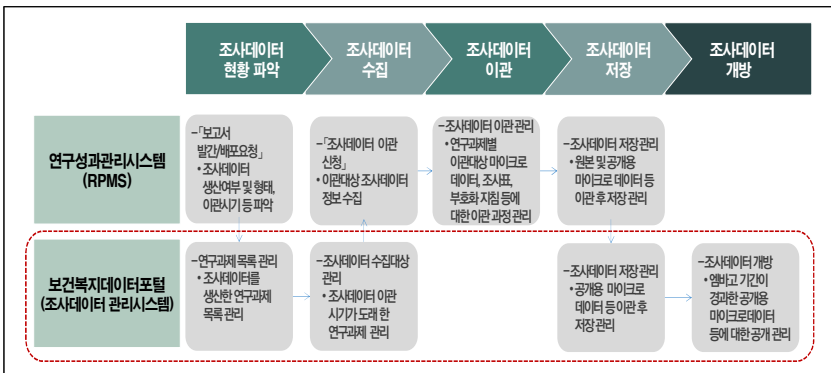
연구성과관리시스템(RPMS)에서 관리하는 데이터 중 조사데이터 관리를 위한 필요 정보를 보건복지데이터포털로 연계할 수 있도록 Open API 기능을 신설하였으며 연구성과관리시스템(RPMS)에서 Open API를 통해 제공되는 정보는 ‘보고서 발간/배포 요청’ 중 조사자료 생산 정보, ‘조사데이터 이관 신청서’ 중 연구 과제 정보와 데이터 정보, 데이터별 메타데이터 정보 등이다.

Research@kihasa에서 관리하는 데이터 중 연구데이터 관리를 위한 필요 정보를 보건복지데이터포털로 연계할 수 있도록 Open API 기능을 일부 수정하였다.

나. 조사데이터 관리시스템 구축

보건복지데이터포털에서는 관련 정보시스템 간 통합 정보 연계를 통해 구축된 데이터를 기반으로 조사데이터 현황을 파악하고 체계적으로 관리(수집, 이관, 저장, 공유(개방) 등)할 수 있는 조사데이터 관리시스템을 구축하였다([그림 5-11] 참조).

[그림 5-11] 조사데이터 관리 단계별 기능 설명

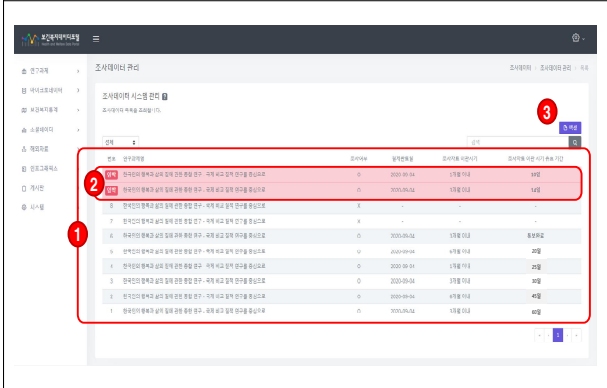


자료: 저자 작성.

150 전사적 연구데이터 관리 체계 마련을 위한 연구

보건복지데이터포털의 조사데이터 시스템 관리 기능에서는 연구성과 관리시스템(RPMS)으로부터 연계된 정보 중 ‘보고서 발간/배포’의 조사 생산 여부 정보를 토대로 연구 과제의 조사데이터 생산 여부, 조사자료 이관 시기 등을 조사데이터 관리자가 쉽게 파악하고 관리할 수 있도록 화면을 구성하였다(그림 5-12) 참조). 관리자는 조사데이터를 생산하였고 보고된 연구 과제 중 조사자료 이관 시기가 임박한 연구 과제 목록을 추출하여, 해당 연구 과제의 조사데이터 이관 신청서 작성을 요청하게 된다.

[그림 5-12] 조사데이터 시스템 관리 화면

조사데이터 시스템 관리 화면	설명
 <p>The screenshot shows a web application interface for managing survey data. The main content is a table with the following columns: '과제명' (Task Name), '상태' (Status), '생산일자' (Production Date), '이관일자' (Transfer Date), and '이관기간' (Transfer Period). The first row is highlighted in red, and a red circle is drawn around the '이관일자' column. Another red circle is drawn around the '이관기간' column. The interface also includes a search bar and a '조회' (Search) button.</p>	<ol style="list-style-type: none"> ① 연구 과제별 조사데이터 생산 여부와 조사데이터 이관 시기 현황 조회 ② 조사데이터 이관 시기가 임박한 연구 과제 목록 ③ 조사데이터 이관 시기가 임박한 연구 과제 엑셀다운로드 기능

자료: 저자 작성.

제4절 연구데이터 활용 활성화 방안

이번 절에서는 연구데이터 활용을 활성화하기 위한 방안을 제시하고자 한다.

첫째, 관련 기관(사이트)에 대한 벤치마킹 및 연계

통계청의 MDIS와 서울대학교 한국사회과학자료원의 KOSSDA는 마이크로데이터 제공에 있어 많은 이들이 찾고 있는 좋은 사이트이다. 이에 반해 보건복지데이터포털은 콘텐츠 측면 및 기능 측면에 있어 한참 뒤떨어져 있다. 그러므로 향후 보건복지데이터포털은 질적, 양적으로 더 많은 마이크로데이터 제공과 더불어 다운로드 서비스뿐 아니라 다양한 서비스를 개발할 필요가 있으며 통계청과 한국사회과학자료원을 또 다른 접근 경로로 활용할 필요가 있다. 또 조사에서 나타난 바와 같이 연구자들은 원내뿐 아니라 외부에서 생산되는 많은 조사데이터들도 필요로 하고 있다. 그러므로 MDIS, KOSSDA 등과 같은 마이크로데이터 제공 사이트, 각종 패널 데이터 제공 사이트 등을 관련 사이트로 제공해 주면 좋을 것이다.

둘째, 조사데이터 이관 및 공개에 따른 인센티브 제도 도입 및 지원 방안 마련

조사데이터 이관과 공개는 공적 자금을 활용하여 공공기관에서 연구를 수행하는 연구자로서는 당연한 책무라 할 수 있다. 그러나 공감대가 충분히 확보되지 않은 현 상황에서는 초반에 연구 평점 부여와 같은 인센티브 제도도 매우 효과적일 수 있다. 또한 데이터 이관 및 공개에 수반되는 연구자들의 업무 부담을 경감시켜 주기 위해 기관 차원의 인력 지원, 더 나아가 연구회 차원의 인력 및 예산 지원도 고려해 볼 수 있다.

셋째, 조사데이터 활용을 장려하기 위한 방안 도입

조사데이터 활용 확산을 위해서는 제공하고 있는 조사데이터를 활용한 논문 경진대회를 개최한다거나 혹은 제공 받은 조사데이터를 활용한 성과를 찾아내어 해당 목록과 성과물을 홈페이지(보건복지데이터포털)에서 제공하여 조사데이터 활용에 대한 동기부여와 아울러 성과물 홍보를 수행해 볼 필요가 있다.

현재 보건복지데이터포털에서 제공하고 있는 조사데이터는 총 29종으로 이들 데이터 활용 현황을 파악하기 위해 2020년 10월부터 약 2달간 자료 수집 요원이 구글 학술검색사이트(scholar.google.co.kr)에서 한국어 웹을 대상으로 조사명을 키워드로 하여 검색¹¹⁾해 본 결과 학술논문지, 학위논문 등에서 해당 조사명을 찾은 건수는 <표 5-5>와 같다. 이들 중에는 원문도 검색되어 보건복지데이터포털에서 목록 제공, 원문 링크 제공 등과 같이 성과물 홍보를 수행할 수 있을 것이다.

<표 5-5> 한국보건사회연구원에서 제공 중인 마이크로데이터 활용 실적

일련 번호	조사데이터명	계	원문이 있는 경우	원문이 없는 경우
계		972	933	39
1	노숙인 등의 실태 조사	1	1	0
2	노인보호 (학대) 실태 조사	5	5	0
3	노인 실태 조사	374	369	5
4	사회문제와 사회통합 실태 조사	2	2	0
5	사회복지시설 종사자 보수수준 및 근로여건 실태 조사	7	7	0
6	사회이동과 사회통합 실태 조사	3	3	0
7	사회통합 및 국민행복 인식 조사	1	1	0

11) 국내의 여러 학술 DB 사이트가 있으나 이번 연구에서는 1개 학술 DB만을 대상으로 하여 국내 학술지로 한정하였으며, 조사명은 현재 조사명을 기준으로 하였기에 조사명이 변경된 경우에는 반영되지 못하였음.

일련 번호	조사데이터명	계	원문이 있는 경우	원문이 없는 경우
8	사회통합 실태 및 국민 인식 조사	3	2	1
9	영아 모성 사망 조사	4	3	1
10	웹다잉에 관한 전국민 인식 조사	1	1	0
11	의료서비스 경험 조사	11	7	4
12	장애인 실태 조사	220	204	16
13	저소득층 자활사업 실태 조사	6	6	0
14	저출산고령화 대응 자영업자 복지수요 조사	1	1	0
15	전국 결혼 및 출산 동향 조사	55	55	0
16	전국 노인장기요양보호 서비스 욕구 실태 조사	17	17	0
17	전국 다문화 실태 조사	70	70	0
18	전국 출산력 및 가족 보건, 복지 실태 조사	70	69	1
19	중년층의 생활 실태 및 복지욕구 조사	4	4	0
20	차상위 계층 실태 조사	3	3	0
21	한국사회의 사회심리적 불안 실태 조사	12	11	1
22	환자 조사	102	92	10

자료: 저자 작성.

넷째, 조사데이터 목록 제공을 통한 활용 장려

인터넷시스템(보건복지데이터포털)을 통한 조사데이터 목록 제공뿐 아니라 기관 내부 직원용 업무시스템에서의 조사데이터 목록 제공도 매우 중요하다. 더욱이 향후 원내, 원외 엠바고 기간을 설정하게 되면 원외에서는 아직까지 비공개하더라도 원내에서는 공개일 수 있어 원내 업무시스템에서의 목록 제공은 매우 필요하다 할 것이다.

다섯째, 발주처와의 협의

현재 수탁 과제에서 생산되는 조사데이터는 원칙적으로 저작권이 발주처에 있으므로 사전 협의 없이 공개하는 것은 무리가 있다. 그러므로 이

를 위해 과제 종료 후, 보고서 발간 심의 시 조사데이터 공개에 대한 사전 승인 절차를 두어 더 많은 조사데이터 공개를 추진해 나갈 필요가 있다. 이때 발주처와의 적극적인 협의 혹은 일괄 협약 등의 논의가 필요하다.

여섯째, 공감대 형성을 위한 환경 조성

제4장 연구자들에 대한 설문조사 결과에서도 나타났듯이 공적 자금을 투입하여 생산한 연구 결과물을 공개하는 것은 당연하다고 생각하면서도 데이터에 대한 오해석과 오용, 개인정보 침해, 데이터 개방을 위한 시간과 노력에 대한 부담감 등으로 인해 본인이 생산한 데이터 공개를 꺼리기도 한다. 이러한 우려들은 앞에서 언급한 전담(담당) 조직 및 인력이 필요 기능을 원활하게 수행함으로써 많은 부분이 해소될 수 있다. 그러나 연구자 개인 차원에서 성과를 전유화하기 위해 공개를 보류하는 경우가 종종 일어난다는 지적도 있다(McMullin, 1985; 신은정 외, 2016 재인용). 그러므로 병행되어야 할 것은 연구 환경에서의 오픈 사이언스(Open Science) 사상에 대한 공감대를 형성하는 것이다. 이를 위해 교육과 홍보가 필요하며 더 많은 조사데이터의 재활용과 병합을 통한 가치를 확인할 필요가 있다.

일곱째, 데이터 영문화 작업 수행

연구데이터는 국내뿐 아니라 해외에서도 수요가 증가하고 있다. 특히 국가 간 비교를 위한 연구나 영문으로 작성하는 논문의 경우 조사자료의 영문화를 필요로 하는 바, 수요가 높은 조사데이터부터 순차적인 영문화 작업이 필요하다. 이때 제공 인프라인 보건복지데이터포털에서도 영문 홈페이지 개발이 병행되어야 할 것이다.

마지막으로 이번 연구는 당 원의 조사데이터 관리 체계상의 문제점을 해결하기 위해 세계 각국에서 공공정책으로 펼치고 있는 오픈 사이언스

정책을 기초삼아 관리 항목과 관리 절차, 관리 조직, 관리 인프라 등을 총괄적으로 제시하였다. 그러나 일차적으로 연구원 자체 과제에서 수행한 조사데이터, 양적 데이터를 대상으로 제시하였는바, 외부에서 위탁받은 수탁 과제에서 수행한 조사데이터, 질적 데이터에 대한 내용은 순차적으로 다루어야 할 뿐 아니라 예·결산 데이터, 재정추계 데이터, 각종 행정 데이터 등과 같은 2차 수집 데이터 관리에 대한 고민이 이어져야 할 것이다. 또 앞에서 제시한 조사데이터 관리 TFT를 운영할 사업이 연속하여 마련되어 연구데이터 관리 체계가 일회성이 아닌 영속성 있는 유의미한 사업으로 남아 있기를 기대한다.



〈국내문헌〉

- 공공기관의 정보공개에 관한 법률, 법률 제14839호 (2017. 7.26. 시행).
- 공공데이터 공통표준용어, 행정안전부고시 제2020-42호. (2020. 8. 11. 시행).
- 공공데이터 관리지침, 행정안전부고시 제2019-71호. (2020. 9. 4. 시행).
- 공공데이터의 제공 및 이용 활성화에 관한 법률, 법률 제17344호 (2020. 12. 10. 시행).
- 과학기술분야 정부출연연구기관 등의 설립·운영 및 육성에 관한 법률, 법률 제 17341호 (2020. 12. 10. 시행).
- 국가과학기술연구회. (2019a). **연구데이터 관리 가이드라인**. 세종: 국가과학기술 연구회, 한국과학기술정보연구원.
- 국가과학기술연구회. (2019b). **출연(연) 연구데이터 관리·활용 방안 연구**. 세종: 국가과학기술연구회, 한국과학기술정보연구원.
- 국가과학기술연구회. (2019c). **연구데이터관리계획(DMP) 가이드라인**. 세종: 국 가과학기술연구회, 한국과학기술정보연구원.
- 국가연구개발사업의 관리 등에 관한 규정, 대통령령 제31251호 (2020. 12. 10. 시행).
- 김용훈, 오영균. (2008). 정부출연연구기관의 지배구조에 관한 연구. **한국행정연 구**, 17(3), 279-297.
- 보건복지데이터포털 홈페이지. (n.d.). **마이크로데이터 전체목록**.
<https://data.kihasa.re.kr/microdata/apply/list>에서 2020. 10. 25. 인출.
- 보건복지데이터포털 홈페이지. (n.d.). **마이크로데이터 이용 안내**.
<https://data.kihasa.re.kr/microdata/guide>에서 2020. 11. 10. 인출.
- 생명공학정책연구센터. (2018). 바이오의학 연구논문의 투명성 및 재현성 개선. **BiolNwatch**, 18-90, 1-2.
- 생명윤리 및 안전에 관한 법률, 법률 제17472호 (2020. 9. 12. 시행).
- 신은정. (2015). 오픈 사이언스(Open Science)에 관한 OECD 논의 동향과 시

사점. **동향과 이슈**, 22, 1-21.

신은정, 안형준, 정원교. (2016). **오픈 사이언스를 위한 연구 성과물 공개 정책과 과제**. 세종: 과학기술정책연구원.

신은정, 안형준, 양현채, 최병삼, 양승우, 정원교, 김수연. (2017). **오픈 사이언스 정책의 도입 및 추진 방안**. 세종: 과학기술정책연구원.

오세영, 윤건, 오균. (2017). **증거기반 정책을 위한 정부의 통계 구축 및 활용에 대한 현황 조사**. 서울: 한국행정연구원.

이민형, 김태양. (2020). 공공연구기관 R&D 협력체계 구축을 위한 주요 혁신방안. **STEPI insight**, 252, 1-30.

저작권법, 법률 제17592호 (2020. 12. 8. 시행).

정부출연연구기관 등의 설립·운영 및 육성에 관한 법률, 법률 제16291호 (2019. 1. 15. 시행).

정영철. (2020). 근거기반 정책 지원을 위한 연구데이터 관리와 개방. **보건복지포럼**, 286, 7-18.

최명석. (2017). [Special Edition 4] 오픈 연구데이터 정책 동향. **ie매거진**, 24(4), 30-35.

최명석, 김현우, 이상환. (2019). 국가 연구데이터 공유의 시작, 데이터 관리계획, **KISTI ISSUE BRIEF**, 13, 1-16.

통계청 MDIS 홈페이지. (n.d.). **제공자료**.

https://mdis.kostat.go.kr/infoData/serviceData.do?curMenuNo=UI_POR_P9007에서 2020. 10. 25. 인출.

통계청 MDIS 홈페이지. (n.d.). **MDIS 소개**.

https://mdis.kostat.go.kr/pageLink.do?link=content/SISMN091002&curMenuNo=UI_POR_P9002 에서 2020. 10. 25. 인출.

한국과학기술정보연구원 기술사업화 플랫폼 홈페이지. (n.d.). **사업화 유망기술 중 [데이터처리] 기관형연구데이터리포지터리시스템**.

http://www.tbm-platform.com/bbs/board.php?ho_id=2019kisti&bo_table=tech&wr_id=5288에서 2021. 1. 6. 인출.

- 한국보건사회연구원. (2012). 「조사자료 처리 및 관리 지침」.
- 한국보건사회연구원. (2016. 6. 9.). 「조사자료 관리 매뉴얼」.
- 한국보건사회연구원. (2019. 9.). 「마이크로데이터 제공 업무 가이드라인」.
- 한국보건사회연구원. (2020. 3.31.). 한국보건사회연구원, 100대 싱크탱크 정치·사회 부문 1위. **한국보건사회연구원 보도자료**.
<https://www.kihasa.re.kr/web/news/report/view.do?ano=10888&menuId=20&bid=79>에서 2020. 12. 17. 인출.
- 한국사회과학자료원 KOSSDA 홈페이지. (n.d.). **양적, 질적 자료DB**.
<https://kossda.snu.ac.kr/data-search/data>에서 2020. 10. 15. 인출.
- 한국사회과학자료원 KOSSDA 홈페이지. (n.d.). **이용안내**.
<https://kossda.snu.ac.kr/component/about/register>에서 2020. 10. 15. 인출.
- 한국정보화진흥원. (2015). 데이터 증거기반(Evidence-Based)의 과학적 정책 수립 방안. **IT & Future Strategy**, 6, 1-19.

<외국문헌>

- Sowaki, H. (2009). 英国におけるエビデンスに基づく教育政策の展開[The Development of Evidence-based Education Policy in England]. *国立教育政策研究所紀要*, 139(4), 153-168.
- Hey, T., Tansley, S., & Tolle, K. (2009). Jim Gray on eScience: A Transformed Scientific Method. *The Fourth Paradigm: Data-Intensive Scientific Discovery*(pp. xvii-xxxi). Newyork: Microsoft Research.
- McMullin, E. (1985). Openness and Secrecy in Science: Some Notes on Early History. *Science, Technology & Human Values*, 10(2), 14-23. doi: 10.1177/016224398501000203
- Merton, R. K. (1942). *The Sociology of Science: Theoretical and*

Empirical Investigations, Chicago: University of Chicago Press.

OECD. (2015). Making Open Science a Reality. *OECD Science, Technology and Industry Policy Papers*, 25, 1-108. doi: 10.1787/5jrs2f963zs1-en

Wilkinson, M. D., M. Dumontier., I. J. Aalbersberg., G. Appleton., M. Axton., A. Baak., N. Blomberg., ... B. Mons. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3(1). 1-9. doi: 10.1038/sdata.2016.18

간행물 회원제 안내

회원제에 대한 특전

- 본 연구원이 발행하는 판매용 보고서는 물론 「보건복지포럼」, 「국제사회보장리뷰」도 무료로 받아보실 수 있으며 일반 서점에서 구입할 수 없는 비매용 간행물은 실비로 제공합니다.
- 가입기간 중 회비가 인상되는 경우라도 추가 부담이 없습니다.

회원 종류

전체 간행물 회원

120,000원

보건 분야 간행물 회원

75,000원

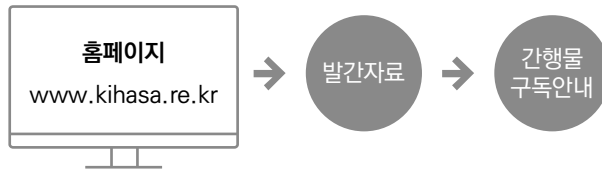
사회 분야 간행물 회원

75,000원

정기 간행물 회원

35,000원

가입방법



문의처

- (30147) 세종특별자치시 시청대로 370 세종국책연구단지
사회정책동 1~5F
간행물 담당자 (Tel: 044-287-8157)

KIHASA 도서 판매처

- 한국경제서적(총판) 02-737-7498
- 영풍문고(종로점) 02-399-5600
- Yes24 <http://www.yes24.com>
- 교보문고(광화문점) 1544-1900
- 알라딘 <http://www.aladdin.co.kr>