

이달의 초점

사회보장분야 행정데이터 구축·연계와 활용 및 활성화 방안

사회보장 행정데이터 활용사례와 향후 과제

|유종성|

사회보장 행정데이터와 사회보장 정책의 근거 강화

|이현주|

사회보장정보시스템 행정데이터 활용 현황과 과제: 가명정보 결합 및 이용을 중심으로

|한은희|

노인 일자리 및 사회활동지원사업 실태조사의 행정데이터 연계 사례와 과제

|천재영|



한국보건사회연구원
KOREA INSTITUTE FOR HEALTH AND SOCIAL AFFAIRS

사회보장정보시스템 행정데이터 활용 현황과 과제: 가명정보 결합 및 이용을 중심으로

Use of Administrative Data from the Social Security
Information System

한은희 한국사회보장정보원 사회보장정보연구소 부연구위원

2020년 1월 개정된 개인정보보호법을 근거로 사회보장정보시스템에 축적된 행정데이터를 가명 처리하여 통계 작성, 과학적 연구, 공익적 기록 보존을 위해 이용하는 것이 가능해졌다. 법적 기반이 마련되었음에도 가명정보 결합 및 이용 절차의 모호성 및 복잡성으로 인해 사회보장정보시스템 행정데이터 활용은 다소 미흡한 실정이다. 이 글에서는 사회보장정보시스템 가명정보 이용 절차 및 가명정보 결합 사례를 검토하고, 가명정보 활용을 활성화하기 위한 과제에 대해 논하였다. 사회보장정보시스템에서 제공 가능한 데이터 목록 및 메타데이터 제공, 사회보장정보시스템 행정데이터에 대한 이해 및 활용 능력 함양, 보존(retention) 가능한 통합데이터 셋 구축 등이 추진되어야 할 것이다. 무엇보다 사회보장정보시스템 행정데이터를 활용한 정책연구의 결과를 반영할 수 있는 증거 기반 정책 평가 및 정책 설계 체계가 확립되어야 한다.

1 들어가며

2010년 1월 사회복지통합관리망(현 사회보장정보시스템)이 중앙정부에서 제공하는 복지급여 및 서비스 대상자의 수급 자격과 급여 이력을 효율적으로 관리하기 위하여 도입되었다. 사회보장정보시스템에 축적된 행정자료는 사회보장 급여 및 서

비스 제공을 지원하기 위해 수집 또는 생성되었으나 정책 평가 및 분석을 위한 활용 가치 또한 매우 높다. 하지만 2020년 이전 개인정보보호법(이하 보호법)하에서는 정보 주체의 동의 없이 개인정보의 목적 외 활용이 불가능하기 때문에 사회보장정보시스템에 축적된 행정데이터의 활용은 소수의 보건복지부 연구용역 또는 집계자료 분석에 한정되었다.

2020년 1월 9일 개인정보보호법이 개정되면서 개인정보를 가명 처리하여 통계 작성, 과학적 연구, 공익적 기록 보존을 위해 활용하는 것이 가능하게 되었다.¹⁾ 특히 행정데이터의 활용 가치를 높이기 위해 이종 데이터 간의 결합이 필수적이기 때문에 가명정보 결합을 지원하기 위한 분야별 결합전문기관도 지정되었다. 2023년 기준 보건복지 분야에서 국민건강보험공단, 심사평가원, 국립암센터, 한국사회보장정보원이 결합전문기관으로 지정되어 운영되고 있다. 이와 같은 변화로 많은 연구자들이 사회보장정보시스템 행정데이터에 관심을 가지고 있다. 일부 연구자들은 가명정보를 활용한 연구를 계획하거나 진행 중이다. 하지만 가명정보 활용 절차가 모호하고 복잡할 뿐 아니라 데이터 신청부터 데이터 반출까지 오랜 시간이 걸리기 때문에 실제 가명정보 활용은 다소 미흡한 실정이다. 이 글에서는 사회보장정보시스템 가명정보 이용 절차와 가명정보 결합 사례를 검토하고, 사회보장정보시스템 가명정보 활용을 활성화하기 위한 과제를 논하고자 한다.

2 사회보장정보시스템 가명정보 이용 절차

보건복지부(이하 복지부)와 한국사회보장정보원(이하 정보원)은 개인정보위원회의 가명정보 제공

가이드라인(2022년 4월 개정)을 바탕으로 사회보장정보시스템 가명정보 제공 업무 절차를 마련하여 안내하고 있다. 사회보장정보시스템 가명정보 제공 절차는 ① 사전준비 및 신청, ② 식별 위험성 검토, ③ 가명 처리, ④ 적정성 검토, ⑤ 가명정보 제공으로 요약된다.

가. 사전준비 및 신청

사회보장 행정데이터를 이용하고자 하는 신청자(이하 신청자)는 먼저 행정데이터를 가명화하여 이용하고자 하는 목적이 통계 작성, 과학적 연구, 공익적 기록 보존에 해당하는지를 명확히 하여야 한다. 신청자는 가명정보 활용 목적을 달성하기 위해 필요한 데이터의 종류, 데이터 추출 범위 등을 탐색·확인하기 위해 정보원에 사전 협의를 요청하고 지원을 받을 수 있다. 가명 처리하고자 하는 행정데이터가 명확해지면 신청자는 복지부의 정보 제공 담당사업과(이하 담당사업과)에 신청서와 연구계획서 등 구비 서류를 준비하여 가명정보 이용 승인을 요청한다. 담당사업과는 신청서와 구비 서류를 검토한 후 미흡한 점이 있으면 신청자에게 보완을 요청할 수 있다. 담당사업과는 가명정보 활용 목적이 통계 작성, 과학적 연구, 공익적 기록 보존에 부합하는지 검토한다(보호법 제28조의 2). 그 외에도 가명정보 활용이 제3자의 권리를 침해할 가능성이

1) 가명 처리는 개인정보의 일부를 삭제하거나 일부 또는 전부를 대체하여 추가 정보 없이는 특정 개인을 알아볼 수 없게 하는 것을 의미한다.

[표 1] 사회보장정보시스템 가명정보 활용 프로세스

구분	업무 절차	주요 내용	처리 주체	제출 서류 또는 산출물
① 사전준비 및 신청	신청서 접수	• 가명정보 활용 목적을 명확히 설정하고 복지부(신청 정보 관련 담당사업과)에 신청서 및 구비서류 제출	신청자	신청서, 연구계획서 등
		• 복지부는 신청서 및 구비 서류를 접수·검토하고 필요 시 보완 요청	복지부	
	가명정보 활용 목적 적합성 검토	• 복지부는 신청자가 요청한 정보의 추출 기준, 항목 정의, 제공 범위 설정 등 정보원과 사전 위험성 검토 진행 • 복지부는 가명정보 활용 목적 타당성, 개인정보 유출 및 재식별 위험 등을 검토하여 자료 활용 승인 여부 결정	복지부	검토 의견서, 검토 결과서
	결과 통보	• 복지부는 자료 활용 승인 검토 결과에 따라 자료 활용 승인 여부를 신청자에게 통보	복지부	
② 식별 위험성 검토	자료 추출 요청 및 가명 처리 위탁	• 복지부 내부 규정 및 절차에 따라 정보원에 자료 추출 요청 • 복지부·정보원 간 가명 처리 업무 위탁	복지부	자료 추출 및 가명 처리 위탁 공문
	자료 추출	• 신청자와 추출 조건을 협의하여 자료 추출 및 검증 수행	정보원	
	식별 위험성 검토	• 가명 정보 처리 목적 달성에 필요한 최소 가명 처리 대상 항목 선정하고, '식별 위험성 검토 결과 보고서' 작성	정보원	식별 위험성 검토 결과 보고서
③ 가명 처리	가명 처리	• 항목별 가명 처리 계획서 작성 • '가명정보 처리 가이드라인'의 가명 처리 절차 수행	정보원	가명 처리 계획서
④ 적정성 검토	가명 처리 적정성 평가	• 내외부 전문가로 구성된 가명 처리 적정성 검토위원회 구성 및 운영	정보원	가명 처리 적정성 검토 결과서
	평가 결과 및 가명정보 제출	• '적정성 검토 종합결과서' 등을 복지부에 제출	정보원	
⑤ 가명정보 제공	가명정보 제공	• 가명 처리 적정성이 확인된 가명정보를 신청자에게 제공 • 가명정보의 재식별 금지, 안전 조치 이행, 목적 외 활용 금지 등 안내 및 협약서 징수	복지부	사후관리 이행 협약서
	자료 분석	• 정보원 폐쇄형 분석 공간에서 가명정보 분석 후 결과만 반출	신청자	
	사후관리	• 가명정보 처리 관련 기록 작성 및 보관 • 가명정보 재식별 금지 및 재식별 가능성 모니터링	정보원	가명정보 제공 관리 대장

자료: 한국사회보장정보원. (2023). '사회보장분야 가명정보 제공 업무처리안내서'를 바탕으로 저자 재구성.

있는지, 가명정보가 아닌 익명정보로 목적 달성이 가능한지 등을 검토하고, 가명 처리 및 제공에 소요 될 것으로 예상되는 인적·물적 자원에 대해 정보원과 협의한다. 적합성 검토가 완료되면 담당사업과는 신청자에게 가명정보 제공 승인 결과를 통보한다.

나. 식별 위험성 검토

가명정보 제공이 승인되면, 담당사업과는 보건복지부의 '보건복지 급여·서비스 기준 관리 및 정보시스템 운영에 관한 규정'에 따라 정보시스템 운영을 담당하는 복지정보운영과에 요청하고, 복지정보

[표 2] 식별 위험성 검토 항목별 조치 가이드라인

구분	검토 항목	상세 내용	조치 가이드라인
데이터 자체 식별 위험성	식별 정보	<ul style="list-style-type: none"> 특정 개인과 직접적으로 연결되는 정보 (예시) 성명, 고유식별 정보, 휴대전화번호, (개인)전자우편 주소, 의료기록 번호, 건강보험 번호 등 	<ul style="list-style-type: none"> 개인을 직접적으로 알아볼 수 있는 식별 정보는 원칙적으로 삭제하여야 하며, 결합 등 이용 목적상 필요한 경우 안전한 방식으로 대체할 수 있는 정보를 생성하여 대체
	식별 가능 정보	<ul style="list-style-type: none"> 다른 항목과 결합하는 경우 식별 가능성이 높아지는 항목 (예시) 성별, 연령, 거주 지역, 국적, 직업, 위치정보 등 	
	특이 정보	<ul style="list-style-type: none"> 전체 데이터에 식별 가능성을 가지고 있는 고유(희소) 값, 편중된 분포를 가지는 단일·다중 이용 항목 (예시) 희귀 성씨 등 특이한 값, 국내 최고령 등 극단값, 특정 데이터 분석 집단에서 희소한 값 등 	<ul style="list-style-type: none"> 특이 정보는 그 정보만으로 개인을 식별할 수 있는 정보는 아니더라도 고유(희소)한 특성 때문에 개인을 알아볼 수 있는 가능성이 높으므로 이용 목적상 반드시 필요하지 않다면 삭제하고 필요 시 상하단 코딩, 범주화 등 처리
	재식별 시 영향도	<ul style="list-style-type: none"> 특정 정보 주체에게 사회적 파장 등 영향도가 높은 항목 사회통념상 차별받을 수 있는 정보 또는 재식별로 인한 불이익이 큰 정보 주체(대중적으로 유명한 사람 등) 	<ul style="list-style-type: none"> 사회통념상 차별받을 수 있는 정보나 대중적으로 유명한 사람 등의 정보는 재식별 시 다른 일반정보와 다르게 개인의 피해와 더불어 사회적 파장이 있을 수 있으므로 반드시 필요하지 않다면 삭제
처리 환경 식별 위험성	이용 및 제공	<ul style="list-style-type: none"> 내부 이용 또는 외부 제공하는 경우 처리자(또는 취급자)가 보유하거나 접근·입수 가능한 정보, 이용 범위 및 유형 등을 고려하여 식별 가능한 항목 보안서약서, 계약서 등을 통해 파악 가능한 범위 정보를 고려하여 식별 가능성 검토 가능 	<ul style="list-style-type: none"> 이용 및 제공의 위험성이 있는 경우 이용자와 제공자가 처리 환경의 위험성 감소 및 안전성 입증 관련 협의 필요
	처리 장소	<ul style="list-style-type: none"> 해당 가명정보 외에 다른 정보의 접근·입수가 제한된 장소 보안서약서, 계약서 등으로 내·외부 활용이 제한된 경우 폐쇄 환경에 준하여 검토 가능 	<ul style="list-style-type: none"> 물리적·관리적·기술적으로 처리 장소의 안전성 확보가 되지 않으면 가명 처리의 수준을 높이거나 별도의 안전한 처리 장소를 모색
	처리 방법	<ul style="list-style-type: none"> 가명정보를 다른 정보와 연계·분석·내부 결합하는 경우 결합 후 식별 가능한 항목 가명정보 반복 제공 시 식별 위험이 높아지는 항목이 있는지 검토 	<ul style="list-style-type: none"> 다른 정보와의 연계·결합 예정이 있는 경우 연계·결합되는 정보와 결합하여 식별 가능성이 높아지는 항목이 있는지 추가 검토 필요 처리 주체가 보유하거나 접근·입수 가능한 정보를 통해 식별 가능한 항목이 있는지 검토 ※ 내·외부 이용 기관 또는 처리자가 보유하고 있는 과거 유사 정보에 대한 수행 경험이나 지식 등은 자체적으로 판단하기 어려우며, 가명정보를 처리하게 되는 관련 담당자에게 사전에 확인 후 체크 필요. 관련 담당자의 확인이 어려운 경우 가명 처리의 수준을 높이는 방법 등으로 위험성을 낮춰야 함

자료: 개인정보위원회. (2022). 가명정보 처리 가이드라인. pp. 16~27.

운영과는 사회보장정보시스템 위탁운영 기관인 정보원에 자료 추출을 요청한다. 이때 담당사업과는 자체적으로 가명 처리를 수행할지 아니면 정보원에

가명 처리를 위탁할지 결정하여 정보원과 가명 처리 업무 위탁을 체결한다. 복지부·정보원 간에 가명 처리 위탁 업무가 처리된 경우에는 정보원이 다음

단계의 업무를 처리하게 된다. 데이터 추출이 완료 되면 가명 처리 대상 정보를 확정하고, 가명 처리 대상 정보의 식별 위험성을 검토하여 가명 처리 방법과 수준을 결정한다(표 2). 식별 위험성은 ① 데이터 자체의 식별 위험성과 ② 처리 환경의 식별 위험성을 구분하여 검토한다.

법 및 수준을 결정하고 항목별 가명 처리 계획을 세운 후 계획에 따라 가명 처리를 수행한다(표 3). 신청자는 목적달성 가능성 검토를 위해 가명처리 수행 전 정보원과 협의를 진행할 수 있다. 만약 가명 처리 방법 및 수준이 적정하지 않다고 판단되는 경우에는 식별 위험성을 재검토할 수 있다.

다. 가명 처리

정보원은 식별 위험성 검토 결과를 기반으로 가명정보 활용 목적 달성을 위해 적절한 가명 처리 방

라. 적정성 평가

가명 처리가 완료되면 가명 처리 결과에 대한 적정성을 최종 검토한다. 가명 처리가 적정하게 수행

[표 3] 항목별 가명 처리 계획 작성 예시 비교

순번	항목명	개인정보 유형	가명처리 계획 A		가명처리 계획 B	
			처리 방법	처리 수준	처리 방법	처리 수준
1	고객 ID	개인 식별 정보	대체	• 일련번호 대체	대체	• 일련번호 대체
2	나이	개인 식별 가능 정보	범주화	• 5세 단위 범주화	범주화	• 10세 단위 범주화
			상하단 코딩	• 13세 미만 삭제 • 90세 이상은 90세 이상 경계치 입력	상하단 코딩	• 20세 미만 삭제 • 80세 이상은 80세 이상 경계치 입력
3	주소	개인 식별 가능 정보	부분 삭제	• 동 단위 이하 삭제	처리 없음	
4	성별	개인 식별 가능 정보	처리 없음		처리 없음	
5	2019년 1월 여행용품 구매액	개인 식별 가능 정보	범주화	• 상단 99.9%를 초과하는 경우 경계치로 변경 • 모든 금액에 대해 1만 단위로 반올림 처리	범주화	• 상단 99.9%를 초과하는 경우 경계치로 변경 • 다음과 같이 금액 범주화 적용 • 0원 • 10만 단위 미만: 1만 단위로 반올림 • 100만 단위 미만: 10만 단위로 반올림 • 1000만 단위 이상: 100만 단위로 반올림
6	2019년 1월 식품류 구매액	개인 식별 가능 정보				
7	2019년 1월 의류 구매액	개인 식별 가능 정보				
8	2019년 1월 취미용품 구매액	개인 식별 가능 정보				
9	2019년 1월 생활용품 구매액	개인 식별 가능 정보				
10	2019년 1월 구매 총금액	개인 식별 가능 정보	범주화	• 각 구매액과 동일한 처리	범주화	• 각 구매액과 동일한 처리

자료: 개인정보위원회. (2022). 가명정보 처리 가이드라인. p. 32 참조.

되었는지 확인하고, 가명 처리 결과가 가명정보의 처리 목적을 달성하기 위해 적절한지 등을 검토하여 가명 처리 적정성 검토 결과서를 작성한다. 가명 처리 적정성 검토는 내부 인원을 활용하여 자체적으로 수행하거나 위원회를 구성하여 외부 전문가를 통하여 수행할 수 있다. 개인정보위원회는 최소 3명 이상의 검토위원회를 구성하는 것을 권고한다. 만약 적정성 검토 과정에서 재식별 가능성이 제기된 경우에는 앞의 3단계(목적 설정 등 사전준비-식별 위험성 평가-가명 처리)를 다시 수행하거나 부분적으로 추가적인 처리를 수행할 수 있다.

마. 데이터 분석 및 사후관리

복지부와 정보원은 제공되는 가명정보의 위험성 정도를 고려하여 제공 방식을 결정한다. 필요 시 출입통제 장치 및 CCTV가 설치되어 있고, 출입관리 시스템이 가동되는 정보원 내 보호구역에서만 데이터 분석이 수행되도록 데이터가 제공될 수 있다. 정보원은 데이터 제공 및 분석 과정에서 특정 개인이 식별될 위험이 있는지 여부를 지속적으로 모니터링하여 가명정보를 안전하게 제공하고자 한다. 신청자는 사전준비 단계에서 수립한 가명정보 관리계획에 따라 제공받은 가명정보를 안전하게 관리해야 한다.

3 사회보장정보시스템 가명정보 결합

가명정보 결합은 내부 결합과 외부 결합으로 구분된다. 내부 결합은 동일한 개인정보 처리자가 보유하고 있는 개인정보를 가명 처리하여 결합하는 것을 의미하고, 외부 결합은 서로 다른 개인정보 처리자가 보유하고 있는 개인정보를 가명 처리하여 결합하는 것을 의미한다. 외부 결합은 제3의 기관(결합전문기관)을 통해서 수행된다. 여기서는 가명정보 외부 결합에 초점을 두었다.

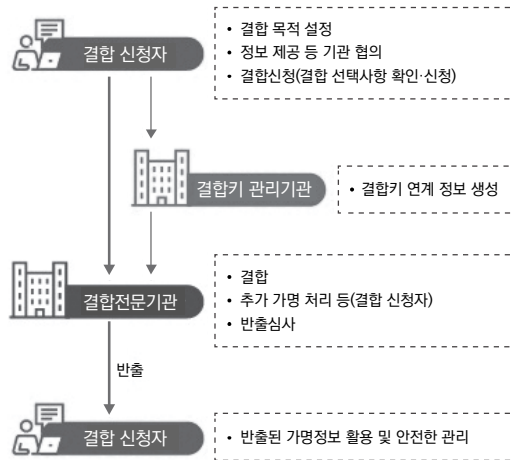
가. 가명정보 결합 절차²⁾

통계 작성, 과학적 연구, 공익적 기록 보존의 목적으로 사회보장정보시스템의 행정데이터(개인정보)와 외부 기관이 보유한 행정데이터(개인정보)를 결합하여 활용하고자 할 경우에는 개인정보위원회 또는 관련 중앙행정기관의 장이 지정한 결합전문기관을 통해 가명 처리 후 결합이 가능하다. 가명정보 결합 절차는 ① 결합 신청자의 결합 신청, ② 결합키 관리기관의 결합키 연계 정보 생성, ③ 결합전문기관의 가명정보 결합 및 반출, ④ 결합 신청자의 반출 정보 활용 및 관리 순이다.

사회보장정보시스템의 행정데이터와 외부 기관의 행정데이터 결합을 추진하고자 하는 연구자는 가명정보 결합 목적을 명확히 하고, 결합 목적을 달성

2) 가명정보 결합에 대해서는 이번 호의 「노인 일자리 및 사회활동 지원사업 실태조사의 행정데이터 연계 사례와 과제」 원고에서 자세히 다루어지므로 여기서는 간략히 기술하였다.

[그림 1] 가명정보 결합 및 반출



자료: 개인정보보호위원회. (2022). 가명정보 처리 가이드라인. p. 39.

하기 위해 이용 가능한 정보를 탐색·확인하기 위해 각 행정데이터 보유 기관과 사전 협의를 진행하여야 한다. 그리고 각 행정데이터 보유 기관의 행정데이터 제공 및 결합에 관한 승인 절차를 따라야 한다.

결합 신청 시에는 결합 신청자뿐 아니라 행정데이터를 제공하는 개인정보 처리자 모두 결합전문기관에 결합 신청서를 제출하여야 한다. 결합 신청자는 가명정보 결합을 수행할 결합전문기관을 선택할 수 있는데, 이때 결합 대상 정보에 대한 전문성, 분석 및 가명 처리에 필요한 인프라, 소요 일정, 가명 처리 또는 분석 지원 여부, 데이터 반출 가능 여부 등을 고려하여 선택하면 된다. 만약 사회보장정보시스템의 데이터와 국민건강보험공단이 보유한 데이터를 결합하고자 한다면 국민건강보험공단 데이

터의 외부 반출이 불가하기 때문에 국민건강보험공단을 결합전문기관으로 지정할 수밖에 없다.

나. 사회보장정보시스템 가명정보 결합 사례

최근 한국사회보장정보원이 결합 전문기관으로서 가명정보 결합을 완료한 과제 중 하나는 소방안전원이 추진한 ‘서울시 취약계층의 주거환경 현황 분석’이다. 소방안전원은 서울시 취약계층의 주거환경 현황 분석을 위하여 사회보장정보시스템의 기초생활보장 생계·의료·주거급여 수급자 정보와 소방안전 빅데이터 플랫폼에서 관리하고 있는 건축물 정보 결합을 추진하였다. 가명정보 결합 논의는 2022년 7월부터 시작되었고, 소방안전원이 보건

복지부 담당사업과에 자료 제공 요청을 시작한 것은 2022년 11월이며, 데이터 결합이 완료된 시점은 2023년 7월이다. 사전준비부터 데이터 결합까지 약 1년이 소요된 것으로 파악된다. 한국사회보장정보원이 결합전문기관으로서 가명정보 결합을 수행한 여러 사업의 평균 소요 기간은 약 7.8개월인 것으로 파악되고 있다(제호영, 2023).

전체 가명정보 결합 과정에서 시간이 가장 많이 소요될 수 있는 파트는 사전준비이다. 사전준비는 가명정보 결합 목적 정의, 목적을 달성할 수 있는 행정데이터의 탐색 및 확인, 각 행정데이터 보유 기관으로부터 가명정보 활용 승인을 받기까지의 과정을 의미한다. 단 사전준비 기간은 연구의 목적, 결합하고자 하는 가명정보의 범위, 정보 보유 기관에 따라 차이가 매우 클 수 있다.

4 사회보장정보시스템 행정데이터 활용 개선을 위한 과제

가. 이용 가능 데이터 목록과 메타데이터 제공

사회보장정보시스템의 가명정보 이용 및 결합 절차에 대해서는 정보원 홈페이지를 통해 안내하고 있다.³⁾ 정보원은 가명정보 결합 기관으로 지정된 이후 최근까지 잠재적인 수요 기관을 타기팅하여 가명정보 결합 및 이용 절차에 대해 안내하고 맞춤

형 컨설팅을 적극적으로 수행하여 왔다(길민권, 2022. 9. 28.). 선행 연구들은 행정데이터 활용이 활성화되기 위해서는 무엇보다 가명정보 이용 및 결합 신청 절차가 명확하고 체계적이며 투명해야 한다고 제언한다(Jones, K. H., Heys, S., Tingay, K. S., Jackson, P., & Dibben, C., 2019). 사회보장정보시스템 가명정보 결합 및 이용 절차에 대한 안내뿐 아니라 신청 및 문의를 할 수 있는 온라인 창구가 마련될 필요가 있다. 무엇보다 사회보장정보시스템에서 제공 가능한 정보 목록과 메타데이터가 제공되어야 한다. 사회복지정보시스템을 통해 관리되는 모든 복지사업에 대해 제공 가능한 데이터 목록과 메타데이터를 단시간에 제공할 수는 없을 것이다. 사회복지 분야의 대표적인 사업부터 시작하거나 최근에 요청되어 제공되었던 데이터부터 목록을 축적하고 메타데이터를 생성하여 제공하는 것을 고려할 수 있다. 활용 가능성과 가치가 낮은 데이터 목록을 많이 제공하는 것보다는 활용 가능성과 활용 가치를 고려하여 제공하는 것이 행정데이터 활용 성과를 높일 수 있다.

나. 행정데이터에 대한 이해와 활용 능력 함양

사회보장정보시스템의 행정데이터를 활용하기 위해 가장 기본적으로 알아야 할 것은 활용하고자

3) 한국사회보장정보원 홈페이지. (n. d.). 가명정보 결합지원. <https://www.ssis.or.kr/lay1/S1T1709C1710/contents.do> 참조.

하는 데이터와 관련된 제도 및 데이터 거버넌스 체계이다. 예를 들어 국민기초생활보장제도가 맞춤형 급여로 분화되면서 급여별 보장 기관이 분화되었기 때문에 생계와 의료급여의 보장 기관은 보건복지부, 주거급여는 국토부, 교육급여는 교육부이다. 맞춤형 급여의 자격 판정은 시군구 통합관리팀에서 수행하기 때문에 자격과 관련된 정보는 사회보장정보시스템에서 관리하고, 급여 관련 정보는 각각의 보장 기관에서 관리하는 구조를 가지고 있다. 따라서 정보 제공 기관도 급여의 종류에 따라, 정보 범위에 따라 달라지게 된다. 사회보장정보시스템 행정데이터 결합 및 이용을 위해서는 제도와 데이터 거버넌스에 대한 이해를 바탕으로 관련된 정보 시스템과 각 정보 시스템이 보유한 데이터의 종류와 범위에 대한 이해가 필요하다.

사회보장정보시스템의 행정데이터는 복지급여 제공 및 수급자 관리를 위해 수집된 데이터이지 연구를 위해 수집된 데이터가 아니다. 이러한 점으로 인해 사회보장정보시스템의 행정데이터를 가명화하여 바로 활용하고자 할 때 연구자는 어려움에 처할 수 있다. 결측치 및 이상치에 대한 데이터 전처리뿐 아니라 연구 목적에 적합한 데이터 형태의 조작(manipulation)을 위해 많은 시간이 소요될 수 있어 관련 기술의 습득이 요구된다. 이러한 이유로 우리보다 앞서 행정데이터 활용을 추진한 영국은 행정데이터 활용을 지원하기 위한 인력과 조직을

마련하여 연구 목적에 맞는 행정데이터를 파악하는 것부터 데이터셋 구축 및 데이터 분석까지 전 과정에서 기술적 지원을 하고, 다양한 교육과 훈련, 네트워킹 기회와 펀딩을 제공하고 있다(한은희, 2020; ADR UK, n.d.).

다. 보전(retention) 가능한 통합데이터셋 구축

최근 영국은 특정 연구 목적을 위해 두 개 이상의 행정데이터를 결합하여 활용하고 연구가 끝난 뒤 파기하는 방식은 비효율적이며 지속가능하지 않다고 판단하고 가명 처리하여 결합한 행정데이터를 보존(retention)할 수 있는 방안을 추진 중이다(한은희, 2020). 사회보장 분야의 행정데이터 활용에서도 연구 목적에 따라 가명정보를 결합하여 이용하고 연구가 끝난 후 파기하는 맞춤형 제공 방식과 지속적으로 연구가 필요한 영역을 선정하고 세부 주제를 포괄할 수 있는 통합데이터셋(Integrated dataset) 구축 방식을 병행할 필요가 있다. 예를 들어 사회보장위원회가 구축하고 있는 사회보장 행정데이터는 보존 가능한 통합데이터셋의 사례이다.⁴⁾ 맞춤형 가명정보 결합 및 이용 경험이 축적됨에 따라 새로운 융합데이터 구축의 필요성이 드러날 수 있다. 사회보장정보시스템의 가명정보를 맞춤형으로 제공함과 동시에 통합데이터를 구축하여 지속적으로 활용하는 방안에 대한 준비가 필요하다.

4) 사회보장 행정데이터에 관해서는 이번 호 「사회보장 행정데이터와 사회보장 정책의 근거 강화」 원고를 참조하기 바란다.

라. 행정데이터 활용한 연구 결과물에 대한 추적 관리

사회보장정보시스템의 가명정보를 활용한 연구 결과물에 대한 추적 관리가 필요하다. 어떤 연구 주제를 위해 어떤 데이터가 활용되었는지, 어떤 분석 방법론이 활용되었는지, 사회보장정보시스템 행정데이터의 한계와 쟁점이 무엇인지에 등에 대한 정보가 축적되고 공유될 필요가 있기 때문이다. 행정데이터를 활용한 분석 결과물은 동료심사(peer-review) 학술지에 발표되는 논문 외에도 정책보고서, 이슈리포트 등 다양한 형태가 있을 수 있다. 사회보장정보시스템 행정데이터를 활용한 다양한 형태의 연구 결과물에 대한 정보가 축적되고 공유된다면 사회보장정보시스템 행정데이터를 활용하고자 하는 연구자들이 활용 가능한 데이터의 범위 및 데이터 활용 방안에 대한 아이디어를 얻을 수 있을 뿐 아니라 통합데이터셋 기획 및 관리 측면에서도 통찰을 얻을 수 있을 것이다.

5 나가며

2020년 개인정보보호법이 개정된 이래 가명정보 결합 및 이용 절차 등이 더욱 명확해지고, 가명정보 활용 사례와 경험들이 축적되고 있다. 이 글에서는 사회보장정보시스템 데이터에 한정하여 가명정보 이용 절차와 가명정보 결합 사례를 검토하고

가명정보 활용을 증진하기 위한 과제와 방안들을 논의하였다. 다른 분야에 비해 복지 분야의 가명정보 활용은 아직까지 활발하지는 못한 편이다. 사회보장정보시스템 가명정보 활용이 활발해지기 위해서는 이용 가능한 데이터 목록 및 메타데이터 제공, 행정데이터에 대한 이해 및 활용 능력 함양, 보존 가능한 통합데이터셋 구축 및 제공, 사회보장정보시스템 가명정보를 활용한 결과물에 대한 추적 관리 등이 병행되어야 한다. 무엇보다 가명정보 활용 결과가 정책적으로 반영되도록 하는 기제가 확립되어야 할 것이다. 근거 기반 정책에 대한 인식이 예전보다는 많이 확산되었고, 「데이터 기반 행정 활성화에 관한 법률」도 제정되었다. 뿐만 아니라 윤석열 정부에서 구현하고자 하는 디지털플랫폼정부의 데이터 기반 행정을 핵심 요소로 하고 있다(윤건, 2023). 사회보장 분야의 증거 기반 정책 평가 및 정책 설계 체계의 확립은 사회보장정보시스템 행정데이터의 이용 및 결합 활성화를 촉발할 것으로 기대된다. 罍

참고문헌

- 개인정보위원회. (2022). **가명정보 처리 가이드라인**.
 길민권. (2022. 9. 28.). 한국사회보장정보원, 가명정보 결합·이용 활성화 지원 위한 맞춤형 설명회 성료.
데일리시큐. <https://www.dailysecu.com/news/articleView.html?idxno=139999>에서 인출.
 윤건. (2023). 공공분야 데이터 융합에 관한 연구: 심층

- 면접을 통한 제약조건 탐색을 중심으로. **한국사회와 행정연구** 33(4). 131-158.
- 제호영. (2023. 11. 17.). 사회복지 분야 가명정보 활용과 과제. **한국정책연합 공동학술대회**. 오송: 2023 사회정책연합 공동학술대회 조직위.
- 한국사회보장정보원. (2023). **사회보장 분야 가명정보 제공 업무처리 안내서**. 내부자료.
- 한은희. (2019). 사회보장 통합 행정데이터 구축 및 활용의 쟁점과 전망. **사회보장정보 이슈리포트 제12호**.
- 한은희. (2020). 영국의 연구를 목적으로 한 행정데이터 연계 활용. **국제사회보장리뷰 제14호**.
- ADR-UK. (n.d.). What is our mission? Retrieved from <https://www.adruk.org/our-mission/our-mission/> 2023. 11. 10.
- Jones, K. H., Heys, S., Tingay, K. S., Jackson, P., & Dibben, C. (2019). The good, the bad, the Clunky: improving the use of administrative data for research. *International Journal of Population Data Science*, 4(1).

Use of Administrative Data from the Social Security Information System

Han, Eunhee

(Korea Social Security Information Service)

The Personal Information Protection Act, as amended in January 2020, allows through pseudonymization for the use of administrative data stored on the social security information system for statistical records, scientific research, and public interest purposes. Despite the legal framework in place, however, the use of administrative data remains limited to an extent due to the unclear and complicated procedures for combining and using pseudonymized data. This article examines specific examples of how pseudonymized social security administrative data has been used and combined and discusses what should be done to further the use of pseudonymized data. This study points out the need for creating a complete list of all the data that can be obtained from the social security information system and providing metadata for these data sets, enhancing people's understanding of and capacity for using social security administrative data, and establishing an integrated dataset that allows for information retention. Most important of all is to establish an evidence-based system of policy evaluation and planning that takes account of the findings of policy research using social security administrative data.