



워킹페이퍼 2023-01

보건복지 조사데이터의 체계적 관리 및 활용 방안 연구

— 한국보건사회연구원 조사자료 관리 사례를 중심으로

박성준
신정우·조용찬



■ 연구진

연구책임자	박성준	한국보건사회연구원 부연구위원
공동연구진	신정우	한국보건사회연구원 연구위원
	조용찬	한국보건사회연구원 연구원

워킹페이퍼 2023-01

보건복지 조사데이터의 체계적 관리 및 활용 방안 연구

- 한국보건사회연구원 조사자료 관리 사례를 중심으로

발행일 2023년 7월
발행인 이태수
발행처 한국보건사회연구원
주소 [30147]세종특별자치시 시청대로 370
세종국책연구단지 사회정책동(1~5층)
전화 대표전화: 044)287-8000
홈페이지 <http://www.kihasa.re.kr>
등록 1999년 4월 27일(제2015-000007호)

© 한국보건사회연구원 2023
ISBN 978-89-6827-962-1 [93510]
<https://doi.org/10.23060/kihasa.i.2023.01>

발|간|사

IT기술의 발달에 따라 각종 정보의 처리와 활용에 대한 중요성이 강조되고 있다. 이에 따라 정부와 공공기관, 관련 연구기관에서는 데이터의 관리와 활용을 위한 정책을 추진하고 있으며, 더 나아가 데이터의 효율적인 제공과 연계 등 가치 창출을 위한 법·제도적 기반을 마련하고 있다.

한국보건사회연구원은 사회보장 및 보건 의료 영역을 비롯한 각종 사회과학 분야의 데이터를 생산·관리하는 대표적인 연구기관이다. 특히 데이터 기반의 연구 수행이 강조됨에 따라 우리 원에서는 방대한 양의 보건·복지 영역에서의 조사데이터를 생산하고 있으며, 이러한 추세는 향후에도 지속될 것으로 예상된다. 이에 대응하고자 우리 원에서는 「조사자료 관리지침」의 전면 개정을 통하여 조사데이터의 관리체계를 구축하였으며, 조사데이터의 활용 활성화를 위한 보건복지데이터포털의 고도화를 지속 추진하고 있다.

이 연구에서는 한국보건사회연구원의 조사데이터 관리 체계를 중심으로 보건복지분야 조사데이터의 체계적 관리와 활용 방안을 제안하고 있다. 특히 우리 원에서 생산한 조사데이터의 관리 현황을 검토함으로써 조사데이터의 체계적 관리를 위한 실질적 과제를 제안하였다는 점에서 의미가 있다. 그 외에도 조사데이터의 관리와 제공과 관련된 법·제도적 환경 변화를 반영하여 조사데이터 관리를 위한 우리 원의 정책 방향을 설정하였다는 점에서 가치를 가진다.

이 연구는 박성준 부연구위원의 주관으로 이루어졌으며, 신정우 연구위원과 조용찬 연구원의 참여로 수행되었다. 또한 우리 원 조사데이터 현황 분석을 위한 데이터 수집에 관련 조직 및 담당자의 기여가 있었다. 그 밖에도 조사데이터 관리와 관련된 타 연구기관의 전문가 및 담당자의 자

문과 의견수렴 등이 연구결과 도출에 반영되었다. 이상 본 연구에 도움을 준 많은 분들에게 감사 인사를 드린다.

이 연구결과는 한국보건사회연구원에서 생산·관리하고 있는 조사데이터의 체계적 관리와 활용의 활성화에 기여할 것으로 예상된다. 더 나아가 우리 원이 사회과학 분야의 데이터 기반 연구추진을 주도적으로 이끌어 갈 수 있는 기반을 다질 수 있을 것으로 기대한다. 마지막으로, 이 보고서의 내용은 연구진의 개별적 연구활동에 따른 결과물이며, 우리 원의 공식적인 의견이 아님을 밝힌다.

2023년 7월

한국보건사회연구원 원장

이 태 수



목 차

KOREA INSTITUTE FOR HEALTH AND SOCIAL AFFAIRS



Abstract	1
요 약	3
제1장 서 론	15
제1절 연구의 배경 및 목적	17
제2절 연구의 내용 및 방법	20
제2장 조사데이터 관리 및 사례 분석	23
제1절 조사데이터 관련 정책 동향 분석	25
제2절 조사데이터 관리 및 활용 사례 검토	68
제3절 소결	90
제3장 조사데이터 관리 실태 분석:	
한국보건사회연구원 조사자료 관리 사례	93
제1절 조사자료 관리 지침 분석	95
제2절 조사데이터 관리 현황 분석	99
제3절 소결	123
제4장 조사데이터의 체계적 관리 및 활용 제고를 위한 방안	125
제1절 조사데이터 관리체계 개선 방향	127
제2절 조사데이터의 체계적 관리 방안	133
제3절 조사데이터의 활용 활성화 방안	170



참고문헌 185

부 록 **189**

부록 1 기관위원회 평가·인증 기준 189

부록 2 통계작성 단계별 체크리스트 192

표 목차



〈요약 표 1〉 조사데이터 관리체계에 따른 각 주체별 역할 요약	10
〈요약 표 2〉 조사데이터 큐레이션 도입을 위한 주체별 역할 설정(안)	12
〈표 2-1〉 OECD주요 국가의 연구데이터 소유·보호 정책 동향	27
〈표 2-2〉 국내 오픈사이언스 관련 주요 정책 목록	29
〈표 2-3〉 연구데이터의 종류	32
〈표 2-4〉 공공데이터 관리지침 조문 체계 구성	35
〈표 2-5〉 개별 포털 등 구축·운영 허용 기준	37
〈표 2-6〉 정보공개제도의 필요성	44
〈표 2-7〉 정보의 사전적 공개: 주요 정책 정보	47
〈표 2-8〉 공공기관위원회 인간대상연구 조사연구용 연구계획서 권고서식 구성	54
〈표 2-9〉 연구데이터 관리 절차	58
〈표 2-10〉 DMP작성 양식 예시	62
〈표 2-11〉 DMP검토 체크리스트	63
〈표 2-12〉 조사통계 진단 지침 사항 요약	65
〈표 2-13〉 연구데이터 리포지터리 운영 현황	69
〈표 2-14〉 데이터 공유 플랫폼 운영 현황	71
〈표 2-15〉 경제·인문사회연구회 소속 연구기관 데이터 제공 웹서비스 현황	73
〈표 4-16〉 조사데이터 관리 사례 요약 구분	91
〈표 3-1〉 연도별, 과제구분별 조사데이터 현황	100
〈표 3-2〉 연도별 조사명 내 단어 빈도(2016년-2019년)	102
〈표 3-3〉 연도별 조사명 내 단어 빈도(2020년-2021년)	103
〈표 3-4〉 연도별 조사예산 현황	107
〈표 3-5〉 연도별 조사데이터 유형 구분	109
〈표 3-6〉 연도별 조사데이터 조사방법 세분화(중복응답)	111
〈표 3-7〉 과제구분별 조사예산 현황	113
〈표 3-8〉 과제유형별 조사데이터 유형 구분	115
〈표 3-9〉 과제유형별 조사데이터 조사방법 세분화(중복응답)	116

〈표 3-10〉 과제구분별 조사정보 현황	118
〈표 3-11〉 과제구분별 IRB심의 및 자료 이관 여부 현황	119
〈표 3-12〉 과제구분별 비공개 처리 사유 세분화(중복응답)	120
〈표 3-13〉 데이터 유형별 비공개 처리 사유 세분화(중복응답)	122
〈표 4-1〉 관리 대상 조사데이터의 범위	128
〈표 4-2〉 조사자료 관리위원회와 생명윤리위원회 비교	145
〈표 4-3〉 조사데이터 관리체계에 따른 각 주체별 역할 요약	157
〈표 4-4〉 통계 데이터 품질진단 차원	162
〈표 4-5〉 데이터 에디팅 기능	165
〈표 4-6〉 데이터 에디팅 구분 및 기능	166
〈표 4-7〉 조사데이터 큐레이션 도입을 위한 주체별 역할 설정	181

그림 목차

KOREA INSTITUTE FOR HEALTH AND SOCIAL AFFAIRS



[요약 그림 1] 조사데이터 관리 방향 설정	7
[요약 그림 2] 조사데이터 제공 단계 세분화(안)	8
[요약 그림 3] 조사데이터 생성 및 관리 체계에 따른 데이터 생성자의 업무 절차	9
[요약 그림 4] 한국보건사회연구원 조사데이터 관리 조직 구성(안)	11
[그림 1-1] 연구의 구성 체계	21
[그림 2-1] 기계 판독이 가능한 형태의 충족 조건	38
[그림 2-2] 개인정보 등 비공개 대상 정보 확인 절차	39
[그림 2-3] 저작권 등 제3자 권리 확인 절차	41
[그림 2-4] 공공데이터 현행화 절차	41
[그림 2-5] 공공데이터 이용불편사항 신고 및 처리	42
[그림 2-6] 공공데이터제공 분쟁조정 처리 절차	43
[그림 2-7] 개인정보·가명정보·익명정보의 비교	48
[그림 2-8] 생명윤리법에 따른 연구대상자의 정보공개 청구 처리 절차 흐름	51
[그림 2-9] 기관위원회 평가 영역	52
[그림 2-10] 기관위원회의 기능	53
[그림 2-11] 연구과제 기획 및 선정 단계에서의 DMP 작성 및 검토	60
[그림 2-12] 연구과제 수행·평가 단계의 DMP 점검	61
[그림 2-13] 한국여성정책연구원 통계결과 및 원시자료 기본정보 양식	78
[그림 2-14] 한국교통연구원 조사수집 자료의 내역 양식	81
[그림 2-15] 한국형사·법무정책연구원 조사자료 데이터베이스 원표	83
[그림 2-16] 한국사회과학자료원 데이터 관리 단계	85
[그림 2-17] 자료 유형별 수집 형식	86
[그림 2-18] 민감 자료의 공유 절차	88
[그림 2-19] 기탁 자료 준비 및 제출 체크리스트	89
[그림 3-1] 연도별, 과제구분별 조사데이터 현황	101
[그림 3-2] 연도별 조사명 워드클라우드 분석 결과	104
[그림 3-3] 2016년~2022년 조사 키워드 간 연결 빈도 시각화 결과	106

[그림 3-4] 연도별 조사예산 현황	108
[그림 3-5] 연도별 조사데이터 유형 구분	109
[그림 3-6] 연도별 조사데이터 조사방법 세분화	112
[그림 3-7] 과제구분별 조사예산 현황	114
[그림 3-8] 과제구분별 조사데이터 유형 구분	115
[그림 3-9] 과제구분별 조사데이터 조사방법 세분화	117
[그림 3-10] 과제구분별 비공개 처리 사유 세분화	121
[그림 3-11] 데이터 유형별 비공개 처리 사유 세분화(중복응답)	122
[그림 4-1] 조사데이터 관리 및 활용 처리 절차	129
[그림 4-2] 조사데이터 관리 방향 설정	132
[그림 4-3] 조사데이터 관리 방향 설정	135
[그림 4-4] 조사데이터 제공 단계 세분화(안)	138
[그림 4-5] 기존 자체과제의 데이터 생산 및 관리 절차	140
[그림 4-6] 모형1: 조사자료 관리위원회 실질화 모형(안)	142
[그림 4-7] 모형2: IRB강화 모형(안)	144
[그림 4-8] 조사데이터 생성 및 관리 체계에 따른 데이터 생성자의 업무 절차	147
[그림 4-9] 기존 수탁과제의 데이터 생산 및 관리 절차	149
[그림 4-10] 수탁과제 관리 모형: 조사자료 관리위원회 실질화 모형 기준	151
[그림 4-11] 주체별 조사데이터 관리 체계(안)	156
[그림 4-12] 한국보건사회연구원 조사데이터 관리 조직 구성(안)	160
[그림 4-13] 공공데이터의 품질진단 방법	163
[그림 4-14] 데이터 에디팅 절차 예시	167
[그림 4-15] 데이터 생애주기를 고려한 데이터 큐레이션 모델	174
[그림 4-16] 데이터 큐레이션의 순차적/비정기적 활동 관계	177



Abstract

A Study on the Management and Utilization of Health and Welfare Survey Data

: Focusing on Survey Data Management Cases in KIHASA

Project Head: Park, Seongjun

To enhance the use and efficient management of research data in the field of health and welfare, this study examined the management and utilization of survey data produced by the Korea Institute for Health and Social Affairs (KIHASA). We analyzed policies and regulations concerning open science and open data. We also reviewed the cases of research institutes that manage open data.

By analyzing the current status of survey data management at KIHASA, we derived future tasks for the management and utilization of survey data. First, as the management and utilization of research data is a responsibility stipulated in the relevant legal framework, a system of data management and utilization needs to be established with a long-term perspective. Second, segmented management and processing criteria should be established that are attuned to the characteristics of different survey datasets. Third, the transparency of the data management system should be secured based on internal and ex-

2 보건복지 조사데이터의 체계적 관리 및 활용 방안 연구

ternal responses to the demand for the provision of research data. Fourth, by improving the accessibility of data, the utilization of research data produced by KIHASA should be promoted.

Keyword : Health and Welfare Data, Survey Data Management, Research Data Utilization, Open Data



1. 연구의 배경 및 목적

IT기술 및 환경의 변화, 각종 정보 활용의 중요성이 강조되고 있는 가운데 데이터에 기반을 둔 정책수립 및 연구수행이 증가하고 있다. 이와 함께 보건·사회보장 영역 또한 다양한 데이터 기반의 정책 추진 및 서비스 제공이 이루어지고 있다. 각 분야의 연구기관에서는 연구데이터에 대한 관리와 활용에 대한 연구를 활발하게 진행하고 있으며, 데이터 관리 및 제공을 위한 플랫폼 구축, 빅데이터 연계 기반 마련, 데이터 거버넌스 체계 도입 등의 시도를 하고 있다.

한국보건사회연구원에서는 보건·복지 영역에서 방대한 양의 조사데이터를 생산하는 대표적인 연구기관이며, 관련 데이터에 대한 관리와 활용에 대한 의무를 가진 공공기관이다. 이에 따라 한국보건사회연구원은 「조사자료 관리지침」의 전면 개정을 통하여 조사데이터의 관리체계를 구축하고자 하였으며, 조사데이터의 활용을 활성화하기 위한 노력을 지속하고 있다. 그러나 본 기관에서 생산된 조사데이터의 체계적인 관리와 보건·사회보장 영역에서의 데이터 활용 활성화를 위한 과제는 여전히 남아 있다.

이상의 논의에 따라 본 연구에서는 1) 한국보건사회연구원에서 생산하는 보건복지분야 조사데이터의 체계적 관리를 위한 방안을 모색하고자 하였으며, 2) 보건복지분야에서의 조사데이터 활용도를 제고하기 위한 선행과제가 무엇인지를 탐색하고자 하였다.

2. 주요 연구결과

가. 조사데이터 관련 정책 배경 및 사례

조사데이터는 연구수행과정에서 생성된 데이터의 하나로, 각종 설문 및 인터뷰 등 다양한 형태의 조사방법을 적용함으로써 생성된 자료를 의미하여, 조사표·설문지·부호화지침서·녹취록·녹음파일 등 양적·질적 데이터를 포함한 조사의 전 과정에서 생성된 다양한 형태의 자료 일체를 뜻한다. 조사데이터의 관리와 활용에 대한 필요성은 국내외 오픈사이언스 정책의 확산을 통하여 강조되고 있으며, 연구데이터의 공유 및 활성화 전략에 영향을 받고 있다.

조사데이터의 관리 및 활용과 관련된 다양한 법적 근거들이 존재하며, 각각의 법령 및 지침을 통하여 정보공개를 통한 국민의 알 권리 보장 및 기관 운영의 투명성 제고, 개인정보의 보호 및 관련 정보의 활용 방안 마련, 인간대상 연구에서의 생명윤리 보호, 통계의 정확도 및 신뢰도 제고, 연구개발정보의 처리와 체계적 관리 등 연구기관의 의무 사항들을 도출할 수 있다. 이에 따라 한국보건사회연구원은 조사데이터의 체계적 관리와 제공에 대한 의무를 이행해야 하며, 조사대상의 보호와 제공 데이터의 신뢰도 제고에 대한 책임 또한 이행할 필요가 있다.

한편, 여러 분야의 연구기관들은 조사데이터의 관리 및 활용을 위한 플랫폼 구축과 데이터관리 체계 도입, 데이터 활용을 위한 빅데이터 연계 기반 마련 등의 노력을 하고 있다. 이와 더불어 조사데이터의 체계적 관리를 위하여 1) 데이터 관리를 전담하는 부서 또는 인력을 배치하고, 2) 조사데이터 계획-생산-제출 과정에서의 절차를 규정하며, 3) 조사데이터에 대한 이관 및 제공에 대한 주요 사항을 심의하는 조직을 구성하는 등

의 시도를 지속하고 있다. 그 외에도 데이터의 수집 및 이관, 관리 등에 대한 의무 사항을 관련 규정에 포함하여 데이터 누락을 최소화하기 위한 노력을 공통적으로 하고 있다.

나. 한국보건사회연구원 조사데이터 관리 실태

한국보건사회연구원(2022a)은 「조사자료 관리지침」에 근거하여 원내 조사데이터의 관리와 관련 행정사항 등을 규정하고 있다. 구체적으로 조사자료의 관리와 제공 등에 관한 사항을 심의하기 위한 별도의 위원회 구성을 규정하고 있으며, 조사데이터 이관과 제공을 위한 데이터 처리 필요성을 명문화하였다. 또한 조사자료 수집 및 관리를 위한 각종 절차와 양식, 체크리스트 등을 구체화하였으며, 특히 데이터관리계획의 도입을 시도하였다는 점에서 의의가 있다. 한편, 조사데이터 관리와 관련된 실무사항의 구체화, 데이터 이관에 요구되는 각종 데이터 처리 방안, 개인정보 및 민감정보에 대한 보호 및 대응 방안 등에 관한 사항들이 추가적으로 보완될 필요가 있다.

한국보건사회연구원은 매년 80건에 이르는 조사데이터를 생산·관리하고 있다. 본 원은 보건·사회보장 영역에서의 인간대상연구를 수행하는 대표적인 기관으로 해당 영역에서의 각종 실태조사, 인식조사 등의 데이터를 생산하고 있다. 또한, 노인·아동·장애인·청년 등 주요 사회 정책의 대상과 코로나·보건의료 등 주요 사회 현안에 대한 조사데이터가 생산·축적되고 있다.

한국보건사회연구원에서 생산·축적하고 있는 조사데이터의 관리 현황을 바탕으로 다음의 사항을 검토할 필요가 있다. 첫째, 데이터의 공개 가능성이 높은 조사데이터를 중심으로 제공범위를 점진적으로 확대하는 방

안을 검토해볼 수 있다. 예를 들어 조사 규모가 큰 자체 연구과제의 경우에 데이터 제공의 가능성이 높다고 볼 수 있다. 둘째, 조사데이터 관리를 위한 사전/사후 조사정보의 수집이 강화될 필요가 있다. 또한, 이를 위한 행정적 지원이 뒷받침될 필요가 있다. 셋째, 데이터 제공 및 이관 사례를 지속적으로 축적하여 데이터 생산 및 관리에 참고할 수 있는 정보가 누적될 수 있도록 해야 한다. 특히 관리위원회를 통한 심의 결과나 이관된 데이터의 개요 등이 공유될 필요가 있다.

다. 조사데이터의 체계적 관리 및 활용

한국보건사회연구원 조사데이터의 체계적 관리를 위한 방향성을 설정하는 것이 중요하며, 여기에는 관련 법률과 원내 지침에 근거를 둘 필요가 있다. 먼저, 조사데이터 생산의 참여 주체이자 데이터의 이용자인 국민은 데이터에 대한 접근 및 청구 권한을 가지며, 자신의 정보가 동의에 따라 적절하게 보호받을 권리를 가진다. 한편, 한국보건사회연구원은 공공데이터의 제공과 데이터 품질관리에 대한 의무를 가지며 인간대상연구기관으로서의 개인정보 보호에 대한 책임을 가진다. 이와 더불어 본 연구원은 공적 자원을 활용하여 생산된 데이터가 보건·사회보장 분야의 발전과 국민의 복지증진에 기여할 수 있도록 그 가치를 제고하는 데 지속적으로 노력할 필요가 있다.

[요약 그림 1] 조사데이터 관리 방향 설정



자료: 저자 작성

조사데이터의 내·외부적 활용가치를 제고하기 위해서는 연구원에서 생산되고 있는 조사데이터 특성을 고려하여 데이터 제공 단계를 세분화할 필요가 있다. 제공 단계는 크게 1) 비공개 관리, 2) 정보공개/목록관리, 3) 결과표 공개, 4) 마이크로데이터 제공으로 구분한다. 한편, 제공 단계의 상향을 위해서는 조사데이터 내의 개인정보, 민감정보, 일시적으로 수집된 개인식별정보, 연구목적의 비식별 정보가 구분될 필요가 있으며, 각 정보의 특성에 따라 적절한 처리를 통하여 관리·제공될 필요가 있다.

8 보건복지 조사데이터의 체계적 관리 및 활용 방안 연구

[요약 그림 2] 조사데이터 제공 단계 세분화(안)

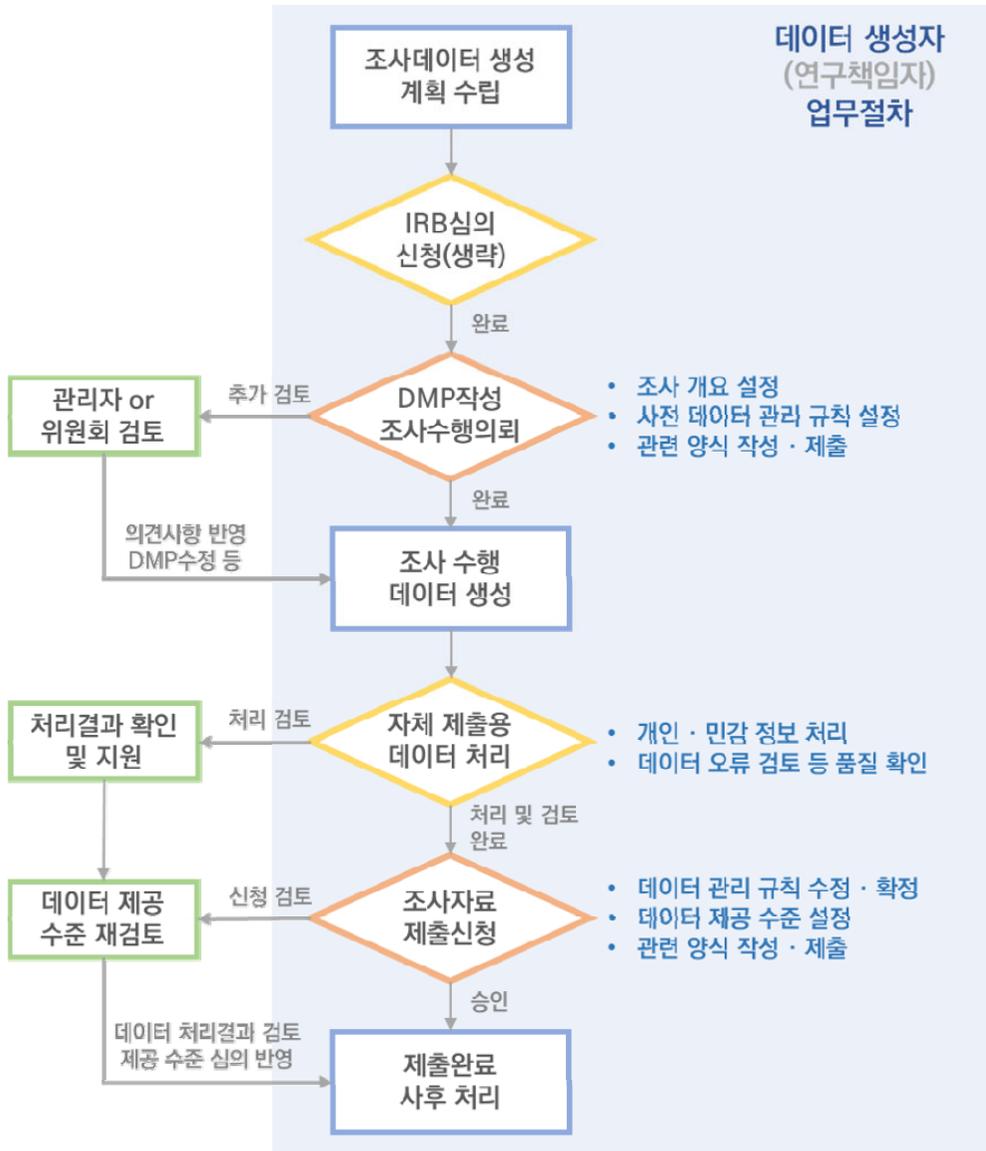
(기존) 비공개/공개 → (개선) 4단계 세분화

관리단계	(1단계) 비공개 관리	(2단계) 정보공개 / 목록 관리	(3단계) 결과표 공개	(4단계) 마이크로데이터 제공
처리	개인정보파기, 익명화 민감정보 가공	(좌동) 메타정보 수집 데이터 관리 관련 정보 축적	(좌동) 보고서 내 결과표 제출 기타 추가 기술표 작성	민감정보 가공 및 엠비고 설정 내검기록 관리
수행업무	<p>관리 및 정보보호</p> <ul style="list-style-type: none"> 비공개 설정 절차 관리 조사계획(안내 등) 이행 점검 보관 대상 여부에 대한 판단 <p>유의사항</p> <ul style="list-style-type: none"> 연구자 개인 저장매체 내 임의적 관리 지양 단, 책임자의 권한(자율성) 부여 	<p>조사정보 공개</p> <ul style="list-style-type: none"> 조사 관련 사항에 대한 공개 내부활용 제고를 위한 정보 투명화 (내부자료 활용, 유사중복 방지, 조사계획 수립 시 참고 자료 등) <p>유의사항</p> <ul style="list-style-type: none"> 원자료는 1단계에 준하여 관리 관리부서와의 협의를 통하여 메타데이터의 정확도 제고 필요 	<p>조사결과 활용</p> <ul style="list-style-type: none"> 보고서에 포함된 결과표에 대한 활용 편의성 제고 각종 보도자료 등 연구성과 확대 전략에 활용 권고 <p>유의사항</p> <ul style="list-style-type: none"> 추가 기술표 작성 시 교차분석 결과에 대한 부서 검토 필요 데이터 요청에 따라 표 추가 제공 	<p>마이크로데이터의 제공</p> <ul style="list-style-type: none"> 개인정보처리 원료, 원자료와 유사한 형태의 데이터 제공 원자료에서의 변동사항 공지 데이터 요청에 대한 사항 관리 <p>유의사항</p> <ul style="list-style-type: none"> 데이터 제공 관련 업무의 일원화 데이터 수정 요청, 내검기록 관리 연구자의 업무과중 사항 고려
활용대상	원내 관리자 연구책임자, 발주처	관리부서, IRB 원내 구성원 전체	(좌동) 원외 연구자, 국민	(좌동) 원외 연구자, 국민
주요대상 (예시)	수탁과제 조사데이터	표적집단 대상, 소수 참여 특정 집단에 대한 민감정보 수집	표적집단 대상(다수) 품질 문제 발생 데이터	민감성 낮은 표적집단 대상(다수) 표집 조사(신뢰수준, 오차 고려)

자료: 저자 작성

조사데이터의 관리체계 구성 시 데이터 생성자 중심의 절차를 구성하면 아래 그림과 같다. 연구책임자는 조사데이터 생성을 위한 계획 수립과 함께 조사데이터에 대한 관리 계획을 작성한다. 조사데이터 생성 이후 연구책임자는 개인정보 및 민감정보가 처리된 제출용 데이터를 제출해야 하며, 데이터의 품질관리를 위한 기본적인 사항을 검토한다. 이때 조사데이터 관리 주체는 추가적인 데이터 처리의 필요성이나 개인·민감정보 처리의 적절성 등을 검토한다. 연구책임자와 조사데이터 관리 주체 간 협의를 통하여 데이터 제공 수준을 결정하고 사후 작성된 데이터관리계획에 근거하여 조사데이터를 관리·제공하게 된다.

[요약 그림 3] 조사데이터 생성 및 관리 체계에 따른 데이터 생성자의 업무 절차



자료: 저자 작성

조사데이터의 관리 절차에 따라 관련 주체들의 역할 설정이 요구된다. 본 연구에서 제안한 조사데이터 관리 주체는 크게 데이터 생성자, 데이터 관리자, 관련 위원회로 구분되며, 각 단계별 역할은 아래 표와 같다.

10 보건복지 조사데이터의 체계적 관리 및 활용 방안 연구

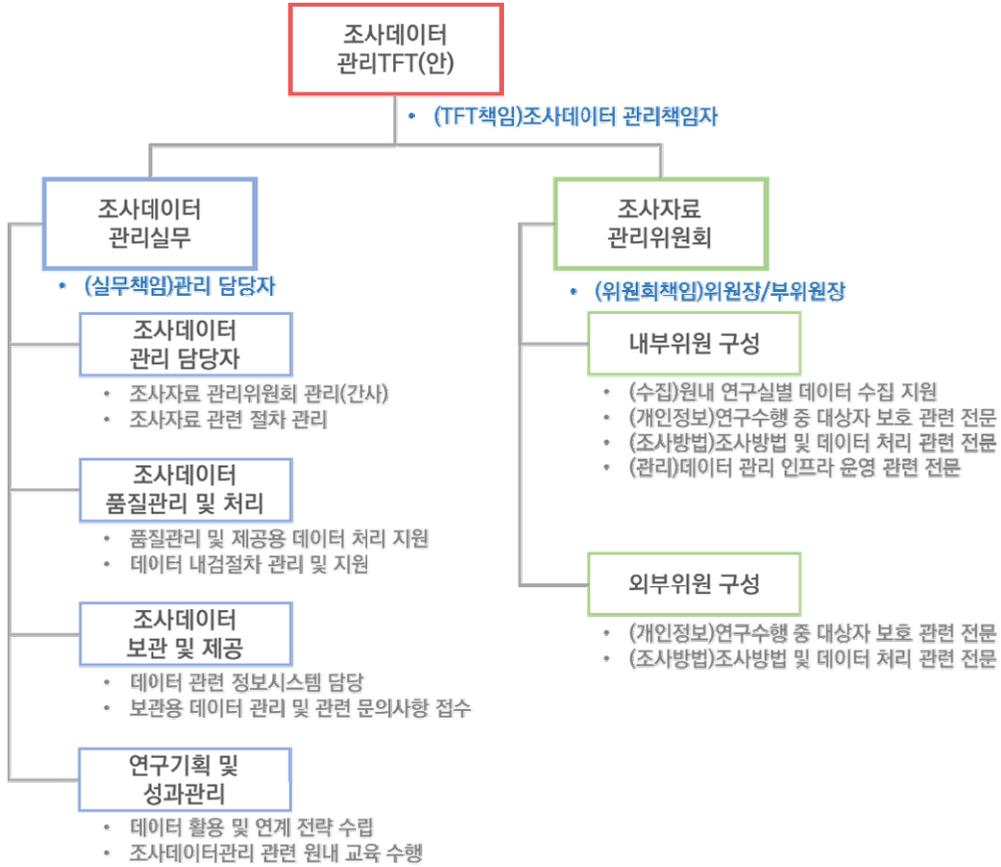
〈요약 표 1〉 조사데이터 관리체계에 따른 각 주체별 역할 요약

절차	데이터 생성자 (원내 연구진)	데이터 관리자 (관련 전담 조직)	관련 위원회 (관리위원회 등)
기획·계획	<ul style="list-style-type: none"> 조사기획, 데이터 관리 계획 작성 조사데이터 관리 관련 세부사항 검토 	<ul style="list-style-type: none"> 조사데이터 관리체계 안내 및 절차 추진 조사관련 사항 지원 및 협의 	<ul style="list-style-type: none"> (필요시)데이터 관리 계획에 대한 자문 (필요시)데이터관리 계획에 대한 사전심의
생산·종료	<ul style="list-style-type: none"> 조사데이터 수집 및 분석 수행 조사데이터 및 관련 자료에 대한 제출 	<ul style="list-style-type: none"> 데이터 생산 관련 자문 데이터 및 관련 자료 수집, 보완 요청 	<ul style="list-style-type: none"> (발생 시)조사 과정 중 연구윤리 및 개인정보 보호 관련 이슈 대응
수집·축적	<ul style="list-style-type: none"> 데이터 제출 전 오류 수정 및 처리 이행 데이터관리계획 이행 및 제공 범위 설정 	<ul style="list-style-type: none"> 데이터 보관·제공 등급 검토 데이터 처리 결과 재검토 및 수정 요청 데이터 목록화, 내검기록, 위원회 심의안건 제출 	<ul style="list-style-type: none"> (안전)심의 안전에 대한 검토 및 의견 제출 조사데이터 관리 현황 등에 대한 모니터링
활용·현행화	<ul style="list-style-type: none"> 데이터 현행화 및 수정 등에 대한 대응 	<ul style="list-style-type: none"> 데이터 제공 지원 데이터 현행화 및 내검 등에 대한 지원 데이터 연계 등 활용 방안 활성화 	<ul style="list-style-type: none"> (요청 시)데이터 제공 관련 심의 안건 처리 조사데이터 현행화 및 수요자 대응 관련 점검

자료: 저자 작성

조사데이터의 체계적 관리와 활용을 위해서는 관련 지침 설정 외에도 조직 및 인력 구성이 필수적이다. 본 연구에서는 조사데이터의 관리를 위한 조직 및 인력 구성을 아래 그림과 같은 제안하였다.

[요약 그림 4] 한국보건사회연구원 조사데이터 관리 조직 구성(안)



주: 조직 구성 및 역할의 내용은 가안이며, 상황에 따라 수정·적용될 수 있음.
 자료: 저자 작성

한편, 한국보건사회연구원 조사데이터의 외부 활용 가능성을 확대하기 위한 방안으로 조사데이터 큐레이션 체계 도입을 제안하였다. 데이터 큐레이션이란 데이터의 생애주기(수집, 관리, 보존, 활용)에 따라 데이터를 적극적으로 관리·보존하고 그 가치를 부여하여 재이용될 수 있도록 하는 일련의 과정을 의미한다. 데이터 큐레이션 도입의 주요 목적은 1) 기존 데이터의 적극적인 관리를 통하여 연구결과를 보존하고, 2) 데이터 중복으로 인한 관리 및 활용 노력의 비효율성을 감축시키며, 3) 데이터의 공유와 활용을 촉진하여 그 가치를 향상시키는 데에 있다.

12 보건복지 조사데이터의 체계적 관리 및 활용 방안 연구

데이터 큐레이션은 데이터의 생성 및 수집, 평가 및 선정, 입수·보존·저장, 이용 및 재이용, 변환 등의 일련의 과정을 통하여 이루어지며 이를 위한 연구원 내 여러 주체들의 기여가 요구된다.

〈요약 표 2〉 조사데이터 큐레이션 도입을 위한 주체별 역할 설정(안)

영역	역할	주체(예시)
데이터 생성자	<ul style="list-style-type: none"> • 주요 조사정보의 입력 및 확인 • 데이터 보관 및 제공을 위한 전처리 • 데이터 활용 범위 설정 및 소유권 사항 검토 • 데이터 활용 계획 및 큐레이션 등에 대한 자문 	연구책임 또는 조사설계
데이터 관리자 (실무)	<ul style="list-style-type: none"> • 주요 조사 정보에 대한 검증 및 데이터화 • 데이터 보관 및 제공의 처리 및 기록 • 데이터 큐레이션 과정에 대한 추진 및 관리 • 데이터 큐레이션 관련 심의 결과 반영 	조사데이터 관리 실무자
데이터 관리자 (책임)	<ul style="list-style-type: none"> • 데이터 큐레이션 정책 방향 설정 및 도입 • 조사자료 관리위원회 주요 심의 사항 결정 • 데이터 큐레이션 절차 추진에 대한 총괄 관리 • 데이터 큐레이션 추진 시 관련 주체의 역할 조정 	원내 데이터 관리 책임
조사자료 관리위원회	<ul style="list-style-type: none"> • 데이터 큐레이션 도입결과에 대한 검토 • 조사데이터 활용 관련 주요 이슈에 대한 심의 • 내외부 위원 및 외부 전문가를 활용한 의견 반영 및 개선사항 논의 	위원장 및 부위원장 중심
정보기술 영역	<ul style="list-style-type: none"> • 데이터 보안 및 안전 관리를 위한 기술 도입 • 자동화 도입 시 시스템의 관리 및 점검 • 데이터 마이그레이션 등 기술 지원 	각 부서 내 관련 업무 수행 인력 중심
학술정보 영역	<ul style="list-style-type: none"> • 조사데이터와 관련 문헌 간 연결 • 국내외 정보기관 네트워크 관리 및 협력 	
대외홍보 영역	<ul style="list-style-type: none"> • 데이터 큐레이션 도입의 홍보 • 데이터 큐레이션 개선을 위한 수요 검토 • 기타 타 기관 연계 방안 적용 등 	
기타 주체	<ul style="list-style-type: none"> • (정보통계영역)데이터 품질 관리 등 지원 • (행정·관리영역)데이터 큐레이션 도입 및 개선을 위한 제도적 사항 검토 • (패널·실태조사 수행영역)주요 활용 데이터에 대한 큐레이션 방안 제안 및 실적 관리 등 	영역별 관련 업무 참여 인력 중심

자료: 저자 작성

3. 결론 및 시사점

본 연구에서는 한국보건사회연구원에서 생산하는 보건복지분야 조사데이터의 체계적인 관리와 활용 활성화를 위한 방안에 대하여 검토하였다. 본 연구의 결과를 통하여 논의된 조사데이터의 관리 및 활용에 관한 주요 시사점을 정리하면 아래와 같다.

첫째, 조사데이터의 관리 및 활용은 관련 법적·제도적 근거에 따른 의무사항이며, 이를 위한 체계 구축은 장기적인 관점에서 추진될 필요가 있다. 한국보건사회연구원은 정보공개를 통한 국민의 알 권리 보장 및 기관 운영의 투명성 제고, 개인정보의 보호 및 관련 정보의 활용, 인간대상연구에서의 생명윤리 보호, 통계의 정확도 및 신뢰도 제고, 연구개발정보의 처리와 체계적 관리 등 연구기관으로서의 의무를 가진다. 따라서 조사데이터의 체계적 관리를 위한 지침 마련, 인적·물적 자원의 활용, 데이터 활용도 제고를 위한 관련 제도 도입 등에 지속적인 노력이 요구된다.

둘째, 조사데이터의 특성을 고려한 세분화된 관리 방안이 마련될 필요가 있다. 조사데이터 내 개인정보·민감정보에 대한 처리, 조사방법 및 조사수행관련 정보 등 조사개요에 대한 정보 수집, 개인이 식별되지 않은 연구 목적의 데이터와의 분리 등이 이루어질 때 조사데이터의 보관, 제공이 보다 원활하게 이루어질 수 있다. 또한, 조사데이터의 관리 현황과 기존 조사데이터의 개요 등의 정보가 원내에 공유되어 내부적 활용 가능성을 제고할 필요가 있다.

셋째, 조사데이터의 관리 체계 투명화를 통하여 정보공개 및 데이터 관리에 대한 내·외부적 대응력 강화가 필요하다. 데이터의 공개/비공개 기준의 설정과 공식화, 민감정보에 대한 기관 차원의 관리체계 도입 등은 데이터를 생산한 연구자가 가지는 책임소재를 분산시키는 기능을 한다.

14 보건복지 조사데이터의 체계적 관리 및 활용 방안 연구

이는 내·외부적인 정보공개 요청이나 조사참여자의 철회 대응, 연구 목적 외 데이터 활용, 데이터 오류에 대한 등에 대한 사안에 대해 연구자를 보호할 수 있는 연구원 차원의 대응력 강화 방안이라 볼 수 있다.

넷째, 조사데이터의 관리 체계 운영에 기반을 둔 데이터 활용 활성화 전략이 요구되며, 점진적인 데이터 활용 범위 확대를 통하여 생산된 데이터의 가치를 재생산하기 위한 노력을 지속해야 한다. 조사데이터를 생산하고 관리하는 주체는 데이터의 공적 가치 확산에 대한 의무를 가진다. 장기적인 관점에서 조사데이터의 가치 제고를 위한 목표와 전략 등이 기관의 설립 목적에 맞게 설정되고 순차적으로 이행될 필요가 있다.

주요 용어 : 보건복지데이터, 조사데이터 관리, 연구데이터 활용, 데이터 개방

사람을
생각하는
사람들



KOREA INSTITUTE FOR HEALTH AND SOCIAL AFFAIRS



제 1 장

서론

제1절 연구의 배경 및 목적

제2절 연구 내용 및 방법

제 1 장 서론

제1절 연구의 배경 및 목적

제4차 산업혁명 시대에 이르러 ‘데이터’는 핵심 자원으로 각광받고 있다. 데이터는 미래 경쟁력을 좌우하는 지식자산으로 인식되고 있으며 보건·사회보장 영역을 포함한 모든 연구 분야에서 데이터의 처리와 분석, 활용을 위한 노력들이 지속되고 있다.

그러나 데이터는 그 자체로서의 가치보다는 사회적 맥락을 반영한 데이터의 수집과 가공, 이용 목적에 따른 분석, 다양한 지식 자원과의 연계 등을 통하여 그 가치가 더욱 의미있게 발휘된다. 이에 따라 단순히 데이터의 양을 증대시키기보다는 데이터를 어떻게 관리하고 연계하며, 적절히 활용할 수 있을지에 대한 고민의 시대로 전환되었다.

IT기술 및 환경의 변화, 각종 정보 활용의 중요성이 강조되고 있는 가운데 데이터에 기반을 둔 정책수립 및 연구수행이 증가하고 있다. 각종 통계정보의 활용과 빅데이터를 통한 분석 등 다양한 형태의 데이터 분석 수요가 증대되고 있다. 각 분야별 연구기관, 정부부처 및 공공기관에서는 데이터의 활용 및 연계에 기반한 정책 추진을 강조하고 있으며, 이러한 데이터 자원을 활용한 미래지향적 가치를 창출하는데 집중하고 있다.

이와 함께 보건·사회보장 영역 또한 다양한 데이터 기반의 정책 추진 및 서비스 제공이 이루어지고 있으며, 보건복지데이터를 소유하거나 활용하는 기관 및 연구자들은 데이터의 활용 가치를 제고하기 위한 방안들을 제시하고 있다. 이러한 사회적 변화에 따라 한국보건사회연구원은 보건복지 분야의 데이터를 생산하고 관리하며 그 가치를 창출하는 기관으

로서의 기능을 강화하고자 ‘연구데이터의 체계적 관리를 위한 연구’를 수행하고 있다. 특히 한국보건사회연구원은 방대한 양의 조사데이터를 생산하는 핵심적인 연구기관이기 때문에 데이터에 대한 체계적 관리와 데이터를 활용한 가치 창출이 무엇보다 중요한 과제로 설정될 필요가 있다.

조사데이터를 포함한 연구데이터 관리 및 활용과 관련된 선행연구들을 살펴보면, 한국과학기술정보연구원을 중심으로 연구데이터의 활용과 관련된 국내·외의 정책 변화들이 검토되었다. 2020년 한국보건사회연구원에서는 전사적 연구데이터 관리 체계 마련을 위한 연구(정영철, 정소희, 이기호, 김은주, 진재현, 안수인, 2020)가 수행되었다. 또한 한국교육과정평가원에서는 2020년 이후 연구데이터의 관리와 활용과 관련된 연구를 활발히 수행하고 있다. 지자체 산하 기관인 제주연구원에서는 자체적인 연구데이터플랫폼 구축을 위한 연구(강영준, 이상운, 박근영, 2022)를 수행한 바가 있다. 이 외에도 데이터의 관리와 제공을 위한 플랫폼 도입, 빅데이터 연계를 위한 거버넌스 체계 구축 방안 등이 다양한 분야에서 시도되고 있다.

한국보건사회연구원의 연구데이터 관리체계는 이상의 선행연구에서 빈번하게 검토되고 있으며, 관련 지침은 타 연구기관의 주요 참고사례로 언급되고 있다. 그러나 다방면의 연구데이터에 대한 체계적 관리와 보건복지 영역에서의 데이터를 활용한 적극적 가치창출에는 여전히 그 역량이 부족한 실정이다. 따라서 현재까지 추진된 데이터 관리 체계 구축의 실태를 진단하고 실질적인 데이터 관리 체계의 안착과 데이터 활용 및 연계를 통한 가치 창출을 활성화하기 위한 구체적 과제를 발굴할 필요가 있다.

이상의 논의에 근거하여 본 연구의 목적은 다음과 같이 설정되었다. 첫째, 한국보건사회연구원에서 생산하는 보건복지분야 조사데이터의 체계적 관리를 위한 방안을 모색하고자 하였다. 보건복지 분야의 각종 데이터

가 체계적으로 관리되고 활용될 수 있는 방안을 검토하고, 이를 위해 요구되는 주요 개선과제들이 무엇인지 도출하고자 하였다. 이를 위해 한국보건사회연구원에서 축적하고 있는 조사데이터 관리목록을 분석하였다.

둘째, 보건복지 조사데이터의 활용을 위한 방향을 설정하고자 하였다. 보건복지분야 조사데이터의 활용도 제고를 위한 선행과제가 무엇인지 탐색하였다. 특히 인간대상의 조사데이터가 안전하게 관리됨과 동시에 적극적인 활용을 통하여 관련 정책분야의 발전에 기여할 수 있는 방안들이 모색될 필요가 있다.

제2절 연구의 내용 및 방법

본 연구의 목적은 보건복지분야 조사데이터의 체계적 관리와 활용 방안을 모색하는 것이다. 이에 따라 구체적으로 조사데이터의 체계적 관리에 요구되는 과제를 도출하고 조사데이터 활용도를 제고하기 위해 고려되어야 하는 법적·제도적 환경을 검토하였다.

본 연구의 내용은 크게 1) 조사데이터 관리 및 관련 사례 분석(제2장), 2) 조사데이터 관리 실태 분석(제3장), 3) 조사데이터 체계적 관리 및 활용 제고를 위한 방안(제4장)으로 나눌 수 있다. 먼저 조사데이터 관리 및 사례 분석에서는 조사데이터 관련 국내외 정책 동향을 검토하였으며, 조사데이터의 개념적 범위를 설정하였다. 또한 조사데이터의 관리에 직·간접적 영향을 미칠 수 있는 각종 법적 제도적 배경을 살펴보았다. 이 외에도 연구데이터의 활용 사례와 타 연구기관의 조사데이터 관리 현황 등을 살펴보았다.

조사데이터 관리 실태 분석에서는 한국보건사회연구원 조사자료 관리 사례를 바탕으로 연구원의 조사데이터 관리 실태를 진단하였다. 먼저, 본원의 관련 지침이 가지는 특성을 분석하고 타 기관의 규정 및 지침과의 관련성을 검토하였다. 또한 조사자료 관리목록에 대한 분석을 수행하여 연구원의 데이터 관리 실태 전반을 살펴보고, 데이터의 특성에 따라 구분되는 관리 현황을 비교 분석하였으며 조사데이터의 특성을 고려한 관리체계 개선 방향을 제시하였다.

조사데이터 체계적 관리 및 활용 제고 방안에서는 기존 조사데이터의 관리체계를 기반으로 향후 조사데이터 관리체계의 고도화를 위하여 요구되는 개선방향을 제시하였다. 또한 조사데이터의 체계적 관리와 활용을 촉진하기 위한 세부적인 과제를 도출하고 이에 대한 대응 방안을 제시하

였다.

이상의 연구내용과 연구방법 및 결과는 아래 그림과 같이 구성되었다.

[그림 1-1] 연구의 구성 체계

목적	보건복지분야 조사데이터의 체계적 관리 및 활용 방안 모색		
세부 목표	조사데이터의 체계적 관리를 위한 과제 도출		조사데이터 활용을 제고하기 위한 전략 검토
연구 내용	조사데이터 관리 및 관련 사례 분석	조사데이터 관리 현황 분석 및 주요 과제 도출	데이터의 활용 가능성 제고를 위한 방안 검토
연구 방법	<ul style="list-style-type: none"> 국내외 데이터 관리 정책 동향 분석 조사데이터 관리 및 활용 사례 분석 	<ul style="list-style-type: none"> 조사데이터 관리 실태 검토 조사데이터 관리 현황 분석 조사데이터의 체계적 관리를 위한 주요 과제 도출 	<ul style="list-style-type: none"> 사례분석 및 실태분석의 함의 도출 관련 전문가·수요자 자문을 통한 의견 수렴 데이터 활용도 제고를 위한 제반 사항 검토
연구 결과	<ul style="list-style-type: none"> 데이터 관리 및 활용 관련 정책적 환경을 고려한 개선방안 도출 조사데이터 체계적 관리 방안에 대한 모색 	<ul style="list-style-type: none"> 실태 분석을 통한 세부적인 개선 사항 도출 안정적인 데이터 관리를 위한 관리체계 방향 설정 	<ul style="list-style-type: none"> 조사데이터 관리체계 관련 주요 쟁점 대응 및 개선 방향 조사데이터 활용도 제고를 위한 방향성 제안

자료: 저자 작성



제2장

조사데이터 관리 및 사례 분석

제1절 조사데이터 관련 정책 동향 분석

제2절 조사데이터 관리 및 활용 사례 검토

제3절 소결

제 2 장 조사데이터 관리 및 사례 분석

제1절 조사데이터 관련 정책 동향 분석

1. 조사데이터 관련 정책적 흐름

가. 오픈사이언스 정책의 확산

2000년대 초반부터 데이터 중심의 연구와 관련 데이터의 공유 및 활용의 활성화 노력이 지속되어 왔다. 대표적으로 OECD에서는 2007년 「공공연구데이터 접근성 제고 지침(Principles and Guidelines for Access to Research Data from Public Funding)」을 마련하였으며(신은정, 안형준, 양현채, 최병삼, 양승우, 정원교, 김수연, 2017. 신은정, 손수정, 서지영, 김영린, 2019에서 재인용), 연구데이터(research data)에 대한 개념화와 이에 대한 접근성 제고 원칙을 제시하였다(신은정, 2018. 신은정 외, 2019에서 재인용). 또한, 2013년 G8정상회담에서는 「오픈데이터 헌장(Open Data Charter)」를 채택함으로써 오픈데이터의 5대 이행 원칙을 정립하였다(G8 Open Data Charter and Technical Annex, 2013) 이 원칙에서는, 1) 모든 데이터는 기본적으로 공개되어야 하고(Open Data by Default), 2) 시의적절하고 포괄적이며 정확한 고품질 데이터가 충분하게 제공되어야 하며(Quality and Quantity), 3) 모두가 해당 데이터를 광범위한 목적으로 활용할 수 있도록 지원해야 하고(Useable by All), 4) 데이터의 개방을 통하여 데이터에 대한 시민의 요구와 참여를 촉진·거버넌스의 활성화에 기여하며(Releasing Data for

Improved Governance), 5) 데이터의 활용을 통해 다양한 분야의 혁신과 발전에 기여(Releasing Data for Innovation)해야 함을 명시하고 있다. 또한, 해당 현장의 전문에서는 데이터의 공개를 통하여 정부와 기업의 투명성을 제고하고 정보의 공유를 통한 거버넌스 활성화가 이루어질 것으로 전망하였다. 이에 따라 데이터의 공유를 통하여 데이터에 대한 시민의 욕구를 충족시키고 혁신과 번영을 위한 데이터 자원 활용에 대해 각 국가의 적극적 정책추진을 제안하였다.

OECD 주요 국가들의 오픈데이터 정책 기조를 살펴보면, 아래 표와 같다. OECD 주요 국가들이 연구데이터 관련 정책은 주로 국가 및 연방정부 중심의 하향식 정책이 이루어진다는 특징을 보인다. 이는 오픈데이터 정책의 추진에 중앙정부의 역할을 강조하고, 데이터에 대한 공유와 활용을 가치창출의 중요한 원천으로 보는 관점이 반영된 것이라 볼 수 있다. 그러나 데이터가 하나의 자산으로 이해되면서 이에 대한 보호 전략이 관리 및 공유 전략과 공존하는 특성을 보이고 있다. 다시 말해 연구데이터에 대한 접근성을 강화함과 동시에 데이터 보호의 중요성과 개인정보 및 저작권 보호에 대한 원칙 및 기준들을 함께 규정되고 있다.

〈표 2-1〉 OECD주요 국가의 연구데이터 소유·보호 정책 동향

구분	정책(계획) 명	정책 유형 (하향식)	주관기관	시행기간	특성
미국	연방데이터 전략	전략 (하향식)	미 연방 예산관리실, 상무부·중기청 등	2019~	·연방정부데이터의 전략적 자산화 ·정부데이터의 효과적 구축 및 민간 활용지원 강화
유럽 연합	디지털단일 시장·데이터경제 정책	정책 (하향식)	유럽위원회(EC) DG CONNECT	2014~	·유럽 전지역의 데이터 자유이동과 디지털 전환 준비 ·데이터거래 및 보호관련 조치 점검
일본	과학기술 혁신 기본계획 (오픈사언스 활성화전략 반영)	계획 (하향식)	내각부·문부과학성	2015~	·연구성과 공유시 지식재산 고려한 개방과 폐쇄의 균형 전략 강조 ·일부 국책사업 데이터관리계획 도입
한국	연구데이터 공유활용 전략	전략 (하향식)	국가과학기술 심의회, 과기정통부	2018~	·혁신성장 위한 연구데이터의 지식자산화·빅데이터화 목표 ·연구데이터 시범사업추진 및 R&D관리규정 개정 (데이터관리계획 도입)

자료: 신은정 외(2019). 공공부문 연구데이터의 소유·활용제도 개선방향. 과학기술정책연구원.

나. 국내 오픈사이언스 정책 추진

우리나라에서에서는 2017년을 기점으로 오픈사이언스 정책이 활발히 추진되기 시작하였으며, 2020년 데이터 3법¹⁾ 이행과 2021년 「데이터 산업진흥 및 이용촉진에 관한 기본법」, 제1차 데이터산업 진흥 기본계획 (국가데이터정책위원회, 2023) 등을 통하여 연구데이터에 대한 오픈사이언스 정책이 구체적으로 논의되었다.

우리나라에서는 ‘연구데이터 공유·활용 전략(국가과학기술심의회, 2018)’을 통하여 연구데이터의 관리·공유·활용에 대한 법적·제도적 기반 마련과 관련 플랫폼 및 거버넌스 형성을 추진하고자 하였다. 또한, 「국가연구개발혁신법」 제정과 관련 행정규칙인 「국가연구개발정보처리기준」을 바탕으로 연구데이터에 대한 정의와 데이터관리계획(DMP)의 도입 등에 근거가 마련되었다.

1) 「개인정보 보호법」, 「정보통신망 이용촉진 및 정보보호 등에 관한 법률」, 「신용정보의 이용 및 보호에 관한 법률」

〈표 2-2〉 국내 오픈사이언스 관련 주요 정책 목록

구분	주요 정책	중앙부처	주요 내용
중장기 계획	4차 산업혁명 대응계획(I-KOREA 4.0) (2017~2022)	과학기술정보 통신부	<ul style="list-style-type: none"> · 연구데이터 공유 활용을 제도화 · (단기) 연구데이터 정의, 데이터관리계획 도입 · (중장기) 추진경험을 바탕으로 국가연구데이터 법제 정비개선 · (필요시 개별입법)국가연구데이터 공유·활용 시범사업 추진
	제2차 보건의료기술육성 기본계획 (2018-2022)	보건복지부	<ul style="list-style-type: none"> · 질병관리본부 국립보건연구원을 국가 보건의료 연구자원 관리의 컨트롤타워로 육성하여 연구자에 대한 종합적·체계적 지원 강화 . · 국가 암 정책집행, 연구데이터·인프라 운용기관으로서 산·학·연·병 협력 네트워크의 허브로 육성
	제3차 융합연구개발 활성화 기본계획 (2018-2027)	과학기술정보 통신부	<ul style="list-style-type: none"> · 국가 연구데이터 플랫폼을 활용한 공동 융합연구 활성화 지원 · 국가 연구데이터 플랫폼 활용 융합연구 활성화
	제4차 연구성과 관리·활용 기본계획 (2021-2025)	과학기술정보 통신부	<ul style="list-style-type: none"> · 연구데이터를 포함하여 연구성과 범위를 유연하게 운영 · 국가연구데이터센터를 통해 연구데이터를 수집·개방 · 분야별 데이터 전문센터(바이오, 소재 등)와 연계하여 데이터 활용 활성화 · 공공 연구성과 개방 공유·촉진·접근성 제고를 위한 이해관계자 의견수렴 · 연구자·연구기관의 인식 제고 및 교육 확대
	제4차 기초연구진흥종합 계획 (2018-2022)	과학기술정보 통신부	<ul style="list-style-type: none"> · 연구데이터 등 기초연구 과정에서 발생 되는 다양한 연구성과·정보를 공유·확산 · 연구데이터를 원스탑으로 검색 하고 활용할 수 있는 국가 연구 데이터 플랫폼 구축 운영 · 연구데이터의 체계적인 관리를 위해 국가연구데이터센터 및 과학 기술 분야별 대분야 소분야 전문센터 체계 구축

30 보건복지 조사데이터의 체계적 관리 및 활용 방안 연구

〈표 2-2〉 국내 오픈사이언스 관련 주요 정책 목록(계속)

구분	주요 정책	주요 내용
과학기술 자문회의	연구데이터 공유·활용전략 (2018.1.19)	<ul style="list-style-type: none"> · 국가 연구데이터 공유·활용 활성화를 통한 연구데이터 지식 자산화 및 빅데이터화 · 연구데이터 관리 체계 구축 및 공유 커뮤니티 형성 촉진 · 국가연구데이터 플랫폼 구축·운영 · 데이터 및 컴퓨팅 활용 R&D인재 성장지원 · 연구데이터 관리·공유·활용 관련 법제도 마련 · 연구데이터의 산업적 활용 촉진, 일자리 창출
	오픈데이터·오픈 사이언스 정책 (2018.12.12)	<ul style="list-style-type: none"> · 공공데이터 오픈과 오픈사이언스를 통해 혁신을 가속화하고, 국가연구개발의 투명성, 연구윤리, 국민신뢰 확보 · 국가연구데이터 공유 및 오픈사이언스 기반 구축 · (비전수립과 법제도 마련) 국가연구데이터 법적 정의, 공개 근거 마련 · (저변확대 지원) 오픈액세스 출판, 오픈소스 확산 · 오픈소스 인재 풀 확충을 위한 교육과 커뮤니티 지원

자료: 이민정(2023). 미국의 「오픈사이언스의 해」 선포와 정책적 시사점. 한국과학기술정보연구원.
 주: 해당 자료 내 내용을 일부 요약

가. 조사데이터의 범위

조사데이터란, 설문·인터뷰 등 다양한 조사 방법을 통하여 수집된 자료를 일컫는다. 또한, 조사데이터는 연구데이터의 하위개념이라 볼 수 있다. 여기서 말하는 연구데이터란, 연구개발과제 수행 과정에서 실시하는 각종 실험, 관찰, 조사 및 분석 등을 통하여 산출된 사실 자료로서 연구결과의 검증에 필수적인 데이터를 말한다(「국가연구개발정보처리기준」 제2조 제6호). 해외에서도 이와 유사하게 연구데이터를 정의하고 있다. OECD에서는 “수치·문자·이미지·음성 등 사실적 데이터이며, 과학연구의 원천으로 사용되는 연구결과의 검증에 필수적인 것”으로 정의하고 있다. 또한 미국정보표준화기구에서는 연구의 전 과정에서 발생하는 데이터로서 연구자가 수집할 정보와 수집 방법·정보처리 분석 계획 모두를 포함하는 개념으로 설명하고 있다(이상환, 2019).

한편, 통계청(2019) 예규 가운데 「지방자치단체의 통계 작성·보급 및 관리에 관한 기본 지침」에서는 통계 업무의 용이성과 통일성을 제고하기 위하여 관련 용어들을 정의하고 있다. 해당 예규에 따르면 “조사통계”란 통계 작성을 목적으로 통계작성 기법을 사용하여 조사한 자료를 통해 작성된 통계를 말한다. 이 외에도 “원자료(Raw Data)²⁾”, “마이크로데이터(Micro Data)³⁾”, “메타데이터(Meta Data)⁴⁾”에 대한 정의를 제시하고 있다.

-
- 2) 통계자료에서 최초 입력한 전산파일 자료로, 입력오류, 조사오류 등이 수정되기 이전 단계의 자료를 말함. 통계청(2019). 지방자치단체의 통계 작성·보급 및 관리에 관한 기본지침. 통계청.
 - 3) 통계조사 원자료에서 입력오류 등을 제거한 기초데이터로, 통계작성 등 데이터 가공에 사용되는 자료를 말함. 통계청(2019). 지방자치단체의 통계 작성·보급 및 관리에 관한 기본지침. 통계청.
 - 4) 통계자료 이용자의 이해를 위한 통계에 대한 설명자료(통계 개요, 조사표, 코드표 등). 통계청(2019). 지방자치단체의 통계 작성·보급 및 관리에 관한 기본지침. 통계청.

한국보건사회연구원(2022a) 「조사자료 관리지침」 제2조 제1호에 따르면 ‘조사자료’란, 연구원의 연구·사업 수행을 목적으로 연구원 내·외부의 면접조사, 전화조사, 우편조사 및 웹조사 등을 통하여 생산된 기계 판독이 가능한 모든 형태의 자료를 의미한다. 즉, 해당 지침에서는 연구를 위한 각종 조사 과정에서 생성된 조사표, 부호화 지침서, 마이크로데이터, 녹취록 및 녹음 파일 등 모든 형태의 자료를 조사자료로 설정하고 있다.

정리하면, 조사데이터는 연구수행과정에서 생성된 데이터의 하나로, 각종 설문 및 인터뷰 등 다양한 형태의 조사방법을 적용함으로써 생성된 자료를 의미하여, 조사표·설문지·부호화지침서·녹취록·녹음파일 등 양적·질적 데이터를 포함한 조사의 전 과정에서 생성된 다양한 형태의 자료 일체를 뜻한다.

〈표 2-3〉 연구데이터의 종류

필요성	주요 내용
실험 데이터	실험장비에서 생산되는 데이터 - 가속기, 화학/바이오 실험데이터
관측 데이터	관측장비를 통해 생산되는 데이터 - 망원경, 전자현미경, 인공위성 등
시뮬레이션 데이터	모델링을 통해 생산되는 데이터 - 기후모델링, 경제전망 모델링
파생 데이터	원천 데이터로부터 재생산된 데이터 - 텍스트마이닝, 3D 모델링
참조 데이터	평가를 거쳐 신뢰성이 공인된 데이터 - 플라즈마 물성표준, 뇌MRI 영상 참조 표준 데이터
조사 데이터	설문조사 등을 통해 생산된 데이터 - 시장조사, 예측조사

자료: 최명석, 김현우, 이상환(2019). 국가 연구데이터 공유의 시작, 데이터 관리계획(Data Management Plan), KISTI ISSUE BRIFE, 13.

정영철 외(2020). 전사적 연구데이터 관리 체계 마련을 위한 연구. 한국보건사회연구원. 재인용

1. 공공데이터의 관리

가. 공공데이터법

공공데이터의 개발에 대한 요구가 점차 커지는 가운데 2013년 「공공데이터의 제공 및 이용 활성화에 관한 법률」이 제정되었다. 이 법률은 공공기관이 보유(이하 공공데이터법)관리하고 있는 데이터의 제공과 활용을 활성화하는 내용을 담고 있으며, 국민의 데이터 이용 권리를 보장하고 그 활용가치를 확대하는데 목적을 가진다(제1조). 여기서 말하는 공공데이터란 “데이터베이스, 전자화된 파일 등 공공기관이 법령 등에서 정하는 목적을 위하여 생성 또는 취득하여 관리하고 있는 광(光) 또는 전자적 방식으로 처리된 자료 또는 정보(제2조 제2호)”을 의미한다. 또한 이를 제공하고 관리하는 주체로서의 공공기관이란 국가기관, 지방자치단체 등을 포함하게 된다.

공공기관은 공공데이터의 제공(이용자가 활용 가능한 형태의 데이터를 접근할 수 있도록 하는 방식(제2조 제4항))에 대한 의무를 가지며, 이때 데이터의 접근과 이용에 대한 평등권을 보장하여야 한다(제3조 제1항 및 제2항). 또한, 공공기관은 타인의 권리를 침해하거나 불법행위에 활용하는 등의 사유가 아니라면 공공데이터의 제공을 금지하거나 제한할 수 없다(제3조 제4항).

공공데이터의 제공 및 활성화는 관련 기본계획 수립을 통하여 공공데이터 제공의 방향, 제공 및 활용 촉진 방안, 품질관리, 관련 법적·제도적 개선사항 등이 구체화된다. 또한 공공기관의 공공데이터 제공 현황에 대한 조사가 이루어질 수 있으며, 행정안전부는 이에 대한 개선 권고, 시정 요구 등의 조치를 취할 수 있다(제 9조 제2항). 이상의 내용은 공공데이터

관리지침을 통하여 보다 구체화된다(제11조).

공공데이터법의 확대로 각 공공기관에서는 공공데이터 제공에 대한 규정 및 지침을 신설하고 공공데이터 제공 및 이용 활성화 업무를 총괄하는 공공데이터제공책임관을 임명하고 있다. 공공데이터제공책임관은 공공데이터의 시책을 총괄조정하고, 기관 내 다른 정책 및 계획 등과 공공데이터 시책을 연계·조정하며, 공공데이터 관리·제공·이용 관련 업무를 지원한다. 또한, 공공데이터의 품질관리를 통하여 국민에게 제공되는 데이터의 질적 제고를 위한 노력에 실질적 업무를 담당하게 된다.

공공데이터법 제3장, 제4장에서는 공공데이터 제공의 기반 조성 및 제공 절차에 대하여 규정하고 있다. 제17조에 따르면, 제공대상 공공데이터는 공공기관이 보유·관리하고 있는 모든 형태의 공공데이터를 말하지만, 「공공기관의 정보공개에 관한 법률」 제9조에 해당되는 비공개 대상정보, 제3자의 권리가 포함되어 정당한 이용 허가를 득하지 않은 정보의 경우에는 제공 범위에서 제외할 수 있다. 그럼에도 불구하고 이상의 문제가 기술적으로 해결 보완가능한 경우에는 부분적인 데이터 제공을 해야한다(제17조 제2항).

나. 공공데이터관리지침⁵⁾

공공데이터관리지침에서는 공공데이터법 제11조에 따라 공공데이터 제공의 주체인 공공기관의 준수사항과 원칙 및 기준 등을 설정하고 있다.

5) 공공데이터활용지원센터(2021a). 공공데이터관리지침.

〈표 2-4〉 공공데이터 관리지침 조문 체계 구성

장/절	조문	
제1장 총칙	제1조(목적), 제2조(정의), 제3조(기본원칙), 제4조(관리주체), 제5조(관리체계)	
제2장 공공데이터의 제공 관리 단계별 기준	제1절 공공데이터의 생성·수집	제6조(공공데이터 생성·수집)
	제2절 공공데이터 제공을 위한 처리	제7조(제공대상 여부의 확인) 제8조(저작물이 포함된 공공데이터의 관리) 제9조(공공데이터 제공방식)
	제3절 공공데이터의 등록	제10조(공공데이터 목록의 등록) 제11조(공공데이터의 등록) 제12조(공공데이터의 현행화)
	제4절 공공데이터의 제공	제13조(공공데이터의 제공 매체) 제14조(개별 포털 등의 운영) 제15조(제공신청에 의한 공공데이터의 제공) 제16조(제공비용의 부담)
	제5절 공공데이터 제공 사후관리	제17조(공공데이터 목록의 제외) 제18조(공공데이터의 제공중단 및 내용변경) 제19조(이용불편사항 신고 및 처리 등) 제20조(공공데이터제공분쟁조정)
제3장 공공데이터 품질관리	제21조(품질관리체계 구축) 제22조(품질관리 계획 수립 및 이행) 제23조(품질진단 및 개선) 제24조(품질관리 평가 및 조치)	
제4장 공공데이터 제공 시 표준 준수	제25조(공공데이터의 표준 관리) 제26조(제공표준) 제27조(제공표준 데이터셋의 관리)	
제5장 민간 중복·유사 서비스 금지	제28조(중복·유사 서비스 개발·제공 금지) 제29조(중복·유사 서비스 정비 기준) 제30조(중복·유사 서비스 정비 방식) 제31조(공공기관의 책무)	

자료: 공공데이터활용지원센터(2021a). 공공데이터 관리지침

공공기관은 공공데이터의 보편적 이용권을 확대하기 위한 조치를 취해야 한다. 더 나아가 공공데이터를 원천데이터로 제공하기 위한 노력을 경주해야 하며, 통계처리한 데이터 등을 제공하는 경우에도 이용자의 활용성을 고려하여 최대한 상세한 설명을 제공할 의무를 가진다(제3조 제2항).

공공기관은 공공데이터 생성 및 수집, 제공을 위한 기반을 마련하여야 하며, 공공데이터의 관리를 위한 체계를 마련하고 공공데이터 제공의 효율성을 제고하기 위한 방안을 모색해야 한다. 이때 저작물이 포함된 공공데이터의 경우에는 제3자의 권리가 침해되지 않도록 권리소재 및 권리자의 이용허락 범위를 명확히 파악해야 한다(제8조). 한편, 공공데이터의 제공은 데이터의 생성주기와 제공환경, 이용자의 활용 편의성 등이 고려되어 적절한 제공방식이 결정될 필요가 있다(제9조).

공공기관은 공공데이터의 등록 및 이용 창구의 일원화를 위하여 공공데이터 포털에 관련 정보를 등록·제공해야 하지만 별도의 데이터 포털이나 데이터 플랫폼을 구축 운영하는 경우에는 개별포털 구축·운영 허용 기준에 맞게 데이터 제공기반을 마련해야 한다.

〈표 2-5〉 개별 포털 등 구축·운영 허용 기준

장/절		조문	
필수 요건	비중복성	공공데이터 포털과의 서비스 기능의 중복이 없을 것	
	연계성	공공데이터 포털과의 시스템 연계 및 데이터 동기화 수행 방안 제공	
선택 요건	법적 근거	개별법에 자체 포털 구축이 명시된 경우	
	특화성	제공 방식	자체 개별포털에서 데이터 가공·처리·분석이 가능한 정보를 제공하는 경우
		제공 형태	제공형태가 특화되어 제공형태와 데이터 간 분리가 어려운 경우
보안성		제한된 접근환경을 통하여 개인 및 민감정보를 활용해야 하며 일부 인가된 접근 권한을 가진 이용자에 한하여 원격 접근을 허용하는 경우	

자료: 공공데이터활용지원센터(2021a). 공공데이터 관리지침

주: 해당 자료 내 내용을 바탕으로 요약 구성

공공기관은 공공데이터의 품질관리 체계를 구축하고 데이터의 품질관리와 관련된 사항(품질관리 목표, 대상, 기준, 과제, 일정, 예산 등)을 계획하여야 한다(제22조). 또한 계획 내 품질관리 기준에 따라 공공데이터를 상시 점검하고 공공데이터 갱신 시기에 맞게 품질진단을 실시하여야 한다.

다. 공공데이터 제공·관리 매뉴얼⁶⁾

공공데이터 제공·관리 매뉴얼은 공공데이터법 제11조 및 공공데이터

6) 공공데이터활용지원센터(2021b). 공공데이터 제공·관리 실무 매뉴얼.

관리 지침에 근거하여 공공기관의 데이터 제공 및 관리 업무의 과정을 구체화하기 위한 목적으로 작성되었다.

공공데이터는 공공데이터 관리지침에 따라 기계 판독이 가능한 형태(소프트웨어로 데이터의 개별 내용 및 구조를 확인하거나 수정·변환·추출 등 가공이 가능한 상태)로 제공되어야 한다.

[그림 2-1] 기계 판독이 가능한 형태의 충족 조건

▶ 기계 판독이 가능한 형태의 포맷 단계별 구분·비교					
구분	1단계	2단계	3단계	4단계	5단계
기계판독이 가능한 형태	미충족포맷 (공공데이터 포털 등록불가)	최소충족포맷	오픈포맷*		
특징	특정 소프트웨어에서 읽을 수만 있는 데이터로 자유로운 수정, 변환 불가	특정 소프트웨어에서 읽고 수정, 변환 가능	최소한 한 가지 이상의 비독점적 소프트웨어에서 읽고 수정, 변환 가능	URI** 기반으로 데이터 속성 특성 관계를 기술하고 있는 데이터 구조	웹상의 다른 데이터와 연결, 공유 가능
예시	PDF	HWP, XLS	CSV, JSON, XML	RDF	LOD***

* 오픈포맷(open format) : 비용 또는 그 밖의 사용에 제약 없이 최소 한 가지 이상의 무료/자유/오픈소스 소프트웨어로 처리(수정, 편집 등)할 수 있는 기계판독 가능한 데이터포맷(관리지침 제2조제4호)
 - 오픈포맷의 형식은 이미지, 텍스트, 테이블 또는 지리데이터 등 데이터 유형에 따라 다를 수 있으며, 비정형 데이터는 국제표준규격을 준수하여야 함
 ※ (참고) <https://opendefinition.org/ofd/>

** URI(Uniform Resource Identifier, 통합자원식별자) : 웹 상의 특정 콘텐츠를 다른 콘텐츠(텍스트, 이미지, 동영상 등)들과 구별하여 인식·확인할 수 있는 고유값(유일식별자)

*** LOD는 파일 포맷이 아니라, RDF와 같은 형식의 데이터를 활용하여 시멘틱 웹을 구현한 웹서비스 방식의 하나임(상세내용은 참고의 설명 참조)

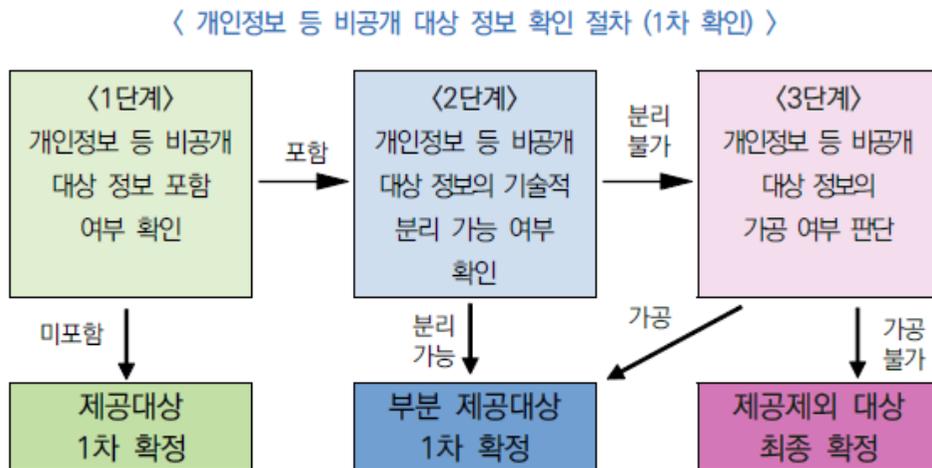
자료: 공공데이터활용지원센터(2021b). 공공데이터 제공·관리 실무 매뉴얼

공공데이터의 제공·관리 단계는 크게 1) 공공데이터 생성·수집 단계, 2) 공공데이터 제공을 위한 처리 단계, 3) 공공데이터 등록 단계, 4)공공데이터 제공 단계, 5) 공공데이터 제공 이후 사후관리 단계로 구분된다. 먼저, 생성·수집단계에서는 생성·수집·취득한 데이터의 제공대상 여부를

정의하고, 기계판독이 가능한 형태로 정비를 해야 한다. 다음으로 데이터 내에 제공제외 대상 정보가 포함되는지, 저작권 문제 등이 발생할 수 있는지 등을 검토하고 이러한 사항에 대한 조치가 이루어진다. 공공데이터를 등록하는 단계에서는 제공대상 공공데이터를 관리 시스템에 등록, 지속적인 현행화 관리를 수행한다. 공공데이터 제공단계에서는 제공대상으로 공표된 공공데이터를 데이터 수요자에게 제공하며, 이후 데이터 수요에 대한 의견 반영, 제공 관련 분쟁 조정 등이 사후에 이루어진다.

공공데이터 관리지침(공공데이터활용지원센터, 2021a)에 따르면, 제공대상이 아닌 정보의 분리제공 기반이 구축될 필요가 있다. 관련 데이터를 데이터베이스화하는 경우에 개인정보, 민감정보, 저작권 포함 정보 등을 제외하여 제공대상에 포함될 수 있도록 해야 한다.

[그림 2-2] 개인정보 등 비공개 대상 정보 확인 절차



자료: 공공데이터활용지원센터(2021b). 공공데이터 제공·관리 실무 매뉴얼

개인정보 등 「공공기관의 정보제공에 관한 법률」 제9조에 해당되는 비공개 대상 정보가 포함(아래에서 구체적으로 설명)되지 않는 경우는 1차적인 제공대상 데이터로 선정된다. 만약 데이터 내 비공개 대상 정보가

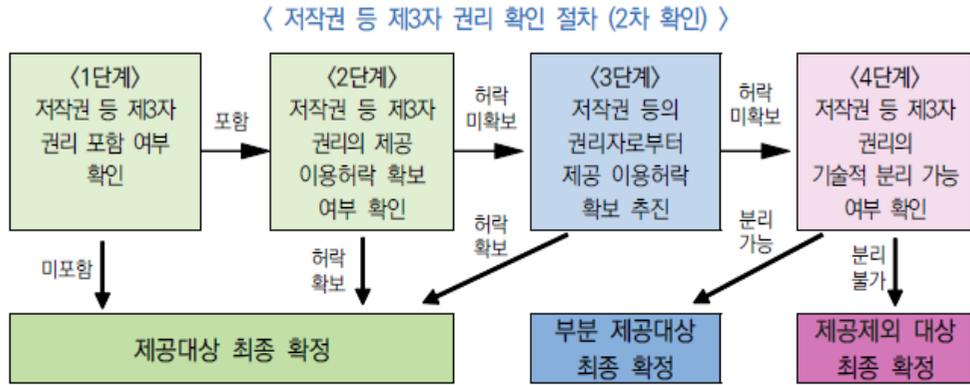
포함된 경우, 기술적 분리⁷⁾가 가능할 때 해당 정보를 분리하고 이를 별도로 기록 관리하여 부분 제공대상으로 확정한다. 만약 기술적 분리가 어려운 비공개 정보를 포함하는 경우에는 비공개정보에 대한 가공⁸⁾을 수행하여 부분 제공 대상으로 선정할 수 있다. 대표적으로 개인정보가 포함된 공공데이터의 경우에는 익명처리를 통하여 제공대상으로 선정될 수 있다. 그럼에도 불구하고 비공개 대상 정보의 가공이 불가능한 경우에는 최종적으로 제공제외 대상으로 선정할 수 있다.

개인정보 등 비공개 대상 정보를 확인한 후에는 해당 데이터가 저작권 등 제3자의 권리를 포함하고 있는지를 검토해야 한다. 만약 제공대상 데이터가 저작권 등 제3자의 권리를 포함하고 있지 않은 경우는 제공대상 범위에 포함될 수 있다. 또한 데이터 권리자의 제공 이용허락을 득한 경우에도 제공대상 범위에 포함될 수 있다. 하지만 권리자의 허락을 확보하지 못한 경우, 해당 공공기관에서는 데이터 권리자로부터의 제공 이용 허락을 득하기 위한 노력이 요구된다. 데이터 권리자로부터의 이용 허가를 받지 못한 데이터는 해당 권리 영역에 대한 기술적 분리가 가능한지 여부를 판단하게 되고, 제공 가능한 데이터와 권리 영역에 해당 데이터 간 분리가 가능한 경우에 한하여 부분적인 제공 대상으로 설정할 수 있다. 그럼에도 불구하고 데이터에 대한 권리가 분리되지 못하는 경우에는 최종적으로 제공제외 대상으로 선정된다.

7) 기술적 분리란 하드웨어 및 소프트웨어와 기술적 지식을 사용하여 분리할 수 있는 정도를 말하며 분리 작업이 기관의 컴퓨터 시스템 운영에 지장을 초래하지 않는 정도의 물리적으로 쉽고 분리 가능한 경우를 말함(제17조 제2항)

8) 생성·수집되어 관리되고 있는 항목 외에 추가적인 정보의 생성이나 변형, 요약, 발췌 등의 처리가 필요한 경우, 이러한 처리를 위한 추가적인 자원이 투입되는 경우를 의미함(제26조 제3항)

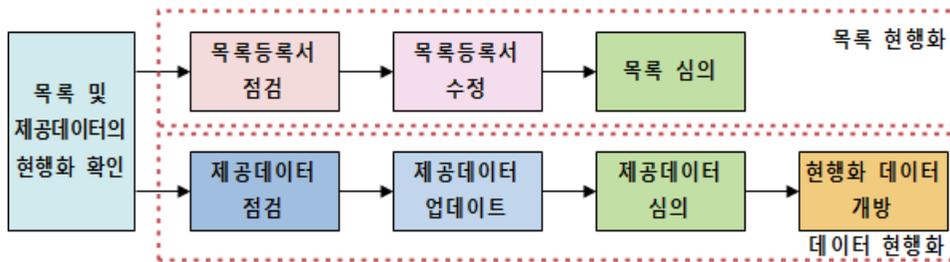
[그림 2-3] 저작권 등 제3자 권리 확인 절차



자료: 공공데이터활용지원센터(2021b). 공공데이터 제공·관리 실무 매뉴얼

공공기관은 이상의 과정을 통하여 공공데이터 제공범위의 확대를 지속 추진해야 하며 이와 함께 공공데이터의 현행화를 통하여 적시성있고 활용 가치가 높은 데이터를 충분히 제공해야 한다. 제공데이터의 경우, 항목이 변경된 시점을 기준으로 10일 이내 변경 등록을 해야 한다(공공데이터의 제공 및 이용 활성화에 관한 법률 시행령 제15조 제2항). 각 공공기관의 실무담당자는 공공데이터의 현행화 수준을 확인하고 그에 맞는 조치를 취할 의무를 가진다.

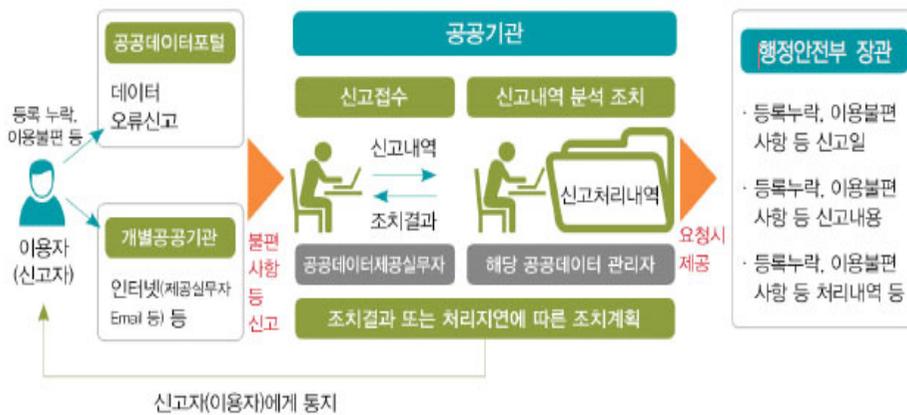
[그림 2-4] 공공데이터 현행화 절차



자료: 공공데이터 제공·관리 실무 매뉴얼(공공데이터활용지원센터, 2021b)
 주: 해당자료 내 내용을 바탕으로 요약 구성

공공기관은 공공데이터 수요에 맞게 적절한 사후조치를 할 의무를 가진다. 다시 말해 데이터 이용자가 신고한 등록 누락, 이용 시 불편사항 등에 대한 시행조치가 이루어져야 한다. 데이터 이용자가 공공데이터 이용과 관련된 불편사항을 신고한 경우, 공공데이터 제공 실무자는 해당 신고내역에 대한 검토 및 조치 이후에 이에 대한 사항을 이용자에게 통지해야 한다. 이와 함께 이용불편사항에 대한 정보(신고일자, 신고내용, 처리내역 등)를 행정안전부에 제출할 수 있다.

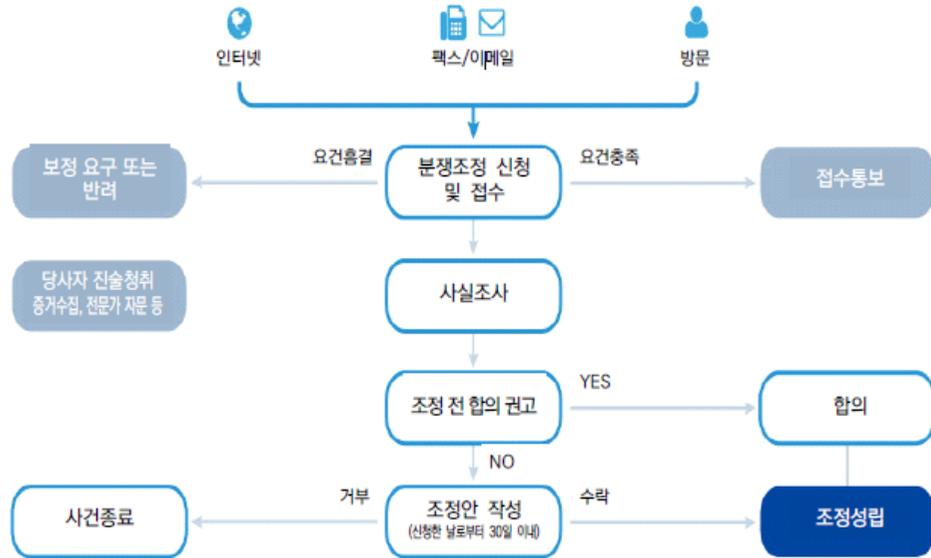
[그림 2-5] 공공데이터 이용불편사항 신고 및 처리



자료: 공공데이터활용지원센터(2021b). 공공데이터 제공·관리 실무 매뉴얼

한편, 공공기관이 공공데이터 이용자의 제공신청을 거부하거나 기 제공 중이던 데이터의 제공중단을 결정한 경우, 데이터 이용자는 제공거부 및 제공 중단 처분에 대한 사항을 공공데이터제공분쟁조정위원회에 분쟁조정신청할 수 있다. 분쟁조정을 통보받은 공공기관은 분쟁조정에 필요한 관련 자료를 제출해야 한다. 분쟁조정대상 데이터에 비공개 대상 정보가 포함된 경우, 제3자 저작권이 포함된 경우에는 기술적 처리를 통하여 개방확대에 노력을 해야 한다. 그 외에 사항에 대해서는 위원회 결정 사항을 따라야 하며, 조정 결정 사항은 데이터 이용자에게 즉시 통지해야 한다.

[그림 2-6] 공공데이터제공 분쟁조정 처리 절차



자료: 공공데이터활용지원센터(2021b). 공공데이터 제공·관리 실무 매뉴얼

2. 공공기관의 정보공개

「공공기관의 정보공개에 관한 법률」(이하 정보공개법)에서는 공공기관이 보유·관리하는 정보에 대한 공개 의무와 관련된 사항을 정의하고 있다. 해당 법률은 국민의 알 권리를 보장하고 국정 운영의 투명성을 확보하기 위한 목적을 가지며(제1조), 공공기관은 법률에 따라 관련 정보공개에 적극적으로 임해야 한다(제3조). 정보공개법 제6조의2(정보공개 담당자의 의무)에 따라 각 공공기관은 정보공개 담당자를 위임하며, 정보공개에 대한 자의적 결정, 고의적 처리 지연, 위법한 공개 거부 및 회피 등을 해선 아니 된다.

〈표 2-6〉 정보공개제도의 필요성

필요성	주요 내용
국민의 알 권리 보장	공공기관에서 보유·관리하고 있는 정보에 대한 접근권을 보장
국민주권주의의 실질적 보장	국정에 대한 정확한 정보 접근성을 보장하여 국민의 주권 보장
국정에 대한 신뢰 제고	정책 추진과정에 대한 투명한 공개를 통하여 정당성 확보 및 책임 행정 구현
공공기관 정보의 균등 배분 필요성 증대	공적 업무 수행을 통하여 생산된 정보를 국민에게 균등하게 제공하여야 하는 의무 이행
공직사회의 부정부패 예방 및 국민의 신뢰 제고	국민의 감시와 비판을 제도화, 투명한 업무수행을 통한 국민의 신뢰를 제고

주: 해당 자료 내 내용을 바탕으로 요약 구성
 자료: 행정안전부(2021) 정보공개 운영 안내서

정보공개 운영 안내서(행정안전부, 2021)에서는 정보공개 대상 기관을 구체적으로 적시하고 있다. 정보공개 대상 기관은 크게 시장형 공기업, 준시장형 공기업, 기금관리형 준정부기관, 위탁집행형 준정부기관, 기타 공공기관으로 구분된다. 이 가운데 기타 공공기관에 경제·인문사회연구회 소속 연구기관 전체가 포함되므로 해당 연구기관들은 공공기관의 정보공개 의무를 가지게 된다.

한편, 정보공개법 제9조에서는 비공개 대상 정보의 범위를 설정하고 있다. 이는 공공데이터 공개를 포함한 각종 연구데이터의 공개범위를 설정하는 중요한 근거가 된다. 제9조 1항에 따르면, 다른 법률에서 설정한 비밀이나 비공개 사항(제1호), 공개 시 국가의 이익을 현저하게 해칠 우려가 있는 정보(제2호), 국민의 생명·신체 및 재산의 보호에 현저한 지장을 초래할 우려가 있다고 인정되는 정보(제3호), 진행중인 재판 관련 정보로 공정한 재판을 받을 권리를 침해한다고 인정될 만한 정보(제4호), 감사·감독·인사관리, 의사결정 및 내부검토 과정에 있는 사항으로 공정한 연구·개발에 현저한 지장을 초래하는 경우(제5호), 「개인정보 보호법」 제2조 제1호에 따른 개인정보로서 사생활의 비밀 또는 자유를 침해할 우려가 있다고 인정되는 정보(제6호), 법인 및 단체의 경영상 비밀의 관한 사항으로 법인 등의 이익을 해칠 우려가 있는 정보(제7호), 특정인에게 이익 또는 불이익을 중 우려가 있다고 인정되는 정보(제8호)가 비공개 대상이 된다. 그러나 제1항 각호의 범위에서 비공개 대상 범위를 설정할 때에는 비공개 세부 기준을 수립하여 정보공개시스템에 이를 공개하여야 한다(제9조 조3항).

비공개 범위 가운데 연구데이터의 생성 및 제공과 상당한 관련성을 가지는 사항은 제3호, 제6호, 제7호라 볼 수 있다. 연구과정 중에 국민의 생명 신체 및 재산의 보호에 현저한 지장을 초래할 우려가 있다고 인정되는

정보가 수집된 경우, 이는 적절한 비공개 대상 범위를 설정하여 공개를 제한⁹⁾할 필요가 있다. 또한 제6호에 따라 개인을 특정하거나 사생활의 비밀 또는 자유를 침해할 우려가 있다고 인정되는 정보¹⁰⁾가 수집된 경우에도 비공개 대상 범위에 포함된다. 다만, 법령에 의하여 열람 가능한 정보, 공표를 목적으로 수집되어 사생활 및 자유를 부당하게 침해하지 않는 정보, 권리구제를 위한 정보, 공적 직무를 수행하는 자의 정보, 공익을 위하여 공개될 수 있는 정보의 한해서는 비공개 대상에서 제외될 수 있다. 한편, 데이터 수집 대상이 법인·단체 등 이익을 추구하는 집단의 경우 이들의 정당한 이익에 영향을 미칠 수 있는 정보는 비공개 대상에 포함될 수 있다. 다만, 정보의 공개를 통하여 국민의 재산 및 생활 보호에 기여할 수 있는 정보의 경우에는 공개가 필요하다고 판단될 수 있다.

정보공개법 시행령 제4조에 따르면, 공공기관은 국민생활에 매우 큰 영향을 미치는 주요 정책정보에 대하여 구체적 공개 범위, 주기, 시기, 방법 등을 미리 정하여 정기적으로 공개해야 함을 명시하고 있다. 여기서 말하는 국민생활에 영향을 미치는 정책정보로는 1) 정책계획 자료, 조사·통계조사 결과 자료, 법령 정보 등 행정계획 및 그 실시상황에 관한 정보, 2) 보건·위생 관련 건강 정보, 환경 및 치안 관련 정보 등 국민의 생명·신체·재산 보호 관련 정보, 3) 사회보장·의료·교통·조세·주거·교육문화 등 국민의 일상생활과 관련된 정보 등이 있다. 이 외에도 정보공개청구 빈도가 높은 정보, 사회적 이슈가 된 각종 현한 사업의 실태조사 결과 등은 공

9) 정보공개 운영 안내서(행정안전부, 2021)에서 제시하고 있는 정보공개/비공개 세부기준 및 유형별 해설(제4장)에 따르면, 국가 연구기관에 허 해한 연구성과는 공개가 원칙이지만, 최종적인 공표까지 전문적 검토가 이루어지지 않은 경우나 공개로 인하여 사회전체에 미치는 영향이 큰 경우에 한하여 공개를 제한할 수 있다.

10) 비공개 대상으로 설정될 수 있는 개인의 민감정보는 다음과 같다(행정안전부, 2021). 성명, 주민등록번호, 통장계좌번호, 급여, 학력이나 직업 등 경력, 활동사항, 사상·양심·종교에 관한 정보, 병원 진료기록 등 개인의 건강상태 및 신체 정보, 재산상황(납세증명서), 개인 평가기록, 가족관계, 다른 사람과의 인적 유대 관계, 신분, 목소리, 초상 등

공기관의 장이 정하여 사전적 공개 범위에 포함시킬 수 있다.

〈표 2-7〉 정보의 사전적 공개: 주요 정책 정보

정보 유형	세부 내용
행정 계획 및 그 실행에 관한 정보	(기본계획) 중장기 기본계획, 행정계획 등 주요 정책자료, (통계·조사) 행정계획의 근거·참고가 되는 통계자료, 조사연구 결과 등, (법령정보) 법령(조례, 지침 포함), 입법예고안, 국회제출 법률안, (기타) 국가 주요 정책 등에 관한 사항
식품·위생, 환경, 복지, 개발사업 등 국민의 생명·신체·재산 보호 관련 정보	(식품) 농산물 안정성 조사, 위해식품 및 유독물 정보 등 각종 검사결과 등, (보건·위생) 식품위생업소 단속 행정처분 등 국민의 건강 관련 정보 등, (환경) 환경관련 검사 및 측정결과, 환경오염물질 배출시설 및 단속 현황 등, (치안) 방범·방재시설 설치내역, 청소년 범죄현황 등 치안상황 관련 정보
교육·의료·교통·조세·건축·상하수도·전기·통신 등 국민의 일상생활과 관련된 정보	(의료·사회보장) 노인·아동·장애인 복지시설 현황, 각종 생활보장급여 수급기준 등, (교통) 교통영향분석 평가 결과, 자전거 도로 현황 등, (조세) 국·지방세제 관련 과세시가 표준액 산정내용, 공시지가 등, (주택·건설) 주택보급률 및 미분양 현황, 건축허가·착공 통계 등, (가스, 상하수도) 단수 공고, 도시가스 공급정보, 하수처리 현황 등, (전기·통신) 정전 공고, 전기 공급 및 이용현황, 통신사업자 현황 등, (교육문화) 학교시설 및 개방현황, 문화재 및 문화공간 현황 등

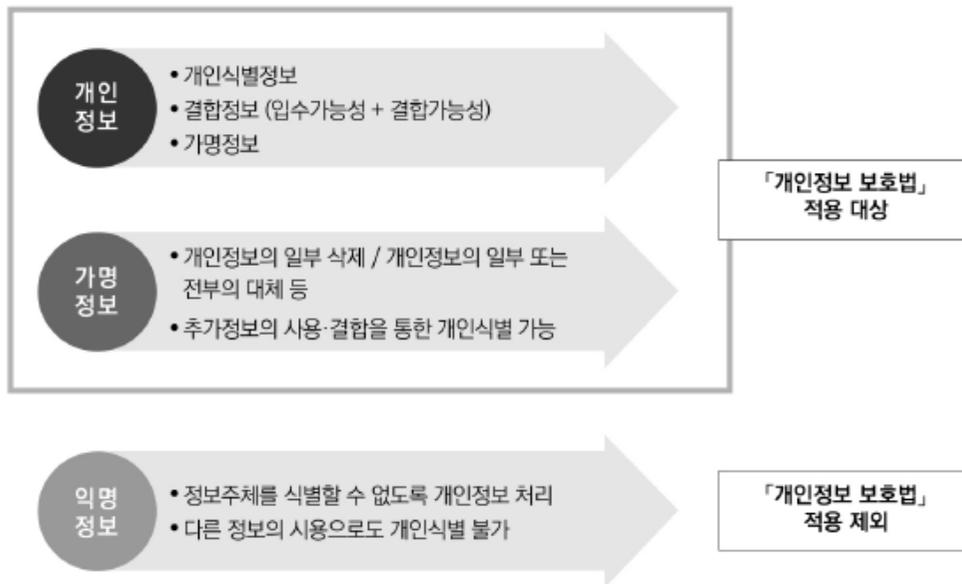
자료: 행정안전부(2021). 정보공개 운영 안내서

주: 해당자료 내 내용을 바탕으로 저자가 작성

3. 개인정보 보호법

「개인정보 보호법」은 개인정보의 처리 및 보호에 관한 사항을 정함으로써 개인의 자유와 권리를 보호하고, 나아가 개인의 존엄과 가치를 구현함을 목적으로 한다(제1조). 여기서 말하는 개인정보는 성명, 주민등록번호 등 개인을 알아볼 수 있는 정보 또는 다른 정보와의 결합을 통하여 쉽게 개인을 특정할 수 있는 정보(제2조 제1호)를 의미한다. 또한, 익명정보는 정보 주체를 식별할 수 없도록 처리된 정보를 의미하며, 다른 정보와의 결합을 통하여서도 개인의 식별이 불가능한 경우를 의미한다. 이 외에도 개인 정보의 일부 삭제 또는 대체(가명처리)되어 추가정보의 결합을 통하여 개인식별이 가능한 정보를 가명정보라 한다. 「개인정보 보호법」에 따르면 개인정보 및 가명정보는 해당 법률에 따라 보호 대상이 되며, 익명정보는 보호 대상에서 제외된다.

[그림 2-7] 개인정보·가명정보·익명정보의 비교



자료: 개인정보보호위원회(2021). 개인정보 보호법령 해석 실무교재.

「개인정보 보호법」에 따라 개인정보의 보호는 다음의 원칙을 따른다(제3조). 1) 개인정보의 처리 목적에 따라 최소한의 개인정보를 수집, 2) 개인정보 처리 목적에 적합하게 활용, 3) 개인정보의 정확성, 완전성, 최신성 보장, 4) 정보주체의 권리 침해 및 위험 가능성을 고려한 안전 관리 대책 수립, 5) 개인정보 처리에 관한 사항을 공개, 열람 청구권 등 정보주체의 권리를 보장, 6) 정보주체의 사생활 침해를 최소화하는 방법으로 개인정보 처리, 7) 익명 또는 가명 처리를 통하여 정보수집의 목적 달성이 이루어지는 경우 익명·가명처리 수행, 8) 개인정보 보호에 대한 책임과 의무를 다하고 정보주체의 신뢰를 얻기 위하여 노력.

정보주체는 개인정보 처리와 관련하여 다음의 권리를 가진다(제4조). 1) 개인정보 처리에 관한 정보를 제공받을 권리, 2) 개인정보 처리에 대한 동의, 그 범위의 결정, 3) 개인정보 처리 여부 확인 및 개인정보에 대한 열람 요구, 4) 개인정보 처리 정지·정정·삭제 및 파기 요구 권리, 5) 개인정보 처리로 인하여 발생한 피해를 신속하고 공정한 절차에 따라 구제받을 권리, 6) 완전히 자동화된 개인정보 처리에 따른 결정을 거부하거나 그에 대한 설명을 요구할 권리.

이상의 내용에 근거하여, 조사데이터 수집 시 「개인정보 보호법」 제20조에 따라 조사참여자에 대한 개인정보 수집의 출처, 처리의 목적, 개인정보 처리 정지를 요구·동의를 철회할 수 있는 권리 사실 등을 의무적으로 알려야 한다. 특히 민감정보(사상·신념, 노동조합·정당의 가입·탈퇴, 정치적 견해, 건강, 성생활 등에 관한 정보, 그 밖에 정보주체의 사생활을 현저히 침해할 우려가 있는 개인정보로서 대통령령으로 정하는 정보)에 대한 처리 시 정보주체에 대한 별도의 동의가 필요하며(제23조 제1호), 민감정보의 공개 가능성 및 비공개 선택 방법 등을 정보주체가 인지하기 쉽도록 알려야 한다(제23조 제3항).

4. 생명윤리 및 안전에 관한 법률

「생명윤리 및 안전에 관한 법률(이하 생명윤리법)」은 인간 대상의 조사를 수행하는 연구분야에 광범위하게 영향을 미치는 법률이다. 생명윤리법은 인간과 인체유래물 등을 연구할 때 인간의 존엄과 가치를 침해하거나 인체에 위해를 끼치는 행위를 방지함으로써 생명윤리 및 국민의 건강과 안전을 확보하기 위함을 목적으로 한다. 생명윤리법 시행령 제2조 제1호에서는 다음과 같은 인간대상 연구의 범위를 설정하고 있다. 1) 사람을 대상으로 물리적으로 개입하는 연구: 연구대상자를 직접 조작하거나 연구대상자의 환경을 조작하여 자료를 얻는 연구, 2) 의사소통, 대인 접촉 등의 상호작용을 통하여 수행하는 연구: 연구대상자의 행동관찰, 대면 설문조사 등으로 자료를 얻는 연구, 3) 개인을 식별할 수 있는 정보를 이용하는 연구: 연구대상자를 직접·간접적으로 식별할 수 있는 정보를 이용하는 연구. 이를 고려하였을 때, 한국보건사회연구원에서 수집하는 조사데이터의 다수가 인간대상 연구의 범위에 포함됨을 알 수 있다.

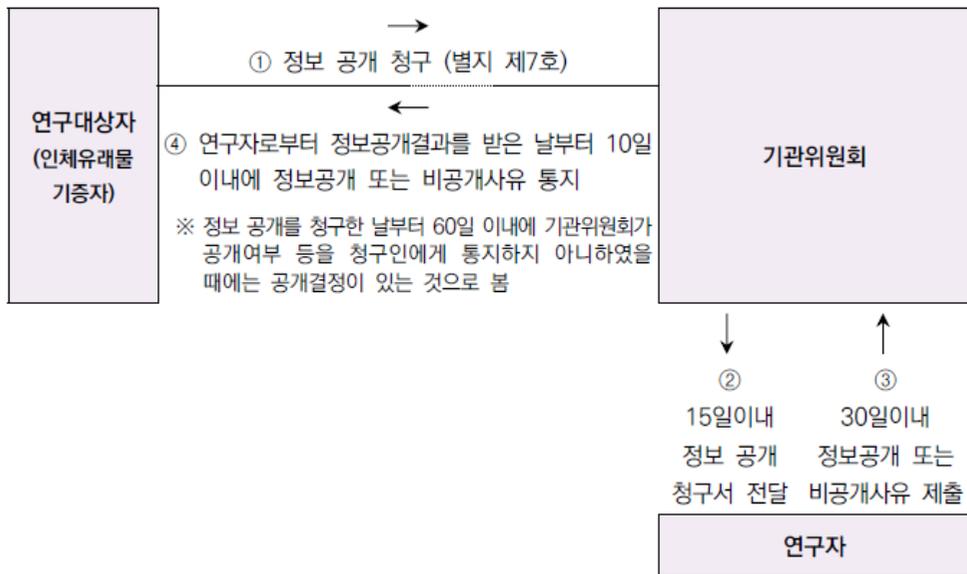
생명윤리법 제10조에서는 기관생명윤리위원회의 설치와 기능에 대하여 규정하고 있다. 기관생명윤리위원회 등록 시 연구의 특성¹¹⁾에 따라 구분될 수 있는데, 한국보건사회연구원은 2013년 7월에 기관생명윤리위원회(이하 기관위원회)에 등록되어 2023년 기준 ‘인간대상연구’를 수행하는 기관으로 등록되어 있다(기관생명윤리위원회 정보포털(2023). 기관생명윤리위원회 등록 현황). 이와 더불어 생명윤리위원회를 개별적으로 운영¹²⁾하는 기관으로 등록된 상태이다.

11) 연구 구분: 인간대상연구, 인체유래물연구, 배아생성의료기관, 배아연구기관, 인체유래물은행, 배아줄기세포주 이용연구

12) 생명윤리위원회의 운영 방식은 개별운영(해당 기관에서 독립적으로 운영하는 형식), 협약운영(공용기관생명윤리위원회 협약을 통하여 위탁 운영), 통합운영(복수의 기관위원회가 설치된 경우에 통합하여 운영)으로 구분된다.

법률 제15조에 따르면, 인간대상 연구를 수행하기 전에 기관위원회에 심의를 받아야 하며, 제16조 제1항에 근거하여 이하의 내용에 대한 서면 동의를 받아야 한다. 1) 인간대상연구의 목적, 2) 연구대상자의 참여 기간, 절차 및 방법, 3) 연구대상자에게 예상되는 위험 및 이익, 4) 개인정보 보호에 관한 사항, 5) 연구 참여에 따른 손실에 대한 보상, 6) 개인정보 제공에 관한 사항, 7) 동의의 철회에 관한 사항. 만약 이상의 내용에 대한 서면 동의를 받은 경우에는 연구대상자의 안전 대책이 보장된 상황에서 익명화된 정보를 제3자에게 제공할 수 있다(제17조, 제18조). 이와 함께 인간대상 연구와 관련된 사항은 기록·보관되어야 하며, 연구대상자가 자신의 정보에 대하여 공개 청구할 수 있으며, 특별한 사유가 없다면 연구자는 그 정보를 공개하여야 한다.

[그림 2-8] 생명윤리법에 따른 연구대상자의 정보공개 청구 처리 절차 흐름



자료: 보건복지부·국가생명윤리정책원(2022). 기관생명윤리위원회 관리 안내.

기관위원회 운영 시 조직 구성과 운영 실적 등을 정기적으로 평가 인증할 수 있으며(법령 제14조), 이때 기관위원회 심의의 적절성, 조사·감독 수행의 적정성, 관련 교육의 적정성, 취약 대상에 대한 보호대책 수립 및 이행, 윤리지침 마련, 위원회 구성의 독립성과 전문성, 표준운영지침, 기록 및 문서관리의 체계성 등을 평가하게 된다(시행령 제10조).

[그림 2-9] 기관위원회 평가 영역



자료: 보건복지부·국가생명윤리정책원(2022). 기관생명윤리위원회 관리 안내.
 주: 생명윤리법 시행령 제10조 제1항의 인증기준은 2개 영역, 5개 범주, 40개 세부평가 기준으로 구분되며, 자세한 사항은 부록 참고

기관위원회는 크게 심의, 조사·감독, 교육·운영의 기능을 수행하게 된다. 기관위원회는 연구계획서의 윤리적·과학적 타당성에 대한 심의를 수행하며, 연구대상자를 통한 데이터 수집의 적법성, 연구대상자의 생명윤리 보호, 개인정보 보호 대책 등을 심의한다. 또한, 심의 대상에 해당되는 연구의 진행과정을 조사하고 감독하는 기능을 수행하며, 그 외에도 연구자 및 담당자에 대한 교육과 취약 대상에 대한 보호대책 마련, 기관 내 연

구 윤리지침 마련 등의 역할을 수행한다.

[그림 2-10] 기관위원회의 기능

1. 연구계획서의 윤리적·과학적 타당성
2. 연구대상자등으로부터 적절한 절차에 따라 동의를 받았는지 여부 심의
3. 연구대상자등의 안전에 관한 사항, 개인정보 보호 대책
4. 그 밖에 위원회에서의 생명윤리 및 안전에 관한 사항



자료: 대한기관윤리심의기구협의회(n.d.). IRB행정간사 심화교육 발표자료

보건복지부 지정 공공기관생명윤리위원회(2023)에서 권고하고 있는 인간대상연구 조사연구용 연구계획서의 주요 구성 내용을 아래 표와 같다. 인간대상의 조사연구를 수행할 경우에는 연구의 배경 및 목적 외에도 연구대상자의 모집의 구체성과 최소한의 개인정보 수집, 연구대상자의 동의(철회 및 철회 포함) 및 연구방법의 구체성, 연구대상자의 위험과 이익, 연구대상자에 대한 안전과 개인정보 보호 대책 수립 등에 대하여 기관생명윤리위원회의 심의를 거쳐야 한다.

〈표 2-8〉 공공기관위원회 인간대상연구 조사연구용 연구계획서 권고서식 구성

구성	설명
연구배경	<ul style="list-style-type: none"> • 연구 배경과 필요성에 대해 선행 연구, 이론적 근거를 기반으로 충분히 기술 필요 ① 연구 분야의 최신 동향, 선행 연구, 인구통계학적 및 역학 정보 등 연구 배경과 연구를 수행해야 하는 필요성에 대한 분명한 설명 ② 연구의 이론적 근거와 이에 근거한 연구 방법에 대한 설명 ③ (필요시) 연구에서 제기된 윤리적 문제나 고려사항에 대한 연구자의 관점, 그리고 그 문제나 고려사항을 어떻게 다룰지에 대한 제안
연구목적	<ul style="list-style-type: none"> • 연구목적은 명확하고 분명하게 기술
연구수행 주체 정보	<ul style="list-style-type: none"> • 연구책임자, 공동연구자, 담당자의 성명과 직명 • 연구실시 기관명 및 주소 • 연구지원 기관
연구기간	<ul style="list-style-type: none"> • 연구 소요 예상 기간(승인일로부터)
연구대상자	<ul style="list-style-type: none"> • 연구대상자의 선정 또는 제외기준에 대한 범위 및 나이, 성별, 사회적 또는 경제적 요인의 기초 하에 모든 제외기준에 대한 정당성 또는 기타 이유에 대한 정당성 기술 • 선정 및 제외기준은 연구대상자 설명문 및 동의서, 모집 문건에도 동일하게 기술 • 연구대상자에 취약한 대상자가 포함되는 경우 아래 사항에 대해 기술 <ul style="list-style-type: none"> 1) 취약한 대상자가 반드시 참여해야 하는 타당성 및 정당성 기술 2) 취약한 대상자의 자발적 참여를 위한 보호 방안 기술 3) 취약한 연구대상자에 대한 위협과 불편함을 최소화하는 방안 기술
예상 연구대상자 수와 산출 근거	<ul style="list-style-type: none"> ① 연구대상자를 직접 모집하는 경우, 중도 탈락률을 고려한 예상 연구대상자 수와 산출 근거 명시 필수 ② 연구에 필요한 연구대상자 수를 통계학적 산출 및 선행 연구에 근거하여 제시 ③ 예상 연구대상자 수는 계획된 연구에서 필요한 결과를 얻을 수 있는 최소한의 수
연구대상자 모집	<ul style="list-style-type: none"> • 연구대상자 모집 절차 및 방법(광고, 인터넷, 선별접촉 등), 모집 장소, 모집 주체를 구체적으로 기술 ① 타 기관과 협조하여 연구대상자를 모집하는 경우, 타기관에 협조 요청 방법 및 절차 등을 구체적으로 기술 ② 연구대상자를 모집하는 동안 개인의 사생활 보호와 비밀 유지를 위한 방법 및 절차 등을 기술

〈표 2-8〉 공공기관위원회 인간대상연구 조사연구용 연구계획서 권고서식 구성(계속)

구성	설명
연구대상자 동의	<ul style="list-style-type: none"> • 연구대상자에게 동의를 획득하는 경우 동의 획득 방법 및 절차, 동의 획득 장소, 동의 획득 시기, 동의 획득 담당자를 구체적으로 기술 <ul style="list-style-type: none"> ① 연구에 대한 설명을 제공하는 시기를 구체적으로 기술 ② 대리인의 동의 취득이 필요한 경우, 법정대리인의 동의 취득 절차에 대해 기술 ③ 비대면 동의를 취득하는 경우, 비대면 동의 취득의 정당성, 동의 취득 방법 및 절차(전자서명, (전자)우편, 택배 등), (전자)문서 관리 및 폐기 방법 등에 대해 기술 ④ 서면동의 면제를 신청하는 경우, 서면동의 면제 사유 반드시 기술
연구방법	<ul style="list-style-type: none"> • 연구 방법(대면 또는 비대면) 및 절차, 수행 횟수 및 장소, 소요시간, 담당자 등을 구체적으로 기술 <ul style="list-style-type: none"> ① 설문조사인 경우 설문지 배부 및 수거 방법에 대해 구체적으로 기술 ② 면담 또는 관찰조사인 경우 녹음(또는 녹화) 여부, 녹음(또는 녹화) 장비, 녹취록 전사 담당자 등을 구체적으로 기술 ③ 비대면 조사인 경우 연구대상자에게 전달할 링크 주소와 링크 전달 방법에 대해 구체적으로 기술 ④ 연구에서 사용하는 설문 또는 면담 척도의 출처에 대해 기술 ⑤ 연구에서 사용하는 설문지나 면담 질문지는 추가 제출 서류에 첨부 필요
관찰 항목	<ul style="list-style-type: none"> • 연구를 통해 얻고자 하는 정보 또는 자료의 내용을 구체적으로 나열 (예시)인구사회학적 정보, 연구수행 척도 문항 등 구성
자료분석과 통계적 방법	<ul style="list-style-type: none"> • 연구를 통해 수집된 자료 또는 정보의 이용하는 방법(통계적 방법 포함)을 구체적으로 기술
동의 철회 및 중도 탈락	<ul style="list-style-type: none"> • 연구 등록 이후 연구대상자의 연구 참여가 제한되는 동의철회 및 중도탈락기준과 해당 연구대상자의 수집된 데이터의 관리 방법을 구체적으로 기술 <ul style="list-style-type: none"> ① 동의철회: 연구대상자가 연구 도중 스스로 참여를 포기하는 경우 ② 중도탈락: 연구대상자의 상태나 기타 이유로 연구 지속이 불가능하다는 연구자의 판단에 의해 연구 참여가 중단되는 경우

〈표 2-8〉 공공기관위원회 인간대상연구 조사연구용 연구계획서 권고서식 구성(계속)

구성	설명
연구대상자의 위험과 이익	<ul style="list-style-type: none"> ● 연구 참여로 인해 연구대상자에게 발생할 수 있는 이익과 위험이나 불편(개인정보 수집 및 활용으로 인한 내용 포함)과 그에 따른 주의사항 및 조치내용 ① 연구에 참여함으로써 연구대상자에게 기대되는 이익 기술 ② 연구대상자에게 직접적으로 발생하는 이익이 없는 경우, 본 연구 참여로 인해 직접적인 이익이 발생하지 않는다는 내용 기술
연구 참여에 따른 보상	<ul style="list-style-type: none"> ● 연구 참여시 연구대상자가 얻게 되는 보상과 보상이 지급되는 기준 ① 중도 탈락 또는 동의 철회 시 참여 정도에 따라 지급되는 보상이 있다면 해당 내용 기술 ② 보상 제공을 위해 수집되는 개인정보 항목과 관리기준(보상 지급 후 즉시 폐기 등) 기술
연구대상자 안전대책 및 개인정보 보호대책	<ul style="list-style-type: none"> ● 연구대상자를 안전하게 보호하기 위한 대책을 마련하고 연구와 관련된 손상이 발생하였을 경우 보상/배상 등 ① (해당시) 연구대상자의 개인정보를 수집하는 경우, 수집하는 개인정보의 항목, 연구자료 보관 장소, 잠금장치, 암호화 여부, 보관과 폐기 방법, 관리자 등에 관한 기술(연구대상자에게서 수집하는 개인정보는 모든 항목을 구체적으로 기술) ② (해당시) 개인식별정보가 수집되는 경우, 해당 항목의 수집 사유와 필요성, 수집 후 분석방법, 보호 방안(익명화 등)에 대해 기술 ③ (해당시) 보상 제공을 위해 수집되는 개인정보가 있는 경우 수집되는 개인정보 항목, 관리기준(보상 지급 후 즉시 폐기 등), 연구목적으로 수집된 정보와 분리하여 보관한다는 내용에 대해 기술 ④ 생명윤리 및 안전에 관한 법률 시행규칙 제15조에 따라 연구 관련 기록은 연구 종료 시점으로부터 3년간 보관해야 한다는 내용 기술 ⑤ 2차적 사용 및 제3자 제공계획 여부에 대해 기술

자료: 보건복지부 지정 공공기관생명윤리위원회(2023). 인간대상연구 조사연구용 권고서식 제1-1-2호
 주: 인간대상연구 조사연구용 연구계획서 내 설명내용을 바탕으로 표 작성

5. 국가연구개발사업의 관리

가. 국가연구개발정보처리기준

「국가연구개발정보처리기준」(과학기술정보통신부고시 제2020-102호)에서는 「국가연구개발혁신법」 제19조 제1항에 따라 연구개발정보의 수집·생산·관리 및 활용 등 연구개발 정보의 처리에 관한 범위, 시기, 방법 등을 규정하고 있다. 해당 고시에서는 연구데이터의 정의와 함께 데이터 관리계획(Data Management Plan)을 정의하고 있다. 먼저, 연구데이터란 위에서 살펴본 바와 같이, 연구개발 과제 수행 중 산출된 데이터를 의미한다. 본 고시의 제23조에 따르면, 연구데이터를 관리할 때 데이터관리계획을 제출해야 함을 명시하고 있다. 여기서 말하는 데이터관리계획이란, 연구데이터의 생산·보존·관리 및 공동 활용 등에 대한 계획을 말한다(제2조). 국가연구개발과제를 선정할 시에 데이터관리계획에 따른 연구데이터의 생산·보존·관리 및 공동 활용가능성을 필수적으로 검토해야 한다.

연구개발기관은 데이터관리계획에 따라 연구데이터를 관리해야 한다. 연구개발기관이란, 「국가연구개발혁신법」 제2조 제3호에 따라 국가연구개발사업을 수행하는 기관을 말한다. 한국보건사회연구원은 연구개발기관 가운데 「정부출연연구기관 등의 설립·운영 및 육성에 관한 법률」 제2조에 따른 정부출연연구기관이므로 연구개발기관에 포함된다. 다시 말해 한국보건사회연구원은 해당 고시에 따라 연구사업 추진 시 연구데이터의 DMP관리를 의무적으로 수행해야 한다.

나. 연구데이터 관리 가이드라인

연구데이터 관리 가이드라인은 연구개발사업을 통하여 생산된 연구데이터를 보유하고 있는 연구기관의 연구데이터 관리 체계를 수립하고 데이터에 대한 수집·관리·공유·활용을 위한 기준을 마련하는데 목적을 가진다. 해당 가이드라인에 따르면 연구데이터는 FAIR(Findable, Accessible, Interoperable, Reusable)의 원칙을 준용해야 한다(국가과학기술연구회, 2019). 1) Findable(검색 가능) : 연구데이터에 메타데이터 등 영구식별자를 할당, 연구데이터 리포지터리에 등록하여 검색 가능하도록 해야 한다. 2) Accessible(접근 가능) : 표준화된 프로토콜을 사용하여 식별자를 통해 연구데이터에 접근 가능해야 한다. Interoperable(상호 운용 가능) : 연구데이터는 표준적인 방식과 어휘를 사용하여 상호 운용 가능하도록 해야 하며, 관련된 다른 데이터에 대한 인용을 포함해야 한다. Reusable(재사용 가능) : 명확한 데이터 라이선스와 프로비넌스 정보를 제공하여 재사용 가능하도록 해야 한다.

〈표 2-9〉 연구데이터 관리 절차

구성	세부구분	내용
연구데이터 관리체계 수립	정책과 절차	<ul style="list-style-type: none"> • 연구기관 내 데이터관리 기준 정의 • 데이터 관리에 필요한 정책과 절차를 수립하여 기관 내 공표
	전담 조직 및 인력	<ul style="list-style-type: none"> • 연구데이터 관리 및 활용 지원을 위한 전문인력 및 조직 구성
	요구사항 정의	<ul style="list-style-type: none"> • 기관에서 생산·보유하고 있는 연구데이터의 기본적인 사항(크기, 종류, 생산주기, 검색 요구사항 등)을 정의 • 체계적 관리와 폭 넓은 공유·활용을 위한 연구데이터 및 프로세스의 표준화 방안 마련
	연구데이터 인프라	<ul style="list-style-type: none"> • 데이터의 안정적 관리와 장기간 보존을 위한 필수적 인프라 제공
	지원 서비스	<ul style="list-style-type: none"> • 연구데이터의 관리·활용을 위한 교육, 자문, 지원 서비스 제공

〈표 2-9〉 연구데이터 관리 절차(계속)

구성	세부구분	내용
연구데이터 수집·저장	연구데이터 정제·표준화	<ul style="list-style-type: none"> 연구데이터 리포지터리 제출 연구기관에서 정의한 연구데이터 관리 기준 설정(표준화)
	메타데이터 작성	<ul style="list-style-type: none"> 연구기관에서 정의한 메타데이터 스키마 준용 장기간 보존 및 활용을 위한 충분한 정보 기술
	제출	<ul style="list-style-type: none"> 연구과제 종료 후 기간 내 데이터 관리 계획에서 지정한 데이터 리포지터리에 데이터 제출 데이터 공개 기준에 따라 비공개/내부공개/대외공개 등을 선택하고 공개기한 등을 지정
	승인 및 검토	<ul style="list-style-type: none"> 제출된 연구데이터에 대한 검토 및 승인 데이터 형식, 메타데이터 기술 내용, 공개 및 라이선스 적합 여부, 민감정보 포함 여부 등 검토
연구데이터 출판	출판	<ul style="list-style-type: none"> 데이터의 활용을 위하여 데이터 리포지터리 등 접근 가능한 형태로 출판 전문가 검토 등 품질 점검 절차 도입 가능
	데이터 식별자 부여	<ul style="list-style-type: none"> DOI(Digital Object Identifier) 등과 같은 식별자 부여를 권장함
	라이선스 적용	<ul style="list-style-type: none"> 대내·외 공개와 저작권 보호를 위한 다이센스 기준과 지원 서비스 제공
연구데이터 공유·활용	접근 및 공유	<ul style="list-style-type: none"> 연구데이터의 접근성 제고를 위하여 표준 프로토콜 및 기술형식에 따라 관리 및 배포 데이터 리포지터리의 검색에 노출되는 데이터 및 메타데이터의 내용을 연구자가 지정 해당 분야의 데이터 검색 사이트 또는 외부 연구데이터 사이트 등과의 연계를 통하여 접근성 제고
	활용	<ul style="list-style-type: none"> 연구데이터의 처리·분석을 위한 다양한 연구환경 제공 타 연구자의 데이터를 활용한 경우, 데이터 인용을 통하여 출처 표기

자료: 국가과학기술연구회(2019a). 연구데이터 관리 가이드라인.

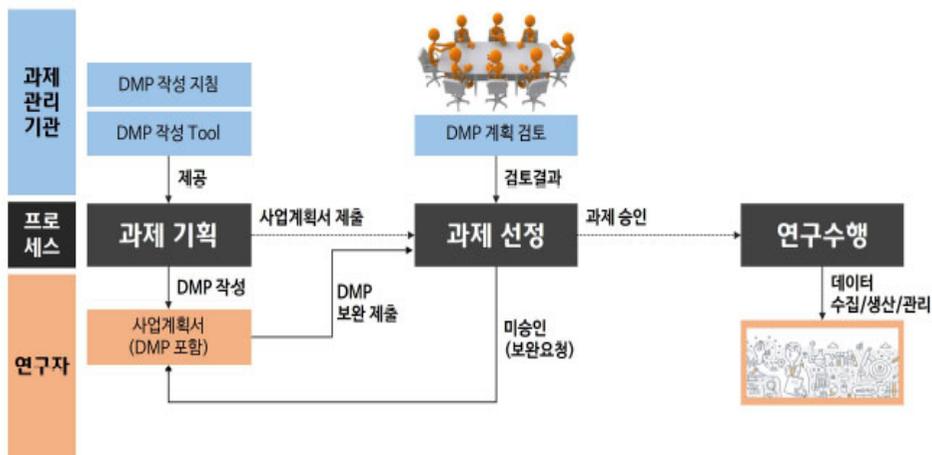
주: 자료 내 내용을 요약 작성함

다. 연구데이터관리계획 가이드라인¹³⁾

연구데이터관리계획(이하 DMP) 가이드라인은 DMP 작성 및 점검 기준을 설정하여 국가연구개발사업을 통하여 생산된 연구데이터의 체계적 관리를 촉진하는데 목적을 가진다.

해당 가이드라인에 따르면, 과제의 기획 및 선정 단계에서는 과제의 사업계획에 DMP를 포함하고 이를 검토하도록 권고하고 있다(아래 그림 참고).

[그림 2-11] 연구과제 기획 및 선정 단계에서의 DMP 작성 및 검토

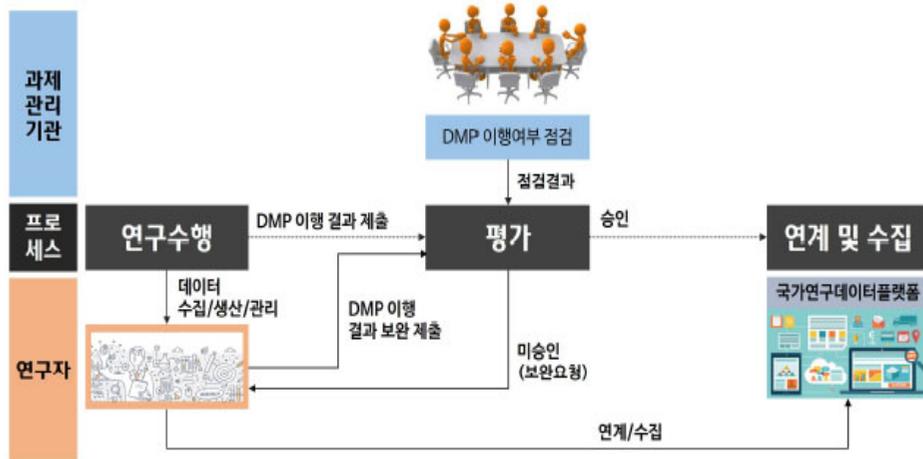


자료: 국가과학기술연구회(2019b). 연구데이터관리계획(DMP) 가이드라인.

또한, 과제를 평가할 시에 DMP에 대한 이행결과를 제출하고 이행 결과에 대한 보완 요청 및 반영 등의 피드백 과정을 통하여 DMP 이행 결과를 심의한다. 이후 DMP 이행 심의가 승인될 경우 데이터 수집 및 활용의 단계로 이행된다(아래 그림 참고).

13) 국가과학기술연구회(2019b). 연구데이터관리계획(DMP) 가이드라인.

[그림 2-12] 연구과제 수행·평가 단계의 DMP 점검



자료: 국가과학기술연구회(2019b). 연구데이터관리계획(DMP) 가이드라인.

DMP작성 시 필수적으로 포함되어야 하는 구성 요소과 그에 대한 세부적인 내용, DMP심의 체크리스트에 대한 사항은 아래 표와 같다. DMP작성 양식와 심의 및 이행점검 체크리스트는 공통적으로 데이터의 개요, 저장 및 장기 보존 계획, 공유 계획, 관리 및 책임자 정보 등을 요구한다. 여기서 데이터의 개요는 데이터의 접근성 제고를 위하여 데이터 수집 및 생산의 규범이나 표준을 따르고 있는지를 검토하는 영역이다. 데이터의 저장 및 장기 보존 계획에서는 데이터의 안전한 보호를 원칙으로 하며, 연구책임자 개인의 저장매체에 데이터를 보관하는 것은 적절하지 않은 것으로 판단하고 있다. 데이터 공유 계획에서는 데이터 공유를 위한 각종 정보의 첨부를 권고하고 있으며, 데이터 활용에 대한 제한이나 유의사항 등을 검토한다. 마지막으로 데이터 공유 책임자 정보는 데이터의 전반적 관리 상황에 대한 점검과 데이터 활용 관련 문제 발생 시 적절한 대응 절차 등을 검토하게 된다.

〈표 2-10〉 DMP작성 양식 예시

구성	세부구분	내용
연구데이터 개요	수집·생산 데이터 및 유형	<ul style="list-style-type: none"> 수집 데이터에 대한 간략한 설명(기존 데이터 혹은 타 데이터 소스 사용 등을 포함) 데이터 유형 및 크기(규모)에 대한 설명
	데이터 수집·생산 방법	<ul style="list-style-type: none"> 어떠한 방법 혹은 표준을 통해 데이터가 수집·생산될 것인지 설명 데이터 입수에 사용될 구체적인 방법에 대해 기재
연구데이터의 저장 및 장기적인 보존 계획	연구 수행 중 데이터 저장·백업	<ul style="list-style-type: none"> 연구 수행 중 데이터를 어디에 저장, 관리할 것인지 설명 데이터의 백업은 구체적으로 어떤 방법을 통해, 어떤 주기로 시행될 것인지? 문제가 생겼을 때 데이터 복원 방법은 무엇인지 등 기재
	연구 종료 후 데이터 보존	<ul style="list-style-type: none"> 연구 종료 후 장기적으로 어떤 장치 혹은 기관에 기탁함으로써 보존할 것인지 기재 연구 분야 또는 연구기관의 데이터 리포지토리가 존재하는지 기재하고 활용 여부 설명 ※ 연구자의 개인 컴퓨터, 연구실 서버 등은 장기적인 보존의 방법으로는 적합하지 않음
연구데이터의 공유 계획	데이터 공유 방법	<ul style="list-style-type: none"> 연구 종료 후 데이터를 언제, 어떤 방법으로 어디에, 어느 범위에서 공유할 것인지 기재 데이터의 공유 시점 기재 (연구 종료 후 2년 내에 공개하는 것이 권장됨) 데이터의 식별, 인용을 하는데 필요한 식별자(예: DOI, Digital Object Identifier)의 입수 및 부착 여부 기재 데이터 파일 이외에도 실제로 데이터를 이해하기 위해 필요한 문서(코드북, 설문지, 활용 시 유의사항 등)
	활용공유·공개 제한	<ul style="list-style-type: none"> 데이터를 공개하거나 공유하는데 제한이 되는 상황, 데이터의 제한이 있는지 기재
데이터관리 및 공유 책임자	책임자 이름	<ul style="list-style-type: none"> 연구책임자 혹은 연구 참여자 중에 데이터의 전반적인 관리, 저장 및 공유에 대한 세부적인 사항을 점검하고 의무 사항을 이행할 연구자의 이름 및 연락처

자료: 국가과학기술연구회(2019b). 연구데이터관리계획(DMP) 가이드라인.

주: 자료 내 내용을 요약 작성함

〈표 2-11〉 DMP검토 체크리스트

구성	내용
연구에서 수집 혹은 생산하고자 하는 데이터의 개요	<ul style="list-style-type: none"> • 데이터 수집 및 생산 방법이 해당 분야의 통상적인 규범이나 표준적인 방법을 따르고 있는가? • 연구 논문의 표나 그림 혹은 보고서의 부록에 수록되는 내용, 예비분석 결과, 논문이나 저술의 초안, 연구노트, 보고서 등은 해당되지 않는데 이런 유형의 데이터가 제시되고 있는가? • 데이터 파일 포맷 등이 가능한 범위에서 공개된 표준을 따르고 있는가?
데이터의 저장 및 장기적인 보존 계획	<ul style="list-style-type: none"> • 구체적인 데이터의 저장 및 장기 보존 계획이 제시되고 있는가? • 데이터 저장 방법이나 저장소는 안전성이 있는가? (연구자의 개인 컴퓨터, 연구실 서버 등은 장기적인 보존의 방법으로는 부적합) • 데이터 저장 및 장기 보존 계획이 실현 가능한가? (필요 경비 등)
데이터의 공개 및 공유 계획 작성	<ul style="list-style-type: none"> • 연구데이터의 공유 계획이 작성되었는가? • 공유 계획이 구체적이고 실현 가능한가? (시기 및 방법) • 해당 분야의 통상적인 규범이나 표준적인 방법으로 공유 계획이 작성되었는가? • 데이터 공유에 제한이 되는 사항(개인정보, 기밀 유지 필요, 지적재산권 및 잠재적인 상업적 가치 보호 등) 등이 파악되고 그에 대한 대안이 제시되고 있는가? • 데이터 파일 이외에 데이터를 이해하기 위해 필요한 자료(예: 설문지, 코드 북, 기술 및 보조 문서 등)가 첨부되었는가? • 데이터의 사용에 대한 제한 여부(라이선싱 조건, 엠바고 및 공유 제한 사항)가 파악되고 있는가?
데이터의 관리 및 공유 책임자 기재	<ul style="list-style-type: none"> • 데이터의 전반적인 관리, 저장 및 공유에 대한 세부적인 사항을 이행할 연구자의 이름 및 연락처가 기재되어 있는가?

자료: 국가과학기술연구회(2019b). 연구데이터관리계획(DMP) 가이드라인.

주: 자료 내 내용을 요약 작성함

5. 통계법

가. 조사통계의 진단

「통계법」은 통계의 작성·보급 및 이용과 그 기반구축 등에 관하여 필요한 사항을 정한 법률이며, 해당 법률의 제15조에서는 통계작성기관을 지정하고 있다. 한국보건사회연구원은 통계작성기관에 해당되어 「통계법」에서 명시하고 있는 통계작성기관으로서의 기능을 수행해야 한다. 법률 제11조에 따르면, 통계작성기관은 자체적인 통계품질진단을 추진해야 한다. 즉, 정기통계품질진단 및 수시통계품질진단 대상 통계 외에 통계작성기관에서 생성하는 통계는 자체적인 품질 진단 절차를 거칠 필요가 있는 것이다.

시행령 제15조에 따르면, 통계청은 자체통계품질진단 지침을 제공해야 한다. 이에 따라 「통계정보보고서」 작성 및 진단 지침에서는 조사통계 작성 및 진단 지침을 제시하고 있다.

〈표 2-12〉 조사통계 진단 지침 사항 요약

단계	내용
통계작성 기획	<ul style="list-style-type: none"> • 통계명, 작성기관, 법적 기준, 조사방법, 조사 및 공표 주기, 통계작성 개관, 통계작성 문서화 • 통계 연혁(개념, 분류, 설계, 과정, 내용, 방법, 표본, 기준년, 가중치 등의 변경 또는 개편이력 관리) • 통계 작성 목적(명확성, 활용 분야, 유사 사례 검토) • 주요 이용자 및 용도(관련성, 유형별 용도 파악), 이용자 의견 수렴(관련성 및 요구반영 결과 제시)
통계설계	<ul style="list-style-type: none"> • 조사내용 및 조사표 설계 <ul style="list-style-type: none"> - 조사항목(관련 정의, 체계성, 비교성) - 적용분류체계(비교성, 표준분류체계 사용 여부) • 조사표 구성(정확성, 비교가능성) <ul style="list-style-type: none"> ※ 조사표 필수 항목: ①조사명, ②조사목적, ③조사의 법적근거, ④국가승인통계로고, ⑤작성승인번호, ⑥응답자 협조사항, ⑦조사협조 감사인사, ⑧조사기관 명시, ⑨응답자 비밀보호정책, ⑩문의사항 연락처 • 조사표 설계 및 변경(적절성, 이력관리), 응답 소요시간 • 모집단 및 표본틀 <ul style="list-style-type: none"> - 목표모집단과 조사모집단의 명확성 및 차이의 적절성 - 표본추출틀의 정확성 및 표본 한계 등에 대한 보완 및 조치결과 • 표본설계 및 관리(정확성, 적절성, 타당성, 모수 및 분산 추정방법)
자료수집	<ul style="list-style-type: none"> • 조사방법의 정확성 및 적절성 • 조사원(채용, 처우, 교육, 업무) • 조사 실시(조사업무 흐름, 조사준비 및 예비조사, 항목별 조사방법) • 조사 질의응답의 체계(운영의 적절성, 응답 오류 사례 추적 및 관리) • 무응답 대처 및 표본 대체 • 사후조사의 수행 • 행정자료의 활용
통계처리 및 분석	<ul style="list-style-type: none"> • 행정자료를 활용한 보완(매칭의 정확성 및 적절성) • 자료 코딩(정확성, 오류 검출 방법), 자료 입력(전산입력 오류 대응 매뉴얼) • 자료내검(내검 방법 및 처리의 적절성, 내검매뉴얼, 에디팅 시스템 구축) • 무응답 수치 제시 및 항목 모응답률 산출 산식 제시 • 오차검증(측정 또는 처리 오차에 대한 추정, 주요 내용 기술) • 통계추정(가중치 조정, 통계추정 산식 및 내용) • 표집오차 추정 방법 및 결과(상대표준오차, 신뢰구간, 오차특성 등 제시) • 지수작성(지수의 산출 산식, 가중치 및 갱신 등) • 기타(계절 조정, 시계열 보정)

〈표 2-12〉 조사통계 진단 지침 사항 요약(계속)

단계	내용
통계공표, 관리 및 이용자 서비스	<ul style="list-style-type: none"> • 공표통계 해석 방법 및 정확성(공표통계의 적절성, 주요 통계표 및 그래프 제시, 해석 방법 및 이용 시 유의사항, 분석결과 관리) • 시의성 및 정시성(공표 시점 제시 및 적절성, 공표일정) • 비교성(통계작성방법/시계열/국가간 비교성) • 일관성(조사항목, 작성주기, 잠정치와 확정치의 차이, 일회성 수정) • 접근성 및 명확성(공표 방법, 통계설명자료 제공, 조사관리정보) • 마이크로데이터 생성·관리 및 서비스(제공, 접근성, 정확성) • 비밀보호 및 보안(조사 관련 비밀사항 보호, 자료 보안 및 접근 제한 설정)
통계기반 및 개선	<ul style="list-style-type: none"> • 기획 및 분석 인력 • 사업예산 • 통계위탁 조사 • 통계품질관리 및 개선(품질제고 계획 및 추진실적 관리)

자료: 통계청(2022). 국가통계 품질관리 매뉴얼: 정기통계품질진단. 통계청.
 주: 자료 내 내용(조사통계 작성 및 진단지침)을 요약 작성함

나. 통계자료의 제공

「통계법」 제28조에 따르면, 통계작성기관은 국민들이 해당 통계 데이터를 신속하고 편리하게 이용할 수 있도록 관련 조치를 취해야 한다. 또한 동법 제29조의2에 따라 통계작성기관은 통계의 보급 및 이용의 활성화를 위하여 통계자료를 보유하고 관리하여야 한다.

특히 동법 제31조에 따라 통계작성기관은 통계이용자의 요청에 통계자료의 활용 목적·내용·범위 등의 타당성을 검토하여 해당통계자료를 제공하여야 한다. 통계자료 제공 시에 필수적으로 개인정보를 위한 처리가 이루어져야 하지만(제3항), 통계응답자 자신이 응답한 자료를 요구하는 경우나 인구총조사 등 전수조사로 인하여 수집된 정보의 경우에는 이러한 처리 대상에서 제외될 수 있다.

한편, 「통계법」 제30조에서는 통계작성기관 간의 통계자료 공유를 명

시하고 있다. 이때에는 특정 개인이나 법인·단체 등을 식별할 수 없는 형태로 자료를 처리한 후에 제공하여야 한다.

6. 데이터기반행정 활성화에 관한 법률

「데이터기반행정 활성화에 관한 법률(이하 데이터기반행정법)」은 데이터 기반 행정의 활성화를 위하여 필요한 사항을 정함으로써 공공기관의 책임성·신뢰성을 높이기 위한 객관적이고 과학적인 행정의 추진을 목적으로 한다. 본 법령에서 말하는 공공기관에는 「정부출연연구기관 등의 설립·운영 및 육성에 관한 법률」에 따라 설립된 정부출연연구기관이 포함되므로, 한국보건사회연구원 또한 해당 법령에 따라 관련 책무를 이행해야 한다.

「데이터기반행정법」 제16조에 따르면, 공공기관은 생성하거나 취득하여 관리하는 데이터에 대한 메타데이터 및 데이터 관계도를 체계적으로 관리하여야 한다. 각 공공기관은 데이터기반행정 책임관을 두어 데이터기반행정 관련 데이터의 연계·제공·공동활용에 관한 업무를 추진해야 한다(제19조 제2항).

그 외에도 공공기관 간의 공동활용의 필요가 인정되는 데이터의 경우(정책수립, 서비스 제공, 안전 등)에는 통합관리 플랫폼에 등록할 수 있으며, 미등록 데이터의 경우에는 상호 요청·제공할 수 있다(제10조).

제2절 조사데이터 관리 및 활용 사례 분석

1. 연구데이터 관리 및 활용 사례

가. 연구데이터 리포지터리

연구데이터의 효율적인 관리와 공유를 위한 방안으로 연구데이터 리포지터리(Research Data Repository)가 활용되고 있다. 연구데이터 리포지터리는 데이터의 수집·관리, 검색 및 활용 등을 지원하기 위한 시스템이라 볼 수 있다.

국내 연구데이터 리포지터리 활용 사례를 살펴보면, 아래 표와 같다. 각 연구기관에서 운영하는 연구데이터 리포지터리는 공통적으로 연구데이터의 수집과 관리, 활용의 효율성 제고를 목적으로 하고 있다. 이에 따라 각 리포지터리 시스템들은 DMP의 등록·관리, 연구데이터의 수집 및 등록 전반(제출, 검토, 등록 및 관리 등), 연구데이터의 검색 및 활용, 타 시스템과의 연동 등의 기능을 탑재하고 있다.

〈표 2-13〉 연구데이터 리포지터리 운영 현황

리포지터리	내용	주요기능	특징
KEI 연구데이터 리포지터리	<ul style="list-style-type: none"> • 한국환경연구원 운영 • 2020년 시범 운영개시 	<ul style="list-style-type: none"> • DMP 등록·관리 • 연구데이터 제출 (제출-검토-승인) • 연구데이터 검색·조회 • 연구데이터 활용 (신청-승인, 다운로드) 	<ul style="list-style-type: none"> • 간편 로그인(SSO) • 연구과제정보 연계를 통한 데이터 제출 컬렉션 자동 생성
지질자원 연구데이터 리포지터리 (GDR)	<ul style="list-style-type: none"> • 한국지질자원연구원운영 • 2015년 운영 개시, 2021년 고도화 	<ul style="list-style-type: none"> • DMP 등록·관리 • 연구데이터 제출 (제출-결재-등록완료) • 연구데이터검색·조회 • 연구데이터 활용 (신청-승인, 다운로드) 	<ul style="list-style-type: none"> • 간편 로그인(SSO) • 연구과제 및 참여 연구원 정보 연계 • 전자결재 시스템 연동 • 기관 고유 분류 체계를 적용한 데이터 등록·검색 • 지리정보 기반 검색
한의학 데이터 리포지터리 (KMDR)	<ul style="list-style-type: none"> • 한국한의학연구원 운영 • 2019년 운영 개시 	<ul style="list-style-type: none"> • DMP 등록·관리 • 연구데이터 제출 (제출-검토-결재-등록완료) • 연구데이터검색·조회 • 연구데이터 활용 (신청-승인, 다운로드) 	<ul style="list-style-type: none"> • 간편 로그인(SSO) • 연구과제 및 참여 연구원 정보 연계 • 전자결재 시스템 연동 • 기관 고유 분류 체계를 적용한 데이터 등록·검색

자료: 송미영, 이미경, 이은경, 이정찬, 서정호, 박지윤(2022). KICE 교육데이터 활용 활성화 방안 탐색 및 시스템 설계. 한국교육과정평가원.

주: 자료 내 표 내용 일부 편집(한국보건사회연구원 사례 생략 등)

나. 데이터 공유 플랫폼

연구데이터의 제공과 활용 활성화를 위하여 데이터 공유 플랫폼을 운영하는 사례들이 있다. 데이터 리포지터리는 데이터의 축적 및 관리가 주된 목적이지만 데이터 공유 플랫폼은 데이터의 공유 및 활용을 주된 목적으로 운영된다. 데이터 공유 플랫폼은 데이터의 활용을 지원하기 위한 공간 제공, 분석 서비스 제공 등을 통하여 데이터 이용자 친화적 기능을 탐

재하고자 한다.

데이터 공유 플랫폼 사례의 특성들을 살펴보면(아래 표 참고), NDIS는 연구데이터 공유 외에도 정책 이슈 모니터링 정보를 제공한다는 점에서 특징적이다. Data ON은 과학기술 분야 연구데이터에 대한 검색을 지원하며, 플랫폼 내 데이터 분석 서비스를 제공하고 있다. KOSSDA는 기탁 자료를 바탕으로 데이터 아카이브를 구축 운영하고 있으며, 마이크로데이터의 활용을 위한 각종 정보 검색이 가능하다. 지오빅데이터 오픈플랫폼은 관련 분야 데이터 검색에 강점을 보이며, DataON 데이터 플랫폼과의 연계를 통하여 데이터 접근성을 높이고 있다.

〈표 2-14〉 데이터 공유 플랫폼 운영 현황

플랫폼	개요	주요기능	특징
NRC 데이터정보 시스템 (NDIS)	<ul style="list-style-type: none"> 경인사연 운영 2021년 운영 개시 	<ul style="list-style-type: none"> 정책이슈 모니터링 (빅)데이터 분석서비스 데이터(정보) 제공 	<ul style="list-style-type: none"> 소속 기관 연구데이터 추가 구축 추진 중 (제공)빅데이터 분석 서비스 및 온라인 분석 도구 제공
국가 연구데이터 플랫폼 (DataON)	<ul style="list-style-type: none"> 과학기술정보통신부·한국과학기술정보연구원 운영 2018년 시범 운영개시 2020년 정식 운영개시 	<ul style="list-style-type: none"> 연구데이터 검색·활용 연구데이터(연계)등록·관리 온라인 분석서비스 커뮤니티 	<ul style="list-style-type: none"> 데이터셋 의미 관계 서비스 제공 기관 리포지터리 연계를 통한 데이터 등록 가능 데이터 저장·공유를 위한 스토리지 제공 온라인 데이터 분석 도구 제공 등
한국사회과학 자료원 (KOSSDA)	<ul style="list-style-type: none"> 서울대 아시아연구소 운영 2006년 데이터아카이빙 시작 2019년 신규 웹사이트 운영 개시 	<ul style="list-style-type: none"> 데이터 검색·제공 데이터 기탁 온라인 통계분석 	<ul style="list-style-type: none"> (제공)기탁 데이터로 운영 데이터-연구 성과물 간 정보 연계 제공 온라인 통계분석 도구 제공
지오빅데이터 오픈플랫폼	<ul style="list-style-type: none"> 한국지질자원연구원 운영 2020년 운영 개시 	<ul style="list-style-type: none"> 연구데이터 등록·검색·제공 지질도 서비스 주제별 정보서비스 	<ul style="list-style-type: none"> (제공)기관 산출 연구데이터 개인 연구자 데이터 등록·관리 가능 다양한 부가 서비스 제공

자료: 송미영 외(2022). KICE 교육데이터 활용 활성화 방안 탐색 및 시스템 설계. 한국교육과정평가원.

주: 자료 내 표 내용 일부 편집

다. 데이터 제공 시스템

경제·인문사회연구회 소속 연구기관의 데이터 제공 시스템 사례를 살펴보면, 아래 표와 같다. 각 연구기관들은 데이터의 접근성 제고를 위하여 웹서비스를 통한 데이터 제공을 추진하고 있으며, 이때 데이터 이용 조건 등에 대한 동의, 데이터 활용 목적 등을 내부적으로 검토하여 데이터이용자에게 데이터를 제공하고 있다. 또한, 주기적으로 작성되는 조사 데이터인 패널데이터 외에도 각 기관에서의 조사연구 수행 결과물을 해당 서비스를 통하여 제공하고 있다.

〈표 2-15〉 경제·인문사회연구회 소속 연구기관 데이터 제공 웹서비스 현황

플랫폼	웹서비스	주요 제공데이터	특징 및 제공
한국청소년 정책연구원	한국아동·청소년데이터 아카이브	<ul style="list-style-type: none"> • 한국아동·청소년패널 조사 등 • 한국아동·청소년통합 조사 • 아동·청소년인권실태 조사 • 청년사회경제실태조사 	<ul style="list-style-type: none"> • 이용조건 동의 · 사용자 정보 입력
육아정책 연구소	한국아동패널	<ul style="list-style-type: none"> • 한국아동패널 	<ul style="list-style-type: none"> • 사용자 소속 및 목적 입 력 → 데이터 검색 및 다 운로드
한국직업능력 연구원	기관 홈페이지 내 '패널연구'	<ul style="list-style-type: none"> • 한국교육고용패널 • 인적자본기업패널 	<ul style="list-style-type: none"> • 이용조건 동의·사용자 정 보 입력 → 데이터 다운 로드
한국교육 개발원	기관 홈페이지 내 '조사자료'	<ul style="list-style-type: none"> • 한국교육중단연구 • 교육과 사회계층 이동 조사 • 평생학습 개인실태 조사 • 학교 교육 실태 및 수준분석 	<ul style="list-style-type: none"> • 조사자료 신청 → 심의 및 승인 → 데이터 제공(이메일) • 연구용 데이터 신청은 홈페이지를 통해 이루어지나, 데이터는 이메일로 제공
한국여성정책 연구원	성인지통계	<ul style="list-style-type: none"> • 성인지통계 • 여성가족패널 • 여성관리자패널 	<ul style="list-style-type: none"> • 통계 데이터 공유 플랫폼 과 패널데이터 제공 플 랫폼 통합
한국형사· 법무정책 연구원	범죄와 형사사법 통계정보	<ul style="list-style-type: none"> • 전국범죄피해조사 • 각종 연구데이터 	<ul style="list-style-type: none"> • 통계 데이터 공유 플랫폼 과 기존 연구데이터 제 공(보고서, MD 등) 통합
한국행정 연구원	기관 홈페이지 내 '연구자료'통계DB'	<ul style="list-style-type: none"> • 각종 조사자료 	<ul style="list-style-type: none"> • 이용 신청서 작성 후 자 료 다운로드 가능

자료: 송미영 외(2022). KICE 교육데이터 활용 활성화 방안 탐색 및 시스템 설계. 한국교육과정평
가원.

주: 자료 내 표를 활용하여 편집, 일부 내용 추가 작성

2. 기관별 조사데이터 관리 사례

각 연구기관 또는 데이터제공기관의 조사데이터 관리 사례를 살펴보고자 하였다. 특히 각 기관의 지침 및 매뉴얼 등 조사데이터의 관리와 관련된 기반을 통하여 그 특성을 검토하고자 하였다.

가. 한국교육개발원

한국교육개발원 내 조사데이터 관련 규정으로는 조사자료 공개지침, 생명윤리심의위원회 운영 지침, 조사도구 및 조사자료 관리 지침 등이 있다. 먼저, 조사자료 관리 및 공개 지침(2023b)에서는 조사자료 공개 절차의 내용을 명시하고 있다. 연구사업 수행 중 수집된 조사 자료 중에 연구책임자가 공개가 필요하다고 판단되는 조사자료는 연구기획실장, 조사분석연구실장의 검토를 거쳐, 부서장, 기획처장, 원장의 결재를 득한 후에 공개하게 된다. 또한, 조사자료를 연구용 데이터와 교육용 데이터를 구분하며, 연구용 데이터는 원자료 제공을 원칙으로 하지만 필요에 따라 일부 정보가 제한될 수 있다.

생명윤리심의위원회 운영 지침(2023)의 주된 내용은 공용기관위원회의 지침을 준수하고 있다. 다만, 한국보건사회연구원 운영 지침과는 달리 각 조문에 심의의 종류를 세분화(제4장 심의 종류)하고 있으며, 제29조에 따라 종료보고 심의를 진행한다. 여기서는 종료보고 심의 신청서, 최종 승인된 연구계획서, 최종연구보고서 등을 제출하게 되며, 이상의 신청 절차는 연구과제 종료 이전에 이루어져야 한다.

한편, 지침 제40조(한국교육개발원, 2023a)에서는 연구대상자가 자신의 정보 공개를 요청할 수 있음을 명시하고 있는데, 이때 연구책임자 및

위원회의 심의 및 관련 절차를 구체화하였다는 점에서 특징적이다.

조사도구 및 조사자료 관리 지침(2020) 제6조에 따르면, 기본연구 등의 연구책임자는 연구종료 결과보고 시 연구·사업 수행 과정에 생성된 조사 자료 목록을 함께 보고해야 하며, 당해 연도 말일까지 조사자료를 연구기획실에 제출하여야 한다. 양적 조사자료의 경우에는 조사도구, 조사 설계 등을 포함한 설명문과 코드북을 함께 제출해야 하며(제7조 제1항), 질적 조사자료의 경우에는 녹취록 또는 녹음파일을 제출해야 한다(제7조 제2항). 조사자료는 연구기획실에서 취합하게 되며(제9조), 해당 연구자료는 연구종료 시점으로부터 최소 5년간 보관된다(제10조).

한편, 제11조(한국교육개발원, 2020)에 따르면, 수탁과제의 경우에 조사자료의 안정적 관리를 위하여 해당 자료를 연구기획실에 제출하여야 한다. 다만 수탁과제의 경우(별도의 관리 체계를 가진 경우)에는 이를 생략할 수 있다. 뿐만 아니라 데이터의 제공은 원칙적으로 발주기관에서 하는 것을 명시해 두었다(제4항).

조사 자료 사용에 대한 안내

□ 조사 자료 신청 절차

- ① 개발원의 조사 자료 공개 페이지의 공개 자료 목록 확인
- ② 조사 자료 신청서 제출[별지서식 1]
 - * 연구 목적으로 공개된 자료의 경우 교육용으로 신청할 수 없음.
 - * 교육용 자료는 변인이나 사례 수가 제한되어 있기 때문에 연구 목적으로 활용할 수 없음.
 - * 연구 목적으로 신청하는 경우 모든 연구진의 정보를 작성하고, 교육용자료 신청은 교육책임자만 작성할 것.
 - * 신청서 내용과 관련하여 개발원은 추가적인 서류를 요청할 수 있음.

- ③ 조사 자료 신청 승인 여부 심의
- ④ 심의 결과 통지 및 자료 제공

□ 조사 자료 사용자의 의무

- ① 개발원의 조사 자료 사용자는 개인, 학교, 단체 등의 식별이 가능한 연구 결과를 공개해서는 안 된다.
- ② 연구목적에 한해 공개된 조사 자료 활용 과정에서 연구 목적, 연구자 등 조사 자료 신청서에 제시된 내용이 변경되는 경우, 개발원에 변경된 신청서를 제출하여 승인을 받아야 한다.
- ③ 개발원이 제공한 조사 자료를 조사 자료 신청서에 제시된 연구자 이외의 타인에게 복사해주거나 양도할 수 없다.
- ④ 개발원이 제공한 조사 자료를 활용한 연구 결과를 발표하는 경우, 자료 출처를 명시해야 하며, 그 결과물을 30일 이내에 개발원에 제출해야 한다.
- ⑤ 교육용 조사 자료의 활용은 조사 자료 신청시 작성된 교육 기간과 개설 강좌에 한하며, 교육 기간 종료 후 교육책임자가 자료를 폐기해야 한다.
- ⑥ 사용자가 개발원의 조사 자료 공개 지침을 위반할 경우, 향후 10년간 개발원의 조사 자료에 대한 공개 신청 및 사용을 제한한다.

□ 관련문의

- 조사 자료 신청, 조사 자료 관련 문의, 조사 자료 활용 서지 사항 정보는 rsdb@kedi.re.kr 이메일을 이용한다.

나. 한국여성정책연구원¹⁴⁾

한국여성정책연구원 통계사무 및 원시자료 서비스 지침에서는 통계작성·보급 및 이용에 따른 사항을 규정하고 있다. 해당 지침에서는 통계사무의 원칙(정확성, 진실성, 적시성, 명확성 등), 통계심의위원회의 운영, 자체통계품질진단 등에 사항을 규정하고 있다.

조사데이터의 제공과 관련하여 제3장 통계자료 제공에서 구체적으로 제시하고 있는데, 통계자료의 제공은 이용자 및 이용방법에 제한없이 제공하는 것을 원칙으로 하고 있으나 사회적으로 증대한 영향을 미칠 가능성이 있는 경우, 비밀 또는 대외비로 분류된 연구의 경우, 공동 및 수탁사업의 경우, 최소 엠바고 기간이 지나지 아니한 경우, 연구가 진행중인 경우 등에 한하여 제공하지 않을 수 있다.

특히 비승인 통계(제15조), 원시자료 제공(제16조)의 경우에는 연구계획서와 데이터 활용 관련 서약서 작성이 요구된다. 또한 통계자료의 관리를 위하여 통계결과 및 원시자료 기본 정보를 수집하여 기록·관리하고 있다. 조사데이터의 경우, 조사방법을 기재해야 하며 조사과정을 통하여 생성된 자료 전체를 제출하여야 한다. 연구책임자는 자료공개여부를 결정하고 원시자료가 공개될 경우에 이에 대한 문의를 할 수 있는 책임자를 지정할 수 있다.

14) 한국여성정책연구원(2021). 통계사무 및 원시자료 서비스 지침

[그림 2-13] 한국여성정책연구원 통계결과 및 원시자료 기본정보 양식

통계결과 및 원시자료 기본정보										
보고서명								과제구분 기본() / 수탁() / 수시()		
연구기간										
연구자										
발간부수		부								
통계자료 정보	통계명칭									
	통계종류		<input type="checkbox"/> 승인통계 <input type="checkbox"/> 기타 ()							
	조사(보고)항목									
	조사(보고)대상		단 위	<input type="checkbox"/> 개인 <input type="checkbox"/> 가구 <input type="checkbox"/> 사업체 <input type="checkbox"/> 기타 ()						
			모 집 단							
			법 위 및 규 모 지 역							
	작성기준시점(기간)		조사(보고)기간		작성주기					
조사(보고)방법		작성형태		<input type="checkbox"/> 조사 <input type="checkbox"/> 보고(업무) <input type="checkbox"/> 가공						
		조사(보고)대상선		<input type="checkbox"/> 전수 <input type="checkbox"/> 확률표본 <input type="checkbox"/> 유의표본 <input type="checkbox"/> 기타 ()						
		자료수집방법		<input type="checkbox"/> 면접조사 <input type="checkbox"/> 전화조사 <input type="checkbox"/> 우편조사(모사전송) <input type="checkbox"/> 온라인조사 <input type="checkbox"/> 집합조사 <input type="checkbox"/> 기타 ()						
제출자료	구분		내용				연구과제관리시스템 탑재여부 및 제출여부			
	보고서전문파일(hwp, PDF)									
	요약문									
	전문가회의자료						-			
	기타자료						-			
보고서인쇄부(3										
보고서명								과제구분 기본() / 수탁() / 수시()		
부)										
파일설계서		유 () 무 ()								
설문지		유 () 무 ()								
코딩북		유 () 무 ()								
자료공개여부		공개가능		원내() / 원외()		공개불가 ()				
원시자료 문의처		성명				연락처				
		소속								
첨부서류 : 1. 조사표 또는 보고서식 2. 결과표 서식 3. 조사요령서 또는 조사사항에 대한 용어해설자료 4. 표본조사의 경우 표본설계내역 5. 통계결과 작성방법(조사항목별 검토사항 등)에 대한 설명자료 6. 통계결과 용어해설자료										

자료: 한국여성정책연구원(2021). 통계사무 및 원시자료 서비스 지침

다. 한국직업능력연구원¹⁵⁾

한국직업능력연구원의 통계자료 관리지침에서는 연구원에서 생산된 실태조사 자료(패널조사 제외, 연구원에서 생산한 각종 연구 조사 자료) 및 패널조사 자료(패널조사 결과물 전체)의 제공 기준을 설정하고 있다. 해당 지침의 제5조에 따르면, 통계자료의 제공 방법은 자료의 제공방법은 웹페이지를 이용한 다운로드, 이동식 저장장치에 복사, 전산자료의 단순열람, 프린트 아웃(Print-out) 등으로 규정된다.

통계자료를 이용하는 자는 통계자료 제공시스템에 접속하여 이용목적, 이용시간 등 관련 양식을 작성하여 제출해야 한다(제6조). 또한 통계자료를 수령한 데이터 이용자는 자료를 임의로 타인에게 복제 및 대여할 수 없으며 연구종료 후 즉시 자료를 파기하여야 한다. 이상의 사항을 위반할 경우에 발생하는 제반 문제에 대해서는 데이터 이용자가 그 책임을 감수하여야 한다. 그 외에도 데이터 이용자는 통계활용 및 분석방법에 대한 문의에 대해 성실하게 협의할 것을 명시하고 있다.

통계자료 관리지침에서는 실태조사 관리책임자를 통향·데이터분석센터(이하 센터)로 지정하고 있으며 관리책임자는 실태조사 자료에 관한 업무를 총괄하게 된다(제7조). 본 연구원에서 수행한 실태조사는 실태조사 최종금 지급 시 원데이터, 설문지, 레이아웃과 함께 결재를 받아야 하며, 외부자료 공개 불가 시 이에 대한 사항을 지정해야 한다. 센터는 결재 시 포함된 정보를 수집 관리해야 한다.

실태조사 자료 가운데 1년이 지난 기본과제의 경우, 수탁과제가 아닌 경우 데이터를 제공할 수 있다. 다만, 연구책임자 또는 부원장의 결재를 득할 필요가 있으며, 연구책임자의 의견에 기초하여 센터에서는 외부 제

15) 한국직업능력연구원(2021). 통계자료 관리 지침.

공여부를 공개한다.

라. 한국교통연구원¹⁶⁾

한국교통연구원 연구조사자료관리규칙은 연구사업에 필요로 하는 자료의 조사, 수집, 이관, 관리, 활용 등에 관한 사항을 규정하고 있다. 여기서 말하는 연구조사자료란 조사 과정을 통하여 정리·집계된 자료, 최종 분석 결과, 외부기관 전산데이터 자료를 말한다.

해당 규칙에 따라 조사계획 수립 시 자료의 이용목적, 자료의 속성(항목, 지역 및 시간적 범위, 표본율), 최종 분석 결과물을 사전에 작성해야 하며, 기존 자료와의 중복을 피하기 위하여 정보자료실과 사전 협의를 거쳐야 한다. 연구책임자는 연구사업 수행 시 조사분석된 자료를 연구사업 종료와 동시에 정보자료실에 이관해야 한다.

조사분석 자료는 정보자료실에서 분류 보관하며, 이 가운데 교통관련 연구사업에 유익하다고 판단되는 자료를 이용자가 다목적으로 활용할 수 있도록 분류하고 데이터베이스화 할 수 있다. 외부기관에서 연구데이터를 활용하고자 할 때에는 사용신청서를 작성, 검토 후 자료를 제공할 수 있다. 한편, 연구사업에 사용중인 기밀 자료는 연구책임자가 보안 관리하여야 함을 원칙으로 한다.

16) 한국교통연구원(2016). 연구조사자료관리규칙.

[그림 2-14] 한국교통연구원 조사수집 자료의 내역 양식

조사 수집 자료의 내역			
자 료 명	자료의 지역적 범위	자료의 시간적 범위	표 본 율(%)
1. 자료 이용 목적:			
2. 자료항목(field)의 속성(항목 값에 대한 내용 및 단위)			
3. 자료의 최종 목적 Layout			
4. 기 타			

자료: 한국교통연구원(2016). 연구조사자료관리규칙

마. 한국형사·법무정책연구원¹⁷⁾

한국형사·법무정책연구원 자료관리 규정에서 조사자료 관리의 사항을 규정하고 있다. 일반적으로 연구기관 내 자료관리규정(또는 지침)이 주로 기관 내 도서관에서 수집하는 도서류 및 전산자료 등의 사항을 규정하고 있는 것과는 차별성을 가진다.

해당 규정 제3조 제2항에 따르면 “연구자료”는 연구과제에 참여한 모든 연구자가 생산한 연구와 관련된 인터뷰 결과, 설문조사 결과, 패널조사 결과 등 데이터베이스(DB)를 말한다. 한편, 해당 규정 제4조에서는 자료관리위원회의 기능을 명시하고 있다. 위원회는 자료의 관리 및 기타 자료에 대한 중요사항을 심의하는 기능을 수행한다. 반면, 2015년 이후 자료관리 규정 내에 자료의 공개 및 서비스에 대한 사항이 삭제되었다.

17) 한국형사·법무정책연구원(2021). 자료관리규정.

[그림 2-15] 한국형사·법무정책연구원 조사자료 데이터베이스 원표

조사자료 데이터베이스 원표	
1. 일련 번호 :	
2. 보고서 번호 :	
3. 등록 일자 :	
4. 갱신 일자 :	
5. 공개 여부 :	
----- 상단은 관리자 기재 -----	
6. 연구과제명 :	
7. 연구진 :	
8. 조사명 :	
※ 데이터베이스 원표는 조사과제별로 각각 작성. (하나의 연구에 여러 개의 조사가 있다면 조사의 수대로 원표 작성)	
9. 조사대상 :	
10. 표본 및 사례수 :	
11. 조사방법(괄호에 ○로 표시. 단, 기타의 경우 해당 내용 기입)	
1) 설문조사() 2) 기록조사() 3) 기타()	
12. 조사기간: 년 월 일 - 년 월 일	
13. 제출파일(괄호안은 숫자로 파일의 개수 기입. 단, 기타의 경우 해당 내용 기입)	
- 조사표: 한글파일()개 MS워드파일()개 엑셀파일()개 기타()	
- 조사결과 데이터(원자료): SPSS파일()개 SAS파일()개 엑셀파일()개 기타()	
- 코드북: 한글파일()개 MS워드파일()개 엑셀파일()개 기타()	
※ 리코딩 작업 등으로 변수 혹은 보기문항을 변경하여 조사표와 조사결과 데이터 파일이 상이한 경우에는 반드시 제출 (조사표 내용과 데이터 파일이 형식이 동일한 경우 제출하지 않아도 되며, 일부분만 상이한 경우 변경되거나 추가된 부분의 코드북만 작성하여 제출가능)	
14. 자료 공개/비공개 여부(괄호에 ○로 표시)	
공개() 부분공개() 비공개()	
15. 비공개/부분공개 사유 (부분공개의 경우, 공개 범위도 함께 기재)	

자료: 한국형사·법무정책연구원(2021). 자료관리규정.

바. 한국사회과학자료원¹⁸⁾

서울대학교 한국사회과학자료원(Korea Social Science Data Archive, 이하 KOSSDA)은 사회과학 전 분야의 연구데이터를 수집, 데이터 베이스화하는 전문 데이터 아카이브이다. KOSSDA는 양적·질적 조사자료 외에도 각종 집합데이터 등을 기탁받아 관리하고 있다.

KOSSDA는 조사데이터를 직접 생산하지 않지만 기탁받은 각종 데이터를 관리하고 공유·활용할 수 있는 플랫폼을 운영하고 있다. 이는 데이터의 소유권을 가진 생산주체와 관리주체 간의 분리를 의미한다. 데이터를 생산한 기관 또는 개인이 데이터의 장기 보관과 데이터의 활용도 제고를 위하여 KOSSDA에 조사데이터를 기탁하게 되는데, 이때 기탁 전 확인절차를 통하여 기탁 대상 조사데이터의 적정성을 판단하게 된다. KOSSDA의 자료기탁 가이드에 따르면, 자료 수집 시 주요 고려 사항은 완결성(데이터가 모든 필수 정보를 포함하고 있는지, 데이터 이해를 위한 필수 정보와 문서가 충분한지), 보존필요성(지속 보존의 가치가 있는지), 잠재적 재활용성(기존의 목적을 넘어선 활용의 가능성), 법적·윤리적 문제(저작권, 데이터 생성자의 사전 동의, 개인정보 및 민감정보의 처리 동의 등)이다. 특히 자료의 재활용에 있어서 연구윤리적 문제가 해결되지 않는 경우에는 데이터의 기탁 또는 공유에 제한을 가진다. 이에 따라 KOSSDA에서는 조사데이터의 수집 전 또는 수집 후에 연구윤리심의를 통한 사전 승인을 권고하고 있다.

KOSSDA의 데이터 관리 단계는 아래 그림과 같으며, 음영 처리된 부분이 기탁 협약 이전에 검토되는 사항 또는 그 절차에 해당된다(그림 참고).

18) 한국사회과학자료원(2023a). 자료 기탁 가이드.

[그림 2-16] 한국사회과학자료원 데이터 관리 단계



자료: 한국사회과학자료원(2023a). 자료 기탁 가이드.

KOSSDA에서는 양적·질적 조사 데이터를 모두 수집·관리·제공¹⁹⁾하고 있다. 다만, 자료의 형태와 관계없이 데이터의 저작권 문제의 처리, 개인정보 보호 및 민감정보의 처리, 데이터에 대한 IRB승인을 필수적으로 요구한다.

19) 자료 공개대상은 일반공개, 회원공개, 제한 공개로 구분되며, 제한공개 경우에는 기탁 기관의 승인을 득하여야 이용할 수 있다.

[그림 2-17] 자료 유형별 수집 형식

• 양적 자료 파일 유형

종류	내용	수집 파일 형식
데이터	최종 버전의 raw data 파일	sav, dta, xlsx, csv 등
설문지	조사표	hwp, doc, pdf 등
보고서	결과보고서, 논문 등	hwp, doc, pdf 등
메타데이터	자료 설명 문서	hwp, doc, xlsx, pdf 등
코드북	코드북, 코딩가이드 등	hwp, doc, xlsx, pdf 등
기타	유저가이드, 서베이리포트, 샘플링 정보, 면접자 지침서, 조사 예산 및 기타 주요 계획 문서, PPT 및 기타 관련 자료, 지도, 산업/직업 분류표, 빈도표, 자료이용문헌목록 등	hwp, doc, xlsx, pdf 등

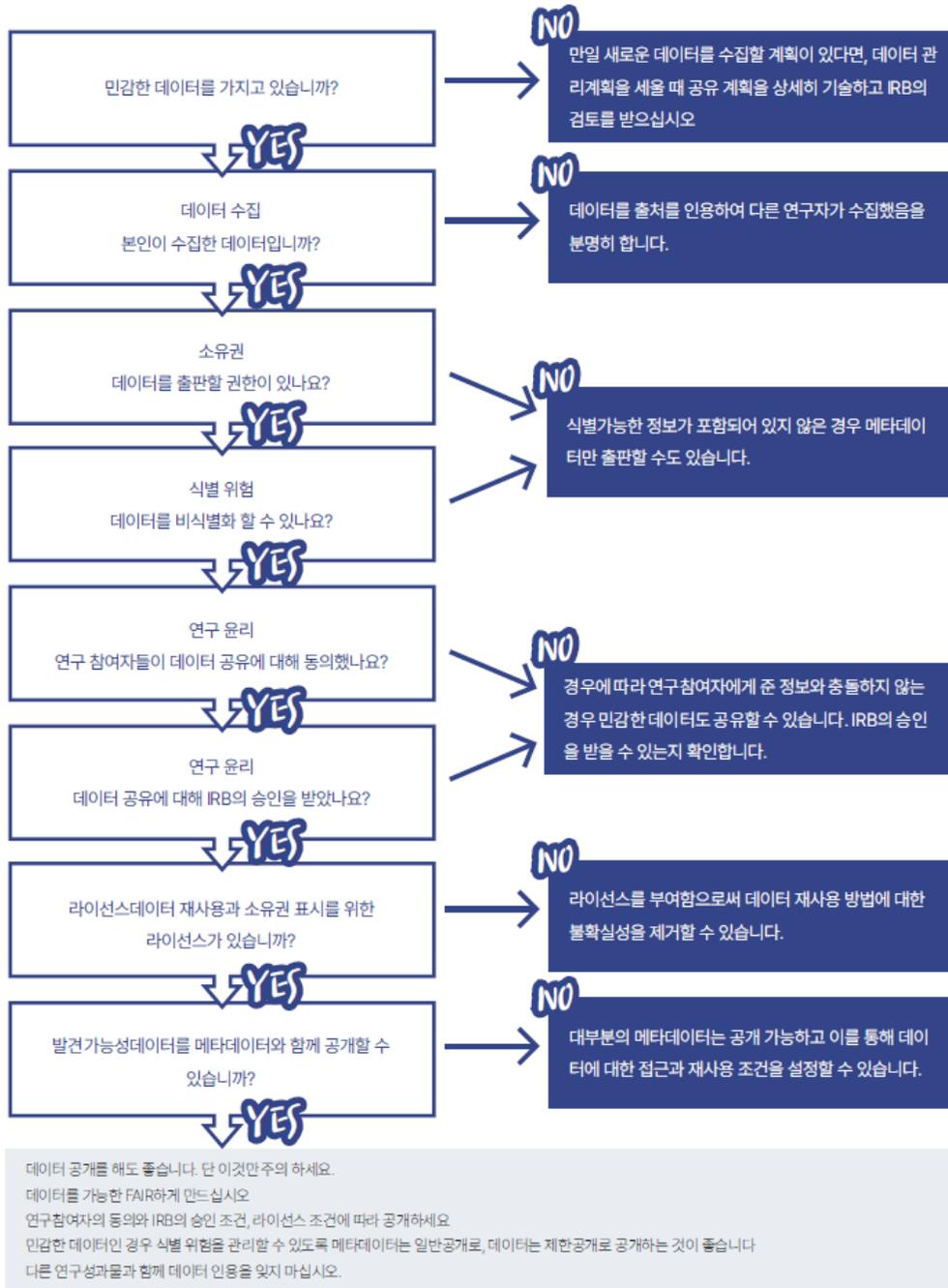
• 질적 자료 파일 유형

종류	내용	수집 파일 형식
인터뷰 자료	전사자료/녹취록 인터뷰 요약문 인터뷰 노트 (인터뷰 조사 시 작성 노트) 질문지 코딩가이드 인터뷰 스케줄	hwp doc pdf
	인터뷰 현장 녹음	wma(권장) .mp3
	인터뷰 현장 촬영 사진	jpeg(권장) .jpg, .tiff, .png, .bmp, .pdf
	인터뷰 현장 촬영 동영상	wmv, avi(권장)
관찰기록	필드노트, 현장 기록 등	hwp, doc, pdf
기록문서	공적, 민간 집단 기록 회의 또는 강연 기록 매스미디어 기록	pdf, jpg
기타	연구보고서(연구배경, 목적, 대상 방법, 사용한 개념 또는 변수에 대한 정의 등 포함) 데이터 파일 목록 사전동의서 저작권 및 라이선스 정보 자료 이용문헌 목록	hwp doc pdf

자료: 한국사회과학자료원(2023a). 자료 기탁 가이드.

데이터의 기탁 시 개인정보 및 민감정보의 처리가 완료된 경우에 한하여 제공 서비스 대상이 될 수 있다. 이때 데이터 관리 계획, IRB승인 여부, 데이터 수집의 출처, 개인식별가능성 여부 확인, 재사용 관련 소유권 표시 사항 등을 검토하게 된다. KOSSDA에서는 데이터 제공자와의 협의를 통하여 개인정보 및 민감정보의 처리를 수행하고 있다.

[그림 2-18] 민감 자료의 공유 절차



자료: 한국사회과학자료원(2023a). 자료 기탁 가이드.

[그림 2-19] 기탁 자료 준비 및 제출 체크리스트

[조사자료 공개 전 체크리스트]

✓ 필수 확인 사항

- Raw data 파일이 맞는가?
때때로 빈도표를 기탁하는 경우 있음
- Raw data 최종 버전의 파일이 맞는가?
결과보고서와 비교할 때 사례수가 차이가 있는 경우 있음
사례수 차이가 있어도 되나 공개용 최종 버전의 파일이 맞는지 확인 필요함
- 설문지와 데이터가 일치하는가?
특정 목적을 위해 일부 변수를 삭제하고 공개할 수는 있으나 때로 설문지의 일부분만 데이터로 남은 경우 있음
- 아이디 변수가 있는가? 아이디 값이 중복되지 않는가?
- 중복 사례가 있는가?
- 민감한 정보 또는 직접적인 식별자가 있는 변수를 익명처리 하였는가? (개인정보 보호 목적)
 - 이름, 전화번호, 우편번호, 주소, GPS 좌표, 민감한 정보를 다루는 데이터의 경우 소속기관명 등
 - 일부 익명처리, 범주화 또는 삭제 권장
- 가중치 변수가 있는 데이터의 경우, 가중치 변수가 포함되어 있는가?

✓ 설문지 체크리스트

- 최종 설문지 파일이 맞는가?
보고서 부록의 설문지(최종)와 기탁한 설문지가 일치하는지 확인 필요
- 설문지 내 연구진 휴대폰 번호가 명시된 경우 이를 익명처리 할 것인가?
삭제 또는 일부 익명처리 권장

✓ 권장 확인 사항

- 변수명이 일관된 원칙으로 부여되었는가?
한글로 된 변수명 지양
- 변수가 설문지와 동일한 순서로 정렬되어 있는가?
- 불필요한 변수, 보고서 작성을 위해 임시로 만든 변수 등을 삭제하였는가?
 - 연령 범주화 변수는 유용하게 사용될 수 있음
- 변수설명, 변수값 설명이 입력되어 있는가?
• 변수설명: 짧고 정확하게 함. 설문문항을 그대로 입력하는 것 지양
• 두 개의 다른 변수에 동일한 변수설명을 사용하지 않음
- 결측값이 일관되게 부여되어 있는가?
DK(실모름)과 NAP(비해당)을 구별가능하다면 별도 값 부여 권장
- 변수값 응답범주가 설문지와 일치하는가?
• 예/아니오 문항에 3, 4번 값에 빈도 있는 경우 등 확인 필요
• 주 평균 근로시간 등에 음수 값이 있는 경우 확인 필요
- 숫자 변수로 변환할 수 있는 문자 변수가 있는가?

자료: 한국사회과학자료원(2023b). 조사자료 체크리스트.

제3절 소결

본 장에서는 조사데이터와 관련된 정책동향을 검토하고 조사데이터 관리 및 활용 사례에 대한 내용들을 살펴보았다. 조사데이터의 관리와 활용에 대한 필요성은 2017년 이후 오픈사이언스 정책의 확산으로 인해 연구데이터의 공유 및 활성화 전략에 영향을 받았다고 볼 수 있으며, 이후 이와 관련된 정책적 배경들이 구체화되었다.

조사데이터의 관리 및 활용과 관련된 다양한 법적 근거들이 존재하며, 각각의 법령 및 지침 등에서는 공공기관의 정보공개를 통한 국민의 알 권리 보장 및 기관 운영의 투명성 제고, 개인정보의 보호 및 관련 정보의 활용 방안, 인간대상 연구에서의 생명윤리 보호, 통계의 정확도 및 신뢰도 제고, 연구개발정보의 처리와 체계적 관리 등 연구기관으로서의 다양한 역할들을 부여하고 있다. 이상의 정책적 배경에 따라 한국보건사회연구원은 조사데이터의 체계적 관리와 제공에 대한 의무뿐만 아니라 조사대상에 대한 보호, 신뢰도 높은 데이터의 생산 등의 의무 또한 이행해야 할 것이다.

한편, 조사데이터의 관리 및 활용 사례 분석을 통해 데이터에 대한 체계적 관리를 위하여 관련 인프라를 구축하고 지속적으로 개선해나가는 사례들을 살펴볼 수 있었다. 데이터 관리의 측면에서는 데이터의 체계적인 축적과 데이터관리 계획의 작성 및 기록이 강조되며, 데이터 활용 활성화를 위한 유연한 형태의 데이터 공유 플랫폼이 구축될 필요가 있다.

다양한 분야의 연구기관들은 데이터에 대한 관리 및 제공을 하여 관련 규정 및 지침 등을 마련하였으며, 이를 정리하면 아래 표와 같다. 조사데이터 관리 사례를 통하여 도출할 수 있는 사항은 다음과 같다. 첫째, 조사데이터의 체계적 관리를 위해서는 전담 부서 또는 전담 인력이 배치되어

야 하며, 이를 기반으로 총괄적 관리 및 업무 절차의 일원화가 요구된다. 둘째, 조사데이터의 신뢰도 제고를 위하여 데이터의 생산 과정에 대한 검토 및 결과에 대한 진단 등의 절차가 보완될 필요가 있다. 셋째, 조사데이터의 관리 및 활용과 관련된 사항을 기관 차원에서 심의할 수 있는 기구가 마련되어야 하며, 실질적인 기능을 수행할 필요가 있다. 넷째, 기관 중심의 데이터 수집 및 축적이 이루어져야 하며, 데이터 누락을 최소화할 수 있는 방안들이 지속적으로 고민되어야 한다.

〈표 4-16〉 조사데이터 관리 사례 요약 구분

구분	내용	해당 기관(비고)
데이터 관련 부서 기능	<ul style="list-style-type: none"> • 관련 통계 생산 및 관리 전담 부서의 총괄 관리 • 조사데이터 기획-생산-결과에 대한 검토 및 지원 업무 • 데이터 관련 업무 절차의 일원화 추진 	한국교육개발원 한국교육과정평가원 한국여성정책연구원 한국직업능력연구원
조사데이터 수집 관리	<ul style="list-style-type: none"> • 조사 전반에 대한 사전-사후 검토 및 의견 반영 • 데이터 민감도 등에 대한 검토 절차 마련 	한국교육개발원(IRB활용) 한국교통연구원(사전협의)
관련 위원회 구성/운영	<ul style="list-style-type: none"> • 관련 위원회 운영을 통한 데이터 관리 및 활용 사항 점검 • 데이터 수집 및 제공 등에 대한 기준 설정 및 심의 	한국여성정책연구원 한국사회과학자료원
데이터 이관 의무	<ul style="list-style-type: none"> • 조사데이터 제출에 대한 기본적 의무 사항 설정 • 데이터 수집 및 관리 주체 설정 및 수집 결과 활용 	한국교육개발원 한국교통연구원 한국형사·법무정책연구원

주: 각 기관별 관련 지침을 중심으로 작성되었으므로 실제 운영 상황과 다를 수 있음.
 자료: 저자 작성



제3장

조사데이터 관리 실태 분석: 한국보건사회연구원 조사자료 관리 사례

제1절 조사자료 관리 지침 분석

제2절 조사데이터 관리 현황 분석

제3절 소결

제3장 조사데이터 관리 실태 분석: 한국보건사회연구원 조사자료 관리 사례

제1절 조사자료 관리 지침 분석

가. 한국보건사회연구원 조사자료 관리지침²⁰⁾의 구성

한국보건사회연구원 「조사자료 관리지침(이하 지침)」은 2012년에 제정되어, 2016년, 2020년에 걸쳐 전문 개정되었다. 특히 2020년 12월 전문 개정을 통하여 현재의 구성을 갖추게 되었다.

먼저, 지침은 한국보건사회연구원 조사자료 관리의 필요사항을 규정하여 조사자료 관리의 체계화 및 이용 활성화를 목적으로 제정되었다(제1조). 제2조에서는 조사자료에 대한 정의를 내리고 있으며(2장 참고), 조사자료의 관리를 조사자료 관리계획 수립과 자료의 제출, 보관, 제공하는 일련의 과정이라 규정하였다. 또한, 조사자료 제공 시 개인정보 및 민감정보 등을 처리한 제공용 조사자료를 생성해야 함을 명시하고 있다. 지침 제3조에 따르면, 연구원에서 생성한 조사자료는 기본적으로 조사자료 관리계획을 수립해야 하며, 연구원에 제출·데이터 제공을 위한 노력을 해야 한다.

제5조와 제6조에서는 조사자료 관리위원회의 구성과 역할에 대하여 명시하고 있다. 조사자료 관리위원회는 조사자료 관리 및 심의에 대한 제반 업무를 전담하여, 구체적으로 1) 연구자의 조사자료 관리계획서 접수 및 등록 업무, 2) 연구자의 조사자료 제출에 따른 보관용 조사자료 취합

20) 한국보건사회연구원(2022a). 조사자료 관리지침.

업무, 3) 보관용 조사자료 보관을 위한 조사자료 관리시스템 운영 업무, 4) 조사자료 제출 확인 및 독려 업무, 5) 조사자료 제공을 위한 사전 준비 업무, 6) 연구원에서 운영하는 정보시스템(웹사이트, 원내업무포털)을 활용한 조사자료 제공 업무, 7) 기타 조사자료 관리에 필요한 제반 업무를 수행한다. 또한 조사자료 전반에 대한 점검 사항, 조사자료의 공개여부, 비공개 자료 요청 사항, 조사자료 관리 감동에 관한 사항 등을 심의한다.

제7조에 따르면, 연구자가 조사를 수행할 시에는 조사자료 관리계획을 사전에 수립하여야 하며, 위원회는 관리계획서에 대한 관리를 수행해야 한다. 이때 개인정보의 처리 및 동의 데이터 공개범위 설정 등에 대한 내용이 조사표에 포함되어 있는지를 검토해야 한다.

연구원에서 생산한 조사자료는 모든 조사데이터는 관리위원회의 결재를 득하여 제출되어야 한다(제10조 제3항). 연구원에서 생산된 조사자료는 기본적으로 공개가 원칙(제11조)이며, 이때 관리위원회와 연구책임자 간의 충분한 협의가 필요하다. 다만 제12조에 따라 정보공개법 제9조에 해당되는 비공개 대상 정보의 경우, 데이터 소유 주체(연구사업 발주 기관)가 데이터 공개를 원하지 않는 경우, 공개 유보기간이 종료되지 않았거나 연구가 진행중인 데이터의 경우, 관리위원회의 심의결과가 결정되지 않은 경우 등은 공개대상에서 제외될 수 있다.

조사자료의 제공 이전에 보관용 조사자료에서 제공용 조사자료로의 변환이 요구된다. 여기서 말하는 제공용 조사자료는 개인정보 및 민감정보 등이 처리된 자료를 의미하며 제공용 조사자료 작성을 위하여 연구책임자와의 협의가 이루어질 필요가 있다. 연구책임자는 제공용 조사자료 작성을 위하여 조사자료 제출 관련 양식과 함께 관련된 각종 문서(조사표, 코드북, 변수구성, 질문지, 음성파일 등)를 동봉하여 제출해야 한다.

나. 한국보건사회연구원 조사자료 관리지침의 특성

한국보건사회연구원의 조사자료 관리지침은 연구원에서 생산한 조사데이터에 대한 별도의 관리지침이라는 점에서 의미가 있다. 제2장에서 살펴본 바와 같이 연구기관에서 생산한 조사데이터에 대한 별도의 규정 및 지침을 마련한 사례는 많지 않다. 더욱이 사회보장 및 보건 영역에서의 연구 대다수가 인간대상의 연구임을 감안할 때 인간대상의 조사자료 관리를 체계적으로 추진하기 위한 지침의 필요성이 더욱 크다고 볼 수 있다.

한국보건사회연구원 조사자료 관리지침의 특성을 1) 전담 조직의 구성, 2) 정보처리 관련 사항, 3) 수집 양식, 4) 제출 절차로 나누어 살펴보았다. 먼저, 전담 조직의 구성과 관련된 사항을 살펴보면, 본 지침에서는 조사자료의 관리에 관한 사항을 별도의 관리위원회를 두어 관련 업무 및 심의를 수행한다. 한국교육개발원, 한국직업능력연구원 등에서는 조사자료와 관련된 업무를 통계 및 데이터 전담부서에 역할로 설정하고 있다. 또한, 한국여성정책연구원에서는 통계심의위원회 혹은 정보공개위원회를 통하여 조사데이터 관련 사항을 검토한다. 이상의 내용을 고려하였을 때, 연구원의 관리위원회는 조사자료 관련 사항에 대한 심의 기능을 특성화하였다는 점에서 의미가 있으나 실제 업무 수행을 위한 부서와의 연결성이 타 연구기관 사례에 비해 떨어질 수 있다.

다음으로 조사데이터의 정보 처리와 관련된 사항을 살펴보면, 연구원 지침에서는 개인정보 및 민감정보에 대한 처리 이후에 제공용 데이터 생산을 하나의 절차로 설정하고 있다. 다만, 개인정보 및 민감정보 처리 방법이나 이를 지원하기 위한 구체적인 내용은 생략되어 있다. 이 경우, 데이터의 제출 결과가 제공용 데이터로의 전환을 의미하게 되기 때문에 이

러한 과정 중에 관리위원회의 검토 또는 승인 절차가 필요하다. 현재의 지침에서는 연구책임자와 관리위원회 간 협의를 명시하고 있는데, 한국 직업능력연구원이나 한국교통연구원과 같이 관련 업무를 전담하는 부서를 특정하지 않을 경우, 데이터 제공과 관련된 협의가 충분히 추진되었는지를 확인하기 어렵다.

조사자료 수집을 위한 양식, 관리 목록, 체크리스트 등에 대한 사항은 타 연구기관에 비하여 상당히 구체적으로 제시되었다는 점에서 의미가 있다. 특히 DMP의 도입은 조사데이터 관리 및 활용에 있어서 중요한 역할을 차지한다. 연구원 지침 내 DMP는 전반적으로 연구데이터관리 가이드라인에서 제시하고 있는 DMP작성 양식을 따르고 있으나 연구종료 후 데이터의 보존 계획, 데이터의 공유방법 등에 대한 내용이 상대적으로 구체화되어있지 않다. 이는 DMP작성 시점에 데이터 보존과 공유 등에 대한 사항이 결정되기 어려운 현실을 반영한 것이라 볼 수 있으나 해당 내용의 비구체성으로 인하여 데이터 제출 이후에 DMP의 활용도가 떨어질 수 있다. 따라서 DMP의 내용과 데이터 제출 결과와의 비교, 제공용 데이터 작성시 DMP의 활용 방안 등이 검토될 필요가 있다.

마지막으로 조사자료 제출 관련 사항을 살펴보면, 조사자료 제출 시 데이터의 공개에 대한 가부를 결정하게 되는데, 이때 데이터의 공개수준을 상향하기 위한 데이터의 처리(익명화, 범주화 등)가 별도로 기재되어 있지 않다. 즉, 조사데이터의 개인정보 및 민감정보에 대한 처리와 민감정보에 대한 관리 사항 등이 명확하게 제시될 수 없는 한계를 가진다. 그러므로 조사자료 제출 시 데이터 제공 범위의 설정과 데이터 처리 방법에 대한 설정, 정보별 보관 및 보호 조치 등이 세분화될 필요가 있다.

제2절 조사데이터 관리 현황 분석

한국보건사회연구원 조사자료 관리 현황을 분석하기 위하여 “조사자료 관리목록(「조사자료 관리지침」 별지 제3호 서식)” 내 정보를 활용하였다. 해당 데이터는 2016년부터 2022년까지 수집된 조사데이터로, 2020년 조사자료 관리지침 전문개정 이후(2020.12.30.) 2016년(조사연도 기준)부터 2022년²¹⁾까지 생산된 조사데이터에 대한 내용이 포함되어 있다.

이하의 분석에 포함된 조사데이터는 총 621건이며, 연구과제의 구분이나 조사데이터의 종류 등과 관계없이 조사자료 관리목록에 기재된 내용을 중심으로 분석이 이루어졌다. 이를 통하여 2016년 이후 생산된 조사데이터의 특성을 파악하고, 조사데이터의 관리 체계와 관련된 문제점 등을 도출하였다.

1. 연도별 조사데이터 관리 현황

가. 연도별 조사데이터 생산 현황

연도별, 과제구분별 조사데이터 현황을 살펴보면, 2016년 79건이 조사데이터가 생산되었으며, 생산데이터의 수가 점차 증가하면서 2020년에는 116개의 조사데이터가 생성되었다. 이후 2022년에는 74건의 조사데이터가 생성된 것으로 보고되고 있다.

2016년부터 2022년까지 수탁과제의 비중이 상대적으로 높은 것을 알 수 있다. 수탁과제를 제외하였을 때에는 일반, 기본, 연개금 사업 순으로

21) 2023년에 조사된 1건의 조사데이터(“공립 노인전문교육원 건립 관련 노인교육 수요조사”)는 이하의 분석에서 2022년에 포함하였다.

조사데이터의 비중이 높은 것으로 나타났다.

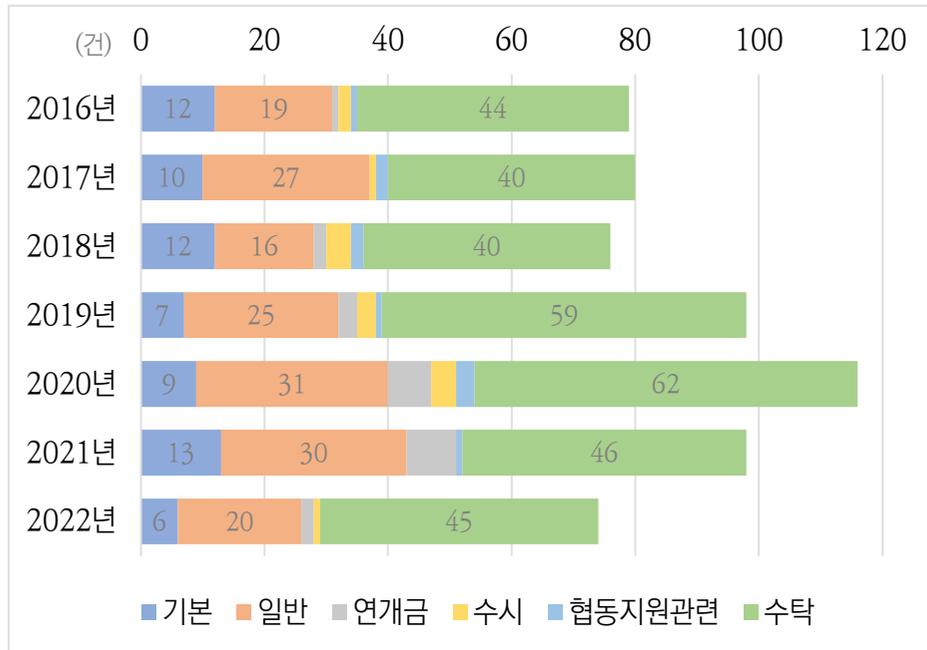
〈표 3-1〉 연도별, 과제구분별 조사데이터 현황

단위: 건(%)

연도	기본	일반	연개금	수시	협동/지원/ 관련	수탁	소계
2016년	12(15.19)	19(24.05)	1(1.27)	2(2.53)	1(1.27)	44(55.7)	79(100)
2017년	10(12.5)	27(33.75)	0(0.00)	1(1.25)	2(2.5)	40(50)	80(100)
2018년	12(15.79)	16(21.05)	2(2.63)	4(5.26)	2(2.63)	40(52.63)	76(100)
2019년	7(7.14)	25(25.51)	3(3.06)	3(3.06)	1(1.02)	59(60.2)	98(100)
2020년	9(7.76)	31(26.72)	7(6.03)	4(3.45)	3(2.59)	62(53.45)	116(100)
2021년	13(13.27)	30(30.61)	8(8.16)	0(0.00)	1(1.02)	46(46.94)	98(100)
2022년	6(8.11)	20(27.03)	2(2.7)	1(1.35)	0(0.00)	45(60.81)	74(100)
합계	69(11.11)	168(27.05)	23(3.7)	15(2.42)	10(1.61)	336(54.11)	621(100)

주: 각 셀의 백분율은 행별 합계를 기준으로 작성됨
 자료: 조사자료 관리목록(한국보건사회연구원, n.d.)을 바탕으로 저자 작성

[그림 3-1] 연도별, 과제구분별 조사데이터 현황



자료: 조사자료 관리목록(한국보건사회연구원, n.d.)을 바탕으로 저자 작성

연도별 조사데이터의 특성을 확인하고자 연도별로 조사명 내 단어 빈도를 분석하였으며, 그 결과는 아래와 같다. 2016년부터 2022년까지 조사명에 가장 빈번하게 활용된 단어는 “조사”, “실태”, “사회”, “인식”이었다. 즉, 여러 사회현상에 대한 실태 분석과 국민 또는 연구대상의 인식 등을 조사하기 위하여 조사데이터가 생성되었음을 알 수 있다. 한편, 2020년에는 감염병 이슈가 불어져 “코로나”라는 키워드의 빈도가 상대적으로 높게 나타났다.

이 외에도 “복지”, “보건”, “의료”, “정책” 등 주요 연구 분야에 대한 키워드, “서비스”, “고령·노인”, “아동”, “돌봄” 등 주요 이슈와 관련된 사항들이 조사데이터 생산의 사유가 된 것으로 파악되었다.

102 보건복지 조사데이터의 체계적 관리 및 활용 방안 연구

〈표 3-2〉 연도별 조사명 내 단어 빈도(2016년-2019년)

순위	2016년		2017년		2018년		2019년	
	단어	빈도(%)	단어	빈도(%)	단어	빈도(%)	단어	빈도(%)
1	조사	91(34.60)	조사	87(29.90)	조사	72(29.03)	조사	100(30.77)
2	실태	16(6.08)	사회	18(6.19)	사회	20(8.06)	실태	33(10.15)
3	인식	15(5.70)	실태	16(5.50)	인식	18(7.26)	사회	24(7.38)
4	전문가	14(5.32)	서비스	15(5.15)	실태	14(5.65)	인식	24(7.38)
5	복지	12(4.56)	정책	14(4.81)	서비스	12(4.84)	복지	18(5.54)
6	사회	12(4.56)	복지	13(4.47)	건강	10(4.03)	전문가	15(4.62)
7	관련	11(4.18)	의료	13(4.47)	의료	10(4.03)	서비스	13(4.00)
8	의료	10(3.80)	인식	13(4.47)	경험	8(3.23)	아동	10(3.08)
9	건강	9(3.42)	전문가	13(4.47)	고령	8(3.23)	정책	9(2.77)
10	관리	9(3.42)	사업	12(4.12)	관련	8(3.23)	관련	8(2.46)
11	평가	8(3.04)	노인	11(3.78)	출산	8(3.23)	국민	8(2.46)
12	현황	8(3.04)	지역	10(3.44)	현황	8(3.23)	방안	8(2.46)
13	국민	7(2.66)	관련	9(3.09)	국민	7(2.82)	사업	8(2.46)
14	노인	7(2.66)	개발	8(2.75)	보건	7(2.82)	의견	8(2.46)
15	설문	7(2.66)	청소년	8(2.75)	전문가	7(2.82)	출산	8(2.46)
16	정책	7(2.66)	건강	7(2.41)	지역	7(2.82)	인력	7(2.15)
17	면접	5(1.90)	가족	6(2.06)	노인	6(2.42)	건강	6(1.85)
18	방안	5(1.90)	경험	6(2.06)	대상	6(2.42)	경험	6(1.85)
19	보건	5(1.90)	설문	6(2.06)	방안	6(2.42)	보건	6(1.85)
20	사업	5(1.90)	시범	6(2.06)	복지	6(2.42)	연구	6(1.85)

주: 각 셀의 백분율은 각 연도의 열별 합계를 기준으로 작성됨
 자료: 조사자료 관리목록(한국보건사회연구원, n.d.)을 바탕으로 저자 작성

〈표 3-3〉 연도별 조사명 내 단어 빈도(2020년-2021년)

순위	2020년		2021년		2022년	
	단어	빈도(%)	단어	빈도(%)	단어	빈도(%)
1	조사	106(31.18)	조사	95(32.76)	조사	72(31.03)
2	실태	24(7.06)	실태	26(8.97)	사회	17(7.33)
3	인식	21(6.18)	인식	18(6.21)	인식	17(7.33)
4	사회	20(5.88)	의료	16(5.52)	실태	16(6.90)
5	국민	14(4.12)	사회	15(5.17)	정책	12(5.17)
6	코로나	14(4.12)	전문가	12(4.14)	서비스	11(4.74)
7	연구	13(3.82)	현황	11(3.79)	경험	8(3.45)
8	현황	12(3.53)	설문	10(3.45)	사업	8(3.45)
9	건강	11(3.24)	대상	9(3.10)	지원	8(3.45)
10	노인	11(3.24)	지원	9(3.10)	건강	7(3.02)
11	복지	11(3.24)	돌봄	8(2.76)	방안	7(3.02)
12	지원	11(3.24)	면접	8(2.76)	복지	7(3.02)
13	서비스	10(2.94)	복지	8(2.76)	의료	7(3.02)
14	전문가	10(2.94)	서비스	7(2.41)	관련	5(2.16)
15	정책	10(2.94)	심층	7(2.41)	설문	5(2.16)
16	대응	9(2.65)	한국	7(2.41)	의견	5(2.16)
17	요구	9(2.65)	건강	6(2.07)	전문가	5(2.16)
18	교육	8(2.35)	보건	6(2.07)	지역	5(2.16)
19	보건	8(2.35)	의견	6(2.07)	청년	5(2.16)
20	설문	8(2.35)	이용	6(2.07)	활동	5(2.16)

주: 각 셀의 백분율은 각 연도의 열별 합계를 기준으로 작성됨
 자료: 조사자료 관리목록(한국보건사회연구원, n.d.)을 바탕으로 저자 작성

[그림 3-2] 연도별 조사명 워드클라우드 분석 결과



주: 단어 가운데 "조사"를 제외한 빈도 분석 결과 활용
 자료: 조사자료 관리목록(한국보건사회연구원, n.d.)을 바탕으로 저자 작성

다음으로 2016년부터 2022년까지 조사명 내 단어의 관계성을 확인하고자 키워드 연결망 분석을 수행하였으며, 이에 Gephi 0.10.1 프로그램 (Bastian, M., S. Heymann, and M. Jacomy., 2009)을 활용하여 시각화하였다. 각 키워드는 방향성이 없는 연결의 빈도를 중심으로 시각화되었으며, 이때 20이상의 연결정도(degree)를 가진 결과에 한하여 제시하였다. 해당 결과의 레이아웃으로는 Force Atlas를 사용하였다(Cherven, 2015).

단순 키워드 빈도 분석 결과와 유사하게 “조사”, “실태”, “사회” 등의 키워드가 빈번하게 나타났다. “조사”는 “노인”, “사회”, “코로나”, “지역” 등과 같은 키워드와 강한 연결관계를 가지고 있으며, “인식”과 “실태”는 “노인”, “아동”, “청년” 등의 정책 대상과 “코로나”, “건강” 등의 주요 이슈와의 연결관계를 가지는 것으로 나타났다. 그 외에도 주요 정책 대상은 각 대상별 주요 현안과의 강한 연결성을 가지고 있었다. 예를 들어 “노인”과 “아동”의 경우에는 “돌봄”, “지역” 등의 키워드에 대한 연결성이 강하게 나타난다는 공통점을 가지지만 “노인”은 “장기요양”, “치매” 등과 “아동”은 “출산”, “보호” 등과의 연결이 빈번하게 나타났다.

나. 연도별 조사데이터의 관리

연도별 데이터 공개 현황을 살펴보면, 약 13%의 조사자료가 조사자료 관리 대장 내에 예산 정보가 기록되지 않거나 부정확한 것으로 나타났다(비예산 조사와 구분). 조사예산 정보 확인불가 사례는 점차 감소하는 양상을 보이며, 1천만원에서 5천만원에 해당되는 조사의 범위가 점차 증가하는 양상을 보였다.

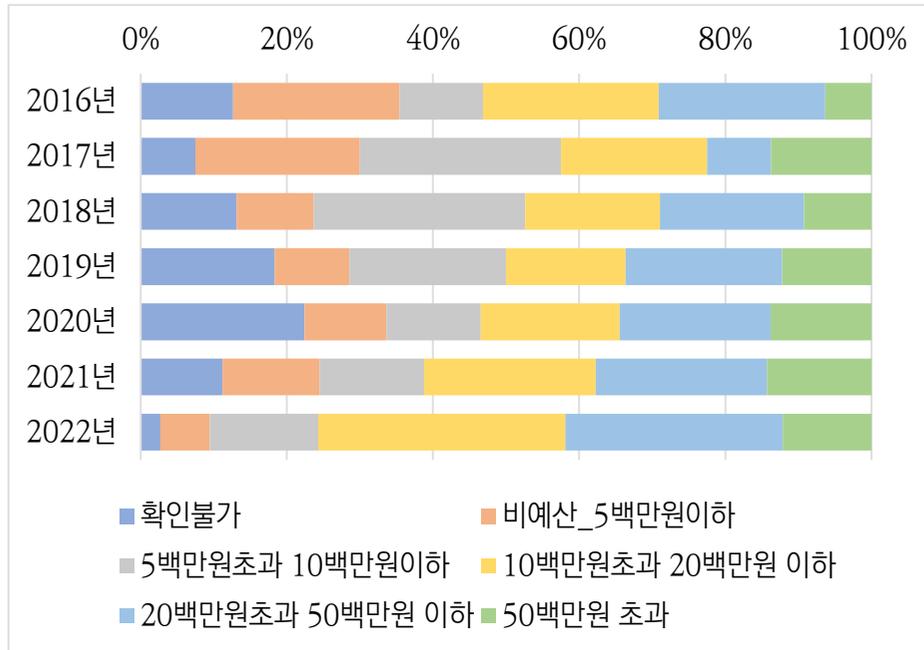
〈표 3-4〉 연도별 조사예산 현황

단위: 건(%)

연도	확인 불가	비예산 5백만원 이하	5백만원 초과 10백만원 이하	10백만원 초과 20백만원 이하	20백만원 초과 50백만원 이하	50백만원 초과
2016년	10(12.66)	18(22.78)	9(11.39)	19(24.05)	18(22.78)	5(6.33)
2017년	6(7.50)	18(22.50)	22(27.50)	16(20.00)	7(8.75)	11(13.75)
2018년	10(13.16)	8(10.53)	22(28.95)	14(18.42)	15(19.74)	7(9.21)
2019년	18(18.37)	10(10.2)	21(21.43)	16(16.33)	21(21.43)	12(12.24)
2020년	26(22.41)	13(11.21)	15(12.93)	22(18.97)	24(20.69)	16(13.79)
2021년	11(11.22)	13(13.27)	14(14.29)	23(23.47)	23(23.47)	14(14.29)
2022년	2(2.70)	5(6.76)	11(14.86)	25(33.78)	22(29.73)	9(12.16)
합계	83(13.37)	85(13.69)	114(18.36)	135(21.74)	130(20.93)	74(11.92)

주: 각 셀의 백분율은 행별 합계를 기준으로 작성됨
 자료: 조사자료 관리목록(한국보건사회연구원, n.d.)을 바탕으로 저자 작성

[그림 3-4] 연도별 조사예산 현황



자료: 조사자료 관리목록(한국보건사회연구원, n.d.)을 바탕으로 저자 작성

연도별 조사데이터의 유형을 살펴보면 약 60% 이상이 설문조사 등과 같은 양적 데이터인 것으로 나타났으며, 약 26%가 심층인터뷰, FGI 등 질적 연구데이터인 것으로 나타났다. 이 가운데 하나의 조사데이터에서 양적·질적 조사결과를 함께 제출한 경우(중복 기입), 데이터의 형태를 확인할 수 없거나 관련 정보가 불충분한 경우(정보 오류)는 전체 조사데이터 중 약 12%를 차지하는 것으로 나타났다.

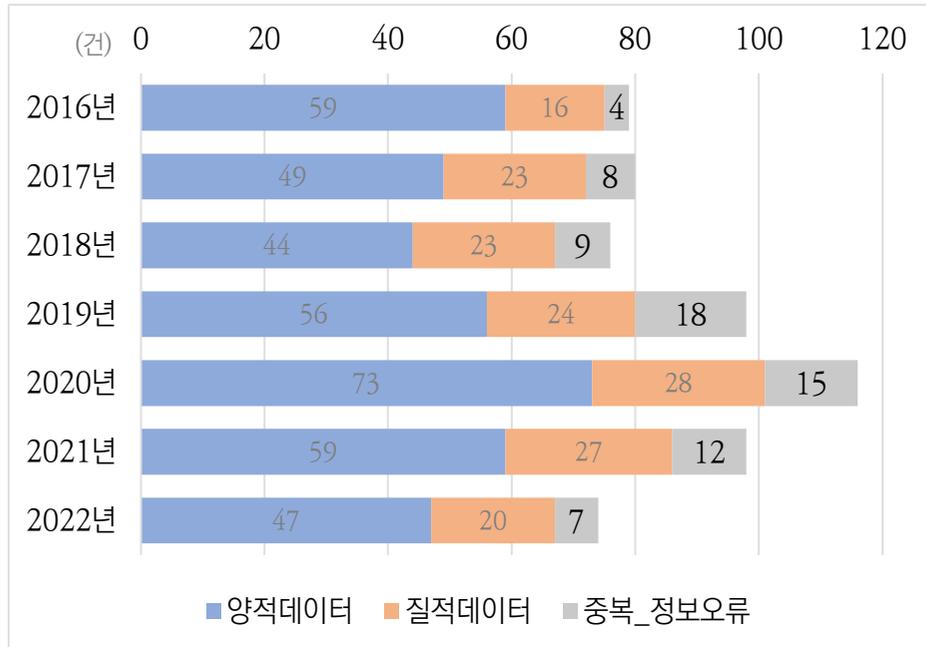
〈표 3-5〉 연도별 조사데이터 유형 구분

단위: 건(%)

연도	양적 데이터	질적 데이터	중복 기입 및 정보 오류
2016년	59(74.68)	16(20.25)	4(5.06)
2017년	49(61.25)	23(28.75)	8(10.00)
2018년	44(57.89)	23(30.26)	9(11.84)
2019년	56(57.14)	24(24.49)	18(18.37)
2020년	73(62.93)	28(24.14)	15(12.93)
2021년	59(60.2)	27(27.55)	12(12.24)
2022년	47(63.51)	20(27.03)	7(9.46)
합계	387(62.32)	161(25.93)	73(11.76)

주: 각 셀의 백분율은 행별 합계를 기준으로 작성됨
 자료: 조사자료 관리목록(한국보건사회연구원, n.d.)을 바탕으로 저자 작성

[그림 3-5] 연도별 조사데이터 유형 구분



자료: 조사자료 관리목록(한국보건사회연구원, n.d.)을 바탕으로 저자 작성

연도별 조사데이터의 조사방법은 아래 표와 같다. 조사자료 수집 시 복수의 조사방법을 활용하는 경우(예시. 설문 우편발송과 메일 발송 동시 진행, FGI수행과 개별 인터뷰 동시 진행 등)를 확인할 필요가 있다. 특히 연구원 지침상 조사데이터의 조사방법은 주관식으로 응답되어 조사방법의 빈도를 확인하기 위해서는 연구책임자가 작성한 내용을 바탕으로 조사 방법을 유추²²⁾해야 한다.

양적 데이터의 경우, 웹 및 모바일 기반의 온라인 조사가 가장 많은 비율을 차지하였으며 그 빈도가 점차 증가하는 양상을 보였다. 그 다음으로 조사원이 직접 방문하는 면접조사, 표집틀에 기반하여 전화로 관련 내용을 조사하는 전화조사 순으로 높게 나타났다. 한편, 질적 데이터의 경우에는 초점집단인터뷰(FGI)가 가장 높은 비율을 차지하는 것으로 나타났다.

22) 예를 들어 전문가 대상의 질적 조사를 수행하는 경우, 메일로 해당 양식을 발송할 때 “이메일 조사”로 기입하게 되어 메일을 통한 설문조사 수행과 중복되는 문제가 발생함.

〈표 3-6〉 연도별 조사데이터 조사방법 세분화(중복응답)

단위: 건(%)

연도	양적데이터					질적데이터		
	온라인 (웹기반)	설문지 발송 (이메일, 팩스)	설문지 발송 (우편)	면접	전화조사	(심층) 인터뷰	FGI	전문가 조사
2016년	20(21.7)	15(16.3)	4(4.3)	16(17.4)	17(18.5)	4(4.3)	7(7.6)	9(9.8)
2017년	13(14.4)	10(11.1)	3(3.3)	25(27.8)	11(12.2)	10(11.1)	7(7.8)	11(12.2)
2018년	13(16.5)	6(7.6)	1(1.3)	20(25.3)	13(16.5)	11(13.9)	8(10.1)	7(8.9)
2019년	21(21.0)	12(12.0)	1(1.0)	27(27.0)	11(11.0)	6(6.0)	12(12.0)	10(10.0)
2020년	46(37.4)	10(8.1)	3(2.4)	21(17.1)	7(5.7)	10(8.1)	21(17.1)	5(4.1)
2021년	48(41.7)	3(2.6)	1(0.9)	20(17.4)	3(2.6)	14(12.2)	20(17.4)	6(5.2)
2022년	41(46.1)	3(3.4)	2(2.2)	8(9.0)	7(7.9)	6(6.7)	15(16.9)	7(7.9)
합계	202(29.4)	59(8.6)	15(2.2)	137(19.9)	69(10.0)	61(8.9)	90(13.1)	55(8.0)

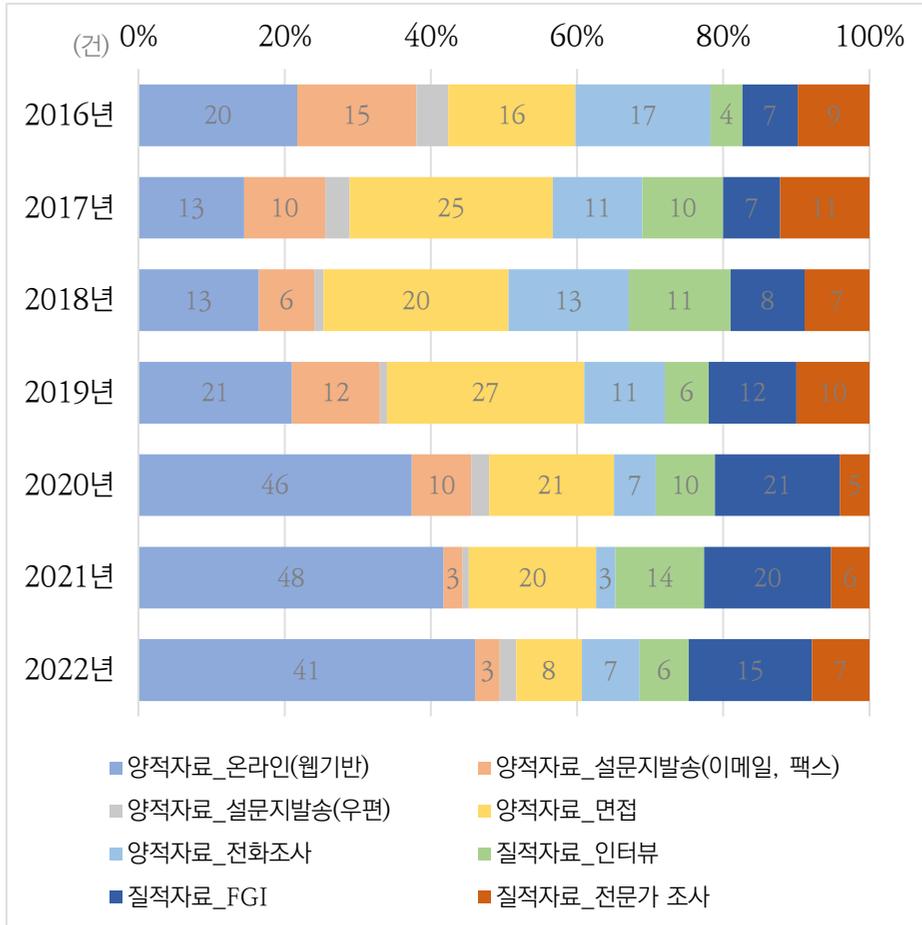
주1: 각 셀의 백분율은 행별 합계를 기준으로 작성됨

주2: 표의 가시성을 위하여 소수점 첫째 자리 표기

자료: 조사자료 관리목록(한국보건사회연구원, n.d.)을 바탕으로 저자 작성

112 보건복지 조사데이터의 체계적 관리 및 활용 방안 연구

[그림 3-6] 연도별 조사데이터 조사방법 세분화



자료: 조사자료 관리목록(한국보건사회연구원, n.d.)을 바탕으로 저자 작성

나. 과제구분별 조사데이터 관리 현황

과제구분별 조사예산을 살펴보면, 일반과제의 경우 대규모의 조사예산이 투입되는 경우가 상대적으로 많았다. 반면 기본과제나 수시과제의 경우에는 상대적으로 조사예산 규모가 작은 것으로 나타났다.

조사예산 규모는 조사대상의 규모와 관계가 있다. 즉, 조사예산이 큰 경우에는 표집의 규모가 크기 때문에 소규모 비확률 표집을 통하여 조사한 데이터보다 공개 가능성이 높다. 일반과제의 공개비율이 상대적으로 높은 것이 이러한 사유에서 기인한 것이라 추정해볼 수 있다.

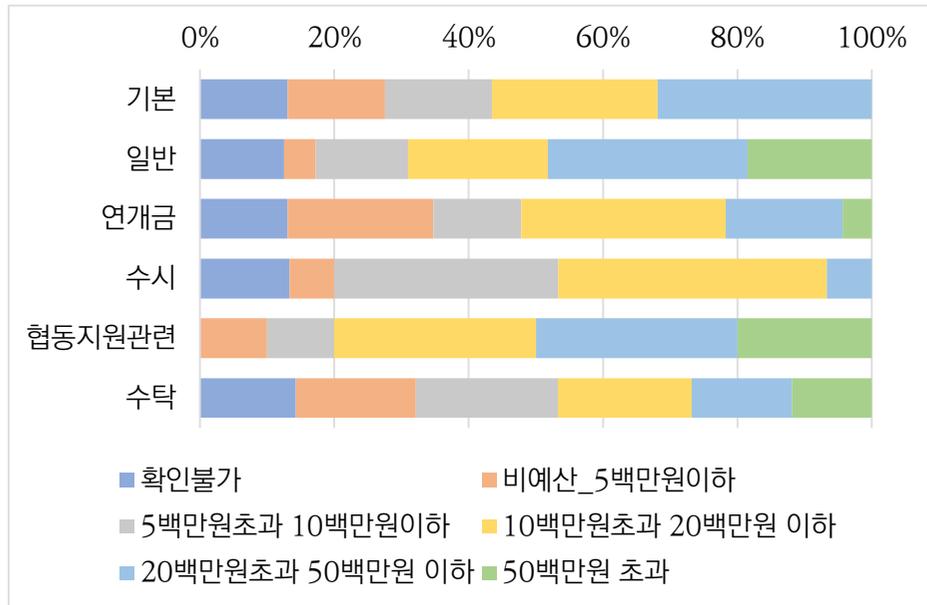
〈표 3-7〉 과제구분별 조사예산 현황

단위: 건(%)

과제 구분	확인 불가	비예산 5백만원 이하	5백만원 초과 10백만원 이하	10백만원 초과 20백만원 이하	20백만원 초과 50백만원 이하	50백만원 초과
기본	9(13.04)	10(14.49)	11(15.94)	17(24.64)	22(31.88)	0(0.00)
일반	21(12.50)	8(4.76)	23(13.69)	35(20.83)	50(29.76)	31(18.45)
연개금	3(13.04)	5(21.74)	3(13.04)	7(30.43)	4(17.39)	1(4.35)
수시	2(13.33)	1(6.67)	5(33.33)	6(40)	1(6.67)	0(0.00)
협동/지원 /관련	0(0.00)	1(10.00)	1(10.00)	3(30.00)	3(30.00)	2(20.00)
수탁	48(14.29)	60(17.86)	71(21.13)	67(19.94)	50(14.88)	40(11.90)
소계	83(13.37)	85(13.69)	114(18.36)	135(21.74)	130(20.93)	74(11.92)

주: 각 셀의 백분율은 행별 합계를 기준으로 작성됨
 자료: 조사자료 관리목록(한국보건사회연구원, n.d.)을 바탕으로 저자 작성

[그림 3-7] 과제구분별 조사예산 현황



자료: 조사자료 관리목록을 바탕으로 저자 작성

조사데이터 유형을 구분하여 살펴보면, 일반과제와 연개금 과제, 수탁 과제의 양적 데이터 수집 비율이 높은 것을 알 수 있다. 일반과제와 연개금 과제의 경우에는 데이터의 소유권이 연구원에 있는 경우가 다수이기 때문에 수탁과제와는 다르게 조사데이터의 제공 가능성이 높다고 볼 수 있다. 반면 기본과제는 상대적으로 질적 데이터가 많고 데이터 유형에 대한 중복 및 오기입 사례가 많다는 특성을 가진다.

연구원에서는 질적 데이터에 대한 개인정보 및 민감정보 처리가 세부적으로 추진되고 있지 않기 때문에 질적 데이터는 제공 대상에서 제외되고 있다. 따라서 기본, 일반, 연개금 과제를 통하여 수집된 양적 데이터가 데이터 제공 대상에 포함될 가능성이 높다.

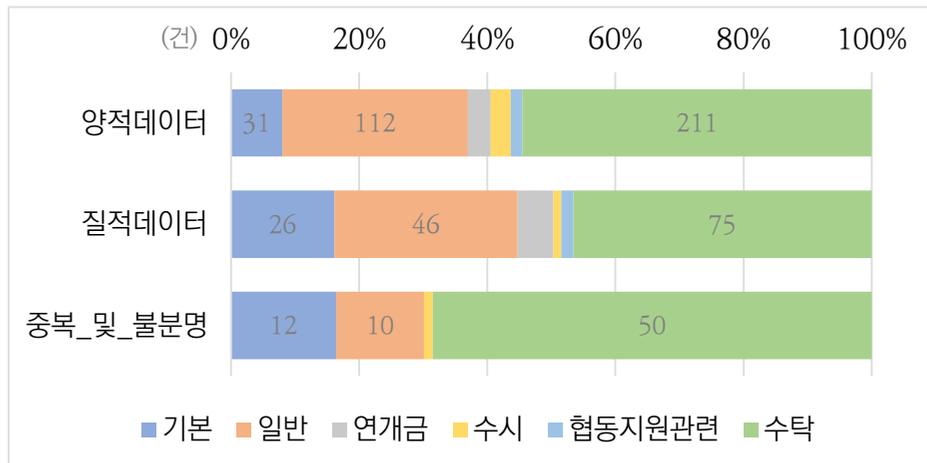
〈표 3-8〉 과제유형별 조사데이터 유형 구분

단위: 건(%)

과제 구분	양적 데이터	질적 데이터	중복 기입 및 정보 오류
기본	31(44.93)	26(37.68)	12(17.39)
일반	112(66.67)	46(27.38)	10(5.95)
연개금	14(60.87)	9(39.13)	0(0.00)
수시	12(80.00)	2(13.33)	1(6.67)
협동/지원/관련	7(70.00)	3(30.00)	0(0.00)
수탁	211(62.8)	75(22.32)	50(14.88)
소계	387(62.32)	161(25.93)	73(11.76)

주: 각 셀의 백분율은 행별 합계를 기준으로 작성됨
 자료: 조사자료 관리목록(한국보건사회연구원, n.d.)을 바탕으로 저자 작성

[그림 3-8] 과제구분별 조사데이터 유형 구분



자료: 조사자료 관리목록(한국보건사회연구원, n.d.)을 바탕으로 저자 작성

과제유형별 조사데이터 조사방법을 살펴보면, 온라인 웹기반 양적 조사의 비율이 높게 나타났다. 수탁과제는 웹기반 설문 및 질적 조사의 비

율이 상대적으로 높은 것을 알 수 있다. 일반과제와 연개금과제는 조사원 방문을 통한 면접조사 비율이 상대적으로 높게 나타났다.

〈표 3-9〉 과제유형별 조사데이터 조사방법 세분화(중복응답)

단위: 건(%)

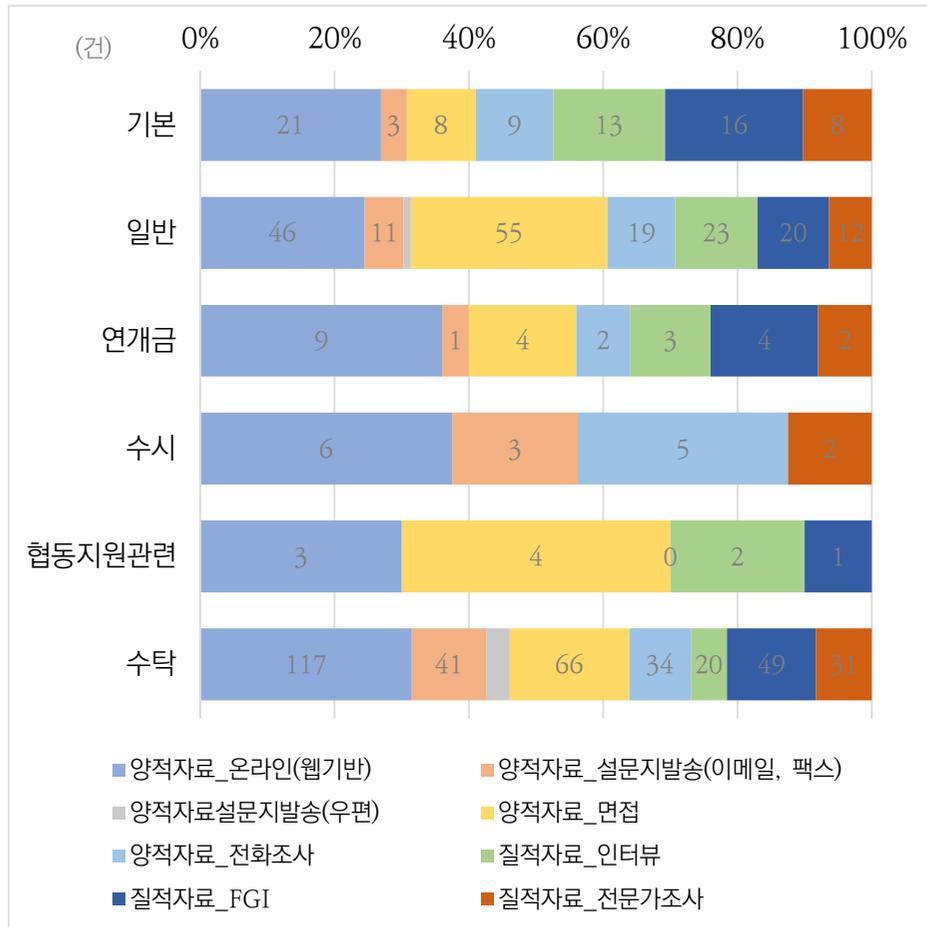
과제 구분	양적데이터					질적데이터		
	온라인 (웹기반)	설문지 발송 (이메일, 팩스)	설문지 발송 (우편)	면접	전화조사	(심층) 인터뷰	FGI	전문가 조사
기본	20(21.7)	15(16.3)	4(4.3)	16(17.4)	17(18.5)	4(4.3)	7(7.6)	9(9.8)
일반	13(14.4)	10(11.1)	3(3.3)	25(27.8)	11(12.2)	10(11.1)	7(7.8)	11(12.2)
연개금	13(16.5)	6(7.6)	1(1.3)	20(25.3)	13(16.5)	11(13.9)	8(10.1)	7(8.9)
수시	21(21.0)	12(12.0)	1(1.0)	27(27.0)	11(11.0)	6(6.0)	12(12.0)	10(10.0)
협동/지 원/관련	46(37.4)	10(8.1)	3(2.4)	21(17.1)	7(5.7)	10(8.1)	21(17.1)	5(4.1)
수탁	48(41.7)	3(2.6)	1(0.9)	20(17.4)	3(2.6)	14(12.2)	20(17.4)	6(5.2)
소계	202(29.4)	59(8.6)	15(2.2)	137(19.9)	69(10)	61(8.9)	90(13.1)	55(8.0)

주1: 각 셀의 백분율은 행별 합계를 기준으로 작성됨

주2: 표의 가시성을 위하여 소수점 첫째 자리 표기

자료: 조사자료 관리목록(한국보건사회연구원, n.d.)을 바탕으로 저자 작성

[그림 3-9] 과제구분별 조사데이터 조사방법 세분화



자료: 조사자료 관리목록(한국보건사회연구원, n.d.)을 바탕으로 저자 작성

과제구분별 조사정보 현황을 살펴보면, 기본과제와 수시과제, 수탁과제 중 조사대상 및 조사방법 정보가 불명확한 사례가 상대적으로 많은 것으로 나타났다.

또한 IRB심의 현황을 살펴보면, 수탁과제의 미심의 비율은 약 72%인 것으로 나타났으며, 그 다음으로 기본과제, 수시과제, 연개금 과제 순으로 IRB미심의율이 높은 것으로 나타났다. 한편, 조사자료의 이관율은 연개금 과제, 수탁과제 순으로 낮게 나타났다. 수탁과제에 비하여 연개금 과제가 미이관율과 공개불가 비율 간의 차이가 상대적으로 큰 것을 미루

어볼 때 연개금 과제의 데이터 이관 현황을 검토해볼 필요가 있음을 유추해볼 수 있다.

〈표 3-10〉 과제구분별 조사정보 현황

단위: 건(%)

과제 구분	조사명 정보		조사대상 정보		조사방법 정보	
	확인	확인 불가	확인	확인 불가	확인	확인 불가
기본	68(98.55)	1(1.45)	59(85.51)	10(14.49)	62(89.86)	7(10.14)
일반	167(99.40)	1(0.60)	159(94.64)	9(5.36)	161(95.83)	7(4.17)
연개금	23(100.00)	0(0.00)	23(100.00)	0(0.00)	23(100.00)	0(0.00)
수시	15(100.00)	0(0.00)	13(86.67)	2(13.33)	13(86.67)	2(13.33)
협동/지원/관련	10(100.00)	0(0.00)	10(100.00)	0(0.00)	10(100.00)	0(0.00)
수탁	330(98.21)	6(1.79)	295(87.80)	41(12.20)	299(88.99)	37(11.01)
소계	613(98.71)	8(1.29)	559(90.02)	62(9.98)	568(91.47)	53(8.53)

주: 각 셀의 백분율은 각 정보의 행별 합계를 기준으로 작성됨
 자료: 조사자료 관리목록(한국보건사회연구원, n.d.)을 바탕으로 저자 작성

〈표 3-11〉 과제구분별 IRB심의 및 자료 이관 여부 현황

단위: 건(%)

단위 건(%)	IRB심의 여부		조사자료 이관 여부	
	미심의	심의 완료	미이관 제출불가	이관 완료
기본	47(68.12)	22(31.88)	17(24.64)	52(75.36)
일반	97(57.74)	71(42.26)	43(25.60)	125(74.40)
연개금	15(65.22)	8(34.78)	10(43.48)	13(56.52)
수시	10(66.67)	5(33.33)	5(33.33)	10(66.67)
협동/지 원/관련	2(20.00)	8(80.00)	1(10.00)	9(90.00)
수탁	242(72.02)	94(27.98)	142(42.26)	194(57.74)
소계	413(66.51)	208(33.49)	218(35.10)	403(64.90)

주: 각 셀의 백분율은 각 정보의 행별 합계를 기준으로 작성됨
 자료: 조사자료 관리목록(한국보건사회연구원, n.d.)을 바탕으로 저자 작성

다. 비공개 조사데이터 특성

연도별 비공개 처리된 조사데이터의 비공개 사유를 정리하면 아래 표와 같다. 비공개 사유는 데이터 제출 시 연구책임자의 판단에 따라 작성하도록 되어 있으며, 「정보공개법」 제9조의 비공개 대상을 기준으로 기재하도록 되어있다. 데이터의 비공개 사유를 지정할 경우에 복수의 사유를 기입하는 경우가 적지 않기 때문에 해당 내용은 중복응답의 문항의 특성을 가진다. 또한, 제9조에 해당되지 않는 사유로 인하여 비공개 처리하는 경우(예시. ‘기타’ 선택 후 주관식 응답)를 정리하기 위하여 빈도가 높은 주관식 응답을 범주화할 필요가 있다.

과제구분별 비공개 처리 사유를 살펴보면, 수탁과제의 경우에는 발주

처의 비동의로 인한 비공개 비율이 가장 높게 나타났다. 그 외의 과제유형에서 나타나는 주된 비공개 사유는 질적 데이터인 경우, 일반화의 한계를 가지는 경우 등이 있었다.

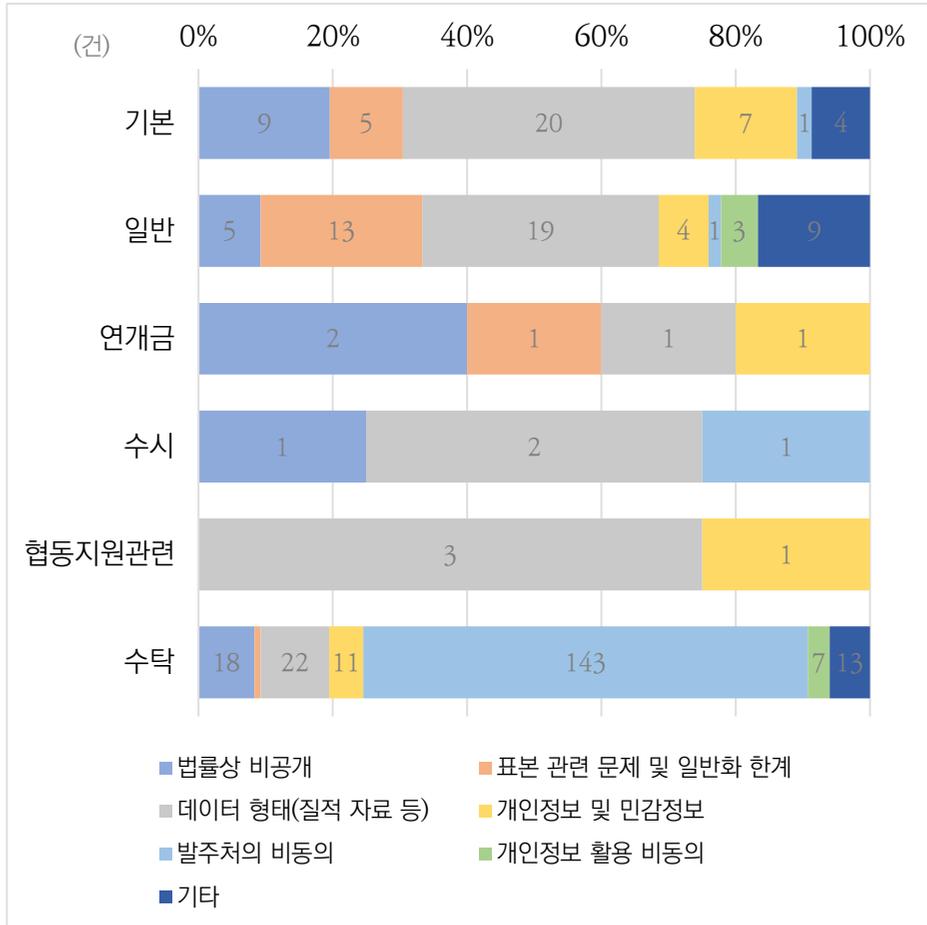
〈표 3-12〉 과제구분별 비공개 처리 사유 세분화(중복응답)

단위: 건(%)

과제 구분	법률상 비공개	표본 관련 문제 및 일반화 한계	데이터 형태(질적 자료 등)	개인정보 및 민감정보	발주처의 비동의	개인정보 활용 비동의	기타
기본	9(19.6)	5(10.9)	20(43.5)	7(15.2)	1(2.2)	0(0.0)	4(8.7)
일반	5(9.3)	13(24.1)	19(35.2)	4(7.4)	1(1.9)	3(5.6)	9(16.7)
연개금	2(40.0)	1(20.0)	1(20.0)	1(20.0)	0(0.0)	0(0.0)	0(0.0)
수시	1(25.0)	0(0.0)	2(50.0)	0(0.0)	1(25.0)	0(0.0)	0(0.0)
협동/지원/관련	0(0.0)	0(0.0)	3(75.0)	1(25.0)	0(0.0)	0(0.0)	0(0.0)
수탁	18(8.3)	2(0.9)	22(10.2)	11(5.1)	143(66.2)	7(3.2)	13(6)
소계	35(10.6)	21(6.4)	67(20.4)	24(7.3)	146(44.4)	10(3)	26(7.9)

주1: 각 셀의 백분율은 행별 합계를 기준으로 작성됨
 주2: 표의 가시성을 위하여 소수점 첫째 자리 표기
 자료: 조사자료 관리목록(한국보건사회연구원, n.d.)을 바탕으로 저자 작성

[그림 3-10] 과제구분별 비공개 처리 사유 세분화



자료: 조사자료 관리목록(한국보건사회연구원, n.d.)을 바탕으로 저자 작성

데이터 유형별 비공개 처리 사유를 살펴보면, 양적 데이터의 경우에는 발주처의 비동의 사례가 가장 많은 것으로 나타났으며, 그 다음으로는 법률상 사유와 일반화 한계의 사유로 비공개 처리된 비율이 높게 나타났다. 한편, 데이터 유형별 비공개 처리 사유에는 발주처의 비동의, 법률상 사유 순으로 비공개 사유가 높게 나타났다. 이는 조사정보가 누락된 수탁과제의 특성이 반영된 것이다.

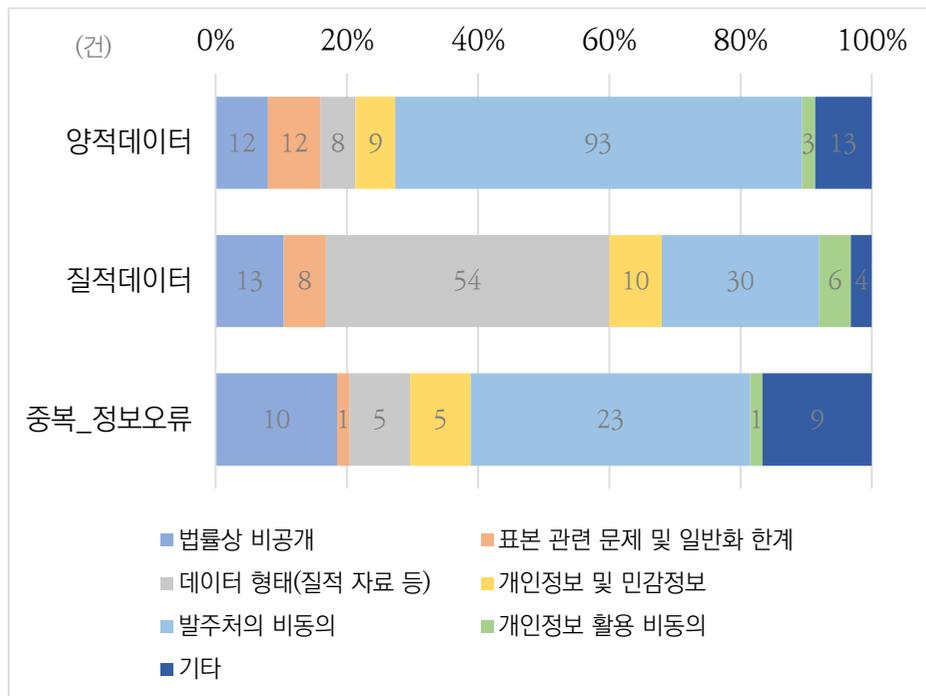
〈표 3-13〉 데이터 유형별 비공개 처리 사유 세분화(중복응답)

단위: 건(%)

데이터 구분	법률상 비공개	표본 관련 문제 및 일반화 한계	데이터 형태(질적 자료 등)	개인정보 및 민감정보	발주처의 비동의	개인정보 활용 비동의	기타
양적 데이터	12(8.0)	12(8.0)	8(5.3)	9(6.0)	93(62.0)	3(2.0)	13(8.7)
질적 데이터	13(10.4)	8(6.4)	54(43.2)	10(8)	30(24)	6(4.8)	4(3.2)
중복 기입 및 정보 오류	10(18.5)	1(1.9)	5(9.3)	5(9.3)	23(42.6)	1(1.9)	9(16.7)

주1: 각 셀의 백분율은 행별 합계를 기준으로 작성됨
 주2: 표의 가시성을 위하여 소수점 첫째 자리 표기
 자료: 조사자료 관리목록(한국보건사회연구원, n.d.)을 바탕으로 저자 작성

[그림 3-11] 데이터 유형별 비공개 처리 사유 세분화(중복응답)



자료: 조사자료 관리목록(한국보건사회연구원, n.d.)을 바탕으로 저자 작성

제3절 소결

본 장에서는 한국보건사회연구원 조사자료 관리지침에 대한 내용 분석과 조사데이터 관리 현황에 대한 사항을 검토하였다. 먼저, 한국보건사회연구원의 「조사자료 관리지침」은 원내 조사데이터의 체계적 관리와 활용에 근거가 되는 지침이며, 여기에서는 조사데이터 관리와 관련된 의무사항, 조사데이터 관리에 대한 각종 행정 사항과 관련 위원회 운영 등을 규정하고 있다. 해당 지침은 조사데이터 관리 전반에 대한 사항을 구체화하였다는 점에서 의의가 있으나 관련 실무와의 연결성, 데이터 보관을 위한 처리, 개인정보 보호 등의 사항에 대한 보완이 필요하다고 볼 수 있다.

한국보건사회연구원은 보건복지 영역에서의 인간대상연구를 수행하는 대표적인 기관으로, 매년 80건에 이르는 조사데이터를 생성하고 있다. 주로 보건·사회보장 영역에서의 각종 실태조사, 인식조사 등의 데이터가 생성되고 있으며, 노인·아동·장애인·청년 등 주요 사회 정책의 대상과 코로나·보건의료 등 주요 사회 현안에 대한 조사데이터가 생산·축적되고 있다.

향후 조사데이터의 관리 및 활용 활성화를 위하여 다음의 사항들을 검토해볼 필요가 있다. 첫째, 제공 가능성이 높은 조사데이터를 중심으로 공개범위를 점진적으로 확장시킬 필요가 있다. 즉, 조사 규모가 큰 자체 과제 가운데 양적 조사를 통하여 데이터를 수집한 경우는 제공 검토 대상 순위를 높게 설정할 수 있다. 한편, 단기간에 이루어진 소규모 조사데이터의 경우에는 정기적인 위원회 심의 안전(또는 보고 안전)으로 상정하여 순차적으로 검토하고, 필요시 전문가의 의견을 받아 공개/비공개 처리할 것을 검토해볼 수 있다.

둘째, 조사데이터의 관리를 위한 사전/사후 정보수집이 강조될 필요가

있다. 데이터관리계획 등 조사데이터 관리 절차에 따른 조사정보 수집에 따라 데이터의 제공/이관 수준이 달라질 수 있으므로 이에 대한 기존 절차가 보다 강화될 필요가 있다. 다시 말해 조사데이터 생성과 제출 시 관련 행정처리에 대한 사항이 지침에 따라 추진될 수 있도록 기관 차원에서의 노력이 요구된다.

셋째, 데이터 제공과 이관에 대한 사례 축적과 가이드라인 제시가 필요하다. 데이터의 특성 또는 연구분야의 특성 등으로 인하여 데이터의 공개 여부를 결정하기 위한 구체적인 조건을 설정하는 것은 상당히 어려운 것이 사실이다. 따라서 데이터 공개 처리와 관련된 사례가 지속적으로 축적되어 대략적인 데이터 공개 처리 범위가 형성되는 것이 중요하다. 초기에는 위에서 설정한 ‘제공 검토 대상’의 우선순위가 높은 데이터를 중심으로 기준을 설정할 수 있다. 이후에는 심의 결과를 통한 공개 사례를 축적하여 데이터의 제공 범위를 확대시킬 필요가 있다. 한편, 공개/비공개/이관 불가 조사데이터에 대한 정보가 공유될 필요가 있다. 데이터 생성자가 데이터 제출과 관련된 사항을 결정할 시에 이러한 정보가 유용하게 활용될 수 있으며, 데이터 관리 행정의 효율성 또한 제고될 수 있다.



제4장

조사데이터의 체계적 관리 및 활용 제고를 위한 방안

제1절 조사데이터 관리체계 개선 방향

제2절 조사데이터의 체계적 관리 방안

제3절 조사데이터의 활용 활성화 방안

제4장 조사데이터의 체계적 관리 및 활용 제고를 위한 방안

제1절 조사데이터 관리체계 개선 방향

1. 기존 조사데이터 관리체계 방향

한국보건사회연구원의 조사데이터 관리는 조사자료 관리지침(2022a)에서 명시하고 있는 기본원칙(제3조)을 따르고 있다. 원칙에 따르면, 연구원 구성원은 조사자료에 대한 보편적 이용권을 보장 및 확대하기 위하여 지속 노력해야 한다. 구체적으로는 1) 연구자가 연구를 위한 조사를 수행할 때는 조사데이터에 대한 관리 등을 사전에 고려하여 조사자료 관리 계획을 수립하여야 한다. 2) 연구자는 연구원에서 생산한 조사자료를 연구원에 제출하여야 한다. 3) 연구원에 제출된 조사자료는 국민의 알 권리를 보장하기 위하여 공개 가능하도록 처리되어야 한다. 4) 조사자료에 대한 수요에 따라 데이터가 제공될 수 있도록 노력해야 한다.

기본적으로 연구원에서 생성한 모든 조사데이터는 관리대상에 포함되기 때문에 데이터의 제출(데이터 이관)은 필수적이라 볼 수 있다. 특히 연구목적으로 수집된 데이터는 공공 저작물에 해당되기 때문에 공공데이터에 해당될 수 있다. 다만, 수탁과제의 경우에는 데이터의 소유권이 발주 기관에 있기 때문에 데이터의 관리 대상에 포함되는지 여부를 사전에 논의할 필요가 있다. 그 외에 사항은 연구책임자의 결정이 영향을 미친다. 연구책임자가 조사데이터가 공개될 시에 발생할 영향에 대하여 적절히 판단할 수 있다는 전제가 반영된 것이다.

〈표 4-1〉 관리 대상 조사데이터의 범위

구분		이관	제공
생산 과제종류	자체과제	○	○
	수탁과제	○ (발주처 요청에 따라 이관 불가일 수 있음)	△ (발주처와 합의 필요)
조사수행 주체	내부	○	○
	외부	○	○
조사규모	모집단	○	○
	표본집단	○	○
데이터 유형	질적	○	△ (연구자 결정)
	양적	○	○
	질적+양적	○	△ (연구자 결정)
공표 여부	공표 이전 데이터	○	△ (연구자 결정)
	공표 데이터	○	○

자료: 정영철 외(2020). 전사적 연구데이터 관리 체계 마련을 위한 연구. 한국보건사회연구원.

기존 조사데이터 관리 체계(정영철 외, 2020)에서는 데이터의 이관 이전에 마이크로데이터의 생성을 강조하고 있다. 다시 말해 원자료에서 나타나는 입력오류, 논리오류, 개인정보 및 민감정보 포함 등의 문제를 적절하게 처리하여 제공 가능한 데이터로 변환하는 과정을 의미한다. 실제 조사데이터 관리 체계에서 데이터 이관 및 데이터 공개 시에 이에 대한 연구책임자의 검토와 데이터 공개 담당자의 교차 점검이 설정되어 있다.

[그림 4-1] 조사데이터 관리 및 활용 처리 절차



자료: 정영철 외(2020). 전사적 연구데이터 관리 체계 마련을 위한 연구. 한국보건사회연구원.

한국보건사회연구원의 조사데이터 관리체계(정영철 외, 2020)에서는 연구수행 과정 중에 생성된 공공데이터임을 강조하고 있으며, 공공데이터로서의 활용을 위하여 연구원에서는 적극 노력할 필요가 있음을 지침을 통하여 명시하고 있다. 이상의 방향성은 향후 조사데이터 관리 방향 설정 시에도 필수적으로 포함되어야 하는 사항이며, 이에 대한 실질적인 노력이 점진적으로 제도화될 수 있도록 정기적인 실태분석과 지침 개정, 연구원 구성원의 참여 등이 함께 이루어질 필요가 있다.

2. 조사데이터 관리체계 방향성 설정

한국보건사회연구원 조사데이터 관리체계의 방향성을 설정하면 아래 그림과 같다. 먼저, 조사데이터의 관리는 크게 데이터 관련 법률에 근거하여 그 필요성이 제시될 수 있다. 연구원에서 생성한 조사데이터는 공공데이터임과 동시에 인간대상의 조사결과물이기 때문에 공개의 대상이면서 조사참여자의 권리에 대한 보호의 대상이 된다. 또한, 연구원에서 생성된 조사데이터는 공공데이터에 해당되므로 이에 대한 국민의 접근성을 제고하는 것이 필수적이다. 한편, 조사참여자는 조사결과에 대한 청구권을 가지며, 자신의 정보가 동의에 따라 적절하게 활용 및 보호되고 있는지를 알 수 있어야 한다. 따라서 한국보건사회연구원에서 생산한 조사데이터는 이상의 법적 배경 아래에서 관리되어야 한다.

한편, 한국보건사회연구원은 공공데이터를 관리 및 제공해야 하는 공공기관에 해당되며, 자체적인 통계를 생산하는 기관(여기서는 조사통계에 한함)으로 등록되어 있기 때문에 제공되는 데이터에 대한 자체적인 질 제고와 함께 데이터 공개 의무를 충분히 이행해야 한다. 이와 더불어 한국보건사회연구원은 자체적인 기관위원회를 운영하는 연구기관으로, 조사참여자에 대한 보호와 조사데이터의 보안 방안으로 지속적으로 점검해야 한다.

이에 따라 한국보건사회연구원은 조사데이터의 체계적 관리를 위하여 조사데이터의 품질 제고를 위한 노력을 해야 하며, 관련 지침을 기반으로 데이터의 타당도와 신뢰도, 정확도 등을 유지·제고하기 위한 실질적 체계를 갖출 필요가 있다. 뿐만 아니라 인간대상 연구기관의 책무로서 개인정보보호를 위한 기반을 강화하여 조사수행 과정 및 조사결과의 활용 중에 발생될 수 있는 개인정보 및 민감정보의 유출이나 정보 보안의 취약성 등

을 사전에 검토해야 한다. 반면에 연구원에서 수행되는 데이터는 공공데이터 해당되므로 이에 대한 제공범위 설정 기준과 그 절차를 투명하게 공개해야 한다. 이때 연구자의 자율성이 보장될 필요가 있으나 정보의 제공 및 개인정보의 보호와 관련된 논의가 필요한 상황에서는 연구원의 조사 데이터 관리 원칙이 적용될 필요가 있다. 마지막으로 연구원은 공적 자원을 활용하여 생산된 조사데이터가 보건 및 사회보장 분야의 발전에 충분히 기여할 수 있도록 그 가치를 제고하는 데 지속 노력할 필요가 있다. 여기에는 데이터 생산의 효율성을 제고하는 방안도 포함될 수 있다.

[그림 4-2] 조사데이터 관리 방향 설정



제2절 조사데이터의 체계적 관리 방안

1. 조사데이터의 체계적 관리를 위한 정보 세분화

조사데이터의 체계적 관리를 위하여 필요한 가장 기초적인 사항은 조사데이터 정보의 구성에 따라 세부적인 관리 방향을 설정하는 것이다. 조사데이터는 크게 5가지의 영역으로 구분될 필요가 있다.

첫째, 연구목적에 따라 수집된 조사참여자의 응답결과이다. 이는 연구 목적으로 활용하고자 조사참여자가 직접 응답한 결과이며, 순수한 조사 목적의 정보를 가지고 있으므로 데이터의 보관 및 품질관리에 대상이며, 공공데이터의 성격을 가진다고 볼 수 있다. 따라서 조사데이터의 보관 사항에 대한 사전 공지가 필수적이다.

둘째, 조사 수행의 목적에 따라 수집된 개인정보이다. 대표적으로 조사 응답자의 인구사회학적 정보가 여기에 해당되는데, 이는 개인을 특정하거나 소수 집단에 대한 정보가 포함될 가능성이 있기 때문에 개인정보 보호법에 근거한 보호 또는 처리의 대상이다. 특히 해당 정보의 수집과 활용, 보관 기간 등은 사전에 조사참여자의 동의를 얻어야 한다.

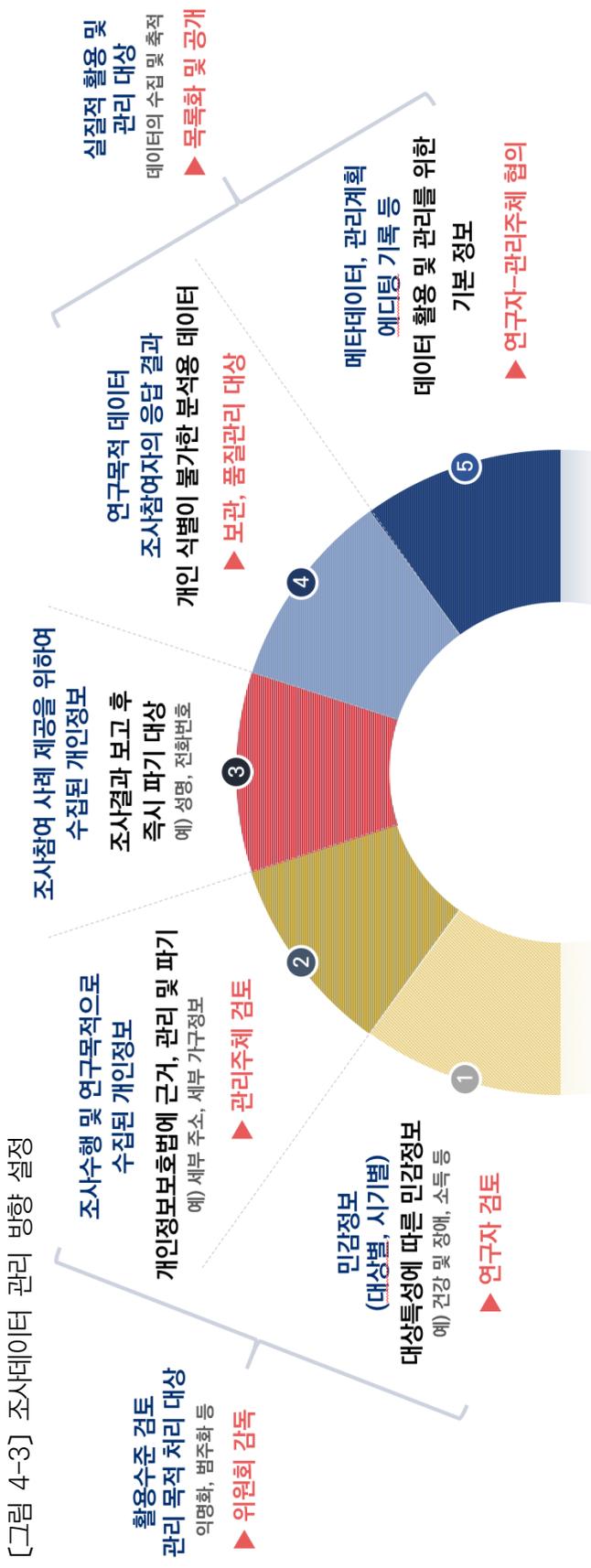
셋째, 조사 수행 목적과 관계없이 수집된 개인정보이다. 대표적인 예로 조사참여 사례 제공을 위하여 수집된 성명 및 전화번호, 계좌번호 등이 있다. 또한 조사에 대한 사후 검토 등을 위하여 수집된 메일 주소나 소속 기관 정보도 이에 포함된다. 이상의 정보는 연구의 종료 또는 데이터 수집 목적이 달성되는 즉시 파기되어야 하며, 보관용 데이터에도 해당 정보가 포함되지 않도록 관리해야 한다.

넷째, 특정 대상 및 시기에 따라 달라질 수 있는 민감정보이다. 상시적인 민감정보로는 개인의 건강 정보, 장애, 구체적인 소득·자산 정보가 있

다. 대상 및 시기에 따라 민감도가 달라질 수 있는 정보에는 특정 시기에 발생한 이슈에 맞게 조사된 결과나 소수의 모집단이 설정된 조사의 경우가 해당된다. 이는 다소 주관적 판단에 의하여 설정될 수 있는 사항이기 때문에 연구책임자의 사전 검토(DMP작성을 통한 검토, IRB심의 절차를 통한 보완 등) 및 사후 공개범위 설정 시에 필수적으로 고려되어야 한다.

다섯째, 조사데이터와 관련된 다양한 형태의 자료이다. 필수적인 관련 자료로는 조사데이터에 대한 메타데이터 정보, 데이터관리계획이 있다. 또한 원자료에서 연구자료로, 보관용 자료에서 제공용 자료로 수정되는 일련의 과정이 내검기록으로 작성되어야 한다. 이는 연구자의 분석 단계에서도 작성될 수 있지만 데이터 이관 이후 내부적인 검토 또는 제공 과정 중에 나타날 수 있는 이의신청 반영 과정에서도 보완될 수 있다.

이상의 내용을 바탕으로 연구원 내 조직 및 구성원의 역할이 구체화될 수 있다. 먼저, 연구책임자는 연구데이터의 이관(첫째)과 관리(다섯째)에 의무를 가지게 되며 민감정보에 대한 사전 판단(넷째)을 적절한 절차에 따라 추진해야 한다. 관리위원회에서는 연구 목적의 조사정보와 조사관련 정보를 수집·축적·관리하고 목록화하여 내·외부적 목적에 따라 공개해야 한다. 또한, 조사 시 수집된 개인정보의 처리 및 관리 등을 지원하고 제공용 데이터 생성 시 발생될 수 있는 문제들을 최소화할 수 있는 방안을 모색해야 한다. 특히 개인정보 및 민감정보의 처리 수준을 심의하거나 시기적·대상적 특성을 반영한 민감도를 고려하여 비공개 여부를 판단하여야 한다.



자료: 저자 작성

2. 데이터 제공 단계의 세분화

현재 조사데이터는 공개와 비공개로 구분하여 관리되어 있다. 조사데이터의 내·외부적 활용가치를 제고하기 위해서는 연구원에서 생산되고 있는 조사데이터 특성을 고려하여 데이터 제공 단계를 세분화할 필요가 있다. 먼저, 제공 단계는 크게 1) 비공개 관리, 2) 정보공개/목록관리, 3) 결과표 공개, 4) 마이크로데이터 제공으로 구분한다.

1단계 비공개 관리 대상은 연구책임자, 관리위원회, 발주처 등의 판단 및 결정에 따라 비공개 대상으로 설정된 데이터이다. 1단계 데이터의 경우 비공개 설정의 근거를 기록(비공개 처리 승인 증빙, 발주처 협의 결과 등)하여야 하며, 해당 결정이 DMP내용과 일치하는지를 확인해야 한다. 다만, 비공개 처리가 데이터이 미이관을 의미하는 것이 아니기 때문에 최소한의 개인정보 보호 조치가 이루어질 필요가 있으며, 연구자 개인의 저장매체에 임의로 저장·관리하는 것을 지양해야 한다.

2단계 정보공개/목록관리는 조사정보(조사의 목적, 내용, 대상 등)를 공개하고 원내 구성원들이 내부적으로 활용할 수 있도록 최소한의 정보를 공개, 목록화하는 단계를 말한다. 조사데이터의 정보가 목록화되고 이러한 정보가 공유될 경우, 데이터 수집의 비효율성이 감소하고 데이터 관리 계획 작성, IRB심의 등 데이터 관리 절차가 원활하게 추진될 수 있다. 예를 들어 조사의 기획 단계에서 원내 조사데이터의 관련 정보를 탐색, 참고하여 DMP, IRB심의 연구계획 등이 작성될 수 있다. 또한 유사·중복 조사를 방지하여 조사데이터 생산의 효율성을 제고할 수 있다. 다만 2단계 관리를 위해서는 조사에 대한 기본적인 사항(DMP, 메타데이터, 내검 기록 등)이 제출되어야 한다.

3단계 결과표 공개 단계에서는 실제 조사결과에 대한 외부 공표가 이

루어진다. 조사데이터의 원자료에서 개인정보·민감정보의 분리 및 처리가 불가능한 경우에 제한된 연구결과를 제공(예를 들어 보고서에 포함된 결과표 또는 보고서에 포함되어 있지 않지만 연구 목적으로 수집된 결과 일부)할 수 있다. 이는 연구성과의 확대에 기여할 수 있으며, 연구자가 아닌 대국민 정보 제공의 측면에서 의미가 있다. 이 단계는 원자료의 공개가 어려우나 연구보고서가 공개 발간된 경우, 데이터에 대한 품질문제가 발생하여 제한된 분석만 수행해야 하는 경우, 특정 이슈에 맞게 편의 표집에 의해 조사가 이루어진 경우 등에 적용될 수 있다. 또한, 질적 데이터의 분석 결과(키워드를 활용한 분석, 인터뷰 결과의 요약 및 도식화 등)도 적용 가능하다.

4단계 마이크로데이터 제공 단계는 개인정보와 민감정보의 처리가 완료되고 데이터의 오류 등이 충분히 수정된 데이터가 해당될 수 있다. 이는 원외 연구자 및 국민 대상으로 MD를 제공하는 것이기 때문에 데이터 요청 및 건의사항 등에 대한 적절한 피드백을 관리하는 것이 중요하다. 대표적으로 데이터 요청 관련 정보(이용자 정보, 이용 목적 등)를 관리하고 데이터의 오류가 사후에 발견될 경우에 이에 대한 조치 내용 및 결과를 기록해야 한다. 특히 데이터 요청에 대한 업무과중이 분산될 수 있도록 데이터 생성자(주로 연구책임자)와 데이터 관리자(주로 관리위원회 담당자)간의 역할이 명확히 설정되어야 한다. 일반적으로 대규모의 양적조사의 경우에 해당 단계에 포함될 가능성이 높다(제3장 참고).

138 보건복지 조사데이터의 체계적 관리 및 활용 방안 연구

[그림 4-4] 조사데이터 제공 단계 세분화(안)

(기준) 비공개/공개 → (개선) 4단계 세분화



자료: 저자 작성

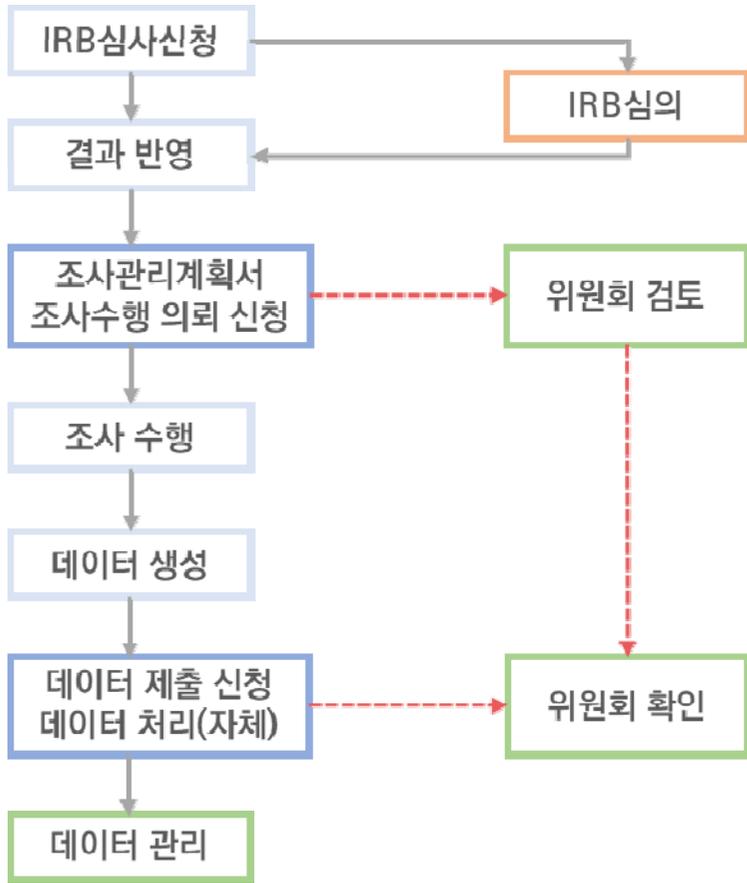
3. 데이터 생성 및 관리 모형 설정

가. 자체과제의 데이터 생성 및 관리

조사데이터의 생성 및 관리의 과정은 조사데이터의 체계적 관리의 시작이다. 특히 조사데이터의 생산 과정의 정교함이 데이터의 신뢰도·타당도와 정보보안 강도 및 적정성을 제고하는데 영향을 미친다. 데이터 생성 및 관리 모형 설정 시에 필수적으로 요구되는 사항은 1) 사회보장재정·데이터 연구실을 중심으로 한 실질적인 데이터 관리 체계 구축, 2) 데이터 품질 관리를 위한 노력과 내검 절차에 대한 관리, 3) 데이터의 관리 및 활용과 관련된 협의 창구의 일원화이다.

먼저, 기존의 조사데이터 생성 과정을 살펴보면 아래 그림과 같다. IRB 심의를 받는 경우에는 조사 수행 전에 IRB심의 절차를 거치게 되며, 자체과제의 경우에 조사관리계획 및 조사수행의뢰서를 작성하게 된다. 이후 조사가 수행되고 데이터가 생성된다. 연구종료 이후 일정 기간 내에 데이터를 이관하게 되는데, 이때 개인정보 및 민감정보의 처리는 연구책임자가 자체적으로 수행하여, 이에 대한 검토는 현재 이루어지고 있지 못한 상황이다. 그러나, 공개 데이터의 경우 관리위원회의 심의가 필수적으로 요구된다.

[그림 4-5] 기존 자체과제의 데이터 생산 및 관리 절차



자료: 저자 작성

현재 지침(한국보건사회연구원, 2022)에 따라 조사자료 관리위원회는 1) 연구자의 조사자료 관리계획서 접수 및 등록 업무, 2) 연구자의 조사자료 제출에 따른 보관용 조사자료 취합 업무, 3) 보관용 조사자료 보관을 위한 조사자료 관리시스템 운영 업무, 4) 조사자료 제출 확인 및 독려 업무, 5) 조사자료 제공을 위한 사전 준비 업무, 6) 연구원에서 운영하는 정보시스템(웹사이트, 원내업무포털)을 활용한 조사자료 제공 업무, 7) 기타 조사자료 관리에 필요한 제반 업무를 수행해야 한다. 그러나 현재는 조사자료 관리에 대한 주요 안건의 심의 처리를 중심으로 기능을 수행하

는 한계가 있다.

조사자료 관리위원회의 기본적인 기능을 수행하기 위해서는 위원회 실무를 담당할 수 있는 전담 인력이 1인 이상 배치되어야 하며, 데이터와 관련된 일련의 절차를 지속적으로 관리하여야 한다. 또한 데이터 생성 시에 작성된 정보에 대해 내외부 자원을 활용하여 검토해야 한다.

조사데이터 생산 및 관리의 체계를 확립하기 위해서는 조사자료 관리 위원회의 기능이 실질화(위원회의 정기적 운영, 전담 인력의 배치 등)될 필요가 있다. 조사데이터 생산 및 관리체계 구축을 위한 조사자료 관리위원회 실질화 모형(안)은 아래 그림과 같다.

관리위원회는 조사수행 및 조사관리계획과 관련된 상시적인 자문의 역할을 수행할 수 있다. 이는 위원회 인력풀을 활용하되, 담당자의 사전 검토에 따라 투입 인력을 결정한다. IRB위원회의 심사 신청 전후로 조사관리 계획서 및 조사수행 의뢰서를 작성하여 관리위원회에 제출한다. 이때 조사관리계획서는 연구책임자의 통보가 아닌 협의를 통하여 확정하는 것을 원칙으로 한다. 만약, 조사관리 계획서 내에 생명윤리 또는 개인정보 보호 등의 이슈가 발생할 경우에 관리위원회의 결과 작성 시 IRB승인 의견을 참고할 수 있다. 다만, 해당 절차에서 두 위원회의 기능²³⁾이 다르기 때문에 연구책임자는 각 검토의견을 반영하여 조사를 추진해야 한다.

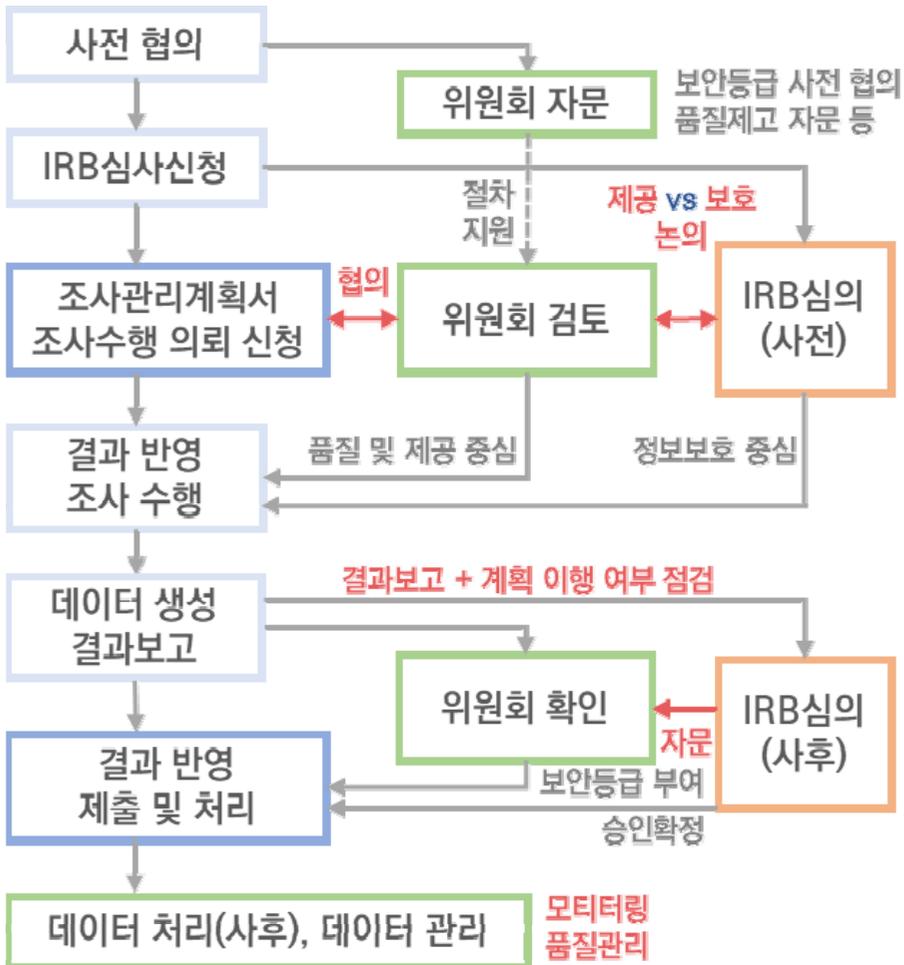
조사데이터 생성 이후에는 IRB에 결과보고 및 종료보고를 3개월 이내에 해야 한다(생명윤리위원회 운영규정 제19조²⁴⁾). 이후 조사데이터를 활용한 연구가 완료된 이후 보관용 데이터 작성, 제공용 데이터 작성이 완료되면 관리위원회에 조사자료 제출 신청서와 함께 조사자료 전체를

23) 조사자료 관리위원회는 조사관리 계획서의 적정성, 데이터 공개 및 관리를 위한 준비 사항, 조사수행의 적절성 등에 대하여 심의한다. 반면, IRB에서는 「생명윤리위원회 운영규정」에 따라 연구계획서의 윤리적 타당성, 연구대상의 안정과 개인정보 보호 대책 심의 등을 검토(제6조)한다.

24) 한국보건사회연구원(2022b). 생명윤리위원회 운영규정

제출한다. 이때 위원회에서는 연구책임자의 의견(DMP 작성 내용과 제출 신청서 내 의견사항)과 IRB의 심의 결과를 고려하여 자료 제공 범위 및 보안등급을 부여한다. 만약 조사자료의 공개나 개인정보보호 및 민감정보 처리와 관련된 심의 사항이 발생할 경우에 관리위원회를 통한 심의를 진행한다. 데이터가 제출된 이후에는 관리위원회의 실무적 차원에서 데이터의 보관과 내검기록 관리 등을 수행한다.

[그림 4-6] 모형1: 조사자료 관리위원회 실질화 모형(안)



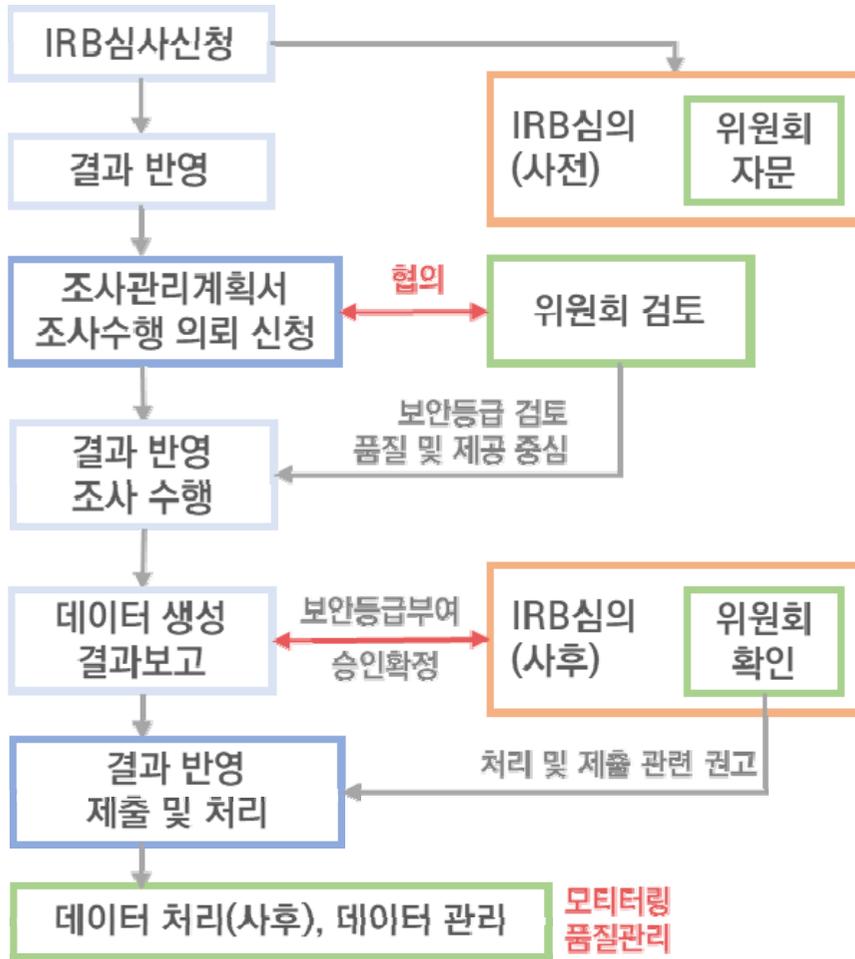
자료: 저자 작성

만약 조사자료 관리위원회가 실무기능을 가지지 못할 경우에는 IRB와 같은 기존의 기구를 활용할 수 있다. 다만, 이 경우는 IRB 운영의 강화(심의 의무 강화, 종료보고 의무 등)가 전제되어야 하며, 관련 규정 및 지침의 개정이 필요하다.

IRB강화 모형(이하 그림 참고)에서는 IRB심의 시 관리위원회의 DMP 검토가 동시에 이루어진다. 그러나 각 위원회의 심의 내용과 운영 방식에 차이가 있으므로 위원회 소속 위원이 동시에 각각의 사항을 심의할 수 없다. DMP검토 시 IRB위원회는 연구윤리와 관련된 검토의견을 관리위원회 위원은 데이터 제공 및 품질관리 관련 의견을 작성한다.

데이터 생성 후에는 조사자료 제출 신청과 종료보고를 함께 추진한다. 이 때에도 DMP검토와 마찬가지로 각 위원회의 역할을 나누어 관련 의견을 작성한다.

[그림 4-7] 모형2: IRB강화 모형(안)



자료: 저자 작성

〈표 4-2〉 조사자료 관리위원회와 생명윤리위원회 비교

비교 영역	조사자료 관리위원회	생명윤리위원회
근거	조사자료 관리지침	생명윤리위원회 운영규정
목적	조사자료 관리 업무의 체계화 및 이용 활성화	인간대상연구에 관한 심의
구성	<ul style="list-style-type: none"> • (위원장)부원장 (부위원장)사회보장재정·데이터연구실장 • (위원구성)조사자료 제출, 공개 및 제공 등 심의 가능한 전문적 지식을 갖춘 자 • (임기) 2년, 연임 가능 	<ul style="list-style-type: none"> • (구성)원장 위촉 5인 이상, 전문성을 가진 외부위원 1명 이상 포함 • (실무)간사 1인 배치, 실무분과위원회 구성 가능
역할	<ul style="list-style-type: none"> • 조사자료 관리계획서 접수 및 등록 업무 • 보관용 조사자료 취합 • 조사자료 관리시스템 운영 • 조사자료 제출 확인 및 독려 • 조사자료 제공을 위한 사전 준비 업무 • 원내 정보시스템을 활용한 조사자료 제공 • 기타 조사자료 관리에 필요한 제반 업무 • 조사자료 관리계획, 제출·보관·제공 등 관리 전반에 대한 정책 수립 	<ul style="list-style-type: none"> • 연구계획서의 윤리적·학문적 타당성 심의 • 연구대상자 등으로부터 적절한 절차에 따라 동의를 받은지 여부 • 연구대상자 등 안전에 관한 사항 심의 • 연구대상자 등 개인정보보호 대책 심의 • 기타 연구 관련 생명윤리 및 안전에 관한 사항 심의 • 인간대상 연구의 진행과정 및 결과에 대한 조사·감독 • 원내 구성원 대상 교육, 취약한 연구 대상자에 대한 보호 대책 수립, 연구자 윤리지침 마련 등
심의	<ul style="list-style-type: none"> • 조사자료 관리 전반에 대한 점검 • 조사자료의 공개 여부 심의 • 비공개 조사자료 공개 요청에 대한 승인 심의 • 제3자 제공에 대한 관리·감독에 관한 사항 심의 	<ul style="list-style-type: none"> • 인간대상연구에 대한 계획서 심사 (인간을 대상으로 물리적으로 개입하는 연구, 상호작용을 통하여 수행하는 연구, 개인 식별 정보를 이용하는 연구) • 연구계획 및 변경, 종료보고 심의
기타 수행 업무	<ul style="list-style-type: none"> • 조사자료 품질관리 • 조사자료 제공을 위한 처리 • 조사자료 활용 관련 성과 측정 및 전략 수립 • 조사데이터 관리 체계 정착을 위한 유인책 마련 • 조사데이터 관리 관련 법률적 자문 	<ul style="list-style-type: none"> • 생명윤리위원회에서 요청한 자문 안전 처리 등 • 심의 대상 제외 범위 설정 • 연구종료 보고에 대한 심의 • 기타 생명윤리법의 사항에 대한 감독

주: 2023년 7월말 기준, 한국보건사회연구원 규정 및 지침을 바탕으로 비교
 자료: 저자 작성

조사데이터 관리체계 구성 시 데이터 생성자(연구책임자)는 아래와 같은 업무절차를 따르게 된다. 먼저, 조사데이터 생성을 위한 계획을 수립하고 연구계획 또는 조사계획에 대한 IRB심의를 신청²⁵⁾한다. IRB심의 결과 확인 이후에 데이터관리계획과 조사수행의뢰서를 작성한다. 이때 조사 개요에 대한 설정과 사전 데이터 관리 규칙의 설정 등이 이루어진다. 조사자료 관리 담당자는 제출양식의 내용을 확인 후에 접수한다. 이때 양식 작성 관련 수정 사항이 발생할 경우에 조사자료 관리 담당자는 일부 수정을 요청할 수 있다. 그러나 조사자료 관리 담당자의 검토를 통하여 결정될 수 없는 사항이 있거나 데이터의 특성상 전문가의 검토가 요구되는 경우에는 조사자료 관리위원회의 위원에게 검토의견을 요청할 수 있다. 이 절차를 통하여 작성된 검토의견은 데이터 생성자에게 전달되어 실제 조사수행 시에 반영될 수 있게끔 지원한다.

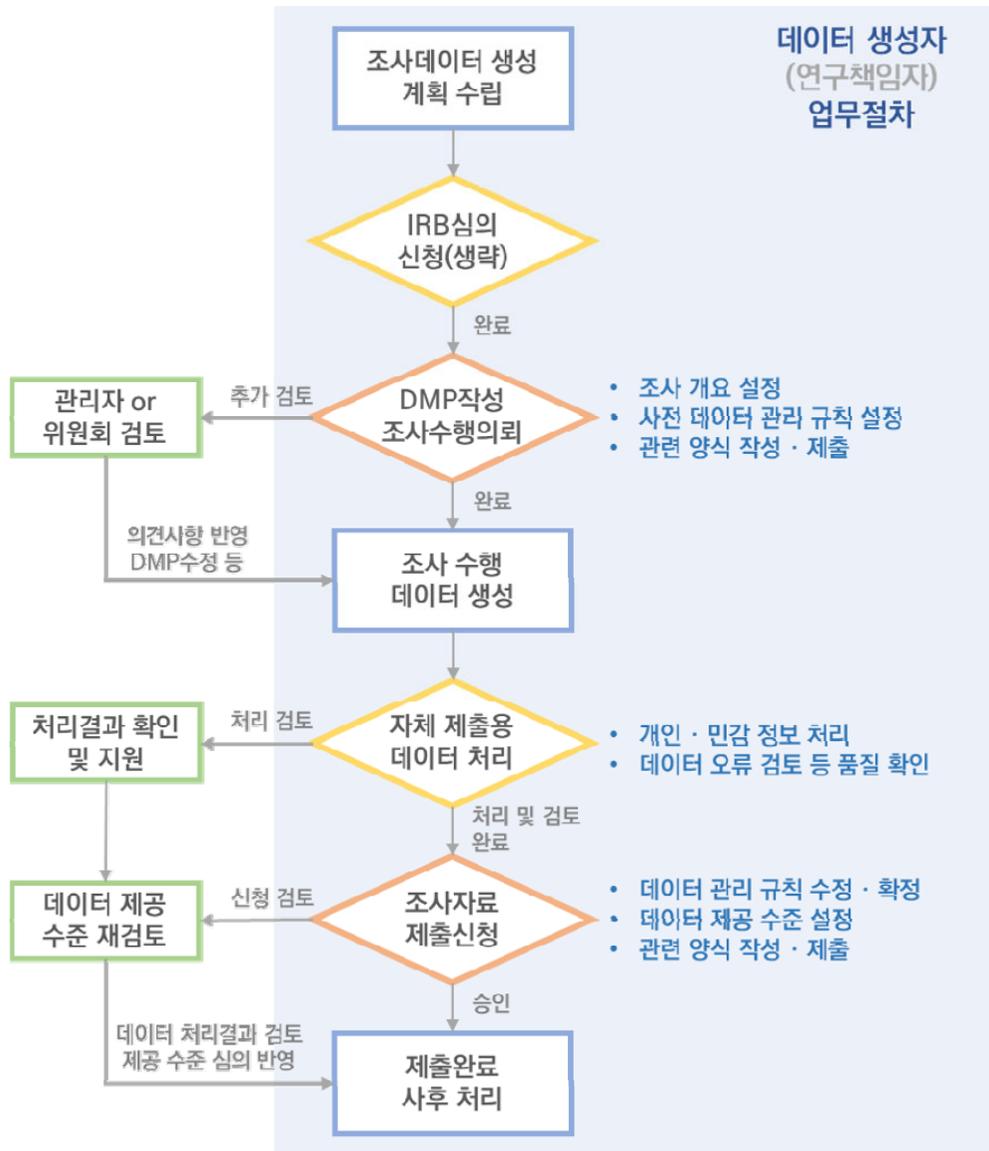
데이터 생성 이후 연구책임자는 자체적으로 제출용 데이터(이관용 데이터)로 원데이터를 수정·처리하게 된다. 대표적으로 개인정보 및 민감정보에 대한 처리 또는 삭제가 이루어지며, 데이터의 품질관리를 위한 기본적인 데이터 오류 사항을 검토(결측 처리, 오타자 수정, 논리적 오류 수정 등)한다. 제출용 데이터에 대한 처리 결과에 대해 관리 책임자는 점검하여 추가적인 데이터 처리가 필요한지 여부를 판단한다. 데이터에 대한 자체 처리가 완료되면 연구책임자는 조사자료 제출신청 양식을 작성하고 조사데이터 관련 자료 전체를 이관한다. 이때 연구책임자의 데이터의 특성을 고려하여 데이터 관리 규칙을 확정하고 데이터 제공 수준을 제안한다.

조사자료 관리 담당자는 양식 작성 내용을 검토 후 제출절차를 완료한다. 만약 데이터 제공 수준 설정 및 데이터 관리 규칙 등에 대한 위원회 심

25) 생명윤리위원회 심의 절차는 본 연구의 범위에서 벗어나므로 구체적인 사항은 생략함.

의가 필요하다고 판단될 경우, 위원을 통한 자문 또는 위원회 개최를 통한 관련 내용 심의 등을 추진할 수 있다. 연구책임자는 심의결과를 반영하여 조사자료 제출 절차를 완료한다.

[그림 4-8] 조사데이터 생성 및 관리 체계에 따른 데이터 생성자의 업무 절차



자료: 저자 작성

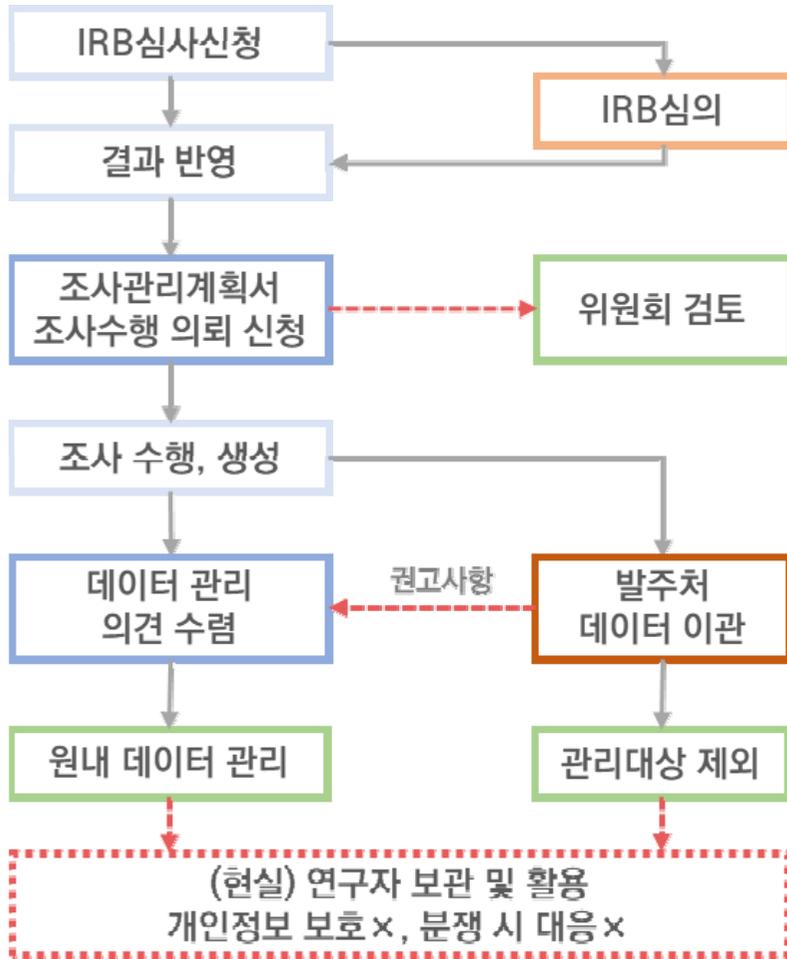
나. 수탁과제의 데이터 생성 및 관리

수탁과제의 경우, 발주기관이 데이터에 대한 소유권을 가지고 있기 때문에 데이터 이관율이 상대적으로 낮다. 그러나 데이터의 소유권이 없다고 하여 조사데이터 생성 체계에 배제될 수는 없다. 조사데이터의 생성을 위한 관리 체계(조사자료 관리와 IRB포함)는 조사데이터의 신뢰도와 타당도, 조사대상자에 대한 보호 및 데이터 활용성 제고를 위한 필수적인 사항이기 때문이다.

일부의 수탁과제는 원내 IRB심의를 받고 조사자료 관리계획에 따른 조사관리 계획서(조사의뢰신청서 제외)를 작성하였다. 다만 관리위원회에서 수탁과제의 DMP를 검토하여 의견을 작성한 사례는 없다.

조사데이터 생성 및 연구 종료 이후 발주처에 데이터를 이관하게 되는데, 이관이 완료되면 연구원에서는 데이터의 보관 의무를 가지지 않는다. 다만, 발주기관의 연구원에서의 데이터 관리 및 제공을 요청할 경우에는 원내 시스템 안에서 데이터를 관리한다. 그러나 공식적인 데이터 이관이 이루어지지 않은 수탁과제 조사데이터의 경우에는 원내 관리체계에서 벗어나기 때문에 데이터의 보안이나 개인정보 및 민감정보의 처리 등이 확인되지 않는다. 특히 연구자가 개인 저장장치에 데이터를 보관하거나 발주기관에서 데이터를 요청할 시 연구자가 비공식적으로 제공하는 등의 사례가 발생할 수 있다. 발주기관으로의 데이터 이관 또는 원내 시스템 내 데이터 관리가 체계적인 절차 안에서 결정되지 않을 경우에 조사데이터 관련 분쟁 발생 시 원 차원에서의 적절한 대응이 어려울 수 있다. 따라서 연구원의 대응력 강화와 연구책임자에 대한 보호, 발주처와의 신뢰 확보 등을 위하여 수탁과제 데이터 생산 및 관리체계가 보다 실질화될 필요가 있다.

[그림 4-9] 기존 수탁과제의 데이터 생산 및 관리 절차



자료: 저자 작성

수탁과제 관리모형을 조사자료 관리위원회 실질화 모형을 기준으로 아래 그림과 같이 개선할 필요가 있다. 데이터 생성 및 결과 검토 이후 데이터의 이관에 대한 발주처와의 사전 협의가 이루어진다. 이때 데이터 보관의 필요성이나 법적 대응을 위한 방안 등에 대하여 설명하고 발주처의 의견에 따라 데이터 이관을 진행한다. 만약 원내 관리체계에 포함될 경우에는 발주처에서 데이터의 관리를 요청한다는 정보를 기록할 필요가 있다. 이후의 절차는 조사자료 제출 및 제출용 데이터 작성, 내검 기록 등 자체

연구 데이터관리 체계와 동일하다. 다만 연구책임자와 발주기관에 한하여 데이터에 접근 가능하며, 데이터 활용 요청 시 발주기관의 허가가 요구된다. 한편, 발주처에서 데이터 관리를 요청하지 않는 경우에는 데이터의 폐기 또는 발주처 이관 내용을 기록할 필요가 있다.

[그림 4-10] 수탁과제 관리 모형: 조사자료 관리위원회 실질화 모형 기준



자료: 저자 작성

4. 데이터관리 체계 및 주체별 역할

가. 데이터관리 체계 주체 및 기반 설정

데이터관리 체계 구성 시 절차별 주체의 역할이 명확히 설정될 필요가 있다. 또한 이러한 역할이 적절하게 수행될 수 있도록 관련 내용을 공식화, 교육 및 매뉴얼 등 제공, 데이터 관리 및 제공 관련 플랫폼 활용 등이 고려되어야 한다.

먼저 데이터 관리 체계의 주체는 크게 데이터 생성자, 데이터 관리자, 데이터 사용자, 관련 위원회로 구분하였다. 먼저 데이터 생성자는 연구원 내 연구진을 의미하여 연구 수행 과정에서 조사데이터를 생성하고 데이터 제출을 위한 사전 검토를 수행한다. 데이터 관리자는 연구원 내 데이터 관련 전담 부서를 의미하며, 조사데이터 관리 체계의 운영에 기여한다. 데이터 사용자는 원내외 조사데이터를 활용할 수 있는 주체를 의미한다. 조사데이터의 공개범위에 따라 사용자가 달라질 수 있으나 공통적으로 모든 데이터는 사용자 친화적인 형식으로 제공 관리되어야 한다. 마지막으로 위원회는 조사데이터 수집과 관련된 각종 위원회와 데이터 관련 책임 제도 등의 주체를 의미한다. 조사데이터의 경우에는 관리위원회의 역할이 중요하지만 그 외에도 생명윤리와 관련된 이슈나 공공데이터 제공 요청에 대한 대응 등 조사데이터의 관리 과정 중에 발생할 수 있는 다양한 사례들은 각 조직별로 대응할 필요가 있다.

한편, 조사데이터 관리의 주요 기반은 앞서 살펴보았던 관련 법령과 규정, 지침, 규칙 등이다. 이와 더불어 조사데이터 관리를 위한 가이드라인 및 매뉴얼 작성, 절차에 대한 안내 및 관련 교육 또한 중요한 기반이 될 수 있다.

나. 조사데이터 관리 절차별 관련 주체의 역할

조사데이터 관리 절차는 크게 1) 기획: 조사 수행 전 계획 수립 및 협의 단계, 2) 생산: 조사데이터 생산 및 관리 정보 작성, 3) 조사데이터 수집 절차 관리 및 데이터 축적, 4) 조사데이터의 제공 및 사후관리로 구분된다. 구체적으로 기획 단계는 기획과 계획으로, 생산 단계는 생산과 종료로, 관리 단계는 수집과 축적으로, 활용 단계는 활용과 현행화로 세분화된다. 이에 대한 구체적 내용은 아래와 같다.

먼저 기획 단계(1-1)에서 생성자는 조사계획 및 제반 사항을 확인하고 개인정보 및 민감정보 등 주요 사안에 대하여 고려하여 조사계획을 수행한다. 관리자는 조사데이터 관리 절차를 연구진에게 안내하며, DMP작성 교육을 수행하고 필요에 따라 연구자의 조사계획 자문을 수행한다. 위원회에서는 연구윤리, 개인정보보호 공공데이터 관리 등의 기반을 검토한다.

계획 단계(1-2)에서 연구자는 실제 조사를 계획하고 이에 따라 DMP를 작성한다. 이때 데이터 관리의 방안 등을 관리위원회와의 협의를 통하여 검토한다. 관리자는 필요에 따라 조사설계를 지원하고 위원회 심의가 필요하거나 위원회 위원의 검토가 필요할 시 관련 행정처리를 추진한다. 위원회에서 조사계획 및 DMP에 대한 검토를 하며, 필요에 따라 보완요청을 할 수 있다. DMP의 내용에 따라 데이터의 제공 수준을 결정할 수 있으며, 연구자의 의견을 적극 반영하되 위원회의 권고사항이 반영될 수 있도록 관리한다.

생산 단계(2-1)에서 생성자는 조사데이터의 생성 및 분석과 더불어 기존 계획의 변경 사항 등을 검토하고 이에 대한 기록을 남겨 기록 단계에서 적절한 검토가 이루어질 수 있도록 대비해야 한다. 관리자는 데이터

생산과 관련된 자문과 지침 내용에 대한 안내, 관련 절차 추진 과정의 추적 등을 지원한다. 위원회는 조사과정 중 논의가 필요한 연구 윤리 사항이나 개인정보보호 관련 문제 등에 대응한다.

종료(2-2)단계에서 생성자는 데이터와 메타데이터를 제출할 수 있는 수준으로 가공해야 한다. 이때 조사수집과 분석과정, 제공용 데이터 작성 과정에서 수정된 사항들이 기록되어야 하며, 데이터와 관련된 각종 정보(조사개요)와 자료(지침상 조사자료에 해당되는 모든 자료)를 정리한다. 관리자는 데이터 관리 관련 양식의 수집·검토를 수행하고, 상황에 따라 이에 대한 보완을 요청할 수 있다. 위원회에서는 조사 종료에 따른 보고 사항을 검토하고 사전에 작성된 DMP와 조사 종료 상황 간의 차이를 확인한다. 그 외에도 조사자료 관리위원회의 의견을 바탕으로 데이터의 품질 관리 관련 사항이 제안될 수 있다.

수집(3-1)단계에서 데이터 생성자는 데이터 제공 등급에 대한 자체적인 고려를 하며, 제공 가능한 데이터로 변환하기 위한 각종 전처리를 수행한다. 이러한 절차는 사전에 작성된 DMP를 준용하되, 조사결과의 특성을 반영하여 일부 변경할 수 있다. 관리자는 데이터 제공 등급 결과와 의견 등을 검토하고 데이터 전처리에 대한 상황을 점검한다. 필요시 추가적인 데이터 처리를 지원할 수 있으며, 외부 전문가의 자문이 필요할 시 이를 지원할 수 있다. 또한 조사결과를 바탕으로 데이터의 품질 수준을 검토하고 최종적으로 데이터 제공 등급을 결정한다.

추적(3-2)단계에서 관리자는 데이터 제공 등급을 확정하고 이를 목록화 하며, 원내에 공유(‘조사데이터 제공단계 세분화’ 참고)한다. 각 등급에 따라 데이터를 관리하고 데이터의 처리 및 수정 과정을 내검절차에 따라 기록 관리한다. 위원회는 데이터 제공 등급 부여에 대한 사항을 심의하고 이에 대한 모니터링의 적정성을 검토한다.

활용(4-1)단계에서 사용자는 데이터 목록을 바탕으로 접근 가능한 데이터의 범위를 확인하고 적정 절차에 따라 데이터를 신청한다. 이때 필수적인 정보를 작성·제출해야 하며, 공개가 제한된 자료의 경우에는 데이터 활용과 관련된 협의 절차를 진행한다. 관리자는 데이터 수요 및 활용에 대한 제반사항을 일원화하여 데이터 제공의 일관성을 확보한다. 만약 데이터 생성자가 데이터 수요를 대응해야 하는 경우가 발생할 때에는 관리자가 이를 중재하여야 한다. 위원회는 데이터활용에 대한 심의(단순 신청건은 포함되지 않음)를 한다. 데이터 요청에 대한 심의 후 데이터 처리 수준의 적정성(공개를 위한 별도의 처리가 필요한 경우)을 고려하여 데이터 제공을 승인한다.

현행화(4-2)단계에서 사용자가 가진 데이터의 현행화와 수정에 대한 권리가 보장될 수 있다. 사용자는 데이터 사용에 대한 의무를 지킬 필요가 있으며, 현행화 및 수정에 대한 요청과 데이터 사용 의무 서약 등은 관리자 중심으로 처리되어야 한다. 특히 관리자는 현행화와 에디팅 과정을 지속적으로 기록·관리할 필요가 있으며, 데이터 수정에 따르는 이슈를 대비해야 한다. 그 외에도 조사데이터의 엠바고 설정, 개인정보 보관기간의 만료 등을 고려하여 조사데이터를 관리한다. 이상의 과정에서 데이터 생성자는 데이터에 대한 사후 보완요청(현행화, 내용 수정 및 추가 등)을 받을 수 있다. 또한 생성자는 해당 데이터가 적절하게 활용되고 있는지를 관리자로부터 안내받을 권리를 가진다.

[그림 4-11] 주제별 조사데이터 관리 체계(안)



〈표 4-3〉 조사데이터 관리체계에 따른 각 주체별 역할 요약

절차	데이터 생성자 (원내 연구진)	데이터 관리자 (관련 전담 조직)	관련 위원회 (관리위원회 등)
기획·계획	<ul style="list-style-type: none"> 조사기획, 데이터 관리 계획 작성 조사데이터 관리 관련 세부사항 검토 	<ul style="list-style-type: none"> 조사데이터 관리체계 안내 및 절차 추진 조사관련 사항 지원 및 협의 	<ul style="list-style-type: none"> (필요시)데이터 관리 계획에 대한 자문 (필요시)데이터관리 계획에 대한 사전심의
생산·종료	<ul style="list-style-type: none"> 조사데이터 수집 및 분석 수행 조사데이터 및 관련 자료에 대한 제출 	<ul style="list-style-type: none"> 데이터 생산 관련 자문 데이터 및 관련 자료 수집, 보완 요청 	<ul style="list-style-type: none"> (발생 시)조사 과정 중 연구윤리 및 개인정보 보호 관련 이슈 대응
수집·축적	<ul style="list-style-type: none"> 데이터 제출 전 오류 수정 및 처리 이행 데이터관리계획 이행 및 제공 범위 설정 	<ul style="list-style-type: none"> 데이터 보관·제공 등급 검토 데이터 처리 결과 재검토 및 수정 요청 데이터 목록화, 내검기록, 위원회 심의안건 제출 	<ul style="list-style-type: none"> (안전)심의 안전에 대한 검토 및 의견 제출 조사데이터 관리 현황 등에 대한 모니터링
활용·현행화	<ul style="list-style-type: none"> 데이터 현행화 및 수정 등에 대한 대응 	<ul style="list-style-type: none"> 데이터 제공 지원 데이터 현행화 및 내검 등에 대한 지원 데이터 연계 등 활용 방안 활성화 	<ul style="list-style-type: none"> (요청 시)데이터 제공 관련 심의 안전 처리 조사데이터 현행화 및 수요자 대응 관련 점검

자료: 저자 작성

다. 조사데이터 관리 조직 구성

한국보건사회연구원 조사데이터의 체계적 관리를 위한 조직 구성(안)은 아래 그림과 같다. 한국보건사회연구원 조사데이터의 관리는 ‘조사데이터 관리 TFT’를 중심으로 이루어지며, 해당 TF의 책임은 조사데이터 관리 책임자가 담당하게 된다. TF구성은 크게 ‘조사데이터 관리실무’와 ‘조사자료 관리위원회’로 구분된다.

‘조사데이터 관리실무’는 조사데이터 관리를 위한 절차, 데이터의 이관 및 처리, 제공용 데이터의 관리 및 활용 등의 역할을 수행한다. 먼저 ‘조사데이터 관리 담당자’는 조사자료 관리지침에 따른 조사자료 관련 절차를 전반적으로 관리한다. 즉, 데이터관리계획 및 제출용 데이터의 수집, 관련 양식 접수 및 수정 등의 역할을 수행하며, 조사자료 관리위원회의 운영과 관련된 실무(가칭 조사자료 관리위원회 간사)를 담당한다. 해당 업무는 1인 이상이 전담할 필요가 있으며, 데이터 관리 담당 역할과 위원회 간사 역할은 분리될 수 있다.

‘조사데이터 품질관리 및 처리’는 조사데이터의 품질에 대한 검토와 제공용 데이터 생성을 위한 개인정보 및 민감정보 처리를 지원하는 역할을 수행한다. 또한, 이관 완료된 데이터 가운데 데이터의 현행화가 필요한 경우나 데이터 생성자 또는 이용자의 요청에 따라 데이터의 수정이 필요한 경우 등에 대응하여 데이터 내검절차를 추진·관리하게 된다. 해당 업무는 원내 데이터 관리 부서 내 1인 이상이 담당하며, 별도의 조사조직이 구성되어 있는 데이터(패널조사, 승인통계 등)에 한하여 해당 조직 내 1인이 지원할 수 있다.

‘조사데이터 보관 및 제공’은 각종 데이터 관리 정보시스템(보건복지데이터포털 등) 담당자 혹은 공공데이터 제공 등의 업무를 담당하는 인력이

전담한다. 특히 데이터 제공 인프라를 통하여 수집된 데이터 수요자의 요청사항(데이터 이용 문의, 데이터 제공 요청, 데이터 수정 등)을 대응하는 역할을 전담한다. 데이터 수요에 대한 대응 체계를 일원화하되 관리실무 조직 내 전담 인력의 협조를 바탕으로 대응 결과를 작성한다.

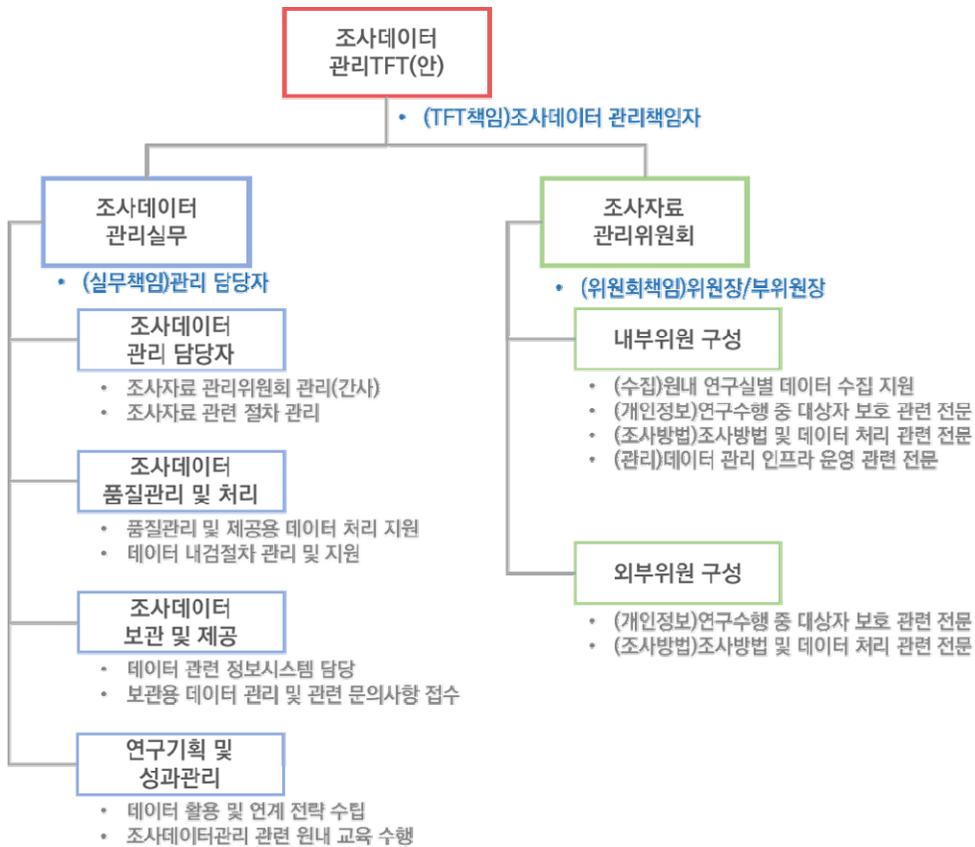
‘연구기획 및 성과관리’는 연구원 내 데이터 활용 및 연계 전략 등을 수립하며, 연구원 내 관련 교육 및 안내가 필요한 경우에 이를 지원한다. 해당 역할은 연구원 내 유사 업무를 수행하는 인력이 겸직할 수 있다. 또한, 조직 구성에 대한 우선순위가 상대적으로 낮으므로 필요에 따라 그 역할을 조정할 수 있다.

‘조사자료 관리위원회’는 관련 지침에 따라 내외부 위원을 구성하게 되며, 이때 조사관리위원회 간사를 두어 조사관리 위원회 구성 및 심의 안건 작성, 심의결과에 대한 반영 등의 실무를 수행할 수 있도록 한다. 조사자료 관리위원회의 구성은 현행 지침에 근거를 두되 아래 그림과 같은 기능을 수행할 수 있도록 전문위원을 배치할 필요가 있다. 첫째, 각 연구실·단별 조사데이터의 수집이 원활하게 이루어질 수 있도록 이를 독려하고 각 연구분야별 이슈에 대하여 검토할 수 있는 위원이 필요하다. 둘째, 조사데이터의 제공 여부를 검토할 시 개인정보보호와 민감정보의 처리 등에 대한 의견을 제시할 수 있는 전문위원이 1인 이상 배치되어야 한다. 셋째, 인간대상 조사에 대한 설계나 조사방법, 데이터의 처리 등에 대한 검토가 가능한 전문위원이 1인 이상 배치되어야 한다. 넷째, 데이터의 제공 시 발생할 수 있는 인적·물적 자원의 문제나 법·제도적 사항 등에 대한 검토가 가능한 전문위원이 1인 이상 배치될 필요가 있다.

한편, 외부위원의 경우에는 개인정보보호 및 민감정보의 처리와 관련된 전문위원, 조사방법 및 데이터 처리 관련 전문위원을 외부위원 구성인원에 맞게 배치하여야 한다. 다만, 내부위원의 구성을 고려하여 필요한

전문위원의 영역을 설정할 수 있다. 예를 들어 내부위원 가운데 조사방법 관련 전문위원의 구성이 상대적으로 적은 경우에 외부위원 전체를 해당 전문위원으로 배치할 수 있다.

[그림 4-12] 한국보건사회연구원 조사데이터 관리 조직 구성(안)



주: 조직 구성 및 역할의 내용은 가안이며, 상황에 따라 수정·적용될 수 있음.
 자료: 저자 작성

5. 조사데이터의 품질관리 및 내검절차

가. 조사데이터의 품질관리

데이터의 품질은 데이터 제공에 상당한 영향을 미친다. 따라서 조사데이터의 품질관리를 통하여 조사데이터의 신뢰성 및 타당성, 적정성 등을 확보하여야 한다. 이는 조사데이터의 제공 가능 범위를 확대하고 데이터의 신뢰성을 제고하는 데 기여할 수 있다.

데이터의 품질이란 데이터의 최신성, 정확성, 상호연계성을 확보하여 사용자에게 유용한 가치를 줄 수 있는 수준을 일컬으며, 품질관리는 품질을 확보하기 위한 목표설정, 진단 및 개선 등 일련의 과정을 말한다.

데이터의 품질관리는 데이터의 생성에서부터 활용까지 전 단계에 걸쳐 이루어질 수 있다. 계획 단계에서는 품질관리에 대한 내부규정 마련, 품질관리 조직 구성 등을 계획한다. 구축단계에서는 데이터의 점검과 구조의 확인, 데이터 산출물의 품질 관리, 연계데이터의 정합성 유지 등을 추진하는 단계이다. 운영 단계는 공공기관이 보유하고 있는 데이터를 관리하는 단계에서 준수해야 할 품질 진단 노력 및 산출물의 현행화 등을 수행하는 단계이다. 마지막으로 활용 단계는 품질 오류 신고나 차기 품질계획의 반영 등의 개선사항을 관리하는 단계이다.

연구책임자는 조사데이터의 생성·수집·관리·활용의 전 단계에서 데이터의 품질이 충분히 확보될 수 있는지를 자체적으로 점검(품질진단 방법)한다. 관리위원회는 자체적인 품질관리가 어려운 경우나 품질관리와 관련된 주요 현안이 발생할 때 이에 대한 심의·결정을 지원할 수 있다.

〈표 4-4〉 통계 데이터 품질진단 차원

진단 차원	개념	내용
관련성 (Relevance)	이용자 요구에 부합되는 정도	<ul style="list-style-type: none"> 통계의 목적을 명확히 설정하고 있는지, 이용자를 파악하고 있는지, 전문가 자문회의나 이용자 만족도 조사 등을 통하여 지속적으로 이용자의 요구를 파악하고 통계에 반영하고 있는지 등을 점검
정확성 (Accuracy)	응답단위에서 주로 발생하는 구조적 원인에 해당하는 오류를 검출하고 처리	<ul style="list-style-type: none"> 표본설계, 표본오차, 비표본오차, 자료 수집방법, 면접소요시간, 임퓨테이션 절차, 마이크로데이터와 공표자료 일치율 여부 등에 대한 상세한 정보를 기술
시의성 및 정시성 (Timeliness and Punctuality)	시의성은 작성기준시점과 결과공표시점 간의 차이를 나타내는 통계의 현실 반영도와 관련된 개념, 정시성은 예고된 공표시기를 정확히 준수하는가에 대한 개념	<ul style="list-style-type: none"> 통계작성 주기, 평균 및 최대 공표, 소요기간(작성기준시점에서 공표일까지 소요기간), 평균 및 최대 공표, 지연기간(공표예정일과 실제공표일 간의 차이), 공표지연 사유
비교성 및 일관성 (Comparability and Coherence)	비교성은 시간적 및 공간적으로 자료가 비교 가능한 정도 일관성은 동일한 경제· 사회현상에 관해 작성된 다른 통계자료와의 유사 또는 근접한 정도	<ul style="list-style-type: none"> 비교성은 통계에서 사용한 개념, 분류, 기준 등이 유사하여 비교 가능한지를 통하여 확인 일관성은 통계 간 결과가 유사한지를 통하여 확인
접근성 및 명확성 (Accessibility and Clarity)	접근성은 이용자가 얼마나 쉽게 통계자료에 접근할 수 있는가 하는 물리적 조건, 명확성은 이용자의 통계자료 이용 편의성 제공 정도	<ul style="list-style-type: none"> 통계자료의 DB화, 간행물 및 보도자료 홈페이지 게시, 모바일 메시지 등 통계작성 과정, 자료 이용방법, 마이크로데이터 이용방법, 적절한 메타자료(주석, 설명, 문서 등) 및 품질정보 등을 제공

주: 자료 내 내용을 표로 요약 작성
 자료: 통계청(2022). 국가통계품질관리 매뉴얼. 통계청.

[그림 4-13] 공공데이터의 품질진단 방법

측정 기법		측정기법 설명
프로파일링	값	<ul style="list-style-type: none"> 데이터 값의 유효성, 정확성 등 데이터 값 자체 오류를 분석하는 방법 - 컬럼분석, 날짜분석, 패턴분석, 코드분석 등을 통해 데이터 값의 정확성을 중심으로 진단
	구조	<ul style="list-style-type: none"> 논리적 데이터 구조의 오류로 인한 일관성, 정합성 등을 확보하지 못하는 결함을 분석하고 진단하는 방법 - 표준화 수준(코드, 도메인 등), 테이블 구조, 정규화 수준, 컬럼 및 관계 정의 등 데이터의 구조적 결함 측정
체크리스트 (설문, 인터뷰)		<ul style="list-style-type: none"> 전반적인 데이터 품질관리 수준과 지표별 데이터 품질 수준을 체크리스트(설문 또는 인터뷰)를 통해 진단하는 방법
업무규칙 진단		<ul style="list-style-type: none"> 법, 규정에 정의된 업무기준(산출식)에 근거하여 데이터가 관리되고 있는지를 진단하는 방법 - 업무규칙(BR; Business Rule)을 준수하고 있는지에 관한 측정 스크립트(SQL 등)를 실행하여 오류 값을 추출
비정형 실측		<ul style="list-style-type: none"> 문서, 이미지, 동영상 등 정형화되어 있지 않은 정보를 사람이 직접 확인(실측)을 통하여 오류 여부를 진단하는 방법 - 별도 도구 없이 직접 정보를 조회하거나 해당 문서를 수기로 확인 등

자료: 공공데이터활용지원센터(2021b). 공공데이터 제공·관리 실무 매뉴얼

나. 조사데이터의 내검 절차

데이터 내검(에디팅)이란 해당 자료 내에 존재하는 각종 오류를 수정하는 일련의 데이터처리 과정을 의미한다(이의규, 2018). 조사데이터의 경우, 데이터의 보관 및 제공 이전에 데이터에 대한 에디팅이 완료되어야 하지만, 데이터의 활용 및 제공 과정 속에서 데이터가 가진 오류들이 발견되는 경우가 발생한다. 이를 대비하고자 데이터의 완결성을 제고하기 위한 내검 절차가 마련되어야 한다.

조사데이터의 내검 절차가 구성될 경우, 데이터의 완결성과 신뢰성이 제고된다. 즉, 내검 과정을 통하여 데이터의 오류가 점차 줄어들고 이에 따라 데이터에 대한 신뢰도가 증가하게 된다. 한편, 내검 절차는 조사데이터 공개의 가능성을 확대하는데 기여한다. 연구책임자는 조사데이터의 제공 시 발생할 수 있는 데이터의 오류로 인하여 공개 여부에 부담을 갖게 된다. 만약 내검 절차가 안정적으로 추진된다면, 데이터 제공에 대한 소극적 태도가 다소 완화될 가능성이 있다.

데이터의 내검은 검토, 선택, 수정의 과정을 거친다. 먼저 검토단계에서는 데이터의 논리적 일관성 등 데이터의 유효성을 검토하게 된다. 또한 데이터를 통한 추정치의 타당성, 체계적 오류에 대한 확인, 등 데이터의 타당성을 검토한다. 그 외에도 통계 생산을 위한 단위의 적정성을 검토한다.

선택 단계에서는 단위·집계 수준·변수 등에 대한 검토가 이루어지며, 수정 단계에서는 변수의 체계적·비체계적 오류 수정, 결측 및 논리적 오류값 수정, 단위 및 집계 수준 수정 등이 이루어진다.

〈표 4-5〉 데이터 에디팅 기능

기능 유형		기능 예시
검토	데이터 유효성 검토	<ul style="list-style-type: none"> • 명백한 오류의 검토 • 관측값 조합의 논리적 일관성 평가 • 데이터 특성 검토
	데이터 타당성 검토	<ul style="list-style-type: none"> • 데이터 값 또는 그 조합의 타당성 검토 • 집계수준 추정치의 타당성 검토 • 의심스러운 집계치의 검토 및 식별 • 체계적 오류의 검토와 식별 • 결합 단위에 대한 집계수준의 검토
	단위 검토	<ul style="list-style-type: none"> • 적합한 단위 검토 • 부적합한 단위 검토 • 영향력이 큰 또는 이상 단위에 대한 스코어 검토 • 단위의 개별레코드 수준의 일관성 검토
선택	단위 선택	<ul style="list-style-type: none"> • 적합한 단위 선택 • 쌍방향 처리/비 쌍방향처리/비 수정 단위 선택 • 영향이 큰 오류에 의해 영향 받는 단위 선택 • 가중치 조정에 의해 처리되어야 하는 이상 단위 선택 • 단위의 구조에 의한 선택 • 집계수준 검토에 의한 단위 선택
	변수 선택	<ul style="list-style-type: none"> • 명백한 오류를 갖는 변수 선택 • 단위 특징에서 오류를 갖는 변수 선택 • 특정한 대체방법 처리를 위한 변수 선택 • 수내검을 위한 영향력 있고 이상한 값 선택 • 비일관된 값들 중에서 오류의 변수(군) 위치 포착 • 각 단위에 대해 오류 변수 위치 포착
수정	변수 수정	<ul style="list-style-type: none"> • 명백한 오류 수정 • 체계적 오류 수정 • 단위 오류 수정 • 위치 포착된 오류의 대체 • 결측 및 무시된(오류) 값의 대체 • 비일관성 수정
	단위 수정	<ul style="list-style-type: none"> • 단위의 처리 • 통계단위 생성 • 단위에 대한 다양한 유형의 매칭 • 단위 연계 결손 처리

자료: 이의규(2018). 통계개발원 2018년 상반기 연구보고서 제1권. 표준 자료처리 모델 GSDEMs의 소개 및 시사점(p.49-87). 통계진흥원.

데이터의 내검은 아래 표와 같이 세부적으로 구분되며, 아래 그림과 같은 절차를 통하여 데이터의 에디팅이 추진될 수 있다.

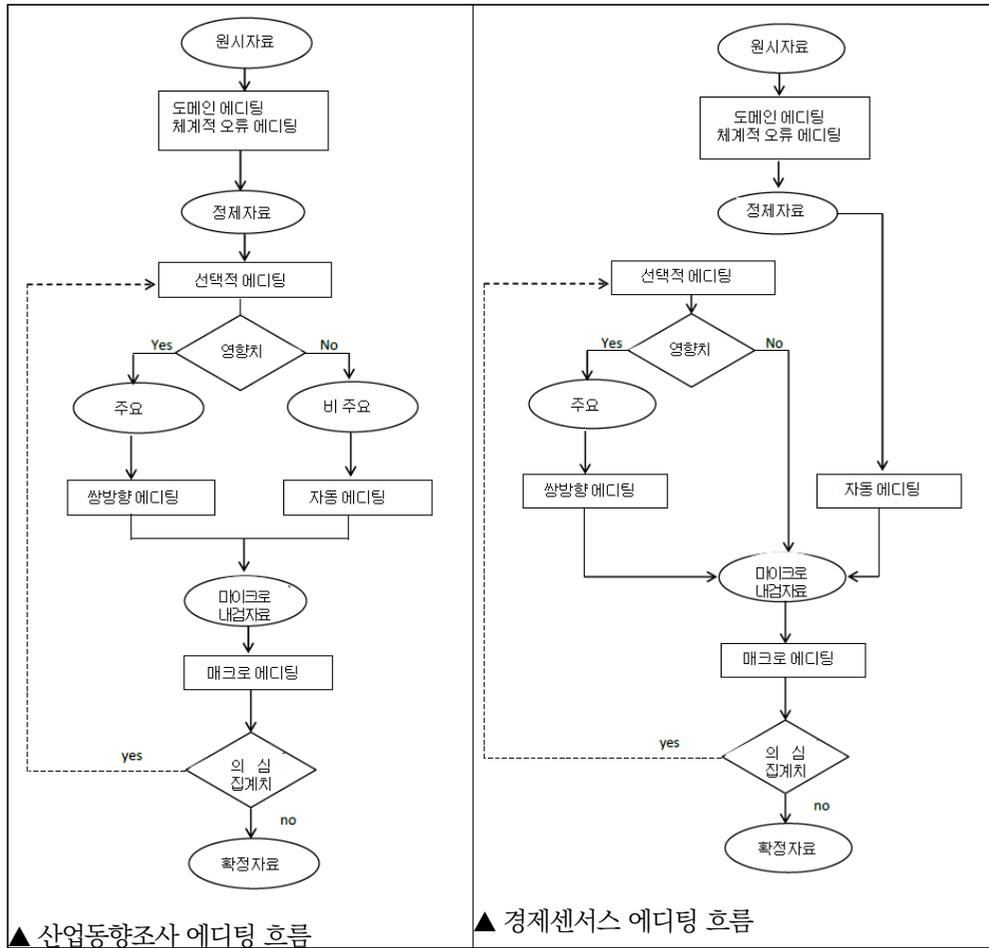
〈표 4-6〉 데이터 에디팅 구분 및 기능

에디팅 단계	개요	기능 및 방법
도메인 에디팅 (Domain Editing)	적격한 단위, 구분변수의 확인 및 선택과 같은 목표모집단과 변수를 정의하는 구조적 정보 객체를 검토	<ul style="list-style-type: none"> • 적격한 단위에 대한 검토와 선택 • 데이터 속성(분류, 조직형태)의 검토와 선택, 수정
체계적 오류 에디팅 (Editing Systematic Errors)	응답단위에서 주로 발생하는 구조적 원인에 해당하는 오류를 검출하고 처리	<ul style="list-style-type: none"> • 명백한 오류의 검토, 선택, 수정 • 체계적 오류의 검토 • 체계적 오류영향 단위 식별 • 연역적 수정, 모형기반 대체
선택적 에디팅 (Selective Editing)	에디팅되지 않을 때 집계치에 큰 영향을 주는 자료를 선택하여 심층 재조사하는 방식	<ul style="list-style-type: none"> • 영향력이 높은 오류에 영향받는 단위 식별 (스코어 계산) • 쌍방향 처리를 위한 단위 선택, 비쌍방향 처리 단위 선택 등(고정 한계값으로 선택)
쌍방향 에디팅 (Interactive Editing)	전산적으로 오류를 검출하고 이에 대한 처리는 사람이 하는 방식	<ul style="list-style-type: none"> • 재접촉, 조사표 검사 등
자동 에디팅 (Automatic Editing)	전적으로 자동화된 방식으로 자료 파일에서 오류 및 결측치를 자동으로 찾고 처리	<ul style="list-style-type: none"> • 내검규칙군에 대해 자료의 일관성 확인 • 각 단위에서 오류변수 위치 포착 • 포착된 오류의 대체 • 결측치 대체
매크로 에디팅 (Macro Editing)	관심영역의 집계치를 비교·검토하는 방법으로 전년 집계치와 같은 참조값과 비교하여 확인하는 방식	<ul style="list-style-type: none"> • 의심스러운 집계값의 검토 및 식별 • 데이터셋 총합비교, 외부자료의 총합 비교, 과거결과 비교

자료: 이익규(2018). 통계개발원 2018년 상반기 연구보고서 제1권. 표준 자료처리 모델 GSDEMs의 소개 및 시사점(p.49-87). 통계진흥원.

주: 자료 내 표 내용을 일부 통합 수정

[그림 4-14] 데이터 에디팅 절차 예시



자료: 이의규(2018). 통계개발원 2018년 상반기 연구보고서 제1권. 표준 자료처리 모델 GSDEMs의 소개 및 시사점(p.49-87). 통계진흥원.

6. 조사데이터 관리체계 구성 시 고려사항

조사데이터 관리체계 구성 시 아래의 사항을 추가적으로 고려해야 한다. 첫째, 조사별 관리코드를 부여할 필요가 있다. 연구 내 복수의 조사데이터가 생성되는 경우 이는 별도 관리가 되어야 한다. 현재는 하나의 연구에서 성격이 다른 복수의 데이터가 수집될 때 하나의 데이터로 관리되는 사례가 확인되고 있다. 특히나 양적 조사와 질적 조사의 데이터 관리는 별도로 이루어져야 하기 때문에 연구단위의 데이터 관리에서 조사 단위의 데이터 관리로 전환될 필요가 있다.

둘째 형태별 관리방안의 세분화가 필요하다. 모든 데이터는 개인정보 및 민감정보를 포함할 수 있기 때문에 개인정보와 민감정보의 처리를 통하여 연구목적 문항의 분리가 필요하다. 한국사회과학자료원의 사례를 살펴보면, 질적 자료의 경우에도 개인정보 및 민감정보, 집단을 특정할 수 있는 정보 등을 처리하여 제공하고 있다. 현 시점에서 질적 자료의 공개는 기술적으로 어려울 수 있으나 향후 개인정보 보호 처리 기능이 강화될 경우에는 질적 자료 내에서의 개인정보와 연구정보 분리가 이루어질 수 있을 것이다.

셋째, 조사데이터의 목록 공개가 필수적이다. 원내 데이터에 대한 전반적인 정보는 원 구성원 모두가 동일하게 접근할 수 있어야 한다. 만약 조사관리와 관련된 정보가 공유된다면, 조사데이터 추진을 위한 다수의 참고자료를 확보할 수 있다. 유사 중복 조사를 사전에 검토할 수 있으며, 조사 기획의 방안들도 연구원 내부에서 공유될 수 있다. 뿐만 아니라 연구원 내부 활용의 활성화를 통하여 안전한 정보 환경에서 데이터의 가치를 제고할 수 있게 된다.

넷째, 연구방법론에 대한 세부적인 정보수집이 필요하다. 현재 양적,

질적 자료로 데이터의 유형이 구분되어 있다. 현재 전문가 자문을 통한 자료수집(녹취록 등)과 델파이, FGI를 통한 자료수집이 명확하게 분류되지 않는 사례가 발생하고 있다. 또한 전문가 대상의 우선순위 분석이나 중요도 분석 등도 데이터 유형 구분이 쉽지 않다. 이와 같이 조사 수행 시 활용될 수 있는 분석 방법들이 사전에 검토되어 조사데이터의 현황을 실시간으로 보여줄 수 있어야 한다.

다섯째, 정보공개 관련 대응력 강화가 요구된다. 데이터 보안 체계 및 연구자 보호체계가 강화될 필요가 있다. 데이터 관리 체계가 강화될 경우에 연구자의 자율성이 제약을 받을 수 있다. 그러나 데이터의 공개/비공개 기준의 설정과 공식화, 민감정보에 대한 기관 차원의 관리체계 도입 등은 데이터를 생산한 연구자의 책임소재를 분산시키는 기능을 한다. 예를 들어 기관에서 생산된 조사데이터가 기관 내 관리체계 밖에 있는 경우에는 연구책임자가 정보공개 요청에 대한 대응, 조사참여자의 정보 동의 철회 대응, 연구 외 목적의 데이터 활용, 데이터 오류에 대한 대응 등에 상당한 부담을 가질 수밖에 없다. 따라서 연구원 중심의 조사데이터 관리체계가 마련되어 연구자에 대한 보호가 보다 강화될 수 있도록 지원해야 한다.

여섯째, 조사데이터의 공개는 법적·제도적 배경 하의 필수적인 절차임이 모든 구성원에게 공유되어야 한다. 제2장에서 살펴본 바와 같이 연구원은 공공데이터의 공개, 개인정보의 보호, 통계생산기능 등의 복합적 책무를 가지고 있다. 연구원의 구성원은 연구원의 책무를 이행할 의무를 가지고 있으므로 조사데이터 관리 체계를 통한 정보공개 활성화는 부차적인 업무로 치부될 수 없다. 한편, 이러한 조직 문화를 확산시키기 위한 데이터 제공 우수사례에 대한 지원 및 홍보, 활용도 제고를 위한 학술대회 추진, 조사수행의 효율성 강화를 위한 유사 중복 조사 모니터링 등이 도입될 필요가 있다.

제3절 조사데이터의 활용 활성화 방안

1. 조사데이터의 내부 활용 방안 검토

조사데이터의 내부 활용을 강화하기 위한 전략으로 1) 조사자료 관리 지침 내 수집 정보에 대한 공개, 2) 데이터 활용 목적에 따른 접근 권한 세분화, 3) 데이터 활용에 절차의 공식화 및 개방화를 검토해 볼 수 있다.

먼저, 조사자료 관리지침에 따라 조사데이터 생성 및 이관 시에 각종 정보가 수집·관리 된다. 대표적으로 지침 내 양식(데이터관리 계획, 조사수행의뢰서, 조사자료 제출 신청서 등)의 작성을 통하여 관련 정보가 수집된다. 조사자료 관리와 관련된 정보의 수집을 촉진하고 조사데이터 수집 계획에 대한 자체적 점검이 이루어지기 위해서는 이러한 수집 정보가 신속하게 공유되어야 한다.

조사자료 관리지침 내 수집 정보의 공유 우선순위를 설정하면 다음과 같다. (1순위) 조사명, 조사대상 정보, 조사방법 정보(표집 방안 포함). 이 정보는 주기적으로 생성되는 조사데이터의 수집 초기에 계획 단계에서 설정되어야 하는 정보로, 조사데이터의 타당성과 신뢰성에 영향을 미칠 수 있는 주요 항목이다. 조사데이터 수집을 계획하는 경우에 유사 조사데이터의 수집 계획 정보가 연구원 내부적으로 충분히 공유될 시에 기존 사례를 참고하여 보다 신뢰성있는 조사데이터 생성 계획이 수립될 수 있다. (2순위) 조사예산, 기간, 실제 조사가 완료된 실적. 조사데이터 수집 계획 시에 기존의 유사 조사가 조사계획과 실행 간에 발생한 문제들을 어떻게 대처하였는지를 내부 정보 공개를 통하여 확인할 필요가 있다. 만약 조사계획과 조사 결과 간의 불일치가 발생할 경우, 이를 보완할 수 있는 방안 또는 사전에 대비할 수 있는 방안 등이 검토되어 이후 조사데이터의 오류

가능성이 최소화되거나 사전에 통제될 수 있도록 해야 한다. (3순위) 조사결과 이관 정보, 조사데이터 공개 범위 설정, 조사데이터의 처리 결과, 조사데이터의 이관 이전에 조사데이터 내 개인정보 및 민감정보의 처리가 조사데이터 생산자의 역할로 주어진다. 조사데이터의 처리에 대한 세부 사항을 구체적으로 설정하는데 어려움이 있으므로 기존 데이터 처리 실적 등을 적극 공유하여 조사데이터 공개의 수준과 처리의 기준선 등이 점진적으로 설정될 수 있도록 지원해야 한다. 특히 민감정보의 경우, 연구책임자의 검토 결과와 데이터 관리자와의 논의 결과 또는 위원회 심의 결과 등이 공유될 경우, 기존 사례를 참고하여 데이터 처리와 공개범위 설정 등에 대한 사항이 이관 전에 충분히 이루어질 수 있다. 이는 데이터의 신뢰성 제고 외에도 데이터 관리 체계의 과부하를 완화하는 데 기여할 수 있는 방안이다.

둘째, 데이터 활용 목적에 따른 접근 권한을 세분화하여 내부 연구진의 데이터 활용 가능성을 제고할 필요가 있다. 현재 데이터의 접근 권한은 양분(모두 접근 가능한 데이터, 생성자 및 관리자만 접근 가능한 데이터) 되어 있다. 그러나 데이터와 관련된 정보를 알지 못하는 경우, 데이터 활용 여부에 대한 검토 자체가 불가능하기 때문에 생성자 및 관리자만 접근 가능한 데이터는 원내 활용의 가능성이 거의 없다고 볼 수 있다. 따라서 조사자료 관리지침 내 수집 정보의 공유와 더불어 제한된 범위에서의 데이터 접근 권한이 부여될 수 있도록 권한의 세분화가 요구된다.

예를 들어 공유된 조사데이터의 정보를 확인하여 특정 조사데이터의 활용 여부를 판단해야 할 경우, 데이터 생성자 또는 데이터 관리자의 감독 하에 데이터 접근 신청자가 데이터의 일부(개인정보 및 민감정보가 삭제된 데이터, 원 데이터가 삭제된 데이터 등, 조사데이터의 형태를 파악하기 위한 용도의 소수 케이스 확인)를 검토할 수 있도록 제한적 접근 권

한을 부여한다. 제한된 접근 권한을 일시적으로 부여받아 데이터의 일부를 검토할 경우, 데이터 활용 여부를 결정하거나 활용 범위를 설정하는데 필요한 정보를 습득할 수 있다. 이러한 과정을 통하여 데이터의 내부적 활용 가능성을 제고할 수 있다.

셋째, 데이터 활용에 절차의 공식화 및 개방화가 필요하다. 현 지침에 따르면 조사데이터의 내부적 활용은 연구책임자의 연구수행에 국한되어 있다. 그러나 조사데이터의 내부적 활용은 개별 연구자의 연구성과 확산, 관련 정책자료 작성, 해당 분야의 동향 파악 등 폭 넓게 이루어질 수 있다. 따라서 데이터 활용의 목적이 데이터 생산자의 의도 또는 연구원의 정보 공유 방침에 벗어나지 않는다면 활용 방법에 대한 제한을 최소화할 필요가 있다. 단 조사데이터 관리체계에 따라 내부적 데이터 활용 절차를 공식화하고 그 과정 및 결과를 투명화할 필요가 있다. 특히 내부적인 활용의 성과가 원내에 적극 공유되어 데이터 가치의 창출과 데이터 활용의 효율성 등을 강조할 필요가 있다. 이는 데이터 생산 및 관리에 따르는 비효율성을 완화하는 데 도움을 줄 수 있으며, 한정된 자원에서 다양한 데이터를 활용할 수 있는 토대를 마련할 수 있는 방안이라 볼 수 있다.

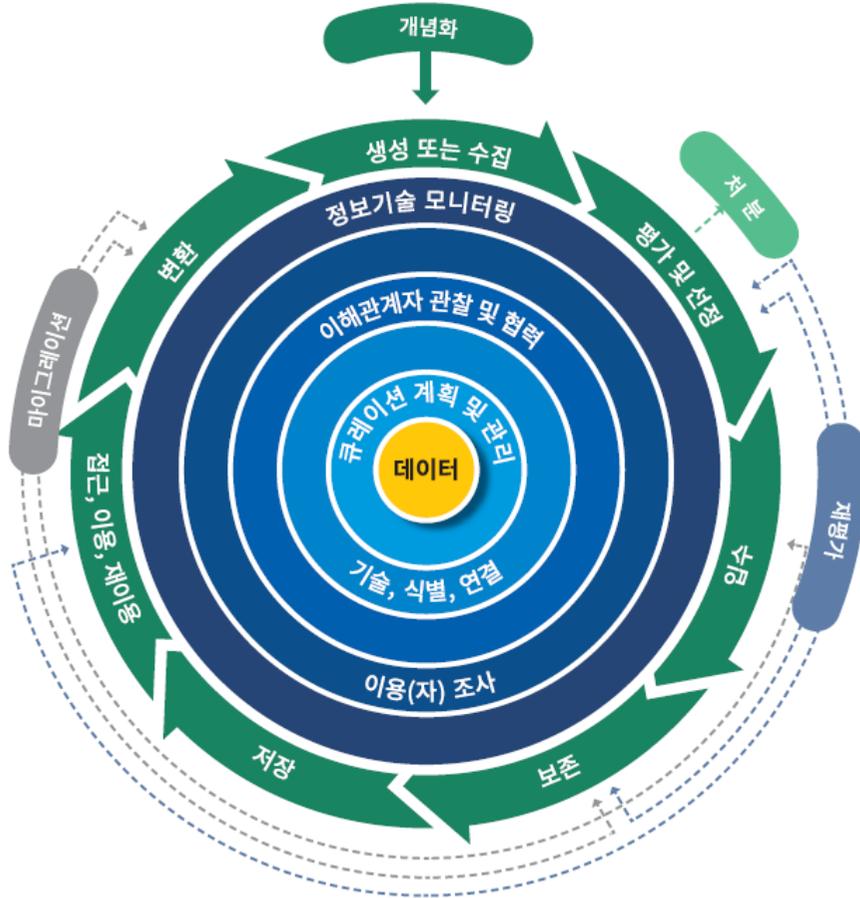
2. 조사데이터 큐레이션 체계 도입

데이터의 체계적 관리 체계에 따라 보건복지 조사데이터가 지속적으로 축적될 경우, 방대한 데이터의 활용을 촉진하기 위한 전략이 요구된다. 이에 따라 문헌정보 분야에서 주목받고 있는 “데이터 큐레이션”이 보건복지 조사데이터의 제공 플랫폼 내 도입이 검토될 수 있다.

데이터 큐레이션이란, 데이터의 생애주기(수집, 관리, 보존, 활용)에 따라 데이터를 적극적으로 관리·보존하고 그 가치를 부여하여 재이용될 수 있도록 하는 일련의 과정을 의미한다(이상헌, 2020; 이혜림, 2020). 원래 큐레이션은 미술관이나 박물관의 기획에 따라 우수한 작품 등을 전시하는 행위를 일컫는다. 이러한 의미가 데이터 분야에 확정되어 양질의 데이터가 정확하게 활용될 수 있도록 하여 그 가치를 제고하는 행위에도 큐레이션이란 용어가 활용되고 있다.

데이터 큐레이션 도입의 주요 목적은 1) 기존 데이터의 적극적인 관리를 통하여 연구결과를 보존하고, 2) 데이터 중복으로 인한 관리 및 활용 노력의 비효율성 감축시키며, 3) 데이터의 공유와 활용을 촉진하여 그 가치를 향상시키는 데에 있다(이혜진, 김재훈, 현미환, 2022).

[그림 4-15] 데이터 생애주기를 고려한 데이터 큐레이션 모델



자료: 이해림(2020). 디지털 큐레이션 가이드라인과 체크리스트. 한국과학기술정보연구원.

가. 데이터 큐레이션 도입을 위한 주요 고려사항

데이터 큐레이션 도입을 위한 주요 고려사항²⁶⁾은 운영 관리, 전주기 활동, 순차적 활동, 비정기적 활동으로 구분될 수 있다.

먼저, 데이터 큐레이션 도입을 위한 기관의 여건과 관련 역량들을 검토할 필요가 있다. 대표적으로 데이터 관리에 대한 계획, 관련 조직과 예산,

26) 「디지털 큐레이션 가이드라인과 체크리스트(2020)」 내 '제2장 디지털 큐레이션 가이드라인'의 내용을 중심으로 요약·기술함.

규정 등의 제도적 기반, 시설설비 등의 물리적 기반 등이 이에 해당된다. 조사데이터의 큐레이션을 위해서는 데이터의 수집과 관리에 투입되는 자원의 현황을 파악할 필요가 있다. 자동화된 프로세스를 도입하지 않는 경우에는 지속적인 인력투입이 불가피하다. 특히 조사데이터의 경우, 데이터의 수집과 관리, 제공을 위한 사후처리 등의 이슈가 발생되므로 인적자원의 투입이 필수적이다. 따라서 데이터를 관리하고 제공하기 위한 사전 작업에 어떠한 인적 자원이 투입될 수 있는지, 그 인적 자원의 활용이 지속성 및 연속성을 가질 수 있는지 검토해보아야 한다.

한편, 조사데이터의 큐레이션을 위한 일련의 제도적 기반이 마련될 필요가 있다. 한국보건사회연구원에서는 조사자료 관리지침을 통하여 조사데이터의 관리에 대한 제도적 근거를 마련하고 있다. 데이터 큐레이션을 통한 조사데이터 활용도 제고를 위해서는 기존의 지침에 데이터 관리 및 제공을 위한 업무를 추가적으로 명시할 필요가 있으며, 기존 위원회의 역할에도 데이터 큐레이션 관련 사항의 개입 및 심의 등의 권한을 부여하여야 한다.

둘째, 데이터 큐레이션을 위한 전주기 활동으로는 큐레이션 계획 수립, 데이터의 문서화 강화, 이해관계자 협력, 수요조사 등이 해당된다. ‘데이터 큐레이션 계획’은 기존 데이터를 어떻게 관리하고 활용할 것인지에 대한 기준을 마련하고 어떠한 일정에 따라 업데이트와 모니터링을 할 것인지 등의 활동을 포함한다. 특히 데이터 생애주기에 따른 체계적 관리 절차 마련이 선행되어야 한다. 이에 더 나아가 데이터 큐레이션 계획에서는 데이터의 관리 및 활용 등의 과정에서 발생할 수 있는 위기에 대한 관리가 추가된다. 예를 들어 데이터의 보존이 어려운 경우, 데이터 제공과 관련된 분쟁이 발생한 경우 등의 문제가 발생한 상황이 해당 계획에 반영될 필요가 있다. 이상의 문제가 발생할 때에 취할 수 있는 조치의 목록화나

대응의 절차, 유관기관 간의 협력 방안 등이 계획을 통하여 설정되어야 한다.

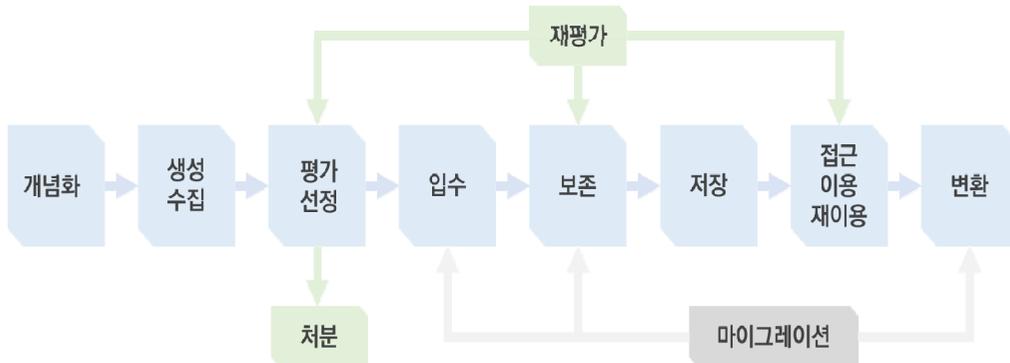
‘데이터의 문서화’는 메타데이터 작성이나 타 데이터와의 연결 등에 필요한 식별자 부여 등의 행위를 말한다. 이때 데이터에 대한 기본적인 정보 수집이 중요하며, 데이터 수요자가 쉽게 이해할 수 있도록 이용자 친화적 측면에서 작성 및 검토되어야 한다. 또한 데이터 큐레이션을 통하여 수요자의 목적에 맞게 데이터가 탐색·활용될 수 있도록 문서화된 데이터 정보가 적절하게 기입되었는지 확인하여야 한다.

한편, 데이터 관리 및 제공, 주요 데이터 이용 주체 등 이해관계자 간의 지속적인 협력체계를 구축할 필요가 있다. 이를 통하여 데이터 관리 및 제공을 효율적으로 처리하고, 다출처 데이터 간 연계를 촉진하며, 잠재적 수요층에 대한 대응력을 강화할 수 있다. 한편, 데이터 수요에 맞는 데이터 큐레이션 전략 수립과 수요 변화를 반영한 데이터 관리 방안 등을 모색하기 위하여 데이터 이용자 대상의 수요조사가 정기적으로 이루어질 필요가 있다. 주로 이용자 대상의 설문조사나 개별·집단 인터뷰가 이루어지며, 그 외에도 데이터 이용 기록 등에 대한 분석이나 관련 문헌 검토 등의 방법이 활용된다.

셋째, 데이터 큐레이션을 위한 순차적 활동으로 개념화, 생성·수집, 평가·선정, 보존·저장, 접근·이용·재이용, 변환의 과정을 거친다. 개념화 단계는 데이터의 생성과 수집 등에 대한 계획을 의미한다. 생성·수집 단계는 데이터의 문서화 과정을 의미한다. 데이터를 생성하는 경우에는 데이터의 제공과 보존을 위한 기준에 따라 데이터를 처리하는 과정을 거치게 된다. 또한 수집 또는 생성된 데이터에 대한 품질관리 검토가 해당 단계에서 이루어진다. 평가·선정 단계에서는 데이터의 선별이 이루어진다. 즉, 생성되는 모든 데이터가 동일한 수준으로 큐레이션이 적용될 수 없기

때문에 데이터 활용 및 관리의 측면에서 그 효율성을 따져 데이터를 선정해야 한다. 데이터 선정 시에 제공 기관의 데이터 제공 목적, 데이터의 고유성 및 신뢰성, 데이터의 중요도 및 활용 가능 수준, 법적 관리 등 관련 이슈 등이 사전에 고려되어야 한다. 보존·저장 단계에서는 데이터 문서화에 대한 검토, 데이터 클리닝, 관리에 필요한 각종 정보의 정리 등이 이루어진다. 접근·이용·재이용 단계에서는 데이터의 접근 가능성을 유지·제고하는 활동이 이루어짐과 동시에 데이터 관련 법적 이슈에 대한 대응과 통계 및 인증 절차 등이 검토된다. 이때에는 공공데이터 제공 관련 법률 등 관련 법령에 근거한 기준 마련이 필수적이다. 마지막으로 변환 단계에서는 데이터 활용의 편의성을 제고하거나 데이터의 장기적 저장을 위한 데이터 포맷의 변환 및 관리가 이루어진다. 또한 쿼리(query)문을 활용한 데이터의 서브셋(Subset) 형성도 이 단계에서 이루어진다.

[그림 4-16] 데이터 큐레이션의 순차적/비정기적 활동 관계



자료: 이혜립(2020). 디지털 큐레이션 가이드라인과 체크리스트. 한국과학기술정보연구원.

넷째, 데이터 큐레이션을 위한 비정기적 활동으로는 데이터 마이그레이션, 재평가, 처분 등이 있다. 데이터 마이그레이션(migration)이란 디지털 콘텐츠를 보존하고 이용자들이 탐색·이용할 수 있도록 변환 및 이전하는 과정을 의미한다. 마이그레이션 시 그에 대한 절차와 방법, 기준 등

이 설정되어야 하며 결과에 대한 기록이 지속적으로 관리되어야 한다. 원본의 정확성과 진본성을 유지하기 위한 품질 검증이 주기적으로 이루어질 필요가 있다. 재평가는 기존의 평가 결과(순차적 활동 중)에 대한 재평가를 의미하며, 이 결과를 통하여 재평가 시점에서의 데이터 활용 가치를 평가하여 큐레이션 될 것인지 처분할 것인지 등을 검토하게 된다. 처분은 순차적 활동 중 평가·선정 단계 혹은 비정기적 활동 중 재평가 단계의 결과이다. 데이터의 장기보존이나 큐레이션의 필요성이 떨어지는 경우 저장소 이전이나 폐기 등의 결정이 이루어진다.

나. 데이터 큐레이션 도입 시 관련 주체의 역할

한국보건사회연구원에서 생성·관리하고 있는 조사데이터의 활용을 제고하고자 데이터 큐레이션을 도입할 경우, 연구원 내 관련 주체별 역할이 부여될 필요가 있다. 각 주체별 역할에 대한 내용은 아래 표와 같다.

먼저, 데이터 생성자는 주요 조사 정보에 대한 정확한 정보를 입력하고 이에 대한 오류 등을 확인할 의무를 가진다. 데이터 관리 체계 내에서 데이터의 이관을 위한 주요 절차들을 이행해야 하며, 데이터 이관 전 개인 정보 및 민감정보 처리에 대한 일차적인 책임을 가진다. 조사데이터가 활용될 것으로 추정되는 범위를 검토하고 데이터의 저작권 등의 문제가 존재하는지 여부를 검토하여 그 결과를 데이터 관리자에게 전달한다. 데이터 큐레이션은 데이터의 수요에 맞게 선별하는 과정이 필요하기 때문에 데이터 생성자 자신의 주요 연구분야를 고려하여 효과적인 큐레이션 전략 등을 제안할 수 있다.

데이터 관리 실무 측면에서 조사데이터에 대한 기본적인 정보와 데이터에 대한 검증 등이 조사데이터 관리 체계에 따라 이루어진다. 이관된 데이터의 보관 또는 제공을 위한 사전 처리가 적절하게 이루어졌는지 확인하고 추가적인 검토가 필요할 시에 이를 위원회 의결사항으로 선정한다. 데이터 큐레이션을 위한 일련의 과정 중에 데이터의 변환 등이 발생할 경우 이에 대한 기록을 하고 지속 관리한다. 그 외에 데이터 큐레이션 정책 및 절차 도입의 전반적인 사항을 추진하고 점검한다.

데이터 관리 책임의 측면에서 데이터 큐레이션 정책의 방향을 설정하고 순차적 도입의 방법을 검토한다. 특히 위원회 차원에서 검토가 필요한 사항을 선별하고 위원회 결과에 대한 반응을 결정한다. 그 외에 데이터 큐레이션 추진과 관련된 주요 주체의 역할을 부여·관리·조정하는 역할을

수행한다.

조사자료 관리위원회에서는 데이터 큐레이션 도입 결과를 검토하고 조사데이터의 큐레이션 시에 발생하는 주요 현안에 대하여 심의한다. 특히 조사데이터와 관련된 내외부 전문가 자원을 활용하여 조사데이터의 활용도 제고 방안, 조사데이터 활용 시 발생할 수 있는 위기 대응, 조사데이터의 질적 관리 및 활용 수준 결정 등에 역할을 수행한다. 그 외에도 데이터 큐레이션 전략의 개선 방안 등에 대해 제안할 수 있다.

연구원 내 정보기술 영역에서는 데이터 관리 및 큐레이션 과정 중 데이터의 보안과 안전 관리를 위한 기술 도입을 지원한다. 데이터 큐레이션의 자동화를 추진할 경우에 이와 관련된 정보기술 지원이 요구되며, 데이터 마이그레이션 시에 필요한 하드웨어 및 소프트웨어 도입 등을 결정한다.

학술정보 영역에서는 조사데이터와 원내외의 관련 문헌 간 연결을 지원하고 국내외 연구기관의 네트워크를 활용하여 조사데이터와 관련된 정보의 접근성을 강화하고 유사 정책을 추진하는 기관 간의 상호 협력체계를 구축 및 관리한다. 대외홍보 영역에서는 조사데이터에 대한 큐레이션 도입을 홍보하고 큐레이션의 개선 방향을 설정하기 위한 수요조사의 데이터 축적을 지원한다. 그 외에도 데이터 제공과 재이용 촉진을 위한 대외 협력 방안을 모색한다.

그 외에도 데이터의 적극적 관리와 활용을 위하여 데이터 품질관리를 지원하고 데이터 큐레이션과 관련된 물적·인적 자원의 관리가 요구된다. 더욱이 연구원에서 생산된 조사데이터 가운데 재이용이 활발한 데이터를 중심으로 큐레이션 전략을 수립하고 변화되는 데이터 수요에 민감하게 반응하기 위한 노력이 지속될 필요가 있다.

〈표 4-7〉 조사데이터 큐레이션 도입을 위한 주체별 역할 설정

영역	역할	주체(예시)
데이터 생성자	<ul style="list-style-type: none"> • 주요 조사정보의 입력 및 확인 • 데이터 보관 및 제공을 위한 전처리 • 데이터 활용 범위 설정 및 소유권 사항 검토 • 데이터 활용 계획 및 큐레이션 등에 대한 자문 	연구책임 또는 조사설계
데이터 관리자 (실무)	<ul style="list-style-type: none"> • 주요 조사 정보에 대한 검증 및 데이터화 • 데이터 보관 및 제공의 처리 및 기록 • 데이터 큐레이션 과정에 대한 추진 및 관리 • 데이터 큐레이션 관련 심의 결과 반영 	조사데이터 관리 실무자
데이터 관리자 (책임)	<ul style="list-style-type: none"> • 데이터 큐레이션 정책 방향 설정 및 도입 • 조사자료 관리위원회 주요 심의 사항 결정 • 데이터 큐레이션 절차 추진에 대한 총괄 관리 • 데이터 큐레이션 추진 시 관련 주체의 역할 조정 	원내 데이터 관리 책임
조사자료 관리위원회	<ul style="list-style-type: none"> • 데이터 큐레이션 도입결과에 대한 검토 • 조사데이터 활용 관련 주요 이슈에 대한 심의 • 내외부 위원 및 외부 전문가를 활용한 의견 반영 및 개선사항 논의 	위원장 및 부위원장 중심
정보기술 영역	<ul style="list-style-type: none"> • 데이터 보안 및 안전 관리를 위한 기술 도입 • 자동화 도입 시 시스템의 관리 및 점검 • 데이터 마이그레이션 등 기술 지원 	각 부서 내 관련 업무 수행 인력 중심
학술정보 영역	<ul style="list-style-type: none"> • 조사데이터와 관련 문헌 간 연결 • 국내외 정보기관 네트워크 관리 및 협력 	
대외홍보 영역	<ul style="list-style-type: none"> • 데이터 큐레이션 도입의 홍보 • 데이터 큐레이션 개선을 위한 수요 검토 • 기타 타 기관 연계 방안 적용 등 	
기타 주체	<ul style="list-style-type: none"> • (정보통계영역)데이터 품질 관리 등 지원 • (행정·관리영역)데이터 큐레이션 도입 및 개선을 위한 제도적 사항 검토 • (패널·실태조사 수행영역)주요 활용 데이터에 대한 큐레이션 방안 제안 및 실적 관리 등 	영역별 관련 업무 참여 인력 중심

자료: 저자 작성

주: 위 내용은 저자의 제안사항으로, 데이터 큐레이션 도입 시 적용할 수 있는 가안에 해당됨.

3. 조사데이터 활용 활성화를 위한 고려사항

조사데이터의 활용을 활성화하기 위한 중점 과제는 조사데이터의 체계적 관리이다. 여기서 말하는 관리란 조사데이터의 기획·계획·생성·이관·보관·제공을 포함한 데이터의 생애 전체를 일컫는다. 만약 조사데이터 체계적으로 관리되지 못할 경우에는 데이터의 내·외부적 활용에 상당한 제약을 받을 수밖에 없다. 예를 들어 적절한 관리 체계 안에서 생성된 데이터는 그 결과에 대한 신뢰수준이 높기 때문에 데이터에 대한 수요를 충족시킬 수 있으며, 향후 재이용의 관점에서도 2차 활용결과의 신뢰도와 타당성을 보장할 수 있게 된다. 또한, 개인정보와 민감정보가 적절하게 처리된 데이터의 경우나 IRB승인을 받은 데이터의 경우에는 조사참여자에 대한 보호 조치가 이루어진 상태이기 때문에 재이용의 제약이 상당 부분 완화될 수 있다. 이처럼 조사데이터의 활용을 활성화하기 위해서는 체계적인 데이터 관리체계를 구축하여 이를 지속적으로 이행해 나가는 것이 무엇보다 중요하다.

조사데이터의 활용을 활성화하기 위해서는 기본적으로 데이터의 제공 범위를 확대하는 것이 필요하다. 그러나 제공 데이터의 수가 증가하여 일정 수준에 이르게 되면 데이터 활용의 비효율성이 발생될 수 있다. 즉, 데이터 제공을 위한 시스템 운영이나 관련 행정처리에 요구되는 자원에 비하여 데이터 활용 정도가 떨어질 수 있다. 그러므로 조사데이터의 활용을 지속적으로 활성화하기 위해서는 조사데이터의 개요 및 통계데이터의 현행화가 정기적으로 이루어져야 하며, 데이터의 수정과 보완이 원활하게 이루어져야 한다. 즉, 데이터 이용자의 수요에 맞게 데이터를 제공해야 하며, 수요자의 의견이 반영될 수 있도록 절차를 마련하는 것이 중요하다. 더 나아가 수요친화적인 조사데이터의 제공(데이터 큐레이션 내용 참

고)을 위한 정기·수시 점검이 이루어질 필요가 있다.

끝으로 공공기관에서 생산된 조사데이터는 연구자의 연구결과물임과 동시에 조사참여자의 정보가 포함된 의미있는 연구자료이며, 공적 활용을 목적으로 생성된 공공데이터이다. 조사데이터를 생산하고 관리하는 주체는 데이터의 공적 가치 확산에 대한 의무를 가진다는 점이 강조될 필요가 있다. 또한, 조사데이터의 활용도를 제고하기 위한 노력은 데이터의 공적 가치를 확대하는 과정이라 볼 수 있으므로 장기적인 관점에서 조사데이터의 가치 제고를 위한 목표와 전략 등이 기관의 설립 목적에 맞게 설정되고 순차적으로 이행될 필요가 있다. 이를 위해서는 기관 내 관련 인적·물적 인프라 구성이 선행되어야 하지만 그 외에 조직 구성원의 적극적인 참여가 독려될 필요가 있다.



- 강영준, 이상운, 박근영(2022). 제주연구데이터플랫폼 구축방안 연구. 제주연구원.
개인정보보호위원회(2021). 개인정보 보호법령 해석 실무교재.
공공데이터활용지원센터(2021a). 공공데이터 관리지침. 행정안전부.
공공데이터활용지원센터(2021b). 공공데이터 제공·관리 실무 매뉴얼. 행정안전부.
공용기관생명윤리위원회(2023). 인간대상연구 조사연구용 연구계획서 권고서식 제 1-1-2호.
국가과학기술연구회(2019a). 연구데이터 관리 가이드라인.
국가과학기술연구회(2019b). 연구데이터관리계획(DMP) 가이드라인.
국가과학기술심의회(2018). 연구데이터 공유·활용 전략(안).
국가데이터정책위원회(2023). 제1차 데이터산업 진흥 기본계획.
국가생명윤리정책원(2023). 기관위원회 평가인증제도 관련 안내자료.
기관생명윤리위원회 정보포털(2023). 기관생명윤리위원회 등록 현황.
대한기관윤리심의기구협의회(n.d.). IRB행정간사 심화교육 발표자료
보건복지부·국가생명윤리정책원(2022). 기관생명윤리위원회 관리 안내.
송미영, 이미경, 이은경, 이정찬, 서정호, 박지윤(2022). KICE 교육데이터 활용
활성화 방안 탐색 및 시스템 설계. 한국교육과정평가원.
신은정, 안형준, 양현채, 최병삼, 양승우, 정원교, 김수연(2017). 오픈사이언스정
책의 도입 및 추진방안. 과학기술정책연구원.
신은정. (2018). 과학기술 연구데이터 공동 활용 기반 마련을 위한 연구. 과학기
술정보통신부 과제 보고서. 과학기술정책연구원.
신은정, 손수정, 서지영, 김영린(2019). 공공부문 연구데이터의 소유·활용제도
개선방향. 과학기술정책연구원.
이민정(2023). 미국의 「오픈사이언스의 해」 선포와 정책적 시사점. 한국과학기
술정보연구원.
이상현(2020). 공공데이터 기록관리 활성화와 큐레이션 활용방안 연구-공공데

- 이터포털, 해안시스템을 중심으로. 기록과 정보· 문화 연구, 11, 115-153.
- 이의규(2018). 통계개발원 2018년 상반기 연구보고서 제1권. 표준 자료처리 모델 GSDEMs의 소개 및 시사점(p.49-87). 통계진흥원.
- 이혜림(2020). 디지털 큐레이션 가이드라인과 체크리스트. 한국과학기술정보연구원.
- 이혜진, 김재훈, 현미환(2022). 디지털 전환 대응 디지털 큐레이션 전략. KISTI Issue Brief. 45. 한국과학기술정보연구원.
- 정영철, 정소희, 이기호, 김은주, 진재현, 안수인(2020). 전사적 연구데이터 관리 체계 마련을 위한 연구. 한국보건사회연구원.
- 최명석, 김현우, 이상환(2019). 국가 연구데이터 공유의 시작, 데이터 관리계획 (Data Management Plan), KISTI ISSUE BRIFE, 13.
- 통계청(2019). 통계작성 단계별 가이드라인.
- 통계청(2022). 국가통계 품질관리 매뉴얼: 정기통계품질진단. 통계청.
- 한국교육개발원(2023a). 생명윤리심의위원회 운영 지침.
- 한국교육개발원(2020). 조사도구 및 조사자료 관리 지침.
- 한국교육개발원(2023b). 조사자료 관리 및 공개 지침.
- 한국교통연구원(2016). 연구조사자료관리규칙.
- 한국보건사회연구원(2022a). 조사자료 관리 지침.
- 한국보건사회연구원(2022b). 생명윤리위원회 운영규정.
- 한국보건사회연구원(n.d.). 조사자료 관리목록. 내부자료.
- 한국사회과학자료원(2023a). 자료 기탁 가이드.
- 한국사회과학자료원(2023b). 조사자료 체크리스트.
- 한국여성정책연구원(2021). 통계사무 및 원시자료 서비스 지침.
- 한국직업능력연구원(2021). 통계자료 관리 지침.
- 한국형사·법무정책연구원(2021). 자료관리규정.
- 행정안전부(2021). 정보공개 운영 안내서.

Bastian, M., S. Heymann, and M. Jacomy. (2009). Gephi: An open source software for exploring and manipulating networks. Proc., In Third international AAAI conference on weblogs and social media, San Jose, California, the Association for the Advancement of Artificial Intelligence, 2 pp.

Cherven, K. (2015). Mastering Gephi network visualization. Packt Publishing.

G8 Open Data Charter and Technical Annex. (2013).

<https://www.gov.uk/government/publications/open-data-charter/g8-open-data-charter-and-technical-annex>에서 2023.07.11. 인출

「공공기관의 정보공개에 관한 법률, 법률 제17690호 (2021).」

「공공데이터의 제공 및 이용 활성화에 관한 법률, 법률 제17344호 (2020).」

「국가연구개발정보처리기준, 과학기술정보통신부고시 제2020-102호 (2021).」

「국가연구개발혁신법, 법률 제17343호 (2021).」

「개인정보 보호법, 법률 제19234호 (2023).」

「데이터기반행정 활성화에 관한 법률, 법률 제17370호 (2020).」

「데이터 산업진흥 및 이용촉진에 관한 기본법, 법률 제18475호 (2022).」

「생명윤리 및 안전에 관한 법률, 법률 제17783호 (2021).」

「신용정보의 이용 및 보호에 관한 법률, 법률 제19234호 (2023).」

「정보통신망 이용촉진 및 정보보호 등에 관한 법률, 법률 제19154호 (2023).」

「정부출연연구기관 등의 설립·운영 및 육성에 관한 법률, 법률 제18432호 (2021).」

「지방자치단체의 통계 작성·보급 및 관리에 관한 기본지침, 통계청 예규 제253호 (2021).」

「통계법, 법률 제17339호 (2020).」



[부록 1] 기관위원회 평가·인증 기준

○ 2개 영역, 5개 범주, 세부 평가기준 40개
(국가생명윤리정책원, 2023년 3월 기준)

2개 영역	5개 범주	40개 평가 기준
기관	1. 설치와 독립성	1.1. 생명윤리법 제10조에 따라 기관에서 수행되는 연구 및 활동의 특성을 고려하여 기관위원회를 설치하고 있다.
		1.2. 기관위원회가 기관장 직속으로 설치되는 등 운영의 독립성이 보장된다.
		1.3. 기관위원회 구성 및 운영의 이해상충을 관리할 수 있는 근거 등을 정하여 운영한다.
		1.4. 기관위원회 심의와 관련하여 부결된 사항을 기관장이 승인할 수 없도록 정하여 운영한다.
	2. 지원	2.1. 기관장은 기관위원회 운영에 필요한 시설 및 공간을 지원한다.
		2.2. 기관장은 기관위원회 운영지원을 위해 필요한 인력 등을 지원한다.
		2.3. 기관장은 기관위원회 운영에 필요한 예산을 배정하고 지원한다.
		2.4. 기관장은 기관 내 연구자 및 종사자에게 필요한 생명윤리(법) 및 기관위원회 관련 교육을 지원한다.
		2.5. 생명윤리법 제11조 ④에 따른 보고 기준 및 절차를 정하여 운영한다.
		2.6. 기관 내 연구자 및 종사자가 생명윤리법에서 규정하는 의무를 이행하는지 여부를 관리한다.

2개 영역	5개 범주	40개 평가 기준
기관 위원회	3. 구성	3.1. 생명윤리법 제11조 ①에 따라 위원을 구성하고 있다. 3.2. 위원 대상 신규교육 및 보수교육에 대한 기준을 정하여 관리한다. 3.3. 기관위원회 운영에 필요한 구성원을 규정하고, 그 역할과 의무를 정하여 운영한다. 3.4. 생명윤리법 제11조 ②에 따라 선출된 위원장이 자신의 역할을 이해하고 적절하게 수행한다. 3.5. 위원의 위촉 및 연임, 해촉 등에 관한 기준 및 절차를 정하여 운영한다.
	4. 운영	4.1. 회의 개최 기준과 시기, 심의를 위한 안전 배포 등 사전 준비에 관한 사항을 정하여 운영한다. 4.2. 생명윤리법 시행규칙 제8조 ②에 따른 회의 성립 요건과 의결 기준을 정하여 운영한다. 4.3. 기관위원회 심의 시 위원 등의 이해상충을 관리할 수 있는 기준과 절차를 정하여 운영한다. 4.4. 심의결과의 유형과 기준, 심의결과에 따른 후속처리 기준과 절차를 정하여 운영한다. 4.5. 회의록의 구성 내용과 작성, 회람, 승인 절차를 정하여 운영한다. 4.6. 심의(심의면제 포함) 결과에 대한 통보 절차 및 양식을 정하여 운영한다. 4.7. 기관위원회 심의 및 심의면제 신청 등 이용에 필요한 제출서류 및 절차를 정하여 기관 내 연구자 및 종사자에게 제공한다. 4.8. 기관위원회 문서(전자문서 포함)에 대한 보관(보안 관리 포함), 이관 또는 폐기 등 관리 기준과 절차를 정하여 운영한다. 4.9. 연구대상자등의 정보공개 요청이 있을 경우 처리할 수 있는 기준과 절차를 정하여 운영한다. 4.10. 기관위원회 운영지원인력(행정간사 등)을 위한 교육을 지원하며, 운영지원인력이 전문성을 갖추고 있다. 4.11. 기관위원회 표준운영지침의 제·개정에 관한 기준과 절차를 정하여 관리한다.

2개 영역	5개 범주	40개 평가 기준
기관 위원회	5. 역할 및 기능	5.1. 기관위원회 유형에 따른 신규 연구과제에 대한 심의 기준과 절차를 정하여 운영한다.
		5.2. 신속심의를 필요한 경우 그 대상과 기준 및 절차를 정하여 운영한다.
		5.3. 인간대상연구 및 인체유래물연구에 대한 심의면제 기준과 절차를 정하여 운영한다.
		5.4. 동의서의 내용과 동의절차, 동의면제에 대한 심의 기준과 절차를 정하여 운영한다.
		5.5. 대리인의 동의가 필요한 경우에 대한 심의 기준과 절차를 정하여 운영한다.
		5.6. 연구대상자등에 미치는 위험 수준을 고려하여 필요한 안전 대책을 심의한다.
		5.7. 연구대상자등의 식별가능성을 고려하여 필요한 개인정보 보호대책을 심의한다.
		5.8. 연구를 통해 얻은 개인정보, 인체유래물등, 배아줄기세포주의 제공 또는 인체유래물등의 폐기·이관에 대한 심의 기준과 절차를 정하여 운영한다.
		5.9. 연구 외 생명윤리법에 따라 기관위원회 심의가 요구되는 사항에 대한 심의 기준과 절차를 정하여 관리한다.
		5.10. 취약한 연구대상자등에 대한 보호방안이 있고 그 준수를 관리한다.
		5.11. 기관 내 연구자 및 종사자를 위한 윤리지침을 마련하여 제공한다.
		5.12. 지속심의 대상 과제에 대한 심의 기준과 절차를 정하여 관리한다.
		5.13. 수행 중인 연구의 계획변경, 종료 및 위반·이탈 등에 대한 보고 및 심의 기준과 절차를 정하여 관리한다.
		5.14. 해당 기관에서 수행 중인 연구 및 활동에 대한 조사감독 기준과 절차를 정하여 관리한다.

자료: 국가생명윤리정책원(2023). 기관위원회 평가인증제도 관련 안내자료.

[부록 2] 통계작성 단계별 체크리스트

I. 기획

1. 통계작성 기획

- 조사 목적이 명확하고 구체적인가?
- 유사·중복 통계 유무를 검토하였는가?
- 전수조사와 표본조사 중 조사방법을 결정하였는가?
- 조사 목적에 맞는 조사 대상이 선정되었는가?
- 공표범위를 결정하였는가?
- 조사 기준시점 및 대상기간을 설정하였는가?
- 조사 기간은 적절하게 산정하였는가?
- 조사 시기는 조사 진행에 차질이 없도록 결정하였는가?
- 조사 대상을 고려한 조사방법(면접, 전화, 배포)을 선택하였는가?
- 통계 작성 기본 계획을 수립하였는가?
- 국가통계 승인(변경 승인) 가능 여부를 사전 검토하였는가?

2. 소요 예산

- 조사준비·실시, 자료처리 및 결과 공표 등 조사 과정을 모두 고려하였는가?
- 조사원 1일 업무량은 적절하게 산정되었는가?
- 조사원의 1인 1일 단가가 적정한가?
- 통계조사답례품은 적절하게 산정되었는가?
- 조사용품류 예산이 적절하게 반영되었는가?
- 시스템개발 비용이 반영되었는가?

II. 설계

1. 조사 항목 선정 및 조사표 설계

- 유사통계 조사 항목 자료와 정책자료·연구자료를 검토하였는가?
- 개념, 용어, 기준 등의 분류체계 및 표준이 있는지 확인하였는가?
- 조사 항목 선정 시 주의사항을 확인하였는가?
- 조사 항목 문구 작성 시 고려사항을 확인하였는가?
- 조사 항목 길이는 적절한가?
- 조사 항목 배열은 적절한가?
- 조사표에 포함되어야 할 주요 내용이 모두 포함되어 있는가?
- 결과표를 고려하여 조사표를 설계하였는가?
- 조사표 사전테스트를 통해 타당성을 검토하였는가?
- 조사표 전문 심의를 거쳤는가?
- 이용 목적에 적합하도록 결과표를 설계하였는가?

2. 자료수집 방법 설계

- 표준화 가이드라인을 참조하여 지침서를 작성하였는가?
- 조사자료 수집 방법은 적절한가?

3. 모집단 및 표본 설계

- 모집단의 정의가 정확히 이루어졌는가?
- 표본 추출률의 구성가능성/포함 정도는 어떠한가?
- 목표오차를 고려하여 표본 규모를 산출하였는가?
- 예산 규모에 맞는 표본 규모를 산출하였는가?
- 조사 목적에 적합한 표본추출 방식을 채택하였는가?
- 표본 설계보고서를 작성하였는가?
- 표본수준 유지를 위해 적절한 표본 관리를 하고 있는가?

II. 설계

4. 자료처리 계획 수립

- 자료처리 계획을 수립하였는가?

5. 통계생산 체계 설계

- 통계생산에 대한 세부적인 계획서를 작성하였는가?
 통계작성 승인 사전 협의를 하였는가?
 통계작성 승인 신청 서류를 준비하였는가?
 통계작성 승인 신청 기한을 준수하였는가?

III. 구축

1. 통계생산 시스템 구성

- 조사원을 모집하고 관리할 수 있는가?
 조사명부를 관리할 수 있는가?
 조사 진행 상황을 입력·확인할 수 있는가?
 조사표를 입력·내검할 수 있는가?
 결과표를 집계할 수 있는가?

2. 시스템 통합테스트

- 각각의 시스템에 오류가 없는지 철저히 점검하였는가?
 자료수집, 자료처리 및 자료 분석에 이르는 일련의 과정을 모두 점검하였는가?

IV. 수집

1. 조사 사전준비

- 사전 확인을 통해 표본 명부를 보완하였는가?
 조사 대상처에 조사에 대한 홍보를 실시하였는가?
 조사표, 지침서, 조사용품 등이 준비되었는가?
 조사용품은 예비량을 포함하여 확보하였는가?

2. 조사원 모집 및 교육

- 조사요원 모집 계획이 적절하게 수립되었는가?
 조사원 관리체계가 구축되어 있는가?
 도급조사원은 일반상해보험에 가입하였는가?
 근로계약 조사요원은 근로기준법을 준수하여 모집하였는가?
 조사원 모집절차는 적법한가?
 조사원 교육 시 필수내용을 빠짐없이 교육하였는가?
 조사 기간 중 교체된 조사원에 대한 교육을 실시하였는가?

3. 현장조사

- 준비조사를 실시하였는가?
 조사원에게 조사표, 조사용품, 통계조사답례품을 적정하게 지급하였는가?
 불응·불능 조사 대상처 처리 방안을 수립 및 실행하고 있는가?
 현장조사 시 발생하는 문제점을 파악하며 조치하고 있는가?
 조사 진행상황을 잘 관리하고 있는가?
 현장조사가 조사 지침에 따라 충실히 수행되고 있는가?
 조사 과정자료는 수집하였는가?

IV. 수집

4. 자료수집 점검 및 완료

- 조사표 입력 전 사전 내검을 수행하였는가?
- 입력요원 교육을 실시하였는가?
- 조사 결과 검증을 수행하였는가?
- 조사 결과 입력 시 내검을 수행하였는가?

V. 처리

1. 코딩 및 자료보완

- 조사 결과 자료를 코딩(부호화) 하였는가?
- 내용검토 규칙을 작성하였는가?
- 입력자료 내검을 실시하였나?

2. 결측치 처리

- 무응답처리 과정의 고려사항을 반영하였는가?
- 무응답 대체 방법이 적절한가?

3. 가중치 계산

- 표본 설계 시 기본 가중치를 계산하였는가?
- 단위무응답이 있는 경우 무응답 가중치를 계산하였는가?
- 가중치를 사후에 보정하였는가?
- 가중치 효과를 분석하였는가?

VI. 분석

1. 통계산출물 검증

- 분석자료 작성에 앞서 자료의 개요를 작성하였는가?
- 관련 통계 등과의 비교 분석을 통해 자료 결과를 검증하고 있는가?
- 공표된 잠정치, 확정치 간의 불일치에 대한 원인을 분석하여 관리하고 있는가?
- 시계열 자료는 연속성이 있으며, 단절이 생길 경우 그 내용을 설명하고 있는가?

2. 통계산출물 확정 및 분석보고서 작성

- 용어, 개요 설명 등 이용자를 위한 설명이 제시되어 있는가?
- 그래프로 표현 시 정확한 정보 전달에 유의하였는가?
- 통계작성 과정별 실시 내용이 빠짐없이 작성되었는가?

VII. 배포

1. 공표 자료 배포

- 결과 자료를 적절한 시점에 공표하고 있는가?
- 모든 이용자가 조사 결과를 동시에 이용할 수 있도록 하고 있는가?
- 정확한 통계 이용을 위해 필요한 사항을 함께 공표하였는가?
- 공표 자료를 철저히 점검하여 신뢰 가능한 자료를 제공하였는가?

VII. 배포

2. 공표 자료 적재·이용

- 통계데이터베이스를 구축하였는가?
- 자료와 메타데이터를 적절한 형식으로 작성하여 데이터베이스에 적재하였는가?
- 자료제공 방법은 적절한가?

IX. 평가

1. 품질 평가 및 개선

- 지속적으로 품질 개선을 위해 노력하고 있는가?

VIII. 보관

1. 통계자료 관리

- (위탁 작성의 경우) 통계 조사가 완료된 후 수탁기관으로부터 조사와 관련된 자료 일체를 제출받고 있는가?
- 조사서류의 보존 방법은 적절한가?

2. 마이크로데이터 서비스

- 메타데이터를 제공하였는가?
- 마이크로데이터의 활용 가능성을 높이기 위해 노력하였는가?
- 마이크로데이터의 이용 목적을 식의한 후 제공하였는가?
- 비밀 보호를 위한 조치가 되어 있는가?

자료: 통계청(2019). 통계작성 단계별 가이드라인.