제 출 문

개인정보보호위원회 위원장 귀하

본 보고서를 「데이터 연계·결합 지원제도 도입방안 연구」의 연구결과 보고서로 제출합니다.

2017년 12월

연 구 기 관 정보인권연구소

총괄책임자 이 은 우(정보인권연구소 이사)

참여연구원 오 병 일(정보인권연구소 이사)

임 진 희(정보인권연구소 연구위원)

장 여 경(정보인권연구소 연구위원)

황 규 만(진보네트워크센터 활동가)

요 약 문

1. 제 목

데이터 연계·결합 지원제도 도입방안 연구

2. 연구 필요성 및 목적

세계 주요 국가들은 공공의 이익을 위한 학술 연구를 활성화하고, 증거에 기반을 둔 정책 수립을 지원하기 위해 데이터 연계를 허용하면서도 정보주체의 개인정보를 보호하고 신뢰할 수 있는 데이터 이용 기반을 만들기 위한 법적·제도적 노력을 하고 있다. 데이터 연계를 통해서 데이터의 질을 향상시키거나, 하나의 데이터 소스로는 알 수 없는 새로운 정보를 생성해낼 수 있는데, 이를 통해 정책 결정 및 연구의 질을 높일 수 있다. 반면, 이와 같은 기법은 개인정보 주체의 개인정보 자기결정권 침해나 사생활의 비밀 침해 등의 문제를 야기할 수 있다.

주요 국가들의 전문적 중계기관들은 안전한 환경에서 개인정보가 포함된 데이터가 결합·연계될 수 있도록 지원한다. 데이터 연계 이슈와 관련하여 국내에서는 개인정보의 비식별화 등 기술적 조치에 대해서 주로 초점을 맞추는 경향이 있다. 그러나 비식별화는 데이터 거버넌스의 한 가지 요소일 뿐이다. 예를 들어, 영국의 행정데이터연구네트워크(ADRN)가 제시하는 5가지 안전 원칙은 데이터 비식별화에 대한 데이터 안전(Safe data)의 개념뿐 아니라 연구 인력(Safe people), 연구 프로젝트(Safe project), 연구 환경(Safe environment/settings), 연구 결과물(Safe results/output)을 포괄하는 보다 폭넓은 개념이다.

주요 국가들은 공익에 기여하면서도 안전한 데이터 거버넌스를 위해서 기술적인 조치뿐 아니라 데이터의 이용과 보호에 관련된 법제, 데이터 접근·연계 정책, 연구기관 혹은 데이터 연계기관의 인증, 심사절차, 데이터 접근 절차 등에 이르는 전반적인 보호 체제를 갖추고 있다.

이 모든 과정이 체계적으로 조율되지 못한다면, 자칫 데이터 연계 및 제공 과정에서 개인정보 침해가 발생하거나, 반대로 공익에 기여할 수 있는 데이터의 활용을 제약할 수 있을 것이다. 따라서 데이터 연계·결합을 활성화하기 위해서는 데이터의 수집·연계·제공에 이르는 전 과정에서 데이터의 활용 및 보호를 위한 데이터 거버넌스 체계가 수립될 필요가 있다.

본 연구는 데이터 연계·결합과 관련된 해외 주요 국가의 데이터 거버넌스 체제를 검토하여 우리 사회에 의미가 있는 시사점을 도출하고자 한다. 또한, 보건의료 및 통계 분야, 그리고 비식별 조치 전문기관을 통한 데이터 연계·결합 현황에 대한 분석을 통해 그 한계 및 문제점을 파악하고 데이터 연계·결합을 위한 제도 도입방안을 제시하고자 한다.

3. 연구의 범위

먼저 데이터 연계의 기본적인 개념과 유형 및 그 필요성 및 위험성을 살펴본 후 데이터 거버넌스의 원칙과 모델을 제시한다.

이어서 해외 주요 국가의 데이터 연계·결합 현황에 대하여 각국 개인정보 보호 법제, 보건의료 분야, 연구 목적의 데이터 연계, 통계 목적의 데이터 연계로 구분하여 검토한 후 시사점을 도출한다.

다음으로 국내 데이터 연계·결합 현황을 살펴보면서 우선 개인정보보호법 등 데이터 연계·결합 관련 국내 법제 및 보건의료, 통계 분야의 현황을 분석한다. 개인정보 비식별 조치 전문기관 및 기타 정부 부처 데이터 연계·결합 현황도 살펴본 후 문제점과 개선 방향을 제시한다.

마지막으로 국내 데이터 연계·결합을 위한 제도 개선 방안을 제안한다.

4. 연구 내용 및 결과

(1) 데이터 연계와 데이터 거버넌스

1) 데이터 연계

'데이터 연계(Data Linkage)'란 두 개 이상의 출처로부터 동일인이나 동일한 사건, 기관, 장소에 연관된 정보를 함께 가져오는 것을 의미한다. 식별자 등을 이용해 정보를 결합함으로써 단일 출처의 정보만으로는 알기 힘든 정보 요소 간의 관계가 밝혀질 가능성이 있다. 데이터 연계는 레코드 연계(record linkage), 데이터 매칭(data matching), 데이터 통합(data integration) 등으로도 불린다.

데이터 연계 방법은 데이터셋의 상태나 연구의 목적 등에 따라 정확 연계, 확률연계, 통계적 연계, 다층 연계 등으로 나뉜다. 연계 데이터는 유형별로 횡단면 조사 데

이터, 코호트 연구와 종적 연구, 등록부 데이터, 기타 행정 데이터 등이 있다.

데이터 연계는 새롭거나 발전된 통계의 생산, 현재 일부 정보가 존재하는 측정값들에 대한 추가적인 정보의 생산, 단일 데이터 소스로부터 얻을 수 있는 것보다 더 많은 단위의 폭넓은 변수 활용, 기존 데이터 소스의 개선이나 검증 가능성, 응답자 부담 감소 등의 측면에서 연구에 이익이 된다.

그러나 데이터 연계는 개인 식별에 관련된 데이터가 더욱 많아지기 때문에, 유출이나 공개 시의 위험성이 더욱 증가하게 된다. 이러한 위험을 낮추기 위해 각국은 연구 데이터셋에 대한 관리는 물론, 연계 프로세스를 단계별로 분리 구축하는 데이터 거버넌스 체계를 채택해 왔다.

2) 데이터의 활용과 보호를 위한 체계

데이터 연계가 수행되기 위해서는 데이터 보유기관으로부터 데이터 접근을 위한 승인을 받는 것에서부터 연구자 등 실제 데이터 이용자에게 개인정보 침해 위험을 최소화하는 방식으로 데이터를 제공하는 일련의 과정이 정비되어 있어야 한다. 데이터(정보)의 활용과 보호 전 과정에 걸쳐 적용되는 원칙, 법제도, 가이드라인 등의 체계를 '데이터 거버넌스'라고 부른다.

국제적으로 제안된 데이터 거버넌스 원칙으로는 OECD와 UN의 경우를 살펴볼 수 있다.

OECD는 환자의 프라이버시를 보호하면서 통계 혹은 연구 목적으로 보건의료 데이터를 이용하기 위해 입법 체계를 비롯한 '보건의료 데이터 거버넌스'를 제안해 왔다. 특히 OECD는 2017년 보건의료 데이터 거버넌스에 대한 이사회 권고를 발표하면서 ① 이해당사자의 관여와 참여, ② 기관 간 협업, ③ 공공부문 보건의료 정보시스템에 대한 보호조치, ④ 개인에 대한 명확한 정보 제공, ⑤ 설명에 기반을 둔 동의 체계 보장, ⑥ 보건의료 데이터를 연구 및 공익적 목적으로 사용하는 데 대한 검토와 승인 절차, ⑦ 공공 정보 체계를 통한 투명성 보장, ⑧ 프라이버시 보호, 정보 보안, 개인정보 이용에 대한 개인의 통제권을 보장하면서 정보 가용성을 증진하는 기술적 수단, ⑨ 감독과 평가 체계, ⑩ 적절한 교육훈련 및 기능 개발, ⑪ 통제권과 보호조치, ⑫ 개인 건강정보를 처리하는 기관들에 대한 인증 또는 인가 체계를 갖출 것을 권고하였다.

UN은 2014년 1월 총회에서 'UN 공식통계 기본 원칙'을 승인하며 통계 기관이 통계 편집을 위해 수집한 개별 데이터는, 엄격한 기밀성이 보장되고 통계 목적으로만 사용되어야 한다는 원칙 등을 밝혔다.

이러한 원칙에 조응하며 구축된 데이터 거버넌스 모델로는 우선 영국 행정데이터연구센터(ADRC)의 경우를 들 수 있다. ADRC는 그 설립 과정에서 ① 영국 4개 권역별 ADRC 설립, ② 연구 목적 행정 데이터 접근과 연계를 규정한 입법, ③ 국가적, 국제적인 모범 관행에 기반한, 전국적 연구 승인 절차, ④ 일반 대중의 참여 촉진, ⑤ 연구 목적 접근과 연계를 위한 재정 지원 등의 원칙을 수립해 왔다.

Rosalyn이 제안한 'DASSL 모델'은 개인정보 보호와 데이터 이용을 통한 공익의 균형을 달성하기 위해, ① 거버넌스, ② 연구 데이터 허브, ③ 제3자 데이터 연계 서비스, ④ 안전시설, ⑤ 연구 지원단, ⑥ 결과물 점검 및 공개 통제, ⑦ 대중 참여 및 소통을 데이터 거버넌스 모델의 주요 요소로 제안하였다.

특히 데이터 연계와 관련한 원칙 및 모델로는 유엔의 2009년 '통계 및 관련 연구 목적을 위해 수행되는 데이터 통합의 기밀성 관련 원칙과 가이드라인'을 살펴볼 필요가 있다. 여기서는 제안된 데이터 연계의 원칙과 관련 가이드라인은 다음과 같다. ① 데이터 통합은 통계 및 관련 연구 목적으로만 국가통계기구 및 국가통계시스템 내의 다른 기구에 의해서 수행되어야 한다. ② 국가통계기구는 자신의 국가통계 임무에 부합하고, 표준 승인 절차를 완료한 이후에만 데이터 통합을 수행해야 한다. ③ 데이터 통합 프로젝트의 공익은 데이터 이용과 관련한 프라이버시 혹은 기밀성 우려와 공식통계 시스템의 완전성에 미치는 위험보다 충분히 더 커야 한다. ④ 응답자에게 하지 않겠다고 특정하게 약속을 한 경우에 데이터는 통합되어서는 안 된다. ⑤ 통합 데이터는 승인된 통계 혹은 연구 목적으로만 이용되어야 하며, 애초에 승인된 목적에서 크게 벗어나는 경우 새로운 표준 승인 절차를 밟아야 한다. ⑥ 연계 데이터셋에 포함되는 단위 레코드 및 데이터 변수의 수가 승인된 목적을 위해 필요한 이상이어서는 안 된다. ⑦ 국가통계기구는 데이터 통합을 개방적이고 투명한 방식으로 수행해야 한다. ⑧ 데이터 통합으로부터 나온 통합 단위 레코드 데이터에 대한 접근은 일반적으로 국가통계기구의 허가된 직원으로 제한된다. 다른 통계적 마이크로데이터에 대한 외부인의 접근은, 명확한 법적 근거에 의해 허가되어야 하고 공식통계를 위한 데이터 사용 목적에 부합해야 한다.

한편 영국 행정데이터작업반은 4가지의 데이터 연계 모델을 제안하였다. 그 가운데 '신뢰할 수 있는 제3자 색인(trusted third party indexing, TTP)' 모델의 경우, 절차 중에 어떤 참가자도 전체 데이터셋의 내용이나 식별 데이터를 다룰 수 없다는 점이 장점으로 꼽힌다.

(2) 해외 주요 국가의 데이터 연계·결합 현황

데이터 연계·결합이 기존에 다른 목적으로 수집된 데이터의 2차적 사용에 해당하는 경우가 많고, 서로 다른 목적으로 수집된 대규모 데이터베이스의 연계·결합 시 정보주체의 동의를 다시 획득하기가 쉽지 않다는 점에서, 데이터 연계·결합에 대한 법적 근거가 반드시 필요하다. 개인정보의 수집 목적 외 처리·제공은 개인정보보호원칙에 벗어난 예외적인 경우로서, 정보주체의 권리에 부정적인 영향을 미칠 수 있기 때문이다.

유럽연합, 영국, 독일, 미국 등 주요 국가 법률에서는 주로 공익을 위한 연구 및 통계 목적으로 데이터 연계·결합을 규정하고 있다.

1) 데이터 연계·결합 관련 해외 법제

가. 유럽연합

유럽연합의 개인정보보호 관련 법제는 '개인정보의 처리와 관련한 개인의 보호와 개인정보의 자유로운 이동에 관한 유럽의회 및 이사회의 지침 95/46/EC' 등을 거쳐 2018년 '일반정보보호규정(GDPR)' 발효를 앞두고 있다.

GDPR은 원칙적으로 개인정보를 연계·결합 등 처리할 때, 정보주체가 특정 목적에 대해 본인의 개인정보 처리를 동의하였거나 정보처리자의 법적 의무를 준수하는데 개인정보 처리가 필요한 경우, 공익상 이유 또는 정보처리자의 공식권한을 행사하기 위한 업무 수행에 개인정보 처리가 필요한 경우 등 제6조에 따라 적법하게 이루어져야 한다고 규정하고 있다.

다만 GDPR 제5조는 공익적인 기록 보존, 과학 및 역사 연구 또는 통계 목적의 개인정보 연계·결합은 원래의 수집 목적과 양립되는 것으로 인정하고 있다. 이 경우 제89조 (1)항에 따라 정보주체의 권리와 자유를 위해 암호화, 가명처리 등 적절한 안전조치를 취할 필요가 있다. 가명처리 정보는 추가 정보를 이용하여 개인을 식별할 수 있는 정보로서 식별할 수 있는 개인정보로 간주된다. 더 이상 개인정보가 아닌 익명정보에는 개인정보보호원칙이 적용되지 않는다. 공익적인 기록 보존, 과학 및 역사 연구 또는 통계 목적을 위해 필요한 경우에 민감정보를 처리하려면 각국 법률에 근거를 두어야 하는데, 이 법률은 추구하는 목적에 비례하고 개인정보보호권의 본질을 존중하며 정보주체의 기본권 및 이익을 보호하기 위해 적절하고 구체적인 조치를 제공해야 한다.

특히 통계 목적의 개인정보 처리와 관련하여 GDPR은 유럽연합과 회원국 통계청이

공식적 통계를 작성하기 위해 수집하는 기밀 정보가 유럽연합 및 회원국 법률로써 보호되어야 한다고 언급하고 있다. 2009년 3월 제정된 '유럽의회 및 각료이사회 규정(EC) No 223/2009 (EU통계규정)'은 통계적 기밀성에 관한 제5장에서 개인정보를 보호하고 있다.

나. 영국

영국의 경우 보건의료서비스에서 2차적 목적으로 기밀 정보(개인 식별이 가능한 정보)를 합법적으로 처리하기 위해서는 해당 조직이 ① 정보주체(환자)로부터 설명에 기반을 둔 동의(informed consent)를 획득하거나 ② 동의를 받지 않아도 되는 법률적 기반이 있어야 한다. 법률적 예외는 일반적으로 '국가보건서비스법(NHS Act 2006)'의 Section 251을 통해서 이루어진다. section 251에 의해 기밀 정보(식별 가능한 환자 정보)에 접근할 수 있으려면, ① 정보 취득자의 목적이 환자 진료의 증진과 관련되어야 하며, ② 공익을 위한 것이어야 하며, ③ 모든 환자로부터 동의를 얻는 것이 불가능하거나 혹은 너무 비용이 많이 들거나, 기술적으로 어려운 경우에만 허용된다. 이 신청은 기밀성 자문 그룹(CAG)에 의해 검토된다. 설명에 기반을 둔 환자들의 동의도 얻지 않았고, section 251에 따른 승인도 얻지 못한 경우, 2차적 데이터의 전송은 익명화되어야 한다.

한편 2017년 봄 입법화된 영국의 '디지털 경제법(Digital Economy Act 2017)'은 제5부 제5장에서 '연구 목적의 공유'를 규정하고 공공기관이 보유한 정보를 연구 목적으로 다른 사람에게 제공될 수 있도록 하였다. 그 정보가 개인정보일 경우, 다음과 같은 조건을 만족해야 한다. ① 그 정보가 특정인을 식별하는 경우, 공개되기 전에 특정인의 신원이 해당 정보 내에서 식별되지 않도록, 그리고 (그 자체로 혹은 다른 정보와 결합하여) 그 정보로부터 특정인의 신원을 추론하는 것이 합리적으로 가능하지 않도록 처리되어야 한다. ② 공개를 위한 그 정보의 처리에 관여하는 모든 사람은 특정인을 식별할 수 있는 우발적인 정보의 공개 위험성을 최소화하고, 그러한 정보의 의도적인 공개를 방지하기 위한 합리적인 조처를 해야 한다. ③ 이 공개는 공공기관에 의해, 혹은 공공기관이 아닌 경우 정보공개를 위한 처리에 관여한 사람에 의해 이루어져야 한다. ④ 정보가 공개되는 연구는 사전 승인을 받아야 한다. ⑤ 공개를 위한 정보의 처리에 관여한 공공기관 및 사람, 정보를 제공받은 사람, 연구 목적으로 그 정보를 이용하는 사람은 사전 승인을 받아야 한다. ⑥ 정보를 공개하거나 그 처리에 관여한 사람은 70조의 실행규약을 유념해야 한다.

다. 독일

독일 보건의료서비스는 '독일사회법전(SGB)'에 의해 규율되고, SGB X은 건강보험 관련 데이터를 포함한 사회 데이터의 보호를 규율하고 있다. 사회 데이터의 전송은 일정한 경우에 허용될 수 있는데, ① 사회서비스 학술 연구 혹은 노동 시장 및 직업에 관한 학술 연구 프로젝트를 위해 필요할 경우, 혹은 ② 공공기관이 자신의 업무와 관련하여 사회서비스 분야의 계획을 위한 프로젝트에 필요한 경우이다. 이때 정보주체의 정당한 이익이 영향을 받지 않아야 하고, 연구 혹은 계획의 공익성이 정보주체의 이익보다 훨씬 커야 한다. 합리적으로 개인의 동의를 얻을 수 있을 때는 해당 개인의 동의 없는 전송이 허용되지 않는다. 위 프로젝트의 개시에 반드시 필요한 정보주체의 성과 이름, 주소, 전화번호, 구조적 특징 또한 설문조사를 위해 전송될 수 있다. 이러한 전송은 최고 연방기관 혹은 해당 데이터가 유래한 지역을 책임지는 주 기관의 사전 승인을 얻어야 한다.

독일 각 주는 2009 '연방암등록데이터법'에 따라 암등록데이터센터(ZfKD)에 암등록 데이터를 제공한다. ZfKD는 신청에 따라 제3자에게 데이터 이용을 허락할 수 있는데 이 경우 정당성, 특히 학술적 이익이 입증되어야 한다. 이용의 범위나 공개는 계약에 의해 규율된다. ZfKD는 일정 기간 후에 통제번호를 삭제하기 때문에, 원 데이터에 오류가 있을 경우에 이를 나중에 식별하거나 수정하는 것이 불가능하다. 실제 개별 데이터로 연결하여 확인할 수 없기 때문이다.

한편, 독일 연방통계의 작성과 관련된 원칙, 조직, 활동을 규율하는 '연방통계법(Bundesstatistikgesetz)'은 13a조에서 통계 목적의 데이터 연계에 관한 사항을 규정하고 있다. 또 제16조(기밀성)에 따라 연방통계 목적으로 제공된 개인에 관한 데이터는 특정한 법률에 명시된 바에 의하지 않고는, 연방통계의 생산을 맡은 공무원 및 공공서비스 종사자에 의해 공개되어서는 안 된다. 다만 관련된 사람이 서면으로 동의한 경우, 공공기관과 관련되거나 연방통계 작성을 위한 법 조항에 근거하여 정보 제공 의무가 있는 경우, 연방통계청 혹은 주 통계청에 의해 다른 응답자의 개별 데이터와 결합되고 통계 결과값으로 제시된 경우, 응답자나 당사자와 관련이 없는 개별 데이터인 경우 등에는 기밀성 의무에서 예외를 인정받는다. 학술 프로젝트의 수행을 목적으로, 연방통계청 및 주 통계청은 고등교육기관 혹은 독립적 학술 연구를 수행하는 다른 기관에 개별 데이터를 제공할 수 있는데, 이 개별 데이터를 통해 응답자 등 개인을 식별하기 위해 비합리적 시간, 비용, 인력이 소요되는 경우, 즉 사실상 익명화된 개별 데이터인 경우에 한한다. 또한, 연방통계청 및 주 통계청의 특별보호구역 내에서, 기밀성을 보호하기 위한 효과적인 조치가 취해진 경우, 공식적으로 익명화된 개별 데이터에 대한 접근을 제공할 수 있다. 공무원, 공공서비스를 위해 특별선서를 한 사

람, 혹은 7항에 따라 기밀성 서약을 한 사람에게만 개별 데이터에 대한 접근 권한이 주어진다. 개별 데이터는 오로지 전송된 목적으로만 사용되어야 한다. 특히 학술 연구 목적으로 전송된 개별 데이터의 경우, 학술 프로젝트가 완료되는 즉시 삭제되어야 한다.

라. 프랑스

프랑스의 경우 '정보, 파일 및 자유에 관한 법률'에 따라, 공공서비스를 관리하는 하나 이상의 법인에 속하고, 서로 다른 공익 목적의 파일들의 연계, 혹은 주목적이 서로 다른 기관에 속하는 파일의 연계 목적으로 자동화된 처리를 하는 경우, 프랑스의 개인정보 감독기구인 CNIL의 허가를 받는다. 또한, 국가등록번호인 사회보장번호(NIR)를 포함한 데이터의 처리도 CNIL의 허가를 받아야 한다. 이때의 정보처리는 오로지 과학 및 역사 연구만을 목적으로 하는 경우에 해당하며, 국가등록번호가 각 연구 프로젝트에 고유한 특정한 임의의 코드로 교체되는 방식으로 사전에 암호화 처리되어야 한다. 암호화 작업 및 그로부터 나온 코드를 통한 파일의 연계는 동일한 처리자가 수행해서는 안 된다. 암호화 작업은 CNIL의 공개된 의견을 받은 후에 국참사원의 시행령으로 규정한 주기로 갱신되어야 한다.

또한, 이 법은 제9장에서 보건의료 분야의 조사, 연구, 평가 목적의 개인정보 처리에 대해 상세하게 규정하고 있다. 보건 분야의 공익적인 연구, 조사 혹은 평가를 목적으로 한 개인정보의 처리는 CNIL의 허가를 받아야 하는데, CNIL은 '공공보건법' L1121-1조에 규정된 인간연구 관련 허가 요청을 위한 자문위원회, 혹은 인간연구 외의 연구 및 평가 신청 승인을 위한 자문위원회의 의견을 받아 결정을 내린다. 신청자가 처리가 예상되는 개인정보 중에 특정 정보의 필요성을 입증할 충분한 증거를 제시하지 못한다면, CNIL은 해당 정보를 보유한 기관에 그 정보의 제공을 금지하고 단지 일부 제한된 데이터의 처리만 허용하도록 할 수 있다.

프랑스 통계법인 '통계 분야의 법적 의무, 조정 및 기밀성에 관한 법률' 제6조는 통계 정보의 기밀성 보호를 다루고 있으며, 사적 성격의 사실과 행동에 관련된 설문조사의 개별 데이터는 원칙적으로 설문이 수행된 후 75년, 혹은 당사자 사망 이후 25년 동안 공개되지 않도록 규정하고 있다. 다만, 공식통계 혹은 학술적·역사적 연구 목적을 위해, 통계 기밀성 위원회의 의견을 청취한 후, 아카이브 행정당국의 결정에 따라 이루어질 경우에는 예외이다.

개인건강정보의 전송 절차는 절대 개인 식별을 허용해서는 안 된다. 다만 '정보, 파일 및 자유에 관한 법률'에 따라, 통계상 직간접적 식별요소가 필요할 경우, 특히 서

로 다른 개인 소스의 데이터 결합 목적을 위해서만 예외가 허용된다. 이 목적을 위해 데이터 처리 허가를 받은 법인의 지정 책임자만이 통계청에 해당하는 INSEE 혹은 공공보건정책에 참여하는 부처의 통계부서에 전송된 개인건강정보를 전송받을 수 있다. 데이터 사용 후에는 개인식별요소가 폐기되어야 한다.

마. 미국

미국에는 공공과 민간을 모두 포괄하는 개인정보 보호법제가 없다. 연방정부가 보유한 개인정보를 규율하는 '프라이버시법'은 연방기관의 기록 시스템에서 유지되는 개인식별정보(Personally Identifiable Information)의 처리를 규율한다. 원칙적으로 어떠한 기관도 개인의 서면 요청 혹은 사전 서면 동의가 없으면 보유 개인정보를 다른 사람이나 기관에 공개해서는 안 된다. 다만, 12가지의 법정 예외를 두고 있는데, title 13 조항에 따른 인구조사, 설문조사, 관련 활동을 계획 혹은 수행하기 위한 목적으로 인구조사국에 제공하는 경우, 해당 기관에 사전에 해당 기록이 오로지 통계 조사 혹은 보고기록으로만 사용될 것임을 서면으로 확인받고 개인식별이 불가능한 형태로 전송할 경우가 이에 포함된다. 이 조항에 따라 공공기관의 기록은 통계 목적으로 인구조사국에 전송될 수 있다.

한편 보건의료 관련 법률인 '건강보험 양도 및 책임법(HIPAA)'은 '보호되는 건강정보(Protected Health Information, PHI)'로서 개인식별이 가능한 건강정보(individually identifiable health information)를 규정하고 있으며, 원칙적으로 이 법 적용을 받는 기관이 치료, 지불, 보건의료 운영 등의 촉진 목적으로만 환자의 서면 허가 없이 개인식별이 가능한 건강정보를 이용 및 공개할 수 있다. 다른 목적으로 이용할 경우에는 정보주체의 서면 동의가 있어야 한다. HIPAA 프라이버시 규칙은 연구 목적의 PHI의 이용 혹은 공개를 다음과 같은 조건 하에 허용한다. 원칙적으로 정보주체의 서면 허가 없이는 PHI를 이용 및 공개할 수 없지만, ① 기관평가위원회 혹은 프라이버시 위원회로부터 허가 예외의 승인을 받은 경우, 연구 준비 평가를 위한 경우, 사망자 정보에 관한 연구인 경우, ② 이름, 주소, 전화번호, 사회보장번호 등 18개 식별자를 제거하는 등 PHI가 비식별화된 경우, ③ 연구, 공중보건, 보건의료 운영의 목적으로 데이터 이용계약이 체결되고 특정한 직접 식별자가 제거된 '제한된 데이터셋'의 형태로 제공될 경우에는 서면 허가 없이 PHI를 연구 목적으로 이용 및 공개할 수 있다. HIPAA 법의 적용을 받는 기관은 제한적이기 때문에 건강보험업체, 보건의료 제공자, 공공기관 등에 소속되지 않은 연구자의 경우 프라이버시 보호를 위한 어떠한 법적 규율도 받지 않는다.

미국의 통계 관련 법률로서 '기밀정보의 보호 및 통계 효율성에 관한 법률'은 미국

통계 관련 기관들 사이에 통계 목적으로 수집되는 정보에 대해 단일한 기밀성 보호 기준을 수립하고, 노동통계국, 경제분석국, 인구조사국 등 지정통계기관 사이에 일부 데이터의 공유를 허용하고 있다. 기밀 정보의 보호를 다루는 이 법의 A절에서는, 통계 목적의 데이터 및 정보는 오로지 통계 목적으로만 이용할 것을 규정하고 있고, 이 정보들은 응답자의 동의가 있을 경우를 제외하고는 통계 목적 외 사용을 위해 개인 식별이 가능한 형태로 제공되지 않으며, 이러한 공개는 해당 기관이 승인했을 경우에만 가능함을 규정하고 있다. '인구조사법'은 미 상무부 장관이 이 법에 따른 업무와 관련된 정보를 다른 부처, 기관 등에 요청할 수 있고, 필요한 기록, 보고 등의 자료 복사본을 주, 시 등 다른 정보 단위 혹은 민간의 개인이나 기관으로부터 획득할 수 있도록 하였다. 정보의 기밀성에 관한 제9조는 법에서 규정한 예외를 제외하고는 누구도 이 법에 따라 수집된 정보를 본래 통계 목적 외의 목적으로 사용해서는 안 되고, 특정한 개인 혹은 기관이 식별되는 방식으로 공개되어서는 안 된다는 점 등을 규정하고 있다.

바. 뉴질랜드

뉴질랜드는 통계청이 개인정보의 수집, 저장, 이용에 관해 '개인정보보호법 1993'에서 규정한 프라이버시 원칙을 준수하도록 하였다. 통계법은 응답자(개인정보제공자)에게만 적용되고, 개인정보보호법은 응답자, 통계청 직원, 연구자 등 통계청 서비스이용자 모두에게 적용된다. 예를 들어 개인정보 공개 제한 원칙에 따라 몇 가지 예외를 제외하고는 개인정보가 제3자에게 공개(제공)되지 않아야 한다. 통계청도 통계법에 따라 응답자의 동의 없이는 통계청 밖으로 응답자의 정보를 제공하지 않는다. 그러나 정부 통계관의 승인에 따라, 비식별화된 정보를 승인된 연구자에게, 승인된 연구 목적을 위해, 안전한 환경에서 제공할 수 있다. 또한, 고유식별자 보호 원칙에 따라 통계 및 연구 목적의 데이터에서 조세번호, 운전면허증 번호, 여권 번호 등 개인식별정보를 제거하며, 통계청의 임의의 고유식별자로 대체된다.

뉴질랜드 '통계법' 제37조(정보 보안)는 통계청이 수집한 정보에 대하여 통계 목적으로만 사용되어야 하고, 개인에 대한 세부 정보 혹은 질의에 대한 응답에는 통계청 직원 외에 접근하거나 공개되지 않아야 하고, 통계청이 공개하는 모든 통계 정보는 (개인이 동의했거나, 합리적으로 예측할 수 없는 불가피한 경우가 아니면) 세부 정보가 식별될 수 있는 방식으로 공개되어서는 안 된다고 규정하였다. 제공자의 허락, 공동 설문조사, 연구 및 통계 목적의 공개, 역사적 문서의 경우에는 예외적으로 마이크로데이터에 대한 접근이 허용된다.

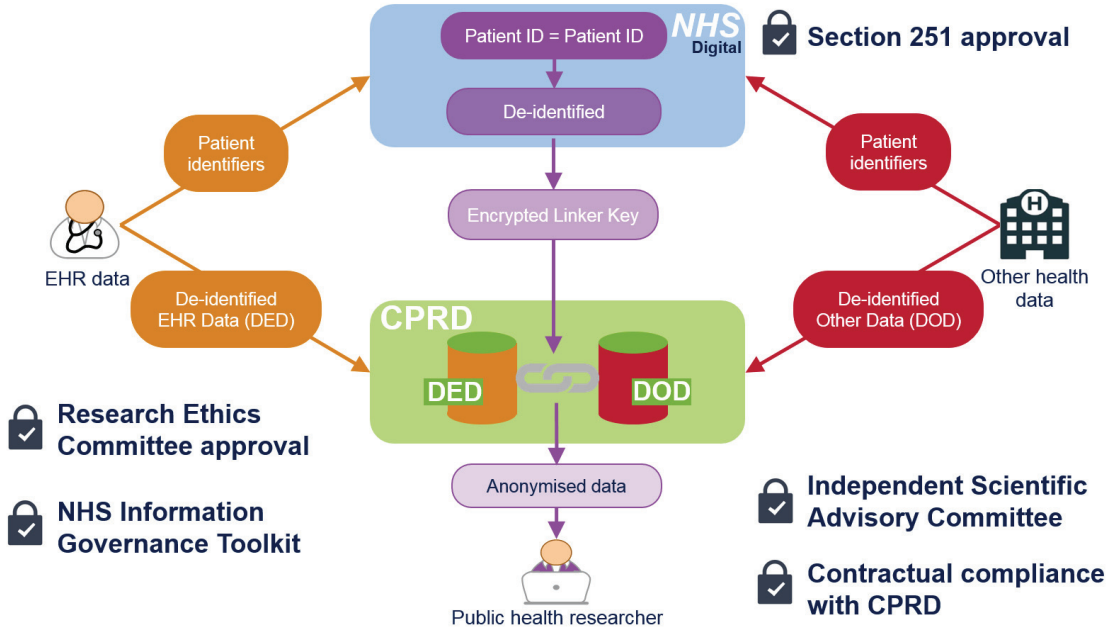
2) 보건의료 분야 데이터 연계 현황

가. 영국

영국은 국가보건서비스(NHS)를 통한 공공영역이 보건의료서비스를 담당하고 있기 때문에 OECD 국가 중 가장 광범한 국가적 보건의료 데이터셋을 보유하고 있다. 다만 데이터 수집 및 접근 체제가 잉글랜드, 웨일즈, 스코틀랜드, 북아일랜드 등 지역별로 분할되어 있다.

잉글랜드 지역 보건의료 분야 데이터 연계는 '임상시험연구데이터링크(CPRD)'가 대표적이다. CPRD는 비영리 연구지원을 목적으로 하는 복지부 산하 기관이며, 1987년부터 공공보건 연구를 위해 익명화된 1차 진료기록을 제공해왔다. CPRD는 NHS 번호, 이름, 전체 생년월일, 주소, 의사들의 진료메모 등은 수집하지 않는다. 일반의는 환자들의 비식별 데이터를 CPRD에 제공할지 여부를 선택할 수 있으며, 개인 환자가 원하지 않을 경우 CPRD에의 개인정보 제공을 거부할 수 있다(opt-out). 데이터 연계는 다음과 같은 절차에 의해서 이루어진다. ① 매년 CPRD는 공공보건연구 목적으로 익명화된 연계 데이터를 제공하는 것에 대해 보건연구당국의 Section 251 규제 승인을 받아야 한다. ② 데이터 연계는, 환자식별정보를 합법적으로 수집할 권한을 가진 잉글랜드의 법정 기구인 NHS Digital에 의해 이루어진다. 즉, NHS Digital 이 '신뢰할 수 있는 제3자(TTP)'의 역할을 수행한다. ③ 연계를 위해, 일반의가 수집한 환자식별정보(NHS 번호, 생년월일, 우편번호, 성별)와 다른 데이터셋의 식별정보가 NHS Digital로 보내진다. ④ NHS Digital은 두 데이터셋에서 환자식별정보를 매칭하여, 환자식별정보를 포함하지 않은 암호화된 연계키(encrypted linker key)를 산출한다. ⑤ NHS Digital은 CPRD가 비식별 데이터셋을 연계할 수 있도록 암호화된 연계키를 CPRD에 보낸다. ⑥ CPRD는 일반의나 NHS Digital로부터 절대 환자식별정보를 받지 않는다. ⑦ 공공보건연구 목적으로 연계 데이터에 접근하고자 하는 연구자는 독립적 과학자문위원회(ISAC)의 승인을 받아야 한다. ⑧ ISAC의 승인에 따라 연구자에게 익명화된 데이터셋을 제공하기 이전에 추가로 암호화한다. 라이선스를 가진 연구자가 ISAC이 승인한 연구를 수행하는 경우에는 CPRD가 보유한 익명화된 1차 진료 데이터베이스에 온라인으로 접근할 수 있다. CPRD 데이터 접근을 위한 라이선스 이용조건은, 데이터 접근이 공익적인 의료연구 목적으로 제한되고, 데이터셋은 승인된 연구로만 사용되어야 하며, 환자, 병원, 의사에 대한 식별 시도가 특히 금지된다는 점 등이 규정되어 있다.

그림 요약-1 CPRD 데이터 연계 절차



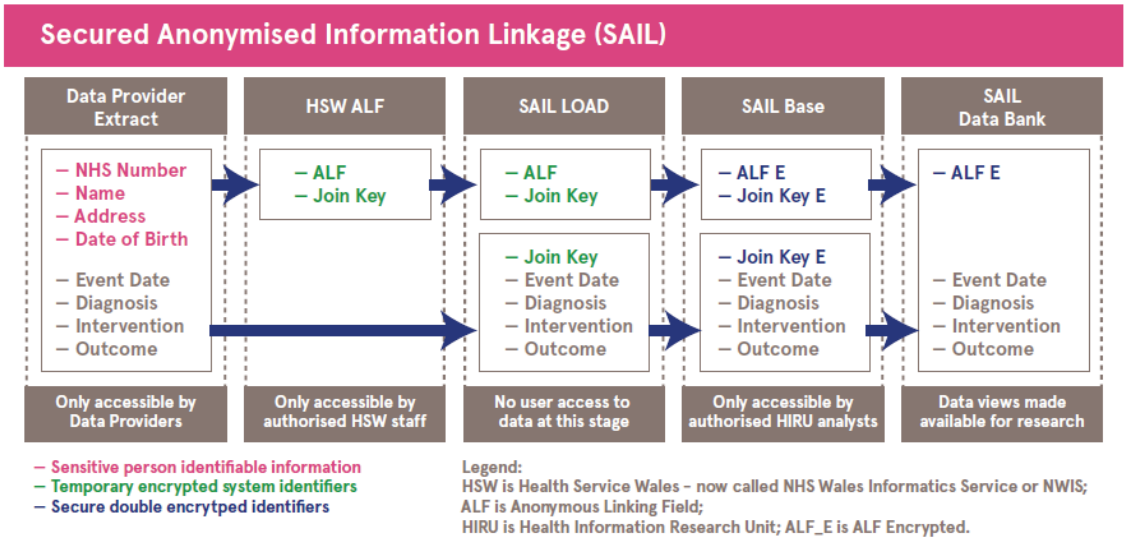
영국의 다른 지역도 CPRD와 유사한 보건의료 데이터 거버넌스 체계를 갖추고 있다.

영국 웨일즈 SAIL DataBank의 경우에는 보건 관련 연구를 위해 익명화된 개인 기반 데이터의 저장 및 이용을 제공한다. SAIL Databank 는 CPRD의 경우와 마찬가지로 개인식별정보를 보유하고 있지 않으며 익명화된 데이터만 보유하고 있는데 연구자의 요청에 따라 데이터 연계를 제공한다. 다만 연계를 위하여 신뢰할 수 있는 제3자(TTP)의 역할을 하는 NHS 웨일즈 정보서비스(NWIS)가 연계를 위한 결합 키값인 익명연계필드(ALF)를 생성한 후 SAIL Databank에 제공한다. SAIL Databank는 ALF를 다시 암호화하여 서로 다른 익명 데이터셋을 연계하는데 사용한다. SAIL Databank 역시 공익에 기여할 수 있는 순수 연구목적으로만 데이터에 대한 접근을 허용하고 있다. 데이터에 접근하기 위해서는 독립적인 정보거버넌스검토패널의 승인을 받아야 한다. 연구 제안서가 승인되면, 모든 연구자는 데이터에 접근하기 전에 적절한 정보 거버넌스에 대한 훈련을 받아야 하며 훈련 자격증은 2년간 유효하다.

데이터 접근을 원하는 연구자들은 훈련을 마친 후 SAIL Gateway 사용 승인을 받는다. SAIL Gateway는 보안이 되는 안전한 환경에서 연구가 수행될 수 있도록 프라이버시 보호를 위한 안전시설이자 원격접근시스템으로 마련된 것이다. 연계 데이터를 바로 제공하기보다는 SAIL Gateway를 통해 접근하도록 하는 것은 혹시 있을 수 있는 연계 공격에 대비하기 위한 것이다. SAIL Gateway 사용이 승인된 연구자들은

SAIL 데이터 접근 계약을 체결하며 연구자가 요청한 특정 데이터는 '읽기전용'으로 접근할 수 있다. 연구가 완료되면 연구자들은 그 개인정보 침해 위험성에 대해 SAIL 데이터 관리자의 검토를 받은 후 자신의 연구결과물을 SAIL Gateway에서 가지고 나갈 수 있다.

그림 요약-2 SAIL 데이터 연계 절차



영국 스코틀랜드 전자데이터연구혁신서비스(eDRIS)도 연구자에 대한 승인 절차를 엄격하게 규정하고 있는데 연구자 훈련, 연구에 대한 승인, 데이터 공유계약을 이행하도록 하고 각 연구는 데이터 보유기관의 허락을 받는 데 필요할 경우 윤리 승인을 획득해야 한다. 특히 이런 절차들과 별도로 연구자는 승인된 기관(Approved Organization) 소속이어야 한다. 공익목적의 연구를 위한 데이터 연계를 지지하면서도 상업적 접근에 대한 대중적 우려가 크기 때문에 '승인된 기관'의 연구자들만 데이터에 접근하도록 한 것인데 승인된 기관은 현재 대학, NHS, 지역 당국 및 스코틀랜드 정부 등 공공영역의 기관으로 제한되어 있으며 영리 업체, 미디어, 로비 그룹, 제3섹터 기관들은 데이터에 직접 접근할 수 없다.

한편, 2016년 7월 영국 보건의료 빅데이터 care.data 프로그램의 운영이 취소되었다. care.data는 기존에 병원 환자 정보를 보유하고 있던 HSCIC의 국가 데이터베이스에 일반의가 보유한 환자정보까지 집적하려는 NHS 잉글랜드의 사업으로 2013년 시작되었다. 일반의 시스템으로부터 추출하는 데이터는 이첩, NHS 처방전, 가족력, 예방접종, 혈액검사결과, 체질량 지수, 흡연/음주습관 등이다. care.data의 데이터는 직접적인 진료 목적이 아닌, 의료 서비스 계획이나 의학 연구 등 2차적 목적을 위해 활용될 예정이었다. NHS 내 기관뿐 아니라, NHS 외부의 제약회사, 보건 자선단체, 대학, 병원

위탁단체, 싱크탱크 및 다른 사기업 등에 제공될 수 있었다. 환자들은 일반의가 보유한 자기 정보를 HSCIC에 전송하는 것에 대해서, 혹은 HSCIC로 전송된 데이터를 제3자에게 전송하는 것에 대해서 거부권(opt-out)을 행사할 수 있다. 그럼에도 불구하고 care.data 사업은 많은 사회적 반발을 가져왔다.

논란이 일자 2015년 9월 영국 복지부는 '복지 질 위원회'에 NHS의 데이터 보안에 대한 검토를, 보건복지를 위한 국가데이터 가디언인 칼디콧에 데이터 보안 및 동의에 대한 독립적 검토를 요청했고, 보고서 권고에 따라 care.data 프로그램이 일단 취소되었다. care.data 사업의 실패는 개인정보의 수집·활용을 촉진하기 위한 사업에서 보안 및 개인정보보호에 대한 대중의 신뢰를 얻는 것이 얼마나 중요한지, 그리고 해당 사업을 추진하는 당국이 사업의 내용을 투명하게 공개하고 관련 이해당사자와 충분히 협의하는 것이 얼마나 중요한지를 보여준다.

나. 호주 PHRN

호주의 인구보건연구네트워크 PHRN은 보건 정보를 안전하게 관리할 수 있는 국가적인 데이터 연계 기반을 구축하기 위해 설립되었다. PHRN 연계 데이터에 접근할 수 있는 연구의 자격 요건은, 공중 보건의 증진·보호·유지에 기여할 것, 보건 서비스의 계획·평가·전달을 촉진할 것, 보건 데이터 수집·보건 관련 데이터 연계·보건 관련 통계의 편집 및 활용과 관련된 연구 방법 증진에 기여할 것 등이다. 연계 데이터를 사용하는 연구 프로젝트는 데이터 연계기구, 각 데이터셋을 보유한 데이터 보유기관들, 인간연구윤리위원회 등 세 기관의 승인을 얻어야 한다.

데이터 연계 절차로는 우선 연구자가 데이터 연계기구에 데이터 연계 프로젝트를 신청한다. 프로젝트가 승인되면 데이터 연계기구는 관련 데이터 보유기관에 데이터 연계를 위한 개인 식별자로서 연계 변수를 요청한다. 데이터 보유기관은 식별정보와 함께 레코드 ID나 환자 ID를 생성하여 데이터 연계기구에 보내는데, 프라이버시 보호를 위해 이 ID는 본래의 레코드 ID나 환자 ID가 아니라 프로젝트 고유의 ID가 권장된다. 데이터 연계기구는 연계 ID인 프로젝트 키를 생성하여 각 데이터 보유기관에 전달한다. 각 데이터 보유기관은 프로젝트 키가 결합된 콘텐츠 데이터 파일을 연구자에게 전달한다.

PHRN에서 개인정보 보호를 위한 중요한 원칙 중 하나가 '분리 원칙(separation principles)'이다. 그 의미는 다음과 같다. ① 연계 데이터와 콘텐츠 데이터의 분리. 연계를 위한 개인정보는 콘텐츠 데이터와 분리되어, 데이터 연계자에게는 연계 ID 생성을 위한 개인정보만 제공된다. ② 기능과 책임의 분리. 데이터 연계 절차와 데이터 보

유 및 추출 기능을 분리한다. 데이터 연계를 수행하는 사람은 연구자와 분리되어 있어야 하며, 콘텐츠 데이터 연구에 참여할 수 없다. 정보의 이용, 제공, 보유 또한 제한된다. 연구자는 특정 프로젝트를 위한 정보에만 접근이 허용되며, 승인받은 방식으로 이용해야 한다. 프로젝트의 연구자들은 신원을 확인받고 승인되어야 하며, 다른 사람에게 정보를 주어서는 안 된다. 정보는 승인받은 기간 동안만 보관되며, 그 이후에는 데이터 보유자에게 반환하거나 삭제된다.

다. 미국

보건의료서비스를 민간 제공자와 보험사가 주도하고 있는 미국은 보건의료 데이터의 수집 및 보관도 다양하게 분산되어 있다. 연구 목적의 데이터 접근을 위한 계약 체결도 개별 업체를 매개로 해야 한다. 보통 사례별로 이용허락이 이루어지며, 연구 출판 전에 연구결과에 대한 사전 승인을 요구할 수 있다. 민간 영역의 데이터 연계는 통상 업체 내에서 이루어지며, 다른 업체와의 데이터 연계에 제한을 두고 있다. 일반적으로 비식별 데이터로부터 개인 식별을 시도하거나 다른 데이터 소스와 연계하는 것을 제한한다.

미국 국가보건통계센터(NCHS)는 자신이 보유한 데이터의 식별자를 제거한 후 공개사용 데이터 파일로 만들어 공개하고 있다. 미국의 '공공보건서비스법' Section 308 (d) 등의 규정에 따라 공개사용 데이터 파일의 이용자는 다음 조건에서 이를 사용해야 한다. ① 데이터셋은 통계 보고 및 분석 목적으로만 사용해야 한다. ② 의도하지 않게 개인 및 기관의 신원이 노출될 경우 이를 사용하지 말고 NCHS 책임자에게 알려야 한다. ③ 이 데이터셋을 개인 식별이 가능한 다른 NCHS의 데이터나 외부 데이터와 연계해서는 안 된다. ④ 공개사용 데이터를 이용하는 것은 위의 법적 요구조건을 준수하겠다는 '데이터 이용자 계약'에 서명한 것이 된다. 또 NCHS는 연구자를 위해 이름, 사회보장번호, 주소 등 직접 식별자는 제거했지만, 지리정보 등 간접 식별자를 포함한 '제한된 데이터(restricted data)'에 대한 접근을 허용하기 위해 연구데이터센터(RDC)를 두고 있다. 연구데이터센터를 이용할 때는 연구 제안서 승인, 승인 범위 안에서 분석 및 연구, 결과물 공개에 대한 제한 등에 대해 연구데이터센터와 협의해야 한다. 개인 수준의 데이터는 센터의 시설로부터 가지고 나갈 수 없으며 전자기기를 사용할 수 없다.

3) 연구 목적 데이터 연계 현황

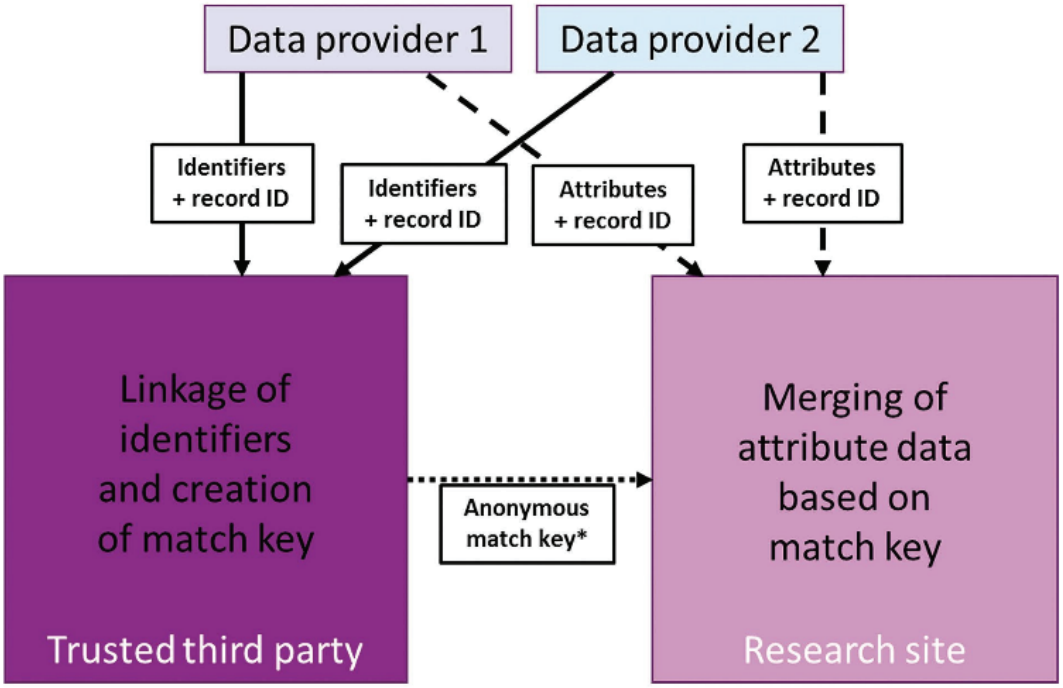
가. 영국 ADRN

영국의 행정데이터연구네트워크(ADRN)는 사회, 경제 연구자들에게 안전한 환경에서, 연계된 비식별 행정 데이터에 대한 접근을 제공하기 위한 네트워크이다. ADRN은 연구자들을 대신하여 요청할 데이터의 범위를 검토하고, 데이터 보유기관과 협의를 진행하며, 신뢰할 수 있는 제3자(TTP)를 통해 데이터를 연계하고, 연계된 비식별 데이터에 접근할 수 있는 보안 환경을 제공하는 기구로서, 행정 데이터를 직접 보유하고 있는 것은 아니다.

ADRN은 영국 데이터 서비스가 만든 '데이터 공유를 위한 5가지 안전 원칙'에 따라 운영된다. 이는 데이터 비식별화 및 TTP 데이터 연계 등 데이터 안전(Safe data)는 물론 인력 안전(Safe people), 연구 안전(Safe project), 환경 안전(Safe environment), 결과물 안전(Safe results) 등의 원칙이다.

ADRN의 데이터 연계는 다음 절차를 따라 이루어진다. ① 연구 제안서의 승인이 이루어지고 연구자가 훈련을 받은 후, ADRN은 프로젝트 관련 데이터의 제공을 위해 데이터 보유기관과 협상을 진행한다. ADRN에 신청하는 모든 연구 프로젝트는, 비영리적 연구 목적이어야 하고, 명확한 과학적 이익과 잠재적인 공익이 있음을 입증해야 하며, 연구 과제에 대한 답을 찾기 위해 부서 수준의 행정 데이터가 필요하다는 것을 보여주어야 한다. 승인 패널은 연구 프로젝트가 윤리적인지, 합법적인지, 실현 가능한지, 과학적 가치가 있는지, 사회에 이익이 되는지를 결정한다. 연구자의 자격요건은 학계, 공공영역, 공동체 혹은 자원봉사 영역, 연구기관 소속이어야 한다. 연구자는 안전한 연구 데이터 이용자 환경 훈련에 참여해야 하고 이용약관과 위반정책에 서명해야 한다. ② 데이터를 보유한 정부 부처는 각 레코드에 고유한 참조번호(reference number)를 부여한다. 그리고 이름, 생년월일 등 사람들을 직접 식별할 수 있는 식별자를 분리한다. ③ 데이터 보유기관들은 식별정보를 고유 참조번호로 대체한 데이터를 ADRC 에 보낸다. 직접 개인을 식별할 수 있는 정보는 데이터와 분리하여 각 레코드의 고유 참조번호와 함께 신뢰할 수 있는 제3자(TTP)에게 보낸다. ④ TTP는 고유 참조번호와 식별정보를 사용하여 이 정보들을 매칭한다. 그리고 개인 식별정보를 삭제한 후 매칭된 고유 참조번호만을 남긴다. ⑤ 색인키(index key)는 서로 다른 데이터 집합에서 어떤 참조번호가 같은 사람과 관련되는지를 보여준다. TTP는 색인키를 ADRC에 보낸다. ⑥ ADRC는 색인키를 사용하여 서로 다른 기관들이 보내온 데이터 집합을 연계한다. 그리고 색인키와 참조번호를 지운 후에 연구자에게 연계된 데이터에 대한 접근을 제공한다.

그림 요약-3 ADRN 식별자와 속성 데이터의 분리



ADRN은 훈련된 연구자들이 연계된 비식별 행정 데이터에 접근할 수 있는 보안 시설을 제공한다. 연구자들은 보안 시설에 전자기기나 필기도구를 갖고 들어가거나 갖고 나올 수 없으며 데이터셋 자체를 가지고 나갈 수도 없다. 연구자가 최종 결과물을 보안 시설로부터 갖고 나가기 이전에, 그 결과물이 승인된 연구 프로젝트에 관련된 것인지, 개인을 직접 식별할 수 있는 정보를 포함하고 있는지 등을 철저하게 검증한다. 연구자들은 연구요약을 ADRN에 제공하며, 결과물을 공개하고 관련 데이터 보유기관에도 제공한다.

이 시스템은 개인 식별정보와 연구 데이터의 분리를 유지한다. 즉, TTP는 단지 식별정보와 참조번호만을 볼 수 있으며, 연구 데이터를 볼 수 없다. ADRN 직원은 단지 연구 데이터와 색인키만을 볼 수 있을 뿐, 개인 식별정보는 볼 수 없다. 연구자는 보안 시설에서 자신이 요청한 데이터만을 볼 수 있으며, 색인키와 개인 식별정보는 볼 수 없다.

나. 독일 GRLC / FDZ

독일레코드연계센터(GRLC)는 사회과학 분야의 학술 연구를 위해, 행정 데이터를 이용한 데이터 연계 활성화를 목적으로 2011년 설립되었다. 연계 프로젝트별로 원 데

이터베이스의 소유권이나 법적 요건이 다르며, GRLC에 데이터 연계를 요청하는 기관이 연계 데이터의 모든 요소에 대한 권한을 갖고 있거나 이용허락을 받았을 경우 연계 데이터 접근에 별다른 제한이 발생하지 않는다. GRLC는 프라이버시를 보호하면서도 효과적으로 데이터 연계를 수행할 수 있는 방법에 대한 연구를 하고 있는데, 이 연구 분야를 프라이버시 보호 데이터 연계(privacy preserving record linkage, PPRL)라고 한다.

독일연방고용국 연구데이터센터(FDZ)는 정부프로젝트 연구자에게 사회보장 및 고용 분야에서 비영리적 실증 연구를 위해 마이크로데이터에 대한 접근을 제공한다. FDZ는 연구자의 데이터 접근을 위해, 현장 이용, 원격 데이터 접근, 학술적 이용 파일 등 세 가지 방법을 제공한다. 이 세 가지 방법은 데이터의 익명화 정도 및 이용조건에서 차이가 있다. 일반적으로 데이터에 대한 접근은 비영리적 연구 목적으로 제한된다.

4) 통계 목적 데이터 연계 현황

가. 미국 Data Linkage Infrastructure

미국은 2016년 3월, 증거기반 정책결정위원회법(Evidence-Based Policymaking Commission Act of 2016)을 통과시켰는데, 관련하여 미국 인구조사국은 평가자 및 정책 분석가의 행정 데이터 접근을 증진하기 위해 데이터 연계기반(Data Linkage Infrastructure)을 확대해 왔다.

인구조사국이 데이터셋을 연계할 경우, 이 연계가 기관의 임무에 부합하려면 가장 좋은 대안(Best Alternative), 공익성, 민감성, 데이터의 기밀성을 고려해야 한다.

데이터 연계기반의 데이터를 사용하기 위해서, 연구자는 우선 제안서를 제출해야 한다. 이 프로젝트가 외부의 데이터를 데이터 연계기반으로 가져올 경우, 프로젝트 제안자는 그 외부 데이터의 이용과 전송을 허가하는 (특히 개인식별정보의 이용을 명시한) 서신을 제출해야 한다. 연구 제안서에 대해서는 학술적인 가치, 실행 가능성, 잠재적인 공개 위험성 등이 평가된다. 인구조사국의 데이터(Title 13) 이용을 요청하는 프로젝트의 경우에는 해당 연구가 인구조사국에 갖는 가치를 입증해야 한다.

행정 기록 프로젝트에 관한 연구자는 현재 인구조사국의 피고용인이거나, 특별선서지위(SSS)를 획득해야 한다. 인구조사국은 자신의 프로그램에 명백히 이익이 되는 작업을 수행하는 개인들에 SSS 자격을 부여하는데, SSS 자격을 가진 개인은 인구조사국 직원과 마찬가지로 평생 데이터를 보호하고 동일한 법적 의무와 처벌을 감수할 것을 선언해야 한다. 인구조사국 직원과 SSS 자격을 가진 개인들은 매년 개최되는 데

이터 관리 훈련과 Title 26/연방조세정보 훈련 등 다른 데이터 보유기관이 요구하는 훈련을 받아야 한다. 또한, 관련 윤리, 기밀성, 프라이버시 보호 절차를 준수해야 한다.

제안서의 승인과 훈련이 완료되면, 연방조사국은 보안 컴퓨팅 환경 내에서 접근을 제공한다. 대부분 연방통계연구데이터센터에서 하게 되는데, 마이크로데이터에 대한 모든 분석은 이 컴퓨팅 환경에서 이루어져야 한다. 연구자들은 승인된 데이터 파일의 읽기전용 비식별화 버전에 접근하게 된다.

연구가 마무리된 후, 그 결과물은 공개되기 전에 개인정보나 기업정보가 포함되어 있지 않은지, 결과물이 애초의 제안서와 일치하는지 검토된다. 인구조사국 데이터의 경우, 공개 회피 사무관 혹은 전면공개평가위원회에서 평가를 수행한다. 관련 데이터 보유기관도 결과물을 검토한다.

나. 네덜란드 SSD 시스템

네덜란드 사회통계데이터셋(SSD) 시스템은 상호 연계되고 표준화된 등록소 및 설문조사시스템이다.

허가받은 연구자들이 엄격한 조건 하에, 네덜란드 통계청이 보유한 상세 데이터에 접근할 수 있다. 통계청 구내 장소에서 접근하거나, 보안 인터넷 접속을 통해 원격 접근할 수 있다. 네덜란드 통계청은 네덜란드 통계법과 프라이버시 관련 법률을 준수해야 한다. 개인정보보호법은 애초 수집 목적 외의 개인정보 처리를 금지하고 있지만, 역사적, 통계적, 학술적 목적으로 개인정보를 처리할 수 있는 예외를 두고 있다.

통계법은 개인정보보호를 위한 규정도 포함하고 있다. 이에 따르면, ① 통계청이 받는 모든 데이터는 통계적 목적으로만 사용되어야 한다. ② 통계청은 데이터 보호를 위한 기술적, 조직적 조처를 해야 한다. ③ 공개된 결과물이 개인정보를 드러내지 않도록 조처를 해야 한다. ④ 통계법에 따른 책임을 수행하는 사람을 제외하고는, 다른 사람에게 데이터를 전달해서는 안 된다. 다만, 이에 대한 예외로서, 통계청은 통계적 혹은 학술적 연구 목적으로 다른 기관에 마이크로데이터를 제공할 수 있다.

다. 캐나다 SDLE

캐나다 통계청의 데이터 연계는 마이크로데이터 연계 지침에 따라 이루어지며, 이는 ① 캐나다 통계청 내에서 수행 중인 데이터 수집 및 방법론적 연구의 설계, 유지, 평가, 연구 및 재설계 지원, ② 연구 조사를 지원하기 위한 총계 혹은 익명 형식의 통

계 정보 제공이라는 두 가지 목적에서 이루어진다.

캐나다 통계청의 사회 데이터 연계환경(SDLE) 프로그램은 데이터 연계를 위한 환경으로 사회경제적 통계 연구를 촉진하기 위한 목적으로 캐나다 전역에 만들어졌으며 고도로 보안을 갖춘 환경이다. 그러나 SDLE가 거대한 통합 데이터베이스는 아니며 그 핵심인 파생기록보관소(DRD)도 단지 기본적인 개인 식별자만을 가진 동적 관계 데이터베이스이다. 데이터 연계를 위해 사용된 SDLE 식별자와 소스색인파일의 레코드 ID 조합은 키 등록부(Key Registry)에 저장된다.

학술적인 가치 및 연구의 실행 가능성, 적용 방법과 분석될 데이터의 관련성, 상세 마이크로데이터에 접근할 필요성의 입증, 연구자의 전문성과 능력 등의 원칙에 따라 연계 데이터를 요구한 연구에 대해 승인이 이루어진다. 연구 승인 후, 연구에 필요한 특정 집단의 레코드 ID와 이에 연계된 파일들의 관련 레코드 ID가 키 등록부에서 추출된다. 추출한 관련 레코드 ID들이 소스데이터 파일에서 개인의 기록을 찾는 데 사용된다. 이 레코드 ID는 분리된 소스데이터파일로부터 선택된 속성정보들을 추출하여 연계 분석 파일을 만들기 위해 사용된다. 이런 방식을 통해 거대한 통합 데이터베이스를 구축할 필요 없이 가상의 연계환경을 제공하게 된다.

캐나다 통계청은 연구자가 요청하는 데이터의 성격에 따라 다양한 데이터를 제공한다. 공개사용이 가능한 마이크로데이터 파일(PUMFs)의 경우에는 공공기관 및 캐나다 통계청이 협약하는 고등교육기관의 교수와 학생 등에 접근이 가능한데 이 파일들은 개인 수준의 데이터가 없는 익명화된 형태이다. 한편 연구를 위해 연구데이터센터(RDC)가 캐나다 전역에 있는데, RDC 데이터는 총계 처리되지 않은, 개인 수준의 데이터에 대한 접근을 제공한다. 그래서 승인된 프로젝트의 연구자만이 접근할 수 있으며, 연구자들은 통계법상 '직원 간주'의 지위를 획득해야 한다. RDC는 캐나다 통계청 직원이 관리하고 있으며, 통계법 조항의 기밀성 규정에 따라 여느 통계청 사무실과 마찬가지로 운영된다. 물리적인 접근이 제한되고, 컴퓨터들은 통계청 외부와 연결되지 않는다.

마이크로데이터에 접근하는 모든 연구자는 캐나다 통계청의 신뢰성 검사를 받고 캐나다 연방 경찰에 보안 검사를 위한 지문도 날인하는 등 보안 절차를 거쳐야 한다. 오리엔테이션을 마친 연구자들은 통계청과 계약을 체결하고 서약한다. 연구자는 캐나다 통계청에 연구결과물을 제출해야 한다.

라. 뉴질랜드 IDI

뉴질랜드 통계청은 자신이 보유한 데이터를 통한 연구를 활성화하기 위해 통합데이

터기반(The Integrated Data Infrastructure, IDI)을 운영하고 있다.

IDI는 ① 모든 소스로부터 데이터를 수집하고, ② 수집된 데이터를 처리·연계하며, ③ 비식별화된 데이터를 연구 목적으로 제공한다. 애초 수집 목적 외의 정보 이용에 대해서는 개인정보보호법을 고려해야 한다. 통계법 1975에 상응하거나 제한하는 명백한 법 조항이 있는 경우에는 프라이버시 감독관과 협의해야 한다.

IDI 데이터에 대한 접근은 통계 목적 혹은 공익목적의 순수한 연구 프로젝트에만 허용된다. 연구자들은 자신의 연구 프로젝트에 필요한 데이터에만, 그리고 보안 데이터랩을 통해서만 데이터에 접근할 수 있다. 기밀성 규칙을 보장하기 위해, 연구자가 데이터랩에서 가져가기를 원하는 모든 연구결과물에 대해 검사가 이루어진다.

데이터 통합은 다음의 원칙을 만족할 경우에만 이루어진다. ① 통합으로 인한 공익이 데이터 이용에 대한 프라이버시 우려와 통계 시스템의 무결성, 원 소스데이터, 다른 정부의 활동에 대한 위험보다 커야 한다. ② 통합 데이터는 통계 혹은 연구 목적으로만 사용되어야 한다. ③ 데이터 통합은 개방적이고 투명한 방식으로 수행되어야 한다. 데이터 통합 행위와 결정에 대해 투명해야 하고, 사람들에게 데이터 통합의 목적과 방법을 적극적으로 알려야 한다. 최소한 각 데이터 통합에 대한 정보를 위험 평가정보와 함께 웹사이트에 게시한다. 또한, 사람들에게 자기 정보에 대한 접근권을 보장한다. ④ 응답자에게 통합하지 않을 것이라고 명확하게 약속한 경우에는 통합하지 않는다. 통계청은 원 정보가 어떻게 수집되었는지, 그 정보가 어떻게 사용될지 사람들에게 얘기했는지 파악한다. 다른 기관으로부터 정보를 수집했을 경우, 그 정보가 통계청과 공유될 것이며 단지 분석·통계·연구 목적으로만 사용될 것임을 사람들에게 알리도록 해당 기관에 요청한다.

통계청은 새로운 데이터셋을 IDI 등에 통합하거나 IDI 외부에서 데이터셋을 통합하는 모든 제안에 대해 프라이버시 및 기밀성 영향평가 등을 수행한다. 이 평가를 수행하는 '전략, 성과 및 프라이버시 고위 자문관'은 프라이버시 감독기구와 협의할 수 있다.

뉴질랜드 통계청은 마이크로데이터에 대한 접근에 대하여 '5가지 안전조치' 체제를 규정하고 있는데, 이는 데이터 비식별화에 대한 데이터 안전(Safe data)는 물론, 인력 안전(Safe people), 연구 안전(Safe projects), 환경 안전(Safe settings), 결과물 안전(Safe output) 등의 원칙이다.

그림 요약-4 IDI의 5가지 안전조치 체제

5) 시사점

해외 사례 분석을 통해 공익목적의 데이터 연계를 위한 모범 관행(Best Practice)을 도출할 수 있다. 데이터 연계에 관련한 국제 모범 관행은 데이터 연계라는 특정 절차만이 아니라, 데이터의 수집, 저장, 연계, 제공을 아우르는 원칙과 이를 구현하기 위한 조직적, 기술적인 체계, 즉 데이터 거버넌스 체제를 포함한다. 통계 및 공익 연구 목적의 데이터 이용을 활성화하면서도, 정보주체의 개인정보를 보호하기 위해서는 데이터 거버넌스 전 과정에서 데이터의 활용 및 개인정보 보호조치가 고려되어야 하기 때문이다.

첫째, OECD 2015년 보고서에서 지적한 바와 같이, 데이터 거버넌스 체제를 구축하는 데 있어서 데이터의 이용 및 개인정보보호를 규율하는 법제는 가장 중요한 요소이다.

세계 주요 국가들은 개인정보 처리의 일환인 데이터 연계에 있어, 개인정보보호법제 및 보건의료 관련 법제, 통계 관련 법제에서 개인정보보호, 연구 및 통계와 같은 공익목적을 위한 개인정보의 활용, 개인정보 활용 시 안전조치 등에 관한 규정을 포함하고 있다.

연구 및 통계 목적으로 개인정보를 제공할 경우에, 제안서의 요건, 이용 주체의 자격(승인된 연구기관 혹은 연구자 요건 등), 이용의 조건(안전시설 내에서의 이용, 계약의 체결, 연구결과물의 검토 등)을 보다 구체적으로 법률에 규정하기도 한다.

대다수 국가의 법률들이 투명성을 강조하고 있다. 데이터 제공기관의 정책, 승인이 필요할 경우 그 기준 및 승인 목록 등을 공개하도록 하였다. 또 대부분의 국가 통계법에서는 개별 정보에 접근할 수 있는 사람을 제한하였다.

둘째, 연구의 공익과 프라이버시 보호의 균형을 추구할 수 있는 '원칙에 기반을 둔, 비례적인 거버넌스'가 필요하다. 좋은 거버넌스를 위해서 정보 거버넌스 기구, 프로젝

트 승인 기구, 연구윤리위원회 등의 기구가 필요한 역할을 수행할 수 있다.

OECD는 이와 같은 검토와 승인 절차의 원칙으로 '증거기반(evidence-based)' 평가가 이루어져야 하고, 객관적이고 공정해야 하며, 적시에 일관성을 촉진하는 방식으로 이루어져야 한다고 권고하였다. 또한, 정보처리가 개인 및 사회에 미치는 편익과 위험성 및 위험성 경감을 평가할 때 전문가들에 의해 독립적이고 학제적인 검토가 이루어져야 한다.

아이슬란드와 덴마크 등 일부 국가에서는 개인정보 감독기구가 연구신청서 승인 기구의 역할을 하고 있다. 프랑스의 경우, 보건 분야에서 공익적인 연구, 조사 혹은 평가를 목적으로 한 개인정보의 처리, 그 목적이 서로 다른 공익을 위한 파일들의 연계 등에 대하여 개인정보 감독기구인 CNIL의 허가를 받고 있다. 이처럼 개인정보 감독기구가 공익목적의 연구 승인 절차에서 일정한 역할을 수행할 필요가 있다.

셋째, 연구 데이터 허브는 데이터를 보유하거나 접근하는 경로로서 역할한다.

데이터 허브는 데이터에 대한 안전한 접근, 이용, 공개에 대한 통제, 검토, 보유, 파기 등의 제반 절차와 관련하여 원 데이터 보유기관 및 규제자와 계약을 체결한다. 개인정보 감독기구는 이 계약에 대해 의견을 제시할 수 있다.

넷째, 데이터 연계의 방식에 있어 개인정보보호를 위하여 '신뢰할 수 있는 제3자 색인(TTP)' 을 비롯한 '분리 원칙(separation principles)'을 보장해야 한다.

이는 연계 데이터와 콘텐츠 데이터를 분리하고, 데이터 연계 절차와 데이터 보유 및 추출 기능을 분리하자는 것이다. 데이터 연계를 수행하는 사람은 연구자와 분리되어 있어야 하며, 콘텐츠 데이터 연구에 참여할 수 없다.

다섯째, 해외 모든 연계기관에서는 개인정보보호 및 보안 조치의 중요성을 강조하고 있다.

이는 데이터 비식별화 및 TTP 등 데이터 안전(Safe data)의 개념을 비롯하여 연구 인력(Safe people), 연구 프로젝트(Safe project), 연구 환경(Safe environment/settings), 연구 결과물(Safe results/output)을 포괄하는 보다 폭넓은 개념으로 제안되고 있다.

각국은 공익성을 보장하기 위해 데이터 제공을 받을 수 있는 연구 프로젝트의 조건을 제한하거나, 연구자 때로는 그 소속 기관의 자격요건을 규정하고 있다. 데이터에 대한 접근은 엄격한 보안 시설 내에서 이루어진다. 보통 보안 시설 내 전자기기 등 기록 가능한 매체의 소지가 제한되며 보안시설 바깥으로 데이터셋의 외부 반출도 제한하고 있다. 연구결과물 또한 개인정보 침해가 없도록 공개되기 전에 철저하게 검토하는 절차를 두고 있다. 또 해외 대부분 기관에서는 각 기관과 연구자 사이에 계약을

체결하거나, 이용조건에 동의하도록 함으로써, 연구자들이 개인 식별을 시도하거나 연구 목적 외로 이용하는 등 이용규칙을 위반한 경우에 제재할 수 있는 조치를 하고 있었다.

더불어 해외의 많은 기관이 개인정보 및 보안 조치의 적절성을 검토하기 위하여 프라이버시영향평가를 수행하거나, 각국 개인정보 감독기구와 협의하고 있다.

여섯째, 여러 나라가 연구의 설계, 승인, 안전한 환경 내 데이터 접근을 위해 연구자들을 돕는 연구 지원단을 두고 있다. 연구자들의 데이터 접근을 촉진하는 거버넌스 체제가 발전된 국가들은 연구 지원단을 매개로 하여 데이터 접근에 관련된 정보를 투명하게 공개하고 있다.

일곱째, 공익적 연구 목적을 위한 데이터의 연계나 제공에 대한 대중의 신뢰를 얻기 위해서는 높은 수준의 투명성과 참여가 필수적이다. 원칙과 절차, 진행된 사업 내용에 대한 정보를 투명하게 공개해야 하며, 정책 결정 과정에 시민들과 다양한 이해관계자들이 참여할 수 있도록 해야 한다.

OECD 역시 공공적인 협의 과정을 통해 광범위한 이해관계자들이 관여하고 참여할 수 있어야 한다고 권고하였다. 또한, 정보공개를 통해 개인정보 처리의 목적, 개인정보 처리를 승인하는데 사용되는 절차와 기준, 데이터 거버넌스 체제의 실행 및 그 효과와 관련된 정보를 제공할 것을 권고하였다. 승인된 연구 프로젝트와 관련한 정보를 공개하는 것은 개인정보 이용에 대한 대중의 신뢰를 높일 수 있다. 대중들이 개인정보가 누구에 의해서, 어떤 목적으로, 어떻게 사용되는지 알 수 있기 때문이다. 대중들의 반대 여론에 따라 추진이 중단된 영국의 care.data 사례는 대중적 소통이 얼마나 중요한지 반증하는 사례로 자주 거론된다.

여덟째, 공공데이터를 영리적 목적의 데이터와 연계하는 것은 제한되어 있다. 해외 대부분 기관에서는 연구자에게 개인 수준의 비식별 데이터에 대한 접근 및 연계를 허용하더라도, 개인정보 침해의 위험보다 큰 공익적 가치를 가지는, 혹은 기관 임무에 부합하는 연구로 제한하고 있으며, 이를 위해 연구 프로젝트의 과학적 가치, 실현 가능성, 프라이버시 침해 가능성 등을 기준으로 검토하고 승인을 하는 절차를 거치고 있다.

OECD는 대부분 국가에서 핵심적인 보건의료 데이터셋의 비식별화된 마이크로데이터에 대해 영리적 기업의 접근을 승인하지 않고 있다고 분석했다. 영리 기관에게 데이터에 대한 접근을 승인하는 경우에도, 이는 공익적인 학술 연구나 통계 목적의 이용으로 제한된다. 이때 OECD는 '공익(public interest)'의 개념을 데이터 보호, 공공보건, 사회적 보호, 보건의료서비스의 관리, 보건의료 연구 및 통계를 포함하는 것으로

보았다.

국내에서도 통계 및 연구 목적의 행정 데이터 연계 및 접근을 활성화하는 방안을 고려함에 있어, 우선 '공익목적'의 연구로 제한하여 신중하게 접근할 필요가 있다.

(3) 국내 데이터 연계·결합 현황

1) 데이터 연계·결합에 대한 국내 법제

가. 데이터 연계·결합 관련 개인정보보호 규율

개인정보자기결정권은 헌법상의 기본권이다. 우리나라 개인정보보호법제의 기본법이자 일반법인 개인정보보호법은 '개인정보의 수집, 생성, 연계, 연동, 기록, 저장, 보유, 가공, 편집, 검색, 출력, 정정(訂正), 복구, 이용, 제공, 공개, 파기(破棄), 그 밖에 이와 유사한 행위'를 개인정보의 '처리'라고 규정하고 있다(개인정보보호법 제2조 제2호). 이처럼 개인정보의 '연계'는 개인정보보호법에서 개인정보 처리의 하나로 명시하고 있고, 결합은 연동과 유사한 개념으로 볼 수 있다. 한편, 개인정보의 연계나 결합으로 인하여 개인에 대한 새로운 정보가 생성되는 것은 개인정보의 수집에 준하는 것으로 볼 수 있다.

처음부터 개인정보를 연계나 결합하는 경우는 당연히 개인정보보호법의 적용대상이 되는데, 애초에는 개인을 식별할 수 없는 정보였는데, 그 정보의 연계나 결합으로 인해서 개인을 식별하거나, 식별 가능성이 생기는 경우에도 연계나 결합으로 인해서 개인정보가 되어 개인정보보호 원칙 및 개인정보 처리에 관한 법률 규정이 적용된다.

개인정보의 처리에 관한 근거로는 크게 ① 개인정보주체의 동의, ② 법률이나 법령의 규정, ③ 공공기관이 업무를 수행하기 위해 불가피한 경우, ④ 기타 적법 요건으로 나누어 볼 수 있으므로 데이터의 연계·결합도 이런 적법 요건을 갖추어야 한다. 개인정보보호법의 적용대상이 되는 데이터 연계·결합 시 정보주체의 권리도 보장되어야 한다.

개인정보인 데이터의 연계나 결합, 연계나 결합의 결과 개인을 식별하거나 식별 가능성이 있는 정보의 연계나 결합을 개인정보보호법 제15조 제1항의 '수집'으로 해석한다면 적법한 연계나 결합은 개인정보보호법 제17조 제1항의 각호의 요건에 해당해야만 가능할 것이다.

한편, 데이터의 연계나 결합이 개인정보의 제3자 제공에 해당하는 경우가 있다. 개인정보주체로부터 개인정보 연계나 결합을 하는 것이 아닌 경우이다. 예를 들어 병원에서 수집한 정보주체의 개인정보를 제3자에게 제공하여 그 정보주체에 대한 정보와

결합하도록 하거나, 연계하도록 하는 경우가 여기에 해당한다. 이와 같은 제3자 제공이 적법하기 위해서는 개인정보보호법 제18조 상의 요건을 충족해야 한다.

또 원칙적으로 개인정보처리자는 정보주체에게 이용·제공의 목적을 고지하고 동의를 받은 범위나 개인정보보호법 또는 다른 법령에 의하여 이용·제공이 허용된 범위를 벗어나서 개인정보를 이용하거나 제공해서는 안 된다(제18조 제1항). 예외적으로 개인정보처리자가 목적 외 이용을 하거나, 목적 외로 제3자에게 제공할 수 있으려면 개인정보보호법 제18조 제2항의 요건을 충족하는 동시에 정보주체 또는 제3자의 이익을 부당하게 침해할 우려가 없어야 한다.

특히 개인정보보호법 제18조 제2항 제4호에 따르면 통계작성 및 학술 연구 등의 목적을 위하여 필요한 경우로서 특정 개인을 알아볼 수 없는 형태로 개인정보를 제공하는 경우는 목적 외 이용이나, 제3자 제공이 허용된다.

데이터 결합이나 연계 시 건강정보와 같은 민감정보를 처리하는 것은 개인정보주체로부터 동의를 받거나, 법령에 민감정보의 처리를 요구하거나 허용하는 규정이 있어야만 가능하다. 예컨대 건강에 관한 정보일 경우에는 당사자의 명백한 동의가 있어야 하는데, 별도의 동의가 있어야 한다.

한편, 우리나라 개인정보보호법제는 각 영역별로 특별법이 제정되어 적용 범위가 매우 복잡하게 얽혀 있다. 특히 전자정부법, 공공데이터법의 경우 서로 다른 정책 목표를 가진 법률들이 모순적으로 존재하면서 사실상 개인정보자기결정권의 가치를 훼손하는 경우가 문제가 된다. 조화로운 해석을 통해 개인정보보호에 있어 정합성 있는 법체계를 유지할 필요가 있다.

나. 전자정부법

전자정부법의 경우, 행정의 생산성, 투명성, 민주성, 공동이용의 확대, 중복투자의 방지를 목적으로 하는데, 전자정부서비스라는 개념을 매개로 행정정보 시스템 연계의 근거 규정이 되고 있다.

그런데 전자정부법은 정보시스템의 연계와 관련하여 필요성, 연계정보의 범위, 안전조치 등에 대한 법적 규율이 미비하고 개인정보보호에 대한 구체적인 기준이나, 평가절차 등이 없어서, 효율성을 이유로 남용될 가능성을 배제할 수 없다. 전자정부법은 행정정보의 공동이용, 전자적 시스템의 연계, 통합 등의 과정에서 개인정보보호법으로 보장되는 개인정보주체의 권리가 침해되지 않도록 하기 위하여 몇 가지 규정을 두고 있지만, 충분하다고 보기 어렵다. 예를 들어 전자정부법은 이용기관이 공동이용센터를 통하여 개인정보가 포함된 행정정보를 공동이용할 때에는 정보주체의 사전동의를 받

아야 하고, 이를 개인정보보호법의 동의로 갈음한다. 그런데 알려야 하는 사항이 개인정보보호법의 고지사항보다 축소되어 있고, 매우 포괄적인 동의 예외규정도 두고 있다.

따라서 전자정부법을 개선하여 그 규정들이 개인정보보호법에 의한 정보주체의 권리를 훼손하는 것이 아니라는 점을 명확히 하고 행정의 효율화는 개인정보보호원칙 및 개인정보보호법과 조화를 이루도록 규율을 하는 것이 바람직하다.

다. 공공데이터의 제공 및 이용 활성화에 관한 법률

공공데이터의 제공 및 이용 활성화에 관한 법률(공공데이터법)은, 공공기관이 보유, 관리하고 있는 공공데이터의 제공 및 이용 활성화를 위한 법률이다. 공공데이터법에 따른 데이터의 연계는, 별도로 제공받은 데이터를 제공받은 자가 직접 연계하는 것도 가능할 수 있고, 제공하는 공공기관에서 데이터를 연계하여 제공하는 것도 가능할 것이다.

공공데이터법은 공공기관의 공공데이터에 대해서 적용할 기본 원칙을 규정하고 있는데, 그 내용은 정보공개법이나 개인정보보호법 상의 원칙과 충돌하는 측면이 있다.

공공데이터에는 정부간행물도 포함되지만, 개인의 민감한 건강정보나 형사 처분에 대한 정보 등도 포함되어 있는데, 이를 구분하지 않고 공공데이터로 묶은 상태에서 마치 문화의 향유나 보편적 서비스의 대상인 전기통신 역무와 유사하게 취급하여 '이용권의 보편적 확대를 위해 필요한 조처를 할 의무'를 부과하거나, '접근과 이용에 대한 평등의 원칙을 보장해야 한다'는 것은 대상별로 적절한 원칙이라 보기 어렵다.

특히 공공데이터법은 비공개 대상정보도 기술적으로 분리 가능하면 분리해서 나머지 정보를 제공하도록 하였는데, '기술적 분리 가능성'이 제공대상 공공데이터와 제공불가 공공데이터의 구분 기준이 될 수 없다. 예를 들어 공공데이터로 공개되는 환자 데이터셋의 경우, 개인 식별정보를 기술적으로 분리했지만 다른 정보와의 연계를 통해서 개인을 식별할 수 있으므로 개인정보로 볼 수 있다. 해당 정보를 공개하는 것이 국민의 알 권리 보장을 위해 필요하여 해당 개인의 개인정보자기결정권보다 우월하다면 이를 공개하는 것이 허용될 수 있겠지만, 이를 제공대상 공공데이터로 취급하여 영리적 목적을 불문하고 민간 활용을 촉진하도록 하는 것은 부적절하다.

따라서 정보공개법, 개인정보보호법과 공공데이터법이 모순되지 않도록 규정해야 하고, 특히 공공데이터법은 그 추진체계 및 개인정보의 보호에 관한 규정들이 개인정보보호법에 모순되지 않도록 해야 한다.

라. 데이터기반행정 활성화에 관한 법률

2017년 5월 8일 행정안전부는 「데이터기반행정 활성화에 관한 법률」 제정안을 입법 예고하였다. 주요 제정이유는, 행정·공공기관이 보유하고 있는 대규모 데이터 및 민간 보유 데이터, 인터넷의 공개된 데이터를 분석·활용하기 위한 원칙과 절차를 마련하는 데 있다.

제정안에 대하여 개인정보보호위원회는 개인정보 침해요인 평가를 하였고, 2017년 7월 결정(제2017-15-125호)에서 다음과 같이 개선을 권고하였다. 행정 데이터에는 국민의 민감한 개인정보가 포함되어 있으므로 이를 이용, 제공, 연계하는 경우 국민의 프라이버시를 침해할 위험이 크므로 정보주체의 명확한 동의를 기반으로 하는 것이 바람직하며 「개인정보 보호법」의 제 규정을 준수해야 한다. 다만 대량의 행정 데이터를 개별적인 동의를 받아 처리하기 어려운 현실적 제약과 행정 데이터 연계를 통해 얻을 수 있는 공공의 이익을 고려하여 「개인정보 보호법」 제18조 제2항 제4호 등에 따라 개인을 알아 볼 수 없는 형태로 데이터 연계를 추진하는 것이 공익을 위해 필요한 것으로 보인다. 그러나 이 경우에도 프라이버시 침해 위험으로부터 개별 국민을 보호하기 위해 「개인정보 보호법」에 따른 명확한 근거가 필요하고 적정한 관리·감독체계 구축, 데이터 처리에 있어 완전한 기능 분리를 통한 개별 국민의 프라이버시 보호와 인적·물적·기술적·관리적 조치를 통한 데이터의 안전성을 확보하여야 하며 연계대상 행정 데이터의 내역 및 연계절차를 국민들이 알 수 있도록 투명하게 공개할 필요가 있다.

이상과 같은 개인정보보호위원회의 권고 내용은 행정 데이터 연계의 공익과 개인정보 보호법에 따른 정보주체의 권리를 조화시키고, 영국 디지털경제법 등 데이터 연계 절차에 대한 국제 모범 사례를 참고하여 이를 국내 관련 입법안에 적절히 반영토록 한 것으로 평가할 수 있다.

2) 국내 보건의료 분야 데이터 연계 현황

개인 병력이나 질병, 건강상태 등에 관한 정보는 가장 민감한 정보로서, 이 정보는 공개될 경우 개인에 대한 사회적 낙인이 될 수도 있고, 고용, 보험은 물론 사회생활에서 차별의 원인이 될 수도 있고, 개인의 민감한 사생활 침해가 될 수도 있다. 의료법, 약사법 등에서도 진료기록과 처방에 대한 기록을 엄격하게 보호하고 있다.

우리나라는 OECD 국가 중 건강정보 수집, 집적, 연계가 높은 수준이다. 치료 목적, 건강보험의 운영 및 관리 목적, 연구 목적, 보건 정책의 평가와 의료서비스의 질 관리 목적 등 다양한 목적으로 건강정보의 연계나 결합이 이루어지고 있다.

문제는 현재 이루어지고 있는 대다수 보건의료정보의 수집, 활용의 경우 사전동의나 법률 규정을 찾기가 어렵다는 점이다. 개인정보보호원칙이 적용되어 관리되고 있다거나, 충분한 거버넌스 구조를 갖추고 있다고 보기도 어렵다.

부득이한 경우 당사자의 동의가 없어도 개인건강정보를 활용한 공익적 목적의 연구가 필요하다는 점은 인정되지만, 현재 법령에 특별한 규정이 없다. 따라서 공익적 목적의 연구와 관련해서는 반드시 필요한 경우 부득이하게 해당 개인의 동의를 받을 수 없을 때 건강정보를 활용하여 연구할 수 있는 규정과, 연구와 관련한 개인정보의 활용과 관련한 안전조치 등을 법제화할 필요가 있다.

가. 건강정보의 수집·이용 관련 법제

개인정보 중 민감정보에 해당하는 건강정보의 수집, 이용, 제공, 보관, 연계나 결합 등에 대해서는 개인정보보호법이 적용되므로, 개인정보보호법 제23조에 따른다. 즉, 민감정보의 처리는 원칙적으로 금지되고, ① 정보주체에게 고지할 사항을 알리고 다른 개인정보의 처리에 대한 동의와 별도로 동의를 받은 경우이거나, ② 법령에서 민감정보의 처리를 요구하거나 허용하는 경우에만 처리가 허용된다.

그리고 민감정보의 수집, 처리 시에는 당사자의 동의를 받았거나 법률의 규정이 있다고 하더라도 최소수집의 원칙, 익명 처리의 원칙 등 개인정보보호 원칙과, 정보를 제공받을 권리, 접근권, 열람 및 정정, 삭제, 처리중지 요구권, 철회권 등 정보주체의 권리가 보장되어야 하고, 침해 시에는 권리구제를 받을 수 있다.

의료기관, 약국, 정신보건기관 등은 치료, 조제 등의 과정에서 환자로부터 방대한 양의 건강정보를 수집하고, 처리한다. 이 정보들은 국민건강보험공단, 심사평가원, 국립암센터 등을 통하여 수집되기도 하고, 데이터 연계 등 2차적 활용으로 이어진다.

의료기관, 약국, 정신보건기관 등의 경우 법령상 기록 작성과 보관의무가 부과되어 있는데, 그 대상 기록은 정신보건법에 의한 진료기록부, 의료법에 의한 처방전, 조산기록부, 간호기록부와 약사법에 의한 조제기록부 등이다. 이 경우는 정보주체의 동의를 받지 않고도 해당 건강정보를 기록하고, 보존할 수 있다. 그 외의 의무기록은 당사자로부터 명시적인 동의를 받아야 할 것이다.

의료법, 약사법, 정신보건법 등은 의료인, 약사 등에 대해서 엄격한 비밀준수 의무를 부과하고 있다. 아울러 의료인, 의료기관의 경우는 환자에 관한 기록, 조제기록부의 정보를 다른 사람에게 열람하게 하거나, 사본을 내주는 등 내용을 확인할 수 있게 하는 행위를 법률에 명시된 경우 외에는 금지하고 있다. 형법도 업무상비밀누설죄로 직무상 알게 된 비밀의 누설이나 공표를 엄격하게 규율하고 있다. 다만 의료법과 약

사법은 환자 정보를 당사자의 동의가 없어도 제3자에게 제공할 수 있는 예외를 한정적으로 열거하고 있는데, 그중 가장 대표적인 것이 국민건강보험, 의료급여의 비용심사, 지급대상 여부 확인, 사후관리, 적정성 평가, 가감지급을 위하여 건강보험공단, 건강보험심사평가원으로 제공하는 경우이다.

의료인이나 의료기관이 환자의 건강정보를 활용하여 연구하고자 할 경우에는 그에 대한 명확한 동의를 얻어야 한다. 아울러 개인정보가 식별될 수 있는 경우 등 필요한 경우에는 생명윤리법에 의한 심의를 거쳐야 한다. 연구 목적으로 개인정보를 활용하면서 데이터의 연계·결합하려고 하는 경우에도 마찬가지이다.

한편, 연구에 활용하는 경우에도 그 자체의 정보뿐만 아니라 다른 정보와 결합하여도 개인을 식별하는 것이 불가능한 정보를 활용하는 경우에는 해당 정보는 개인정보로 볼 수 없을 것이므로 해당 정보주체로부터 동의를 얻을 필요는 없다.

환자와 의료기관 사이 건강정보의 수집, 이용 및 데이터 연계 등에 있어, 현행 의료법, 약사법, 정신보건법 등의 규정에서 건강정보의 수집 근거를 명확히 하고, 개별 의료기관에서 동의를 구할 때는 정보 수집 시 민감정보 수집을 별도로 고지하고, 수집 항목, 보유 기간, 제3자 제공 등에 대해서 분명하게 명기한 표준 동의서 등을 작성하는 등 제도개선이 필요하다.

건강정보의 2차적 수집 및 활용의 경우로는 국민건강보험공단, 건강보험심사평가원 등이 의료기관으로부터 수집하여 보유하고 있는 건강정보, 보건의료 통계나 조사자료, 질병관리본부의 연구 목적 건강정보, 암 등록정보, 전염병 정보, 건강검진 정보 등을 들 수 있다.

그러나 의료기관으로부터 건강보험공단, 건강보험심사평가원 등이 수집 및 활용, 질병관리본부, 국립암센터 등이 수집하는 정보 및 그 2차적 이용에 대해서는 법적 근거가 매우 미비한 상황이다.

특히 민감정보는 법령에서 구체적으로 민감정보의 종류를 명시하여 처리의 근거를 규정해야 한다.

우리나라는 의료 사회보장제도 운영에 따라 국민건강보험, 의료급여, 노인장기요양보험, 산업재해보험 등에서 건강정보의 수집 및 이용이 이루어지고 있다. 국민건강보험공단은 산업재해보험을 제외한 이들 사회보험에서 의료기록을 제공받고, 이용하고 있다.

나. 의료 사회보장제도와 데이터 연계

가) 국민건강보험공단의 건강정보 수집, 활용과 데이터 연계

국민건강보험공단은 건강보험과 노인장기요양보험의 보험자, 의료급여의 수탁기관, 국민건강증진기관으로서 역할을 수행하는 과정에서 (i) 가입자 및 피부양자의 자격 관리를 위해 필요한 자료, (ii)보험료 부과, 징수를 위해 필요한 자료, (iii) 보험급여의 관리를 위해 필요한 자료, (iv) 보험급여 비용의 지급을 위해 필요한 자료, (v) 건강증진과 예방사업을 위해 필요한 자료를 광범위하게 수집하게 된다.

(i) 국민건강보험공단은 국민건강보험법, 노인장기요양보험법과 의료급여법에 의하여 가입자 및 피부양자, 수급자의 자격 관리를 업무로 관장하도록 수권을 받았기 때문에, 이를 근거로 가입자, 수급자 자격 관리를 위해서 다양한 정보를 수집하고 대부분 영구로 보관한다.

현재 자격정보를 다른 정보와 연계하거나, 결합하는 것과 관련해서는 법령에 구체적 규정은 없다. 다만, 국민건강보험공단이 공개한 개인정보파일에서는 다른 정보시스템과 연계하고 있다고 밝히고 있다. 그러나 구체적으로 어떤 정보시스템과 연계하고 있는지에 대해서는 밝히지 않고 있다.

자격정보 수집과 관련하여 다음과 같은 제도개선이 필요하다. ① 수집하는 정보의 범위를 법령이나 고시 등으로 규율하고, 공개할 필요가 있다. ② 수집하는 정보는 자격 정보로써 필요한 최소한의 범위로 한다. ③ 보유 기간은 목적 달성에 필요한 최소한의 기간으로 하여, 목적 달성 후 즉각 폐기하도록 한다. ④ 보유 기간을 법령이나 고시 등으로 명확하게 규율하는 것이 바람직하다. ⑤ 자격정보를 자격 확인 외의 목적으로 활용하는 것은 목적 외 활용으로 부당하다. 예를 들어 자격정보를 취업 후 학자금 상환 특별법에 따라 국세청에 제공하도록 하는 것은 그 타당성을 수긍하기 어렵다. ⑥ 자격정보에 대한 개인정보 처리방침 등 개인정보의 처리에 관한 사항을 공개하고, 개인정보주체의 열람권, 정정권 등을 보장하여 개인정보의 정확성, 완전성 및 최신성이 보장되도록 하여야 한다. ⑦ 개인정보처리자는 정보주체의 사생활 침해를 최소화하는 방법으로 개인정보를 처리하고, 개인정보의 익명 처리가 가능한 경우에는 익명에 의하여 처리될 수 있도록 하여야 한다.

(ii) 국민건강보험공단은 보험료, 장기요양보험료의 부과 및 징수를 위해서도 방대한 자료를 수집한다. 직장의료보험 가입자의 경우에는 직장과 소득에 대한 상세한 자료, 지역의료보험 가입자인 경우는 재산이나 소득에 대한 상세한 자료를 수집하고 대부분 준영구로 보관한다.

현재 보험료 부과, 징수 정보를 다른 정보와 연계하거나, 결합하는 것과 관련해서도

법령에 구체적 규정은 없다. 다만 국민건강보험공단이 공개한 개인정보파일에서 다른 정보시스템과 연계하고 있다고 밝히고 있지만, 역시 구체적으로 어떤 정보시스템과 연계하고 있는지에 대해서는 밝히지 않고 있다.

보험료 부과, 징수와 관련된 정보 수집과 관련하여 다음과 같은 제도개선이 필요하다. ① 수집하는 정보의 범위를 법령이나 고시 등으로 규율하고, 공개할 필요가 있다. ② 수집하는 정보는 자격 정보로써 필요한 최소한의 범위로 한다. ③ 보유 기간은 목적 달성에 필요한 최소한의 기간으로 하여, 목적 달성 후 즉각 폐기하도록 한다. ④ 보유 기간을 법령이나 고시 등으로 명확하게 규율하는 것이 바람직하다. ⑤ 보험료 부과, 징수에 관한 정보를 보험료 부과. 징수 목적 외의 목적으로 활용하는 것은 부당하다.

(iii) 보험급여(의료급여) 관리 및 (iv) 비용지급과 관련하여, 우리나라 건강보험이나 의료급여는 모든 진료 행위에 대한 상세한 내역이 제공되어야만 보수를 산정할 수 있으므로, 요양기관은 그에 대한 상세한 정보를 국민건강보험공단과 심사평가원에 모두 제공해야만 한다. 요양급여비용 청구명세서 정보는 5년 보존되지만 5년 후에도 삭제하지 않고 국민건강DW정보에 보유되며, 요양급여내역 등의 정보는 5년 또는 10년씩 보유되고, 상해 등 고의 또는 중대한 과실의 경우에는 사고발생경위, 사법기관에서 제공받은 처분결과, 가해자 성명 등에 대한 정보를 수집하여 준영구 보유하고 있다.

현재 요양급여 청구정보 등과 다른 정보를 연계하거나, 결합하는 것과 관련해서는 법령에 구체적 규정이 없는데, 국민건강보험공단은 다른 정보시스템과 연계하고 있다고 밝히고 있다. 구체적으로 어떤 정보시스템과 연계하고 있는지에 대해서는 밝히지 않고 있다.

보험급여(의료급여) 관리 및 비용지급과 관련하여 다음과 같은 제도개선이 필요하다. ① 수집하는 정보의 범위를 법령이나 고시 등으로 규율하고, 공개할 필요가 있다. ② 수집하는 정보는 자격 정보로써 필요한 최소한의 범위로 한다. ③ 보유 기간은 목적 달성에 필요한 최소한의 기간으로 하여, 목적 달성 후 즉각 폐기하도록 한다. ④ 보유 기간을 법령이나 고시 등으로 명확하게 규율하는 것이 바람직하다. ⑤ 보험료 부과, 징수에 관한 정보를 보험료 부과. 징수 목적 외의 목적으로 활용하는 것은 부당하다.

(v) 국민건강보험공단은 건강증진과 예방 등의 업무로 전 국민에 정기적으로 건강검진을 하고 이 과정에서 사실상 전 국민의 정기적인 건강검진 자료가 국민건강보험공단에 수집되어 영구 보관된다. 다만 만성질환자와 전단계자에 대해서는 사후관리대상자로 별도로 관리하고 있다(10년).

현재 건강검진 정보 등을 다른 정보와 연계·결합하는 것에 관해서는 법령에 구체적 규정은 없다. 다만, 국민건강보험공단이 공개한 개인정보파일에서는 다른 정보시스템과 연계하고 있다고 밝히고 있고, 건강검진대상자 및 검진결과 내역의 경우, 영구적으로 보유하면서 보건복지부, 국립암센터, 사회보장기관에 정보를 제공하여 연계하고 있다. 그 외에 어떤 정보와 연계하고 있는지에 대해서는 밝히지 않고 있다.

건강검진 자료수집과 관련하여 다음과 같은 제도개선이 필요하다. ① 수집하는 정보의 범위를 법령이나 고시 등으로 규율하고, 공개할 필요가 있다. ② 수집하는 정보는 필요한 최소한의 범위로 한다. ③ 보유 기간은 목적 달성에 필요한 최소한의 기간으로 하여, 법령이나 고시로 명시하고, 목적 달성 후 즉각 폐기하도록 한다. ④ 데이터의 연계, 제3자 제공 등을 엄격하게 제한해야 한다.

나) 건강보험심사평가원의 건강정보 수집, 활용

건강보험심사평가원은 의료공급자가 진료비를 청구하면 국민건강보험법 등에서 정한 기준에 의해 진료비와 진료 내역이 올바르게 청구되었는지, 의·약학적으로 타당하고 비용 효과적으로 이루어졌는지 확인한다. 이 과정에서 건강보험심사평가원은 방대한 양의 건강정보를 수집, 처리한다. 심사평가원이 수집하는 정보는 범주를 나누어 보면, ① 진료정보, ② 의약품 정보, ③ 치료재료 정보, ④ 의료자원 정보, ⑤ 비급여 정보, ⑥ 평가정보로 나누어 볼 수 있다.

진료비 심사 청구에 관한 정보는 당해 목적의 범위에서 보유하는 것이 바람직한데, 준영구로 보유하는 것은 문제이다. 따라서 청구명세서의 항목도 필수 불가결한 것으로 검토하여 제도를 개선하는 것이 바람직하다.

다) 건강보험의 건전성 유지, 연구 목적 2차적 이용

국민건강보험공단은 보험급여의 관리, 건강보험에 관한 조사연구, 건강관리 사업을 업무로 담당하도록 법률에 규정되어 있다. 현재 국민건강보험공단은 이를 근거로 연구 목적이나 2차적 이용을 목적으로 다양한 정보를 연계·결합하여 활용하고 있다.

(i) 국민건강정보 데이터베이스

국민건강보험공단은 현재 40,833,684,272명의 개인에 대한 건강정보를 삭제하지 않고 보유하고 있다고 한다. 국민건강보험공단은 이 정보 파일에 대해서 '전 국민의 자격 및 보험료, 건강검진결과, 진료내역, 노인장기요양보험 자료, 요양기관 현황, 암 및

희귀난치성질환자 등록정보 등 1조 3천억 건에 달하는 방대한 빅데이터를 말한다'고 설명하고 있다.

국민건강보험공단이 보유한 △건강보험 적용인구, 세대정보 △사업장 및 지역보험료 부과 정보, 징수, 체납정보 △일반건강검진, 암 검진, 생애 전환기 검진 △노인장기요양 인정신청 및 급여 △요양급여비용 지급 및 중증암환자 등록, 현금급여 지급 △의료급여 적용, 세대, 급여비용 등의 정보는 주민등록번호나 건강보험증번호와 같이 고유식별정보를 모두 포함하고 있어서 다른 정보와의 연계도 가능하다. 아울러 이 정보는 국민건강보험공단이 자격정보, 징수 정보, 심사 청구 명세서 정보, 요양급여내역, 건강검진정보, 희귀질병 정도 등 국민건강보험공단이 각각 다른 목적으로 수집하여 보유하고 있는 정보를 연계한 것으로 보인다.

그런데 다양한 목적으로 국민건강정보 데이터베이스로 수집된 개인의 건강정보와 자격정보, 징수 정보 등의 개인정보가 연계·결합되고, 이용, 제공되는 것에 대하여 해당 개인에게 동의를 받은 사실이 없다. 따라서 법적 근거가 있어야 하는데 국민건강보험법 제14조 제1항 제9호에서 국민건강보험공단의 업무로 '건강보험에 관한 조사연구 및 국제협력'이 규정되어 있지만, 이 규정을 개인정보보호법 제23조 제1항의 '민감정보의 처리를 허용하는 법령의 규정'으로 보기에는 난점이 있다.

2차적 이용과 관련하여 다음과 같은 제도개선이 바람직하다. ① 연구 목적의 건강정보 활용을 위해서는 개인정보주체의 동의가 없는 한 현재의 국민건강보험법에 의한 개인정보 활용은 적법한 근거라고 보기 어렵다. ② 모든 개인의 건강정보를 포괄적으로 연계하여 준영구로 보유하는 것은 개인정보보호원칙을 위반한다. ③ 연구 목적의 건강정보 활용을 하려면 연구 목적에 필요한 최소한의 정보를 보유하고, 폐기해야 한다. ④ 연구 목적의 건강정보 활용을 위해 최소한의 정보를 활용하더라도 익명화 등 안전조치를 취하는 것이 바람직하다.

(ii) 5종의 코호트 DB

국민건강보험공단은 5종의 코호트 정보를 연구 목적으로 생성하여 보유, 활용하고 있는데, 코호트 정보는 개인에 대한 추적정보이기 때문에 그 민감성이 더욱 크다.

규모가 가장 큰 전 국민 표준코호트DB의 경우 전 국민의 2%에 달하는 100만 명을 추출하여, 무려 14년간의 사회, 경제적 현황과 장애, 사망, 의료이용 현황, 모든 진료내역은 물론, 건강검진 등의 상세한 정보(자격 및 보험료, 출생 및 사망, 모든 진료내역, 건강검진 내역)를 데이터베이스로 구축한 것이다. 또 건강검진 코호트 DB는 수검자 중 10%, 노인 코호트는 대상자 중 10%, 직장여성 코호트 DB는 5%, 영유아검진

코호트 DB 5%를 추출하는 등 그 대상도 방대하고, 코호트 정보를 추적하여 축적한 기간이 14년 ~ 8년간으로 장기간인 데다가, 수록 정보도 사회경제적인 정보와 의료와 건강검진 정보까지 포괄하고 있어서 개인의 내밀한 사생활까지 고스란히 담겨져 있다.

5종의 코호트 DB의 보유 목적은 연구 목적이라고 밝히고 있다. 일반적으로 코호트 연구는 당사자의 동의를 받아서 진행한다. 그러나 국민건강보험공단은 어디에서도 당사자들에게 동의를 받았다는 정보나 받은 동의의 내용을 밝히고 있지 않다. 그뿐만 아니라 해당 정보주체에게 코호트DB로 해당 정보주체의 여러 가지 정보가 연계되어 추적, 축적되어 있고, 그 정보들이 제3자에게도 제공되고 있다는 점에 대해서 알리고 있지 않은 것으로 보인다. '건강보험에 관한 조사연구'(국민건강보험법 제14조 제1항 제9호)를 법적 근거로 보기도 어렵다. 국민건강보험공단은 코호트 자료를 만들면서 소위 '비식별 조치'를 했다고 하지만 이는 가명화와 그룹핑, 민감 상병의 마스킹 수준으로, 개인정보보호법의 적용이 배제되는 '개인을 알아볼 수 없게 조치한 것'으로 보기는 어렵다.

(iii) 건강보험심사평가원의 DW시스템, 빅데이터 분석 DB 등

건강보험심사평가원은 심사기준 및 평가 기준의 개발, 그에 따른 업무와 관련된 조사연구 및 국제협력, 요양급여비용의 심사 청구와 관련된 소프트웨어의 개발·공급·검사 등 전산 관리, 환자 분류체계의 개발·관리 등의 업무를 담당하는데, 이를 위하여 건강정보를 활용할 수 있는지가 문제 된다.

건강보험심사평가원은 DW시스템(명세서), 빅데이터 분석 DB 등의 정보를 준영구로 보유하고 있다고 한다. 이 정보들은 데이터 연계로 축적되고 있는 것으로 보인다. DW 시스템이라는 명목으로 139억명의 민감한 건강정보를 준영구로 보유하고, 5,300만 명의 빅데이터 분석 DB를 보유하는 것은 법적 근거를 찾기 어렵고, 개인정보보호 원칙에 어긋나는 것으로 문제이다.

(iv) 건강보험심사평가원의 환자데이터셋

건강보험심사평가원은 2009년부터 매년 1년간 진료받은 환자를 대상으로 표본 추출을 하여 환자데이터셋이라는 이름으로 공개를 하고 있다. 입원환자 데이터셋은 입원환자의 13%에 달하는 100만 명 정도, 전체 환자 데이터셋은 전체 환자의 3%인 140만 명, 고령환자 데이터셋은 전체 고령환자(60세 이상)의 20%인 100만 명 정도, 소아청소년환자데이터셋은 소아청소년환자(20세 미만)의 10%인 약 100만 명 정도의 환자

데이터셋을 매년 추출하여 작성하는 것이다.

환자데이터셋에 대해서 주민등록번호를 대체키로 변환하고, 가명처리, 데이터삭제, 마스킹 등의 조치를 했다고 한다. 그런데 이와 같은 비식별 조치로는 환자 정보의 일부를 가지고 있거나, 언론, 소셜미디어 등의 정보와 결합하는 경우 재식별이 가능할 수 있다.

공공데이터로 개방하여 공개하기 위해 구축된 환자데이터셋은 민감한 개인정보이므로 이를 연계하려면 해당 환자로부터 그와 같은 개인정보의 연계에 대해서 명시적인 동의를 받거나, 법적 근거가 필요하다. 건강보험심사평가원은 '공공데이터의 제공 및 이용 활성화에 관한 법률'에 근거하여 비식별 조치를 했으므로 제공할 수 있다고 보았다. 그러나 공공데이터법이 제공의 근거가 될 수 없으며, 비식별 조치라는 것은 개인을 식별할 수 있는 정보로서 개인정보보보협 제23조의 규율 대상이 된다. 따라서 그에 대한 명확한 법적 규율이 없는 한 민감한 개인정보인 건강정보의 제공은 허용되지 않는다.

라) 보건의료 빅데이터 개방시스템을 통한 데이터의 연계

보건의료 빅데이터 개방시스템으로 국민건강보험공단은 국민건강보험자료 공유서비스를, 건강보험심사평가원은 보건의료빅데이터 개방시스템을 운영하고 있다.

(i) 국민건강보험공단의 국민건강보험자료 공유서비스

국민건강보험공단은 국민건강보험자료 공유서비스를 통해 자신이 수집·보유·관리하는 건강보험 및 장기요양보험 관련 자료를 학술연구 및 정책연구 목적으로 제공하고 있다. 표본연구 DB 외에도 맞춤형 자료의 경우 외부 기관 자료와 연계 서비스도 제공하고 있다. 그러나 현행 법령상 건보공단이 제공받은 자료를 계속 보유하면서 개인정보를 연구에 제공할 수 있는 법적 근거는 희박해 보인다.

(ii) 건강보험심사평가원의 보건의료빅데이터 개방시스템 운영

건강보험심사평가원은 심사평가 등의 업무와 관련하여 수집, 보유하고 있는 국민의 건강정보를 '공공데이터의 제공 및 이용 활성화에 관한 법률'의 '공공데이터'로 보고, 보건의료빅데이터개방시스템을 통하여 개방하고 있다.

자료제공대상은 의료기관 종사자, 학술 연구 수행기관, 제약업체 및 의료기기 업체 등 산업계는 물론이고, 컨설팅업체 및 예비 창업자 등도 포함하고 있다. 영리 여부를

불문하고, 누구에게나 개방하며, 실제로 민간보험회사에도 국민의 진료내역 등을 제공하였음이 밝혀지기도 했다.

진료정보, 의약품 정보, 치료재료정보, 의료자원정보 외에도 맞춤형 자료의 경우, 요청되는 정보들이 특정한 상병이나 특정한 약품 수진자 전부에 대한 정보 등 충분히 개인이 식별될 수 있는 정보들이다.

현재 보건의료빅데이터 개방시스템에서는 과제목록을 공개하고 있는데, 전체 625건의 과제 중 공개하고 있는 것은 단 2개에 불과하고, 나머지는 모두 비공개로 신청하고 있다.

연구자는 외부기관의 자료를 연계하기도 하는데, 외부기관의 자료원을 개인의 고유식별정보를 이용하여 심사평가원의 청구자료와 연계하여 분석을 수행하고자 할 경우, 외부기관이 민감정보, 고유식별정보, 주민등록번호를 심사평가원에 제공하기 위해서는 제3자 제공에 대한 법적 근거가 있어야 한다. 개인정보보호법과 생명윤리 및 안전에 관한 법률에 따른 서면 동의를 받고, 기관위원회의 심의를 받아야만 민감정보, 고유식별정보, 주민등록번호 처리가 가능하고, 제3자에게 제공이 가능하다.

다. 기타 건강정보와 데이터 연계

법률적 의무를 부과하는 암 정보 등록제도에 대해서는 그 필요성이 인정된다. 그럼에도 국립암센터와 지역의 대학병원들이 암 등록사업을 수행함에 있어서 환자들의 암 관련 정보들이 개인정보보호법 적용대상에서 제외되고, 통계법에는 통계자료의 비밀보호 규정이 미비하고, 암관리법도 비밀보호규정이 미비하기 때문에 암 등록사업의 비밀보호와 개인정보보호가 충분하다고 보기 어렵다.

암관리법은 보건복지부장관 요구하는 '암등록통계사업에 필요한 자료'의 제출(암 환자의 진료와 관련된 자료 및 의무기록 등)이나 의견의 진술에 있어 그 수집대상 정보가 무엇이고 요구 상대방이 누구인지를 명확하게 규율하고 있지 않다. 또 보건복지부장관이 자료나 의견 진술을 요구할 경우 자료제출을 요구받은 자는 특별한 사유가 없으면 요구에 따라야 한다는 규정을 두고 있는데, 환자 본인의 거절권이 보장되는 것이 바람직하다. 특히 암 정보는 매우 민감한 정보이고 유출되는 경우 해당자의 고용, 보험 등 사회생활에서 차별과 불이익이 우려되므로 암등록통계사업은 명확하게 규정된 법령에 따라서 이루어져야 한다. 또 암등록통계사업을 위해 수집한 환자들의 암 관련 정보들은 암등록통계사업의 목적으로만 활용될 수 있도록 규정할 필요가 있다. 수집된 정보는 개인정보보호법이 적용 배제되고 개인정보보호 관련 규정이 매우 미비한 통계법만 적용되는데 특히 비밀보호에 대한 개인정보보호 규정의 보완이 필요하

다.

한편 중앙암등록본부의 자료는 통계법의 적용대상이 되므로, 비밀보호의 대상이 되고, 통계나 연구 목적 외의 목적으로 활용되어서도 안 된다. 다만 현재 중앙암등록본부에 등록된 원시자료 및 주문형 자료는 개인정보보호법 및 의료법 제20조에 준하여 '등록환자의 개인정보가 보호될 수 있는 범위 내에서 제공하고 있다'고 하는데, 어느 정도의 비식별 조치를 했어야 '개인정보가 보호될 수 있는 범위'가 될 수 있는 것인지는 논란이 된다. 의료정보의 경우는 비식별화를 하더라도 일부의 주소, 초진 연월일, 입원일, 퇴원일, 치료시행 여부, 진단경로, 원발부위, 전이부위, 조직학적 진단명, 진단방법, 병기 분류, 분화도, 사망연월일, 사망원인 등의 자료만으로도 특정인을 식별해 낼 수 있다.

국립암센터는 암관리법의 암등록통계사업, 암 검진사업, 암 환자의 의료비 지원사업 등, 완화의료 사업 등 4개 사업별 DB를 연계, 통합하여 통합 데이터베이스로 구축하고, 이를 활용해 시범서비스를 개발하고 있다. 그러나 이와 같은 4개 사업 DB의 연계, 구축에 대한 적법한 근거가 부족하다. 또 국민건강보험공단과 국립암센터는 공동연구 협약을 체결하고, 국립암센터 암 등록자료, 건강보험공단 암 환자 자료, 통계청 사망자료 등 연계를 통하여 암 진단 및 치료 효과, 국가 암 관리사업 효과 평가 등을 수행하고자 관련 자료의 수집 및 연계와 분석 적합성 점검을 진행하였다. 그런데 이와 같이 대규모로 민감한 건강정보를 연계하는 것과 관련하여 법적 근거가 미비하다. 비록 암 관련 연구 과제의 해결이 중요하기는 하지만, 10여 년간의 민감한 개인정보가 환자에게 아무런 선택권도 없이 통합되어 만들어졌다는 것은 문제의 소지가 있다.

암 등록사업의 필요성을 인정하더라도 암 등록자료는 개인의 민감정보에 해당한다. 따라서 해당 정보주체의 동의 없이 민감한 개인정보를 연구 목적으로 제공하는 것은 적법한 법적 근거를 갖출 필요가 있다.

라. 인간대상연구 등 개인정보 활용 연구

생명윤리법에서 규율하는 인간대상연구와 인체 유래물 연구도 연구의 대상이 되는 정보가 개인정보에 해당할 가능성이 크다. 생명윤리법은 연구대상자의 보호를 위한 조치로 피험자의 권리와 안전에 대한 고려, 불완전 능력자와 취약계층에 대한 보호, 개인정보보호 및 기록의 유지와 정보공개 원칙 등을 규정하고 있다. 또 인간대상연구를 하려는 자나 인체 유래물 연구를 하려는 자는 연구를 하기 전에 연구계획서를 작성하여 기관위원회의 심의를 받아야 한다.

인간대상연구, 인체 유래물 연구를 하기 전 연구대상자, 기증자로부터 서면 동의를

받는 것이 원칙이다. 다만 ① 연구대상자의 동의를 받는 것이 연구 진행과정에서 현실적으로 불가능하거나 연구의 타당성에 심각한 영향을 미친다고 판단되는 경우로서, ② 연구대상자의 동의 거부를 추정할 만한 사유가 없고, 동의를 면제하여도 연구대상자에게 미치는 위험이 극히 낮은 경우 서면 동의를 면제할 수 있다.

제3자 제공의 경우 기증자나 연구대상자로부터 인체 유래물 및 개인정보 제공에 대하여 서면 동의를 받은 경우 기관위원회의 심의를 거쳐 인체 유래물 은행이나 다른 연구자 등에 제공이 가능하다. 이때 기증자 및 연구대상자가 개인식별정보를 포함하는 것에 동의하지 않았다면 익명화를 해서 제공해야 한다. 연구자는 이에 대한 기록을 작성, 보관해야 한다.

3) 국내 통계 분야 데이터의 연계 현황

통계 분야에서도 데이터 연계·결합의 요구가 늘어나고 있다. 통계청은 빅데이터 활용을 주요 과제로 선정하고 있다. 통계와 관련한 법제로도 개인정보보호법과 통계법이 있고, 각 부처의 훈령으로 통계관리규정을 두고 있지만, 다양한 데이터 연계와 통합, 특히 민간분야와의 연계 등을 시도하는 상황에서 그에 대한 규율이 매우 미흡하다.

가. 통계 제도와 기본 원칙

우리나라 통계법은 제2조와 제33조에서 통계의 비밀보호에 대해 규율하고 있다. 그런데 이 규정은 '비밀'에 해당하는 자료에 대해서만 보호가 이루어지는 것으로 되어있어서, 이것만으로는 통계단위의 신뢰를 보호할 수 없다. 반면 해외의 입법들은 통계의 작성과정에서 수집한 개인에 관한 정보는 그것이 비밀에 해당하는지 여부를 불문하고 이를 비밀로 보호하고 있다.

또 통계법은 통계자료의 제3자 제공과 관련하여, 통계작성기관인 제3자가 통계작성기관에 통계작성을 위해 통계자료를 요청하여, 통계자료를 제공하는 경우는 '통계주체를 식별할 수 없도록 해야 한다'는 내용의 규정을 두고 있다. '특정의 개인이나 법인 또는 단체 등을 식별할 수 없는 형태로 통계자료를 처리한 후 제공하여야 한다'는 것이다. 그러나 '특정의 개인이나 법인 또는 단체 등을 식별할 수 없는 형태로 통계자료를 처리하는 것'은 식별 가능성을 제거하는 것이 아니므로 비밀보호로 충분하지 못하다.

통계의 비밀보호를 위해서는 통계법이 다음과 같이 개선될 필요가 있다. 통계의 비밀을 보호한다는 점을 명백하게 밝혀야 한다. 통계의 공표 시 통계단위가 식별되지

않도록 해야 한다는 점 또한 밝혀야 한다. 그 예외로는 유럽연합의 경우처럼 법령의 규정에 의한 경우로 한정하는 것이 좋을 것이다. 통계는 반드시 통계 목적으로만 활용되어야 함을 명백하게 밝혀야 한다. 통계자료에 관한 연구 목적의 활용에 대해서도 분명한 규율을 제정하는 것이 바람직하다.

나. 통계와 개인정보보호

개인정보보호법 제58조는 “공공기관이 처리하는 개인정보 중 통계법에 따라 수집되는 개인정보에 대해서 개인정보보호법 제3장 ~ 제6장을 배제한다.”는 규정을 두고 있다. 이와 관련하여 개인정보보호법의 적용이 배제되는 경우가 어떤 경우인지 해석에 논란의 여지가 있다.

물론 통계법에 따라 통계작성을 위한 개인정보 수집 시 개인정보보호법 제58조 제1항에 해당하여 개인정보보호법 제3장 ~ 제7장의 적용이 배제되더라도, 제58조 제4항에 의하여 목적에 따른 필요 최소한의 범위의 개인정보를 수집해야 하고, 목적 범위를 넘는 개인정보 수집을 해서는 안 되고, 목적 달성 후에는 즉시 파기해야 한다. 개인정보의 안전한 관리를 위하여 필요한 기술적·관리적 및 물리적 보호조치도 갖추어야 한다. 또한, 개인정보보호법 제3조 개인정보보호 원칙 및 제4조 정보주체의 권리보장도 준수되어야 한다.

해외의 법제는 통계작성을 위하여 개인정보를 처리하는 과정에서 개인정보보호를 개인정보보호법제와 통계 법제에서 함께 규율하는 경우가 대부분이고, 통계작성을 위한 개인정보의 수집이나, 통계자료에 포함된 개인정보의 보유, 이용, 제3자 제공 등과 관련해서도 필요한 규정에 대해서만 특칙을 두고 있는 경우가 대부분이다. 우리의 제도도 통계의 작성 및 활용에 있어 공공기관의 법령상 의무를 수행하는 과정에서 개인정보를 수집, 처리, 활용하는 경우와 마찬가지로 규율하도록 개선될 필요가 있다.

다. 통계 작성과정의 데이터 연계

통계의 작성과정은 통계 조사의 기획과 설계, 자료수집, 자료처리, 자료의 정리 및 공급의 절차로 구분할 수 있다.

우선 통계는 민감한 개인정보를 수집하여 작성하게 된다. 승인 통계의 경우는 통계법에서 개인정보의 수집을 강제하는 규정을 두고 있고, 이용, 제3자 제공을 허용하는 포괄적인 규정을 두고 있는 반면, 그에 대한 보호조치들은 규정이 매우 미비하다. 따라서 통계로 승인하는 절차는 매우 엄격하거나, 영향평가의 대상이 되거나, 법령의 체계를 갖춘 것이어야 바람직할 것이다.

현재 통계청은 통계작성 승인제도를 통계의 신뢰성 확보, 중복 방지의 관점에만 주안점을 두고 운영하고 있다. 통계작성을 승인할 때, 개인정보보호 원칙을 준수하고 있는지가 검토의 기준 중 하나가 되어야 한다. 현재 통계승인이나 지정통계의 지정을 결정하는 통계조정심의위원회는 통계청 내부의 조직으로 독립성과 개인정보보호에 관한 전문성을 갖추고 있다고 보기 어렵다. 성 평등을 위한 성별통계와 관련하여 여성가족부장관에게 의견조회를 의무화한 것처럼 개인정보보호위원회의 의견조회를 의무화하거나, 의견제시를 할 수 있는 제도적인 방안을 마련하는 것이 좋다.

현재 통계법에서는 데이터 연계에 대해서 특별한 규정은 두고 있지 않고, 연계의 요건에 대해서는 특별한 규정은 없다. 다만, 등록센서스 자료 수집 시 행정자료의 정보보호를 위한 운영규정·개인정보보호지침 등을 제정 운영하면서 행정자료 접근을 엄격히 제한하고, 주민등록번호·외국인등록번호 등 개인식별번호는 복원 불가능한 가상번호로 대체 후 즉시 삭제하며, 인터넷과 분리된 업무전용망에서 행정자료 통합관리 시스템을 통해 관리하고 있다고 한다.

임금근로일자리 행정통계의 사례에서처럼 행정통계를 작성하는 과정에서 여러 행정자료를 연계하여 활용하는 경우가 많다. 패널 조사 등 조사통계의 작성과정에서도 조사자료와 행정자료의 연계가 이루어진다. 예를 들어 의료패널도 국민건강보험공단의 건강보험자료와 연계를 한다. 이 외에도 통계청은 조사자료와 행정자료의 연계를 적극 추진하고 있다. 연계 대상인 행정자료로는 국세청 근로소득·사업소득·금융소득 자료 및 보건복지부 기초연금·생계급여·주거급여·장애인연금 자료 등 약 20여 종을 들었다. 인구총조사, 경제총조사 등 전수조사통계에도 데이터 연계는 광범위하게 활용된다. 인구총조사는 행자부 등 13개 정부기관 24종 행정자료를 융합한 등록센서스로 전환하여, 행정자료를 활용한 후 20% 표본조사를 하는 방식으로 변경되었다. 경제총조사도 국세청 사업자등록자료, 과세자료 등 8개 기관 20종 행정자료를 활용하여, 행정자료를 연계하고 있다.

통계의 원천이 되는 자료를 통계자료 혹은 마이크로데이터라고 부른다.

통계자료의 보유 및 관리에 대한 규율과 관련하여서, 현재 통계법은 ‘통계작성기관의 장은 통계의 보급 및 이용의 활성화를 위하여 통계자료를 보유·관리하여야 한다.’는 규정만을 두고 있고(통계법 제29조의 2), 시행령은 ‘전산 데이터베이스 등의 매체에 유실되지 않도록 보유·관리하여야 한다.’는 규정만을 두고 있다(시행령 제45조의 2 제1항).

통계자료의 비밀보호를 위해서 통계자료의 보유와 관리에 대한 규율을 마련할 필요가 있다. 특히 현재 개인정보보호법이 통계와 관련한 정보에 대한 개인정보보호법의

적용을 배제하고 있기 때문에 더더욱 통계법에 그에 관한 규정을 두어야 한다. 통계자료의 목적 달성 후 폐기 의무, 통계자료를 보유하는 동안에도 통계자료의 개인식별자를 대체번호 등으로 치환하고, 연계정보는 별도로 보관하면서 대체식별자로 처리된 통계자료를 활용하는 등의 안전조치가 바람직하다. 통계자료를 보유하는 동안 물적, 조직적, 관리적 안전조치를 유지하도록 하고, 특히 통계의 기밀유지를 위한 여러 규정을 마련할 필요가 있다.

라. 통계자료의 활용

통계법상 통계자료의 활용에 관한 규정은 통계작성의 목적을 위해서 활용하는 경우와 기타의 경우로 나누어 볼 수 있다. 통계의 작성을 위해서 필요한 경우는 통계법 제30조에서 규율하고 있다.

통계법은 통계작성 목적 외의 통계자료 활용에 대해서 제31조에 규정하면서, 특정의 개인이나 법인 또는 단체 등을 식별할 수 없는 형태로 통계자료를 처리한 후에는 제공하도록 규정하였다. 통계법은 통계자료의 이용신청을 할 수 있는 자를 제한하지 않고 있는데, 다만 해당 통계자료를 다른 자료와 대응 또는 연계함으로써 특정의 개인이나 법인 또는 단체 등의 식별이 가능하게 되는 경우, 사업체의 영업상 비밀을 침해하게 되는 경우에는 통계작성기관의 장이 통계자료를 제공하지 않을 수 있을 뿐이다.

따라서 통계자료가 특정 개인을 식별할 수 있는 형태로 제3자에게 제공되는 것은 엄격하게 금지해야 하고, 다만 예외적으로 이를 인정하려면 공공 연구의 목적으로 통계자료를 활용하는 경우로 제한하는 것이 바람직하다. 이 경우에도 해당 공공 연구는 국가기관이나 그에 준하는 신뢰성을 갖는 연구기관의 연구이어야 하고, 연구계획서에 대한 기관윤리위원회 등의 심의를 거치도록 하는 것이 바람직하다. 그리고 해당 통계자료를 제공받을 수 있는 시설을 제한하고, 통계자료의 유출이나 남용을 방지하기 위한 안전조치를 마련하는 것이 필요하다.

통계청 훈령인 국가통계자료제공 규정에 따르면 각 통계작성기관에 통계자료의 제공과 관련된 사항을 심의하기 위하여 통계자료제공심의회를 두도록 했다. 그런데 심의회는 통계자료의 제공에 대한 심의를 하는데, 심의회의 구성과 운영의 독립성은 거의 보장되지 않고 있다. 통계자료의 제공에 대한 심의는 독립성과 전문성이 보장될 수 있도록 구성하고, 운영하는 것이 바람직할 것이다. 통계자료의 비밀보호와 관련해서는 '실질적으로 식별되지 않도록'이라는 기준은 매우 불충분하므로, 자료제공 요청의 목적이나 제공 요청대상 정보의 내용에 따라서 세분화할 필요가 있다.

통계청은 통계청에서 작성하는 마이크로데이터 뿐만 아니라 정부 각 부처, 지자체, 연구기관 등 타 통계작성기관의 마이크로데이터를 한 곳에 모아 마이크로데이터 통합서비스라는 이름(MDIS)으로 제공하고 있다. 마이크로데이터의 경우 제공되는 형태에 따라서는 데이터연계나 데이터 결합이 가능한 경우도 있다.

마. 통계청의 데이터 연계·융합 활성화 추진전략

통계청은 데이터 연계, 융합을 활성화하려는 방안을 추진하고 있는 상황이다. 특히 통계청은 인구/가구, 사업체, 주택/건물, 경제활동 등 4개 분야의 통계작성 명부를 구축하고 있으며 등록부 간 연계로 범용 종합등록부의 구축을 추진할 계획이다.

통계청은 최근 공공데이터와 민간데이터 간 연계로 새로운 통계 정보 작성을 추진하고 있다. 그 시범적 사업으로 통계청 공공데이터와 민간기관 신용정보회사(코리아크레딧뷰로)의 빅데이터를 연계·분석하여 통계를 작성하였다.

통계청과 KCB의 데이터 연계는 신혼부부 데이터와 신용정보를 연계한 것인데, 통계청은 이를 저출산 대책 등 지원이 필요한 신혼부부에 관한 통계를 구축·분석한 것이라고 설명한다. 통계생산의 목적은 타당하다고 보이지만, 이 통계를 생산하는 과정에서 개인정보 침해를 최소화하려는 조치 등이 마련되어 있어야 하고, 절차가 통제되어야 한다.

통계청의 데이터로는 인구주택총조사의 등록센서스 정보(개인과 가구와 관련된 다양한 정보), 인구동향정보의 혼인 관련 정보, 국적 정보, 경제활동 정보로 임금근로, 일자리통계, 4대보험 정보, 근로소득 정보, 그 밖에 통계청의 각종 조사정보도 연계되었다. 이처럼 통계청이 신혼부부를 추출하여 표본을 구성하고, 이 데이터와 민간기업인 KCB의 데이터(소득, 신용등급, 대출잔액, 연체금액, 부채상환액, 신용카드 사용액 등)가 연계되었다. 물론 해당 신혼부부에게는 동의를 얻지 않은 것이다. 통계청은 KCB가 데이터 연계를 하고 연계키를 삭제하였으므로 법적으로 문제가 없다고 주장하나, KCB에게 데이터를 제공한 것이 현행법률상으로도 근거가 없는 것으로 보인다.

이와 같은 정보들을 연계하는 것이 타당한지에 대한 평가가 이루어졌는지도 알 수 없다. 통계청에서 통계작성을 위해서 데이터 연계를 할 때 관련 심사와 승인을 할 수 있는 독립적이고, 전문적인 기관이 필요하다.

통계청이 사전에 그 내용을 공개하지도 않고, 민간기업과의 협력을 통해서 서로 통계주체의 민감한 개인정보를 주고받은 것은 적절한 것으로 평가되기 어렵다. 특히 통계의 비밀보호는 엄격히 법률로서 보호되어야 하는바 법령의 규정도 없이 협력사업의 명목으로 이루어질 사안으로 보기는 어렵다.

통계자료의 공유나 활용은 엄격하게 통계의 비밀보호와 개인정보보호의 관점에서 규율되어야 한다. 그 목적은 공익목적의 연구로 제한되어야 하고, 그 절차도 비밀보호와 개인정보보호에 충분하도록 준수되어야 한다.

4) 개인정보 비식별 조치 전문기관을 통한 민간데이터 연계·결합 지원

2016년 7월 관계부처 합동으로 「범정부 개인정보 비식별 조치 가이드라인」('비식별 가이드라인')을 발간하였다. 비식별 가이드라인은 빅데이터 분석에 활용하기 위해 서로 다른 정보집합물(데이터셋)을 결합하는 공공기관 및 민간기업의 업무를 지원하기 위하여 개인정보 비식별 조치 '전문기관'을 설립하도록 하였다.

가이드라인은 서로 다른 기업이 보유한 DB를 결합하는 과정에서 권장 방식대로 적정하게 비식별 조치가 이루어진 경우에는 개인정보가 아닌 것으로 추정되고, 이를 전문기관이 결합하는 것도 현행 개인정보보호법 상 목적 외 이용·제공 조항에 위반되지 않는다고 보았다.

2016년 8월 각 부처는 국가정보자원관리원(공공기관), 한국인터넷진흥원(공공기관, 통신), 한국정보화진흥원(통신), 금융보안원(금융), 한국신용정보원(금융), 사회보장정보원(보건·복지), 한국교육학술정보원(교육) 등을 분야별 개인정보 비식별 조치 전문기관으로 지정하였으며, 9월 한국인터넷진흥원에 '개인정보 비식별 조치 지원센터'가 설치되었다.

분야별 전문기관을 통해 데이터 연계·결합이 이루어지는 절차와 운영의 주요 내용은 다음과 같다. 첫째, 정보 집합물 결합을 원하는 A 회사와 B 회사가 같은 알고리즘을 적용하여 식별자를 임시 대체키로 전환하고, 결합 상대 정보 집합물도 비식별 조치 및 적정성 평가를 수행한다. 둘째, 비식별 조치된 정보를 전문기관에 제공하고 결합을 요청한다. 셋째, 임시 대체키를 활용하여 전문기관에서 결합을 수행한다. 넷째, 전문기관이 결합을 수행한 후 임시 대체키를 삭제한다. 다섯째, 전문기관이 결합한 결합 DB를 필요한 기업에게 제공한다.

지원기관 제도가 시작된 2016년 8월부터 2017년 9월까지 26차례에 걸쳐 총 347,522,005건의 민간기업의 데이터가 결합된 것으로 나타났다.

국내 전문기관을 통한 데이터 연계·결합 지원제도의 현황과 문제점은 다음과 같이 요약할 수 있다.

첫째, 결합 목적에 있어서 제한을 두고 있지 않다. 해외 데이터 연계 제도의 경우 공익, 연구, 통계 목적을 명확하게 한정하고 그에 대한 심사도 엄격한 데 비하여, 국

내 전문기관이 수행한 대부분 사례에서 데이터 결합의 목적은 대출 심사나 마케팅 등 민간기업의 영리적 목적에 할애되어 있다.

둘째, 데이터 결합 단계별로 기능 분리가 이루어지고 있지 못하다. 개인정보보호위원회는 개인의 프라이버시 보호를 위해서는 데이터 연계과정에서 누구도 관련 데이터에 포함된 정보주체를 알아볼 수 없도록 하여야 하고, 이를 위해서 데이터 연계절차에 관여하는 각 기관의 기능이 분리되어야 하며 연계되는 데이터의 일부를 보유하고 있는 데이터 제공기관은 해당 과정에 참여할 수 없도록 해야 한다고 지적한 바 있다. 반면 국내 전문기관 제도의 경우 임시 대체키를 데이터 보유기관이 스스로 생성하여 그 메커니즘을 알도록 하였으며, 데이터 일부를 보유하고 있는 보유기관에 대규모로 결합된 데이터를 반출하고 해당 데이터의 재식별 검사도 데이터 보유기관이 스스로 수행하는 데 맡기고 있었다.

셋째, 투명성이 부족하다. 해외의 경우 데이터 연계가 승인된 모든 개인과 연구 프로젝트에 대한 등록사항이 공개되는 것이 원칙이다. 반면 국내 전문기관 제도의 경우 별도의 승인 요건이나 절차를 두고 있지 않음에도 데이터 결합 목적이나 결합 기업에 대한 정보가 공개되어 있지 않다.

넷째, 개인정보 보호법을 준수하고 있지 않다. 비식별화 가이드라인의 경우 개인정보 보호법의 보호 대상이 되는 '다른 정보와 쉽게 결합하여' 알아볼 수 있는 개인정보를 매우 좁게 해석하고 비식별 조치를 취한 데이터셋에 대하여 개인정보가 아닌 것으로 추정하였다. 그러나 법원은 결합의 용이성과 관련하여 "구하기 쉬운지 어려운지와는 상관없이 해당 정보와 다른 정보가 특별한 어려움 없이 쉽게 결합하여 특정 개인을 알아볼 수 있게 되는 것"이라 판시한 바 있다. 전문기관을 통해 자사 보유 데이터셋을 다른 기업의 데이터셋과 결합시켜 돌려받은 기업으로서는 비식별 조치 이전 상태의 데이터셋 원본 또한 가지고 있으며, 임시 대체키를 직접 생성한 상태이므로 해당 데이터셋에서 개인을 알아볼 가능성이 있다. 이런 상태에서 특정한 비식별 조치를 기술적으로 취했다는 이유만으로 해당 데이터셋 모두를 개인정보가 아닌 것으로 추정한 것은 개인정보 보호법제의 적용을 자의적으로 면제한 것이다.

다섯째, 앞서의 문제점들이 종합적으로 작용한다면 비식별 정보의 재식별 위험성도 매우 높아진다. 실제 결합사례에서 비식별 조치가 개인을 더이상 식별하기 불가능한 수준에 이르렀다고 보기 어려운 경우도 있었다. 특히 데이터셋 일부를 보유한 기관에 결합데이터를 활용하려는 동기와 이해관계가 크게 작용하고 있을 때 비식별 조치는 매우 소극적으로 이루어질 수밖에 없다. 해외 사례에서처럼 절차적으로 어떤 참여기관도 전체 데이터셋의 내용이나 식별 데이터를 다룰 수 없도록 기능을 분리하지 않은 상태에서, 보유기업 스스로의 의지만으로 재식별화에 대한 억지 효과가 달성될 것이

라고 기대하는 것은 무리이다.

5) 기타 정부 부처별 민간 데이터 연계·결합

가. 미래창조과학부

(구)미래창조과학부는 2013년부터 매년 「빅데이터 시범사업」을 실시해 왔으며, 2014년부터 「플래그십 프로젝트」 또한 공모해 왔다. 이들 시범사업 가운데 데이터 연계·결합 사례로는 2015년 비씨카드 컨소시움의 「빅데이터 시범사업」과 2016년 SK텔레콤의 「플래그십 프로젝트」를 들 수 있다.

2015년 비씨카드 컨소시움의 빅데이터 시범사업은 소셜 빅데이터와 카드 결제정보를 연계하여 소비 트렌드를 추출하고 트렌드 프로파일링 작업을 통해 신용카드 분야에서 타켓 마케팅을 실시하는 데 목표가 있었다. 컨소시움은 이 사업이 소셜 데이터와 카드사 내부의 빅데이터를 업계 최초로 결합하여 성과를 증명하였다고 자평하였다.

시범사업 결합에 사용된 BC카드의 신용카드 거래 데이터로는, 3천 6백만 명 유효카드 회원의 연령·성별·주소·연추정소득·전화번호·외국인여부·자동차보유여부·취미 등의 카드 회원 데이터, 261만 개 가맹점의 가맹점명·사업자주소·가맹점주소·전화번호·대표자정보·연평균 매출금액 등의 가맹점 정보 데이터 및 은행, 증권, 카드 등 30개 고객사 네트워크 데이터 등이었다. 결합에 사용된 소셜미디어 데이터로는, 국내 4,000개 뉴스사이트, 포털사이트 카페·블로그·지식인·댓글, 트위터·페이스북 페이지, 모네타·뽐뿌재테크 등 금융 커뮤니티/게시판, 통계청·서울시 공공데이터 및 지역단체 사이트 등 시드 웹사이트에서 수집한 정보였다.

이 사례의 경우 비록 개인 단위로 결합이 이루어지지는 않았으나 이와 같은 방식의 데이터 결합 사례는 소셜 데이터 수집 및 이용의 문제가 쟁점이 될 수 있다. 비록 정보주체인 소셜 계정 이용자가 스스로 공개한 정보이기는 하지만, 빅데이터 처리로부터 정보주체의 거부권이 보장될 필요가 있다. 현행 우리나라 개인정보보호법제의 경우 이 사업과 같은 프로파일링 처리로부터 정보주체의 권리를 보호하는 규율이 미흡하다.

SK텔레콤은 2016년도 미래성장동력 플래그십 프로젝트 사업으로 '개인정보 비식별 자료 생성·유통의 현장적용을 위한 실증 적용과제'를 수행하면서 SK텔레콤 가입자 데이터셋과 SCI 평가정보 데이터셋을 결합하였다.

SK텔레콤은 중금리 대출 이용자의 신용도 향상 가능성을 검증하기 위한 목적으로 SK텔레콤 가입자 중 제3자 데이터 제공동의를 한 서울지역 가입자를 대상으로 SCI

와 직접 데이터 결합을 진행하였다. 이때 SK텔레콤의 결합 대상은 45개 항목이었으며, 비식별 조치 후 SK텔레콤 결합 대상자 3,754,040건과 SCI 결합 대상자 15,555,049건을 연계한 결과 970,553건의 결합(동기화)이 성공하였다. 이 결합은 별도의 임시 대체키를 생성하지 않고 가입자 성, 연령대, 가입자 성별을 키값으로 하여 결합을 수행하였으며 전문기관 개입 없이 자체적으로 결합을 수행하였다.

이와 같은 방식의 데이터 결합사례의 문제점은 개인정보 보호법의 준수 여부가 불명확하다는 것이다. 별도의 대체키 없이 데이터셋에 포함된 가입자 성, 연령대, 가입자 성별을 키값으로 하여 각 데이터셋에 속한 개인 1명을 직접 연계한 경우, 이 유일한 키값을 통해서도 원본 데이터셋에 포함된 개인을 알아볼 수 있다.

나. 국토교통부

국토교통부는 교통 분야 민간 데이터에 대한 접근을 위하여 법 제도적 장치 마련을 추진해 왔다. 2015년 12월 개정된 「대중교통의 육성 및 이용촉진에 관한 법률」(대중교통법)의 경우 교통카드데이터 수집 및 제공 체계를 규정하고 지속적 활용을 위한 교통카드데이터 통합정보시스템의 구축방안을 마련하였다. 대중교통법에서는 이 시스템에서 수집 및 활용하는 교통카드데이터에 대하여 '이용자를 알아볼 수 없는 형태로 가공한 자료'로 규정하고 있으며, 개별 교통카드 정산사업자는 교통카드 번호를 암호화하여 16~64자리의 가상번호로 변환한 뒤 이를 통합정보시스템에 제공한다.

2016년 12월까지 이루어진 교통카드빅데이터 통합정보시스템 1단계 구축사업으로, 교통카드 정산사업자인 한국스마트카드 수도권 교통카드데이터 및 기반데이터가 연계되었다. 2017년 2단계 구축사업으로 그 외 교통카드 정산사업자의 교통카드데이터 및 기반데이터를 연계하고, 특정부문사업자인 한국철도공사, SR, 시외버스, 고속버스의 교통카드데이터 및 기반데이터와 발권데이터에 대한 연계를 추진하고 있다.

연계된 교통카드데이터는 한국교통연구원 등 교통관련기관, 국가기관 및 지방자치단체 등에 제공되고, 한국교통연구원은 이와 같은 과정을 통해 수집한 교통데이터를 업무협약(MOU)을 통해 삼성카드가 보유한 소비데이터와 연계하는 사업을 추진 중이다.

국가 차원에서 교통카드데이터를 수집하고 이를 집적하여 빅데이터 통합정보시스템을 구축하는 정책은 본래 대중교통 현황조사라는 공익목적으로 추진되어 왔다. 그러나 대중교통법은 교통카드번호를 암호화하는 방식으로 안전조치를 취한 후에는 아무런 제한 없이 영리적 기업을 비롯하여 "교통카드데이터를 이용하려는 자" 모두에게 제공할 것을 규정하고 있고 공공과 민간이 보유한 교통카드데이터의 광범위한 연계

또한 예상되고 있다. 공공정책을 위해 교통 빅데이터를 구축 및 이용하고자 한다면 그 목적에 부합하도록 보관 기간은 물론 수집 및 제공을 제한하는 등 데이터 생애주기별로 세심하게 설계할 필요가 있다.

(5) 정책 제안

1) 관련 법제의 정비

첫째, 데이터 연계·결합 관련 법제의 정합성 유지.

현재 개인정보보호법, 전자정부법, 공공데이터법에는 개인정보의 제공, 공개, 연계·결합 등에 관하여 동일한 영역을 법률마다 상호 모순되게 규정하고 있다. 전자정부법은 행정정보 시스템의 연계·통합 등을 주로 효율성의 측면에서 평가하고 추진하는 법률이고, 공공데이터법은 공공데이터 이용 활성화를 목적으로 하고 있으므로, 이로 인해서 개인정보보호 원칙이 훼손되거나 정보주체의 개인정보자기결정권이 침해될 수 없다는 점을 분명히 해야 한다. 이런 관점에서 세 법의 관계를 설정하고, 정합성이 있도록 개선해 나가야 한다.

둘째, 개인정보보호법 정비: 연구 및 통계목적 개인정보 처리.

연구 및 통계 목적의 개인정보 처리를 위해 보다 구체적인 원칙과 조건을 포함하는 방향으로 현행 개인정보보호법을 개정할 필요가 있다.

① 공익을 위한 유지보존의 목적, 학술적·역사적 연구의 목적 또는 통계 목적의 개인정보 처리 시 정보주체의 권리를 보호하기 위해 적절한 안전조치를 취해야 한다. 이를 조건으로 최초 개인정보 수집 시 이에 대한 동의를 별도로 받지 않았어도 이와 같은 목적의 개인정보 처리를 허용하는 규정을 마련할 필요도 있다. 단, 이는 동의를 얻는 것이 현실적으로 불가능하거나, 지나치게 비용이 많이 들거나 기술적으로 어려운 경우로 한정하는 것이 바람직할 것이다.

② 개인정보가 학술적·역사적 연구 목적, 또는 통계 목적으로 처리되는 경우, 열람권, 수정권, 처리제한권, 처리 거부권, 처리에 대한 안전조치는 일부 제한될 수 있다.

한편, 개인정보보호법 제58조에서 통계법 등에 따라 수집되는 개인정보에 대해 제3장부터 7장까지 일률적으로 적용 배제하도록 한 것은 수정될 필요가 있다. 일부 예외적인 규정 외에는 개인정보보호법의 개인정보보호원칙이 모두 적용되어야 한다. 또 연구 및 통계 목적으로 제공된 정보는 해당 목적으로만 사용되어야 한다.

셋째, 보건의료 관련 법률 및 통계법 등의 정비.

① 보건의료 관련 법률의 정비

국민의 건강정보가 방대하게 수집·활용·연계되고 있음에도 불구하고, 국민건강보험법, 암관리법, 건강검진기본법 등 현행 보건의료 관련 법률에서는 건강정보의 수집, 활용, 제공 등의 근거가 부재하거나 모호한 경우가 많았다. 또한, 수집된 건강정보를 (준)영구적으로 보유하고 있는 등 개인정보보호 원칙에서 벗어나는 경우도 많았다. 건강정보 데이터의 연계·결합이 활성화되기 위해서라도 건강정보의 수집부터 활용·제공 등 전반적인 데이터 거버넌스 체제가 정비될 필요가 있다. 무엇보다 현행 보건의료 관련 법률을 재검토하여 법적 근거를 명확하게 하고, 개인정보보호 원칙을 위반하는 건강정보의 수집 관행을 바로잡을 필요가 있다.

② 통계법의 정비

현재 통계법의 규정은 통계의 비밀보호가 충분하지 못한 상황이다. 통계자료의 목적 달성 후 폐기 의무, 통계자료를 보유하는 동안에도 통계자료의 개인식별자를 대체번호 등으로 치환하고, 연계정보는 별도로 보관하면서 대체식별자로 처리된 통계자료를 활용하는 등의 안전조치를 취하도록 해야 한다. 통계자료를 보유하는 동안 물적, 조직적, 관리적 안전조치를 유지하도록 하고, 특히 통계의 기밀유지를 위한 여러 규정을 마련할 필요가 있다. 비밀보호의 예외적인 경우로 학술적 연구와 통계 목적의 통계자료 활용에 대해서는 그 절차와 기준에 대한 엄격한 법적 근거를 갖추는 것이 필요하다.

③ 연구 및 통계 목적의 데이터 활용 및 연계에 대한 법적 규율 마련

연구 및 통계 목적의 데이터 활용 및 연계에 대해서 그 활용을 허용하면서도 분명한 규율을 마련하여 엄격한 요건 하에서 이루어질 수 있도록 할 필요가 있다. 공익적 필요가 큰 반면에 당사자의 동의를 받는 것이 극히 곤란한 부득이한 경우에는 당사자의 동의 없는 데이터 활용 및 연계를 허용할 수 있겠지만, 비밀보호와 개인정보보호를 위한 다양한 안전장치를 마련해야 한다. (i) 접근이 승인된 연구기관에 의해 이루어질 것 (ii) 적절한 연구 제안서가 제출될 것 (iii) 학술 목적으로 요청된 기밀 정보의 유형이 적시될 것 (iv) (통계작성기관 등이) 인가한 접근 시설에서 접근이 제공될 것 (v) (학술 목적으로 통계자료에 대한 접근을 허용하는 경우) 해당 정보를 제공한 해당 통계작성기관이 승인할 것

데이터 연계와 관련해서는 유엔의 <통계 및 관련 연구 목적을 위해 수행되는 데이터 통합의 기밀성 관련 원칙과 가이드라인>을 각국의 법제 및 가이드라인에 반영할 수 있을 것이다. 뉴질랜드에서 2006년에 <데이터 통합 매뉴얼>의 두 번째 버전을 만

들 때, 이 원칙과 가이드라인을 반영한 것을 참고할 필요가 있다.

대다수 국가의 법률들을 참고하여 투명성에 관한 규정도 마련하는 것이 바람직하다. 개인정보 감독기구가 데이터 거버넌스와 관련한 원칙 및 정책 수립, 다른 거버넌스 기구에 대한 자문 등 국내 데이터 거버넌스 체제 수립에 적극적인 역할을 할 필요가 있다.

2) 데이터 거버넌스 체제의 구축

국내에서는 개인정보의 비식별화 등 데이터의 개인 식별성을 어떻게 최소화할 것인지에 대해서만 초점을 맞추는 경향이 있다. 그러나 비식별화는 전반적인 거버넌스의 하나의 요소일 뿐이다. 데이터의 이용과 보호에 관련된 법제, 데이터 접근·연계 정책, 연구기관 혹은 데이터 연계기관의 인증, 심사절차, 데이터 접근 절차 등에 이르는 전반적인 거버넌스 체제를 갖추도록 노력할 필요가 있다. 제1절에서 다루었던 법제 개선과 함께, 전반적인 데이터 거버넌스 체제를 국가적인 차원에서 고민할 필요가 있다.

① 데이터 거버넌스 프레임워크

OECD 데이터 거버넌스 프레임워크의 8가지 핵심요소 등을 참고하여 보건의료, 통계 등 주요 분야에서 데이터의 수집, 이용, 제공, 연계 등 전반에 걸쳐 개인정보의 활용과 보호의 균형을 맞출 수 있는 데이터 거버넌스 프레임워크를 도입할 필요가 있다.

② 데이터 거버넌스 기구

정보 거버넌스 기구, 프로젝트 승인 기구, 연구윤리위원회 등 데이터 거버넌스 기구가 기능별로 설치될 필요가 있다. 개인정보보호위원회는 데이터 거버넌스와 관련한 원칙 및 정책 수립, 다른 거버넌스 기구에 대한 자문 등 국내 데이터 거버넌스 체제 수립에 적극적인 역할을 할 필요가 있다.

③ 연구 데이터 허브

데이터 연계·결합을 통한 데이터의 활용도를 높이고자 한다면, 연구 데이터 허브의 설치를 고려할 수 있다. 다만 과도한 개인정보 집적을 피할 수 있는 모델을 신중히 검토할 필요가 있다.

④ 데이터 연계 모델

콘텐츠 데이터의 보유, 연계키 생성, 데이터의 연계, 데이터의 제공 등을 담당하는 사람 혹은 기관의 분리를 통해 개인정보 침해의 위험성을 최소화하는 데이터 연계 모델을 도입할 필요가 있다. 해외의 많은 사례에서 ‘신뢰할 수 있는 제3자(TTP)’ 모델

은 연계키를 생성하는 기관을 데이터 연계 및 접근을 제공하는 기관과 분리하고 있다는 점을 참고할 수 있다.

⑤ 안전조치

데이터의 수집, 저장, 연계, 제공 등 전 과정에 걸쳐서 개인정보 보호 및 보안을 위한 조치들이 취해져야 한다. 비식별화 등 데이터에 대한 안전조치(safe data) 뿐 아니라 연구자(safe people), 연구 프로젝트(safe project), 환경(safe environment), 결과물(safe results) 등에 대한 안전조치가 취해져야 한다. 다만 비식별 조치 등 안전조치를 취했다는 이유로 개인정보 보호법이 규정한 관련 책임이 일률적으로 면제되는 것은 아닐 것이다.

⑥ 연구지원단

연구의 설계, 승인, 안전한 환경에서의 데이터 접근에 있어서 연구자들을 돕는 창구 역할을 하는 연구지원단의 설치를 고려할 수 있다.

⑦ 투명성과 대중 참여

대중의 신뢰를 얻기 위해서는 높은 수준의 투명성과 참여가 필수적이다. 즉, 원칙과 절차, 진행된 사업 내용에 대한 정보를 투명하게 공개해야 하며, 정책 결정 과정에 시민들과 다양한 이해당사자들이 참여할 수 있도록 해야 한다.

5. 기대효과

현재 정부는 빅데이터, 인공지능, 사물인터넷 등 4차 산업혁명에 대응한 기술 육성 및 산업 진흥을 위해 노력하고 있다. 그 일환으로 행정안전부, 과학기술정보통신부, 보건복지부, 통계청 등 제반 정부 부처에서 나름의 빅데이터 관련 사업을 추진하고 있다. 이를 위해 정부는 비식별 가이드라인을 발표하기도 했지만, 오히려 개인정보 침해 논란은 증폭되고 있다. 보다 건설적인 논의를 위해서는 비식별 조치에 대한 논란을 넘어, 데이터의 이용과 보호에 관련된 법제, 데이터 접근·연계 정책, 연구기관 혹은 데이터 연계기관의 인증, 심사절차, 데이터 접근 절차 등에 이르는 전반적인 데이터 거버넌스 체제를 어떻게 구축할 것인지 고민할 필요가 있다. 이번 보고서가 정부의 정책 추진 및 국회의 입법 과정에서 증거기반 정책 결정에 참고될 수 있기를 기대한다.

목 차

표 목 차

그 림 목 차

제 1 장 서 론

제 1 절 연구의 필요성 및 목적

인공지능, 빅데이터 등의 기술 발전으로 데이터의 수집·분석·활용이 지능정보사회에서 국가경쟁력 강화의 중요한 요소로 인식되고 있다. 한국 정부 역시 2016년 12월 29일 발표한 <제4차 산업혁명에 대응한 지능정보사회 중장기 종합대책>에서 '미래 경쟁력 원천인 데이터 자원의 가치 창출'을 추진과제, '국가데이터 관리체계를 확립하여 기계가 학습할 수 있는 대규모 데이터 기반 구축'을 세부 목표로 설정한 바 있다. 이를 위해 정부 보유 공공데이터는 기계학습이 가능한 오픈 포맷으로 전환·개방하고, 공공기관 보유 데이터도 발굴·개방 확대할 계획이다. 의료·특허·언어 등 민간 활용도가 높은 데이터를 정부 지원을 통해 기계학습이 가능한 형태의 데이터셋(Data Set)으로 구축·제공하고, 스마트시티에서 생산된 사물인터넷(IoT) 기반 센서데이터의 체계적 축적·활용을 위해 민·관 데이터 플랫폼과 연계, 제공하는 데이터 개방체계를 구축한다는 것이다.

공공정책 결정이나 연구에서 자료의 정확한 분석은 가장 중요한 관건이다. 이에 따라, 공공기관이 보유한 다종의 데이터 또는 다양한 원천(Source)의 데이터를 연계 또는 결합하는 데이터 연계(Data Linkage) 기법이 발전하고 있으며, 다양한 시도가 이루어지고 있다.

데이터 연계란 서로 다른 복수의 데이터 파일을 결합하여 보다 풍부한 정보를 제공해줄 수 있는 하나의 완전한 통합데이터를 만드는 방법을 말한다. 데이터 연계를 통해 데이터의 질을 향상하거나, 하나의 데이터 소스로는 알 수 없는 새로운 정보를 생성해낼 수 있는데, 이를 통해 정책 결정 및 연구의 질을 높일 수 있다. 반면, 이와 같은 기법은 개인정보주체의 개인정보 자기결정권 침해나 사생활의 비밀 침해 등의 문제를 야기할 수 있다. 데이터 연계에 따른 통계적 문제의 발생 가능성, 기술적·운영상의 문제점, 법적·윤리적·문화적 장벽과 같은 제도적 문제점 등도 존재한다. 일부는 기술 발전이나 경험의 축적을 통해 해결되고 있지만, 특히 제도적 문제점은 여전히 논란이 되고 있는 문제이다.

세계 주요 국가들은 공공의 이익을 위한 학술 연구를 활성화하고, 증거에 기반한 정책 수립을 지원하기 위해 데이터 연계를 허용하면서도 정보주체의 개인정보를 보호하고 신뢰할 수 있는 데이터 이용 기반을 만들기 위한 법적·제도적 노력을 하고 있다. 이러한 노력은 특히 보건의료 연구 공동체에서 발전하였는데, 예를 들어 영국에는 지역별로 잉글랜드의 임상시험연구데이터링크(CPRD), 웨일즈의 SAIL Databank, 스

코틀랜드의 eDRIS 등이 보건의료 데이터의 연계 및 접근을 제공하고 있다. 호주에서도 호주 전역에 걸쳐 보건 정보를 안전하게 관리할 수 있는 국가적인 데이터 연계기반을 구축하기 위해 인구보건연구네트워크(PHRN)가 설립되었다.

이러한 노력은 보건의료 영역을 넘어 여타 사회경제 분야로 확산되고 있는데, 영국의 행정데이터연구네트워크(ADRN)가 대표적이다. ADRN은 훈련된 사회경제적 연구자들에게 안전한 환경에서 연계되고 비식별화된 행정 데이터에 대한 접근을 허용하고 있다. 또한, 많은 나라에서 국가통계기관이 통계·연구 목적의 데이터 연계 및 제공의 중심적인 역할을 하고 있다. 이미 수집된 데이터의 활용도를 높이고 설문조사의 비용과 부담을 줄이기 위해 행정 데이터의 연구 및 통계 목적의 2차적 활용이 증가하고 있기 때문이다. 예를 들어, 미국의 데이터 연계 기반(Data Linkage Infrastructure)은 정책 분석과 연구를 위한 데이터 발견 및 안전한 분석적 접근을 가능하게 하는 토대로서, 미국 인구조사국(Census Bureau)이 담당하고 있다. 캐나다 통계청의 사회데이터연계환경(SDLE), 뉴질랜드의 통합데이터기반(IDI) 역시 통계·연구 목적의 데이터 연계와 접근을 제공하기 위해 구축된 인프라이다.

이러한 전문적 중계기관들은 안전한 환경에서 개인정보가 포함된 데이터가 결합·연계될 수 있도록 지원한다. 각 기관에 따라 형태가 조금씩 다르기는 하지만, 이들 중계기관은 데이터 보유기관들의 허브로서의 역할, 연구자나 연구 프로젝트에 대한 승인 및 지원, 데이터 연계·결합 지원, 데이터에 접근할 수 있는 보안 환경의 제공 등의 역할을 하고 있다.

물론 개인정보를 포함한 데이터를 연구·통계 목적 등으로 2차적 이용을 할 수 있으려면 법적 근거가 필요하다. 이는 애초 수집목적 외의 개인정보 처리에 해당할 수 있기 때문이다. 유럽연합의 경우 2018년 시행 예정인 GDPR에서 공익을 위한 유지보존의 목적, 학술·역사적 연구의 목적 또는 통계 목적을 위한 개인정보의 수집목적 외 처리를 인정하고 있으며, 다만 가명처리 등 안전조치를 취하도록 하고 있다. 다른 나라들 역시 개인정보 보호 법제에 연구·통계 목적의 개인정보 처리에 관한 규정을 두고 있으며, 보건의료 관련법제나 통계법 등에서 보다 구체적인 규정을 포함하고 있다. 유엔 역시 데이터 연계와 관련한 원칙 및 가이드라인으로 <통계 및 관련 연구목적을 위해 수행되는 데이터 통합의 기밀성 관련 원칙과 가이드라인>을 두고 있다.

그러나 법에서 현실의 복잡 다양한 모든 경우를 규정하는 것에는 한계가 있다. 또한, 데이터 연계를 위해서는 데이터 보유기관의 승인, 연계, 데이터 제공, 결과물의 검토 등 다양한 절차를 거쳐야 한다. 이 모든 과정이 체계적으로 조율되지 못한다면, 자칫 데이터 연계 및 제공 과정에서 개인정보 침해가 발생하거나, 반대로 공익에 기여할 수 있는 데이터의 활용을 제약할 수 있을 것이다. 따라서 데이터 연계·결합을 활

성화하기 위해서는 데이터의 수집·연계·제공에 이르는 전 과정에서 데이터의 활용 및 보호를 위한 데이터 거버넌스 체계가 수립될 필요가 있다. 관련하여 경제협력개발기구(OECD)는 환자의 프라이버시를 보호하면서 통계 혹은 연구목적으로 보건의료 데이터를 이용하기 위한 데이터 거버넌스 프레임워크를 제안하였으며, 2017년에는 건강 데이터 거버넌스에 대한 이사회 권고를 발표하기도 했다.

한국의 개인정보보호법도 '통계작성 및 학술 연구 등의 목적을 위하여 필요한 경우로서 특정 개인을 알아볼 수 없는 형태로 개인정보를 제공하는 경우'(제18조 2항 4호), 그리고 공공기관이 법률상 소관 업무 수행이 불가능한 경우에 한해 보호위원회의 심의·의결을 거쳐 제3자 제공을 허용하고 있다. (법 제18조 2항 5호) 그러나 해외 법제에 비해 구체성이 부족하고 해석상의 논란을 낳고 있다.

한편 행정자치부는 2016년 6월 30일 발표한 <개인정보 비식별 조치 가이드라인>에 따라 한국인터넷진흥원(통신·공공), 한국정보화진흥원(통신), 금융보안원(금융), 한국신용정보원(금융) 등을 비식별조치 전문기관으로 지정한 바 있다. 또한 <제4차 산업혁명에 대응한 지능정보사회 중장기 종합대책>에서는 기업들이 자유롭게 데이터 결합을 테스트할 수 있도록, 공공데이터 수요가 높은 의료·통계 분야 등의 데이터를 제한된 장소(통계청, 건강보험심사평가원, 건강보험공단 등)에서 접근·분석할 수 있는 '데이터 프리존'을 구축·운영하고 관련 법 제도를 지원한다는 계획을 밝혔다.

그러나 <개인정보 비식별 조치 가이드라인>의 경우, 가이드라인에 따라 적정하게 비식별 조치가 된 정보는 현행 개인정보보호법제에도 불구하고 개인정보가 아닌 것으로 추정하고 정보주체로부터의 별도 동의 없이 해당 정보를 이용하거나 제3자에게 제공할 수 있도록 하고 있는데, 이는 개인정보보호법의 규정과 충돌한다는 논란이 제기되고 있다. 또한, 데이터 연계의 목적, 데이터 연계의 과정, 운영상의 투명성 등의 측면에서 문제점이 지적되고 있다. 특히 국내에서 데이터 연계를 둘러싼 논란은 개인정보의 비식별화 등등 데이터의 개인 식별성을 어떻게 최소화할 것인지에 대한 기술적 측면만이 과도하게 부각되고 있다. 그러나 비식별화는 데이터 거버넌스의 한 가지 요소일 뿐이다. 예를 들어, 영국의 행정데이터연구네트워크(ADRN)가 제시하는 5가지 안전 원칙은 데이터 비식별화에 대한 Safe data의 개념뿐 아니라 연구 인력(Safe people), 연구 프로젝트(Safe project), 연구 환경(Safe environment/settings), 연구 결과물(Safe results/output)을 포괄하는 보다 폭넓은 개념이다. 주요 국가들은 공익에 기여하면서도 안전한 데이터 거버넌스를 위해서 기술적인 조치뿐 아니라 데이터의 이용과 보호에 관련된 법제, 데이터 접근·연계 정책, 연구기관 혹은 데이터 연계기관의 인증, 심사절차, 데이터 접근 절차 등에 이르는 전반적인 보호 체제를 갖추고 있다. 한국에서도 개인정보 보호 법제를 포함한 데이터 거버넌스 체계 전반으로 사회적 논

의가 확대될 필요가 있다.

데이터의 효과적인 활용을 위해서는 대중들의 신뢰가 절대적으로 중요하다. 특히 빅데이터 기술은 보다 효과적인 공공정책 결정, 소비자 편의 등 다양한 혜택을 가져올 수 있지만, 이에 따른 부정적인 영향에 대한 우려도 제기되고 있다. 예를 들어, 유럽개인정보감독관은 “빅데이터 기술 환경에서 투명성 부족, 정보 불균형, 개인정보 보호 원칙 침해의 문제가 심각하므로 국가적 대책이 필요하다”고 지적한 바 있다. 미국 대통령실의 빅데이터 보고서에서도 지적하고 있듯이, 프라이버시 및 보안에 대한 국민의 우려는 기술 확산의 걸림돌이 될 수 있다. 즉, 빅데이터 산업의 발전이나 데이터 연계를 통한 연구 활성화를 위해서도 개인정보 보호에 대한 국민의 신뢰를 확보할 필요가 있다. 영국의 경우 보건의료 데이터의 2차적 활용을 극대화하기 위한 care.data 사업이 국민적 논란 끝에 2016년 운영이 중단되었다. 영국 정보위원회(ICO)는 이 문제를 언급하며 빅데이터 처리의 투명성 부족이 대중적 신뢰 부족으로 이어져 공공데이터 공유에도 장벽이 될 수 있다고 지적하였다. 데이터의 활용과 보호를 위한 거버넌스 체제의 구축은 대중적인 신뢰를 얻기 위한 필요조건이다.

본 연구는 데이터 연계·결합과 관련된 해외 주요 국가의 데이터 거버넌스 체제를 검토하여 우리 사회에 의미가 있는 시사점을 도출하고자 한다. 또한, 보건의료 및 통계 분야, 그리고 비식별조치 전문기관을 통한 데이터 연계·결합 현황에 대한 분석을 통해 그 한계 및 문제점을 파악하고 데이터 연계·결합을 위한 제도 도입방안을 제시하고자 한다.

제2절 연구의 범위

제2장에서는 기본적인 개념에 대한 이해를 위해 데이터 연계의 개념과 유형을 설명하고, 데이터 연계가 왜 필요한지, 이에 따른 개인정보의 위험은 무엇인지, 그리고 데이터 연계과정에서 나타날 수 있는 문제점과 과제는 무엇인지 검토한다. 이어, 데이터 거버넌스의 원칙과 모델, 그리고 보다 구체적으로 데이터 연계의 원칙과 모델을 검토한다. 이러한 원칙과 모델은 제3장과 제4장에서 국내외 현황분석을 위한 틀이 될 것이다.

제3장은 해외 주요 국가의 데이터 연계·결합 현황을 분석한다. 우선 제1절에서 유럽연합 등 해외 주요 국가의 개인정보 보호법제, 보건의료 관련법제, 통계법 등을 분석한 후, 보건의료 분야(제2절), 연구목적의 데이터 연계(제3절), 통계 목적의 데이터 연계(제4절)로 구분하여 해외 주요 국가의 사례를 검토한다. 그리고 제5절에서 해외 사례에서의 시사점을 도출하고자 한다.

제4장에서는 국내 데이터 연계·결합 현황을 분석한다. 제3장과 마찬가지로, 제1절에서는 개인정보보호법 등 데이터 연계·결합 관련 국내 법제를 분석하고, 제2절 및 제3절에서 보건의료 분야 및 통계 분야의 현황을 검토한다. 제4절 및 제5절에서는 개인정보 비식별조치 전문기관 및 기타 정부 부처에서의 데이터 연계·결합 현황을 검토한다. 각 분야의 분석 과정에서 해당 영역에서의 문제점과 개선 방향을 제시하였다.

마지막으로 제5장에서는 앞 장에서의 국내외 현황분석을 기반으로 국내 데이터 연계·결합을 위한 제도개선 방안을 제안하였다.

제2장 데이터 연계와 데이터 거버넌스

제1절 데이터 연계[1)]

1. 데이터 연계의 개념과 유형

(1) 데이터 연계의 개념

데이터 연계(Data Linkage)란 두 개 이상의 출처로부터 동일인이나 동일한 사건, 기관, 장소에 연관된 정보를 함께 가져오는 것을 의미한다. 정보를 결합함으로써 단일 출처의 정보만으로는 알기 힘든 정보 요소 간의 관계가 밝혀질 가능성이 있다. 데이터 연계는 레코드 연계(record linkage), 데이터 매칭(data matching), 데이터 통합(data integration) 등으로도 불린다.[2)]

데이터를 연결할 때 사용되는 변수들은 식별 변수와 관심변수, 크게 두 종류로 구분된다. 식별 변수(Identifying variables)는 이름, 주소, 의료보험번호 등과 같이 개인 식별에 사용되는 변수로서 서로 다른 데이터셋의 연계에 사용된다. 관심변수(Variables of interest)는 나이, 성별, 수입, 질병, 직업 등 연구의 주된 관심사가 되는 변수들을 의미한다.

식별자에는 직접적인 식별자(direct identifiers)와 간접적인 식별자(indirect identifiers) 두 종류가 있다. 직접적인 식별자는 이름, 주민등록번호, 주소 등과 같이 개인을 정확히 구별해낼 수 있는 식별자이며, 데이터들을 연결할 때만 가치가 있는 정보로 연구자들은 별로 관심을 두지 않는 대상이다. 하지만, 직접 식별자는 개인 식별이 가능하므로 매우 민감한 정보일 수 있다. 그러므로 연구목적의 데이터셋에서 직접적인 식별자들은 제거한 후에 연구자들이 접근하도록 하는 것이 일반적이다.

간접적인 식별자는 다른 정보와의 결합을 통해서만 개인을 구별해낼 수 있다. 대부분의 간접적인 식별자는 위의 관심변수와 중첩되는 경우가 많다. 예를 들어, 나이, 성별, 인종과 같은 정보는 개인을 구별하는 데도 사용되지만 중요한 설명요인이 되는 경우가 많다. 간접적인 식별자를 통해서도 개인이 재식별될 가능성이 있는데, 특히 데

---

1) 이 절의 내용은 Wellcome trust (2015)의 일부 내용을 요약·발췌하여 작성한 것이며, 필요에 따라 다른 문헌도 참고하였다.

2) UN 유럽경제위원회(UNECE)의 '통계 및 관련 연구목적을 위해 수행되는 데이터 통합의 기밀성 관련 원칙과 가이드라인'에서는 데이터 매칭과 데이터 통합을 구분하고 있다. 데이터 통합은 "새로운 결과물 생산을 위해 두 개 이상 소스의 데이터를 결합하는 과정"으로, 데이터 매칭은 "서로 다른 소스의 마이크로데이터를 그 소스들에 존재하는 공통 특성에 기반을 두어 연계하는 것"으로 정의하고 있다.

이터셋이 연계될 경우 이러한 특징적인 정보들이 증가함에 따라 개인 재식별 가능성이 증가하게 된다.

(2) 데이터 연계의 유형

데이터셋의 상태나 연구의 목적 등에 따라 다양한 데이터 연계 방법을 활용할 수 있다.

가. 정확 연계(exact/deterministic linking)

정확 연계는 두 개의 데이터 출처가 고유한 참조번호(reference number)를 공유하고 있는 경우에 가능하다. 예를 들어, 영국에서는 국가보건서비스(NHS) 번호를 이용해서 NHS 의료기록들을 연결할 수 있다. 이론적으로 정확 연계의 장점은 확실하고 단순하다는 점이다. 이때 연결에 사용되는 고유식별자가 다른 개인정보를 포함하지 않는 임의번호인 경우가 좋다. 혹시 의도하지 않게 데이터가 유출될 경우, 개인정보를 포함하고 있는 고유식별자보다는 개인정보 노출의 위험성이 낮아지기 때문이다.[3)]

정확 연계는 연계 필드(match field)의 고유성과 함께, 그 데이터가 정확하다는 것을 전제로 한다. 연계 필드의 데이터에 오류가 있다면, 당연히 연계로 생성된 데이터셋에 오류가 발생할 수밖에 없기 때문이다. 이는 연계 필드를 생성하는 기관이 얼마나 많은 자원을 갖고 있는지와 정확한 연계의 중요성 정도에 달려있다. 예를 들어, 카드번호와 같이 데이터값의 정확성이 중요한 경우에는 값의 정확성을 즉시 자체 점검할 수 있는 메커니즘을 가지고 있다.

정확 연계가 반드시 임의의 참조번호에 의해서만 이루어지는 것은 아니다. 경우에 따라 이름이나 생년월일 등의 정보를 통해서도 충분히 연계가 가능할 수도 있다. 그러나 이 경우에는 연계 필드가 고유하고, 정확하다는 전제가 취약해질 수 있다.

나. 확률연계(probabilistic matching)

확률연계는 두 개 이상의 데이터셋에서 식별 변수에 해당하는 값을 비교하여 두 레코드가 동일인에 관한 기록일 가능성을 추정하는 방식이다. 이 방법은 데이터가 부정확하거나 불완전할 수 있다는 점, 그리고 데이터 출처마다 값이 다른 포맷으로 입력

3) 우리나라의 주민등록번호는 단순한 개인 식별자에서 더 나아가, 생년월일, 성별, 출생지 등 개인정보를 포함하고 있고, 주민등록번호를 통해 또 다른 개인정보와 연결되어 결과적으로 개인정보를 통합하는 연결자(key data)로 사용되고 있다. 따라서 우리나라의 경우 주민등록번호가 정확 연계의 유일 식별자로 사용되고 사고로 데이터가 공개된다면 큰 문제가 발생할 수 있다.

되었을 수도 있다는 점을 전제로 한다.

서로 다른 데이터 소스가 정확 연계를 위해 동일한 참조번호를 사용하기 위해서는 기관 간의 조정이 필요하지만, 이름, 주소, 나이, 성별 등은 많은 데이터 소스에서 일반적으로 사용하는 정보들이다. 서로 다른 데이터셋에 공통적인 고유한 참조번호가 없거나, 혹은 법적인 문제를 포함한 여러 가지 이유로 사용하기 힘들 경우 확률연계를 수행할 수 있다.

두 레코드가 동일인에 관한 것일 확률을 측정하는 것이므로 오류 가능성이 존재한다. 이러한 오류 가능성은 동일인에 관련된 레코드가 아님에도 연결하는 경우와 실제로 연결되는 레코드가 있음에도 연계되지 않는 경우가 있을 수 있는데, 전자를 1종 오류(false positives), 후자를 2종 오류(false negatives)라고 부른다.

표 2-1 데이터 연계 오류

연계 \ 실제	True Match	True Non-Match
Link	True Positive	False Positive (1종 오류)
Non-Link	False Negative (2종 오류)	True Negative

* 출처: 오미애 등 (2014)

확률연계를 수행하는 소프트웨어는 데이터를 사용자가 설정한 일정한 오차범위를 갖고, '매칭됨(matched)', '매칭되지 않음(unmatched)', '불확실(uncertain)' 등으로 분류하게 된다. 그 목적은 사용자가 '매칭됨' 및 '매칭되지 않음'은 유효하다고 생각하고, '불확실' 영역에 초점을 맞추어 검토하도록 하는 것이다. 이렇게 사람이 직접 개입함으로써(이를 clerical matching이라고 한다) 연계율을 높일 수 있다.

확률연계는 정확 연계보다 다소 주관적인 절차가 더 많을 수밖에 없다. 사람들이 많은 결정을 내려주어야 하기 때문이다. 예를 들어, 어떠한 변수의 조합을 사용하는 것이 적절한가, 연계 여부 결정을 위한 요구조건은 얼마나 엄격하게 설정되어야 하는가[4], 일치하지 않는 값들은 오류로 처리할 것인가, 어떤 변수에서의 불일치는 다른 것보다 중요한 것으로 봐야 할 것인가.[5] 또한, 확률연계를 통한 분석을 다른 연구자가 검증할 수 있도록 하기 위해서는 이러한 결정이 투명하고, 기록되고, 일관된 방식으로 이루어질 필요가 있다.

4) 예를 들어, '매칭됨'으로 판단하는 요구조건을 엄격하게 설정한다면 1종 오류는 줄일 수 있겠지만, 2종 오류는 증가할 것이다. 그 반대 역시 마찬가지다.

5) 예를 들어, 성별보다는 주소의 경우에는 같은 값이라도 훨씬 다양한 형태의 값을 가질 가능성이 클 것이다. 즉, 성별에서의 불일치가 주소에서의 불일치보다 중요하다고 판단할 수 있다.

확률연계는 그 속성상 정확 연계보다는 오류에 덜 민감하다. 그럼에도 불구하고, 데이터 연계 작업 전에 데이터 정제(data cleaning) 작업이 선행되어야 할 수 있다. 예를 들어, 문자열 데이터의 양식을 통일하거나 불필요한 부분을 제거하는 절차가 필요하다. 또한, 문자열을 비교하는 알고리즘의 특성은 연계 결과에 큰 영향을 미칠 수 있다. 예를 들어, 소리 기반 연계 알고리즘은 'beat' 와 'beath'보다는 'Jon'과 'John'이 더 가깝다고 판단할 수 있겠지만, 바이그램(bigram) 분석 알고리즘('be/ea/at'와 'be/ea/at/th'처럼 문자열을 문자 쌍들로 분리하여 분석하는 방식)에서는 반대가 될 것이다.

다. 통계적 연계(statistical linking)

정확 연계 및 확률연계가 정확한 두 개인을 연계하는 것이라면, 통계적 연계는 서로 다른 개인에 관한 두 개의 레코드가 마치 동일인에 관한 레코드인 것처럼 연계하여 분석하기 위해 개발된 방법으로 데이터 퓨전(data fusion)이라고도 한다. 예를 들어, 50세 경상도 남성인 홍길동의 어떤 특성(교육수준, 정치적 성향 등)이 다른 50세 경상도 남성과 유사하다면, 홍길동의 의료 데이터를 그와 유사한 다른 누군가의 정보와 연계했을 때 통계적으로 유의미한 결과를 얻을 수 있을 것이다. 공중보건 분야에서는 연구자가 정책 평가를 위해서 시뮬레이션 모델을 만들기 위해 사용한다.

이 방법의 장점은 연계의 질이 덜 중요해진다는 점이다. 그러나 이 방법은 많은 통계적 가정에 의존하고 있다. 우선 연계할 여러 데이터의 표본은 동일한 모집단에서 조사되었다고 가정한다. 또한, 매칭 변수가 주어졌을 때, 두 데이터셋의 관심변수가 서로 독립적이어야 한다는 것이다. 이는 하나의 연계 후보가 다른 것과 유사하다는 전제를 위해 필요하다.

그러나 통계적 연계의 문제점은 그 전제가 충족되지 않았을 경우, 분석의 결과를 신뢰할 수 없게 된다는 점이다. 두 데이터셋의 조사 혹은 생산 시점이 다를 경우, 두 데이터의 모집단이 다르거나 가중치가 다른 경우, 공통 변수 선택 등 고려해야 할 문제가 많으며, 연계 데이터의 품질을 어떻게 측정할 것인지도 중요한 문제이다. (오미애 등, 2014)

라. 다층 연계(multilevel linking)

데이터 연계가 반드시 개인 레벨일 필요 없이 환경정보와 연결되어도 유용성이 높다. 즉, 개별 단위(사람과 사람, 단체와 단체)의 연계가 아니라, 서로 다른 차원 사이에 '수직적으로'(예를 들어 개인을 의사 혹은 병원과 연계), 혹은 '수평적으로'(개인을

작은 지역의 데이터와 연계) 연계할 수도 있다. 예를 들어, HIV 감염 데이터를 지역적 데이터와 연계할 경우, 지역에 따라 다른 특성을 보여줄 수 있다. 다층연계를 사용할 경우 정확도가 높으면서도 보안상의 위험도는 낮출 수 있다.

(3) 연계 데이터 유형별 특징

가. 횡단면조사 데이터(Cross-sectional survey data)

사회경제적 데이터를 수집할 때는 주로 설문을 이용하게 되는데, 통계를 낼 목적이므로 설문조사 결과 데이터는 표면적으로는 깨끗하다. 즉, 통상적인 표준에 따라, 통상적인 정의와 메타데이터와 함께 수집·생산된다.

그러나 설문 데이터는 주로 표본조사를 하므로 이 데이터가 관심 인구의 대표성을 갖고 있는지가 관건이 된다. 적은 비용으로 가치 있는 데이터를 추출하기 위해, 군집화(clustering, 특정 지역 혹은 그룹에 초점을 맞춘다)와 계층화(어떤 외부 특성에 따라 다른 샘플링 방법을 사용) 기법이 사용된다.

설문조사 데이터는 한번 수집되면 그 정확성을 검증하기 힘들다. 설문 응답자를 추적해서 확인하는 것이 비용이 많이 들어 사실상 불가능하기 때문이다. 또한, 설문조사 데이터는 일반적으로 가명화되기 때문에 데이터 연계 시 문제가 발생할 수도 있다. 예를 들어 연령이 42세가 아니라 43세로 잘못 기록되었을 때, 이는 통계적으로는 큰 문제가 아닐 수 있지만, 데이터 연계 시에는 문제를 야기할 수 있다.

나. 코호트 연구와 종적 연구(Cohort studies and longitudinal studies)

코호트 연구에서는 코호트 기획자가 피험자를 반복적으로 인터뷰하게 되므로 응답자와 항상 연락할 수 있도록 노력하게 된다. 이런 상황에서는 추가 검사나 조회 추적 메커니즘이 제공되므로 데이터의 품질이 높아질 수 있다. 통계 측면에서 코호트 연구 데이터는 유리한 측면을 가진다.

코호트 연구의 단점은 비용이 많이 든다는 것인데, 오랫동안 복잡한 데이터를 수집하고 관리하는데 드는 비용과 피험자(참가자)에 대해 후속 조치를 하는데 드는 비용이 크기 때문이다. 한 응답자를 다른 유사한 응답자로 대체하는 것은 고려되지 않는다. 이 때문에 코호트 연구는 횡단면 설문조사에 비해 규모가 작은 경향이 있다.

연계 측면에서 본다면 코호트 연구는 횡단면 연구보다 연계가 용이하다. 피험자에 대한 정확한 식별정보를 유지하는 것이 코호트를 유지하는 데 필수적이기 때문이다. 데이터 연계를 하게 되면 코호트에 관한 향후 연구에 도움을 줄 수 있다. 주된 문제

점은 코호트 내의 사람이 이탈할 수 있다는 점이다. 예를 들어, 사망했거나 이사를 했거나 이름을 바꾸었을 수 있다.

다. 등록부 데이터(Register data)

나라별로 다양한 인구 등록부를 보유하고 있다. 인구 등록부에는 시민들의 ID카드를 시스템화하여 관리하는 방식의 범용적 등록부, 의료분야에서 암환자에게만 식별번호를 부여하여 관리하는 것과 같은 분야별 등록부가 있다.

등록부의 목적은 해당 인구에게 서비스를 제공하기 위한 것이기 때문에, 이 등록부는 포괄적이며 데이터가 정확해야 한다. 그래서 등록부는 통계적인 유용성을 가질 수 있다. 설문조사 대상이나 코호트 대상을 선정할 때 편향을 줄일 수 있고 실질적인 실험군(treatment group)과 대조군(control group)을 용이하게 만들어낼 수 있다.

국가에 따라 등록부에 개인별 ID 번호를 갖고 있는 경우에는 데이터 연계를 빠르고 정확하게 할 수 있다. 다른 ID를 사용하게 되었을 때 지속적으로 업데이트하도록 등록부를 설계함으로써 항시적으로 연계를 용이하게 할 수 있다. 북유럽 국가들이 가장 크고 범용적인 등록부 시스템을 운영하고 있다.

라. 기타 행정데이터(Other administrative data)

행정데이터는 일반적인 행정운영을 통해 수집되며, 관심 분야의 인구 전체를 대상으로 하는 경우가 많다. 따라서 행정데이터를 연계할 경우, 연구를 위해 사례의 수를 줄일 필요가 없다. 행정데이터는 앞서 언급한 바와 같이 통계의 실험군과 대조군을 편향 없이 만드는 데 활용됨으로써 조사결과의 신뢰성을 증진할 수 있다.

행정데이터의 활용은 설문조사에 비해 훨씬 비용이 적게 들고, 응답자에게도 설문에 응해야 하는 부담을 없애주는 효율적인 방법이다. 점점 더 많은 행정기관이 디지털 형태로 정보를 수집하고, 행정 목적으로 개인 식별 번호를 사용하고 있기 때문에, 서로 다른 부처의 정보들을 연계시키는 것도 용이하다. 또한, 사회 변화에 따라 소규모 그룹이나 발생 빈도가 낮은 사건들에 대한 정보 수요도 증가하였고, 정기적으로 수집되기 때문에 종적 연구에도 유용하다. 행정데이터는 설문조사보다 정확하고, 범죄행위와 같이 사회적 가치문제 때문에 설문조사에서 파악하기 힘든 측면을 파악할 수도 있다. (B.F.M. Bakker et al., 2014)

그래서 각 국가는 점차 설문조사 대신 행정데이터를 통계 및 연구 목적으로 활용하려는 경향을 보이고 있다. ‘유럽의회 및 각료이사회 규정(EC) No 223/2009 (EU통계

규정)'은 24조에서 유럽통계에 필요한 데이터를 얻기 위해 행정데이터 소스에 접근할 필요성을 규정하고 있다. 독일의 연방통계법 5a조(section 5a)는 연방통계청에서 연방통계를 작성하거나 수정하기 전에, 공공행정기관이 관련 연방통계를 생산하기 위해 사용할 수 있는 양질의 데이터를 가지고 있는지 검토하도록 하고 있다. 연방통계청이 행정데이터의 적절성을 확인하면, 이 데이터는 관련 연방통계의 편집에 이용되어야 한다. 미국의 경우에도 인구조사법에서 미 상무부 장관이 이 법에 따른 업무와 관련된 정보를 다른 부처, 기관 등에 요청할 수 있고, 인구조사 및 설문조사의 효율적이고 경제적인 수행을 위해 필요한 기록, 보고 등의 자료 복사본을 주(stats), 시(cities) 등 다른 정보 단위 혹은 민간의 개인이나 기관으로부터 획득할 수 있도록 하고 있다. 또한, 가능한 한, 그리고 통계의 종류, 적시성, 질, 범위 등과 일치하는 한도에서, 직접적인 조사 대신 다른 소스로부터 입수 가능한 정보를 획득하여 사용해야 한다고 규정하고 있다. 즉, 통계 작성에 있어서 직접적인 설문조사보다, 행정데이터 및 민간의 기존 데이터를 활용하도록 하는 것이다. 네덜란드의 경우에도 설문조사에 대한 무응답률이 높아지자 통계 작성에 있어 행정데이터의 활용도를 높여가고 있다.[6)]

그러나 행정데이터는 의미(semantics), 질, 변수의 범위 측면에서 약점을 가지고 있다. 행정데이터는 애초에 통계 목적으로 수집되는 것이 아니라 운영 목적으로 수집된다. 예를 들어, 의사가 환자의 진료 결과를 기록할 때, 같은 환자에 대해서도 의사에 따라 다르게 기록할 수 있다. 변수의 범위도 행정 목적에 필요한 한도에서 정보를 수집하기 때문에 제한적일 수 있다. 예를 들어 진료 기록에는 환자의 사회·경제적 조건에 대한 정보가 없을 수 있으며, 조세 데이터에는 인종적 특성과 관련한 정보가 필요하지 않기 때문이다. 오히려 목적 범위 외의 정보를 수집하는 것은 관련 법 위반의 가능성이 있다.

질적 측면에서도 행정데이터는 많은 사람이 오랜 기간에 걸쳐 데이터를 수집하기 때문에 여러 가지 오류가 발생할 수 있다. 반면, 행정데이터는 업무 수행과정에서 해당 오류가 발견되고 수정될 가능성이 크기 때문에 오히려 질이 높다는 주장도 있다. 그래서 해당 데이터의 정확성보다는 행정데이터가 연구나 통계 목적에 필요한 데이터를 보유하고 있는지 여부가 문제일 수 있다.

6) 유럽, 독일, 미국, 네덜란드 등의 관련 법제 및 데이터 연계 현황에 대한 상세 내용은 3장 참조.

2. 데이터 연계의 필요성과 과제

(1) 데이터 연계의 필요성

전술했듯이, 데이터 연계는 단일 데이터 소스에서는 파악하기 힘든 정보 요소 간의 관계를 파악할 수 있도록 해준다. 특히 역학 연구에 있어서 데이터 연계는 필수적인데, 2015년 웰컴트러스트(Wellcome Trust)의 보고서는 그 이유를 다음과 같이 제시하고 있다. 이 보고서는 보건의료 분야 연구를 위한 데이터 연계에 초점을 맞추고 있지만, 데이터 연계의 가치는 여타 분야에도 적용 가능할 것으로 보인다.

첫째, 많은 통계 연구는 '실험군'과 '대조군'을 요구하는 경우가 많다. 그런데 단일 소스는 실험군 혹은 대조군만 있는 경우가 많은데, 연계 데이터는 이를 구분할 수 있도록 도울 수 있다.

둘째, 연구 가능한 주제 영역의 범위를 확대한다. 특정 연구를 위해 수집된 데이터는 조금 다른 연구 주제를 해결하는데 제한적일 수 있다. 다양한 소스의 데이터를 결합함으로써 서로 다른 영역 간의 상호 관련 효과를 탐색할 수 있다. 예를 들어, 의료 데이터와 사회경제적 데이터를 결합하여, 주거 환경이나 경제적 상황 등이 특정 질병 발생에 어떻게 기여하는지 파악할 수 있다.

셋째, 데이터 연계는 종적 연구를 가능하게 한다. 특히 보건의료 연구의 경우 장기간에 걸쳐 추적 조사해야 할 경우가 많은데, 단일한 데이터베이스를 통해 장기간의 관련 사건(event)들을 추적하는 것은 응답자에게 과도한 부담을 주거나 비용이 많이 든다. 재입원 데이터, 처방 데이터 등 다른 추가 정보를 통해 정보의 정확도를 향상하고 응답자의 부담을 줄일 수 있다.

넷째, 회고 분석(retrospective analysis)이나 기대 분석(prospective analysis)에 유용하다. 어떤 질환은 명백하지 않고 오랜 기간 후에나 발병하는 경우가 있다. 반면, 어떤 질병은 그것에 영향을 미치는 요소가 환자의 과거 이력에 기인하는 경우가 많다. 이 경우 질병에 관한 연구를 위해서는 다른 목적으로 수집된 과거 정보(행정데이터, 등록 데이터 등)가 필요할 경우가 있다. 또한 전 기간에 걸쳐 데이터가 수집되는 경우, 기억에 의존한 설문 응답의 오류를 수정할 수 있다.

다섯째, 통계 기반을 확대할 수 있다. 예를 들어, 동반이환(두 만성질환을 동시에 앓고 있는 상태, co-morbidity)의 경우에는 서로 다른 사회경제적 요인이 관련되거나 동시에 여러 건강 문제가 발생할 수 있는데, 각 데이터 수집기관에서 수집한 단일 데이터 소스로는 이를 파악하기 힘들다.

여섯째, 서로 다른 데이터셋을 결합함으로써 데이터 일관성을 검증할 수 있고, 혹은 빠져있는 데이터를 채워 넣을 수 있다.

일곱째, 드문 사건의 분석에 유용하다. 그 속성상 드물게 발생하는 사건의 경우에는 단일한 데이터 소스로부터 충분한 정보를 이끌어내기 힘들다. 예를 들어, 드물게 발생하는 암에 대한 하나의 케이스를 20개의 병원이 가지고 있을 경우를 생각해보면, 하나의 병원에서는 자신의 데이터만으로 유의미한 분석을 하기 힘들지만, 데이터 연계를 통해 공통점을 찾아낼 수 있다.

여덟째, 다수준 모델링 기법을 통해 환경적 요소를 찾아낼 수 있다. 즉, 개인정보를 그룹, 지역, 시스템 등의 정보와 결합하여, 사회 내 구조를 반영할 수 있는 기여 요소를 이끌어낼 수 있다.

아홉째, 데이터 연계를 통해 시뮬레이션 모델과 같은 인구 수준의 분석 도구를 개발할 수 있다.

열째, 시의적절한 데이터 분석에 도움이 된다. 기존 데이터에 대한 접근 허락을 얻는 것도 시간이 걸리기는 하지만, 데이터 연계를 통해 설문조사보다는 데이터 수집 시간을 절약할 수 있어서 보다 시의적절하게 분석을 수행할 수 있다.

열한째, 데이터 수집에는 많은 비용이 드는데, 기존에 수집된 데이터를 재활용함으로써 공익에 기여할 수 있다.

열두째, 국가 간에 데이터를 연계함으로써 국가적 환경 효과를 비교할 수 있다. 또한, 드문 질병의 경우에는 국가 간의 데이터 연계가 충분한 사례를 수집하는 유일한 방법이 되기도 한다.

열셋째, 데이터 연계는 학제간 연구를 촉진할 수 있다. 특히 역학의 경우에는 바이러스나 박테리아뿐만 아니라, 사회경제적 요인이 공중보건에 영향을 미치는 경우가 많기 때문에, 학제간 연구가 질병의 원인과 효과를 식별하는 데 도움이 된다.

한편, 유엔 유럽경제위원회의 '통계 및 관련 연구 목적을 위해 수행되는 데이터 통합의 기밀성 관련 원칙과 가이드라인'에서는 데이터 연계(통합)의 이익을 다음과 같이 규정하고 있다.[7)]

- 새롭거나 발전된 통계의 생산
- 현재 일부 정보가 존재하는 상태에서의 측정값들을 위해 추가적인 개별 정보의

7) Principles and Guidelines on Confidentiality Aspects of Data Integration Undertaken for Statistical or Related Research Purposes, 6조

생산

- 단일 데이터 소스로부터 얻을 수 있는 것보다 더 많은 수의 단위들의 폭넓은 변수를 포함할 수 있는 통합 마이크로데이터(개인 수준의 데이터)를 이용한 연구의 수행 능력
- 기존 데이터 소스를 개선하거나 검증할 수 있을 가능성
- 응답자의 부담을 줄여줄 수 있는 가능성

(2) 프라이버시 침해와 보안

민감한 데이터를 사용하는 연구에서는 데이터를 안전하게 사용하고 있는가 하는 점이 중요 이슈가 될 수 있다. 애초 수집목적과 다르게 개인정보가 사용되거나 유통됨으로써 개인이 식별되거나 정보주체의 의도와 다르게 사용될 경우 해당 정보주체에게 크고 작은 정신적, 금전적, 신체적 손해를 가져올 수 있기 때문이다. 예를 들어, 희귀질병을 앓고 있는 사실이 주변 사람들에게 알려진다면 환자에게 정신적 고통과 대인관계에 있어서 피해를 야기할 수 있다. 의료 정보가 보험회사에 유출될 경우 보험 가입자에게 경제적인 손해를 야기할 수도 있다.

앞서 행정데이터 활용의 유용성에 대해 언급했지만, 반면 행정데이터를 연구 등 2차적인 목적으로 활용하는 것은 애초 수집목적 외 이용으로 정보주체의 기대를 벗어나 남용될 가능성이 존재한다. 또한, 서로 다른 데이터셋의 연계는 해당 데이터셋의 애초 수집목적을 벗어난 이용일 가능성이 높아진다. 따라서 데이터의 2차적 활용 및 데이터 연계는 이와 같은 프라이버시 침해보다 큰 공익적인 목적이 있으면 정당화되며, 그러한 2차적 이용을 허용한다고 하더라도 프라이버시 침해를 최소화할 수 있는 엄격한 안전장치가 전제되어야 한다.

보안(security)은 프라이버시를 보호하기 위한 요소이지만, 개인정보를 넘어선 모든 정보를 포괄하는 의미이다. 국제표준기구(ISO)는 '정보보안'을 "정보의 기밀성, 무결성, 가용성을 유지하는 것"이라고 정의하고 있다. 또한, 다른 속성으로 진실성, 책임성, 부인방지, 신뢰성이 포함될 수 있다.[8] 데이터 보안(혹은 정보보안)은 비단 내부나 외부의 공격자뿐만 아니라 의도치 않은 데이터 유실의 방지 등까지 포함하지만, 데이터 활용과 관련한 보안 문제는 크게 두 가지로 구분할 수 있을 것이다. 첫째 보안이 요구되는 데이터가 본의 아니게 유출되는 경우(accidental-release), 둘째 연구자가 의

8) "Preservation of confidentiality, integrity and availability of information. Note: In addition, other properties, such as authenticity, accountability, non-repudiation and reliability can also be involved." (ISO/IEC 27000:2009)

도적으로 데이터에서 누군가를 식별해내려고 하는 경우(deliberate-release)이다.

후자의 경우에는 현실에서의 거의 발생하지 않아 왔는데, 이에는 몇 가지 이유가 있다. 우선 연구 데이터셋은 많은 사람이 식별 데이터에 접근하는 행정데이터보다 관리 및 통제가 쉽다. 둘째, 연구 데이터의 경우 가능한 비식별 상태로 제공되기 때문에 데이터 유실의 위험이 적다. 셋째, 대부분의 연구자가 광범위한 데이터 보호 훈련을 받으며, 데이터 관리 계획은 윤리위원회 승인의 중요 요소이다. 넷째, 연구 공동체는 서로 다른 수준의 민감성을 가진 데이터를 다양한 컴퓨터 환경에서 관리하는 기술적 해결책을 개발해왔다. 즉, 연구 목적 데이터 이용과 관련한 거버넌스 체제가 잘 구축되어 있을 경우, 개인정보 침해의 위험성이 낮아질 수 있다. 반면, 이러한 거버넌스 체제가 제대로 구축되어 있지 않거나, 영리적 목적 등 다른 이해관계가 개입할 경우, 혹은 데이터에 대한 식별성이 높아질수록 그 위험성은 커질 것이다.

데이터 연계는 개인 식별에 관련된 데이터가 보다 많아지기 때문에, 유출이나 공개 시의 위험성이 더욱 증가하게 된다. 이러한 위험을 낮추기 위해 실무에서는 대부분 시설에서 연구 데이터셋 자체와 연계 프로세스를 별도로 분리하는 모델을 채택하고 있다.

(3) 데이터 연계를 위한 과제

데이터 연계의 유용성에도 불구하고, 실제 데이터 연계를 수행하기 위해서는 여러 가지 해결해야 할 장벽이 존재한다. 웰컴트러스트의 보고서는 이를 통계적 이슈, 운영/기술적 이슈, 제도적 이슈 등 세 가지 측면에서 분석하고 있다.

가. 통계적 이슈

통계적 이슈는 데이터의 부족, 연계 필드가 없거나 질이 낮은 경우, 데이터의 편향성, 부적절한 가정, 행정데이터에서 대조군의 부족, 데이터 수집 시점의 불일치 등의 문제를 포함한다.

모든 연구 데이터가 일정한 한계가 있지만, 연계 데이터의 경우에는 추가적인 문제가 발생할 수 있다. 하나의 데이터셋을 분석할 경우에는 일정한 측정 오류나 일관성의 부족은 큰 문제가 아닐 수 있지만, 데이터 연계 시에는 값의 작은 차이가 연계 비율에 큰 영향을 미칠 수 있다. 연계 필드의 일관성도 문제가 된다. 예를 들어, 하나의 데이터셋은 정확한 나이가, 다른 데이터셋은 5년 단위로 되어 있을 경우 문제가 발생할 수 있다.

또한, 데이터 연계를 통해 연계 데이터의 유용성이 향상될 수도 있지만, 반대로 사용 가능한 데이터의 양이 크게 줄어들 수도 있다. 연계를 통해 생성된 데이터셋은 두 소스 중 작은 것의 사이즈일 수 있으며, 이 경우 소스 데이터보다 실험군의 대표성이 더 상실된다. 연계 데이터의 특성이 얼마나 잘 알려져 있는지도 문제가 된다. 특히 두 데이터셋이 모두 표본 데이터일 경우, 데이터에 대한 전제 자체가 유지되지 않는다면 연계된 데이터의 특성에 대해서는 거의 알 수가 없게 된다.

이러한 통계적 이슈에 관한 연구는 계속 진행되고 있지만, 대체로 데이터 연계 이론은 안정되어 있고 논란의 여지가 많지는 않다. 이미 데이터 연계를 위한 제품들이 많이 나와 있다.

나. 운영적, 기술적 이슈

연구 목적으로 연계 데이터를 얻기 위한 프로젝트를 수행할 때, 다음과 같은 단계를 밟아야 한다.

- 연계에 대한 허가 획득
- 데이터 보유 계약의 체결
- 데이터 획득
- 연구자에의 접근 제공
- 연구에서 연계 데이터 이용

단일 데이터 소스를 이용하는 것과 달리, 연계 데이터를 이용하는 절차는 복잡해진다. 왜냐하면, 데이터를 보유한 다수 기관의 허가를 받아야 하는데, 단순히 허가를 받아야 하는 기관의 수만 증가한 것이 아니라, 기관마다 관심사나 목적, 보안에 대한 관점, 데이터 제공 시 위험에 대한 인식 정도, 연구에 대한 이해 등이 모두 다르기 때문이다.

예를 들어, 일본의 경우에는 국가 수준, 지역 수준, 그리고 공기업을 규율하는 법제가 달라서 서로 다른 기구 사이에 데이터 공유협약을 체결하거나 데이터 연계를 수행하는데 어려움을 겪고 있다. 미국에서도 국립보건통계센터(NCHS)와 관련된 데이터 공유 협정을 협의하는 데 몇 년씩 걸리고, 참여 기관의 법제담당자가 관여하게 된다. 데이터 연계 프로젝트가 계속 진행될 경우, 새로운 NCHS의 데이터가 연계될 때마다 기존 협정 문서를 다시 검토해야 한다. 법, 기술, 담당자가 변화할 때마다 데이터 공유 협정이 변경된다. (OECD, 2015)

또 하나의 이슈는 데이터 연계를 어떻게 수행할 것인가의 문제이다. 데이터 연계를

위해서는 식별 데이터에 대한 접근이 필요한데, 이 경우 한 기관이 다른 기관의 식별 데이터에 접근해야 하는 문제가 발생하게 된다. 이에 대해서는 여러 해결책이 제시되어 왔는데, 대표적인 것이 '신뢰할 수 있는 제3자(Trusted Third Party, TTP)'를 사용하는 방법이다. 즉, 제3의 기관이 관심변수를 제외한 식별 데이터만을 받아 연계에 필요한 익명의 연계 필드를 생성하고, 각 데이터 보유기관은 개인 식별자를 익명의 연계 필드로 대체하는 방법이다. TTP 모델은 선진국에서는 데이터 연계를 위한 전형적인 방법이 되고 있다.

그러나 이와 같은 제도적 장치를 하고 있지 않을 경우 데이터 연계에 실질적인 어려움이 발생할 수밖에 없다. 예를 들어, 싱가포르의 경우 데이터 연계를 수행할 TTP 및 데이터 익명화 체제가 부재하여 기관 간의 데이터 공유에 어려움을 겪고 있다. (OECD, 2015)

연구자에게 기밀 데이터에 어떻게 접근하도록 할 것인지도 문제가 된다. 이와 관련하여 기밀 데이터의 익명화, 연구자에 대한 승인, 데이터에 접근하기 위한 보안 시설, 데이터에 대한 원격접근 등의 세부 이슈들이 있을 수 있다. 해외의 주요 데이터 연계 중계기관들은 이를 위한 다양한 해결책들을 이미 채택하고 있다. 보다 구체적인 해외 사례들은 제3장에서 살펴볼 것이다.

최근에는 가공 데이터(synthetic data)에 대한 관심도 증가하고 있다. 즉, 실제 데이터와 같은 속성을 가지고 있으며, 통계 모델로부터 만들어진 가공의 데이터셋을 이용하도록 하는 것이다. 이는 실제 데이터가 아니므로 배포해도 안전한 것으로 여겨진다. 가공 데이터셋을 더 실제적으로 보이게 하려고 실제 데이터와 섞을 수도 있는데, 실제 데이터를 더 많이 사용할수록 위험성은 커지게 된다. 또한, 실제 보건의료 실태를 정확하게 평가하는 것이 중요한 연구 영역에서는 가공 데이터의 유용성은 낮다.

데이터 이용과 관련하여, 연계 데이터는 단일 소스 데이터에 비해 추가적인 문제가 발생한다. 예를 들어, 서로 다른 데이터 소스의 데이터가 서로 다른 기간에 수집된 것일 수 있고, 데이터 형식이 다를 수도 있으며, 샘플의 특성이 다를 수도 있다. 행정데이터의 경우에는 구문적(semantic) 문제가 있을 수 있어, 행정데이터로부터 연계에 필요한 데이터를 어떻게 추출할 것인지도 문제가 된다. 설문조사 데이터는 수집의 시작과 끝나는 시점이 명확하지만, 행정데이터의 경우 계속 업데이트되는 경향이 있으므로 데이터가 계속 변화할 수 있다. 반복적으로 데이터를 요청하는 경우 연계 데이터는 다른 결과를 산출할 수 있다.

연계 데이터의 경우, 하나의 데이터 보유자와만 관련된 것이 아니므로 데이터 보유자가 연계 데이터에 대한 전문성이 없고, 그래서 연계 데이터에 대한 지원은 누가 할 것인가의 문제도 존재한다. 하나의 해결책은 데이터에 대한 전문성을 가진 데이터 관

리자(데이터 연계 중계기관)를 두는 것이다. 이는 연구자 참여를 증진시킬 수 있다는 이점이 있지만, 이를 위한 별도의 비용을 고려해야 한다.

데이터 연계의 운영적 이슈들은 다소 복잡하기는 하지만, 주요 문제점들에 대한 해결책은 이미 나와 있는 상황이다.

다. 제도적 이슈

통계적, 운영적 이슈와 달리 제도적인 이슈들은 여전히 논쟁적인 부분이 많이 남아 있다. 제도적인 이슈는 크게 법적 이슈, 윤리적 우려, 문화적 장벽으로 구분할 수 있다.

법적 이슈는 누가, 어떠한 권한으로, 어떤 목적으로, 얼마나 오랫동안 데이터를 이용할 수 있는지에 대한 법적 규율의 문제이다. 통상적으로 현행 법제는 데이터에 대한 접근·이용을 위해서 정보주체의 동의를 얻거나, 동의 없이 개인정보에 접근할 수 있는 다른 법적 근거가 있어야 함을 요구한다.

정보주체의 동의는 개인의 기밀 데이터 관리 및 이용을 위한 윤리적·법적 근거를 제공한다. 그러나 동의를 얻는 데 있어서 실질적, 윤리적, 통계적 측면의 문제가 존재할 수 있다. 우선 동의를 얻는 것이 매우 어렵거나 비현실적일 수 있다. 예를 들어, 기존의 데이터를 애초 수집목적 외로 이용할 경우, 해당 정보주체를 찾아 동의를 받는 것은 현실적으로 쉽지 않은 일이다. 또한, 연구 목적으로 개인정보를 이용할 때, 동의를 받는 과정 자체가 데이터의 편향성을 가져오기도 한다. 동의를 한 사람과 하지 않은 사람의 성향이 반영될 수밖에 없기 때문이다. 유전자 정보처럼 개인이 동의하더라도 동의하지 않은 다른 가족의 정보를 드러낼 수도 있다. 동의를 얻는 과정에서 샘플 선택의 기준이 노출되어 기밀성을 침해할 수도 있다.

또한, 동의가 무엇을 의미하는지도 연구에 영향을 미친다. 좁은 의미의 설명 후 동의인가, 포괄적 동의인가, 동의가 옵트인(opt-in)이어야 하는가, 혹은 옵트아웃(opt-out)이어도 되는가 등에 따라 동의자의 수 및 연구에 미치는 제약이 크게 달라질 것이다.

이러한 문제 때문에 통상 연구 및 통계 목적으로는 정보주체의 동의 외에 데이터에 접근할 방법을 법적으로 허용해주고 있다.

법제가 명확하게 정리되었다고 하더라도, 데이터 연계 시 다수의 관할권(jurisdiction)의 승인을 얻어야 하는 문제가 남아있다. 특히 주의 자치권이 강한, 분산화된 시스템을 가지고 있는 국가인 경우에는 더욱 그러하다. 예를 들면, 한 지역의 윤리위원회의 승인을 얻더라도, 다른 지역에서 이를 인정할 것인지가 문제가 된다. 각

기관은 자신이 해당 연구에 대해 관할권을 가지고 있다고 생각하며, 다른 기관의 결정을 거부할 수 있다. 데이터 연계는 학제간 연구가 이루어질 수 있기 때문에, 서로 다른 관할권의 승인이 문제가 될 가능성이 크다.

또한, 법에 대한 해석 문제가 있을 수 있고, 대부분의 연구자는 법 전문가가 아니므로 오랜 동안의 관행이 법적 요건인 것으로 오해될 수 있다. 특히 법제가 모호한 경우에는 해당 기관들이 법제를 나름대로 해석하여 결정해야 할 책임을 지게 된다.

기밀성, 익명성 등의 개념을 어떻게 정의할 것인가에 따라 데이터에 대한 접근에 영향을 미칠 수 있다. 이를 엄격하게 해석할수록 데이터에 대한 접근이 제한될 것이다. 결국, 법제의 핵심적인 부분은 사람의 해석에 따라 달라질 수밖에 없으며, 두 기관이 연계 데이터의 기밀성에 대해 다른 판단을 내릴 수도 있다.

윤리적 문제는 개인정보에 대한 개인의 권리와 데이터 이용이 가져올 사회의 이익(혹은 사회 전체의 이익을 대변해야 하는 정부의 의무) 간의 균형을 어떻게 평가할 것인지의 문제이다. 이와 관련하여 보호자적 관점(즉, 정부는 사회에 무엇이 이익인지 알고 있다), 사회 계약적 관점(사회의 일부로서 모든 사람이 사회에 기여할 필요가 있다), 이기주의적 관점(연구를 위해 약간의 프라이버시를 희생하는 것은 결국 나에게 큰 이익이 된다) 등 다양한 철학적 논의가 존재한다.

문화적 장벽은 데이터 공유에 대한 공중, 데이터 보유자, 학술 연구자 등의 태도나 관점과 관련된다. 일반 대중들이 데이터 공유 및 연계에 대해 어떠한 태도를 보이는가는 연구 목적의 개인정보 이용에 심대한 영향을 미친다. 일반 공중은 연구 목적, 특히 학술 기관에 의한 연구 목적으로 개인정보가 이용되는 것에 대해 우호적인 태도를 보이기도 한다. 반면, 개인정보 유출의 우려, 정부의 빅브라더화에 대한 우려, 보험회사 등 영리적 목적의 개인정보 이용에 대한 우려 등을 가지고 있다. 이러한 대중들의 태도는 미디어의 영향을 받으며, 개인정보 유출 사고 등 특정한 국면에 따라 달라지기도 한다. 또한, 데이터 보유기관에 대한 신뢰가 큰 영향을 미친다.

데이터 보유기관은 위험 회피적 속성을 가지고 있다. 일반적으로 데이터 보유자는 연구 공동체보다 위험에 대해 더 보수적으로 접근한다. 이는 연구의 이익 및 기밀성 침해의 손실이 행위자 각각에 미치는 영향을 반영한다. 예를 들어, 일반의들 역시 연구의 가치를 인정하기는 하지만, 연구 목적으로 환자들의 데이터를 제공하는 것보다 환자들의 기밀성을 보호할 책임을 더 중요하게 인식할 것이다.

학제간의 차이도 하나의 장벽으로 작용한다. 학문 영역에서 표준화된 작업 관행은 학제간보다는 통상 같은 분야의 사람들과 일하는 것을 중심에 두고 있기 때문이다.

데이터 연계는 학제간 협업을 추동할 수 있지만, 이러한 장벽을 어떻게 극복할 것인지가 문제이다.

이러한 데이터 연계의 장벽은 국가마다 다르게 작용할 수 있다. 예를 들어 저소득국의 경우에는 데이터 부족이 가장 큰 문제가 된다. 반면 통합적 보건의료 서비스를 가진 고소득국의 경우 데이터 공유에 대한 제도적 장벽이 클 수 있다.

지금까지 데이터 연계의 개념과 그 필요성, 위험성, 그리고 데이터 연계를 위한 과제를 검토해보았다. 제3장에서 데이터 연계와 관련된 해외 사례를 검토하기에 앞서, 제2장 2절에서는 데이터의 안전한 활용과 보호를 위한 거버넌스의 원칙과 모델을 살펴보려고 한다.

제2절 데이터의 활용과 보호를 위한 체계

앞 절에서 살펴본 바와 같이 데이터 연계가 실제로 수행되기 위해서는 조직적, 제도적인 문제가 해결되어야 한다. 즉, 데이터 보유기관으로부터 데이터 접근을 위한 승인을 받는 것에서부터 연구자 등 실제 데이터 이용자에게 개인정보 침해 위험을 최소화하는 방식으로 데이터를 제공하는 일련의 과정이 정비되어 있어야 한다. 또한, 이러한 전 과정에 걸쳐서 적용되어야 할 원칙, 법제도, 가이드라인 등이 뒷받침되어야 할 것이다. 이와같은 데이터(정보)의 활용과 보호를 위한 체계를 데이터(정보) 거버넌스라고 할 수 있다.

이 절에서는 데이터 연계를 포함한 데이터 거버넌스의 원칙에 대한 국제적인 제안을 먼저 검토한 후, 보다 구체적인 데이터 거버넌스 및 데이터 연계 모델을 살펴보고자 한다.

1. 데이터 거버넌스의 원칙

국제적인 데이터 거버넌스의 원칙으로는 OECD의 데이터 거버넌스 프레임워크와 UN 공식통계 기본원칙이 있다. OECD의 데이터 거버넌스 프레임워크는 보건의료 데이터에, UN 공식통계 기본원칙은 '통계'에 초점을 맞추고 있지만, 전반적인 데이터 거버넌스 원칙을 도출하는데 참조할 만하다. 또한, UN은 데이터 연계(통합)에 초점을 맞춘 가이드라인으로 '통계 및 관련 연구 목적을 위해 수행되는 데이터 통합의 기밀성 관련 원칙과 가이드라인'을 두고 있다.

(1) OECD의 건강 데이터 거버넌스

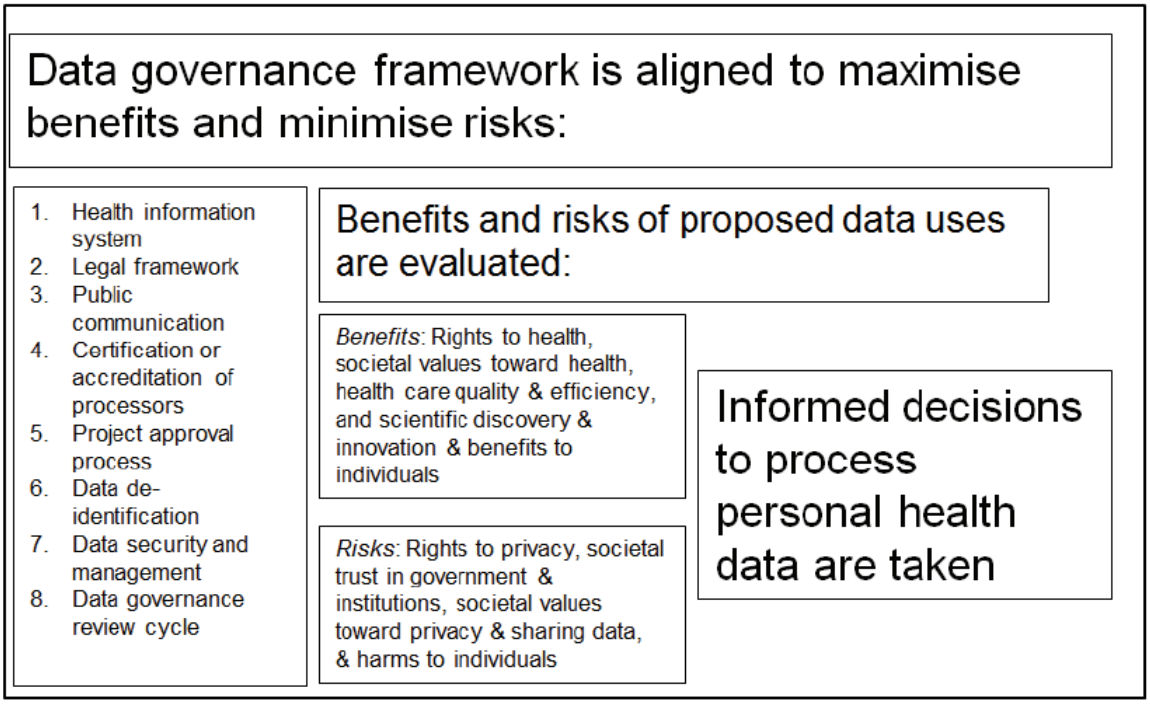
경제협력개발기구(OECD)는 환자의 프라이버시를 보호하면서 통계 혹은 연구 목적으로 보건의료 데이터를 이용하기 위한 데이터 거버넌스 프레임워크를 제안해 왔다. 2015년 보고서에서는 데이터 거버넌스 프레임워크의 주요 요소들과 프라이버시 보호적인 건강정보 사용을 위한 입법 체계 주요 요소를 제안하였고, 특히 2017년에는 건강 데이터 거버넌스에 대한 이사회 권고를 발표하였다.

가. OECD 데이터 거버넌스 프레임워크

OECD는 2015년 보고서에서 보건의료 증진을 위한 데이터의 활용성을 높이면서도, 프라이버시를 보호하기 위한 데이터 거버넌스 프레임워크를 제시하였다. (OECD, 2015) 만성적인 질환에 시달리는 환자의 치료와 서비스의 효과성을 측정하고 비교하기 위해, 그리고 보건의료 서비스 전달체계의 모델을 재설계하고 평가하기 위해서는 데이터의 더 나은 이용이 필요하다. 보건의료의 경로 및 결과에 대한 양질의 데이터는 발견과 혁신을 지원할 수 있다. 반면, 건강 데이터 이용은 환자의 프라이버시를 위험에 처하게 할 수 있다. 특히 데이터가 연계될 경우, 연계 데이터셋은 원 데이터셋보다 정보주체에 대한 더 많은 정보를 제공한다. 그 결과 연계 데이터는 그것이 유출되거나 남용되었을 때 정보주체에 더 많은 해를 입힐 수 있다.

따라서 통계 혹은 연구 목적의 이용에 대한 정책 결정이 사회적 위험과 이익을 모두 고려하면서 내려져야 한다. 이러한 정책 결정을 위해 중요한 것은 데이터 거버넌스이다. 연구 및 통계 목적의 데이터 이용을 위한 최적의 정책 결정은 이익을 극대화하고 위험을 최소화하기 위한 포괄적인 데이터 거버넌스 프레임워크가 있을 때만 성취될 수 있다.

그림 2-1 OECD의 데이터 거버넌스 프레임워크



* 출처: OECD (2015)

OECD 데이터 거버넌스 프레임워크의 8가지 핵심요소는 다음과 같다.

첫째, 건강 정보 시스템은 더 나은 보건의료 및 결과를 위한 연구 혁신과 함께, 보건의료의 질 및 시스템 성능의 모니터링과 증진을 지원해야 한다.

둘째, 공공보건, 연구 및 통계적 목적의 데이터 처리 및 2차적 이용은 개인정보 보호를 위한 법제에서 명시하고 있는 안전조치를 조건으로 허용된다.

셋째, 개인건강 데이터의 수집 및 처리와 관련하여 공중에게 정보를 제공하고 협의한다.

넷째, 연구 및 통계 목적의 건강 데이터의 처리를 위한 인증/승인 절차를 구현한다.

다섯째, 프로젝트 승인 절차는 공정하고 투명하며, 의사결정은 독립적이고 다학제적인 프로젝트 심의 기구의 지원을 받는다.

여섯째, 환자 데이터 프라이버시 보호를 위해 데이터 비식별화 모범 관행이 적용된다.

일곱째, 재식별 및 위반 위험을 줄이기 위해 데이터 보안 및 관리의 모범 관행이 적용된다.

여덟째, 새로운 데이터 소스와 기술이 도입됨에 따라 사회적 이익을 극대화하고 사회적 위험을 최소화하기 위해, 거버넌스 메커니즘은 국제적인 수준에서 정기적으로 검토된다.

데이터 이용을 극대화하면서도 위험성을 최소화하기 위해서는 보건부, 법무부, 개인정보 감독기구의 효과적인 협력이 필수적이다. 또한, 이 프레임워크가 사회적인 가치와 우선순위를 반영하기 위해서, 정부는 데이터의 이용 및 개발과 관련하여 이해당사자들과 투명하고 개방적인 소통 채널을 마련해야 한다.

OECD는 위와 같은 요소들이 보건의료 정보기반을 구축하기 위한 국내 입법에 포함되는 것이 바람직하다고 제안하고 있다. 개인 건강정보의 처리를 포함하는 연구를 승인할지 여부에 대해 균형 잡힌 의사결정을 지원하기 위해 법적으로 데이터 거버넌스 프레임워크를 수립해야 한다는 것이다.

나. 프라이버시 보호적인 건강정보 사용을 위한 입법 체계 주요 요소

데이터 거버넌스 체제를 구축하는 데 있어서, 데이터의 이용 및 개인정보 보호를 규율하는 법제가 가장 중요한 부분에 해당한다는 점은 명확하다. OECD 보고서는 건

강정보의 이용 및 보호와 관련한 각국의 법제를 검토한 후, 프라이버시 보호적인 건강정보의 이용을 위한 핵심적인 법적 요소를 제안하고 있다.

이에 따르면, 연구 및 통계 목적의 보건의료 데이터의 처리 및 2차적 이용은 개인정보 보호를 위한 법제에서 규정하고 있는 안전조치를 조건으로 허용된다. 관련 법제는 반드시 다음과 같은 내용을 포함해야 한다.

첫째, OECD 프라이버시 프레임워크에 명시된 프라이버시 보호 기본원칙을 반영해야 한다(OECD, 2013).

둘째, 모든 데이터 소스, 모든 데이터 보유기관 및 처리자를 포괄해야 한다.

셋째, 독립적이고 다학제적인 프로젝트 승인 기관 등 공정하고 투명한 프로젝트 승인 절차를 갖추어야 한다.

넷째, 승인 절차에 따라, 공중보건, 공익적인 연구와 통계를 위한 개인 건강정보의 사용을 허용해야 한다.

다섯째, 동의에 의하건, 동의 예외나 특별 권한에 의하건, 추가적으로 승인된 통계, 연구 프로젝트를 위한 데이터 처리를 허용해야 한다. 원칙적으로 데이터의 추가적인 이용을 정당한 목적이라고 간주할 수 있는 활동은 정부 통계 및 연구뿐이다.

여섯째, 미래의 승인된 연구와 통계 목적으로 사용되는 데이터셋에서 환자가 자신의 개인정보가 포함되는 것에 대해 거부권(옵트아웃)을 갖는 경우, 가능하다면 환자가 선택사항을 간소하게 표현하고 유지할 수 있는 적절한 기술 등, 환자가 자신의 권리를 행사할 수 있는 실질적인 수단이 있어야 한다.

일곱째, 개인건강 데이터셋의 승인된 목적으로의 연계를 허용해야 한다. (기록 연계)

여덟째, 승인된 데이터 연계 프로젝트나 정부 통계를 위해 연계 가능한 데이터를 공공기관 간에 공유하는 것은 허용된다.

아홉째, 공공기관과 '신뢰할 수 있는 제3자 기관'(TTP)은 미래의 승인된 데이터 연계 프로젝트나 정부 통계가 가능하도록 데이터 재식별화 키를 안전하게 보관하는 것이 허용된다.

열째, 연구 및 통계 프로젝트를 위해 사회의 모든 부문의 신청자에 의한 비식별화된 개인 단위 건강정보에 대한 공유 및 접근이 허용된다. 다만 프라이버시 및 보안을 위한 보호조치 및 재식별화를 방지하는 승인 절차를 조건으로 한다.

열한째, 연구 및 통계 프로젝트를 위해 외국인 신청자에 의한 비식별화된 개인 단위 건강정보에 대한 공유 및 접근이 허용된다. 다만 해당 국가의 입법 시스템이 국내

개인정보 보호 기준에 적절하게 부합해야 하고, 프라이버시 및 보안을 위한 보호조치 및 재식별화를 방지하는 승인 절차를 조건으로 한다.

열두째, 개인건강정보 처리 승인에 대한 모든 신청사항과 승인 결정이 일반에게 공개되어야 한다.

다. 건강 데이터 거버넌스에 대한 OECD 이사회 권고

2016년 12월 13일 OECD 이사회는 건강 데이터 거버넌스(Health Data Governance)에 대한 권고안을 채택하였으며, 2017년 1월 17일 파리에서 열린 OECD 보건부 장관회의에서 환영을 받았다. 이 권고안은 프라이버시를 보호하면서도 공익 목적의 건강정보의 이용을 가능하게 하는 국가적 건강 데이터 거버넌스 프레임워크를 개발, 구현할 것을 각국에 권고하고 있다.

이 권고안은 건강 데이터 거버넌스 프레임워크가 제공해야 할, 다음과 같은 12가지 원칙으로 이루어져 있다.

첫째, 공개적 협의를 통해 광범한 이해당사자들의 개입과 참여를 제공해야 한다. 이를 통해 이 프레임워크 하의 개인 건강 데이터의 처리가 공익에 복무하고, 사회적 가치 및 자신의 데이터가 보호될 것이라는, 그리고 건강 시스템 관리, 연구, 통계 등 공익 목적으로 사용될 것이라는 개인의 합리적 기대와 일치하도록 보장한다.

둘째, 정부 내 조정 및 공공 혹은 민간 영역에서 개인건강 데이터를 처리하는 기관 사이의 협력을 증진해야 한다. 이러한 협력은 공통의 데이터 요소 및 포맷, 질(quality)의 보증, 데이터의 상호호환성 표준을 권장해야 하며, 프라이버시 및 데이터 보안을 보호하면서도 공익 목적을 위한 데이터 공유에 대한 장벽을 최소화하기 위한 공통의 정책과 절차를 권장해야 한다.

셋째, 공익에 복무하고 공익을 보호하기 위하여, 개인건강 데이터의 처리에 사용되는 공공 영역의 건강 데이터 시스템의 역량을 평가해야 한다. 이러한 평가는 프라이버시와 데이터 보안의 보호뿐만 아니라, 데이터의 획득 가능성, 질, 사용 적합성, 접근 가능성을 포함해야 하고, 또한 적절한 보호조치를 조건으로 공익 목적으로 허용되는 데이터 처리 요소, 특히 데이터셋의 전송 및 연계를 포함해야 한다.

넷째, 개인에게 명확하게 정보를 제공해야 한다. 이러한 정보 제공은, 개인으로부터 개인건강 데이터가 수집되는 경우, 제3자에 의한 합법적인 접근을 포함한 개인건강 데이터의 처리에 대한 정보, 그 처리의 목적, 이익, 법적 근거가 명확하고, 정확하며, 쉽게 이해할 수 있고 눈에 잘 띄는 용어로 공개될 수 있도록 보장한다. 또한, 개인들은 개인건강 데이터에 대한 중대한 침해나 기타 남용에 대해 시의적절하게 고지받아

야 한다. 개별적 고지가 힘들 경우, 이러한 고지는 효과적인 공개적 홍보를 통해 이루어질 수 있다.

다섯째, 설명에 기반한 동의를 받아야 하며, 적절한 대안을 제공해야 한다. 그 동의의 메커니즘은 아래와 같은 방식으로 이루어져야 한다. 개인건강 데이터 처리에 대한 개별적 동의가 요구되는지, 그렇다면 이러한 결정을 하기 위한 기준은 무엇인지, 무엇이 유효한 동의이고 어떻게 철회할 수 있는지, 동의를 얻는 것이 불가능하거나 비현실적이거나 건강 관련 공익 목적의 달성에 부합하지 않는 상황을 포함하여 동의에 대한 합법적인 대안과 예외는 무엇인지, 그리고 그 처리가 이 권고에 부합하는 안전조치 하에서 이루어져야 한다는 것에 대해 명확해야 한다.

또한, 개인건강 데이터에 대한 처리가 동의에 기반을 두는 경우, 그러한 동의는 설명 후에 자유롭게 이루어져야만, 그리고 미래의 데이터 이용을 위한 동의를 제공하거나 철회할 수 있는 명확하고, 분명하며 사용하기 쉬운 방법을 개인들에게 제공해야만 유효하다는 것을 포함해야 한다.

개인건강 데이터의 처리가 동의에 기반을 두지 않을 경우, 실행 가능한 한 다음과 같이 이루어져야 한다. 즉, 개인은 특정한 상황에서 그 처리를 거부할 권한 및 자신의 정보가 연구 등의 목적으로 공유될 것을 적극적으로 요청할 권한을 포함하여 개인건강 데이터 처리에 대한 선호를 표현할 수 있어야 한다. 또한, 데이터 처리의 거부 혹은 요청이 존중되지 않을 경우, 개인들은 관련 법적 근거를 포함하여 그 이유를 제공받아야 한다.

여섯째, 연구 및 다른 건강 관련 공익 목적으로 개인건강 데이터를 사용하는 것에 대해 적절한 검토 및 승인 절차를 제공해야 한다. 이러한 검토 및 승인 절차는 a) 제안된 사용이 공익적인지에 대한 증거에 기반한 평가를 포함해야 하고, b) 견고하고 객관적이고 공정해야 하며, c) 적시에 그리고 결과물의 일관성을 증진하는 방식으로 운영되어야 하고, d) 정당한 이익을 보호하면서 투명하게 운영되어야 하며, e) 그 처리와 위험 최소화에 대해 개인과 사회에 미치는 이익과 위험을 평가하는 데 필요한 전문성을 가진 사람들에 의한, 독립적이고 다학제적인 검토에 의해 지원을 받아야 한다.

일곱째, 건강 데이터의 프라이버시 및 보안의 보호, 혹은 조직의 상업적 또는 다른 정당한 이익을 침해하지 않는, 정보공개 방식을 통해 투명성을 제공해야 한다. 이러한 공개 정보는 다음과 같은 요소를 포함한다. a) 개인건강 정보 처리의 목적, 건강 관련 공익, 법적 기반, b) 처리 승인의 절차와 기준, 승인된 데이터 취득자의 범주 목록을 포함한 승인 결정의 요약문, c) 건강 데이터 거버넌스 프레임워크의 이행에 대한 정보 및 얼마나 효과적이었는지.

여덟째, 개인건강 데이터의 재사용 및 분석을 가능하게 하면서, 동시에 프라이버시 및 보안을 보호하고 자신의 데이터에 대한 개인의 통제권을 촉진하는 기술의 잠재력을 극대화하고 그 개발을 증진해야 한다.

아홉째, 모니터링 및 평가 메커니즘을 제공해야 한다. 이러한 메커니즘은 개인건강 데이터 사용이 의도된 공익 목적에 부합하는지, 그러한 사용으로부터 예상된 이익이 달성되었는지 여부, 그리고 프라이버시, 데이터 보안 보호를 위한 국가적 요구조건을 준수하지 않거나, 데이터 유출이나 남용 등 어떤 부정적 결과가 발생했는지 여부에 대해 평가하고, 그러한 평가의 결과를 계속적인 개선 절차에 활용해야 한다. 이러한 개선 절차는 개인건강 데이터의 획득 가능성, 보건 연구 및 관련 활동의 필요성, 공공정책적 필요성에 대한 정기적인 검토, 그리고 프라이버시, 개인 건강정보 보호, 데이터 거버넌스와 관련된 보안 위험성의 관리를 위한 정책 및 관행의 정기적인 평가와 업데이트를 포함한다. 또한, 이러한 메커니즘은 사용되고 있는 기술의 역량, 신뢰성, 취약성을 정기적으로 검토하고 평가하기 위한 개인건강 데이터의 처리를 촉진해야 한다.

열째, 지배적인 표준 및 데이터 처리 기술에 따라, 프라이버시 및 보안 조치에 대한 적절한 훈련 및 기술 개발을 수립해야 한다.

열한째, 통제 및 안전조치의 이행을 제공해야 한다. 이러한 통제 및 안전조치는 다음을 포함한다. a) 적절한 감독 메커니즘을 수반하는, 개인건강 데이터 처리에 대한 명확하고 견고한 일련의 책임성을 제공해야 한다. b) 개인건강 데이터가 자신의 역할과 책임에 부합하며, 적용 가능한 전문가 행위 규약을 준수하는, 모든 직원을 대상으로 한 적절한 데이터 프라이버시 및 보안 훈련을 시키는 조직에 의해서만 처리되거나, 혹은 그러한 조직의 책임으로 하는 요구조건을 수립해야 한다. c) 개인건강 데이터를 처리하는 조직으로 하여금 조직의 정보보안 프로그램을 조정하고 책임지는 직원을 지정하도록 해야 한다. 이러한 프로그램은 그 조직 및 직원들에게 프라이버시 및 데이터 보안의 보호를 위한 법적 의무에 대한 정보를 제공하는 것을 포함한다.

d) 가능한 한 건강 데이터의 공익 목적의 유용성을 유지하면서도, 프라이버시 및 보안 보호를 위한 기술적, 물리적, 조직적 조치를 포함해야 한다. 이러한 조치는, 승인될 경우 재식별을 허용하면서도, 개인건강 데이터의 비식별화 등을 통해 개인 식별을 제한하고, 데이터의 목적한 사용을 고려하는 메커니즘을 포함한다. 연구 등 공익 목적의 미래 데이터 분석을 수행하기 위해, 적절할 경우 재식별이 승인될 수 있다. 또한, 이러한 조치는 이익 극대화 및 위험 최소화를 위한 처리를 목적으로 개인건강 데이터를 제 3자에게 공유할 때의 계약을 포함한다. 이러한 계약은 데이터의 안전한 전송을 위한 방법을 명시해야 하고, 위반을 효과적으로 제재하기 위한 적절한 조치를 포함해

야 한다. 또한, 이러한 조치는 실행 가능하고 적절할 경우 보안 데이터 접근 센터와 원격 데이터 접근 시설과 같은 제3자에 대한 데이터 전송에 대한 대안을 고려해야 하고, 개인건강 데이터에 접근하는 개인에 대한 엄격한 신원 확인 및 인증을 포함한다.

열두째, 개인건강 데이터를 처리하는 기관에 건강 데이터 거버넌스의 국가적 기대를 충족하고 있음을 입증할 것을 요구해야 한다. 인증 혹은 승인이 표준 이행에 도움이 되고, 인정된 거버넌스 표준을 준수하는 역량을 입증할 수 있다면, 개인건강 데이터를 처리하는 기관에 대한 인증 혹은 승인을 포함할 수 있다.

(2) UN 공식통계의 기본원칙

사회 각 부문에서 수집되는 데이터의 공유 및 활용이 가장 활발한 영역 중 하나가 통계 작성 분야이다. 국가 및 사회가 스스로를 가장 잘 이해할 수 있는 방법 중 하나로 공식통계의 중요성은 갈수록 중요해지고 있으며, 통계는 그 자체로 정부를 포함한 사회 각 주체의 정책 결정에 사용될 수 있고, 혹은 또 다른 연구의 재료로 활용될 수도 있다. 그러나 통계 작성 과정에서 사용되는 데이터에는 개인정보도 포함될 수 있기 때문에 통계 작성 과정의 전반에 걸친 데이터 거버넌스는 무척 중요하다고 할 것이다. 각국은 개인정보 보호 관련 법제 및 통계 관련 법제를 통해 인구조사, 설문조사, 데이터 기밀성 등과 관련한 원칙과 절차를 규정하고 있는데, 이러한 통계 작성의 기본원칙은 UN 차원에서도 수립되어 있다. 주요 국가의 통계 관련 법률은 제3장에서 검토하도록 한다.

1980년대 후반, 중앙 유럽 국가들이 계획 경제에서 시장 경제 체제로 전환되면서 공식통계를 규율한 일련의 원칙의 필요성이 부각되었다. 이에 1991년 유럽 통계전문가 회의(the Conference of European Statisticians)는 공식통계 기본원칙(CES/702)을 채택했으며, 이어 1992년 UN 유럽경제위원회(UNECE)의 장관급 회의에서 채택되었다. 이후 국제적인 협의 과정을 거쳐, UN 통계 위원회는 1994년 4월, 특별 세션에서 거의 같은 내용으로 'UN 공식통계 기본원칙'을 채택했다. 이 원칙이 2013년 7월 24일 경제사회이사회에서, 2014년 1월 UN 총회에서 승인되었다. (A/RES/68/261)[9]

UN은 이 결의안에서 공식통계의 완전성에 대한 대중의 신뢰가 기본적인 가치와 원칙에 대한 존중에 의존하며, 이런 맥락에서 통계 기관의 전문적 독립성과 책임성이 핵심적임을 지적하고 있다. 또한, 통계 작업의 기본적 가치와 원칙이 법 및 제도적 체제에 의해 보장되어야 함을 강조하고 있다.

---

9) https://unstats.un.org/unsd/dnss/gp/fundprinciples.aspx

UN 공식통계의 10가지 기본원칙은 다음과 같다.

원칙 1. 공식통계는 정부, 경제, 공중에게 경제, 인구, 사회 및 환경 상황에 대한 데이터를 제공함으로써 민주 사회의 정보 시스템에 필수적인 요소를 제공한다. 이러한 목적을 위해, 실제 활용성 요건을 충족시키는 공식통계는 공공 정보에 대한 시민들의 권리를 존중하면서 공식통계기관에 의해 편집되고 제공된다.

원칙 2. 공식통계에 대한 신뢰성을 유지하기 위해, 통계 기관은 과학적 원칙 및 전문적 윤리를 포함하여 철저한 전문적 고려에 따라서, 통계 데이터의 수집, 처리, 저장 및 제출을 위한 방법 및 절차에 관해 결정할 필요가 있다.

원칙 3. 데이터의 정확한 해석을 촉진하기 위해, 통계 기관은 통계의 소스, 방법, 절차에 대한 과학적 표준에 따라 정보를 제시한다.

원칙 4. 통계 기관은 통계의 잘못된 해석 및 오용에 대해 의견을 제시할 권리가 있다.

원칙 5. 통계 목적의 데이터는 통계적 설문조사든 행정 기록이든, 모든 형태의 소스로부터 가져올 수 있다. 통계 기관은 질, 적시성, 비용 및 응답자의 부담을 고려하여 소스를 선택한다.

원칙 6. 통계 기관이 통계 편집을 위해 수집한 개별 데이터는, 그것이 자연인이든 법인이든, 엄격한 기밀성이 보장되고 통계 목적으로만 사용된다.

원칙 7. 통계 시스템을 규율하는 법, 규정, 조치는 공개된다.

원칙 8. 국내 통계 기관들 사이의 조정은 통계 시스템의 일관성과 효율성 달성에 필수적이다.

원칙 9. 각국의 통계 기관이 국제적인 개념, 분류, 방법을 사용하는 것은 모든 공식적 수준에서 통계 시스템의 일관성과 효율성을 촉진한다.

원칙 10. 통계에서의 양자 간, 다자간 협력은 모든 국가에서 공식통계시스템의 증진에 기여한다.

(3) UN의 데이터 연계 원칙

데이터 연계와 관련한 원칙 및 가이드라인으로는 유엔의 '통계 및 관련 연구 목적을 위해 수행되는 데이터 통합의 기밀성 관련 원칙과 가이드라인'이 있다.[10] 이 원칙

10) Principles and Guidelines on Confidentiality Aspects of Data Integration Undertaken for Statistical or Related Research Purposes, https://www.unece.org/fileadmin/DAM/stats/publications/Confidentiality_aspects_data_integration.pdf.

과 가이드라인은 2009년 6월 유럽 통계전문가 회의에서 승인(endorse)된 것으로, 국가통계기구(National Statistical Organizations, NSOs)에 의해 수행되는 데이터 통합 작업에 적용된다.

이보다 앞서 유엔 유럽경제위원회(United Nations Economic Commission for Europe, UNECE)는 1992년 4월 15일, 유럽경제위원회 지역에서 공식통계에 관한 기본원칙에 대한 결정을 채택한 바 있는데, 이 기본원칙 6조는 “통계 편집 목적으로 통계 기관에 의해 수집된 개별 데이터는, 그것이 자연인이든 법인이든, 오로지 통계 목적으로만 사용되도록 엄격히 기밀로 보호되어야 한다”고 규정하고 있다.

통합 데이터의 이용은 단일 소스 데이터셋에 비교하여 추가적인 법적, 정책적 우려가 제기될 수 있다. 이러한 우려는 특히 프라이버시 및 개인정보 보호 요구조건과 관련된다. 그래서 통합 데이터셋에도 공식통계의 원칙이 적용되어야 하며, 2009년 원칙과 가이드라인은 공식통계 원칙 6조를 확대한 것이다. 통계 및 연구 목적의 통합 데이터셋의 생성 및 사용의 법적, 기밀성 측면을 평가하고 완화하기 위한 공통의 프레임워크를 제공한다.

이 문서에서는 데이터 연계(통합)에 대한 8가지 원칙과 함께, 각 원칙과 관련된 구체적인 가이드라인을 제시하고 있다.

원칙 1. 데이터 통합은 통계 및 관련 연구 목적으로만 국가통계기구 (및 국가통계 시스템 내의 다른 기구)에 의해서 수행되어야 한다.

가이드라인 :

(a) 이 원칙은 통계법 및/혹은 개인정보 보호 법제에 반영되어야 하며, 정부는 이를 철저히 존중해야 한다.

(b) 명확한 법적 보호가 존재하지 않을 경우, 국가통계기구는 자연인 및 법인과 관련된 데이터 통합을 하지 말아야 한다.

(c) 국가 법률이 다르게 규정하지 않는 이상, 국가 혹은 주 정부의 부처, 공공기관이 보유한 행정적 혹은 통계적 소스로부터 나온 기존 데이터의 국가통계기구에 의한 통계, 연구 목적의 이용은 특정 자연인 혹은 법인의 프라이버시를 침해하지 않는다.

원칙 2. 국가통계기구는 자신의 국가통계 임무에 부합하고, 표준 승인 절차를 완료한 이후에만 데이터 통합을 수행해야 한다. (예를 들어, 업무예시)

가이드라인 :

(a) 국가통계기구가 자연인과 관련된 행정적 혹은 규제적 목적의 데이터 이용을 포함하여, 통계 및 관련 연구 목적 이상의 임무를 가지고 있을 경우, 법에서 특별히 허가하지 않는 이상, 이 단위들에 관련된 통계 혹은 관련 연구 목적의 데이터 통합을 수행해서는 안 된다.

(b) 통계 목적으로 새로운 설문조사를 수행하기 전에, 국가통계기구가 이미 접근할 수 있는 데이터 소스의 통합이 대안적인 방안이 될 수 있는지 여부에 대해 고려해야 한다.

(c) 새로운 데이터 통합 제안은 표준 승인 절차를 따라야 한다. 이는 공식적인 업무 예시(business case) 형식일 수 있다. 업무예시 개요의 예시가 부록으로 있으나 각국은 데이터 통합 프로젝트의 승인 절차를 위한 자신만의 템플릿을 수립해야 한다. 승인 절차는 데이터 통합이 얼마나 공식통계의 생산 및 증진, 혹은 관련 연구에 기여할 수 있느냐를 확인할 수 있어야 한다.

원칙 3. 데이터 통합 프로젝트의 공익은 데이터 이용과 관련한 프라이버시 혹은 기밀성 우려와 공식통계시스템의 완전성에 미치는 위험보다 충분히 더 커야 한다.

가이드라인 :

(a) 데이터 통합은 안전한 환경에서, 그리고 공식통계시스템의 완전성에 위험을 가하지 않는 방식으로 이루어져야 한다.

(b) 법령 혹은 표준 승인 절차에 의하지 않고는, 통합되는 데이터와 관련된 직접 식별자는 통합 절차 완료 후 가능한 한 빨리 제거되어야 한다.

(c) 적절한 경우, 표준 승인 절차의 일부로서, 모든 이익, 프라이버시 우려 및 위험성이 국가통계기관에 의해 파악되고 적절히 고려될 수 있도록 보장할 책임을 가진 기구와 협의가 이뤄져야 한다. 이익의 목록은 통합 데이터셋의 장기간 보유 혹은 시간의 계획된 연장의 결과들을 포함해야 한다.

(d) 일부 국가에서는, 표준 승인 절차가 프라이버시 영향평가를 포함해야 한다는 것이 법적 요구조건이다.

(e) 합리적이고 실행 가능하다면, 데이터 제공자의 동의를 얻어야 한다.

(f) 또한 프라이버시 및 기밀성 개념은 간접적 식별(일반적으로 흔하지 않은 특성을 가진 단위의 경우)의 위험성과, 소스들보다 더 광범위한 변수를 포함하고 있는 통합 데이터셋의 증대된 민감성에 대한 신중한 관리를 요구한다.

원칙 4. 응답자에게 하지 않겠다고 특정하게 약속을 한 경우에 데이터는 통합되어서는 안 된다.

가이드라인 :

(a) (데이터 통합 업무예시와 같은) 표준 승인 절차는 자신의 데이터가 이용될 목적에 관해 응답자에게 어떠한 약속을 했는지를 조사해야 한다. 국가통계기구의 장은 제안서가 그러한 약속에 부합하지 않는 경우 데이터 통합 제안을 승인해서는 안 된다.

원칙 5. 통합 데이터는 승인된 통계 혹은 연구 목적으로만 이용되어야 하며, 애초에 승인된 목적에서 크게 벗어나는 경우 새로운 표준 승인 절차를 밟아야 한다.

가이드라인 :

법에 규정되어 있지 않으면, 다음과 같은 경우 새로 승인을 받아야 한다.

(a) 통합 절차에 사용된 소스(데이터셋)가 크게 변경되었을 경우 (예를 들어, 단위의 범주가 추가 혹은 삭제된 경우, 혹은 변수의 형식에 변경이 있는 경우), 혹은 통합 절차에 새로운 소스의 추가가 제안된 경우.

(b) 통합이 포괄하는 특성의 수를 상당히 확대하려는 경우.

(c) 통합 절차가 포괄하는 단위의 수를 상당히 확대하려는 경우 (예를 들어, 경제의 일부에서 전체 부문으로의 확대)

(d) 통합 방법이 변경되고(예를 들어 통계적 매칭에서 정확 매칭으로), 이러한 변경으로 자연인/법인의 노출 위험이 크게 변하는 경우

(e) 통합으로 인한 결과 데이터셋을 또 다른 공식통계 목적 혹은 외부의 연구자가 제안한 연구 목적으로 사용하려고 할 때, 이것이 애초의 표준 승인 절차에 포함되지 않았었던 경우.

원칙 6. 연계 데이터셋에 포함되는 단위 레코드 및 데이터 변수의 수가 승인된 목적을 위해 필요한 이상이어서는 안 된다.

가이드라인 :

(a) '승인된 목적'은 데이터 통합 업무예시에서 승인된 것이다. 단지 이 목적을 위해 필요한 데이터 변수만이 승인된 데이터 통합 작업을 위한 데이터셋에 포함되어야 한다.

(b) 통합되는 단위 레코드의 수는 승인된 목적에 필요한 최소한이어야 한다. (예를

들어, 데이터 소스의 전체를 포괄하는 샘플의 통합에 주의를 기울여야 한다.)

원칙 7. 국가통계기구는 데이터 통합을 개방적이고 투명한 방식으로 수행해야 한다.

가이드라인 :

(a) 국가통계기구가 수행하는 데이터 통합 작업의 개요뿐만 아니라, 데이터 통합에 관한 국가통계기구의 정책은 공개되어야 한다.

(b) 데이터 통합 작업의 주요 통계 결과는 공개적으로 접근 가능해야 한다. 데이터 통합 작업이 공식통계의 생산 증진을 위해 사용될 때 (예를 들어, 질의 개선을 통해), 그 공식통계의 출판은 이 조건을 충족한다. 통합 데이터베이스로부터 공개되는 통계의 메타데이터는 데이터 통합에 사용된 원래 데이터 소스에 대한 정보를 포함해야 한다.

(c) 법에서 허락하지 않는 이상, 행정기관은 합리적이고 실현 가능한 한, 응답자에게 그들의 정보가 일반적으로 통계 혹은 연구 목적으로 사용됨을 고지해야 한다.

원칙 8. 데이터 통합으로부터 나온 통합 단위 레코드 데이터(그러나 식별자는 포함하지 않은 데이터)에 대한 접근은 일반적으로 국가통계기구의 허가된 직원으로 제한된다. 다른 통계적 마이크로데이터와 관련하여, 외부의 사람에 대한 접근 허가 제안은 명확한 법적 근거가 있어야 하고, 공식통계를 위한 데이터 사용 목적에 부합해야 한다. 접근을 허가받은 사람은 그 사용이 승인된 제안서에 부합하고 허가받지 않은 사람은 데이터셋에 접근하지 않으리라는 것에 대해 법적으로 유효한 기관 및 실행상(logistical) 보장을 제공해야 한다.

가이드라인 :

(a) 가이드라인 2(b)에 따른 업무예시가 국가통계기관에 의해 승인될 경우, 국가통계기관이 수행한 통계로부터 나온 통합 마이크로데이터는 같은 국가통계 시스템의, 혹은 적절한 국내법적 근거가 있다면 초국가(supra-national) 통계 시스템의, 다른 공식통계 생산자에 의해 통계 혹은 관련 연구 목적으로 사용될 수 있다. 승인은 통계 및 관련 연구 행위가 행정 목적의 데이터 수집 혹은 처리와 조직적으로 엄격하게 분리되었는지 여부에 대한 고려를 포함해야 한다.

(b) 국가통계기구는 데이터 보유기관이 행정 혹은 규제 목적을 수행하는 데 도움이 될 경우, 통합 데이터 내의 변수에 대한 정보를 데이터 보유기관에 제공해야 한다.

(c) 사업 사례가 국가통계기구에 의해 승인된 경우, 외부 연구자는 유럽 통계전문가

회의 가이드라인(Conference of European Statisticians guidelines)인 “통계 기밀성 관리와 마이크로데이터 접근”에 따라 통합 데이터셋의 마이크로데이터에 대한 접근을 허가받을 수 있다.

(d) 수령자의 의무는 계약서에 적시되어야 하며, 수령자에 대한 기밀성 규칙의 위반은 법에 명시된 제재를 받아야 하며, 수령자 및 (적절한 경우) 후원 기관에 집행되어야 한다.

또한, 위의 8개의 원칙과 함께, 두 개의 기준을 덧붙이고 있다. 첫째, 데이터 통합은 그것이 소스 데이터 수집물의 완전성(integrity)에 실질적으로 위협이 될 경우에는 시행해서는 안 된다. 예를 들어, 응답률을 감소시킬 위험이 제기될 경우가 이에 해당한다. 둘째, 연구 목적의 데이터 통합은 이것이 공익에 복무한다는 것을 승인 절차를 통해 정당화할 수 있을 경우에만 고려되어야 한다.

한편, 이 문서의 부록으로는 업무예시(business case)의 사례를 제공하고 있다. 원칙 2에서는 새로운 데이터 통합 제안의 경우 표준 승인 절차를 따라야 하며, 이는 공식적인 사업 사례의 형식을 띨 수 있다고 제안하고 있는데, 이 경우 사업 사례는 다음과 같은 주제 영역을 다룰 것을 제안하고 있다.

(A) 목적

사업 사례는 통합 데이터가 사용될 목적을 기술해야 한다.

(B) 공식통계에의 이익

제안된 프로젝트가 어떻게 공식통계를 생산 혹은 개선시킬지 기술해야 한다. 공식통계의 개선은 정확성, 신뢰성, 관련성, 적시성, 일관성, 그리고 통계, 개념, 정의 혹은 통계를 생산하는 데 사용되거나 비용, 응답자 부담의 감소에 사용된 방법의 포괄범위의 개선을 포함할 수 있다.

(C) 다른 이익

그 프로젝트로부터 누가, 어떻게 이익을 얻을지 기술해야 한다.

(D) 위험 평가

기밀성에의 위험, 데이터 소스의 완전성에의 위험, 기타 관련 위험에 대한 평가 및 이러한 위험들을 어떻게 관리할 것인지를 포함해야 한다.

(E) 보유

사용 목적을 위해 통합 데이터셋을 얼마나 오랫동안 보유할 필요가 있는지 적시해야 한다. 이러한 보유는 정기적인 검토를 받을 수 있다.

(F) 데이터 소스

제안서는 데이터 통합을 위해 어떠한 데이터 소스가 사용될 것인지 기술해야 한다. 이는 제안된 소스 기관을 열거하고 각 기관으로부터 받을 데이터를 일상적인 용어로 기술해야 한다. 소스 데이터의 수집을 규율하는 법률이 데이터 통합 프로젝트에 미치는 함의가 작성되어야 한다.

(G) 대안들

비용, 질, 준수 부담의 최소화의 측면에서 다른 가능한 대안보다 왜 데이터 통합이 나은지에 대해 설명되어야 한다.

(H) 이해당사자

사업 사례는 데이터 통합 프로젝트의 모든 핵심적 이해당사자(내외부 모두) 및 그들과의 협의 결과를 작성해야 한다.

(I) 이름과 주소의 보유

데이터 통합 프로젝트가 연계를 위해 개인의 이름과 주소를 보유할 필요가 있을 경우, 얼마나 오랫동안 보유해야 하는지 명시해야 한다.

(J) 검토의 주기

사업 사례는 데이터 통합에 대한 검토가 어느 정도의 주기로 이루어질지 명시해야 한다.

(K) 프라이버시 영향평가

국내 법률과 관련 국가통계기구의 정책이 예외로 인정하지 않는 한, 프라이버시 영향평가가 완료되어야 한다. 또한, 프라이버시는 일반적으로 자연인에 관련되지만, 또한 어떤 기업 혹은 산업의 경우에는 법인과 관련될 수도 있다는 것을 주의해야 한다. 예를 들어, 일부 국가의 경우 농장과 같은 어떤 비법인 기업은 프라이버시를 고려할 필요가 있다.

2. 데이터 거버넌스 모델

(1) 영국 '행정데이터 작업반'의 보고서

가. 행정데이터연구센터의 구축

영국은 행정데이터의 연구 및 정책 목적의 접근 및 연계를 증진하기 위한 목적으로 2011년 12월에 '행정데이터 작업반(Administrative Data Taskforce)'을 구성하였다. 이 작업반은 경제사회연구위원회가 의학연구위원회 및 웰컴트러스트와의 협력하에 주도하였다. (Administrative Data Taskforce, 2012)

2012년 12월, 작업반은 <연구 및 정책을 위한 접근 증진(Improving Access for Research and Policy)>이라는 보고서를 발표하였다. 이 보고서는 행정데이터의 접근 및 연계 서비스를 제공할 행정데이터연구센터(ADRC)의 설립을 포함하여, 다섯 가지의 권고를 하였다.

첫째, 영국 내 4개 국가에 각각 행정데이터연구센터(ADRC)를 설립해야 한다.

둘째, 행정데이터에 관한 연구 목적 접근을 촉진하고 부처 간 데이터 연계를 효과적으로 할 수 있도록 허용할 입법이 이루어져야 한다.

셋째, 국가적, 국제적인 모범 관행에 기반하여, 영국 전역에 걸친 단일한 연구 승인 절차가 마련되어야 한다.

넷째, 일반 대중의 참여를 촉진할 전략이 수립되어야 한다.

다섯째, 연구 목적의 접근과 행정데이터 연계를 증진할 수 있는 충분한 재정 지원이 이루어져야 한다.

또한, 작업반은 행정데이터연구센터가 각 지역에 맞게 유연할 수 있음을 전제하면서도, 행정데이터연구센터가 갖추어야 할 최소 요건을 다음과 같이 제시하고 있다.

- 데이터에 접근할 수 있는 보안 시설을 갖출 것
- 데이터 보안을 보장할 것
- 법적 권한 측면에서 각 정부부처의 다양한 조건에 대응할 수 있는 유연성을 갖출 것
- 행정데이터의 연구적 가치 및 데이터 연계의 가치에 대한 이해 증진을 위한 자체 연구 역량을 갖출 것

- 데이터 관리 및 통계 분석 지원 기능을 갖출 것

- 강력한 하드웨어, 분석 소프트웨어 등 연구를 지원할 수 있는 환경을 갖출 것

- 기록 유지 시스템을 통해 모든 행위에 대해 감사(audit)가 가능하도록 할 것. 또한, 개인정보 침해가 없도록 연구 결과물에 대한 통제 시스템을 갖출 것

- 정기적으로 외부 감사를 받을 수 있는 체제를 가져야 함

- 연구자들이 데이터 접근에 걸리는 시간, 정부부처에 미치는 효율성, 증가된 연구결과물 등 성과 지표를 수립하고 이를 모니터링할 수 있는 시스템을 갖출 것

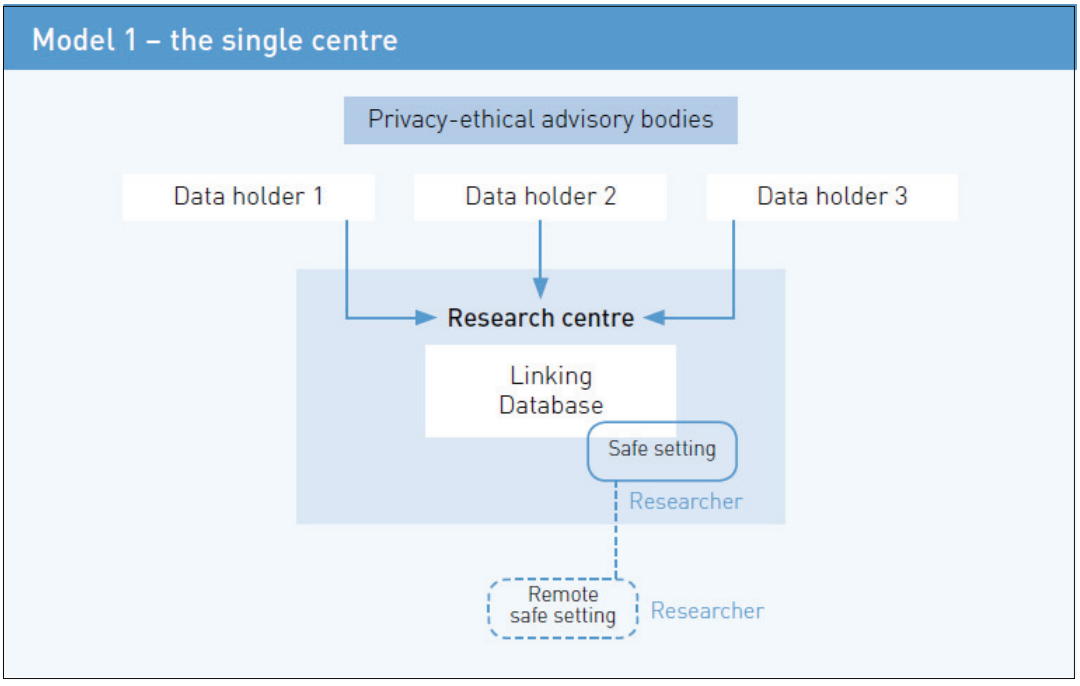
이 보고서 발표 이후, 영국에는 행정데이터연구네트워크(ADRN)이 만들어졌다. 이에 대해서는 제3장에서 자세히 검토하기로 한다.

나. 데이터 연계 모델

영국 '행정데이터 작업반(Administrative Data Taskforce)'은 이 보고서에서 4가지의 데이터 연계 모델을 검토한 바 있다.

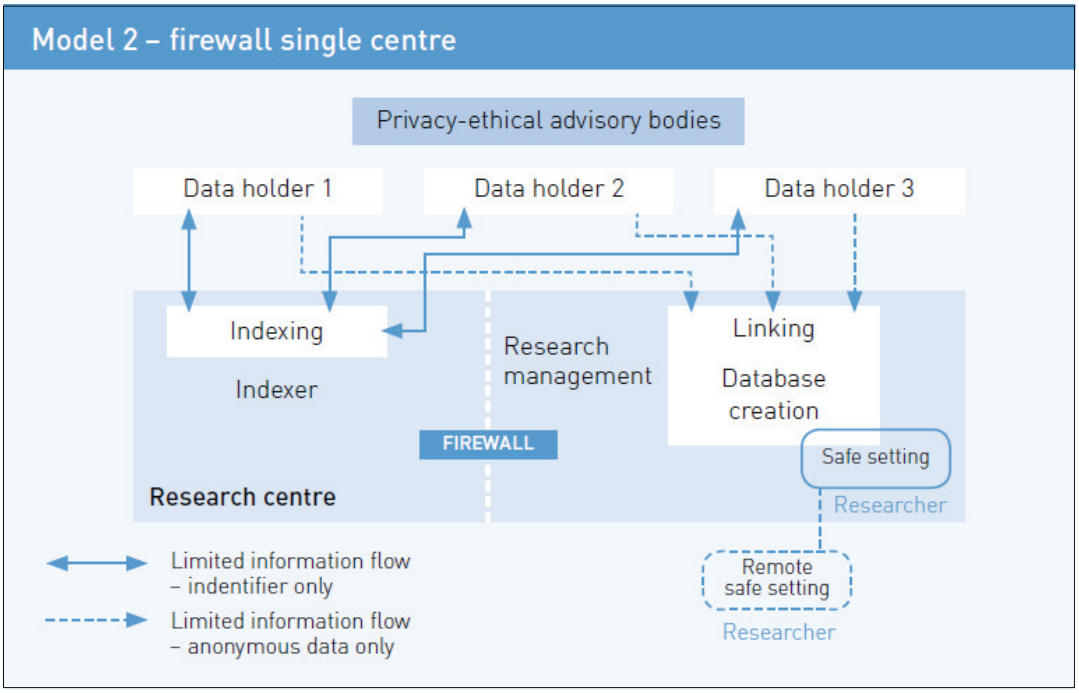
첫 번째는 '단일 센터(the single centre)'모델로서, 신뢰할 수 있는 연구센터에서 서로 다른 기관으로부터 데이터셋을 받아 연계를 수행하고, 연계 데이터를 안전한 환경에서 허가된 연구자가 접근하도록 한다.

그림 2-2 단일 센터 모델



두번째 모델은 '방화벽 단일 센터(firewall single center)' 모델인데, 하나의 센터 내에서 데이터 연계 작업(서로 다른 데이터셋을 연계할 수 있는 색인 생성)과 데이터셋 통합 및 제공 작업이 이루어지지만, 이 두 기능이 기관 내부에서 서로 엄격하게 분리된다. 데이터 보유기관은 센터 내의 색인팀(indexer)에 데이터 연계에 필요한 식별 정보만을 보내게 되며, 색인팀은 데이터 연계 후 각 레코드에 고유한 연구 식별자(study identifier) 혹은 색인키(색인 ID)를 붙여서 데이터 보유기관에 보내게 된다. 데이터 보유기관은 이 연구 식별자와 속성정보를 포함한 (그러나 다른 개인 식별자는 포함하지 않은) 데이터셋을 센터 내의 연구 관리(research management)팀에 보낸다. 연구 관리팀은 연구 식별자를 매개로 데이터셋을 통합한 후 연구자에게 제공하게 된다.

그림 2-3 방화벽 단일 센터 모델



세 번째 모델은 '신뢰할 수 있는 제3자 색인(trusted third party indexing)' 모델이다. 두 번째 모델에서 색인 기능을 센터 내부가 아니라, 완전히 다른 기관에서 수행하게 된다. 신뢰할 수 있는 제3자(TTP)는 연구센터에 어떠한 데이터도 보내지 않으며, TTP와 연구센터 사이의 정보 교환은 제한된다. 연구센터는 데이터 보유기관으로부터 색인 키와 익명화된 데이터셋을 받게 된다. 결국, TTP는 두 데이터셋의 연계를 위한 색인키만 생성할 뿐 속성정보는 볼 수 없고, 데이터 보유기관 역시 연계되는 다른 데이터 보유기관의 데이터에 접근할 수 없으며, 연계기관 혹은 연구자는 연계된 속성정보만을 볼 수 있을 뿐 개인을 식별할 수 있는 식별 정보에는 접근할 수 없게 된다.

그림 2-4 신뢰할 수 있는 제3자 색인 모델

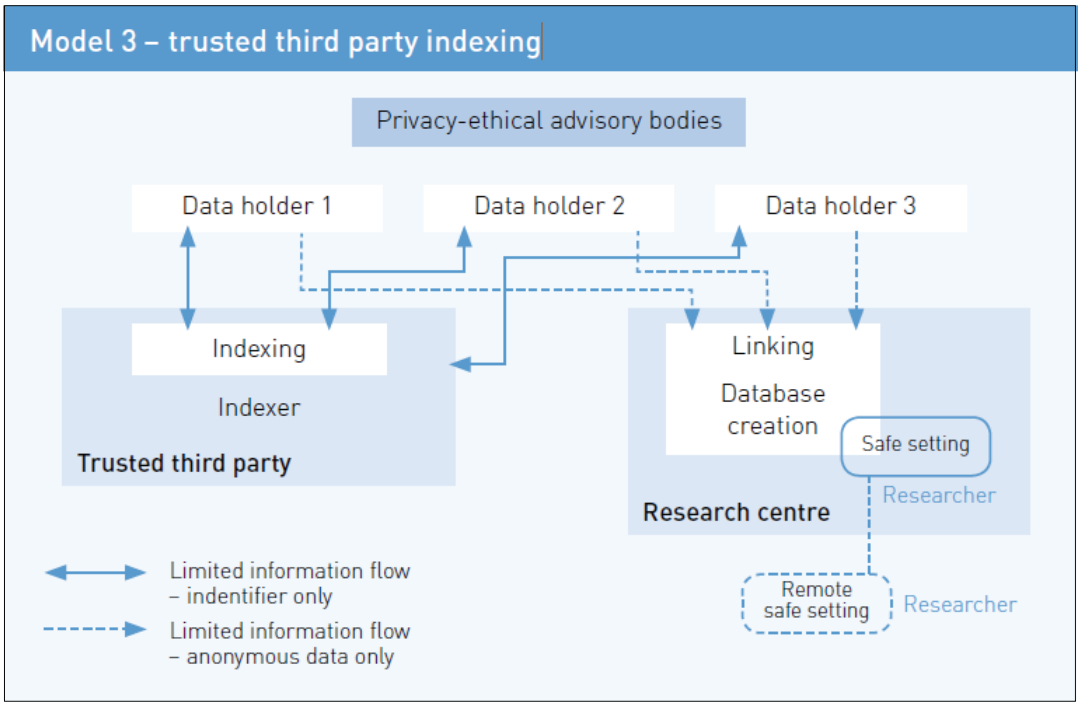
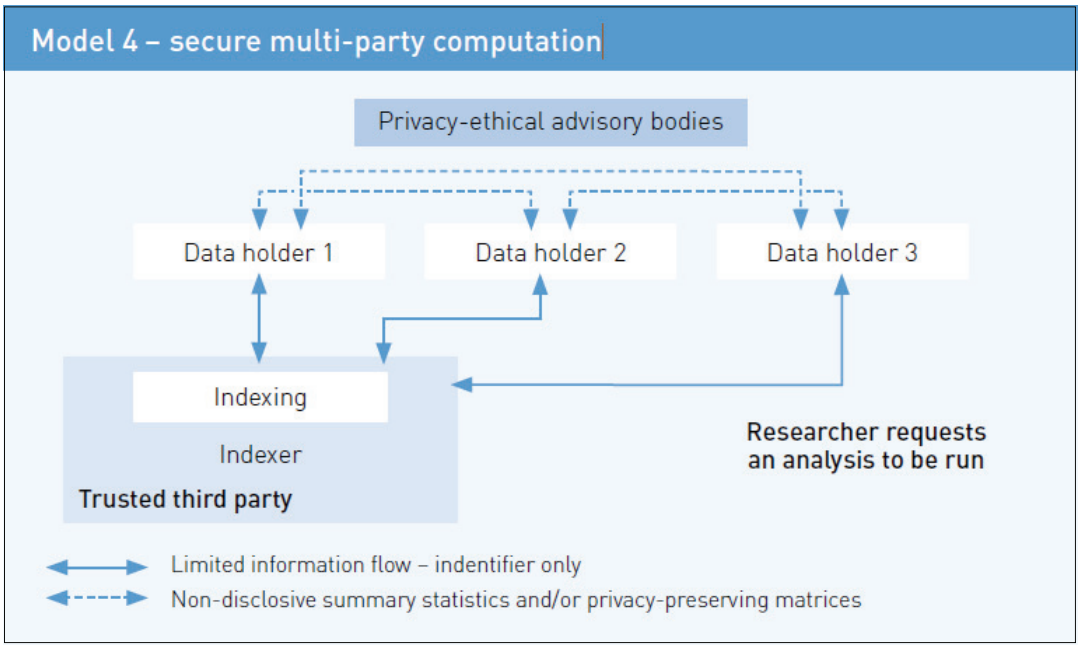


그림 2-5 다자간 보안 계산 모델



네 번째는 '다자간 보안 계산(secure multi-party computation)' 모델이다. 이 모델에는 연구센터가 존재하지 않는다. 연구자가 일련의 데이터 보유기관에 데이터를 요청하면, 데이터 보유기관이 서로 간에 보안이 되는 방식으로 데이터셋을 산출해서 그 결과물을 연구자에게 제공하는 방식이다. 공통의 식별자가 필요하기 때문에 '신뢰할

수 있는 제3자'가 색인을 위해 필요할 수 있다. 데이터 보유기관 사이에 서로 다른 기관의 속성정보를 노출하지 않으면서도 연계 데이터셋을 생성해야 하기 때문에, 데이터 보유기관 사이에 보안이 되는 방식으로 데이터셋 생성을 위한 계산이 이루어진다.

행정데이터작업반은 각 모델의 장단점을 다음과 같이 정리하고 있다.

표 2-2 데이터 연계 모델의 장단점

	장점	단점
모델 1 : 단일센터	- 데이터 전송 최소화 (전송 과정의 위험성 최소화) - 단일 기관이므로 감독 용이 - 단일 기관이므로 효율적 - 단일 기관에서 처리되므로, 연계의 질 및 편향성 평가 측정이 효과적임. - 시간이 지나면서 메타데이터, 프로그램 수집 가능. 행정 데이터셋에 대한 더 깊은 이해 촉진.	- 부정행위를 막을 수 있는 가시적 장벽이 없음. - 개인정보 익명화를 보장할 수 있는 구조가 없음. - 연구센터의 정직성에 완전히 의존해야 함.
모델 2 : 방화벽 단일센터	- 부정행위를 막을 수 있는 구조적 장벽 제공 - 단일 기관이므로 감독 용이 - 단일 기관이므로 효율적 - 단일 기관에서 처리되므로, 연계의 질 및 편향성 평가 측정이 효과적임. - 시간이 지나면서 메타데이터, 프로그램 수집 가능. 행정 데이터셋에 대한 더 깊은 이해 촉진.	- 부정행위를 막을 수 있는 장벽을 제공하기는 하지만, 외부적으로 가시적이지는 않음. - 연구센터가 기능 분리(방화벽)를 철저히 준수하고 있는지에 좌우됨. - 데이터 전송이 많아짐. (전송 과정의 위험성 증가)
모델 3 : 신뢰할 수 있는 제3자 색인	- 부정행위를 막을 수 있는 가시적인 구조적 장벽 제공 - 개인정보 익명화를 보장할 수 있는 구조 존재. - 시간이 지나면서 메타데이터, 프로그램 수집 가능. 행정 데이터셋에 대한 더 깊은 이해 촉진.	- 많은 기관이 관련되어 감독이 힘들어짐. - 많은 기관이 관련되어 비효율적. - 데이터 전송이 많아짐. (전송 과정의 위험성 증가) - 두 기관에 의해 처리되므로, 연계의 질 및 편향성 평가 측정 곤란.
모델 4 : 다자간 보안 계산	- 부정행위를 막을 수 있는 가시적인 구조적 장벽 제공 (개인 수준 데이터 접근 불가) - 개인정보 전송이 없음. (프라이버시 보호가 되는 매트릭스만 전송되며, 결과물이 드러나지 않도록 분석 방식의 통제 가능)	- 컴퓨터/통계 과학에서 아직 개발 중인 분야임. - 분석이 불가능한 부분도 (아직) 있음.

영국 행정데이터작업반은 ADRN의 모델로서 세 번째 TTP 모델이 가장 신뢰할만하다고 보았다.

그러나 TTP 모델에서도 민감한 개인 식별자가 데이터 보유자의 통제를 벗어날 가능성이 있다. 데이터 보유자가 TTP에 개인 식별자를 전달하기 때문이다. 이를 피하기 위해 '신뢰할 수 없는 제3자(Untrusted Third Party, UTP)' 모델을 사용할 수 있다. UTP 모델에서는 개인 식별자를 알려진 절차에 따라 임의의 식별자로 변환한 후에, 제3자에게 전달되게 된다. 이때 임의의 식별자를 개인 식별자로 역으로 변환할 수는 없다. 이후의 절차는 TTP와 같다.

이 방법은 개인 식별자가 데이터 보유자를 벗어나는 것을 허용하지 않기 때문에, 종종 프라이버시 보호 레코드 연계(Privacy Preserving Record Linkage, PPRL)라고도 부른다. 프라이버시 측면에서는 매력적이지만, 실제 구현하는 데 어려움도 있다. 가장 큰 단점은 확률연계에 사용하기 쉽지 않다는 것이다. 확률연계의 경우, 연계자(linker)가 실제 데이터를 볼 수 없기 때문에 동일인에 대한 약간 다른 값을 구별하기 힘들기 때문이다. (Wellcome trust, 2015)

해외 각국이 채택하고 있는 데이터 연계 방식은 조금씩 다르다. 제3장에서 각국의 데이터 연계의 사례를 구체적으로 살펴볼 것이다.

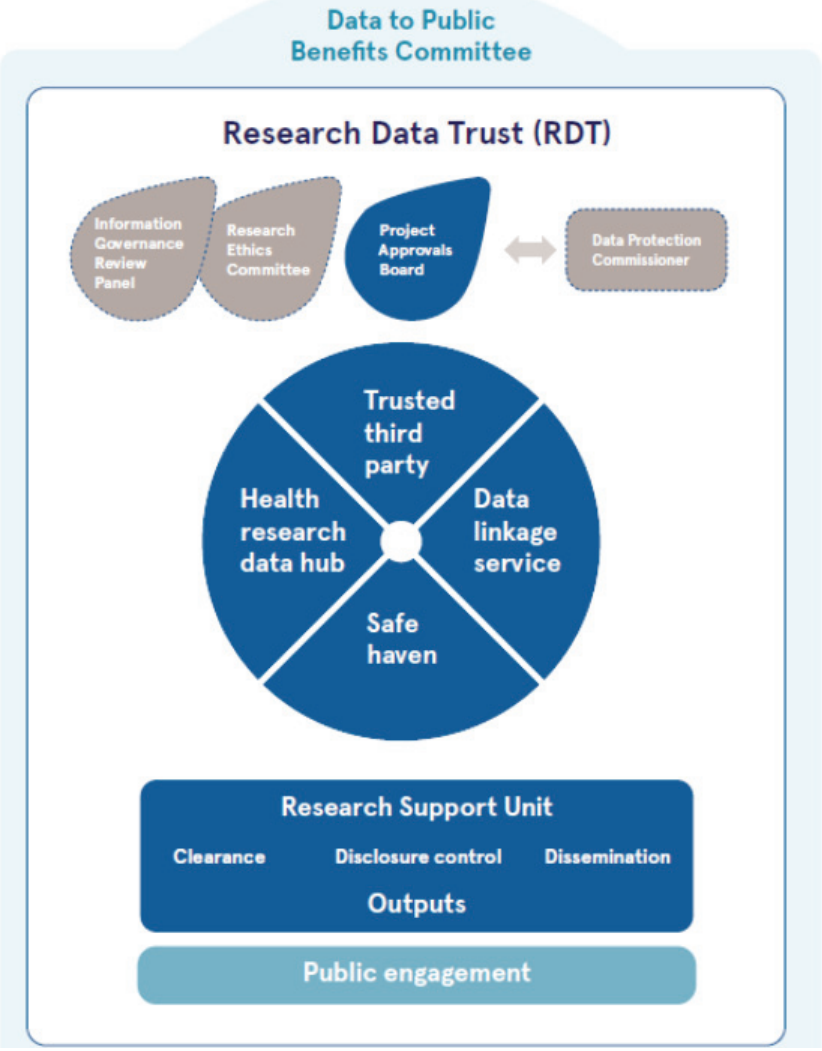
(2) DASSL 모델

전반적인 데이터 거버넌스 모델로서 Rosalyn이 제안한 DASSL Model 역시 참조할 만하다. DASSL은 데이터 접근, 저장, 공유 및 연계(Data Access, storing, sharing and linkage)를 의미하는데, Rosalyn은 해당 논문에서 아일랜드의 보건의료 데이터 거버넌스 체제 구축을 위해 이 모델을 제안하였다. Rosalyn은 이 논문에서 다른 나라의 실제 사례를 검토한 후, 이를 근거로 DASSL 모델을 고안하였다. (Rosalyn Moran, 2016)

DASSL 모델은 다음과 같은 7개의 주요 요소로 구성된다.

1. 거버넌스
2. (보건) 연구 데이터 허브
3. 제3자 데이터 연계 서비스
4. 안전시설(safe haven)
5. 연구 지원단(Research Support Unit)
6. 결과물 점검 및 공개 통제(disclosure control)
7. 대중 참여 및 소통

그림 2-6 DASSL Model



* 출처: Rosalyn Moran (2016)

개인정보 보호와 데이터 이용을 통한 공익의 균형을 달성하기 위해, 원칙에 기반한, 비례적인, 위험 기반의 접근을 채택함으로써 데이터의 공유 및 연계를 포함한 연구 프로젝트에 대한 최적의 거버넌스가 이루어질 수 있다. 이를 위한 구조로서 정보 거버넌스 검토 패널, 연구 윤리위원회, 프로젝트 승인 위원회 등이 있을 수 있다.

연구 데이터 허브는 데이터 보유자와의 계약과 합의된 거버넌스 체제를 통해 이미 수집된 데이터에 대한 안전한 접근을 제공, 활성화하는 역할을 한다. 제3자 데이터 연계 서비스는 개인정보를 포함하고 있는 데이터셋에 대한 연계 서비스를 제공한다.

안전 시설은 연구자가 개인정보를 안전하게 접근·처리할 수 있도록 지원하는 보안 시설이다. 연구지원단은 데이터 접근, 연계와 관련하여 연구자에게 다양한 지원을 제

공한다. 결과물 검토 및 공개 통제는 연구 결과물에 식별될 수 있는 개인정보가 포함되지 않도록 고도로 훈련된 통계전문가가 철저하게 검증하는 과정이다.

마지막으로 민감한 데이터 사용에 대한 대중들의 신뢰를 높이기 위해, 대중에 대한 교육, 협의, 참여를 촉진하는 노력이 필요하다.

그러나 이는 하나의 모델일 뿐, 설립 과정에서 다양한 이해 관계자의 참여 속에 지역의 실정에 맞는 현실적인 방법을 찾아가야 한다.

제3장 해외 주요 국가의 데이터 연계·결합 현황

제1절 데이터 연계·결합 관련 해외 법제

국내 개인정보보호법은 개인정보의 “처리”를 “개인정보의 수집, 생성, 연계, 연동, 기록, 저장, 보유, 가공, 편집, 검색, 출력, 정정(訂正), 복구, 이용, 제공, 공개, 파기(破棄), 그 밖에 이와 유사한 행위를 말한다”고 정의하고 있다. (제2조 2호) 유럽 개인정보보호규정(GDPR)은 제4조(2)에서 처리(processing)를 “별개 또는 일련의(sets of) 개인정보의, 수집, 기록, 조직, 구성, 저장, 개조, 정정, 검색, 참조, 사용, 이전을 통한 제공, 배포나 정렬 또는 결합, 제한, 삭제, 파기와 그 밖에 가능한 모든 별개 또는 일련의(sets of) 작업(operation)을 의미한다. 이 경우 처리는 자동화(automated) 수단 또는 비자동화 수단에 의해 행해지는 작업 모두를 포함한다”고 정의하고 있다. 이러한 정의에 따르면, 개인정보를 포함한 데이터의 연계·결합은 개인정보의 처리에 해당한다. 또한, 서로 다른 목적으로 수집된 데이터를 연계하는 것은 개인정보의 목적 외 처리·제공에 해당한다.

데이터의 연계·결합이 정보주체의 동의를 전제로 이루어진다면 개인정보보호법 상 문제는 없을 것이다. 그러나 데이터 연계·결합이 기존에 다른 목적으로 수집된 데이터의 2차적 사용에 해당하는 경우가 많다는 점을 고려하면, 그리고 특정 목적으로 한정된 개인들만을 대상으로 한 매칭(matching)이 아니라 서로 다른 목적으로 수집된 대규모 데이터베이스의 연계·결합의 경우에는 수집 이후에 정보주체의 동의를 다시 획득하기는 쉽지 않을 것이다.

정보주체의 동의를 받지 않고 애초 수집목적 외로 처리·제공하기 위해서는 법적 근거가 있어야 한다. 개인정보의 수집목적 외 처리·제공은 개인정보보호원칙에 벗어난 예외적인 경우로서, 정보주체의 권리에 부정적인 영향을 미칠 수 있기 때문이다. 따라서 그러한 예외는 매우 제한적으로 허용되어야 하며, 동시에 정보주체의 권리에 미치는 부정적 영향을 최소화하기 위한 안전장치가 마련되어야 한다. 이러한 법적 근거는 개인정보보호법 내에 존재할 수도 있고, 특정 분야를 규율하는 다른 법률에 포함될 수도 있다.

제2절부터 구체적인 사례를 통해서 살펴보겠지만, 해외에서 데이터 연계·결합은 주로 공익을 위한 연구 및 통계 목적으로 이루어지고 있다. 특히 보건의료 증진을 위한 학술연구 분야에서 데이터 연계·결합에 대한 요구가 먼저 시작되었다.

데이터 연계·결합 관련 해외 법제는 유럽연합, 영국, 독일, 미국 등 주요 국가의 개인정보 보호 법제, 보건의료 관련 법제, 통계 관련 법제를 중심으로 검토하였으며, 기

타 참조할만한 국가의 사례도 포함하였다. 이들 법제에서 데이터 연계 결합과 관련된 조항을 중심으로 살펴보았다.

1. 유럽연합

(1) 유럽연합의 개인정보 보호 관련 법제

유럽연합은 일찍이 개인정보 보호에 관한 국제적인 협력을 이루어 왔다. 1948년의 UN세계 인권선언을 유럽 역내에서 구체화하여 1953년에 유럽인권조약을 체결했고, 1980년에는 OECD 프라이버시에 관한 국제적인 원칙을 확인했으며, 1981년 개인정보 보호에 관한 유일한 국제규범이었던 유럽연합조약 제108호를 채택하였다. 이런 노력으로 유럽연합 역내 각 국가의 개인정보 보호 법제가 갖추어져 나가기 시작하였으며, 유럽연합 역내의 개인정보 보호의 통일성을 기하기 위해 1995년에 마침내 개인정보보호지침을 제정하였다. 이 지침으로 유럽연합 각국의 개인정보보호법은 통일적인 수준을 갖추어 나가게 되었고, 유럽연합의 개인정보보호규범은 가장 포괄적인 국제적인 기준으로 자리매김하게 되었다.

유럽연합은 여기에서 그치지 않고 2012년부터 통일적인 법규 제정을 위한 노력을 하기 시작하였고, 마침내 2017년 일반정보보호규정(General Data Protection Regulation, GDPR)을 제정하였고, 2018년 5월 25일부터 시행되게 된다.

유럽연합의 개인정보 보호에 대한 최초의 국제적인 규범은 1948년의 UN세계 인권선언(Universal Declaration of Human Rights, UDHR) 제12조[11]를 들 수 있다. 뒤이어 1953년 유럽평의회[12]가 채택하여 시행하고 있는 유럽인권조약(The European Convention on Human Rights) 제8조[13][14]는 유럽인권조약에 의해 설립된 유럽인권재판소를 통해서 여러 판결로 정보보호에 관한 원칙을 확인해 주고 있고, 개인정보 보호가 핵심적인 인권으로 근원적 가치를 가지고 있다는 점을 일깨워 주는 역할을 하고 있다. 특히 유럽인권재판소는 공적 기관에 의한 통신의 도청[15], 여러 가지 유형의 감

11) 사생활 및 가족생활의 존중에 관한 조항으로 개인의 사적 영역을 국가의 침해로부터 보호받을 권리를 규정한 최초의 국제법규이다.

12) 2013년 유럽평의회의 회원국은 47개국이고, 그중 28개국은 유럽연합 회원국이다.

13) 체약 국가들은 ECHR을 준수할 국제적 의무를 지고 있다. 유럽평의회 모든 회원국은 국가법에 ECHR을 도입하였거나 실효성을 부여하였다. 체약 당사국들이 ECHR에 의한 의무를 준수할 것을 보장하기 위하여, 유럽인권재판소(ECtHR)가 설립되었으며, 여기에는 조약 위반을 주장하는 개인, 개인의 그룹, NGO 또는 법인들이 소송을 제기할 수 있다.

14) 제8조는 사생활 및 가족생활, 가정과 교신의 존중권을 보장하고 있으며, 이 권리의 제한이 허용되는 조건을 규정하고 있다.

15) ECtHR, Malone v. the United Kingdom, No. 8691/79, 2 August 1984; ECtHR, Copland v. the

시[16]와 개인정보의 저장에 대한 보호문제[17]들을 심리하여 많은 판결을 내려왔다. 이를 통해서 유럽인권재판소는, 유럽인권조약 제8조에 의해 국가들은 동 조약상의 권리를 침해하는 어떠한 행위도 금지하도록 하는 의무를 부담하고 있을 뿐만 아니라, 일정한 상황에서 효과적으로 사생활과 가족생활의 존중을 적극적으로 보장할 의무도 부담하고 있다는 점을 명확히 해왔다.[18]

1981년에 서명을 위해 개방된 유럽평의회 조약 제108호(Council of Europe Convention 108) '개인정보의 자동처리와 관련한 개인의 보호를 위한 조약'은 개인정보 보호 분야에서 법적 구속력을 가진 유일한 국제규범이었다.

조약 제108호는 사적 영역과 사법기관 및 법집행기관에 의한 정보 처리와 같은 공적 영역에 의해 수행된 모든 정보 처리에 적용된다. 동 조약은 개인정보의 수집 및 처리에 수반될 수 있는 남용에 대해 개인을 보호하며, 그와 동시에 국경을 넘는 개인정보의 유통을 규제하고자 하는 것이다.

이 조약은 개인정보의 수집 및 처리에 관한 원칙을 규정하고 있는데, 구체화된 정당한 목적을 위해 저장되고, 이들 목적과 양립 불가능한 목적을 위해 사용되지 않을 것을 명시하고 있다. 그리고 필요한 기간 이상으로 보관되지 않아야 한다. 개인정보의 질과 관련해서는 정확하고, 적정하고 관련성이 있으며 과도해서는 안 된다(비례성). 인종, 정치, 건강, 종교, 성생활 또는 범죄기록에 관한 정보와 같은 민감정보는 적절한 법적 안전장치가 없는 경우에는 처리할 수 없도록 규정하고 있다.

동 조약은 또한 개인이 자기에 관한 정보가 저장된다는 사실을 알며, 필요한 경우에, 그 정보를 정정하게 할 권리를 보장하고 있다. 동 조약에서 규정된 권리에 대한 제한은 국가안보 또는 국가방위와 같은 우월한 이익이 문제 되는 경우에만 가능하다.

동 조약은 조약 당사국 간의 개인정보의 자유로운 유통을 규정하고 있지만, 또한 법적 규제가 동등한 보호를 제공하고 있지 않은 국가에의 유통에 대한 제한도 규정하고 있고, 2001년에는 조약 제108호 추가의정서로 국가정보보호감독기관의 의무적 설립에 관한 규정을 도입하였다.

유럽연합은 2000년에 유럽연합기본권헌장(The Charter of Fundamental Rights of the European Union)을 제정, 공포하였는데, 사생활과 가족생활의 존중(제7조)을 보장

United Kingdom, No. 62617/00, 3 April 2007.

16) ECtHR, Klass and Others v. Germany, No. 5029/71, 6 September 1978; ECtHR, Uzun v. Germany, No. 35623/05, 2 September 2010.

17) 유럽인권재판소 Leander v. Sweden, No. 9248/81, 26 March 1987; 유럽인권재판소 S. and Marper v. the United Kingdom, Nos. 30562/04 and 30566/04, 4 December 2008.

18) 유럽인권재판소 I. v. Finland, No. 20511/03, 17 July 2008; 유럽인권재판소 K.U. v. Finland, No. 2872/02, 2 December 2008

할 뿐만 아니라 정보보호권(제8조)을 규정하여, 이러한 보호의 수준을 EU법상의 기본권의 수준으로 명시적으로 높이고 있다.

(2) 개인정보의 처리와 관련한 개인의 보호와 개인정보의 자유로운 이동에 관한 유럽의회 및 이사회의 지침 95/46/EC(정보보호지침)

가. 지침의 역할

규칙, 지침, 결정은 제2차 유럽연합법이라고도 불리는데[19], 조약들에 의해 권한을 부여받은 EU기관들에 의해 채택된다. 유럽연합은 1995년에 유럽연합 내에서 통일적이고 높은 수준의 개인정보 보호를 실현하기 위해서 '개인정보의 처리와 관련한 개인의 보호와 개인정보의 자유로운 이동에 관한 지침 95/46/EC(정보보호지침)'을 제정한다. 이 지침은 유럽연합 회원국이 지침에 부합하는 국내 이행입법을 제정할 의무를 부과한다. 현재 유럽연합 각국은 지침의 규정을 모두 국내법으로 이행입법을 제정할 의무를 다했다고 평가되었다.

나. 개인정보 처리에 관한 원칙

유럽연합 지침은 식별되거나, 식별될 수 있는 자연인에 대한 모든 정보를 '개인정보(personal data)'라고 정의한다. 식별 가능하다는 것은 직접적·간접적, 특히 신원증명번호 또는 신체적·생리적·정신적·경제적·문화적 또는 사회적 동일성에 관한 하나 또는 그 이상의 요인을 참조하여 그 신원을 알 수 있는 경우를 말한다.

지침은 자동화 수단의 여부와 관계없이 이전, 보급, 기타의 사용, 정렬·조합, 방지, 삭제 또는 파괴와 같은 방법에 의하여 수집, 기록, 조직, 저장, 채용, 수정, 복구, 참조, 사용, 공개와 같은 개인정보를 실행하는 작업 또는 작업의 집합(set of operations)을 '처리'라고 정의한다.[20]

따라서 개인을 식별하거나 식별할 수 있는 가능성이 있는 정보를 연계하는 경우에는 개인정보의 처리에 해당한다. 지침은 개인정보의 처리에 해당할 경우에는 다음과 같은 원칙을 준수해야 한다고 규정하고 있다.[21]

---

19) 유럽연합조약(TEU)과 유럽연합운영조약(TFEU)을 제1차 유럽연합법이라 부른다.

20) 정보보호지침 Article 2 (b)

21) 정보보호지침 Article 6

(a) 공정하고 적법하게 처리될 것

(b) 특정되고, 명백하고, 정당한 목적을 위해 수집되어야 하고, 당해 목적과 모순되는 방법에 의하여 재처리를 하지 않아야 한다. 다만, 역사적, 통계적 또는 과학적 목적을 위한 정보의 재처리는 회원국이 적절한 보호조건을 규정하는 한 모순되는 것으로 간주되지 아니한다.

(c) 개인정보 수집 및 재처리의 목적에 비추어 적절하고 관련 있고 지나치지 않을 것

(d) 정확할 것 및 필요하다면 최신 정보로 갱신할 것. 정보가 수집되었을 당시의 목적에 비추어, 부정확 또는 불완전한 정보가 재처리, 삭제, 교정되는 것을 보장하기 위한 모든 합리적인 조치가 취해져야 한다.

(e) 수집된 정보의 목적 또는 재처리를 위한 목적에 필요한 그 이상으로 정보주체의 신원 확인을 허용하지 않는 형식을 유지한다. 회원국은 역사적, 통계적 또는 과학적 이용을 위하여 더 오랜 기간 저장될 개인정보에 대해서는 적절한 보호조치(safeguards)를 취해야 한다.

다. 적법한 정보 처리의 요건

지침은 데이터의 연계가 개인정보의 처리에 해당할 경우에는 적법한 정보 처리가 되기 위해서는 아래의 요건 중 어느 하나의 요건을 충족해야 한다고 규정하고 있다.[22]

(a) 정보주체가 그의 동의를 명확하게 표시한 경우

동의는 자유롭게 제공되고 구체적이며 충분한 설명 후에 주어진(informed) 정보주체의 의사표시이어야 한다. 자유로운 동의란 정보주체가 실제로 선택을 할 수 있고, 동의하지 않는다 할지라도 기만, 협박, 강요 또는 상당히 부정적인 결과의 위험이 없어야 한다. 자유로운 동의는 또한 동의를 확보하는 관리자와 동의를 제공하는 정보주체 간에 중대한 경제적 또는 다른 불균형이 존재하는 종속적 상황에서 위협받을 수 있다. 정보주체는 의사결정을 하기 전에 충분한 설명을 들어야 한다. 또한, 동의가 유효하기 위해서는 특정된 것이어야 한다. 평균적인 정보주체의 합리적인 기대에 비추어 원래의 동의가 주어졌을 때 합리적으로 예상할 수 없었던 정도로 처리 작용이 추가되거나 변경되려고 한다면 정보주체는 다시 동의를 요청받아야 한다.

(b) 정보주체가 당사자인 계약의 이행에 필요한 처리 또는 계약을 체결하기 전에

22) 정보보호지침 Article 7

정보주체의 요구에 따른 조치를 취하기 위하여 필요한 처리

예컨대, 일방 당사자가 계약을 체결하고자 하지만 몇 가지 체크를 해야 할 사항이 남아 있기 때문에 그렇게 하지 못한 경우, 만일 일방 당사자가 이러한 목적을 위하여 정보를 처리할 필요가 있다면 그러한 처리는 그것이 "계약을 체결하기 전에 정보주체의 요청에 따라서 조치를 취하기 위한" 것인 한 정당하다.

(c) 관리자가 적용대상인 법적 의무를 준수하기 위하여 필요한 처리

예컨대, 의사들과 병원들은 수년 동안 환자들의 치료에 대한 정보를 저장할 법적 의무를 부담하고, 고용주들은 사회보험과 과세의 이유로 고용인들에 대한 정보를 처리하여야 하고, 사업가들은 과세 이유로 고객들에 관한 정보를 처리하여야 한다.

(d) 정보주체의 중대한 이익을 보호하기 위하여 필요한 처리

예컨대, 건강정보나 행방불명된 사람에 대한 정보의 적법한 이용의 근거로 될 수 있다.

(e) 공공의 이익을 위하여 또는, 관리자 또는 정보의 공개를 받는 제3자가 그에게 유보된 공적 권한을 행사함에 있어 수행할 직무의 이행을 위하여 필요한 처리

(f) 관리자 또는 정보의 공개를 받는 제3자에 의하여 추구되는 정당한 이익의 목적을 위하여 필요한 처리. 다만, 당해 이익보다 정보주체의 기본권과 자유를 목적으로 한 이익이 우선되는 경우에는 제외한다.

라. 민감정보의 처리에 관한 규율

지침은 민감정보 처리에 관한 특별한 규정을 두고 있는바, 데이터의 연계가 민감정보의 처리에 해당하는 경우에는 특별규정을 준수해야 한다. 민감정보란 민족 또는 인종적 기원, 정치적 성향, 종교 또는 철학적 신념, 노동조합 회원자격이 나타나는 개인정보, 건강 또는 성생활에 관련된 정보를 말한다. 이와 같은 민감정보를 처리할 수 있는 요건은 다음과 같다.[23)]

(a) 정보주체가 당해 정보의 처리에 명시적으로 동의한 경우. 다만, 각국 법에서 정보주체의 동의에도 불구하고 민감정보의 처리 금지를 규정한 경우는 동의가 적법한 요건이 될 수 없다. 지침은 민감정보인 경우는 정보주체의 동의가 명시적이어야 한다고 규정하여, 암묵적 동의를 배제하고 있다.

(b) 고용법 분야와 관련하여, 관리자의 의무와 특정권리를 수행할 목적에 필요한 처리가 적절한 보호조건을 규정한 국내법에 의하여 수권 된 경우

23) 정보보호지침 Article 8

(c) 정보주체가 신체적 또는 법적으로 동의할 수 없는 경우 정보주체 또는 다른 사람의 중대한 이익을 보호하기 위하여 필요한 처리의 경우

(d) 정치적, 철학적, 종교적 또는 노동조합의 목적을 수행하는 재단, 사단 또는 기타 비영리단체에게 적절하게 보장된 정당한 활동의 과정에서 수행되는 처리. 다만, 단체의 회원 또는 당해 단체의 목적과 관련한 정규적인 접촉을 가지는 사람에 관련된 처리에 한하고, 정보가 정보주체의 동의가 없이는 제3자에게 공개되지 않을 것을 조건으로 한다.

(e) 정보주체에 의하여 공공연히 공개된 정보에 대한 처리 또는 소송의 제기, 소송의 수행 또는 소송의 방어에 필요한 처리

한편, 지침은 각국의 법률로 민감정보 처리제한이 적용되지 않는 경우를 규정할 수 있다고 규정하고 있다. 즉, 정보의 처리가 예방의학, 의학적 진단, 의료보호 또는 치료의 제공 또는, 건강관리 서비스의 운영 등의 목적을 위하여 요구되는 정보의 처리의 경우, 그리고 국가의 관할 기관에 의해 수립된 국내법이나 규칙에 따라 전문적 비밀유지 의무를 준수해야 하는 보건 전문가, 혹은 동등한 비밀유지 의무를 가진 다른 자에 의해 정보가 처리되는 경우에는 민감정보 처리제한에 관한 규정이 적용되지 않는다.

아울러 지침은 회원국에 그 밖의 적용제외 규정을 추가로 제정할 권한을 부여하고 있는데, 그 요건은 (i) 중요한 공익을 위하여야 한다는 조건과 (ii) 국가법 또는 감독기관의 결정에 의해 규정되어야 하고, (iii) 국가법 또는 감독기관의 결정이 정보주체들의 이익을 실효적으로 보호하기 위하여 필요한 안전장치를 포함하는 것을 조건으로 한다.

범죄, 범죄의 평결 또는 보안 조치와 관련한 정보도 민감정보에 해당하여 그 처리는 몇 가지 요건을 충족해야 하는데, (i) 공공기관의 통제 하에서만 수행되어야하고, 혹은 (ii) 적절한 특정의 보호조건이 국내법에 규정된 경우, 그러한 국내법에 따라 예외가 인정될 수 있다.

결국, 개인을 식별하거나 식별할 가능성이 있는 데이터의 연계·결합에 대해서는 적법한 처리를 위한 목적으로 충족하는 경우에만 그 처리가 인정되는데, 처리하는 정보가 민감정보에 해당하는 경우에는 민감정보 처리에 관한 부가적인 요건을 충족하는 경우에만 처리할 수 있다. 이러한 경우에도 개인정보 처리에 관한 원칙을 준수해야 한다.

마. 개인정보주체의 권리

지침은 개인정보의 처리와 관련하여 여러 가지 권리를 보장하고 있다.

첫째, 정보주체에게는 정보(information)를 제공받을 권리가 보장된다. 이는 정보주체로부터의 요청에 의하는 것이 아니라, 정보주체가 정보에 관심을 표시하는지 여부에 관계없이 정보처리자가 주도적으로 준수하여야 한다.[24]

정보주체에게 제공해야 하는 정보에는 관리자와 그 대리인의 신원, 정보의 처리목적, 정보의 수령인 또는 그 범주, 질문에 대한 응답이 의무적인지 임의적인지, 응답을 거부한 경우에 가능한 결과, 정보주체와 관련한 정보의 열람권과 정정요구권의 존재 등이다. 특히 정보가 수집되는 특정한 상황, 정보주체에 관한 공정한 처리의 보장과 관련하여 더 많은 정보를 필요로 하는 경우에는 그 외의 추가적인 정보도 제공해야 한다.

당해 고지가 불가능하거나 고지를 위하여 불합리한 노력이 필요하거나 법에 의하여 명백하게 공개가 정하여진 경우, 특히 통계적 목적을 위한 또는 역사적 또는 과학적 연구를 목적으로 처리하는 경우에는 고지의무가 면제된다. 이 경우 적절한 보호조치를 제공해야 한다.

정보(data)가 정보주체로부터 수집되는 경우에, 정보(information)는 적어도 수집 시에 제공되어야 하고, 제3자로부터 수집되는 경우에는 적어도 관리자가 그 정보(data)를 기록하는 순간이나 그 정보가 최초로 제3자에게 공개되기 전에 제공되어야 한다.

둘째, 정보주체에게는 접근권이 보장된다.[25]

지침은 회원국들은 모든 정보주체에게 그들의 개인정보(personal data)와 정보(information)에의 접근권을 보장하여야 한다고 규정한다. 특히 모든 정보주체는 자기와 관련되는 정보가 처리되고 있는지 여부에 관한 확인과 적어도 아래의 정보를 관리자로부터 취득할 권리를 가진다. 즉, 처리의 목적, 관련 정보의 범주, 처리 중인 정보, 정보가 공개된 수취인 또는 수취인의 범주, 처리 중인 정보의 출처에 관한 이용 가능한 정보, 자동화된 결정의 경우 정보의 자동적 처리에 포함된 로직. 각 국가법은 예컨대, 정보 처리를 인정하는 법적 근거를 인용하는 등 관리자가 제공하는 정보를 부가할 수 있다.

셋째, 정보의 정정, 삭제 및 차단권이 보장된다.

정보주체는 자기정보의 처리가 특히 그 정보의 부정확성 또는 불완전성으로 인하여

---

24) 정보보호지침 Article 10

25) 정보보호지침 Article 12

지침 조항을 준수하지 않는다고 생각한다면, 관리자로부터 정보의 정정, 삭제 또는 차단을 얻을 국가법에 의한 권리를 가져야 한다.

바. 익명화 정보

정보보유제한의 원칙에 따르면, 정보는 “그 정보가 수집된 또는 그에 이어 처리되는 목적을 위해 필요한 기간보다 길지 않은 기간 동안 정보주체의 식별을 허용하는 형태로” 보유되어야 한다. 그러므로 정보는 그것이 오래되어 더 이상 원래의 목적에 사용되지 않게 된 후에도 관리자가 그 정보를 저장하기를 원한다면 익명화되어야 한다. 익명화란 모든 식별요소가 제거되어 정보주체가 더 이상 식별되지 않도록 하는 것인데, 이때 식별 가능성을 판단하기 위해 처리자 혹은 다른 누군가가 합리적으로 사용할 수 있는 모든 수단이 고려되어야 한다. 익명화된 정보는 더 이상 개인정보가 아니다.[26)]

사. 역사적, 통계적 또는 과학적 목적을 위한 정보 처리에 대한 특별규정

정보보호지침은 “역사적, 통계적 또는 과학적 목적을 위해 정보를 추가적으로 처리하는 것은 회원국들이 적절한 안전장치를 제공한다면 양립 불가능한 것으로 간주되지 않는다”고 명시적으로 선언한다.

장래의 과학적, 역사적 또는 통계적 이용을 위해 정보를 보유하는 것도 정보보호지침에서 정보보유 제한의 원칙을 명시적으로 제외하고 있다. 그러나 개인정보의 이러한 저장과 이용에는 국가법에 의한 특별한 안전장치가 수반되어야 한다.[27)]

지침은 통계적 목적 또는 역사적·과학적 연구 목적을 위한 처리의 경우에는 정보주체에 대한 정보제공의무를 면제하고 있다.

국가는 정보주체의 프라이버시를 침해할 위험성이 명백히 없다면, 데이터가 오로지 과학적 연구 목적으로 처리되거나 혹은 통계 생성만을 위해 필요한 기간 동안 개인적 형태로 보관될 경우, 입법으로 정보주체의 접근권을 제한할 수 있다. 다만, 적절한 법적 보호조치를 취해야 하며, 특히 그 데이터가 특정 개인에 관한 조치나 결정에 사용되어서는 안 된다.[28)]

---

26) 정보보호지침 Recital 26.

27) 정보보호지침 Article 6 (1) (e)

28) 정보보호지침 Article 13 (2)

(3) 일반정보보호규정(GDPR)

가. 개인정보의 처리와 데이터 연계

GDPR 제4조 '정의'(2)에서는 개인정보의 '처리(processing)'를 "별개 또는 일련의(sets of) 개인정보의, 수집, 기록, 조직, 구성, 저장, 개조, 정정, 검색, 참조, 사용, 이전을 통한 제공, 배포나 정렬 또는 결합(combination), 제한, 삭제, 파기와 그 밖에 가능한 모든 별개 또는 일련의(sets of) 작업(operation)을 의미한다. 이 경우 처리는 자동화(automated) 수단 또는 비자동화 수단에 의해 행해지는 작업 모두를 포함한다"고 정의하고 있어, 개인정보의 연계·결합이 처리에 해당함은 명확하다.

데이터 연계·결합을 포함한 개인정보의 처리는 GDPR 제5조 개인정보 처리의 원칙에 근거하여, 제6조에 따라 적법하게 이루어져야 한다. 제6조는 개인정보의 처리가 적법성을 인정받을 수 있는 경우를 다음과 같이 규정하고 있다.[29)]

(a) 정보주체가 하나 또는 그 이상의 특정 목적에 대해 본인의 개인정보 처리를 동의한 경우

(b) 정보주체가 계약 당사자가 되는 계약을 이행하는 경우, 또는 정보주체가 계약 체결 전에 조치를 요청하여 개인정보 처리가 필요한 경우

(c) 정보처리자의 법적 의무를 준수하는데 개인정보 처리가 필요한 경우

(d) 정보주체 또는 제3자의 생명에 관한 이익을 보호하기 위해 개인정보 처리가 필요한 경우

(e) 공익상 이유 또는 정보처리자의 공식권한을 행사하기 위한 업무수행에 개인정보 처리가 필요한 경우

(f) 정보처리자 또는 제3자의 정당한 이익을 달성하기 위하여 필요한 경우. 이 경우, 특히 정보주체가 아동일 때, 개인정보처리자의 정당한 이익이, 개인정보의 보호가 요구되는 정보주체의 이익 또는 기본권 및 자유보다 우선 되어서는 아니 된다.

제1항 (f)는 공공기관이 해당 기관 업무수행을 위하여 개인정보를 처리할 때는 적용되지 않는다.

보건의료 당국이나 통계청 등에서 데이터 연계·결합을 수행할 경우, 위 6조 1항의 (c), (e)가 적용될 수 있다. 6조는 (c)와(e)에서의 개인정보 처리는 유럽연합 법률 혹은 유럽연합 회원국의 법률에 근거하도록 하고 있다.

한편, 정보보호지침과 다르게, GDPR의 경우에는 개별 조항에서 뿐만 아니라, 제9장

29) GDPR Article 6 (1)

특정 정보 처리 상황에 관한 규정, 제89조에서 통계 및 연구 목적의 개인정보 처리와 관련된 사항을 다루고 있다. 그 구체적인 내용은 아래에서 서술한다.

나. 개인정보 처리에 관한 원칙

GDPR 제5조는 개인정보의 처리 원칙으로 적법성·공정성·투명성, 목적 제한, 데이터 최소화, 정확성, 보관 기간 제한, 무결성과 기밀성, 책임성 등을 규정하고 있다. 원칙적으로 개인정보는 “명시적이고 적법한 특정 목적을 위해 수집되어야 하고, 해당 목적과 양립하지 않는 방식으로 추가 처리 되어서는 안 된다.” 이에 따르면, 서로 다른 목적으로 수집된 개인정보를 연계해서는 안 될 것이다. 다만, “공익적인 기록 보존, 과학 및 역사 연구 또는 통계 목적을 위하여 개인정보를 추가 처리한 때는 제89조 1항에 따라 원래의 목적과 양립된다고 본다.” 즉, 공익적인 기록 보존, 과학 및 역사 연구 또는 통계 목적의 개인정보 연계·결합은 예외로 인정하고 있다. 그러나 이 경우에도 제89조 1항에 따라 정보주체의 권리와 자유를 위해 적절한 안전조치를 취할 필요가 있다.

또한, 필요 이상으로 개인정보를 보관하지 않는 것이 원칙이지만, 공익적인 기록 보존, 과학 및 역사 연구 또는 통계 목적을 위해 개인정보를 처리하는 경우에는 보유기간이 연장될 수 있다.

개인정보보호원칙은 식별되었거나 식별될 수 있는 개인에 관한 일체의 정보에 적용될 수 있으며, GDPR은 가명처리 정보는 추가 정보를 이용하여 개인을 식별할 수 있는 정보로서 식별할 수 있는 개인정보로 간주되어야 한다고 한다. 익명 정보에는 개인정보보호원칙이 적용되지 않는다. 따라서 이 법은 통계 목적 및 연구 목적 등을 위한 익명 정보의 처리에는 적용되지 않는다.[30]

다. 민감정보의 처리

연계·결합되는 데이터에 민감정보가 포함될 수 있는데, 이때 GDPR 제9조 민감정보의 처리에 관한 규정이 적용된다. 원칙적으로 민감정보, 즉 인종·민족, 정치적 견해, 종교·철학적 신념, 노동조합의 가입 여부를 나타내는 개인정보, 유전자 정보, 자연인을 고유하게 식별할 수 있는 생체정보, 건강정보, 성생활·성적 취향에 관한 정보의 처리는 금지된다. 다만, 몇 가지 경우에는 민감정보를 처리할 수 있다. 예를 들어, 정보주체가 ‘명백한 동의’를 제공하는 경우[31], 일정한 조건 하에서 의료 및 의학적인 목적

30) GDPR recital 26

31) GDPR Article 9 (2) (a)

으로[32] 처리하는 경우 등이 그렇다.

또한, 제89조 1항에 따라 공익적인 기록 보존, 과학 및 역사 연구 또는 통계 목적을 위해 필요한 경우에도 민감정보를 처리할 수 있는데, 이 경우 유럽연합·회원국 법률에 근거해야 하는데, 이 법률은 추구하는 목적에 비례하고 개인정보보호권의 본질을 존중하며 정보주체의 기본권 및 이익을 보호하기 위해 적절하고 구체적인 조치를 제공하는 것이어야 한다.[33]

라. 정보주체로부터 개인정보가 수집되지 않은 경우

데이터 연계·결합은 행정데이터 등 이미 수집된 데이터를 이용하는 경우가 많다. GDPR 제14조는 '정보주체로부터 개인정보가 수집되지 않은 경우 제공되는 정보'에 대해서 규정하고 있다. 즉, 개인정보가 정보주체로부터 수집되지 않은 경우, 정보처리자는 다음과 같은 정보를 정보주체에게 제공해야 한다.

(a) 정보처리자 또는, 가능한 경우, 정보처리자 대리인의 신원 및 상세 연락처;

(b) 해당되는 경우, 개인정보 보호 담당관의 상세 연락처;

(c) 해당 개인정보의 예정된 처리의 목적뿐 아니라 처리의 법적 근거;

(d) 관련 개인정보의 범주;

(e) 해당되는 경우, 개인정보의 수령인 또는 수령인의 범주;

(f) 해당되는 경우, 정보처리자가 제3국이나 국제기구의 수령인에게 개인정보를 이전할 예정이라는 사실과 집행위원회의 적합성 결정의 여부, 또는 제46조나 제47조, 또는 제49조의 1항의 두 번째 단락에 규정된 이전의 경우나 해당 이전이 공개되는 경우, 적절하고 적합한 보호수단과 이에 대한 사본을 입수하기 위한 수단;

그러나 통계 및 연구 목적의 데이터 연계 시 현실적으로 정보주체에게 이러한 정보를 제공하기 쉽지 않을 것이다. 제14조 5항은 정보주체에게 정보 제공을 하지 않아도 되는 몇 가지 예외를 규정하고 있다. 공익상의 기록보관 목적이나 과학 및 역사 연구 목적 또는 통계 목적의 처리를 할 때, 제89조 1항에 규정된 조건 및 안전조치를 전제로, "해당 정보의 제공이 불가능하거나 과도한 노력이 수반되어야 하는 경우, 또는 본 조문의 제1항에 규정된 의무(즉, 정보 제공의 의무)가 불가능하다고 생각되거나 관련 처리의 목적 달성을 심각하게 저해하는 경우"가 이에 포함된다. 다만, 이 경우 정보처리자는 해당 정보의 공개 등, 정보주체의 권리와 자유 그리고 정당한 이익을 보호하

32) GDPR Article 9 (2) (h)(i)

33) GDPR Article 9 (2) (j)

기 위해 적절한 조치를 취해야 한다.[34] 이와 관련해 정보주체의 인원수, 해당 개인정보의 생성 시점 및 채택된 모든 적절한 보호수단이 고려될 수 있다.[35]

마. 삭제권("잊힐 권리")

GDPR 제17조는 정보주체의 삭제권을 규정하고 있다. 정보주체는 본인에 관한 개인정보의 삭제를 정보처리자에게 요청할 권리를 가지며, 정보처리자는 일정한 경우 부당한 지체 없이 개인정보를 삭제할 의무를 갖는다. 그러나 정보주체의 삭제권이 제한되는 경우가 있다. 예를 들어, 공중보건 분야의 공익상 이유[36]로, 또는 89조 1항에 따라 공익상의 기록 보존 목적이나 과학·역사 연구 목적 또는 통계 목적으로, 제1항에 규정된 권리가 불가능하다고 생각되거나 해당 처리의 목적 달성을 심각하게 저해할 가능성이 있는 경우[37] 등이다.

바. 역사적, 통계적 또는 과학적 목적을 위한 정보 처리에 대한 특별규정

GDPR 제9장은 개인정보 처리와 표현의 자유와의 관계, 공식 문서 공개와의 관계, 국가 식별 번호의 처리 등 특정 정보 처리 상황에 관한 규정을 두고 있다. 앞서 개별 조항에서 통계 및 연구 목적의 개인정보 처리 예외를 두고 있는 것과 함께, 제89조에서는 이 문제 자체를 다루고 있다.

이에 따르면, 공익을 위한 유지보존의 목적, 과학이나 역사적 연구의 목적 또는 통계 목적의 개인정보 처리의 경우 정보주체의 권리와 자유를 위해 적절한 안전조치를 취해야 한다. 이러한 안전조치는 데이터 최소화 원칙을 보장하기 위한 기술적·조직적 조치의 구비를 보장하는 것이어야 한다. 가명처리가 그러한 목적에 부합한다면, 기술적·조직적 조치에 포함될 수 있다. 정보주체의 식별을 할 수 없거나 더 이상 허용하지 않는 방식의 추가 처리를 통해 이러한 목적이 달성될 수 있다면, 그러한 방식을 채택해야 한다.

제89조의 2항과 3항은 개인정보가 과학이나 역사적 연구 목적, 또는 통계 목적으로 처리되는 경우(2항) 및 공익을 위한 유지보존 목적으로 처리되는 경우(3항), 일부 권리로 인해 특정 목적의 달성이 불가능하거나 심각하게 저해될 가능성이 있고 적용의 일부 제외가 그 같은 목적의 충족에 요구되는 한, 유럽연합 또는 회원국의 법률로 해

34) GDPR Article 14 (5) (b)
35) GDPR recital 62
36) GDPR Article 17 (3) (c)
37) GDPR Article 17 (3) (d)

당 권리의 적용을 일부 제외할 수 있도록 하고 있다. 즉, 연구 및 통계 목적으로 개인정보를 처리할 때(2항)에는 정보주체의 열람권(15조), 수정권(16조), 처리에 대한 제한권(18조), 개인정보 처리에 반대할 권리(21조) 및 89조 1항의 조건 및 안전조치에 따른 권리의 적용을 일부 제한할 수 있으며, 공익을 위한 유지보존 목적으로 처리할 때(3항)에는 이에 더하여 개인정보의 수정이나 삭제 또는 처리의 제한에 관한 고지 의무(19조), 본인의 개인정보 이전권(20조)도 제한할 수 있다.

GDPR은 해설(recital) 부분에서 통계 목적의 개인정보 처리에 대해 상세히 설명하고 있다. 유럽연합 또는 회원국의 법률은 GDPR의 한도 내에서 통계 내용, 접근(access) 통제, 통계 목적의 개인정보 처리에 대한 세부사항 및 정보주체의 권리와 자유를 보호하고 통계의 신뢰성을 보장하기 위한 적절한 조치를 결정해야 한다. 통계 목적은 통계 조사나 통계 결과를 작성하는 데 필요한 개인정보의 수집 및 처리의 작업 일체를 의미한다. 그 통계 결과는 과학적 연구 목적 등 다른 목적을 위해 추가적으로 활용될 수 있다. 통계 목적에는 통계 목적으로의 정보 처리 결과가 개인정보가 아닌 총계 데이터(aggregate data)이며 이 결과나 개인정보가 다른 특정 개인에 관한 조치나 결정을 지지하는 데 활용되지 않는다는 점이 내포되어 있다.[38]

(4) 유럽연합의 통계 관련 법제

GDPR은 해설 부분에서 유럽연합과 회원국 통계청이 공식적 통계를 작성하기 위해 수집하는 기밀 정보는 보호되어야 한다고 언급하고 있다. 유럽연합의 통계는 유럽연합 기능에 관한 조약(TFEU) 제338조(2)에 규정된 통계 원칙에 부합하여 개발, 작성 및 유포되어야 하고 회원국 통계 또한 회원국 법률을 준수하여야 한다. 유럽의회 및 각료이사회 규정(EC) No 223/2009는 유럽연합 통계에 있어 통계의 신뢰성에 대한 추가 세부사항을 규정하고 있다.[39]

2009년 3월 제정된 '유럽의회 및 각료이사회 규정(EC) No 223/2009 (EU통계규정)'은 유럽연합 통계에 관한 일반법의 지위를 가지며, 전 6장 29개 조로 구성되어 있다. 이 중 개인정보 보호와 관련이 있는 부분은 제5장(통계적 기밀성)이다.[40]

가. 개인정보와 기밀 정보

EU통계규정의 본문 규정에서 '개인정보'를 직접 언급하고 있지는 않다. 그러나 해

38) GDPR recital 162

39) GDPR recital 163

40) 이에 대한 자세한 내용은 전남대산학협력단(2016) 참조.

설 부분에서 이 규정은 EU 기본권 헌장 7조 및 8조에서 규정하고 있는 '사생활 및 가족생활의 존중 권리' 및 '개인정보 보호 권리'를 보장한다고 명시하고 있고,[41] 또한, 유럽통계와 관련한 개인정보의 처리와 관련하여 개인의 보호를 보장한다고 하며, 95년 정보보호지침 등을 언급하고 있다.[42] 본문 규정에서는 '개인정보' 대신 '기밀 정보(confidential data)'라는 개념을 사용하고 있는데, 이는 "통계단위(statistical Units)가 직접 혹은 간접적으로 식별됨으로써 개별 정보(individual information)를 드러낼 수 있는 데이터"를 의미한다. 통계단위가 식별될 수 있는지 여부를 결정하기 위해, 제3자가 통계단위 식별을 위해 합리적으로 사용할 수 있는 모든 관련 수단을 고려해야 한다.[43] 업체의 정보와 같이 기밀 정보가 모두 개인정보는 아니겠지만, 개인정보는 기밀 정보에 해당한다고 볼 수 있다.

EU통계규정 제5장은 이러한 기밀 정보의 보호와 관련된 내용을 다루고 있는데, 통계 목적으로 데이터를 연계할 경우, 이 장의 규제가 적용된다.

나. 기밀 정보의 보호

제20조는 기밀 정보의 보호를 규정하고 있는데, 기밀 정보가 오로지 통계 목적으로만 사용되고 불법적인 공개를 방지하기 위해 다음과 같은 규칙과 조치가 적용되어야 한다. 우선 유럽통계 생산을 위해 독점적으로 획득한 기밀 정보는, 통계단위가 다른 목적의 사용에 명확하게(unmabiguously) 동의하지 않았다면, 국가통계기구(NSI), 다른 국가 당국, 위원회(EuroStat)(이하 NSI 등)에 의해서 오로지 통계 목적으로만 사용되어야 한다.

통계단위를 식별할 수 있는 통계 결과의 보급은 NSI 등에 의해 다음과 같은 예외적인 경우에만 이루어질 수 있다. 첫째, 특정한 조건과 양식이 유럽 의회 및 이사회의 법령에 의해서 결정되고, 통계단위가 요청했을 때 통계적 기밀성을 훼손하지 않는 방식으로 통계 결과가 수정될 수 있는 경우. 둘째, 통계단위가 데이터의 공개에 명확하게 동의를 한 경우.

NSI 등은 기밀 정보의 물리적, 논리적 보호를 위해 필요한 모든 규제적, 행정적, 기술적, 조직적 조치를 취해야 한다. (통계적 노출 제어) 또한, NSI 등은 기밀 정보의 물리적, 논리적 보호와 관련된 원칙과 가이드라인의 조화를 위해 필요한 모든 조치를 취해야 한다. 이러한 조치는 27조 2항에서 언급한 규제 절차에 따라 위원회(Eurostat)가 채택한다. NSI 등의 간부 및 직원은 직무종료 이후에도 기밀 유지의무를 준수해야

41) EU통계규정 recital 21
42) EU통계규정 recital 22
43) EU통계규정 Article 3 (7)

한다.

다. 기밀 정보의 이전(transfer)

21조는 기밀 정보의 이전을 규정한다. 유럽통계시스템(ESS) 당국(앞서 언급한 NSI, 기타 국가 당국, Eurostat 등) 사이에서 유럽통계의 개발, 생산, 보급을 위해 기밀 정보의 이전이 발생할 수 있다. 첫 이전 이후의 추가적인 이전을 위해서는 해당 정보의 수집 기관에 의한 명확한 승인이 필요하다. 유럽의회와 이사회의 법률이 그러한 이전을 규정한 경우, 통계적 기밀성에 대한 국가 규칙이 기밀 정보의 이전을 막아서는 안 된다. 기밀 정보의 이전은 오로지 통계 목적을 위해서만 이루어져야 하며, 특정 업무영역 내에서 통계 활동에 종사하는 직원만이 접근할 수 있다. 통계적 기밀성에 관한 조항은 ESS 당국 사이 및 ESS 당국과 ESCB(유럽중앙은행시스템) 사이의 모든 기밀 정보 이전에 적용된다.

라. 위원회(Eurostat) 내의 기밀 정보 보호

기밀 정보는 특정 업무영역 내의 위원회 간부만이 접근할 수 있다. (22조 1항) 다만, 예외적인 경우로, 위원회는 특정 업무영역 내의 계약에 따라, 다른 직원 혹은 위원회를 위해 일하는 다른 사람에게 접근을 허락할 수 있다. 기밀 정보에 접근하는 사람들은 오로지 통계 목적으로만 사용해야 하며, 직무종료 후에도 이러한 제한이 적용된다.

마. 학술적 목적의 기밀 정보 접근

23조는 과학적 목적의 기밀 정보 접근을 규정한다. 통계단위를 간접 식별할 수 있는 기밀 정보에 대한 접근은 각각의 관할 영역에서 NSI 등에 의한 학술적(scientific) 목적을 위해 통계적 분석을 수행하는 연구자에게 허용될 수 있다. 해당 정보가 위원회에 이전된 경우, 그 정보를 제공한 NSI 혹은 다른 국가 당국의 승인이 필요하다. EU 차원에서의 접근을 위한 양식, 규칙, 조건 등은 위원회에 의해 수립되어야 한다. 이 규정을 보완함으로써, 핵심적이지 않은 요소들을 개정하기 위한 조치들은 27(3)조에 따른 규제 절차에 따라 채택되어야 한다.

바. 행정 기록에 대한 접근

24조는 유럽통계에 필요한 데이터를 얻기 위해 행정데이터 소스에 접근할 필요성을

규정하고 있다. 행정데이터의 이용은 응답자의 부담을 줄여준다. 효과적인 접근을 위한 실질적인 체제 및 조건은 필요할 경우 각 회원국과 위원회에 의해서 각자의 관할 범위 내에서 결정되어야 한다.

사. EU통계규정 5장의 기타 조항

25조는 일반인이 합법적으로 접근할 수 있는 소스로부터 얻은 정보는 기밀 정보로 간주되지 않는다는 것, 26조는 통계적 기밀성 침해를 방지하고 처벌할 수 있는 적절한 조치를 취해야 함을 규정하고 있다.

아. 학술적 목적의 기밀 정보에의 접근 규정

앞서 언급한 EU통계규정 23조(학술적 목적의 기밀 정보 접근)의 집행을 위해 유럽위원회는 위원회규정 557/2013(COMMISSION REGULATION (EU) No 557/2013 of 17 June 2013)을 두고 있다. 이 규정은 유럽통계를 위해 수집된 기밀 정보에 대해 학술 연구자들의 접근을 증진하여 수집된 데이터의 이익을 극대화하기 위한 목적으로 만들어졌으며[44], 총 11개의 조문으로 구성되어 있다.

제2조는 여러 개념 정의를 다루고 있다. '학술적 목적의 기밀 정보(confidential data for scientific purposes)'란 통계단위의 간접식별이 가능한 정보를 의미하며, 보안사용파일(secure-use files) 혹은 학술사용파일(scientific-use files)의 형태를 가지고 있다. 학술사용파일은 통계단위의 식별 위험성을 줄이기 위해 현재의 모범 관행에 따라 적절한 수준의 '통계적 노출 제어(statistical disclosure control)' 방법이 적용된 기밀 정보를 의미하며, 보안사용파일은 이러한 방법이 적용되지 않은 정보를 의미한다. 즉, 보안사용파일은 통계단위가 식별될 가능성이 높은 정보이다.

제3조는 일반 원칙을 담고 있는데, 위원회(eurostat)가 학술 목적으로 기밀 정보를 제공할 경우, 다음과 같은 조건을 만족해야 한다.

(a) 접근이 승인된 연구기관에 의해 이루어질 것

(b) 적절한 연구 제안서가 제출될 것

(c) 학술 목적으로 요청된 기밀 정보의 유형이 적시될 것

(d) 위원회(Eurostat) 혹은 위원회가 인가한 다른 접근 시설에서 접근이 제공될 것

(e) 해당 정보를 제공한 관련 국가 통계 당국이 승인할 것

---

44) 위원회규정 557/2013 recital 2

연구기관이 승인을 받기 위해서는 다음과 같은 기준을 만족해야 한다. (4조)

(a) 기관의 목적 : 법령, 임무 혹은 기타 목적의 선언에 근거하여 기관의 목적에 대한 평가가 수행되어야 한다. 기관의 목적에 연구가 포함되어야 한다.

(b) 양질의 연구를 생산하고 공개적으로 이용 가능 하도록 하는 기관으로서 확립된 기록이나 명성 : 출판물이나 연구 프로젝트의 목록에 기반하여 연구 프로젝트를 수행하는 기관의 경험이 평가되어야 한다.

(c) 연구를 위한 내부 조직체제 : 연구기관은 법인과 분리된, 연구 혹은 조직 내 연구부서에 초점을 맞춘 기관이어야 한다. 연구기관은 학술적 결론을 내는 데 있어 독립적, 자율적이어야 하며, 자신이 속한 기구의 정책부서와 분리되어야 한다.

(d) 정보보안을 보장할 안전조치의 구비 : 연구기관은 정보보안을 보장할 기술 및 인프라 요구조건을 충족해야 한다.

또한, 기밀 정보에 접근하는 모든 연구자를 포함하는, 그리고 접근 조건, 연구자의 의무, 기밀성 보호조치, 위반 시 제재 등을 명시한 기밀성 동의서가 연구기관의 지정된 대표자에 의해 서명되어야 한다. 연구기관에 대한 평가 보고서는 국가통계 당국에 제공되어야 하며, 위원회(Eurostat)는 승인된 연구기관의 목록을 웹사이트를 통해 공개하고, 정기적으로 재평가해야 한다.

제5조는 연구 제안서의 세부 요건을 다루고 있으며, 연구자들이 서명한 개별적인 기밀성 선언이 첨부되어야 함을 요구하고 있다. 각 연구 제안서는 해당 기밀 정보를 이전한 국가통계 당국의 승인을 얻어야 한다. (제6조)

제7조는 앞서 언급한 보안사용파일 및 학술사용파일의 접근 조건을 정하고 있다. 보안사용파일이 허가되기 위해서는 연구 결과가 기밀 정보를 드러내지 않을 것을 보장하는 사전 점검을 받아야 하며, 위원회(Eurostat)의 접근 시설 내에서 혹은 위원회가 인가한 다른 접근 시설 내에서만 보안사용파일에 접근할 수 있다. 반면, 학술사용파일의 경우에는 연구기관 내에 적절한 안전조치가 구비되어 있다면 제공이 허용된다. 위원회는 필요한 안전조치에 대한 정보를 공개해야 한다. 위원회는 국가통계 당국과 협력하여, 서로 다른 유형의 학술 목적 기밀 정보를 대상으로 하는 연구용 데이터 셋을 준비해야 하며, 이때 기밀 정보의 불법적인 공개의 위험성과 영향을 고려해야 한다.

제8조는 접근 시설에 대한 세부사항을 규정하고 있다. 접근 시설은 국가통계 당국 내에 위치해야 한다. 다만, 국가통계 당국의 명시적인 사전 승인에 따라 외부에 위치할 수도 있다. 접근 시설에 대한 인가는 접근 시설의 목적, 조직 구조, 데이터 보안 및 데이터 관리 표준 등의 기준에 기반하여 이루어져야 한다. 위원회(Eurostat)가

ESS 위원회와 협력하여 접근 시설 평가를 위한 가이드라인을 수립해야 하고, 평가 보고서는 국가통계 당국에 제공되어야 한다. 이 보고서에는 해당 접근 시설에서 제공되는 기밀 정보의 유형에 대한 권고가 포함되어야 한다. 위원회(Eurostat)는 접근 시설 인가의 결정 전에 ESS 위원회와 협의해야 한다. 접근 시설 혹은 이를 호스팅하는 기관의 지정된 대표자와 위원회(Eurostat) 사이에 기밀 정보의 보호 및 조직적 조치에 관한 접근 시설의 의무를 결정하는 계약이 체결되어야 한다. 위원회는 인가된 접근 시설의 목록을 웹사이트에 공개해야 한다.

2. 영국

(1) 개인정보 보호 관련 법제

영국의 개인정보 처리에 관한 기본적인 법률은 데이터 보호법(Data Protection Act 1998, DPA)이다. 데이터 보호법은 유럽연합의 개인정보 보호에 관한 지침인 정보보호지침(Directive 95/46/EC)의 국내 이행입법이다.

DPA는 지침의 내용과 대동소이한데, 개인정보 처리에 관한 8가지 원칙을 규정하고 있다.[45]

1. 개인 데이터는 공정하고 적법하게 처리되어야 하며, 아래의 요건을 충족하지 못하는 경우에는 처리되어서는 안 된다.

(a) 최소한 아래의 적법한 처리 요건의 하나 이상을 충족해야 한다.

(i) 정보주체의 동의, (ii) 정보주체가 당사자인 계약의 이행, 계약을 체결하기 위해 데이터 주체가 요청하는 조치를 취하기 위하여 필요한 경우, (iii) 계약에 의해 부과된 의무 이외에 정보처리자가 준수해야 하는 법적 의무를 준수해야 할 필요가 있는 경우, (iv) 정보주체의 중대한 이익을 보호하기 위하여 필요한 경우, (v) 사법 행정, 의회의 기능 수행, 법령에 따라 어떤 사람에게 주어지는 임무의 수행, 국왕이나, 각료(Minister of Crown; 또는 왕실장관), 행정 부서의 기능을 수행하기 위하여, 개인이 공익을 위해 공적 성격을 지닌 기타 임무의 수행을 위해 필요한 경우, (vi) 정보처리자나 제3자, 또는 정보를 공개할 당사자에 의해 추구되는 법적 이익을 위해 필요한 경우. 단, 정보주체의 권리와 자유 또는 법적 이익을 해치지 않고는 정보 처리를 할 수 없는 경우는 제외. (vi) 테러방지법, 형사절차법의 이행을 위해 정보공개가 필요한 경우.

---

45) DPA, SCHEDULE 1 The data protection principles

특히 민감정보는 다음과 같은 요건 하에서만 처리가 허용된다.

(i) 정보주체의 명백한 동의, (ii) 처리가 고용과 관련하여 정보처리자에게 주어지거나 부과된 권리나 의무를 행사하기 위해 필요한 경우, (iii) 정보주체나 그 대리인이 동의할 수 없는 경우나 정보처리자가 분명하게 정보주체의 동의를 얻을 수 없을 것이라고 보이는 경우로서 정보주체나 타인의 중대한 이익을 보호하기 위해 필요한 경우, (iv) 정보주체나 그의 대리인이 타당한 이유 없이 동의를 거부한 경우 타인의 중대한 이익을 보호하기 위해 필요한 경우, (v) 비영리공익단체의 업무 수행상 필요하며 정보주체의 권리보호를 위한 적절한 조치가 취해지는 경우, (vi) 정보주체의 의도적 행위에 의해 관련 정보가 이미 공개된 경우, (vii) 소송과 관련하여, (iix) 사법행정상 필요, 법률상 특정인에게 부여된 기능의 수행상 필요, 국가 및 정부의 기능 수행상 필요한 경우, (ix) 의료진이나 그에 상당하는 묵비의무를 갖는 사람이 의료목적상 필요한 경우, (x) 인종 내지 민족 간의 기회 및 대우의 평등을 도모하기 위해 관련 자료가 필요한 경우, (xi) 기타 장관령에 의해 규정된 상황에서 개인정보가 처리되는 경우

이때, “의료 목적”은 예방의학, 의학적 진단, 의학 연구, 돌봄 및 치료 제공, 보건의료 서비스의 관리를 포함한다.[46]

2. 개인 데이터는 하나 이상의 명시된 적법한 목적으로만 취득되어야 하며, 그와 같은 목적과 부합되지 않는 방식으로 처리되어서는 안 된다.

3. 개인 데이터는 처리되어야 하는 목적에 적합해야 하며, 그 목적과 관련이 있어야 하며, 이를 넘어서서는 안 된다.

4. 개인 데이터는 정확하고 가능한 한 최신의 상태로 보존되어야 한다.

5. 개인 데이터는 그 목적상 필요한 기간 이후에 보존되어서는 안 된다.

6. 개인 데이터는 본 법령에 따른 정보주체의 권리와 일치하게 처리되어야 한다.

7. 개인 데이터의 무자격 또는 불법적 처리와 사고로 인한 손실, 파괴나 훼손에 대해서는 적절하고 조직적인 조치가 취해져야 한다.

8. 유럽 경제 구역 밖의 나라나 지역에서 개인 데이터의 처리에 정보주체의 권리와 자유를 적절한 수준으로 보장하지 않는다면 개인 데이터는 그 국가나 지역으로 이동시킬 수 없다.

정보보호법은 연구, 역사 및 통계 목적의 정보 처리에 대해서는 일정한 요건 하에서 개인정보 보호의 원칙 중 일부를 배제하고 있다.[47]

---

46) DPA, SCHEDULE 3, 8(2)

47) DPA, 33 Research, history and statistics.

적용이 배제되는 원칙은 목적 제한의 원칙, 보유 기간 제한의 원칙, 개인정보주체의 접근권이다. 관련 조건을 준수하여 연구, 통계 목적으로만 처리하는 경우에는 최초 수집된 목적과 양립할 수 있는 목적의 처리로 인정된다. 보유 기간 제한의 원칙이 적용되려면 관련 조건을 준수하면서 연구 목적으로만 처리되는 경우이어야 한다. 접근권이 배제되려면 개인정보가 오로지 연구 목적으로만 처리되고, 관련 조건에서 처리되고, 연구 결과나 통계 결과가 정보 주체들이나 정보주체의 누구라도 식별할 수 있도록 만들어지지 않아야 한다. 여기서 '관련 조건(the relevant conditions)'은 연구 목적의 정보처리가 특정 개인에 대한 조치나 판단을 돕기 위한 처리가 아니어야 하고, 어떤 개인에게도 중대한 손해나 정신적인 침해를 가져올 수 있는 방법으로 처리되어서는 안 된다는 것을 의미한다.

(2) 인권법 (Human Rights Act 1998)

영국 인권법은 2000년 10월 2일 법제화되었으며, 유럽인권조약(European Convention on Human Rights)을 영국 법제에 반영한 것이다. 인권법은 공공기관에만 적용되는데, 이에 따라 NHS와 보건의료 기관에도 적용된다. 인권법의 SCHEDULE1 8조는 가족 및 사생활을 존중받을 권리(Right to respect for private and family life)를 규정하고 있다. 이에 따라 자신의 의료 정보에 대한 환자들의 권리 역시 도출된다. 또한, 8조는 보통법상 기밀 의무도 반영하고 있는 것으로 해석되는데, 이에 따라 환자들은 자신의 건강 기록의 기밀성을 유지할 권리를 가진다. 보건의료 정보를 수집, 보유하는 영국의 정보센터들은 이에 따라 환자들의 정보를 동의 없이 공개해서는 안 되며, 환자 정보는 안전하게 보관되어야 한다. 8조는 법에 따라, 그리고 건강 혹은 도덕(moral) 보호 등 민주 사회의 필요에 따른 목적을 제외하고는 공공기관이 이러한 권리를 침해할 수 없다고 하고 있다.

(3) 보건의료 관련 법제

가. 보건사회복지법(Health and Social Care Act 2012)

보건사회복지법 2012는 2012년 3월 제정되어 2013년 4월 1일 발효되었다. 이 법은 영국의 국립보건서비스(NHS)의 구조를 근본적으로 바꾸는 내용을 포함하고 있어 이를 둘러싸고 큰 논란이 벌어졌다. 영국 내에서 보건의료 정보에 대한 국가적인 정보서비스를 제공하는 '보건 및 사회복지 정보센터(Health and Social Care Information Centre, HSCIC)도 기존 기구를 통폐합하여 이때 재설립되게 되었다. (HSCIC는 2016

년 7월 이름을 NHS Digital로 변경하였다.) HSCIC는 NHS 잉글랜드의 지시를 받으며, 보건 및 사회복지 기구들로 하여금 HSCIC에 정보를 제공하도록 한다. 또한, 개인 식별이 안되는 통계정보를 발행하며, 일정한 조건하에 환자식별이 가능한 정보를 포함한 정보를 제공할 수 있다.[48]

나. 국가보건서비스법(NHS Act 2006)

2차적 목적으로 사용되는 기밀 정보(즉, 개인 식별이 가능한 정보)를 합법적으로 처리하기 위해서는 해당 조직이 정보주체(즉, 환자)로부터 설명 후 동의를 획득하거나, 혹은 동의를 받지 않아도 되는 법률적 기반이 있어야 한다.

법적 예외는 일반적으로 NHS법 2006의 Section 251을 통해서 이루어진다. 보건및사회복지법(Health and Social Care Act) 2001의 Section 60이 재입법 된 NHS법 2006의 Section 251은 보건부 장관(Secretary of State for Health)이 특정한 의료적 목적으로 보통법상 기밀유지의무의 예외를 둘 수 있는 규정이다. 이 규정을 '보건서비스(환자정보통제)규정 2002'[49]라고 부른다. 'section 251 지원 혹은 승인'은 이 규정에 따른 권한 하에 주어지는 승인을 의미한다. NHS의 보건연구당국(HRA)는 (보건사회복지법 2012에 따라) 2013년 4월에 기밀성자문그룹(Confidentiality Advisory Group, CAG)을 설립하며, Section 251에 대한 책임을 갖게 되었다.[50]

section 251 의 지원을 받아 기밀 정보(식별 가능한 환자 정보)에 접근할 수 있으려면, 정보 취득자의 목적이 환자 진료의 증진과 관련되어야 하며, 공익을 위한 것이어야 하며, 모든 환자로부터 동의를 얻는 것이 불가능하거나 혹은 너무 비용이 많이 들거나, 기술적으로 어려운 경우에만 허용된다. 이 신청은 기밀성 자문 그룹(CAG)에 의해 검토된다. 환자들의 설명 후 동의도 얻지 않았고, section 251에 따른 승인도 얻지 못한 경우, 2차적 데이터의 이전은 익명화되어야 한다.

(4) 디지털경제법

2017년 봄 입법화된 영국의 디지털 경제법(Digital Economy Act 2017)[51]은 디지털 서비스에 대한 접근, 디지털 기반, 온라인 음란물 규제, 저작권 규제, 전자 정부 등 다

48) https://medconfidential.org/whats-the-story/ 참고.

49) Health Service (Control of Patient Information) Regulations 2002, http://www.legislation.gov.uk/uksi/2002/1438/pdfs/uksi_20021438_en.pdf

50) http://www.hra.nhs.uk/about-the-hra/our-committees/section-251/what-is-section-251/

51) https://www.legislation.gov.uk/ukpga/2017/30/contents

양한 디지털 정책 이슈를 포괄하고 있다.

디지털 경제법은 제정 과정에서 많은 반발에 부딪혔다. 저작권 침해에 대한 형사처벌, 온라인 음란물 접근 제한을 위한 연령 확인 등의 문제도 있었지만, 제5부(part 5) '전자정부(Digital Government)'에서 규정하고 있는 정부 내 광범위한 정보공유에 대해서도 시민들의 개인정보 침해 우려가 제기되었다.[52] 영국 내각부(Cabinet Office) 프라이버시 및 소비자 자문그룹 (Privacy and Consumer Advisory Group, PCAG)의 제리 피쉔덴(Jerry Fishenden)은 이 법에 항의하여 사임하기도 하였다. PCAG가 이 법의 문제를 지적하며 내각부 장관에 계속 의견을 표명했음에도 불구하고 계속 무시를 당했기 때문이다.[53]

제5부 전자정부 부분은 더 나은 공공서비스 제공을 명분으로 데이터를 더 효율적으로 활용하는 것을 목적으로 하고 있는데, 제리 피쉔덴은 이 법의 정의가 명확하지 않고, 정보공유에 대한 과거의 모델에 기반을 두고 있다고 비판하고 있다. 즉, 이 법에서 규정한 공개(disclose)의 의미가 서로 다른 기관 간에 시스템적으로 연동한다는 것인지, 정보를 복제해서 제공한다는 것인지, 일시적으로 특정 정보에 대한 접근만을 허용한다는 것인지 등이 명확하지 않은데, 구체적인 구현 방식에 따라 보안 및 프라이버시에 미치는 영향이 크게 달라짐에도 불구하고 법에서 이것을 구체적으로 규정하지 않고 있다는 것이다. 또한, 디지털 환경에서 보다 많은 개인과 조직이 개인정보에 접근할 수 있다면, 이 법이 목적하는 바와 달리 '사기(fraud)'를 방지하기보다는 증가시킬 수 있다고 지적한다. 또한, 이 법은 시민들의 자기정보에 대한 통제권을 전반적으로 약화시키고 있다며, 보안 및 개인정보 보호의 약화는 경제 및 사회에 부정적 영향을 줄 것이라고 경고하고 있다. 그는 이러한 우려를 해소하기 위해서 법적이고 기술적인 세부사항들이 규정될 필요가 있다고 제안한다.[54]

영국의 정보인권 단체인 오픈라이츠그룹(Open Rights Group)도 공공기관에 대한 사기 방지를 위한 데이터 공유, 공공서비스를 위한 프로파일링, 국세청에 의한 데이터 공유 등에 대해 개인정보 침해 우려를 나타냈다. 이들 역시 안전조치를 법이 아니라 실행 규약에서 정하는 문제를 지적하며, 기본적인 안전조치에 대해 법제화할 것을 요구하고 있다. 다만, 이러한 안전조치가 제대로 취해진다면, 연구나 통계 목적으로 데이터를 공유하는 것에 대해서는 동의하는 입장이다.[55]

---

52) 제5부는 다음과 같은 7개의 장으로 구성되어 있다. 1장. 공공서비스 전달, 2장 시민 등록, 3장 공공 분야에의 부채, 4장 공공 분야에 대한 사기, 5장 연구 목적의 공유, 6장 세무당국에 의한 공개, 7장 통계

53) https://ntouk.wordpress.com/2017/05/03/the-canary-that-ceased-to-be/

54) Jerry Fishenden (2016). "Submission on Part 5 of the Digital Economy Bill: "Digital Government"". 2016.10.5. https://publications.parliament.uk/pa/cm201617/cmpublic/digitaleconomy/memo/DEB04.pdf

제5부 전자정부의 5장은 '연구 목적의 공유'를 규율하고 있다. 이에 따르면 공공기관이 보유한 정보는 연구 목적으로 다른 사람에게 제공될 수 있다. 그 정보가 개인정보일 경우, 다음과 같은 조건을 만족해야 한다. 첫째, 그 정보가 특정인을 식별하는 경우, 공개되기 전에 특정인의 신원이 해당 정보 내에서 식별되지 않도록, 그리고 (그 자체로 혹은 다른 정보와 결합하여) 그 정보로부터 특정인의 신원을 추론하는 것이 합리적으로 가능하지 않도록 처리되어야 한다. 둘째, 공개를 위한 그 정보의 처리에 관여하는 모든 사람은 특정인을 식별할 수 있는 우발적인 정보의 공개 위험성을 최소화하고, 그러한 정보의 의도적인 공개를 방지하기 위한 합리적인 조치를 취해야 한다. 셋째 이 공개는 공공기관에 의해, 혹은 공공기관이 아닌 경우 정보공개를 위한 처리에 관여한 사람에 의해 이루어져야 한다. 넷째, 정보가 공개되는 연구는 (71조에 따라) 승인된 것이어야 한다. 다섯째, 공개를 위한 정보의 처리에 관여한 (공공기관을 포함한) 사람, 정보를 제공받은 사람, 연구 목적으로 그 정보를 이용하는 사람은 (71조에 따라) 승인을 받아야 한다. 여섯째, 정보를 공개하거나 그 처리에 관여한 사람은 70조의 실행 규약을 유념해야 한다.[56]

71조(section 71)는 제5장의 목적을 위한 승인을 규정하고 있는데, 이는 통계 위원회(the Statistics Board)가 담당한다. 통계 위원회는 승인을 위한 조건을 수립하고 공개해야 한다. 70조(section 70)는 실행 규약을 규정하는데, 이 역시 통계 위원회가 담당한다. 통계 위원회는 64조에 따른 개인정보의 공개, 정보의 처리, 공개된 개인정보의 보유 혹은 이용에 대한 실행 규약을 공표해야 한다.

이 장의 규율하의 공공기관에서 공공서비스 혹은 성인 사회복지의 제공과 관련한 기능만을 하는 기관은 제외한다.[57]

3. 독일

(1) 개인정보 보호 법제

독일에서는 1971년에 헤센 주에서 세계 최초로 개인정보보호법이 제정되었고, 연방 차원의 개인정보보호법은 1978년 처음 제정되었다. 1983년 독일헌법재판소는 인구조사와 관련한 판결에서 '개인정보 자기결정권'을 기본권으로 인정하였다. 연방정보보호

55) ORG's response to data sharing consultation, https://www.openrightsgroup.org/ourwork/reports/orgs-response-to-data-sharing-consultation

56) Digital Economy Act 2017, section 64

57) Digital Economy Act 2017, section 73 (2)

법(Bundesdatenschutzgesetz, BDSG)은 95년 유럽연합 정보보호지침을 수용하여 2003년 1월 14일 개정되었고, 2017년 4월에는 GDPR을 수용하여 완전히 새로운 연방정보보호법이 만들어졌다. 새 연방정보보호법은 GDPR과 함께, 2018년 5월 25일 시행될 예정이다.[58] 독일에서 주 차원의 개인정보보호법은 각 주의 공공부문만을 대상으로 하지만, 연방법은 공공 및 민간 부문 전 영역에 적용된다.

독일 연방정보보호법은 개인정보의 '처리(prcessing)'를 "개인정보의 저장, 수정, 이전, 차단, 삭제"로, 그 밖의 개인정보의 활용은 '이용(use)'으로 정의하고 있다. 개인정보의 연계·결합을 위해 개인정보의 저장·수정·이전·차단·삭제 등이 수반될 수밖에 없다는 점에서 개인정보의 처리로 봐야 할 것이다. 또한, 연방정보보호법은 '익명화(Rendering anonymous)' 및 '가명화(Aliasing)' 등에 대한 정의도 두고 있다. 3a조(section 3a)에서는 개인정보의 수집·처리·이용을 가능한 최소화하고, 가능한 한 가명화 또는 익명화할 것을 규정하고 있다.

연방정보보호법은 학술연구 목적의 개인정보 처리와 관련된 많은 규정을 포함하고 있다. 그런데 특이하게도 통계 목적의 개인정보 처리에 관해서는 규정하고 있지 않다. 따라서 통계 목적의 개인정보 처리에 대해서도 연방정보보호법이 동일하게 적용되지만, 통계 관련법에서 개인정보 처리와 관련해서 별도의 규정이 있을 경우에는 해당 규정을 따르게 된다. 연방정보보호법 1조 3항은 다른 연방 법 규정이 개인정보에 적용되는 한, 이 법에 우선한다고 규정하고 있다.[59]

개인정보의 수집·처리·이용은 개인이 동의했거나, 법적 근거가 있는 경우에만 허용된다.[60] 동의는 정보주체의 자유로운 결정에 근거해야 하고, 정보주체에게 수집·처리·이용의 목적과 (개별 사례의 상황에 따라, 혹은 요청에 의해) 동의하지 않았을 경우의 결과에 대해 고지해야 한다. 동의는 서면 동의가 원칙이지만, 다른 형식이 허용되어야 할 특정 상황이 있을 수 있다. 특히 학술적 연구 분야에서는 서면 동의를 얻어야 할 경우 연구 목적이 심각하게 훼손될 수 있는 경우도 존재한다. 이 경우 정보주체에게 고지하는 위의 정보와 함께, 연구 목적이 훼손되는 이유가 서면으로 기록되어야 한다.[61]

2부(part 2)에서는 공공기관에서의 데이터 처리에 관해 규정하고 있는데, 연구 프로젝트 수행을 통한 학술적 이익이 정보주체의 이익보다 훨씬 큰 경우, 그리고 연구의 목적이 다른 방법으로 달성될 수 없거나 비례적이지 않은 노력이 필요할 경우 학술

58) 아래 분석은 기존 연방정보보호법을 기준으로 한 것이다.
59) 연방정보보호법 section 1 (3)
60) 연방정보보호법 section 4
61) 연방정보보호법 section 4a

연구의 목적으로 필요한 민감정보를 수집할 수 있고[62], 수집목적 외로 개인정보[63] 및 민감정보를 저장·수정·이용하는 것[64]이 허용된다. 수집목적 외로 민감정보의 저장·수정·이용을 허용할 것인지에 대한 상황을 평가할 때 연구 프로젝트의 과학적 이익을 공익의 맥락에서 특별히 고려하는 것이다.

공공기관의 개인정보 처리와 관련하여 20조(section 20)는 개인정보의 정정·삭제·차단 및 거부권을 규정하고 있는데, 정보수집자 혹은 제3자의 이익이 훨씬 더 중요하고, 학술적 목적, 증거로서의 사용 목적 혹은 다른 이유를 위해 필수불가결한 경우, 해당 정보를 차단하지 않으면 그러한 목적을 위한 개인정보의 이전 혹은 이용이 허용될만한 경우에는 정보주체의 동의 없이 차단된 데이터를 이전 혹은 사용할 수 있도록 하고 있다.

3부는 민간기구 및 경쟁에 참여하는 공기업에서의 데이터 처리를 규정한다. 제28조(section 28)는 연구 프로젝트를 수행하는 학술적 이익이 정보주체의 이익보다 훨씬 크고, 연구 목적이 다른 수단으로 달성될 수 없거나 비례적이지 않은 노력을 요구할 경우에는 연구기관의 학술연구 수행을 위해 개인정보의 수집목적 외 다른 목적으로의 이전 및 이용을 허용하고 있다.[65]

익명화된 형식으로 개인정보를 이전하기 위해 사업 과정에서 개인정보를 수집·저장한 경우, 개인을 식별할 수 있게 하는 속성들은 분리해서 저장해야 한다. 이러한 속성들은 저장 혹은 학술 목적으로 필요할 경우에만 다른 정보들과 결합될 수 있다.[66] 33조(section 33)은 정보주체에의 고지를 규정하고 있는데, 학술 연구의 목적으로 저장 및 이전이 필요하고 고지에 비례적이지 않은 노력이 드는 경우 고지 의무가 면제된다. 35조(section 35)는 정보주체의 정정·삭제·차단권을 규정하고 있는데, 위의 20조와 같은 조건에서 차단된 정보가 정보주체의 동의 없이 이전 혹은 이용될 수 있다.

4부(part 4)는 특별 조항을 포함하고 있다. 40조는 연구기관에서의 개인정보의 처리 및 이용을 규정하고 있는데, 학술연구 목적으로 수집 혹은 저장된 개인정보는 오로지 그러한 목적으로만 처리·이용되어야 한다. 연구 목적이 허용하는 한, 가능한 한 빨리 개인정보는 익명화되어야 한다. 그때까지 개인 식별을 가능하게 하는 속성들은 분리해서 보관되어야 한다. 이 속성들은 연구 목적에 필요한 한도 내에서만 다른 정보들과 결합될 수 있다. 학술적 연구를 수행하는 기구는 다음과 같은 조건에서만 개인정보를 공개할 수 있는데, 첫째 정보주체가 동의한 경우, 둘째 동시대 사건에 관한

62) 연방정보보호법 section 13 (2) (8)
63) 연방정보보호법 section 14 (2) (9)
64) 연방정보보호법 section 14 (5) (2)
65) 연방정보보호법 section 28 (2) (3)
66) 연방정보보호법 section 30 (1)

연구 결과를 보여주기 위해 불가피한 경우이다.[67]

(2) 보건의료 관련 법제

독일의 보건의료는 법정건강보험시스템에 기반을 두고 있으며, 이는 연금 및 실업 급여 등을 포함한 사회보험의 일부이다. 이는 독일사회법전(Social Code Book, SGB)에 의해 규율된다. 독일사회법전 5장(SGB V)은 법정건강보험을, 10장(SGB X)은 행정절차 및 사회데이터 보호를 다루고 있다.

사회데이터는 특정한 혹은 식별 가능한 자연인(정보주체)의 개인적 혹은 물질적 환경에 대한 세부사항을 의미하며, (SGB I 35조에 명시된 바와 같이) 기구에 의해 이 법전의 의무와 관련하여 수집·처리·사용되는 것을 의미한다.[68] SGB I 35조는 사회데이터의 기밀성을 규정하고 있다. 즉, 모든 사람은 자신과 관련된 사회데이터가 허가 없이 수집·처리·사용되지 않을 권리(사회적 비밀)를 가진다. 사회적 비밀의 보호는 서비스제공자 내에서도, 단지 허가된 사람에게만 사회데이터가 접근 가능하거나 전달될 수 있도록 하는 의무를 포함한다.

SGB V는 법정건강보험을 규율한다. 건강보험회사는 단지 건강보험의 목적을 위해서만 사회 데이터를 수집·저장할 수 있다.[69] 건강보험기금 및 법정건강보험의사협회는 감독기관의 허가를 받아, 데이터를 임시적 혹은 제한적 연구 프로젝트를 위해 제공할 수 있는데, 특히 역학 조사결과, 질병과 작업 환경 사이의 관계에 대한 조사, 지역적 질병에 대한 지식 등의 연구 프로젝트가 포함된다. 이때 사회데이터는 익명화되어야 한다.[70]

SGB X은 (건강보험 관련 데이터를 포함한) 사회데이터의 보호를 규율하고 있다. 사회데이터의 이전은 일정한 경우에 허용될 수 있는데, 사회 서비스 학술연구 혹은 노동 시장 및 직업에 관한 학술 연구 프로젝트를 위해 필요할 경우, 혹은 공공기관이 자신의 업무와 관련하여 사회 서비스 분야의 계획을 위한 프로젝트에 필요한 경우이다. 이때 정보주체의 정당한 이익이 영향을 받지 않아야 하고, 연구 혹은 계획의 공익성이 정보주체의 이익보다 훨씬 커야 한다. 합리적으로 개인의 동의를 얻을 수 있을 경우에는 해당 개인의 동의 없는 이전은 허용되지 않는다. 위 프로젝트의 개시에 반드시 필요한 정보주체의 성과 이름, 주소, 전화번호, 구조적 특징 또한 설문조사를 위

67) 학술적·역사적 연구 및 통계 목적의 데이터 처리와 관련하여 새로운 연방정보보호법에서는 section 27에서, 아카이빙 목적의 데이터 처리는 section 28 규정하고 있다.

68) SGB X, §67 (1)

69) SGB V §284 (1)

70) SGB V §287

해 이전될 수 있다.[71] 이러한 이전은 최고 연방기관 혹은 해당 데이터가 유래한 지역을 책임지는 주 기관의 사전 승인을 얻어야 한다.

독일 내 각 주의 암 등록 데이터의 암등록데이터센터(ZfKD)에의 제공은 2009 연방암등록데이터법(BKRG)에 따라 수행된다. 이 법에 따라 암등록데이터센터는 로베르트 코흐 연구소(Robert Koch Institute)에 설립되었다.[72] 각 주는 성별, 생년월일, 지역코드 앞 5자리 등 개인정보와 종양 진단정보 등을 암등록데이터센터에 이전한다. 국가적 암등록소를 상호 비교하기 위해 모든 암등록소는 공통의 절차에 따라 각 개인에게 고유한 통제번호를 부여한다. 이 통제번호는 암등록데이터센터 내에서 다른 데이터와 분리되어 저장되며, 통합 목적으로만 결합할 수 있다. 통합 후, 늦어도 이전 후 3년 이내에, 이 통제번호는 삭제된다.[73] 암등록데이터센터는 신청에 따라 제3자에게 데이터 이용을 허락할 수 있다. 이 경우 정당성, 특히 학술적 이익이 신뢰성 있게 입증되어야 한다. 신청서는 특히 그 목적 및 이용 범위에 있어서 입증될 수 있어야 하며, 자문위원회에 제출되어진다. 이용의 범위나 공개는 계약에 의해 규율된다.[74] 암등록데이터센터는 일정 기간 후에 통제번호를 삭제하기 때문에, 원 데이터에 오류가 있을 경우에 이를 나중에 식별하거나 수정하는 것은 불가능하다. 실제 개별 데이터로 연결하여 확인할 수 없기 때문이다. 또한, 개인정보가 변경되거나 오류가 있는 경우, 기존 데이터와 연계가 되지 않거나 중복적인 엔트리가 발생할 수 있다. (OHE Consulting Report, 2015)

(3) 연방통계법

독일 연방통계법(Bundesstatistikgesetz, BStatG)은 연방통계의 작성과 관련된 원칙, 조직, 활동을 규율한다.

가. 데이터 연계

연방통계법 13a조(section 13a)는 통계 목적의 데이터 연계에 관한 사항을 규정하고 있다. 이에 따르면, 추가적인 통계 설문조사를 수행하지 않고 통계정보를 얻기 위한 목적으로, 그리고 13조(1)에 명시된 목적[75]을 달성하기 위하여 데이터 연계가 필요한

71) SGB X §75
72) BKRG §1 (1)
73) BKRG §4
74) BKRG §5

한도에서, 다음과 같은 데이터를 연계할 수 있다.

첫째, 독일연방은행에 의해 편집된 통계의 데이터를 포함하여, 기업, 기관 및 지역 기구의 경제 및 환경 통계의 데이터

둘째, 통계 등록부의 데이터

셋째, 행정데이터 이용에 관한 법률에서 명시한 데이터

넷째, 일반적으로 접근 가능한 소스로부터 연방통계청 및 주 통계청에 의해 획득된 데이터

이러한 목적을 위해 독일연방은행은 자신의 경제 통계 데이터를 연방통계청에 이전할 수 있다. 데이터가 연계될 경우 (통계등록법의 1조(1), 4번째 문장에서 명시한) 식별 번호는 설문조사 변수에 대한 정보를 포함하고 있는 데이터 기록 내에 30년까지 저장할 수 있다. 저장 기간이 만료된 후 식별 번호는 삭제되어야 한다. 각 설문조사가 완료되었을 때 그 기간이 시작된다.

나. 행정데이터의 이용

연방통계법 5a조(section 5a)는 연방통계청에서 연방통계를 작성하거나 수정하기 전에, 공공행정기관이 관련 연방통계를 생산하기 위해 사용할 수 있는 양질의 데이터를 가지고 있는지 검토하도록 하고 있다. 연방행정기관 및 주 법에 따라 공공행정업무를 수행하는 기관은 적절성 확인을 위해, 요청에 의해 연방통계청에 행정데이터의 출처, 구조, 콘텐츠에 대한 정보와 다른 관련 메타데이터를 이전해야 한다. 연방통계청이 행정데이터의 적절성을 확인하면, 이 데이터는 관련 연방통계의 편집에 이용되어야 한다. 데이터 이전은 연방통계의 작성 혹은 수정을 위한 법 조항에 의해 규제되어야 한다.

다. 기밀성

개인정보와 가장 관련성이 높은 조항은 연방통계법 제16조(기밀성)이다. 연방통계 목적으로 제공된 개인에 관한 데이터는 특정한 법률에 명시된 바에 의하지 않고는, 연방통계의 생산을 맡은 공무원 및 공공서비스 종사자에 의해 공개되어서는 안 된다. 이러한 기밀성 의무는 업무 종료 이후에도 계속된다. 그러나 기밀성 의무는 다음의 경에는 적용되지 않는다. 첫째, 다른 형태의 동의가 적합하다는 특별한 사정이 없는

75) 연방통계법 13조(1)은 연방통계의 준비 및 생선을 위해, 그리고 평가 목적으로, 연방통계청은 EU통계규정 및 통계등록법에 따라 통계적 목적의 사업 등록부(통계 등록부)를 유지해야 한다고 규정하고 있다.

한, 관련된 사람이 서면으로 동의한 경우. 둘째, 해당 데이터가 15조 1항에서 규정된 공공기관과 관련되거나, 연방통계 작성을 위한 법 조항에 근거하여 정보 제공 의무가 있는 경우에, 일반적으로 접근 가능한 소스로부터 얻은 개별 데이터인 경우. 셋째, 연방통계청 혹은 주 통계청에 의해 다른 응답자의 개별 데이터와 결합되고 통계 결과로 제시된 개별 데이터인 경우. 넷째, 응답자나 당사자와 관련이 없는 개별 데이터인 경우.

연방통계 작성에 필요한 한도에서 통계 작성을 맡은 사람 및 기관 사이의 개별 데이터 이전은 허용된다.

학술 프로젝트의 수행을 목적으로, 연방통계청 및 주 통계청은 고등교육기관 혹은 독립적 학술연구를 수행하는 다른 기관에 개별 데이터를 제공할 수 있는데, 이 개별 데이터를 통해 응답자 등 개인을 식별하기 위해 비합리적 시간, 비용, 인력이 소요되는 경우, 즉 사실상 익명화된 개별 데이터인 경우에 한한다. 또한, 연방통계청 및 주 통계청의 특별보호구역 내에서, 기밀성을 보호하기 위한 효과적인 조치가 취해진 경우, 공식적으로 익명화된 개별 데이터에 대한 접근을 제공할 수 있다. 공무원, 공공서비스를 위해 특별선서를 한 사람, 혹은 7항에 따라 기밀성 서약을 한 사람에게만 개별 데이터에 대한 접근 권한이 주어진다.[76] 7항은 (공무원 및 공공서비스를 위해 특별선서를 한 사람이 아닌) 개별 데이터를 제공받는 사람은 이전 전에 기밀성 서약을 해야 한다고 규정하고 있다.

개별 데이터는 오로지 이전된 목적으로만 사용되어야 한다. 특히 학술연구 목적으로 이전된 개별 데이터의 경우, 학술 프로젝트가 완료되는 즉시 삭제되어야 한다. 개별 데이터를 이전받는 기관은 오로지 권한 있는 사람들만 개별 데이터에 접근할 수 있도록 조직적·기술적 조치를 취해야 한다.[77]

연방통계법 21조는 재식별 금지를 규정하고 있다. 연방통계의 개별 데이터를 통계 작성 외의 목적으로 사람, 기업, 기관, 지역 단위의 식별을 위해 다른 정보와 연계하는 것은 금지된다.

4. 프랑스

(1) 개인정보 보호 법제

프랑스에서 개인정보 보호를 위한 기본법은 1978년 1월 6일 정보, 파일 및 자유에

76) 연방통계법 section 16 (6)
77) 연방통계법 section 16 (8)

관한 법률(Loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés)이다. 이 법률은 95년 EU 정보보호지침을 수용하여 2004년 8월 6일 대폭 개정되었고, 2018년 GDPR 시행에 맞추어 2016년 10월 7일 전자공화국법(Loi n°2016-1321 du 7 octobre 2016 pour une République numérique)이 통과되면서 개인정보 보호 관련 규정도 개정되었다.

제6조는 개인정보 처리의 원칙을 규정하고 있는데, 개인정보는 공정하고 적법한 방식으로 수집·처리되어야 하며(1항), 특정되고 명확하며 적법한 목적으로 수집되고 이와 양립하지 않는 방식으로 추가 처리되어서는 안 된다. 그러나 4장(데이터 처리 개시 전의 절차), 5장(개인정보처리자의 의무 및 개인의 권리)의 섹션1(개인정보처리자의 의무), 9장(보건 분야의 연구, 학술, 평가 목적으로의 개인정보 처리)에서 규정한 원칙과 절차를 준수하며 정보주체에 관한 결정을 하기 위해 이용되지 않는다면, 통계 목적, 혹은 학술적·역사적 연구 목적으로 추가 처리하는 것은 애초의 수집목적과 양립하는 것으로 본다. (2항)

8조는 민감정보의 수집·처리에 관해 규정하고 있는데, 민감정보의 수집 및 처리는 원칙적으로 금지(1항)되지만 8조 2항에 따른 경우에는 예외인데, 이 법 25조의 조건 하에서 수행되고 '통계 분야의 법적 의무, 조정 및 기밀성에 관한 법률(Act n° 51-711 of June 7th, 1951 on Legal Obligation, Coordination and Confidentiality in the Field of Statistics, 프랑스 통계법)'에 따라 국가통계정보위원회와의 협의 후에 국가통계청(INSEE) 혹은 부처통계서비스에 의해 수행되는 통계처리, 혹은 9장에서 규정하는 방식으로 보건 분야의 연구, 조사, 평가에 필요한 처리가 이에 포함된다.

32조는 개인정보처리자의 통지 의무를 규정하고 있는데, 애초에 다른 목적으로 수집된 개인정보를 문화유산법(the Heritage code) 제2권(Book II)에 규정된 바에 따른 역사적, 통계적, 학술적 목적으로의 보유를 위해, 혹은 프랑스 통계법 7조의2(7bis)의 규정에 따른 통계 목적의 개인정보 재이용에 필요한 처리에는 이러한 의무가 면제된다. (3항) 역사적·통계적·학술적 목적의 처리를 위해 개인정보는 애초 수집목적에 필요한 기간 이상으로 저장될 수 있다. 보존되는 데이터의 결정은 문화유산법 L212-3조의 조건에 따른다. (36조) 39조는 정보주체가 자신의 개인정보가 처리되고 있는지 여부 등에 대해 개인정보처리자에게 질의할 권리를 규정하고 있다. 그런데 개인정보가 명백히 해당 정보주체의 프라이버시 침해 위험이 없는 형식으로, 오로지 통계 작성을 위한 목적, 혹은 학술적·역사적 연구를 위해 필요한 기간 이상 저장되지 않는 경우에는 이 조항이 적용되지 않는다.

주목할 점은 프랑스에서는 공공서비스를 관리하는 하나 이상의 법인에 속하고, 서로 다른 공익 목적의 파일들 연계, 혹은 주목적이 서로 다른 기관에 속하는 파일의

연계 목적으로 자동화된 처리를 하는 경우, 프랑스의 개인정보 감독기구인 CNIL의 허가를 받도록 하고 있다는 점이다. (25조 1항 5호) 25조는 CNIL의 허가가 필요한 처리를 규정하고 있는데, 이러한 파일의 연계 목적의 자동화된 처리뿐만 아니라, 민감정보의 처리, 유전 정보의 자동화된 처리 등이 이에 포함된다.

또한, 제27조 1항의 1, 2항의 1과2 규정의 예외로, 국가등록번호인 사회보장번호(Numéro de sécurité sociale, NIR)를 포함한 데이터의 처리도 CNIL의 허가를 받아야 한다. 이때의 정보 처리는 오로지 과학 및 역사 연구만을 목적으로 하는 경우에 해당하며, 국가등록번호가 각 연구 프로젝트에 고유한 특정한 임의의 코드로 교체되는 방식으로 사전에 암호화 처리되어야 한다. 암호화 작업 및 그로부터 나온 코드를 통한 파일의 연계는 동일한 처리자가 수행해서는 안 된다. 암호화 작업은 CNIL의 공개된 의견을 받은 후에 국참사원(Conseil d'État)[78]의 시행령(decree)으로 규정한 주기로 갱신되어야 한다. (25조 1항 9호)

26조(국가안보 및 형사 관련 처리)와 27조(국가등록번호 등의 처리 등)는 25조에 규정된 사항 중 CNIL의 허가가 필요 없는 예외를 규정하고 있다. 또한, 이 법은 제9장에서 보건의료 분야의 연구, 조사, 평가 목적의 개인정보 처리에 대해 상세하게 규정하고 있다. 53조는 일정한 경우(예를 들어, 환자치료 목적의 개인정보 처리 등)를 제외하고는 보건의료 분야의 연구 및 조사, 치료·예방 활동의 평가 및 분석을 목적으로 한 개인정보의 자동처리에 이 법이 적용됨을 밝히고 있다. 보건 분야의 공익적인 연구, 조사 혹은 평가를 목적으로 한 개인정보의 처리는 CNIL의 허가를 받아야 하는데(54조 1항), CNIL은 공공보건법(Public Health Code) L1121-1조에 규정된 인간연구 관련 승인 요청을 위한, L.1123-6조에 규정된 개인의 보호 권한을 갖는 자문위원회, 혹은 인간연구 외의 연구 및 평가 신청 승인을 위한, 연구, 조사 및 평가를 위한 자문위원회의 의견을 받아 결정을 내린다. 자문위원회는 1달 이내에 연구 방법, 해당 개인정보 이용의 필요성, 프로젝트의 목적 및 학술적 질과 해당 개인정보의 연관성 등에 대한 의견을 제시해야 한다. 필요한 경우 위원회는 신청자에게 프로젝트의 수정을 권고할 수 있다. (54조 2항) 각 요청에 대해, CNIL은 이 조항의 적용에 대한 신청자의 확약과 신청자의 요청이 임무 및 목적에 부합한다는 것을 검증한다. 신청자가 처리가 예상되는 개인정보 중에 특정 정보의 필요성을 입증할 충분한 증거를 제시하지 못한다면, CNIL은 해당 정보를 보유한 기관에 그 정보의 제공을 금지하고 단지 일부 제한된 데이터의 처리만 허용하도록 할 수 있다. CNIL은 처리에 필요한 데이터 보유기간을 결정하고 안전성과 기밀성을 보장하기 위해 취해진 조치를 평가한다. (54조 3항) CNIL은 조사 절차를 단순화하기 위해 가장 통상적으로 사용되는 방식을 '기준

78) 프랑스 최고 행정법원

방법'으로 만들어 승인, 공표할 수 있다. (54조 4항)

보건 전문가는, 기밀성 의무에도 불구하고, 자신이 보유한 개인정보를 53조에 따라 허가받은 개인정보 처리를 위해 이전할 수 있다. 이 개인정보가 개인 식별이 가능할 경우, 개인정보의 이전은 기밀성을 보장할 수 있도록 해야 하며, CNIL은 기술적 절차에 대한 권고안을 채택할 수 있다. 데이터 처리의 결과가 제시될 경우, 개인을 직간접적으로 식별할 수 있어서는 안 된다. 관리자는 정보 및 그 처리의 보안을 보장해야 한다. 개인건강정보를 처리하는 사람 및 데이터에 접근하는 사람은 기밀성 의무를 지며 이를 위반할 경우 형법 226-13조에 따라 형사처벌을 받는다. (55조)

(2) 보건의료 관련 규정

프랑스에는 2004년 8월 13일 건강보험법(Loi n° 2004-810 du 13 août 2004 relative à l'assurance maladie)에 의해 개인건강기록(dossier médical personnel, DMP)이 시행되었다. 이는 프랑스의 국가적인 전자건강기록으로서 4가지 요소를 가지고 있다. 국가적인 체제의 수립 목적, 데이터 보유기관에 대한 형식(formalities), 환자에 관한 (동의 및 식별) 양식, 보건 전문가의 식별 등이다. 그러나 각 보건 전문가나 기관에 따라 다른 전자건강기록들도 사용되고 있다. DMP는 2006년 첫 시범사업 이후, 2011년 4개 지역에서 공식적으로 시작되었다. 현재 프랑스 전체 영역을 포괄하고 있고, 국가법령에 의해 규율되지만, 아직 일반적으로 사용되는 단계는 아니다. (Milieu, 2014)

환자들이 DMP의 생성에 동의하면[79], 국가 건강보험(assurance maladie)의 지원을 받는 모든 환자는 무료 DMP를 가질 수 있다.

데이터 연계를 용이하게 하기 위해서는 고유식별번호가 필요한데, 프랑스의 모든 시민은 국가등록번호인 사회보장번호를 가지고 있다. 이 번호는 보건의료당국에서 건강보험카드(carte vitale)를 발급하는 데 사용하는데, 병원에서는 사용되지 않는다. 전자의료기록에 사용하기에는 너무 민감하기 때문이다. (OHE Consulting Report, 2015) 그래서 DMP 역시 사회보장번호가 아니라 국가건강식별번호(Identifiant National de Sante, INS)를 통해서 생성되는데, INS는 국가건강보험 수혜자에게 부여되는 번호이다. (Milieu, 2014)

보건 수혜자에게 주어지는 건강 관련 관리에 필요하다면 진단 혹은 치료 등 어떤 정보도 DMP에 포함될 수 있다. DMP는 환자가 소유하는 것으로 설계되어있다. 환자들은 이 기록에 접근할 수 있으며, 어떤 보건 전문가가 이에 접근하는지 모니터할 수

79) Public Health Code (Code de la santé publique), Article L.1111-8.

있다. 또한, 그 기록 중 공유될 수 있는 요소를 지정할 수 있다. 그러나 법에 따라 환자가 동의할 수 없는 응급 상황에는 의사가 이에 접근할 수 있도록 허용하고 있다. (OHE Consulting Report, 2015)

공공보건법전(Code de la santé publique) 및 관련 수정법들은 DMP의 생성부터 데이터 보유기관이 승인되는 절차까지 DMP와 관련된 대부분의 조항을 포함하고 있다. 사회보장법전(Code de la sécurité sociale) 및 관련 수정법들은 DMP 체제의 원칙과 목적을 반영하고 있다. 또한, 의료 행위와 관련된 프라이버시 및 비밀보호 권리와 관련된 조항을 포함하고 있다.

2009년 9월 8일 공익 그룹 설립 협정의 승인 명령(Arrêté du 8 septembre 2009 portant approbation de la convention constitutive d'un groupement d'intérêt public)은 국가건강공유정보시스템국(National Agency of health shared information systems)인 ASIP Santé의 설립을 규정하고 있는데, 이 기구는 비영리 법정기구로서 보건 및 의학사회 분야의 공유정보시스템 개발을 증진하고, 보건 서비스의 질 향상을 목적으로 한다. (Milieu, 2014)

한편, 앞서 보았다시피, 프랑스의 개인정보 보호 법제에서 전자건강기록의 처리, 전자파일 보유와 관련된 안전조치 및 연구 목적 활용에 대해 규율하고 있다.

공공보건법전 및 바이오윤리법(Loi de Bioethique, 2004-2011)에 의하면, 지역 윤리위원회는 개입 연구, 돌봄 연구 표준, 의학적 및 다른 보건제품, 유전학, 생리학 등 추가적인 연구 영역의 승인을 책임진다. 현재 39개의 지역 윤리위원회가 존재하는데, 위원회는 생의학, 윤리, 법, 사회과학, 의학, 환자 등 다양한 분야의 28명으로 구성된다. 국가적 차원에서는 국가협의윤리위원회(National Consultative Ethics Committee) 존재하며, 논란이 되는 이슈를 검토한다. (OHE Consulting Report, 2015)

(3) 통계법

프랑스 통계법은 '통계 분야의 법적 의무, 조정 및 기밀성에 관한 법률(Act n° 51-711 of June 7th, 1951 on Legal Obligation, Coordination and Confidentiality in the Field of Statistics)'이고, 통계청 업무는 INSEE(National Institute of Statistics and Economic Studies)가 맡고 있다.

통계법 6조는 통계정보의 기밀성 보호를 다루고 있다. 이에 따르면, 사생활 및 가족생활과 관련된, 보다 일반적으로 사적 성격의 사실과 행동에 관련된, 설문조사의 개별 데이터는 원칙적으로 아카이브 보관 서비스에 의해 설문이 수행된 후 75년, 혹은 당사자 사망 이후 25년 동안(후자가 먼저 도래할 경우) 공개되지 아니한다. 다만, 공식

통계 혹은 학술적·역사적 연구 목적을 위한 요청에 따라, '통계 기밀성 위원회'의 의견을 요청하여 청취한 후, 아카이브 행정당국의 결정에 따라 이루어질 경우에는 예외이다. 설문조사의 경제적·재정적 성격의 개별 데이터에도 유사한 규정이 적용된다. 이 정보는 조세 혹은 경제적 처벌 목적으로 사용되어서는 안 된다. 이 법에 따라 설문조사를 수행하는 직원은 기밀성 의무를 지며, 이를 위반할 경우 형법 226-13 및 226-14조에 따라 처벌된다. 이 법에 따라 수행된 통계적 인구조사나 설문조사는 '공공 아카이브'로 간주된다. 6조의2(6bis)는 통계 기밀성 위원회에 관한 사항을 규정하고 있다.

7조의2(7bis)는 데이터의 이전과 연계를 규율하고 있다. (법에서 따로 규정하고 있지 않다면) 국가통계위원회의 자문과 경제 장관의 요청에 따라, 공공서비스 수행과정에서 수집된, 성생활 관련 데이터를 제외한 자연인의 정보 및 법인의 정보는, 통계 작성만을 위한 목적으로 국가통계청(INSEE) 혹은 통계부장관에 이전된다. 이 중 개인건강정보의 경우 보건부 장관의 요청에 따라 국가통계청(INSEE) 혹은 공중보건 정책의 평가, 실행, 결정에 관여하는 부처들에게 제공될 수 있으나, 이는 오직 인구보건상태, 공공보건정책 및 국민의 질병과 관련한 사회보장 시스템에 의한 공공보건 관리에 관한 통계 작성만을 목적으로 이전된다.

개인건강정보의 이전 절차는 절대 개인 식별을 허용해서는 안 된다. 다만, '1978년 1월 6일 정보, 파일 및 자유에 관한 법률'의 규정에 따라, 통계 편집 조건상 직간접적 식별요소가 필요할 경우, 특히 개인의 샘플을 추출하여 서로 다른 소스의 데이터 결합의 목적을 위해서만 허용된다. 이 목적을 위해 데이터 처리 허가를 받은 법인이 지정한 책임자만이 INSEE 혹은 공공보건정책의 정의, 수행, 평가에 참여하는 부처의 통계부서에 이전된 개인건강정보를 받을 수 있다. 이 데이터의 사용 후에 개인 식별요소는 폐기되어야 한다.

개인정보의 이전은 개인정보 보호법제의 적용을 받는다. 규제법 및 (이러한 이전이 두 개의 서로 다른 법인 사이에서 일어날 경우) 정보 제공자와 수신자의 계약은 이전 절차, 데이터 처리의 목적, 통계처리 목적으로의 사용 후 처리 등에 관해 명시해야 한다. 법인에 관한 정보의 이전은 경제 장관 및 관련 분야 장관이 공동으로 허가해야 한다. 국가통계청 및 통계부처의 직원은 그들이 업무상 알게 된 정보에 대해 기밀을 유지할 의무가 있다.

통계 기밀성 위원회는, 필요하다면 행정부 혹은 관련 정보를 수집한 법인의 의견을 받아, 학술적 연구 목적 혹은 경제 연구 목적으로 국가통계청 및 통계부처에 이전된 개별 데이터에 대한 접근에 대해 권고할 권한이 있다. (7ter조)

5. 미국

(1) 개인정보 보호 법제

미국에는 공공과 민간을 모두 포괄하는 개인정보 보호법제가 없다. 연방정부가 보유한 개인정보를 규율하는 프라이버시법(The Privacy Act)이 1974년 제정되었고, 민간 영역에서는 부문별로 개인정보를 보호하는 개별법을 두고 있다.

프라이버시법은 연방기관의 기록 시스템에서 유지되는 개인 식별 정보(Personally Identifiable Information)의 수집, 유지, 이용, 배포를 규율한다. 개인정보 공개의 조건[80], 정보주체의 접근·정정·삭제 요구권[81], 개인정보 수집 기관의 의무[82] 등을 규정하고 있다.

원칙적으로 어떠한 기관도 개인의 서면 요청 혹은 사전 서면 동의가 없으면, 보유하고 있는 개인정보를 다른 사람이나 기관에 공개해서는 안 된다. 다만, 12가지의 법정 예외를 두고 있는데, title 13 조항에 따른 인구조사, 설문조사, 관련 활동을 계획 혹은 수행하기 위한 목적으로 인구조사국에 제공하는 경우, 해당 기록이 오로지 통계 조사 혹은 보고기록으로만 사용될 것임을 해당 기관에 사전에 적절한 서면 확인을 받은 수신자에게 개인 식별이 불가능한 형태로 이전할 경우가 이에 포함된다. 이 조항에 따라 공공기관의 기록은 통계 목적으로 인구조사국에 이전될 수 있다.

프라이버시법은 연방기관 기록 시스템 내의 어떠한 기록도, 보유기관과 수령 기관 혹은 비연방기관 사이의 서면 계약 없이, 컴퓨터 매칭 프로그램 사용을 목적으로 수령 기관 혹은 비연방기관에게 제공되지 않는다고 규정하고 있다. 이 계약은 a) 프로그램 실행의 목적 및 법적 근거, b) 프로그램의 정당성 및 예상 결과, c) 사용될 각 데이터 요소, 매칭될 대략의 레코드 수, 시작일과 완료일 등 매칭될 레코드에 대한 설명, d) 적용 시점의 개별적인 고지 절차 및 그 이후 정기적인 고지, e) 매칭 프로그램에서 생성된 정보의 검증 절차, f) 수령 기관 등에서 생성된 식별 가능한 레코드의 보유 및 적시의 삭제 절차, g) 매칭되는 레코드와 그 결과의 행정적, 기술적, 물리적 보안을 보장할 절차, h) 레코드의 복제 및 재제공 금지, i) 레코드 사용을 규율하는 절차, j) 레코드의 정확성에 대한 평가 정보, k) 감사원장(Comptroller General)이 계약 이행의 감시를 위해 필요하다고 생각하는 모든 기록에 접근할 수 있다는 것 등을 포함한다.[83]

---

80) 5 U.S.C. § 552a (b)
81) 5 U.S.C. § 552a (d)
82) 5 U.S.C. § 552a (e)
83) 5 U.S.C. § 552a (o) (1)

(2) 보건의료 관련 법제

가. HIPAA

미국에서 보건의료 영역의 개인정보와 관련된 기본적인 법률은 건강보험 양도 및 책임법(The Health Insurance Portability and Accountability Act of 1996, HIPAA)이다. 미국 보건복지부(HHS)는 HIPAA Title II에 따라 프라이버시 규칙(Privacy Rule)과 보안 규칙(Security Rule)을 제정했다. 프라이버시 규칙은 개인 의료기록 및 다른 개인의료정보의 보호를 위한 국가적인 표준을 설정한다. 이 규칙은 개인의료정보를 보호하기 위한 적절한 안전조치를 요구하며, 환자들의 허가 없이 그러한 정보가 이용 및 공개될 경우의 한계와 조건을 설정하고 있다. 또한, 환자들에게 자신의 건강 기록을 검토하고 복사본을 취득할 권리 및 수정을 요구할 권리 등 의료 정보에 대한 환자들의 권리를 규정하고 있다. 이 규칙은 연방규제코드(CFR) Part 160 및 Part 164의 Subparts A 및 E에 포함되어 있다.[84]

이 규칙은 건강보험(health plan), 보건의료데이터저장소(health care clearinghouse), 전자적으로 거래와 연관된 건강정보를 이전하는 보건의료제공자에 적용된다. 또한, 건강정보기관 등 제휴사업자에게도 적용될 수 있다.[85] 이 규칙에서 규정하고 있는 '건강 정보(health information)'란 보건의료제공자, 건강보험, 공공보건기관, 고용주, 생명보험사, 학교 및 대학 혹은 다른 보건의료데이터저장소가 생성 혹은 취득하는 정보, 그리고 개인의 신체적·정신적 건강이나 조건, 개인에게 보건의료 서비스의 제공, 보건의료 서비스 제공에 대한 지불과 관련된 정보로서 유전 정보를 포함하며, 구술된 것이든 어떤 형태의 미디어에 기록된 것이든 포함한다. '보호되는 건강정보(Protected Health Information, PHI)'는 개인 식별이 가능한 건강정보(individually identifiable health information)를 의미한다.[86]

HIPAA법의 적용을 받는 기관은 치료, 지불, 보건의료 운영 등의 촉진을 목적으로만 환자의 서면 허가 없이 PHI를 이용 및 공개할 수 있다.[87] 다른 목적으로 이용할 경우에는 정보주체의 서면 동의가 있어야 한다.

HIPAA는 PHI를 이용, 생성, 공개하는 모든 연구에 적용된다. 의료기록 평가 혹은 새로운 의료기록 생성을 포함하는 과거의 혹은 향후의 데이터 수집에 적용된다. 프라이버시 규칙은 연구 목적의 PHI 이용 혹은 공개를 다음과 같은 조건 하에 허용한다.

---

84) HHS. "The HIPAA Privacy Rule". https://www.hhs.gov/hipaa/for-professionals/privacy/index.html

85) 45 CFR 164.104

86) 45 CFR 160.103

87) 45 CFR 164.502

원칙적으로 정보주체의 서면 허가(authorization) 없이는 PHI를 이용 및 공개할 수 없다.[88] 다만, 다음과 같은 세 가지 경우에는 서면 허가 없이 PHI를 연구[89] 목적으로 이용 및 공개할 수 있다.

첫째, 기관평가위원회(Institutional Review Board, IRB) 혹은 프라이버시 위원회(Privacy board)로부터 허가 예외의 승인을 받은 경우, 연구 준비 평가(Reviews preparatory to research)를 위한 경우, 사망자 정보에 관한 연구인 경우[90]

둘째, HIPAA가 설정한 표준에 따라 PHI가 비식별화된 경우[91]

셋째, 연구자와 보유기관 사이에 누가 데이터를 이용 혹은 취득할 수 있는지 제한하고, 연구자가 개인과 접촉하는 것을 금지하는 내용의 데이터 이용계약이 체결되고, 특정한 직접 식별자가 제거된 상태로, '제한된 데이터셋(Limited data set)'의 형태로 제공될 경우[92]

첫 번째 경우에서 기관평가위원회(the Institutional Review Board, IRB) 혹은 프라이버시 위원회는 최소한 다음과 같은 기준으로 승인 여부를 평가한다.

- PHI의 이용 혹은 공개가 개인의 프라이버시에 단지 최소한의 위험만 야기할 경우. 이를 위해 개인 식별자의 부적절한 사용 및 공개로부터 보호할 적절한 계획, 연구 수행 후 가능한 한 빨리 식별자를 삭제할 적절한 계획, PHI가 재사용되거나 제3자에게 제공되지 않을 것이라는 적절한 서면 확인이 필요하다.

- 면제 혹은 변경 없이 사실상 연구가 수행될 수 없는 경우

- PHI에 대한 접근 및 이용 없이 사실상 연구가 수행될 수 없는 경우 면제나 변경이 정보주체의 권리 및 복지에 부정적 영향을 주지 않을 경우

두 번째, HIPAA가 설정한 표준에 따라 PHI가 비식별화된 경우 건강정보는 더 이상 PHI로 간주되지 않는다. 건강정보가 개인을 식별하지 않고 그 정보가 개인 식별에 사용될 수 있다고 믿을만한 합리적인 근거가 없는 경우에는 PHI로 보지 않는다. 프라이버시 규칙은 건강정보가 더 이상 PHI로 간주되지 않도록 비식별화하기 위한 두 가

88) 45 CFR 164.508

89) 여기서 '연구(research)'란 일반적인 지식을 발전시키거나 기여하기 위해 고안된, 연구 개발, 테스트 및 평가를 포함하는, 체계적인 조사를 의미한다. (45 CFR 164.501)

90) 45 CFR 164.512 (i)

91) 45 CFR 164.514 (b)

92) 45 CFR 164.514 (e)

지 방법을 제시한다.

첫째는 전문가 결정 방법으로 “일반적으로 수용되는 통계적, 과학적 원칙 및 정보의 비식별화 방법에 적절한 지식과 경험을 가진 사람”이 통계적, 과학적 원칙과 방법을 적용하여, 해당 정보가 단독으로 혹은 다른 합리적으로 접근 가능한 정보와 결합하여, 취득자에 의해 식별될 위험이 “매우 작음(very small)”을 판단하는 방법이다. 그러한 결정을 정당화하는 분석의 방법과 결과는 문서화되어야 한다.

두 번째는 세이프하버(safe harbour) 방법으로 18개의 고유식별자를 제거하는 방법이다. 이는 해당 개인뿐만 아니라, 친척, 고용주, 다른 가족 구성원의 식별자를 포함한다. 이 18개의 개인 식별자는 다음과 같다. 이 정보의 이용기관은 해당 정보가 단독으로, 혹은 다른 정보와 결합하여 개인이 식별될 수 있다는 것을 실제로 알 수 없어야 한다.

(A) 이름

(B) 거리 주소, 시, 군, 구역, 우편번호, 이와 동등한 지역코드를 포함하여 주보다 작은 모든 지리적 구분. 다만, 인구조사국으로부터 현재 얻을 수 있는 데이터에 따라, 앞의 세 자리 수가 동일한 모든 우편번호(ZIP 코드)를 결합하여 형성된 지역이 2만 명을 넘거나, 2만 명보다 적은 모든 지역 단위의 첫 세 자리 우편번호를 000으로 변경한 경우, 우편번호 앞 세 자리는 예외로 함.

(C) 개인과 직접 관련된 (연도를 제외한) 모든 일자 요소. 출생일, 입원일, 퇴원일, 사망일을 포함. 그리고 89세 이상의 모든 나이, 그러한 나이를 가리키는 (연도를 포함한) 일자의 모든 요소. 다만, 이 경우 90 이상의 단일 범주로 총계화될 수 있음.

(D) 전화번호

(E) 팩스 번호

(F) 전자메일주소

(G) 사회보장번호

(H) 의료기록번호

(I) 건강보험수혜자 번호

(J) 계좌번호

(K) 증명서/면허증 번호

(L) 차량 번호판 번호를 포함한 차량 식별자 및 일련번호

(M) 기기 식별자 및 일련번호

(N) URLs

(O) IP 주소

(P) 지문 및 성문을 포함한 생체식별정보

(Q) 얼굴전체사진 및 그에 준하는 이미지

(R) (section (c) (재식별)에서 허용하고 있는 것을 제외한) 다른 고유식별번호, 특징, 코드

그런데 비식별 데이터에 대해 이후에 재식별될 수 있도록 코드나 식별을 위한 다른 방법을 적용할 수 있다. 다만, 그 코드나 다른 방법이 개인과 관련된 정보로부터 추정할 수 없어야 하며, 이를 다른 목적으로 사용하거나 재식별 방법을 공개해서는 안 된다.

'제한된 데이터셋'은 앞의 비식별 정보보다 개인 식별성이 강한 정보이다. 이 조항은 식별자를 모두 제거할 경우 연구의 수행이나 데이터 연계 등이 불가능해질 경우가 있을 수 있기 때문에 마련된 것이다. '제한된 데이터셋'은 다음과 같이 16개의 '직접식별자'를 제거한 것이다. 앞의 비식별조치와 비교해보면, 세부 주소 정보, 일자 정보, 기타 식별자 정보가 추가될 수 있다.

(1) 이름
(2) 도시, 주, 우편번호를 제외한 주소
(3) 전화번호
(4) 팩스 번호
(5) 전자메일주소
(6) 사회보장번호
(7) 의료기록번호
(8) 건강보험수혜자 번호
(9) 계좌번호
(10) 증명서/면허증 번호
(11) 차량 번호판 번호를 포함한 차량 식별자 및 일련번호
(12) 기기 식별자 및 일련번호
(13) URLs
(14) IP 주소
(15) 지문 및 성문을 포함한 생체식별정보
(16) 얼굴전체사진 및 그에 준하는 이미지

다만, 제한된 데이터셋은 연구, 공중보건, 보건의료 운영의 목적으로만 이용 및 공개될 수 있으며, 데이터 취득자와 데이터 이용계약을 체결해야 한다. 데이터 이용계약은 다음과 같은 내용을 포함해야 한다.

- 데이터 취득자에게 허용된 이용과 공개의 범위
- 허가받은 사람이 누구인지
- 데이터 취득자가 준수해야 할 내용. 즉 데이터 취득자는 허용된 이상으로 이용하

거나 공개하지 않아야 하고, 적절한 안전조치를 취해야 하며, 데이터 이용계약에서 벗어난 이용 및 공개가 이루어질 경우 보고해야 한다. 또한, 중개자에게도 똑같은 제한과 조건이 적용되며, 개인을 식별하거나 접촉하지 말아야 한다.

HIPAA가 개인 건강정보를 보유한 모든 보유기관에 적용되는 것은 아니다. 예를 들어, 국가보건기구(National Institutes of Health, NIH)가 보유한 개인 건강정보에는 HIPAA가 적용되지 않는다. 다만, NIH는 공공기관이기 때문에, 프라이버시법이 적용된다. 그러나 비의료 건강 서비스 제공자(체육관, 영양 상담사 등), 민간 연구소, 유전자 테스트 서비스, 마케팅 설문에 포함된 환자 정보, 건강관리 애플리케이션, 소셜미디어, 은행 등에 포함된 건강정보는 어떠한 법적 보호도 받지 못하고 있다. 이렇게 규제받지 않는 데이터 보유기관의 식별 데이터들이 종종 마케팅 목적으로 팔리기도 한다. (OECD, 2015)

건강보험업체, 보건의료제공자, 공공기관 등에 소속되지 않은 연구자의 경우 프라이버시 보호를 위한 어떠한 법적 규율도 받지 않고 있다. 이는 대학이나 영리/비영리 연구기관에 소속된 대부분의 연구자를 포함한다. 또한, HIPAA의 규제를 받는 데이터가 적용대상 기관으로부터 적용대상이 아닌 연구자에게 전달되었을 때, 더 이상 HIPAA의 적용을 받지 않게 된다. 다만, 인간 대상 의료 연구의 경우 '공통 규칙(Common Rule)'의 적용을 받게 된다. (OECD, 2015)

HIPAA의 보안 규칙은 HIPAA의 적용대상 기관이 생성, 취득, 사용, 유지하는 전자적 개인의료정보의 보호를 위한 국가적인 표준을 설정한다. 이 규칙은 전자적 PHI의 기밀성, 무결성 및 보안을 보장하기 위한 적절한 행정적, 물리적, 기술적 안전조치를 취할 것을 요구한다. 이 규칙은 45 CFR part 160과 part 164의 subpart A 및 C에 포함되어 있다.[93]

나. 공통규칙

연구 목적의 보건의료 데이터 이용에는 "공통 규칙(Common Rule)"이라 불리는 인간 주체의 보호를 위한 연방정책(The Federal Policy for the Protection of Human Subjects)이 적용될 수 있다. 이는 1991년에 발표되어 15개 연방 부처 및 기관의 규칙으로 제정되었다. (보건복지부의 경우 연방규제코드 Title 45, part 46, subpart A에

93) HHS. "The HIPAA Security Rule".
https://www.hhs.gov/hipaa/for-professionals/security/index.html

포함) 이는 이 부처 및 기관에 의해 수행, 지원, 혹은 규제를 받는 인간 주체 연구에 적용된다.

보건복지부(HHS) 규정은 subparts B, C, D를 포함하는데, 이는 임산부·인간 태아·신생아, 죄수, 어린이에게 추가적인 보호를 제공한다. 또한, 연방규제코드 Title 45, part 46은 윤리평가위원회(IRB)의 설립을 규정하고 있다. IRB는 각 기관 내에서 수행되는 인간 주체연구의 평가를 위한 행정위원회로서, 모든 연구 행위의 승인, 수정 요구, 불승인 등을 할 권한을 가진다. IRB는 인간 주체연구의 승인을 위해 다음의 기준에 근거해 판단한다. (1) 위험 최소화 (2) 위험/이익 비교 (3) 적절한 주체 선정 (4) 정보기반 동의 (5) 안전 보장을 위한 데이터 모니터링 (6) 프라이버시 보호 및 기밀성 (7) 취약한 주체에 대한 보호. (OHE Consulting Report, 2015)

다. 국가연구법 (The National Research act)

1974년 만들어진 국가연구법에 따라 바이오의학 및 행동 연구의 인간 주체 보호를 위한 국가위원회(이는 이후 의학, 바이오의학, 행동 연구의 윤리적 문제 연구를 위한 대통령 위원회로 계승되었다)가 설립되었고 인간 주체연구 및 의약품의 인간 실험 이용의 감독 및 규제를 위한 가이드라인을 개발했다. 국가위원회가 발표한 벨몬트 보고서는 향후 HIPAA 및 공동규칙에 영향을 미쳤는데, 다음과 같은 세 가지 윤리원칙을 제시하고 있다. (OHE Consulting Report, 2015)

- 인간 존중 : 연구는 모든 인간의 자율성을 보호하고 정보기반 동의를 허용해야 한다.

- 자애로움 : 연구는 “해를 끼치지 말아야 하며”, 연구 주체에 대한 이익을 극대화하고 위험을 최소화해야 한다.

- 정의 : 연구는 합리적이고, 비착취적이며, 잠재적 연구 참여자들에게 공정하고 동등하게 비용과 이익을 분배해야 한다.

(3) 통계 관련 법제

가. 기밀 정보의 보호 및 통계 효율성에 관한 법률(CIPSEA)

기밀 정보의 보호 및 통계 효율성에 관한 법률(the Confidential Information Protection and Statistical Efficiency Act, CIPSEA)은 2002년 전자정부법(E-government Act of 2002)의 제5장으로 입법되었다. 이 법률은 미국의 통계 관련 기관들 사이에 통계 목적으로 수집되는 정보에 대해 단일한 기밀성 보호 기준을 수립

하고, 노동통계국(the Bureau of Labor Statistics), 경제분석국(the Bureau of Economic Analysis), 인구조사국(the Census Bureau) 등 지정통계기관[94] 사이에 일부 데이터의 공유를 허용하고 있다. (전남대산학협력단, 2016)

이 법의 A절(Subtitle A)에서는 기밀 정보의 보호를 다루고 있는데, 정보보호에 대한 공중의 신뢰 저하는 통계 분석과의 정확성과 완전성에 부정적 영향을 미칠 수 있기 때문에, 정보 기밀성이 통계 프로그램에 대한 공중의 협력을 얻는데 필수적임을 인식하고 있다. 이에 따라 이 법의 목적으로 첫째 통계 목적으로 기관에 제공된 개인 및 단체의 정보가 오로지 통계 목적으로만 사용될 것을 보장할 것, 둘째 정보를 제공한 개인 및 단체의 정보가 이 법에서 허용하지 않는 한, 식별 가능한 형태로 공개되거나 다른 목적으로 사용되지 않는 것을 보장할 것, 셋째 접근 및 이용 통제를 통해 개별 식별이 가능한 정보의 기밀성을 보장할 것을 규정하고 있다.[95]

이러한 목적에 맞게, 통계 목적의 데이터 및 정보는 오로지 통계 목적으로만 이용할 것을 규정하고 있고, 이 정보들은 응답자의 동의가 있을 경우를 제외하고는 통계 목적 외의 사용을 위해 개인 식별이 가능한 형태로 제공되지 않으며, 이러한 공개는 해당 기관이 승인했을 경우에만 가능함을 규정하고 있다. 또한, 통계 기관은 통계 외 목적으로 자신이 수집한 정보를 명확하게 구분해야 하며, 데이터 수집 전에 비통계적 목적으로 사용된다는 사실을 공중에 고지해야 한다.[96]

한편, 위에서 언급한 3개의 지정통계기관은 서면 계약 하에 '사업 데이터(business data)'를 식별 가능한 형태로 상호 공유할 수 있다.[97] 다만, 사업 데이터의 기밀성을 보장하기 위한 적절한 보안 절차가 취해져야 한다. 여기서 '사업 데이터'는 사업체, 면세 단체, 정부기관에 대한 운영 및 재정 데이터를 의미한다.[98]

나. 인구조사법 (Census Law)

인구조사와 관련해서는 인구조사법(Census Law, Title 13 United States Code)이 규율하고 있다.

이 법은 (인구조사국의 상급기관인) 미 상무부 장관이 이 법에 따른 업무와 관련된 정보를 다른 부처, 기관 등에 요청할 수 있고, 인구조사 및 설문조사의 효율적이고 경제적인 수행을 위해 필요한 기록, 보고 등의 자료 복사본을 주(stats), 시(cities) 등

94) CIPSEA, SEC. 522. DESIGNATION OF STATISTICAL AGENCIES.
95) CIPSEA, SEC. 511
96) CIPSEA, SEC. 512
97) CIPSEA, SEC. 524
98) CIPSEA, SEC. 502

다른 정보 단위 혹은 민간의 개인이나 기관으로부터 획득할 수 있도록 하고 있다. 또한, 가능한 한, 그리고 통계의 종류, 적시성, 질, 범위 등과 일치하는 한도에서, 직접적인 조사대신 앞서 규정한 바에 따른 소스로부터 입수 가능한 정보를 획득하여 사용해야 한다고 규정하고 있다.[99] 즉, 통계 작성에 있어서 직접적인 설문조사보다, 행정데이터 및 민간의 기존 데이터를 활용하도록 하는 것이다.

9조에서는 정보의 기밀성을 규정하고 있는데, 이 법에서 규정한 예외를 제외하고는 누구도 이 법에 따라 수집된 정보를 그것이 제공된 통계 목적 외의 목적으로 사용해서는 안 되고, 특정한 개인 혹은 기관이 식별되는 방식으로 공개되어서는 안 되며, 선서를 한 책임자(sworn officers)나 인구조사국의 직원 등을 제외하고는 누구도 개별 보고서를 검토하도록 허용되어서는 안 된다.[100]

6. 뉴질랜드

뉴질랜드 통계청의 업무는 아래와 같이 통계법 등 여러 법률에 의해 규율을 받는다.

- 통계법(Statistics Act) 1975[101]
- 개인정보보호법(Privacy Act) 1993[102]
- 공공정보법 (Official Information Act) 1982
- 공공기록법 (Public Records Act) 2005
- 뉴질랜드 데이터 및 정보관리 원칙

(1) 개인정보보호법

통계청은 개인정보의 수집, 저장, 이용에 관해 개인정보보호법 1993에서 규정한 프라이버시 원칙을 준수한다. 통계법은 응답자(개인 정보 제공자)에게만 적용되지만, 개인정보보호법은 응답자, 통계청 직원, 연구자 등 통계청 서비스이용자 모두에게 적용된다.

---

99) 13 U.S.C. § 6

100) 13 U.S.C. § 9

101) 뉴질랜드통계법 1975, http://www.legislation.govt.nz/act/public/1975/0001/latest/DLM430705.html (이 법의 최근 개정은 2013년에 이루어졌다.)

102) 개인정보보호법 1993 http://www.legislation.govt.nz/act/public/1993/0028/latest/DLM296639.html

개인정보보호법 제2장(part 2)은 12개의 정보프라이버시 원칙을 규정하고 있다.[103] 통계청의 '프라이버시 및 기밀성 가이드라인'[104]은 12원칙과 관련하여 개인정보호법과 통계법이 어떻게 적용되는지 설명하고 있다.

- 원칙 1 (개인정보 수집의 목적)

합법적인 목적을 위해 필요한 개인정보만이 수집되어야 한다. 통계청은 통계법에 따라 통계 및 연구 목적으로 정보를 수집할 권한을 가지며, 통계 및 연구 목적에 필요한 경우에만 정보를 수집한다.

- 원칙 2 (개인정보의 소스)

개인정보의 수집은 해당 개인에게 직접 수집해야 한다. 다만, 몇 가지 예외 규정이 있는데, 공개적으로 접근 가능한 정보이거나, 정보주체가 다른 사람으로부터의 수집을 허용한 경우 등이다. 또한, 정보가 개인이 식별되지 않는 방식으로 사용되거나, 통계 혹은 연구 목적으로 사용되고 개인을 식별하지 않는 방식으로 공개될 경우도 예외 규정에 포함된다.

뉴질랜드 통계청은 통계 조사를 수행할 때 직접 정보를 수집한다. 장애인이 직접 응답하기 힘든 경우와 같이, 때로는 관련된 사람을 통해 수집할 필요도 있다. 개인정보보호법에 따라 다른 기관으로부터도 정보를 수집하며, 공식통계 생산 및 연구 목적으로 이를 결합한다. 그러나 비식별화 및 기밀성 보장을 통해 기밀 정보가 공개되지 않도록 보호한다. 고용이나 이용자와의 소통 등 다른 맥락에서 기밀 정보를 수집할 경우, 해당 개인 혹은 기관으로부터 직접 수집하거나, 심판관 등 다른 사람으로부터 계약에 따라 수집한다.

- 원칙 3 (정보주체로부터의 수집)

개인정보를 수집하기 전에, 개인정보의 수집 사실, 수집목적, 정보 취득자, 수집·보유자의 이름 및 주소, 관련 법률 및 강제성 여부, 수집 거부 시 결과, 정보주체의 정보 접근권 등 권리 등을 알려야 한다. 정보가 개인이 식별되지 않는 방식으로 사용되거나, 통계 혹은 연구 목적으로 사용되고 개인을 식별하지 않는 방식으로 공개될 경우는 예외이다.

'기본 통계(Tier 1 Statistic)[105] 생산자를 위한 원칙 및 규약'은 사람들에게 수집 정

---

103) A quick tour of the privacy principles
https://www.privacy.org.nz/news-and-publications/guidance-resources/a-quick-tour-of-the-privacy-principles/

104) Privacy and confidentiality guidelines,
http://www.stats.govt.nz/about_us/legisln-policies-protocols/privacy-confidentiality-gdlns.aspx

105) 뉴질랜드는 공식통계 중 가장 중요한 통계를 'Tier 1' 통계로 지정하고 있다. 이는 국내 통계법

보가 어떤 목적으로 이용되는지 알리도록 하고 있다. 애초 수집목적과 관계없는 이용의 세부 내용을 예측하기는 힘들지만, 관련 없는 통계 및 연구 목적으로 이용될 가능성에 관해 설명한다. 또한, 기밀 정보를 수집하는 다른 기관에도 해당 정보가 다른 데이터와 통계 및 연구 목적으로 통합될 수 있음을 사람들에게 알리도록 요청한다.

- 원칙 4 (개인정보 수집 방법)

개인정보는 불법적인 방식으로, 혹은 공정하지 않거나 개인의 사생활을 부당하게 침해하는 방식으로 수집되어서는 안 된다. 통계청은 기밀 정보의 수집이 사람들의 삶을 일정하게 침범하는 것임을 유념하며, 신중하게 접근한다. 정보의 민감성에 대한 사람들의 견해, 특정 그룹의 사람들의 특성들을 고려하며, 정보를 수집할 때 영향을 최소화한다.

- 원칙 5 (개인정보의 저장 및 보안)

개인정보는 유실, 무단 접근·이용·수정·공개, 남용이 없도록 안전하게 보관되어야 한다. 이에 따라 통계청은 안전한 저장을 위한 보안 원칙과 절차를 적용한다. 보관된 정보의 공개(제공)는 승인된 이용자, 허가된 이용으로 제한한다. 직원들도 자신의 직무에 필요한 정보에만 접근할 수 있다.

- 원칙 6 (개인정보에 대한 접근)

정보주체는 (해당 개인정보를 바로 추출할 수 있는) 개인정보 보유자가 보유하고 있는 자신의 개인정보에 대한 접근권을 가진다. 관련하여 개인정보보호법 제4장(part 4)은 개인정보에 대한 접근권을 제한할 수 있는 근거 규정들을 포함하고 있으며, 5장은 개인정보의 접근 및 수정과 관련된 절차 규정을 두고 있다.

- 원칙 7 (개인정보의 수정)

정보주체는 개인정보 보유자가 보유하고 있는 자신의 개인정보에 대해 수정을 요청할 권리를 가진다. 개인정보보호법 7조(예외 규정)는 원칙 7은 통계법에 따라 수집되고, 통계부(Department of Statistics)가 보유한 정보에는 적용되지 않는다고 하고 있다.

정보주체의 접근권에 대한 개인정보보호법 규정이 통계청에 직접 적용되는 것은 아니지만, 통계법은 정부 통계관에 그러한 요청에 동의할 재량을 부여하고 있다. 뉴질랜

제15조에 따른 '지정통계'와 유사하지만, 그 범위가 더 엄격해 보인다. Tier 1 통계는 △ 중요한 의사결정에 필수적이고, △ 공익성이 크고, △ '기본 통계 생산자를 위한 원칙과 규약'에 따라, 공정성 및 통계의 질에 대한 기대를 충족시키며, △ 장기적 데이터 지속성을 요구하고, △ 국제적인 호환성이 있으며, △ 국제 통계 의무를 충족시킨다. 이는 5년마다 갱신되는데, 최근의 목록은 2012년에 지정된 것으로 160여 개가 'Tier 1' 통계로 지정되어 있다. http://www.stats.govt.nz/about_us/who-we-are/home-statisphere/tier-1.aspx

드 통계청의 정책은 가능하다면, 정보주체에게 자기정보에 대한 접근권을 부여하고자 한다. 이것이 투명성과 사람들의 신뢰를 유지할 수 있다고 보기 때문이다. 그러나 현실적으로 데이터 연계와 비식별화로 인해 특정 정보주체의 정보를 찾는 것이 어려울 수도 있다. 그러나 해당 정보를 제공한 원 데이터 보유기관은 보다 쉽게 해당 정보를 추출할 수 있기 때문에 통계청은 정보주체의 요청을 해당 기관에 전달한다. 통계청은 요청자의 신원을 확인해야 하며, 정보가 통계법에 따라 수집된 경우, 열람 요청은 정부 통계관에 의해 허가된다. 요청자 외의 다른 사람에 대한 정보 접근의 요청은 '공공정보법'에 따라 처리한다.

- 원칙 8 (사용 전 개인정보의 정확성 등 점검)

개인정보의 사용 전에 정확성, 최신성, 완전성, 관련성, 그리고 오도될 여지가 없는지를 점검해야 한다. 통계청은 통계법에 따라 수집된 행정데이터가 해당 목적에 맞게 이용될 수 있도록 노력한다. 통계청이 보유하고 있는 다른 기밀 정보는 정확성을 보장하기 위해 주기적으로 검토한다. 보유하고 있는 정보들을 점검하여 최신성이 유지되도록 한다.

- 원칙 9 (필요 이상의 개인정보 보관 금지)

통계청은 장기적인 통계 및 연구적 가치가 있는 경우가 아니면, 애초 목적을 넘어서 기밀 정보를 보유하지 않는다. 보유하고 있는 모든 정보에는 '정보 및 데이터 관리정책'을 적용한다. 정보의 보관, 유지, 처분은 정보관리자가 담당하며, 적절한 승인을 필요로 한다.

- 원칙 10 (개인정보의 이용 제한)

수집된 개인정보는 원칙적으로 애초 수집목적 외로 이용되어서는 안 된다. 다만, 몇 가지 예외가 있는데, 정보가 개인이 식별되지 않는 방식으로 사용되거나, 통계 혹은 연구 목적으로 사용되고 개인을 식별하지 않는 방식으로 공개될 경우도 이러한 예외에 포함된다. 이에 따라, 통계청은 애초 수집목적에서 벗어나더라도 개인정보를 (비식별화될 수 있는 경우) 통계 및 연구 목적으로 이용할 수 있다. 이는 다른 기관으로부터 수집된 정보에도 적용된다. 통계 및 연구 목적 외에는, 해당 개인이나 기관의 허락이 없이는, 애초 수집목적 외로 기밀 정보를 이용하지 않는다. 또한, 수집된 정보를 어떻게 이용하는지에 대한 정보를 제공하려고 노력한다.

- 원칙 11 (개인정보 공개 제한)

개인정보의 공개(제공)가 정보의 취득과 관련된 목적의 하나 혹은 직접 관련된 경우 등 몇 가지 예외를 제외하고는 개인정보는 제3자에게 공개(제공)되지 아니한다. 정보가 개인이 식별되지 않는 방식으로 사용되거나, 통계 혹은 연구 목적으로 사용되

고 개인을 식별하지 않는 방식으로 공개될 경우도 이러한 예외에 포함된다.

통계청은 통계법에 따라 응답자의 동의 없이는 통계청 밖으로 응답자의 정보를 제공하지 않는다. 그러나 정부 통계관의 승인에 따라, 비식별화된 정보를 승인된 연구자에게, 승인된 연구 목적을 위해, 안전한 환경에서 제공할 수 있다.

수집목적에 공개가 포함된 경우가 아니라면, 다른 기밀 정보들을 공개하지 않는다. 예를 들어, 관리자는 직원 관리를 위해 직원 정보를 볼 수 있다. 그러나 그 정보를 통계청 외부에 (해당 직원의 허락 없이) 제공하지는 않는다.

- 원칙 12 (고유식별자)

업무의 효율적인 수행에 필요하지 않으면 개인에게 고유식별자를 부여해서는 안 된다. 고유식별자를 부여한 목적 외로 개인에게 고유식별자의 공개(제공)를 요구해서는 안 된다. 통계청은 IRD 번호(조세번호), NHI 번호(국가건강색인 번호), 은행 고객 번호, 운전면허증 번호, 여권 번호 등 고유식별자의 경우, 종적 연계나 데이터 통합을 위해 암호화된 고유식별자를 이용한다. 뉴질랜드 사업자 번호와 같은 어떤 고유식별자는 다수의 기관에 의해 사용되는 것을 목적으로 하기 때문에 가능하면 이용되어야 한다. 고유식별자는 연구나 분석에 사용되는 데이터와 분리해서 보관하며, 접근을 제한한다. 통계 및 연구 목적의 데이터에서는 모든 개인 식별 정보를 제거하며, 고유식별자는 통계청의 임의의 고유식별자로 대체된다. 직원, 이용자, 다른 기관 등 응답자 외의 사람들을 상대할 때에는, 관련 업무를 효율적으로 수행하기 위해 (예를 들어 직원 번호와 같은) 고유식별자를 부여할 수 있다.

위의 원칙 6(개인정보에 대한 접근)과 원칙 11 (개인정보 공개 제한)은 다른 법령에 의해서 제한되지 않는다. (개인정보보호법 제7조) 다른 원칙들은 다른 법령에 의해 제한될 수 있다. 개인정보보호법 제7장(part 7)은 '공공 등록소 개인정보'에 대한 장인데, 위의 개인정보보호원칙은 이에는 적용되지 않는다.

(2) 통계법

통계법은 정부통계관의 역할, 공식통계의 개념, 설문조사 및 인구조사 수행 절차, 수집된 정보를 안전하게 보관하기 위한 규칙 등을 규정하고 있다. 통계청은 통계법 하의 규제 권한을 가지고 있다. 이 법에 따라 기업 및 가정은 통계청이 요청한 정보를 제공할 법적 의무를 지며, 통계청은 수집된 정보의 기밀성을 보호할 책무가 있다. 통계청 직원 역시 비밀서약서에 서명(통계법 21조)해야 한다.

통계법 37조(정보보안)는 통계청이 수집한 정보의 보호를 다룬다. 이 조항은 해당 정보에 대해 다음 사항을 요구하고 있다.

- 통계 목적으로만 사용되어야 함.

- 개인에 대한 세부 정보 혹은 질의에 대한 응답에는 통계청 직원 외에는 접근하거나 공개되지 않음.

- 통계청이 공개하는 모든 통계정보는 (개인이 동의했거나, 합리적으로 예측할 수 없는 불가피한 경우가 아니면) 세부 정보가 식별될 수 있는 방식으로 공개되어서는 안 됨.

이어, 이에 대한 예외로서, 마이크로데이터에 대한 접근을 허용하는 4개의 조항을 포함하고 있다.

- 제공자의 허락 : 37A조(특정 정보의 공개를 위한 통계관의 허가)는 정부 통계관이 정보를 공개할 수 있는 몇 가지 경우를 규정하고 있다. 이에는 정보 제공자 혹은 특정 사업의 책임자가 서면으로 공개에 동의한 정보, 법이나 공공문서에 의해 공개적으로 접근 가능한 정보가 포함된다. 어떤 경우에는 뉴질랜드 통계청이 통계법에 따라 수집되지 않은 데이터셋을 보유하고 있지만, 이 법의 관련 조항 및 적절한 정책이 이에 적용될 필요가 있다.

- 공동 설문조사(Joint Surveys) : 37B조(공동으로 수집된 정보의 공개)에 의해, 뉴질랜드 통계청과 다른 정부부처, 지방정부 혹은 법정기구에 의해 공동으로 수집된 정보는 각 기관 사이에 공유될 수 있다. 통계청은 어떤 기관에 수집된 정보를 공유하는지 응답자에게 알려야 하며, 응답자가 원하지 않을 경우 이를 존중해야 한다. 데이터의 수집 및 처리를 담당하는 다른 기관의 피고용인은 비밀서약서에 서명해야 한다. 이 조항의 의무는 공동 수집에 관여한 모든 기관에 적용된다.

- 연구 및 통계 목적의 공개 : 통계법 37C조(section 37C, 순수 연구 및 통계 목적을 위한 개인 수준 데이터의 공개)에 의해, 정부 통계관은 일정한 조건 하에 개인 수준의 데이터(즉, 마이크로데이터)를 제공할 수 있다. 그 조건은 △ 오로지 공익을 위한 연구 및 통계 목적으로만 사용될 것, △ 해당 연구자가 필요한 연구 경험, 지식, 기술을 가지고 있을 것, △ 이름 및 주소는 삭제할 것, △ 해당 연구 및 프로젝트에 관여하고 있는 모든 사람이 비밀서약서에 서명할 것, △ 정보가 항상 안전하게 보관되어야 할 것 등이다. 공개된 연구 결과는 통계청이 공개할 수 있는 것 이상의 정보를 포함해서는 안 된다. 모든 관련 연구자들은 해당 정보를 연구 및 통계 목적으로만 사용해야 하며, 정부 통계관이 지시에 따라야 한다.

- 역사적 문서 : 37D조에 의해, 정부 통계관이 역사적 문서로 분류한 경우에, 그는 100년 이후에 개인 수준의 데이터 공개를 허가할 수 있다.

(3) 공공정보법

공식통계에 대한 접근은 공공정보법의 대상이 된다. 일반적인 원칙은 정보는 요청에 의해 접근할 수 있어야 한다는 것이다. 공식통계는 공개되기 때문에 일반 대중에게 접근 가능하다. 이 법은 방법과 관행이 통계 이용자에게 제공될 것을 요구하지만, 개인이나 기관에 대한 정보가 공개되어야 하는 것은 아니다.

(4) 공공기록법

통계기록의 생성, 유지, 폐기는 공공기록법에 따른다. 최고 기록관(The Chief Archivist)이 공공 기록물의 폐기를 허가할 권한을 가진다.

제2절 보건의료 분야 데이터 연계 현황

1. 영국

(1) 개요

영국의 보건의료 시스템은 국가보건서비스(NHS)를 통한 공공 영역이 주로 담당하고 있다. 따라서 방대한 보건의료 데이터가 NHS에 의해 수집, 관리되고 있고, OECD 국가 중 가장 광범한 국가적 보건의료 데이터셋을 보유하고 있는 나라이다. 영국에서 일상적으로 수집되는 데이터는 1차 진료, 2차 진료, 사회복지, 처방, 환자 경험, 공공보건, 수많은 특정 질병 영역에 대한 국가적 감사(national audit)를 포함한다.

그러나 영국에서의 데이터 수집 및 접근 체제는 잉글랜드, 웨일즈, 스코틀랜드, 북아일랜드 등 영국 내 각 지역별로 조직이 되어 있다. 잉글랜드의 경우, NHS Digital (구 HSCIC)이 보건복지 데이터를 위한 국가적인 수집·제공자이다. NHS Digital은 광범한 병원진료통계(Hospital Episode Statistics, HES) 데이터를 보유하고 있는데, 병원의 비용이 NHS에 제출되는 정보에 기반을 두고 있기 때문이다. CPRD는 잉글랜드의 관찰 및 개입 연구를 위한 서비스로 익명화된 1차 진료 데이터를 연구 목적으로 제공한다. NHS Digital이 CPRD를 위한 연계 서비스를 제공한다.

웨일즈에서는 SAIL Databank가 일상적으로 수집되는 국가적 보건 데이터의 중앙 접속점의 역할을 한다. NHS Wales, 국가통계청(ONS) 등으로부터 데이터를 제공받는다. 스코틀랜드에서는 NHS National Service Scotland(NSS)를 대신해서, 정보서비스부(Information Service Division, ISD Scotland)가 광범한 건강 관련 행정 데이터를 수집한다. eDRIS는 연구 목적의 보건의료 데이터 접근 및 연계를 제공하기 위한 ISD의 서비스이다. 북아일랜드는 북아일랜드 보건복지(HSCNI)가 보건 서비스의 정보 허브이다. 데이터는 지역 데이터웨어하우스를 통해 수집되며, Honest Broker Service를 통해 접근이 관리된다.

영국에서는 고유한 환자 식별자가 있기 때문에 데이터 연계가 용이하다. 고유 환자 식별자는 영국 전역에 걸쳐서 NHS의 정보 시스템에 활용된다. 잉글랜드와 웨일즈에서 사용되는 환자 식별자는 NHS 번호(NHS number)인데, 이는 10개의 숫자로 된 코드로 전자건강기록을 관리하는 데 사용된다. 스코틀랜드는 공동체보건인덱스(Community Health Index, CHI)를, 북아일랜드는 보건복지번호(Health and Care Number, HCN)를 사용하는데, 이 두 번호 모두 NHS 번호와 포맷이 같으며, 다만 중복을 피하기 위해 10개 숫자의 부여 범위만 다르다. 이 번호들은 보건의료 서비스 제

공을 위해서만 사용되며, 다른 공공서비스나 조세와 연결되지는 않는다. (OHE Consulting Report, 2015)

영국 보건부는 2012년 5월, “정보의 힘 : 우리가 필요한 보건 정보를 우리 모두가 통제할 수 있도록 함(The power of information: Putting all of us in control of the health and care information we need)”이라는 전략을 발표하였다.[106] 이 보고서는 진료 및 환자 수준 데이터를 한번 수집하면, 프라이버시 보호 규정에 따라 사용, 재사용, 공유될 수 있는 정보 시스템에 대한 10년의 비젼을 수립하였다. 그 목적은 NHS 및 사회복지 서비스 내의 정보공유를 통해 서비스를 지원하고, 연구 및 통계 목적으로 데이터를 제공하며, 환자들이 자신의 기록에 접근할 수 있도록 하는 것이다. 이 전략은 연구 목적의 데이터 연계도 포함하고 있다. 이 전략으로부터 영국 사회에서 논란이 되었던 care.data 프로그램이 나왔는데, 이에 대해서는 후술하기로 한다. (OECD, 2015)

(2) 잉글랜드 CPRD[107]

영국의 임상시험연구데이터링크(Clinical Practice Research Datalink, CPRD)는 비영리 연구 지원을 목적으로 하는 복지부(Department of Health) 산하 기관이다. 영국 NHS 국가보건연구소(NIHR)과 의약보건제품규제기구(MHRA)가 공동으로 재정을 지원한다.

CPRD는 1987년부터 공공보건 연구를 위해 익명화된 1차 진료 기록을 제공해왔다. 현재 CPRD 데이터를 이용한 연구 출판물은 1700여 개에 달한다. 또한, CPRD는 1차 진료 데이터를 임상시험에도 사용하고 있다. 이와 같은 e-health 연구 서비스는 MHRA의 일반의연구데이터베이스(General Practice Research Database, GPRD)와 보건부의 NIHR 연구역량프로그램(Research Capability Programme, RCP)을 결합한 것이다.[108]

CPRD는 법적 계약이나 승인에 따라, 영국 및 전 세계의 학계, 의약/바이오기술/기기 및 연구전문조직(Contract Research Organization, CRO)의 연구자들에게 다음과 같은 3가지 핵심적인 서비스를 제공한다.

---

106) https://data.gov.uk/sites/default/files/DH%20Open%20Data%20Strategy_10.pdf

107) 이하 내용은 CPRD 홈페이지(Clinical Practice Research Datalink)를 참고한 것이다.

108) CPRD는 그 이전의 GPRD(General Practice Research Database)를 기반으로 2012년 4월에 설립되었다. GPRD는 1차 진료 데이터베이스로 의약및보건제품규제기구(MHRA)에 의해 운영이 되어 왔으며, 현재는 CPRD의 일부가 되었다.
https://en.wikipedia.org/wiki/Clinical_Practice_Research_Datalink

- 임상시험, 바이오 샘플 수집, 치료결과(Patient Reported Outcomes, PROs)를 위한 중재서비스 및 IT 시스템
- 연구 서비스 : 약물 역학, 약물 경제학, 결과 및 위험편익
- 관찰 데이터 : NHS 및 다른 보건 관련 데이터와 연계된 데이터에의 접근 (적절하게 익명화된 데이터)

가. 데이터 연계

1차 진료 데이터를 2차 진료나 질병 등록부와 같은 다른 보건 데이터셋과 연계하는 것은 데이터를 통해 해결할 수 있는 의료 연구의 폭을 크게 넓힐 수 있다. CPRD는 오랫동안 다른 데이터셋과 연계된 1차 진료 데이터를 제공해왔다. 데이터 연계는 다음과 같은 절차에 의해서 이루어진다.

① 매년 CPRD는 공공보건연구 목적으로 익명화된 연계 데이터를 제공하는 것에 대해 보건연구당국의 Section 251 규제 승인을 받아야 한다.[109)]

② 데이터 연계는, 환자식별정보를 합법적으로 수집할 권한을 가진 잉글랜드의 법정기구인 NHS Digital에 의해 이루어진다. 즉, NHS Digital 이 '신뢰할 수 있는 제3자(TTP)'의 역할을 수행한다.

③ 연계를 위해, 일반의가 수집한 환자식별정보(NHS 번호, 생년월일, 우편번호, 성별)와 다른 데이터셋의 식별 정보가 NHS Digital로 보내진다.

④ NHS Digital은 두 데이터셋에서 환자식별정보를 매칭하여, 환자식별정보를 포함하지 않은 암호화된 연계키(encrypted linker key)를 산출한다.

⑤ NHS Digital은 CPRD가 비식별 데이터셋을 연계할 수 있도록 암호화된 연계키를 CPRD에 보낸다.

⑥ CPRD는 일반의나 NHS Digital로부터 절대 환자식별정보는 받지 않는다.

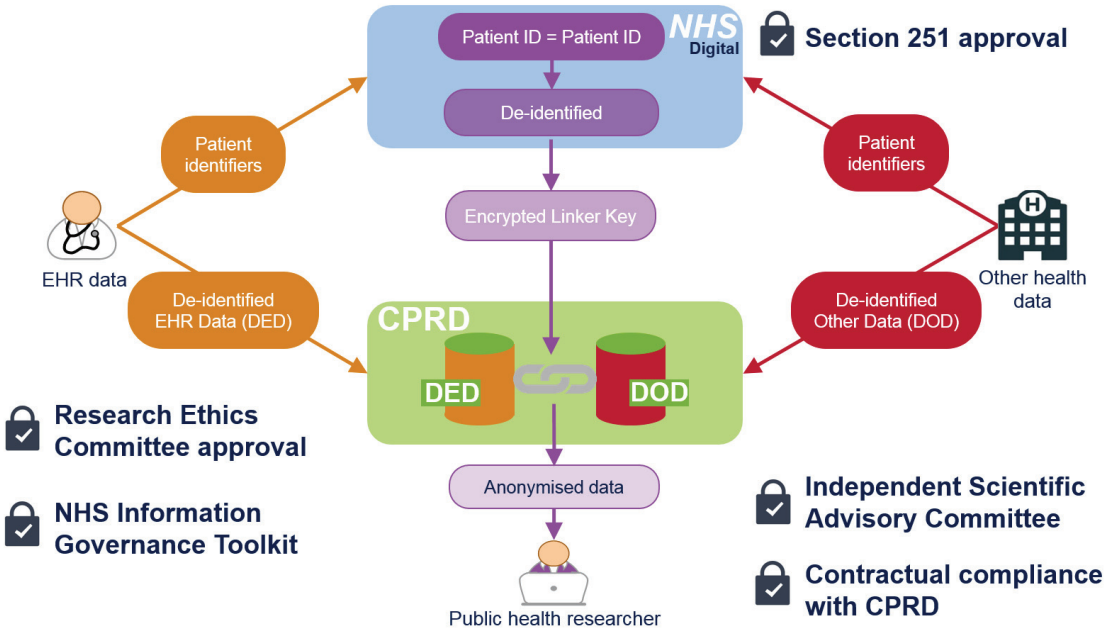
⑦ 공공보건연구 목적으로 연계 데이터에 접근하고자 하는 연구자는 독립적 과학자문위원회(Independent Scientific Advisory Committee, ISAC)의 승인을 받아야 한다.

⑧ ISAC의 승인에 따라 연구자에게 익명화된 데이터셋으로 제공하기 이전에 추가적으로 암호화한다.

---

109) 보건및사회복지법(Health and Social Care Act) 2001의 Section 60이 재입법된 NHS법 2006의 Section 251은 보건부 장관(Secretary of State for Health)이 특정한 의료적 목적으로 보통법 상 기밀유지의무의 예외를 둘 수 있는 규정이다. 이 규정을 보건서비스(환자정보통제)규정 2002라고 부른다. 'section 251 지원 혹은 승인'은 이 규정에 따른 권한 하에 주어지는 승인을 의미한다. NHS의 보건연구당국(HRA)는 2013년 4월에 기밀성자문그룹(Confidentiality Advisory Group, CAG)을 설립하며, Section 251에 대한 책임을 갖게 되었다.
http://www.hra.nhs.uk/about-the-hra/our-committees/section-251/what-is-section-251/

그림 3-1 CPRD 데이터 연계 절차



* 출처: CPRD 홈페이지 https://www.cprd.com/safeguardingpatientdata/

일반의는 CPRD에 환자 데이터 제공을 동의할 때, 이 절차에 따라 환자 데이터가 연계될 수 있다는 것에도 동의해야 한다. CPRD는 잉글랜드의 데이터만 연계할 수 있다. 스코틀랜드, 웨일즈, 북아일랜드의 일반의들은 CPRD가 비식별 환자 데이터를 추출할 수 있다는 것에 동의할 필요가 있다.

나. 연구 데이터의 신청과 승인

가) 연구계획서의 작성

CPRD 데이터에 대한 접근 및 요청을 하고자 하는 연구자는 독립적 과학자문위원회(ISAC) 에 연구계획서를 제출해야 한다. 연구계획서를 쓰기 전에 ISAC 사무국과 연구의 실행 가능성에 대해 협의할 것이 권고된다. 연구계획서는 연구팀의 경험과 전문성에 관한 내용도 포함된다. 또한, 연구를 책임지며, 신청서에 서명할 선임연구원급의 연구책임자와 연락담당자를 포함하며, 신청자의 이력서를 제출해야 한다.

나) ISAC에 의한 연구계획서 승인

ISAC는 2006년 보건부 장관에 의해 설립된 비법정 전문가 자문기구이며, CPRD에

의해 제공되는 데이터에 대한 접근 요청과 관련한 자문을 제공한다. 2015년까지 복지부의 임명위원회(Appointments Commission)가 위원들을 임명했으며, 현재는 MHRA가 직접 임명한다.

ISAC은 다음과 같은 역할을 한다.

① 위원회는 CPRD의 익명화된 환자-수준 데이터의 사용을 요청하는 연구의 현실성, 질, 보건 의료적 가치에 대해 MHRA에 자문을 제공한다.

② CPRD 데이터에 대한 접근 및 이용을 요청하는 연구계획서의 학술적 (의학적, 역학적, 방법론적) 가치에 대해 적시에, 높은 수준의 동료 평가를 제공한다.

③ 기밀성 자문그룹 혹은 연구윤리위원회의 의견을 고려하여, 연구 과정에서 CPRD 데이터의 접근 중에 발생할 수 있는 중요한 윤리적 이슈와 기밀성 이슈들을 강조한다.

④ CPRD 데이터의 접근 및 이용을 요청하는 연구계획서의 작성과 관련한 가이드의 과학적 내용에 자문하고 기여한다.

⑤ 일관성, 효율성, 높은 수준의 동료 평가를 보장하기 위해 위원회의 내부 활동을 평가한다.

⑥ MHRA와 CPRD가 요청한 특정 이슈에 대한 자문을 제공한다.

연구 목적으로 익명화된 환자 수준 데이터에 대한 접근을 요청하는 경우 ISAC의 승인이 필요하다. 이는 영국 1차 진료 데이터와 병원진료통계(Hospital Episode Statistics, HES)와 같은 연계 데이터를 포함한다. 공개를 목적으로 총계 데이터 생성을 위해 환자 수준 데이터를 이용하는 경우에도 승인을 받아야 한다.

보건연구당국(HRA)에 의해 설립된 기밀성자문그룹(Confidentiality Advisory Group, CAG)은 환자 기밀성과 관련하여 ISAC을 자문한다. CAG의 지침에 따라, ISAC은 CPRD와 다른 데이터 소스의 연계가 필요한 승인된 연구계획서의 요약본을 공개해야 한다. 또한, CPRD를 통해 얻을 수 있는 데이터 소스의 특정한 연계를 요청하는 연구자의 경우에도 ISAC의 승인이 필요하다. 이와 같은 요청은 제안된 연구 및 프로그램의 맥락 안에 있어야 한다. ISAC 승인은 추가적인 환자 정보(예를 들어, 병원의 서류 기록을 포함한 익명화된 환자 기록, 일반의나 환자에 대한 설문조사 등)에 대한 접근을 얻기 위해서도 필요하다.

그러나 다음의 경우에는 ISAC의 승인이 면제된다. 첫째, 공개를 목적으로 하지 않는, 총계 데이터 생성을 위한 환자 수준 데이터의 이용. 여기서 공개(publication)란 데이터 요청 기관 외의 제3자에게 유통하는 것을 의미한다. 둘째, 규제 제안/요청을

지원하기 위한 간단한 설명 정보 생산을 위한 환자 수준 데이터 이용.

다) CPRD의 지원

CPRD는 역학 및 통계학에 전문성을 가진 내부 연구팀이 있으며, 연구자에게 연구 계획서의 작성, 연구 승인의 획득, 데이터 추출 및 분석, 보고서 작성 및 출판 등에 관한 지원을 제공한다.

다. 데이터에 대한 접근

CPRD가 보유한 1차 진료 데이터베이스를 GOLD라고 부른다. GOLD는 등록된 1차 진료 기관의 익명화된, 종적 진료 데이터를 보유하고 있다. CPRD GOLD는 라이선스를 가진 연구자가 ISAC이 승인한 연구를 수행할 수 있도록 온라인 접근을 제공하고 있다. CPRD GOLD는 질병 및 의약품 코드 사전 및 환자 코호트를 정의할 수 있는 질의 도구를 제공한다. 추출 도구는 코호트나 통제 그룹별로 데이터를 추출할 수 있도록 한다.

CPRD 데이터에 대한 접근을 위해서는 라이선스 계약을 맺어야 하는데, 이는 상세한 이용조건을 규정하고 있다. 그 핵심적인 내용은 다음과 같다.

CPRD의 데이터 접근 라이선스는 양도할 수 없다.

데이터의 제공은 공익적인 의료 연구 목적으로 제한된다.

다른 사람에게 공개되는 모든 연구에는 ISAC 승인이 필요하다.

데이터셋은 승인된 연구로만 사용되어야 한다.

환자, 병원, 의사에 대한 식별 시도는 특히 금지된다.

CRPD는 데이터에 접근하는 기관을 감사(audit)할 권한을 가진다.

다운로드된 데이터의 무단 이용 및 복제를 방지할 수 있는 적절한 보안 조치를 취해야 한다.

라. 개인정보 보호

CPRD는 일반의(GP)로부터 수집하는 환자들의 개인정보 보호를 위해서 다음과 같이 다양한 조치를 취하고 있다.

- 일반의나 다른 소스로부터 수집하는 데이터에 환자들의 식별정보는 포함되지 않

는다. 특히 NHS 번호, 이름, 전체 생년월일, 주소, 의사들의 진료메모(free text medical notes) 등은 수집하지 않는다.

- 단지 비식별화되고 암호화된 환자 데이터만이 CPRD로 전송된다.
- CPRD는 공공보건연구를 위한 환자 데이터의 수집과 제공을 위해, 매년 윤리 승인(ethics approval)을 받는다.[110]
- 일반의는 환자들의 비식별 데이터를 CPRD에 제공할지 여부를 선택할 수 있다.
- 개인 환자가 원하지 않을 경우 CPRD에의 개인정보 제공을 거부할 수 있다. (opt-out)
- CPRD가 보유한 데이터에 접근하기 위한 연구 요청은 ISAC의 심의를 받는다.
- 공공보건 연구를 수행하는 실제 연구자만이 데이터에 접근할 수 있다.
- ISAC의 승인에 따라, 데이터가 익명화된 데이터셋으로 연구자에게 제공되기 전에 추가적으로 암호화된다.
- 연구자가 데이터 사용에 대한 이용조건을 준수하도록 계약을 통해 강제한다.

(3) 웨일즈 SAIL Databank[111]

보안익명정보연계(Secure Anonymised Information Linkage) 데이터뱅크(SAIL Databank)는 영국 웨일즈에 기반하고 있으며, 보건 관련 연구를 위해 익명화된 개인 기반의 데이터의 저장 및 이용을 제공한다. 웨일즈 정부의 웨일즈보건의료연구(Health and Care Research Wales)의 재정 지원을 받고 있으며, 스완지 의과대학의 보건정보연구단(Health Informatics Research Unit, HIRU)이 관리하고 있다.

2006년에 웨일즈보건의료연구의 재정 지원으로 HIRU가 설립되었고, 2007년에 서부 웨일즈 스완지 지역에서 SAIL Databank 시범사업이 시작되었다. 2011년에는 원격접근을 위한 SAIL Gateway가 개발되었다. 2015년에는 eHealth와 행정데이터 연구, 훈련 및 개발 센터로서 데이터과학빌딩(Data Science Building, DSB)이 문을 열었는데, 여기에 Farr Institute와 ADRC Wales가 들어왔다.

SAIL Databank는 국가통계청(ONS), 연례 출생 및 사망정보, NHS 웨일즈 정보서비스, 응급부서 데이터셋, 국가공동체 아동보건 데이터베이스, 외래환자 데이터셋, 웨

110) 보건연구당국(Health Research Authority, HRA) 승인은 거버넌스와 법적 준수를 평가하는 NHS 잉글랜드의 절차로서, HRA 직원이 독립적인 연구윤리위원회의 의견을 받아 수행한다. http://www.hra.nhs.uk/research-community/hra-approval-the-new-process-for-the-nhs-in-england/

111) 이하 내용은 SAIL Databank 홈페이지를 참고한 것이다. https://saildatabank.com/

일즈 환자치료 데이터베이스(PEDW), 웨일즈 인구정보 서비스, 웨일즈 공공보건(Public Health Wales), 웨일즈 암정보감시단(WCISU), 선천성기형 등록정보서비스, 일반의 1차 진료기록 등의 데이터를 보유하고 있다. 그러나 SAIL Databank는 개인 식별정보를 보유하고 있지 않으며 익명화된 데이터만 보유하고 있다. 이 과정에서 NHS 웨일즈 정보서비스(NWIS)가 신뢰할 수 있는 제3자(TTP)의 역할을 한다. SAIL Databank의 데이터셋은 데이터 보유기관의 추가적인 허가가 필요한 데이터셋과 그렇지 않은 데이터셋으로 나누어진다.

SAIL Databank는 연구자의 요청에 따라 데이터 연계를 제공하고 있다. 또한, 다른 기관이나 연구자들이 SAIL Databank에 자신이 보유하고 있는 데이터를 제공할 수 있도록 하고 있다.

가. 데이터 연계 및 제공 절차

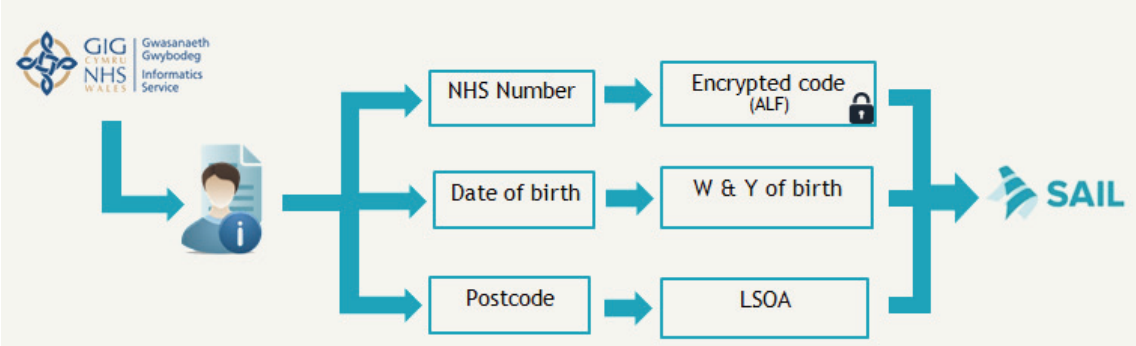
① 1단계 : 데이터 제공기관에 대한 지원

의료 서비스를 제공하는 데이터 제공기관들은 환자들의 개인정보 및 의료 데이터를 보유하고 있다. SAIL Databank는 데이터 제공기관들이 충분한 정보에 기반하여 데이터 제공 여부를 결정할 수 있도록 지원한다. 데이터 제공이 승인되면, SAIL Databank는 정보들이 익명화된 형태로 전송될 수 있도록 기술적인 지원을 제공한다.

② 2단계 : 데이터 익명화

데이터 제공기관은 자신의 데이터를 인구정보 부분(demographic component)과 콘텐츠로 분리한다. 이름, 주소, 성별, 생년월일, NHS 번호 등을 포함한 인구정보는 NWIS로 보내진다. 데이터셋에 NHS 번호가 없을 경우, NWIS는 웨일즈 인구정보서비스에서 찾아 추가한다. NHS 번호는 고유한 비식별 코드의 생성에 사용되는데, 이를 익명연계필드(Anonymous Linking Field , ALF)라고 부른다. ALF는 데이터셋의 각 개인에게 부여된다. ALF와 출생 주(week of birth), 성별, 거주지역(LSOA) 등 최소한의 인구정보가 SAIL Databank에 보내진다. 보건 관련 기록을 담고 있는 콘텐츠(예를 들어, 입원 기관, 처방 약, 시험 결과 등)는 SAIL Databank로 직접 보내진다.

그림 3-2 NWIS와 데이터 익명화

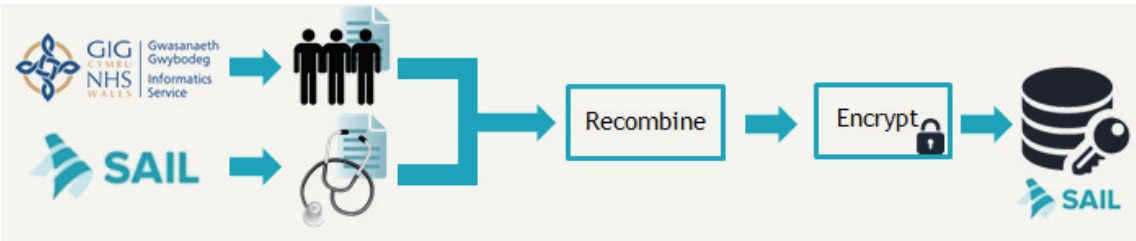


* 출처: SAIL DATABANK 홈페이지 https://saildatabank.com/faq/

③ 3단계 : 데이터 연계

데이터 보유기관이 인구정보와 콘텐츠 정보로 데이터셋을 분리할 때, 그들은 '결합키(join key)'를 각 분리된 데이터셋에 부여한다. 이 키는 그 자체로 아무런 의미도 없으며, SAIL Databank에서 두 부분을 다시 결합하기 위한 용도로 사용된다. ALF는 이미 암호화된 코드이지만, SAIL Databank는 추가적인 보호조치로 이를 다시 암호화한다. 이에 따라 SAIL Databank와 NWIS, 양자 모두 환자 식별자를 복호화할 수 없다. ALF는 서로 다른 익명 데이터셋을 연계하는 데 사용된다.

그림 3-3 SAIL Databank의 데이터 연계

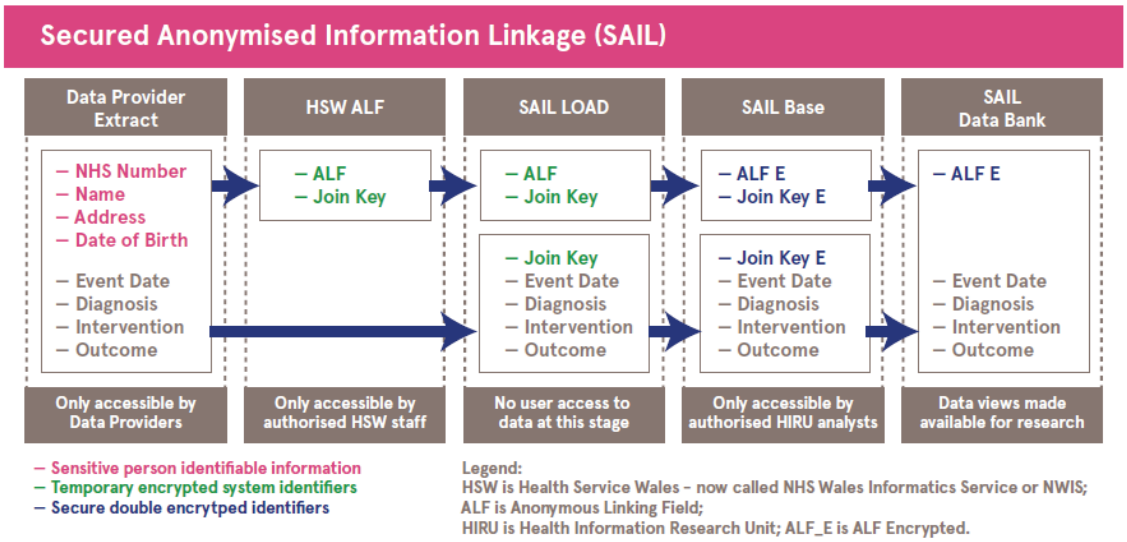


* 출처: SAIL DATABANK 홈페이지 https://saildatabank.com/faq/

④ 4단계 : 연구자에의 제공

지금까지 설명한 데이터 연계절차를 데이터 흐름에 따라 그림으로 나타내면 아래 그림과 같다. 데이터 연계 과정에서 ALF와 Join Key 두 개의 임의의 키가 형성되는데, ALF는 서로 다른 데이터셋의 연계를 위해 사용되며, Join Key는 두 개로 나누어진 인구정보(식별정보)와 콘텐츠를 다시 결합하는 데 사용된다.

그림 3-4 SAIL 데이터 연계 절차



* 출처: Rosalyn Moran (2016)

나. 연구 제안서의 신청과 승인

연구 제안서의 제출은 2단계로 이루어진다. 우선, 연구자들은 제안서 초안과 연구 범위 문서(Initial application and scoping document)를 통해 SAIL Databank의 분석가와 프로젝트에 대해 협의를 하게 된다. 프로젝트의 아이디어와 실현 가능성, SAIL Databank 관련 연구 활동, 연구 스케줄, 자원의 이용과 비용 등에 관해 협의한다.

SAIL Databank는 공익에 기여할 수 있는 순수 연구 목적으로만 데이터에 대한 접근을 허용하고 있다. 데이터에 접근하기 위해서는 독립적인 정보거버넌스검토패널(Information Governance Review Panel, IGRP)의 승인을 받아야 한다. IGRP는 영국의학협회, 웨일즈 공공보건, 국가연구윤리서비스, NHS 웨일즈 정보서비스, 데이터 연계 연구를 위한 소비자패널의 대표자들로 구성된다. IGRP는 SAIL Databank 데이터의 적절한 사용을 보장하기 위해 각 프로젝트를 신중하게 검토한다.

연구 제안서가 승인되면 모든 연구자는 데이터에 접근하기 전에 적절한 정보 거버넌스에 대한 훈련을 받아야 한다. 현재 SAIL Databank는 의학연구위원회(Medical Research Council, MRC)의 규제지원센터의 '연구 데이터 및 기밀성 원격교육'과 ADRN의 'SURE training'을 훈련 과정으로 인정하고 있다.[112] 유사한 주제를 다루는 다른 훈련 프로그램을 수행한 경우, 신청서에 그 과정의 내용을 기재하도록 한다. 정

112) SURE 훈련은 안전한 연구 데이터 이용자 환경 훈련(Safe Users of Research data Environment Training)의 약자이다.

보 거버넌스 훈련은 현재 통용되는 것이어야 하며, 2년마다 업데이트되어야 한다. 연구 기간 중에 훈련 자격증이 만료될 경우 이를 갱신해야 한다.

다. 데이터에 대한 접근 - SAIL Gateway

훈련이 끝난 후 원격접근을 원하는 연구자들은 SAIL Gateway 사용 승인을 받아야 한다. SAIL Gateway는 프라이버시 보호를 위한 안전시설(safe havens)이자 원격접근 시스템이다. 이를 통해 연구가 보안이 되는 안전한 환경에서 수행되도록 한다. 연계 데이터를 바로 제공하기보다는 SAIL Gateway를 통해 접근하도록 하는데, 이는 혹시 있을 수 있는 연계 공격(linkage attack)에 대비하기 위한 것이다.

연구자들은 SAIL Gateway를 통해서 언제든, 전 세계 어디서든 윈도 데스크탑 환경에서 원격 접근하여, 사전에 설치된 통계 패키지를 통해 분석을 수행할 수 있다. SAIL Gateway에 접근하기 위해 연구자들은 SAIL 데이터 접근 계약을 체결해야 한다. 이는 모든 SAIL 데이터 이용자가 준수해야 하는 사용 조건을 규정하고 있다. 또한, 이용자 이력서 복사본과 함께 Gateway 계정 요청 폼을 작성해야 한다. 이후 연구자가 선택한 데스크탑 컴퓨터를 통해 SAIL Gateway에 로그인할 수 있는 계정 정보를 받게 된다. 보안원격접근은 VPN을 통해 이루어지며, 연구자가 요청한 특정한 데이터는 '읽기전용'으로 접근할 수 있다.

또한, 다음과 같이 추가적인 보안 조치가 취해진다. 우선, SAIL Databank에의 모든 접근은 세세하게 모니터 된다. 모든 데이터가 익명화된 상태이지만, SAIL Databank는 추가적인 보호조치를 취하기도 한다. 예를 들어, 드문 질병 사례에 관한 연구의 경우, 단지 몇몇 사례밖에 없기 때문에 개별 기록보다는 그룹으로 총계 처리된 형태(예컨대 연령 그룹화)로 제공된다. 프로젝트 후에 데이터셋은 아카이브로 기록된다.

연구가 완료된 후, 연구자들은 SAIL 데이터 관리자의 검토를 받은 다음에야 자신의 연구결과물을 SAIL Gateway로부터 가지고 나갈 수 있다. SAIL 데이터 관리자는 연구결과물이 개인정보 침해의 위험성이 없는지 평가한다.

라. 외부 데이터의 연계

승인을 받으면, SAIL Databank의 데이터와 외부의 데이터를 연계할 수 있다. 이를 원하는 연구자들은 우선 데이터 접근 요청폼(Data Acquisition Request Form)과 신규 데이터셋 검사폼(New Dataset Scoping Form)을 작성해야 한다. 후자는 필수적인 것은 아니지만, SAIL Databank에 업로드할 데이터셋의 메타 데이터를 포함하기 위해 작성을 권고하고 있다.

SAIL 데이터관리위원회는 이 요청이 운영정책과 일치하는지 검토한 후, SAIL 데이터 공유 계약을 보내준다. SAIL 데이터 공유 계약에 서명하면, 연구자는 데이터를 SAIL Databank에 업로드할 수 있게 된다.

업로드하는 절차는 다음과 같다.

① 자신이 보유하고 있는 데이터셋을 개인정보를 포함하고 있는 개인식별 데이터(File1)와 콘텐츠 데이터, 즉 임상/행정데이터(File2)로 분리한다. 두 개의 파일은 이후 재결합을 통해 익명화된 데이터셋을 만들기 위해, 고유한 레코드ID 번호를 가지고 있어야 한다.

② NWIS와 SAIL Databank에 접근할 수 있는 두 개의 계정을 만들어야 한다.

③ 두 계정에 로그인해서 각각 파일을 업로드한다.

마. 개인정보 보호 및 보안

SAIL Databank는 2007년 시스템 구축 단계에서부터 '프라이버시 중심설계(privacy by design)'에 기반하여 견고한 거버넌스 모델을 구축해왔다. 이는 다음과 같은 원칙에 기반하고 있다.

- 안전한 데이터 전송
- 신뢰할 수 있는 레코드 매칭
- 식별 데이터의 익명화와 암호화
- 노출 제어 (disclosure control)
- 데이터 접근 통제
- 제안서와 결과에 대한 면밀한 검토
- 정보 거버넌스에 대한 외부 인증

SAIL Databank는 암호화된 익명연계필드(ALF)를 통해 데이터 연계를 지원하지만, 개인식별 정보는 보유하고 있지 않다. NWIS가 SAIL Databank를 위한 신뢰할 수 있는 제3자(TTP) 역할을 한다. 데이터에 접근하기 위해서는 IGRP의 승인을 받아야 하며, SAIL Gateway라는 보안 원격접근시스템을 통해서만 접근할 수 있다. 연구결과물을 가지고 나가기 전에는 SAIL 데이터 관리자의 검토를 받아야 한다. 또한, 보안 평가를 위해, 2015년 10월 SAIL Databank는 ISO 27001 정보보안관리시스템(ISMS) 인증을 받았다.

(4) 스코틀랜드 eDRIS[113]

전자데이터연구혁신서비스(The electronic Data Research and Innovation Service, eDRIS)는 NHS 스코틀랜드 국가서비스(National Service Scotland, NSS) 산하 스코틀랜드 정보서비스부(Information Services Division, ISD)가 제공하는 서비스의 하나로서, 연구자들이 안전한 환경에서 데이터에 대한 신청, 승인, 접근을 할 수 있도록 지원하는 단일 창구의 기능을 하고 있다. eDRIS는 Farr Institute Scotland, 행정데이터연구센터(ADRC), 스코틀랜드 정부의 연계 프로젝트를 지원한다.

eDRIS가 제공하는 서비스는 다음과 같다.

- 전 과정에 걸친 연구 코디네이터 지정
- 연구 설계 지원
- 코딩, 용어, 메타 데이터, 연구의 실현 가능성에 대한 전문가 자문
- 데이터 접근을 위해 필요한 허가의 획득 지원
- 연구결과물 및 타임라인에 대한 협의
- 데이터 확보를 위한 데이터 제공기관과의 연결
- 스코틀랜드 국가서비스(NSS)의 안전시설(Safe haven) 내에서의 데이터 접근
- (필요할 경우) 데이터에 대한 분석, 해석, 정보제공

eDRIS는 연구 전 과정에 걸쳐서 연구자를 지원하는 '연구 코디네이터(Research Coordinator)를 제공하고 있다. 연구 코디네이터는 eDRIS의 절차, 활용 가능한 데이터 선정, 승인 신청, 안전시설에의 접근, 최종결과물의 공개까지 모든 과정에 있어서 연구자를 지원한다.

ISD는 스코틀랜드의 500만 명 이상의 NHS 보건 데이터 및 보건 관련 데이터를 보유하고 있으며, 출생 전부터 엄마의 출산 전 기록, 사망 등록부까지 개인의 전체 생애에 대한 데이터도 일부 가지고 있다. 홈페이지를 통해 제공하는 국가데이터목록(National Data Catalogue)에서 어떤 데이터를 활용할 수 있는지 찾을 수 있다.[114] 또한, ISD는 행정데이터연구네트워크(ADRN)에 주요 보건 데이터셋에 대한 정보를 제

113) 이하 내용은 ISD Scotland의 eDRIS 홈페이지를 참조한 것이다.
http://www.isdscotland.org/Products-and-Services/eDRIS/

114) http://www.ndc.scot.nhs.uk/National-Datasets/Full-A-Z/index.asp

공하고 있다.

가. 연구자에 대한 승인 절차

연구자로서 승인을 받기 위해서는 다음과 같은 5가지 요건이 필요하다.

- 모든 이용자는 적절한 정보 거버넌스 훈련을 받았음을 입증해야 한다.
- 연구가 적절한 승인을 받아야 하며, 데이터 공유계약을 체결해야 한다. 모든 연구는 데이터 보유기관의 허락을 받아야 하며, 필요할 경우 윤리 승인을 획득해야 한다.
- NHS 기밀성 실행규약(NHS Confidentiality Code of Practice)을 읽어야 한다.[115]
- 승인된 기관(Approved Organization) 소속이어야 한다.
- 스코틀랜드 국가서비스(NSS)의 안전시설을 이용할 경우, eDRIS 이용자 계약에 서명해야 한다. 이 문서는 안전시설 이용과 관련된 모든 요건을 서술하고 있다. 이는 이용자를 고용하거나 지원하는 기관의 권한 있는 사람이 서명해야 한다.

연구자가 받는 훈련은 앞서 설명한 SAIL Databank 에서와 같이 의학연구위원회(MRC) 규제지원센터의 연구 데이터 및 기밀성 원격교육, ADRN의 '안전한 연구 데이터 이용자 환경 훈련(SURE Training)', 그리고 보건 및 사회보장 정보센터의 정보 거버넌스 훈련도구(Health and Social Care Information Centre Information Governance Training Tool) 등을 통해 수행할 수 있다. 또한, 3년마다 이 훈련은 업데이트되어야 하며, 연구 기간 중에 훈련 자격증이 만료된 경우에는 다시 받아야 한다.

승인된 기관(Approved Organization)은 현재 대학, NHS, 지역 당국 및 스코틀랜드 정부 등 공공영역의 기관으로 제한되어 있다. 공익목적의 연구를 위한 데이터 연계를 지지하면서도 상업적 접근에 대한 대중적 우려가 크기 때문에, '승인된 기관'의 연구자들만 데이터에 접근하도록 한 것이다. 영리 업체, 미디어, 로비 그룹, 제3섹터 기관들은 데이터에 직접 접근할 수 없다. 이들 기관의 연구자들의 경우에는 승인된 기관과 협력 관계를 맺거나, eDRIS가 자신들을 대신해서 분석을 수행하도록 할 수 있으며, 개인정보가 포함되어 있지 않은(non-disclosive) 결과물을 받는다.

---

115) Confidentiality: NHS Code of Practice
https://www.gov.uk/government/publications/confidentiality-nhs-code-of-practice

나. 데이터 연계에 대한 승인 절차

NHS 스코틀랜드, ISD 혹은 스코틀랜드 통계청(National Records Scotland, NRS)이 보유한 개인 수준의 건강 데이터에 접근을 요청하거나 혹은 다른 데이터셋과의 연계를 위한 경우, '보건 및 사회복지를 위한 공익과 프라이버시 패널(Public Benefit and Privacy Panel for Health and Social Care, PBPP)'에 승인을 신청해야 한다.[116] 만일 연구가 건강 데이터를 포함하지 않을 경우에는 다른 기관의 승인 절차를 밟아야 하며, 이에 대해 연구 코디네이터가 안내해 준다.

eDRIS는 패널에 대한 모든 신청의 단일한 창구 역할을 한다. 모든 신청은 eDRIS 팀 내의 신청 코디네이터(Application Coordinator)가 검토하고, 신청자를 대신하여 제출한다. 이들은 신청자가 목적에 맞게 폼을 작성하도록 지원하여, 패널들이 빠르고 효율적으로 검토할 수 있도록 한다.

NHS Board의 구역 내에서 거주하는 환자나 그 지역에서 치료받은 환자의 데이터 추출을 요청하는 NHS Board 피고용인의 경우에는 패널에 신청하지 않아도 된다.[117] 그 대신, 기밀성 선서(confidentiality statement)를 작성하고 지역의 칼디콧가디언(Caldicott Guardian)[118]의 서명을 받아야 한다.

모든 연구 조사에는 윤리 승인이 필요한데, 다만 다음의 조건을 만족하는 경우에만 예외이다. 즉, 해당 연구가 ① NSS에 의해 제공되고 통제되는 비식별 데이터만을 사용하고, ② 연구의 대상과 어떠한 접촉도 하지 않으며, ③ 과학적 동료 평가를 수행하고, ④ 데이터를 안전시설에서 보유하고 접근할 경우이다.

데이터 접근에 필요한 모든 승인을 받은 후 eDRIS의 연구 코디네이터에게 연락하면, 연구 코디네이터가 데이터 추출 및 연계를 처리해주며, 만일 NSS의 안전시설을 이용한다면 연구를 위한 공간 마련을 요청해 준다.

---

116) 2015년 3월 1일부터 NHS 스코틀랜드의 데이터 사용에 대한 신청서 검토를 위한 체제가 변화되었는데, 기존의 공동체건강색인자문그룹(Community Health Index Advisory Group, CHIAG), NHS 스코틀랜드 국가서비스 프라이버시 자문위원회(Privacy Advisory Committee, PAC), 국가칼디콧가디언(National Caldicott Guardians)에 대한 신청 절차가 공익과프라이버시패널(Public Benefit and Privacy Panel)로 합쳐졌다. http://www.informationgovernance.scot.nhs.uk/pbpphsc/

117) NHS 스코틀랜드는 지역 주민의 건강 보호 및 증진, 그리고 서비스 전달을 책임지는 14개의 지역 NHS Boards들과 이들을 지원하는 7개의 특별 NHS Board 및 한 개의 공공보건기구로 구성된다. http://www.gov.scot/Topics/Health/NHS-Workforce/NHS-Boards

118) 칼디콧가디언(Caldicott Guardian)은 보건의료정보의 기밀성을 보호하고 그것이 적절히 이용되도록 책임지는 고위급 인사이다. https://www.gov.uk/government/groups/uk-caldicott-guardian-council

다. 데이터 연계절차

데이터는 하워드 뉴컴(Howard Newcombe) 원칙에 기반한 확률 매칭 기법을 사용하여 연계된다. 데이터 연계는 신뢰할 수 있는 제3자(TTP)를 통해 수행되고, 'Population Spine'이라는 시스템이 중간연계도구로 사용된다. Population Spine은 NHS 스코틀랜드와 접촉이 있었던, 스코틀랜드의 모든 개인의 개인 식별자를 보유하고 있다. 데이터 연계의 절차는 다음과 같다.

① 데이터 제공기관이 스코틀랜드 통계청(NRS)이 운영하는 색인팀(indexing team)에 개인 식별자와 자체 레코드 ID번호(아래 그림의 RealSourceID)를 제공한다.

② 색인팀은 복잡한 알고리즘을 사용해 그 식별자와 'Population Spine'의 확률 매칭을 시행한다.

③ 데이터 제공기관은 자체 레코드 ID 번호(RealSourceID)와 해당 데이터셋에 특정된 고유한 색인 ID 번호(indexed source ID)를 포함한 파일을 돌려받게 된다. 색인 ID번호는 색인팀이 생성한 것이다.

④ 데이터 제공기관은 색인팀으로부터 받은 색인 ID번호(이 번호가 연계에 사용된다)를 데이터셋의 콘텐츠에 붙인다. 그리고 연구 코디네이터에게 보낸다.

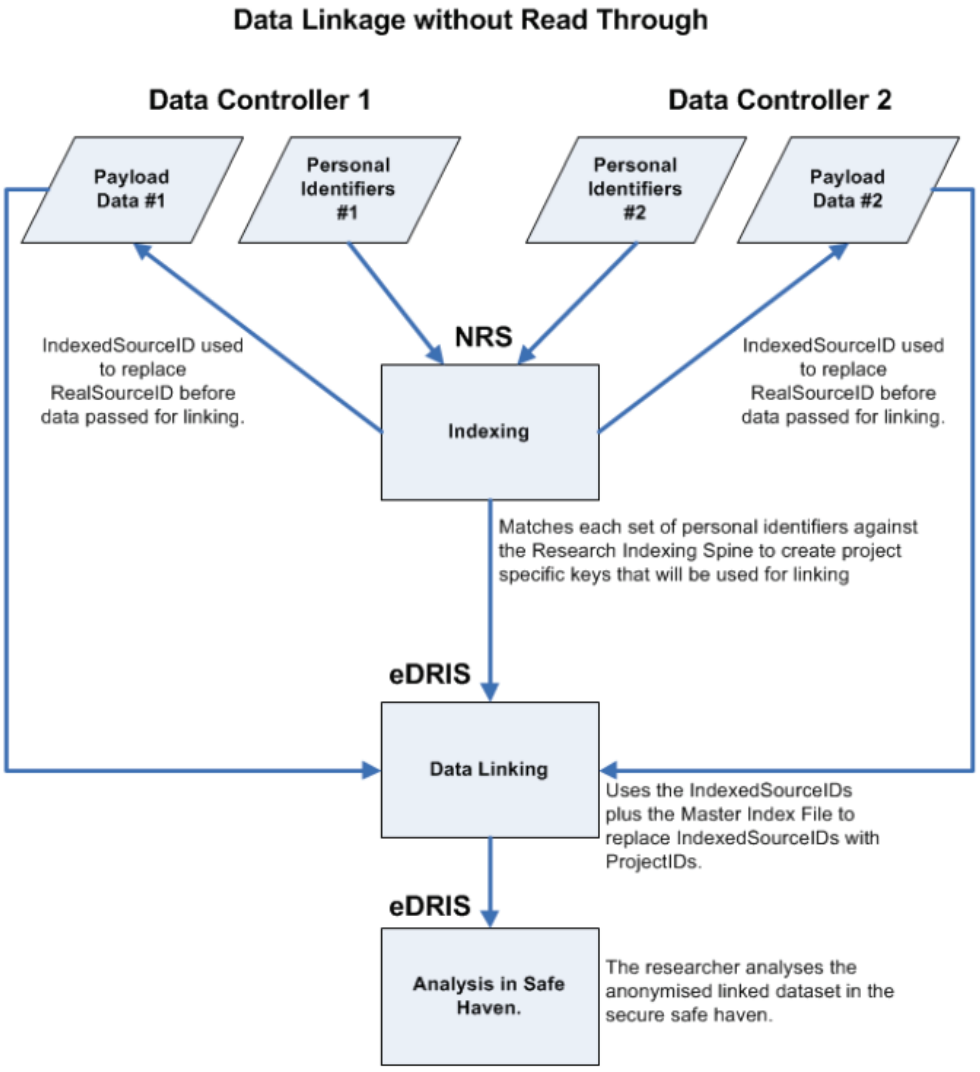
⑤ 연구 코디네이터는 합의된 데이터를 받은 것인지 확인한 후, 그 파일을 연계 에이전트(linkage agent)에 보낸다. 연계 에이전트는 연계를 수행하는 자동화된 컴퓨터 프로그램이다.

⑥ 연계 에이전트는 2개의 파일을 받게 되는데, 하나는 모든 데이터셋과 색인 ID 번호를 가진 파일이고, 또 하나는 프로젝트 ID와 색인 ID 번호를 포함한 마스터색인 파일(Master Index File)이다.

⑦ 연계 에이전트는 모든 데이터셋의 색인 ID 번호를 각 콘텐츠 데이터 파일에 대한 프로젝트 ID로 대체한다.

이를 그림으로 나타내면 아래 그림과 같다.

그림 3-5 eDRIS 데이터 연계 절차 및 기능의 분리



* 출처: NRS (2015)

라. 안전시설(safe haven)에서의 연구수행

데이터 연계 프레임워크 가이드 원칙[119]에 따라 개인 수준의 데이터는 안전한 환경에 보관되어야 한다. 연구자는 NSS의 안전시설에서 데이터에 접근하고 연구를 수행하게 된다.

이 안전시설은 고성능 컴퓨팅 서비스와 보안 분석 환경, 보안 파일전송 프로토콜, 다양한 분석 소프트웨어를 제공한다. 이에 대한 접근은 물리적으로 안전한 접속장소(access point)에서 하게 되는데, 이는 현재 스코틀랜드 Farr Institute에 위치해 있다.

119) Guiding Principles for Data Linkage http://www.gov.scot/Resource/0040/00407739.pdf

이곳을 사용하기 원하는 연구자들은 eDRIS의 연구 코디네이터에 연락하여 예약해야 한다.

데이터 보유기관이 허용한 경우에는 승인된 기관의 PC나 노트북을 통해 원격으로 접속할 수도 있다. 이는 VPN을 통해 이루어지며 2단계 인증(2factor authentication)을 거치게 되는데, 그중 하나는 연구자의 휴대전화로 전송되는 접근 코드이다.

안전시설에는 CD나 USB와 같은 외부 기기를 가지고 들어갈 수 없으며, 물리적인 안전시설에서든 원격으로 접속하든 데이터를 가지고 나올 수는 없다. 모니터링 소프트웨어가 외부 기기를 통한 소프트웨어나 데이터의 다운로드나 업로드를 차단하고 기록에 남긴다.

신청할 때 명시한 기간 동안 데이터에 접근할 수 있으며, 종료 한 달 전에 eDRIS 연구 코디네이터가 연구 장소의 모든 데이터가 아카이브되어도 좋은지 물어본다. 연구 코디네이터와 협의하여 연기할 수도 있다. 데이터는 아카이브에 일정 기간 저장되어 있다가 파기되는데, 아카이브 기간도 계약에 의해 연장될 수 있다. 향후 추가적인 분석이 필요할 경우, 아카이브된 데이터를 다시 가져와야 할 수도 있기 때문이다.

마. 연구결과물에 대한 처리

연구자의 연구결과물은 공개되기 전에 eDRIS 연구 코디네이터가 공개 위험성에 대해 검토한 후에, 연구자에게 이메일로 보내준다.

바. 정보 보안

eDRIS는 데이터의 보안 및 기밀성을 보장하기 위한 다양한 절차를 마련하고 있다.

- 최신의 데이터 보안 기술 적용 및 백업
- 연구자에 대한 승인 및 안전시설 등 정보 거버넌스 절차
- 데이터 제공기관과 안전시설 사이에 보안 파일전송 프로토콜 이용
- 익명화 : 역할의 분리와 색인 절차를 적용하여 개인 식별자와 속성/콘텐츠 정보의 분리
- 데이터 연계 : 프라이버시 보호 프로토콜 및 직접/확률 매칭 적용.
- 안전 환경 (안전시설)
- 기밀성 : 기밀 데이터에 대한 접근 제한. 그래픽, 테이블, 회귀분석과 같은 통계적 결과물에 대해 공개 전에 '통계적 노출제어 방법(statistical disclosure control

method)'[120] 적용.

(5) 북아일랜드 BSO Honest Broker Service[121]

영국 북아일랜드의 보건복지(Health and Social Care, HSC) 사업서비스기구(Business Service Organization, BSO)는 북아일랜드의 보건복지 서비스 부문에 대해 재정, 인력, 법률, 정보통신 등 다양한 지원을 제공한다.

Honest Broker Service(HBS)는 보건복지 데이터에 대한 접근 촉진과 데이터 분석의 지원을 목적으로 2014년 6월 20일에 시작되었다.(HSC, 2015) HBS의 수립 과정에서, 당시의 보건복지 및 공공안전부는 개인정보감독기구인 ICO와 협의하였고, 양해각서(MOU)에 대한 ICO의 의견이 최종 문서의 설계와 내용에 반영되었다. 이 MOU는 북아일랜드의 HSC 관련 조직들[122]에 의해 서명되었다. HBS는 ICO의 익명화 실행규약(Anonymisation Code of Practice)에 따라 수립되었으며, ICO는 계속적으로 이 서비스를 관장하는 이사회(Honest Broker Governance Board, HBGB)에 참여하고 있다.

HBS는 보건부(Department of Health)와 HSC 조직에 익명화, 총계화 혹은 어떤 경우에는 가명화된 보건복지 데이터에 대한 접근을 제공하고, 윤리 승인을 받은 보건복지 연구에 익명 데이터에 대한 접근을 제공한다.

HBS의 2014.6-2015.5 연례보고서에 따르면, HBS는 다음과 같은 목표를 가지고 있다.

- 공공보건의 증진, 보호, 유지에 기여할 수 있는 윤리 승인된 연구의 촉진
- 보건서비스의 계획, 평가, 전달 촉진
- 보건서비스 전반에 걸쳐 위의 목표와 관련된 활동 지원
- 보건데이터 수집, 보건 관련 데이터의 연계, 보건 관련 통계의 편집 및 이용 일반에 관한 지식에 기여
- 적절한 거버넌스에 따라, 위의 목표와 관련된 활동 결과물을 공익목적으로 공개하고 개방적이고 공정한 지식에 기여

---

120) 데이터 기반 연구에서, 설문조사나 행정데이터의 분석 결과로부터 개인이나 단체가 식별되지 않도록 하는 기법을 말한다. https://en.wikipedia.org/wiki/Statistical_disclosure_control (2017년 10월 9일 접근)

121) 이하 내용은 BSO의 Honest Broker Service 페이지를 참조한 것이다. HSC Business Service Organization, Honest Broker Service. http://www.hscbusiness.hscni.net/services/2454.htm

122) 북아일랜드의 보건복지서비스(HSC)는 많은 조직이 관련되어 있는데, 그 구조는 다음 사이트에서 참조할 수 있다. http://online.hscni.net/home/hsc-structure/

연례보고서에 따르면, 현재 HBS의 4가지 주요 활동은 다음과 같다. 이를 보면, HBS의 운영이 아직 초기 단계임을 알 수 있다[123].

① 인프라
- 기술적 역량 있는 직원의 채용
- 지역 데이터웨어하우스의 범위를 확대하기 위해 기존 작업 프로그램과 부합하는 새로운 데이터셋에 대한 접근 허용
- HBS 웹사이트와 메타 데이터 개발
- 운영을 위한 HBS 워킹그룹 설립
- 안전한 연구 보안시설(safe haven) 구축
- 서비스 역량의 시연을 위한 HBS 시범 연구에 대한 보고서 생산

② 보안 및 거버넌스
- 연구 신청서 승인 및 서비스 운영 감독을 위한 운영 규칙 및 절차를 포함한 이사회(HBGB) 설립
- HBS가 북아일랜드 및 영국의 모든 법적 요건과 모범 관행을 따르도록 보장

③ 소통
- HBS 서비스 개시
- 서비스에 대한 인식 및 지원 확대를 위한 참여, 협력, 소통
- 서비스의 효과적인 이용을 보장하기 위해 HSC 내, 산업계, 고등교육기관과의 협력 관계 구축
- 유사 기관과의 협력 관계 구축

④ 운영 절차
- 수수료 정책을 포함하여, 탄탄한 연구 신청 절차의 수립
- HBS에의 연구 신청을 위한 가이드라인 및 규칙 등 개발
- HSC데이터를 평가하는 HSC 기구의 절차 및 표준에 대한 동의 및 문서화

북아일랜드 종적연구(Northern Ireland Longitudinal Study, NILS), 북아일랜드의 행정데이터연구센터(ADRC) 등이 HBS와 유사한 서비스, 즉 보건의료 데이터에 대한 접근 및 연계 서비스를 제공하고 있다. HBS가 보건데이터만 제공하는 데 반해, ADRC는 보건복지 데이터 및 다른 행정데이터도 제공하고 있으며, NILS는 인구조사 및 일반등록소 데이터와 연계된 보건복지 데이터를 포함한다.[124]

가. 거버넌스

HBS는 HSC BSO의 고객관리 및 성과 책임자(Director of Customer Care & Performance)의 지휘 하에 운영되고 있으며, 운영 책임자 산하 정보연구단의 지원을 받는다. 이사회(HBGB)는 모든 HBS의 운영 규칙과 절차를 감독하고, 연구 신청서를

123) HBS 2014.6-2015.5 연례보고서.

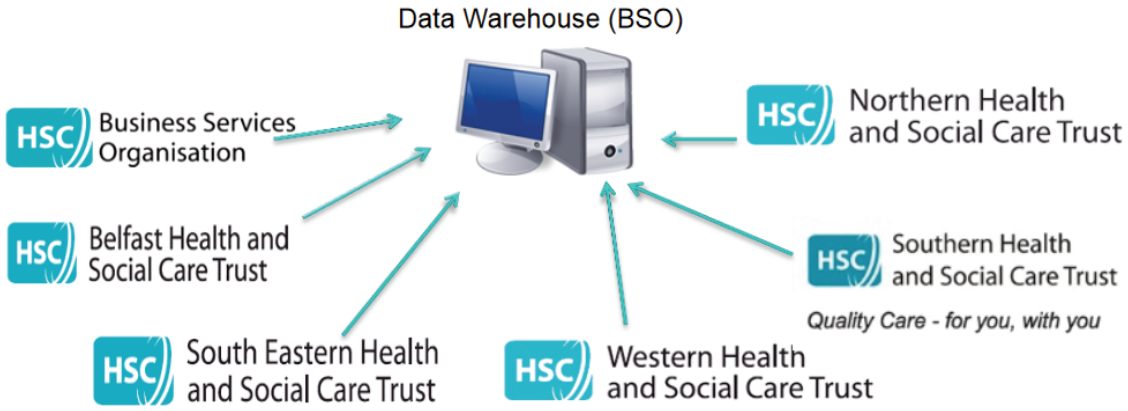
124) Similarities and differences between: HBS, ADRC, Open Data, and NILS, http://www.hscbusiness.hscni.net/pdf/Similarities_and_differences_Final_Version.pdf

승인하는 역할을 한다. 이사회는 지역 데이터웨어하우스(Regional Data Warehouse, RDW) 내의 데이터관리자, HSC 조직들의 주요 데이터 이용자들의 대표로 구성된다.

나. 데이터에 대한 접근 및 연계

HSC의 지역 데이터웨어하우스는 다양한 HSC 정보 시스템으로부터 정기적으로 데이터를 받아 대규모 데이터를 안전하게 관리하는 곳이다. 이는 BSO에 의해 관리되며 지역 데이터센터 내에 위치하고 있다. 허가된 이용자가 정보를 안전하게 관리하고 개인정보 보호 법제를 준수하도록 보장한다. 또한, 사업정보 도구 및 관리정보 포털을 통해 저장된 비식별 데이터들의 분석을 지원한다. HSC 관련 기관들은 직접적인 환자 치료를 위해 수집된 자신의 데이터에 접근할 수 있으며, 다른 기관들은 데이터 접근 계약이나 HBS를 통해 허가된 접근만이 이루어진다. 즉, HSC는 보건복지 및 연구를 위해 연계되고 비식별화된 웨어하우스 데이터에 대한 안전한 접근을 제공하는 효율적인 수단으로서 역할 한다. 현재의 법률 하에서, HSC는 환자 식별정보는 누구와도 공유하지 않는다.

그림 3-6 HSC 지역 데이터웨어하우스



* 출처: Siobhán Morgan (2016)

HBS는 홈페이지를 통해 보건부와 HSC용, 일반 연구자용 서비스 신청서를 각각 제공하고 있다. 연구자에게는 HBS의 보안시설(safe haven)을 통해 익명화된 데이터로만 제공되며, 데이터 접근을 위해서는 윤리 승인을 얻어야 하고, '안전한 연구 데이터 이용자 환경 훈련(SURE training)'도 받아야 한다. 영리기업 역시 이 서비스를 이용할 수 있다. HBS는 데이터 연계 서비스도 제공하는데, 구체적인 데이터 연계절차는 홈페이지를 통해 설명되어 있지 않았다.

(6) care.data 논란

care.data는 일반의(GP)가 보유한 환자정보를 HSCIC의 국가 데이터베이스에 집적하려는 NHS 잉글랜드의 사업으로 2013년에 시작되었다. 진료정보의 활용을 통해 서로 다른 영역을 통해 제공되는 진료에 대한 전반적인 상황을 파악하고, 개선이나 투자가 필요한 부분을 찾아내는 것이 이 사업의 목적이다.

NHS는 이미 병원의 환자 데이터를 수집해왔다. care.data는 이를 일반의가 보유하는 데이터로 확대하고자 한 것인데, 이것이 문제가 되는 이유는 영국에서 대부분의 사람이 병원보다 일반의에게 훨씬 많이 가고, 일반적으로 일반의는 사람들의 일생동안의 조건, 처방, 가족력, 혈액 검사, 이첩(referral) 등에 대한 기록을 보유하고 있기 때문이다.

일반의의 시스템으로부터 추출하는 데이터는 이첩(referral), NHS 처방전, 가족력, 예방접종, 혈액검사결과, 체질량 지수, 흡연/음주습관 등이다. 이 정보들은 문장 형식이 아니라 코드 형식으로 추출되며, 생년월일, 우편번호, NHS 번호, 성별 등의 식별자가 연계를 위해 필요하다. 상세 정보는 다른 보건의료 데이터와의 연계를 위해 가명화되고 식별자는 분리되어 저장된다. 데이터는 매달 업데이트되고 잉글랜드의 모든 환자 데이터를 포괄한다.

care.data의 데이터는 직접적인 진료 목적이 아닌, 의료 서비스의 계획이나 의학 연구 등 2차적 목적을 위해 활용된다. NHS 내 기관뿐만 아니라, NHS 외부의 (승인된) 제약회사, 보건 자선단체, 대학, 병원 위탁단체, 싱크탱크 및 다른 사기업 등에 제공될 수 있다.[125]

환자들은 일반의가 보유한 자기 정보를 HSCIC에 이전하는 것에 대해서, 혹은 HSCIC로 이전된 데이터를 연구자 등 제3자에게 이전하는 것에 대해서 거부권(opt-out)을 행사할 수 있다. 그럼에도 불구하고 care.data 사업은 많은 사회적 반발을 가져왔다. 정보주체의 동의 없이 민감한 의료정보가 HSCIC에 집적되고 연구자 등 제3자에게 제공됨으로써 정보주체 통제권 약화, 해킹 등을 통한 개인정보 유출, 보험회사 등 영리적인 목적을 위한 제공, 비식별 데이터의 재식별 가능성 등 개인정보 침해에 대한 여러 우려가 제기되었기 때문이다. 특히 HSCIC가 보험사에 비식별 데이터를 제공했다는 것이 미디어를 통해 알려지면서 care.data 사업은 대중의 큰 반발을 야기하게 되었다.

2015년 9월 영국 복지부는 '복지 품질 위원회(Care Quality Commission, CQC)'에

125) WIRED. "A simple guide to Care.data". 2014.2.7.
http://www.wired.co.uk/article/a-simple-guide-to-care-data

NHS의 데이터 보안에 대한 검토를, 그리고 보건복지를 위한 국가 데이터가디언인 칼 디콧(Dame Fiona Caldicott)에게 데이터 보안 및 동의에 대한 독립적인 검토를 수행해 달라고 요청했다. 이들의 보고서는 2016년 7월에 발표되었다.

많은 논란 끝에, 보고서의 권고에 따라 결국 2016년 7월 care.data 프로그램은 일단 취소되었다. 그러나 향후 좀 더 보완된 방식으로 의료정보의 공유 및 활용을 촉진하는 사업이 시행될 것으로 보인다.[126]

사업을 시작하면서 영국 정부는 일반의 진료소에 포스터를 붙이는 것으로 충분하다고 생각했다. 그러나 이 사업에 대한 비판이 제기되자, 이후 홍보 리플릿을 배포했다.[127] 그러나 이 리플릿 역시 사업에 대한 홍보가 중심일 뿐 이용자가 어떻게 거부권을 행사할 수 있는지에 대한 상세한 정보를 포함하고 있지 않았다. 리플릿 배포 이후 언론에도 보도되기 시작하고 논란은 더욱 확산되었다.[128]

care.data 사업의 실패는 개인정보의 수집·활용을 촉진하기 위한 사업에서 보안 및 개인정보 보호에 대한 대중들의 신뢰를 얻는 것이 얼마나 중요한지, 그리고 해당 사업을 추진하는 당국이 사업의 내용을 투명하게 공개하고 관련 이해당사자와 충분히 협의하는 것이 얼마나 중요한지를 보여주고 있다.

2. 호주 PHRN[129]

(1) 개요

호주의 인구보건연구네트워크 PHRN(Population Health Research Network)은 호주 전역에 걸쳐 보건 정보를 안전하게 관리할 수 있는 국가적인 데이터 연계 기반을 구축하기 위해 설립되었다. 호주에서는 1990년대 중반부터 일부 지역에서 데이터 연계가 이루어져 왔는데, PHRN의 설립을 통해 호주의 각 주(state) 및 준주(territory)[130]에서 운영되고 있는 데이터연계기구(Data Linkage Unit)들이 인구보건연구를 보다 포괄적이고 효과적으로 수행할 수 있게 되었다. PHRN은 호주 정부의 국가협력연구기반전략(NCRIS) 프로그램으로부터 자금을 지원받아 2009년부터 운영을 시작했으며,

126) https://www.gov.uk/government/speeches/review-of-health-and-care-data-security-and-consent

127) https://www.england.nhs.uk/wp-content/uploads/2014/01/cd-leaflet-01-14.pdf

128) https://medconfidential.org/whats-the-story/

129) 아래 내용은 호주 PHRN 홈페이지를 참조한 것이다. http://www.phrn.org.au

130) 호주 연방은 6개의 주(state)와 2개의 준주(territory), 6개의 특별지역으로 이루어져 있다. https://ko.wikipedia.org/wiki/%EC%98%A4%EC%8A%A4%ED%8A%B8%EB%A0%88%EC%9D%BC%EB%A6%AC%EC%95%84 (2017.9.9 확인)

호주 정부뿐만 아니라, 주 정부 및 연구소와 대학의 지원도 받고 있다.

PHRN은 다양한 협력자들의 네트워크로 구성된다. 서호주 대학(University of Western Australia)이 주관기관(lead agent)인데, 참여자들과 계약 하에 데이터 연계 기반을 구축하고 있다. 이에는 다음과 같은 기관들이 포함된다.

- 각 주 및 준주(territory)에서 서비스를 제공하는 데이터연계기구의 네트워크
- 주간, 연방(Commonwealth) 간, 주와 연방 간 데이터 연계를 제공하는 국가데이터 연계기구
- 안전한 원격접근연구소
- 보안파일전송시스템
- 국가조정사무소

PHRN은 각 주 및 준주의 6개의 데이터연계기구와 협력을 하고 있다.

표 3-1 호주 데이터연계기구(Data Linkage Unit)

데이터 연계 기구	위치	관할권
AIHW	호주 보건복지연구소	전지역
Data Linkage Branch	서호주 보건부	서호주
SA-NT DataLink	남호주 대학	남호주 및 북부준주
CHeReL	뉴사우스웨일즈주 보건부	뉴사우스웨일즈 및 호주 수도 준주
Center for Victorian Data Linkage	빅토리아주 보건복지부	빅토리아주
Tasmanian Data Linkage Unit	태즈메이니아 대학	태즈메이니아주
Data Linkage Queensland (DLQ)	퀸즈랜드 보건부	퀸즈랜드주

* 출처: PHRN 홈페이지 http://www.phrn.org.au/about-us/who-is-involved/australian-data-linkage-units/

PHRN에서 활용하는 데이터는 행정데이터와 프로젝트 기반 데이터(project specific datasets)로 구분할 수 있다. 다만, 데이터연계기구 자체가 개인으로부터 데이터를 수집하는 것은 아니다.

행정데이터란 서비스 제공 과정에서 일상적으로 수집되는 데이터를 의미하는데, 정부 부처나 기관에 의해 수집되며 일반적으로 법에 의해 그 수집이 허가된다. 특히 의료 제공자들은 의무적으로 보고하도록 되어 있어서 환자들의 동의 없이 데이터 수집이 가능하다. 이 정보들은 각 기관의 안전한 데이터 보관소에 저장되며, 데이터 보유기관(data custodian)의 통제 하에서 접근할 수 있다. 행정데이터에는 출생등록소, 사

망등록소, 암등록소, 응급부서나 입원환자의 데이터 등이 포함된다.

때때로 정부나 기관은 특정한 목적을 위해 건강 행태에 대한 설문조사 등을 통해 정보를 수집하는데, 이것이 프로젝트 기반 데이터이다. 연구자들은 승인 하에 이러한 데이터를 행정데이터나 다른 프로젝트의 데이터와 연계해줄 것을 데이터연계기구에 요청할 수 있다. 이러한 데이터들은 연구 목적으로, 정보주체의 동의를 받아 수집되며, 때로는 다른 정보에 대한 접근을 위한 동의가 포함되기도 한다.

(2) 거버넌스

서호주 대학이 주관기관으로서 기금지원계약에 명시된 보고 및 책무성 요건에 따라 PHRN에 대한 전반적인 책임을 진다. PHRN의 이사회(board)가 감독 및 전략적 방향을 제공한다. PHRN 이사회는 PHRN 참여기관협의회(Participant Council)의 자문을 받는다.

PHRN은 그 자체로 하나의 기관이 아니며, 참여기관 각자의 역할과 책임이 일련의 계약을 통해 정의된다. 이러한 계약에 따라 각 기관은 1988 프라이버시법(Privacy Act 1988)의 정보프라이버시 원칙, 각 주의 프라이버시 법령, 그리고 PHRN의 정책을 준수해야 한다. 기관 간의 계약뿐만 아니라, 개인정보에 접근하는 데이터 연계 직원과 연구자들도 보안 각서(Confidentiality Agreements)에 서명해야 한다.

(3) 연구의 승인

연계 데이터에 접근할 수 있는 연구의 자격 요건은 다음과 같다.

- 공중 보건의 증진, 보호, 유지에 기여할 수 있는 연구를 촉진할 것
- 보건서비스의 계획, 평가, 전달을 촉진할 것
- 일반적으로 보건 데이터 수집, 보건 관련 데이터의 연계, 보건 관련 통계의 편집 및 활용과 관련된 연구 방법에 대한 지식에 기여할 것.

또한, 연구자는 다음과 같은 자격 요건을 갖추어야 한다.

- 제안된 연구를 수행할 수 있는 적절한 경험, 자격, 시설, 자금을 갖춘 연구자
- 적절한 경험 및 자격을 갖춘 연구팀의 일원인 학생 및 초기 경력 데이터 이용자
- 프로젝트의 특성 및 요구되는 데이터의 형태에 따라서는, 국제 협력자.

통상 선착순에 의해 데이터 접근의 우선순위가 정해지지만, 다음과 같은 요소가 고려될 수 있다. 즉, 데이터 입수 가능성, 프로젝트의 복잡성/기술적 실행 가능성, 공익성, 기금과 같은 자원의 입수 가능성, 호주 보건 장관 회의에 의해 결정되는 국가 보건 우선순위 분야, 전략적 우선순위 등이다.

연계 데이터를 사용하는 연구 프로젝트는 데이터연계기구, 각 데이터셋을 보유한 데이터 보유기관들, 인간연구윤리위원회(HREC) 등 세 기관의 승인을 얻어야 한다. 승인 절차, 순서, 그리고 데이터 보유기관의 승인을 얻기 위한 지원의 수준은 데이터 연계기구에 따라 다르다.

(4) 데이터 연계절차

① 1단계 : 신청 절차 (application process)

연구자는 데이터 연계를 신청할 경우, 데이터 보유기관에 다른 데이터셋과의 연계를 위한 식별 정보(이름, 주소, 생년월일 등)의 제공 및 연구 분석을 위한 진료 및 다른 데이터의 제공에 대해서 승인을 요청해야 한다. 데이터 보유기관으로부터 데이터 연계기구는 식별정보만을, 연구자는 콘텐츠 정보만을 받게 된다.

② 2단계 : 식별정보의 데이터연계기구에의 제공

프로젝트가 승인되면, 데이터연계기구는 관련 데이터 보유기관에 연계 변수(linkage variables)를 요청한다. 연계변수란 이름, 생년월일, 성별, 주소, 기록일, 기타 병원진료 기록번호 등과 같은 개인 식별자인데, 연계변수가 많을수록 연계의 질이 높아진다. 연계변수는 연계자가 데이터셋을 서로 연계하는 데 사용된다. 특정 프로젝트를 위한 연계변수는 신청 과정에서 '데이터 신청양식'에 명시된다.

데이터연계기구와 데이터 보유기관은 데이터 추출 절차 이전에 추출 계획을 준비한다. 이는 프로젝트의 복잡성과 접근 가능한 변수에 따라 달라진다. 어떤 프로젝트는 코호트 선택을 결정하기 위해 몇 번의 반복적 과정을 요구하게 된다.

식별정보와 함께 데이터연계기구는 데이터 보유기관에 하나 이상의 필드를 요구한다.

- 레코드 ID : 데이터연계기구에 보내지는 개별 레코드의 식별자.
- 환자 ID 혹은 개인 ID : 데이터 보유기관 DB 내의 각 개인 식별자. 고유한 개인 식별자가 없을 경우 환자 ID 필드는 레코드 ID로 설정됨.

프라이버시 보호를 위해, 데이터 보유기관은 본래의 레코드 ID나 환자 ID가 아니라, 프로젝트 고유의 ID를 생성하여 데이터연계기구에 보낼 것을 권고하고 있다. 이는 암호 소프트웨어를 사용하거나 자동번호 생성을 통해서 가능하다. 예를 들어, 한 사람에 대해 다수의 레코드가 있고 고유한 개인번호가 있을 경우, 환자 ID 필드는 다음과 같이 할 수 있음.

- 고유한 개인번호를 가진 레코드 목록 생성
- 각 개인번호에 자동번호 적용
- 환자 ID 필드에 자동번호 설정

프로젝트 완료 시까지 자동번호와 레코드 ID 혹은 환자 ID의 매핑을 유지한다. 이렇게 하지 않으면 데이터연계기구에서 생성한 정보와 원 데이터를 합칠 수 없다.

데이터연계기구에 보내지는 데이터는 식별정보이다. 모든 데이터는 제공되기 전에 암호화되어야 한다. 데이터연계기구에는 CD나 USB에 저장해서 인편으로 전달하는 방법, CD나 USB에 저장해서 등록된 장소로 전달하는 방법, Secure data transfer(SUFEX)로 전달하는 방법(아래 참고) 중 하나의 방법을 사용하여 전달한다.

③ 3단계 : 데이터 보유기관에 매핑 파일 전달

데이터 연계가 완료되면, 데이터연계기구는 각 데이터셋을 위한 '프로젝트 키'를 생성한다. 이는 각 데이터 보유기관에 전달되어 콘텐츠 데이터에 덧붙여진다. 프로젝트 키는 마스터 연계키를 암호화한 것이며, 포맷이나 필드 수는 데이터연계기구에 따라 달라진다. 통상 프로젝트 키는 프로젝트 개인번호(Project Person Number, PPN), 레코드 ID, 프로젝트 사건 번호(Project Event number, PEN) 등으로 구성된다.

다음, 데이터 보유기관은 다음과 같은 작업을 수행한다. 우선, 앞서 생성한 매핑 테이블을 통해 프로젝트 키 파일 내의 레코드 ID를 보유기관의 내부 레코드 ID로 변환한다. 그리고 레코드 ID를 이용하여 PPN과 PEN을 데이터셋의 레코드에 결합한다. 다음 데이터베이스에서 프로젝트에 필요한 PPN, PEN, 콘텐츠 데이터를 뽑아낸다. 마지막으로 PPN, PEN과 승인된 콘텐츠 데이터가 포함된 파일을 연구자에게 전달한다.

④ 4단계 : 추출 데이터를 연구자에게 전달

추출 데이터를 누가 전달할 것인지는 데이터연계기구의 운영 모델이나 데이터 보유기관의 선호에 따라 달라진다. 연구자들은 자신들이 요청한 데이터를 데이터 보유기관 각각으로부터, 혹은 데이터연계기구로부터 비식별화된 데이터로 받는다. 일부 데이

터연계기구는 연구자에게 전달하기 전에 데이터의 준비를 지원하는데, 추출된 데이터의 연계 전 점검, 데이터에 파생 변수의 부가, 데이터의 연계, 연계 후 점검, 연구자에의 제공 등이다.

추출된 데이터를 연구자에게 전달하는 몇 가지 방법은 PHRN 데이터 이전 협약(PHRN Data Transfer Agreement)에 규정되어 있는데, SUFEX를 통한 방법, 데이터연계기구의 보안파일전송시설을 통한 방법, SURE를 통한 방법, 암호화 디스크를 통한 방법 등이 있다.

- Secure data transfer (SUFEX) : SUFEX는 PHRN이 제공하는 보안 파일전송 서비스이다. 이용자는 언제 어디서나 SUFEX를 이용하여 파일을 주고받을 수 있다. 등록된 이용자는 개인 로그인 계정을 부여받으며 다른 등록 이용자뿐만 아니라, 등록되지 않은 이용자와도 파일을 주고받을 수 있다. SUFEX는 컬틴(curtin) 대학에 있는 데이터연계센터에서 운영하고 있다.
- Secure Unified Research Environment (SURE)[131] : SURE는 연구자들이 승인된 데이터에 접근할 수 있는 원격접근 컴퓨팅 환경이다. 연구자의 컴퓨터 화면에는 원격지의 가상 컴퓨터의 복사본이 뜨게 된다. SURE 내에서 연구자들은 SURE에 의해 통제되는 가상 컴퓨터를 할당받는다. 또한 SURE는 고용량 데이터 저장 공간, 백업 기능, 연구자가 활용할 수 있는 분석 소프트웨어를 보유하고 있다. SURE에 접근하기 위해 이용자는 등록폼을 작성하고 훈련을 받아야 하며, 이용약관에 서명해야 한다. SURE를 통해 파일을 주고받기 위해서는 '큐레이티드 게이트웨이(Curated Gateway)'라는 포털을 거쳐야 한다. 연계 데이터가 SURE에서 제공될 경우, 연구 승인 시 허용된 데이터 보유 기간 만료 후 SURE에 대한 접근이 제한된다.
- 암호화 디스크 : 암호화된 파일을 디스크로 구워서 연구자에게 전달할 수도 있다. 비밀번호는 별도의 통로로 연구자에게 전달된다. 가능하면 연구자는 데이터 보유기관이나 데이터연계기구로부터 직접 전달받아야 한다.

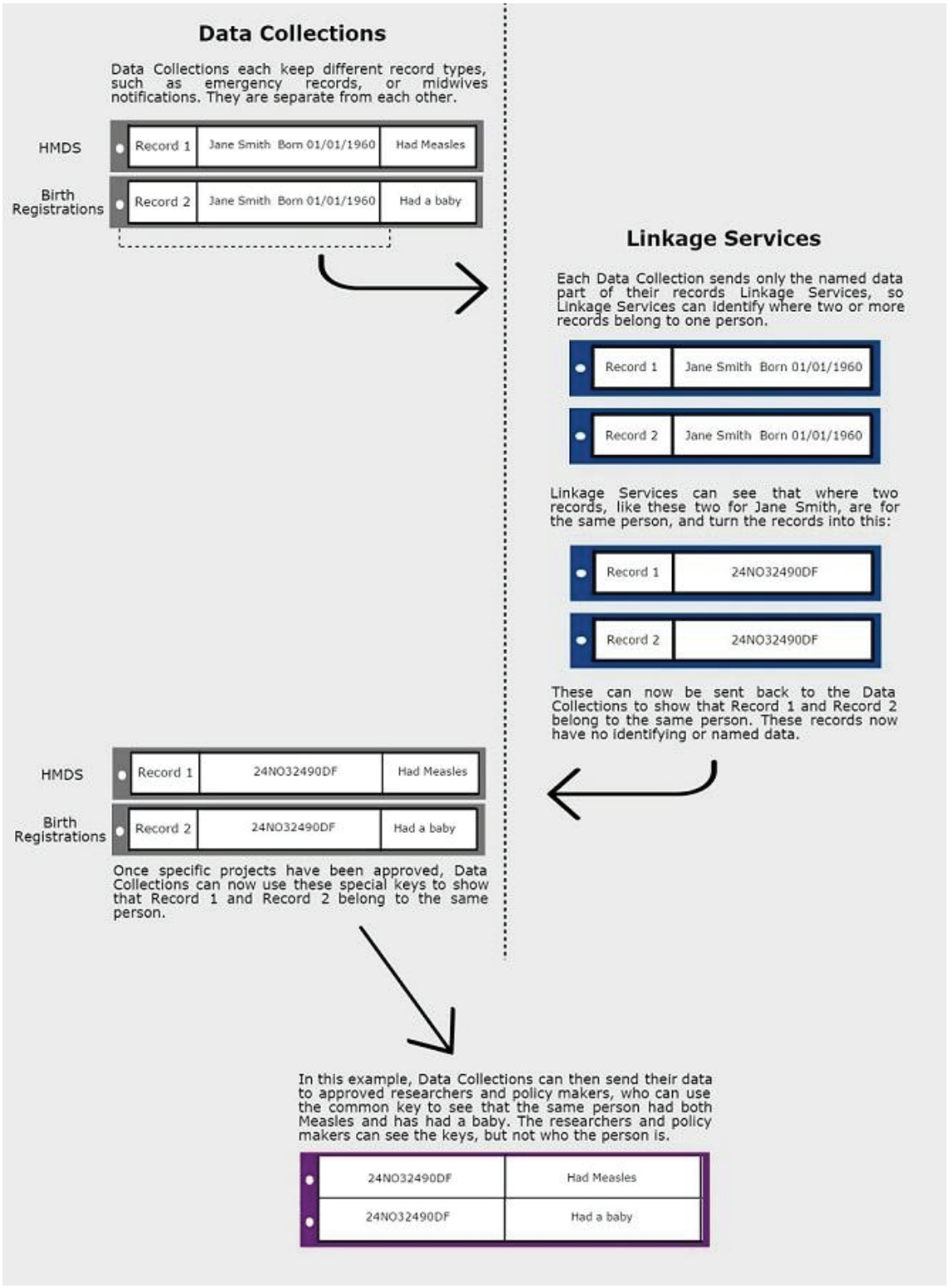
⑤ 5단계 : 데이터 연계 프로젝트의 모니터링.

데이터 보유기관은 연구자가 합의된 조건을 준수하는지 모니터링할 책임이 있다. 데이터연계기구가 제공하는 서비스 중의 하나가 프로젝트 모니터링이다. 이는 연구결과물(출판물) 및 데이터 삭제 모니터링, 프로젝트 종료 시 아카이빙 등을 포함한다.

PHRN의 데이터 연계절차를 그림으로 나타내면 다음과 같다.

131) 여기서 SURE는 ADRN의 '안전한 연구 데이터 이용자 환경 훈련(Safe Users of Research data Environment Training)'과는 다른 의미이다.

그림 3-7 PHRN 데이터 연계 절차



* 출처: Data Linkage Western Australia, http://www.datalinkage-wa.org/what-is-data-linkage

(5) 개인정보 보호 정책

PHRN은 개인정보 보호 및 보안을 위해 여러 절차를 마련해두고 있다.

- 프라이버시, 보안, 통신 및 정보관리 정책
- 프라이버시 영향평가
- 프라이버시 전문가의 법적 자문
- 소비자 대표의 자문과 훈련, 높은 수준의 보안 기준에 맞춘 첨단 기술의 이용.
- 교육 및 정보제공을 위한 웹사이트
- 책무성을 보장하기 위한 거버넌스 및 관리 시스템

개인정보 보호를 위한 중요한 원칙 중 하나가 '분리 원칙(separation principles)'이다. 첫째, 연계 데이터와 콘텐츠 데이터의 분리. 연계를 위한 개인정보는 콘텐츠 데이터와 분리되어, DLU의 데이터 연계자에게는 연계 ID 생성을 위한 개인정보만 제공된다. 둘째, 기능과 책임의 분리. 데이터 연계절차와 데이터 보유 및 추출 기능을 분리한다. 데이터 연계를 수행하는 사람은 연구자와 분리되어 있어야 하며, 콘텐츠 데이터 연구에 참여할 수 없다.

정보의 이용, 제공, 보유 또한 제한된다. 연구자는 특정 프로젝트를 위한 정보에만 접근이 허용되며, 승인받은 방식으로 이용해야 한다. 프로젝트의 연구자들은 신원을 확인받고 승인되어야 하며, 다른 사람에게 정보를 주어서는 안 된다. 또한, 정보는 승인받은 기간 동안만 보관되며, 그 이후에는 데이터 보유자에게 반환하거나 삭제된다. 이러한 조건은 연구자가 데이터 보유자와의 계약, HREC 승인 조건, 혹은 DLU와의 합의에 의해 관리된다. HREC와 데이터 보유자는 연구자가 합의된 조건들을 준수하는지 감사, 감독, 점검할 권한을 가진다.

PHRN은 개인정보 보호정책의 효과적인 이행을 위해 연구자에 대한 적절한 훈련을 제공하고 있으며, 이러한 훈련은 HREC 구성원들에게도 제공된다. 또한, PHRN은 네트워크 개발의 많은 단계에서 프라이버시 영향평가를 수행하고 있다.

3. 미국

미국의 보건의료 시스템은 민간의 보건의료 서비스 제공자와 보험사가 주도하고 있다. 미국 국민의 약 3~40% 정도가 메디케어(Medicare)나 메디케이드(Medicaid)와 같

은 공공 의료 시스템의 지원을 받고 있지만 대부분의 미국인은 민간의료보험을 이용하고 있다. 민간의료보험 가입자의 8~90%는 고용주 지원 프로그램(employer-sponsored coverage)이다. (OHE Consulting Report, 2015)

따라서 보건의료 데이터의 수집 및 보관도 다양하게 분산되어 있다. 민간 영역에서는 Truven Health Analytics, HealthCore, Humana 등의 민간기업들이 보건의료 데이터를 수집 및 제공하고 있다. 연구 목적의 데이터 접근을 위한 계약 체결도 이와 같은 개별 업체를 매개로 해야 한다. 보통 사례별로 이용허락이 이루어지며 이용 대가를 지불해야 한다. 또한, 업체들은 연구 출판 전에 연구 결과에 대한 사전 승인을 요구할 수 있다. 통상 민간 영역에서의 데이터 연계는 이러한 업체 내에서 이루어지며, 이들은 다른 업체와의 데이터 연계에는 제한을 두고 있다. 일반적으로 비식별 데이터로부터 개인식별을 시도하거나 다른 데이터 소스와 연계하는 것을 제한한다. (OHE Consulting Report, 2015)

미국의 보건의료 시스템이 다양한 만큼, 데이터에 대한 접근 및 이용 방식 역시 데이터 보유기관에 따라 천차만별이다. 여기서는 주로 공공기관에서의 데이터 수집·연계, 연구 목적의 제공과 관련된 몇 가지 사례를 살펴보려고 한다.

(1) 국가보건통계센터(NCHS)의 데이터 연계[132)]

국가보건통계센터(The National Center for Health Statistics, NCHS)는 보건복지부(Department of Health and Human Services., HHS) 질병통제예방센터(The Center for Disease Control and Prevention, CDC) 산하 기관으로 건강 증진을 목적으로 한 통계 정보를 제공한다.

가. NCHS 데이터 및 데이터 연계

NCHS는 보건의료 관련 인구조사 데이터, 출생·사망 기록(Vital Records), 의료기관에서 제공한 데이터 등을 보유하고 있다.

또한, NCHS의 인구기반 설문조사 데이터의 과학적 가치를 극대화하기 위해 데이터 연계 프로그램을 개발하고 있다. 현재 NCHS는 자신이 보유한 다양한 설문조사 데이터와 국가 사망 인덱스(National Death ndex, NDI), 메디케어 및 메디케이드 서비스 센터 (CMS), 미국 신장 데이터 시스템(USRDS), 사회보장국(SSA), 주거 및 도

132) 아래 내용은 NCHS의 데이터 연계 홈페이지를 참조한 것이다.
https://www.cdc.gov/nchs/data-linkage/index.htm

시개발부(HUD) 등 기관의 행정데이터를 연계하고 있다.

NCHS는 자신이 보유한 데이터를 공개사용 데이터 파일(public-use data files)로 만들어 공개하고 있다. 이 파일은 질병통제예방센터(CDC)의 FTP 서버를 통해 다운로드 받을 수 있다. 이용자는 데이터셋, 관련 문서, 질문표 등에 접근할 수 있다.

미국의 공공보건서비스법(The Public Health Service Act) Section 308 (d)는 NCHS와 CDC가 수집한 데이터는 단지 보건통계보고 및 분석 목적으로만 사용할 수 있다고 규정하고 있다. NCHS는 정보주체가 식별되지 않도록 하기 위해, 모든 직접 식별자와 식별을 가능하게 할 수 있는 특성들은 공개 데이터셋에서 제거한다. 개인이나 기관에 대한 의도적인 식별이나 공개는 정보 제공자의 기밀성 보장 의무를 위반한 것이 된다. 이러한 규정에 따라, 공개사용 데이터 파일의 이용자는 다음과 같은 조건에서 사용해야 한다. 첫째, 데이터셋은 통계 보고 및 분석 목적으로만 사용해야 한다. 둘째, 의도하지 않게 개인 및 기관의 신원이 노출될 경우 이를 사용하지 말고, NCHS 책임자에게 알려야 한다. 셋째, 이 데이터셋을 개인식별이 가능한 다른 NCHS의 데이터나 외부 데이터와 연계해서는 안 된다. 넷째, 공개사용 데이터를 이용하는 것은 위의 법적 요구조건을 준수하겠다는 '데이터 이용자 계약(Data User Agreement)'에 서명한 것이 된다.

나. NCHS 연구데이터센터

NCHS는 연구자에게 '제한된 데이터(restricted data)'에 대한 접근을 허용하기 위해 연구데이터센터(RDC)를 두고 있다. 연구데이터센터는 NCHS의 데이터뿐만 아니라, 보건복지부 내의 다양한 그룹의 제한된 데이터도 제공하고 있다. 연구자들은 이 데이터에 접근하기 위해 연구제안서를 제출하고 승인을 받아야 한다.

가) 제한된 데이터

연구데이터센터를 통해 접근할 수 있는 '제한된 데이터'는 이름, 사회보장번호, 주소 등 직접 식별자는 제거되었지만, 지리정보 등 간접 식별자를 포함하고 있는 데이터이다.

나) 연구제안서 신청 절차

연구제안서 신청은 다음과 같은 절차에 따라 이루어진다.

- 1단계 : 필요한 '제한된 데이터' 결정.

- 2단계 : 선호하는 접근방법 결정. (접근방법에 대해서는 아래 '접근방법' 참조)
- 3단계 : 연구제안서 초안 작성
- 4단계 : 연구 제안서 제출
- 5단계 : 검토위원회(review committee)의 검토
- 6단계 : 내용이 변경될 경우에는 연구제안서 갱신

연구제안서의 핵심 부분은 '데이터 사전(data dictionary)'이다. 이는 제안서 검토 과정에서 프로젝트의 공개 위험성을 평가하고, 프로젝트가 승인된 후 데이터셋을 생성하기 위해 사용된다. 데이터 사전은 세 가지 부분으로 이루어진다.

첫째, 공개 데이터 : 공개사용 데이터에서 연구에 필요한 변수만을 선택한다.

둘째, 제한된 데이터 : NCHS는 홈페이지를 통해 접근할 수 있는 제한된 데이터의 목록과 정보를 제공하고 있다.[133]

셋째, NCHS 데이터 외의 데이터 : 다른 데이터 소스의 변수들을 추가할 수 있다.

연구제안서에는 데이터셋 통합에 사용될 변수들을 명시하도록 하고 있다. 이 변수들을 사용하여 NCHS가 연계된 데이터셋을 제공하는 것으로 보인다.

연구 제안서가 접수되면, 연구데이터센터 책임자는 분석가(Analyst)를 지명한다. 이 분석가는 연구자와의 1차적인 소통 창구가 된다. 이 분석가가 하는 역할은 다음과 같다.

- 제안서에 대한 검토 준비
- 연구자를 위한 분석 데이터셋 생성
- 수수료 수납
- NCHS 기밀성 요구조건 수납
- 인구조사국에 연구자의 데이터셋 이전
- 연구자의 연구결과물의 공개 위험성 검토

분석가는 연구제안서의 완성도 및 실현 가능성을 검토하고, 필요할 경우 연구자와

133) Restricted NCHS Variables, https://www.cdc.gov/rdc/B1DataType/Dt122.htm

협의하여 연구제안서를 변경한다. 연구제안서가 완성되면, 분석가는 이를 검토위원회에 보낸다. 검토위원회는 분석가, 1명 이상의 데이터 시스템 대표자, 기밀성 담당관 등으로 구성된다. 검토위원회는 요청한 데이터의 적절성 및 입수 가능성, 공개 위험성, 기술적 실현 가능성, 공공보건에 미치는 이익 및 데이터 보유기관의 임무와의 일치 여부 등의 기준에 따라 연구제안서를 평가한다.

다) 접근방법

NCHS는 제한된 데이터에 접근할 수 있는 몇 가지 방법을 제공하고 있는데, NCHS의 연구데이터센터, 연방통계연구데이터센터(FSRDC), 원격접근 등이다. 연구자들은 이 방법 중에 선택할 수 있는데, 프로젝트나 데이터셋의 성격에 따라 특정한 방법을 통한 접근이 제한되는 경우도 있다.

NCHS의 연구데이터센터는 NCHS 본부가 있는 하이엇츠빌(Hyattsville), 아틀랜터의 CDC의 세기센터캠퍼스, 워싱턴 DC의 보건복지부 본부에 위치해있다. SAS, Stata 등 분석 소프트웨어도 갖추고 있다. 연방통계연구데이터센터는 미국 인구조사국에 의해 운영되며 미국 전역에 걸쳐 존재한다.

원격접근 시스템을 이용하는 경우에는 개인 수준의 데이터에 접근할 수 없으며 제한된 데이터를 분석할 수 있는 자동 시스템에 코드를 제출하는 방식으로 이루어진다. NCHS가 보유한 모든 데이터셋에 적용되는 것이 아니기 때문에 연구에 필요한 데이터셋에 따라 원격접근이 제한될 수 있다. FTP로 코드를 보내고 이메일로 분석 결과를 받게 된다.

연구데이터센터를 이용할 때는 다음의 규칙을 따라야 한다.

첫째, 분석 및 연구결과물은 승인된 제안서의 범위 내에 있어야 한다. 연구데이터센터는 출판에 필요한 범위로 결과물의 공개를 제한할 수 있다. 분석 계획에 변경이 발생할 경우 연구데이터센터의 분석가와 협의해야 한다.

둘째, 다수의 연구 프로젝트를 진행하고 있을 경우, 한 번에 하나씩만 작업이 허용된다.

셋째, 모든 코드, 결과물, 노트 등은 연구데이터센터에서 가지고 나가기 전에 센터의 분석가(RDC Analyst)에 의해 '공개 검토'를 받아야 한다. 개인 수준의 데이터는 센터의 시설로부터 가지고 나갈 수 없다.

넷째, 연구데이터센터는 시설로 가지고 들어오거나 가지고 나가는 모든 자료를 검색할 권한을 가진다. 휴대전화, 노트북, 카메라 등 전자기기는 센터로 가지고 올 수

없다.

다섯째, 개인이나 기관을 식별할 수 있는 것들을 연구데이터센터로 가지고 올 수 없다. 코드 내에 재식별을 가능하게 하는 내용을 포함하려고 해서는 안 된다.

여섯째, '공개 검토' 전에 연구자들은 잠재적인 공개 위험성이나 불필요한 정보가 없는지 자신의 결과물을 검토해야 한다. 우선, 제출한 결과물의 내용 및 연구 목적을 센터 분석가와 협의할 것이 권고된다. 이에는 출판물에 포함될 실제 테이블들을 포함할 것을 연구자에게 강하게 권고하고 있다. 또한, 결과물은 사람이 읽을 수 있는 평문 파일로 제출해야 한다. 결과물의 추가적인 포맷으로 제출은 센터 분석가의 재량에 따른다. 연구자가 중간 결과물의 검토를 원할 경우 이를 요청할 수 있다. 중간 결과물은 센터 내에서 생성되고 사용될 수 있지만, 공개해서는 안 된다.

일곱째, 공개 검토를 통과하지 못할 금지된 정보를 취득하려고 할 때는 계정이 즉시 정지되고 법적 조치가 취해질 수 있다.

라) 연구결과물 검토

앞서 언급했듯이, 연구결과물은 출판되기 전에 연구데이터센터에 의해 '공개 검토'를 받는다. NCHS는 연구자의 연구결과물 출판을 위한 가이드라인을 제공하고 있으며, 연구자에게 이 가이드라인에 따라 원고를 작성할 것을 권고하고 있다.[134]

많은 저널에서 연구자들의 논문에 접근 가능한 데이터를 포함할 것을 요구하고 있지만, 연구데이터센터의 데이터는 논문에 포함하기 힘들다. 그 대신, 연구데이터센터는 데이터 사전(data dictionary) 및 통합(연계) 절차를 논문에 포함할 것을 연구자에게 권고하고 있다.

또한, 연구결과물에는 개인 혹은 기관을 식별할 수 있는 정보, (해당 지역을 측정할 수 있는 허락을 별도로 받지 않는 한) 응답자가 살고 있는 지역을 식별할 수 있는 정보, 정확한 날짜, 위의 정보들을 의도하지 않게 드러낼 수 있는 메커니즘 등이 포함되어서는 안 된다. NCHS 데이터를 외부의 데이터 소스와 결합할 경우, 지역, 날짜, 개인들의 식별이 가능해질 수 있음을 주의해야 한다.

마) 기밀성 보장

NCHS는 데이터의 기밀성을 보장하기 위해 다음과 같은 세 가지 조치를 취하고 있다. 첫째, 기밀성 교육(confidentiality orientation). 10~20분 정도의 사전 교육을 완료

134) Publication Guidelines https://www.cdc.gov/rdc/b6pubeyond/pub600.htm

해야 한다. 교육 후에 퀴즈를 풀어야 하는데, 만점을 받아야 한다. 둘째, 기밀성 양식. 기밀성 교육 확인증과 함께, 각 프로젝트마다 기밀성 양식을 제출해야 한다. '연구데이터센터의 기밀 데이터 접근 조건에 관한 계약'과 '지정요원자격(designated agent status)' 부여를 위한 개인정보 제공양식이 있다. 셋째, 공개 검토.

(2) Healthdata.gov[135]

미국 보건복지부는 보건의료 데이터에 대한 민간기업, 연구자, 정책결정자에 대한 접근을 촉진하여 그 활용가치를 높이기 위해 Healthdata.gov를 운영하고 있다. 이는 메디케어 및 메디케이드 서비스 센터(CMS), 질병통제예방센터(CDC), 국립보건원(NIH) 등 보건복지부 산하 기관들의 데이터를 더 많이, 더 쉽게 공중에 제공하고자 하는 목적에서 만들어졌다.

Healthdata.gov는 보건의료, 메디케어, 병원, 입원 등 다양한 주제의 데이터셋을 제공하고 있는데, 이는 병상 간호 제공자의 질 정보, 국가적 보건의료 서비스 제공자 정보, 최신의 의학 및 과학적 지식, 소비자 제품 데이터, 공동체의 건강 성과 정보, 정부의 지출 데이터 등을 포괄한다. 또한, 개발자들이 이 데이터를 사용할 수 있도록, 기계 판독이 가능하고 다운로드할 수 있으며, API를 통해 접근할 수 있는 형식으로 제공하고 있다.

보건복지부의 이 프로젝트는 2010년 Health Data Initiative 사업으로부터 시작되었다. 이 사업의 목적은 보건의료 및 그 시스템의 성과에 대한 인식을 높이고, 건강을 증진하기 위한 공동체의 행동을 촉구하기 위한 프로그램의 개발을 위해 혁신가들의 보건의료 데이터 활용을 촉진하기 위한 것이었다. 또한, 이 사업은 보다 폭넓은 차원에서 오바마 정부의 '공개 데이터 정책(Open Data Policy)'의 일환이기도 하다. 오바마 대통령은 2013년 5월 9일, 미 행정관리예산국(OMB)로 하여금 전 연방정부에 걸쳐 공개 데이터 정책을 발표하도록 하는 행정명령을 내렸다. 이는 공공정보의 상호호환성과 개방성을 증진하고 공공정보를 공중에게 보다 접근 가능하도록 만들기 위한 것이었다. 이 데이터들은 오픈 데이터 사이트(data.gov)를 통해 제공되고 있다.

2017년 10월 현재, healthdata.gov는 8400여 자료(resources), 3500여 데이터셋을 제공하고 있다. 이 데이터는 누구나 접근할 수 있도록 공개되어 있기 때문에 개인을 식별할 수 있는 자료는 포함되어 있지 않다. 각 데이터마다 라이선스가 적용되어 있는데, 대부분 오픈데이터커먼즈(Open Data Commons, ODC)의 오픈데이터베이스라이선스(Open Database License, ODbL)가 적용되어 있다.[136]

---

135) 아래 내용은 healthdata.gov 홈페이지를 참조한 것이다. http://healthdata.gov/

Healthdata.gov에는 데이터 연계와 관련한 특별한 메뉴는 없다. 다만, 이 사이트를 통해 제공되는 데이터셋 중에는 이미 연계된 데이터셋도 존재한다. 예를 들어, 메디케어 및 메디케이드 서비스 센터(CMS)의 메디케이드분석추출물(Medicaid Analytic eXtract, MAX)파일과 NCHS 등의 설문조사 데이터의 연계 데이터가 제공되고 있다.

4. 기타 국가들

(1) 독일

독일은 법정 건강보험 시스템을 갖고 있으며, 이는 연금 및 실업급여 등을 포함한 사회보험의 일부이다. 1차 의료에서 전자건강기록(EHR)의 광범위한 이용으로 풍부한 전자 데이터를 보유하고 있다. 다만, 2차 의료에서는 전자건강기록이 다소 제한적으로 이용되고 있다. 또한, 건강보험 지급자가 보유하고 있는 거대한 청구 데이터셋 역시 보유하고 있다. 그러나 병원 진료기록을 보유하고 있는 중앙 데이터베이스는 없다. 전자 환자건강기록에 포함되는 최소한의 기준은 없으며, 보건의료전문가 조직에 의해 정의되고 관리된다. 그리고 주로 직접 진료를 목적으로만 사용된다. (OHE Consulting Report, 2015)

앞서 독일 법제를 다룬 장에서 본 바와 같이, 데이터의 제공은 동의가 있어야 하며 학술적 연구 목적의 데이터 제공은 동의가 불가능하거나 비례적이지 않은 노력이 요구되며 과학적 연구 이익이 프라이버시보다 월등히 중요할 경우에 한해서 제한적인 조건에 따라 이루어진다.

OECD(2013)에 따르면, 데이터 연계 프로젝트는 국가적 수준에서보다 주 수준에서, 그리고 법에 따라 허용된 경우에 수행된다. 연구 프로젝트를 위한 서로 다른 주의 데이터 통합 혹은 암등록 데이터를 다른 데이터 소스와 링크하는 것은 개별 주의 허가를 필요로 한다. 현재 데이터 연계가 수행되는 곳에서 데이터 연계는 제3자를 통한 공통의 가명화 키를 통해 수행된다. 단지 비식별화된 데이터만이 연구자에게 제공된다. (OHE Consulting Report, 2015)

정확 연계(deterministric linkage)를 위해서는 고유식별자가 필요하다. 보건의료 영역에서 독일은 건강보험번호(health insurance number)를 사용하는데, 이는 가입자

136) ODbL 라이선스는 누구나 자유롭게 데이터베이스를 제공, 이용할 수 있고, 데이터베이스로부터 무언가를 생산할 수 있으며, 데이터베이스를 수정, 변환할 수 있도록 허용한다. 다만, 원 데이터베이스의 출처와 라이선스를 표시하고, 수정된 버전에 동일한 라이선스를 적용해야 하며, 재배포할 경우 공유를 제한하는 기술적 조치를 취해서는 안 된다.
https://opendatacommons.org/licenses/odbl/summary/

식별을 위한 불변 부분과 가변 부분으로 구분되어 있다.[137] 2012년 이후 전자건강카드 도입과 함께 불변 부분인 앞 10자리는 평생 변하지 않은 상태로 남아있게 되었다.[138] 독일의 전자건강카드(elektronische Gesundheitskarte, eGK)[139] 도입 계획은 2003년 11월 14일 법정건강보험현대화법에 따라 2006년에 시행하는 것으로 수립된 바 있다. 그러나 카드의 저장소에 환자정보를 저장하는 것에 대한 개인정보 침해 우려가 제기되어 수차례 연기되고, 기능 축소가 발표되기도 했다. 그러나 2013년 10월 현재, 95%의 가입자가 eGK를 가지고 있다고 한다.[140]

eGK의 시행 지연에서도 볼 수 있다시피, 독일은 개인정보 보호에 대한 사회적 인식이 높다. 또 영국 등과 달리 연구 목적의 데이터 연계를 위한 거버넌스 체제도 발전되어 있지 않은 것으로 보인다.

(2) 프랑스

프랑스의 보건의료 시스템은 광범위한 사회보장 시스템의 일부로서 국가적인 건강보험 프로그램에 의해 관리된다. 따라서 건강보험 청구서(claims)에 의해 일상적으로 수집되는 광범한 데이터 소스를 보유하고 있다. 국가건강보험정보시스템(Système National d'Information Inter-Régime de l'Assurance Maladie, SNIIRAM)이 응급 및 1차 의료 관련한 보험청구 데이터를 보유하고 있는데, 여기에는 진료(내용은 제외), 절차, 의약품, 진단 시험(결과는 제외), 의료기기, 환자 개인정보(나이, 생년월일 등) 등이 포함된다. 의학정보시스템프로그램(Programme de Méedicalisation des Systèeme d'Information, PMSI)은 2차 의료와 관련한 국가적인 병원 기록 데이터베이스인데, 퇴원 데이터, 진단, 의학적 절차(medical procedure), 입원 기간, 의약품 및 의료기기 등의 정보를 포함한다. (OHE Consulting Report, 2015)

OECD 보고서(2013)에 따르면, 프랑스에서 몇 개 데이터베이스를 연계하는 프로젝트, 특히 1차 의료 데이터(SNIIRAM)를 입원환자 데이터(PMSI) 혹은 설문조사 데이터(ESPS)와 연계하는 프로젝트가 정기적으로 수행되었다. 그러나 전반적으로 연구 목적의 데이터 연계가 활성화되어 있지는 않다. (OHE Consulting Report, 2015)

그 원인 중 하나는 연계에 사용될 수 있는 식별번호의 부재에 기인한 것으로 보인다. 앞서 프랑스 법제를 다룬 장에서 언급했듯이, 프랑스에서는 모든 시민이 가지고

137) SGB V § 290 (1)

138) https://de.wikipedia.org/wiki/Krankenversichertennummer (2017.11.11 방문)

139) 전자건강카드는 SGB V § 291에서 규정하고 있다.

140) https://de.wikipedia.org/wiki/Elektronische_Gesundheitskarte (2017.11.11 방문)

있는 사회보장번호(NIR)를 병원에서 수집할 수 없으며, 병원에서 사용하는 식별번호는 다양하다. 대신 의료기록 목적의 국가식별번호인 INS가 사용되고 있다.

의료기록 연계를 위해 (익명화된 NIR을 가지고 있는) 건강보험기록과 INS를 가진 의료기록을 연계하는 키를 제3자가 보유하는 방안과 건강보험 시스템이 같은 INS를 사용하는 방안이 제출된 바 있다. 그런데 공공건강 고위급위원회(Haut Conseil de la Sante Publique – HCSP)의 2012년 보고서는 두 번째 방안에 따른 변경을 제안했다고 한다. 이 경우 의료기록과 그 외(의료기록이 아닌) 기록들의 연계가 제한될 수 있다. (OHE Consulting Report, 2015)

앞서 보았듯이, 서로 다른 목적의 데이터베이스의 연계를 위한 자동화된 처리, 그리고 개인건강정보가 필요한 연구 프로젝트의 수행을 위해서는 개인정보 감독기구인 CNIL의 허가가 필요하다. 비정부 연구자는 CNIL의 승인뿐만 아니라 국가통계정보위원회의 승인도 얻어야 한다. 또한, 서로 다른 국가 간에 개인정보의 공유를 포함하는 프로젝트 역시 CNIL의 승인을 받아야 하며, 프랑스와 EU 밖의 국가 간에 데이터가 공유될 경우, 해당 국가가 프랑스에 준하는 개인정보 보호체제를 가지고 있어야 한다. (OHE Consulting Report, 2015)

2013년 OECD 보고서에 따르면, 광범위한 건강보험 데이터에도 불구하고, 동료평가 저널에 의해 측정된 프랑스의 연구결과물은 다른 나라에 비해 적다고 한다. 프랑스에서는 개인건강정보의 프라이버시 보호가 강력히 강조되고 있으며, 연구 목적의 접근은 제한적이다. (OHE Consulting Report, 2015)

제3절 연구 목적 데이터 연계 현황

영국 등 여러 국가에서 연구 및 통계 등 공익목적으로 서로 다른 데이터를 연계할 수 있는 거버넌스 체제를 마련하고 있다. 앞 절에서 살펴본 바와 같이, 이러한 거버넌스 체제는 보건의료 연구 분야를 중심으로 형성되었지만, 점차 보건의료 외 영역으로 확대되고 있다. 각 국가는 정부나 공공기관이 보유한 행정데이터, 혹은 공공 및 민간 영역의 설문조사 데이터를 상호 연계할 수 있는 시스템을 제공하고 있는데, 앞 절에서와같이 보건의료 분야의 데이터 허브가 존재하기도 하고, 통계청이 이러한 역할을 수행하기도 하며, 데이터 연계를 위한 별도의 기관을 두기도 한다. 통계기관은 자신의 고유한 목적인 통계 생산을 위해 데이터 연계를 활용하기도 하며, 외부의 연구자들에게 연구 목적의 데이터 접근 및 연계 서비스를 제공하기도 한다. 이 절에서는 통계기관 외의 연구 목적 데이터 연계 사례들을 검토하고자 하며, 제4절에서는 통계기관을 중심으로 이루어지는 통계 및 연구 목적의 데이터 연계 사례를 검토한다.

1. 영국 ADRN[141)]

영국의 행정데이터연구네트워크(Administrative Data Research Network, ADRN)는 사회, 경제 연구자들에게 안전한 환경에서, 연계된 비식별 행정데이터에 대한 접근을 제공하기 위한 네트워크이다. 이를 통해 사회에 대한 지식과 이해를 증진시키고, 정책결정자에게 도움을 주고자 한다.

ADRN은 2012년 발간된 행정데이터 작업반(Administrative Data Taskforce, ADT)의 보고서에 따라 설립되었다.[142)] 이 보고서는 공공 정책을 지원하고 궁극적으로 사회를 발전시킬 수 있는, 정부 부처와 연구자들의 역량을 강화하기 위해 영국 전역에서 연구에 행정데이터가 보다 효과적으로 사용될 필요가 있다고 지적하며, 사회경제적 연구를 위해 데이터 사용을 원하는 연구자들이 접근할 수 있는 단일한 창구(point of access) 역할을 할 네트워크를 만들 필요성이 있음을 제안하였다. 이 제안에 따라 영국 경제사회연구위원회(ESRC)는 ADRN을 지원하기로 결정하였다.

ADRN은 다음과 같은 역할을 한다.

- 연구자가 승인 패널(Approvals Panel)에 제출할 연구 제안서를 준비하도록 도움을 준다.

---

141) 아래 내용은 영국 ADRN 홈페이지를 참조한 것이다. https://www.adrn.ac.uk/

142) 행정데이터 TF는 경제사회연구위원회(ESRC), 의학연구위원회(MRC), 웰컴트러스트에 의해 만들어진 워킹그룹이다.
http://www.esrc.ac.uk/research/our-research/administrative-data-research-network/administrative-data-taskforce-adt/

- 승인 패널은 연구 프로젝트가 윤리적인지, 합법적인지, 실현 가능한지, 과학적 가치가 있는지, 사회에 이익이 되는지 여부를 결정한다.
- 연구자들이 비식별 데이터를 안전하고 합법적이며 책임감 있게 사용할 수 있도록 훈련을 시키며, 동시에 데이터 보유기관과 협상을 진행한다.
- 훈련된 경제사회 연구자들에게 안전한 환경에서 연계된 비식별 행정데이터에 대한 접근을 제공한다.
- 연구자가 최종결과물을 안전한 환경(safe havens)으로부터 갖고 나가기 이전에, 그 결과물이 승인된 연구 프로젝트에 관련된 것인지, 개인을 직접 식별할 수 있는 정보를 포함하고 있는지 등을 철저하게 검증한다.
- 연구자들은 연구요약을 ADRN에 제공하며, 결과물을 공중 및 관련 데이터 보유기관에 제공한다.

ADRN이 행정데이터를 보유하고 있는 것은 아니다. ADRN은 연구자들을 대신하여 요청할 데이터의 범위를 검토하고, 데이터 보유기관과 협의를 진행하며, 신뢰할 수 있는 제3자(TTP)를 통해 데이터를 연계하고, 연계된 비식별 데이터에 접근할 수 있는 보안 환경을 제공한다.

(1) ADRN의 구조와 거버넌스

ADRN의 네트워크는 영국 내 각 자치국가에 기반을 둔 4개의 행정데이터연구센터(Administrative Data Research Centres, ADRC), 네트워크의 조정 역할을 담당하는 행정데이터서비스(Administrative Data Service, ADS), 그리고 유관 공공기관 등으로 구성되어 있다.

- 행정데이터서비스(ADS) : ADS는 네트워크의 조정 역할을 담당하며, 개인정보 보호를 위한 높은 수준의 표준과 관행을 유지할 수 있도록 안전한 기반시설, 절차, 규약, 훈련, 거버넌스를 보장하는 역할을 한다. 에섹스(Essex) 대학에 기반을 두고 있으며, 맨체스터, 옥스퍼드, 웨스트 잉글랜드, 에딘버러 대학과 파트너 관계를 맺고 있다.
- 행정데이터연구센터(ADRC) : 연구를 위한 전문지식을 제공하며, 국가별로 다음의 대학교가 주도하고 있다.
  - 잉글랜드: 사우스햄튼 대학
  - 북아일랜드: 퀸즈대학교 벨파스트

- 스코틀랜드: 에딘버러 대학
- 웨일즈: 스완지 대학

- 국가별 통계당국 (국가별 ADRC와 파트너쉽을 맺고 있다.)
- 정부 부처 및 기관 (데이터 보유기관)
- 경제사회연구위원회 ESRC (기금제공자)
- 영국 통계청 (네트워크의 이사회 주도)

ADRN의 거버넌스 구조는 자문이사회, 사무국장 그룹, 운영그룹 등으로 구성되어 있다.

- 자문이사회(Advisory Board) : 영국 경제사회연구위원회(ESRC)는 (2018년 9월로 예정된) 투자 기간 동안 ADRN을 감독할 새 이사회 구성했다. 이사회는 더 장기적인 미래에 대해서도 조언한다. ESCR 위원회 멤버인 옥스퍼드 대학교수 melinda Mills가 의장을 맡고 있다.
- 사무국장 그룹(Director Group) : ADRN의 네트워크의 정책, 전략, 자원배분, 혁신을 담당하는 등 네트워크의 관리를 맡고 있다. 각 ADRC와 ADS의 사무국장, ESRC(참관), 이사회 담당자(참관), 운영그룹 의장으로 구성된다. 이 그룹은 ESRC에 보고를 하며, ESRC은 상업혁신기술부에 책임을 지고 있다.
- 운영그룹(Operating Group) : 이 그룹은 일상적인 운영과 사무국장 그룹이 결정한 전략 방향을 어떻게 이행할 것인지에 대한 집단적 의사결정을 담당한다. 각 ADRC에서 2명, ADS 2명으로 구성되며, ESRC의 고위급 정책 관리자도 참여한다.

(2) 행정데이터 활용 절차

행정데이터를 요청하는 연구자는 다음과 같은 자격조건을 만족해야 하며, ① 신청서 작성, ② 승인 패널, ③ 데이터 협상 및 데이터 준비, ④ 분석 시작 등 4단계에 걸친 신청 절차를 밟아야 한다.

가. 자격조건

ADRN에 신청하는 모든 연구 프로젝트는 한 곳 이상의 데이터 소스를 포함한, 비식별 행정데이터에 접근할 필요성이 있음을 입증해야 한다. 모든 데이터 소스가 행정

데이터일 필요는 없으며, 행정데이터와 설문조사 데이터의 연계도 가능하다.

연구 프로젝트는 ① 비영리적 연구 목적이어야 하고, ② 명확한 과학적 이익과 잠재적인 공익이 있음을 입증해야 하며, ③ 연구 과제에 대한 답을 찾기 위해 부서 수준의(unit-level) 행정데이터가 필요하다는 것을 보여주어야 한다.

ADRN은 다른 데이터 서비스, 예를 들어, FARR 연구소, 영국 데이터 서비스(UKDS)의 시큐어랩, 종적 연구소(Longitudinal Studies), HMRC 데이터랩 등에서 더 나은 데이터를 얻을 수 있는지, 그리고 정부 부처의 일반적인 업무의 일부로서 해당 연구가 수행되지는 않는지 등을 점검한다.

연구자의 자격 요건도 규정하고 있다. 우선 연구자는 영국 연구기금위원회(Research Councils UK funding)의 자격 요건을 갖춘, 학계, 공공영역, 공동체 혹은 자원봉사 영역, 연구기관 소속이어야 한다. 또한, 독립적으로 혹은 적절한 감독자나 연구팀장의 지시 하에 연구를 수행할 수 있어야 한다. SURE 훈련에 참여해야 하며, 이용약관과 위반정책에 서명해야 한다. 연구자는 자신의 기관에 소속된 기관 보증인의 이름을 제공해야 하며, 기관 보증인은 신청서에 서명해야 한다. SURE 훈련은 안전한 연구 데이터 이용자 환경 훈련(Safe Users of Research data Environment Training)의 약자이다. 이 훈련은 연구자가 행정데이터의 활용과 관련한 민감성을 이해하고, 민감 데이터를 안전하고 합법적이며 책임성 있게 다루는 방법을 훈련시킨다. 하루 일정의 교육과정이며 완료 후 테스트를 받아야 한다. 이 훈련은 ADRC에서 제공한다.

나. 신청서 작성

연구자는 자신의 연구 제안서를 갖고 ADRN의 이용자 서비스팀에 연락한다. 이용자 서비스팀은 우선 다른 기관을 통해서 필요한 데이터에 접근할 수 있는지 검토한다. 다른 곳에서 데이터를 얻을 수 없다면, ADRN 직원은 연구 제안서를 검토한 후 신청서 작성을 위해 프로젝트에 대해 연구자와 협의한다. ADRN 직원이 프로젝트에 대해 명확하게 이해하고, 연구자가 ADRN에 무엇을 기대하는지 정확히 알게 되면, 프로젝트 신청양식을 보내준다. 이는 승인 패널이 프로젝트를 심사하고, ADRN이 데이터 보유자들과 협상하기 위한 목적으로 이용된다. ADRN은 프로젝트 신청양식과 함께 훈련을 위한 연구자 신청양식을 보내준다.

다. 승인패널(Approvals Panel)

승인 패널은 독립적 전문가 및 비전문가로 구성된다.[143] 승인 패널은 매월 한번 회

의를 한다. 승인 패널은 해당 연구 프로젝트의 실현 가능성, 학술적 가치, 공익성, 프라이버시에 대한 영향 등의 기준으로 평가한다.

또한, 적절한 윤리평가를 수행했는지 점검한다. 필요하다면, 국가통계학자데이터윤리자문위원회(National Statistician's Data Ethics Advisory Committee, NSDEC)를 이용할 수 있다.

승인 패널의 결정이 내려지면, ADRN 직원은 이를 연구자에게 통보한다. 보류 결정이 내려지고 추가적인 정보가 필요하다면, ADRN은 연구자를 도와 신청서를 보완하도록 한다. 프로젝트가 기각되면 항소 절차를 알려주며, 승인되면 SURE 훈련을 받도록 한다.

라. 데이터 협상 및 데이터 준비

ADRN은 연구자를 대신하여 데이터를 보유한 정부 부처에 연락하여 협상을 진행하며 그들과 데이터 공유협약을 맺는다. 데이터 보유기관인 정부 부처는 어떠한 제한을 설정하거나 데이터의 안전성을 위한 추가적인 요청을 할 수 있다.

이후 ADRN 직원은 서로 다른 정부 부처의 데이터를 받아 연계를 수행하는데 TTP를 통해 적절히 비식별 조치가 이루어지도록 보장한다.

마. 분석 단계

데이터가 연계되고 비식별화되면 훈련을 마친 연구자들은 보안 환경에서 데이터에 접근할 수 있는데, 이 시설은 각 국가의 ADRC에서 제공한다. 이 시설들은 연구자에게 데이터 분석을 위한 통계 소프트웨어의 사용만을 허락한다. 연구자들은 방에서 핸드폰, 메모리 카드, 볼펜이나 종이 등 아무것도 갖고 들어가거나 갖고 나올 수 없다. 또한, 데이터를 복제하거나 다운로드 받거나 전송할 수 없다. 다만, ADRN이나 데이터 보유자의 승인이 있는 경우에는 자기 기관의 시설을 이용할 수도 있다. 연구자들이 분석을 끝내면, ADRN 직원들은 최종결과물을 연구자에게 제공하기 전에, 반드시 연구 자료들을 안전하게 삭제해야 한다. 연구자들은 데이터셋 자체를 보안 환경에서 가지고 나갈 수는 없다.

143) 승인 패널은 의장, 통계청(ONS) 등 데이터 제공자 3명, 학계 3명, 프라이버시 전문가 1명, 일반인 2명으로 구성된다. https://www.adrn.ac.uk/get-data/approvals-panel/

그림 3-8 ADRN의 안전시설 위치

* 출처: ADRN 홈페이지 https://www.adrn.ac.uk/get-data/secure-access/

(3) 데이터 연계

개인정보 보호를 위해, 서로 다른 데이터의 연계는 다음 절차를 따라 이루어진다.

① 1 단계 : 연구 제안서의 승인이 이루어지고 연구자가 훈련을 받은 후, ADRN은 프로젝트와 관련된 데이터의 제공을 위해 데이터 보유기관과 협상을 진행한다.

② 2 단계 : 데이터 보유기관(데이터를 보유한 정부 부처)은 각 레코드에 고유한 참조번호(reference number)를 부여한다. 그리고 이름, 생년월일 등 사람들을 직접 식별할 수 있는 식별자를 분리한다.

③ 3-1 단계: 데이터 보유기관들은 식별정보를 고유 참조번호로 대체한 데이터를 ADRC 중 하나로 보낸다.

④ 3-2 단계 : 동시에, 직접 개인을 식별할 수 있는 정보는 각 레코드의 고유 참조번호와 함께 신뢰할 수 있는 제3자(TTP)에게 보낸다. 그러나 연구 데이터는 포함되지 않는다.

⑤ 4 단계 : TTP는 고유 참조번호와 식별정보를 사용하여 이 정보들을 매칭한다. 그리고 개인 식별정보를 삭제한 후 매칭된 고유 참조번호만을 남긴다.

⑥ 5 단계: 색인키(index key)는 서로 다른 데이터 집합에서 어떤 참조번호가 같은 사람과 관련되는지를 보여준다. TTP는 색인키를 ADRC에 보낸다.

⑦ 6 단계: ADRC는 색인키를 사용하여 서로 다른 기관들이 보내온 데이터 집합을 연계한다. 그리고 색인키와 참조번호를 지운 후에 연구자에게 연계된 데이터에 대

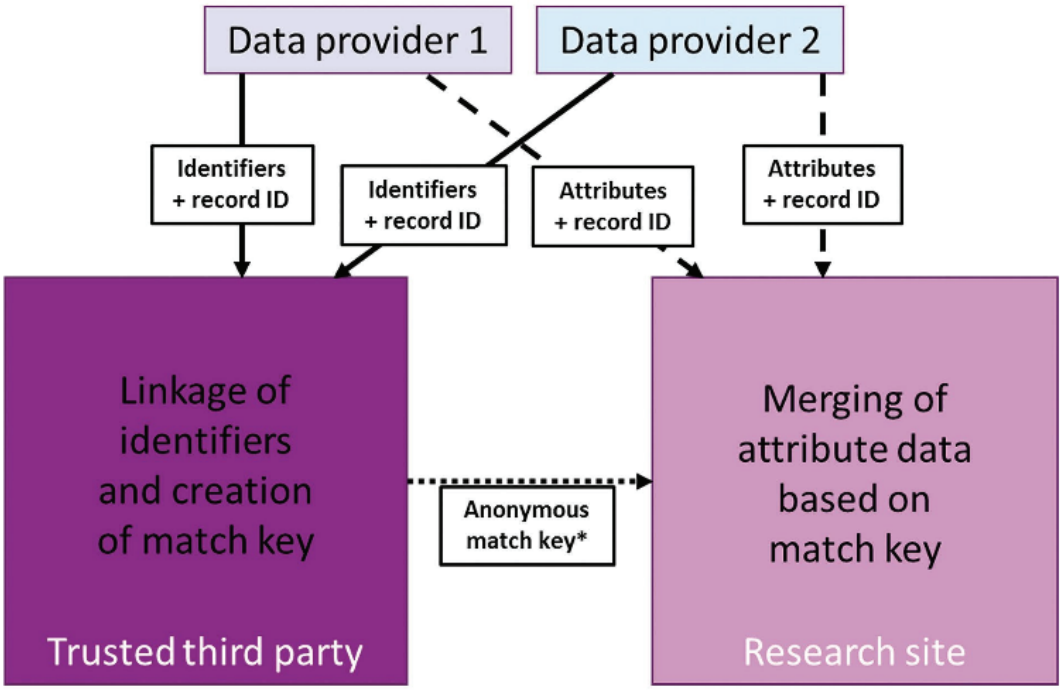
한 접근을 제공한다.

이 시스템은 개인 식별정보와 연구 데이터의 분리를 유지한다. 즉, TTP는 단지 식별정보와 참조번호만을 볼 수 있으며, 연구 데이터를 볼 수 없다. ADRN 직원은 단지 연구 데이터와 색인키만을 볼 수 있을 뿐, 개인 식별정보는 볼 수 없다. 연구자는 보안시설에서 자신이 요청한 데이터만을 볼 수 있으며, 색인키와 개인 식별정보는 볼 수 없다.

여기서 신뢰할 수 있는 제3자(TTP)는 데이터 매칭을 위한 보안시설을 가지고 있는 조직을 의미하는데, 국가통계청(Office for National Statistics, ONS)이나 북아일랜드 통계연구소(Northern Ireland Statistics and Research Agency, NISRA) 등이 이에 해당한다.

만일 어떤 기관이 TTP로서의 역할과 ADRC의 파트너 역할을 동시에 수행하고 있다면, 두 업무는 철저하게 분리되어야 한다. 해당 조직의 서로 다른 부서에서 완전히 다른 직원들이 각각의 역할을 수행해야 한다. 이러한 역할 분리를 통해 데이터 기밀성이 유지될 수 있도록 기술적, 운영적 통제를 해야 한다.

그림 3-9 ADRN 식별자와 속성 데이터의 분리



* 출처: Katie Harron (2016)

(4) 보안과 개인정보 보호

ADRN은 영국 데이터 서비스가 만든 '데이터 공유를 위한 5가지 안전 원칙'에 따라 운영된다.

- Safe people : 연구자의 자격 요건 검증 및 훈련
- Safe project : 독립된 승인 패널에 의한 심사 및 윤리평가
- Safe environment : 안전한 공간에서 연구가 수행될 수 있도록 보안시설 운영
- Safe data : 데이터 비식별화 및 TTP에 의한 데이터 연계
- Safe results : 연구결과물에 대해 엄격한 통계적 노출 검토(statistical disclosure checks) 수행

2. 독일 GRLC / FDZ

독일은 연방 구조이기 때문에, 대부분의 데이터베이스는 각 연방 주(federal state)에 흩어져있다. 그래서 전체 인구를 포괄하는 행정데이터베이스는 드물다. 단지 사회보장 행정 등록소만이 전체 인구를 포괄하는 데이터베이스로 이용되어 왔는데, 이조차도 완전한 것은 아니다. 독일에서는 암 등록소를 포함한 의학 연구 데이터베이스조차 국가 전체가 아니라 연방 주 내에서 운영된다. (Antoni 외, 2017)

독일레코드연계센터(German Record Linkage Center, GRLC)는 사회과학 분야의 학술 연구를 위해, 행정데이터를 이용한 데이터 연계 활성화를 목적으로 2011년에 설립되었다.

GRLC는 독일연방고용국(BA)의 연구데이터센터(FDZ)와 협력 관계를 맺고 있는데, FDZ는 연구자에게 사회보장 및 고용 분야에서 비영리적 실증 연구를 위한 마이크로데이터에 대한 접근을 제공한다.

(1) 독일레코드연계센터 GRLC

독일에서는 연방 구조로 인한 행정데이터베이스의 분산과 강력한 개인정보보호법에 의해, 서로 다른 데이터의 연계를 위한 법적 요구조건을 만족시키기 위해 프로젝트별로 협의가 되어야 했다. 아직 독일에서는 일반적인 목적의 중앙 데이터 연계기관은 없다. 서로 다른 연방 주들 사이의 실질적이고 정기적인 데이터 연계가 없기 때문에 데이터 보유기관들은 종종 연구를 위한 데이터 신청을 거부하는데, 데이터 연계를 위

한 모델이 없다는 것도 하나의 장애물이다. 그래서 GRLC는 독일 내에서 데이터 연계 활성화를 위한 모델을 만들고자 노력해왔다.

사회과학 분야의 데이터 연계를 활성화하기 위하여 라이너 슈넬(Rainer Schnell)과 스테판 벤더(Stefan Bender)는 2010년에 독일연구재단(German Research Foundation, DFG)에 연구 기금을 신청했다. 라이너 슈넬은 뒤스부르크-에센(Duisburg-Essen) 대학교 교수이며, 스테판 벤더는 고용연구소(Institute for Employment Research, IAB)에 있는 독일연방고용국(BA)의 연구데이터센터(FDZ)의 전 책임자이다. 이들은 2011년에 기금을 받아 독일에서 처음으로 학술 연구 목적의 데이터 연계 센터인 GRLC를 설립하였다. 이 단체는 데이터 연계 방법에 관한 연구와 개발을 하면서, IAB를 통해 학술기관의 의뢰를 받아 상담 및 데이터 연계 수행과 같은 서비스를 제공한다. (Antoni 외, 2017)

GRLC의 활동은 뉘렘베르크의 FDZ와 뒤스부르크-에센 대학 등 두 지역에서 이루어진다. FDZ팀은 데이터 연계 서비스 제공에, 뒤스부르크-에센 대학팀은 데이터 연계와 관련한 연구에 초점을 맞추고 있다. (Johanna Eberle, 2014)

GRLC는 프라이버시를 보호하면서도 효과적으로 데이터 연계를 수행할 방법에 대한 연구를 하고 있는데, 이 연구 분야를 프라이버시 보호 데이터 연계(privacy preserving record linkage, PPRL)라고 한다. 또한, 뒤스부르크-에센 대학의 슈넬 교수 연구팀은 Merge Toolbox(MTB)라는 데이터 연계를 위한 자바 프로그램을 개발해왔는데, PPPL 루틴 역시 이 프로그램에 포함되어 있다.

데이터 연계 알고리즘에 관한 연구 및 프로그램 개발 외에 GRLC는 데이터 연계 서비스의 제공, 데이터 연계 프로젝트의 설계 및 구현을 위한 자문, 워크샵과 강의 등을 통한 데이터 연계와 관련한 훈련, 데이터 연계에 대한 온라인 정보 포털 운영 등 다양한 서비스를 제공하고 있다. 현재 GRLC의 홈페이지에는 GRLC의 연구 논문, 수행한 데이터 연계 프로젝트 목록, MTB와 같은 데이터 연계 프로그램 등을 제공하고 있다.

GRLC는 다른 학술기관 등 제3자를 대신해서 수많은 데이터 연계 프로젝트를 수행해왔다. 연계를 위한 데이터 소스는 IAB의 행정데이터를 포함하여 다양하다. 연계 프로젝트별로 원 데이터베이스의 소유권이나 법적 요건이 다르기 때문에, GRLC 홈페이지에 데이터 연계를 위한 일반화된 절차가 나와 있지는 않다. 일반적으로 GRLC에 데이터 연계를 요청하는 기관이 연계 데이터의 모든 요소에 대한 권한을 갖고 있거나 이용허락을 받았다면, 연계 데이터에 대한 접근에 별다른 제한이 발생하지 않는다. 그러나 프로젝트에 따라 연계 데이터의 일부 혹은 전체 구성요소에 대한 접근이 제한될 수도 있다. (Antoni 외, 2017)

FDZ에서의 데이터 연계 작업은 GRLC에 의해서 수행되거나 GRLC의 자문을 받아 FDZ 직원이 수행한다. FDZ의 연계된 표준 데이터셋에 대한 접근은 FDZ의 보안 인프라를 통해 제공된다. (FDZ에서의 데이터 연계에 대한 자세한 내용은 아래에서 설명하도록 한다.) FDZ의 표준 데이터셋은 학술 연구자에게 무료로 접근이 제공되지만, GRLC가 요청을 받아 연계한 데이터셋에는 적용되지 않는다. 앞서 얘기했듯이, 연계 데이터에 대한 모든 요소에 대한 권한을 가진 연구자에 의한 접근은 데이터 연계에 의한 영향을 받지 않는다. 이 경우 연계 데이터는 해당 데이터에 의해 허용되는 모든 환경에서 저장, 접근될 수 있다. 그러나 IAB의 데이터와 외부 데이터가 연계될 경우, 연계 데이터는 FDZ의 안전한 컴퓨팅 환경을 벗어날 수 없다. 또한, 해당 연구결과물 역시 FDZ에 의해 엄격하게 검토된다.

(2) 독일연방고용국 연구데이터센터(FDZ)[144]

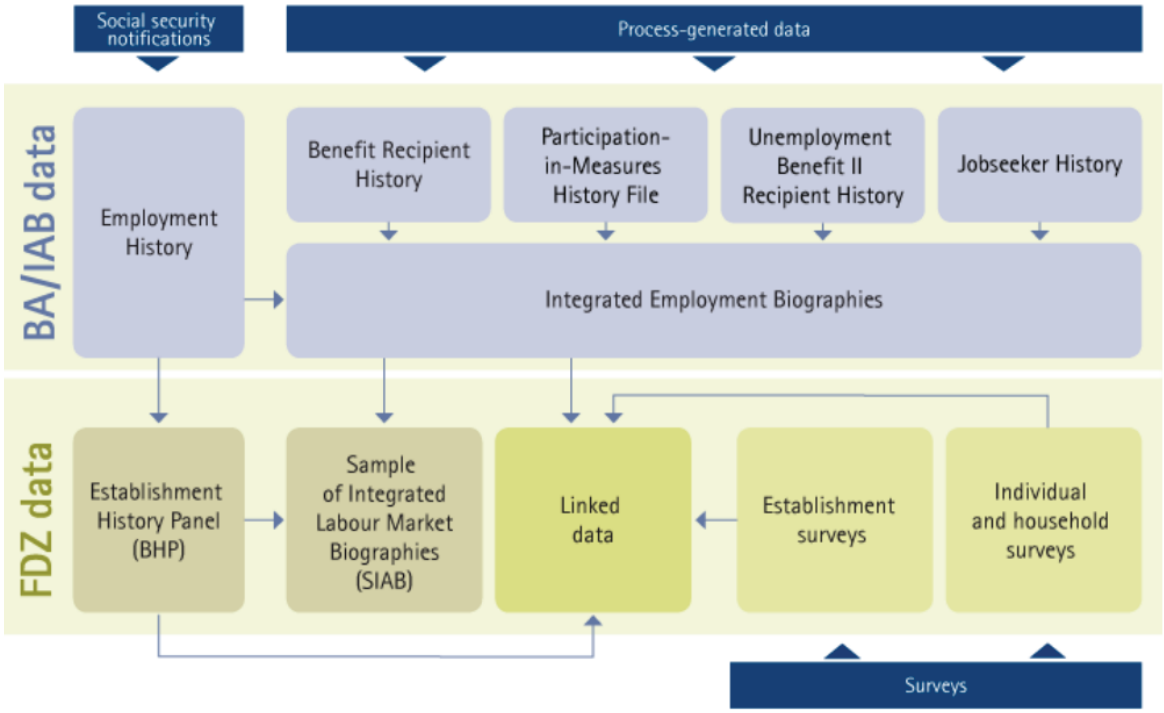
가. FDZ의 데이터

고용연구소(Institute for Employment Research, IAB)에 있는, 독일 연방고용국(BA)의 연구데이터센터(FDZ)는 연구자에게 사회보장 및 고용 분야에서 비영리적 실증 연구를 위한 마이크로데이터에 대한 접근을 제공한다.

FDZ는 개인, 가정, 기관(establishment)에 대한 데이터뿐만 아니라, 기관 및 개인정보 양자로 구성된 데이터를 제공한다. FDZ의 데이터는 세 가지 소스로부터 온다. 사회보장 시스템의 통지 과정 및 연방고용국의 내부 절차 과정으로부터 데이터가 생성된다. 또한, IAB는 자체 설문조사를 수행하여 데이터를 획득한다.

144) 아래 내용은 FDZ 홈페이지를 참조한 것이다. http://fdz.iab.de/en.aspx

그림 3-10 FDZ에서의 데이터 수집 과정



* 출처: FDZ 홈페이지 http://fdz.iab.de/en/FDZ_Overview_of_Data.aspx

사회보장 통지는 IAB에 기록되고, 고용 이력(Employment History)이라는 이력 데이터셋의 형태로 편집된다. 기관 이력 패널(Establishment History Panel, BHP)은 고용 이력을 기관 수준에서 통합하여 생성된다. IAB는 연방고용국의 내부 절차로부터 생성되는 데이터를 보유하고 있다. 이 데이터들은 고용 이력과 연계 가능하며, '통합 고용 전기(Integrated Employment Biographies)'를 형성한다. 이것이 FDZ 데이터셋의 핵심적인 구성요소를 이룬다.

나. FDZ에서의 데이터 연계

과정 생성 데이터(process-generated data)가 설문조사 데이터와 결합하여 연계 데이터셋이 생성된다. 이를 통해 서로 다른 소스로부터 얻어진 기관, 개인, 가정 데이터가 결합된다. 현재 FDZ는 IAB의 연계된 고용주-고용인 데이터 (LIAB) 등 다양한 연계 데이터셋을 제공하고 있다.

다. 데이터 접근

FDZ는 연구자에게 데이터 접근을 위해, 현장 이용(On-site Use), 원격 데이터 접근

(Remote Data Access / JoSuA), 학술적 이용파일(Scientific Use Files) 등 세 가지 방법을 제공한다. 이 세 가지 방법은 데이터의 익명화 정도 및 이용조건에서 차이가 있다. 일반적으로 데이터에 대한 접근은 비영리적 연구 목적으로 제한되며, 데이터 접근 신청은 접수 순서대로 처리된다.

가) 현장 이용 (on-site use)

약하게 익명화된 데이터의 이용은 개인정보 보호 법제의 규제를 받는다. 이 규정에 따라 데이터는 단지 FDZ가 방문 연구자를 위해 제공한 작업 공간에서만 분석될 수 있다. FDZ는 뉘른베르크(Nuremberg)에 위치한 곳에서만 직원들이 자문 서비스를 제공한다. 이 작업 공간은 독일 내에서는 뉘른베르크, 베를린, 브레멘, 드레스덴, 뒤셀도르프, 하노버, 만하임에 위치해있으며, 해외에서는 미국의 앤아버, 코넬, 버클리, 하버드, 로스앤젤레스, 프린스턴, 영국의 에섹스, 런던에 위치해있다.

데이터에 대한 접근 권한을 얻기 위해 다음과 같은 절차를 거친다.

① 신청서 작성 : 방문 연구자로 방문 신청을 하기 위해서는 '현장 이용' 신청서를 작성해야 한다. 제출된 신청서는 독일사회법전(SGB) X, 75조 조항을 준수하고 있는지에 대해 FDZ 가 연방노동사회부의 협력을 얻어 검증한다.

② 이용 계약 : 요청이 승인되면, 연구기관과 FDZ 사이에 연구 계약이 체결된다. 이 계약은 연구자에게 요청한 데이터를 '현장 이용'을 통해 분석할 권한을 부여한다.

③ 지정 : 계약 체결 후에 연구자는 해당 장소에 방문할 약속을 잡을 수 있다. 방문 날짜는 통상 2주 전에 통보된다.

④ '현장 이용' 준비 : FDZ에서의 '현장 이용'은 FDZ가 제공한 가이드라인을 따라야 한다.[145] 연구자들은 FDZ가 제공한 테스트 데이터와 문서를 바탕으로 방문 전에 미리 프로그램을 준비할 것이 권고된다. 방문 전에 프로그램을 JoSuA(Job Submission Application) 소프트웨어를 통해 업로드한다.

⑤ '현장 이용' 후 : 결과물은 개인정보 보호 법제 준수 여부를 검토받게 된다. 연구기간 이후에는 원격 데이터 접근을 통해 프로젝트를 위한 추가 분석을 할 수 있다.

⑥ 연구결과물의 공개 : FDZ 데이터에 기반한 연구자의 출판물은 사용된 데이터셋에 대한 정보(이터셋의 이름과 데이터 접근방법 등)를 포함해야 한다. 모든 출판물의 복사본을 디지털 혹은 출력물 형태로 FDZ에 제출해야 한다.

---

145) 현장 이용 및 원격 데이터 접근에 관한 FDZ 가이드라인 (Guidelines of the FDZ of the BA at the IAB as to Remote Data Access and On-site Use)
http://doku.iab.de/fdz/access/Vorgaben_DAFE_EN.PDF

나) 원격 데이터 접근 (Remote Data Access)

원격 데이터 접근은 연구자가 테스트 데이터를 기반으로 데이터 분석 통계 소프트웨어인 Stata로 프로그램을 개발하는 것을 의미한다. 여기서는 노동연구소(Study of Labor, IZA)가 개발한 JoSuA(Job Submission Application) 소프트웨어를 이용한다. 이용자는 JoSuA를 이용해 프로그램을 업로드한다. 개인정보 보호 법제의 준수 여부를 검증한 후 결과물도 JoSuA에서 접근할 수 있다.

데이터에 대한 접근 권한을 얻기 위해 다음과 같은 절차를 거친다. 우선 '원격 데이터 접근' 신청양식을 작성해서 제출한다. 신청서가 승인되면, 연구자와 FDZ 사이에 이용 계약이 체결된다. 이 계약에 따라 연구자에게 JoSuA를 통해 요청한 데이터를 분석할 자격이 주어진다. FDZ 데이터에 기반한 연구자의 출판물은 사용된 데이터셋에 대한 정보를 포함해야 하며, 모든 출판물의 복사본은 FDZ에 제출해야 한다.

다) 학술적 이용 파일(Scientific Use Files, SUF)

SUF는 분석을 위한 학술기관의 연구자들에게 제공되는 사실상 익명화된 데이터셋이다. 기밀성 요건 때문에 SUF는 더 적은 수의, 더 총계화된 변수들을 포함하고 있다.

SUF를 신청하기 위해서는 우선 신청서를 작성해야 한다. 신청서가 검증된 후에 FDZ는 연구자에게 데이터 제공을 허용할지 여부를 결정한다. 허가가 되면, 데이터 이용 계약이 이용자에게 보내지는데, 이를 작성해서 서명한 후에 FDZ에 보내면 된다. 이후 보안 인터넷 연결을 통해 교환 서버에서 SUF를 다운로드 받을 수 있다. FDZ 데이터에 기반한 연구자의 출판물은 사용된 데이터셋에 대한 정보를 포함해야 하며, 모든 출판물의 복사본은 FDZ에 제출해야 한다.

데이터 이용은 이용 계약 기간 내로 제한되기 때문에, 데이터의 모든 추출물 및 복사본은 삭제해야 한다. 기관장 혹은 데이터보호관에 의해 서명된 데이터 삭제 확인서를 FDZ에 제출해야 한다.

라. 관련 법제

FDZ는 개인 수준의 데이터 및 기관 데이터를 제공하는데, 이는 개인정보 보호 법제의 규제를 받는다. 데이터 접근에 대해 기본적으로는 독일사회법전(SGB) X의 67조가 주로 적용된다. 이 규제에 따라 FDZ는 연구자에게 데이터 접근을 위한 세 가지 방법을 제공한다. 각각의 경우에 데이터셋 분석을 위한 요구조건은 다음과 같다.

가) 현장 이용(On-site use)

FDZ가 제공하는 개인 데이터는 SGB X, 67조 (1)에 기반한 '사회 데이터(social data)'에 해당한다. 이 데이터는 자연인을 식별하거나 식별할 수 있는, 개인 혹은 물질적 환경에 대한 개인정보를 포함한다. SGB II 및 III에서 규정한 연방고용국(BA)의 책임 범위 내에서, 데이터는 BA에 의해 수집 및 처리될 수 있고, 이에 따라 고용연구소(IAB)에서의 연구 목적으로 편집될 수 있다. 사회 데이터는 특히 민감하며, 따라서 SGB I, 35조 (1)의 기밀성 규제를 받는다. 이에 따라 사회 데이터에 대한 접근은 독일 SGB 조항의 적용을 받게 된다. 기관 데이터 또한 사회 데이터로 취급된다.

SGB X, 75조에 따라 연구자들은 다음 조건에서 사회 데이터에 대한 접근이 허용된다. 첫째 사회적 이익 혹은 사회적 노동시장 연구를 위한 연구 프로젝트에 필요한 데이터, 둘째 기밀성에 대한 개인의 이익보다 훨씬 큰 공공 연구의 이익, 셋째 합리적인 상황 하에서 사회 데이터의 이전에 개인의 허가를 받기 힘든 경우, 넷째 데이터의 이용은 연방노동사회부(Federal Ministry of Labour and Social Affairs)의 허가를 받아야 한다. 또한, 데이터는 상업적 목적으로 이용되어서는 안 된다.

나) 원격 데이터 접근

원격 데이터 접근은 공익을 위한 과학적 연구 프로젝트에 허용된다. 상업적 배경이 있는 요청은 고려되지 않는다. 의심이 드는 경우, 프로젝트에 대한 더 상세한 설명이 요청된다.

다) 학술적 이용 파일 (Scientific Use Files)

학술적 이용 파일은 사실상 익명화된 데이터셋이다. SGB III 7장 282조에서 규정하고 있는 학술기관에 제공된다. 데이터는 여전히 개인 데이터이기 때문에 다음의 조건을 만족해야 한다. 첫째 데이터는 노동시장연구 분야의 임시적 연구를 위해서만 분석되어야 한다. 둘째 데이터는 독립적 과학적 연구에만 사용되어야 한다. 셋째 연구 목적이 다른 데이터셋의 분석을 통해서는 수행될 수 없다. 넷째 조직적, 기술적 조치에 의해 데이터 기밀성이 보장되어야 한다. 다섯째 데이터가 특정 연구 프로젝트의 이용계약서에 명기된 사람에 의해서만 이용되어야 한다. 다른 데이터셋과의 연계나 재식별하려는 시도는 금지된다. 여섯째 데이터에 접근하려는 사람은 연방 개인정보보호법 섹션 5[146]를 준수하며, 공무원이거나 정부 업무를 특별히 수행할 의무가 있는 사람

146) 독일 연방 개인정보보호법 section 5는 '데이터 처리를 위해 고용된 사람은 허가 없이 개인정보를 수집, 처리, 사용해서는 안 된다'는 데이터 기밀성을 규정하고 있다.

혹은 그에 따른 지시를 받거나 의무를 수행하는 사람이어야 한다.

라) 기밀성

FDZ 직원은 방문자의 연구 이슈 및 방법에 대해 알고 있는 사실은 자문 서비스 제공, FDZ 서비스의 개선, 개인정보 보호 규제 준수를 위한 목적 외로 사용해서는 안 된다. IAB 및 BA의 다른 직원은 방문자 연구 활동에 대한 정보에 접근할 수 없다. 그러나 이 규정은 방문 연구자와 IAB 혹은 FDZ 연구자의 협력 연구에는 적용되지 아니한다.

제4절 통계 목적 데이터 연계 현황

1. 미국 Data Linkage Infrastructure[147)]

미국 인구조사국(Census Bureau)은 통계 생성을 위해 광범위한 행정데이터를 이용하고 있다. 또한, 인구조사 및 설문조사 데이터를 포함한 다양한 소스의 데이터를 연계하고 통합할 권한을 가지고 있다. 외부의 데이터를 사용함으로써 응답자의 부담을 줄이고 통계 및 분석의 범위와 깊이를 증가시킬 수 있다. 인구조사국은 1940년대부터 외부의 데이터 소스를 활용해왔다.

인구조사국은 평가자 및 정책 분석가의 행정데이터 접근을 증진하기 위해 데이터 연계 기반(Data linkage infrastructure)을 확대해왔다. 연방 및 연방의 지원을 받은 행정데이터 소스를 신속하게 획득하고, 데이터의 문서화와 연계 기술을 증진시키며, 거버넌스, 프라이버시 보호, 데이터에 대한 보안 접근을 위한 시스템을 활용해왔다. 이러한 인구조사국의 활동은 효과적인 정책 결정을 지원하기 위해 적시에 편향되지 않은 데이터를 제공함으로써 연방통계시스템의 목적 달성을 지원한다.

미국은 2016년 3월, 증거기반 정책결정위원회법(Evidence-Based Policymaking Commission Act of 2016)을 통과시켰는데, 데이터 연계기반은 이를 지원한다. 이 법은 연방 프로그램에 대한 행정데이터, 설문조사 데이터, 일련의 관련 통계 데이터를 프로그램 평가, 비용편익 분석, 정책 관련 연구를 위해 접근할 수 있도록 포괄적인 연구를 수행하는 위원회를 설립하도록 하고 있다.

인구조사국은 통계 목적으로 정부 기록에 접근하고 데이터를 수집할 수 있는 권한을 가지고 있다. 데이터 연계기반은 정책 분석과 연구를 위한 데이터를 확보하고 안전하게 분석적인 접근을 할 수 있도록 한다. 데이터 연계기반은 정부 기록을 확보하기 위한 규약, 파일에 대한 메타데이터와 문서화, 연계 방법 및 결과, 데이터 저장(warehousing) 및 제공, 기존 연계 결과의 공유 등을 포함한다.

(1) 데이터 연계 방법

파일이 입수되어 인구조사국에 이전되면, 파일 콘텐츠 목록화 책임을 가진 소수의 직원만이 이에 접근할 수 있으며 기본적인 품질관리(Quality control) 점검을 한다. 이 직원은 보안이 된 물리적 공간에서 작업하며, 컴퓨터 자원은 인구조사국 방화벽 내의

147) 아래 내용은 미국 인구조사국 홈페이지의 Data Linkage Infrastructure를 참조한 것이다. https://www.census.gov/about/adrm/linkage.html

고도로 제한된 공간에 위치해있다. 데이터 처리 및 비식별화 직원은 해당 파일이 법적 계약 요건에 맞는지 여부를 확인한다. 또한, 변수 및 문서들이 그것을 사용할 수 있는 기본적 무결성을 갖고 있는지 확인한다.

서로 다른 데이터셋을 연계하기 위해, 인구조사국 행정기록연구응용센터(Center for Administrative Records Research and Applications, CARRA)는 개인식별 확인시스템(Person identification Validation System, PVS)을 통해 서로 다른 개인식별을 위한 식별 보호키(Protected Identification Key, PIK)을 부여하는데 이 PIK가 데이터 연계에 이용된다.

PVS는 '입수한 파일(incoming file)'을 '참조 파일(reference file)'과 매칭하기 위해, '입수한 파일'에 포함된 이름, 주소, 생년월일, 사회보장번호와 같은 개인 식별번호에 기반한 확률적 연계 방법을 활용한다. PVS 시스템을 통한 매칭은 '참조파일'이 필요한데, 참조파일은 사회보장국(Social Security Administration, SSA)의 숫자식별파일(Numerical Identification file, SSA Numident)로부터 만들어진다. 이 숫자식별파일은 하나의 사회보장번호(SSN)에 대응한 모든 기록을 가지고 있는데, 이로부터 인구조사국의 숫자식별파일(Census Numident), 즉 참조파일이 만들어진다. 참조파일은 각 SSN에 해당하는 하나의 레코드를 보유하고 있으며, 생년월일과 이름 등 모든 변종을 별도의 파일로 보유한다. 상응하는 개인 레코드에는 고유한 식별보호키(PIK)가 부여되며, 이것이 PVS 시스템을 이용하는 모든 파일의 개인 연계키로 활용되는 것이다. '입수한 파일'과 '참조파일'의 연계가 이루어지면, '입수한 파일'에 PIK가 덧붙여진다. (Deborah Wagner, Mary Layne, 2014)

또한, CARRA는 주소 매칭도 할 수 있다. 인구조사국의 경제연구센터(Center for Economic Studies, CES)는 서로 다른 데이터셋의 기업들을 연계하는데 조세 ID, 고용인식별번호(Employer Identification Number, EIN)을 이용한다. 연계절차 자체는 일회적인 서비스이다. 데이터 획득과 연계 계약의 실행에는 5-12개월 정도가 소요된다.

(2) 데이터 접근 절차

가. 신청서 제출 및 심사

데이터 연계기반의 데이터를 사용하기 위해서, 연구자는 우선 제안서를 제출해야 한다. 제안서에는 프로젝트의 방법론과 목적, 예상 결과물, 인구조사국에 갖는 가치, 필요한 데이터셋 등의 내용이 포함되어야 한다. 이 프로젝트가 외부의 데이터를 데이터 연계기반으로 가져올 경우, 프로젝트 제안자는 그 외부 데이터의 이용과 이전을 허가하는 (특히 개인식별정보의 이용을 명시한) 서신을 제출해야 한다.

제안서는 학술적인 가치, 실행 가능성, 잠재적인 공개 위험성 등이 평가된다. 인구조사국의 데이터(Title 13 데이터) 이용을 요청하는 프로젝트의 경우에는 Title 13의 가치, 즉 해당 연구가 인구조사국에 갖는 가치를 입증해야 한다.

승인 절차는 인구조사국의 정책조정부(Census Bureau's Policy Coordination Office)가 관리한다. 평가는 정책조정부 대표, 센서스 소재 전문가(Census Subject Matter Experts, SMEs), 집행부장(Executive-level Division Chief), 지역 정보 보유자 등이 수행한다. 요청하는 데이터가 Title 13의 데이터가 아닐 경우, 인구조사국은 (해당 기관과 인구조사국의 협약에 따라) 다른 데이터 보유기관의 승인을 지원한다.

나. 데이터 이용자에 대한 훈련

행정 기록 프로젝트에 관한 연구자는 인구조사국의 현재의 피고용인이거나, 특별선서지위(Special Sworn Status, SSS)를 획득해야 한다. 인구조사국은 자신의 프로그램에 명백히 이익이 되는 작업을 수행하는 개인들에 이 자격을 부여한다. SSS 자격을 가진 개인은 인구조사국 직원과 마찬가지로 평생 데이터를 보호하고 동일한 법적 의무와 처벌을 감수할 것을 선언해야 한다.

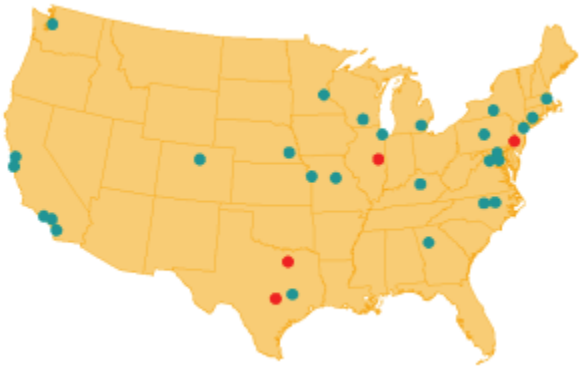
인구조사국 직원과 SSS 자격을 가진 개인들은 매년 개최되는 데이터관리 훈련(Data Stewardship Training)과 Title 26/연방조세정보(Federal Tax Information) 훈련과같이 다른 데이터 보유기관이 요구하는 훈련을 받아야 한다. 또한, 관련 윤리, 기밀성, 프라이버시 보호 절차를 준수해야 한다.

다. 데이터 접근

제안서의 승인과 훈련이 완료되면, 연방조사국은 보안 컴퓨팅 환경 내에서 접근을 제공한다. 대부분의 경우 연방통계연구데이터센터(Federal Statistical Research Data Centers, FSRDCs)에서 하게 되는데, 원격접근이 허용되는 경우도 있다. 연구자들은 승인된 데이터 파일의 읽기전용의(read-only) 비식별화된 버전에 접근하게 된다. 그들은 승인된 연구자들이 공유하는 프로젝트 단위의 작업 공간에서 모든 작업을 수행하게 된다. 마이크로데이터에 대한 모든 분석은 이 컴퓨팅 환경에서 이루어져야 한다.

FSRDC는 인구조사국에 의해 관리되는 시설로서, 연방통계청(federal statistical agencies)과 주요 연구기관들의 협력 관계로 운영된다. 사용이 제한된 마이크로데이터(restricted-use microdata)를 통계 목적으로만 접근을 허가하는 보안시설이다. 현재 25개의 연구데이터센터(RDC)가 있으며, 50여 개의 대학, 비영리연구소, 정부기관과 협력 관계를 맺고 있다.

그림 3-11 미국 FSRDC의 위치



* 출처: 미국 인구조사국 홈페이지 https://www.census.gov/fsrdc

라. 연구결과물 평가 및 공개

연구가 마무리된 후, 그 결과물은 공개되기 전에 어떠한 개인정보나 기업정보가 포함되어 있지 않은지, 그리고 결과물이 애초의 제안서와 일치하는지 검토되어야 한다. 인구조사국 데이터의 경우, 공개 회피 사무관(Disclosure Avoidance Officer) 혹은 전면공개평가위원회(Full Disclosure Review Board, DRB)에서 평가를 수행한다. 인구조사국은 모든 관련 데이터 보유기관에도 검토할 수 있도록 지원한다. 결과물이 평가를 통과하면, 공개된 결과물은 보안 컴퓨팅 환경 외부의 연구팀에 (통상적으로 이메일로) 제공된다. 이제 연구팀은 보고서나 발표물 등을 생산할 수 있게 된다. 인구조사국과 데이터 보유기관과의 계약에 따라서는, 최종결과물이 최초 제안 프로젝트와 일치하는지 여부를 확인하기 위해, 데이터 제공기관이나 인구조사국에 의해 최종 심사를 받아야 할 수도 있다.

(3) 법적 근거 및 정책

인구조사국의 행정데이터 제공에는 인구조사법(Census Law, Titile 13, US code) 및 프라이버시법(the privacy act) 등이 관련된다. 연방법은 인구조사국이 다른 연방기관, 주, 부족, 지방정부, 민간단체에서 앞서 수집한 기록을 사용할 권한을 부여하고, 직접 조사를 수행하는 대신 이러한 정보를 찾도록 하고 있다. 다른 기관에서 수집한 기존 정보들을 사용하는 것이 재정을 절약하고 부담을 줄여줄 수 있기 때문이다. 또한, 인구조사국이 자신이 직접 수집한 정보들과 마찬가지로, 이 정보들의 기밀성을 보호하도록 하고 있다. 또한, 행정데이터의 이용과 개인정보 보호를 위한 다수의 연방

지침들이 존재한다.

인구조사국은 개인정보 보호 및 데이터 연계를 위한 많은 정책을 가지고 있다.

- 프라이버시 원칙 : 인구조사국의 데이터 수집, 관리, 제공에 대한 윤리적 원칙을 수립한다.
- 기록 연계 정책 : 인구조사국의 법적 권한 및 임무에 비추어 어떠한 프로젝트가 적절한지에 대한 기준을 제공한다. 예를 들어, 가장 좋은 대안(best alternative), 공익, 민감성, 개방성, 일관된 평가 및 추적 등이 이에 포함된다.
- 프로젝트를 위한 인구조사국의 가치: 프로젝트가 인구조사국의 프로그램에 어떻게 관련되고 기여하는지에 대한 기준을 수립한다. 예를 들어, 인구 추계, 인구조사/설문조사 프레임, 인구통계 및 경제추계 생산을 위해 사용되는 모델 등이다.
- 연구 제안서 평가 절차 : 연구 프로젝트 제안서 평가를 위한 최소한의 요건을 수립한다. 모든 연구 제안서는 인구조사국의 업무 필요성(인구조사국의 전략 계획 및 업무 범위 관리에 어떻게 일치하는지를 포함하여), 과학적 타당성, 인구조사국의 임무와 명성에 해를 끼치지 않는다는 점 등을 충족해야 한다.
- 인구조사국의 '13-17 전략 계획: 몇몇 전략적 계획은 행정 기록의 이용과 직접적으로 관련되는데, 예를 들어 "기존에 연계되지 않았던 데이터를 결합하여, 국민과 경제에 대한 더 깊은 통찰을 제공하는 새로운 정보제품의 생산을 위해, 행정 기록의 이용을 포함하여 기존 소스로부터 데이터를 통합"하라는 지침이 이에 해당한다. 또 다른 계획은 '빅 데이터'의 연구 목적 이용과 인구조사국의 임무를 지원하기 위한 효율성의 활용을 지원한다.

인구조사국이 데이터셋을 연계할 경우, 그것이 기관의 임무를 지원하는지와 다음과 같은 기준을 고려한다.

첫째, 가장 좋은 대안(Best Alternative) : 비용, 응답자 부담, 시의성, 데이터 질 등을 고려할 때 기록 연계가 가장 좋은 대안인지 판단한다.

둘째, 공익성 (Public Good Determination) : 데이터 연계로 인한 프라이버시 위험과 그로 인해 얻을 수 있는 공익을 비교한다. 또한, 그 위험성을 최소화하기 위한 절차를 이행한다.

셋째, 민감성(Sensitivity) : 특정한 연계가 개인 프라이버시에 미치는 위험 수준에 대한 대중의 인식을 평가하고, 적절한 수준의 검토 및 추적을 해야 한다.

넷째, 데이터 연계 및 행정 기록의 이용은 데이터의 기밀성을 보호해야 한다. 인구

조사국에 의해 획득된 모든 개인 및 기업 정보, 그리고 연계 결과는 USC Title 13에 따라 법으로 보호된다.

2. 네덜란드 SSD 시스템[148]

사회통계데이터셋(social statistical datasets, SSD) 시스템은 상호 연계되고 표준화된 등록소 및 설문조사시스템이다. SSD는 개인, 가정, 직업, 보조금, 연금, 교육, 의료, 범죄, 주거, 교통 등과 관련된 풍부한 정보를 포함하고 있다.

네덜란드에서는 90년대까지 개인, 가정, 기업에 대한 통계는 대부분 설문조사(서베이)를 통해 이루어졌다. 그러나 설문조사에 대한 무응답률이 갈수록 높아지면서 이의 신뢰성에 대한 문제가 제기되었으며, 이에 따라 행정데이터의 활용도가 높아지게 되었는데, 행정데이터는 설문조사의 무응답 편향을 시정하기 위한 용도로 사용되기도 하고, 혹은 설문조사를 아예 대체하기도 했다.

네덜란드 통계청은 1996년에 행정데이터와 설문조사 데이터의 연계에 대한 실험을 진행하였다. 인구 등록소 및 피고용인보험계획의 데이터가 노동력 설문조사 데이터와 연계·처리되었는데, 그 결과는 성공적이었다. 이는 SSD의 개발로 이어졌고, 이후 SSD는 엄청나게 확대되었다. 전통적인 인구조사(센서스)를 수행하는데 3억 유로가 소요된 것에 반해, SSD에 기반해 시행된 2001년 인구조사는 300만 유로밖에 들지 않았다.

SSD는 조직, 과정, 메타 데이터, 소프트웨어 도구, 표준화 및 조정 원칙, 절차 및 프라이버시 보호조치 등과 밀접하게 연결되어 있다. SSD 의 핵심 요소는 다음과 같다.

첫째, 데이터는 표준화된 방식으로 중앙에 저장된다. 둘째, 개인, 건물, 가정, 기업 등 서로 다른 단위 형태가 지정 연계키를 통해 쉽게 연계된다. 셋째, 일관성 있는 결과물을 위해서는 조정이 핵심적이다. 조직적, 기술적, 콘텐츠 관련 조정이 이루어진다.

한편, 네덜란드 통계청은 2003년에서 2006년에 경제통계 데이터베이스를 개발하려고 했으나 성공하지 못했다. 그 이유는 첫째 당시 충분히 많은 행정데이터가 없었다는 점, 둘째 기업 통계에 관한 데이터셋을 미시적인 수준에서 서로 연계하는 것이 어려웠다는 점 때문이다. 그러나 최근 많은 행정데이터가 접근 가능해졌기 때문에 네덜란드는 2016년에 서로 다른 경제 데이터의 통합을 위한 시스템 개발을 목표로 프로젝트를 시작하였다.

148) B.F.M. Bakker et al. (2014)를 주로 참조한 것이다.

(1) SSD 시스템의 데이터 연계

가. 개요

설문조사 기반에서 행정데이터 기반으로의 변화 과정은 서로 다른 데이터 소스의 연계·결합 과정과 동시에 추진될 수밖에 없었다. 왜냐하면, 설문조사와 다르게 행정데이터의 등록소는 주로 한정적인 변수만을 갖고 있었기 때문에, 하나의 소스만을 갖고는 통계를 위한 범위나 깊이가 제한적이었기 때문이다.

SSD의 핵심은 중앙데이터도서관인데, 네덜란드 통계청의 사회경제공간통계부(Division of Socioeconomic and Spatial Statistics)가 유지 관리한다. 이 부서 내의 각 조직 단위(Organizational Unit)에서 고용, 사회보장 등 특정 주제의 통계를 책임지는데, 등록부 데이터(register data)의 수집, 편집, 귀속을 담당한다. 행정 등록부 데이터는 통계 목적으로 수집되지 않기 때문에, 풀질 관리를 위해 광범위한 처리가 필요한데, 등록부 처리는 SSD의 관할은 아니다.

등록부 처리 후에 표준화된 연계키가 부여된다. 이는 서로 다른 통계 등록부를 효과적으로 연계하기 위해 사용된다. 통계 등록부(statistical registers)는 표준화되어 중앙데이터도서관에 저장된다. 메타데이터는 중앙메타데이터보관소에 저장된다. 통계 등록부를 표준화된 형태로 중앙도서관에 저장함으로써 조직 단위들이 정보를 공유하기 용이해진다. 통계 등록부는 서로 연계되어 필요한 범위 및 깊이를 가진 통계 결과물을 생산할 수 있게 된다. 외부의 연구자들이 연구 목적으로 SSD에 저장되어 있는 통계 등록부에 접근하고 연계할 수 있다.

SSD 시스템 개요를 그림으로 나타내면 <그림 3-12>와 같다.

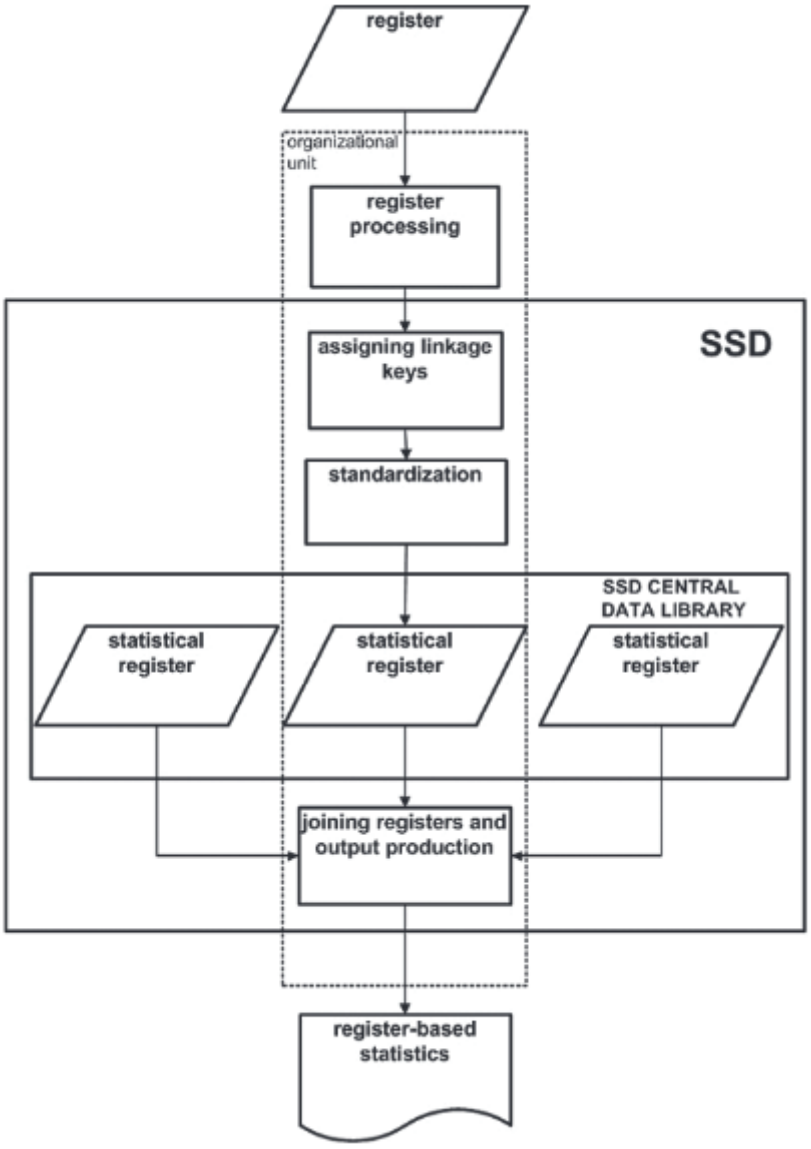
나. 연계키의 부여

중앙 통계의 단위 형태(unit type)는 사람, 가정, 건물, 조직이다. 각 단위 형태에 관한 정보들을 연계하기 위해서, 모든 단위 형태에 연계키가 부여된다.

네덜란드 통계청이 처리하는 대부분의 행정 등록부는 시민서비스번호(Citizen Service number)[149]라는 고유한 개인 식별자를 가지고 있으며, 등록부에 따라 생년월일, 이름, 주소 등 개인 식별자를 포함하고 있다. 프라이버시 보호와 효과적인 연계를 위해 개인식별자는 PIN(개인식별번호)과 AIN(주소식별번호) 연계키로 대체된다. 이들 연계키는 개인을 직접 식별할 수 없는 임의의 키이다.

149) 이전에는 사회보장재정번호(social security and fiscal number, SoFi number)라고 불렀다.

그림 3-12 네덜란드 SSD 시스템 개요



* 출처: B.F.M. Bakker et al. (2014)

네덜란드 인구등록소(population register)는 이러한 키 부여에 핵심적인 역할을 한다. 인구등록소는 등록된 모든 거주자에 대한 식별자와 인구 정보를 보유하고 있는데, 1995년부터 누적된 인구등록소 데이터가 통계청이 개인에 대한 중앙 연계 파일(central linkage file of persons, CLFP)을 유지하는 데 활용된다. CLFP에서 사람들은 PIN과 AIN을 부여받는다. 사람과 관련된 정보를 가지고 있는 모든 등록부는 개인 식별자에 기반하여[150] CLFP에 연계되는데, CLFP에서 PIN과 AIN을 가져오고 생년

150) 시민서비스번호가 있는 경우 이를 이용하여 CLFP에 연계되며, 없는 경우에는 생년월일, 성별, 주소 등 다른 식별자를 통해 연계된다. 시민서비스번호를 통해 연계할 경우 100%에 가깝게 연계될

월일을 제외한 원래의 개인 식별자는 등록부에서 삭제된다.

가정 역시 통계에 있어서 중요한 단위이지만, 대부분 국가에서 행정 등록부에 가정은 존재하지 않는다. 네덜란드 역시 마찬가지이며, 그래서 가정의 경우에는 가족관계, 동거, 경제적 관계 등 다양한 등록부에 기반하여 처리되며 가정식별번호(household identification numbers, HIN)가 부여된다. 전통적인 가정의 경우에는 주소 및 가족관계 정보에 따라, 비혼 커플의 경우에는 경제적 정보에 따라, 기타 가정들은 노동력 설문조사로 수집된 가정 정보에 기반하여 추정된다.

회사나 비영리단체의 경우에는 단체식별번호(organization identification number, OIN)가 부여된다. 대표적으로 직업 등록부의 경우에는 조세번호를 포함하고 있는데, 이 조세번호가 기업 등록부와 연계되어 OIN으로 대체된다. OIN은 산업분류, 기업 규모 등 기업의 특성을 포함하는 사회 통계를 위해 활용된다.

다. 표준화

효율적인 데이터 공유는 높은 수준의 표준화를 요구한다. 네덜란드 통계청은 데이터 생산 절차를 (재)설계하기 위한 공통 구조 프레임워크(architectural framework)를 적용하고 있다. 이 단계에서의 표준화는 주로 통계 등록부의 포맷과 이름, 그에 상응하는 의무적인 메타데이터 파일, 연계키의 데이터 타입 등과 관련된다. 표준화 오류를 체크하기 위해 소프트웨어 도구를 통한 점검이 이루어진다.

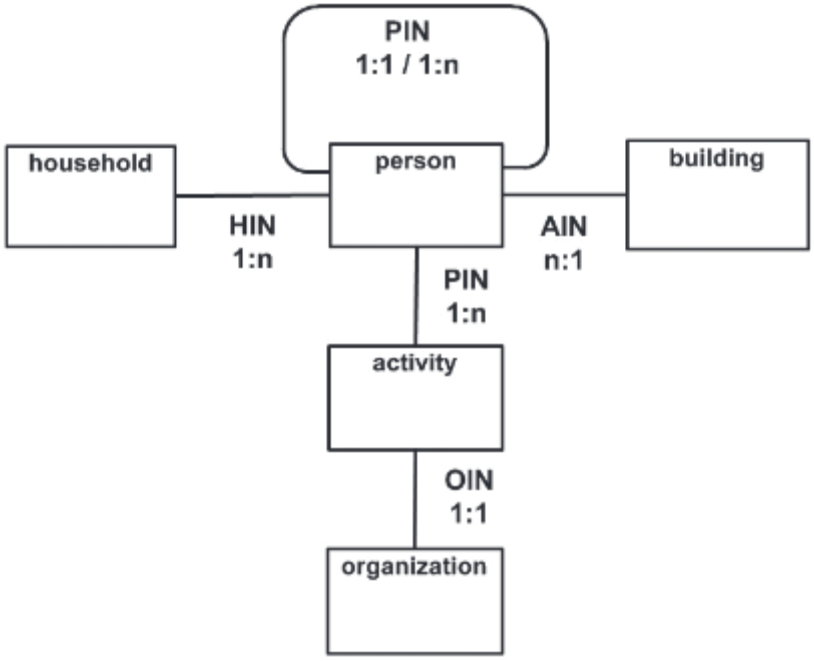
라. 중앙데이터도서관

SSD의 중앙데이터도서관은 통계 결과물을 산출하기 위한 실제 데이터를 보유하고 있다. 그 핵심은 효율적인 데이터 공유와 이를 위한 조정이다.

SSD는 50여 개 이상의 행정 등록부를 보유하고 있다. 중앙데이터도서관의 콘텐츠를 개념적으로 보면, 통계적 단위로서 객체 형태(object type)의 관계(relation)로 볼 수 있는데, 앞서 언급한 사람, 가정, 건물, 조직 등 4개 단위 형태가 있으며 여기에 활동(activity)이라는 객체 형태가 추가된다. 예를 들어 '직업을 갖다' '교육프로그램에 등록하다' 등이 활동이 될 수 있다. '직업을 갖는다'는 활동은 개인 및 기업(조직)과 관계가 형성된다.

수 있다. 연계 처리의 첫 번째 단계는 모든 식별자값이 동일할 경우 연계되며, 연계되지 않은 나머지는 두 번째 단계로 넘어가는데, 여기서는 일정한 차이가 허용된다. 연계키 부여는 결정론적(deterministic) 방법을 사용하지만, 두 번째 단계는 일정한 오차를 허용하므로 확률적(probabilistic) 연계와 유사한 결과가 된다.

그림 3-13 네덜란드 SSD 등록부 시스템의 개념적 모델



* 출처: B.F.M. Bakker et al. (2014)

또한, 데이터의 공유가 효율적으로 이루어지기 위해서는 통계 등록부가 잘 조직이 되어야 하는데, 이를 위한 '조정(coordination)'이 핵심적으로 중요하다. 조정이 제대로 이루어지지 않을 경우, 같은 변수가 서로 다른 이름으로 저장된다거나, 통계 등록부가 서로 다른 파일 포맷으로 저장되는 등의 문제가 발생할 수 있다. SSD는 조직적, 기술적, 콘텐츠 관련, 산출물 관련 조정 등 4가지의 조정 역할을 수행한다.

SSD의 조정 역할은 다음과 같다. 전술했듯이, 통계청의 사회경제공간통계부는 특정한 영역을 담당하는 많은 조직 단위로 구성되는데, 각 조직 단위는 등록부 처리와 그 결과물인 통계 등록부를 SSD의 중앙데이터도서관에 저장하는 일을 수행한다.

두 개의 중앙 조직 단위가 이들의 업무를 지원하는데, 그 하나는 통계 등록부의 연계키를 부여하는 일을 담당하며, 이를 위해 CLFP를 유지하고 매칭 알고리즘을 개발·적용한다. 또 다른 조직 단위는 SSD의 무결성과 콘텐츠의 효율적 활용을 위한 여러 가지 활동을 담당한다. 예를 들어, 서로 다른 통계 등록부의 세부 연계의 수행, 소프트웨어 도구의 개발, SSD 원칙에 대한 교육 제공 등이다.

또한, 두 개의 자문 기구가 있다. 첫째는 모든 조직 단위의 대표자들이 참여하는 자문 기구로서 SSD의 콘텐츠 및 기술 측면의 조정을 목표로 한다. 또 하나는 운영위원회로서 SSD를 감독하고 분쟁을 해결한다.

기술적 조정은 주로 표준화와 관계된다. 파일 포맷, 연계키의 데이터포맷, 명명(naming) 규칙, 메타데이터, IT 기반 및 계획 도구 등 모든 것들이 표준화된다. 메타데이터와 그 구조는 통계 데이터의 올바른 처리 및 이해를 위해 중요하다. 메타데이터는 중앙 메타데이터 저장소에 저장되며, 통계 등록부는 등록부 이름에 기반하여 해당 메타데이터 파일과 일대일 대응되고, 마찬가지로 변수들도 변수명을 기반으로 메타데이터와 관련된다.

콘텐츠 관련 조정은 다음과 같은 절차에 관련된다. 첫째, 새로운 통계 등록부가 만들어지거나, 기존 등록부의 수정이 이루어지면, 그 세부 내용을 모든 조직 단위에 보내서 의견을 받게 된다. 둘째, 각 조직 단위는 표준화된 계획 도구를 이용해 자신만의 시간표(timetable)를 만들게 되는데, 이는 중앙의 생산 계획에 통합된다. 셋째, 자주 업데이트되는 전통적인 등록부의 경우, 모든 통계에 적용될 '조정 버전'이 만들어지게 된다. 예를 들어, 인구조사 등록부의 경우 매월 업데이트되지만, 통계의 일관성 및 재생산가능성(reproducibility)을 위해 매해 '조정 버전'을 지정하게 된다. 넷째, 중앙 메타데이터 저장소에 표준화된 분류 및 그룹화가 저장된다. 다섯째, 상호 일관성 있고, 종적으로 일관성 있는 통계 등록부의 세트를 구축하기 위해 세부 통합(micro-integration)이 수행된다. 이는 측정 오류(measurement error)나 대표성 오류(representation error)를 교정하기 위한 것이다.

산출물 관련 조정은 중복 작업이나 데이터의 부적절한 사용을 방지하기 위한 것이다. 한 조직 단위의 전문가가 자신들의 통계 등록부를 이용한 다른 조직 단위의 결과물을 검토할 수 있어야 한다.

마. 통계 결과물의 생산

SSD의 통계 등록부는 표준화되어 있고 공통의 연계키를 가지고 있기 때문에, 이들 통계 등록부를 결합하여 결과물을 산출하는 것은 명확한 과정이다. 예를 들어, 개인들과 가정, 직업, 혹은 거주지를 연계하고자 하면, 각각 HIN, PIN, AIN을 연계키로 사용한다. 결과물 산출을 위해 두 개의 소프트웨어 도구를 사용할 수 있다. 첫째는 SSD 도서관의 콘텐츠 목록을 산출하는 도구이다. 관할 조직 단위로 조직화된, 메타데이터를 포함한, 사용 가능한 통계 등록부 목록을 보여준다. 두 번째 도구는 필요한 데이터 셋을 생성하는 데 이용된다. 이용자가 필요한 인구 및 변수를 특정하면, 이 도구는 SSD 메타데이터 보관소를 참조하여 결합이 필요한 통계 등록부를 결정한다. 그리고 선택 및 연계 과정을 거쳐 원하는 데이터셋을 보여준다. 데이터셋의 분석과 결과물 생산은 목적에 적합한 소프트웨어를 사용한다.

(2) 연구 목적의 통계 데이터 제공

엄격한 조건 하에, 허가받은 연구자들은 네덜란드 통계청이 보유한 상세한 데이터에 접근할 수 있다. 통계청 구내의 장소에서 접근하거나, 혹은 보안 인터넷 접속을 통해 원격 접근할 수도 있다. 현재 전 세계 300개의 프로젝트를 위해 700명의 연구자가 원격접근을 활용하고 있다. 파일은 통계청 서버에 저장되어 있으며, 연구자들은 보안 인터넷 접속을 통해 원격으로 분석을 수행한다. 네덜란드 통계청은 보다 유연한 원격 접근 서비스를 제공하기 위해 2017년 1월 18일에 시범 프로그램(pilot program)을 시작하였다.[151] 현재 원격접근을 위해서 연구자들은 폐쇄된 공간에 분리된 컴퓨터가 필요하다. 그들은 특정 컴퓨터에 연결된 지문인식기를 통해 프로젝트 환경에 원격으로 접근할 수 있다. 새 시범 프로젝트는 연구자가 문자 메시지로 토큰과 코드를 받아 어떤 작업실에서도 데이터에 접근할 수 있도록 하려 한다.

(3) 법적 근거 및 프라이버시 보호

1970년대에 네덜란드에서 프라이버시 보호에 대한 우려가 커졌는데, 그 논쟁을 촉발시킨 것이 1971년 인구 및 주택 총조사였다. 결국, 1981년 총조사는 연기되었는데 이는 무응답률이 26%에 이를 것으로 예상되었기 때문이다. 결국, 전통적인 방식의 총조사는 중단되었다. 그리고 이 논쟁을 계기로 개인정보보호법 제정 작업이 시작되었다.

네덜란드 통계청의 법적 근거는 네덜란드 통계법(Statistics Netherlands Act)인데, 이 법은 통계청이 가능하면 정부 기관의 행정데이터를 사용하도록 하고, 그럴 수 있는 권한을 부여하고 있다. 또한, 통계청이 시민서비스번호를 사용할 수 있도록 허가하고 있다.

통계법은 개인정보 보호를 위한 규정도 포함하고 있다. 이에 따르면, 첫째, 통계청이 받는 모든 데이터는 통계적 목적으로만 사용되어야 한다. 둘째, 통계청은 데이터 보호를 위한 기술적, 조직적 조치를 취해야 한다. 셋째, 공개된 결과물이 개인정보를 드러내지 않도록 조치를 취해야 한다. 넷째, 통계법에 따른 책임을 수행하는 사람을 제외하고는, 다른 사람에게 데이터를 전달해서는 안 된다. 다만, 이에 대한 예외로서, 통계청은 통계적 혹은 학술적 연구 목적으로 다른 기관에 마이크로데이터를 제공할 수 있다. 통계법은 어떤 기관이 통계적·학술적 연구를 수행하는지를 규정하고 있지만,

151) CBS, More flexible access to CBS microdata for researchers, 2017.4.6. https://www.cbs.nl/en-gb/corporate/2017/04/more-flexible-access-to-cbs-microdata-for-researchers

다른 조직이나 기관도 허가를 신청할 수 있으며, 통계청의 독립적인 감독기관인 중앙통계위원회(Central Commission for Statistics)가 심사한다.

네덜란드 통계법 외에 프라이버시 관련 법률을 준수해야 한다. 개인정보를 포함한 등록부의 유지 및 활용을 규제하는 최초의 법률로서, 1998년에 개인정보등록법(Act on Personal Data Registrations, WPR)이 제정되었고, 2001년에 네덜란드 개인정보보호법(Netherlands Data Protection Act, WBP)으로 대체되었다. 개인정보 감독을 위해 개인정보감독기구(Data Protection Authority, DPA)를 두고 있다.

개인정보보호법은 애초 수집 목적 외의 개인정보 처리를 금지하고 있지만, 역사적, 통계적, 학술적 목적으로 개인정보를 처리할 수 있는 예외를 두고 있다. 통계청이 행정데이터의 처리 및 통계 등록부를 전달할 수 있는 것도 이 예외규정에 따른 것이다. 수집된 개인정보를 필요 이상으로 보유하는 것도 금지되지만, 이 역시 같은 목적으로 예외가 적용된다. 이 예외규정의 적용을 위해서는 프라이버시를 보호하고, 개인정보를 명시된 목적 외로 사용하는 것을 방지하기 위한 적절한 조치를 취해야 한다. 또한, 모든 개인정보 처리는 개인정보보호감독기구 혹은 각 조직 내에서 지정한 개인정보보호책임자(data protection officer)에게 보고되어야 한다. 통계청 내에서도 개인정보보호법의 적용과 준수 여부를 감독하는 책임자가 지정되어 있다.

또한, 데이터 보안을 목적으로 한, 다음과 같은 조치를 취해야 한다.

첫째, 통계 데이터에서 개인식별자를 제거하고 공통이 PIN 및 AIN 연계 키로 대체한다. 둘째, 중앙데이터도서관에의 접근을 통제한다. 업무를 위해 SSD 데이터를 필요로 하는 직원만이 접근 가능하며, SSD 네트워크의 일부, 그리고 필요한 데이터에만 접근이 허용된다. SSD에 접근하고 있는 사람은 동시에 네트워크의 다른 부분에 대한 접근이 동시에 허용되지 않는다. 셋째 SSD 외부로의 데이터 유출을 막기 위해 이메일 기능이 제한된다. 데이터도서관에 접근하는 직원들은 이메일에 파일을 첨부할 수 없다. 넷째 통계청 건물에 출입하는 사람들은 엄격한 검사를 받는다. 모든 직원은 신원 확인을 위한 ID 카드를 지참해야 하며, 방문자는 항상 신분증을 통한 신원 확인을 받고 직원을 동반해야 한다.

3. 캐나다 SDLE

캐나다 통계청(Statistics Canada)[152]은 가능하면 추가적인 설문조사를 수행하는 것보다 개인, 기업, 기관 등이 이미 통계청이나 다른 정부부처에 제공한 정보들을 사용

152) 이하 SDLE에 대한 소개는 캐나다 통계청 홈페이지(http://www.statcan.gc.ca)를 참고한 것이다.

한다. 서로 다른 소스의 데이터를 연계하는 것은 통계 데이터의 설계, 생산, 분석, 평가에 있어서 유용하고 비용 효과적인 도구가 될 수 있고, 비용, 시간, 응답자 부담을 줄여줄 수 있으며, 때로는 중요한 통계정보를 생산하는 유일한 수단이 될 수도 있기 때문이다.

캐나다 통계청의 데이터 연계는 마이크로데이터 연계 지침에 따라 이루어지며, 다음과 같은 두 가지 목적을 가진다. 첫째, 캐나다 통계청 내에서 수행 중인 데이터 수집 및 방법론적 연구의 설계, 유지, 평가, 연구 및 재설계 지원. 둘째, 연구 조사를 지원하기 위한 총계 혹은 익명 형식의 통계정보 제공.

캐나다 통계청은 사전 승인된 특정 형식의 연계를 보유하고 있는데, 이는 프라이버시 위험이나 잠재적인 이해 충돌의 가능성이 작을 경우, 그리고 프라이버시 및 기밀성에 대한 위험을 최소화하기 위한 절차가 적용된 경우이다. 그 외의 마이크로데이터 연계는 상급 관리자에게 제안서를 제출하여 평가 및 승인 절차를 밟아야 한다. 캐나다 통계청은 2000년 1월부터 마이크로데이터 연계 지침에 따라 평가되고 승인된 데이터 연계 목록을 홈페이지를 통해 공개하고 있다.

캐나다 통계청의 사회데이터연계환경(Social Data Linkage Environment, SDLE)[153] 프로그램은 데이터 연계를 위한 환경으로 캐나다 전역에서 사회경제적 통계 연구를 촉진하기 위한 목적으로 만들어졌다. 데이터 연계를 통해 중요한 연구 과제를 해결하고 사회경제 정책에 정보를 제공하기 위해 기존의 행정데이터와 설문조사 데이터의 혁신적 활용을 증진하고자 한다. 또한, 추가적인 데이터의 수집 없이, 연계된 분석 데이터 파일의 생성을 통해, 보건, 사법, 교육, 수입과 같은 여러 영역에 걸친 데이터 통합의 잠재력을 확장하고자 한다.

SDLE 프로그램은 다음과 같은 역할을 한다. 첫째, (완료된 종적 설문조사와의 관련성을 유지하면서) 새로운 데이터의 수집 없이 캐나다 통계청의 기존 설문조사와의 관련성을 증가시킨다. 둘째, 행정데이터의 활용을 크게 증가시킨다. 셋째, 추가적인 데이터 수집 없이 새로운 정보를 생성한다. 넷째, 높은 수준의 프라이버시 및 데이터 보안을 유지한다. 다섯째, 데이터 연계 절차 및 방법에 표준화된 접근을 증진한다.

SDLE의 처리 시스템 및 누적된 연계 결과의 활용을 통해 데이터 연계를 위한 파일의 준비 및 관리를 둘러싼 절차가 더 효율적이고 적시에 이루어지게 된다.

153) http://www.statcan.gc.ca/eng/sdle/index

(1) SDLE를 통한 데이터 연계

SDLE는 사회적 분석을 위해 연계된 인구 데이터 파일을 생성하기 위한 고도로 보안이 갖춰진 환경이다. 그러나 SDLE가 거대한 통합 데이터베이스는 아니다. SDLE의 핵심에는 파생기록보관소(Derived Record Depository, DRD)가 있다. DRD는 단지 기본적인 개인 식별자만을 담고 있는 국가적인 동적 관계 데이터베이스(dynamic relational database)이다.

DRD는 고유한 개인 목록을 생성하기 위한 목적으로 캐나다 통계청이 선정한 소스 색인 파일들(source index files)을 연계하여 생성된다. 이 소스 색인 파일들은 속성정보는 포함하고 있지 않으며, 단지 개인 식별정보만을 가지고 있다. DRD 구축을 위해 T1 개인 마스터 파일(조세), 캐나다 아동 세금 혜택 파일(Canadian Child Tax Benefits, CCTB), 캐나다 인구 동태 통계(출생 및 사망 데이터베이스), 입국 파일 등의 소스 색인 파일이 이용된다.

이 파일들을 SDLE로 가져와서 처리하고, DRD와 단 한 번 연계된다. 소스색인파일들은 G-Link라는 소프트웨어를 사용하여 확률연계를 통해 DRD에 연계되거나, SAS 스크립트를 사용하여 확정연계 방식으로 연계된다. 확정연계는 양 파일이 공유하고 있는 고유식별자에 기반하여 연계가 수행되며, 확률연계는 이름, 성별, 생년월일, 우편번호 등 고유식별자가 아닌 식별자를 통해 연계된다.

DRD의 각 개인은 익명화된 SDLE 식별자(SDLE idenifier)를 부여받게 된다. SDLE 식별자는 임의적으로 부여되며 SDLE 밖에서는 아무런 가치도 갖지 않는다. 소스 색인 파일들의 갱신 내용은 DRD에 지속적으로 연계된다.

DRD에는 단지 개인 식별정보만이 저장된다. DRD에 저장되는 개인정보는 성, 이름, 생년월일, 성별, 사회보장번호(Social Insurance Number), 부모 이름, 결혼 여부, (우편번호를 포함한) 주소, 전화번호, 이민(입국, 출국)날짜, 사망일 등이다. 데이터 연계를 위해 사용된 SDLE 식별자와 소스색인파일의 레코드 ID 조합은 키등록부(Key Registry)에 저장된다.

연계 데이터를 요구하는 연구가 승인되면 특정 집단의 레코드 ID와 이와 연계된 파일들의 관련 레코드 ID가 키등록부에서 추출된다. 추출한 관련 레코드 ID들이 소스 데이터 파일에서 개인의 기록을 찾는 데 사용된다. 이 레코드 ID는 분리된 소스데이터 파일로부터 선택된 속성정보들을 추출하여 연계 분석 파일을 만들기 위해 사용된다. 이런 방식을 통해 거대한 통합 데이터베이스를 구축할 필요 없이 가상의 연계 환경을 제공하게 된다.

소스데이터 파일의 복잡성에 따라 연계분석 파일을 어떻게 구조화할 것인지 결정되

며, 연계데이터의 품질도 평가되어야 한다. SDLE에서 연계되는 데이터들은 다음 두 종류의 유효성 검증을 받는다. 첫째, 데이터 연계 평가 : DRD와의 연계율을 어떻게 되는가? 연계는 유효한가? 둘째, 연계분석 파일의 평가 : 연계된 해당 주제의 관점에서 적절한가? 연계 과정을 통해 어떤 편향이 있었는가? 연구 대상이 되는 인구를 적절히 대표하고 있는가? 이와 같은 파일 구조의 결정과 데이터 품질 측정은 문서화되고 최종 분석에서 고려될 필요가 있다.

<그림 3-14>는 SDLE의 구조를 나타낸다.

SDLE는 필요할 경우 다음과 같은 서비스도 제공하고 있다.

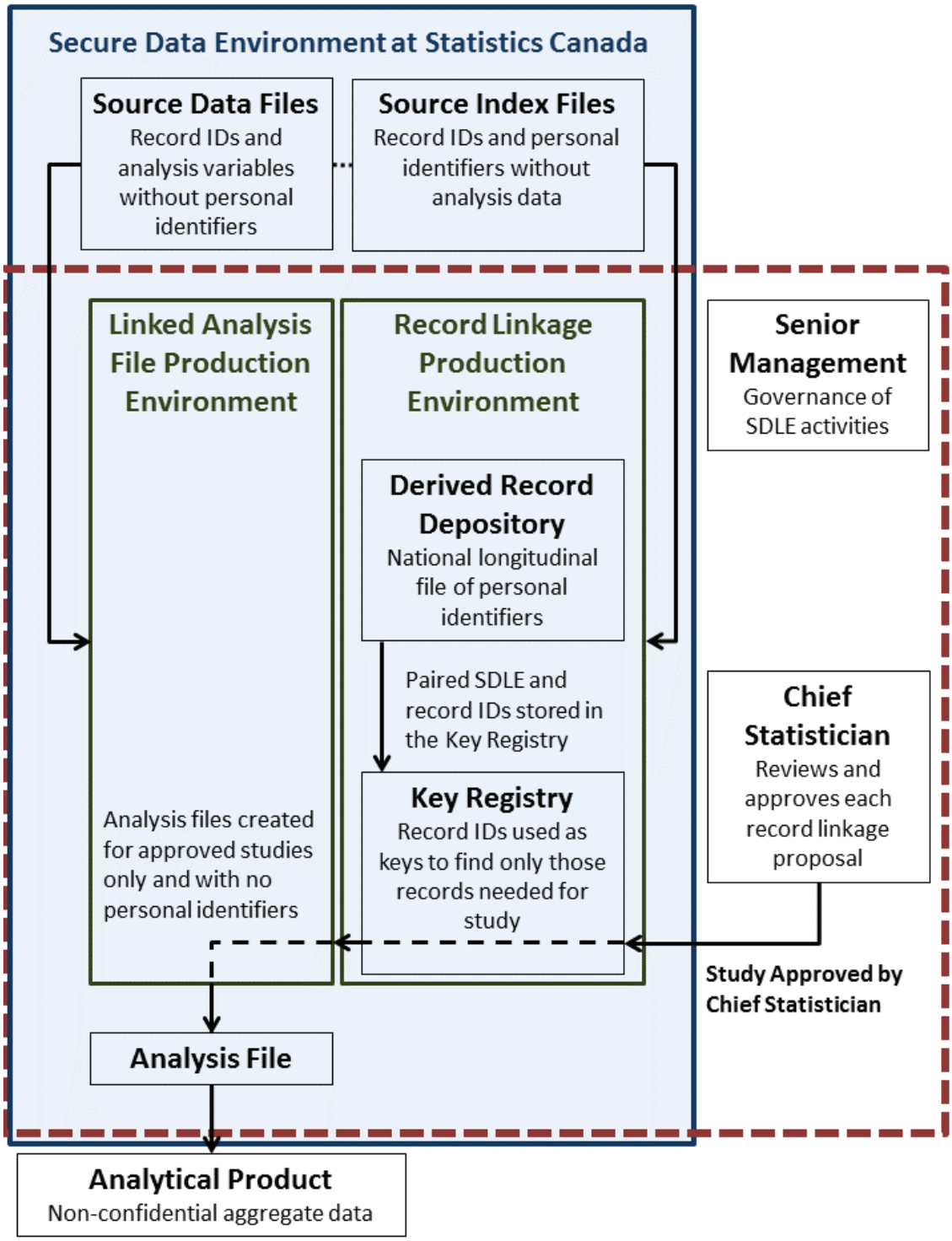
- 프로젝트 실현 가능성에 대한 평가
- 데이터 소스, 분석적 한계, 유효성에 대한 조언
- 해당 주제의 전문가 소개
- 승인 절차 지원
- 통상적 연계분석 파일의 구축
- 훈련 및 홍보 제공

(2) 데이터에의 접근

캐나다 통계청은 연구자가 요청하는 데이터의 성격에 따라 다양한 데이터 접근 방법을 제공하고 있다.

- Data Liberation Initiative (DLI) : 고등 교육기관과 캐나다 통계청의 협약에 의해 진행되고 있는 사업으로, 고등교육기관에 광범한 데이터와 메타데이터를 제공한다. 교수와 학생들은 공개사용이 가능한 마이크로데이터 파일(public use microdata files, PUMFs), 데이터베이스, 지리 파일에 무제한 접근 가능하다.
- PUMFs 모음에 대한 접근 : 통계청의 공개사용이 가능한 마이크로데이터 파일들에 대해 가입 방식으로 기관들의 접근을 허용하고 있다. 이 파일들은 총계 처리되지 않은(non-aggregated) 데이터를 익명화된 방식으로 보유하고 있다.
- 연구데이터센터(Research Data Centres, RDC) 프로그램 : RDC 프로그램은 대학의 안전한 환경 내에서 행정 마이크로데이터 파일과 인구 및 가구 설문조사 데이터에 대한 직접 접근을 제공한다. 이 센터는 통계청 직원이 관리하며, 승인된 프로젝트의 연구자에게만 제공된다. 연구자들은 통계법상 '직원 간주'의 지위를 획득해야 한다. RDC는 캐나다 전국에 위치해 있다.

그림 3-14 캐나다 SDLE 개요 다이어그램



* 출처: 캐나다 통계청 홈페이지 http://www.statcan.gc.ca/eng/sdle/overview

- 연방연구데이터센터(Federal Research Data Centre, FRDC) 프로그램 : RDC와 유사하며, 연방 부처의 직원들이 안전한 공간에서 복잡한 통계 분석, 연구를 할 수 있도록 지원한다.
- 캐나다 데이터 개발 및 경제 연구 센터(Canadian Centre for Data Development and Economic Research, CDER) : CDER는 연구자에게 분석적 연구를 위해 기업 및 경제 마이크로데이터 파일에 대한 직접 접근을 제공한다. CDER은 오타와의 통계청 본사에 위치해있다. 이 데이터에 대한 접근 역시 승인된 프로젝트의 연구자에게 제공되며, 연구자들은 통계법상 '직원 간주'의 지위를 획득해야 한다.
- 실시간 원격접근(Real Time Remote Access, RTRA) 시스템 : RTRA 시스템은 온라인을 통해 이용자들이 안전한 공간에 위치해있는 마이크로데이터 파일에 대해 실시간으로 SAS 프로그램을 실행할 수 있도록 한다. 연구자들은 마이크로데이터에 직접 접근하여 콘텐츠를 볼 수는 없으며, 대신 SAS 프로그램을 통해 결과물만을 추출할 수 있다. 따라서 통계법상 '직원 간주'의 지위를 획득할 필요는 없으며, 연구제안서를 제출할 필요도 없다.
- 보건 설문조사 데이터에 대한 원격접근 : 보건 통계청에 의해 제공되는 서비스로 다변수 분석을 포함한 프로젝트에 적합하다. PUMF 데이터가 연구에 충분하지 않고 RDC나 RTRA 서비스에 대한 접근이 가능하지 않을 경우, 이 서비스를 신청할 수 있다. 프로젝트가 승인되면 연구자는 자신들의 프로그램을 개발하고 실험할 수 있는 모조 데이터를 받게 된다. 그리고 이 프로그램을 보건 통계부에 전송하면, 안전한 데이터 서버에서 이 프로그램이 실행된다. 결과물은 공개 및 기밀성 조건에 맞는지 평가되며, 결과물이 승인되면 연구자에게 보내진다.
- 바이오 은행(Biobank) : 캐나다 보건측정 설문조사(Canadian Health Measures Survey, CHMS)의 일환으로 혈액, 소변, DNA 샘플이 참여자 동의하에 수집된다. 이 샘플은 향후 보건 연구를 위해 CHMS 바이오 은행에 저장된다. 이를 이용한 모든 프로젝트 결과물은 RDC를 통해 공개되어야 한다. 공개된 결과물은 총계화된 형태로만 공개되어야 하며, 개인 식별 정보가 공개되어서는 안 된다.

(3) 연구데이터센터와 연구제안서 신청 절차

1998년 캐나다 사회통계사업(Canadian Initiative on Social Statistics)은 캐나다 내의 연구 공동체들이 당면한 문제들에 대해 연구를 했는데, 그 권고 중의 하나가 연구자들에게 캐나다 통계청의 마이크로데이터 파일에 대한 접근을 증진할 수 있는 연구시설을 구축하라는 것이었다. 이에 따라 구축된 것이 연구데이터센터(RDC)인데, 이는

캐나다 통계청, 사회과학인문연구위원회(Social Sciences and Humanities Research Council, SSHRC) 및 대학들의 협력 사업이다.

RDC는 캐나다 전역에 위치해있는데, 29개의 대학이 캠퍼스 내에 RDC를 유치하고 있고, 연방 부처를 위해 오타와에도 설립되어 있다. 캐나다 통계청 직원이 관리하고 있으며, 통계법 조항의 기밀성 규정에 따라 여느 통계청 사무실과 마찬가지로 운영된다. 물리적인 접근이 제한되고, 컴퓨터들은 통계청 외부와 연결되지 않는다. RDC들과 그 지부들이 RDC 네트워크를 구성하고 있다.

RDC는 연구자에게 안전한 환경에서 인구 및 가정 설문조사, 행정데이터, 연계 데이터를 마이크로데이터 수준에서 제공한다. PUMF가 개인 수준의 데이터가 없는 총계화된 데이터인 반면, RDC는 총계 처리되지 않은, 개인 수준의 데이터에 대한 접근을 제공한다. 그래서 승인된 프로젝트의 연구자만이 접근할 수 있으며, 연구자들은 통계법상 '직원 간주'의 지위를 획득해야 한다.

RDC에 대한 데이터 접근 신청 절차는 연구 책임자의 소속이나 연구의 형태에 따라 달라진다. 학술 연구자인지, 정부 지원 연구자인지에 따라 달라지며, 학술 연구자인 경우에도 교수, 박사과정, 연구 프로그램 등에 따라 달라진다.

예를 들어, 고등교육기관의 교수 혹은 직원이고, 정부 지원을 받지 않는 연구를 수행하는 경우, 다음과 같은 절차에 따라 신청하게 된다.

- Step 1: 연구 제안서 초안작성
- Step 2: SSHRC 웹사이트를 통해 온라인 신청양식 작성
- Step 3: 제안서에 대한 평가
- Step 4: 보안 검사절차 완료
- Step 5: 캐나다 통계청과 마이크로데이터 연구 계약서 서명
- Step 6: 결과물 제출

RDC의 분석가가 제안서가 필요한 요소들을 모두 포함하고 있는지, 해당 프로젝트가 상세한 마이크로데이터 접근에 적절한지 등을 검토하고 조언을 제공한다. 각 제안서는 두 명의 학계 연구자와 통계청의 해당 주제 전문가에 의해 평가된다. 제안서의 승인은 학술적인 가치 및 연구의 실행 가능성, 적용 방법과 분석될 데이터의 관련성, 상세 마이크로데이터에 접근할 필요성의 입증, 연구자의 전문성과 능력 등의 원칙에 따라 이루어진다.

프로젝트가 승인되면 몇 가지 보안 절차를 밟아야 한다. 첫째, 캐나다 통계청은 마

이크로데이터에 접근하는 모든 연구자에 대한 신뢰성 검사(Reliability Check)를 수행한다. 둘째, 연구자들은 RDC 분석가가 있는 자리에서 보안 검사 양식을 작성해야 한다. 셋째, RDC 분석가가 이 양식을 통계청에 보내면 통계청은 이를 검토하고 그 결과를 연구자에게 통보해준다. 넷째, 2016년 12월 1일부터 캐나다 연방경찰(RCMP)는 (RDC 연구자와 직원을 포함한) 모든 연방 공공 공무원에게 보안 검사의 일환으로 지문을 찍게 하고 있다. 추가적으로 캐나다 재무위원회(Treasury Board of Canada)는 신용 검사를 요구하고 있다. 다섯째, RDC 분석가는 연구자들을 RDC에 초청해서 절차를 설명하는 오리엔테이션을 갖는다. 이 세션에서 연구자들은 캐나다 통계청과 계약을 체결하고 기밀성 서약을 서약(Oath or Affirmation of Office and Secrecy)을 하게 된다.

일반적으로 보안 검사 양식을 작성할 때, 연구자들은 사진, 5년 동안의 주소 기록, 캐나다 국적이 아닌 사람은 취업 비자와 캐나다 거주 증명서 등을 제출해야 한다.

제안서 승인 및 보안 절차 후에 연구자들은 통계청과 계약을 하게 되는데, 계약서에 서명하고 서약서를 받으면 '직원 간주' 지위를 얻게 된다. 계약서에는 캐나다 통계청으로부터 제공받을 데이터셋, 제안서에 명시된 연구 프로젝트의 목적과 범위, 프로젝트 시작 및 완료일, RDC의 보안 및 기밀성 요구조건을 준수한다는 연구자 동의서, 최종 결과물을 계약 종료일에 통계청에 제공한다는 동의서 등의 내용이 포함된다.

연구자와 캐나다 통계청의 계약에 따라 연구 결과물을 제출해야 한다. 연구 결과물은 캐나다 통계청의 최고연구책임자가 주관하는 RDC 워킹 페이퍼, 동료 평가된 저널 논문, 책(혹은 일부 챕터), 학위 논문, 연구보고서 등의 형태가 될 수 있으며, 다른 형식이 필요하다면 RDC 분석가와 협의한다.

(4) 개인정보 보호

캐나다 통계법에 따라 캐나다 통계청에 제공되는 모든 정보는 기밀성이 유지되며, 단지 통계 목적으로만 이용된다. 개인정보 보호를 위해 캐나다 통계청은 개인정보 감독기구(Office of the Privacy Commissioner)의 자문을 받을 뿐만 아니라, 프라이버시 영향평가를 수행한다.

연계분석 파일은 민감한 통계정보로 볼 수 있고 통계법의 기밀성 요구조건을 따라야 한다. 프라이버시 침해 위험을 줄이기 위해, SDLE의 소스 파일은 소스 색인 파일과 소스데이터 파일로 분리된다. 또한, 소스 색인 파일을 사용하는 데이터 연계 환경은 소스데이터 파일을 사용하는 데이터 통합 및 분석 환경과 분리되어 있다. 즉, 데이터 연계를 수행하는 캐나다 통계청의 직원들은 단지 연계에 필요한 기본적인 개인 식

별자에만 접근할 수 있으며, 분석 파일을 생성하는 직원들은 단지 개인 식별자가 제거된 데이터에만 접근할 수 있다. 서로 다른 소스의 데이터를 연계분석데이터 파일로 만들기 위해 익명키(anonymous keys)가 사용된다. 또한, 분석 작업을 위한 데이터 접근이 필요한 승인된 직원만이 연계분석 파일에 접근이 허용된다.

4. 뉴질랜드 IDI[154)]

뉴질랜드 통계청은 2개 이상의 서로 다른 소스의 데이터셋 연계를 데이터 통합(data integration)이라고 부른다. 뉴질랜드 통계청도 다양한 부문 간 관계에 대한 연구, 비용과 시간의 절약 등을 위해 데이터 연계를 적극적으로 활용하고 있다. 홈페이지를 통해 데이터통합매뉴얼[155)] 등 데이터 연계 및 통합과 관련된 법적, 기술적 정보를 제공하고 있다. 또한, 뉴질랜드 통계청은 자신이 보유한 데이터를 통한 연구를 활성화하기 위해 통합데이터기반(The Integrated Data Infrastructure, IDI)[156)]을 운영하고 있는데, 이는 사람 및 가정에 대한 마이크로데이터를 포함하고 있는 연구 데이터베이스이다.

1997년 내각 지침(Cabinet directive)에서 뉴질랜드 통계청이 기관 간 데이터 통합을 위한 적절한 조직임을 언급한 이래, 뉴질랜드 통계청은 10여 년 이상 데이터 통합 업무를 수행해왔다. 초기의 데이터 통합 프로젝트는 특정한 목적을 위해 수행되었고, 프로젝트마다 각각 분리된 환경에 보관되었기 때문에, 복잡한 연계 작업이나 연구 투자를 되풀이하기 쉽지 않았다.

2011년 내각은 노동부의 이주 데이터를 통계청이 관리하는 통합 데이터셋에 통합하자는 제안에 동의했고, 그 결과 통계청은 기존의 분리된 통합 프로젝트를 이주 데이터와 통합하였는데, 이것이 IDI의 원형이 된다. 이 시점부터 통계청은 일회적인 데이터 통합에서 데이터 통합 서비스 제공으로 나아가기 시작했다.

2013년 내각은 기존 데이터셋의 공유를 통한 정부 역량의 강화를 통해 더 나은 서비스 제공이 가능하다는 것에 합의했고, 부처 간 데이터 공유를 위한 해결책으로 IDI를 확대하기로 하였다.

---

154) 이하 내용은 뉴질랜드 통계청 홈페이지(http://www.stats.govt.nz/)를 참고한 것이다.

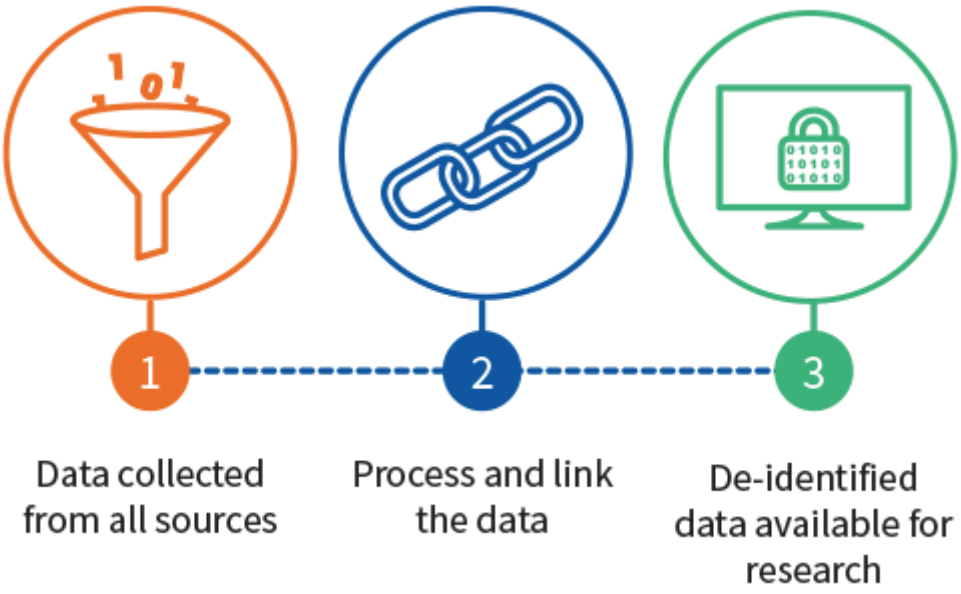
155) 데이터통합매뉴얼(Data integration manual: 2nd edition)은 뉴질랜드 통계청의 데이터 통합에 대한 가이드이다. 이 가이드는 지난 경험으로부터 모범 사례와 통찰을 공유하기 위한 목적으로 만들어졌다. 데이터 통합의 기본 개념, 이론, 절차 및 실질적인 조언을 담고 있다.
http://www.stats.govt.nz/methods/data-integration/data-integration-manual-2edn.aspx

156) 뉴질랜드 통계청 홈페이지 내에서 IDI에 대한 설명 페이지는 다음과 같다.
http://www.stats.govt.nz/browse_for_stats/snapshots-of-nz/integrated-data-infrastructure.aspx

(1) IDI의 작동방식

아래 그림은 IDI가 어떻게 작동하는지에 대한 개략적인 흐름을 보여준다. 1단계로서 모든 소스로부터 데이터가 수집되고, 2단계로 수집된 데이터를 처리·연계하며, 3단계로 비식별화된 데이터를 연구 목적으로 제공한다.

그림 3-15 뉴질랜드 IDI의 작동방식



* 출처: 뉴질랜드 통계청
http://www.stats.govt.nz/browse_for_stats/snapshots-of-nz/integrated-data-infrastructure/idi-how-it-works.aspx

가. 데이터 수집

다양한 정부부처, 통계청의 인구조사, 비정부기구의 데이터들이 IDI로 안전하게 제공된다. 이름과 같은 개인 식별정보를 포함하고 있는 데이터에 대한 접근은 데이터 처리를 담당하는 필수 IDI 직원으로 접근이 제한된다. IDI의 대부분의 데이터는 분기마다 업데이트가 이루어진다.

나. 연계 사전작업(Pre-linking)

원 데이터 보유기관으로부터 데이터를 받으면, 우선 해당 데이터가 데이터 제공 계약에 부합하는지 검토하고 연계를 위한 준비를 한다. 연계 변수가 표준 포맷으로 제시되어 있는지 확인하고, 유효하지 않은 문자를 제거하기 위한 텍스트 처리를 적용한다. 지오코딩(geocoding)을 위해 주소 정보도 표준화한다. 지오코딩은 메시블록

(meshblocks)이라 부르는 작은 그래픽 단위를 통해 연구자가 위치 정보를 활용할 수 있도록 하는데, 특정한 주소는 익명화된다.

다. 데이터 연계

각 데이터셋에서 활용 가능한 변수에 따라 서로 다른 데이터 연계 방법이 적용된다. IDI에서는 확정연계 및 확률연계 방법이 모두 사용된다. 국가건강색인(National Health Index, NHI) 번호와 같은 공통된 고유 개인 식별자가 있을 경우 확정연계 방법이 이용된다. 그렇지 않을 경우, 이름, 생년월일, 성별 등 인구 정보 변수를 이용한 확률연계 방법이 이용된다. 오류를 줄이기 위해, 업데이트마다 연계 품질 검사가 수행된다. 연계를 책임지는 직원은 변수의 일부(subset)에만 접근할 수 있으며, 고유식별자는 연계 전에 암호화된다.

라. 연구를 위한 비식별 데이터 제공

이름, 주소 등 개인 식별정보는 삭제되며, IRD 번호(조세번호), NHI 번호 등 고유식별자는 암호화(다른 번호로 대체)된다. 연구를 위해 다음과 같은 세 개의 SQL 데이터베이스가 제공된다.

첫째, 온전한 데이터베이스(clean database) : 통상 데이터 제공기관의 이름순으로 정렬된 모든 데이터 테이블을 포함하고 있다. 또한, 서로 다른 소스의 정보를 연계한 데이터도 제공된다. 연구자들은 이 모든 데이터셋에 접근할 수는 없으며, 연구에 필요한 데이터셋에만 접근이 허용된다.

둘째, 메타데이터베이스 : 특정 데이터 모음에 대한 검색 코드와 분류를 포함하고 있다.

셋째, 작업장 데이터베이스 (sandpit database) : 연구자들이 프로젝트팀원들과 테이블, 데이터셋, 프로그래밍 코드 등을 공유할 수 있는 공간이다.

공익 목적의 연구 프로젝트에만 IDI에 대한 접근이 허용된다. 연구자들은 자신의 연구 프로젝트에 필요한 데이터에만, 그리고 보안 데이터랩(secure Data Lab)을 통해서만 데이터에 접근할 수 있다.

마. 데이터랩

연구자들은 보안 데이터랩 환경에서만 마이크로데이터에 접근할 수 있다. 이 공간에서 연구자들은 프로그램 실행, 데이터셋의 생성, 자료 저장, 다른 동료 연구자와의

공유, 추가적인 메타데이터에 대한 접근 등을 할 수 있다. 그러나 데이터랩의 컴퓨터는 인터넷 및 프린터 연결이 되지 않는다.

데이터랩은 뉴질랜드 통계청에 위치해있지만, 연구자들은 자신의 작업공간에서 데이터랩 서버로의 보안 연결을 신청할 수 있다. 승인된다면, 통계청 사무실에서와 마찬가지의 소프트웨어 및 데이터가 제공된다. 이러한 원격 시설은 6개월 이상 소요되는 프로젝트, 그리고 과거에 기밀성과 보안 관련하여 좋은 기록을 갖고 있는 경우에 고려될 수 있다. 실현 가능성, 보안, 모니터링과 관련한 엄격한 조건을 충족해야 하는데, 예를 들어 데이터랩은 공공공간이나 사적 주거공간에는 설치될 수 없다.

뉴질랜드 통계청은 2017년 8월 16일, 시드니의 언스트앤영(Ernst&Young) 부지에 첫 국제 데이터랩을 승인했다. 언스트앤영의 분석가는 사회개발부 및 취약아동부와 함께 수행하는 프로젝트를 위해 뉴질랜드 통계청의 IDI를 사용하고 있다.

(2) 데이터 연계의 원칙과 절차

통계청은 두 개의 마이크로데이터 통합 연구 데이터베이스를 보유하고 있다. 하나는 개인 및 가정에 대한 마이크로데이터를 보유하고 있는 IDI, 다른 하나는 종적 기업 데이터베이스(Longitudinal Business Database, LBD)이다. 서로 다른 데이터셋이 연계 및 통합될 경우, 새로운 통계·연구적 가치를 창출할 수 있는 반면, 개인 식별의 위험성이 증가할 수 있고 이에 대한 대중의 우려가 커질 수 있다. 따라서 뉴질랜드 통계청은 데이터 통합 가이드라인(Data Integration Guideline)[157]의 원칙과 절차에 따라 데이터 연계를 수행하고 있다.

데이터 통합은 다음의 원칙을 만족할 경우에만 이루어진다.

첫째, 통합으로 인한 공익이 데이터 이용에 대한 프라이버시 우려와 통계 시스템의 무결성, 원(original) 소스데이터, 다른 정부의 활동에 대한 위험보다 커야 한다.

둘째, 통합 데이터는 통계 혹은 연구 목적으로만 사용되어야 한다.

셋째, 데이터 통합은 개방적이고 투명한 방식으로 수행되어야 한다.

넷째, 응답자에게 통합하지 않을 것이라고 명확하게 약속한 경우에는 통합하지 않는다.

이 원칙의 준수를 보장하기 위해 아래와 같은 절차에 따라 데이터 통합을 하게 된

157) Data Integration Guideline,
http://archive.stats.govt.nz/about_us/legisln-policies-protocols/data-integration-gdlns.aspx

다. 이는 총계적 통계 생산, 다른 통계 결과물 생산, 마이크로데이터 연구 목적의 데이터 통합에 모두 적용된다.

첫째, 데이터 통합의 잠재적 이익을 확인한다. 일반적인 언급이 아니라, 해당 데이터 통합과 관련된 이익과 위험이 명확하게 확인되고, 설명되고, 소통되어야 한다.

둘째, 위험 평가 및 위험 관리에 필요한 모든 관련 법령 및 규제 의무의 준수를 준비한다. 서로 다른 조직의 데이터, 그리고 행정·설문·상업적 데이터를 통합할 경우, 그 크기와 데이터의 성격, 그것이 애초에 다른 목적으로 수집되었다는 사실 때문에 설문 데이터의 결합보다 더 위험할 수 있다. 합리적으로 예측 가능한 모든 위험과 그것을 방지할 수 있는 방법을 확인해야 하며, 이에 대한 분석을 데이터 통합 제안서에 포함해야 한다. 프라이버시 및 기밀성 영향 분석 템플릿과 영향평가 가이드를 참고한다.

셋째, 통합 데이터는 통계 및 연구 목적으로만 사용되어야 한다. 따라서 이에 필요한 데이터만 통합되어야 한다. 접근을 위한 기준과 해당 데이터 통합에 고유한 특별한 고려사항들을 문서화한다. 기밀 정보의 공유 혹은 통합 방식에 반대하는 사람 혹은 그룹이 있는지, 뉴질랜드 통계청과 원 데이터 보유기관의 명성에 어떠한 영향을 미칠 수 있는지 고려한다.

넷째, 일반적인 발행 관행과 마이크로데이터 접근 관행을 통합된 데이터에 적용한다. 통합 데이터 역시 그러한 관행의 예외가 아니다.

다섯째, 데이터 통합은 개방적이고 투명한 방식으로 수행되어야 한다. 데이터 통합 행위와 결정에 대해 투명해야 하고, 사람들에게 데이터 통합의 목적과 방법을 적극적으로 알려야 한다. 최소한 각 데이터 통합에 대한 정보를 위험 평가 정보와 함께 웹사이트에 게시한다. 또한, 사람들에게 자기 정보에 대한 접근권을 보장한다.

여섯째, 원 정보가 어떻게 수집되었는지, 사람들에게 그 정보가 어떻게 사용될지를 얘기했는지 파악한다. 다른 기관으로부터 정보를 수집했을 경우, 그 정보가 통계청과 공유될 것이며 단지 분석·통계·연구 목적으로만 사용될 것임을 사람들에게 알리도록 해당 기관에 요청한다. 종종 수집 시 예상하지 못한 목적으로 데이터를 사용하게 되는데, 사람들이 제안된 사용을 어떻게 볼지 세심하게 고려한다면, 이것 때문에 데이터 통합을 하지 못하게 되는 것은 아니다.

일곱째, 통계청은 새로운 데이터셋의 IDI 혹은 LBD에 통합하는 모든 제안에 우선순위를 두며, 이 제안들은 업무사례(business case), 프라이버시 및 기밀성 영향평가를 따라야 한다. 이 평가에 대한 검토는 '전략, 성과 및 프라이버시 고위 자문관(the senior advisor, strategy, performance and privacy)'이 수행하는데, 이 자문관은 프라이버시 감독기구와 협의할 수 있다. 검토 후 정부통계관 혹은 그 위임을 받은 사람이

승인한다. 마이크로데이터에 대한 접근은 '마이크로데이터 접근 가이드라인'을 적용한다.

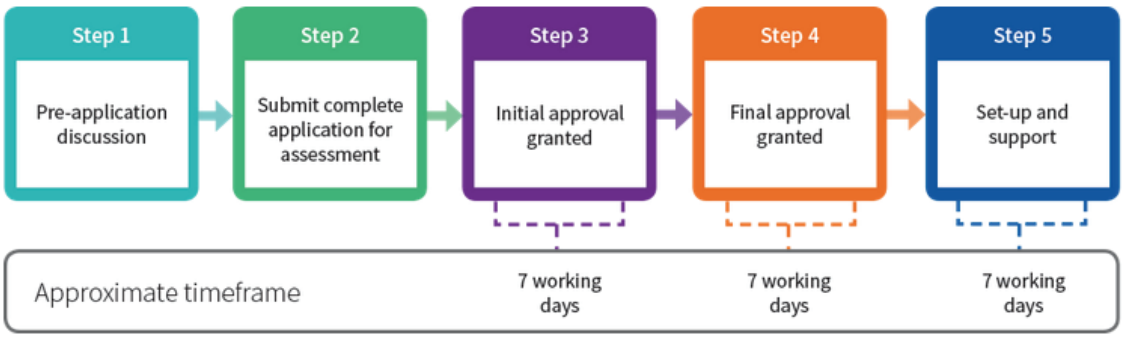
여덟째, IDI나 LBD의 외부에서 데이터셋 통합에 대한 모든 제안은 프라이버시 및 기밀성 영향평가(요약 혹은 전체)를 포함해야 한다. 이 평가는 '전략, 성과 및 프라이버시 고위 자문관'의 검토를 받고, 정부 통계관(혹은 그 위임자)의 승인을 받는다. 통합과정에서 마이크로데이터에 대한 접근이 필요한 경우 해당 가이드라인을 적용한다.

아홉째, 해당 데이터 통합의 이익과 위험에 대한 협의가 있어야 하며, 그 결과 및 결론을 문서화해야 한다. 이 협의는 일반 대중(웹사이트에의 정보 제공을 통해서 할 수 있다.), 외부 데이터 보유기관 및 해당 기관의 관리자, (통계청이 수집하는 경우) 적절한 주제 영역 및 데이터 관리자, 해당 주제 영역에 의해 영향을 받거나 관련된 사람, 적절한 부(deputy) 정부통계관 및 고위 관리자, 전략 성과 및 프라이버시 고위 자문관, 개인정보 감독기구, 통계청 응답자 옹호관(respondent advocate)[158], 통계 기법 전문가 등의 이해당사자들을 포함한다.

(3) IDI 데이터에 대한 접근 신청 절차

공익을 위한 연구를 목적으로 마이크로데이터에 접근하고자 하는 뉴질랜드의 연구자는 데이터랩 서비스를 통해 IDI의 데이터, 종적 기업 데이터베이스(LBD), 기타 통계청의 인구조사 데이터에 접근할 수 있다. 새로운 프로젝트를 수행하거나 기존 프로젝트의 변경을 원하는 연구자는 아래 그림과 같이 신청 절차를 밟아야 한다.

그림 3-16 뉴질랜드 IDI 데이터랩 신청절차



* 출처: 뉴질랜드 통계청 http://www.stats.govt.nz/tools_and_services/microdata-access/data-lab.aspx

158) 응답자 옹호관은 응답자를 대신해서 그들의 불만을 처리하는 역할을 하며, 모든 설문 참여자에 무료로 제공된다. 이 역할은 통계청 내의 수집 활동과 분리되어 있다. http://www.stats.govt.nz/survey-participants/contact-us-about-our-surveys/respondent-advocate.aspx

가. 1단계 : 신청 전 협의(Pre-application discussion)

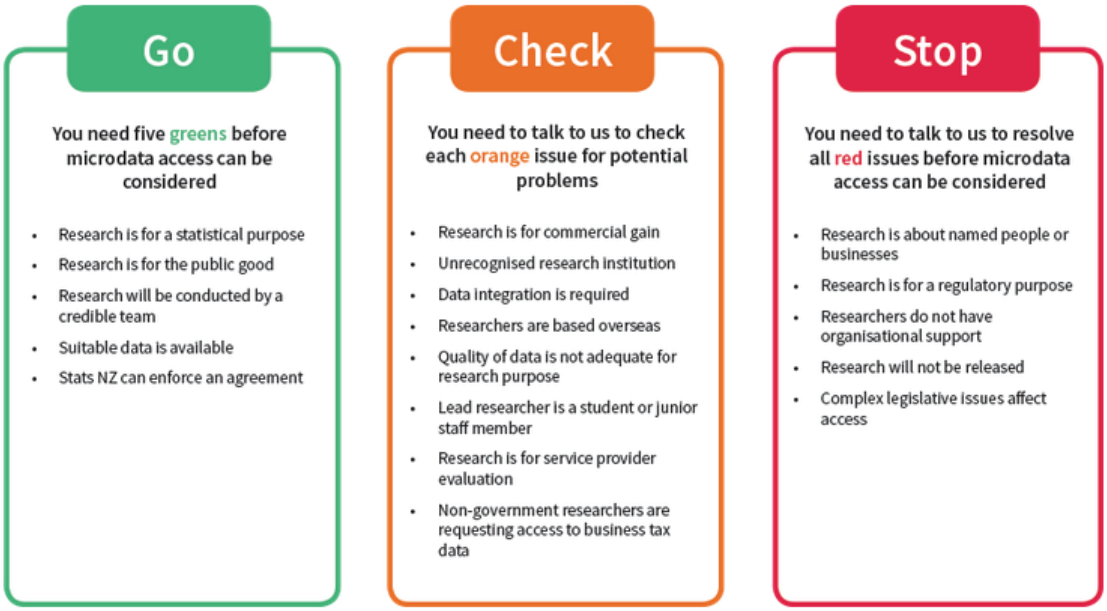
신청서 제출 전에 제안할 프로젝트에 대해 데이터랩 직원과 협의할 것이 권고된다. 데이터랩은 연구의 실현 가능성, 제안된 프로젝트에 어떠한 데이터가 적절한지 여부, 자문을 받을 수 있는 전문가, 신청서에서 해결해야 할 문제점 등에 관한 조언을 제공한다. 데이터랩 신청 전에 다음과 같은 문제들을 고려할 필요가 있다.

- 마이크로데이터 접근 전에 고려해야 할 점
  - 연구는 통계 목적이어야 함
  - 연구는 공익 목적이어야 함
  - 연구가 신뢰할 수 있는 연구팀에 의해 수행되어야 함
  - 적절한 데이터에 접근할 수 있어야 함
  - 뉴질랜드 통계청은 계약을 집행할 수 있음.
- 잠재적인 문제점에 대한 검토
  - 연구가 상업적 이익을 목적으로 하는가
  - 공인되지 않은(unrecognized) 연구 기관인가
  - 데이터 통합이 필요한가
  - 연구자가 해외에 살고 있는가
  - 데이터의 질이 연구 목적에 적절한가
  - 연구 책임자가 학생 혹은 하급 직원인가
  - 연구가 서비스제공자 평가를 위한 것인가
  - 비정부연구자가 기업 조세 데이터에 대한 접근을 요청하는가
- 마이크로데이터에 대한 접근을 위해 해결해야만 하는 문제
  - 연구가 유명한 사람 혹은 기업에 대한 것인가
  - 연구가 규제 목적인가
  - 연구가 조직적 지원을 받지 못하고 있는가
  - 연구가 공개되지 않을 예정인가
  - 복잡한 법적 이슈가 접근에 영향을 주고 있는가

나. 2단계 : 신청서 제출

새로운 프로젝트의 신청인지, 기존 프로젝트의 변경인지에 따라 신청서가 다르다. 새로운 프로젝트의 신청인 경우, 신청서는 신청인 및 신청자의 조직, 제안하고자 하는 프로젝트, 어떠한 소프트웨어가 필요한지, 요청하는 데이터셋 등의 내용을 포함하게 된다.

그림 3-17 데이터랩 신청을 준비할 때 인식해야 할 점



* 출처: 뉴질랜드 통계청 http://www.stats.govt.nz/tools_and_services/microdata-access/data-lab.aspx

다. 3~4단계 : 신청 승인

신청서는 '마이크로데이터 접근 규약(microdata access protocols)'에 근거해서 평가된다. 이 규약은 다음과 같은 6가지 원칙에 근거하여 연구 제안서를 평가한다.

첫째, 통계 목적, 혹은 순수한 연구 목적이어야 한다.

둘째, 통계법 및 다른 관련 법제를 준수해야 한다.

셋째, 마이크로데이터에 대한 접근은 정부 통계관(Government Statistician)이 재량권을 가지고 있다.

넷째, 응답자의 기밀을 보호해야 한다.

다섯째, 뉴질랜드 통계청의 응답자와의 관계에 부정적 영향을 미쳐서는 안 된다.

여섯째, 요청에 대한 결정은 투명한 절차를 통해서 이루어져야 한다.

뉴질랜드 통계청이 보유한 데이터셋 중 일부는 통계법 외의 법률에 의해 수집된 것이 있다. 조세 데이터에 대한 접근 등은 관할 기관의 최고 책임자의 동의를 얻어야 한다. 애초 수집목적 외의 정보 이용에 대해서는 개인정보보호법을 고려해야 한다. 통계법 1975에 상응하거나 제한하는 명백한 법 조항이 있는 경우에는 프라이버시 감독관과 협의해야 한다. 또한, 뉴질랜드 통계청은 마이크로데이터에 대한 접근이 일정한 조건 하에서 이루어질 수 있도록 '5가지 안전조치' 체제('Five Safes Framework)를

갖추고 있다. (이에 대해서는 아래에서 서술하도록 한다.)

데이터랩의 주제, 법률, 방법론팀이 신청서를 1차 승인한다. 일부 데이터 제공기관은 자신들의 데이터를 사용하는 신청서를 검토하고자 요청하기도 한다. 이들은 데이터 사용에 대한 정보나 제공을 제공할 수 있으며, 프로젝트에 영향을 미칠 수 있는 경고를 하기도 한다. 이 기간 동안 지정된 심판관이 연구자에게 연락을 담당한다. 신청에 대한 최종 승인은 정부통계관 혹은 그의 위임을 받은 사람이 담당한다.

라. 5단계 : 준비 및 지원

프로젝트가 승인되면, 데이터랩은 다음과 같이 연구자를 위한 준비 작업을 한다.

① 계약서 및 비밀서약을 포함한 양식에의 서명

② 새로운 연구자를 위한 기밀성 훈련

③ 이용자 프로필 생성 및 프로젝트 공간 준비

④ (필요할 경우) 로그인을 위한 세부사항 및 접근 카드 준비

⑤ 연구에 도움이 되는 자료 제공

마. 결과물 검토

연구자가 기밀성 규칙을 제대로 수행했는지 보장하기 위해, 데이터랩에서 가져가기를 원하는 모든 연구 결과물은 검사된다. 마이크로데이터에 대한 접근은 통계법 1975의 적용을 받으며, 이에 따라 뉴질랜드 통계청은 기밀이 보장되는 개인 및 기업 정보를 보호할 법적 의무가 있다. 뉴질랜드 통계청은 '마이크로데이터 결과물 가이드(Microdata output guide)'를 통해 상세한 지침을 제공하고 있다.

(4) 개인정보 보호 및 보안 조치

뉴질랜드 통계청은 데이터 연계 과정에서의 개인정보 보호를 위해 다음과 같은 원칙을 가지고 있다. 첫째, 연계 데이터는 공식적인 통계 및 승인된 통계 연구의 생산을 위해서만 사용되어야 한다. 둘째, 프로젝트는 원 데이터 소스의 보유기관의 허가를 받아야 하며, 원 소스데이터의 무결성이 유지되어야 한다. 셋째, 뉴질랜드 통계청 직원에 의해 수행되는 데이터 연계는 통계법 1975, 개인정보보호법(the Privacy Act) 1993, 그리고 다른 관련 법제 및 통계청의 데이터통합정책에 따라 수행된다. 넷째, 프로젝트는 통계청이 사용하는 방법에 대한 대중적 신뢰를 위험에 빠뜨려서는 안 된다.

다섯째, 프라이버시영향평가를 통해 프라이버시 위협이 평가된다.

또한, 뉴질랜드 통계청은 마이크로데이터에 대한 접근이 다음과 같은 조건 하에서 이루어질 수 있도록 '5가지 안전조치' 체제('Five Safes Framework)를 갖추고 있다.

그림 3-18 IDI의 5가지 안전조치 체제

* 출처: 뉴질랜드 통계청
http://www.stats.govt.nz/browse_for_stats/snapshots-of-nz/integrated-data-infrastructure/keep-data-safe.aspx

- Safe people : 연구자들은 데이터를 적절하게 활용하고 절차를 따를 것이라는 신뢰가 있어야 한다. 연구자들은 데이터 접근 전에 신원 검사를 받는다. 통계법 1975에 따른 비밀서약서에 서명해야 하고, 통계청의 규칙 및 규약을 준수해야 한다. 비밀서약서는 평생 적용된다. 이를 위반할 경우 접근이 금지되고 블랙리스트에 등재되며 기소될 수 있다.
- Safe projects : 프로젝트는 통계 및 공익 목적이어야 한다. 연구는 개인이 아니라 그룹의 분석으로 제한된다. 정부통계관이나 그 위임을 받은 사람이 모든 연구 제안서를 승인해야 한다.
- Safe settings : 데이터에 대한 무단 접근을 방지하기 위한 안전한 환경에서 연구가 이루어져야 한다. 데이터는 보안 데이터랩 환경에서 접근 가능하며 컴퓨터는 네트워크에 연결되지 않고 단지 통계청 직원만이 연구자에게 데이터를 제공할 수 있다.
- Safe data : 데이터는 근본적으로 노출의 위험을 제한하기 위해 비식별화된다. 연구자는 단지 연구와 관련된 데이터에만 접근할 수 있다.
- Safe output : 생산된 통계 결과는 식별 가능한 정보를 포함해서는 안 된다. 연구자는 데이터랩에서 결과물을 가지고 나오기 전에 기밀성 처리를 해야 하며, 통계청 직원이 이를 다시 검토한다.

뉴질랜드 통계청은 IDI에 대한 프라이버시영향평가를 시행하고 있다. IDI 전반에 대한 프라이버시영향평가(2017년 7월 공개됨)와 함께 새로운 데이터가 추가될 때마다 프라이버시영향평가를 실시하고 이를 공개하고 있다.[159] 또한, 시민들은 IDI가 보유한 자신의 개인정보 제공을 요청할 수 있다. 개인정보보호법 1993이 이러한 권리를 보장하고 있다. 뉴질랜드 통계청은 최고프라이버시책임자(Chief Privacy Officer)를 두고 있으며, 고위프라이버시자문관(Senior Advisor, Privacy)의 지원을 받는다. 최고프라이버시책임자는 정보프라이버시, 보안, 기밀성 거버넌스 그룹의 의장을 맡는다. 이 그룹은 모니터링, 평가, 보고 등 통계청의 개인정보보호법 준수 여부를 감독한다. 고위프라이버시자문관은 개인정보 감독기구(the Office of the Privacy Commissioner)와 긴밀히 협력한다.

159) 다음 페이지에서 프라이버시영향평가 결과를 볼 수 있다.
http://www.stats.govt.nz/browse_for_stats/snapshots-of-nz/integrated-data-infrastructure/keep-data-safe/privacy-impact-assessments.aspx

제5절 시사점

지금까지의 해외 사례 분석을 통해 데이터 연계를 위한 모범 관행(Best Practice)을 도출할 필요가 있다. 이는 데이터 연계라는 특정 절차만이 아니라, 데이터 연계를 포함한 데이터의 수집, 저장, 연계, 제공을 아우르는 원칙과 이를 구현하기 위한 조직적, 기술적인 체계, 즉 데이터 거버넌스 체제를 포함한다. 통계 및 공익 연구 목적의 데이터 이용을 활성화하면서도, 정보주체의 개인정보를 보호하기 위해서는 데이터 거버넌스 전 과정에서 데이터의 활용 및 개인정보 보호조치가 고려되어야 하기 때문이다. 데이터 연계절차는 그러한 거버넌스의 일부분일 뿐이다. 데이터 연계를 포함한 각국의 데이터 거버넌스는 데이터의 이용과 개인정보 보호를 조화시키기 위해 다양한 수준에서 원칙과 절차를 마련하고 있다.

아래에서는 주요한 몇 가지 이슈별로 해외 사례에서 참조할 수 있는 모범 관행을 검토해보고자 한다.

1. 데이터의 보호 및 활용 관련 법제

OECD의 2015년 보고서에서 지적한 바와 같이, 데이터 거버넌스 체제를 구축하는 데 있어서, 데이터의 이용 및 개인정보 보호를 규율하는 법제는 가장 중요한 요소이다. 3장 1절에서 살펴본 바와 같이 세계 주요 국가들은 개인정보 보호 법제 및 보건의료 관련 법제, 통계 관련 법제에서 개인정보 보호, 연구 및 통계와 같은 공익 목적을 위한 개인정보의 활용, 개인정보 활용 시 안전조치 등에 대한 규정을 포함하고 있다.

프랑스의 개인정보보호법인 '1978년 1월 6일 정보, 파일 및 자유에 관한 법률'을 제외하고는 개인정보 보호 법제에서 데이터 연계에 대해 명시하고 있는 경우는 없었다. 다만, 데이터 연계는 개인정보의 '처리'로서 일반적인 개인정보 보호 법제의 적용을 받는다. 유럽연합의 일반정보보호규정(GDPR)을 비롯하여 대부분 국가에서 공익을 위한 아카이브, 학술 연구, 통계를 목적으로 일정하게 개인정보의 활용을 허용하고 정보주체의 권리를 제한하는 규정을 두고 있었다. 그러나 학술 및 통계 목적의 활용이라고 무조건 개인정보 보호의 예외가 되는 것은 아니다.

GDPR은 공익적인 기록 보존, 과학 및 역사 연구 또는 통계 목적을 위해 필요한 경우 민감정보를 처리할 수 있도록 하면서도 법률에 근거하고, 이 법률은 추구하는 목

적에 비례하고 개인정보보호권의 본질을 존중하며 정보주체의 기본권 및 이익을 보호하기 위해 적절하고 구체적인 조치를 제공하는 것이어야 한다는 요건을 두고 있다.

동의를 얻는 것이 현실적으로 불가능하거나, 지나치게 비용이 많이 들거나 기술적으로 어려운 경우로 한정하기도 한다. GDPR 제89조의 2항 및 3항, 영국 NHS법 2006의 Section 251 지원 조항, 독일 연방정보보호법 제28조, 미국 HIPAA에서 기관평가위원회의 이용 승인 조건 등이 그렇다. 이와 함께 독일 연방정보보호법과 같이 학술적 이익이 개인정보 침해의 위험보다 훨씬 클 것을 요구하기도 한다.

대부분 법률에서 연구 및 통계 목적으로 개인정보를 제공하는 경우, 적절한 안전조치를 취할 것을 전제로 하고 있다. 예를 들어, GDPR은 이러한 안전조치가 가명화를 포함하여 데이터 최소화 원칙을 보장하기 위한 기술적·조직적 조치의 구비를 보장하는 것이어야 한다고 규정하고 있다. 또한, 연구 및 통계 목적으로 제공된 정보는 해당 목적으로만 사용되어야 한다.

연구 및 통계 목적으로 개인정보를 제공할 경우에, 제안서의 요건, 이용 주체의 자격(예를 들어, 승인된 연구기관 혹은 연구자 요건), 이용의 조건(안전시설 내에서의 이용, 계약의 체결, 연구 결과물의 검토) 등을 보다 구체적으로 규정하기도 한다. 예를 들어, EU 통계 규정은 연구기관의 승인, 연구 제안서의 적절성, 기밀 정보의 유형을 적시할 것, 인가된 접근시설에서의 이용, 국가통계 당국의 승인, 기밀성 동의서 작성 등의 세부사항을 규정하고 있다. 독일 연방통계법도 고등교육기관, 혹은 기밀성 서약을 한 사람 등으로 개인정보를 제공받을 수 있는 주체의 제한, 통계청의 특별보호구역 내 등 접근장소의 제한 등의 규정을 두고 있다. 뉴질랜드 통계법 역시 정부통계관의 승인, 연구자의 승인, 연구 목적의 승인, 안전한 환경에서 제공 등의 요건을 두고 있다.

또한, 대다수 국가의 법률들이 투명성을 강조하고 있다. 데이터 제공기관의 정책, 승인이 필요할 경우 그 기준 및 승인 목록 등을 공개하도록 하고 있다. 또한, 대부분의 국가 통계법에서는 개별 정보에 접근할 수 있는 사람을 제한하고 있다.

프랑스의 경우, 보건 분야의 공익적인 연구, 조사 혹은 평가를 목적으로 한 개인정보의 처리, 그 목적이 서로 다른 공익을 위한 파일들의 연계 등은 개인정보 감독기구인 CNIL의 허가를 받도록 하고 있어, 이와 같은 공익 목적의 연구 승인 기관으로서 감독기구의 역할이 매우 크다. 독일의 경우에는 통계법에서 데이터 연계에 대해서 구체적으로 규정하고 있다.

관련 법제의 내용과 함께, 국가 전체적인 법체계도 개인정보의 보호 및 활용에 큰 영향을 미친다. 즉, 각국은 개인정보 보호 법제, 통계 관련 법제, 보건의료 관련 법제

등을 통해서 연구 및 통계 목적 등을 위한 개인정보의 보호 및 활용을 규정하고 있는데, 이러한 법제의 통일성과 일관성이 없을 경우 법적 규율에서 벗어나는 경우가 발생할 수 있다. 특히 사회 전 영역을 포괄하는 개인정보 보호법제가 없는 경우, 혹은 연방제적 특성이 강해 각 주마다 법적 규율이 다른 경우, 법적 규율 체제가 복잡해지고 공백이 발생할 가능성이 커지게 된다. 예를 들면, 미국의 경우 사회 전 영역을 포괄하는 개인정보 보호 법제가 부재하고, 개인 건강정보의 경우 주로 HIPAA의 규율 대상이 되지만, 일부 민간 영역의 개인 건강정보의 경우 HIPAA의 규율 대상에서 제외되어 있다. (OECD, 2015)

정리하자면, 다음과 같은 요소를 개인정보 보호 법제 등 관련 입법에 반영할 필요가 있다.

첫째, 특정한 경우 개인정보의 공익적인 활용을 위해 일정하게 정보주체의 권리를 제한할 수 있다. 대부분 국가에서 이는 공익적인 아카이브, 학술 연구, 통계 목적으로 규정하고 있다. 연구 및 통계 목적인 경우에도 동의를 받는 것이 현실적으로 불가능한 경우로 제한할 필요가 있다.

둘째, 공익적인 목적으로 개인정보를 제공·활용할 경우에도 개인정보 보호를 위한 안전조치가 취해져야 한다. 정보보안을 위한 기술적·조직적 조치가 취해져야 하며, 연구 목적에 필요한 이상으로 개인 식별이 가능한 정보를 이용해서는 안 된다.

셋째, 정보를 제공받는 목적(공익 목적의 학술 연구 및 통계), 데이터 이용 주체(연구기관, 연구자)의 조건, 데이터 접근의 조건(안전시설), 데이터 활용의 조건(안전조치, 계약, 연구 결과물의 공개 전 검토 등), 제반 정책 및 승인 절차 등을 구체적으로 규정할 필요가 있다.

넷째, 데이터 연계와 관련해서는 유엔의 <통계 및 관련 연구 목적을 위해 수행되는 데이터 통합의 기밀성 관련 원칙과 가이드라인>을 각국의 법제 및 가이드라인 반영할 수 있을 것이다. 예를 들어, 뉴질랜드의 경우 데이터 연계를 위한 <데이터 통합 매뉴얼(Data Integration Manual)>을 작성하였는데 2006년의 첫 번째 매뉴얼을 업그레이드한 2015년의 두 번째 매뉴얼[160]은 이 원칙과 가이드라인을 반영한 것이다.

한편, OECD는 2015년 보고서에서 프라이버시 보호적인 건강정보 사용을 위한 입법 체계 주요 요소를 제안하고 있는데, 이는 제2장 2절에서 다룬 바 있다.

물론 법에서는 이와 같은 큰 틀에서의 원칙을 규정할 뿐, 구체적으로 특정 연구가 얼마나 공익적인지, 연구기관이나 연구자가 자격을 갖추었는지, 기술적·조직적 안전조

160) http://www.stats.govt.nz/methods/data-integration/data-integration-manual-2edn.aspx

치는 제대로 갖추어졌는지 등에 대한 판단이 비례적으로 이루어지기 위해서는 데이터 거버넌스 체제가 제대로 갖춰질 필요가 있다.

2. 거버넌스

데이터 거버넌스 체제는 개인정보 보호 법제를 포함한 각국의 법을 준수해야 한다. 그러나 법적 규제만으로 데이터 이용과 관련한 모든 문제를 다룰 수는 없다. 법 규정만으로 변화하는 사회 및 기술 환경에서 이루어지는 복잡한 연구의 적법성 여부를 일일이 판단하기는 쉽지 않기 때문이다. 따라서 좋은 거버넌스 체제는 법제로 규정하기 힘든 '회색 영역(gray area)'을 다룰 수 있어야 한다. 위험에 대한 평가에 기반하여 데이터의 접근, 공유, 연계 제안을 판단할 수 있는 적절한 기구가 필요하다. Rosalyn은 원칙에 기반하여, 연구의 공익과 프라이버시 보호의 균형을 추구할 수 있는 '원칙에 기반한, 비례적인 거버넌스'가 필요하다고 제안한다. (Rosalyn Moran, 2016)

좋은 거버넌스를 위해서는 데이터의 접근, 이용, 관리, 공유 등을 위한 원칙이 세워져야 한다. 이와 관련해서는 제2장 2절에서 다룬 바 있는, OECD의 데이터 거버넌스 프레임워크, 건강 데이터 거버넌스에 대한 OECD 이사회의 권고, UN의 공식통계 기본원칙, 유엔의 <통계 및 관련 연구 목적을 위해 수행되는 데이터 통합의 기밀성 관련 원칙과 가이드라인> 등을 참고할 수 있을 것이다.

해외 사례에서 볼 수 있다시피, 이러한 거버넌스 기구는 역할에 따라 다음과 같이 구분할 수 있다.

첫째, 정보 거버넌스 기구

둘째, 프로젝트 승인 기구

셋째, 연구윤리위원회

스코틀랜드의 '공익과 프라이버시 패널(Public Benefit and Privacy Panel for Health and Social Care, PBPP)'의 경우처럼 첫 번째, 두 번째 역할을 함께 하는 기구도 있을 수 있다. PBPP는 스코틀랜드에서 보건의료 관련 프라이버시 및 정보 거버넌스를 관장하고, 보건의료 데이터의 연구 목적 이용을 검토, 승인하는 역할을 한다. PBPP는 자신의 활동에 관한 원칙 문서(Principles that guide the PBPP)를 두고 있다.[161] 이 원칙문서는 NHS 스코틀랜드의 데이터 사용 신청을 고려할 때 적용이 되는데, 다음과 같은 원칙으로 이루어져 있다. △ 프라이버시, △ 공익, △ 적절한 과학

161) PBPP, Guiding Principles and Policy for Decision Making.
http://www.informationgovernance.scot.nhs.uk/pbpphsc/home/about-the-panel/

(appropriate science) : 해당 연구가 과학적으로 타당하며, 윤리적인 문제가 없음을 입증해야 함, △ 동의 : 가능할 경우, 정보주체의 동의를 얻어야 함, △ 투명성, △ 익명화, △ 프라이버시에의 영향, △ 안전조치, △ 보안, △ 비례성, △ 전례(precedent).

영국의 ADRN의 경우에는 자문이사회, 사무국장 그룹, 운영그룹과 같은 거버넌스 구조를 가지고 있고, 독립적 전문가 및 비전문가로 구성된 승인 패널이 연구 제안서의 심사를 담당한다. 호주 PHRN의 경우에는 이사회가 감독 및 전략적인 방향을 설정하며, 연계 데이터에의 접근을 위해 데이터 연계기구, 데이터 보유기관, 인간연구윤리위원회의 승인이 필요하다.

대부분 국가가 (데이터 연계와 무관하게) 독자적인 연구윤리위원회를 두고 있으며, 연구 주제에 따라 프로젝트에 대한 승인 조건으로 연구윤리위원회의 승인을 얻도록 하고 있다. 한국의 경우에도 <생명윤리 및 안전에 관한 법률>에 따라 인간 대상 연구를 하는 경우 '기관생명윤리위원회'의 승인을 받도록 하고 있다.

아이슬란드와 덴마크 등 일부 국가에서는 개인정보 감독기구가 연구신청서 승인 기구의 역할을 하고 있다. 아이슬란드는 개인정보감독기구가 관련 데이터 보유기관의 확인 및 국가바이오윤리위원회의 의견을 고려하며 연구신청서의 승인을 담당하고 있다. 덴마크에서는 개인정보 감독기구가 ① 등록소(registry)를 다른 등록소나 데이터셋에 연계, ② 새로운 등록소 생성에 대한 요청을 승인한다. 덴마크에는 2개의 연구윤리위원회가 있으며, 하나는 건강 데이터, 다른 하나는 다른 데이터를 담당한다. 일부 등록소의 경우에는 개인정보 감독기구 승인 전에 연구윤리위원회 승인을 요구하기도 한다. (OECD, 2015) 앞서 언급했듯이, 프랑스에서는 보건 분야의 공익적인 연구, 조사 혹은 평가를 목적으로 한 개인정보의 처리, 그 목적이 서로 다른 공익을 위한 파일들의 연계 등은 개인정보 감독기구인 CNIL의 허가를 받도록 하고 있다.

OECD는 이와 같은 검토와 승인 절차의 원칙으로 '증거기반(evidence-based)' 평가가 이루어져야 하고, 객관적이고 공정해야 하며, 적시에 일관성을 촉진하는 방식으로 이루어져야 한다고 권고하였다. 또한, 정보 처리가 개인 및 사회에 미치는 편익과 위험성, 그리고 위험성 경감을 평가하는 데 필요한 전문성을 가진 이들에 의해 수행되는 독립적이고 학제적인 검토가 이루어져야 함을 포함하고 있다. (OECD, 2017) 영국의 행정데이터작업반도 이러한 인증(accreditation) 절차가 각 기관별로 이루어질 경우 평가의 일관성이 떨어질 수 있음을 우려하면, 전국적인 인증 절차를 마련할 것을 권고하고 있다.

데이터 거버넌스 기구는 데이터의 수집·이용·제공과 관련한 구체적인 원칙·정책·절차 등을 가이드라인이나 매뉴얼로 정리하고 공개할 필요가 있다. 앞서 보았듯이, 뉴질랜드의 경우 데이터 연계를 위한 <데이터 통합 매뉴얼(Data Integration Manual)>을

작성하여 공개하고 있으며, 스코틀랜드의 '공익과 프라이버시 패널(PBPP)' 역시 자신의 활동과 관련한 원칙문서를 두고 있다. 독일에서는 사회 의료 및 예방 협회(German Society for Social Medicine and Prevention, DGSMP) 산하에 구성된 '설문조사 및 2차 데이터 이용을 위한 워킹그룹(Working Group for the Survey and Utilization of Secondary Data , AGENS)'이 '데이터 2차 분석 모범 사례'를 제시하고 있다.[162]

국내에서는 개인정보의 비식별화 등 데이터의 개인 식별성을 어떻게 최소화할 것인지에 대해서만 초점을 맞추는 경향이 있다. 그러나 비식별화는 전반적인 거버넌스의 하나의 요소일 뿐이다. 데이터의 이용과 보호에 관련된 법제, 데이터 접근·연계 정책, 연구기관 혹은 데이터 연계기관의 인증, 심사절차, 데이터 접근 절차 등에 이르는 전반적인 거버넌스 체제를 갖추도록 노력할 필요가 있다.

3. 연구 데이터 허브

연구 데이터 허브는 데이터를 보유하거나 접근할 수 있도록 하는 역할을 한다. 데이터에 체계적으로 접근할 수 있는 체계가 없다면, 연구자들은 데이터에 접근하기 위하여 해당 데이터 보유기관을 개별적으로 접촉해야 하고, 이에 필요한 시간과 비용 때문에 현실적으로 연구를 제약하는 요인으로 작용할 수밖에 없다. 또한, 각 데이터 보유기관이나 데이터의 성격에 따른 법적 요건을 판단해야 하고, 기관마다 그 판단이 달라지거나 혹은 법적 문제를 우려해서 각 기관이 데이터를 제공하는 것을 주저할 수도 있다. 연구 데이터 허브는 연구자를 대신해서 데이터 보유기관과 데이터 접근에 대해 협의하고, 데이터 보유기관에 법적 자문을 제공할 수 있다. 또한, 데이터 보유기관으로부터 데이터를 제공받아 안전하게 보유, 관리하고, 연구자가 안전한 환경에서 데이터에 접근할 수 있도록 하는 역할을 한다.

연구 데이터 허브는 두 가지 형태가 있을 수 있는데, 직접 데이터를 보유하지는 않고 연구 프로젝트별로 데이터 보유기관으로부터 데이터를 받아 접근을 매개하는 연합형(federated type)과 각국의 통계청과 같이 원래의 데이터 제공기관으로부터 일상적으로 데이터를 제공받아 보유, 관리하는 중앙형(centralized type) 모델이 있을 수 있다. (Rosalyn Moran, 2016)

앞서 살펴보았던 사례에서 잉글랜드의 CPRD, 웨일즈 SAIL Databank, 미국 연방조사국의 DLI, 네덜란드 통계청의 SSD는 비식별화된 데이터를 직접 보유하고 있는 중

162) Good Practice in Secondary Data Analysis. 자세한 내용은 부록 참조.

앙형 모델(centralized model)이고, 스코틀랜드의 eDRIS, 북아일랜드의 BSO HBS, 호주 PHRN 등은 데이터를 직접 보유하지는 않고 데이터 보유기관을 통해 연계 및 접근 서비스를 제공하는 연합형 모델(federated model)이라고 할 수 있다.

데이터 허브는 데이터에 대한 안전한 접근, 이용, 공개에 대한 통제, 검토, 보유, 파기 등의 제반 절차와 관련하여 원 데이터 보유기관 및 규제자와 계약을 체결한다. 개인정보 감독기구는 이 계약에 대해 의견을 제시할 수 있다.

영국의 행정데이터 작업반은 행정데이터연구센터(ADRC)가 데이터 연계 방법에 관련된 이슈를 탐구할 뿐 아니라 이 데이터를 이용하여 연구할 수 있는 역량을 갖춰야 한다고 권고하고 있는데, 실제로 많은 해외 연계기관에서 다른 연구자의 연구를 지원하는 것과 동시에 자체적으로 연구를 수행하고 있다.

데이터 공유의 활성화를 위해서는 데이터 보유기관 사이의 조정(co-ordination)이 중요하다는 것도 주목할 필요가 있다. OECD는 기관 간의 협력을 통해, 공통의 데이터 요소 및 포맷, 데이터 품질의 보장, 데이터 상호운용성 표준 등을 촉진하고, 데이터 공유의 장벽을 최소화하기 위한 공통의 정책과 절차를 개발할 것을 촉구하고 있다. 또한, 데이터 보유기관의 데이터 시스템도 검토되어야 하는데, 이는 개인정보 및 데이터 보안의 보호, 데이터의 접근 가능성, 데이터 품질과 활용 적합성, 접근성을 포함하며, 데이터셋 이전이나 연계가 허용될 수 있는 요소들이 무엇인지 역시 포함해야 한다. (OECD, 2017)

4. 데이터 연계의 방식

제2장에서 영국 '행정데이터 작업반'이 검토한 4가지 데이터 연계 모델을 살펴본 바 있다. 앞서 본 해외 사례에서는 주로 영국의 기관들이 '신뢰할 수 있는 제3자 색인(trusted third party indexing)' 모델(TTP)을 채택하고 있다.[163]

다만, 구체적인 데이터 연계 방식은 기관마다 조금씩 다르다. 통상적으로 TTP는 데이터 연계를 수행하는 기관과 분리되어 있다. 영국 잉글랜드의 CRPD의 경우에는 NHS Digital, 웨일즈의 SAIL Databank의 경우에는 NHS 웨일즈 정보서비스(NWIS),

163) 영국의 경우, 2014년 3월부터 정부의 데이터 공유 정책에 대해 정부, 기업, 시민사회 등 각 이해당사자가 참여하는 협의가 1년 동안 진행되었다. 2015년 3월에 협의 결과를 담은 최종 보고서가 발표되었는데, 정부를 포함하여 협의에 참여한 대표자들은 연구 목적의 데이터 연계를 위한 공공기관이 필요하다는 것에 합의하였다. 다만, 이는 데이터 연계가 '신뢰할 수 있는 제3자(TTP)'에 의해 수행될 것을 전제로 하였다. 이는 비식별 데이터가 보안 접속 시설 내에서 연계되고 통제된 조건에서 연구자에게 제공되는 절차를 포함한다. 또한, TTP, 연구자, 연구 주제 모두 법에 규정된 시스템 하의 인증 기관에 의한 승인을 받아야 하며, 연구는 '공익'에 기여한다는 특정한 조건을 충족해야 한다. (http://datasharing.org.uk/conclusions/index.html)

스코틀랜드의 eDRIS는 스코틀랜드 통계청인 NRS가 TTP의 역할을 하고 있다.

그러나 호주 PHRN의 경우에는 외부 기관을 TTP로 두기보다는 PHRN의 데이터연계기구(Data Linkage Unit) 내에서 연계를 담당하는 부서와 고객 서비스 및 데이터 전달을 담당하는 부서를 엄격하게 분리하고 있다. (영국 '행정데이터 작업반'의 구분에 따르면, '방화벽 단일 센터' 모델이라고 할 수 있다.) 개인정보 보호를 위한 이 원칙을 '분리 원칙(separation principles)'이라고 한다.

- 연계 데이터와 콘텐츠 데이터의 분리 : 연계를 위한 개인정보는 콘텐츠 데이터와 분리되어, DLU의 데이터 연계자에게는 연계 ID 생성을 위한 개인정보만 제공된다.
- 기능과 책임의 분리 : 데이터 연계절차와 데이터 보유 및 추출 기능을 분리한다. 데이터 연계를 수행하는 사람은 연구자와 분리되어 있어야 하며, 콘텐츠 데이터 연구에 참여할 수 없다.

미국의 데이터연계기반(DLI), 네덜란드 통계청의 SSD, 캐나다 통계청의 SDLE, 뉴질랜드 통계청의 IDI의 경우에도 외부 기관을 TTP로 두지 않고, 기관 내부에서 연계를 수행하고 있다. 이는 통계 데이터 수집 자체를 목적으로 하는 기관의 특성 때문인 것으로 보인다. 그러나 이 기관들 역시 개인정보 보호를 위한 별도의 조치를 취하고 있다. 미국 데이터연계기반을 운영하는 인구조사국은 제한된 담당자만 개인정보에 접근하도록 하고 있고 업무에 필요한 작업만 해야 하며, 법에 따라 기밀성 서약을 해야 한다. 네덜란드 통계청의 SSD 역시 업무를 위해 SSD 데이터를 필요로 하는 직원만이 접근 가능하며, SSD 네트워크의 일부, 그리고 필요한 데이터에만 접근이 허용되며, SSD에 접근하고 있는 사람은 동시에 네트워크의 다른 부분에 대한 접근이 동시에 허용되지 않는 등 조직적, 기술적 보안 조치를 취하고 있다.

5. 개인정보 보호 및 보안 조치

앞서 살펴본 모든 기관에서는 개인정보 보호 및 보안 조치의 중요성을 강조하고 있다. 통계 및 연구 목적으로 데이터의 제공과 이용을 활성화하기 위해서라도, 이것이 개인정보 침해를 야기하지 않을 것이라는 신뢰가 핵심적이기 때문이다. 따라서 데이터의 수집, 저장, 연계, 제공 등 전 과정에 걸쳐서 개인정보 보호 및 보안을 위한 조치들이 취해질 필요가 있다. 각국의 데이터 제공기관은 이를 위한 원칙과 이를 각 단계에서 구현하기 위한 방안을 취하고 있다.

예를 들어, 영국 ADRN은 5가지 안전 원칙을 제시하고 있다.

- Safe people : 연구자의 자격요건 검증 및 훈련
- Safe project : 독립된 승인 패널에 의한 심사 및 윤리 평가
- Safe environment : 안전한 공간에서 연구가 수행될 수 있도록 보안 시설 운영
- Safe data : 데이터 비식별화 및 TTP에 의한 데이터 연계
- Safe results : 연구 결과물에 대해 엄격한 통계적 공개 검토(statistical disclosure checks) 수행

뉴질랜드 통계청 역시 연구자(Safe people), 프로젝트(Safe project), 환경(Safe settings), 데이터(Safe data), 결과물(Safe output)의 안전 등 영국 ADRN과 유사한 '5가지 안전조치' 체제('Five Safes Framework)를 갖추고 있다.

이러한 안전조치를 좀 더 세부적으로 들여다보면 다음과 같다.

(1) 연구 프로젝트의 평가

수집된 개인정보는 원칙적으로 애초 수집목적 외로 사용할 수 없다. 다만, 연구 및 통계 목적인 경우, 예외적으로 수집목적 외 이용을 허용하고 있는데, 이 경우에도 개인정보 침해 위험성 및 데이터 이용을 통해 얻을 수 있는 공익성에 대해 엄격한 심사를 하고 있으며, 개인정보 비식별화나 보안 조치 등 개인정보 보호를 위한 제반 조치를 취하도록 하고 있다.

이를 위해 각국은 데이터 제공을 받을 수 있는 연구 프로젝트의 조건을 제한하거나, 연구자의 자격요건을 규정하고 있다. ADRN은 연구 프로젝트를 해당 연구의 실현 가능성, 학술적 가치, 공익성, 프라이버시에 대한 영향을 기준으로 평가한다. SAIL Databank 역시 '공익에 기여할 수 있는 순수 연구 목적'으로만 접근을 제한하고 있다. 호주의 PHRN도 연계 데이터에 접근할 수 있는 연구의 자격요건을 다음과 같이 규정하고 있다.

- 공중 보건의 증진, 보호, 유지에 기여할 수 있는 연구를 촉진할 것
- 보건 서비스의 계획, 평가, 전달을 촉진할 것
- 일반적으로 보건 데이터 수집, 보건 관련 데이터의 연계, 보건 관련 통계의 편집 및 활용과 관련된 연구 방법에 대한 지식에 기여할 것.

연구 프로젝트의 승인과 별개로, 보건의료와 같은 특정 분야의 연구와 관련해서는 연구윤리위원회의 승인이 필요하다.

(2) 연구자 및 연구기관의 자격요건

연구자의 자격요건 역시 데이터를 적절하게 다룰 수 있는 능력을 갖춘 사람으로 제한하고 있다. 해외 대부분 기관에서는 데이터에 접근하기 전에 적절한 훈련을 받을 것을 요구하고 있다. 실제 연구자들이 접근하는 것은 비식별화된 데이터라고 할지라도 여전히 재식별될 위험이 남아있기 때문이다. 따라서 보안 시설 내에서만 이 데이터에 접근하도록 하는 한편, 연구자들이 개인정보의 의미, 법적·윤리적 맥락, 연구수행 과정에서 준수해야 할 행동 규범들에 대해 숙지하고 있을 필요가 있다. ADRN의 경우, SURE(Safe Users of Research data Environment) 훈련 프로그램을 운영하며, 이에 참여할 것을 요구하고 있다. 웨일즈의 SAIL Databank, 스코틀랜드의 eDRIS 역시 SURE나 혹은 이에 준하는 훈련을 받도록 하고 있다. 미국 인구조사국은 자 기관의 피고용인과 특별선서지위(Special Sworn Status, SSS)를 획득한 연구자에게만 데이터 접근을 허용하고 있는데, 이들은 매년 개최되는 데이터 관리 훈련 등을 이수해야 한다. 이러한 훈련은 한 번에 그치는 것이 아니라, 정기적으로 수행해야 한다.

이와 함께 통상 데이터에 접근하는 연구자들은 데이터를 제공하는 기관과 계약을 체결해야 하며, 이를 위반할 경우 법적 처벌을 받게 된다. 연구자뿐만 아니라, 기관 내에서 기밀 데이터에 접근할 수 있는 사람도 담당 직원으로 제한된다.

연구자의 소속 기관의 지위를 제한하는 경우도 있다. 예를 들어, EU 통계 규정은 위원회(Eurostat)가 기밀정보를 제공할 경우, 접근이 승인된 연구기관에 이루어질 것을 조건으로 하고 있다. 스코틀랜드 eDRIS의 경우에도 연구자가 승인된 기관(approved organization) 소속이어야 한다. 승인된 기관은 현재 대학, NHS, 지역 당국 및 스코틀랜드 정부 등 공공영역의 기관으로 제한되어 있다.

(3) 안전시설(Safe haven)

데이터에 대한 접근은 엄격한 안전시설(safe haven) 내에서 이루어진다. 스코틀랜드 정부는 <스코틀랜드 안전시설 헌장(A Charter for Safe Havens in Scotland) >에서 Safe Haven을 “기밀 개인정보가 안전하게(safely and securely) 처리, 분석, 이전될 수 있도록 보장하기 위해, 기관과 합의된 일련의 행정계약 하에서 일하는 훈련된 전문가의 지원을 받는, 보안이 되는 물리적 혹은 원격접근 환경”으로 정의하고 있다. (The Scottish Government, 2015)

안전시설은 데이터 허브의 시설을 이용하는 경우, 외부의 연구소나 대학의 보안 시설을 이용하는 경우, 원격접근을 허용하는 경우 등으로 나누어진다.

안전시설은 데이터 보안을 위한 엄격한 보안 조치가 취해지며, 데이터의 이전도 암호화된 보안 전송을 이용한다. 보통 보안 시설 내 전자기기 등 기록 가능한 매체의 소지가 제한되며 보안 시설 외부에서 데이터의 반입이나 반출도 제한하고 있다. 보안 시설 내에서 연구자들의 행위는 향후에 감사를 받을 수 있도록 모두 기록된다.

원격접근을 허용하는 경우에도 이용자에 대한 인증, VPN을 통한 보안 접근, 로그 기록의 모니터링 등 보안 조치가 이루어진다. 캐나다의 SDLE나 독일 FDZ와 같이, 원격접근의 경우에는 세부 데이터에 직접 접근하는 것이 아니라, 자신의 분석에 필요한 코드나 자료를 입력하고 분석 결과만을 받는 식으로 운영되기 한다.

안전시설이 비단 데이터 보호만을 목적으로 하는 것은 아니다. 보통 연구데이터센터라고 불리는 해외 연계기관의 안전시설은 연구자들의 연구를 지원할 수 있는, 강력한 하드웨어나 분석 소프트웨어와 같은 자원과 기술을 제공하기도 한다.

(4) 결과물 점검 및 공개 통제

연구 결과물은 개인정보 침해가 없도록 공개되기 전에 통계 전문가에 의해 철저하게 검토되어야 한다. 이와 관련하여 유럽통계시스템 체제(Network of Excellence in the European Statistical System)의 통계적 노출 제어(Statistical Disclosure Control)는 연구 결과물 검토와 관련한 조직적, 절차적 측면에 대한 권고 및 모범 관행에 대한 상세한 가이드를 제공하고 있다.

앞서 살펴본 대부분의 해외 사례에서, 각 기관은 연구자가 안전시설 내에서 연구 결과물을 가지고 나가기 전에, 연구 결과물을 검토하는 절차를 가지고 있었다. EU통계규정 역시 국가통계기구(NSI)가 기밀 정보의 물리적, 논리적 보호를 위해 필요한 모든 규제적, 행정적, 기술적, 조직적 조치를 취하도록, 즉 통계적 노출 제어를 하도록 하고 있다.

일부 데이터 보유자는 데이터가 정확하게 활용되었는지, 혹은 데이터 보유자에 대한 감사표시(acknowledgement)가 제대로 되었는지 확인하기 위해, 최종 결과물을 스스로 검토하기를 원할 수 있다. 또한, 연구자들은 자신의 공개된 연구 결과물(출판물, PPT)을 연구 데이터 허브에 제공하거나 공개하도록 요구받을 수 있다. 이는 투명성에 기여하고, 연구자나 정책결정자를 위한 정보 자원으로 활용될 수 있다.

(5) 보안 조치

데이터 암호화, 데이터 전송 시 보안, 접근 통제 등 데이터의 수집, 보관, 이전, 제

공 등 전 과정에 걸쳐 보안 조치가 취해질 필요가 있다. 이는 관련 법제에서 연구·통계 목적으로 개인정보를 제공할 때 기본적으로 요구하고 있는 조건이다.

예를 들어, SAIL Databank는 시스템 구축 단계에서부터 '프라이버시 중심설계(privacy by design)'에 기반하여 견고한 거버넌스 모델을 구축해왔는데, 이는 다음과 같은 원칙에 기반하고 있다.

- 안전한 데이터 전송 secure data transportation
- 신뢰할 수 있는 레코드 매칭 reliable record matching
- 식별 데이터의 익명화와 암호화 anonymization and encryption of identifiable data
- 공개 통제 disclosure control
- 데이터 접근에 대한 통제 controlling data access
- 제안서와 결과에 대한 면밀한 검토 scrutiny of proposal and results
- 정보 거버넌스에 대한 외부 인증 external IG verification

(6) 연구자와의 계약

해외 대부분의 기관에서는 각 기관과 연구자 사이에 계약을 체결하거나, 이용조건에 동의하도록 함으로써, 연구자들이 개인 식별을 시도하거나 연구 목적 외로 이용하는 등 이용규칙을 위반한 경우에 제재할 수 있는 조치를 취하고 있었다.

(7) 프라이버시 영향평가

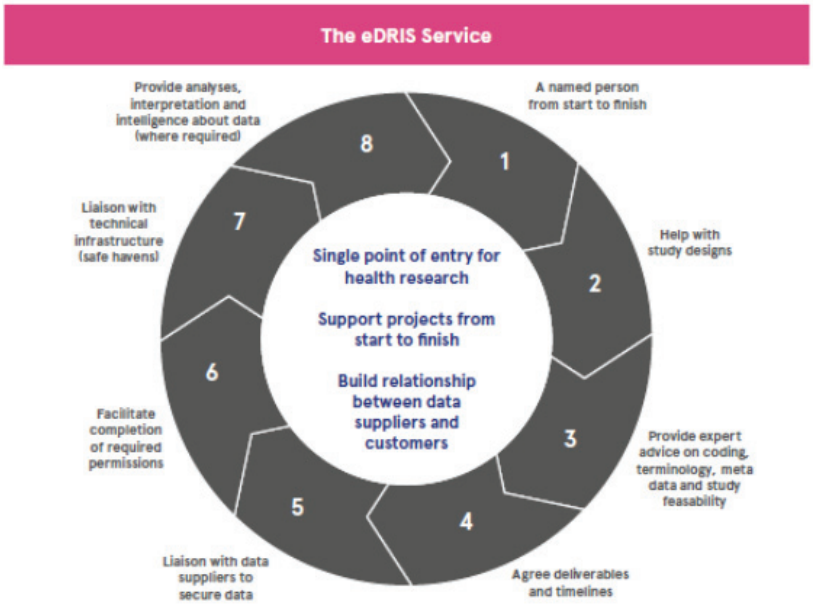
해외의 많은 기관이 개인정보 및 보안 조치의 적절성을 검토하기 위하여 프라이버시영향평가를 수행하거나, 각국의 개인정보 감독기구와 협의를 하고 있다. 예를 들어, 호주 PHRN, 캐나다 SDLE, 뉴질랜드 IDI 등에서 프라이버시 영향평가를 받고, 이를 공개하고 있었다.

6. 연구 지원단(Research Support Unit, RSU)

연구 지원단은 연구의 설계, 승인, 안전한 환경에서의 데이터 접근에 있어서 연구자들을 돕는 단일한 창구 역할을 한다. 이 장에서 소개한 대다수의 기관은 데이터 허브의 역할 뿐만 아니라 연구자들의 연구를 지원할 수 있는 다양한 역할을 하고 있었다.

영국 스코틀랜드의 eDRIS가 대표적인 사례이다. eDRIS는 전 과정에 걸쳐 연구 코디네이터를 지정하여, 연구설계 지원, 연구에 대한 전문가 자문, 데이터 접근에 필요한 허가 획득 지원, 데이터 제공기관과의 연결, 보안 시설을 통한 데이터 접근 등 제반 서비스를 제공하고 있다.

그림 3-19 eDRIS 의 연구지원 서비스



* 출처: Rosalyn Moran (2016)

Rosalyn은 RSU의 기능들을 다음과 같이 제안하고 있는데, 이 장에서 살펴본 각 기관은 이러한 기능들의 일부를 수행하고 있다.

- 양질의, 문서화된 데이터의 제공을 위한 데이터 제공자에 대한 지원 제공.
- 데이터 제공자와 RSU 사이의 데이터 공유계약의 체결
- 연구 프로젝트의 실현 가능성에 대해 연구자에게 가이드 제공
- 연구윤리위원회와 다른 거버넌스 기구에 연구 제안서를 제출하기 전에 조언과 의견 제공

- 연구자를 위한 훈련과 '승인된 연구자'로서 연구자 자격 획득 관리 및 제공
- 연구자와 RSU 사이에 허가 계약서 체결
- 보안 시설에의 접근 제공
- 보안 시설 이용의 감독
- 연구 데이터에 대한 접근 및 연구수행 지원
- 연구자를 위해 해당 연구에 필요한 데이터 생성
- 데이터 공개 절차 및 연구 결과물 검토
- 프로젝트 마무리
- DASSL 요소의 웹사이트 관리
- 대중 참여 및 소통 관리

연구자들의 데이터 접근 촉진을 위한 거버넌스 체제가 발전된 국가들은 연구지원단을 매개로 하여 데이터 접근에 관련된 정보를 투명하게 공개하고 있다. 즉, 홈페이지를 통해 데이터 접근을 위한 신청, 프로젝트 승인 요구조건, 승인 단계, 승인된 신청자를 위한 법적, 실질적 요구조건 등을 공개하고 있다. 이는 연구자에게 보다 공정한 데이터 접근의 기회를 제공할뿐더러, 데이터 이용에 대한 대중들의 신뢰도를 향상시킬 수 있다. (OECD, 2015) 이 장에서 소개한 각국의 기관들은 이러한 정보공개가 충실하게 이루어지고 있는 사례라고 볼 수 있다.

7. 대중 참여와 소통

공익적 연구 목적을 위한 데이터의 연계나 제공은 정보주체의 동의에 기반한 정보의 제공이나 애초 수집목적 내의 이용 등 일반적인 개인정보 보호원칙을 벗어난 이용에 해당한다. 따라서 이러한 이용이 사회적으로 용인되기 위해서는 데이터 제공 및 이용 과정 전반에 걸쳐서 개인정보 보호 및 보안이 지켜지고 있다는 것에 대한 일반 공중의 신뢰가 필수적이다. 그리고 이러한 대중의 신뢰를 얻기 위해서는 높은 수준의 투명성과 참여가 필수적이다. 즉, 원칙과 절차, 진행된 사업 내용에 대한 정보를 투명하게 공개해야 하며, 정책 결정 과정에 시민들과 다양한 이해당사자들이 참여할 수 있도록 해야 한다.

OECD 역시 공공적인 협의 과정을 통해서 광범위한 이해당사자들이 관여하고 참여

(engagement and participation)할 수 있도록 해야 한다고 권고하고 있다. 또한, 정보의 공개를 통해 개인정보 처리의 목적, 개인정보 처리를 승인하는 데 사용되는 절차와 기준, (보건의료) 데이터 거버넌스 체제의 실행 및 그 효과와 관련된 정보를 제공할 것을 권고하고 있다. (OECD, 2017)

승인된 연구 프로젝트와 관련한 정보를 공개하는 것은 개인정보 이용에 대한 대중의 신뢰를 높일 수 있다. 대중들이 개인정보가 누구에 의해서, 어떤 목적으로, 어떻게 사용되는지 알 수 있기 때문이다. (OECD, 2015) 예를 들어, EU통계규정은 위원회(Eurostat)가 승인된 연구기관의 목록을 웹사이트를 통해 공개하고, 정기적으로 재평가해야 하도록 하고 있다. 영국 CPRD의 경우, 연구 계획서의 승인을 담당하는 독립적 과학자문위원회(ISAC)는 CPRD와 다른 데이터 소스의 연계가 필요한 승인된 연구 계획서의 요약본을 공개하고 있다. 북아일랜드 HBS 역시 활동 결과물을 공익 목적으로 공개하고 개방적이고 공정한 지식에 기여하는 것을 목표로 하고 있다. 캐나다 통계청은 2000년 1월부터 마이크로데이터 연계 지침에 따라 평가되고 승인된 데이터 연계 목록을 홈페이지를 통해 공개하고 있다.

행정데이터작업반 역시 대중들의 참여를 강조하고 있는데, 이를 위해 연구 목적의 행정데이터 이용의 필요성에 대한 대중들의 인식을 높이기 위한 홍보, 정책 결정 과정에서 대중들과의 협의 및 이사회 등 결정단위에의 참여, 이를 위해 ADRC 내에 대중 참여와 소통을 담당할 사무관을 둘 것 등을 제안하고 있다.

대중들의 반대 여론에 따라 추진이 중단된 영국의 care.data 사례는 대중들과의 소통이 얼마나 중요한지 반증하는 사례로 자주 거론되고 있다.

8. 영리적 목적의 데이터 접근 및 연계

소비자들의 구매 이력, 금융 정보, 통신 내역, SNS 정보 등 기업들 역시 방대한 데이터를 보유하고 있으며, 빅데이터 기술의 발전과 함께 기업이 보유한 데이터에 대한 분석 및 활용 필요성도 증가하고 있다. 기업이 보유한 데이터와 공공기관이 보유한 행정데이터의 결합을 통해서 과학적 연구 및 공공 정책을 위한 가치 있는 결과가 산출될 수도 있다.

영국 행정데이터작업반의 보고서에서도 이 문제를 검토하고 있다. 영국 행정데이터작업반은 상업적인 데이터와 행정데이터 연계는 기업의 운영 효율성을 증진시킬 수 있는 정보를 제공할 수 있고, 공적 지원을 받은 연구자가 기업의 데이터에 접근하고 행정데이터와 연계함으로써 공익적 가치를 가지는 연구가 수행될 수 있는 잠재력이

있다고 보았다. 이와 함께, 영리적 목적의 행정데이터 접근 및 연계가 공공기관의 명성에 해를 끼칠 가능성도 고려할 필요가 있다고 한다. 행정데이터작업반은 자신들의 제안은 '공익 목적의 연구'에 한정하고 있음을 밝히며, 추후 ADRC의 거버넌스 기구에서 이 이슈를 검토하고 가이드라인을 만들 것을 권고하고 있다.

기업들의 행정데이터 접근이나 연계를 허용하더라도, 타겟 마케팅과 같은 영리 목적까지 허용할 것인지, 기업 연구소 등이 수행하더라도 공익적 가치가 있는 연구에 한정할 것인지, 기업이 재정을 지원하는 학술기관의 연구는 어떻게 판단할 것인지, 해당 연구 결과물은 공개하도록 할 것인지 등 세부적인 이슈들에 대한 검토가 필요해 보인다. 예를 들어, 영국의 CPRD의 경우 <CPRD 접근 라이선스 표준안>[164]에서 자신이 취득한 데이터나 정보를 a) 환자식별, 접촉, 타게팅, b) 일반개업의나 일반의료 행위의 식별, 프로파일링, 접촉, 타게팅, c) 광고 및 영업 효과 연구 등에 사용해서는 안 된다고 규정하고 있다.

스웨덴의 경우, 상업적 기관은 대학 연구자와 협력할 때 승인을 받기 쉬우며, 이 경우 대학 연구자가 데이터 접근에 대한 승인을 받고 개인 식별이 불가능한 연구 결과만을 상업적 기관에 제공하게 된다. 미국에서도 마케팅 목적의 이용의 경우 HIPAA에 의한 예외를 인정받지 못한다. 즉, 마케팅 목적으로 환자 정보를 이용할 경우에는 환자들의 동의가 필요하다. (OECD, 2015)

OECD(2015)는 대부분 국가에서 모든 핵심적인 보건의료 데이터셋의 비식별화된 마이크로데이터에 대해 영리적 기업의 접근 승인은 허용하지 않고 있다고 분석했다. 영리 기관에 데이터에 대한 접근을 승인하는 경우에도, 이는 공익적인 학술 연구나 통계 목적의 이용으로 제한된다. 이때 OECD는 '공익(public interest)'의 개념을 데이터 보호, 공공 보건, 사회적 보호(social protection), 보건의료 서비스의 관리, 보건의료 연구 및 통계를 포함하는 것으로 본다. (OECD, 2015)[165]

영리/비영리 여부를 불문하고 연구 프로젝트의 공익성을 근거로 데이터 접근에 대한 승인 여부를 결정하는 경우도 있지만, 어떤 국가들은 대학, 비영리, 공공부문의 신청자와 영리 부문의 신청자를 구분하는 경우도 있다. 예를 들어, 캐나다, 아이슬란드, 영국 웨일즈, 네덜란드, 미국 등이 이에 해당한다. 아이슬란드의 경우 영리 부문의 신청자는 데이터 접근의 승인을 받지 못하며, 단지 총계 데이터에만 접근할 수 있다. 영리 부문의 신청자가 비식별화된 등록부 혹은 연계 데이터에 접근하기 위해서는 정보주체의 동의를 얻어야 한다. 영국 웨일즈의 SAIL Databank의 경우에도 학계 및 비영리 부문의 연구자에게 비식별 개별 데이터에 대한 접근을 허용하고 있다. 다만,

164) 표준안의 전체 내용은 부록 참조.

165) OECD, 2015. p60.

H-Health 산업 혁신 센터를 통해 총계 데이터에 대한 접근을 허용한다. (OECD, 2015)

앞서 살펴보았듯이, 아직까지 해외 대부분의 기관에서는 연구자에게 개인 수준의 비식별 데이터에 대한 접근 및 연계를 허용하더라도, 개인정보 침해의 위험보다 큰 공익적 가치를 가지는, 혹은 기관의 임무(mission)에 부합하는 연구로 제한하고 있으며, 이를 위해 연구 프로젝트의 과학적 가치, 실현 가능성, 프라이버시 침해 가능성 등을 기준으로 검토하고 승인을 하는 절차를 거치고 있다.

그러나 (후술할 국내 '비식별조치 전문기관'과 같이) 기업들이 보유한 데이터베이스 사이의 연계를 공공기관이 지원하고 있는 사례를 찾아보기 힘들었다. 국내에서도 통계 및 연구 목적의 행정데이터 연계 및 접근을 활성화하는 방안을 고려할 때, 우선 '공익 목적'의 연구로 제한하여 신중하게 접근할 필요가 있다.

다음 <표3-2>는 앞서 설명한 주요 이슈별로 해외 사례를 비교한 것이다.

표 3-2 주요 해외 사례의 특징 비교

	연구 승인	데이터 허브	데이터 연계 모델	안전시설 (Safe Havens)	목적제한 (통계, 공익연구)	연구자 훈련	연구 결과물 검토	비고
잉글랜드 CPRD	ISAC[166]	중앙형	TTP (NHS Digital)	(온라인 접근)	○			매년 보건연구 당국의 Section 251 규제 승인, 환자들의 Opt-out 권리보장
웨일즈 SAIL Databank	IGRP[167]	중앙형	TTP (NWIS)[168]	SAIL gateway (원격접근 가능)	○	○	○	프라이버시 중심설계에 기반한 모델 구축
스코틀랜드 eDRIS	데이터보유기관 / PBPP[169]	연합형	NRS[170]	NSS National Safe Haven (원격접근 가능)	○	○	○	
북아일랜드 BSO HBS	HBGB[171]	연합형		HBS 내	○	○		연구 결과물 공개 필요
호주 PHRN	DLU[172], 데이터보유기관	연합형	내부 분리	SURE[173] (원격접근 가능)	○	○	○	프라이버시 영향평가 수행
미국 NCHS	검토위원회	중앙형		RDC[174]	○	○	○	
영국 ADRN	승인패널	연합형	TTP (지역별)	각 지역 ADRC	○	○	○	연구 결과물 공개
미국 DLI	PCO[175]	중앙형	내부 분리	FSRDCs[176] (원격접근 가능)	○	○	○	연구자는 특별선서지위(Special Sworn Status, SSS) 획득 필요
네덜란드 SSD	CCS[177]	중앙형	내부 분리	통계청 구내 혹은 원격접근	○			

캐나다 SDLE	통계청 / 최고 통계관	중앙형	내부 분리	RDC	○	○		연구자는 통계법상 '직원 간주' 지위 획득 필요 / 개인정보감독기구의 자문, 프라이버시 영향평가 수행
뉴질랜드 IDI	정부통계관/ 데이터보유기관	중앙형	내부 분리	보안 데이터랩 (원격접근 가능)	○	○	○	프라이버시 영향평가 수행
독일 GRLC/FDZ	연방노동 사회부	중앙형	내부 분리	FDZ 보안인프라 원격접근 (JoSuA)	○		○	

166) 독립적 과학자문위원회, Independent Scientific Advisory Committee
167) 독립적인 정보거버넌스검토패널, Information Governance Review Panel
168) NHS 웨일즈 정보서비스, NHS Wales Informatics Service
169) 공익과프라이버시패널, Public Benefit and Privacy Panel
170) 스코틀랜드 통계청, National Records of Scotland
171) Honest Broker Governance Board
172) 데이터 연계기구, Data Linkage Unit
173) 보안통합연구환경, Secure Unified Research Environment
174) 연구데이터센터, Research Data Center
175) 인구조사국 정책조정부, Census Bureau's Policy Coordination Office
176) 연방통계연구데이터센터, Federal Statistical Research Data Centers
177) 통계청의 독립적 감독기관인 중앙통계위원회, Central Commission for Statistics

제4장 국내 데이터 연계·결합 현황

제1절 데이터 연계·결합 관련 국내 법제

1. 개요

우리나라에서도 데이터 연계·결합은 널리 이루어져 왔는데, 특히 최근 빅데이터 분석 기술의 발달, 사물인터넷 기술의 발달 등 정보처리 기술의 고도화로 인해 데이터 연계·결합의 시도는 폭발적으로 늘어나는 추세다.

아래에서 보는 것처럼 우리나라 법제에서도 데이터의 연계·결합으로 인해 개인이 식별되거나 식별될 가능성이 있는 경우에는 원칙적으로 개인정보보호법의 적용대상이 되고 개인정보의 처리에 관한 근거가 있어야 하며 개인정보보호법의 여러 규정을 준수해야 한다.

개인정보의 처리에 관한 근거로는 크게 ① 개인정보주체의 동의, ② 법률이나 법령의 규정, ③ 공공기관이 업무를 수행하기 위해 불가피한 경우, ④ 기타 적법 요건으로 나누어 볼 수 있으므로 데이터의 연계·결합도 이런 적법 요건을 갖추어야 한다.

데이터 연계·결합은 공공부문과 민간 부문을 가리지 않고 왕성하게 이루어지고 있다. 그중 공공 부문의 경우는 ① 사회복지, 식품위생, 주택, 환경, 도로교통, 병무, 교육 등 각종 영역에서 행정서비스 제공의 요건 및 자격을 확인하거나 서비스 제공을 위하여, ② 공공복리나 질서유지, 기타 행정상의 규제를 준수하고 있는지 여부를 확인하기 위한 목적으로, ③ 정책 결정을 위한 분석을 하기 위한 목적으로, ④ 행정서비스의 질을 관리하거나, 질을 향상시키기 위해 또는 재정 건전성을 유지하는 등 관리 목적으로, ⑤ 연구를 위한 목적으로 이루어지기도 하는 등 다양한 영역에서 다양한 목적으로 이루어지고 있다.

민간 분야에서도 ① 기업의 내부적인 고용관계, ② 기업과 소비자 또는 서비스 이용자 사이, ③ 학술, 종교, 친목 단체 등 비영리단체를 포함한 다양한 영역에서 데이터 연계·결합이 활발하게 이루어지고 있다. 특히 기업들은 정보처리 기술의 고도화로 인해 용역이나 상품의 판매, 혹은 판매 촉진을 위한 고객 분석에 경쟁적으로 데이터 연계·결합을 시도하고 있다.

이하에서는 데이터 연계·결합과 관련한 법률 규정을 분석하고, 그중 데이터 연계·결합과 관련하여 필요성이 아주 크고, 제도의 도입이나 개선을 위한 요구가 강력하게 제기되고 있는 영역으로서 보건의료 분야 및 통계 분야를 검토한다. 이 분야는 다른

분야보다도 데이터 연계·결합의 공공적인 필요성과 이익이 큰 분야이다. 그래서 해외 각국에서도 데이터 연계·결합에 대한 나름의 법률과 제도 및 거버넌스를 발전시켜온 분야이다. 특히 우리나라에서도 부처별로 이 분야에 대한 제도 개선방안을 경쟁적으로 발표하고 있기 때문에 바람직한 법률과 제도, 거버넌스의 발전 방향을 검토해 볼 실익이 있다.

민간 분야의 경우는 데이터 연계·결합의 급증에 따라서 그 위험성과 개인정보주체의 자기결정권을 보장해 주기 위한 제도 개선 방향을 간략하게 검토해 보고, 우리나라에서 데이터 연계·결합의 독특한 방안으로 도입, 시행되고 있는 '전문기관을 통한 데이터 연계·결합'의 사례를 검토하고, 그에 따르는 법적 문제를 분석한다.

2. 데이터 연계·결합 관련 개인정보보호 규율

(1) 국내 개인정보보호 법제의 특징

가. 헌법상의 기본권의 지위를 갖는 개인정보 자기결정권

우리나라는 개인정보 자기결정권이 헌법에 명시되어 있지는 않지만, 이를 헌법의 기본권으로부터 도출되는 권리로 보는데 사실상 이견이 없다. 개인정보 자기결정권은 인간의 존엄과 가치, 행복추구권을 규정한 헌법 제10조 제1문에서 도출되는 일반적 인격권 및 헌법 제17조의 사생활의 비밀과 자유에 의하여 보장되는 권리라는 것이다.[178]

개인정보 자기결정권은 자신에 관한 정보가 언제 누구에게 어느 범위까지 알려지고 또 이용되도록 할 것인지를 그 정보주체가 스스로 결정할 수 있는 권리이기 때문에 개인정보를 대상으로 한 조사·수집·보관·처리·이용 등의 행위는 모두 원칙적으로 개인정보 자기결정권에 대한 제한에 해당한다.[179]

이와 같이 개인정보 자기결정권을 헌법상의 기본권으로 인정하고 있기 때문에, 개인정보 자기결정권 또는 그에 대한 제한을 내용으로 하는 입법 또는 해석은 헌법상 기본권 보장과 관련한 해석 원칙에 입각하게 된다.

나. 일반법과 각 영역의 특별법 사이의 목표 혼동과 모순

우리나라 개인정보 보호 법제는 일반법인 개인정보보호법이 있고, 영역별로 특별법

178) 헌법재판소 2005. 5. 26. 99헌마513·2004헌마190 결정 등, 대법원 2014. 7. 24, 선고, 2012다49933, 판결 등

179) 대법원 2014. 7. 24. 선고 2012다49933 판결, 대법원 2016. 8. 17. 선고 2014다235080 판결 등

이 제정되어 있지만, 일반법과 특별법의 적용 범위가 매우 복잡하게 얽혀 있다. 특히 분야별로 감독권한이 나뉘어 있는데다, 감독권한을 가진 기관 사이의 권한 배분이 애매하고, 일반법인 개인정보보호법의 관장 기관인 개인정보보호위원회의 권한이 그에 걸맞게 배분되어 있지 못하여 개인정보 보호법제의 통일성을 유지하기 매우 어렵게 되어 있다.

예를 들어 개인정보보호법에 의하면 개인정보보호위원회는 개인정보보호 기본계획 및 시행계획의 의결, 개인정보 보호와 관련된 정책, 제도 및 법령의 개선에 관한 사항, 개인정보 보호에 관한 법령의 해석·운용에 관한 사항의 의결을 하는 기관인데, 정보통신서비스 제공자와 이용자 사이의 '정보통신망 이용촉진 및 정보보호 등에 관한 법률'(이하 '정보통신망법') 적용 대상인 분야와 '신용정보의 이용 및 보호에 관한 법률'(이하 '신용정보법') 적용 대상인 분야의 경우, 정책. 제도, 법령 개선, 법령의 해석, 운용의 기본계획이 일관성 있게 수립, 집행되고 있지 못하고 있다.

여기에 덧붙여 서로 다른 정책 목표를 가진 법률들이 모순적으로 존재하면서 사실상 개인정보 자기결정권의 가치를 크게 훼손하는 현상이 나타나고 있다. 예를 들어 '공공기관의 정보공개에 관한 법률'(이하 '정보공개법'), 전자정부법, 공공데이터의 제공 및 이용 활성화에 관한 법률(이하 '공공데이터법')과 개인정보보호법의 관계이다.

표 4-1 분야별 법제 현황

범주	개인정보보호	정보공개	정보활용
해당 법률	개인정보보호법 정보통신망법 신용정보법 위치정보법 기타	정보공개법	전자정부법 공공데이터의 제공 및 이용 활성화에 관한 법률 데이터기반행정
목적	개인정보의 보호, 적법한 활용	정보의 공개를 통한 알권리, 투명성	행정편의, 데이터 활용

다. 데이터 연계·결합에 대한 법적 규율

개인정보에 관한 규율이 일반법과 특별법, 가치를 달리하는 정보공개법, 전자정부법, 공공데이터법 등과의 관계 속에서 해석되어야 하기 때문에 데이터 연계나 결합에 대한 법적 규율을 해석하기가 매우 복잡하다.

영역별로 법 적용이 달라질 뿐만 아니라, 특히 공공분야의 경우 정보공개나 공공데이터의 활용을 근거로 한 데이터의 연계나 결합이 별도로 허용되고, 전자정부법과 관련해서도 행정정보의 공동이용이라는 견지에서 데이터의 연계나 결합을 허용하는 별도의 규율을 두고 있기 때문이다. 이로 인해서 개인정보보호법과의 관계에 있어 규정의 통일적인 해석이 어렵고, 개인정보보호법의 개인정보 보호원칙과 규율들이 훼손되

는 결과를 빚고 있다.

개인정보 자기결정권이 가장 중요한 개인의 기본권 중 하나라는 점에서 그 지위에 걸맞도록 개인정보 자기결정권과 정보공개나 공공데이터의 활용, 행정 효율성과 행정정보 공동활용 등의 가치가 일관되고 통일적으로 해석되고, 통일적 추진체계나 컨트롤타워를 갖출 필요가 있다.

(2) 개인정보보호 법제에서 데이터 연계·결합의 취급

가. 데이터 연계·결합은 처리에 해당

우리나라 개인정보 보호법제의 기본법이자 일반법인 개인정보보호법은 개인정보를 '살아 있는 개인에 관한 정보로서 성명, 주민등록번호 및 영상 등을 통하여 개인을 알아볼 수 있는 정보(해당 정보만으로는 특정 개인을 알아볼 수 없더라도 다른 정보와 쉽게 결합하여 알아볼 수 있는 것을 포함한다)를 말한다'고 정의하고 있다(개인정보보호법 제2조 제1호). 개인정보보호법의 수범자는 대부분 개인정보처리자인데, 업무를 목적으로 개인정보파일을 운용하기 위하여 스스로 또는 다른 사람을 통하여 개인정보를 처리하는 공공기관, 법인, 단체 및 개인 등을 말한다(개인정보보호법 제2조 제5호).

'살아 있는 개인에 관한 정보'이어야 하기 때문에 사망한 사람인 경우는 개인정보의 주체가 되지 않는다. 따라서 사망한 사람에 대한 건강정보를 보유하는 것은 개인정보보호법의 적용대상이 되지는 않는다. 그러나 그 정보가 살아 있는 개인에 관한 정보를 제공해 주는 경우에는 그 살아 있는 개인에 대한 개인정보로 볼 수 있다.

여기서 '개인정보파일'이란 '개인정보를 쉽게 검색할 수 있도록 일정한 규칙에 따라 체계적으로 배열하거나 구성한 개인정보의 집합물(集合物)'을 말한다. (개인정보보호법 제2조 제4호)

개인정보보호법은 '개인정보의 수집, 생성, 연계, 연동, 기록, 저장, 보유, 가공, 편집, 검색, 출력, 정정(訂正), 복구, 이용, 제공, 공개, 파기(破棄), 그 밖에 이와 유사한 행위'를 개인정보의 '처리'라고 규정하고 있는데(개인정보보호법 제2조 제2호), 이처럼 개인정보의 '연계'는 개인정보보호법에서 개인정보 처리의 하나로 명시하고 있고, 결합은 명시하고 있지는 않지만, 연동이 이와 유사한 개념으로 볼 수 있다. 따라서 개인정보의 연계나 결합은 개인정보의 처리로서 법률의 규정을 준수해야 한다. 한편, 개인정보의 연계나 결합으로 인하여 개인에 대한 새로운 정보가 생성되는 것은 개인정보의 수집에 준하는 것으로 볼 수 있다.

처음부터 개인정보를 연계·결합하는 경우는 당연히 개인정보보호법의 적용대상이

되는데, 애초에는 개인을 식별할 수 없는 정보였으나 그 정보의 연계나 결합으로 인해서 개인을 식별하거나 식별 가능성이 생기는 경우에도 연계나 결합으로 인해서 개인정보가 되어 개인정보보호법의 적용대상이 된다.

이처럼 개인정보인 데이터의 연계·결합, 또는 데이터의 연계·결합으로 인해 개인을 식별하거나 식별 가능해지는 경우(개인정보보호법의 적용대상인 연계·결합)에는 개인정보보호법의 개인정보 처리에 대한 법률 규정이 적용된다.

나. 데이터 연계·결합과 정보처리의 원칙

개인정보보호법은 제3조에서 8가지 개인정보 보호원칙을 규정하고 있는데, 개인정보보호법의 적용대상인 데이터의 연계·결합 시에 이 원칙이 적용될 것인지가 해석상 논란의 소지가 있다. 그 여덟 가지 개인정보 보호원칙은 다음과 같다(제3조).

① 개인정보처리자는 개인정보의 처리 목적을 명확하게 하여야 하고 그 목적에 필요한 범위에서 최소한의 개인정보만을 적법하고 정당하게 수집하여야 한다. 따라서 과도한 개인정보의 연계와 결합은 개인정보 처리원칙에 위반되는 것으로 볼 수 있다.

이와 관련하여 개인정보보호법은 개인정보처리자는 보유 기간의 경과, 개인정보의 처리목적 달성 등 그 개인정보가 불필요하게 되었을 때에는 지체 없이 그 개인정보를 파기하여야 한다는 규정을 두고 있다. 다만, 다른 법령에 따라 보존하여야 하는 경우에는 그러하지 아니한데, 이 경우 해당 개인정보 또는 개인정보파일을 다른 개인정보와 분리해서 저장·관리하여야 한다.

② 개인정보처리자는 개인정보의 처리 목적에 필요한 범위에서 적합하게 개인정보를 처리하여야 하며, 그 목적 외의 용도로 활용하여서는 아니 된다. 따라서 개인정보 데이터의 연계한 후 이를 다른 목적으로 전용해서는 안 된다.

③ 개인정보처리자는 개인정보의 처리 목적에 필요한 범위에서 개인정보의 정확성, 완전성 및 최신성이 보장되도록 하여야 한다.

④ 개인정보처리자는 개인정보의 처리 방법 및 종류 등에 따라 정보주체의 권리가 침해받을 가능성과 그 위험 정도를 고려하여 개인정보를 안전하게 관리하여야 한다. 안전한 관리는 위험성과 가능성을 고려하여 그 수준에 맞는 관리가 이루어져야 한다는 것이다.

⑤ 개인정보처리자는 개인정보 처리방침 등 개인정보의 처리에 관한 사항을 공개하여야 하며, 열람청구권 등 정보주체의 권리를 보장하여야 한다. 이와 같은 공개 및 열람청구권은 개인정보 주체가 모르는 상태에서 결합을 통해서 새로운 개인정보가 생성

된다는 점에서 특히 중요하다.

⑥ 개인정보처리자는 정보주체의 사생활 침해를 최소화하는 방법으로 개인정보를 처리하여야 한다. 따라서 데이터의 연계·결합이 필요한 경우에도 사생활 침해를 최소화할 방안을 모색하여 처리해야 한다.

⑦ 개인정보처리자는 개인정보의 익명처리가 가능한 경우에는 익명으로 처리될 수 있도록 하여야 한다. 개인정보의 결합·연계의 경우도 마찬가지로 익명처리가 가능하다면 익명처리를 원칙으로 한다.

⑧ 개인정보처리자는 이 법 및 관계 법령에서 규정하고 있는 책임과 의무를 준수하고 실천함으로써 정보주체의 신뢰를 얻기 위하여 노력하여야 한다.

그런데 이 원칙들은 개인정보보호법 제3장(개인정보의 처리), 제4장(개인정보의 안전한 관리)에서 개별 조항으로 다시 구체적으로 규정되어 있다.[180] 특히 최근 대법원은 홈플러스 개인정보 유출 사건에서 '개인정보 처리원칙 위반은 개인정보보호법을 위반하는 것'이라고 판시하여, 개인정보 처리원칙이 단순히 선언적 규정에 그치는 것은 아니라는 점을 분명히 하고 있다.

> 이 사건 경품행사에 응모한 고객들은 응모권 뒷면과 인터넷 응모화면에 기재되어 있는 '개인정보 수집 및 제3자 제공 동의' 등 사항이 경품행사 진행을 위하여 필요한 것으로 받아들일 가능성이 크다. 그런데 응모권에 따라서는 경품추첨 사실을 알리는 데 필요한 개인정보와 관련 없는 '응모자의 성별, 자녀 수, 동거 여부' 등 사생활의 비밀에 관한 정보와 심지어는 주민등록번호와 같은 고유식별정보까지 수집하면서 이에 관한 동의를 하지 않을 때는 응모가 되지 아니하거나 경품 추첨에서 제외된다고 고지하고 있다. 이는 개인정보처리자가 정당한 목적으로 개인정보를 수집하는 경우라 하더라도 그 목적에 필요한 최소한의 개인정보 수집에 그쳐야 하고 이에 동의하지 아니한다는 이유로 정보주체에게 재화 또는 서비스의 제공을 거부하여서는 안 된다는 개인정보 보호 원칙(개인정보보호법 제3조 제1항)과 개인정보보호법 규정에 위반되는 것이다. (대법원 2017.4.7, 선고, 2016도13263, 판결)

다. 데이터 연계·결합 시의 정보주체의 권리

개인정보보호법의 적용대상이 되는 데이터 연계·결합 시 정보주체의 권리는 보장되어야 한다(개인정보보호법 제4조). 개인정보보호법이 보장하고 있는 권리는 1. 개인정

---

180) 그래서 개인정보보호법이 제58조에서 개인정보보호법 제3장부터 제7장까지를 적용하지 않는다고 한 경우에, 즉, 공공기관이 처리하는 개인정보 중 '통계법'에 따라 수집되는 개인정보, 국가안전보장과 관련된 정보 분석을 목적으로 수집 또는 제공 요청되는 개인정보, 공중위생 등 공공의 안전과 안녕을 위하여 긴급히 필요한 경우로서 일시적으로 처리되는 개인정보, 언론, 종교단체, 정당이 각각 취재·보도, 선교, 선거 입후보자 추천 등 고유 목적을 달성하기 위하여 수집·이용하는 개인정보인 경우에도 제3조의 원칙을 준수해야 하는지, 아니면 제3장, 제4장이 적용 배제되므로 적용되지 않는다고 볼 것인지가 문제 되는데, 기본 원칙으로는 적용된다고 볼 것이다. 이 경우 '개인정보처리자는 개인정보 처리방침 등 개인정보의 처리에 관한 사항을 공개하여야 하며, 열람청구권 등 정보주체의 권리를 보장하여야 한다'는 보호원칙은 어떻게 준수해야 할지가 논란이 될 수 있다.

보의 처리에 관한 정보를 제공받을 권리, 2. 개인정보의 처리에 관한 동의 여부, 동의 범위 등을 선택하고 결정할 권리, 3. 개인정보의 처리 여부를 확인하고 개인정보에 대하여 열람(사본의 발급을 포함한다. 이하 같다)을 요구할 권리, 4. 개인정보의 처리 정지, 정정·삭제 및 파기를 요구할 권리, 5. 개인정보의 처리로 인하여 발생한 피해를 신속하고 공정한 절차에 따라 구제받을 권리이다. 개인정보호보법은 이를 개별 조문으로 구체적으로 규정하고 있다.

① 개인정보주체는 개인정보처리자로부터 개인정보 처리에 관한 정보를 제공받을 권리가 있다. 이와 관련하여 개인정보보호법은 당사자의 동의를 얻어서 개인정보를 수집하거나(제15조), 당사자의 동의를 얻어서 개인정보를 제3자에게 제공하거나(제17조), 당사자의 동의를 얻어서 수집 당시의 목적 외의 목적으로 이용하거나 제3자에게 제공할 경우(제18조) 정보주체에게 고지할 사항을 규정하고 있다.

아울러 정보주체 이외로부터 수집한 개인정보의 수집 출처 등을 고지할 의무도 부과하고 있다. 즉, 개인정보처리자가 정보주체 이외로부터 수집한 개인정보를 처리하는 때에는 정보주체의 요구가 있으면 즉시 수집 출처, 처리 목적, 개인정보 처리의 정지를 요구할 권리가 있다는 사실을 정보주체에게 알려야 한다(제20조).

② 개인정보주체는 개인정보처리자가 처리하는 자신의 개인정보에 대한 열람을 해당 개인정보처리자에게 요구할 수 있는 개인정보열람청구권을 갖는다(제35조).

③ 개인정보주체는 개인정보처리자에게 그 개인정보의 정정 또는 삭제를 요구할 수 있는 개인정보의 정정, 삭제청구권(제36조)을 갖는다.

④ 개인정보주체는 개인정보처리자에 대하여 자신의 개인정보 처리의 정지를 요구할 수 있는 개인정보 처리정지 및 파기 등의 청구권(제37조)이 보장된다.

⑤ 개인정보주체는 손해배상청구나 법정손해배상을 청구할 권리(제39조, 제39조의2), 개인정보 분쟁조정위원회에 분쟁조정신청을 할 권리(제43조 등), 집단분쟁조정신청(제49조)이나 단체소송(제51조)을 제기할 수 있는 권리 등이 보장된다.

(3) 적법한 개인정보 연계·결합의 요건

가. 개요

개인정보보호법은 ‘개인정보처리자는 다음 각 호의 어느 하나에 해당하는 경우에는 개인정보를 수집할 수 있으며 그 수집 목적의 범위에서 이용할 수 있다’(제15조 제1항의 요건)고 하여 적법한 개인정보 수집의 요건과 ‘개인정보처리자는 다음 각 호의 어느 하나에 해당되는 경우에는 정보주체의 개인정보를 제3자에게 제공(공유를 포

함한다. 이하 같다)할 수 있다'(제17조 제1항)고 하여 적법한 제3자 제공의 요건과 '개인정보처리자는 다음 각 호의 어느 하나에 해당하는 경우에는 정보주체 또는 제3자의 이익을 부당하게 침해할 우려가 있을 때를 제외하고는 개인정보를 목적 외의 용도로 이용하거나 이를 제3자에게 제공할 수 있다'(제18조 제2항)고 하여 적법한 목적 외 이용 및 제3자 제공의 요건을 규정하고 있다.

나. 적법한 수집의 요건

개인정보인 데이터의 연계·결합, 혹은 연계나 결합의 결과 개인을 식별하거나 식별가능성이 있는 정보의 연계나 결합이 개인정보보호법 제15조 제1항의 '수집'으로 해석된다면 이 경우 개인정보보호법 제17조 제1항의 각호의 요건에 해당해야만 연계나 결합이 가능할 것이다. 개인정보보호법의 취지가 개인정보주체의 자유와 권리를 보호하고, 나아가 개인의 존엄과 가치를 구현함을 목적으로 한다는 점에 비추어 본다면, 제15조 제1항의 '수집'이란 개인정보가 새롭게 생성되는 모든 경우를 포괄하는 것으로 해석하는 것이 타당할 것이다. 그렇다면 연계나 결합은 아래의 요건에 해당하는 경우에만 가능하다.

첫째가 정보주체의 동의이다(제15조 제1호 사유). 정보처리자는 정보주체의 동의를 얻어서 데이터 연계·결합을 할 수 있다. 물론 이때에도 처리원칙은 준수해야 한다. 즉, 동의에도 불구하고, 개인정보 최소수집의 원칙 등 개인정보보호법 제3조의 원칙은 준수해야 한다. 이는 개인정보보호법 제16조에 분명히 표현되어 있다.[181] 법령에서는 수집이라고 표기되어 있지만, '데이터 연계'나 '결합'에 의해 개인정보를 생성하는 경우에도 수집과 마찬가지로 필요 최소한의 연계나 결합을 해야 한다. 그리고 연계나 결합에 대해서 동의를 받을 때는 필요 최소한의 정보 외의 개인정보 연계·결합에는 동의하지 않을 수 있다는 사실을 구체적으로 알려야 한다. 그리고 정보주체가 필요한 최소한의 정보 외의 개인정보 연계·결합에 동의하지 아니한다는 이유로 정보주체에게 재화 또는 서비스의 제공을 거부해서는 안 된다(개인정보보호법 제16조).

민간 분야에서 데이터의 연계나 결합은 다른 두 사이트의 계정을 통합하는 경우를 들 수 있다. 이때에도 과거의 모든 데이터를 연계나 결합하는 것을 허용해야만 해당 서비스를 이용할 수 있다고 한다면 이는 최소수집의 원칙을 위반하는 것으로 볼 수도 있다.

동의를 받을 때는 개인정보의 연계·결합의 목적, 연계·결합하려는 개인정보의 항목,

181) 개인정보처리자가 동의 등 개인정보보호법 제15조 제1항 각호 어느 하나의 사유로 개인정보를 수집하는 경우에는 그 목적에 필요한 최소한의 개인정보를 수집해야 한다.

연계·결합한 개인정보의 보유 및 이용 기간, 연계·결합에 대해 동의를 거부할 권리가 있다는 사실 및 동의 거부에 따른 불이익이 있는 경우에는 그 불이익의 내용을 알리고 동의를 받아야 한다(제15조 제2항).

개인정보보호법은 정보주체의 동의를 받는 방법도 규정하고 있다. 개인정보보호법은 각각의 동의 사항을 구분하여 정보주체가 이를 명확하게 인지할 수 있도록 알리고 각각 동의를 받도록 규정하고 있다(제22조 제1항). 이를 서면으로 받을 때에는 개인정보의 연계·결합의 목적, 개인정보의 항목 등 대통령령으로 정하는 중요한 내용을 행정안전부령으로 정하는 방법[182]에 따라 명확히 표시하여 알아보기 쉽게 하여야 한다.

이때에도 정보주체와의 계약 체결 등을 위하여 정보주체의 동의 없이 처리할 수 있는 개인정보와 정보주체의 동의가 필요한 개인정보를 구분하여야 한다. 이 경우 동의 없이 처리할 수 있는 개인정보라는 입증책임은 개인정보처리자가 부담한다(제22조 제3항). 특히 개인정보의 연계나 결합이 정보주체에게 재화나 서비스를 홍보하거나 판매를 권유하기 위하여 이루어지는 경우에는 정보주체가 이를 명확하게 인지할 수 있도록 알리고 동의를 받아야 한다(제22조 제4항). 아울러 개인정보처리자는 정보주체가 선택적으로 동의할 수 있는 사항을 동의하지 아니한다는 이유로 재화 또는 서비스의 제공을 거부하여서는 안 된다(제22조 제5항).

둘째, 법률에 특별한 규정이 있거나 법령상 의무를 준수하기 위하여 불가피한 경우(제15조 제1항 제2호 사유). 법률에 특별한 규정이 있는 경우로는 형사소송법, 민사소송법, 신용정보법, 보험업법, 국민건강보험법, 의료법, 약사법 등에서 개인정보를 수집할 수 있는 권한을 부여하고 있는 경우를 들 수 있다. 법령상 의무를 준수하기 위하여 불가피한 경우로는 법령에서 본인확인 의무를 부과하고 있거나, 연령확인 의무를 부과하고 있거나, 특정한 정보의 보유 의무를 규정하고 있는 경우를 들 수 있다.

개인정보의 연계나 결합을 법률에서 특별히 규정하고 있는 경우로는 건강보험 서비

182) 개인정보보호법 시행령은 동의를 받는 방법을 다음 어느 하나의 방법으로 규정하고 있다(제17조 제1항).
1. 동의 내용이 적힌 서면을 정보주체에게 직접 발급하거나 우편 또는 팩스 등의 방법으로 전달하고, 정보주체가 서명하거나 날인한 동의서를 받는 방법
2. 전화를 통하여 동의 내용을 정보주체에게 알리고 동의의 의사표시를 확인하는 방법
3. 전화를 통하여 동의 내용을 정보주체에게 알리고 정보주체에게 인터넷주소 등을 통하여 동의 사항을 확인하도록 한 후 다시 전화를 통하여 그 동의 사항에 대한 동의의 의사표시를 확인하는 방법
4. 인터넷 홈페이지 등에 동의 내용을 게재하고 정보주체가 동의 여부를 표시하도록 하는 방법
5. 동의 내용이 적힌 전자우편을 발송하여 정보주체로부터 동의의 의사표시가 적힌 전자우편을 받는 방법
6. 그 밖에 제1호부터 제5호까지의 규정에 따른 방법에 준하는 방법으로 동의 내용을 알리고 동의의 의사표시를 확인하는 방법

스를 활용하기 위해서 자격정보를 결합하는 경우가 해당한다. 법령상의 의무를 준수하기 위해 불가피한 경우도 마찬가지다.

셋째, 공공기관이 법령 등에서 정하는 소관 업무의 수행을 위하여 불가피한 경우. 공공기관이 법령의 소관 업무를 수행하기 위해 불가피하게 데이터의 연계나 결합을 해야 한다면 정보주체의 동의 없이 할 수 있다. '법령 등에서 정하는 소관 업무'란 '정부조직법' 및 각 기관별 직제·직제규칙, 개별 조직법 등에서 정하고 있는 소관 사무 이외에, '주민등록법', '국세기본법', '의료법', '국민건강보험법' 등 소관법령에 의해서 부여된 권한과 의무, 지방자치단체의 경우 조례에서 정하고 있는 업무 등을 의미한다. '불가피한 경우'란 개인정보를 수집하지 아니하고는 법령 등에서 해당 공공기관에 부여하고 있는 권한의 행사나 의무의 이행이 불가능하거나 다른 방법을 사용하여 소관 업무를 수행하는 것이 현저히 곤란한 경우를 의미한다고 한다. 예를 들어 인사혁신처가 '정부조직법' 제22조의3, '인사혁신처와 그 소속기관 직제' 및 '인사혁신처와 그 소속기관 직제 시행규칙'에 따라 공무원의 인사·윤리·복무·연금 등 관리를 위해 공무원 인사 관련파일을 수집·이용하거나 국가인재데이터베이스 시스템을 구축·운영하는 경우, 국민건강보험공단이 '국민건강보험법' 제14조에 따라 보험급여관리 등을 위하여 진료내역 등을 수집·이용하는 경우가 여기에 해당할 수 있다[183].

넷째, 정보주체와의 계약의 체결 및 이행을 위하여 불가피하게 필요한 경우. 이는 정보주체의 이익을 위한 것이기도 하고, 계약의 체결과 이행을 위해 불가피한 정보임에도 별도의 개인정보 수집에 대한 동의를 하도록 할 경우 거래비용을 지나치게 늘리게 되므로 정보주체의 동의가 없어도 처리가 가능하도록 한 것이다. 예를 들어 보험회사가 계약 체결을 위해 청약자의 자동차사고 이력, 다른 유사보험의 가입 여부 등에 관한 정보를 수집하는 경우, 고객이 주문한 상품을 배송하기 위하여 주소, 연락처 정보를 수집하는 경우가 이에 해당한다.[184]

다섯째, 정보주체 또는 그 법정대리인이 의사표시를 할 수 없는 상태에 있거나 주소불명 등으로 사전동의를 받을 수 없는 경우로서 명백히 정보주체 또는 제3자의 급박한 생명·신체·재산상의 이익을 위하여 필요하다고 인정되는 경우. 예를 들어 조난·홍수 등으로 실종되거나 고립된 사람을 구조하기 위하여 연락처, 주소, 위치정보 등 개인정보를 수집하는 경우, 아파트에 화재가 발생하였을 때 집안에 있는 자녀를 구하기 위해 해당 자녀 또는 부모의 이동전화번호를 수집하는 경우, 의식불명이나 중태에 빠진 환자의 수술 등 의료조치를 위하여 개인정보를 수집하는 경우, 고객이 전화사기(보이스피싱)에 걸린 것으로 보여 은행이 임시로 자금이체를 중단시키고 고객에게 사

183) 행정자치부(2016. 12), '개인정보보호 법령 및 지침·고시 해설', p77.
184) 앞의 자료, p78.

실 확인을 하고자 하는 경우 등이 이에 해당한다고 본다.[185]

여섯째, 개인정보처리자의 정당한 이익을 달성하기 위하여 필요한 경우로서 명백하게 정보주체의 권리보다 우선하는 경우.

다. 적법한 제3자 제공의 요건

데이터의 연계나 결합이 개인정보의 제3자 제공에 해당하는 경우가 있다. 개인정보 주체로부터 개인정보 연계나 결합을 하는 것이 아닌 경우이다. 예를 들어 병원에서 수집한 정보주체의 개인정보를 제3자에게 제공하여 그 정보주체에 대한 정보와 결합·연계하도록 하는 경우가 여기에 해당한다. 이 경우 개인정보의 제3자 제공이 이루어진다. 이와 같은 제3자 제공이 적법하기 위해서는 다음의 요건을 충족해야 한다.

첫째, 당사자의 동의. 정보주체가 동의하는 경우에는 데이터 연계·결합을 위한 개인정보 제3자 제공이 허용된다. 이때에는 개인정보를 제공받는 자, 개인정보를 제공받는 자의 개인정보 이용 목적, 제공하는 개인정보의 항목, 개인정보를 제공받는 자의 개인정보 보유 및 이용 기간, 동의를 거부할 권리가 있다는 사실 및 동의 거부에 따른 불이익이 있는 경우에는 그 불이익의 내용을 알리고 동의를 얻어야 한다.

둘째, 법률에 특별한 규정이 있거나 법령상 의무준수를 위해 불가피하게 수집한 경우로서 그 수집 목적 범위 내에서 개인정보를 제공하는 경우. 법률에 특별한 규정이 있는 경우의 예를 들면, 시·군·구의 장의 공직선거 입후보자에 대한 선거인명부 교부(‘공직선거법’ 제46조), 보험요율산출기관의 보험회사에 대한 보험계약자 교통법규위반 개인정보 제공(‘보험업법’ 제176조) 등의 경우 개인정보처리자는 정보주체의 동의 없이 개인정보를 관계 당사자에게 제공할 수 있다.[186] 법령상 의무준수를 위해 불가피한 경우는 학원을 설립·운영하려는 자의 강사명단 등을 교육감에게 등록하여야 하는 의무(‘학원의 설립·운영 및 과외교습에 관한 법률’ 제6조), 소득지급자의 소득귀속자에 대한 원천징수의무 및 원천징수이행상황신고의무(‘소득세법’ 제127조 및 제128조) 등이다.[187]

셋째, 공공기관이 법령 등에서 정하는 소관 업무 수행을 위해 불가피하게 수집한 경우로서 그 수집 목적 범위 내에서 개인정보를 제공하는 경우. 공공기관의 경우에는 개인정보를 수집할 수 있도록 명시적으로 허용하는 법률 규정이 없더라도 법령 등에서 소관 업무를 정하고 있고 그 소관 업무의 수행을 위하여 불가피하게 개인정보를

185) 앞의 자료, p80.
186) 앞의 자료, p92.
187) 앞의 자료, p93.

수집할 수밖에 없는 경우에는 정보주체의 동의 없이 개인정보의 수집이 허용된다.[188]

넷째, 급박한 생명·신체·재산상 이익을 위하여 필요한 경우. 정보주체 또는 그 법정대리인이 의사표시를 할 수 없는 상태에 있거나 주소불명 등으로 사전동의를 받을 수 없는 경우로서 명백히 제3자의 급박한 생명·신체·재산상의 이익을 위하여 필요하다고 인정되어 개인정보를 수집하였다면 그 수집 목적 범위에서 정보주체의 동의 없이 개인정보를 제3자에게 제공할 수 있다. 동사무소나 경찰관서가 시급히 수술 등의 의료조치가 필요한 교통사고 환자의 연락처를 의료기관에 알려주는 행위가 이에 속한다고 한다.[189]

라. 적법한 목적 외 이용, 제공의 요건

원칙적으로 개인정보처리자는 정보주체에게 이용·제공의 목적을 고지하고 동의를 받은 범위나 개인정보보호법 또는 다른 법령에 의하여 이용·제공이 허용된 범위를 벗어나서 개인정보를 이용하거나 제공해서는 안 된다(제18조 제1항).

예외적으로 개인정보처리자가 목적 외 이용을 하거나, 목적 외로 제3자에게 제공할 수 있으려면 아래의 요건을 충족하는 동시에 정보주체 또는 제3자의 이익을 부당하게 침해할 우려가 없어야 한다(제18조 제2항). 그리고 이 경우에도 개인정보처리자는 개인정보를 제공받는 자에게 이용 목적, 이용 방법, 그 밖에 필요한 사항에 대하여 제한을 하거나, 개인정보의 안전성 확보를 위하여 필요한 조치를 마련하도록 요청하여야 한다. 이 경우 요청을 받은 자는 개인정보의 안전성 확보를 위하여 필요한 조치를 하여야 한다(제18조 제5항).

첫째, 정보주체의 별도의 동의가 있는 경우. 별도의 동의이어야 하므로 개인정보를 수집 목적을 넘어 이용하거나, 제공하는 경우 다른 개인정보의 처리에 대한 동의와 분리해서 목적 외 이용·제공에 대한 동의를 따로 받아야 한다. 이때에도 개인정보를 제공받는 자, 개인정보의 이용 목적(제공 시에는 제공받는 자의 이용 목적을 말한다), 이용 또는 제공하는 개인정보의 항목, 개인정보의 보유 및 이용 기간(제공 시에는 제공받는 자의 보유 및 이용 기간을 말한다), 동의를 거부할 권리가 있다는 사실 및 동의 거부에 따른 불이익이 있는 경우에는 그 불이익의 내용을 알리고 동의를 받아야 한다.

둘째, 다른 법률에 특별한 규정이 있는 경우. 그 예로는 소득세법 제170조에 따른 세무공무원의 조사, 질문, 감사원법 제27조에 따른 감사원의 자료 요구, 국가유공자

188) 앞의 자료, p94.
189) 앞의 자료, p95.

등 예우 및 지원에 관한 법률 제77조에 따른 국가보훈처장의 자료제공 요구, 병역법 제81조 제2항에 따른 병무청장의 자료제공 요구, 부패방지 및 국민권익위원회 설치와 운영에 관한 법률 제42조 제1항 및 제3항에 따른 국민권익위원회의 자료제출 요청 등을 들 수 있다.[190]

셋째, 정보주체 또는 그 법정대리인이 의사표시를 할 수 없는 상태에 있거나 주소불명 등으로 사전동의를 받을 수 없는 경우로서 명백히 정보주체 또는 제3자의 급박한 생명·신체·재산의 이익을 위하여 필요하다고 인정되는 경우.

넷째, 통계작성 및 학술연구 등의 목적. 통계작성 및 학술연구 등의 목적을 위하여 필요한 경우로서 특정 개인을 알아볼 수 없는 형태로 개인정보를 제공하는 경우는 목적 외 이용이나, 제3자 제공이 허용된다(제18조 제2항 제4호).

이는 통계작성, 학술연구의 목적이나 이에 준하는 목적으로 제한된다. 논자에 따라서는 제공 목적이 통계작성 또는 학술연구를 위한 것이면 되고 그 결과물을 어디에 어떤 목적으로 이용할 것인지는 제한이 없다고 보기도 한다. 즉 그 결과물은 공공기관의 공공정책 수립 목적은 물론 민간기업의 시장분석, 경영전략 수립, 신상품 개발, 서비스 개선 등의 목적으로 이용하는 것도 가능하다고 본다.[191] 그러나 이는 부당한 해석이다. 여기서 목적 외 이용을 예외적으로 허용하면서 그 목적 범위를 통계작성이나 학술연구 등으로 제한하고 있는 취지에 비춰 본다면 통계작성, 학술연구 외의 목적으로 범위를 확장하는 것은 법률의 규정 취지를 벗어난다고 볼 것이다. 제공하는 목적뿐만 아니라 제공받아 사용하는 목적도 통계작성, 학술연구의 목적으로 한정하고 있다고 보는 것이 자연스럽다. 그리고 개인정보를 제공받은 자가 또 다른 제3자에게 개인정보를 제공하는 것은 새로운 개인정보의 처리로 보아야 할 것인데, 해당 조항에서는 활용 목적과 특정 개인을 알아볼 수 없는 형태로 개인정보를 제공한다는 개인정보의 처리 방법 두 가지 요건 모두를 충족해야만 한다. 따라서 이를 다른 목적으로 제공하는 것은 허용되지 않는다고 볼 것이다. 여기서 특정 개인을 알아볼 수 없는 형태로 가공했다는 것은 재식별 가능성이 없거나, 재식별이 합리적으로 기대하기 어려운 경우를 말한다고 볼 것이다.

논자에 따라서는 제공하는 사람은 특정 개인을 알아볼 수 있어도 제공받는 사람이 합리적인 노력을 기울여도 특정 개인을 알아볼 수 없도록 가공되었다면 “특정 개인을 알아볼 수 없는 형태로 개인정보”를 제공하는 것에 해당한다고 보는 입장도 있는데,[192] 식별 가능성이란 제공을 받는 사람을 기준으로만 판단해서는 안 되고, 제3자가

190) 앞의 자료, p104.
191) 앞의 자료, p104.
192) 앞의 자료, p104.

식별이 가능한 경우에도 식별 가능성이 있다고 보아야 할 것이다. 따라서 제공하는 사람, 제공받는 사람, 합리적 범위의 제3자를 포함하여 재식별 가능성을 평가해야 할 것이다.[193]

다섯째, 개인정보를 목적 외의 용도로 이용하거나 이를 제3자에게 제공하지 아니하면 다른 법률에서 정하는 소관 업무를 수행할 수 없는 경우로서 보호위원회의 심의·의결을 거친 경우.

여섯째, 조약, 그 밖의 국제협정의 이행을 위하여 외국정부 또는 국제기구에 제공하기 위하여 필요한 경우.

일곱째, 범죄의 수사와 공소의 제기 및 유지를 위하여 필요한 경우.

개인정보처리자는 보유한 개인정보를 수집·이용 목적 이외의 용도로 제공하기 위해서는 다른 법률의 특별한 규정이 있거나 정보주체의 동의를 받아야 한다. 즉 공공기관 외의 개인정보처리자에 대해서는 비록 범죄수사 목적이라 하더라도 원칙적으로 형사소송법 등의 규정에 따라서만 개인정보 제공을 요구할 수 있다. 그러나 공공기관의 경우 수사기관이 범죄수사, 공소제기 및 유지를 위해서 필요하다고 요청하는 경우 해당 개인정보를 정보주체의 별도의 동의 없이 제공할 수 있다. 이는 범죄수사 편의를 위해 공공기관이 보유한 개인정보에 대해서는 정보주체의 동의 없이 목적 외로 이용 또는 제공할 수 있게 하기 위한 것이다.[194]

여덟째, 법원의 재판업무 수행을 위하여 필요한 경우. 법원은 형사소송법이나 민사소송법에 의하여 공공기관이 보유한 개인정보의 제공을 요청할 수 있다. 이와 같이 법률의 규정에 의한 경우 외에도 개인정보보호법은 법원이 재판업무 수행을 위해 필요한 경우 개인정보의 목적 외 이용을 허용하고 있다.

아홉째, 형 및 감호, 보호처분의 집행을 위하여 필요한 경우.

한편 이와 같이 공공기관이 개인정보를 목적 외로 이용하거나 제3자에게 제공하는 경우에는 1개월 이내에 목적 외 이용·제공의 법적 근거, 이용 또는 제공 일자·목적·항목에 관하여 관보 또는 인터넷 홈페이지에 게재하여 공고하여야 한다. 다만, 정보주체의 동의를 받거나 범죄수사와 공소제기 및 유지를 위해 개인정보를 목적 외로 이용하

193) 개인정보에서 식별성을 제거하여 더 이상 재식별이 불가능해지는 경우에는 개인정보보호법의 적용이 배제되는데, 흔히 이를 '익명화', 그 결과물을 '익명화된 정보'로 부른다. 반면 식별성을 제거하는 것을 '식별성 제거'(de-identification)라고 부르는데, 여기에는 익명화 또는 익명화된 정보도 포함된다. 그러나 국내에서는 '비식별 조치'라는 용어를 사용하면서 이를 '익명화' 혹은 '익명화된 정보'로 사용하고 있어 혼란이 야기되고 있다. 이 글에서 사용된 '비식별'이라는 개념은 영어의 de-identification을 의미하는 것으로 가명화(Pseudonymization)와 익명화(anonymization)를 포함하는 것으로 본다.

194) 앞의 자료, p105.

거나 제공하는 경우에는 그러하지 아니하다(제18조 제4항).

마. 민감정보인 경우

개인정보보호법은 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(대통령령은 유전자검사 등의 결과로 얻어진 유전정보와 '형의 실효 등에 관한 법률' 제2조 제5호에 따른 범죄경력자료에 해당하는 정보를 추가하고 있다.[195])를 민감정보로 규정하고(제23조 제1항), 그에 대해서는 좀 더 특별한 규율을 하고 있다.

건강에 대한 정보는 민감정보로 분류되는데, 진료기록, 의약품 처방 기록 등이 포함된다. 혈압, 체온, 맥박 등의 정보도 개인의 건강상태를 알 수 있게 해 주는 정보이기 때문에 민감정보에 해당한다. 민감정보의 처리는 원칙적으로 금지되며, 정보주체에게 고지할 사항을 알리고 다른 개인정보의 처리에 대한 동의와 별도로 동의를 받은 경우이거나, 법령에서 민감정보의 처리를 요구하거나 허용하는 경우에만 처리가 허용된다(제23조 제1항 제1호, 제2호). 따라서 계약의 체결이나 이행을 위해서 필요한 경우나, 개인정보처리자의 정당한 이익을 보호하기 위하여 필요한 경우라는 이유로는 당사자의 동의 없는 민감정보의 처리가 허용되지 않는다.

한편 개인정보처리자가 민감정보를 처리하는 경우에는 그 민감정보가 분실·도난·유출·위조·변조 또는 훼손되지 아니하도록 개인정보보호법 제29조에 따른 안전성 확보에 필요한 조치를 하여야 한다(제23조 제2항). 데이터 결합이나 연계 시 건강정보와 같은 민감정보를 처리하는 것은 개인정보주체로부터 별도의 동의를 받거나, 법령에 민감정보의 처리를 요구하거나 허용하는 규정이 있어야만 가능하다.

3. 전자정부법

(1) 효율을 우선시하는 전자정부의 원칙

개인정보가 포함되어 있거나 연계를 통해 개인정보가 생성 또는 개인이 식별될 수

195) 다만 유전자검사 등의 결과로 얻어진 유전정보와 범죄경력자료에 해당하는 정보는 공공기관이 ① 개인정보를 목적 외의 용도로 이용하거나 이를 제3자에게 제공하지 아니하면 다른 법률에서 정하는 소관 업무를 수행할 수 없는 경우로서 보호위원회의 심의·의결을 거친 경우, ② 조약, 그 밖의 국제협정의 이행을 위하여 외국정부 또는 국제기구에 제공하기 위하여 필요한 경우, ③ 범죄의 수사와 공소의 제기 및 유지를 위하여 필요한 경우, ④ 법원의 재판업무 수행을 위하여 필요한 경우, ⑤ 형(刑) 및 감호, 보호처분의 집행을 위하여 필요한 경우에는 민감정보로 보지 않는다(개인정보보호법 시행령 제18조).

있는 행정정보의 연계는 개인정보보호법상 개인정보의 '처리'에 해당하므로 개인정보보호법과 개별 근거 법률에 의해 규율되게 된다. 그런데 현실에 있어서는 전자정부법이 행정정보 연계의 중요한 근거 법률로 작용하고 있다.

전자정부법은 전자정부[196]를 효율적으로 구현하고, 행정의 생산성, 투명성 및 민주성을 높여 국민의 삶의 질을 향상시키는 것을 목적으로 제정되었다(제1조). 전자정부법은 처음에는 행정기관에만 국한되었으나, 범위를 넓혀서 현재는 공기업, 공사, 각급 학교, 특별법에 따라 설립된 특수법인[197], 연구기관[198]도 포함하고 있어서 단순히 행정기관에 국한되는 법률로 보기도 어렵다.

그런데 전자정부법은 '전자정부의 원칙'(행정기관등이 전자정부의 구현·운영 및 발전을 추진할 때 우선적으로 고려하고 이에 필요한 대책을 마련하여야 할 원칙)이라는 것을 제시하고 있는데(제4조), 여기에는 주로 개인정보보호법의 원칙과 상반되는 원칙들이 포함되어 있다. 즉, 전자정부법이 제시하는 전자정부의 원칙은 ① 대민서비스의 전자화 및 국민편익의 증진, ② 행정업무의 혁신 및 생산성·효율성의 향상, ③ 정보시스템의 안전성·신뢰성의 확보, ④ 행정정보의 공개 및 공동이용의 확대, ⑤ 중복투자의 방지 및 상호운용성의 증진, ⑥ 개인정보 및 사생활의 보호이다. 마지막 ⑥의 개인정보 보호를 제외한 나머지 5가지 원칙은 사실상 개인정보보호 원칙과 상반되는 원칙들이다(제4조).

이처럼 전자정부법이 개인정보의 보호와 상반되는 가치인 행정의 생산성, 투명성, 민주성, 공동이용의 확대, 중복투자의 방지를 목적으로 하고 있기 때문에 그와 같은 가치를 구체화한 전자정부법의 행정정보 공동이용, 통합시스템 구축 등에 대한 규정과 개인정보보호법의 규정의 관계를 어떻게 해석하고, 어디에 우선순위를 둘 것인가 등이 문제가 된다.

(2) 전자정부법과 개인정보보호법의 관계, 해석의 우선순위

가. 두 법률의 비교

전자정부법은 전자정부의 원칙을 규정하고 있고(전자정부법 제4조), 국가정보화전략위원회와 중장기 전자정부기본계획에 대한 규정을 두고 있는데(제5조), 개인정보보호

196) 전자정부는 '정보기술을 활용하여 행정기관 및 공공기관의 업무를 전자화하여 이들 상호 간의 행정업무 및 국민에 대한 행정업무를 효율적으로 수행하는 정부'를 의미한다(전자정부법 제2조 제1호).

197) 한국은행(한국은행법), 국립암센터(국립암센터법), 한국해운조합(한국해운조합법), 한국방송공사(방송법 제43조), 예금보험공사(예금자보호법 제3조) 등이 그 예이다. 행정안전부(2010. 8), '전자정부법의 이해와 해설', p39.

198) 한국전자통신연구원, 한국화학연구원, 한국행정연구원, 한국조세연구원 등. 앞의 자료, p40.

법은 개인정보보호 원칙을 규정하고 있고(개인정보보호법 제4조), 개인정보보호위원회와 개인정보 보호 기본계획과 시행계획을 규정하고 있다(제9조, 제10조).

중장기 전자정부기본계획에는 행정정보 공동이용의 확대 및 안전성 확보, 정보기술 아키텍처의 도입 및 활용, 정보자원의 효율적 관리, 전자정부 표준화 및 공유서비스의 확대 등에 대한 계획 등이 포함되어 있다. 그래서 양자가 충돌될 경우, 양자의 관계가 문제 된다.

표 4-2 개인정보보호법과 전자정부법의 비교

	개인정보보호법	전자정부법
목적	개인정보의 처리 및 보호에 관한 사항을 정함으로써 개인의 자유와 권리를 보호하고, 나아가 개인의 존엄과 가치를 구현함을 목적으로 함.	행정업무의 전자적 처리를 위한 기본원칙, 절차 및 추진방법 등을 규정함으로써 전자정부를 효율적으로 구현하고, 행정의 생산성, 투명성 및 민주성을 높여 국민의 삶의 질을 향상시키는 것을 목적으로 함
원칙	개인정보보호 원칙	전자정부 기본원칙
수립 계획	개인정보 보호 기본계획, 시행계획	중장기 전자정부기본계획
수립 주체	개인정보보호위원회	국가정보화전략위원회
주요 규정	개인정보보호 원칙, 정보주체의 권리, 개인정보의 처리 개인정보의 안전한 관리 정보주체의 권리 보장 개인정보분쟁조정위원회 개인정보 단체소송	행정정보의 공동이용에 관한 규정 정보자원의 효율적 관리에 관한 규정 전자정부 표준화 및 공유서비스의 확대 등에 대한 규정

나. 양자의 모순, 긴장 관계

개인정보보호법은 개인정보 자기결정권을 바탕으로 한 개인의 권리를 구체화한 법률인 반면, 전자정부법은 행정업무의 효율화를 위한 법률이다.

전자정부법의 행정정보 공동이용이나 행정정보시스템의 연계·통합에 대한 규정은 공동이용의 대상이 되는 정보처리의 내용에 따른 처리를 규정하고 있는 '실체법'이라기보다는 행정정보의 처리에 대한 '절차법'에 속한다고 볼 수 있다.

이런 점을 고려한다면, 행정의 효율화라는 이익, 그에 따른 전자정부법의 행정정보 처리의 효율화를 위한 여러 규정은 국민의 헌법상 기본권인 개인정보보호원칙을 훼손하지 않는 범위에서만, 개인정보 보호원칙과 조화를 이루는 범위로만 한정적으로 규정되고, 그 원칙을 준수하는 범위에서 인정되어야 한다. 실제로 각국은 행정의 효율화를 위한 입법을 마련하면서도 개인정보보호법에서 규율한 권리들이 이로 인해서 훼손될 수 없음을 밝히기도 한다.

그런데 전자정부법에는 '행정기관등의 대민서비스 및 행정관리의 전자화, 행정정보

의 공동이용 등 전자정부의 구현·운영 및 발전에 관하여 다른 법률에 특별한 규정이 있는 경우를 제외하고는 이 법에서 정하는 바에 따른다'(제6조)는 규정을 두어 행정정부의 공동이용에 관한 규정이 개인정보보호법의 규정에 우선하는 것으로 해석될 소지가 있다. 그러나 이 규정은 위에서 본 바와 같은 이유로 엄격하게 제한적으로 해석되어야 한다.

나아가 국가정보화전략위원회의 심의를 거쳐 관장기관의 장(행정안전부, 국회사무처, 법원행정처 등)이 수립하도록 하고 있는 중장기 전자정부기본계획의 경우[199], 그 중 개인정보의 보호와 관련된 내용은 개인정보보호위원회의 심의를 요하거나, 개인정보보호 기본계획과 부합하도록 할 필요가 있다. 그러나 현재의 법제에서는 이 점이 보장되어 있지 않다. 그리고 예를 들어 행정정보의 통합, 공동이용에 대한 내용은 개인정보보호법을 훼손하지 않도록 마련되도록 해야 하고 그에 대한 통제를 하는 것이 바람직하다.[200] 전자정부법에 따라 제정하는 전자정부기본계획에 포함되는 개인정보의 활용에 관한 내용은 개인정보보호기본계획과 부합해야 한다.

다. 전자정부 서비스와 관련한 규정들의 해석

① 전자정부 서비스를 위한 행정정보의 연계

전자정부법은 행정기관등이 전자정부를 통하여 다른 행정기관등 및 국민, 기업 등에 제공하는 행정서비스를 '전자정부 서비스'라고 통칭하면서(제2조 제5호), 그 서비스를 제공하기 위하여 여러 가지 행정정보나 정보시스템을 연계할 수 있다는 규정을 두고 있다. 전자정부법이 전자정부 서비스라는 개념을 매개로 행정정보 시스템 연계의 근거 규정이 되는 것이다.

그런데 전자정부서비스의 제공 및 이용촉진과 관련하여 전자정부법은 이용의 촉진, 유비쿼터스 서비스의 제공 등에 대해서는 규정하면서도 프라이버시의 보호, 개인정보 자기결정권의 보호 및 보장 등에 대해서는 구체적인 언급이 없는 등 균형 있는 법률로 보기 어려운 점이 있음을 부인할 수 없다.[201]

---

199) 2015년에 5년 단위의 중장기 전자정부기본계획('16~'20)을 전자정부 2020 기본계획이라는 이름으로 수립했다.

200) 예를 들어 2020 기본계획에는 국민주도의 융합형 개인 맞춤형 서비스 확대, 데이터 기반의 미래 디지털 행정체계 혁신, 신뢰 기반의 미래 디지털 인프라 확충 등의 내용이 들어 있다. 구체적 내용으로는 정부·공공기관 및 민간에서 보유한 데이터·정보의 집적 및 분석하여 서비스 및 정책에 활용하는 데이터 기반 선제적·과학적 행정체계 구축(빅데이터 공통기반(HW/SW, 플랫폼 등) 확충, 빅데이터 기반 과학적 의사결정체계 구축, 국토·복지·고용 등 분야별 빅데이터/데이터 기반 마련 및 서비스 확대, 지자체 빅데이터 공통기반 마련 및 활용서비스 구축 등) 등의 내용이 들어 있다.

201) 예를 들어 전자정부서비스와 관련해서는 아래와 같이 제공과 이용촉진에 대한 규정이 주를 이루고 있다.
제16조(전자정부서비스 개발·제공)

② 통합전자민원창구와 정보시스템 연계

예를 들어 통합전자민원창구에 관한 규정(제9조)과 통합전자민원창구를 통한 생활정보의 제공에 대한 규정(제9조의2)은 통합전자민원창구와 행정기관, 소속기관, 지자체, 공공기관의 정보시스템을 연계할 수 있도록 하는 근거 규정이 되고 있다. 즉, 중앙사무관장기관의 장은 행정기관등의 전자민원창구의 설치·운영을 지원하고 이를 연계하여 통합전자민원창구를 설치·운영할 수 있다고 규정하고, 행정기관등의 장은 민원인이 해당 기관을 직접 방문하지 아니하고도 민원사항 등을 처리할 수 있도록 관계 법령의 개선, 필요한 시설 및 시스템의 구축 등 제반 여건을 마련하여야 한다는 규정도 두고 있다.

나아가 행정안전부장관은 민원인에게 중앙행정기관과 그 소속기관, 지방자치단체 및 공공기관이 보유한 본인의 건강검진일, 예방접종일, 운전면허갱신일 등 생활정보를 열람할 수 있는 서비스를 제공할 수 있다고 하고, 이 경우 행정안전부장관은 다른 중앙행정기관등의 장과 협의하여 제9조 제3항에 따른 통합전자민원창구와 다른 중앙행정기관등의 정보시스템을 연계할 수 있다고도 규정하고 있다. 생활정보 열람서비스의 종류는 행정안전부장관이 결정, 고시하는데[202], 해당 정보의 제공을 위해 정보시스템이 연계되는 것이다. 그런데 해당 정보들은 개인의 민감한 정보들로서 정보시스템의 연계로 인한 여러 가지 영향들이 평가되어야 할 것들이다. 그런데 정보시스템의 연계

---

① 행정기관등의 장은 국민의 복지향상 및 편익증진, 국민생활의 안전보장, 창업 및 공장설립 등 기업활동의 촉진 등을 위한 전자정부서비스를 개발하여 제공하고 이를 지속적으로 보완·발전시키기 위한 대책을 마련하여야 한다.

② 행정기관등의 장은 전자정부서비스 이용자가 손쉽게 전자정부서비스에 접근하여 안전하고 편리하게 활용할 수 있도록 하여야 하며, 제공되는 전자정부서비스는 최신의 것이 되도록 하여야 한다.

③ 행정기관등의 장은 전자정부서비스를 개발·제공할 때 전자정부서비스 이용자의 요구사항 및 편익을 고려하여야 한다.

202) 현재는 다음과 같다.

1. 일반건강검진, 생애전환기 건강진단, 암검진, 영유아 건강검진 등 건강검진 정보
2. 예방접종명, 예방접종일 등 국가필수 예방접종 정보
3. 국민·공무원·사립학교교직원 연금예상액 등 연금 정보
4. 국세청 종합소득세 신고안내 정보
5. 재산세, 주민세, 등록면허세, 자동차세 등 정기분에 대한 납부안내 정보
6. 징병검사일, 입영일, 동원훈련일, 민방위훈련일 등 병역(훈련) 정보
7. 경찰청 교통위반 범칙금·과태료 등 납부안내 정보
8. 운전면허갱신기간, 정지기간 등 운전면허 정보
9. 국세, 지방세, 건강보험료, 국민연금보험료 미환급금 정보
10. 휴면예금관리재단에 출연된 휴면예금, 휴면보험금 안내 정보 <신설 2016.1.8.>
11. 자동차 검사기간, 자동차 압류 등 자동차 정보 <신설 2016.1.8.>
12. 고속도로미납통행료, 주정차위반 과태료 납부안내 정보 <신설 2016.1.8.>
13. 건축물에너지사용량등급, 근로·자녀장려금 대상자, 공공임대 예비입주자순위, 여권만료일 안내 등 생활지원 정보 <신설 2016.1.8.>
14. 일반상환·든든·농어촌대학생·공무원연금 대여학자금 등 학자금 대출내역 정보 <신설 2016.1.8.>
15. 보금자리론, 디딤돌대출, 아낌e대출, 노후긴급자금(실버론), 주택연금잔액 등 생활금융 정보

와 관련하여 필요성, 연계정보의 범위, 안전조치 등에 대한 법적 규율은 매우 미비한 실정이다.

그림 4-1 생활정보서비스 홈페이지

* 출처: 민원24 https://www.minwon.go.kr/main?a=AA210LifeSvcInfoApp

라. 행정정보 공동이용 우선의 원칙과 행정정보 연계

전자정부법은 행정정보의 공동이용에 관한 장(제4장)을 두고 있으며, '행정기관등의 장은 수집·보유하고 있는 행정정보를 필요로 하는 다른 행정기관등과 공동으로 이용하여야 하며, 다른 행정기관등으로부터 신뢰할 수 있는 행정정보를 제공받을 수 있는 경우에는 같은 내용의 정보를 따로 수집하여서는 아니 된다'(제36조)는 규정을 두고 있다.

그런데 이와 같은 중복 수집의 금지는 행정의 효율성이라는 측면에서는 정당화될 수 있지만 효율성만 내세워서 일의적으로 판단할 문제는 아니다. '다른 행정기관 등으로부터 신뢰할 수 있는 행정정보를 제공받을 수 있는 경우'의 의미가 문제이다. 실제로 수집하려고 하는 행정정보는 개인정보주체가 상황에 따라서는 답변을 거부할 수 있는 것도 있다. 수집대상인 정보의 내용이나 맥락에 따라서는 공동이용이 부적절한 경우도 있을 수 있다. 이 원칙은 자칫 개인정보주체의 권리를 침해할 우려도 있는 원칙이므로 정보주체의 권리가 침해되지 않도록 조화롭게 해석할 필요가 있다.

한편 행정정보의 공동이용 대상에는 사기업인 은행업, 신용카드사나 신용정보업자나 그 협회 등도 포함되어 있는데, 이는 은행, 신용카드사, 신용정보업자에 대한 특혜로 비쳐질 수 있다.

마. 행정정보 공동이용센터의 구축·이용

전자정부법은 행정안전부장관으로 하여금 행정정보의 원활한 공동이용을 위하여 행정안전부장관 소속으로 행정정보 공동이용센터를 두고 공동이용에 필요한 시책을 추진하도록 하고 있다(제37조 제1항). 법은 행정정보를 공동으로 이용하는 기관은 정당한 사유가 없으면 공동이용센터를 통하여 행정정보를 공동이용하여야 한다(제37조 제2항)는 규정까지 두고 있다.

행정정보 공동이용센터를 통하여 공동으로 이용할 수 있는 행정정보는 민원사항 등의 처리를 위하여 필요한 행정정보, 통계정보, 문헌정보, 정책정보 등 행정업무의 수행에 참고가 되는 행정정보, 행정기관 등이 법령 등에서 정하는 소관 업무의 수행을 위하여 불가피하게 필요하다고 인정하는 행정정보이다(제38조).

그 중 '민원 처리에 관한 법률' 제36조 제1항에 따른 민원처리기준표에 올라있는 민원의 처리를 위하여 행정정보 공동이용이 필요한 경우는 개인정보보호위원회의 심의, 의결을 생략하고 공동이용의 승인을 할 수 있다고 하고 있는데(제39조 제5항 제2호), 개념이 포괄적인 민원의 처리를 위해 공동이용을 일률적으로 허용하는 것보다는 그에 대한 안전장치의 마련 등 규제와 감독이 필요하다. 행정안전부장관이 고시한 공동이

용 대상 행정정보는 아래와 같다.

그림 4-2 행정안전부장관이 고시한 공동이용 대상 행정정보

정보보유기관	공동이용 대상 행정정보
과학기술정보통신부(4)	소프트웨어사업자신고확인서, 정보통신공사업등록증, 정보통신기술자경력수첩, 정보통신감리원자격증
교육부(2)	검정고시 합격증명서, 고등학교졸업증명서
외교부(2)	여권, 해외이주신고확인서
법무부(4)	출입국에관한사실증명, 외국인등록사실증명, 국내거소신고사실증명, 외국인의부동산등기등록증명서
행정안전부(10)	주민등록표 등·초본, 지방세납세증명서, 지방세세목별과세(납세)증명서(자동차세),지방세세목별과세(납세)증명서(재산세), 상훈수여증명서, 인감증명서, 국외이주신고증명서, 지방세납부확인서(등록면허세면허분) 주민등록전입세대, 본인서명사실확인서
농림축산식품부(2)	축산업등록증, 농업경영체증명서
산업통상자원부(8)	공장등록증명서, 석유판매업등록증, 전기안전점검확인서, 전기공사업등록증, 전기공사기술자경력수첩, 전기공사업등록관리대장, 공장(신설,증설,이전,업종변경,제조시설)승인(변경승인)서, 산업단지입주계약(계약변경)신청(확인)서
보건복지부(13)	국민기초생활수급자증명서, 장애인증명서, 약사면허증, 영양사면허증, 의료기사면허증(안경사/방사선사), 의료면허증(의사/치과의사/한의사/간호사), 전문의자격증(의사/치과의사/한의사), 요양보호사자격증, 의료기관개설신고증명서, 어린이집인가증, 장애인연금(경증)장애수당장애아동수당수급자확인서, 의료기관개설허가증, 건강진단결과서
환경부(6)	사업장폐기물배출자신고증명서, 폐수배출시설설치허가증(신고증명서), 폐기물수집운반업허가증, 폐기물(중간/최종/종합)처리업허가증, 폐기물처리시설설치승인서, 폐기물처리시설설치신고증명서
고용노동부(1)	국가기술자격취득사항확인서
여성가족부(1)	한부모가족증명서

정보보유기관	공동이용 대상 행정정보
국토교통부(33)	개별공시지가확인서, 건축물대장, 건축물사용승인서, 건설기계등록원부, 주택건설사업사용검사필증, 자동차등록원부, 이륜자동차사용신고필증, 건축·대수선·용도변경허가서, 토지(임야)대장, 지적도, 임야도, 건설업등록증, 토지이용계획확인서, 건설기계등록증, 건설기계검사증, 건설기계사업등록증, 건축사업무신고필증, 자동차등록증, 부동산등기용등록번호증명서, 토지거래계약허가증, 자동차말소등록사실증명서, 임대사업자등록증, 임시운행허가증, 개별주택가격확인서, 공동주택가격확인서, 착공신고필증, 건축·대수선·용도변경신고필증, 가설건축물관리대장, 위반건축물관리대장, 부동산종합증명서(토지), 부동산종합증명서(토지, 건축물), 부동산종합증명서(토지, 집합건물), 주거급여수급자정보
해양수산부(8)	선박원부, 선박검사증서, 선박국적증서(상선), 선박국적증서(어선), 어선등록필증, 어업면허증, 선적증서, 해상화물운송사업등록증
국가보훈처(4)	국가유공자(유족)확인원, 취업지원대상자증명서, 대학수업료등면제대상자증명서, 교육지원대상자증명서
국세청(8)	(국세)납세증명서, 소득금액증명, 납세사실증명, 휴업사실증명, 사업자등록증명, 폐업사실증명, 부가가치세과세표준증명원, 표준재무제표증명(법인·개인)
관세청(3)	수출신고필증, 수입신고필증, 관세납세증명서
병무청(2)	병적증명서, 징병신체검사결과통보서(시력)
경찰청(2)	자동차운전면허증, 운전경력증명서
소방청(4)	안전시설등완비증명서, 화재증명원, 소방시설업등록증, 소방시설완공검사증명서
해양경찰청(4)	선박출항·입항신고사실확인서(개별), 선박출항·입항신고사실확인서(총괄), 선원승선신고사실확인서, 폐기물위탁·처리신고증명서
특허청(4)	특허등록원부, 실용신안등록원부, 디자인등록원부, 상표등록원부
중소벤처기업부(3)	벤처기업확인서, 메인비즈확인서, 이노비즈확인서
대법원(4)	법인등기사항증명서, 건물등기사항증명서, 토지등기사항증명서, 가족관계등록전산정보

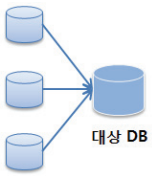
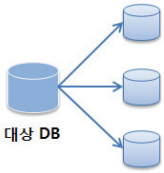
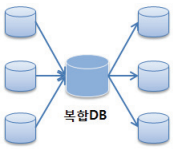
정보보유기관	공동이용 대상 행정정보
국민건강보험공단(8)	건강보험자격확인서, 건강보험자격득실확인서, 건강·장기요양보험료납부확인서(개인), 사업장건강·장기요양보험료납부확인서, 차상위본인부담경감대상자증명서, 건강검진결과통보서 , **4대사회보험료 완납증명서,** **건강 · 연금보험료완납(납부)증명서**
공무원연금공단(1)	공무원연금내역서
국민연금공단(3)	연금산정용가입내역확인서, 사업장국민연금보험료월별납부증명, 국민연금가입자가입증명
사립학교교직원연금공단(1)	연금법적용대상교직원확인서
근로복지공단(3)	고용보험료완납증명원, 산재보험료완납증명원, 산재보험급여지급확인원
한국토지주택공사(1)	공공임대주택정보
한국가스안전공사(1)	액화석유가스사용시설완성검사증명서(발급 확인서)
국가평생교육진흥원(1)	학점은행제학위증명(전문학사, 학사)
한국사회적기업진흥원(1)	사회적기업인증서
대한상공회의소(1)	국가기술자격증

바. 전자적 시스템의 상호 연계·통합

전자정부법은 각 중앙행정기관이 소관 전자적 시스템을 행정 효율성 제고와 대민서비스의 통합적, 효율적 제공을 위해 다른 중앙행정기관등의 전자적 시스템과 상호 연계하거나 통합할 수 있다는 규정을 두고(제30조의 2), 이때는 행정정보 간 상호 연관성 및 연계·통합으로 인한 기대효과를 고려하여야 한다고만 규정하고 있다.

공공기관이 데이터베이스를 구축하는 유형은 독립적으로 운영되는 자체생성형 시스템, 다른 데이터베이스로부터 연계를 통해 정보를 수집, 활용하는 수집형 데이터베이스, 연계를 통해 다른 데이터베이스에 정보를 제공하는 제공형 데이터베이스, 연계를 통해서 수집하기도 하고, 제공하기도 하는 복합형 데이터베이스로 나누어볼 수 있다.

표 4-3 공공기관 데이터베이스 구축 유형

유형 구분	유형 정의	중점 품질관리 대상
자체생성형  	○ 다른 DB와 연계 없이 자체적으로 정보를 생성 활용하는 DB ○ 독립적으로 시스템이 운영되며, 자체 DB에서 정보의 생성부터 소멸까지가 모두 이루어짐 예) 국가생물자원통합DB, 제주의 민속문화 DB 등	○ 데이터 값의 정확성, 유효성, 적시성, 유용성 관리가 중요
수집형  	○ 다른 DB로부터 연계를 통해 정보를 수집・활용하는 DB ○ 외부 DB로부터 정보를 수집・활용하여 다양한 집계, 통계 등의 목적 있는 DB를 생성 예) 국가통계통합DB 등	○ 수집 데이터 구조의 일관성, 값의 유효성, 보안성, 적시성 관리가 중요
제공형  	○ 생성・수집을 통해 보유하고 있는 정보에 대하여 연계를 통해 다른 DB에 제공하는 DB ○ 정보를 생성・관리하는 DB로서 정보를 외부 DB에 제공 예) 주민등록DB, 병적 DB, 토지대장 DB 등	○ 제공 데이터 구조의 일관성, 값의 정확성, 적시성, 보안성 관리가 중요
복합형  	○ 정보 생성, 수집, 제공이 다른 DB와 상호 연계를 통해 복합적으로 구성되는 DB 예) 연계・통합이 늘어나는 많은 공공기관의 DB가 해당 (국민연금DB, 건강진료정보심사 DB 등)	○ 타 유형의 중점 품질관리 대상 모두 적용

그런데 데이터 시스템의 연계는 새로운 개인정보의 생성에 준하여, 원칙적으로 정보주체의 동의가 필요한 것이다. 동의가 없어도 되는 법적 근거가 있는지를 검토하고, 개인정보의 영향평가를 하여 연계로 인한 효과를 면밀하게 평가하여, 연계할 것인지 여부, 연계대상 정보 내용과 범위, 연계방식, 안전조치, 비용부담 등에 관한 구체적 사항을 정하는 것이 바람직하다.

특히 중앙행정기관이 운영하고 있는 전자적 시스템을 다른 행정기관과 상호 연계·통합하는 것은 단순히 다른 행정기관으로부터 개인정보를 제공받는 것과는 근본적으로 다른 문제이다.

일반적으로 개별적인 정보 제공을 받는 것에 비해서 전자적 시스템의 연계·통합은 개인정보주체에게 미치는 영향이 크기 때문이다. 연계·통합은 그 목적, 연계대상 정보 내용과 범위, 연계방식 등에 있어서 천차만별이며 그 효과도 하나로 볼 수 없다.

사. 데이터활용공통기반시스템 구축·활용

전자정부법은 행정안전부장관이 '데이터활용공통기반시스템'을 구축·활용하도록 하고 있는데(제30조의 3), 이는 전자적 시스템을 통하여 수집·관리되는 데이터를 공동으로 활용하기 위한 시스템이다. 이를 통해서 각 중앙행정기관등의 장은 다른 중앙행정기관등의 장이 수집·관리하는 데이터를 공동으로 활용할 수 있다.

한편 시행령은 데이터활용공통기반시스템을 통하여 활용 가능한 데이터는 '공공기관의 정보공개에 관한 법률' 제9조에 따른 비공개대상정보가 포함되지 아니한 데이터로 한다고 제한하고 있다(제35조의 3 제1항). 그리고 데이터활용공통기반시스템을 통하여 수집·활용 가능한 공개된 인터넷 데이터는 법인·단체 또는 개인의 정당한 이익을 현저히 침해할 우려가 있다고 인정되는 데이터에 해당하지 아니한 데이터로 한다. 수집·활용하려는 데이터에 '저작권법' 및 그 밖의 다른 법령에서 보호하고 있는 제3자의 권리가 포함되어 있는 경우 해당 법령에 따른 정당한 이용허락을 받아야 한다(제2항, 제3항). 데이터활용공통기반시스템을 통하여 데이터를 활용하려는 기관의 장은 그 목적과 활용하려는 데이터 내용 등을 구체적으로 작성하여 행정안전부장관에게 데이터 활용을 신청할 수 있다. 활용 제외되는 것으로는 국가안전보장 등의 공익이나 국민의 권리를 현저히 침해할 소지가 있는 경우와 업무 관련성이 없는 경우이다(시행령 제35조의 4).

그런데 문제가 되는 것은 정보공개법의 비공개대상정보의 범위와 데이터활용공통기반시스템의 공동활용 대상이 되는 정보의 범위가 다름에도 불구하고, 이를 동일시하고 있는 점이다. 정보공개법의 비공개대상정보는 정보공개의 정당한 이익이 있는 경우를 고려한 것으로 그 목적과 데이터활용공통기반시스템을 통한 공동활용의 목적은 다르므로 이를 목적이 다른 법률에서 동일하게 사용해서는 안 된다.

아. 행정정보의 제공과 공동이용, 시스템 연계·통합 등의 다양한 규정

이상으로 본 바와 같이 전자정부법은 행정의 효율성 제고라는 측면에서 행정정보의 연계를 포함하는 행정정보의 공동이용에 대한 여러 규정을 두고 있다. 그런데 행정정보의 제공, 공동이용, 시스템의 연계·통합 등은 구체적인 이용형태나 방식, 절차 등에 따라서 법적 의미와 실제적인 위험 등에 있어서도 천차만별이므로 각각의 내용에 대해서는 개별적인 검토가 필요하다.

예를 들어 (i) 행정정보의 개별적인 제공과 (ii) 행정정보의 공공기관 간 또는 민간과의 공동이용, 더 나아가 (iii) 행정정보시스템의 연계나 (iv) 행정정보 시스템의 통합 구축은 해당 행정정보에 대한 접근권한의 관리, 이용의 양상, 이용의 방식 등에 있어

서 현저한 차이가 있다. 예컨대, 주민등록증 발급 시 수집하는 지문정보의 법률상 수집·관리주체인 행정안전부가 해당 지문정보 데이터베이스를 경찰청과 공동관리를 한다면 이는 지문정보 데이터베이스에 수록된 모든 국민의 정보를 경찰청에 제공한 것과 같이 볼 수도 있다. CCTV의 설치주체가 각 자치구와 지방경찰서로 나뉘어 있는데, CCTV 통합관제센터 시스템을 행정안전부나 경찰청에서 연계하거나 통합 관리하는 경우도 연계나 통합의 방식과 유형에 따라서 법적으로 허용되기도 하고, 허용되지 않을 수 있는 것도 마찬가지이다.

이처럼 행정정보의 제공, 공동이용, 시스템의 연계·통합 등은 구체적인 이용형태나 방식, 절차 등에 따라서 법적 의미와 실제적인 위험 등에 있어서도 천차만별이므로 그에 대한 개인정보보호 측면에서의 면밀한 검토가 필요하다. 그런데 전자정부법은 그에 대한 구체적인 기준이나, 평가 절차 등이 없어서 효율성을 이유로 남용될 가능성을 배제할 수 없다.

자. 개인정보 보호원칙

전자정부법은 행정정보의 공동이용, 전자적 시스템의 연계·통합 등의 과정에서 개인정보보호법으로 보장되는 개인정보주체의 권리가 침해되지 않도록 하기 위하여 몇 가지 규정을 두고 있지만 충분하다고 보기는 어렵다.

개인정보주체의 동의권 보호를 위해서 이용기관이 공동이용센터를 통하여 개인정보가 포함된 행정정보를 공동이용할 때에는 정보주체가 다음 각 호의 사항을 알 수 있도록 정보주체의 사전동의를 받아야 하는데, 이를 개인정보보호법의 동의로 갈음한다(제42조).

그런데 알려야 하는 사항이 공동이용의 목적, 공동이용 대상 행정정보 및 이용범위, 공동이용 대상 이용기관의 명칭이어서 개인정보보호법의 고지사항보다 축소되어 있다(제42조 제1항). 게다가 아래와 같이 매우 포괄적인 동의 예외규정을 두고 있다. 아래의 경우에 해당하여 정보주체의 사전동의를 받을 수 없거나 동의를 받는 것이 부적절하다고 인정되면 이용기관은 그 행정정보를 공동이용한 후 사후 공고 등의 절차만 거치면 된다고 하고 있다(제42조 제2항). 물론 행정안전부장관은 이와 같이 정보주체의 사전동의 없이 공동이용할 수 있는 업무와 행정정보의 구체적인 범위를 대통령령으로 정하는 바에 따라 공개하여야 한다고는 하고 있지만(제42조 제3항) 법령의 근거 없는 개인정보주체의 권리 제한임은 분명하다.

1. 정보주체의 생명 또는 신체를 보호하기 위하여 긴급하게 공동이용할 필요가 있는 경우

2. 법령에 따라 정보주체에게 의무를 부과하거나 권리·이익을 취소·철회하는 업무를 수행하기 위하여 공동이용이 불가피한 경우

3. 법령을 위반한 정보주체에 대한 조사 또는 처벌 등 제재와 관련된 업무를 수행하기 위하여 공동이용이 불가피한 경우

4. 그 밖에 법령에서 정하는 업무를 수행함에 있어서 정보주체의 사전동의를 받는 것이 그 업무 또는 정보의 성질에 비추어 현저히 부적합하다고 인정되는 경우로서 대통령령으로 정하는 경우

특히 이 동의를 개인정보보호법의 민감정보와 고유식별정보에 대한 동의로도 보고 있다.

차. 제도 개선

전자정부법을 통한 행정정보의 연계와 관련한 규율들에 대해서는 다음과 같은 제도 개선을 고려해 볼 수 있을 것이다.

- 전자정부법이 행정의 효율성을 달성하기 위한 법령이라는 점에서 전자정부법의 규정들이 개인정보보호법에 의한 정보주체의 권리를 훼손하는 것이 아니라는 점을 밝혀 놓는 것이 바람직할 것이다.
- 전자정부기본계획은 개인정보보호 기본계획의 내용에 부합해야 한다는 점도 밝혀 놓는 것이 바람직할 것이다.
- 행정정보, 데이터 시스템의 공동이용, 공유, 연계, 통합 등은 기술적인 용어로서 구체적인 개인정보의 수집, 이용, 보유, 관리 등의 내용이 검토되고, 개인정보의 안전이 충분히 보장되어야 한다.
- 행정정보 공동이용 우선의 원칙이나, 행정정보 의무적 공동이용 등은 지나치게 행정 효율만을 추구할 것이 아니라 개인정보 보호원칙을 준수하여야 한다.
- 개인정보주체의 사전동의를 형식화하는 것은 개인정보보호법을 무력화하는 것으로 부당하므로, 개인정보보호법에 의하도록 하는 것이 바람직하다.

4. 공공데이터의 제공 및 이용 활성화에 관한 법률

(1) 개요

공공데이터의 제공 및 이용 활성화에 관한 법률(이하 '공공데이터법')은 공공데이터의 제공을 통한 이용 활성화를 목적으로 제정되었다. 제정 목적에서부터 '스마트폰 대중화에 따라 교통, 기상, 공간, 복지, 보건, 식품, 관광, 환경 등 국민의 생활 전반에 걸쳐 생성된 공공데이터는 스마트산업의 핵심자원으로서 그 중요성이 부각되고 있는 바, 공공데이터를 국민이 최우선으로 이용할 수 있도록 보장하고, 공공기관에 공공데

이터 제공의무를 부여하며, 효과적인 민간제공과 이용 활성화를 지원할 수 있는 법적 근거를 마련함으로써 공공데이터가 민간의 창의성과 결합하여 고부가가치 신산업으로 발전할 수 있는 기반을 마련하고, 신규 일자리를 창출하며, 정부의 행정 혁신으로 국민의 삶의 질을 향상시키는 데 이바지하려는 것'[203]을 목적으로 한다고 밝히고 있다.

그런데 공공데이터법은 아래에서 살펴볼 것처럼 제공받은 공공데이터 끼리의 연계, 공공데이터와 민간이 보유하고 있는 정보와의 연계 가능성을 내포하고 있다.

(2) 개인정보보호의 원칙과 모순되는 기본 원칙과 추진체계

가. 다종다양한 공공기관의 공공데이터

공공데이터법은 공공기관이 보유, 관리하고 있는 공공데이터의 제공 및 이용 활성화를 위한 법률인데, 공공기관의 범위, 공공데이터의 범위가 매우 넓다. 즉, 공공기관이란 국가기관, 지방자치단체는 물론 각급 학교, 공기업 등까지 포함하여 매우 폭넓게 정의하고 있고, 공공데이터에는 행정정보는 물론, 공공기관이 생산한 정보나 전자기록물 등이 모두 포함된다.

따라서 여기에는 정부에서 간행하는 간행물도 포함되지만, 공공기관에서 보유하고 있는 개인정보도 포함된다. 특히 개인의 건강정보나 형사처벌에 대한 정보 등과 같은 민감한 정보도 포함되어 있다. 2017년 12월 11일 현재 공공데이터포털에는 24.517건의 데이터가 공개되어 있다고 한다.[204]

나. 공공데이터의 종류를 불문한 기본원칙의 문제점

공공데이터법은 공공기관의 공공데이터에 대해서 적용할 기본원칙을 규정하고 있는데, 그 내용은 정보공개법이나, 개인정보보호법과 복잡한 충돌을 야기하고 있다.

기본원칙은 ① 공공기관은 누구든지 공공데이터를 편리하게 이용할 수 있도록 노력하여야 하며, 이용권의 보편적 확대를 위하여 필요한 조치를 취하여야 한다. ② 공공기관은 공공데이터에 관한 국민의 접근과 이용에 있어서 평등의 원칙을 보장하여야

---

203) 공공데이터법 제정이유(2013.7.30., 제정)

204) 데이터 범주는 아래와 같다.

건축정보, 국민건강정보, 상권정보, 수산정보, 실시간수도정보, 농수축산 경락 및 조사가격정보, 등산로 및 국가생물종 정보, 부동산종합정보, 통합재정정보, 지방행정정보, 법령정보, 부동산거래관리정보, 지방재정정보, 식의약품종합정보, 도로명주소정보, 산업재산권정보, 노동보험정보, 국가통계통합정보, 재난관리정보, 교육행정정보, 음식물쓰레기정보, 국가공간정보, 해양공간정보, 기상정보, 국가종합전자조달정보, 수출입, 역통계, 국민의료정보, 국민연금정보, 고용보험정보, 도시계획정보, 사회보장정보, 산업기술정보, 교통사고정보, 지진대피소정보, 생태자연도, 일자리종합정보

한다. ③ 공공기관은 정보통신망을 통하여 일반에 공개된 공공데이터에 관하여 제28조 제1항 각 호의 경우를 제외하고는 이용자의 접근제한이나 차단 등 이용저해행위를 하여서는 아니 된다. ④ 공공기관은 다른 법률에 특별한 규정이 있는 경우 또는 제28조 제1항 각 호의 경우를 제외하고는 공공데이터의 영리적 이용인 경우에도 이를 금지 또는 제한하여서는 아니 된다. ⑤ 이용자는 공공데이터를 이용하는 경우 국가안전보장 등 공익이나 타인의 권리를 침해하지 아니하도록 법령이나 이용조건 등에 따른 의무를 준수하여야 하며, 신의에 따라 성실하게 이용하여야 한다는 것이다.

그런데 앞에서도 본 것처럼 공공데이터에는 정부간행물도 포함되지만, 개인의 민감한 건강정보나, 형사처벌에 대한 정보 등도 포함되어 있는데, 이를 구분하지 않고 공공데이터로 묶은 상태에서 이를 마치 문화의 향유나, 보편적 서비스의 대상인 전기통신역무와 유사하게 취급하여 '이용권의 보편적 확대를 위해 필요한 조치를 취할 의무'를 부과하거나, '접근과 이용에 대한 평등의 원칙을 보장해야 한다'는 것은 그 대상별로는 적절한 원칙으로 보기 어렵다.

실제로 정보공개법은 공공기관이 보유·관리하는 정보는 국민의 알 권리 보장 등을 위하여 이 법에서 정하는 바에 따라 적극적으로 공개하여야 한다는 정보공개의 원칙을 규정하고 있을 뿐인데(제3조), 공공데이터법은 이를 넘어서는 원칙을 제시하고 있는 것이다.

특히 개인정보보호법, 정보공개법, 공공데이터법 사이에는 개인정보보호법의 개인정보, 정보공개법에서 '비공개 대상정보'(제9조)를 제외한 정보공개의 대상인 공공기관의 정보(제14조의 부분공개 포함), 공공데이터법의 제공대상 공공데이터(제17조, 제2항의 기술적 분리 포함)의 관계가 각기 다른 용어를 사용하고, 각기 다른 기준을 가지고 있기 때문에 매우 복잡한 문제를 야기하고 있다.

특히 공공데이터 제공과 이용에 따른 면책(제36조) 규정은 매우 이례적일 뿐만 아니라, 개인정보보호법의 체계와 모순을 이루고 있다.

다. 추진체계

한편 공공데이터법은 국무총리 소속으로 공공데이터전략위원회 설치(제5조), 3년마다 공공데이터 제공 및 이용 활성화에 관한 기본계획 수립((제7조, 제8조), 안전행정부장관의 공공기관 대상 공공데이터의 제공기반조성, 제공현황 등 제공 운영실태 평가, 보고(제9조), 한국정보화진흥원에 공공데이터활용지원센터 설치·운영(제13조), 공공데이터 목록 등록의무(제18조), 공표된 제공대상 공공데이터의 경우 소관 공공기관이나 공공데이터 포털 등에서 별도의 신청절차 없이 제공받을 수 있도록 함(제26조),

공공기관의 공공데이터 제공거부 및 제공중단에 관한 분쟁조정을 하게 하기 위한 공공데이터제공분쟁조정위원회 설치(제29조) 등의 규정을 두고 있는데, 이와 같은 추진체계는 개인정보보호법의 추진체계와 상호 정합성을 갖도록 조정이 필요하다.

(3) 제공대상 공공데이터, 정보공개 대상정보, 개인정보의 관계

가. 개인정보보호법과 정보공개법의 충돌

개인정보보호법은 공공기관이 보유하고 있는 개인정보에 대해서 공공기관이 개인정보를 목적 외의 용도로 이용하거나 이를 제3자에게 제공하지 아니하면 다른 법률에서 정하는 소관 업무를 수행할 수 없는 경우로서 보호위원회의 심의·의결을 거친 경우와 같이 예외적인 경우가 아니라면 원칙적으로 당사자의 동의가 없이 제3자에게 공개하거나 제공하는 것을 허용하지 않고 있다(개인정보보호법 제18조)[205]. 그리고 개인정보는 수집 목적을 달성하는 경우 삭제하도록 하고 있다.

반면, 공공기관의 정보공개법에 관한 법률(이하 '정보공개법')은 비공개 대상정보가 아니라면 정보를 공개해야 한다.

일반적으로 타인의 개인정보는 이를 공개하는 것이 개인정보보호법의 위반이 되는 경우에는 이를 비공개 대상으로 하되, '알 권리' 보장을 위해서 공개를 해야 할 공익적 가치가 크다면 공개대상이 될 수 있다고 규정을 하는데, 우리의 정보공개법은 이와는 다른 입법을 하고 있다. 정보공개법의 비공개 대상정보는 개인정보보다 범위가 좁게 규정하면서 '공익적 필요'와의 이익형량을 생략하고 있다. 즉, '해당 정보에 포함되어 있는 성명·주민등록번호 등 개인에 관한 사항으로서 공개될 경우 사생활의 비밀 또는 자유를 침해할 우려가 있다고 인정되는 정보'를 비공개정보로 하고 있다.[206] 헌

---

205) 다음과 같은 예외적인 경우로서 정보주체 또는 제3자의 이익을 부당하게 침해할 우려가 없는 경우
1. 정보주체로부터 별도의 동의를 받은 경우
2. 다른 법률에 특별한 규정이 있는 경우
3. 정보주체 또는 그 법정대리인이 의사표시를 할 수 없는 상태에 있거나 주소불명 등으로 사전동의를 받을 수 없는 경우로서 명백히 정보주체 또는 제3자의 급박한 생명, 신체, 재산의 이익을 위하여 필요하다고 인정되는 경우
4. 통계작성 및 학술연구 등의 목적을 위하여 필요한 경우로서 특정 개인을 알아볼 수 없는 형태로 개인정보를 제공하는 경우
5. 개인정보를 목적 외의 용도로 이용하거나 이를 제3자에게 제공하지 아니하면 다른 법률에서 정하는 소관 업무를 수행할 수 없는 경우로서 보호위원회의 심의·의결을 거친 경우
6. 조약, 그 밖의 국제협정의 이행을 위하여 외국정부 또는 국제기구에 제공하기 위하여 필요한 경우
7. 범죄의 수사와 공소의 제기 및 유지를 위하여 필요한 경우
8. 법원의 재판업무 수행을 위하여 필요한 경우
9. 형(刑) 및 감호, 보호처분의 집행을 위하여 필요한 경우

206) 다만, 다음 각 목에 열거한 개인에 관한 정보는 제외한다.

법적 가치에 입각해서 본다면 정보공개를 통해서 개인정보가 공개되는 것은 개인정보 자기결정권의 침해가 될 테지만, 알 권리에 의한 공개의 이익이 더 크다면 정보공개의 대상이 된다고 보는 것이 우리나라 헌법상 기본권의 이익형량 구조에 부합할 것이다. 개인정보 자기결정권과 사생활의 비밀과 자유는 보호대상과 범위가 다르기 때문이다. 따라서 현재의 입법 태도보다는 (i) 개인정보는 비공개 대상이지만, (ii) 공익적 필요와의 이익형량을 통해서 공개할 수 있다고 규정하는 것이 명확한 논리구조가 될 것이다. 그러나 실제로는 운용에 있어서는 큰 차이가 없도록 운영되고 있는 것으로 보인다.

나. 제공대상 공공데이터와 개인정보의 관계

공공데이터법은 원칙적으로 정보공개법에 의한 비공개 대상정보를 제외한 정보는 제공대상 공공데이터로 본다(제17조 제1항). 아울러 공공데이터법은 '비공개 대상정보를 기술적으로 분리할 수 있는 때에는 이에 해당하는 부분을 제외한 공공데이터를 제공하여야 한다'(제17조 제2항)는 규정을 두고 있다. 이와 관련하여 다음과 같은 문제가 있다.

첫째, 정보공개법과 공공데이터법의 제정 목적의 차이를 고려하지 않은 규정이다. 정보공개법은 알 권리를 보장하고, 행정의 투명화를 통한 민주주의 가치를 실현하기 위한 것이다. 따라서 정보공개 청구자, 정보공개 청구의 목적에 따라서 정보공개를 할지 여부가 결정된다.

예를 들어 대법원은 우리나라의 교육실태를 연구하기 위한 목적으로 교육인적자원부장관에게 학업성취도평가정보와 수능시험정보에 대한 정보공개를 신청한 사건에서 전자와 후자의 정보가 공개될 경우의 영향, 연구의 목적 등을 고려한 판단을 하고 있다(대법원 2010.2.25, 선고, 2007두9877, 판결). 다종다양한 정보에 대해서 정보공개청구가 있는 경우, 해당 정보공개청구의 목적이 무엇인지, 청구자가 누구인지를 고려해서 알 권리와의 이익형량을 해야 한다는 것이 대법원의 판결 취지이다. 따라서 공공데이터법에서 정보공개법의 규정을 그대로 들어서 제공대상 공공데이터 여부를 판단하도록 하고 있는 것은 문제가 있다. 실제로 공공데이터법은 공공데이터 분쟁조정위원회를 두고 있는데, 정보공개심의위원회와 별개의 조직으로 판단의 정합성 등에 있

가. 법령에서 정하는 바에 따라 열람할 수 있는 정보
나. 공공기관이 공표를 목적으로 작성하거나 취득한 정보로서 사생활의 비밀 또는 자유를 부당하게 침해하지 아니하는 정보
다. 공공기관이 작성하거나 취득한 정보로서 공개하는 것이 공익이나 개인의 권리 구제를 위하여 필요하다고 인정되는 정보
라. 직무를 수행한 공무원의 성명·직위
마. 공개하는 것이 공익을 위하여 필요한 경우로서 법령에 따라 국가 또는 지방자치단체가 업무의 일부를 위탁 또는 위촉한 개인의 성명·직업

어서도 문제가 될 수 있다.

둘째, 공공데이터법은 비공개 대상정보를 '기술적으로 분리할 수 있는지'를 기준으로 해서 기술적으로 분리 가능하면 분리해서 나머지 정보는 제공하도록 하고 있는데, '기술적 분리 가능성'이 제공대상 공공데이터와 제공불가 공공데이터의 구분 기준이 될 수 없다. 이에 대해서는 항을 나누어서 검토한다.

다. 기술적 분리 가능성과 '익명정보' 여부

기술적인 분리로 해당 정보가 '익명정보'로 되는 것은 아니다. 기술적인 분리 가능성이 있어서 '비공개 대상정보'를 기술적으로 분리해 내더라도 남은 정보에서 다른 정보와의 조합을 통해서 개인을 유추할 수 있다면 기술적으로 분리된 나머지 정보는 여전히 개인을 식별할 가능성이 있는 개인정보이기 때문이다.

이 경우 해당 정보를 공개하는 것이 국민의 알 권리 보장을 위해 필요하여 해당 개인의 개인정보 자기결정권보다 우월하다면 이를 공개하는 것은 허용될 수 있겠지만, 이를 제공대상 공공데이터로 취급하여 영리적 목적을 불문하고, 민간 활용을 촉진하도록 하는 것은 부적절하다.

라. 구체적 사례

예를 들면, 공공데이터 포털 사이트에서 2016년 교통사망사고 정보(4,119건)를 공공데이터로 공개하고 있는데, 아래와 같이 '비공개 대상정보'라고 할 수 있는 교통사망사고의 사고 관련 당사자들의 이름이나 주민등록번호 등과 같은 정보를 기술적으로 분리한 후 나머지 정보인 발생시간, 발생장소, 사고 유형과 과실 등에 대해서 모든 정보를 그대로 공개하고 있다. 그런데 공개된 나머지 정보를 가지고도 사고 관련 신문기사 등을 통해서 개인을 식별할 수 있으므로 개인정보로 볼 수 있다.

실제로 공공데이터제공분쟁조정위원회에 경찰청의 교통사고 통계 원시데이터를 공개해 달라는 취지로 분쟁조정신청이 되었었는데(사건번호 2015-011, 경찰청 교통사고 통계 원표 데이터 사건), 분쟁조정위원회는 교통사고 통계원표 항목 중에서 교통사고 당사자의 주민등록번호, 운전면허번호, 차량등록번호, 피해자의 주민등록번호는 개인식별정보에 해당하는 정보로써 비공개대상정보에 해당하지만, 공공데이터법 제17조 제2항에 따르면 동조 제1항에 해당하는 비공개 대상정보가 포함되어 있는 데이터의 경우 해당 정보를 기술적으로 분리할 수 있으면 분리하여 제공하도록 정하고 있는바, 교통사고 통계 원시데이터 중 개인정보를 기술적으로 분리할 수 있으면 분리하여 제공하여야 한다'고 판단했다.

표 4-4 2016년 교통사망사고정보

발생년	발생년월일시	발생분	주야	요일	사망자 수	사상자 수
2016	2016122320	35	야간	금	1	1
2016	2016122517	48	주간	일	1	1

중상자 수	경상자 수	부상신고자 수	발생지시도	발생지 시군구	사고유형_ 대분류	사고유형_ 중분류
0	0	0	경기	광주시	차대사람	기타
0	0	0	서울	금천구	차대사람	횡단 중

사고유형	법규위반_대분류	법규위반	도로형태_대분류	도로형태
기타	운전자법규위반	안전운전 의무 불이행	교차로	교차로 내
횡단 중	운전자법규위반	안전운전 의무 불이행	단일로	기타단일로

당사자 종별_1당_ 대분류	당사자 종별_1당	당사자 종별_2당_ 대분류	당사자 종별_2당	발생위치X _UTMK	발생위치Y _UTMK	경도	위도
승용차	중형	보행자	보행자	981731	1932086	127.2936	37.38769
승용차	중형	보행자	보행자	945989	1942347	126.8891	37.47878

실제로 경찰청은 분쟁조정위원회에서 조정부 회의 이후, 내부 검토 결과, 68개의 교통사고 통계원표 항목 중 발생일시, 요일, 발생 시군구, 사망자 수, 중상자 수, 경상자 수, 부상신고자 수, 사고 유형, 성별, 연령, 당사자 종별, 차량용도, 도로종류, 차도 폭, 기상상태 등 총 15개 항목에 대해서 제공할 수 있다고 하였고, 좌표정보의 정확도 제고 후 발생 위치정보도 제공하기로 했다. 그 결과 위와 같은 정보가 공공데이터로 공개되고 있는 것이다. 그러나 교통사고조사규칙 제36조에 의하더라도 교통사고 조사과정에서 취득한 사고 관련자의 인적사항 등 개인정보는 사고조사목적 또는 법령에 의하여 이용되거나 제공되는 경우 외에 다른 용도로 사용할 수 없도록 정하고 있는데, 위와 같은 정보는 손쉽게 재식별이 가능하므로 교통사고조사규칙에 위반되는 것으로도 볼 것이다.

또 다른 사례로는 공공데이터로 공개되는 환자 데이터셋을 들 수 있다. 환자 데이터셋도 개인 식별정보를 기술적으로 분리했지만 다른 정보와의 연계를 통해서 개인을 식별할 수 있으므로 개인정보로 볼 수 있다.

마. 3자의 복잡한 관계

이와 같이 개인정보와 비공개 대상정보, 제공대상 공공데이터는 그 관계가 매우 복잡한데, 특히 공공데이터법이 제공대상 공공데이터로 규정하고 있는 것에는 알 권리

충족을 위해 공개가 필요한 정보가 아니면서, 개인을 식별할 수 있는 개인정보가 포함될 가능성이 매우 크다. 이런 점에서 현행 공공데이터법은 개인정보보호법, 정보공개법과 조화를 이루고 있다고 보기 어렵다.

(4) 공공데이터법과 개인정보보호법의 충돌

가. 공공데이터법은 개인정보보호법의 특별법인지

이와 같이 공공데이터법은 개인정보보호법과 충돌하고 있는데 공공데이터법이 개인정보보호법 제6조에서 정한 '개인정보의 보호에 관하여 다른 법률의 특별한 규정'으로 볼 수 있을지는 문제이다(제6조). 만약 이에 해당한다고 본다면, 공공데이터전략위원회/공공데이터의 제공 및 이용 활성화에 관한 기본계획과 개인정보보호위원회/개인정보 보호 기본계획의 충돌 가능성, 각 분쟁조정의 모순 가능성 측면에서 문제가 발생한다.

나. 각 위원회 및 기본계획의 충돌 가능성

공공데이터법은 공공데이터전략위원회를 두고, 공공데이터의 제공 및 이용 활성화에 관한 기본계획을 수립하도록 하고 있는데, 공공데이터법과 개인정보보호법이 서로 충돌하고 있는바, 개인정보보호위원회와 개인정보 보호 기본계획과의 관계가 문제 된다.

공공데이터전략위원회는 공공데이터에 관한 정부의 주요 정책과 계획을 심의·조정하고 그 추진사항을 점검·평가하기 위하여 국무총리 소속으로 두고 있고(제5조), 여기서는 공공데이터의 제공 및 이용 활성화에 관한 기본계획과 시행계획의 수립·변경에 관한 사항, 부문 계획의 작성지침, 실태조사 결과에 대한 개선 또는 시정 권고에 관한 사항, 제공대상 공공데이터 목록의 심의·의결 및 목록 공표에 관한 사항, 공공데이터 목록의 제외에 관한 사항, 공공데이터의 제공 및 이용과 관련된 정책 및 제도 개선에 관한 사항 등을 심의한다(제6조). 한편 공공데이터의 제공 및 이용 활성화에 관한 기본계획에는 공공데이터 제공 및 이용 활성화의 기본목표와 추진방향, 공공데이터의 제공형태 및 제공방안에 관한 사항, 공공데이터의 등록 및 이용 현황, 제공 및 이용 가능한 공공데이터의 확대, 공공데이터의 민간 활용 촉진에 관한 사항, 공공데이터의 품질관리에 관한 사항, 공공데이터 관련 제도 및 법령의 개선에 관한 사항 등을 포함해야 한다. 따라서 공공데이터전략위원회는 공공데이터법에 따라서 개인정보보호위원회의 개인정보보호 기본계획과 모순되는 계획을 수립할 가능성이 매우 크다.

다. 각 분쟁조정의 모순 가능성

공공데이터제공분쟁조정위원회, 개인정보분쟁조정위원회, 정보공개심의위원회의 관계도 문제이다. 서로 다른 기준에 의하여 분쟁조정이 이루어지므로 결론에 상호 모순이 발생할 가능성이 크다.

공공데이터 포털에서는 국민 알 권리 목적의 정보열람이나 정보공개 청구, 민원신청 등은 정보공개포털(www.open.go.kr)에서 접수하며, 이를 공공데이터 포털에서 신청하는 경우 신청인에게 통보(메일)하고 종결 처리한다고 소개하고 있는데,[207] 무엇이 알 권리 목적의 정보열람 청구인지, 민원신청인지 구별하기는 곤란하다.

(5) 공공데이터의 제공과 데이터 연계

공공데이터법에 따른 데이터의 연계는 두 가지 연계의 가능성이 있다. 첫째는 별도로 제공받은 데이터를 직접 연계하는 것이다. 예를 들어 연계가 가능한 정보가 각각 포함된 정보를 제공받아 연계를 시도할 수 있다. 둘째는 제공하는 공공기관에서 데이터를 연계하여 제공하는 것이다.

현재 공공데이터법은 공공데이터의 연계와 관련하여 특별한 규정은 두고 있지 않고, 다만 데이터 연계가 가능할 수 있는 규정만 두고 있다. 즉, 공공데이터법에는 공공데이터 포털의 운영과 관련하여 행정안전부장관은 공공기관의 장에게 공공데이터 포털의 구축과 운영에 필요한 공공데이터의 연계, 제공 등의 협력을 요청할 수 있다. 이 경우 요청을 받은 공공기관의 장은 특별한 사유가 없는 한 이에 따라야 한다는 규정을 두고 있다(제21조 제2항). 시행령에서는 공공기관의 장은 법 제21조 제1항에 따른 공공데이터 포털에 연계·제공하는 소관 공공데이터의 최신성, 정확성 및 상호연계성이 유지되도록 하여야 한다고 규정하고 있다(제16조).

이에 의하면 별도로 제공받은 데이터를 제공받은 자가 직접 연계하는 것도 가능할 수 있고, 제공하는 공공기관에서 데이터를 연계하여 제공하는 것도 가능할 것이다.

(6) 공공데이터 제공과 개인정보 자기결정권 침해 가능성

공공데이터법은 공공기관의 공공데이터를 정보공개법의 비공개정보가 아닐 경우 공개하고 있고, 비공개정보가 포함되어 있더라도 기술적 분리 가능성이 있다면 제공이 가능하다고 규정하고 있다. 그런데 앞에서도 본 것처럼 비공개 대상정보를 다른 정보

207) 공공데이터 제공신청 소개(https://www.data.go.kr/participation/provdReqst/dcIndex.do)

와 기술적으로 분리하여 제거하고 나머지 정보를 제공하더라도 나머지 정보만으로 특정 개인을 추론할 수 있다면 나머지 정보는 특정 개인의 개인정보가 된다. 이 경우 공공데이터법에 따른 공공데이터 제공은 개인정보 자기결정권의 침해로 볼 것이다.

(7) 개선 방향

- 공공데이터법의 추진체계가 개인정보보호 추진체계와 모순되지 않도록 해야 하고, 개인정보의 보호에 관하여는 공공데이터법이 개인정보보호법에 모순되지 않도록 해야 한다.
- 공공데이터의 이용, 제공에 대한 원칙은 분야별로 달라야 한다.
- 정보공개법, 개인정보보호법과 공공데이터법이 모순되지 않도록 규정해야 한다.
- 특히 공공데이터법에서 '기술적 분리 가능성'을 기준으로 제공 여부를 판단하는 것은 개인정보보호법제와 부합하지 않는 개념이다.

5. 데이터기반행정 활성화에 관한 법률

(1) 입법 예고된 법안의 내용

2017년 5월 8일 행정안전부는 '데이터기반행정 활성화에 관한 법률' 제정안(이하 제정안)을 입법 예고하였다. 주요 제정이유는, 행정·공공기관이 보유하고 있는 대규모 데이터 및 민간 보유 데이터, 인터넷의 공개된 데이터를 분석·활용하기 위한 원칙과 절차를 마련하는 데 있다.

제정안은 이를 위하여 우선 데이터기반행정 활성화 목적, 적용 범위를 규정하였다(안 제1조~제2조). 공공기관에 제도 및 환경 정비 등 기반 조성, 민간전문가 참여로 객관성·정확성·공정성 제고, 개인정보보호, 관련 정보공개, 재원 마련 등 책무를 부여하고(안 제3조), 데이터기반행정 대상 업무로는 업무혁신, 행정서비스 향상, 실시간대응, 위험요소 제거, 사전예측 등 데이터 분석·활용이 필요한 분야에서 발굴·추진하도록 하였다(안 제4조). 추진체계로는 데이터 제공 등에 대한 심의를 위해 국무총리 소속 데이터기반행정활성화위원회를 설치하고, 행정자치부[행정안전부] 장관이 기본계획을, 공공기관의 장이 시행계획 등을 수립하도록 하였다(안 제6조～제8조).

추진 절차는 다음과 같다(안 제9조～제11조). 공공기관은 자체적으로 발굴·선정하는 과제 및 위원회가 지정하는 과제에 대한 조치계획을 위원회에 제출하고 결과에 따라 실행계획을 수립·추진한다. 공공기관은 자체과제 및 지정과제를 위해 필요한 경우 타 공공기관의 데이터를 요청할 수 있도록 하고, 위원회에 갈등 조정 역할을 부여하였다.

공공기관은 소관 업무를 처리하기 위해 필요한 경우 민간데이터를 요청할 수 있으며, 대가 등에 대한 사항은 협약을 통해 결정하도록 하였다. 제정안은 그 밖의 기반 구축에 대해서도 규정하였다(안 제12조～제22조). 공공기관은 생성·수집하여 보유하는 데이터에 대한 메타데이터 및 데이터관계도를 체계적으로 관리하고, 행정자치부[행정안전부] 장관은 이를 통합·연계한 중앙메타데이터관리시스템과 데이터맵을 구축·운영한다. 또 장관은 데이터기반행정을 위해 필요한 데이터의 연계·수집·저장·분석 및 분석결과의 공유·활용을 위한 정부통합데이터관리플랫폼을 구축한다. 공공기관은 데이터기반행정에 관한 업무를 총괄하는 책임관을 지정하고 자체 데이터분석센터를 설치·운영하며, 행정자치부[행정안전부] 장관은 정부통합데이터분석센터를 설치·운영하고 데이터기반행정 전문기관을 지정·운영한다. 공공기관은 자체 점검·평가를 실시하고, 행정자치부[행정안전부]는 시범사업, 데이터기반행정 관련 표준화, 우수사례 발굴·홍보·포상, 표준분석모델 정립·확산 등을 추진한다. 데이터기반행정 활성화를 위한 민간과의 협력 및 국제협력 사항도 규정하였다.

(2) 평가

이상과 같은 제정안에 대하여 개인정보보호위원회는 개인정보보호법 제8조의2에 따른 개인정보 침해요인 평가를 실시하였고, 2017년 7월 결정(제2017-15-125호)에서 다음과 같이 개선을 권고하였다.

행정 데이터에는 국민의 민감한 개인정보가 포함되어 있으므로 이를 이용, 제공, 연계하는 경우 국민의 프라이버시를 침해할 위험이 크므로 정보주체의 명확한 동의를 기반으로 하는 것이 바람직하며 개인정보보호법의 제 규정을 준수해야 한다. 다만 대량의 행정 데이터를 개별적인 동의를 받아 처리하기 어려운 현실적 제약과 행정 데이터 연계를 통해 얻을 수 있는 공공의 이익을 고려하여 개인정보보호법 제18조 제2항 제4호 등에 따라 개인을 알아볼 수 없는 형태로 데이터 연계를 추진하는 것은 공익을 위해 필요한 것으로 보인다. 그러나 이 경우에 있어서도 프라이버시 침해 위험으로부터 개별 국민을 보호하기 위해 개인정보보호법에 따른 명확한 근거가 필요하고 적정한 관리·감독체계 구축, 데이터 처리에 있어서 완전한 기능 분리를 통한 개별 국민의 프라이버시 보호와 인적·물적·기술적·관리적 조치를 통한 데이터의 안전성을 확보하여야 하며 연계대상 행정 데이터의 내역 및 연계절차를 국민이 알 수 있도록 투명하게 공개할 필요가 있다.

이를 위하여 첫째, 제정안에 따른 데이터 요청 및 처리에 있어서 개인정보의 처리 및 보호에 관한 사항에 대하여는 개인정보보호법에서 정하는 바에 따른다는 내용을

추가하여 개인정보보호법이 우선하여 적용됨을 명확히 하여야 한다(안 제5조 및 제10조).

둘째, 데이터 연계를 위한 연결키로서 주민등록번호를 사용할 경우에는 법령에 주민등록번호를 처리할 수 있는 구체적·개별적인 근거가 있어야 하고 주민등록번호 이외에 성명, 생년월일, 주소 등의 식별자를 사용하는 경우에는 해당 정보주체의 동의, 법률의 특별한 규정, 보호위원회의 의결 등 개인정보보호법 제18조 제2항 각 호의 근거가 필요하다는 점을 명확히 해야 한다.

셋째, '소관 업무를 처리하기 위하여 필요한 경우' 공공기관이 공공기관 이외의 법인·단체 또는 개인이 보유한 민간데이터의 제공을 요청할 수 있도록 한 조항(안 제10조 제5항)은 개인정보보호법 제15조 제1항 제3호의 '소관 업무 수행을 위하여 불가피한 경우'와 정합성을 유지하도록 변경해야 한다.

넷째, 데이터기반행정 관련 시스템 구축 전에 개인정보보호법 제33조에 따라 개인정보 영향평가를 수행하여야 하고 동 제정안 시행 후 개별 과제 시행에 있어서도 법상 요건에 해당할 경우에는 개인정보 영향평가를 수행하여야 함을 명확히 해야 한다.

다섯째, 데이터기반행정활성화위원회가 데이터기반행정과제 선정 및 승인을 함에 있어 과제 수행에 있어 공익적 견지에서 데이터 연계가 불가피한지 여부를 심사하고 개인정보보호법 제18조 제2항 각 호의 요건을 준수하고 있는지 여부를 확인하여 필요시 보호위원회의 의결을 거치도록 하는 등 필요한 조치를 할 수 있도록 관련 체계를 보완해야 한다.

여섯째, 데이터기반행정 과제 선정 및 발굴 의무화 내용은 데이터기반행정을 선언적으로 권장하는 선에서 그치고 데이터 연계는 공공기관의 신청을 기반으로 해야 한다.

일곱째, 데이터 연계과정에서 누구도 관련 데이터에 포함된 정보주체를 알아 볼 수 없도록 공공기관 데이터분석센터, 행정자치부 정부통합데이터분석센터(안 제16조) 및 데이터기반행정 전문기관(안 제17조)의 기능이 분리되어 데이터 연계 기능 분리 원칙에 부합하도록 해야 한다. 이때 행정데이터 제공기관으로서 공공기관 데이터분석센터는 연계대상 행정데이터를 식별정보와 속성정보로 분리하여 식별정보를 전문기관에 제공하도록 하고 이 경우 개인정보 제공의 법적요건을 충족하도록 하고, 전문기관은 복수의 공공기관 데이터분석센터로부터 제공받은 연계대상 행정데이터의 식별정보에서 동일인을 확인하여 임시대체키를 부여하고 해당 임시대체키를 각각의 공공기관 데이터분석센터에 다시 제공하도록 해야 한다. 정부통합데이터분석센터는 연계대상 행정데이터를 보유한 각각의 분석센터로부터 임시대체키와 속성정보를 제공받아 행정데

이터를 연계하고 데이터 연계 결과의 익명성을 검증한 후 데이터 연계를 요청한 공공기관의 데이터분석센터에 제공하도록 해야 하고, 데이터 연계를 요청한 공공기관의 데이터분석센터는 정부통합데이터분석센터의 보안시설 내에서 데이터 연계 결과를 열람, 연구에 활용하도록 하고 다만 공공기관의 데이터분석센터가 데이터 보안을 위해 정부통합데이터분석센터와 동일한 시설 및 인적 요건을 갖춘 경우에는 데이터 연계 결과를 제공할 수 있도록 해야 한다.

여덟째, 정부통합데이터분석센터, 데이터분석센터 및 데이터기반행정 전문기관 설치·운영에 있어 보안대책을 보완해야 한다. 데이터 보관 시설 및 장소를 보안구역으로 설정하고 보안 서약 및 관련 기관으로부터 허가를 받은 제한된 인원에 대해서만 해당 구역 및 데이터에 대한 접근을 허용하는 등 엄격한 접근통제를 실시하고, 데이터 전송은 보안이 확보된 전송망을 이용하고 전송 전에 데이터의 안전성을 확인해야 하며, 연계 데이터 등 중요 데이터가 보관되는 서버는 다른 정보를 보관하는 서버와 분리하고 유출에 대한 방어 수단을 구비해야 하고, 기타 개인정보보호법 제29조에 따른 안전조치의무를 준수해야 한다.

이상과 같은 개인정보보호위원회의 권고 내용은 행정 데이터 연계의 공익과 개인정보보호법에 따른 정보주체의 권리를 조화시키고, 영국 디지털경제법 등 데이터 연계 절차에 대한 국제 모범 사례를 참고하여 이를 국내 관련 입법안에 적절히 반영토록 한 것으로 평가할 수 있다. 특히 이러한 권고 배경에는 데이터 연계에 대한 안전대책이 없는 상황에서 재식별 위험이 상존함에도 불구하고 데이터 연계를 강행할 경우 재식별에 따른 개인정보 침해 우려가 있다는 문제의식이 내포되어 있다.

6. 데이터 연계·결합 관련 개인정보 보호법제의 정합성

(1) 개인정보보호법, 전자정부법, 공공데이터법의 조화로운 해석

현재 개인정보보호법, 전자정부법, 공공데이터법에는 개인정보의 제공, 공개, 연계·결합 등에 관하여 동일한 영역을 법률에 따라 상호 모순되게 규정하고 있는 점이 있는데 이 모순은 우리나라 개인정보 보호 법제에서 의도된 것으로 보기 어렵다. 오히려 이로 인해 개인정보보호의 기본 원칙이 심각하게 훼손되고 있으므로, 3법의 관계를 명확하게 정리할 필요가 있다.

전자정부법은 행정정보 시스템의 연계, 통합 등을 주로 효율성의 측면에서 평가하고 추진하는 법률이므로 개인정보보호 원칙이 훼손되거나, 정보주체의 개인정보 자기결정권이 이로 인해서 침해될 수 없다는 점을 분명히 해야 한다.

공공데이터법은 공공데이터 이용 활성화를 목적으로 하고 있지만, 이로 인해서 개인정보보호 원칙이 훼손되거나 정보주체의 개인정보 자기결정권이 침해될 수 없다는 점을 분명히 해야 한다. 이런 관점에서 세 법의 관계를 설정하고 정합성을 갖추도록 정비, 개선해 나가는 것이 바람직할 것이다.

(2) 개인정보보호에 관한 정합성 있는 법체계 유지 필요

전자정부법은 특정한 행정기관의 행정 작용에 대한 법률이 아니고 행정업무의 전자적 처리를 위한 기본 원칙, 절차 및 추진방법 등을 규정함으로써 전자정부를 효율적으로 구현하고 행정의 생산성, 투명성 및 민주성을 높여 국민의 삶의 질을 향상시키는 것을 목적으로 하는 법률인데, 실제로는 공공부문에서 데이터의 연계·결합이 이루어지게 하는 근거 법률로 기능하고 있다. 이런 점에서 전자정부법을 본래의 목적에 부합하도록 해석·적용할 필요가 있다.

공공데이터법은 그 대상이 되는 정보가 공공기관이 보유·관리하는 데이터로서 매우 범위가 넓은데 그 안에는 민감한 개인정보도 포함되어 있다. 공공데이터법은 국민의 공공데이터에 대한 이용권을 보장하고 공공데이터의 민간 활용을 통한 삶의 질 향상과 국민경제 발전에 이바지함을 목적으로 하는 것이므로 국민의 이용권이 개인정보 자기결정권을 초월하는 가치를 갖는 경우로 본래의 목적에 부합하도록 해석·적용할 필요가 있다. 공공데이터법은 기술적 분리 가능성을 근거로 기술적 분리 후의 공공데이터를 제공하도록 하고 있는데 기술적 분리가 이루어진 공공데이터는 여전히 연계·결합이나 개인식별의 가능성이 매우 크므로 이런 위험에 대한 대응이 필요하다.

제2절 국내 보건의료 분야 데이터 연계 현황

1. 개요

개인의 병력이나 질병, 현재의 건강상태 등에 관한 정보는 개인정보 중 가장 민감한 정보 중의 하나다. 이 정보는 공개될 경우 개인에 대한 사회적 낙인이 될 수도 있고, 고용, 보험은 물론 사회생활에서 차별의 원인이 될 수도 있고, 개인의 민감한 사생활 침해가 될 수도 있다.[208] 그래서 개인정보보호법은 건강정보를 민감정보로 보다 더 엄격하게 규율하고 있으며(제23조) 의료법, 약사법 등에서도 진료기록과 처방에 대한 기록을 엄격하게 보호하고 있다.

반면, 건강정보의 집적이나 활용은 환자 개인에게는 질 높은 치료를 제공하기 위해서 필요하고 환자들의 소통, 투명성, 경험, 응답 등에 필수적이기도 하다. 공공의 목적으로는 새로운 치료법과 연구개발을 위해 필요하고 각국에서 재정적 제한을 받는 보건정책 추진에 필수적이기도 하다.[209] 이처럼 건강정보의 활용에 대한 요구도 매우 크다.

그래서 실제로 건강정보의 연계나 결합은 매우 빈번하게 일어나고 있는데 치료 목적, 건강보험의 운영 및 관리 목적, 연구 목적, 보건정책의 평가와 의료서비스의 질 관리 목적 등 다양한 방면의 목적으로 연계나 결합이 이루어지고 있고 개인의 의료기관을 달리하는 건강정보 사이의 연계나 결합, 건강정보와 행정정보의 연계, 건강정보와 조사정보의 연계가 이루어지기도 한다.

실제로 우리나라는 의료부문에서 국민건강보험, 의료급여, 노인장기요양보험 등 사회보장제도의 운영과 관련하여 국민건강보험공단과 건강보험심사평가원이 정보를 수집, 관리하면서 이 정보를 바탕으로 다양한 2차적 활용을 하고 있다. 최근 발표된 OECD 연구보고서의 조사에서도 건강정보의 연계가 조사대상국 중 아이슬란드 다음으로 높은 수준으로 밀도 높게 이루어지고 있다는 것이 드러난 것처럼 우리나라는 다른 나라의 경우와 비교해 보더라도 매우 높은 수준의 건강정보 수집과 집적, 연계가 이루어지고 있다. 그뿐 아니라 질병관리본부나 국립암센터, 국립보건원 등은 암 환자 정보나 코호트, 인체 은행 등 연구기반정보를 갖추고 있고, 이를 데이터 연계나 결합을 시도하면서 활용하고 있고, 그 활용의 폭을 넓히고자 하고 있다.

---

208) Institute of Medicine (US) Committee on Health Research and the Privacy of Health Information: The HIPAA Privacy Rule(2009), "Beyond the HIPAA Privacy Rule: Enhancing Privacy, Improving Health Through Research", Nass SJ, Levit LA, Gostin LO, editors, Washington (DC): National Academies Press (US), https://www.ncbi.nlm.nih.gov/books/NBK9579.

209) OECD(2015), "Health Data Governance : Privacy, Monitoring and Research".

이런 현실과 달리 건강정보는 민감정보로서 법령에서 민감정보의 수집이나 활용을 요구 혹은 허용하거나, 당사자의 명시적 동의가 있어야 수집이나 활용이 가능한 것인데 그에 대한 법률 규정은 찾기 어렵고, 개인정보보호원칙이 적용되어 관리되고 있다거나 충분한 거버넌스 구조를 갖추고 있다고 보기도 어렵다.

건강정보를 공익적 목적의 연구에 활용하는 것이나, 부득이한 경우 당사자의 동의가 없어도 개인정보를 활용한 공익적 목적의 연구가 필요하다는 점은 인정되는데 현재 법령에 특별한 규정이 없다. 따라서 공익적 목적의 연구와 관련해서는 반드시 필요한 경우에는 부득이하게 해당 개인의 동의를 받을 수 없는 경우에도 건강정보를 활용하여 연구할 수 있는 규정과 연구와 관련한 개인정보의 활용과 관련한 안전조치 등을 법제화할 필요가 있다.

그런데 현실에서는 통일적인 기준 없이 위법의 소지가 있는 방식으로 건강정보가 연구에 활용되고 있어서 시급하게 대책을 마련할 필요가 있다. 보건복지부는 보건의료 빅데이터 추진전략(2018-2022)을 마련하고 있는데, 그중 보건의료 빅데이터 활용 서비스 개발과 보건의료 빅데이터 개방·연계 강화가 핵심 과제를 이루고 있다. 이와 관련해서도 현재의 실태를 점검하고, 제도적 방안을 마련하는 것이 급선무이다.

이하에서는 우리나라의 건강정보에 대한 법적 규율을 개관하고 의료기관이나 약국을 넘는 건강정보의 집적과 집적된 건강정보의 2차적 가공, 연계, 이용에 관한 법제도 및 의료 연구에 관한 법 제도와 현황을 살펴본다. 다음으로 그동안 한국보건사회연구원, 보건복지부 등이 추진해 온 데이터 연계·결합의 제도 개선 방안과 보건복지부가 추진하고 있는 보건의료 빅데이터 추진전략을 평가하고, 국내외 제도에 대한 검토를 통해서 얻은 시사점을 바탕으로 법, 제도, 절차, 관행 등의 데이터 거버넌스의 개선방안을 제안한다.

2. 건강정보 수집·이용 관련 법제

(1) 개요

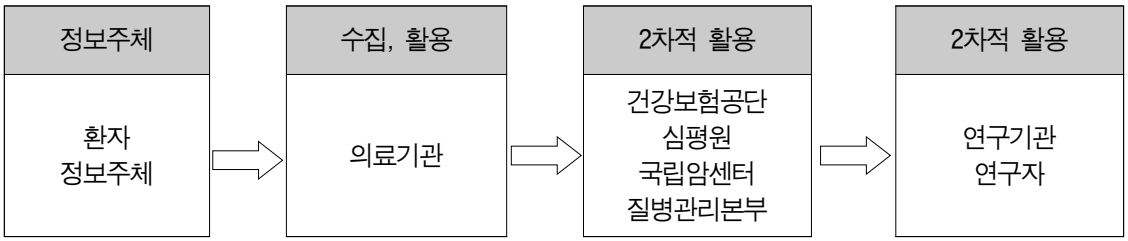
개인정보에 해당하는 건강정보의 처리에 대해 규율하는 일반법으로는 개인정보보호법이 있다. 그리고 개별법으로는 부문별로 의료법, 약사법, 정신보건법 등이, 의료 관련 사회보장제도의 구성과 운영에 대한 법률인 국민건강보험법, 의료급여법, 산업재해보상보험법 등이 있는데, 이들 법률에는 건강정보의 처리와 관련한 규정들이 있다. 그리고 연구와 관련한 법령으로 생명윤리 및 안전에 관한 법률, 생명연구자원의 확보·관리 및 활용에 관한 법률 등이 있다.[210]

이들 법령을 분류해 보면, 모든 영역에 적용되는 일반법과 개인과 의료기관, 약국, 정신보건기관 등과의 관계에서 개인정보인 의료정보의 수집·이용을 규율하는 것과 사회보험 등의 목적으로 의료정보가 제3자에게 제공되어 활용되는 것에 대해서 규율하는 것이 있다. 특히 연구 목적의 개인정보 활용에 대해서는 특별한 규율이 적용된다.

표 4-5 건강정보 관련 법제의 현황과 개요

구분	법률	개요
일반법	개인정보보호법	개인정보보호에 관한 규율
의약기관, 의료인/약사	의료법 약사법 정신보건법 등	의사, 치과의사, 간호사, 조산사 등 약사, 한약사 등 정신보건기관
의료 관련 사회보장제도	국민건강보험법 의료급여법 노인장기요양보험법	국민건강보험 의료급여 노인장기요양보험
연구에 관한 규율	생명윤리법	연구에 대한 규율
기타	전염병관리법 암관리법 건강검진기본법	건강정보수집에 대한 규정

표 4-6 건강정보의 흐름



(2) 개인정보보호법의 건강정보 수집·처리에 대한 규율

개인을 식별할 수 있는 정보의 처리에 대해서는 개인정보보호법이 적용된다. 개인정보보호법은 개인의 건강에 관한 정보를 사상·신념, 노동조합·정당의 가입·탈퇴, 정치적 견해, 성생활 등에 관한 정보와 함께 정보주체의 사생활을 현저히 침해할 우려가 있는 정보인 '민감정보'로 특별하게 규율하고 있다(제23조 제1항).[211] 덧붙여 유전

210) 그 밖에도 공공보건의료에 관한 법률, 군보건의료법, 보건의료기본법, 농어촌 보건의료 특별법, 보건의료기술진흥법, 응급의료법, 의료기기법, 장애인건강권 및 의료접근성 보장에 관한 법률, 지방의료원법, 국립중앙의료원법, 감염병의 예방 및 관리에 관한 법률, 가축전염병 예방법, 건강검진기본법, 국민건강증진법, 정신건강증진 및 정신질환자 복지서비스 지원에 관한 법률 등이 있다. 이들 법률은 건강정보의 수집, 이용에 대한 근거 규정이 될 수 있다. 의료 연구와 관련해서도 제대혈 관리 및 연구에 관한 법률, 질병관리본부 시험의뢰규칙, 천연물신약 연구개발 촉진법 등도 있다.

211) 법제처 법령해석도 개인정보보호법의 규정 체계를 고려할 때, 민감정보의 처리에 관한

자검사 등의 결과로 얻어진 유전정보도 민감정보로 규정하고 있다(시행령 제18조).[212]

개인정보 중 민감정보에 해당하는 건강정보의 수집, 이용, 제공, 보관, 연계나 결합 등에 대해서는 개인정보보호법이 적용되므로 개인정보보호법 제23조에 따른다. 즉, '민감정보의 처리는 원칙적으로 금지되고 ① 정보주체에게 고지할 사항을 알리고 다른 개인정보의 처리에 대한 동의와 별도로 동의를 받은 경우이거나, ② 법령에서 민감정보의 처리를 요구하거나 허용하는 경우에만 처리가 허용된다'(제23조 제1항 제1호, 제2호).

여기에서 '법령에서 민감정보의 처리를 요구하거나 허용하는 경우'의 의미를 어떻게 볼 것인지가 문제인데, 이와 관련하여 법제처는 '법령에서 민감정보의 처리가 필요한 사무와 민감정보의 종류를 명시적으로 열거하고 민감정보의 처리를 요구하거나 허용하는 경우로 제한되는 것'(법제처 2014. 9. 5. 회신 14-0440 해석례)이라고 보고 있다. 즉, 법령에서 민감정보의 처리가 필요한 사무와 민감정보의 종류를 명시적으로 열거하고 민감정보의 처리를 요구하거나 허용하는 경우라야 개인정보보호법 제23조 제2호에 따라 민감정보를 처리할 수 있는 것으로 볼 수 있다는 것이다(16-0022회신일자 2016-04-14). 개인정보보호법 해설서(2016년, 행정자치부)도 이를 '법령에서 민감정보의 종류를 열거하고 그 처리를 요구하고 있는 경우(법정 서식에 민감정보 기재사항이 있는 경우도 포함)'로 매우 엄격하게 보고 있다.[213] 이와 같은 해석에 의하면, 최소한 법령에서 민감정보의 종류가 열거되지 않은 경우는 민감정보 처리가 불가능하다고 볼 것이다. 법령의 규정으로는 예를 들어 의료법은 응급상황이나, 환자가 의식이 없어서 동의할 수 없는 상황에서 의료인이나 의료기관의 장이 환자의 동의 없이도 다른 의료인이나 의료기관의 장에게 환자의 건강정보를 전송할 수 있다는 규정 등이 있다.

이처럼 민감정보인 경우에는 일반적인 개인정보에 적용되는 예외 규정들(정보주체의 동의를 받지 않고도 개인정보의 처리가 가능하게 하는 여러 가지 예외규정)인 개인정보보호법 제15조 제1항 제2호 ~ 제6호 즉, 계약의 체결이나 이행을 위해서 불가피하게 필요한 경우나, 정보주체의 정당한 이익을 보호하기 위하여 필요한 경우라는 등의 이유로는 정보주체의 동의 없이 건강정보 처리가 허용되지 않는다.

한편 개인정보보호법은 개인정보처리자가 민감정보를 처리하는 경우에는 그 민감정보가 분실·도난·유출·위조·변조 또는 훼손되지 아니하도록 개인정보보호법 제29조에

개인정보보호법 제23조는 개인정보 수집·이용에 관한 규정인 같은 법 제15조 제1항 등에 대한 특례규정으로 보아야 한다고 한다(법제처 2014. 9. 5. 회신 14-0440 해석례 등).

212) 단, 유전정보는 공공기관이 개인정보를 목적 외의 용도로 이용하거나 이를 제3자에게 제공하지 아니하면 다른 법률에서 정하는 소관 업무를 수행할 수 없는 경우로서 보호위원회의 심의·의결을 거친 경우에는 민감정보로 보지 않는다(시행령 제18조).

213) 예를 들어 총포·도검·화약류 등 단속법 시행규칙 [별지 제10호의3서식]의 병력(病歷) 신고 및 개인정보 이용 동의서에 의한 정신분열증, 정동장애 등의 치료 사실 여부의 확인.

따른 안전성 확보에 필요한 조치를 하도록 하고 있다(제23조 제2항).

그리고 민감정보의 수집·처리 시에는 당사자의 동의를 받았거나 법률의 규정이 있다고 하더라도 최소수집의 원칙, 익명처리의 원칙 등 개인정보 보호 원칙과, 정보를 제공받을 권리, 접근권, 열람 및 정정, 삭제, 처리중지 요구권, 철회권 등의 정보주체의 권리는 보장되어야 하고, 침해 시에는 권리 구제를 받을 수 있다.

(3) 환자와 의료기관 사이의 건강정보 수집·이용 및 데이터 연계

가. 개요

의료기관, 약국, 정신보건기관 등은 환자로부터 치료, 조제 등의 과정에서 방대한 양의 건강정보를 수집하고, 처리한다. 이 정보들은 아래에서 보는 것처럼 국민건강보험공단, 심사평가원, 국립암센터 등을 통하여 수집되어 2차적 활용이 되기도 하고, 데이터 연계 등이 이루어지기도 하므로, 이와 같은 개인정보의 2차적 활용의 기원이 되고 있다. 따라서 기본적으로 1차적 수집, 활용의 적법성이 담보되어야 한다.

이와 관련하여 개인정보보호법은 민감정보의 처리에 대해서 강화된 규율을 두고 있으며, 별도로 의료법, 약사법 등은 의료기관, 약국, 정신보건기관 등이 건강정보의 비밀보호 및 안전조치를 하도록 특별한 규정을 두고 있다.

나. 의료기관 등의 법령상 기록의무 및 보존의무

의료기관, 약국, 정신보건기관 등의 경우 환자로부터 건강정보를 수집하는 것과 관련해서는 법령상 기록 및 보존의무가 부과되어 있다. 이 경우는 정보주체의 동의를 받지 않고도 해당 건강정보를 기록하고 보존할 수 있다.

의료법, 약사법, 정신보건법(정신건강증진 및 정신질환자 복지서비스 지원에 관한 법률) 등은 의료기관이나 의료인, 약사, 한약사가 의료행위나 조제 과정에서 환자에 관한 일정한 건강정보를 의무적으로 기록하고 보관하도록 의무를 부여하고 있다(의료법 제22조, 약사법 제30조, 정신보건법 제30조).[214]

예를 들어 의료법은 의료인에게 진료기록부, 조산기록부, 간호기록부, 그 밖의 진료에 관한 기록(이하 '진료기록부 등')을 갖추어 두고 환자의 주된 증상, 진단 및 치료내

214) 의료법 등이 진료기록부 등의 작성, 보존의무를 규정하고 있는 취지는 (i) 진료를 담당하는 의사 자신으로 하여금 환자의 상태와 치료의 경과에 관한 정보를 빠뜨리지 않고 정확하게 기록, 보존하여 이를 그 이후 계속되는 환자치료에 이용하도록 함과 아울러 (ii) 다른 의료관련 종사자들에게도 그 정보를 제공하여 환자로 하여금 적정한 의료를 제공받을 수 있도록 하고, (iii) 의료행위가 종료된 이후에는 그 의료행위의 적정성을 판단하는 자료로 사용할 수 있도록 하고자 함이다(대법원 1998. 1. 23. 선고 97도2124 판결).

용 등 의료행위에 관한 사항과 의견을 상세히 기록하고 서명하도록 하고 이를 일정한 기간 보존하도록 하고 있다(의료법 제22조). 의료법은 진료기록부 등에 기록할 사항을 보건복지부령으로 정해 놓았는데, 예를 들어 진료기록부에는 진료를 받은 사람의 주소·성명·연락처·주민등록번호 등 인적사항, 주된 증상, 진단결과 또는 진단명, 치료내용(주사·투약·처치 등), 진료 일시(日時) 등을 기재하도록 하고 있다.[215] 의료법은 진료기록부 등을 보존하지 않은 의료인에 대하여 자격정지(의료법 제66조), 형벌(의료법 제90조) 등의 제재를 가할 수 있도록 규정하고 있다.

이와 같이 작성과 보관의무의 대상이 되는 것은 정신보건법에 의한 진료기록부[216], 의료법에 의한 처방전[217], 조산기록부[218], 간호기록부[219]와 약사법에 의한 조제기록부[220] 등이다. 이와 같은 규정들이 개인정보보호법에서 민감정보 처리의 허용규정인 법령에서 민감정보의 처리를 허용한 규정으로 볼 것이라는 점은 분명한데, 수집하는 민감정보의 항목과 내용을 얼마나 구체적으로 규율한 것인지 등에 대해서는 해석의 여지가 있다. 한편 그 외의 의무기록은 특별한 양식을 법률로 규정하고 있지는 않다. 이 경우 의료기관, 의사, 약사가 건강정보를 당사자로부터 치료 목적을 위하여 수집하고 기록하고 보관하려면 아래에서 보는 바와 같이 당사자로부터 명시적인 동의를 받아야 할 것이다.

다. 의료기관에서 정보주체의 명시적 동의에 의한 수집·처리

개인정보보호법이 건강정보의 처리를 위해서는 별도의 명시적 동의가 필요하다고

---

215) 의료법 시행규칙 제14조 1. 진료기록부 - 진료를 받은 사람의 주소·성명·연락처·주민등록번호 등 인적사항, 주된 증상. 이 경우 의사가 필요하다고 인정하면 주된 증상과 관련한 병력(病歷)·가족력(家族歷)을 추가로 기록할 수 있다. 진단결과 또는 진단명, 진료경과(외래환자는 재진 환자로서 증상·상태, 치료내용이 변동되어 의사가 그 변동을 기록할 필요가 있다고 인정하는 환자만 해당한다), 치료 내용(주사·투약·처치 등), 진료 일시(日時).

216) 입원 등 당시의 대면 진단 내용, 퇴원 등의 의사 확인, 퇴원 등의 신청 일시 및 퇴원 등의 거부 사유, 입원 등의 기간 연장에 대한 심사 청구 및 결과, 투약 등의 치료내용을 적은 진료기록, 특수치료에 관한 협의체의 회의 내용, 통신과 면회의 자유 제한의 사유 및 내용, 격리시키거나 묶는 등의 신체적 제한의 사유 및 내용, 작업요법의 내용 및 결과, 그 밖에 보건복지부령으로 정하는 사항.

217) 의료법 시행규칙 제12조(처방전의 기재사항 등) - 환자의 성명 및 주민등록번호, 의료기관의 명칭, 전화번호 및 팩스 번호, 질병 분류기호, 의료인의 성명·면허종류 및 번호, 처방 의약품의 명칭(일반명칭, 제품명이나 '약사법' 제51조에 따른 대한민국약전에서 정한 명칭을 말한다)·분량·용법 및 용량, 처방전 발급 연월일 및 사용 기간, 의약품 조제 시 참고 사항.

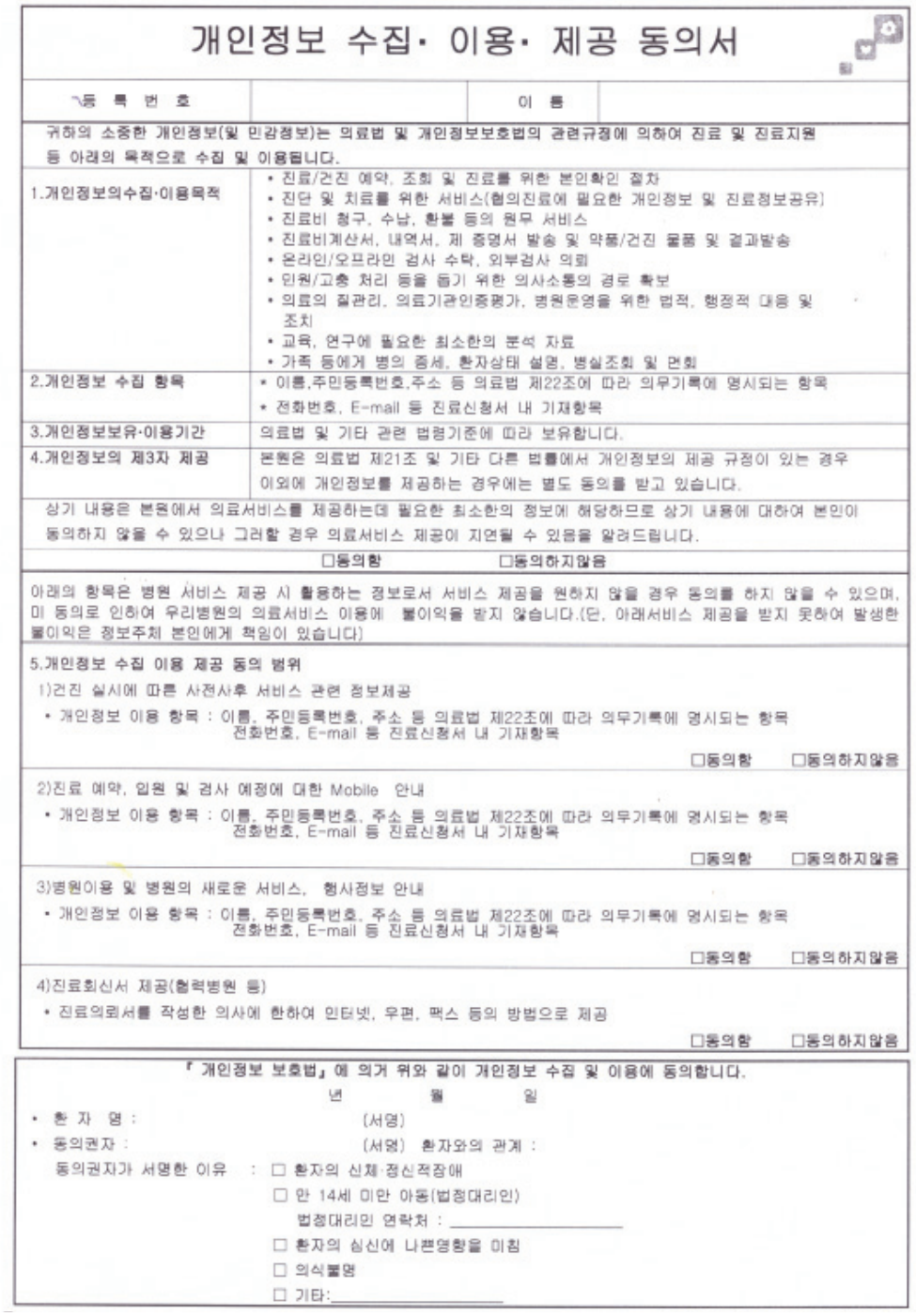
218) 의료법 시행규칙 제14조 2. 조산기록부 - 조산을 받은 자의 주소·성명·연락처·주민등록번호 등 인적사항, 생·사산별(生·死産別) 분만 횟수, 임신 후의 경과와 그에 대한 소견, 임신 중 의사에 의한 건강진단의 유무(결핵·성병에 관한 검사를 포함한다), 분만 장소 및 분만 연월일시분(年月日時分), 분만의 경과 및 그 처치, 산아(産兒) 수와 그 성별 및 생·사의 구별, 산아와 태아부속물에 대한 소견, 산후의 의사의 건강진단 유무.

219) 의료법 시행규칙 제14조 3. 간호기록부 - 간호를 받는 사람의 성명, 체온·맥박·호흡·혈압에 관한 사항, 투약에 관한 사항, 섭취 및 배설물에 관한 사항, 처치와 간호에 관한 사항, 간호 일시(日時).

220) 환자의 인적사항, 조제 연월일, 처방 약품명과 일수, 조제 내용 및 복약지도 내용.

규정하고 있으므로 의료기관이나 의료인이 환자로부터 건강정보를 수집하기 위해서는 법령상 건강정보 처리에 관한 규정이 있지 않는 한 정보주체의 명시적 동의가 필요하다. 의료기관이나 의료인들은 일반적으로 아래와 같은 개인정보 수집·이용·제공 동의서를 작성하여 환자 본인의 동의를 받고 있다.

그림 4-3 의료기관 개인정보 수집・이용・제공 동의서 예시



개인정보 수집· 이용· 제공 동의서

등 록 번 호		이 름	

귀하의 소중한 개인정보(및 민감정보)는 의료법 및 개인정보보호법의 관련규정에 의하여 진료 및 진료지원 등 아래의 목적으로 수집 및 이용됩니다.

1.개인정보의수집·이용목적	• 진료/건진 예약, 조회 및 진료를 위한 본인확인 절차 • 진단 및 치료를 위한 서비스(협의진료에 필요한 개인정보 및 진료정보공유) • 진료비 청구, 수납, 환불 등의 원무 서비스 • 진료비계산서, 내역서, 제 증명서 발송 및 약품/건진 물품 및 결과발송 • 온라인/오프라인 검사 수탁, 외부검사 의뢰 • 민원/고충 처리 등을 돕기 위한 의사소통의 경로 확보 • 의료의 질관리, 의료기관인증평가, 병원운영을 위한 법적, 행정적 대응 및 조치 • 교육, 연구에 필요한 최소한의 분석 자료 • 가족 등에게 병의 증세, 환자상태 설명, 병실조회 및 면회
2.개인정보 수집 항목	• 이름,주민등록번호,주소 등 의료법 제22조에 따라 의무기록에 명시되는 항목 • 전화번호, E-mail 등 진료신청서 내 기재항목
3.개인정보보유·이용기간	의료법 및 기타 관련 법령기준에 따라 보유합니다.
4.개인정보의 제3자 제공	본원은 의료법 제21조 및 기타 다른 법률에서 개인정보의 제공 규정이 있는 경우 이외에 개인정보를 제공하는 경우에는 별도 동의를 받고 있습니다.

상기 내용은 본원에서 의료서비스를 제공하는데 필요한 최소한의 정보에 해당하므로 상기 내용에 대하여 본인이 동의하지 않을 수 있으나 그러할 경우 의료서비스 제공이 지연될 수 있음을 알려드립니다.

□동의함 □동의하지않음

아래의 항목은 병원 서비스 제공 시 활용하는 정보로서 서비스 제공을 원하지 않을 경우 동의를 하지 않을 수 있으며, 미 동의로 인하여 우리병원의 의료서비스 이용에 불이익을 받지 않습니다.(단, 아래서비스 제공을 받지 못하여 발생한 불이익은 정보주체 본인에게 책임이 있습니다)

5.개인정보 수집 이용 제공 동의 범위

1)건진 실시에 따른 사전사후 서비스 관련 정보제공

• 개인정보 이용 항목 : 이름, 주민등록번호, 주소 등 의료법 제22조에 따라 의무기록에 명시되는 항목 전화번호, E-mail 등 진료신청서 내 기재항목

□동의함 □동의하지않음

2)진료 예약, 입원 및 검사 예정에 대한 Mobile 안내

• 개인정보 이용 항목 : 이름, 주민등록번호, 주소 등 의료법 제22조에 따라 의무기록에 명시되는 항목 전화번호, E-mail 등 진료신청서 내 기재항목

□동의함 □동의하지않음

3)병원이용 및 병원의 새로운 서비스, 행사정보 안내

• 개인정보 이용 항목 : 이름, 주민등록번호, 주소 등 의료법 제22조에 따라 의무기록에 명시되는 항목 전화번호, E-mail 등 진료신청서 내 기재항목

□동의함 □동의하지않음

4)진료회신서 제공(협력병원 등)

• 진료의뢰서를 작성한 의사에 한하여 인터넷, 우편, 팩스 등의 방법으로 제공

□동의함 □동의하지않음

『 개인정보 보호법』에 의거 위와 같이 개인정보 수집 및 이용에 동의합니다.

년 월 일

• 환 자 명 : (서명)

• 동의권자 : (서명) 환자와의 관계 :

동의권자가 서명한 이유 : □ 환자의 신체·정신적장애
□ 만 14세 미만 아동(법정대리인)
법정대리인 연락처 : _____
□ 환자의 심신에 나쁜영향을 미침
□ 의식불명
□ 기타:_____

라. 비밀준수 의무와 제3자 제공

의료법(제19조[221], 제69조[222]), 약사법(제68조의9[223], 제87조[224]), 정신보건법(제71조)[225] 등은 의료인, 약사 등에 대해서 엄격한 비밀준수 의무를 부과하고 있다. 아울러 의료인, 의료기관의 경우는 환자에 관한 기록, 조제기록부의 정보를 다른 사람에게 열람하게 하거나, 사본을 내주는 등 내용을 확인할 수 있게 하는 행위를 법률에 명시된 경우 외에는 금지하고 있다(의료법 제21조 제2항, 제3항, 약사법 제30조 제3항). 형법도 업무상비밀누설죄(제371조 제1항)로 직무상 알게 된 비밀의 누설이나 공표를 엄격하게 규율하고 있다.[226]

한편 의료법과 약사법은 환자 정보를 당사자의 동의가 없어도 제3자에게 제공할 수 있는 예외를 한정적으로 열거하고 있는데, 아래와 같다.

---

221) 제19조(정보 누설 금지) ①의료인이나 의료기관 종사자는 이 법이나 다른 법령에 특별히 규정된 경우 외에는 의료·조산 또는 간호업무나 제17조에 따른 진단서·검안서·증명서 작성·교부 업무, 제18조에 따른 처방전 작성·교부 업무, 제21조에 따른 진료기록 열람·사본 교부 업무, 제22조 제2항에 따른 진료기록부등 보존 업무 및 제23조에 따른 전자의무기록 작성·보관·관리 업무를 하면서 알게 된 다른 사람의 정보를 누설하거나 발표하지 못한다. <개정 2016.5.29.>
② 제58조제2항에 따라 의료기관 인증에 관한 업무에 종사하는 자 또는 종사하였던 자는 그 업무를 하면서 알게 된 정보를 다른 사람에게 누설하거나 부당한 목적으로 사용하여서는 아니 된다.

222) 제69조(의료지도원) ③의료지도원 및 그 밖의 공무원은 직무를 통하여 알게 된 의료기관, 의료인, 환자의 비밀을 누설하지 못한다.

223) 제68조의9(비밀유지의무) 의약품안전관리원의 임원이나 직원 또는 그 직에 있었던 자는 직무상 알게 된 비밀을 누설하여서는 아니 된다.

224) 제87조(비밀 누설 금지) ①약사·한약사는 이 법 또는 다른 법령에 규정된 경우 외에는 의약품을 조제·판매하면서 알게 된 타인의 비밀을 누설하여서는 아니 된다. <개정 2007.10.17.>
② 제47조의3제2항에 따라 의약품 품목허가를 받은 자·수입자 및 의약품 도매상 등의 영업에 관한 비밀을 업무상 알게 된 자는 그 비밀을 타인에게 누설하거나 업무목적 외의 용도로 사용하여서는 아니 된다.

225) 제71조(비밀누설의 금지) 정신질환자 또는 정신건강증진시설과 관련된 직무를 수행하고 있거나 수행하였던 사람은 그 직무의 수행과 관련하여 알게 된 다른 사람의 비밀을 누설하거나 공표하여서는 아니 된다.

226) 제317조(업무상비밀누설) ①의사, 한의사, 치과의사, 약제사, 약종상, 조산사, 변호사, 변리사, 공인회계사, 공증인, 대서업자나 그 직무상 보조자 또는 차등의 직에 있던 자가 그 직무처리중 지득한 타인의 비밀을 누설한 때에는 3년 이하의 징역이나 금고, 10년 이하의 자격정지 또는 700만원 이하의 벌금에 처한다.

표 4-7 의료법에서 환자의 동의 없이 제3자 제공 가능한 경우

요건	제3자
급여비용 심사·지급·대상 여부 확인·사후관리 및 요양급여의 적정성 평가·가감지급 등을 위하여	국민건강보험공단 또는 건강보험심사평가원
의료급여 수급권자 확인, 급여비용의 심사·지급, 사후관리 등 의료급여 업무를 위하여	보장기관(시·군·구), 국민건강보험공단, 건강보험심사평가원
보험급여를 받는 근로자를 진료한 산재보험 의료기관(의사를 포함한다)에 대하여 그 근로자의 진료에 관한 보고 또는 서류 등 제출을 요구하거나 조사하는 경우	근로복지공단
의료기관으로부터 자동차보험진료수가를 청구받은 보험회사 등이 그 의료기관에 대하여 관계 진료기록의 열람을 청구한 경우	의료기관으로부터 자동차보험진료수가를 청구받은 보험회사 등
병역판정검사와 관련하여 질병 또는 심신장애의 확인을 위하여 필요하다고 인정하여 의료기관의 장에게 병역판정검사대상자의 진료기록·치료 관련 기록의 제출을 요구한 경우	지방병무청장
부양가족연금, 장애연금 및 유족연금 급여의 지급심사와 관련하여 가입자 또는 가입자였던 사람을 진료한 의료기관에 해당 진료에 관한 사항의 열람 또는 사본 교부를 요청하는 경우	국민연금공단
공무상 요양비, 재해부조금, 장해급여 및 유족급여의 지급심사와 관련하여 공무원 또는 공무원이었던 자를 진료한 의료기관에 해당 진료에 관한 사항의 열람 또는 사본 교부를 요청하는 경우	공무원연금공단
장애 정도에 관한 심사와 관련하여 장애인 등록을 신청한 사람 및 장애인으로 등록한 사람을 진료한 의료기관에 해당 진료에 관한 사항의 열람 또는 사본 교부를 요청하는 경우	공공기관의 장
감염병의 역학조사 및 예방접종에 관한 역학조사를 위하여 필요하다고 인정하여 의료기관의 장에게 감염병 환자 등의 진료기록 및 예방접종을 받은 사람의 예방접종 후 이상반응에 관한 진료기록의 제출을 요청하는 경우	보건복지부장관, 질병관리본부장, 시·도지사 또는 시장·군수·구청장

표 4-8 약사법에서 환자의 동의 없이 제3자 제공 가능한 경우

요건	제3자
급여비용 심사·지급·대상 여부 확인·사후관리 및 요양급여의 적정성 평가·가감지급 등을 위하여	국민건강보험공단 또는 건강보험심사평가원
의료급여 수급권자 확인, 급여비용의 심사·지급, 사후관리 등 의료급여 업무를 위하여	보장기관(시·군·구), 국민건강보험공단, 건강보험심사평가원
형사소송법의 영장	수사기관
민사소송법의 문서제출명령	법원

그중 가장 대표적인 것이 국민건강보험, 의료급여의 비용심사, 지급 대상 여부 확인, 사후관리, 적정성 평가, 가감지급을 위하여 건강보험공단, 건강보험심사평가원으로 제공하는 경우이다.

마. 의료기관 등에서 연구 목적 개인정보 활용

의료인이나 의료기관은 의료나 의학에 관한 연구를 할 수 있다. 의료법은 의료법인이 수행할 수 있는 부대사업으로 의료나 의학에 관한 조사, 연구를 들고 있기도 하다(의료법 제49조 제1항 제2호). 의료인이나 의료기관이 환자의 건강정보를 활용하여 연구하고자 할 경우에는 그에 대한 명확한 동의를 얻어야 한다. 아울러 개인정보가 식별될 수 있는 경우 등 필요한 경우에는 생명윤리법에 의한 심의를 거쳐야 한다. 생명윤리법에 의한 심의에 대해서는 별도의 장으로 서술한다.

연구 목적으로 개인정보를 활용하면서 데이터의 연계·결합을 하려고 하는 경우에도 마찬가지이다. 개인정보주체에게 데이터의 연계·결합에 대해서 어떤 정보를 연계·결합하려고 하는 것인지, 그 목적은 무엇인지 등에 대해서 명확하게 고지하고, 분명한 동의를 얻어야 한다.

한편, 학회의 세미나 등 외부에 발표하는 연구뿐 만 아니라 의료기관 내부에서 세미나를 하는 경우에도 원칙적으로 해당 개인정보주체로부터 명시적인 동의를 얻어야 한다. 연구에 활용하는 경우에도 그 자체의 정보뿐만 아니라 다른 정보와 결합하여도 개인을 식별하는 것이 불가능한 경우에는 해당 정보를 개인정보로 볼 수 없을 것이므로 정보주체로부터 동의를 얻을 필요는 없다.

바. 개선 방향

이와 관련하여 다음과 같은 제도 개선이 필요하다.

1. 건강정보의 수집과 관련한 현행 의료법, 약사법, 정신보건법 등의 규정에서 민감정보인 건강정보의 수집 근거를 명확하게 하는 것이 해석의 논란을 없앨 수 있을 것이다.

2. 개별 의료기관 등에서 민감정보를 수집하는 것과 관련한 동의서는 민감정보 수집을 별도로 분명하게 고지하고, 수집항목이나, 수집된 항목의 보유 기간, 제3자 제공 등에 대해서도 분명하게 명기하고 그에 대한 동의를 얻을 수 있도록 표준 동의서 등을 작성하여 활용하도록 할 필요가 있다.

3. 개별 의료기관에서 건강보험 등과 관련하여 국민건강보험공단이나, 심사평가원에 심사 청구를 위해서 정보가 제공되는 것에 대하여 제공되는 정보의 내용, 제공되는 목적, 보유 기간 등에 대해서 구체적으로 고지할 수 있도록 하는 것이 바람직하다.

4. 의료기관 내부에서 내부 세미나 등의 목적으로 환자의 건강정보를 활용하는 것과 관련해서는 동의서 등을 분명하게 마련하여 활용할 수 있도록 하는 것이 바람직할 것이다.

5. 데이터의 연계와 관련해서도 환자의 진료기록과 환자의 기타 정보를 연계하는 것에 대해서는 이를 명시할 필요가 있다.

(4) 건강정보의 2차적 수집 및 활용

가. 개별 의료기관 외부의 건강정보 수집·활용

우리나라의 경우 개별 의료기관이나 약국 외부에서 국민의 건강 관련 정보가 막대하게 수집되어 활용되고 있다. 이는 대체로 건강정보의 2차적 활용이라고 볼 수 있다. 현재 개별 의료기관이나 약국 등의 외부에서 수집·활용되고 있는 국민의 건강 관련 정보들을 유형별로 분류해 보면 다음과 같다.

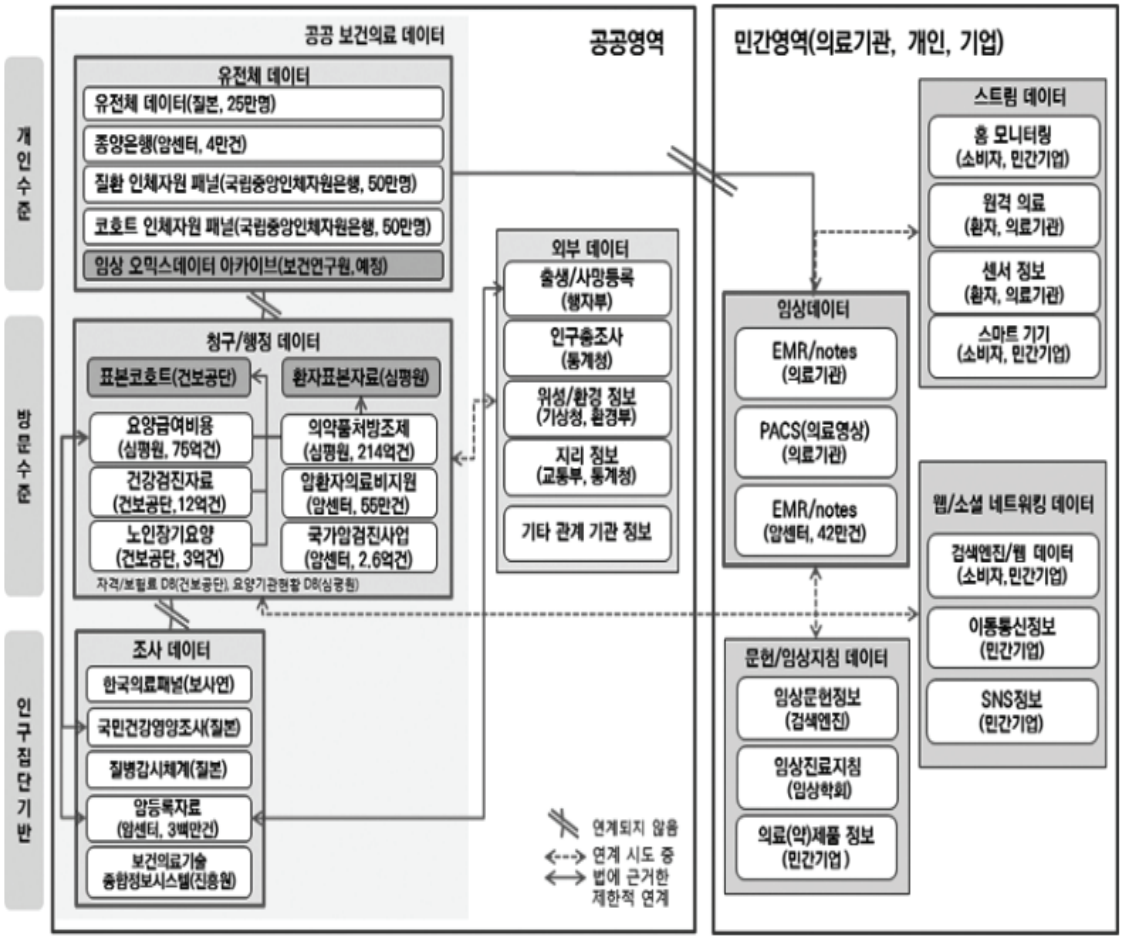
첫째, 국민건강보험공단, 건강보험심사평가원 등이 보유하고 있는 건강정보. 이는 사회보험의 운영을 위해 수집된 정보들인데, 추가로 이를 연계하거나 가공하여 생성한 정보들이 있다.

둘째, 보건의료 통계나 조사자료. 이는 통계법이나 개별법에 의하여 보건의료와 관련한 정책의 수립, 운용을 위해 수집·활용하고 있는 것들이다.

셋째, 질병관리본부의 연구 목적 건강정보. 질병관리본부는 대규모의 코호트 사업을 통해서 방대한 양의 코호트 정보를 축적해 놓고 있고 인체 유래물 정보 등도 집적하고 있다. 이는 연구 목적으로 수집·활용되는 것들이다.

넷째, 암 등록정보, 전염병 정보, 건강검진 정보 등. 암관리법이나 전염병관리법, 건강검진기본법 등에 의해 법정 의무 수집대상 등으로 지정하여 수집·활용하는 정보이다.

그림 4-4 우리나라 보건의료 데이터와 연계 현황



* 출처: 강희정 외 (2015). 보건의료 빅데이터 활용을 위한 기본계획 수립 연구. 보건복지부, 한국보건사회연구원. p.356.

표 4-9 보건의료 주요 빅데이터 현황

기관	건강정보
국민건강보험공단	요양급여내역 DB, 전 국민 건강정보 DB, 건강검진 DB, 코호트 자료 DB
건강보험심사평가원	요양급여비용 청구명세서 DB 요양기관현황DB 의약품안전사용정보(DUR) DB 의약품유통정보 DB 병원평가정보 DB
질병관리본부 (국립보건연구원 포함)	국민건강영양조사, 지역사회건강조사, 청소년건강행태온라인조사 등 임상오믹스 데이터 아카이브 임상오믹스 포스트게놈 다부처 유전체 사업 DB 등 유전체 정보 집적
국립암센터	암 등록 통계 검진자코호트 DB 국가암검진사업 정보시스템 DB 암 환자 의료비지원 정보시스템 DB 말기 암 DB EMR DB 종양은행 DB

표 4-10 보건복지 분야 국가승인통계현황

통계 종류	분야	통계 명칭
조사 (42종)	보건 (23)	건강보험환자 진료비 실태조사, 국민건강영양조사, 국민보건의료실태조사, 근로환경조사, 병원경영실태조사, 아동구강건강 실태조사, 의료기관별 급여 적정성 평가 현황, 의료기기 제조/유통 조사, 의약품/의료기기연구개발실태조사, 작업환경실태조사, 전국민장내기생충감염실태조사, 전국출산력 및 가족보건복지실태조사, 정신질환실태조사, 지역사회건강조사, 청소년건강행태 온라인조사, 퇴원손상심층조사, 한국의료패널조사, 한국인인체치수조사, 한방의료이용 및 한약소비실태조사, 한의약산업실태조사, 화장품제조유통조사, 환자조사
	복지 (19)	가정폭력실태조사, 고령화연구패널조사, 국민노후보장패널조사, 기업 및 공공기관의 가족친화수준조사, 남해군 노인실태조사, 노인실태조사, 노후준비실태조사, 보육실태조사, 복지욕구조사, 사회서비스수요/공급실태조사, 생명보험성향조사, 서울특별시 복지실태조사, 성폭력실태조사, 아동종합실태조사, 장애인생활체육실태조사, 장애인실태조사, 장애인편의시설설치현황조사, 한국복지패널조사, 한부모가족실태조사
보고 (34종)	보건 (21)	HIV/AIDS 신고현황, 건강검진통계, 건강보험 주요 수술통계, 건강보험통계, 결핵현황, 공중위생관계업소 실태보고, 근로자건강진단실시상황보고, 급성심장정지조사, 노인장기요양보험통계, 법정감염병발생보고, 보건소 및 보건지소운영현황, 수입식품현황, 식품 및 식품첨가물 생산실적, 식품수거검사실적, 암등록통계, 완제의약품 유통정보통계, 응급의료현황통계, 의료기기생산실적, 전국예방접종률조사, 지역별 의료이용통계, 학생건강검사통계보고
	복지 (13)	가정위탁 국내 입양소년소녀 가장현황, 국민기초생활보장 수급자현황, 국민연금통계, 노인복지시설현황, 노인학대현황, 보훈보상금지급현황, 산업재해현황, 산재보험통계, 아동복지시설 보호아동 및 종사자현황보고, 어린이집 및 이용자통계, 요보호 아동현황보고, 장애인현황, 학대피해 아동보호현황
가공 (5종)	보건 (4)	국민의료비추계 및 국민보건계정, 사망원인통계, 어린이 식생활안전지수, 의약품소비량 및 판매액 통계
	복지 (1)	한국의 사회복지지출

그런데 환자로부터 의료기관이 건강정보를 수집하여 활용하는 것에 대해서는 근거 규정이 분명하고, 그에 대한 비밀보호, 정보주체의 동의를 받지 않고 활용할 수 있는 경우 등에 대해서 법령이 엄격하게 규정하고 있는 등 법 제도가 마련되어 있는 반면 의료기관으로부터 건강보험공단, 건강보험심사평가원 등이 수집하는 정보, 그 정보의 활용, 질병관리본부, 국립암센터 등이 수집하는 정보와 그 정보의 2차적 이용에 대해서는 법적 근거가 매우 미비한 상황이다. 특히 민감정보는 법령에서 구체적으로 민감정보의 종류를 명시하여 처리의 근거를 규정해야 처리가 가능하다는 해석에 비춰 본다면 현재의 상태는 '법적 근거가 없다'고 볼 수도 있는 상황이다.

나. 의료 사회보장제도 운영에 따른 건강정보 수집·이용

① 우리나라 사회보장제도 개요

우리나라 헌법 제34조 제1항 및 제2항은 국민의 인간다운 생활을 할 권리와 이를

실현하기 위한 국가의 사회복지 증진의무를 규정하고 있다. 이에 근거한 사회보장제도로 우리나라 사회보장기본법은 사회보험, 공공부조, 사회서비스를 실시하고 있다.

즉, '출산, 양육, 실업, 노령, 장애, 질병, 빈곤 및 사망 등의 사회적 위험으로부터 모든 국민을 보호하고 국민 삶의 질을 향상시키는 데 필요한 소득·서비스를 보장하는 제도'로 사회보험, 공공부조, 사회서비스'를 실시하고 있는 것이다.

그 중 사회보험은 질병・사망・노령・실업 기타 신체장애 등 국민에게 발생하는 사회적 위험을 보험방식에 의하여 대처함으로써 국민건강과 소득을 보장하는 제도이다. 사회보험은 운영과 방법론에서 보험기술과 보험원리를 따르고 있다는 점에서 공공부조와 차이가 있지만, 사회의 연대성과 강제성이 적용된다는 점에서 사보험과 다르다.

사회보험의 대상은 업무상의 재해·질병·분만·폐질(장애)·사망·유족·노령 및 실업 등인데, 업무상의 재해에 대해서는 산업재해보상보험, 질병과 부상에 대해서는 건강보험, 폐질·사망·노령 등에 대해서는 연금보험, 실업에 대해서는 고용보험제도가 있으며 이를 4대 사회보험이라 한다.

공공부조는 국가 및 지방자치단체의 책임하에 생활유지 능력이 없거나 생활이 어려운 국민의 최저생활을 보장하고 자립을 지원하는 제도를 말한다.

사회서비스는 국가·지방자치단체 및 민간부문의 도움이 필요한 모든 국민에게 복지, 보건의료, 교육, 고용, 주거, 문화, 환경 등의 분야에서 인간다운 생활을 보장하고 상담, 재활, 돌봄, 정보의 제공, 관련 시설의 이용, 역량 개발, 사회참여 지원 등을 통하여 국민의 삶의 질이 향상되도록 지원하는 제도이다.

표 4-11 우리나라의 사회보장제도

분류	해당 사회보장제도	
사회보험	- 국민연금[227] - 노인장기요양보험[228] - 산업재해보상보험[229]	- 건강보험[230] - 고용보험[231]
공공부조	- 생계급여[232] - 주거급여[233]	- 의료급여[234] - 교육급여
사회서비스	- 노인복지서비스[235] - 아동복지서비스[236][237] - 가정복지서비스[238]	- 장애인복지서비스[239] - 영유아복지서비스[240]

227) 국민의 노령·폐질 또는 사망에 대하여 연금급여를 실시함으로써 국민의 생활안정과 복지증진에 기여함을 목적으로 1973년 국민복지연금법으로 제정되어, 1986년 명칭이 변경되었다. 특수직역연금 도입 이후 일반 국민을 위한 연금이 도입됨으로 노후의 기본적 생활을 위한 토대가 마련되었다.

228) 고령이나 노인성 질병 등의 사유로 일상생활을 혼자서 수행하기 어려운 노인등에게 제공하는 신체활동 또는 가사활동 지원 등의 장기요양급여에 관한 사항을 규정하여 노후의 건강증진 및 생활안정을 도모하고 그 가족의 부담을 덜고자하는 취지로 2007년 4월 제정되어, 2008년 7월

② 국민건강보험, 의료급여, 노인장기요양보험, 산업재해보험

우리나라의 의료 관련 사회보장제도로는 국민건강보험, 의료급여, 노인장기요양보

시행되었다.

229) 1963년 11월에 제정된 이 법에 따른 산업재해보상보험제도는 우리나라 최초의 사회보험제도로서 산업재해를 입은 근로자에 대한 치료와 생활보장의 기능을 수행해왔다. '산업재해보상보험법'은 '근로기준법'의 재해보상제도를 기초로 하여 보험급여를 통한 재해보상제도를 꾸준히 발전시켜 왔다.

230) 국민의 질병·부상에 대한 예방·진단·치료·재활과 출산·사망 및 건강증진에 대하여 보험급여를 실시함으로써 국민건강을 향상시키고 사회보장을 증진함을 목적으로 하는 '국민건강보험법'이 국민건강보험제도를 구체화하였다. 이 법은 의료보험제도의 통합 운영에 따라 종전의 '의료보험법'과 '국민의료보험법'을 대체하여 1999년 제정되었다.

231) 실업의 예방, 고용의 촉진 및 근로자의 직업능력의 개발과 향상을 꾀하고, 국가의 직업지도와 직업소개 기능을 강화하며, 근로자가 실업한 경우에 생활에 필요한 급여를 실시하여 근로자의 생활안정과 구직 활동을 촉진함으로써 경제·사회 발전에 이바지하는 것을 목적으로 '고용보험법'은 1993년 12월에 제정되어 1995년 7월 1일부터 시행되고 있다.

232) 저소득 국민, 영세 도시빈민, 실업자등을 지원하여 빈곤문제에 대한 사회안전망의 기초를 튼튼히 하는 한편, 빈곤가구별로 자활지원계획을 수립하고 그에 맞는 자활급여를 실시함으로써 빈곤의 장기화를 방지하기 위하여 1999년 9월 제정되었다. 기초생활보장제도를 효과적이고 효율적인 맞춤형 빈곤정책으로 전환하여 지원대상을 확대하고 일할수록 유리한 급여체계를 마련함으로써 탈수급 유인을 촉진하고 빈곤예방기능을 강화하는 한편, 현행 제도의 운영상 나타난 미비점을 전반적으로 개선·보완하고자 2014년 12월 맞춤형 급여체계로 개편하였다.

233) 2014년 국민기초생활 보장법의 개정과 더불어 신설된 법이다. 주거급여의 지급 근거와 더불어 국민의 주거안정과 주거수준 향상에 이바지함을 목적으로 국가와 지방자치단체가 주거급여 관련 정책을 수립하도록 제정하였다.

234) 1977년 제정된 의료급여법에서 국민기초생활 보장법 외에도 재해구호법 등의 법률에 따라 생활이 어려운 사람에게 의료급여를 제공하도록 한다.

235) 1981년 제정된 노인복지법은 노인복지시설의 설치근거를 마련하고 이후 1989년 개정된 노인복지법에서 노령수당지급 근거를 마련하였다.

236) 1962년 아동복리법으로 제정된 이후 1981년 아동복지법으로 개정되면서 요구호아동 위주로 되어있던 대상범위를 전체아동으로 확대하였다. 아동이 건강하게 출생하여 행복하고 안전하게 자랄 수 있도록 아동의 복지를 보장하는 것을 목적으로 한다.

237)국가와 지방자치단체가 장애아동의 특별한 복지적 욕구에 적합한 지원을 통합적으로 제공함으로써 장애아동이 안정된 가정생활 속에서 건강하게 성장하고 사회에 활발하게 참여할 수 있도록 하며, 장애아동 가족의 부담을 줄이고자 2012년 장애아동복지지원법을 제정. 장애아동지원센터의 설치 근거 및 의료적 지원 및 복지지원 근거를 마련하였다.

238) 건강가정기본법이 건강한 가정생활의 영위와 가족의 유지 및 발전을 위한 국민의 권리·의무와 국가 및 지방자치단체 등의 책임을 명백히 하고, 가정문제의 적절한 해결방안을 강구하며 가족구성원의 복지증진에 이바지할 수 있는 지원정책을 강화하고자 2005년에 제정되고, 1989년 모자복지법이 제정되어, 모자가정의 복지급여지급, 복지자금 대여, 모자복지시설에 대해 규정하였다. 이후 2003년 모부자복지법으로 확대 개정한 뒤 2008년 한부모가족지원법으로 변경하고, 학비 지원 등의 근거마련 및 조손가족도 이법의 보호 대상자로 포함하여 한부모가족의 생활 안정과 복지 증진을 도모한다.

239) 심신장애자복지법이 1981년 6월 제정된 이후 1989년 장애인복지법으로 전부 개정되어 장애인 등록제를 시행하고, 장애발생 예방과 장애인의 의료·교육·직업재활·생활환경개선 등에 관한 사업을 정하여 장애인복지대책을 종합적으로 추진하며, 장애인의 자립생활·보호 및 수당지급 등에 관하여 필요한 사항을 정한다.

240)1991년 제정된 영유아보육법에서 보육시설의 종류에 대한 규정 등을 통해 아동복지법에 의한 탁아사업의 설립주체 제한을 극복하여 보육시설의 조속한 확대를 통해 영유아의 건전한 보호 교육 및 보호자의 경제적·사회적 활동이 원활하게 이루어지도록 함으로써 영유아 및 가정의 복지 증진에 기여하도록 하였다.

험, 산업재해보험 등을 들 수 있다. 국민건강보험은 “국민보건을 향상시키고 사회보장을 증진하”는 것을 목적으로 하여(국민건강보험법 제1조 전단), 국민에게 “질병·부상에 대한 예방·진단·치료·재활과 출산·사망 및 건강증진에 대하여 보험급여를 실시”하는 것이다(국민건강보험법 제1조 후단). 건강보험은 계층 간 또는 질병 고위험군과 저위험군 간의 위험 분산이라는 사회적 상호부조의 성격(젊은 사람이 노약자를 부양하고, 건강한 사람이 환자를 부양함)을 지닌 사회보험제도이기 때문에, 국민 모두(의료급여 등 제외)를 대상으로 한다(의무가입, 당연 가입). 그래서 건강보험은 가입자 및 사용자로부터 징수한 보험료와 국고 및 국민건강증진기금 등 정부지원금을 그 재원으로 한다. 반면 의료급여는 생활이 어려운 사람(기초생활보장수급자)에게 질병·부상·출산 등에 대하여 제공하는 의료서비스이다(의료급여법 제1조). 수급대상자는 1종 수급권자와 2종 수급권자로 나뉘는데, 국민기초생활보장 수급자(근로무능력가구, 희귀난치성질환등록자, 중증질환등록자, 시설수급자), 타법적용자(이재민, 의사상자, 국가유공자, 중요무형문화제 보유자, 북한이탈주민, 5.18민주화운동 관련자, 18세 미만 입양아동, 노숙인보호시설입소자 등) 등이다.

노인장기요양보험은 고령이나 노인성 질병 등의 사유로 일상생활을 혼자서 수행하기 어려운 노인 등에게 신체활동 또는 가사활동 지원 등의 장기요양급여를 제공하는 것이다(노인장기요양보험법 제1조). 수급대상자는 소득에 관계없이 65세 이상의 노인과, 65세 미만자로서 노인성으로 인한 기능장애자(치매·뇌혈관성 질환, 파킨슨병) 가운데 표준화된 평가판정에 의거하여 중증도가 비교적 높은 경우(6개월 이상 동안 혼자서 일상생활을 수행하기 어렵다고 인정되는 경우)이다.

산업재해보상보험은 근로자의 업무상 재해를 보상하고, 산업재해(업무상 재해, 부상, 질병, 사망)를 당한 근로자에게는 요양급여, 휴업급여, 장해급여, 간병급여, 직업재활급여 등을 제공하는 사회보험제도이다(제1조, 제36조).

표 4-12 국내 보험제도

보험제도	보험자(보장기관)	관련 법률
국민건강보험	국민건강보험공단	국민건강보험법
의료급여	지방자치단체장(특별시장ㆍ광역시장ㆍ도지사와 시장ㆍ군수ㆍ구청장) 시장, 군수, 구청장의 업무 중 수급권자의 관리, 급여비용의 심사ㆍ조정 및 지급 업무 등 의료급여에 관한 업무를 국민건강보험공단에 위탁.	의료급여법
노인장기요양보험	국민건강보험공단	노인장기요양보험법
산업재해보상보험	근로복지공단	산업재해보상보험법

그런데 국민건강보험공단은 국민건강보험, 노인장기요양보험의 보험자로서 의료급여의 경우는 위탁을 받아서 업무를 처리하는 자로서 의료기록을 제공받고, 이용하게

된다. 이와 같은 사회보장제도 운영과 관련한 보건의료 정보 수집과 활용에 대해서는 별도의 장으로 검토한다.

3. 의료 사회보장제도와 데이터 연계

(1) 국민건강보험공단

가. 개요

국민건강보험공단은 국민건강보험법에 의하여 설립된 법인인데, 국민건강보험의 보험자로 지정된 자이고(국민건강보험법 제13조), 노인장기요양보험의 보험자이기도 하다(노인장기요양보험법 제48조). 그뿐 아니라 의료급여에 관한 업무를 시장, 군수, 구청장으로부터 위탁받아 수행하는 전문기관이기도 하다(의료급여법 제33조 제2항). 그 외에 국민건강보험공단은 국민건강보험법에 의한 부수적인 업무로 국민의 건강증진사업과 같은 업무도 수행하고 있다.[241]

국민건강보험공단은 건강보험과 노인장기요양보험의 보험자, 의료급여의 수탁기관, 국민건강증진기관으로서 역할을 수행하는 과정에서 (i) 가입자 및 피부양자의 자격관리를 위해 필요한 자료, (ii) 보험료 부과, 징수를 위해 필요한 자료, (iii) 보험급여의 관리를 위해 필요한 자료, (iv) 보험급여 비용의 지급을 위해 필요한 자료, (v) 건강증진과 예방사업을 위해 필요한 자료를 광범위하게 수집하게 된다.

국민건강보험공단에 축적되는 정보는 한국의 국민건강보험이 모든 의료 공급자와 전 국민을 의무적으로 가입하도록 하는 강제성을 가지고 있다는 점, 지역별, 직업별 보험자가 나뉘어 있는 대부분의 국가와 달리 국민건강보험공단이라는 단일한 보험자로 이루어져 있다는 점, 주된 보수지불제도가 행위별 수가제이기 때문에 다른 여타의 제도에 비하여 압도적으로 많은 건강보험 데이터가 수집되고 있다는 평가를 받고 있다.[242] 그뿐 아니라 한국은 다른 국가와 다르게 전 국민을 대상으로 하는 건강검진서

---

241) 1. 가입자 및 피부양자의 자격 관리
2. 보험료와 그 밖에 이 법에 따른 징수금의 부과·징수
3. 보험급여의 관리
4. 가입자 및 피부양자의 건강 유지와 증진을 위하여 필요한 예방사업
5. 보험급여 비용의 지급
6. 자산의 관리·운영 및 증식사업
7. 의료시설의 운영
8. 건강보험에 관한 교육훈련 및 홍보
9. 건강보험에 관한 조사연구 및 국제협력
10. 이 법에서 공단의 업무로 정하고 있는 사항
11. 국민연금법, 고용보험 및 산업재해보상보험의 보험료징수 등에 관한 법률, 임금채권보장법 및 석면피해구제법(이하 "징수위탁근거법"이라 한다)에 따라 위탁받은 업무
12. 그 밖에 이 법 또는 다른 법령에 따라 위탁받은 업무
13. 그 밖에 건강보험과 관련하여 보건복지부장관이 필요하다고 인정한 업무

비스제도를 운영하고 있고, 암 검진 제도를 운영하는 국가는 많으나 한국처럼 매년 1천만 명 넘는 국민의 혈압, 혈당, 심지어 중성지방, HDL까지 측정하여 결과 값을 축적하는 곳은 극히 드물다고 한다. 아울러 이러한 건강보험 데이터들은 출생부터 한국인 개인에게 고유하게 부여된 주민등록번호를 통해 쉽게 연계될 수 있다는 점, 건강보험 운영에 필요한 개인의 출생, 사망, 주소, 직장, 장애, 소득 등 정보들은 행정전산망을 통해 주민등록번호를 기준으로 건강보험 데이터와 연계될 수 있다는 점을 건강보험공단이 수집하는 데이터가 방대한 이유로 들고 있다. 현재 이러한 모든 데이터는 1990년대부터 급속하게 진행된 정보화의 흐름을 타고 견실하게 구축된 IT 인프라를 통해 데이터베이스 형태로 실시간으로 구축되어 간다고 한다.[243]

나. 자격정보

① 자격 관리 정보 수집의 근거

건강보험은 건강보험의 가입자와 피부양자를 대상으로 하는데(국민건강보험법 제5조 제1항), 가입자는 직장가입자(원칙적으로 모든 사업장의 근로자 및 사용자와 공무원 및 교직원)와 지역가입자(가입자 중 직장가입자와 그 피부양자를 제외한 자)로 구분되며(국민건강보험법 제5조, 제6조), 여기에 직장가입자의 피부양자를 합하면 건강보험 적용대상이 된다. 의료급여는 의료급여 수급자를 대상으로 하는 것인데, 수급자는 법정 요건에 따라서 1종 수급권자와 2종 수급권자로 나뉜다.

장기요양보험의 수급대상자는 ① 소득에 관계없이 65세 이상의 노인과, ② 65세 미만자로서 노인성으로 인한 기능장애자(치매・뇌혈관성 질환, 파킨슨병)가운데 표준화된 평가판정에 의거하여 중증도가 비교적 높은 경우(6개월 이상 동안 혼자서 일상생활을 수행하기 어렵다고 인정되는 경우)이다.[244]

국민건강보험공단은 국민건강보험법, 노인장기요양보험법과 의료급여법에 의하여 가입자 및 피부양자, 수급자의 자격 관리를 업무로 관장하도록 수권을 받았기 때문에 이를 근거로 가입자, 수급자 자격 관리를 위해서 다양한 정보를 수집하게 된다.

이와 관련하여 국민건강보험법은 국민건강보험공단이 사용자, 직장가입자 및 세대주에게 가입자의 거주지 변경, 가입자의 보수·소득, 그 밖에 건강보험사업을 위하여

---

242) 박종헌(2016), "건강보험공단 빅데이터 활용 방법", 제5회 임상연구방법론 워크숍.

243) 앞의 논문 p32.

244) 국민건강보험 가입자 및 그 피부양자나 의료급여수급권자 누구나 장기요양급여를 받을 수 있는 것은 아니고, 일정한 절차(신청 → 방문 인정조사 → 등급판정)에 따라 장기요양급여를 받을 수 있는 권리(수급권)가 부여된 사람만이 급여를 받게 되는데, 이를 장기요양인정이라고 한다. 2011년 10월 현재 장기요양인정자는 324,227명으로 65세 이상의 노인 인구의 5.8% 정도가 장기요양보험의 혜택을 받고 있다.

필요한 사항을 신고하게 하거나 관계 서류를 제출하게 할 수 있다고 규정하고(제94조), 공단과 심사평가원이 국가, 지방자치단체, 요양기관, 보험회사 및 보험료율 산출기관, 공공기관, 그 밖의 공공단체 등에 대하여 가입자 및 피부양자의 자격 관리, 보험료의 부과·징수, 보험급여의 관리 등 건강보험사업의 수행을 위하여 필요한 자료(공단), 요양급여비용을 심사하고 요양급여의 적정성을 평가하기 위하여 필요한 자료(심사평가원)를 요청할 수 있다는 규정을 두고 구체적인 자료를 대통령령으로 정하고 있다(제96조).

② 수집하는 자격 관리 정보와 보유 기간

대통령령에서 국민건강보험공단이 제출을 요청할 수 있는 자료로 적시한 것 중 아래의 항목들이 자격 관리에 관한 정보로 볼 수 있다.

- ‘주민등록법’에 따른 주민등록자료
- ‘가족관계의 등록 등에 관한 법률’에 따른 가족관계등록 전산정보자료와 가족관계기록사항에 관한 증명서
- ‘출입국관리법’에 따른 외국인등록자료 및 국민·외국인의 출입국자료
- ‘재외동포의 출입국과 법적 지위에 관한 법률’에 따른 재외국민 및 외국국적동포의 국내거소신고자료
- ‘병역법’에 따른 병역 복무자료
- ‘형의 집행 및 수용자의 처우에 관한 법률’, ‘보호소년 등의 처우에 관한 법률’, ‘치료감호 등에 관한 법률’에 따른 시설의 입·출소(원)자 성명 및 주민등록번호, 입·출소(원)일, 수용 여부 자료

현재 국민건강보험공단은 다양한 자격 관리 정보를 수집, 보유하고 있는데, 그 중 ‘자격상세내역’이라는 명칭의 개인정보파일의 경우, 그 구성은 아래와 같다.[245] 보유

245) 그 외에도 ▲ 주민등록번호/성명 변경내역(보유기간 : 영구)은 보유 개인정보주체 수 4,889,472명, 개인정보 내역은 ‘이름:필수, 생년월일:필수, 주민등록번호:필수, 기타 주민등록번호 및 성명 변경내역 관련 정보’. ▲ 급여정지 및 급여정지 해제(보유 기간 영구)는 보유 개인정보주체 수 6,627,023명, 개인정보 내역은 ‘이름:필수, 생년월일:필수, 주민등록번호:필수, 외국인등록번호:필수, 기타 현역병 및 보충역 군입대 관련 정보’. ▲ 지역가입자 부과내역(보유기간 : 준영구)은 보유 개인정보주체 수 2,005,975,424명, 개인정보 내역은 ‘이름:필수, 생년월일:필수, 주민등록번호:필수, 외국인등록번호:필수, 기타 건강보험료액, 요양보험료액, 등급별 점수(생활 수준 및 경제활동 참가율, 재산, 소득, 자동차), 정지보험료, 지역보험료경감정보(경감금액, 경감률, 경감사유), 정산반영정보(정산반영금액, 정산부과금액), 전·월세 부과 정보(무상거주, 전세금, 월세금, 월세 보증금, 직권 전·월세, 부채 금액), 자동차 등록정보(자동차 세액), 세목(금액)’. ▲ 직장가입자 부과내역(보유기간 : 준영구)은 보유 개인정보주체 수 2,572,519,551명, 개인정보 내역은 ‘이름:필수, 생년월일:필수, 주민등록번호:필수, 외국인등록번호:필수, 기타 건강보험료, 요양보험료, 사업장 관련 정보, 보수월액, 소득월액, 산정 보험료, 감면보험료, 면제보험료, 선납보험료, 가입자부담보험료, 사용자부담보험료, 부과금액, 보수총액, 근무 월수, 유예기간, 감면코드, 정산보험료, 소득종류, 소득금액, 복무 시작일, 복무종료일’이다.

기간은 '영구'로 되어있다.

표 4-13 자격상세내역 개인정보파일

개인정보파일 명칭		자격상세내역
개인정보파일 운영 근거		국민건강보험법 제14조(업무 등), 국민건강보험법 시행령 제81조(민감정보 및 고유식별정보의 처리)
개인정보파일 운영 목적		건강보험 및 장기요양 업무의 관리·운영
개인정보파일에 기록되는 개인정보의 항목	정보주체 (개인정보를 수집하는 본인 등)	필수: 이름, 집주소, 직장주소, 집연락처, 직장연락처, 핸드폰, 생년월일, 고유식별번호(주민등록번호, 외국인등록번호) 기타: (주민전산정보, 가입자 건강보험 가입이력 정보, 차상위본인부담 경감대상자 자격 관리 정보, 장애인관리 정보, 외국인등록정보, 재외국민거소신고정보, 외국국적동포거소신고정보, 국가유공자정보)
	법정대리인 (14세 미만 보호자 등)	
개인정보의 처리 방법		오프라인 수집(개인의 신청서를 통한 서면 수집) 온라인 수집(홈페이지 회원신청, 전자접수 등) 시스템 연계를 통한 수집 기타(팩스, 4대 포털, EDI)
개인정보의 보유 기간		영구
개인정보를 통상적 또는 반복적으로 제공하는 경우	제공받는자	건강보험심사평가원, 공무원연금공단, 국가보훈처, 국민연금공단, 국방부, 국세청, 근로복지공단, 노동부, 사학연금, 행정자치부, 4대사회보험정보연계센터, 보건복지부(행복e음)
	근거	건강보험심사평가원 : 국민건강보험법 제96조 / 공무원연금공단 : 공무원연금법 제85조 / 국가보훈처 : 국가보훈법 제17조 / 국민연금공단 : 국민연금법 제123조 / 국방부 : 군인연금법 제20조 / 국세청 : 취업후 학자금상환 특별법, 조세특례제한법 제100조의13 / 근로복지공단 : 고용산재보험료징수법 제40조 / 노동부 : 고용정책기본법 제18조 / 사학연금 : 사립학교교직원연금법 제19조 / 행정자치부 : 지방세기본법 제134조의6 / 4대사회보험정보연계센터 : 2001.1월 발족한 전자정부특위에서 전자정부 11대 과제(G4C, 교육행정정보화(NEIS), 4대사회보험 정보연계 등)의 하나로 4대 사회보험 정보연계 시스템 구축, 보건복지부고시 제2010-63호, 노동부고시 제2010-11호 / 보건복지부(행복e음) : 사회복지사업법 제33조의3 제4항, 사회복지사업법 시행령 제18조2 제1항
	개인정보의 범위	건강보험심사평가원 : 가입자전산정보 / 공무원연금공단, 국가보훈처, 국민연금공단, 국방부, 국세청, 근로복지공단, 노동부, 사학연금, 행정자치부 : 직장가입자 취득및 상실정보, 사업장정보 / 4대사회보험정보연계센터 : 사업장 가입자 취득 및 상실 정보 / 보건복지부(행복e음) : 차상위본인부담 경감 대상자 자격관리 정보
개인정보파일로 보유하고 있는 개인정보의 정보주체 수		566,978,196건
타 정보시스템과의 연계 여부		예
해당 공공기관에서 개인정보 처리 관련 업무를 담당하는 부서	범위	해당 업무부서의 담당자와 시스템 관리자만이 접근하여 활용할 수 있다.
	공동사용부서	본부, 지역본부, 지사
개인정보의 열람 요구를 접수		국민건강보험공단 본부(자격부과실), 지역본부, 지사별 자격담당 부서

· 처리하는 부서		
개인정보파일에서 열람을 제한하거나 거절할 수 있는 개인정보의 범위 및 그 사유	개인정보의 범위	없음
	사유	

③ 열람청구권 등의 보장

한편 국민건강보험법은 가입자 및 피부양자의 자격, 보험료등, 보험급여, 보험급여 비용에 관한 공단의 처분에 이의가 있는 자는 공단에 이의신청을 할 수 있다는 규정을 두고 있다(제87조)[246]. 그런데 이는 구체적인 처분에 대한 이의로서 정보주체가 자신의 정보에 대한 열람이나 정정을 청구하는 것과는 차이가 있다. 따라서 처분에 대하여 이의신청을 보장하는 것과는 별도로 개인정보보호법에 의한 열람이나 정정청구권이 보장되어야 할 것이다. 실제로 국민건강보험공단의 개인정보취급방침과 개인정보파일[247]에는 열람 요구를 접수, 처리하는 곳으로 자격부과실로 소개하고 있다.[248] 현재 국민건강보험법상으로는 열람청구권이 보장된다는 점에 대한 규정은 없다.

④ 자격정보와 다른 정보의 연계·결합

현재 자격정보를 다른 정보와 연계하거나, 결합하는 것과 관련해서는 법령에 구체적 규정은 없다. 다만, 국민건강보험공단이 공개한 개인정보파일에서는 다른 정보시스템과 연계하고 있다고 밝히고 있다. 그러나 구체적으로 어떤 정보시스템과 연계하고 있는지에 대해서는 밝히지 않고 있다.

---

246) 제87조(이의신청) ① 가입자 및 피부양자의 자격, 보험료등, 보험급여, 보험급여 비용에 관한 공단의 처분에 이의가 있는 자는 공단에 이의신청을 할 수 있다.
② 요양급여비용 및 요양급여의 적정성 평가 등에 관한 심사평가원의 처분에 이의가 있는 공단, 요양기관 또는 그 밖의 자는 심사평가원에 이의신청을 할 수 있다.
③ 제1항 및 제2항에 따른 이의신청(이하 "이의신청"이라 한다)은 처분이 있음을 안 날부터 90일 이내에 문서(전자문서를 포함한다)로 하여야 하며 처분이 있은 날부터 180일을 지나면 제기하지 못한다. 다만, 정당한 사유로 그 기간에 이의신청을 할 수 없었음을 소명한 경우에는 그러하지 아니하다.
④ 제3항 본문에도 불구하고 요양기관이 제48조에 따른 심사평가원의 확인에 대하여 이의신청을 하려면 같은 조 제2항에 따라 통보받은 날부터 30일 이내에 하여야 한다.
⑤ 제1항부터 제4항까지에서 규정한 사항 외에 이의신청의 방법·결정 및 그 결정의 통지 등에 필요한 사항은 대통령령으로 정한다.

247) 이는 개인정보보호법에 의해 공개하고 있는 것이다. 국민건강보험공단의 개인정보파일은 국민건강보험공단 홈페이지의 '개인정보취급방침'과 개인정보보호포털의 개인정보열람 카테고리에 게시되어 있다.

248) 반면 건강보험심사평가원의 경우는 열람청구권을 행사할 수 있다는 점을 밝히고 있지 않다.

따라서 다음과 같이 데이터의 연계·결합에 대해서 나누어 볼 수 있을 것이다.

첫째, 사회보험의 제공 및 유지 목적으로 자격정보와 다른 정보를 연계·결합할 필요가 있는 경우가 있을 수 있다. 이때는 법적 근거가 별도로 없더라도 데이터의 연계·결합은 허용된다고 볼 수도 있을 것이다. 다만 개인정보보호의 원칙에 따라서 연계·결합이 이루어져야 할 것이다.

둘째, 그 외의 목적으로 데이터의 연계·결합이 이루어지는 경우는 원칙적으로는 해당 데이터의 연계·결합에 대한 법령의 명확한 규정이 마련되는 것이 바람직할 것이다. 이 경우 그 목적이 무엇인지에 따라 개인정보보호법의 원칙에 따른 연계·결합이 고려될 수 있을 것이다.

⑤ 개인정보 보호 원칙 준수 여부

현재 건강보험공단에서 자격정보를 수집하여 보유하는 것과 관련해서는 다음과 같이 개인정보 보호 원칙에 부합하지 않는 점이 있다.

첫째, 자격정보가 필요한 범위에서 최소한으로 수집하고 있는 것으로 보기 어려운 정보들이 있다. 예를 들어 연락처 정보도 집연락처, 직장연락처, 핸드폰 등을 모두 필수정보로 하고 있는 것은 최소수집의 원칙에 위반되는 것으로 보인다. 최소수집의 원칙에 따라서 수집 정보를 재조정할 필요가 있다.

둘째, 보유 기간을 영구로 하는 것은 큰 문제다. 특히 자격정보에는 '형의 집행 및 수용자의 처우에 관한 법률', '보호소년 등의 처우에 관한 법률', '치료감호 등에 관한 법률'에 따른 시설의 입·출소(원)자 성명 및 주민등록번호, 입·출소(원)일, 수용 여부 자료와 같이 매우 민감한 정보가 있다. 따라서 이와 같은 과거의 자격정보를 폐기하지 않고 영구적으로 보유하고 있는 것은 문제가 있다. 자격정보 보유를 필요한 최소한의 기한으로 할 필요가 있고 전과정보에 준하는 정보의 경우 꼭 필요한 정보만을 수집하고 즉각 폐기해야 한다.

셋째, 목적 외 활용과 관련하여 예를 들어 자격정보를 취업후 학자금상환 특별법에 따라 국세청에 제공하도록 하는 것은 사회보험 서비스의 자격을 확인하기 위하여 수집된 정보를 대출금 상환을 위한 목적으로 전용하는 것으로서 그 타당성을 수긍하기 어렵다.

넷째, 개인정보처리자는 개인정보의 처리 목적에 필요한 범위에서 개인정보의 정확성, 완전성 및 최신성이 보장되도록 하여야 하고, 개인정보 처리방침 등 개인정보의 처리에 관한 사항을 공개하여야 하며, 열람청구권 등 정보주체의 권리를 보장하여야

한다.

다섯째, 개인정보처리자는 정보주체의 사생활 침해를 최소화하는 방법으로 개인정보를 처리하여야 하고, 개인정보의 익명처리가 가능한 경우에는 익명에 의하여 처리될 수 있도록 하여야 한다.

⑥ 개선 방향

이와 관련하여 다음과 같은 제도 개선이 필요하다.

1. 수집하는 정보의 범위를 법령이나 고시 등으로 규율하고, 공개할 필요가 있다.
2. 수집하는 정보는 자격 정보로써 필요한 최소한의 범위로 한다.
3. 보유 기간은 목적 달성에 필요한 최소한의 기간으로 하여, 목적 달성 후 즉각 폐기하도록 한다.
4. 보유 기간을 법령이나 고시 등으로 명확하게 규율하는 것이 바람직하다.
5. 자격정보를 자격 확인 외의 목적으로 활용하는 것은 목적 외 활용으로 부당하다. 예를 들어 자격정보를 취업후 학자금상환 특별법에 따라 국세청에 제공하도록 하는 것은 그 타당성을 수긍하기 어렵다.
6. 자격정보에 대한 개인정보 처리방침 등 개인정보의 처리에 관한 사항을 공개하고, 개인정보주체의 열람권, 정정권 등을 보장하여 개인정보의 정확성, 완전성 및 최신성이 보장되도록 하여야 한다.
7. 개인정보처리자는 정보주체의 사생활 침해를 최소화하는 방법으로 개인정보를 처리하고, 개인정보의 익명처리가 가능한 경우에는 익명에 의하여 처리될 수 있도록 하여야 한다.

다. 보험료, 장기요양보험료 부과, 징수 관련 수집 정보

① 보험료, 장기요양보험료 부과, 징수 관련 정보 수집의 근거

국민건강보험공단은 보험료, 장기요양보험료의 부과 및 징수를 위해서도 방대한 자료를 수집한다. 건강보험의 보험료 결정은 개별 요율이 아닌 집단 요율에 따르는데, 전년도 급여비 지출액을 산출하고, 국고부담금 등을 공제한 후, 나머지 금액을 전체 가입자가 경제적 부담능력에 따라 일정률을 균등분담하는 방식이다. 그래서 임금근로자는 근로자가 일정기간동안 지급받은 보수를 기준으로 산정한 보수월액에 보험료율(6.12%)을 곱하여 산정한 금액을 가입자와 사용자가 각각 1/2씩 부담하고 있다(소득비례정률제). 지역가입자는 성, 연령, 성별, 장애, 재산, 소득을 점수화하여 이를 합산한 점수로 산정한 점수(보험료부과점수)에 점수당 단가를 곱하여 보험료를 부담한다. 장기요양보험료는 건강보험료로 산정한 보험료액에서 경감, 면제 비용을 공제한 금액에 장기요양보험료율을 곱하여 산정한다.

건강보험의 보험료 부과를 위해 국민건강보험공단은 직장의료보험 가입자의 경우에

는 직장과 소득에 대한 상세한 자료, 지역의료보험 가입자인 경우는 재산이나 소득에 대한 상세한 자료를 수집한다.

국민건강보험공단은 국민건강보험법에 의하여 보험료 부과·징수를 업무로 관장하도록 수권을 받았기 때문에 이에 근거하여 정보를 수집하게 된다.

② 수집하는 정보 및 보유 기간

국민건강보험공단은 아래와 같이 지역가입자 부과내역 정보, 직장가입자 부과내역 정보 파일의 경우 아래와 같은 개인정보를 수집하고 있다. 보존 기간은 준영구라고 한다.

표 4-14 국민건강보험공단이 수집하고 있는 개인정보

개인정보파일 명칭	개인정보 내역	파일 운영 목적
지역가입자 부과내역 준영구 2,005,975,424명	이름:필수, 생년월일:필수, 주민등록번호:필수, 외국인등록번호:필수, 기타(건강보험료, 요양보험료, 등급별 점수(생활수준 및 경제활동 참가율, 재산, 소득, 자동차), 정지보험료, 경감관련정보(경감금액, 경감률, 경감사유), 정산반영정보(정산반영금액, 정산부과금액), 전·월세 부과 정보(무상거주, 전세금, 월세금, 월세 보증금, 직권 전·월세, 부채 금액), 자동차 등록정보(자동차 세액), 세목(금액))	건강보험 및 장기요양 업무의 관리운영
직장가입자 부과내역 준영구 2,572,519,551명	이름:필수, 생년월일:필수, 주민등록번호:필수, 외국인등록번호:필수, 기타 (직장보험료, 요양보험료, 사업장관련정보, 보수월액, 소득월액, 산정보험료, 감면보험료, 면제보험료, 선납보험료, 가입자부담보험료, 사용자부담보험료, 부과금액, 보수총액, 근무 월수, 유예기간, 감면코드, 정산보험료, 소득종류, 소득금액, 복무 시작일, 복무종료일)	건강보험 및 장기요양 업무의 관리운영
가입자 전화번호 등록내역 준영구 33,847,751명	이름:필수, E-Mail:필수, 집연락처:필수, 직장연락처:필수, 핸드폰(연락처):필수, 주민등록번호:필수	건강보험 및 노인장기요양 업무의 관리·운영
요양기관 진료비 압류내역 30년 164,587명	이름:필수, 생년월일:필수, 주민등록번호:필수, 기타 (요양기관기호, 요양기관명, 사건번호, 송달일자, 채권청구금액, 변제금액, 금융기관명, 계좌번호)	금전채권에 대한 강제집행으로 제3채무자의 업무 수행
독촉고지내역 준영구 1,336,884,439명	이름:필수, 집주소:필수, 직장주소:필수, E-Mail:필수, 집연락처:필수, 직장연락처:필수, 핸드폰(연락처):필수, 주민등록번호:필수, 외국인등록번호:필수, 기타(증번호, NPS번호, 사업자등록번호 등 사업장정보, 체납보험료 등 체납정보)	4대 보험(건강·연금·고용·산재보험)의 독촉 고지 업무 수행
보험료 납부내역 준영구	이름:필수, 집주소:필수, 직장주소:필수, 집연락처:필수, 직장연락처:필수, 핸드폰(연락처):필수, 생년월일:필수, 주민등록번호:필수, 외국인등록번호:필수, 기타(가입자정보,	건강보험 및 장기요양보험,

3,604,301,957명	납부자번호, 증번호, NPS번호), 취득일, 상실일, 가입자구분, 사업장관리번호, 사업장기호), 사업장정보(사업장관리번호, 통합납부자번호, 사업자등록번호, 회계, 차수, NPS번호, 사업장원부번호, 단위사업장내역, 가입일, 탈퇴일, 대표자, 법인번호, 전화번호, 직종), 창구수납현황조회(가입자정보, 사업장정보, 카드정보, 수납정보, 고지정보, 체납정보, 취소정보 ,수표이력조회내역, 카드승인정보, 카드청구정보, 본부전송내역, 대납정보), 카드이체입금관리, 수납종류별입금내역, 수납종류별통계현황, 표준OCR 징수정보수신현황, 창구수납일일현황, 창구수납일일결산, 징수포털 실시간 입금내역, 징수포털 카드청구내역(카드정보, 승인정보, 가입자정보), 2D 실시간 납부내역(납부자정보, 고지정보, 수납정보, 가입자정보, 사업장정보), CD/ATM 수납내역(가입자정보, 수납정보, 고지정보, 계좌정보), 전자수납 실시간 납부내역(전자납부번호, 수납정보, 고지정보, 취소정보), 전자수납 수납내역(납부자정보, 고객정보,고지정보,수납정보,수수료정보, 계좌정보), 무고지 실시간 납부내역(납부자정보, 사업장정보, 가입자정보, 고지정보, 수납정보), 신용카드일일청구결과내역, 기금정산취소내역 재반영지역가상계좌 고지생성전송 내역(고지내역, 체납내역, 납부내역, 세부내역, 신청자 정보), 고지생성 조회취소내역(신청자정보, 납부일자, 입금액), 가상계좌 발급현황(은행, 가상계좌번호, 가입자번호), 지역가상계좌 미입금내역(은행명,가상계좌번호,가입자번호,세대주명,전화번호,핸드폰번호), 지역가상계좌 하이뱅크 입금조회(은행, 가입자번호, 가상계좌번호, 입금액), 표준 OCR 1차수납정보 내역(고객번호,상세내역,징수납부자번호,색인번호), 표준OCR 과오정정내역(납부자번호, 과오정정금액), 표준OCR 등 납부반영 오류건 재반영(증번호, 건별수정 및 오류처리내역), 사업장정보(사업자등록번호,법인등록번호,대표자,전화번호,주소,가입일자,탈퇴일자), 납부내역 (사업장기호, 금융기관지점,납부일자,납부보험료,납부처리지사), 납부취소내역(통합납부자번호,사업장관리번호,사업장기호,수납내역,수납금액,납부일자), 지역납부 상세조회내역(증번호, 고지년월, 수납유형, 수납지점, 자산내역, 물건코드, 자산등록일, 납부상세, 납부취소, 납부처리), 지역납부취소내역(건강보험료, 연체금, 납부구분, 납부취소일, 수납유형, 수납일자), 납부반영내역(고지연월, 고지유형, 납부형태, 은행명, 지점명, 납부일, 납부금액), 국가부담보험료 수납내역, 납부상세내역(통합납부자번호, 사업장명, 수납내역, 수납일자), 국가부담보험료 수납취소처리내역(통합사업장번호,사업장기호,수납유형,수납일자), 납부내역조회(지사, 증번호, 사업장기호, 사업장명, 우편번호, 주소, 취득일자, 상실일자, 부과시작월, 부과종료월, 지역납부내역, 직장납부내역), 기금정산 취소내역 재반영, 기금정산납부취소내역(가입자정보, 납부내역, 취소내역), 기금정산 재반영 납부내역 (직역구분, 통합가입자번호, 보험구분,수납유형,납부반영보험료,납부반영일자)환급금 선납 및 상계충당내역(가입자정보, 사업장정보, 충당내역, 처리내역, 취소내역), 환급금 발생내역(가입자정보, 사업장정보, 발생내역, 잔액, 선납 및 상계충당, 지금, 지급의뢰내역, 환입, 타보험지급대체, 소액 및 소멸시효, 취소), 환급금 지급계좌 등록내역(가입자정보, 사업장정보,	국민연금보험, 고용보험, 산업재해보상보험, 임금채권보장부담금, 석면피해구제분담금 업무의 관리·운영

	신청인정보, 보험료, 계좌정보, 예금주정보), 환급금 지급내역(가입자정보, 사업장정보, 계좌정보, 예금주정보, 지급정보, 지급액, 지급결과), 환급금 안내문 생성 및 발급내역(가입자정보, 사업장정보, 보험료, 연체금, 이자, 소멸시효, 발급정보, 수신정보), 소액 및 소멸시효내역(가입자정보, 사업장정보, 잡수입처리, 소멸시효정보, 취소내역), 환급금 환입내역 (사유,금액,처리내역,취소내역,지급내역), 저소득 취약계층내역, 반환결정 내역(가입자정보, 사업장정보, 반환결정정보,지급내역,타보험지급대체내역,취소내역), 지급보류 및 취소내역(과납내역, 보류내역, 취소내역), 타보험 지급대체내역, 이자내역, 환급금 지급대상 확인내용(발생사유, 발생금액, 잔액, 보류내역), 환급금 지급결정 등록내용(지급의뢰정보, 계좌정보, 예금주정보, 입금정보, 신청정보, 지급정보), 환급처리 및 결과조회 내역(가입자정보, 사업장정보, 예금주정보, 계좌정보, 지금결과처리내역, 타공단전송결과, 입금정보, 지급정보, 신청정보), 무통장 계좌 입금내역(거래일자, 입금금액, 입금인, 입금은행), 무통장 계좌 납부반영 내역(가입자정보, 사업장정보, 계좌정보, 납부반영정보), 납부확인서 발급내역(발급번호, 발급방법, 발급서식, 가입자정보, 사업장정보, 보험료 정보), 납부확인서 신청내역 (신청인정보, 전화번호, 팩스번호, 신청방법, 신청서식, 용도, 주소, 제공방법), 신청인 개인정보 확인내역(전 직장명, 자동이체정보, 세대정보, 취득일, 취득신고일, 상실일, 상실신고일, 주소), 행정정보공동이용 열람이력(가입자정보, 사업장정보, 보험료 정보, 이용기관정보), 사회복지통합망 조회이력(가입자정보, 사업장정보, 보험료 정보)	

③ 보험료 부과, 징수 정보와 다른 정보의 연계·결합

현재 보험료 부과, 징수 정보를 다른 정보와 연계하거나 결합하는 것과 관련해서도 법령에 구체적 규정은 없다. 국민건강보험공단이 공개한 개인정보파일에서는 다른 정보시스템과 연계하고 있다고 밝히고 있지만 역시 구체적으로 어떤 정보시스템과 연계하고 있는지에 대해서는 밝히지 않고 있다. 따라서 이 경우도 징수 정보와 마찬가지로 데이터의 연계·결합에 대해서 두 가지로 나누어 볼 수 있을 것이다.

첫째, 사회보험의 제공 및 유지 목적으로 다른 정보와 연계·결합할 필요가 있는 경우는, 법적 근거가 별도로 없더라도 데이터의 연계·결합은 허용된다고 볼 수도 있을 것이다. 다만 개인정보보호의 원칙에 따라서 연계·결합이 이루어져야 할 것이다.

둘째, 그 외의 목적으로 데이터의 연계·결합이 이루어지는 경우는 원칙적으로는 해당 데이터의 연계·결합에 대한 법령의 명확한 규정이 마련되는 것이 바람직할 것이다. 이 경우 그 목적이 무엇인지에 따라 개인정보보호법의 원칙에 따른 연계·결합이 고려될 수 있을 것이다.

④ 개인정보 보호 원칙 준수 여부

보험료 부과. 징수 정보를 수집하여 준영구적으로 보유하는 것이 개인정보 보호 원칙에 부합하는 것인지는 의문이 있다.

첫째, 목적 명확화 및 최소수집의 원칙과 관련하여, 보험료 부과, 징수에 관하여 수집하는 정보의 범위, 보유 기간을 정해서 법령이나 고시 등으로 공개할 필요가 있다. 현재는 수집하는 정보의 범위가 알려지지 않고, 보유 기간은 모든 정보를 기간도 애매한 준영구로 하고 있어서 문제이다.

둘째, 목적 외 이용의 금지와 관련해서도 보험료 부과, 징수 정보를 준영구로 보유하면서 보험료 부과, 징수의 목적 외로 건강정보와 연계하여 활용하는 것은 목적 외 활용에 해당하는데, 이를 정당화할 근거가 명확하지 않다.

셋째, 개인정보의 정확성, 완전성 및 최신성 보장 및 열람청구권 등 정보주체의 권리 보장과 관련해서도 문제가 있다.

넷째, 사생활 침해 최소화 및 익명처리의 원칙과 관련해서도 문제의 소지가 있다.

⑤ 개선 방향

이와 관련하여 다음과 같은 제도 개선이 필요하다.

1. 수집하는 정보의 범위를 법령이나 고시 등으로 규율하고, 공개할 필요가 있다.
2. 수집하는 정보는 자격 정보로써 필요한 최소한의 범위로 한다.
3. 보유 기간은 목적 달성에 필요한 최소한의 기간으로 하여, 목적 달성 후 즉각 폐기하도록 한다.
4. 보유 기간을 법령이나 고시 등으로 명확하게 규율하는 것이 바람직하다.
5. 보험료 부과, 징수에 관한 정보를 보험료 부과. 징수 목적 외의 목적으로 활용하는 것은 부당하다.

라. 요양급여 청구명세서 정보 DB 등

① 급여 비용지급 등과 관련된 정보 수집의 근거 - 심사평가 정보

국민건강보험과 의료급여의 보장 내용인 보험급여는 원칙적으로 현물급여인 요양급여(국민건강보험법 제41조)나 건강진단(국민건강보험법 제49조) 등으로 제공되는데, 예외적으로 현금급여인 요양비(국민건강보험법 46조) 등이 제공되기도 한다.

요양급여는 가입자 및 피부양자나 수급자의 질병·부상·출산 시에 제공하는데, 국민건강보험법과 의료급여법은 요양급여로 ① 진찰·검사, ② 약제·치료재료의 지급, ③

처치·수술 기타의 치료, ④ 예방·재활, ⑤ 입원, ⑥ 간호, ⑦ 이송의 7종을 규정하고 있다(국민건강보험법 제41조).

요양급여의 제공방법으로는 직접제공방법과 간접제공방법이 있는데, 우리나라는 간접제공방법에 의하며, 그 중 제3자 지불제도에 의한다. 그래서 (i) 가입자나 피부양자 또는 수급자가 필요할 때 요양기관에서 의료서비스(요양급여)를 이용하고, (ii) 요양기관이 보험자인 국민건강보험공단에 요양급여에 대한 비용을 청구하면, (iii) 요양급여비용을 심사하여, 보험자가 요양기관에 요양급여비용을 지불하게 된다.[249] 그런데 요양급여비용의 심사 과정은 건강보험심사평가원에서 담당하고 있다.

즉, 요양급여비용을 청구하려는 요양기관은 심사평가원에 요양급여비용의 심사 청구를 하여야 하고, 심사 청구를 받은 심사평가원은 이를 심사한 후 지체 없이 그 내용을 공단과 요양기관에 알리면, 심사 내용을 통보받은 공단이 그 내용에 따라 요양급여비용을 요양기관에 지급하게 된다(제47조). 이때 심사 청구를 하기 위해서는 요양급여의 내용을 국민건강보험공단과 심사평가원에 제공하여야 하는데, 요양급여의 내용은 바로 건강정보로서 민감정보에 해당한다.

한편, 건강보험심사평가원이 요양기관에 대하여 요양급여의 적정성 평가를 하여(제63조), 그 결과에 따라 요양급여비용을 가감하여 지급하는 성과지불제도를 운영하고 있다(제47조 제5항). 그에 따라 요양기관의 적정성을 평가하기 위한 정보의 수집도 건강보험심사평가원에 의해 이루어지게 된다.

건강보험이나 의료급여에서 의료공급의 대가를 결정하는 것을 '진료보수 지불제도'라고 부르는데, 우리나라는 진료비의 경우는 기본적으로 행위별수가제를 채택하면서 포괄수가제를 일부 반영하고 있다. 따라서 모든 행위에 대한 상세한 내역이 국민건강보험공단에 제공되어야만 보수를 산정할 수 있으므로, 요양기관은 그에 대한 상세한 정보를 모두 제공해야만 한다. 약제비의 경우는 1999년 11월 15일부터 실거래가상환제를 포함하여 사후 사용량 약가 연동제 등 몇 가지 제도를 시행하고 있으므로 그에 대한 정보가 청구를 위해서 제공되어야 한다. 국민건강보험법과 의료급여법, 노인장기요양보험법은 보험급여의 관리, 보험급여 비용의 지급의 업무를 국민건강보험공단에서 관장하도록 하고 있으므로 이를 근거로 관련 정보를 수집하게 된다.

---

249) 의료행위나 약제, 치료제에 대한 대가인 요양급여비용은 기본적으로 보험자인 국민건강보험공단이 부담하지만, 예산상의 문제로 그 비용의 일부를 본인에게도 부담하도록 하고 있다('본인일부부담금'). 요양급여비용 중 본인이 부담할 비용의 부담율 및 부담액은 국민건강보험법 시행령 등에서 규정하고 있는데, 우리나라의 경우는 정률제를 기본으로 하고, 상한제, 경감제 등을 혼용하고 있다.

② 수집하는 정보의 내용과 보유 기간

앞에서도 본 것처럼 국민건강보험공단이나 건강보험심사평가원이 요양급여의 제공과 관련하여 수집하는 정보는 건강정보로서 민감정보에 해당하므로 민감정보의 제공에 대한 명확한 법률의 규정이 있어야 하는데, 국민건강보험법, 의료급여법, 장기요양보험법 등의 심사청구에 대한 규정이 이에 해당한다.

한편, 우리나라 건강보험, 의료급여, 노인장기요양보험은 요양급여 방법의 특수성과 진료비와 약제비 산정과 지불, 적정성 담보제도의 특수성이 있는데, 요양급여비용의 심사, 본인부담금의 책정과 부과(상한제, 경감제 적용), 행위별수가제와 포괄수가제 적정성 평가, 실거래가 상환제에 따른 약제비 지급, 부당 청구 방지 위한 심사 및 사후조사, 사용량 약가 반영 등으로 인하여 포괄수가제나 총액계약제 등과 비교하여 보험급여 관리 및 비용지급을 위하여 더 방대한 자료를 수집·처리해야 한다.

즉, 직접 요양급여를 제공하는 NHS 방식이 아닌 보험방식이기 때문에 보험자인 국민건강보험공단이 집중해서 모든 요양급여 비용에 대한 청구를 심사해야 하고 나아가 적정성 평가도 하게 되므로 그에 대한 모든 정보를 수집해야 하는 것이다. 게다가 진료보수 지불제도도 총액계약제나 포괄수가제, 인두제, 봉급제가 아닌 행위별 수가제와 실거래가 상환제를 택하고 있기 때문에 다른 나라의 국민건강보험의 보험자에 비해서 각 행위별로 상세한 정보를 국민건강보험공단이 수집하게 된다. 따라서 우리나라 국민건강보험공단은 외국의 다른 어느 보험자나 서비스에 비해서 집중하여 수집·처리하는 정보의 양이 훨씬 많게 된다.

이 과정에서 요양기관으로부터 국민건강보험공단이 수집하는 정보로 대표적인 것이 요양급여비용 청구명세서 정보이다. 국민건강보험공단으로부터 해당 업무를 위임받은 건강보험심사평가원은 1년에 14.4억 건의 청구정보를 심사한다고 한다.

요양급여 청구명세서 정보는 의·치과 명세서, 한방 명세서, 보건기관 명세서, 약국 명세서, 정신과 정액 의료급여비용명세서, 요양병원 명세서, 질병군(DRG) 요양급여비용 명세서 등, 요양기관 현황통보서, 일반현황[250], 시설현황[251], 운영현황[252], 인력현황[253], 장비현황[254] 등 19종의 명세서가 포함된다고 한다. 명세서 자료는 아래와 같이

250) 일반현황 : 개설자 인적사항, 전문의자격, 주소, 설립형태, 진료과목, 입원환자 간호관리료 등급, 입원환자 식대 및 중환자실 현황 등.

251) 시설현황 : 병실 및 병상현황 입원실, 분만실, 신생아실, 모자동실, 수술실, 응급실, 인공신장실, 물리치료실, 중환자실, 무균치료실, 격리병실 등.

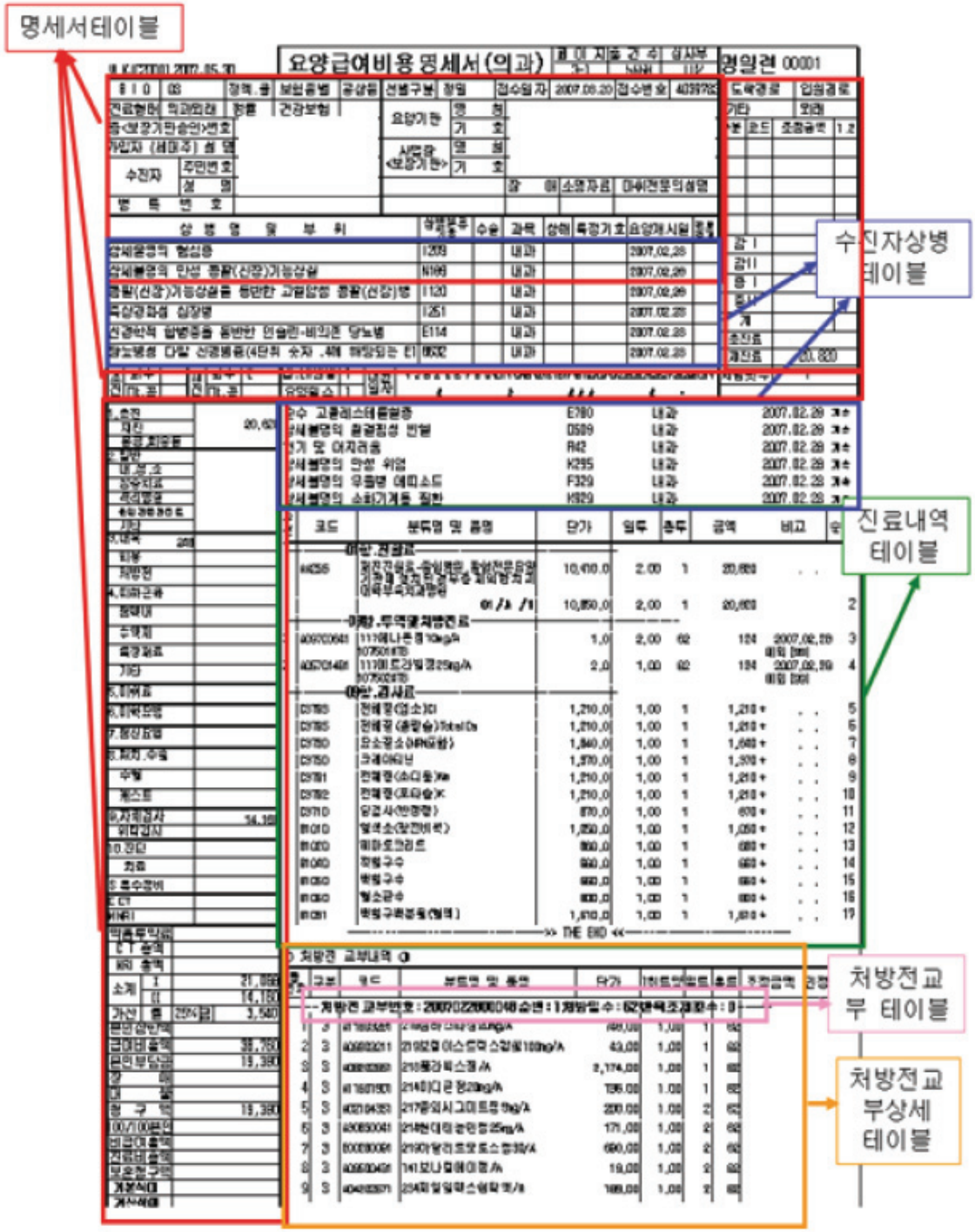
252) 운영현황 : 의약분업 예외기관, 응급의료기관, 개방병원기관, 가정간호실시기관, 인공와우기관, 촉탁기관 등.

253) 인원현황 : 의사, 치과의사, 한의사, 약사, 간호사, 임상병리사, 방사선사, 영양사 등 24분류 의료인력.

254) 장비현황 : CT, MRI, PET등 방사선진단 및 치료, 검사, 이학요법등 167종.

구성되어 있다고 한다.

그림 4-5 요양급여비용 청구명세서의 테이블 구성



* 출처: 보건의료 근거 생산을 위한 건강보험 청구자료 분석 매뉴얼 (2017.7)

자료의 구조는 아래와 같이 명세서 조인키와 처방전 교부번호로 연결돼 있다고 한다.[255]

그림 4-6 건강보험 청구자료 테이블 구조

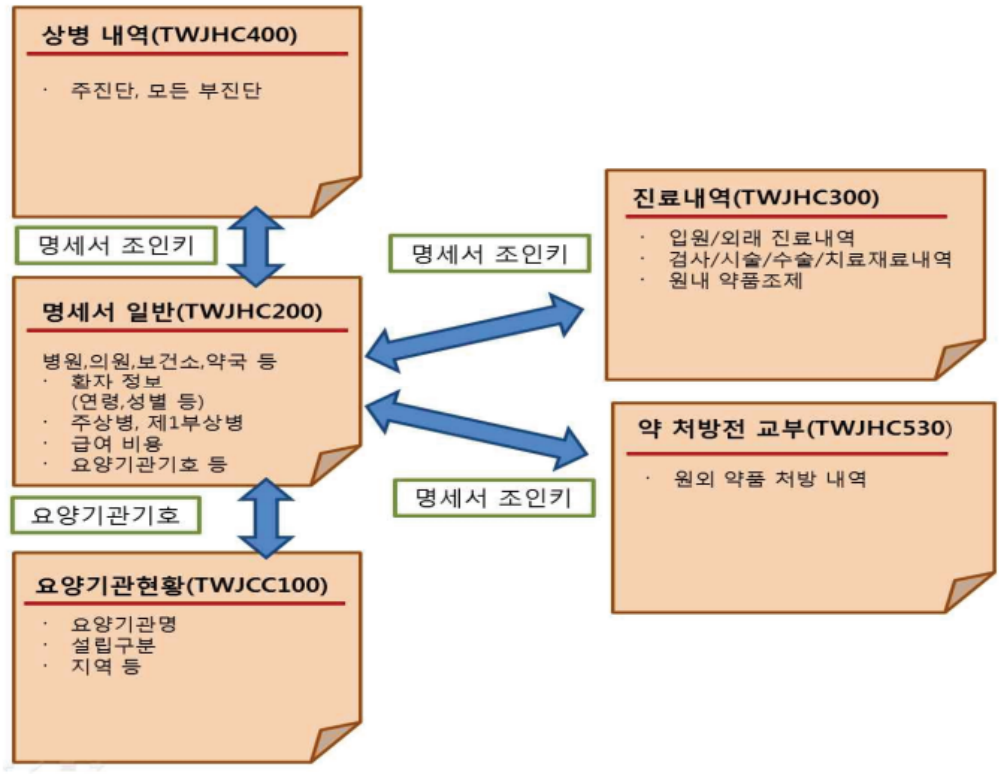


표 4-15 명세서 일반내역 자료

항목명	세부내용
명세서조인키	명세서 레벨의 팩트간 조인 Key(각 자료 간 연결키)
보험자코드	4: 건강보험, 5:의료급여, 7: 국비보훈(상이처,무자격자)
수진자식별대체키	주민등록번호 대체번호
성별구분	9:기타, 1:남, 2:여
수진자연령	0.2: 28일미만, 0.9: 28일이상1세미만 , 999: 연령산출 불가
수진자통계연령	매년 말일 기준의 수진자 연령
수진자구분코드	0:일반 , 1:신생아 , 2:행려
요양기관식별대체키	요양기관기호 대체번호
요양종별코드	01 상급종합병원, 11 종합병원, 21 병원, 28 요양병원, 31 의원,41 치과병원, 51 치과의원 , …. 81:약국
시도코드	11 서울, 21 부산, 22 인천, 23 대구, 24 광주, 25 대전, 26 울산, 31 경기, 32 강원, ... 39 제주
지역형태코드	01:특별시, 02:광역시, 03:구가있는시, 04:시, 05:군, 06:특별시의군, 07:광역시의군, 08:구가있는시군, 09:시의군
서식구분코드	021: 의과입원 031: 의과외래 041: 치과입원 051: 치과외래 …. 121: 한방입원 131: 한방외래

255) 안동대학교 산학협력단(2016), "의료정보안내서", 보건의료빅데이터 활용 고도화 방안 연구 부록, p19.

주상병코드	주된 상병분류기호(약국 증상분류기호, 한방상병코드, 의과:표준질병사인분류코드), (배제진단 제외)
부상병코드	주된 상병분류기호외의 추가 상병분류기호 DW에서 생성 된 ABC구분추가(배제진단 제외)
진료과목코드	요양기관에서 청구시 기재한 진료과목코드, 의원급 표시과목코드
표시과목코드	의원급의 경우 요양기관현황의 표시과목코드, 병원급이상은 청구된 진료과목코드 기준
내원일표시코드	요양기관에 실제 내원한 일자. 31 occurs.(외래서식의 경우 내원일짜 위치에 9 표시)
당월요양개시일자	진료를 받기 시작한 일자, 조제투여일자, 최초내방일자(청구명세서 단위)
요양종료일자	수진자가 진료 받기를 종료한 일자
최초입원일자	그 상병진료를 위하여 그달에 최초 내원한 년/월/일(분리청구시 기재)
입내원일수	수진자가 진료를 받기위해 요양기관에 내원한 일수(초진+재진). 입원일수
요양일수	수진자를 진료한 총일수
원외처방일수	처방전 발행 내역 중 처방일수의 합계
원외처방약제비	처방전 발행 내역 중 약제비 금액의 합계
원외처방전건수	처방전을 발행한 건수
청구요양급여비용총액	수진자 부담 금액과 보험자 부담 금액을 합한 요양급여비용총액
청구본인부담금	진료형태별로 요양급여비용명세서의 본인일부부담금을 합하여 기재
청구보험자부담금	진료형태별로 요양급여비용명세서의 청구액을 합하여 기재
심결요양급여비용총액	심사결과 수진자 부담 금액과 보험자 부담 금액을 합한 요양급여비용총액
심결본인부담금	심사결과 수진자 본인이 부담해야 될 금액
심결보험자부담금	심사결과 보험자가 부담해야 될 금액
공상구분	0: 정상건 1: 공상건(공교공단) 3: 보훈감면30% 4: 보훈국비 5: 보훈감면50% 6: 보훈감면60% ….
상해외인구분코드	한국표준질병사인분류에 의거 상병의 원인에 해당하는 분류기호중 영문 첫 자리(V, W, X, Y)만 기재
특정기호구분	V001:인공신장투석 …. V008:가정간호 ….
진료결과구분	1: 계속 2: 이송 3: 회송 4: 사망 5: 기타 9: 퇴원
입원도착경로구분	도착경로+입원경로: 도착경로(1:타기관경유, 2:응급구조대, 3:기타) 입원경로(1:응급실, 2:외래)

표 4-16 명세서 진료내역 자료

항목명	세부내용
명세서조인키	명세서 레벨의 팩트간 조인 Key
항코드	01:진찰료, 02:입원료, 03:투약료, 04:주사료, 05:마취료, 06:이학요법료, 07:정신요법료, ….
분류코드구분	1:수가, 2:준용, 3:국산보험등재약, 4:수입약/원료약/조제(제제)약, 7:협약재료, 8:일반재료 (한방 - A: 수가, B: 준용수가, C: 약가, H:치료재료)
분류코드	수가(행위)코드, 약품코드, 재료코드 등 진료내역코드
일반명코드	약가 주성분 등재약, 수입약의 일반명 코드
1회투약량	1회 환자에게 투약한량
1일투약량	1일 환자에게 투약한량
일일투여량 또는 실시회수	1일 투여량(소수 셋째자리에서 4사5입)을 기재, 의약품 및 처방내역 의약품의 경우 1일 투약횟수 기재
총투여일수 또는 실시회수	총투여일수 또는 실시 횟수
총사용량 또는 실시횟수	1회투약량 * 1일투약량 * 총투여일수또는실시횟수
단가	분류코드별 단가

금액	단가 * (일일투여량 또는 실시횟수) * (총투여일수 또는 실시횟수) = 금액
가산적용금액	2란 행위 가산이 있는 경우 요양기관 종별 가산율을 적용한 금액
단가변경일자	약국:건강보험 진료 수가기준과 청구단가가 상이한 경우 최초 투여일자 기재 (수탁기관검사의뢰일) 의과, 치과:단가변경일자

표 4-17 수진자 상병내역 자료

항목명	세부내용
명세서조인키	명세서 레벨의 팩트간 조인 Key
일련번호	상병 순서번호
상병코드	통계청 고시 제1993-3호(1993.11.20)에 의거 한국표준질병사인분류의 상병분류기호 (약국 증상분류기호, 한방상병코드, 의과:표준질병사인 분류코드) 자릿수1 - A:서식이 보건기관이고 진료과가 한방이 아닌 경우이거나 B, C가 아닌 경우, B:서식이 보건기관이고 진료과가 한방인 경우이거나 서식이 한방인 경우, C:서식이 직접조제인 경우 $:상병기호가 ''이거나 마스터에 코드가 없는 경우 발생
상병분류구분	1: 주상병, 2: 부상병, 3: 배제진단(확진이전의 R/O상병 등)
진료과목코드	진료를 받은 진료과목(병원급 이상) 또는 상병명에 해당하는 진료과목(의원급) 을 기재하되, 진료과목이 2개 이상에 해당되는 경우 상병별로 모두 기재 01 : 내과 02 : 신경과 03 : 정신과 04 : 외과 …
요양개시일자	요양기관에 그 상병 진료를 위하여 그 달에 최초 내원한 년월일을 기재. 단, 수진자의 요양급여비용명세서가 동일 청구서에서 정액, 정률 등으로 명세서가 분리되는 경우는 해당 요양급여비용명세서의 최초 진료일자를 기재
내과세부전문과목	00: 내과통합 01: 소화기내과 02: 순환기내과 03: 호흡기내과 04: 내분비.대사내과 05: 신장내과 06: 혈액종양내과 07: 감염내과 08: 알레르기내과 09: 류마티스내과 11: 소아감염 12: 소아내분비

표 4-18 원외처방 상세내역 자료

항목명	세부내용
명세서조인키	명세서 레벨의 팩트간 조인 Key
처방전교부번호	처방전교부일자+일련번호
처방전일련번호	진료기관에서 처방전 발행시 부여한 교부번호에 대한 일련번호
분류코드구분	명세서의 분류 코드 구분(1:수가 3:약가 4:수입,원료,조제,제제 7:협약 8:일반재료)
분류코드	약품코드
일반명코드	약가 주성분 등재약, 수입약의 일반명 코드
1회투약량	1회 환자에게 투약한량
1일투여횟수	처방전의 1일투여횟수
총투여일수	총투여일수 또는 실시 횟수
총사용량	1회투약량 * 1일투여횟수 * 총투여일수
단가	분류코드별 단가(상한가)
금액	단가 * (일일투여량 또는 실시횟수) * (총투여일수 또는 실시횟수) = 금액

요양급여비용 청구명세서 정보는 요양급여비용의 청구와 지급을 위한 정보이기 때문에 원칙적으로는 청구와 지급이 마쳐지면 삭제되어야 한다. 그런데 현재는 5년을 보존하고 있다. 그러나 5년이 지나도 이를 삭제하지 않고 국민건강DW정보에 그대로 보유하고 있는 것으로 보인다.

현재 국민건강보험공단에는 요양급여청구명세서 자료 외에도 아래와 같이 요양급여 내역 등의 정보가 수집되어 5년 또는 10년씩 보유되고 있다고 한다.

표 4-19 국민건강보험공단 개인정보 보유 내역

개인정보파일 명칭	개인정보 내역	파일 운영 목적
요양급여비용청구 명세서 5년 11,040,724,000명	이름:필수, 주민등록번호:필수, 외국인등록번호:필수, 건강:필수, 성생활:필수, 유전정보:필수, 기타:필수, 기타 (청구서, 명세서, 진료내역, 상병내역, 처방전교부내역, 처방전교부상세내역, 심사조정내역)	보험급여 사전·사후관리 활용
의료급여내역 10년661,740,265명	이름:필수, 주민번호:필수, 기타, 기타 (지급대상, 직역, 인정등급, 장기요양관리번호, 증번호, 서비스 종별, 청구관련 자료, 심사결정요양급여비용총액, 심사결정본인부담금, 진료일수) 지급년월, 지급일련번호, 상세지급일련번호, 수급자성명, 서비스종류, 서비스시작일, 서비스종료일, 서비스횟수, 수가코드, 수가명	의료급여, 보험급여, 보험급여비용의 지급
현금급여비 지급내역(본인부담금환급금, 본인부담액보상금, 현금급여 포함)10년 2,006,024 건	이름:필수, 집연락처:필수, 핸드폰(연락처):필수, 생년월일:필수, 주민번호:필수, 건강, 기타 (계좌번호, 금융기관, 내구연한, 민원구입액, 보장구구입일자, 보장구구입처,보장구코드,본인부담액,사고경위,사망구분,사망상병, 사망일,사업장기호,사업장명,사용제품종류,상병코드,수진자명,수진자주민번호,실소요액,실종자여부,심사결정액,예금주관계,예금주성명,예금주주민번호,요양기관,의뢰일,자격상실여부,장애등급,장애발생경위,장애발생유형,장애종류,접수유형,접수일,중복장애여부, 증번호,지급방법,지급상한액,지급액,처리지사,청구사유,청구인관계,청구인성명,청구인전화번호,청구인주민번호,청구인주소,최초지급,최초지급,출산구분,출산일,출산자명,출산자주민번호,출산장소, 투약개시일,투약기간)	현금급여비 지급내역 관리로 지급대상자에게 적정 지급여부 확인
임신 · 출산 진료비 지원신청내역 5년 3,256,198명	이름:필수, E-Mail:필수, 집연락처:필수, 핸드폰(연락처):필수, 주민번호:필수, 외국인등록번호:필수, 건강:필수, 기타(수진자 성명, 전화번호, 휴대폰번호, 임신확인일, 사실확인일, 분만예정일, 다태아구분, 요양기관기호, 요양기관명, 의사면허번호, 담당의사명, 신청자 성명, 수진자와의 관계, 신청인 전화번호, 신청일자)	임신·출산 진료비지원 대상자의 급여내역 및 카드발급 관리
요양급여내역 10년 54,760,779명	이름, 주민번호, 외국인등록번호, 건강, 성생활, 유전정보, 기타, 기타 (청구서내역, 명세서내역, 심사조정내역)	건강보험 지급업무의 관리・운영

한편, 고의 또는 중대한 과실로 인한 범죄행위에 그 원인이 있거나 고의로 사고를 일으킨 경우나, 고의 또는 중대한 과실로 공단이나 요양기관의 요양에 관한 지시에 따르지 아니한 경우에는 보험급여를 지급하지 않으므로, 국민건강보험공단은 예를 들어 상해로 인한 치료의 경우에는 사고발생경위, 사법기관에서 제공받은 처분결과, 가해자 성명 등에 대한 정보를 수집하고 있는데 준영구로 보존하고 있다. 현재 31,510,974명에 대한 정보가 수집되어 있다. 특히 이 정보는 일종의 형사처벌 기록과도 같은 것인데 준영구로 보유하고 있다.

표 4-20 급여사후관리 결정내역

개인정보파일 명칭	개인정보 내역	파일 운영 목적
급여사후관리 결정내역(상해요인 사전·사후 결정내역, 부당수급, 체납후진료 포함) 준영구 31,510,974명	이름:필수, 집주소:필수, 직장주소:필수, 집연락처:필수, 핸드폰(연락처):필수, 주민번호:필수, 외국인등록번호:필수, 건강:필수, 기타 (전산관리번호, 결정번호, 건강보험증번호, 사업장명칭, 요양기관명칭, 사유발생일, 진료개시일, 진료종료일, 입원·외래구분, 공단부담금, 본인부담금, 사전상한적용금액, 적용금액, 주상병기호, 주상병명, 부상병기호, 부상병명, 상해상병기호, 상해상병명, 납부의무자, 결정금액, 사고발생경위, 결정근거및사유, 사법기관에서 제공받은 처분결과, 가해자 성명, 가해자 주민번호, 가해자 연락처)	건강보험 급여사후업무의 관리·운영

③ 요양급여 청구정보 등과 다른 정보의 연계·결합

현재 요양급여 청구정보 등과 다른 정보를 연계하거나, 결합하는 것과 관련해서는 법령에 구체적 규정이 없는데, 국민건강보험공단은 다른 정보시스템과 연계하고 있다고 밝히고 있다. 물론 구체적으로 어떤 정보시스템과 연계하고 있는지에 대해서는 밝히지 않고 있다.

요양급여 청구정보와 다른 정보의 연계·결합에 대해서도 다음과 같이 두 가지로 경우를 나누어 볼 수 있을 것이다.

첫째, 사회보험의 제공 및 유지 목적으로 자격정보와 다른 정보를 연계·결합할 필요가 있는 경우. 이때는 법적 근거가 별도로 없더라도 데이터의 연계·결합은 허용된다고 볼 수도 있을 것이다. 다만 개인정보보호의 원칙에 따라서 연계·결합이 이루어져야 할 것이다.

둘째, 그 외의 목적으로 데이터의 연계·결합이 이루어지는 경우. 원칙적으로는 해당 데이터의 연계·결합에 대한 법령의 명확한 규정이 마련되는 것이 바람직할 것이다. 이 경우 그 목적이 무엇인지에 따라 개인정보보호법의 원칙에 따른 연계·결합이 고려될 수 있을 것이다.

④ 정보 수집의 적정성과 개인정보 보호원칙 준수 여부

국민건강보험공단과 건강보험심사평가원이 전 국민의 국민건강보험, 의료급여, 장기요양보험의 요양급여에 관한 정보를 각 요양기관으로부터 수집하여 처리하는 것은 전 국민의 민감한 건강정보를 단일한 기관에서 집적하여 처리하고 있다는 점에서 애초부터 매우 큰 위험을 내포하고 있는 것이다.

따라서 건강보험공단이나 심사평가원이 전 국민의 건강정보를 수집하여 처리하는 것은 엄격하게 요양급여에 대한 심사 목적으로 한정해야 한다. 수집하는 정보는 요양

급여 심사에 필요한 정보로 한정되어야 하고, 보유 기간도 목적 달성에 필요한 범위로 제한되어야 하고, 목적을 달성한 후에는 즉시 폐기하여야 한다.

그런데 현재 수집하는 정보가 요양급여 심사에 필요한 정보로 한정되어 있는지는 검토가 필요하다. 보유 기간도 현재는 5년, 10년으로 되어있는데 심사 청구를 마친 후에도 5년이나 10년간 보유하는 것은 지나치게 장기간 보유하는 것이 되어 개인정보 보호 원칙에 부합하는 것인지는 의문이 있다. 특히 요양급여와 관련된 모든 정보가 사실상 영구적으로 폐기되지 않고 저장되고 있는 것으로 보이는데, 이는 심각한 문제가 있다.

⑤ 개선 방향

이와 관련하여 다음과 같은 제도 개선이 필요하다.

1. 수집하는 정보의 범위를 법령이나 고시 등으로 규율하고, 공개할 필요가 있다.
2. 수집하는 정보는 자격 정보로써 필요한 최소한의 범위로 한다.
3. 보유 기간은 목적 달성에 필요한 최소한의 기간으로 하여, 목적 달성 후 즉각 폐기하도록 한다.
4. 보유 기간을 법령이나 고시 등으로 명확하게 규율하는 것이 바람직하다.
5. 보험료 부과, 징수에 관한 정보를 보험료 부과. 징수 목적 외의 목적으로 활용하는 것은 부당하다.

마. 건강증진과 예방 등의 업무와 수집하는 정보

① 개요

국민의 건강증진과 질병의 예방사업도 매우 중요한 사업이다. 이 과정에서 국민의 건강정보가 수집될 수 있다. 이 정보들 역시 매우 민감한 정보들이다. 이 정보가 수집의 목적에 부합하게 활용되는 것은 무엇인지, 만약 전체 국민의 건강을 증진할 수 있도록 2차적으로 활용될 수 있다면, 그 장단점을 고려하여 활용의 범위나 필요한 안전조치들이 분명하게 마련되어야 할 것이다.

국가나 지방자치단체가 국민을 대상으로 하여 시행하는 건강검진사업은 국민건강보험법에 의한 건강검진을 비롯해서 여러 법률에서 정하고 있는 건강검진이 있다.[256] 건강검진기본법은 모든 국민은 자신이 받은 국가건강검진의 내용과 그 결과에 대하여

256) '모자보건법'에 따른 영유아에 대한 건강검진, '영유아보육법'에 따른 영유아에 대한 건강검진, '학교보건법'에 따른 초·중·고등학교 학생의 건강검사, '청소년복지지원법'에 따른 청소년 건강진단, '국민건강보험법'에 따른 건강검진, '산업안전보건법'에 따른 일반건강진단, '의료급여법'에 따른 건강검진, '암관리법'에 따른 암검진, '노인복지법'에 따른 건강진단.

설명을 들을 권리를 가지며, '공공기관의 정보공개에 관한 법률'로 정하는 바에 따라 국가와 지방자치단체에 대하여 국가건강검진에 관한 정보의 공개를 청구할 권리를 가진다고 규정하고 있다(제4조 제2항).

② 정보 수집의 근거

국민건강보험공단은 건강증진과 예방에 관한 업무도 수행하고 있다. 국민건강보험공단은 가입자와 피부양자에 대하여 질병의 조기 발견과 그에 따른 요양급여를 하기 위하여 건강검진을 실시한다(국민건강보험법 제52조 제1항). 이 과정에서 사실상 전 국민의 정기적인 건강검진 자료가 국민건강보험공단에 수집된다.

③ 수집하는 정보의 범위와 보유 기간

국민건강보험공단이 공개한 개인정보파일의 내역에 의하면 이 자료는 영구 보존된다고 하며, 현재 916,864,439명에 대한 정보가 보관되어 있다고 한다. 그 내역은 검진의 모든 결과, 문진 내역, 소견 및 조치 등이 포함된다. 그중 만성질환자와 전단계자에 대해서는 사후관리대상자로 별도로 관리하는데, 현재 77,134,811명의 정보가 보관되어 있으며, 그 내역은 아래와 같이 가족관계나 교육 정도, 직업 등의 정보는 물론, 검진결과, 암 검진 결과, 특별한 질환 여부 등에 대한 정보도 수집된다. 이 정보는 10년간 보관된다.

표 4-21 건강증진과 예방 등의 업무와 수집하는 정보

개인정보파일 명칭	개인정보 내역	파일 운영 목적
건강검진 대상자 및 검진결과 내역 영구 916,864,439명	이름:필수, 집주소:필수, E-Mail:필수, 집연락처:필수, 핸드폰(연락처):필수, 주민번호:필수, 외국인등록번호:필수, 건강, 기타(대상자・수검자・검진의사 성명 및 주민번호, 일반・암검진・영유아 검진결과 및 종합판정, 문진내역, 소견 및 조치, 검진기관명, 검진의사 면허번호, 검진(판정)일, 개인정보활용동의서)	가입자 및 피부양자 건강검진 실시 및 평생 건강관리 체계 수행
검진사후관리 대상자 내역 10년 123,764,021명	- 이름, 집주소, 직장주소, E-mail, 집연락처, 직장연락처, 핸드폰, 생년월일 - 고유식별번호(주민번호, 외국인등록번호) - 민감정보(검진결과, 건강, 기타) - 기타 : 가입자 및 세대주 정보(보험료, 성명, 주민번호, 관계코드, 취득일자, 동거가족구분, 성별, 연령, 건강보험증번호, 가족수), 결혼상태, 교육정도, 직업, 사업장기호, 차상위 여부, 차상위구분, 차상위종별, 1차일반건강검진및생애전환기건강진단결과, 2차일반건강검진및생애전환기건강진단결과, 암검진결과, 대사증후군 상담관리, 유질환(집중관리)군 상담관리,	고혈압, 당뇨병, 대사증후군 등 만성질환자 및 질환 전단계 대상자의 건강관리, 건강증진, 생활습관개선,

	HCA발급관리, HRA발급관리, 영유아검진결과, 영유아발달평가결과, 영유아상담관리,세대조회,정밀검사비지원대상여부등)	보건교육 등 예방사업 수행
합리적의료이용지원 및 적정투약관리 대상자 관리내역 10년 8,092,626명	이름:필수, 집주소:필수, 직장주소:필수, E-Mail:필수, 집연락처:필수, 직장연락처:필수, 생년월일:필수, 주민번호, 건강:필수, 기타(가루약종, 가입자 구분, 가입자 주민번호SEQ, 가입자와의 관계, 거주지지사, 계단 오르기, 공단부담금, 내원일수, 단골병의원명, 단골약국명, 대상자 동거가족, 대상자 성별, 대상자 연령, 목욕하기, 보건복지부고시번호, 복용약물 수의 감소, 복용약물 종류의 감소 봉지, 생활습관기타, 소속지사, 수진자 성명, 수진자 주민번호, 수진자 주민번호SEQ, 시럽종, 식사량, 식사정도, 신장, 알, 알약종, 약국급여비, 약국이용이관 수, 약국조제일수, 약국진료비, 약국진료일수, 약국투약일수, 옷 벗고입기, 외래 내원일수, 외래 진료일수, 외래급여비, 외래이용기관 수, 외래이용일수 감소, 외래진료건수, 외래진료비, 외래진료일수, 요양기관 중복경로, 요양기호, 우편번호, 운동방법, 운동정도, 운동횟수, 월보험료, 음주량, 음주정도, 의료기관수 감소, 의료기관이용양상 기타, 일상생활수행능력 기타특이사항, 입원급여비, 입원이용기관 수, 입원일수, 입원진료건수, 입원진료비, 입원진료일수,중복간격(일), 중복투약일수, 증번호, 직역구분, 진료일자,처방(조제)기관 2, 처방(조제)기관1, 처방(조제)기관명1, 처방(조제)기관명2, 처방(조제)일자, 체중, 총의료기관 이용기관 수, 피면담자 관계, 혼자잠자리살피기, 휴대전화번호, 흡연량, 흡연정도, BMI지수, cc, 부상병, 비만상태, 상담의사 진료과명, 성분명1(영문), 성분명1(한글), 성분명 2(영문), 성분명2(한글), 약물 오.남용 이력, 약품명1, 약품명2, 약품코드1, 약품코드2, 주상병, 298상병)	가입자의 합리적 의료이용지원 및 적정투약관리(건강유지 및 증진 위하여 필요한 예방사업)
만성질환자 건강지원서비스 준영구 626,163,364명	필수: 이름, 집주소, 직장주소, E-mail, 집연락처, 직장연락처, 핸드폰, 생년월일, 고유식별번호(주민번호), 민감정보(건강, 성생활, 유전정보, 기타) 기타:등록자관리번호,등록자건강보험증번호,등록자관리지사, 서비스신청일자, 등록자사업장기호, 등록자보험료, 등록자면담방법, 등록자가입경로, 등록자진행상태, 등록자특이사항, 자가측정기 신청여부, 검진사후이관여부, 의원급만성질환관리제참여여부, 지역자원보건소여부, 제3자정보제공동의여부, 서비스제공면담결과구분, 면담일시, 기본상담일시, 면담상세내역, 측정기대여 및 반납일자, 서비스진행상태, 발송SMS내용, SMS발송전화번호, 안내문발송일자, 동거가족유무, 직업유무, 직업코드, 요구사정문항별 답변내용, 문제목록명, 면담내용, 피상담자반응구분, 개인별수행활동내용	만성질환자의 건강 유지 및 증진을 위한 건강지원서비 스 서비스 제공
학교밖청소년 건강검진 대상자 및 검진결과 내역 영구 19,380명	이름, 집주소, 집연락처, 핸드폰, 고유식별번호(주민번호) 민감정보(건강), 기타(대상자・수검자・검진의사 성명 및 주민번호, 검진결과 및 종합판정, 문진내역, 소견 및 조치, 검진기관명, 검진의사 면허번호, 검진(판정)일)	

④ 정보의 보유·활용에 대한 규정

건강검진기본법은 검진자료의 활용에 대한 규정을 두고 검진자료를 활용할 수 있는 범위를 다음과 같이 정하고 있다.

> 건강검진기본법 제18조(검진자료의 활용) ① 보건복지부장관 및 관계 중앙행정기관의 장은 국가건강검진을 통하여 얻은 검진자료를 다음 각 호의 목적으로 활용할 수 있다. <개정 2010.1.18.>
> 1. 건강정책 수립 및 이를 위한 통계자료의 작성
> 2. 지역사회 건강증진사업
> 3. 만성질환 관리 및 지원 사업
> 4. 국가건강검진 검사항목 및 검진주기의 평가 및 지침 개발
> 5. 국가건강검진제도 개선 및 평가를 위한 연구사업
> ② 보건복지부장관은 검진자료를 활용하여 건강상태 및 질병에 관한 통계를 생산하여 발표할 수 있다. <개정 2010.1.18.>
> ③ 제1항에 따라 검진자료를 활용함에 있어서 개인의 사생활의 비밀을 침해하지 아니하도록 정보를 보호하여야 한다.
> ④ 검진자료의 수집, 관리 및 통계의 작성이나 개인정보 및 사생활 보호 등에 필요한 세부사항은 대통령령으로 정한다.
>
> 건강검진기본법 시행령 제11조(검진자료의 수집·관리 및 통계의 작성) ① 법 제18조에 따른 건강검진자료를 활용한 통계의 작성은 「통계법」을 준용한다.
> ② 보건복지부장관 및 관계 중앙행정기관의 장은 법 제18조에 따라 개인정보가 포함된 건강검진자료를 활용하려는 경우에는 개인정보의 활용에 관하여 검진대상자의 동의를 받아야 한다.

규정에 의하면 검진자료는 통계를 생산하여 발표하는 데 활용될 수 있는데, 사생활의 비밀이 침해되지 않도록 해야 하고 보건복지부장관 및 관계 중앙행정기관의 장이 개인정보가 포함된 건강검진자료를 활용하려는 경우에는 검진대상자의 동의를 받아야 한다. 국민건강보험공단이 보유하는 건강검진자료도 건강검진기본법의 검진자료에 해당하므로 건강검진기본법의 규정이 적용될 것이다.

⑤ 건강검진 정보 등과 다른 정보의 연계·결합

현재 건강검진 정보 등을 다른 정보와 연계·결합하는 것에 대해서는 법령에 구체적 규정은 없다. 다만, 국민건강보험공단이 공개한 개인정보파일에서는 다른 정보시스템과 연계하고 있다고 밝히고 있고, 건강검진대상자 및 검진결과 내역의 경우 영구적으로 보유하면서 보건복지부, 국립암센터, 사회보장기관에 정보를 제공하여 연계하고 있다고 밝히고 있다. 그런데 그 외에 어떤 정보와 연계하고 있는지에 대해서는 밝히지 않고 있다. 건강검진 정보 등과 관련해서도 데이터의 연계·결합에 대해서 경우를 나누어 볼 수 있을 것이다.

첫째, 사회보험의 제공 및 유지 목적으로 자격정보와 다른 정보를 연계·결합할 필

요가 있는 경우가 있을 수 있다. 이때는 법적 근거가 별도로 없더라도 데이터의 연계·결합은 허용된다고 볼 수도 있을 것이다. 다만 개인정보보호의 원칙에 따라서 연계·결합이 이루어져야 할 것이다.

둘째, 그 외의 목적으로 데이터의 연계·결합이 이루어지는 경우는 원칙적으로는 해당 데이터의 연계·결합에 대한 법령의 명확한 규정이 마련되는 것이 바람직할 것이다. 이 경우 그 목적이 무엇인지에 따라 개인정보보호법의 원칙에 따른 연계·결합이 고려될 수 있을 것이다.

⑥ 정보 수집·보유의 적정성과 개인정보 보호원칙 준수 여부

첫째, 건강보험공단이나 보건복지부가 전 국민의 건강검진 정보를 수집하여 보유하고 있는 것이 타당한지, 특히 이를 영구적으로 보유하는 것이 타당한지는 의문이다. 최소수집의 원칙에 비추어 보았을 때, 건강보험공단이 전 국민의 건강검진 정보를 검진결과에 따른 특별한 조치를 취하는 데 사용하는 경우 외에는 이를 보유하는 것은 그 민감성에 비추어 문제가 있어 보인다. 특히 건강검진정보가 여러 범주로 나누어져 있는데, 만성질환자, 합리적 의료이용지원 및 적정투약관리 대상자, 학교 밖 청소년 등 매우 민감한 정보가 아닐 수 없다.

둘째, 건강검진 정보를 다른 정보와 연계하는 것도 신중해야 할 문제이다. 따라서 표본코호트 등의 정보로 임의로 정보를 연계하는 것은 법적 근거가 있어야 한다.

셋째, 개인정보의 정확성, 완전성 및 최신성 보장 및 열람청구권 등 정보주체의 권리 보장과 관련해서도 문제가 있다.

넷째, 건강검진 정보의 민감성에 비춰 본다면, 사생활 침해 최소화 및 익명처리의 원칙과 관련해서도 문제의 소지가 있다.

⑦ 개선 방향

이와 관련하여 다음과 같은 제도 개선이 필요하다.

1. 수집하는 정보의 범위를 법령이나 고시 등으로 규율하고, 공개할 필요가 있다.
2. 수집하는 정보는 필요한 최소한의 범위로 한다.
3. 보유 기간은 목적 달성에 필요한 최소한의 기간으로 하여, 법령이나 고시로 명시하고, 목적 달성 후 즉각 폐기하도록 한다.
4. 데이터의 연계, 제3자 제공 등을 엄격하게 제한해야 한다.

바. 보론 - 국세청의 과세정보와 과세자료의 비밀보호를 위한 법령의 규정

① 과세정보나 과세자료의 비밀보호 규정

우리 법률은 세금의 부과. 징수를 위하여 수집하는 자료에 대해서 목적 외 이용을 별도로 규정하는 법률 규정을 두고 있다. 즉, 국세기본법과 지방세기본법은 세무공무원에 대하여 세금의 부과, 징수를 위하여 업무상 취득한 자료 등인 과세정보에 대해서 목적 외 이용을 금지하는 등 비밀보호 의무를 규정하고 있다. 목적 외 용도로의 이용에 대해서는 영장주의를 적용하고, 엄격하게 제한하고 있다.

표 4-22 국세기본법의 입법례(제81조 비밀유지)

제81조의13(비밀 유지) ① 세무공무원은 납세자가 세법에서 정한 납세의무를 이행하기 위하여 제출한 자료나 국세의 부과·징수를 위하여 업무상 취득한 자료 등(이하 "과세정보"라 한다)을 타인에게 제공 또는 누설하거나 목적 외의 용도로 사용해서는 아니 된다. 다만, 다음 각 호의 어느 하나에 해당하는 경우에는 그 사용 목적에 맞는 범위에서 납세자의 과세정보를 제공할 수 있다. <개정 2014.1.1>
1. 지방자치단체 등이 법률에서 정하는 조세의 부과·징수 등을 위하여 사용할 목적으로 과세정보를 요구하는 경우
2. 국가기관이 조세쟁송이나 조세범 소추(訴追)를 위하여 과세정보를 요구하는 경우
3. 법원의 제출명령 또는 법관이 발부한 영장에 의하여 과세정보를 요구하는 경우
4. 세무공무원 간에 국세의 부과·징수 또는 질문·검사에 필요한 과세정보를 요구하는 경우
5. 통계청장이 국가통계작성 목적으로 과세정보를 요구하는 경우
6. 「사회보장기본법」 제3조제2호에 따른 사회보험의 운영을 목적으로 설립된 기관이 관계 법률에 따른 소관 업무를 수행하기 위하여 과세정보를 요구하는 경우
7. 국가행정기관, 지방자치단체 또는 「공공기관의 운영에 관한 법률」에 따른 공공기관이 급부·지원 등을 위한 자격의 조사·심사 등에 필요한 과세정보를 당사자의 동의를 받아 요구하는 경우
8. 다른 법률의 규정에 따라 과세정보를 요구하는 경우

② 제1항제1호·제2호 및 제5호부터 제8호까지의 규정에 따라 과세정보의 제공을 요구하는 자는 문서로 해당 세무관서의 장에게 요구하여야 한다. <개정 2014.1.1>

③ 세무공무원은 제1항 및 제2항을 위반하여 과세정보의 제공을 요구받으면 그 요구를 거부하여야 한다.

④ 제1항에 따라 과세정보를 알게 된 사람은 이를 타인에게 제공 또는 누설하거나 그 목적 외의 용도로 사용해서는 아니 된다.

⑤ 이 조에 따라 과세정보를 제공받아 알게 된 사람 중 공무원이 아닌 사람은 「형법」이나 그 밖의 법률에 따른 벌칙을 적용할 때에는 공무원으로 본다.

근거과세와 공평 과세를 실현하고 세무행정의 과학화와 성실한 납세 풍토를 조성하기 위해서 과세자료의 제출을 과세자료를 보유한 기관에 의무화하고 있는데, 이들 과세자료제출기관이 제출한 과세자료에 대해서도 목적 외 용도로 사용하지 못하도록 하고, 비밀 유지 의무를 부과하는 등의 법률 규정을 두고 있다(과세자료의 제출 및 관리에 관한 법률 제11조, 지방세기본법 제132조). 역시 목적 외 용도로의 이용에 대해서는 영장주의를 적용하고, 엄격하게 제한하고 있다.

표 4-23 과세자료의 제출 및 관리에 관한 법률의 입법례

제11조(비밀유지 의무) ① 세무관서의 소속 공무원은 이 법에 따라 제출받은 과세자료(제6조에 따라 제출받은 금융거래정보 및 제8조에 따라 수집한 자료를 포함한다)를 타인에게 제공하거나 누설하거나 목적 외의 용도로 사용하여서는 아니 된다. 다만, 「국세기본법」 제81조의13제1항 단서 및 같은 조 제2항에 따라 제공하는 경우에는 그러하지 아니하다. ② 세무관서의 소속 공무원은 제1항을 위반하는 과세자료의 제공을 요구받으면 이를 거부하여야 한다. ③ 제1항 단서에 따라 과세자료를 제공받은 자는 이를 타인에게 제공하거나 누설하거나 그 목적 외의 용도로 사용하여서는 아니 된다.

반면, 개별 의료기관이 수집한 건강정보에 대해서는 비밀보호 조항이나 형법의 비밀침해죄 등의 보호대상으로 입법되어 있고, 제3자 제공에 대해서도 법률로서 엄격하게 제한하고 있지만, 이를 국민건강보험공단이나 건강보험심사평가원이 업무처리를 위하여 수집한 것에 대해서는 그와 같이 법률로서 제3자 제공에 대해서 제한하는 규정이 없다. 단, 국민건강보험법은 공단, 심사평가원 및 대행청구단체에 종사하였던 사람 또는 종사하는 사람이 가입자 및 피부양자의 개인정보를 직무상 목적 외의 용도로 이용하거나 정당한 사유 없이 제3자에게 제공하는 행위에 대해서만 처벌규정을 두고 있을 뿐이다.

국민건강보험공단은 매우 민감한 건강정보를 수집, 처리하고 있다는 점에서 수집한 정보를 목적 외로 이용해서는 안 된다는 법률의 규정을 마련하는 것이 바람직할 것이다.

② 제도 개선

국민건강보험공단과 심사평가원은 과세정보나 과세자료에 대한 비밀보호, 목적 외 이용의 금지에 대한 법률 규정을 명확하게 두고 있는 점에 비추어 아래와 같은 제도 개선이 바람직하다.

국민건강보험법에 국민건강보험공단과 심사평가원이 수집한 정보에 대해서 원칙적으로 목적 외 용도로 사용해서는 안 된다는 점을 밝히고, 목적 외 이용이 허용되는 경우에 대해서는 허용되는 사유를 미리 명시적으로 규정해 놓는 것이 바람직하다.

(2) 건강보험심사평가원

가. 건강보험심사평가원의 업무

건강보험심사평가원은 요양급여비용을 심사하고 요양급여의 적정성을 평가하기 위

하여 국민건강보험법에 의해 설립된 법인이다(국민건강보험법 제62조).

국민건강보험법은 그 외에도 심사평가원에 대하여 심사기준 및 평가 기준의 개발, 요양급여비용 심사, 요양급여 적정성 평가 업무와 관련된 조사연구 및 국제협력, 의료급여법, 노인장기요양보험법에 의해 위탁받은 급여비용 심사와 의료의 적정성에 대한 평가 업무, 심사 청구와 관련된 소프트웨어의 개발·공급·검사 등 전산 관리도 하도록 하고 있고, 요양급여의 적정성 평가 결과의 공개 업무도 하도록 하고 있다(제63조, 시행령 제28조).

실제로 국민건강보험법 제14조, 제47조, 제48조 및 제63조에 의하여 의료기관과 약국은 급여비용 심사·지급·대상 여부 확인·사후관리 및 요양급여의 적정성 평가·가감지급 등을 위하여 건강정보를 국민건강보험공단 또는 건강보험심사평가원으로 제공하도록 하고 있다. 의료급여법도 요양기관이 요양급여비용을 청구하기 위해서 공단과 심사평가원에 요양급여의 내역을 제공하도록 하고 있고, 노인장기요양보험법도 공단과 심사평가원에 내역을 제공하도록 하고 있다.

그림 4-7 건강보험심사평가원의 업무

나. 건강보험심사평가원에서 수집하는 정보

이런 법적 근거를 바탕으로 건강보험심사평가원이 수집하는 정보는 범주를 나누어 보면, ① 진료정보, ② 의약품 정보,[257] ③ 치료재료 정보, ④ 의료자원 정보, ⑤ 비급

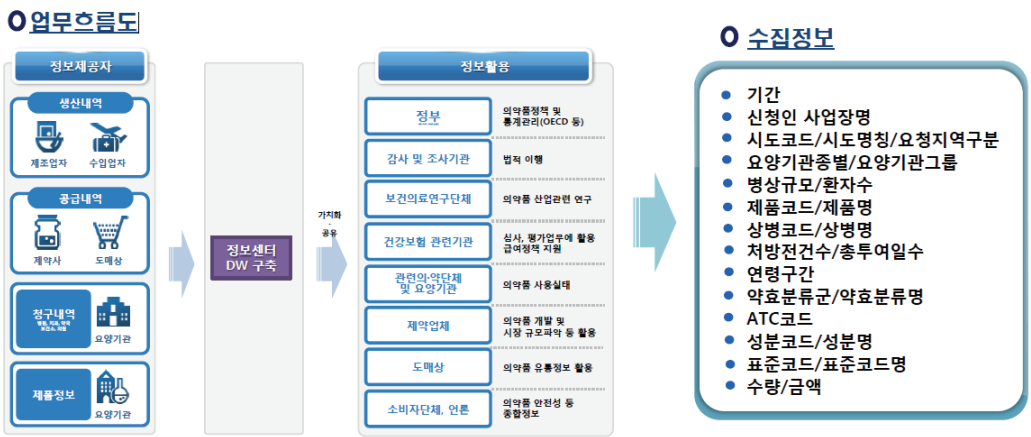
257) 예를 들어 건강보험심사평가원은 의약품 유통정보 표준화 및 선진화를 통한 공정하고 투명한 의약품 유통환경 조성하기 위하여 의약품 유통정보 DB도 구축, 운영하고 있다. 여기서 수집하는 정보는 아래와 같다.

여 정보, ⑥ 평가정보로 나누어 볼 수 있다. 국민건강보험법 시행령은 아래 표와 같은 정보를 제출해 달라고 요청할 수 있다고 규정하고 있다.

표 4-24 심사평가원이 요청할 수 있는 자료

국민건강보험법 시행령 제69조의 2 제1항, 별표 4의3
가. 법 제96조의2에 따라 요양기관이 보존하여야 하는 서류 나. 「의료법」 제22조에 따라 보존하여야 하는 진료에 관한 기록 다. 「약사법」 제29조 및 제30조에 따라 보존하여야 하는 처방전 및 조제기록부 라. 가목부터 다목까지에서 규정한 사항 외에 요양급여의 내용에 관한 자료 및 이를 증명하는 서류 마. 법 제43조에 따른 신고사항 등 요양기관의 현황과 관련한 사실을 확인하기 위해 필요한 자료 바. 「약사법」 및 「의료기기법」에 따른 약제·치료재료·의료기기의 제조·수입·판매·도매 업무를 하는 자의 제조·수입·판매·도매 현황 및 관련 서류, 원가 관련 자료 등 요양급여비용의 결정·조정과 관련한 자료 사. 「주민등록법」에 따른 주민등록자료 아. 「출입국관리법」에 따른 출입국자료 자. 「국세기본법」 또는 「지방세기본법」에 따른 과세자료 차. 「국민건강보험법」, 「의료급여법」, 「보훈보상대상자 지원에 관한 법률」, 「산업재해보상보험법」 등에 따른 자격, 급여제공 또는 비용지원, 급여의 제한·정지에 대한 자료 카. 다음 각 목의 자에 대한 면허, 자격 및 행정처분 등에 대한 자료 1) 「의료법」에 따른 의사, 치과의사, 한의사, 조산사, 간호사 및 간호조무사 2) 「약사법」에 따른 약사 및 한약사 3) 「의료기사 등에 관한 법률」에 따른 임상병리사, 방사선사, 물리치료사, 작업치료사, 치과기공사, 치과위생사 및 의무기록사 4) 「사회복지사업법」에 따른 사회복지사 5) 「국민영양관리법」에 따른 영양사 6) 「식품위생법」에 따른 조리사 7) 「정신보건법」에 따른 정신보건임상심리사, 정신보건간호사 및 정신보건사회복지사 8) 「원자력안전법」에 따른 방사성동위원소취급자 및 방사선취급감독자 9) 그 밖에 다른 법령에 따라 면허를 받거나 자격을 인정받은 자로서 요양급여 관련 업무에 종사하는 자 타. 요양기관, 「의료급여법」에 따른 의료급여기관, 「의료법」에 따른 의료기관, 「약사법」에 따른 의약품도매상, 의약품·의약외품의 제조업자·품목허가를 받은 자·수입자·판매업자, 「의료기기법」

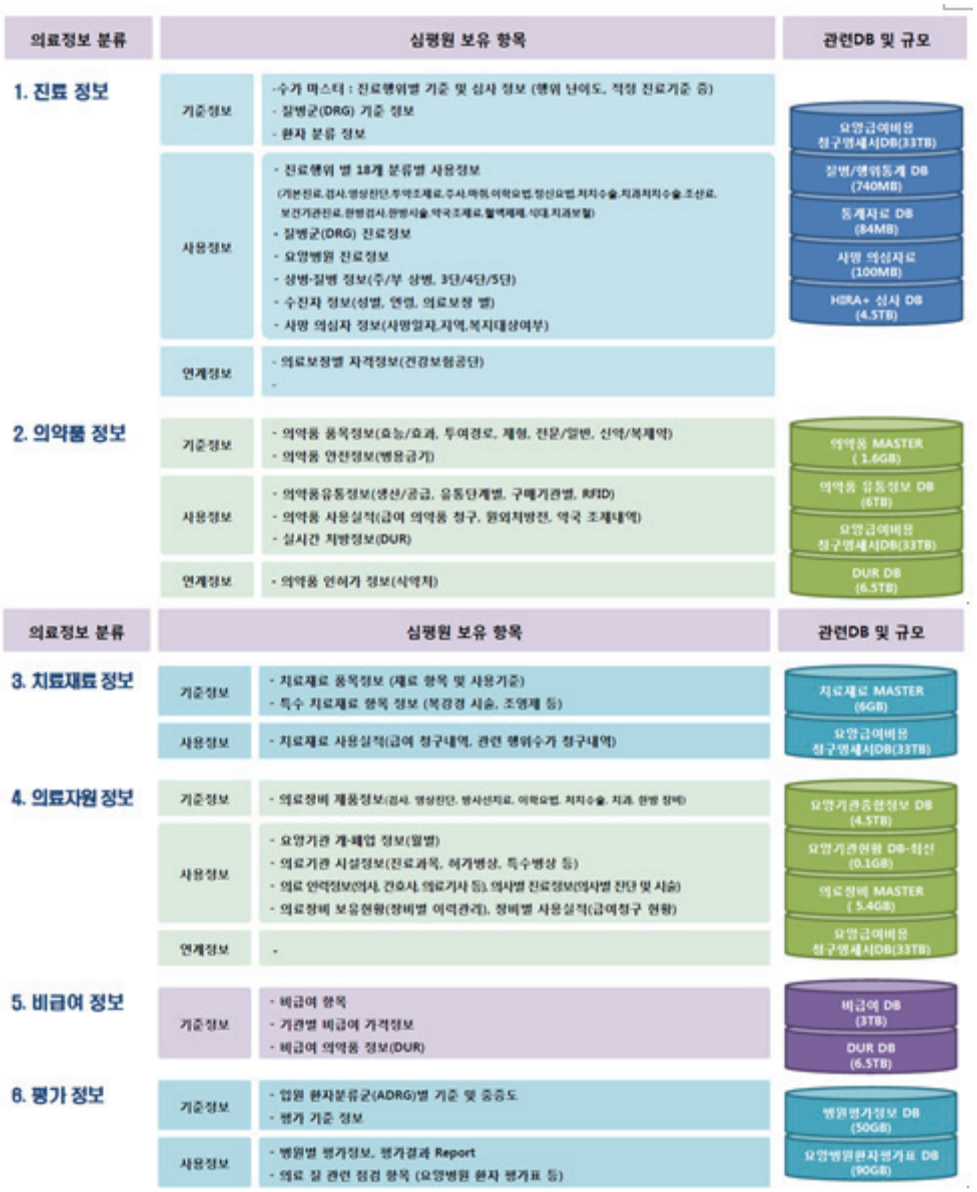
그림 4-8 건강보험심사평가원 수집정보



에 따른 의료기기취급자, 「식품위생법」에 따른 집단급식소 운영자, 「마약류 관리에 관한 법률」에 따른 마약류취급자 등에 대한 업무정지·허가취소 등 처분에 대한 자료
파. 그 밖에 국가, 지방자치단체, 요양기관, 「보험업법」에 따른 보험회사 및 보험료율 산출 기관, 「공공기관의 운영에 관한 법률」에 따른 공공기관, 그 밖의 공공단체 등이 보유한 자료로서 법 제63조제1항 각 호의 업무를 위해 필요한 자료

심사평가원이 소개하고 있는 정보의 규모와 항목은 다음과 같다.

그림 4-9 건강보험심사평가원 정보의 규모와 항목



* 출처: 건강보험심사평가원 보유 정보 현황과 활용(건강보험심사평가원 1차 교육자료, 2015.8) p.14-15.

건강보험심사평가원이 수집·처리하는 정보의 특징은 다음과 같다고 한다.[258] 전 국민, 모든 요양기관의 의료서비스 내용을 대표하는 기본 자료로서 일반화가 쉽다. 행위별수가제를 기본으로 청구데이터를 구축하여 처방약을 포함한 세부적인 의료행위 내역을 포함하고 있다. 희귀질환, 합병증, 약물 부작용, 희소하게 이용되는 진료행위에 대한 일반 정보 확보가 가능하다. 제한적이고 실험적 환경이 아닌 실제 의료환경·제도를 반영한 정보이다.

다. 진료비 심사와 청구정보

심사평가원은 의료 공급자가 진료비를 청구하면 국민건강보험법 등에서 정한 기준에 의해 진료비와 진료 내역이 올바르게 청구되었는지, 의·약학적으로 타당하고 비용효과적으로 이루어졌는지 확인한다. 이 과정에서 건강보험심사평가원은 방대한 양의 건강정보를 수집, 처리한다.

표 4-25 건강보험심사평가원의 청구명세서 정보

청구명세서 준영구 건강보험 17,550,432,183명 의료급여 53,124,603 명 보훈국비환자 174,513명	이름:필수, 건강:필수 (요양기호, 보험자코드, 청구명일련번호, 청구형태코드, 서식번호, 수진자주민번호, 수진자성명, 증번호, 가입자성명, 내원일수, 청구정액정률구분, 공상구분, 의료급여종별코드, 청구대불금, 청구장애인기금, 청구요양급여비용총액, 청구본인부담금, 청구보험자부담금, 진료결과구분, 내원일자, 요양일수, 입원경로구분, 처방전발행기관, 당월요양개시일자, 진료과목코드, 전접수번호, 청구전명일련번호, 청구구분, 전지급불능사유코드, 최초입원일자, 재료및항독소구분, 재료 및 항독소금액, 물리치료실시횟수, 야간조산여부, 다태아구분, 총처방일수, 투약일수, 의료급여정액제구분, 청구적용수가구분, 청구본인부담상한액초과금, 진찰횟수, 청구진료비총액, 청구보훈부담액, 청구100분의100본인부담액, 청구비급여액, 본인부담적용코드, 진료확인번호갯수, 입·퇴원 방법, 내원시 증상 및 발생시각, 수술·치료(약제투여) 전후 환자상태, 약제 투여시각, 검사 종류 및 내용, 수술·치료 형태 및 시각(54))

진료비 심사 및 청구정보의 보유 및 활용과 관련하여 다음과 같은 제도 개선이 바람직하다.

- 진료비 심사 청구에 관한 정보는 당해 목적의 범위에서 보유하는 것이 바람직한데, 준영구로 보유하는 것은 문제이다.
- 청구명세서의 항목도 필수 불가결한 것으로 검토할 필요도 있다.

258) 건강보험심사평가원 보유 정보 현황과 활용(건강보험심사평가원 1차 교육자료, 2015.8) p.6.

(3) 건강보험의 건전성 유지, 연구 목적 2차적 이용

가. 법적 근거

국민건강보험공단은 보험급여의 관리, 건강보험에 관한 조사연구, 건강관리 사업[259]을 업무로 담당하도록 법률에 규정되어 있다. 현재 국민건강보험공단은 이를 근거로 연구 목적이나 2차적 이용을 목적으로 하여 다양한 정보를 연계·결합 등의 방법으로 생성하여 활용하고 있다.

한편 건강보험심사평가원은 심사기준 및 평가 기준의 개발, 그에 따른 업무와 관련된 조사연구 및 국제협력, 요양급여비용의 심사 청구와 관련된 소프트웨어의 개발·공급·검사 등 전산 관리, 환자 분류체계의 개발·관리 등의 업무를 관장하는데, 이를 위하여 건강정보를 활용할 수 있는지가 문제 된다.

나. 국민건강정보 데이터베이스

① 보유 정보 및 보유 기간

국민건강보험공단은 현재 40,833,684,272명의 개인에 대한 건강정보를 삭제하지 않고 보유하고 있다고 한다. 국민건강보험공단은 이 정보파일에 대해서 '전 국민의 자격 및 보험료, 건강검진결과, 진료내역, 노인장기요양보험 자료, 요양기관 현황, 암 및 희귀난치성질환자 등록정보 등 1조 3천억 건에 달하는 방대한 빅데이터를 말한다'고 설명하고 있다.

국민건강보험공단이 공개한 개인정보파일에 기재된 보유 항목은 아래와 같다. 해당 항목은 해당 공공기관에서 개인정보 처리 관련 업무를 담당하는 부서와 일부 관련부서에서 공동으로 활용하고 있다고 한다. 그리고 해당 정보는 시스템 연계를 통해서 수집하고 있다고 한다. 이와 관련한 개인정보의 열람 요구를 접수·처리하는 부서는 국민건강보험공단 본부, 지역본부, 지사라고 한다.

표 4-26 국민건강정보 데이터베이스

개인정보파일 명칭	개인정보 내역	파일 운영 목적
DW건강보험통계 DB	○ 건강보험 적용인구, 세대정보 : 주민등록번호, 장애등급, 장애유형, 장애등록일자, 체류자격, 국가, 주소, 실거주지주소, 건강보험증번호,	국민건강

259) 가입자 및 피부양자의 건강관리를 위한 전자적 건강정보시스템의 구축·운영, 생애주기별·사업장별·직능별 건강관리 프로그램 또는 서비스의 개발 및 제공, 연령별·성별·직업별 주요 질환에 대한 정보 수집, 분석·연구 및 관리방안 제공, 고혈압·당뇨 등 주요 만성질환에 대한 정보 제공 및 건강관리 지원, 지역보건의료기관과의 연계·협력을 통한 지역별 건강관리 사업 지원.

준영구 40,833,684,272 건	사업장, 행정전산망세대주주민등록번호, 국가유공자, 국가유공자자료구분, 장애인세대여부, 단독세대여부, 외국인세대구분, 만성희귀질환세대여부,건강보험적용세대수, 개인성명, 출입국일자, 출국사유, 출입국구분, 전역여부, 군기관유형, 군복무유형, 군전환복무여부, 군입대년월, 군전역년월, 국가유공자주민등록번호, 국가유공자명, 국가유공자등록일자, 국가유공자제외일자, 상이등급세부사항, 자격취득일자, 자격상실일자, 자격취득구분, 상실구분, 세대주구분, 보험급여정지사유, 시도, 시군구, 행정동, 법정동, 리 ○ 사업장 및 지역보험료 부과정보, 징수, 체납 정보 : 사업장, 사업자등록번호, 개업일자, 대표자주민등록번호, 사업자소득구분, 종합소득종류, 사업자소득금액, 건강보험증번호, 주소, 총소득금액, 자동차최초등록일자, 자동차연식구분, 변경일자, 지역보험료자동차번호, 차종, 자동차용도, 자동차배기량, 자동차적재량, 승차정원수, 자동차연식구분, 자동차세액, 총재산금액, 보험료수납일자, 총납부금액, 징수당월보험료납부금액, 과년도보험료납부금액, 환급지급보험료, 환급지급가산금, 보험료경감, 보험료면제구분, 직장보수월액, 직장산정보험료, 직장경감보험료, 직장면제보험료, 직장조정특례감면보험료, 직장가입자분할고지보험료, 직장수시정산추가보험료, 직장수시정산반환보험료, 직장퇴직정산보험료, 직장중간정산보험료, 직장연말정산보험료, 직장본인부담금환급금, 직장가입자부담보험료, 직장사업장부담보험료, 직장총고지보험료, 직장국가부담보험료, 건강보험장기요양구분, 상계충당금액, 법인상계충당금액, 상계충당보험료, 직장납부보험료, 직장납부가산금, 직장보험료과오납금액, 직장가산금과오납금액, 국가부담납부보험료, 지역정기산정보험료, 지역산정보험료, 지역경감보험료, 지역정지보험료, 지역면제보험료, 지역도서벽지경감금액, 지역농어촌경감금액, 지역농어업인경감금액, 지역세대경감금액, 지역경감반영보험료, 지역정기부과보험료, 지역정산반영보험료, 지역정산부과보험료, 지역고지보험료, 지역최종부과보험료, 지역총체납보험료, 결손처분보험료, 압류상태, 압류유형구분, 압류일자, 압류취소사유, 압류취소일자 ○ 일반건강검진, 암검진, 생애전환기 검진 : 주민등록번호, 건강검진유형코드, 외국인여부ID, 체류자격코드, 국가코드, 건강검진일자, 검진비지급일자, 장애유형코드, 장애등급코드, 의사면허번호, 의사주민등록번호, 일반 건강검진 결과코드, 일반건강검진 문진결과코드, 생애전환기 검진결과코드, 생애전환기 문진결과코드, 생애전환기 암검진 결과코드, 생애전환기 건강검진암검사문진 결과코드, 영유아건강검진 대상자, 과거병력 문진내역코드, 영유아검진구강검사 결과코드, 건강검진구강검사 결과 코드, 건강검진대상자코드, 암검진대상자코드, 의료급여암검사대상자코드, 생애전환기건강검진대상자코드 ○ 노인장기요양 인정신청 및 급여 : 장기요양관리번호, 개인ID, 주민등록번호, 인정여부, 1차판정등급코드, 인정등급코드, 신청일자, 대상자구분코드, 인정등급판정일자, 의료급여보장기관코드, 의료급여시설기관코드, 의료급여종별코드, 의료급여유형코드, 거주지사시도시군구코드, 장애등급코드, 장애유형코드, 장애등록일자, 체류자격코드, 국가코드, 직역상세코드, 동거여부ID, 동거인관계코드,	증진과 건강보험의 지속가능성 관련 연구목적

	신청자건강보험증번호, 가입자구분코드, 만성희귀질환세대여부ID, 산정보험료, 보수월액, 도서벽지거주여부, 자격취득일자, 자격취득사유코드, 자격상실일자, 상실사유코드, 사망여부ID, 사업장관리번호, 주거형태코드, 동거인여부, 동거인유형별코드, 자격취득구분코드, 실거주지주소코드, 노인장기요양신청일, 장기요양보호자관계코드, 인정변경신청사유코드, 노인성질병코드, 장기요양변경전등급분류코드, 유효기간시작일자, 유효기간종료일자, 대리인유형구분코드, 대리인신청관계기타코드, 의사소견서발급기관기호, 의사소견서발급구분코드, 요양기관진료과목코드, 의사소견서치매노인코드, 인정조사치매여부, 종전등급코드, 장기요양인정점수, 건강보험재산상태코드, 월소득금액, 주요질병코드, 장기요양주요질병코드, 감경적용기초코드, 감경적용여부, 2차판정급여종류, 질환 여부, 갱신이전인정등급코드, 판정결과, 장기요양인정조사표코드, 의사소견서코드, 장기요양기관 인력현황코드, 요양보호사발급대장코드, 복지용구 정보, 한도관리내역, 이용상담내역,급여계약내역서 정보, 이용지원 상담내용을 관리, 노인장기요양급여비용 명세서코드, 노인장기요양급여비용 서비스내역코드 ○ 요양급여비용 지급 및 중증암환자 등록, 현금급여 지급 : 건강보험증번호, 주민등록번호, 현금급여지급현황코드, 급여사후결정코드, 가입자피부양자구분, 장애등급, 장애유형, 체류자격, 외국인여부ID, 주소, 사업장, 진료결과, 요양급여구분, 진료과목, 공상산재구분, 수술구분, 정액정률구분, 진료개시일자, 요양기관기호, 심사결정DRG번호, 요양급여입내원일수, 요양급여투약일수, 심사결정총진료비, 심사결정보험자부담금, 심사결정본인부담금, 심사결정보험자부담금차감액, 심사결정본인부담액상한초과금액, 심사결정식대기본금액, 심사결정식대가산금액, 본인부담금환급금, 심사조정건수, 요양급여비용총삭감액, 산정특례대상자 등록번호, 주상병, 부상병, 공무상요양비지급코드, 중증암등록환자코드 ○ 의료급여 적용, 세대, 급여비용 : 주민등록번호, 국가유공자,국가유공자자료구분, 의료급여종별, 의료급여유형, 행정동의료보호기관, 의료급여증번호, 세대주주민등록번호, 세대주나이, 세대주성별, 장애등급, 장애유형, 체류자격, 외국인여부ID, 주소, 진료결과, 요양급여구분, 진료과목, 공상산재구분, 수술구분, 정액정률구분, 진료개시일자, 요양기관기호, 심사결정DRG번호, 요양급여입내원일수, 요양급여투약일수, 심사결정총진료비, 심사결정보험자부담금, 심사결정본인부담금, 심사결정보험자부담금차감액, 심사결정본인부담액상한초과금액, 심사결정식대기본금액, 심사결정식대가산금액, 본인부담금환급금, 심사조정건수, 요양급여비용총삭감액, 산정특례대상자 등록번호, 주상병, 부상병	

이 정보는 주민등록번호와 건강보험증번호와 같이 고유식별정보인 개인식별정보를 모두 포함하고 있기 때문에 다른 정보와의 연계도 가능한 것으로 보인다.

아울러 이 정보는 국민건강보험공단이 자격정보, 징수정보, 심사청구 명세서 정보,

요양급여내역, 건강검진정보, 희귀질병 정도 등 국민건강보험공단이 각각 다른 목적으로 수집하여 보유하고 있는 정보를 연계한 것으로 보인다.

이는 자격정보나 징수를 위한 정보, 행정정보는 물론, 건강검진 정보를 포함해서 그야말로 전 국민의 모든 건강 관련 정보와 사회적 활동과 재산 등에 관련된 정보가 망라된 것으로 볼 수 있다.

② 보유 목적

국민건강보험공단은 국민건강 증진과 건강보험의 지속가능성 관련 연구목적이라고 목적을 밝히고 있다. 그에 대한 법적 근거로는 9호(9. 건강보험에 관한 조사연구 및 국제협력)를 들고 있다.

③ 개인정보 수집 및 연계에 대한 동의 여부와 보유의 적법성

민감정보의 처리는 정보주체에게 각 고지사항을 알리고 다른 개인정보의 처리에 대한 동의와 별도로 동의를 받은 경우나, 법령에서 민감정보의 처리를 요구하거나 허용하는 경우에만 허용된다(개인정보보호법 제23조 제1항).

그런데 국민건강정보 데이터베이스로 다양한 목적으로 수집된 개인의 건강정보와 자격정보, 징수정보 등의 개인정보가 연계·결합되고 이용·제공되는 것에 대하여 해당 개인에게 동의를 받은 사실은 없다. 따라서 법적 근거가 있어야 하는데 국민건강보험법 제14조 제1항 제9호에서 국민건강보험공단의 업무로 '건강보험에 관한 조사연구 및 국제협력'이 규정되어 있지만, 이 규정이 개인정보보호법 제23조 제1항의 '민감정보의 처리를 허용하는 법령의 규정'으로 보기에는 난점이 있다.

④ 개인정보 보호 원칙의 준수 여부

아울러 이 규정을 법적 근거로 본다고 하더라도 개인정보 보호 원칙은 준수해야 하는데, 다음과 같은 점에서 문제가 있다.

첫째, 목적 적합성과 최소수집의 원칙을 준수했다고 보기 어렵다. 연구 목적을 위하여 모든 국민의 건강정보를 연계·결합하는 것이 필요한지는 의문이고, 익명처리가 되지 않은 상태로 영구보존되어야 하는지도 의문이다.

둘째, 개인정보주체에게 정보를 제공할 의무를 준수했다고 보기도 어렵다. 개인정보의 수집·보유, 연계의 범위, 연계의 목적, 보유 기간 등에 대한 정보 제공이 이루어지지 않고 있다. 따라서 열람, 삭제, 정정 등의 권리를 행사할 수 있는 가능성도 차단되

고 있다.

셋째, 개인정보주체의 처리중지나 삭제에 대한 권리가 보장되어 있지 않다. 연구 목적의 데이터베이스를 구축하는 것이라면 개인정보주체의 선택에 따라서 처리중지나 삭제를 요구할 수 있도록 권리를 보장해 주는 것이 바람직하다.

⑤ 제도 개선

이와 관련하여 다음과 같은 제도 개선이 바람직하다.

1. 연구 목적의 건강정보 활용을 위해서는 개인정보주체의 동의가 없는 한 현재의 국민건강보험법에 의한 개인정보 활용은 적법한 근거라고 보기 어렵다.
2. 모든 개인의 건강정보를 포괄적으로 연계하여 준영구로 보유하는 것은 개인정보보호원칙에 위반한다.
3. 연구 목적의 건강정보 활용을 하려면 연구목적에 필요한 최소한의 정보를 보유하고, 폐기해야 한다.
4. 연구 목적의 건강정보 활용을 위해 최소한의 정보를 활용하더라도 익명화 등 안전조치를 취하는 것이 바람직하다.

다. 코호트 DB

① 코호트 DB의 특징

코호트 연구(Cohort study, 추적 연구)는 전향성 추적조사에 의한 연구를 말한다. 보통 특정 요인에 노출된 집단과 노출되지 않은 집단을 추적하여 어떤 원인이 어떤 결과를 가져오는가를 추적하여 연구하는 방법이다. 코호트 연구는 비교 위험도와 귀속 위험도를 직접 측정할 수 있고, 객관적이며, 부수적으로 다른 질환과의 관계 파악이 가능하며, 시간적인 선후관계를 알 수 있다는 점 등이 장점으로 꼽히는 반면, 시간과 비용이 많이 들고, 대상자가 중도에 탈락하기 쉽다는 점이 단점으로 꼽힌다.

코호트 정보는 개인에 대한 추적정보이기 때문에 그 민감성이 더욱 크다.

② 현재 국민건강보험공단이 수집·축적하고 있는 코호트 DB

현재 국민건강보험공단은 아래와 같이 5종의 코호트 정보를 생성하여 보유·활용하고 있다. 그 내용을 보면 5종의 코호트 DB 중 규모가 가장 큰 전 국민 표준코호트 DB의 경우, 전 국민의 2%에 달하는 100만 명을 추출하여, 무려 14년간의 사회, 경제적 현황과 장애, 사망, 의료이용 현황, 모든 진료내역은 물론, 건강검진 등의 상세한 정보(자격 및 보험료, 출생 및 사망, 모든 진료내역, 건강검진 내역)를 데이터베이스로 구축한 것이다.

표 4-27 코호트 자료 DB

	대상	추적 기간	자료 건수(천건)	내용
표본코호트DB	2006년 1년간 건강보험가입자 및 의료급여수급권자 자격을 유지한 전국민 중 2% 추출(100만명) 전국민 모집단의 2%, 성・연령・가입자구분・보험료 분위・지역별 층화추출	2002 ~ 2015년(14년)	2,619,397	사회・경제적 현황(자격 및 보험료, 장애 및 사망), 의료이용 현황(진료 및 건강검진), 요양기관 현황
건강검진 코호트DB	2002년 자격유지자 중 2002~2003년 40~79세 일반건강검진 수검자 중 약 51만명(모집단의 10%)	2002 ~ 2015년 (14개년)	2,087,629	사회・경제적 자격 변수(장애 및 사망 포함), 의료이용(진료 및 건강검진)현황, 요양기관 현황
노인코호트DB	2002년 자격유지자 중 만 60세 이상 대상자 중 약 55만명 (모집단의 10%)	2002 ~ 2015년 (14개년)	2,749,045	사회・경제적 자격 변수(장애 및 사망 포함), 의료이용(진료 및 건강검진)현황, 요양기관 현황, 노인장기요양 서비스 현황
영유아검진코호트DB	2007년 자격유지자 중 15~64세(생산가능인구) 직장 여성 약 18만명(가입자 약 360만 명의 5% 무작위 추출)	2007 ~ 2015년 (9개년)	233,688	사회・경제적 자격 변수(장애 및 사업장 포함), 의료이용(진료 및 건강검진)현황, 요양기관 현황
직장여성 코호트DB	영유아검진 1~2차를 1회 이상 받은 전체 수검자 중 2008~2012년 출생자를 추출하여 각 출생연도별 5% 표본 추출	2008 ~ 2015년 (8개년)	368,226	사회・경제적 자격 변수(장애 및 사망), 의료이용(진료 및 건강검진)현황, 요양기관 현황

정보의 범위는 건강정보뿐만 아니라 자격과 징수정보, 건강검진 정보를 포함하여 연계 가능한 정보를 집대성하여 연계한 것이다.

자료구축 시 성, 연령, 지역(읍면동), 건강보험가입구분, 건강보험 부과기준 소득분위를 층으로 비례배분하여 계통추출된 자료이다. 표본코호트DB는 크게 자격DB, 진료DB, 건강검진DB, 요양기관DB로 구성되어 있다. 자격DB는 건강보험가입자 및 의료급여수급권자의 성, 연령대, 지역, 가입자 구분, 소득분위 등 대상자의 사회경제적 변수 및 장애, 사망관련 총 14개 변수로 구성되어 있다. 여기서 외국인은 제외되어 있다. 추가적으로 통계청 사망원인(세분류, 중분류) 시군구 자료가 있는데 이것은 필요한 경우 공단에서 검토 후에 제공한다. 진료DB는 대상자가 요양기관에 방문하여 진료 등을 받은 내역에 대해 요양기관으로부터 요양급여가 청구된 자료로 외과 보건기관, 치과·한방, 약국명세서, 진료내역, 상병내역, 처방교부상세내역의 10개의 세부DB

로 구성되어 있다. 건강검진DB는 건강검진 주요 결과 및 문진에 의한 생활습관 및 행태관련 자료로 구성되어 있다. 2009년 검진제도 개편에 따른 주요검진 및 문진항목 변경으로 인해 2009년 이전과 이후의 자료로 구성되어 있다. 요양기관DB는 요양기관의 종별, 설립구분별, 지역(시도)별 현황 및 시설, 장비, 입력관련 자료로 총 10개의 변수로 구성되어 있다.[260]

국민건강보험공단은 이 코호트 자료에 대해서 전 국민의 진료내역을 담고 있어 세계에서도 찾아보기 힘든 상당한 가치를 지닌 데이터들이라고 소개하고 있다.

건강검진 코호트DB는 수검자 중 10%, 노인코호트는 대상자 중 10%, 직장여성 코호트 DB는 5%, 영유아 검진 코호트 DB 5%를 추출하는 등 그 대상도 방대하고, 코호트 정보를 추적하여 축적한 기간이 14년 ~ 8년간으로 장기간인 데다가, 수록 정보도 사회경제적인 정보와 의료와 건강검진 정보까지 포괄하고 있어서 그야말로 개인의 내밀한 사생활이나 밝혀져서는 곤란한 민감한 정보가 고스란히 담겨 있는 정보이다.

국가에서 한 개인에 대해서 이와 같은 정보를 연계하여 보유해도 무방한 것인지 의문이 들 수 있는 정보들이다.

③ 보유 목적

5종의 코호트 DB의 보유 목적은 연구 목적이라고 밝히고 있다.

④ 정보주체의 동의 여부, 정보주체에게 통지 여부

일반적으로 코호트 연구는 당사자의 동의를 받아서 진행한다. 예를 들어 한국보건연구원에서 하는 코호트 사업인 한국인유전체역학조사사업(이하 'KoGES')은 동의를 기반으로 이루어진다.[261] 그래서 40~69세 남, 녀 일반 인구 집단을 기반조사 대상자로 하는데, 먼저 사업 참여와 시료 수집에 대한 동의를 받은 후 동의를 한 대상자에게 건강이나 질병력, 생활습관 (운동, 식습관) 관련한 다양한 설문조사, 신체계측, 검진을 통한 자료 수집, 혈액과 뇨를 채취한 임상검사 자료와 인체자원(혈청, 혈장, DNA)을 확보하는데, 이를 2~4년 주기로 반복 추적한다.

그러나 국민건강보험공단이 구축한 5개의 코호트 DB는 당사자의 동의를 기반으로 하지 않고, 일방적으로 추출한 자료를 바탕으로 생성된 것으로 보인다. 그 어디에서도 당사자들에게 동의를 받았다는 정보나 받은 동의의 내용을 밝히고 있지 않다.

---

260) 표본코호트DB(국민건강보험)와 조사자료(통계청)의 자료연계에 따른 심층분석 - 진료자료 및 건강검진자료와 생활시간조사의 확률적 연계 가능성 검토- 임찬수 · 최재혁 · 김경미, p41.

261) '한국인 유전체역학조사사업' 수집자료 통합·정제 지침서 공개(질병관리본부 유전체센터 유전체역학과 송대섭, 김연정), p1.

그뿐만 아니라 해당 정보주체에게 코호트DB로 해당 정보주체의 여러 가지 정보가 연계되어 추적, 축적되어 있고, 그 정보들이 제3자에게도 제공되고 있다는 점에 대해서 알리고 있지 않은 것으로 보인다.

⑤ 코호트 자료의 개인식별 가능성

국민건강보험공단은 코호트 자료를 만들면서 소위 '비식별 조치'를 했다고 한다. 코호트 자료를 제3자에게 제공할 때 비식별 조치를 하는 것인지, 아니면 보유하는 모든 코호트 DB가 비식별 조치가 이루어진 것인지는 밝히고 있지 않지만 코호트 대상자의 정보가 해마다 누적되는 것으로 볼 때는 내부의 코호트 DB는 비식별화를 하지 않은 상태이고, 외부 공개를 할 때 비식별화를 하는 것으로 보인다.

그런데 코호트 자료에서 사용한 소위 '비식별 조치'는 3단계 비식별 조치로 설명하는데, 1단계로 대체변수 사용(가명화), 2단계로 그룹화, 3단계로 민감상병 및 민감수가 코드 마스킹 기법이라고 한다.[262]

그림 4-10 표본코호트 DB 비식별화 방법

구체적으로는 주민등록번호를 일련번호로 바꾸고(가명화) 요양기관도 식별번호를 대체하고 민감상병의 경우 법정 감염병이라는 사실만 표시하고 구체적인 감염병의 내용을 마스킹한 것과 같다.

262) 박숙희(2015), "표본연구DB 개방, 성과 및 향후 비전", 건강보험 빅데이터 개방, 2차년도 성과 공유 심포지엄.

그러나 코호트 자료에 사용된 비식별조치는 매우 불완전하기 때문에 아주 손쉽게 재식별이 가능하다.

첫째, 일반적으로 가명화는 개별화, 연결 가능성, 추론 가능성 모두 아주 크게 남아 있기 때문에 익명화로 볼 수 없다. 따라서 요양기관을 표시하지 않고 식별번호로 대체해도 해당 요양기관을 추론하는 것은 아주 손쉬운 일이다.

둘째, 그룹화와 마스킹 기법도 개별화(Single out), 연결 가능성(Linkability), 추론 가능성(Inference)으로 인한 재식별화의 위험성을 안고 있는데 개인의 의료기록을 모두 담고 있는 경우라면 특정인의 의료정보를 일부라도 보유하고 있다면 특정인을 추론할 수 있다.

그림 4-11 비식별 조치별 개별화, 연결 가능성, 추론 가능성

	Is Singling out still a risk?	Is Linkability still a risk?	Is Inference still a risk?
Pseudonymisation	Yes	Yes	Yes
Noise addition	Yes	May not	May not
Substitution	Yes	Yes	May not
Aggregation or K-anonymity	No	Yes	Yes
L-diversity	No	Yes	May not
Differential privacy	May not	May not	May not
Hashing/Tokenization	Yes	Yes	May not

* 출처: 유럽연합 29조 작업반, Opinion 05/2014 on Anonymisation Techniques, p24.

따라서 이를 개인정보보호법의 적용이 배제되는 '개인을 알아볼 수 없게 조치한 것'으로 보기는 어렵다.

⑥ 개인정보 수집·연계에 대한 동의 여부와 보유의 적법성

코호트 자료 DB를 개인을 식별할 수 있는 개인정보로 본다면 민감정보이기 때문에 당사자의 동의가 없이 이를 작성하는 행위, 공개하는 행위는 법적 근거를 찾기 어렵다. 민감정보의 처리는 정보주체에게 각 고지사항을 알리고 다른 개인정보의 처리에 대한 동의와 별도로 동의를 받은 경우나, 법령에서 민감정보의 처리를 요구하거나 허용하는 경우에만 허용되기 때문이다(개인정보보호법 제23조 제1항).

그런데 국민건강정보 데이터베이스로 다양한 목적으로 수집된 개인의 건강정보와 자격정보, 징수정보 등의 개인정보가 연계·결합되고 이용되고, 특히 코호트 DB로 가공되어 연구용으로 제공되는 것에 대하여 해당 개인에게 동의를 받은 사실은 없다.

따라서 법적 근거가 있어야 하는데 앞서 본 것처럼 국민건강보험법 제14조 제1항 제9호의 '건강보험에 관한 조사연구'를 근거로 하기에는 난점이 있다.

⑦ 개인정보 보호 원칙의 준수 여부

이 규정을 법적 근거로 본다고 하더라도 표본 코호트 DB는 다음과 같은 점에서 문제가 있다.

첫째, 연구 목적으로 코호트 DB를 생성할 필요가 있다고 하더라도 당사자의 동의 없이 일방적으로 코호트 DB를 작성하는 것이 정당화될 수 있는지 검토가 필요하다. 동의를 거치는 것이 불가능하거나, 동의를 받으면 연구 결과에 영향을 미친다고 보기 어렵기 때문이다.

둘째, 코호트 DB의 목적에 비추어 연계된 정보들이 적합성과 최소수집의 원칙을 준수했다고 보기 어렵다. 연구 목적을 위하여 자격정보나 징수정보 등의 정보까지 연계·결합하는 것이 필요한지 의문이다.

셋째, 개인정보주체에게 정보를 제공할 의무를 준수했다고 보기도 어렵다. 정보주체들에게 개인정보의 수집·보유, 연계의 범위, 연계의 목적, 보유 기간 등에 대한 정보 제공이 이루어지지 않고 있다. 따라서 열람, 삭제, 정정 등의 권리를 행사할 수 있는 가능성도 차단되고 있다.

넷째, 개인정보주체의 처리중지나 삭제에 대한 권리가 보장되어 있지 않다. 연구 목적의 데이터베이스를 구축하는 것이라면 개인정보주체의 선택에 따라서 처리중지나 삭제를 요구할 수 있도록 권리를 보장해 주는 것이 최소한의 권리로 보장되어야 한다.

다섯째, 코호트 DB의 익명화 조치가 부실하다. 가명화와 그룹핑, 민감 상병의 마스킹 수준으로는 재식별 가능성은 거의 완전하다고 볼 수 있다.

라. 건강보험심사평가원의 DW시스템, 빅데이터 분석 DB 등

건강보험심사평가원은 아래와 같이 DW시스템(명세서)(준영구), 빅데이터 분석 DB(준영구) 등의 정보를 준영구로 보유하고 있다고 한다. 이 정보들은 데이터 연계로 축적되고 있는 것으로 보인다.

표 4-28 건강보험심사평가원이 보유하고 있는 개인정보파일 중 일부

DW 시스템 13,993,418,243명	이름:필수,건강:필수(수신자식별대체키,요양기호,보험자코드,청구명일련번호,청구형태코드,서식번호,수진자성명,증번호,가입자성명,내원일수,청구정액정률구분,공상구분,의료급여종별코드,청구대불금,청구장애인기금,청구요양급여비용총액,청구본인부담금,청구보험자부담금,진료결과구분,내원일자,요양일수,입원경로구분,처방전발행기관,당월요양개시일자,진료과목코드,전접수번호,청구전명일련번호,청구구분,전지급불능사유코드,최초입원일자,재료및항독소구분,재료 및 항독소금액,물리치료실시횟수,야간조산여부,다태아구분,총처방일수,투약일수,의료급여정액제구분,청구적용수가구분,청구본인부담상한액초과금,진찰횟수,청구진료비총액,청구보훈부담액,청구100분의100본인부담액,청구비급여액,본인부담적용코드,진료확인번호갯수,입·퇴원 방법,내원시 증상 및 발생시각,수술·치료(약제투여) 전후 환자상태,약제 투여시각,검사 종류 및 내용,수술·치료 형태 및 시각(54))
빅데이터분석 DB 53,092,815 명	건강:필수, 기타 (H-PIN, 성별, 진료비청구심사정보)

그런데 이 정보들을 준영구의 기간으로 하여 보유하고 있는 것은 개인정보보호의 원칙에 부합하는 것으로 보기 어렵다. 그리고 건강보험심사평가원의 업무에 비춰 보거나, 법령의 명시적인 규정이 있는 경우에나 민감정보를 수집, 처리할 수 있다는 점에서 적법한 것으로 보기도 어렵다.

특히 심사 청구를 위해 제출된 자료를 심사 청구를 마친 후에도 영구적으로 보유하고 있다는 것은 문제가 있다. 국민건강보험법은 요양기관에도 요양급여가 끝난 날부터 요양급여비용의 청구에 관한 서류를 5년간 보존하도록 하고 있고, 약국은 3년간 보존하도록 하고 있다. 국민건강보험법의 5년, 3년도 적정한 것인지 의문이나, 심사청구자료를 준영구로 보유하는 것은 매우 심각한 문제이다.

DW 시스템이라는 명목으로 139억 명의 민감한 건강정보를 준영구로 보유하고, 5,300만 명의 빅데이터 분석 DB를 보유하는 것도 법적 근거를 찾기 어렵고, 개인정보보호원칙에 어긋나는 것으로 문제이다.

표 4-29 건강보험심사평가원이 보유하고 있는 처방전정보

처방전 정보 8,514,112,431명	이름:필수, 기타 (이름, 기타 (성명, 대체키, 요양기관기호, 의·약사면허번호, 전화번호, 처방일자, 조제일자, 보험자구분, 처방조제 유형구분, 진료과목 코드, 질병분류기호, 임부여부, 특정기호, 팩스번호, 처방전 사용기간, 주사제 처방내역, 용법(전, 간, 후), 조제 시 참고사항, 처방의 변경 수정 확인 대체 시 내용, 약품코드, 약품명, 성분코드, 성분명, 1회 투여량, 1일 투여횟수, 총 투여일수, 급여구분, 원내·원외 구분(29)))

마. 건강보험심사평가원의 환자데이터셋

① 구축 및 보유 현황

건강보험심사평가원은 2009년부터 매년 1년간 진료받은 환자를 대상으로 표본 추출을 하여 환자데이터셋이라는 이름으로 공개를 하고 있다. 입원환자 데이터셋은 입원환자의 13%에 달하는 100만 명 정도, 전체 환자 데이터셋은 전체 환자의 3%인 140만 명, 고령환자 데이터셋은 전체 고령환자(60세 이상)의 20%인 100만 명 정도, 소아청소년환자데이터셋은 소아청소년환자(20세 미만)의 10%인 약 100만 명 정도의 환자데이터셋을 매년 추출하여 작성하는 것이다. 매년 합계 약 450만 명의 환자데이터셋이 구축되어 공개되는 것이다.

표 4-30 환자데이터셋 현황

데이터셋	구축 기간	추출된 환자 수
입원환자데이터셋	2009년 ~ 2016년	입원환자 추출비율 13% (약 100만 명)
전체환자데이터셋	2009년 ~ 2016년	전체환자 추출비율 3% (약 140만 명)
고령환자데이터셋	2009년 ~ 2016년	고령환자(65세 이상) 추출비율 20% (약 100만 명)
소아청소년환자데이터셋	2009년 ~ 2016년	소아청소년환자(20세 미만) 추출비율 10% (약 110만 명)

② 수록 정보

환자데이터셋은 해당 환자의 1년간의 청구명세서 등의 내역을 연계하여 작성한 것이다. 따라서 여기에는 1년 동안의 국민건강보험이나 의료급여, 장기요양보험을 통해서 제공받은 요양급여 내역이 모두 포함되어 있는 것이다.

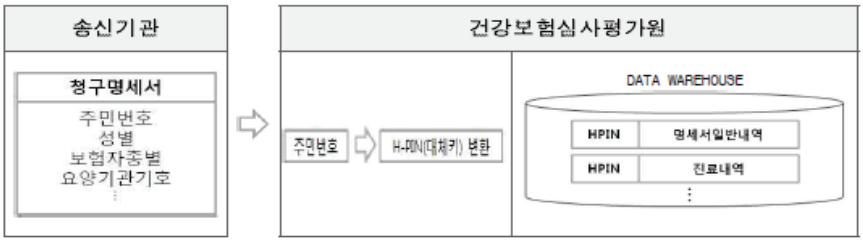
③ 비식별 처리

환자데이터셋에 대해서 주민등록번호를 대체키로 변환하고, 가명처리, 데이터삭제, 마스킹 등의 조치를 취했다고 한다. 그런데 이와 같은 비식별조치는 환자의 정보를 일부만 가지고 있거나, 언론, 소셜미디어 등의 정보와 결합하여 손쉽게 재식별이 가능하다.

그림 4-12 환자데이터셋 비식별 조치

○ 원천DB에서 환자이름 삭제, 주민번호를 H-PIN(대체키)으로 변환

< 원천DB 1차 비식별화 >



< 2차 비식별화 조치 방법 >

처리기법	조치 내용
가명처리	· 1차 비식별화 대체키에 임의의 난수 부여 후 1차 비식별자 삭제 · 요양기관 기호 및 명세서 번호 임의의 난수 부여
데이터 삭제	· 요양기관 주소는 시·도까지 제공 · 법정 전염병 및 100세 이상 환자 삭제 · 제약업체 영업정보 보호를 위하여 의약품코드 삭제 · 의약품 주성분 중 제형정보 일부 삭제
데이터 마스킹	· 치료재료업체의 영업정보 보호를 위하여 업체를 식별할 수 있는 세분류코드 마스킹

④ 법적 근거

공공데이터로 개방하여 공개하기 위해 구축된 환자데이터셋은 민감한 개인정보이므로 이를 연계하기 위해서는 해당 환자로부터 그와 같은 개인정보의 연계에 대해서 명시적인 동의를 받거나, 법적 근거가 필요하다. 그런데 환자로부터 동의를 받고 추진하지는 않았으며, 환자데이터셋으로 구축하기 위한 데이터 연계의 법적 근거는 찾아볼 수 없다. 환자의 건강정보를 구축하여 공개하는 것에 대한 법적 근거가 없기 때문이다.

아울러 건강보험심사평가원에서 환자데이터셋을 환자의 동의 없이 공공데이터로 제공하는 것과 관련하여서도 법적 근거가 있어야 하는데, 법적 근거를 찾을 수 없다.

건강보험심사평가원은 '공공데이터의 제공 및 이용 활성화에 관한 법률'에 근거하여 비식별조치를 했으므로 제공이 가능하다고 보았다[263]. 그러나 공공데이터법이 제공의 근거가 될 수 없으며, 비식별조치라는 것은 개인을 식별할 수 있는 정보로서 개인정

263) "건강보험심사평가원, 지난 3년간 민간보험사에 6천만명분 진료데이터 넘겨", 국회의원 정춘숙 보도자료(2017. 10. 24).

보보보협 제23조의 규율 대상이 된다. 따라서 그에 대한 명확한 법적 규율이 없는 한 민감한 개인정보인 건강정보의 제공은 허용되지 않는 것이다.

(4) 보건의료 빅데이터 개방시스템을 통한 데이터의 연계

가. 국민건강보험자료 공유서비스

① 개요

국민건강보험공단은 자신이 수집·보유·관리하는 건강보험 및 장기요양보험 관련 자료를 인터넷을 활용하여 학술연구 및 정책연구를 수행할 수 있도록 제공하고 있다.

공유서비스는 표본연구 DB와 맞춤형 연구 DB로 구성된다고 한다. 국민건강정보자료 제공 운영규정에 의하면 “표본연구DB”란 공단이 그 업무와 관련하여 보유하고 있는 정보에서 표본을 추출하여 정보주체를 알아볼 수 없도록 조치한 후 주제별로 규격화한 자료를 말하고, “맞춤형연구DB”란 공단이 그 업무와 관련하여 보유하고 있는 정보를 신청자의 연구 목적에 따라 추출·요약·가공하여 정보주체를 알아볼 수 없도록 조치한 후 구축한 자료를 말한다고 정의하고 있다.

② 법적 근거

그러나 현행 법령상 건보공단이 제공받은 자료를 계속 보유하면서 개인정보를 연구에 활용할 수 있는 법적 근거는 희박해 보인다.

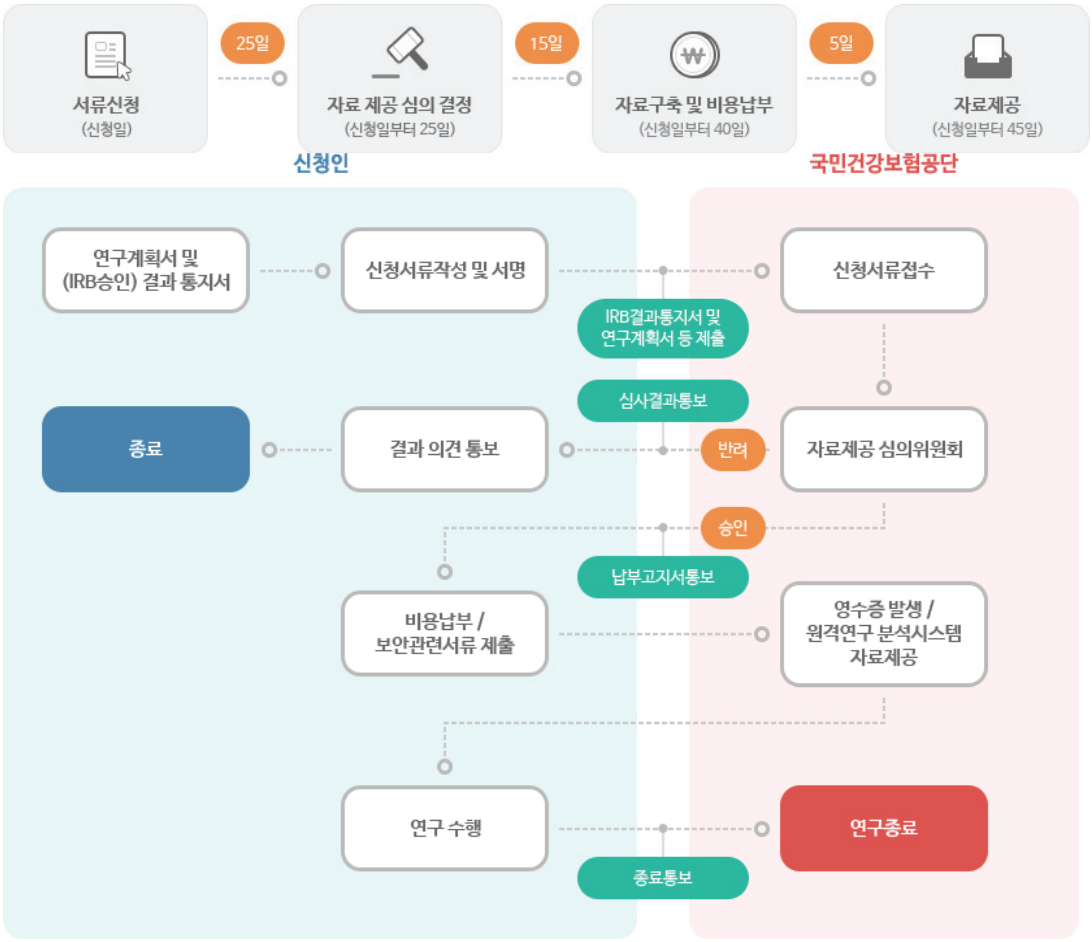
③ 운영현황

- 자료 신청과 제공 결정 절차

자료의 신청과 신청에 대한 심의와 결정은 아래와 같이 이루어진다.[264]

264) https://nhiss.nhis.or.kr/bd/ab/bdaba001cv.do

그림 4-13 국민건강보험공단 공유서비스 신청절차



- 신청할 수 있는 자와 신청의 목적

운영규정은 국민건강정보자료를 이용할 수 있는 자를 중앙행정기관, 지방자치단체, 공공기관, 공공연구기관, 이런 기관 등과의 계약에 따라 해당 기관 등이 주관하는 연구를 수행하는 자, 공단과 체결한 계약에 따라 연구를 수행하는 기관 또는 사람, 이상의 기관 등에 소속된 자로서 학술지에 논문을 투고하기 위하여 연구를 수행하는 자, 학위논문을 위하여 연구를 수행하는 자, 그 밖에 공익적 목적의 연구를 수행하는 자(이 경우에는 표본연구DB에 한정한다)로 제한하고 있다.

- 위원회

국민건강정보자료의 제공 및 이용 등에 관한 승인, 제공방법 등의 사항을 심의·의결하기 위하여 공단에 국민건강정보자료 제공 심의위원회를 둔다.

심의위원회는 위원장을 포함하여 11명 이내의 위원으로 구성한다. 위원장은 공단 각 부서(본부 각 실, 건강보험정책연구원 및 국민건강보험공단인재개발원을 말한다) 중 '직제규정'에 따라 국민건강정보자료에 관한 제공 업무를 주관하는 부서의 장으로

하고, 위원회의 위원은 공단 각 부서의 장이 추천하는 사람 중에서 위원장이 지명하는 5명 이내의 사람, 관련 학계, 유관기관, 시민단체 등에서 추천하는 5명 이내의 사람 중에서 이사장이 임명하거나 위촉하는 사람으로 한다. 결국, 위원회는 과반수가 내부 직원으로 구성되어 있다.

④ 맞춤형 자료

맞춤형 자료는 사전 협의를 통해서 제공 가능한 형태의 자료를 생성하여 제공하는 것이다. 제공대상은 국가기관 및 지방자치단체, 공공기관, 정책연구나 학술연구를 수행하는 기관 또는 사람, 공단과 체결한 협약(MOU, Memorandum of understanding) 등에 따라 연구를 수행하는 기관 또는 사람, 기타 연구를 수행하는 기관 또는 사람이다. 건강보험공단 내 빅데이터 분석센터에서만 자료 열람 및 분석이 가능하다.

⑤ 자료의 연계

국민건강보험공단에서 공개하고 있는 데이터의 경우는 식별자를 포함하고 있기 때문에 자료 연계가 가능하다. 자료의 연계는 국민건강보험공단 내부 자료끼리의 연계도 가능하지만, 그 외의 자료와도 연계가 가능하다.

'건강보험 빅데이터와 통계청 사망원인 자료 연계' 자료를 바탕으로 현재 맞춤형 자료의 연계 절차와 요건을 보면 다음과 같다.

그림 4-14 국민건강보험공단 맞춤형 자료 연계 흐름도

* 주황색 1. 4. 5 번은 신청자 본인이 직접 수행

- 자료 신청: 건강보험 자료에 사망원인을 연계하여 연구분석을 하고자 할 경우 국민건강보험공단에 맞춤형 연구 DB를 신청하면서 연구계획서와 연구요약서에 통계청 사망원인 연계 내용을 포함한다.

- 자료 연계를 위한 데이터셋 완성 :이용자는 승인받은 연구과제의 맞춤형 연구 DB를 이용하여 1차 분석을 마친 후 사망원인 자료 연계를 위한 최종 데이터셋을 완

성한다. 최종 데이터셋은 건보DB 중 사망이 확인된 데이터에 한하여 구성하며, 정확한 데이터 확인을 위해 사망 관련 변수(예. 사망 일자)를 포함하여 생성한다. 이때 '개인식별변수'와 '자료 간 연결변수'가 포함된다.

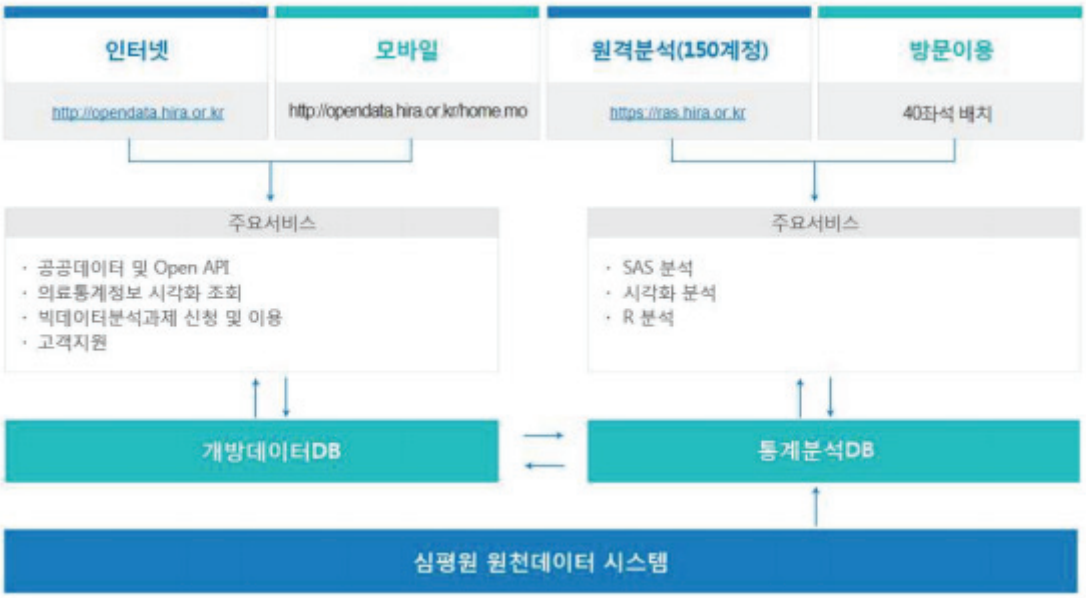
- 연계 신청 : 연구 데이터 간 자료연결이 필요한 변수의 경우 통계청에서는 새로운 대체키를 부여하여 연계작업을 진행한다. 이용자는 데이터셋 완성 단계에서 통계청 MDIS 홈페이지(mdis.kostat.go.kr)를 방문하여 직접 사망원인 연계 신청을 하고, 테이블 정의서 파일 등을 작성하여 공단 반출 요청 폴더에 저장한 후 공단 담당자에게 신청한다. 공단은 이용자가 작성한 최종 데이터셋과 테이블정의서를 통계청 MDIS에 제출한다.

- 연계 진행 : 통계청은 MDIS 신청자료(IRB 승인서, 연구계획서, 식별정보 위탁처리 신청서 첨부)와 건강보험공단 담당자의 연계요청자료(최종 데이터셋, 테이블정의서)를 검토 후 사망원인자료와 연계를 진행한다.

- 통계청은 공단 신청자료에 대해 '반출승인' 처리하고, ⑨ 공단은 '반출승인'된 데이터를 제공받아 연구자의 분석용 컴퓨터에 자료를 제공한다.

나. 건강보험심사평가원의 보건의료빅데이터개방시스템 운영[265]

그림 4-15 보건의료빅데이터개방시스템 구성



265) 건강보험심사평가원(2017), "보건의료빅데이터센터_이용가이드".

① 개요

심사평가원은 심사평가 등의 업무와 관련하여 수집·보유하고 있는 국민의 건강정보를 '공공데이터의 제공 및 이용 활성화에 관한 법률'의 '공공데이터'로 보고, 보건의료빅데이터개방시스템을 통하여 개방하고 있다. 심사평가원은 보건의료 빅데이터인 의료정보(의료기관 청구자료 기반의 자료로 전 국민의 진료정보, 의약품 정보, 치료재료정보, 의료자원정보 등을 분석·정제한 데이터를 말한다고 정의한다)를 편리하고 손쉽게 활용할 수 있도록 국민에게 개방하고 있으며, 이들 공공데이터는 누구나 이용할 수 있고, 영리 목적의 이용을 포함한 자유로운 활용이 보장된다고 소개하고 있다. 건강보험심사평가원은 운영의 목적에서 의약계·학계·산업계 등을 대상으로 전 국민의 진료내역, 의료자원 현황, 의약품 유통정보 등 보건의료 빅데이터를 개방하여 민간의 공공데이터이용 활성화를 위함이라고 밝히고 있다.

② 운영 현황

- 보건의료빅데이터개방시스템의 구성

보건의료빅데이터개방시스템은 1) 개방데이터 DB를 이용한 인터넷 및 모바일을 통한 데이터 및 통계자료 조회 서비스와 2) 통계분석 DB를 이용한 원격접속시스템 및 보건의료빅데이터센터(이하 '빅데이터센터') 운영으로 구성되어 있다.

원격접속시스템은 이용자가 원하는 장소에서 원격접속(가상화PC)을 통한 분석지원서비스로 주로 유관기관 및 학계의 연구목적의 자료 분석 등을 위해 이용되고 있으며, 총 150 계정이 운용되고 있다(2016년 12월 기준).

빅데이터센터는 유관기관 및 학계의 연구목적의 자료 분석 및 의료계 및 산업계의 R&D 개발을 위한 공간으로, 본원 및 전국에 총 40석을 운영하고 있다. 산업계자료는 심평원에서 운영하고 있는 빅데이터센터에 방문하여 분석을 수행하고 있으며, 연구과제는 IRB 심의 통보서, 기관 간 자료요청 공문, 보안각서, 보안준수확약서 등의 정보보안을 준수하는 조건으로 원격분석 지원을 이용하고 있다. 원격통계분석시스템 접속 시에는 이용자 ID, 공인IP, 사설IP, MAC 주소, 공인인증서 등을 통해 승인된 이용자만 접속할 수 있다. 제공데이터는 온라인, 전자매체 등의 방법으로 제공되며, 개인정보보호를 위한 대체키 변환 등을 통해 제공된다고 한다.

- 자료제공 체계

자료제공대상은 아래와 같이 연구자료와 산업체 자료로 나누고 있다. 의료기관 종

사자, 학술연구 수행기관, 제약업체 및 의료기기 업체 등 산업계는 물론이고, 컨설팅 업체 및 예비 창업자 등도 포함하고 있다. 영리 여부를 불문하고, 국민 누구에게나 개방하며, 실제로 민간보험회사에도 국민의 진료내역 등을 제공하였음이 밝혀지기도 했다.

표 4-31 보건의료빅데이터 개방시스템 자료제공 대상

연구자료	・국가, 지방자치단체 및 정부 산하기관 ・연구 중심 병원 및 학술 연구 수행기관 등 연구목적으로 신청하는 자
산업체자료	・의약품 품목 허가를 받은 자 ・제약, 의료기기 산업 ・컨설팅 회사 및 예비 창업자 등

자료제공범위는 맞춤형 자료는 건강보험심사평가원에서 제공 가능한 형태의 자료 제공하고(연계·결합), 보건의료 자료제공 시에는 건강보험자료 범위 내에서 제공한다고 한다.

자료제공 기준은 다음과 같다고 한다. 비공개대상정보가 포함된 경우는 제공을 제한하지만, 기술적으로 분리할 수 있는 경우는 비공개대상정보를 제외한 공공데이터는 부분제공하고, '주민등록번호, 성명, 주소, 전화번호, 요양기관기호(명칭) 등 개인정보와 개별 법인·단체 등의 정보가 식별 불가능한 형태로 제공'한다는 것이다.

표 4-32 보건의료빅데이터 개방시스템 자료제공 기준

- 공공데이터에 개인정보 등 비공개대상정보가 포함된 경우 (공공기관의 정보공개에 관한 법률) 제공 제한
- 주민등록번호, 성명, 주소, 전화번호, 요양기관기호(명칭) 등 개인정보와 개별 법인・단체 등의 정보가 식별 불가능한 형태로 제공
- 개인정보 등 비공개대상정보를 기술적으로 분리할 수 있는 경우 개인정보 등 비공개대상정보를 제외한 공공데이터는 부분제공
- 공공데이터의 정보가 다른 정보와 결합하여 개인 식별이 가능한 경우 제공 제한

- 제공자료 내역

심사평가원에서 제공하는 자료는 진료정보, 의약품 정보, 치료재료정보, 의료자원정보로 크게 구분할 수 있으며, 진료정보에는 전 국민의 의료이용이 모두 포함되어 있는 건강보험 청구자료[266]가 포함되어 있다.

266) 청구자료의 장점으로는 1) 요양기관을 방문한 환자의 비급여를 제외한 모든 의료서비스 이용 정보가 포함되어 분석결과의 일반화가 가능하고, 2) 의료서비스 내역(시술/수술, 검사, 약제처방, 치료재료 등), 서비스별 급여비용 등이 포함되어, 환자가 받은 개별적인 의료서비스를 확인 가능하며, 3) 환자 고유식별자가 자료에 포함되어 환자의 추적 자료 구축이 가능하다는 점을 든다.

표 4-33 보건의료빅데이터 개방시스템 제공자료 내역

구분	주요정보	DB명칭
진료정보	청구명세서 정보 의료행위 정의 및 환자 분류 정보 수가마스터 정보 질병군(DRG)·요양병원 수가 마스터정보 의료행위별 심사기준 정보(보장 범위) 의료행위 18개 분류별 진료규모 정보(진료량, 금액) 질병군(DRG) 및 요양병원 진료규모 정보 질병정보(주상병) 및 질병단위 진료규모 정보	청구명세서 DB 정보분석 DB 통계자료 DB 질병통계 DB 행위통계 DB
의약품 정보	급여의약품 마스터 정보 급여의약품 사용 정보	의약품DB
치료재료 정보	치료재료 마스터 정보 치료재료별 사용정보 특수 재료관련 정보(복강경시술, 조영제 등)	치료재료 DB
의료자원 정보	요양기관 개폐업 정보 의료기관 시설 정보(병상, 집중치료실, 수술실 등) 인력(의사, 간호사, 의료기사 등) 현황 장비 보유현황(장비별 이력관리)	요양기관현황 DB 요양기관 종합정보 DB

구체적으로 살펴보면, 유관기관 및 학계의 연구목적의 자료 분석 등을 위해 이용되고 있는 원격시스템에서 제공하고 있는 테이블과 변수의 경우는 매우 상세하며, 명세서 조인키와 같은 정보는 고유키로서 중복이 없고, 명세서일반내역테이블, 진료내역테이블, 수진자상병내역테이블, 원외처방전상세내역테이블, 요양기관테이블 등 모든 정보를 망라하고 있다.

한편 의료계와 산업계를 위한 제공자료의 경우는 아래와 같은 자료를 제공하고 있다고 한다.

그림 4-16 보건의료빅데이터 개방시스템 제공자료 (의료계와 산업계)

한글변수명	영문변수명	변수유형	변수 크기	변수설명
서식코드	FOM_CD	CHAR	2	021: 의과입원, 031: 의과외래, 041: 치과입원, 051: 치과외래, 061: 조산원입원, 071: 보건기관입원의과, 072: 보건기관입원치과, 073: 보건기관입원한방, 081: 보건기관외래의과, 082: 보건기관외래치과, 083: 보건기관외래한방, 091: 정신과낮병동, 101: 정신과입원, 111: 정신과외래, 121: 한방입원, 131: 한방외래, 151: 의료급여혈액투석정액, 201: 약국직접조제, 211: 약국처방조제, 991: 조산원외래
통합분류코드 (제품명)	UNI_DIV_CD	CHAR	17	주성분코드로 요청 시 일부분만 제공
약효분류코드	MEFT_DIV_NO	CHAR	3	약효가 유사한 동일 효능(약효)군
ATC코드	WHO_ATC_CD	CHAR	7	WHO에서 개발한 국제의약품 분류 코드
주성분코드	GNL_NM_CD	CHAR	9	주성분코드(일반명코드)
진료과목코드	DGSBJT_CD	CHAR	2	진료를 받은 진료과목(병원급 이상) 또는 상병명에 해당하는 진료과목(의원급)
표시과목코드	SHW_SBJT_CD	CHAR	2	요양기관이 외부에 표시하기 위해 정의한 표시과목
내과세부과목코드	IFLD_DTL_SPC_SBJT_CD	CHAR	2	내과 진료과목 중 '세부전문의 제도인증 규정(대한의학회)'에 따라 인증받은 세부전문과목을 운영하고 있는 종합병원, 상급종합병원의 경우 진료를 받은 세부전문과목
내원일수	VST_DDCNT	INTEGER	4	입원 또는 내원하여 진료를 받은 실 일수
1회투약량	FQ1_MDCT_QTY	NUMERIC	18	1회투약량
1일투여횟수	DY1_INJC_QTY_EXEC_FQ	NUMERIC	18	1일투여횟수
총투여일수	TOT_INJC_DDCNT_EXEC_FQ	INTEGER	4	총 투여일수 또는 실시횟수
총사용량	TOT_USE_QTY_OR_EXEC_FQ	NUMERIC	18	1회투약량 * 1일투여횟수 * 총투약일수
금액	AMT	BIGINT	8	단가 * 1회 투약량 * 1일 투여량 (투여(실시)횟수) * 총투여일수(실시횟수)
단가	UNPRC	BIGINT	8	「약제 및 치료재료의 비용에 대한 결정기준」에 따른 단가

한글변수명	영문변수명	변수유형	변수크기	변수설명
명세서일련번호	MID	CHAR	20	명세서당 부여된 일련번호(임의의 연번 부여)
요양개시년월	RECU_FR_YM	CHAR	6	외래인 경우 내원한 진료년월 입원인 경우 해당 명세서의 입원진료년월
심사년월	RV_YM	CHAR	6	건강보험심사평가원에서 심사가 발생한 년월
처방진료구분코드	PRSC_DIAG_TP_CD	CHAR	1	처방(C), 진료(J) 구분코드
DW주상병코드	DW_MSICK_CD	CHAR	7	양방(A) + 상병코드 or 한방(B) + 상병코드를 나타내는 주상병코드
DW부상병코드	DW_SSICK_CD	CHAR	7	양방(A) + 상병코드 or 한방(B) + 상병코드를 나타내는 부상병코드
수진자일련번호	JID	CHAR	20	수진자당 부여된 일련번호(임의의 연번 부여)
수진자연령	PAT_AGE	NUMERIC	4.1	5세 단위로 범주화하여 제공
성별	SEX_TP_CD	CHAR	1	수진자 수급자의 성별구분 1: 남, 2: 여, 9: 기타
보험자종별	INSUP_TP	CHAR	1	4: 건강보험, 5: 의료급여, 7: 보훈, 9: 무료진료
요양기관일련번호	YID	CHAR	20	요양기관당 부여된 일련번호 (월별 별도의 임의의 연번 부여)
종별코드	L_CD	CHAR	2	01: 상급종합병원, 11: 종합병원, 21: 병원, 28: 요양병원, 29: 정신요양병원, 31: 의원, 41: 치과병원, 51: 치과의원, 61: 조산원, 71: 보건소, 72: 보건지소, 73: 보건진료소, 74: 모자보건센타, 75: 보건의료원, 81: 약국, 92: 한방병원, 93: 한의원
지역코드	PLC_CD	CHAR	6	요양기관 소재한 지역(시, 군, 구)의 코드
총병상수	TOT_SBD_CNT	INTEGER	4	요양기관의 총 병상수로 범주화하여 제공

- 자료의 신청과 이용절차

아래와 같이 자료이용 문의 → 자료요청 → 제공 여부 결정 → 자료생산 → 통보 → 좌석 배정 → 분석의 절차로 소개하고 있다.[267]

267) 심사평가원에 이메일을 통해 상담신청을 하면, 심사평가원에서 상담신청내용을 검토하며, 검토가 끝나면 신청자가 홈페이지에서 이용신청을 한다. 상담신청부터 홈페이지에 이용신청을 하기까지 대략 10일 소요되며, 홈페이지에 신청을 완료하면 심사평가원에서는 자료요청을 접수하고 신청한 자료를 추출하여 원격시스템을 통해 제공한다. 자료 신청부터 데이터 분석 자료를 원격시스템을 통해 제공하기까지 보통 2~4주 정도 소요된다고 한다.

그림 4-17 보건의료빅데이터센터 이용절차

그림 4-18 공공데이터 제공 신청서

공공데이터 제공 신청서

접수번호		접수일	처리기간	일
신청인	성명 (단체명 및 대표자 성명)		생년월일	
	주소(소재지)		사업자(법인·단체) 등록번호	
	전화번호		전자우편주소	
신청내용	공공데이터 명칭			
	공공데이터 내용			

「공공데이터의 제공 및 이용 활성화에 관한 법률」 제27조제1항, 같은 법 시행령 제26조제1항 및 제3항에 따라 공공데이터의 제공을 신청합니다.

년 월 일

신청인 (서명 또는 인)

건강보험심사평가원장 귀하

접 수 증

접수번호		신청인 성명	
접수부서		접수자 성명	(서명 또는 인)

귀하의 신청서는 위와 같이 접수되었습니다.

년 월 일

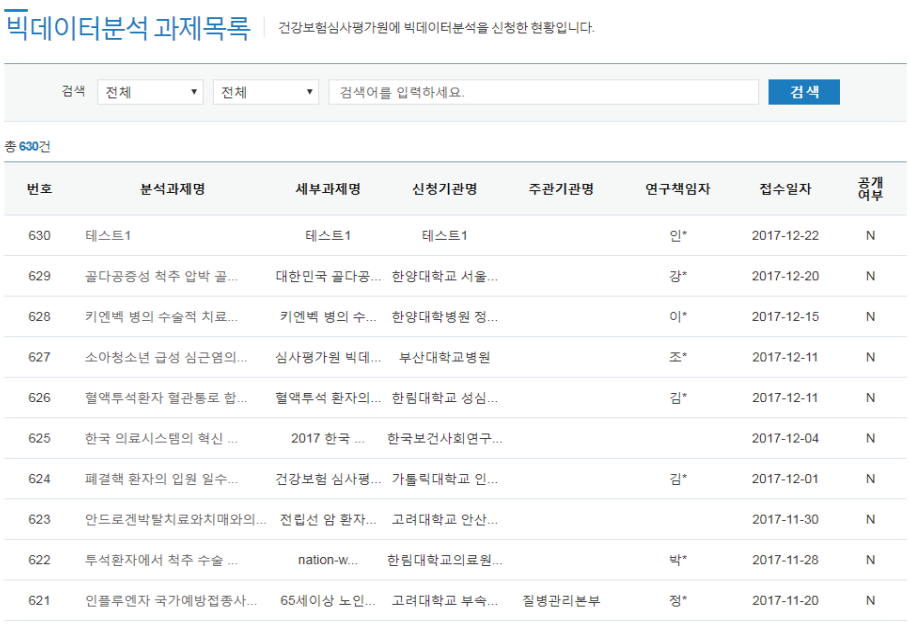
건강보험심사평가원장 직인

210mm×297mm[백상지 (80g/㎡) 또는 중질지(80g/㎡)]

- 자료신청서 : 건강보험심사평가원은 아래와 같은 공공데이터 제공신청서로 자료요청을 받고 있다. 특별히 연구를 위한 목적이 아닌 경우에도 제공하는 것으로 되어 있다. 한편 연구자료의 경우는 맞춤형 자료로 제공하는데, 이때는 연구과제수행개요서를 제출하도록 하고 있다. 이때 신청서를 제출하면서 요청하는 정보를 기재하도록 하고 있는데, 그 내용은 환자 개인을 충분히 식별할 수 있는 내용이다.[268] 그리고 산출조건도 특정한 상병이나, 특정한 약품의 수진자 전부에 대한 정보를 추출하도록 하는 등 충분히 개인이 식별될 수 있는 정보들이다.

- 연구과제의 공개 여부 : 현재 보건의료빅데이터개방시스템에서는 과제목록을 공개하고 있는데, 전체 625건의 과제 중 공개하고 있는 것은 단 2개에 불과하고, 나머지는 모두 비공개로 신청하고 있다. 그러나 공개로 신청한 과제도 결과보고서는 비공개로 하고 있다.[269]

그림 4-19 보건의료빅데이터 개방시스템 과제목록



빅데이터분석 과제목록 | 건강보험심사평가원에 빅데이터분석을 신청한 현황입니다.

검색 전체 전체 검색어를 입력하세요. 검색

총 630건

번호	분석과제명	세부과제명	신청기관명	주관기관명	연구책임자	접수일자	공개여부
630	테스트1	테스트1	테스트1		인*	2017-12-22	N
629	골다공증성 척추 압박 골...	대한민국 골다공...	한양대학교 서울...		강*	2017-12-20	N
628	키엔벡 병의 수술적 치료...	키엔벡 병의 수...	한양대학병원 정...		이*	2017-12-15	N
627	소아청소년 급성 심근염의...	심사평가원 빅데...	부산대학교병원		조*	2017-12-11	N
626	혈액투석환자 혈관통로 합...	혈액투석 환자의...	한림대학교 성심...		김*	2017-12-11	N
625	한국 의료시스템의 혁신 ...	2017 한국 ...	한국보건사회연구...			2017-12-04	N
624	폐결핵 환자의 입원 일수...	건강보험 심사평...	가톨릭대학교 인...		김*	2017-12-01	N
623	안드로겐박탈치료와치매와의...	전립선 암 환자...	고려대학교 안산...			2017-11-30	N
622	투석환자에서 척추 수술 ...	nation-w...	한림대학교의료원...		박*	2017-11-28	N
621	인플루엔자 국가예방접종사...	65세이상 노인...	고려대학교 부속...	질병관리본부	정*	2017-11-20	N

---

268) 포함 여부를 선택할 수 있는 것들은 명세서 일반내역, 진료내역, 수진자 상병내역, 원외처방전 상세내약 등인데, 예를 들어 명세서 일반내역의 경우만 보더라도 다음과 같이 특정 개인을 충분히 식별할 수 있는 내용이다.

수진자개인식별대체키, 성별구분, 수진자연령, 수진자통계연령, 요양기관식별대체키, 요양기관종별구분코드, 지역(시도)코드, 서식구분코드, 주상병코드, 부상병코드, 진료과목코드, 요양개시일자, 요양종료일자, 최초입원일자, 입내원일수, 요양일수, 원외처방일수, 원외처방약제비, 원외처방전건수, 심사결정요양급여비용총액, 심사결정본인부담금, 심사결정보험자부담금, 심사결정100분의100미만 총액, 수술여부, 공상구분코드, 특정기호구분코드, 상해외인구분코드, 진료결과구분코드, 입원도착경로구분코드, 의료급여종별코드, 청구형태코드, 청구구분코드, 심사년월, 표시과목코드.

269) 공개하고 있는 과제는 '영아기 타이레놀 복용이 천식 발생 및 호흡기질환에 미치는 영향'(순천향대학교, 주관기관 한국보건의료연구원), '2015 국내 혈액수급감시 통계 연보'(질병관리본부, 김준년) 단 두 건인데, 두 건도 모두 결과보고서는 비공개를 선택하였다.

표 4-34 연구수행개요서 첨부자료

<table>
<tr><td>진료기간</td><td colspan="6"></td></tr>
<tr><td>보험자종별구분</td><td colspan="6">☐4: 건강보험 ☐5: 의료급여 ☐7: 국비보훈</td></tr>
<tr><td>요양기관종별구분</td><td colspan="6">
<table>
<tr><th>포함여부</th><th>종별코드</th><th>요양종별명칭</th><th>포함여부</th><th>종별코드</th><th>요양종별명칭</th></tr>
<tr><td>☐</td><td>01</td><td>상급종합병원</td><td>☐</td><td>71</td><td>보건소</td></tr>
<tr><td>☐</td><td>11</td><td>종합병원</td><td>☐</td><td>72</td><td>보건지소</td></tr>
<tr><td>☐</td><td>21</td><td>병원</td><td>☐</td><td>73</td><td>보건진료소</td></tr>
<tr><td>☐</td><td>28</td><td>요양병원</td><td>☐</td><td>74</td><td>모자보건센타</td></tr>
<tr><td>☐</td><td>31</td><td>의원</td><td>☐</td><td>75</td><td>보건의료원</td></tr>
<tr><td>☐</td><td>41</td><td>치과병원</td><td>☐</td><td>81</td><td>약국</td></tr>
<tr><td>☐</td><td>51</td><td>치과의원</td><td>☐</td><td>92</td><td>한방병원</td></tr>
<tr><td>☐</td><td>61</td><td>조산원</td><td>☐</td><td>93</td><td>한의원</td></tr>
</table>
</td></tr>
<tr><td>서식구분</td><td colspan="6">
<table>
<tr><th>포함여부</th><th>서식구분</th><th>서식구분명칭</th><th>포함여부</th><th>서식구분</th><th>서식구분명칭</th></tr>
<tr><td>☐</td><td>021</td><td>의과입원</td><td>☐</td><td>091</td><td>정신과낮병동</td></tr>
<tr><td>☐</td><td>031</td><td>의과외래</td><td>☐</td><td>101</td><td>정신과입원</td></tr>
<tr><td>☐</td><td>041</td><td>치과입원</td><td>☐</td><td>111</td><td>정신과외래</td></tr>
<tr><td>☐</td><td>051</td><td>치과외래</td><td>☐</td><td>121</td><td>한방입원</td></tr>
<tr><td>☐</td><td>071</td><td>보건기관 입원의과</td><td>☐</td><td>131</td><td>한방외래</td></tr>
<tr><td>☐</td><td>072</td><td>보건기관 입원치과</td><td>☐</td><td>201</td><td>직접조제(약국)</td></tr>
<tr><td>☐</td><td>073</td><td>보건기관 입원한방</td><td>☐</td><td>211</td><td>처방조제(약국)</td></tr>
<tr><td>☐</td><td>081</td><td>보건기관 외래의과</td><td>☐</td><td>061</td><td>조산원 입원</td></tr>
<tr><td>☐</td><td>082</td><td>보건기관 외래치과</td><td>☐</td><td>991</td><td>조산원 외래</td></tr>
<tr><td>☐</td><td>083</td><td>보건기관 외래한방</td><td></td><td></td><td></td></tr>
</table>
</td></tr>
<tr><td>요양기관 소재지역</td><td colspan="6"></td></tr>
<tr><td>수진자 성별</td><td colspan="6">☐전체 ☐1:남 ☐2:여</td></tr>
<tr><td>수진자 연령</td><td colspan="6"></td></tr>
<tr><td rowspan="3">상병코드 조건</td><td rowspan="2">상병코드</td><td>하위코드 포함</td><td colspan="4"></td></tr>
<tr><td>하위코드 미포함</td><td colspan="4"></td></tr>
<tr><td>상병순위</td><td colspan="4">예 (1:주상병, 2:부상병, 10:10순위내, ... 전체)</td></tr>
<tr><td>약품코드 조건
(성분명코드)</td><td colspan="6"></td></tr>
<tr><td>행위코드 조건
(처치,수술,검사)</td><td colspan="6"></td></tr>
<tr><td>산출 기준</td><td colspan="6">☐ 산출조건을 만족하는 수진자의 모든 진료내역 추출
☐ 산출조건을 만족하는 해당 명세서 추출</td></tr>
<tr><td rowspan="2">기타 조건</td><td colspan="3">☐ DRG청구 제외</td><td colspan="3">☐ 추가청구 제외</td></tr>
<tr><td colspan="6"></td></tr>
</table>

③ 자료의 연계

- 자료 연계의 법적 근거

연구자는 외부기관의 자료를 연계하기도 하는데, 외부기관의 자료원을 개인의 고유

식별정보를 이용하여 심사평가원의 청구자료와 연계하여 분석을 수행하고자 할 경우, 외부기관이 민감정보, 고유식별정보, 주민등록번호를 심사평가원에 제공하기 위해서는 제3자 제공에 대한 법적 근거가 있어야 한다. 개인정보보호법과 생명윤리 및 안전에 관한 법률에 따른 서면 동의를 받고, 기관위원회의 심의를 받아야만 민감정보, 고유식별정보, 주민등록번호 처리가 가능하고, 제3자에게 제공이 가능하다.

환자를 대상으로 수집한 자료를 청구자료와 연계하여 분석하고자 할 경우, 외부기관에서는 ① 환자들을 대상으로 민감정보, 고유식별정보, 주민등록번호 활용, 제3자에 제공에 대한 별도 서면 동의를 취득해야 한다. 그리고 ② 기관위원회의 심의를 거쳐 승인을 받아야 한다. 이 경우 외부기관은 환자 자료와 주민등록번호를 심사평가원에 반입하여 주민등록번호를 기준으로 자료 연계를 시도할 수 있다고 한다.

'생명윤리 및 안전에 관한 법률' 제16조는 다음 두 가지 요소를 모두 갖춘 경우 기관위원회의 승인을 받아 연구대상자의 서면 동의를 면제할 수 있다고 규정하고 있는데, 그 두 가지 요소는 1. 연구대상자의 동의를 받는 것이 연구 진행과정에서 현실적으로 불가능하거나 연구의 타당성에 심각한 영향을 미친다고 판단되는 경우 2. 연구대상자의 동의 거부를 추정할 만한 사유가 없고, 동의를 면제하여도 연구대상자에게 미치는 위험이 극히 낮은 경우이다. 이 경우 민감정보, 고유식별정보, 주민등록번호 활용, 제3자에 제공에 대한 별도 서면동의서를 받지 않아도 되는지가 문제 되는데, 이 경우도 개인정보보호법의 적용이 배제되는 것은 아니므로 개인정보보호법에 따른 동의는 받아야 할 것이다.[270]

한편, 심평원에서 건강정보를 제3자에게 개방하는 것은 해당 정보주체의 동의가 있는 경우라면 허용될 것이다. 서면 동의는 목적과 사용범위를 명확하게 밝히고, 분명한 동의를 받아야 할 것이다. 그런데 현재 심평원은 해당 정보주체로부터 서면 동의를 받는 절차를 거치지 않고, 빅데이터 개방시스템을 통해서 데이터베이스를 통해 의료정보를 공개하고 있다.

- 자료의 연계에 대한 특별한 규정

현재 심평원의 경우 자료의 연계에 대한 특별한 규정은 두고 있지 않다.

- 자료의 연계 사례

2014년 제2사분기부터 2016년 제3사분기까지 원격접속시스템을 이용하는 자료이용을 신청한 142건을 대상으로 자료제공 시스템 운영 현황을 분석했는데, 자료 신청 기관의 자료 활용 목적으로는 학술연구가 전체 142건 중 128건(90.1%)으로 가장 많았으

270) 에서는 서면동의서를 받지 않아도 된다고 보았으나, 개인정보보호법은 적용될 것이므로 생명윤리법의 서면동의서를 받을 의무만 적용 배제되는 것으로 볼 것이다.

며, 정책 결정이 8건(5.6%), 마케팅정보 1건(0.7%) 그리고 병원경영 5건(3.5%)으로 파악되었다고 한다. 이 사례에서 전체 142건 중 외부자료와 연계를 신청한 건은 총 6건(42. %)이며, 기상/환경자료의 일자와 연계를 신청한 건이 3건(2.1%), 청구자료대상자에 대한 사망 일자를 신청한 건이 2건(1.4%)이었다. 외부자료를 주민등록번호와 함께 심사평가원으로 반입하여 해당하는 대상자들의 청구자료를 추출한 건은 1건(0.7%)으로 파악되었다고 한다.[271]

그림 4-20 건강보험심사평가원 원격접속시스템 자료요청의 범위

구분		빈도	백분율(%)
명세서/전체	해당 코드를 가진 명세서만 요청	54	38.0
	해당 코드를 가지는 대상자의 전체자료	88	62.0
외부자료 연계		6	4.2
	기상/환경자료	3	2.1
	사망일자	2	1.4
	외부기관자료	1	0.7

4. 기타 건강정보와 데이터 연계

(1) 국립암센터 암 정보 DB

가. 개요

암은 사망원인 1위이며, 질병 부담 1위로서 우리 국민의 건강을 위협하는 대표적인 질환으로 손꼽히고 있다.[272] 암에 대한 정보를 수집, 분석하여 체계적인 관리로 조기치료를 통한 국민건강 증진, 암의 발병원인 분석 등을 위해서 많은 나라에서는 지역기반 암 정보 데이터베이스(Population based Cancer Registry)를 구축하고 있다.

법률적 의무를 부과하는 암 정보 등록제도에 대해서는 프라이버시 침해의 위험과 위헌성 논의가 있기도 하지만, 그 필요성을 인정하고 있다.[273] 암 정보 등록의 목적과 효과는 (i) 정책연구나 암에 대한 과학적 연구에 도움이 된다는 점, (ii) 치료의 가이드라인을 작성하는 데 도움이 되고, (iii) 환자권리를 옹호하는 데 도움이 되기 때문이라고 한다. 실제로 암 등록 정보를 통해서 역학적 연구를 하고, 암의 원인을 밝히고,

271) 안동대학교 산학협력단(2016), 앞의 자료, p13.

272) 2012년 현재 인구 10만 명당 149.0명이 사망함으로써 사망원인 중 1위를 차지하고 있으며, 질병부담 역시 10만 명당 1,525 DALYs로 주요 질병군 중 가장 높은 것으로 나타나고 있다. 이덕형, "암 빅데이터 플랫폼 구축 사업 기획 연구", 보건복지부, 2014. p.58.

273) Robert H McLaughlin, Christina A Clarke, LaVera M Crawley and Sally L Glaser (2010), "Are Cancer Registries Unconstitutional?".

암질환에 대한 계획과 의료기관들의 암 치료에 대한 효과성 평가, 사회와 지역의 암 발생 특징과 부담 확인, 암에 대한 다양한 연구 등을 할 수 있게 된다.[274]

우리나라도 암관리법을 제정(2003)하여 보건복지부장관으로 하여금 암등록통계사업을 비롯한 암 정보 사업을 시행하도록 하고 있다. 암등록통계사업은 암 발생 위험요인과 암의 발생 및 치료에 관한 자료를 지속적이고 체계적으로 수집·분석하여 암 발생률, 생존율 등의 통계를 산출하기 위한 등록·관리·조사사업이다.

나. 암 정보 등록의무와 비밀보호

암관리법은 보건복지부장관이 암 환자를 진단·치료하는 의료인 또는 의료기관, '국민건강보험법'에 따른 국민건강보험공단 및 건강보험심사평가원, 그 밖에 암에 관한 사업을 하는 법인·기관·단체에 대하여 암등록통계사업에 필요한 자료의 제출(암 환자의 진료와 관련된 자료 및 의무기록 등)이나 의견의 진술 등을 요구할 수 있다고 규정하고, 자료의 제출을 요구받은 자는 특별한 사유가 없으면 요구에 따라야 한다고 규정하고 있다.[275]

그리고 암관리법은 이때 처리되는 개인정보가 통계법을 준용하고, 개인정보보호법 제58조 제1항에 따라 개인정보보호법이 적용되지 않는다고 규정하고 있다. (암관리법 제14조). 이 규정은 개인정보보호법 제3장 ~ 제7장까지만 적용이 배제된다는 것으로 해석해야 할 것이다.

> 제14조(암등록통계사업) ① 보건복지부장관은 암 발생 위험 요인과 암의 발생 및 치료에 관한 자료를 지속적이고 체계적으로 수집·분석하여 암 발생률, 생존율 등의 통계를 산출하기 위한 등록·관리·조사사업(이하 "암등록통계사업"이라 한다)을 시행하여야 한다. 이 경우 통계자료의 수집 및 통계의 작성 등에 관하여는 「통계법」을 준용하며, 통계의 산출을 위하여 처리되는 개인정보는 「개인정보보호법」 제58조제1항에 따라 같은 법이 적용되지 아니하는 개인정보로 본다. <개정 2011.3.29.>
> ② 보건복지부장관은 암환자를 진단·치료하는 의료인 또는 의료기관, 「국민건강보험법」에 따른 국

274) 배종면(2002), "국가 암등록사업 현황 및 향후 발전 방향", 국립암센터, 제4회 심포지움(2002년 2월).

275) 실제로 우리나라의 암등록사업은 다음과 같이 이루어진다고 한다(2014년 국가암등록통계 연례보고서, p2).
중앙암등록본부에서는 [국가암발생데이터베이스(DB)]를 구축하기 위하여 다음과 같이 자료원을 이용한다. 가장 주축이 되는 자료원은 1988년부터 2014년까지 우리나라의 암등록병원에서 신규로 진단 보고된 암환자의 DB를 통합한 중앙암등록모(母)DB이다. 이를 기본으로, 각 지역암등록사업의 등록DB, [암발생통계조사] DB, 전문 학회 및 연구회의 암종별 등록DB를 모두 통합한 자료와 사망진단서에서만 암으로 확인가능한 자료(Death Certificate Only: DCO)를 포함하여 우리나라 전체 암발생자인 [국가암발생DB]를 구축하였다. [암발생통계조사]란 국민건강보험공단의 중증질환(암)등록자료 및 암수진자료 중 중앙암등록모(母)DB에 등록되지 않은 암발생 추정자를 대상으로 한 의무기록조사이다. [암발생통계조사]는 관할 지역에 따라 중앙암등록본부와 지역암등록본부에서 조사를 실시하였으며, 의무기록 조사결과 암발생자로 확인되면 [국가암발생DB]에 추가하였다.

민건강보험공단 및 건강보험심사평가원, 그 밖에 암에 관한 사업을 하는 법인・기관・단체에 대하여 보건복지부령으로 정하는 바에 따라 암등록통계사업에 필요한 자료의 제출이나 의견의 진술 등을 요구할 수 있다. 이 경우 자료의 제출을 요구받은 자는 특별한 사유가 없으면 요구에 따라야 한다.
③ 보건복지부장관은 암등록통계사업과 관련하여 고유식별정보를 처리하는 경우에는 개인정보 보호를 위하여 보건복지부령으로 정하는 바에 따라 암호화 등 안전성 확보에 필요한 조치를 하여야 한다.

암관리법은 암등록통계사업을 위한 자료의 수집, 분석 관리 등을 하는 중앙암등록본부와 지역암등록본부를 지정하도록 하고 있다.

제17조(중앙암등록본부 및 지역암등록본부의 지정 등) ① 보건복지부장관은 다음 각 호의 업무를 수행하기 위하여 국립암센터 또는 보건복지부령으로 정하는 시설・인력・장비 등의 기준을 충족하는 암 전문 연구기관 중 1개 기관을 중앙암등록본부로 지정할 수 있다.
1. 암 발생율 및 생존율 등 암 통계 산출을 위한 자료의 수집・분석・관리
2. 암등록통계사업과 관련한 조사・연구사업
3. 암등록통계사업과 관련한 교육훈련・국제협력
4. 지역암등록본부 지원
5. 그 밖에 암등록통계사업과 관련하여 보건복지부장관이 필요하다고 인정하는 사업
② 보건복지부장관은 다음 각 호의 업무를 수행하기 위하여 제19조에 따른 지역암센터 또는 보건복지부령으로 정하는 시설・인력・장비 등의 기준을 충족하는 관련 전문기관 중 1개 기관을 특별시・광역시・특별자치시・도・특별자치도(이하 “시・도“라 한다)별 지역암등록본부로 지정할 수 있다. <개정 2015.12.29.>
1. 해당 지역의 암 발생률 및 생존율 등 암 통계 산출을 위한 자료의 수집・분석・관리
2. 해당 지역의 암등록통계사업과 관련한 조사・연구사업
3. 그 밖에 암등록통계사업과 관련하여 보건복지부장관 또는 중앙암등록본부의 장이 필요하다고 인정하는 사업
③ 보건복지부장관은 중앙암등록본부 및 지역암등록본부가 다음 각 호의 어느 하나에 해당하는 경우에는 그 지정을 취소할 수 있다.
1. 제1항 또는 제2항에 따른 업무를 수행하지 아니하거나 제18조에 따른 지도・감독을 따르지 아니한 경우
2. 제1항 또는 제2항에 따른 지정 기준에 미달한 경우
3. 그 밖에 대통령령으로 정하는 사유에 해당한 경우
④ 중앙암등록본부 및 지역암등록본부의 지정 절차 등에 관하여 필요한 사항은 보건복지부령으로 정한다.

그런데 암관리법은 암등록통계사업을 수행하면서 수집되는 정보에 대한 비밀보호와 관련해서 중앙암등록본부(국립암센터)와 지역암등록본부(지역암센터)로 하여금 ‘암등록통계사업으로 수집된 건강정보는 암등록통계사업의 목적으로만 사용되어야 한다’의 비밀보호 의무에 대해서는 아무런 규정을 두지 않고 있다.

다만 ‘암관리사업에 종사하거나 종사하였던 사람은 개인정보보호법 제18조 제2항에 따른 경우를 제외하고는 업무상 알게 된 개인정보를 타인에게 제공 또는 누설하거나 목적 외의 용도로 사용하여서는 아니 된다.’는 규정을 두고 있을 뿐이다.

제49조(개인정보의 목적 외 사용 금지) 이 법에 따라 암관리사업에 종사하거나 종사하였던 사람은 「개인정보보호법」 제18조제2항에 따른 경우를 제외하고는 업무상 알게 된 개인정보를 타인에게 제공 또는 누설하거나 목적 외의 용도로 사용하여서는 아니 된다.

이러한 암관리법은 다음과 같은 문제가 있다.

첫째, 암관리법은 보건복지부장관이 '암등록통계사업에 필요한 자료'의 제출(암 환자의 진료와 관련된 자료 및 의무기록 등)이나 의견의 진술 등을 요구할 수 있다고 아주 포괄적으로 규정하고 있을 뿐, 해당 자료가 무엇인지를 시행령이나 고시에서도 규정하지 않고 있다. 특히 수집되는 정보는 개인정보보호법의 적용대상에서 배제하고 있으므로 수집 대상 정보가 무엇인지를 명확하게 법령이나 고시로 규율할 필요가 있다.

둘째, 보건복지부장관이 자료를 요구할 수 있는 상대방에 의료인 또는 의료기관, 국민건강보험공단 및 건강보험심사평가원은 물론 암에 관한 사업을 하는 법인·기관·단체를 포함하고 있는데, 그 대상이 명확하게 한정되어 있지 않다.

셋째, 보건복지부장관이 자료나 의견 진술을 요구할 경우 자료 제출을 요구받은 자는 특별한 사유가 없으면 요구에 따라야 한다는 규정을 두고 있는데, 제출 여부를 판단하는 자가 환자 본인이 아니라 의료기관, 의료인, 건강보험공단 등으로 되어 있을 뿐이다. 환자 본인의 거절권이 보장되는 것이 바람직하다.

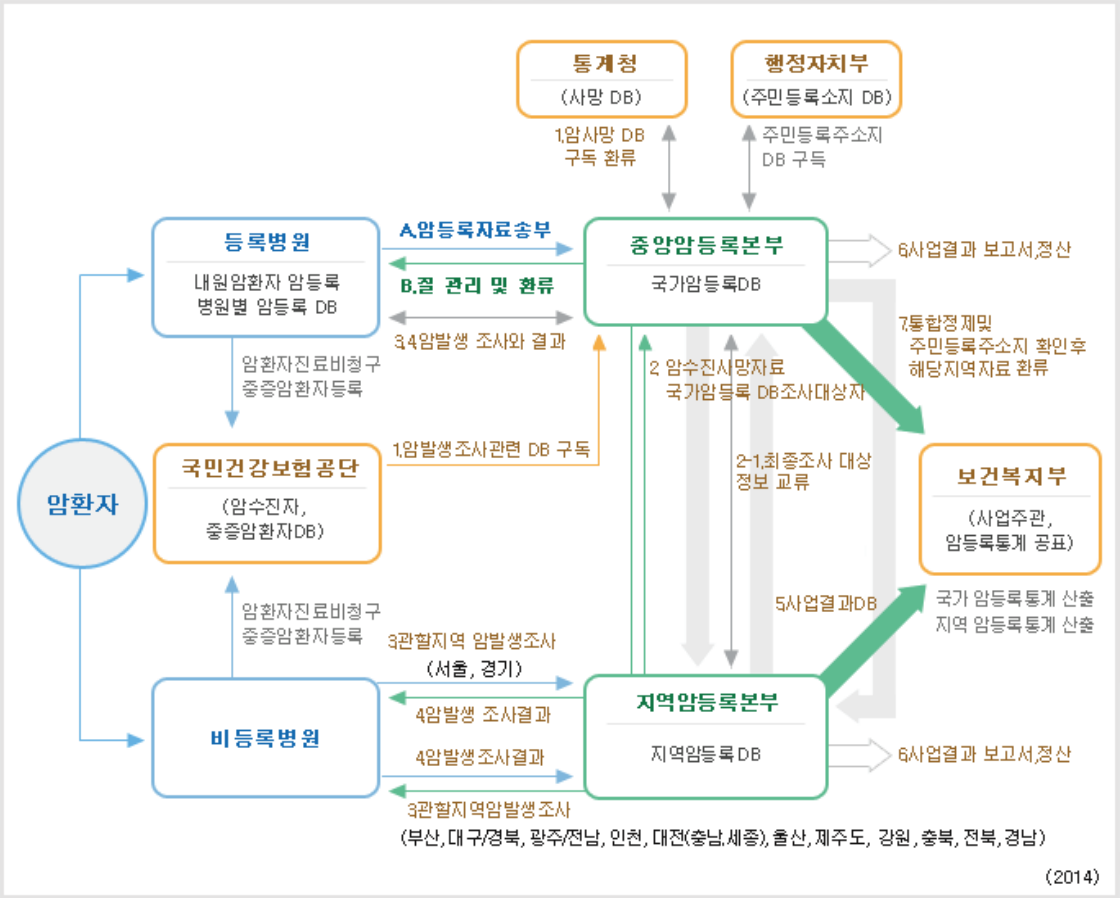
넷째, 수집되는 정보는 매우 민감한 정보이고, 특히 그 정보가 유출되는 경우 해당자는 사회적으로 고용, 보험에서는 물론 사회 전반의 생활에서 차별과 불이익을 입게 될 수도 있다. 이런 점을 고려한다면 암등록통계사업은 명확하게 규정된 법령에 입각해서 이루어져야 한다.

다섯째, 암등록통계사업을 위해 수집한 환자들의 암 관련 정보들은 암등록통계사업의 목적으로만 활용될 수 있도록 규정할 필요가 있다. 그런데 국립암센터는 수집한 정보를 확장하거나, 연계하려는 방안을 구상하고 있다. 그중에는 암등록통계사업의 범위를 넘는 것도 있다.[276]

276) 즉, 지역사회 일반인구 대상 코호트(총 17,715명)와 국립암센터 방문 검진자 대상 코호트(누적 42,784명) 자료 활용을 통하여 생활습관, 유전, 환경 등 암관련 위험요인 평가와 관련 정책의 기초자료 제공, 지역사회 코호트 및 검진자 코호트의 경우 대상자로부터 수집된 다양한 설문정보(인구사회경제 학적 특성, 건강 행태, 보유 증상 및 질환, 기초 건강검진 정보 등)와 함께 혈액, 세포, 소변 등 다양한 생체시료를 보유하고 있어 현재의 기초, 임상, 역학 등 다양한 영역의 연구지원 가능은 물론 미래 연구에 활용할 수 있는 잠재적 리소스, 코호트 보유 생체시료 활용한 유전체, 단백체, 대사체 연구 등에서 분석 결과 제시된 관련 정보는 환류하여 추가적인 정보로 확보하는 한편, 암(환자의) 조직 등 생물학적 검체의 체계적이고 효율적인 보관 및 관리, 보관 관리되는 검체(조직)와 관련 환자 임상정보 및 병리학적 정보를 데이터베이스화, DNA, RNA, protein 등의 분리 및 보관 등 국립암센터 병원 내원 환자를 대상으로 한 종양은행 구축 및 활용을 통하여 암관련 유전체, 단백체, 대사체 연구 등 수행 지원. 타 기관과의 연계를 통한 전국적

여섯째, 수집된 정보는 개인정보보호법이 적용 배제되고, 개인정보보호 관련 규정이 매우 미비한 통계법만 적용되므로 특히 비밀보호에 대한 개인정보보호 규정의 보완이 필요하다. 현재 암관리법에 의해서 처리되는 정보들은 민감정보임에도 불구하고 개인정보보호가 매우 미비한 형편이다.

그림 4-21 암등록통계사업의 개요



* 출처: 국립암센터 http://www.ncc.re.kr/main.ncc?uri=hq_register01_1

다. 국립암센터 및 지역암등록본부

국립암센터는 중앙암등록본부로 지정되어, 암 발생률 및 생존율 등 암 통계 산출을 위한 자료의 수집·분석·관리, 암등록통계사업과 관련한 조사·연구사업, 암등록통계사업과 관련한 교육훈련·국제협력, 지역암등록본부 지원, 그 밖에 암등록통계사업과 관련하여 보건복지부장관이 필요하다고 인정하는 사업을 수행하고 있다(제17조).

국립암센터와 지역의 대학병원들이 지역암등록본부로 암등록사업을 수행함에 있어서 환자들의 암 관련 정보들은 개인정보보호법 적용대상에서 제외되고, 통계법에는

종양은행 네트워크의 설립 및 운영 근간 마련 및 전향적 확대와 활용 방안 구상 중이라고 한다.

통계자료의 비밀보호 규정이 미비하고, 암관리법도 비밀보호규정이 미비하므로 암등록사업의 비밀보호와 개인정보보호는 충분하다고 보기 어렵다.

한편 보건복지부장관은 암에 관한 정보를 지속적이고 체계적으로 구축하여 효율적으로 국민에게 제공하는 암정보사업도 시행한다. 암정보사업에는 암에 관한 각종 정보 수집, 데이터베이스 구축 및 관리, 국민에 대한 암 정보 제공 및 상담, 암에 관한 교육자료 개발 및 교육·홍보, 그 밖에 암 정보 데이터베이스 구축과 제공에 관하여 보건복지부장관이 필요하다고 인정하는 업무가 포함되는데, 국립암센터[277]가 이를 수행하고 있다(제15조).

라. 암 발생 자료

한국중앙암등록본부 암 발생 자료는 암 발생 위험요인, 암의 발생 및 처치에 대한 자료를 체계적으로 수집·분석하여 암의 발생률, 생존율, 유병률 등에 대한 통계를 산출하기 위한 자료라고 한다. 이를 위해 여러 자료를 연계하여 작성한다.

즉, 각급 병원으로부터 등록받은 암 환자 자료를 기반으로 국민건강보험공단의 암검진수검 정보, 중증질환자등록(산정 특례) 정보, 통계청의 사망원인 자료 등과의 연계 및 의무기록 조사를 통해 등록되지 않은 암 발생자 및 암 사망자 확인 후 암 통계를 산출한다. 이때 자료수집 대상은 조직학적 확진 여부에 관계없이 병원에서 진단 또는 치료받은 암 환자라고 한다.[278]

암관리법 제14조 제2항은 '보건복지부장관은 암 환자를 진단·치료하는 의료인 또는 의료기관, 「국민건강보험법」에 따른 국민건강보험공단 및 건강보험심사평가원, 그 밖에 암에 관한 사업을 하는 법인·기관·단체에 대하여 보건복지부령으로 정하는 바에 따라 암등록통계사업에 필요한 자료의 제출이나 의견의 진술 등을 요구할 수 있다.'고 규정하고 있으므로, 이 규정을 근거로 필요한 자료의 제출을 요구할 수 있다. 이는 법령의 근거가 되므로 환자의 동의가 없어도 민감정보를 수집할 수 있다. 그러나 이와 같이 수집한 정보를 연계해도 되는지는 의문이다.

---

277) 암관리법에 의하여 암에 관한 전문적인 연구와 암환자의 진료 등을 위하여 국립암센터를 법인으로 설립·운영하고 있다(제27조).

278) 이덕형, 앞의 논문, p85.

그림 4-22 중앙암등록본부 암발생 자료수집 및 활용 체계

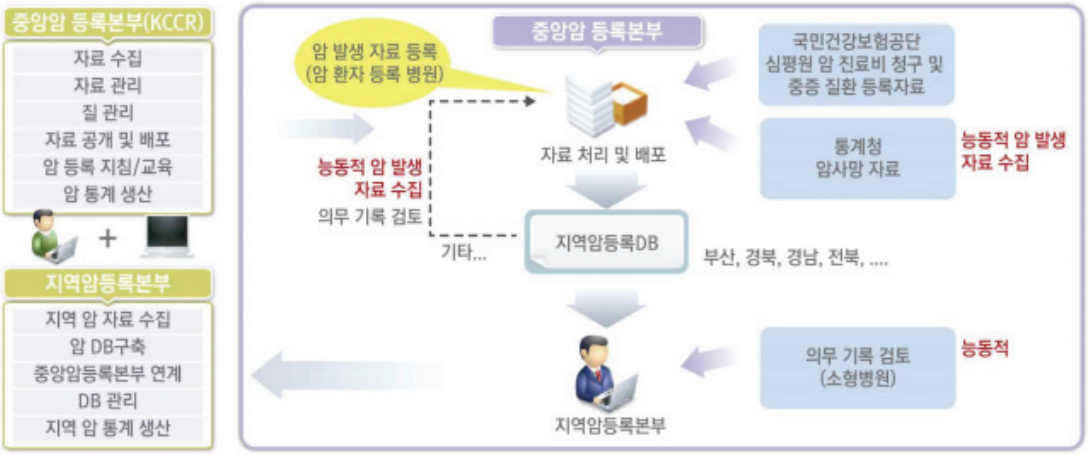


표 4-35 중앙암등록본부 암 발생 자료 수집 항목

항목		주요 조사내용
인적정보		환자등록번호, 출생년도, 나이, 주민등록번호, 이름, 행려환자, 직업, 주소
질 병 정보	진료	초진연월일, 입원일, 퇴원일, 치료시행여부, 진단경로, 원발부위, 전이부위
	조 직 학 적 진단	조직학적 진단명, 진단방법, 병기분류(SEER summary stage code), 분화도(2012년 발생자부터 수집)
	사망	사망년월일, 사망원인
입력정보		입력일, 입력자

그림 4-23 중앙암등록 자료 구조 (2012년)

번호	변수명	유형	자리수	설 명	예	비고
1	HCODE	문자	4	병원번호	국립암센터 → 0177	
2	PTNO	문자	15	환자등록번호	02428	
3	AGE	문자	3	나이	자동계산항목(56)	
4	REGNO	문자	13	주민등록번호	******-1******	
5	NAME	문자	20	이름	홍길동	
6	SEX	문자	1	성별	남→1, 여→2	
7	FRG	문자	1	외국인	외국인→1, 내국인→0, 모름→9	2005
8	HMLS	문자	1	행려환자	행려환자→1, 일반환자→0, 모름→9	2005
9	JOB	문자	2	직업분류	판매원→04, 분류불능→10	
10	JOB1	문자	20	직업이 '분류불능(10)' 일 때 직접 기술	(분류되지 않은 직업명)	
11	UNKNOWJOB	문자	20	각 직업의 구체적 기술	교사, 자영업자	
12	ADR1	문자	50	주소(시, 구, 군, 동, 읍, 면)	고양시 일산동구 마두1동	
13	ADR2	문자	50	세부주소	809번지	
14	ZIP	문자	6	우편번호	154895	
15	FDX	문자	8	초진연월일	20061101	
16	VISITPATH	문자	1	진단경로	증상발현→3	2012
17	VISITPATH_DESC	문자	50	진단경로가 "5. 기타" 인 경우 세부설명 입력		2012
18	TCODE	문자	4	원발부위코드	C220	
19	TCODE1	문자	100	원발부위코드 설명	Liver	
20	TCODE2	문자	100	원발부위코드가 코드설명내용과 다른경우 입력	Hepatic, NOS	
21	LATER	문자	1	편측성	Right→1, Left→2	2012
22	MCODE	문자	5	조직학적진단명 코드	81603	
23	MCODE1	문자	100	조직학적진단명 코드 설명	Cholangiocarcinoma	
24	MCODE2	문자	100	조직학적진단명이 코드설명과 다른경우 입력	Bile duct carcinoma	
25	BUNHWADO	문자	1	분화도 (grade)	well differentiated→1	2012
26	EXPIRE	문자	8	사망연월일	20061127	
27	SAIN	문자	4	사망원인, ICD-10(KCD-6 , 한국표준질병사인분류 6차 개정판) 사용	C169	2007
28	METHOD	문자	1	진단방법	임상검사(CT)→2	
29	METHOD2	문자	4	진단방법이 임상검사(2)인 경우 세부사항표시 (SONO, CT, MRI, 기타)	CT→0100, SONO,MRI→1010	
30	METHODETC	문자	20	진단방법이 임상검사(2)의 '기타' 사항인 경우 기술	angiogram	
31	TXCHECK	문자	1	치료 시행 여부	시행→1, 시행하지 않음→0	2007
32	TX	문자	5	수술,화학요법,방사선요법,면역요법,호르몬요법	11000, 01000	
33	TXETC	문자	50	기타치료	TEA, PTBD	
34	REGYEAR	문자	2	출생 연대(century) 표기	2000년대생→20	
35	ADMISDATE	문자	8	입원일	20060117	
36	DISDATE	문자	8	퇴원일	20060131	
37	DBYEAR	문자	4	자료등록년도	2007	
38	INPUTDAY	문자	8	입력일	20061127	
39	INPUTMAN	문자	20	입력자	아무개	
40	LICENSENO	문자	5	의무기록사 면허번호	99999	2005
41	SEERCODE	문자	1	SEER SUMMARY STAGE CODE	in situ→0	
42	SEERNAME	문자	80	SEER SUMMARY STAGE CODE 설명		
43	STAGECODE	문자	20	TNM, FIGO, DUKES, JEWETT, GLEASON'S SCORE 등	TNM	
44	STAGEDESC	문자	40	각 해당하는 병기 STAGE 직접 입력	T2N0M0	
45	METACODE 1	문자	4	원격전이된 부위 ICD-O-3 코드로 입력 1	Lung→C349	2012
46	METACODE 2	문자	4	원격전이된 부위 ICD-O-3 코드로 입력 2	Brain→C719	2012
47	METACODE 3	문자	4	원격전이된 부위 ICD-O-3 코드로 입력 3	Bone→C419	2012
총 계			977			

* 암등록병원자료는 국가암등록통계시스템(http://ncrs.cancer.go.kr/index.do)을 통하여 개별 입력하거나 병원 자체 암등록시스템 또는 엑셀파일로 입력된 암등록 대용량 자료를 송부

** 암발생자 확인을 위한 자료원은 의무기록(외래기록지, 입원기록지, 응급실기록지), 영상의학 보고서(신환목록, 치료기록, 방사선 치료, Xray, Ultrasonogram, CT, MRI), 혈액검사 보고서, 병리보고서 (병리의사들이 사용하는 SNOMED code검토), 세포검사 보고서, 부검보고서, 사망진단서, 중증질환(암) 등록자료 등임

마. 자료의 공개

중앙 암등록본부의 자료는 통계법의 적용대상이 되므로 비밀보호의 대상이 되고 통계나 연구목적 외의 목적으로 활용되어서도 안 된다.

다만 현재 중앙암등록본부에 등록된 원시 자료 및 주문형 자료는 개인정보보호법 및 의료법 제20조에 준하여 '등록환자의 개인정보가 보호될 수 있는 범위 내에서 제공하고 있다'고 하는데 이는 소위 '비식별 조치'를 취한다는 의미이다. 그런데 어느 정도의 비식별 조치를 취했어야 '개인정보가 보호될 수 있는 범위'가 될 수 있는 것인지는 논란이 된다. 의료정보의 경우는 비식별화를 하더라도 일부의 주소, 초진연월일, 입원일, 퇴원일, 치료시행여부, 진단경로, 원발부위, 전이부위, 조직학적 진단명, 진단방법, 병기분류(SEER summary stage code), 분화도, 사망연월일, 사망원인 등의 자료만으로도 특정인을 식별해 낼 수 있다. 따라서 이런 행위는 법적으로 문제의 소지가 있다. 만약 대상자의 동의서가 있다면 외부자료를 원시 자료에 연계하여 제공하고 있다고 한다.

사망원인자료의 제공은 1997~2012년까지 온라인으로 가능하며, 1991-1996까지의 원시 자료는 기관과 논의해야 한다고 한다. 사망자 자료 연계와 관련하여 사망원인통계는 개인식별정보(주민등록번호, 성명, 주소 등)는 제공하지 않는다고 한다. 그러나 개인식별정보가 없어도 의료기록은 특이성이 있으므로 충분히 식별 가능성이 있다.

한편, 국가통계작성을 목적으로 공공기관 및 사망원인통계 작성협의체가 개인식별정보를 요청하는 경우에는 인구동향과에서 이용 목적 및 최소 정보 제공 원칙에 따라 내용을 검토한 후 제공 여부 및 제공방식을 결정한다고 한다. 사망원인통계작성 협의체는 근로복지공단, 국립암센터, 국민건강보험공단, 산업안정공단, 질병관리본부 등으로 구성되어 있다고 한다. 일부 항목에 대해서는 개인정보보호를 위해 개인 식별이 불가능하도록 5건 이하 수치를 제공하지 않고 있다고 한다.[279]

바. 국가암관리사업본부의 사업별 DB 연계

국립암센터는 암관리법의 4개 사업별 DB를 연계·통합하여 통합데이터베이스로 구축하고 이를 활용해 시범서비스를 개발하고 있다고 한다. 즉, 암등록통계사업(제14조), 암검진사업(제11조), 암환자의 의료비 지원사업 등(제13조), 완화의료사업(제21조) 등 보건복지부가 암관리법에 따라서 국립암센터에게 위탁한 사업의 DB를 하나로 통합하는 것이다. 이와 같은 4개 사업 DB의 연계·구축이 적법한 것인지는 논란이다.

---

279) 이덕형, 앞의 논문, p88~89.

그림 4-24 국립암센터 내 4개 사업과의 데이터 통합

이와 관련해서 정부법무공단은 국가암관리사업본부 4개 사업 연계 DB 구축이 암관리법 제14조 제2항에 의해 가능하다는 의견을 제시했다고 한다. 암관리법 제14조 제2항은 암등록통계사업에 필요한 자료나 의견을 제출받을 수 있는 근거인데, 암등록통계사업은 '암 발생 위험요인과 암의 발생 및 치료에 관한 자료를 지속적이고 체계적으로 수집·분석하여 암 발생률, 생존율 등의 통계를 산출하기 위한 등록·관리·조사사업'이므로 암등록통계사업, 암검진사업(제11조), 완화의료사업(제21조)은 암 발생 및 치료에 관한 직접적인 자료이고 암 환자 의료비 지원사업(제13조)도 암 치료와 직·간접적으로 관련된 것이어서 암등록통계사업의 범위에 포함될 수 있다는 것이다.[280]

그러나 암등록통계사업의 근거 규정에 따라서 이미 암등록통계 DB를 별도로 구축해서 운용하고 있으면서 그와 별도로 목적이 다른 3개 사업을 위한 DB를 구축하고 있는 상황에서는 각각 DB 구축의 목적이 다를 것임에도 별도의 법적 근거 없이, 개별 정보의 필요 여부 판단도 없이 DB를 통합하는 것이 적법한지는 의문의 여지가 있다.

사. 암등록정보와 다른 정보와의 연계

국민건강보험공단과 국립암센터는 2013. 5. 1 ~ 2016. 4. 30 '한국 암 성과 연구의

280) 이덕형, 앞의 논문, p80.

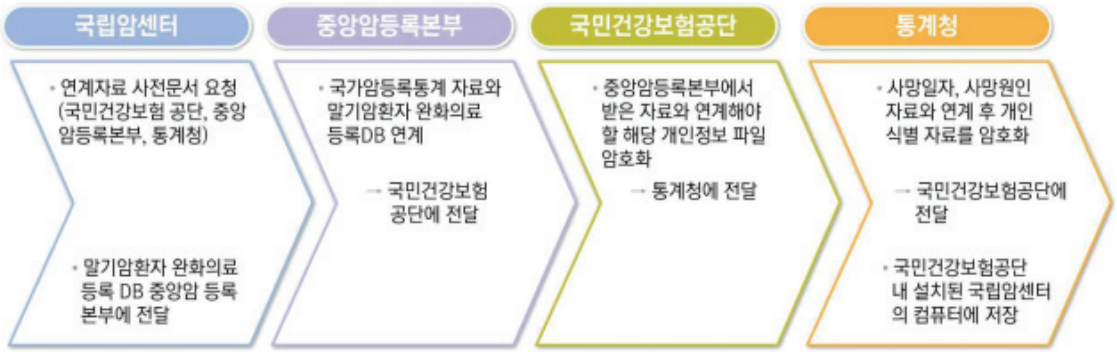
체계화'라는 공동연구 협약을 체결하고, 국립암센터 암등록자료, 건강보험공단 암 환자 자료, 통계청 사망자료 등 연계를 통하여 암 진단 및 치료 효과, 국가암관리사업 효과 평가 등을 수행하고자 관련 자료의 수집 및 연계와 분석 적합성 점검을 진행하였다고 한다.[281]

그림 4-25 암등록정보와 건보공단의 진료 정보의 연계

이를 통해 12년간(2001~2012)의 국립암센터의 암 등록·완화의료등록 자료와 국민건강보험공단의 검진·자격·의료이용자료, 통계청의 사망원인 자료를 연계하여 대규모 국가 단위의 암 종합 DB(2001~2012)를 구축했다고 한다.[282]

이 종합 DB는 아래와 같이 보건복지부 질병정책과의 공문을 근거로 연구 관련 자료 및 협조 요청을 하여 암 종합 DB를 연계하고, 병합했다고 한다.

그림 4-26 암 종합 DB 병합 절차



281) 이덕형, 앞의 논문, p98.
282) 이덕형, 앞의 논문, p97.

이와 같이 생성한 국가 단위의 암 종합정보 DB는 연계한 행정자료가 진료비 청구와 심사를 목적으로 생산되는 자료로서 '비용 부분'에 있어 정확도가 상대적으로 높고 자료의 포괄성과 시의성, 추적조사 등에 보완적으로 연계 및 활용이 가능하고 2001년부터 2010년까지 상대적으로 장기간을 대상으로 구축되었고 국민건강보험 의료이용자료의 특성상 의료 이용자의 건강상태와 관련된 정보가 없다는 단점을 중앙암등록본부 암 발생 자료의 암 병기 자료로 보완 가능하다는 점을 장점으로 뽑고 있다. 한계로는 보험 청구자료이기 때문에 질병코드, 청구 진단명 등에 정확도가 낮아 이차 행정자료를 활용한 연구 분야에 타당도 문제가 발생할 수 있다는 점, 급여가 아닌 비급여 자료와 치료 및 검사의 결과자료와 삶의 질·만족도 등의 설문자료가 없다는 점을 들고 있다.[283]

그런데 이와 같이 대규모로 민감한 건강정보를 연계하는 것과 관련하여 법적 근거는 미비하다. 암 환자의 정보 등은 매우 민감한 정보이기 때문에 법적 근거가 명확해야만 데이터의 연계가 허용되어야 하는데, 이에 대한 법적 근거로 볼 수 있는 것이 없기 때문이다. 특히 본 DB 연계는 암관리법 제14조 제2항이 적용될 수 있는 조건으로 보기도 어렵다. 물론 아래에서 보는 것처럼 암과 관련한 연구과제의 해결이 중요하고, 연구가 필요하기는 하지만, 10여 년의 환자의 민감한 개인정보가 환자에게 아무런 선택권도 없이 통합되어 만들어졌다는 것은 문제의 소지가 있다.

그림 4-27 암 예방부터 시기에 따른 연구 주제

시기	주제	자료원	output	outcomes
암 예방 및 발생	생활습관과 암 예방관련성	공단-암(일반검진/청구자료 암등록본부-등록자료	• 생활습관, 동반질환 및 약물복용에 따른 암 발생위험 및 예방 관련성	• 국가암예방정책 변화에 따른 암 발생율 감소 및 경제적 효과 • 국가암검진사업 후 발견율 및 병기 변화, 생존율과 관련 요인 분석 • 국가정책(보장성) 전/후에 따른 효율성, 형평성, 치료 성과(생존율, 경제적 영향 등)의 평가 • 암정복 10개년계획 결과 평가 • 국가암관리정책의 우선순위에 대한 비용-효과 분석 ※국가단위 암 관리체계 구축을 통한 국민의 건강 및 삶의 질 향상 위한 근거자료 제시 및 정책 제언
	암 검진 시 암 발견(시기)	공단-암(일반)검진 암등록본부-등록자료 통계청-사망자료	• 검진 수검에 따른 암 발견율 및 병기 변화, 생존율	
암 치료	암 진단후 치료 패턴 및 성과	공단-청구자료 암등록본부-등록자료 통계청-사망자료	• 암 치료의 접근성 • 암 치료에 따른 성과 • 동반질환관리 패턴 및 성과	
생존자 관리	암 진단후 생존율 및 2차암 발생	공단-암(일반)검진/청구자료 암등록본부-등록자료 통계청-사망자료	• 생활습관, 동반질환 및 약물복용 등에 따른 암 생존율 및 2차암 발생율 분석 과 新 예측 모형	
사망	말기 및 사망	공단-청구자료 암등록본부-등록자료 통계청-사망자료 암센터-eVelos말기암환자정보	• 말기암환자 관리 • 사망 전 의료이용행태 및 의료비	
경제적 영향	보건경제적 영향	공단-암(일반)검진/청구자료 암등록본부-등록자료 통계청-사망자료	• 암예방, 암검진, 보장성 강화 정책 등 국가암관리정책의 경제적 영향	
연속성	암진단부터 삶의 마무리	공단-암(일반)검진/청구자료 암등록본부-등록자료 통계청-사망자료	• 암진단부터 치료, 생존자 관리, 삶의 마무리까지의 10년간의 변화 및 예측	

283) 이덕형, 앞의 논문, p98~99.

아. 암등록자료의 제공과 활용

암등록자료에 대한 신청은 연구계획서를 첨부하여 신청하도록 하고 있다. 아울러 연구윤리심의위원회 승인이 필요한 경우는 승인서도 제출하도록 한다.[284] 여기에서도 확인할 수 있는 것처럼 식별자료를 포함해서 자료를 연계하는 것도 가능하다.

표 4-36 암등록자료 신청 서식 : 자료 구분 및 요청내용

자료 구분	식별자료 이용여부	□ 예(자료연계)	□ 아니오(통계표)
요청 내용 및 범위			
이용목적			

표 4-37 암등록자료 신청 서식 : 자료연계

연구비 제공기관	기관명		연구비 유형	
연구기간	시작일		종료일	
연구윤리심의위원회 승인여부		□ 승인	□ 미승인	□ 해당사항 없음
자료 출처	기관명		식별자료 이용 동의 획득여부	□ 예 □ 아니오

[자료이용 시 준수사항]
- 개별 자료에 의거 알게 된 사항에 대한 누설금지
  ·다른 자료와 연계를 통해 알게 된 특정 개인·법인·단체의 정보 포함
- 이용목적 이외의 사용금지
- 제공 자료의 복제 및 대여금지

※ 위 사항을 위반 할 경우 중앙암등록본부는 이용자에게 자료의 폐기 또는 반환을 요구할 수 있으며, 향후 자료제공 이용에 제한될 수 있음을 알려드립니다.

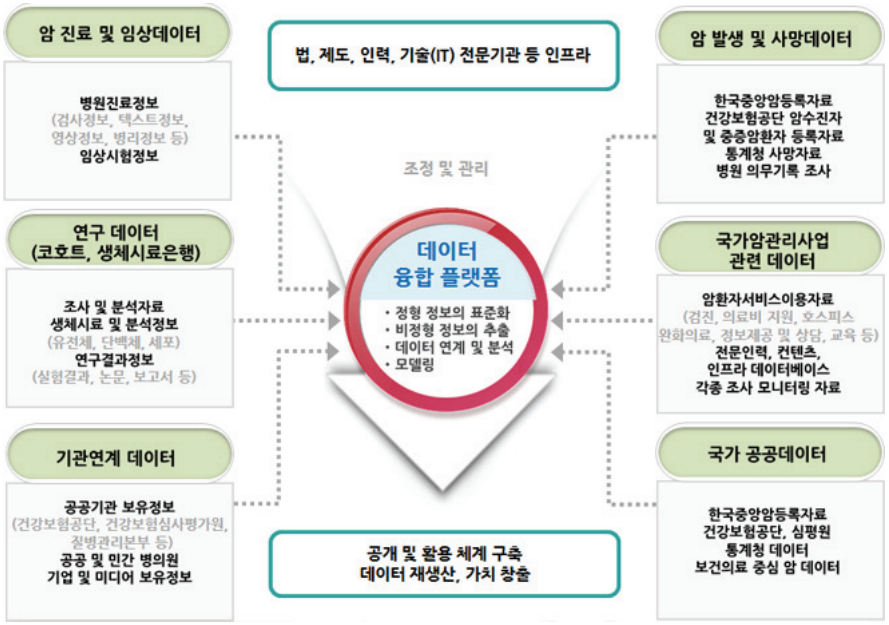
자. 암 빅데이터 구축 사업

국립암센터는 현재 아래와 같은 암 빅데이터센터를 구축하고 있다고 한다. 아래의 개요에 의하면 암 진료 및 임상데이터, 연구 데이터, 기관 연계 데이터, 국가 암관리 사업 관련데이터를 연계·융합하여 활용하는 것을 목적으로 제시하고 있다.[285] 그런데 이와 같은 연계는 법적 근거가 희박한 것이다.

284) 국립암센터 홈페이지 참조

285) 이덕형, 앞의 논문, p11; http://www.ncc.re.kr/main.ncc?uri=hq_giddata

그림 4-28 암 빅데이터센터의 개요



차. 평가

암등록사업의 필요성을 인정하더라도 암등록자료는 개인의 민감정보에 해당한다. 따라서 해당 정보주체의 동의 없이 민감한 개인정보를 연구목적으로 제공하는 것은 적법한 법적 근거를 갖출 필요가 있다.

(2) 국립보건연구원의 연구자원 인프라

국립보건원은 연구자원 인프라 사업으로 19개 코호트 사업, 한국인체자원은행사업, 국가병원체자원은행, 매개체자원, 국립줄기세포재생센터(국가줄기세포은행) 사업을 수행하고 있다.

가. 19개 코호트 사업

① 사업의 개요

의학/보건학 분야에서 연구자 주도의 코호트가 구축되어 왔지만, 최근 질병관리본부가 주관하는 한국인유전체역학조사 자료가 향후 국가 수준의 주요 코호트 자료로 기능할 가능성이 크다. 2003년도에 안성, 안산코호트 자료(매 2년마다 추적 중)를 비롯하여 다양한 인구집단을 대상으로 하여 코호트를 구축하여, 2007년 현재 약 10만 명의 자료가 구축되어 있다. 2008년까지 20만 명 자료가 구축 예정이다. 이 조사자료

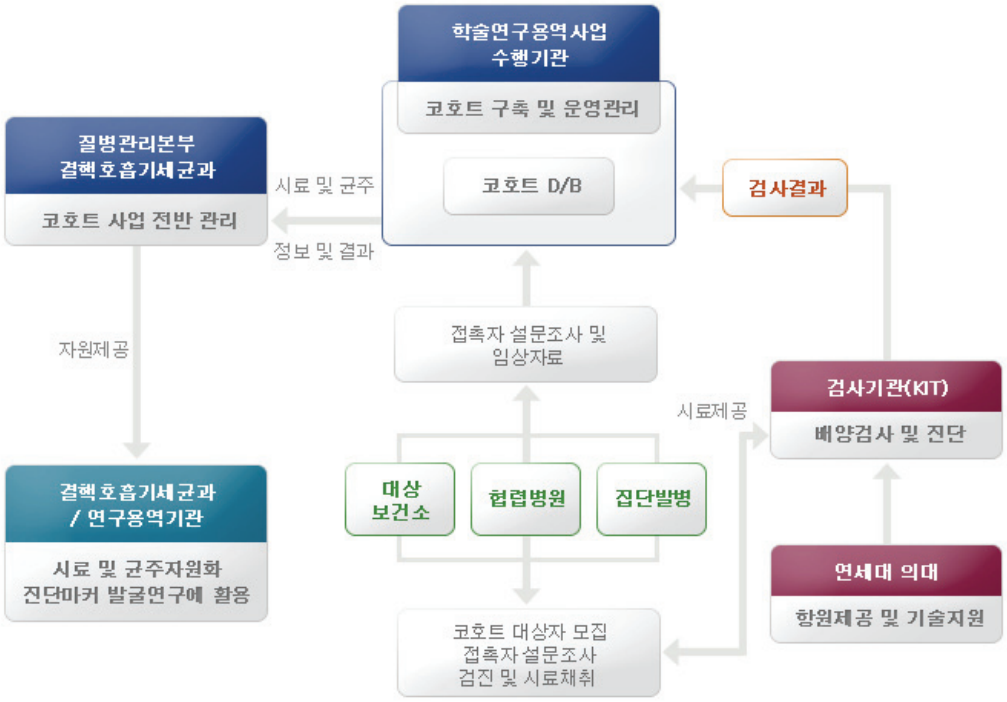
의 경우 추적조사 여부가 매우 중요한 쟁점이 될 전망인데, 안성, 안산코호트 등 일부를 제외할 경우, 상당수 자료, 특히 검진기관을 기반으로 한 자료는 기존 2차 자료(사망자료, 건강보험자료, 암등록자료)와의 연계 연구가 주요한 추적 방법이 될 가능성이 크다[286].

표 4-38 국립보건연구원 19개 코호트 사업의 분류

일반인대상 코호트	지역사회 기반 코호트 (안산・안성)농촌 기반 코호트도시 기반 코호트
환자군 코호트	HCV에 의한 간질환 코호트HIV 코호트HPV 코호트결핵 고위험군 코호트노인 천식 코호트알레르기 비염 코호트주선천성기형 코호트급성심근경색 질환 코호트다낭성난소증후군 질환 코호트심부전 질환 코호트
특수집단 및 모델 코호트	국내 이주자 및 국제 협력 코호트 II국제 협력 코호트쌍둥이 및 가족 코호트소아호흡기・알레르기질환 장기추적 코호트소아비만 및 대사질환 코호트노인질환 예방관리 코호트

② 코호트 사업의 절차[287]

그림 4-29 코호트 사업의 절차-결핵고위험군 코호트의 경우



286) 정영호·고숙자·이용갑·서남규·태윤희·이원영·이경용·김범수·강영호(2009), "한국의료패널의 활용과 기대효과", 한국보건사회연구원·국민건강보험공단, p171.

287) http://www.nih.go.kr/NIH_NEW/contents/NihKrContentView.jsp?cid=19755&menuIds=HOME005-MNU0849-MNU0863-MNU1036

이 조사에서는 조사참여자로부터 광범한 건강위험요인에 대한 정보를 취득하였는데, 유전정보(DNA)를 비롯하여 기저 건강수준, 건강행태, 식이에 대한 매우 충실한 자료가 구축되었다. 하지만, 의료서비스 연구의 측면에서는 기저 수준의 의료서비스 이용(이용량 및 의료비)과 관련된 정보는 취약하다는 문제점이 있다.

③ 데이터 연계

코호트 조사의 경우 주민등록번호 정보가 거의 모든 대상자로부터 수집되어 있으므로 2차 자료와의 연계가 가능하다. 국민건강보험공단이나 심평원의 데이터베이스와 연계하여 다양한 정보의 추출이 가능하다.

특히 의료서비스 이용을 결과변수로 한 연구의 가능성은 크게 열려 있다고 할 수 있고, 표본 수가 매우 크므로 매우 희귀한 결과변수에 대한 연구의 가능성에도 주목할 필요가 있다.[288]

나. 한국인체자원은행사업

인체 자원은 인체유래물(인체로부터 수집하거나 채취한 조직·세포·혈액·체액 등 인체 구성물 또는 이들로부터 분리된 혈청, 혈장, 염색체, DNA, RNA, 단백질 등)과 해당 인체 유래물 기증자의 임상, 역학정보 및 이로부터 분석된 유전정보 등을 말한다.

국립중앙인체자원은행(국립보건원)과 17개 대학병원소재 인체자원단위은행이 한국인체자원은행네트워크(Korea Biobank Network, KBN)를 구성하여, 대규모 인구집단 기반(중앙은행) 및 질병 기반 인체 자원(17개 단위은행)을 수집, 관리하여 국내 연구자들에게 분양하고 있다.[289]

다. 국립줄기세포재생센터(국가줄기세포은행)

국가줄기세포은행에서는 줄기세포주를 수집하고 각 세포주의 특성분석을 통해 국제표준의 고품질 줄기세포주를 대량 배양, 보관하여 연구자들에게 분양하는 중앙저장고의 역할을 담당한다. 줄기세포주를 등록할 때는 줄기세포주 수립에 이용된 잔여배아 이용에 관한 동의서 등을 제출해야 한다.

---

288) 정영호 외, 앞의 논문, p171.

289) 인체유래물이 법률 용어임에도 불구하고 인체자원이라는 용어를 사용하는 것은 혼동을 야기할 우려가 있다.

5. 인간대상연구 등 개인정보 활용 연구

(1) 개인정보를 수집·활용하는 연구에 대한 규율

개인정보보호법은 개인정보를 활용하는 연구나 건강정보를 활용하는 연구에 대하여 한 개의 조문을 두고 있다. “개인정보처리자는 학술연구의 목적을 위하여 필요한 경우로서 특정 개인을 알아볼 수 없는 형태로 개인정보를 제공하는 경우에는 정보주체 또는 제3자의 이익을 부당하게 침해할 우려가 있을 때를 제외하고는 개인정보를 목적 외의 용도로 이용하거나 이를 제3자에게 제공할 수 있다.”는 규정이다(제18조 제2항 제4호). 이 규정이 적용되는 요건은 다음과 같다.

개인정보처리자이어야 한다. 따라서 특정 개인을 알아볼 수 없는 형태로 개인정보를 제공하는 경우가 아니라면 개인정보처리자에 해당하는 경우에는 개인정보보호법의 규정을 준수해야 한다. 즉, 연구자를 개인정보처리자로 볼 수 있는 경우-업무를 목적으로 개인정보를 쉽게 검색할 수 있도록 일정한 규칙에 따라 체계적으로 배열하거나 구성한 개인정보의 집합물(集合物)을 운용하기 위하여 스스로 또는 다른 사람을 통하여 개인정보를 처리하는 공공기관, 법인, 단체 및 개인 등에 해당하는 경우-라면 개인정보보호법을 준수해야 한다. 따라서 개인정보주체를 알아볼 수 있는 개인정보를 수집하거나 활용하는 연구를 할 경우에는 수집목적, 수집하는 개인정보, 보유기간, 제3자 제공 등에 대하여 고지하고 동의를 받아야 한다.

(2) 생명윤리법의 규율

그러나 생명윤리 및 안전에 관한 법률(약칭 ‘생명윤리법’)에서는 특별한 규율을 하고 있다. 생명윤리법에서 규율하는 연구는 인간대상연구와 인체 유래물 연구인데(제2조 제1호, 제2호), 두 가지 연구의 경우 연구 과정에서 수집되는 정보나, 연구의 대상이 되는 정보가 개인정보에 해당할 가능성이 크다.

생명윤리법의 규율대상인 인간대상연구는 ① 사람을 대상으로 물리적으로 개입하는 연구(중재연구),[290] ② 의사소통, 대인접촉 등의 상호작용을 통하여 수행하는 연구(상

290) 중재(intervention) 연구 즉, 사람을 대상으로 물리적으로 개입하는 연구란, "연구"를 위해 연구대상자에게 어떤 침습적 행위(식품, 의약품 등의 섭취, 혈액채취 등)를 하거나 연구대상자의 환경을 조작(시각, 청각 등에 자극 또는 스트레스 유발) 물리적 개입이 포함된 연구를 수행하고 그 결과를 얻어 연구에 이용하는 것으로 다음의 연구들이 해당할 수 있다. (1) 「약사법 시행규칙」내지 「의료기기법 시행규칙」에 따라 승인된 임상시험계획서에 따라 수행되는 의약품 또는 의료기기를 이용한 임상시험 (2) 식품의약품안전청 고시 「생물학적 동등성시험 관리기준」에 따라 시험기관 에서 수행되는 생물학적 동등성시험 (3) 그 밖에 화장품·건강기능식품·생의약제·생물학적 제제 등에 대한 안전성·효능·효과를 보기 위해 해당

호작용연구)[291], ③ 개인을 식별할 수 있는 정보를 이용하는 연구(개인식별정보를 포함하는 연구)[292] 이다.[293] 중재연구나 상호작용연구는 그 과정에서 수집되는 정보가 개인을 식별할 수 있는 개인정보에 해당하므로 인간대상연구는 모두 개인정보를 수집, 활용하는 연구로 볼 수 있다.

그림 4-30 인간대상 연구등의 범위

* 출처: 공용 IRB 운영 모델 개발, 한국보건의료연구원 (2010), p3.

생명윤리법의 규율대상인 인체 유래물연구는 인체로부터 수집하거나 채취한 조직·세포·혈액·체액 등 인체 구성물 또는 이들로부터 분리된 혈청, 혈장, 염색체, DNA, RNA, 단백질 등을 (연구자가) 직접 조사·분석하여 수행하는 연구를 말한다.

가. 인간대상 연구, 인체 유래물 연구에 대한 규율

① 연구대상자의 보호를 위한 조치로 피험자의 권리와 안전에 대한 고려, 불완전

물질을 직접 연구대상자에게 적용한 후 그로부터 얻은 정보를 이용하는 연구 4) 그 밖에 소음, 물리적 자극 등으로 연구대상자의 환경을 조작하여 얻은 정보를 이용하는 연구 등 실험적 연구.

291) 상호작용(interaction)을 통한 연구 즉, 의사사통, 대인접촉 등의 상호작용을 통하여 수행하는 연구란, "연구"를 위해 연구대상자를 선정하고 연구대상자의 대면을 통한 설문조사나 행동관찰 등으로 자료를 얻어 그 정보를 이용하는 연구로 다음과 같은 연구의 유형을 말한다. (1) 연구를 위해 연구대상자의 행동관찰 등을 수행하여 자료를 얻는 연구 (2) 연구를 위해 연구대상자를 대면하며 설문조사 등을 통해 자료를 얻는 연구 (3) 그 밖에 연구를 위해 연구대상자를 접촉하고 조사 및 관찰 등을 수행하는 연구.

292) 개인식별정보를 포함한 연구란, 위의 두 연구 유형처럼 연구대상자를 직접 대면하거나 연구대상자로부터 정보를 직접 수집하지는 않지만, 연구대상자를 직·간접적으로 식별할 수 있는 자료를 이용하는 연구를 말한다.

293) 복합적으로 수행되는 연구는 실험적 연구와 조사 등을 통한 정보를 이용하는 연구가 복합적으로 수행되거나, 관찰 연구가 포함된 연구 등 위의 연구유형들이 복합적으로 수행되는 연구로 인간대상연구에 해당한다.

능력자와 취약계층에 대한 보호, 개인정보보호 및 기록의 유지와 정보공개 원칙 등을 규정하고 있다.

② 기관위원회의 심의

인간대상 연구를 하려는 자나 인체 유래물 연구를 하려는 자는 연구를 하기 전에 연구계획서를 작성하여 기관위원회의 심의를 받아야 한다. 연구대상자 및 공공에 미치는 위험이 미미한 경우에 한하여 각 기관위원회가 심의면제를 확인해줄 수 있다(법 제15조 제2항, 시행규칙 제13조). 기관위원회의 심의를 면제할 수 있는 인간대상 연구는 일반대중에게 공개된 정보를 이용하는 연구이거나, 연구대상자에 대한 개인식별정보를 수집하거나 기록하지 않는 연구를 기본 대상으로 하는데, 그 외에도 추가 요건을 충족해야 한다.[294] 기관위원회의 심의를 면제할 수 있는 인간대상 연구는 연구자가 개인정보를 수집・기록하지 않은 연구이어야 하고[295], 그 외에 추가적인 요건을 충족해야 한다.[296]

---

294) (1) 연구대상자로 “취약한 환경의 피험자”를 포함하지 않는 연구로서 연구대상자를 직접 조작하거나 연구대상자의 환경을 조작하여 얻은 자료(data)를 이용하는 연구라 할지라도 다음 각각의 경우에는 심의를 면제한다.
① 약물투여나 혈액채취 등의 침습적 행위가 개입되지 않은 연구,
② 신체적 변화가 초래되지 않는 단순 접촉 측정장비 또는 관찰장비만을 사용하는 연구,
③ 「식품위생법」 시행규칙 제3조에 따라 판매 등이 허용되는 식품을 이용하여 맛 또는 질을 평가하는 연구, 또는
④ 「화장품법」 제8조 제1항 및 제2항에 따른 안전기준에 적합한 화장품을 이용하여 사용감 또는 만족도 등을 조사하는 연구
(2) 연구대상자로 “취약한 환경의 피험자”를 포함하지 않는 연구로서 의사소통이나 대인접촉 등의 상호작용 즉, 연구대상자 대면을 통한 설문조사나, 연구대상자의 행동관찰 등을 통해 얻은 자료(data)를 이용하는 연구라 할지라도, 그 연구대상자가 불특정하며, 연구대상자로부터 “민감정보”를 수집하거나 기록하지 않는 연구는 심의를 면제한다.
(3) (1)과 (2)에 해당하지 않더라도, 연구대상자를 직접 또는 간접적으로 식별할 수 있는 정보를 포함하고 있는 정보(information)를 이용하는 연구로서 이때 연구대상자 등에 관한 정보가 이미 생성된 기존의 자료나 문서를 이용하는 연구는 심의를 면제한다.
(″취약한 환경에 있는 시험대상자″(Vulnerable Subjects)란 임상시험 참여와 관련한 이익에 대한 기대 또는 참여를 거부하는 경우 조직위계상 상급자로부터 받게 될 불이익에 대한 우려가 자발적인 참여 결정에 영향을 줄 가능성이 있는 대상자(의과대학·한의과대학·약학대학·치과대학·간호대학의 학생, 의료기관·연구소의 근무자, 제약회사의 직원, 군인 등을 말한다), 불치병에 걸린 사람, 제27조에 따른 집단시설에 수용되어 있는 사람, 실업자, 빈곤자, 응급상황에 처한 환자, 소수 인종, 부랑인, 노숙자, 난민, 미성년자 및 자유의사에 따른 동의를 할 수 없는 대상자를 말한다.)

295) ① 인체유래물은행이 수집・보관하고 있는 인체유래물과 그로부터 얻은 유전정보(이하 “인체유래물등”이라 한다)를 제공받아 사용하는 연구로서 인체유래물등을 제공한 인체유래물은행을 통하지 않으면 개인정보를 확인할 수 없는 연구
② 의료기관에서 치료 및 진단을 목적으로 사용하고 남은 인체유래물등을 이용하여 정확도 검사 등 검사실 정도관리 및 검사법평가 등을 수행하는 연구
③ 인체유래물을 직접 채취하지 않는 경우로서 일반 대중이 이용할 수 있도록 인체유래물로부터 분리・가공된 연구재료(병원 체, 세포주 등을 포함한다)를 사용하는 연구
④ 연구자가 인체유래물 기증자의 개인식별정보를 알 수 없으며, 연구를 통해 얻어진 결과가 기증자 개인의 유전적 특징과 관계가 없는 연구. 다만, 배아줄기세포주를 이용한 연구는 제외한다.

296) 그 외에 ‘초・중등교육법’ 제2조 및 ‘고등교육법’ 제2조에 따른 학교와 보건복지부장관이 정하는

기관위원회 심의대상인 경우 연구자는 다음 사항이 포함된 연구계획서를 작성하여 신청서류와 함께 기관위원회에 심의를 신청한다.

1) 연구의 목적
2) 연구 배경(선행연구 포함)
3) 인체유래물기증자 선정, 예상 수, 산출 근거, 모집 및 동의 과정 등에 관한 사항
4) 연구 내용 및 방법, 조사 도구(해당하는 경우) 등
5) 연구로 인해 수집되는 자료 및 정보 등 관찰 항목에 관한 사항
6) 연구로 인한 연구대상자의 위험과 이익, 보상 등에 대한 사항
7) 인체유래물기증자 안전대책 및 개인정보보호대책에 관한 사항
8) 평가 기준 및 방법, 자료 분석 등 통계적 측면에 관한 사항
9) 연구 수행 장소 및 연구 참여기간, 연구자(연구책임자 및 공동연구자)에 관한 사항 등
10) 그 밖에 기관위원회 또는 공용위원회가 심의를 위해 요청하는 내용

따라서 건강정보를 활용하면서 인간대상 연구, 인체 유래물 연구에 해당하는 경우에는 생명윤리법이 추가로 적용되어 기관위원회의 연구 심의를 거쳐야 한다.

한편 연구자는 아래와 같은 기록물을 3년간 보관해야 한다. 보관 기간이 지난 문서 중 개인정보에 관한 사항은 파기하여야 한다.

1) 연구계획서 및 법 해당 연구를 심의한 기관위원회의 심의 결과(변경되었을 경우에는 변경된 연구계획서와 심의 결과를 포함한다)
2) 인체유래물기증자로부터 받은 서면동의서 또는 같은 조 제3항에 따른 기관위원회의 서면동의 면제 승인서
3) 개인정보의 수집·이용 및 제공 현황
4) 연구 결과물 등이 포함된 연구 종료 보고서 및 법 제10조 제3항 제2호에 따른 연구의 진행과정 및 결과에 대한 기관위원회의 조사·감독 결과

③ 연구대상자, 기증자로부터의 서면 동의

생명윤리법은 인간대상 연구, 인체 유래물 연구를 하기 전에 연구대상자, 기증자로부터 서면 동의를 받도록 하고 있다.

교육기관에서 통상적인 교육과정의 범위에서 실무와 관련하여 수행하는 연구, 공중보건상 긴급한 조치가 필요한 상황에서 국가 또는 지방자치단체가 직접 수행하거나 위탁한 연구 다만, 이 경우 해당 기관의 장은 보건복지부장관이 지정하여 고시하는 공용위원회에 연구 종료 전 연구의 진행 상황을 통보하여야 한다.

표 4-39 생명윤리법 연구 서면동의 항목

인간대상 연구	인체유래물 연구
1. 인간대상연구의 목적 2. 연구대상자의 참여 기간, 절차 및 방법 3. 연구대상자에게 예상되는 위험 및 이득 4. 개인정보 보호에 관한 사항 5. 연구 참여에 따른 손실에 대한 보상 6. 개인정보 제공에 관한 사항 7. 동의의 철회에 관한 사항 8. 그 밖에 기관위원회가 필요하다고 인정하는 사항	1. 인체유래물연구의 목적 2. 개인정보의 보호 및 처리에 관한 사항 3. 인체유래물의 보존 및 폐기 등에 관한 사항 4. 인체유래물과 그로부터 얻은 유전정보(이하 “인체유래물등“이라 한다)의 제공에 관한 사항 5. 동의의 철회, 동의 철회 시 인체유래물등의 처리, 인체유래물 기증자의 권리, 연구 목적의 변경, 그 밖에 보건복지부령으로 정하는 사항

그런데 (i) 연구대상자의 동의를 받는 것이 연구 진행과정에서 현실적으로 불가능하거나 연구의 타당성에 심각한 영향을 미친다고 판단되는 경우로서, (ii) 연구대상자의 동의 거부를 추정할 만한 사유가 없고, 동의를 면제하여도 연구대상자에게 미치는 위험이 극히 낮은 경우에는 기관위원회의 승인을 받아 연구대상자의 서면 동의를 면제할 수 있다.

④ 연구대상자에 대한 안전조치

인간대상 연구자는 사전에 연구 및 연구 환경이 연구대상자에게 미칠 신체적·정신적 영향을 평가하고 안전대책을 마련해야 한다. 그리고 수행 중인 연구가 개인 및 사회에 중대한 해악(害惡)을 초래할 가능성이 있을 때는 이를 즉시 소속 기관의 장에게 보고하고 적절한 조치를 하여야 하고, 질병의 진단이나 치료, 예방과 관련된 연구에서 연구대상자에게 의학적으로 필요한 치료를 지연하거나 진단 및 예방의 기회를 박탈하여서는 안 된다(제17조).

⑤ 제3자 제공에 대한 규정

인체 유래물 연구자는 인체 유래물 기증자로부터 인체 유래물등을 제공하는 것에 대하여 서면 동의를 받은 경우에만 기관위원회의 심의를 거쳐 인체 유래물 등을 인체유래물 은행이나 다른 연구자에게 제공할 수 있다. 이 경우 인체 유래물 기증자가 개인식별정보를 포함하는 것에 동의한 경우가 아니면 인체 유래물등을 다른 연구자에게 제공하는 경우에는 익명화하여야 한다(제38조). 인간대상 연구자도 연구대상자가 개인정보를 제공하는 것에 대하여 서면 동의를 한 경우에만 기관위원회의 심의를 거쳐 개인정보를 제3자에게 제공할 수 있다. 이때 연구대상자가 개인식별정보를 포함하는 것에 동의하지 않았다면 익명화를 해서 제공해야 한다(제18조). 연구자는 이에 대한 기록을 작성, 보관해야 한다.

그림 4-31 인체 유래물 연구 동의서

■ 생명윤리 및 안전에 관한 법률 시행규칙 [별지 제34호서식]

인체유래물 연구 동의서

(앞쪽)

동의서 관리번호	

인체유래물 기증자	성 명	생년월일
	주 소	
	전화번호	성별
법정대리인	성 명	관계
	전화번호	
연구책임자	성 명	
	전화번호	

이 동의서는 귀하로부터 수집된 인체유래물등(인체유래물과 그로부터 얻은 유전정보를 말합니다)을 질병의 진단 및 치료법 개발 등의 연구에 활용하기 위한 것입니다. 동의는 자발적으로 이루어지므로 아래의 내용을 읽고 궁금한 사항은 상담자에게 묻고 질문할 기회를 가지고 충분히 생각한 후 결정하시기 바라며, 이 동의서에 대한 동의 여부는 귀하의 향후 검사 및 치료 등에 어떤 영향도 미치지 않습니다.

1. 인체유래물이란 인체로부터 수집하거나 채취한 조직·세포·혈액·체액 등 인체 구성물 또는 이들로부터 분리된 혈청, 혈장, 염색체, DNA, RNA, 단백질 등을 말하며, 귀하의 인체유래물을 채취하기 전에 채취 방법 및 과정에 관한 설명을 충분히 들어야 합니다.
2. 귀하가 귀하의 인체유래물등을 아래의 연구 목적에 이용하도록 동의하는 경우, 귀하의 인체유래물등의 보존기간, 다른 사람 또는 다른 연구 목적에 대한 제공 여부, 제공 시 개인정보 처리에 관한 사항 및 폐기 등을 결정할 수 있습니다. 또한 동의한 사항에 대해 언제든지 동의를 철회할 수 있습니다. 이 경우 연구의 특성에 따라 철회 전까지 수집된 귀하의 인체유래물등과 기록 및 정보 등의 처리방법이 달라질 수 있으므로 연구자로부터 별도의 설명문 등을 통해 정보를 받으실 것입니다.
3. 귀하는 이 연구 참여와 관련하여 귀하의 동의서 및 귀하의 인체유래물등의 제공 및 폐기 등에 관한 기록을 본인 또는 법정대리인을 통하여 언제든지 열람할 수 있습니다.
4. 귀하가 결정한 보존기간이 지난 인체유래물은 「폐기물관리법」 제13조에 따른 기준 및 방법에 따라 폐기되며, 해당 기관의 휴업·폐업 등 해당 연구가 비정상적으로 종료될 때에는 법에서 정한 절차에 따라 인체유래물등을 이관할 것입니다.
5. 귀하의 인체유래물등을 이용하는 연구는 「생명윤리 및 안전에 관한 법률」에 따라 해당 기관의 기관생명윤리위원회의 승인 후 진행될 것이며 해당 기관 및 연구자는 귀하의 개인정보 보호를 위하여 필요한 조치를 취할 것입니다.
6. 귀하의 인체유래물등을 이용한 연구결과에 따른 새로운 약품이나 진단도구 등 상품개발 및 특허출원 등에 대해서는 귀하의 권리를 주장할 수 없으며, 귀하가 제공한 인체유래물등을 이용한 연구는 학회와 학술지에 연구자의 이름으로 발표되고 귀하의 개인정보는 드러나지 않을 것입니다.

※ 위의 모든 사항에 대해 충분한 설명을 듣고, 작성된 동의서 사본을 1부 받아야 합니다.

동의 내용	연구 목적	
	인체유래물 종류 및 수량	
	인체유래물 보존기간	1. 영구보존 [ ] 2. 동의 후 [ ] 년
	보존 기간 내 2차적 사용을 위한 제공 여부	1. 유사한 연구 범위 안에서만 제공하는 것에 동의합니다. [ ] 2. 포괄적 연구 목적으로 제공하는 것에 동의합니다 [ ] 3. 동의하지 않습니다. [ ]
	2차적 사용을 위한 제공 시 개인식별정보 포함 여부	1. 개인식별정보 포함 [ ] 2. 개인식별정보 불포함 [ ]

210mm×297mm[백상지 80g/㎡(재활용품)]

⑥ 정보공개에 관한 권리

생명윤리법은 인체 유래물 기증자등의 정보공개에 관한 권리도 보장하고 있다. 그래서 인체 유래물 기증자등(법정대리인 포함)은 자신에 관한 정보의 공개를 청구할 수 있고, 그 경우 연구자는 특별한 사유가 없으면 공개 청구받은 정보를 공개하여야 한다. 정보공개절차는 기관위원회를 통해서 이루어진다.

나. 기관위원회의 구성과 운영

생명윤리법은 기관위원회의 요건에 대해서도 규정하고 있다. 기관위원회는 기관의 장이 위원을 위촉하는데, 위원장 1명을 포함하여 5명 이상의 위원으로 구성하되, 하나의 성(性)으로만 구성할 수 없으며, 사회적·윤리적 타당성을 평가할 수 있는 경험과 지식을 갖춘 사람 1명 이상과 그 기관에 종사하지 아니하는 사람 1명 이상이 포함되도록 했다. 기관위원회의 독립성 보장을 위해서 생명윤리법은 기관의 장은 기관위원회가 독립성을 유지할 수 있도록 행정적·재정적 지원을 하여야 한다는 규정을 두고 있다(법 제11조 제5항). 그리고 기관위원회를 설치한 기관의 장은 기관위원회의 업무를 수행하기 위하여 기관위원회 표준운영지침을 마련하여야 한다(시행규칙 제8조 제4항). 아울러 기관의 장은 해당 기관에서 수행하는 연구 등에서 생명윤리 또는 안전에 중대한 위해가 발생하거나 발생할 우려가 있는 경우에 지체 없이 기관위원회를 소집하여야 한다고 규정하고, 기관의 장이 그 결과를 보건복지부장관에게 보고하도록 하여 기관위원회를 생명윤리나 안전에 관한 판단주체로 위상을 보장하고 있다.

기관위원회는 연구계획서 심의 외에도 연구 진행과정과 결과에 대한 조사, 감독, 해당 기관의 연구자 및 종사자 교육이나 윤리지침 마련 등의 활동을 수행한다(제10조 제3항).

1. 다음 각 목에 해당하는 사항의 심의
가. 연구계획서의 윤리적·과학적 타당성
나. 연구대상자등으로부터 적법한 절차에 따라 동의를 받았는지 여부
다. 연구대상자등의 안전에 관한 사항
라. 연구대상자등의 개인정보 보호 대책
마. 그 밖에 기관에서의 생명윤리 및 안전에 관한 사항
2. 해당 기관에서 수행 중인 연구의 진행과정 및 결과에 대한 조사·감독
3. 그 밖에 생명윤리 및 안전을 위한 다음 각 목의 활동
가. 해당 기관의 연구자 및 종사자 교육
나. 취약한 연구대상자등의 보호 대책 수립
다. 연구자를 위한 윤리지침 마련

기관위원회는 보건복지부장관에게 등록을 해야 하고, 보건복지부장관은 기관위원회의 운영실태 등에 대한 조사를 할 수 있는데, 3년에 한 번씩 실태조사를 한다. 기관위원회는 평가와 인증을 받을 수 있게 되어 있다. 한편 보건복지부는 보건복지부 지정 공용기관생명윤리위원회 표준운영지침서를 작성하여 공개하고 있다.

이상과 같은 국내 보건의료 분야 데이터 연계 현황을 검토한 결과, 보건의료 분야의 데이터 거버넌스에 대한 종합적인 정책 제안이 필요해 보인다. 이는 제5장에서 다룰 예정이다.

제3절 국내 통계 분야 데이터의 연계 현황

1. 개요

통계는 합리적인 국가 정책 결정을 위한 매우 중요하고, 필수불가결한 수단이다. 그런데 통계에는 불가피하게 민감한 개인의 정보를 수집하는 경우가 많다. 통계 분야에서도 데이터 연계, 결합의 요구는 늘어나고 있다. 통계청은 빅데이터 활용을 주요한 과제로 선정하고 있다.

그런데 통계는 통계의 작성 단계에서 다른 행정정보를 활용하여 연계하거나, 다른 통계자료를 연계하는 데이터 연계나 결합의 수요가 있다. 통계 작성과정에서 취득한 통계자료(마이크로데이터)의 활용을 극대화하기 위한 여러 가지 노력이 이루어지고 있는데, 그중 다른 자료와 연계를 해서 가용성을 극대화하는 것이 있다.

통계청은 최근 빅데이터를 활용한 통계 작성을 위해 통계데이터허브국에 정원 12명의 '빅데이터통계과'를 신설하고(2015. 10.) 빅데이터 통계기획, 생산, 시스템구축을 담당하도록 하고 있다. 공공데이터와 민간 빅데이터 간 연계로 새로운 가치 창출, 사회가 필요로 하는 다양한 통계적 정보 작성 및 제공, 연계와 활용을 강화하겠다고 입장을 밝히고 있다. 마이크로데이터 이용센터를 운영하고 이를 통해서 통계의 연구목적 활용을 꾀하고 있다. 건보공단 자료와의 연계를 하기도 한다. 그동안 공공데이터를 연계·활용한 통계 작성 경험을 바탕으로 공공데이터와 민간데이터 간 연계로 새로운 통계정보 작성을 추진하고 있기도 하다.

통계와 관련한 법제로도 개인정보보호법과 통계법이 있고 각 부처의 훈령으로 통계관리규정을 두고 있지만 다양한 데이터 연계와 통합, 특히 민간분야와의 연계를 시도하는 상황에서 그에 대한 규율은 매우 미흡하다.

이하에서는 통계와 관련한 법제도, 통계작성과정의 데이터 연계, 통계자료의 활용방법으로 데이터 연계, 통계청의 마이크로데이터 이용 센터를 분석하고 바람직한 대안을 제시하고자 한다.

2. 통계 제도와 기본원칙

(1) 통계와 통계 제도

통계란 '집단에 대해 집단의 현상을 체계적인 숫자로 표현한 것'[297] 또는 '시간, 공

간 및 속성이 규정된 집단에 대하여 집단의 현상을 체계적인 숫자에 의하여 표현한 것'[298]을 말한다. 통계는 집단에 관한 어떠한 정보를 전달하는 숫자로서 집단을 구성하는 개체를 특정할 수 있는 고유한 정보는 제거되어 있다. 통계수치에는 개체를 식별하는 정보가 필요하지 않으며 어떤 개체가 어떤 값을 가지고 있느냐 하는 것은 무의미하며, 이런 통계의 특성을 익명성이라고 표현하기도 한다.[299] 통계를 넓게 본다면 민간 기업에서 실시하는 통계까지 볼 수도 있지만, 본 연구에서는 통계법 적용대상이 되는 통계를 주로 검토한다.

우리나라 통계법은 통계를 통계작성기관이 산업·물가·인구·주택·문화·환경 등 특정의 집단이나 대상 등에 관하여 작성하는 수량적 정보로서 정부정책의 수립·평가, 경제·사회현상의 연구·분석 등에 활용할 목적으로 작성하는 것을 말한다고 정의하고 있다. 우리나라의 경우 직접 또는 위임이나 위탁으로 통계를 작성하는데 각 중앙부처를 비롯한 중앙행정기관, 지방자치단체, 민간기관 등 411개 기관이 통계작성기관으로 통계청의 지정을 받았고 391개 기관이 통계를 작성하였다고 한다(2017년 9월 현재, 통계청 검색결과).[300]

통계를 운용하는 제도는 집중형과 분산형으로 나누고 있는데 우리나라는 분산형으로 각 기관에서 통계를 작성하며 통계청은 조정기관의 역할을 하고 있다.

표 4-40 집중형과 분산형 통계 제도의 장·단점 비교

유형	집중형	분산형
주요 특징	국가기본통계를 단일화된 통계전문기관에서 작성 부처간 통계연락기구의 설치	부처별로 필요한 통계를 개별적 작성 통계조정기관 설치
해당 국가	캐나다, 스웨덴, 핀란드, 네덜란드, 호주, 인도네시아 등	한국, 미국, 일본, 영국, 대만 등
장점	통계의 균형적 개발과 유기적 체계 확보 통계의 객관성 및 신뢰도 제고 통계전문인력의 집중적 활용	분야별 전문지식을 관련 통계 개발에 활용 가능 통계수요에 신속히 대응
단점	관련행정분야별 전문지식 활용 미흡 통계수요에 대한 신속한 대응 어려움	중복작성 등으로 인력과 예산의 낭비 초래 체계적 통계 개발의 제약

* 통계교육원(2015), p77; 통계행정편람(2016, 통계청), p18 종합.

297) 통계교육원(2015), "국가통계의 이해", p9.

298) 통계행정편람(2011, 통계청), p11.

299) 통계행정편람(2016, 통계청), p12.

300) 2016년 통계에 의하면 작성기관의 통계인력(통계업무를 전담하거나 주로 통계업무에 종사하는 인력)은 총 4,802명이라고 한다.
http://narastat.kr/pms/pub/scs/aim/selectInsttManageInfoList.do 참조.

통계는 다양한 분야에서 작성, 활용되고 있다. 정부와 중앙행정기관, 지자체가 아닌 지정기관의 통계도 182종이나 되고 그중에는 금융업, 공단 등에서도 많은 통계를 작성하고 있음을 알 수 있다. 분야별로는 보건·사회복지 통계가 261종으로 가장 많고 경기·기업경영과 관련된 통계가 89종으로 그다음이다.

표 4-41 부문별 통계작성 현황

구분	작성기관 수	작성통계수	종류별		작성방법별		
			지정	일반	조사	보고	가공
계	411	1052	93	959	473	468	111
정부기관	305	870	76	794	369	412	89
중앙행정기관	45	378	59	319	189	150	39
통계청	1	60	38	22	37	2	21
이외기관	44	318	21	297	152	148	18
지방자치단체	260	492	17	475	180	262	50
지정기관	106	182	17	165	104	56	22
금융기관	8	23	10	13	10	5	8
공사/공단	31	53	-	53	21	30	2
- 연구기관	21	40	2	38	32	4	4
- 협회/조합	25	31	3	28	24	5	2
- 기타기관	21	35	2	33	17	12	6

(2) 통계 작성 방법에 따른 분류

통계는 그 작성방법에 따라서 조사통계와 행정통계(보고통계)로 나눌 수 있다. 조사통계는 통계의 작성을 주목적으로 조사를 실시하여 얻어진 통계를 말하며[301] 제1의 통계라고도 한다. 조사통계는 조사대상 집단의 모든 단위를 조사하는 전수조사와 집단의 모든 구성단위를 전부 조사하는 대신 일부만을 조사하고서도 전부 조사하는 것과 같은 자료를 얻는 표본조사로 나누어진다.[302] 행정통계는 보고 또는 가공 방식으로 통계를 작성한다. 그중 보고통계는 법령에 의한 개인 및 단체의 신고, 보고, 신청, 인허가 등과 같이 다른 행정업무에 수반하여 수집된 자료로부터 통계를 작성한 것을 말하며[303] 제2의 통계라고도 한다. '가공통계'란 한 종류 이상의 통계와 추가로 수집한 통계자료 또는 행정자료를 이용하여 작성한 통계를 말한다.[304]

301) 통계승인업무 처리지침 제2조 제1호 "조사통계"란 통계 작성을 목적으로 통계작성기법을 사용하여 조사한 자료를 통해 작성한 통계를 말한다.

302) 통계행정편람(2016, 통계청), p15.

303) 통계승인업무 처리지침 제2조 제2호 "보고통계"란 신고, 보고, 신청, 인허가 등과 같이 다른 행정업무에 수반하여 수집된 자료를 이용하여 작성한 통계를 말한다.

304) 통계승인업무 처리지침 제2조 제3호.

표 4-42 통계작성방법에 따른 통계의 분류

구분	조사통계		행정통계
	전수조사(Census)	표본조사(Survey)	
포괄범위	모집단 전체	모집단 대표 표본	모집단 전체 (포괄범위 차이 발생 가능)
내용	광범위한 자료	심층 분석에 적합	행정 목적에 따라 제한적
품질관리	비표본오차	표본오차, 비표본오차	시스템 차원에서 접근
비용	고비용	상대적 저렴	저비용
주기	장기(5년, 10년 등)	월간, 분기 또는 연간	행정 프로그램에 좌우 (대부분 연간)
시의성	조사 종료 1년 후 사용	정기조사: 몇 주 후 산출 임시조사: 몇 달 내 산출	자료입수 후 3개월 전후 결과 공표
안정성	이용자 요구에 따라 통계청이 관리	시계열 확보로 인하여 조사 변동 가능성 낮음	법 · 제도, 정책 변화 시 변동
응답부담	전 국민, 전체 가구	낮음	없음

(3) 통계의 기본원칙과 통계의 비밀보호

가. 통계의 기본원칙 중 하나인 비밀보호의 원칙

통계기관이 수집한 정보에 대한 비밀보호는 통계법에 특유한 원칙으로 전 세계 각국에서 인정되고 있는 원칙이다.

예를 들어 UN은 공공통계에 대한 기본원칙(Fundamental Principles of Official Statistics)을 UN총회에서 결의하였는데, 그 중 '통계기관이 수집한 개인이나 법인에 대한 개별적인 자료를 엄격하게 비밀로 보호해야 하고, 통계 목적으로만 사용해야 한다'는 비밀보호 규정을 두고 있다. UN 기본원칙은 개인정보보다도 폭넓은 규정으로 '개인과 법인에 대한 개별적인 자료'라는 개념을 사용하고 있다.

> 제6조 통계기관에 의해 수집된 개인이나 법인에 대한 개별적인 자료는 비밀이 엄격하게 보호되어야 하며, 통계 목적으로만 사용되어야 한다.

유럽연합의 통계규정도 마찬가지이다. 통계의 원칙으로 통계적 기밀유지를 들고 있다. 여기서도 개인정보보다 더 폭넓게 '개인과 가정, 경제주체 등의 통계단위(statistical unit)'가 식별되거나 식별될 가능성이 있는 정보를 '기밀정보(confidential data)'로 정의하고 그에 대한 비밀을 엄격하게 보호하고 있다.

통계단위의 기밀정보는 통계 목적 외의 목적으로 사용되어서는 안 된다. '통계적 기밀성(statistical confidentiality)'은 통계적 목적으로 직접 획득하거나 행정 혹은 다른 소스로부터 간접적으로 획득한 단일한 통계적 단위와 관련된 기밀 데이터의 보호를 의미하며 비통계적 목적으로 획득한 데이터를 이용하거나 불법적인 공개를 금지함을 뜻한다.

유럽연합 뿐만 아니라 유럽 외의 국가들에서도 통계와 관련한 비밀보호 규정을 두고 있다. 예를 들어 오스트레일리아는 통계의 비밀보호에 대한 규율을 두고 있다.

우리 통계법도 기본이념으로 공공성, 정확성·시의성·일관성, 중립성과 함께 비밀보호를 들고 있기는 한데, 아래에서 살펴볼 것처럼 그 비밀보호는 외국의 입법례와는 다르게 모든 통계의 비밀을 보호해 주는 것이 아니라 '비밀'정보인 통계의 비밀만 보호해 주는 것으로 범위가 아주 좁게 규정하고 있어서 통계 비밀보호로 보기 어렵다.

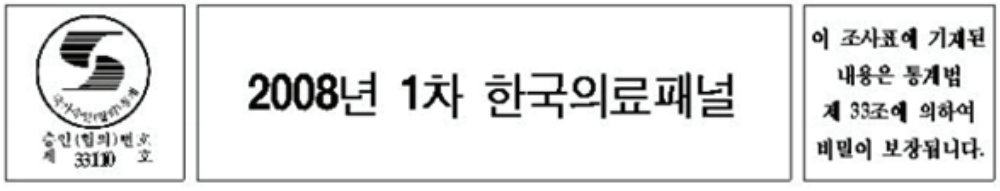
나. 통계의 비밀보호가 필요한 이유

통계기관이 수집한 정보에 대한 비밀보호가 필요한 이유로는 다음을 들 수 있다.

첫째, 통계의 기본원칙인 공정성과 신뢰성, 통계의 질을 유지하기 위해서는[305] 통계제공자의 신뢰가 필요하고, 통계제공자의 협력과 선의에 의존해야 하는데, 이와 같은 통계의 신뢰성과 품질을 유지하려면 통계에 대한 비밀을 보호해 주어야 하기 때문이다.[306]

예를 들어 공공기관 운영에 관한 법률에 따라 실시하는 공공기관 고객만족도 조사의 경우는 서비스를 직접 제공받는 고객을 대상으로 매년 실시하는 조사인데, 만족도 조사의 결과가 비밀로 유지된다는 전제가 없을 경우 조사의 신뢰성이 보장되지 않을 것이다. 실제로 통계조사지에는 통계법 제33조에 의해 비밀이 보장된다는 내용을 기재하고 있다.

그림 4-32 제1차 한국의료패널의 안내문 1페이지



305) 예를 들어 유럽연합은 '전문적인 독립성(professional independence)', '공정성(impartiality)', '객관성(objectivity)', '신뢰성(reliability)', '통계 기밀성(statistical confidentiality)'과 '비용 효과성(cost effectiveness)을 기본원리로 삼고 있다.

306) http://www.abs.gov.au/websitedbs/a3121120.nsf/home/statistical+language+-+confidentiality

그러나 우리나라의 경우 표시된 내용과는 달리 비밀이 보호되지 않을 수 있다. 오히려 이 표시는 응답자를 오도하는 내용이 될 수도 있는 상황이다.

또 다른 하나는 통계가 개인이나 법인에 관한 민감한 정보의 수집, 가공, 처리를 동반하고 있기 때문에[307] 해당 주체의 개인정보나 영업비밀 등을 보호하기 위함이다.

다. 통계의 비밀보호의 구체적 내용

각국에서 통계의 비밀보호를 보장하기 위해서 입법 등으로 규정하고 있는데, 예를 들어 유럽연합의 경우 유럽연합 통계규칙에서 규정하고 있는 통계의 비밀보호에 대한 내용은 다음과 같다.

첫째, 통계의 공개 시 통계단위의 주체를 식별할 수 있도록 공개해서는 안 된다는 것이다. 통계기관은 기밀성 보호를 위해 필요한 모든 물리적 및 논리적 보호(통계적 공개통제, Statistical disclosure control, SDC)를 보장하기 위해 필요한 모든 규제적, 행정적, 기술적 및 조직적 조치를 취하여야 한다(Commission Regulation (EC) No 223/2009, 제20조). 이와 같이 유럽연합의 경우는 통계적 공개통제를 통해서 공개된 통계정보에서 통계단위가 식별되거나 식별될 가능성이 있도록 해서는 안 된다는 것이 법적 의무이다.

비밀보호의 예외는 두 가지가 있는데, 하나는 별도의 법률[308]에서 명시하는 경우로서 소위 '수동적 비밀보호'(passive confidentiality)라고 불리는 것이다.[309] 별도의 법률에서 명시하는 경우로서 통계주체의 비밀보호 요청이 있지 않으면 통계주체가 식별되는 방식으로 노출되는 것을 허용한다(Regulation (EC) No 223/2009 20조 제1항). 이는 주로 무역 통계에서 취하는 입법이라고 한다. 물론 나머지 영역에서는 '능동적 비밀보호'(active confidentiality)라고 하여 통계주체의 요구가 있지 않더라도 노출되지 않도록 하여야 한다.[310]

또 다른 예외는 '면제 방식'(waiver approach)이다. 통계주체가 공개하는 것을 허용

---

307) 수집되는 개인정보의 항목은 매우 민감한 것들이 많다. 특히 보고통계의 경우는 당사자가 모르는 상태에서 개인정보가 수집되는 경우가 많다. 예를 들어 보고통계의 하나인 범죄통계의 경우 수집되는 항목은 고령범죄자, 공무원범죄자, 공범관계, 공범수, 구속·불구속, 금전소비용도, 소년범죄자, 발생비, 범행 후 은신처, 범행도구, 보호처분, 생활정도, 외국인범죄자 국적, 재범, 재범기간, 전과, 정신장애범죄자, 처분결과, 피해자 성별 및 연령, 피해자와의 관계 등의 정보를 수집한다.

308) 예를 들어 유럽 비회원국과의 무역에 관한 통계에 대해 규정하는 Regulation (EC) No 471/2009 제10조.

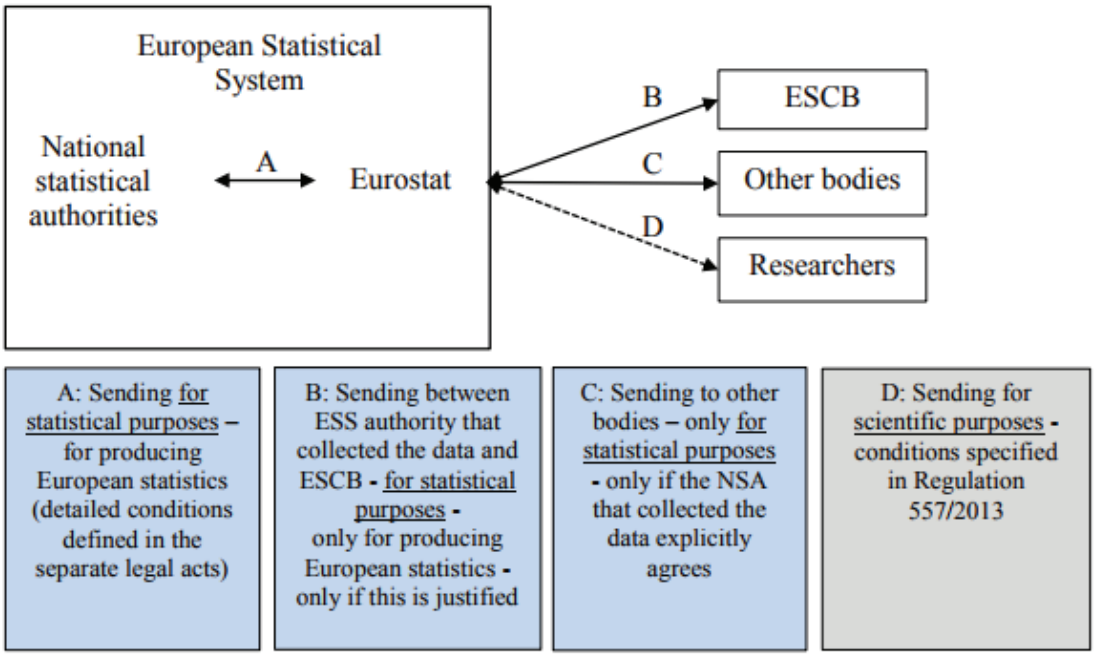
309) European Business Statistics (EBS) Manual "statistical disclosure control"(European Commission), p.4

310) 앞의 자료 p.4

하여 비밀보호 의무에 대해 면제를 해 주면 통계주체를 식별할 가능성이 있는 정보의 공개가 허용된다(제20조 제2항).[311]

둘째, 통계자료는 통계의 목적으로만 활용되어야 하고, 통계단위 주체를 식별할 수 있는 통계자료를 제3자에게 제공하거나 공개해서는 안 된다는 것이다. 제3자에게 통계단위 주체가 식별될 수 있는 정보가 제공되는 것은 통계 목적으로 통계기관에 제공되는 것만이 예외로 한다. 연구자나 연구기관에 대해 과학연구 목적으로 제공되는 것도 통계 목적으로 본다.

그림 4-33 유럽통계시스템에서 기밀정보의 전송



* 출처: European Business Statistics (EBS) Manual "statistical disclosure control"(European Commission), 6 페이지

단, 통계의 비밀보호와 관련해서 공중에게 적법하게 공개된 자료로부터 수집되고 법률에 의해서 공중에게 계속 공개되고 있는 데이터는 통계의 배포와 관련해서 비밀유지의 대상에서 제외된다(제25조). 한편, 여기서 보호대상이 되는 정보는 통계단위가 식별되는 정보이다. 통계단위란 개인뿐만 아니라 가구나 기업, 단체 등도 포함된다.

셋째, 기밀정보의 제공 과정에서 비밀보호에 대한 규율도 두고 있다. 예를 들어 유럽연합 통계규칙은 '정보를 수집한 ESS기관으로부터 다른 ESS기관으로의 기밀정보의 전송은 이러한 전송이 유럽통계의 효율적인 개발, 작성 및 보급을 위하거나 유럽

311) 앞의 자료 p.4

통계의 품질 향상을 위하여 필요하다면 이루어질 수 있다. 정보를 수집한 ESS기관과 ESCB 회원 간에 기밀정보를 전송하는 것은 이 전송이 ESS의 각각의 관할범위 내에서 유럽통계의 효율적인 개발, 생산 및 보급 또는 유럽통계의 품질 향상을 위해 필요하다면, 그리고 이러한 필요성이 입증되었다면 가능하다. 첫 번째 전송을 넘어서는 추가 전송에는 그 정보를 수집한 기관의 명시적인 승인이 요구되어야 한다'고 비밀보호를 위한 조건을 제시하고 있다. 아울러 '통계 기밀성에 관한 국가법령은 유럽의회와 이사회의 법령이 그러한 정보의 전송을 규정한 경우 기밀정보의 전송을 방지하기 위하여 발동되어서는 안 되고 전송된 기밀정보는 오로지 통계 목적으로만 사용되어야 하며 특정 업무영역 내에서 통계작업을 수행하는 직원만이 액세스할 수 있다는 점, 유럽연합 통계규칙에 규정된 통계적 기밀성 조항은 ESS내에서 그리고 ESS와 ESCB 간에서 전송된 모든 기밀정보에 적용되어야 한다는 점 등도 규정하고 있다.

넷째, 통계의 기밀보호는 통계기관의 기밀정보 보호에 대한 규정도 마련하고 있다. 예를 들어 유럽연합의 통계규칙은 '기밀정보는 본 규칙의 규정에 따라서 특정 업무영역 내에서 위원회(Eurostat)의 간부들에게만 접근 가능해야 하며, 위원회(Eurostat)는 예외적인 경우에 기밀정보에 대한 액세스 권한을 다른 직원들 및 그들의 특정 업무영역 내에서 계약에 따라서 위원회(Eurostat)를 위하여 일하는 다른 자연인에게 부여할 수 있다. 기밀정보에 접근할 수 있는 사람은 오로지 통계 목적으로만 이 정보를 사용해야 하며 직무 종료 이후에도 이러한 제한을 받아야 한다'고 규정하고 있다.

다섯째, 학술 목적으로 하는 기밀정보에의 접근에 대한 규율이다. 각국은 통계의 기밀보호가 학술 목적을 위해 통계분석을 수행하는 연구자들에게 제한적으로 허용될 수 있음을 규정하고 있다. 예를 들어 유럽연합의 통계규칙은 '통계단위의 간접 식별만을 허용하는 기밀정보에 대한 접근은 위원회(Eurostat) 또는 NSI나 기타 국가기관에 의하여 그들 각각의 관할영역 내에서 학술 목적을 위해 통계분석을 수행하는 연구자들에게 부여될 수 있다. 정보가 위원회(Eurostat)에 전송된 경우, 그 정보를 제공한 NSI 또는 기타 국가기관의 승인이 필요하다. EU 차원에서의 접근방식, 규칙 및 조건은 유럽위원회에 의해 수립되어야 한다'는 규정을 두고 있다.

라. 우리나라 통계법의 통계 비밀보호에 관한 규율

우리나라 통계법은 제2조와 제33조의 단 두 개의 조문에서 통계의 비밀보호에 대해 규율하고 있다.

통계법 제2조(기본이념)는 '통계는 개인이나 법인 또는 단체 등의 비밀이 보장되는 범위 안에서 널리 보급·이용되어야 한다'고 규정하고 있다. 이 규정은 1999년 통계법

개정(법률 제5691호) 시 통계의 기본이념이 최초 도입된 후[312] 2007년 개정(법률 제8387호) 시에 비밀보호의 내용이 추가로 포함된 것이다.

> 제2조(기본이념) ③통계는 개인이나 법인 또는 단체 등의 비밀이 보장되는 범위 안에서 널리 보급·이용되어야 한다.
> 제33조 (비밀의 보호) ①통계의 작성과정에서 알려진 사항으로서 개인이나 법인 또는 단체 등의 비밀에 속하는 사항은 보호되어야 한다.
> ②통계의 작성을 위하여 수집된 개인이나 법인 또는 단체 등의 비밀에 속하는 자료는 통계 작성 외의 목적으로 사용되어서는 아니 된다.

그런데 이 규정은 '비밀'에 해당하는 자료에 대해서만 보호가 이루어지는 것으로 되어 있어서 이것만으로는 통계단위의 신뢰를 보호할 수 없다.

실제로 해외의 입법들은 통계의 작성과정에서 수집한 개인에 관한 정보는 그것이 비밀에 해당하는지 여부를 불문하고 이를 비밀로 보호하고 있는데, 반면 우리의 통계법은 통계를 통해서 비밀침해를 하는 것만을 금지하고 있다. 즉, 통계를 통해서 개인의 비밀이 침해되지만 않으면 통계를 통해서 수집된 개인의 정보가 공개되어도 문제가 없다고 규정하고 있다. 통계 작성 과정에서 수집된 정보는 그 내용이 비밀인지 여부를 떠나서 통계작성기관이 취득한 해당 개인에 대한 모든 정보(단, 공공에게 접근하도록 공개한 정보는 제외)를 비밀로 보호해야 한다. 이 점을 분명하게 규정하는 방식으로 법률 개선이 필요하다.

> 통계의 비밀보호는 '통계기관에 의해 수집된 개인이나 법인에 대한 개별적인 자료는 비밀이 엄격하게 보호되어야 하며, 통계 목적으로만 사용되어야 한다'는 것으로 분명하게 규율한다.

나아가 현행 통계법은 통계의 비밀보호의 각 영역별 내용을 규율하고 있지 않다. 이를 구체적으로 영역별로 살펴보면 다음과 같다.

① 통계의 공표 시 비밀보호와 관련하여

무엇보다도 현행 통계법은 가장 기본적인 통계의 공표와 관련한 비밀보호 원칙을 규정하고 있지 않다. 즉, 통계를 공표할 때 통계주체를 식별할 수 있게 공개해서는 안 된다는 규율은 두지 않고 있다. 현행 통계법은 지체 없이 공표해야 한다는 규정만을 두고 있을 뿐이다.[313]

---

312) 이때는 '통계는 각종 의사결정을 합리적으로 수행하기 위한 자원으로서 사회발전에 기여할 수 있도록 과학적인 방법에 의하여 생산되고 공정하게 이용되어야 한다.'는 것이었다.

313) 그 외 통계의 공표에 대한 법률 규정이 있기는 한데, 이는 통계의 비밀보호에 관한 규정은

제27조 (통계의 공표) ①통계작성기관의 장은 통계를 작성한 때에는 그 결과를 지체 없이 공표하여야 한다.
②통계작성기관의 장은 제1항에 따라 통계를 공표하는 때에는 통계이용자가 통계를 정확하게 이용할 수 있도록 조사의 대상·방법 등 필요한 사항을 함께 공표하여야 한다.
③제1항에도 불구하고 통계작성기관의 장은 작성한 통계가 다음 각 호의 어느 하나에 해당하는 경우에는 통계를 공표하지 아니할 수 있다. 이 경우 미리 통계청장의 승인을 받아야 한다.
1. 공표할 경우 국가안전보장·질서유지 또는 공공복리에 현저한 지장을 초래할 것으로 인정되는 경우
2. 통계의 신뢰성이 낮아 그 이용에 혼란이 초래될 것으로 인정되는 경우
3. 그 밖에 통계를 공표하지 아니할 필요가 있다고 인정되는 상당한 이유가 있는 경우
④통계작성기관의 장은 제3항에 따라 공표하지 아니한 통계로서 그 사유가 소멸되었다고 인정되는 때에는 이를 공표하여야 한다. 이 경우 미리 통계청장과 협의하여야 한다.
⑤통계작성기관의 장은 제1항 또는 제4항에 따라 통계를 공표한 때에는 지체 없이 그 결과를 통계청장에게 제출하여야 한다.

그 결과 현재 통계 공표 시 각 통계기관의 자의적 판단에 따르게 되고 통계주체가 노출되어도 이를 규율할 수 없는 상태이다. 그래서 통계의 비밀이 노출되는 문제가 발생하고 있다.

예를 들어 범죄분석통계는 범죄의 현황과 범죄심리 및 그 양적·질적 변화를 조사

아니다.
제27조의2(통계작성·공표 과정에서의 영향력 행사, 누설 및 목적 외 사용의 금지 등)
① 누구든지 정당한 사유 없이 통계작성기관에서 작성 중인 통계(통계작성기관의 결재권자로부터 결재를 받기 전의 통계로 이를 서술한 정보와 통계자료를 포함한다. 이하 같다) 또는 작성된 통계(통계작성기관의 결재권자로부터 결재를 받은 통계로 이를 서술한 정보와 통계자료를 포함한다. 이하 같다)를 공표 전에 변경하거나 공표 예정 일시를 조정할 목적으로 통계종사자(통계작성기관으로부터 통계 작성업무의 전부 또는 일부를 위탁받아 그 업무에 종사하는 자를 포함한다)에게 영향력을 행사해서는 아니 된다.
② 누구든지 통계작성기관에서 작성 중인 통계 또는 작성된 통계를 공표 전에 제공 또는 누설하거나 목적 외의 용도로 사용해서는 아니 된다. 다만, 다음 각 호의 어느 하나에 해당하는 경우에는 공표 전에 제공할 수 있다.
1. 통계작성기관이 새로운 통계를 작성하거나 기존의 통계를 변경하기 위하여 관계 기관(해당 통계의 대상이 되는 산업·물가·인구·주택·문화·환경 등과 관련된 기관을 말한다. 이하 같다) 및 전문가에게 의견을 구하거나 공청회를 개최할 때에 작성 중인 통계를 제공하는 경우
2. 다음 각 목의 어느 하나에 해당하는 경우로서 통계작성기관이 관계 기관에 작성된 통계를 제공하는 경우
가. 행정자료를 단순 집계하여 작성하는 통계를 제공하는 경우
나. 관계 기관이 업무수행을 위하여 필요하다고 요청하는 경우
3. 다른 기관으로부터 위임·위탁을 받아 작성된 통계를 통계작성기관이 그 위임·위탁 기관에 제공하는 경우
③ 통계작성기관은 제2항제2호나목에 따라 작성된 통계를 제공하는 경우 내용, 일시, 제공자, 제공방법, 제공받은 기관명 및 담당자를 기록한 후 이를 증명할 수 있는 자료를 첨부하여 5년 동안 보존하여야 한다.
④ 통계작성기관은 제2항제2호나목에 따라 작성된 통계를 제공하는 경우 공표 예정일 전날 낮 12시 이후에 제공하여야 한다. 다만, 국제기구의 요청을 받아 통계를 제출하는 등 국제협력을 위하여 필요하거나 경제위기, 시장불안 등으로 관계 기관의 대응이 시급하다고 인정하는 경우 등 대통령령으로 정하는 경우는 그러하지 아니하다.
⑤ 통계작성기관은 제2항제2호나목에 따라 작성된 통계를 제공하는 경우 매년 3월 31일까지 전년도의 작성된 통계와 공표된 통계를 비교·점검한 결과(통계 내용, 공표 예정 일시의 변경 여부 및 변경 시 해당 사유를 말한다)를 대통령령으로 정하는 바에 따라 공개하고, 통계청장에게 제출하여야 한다.

측정하여 범죄의 예방과 검거대책수립 등 형사대책 및 사회정책 수립에 기초자료제공을 위해 작성되는 통계이다. 이 통계는 전국 각급 수사기관(검찰, 경찰, 특별사법경찰)에서 범죄사건을 수사하면서 작성·전산 입력한 각 범죄통계원표(발생통계원표, 검거통계원표, 피의자통계원표)를 토대로 작성하는데, 보고 항목은 발생 통계원표의 발생지 조사기관, 죄명, 장소, 수법 등 21개 항목, 검거 통계원표의 검거 조사기관, 검거인원, 검거유형 등 19개 항목, 피의자 통계원표(사건표)의 검찰 34개 항목, 경찰 특별사법경찰 33개 항목으로 구성되어 있다.

대검찰청은 이를 범죄분석이라는 발간물과 국가통계포털을 통해 공개하고 있는데 공개항목을 보면 항목이 세분화되어 있어서 1명 단위로 공개되는 경우가 많다. 이 경우 개인에 대한 민감한 개인정보가 통계 공표를 통해서 노출되게 되는 것이다. 예를 들어 '공무원 범죄자 현황'을 보면, (i) 소속기관이 중앙부처별, 직급별로 세분화되어 있어서 1명이 고스란히 드러나는 경우가 많다. (ii) 범죄도 죄명이 절도, 장물, 사기, 횡령, 배임, 손괴, 살인, 강도, 방화, 강간, 간통, 혼인빙자간음, 기타 음란행위, 도박과 복표, 등은 물론 각종 특별법에 의한 범죄가 세분화되어 있어서 1명이 고스란히 드러나는 경우가 많다. (iii) 이와 같이 1명 단위로 공개된 범죄별로 범행동기(생활비 마련, 유흥비 마련, 도박비 마련, 허영·사치심, 치부·가정불화, 호기심·유혹, 우발적, 현실불만, 부주의, 사행심, 보복(신고, 고소, 수사협조, 증언, 기타))도 세분화되어 있고, (iv) 1명 단위로 공개된 죄명별로 생활 정도, 혼인관계(동거, 이혼, 사별, 미혼, 미상), (v) 1명 단위로 공개된 죄명별 처분결과, (vi) 1명 단위로 공개된 죄명별 전과자 범행 시의 연령, 전회 처분, 전과자 보호처분 상황, 재범기간 및 종류, 전과자 마약류 등 상용여부, 범행 시 정신상태(정상, 정신이상 정신박약 기타정신장애, 주취, 월경 시 이상) 등의 민감한 정보가 개인을 식별 가능하게 공개되고 있다.

그림 4-34 공무원 범죄자 현황 자료 예시

공 무 원 범 죄 자

단위 : 명

2010년	계	국가										
		소 계	감사원	국가정보원	법제처	국가보훈처	국무총리실	국정홍보처	외교통상부	기획재정부	국세청	관세청
계	13,581	2,173	5	4	1	12	1	-	7	35	130	35
	① 100.0	16.0	0.0	0.0	0.0	0.1	0.0	0.0	0.1	0.3	1.0	0.3
형법범계	7,017	1,120	2	1	1	8	1	-	4	17	66	13
	① 100.0	16.0	0.0	0.0	0.0	0.1	0.0	0.0	0.1	0.2	0.9	0.2
재산범죄	1,240	136	-	-	-	2	-	-	2	5	7	1
	① 100.0	11.0	0.0	0.0	0.0	0.2	0.0	0.0	0.2	0.4	0.6	0.1
절도	126	27	-	-	-	1	-	-	-	1	3	-
장물	-	-	-	-	-	-	-	-	-	-	-	-
사기	373	59	-	-	-	1	-	-	1	2	3	1
횡령	375	14	-	-	-	-	-	-	-	2	-	-
배임	228	2	-	-	-	-	-	-	-	-	-	-
손괴	138	34	-	-	-	-	-	-	1	-	1	-
강력범죄(흉악)	158	34	-	-	-	-	-	-	-	-	4	1
	① 100.0	21.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.5	0.6
살인	6	4	-	-	-	-	-	-	-	-	-	-
강도	5	2	-	-	-	-	-	-	-	-	-	-
방화	3	1	-	-	-	-	-	-	-	-	-	-
강간	144	27	-	-	-	-	-	-	-	-	4	1
강력범죄(폭력)	1,451	296	2	1	1	1	1	-	1	7	17	2
	① 100.0	20.4	0.1	0.1	0.1	0.1	0.1	0.0	0.1	0.5	1.2	0.1
폭행	640	106	1	1	1	-	-	-	-	3	7	-
상해	445	94	1	-	-	-	1	-	1	3	6	-
협박	38	5	-	-	-	-	-	-	-	-	-	1
공갈	11	4	-	-	-	-	-	-	-	-	-	-
약취와유인	-	-	-	-	-	-	-	-	-	-	-	-
체포와감금	8	8	-	-	-	-	-	-	-	-	-	-
[illegible]	2	1	-	-	-	-	-	-	-	-	-	-
[illegible]	307	78	-	-	-	1	-	-	-	1	4	1
위조범죄	1,017	77	-	-	-	2	-	-	-	2	7	-
	① 100.0	7.6	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.2	0.7	0.0
통화	2	-	-	-	-	-	-	-	-	-	-	-
유가증권·인지·우표	1	-	-	-	-	-	-	-	-	-	-	-
문서	1,012	77	-	-	-	2	-	-	-	2	7	-
인장	2	-	-	-	-	-	-	-	-	-	-	-
공무원범죄	2,176	392	-	-	-	2	-	-	-	-	19	7
	① 100.0	18.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.9	0.3
직무유기	896	107	-	-	-	1	-	-	-	-	1	-
직권남용	401	184	-	-	-	1	-	-	-	-	3	-
수뢰	839	94	-	-	-	-	-	-	-	-	14	7
증뢰	40	7	-	-	-	-	-	-	-	-	1	-
풍속범죄	189	39	-	-	-	-	-	-	-	-	1	-
	① 100.0	20.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.5	0.0
간통	53	12	-	-	-	-	-	-	-	-	-	-
음화반포등	-	-	-	-	-	-	-	-	-	-	-	-
도박과복표	127	24	-	-	-	-	-	-	-	-	1	-
기타음란행위	5	2	-	-	-	-	-	-	-	-	-	-
신앙	4	1	-	-	-	-	-	-	-	-	-	-
과실범죄	50	10	-	-	-	-	-	-	-	-	-	-
	① 100.0	20.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
과실치사상	12	3	-	-	-	-	-	-	-	-	-	-
업무상과실치사상	34	7	-	-	-	-	-	-	-	-	-	-
실화	4	-	-	-	-	-	-	-	-	-	-	-
기타형법범죄	736	136	-	-	-	1	-	-	1	3	11	2
	① 100.0	18.5	0.0	0.0	0.0	0.1	0.0	0.0	0.1	0.4	1.5	0.3
명예에관한죄	203	41	-	-	-	-	-	-	1	1	1	1
비밀침해	11	1	-	-	-	-	-	-	-	-	-	-
신용업무방해	93	19	-	-	-	1	-	-	-	-	3	1
주거침입	36	7	-	-	-	-	-	-	-	-	1	-
비밀침해	-	-	-	-	-	-	-	-	-	-	-	-
유가증권에관한죄	2	-	-	-	-	-	-	-	-	-	-	-
낙태의죄	-	-	-	-	-	-	-	-	-	-	-	-
아편에관한죄	-	-	-	-	-	-	-	-	-	-	-	-
방화와실화의죄	5	-	-	-	-	-	-	-	-	-	-	-
공무방해에관한죄	271	38	-	-	-	-	-	-	-	2	3	-
주거침입등인질의죄	9	4	-	-	-	-	-	-	-	-	1	-
도주와범인은닉의죄	63	14	-	-	-	-	-	-	-	-	1	-
위증과증거인멸의죄	35	10	-	-	-	-	-	-	-	-	1	-
무고의죄	1	1	-	-	-	-	-	-	-	-	-	-
공안을해하는죄	-	-	-	-	-	-	-	-	-	-	-	-
내란의죄	-	-	-	-	-	-	-	-	-	-	-	-
폭발물에관한죄	2	1	-	-	-	-	-	-	-	-	-	-
공수와수리에관한죄	5	-	-	-	-	-	-	-	-	-	-	-
기타	-	-	-	-	-	-	-	-	-	-	-	-

주) 1. ①은 백분율(%), 2. 특정범죄가중처벌등에관한법률로 가중처벌되는 경우 포함

그 외에도 소년범죄자, 소녀범죄자, 학생범죄자 등에 대해서도 마찬가지로 개인이 식별될 수 있도록 공개하고 있다.

이런 문제를 시정하기 위해서는 통계법에서 통계의 공표와 관련해서 통계주체가 식별되거나 식별될 가능성이 있게 공표되어서는 안 된다는 규정을 두어야 한다. 이 경

우 통계 노출통제 기법도 발달할 것이다.

한편, 통계 공표 시의 비밀보호의 예외 규정을 둘지도 검토되어야 할 텐데, 유럽연합의 예처럼 당사자의 동의가 있는 경우나 예외적으로 법인이나 사업체의 경제활동과 관련한 통계에서 당사자의 배제 요청이 없는 경우, 공개된 행정자료를 활용하는 경우에는 식별 가능한 공표도 허용하는 수준이 적절할 것이다.

② 통계자료의 제3자 제공과 관련한 비밀보호와 관련하여

현행 통계법에는 통계자료의 제3자 제공과 관련한 두 개의 조문이 있다. 제30조는 통계작성기관이 통계 작성을 위해 다른 통계작성기관에 통계자료의 제공을 요청하는 경우를 규정하고 있고, 제31조는 제3자가 통계 목적 외의 목적으로 통계자료를 요청하는 경우를 규정하고 있다.

제30조 (통계자료의 제공) ①통계작성기관의 장은 통계의 작성을 위하여 필요한 경우에는 다른 통계작성기관에 통계자료의 제공을 요청할 수 있다. 이 경우 요청을 받은 통계작성기관의 장은 특별한 사유가 없는 한 이에 응하여야 한다.
②통계작성기관의 장은 다른 통계작성기관의 장으로부터 제1항에 따라 통계자료를 제공하는 때에는 특정의 개인이나 법인 또는 단체 등을 식별할 수 없는 형태로 통계자료를 처리한 후 제공하여야 한다. 다만, 다른 통계작성기관의 장이 통계의 작성을 위한 방문조사·전화조사·우편조사 등에 따른 표본조사의 표본으로 사용하기 위하여 제1항에 따른 요청을 하는 때에는 특정의 개인이나 법인 또는 단체 등이 식별되는 형태로 통계자료를 제공할 수 있다.
③제2항에 따라 통계작성기관으로부터 제공받은 통계자료는 이를 제공받은 목적 외의 목적으로 사용하거나 다른 자에게 제공하여서는 아니 된다.
④통계자료의 제공방법 등에 관하여 필요한 사항은 대통령령으로 정한다.

제31조 (통계자료의 이용) ①특정의 대상에 관한 수량적 정보를 작성하거나 학술연구를 위한 목적으로 통계자료를 이용하고자 하는 자는 대통령령으로 정하는 바에 따라 통계작성기관의 장에게 통계자료의 제공을 신청할 수 있다.
②통계작성기관의 장은 제1항에 따른 신청을 받은 때에는 통계자료의 사용목적·내용 및 범위의 타당성을 심사하여 타당하다고 판단되는 경우에는 이를 제공하여야 한다. 이 경우 통계작성기관의 장은 특정의 개인이나 법인 또는 단체 등을 식별할 수 없는 형태로 통계자료를 처리한 후 제공하여야 한다.
③제2항에도 불구하고 통계작성기관의 장은 해당 통계자료를 다른 자료와 대응 또는 연계함으로써 특정의 개인이나 법인 또는 단체 등의 식별이 가능하게 되는 경우에는 통계자료를 제공하지 아니할 수 있다.
④제2항에 따라 통계작성기관으로부터 제공받은 통계자료는 이를 제공받은 목적 외의 목적으로 사용하거나 다른 자에게 제공하여서는 아니 된다.
⑤통계자료의 제공방법 등에 관하여 필요한 사항은 대통령령으로 정한다.

통계작성기관인 제3자가 통계작성기관에게 통계 작성을 위해 통계자료를 요청하여 통계자료를 제공하는 경우는 '통계주체를 식별할 수 없도록 해야 한다'는 내용의 규정을 두고 있다. '특정의 개인이나 법인 또는 단체 등을 식별할 수 없는 형태로 통계자

료를 처리한 후 제공하여야 한다'는 것이다. 그러나 '특정의 개인이나 법인 또는 단체 등을 식별할 수 없는 형태로 통계자료를 처리하는 것'은 식별 가능성을 제거하는 것이 아니므로 비밀보호로 충분하지 못하다. 예를 들어 국민건강보험공단이 보유하고 있는 통계자료로 요양급여 내역(즉, 의료서비스 이용 내역)이 있다고 할 때, 제3자인 국세청에서 요양급여 내역에 대한 통계자료를 요구한다면 특정 개인이나 법인 또는 단체 등을 식별할 수 없도록 주민등록번호, 이름 등을 삭제하거나 임의의 일련번호로 대체하여 처리하는 경우 그 자체로는 식별할 수 없는 형태로 통계자료를 처리한 것이 되지만, 이를 제공받은 국세청에서는 신용카드의 결제내역으로 각 개인을 식별해 낼 수도 있다. 따라서 식별할 수 없는 형태로 통계자료를 처리하는 것만으로는 다른 정보와의 결합을 통한 식별 가능성을 막지 못한다.

반면 그 외 목적의 통계자료를 제공해 달라고 요구하는 경우를 규정한 통계법 제31조에서는 '다른 자료와의 연계를 통해서 개인이 식별될 수 있는 경우에는 제공하지 아니할 수 있다'는 규정을 두고 있다. 통계주체의 비밀보호를 위한 것임에도 제공 여부의 판단을 아무런 기준도 마련해 놓지 않고 통계작성기관의 판단에만 맡기는 방식의 현행 규율은 부당하다.

마. 통계의 비밀보호를 위한 통계법의 개정 방안

통계의 비밀보호를 위해서는 통계법이 다음과 같이 개정되어야 한다.

첫째, 통계의 비밀을 보호한다는 점을 명백하게 밝혀야 한다.

둘째, 통계의 공표 시 통계단위가 식별되지 않도록 해야 한다는 점을 명백하게 밝혀야 한다. 그 예외로는 유럽연합과 같이 법령의 규정에 의한 경우로 한정하는 것이 좋을 것이다.

셋째, 통계는 반드시 통계 목적으로만 활용되어야 함을 명백하게 밝혀야 한다.

넷째, 통계자료에 대한 연구 목적의 활용에 대해서도 분명한 규율을 제정하는 것이 바람직하다.

3. 통계와 개인정보 보호

(1) 통계와 개인정보 보호의 필요성

통계가 작성되는 과정에서는 민감한 개인정보가 수집되고 이용되며 이를 활용하는 과정에서도 개인정보가 처리된다. 따라서 통계의 작성·활용 과정에서 당연히 개인정

보 보호가 이루어져야 한다. 그런데 우리나라 통계법과 개인정보 보호법은 통계에 관한 개인정보 보호법의 적용을 배제하고 있어서 많은 문제점을 드러내고 있다.

(2) 개인정보 보호법 제58조의 문제점

가. 개인정보 보호법 제58조

개인정보 보호법 제58조는 “공공기관이 처리하는 개인정보 중 통계법에 따라 수집되는 개인정보에 대해서 개인정보 보호법 제3장 ~ 제6장을 배제한다.”는 규정을 두고 있다.

> 제58조(적용의 일부 제외) ① 다음 각 호의 어느 하나에 해당하는 개인정보에 관하여는 제3장부터 제7장까지를 적용하지 아니한다.
> 1. 공공기관이 처리하는 개인정보 중 「통계법」에 따라 수집되는 개인정보

이와 관련하여 개인정보 보호법의 적용이 배제되는 경우가 어떤 경우인지 해석에 논란의 여지가 있다.

나. 개인정보 보호법 제58조 적용 요건

개인정보 보호법 제58조의 요건을 나누어 보면 (i) 공공기관이 처리하는 개인정보 중 (ii) 통계법에 따라서 수집되는 개인정보를 대상으로 하고 있다.

첫째, 분명한 것은 통계법에 의한 수집인 경우만을 그 대상으로 할 뿐, 통계법에 의하지 않을 경우는 개인정보 보호법이 적용된다는 점이다.

반면 통계법에 따라서 개인정보를 수집하는 경우만이 그 대상(적용 배제의 대상)이 될지, 통계법에 따라서 수집되기만 하면 그 정보를 처리하는 모든 경우가 그 대상이 될지는 문구가 명확하지 않아서 논란이 있을 수 있다. 가능한 두 가지 해석이 있다.

하나는 개인정보를 수집하는 경우만이 적용 배제의 대상이라고 해석하는 것이다. 이 경우 통계법에 의해 개인정보를 수집하는 행위에 대해서만 개인정보 보호법 적용이 배제되고 이 정보를 다른 개인정보와 연계·결합하거나, 이 정보를 제3자에게 제공하는 경우에는 개인정보 보호법이 적용된다. 물론 이 경우에도 개인정보 보호법 제18조 제2항 제4호에 의하여 ‘통계 작성 및 학술연구 등의 목적을 위하여 필요한 경우로서 특정 개인을 알아볼 수 없는 형태로 개인정보를 제공하는 경우’는 목적 외로 이용할 수 있고 제3자에게 제공할 수 있다. 통계 작성이나 학술연구 등의 목적에 필요하

여 특정 개인을 알아볼 수 없는 형태로 개인정보를 제공하는 경우가 아니라면 개인정보 보호법의 규정에 따라서 별도의 동의를 받아야만 목적 외 이용이나 제3자 제공 등이 허용된다고 볼 것이다.

다른 하나는 개인정보를 수집하는 경우뿐 만 아니라 수집된 개인정보를 다른 개인정보와 결합하거나 제3자에게 제공하는 등 공공기관이 그 개인정보를 처리하는 모든 경우에 개인정보 보호법 적용이 배제된다는 견해이다. 이 경우는 통계법이 적용될 것이다.

전자와 같이 해석하는 것이 법문에서 자연스러워 보인다. 그 이유는 (i) '다음 각호의 하나에 해당하는 개인정보에 관하여는 제3장부터 제7장까지를 적용하지 아니한다'고 규정하고 있는데, '통계법에 따라 수집되는 개인정보'라고 하였으므로 그 대상 개인정보는 통계법에 따라 수집되는 개인정보로 해석하는 것이 자연스럽기 때문이고, (ii) 제2호부터 제4호를 보면 '국가안전보장과 관련된 정보분석을 목적으로 수집 또는 제공 요청되는 개인정보'(제2호), '공중위생 등 공공의 안전과 안녕을 위하여 긴급히 필요한 경우로서 일시적으로 처리되는 개인정보'(제3호), '언론, 종교단체, 정당이 각각 취재·보도, 선교, 선거 입후보자 추천 등 고유 목적을 달성하기 위하여 수집·이용하는 개인정보'라고 하여, 제2호는 수집 또는 제공 요청을 대상으로, 제3호는 처리되는 개인정보를, 제4호는 수집·이용하는 개인정보를 대상으로 하고 있음을 구별해서 사용하고 있기 때문에 제1호는 '수집하는 개인정보'만을 대상으로 한다고 보는 것이 자연스럽기 때문이다.

표 4-43 개인정보 보호법과 통계법의 적용범위

적용범위		적용 법률
통계법에 따라 개인정보를 수집하는 경우	공공기관으로부터 수집	개인정보 보호법(제3장~7장 제외), 통계법
	그 밖의 개인, 법인, 단체로부터 수집	개인정보 보호법, 통계법
수집한 정보를 활용하는 경우(결합, 연계, 제3자 제공 등)	통계, 연구 목적으로만 개인을 알아볼 수 없게 하여 제공하는 경우	개인정보 보호법 제18조 제2항 제4호, 개인정보 보호법(3장~7장 적용), 통계법
	개인을 알아볼 수 있는 경우	개인정보 보호법(3장~7장 적용), 통계법

다. 개인정보 보호법 제58조 제1호 해당 시 적용되는 개인정보 보호법 규율

한편 개인정보 보호법 제58조 제1항에 해당하여 개인정보 보호법 제3장 ~ 제7장의 적용이 배제되더라도 다음과 같은 규정은 적용된다.

① 제58조 제4항에 따른 준수의무

개인정보 보호법 제58조 제4항은 “제58조 제1항에 의해 개인정보 보호법 제3장 ~ 제7장의 규정이 적용되지 않더라도 개인정보처리자는 개인정보를 처리할 때, (i) 그 목적을 위하여 필요한 범위에서 최소한의 기간에 최소한의 개인정보만을 처리하여야 하며, (ii) 개인정보의 안전한 관리를 위하여 필요한 기술적·관리적 및 물리적 보호조치, (iii) 개인정보의 처리에 관한 고충처리, (iv) 그 밖에 개인정보의 적절한 처리를 위하여 필요한 조치를 마련하여야 한다”는 규정을 두고 있다.

따라서 통계법에 따라 통계 작성을 위한 개인정보 수집 시에도 목적에 따른 필요 최소한의 범위의 개인정보를 수집해야 하고 목적 범위를 넘는 개인정보 수집을 해서는 안 되며 목적 달성 후에는 즉시 파기해야 한다. 그리고 개인정보의 안전한 관리를 위하여 필요한 기술적·관리적 및 물리적 보호조치도 갖추어야 한다.

개인정보 처리에 관한 고충처리의 경우는 그 범위가 문제인데, 개인정보의 열람, 정정, 처리정지, 삭제를 요구할 권리를 보장해 주어야 할 것이다. 그 밖의 필요한 조치로는 해당 개인정보의 성질 등에 따라 취하는 조치로 암호화 조치 등을 예로 들 수 있을 것이다.

② 개인정보 보호법 제3조 개인정보 보호 원칙

한편 개인정보 보호법 제3조의 규정은 제1장에 규정되어 있으므로 이를 준수해야 한다.

제3조(개인정보 보호 원칙) ① 개인정보처리자는 개인정보의 처리 목적을 명확하게 하여야 하고 그 목적에 필요한 범위에서 최소한의 개인정보만을 적법하고 정당하게 수집하여야 한다.
② 개인정보처리자는 개인정보의 처리 목적에 필요한 범위에서 적합하게 개인정보를 처리하여야 하며, 그 목적 외의 용도로 활용하여서는 아니 된다.
③ 개인정보처리자는 개인정보의 처리 목적에 필요한 범위에서 개인정보의 정확성, 완전성 및 최신성이 보장되도록 하여야 한다.
④ 개인정보처리자는 개인정보의 처리 방법 및 종류 등에 따라 정보주체의 권리가 침해받을 가능성과 그 위험 정도를 고려하여 개인정보를 안전하게 관리하여야 한다.
⑤ 개인정보처리자는 개인정보 처리방침 등 개인정보의 처리에 관한 사항을 공개하여야 하며, 열람청구권 등 정보주체의 권리를 보장하여야 한다.
⑥ 개인정보처리자는 정보주체의 사생활 침해를 최소화하는 방법으로 개인정보를 처리하여야 한다.
⑦ 개인정보처리자는 개인정보의 익명처리가 가능한 경우에는 익명에 의하여 처리될 수 있도록 하여야 한다.
⑧ 개인정보처리자는 이 법 및 관계 법령에서 규정하고 있는 책임과 의무를 준수하고 실천함으로써 정보주체의 신뢰를 얻기 위하여 노력하여야 한다.

③ 개인정보 보호법 제4조 정보주체의 권리 보장

개인정보 보호법 제4조에 따라서 정보주체의 권리도 보장해야 한다. 따라서 개인정보의 처리에 관한 정보를 제공받을 권리를 보장해 주어야 하고, 개인정보의 처리정지, 정정, 삭제 및 파기를 요구할 권리를 보장해 주어야 한다.

제4조(정보주체의 권리) 정보주체는 자신의 개인정보 처리와 관련하여 다음 각 호의 권리를 가진다. 1. 개인정보의 처리에 관한 정보를 제공받을 권리 2. 개인정보의 처리에 관한 동의 여부, 동의 범위 등을 선택하고 결정할 권리 3. 개인정보의 처리 여부를 확인하고 개인정보에 대하여 열람(사본의 발급을 포함한다. 이하 같다)을 요구할 권리 4. 개인정보의 처리 정지, 정정·삭제 및 파기를 요구할 권리 5. 개인정보의 처리로 인하여 발생한 피해를 신속하고 공정한 절차에 따라 구제받을 권리

라. 개인정보 보호법 적용 배제 조항의 문제점

개인정보 보호법에서 개인정보 보호법 제3장 ~ 제7장의 규율을 일률적으로 배제하는 경우로 통계법에 따라 수집되는 개인정보를 포함시킨 것은 적절한 비례의 원칙을 지킨 것으로 보기는 어렵다.

물론 개인정보 보호법 제58조의 제4항, 제3조, 제4조에 따라서 개인정보 보호원칙이나 개인정보주체의 권리가 일부 보장된다고 볼 수 있겠지만, 그와 같은 규정만으로는 개인정보 보호가 충분하다고 보기 어렵다.

실제로 공공기관의 활동에 필요한 개인정보 수집에 대해서는 개인정보 보호법 제15조 제1항에서 법률에 특별한 규정이 있거나 법령상 의무를 준수하기 위하여 불가피한 경우나(제2호), 공공기관이 법령 등에서 정하는 소관 업무의 수행을 위하여 불가피한 경우(제3호)에 개인정보를 수집할 수 있다고 규정하고 있고, 이 규정을 통해서 얼마든지 개인정보의 수집이나 이용을 하면서 업무수행을 할 수 있다. 이런 경우는 수집한 목적 범위 내에서 제3자에게도 제공할 수 있다고도 규정하고 있다(제17조 제1항 제2호). 목적 외 이용이나 목적 외로 제3자에게 제공할 수도 있는데, 예를 들어 법률에 특별한 규정이 있는 경우나(제18조 제2항 제2호), 개인정보를 목적 외의 용도로 이용하거나 이를 제3자에게 제공하지 아니하면 다른 법률에서 정하는 소관 업무를 수행할 수 없는 경우로서 보호위원회의 심의·의결을 거친 경우(제5호), 조약 그 밖의 국제협정의 이행을 위하여 외국정부 또는 국제기구에 제공하기 위하여 필요한 경우(제6호), 범죄의 수사와 공소의 제기 및 유지를 위하여 필요한 경우(제7호) 등이다.

이런 경우에 개인정보 보호법에 따라서 정보주체의 요구가 있는 경우 개인정보의 수집 출처의 고지 의무(제20조 제1항), 개인정보의 파기 의무(제21조), 민감정보의 처

리제한 의무(제23조), 고유식별정보의 처리제한(제24조), 업무위탁에 따른 개인정보의 처리제한(제27조), 개인정보취급자에 대한 감독(제28조), 개인정보 안전조치의무(제29조)), 개인정보처리방침의 수립 및 공개의무(제30조), 개인정보 보호책임자의 지정(제31조), 개인정보파일의 등록 및 공개(제32조), 개인정보 영향평가(제33조), 개인정보 유출 통지(제34조), 개인정보의 열람(제35조), 개인정보의 정정·삭제(제36조), 개인정보의 처리정지(제37조), 손해배상책임(제39조), 개인정보 분쟁조정(제40조) 등의 규정이 적용되고 있다. 이런 규정들이 적용되어도 행정 목적을 달성하면서 개인정보를 이용하는 데는 아무런 문제도 없다. 반면 현재 통계법에 의하면 이와 같은 규정들이 전면적으로 배제된다. 이는 통계의 작성과 관련한 개인정보 보호의 중대한 흠결이다.

표 4-44 개인정보 보호법과 통계법의 개인정보 보호 관련 규정의 문제점

법률	관련 규정	검 토
개인정보 보호법	제58조 제1호 : 개인정보 보호법 적용 배제 제18조 제2항 : 목적 외 이용, 제3자 제공 허용	개인정보 보호법 적용 배제의 범위 모호 개인정보 보호법 적용 배제 타당성 부족
통계법령	통계의 승인 통계자료(개인정보)의 수집 - 통계 작성에 관한 협조 - 행정자료의 제공 - 사법기관 등의 자료제공 - 자료제출 요구 - 실지조사 등 통계의 공표 통계의 목적 외 사용 금지 통계자료의 보유 및 관리 통계자료의 제공 통계자료의 이용 통계응답자의 응답 의무	통계 작성을 위한 자료의 수집, 자료의 이용, 제공에 대한 규정 미비. * 개인정보의 수집, 출처의 고지 의무(개인정보 보호법 제20조 제1항, 이하 개인정보 보호법의 해당 조항), 개인정보의 파기 의무(제21조), 민감정보의 처리제한 의무(제23조), 고유식별정보의 처리제한(제24조), 업무위탁에 따른 개인정보의 처리제한(제27조), 개인정보취급자에 대한 감독(제28조), 개인정보 안전조치의무(제29조)), 개인정보처리방침의 수립 및 공개의무(제30조), 개인정보 보호책임자의 지정(제31조), 개인정보파일의 등록 및 공개(제32조), 개인정보 영향평가(제33조), 개인정보 유출 통지(제34조), 개인정보의 열람(제35조), 개인정보의 정정, 삭제(제36조), 개인정보의 처리정지(제37조), 손해배상책임(제39조), 개인정보분쟁조정(제40조) 등

실제로 해외의 법제에서도 대부분 국가는 통계 작성을 위하여 개인정보를 처리하는 과정에서 개인정보 보호를 개인정보 보호법제와 통계법제에서 함께 규율하는 경우가 대부분이고, 통계 작성을 위한 개인정보의 수집이나 통계자료에 포함된 개인정보의 보유, 이용, 제3자 제공 등과 관련해서도 필요한 규정에 대해서만 특칙을 두고 있는 경우가 대부분이다.

마. 제도 개선방향

통계의 작성 및 활용에 관하여 개인정보 보호법의 주요 규정(예를 들어 제3장 ~ 제7장)의 적용을 전면적으로 배제하는 것은 전혀 합리적 논거를 찾기 어렵다. 실제로 해외의 입법례에서도 통계에서 개인정보 보호법의 주요 규정을 일괄적으로 배제하는 입법을 취하고 있지 않다. 통계의 작성 및 활용도 공공기관의 법령상 의무를 수행하는 과정에서 개인정보를 수집, 처리, 활용하는 경우와 마찬가지로 규율해도 충분하다.

통계법에 의한 통계에 대해서는 정보주체의 동의가 없어도 개인정보를 수집할 수 있고 목적 외 이용도 통계법의 규정이 있는 경우에는 허용되고 제3자 제공도 통계법의 규정이 있는 경우는 허용되는 것으로 규율하는 것이 바람직하다. 다만, 통계법에 의하여 처리되는 개인정보에 대해서는 정보주체에게 고지할 의무의 완화(제20조), 개인정보의 정정·삭제(제36조)에 대한 예외, 개인정보의 처리정지(제37조)에 대한 예외 등을 규정할 수는 있을 것이다.

(3) 통계 작성 목적의 개인정보 제공 규정

개인정보 보호법은 통계 작성 및 학술연구 등의 목적을 위하여 필요한 경우로서 특정 개인을 알아볼 수 없는 형태로 개인정보를 제공하는 경우에는 정보주체 또는 제3자의 이익을 부당하게 침해할 우려가 있을 때를 제외하고는 개인정보를 목적 외의 용도로 이용하거나 이를 제3자에게 제공할 수 있다고 규정하고 있다(제18조 제2항 제4호).

> 제18조(개인정보의 목적 외 이용·제공 제한) ① 개인정보처리자는 개인정보를 제15조제1항에 따른 범위를 초과하여 이용하거나 제17조제1항 및 제3항에 따른 범위를 초과하여 제3자에게 제공하여서는 아니 된다.
> ② 제1항에도 불구하고 개인정보처리자는 다음 각 호의 어느 하나에 해당하는 경우에는 정보주체 또는 제3자의 이익을 부당하게 침해할 우려가 있을 때를 제외하고는 개인정보를 목적 외의 용도로 이용하거나 이를 제3자에게 제공할 수 있다. 다만, 제5호부터 제9호까지의 경우는 공공기관의 경우로 한정한다.
> 4. 통계작성 및 학술연구 등의 목적을 위하여 필요한 경우로서 특정 개인을 알아볼 수 없는 형태로 개인정보를 제공하는 경우

이때 개인정보처리자는 공공기관이거나 기업을 불문한다. 따라서 기업이 자신이 보유하고 있는 개인정보를 통계 작성 및 학술연구 등의 목적을 위하여 특정 개인을 알아볼 수 없는 형태로 개인정보를 제공하는 경우에는 해당 개인정보주체에게 동의를 받지 않고도 제3자에게 제공 가능하다. 이와 같이 개인정보주체의 동의가 필요 없는 경우인 '특정 개인을 알아볼 수 없는 형태'란 무슨 의미인지가 해석상 논란이 있다.

한 견해는 이를 ‘익명화’를 의미하는 것으로 해석한다. 즉, 개인정보를 제공대상이 되는 자뿐만 아니라 그 외의 제3자를 포함하여 특정 개인을 알아볼 수 있는지를 판단해야 하고 다른 제3자로부터 입수 가능한 정보와 결합하여 개인을 알아볼 수 있는 경우도 특정 개인을 알아볼 수 있는 경우로 보아야 하며, 향후의 재식별 가능성까지도 포함하여 더 이상 개인을 식별할 수 없도록 익명화한 경우에만 이 규정을 적용할 수 있다고 본다.

다른 견해는 수령자가 자신이 보유하고 있는 정보를 바탕으로 특정 개인을 알아볼 수 없다면 이 규정의 적용대상이 된다고 해석한다. 이 견해에 의하면 예를 들어 가명화된 개인정보도 정보주체의 동의 없이 제3자에게 제공할 수 있다는 결론에 이를 수도 있다. 즉, 수령자가 가명화된 개인정보로부터 특정 개인을 알아볼 수 없다면 당사자의 동의를 받지 않아도 된다는 것이다.

이 규정의 범위가 넓은 점, 수령자가 해당 개인정보를 또 다른 제3자에게 제공하는 것에 대한 금지 규정도 두고 있지 않은 점에 비춰 본다면 당사자의 동의를 받지 않아도 된다고 하려면 ‘현재는 물론 장래를 포함하여 수령자는 물론 제3자를 포함하여 입수 가능한 다른 정보와 결합하여 특정 개인을 알아볼 가능성이 있는 경우나 기술발전으로 재식별 가능성이 있는 경우’를 배제하는 것, 즉 ‘익명화’를 의미한다고 해석해야 한다.

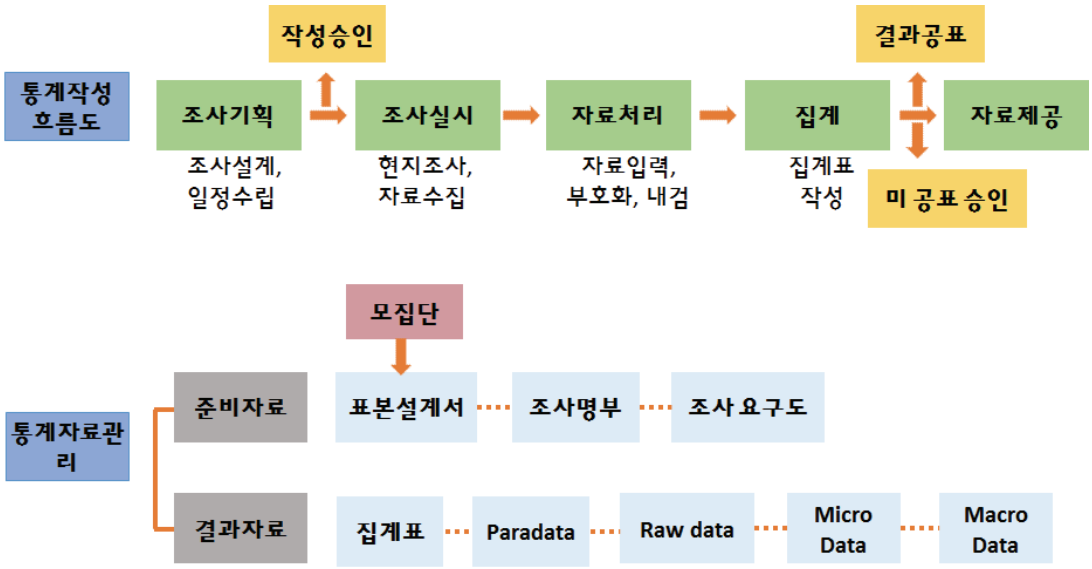
4. 통계 작성과정의 데이터 연계

(1) 통계의 작성과정

실제 통계의 작성과정은 통계조사의 기획과 설계, 자료수집, 자료처리, 자료의 정리 및 공급의 절차로 구분할 수 있다. 자료수집은 paradata → raw data → micro data → macro data의 과정을 거친다.[314] 통계를 위한 자료로는 조사자료 및 행정자료가 있을 수 있다. 실제로 통계가 작성되는 과정을 통해서 통계의 작성·활용과 관련된 규율체계에 대해서 검토해 본다.

314) 김두만(2015), “국가통계작성 기획 및 승인관리”, 제5회 국가통계방법론 심포지엄.

그림 4-35 통계작성 관리 흐름도



(2) 통계 승인에 대한 통제의 필요성

가. 통계 승인의 의미

통계는 민감한 개인정보를 수집하여 작성하게 된다. 예를 들어 5년마다 실시하는 가족실태조사의 경우 2015년의 조사결과에는 전국 5,018가구(가구별로는 조사구번호, 거처번호, 가구번호, 가구원 번호가 할당된다)에서 조사한 여러 가지 민감한 항목이 포함되어 있다. 즉, Ⅰ. 개인관련 사항, Ⅱ. 가족에 대한 인식과 태도, Ⅲ. 가족 형성 및 변화, Ⅳ. 가족 관계, Ⅴ. 일, Ⅵ. 가족생활 및 노후준비, Ⅶ. 정책에 대한 인식, Ⅷ. 가족 돌봄, Ⅸ. 가구 특성과 같은 10가지 카테고리의 73개 항목에 대한 조사내용이 담겨 있다.[315]

315) [Ⅰ. 개인관련사항] 1. 성별 2. 생년월일 3. 교육수준
[Ⅱ.가족에 대한 인식과 태도] 4. 가족범위 5. 의지대상 6. 삶의 방식이나 가족형태 7. 결혼문화 8. 성역할 9. 부모부양 및 부모의 책임 10. 자녀와 자녀양육 11. 계층
[Ⅲ. 가족 형성 및 변화] 12. 남녀 적정 결혼 연령 13. 배우자 선택요건 14. 결혼비용 15. 혼인상태 16 현재 자녀 수 17.자녀계획 18. 적정 자녀수 19. 입양의향
[Ⅳ. 가족 관계] [부부 관계] 20. 부부의사 결정 21. 대화시간 22. 배우자와의 갈등 정도 23. 이혼 고려 여부 등 24. 갈등해결 방식 25. 자녀 돌봄 분담 정도 26. 부부관계 만족도 [청소년 자녀와의 관계] 27. 자녀 양육의 어려움 28. 자녀와의 관계 정도 29. 자녀관계 만족도 [성인 자녀와의 관계] 30. 자녀와 상호도움 여부 등 31. 자녀와 접촉 빈도 32. 성인자녀와의 관계 33. 자녀와의 관계만족도 [청소년 자녀 대상 부모와의 관계] 34. 부모님 생존 여부 35. 부모와의 관계 36. 부모관계 만족도 [성인자녀 대상 부모와의 관계] 37. 부모님 동거여부 38. 부모님 접촉 빈도 39. 부모와 관계 정도 40. 부모와 상호 도움 여부 등 41. 부모님 생활비 마련 42. 부모와의 관계 만족도
[Ⅴ. 일] 43. 근로여부 44. 고용형태 45. 근무시간 46. 정기적휴무 47. 일·가정 갈등 경험 정도
[Ⅵ. 가족 생활 및 노후준비] 48. 가사노동 여부 등 49. 가족 여가활동 50. 가족평균 여가시간 51. 여가시간 충분도 52. 생애설계 53. 원하는 노후거처 및 노후 동반자 54. 노후 경제적 준비 여부
[Ⅶ. 정책에 대한 인식] 55. 제도인식 정도 56. 일가정 조화지원책 57.남성의 가사/육아 참여확대

현재 통계법에 의해 승인받은 통계는 총 1,052종으로 지정통계 93종, 일반통계 959종이라고 하는데, 이들 통계는 가족실태조사와 정도의 차이는 있지만 민감한 개인정보나 법인 또는 단체에 관한 정보를 담고 있다.

지정통계란 통계작성기관의 장이 정부의 각종 정책의 수립·평가 또는 다른 통계의 작성 등에 널리 활용되는 통계로서 법이 규정하고 있는 요건[316]에 해당하는 통계를 통계청장에게 신청하여 통계청장이 지정·고시하는 통계이다(제17조).[317]

일반통계는 지정통계로 지정되지 않은 승인통계를 말한다. 통계작성기관은 작성하고자 하는 통계를 그 명칭, 종류, 목적, 조사대상, 조사방법, 통계표 서식, 조사사항의 성별구분 등 대통령령으로 정하는 사항에 관하여 미리 통계청장의 승인을 받아야 하는데(제18조) 지정통계로 지정되지 않은 승인통계를 일반통계라고 한다.

승인 통계는 통계법에 의해서 개인정보 보호법의 적용이 배제되고 통계법의 적용대상이 된다. 통계법은 승인 통계의 경우는 개인정보의 수집을 강제하는 규정을 두고 있고 이용·제3자 제공을 허용하는 포괄적인 규정을 두고 있는 반면, 그에 대한 보호조치들은 규정이 매우 미비하다. 따라서 현재의 상태에서 통계로 승인하는 절차는 (i) 개인정보의 수집·제공을 허용하고, 심지어는 개인정보 제공에 대한 법적 의무를 부과하며, (ii) 개인정보의 제3자 제공과 활용이 법적으로 의무가 되거나 허용하고, (iii) 개인정보의 보호가 통계법의 수준에 머무르게 되는 것을 허용하는 것을 의미한다.

이런 점을 고려한다면 통계로 승인하는 절차는 매우 엄격하거나 영향평가의 대상이 되거나 법령의 체계를 갖춘 것이어야 바람직할 것이다.

정책 58.작은 결혼식 59. 이웃간 교류 활성화
[Ⅷ. 가족 돌봄] [영유아 자녀 돌봄] 61. 영유아 자녀 다니는 기관 등 62. 긴급양육 도움자 [초등학생 돌봄] 64. 방과 후 시간을 보내는 장소 65. 자녀돌봄서비스 필요 시간 [함께 사는 가족 돌봄] 66. 함께 사는 가족원 돌봄
[Ⅸ. 가구 특성] [가족형태] 67. 가족형태 68. 가정 건강 [주택특성] 69. 주택 유형 70. 점유형태 [경제상태] 71. 가족 소득 72. 가구 지출 73. 경제 상태

316) 지정통계는 ① 전국을 대상으로 작성하는 통계 ② 지역발전을 위한 정책수립 및 평가의 기초자료가 되는 통계 ③ 다른 통계의 모집단자료로 활용 가능한 통계 ④ 국제연합 등 국제기구에서 권고하는 통일된 기준 및 작성방법에 따라 작성하는 통계 ⑤ 그 밖에 지정통계로 지정할 필요가 있다고 통계청장이 인정하는 통계가 그 대상이다(「통계법」 제17조 제1항).

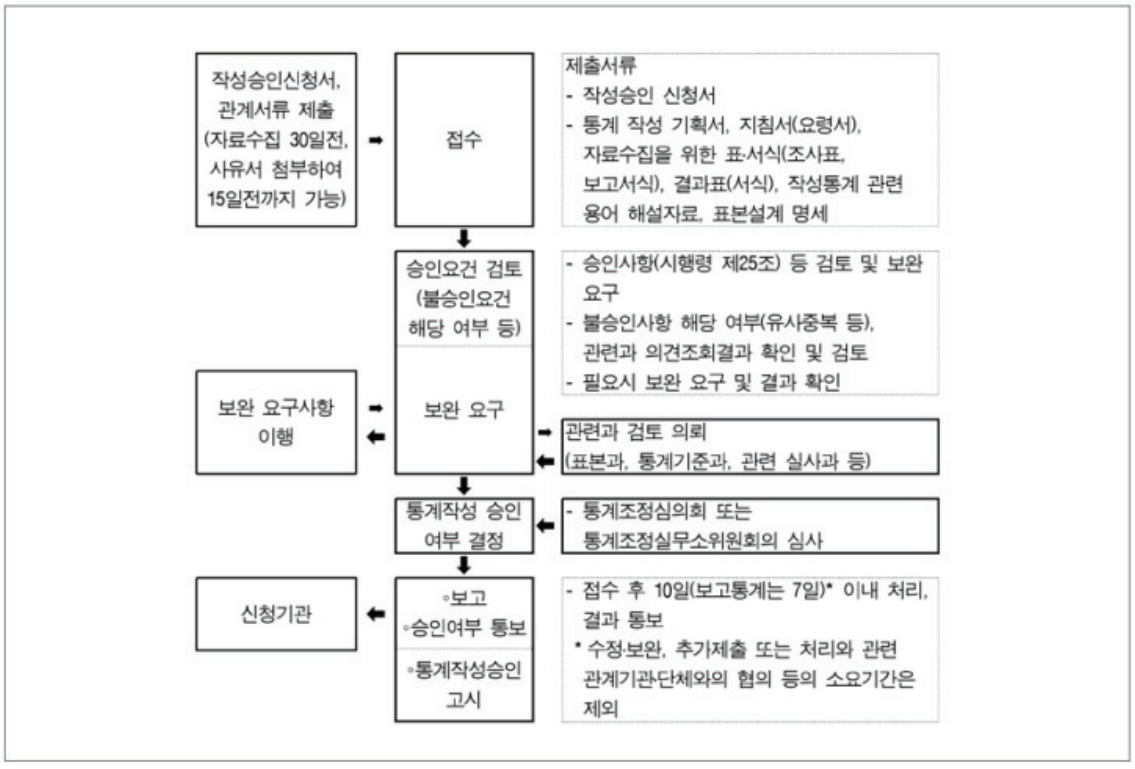
317) 통계청장에 의해 지정통계로 지정, 고시되면, 관련 개인이나 법인, 단체 등에 지정통계의 작성을 위해 필요하다고 인정되는 관계 자료의 제출을 통계청장을 통해서 요구할 수 있고, 통계 작성을 위한 조사 또는 확인이 필요한 경우에는 관계 자료의 제출을 요구할 수 있으며, 이 경우 자료제출을 요구받거나 질문을 받은 자는 정당한 사유가 없는 한 이에 응해야 한다(제25조, 제26조). 만약 응하지 않을 경우에는 100만원 이하의 과태료가 부과될 수 있다.

나. 통계 승인 제도의 목적

통계법은 통계를 작성하기 위해서는 통계청장으로부터 통계 작성에 대한 승인을 받거나 협의(다른 법률에 따라 통계를 작성하는 경우)를 하도록 하고 있다. 통계로 승인되면 아래에서 보는 바와 같은 효과를 누리기 때문에 이는 개인정보 제공에 대한 의무, 수집 및 이용의 근거, 통계법 적용의 대상이 되게 하는 것이므로 개인정보 보호법의 원칙에 따른 심사가 반드시 필요하다.

그런데 통계청은 통계 작성 승인제도의 목적을 통계의 신뢰성을 확보하고 통계의 중복작성으로 인한 예산과 인력의 낭비 및 자료 이용 상의 혼란을 방지하는데 주안점을 두고 있다. 따라서 통계 승인의 기준은 통계의 신뢰성, 통계의 중복 방지에 맞추고 있다.

그림 4-36 통계작성 승인(협의) 절차



* 출처: 통계교육원 (2015), "국가통계의 이해", p100.

다. 통계 작성 승인 기관 - 통계조정심의위원회

통계 작성 승인은 통계조정심의위원회에서 이루어지는데, 위원장은 통계정책국장으로 하며 위원은 통계정책과장, 통계조정과장, 품질관리과장, 표본과장, 해당 통계조정 업무와 관련 있다고 위원장이 지정하는 담당과장, 민간 전문가 2명으로 한다. 민간 전

문가는 통계 또는 관련 분야에 전문지식을 갖춘 사람으로 위원장이 위촉하고 임기는 2년으로 한다.[318] 한편 통계조정심의위원회에서는 지정통계의 지정과 지정취소도 담당한다.

라. 승인의 심사기준과 절차

실제로 통계작성기관이 수집하는 통계자료는 통계청장의 승인만 받으면 되는데,[319] 통계청장이 승인을 거부할 수 있는 사유로는 첫째, 이미 승인을 받은 다른 통계와 조사 또는 보고의 대상·목적 및 방법 등 그 내용이 동일 또는 유사하다고 인정되는 경우, 둘째, 표본규모가 지나치게 작거나 검증된 통계작성기법을 사용하지 아니하여 통계의 신뢰성을 확보할 수 없다고 인정되는 경우, 셋째, 조사 또는 보고의 대상 또는 목적 등이 특정 이익집단 또는 특정 부문에 편중되거나 영리적인 목적으로 작성되는 등 공공의 이익을 목적으로 작성된다고 보기 어려운 경우는 승인하지 않을 수 있다고 하여 신뢰성과 중복, 공공의 이익만을 규정하고 있다.

통계승인업무 처리지침은 주요 심사기준을 제시하고 있는데 통계의 정확성, 신뢰성, 중복 여부를 심사기준으로 제시하고 있다.

< 통계작성 승인의 세부적인 심사사항>
- 통계법(3조)에서 규정한 통계의 정의 및 범위에 해당되는지 여부
- 기존 승인통계와 유사·중복이 아닌지 여부
- 통계청장이 작성·고시한 표준분류를 사용하는지 여부
- 통계의 신뢰성에 문제가 없는지 여부
  - 표본조사인 경우 표본설계 내역과 추정방법에 대해 중점 심사
  - 보고통계는 기초자료의 대표성을 중점 심사
  - 가공통계는 기초자료의 신뢰성 및 가공기법 등을 심사
- 통계의 공표 및 관리 책임성 여부

마. 심사기준과 개인정보 보호

통계 승인 시에 개인정보의 보호에 관한 문제는 심사기준에도 없고 통계조정심의위

318) 통계승인업무 처리지침(제193호, 2013. 4. 2.) 제16조. 운영지침은 전문성 강화를 위해 심의위원회는 “통계조정 전문가풀”을 운영하는데, 표본설계, 통계기준, 조사방법, 현장조사, 통계생산에 전문성을 가지고 통계조정 관련 컨설팅을 할 수 있는 3명 내외의 청내 전문가로 구성한다. 전문가풀의 전문가는 심의회 및 통계조정실무소위원회에 참석하여 의견을 개진할 수 있다(제20조).

319) 통계작성기관의 장은 통계 작성을 위한 조사·보고 등 자료 수집을 시작하기 전에 ① 통계의 명칭 및 종류 ② 통계의 작성 목적 ③ 통계작성의 사항(작성의 사항이 자연인이면 성별 구분을 포함). ④ 통계작성의 대상(성별 구분을 포함). ⑤ 통계작성의 기준시점·기간 및 주기. ⑥ 통계작성의 방법 ⑦ 자료수집체계 ⑧ 통계작성에 사용하려는 분류 또는 기준(통계법 제22조제2항에 따른 표준분류 또는 미리 통계청장의 동의를 받은 다른 기준을 말한다) ⑨ 조사표, 보고서식 및 통계표 등 통계의 작성이나 공표와 관련된 서식.

원회의 위원에도 해당 분야 전문가가 없으며 그에 대한 의견조회 절차도 없다.

바. 여성가족부의 의견 반영 제도

반면, 통계 승인 시에 여성가족부의 검토 의견을 받아서 반영하도록 하는 규정은 있다. 국가나 지방자치단체의 정책수립과 시행에서 성 평등을 실현하는 것을 목적으로 성별영향분석평가를 실시하도록 하는 성별영향평가분석법이 제정되었는데 분석평가를 할 때 성별통계를 고려하도록 하고 있다.[320] 그리고 여성발전기본법은 국가와 지방자치단체가 인적(人的) 통계를 작성하는 경우에는 성별을 주요 분석 단위에 포함시켜야 한다는 규정을 두고 있다.[321]

이런 규정에 기초하여 통계청은 통계작성기관이 성별 구분 예외통계로 요청하는 경우에 이를 승인하는 절차를 별도로 규정하고 있다.[322] 그 기준은 작성 가능성과 정책 중요도인데, 작성 가능성은 통계청이 판단하고 정책 중요도는 여성가족부의 의견을 반영한다. 이를 위해 여성가족부에 요청통계의 성별 구분이 정책에 중요한지 여부에 대한 의견을 요청하여야 한다. 통계청은 통계의 작성 가능성에 대한 판단 기준도 상세하게 제정해서 운용하고 있다.[323]

사. 제도 개선 방안

통계 작성 승인 시 자료수집의 범위, 대상, 목적 등을 승인하게 되는데, 이때 개인

---

320) 성별영향평가분석법 제6조

321) 여성발전기본법 제13조 제3항

322) 통계승인업무 처리지침 제11조 제3항. 예외통계 승인 절차는 다음과 같다.
1. 성별구분 예외통계로 승인을 받고자 하는 통계작성기관은 승인 신청 시 그 사유를 소명하여야 하고, 통계청은 필요시 추가 자료를 요청할 수 있다.
2. 통계작성기관의 요청이 있는 경우 여성가족부에 요청통계의 성별구분이 정책에 중요한지 여부에 대한 의견을 요청하여야 한다.
3. 통계의 성별구분 작성 가능성이 낮거나 정책 중요도가 낮은 통계는 작성예외로 승인하되, 작성 가능성이 낮고 정책 중요도가 높은 통계는 중장기적 개선의견을 권고한다.
4. 통계 작성 및 변경 승인 시 성별구분 통계작성 예외를 요청하는 경우 검토 및 협의기간을 고려하여 다른 승인사항의 결정과 구분하여 진행할 수 있다. 이 경우 성별구분의 예외는 결정되지 않은 것으로 본다.

323) 통계승인업무 처리지침 제11조 제2항 작성 가능성 기준은 다음과 같다.
1. 자료수집 도구의 부적합성
- 개념, 정의, 방법이 편향되어 성별 구분이 부적합한 경우
- 조사모집단의 성별 왜도가 커서 적합한 통계 생산이 어려울 경우 등
2. 현실적 곤란
- 불특정 다수 등을 대상으로 하는 등 통제가 불가능한 경우
- 성별 구분의 응답가중으로 인해 무응답이 현저히 높아질 경우 등
3. 법령상 곤란
- 개별법령에 서식이 지정되어 구분 작성이 불가능한 경우
4. 구분 부적합
- 성별차이와 남녀의 역할과 기여를 나타내는 사항과 무관하여 조사의 실익이 없는 경우

정보 보호 원칙을 준수하고 있는지가 검토의 기준 중 하나가 되어야 한다. 그러나 현재 통계청은 통계작성 승인제도를 통계의 신뢰성 확보, 중복 방지의 관점에만 주안점을 두고 운영하고 있다. 앞서 본 것처럼 개인정보 보호 원칙에 부합하는지 여부도 중요한 판단 기준으로 명시하고 구체적인 판단요소를 마련해 놓는 것이 바람직하다.

현재 통계 승인이나 지정통계의 지정을 결정하는 통계조정심의위원회는 통계청 내부의 조직으로 독립성과 개인정보 보호에 관한 전문성을 갖추고 있다고 보기 어렵다. 앞에서 성 평등을 위한 성별통계와 관련하여 여성가족부장관에게 의견조회를 의무화한 것처럼 개인정보 보호위원회의 의견조회를 의무화하거나 의견제시를 할 수 있는 제도적인 방안을 마련하는 것이 좋다.

(3) 통계 작성과정에 대한 통계법의 규정 검토

통계법에 따라서 공공기관이 처리하는 개인정보를 수집하는 경우에는 개인정보 보호법 제3장 ~ 제7장의 규정을 배제하고 있다. 따라서 이 경우는 통계법만이 적용된다.

통계법은 통계작성기관에서 통계 작성을 위한 개인정보를 비롯한 자료를 수집하는 것에 대하여 여러 가지 권한을 부여하고 있는데 데이터 연계를 할 수 있는 근거 규정이 되기도 한다. 크게 자료수집의 방법으로는 ① 공공기관의 행정자료 이용, ② 형사사법정보 이용, ③ 자료제출명령, ④ 직접 조사를 규정하고 있다. 통계 작성을 할 때 데이터 연계를 하거나, 통계자료를 데이터 연계를 통해서 활용하는 경우가 있다. 데이터 연계는 비용절감, 정확성 향상, 분석의 깊이에 도움을 준다.

가. 행정자료, 형사사법정보의 활용

첫째, 통계청은 통계작성기관이 통계를 작성하기 위해 공공기관의 행정자료를 활용할 수 있다고 규정하고 있다. 행정자료는 공공기관이 직무상 작성·취득하여 관리하고 있는 문서·대장 및 도면과 데이터베이스 등 전산자료로 통계자료(통계작성기관이 통계의 작성을 위하여 수집·취득 또는 사용한 자료(데이터베이스 등 전산자료를 포함한다)를 제외한 것을 말한다. 특히 통계법은 통계작성기관이 통계 작성을 위해 관계 자료가 필요할 때 먼저 공공기관의 행정자료로 그 목적 달성이 가능한지를 판단하도록 하여, 행정자료로 가능한 경우는 공공기관에 행정자료를 요청하여 제공받아서 통계를 작성하도록 한다(제24조).

통계작성기관은 여러 공공기관으로부터 행정자료를 받아서 데이터 연계를 하는 것도 가능하고 조사자료와 행정자료를 연계하는 것도 가능할 것이다. 공공기관의 장은

행정자료의 제공을 요청받은 때는 행정자료를 제공해야 하는데, 제공을 거부할 수 있는 사유로는 국가기밀, 개인과 기업의 중대한 비밀의 침해 등인데 구체적으로는 아래와 같은 사유이다.

1. 국가안전보장・국방・통일・외교관계 등에 중대한 영향을 미치는 국가기밀에 관한 행정자료로서 통계의 작성을 위하여 제공되면 국가의 중대한 이익을 현저히 침해할 우려가 있다고 인정할만한 상당한 이유가 있는 경우
2. 진행 중인 재판에 관련되거나 범죄의 예방, 수사, 공소의 제기 및 유지, 형의 집행, 교정, 보안처분에 관한 행정자료로서 통계의 작성을 위하여 제공되면 직무수행을 현저히 곤란하게 하거나 형사피고인의 공정한 재판을 받을 권리를 침해한다고 인정할 만한 상당한 이유가 있는 경우
3. 개인이나 기업의 신제품 개발, 신기술 연구 또는 상당한 노력으로 비밀로 유지・관리되고 있는 생산방법이나 판매방법에 관한 행정자료로서 통계의 작성을 위하여 제공되면 개인이나 기업의 중대한 영업상의 비밀을 현저히 침해할 우려가 있다고 인정할 만한 상당한 이유가 있는 경우
4. 개인의 정치적, 종교적 또는 성적 성향이나 생활에 관한 행정자료로서 통계의 작성을 위하여 제공되면 개인의 생명이나 신체, 재산의 보호에 현저한 지장을 줄 우려가 있다고 인정할 만한 상당한 이유가 있는 경우

행정자료를 제공받아서 데이터 연계를 하여 통계를 작성하는 경우 제공기관의 장은 요청기관의 장에게 행정자료에 포함되어 있는 개인이나 법인 또는 단체 등의 정보를 보호하기 위하여 사용방법·사용부서나 그 밖에 필요한 사항에 대하여 제한을 하거나 행정자료의 안전성 확보를 위하여 필요한 조치(정보보호조치)를 강구하도록 요청할 수 있다. 이와 같이 공공기관으로부터 제공받은 행정자료는 이를 통계 작성 외의 목적으로 사용하거나 다른 자에게 제공하여서는 아니 된다.

그리고 행정자료 제공기관의 장은 요청기관의 장이 제3항에 따라 요청한 정보보호조치를 하지 아니하거나 제4항을 위반하는 경우에는 행정자료의 제공을 중지 또는 제한할 수 있다.

둘째, 행정정보뿐만 아니라 형사사법정보도 활용할 수 있다(제24조의 2). 여기에는 가족관계등록전산자료, 사망원인통계 등이 있다.

나. 자료제출명령, 직접 조사

행정자료로 불가능한 경우에는 개인이나 법인 또는 단체 등에 관계 자료의 제출을 명할 수 있다. 이때 자료제출명령을 받는 자의 성명 또는 명칭 및 주소, 지정통계의 명칭과 통계작성승인번호 또는 통계작성협의번호, 제출할 자료의 내용, 자료제출의 기한과 방법, 자료를 제출하여야 할 중앙행정기관 또는 지방자치단체의 명칭, 담당부서 및 담당자, 자료제출명령에 응하지 아니하는 경우 법 제41조에 따른 과태료처분 안내를 자료제출명령에 포함시켜야 한다(통계법 시행령 제40조). 그런데 이 경우 개인정

보 보호법의 규정이 중첩 적용될지, 아니면 적용이 되지 않을지는 논란이 있을 수 있다. 개인정보 보호법은 그 대상을 '공공기관이 처리하는 개인정보를 통계법에 의하여 수집하는 경우'로 한정하고 있으므로 공공기관이 처리하는 개인정보가 아니라 개인이나 법인 또는 단체가 보유하고 있는 개인정보라면 개인정보 보호법의 적용이 배제되지 않고 중첩 적용된다고 볼 소지가 있다.

통계법은 통계응답자에게 직접 응답을 들어서 수집하는 권한도 부여하고 있다. 나아가 통계법은 통계응답자에게 성실응답의무를 부과하고 있다. 즉, 통계응답자는 통계의 작성에 관한 사무에 종사하는 자로부터 통계의 작성을 목적으로 질문 또는 자료제출 등의 요구를 받은 때에는 신뢰성 있는 통계가 작성될 수 있도록 조사사항에 대하여 성실하게 응답하여야 한다는 의무를 두고 있다. (제32조) 이는 법령상의 의무이므로 개인정보 수집에 대한 동의가 필요하지 않다. 그러나 개인정보 보호법의 규정은 중첩 적용될 것이다. 이런 경우에 수집한 개인정보를 데이터 연계를 통해서 활용할 수 있다.

(4) 통계데이터 연계의 규율과 현황

가. 데이터 연계에 대한 규율

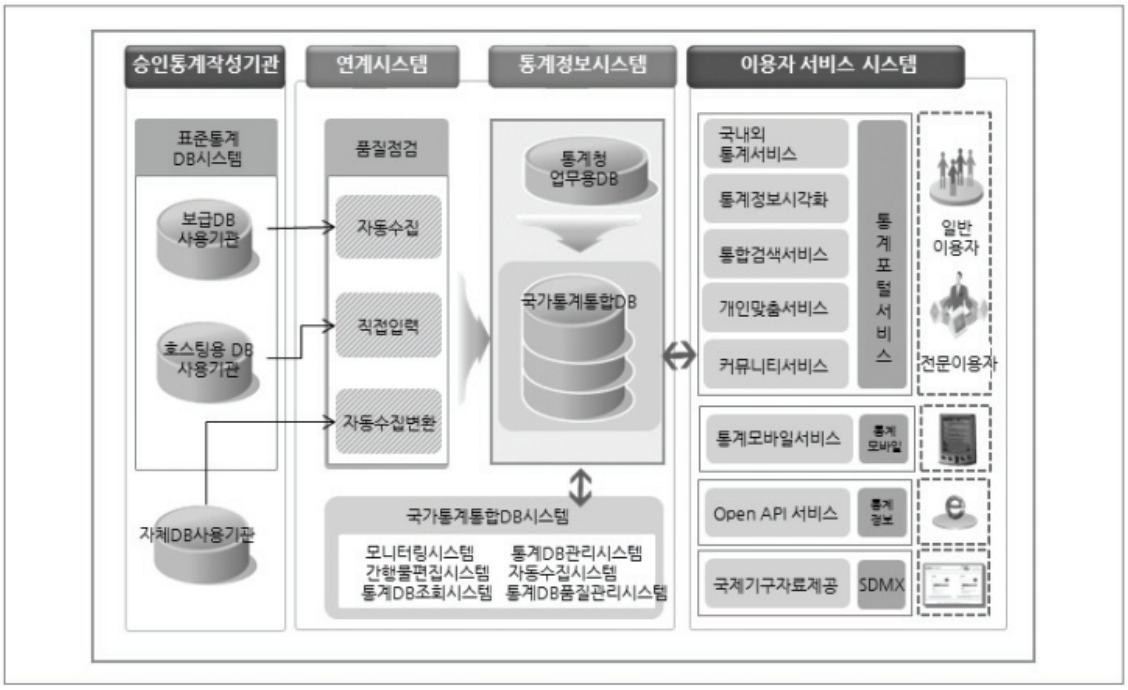
현재 통계법에서는 데이터 연계에 대해서 특별한 규정은 두고 있지 않다. 다만, 등록 센서스 자료수집 관련 행정자료 정보보호를 위한 운영규정 및 개인정보 보호지침을 제정·운영하고 있다고 한다.

그래서 행정자료 접근을 엄격히 제한하고 주민등록번호·외국인등록번호 등 개인식별번호는 복원 불가능한 가상번호로 대체 후 즉시 삭제하며 인터넷과 분리된 업무전용망에서 행정자료 통합관리 시스템을 통해 관리하고 있다고 한다. [324]

---

324) 인구주택총조사 규칙 제6조의3(행정자료 등의 안전관리) ① 통계청장은 제6조의2에 따른 자료에 「개인정보 보호법 시행령」 제19조에 따른 고유식별정보가 포함된 경우에는 복원 불가능한 가상번호를 부여하고 고유식별 정보는 즉시 삭제하여야 한다. ② 통계청장은 제6조의2에 따른 자료를 안전하게 입수하고 관리하기 위한 전산시스템을 운영하여야 한다.

그림 4-37 국가통계통합DB 시스템 구성도



* 출처: 통계교육원 (2015), "국가통계의 이해", p321.

나. 데이터 연계의 현황

① 행정통계 작성과 데이터 연계

행정통계를 작성하는 과정에서 행정자료 등의 데이터를 활용하는 경우가 많다. 단일의 행정자료만을 사용하기보다는 여러 행정자료를 연계하는 경우가 많다.

연계는 정확 매칭에 의한 방법이나 확률매칭에 의한 방법으로 이루어질 수 있다. 예를 들어 정확 매칭은 연계변수로 주민등록번호, 이름, 생년월일, 성별, 주소 등을 사용하는 것이고 확률매칭은 이름, 나이, 성별, 주소를 매칭 값으로 하되 동일주소 내 유사 발음은 동일인으로 간주하며 나이는 양력환산에 따른 오차범위를 ±3세로 하여 비교하는 것과 같다.

그림 4-38 행정자료를 활용한 통계생산 절차

* 출처: 통계교육원 (2015), "국가통계의 이해", p373.

실제로 '임금근로일자리 행정통계'라는 행정통계[325]를 작성하는 과정을 살펴보면 이 행정통계를 작성하기 위해서 9종의 행정자료를 활용했다고 한다.

표 4-45 임금근로일자리 행정통계

	행정자료명	보유기관
사회보험(4종)	건강보험	건강보험공단
	국민연금	국민연금공단
	고용보험	고용노동부
	산재보험	
과세자료(3종)	근로소득지급명세서	국세청
	법인세	
	부가가치세	
기업체 사업체 등록자료	사업자등록자료	국세청
	법인등기자료	대법원

9종의 행정자료는 연계를 통해서 통계자료로 수집되고 통계가 작성된다. 일자리통계의 경우는 종사자와 기업체 정보를 연계하는 작업 절차를 거쳤다고 한다. 즉, 통계작성 모집단 및 항목 생성의 단계에서 일자리 및 사업자 DB를 구축하고 인적 항목을

325) 일자리통계수요가 증가하고, 기업체 종사자를 연계한 통계가 미흡하고, 조사환경이 악화되었고, 조사비용과 응답 부담을 고려하여 행정자료를 활용한 일자리 통계가 기획, 개발되었다고 한다.

생성하고 기업체 항목을 생성한 후 기업체 항목과 인적 항목을 통합하는 과정을 거쳤다고 한다.

또 다른 예로 사망원인통계를 작성하는 과정을 예로 들어보면, 다음과 같은 행정자료를 연계하여 작성된다고 한다, 사망원인통계는 우리 국민의 정확한 사망원인 구조를 파악하여 국민복지 및 보건의료 정책수립을 위한 기초자료 제공을 위해 공표되는 것이다. 사망원인통계는 전국 시군구 및 읍면동에 신고된 사망신고서 및 의료기관에서 발생한 영아, 모성, 태아 사망 건에 대한 조사자료를 연계하여 작성된다. 의료기관의 정보를 반영하기 위해서 건강보험공단의 건강보험자료를 취득하여 질환 및 질병, 병태를 파악한다. 또한 외부의 요인을 파악하기 위해 근로복지공단의 산재보험자료를 통해 운수사고 등에 대한 정보를 연계한다. 이를 통해 사망원인을 파악하고 통계를 작성한다. 사망원인통계는 사망자에 대하여 조사되는 자료로 개인정보 보호법에 해당되지 않아 주민등록번호를 키변수로 이용하여 연계를 진행한다고 한다.

② 조사통계 작성과정에서 데이터 연계 - 의료패널의 사례

조사통계의 작성과정에서도 데이터 연계를 활용한다. 조사통계 중 대표적인 것이 패널 조사[326]이다. 패널 조사에서 조사자료와 행정자료의 연계가 이루어지는 경우가 많다. 예를 들어 의료패널도 국민건강보험공단의 건강보험자료와 연계를 한다. 의료패널은 국민 개개인 및 가구 단위의 의료비 규모 산출, 의료비 재원 분석, 보건의료서비스 수요자의 이용행태 분석, 보건의료정책수립을 위한 보건복지관련 지표 생산, 건강보험급여자료와의 연계를 통한 의료비 데이터생산의 완전성 구축, 국민 의료비 산출 및 변화 양상 추적, 주기적·종단적 데이터를 분석하여 의료비의 흐름(인과관계) 분석 등을 목적으로 하는 대표적인 조사통계이다.

의료패널의 기획 단계부터 조사결과를 국민건강보험공단의 자격·급여 자료를 연계하여 신뢰성 있는 데이터의 생산을 계획하였고 조사자료와 국민건강보험공단과의 자료 연계가 이루어지면 한국의료패널 데이터의 신뢰도가 더욱 제고될 것으로 기대하고 폭넓고 심도 있는 분석이 한층 가능해질 것으로 보았다.[327]

---

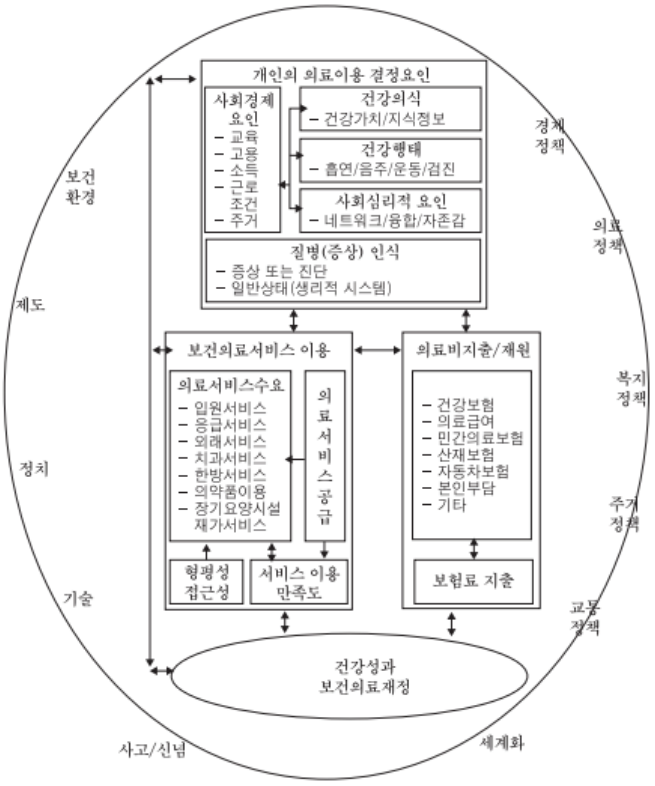
326) 패널조사는 일정 시점을 기준으로 한 횡단면조사의 단점을 보완하고 장기간에 걸친 동적 변화를 포착하는 데 유용한 자료이다. 우리나라는 다양한 패널조사가 있는데, 대표적으로 활용되는 분야가 복지 패널이다. 대표적인 복지패널은 다음과 같다.
한국아동패널/한국복지패널/한국노동패널조사/한국교육고용패널조사/한국아동·청소년패널조사/한국청소년패널조사/한국청년패널조사/고령화연구패널조사/국민노후보장패널/가계금융·복지조사/장애인고용패널조사/여성가족패널조사/여성관리자패널조사/한국의료패널/재정패널조사/사업체패널조사/인적자본기업패널조사.

327) 정영호·고숙자·이용갑·서남규·태윤희·이원영·이경용·김범수·강영호(2009), “한국의료패널의 활용과 기대효과”, 한국보건사회연구원·국민건강보험공단, p40.

의료패널의 경우, 해당 설문은 응급, 입원, 외래서비스 이용 및 본인 부담, 재원 등이 포함되어 있으며, 부가조사내용에는 흡연, 음주, 신체활동, 정신건강, 삶의 질 등이 포함되어 있다. 조사원이 직접 방문하여 질문하고 응답을 기록하는 면접타계식(face-to-face interview)이다. 둘째, 응답자의 기억이 쉽게 상기되도록 하기 위해 기억보조장치의 일환으로 가구원들의 의료이용 및 의료비에 관한 건강가계부를 작성하도록 한다.

이 통계를 작성하는 과정에서 개인식별정보로 주민등록번호를 수집하고 있는데 이를 통해서 행정데이터나 의료데이터와 연계를 하게 된다.[328]

그림 4-39 한국의료패널의 개념적 틀



* 출처: 정영호 외, 앞의 논문, p28.

한국의료패널은 아예 그 구축 목적의 하나로 건강보험자료와의 연계를 두었다. 건강보험자료와의 연계는 정확 연계와 확률연계가 있을 수 있는데, 주민등록번호, 건강보험증 번호를 활용하여 정확 연계가 가능하나 모든 가구원으로부터 개인식별번호 수집은 쉽지 않으므로 확률연계도 활용되어야 한다고 한다.[329] 패널조사 자료의 경우

328) 한국의료패널 조사-건강보험DB 연계지침서(Ver 1.0).

329) 도세록·김진수·강성홍·고혜연·신은숙(2012), "의료이용 통계생산 개선에 관한 연구", 한국보건사회연구원, P124.

개인식별자를 이용하여 다양한 2차 자료원(건강보험자료, 사망신고자료, 암 등록자료, 출생신고자료 등)과의 연계 연구의 가능성이 있으므로 연계를 위한 윤리적 고려사항에 대한 준비도 필요할 것으로 보인다(자료 연계에 대한 동의서, 자료 연계 연구에 대한 연구윤리위원회의 승인, 동의서에 연구대상자의 연구 참여 철회 권한에 대한 명기).[330]

이 외에도 통계청은 조사자료와 행정자료의 연계를 적극적으로 추진하고 있다. 2017년 기본계획에서도 국세 자료 및 저소득층 소득자료 등 행정자료와 조사자료 연계 활용 기준 마련 및 행정자료 이용에 따른 소득 관련 지표 분석·검토를 과제로 제시하고 있다. 연계 대상인 행정자료로는 국세청 근로소득·사업소득·금융소득 자료 및 보건복지부 기초연금·생계급여·주거급여·장애인연금 자료 등 약 20여 종을 들었다.

③ 전수조사통계와 데이터 연계

전수조사통계로는 인구총조사와 경제총조사가 가장 대표적이다. 전수조사통계에도 데이터 연계는 광범위하게 활용된다.

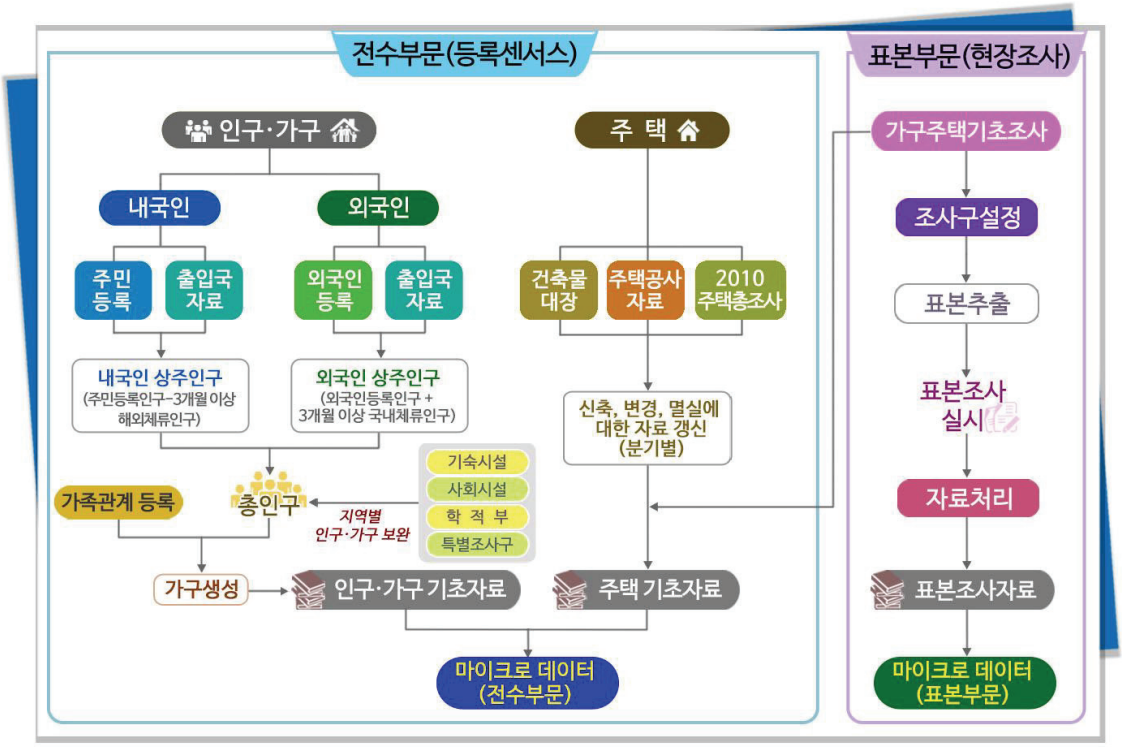
인구총조사는 인구 및 가구의 실태를 파악하기 위한 것으로 통계법에 근거를 두고 있다. 인구총조사는 그동안 약 90여 년간 '현장조사 방식'으로 실시해 오던 것을 행정자료를 기반으로 한 등록센서스로 변경하였다. 그래서 2015년에는 행자부 등 13개 정부기관 24종 행정자료를 융합한 등록센서스로 전환하여 행정자료를 활용한 후 20% 표본조사를 하는 방식으로 변경했다.

2015년의 인구총조사의 경우 인구·가구 및 주택에 관한 모집단 기초자료를 구축하기 위해 주민등록부(행자부), 가족관계등록부(대법원), 출입국자료(법무부), 외국인등록부(법무부), 건축물대장(국토부), 주택공시가격자료(국토부), 도로명주소자료(행자부)가 활용되었다. 이와 함께 자료 보완을 목적으로 사회시설 및 이용자명부(복지부), 학적부(교육부 및 각 대학), 군인명부(국방부), 전기시설명부(한국전력공사) 등 14종의 행정자료를 연계·활용했다. 행정자료는 11월 1일 0시를 기준으로 각 기관으로부터 매년 제공받고 있다.

이 과정에서도 데이터 연계를 이용하게 된다. 인구총조사와 주택총조사의 자료수집은 행정자료, 사법자료와 면접조사, 인터넷을 통한 조사자료를 연계하고 있다.

330) 정영호 외, 앞의 논문, p188.

그림 4-40 인구총조사 자료 연계



경제총조사는 매 5년마다 우리나라 전체 산업에 대한 고용, 생산, 투입(비용) 등에 관한 구조를 파악하기 위하여 동일시점에 통일된 조사기준으로 실시하는 대규모 전수조사이다. 산업총조사와 서비스업총조사를 통합한 것이다. 여기에도 국세청 사업자등록자료, 과세자료 등 8개 기관 20종 행정자료를 활용하여 행정자료를 연계하고 있다.

④ 민간분야 통계와의 데이터 연계를 통한 통계 작성

최근 공공데이터와 민간 빅데이터를 연계하여 통계를 작성하는 방법도 활용되고 있다. 통계청은 우선 시범적 사업으로 통계청 공공데이터인 통계조사 및 공공 행정자료와 민간기관 신용정보회사(코리아크레딧뷰로, KCB)의 빅데이터인 신용 자료를 연계·분석을 하여 통계를 작성하였다고 한다. 이 통계는 저출산 대책 등 지원을 위한 신혼부부에 관한 통계로 DB로 구축·분석하였다고 한다.

통계청은 이 밖에도 신용보증재단중앙회의 소상공인 개인별 자료 연계(소상공인 창폐업 정보 및 가구의 특성분석, 자영업자 보증포화지수 등 개발), 신용카드사의 카드승인액 자료 연계 및 물가(POS데이터) 자료 활용(개인별·가구별 소비현황, 카드승인액 지수 등 민생지표 작성, 물가정보, 현재 여신금융협회를 통해 전체카드사의 일단위

업종별·지역별 카드승인액 자료 입수 중이며 개별 카드사의 개인별 신용카드 사용자료 입수 협의 중 통신사의 위치정보와 공공데이터 연계) 등을 추진하고 있다고 한다.

(5) 통계자료의 보유 및 관리에 대한 규율

가. 법령의 규정

통계의 원천이 되는 자료를 통계법은 통계자료라고 칭하는데 학술적으로는 이를 마이크로데이터라고 부른다. 마이크로데이터란 통계조사 원자료(raw data)에서 조사 오류, 입력 오류 등을 수정한 개인, 가구, 사업체 등 특성에 관한 자료다. 마이크로데이터를 통해서는 이미 공표된 통계표뿐만 아니라 다양한 관점의 새로운 분석자료(종단, 횡단 시 계열자료 등)를 만들어 활용할 수 있다.

현재 통계법은 '통계작성기관의 장은 통계의 보급 및 이용의 활성화를 위하여 통계자료를 보유·관리하여야 한다'는 규정만을 두고 있고(통계법 제29조의 2), 시행령은 '전산 데이터베이스 등의 매체에 유실되지 않도록 보유·관리하여야 한다'는 규정만을 두고 있다(시행령 제45조의 2 제1항).

보유 및 관리의 운영기준은 각 통계작성부서에서 자체적으로 제정하게 되어 있는데, 그에 따라 제정된 각 부처의 통계관리규정(예규)에서는 예를 들어 보건복지부의 통계관리규정의 경우를 보아도 통계자료의 보유에 대한 특별한 규정은 두고 있지 않다(보건복지부 통계관리규정 제17조).

나. 제도개선

통계자료의 비밀보호를 위해서 통계자료의 보유와 관리에 대한 규율을 마련할 필요가 있다. 특히 현재 개인정보 보호법이 통계와 관련한 정보에 대한 개인정보 보호법의 적용을 배제하고 있기 때문에 더더욱 통계법에 그에 대한 규정을 두어야 한다.

통계자료의 목적 달성 후 폐기 의무 및 통계자료를 보유하는 동안에도 통계자료의 개인 식별자를 대체번호 등으로 치환하고 연계정보는 별도로 보관하면서 대체식별자로 처리된 통계자료를 활용하는 등의 안전조치가 바람직하다.

통계자료를 보유하는 동안 물리적, 조직적, 관리적 안전조치를 유지하도록 하고, 특히 통계의 기밀유지를 위한 여러 규정을 마련할 필요가 있다.

5. 통계자료의 활용

(1) 통계자료의 활용에 대한 규율

현행 통계법상 통계자료의 활용에 대한 규정은 통계 작성의 목적을 위해서 활용하는 경우와 기타의 경우로 나누어 볼 수 있다.

가. 통계의 작성을 위해 필요한 경우

① 통계법의 규율

통계의 작성을 위해서 필요한 경우는 앞에서도 본 것처럼 통계법 제30조에서 규율하고 있다.

> 제30조(통계자료의 제공) ①통계작성기관의 장은 통계의 작성을 위하여 필요한 경우에는 다른 통계작성기관에 통계자료의 제공을 요청할 수 있다. 이 경우 요청을 받은 통계작성기관의 장은 특별한 사유가 없는 한 이에 응하여야 한다.
> ②통계작성기관의 장은 다른 통계작성기관의 장에게 제1항에 따라 통계자료를 제공하는 때에는 특정의 개인이나 법인 또는 단체 등을 식별할 수 없는 형태로 통계자료를 처리한 후 제공하여야 한다. 다만, 다른 통계작성기관의 장이 다음 각 호의 어느 하나에 해당하는 조사의 표본으로 사용하기 위하여 제1항에 따른 요청을 하는 때에는 특정의 개인이나 법인 또는 단체 등이 식별되는 형태로 통계자료를 제공할 수 있다. <개정 2012.12.18.>
> 1. 제18조제1항 또는 제20조제1항에 따라 승인을 받거나 협의를 거친 통계의 작성을 위하여 실시하는 조사
> 2. 제18조제1항에 따른 승인을 신청하거나 제20조제1항에 따른 협의를 요청한 후 그 통계의 작성을 위하여 예비적으로 실시하는 조사
> ③제2항에 따라 통계작성기관으로부터 제공받은 통계자료는 이를 제공받은 목적 외의 목적으로 사용하거나 다른 자에게 제공하여서는 아니 된다.
> ④통계자료의 제공방법 등에 관하여 필요한 사항은 대통령령으로 정한다.

② 개인이나 법인 단체를 식별할 수 없는 형태로 제공한다는 의미

통계법에서 '개인이나 법인 단체를 식별할 수 없는 형태로 제공한다'는 것은 개인을 식별할 가능성이 완전히 제거된 것을 의미하지는 않는다. 제31조 제3항에서 언급하고 있는 것처럼 다른 자료와 연계되어 개인을 식별할 수도 있게 되기 때문이다. 따라서 이와 같은 방식으로 제공하는 것은 비밀유지가 되는 것으로 볼 수 없다.

③ 통계 작성을 위해 다른 통계자료의 활용에 대한 통제

통계법은 통계작성기관의 장이 통계작성을 위해 필요한 경우에는 다른 통계작성기

관의 장에게 통계자료의 제공을 요청할 수 있다고 규정하고 있는데 그에 대한 요건과 판단 기준에 대해서는 규정된 바가 없다.

통계자료의 비밀보호가 이루어질 수 있어야 통계주체의 신뢰를 얻을 수 있는 것이므로 통계법 제30조에서 통계 작성에 필요한 경우 다른 통계작성기관의 통계자료를 요청하여 활용할 수 있도록 하는 것은 남용될 경우 통계주체의 신뢰를 훼손할 것이므로 엄격한 제한이 필요하다.

제공을 허용할지의 판단 기준으로는 통계가 공익을 위하여 반드시 필요한 것인지, 통계자료 요청이 부득이한지, 통계단위의 통계비밀 유지에 대한 기대, 제공된 통계자료가 다른 정보와 결합하게 되는지 여부, 새로운 통계 작성을 위해 통계자료를 제공받는 것이 반드시 필요한지 등 개인정보 보호의 필요성과 이익형량 등을 종합적으로 고려해야 할 것이다.

나. 통계 작성 목적 외의 통계자료 활용에 대한 규율

① 통계법의 규정

현행 통계법은 통계 작성 목적 외의 통계자료 활용에 대해서 규정을 두고 있는데 통계의 비밀보호 원칙에 비추어 문제가 있다.

> 제31조(통계자료의 이용) ①통계자료를 이용하고자 하는 자는 대통령령으로 정하는 바에 따라 통계작성기관의 장에게 통계자료의 제공을 신청할 수 있다. <개정 2016.1.27.>
> ②통계작성기관의 장은 제1항에 따른 신청을 받은 때에는 통계자료의 사용목적·내용 및 범위의 타당성을 심사하여 타당하다고 판단되고, 영업상 비밀을 침해할 가능성이 없는 경우에는 이를 제공하여야 한다. 이 경우 통계작성기관의 장은 다음 각 호의 경우를 제외하고는 특정의 개인이나 법인 또는 단체 등을 식별할 수 없는 형태로 통계자료를 처리한 후 제공하여야 한다. <개정 2016.1.27.>
> 1. 통계응답자가 자신이 응답한 자료를 요구하는 경우
> 2. 총조사 및 제18조에 따라 통계청장의 승인을 받아 작성하는 통계 중 사업체를 대상으로 하는 전수조사를 통하여 취득한 정보로서 사업체 명, 업종, 주소 등 대통령령으로 정하는 정보를 제공하는 경우
> ③제2항에도 불구하고 통계작성기관의 장은 해당 통계자료를 다른 자료와 대응 또는 연계함으로써 다음 각 호의 어느 하나에 해당하는 경우에는 통계자료를 제공하지 아니할 수 있다. <개정 2016.1.27.>
> 1. 특정의 개인이나 법인 또는 단체 등의 식별이 가능하게 되는 경우
> 2. 사업체의 영업상 비밀을 침해하게 되는 경우
> ④제2항에 따라 통계작성기관으로부터 제공받은 통계자료는 이를 제공받은 목적 외의 목적으로 사용하거나 다른 자에게 제공하여서는 아니 된다.
> ⑤통계자료의 제공방법 등에 관하여 필요한 사항은 대통령령으로 정한다.

② 통계자료의 이용에 대한 규율과 통계의 비밀보호 원칙

통계법은 통계자료의 이용신청을 할 수 있는 자를 제한하지 않고 있다. 통계자료를

이용하고자 하는 자는 대통령령으로 정하는 바에 따라 통계작성기관의 장에게 통계자료의 사용 목적을 밝히고, 통계자료의 제공을 신청할 수 있다.[331]

통계작성기관의 장은 통계자료의 사용 목적·내용 및 범위의 타당성을 심사하여 타당하다고 판단되고 영업상 비밀을 침해할 가능성이 없는 경우에는 이를 제공하여야 한다고 규정하고 있다. 이때 특정 개인이나 법인을 식별할 수 없는 형태로 통계자료를 처리한 후 제공하여야 한다. 다만, 해당 통계자료를 다른 자료와 대응 또는 연계함으로써 특정의 개인이나 법인 또는 단체 등의 식별이 가능하게 되는 경우나 사업체의 영업상 비밀을 침해하게 되는 경우에는 통계자료를 제공하지 아니할 수 있다고 규정하고 있다. 통계작성기관의 장은 다른 정보와 결합하여 식별이 가능해지는 경우에도 통계자료를 제공하지 않을 수도 있고 제공할 수도 있으므로, 식별 가능성 여부는 통계작성기관의 장이 통계자료를 제공하지 않으려고 할 경우에 제공을 거부할 수 있는 논거가 될 뿐이다. 즉, 통계작성기관의 장은 누구라도 통계자료를 요청하는 경우 특정 개인을 식별할 수 없는 형태로 처리한 후 이를 제공할 수 있는 것이다. 공익적 목적이나 연구목적 등의 제한도 없이, 어느 누가 요청하더라도 제공 여부는 통계작성기관의 전적인 자유에 맡기고 있다

이런 태도는 통계는 통계 목적으로만 활용되어야 하며 개인 식별 가능성이 있는 경우에는 공개하거나 제3자에게 제공되어서는 안 된다는 통계의 비밀보호의 원칙에 대한 중대한 침해이다.

---

331) 통계법 시행령은 통계자료의 신청에 대한 규정을 두고 있다.
제47조(통계이용자의 통계자료 신청과 제공) ① 법 제31조에 따라 통계자료를 이용하려는 자(이하 "통계이용자"이라 한다)는 다음 각 호의 사항을 문서(전자문서를 포함한다)에 적어 신청하여야 한다. <개정 2016.7.26.>
1. 통계이용자의 이름(기관등인 경우에는 기관등의 명칭) 및 주소
2. 통계자료의 명칭
3. 통계자료의 사용목적
4. 통계자료의 내용 및 범위
5. 통계자료의 제공방법
6. 통계자료의 보호방법
② 통계작성기관의 장은 제1항에 따른 신청을 받은 날부터 30일 이내에 법 제31조제2항에 따른 심사의 결과에 따라 통계자료 제공 여부 등을 결정한 후 통계이용자에게 알려야 한다. 다만, 통계자료의 가공·처리 등 부득이한 사유로 30일 이내에 통계자료를 제공할 수 없으면 통계이용자와 협의하여 그 제공기간을 10일 이내의 범위에서 연장할 수 있다.
③ 법 제31조제2항2호에서 "사업체 명, 업종, 주소 등 대통령령으로 정하는 정보"란 사업체의 상호·업종·주소 및 전화번호를 말한다. <신설 2016.7.26.>
④ 통계이용자에게 통계자료를 제공하는 방법에 관하여는 제46조제3항을 준용한다. 이 경우 "요청기관의 장"은 "통계이용자"로, "제공기관"은 "통계작성기관"으로 본다. <개정 2016.7.26.>
⑤ 제2항에 따른 통계자료 제공에 드는 경비나 수수료는 통계이용자가 부담하는 것을 원칙으로 한다. <개정 2016.7.26.>

③ 특정 개인을 식별할 수 없는 형태의 의미

특정 개인을 식별할 수 없는 형태라는 것은 가명화나 범주화 등의 비식별 처리를 의미한다고 볼 것이다. 이렇게 처리하더라도 다른 정보와 결합하여 특정 개인을 식별할 가능성이 있다. 실제로 통계법 제31조 제3항 제1호는 특정 개인을 식별할 수 없는 형태로 통계자료를 처리했으나 다른 자료와 대응 또는 연계함으로써 특정 개인이나 법인이 식별 가능하게 되는 경우를 규정하고 있다.

④ 제도의 개선

통계자료를 아무런 기준도 없이 특정 개인이 식별될 수 있는 상태로 제3자에게 제공할 수 있도록 허용하는 것은 통계의 비밀보호 원칙에 대한 중대한 침해가 아닐 수 없다. 이 경우 통계단위의 신뢰는 얻을 수 없다.

따라서 통계자료가 특정 개인을 식별할 수 있는 형태로 제3자에게 제공되는 것은 엄격하게 금지해야 하고, 다만 예외적으로 이를 인정하려면 공공 연구의 목적으로 통계자료를 활용하는 경우로 제한하는 것이 바람직하다. 이 경우에도 해당 공공 연구는 국가기관이나 그에 준하는 신뢰성을 갖는 연구기관의 연구이어야 하고 연구계획서에 대한 기관윤리위원회 등의 심의를 거치도록 하는 것이 바람직하다. 그리고 해당 통계자료를 제공받을 수 있는 시설을 제한하고 통계자료의 유출이나 남용을 방지하기 위한 안전조치를 마련하는 것이 필요하다.

다. 국가통계 자료제공 규정

① 통계자료제공심의회의 독립성

한편 통계작성기관이 생산·보유하고 있는 통계자료를 통계법과 통계법 시행령[332]에 의해 외부에 제공함에 있어 필요한 사항을 정함을 목적으로 통계청 훈령인 국가통계 자료제공 규정이 제정되어 있다. 이 규정은 각 통계작성기관에 통계자료의 제공과 관련된 사항을 심의하기 위하여 통계자료제공심의회(이하 "심의회"라 한다)를 두도록 했다. 심의회는 통계자료의 제공에 대한 심의를 하는데 심의회의 구성과 운영의 독립성은 거의 보장되지 않고 있다. 규정은 심의회를 의장 1명, 위원 15명, 간사 1명으로 구성하도록 하고 있는데, 의장을 통계작성기관의 장이 위임한 부서장으로 하고 위원을 의장이 지정하도록 하고 있다. 규정은 단지 '심의회의 기능을 효율적인 운영을 위

332) 통계법 제27조, 제30조, 제31조, 제33조, 제37조와 같은 법 시행령 제42조, 제46조부터 제50조까지, 제52조.

하여 필요에 따라 내부 또는 외부 전문가를 추가로 위원 또는 자문위원으로 위촉할 수 있다'는 규정만을 두고 있다(제6조).[333] 규정은 심의회를 의장이 필요하다고 인정한 사항을 심의한다고 하고 의장이 회의를 소집한다고만 하고 있다.

② 통계자료의 비밀보호

한편 규정은 자료제공의 범위에 대해서 통계의 비밀보호를 실질적으로 보장하지 못하는 내용으로 규정하고 있다. 즉, 규정은 통계기초자료 및 미공표 집계자료를 '개인, 가구, 사업체, 법인 또는 단체 등 개별 자료가 실질적으로 식별되지 않도록 하여 제공'한다고 규정하고 있는데, '실질적으로 식별되지 않도록'이라는 매우 불충분한 수준의 비밀보호를 하고 있다(제10조). 더구나 이와 같은 통계자료의 요청은 연구목적이나 통계작성을 위한 목적으로 제한하고 있지도 않다.

③ 통계청의 통합 제공

규정은 통계기초자료(마이크로데이터) 등 공표외 자료는 통계청에서 통합하여 제공한다고 규정하고 있다(제16조).

④ 제도개선 사항

통계자료의 제공에 대한 심의는 독립성과 전문성이 보장될 수 있도록 구성하고 운영하는 것이 바람직할 것이다. 통계자료의 비밀보호와 관련해서는 '실질적으로 식별되지 않도록'이라는 기준은 매우 불충분하므로 자료제공 요청의 목적이나 제공 요청대상 정보의 내용에 따라서 세분화할 필요가 있다.

라. 통계의 보급을 위한 데이터베이스 자료제출 등

한편 통계법은 통계자료의 이용을 촉진하기 위한 규정을 두고 있는데, 통계청장이 통계데이터베이스의 구축 등을 할 수 있도록 하고, '통계청장은 통계데이터베이스의 구축·연계 및 통합 등을 위하여 필요한 경우에는 통계작성기관이 보유하는 데이터베이스자료 등 세부적인 통계 관련 자료를 제출하도록 요구할 수 있다. 이 경우 요구를

333) 심의회의 심의 안건은 아래와 같다(규정 제5조).
1. 공표외 자료의 보유·관리에 관한 사항
2. 공표외 자료의 제공방법 및 제한에 관한 사항
3. 통계자료의 제공 수수료에 관한 사항
4. 자료제공업무의 효율성 향상과 이용의 편의성 제고를 위한 제도개선 및 운영방법에 관한 사항

받은 통계작성기관의 장은 특별한 사유가 없는 한 이에 응하여야 한다'는 규정을 두고 있다(통계법 제28조 제3항). 이 규정이 통계자료를 통계청장이 제공받을 수 있는 규정으로 해석하는 경우, 이는 통계의 비밀보호에 위반하는 것으로 볼 수 있다.

(2) 통계청의 통계자료 제공 서비스

가. 서비스 개요

통계청은 통계청에서 작성하는 마이크로데이터뿐만 아니라 정부 각 부처, 지자체, 연구기관 등 타 통계작성기관의 마이크로데이터를 한곳에 모아 마이크로데이터 통합서비스라는 이름(영문명은 MDIS, MicroData Integrated Service)으로 제공하고 있다. 마이크로데이터 서비스는 1993년부터 시작되었는데 2014년부터는 통합서비스로 제공하고 있다. 서비스의 주요 연혁은 아래와 같다.

- 1993년 마이크로데이터 서비스 개시(오프라인)
- 2006년 MDIS 서비스 개시(온라인), 자료이용센터(RDC) 개소,
- 2008년 DW 기반 MDIS 서비스 개편,
- 2010년 원격접근서비스 개시,
- 2014년 마이크로데이터 통합서비스(MDIS) 구축 1단계,
- 2015년 2단계, 2016년 구축 3단계
- 마이크로데이터 통합서비스시스템 구축 및 서비스
- (마이크로데이터 통합DB) 사업체노동력조사(고용노동부) 등 15종을 포함하여 □년까지 150개 기관 279종 통계의 마이크로데이터 통합DB를 구축하고, 146종에 대한 서비스 실시

마이크로데이터를 통해서는 이미 공표된 통계표뿐만 아니라 다양한 관점의 새로운 분석자료(종단, 횡단 시계열 자료 등)를 만들어 활용할 수 있다. 예를 들면 한 나라의 소득 불평등도를 나타내는 지니계수의 경우 마이크로데이터를 활용해 소득분위별 지니계수를 작성하고 시계열 자료를 분석해 소득 불평등 완화정책 수립에 활용할 수 있다. 현재 통계청이 MDIS 홈페이지(mdis.kostat.go.kr)에서 제공하는 마이크로데이터는 77종이다(통계청이 제공하는 마이크로데이터 40종, 보건복지부 등 타 통계작성기관의 통계 37종). 마이크로데이터의 경우 제공되는 형태에 따라서는 데이터 연계나 데이터 결합이 가능한 경우도 있다.

표 4-46 연도별 통합DB 구축 및 서비스 현황 (단위:종)

구분	'14년	'15년	'16년	'17년
구축(누계)	71	72(143)	121(264)	15(279)
서비스(누계)	45	16(61)	30(91)	55(146)

* 마이크로데이터 이용 건수 : ('14) 9,483 → ('15) 14,398 → ('16) 21,885(9월말 기준)

나. 서비스 종류

마이크로데이터 통합서비스에서 제공되는 서비스는 아래와 같다.

표 4-47 마이크로데이터 통합서비스에서 제공되는 서비스의 종류

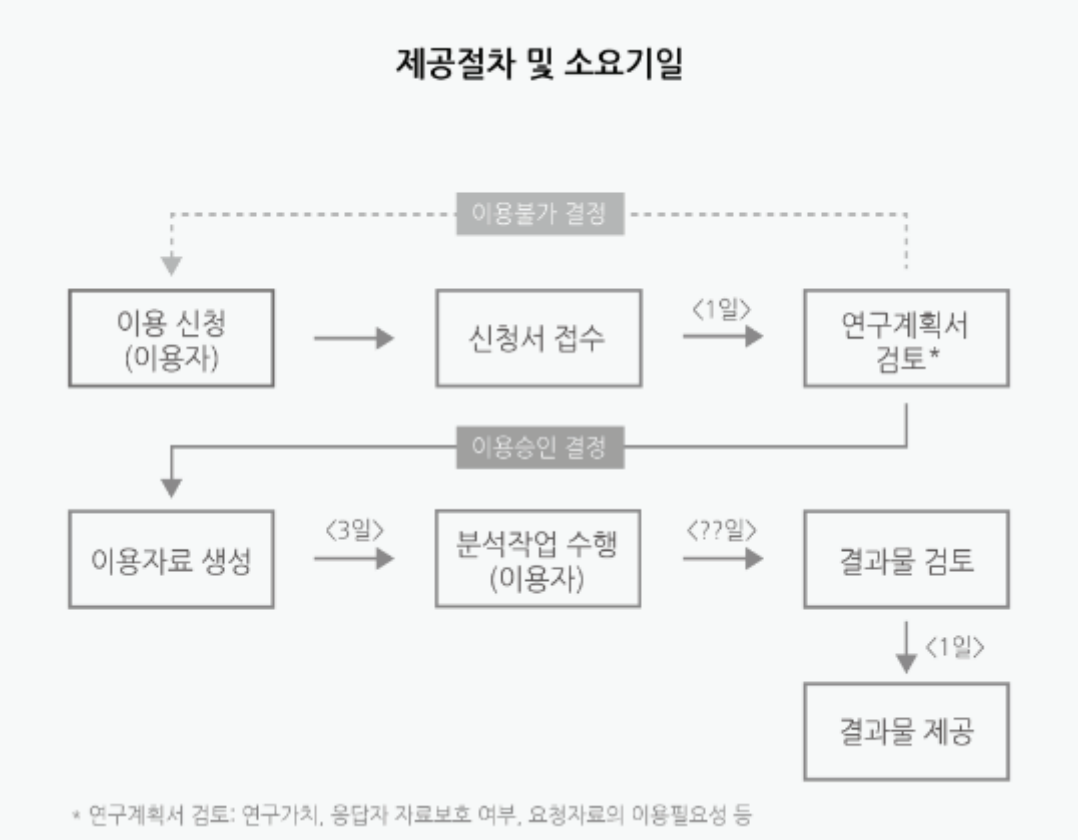
서비스	설 명
추출·다운로드	이용자가 원하는 공공용 마이크로데이터를 개인PC에 직접 다운로드 받아 분석 · 활용할 수 있는 방법
RAS(원격접근서비스)	이용자가 집, 사무실 등에서 직접 인터넷을 통해 승인용 마이크로데이터를 분석할 수 있는 서비스 승인용 마이크로데이터는 개인 PC에 다운로드 되지 않고, 가상의 서버에서 분석한 뒤 분석결과표만 통계청의 반출승인을 받아 개인PC에 다운로드 할 수 있음.
MDAC(마이크로데이터 이용센터)서비스	통계망이 차단된 지정된 장소에서 승인용 또는 특수목적용 마이크로데이터를 이용하여 심층적인 경제, 사회적 현상을 분석할 수 있도록 지원하는 서비스 마이크로데이터는 복사해 갈 수 없으며 분석결과표만 통계청의 반출승인을 받아 이용자에게 전달
주문형서비스	MDIS의 추출 및 다운로드에서 제공하는 공용 마이크로데이터를 이용하여 이용자가 원하는 집계표(분석결과표)를 작성해 주는 서비스
마이크로데이터 CD	공공용 마이크로데이터(모든 항목)를 CD로 제작하여 제공하는 방법

* MDAC 장소 : 통계청 본청(대전), 한국통계진흥원(성남), 한국개발연구원(세종), 경인청 나라셈도서관(서울 강남), 서강대학교(서울 강북)
* 특수목적용 이용센터 (RDC) •매스킹 되지 않은 자료 분석 •비밀, 민감정보 등 승인용 미제공 항목검토 후 제공

다. 제공절차

서비스의 제공절차는 아래와 같다. 신청서 접수 후 연구계획서를 검토하고, 이용자료를 생성하여 주고, 이를 분석하도록 하고, 결과물을 검토한 후 결과물을 제공하는 것이다. 통계청은 통계조사 응답자가 제3자에 의해 식별되거나 비밀정보가 노출되지 않도록 통계적 노출관리기법을 적용하여 마이크로데이터를 제공하고 있으며 자료제공 범위를 정하기 위하여 통계자료제공심의회를 운영하고 있다고 한다.

그림 4-41 마이크로데이터 통합서비스 제공절차



라. 이용실적

마이크로데이터 이용실적은 2010년 4,133건에서 2013년 9,253건으로 급증했다고 한다. 2011년 통계청에서 마이크로데이터 이용 전문가를 대상으로 조사한 결과, 자료연계 분석 경험이 36.7%였으며 향후 다양한 분야의 2차 자료를 요구하고 있는 것으로 나타났다고 한다(국가통계 마이크로데이터 서비스 개선방안, 2012. 3. 28. 제6차 국가통계위원회, 통계청장).

6. 통계청의 데이터 연계·융합 활성화 추진전략

(1) 통계청 2017년 사업계획

통계청은 2017년 사업계획을 통해서 데이터 연계를 활성화기 위한 방안을 제시하고 있다.

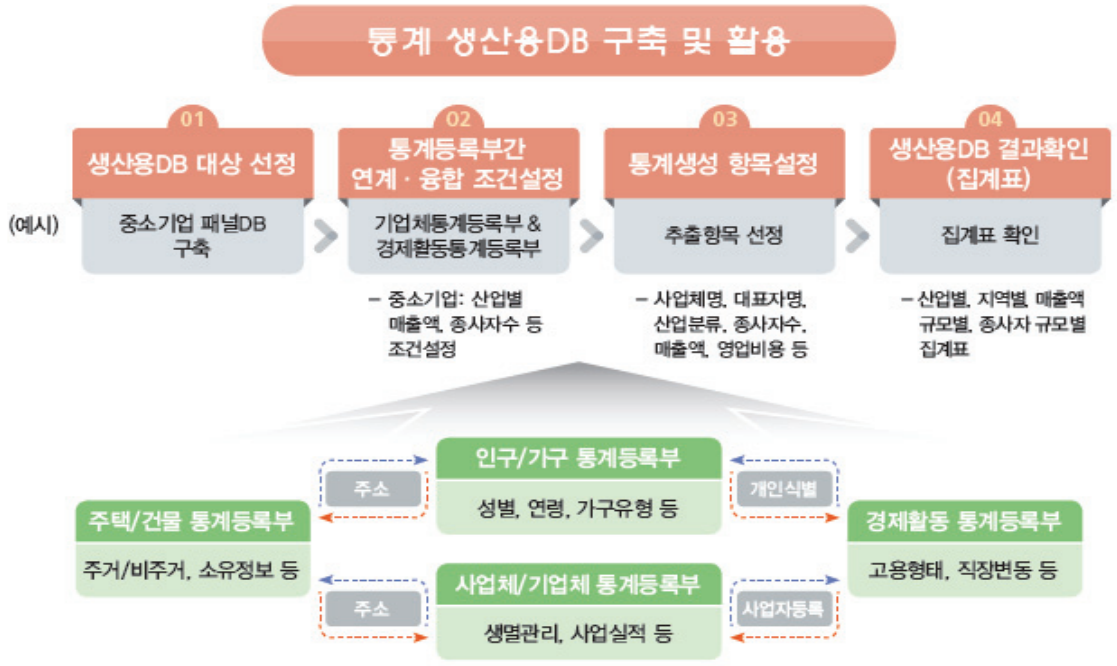
첫째, 데이터 공유 기반 강화로 데이터를 원활하게 공유 및 연계할 수 있도록 관련

제도 마련하겠다고 한다. 그 구체적 방안으로 행정자료 서식 개정 및 표준화로 데이터 형식을 통일하고 이종 데이터 간 연계를 지원하겠다는 것이다. 그리고 행정자료, 통신자료, 신용카드 등 데이터 원천별로 데이터를 축적하고 특성 및 활용사례 등을 정리하여 제공하여 데이터 간 연계 및 활용사례를 통해 데이터 공유를 위한 동력 창출하겠다는 것이다.

둘째, 데이터 활용 지원으로 개인정보 침해 없이 원활하게 데이터 간 연계를 지원하기 위한 데이터 인프라 확충하겠다고 한다. 보안과 처리 기술을 제공함으로써 개인정보 침해 없이 데이터 연계·활용을 지원하는 통계데이터센터 운영하겠다는 것이다. 특히 데이터 간 원활한 연계를 위한 기초 장치로서 통계고유번호를 도입하겠다고 한다. 이는 개인정보를 포함하지 않으면서도 데이터 간 연계가 가능하도록 마련하겠다는 것으로 주민등록번호 활용 연계 제약에 따라 통계 작성을 위한 새로운 비식별화 연계키라고 한다.

셋째, 범용 DB 구축이다. 다른 데이터와 연계·융합하여 다양한 통계를 생산할 수 있는 범용 DB 구축 추진하겠다는 것이다. 이를 위해 4개 분야별 통계등록부 구축을 마무리하고 등록부 간 연계로 종합등록부를 구축 추진하겠다는 것이다. 인구/가구, 사업체, 주택/건물, 경제활동 등 4개 분야의 통계 작성 명부를 위해 구축한다는 것이다. 이는 행정자료를 이용하여 작성한 명부 자료(Statistical Register)로서 조사통계에서의 모집단과 같은 기능을 수행하겠다고 한다. 통계등록부의 변동내역(개인의 경제활동상태 변화, 사업체의 생멸 등) 관리체계 구축 및 등록부 간 연계 Key 고도화, 연계된 등록부를 활용하여 수요자가 원하는 사회·경제 융합형 통계, 패널통계 등 다양한 통계생산용 DB를 구축·제공하겠다는 것이다.

그림 4-42 통계청, 데이터 연계·융합 활성화 전략



☐ (법령 개정) 행정자료 활용 및 보안 강화, 행정자료 표준화 등을 반영한 통계법 개정 검토·추진
* 기존 「통계법」은 행정자료의 제공 규정만을 두고 있어 행정자료 이용을 활성화하고, 관련 한계를 극복하기 위해 필요한 조항 반영

(2) 통계청의 민간 빅데이터 데이터 연계

가. 통계청의 데이터 연계 추진

통계청은 빅데이터를 활용한 통계 작성을 위해 통계데이터허브국에 15명의 정원으로 빅데이터 통계기획, 생산, 시스템구축의 업무를 하는 '빅데이터통계과'를 신설하고(2015. 10.), 공공데이터와 민간 빅데이터 간 연계로 새로운 가치를 창출하겠다는 포부를 밝히고 있다. 이를 통해서 사회가 필요로 하는 다양한 통계적 정보 작성 및 제공을 하겠다는 것이다.

그리고 새롭고 높은 부가가치 창출을 위해서는 개별 데이터 활용보다는 데이터 간 연계가 필요하다는 인식이 확산되어 연계를 위한 전제로서 비식별화 조치에 관해 미국·일본·유럽 등은 물론 국내도 관계부처 공동으로 가이드라인 작성 및 공표를 하고 있는데, 통계청은 그동안 공공데이터를 연계·활용한 통계 작성 경험을 바탕으로 공공데이터와 민간데이터 간 연계로 새로운 통계정보 작성을 추진하고 있다는 것이다.

통계청은 아래에서 보는 것처럼 KCB와 연계 분석을 하였다고 하는데, 그 외에도 신규 데이터 연계 시도를 하고 있다고 한다. 예를 들어 신용보증재단중앙회의 소상공인 개인별 자료와 연계하려고 한다고 한다. 이를 통해서 소상공인 창폐업 정보 및 가구의 특성분석, 자영업자 보증포화지수 등을 개발하고자 한다는 것이다.

그 외에도 신용카드사의 카드승인액 자료와 연계하는 것과 물가(POS데이터) 자료로 활용하려고 한다는 것이다. 이는 신용카드사의 정보와 연계하여 개인별·가구별 소비현황, 카드승인액 지수 등 민생지표를 작성하고 물가정보로 활용하려고 한다는 것이다. 이와 관련해서는 현재 여신금융협회를 통해 전체카드사의 일단위 업종별·지역별 카드승인액 자료를 입수 중이며 개별 카드사의 개인별 신용카드 사용자료 입수를 협의하고 있다고 한다.

또 다른 시도로는 통신사의 위치정보와 공공데이터를 연계하는 것도 추진한다고 한다. 그래서 출퇴근 거리·시간·인구수 등 교통시설 공급 정책에 필요한 정보를 확보하고 등록센서스 등의 실거주 여부 확인, 유동인구 등 확대된 개념의 인구 정보를 제공하고자 한다는 것이다.

나. 통계청과 코리아크레딧뷰로(KCB)의 연계 사업

① 협력사업의 개요

통계청은 이 사업을 공공데이터와 민간 빅데이터 간 연계·융합을 통한 활용사례 발굴 및 연계 DB 구축·분석 사업으로 KCB와 추진했다고 한다. 2016년 1월에 KCB와

‘통계청과 KCB 간의 빅데이터 기반 가계부채 관련 연구를 위한 업무협력 약정서’를 체결하여 사업을 추진했다고 한다. 연계 사업의 상대방인 KCB는 은행, 보험사 등이 출자하여 설립한 회사로 신용정보를 가공·분석하여 판매하는 민간회사이다.

약정서의 내용에 의하면 공공데이터와 민간 빅데이터를 연계하여 시의성 있는 가계부채 관련 연구를 목적으로 하여, 상호 협력의 대상은 ① 빅데이터 기반 부채 관련 분석 공동연구, ② 가계부채 관련 데이터 연계 방안 연구, ③ 가계부채 정보분석 및 활용을 위한 자료의 공동 이용, ④ 기타 양 기관의 발전에 필요한 사항으로 하였다.

양 기관은 협력 업무의 효율적 추진을 위해 업무 관련 실무자를 중심으로 협의체를 구성하여 운영하였고, 특히 약정서의 업무협력을 통해 취득한 정보를 상대방의 동의 없이 외부에 공개하거나 제공하는 것을 금지하였다. 약정서의 업무협력을 통해 취득한 정보의 비밀유지는 약정서의 유효기간 종료 이후에도 동일하게 적용된다고 하였다. 이에 따라 양 기관은 부채 정보분석 및 활용을 위해 보유 자료의 제공을 요청한 경우 관련법규에서 허용하는 범위에서 적극적으로 협조한다고 하였다.

② 연계 분석의 목적

통계청은 데이터 연계의 목적으로 신혼부부의 신용정보를 이용한 통계데이터의 생산을 들었다.

③ 연계에 대한 승인

통계청이 발표한 자료에서는 연계에 대한 승인 과정은 전혀 공개되지 않았다. 이는 지정통계가 아니기 때문에 특별한 승인 과정을 거치지 않은 것으로 보인다. 생산하고자 하는 통계의 목적, 공공성, 자료 설계의 적정성, 개인정보 처리의 적정성 등이 투명하고 공정하게 평가되었어야 한다.

통계청이 밝힌 바에 의하면 통계청과 KCB의 데이터 연계는 신혼부부 데이터와 신용정보를 연계한 것인데, 통계청은 이를 저출산 대책 등 지원을 위한 신혼부부에 관한 통계를 우선적으로 DB로 구축•분석한 것이라고 한다. 통계생산의 목적은 타당하다고 보이지만, 이 통계를 생산하는 과정에서 개인정보 침해를 최소화하기 위한 조치 등이 마련되어 있어야 하고, 절차가 통제되어야 한다.

연계하는 데이터는 인구주택총조사의 등록센서스 정보(개인과 가구와 관련된 다양한 정보), 인구동향정보의 혼인 관련 정보, 국적 정보, 경제활동 정보로 임금근로, 일자리통계, 4대 보험 정보, 근로소득 정보, 그 밖에 통계청의 각종 조사정보도 연계했

다고 한다. 이와 같은 정보들을 연계하는 것이 타당한지에 대한 평가가 이루어졌는지는 알 수 없다. 그러나 통계청에서 통계 작성을 위해서 데이터 연계를 하고자 할 때는 데이터 연계와 관련한 심사와 승인을 할 수 있는 독립적이고, 전문적인 기관이 필요하다.

④ 연계의 방법

본 연계에서는 통계청이 KCB라는 민간 업체들에 데이터를 제공하고 KCB가 데이터 연계를 한 후 연계키를 삭제한 후 데이터베이스를 양 기관이 공유하고 분석했다는 것이다. 이와 관련하여 통계청은 KCB가 데이터 연계를 하고 연계키를 삭제하였으므로 법적으로 문제가 없다고 주장하나 KCB에게 데이터를 제공한 것은 현행법률 상으로는 근거가 없는 것으로 보인다.

표 4-48 통계청과 KCB의 연계 방법

구분	관련 여건	보완 조치 내용
연계를 위한 데이터 제공방식	• 관련법상 제3자 제공 금지 규정 ⇒ KCB는 연계를 위한 1차 암호화 데이터의 제공에 한계 • 통계법상 통계자료 제공 가능 ⇒ 통계법상 식별할 수 없는 형태로 통계자료를 처리 후 제공 (제31조)	• 통계청이 KCB에 데이터 제공 • KCB가 데이터간 연계 및 연계키 삭제 후 분석용 DB 작성 → 양기관 공유 * 분석용 DB는 개인정보가 아님 (법률 자문 결과)
데이터 관리 및 분석방식	• 상호간에 소관 데이터의 유출 우려	• 통계청의 독립된 제한 공간(데이터 센터)에서 자료 분석 및 통계 작성 → 민간기관의 서버를 가지고 입주 • 자료 접근권자 인가

연계한 데이터는 통계청 통계조사 및 공공 행정자료와 신용정보회사의 신용 자료인데 그 내용은 아래와 같다고 한다.

표 4-49 통계청과 KCB 연계 데이터

자료원	활용 항목	자료원	활용 항목
인구동향통계	혼인날짜, 연령, 직업, 학력, 출산자녀수	임금근로 일자리통계	전년도보수총액
인총 및 등록센서스	주택형태, 가구 유형	사업소득신고	매출액
주택소유통계	보유주택수	민간 부채자료	소득, 신용등급, 대출잔액, 연체금액, 부채상환액, 카드사용액

연계데이터의 구성항목을 도식화하면 아래와 같다.

그림 4-43 신혼부부 신용DB 자료연계 구성항목

통계청은 신혼부부를 추출하여 표본을 구성하고, 이들의 정보와 민간데이터인 KCB의 데이터(소득, 신용등급, 대출잔액, 연체금액, 부채상환액, 신용카드 사용액 등)를 연계한 것이다. 물론 해당 신혼부부에게는 동의를 얻지 않은 것이다. 분석한 내용은 신혼부부 가구의 일반현황, 사회경제적 특성, 부채 및 소득현황, 종단분석(동태적 분석), 자녀수, 직업 변동 및 대출잔액 변동, 소득이용 및 빈곤탈출 현황 등이라고 한다.

이 시범사업과 관련하여 통계청은 다음과 같은 조치를 취했다고 한다.

그림 4-44 통계청 연계방식과 비식별 가이드라인과의 차이점

■ 개인정보 비식별 가이드라인과의 차이점

	통계청 연계 방식	가이드라인 방식
비식별화 형태	◇ 식별자의 비식별화 조치 ◇ 속성 정보의 비식별화 미적용 ⇒ 구체적 분석 가능	◇ 식별자의 비식별화 조치 ◇ 속성 정보의 비식별화 적용 ⇒ 활용 범위 제한
연계DB 관리방식	◇ 물리적 공간 및 자료 활용 형태 (ex. 집계표) 제한	◇ 공유 기관간 자료의 개별 활용 (제3자 제공 금지)

다. 평가

통계청이 사전에 그 내용을 공개하지도 않고 민간 기업과의 협력을 통해서 서로 통계주체의 민감한 개인정보를 주고받은 것은 적절한 것으로 평가되기 어렵다. 특히 통계의 비밀보호는 엄격히 법률로서 보호되어야 하는 바, 법령의 규정도 없이 협력사업의 명목으로 이루어질 사안으로 보기는 어렵다.

통계자료의 공유나 활용은 엄격하게 통계의 비밀보호와 개인정보 보호의 관점에서 규율되어야 한다. 그 목적은 공익 목적의 연구로 제한되어야 하고 그 절차도 비밀보호와 개인정보 보호에 충분하도록 준수되어야 한다.

이런 점에서 본 협력사업은 바람직한 데이터 연계의 모델로 보기 어렵다. 오히려 지양해야 할 데이터 연계의 모델이라고 보지 않을 수 없다.

제4절 개인정보 비식별 조치 전문기관을 통한 민간데이터 연계·결합 지원

1. 개요

2016년 7월 관계부처 합동으로 '범정부 개인정보 비식별 조치 가이드라인'(이하 '비식별 가이드라인')을 발간하였다. 비식별 가이드라인은 빅데이터 분석에 활용하기 위해 서로 다른 정보집합물(데이터셋)을 결합하는 공공기관 및 민간 기업의 업무를 지원하기 위하여 개인정보 비식별 조치 '전문기관'을 설립하도록 하였다.

이와 같은 전문기관의 설치 및 운영은 다음과 같은 법률 해석에 기반하여 이루어졌다. 현행 개인정보 보호법의 경우 개인정보의 목적 외 이용·제공 제한을 엄격하게 제한하고 있으며(동법 제18조), 데이터 연계·결합의 경우에도 이 조항의 적용을 받는다. 그런데 개인정보 보호법상 개인정보의 정의에는 개인을 '알아볼 수 있는' 정보와 '다른 정보와 쉽게 결합하여' 알아볼 수 있는 정보가 포함되어 있다(동법 제2조제1호). 비식별 가이드라인은 빅데이터 활용에서 개인정보 범위에 대한 확대 해석이 우려된다고 보고, 개인정보 보호법의 적용대상이 되는 '개인정보'의 범위 중 다른 정보와 쉽게 결합하여 '알아볼 수 있는' 자의 주체를 해당 '정보를 처리하는 자'로, '다른 정보와 쉽게 결합하여'의 의미를 '결합대상이 될 다른 정보의 입수 가능성이 있어야 하고, 또 결합 가능성도 높아야 한다'는 것으로 해석하였다(비식별 가이드라인: 4). 가이드라인은 개인을 식별할 수 있는 요소(식별자, 속성자)의 전부 또는 일부를 삭제하거나 대체하는 등의 방법을 통하여 개인을 알아볼 수 없도록 조치한 정보는 비식별 정보로 규정하고, 이 가이드라인에 따라 비식별 조치한 정보는 '개인정보가 아닌 것으로 추정'하였다.

> 비식별 정보가 개인정보에 해당하는지 여부가 의문이 있을 수 있으나, 가이드라인에 따라 적정하게 비식별 조치가 된 정보는 더 이상 특정 개인을 알아볼 수가 없으므로 개인정보가 아닌 것으로 추정됩니다 … 비식별 정보는 개인정보가 아닌 정보로 추정되므로 정보주체로부터의 별도 동의없이 해당 정보를 이용하거나 제3자에게 제공할 수 있습니다. (범정부 개인정보 비식별 조치 가이드라인: 57)

전문기관의 정보 집합물 결합지원 서비스도 이러한 법률 해석에 기반하여 운영된다. 즉, 서로 다른 기업이 보유한 DB를 결합하는 과정에서 정부 가이드라인이 권장하는 방식대로 적정하게 비식별조치가 이루어진 경우에는 개인정보가 아닌 것으로 추정되고, 이를 전문기관이 결합하는 것도 현행 법률상 목적 외 이용·제공 조항에 위반되지 않는다고 보는 것이다. 의도적으로 비식별 정보가 쉽게 재식별될 수 있도록 이용·

제공하거나, 재식별 정보를 보호조치 없이 보관·이용·제공한 경우에는 관련 법령에 따른 벌칙이나 과태료가 부과될 수 있지만, 가이드라인에 따라 비식별 조치를 실시하였다면 고의성 없는 재식별 사실만으로 책임을 부과할 수는 없다고 보았다(비식별 가이드라인: 75).

특히 정보 집합물 결합에 있어 전문기관의 핵심 기능은 임시 대체키를 활용한 연계기관으로서의 역할이다. 비식별 가이드라인은 이를 다음과 같이 설명하고 있다(비식별 가이드라인: 18). 빅데이터 분석에 활용하기 위해 서로 다른 사업자가 보유하고 있는 정보 집합물을 결합하는 경우, 정보주체를 알아볼 수 있는 식별자 그 자체를 매칭키로 사용하는 것은 개인정보에 대한 목적 외 이용·제공 제한 등 현행법(개인정보 보호법 제18조) 위반 소지가 있다. 따라서 정보 집합물 간 결합·분석을 위해서는 결합과정에서만 임시로 매칭키 역할을 하는 '임시 대체키'의 활용이 필요하다. 임시 대체키를 활용한 결합을 허용하는 경우 무분별한 결합을 통한 개인정보 침해 소지를 방지하기 위해 제3의 공공기관에서만 결합하도록 하는 등 지원 및 관리체계가 필요하다. 이때 결합을 지원하고 관리하는 제3의 공공기관이 비식별 조치 전문기관이다.

개인정보 비식별 조치 전문기관을 통한 데이터 연계·결합의 경우 목적을 제한하고 있지 않다. 신청기관 역시 '정보 집합물 결합지원 등을 받고자 하는 기관 또는 사업자 등'으로 폭넓게 규정되어 있다(개인정보 비식별 조치 지원기관 운영지침 제2조 제6호). 정부 비식별 가이드라인은 기업 간 정보 집합물 결합을 지원하는 데 초점을 두고 있다. 가이드라인은 이와 같은 데이터 결합의 기대 효과를 <표 4-50>과 같은 예를 들어 설명하였다. 실제로 2016년 8월 각 전문기관이 지정된 후로부터 2017년 9월까지 전문기관을 이용한 기관은 대부분은 민간 기업들이었다.

가이드라인은 각 소관부처 책임 하에 분야별로 전문기관을 정하여 운영하도록 하였다. 즉, 소관 부처가 한국인터넷진흥원, 한국신용정보원, 금융보안원, 사회보장정보원, 한국정보화진흥원 중에서 공문으로 분야별 전문기관을 지정·공표하여 운영하고, 필요시 추가 지정하도록 하였다.

분야별 전문기관은 가이드라인에 따른 개별 기업의 개인정보 비식별 조치를 지원함과 동시에, 기업 간 정보 집합물의 결합을 지원한다. 우선 전문기관은 기업의 비식별 조치를 지원하는데, 그 역할로는 △ 비식별 조치 적정성 평가단 풀(비식별 조치 기법 전문가, 법률 전문가 등) 구성·운영 △ 의료, 복지, 교육, 금융·신용, 통신, 유통, 공공·기타 분과 등 산업별로 필수적인 비식별 조치 이행 권고(k-익명성 수치 등) △ 비식별 조치 적정성 실태 점검 등을 수행한다. 다음으로 각 전문기관은 기업 간 정보 집합물의 결합을 지원하는데, 정보 집합물 결합지원 업무 전반에 있어 개인을 식별하려는 일체의 시도가 금지되고 정보 집합물 결합 및 정보 제공 완료 후 모든 정보를 파

기하도록 하였다.

표 4-50 전문기관을 통한 데이터 결합의 예시

예시	신청기관	상대기관	결합 목적	결합 활용(예시)
1	○○증권	△△은행, ◇◇보험	신상품 개발에 활용	▶ △△은행, ◇◇보험 등은 보유하고 있는 다양한 신용정보를 비식별 조치한 후 ○○증권에 제공 ▶ ○○증권은 제공받은 자료를 빅데이터 분석하여 &로보어드바이저&, &ISA& 등 다양한 신상품 개발에 활용하고 국내 및 해외시장 개척을 추진
2	신생 스타트업인 ◇◇사	□□은행	새로운 비즈니스 모델 개발에 활용	▶ □□은행은 보유하고 있는 학력·연령·성별 첫 직장, 이직경로, 연봉 등의 정보를 비식별 조치하여 신생 기업인 ◇◇사에 제공 ▶ ◇◇사는 기존의 헤드헌팅 회사와 차별화된 &첫 직장부터 퇴직 후까지 커리어 관리프로그램&을 제공하는 비즈니스 모델을 개발하여 활용
3	○○제약회사	△△심사평가원	××신약개발 연구에 활용	▶ △△심사평가원이 특정 질병 환자의 연령과 성별에 따른 진료기록을 충분히 비식별 조치한 후, ○○제약회사에게 제공 ▶ ○○제약회사는 해당 정보를 활용하여 ××병의 발병원인 및 치유 원인을 분석하여 신약을 개발, 수입 약품 대비 20% 저렴한 가격으로 판매
4	□□홈쇼핑	◇◇카드사	우수고객 마케팅 전략 수립에 활용	▶ □□홈쇼핑과 ◇◇카드사는 고객 전화번호와 카드 결제정보를 각각 복원되지 않는 알고리즘으로 비식별 조치하여 A전문기관에 제공하고 A전문기관은 두 정보를 결합한 후, □□홈쇼핑에게 제공 ▶ 비식별 조치된 고객의 결제정보를 통해 □□홈쇼핑은 우수고객이 선호하는 물품을 특정 시간대에 할인 행사를 실시하는 마케팅 전략 수립

* 출처: 범정부 개인정보 비식별 조치 가이드라인, p25

기업 간에 정보 집합물을 결합할 때 전문기관 선택 기준은 분야별로 이루어진다. 즉, △ 동일 산업 분야 내 기업 간 결합은 해당 분야 전문기관에서 결합지원 △ 이종 산업 간 결합은 대량의 정보 집합물을 결합하고자 하는 기업이 속해 있는 분야별 전문기관 또는 사업자 간 상호 협의하여 선정한 전문기관에서 수행 △ 당해 산업을 지원해 주는 전문기관이 없는 경우에는 한국인터넷진흥원 또는 한국정보화진흥원에서 지원하도록 하였다(개인정보 비식별 조치 지원기관 운영지침 제9조). 전문기관에 대한 세부 이용기준은 각 부처에서 마련·시행한다.

한국인터넷진흥원(KISA)에는 '개인정보 비식별 지원센터'를 설치·운영하여 △ 분야별 전문기관 운영 가이드라인 마련 및 실태 점검, △ 전문기관 실무협의회 운영, △ 분야별 평가단 풀 관리 및 교육, △ '개인정보 비식별 조치 가이드라인' 업데이트 및 활용 지원 등 전문기관 간의 비식별 조치 지원 업무를 전반적으로 점검 및 조율하도

록 하였다.

2016년 8월 각 부처에 의해 개인정보 비식별 조치 전문기관이 지정되었으며, 9월 한국인터넷진흥원에도 '개인정보 비식별 조치 지원센터'가 설치되었다. 그 현황은 <그림 4-45>와 같다.

그림 4-45 부처별 개인정보 비식별 조치 전문기관 지정 현황 (2016년 9월)

* 출처: "KISA, 기업의 개인정보 비식별조치 본격 지원한다", 한국인터넷진흥원 보도자료(2016. 9. 20.)

2016년 8월 24일 '비식별 조치 전문기관 협의회'가 발족하였다. 분야별 전문기관으로 구성된 협의회는 그 의장을 한국인터넷진흥원 개인정보 보호본부장이 맡고 한국인터넷진흥원 개인정보 비식별 조치 지원센터에 사무국을 두었다.

2016년 9월 21일에는 한국인터넷진흥원을 비롯한 전문기관들 공동으로 '개인정보 비식별 조치 및 결합지원 서비스 설명회'를 개최하고 △ 비식별 조치 및 적정성 평가 절차·방법 △ 정보 집합물 결합지원 서비스 안내 및 신청방법 △ 비식별 조치 솔루션 등을 안내하고 9월 말부터 기업의 비식별 조치 지원 및 정보 집합물 결합지원 서비스 이용이 가능하도록 하였다.

2. 절차 및 운영

한국인터넷진흥원은 2016년 8월 '비식별 조치 전문기관 협의회'를 통해 비식별 조치 적정성 실태 점검, 정보 집합물 결합 시 준수사항 등을 포함한 '개인정보 비식별 조치 지원기관 운영지침'(이하 지원기관 운영지침)을 마련하였다.

비식별 가이드라인 및 지원기관 운영지침에서 서술한 데이터 연계 및 결합의 절차와 운영의 주요 내용은 다음과 같다.

전문기관을 통한 정보 집합물 결합 절차는 <그림 4-46>과 같은 단계를 거친다. 첫째, 정보 집합물 결합을 원하는 A 회사와 B 회사가 같은 알고리즘을 적용하여 식별

자를 임시 대체키로 전환하고, 결합 상대 정보 집합물도 비식별 조치 및 적정성 평가를 수행하도록 한다. '임시 대체키'를 생성할 시 동 대체키에 잡음을 추가하거나, 2개 이상의 식별자를 활용할 경우 식별자 중 일부를 조합하여 불법적 복호화 또는 원본 정보와 결합 시에도 개인을 식별할 수 없도록 조치한다. 둘째, 비식별 조치된 정보를 전문기관에 제공하고 결합을 요청한다. 이 경우 전문기관은 제공받은 비식별 정보를 통해 특정 개인을 식별하는 것이 불가하다. 셋째, 임시 대체키를 활용하여 전문기관에서 결합을 수행한다. 넷째, 전문기관이 결합을 수행한 후 임시 대체키를 삭제한다. 다섯째, 전문기관이 결합한 결합 DB를 필요한 기업에 제공한다. 전문기관은 제공 후 이를 파기 조치한다. 가이드라인은 이때 결합 DB는 임시 대체키가 삭제된 후 제공되었으므로 A와 B도 결합 DB를 통해 특정 개인의 식별이 어렵다고 보았다.

결합 과정에서 신청기업과 전문기관은 다음 사항을 준수해야 한다. 우선 신청기업 A사와 B사는 분야별 전문기관과 임시 대체키 생성 알고리즘에 대한 정보를 공유하는 것이 금지된다. 임시 대체키 생성을 위해 주민등록번호를 활용하는 것은 금지된다. 다른 정보와의 결합을 위해 임시 대체키를 활용하는 경우, k-익명성 값은 임시 대체키를 제외하고 산출해야 한다. 전문기관으로부터 결합 DB를 제공받은 기업은 이용 전에 반드시 적정성 평가를 수행해야 한다. 전문기관의 경우 결합 과정에서 재식별 발생 시 해당 정보를 즉시 파기해야 한다(비식별 가이드라인: 20).

그림 4-46 비식별 가이드라인에 따른 기업 정보집합물 결합 절차

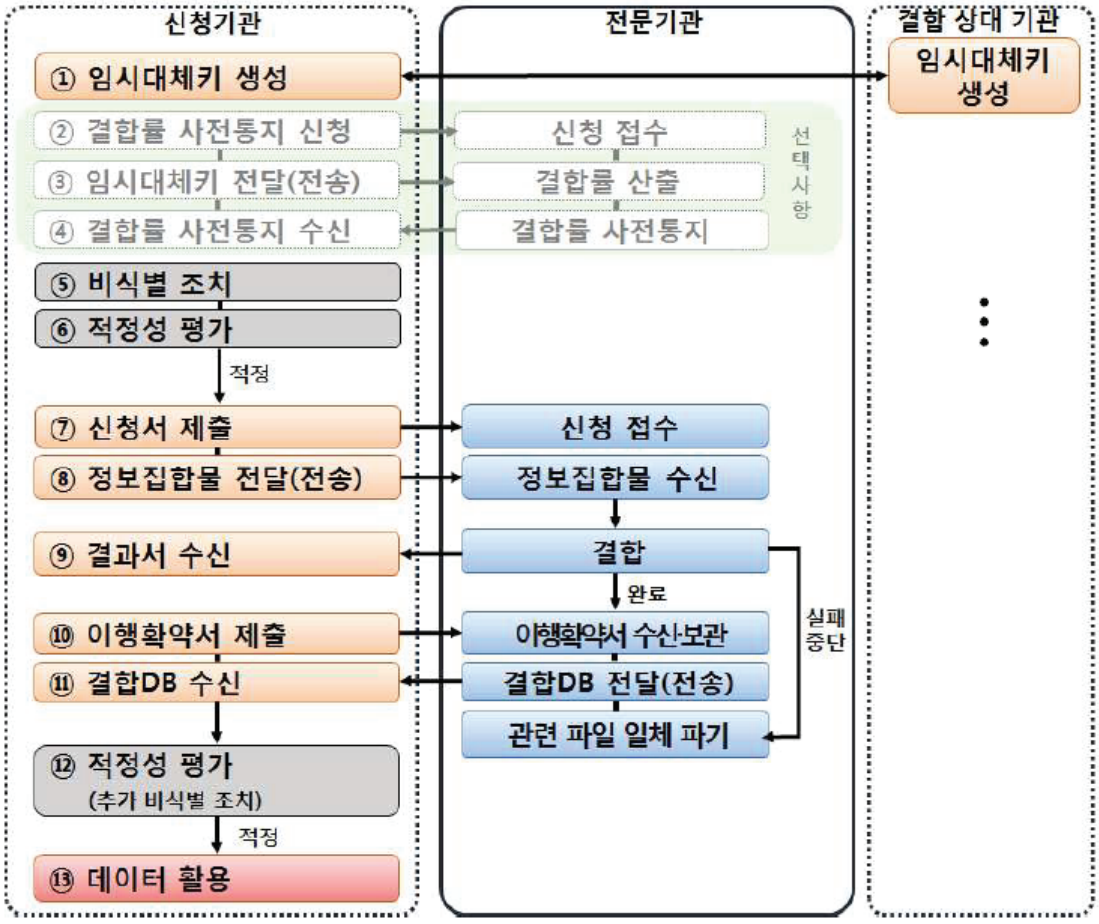
* 출처: 범정부 개인정보 비식별 조치 가이드라인, p20

또 전문기관은 정보 집합물 결합지원 업무의 처리 과정에서 개인을 식별하려는 일체의 시도를 하여서는 안 된다. 결합된 정보를 신청기업에 제공한 이후에는 관련 정

보 일체를 즉시 파기하여야 한다. 전문기관은 업무 처리 과정에서 정보 집합물에 주민등록번호가 포함되어 있는 등 비식별 조치가 충분하지 않음을 알게 된 경우에는 해당 정보를 즉시 파기하고, 신청기업에 결합대상 정보 집합물을 적정하게 비식별 조치한 후 다시 제출하도록 요청하여야 한다. 신청기업으로부터 결합대상 정보를 제공받거나, 결합정보를 신청기업에 제공하는 경우에는 안전성 확보에 필요한 조치를 하여야 한다. 정보 집합물 결합 신청 및 처리 현황 등을 관리할 수 있도록 '정보 집합물 결합 관리대장'을 작성·관리하여야 한다(지원기관 운영지침 제10조).

그 밖에도 전문기관은 정보 집합물 결합지원을 할 때 이를 안전하게 처리 및 관리하기 위하여 다음과 같은 보호조치를 취하고 있다. 전문기관은 지원과정에 참여한 인력, 장비 및 각종 보고서 등에 대한 보안대책을 수립·시행하여야 하며 그 이행실태에 대한 정기적인 점검을 통하여 필요한 경우 즉각 개선 조치를 하여야 한다. 또 비식별 조치 지원 관련 업무수행 인력으로부터 보안각서를 징구하여 보관하여야 하며 그 업무수행 인력은 비식별 조치 관련 업무 수행과정에서 취득한 자료와 정보를 외부에 누설하여서는 아니 된다(지원기관 운영지침 제19조).

그림 4-47 전문기관 정보집합물 결합 세부절차



* 출처: 한국신용정보원·금융보안원, "정보집합물 결합 안내서(2017. 3)", p4.

이상과 같은 전문기관을 통한 정보 집합물 결합 절차를 단계별로 보다 세부적으로 살펴보면 다음과 같다.

우선 결합 신청기업은 상대기업과 결합에 대한 세부사항과 임시 대체키 생성방법을 협의한다. 이때 △ 결합DB의 활용범위, 재제공 또는 가공 여부 및 범위, 수수료, 보유기간을 포함하여 이용 목적을 구체적으로 협의 △ 데이터 활용도를 높일 수 있는 비식별 조치 기법 논의 △ 임시 대체키 생성 시 사용할 식별자 및 임시 대체키 생성 알고리즘 등 임시 대체키 생성방법을 결정한다.

신청기업은 이와 같은 방식으로 결합 상대기업과 협의하여 동일한 방식으로 임시 대체키를 생성한 후 결합대상 데이터에 추가한다.

정보 집합물을 결합할 때 전문기관에 실제 식별자가 노출되지 않도록 식별자를 대신하기 위해 임시로 생성한 매칭키를 의미하는 임시 대체키는 정보주체를 유일하게 구분할 수 있어야 하며 그 생성 알고리즘이 비가역적 알고리즘(안전한 해시알고리즘), 안전한 암호알고리즘 등을 활용하여 원본값 복원이 충분히 어려워야 한다. 임시 대체키 생성을 위한 입력 값으로 주민등록번호는 사용할 수 없다. 임시 대체키의 다른 예로는 전화번호, 이름 조합이나 전화번호 뒤 4자리, 주소의 조합을 들 수 있다.

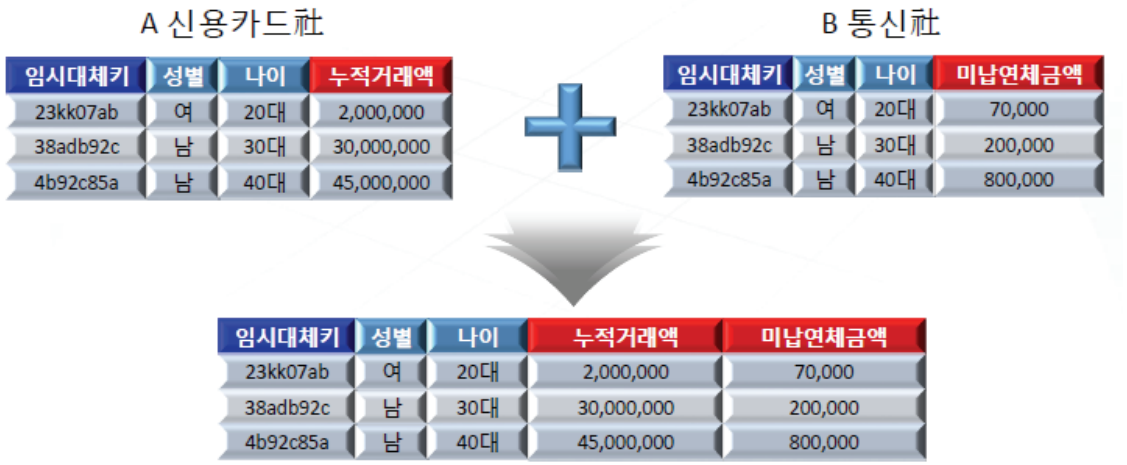
신청기업 A와 상대기업 B의 정보 집합물에 대한 임시 대체키 생성 예제는 다음과 같이 그려볼 수 있다.

그림 4-48 임시 대체키 생성 예시

* 출처: 금융보안원, "금융분야 비식별 조치 지원 전문기관 정보집합물 결합", 금융권 개인정보 비식별 조치 세미나(2016. 9. 9), p8.

신청기업 A와 상대기업 B의 정보 집합물을 임시 대체키를 이용하여 결합한 결과물의 예제는 다음과 같이 그려볼 수 있다.

그림 4-49 임시 대체키 이용 결합 과정 예시



* 출처: 금융보안원, 앞의 자료, 금융권 개인정보 비식별 조치 세미나(2016. 9. 9), p2.

위 정보 집합물은 전문기관에 제공되기 전에 비식별 조치가 이루어진다. 비식별 조치 기준은 식별자와 속성자로 나누어 볼 수 있다. '식별자'는 개인 또는 개인과 관련한 사물에 고유하게 부여된 값 또는 이름이며, '속성자'는 다른 정보와 쉽게 결합하는 경우 특정 개인을 알아볼 수도 있는 정보를 의미한다. 비식별 가이드라인에서는 비식별 조치 단계에서는 정보 집합물에 포함된 식별자의 경우 원칙적으로 삭제 조치하도록 하였다. 다만 데이터 이용 목적상 반드시 필요한 식별자는 비식별 조치 후 활용할 수 있다고 허용하였다. 정보 집합물에 포함된 속성자도 데이터 이용 목적과 관련이 없는 경우에는 원칙적으로 삭제 조치하도록 하였다. 다만 데이터 이용 목적과 관련이 있는 속성자는 가명처리, 총계처리 등의 기법을 활용하여 비식별 조치하도록 하였다. 희귀병명, 희귀경력 등의 속성자는 구체적인 상황에 따라 개인 식별 가능성이 매우 크므로 엄격한 비식별 조치가 필요하다고 하였다. 일반상병명이나 신용등급 등이 이용 목적에 요긴한 속성자의 경우에는 비식별 조치를 미적용하기도 한다. 적정성 평가 시에는 기본적으로 k-익명성 모델을 활용한다.

이와 같은 비식별 조치가 의료기관에 적용된 예시는 다음 <그림 4-50>과 같이 살펴볼 수 있다. 이 예시에서 이름은 식별자로서 삭제되었고, 준식별자인 혈액형은 5-익명성에 준하여 마스킹 되었으며, 또 다른 준식별자인 키와 몸무게는 10단위로 범주화되었다. 병명은 민감속성정보이지만 이용 목적상 필요하므로 비식별 조치를 미적용하였다.

그림 4-50 비식별 조치 예시 (의료기관)

	이름	혈액형	키	몸무게	병명
구분	식별자	준식별자	준식별자	준식별자	민감속성정보
비식별 조치 기준 및 수치	-	k-익명성 (5)			l-다양성 (3)
비식별 조치 기법·기술	삭제	마스킹 (* 처리)	범주화 (범위방법)	범주화 (범위방법)	미적용

혈액형	키	몸무게	병명
*	180~190	80~90	우울증
*	160~170	50~60	-

* 출처: 한국신용정보원·금융보안원, "개인정보 비식별 조치 적정성 평가 안내서(2016. 9.)", p15.

비식별 조치가 금융기관에 적용된 또 다른 예시는 <그림 4-51>과 같이 살펴볼 수 있다. 이 예시에서 이름은 식별자로서 삭제되었고 속성자인 직업은 5-익명성에 준하여 범주화되었으며, 성별은 1 혹은 2로 가명화되었고 고객신용등급은 민감속성정보이지만 이용 목적상 필요하므로 비식별 조치를 미적용하였으며, 다른 민감속성정보인 대출액 합계는 1백만 원 단위로 범주화되었다.

신청기업은 결합 신청 전에 결합대상 정보 집합물의 비식별 조치 및 적정성 평가를 이와 같이 완료하여야 한다. 다만 신청기업은 정보 집합물 결합 추진의 타당성 여부를 검토하기 위하여 정보 집합물 결합을 위한 비식별 조치 및 적정성 평가를 수행하기 전에 전문기관에 의뢰하여 결합률을 사전통지 받은 후 결합 추진 여부를 결정할 수 있다. 결합률 사전통지를 신청하는 각 신청기업은 신청이 접수되면 임시 대체키를 전문기관에 안전한 방법으로 전달 또는 전송한다. 전문기관은 임시 대체키를 사용하여 정보 집합물 결합 시 예상되는 결합률을 산출하여 그 결과를 신청기업에 통지한다. 정보 집합물 결합을 원하는 신청기업은 결합 상대 정보 집합물 정보, 결합 완료 요청 일자, 결합정보 이용 목적 등을 기재한 '정보 집합물 결합지원 신청서'를 작성하여 분야별 전문기관 홈페이지 및 이메일 등을 통해 제출한다.

그림 4-51 비식별 조치 예시 (금융기관)

항목	구분	비식별 조치 기법·기술	비식별 조치 기준 및 수치
이름	식별자	삭제	-
직업	준식별자	범주화	k-익명성(5)
성별	준식별자	가명화	
고객등급	민감속성정보	미적용	l-다양성(3)
대출액 합계	민감속성정보	라운딩, 범주화	

항목 (조치기법)	원본 데이터	비식별 조치 데이터
직업 (범주화)	치과의사, 한의사	의사
	판사, 검사	법조인
	회사원, 공무원	급여소득자
	:	:
	주부, 학생	무직
성별 (가명화)	남자	1
	여자	2
대출액 합계 (라운딩 및 범주화)	1~1,000,000(원)	1,000,000(원)
	1,000,000~2,000,000(원)	2,000,000(원)
	:	:
	748,000,000~ 749,000,000(원)	749,000,000(원)
	749,000,000(원) 이상	750,000,000(원)

* 출처: 한국신용정보원·금융보안원, "개인정보 비식별 조치 적정성 평가 안내서(2017. 4)", p16.

신청기업은 전문기관에 결합 신청서를 제출하여 신청이 접수되면, 결합대상 정보 집합물을 저장 매체 또는 정보통신망을 통해 안전한 방법으로 전문기관에 전달한다.

전문기관은 전달받은 복수의 정보 집합물을 임시 대체키를 기준으로 결합하여 신청기업에 전달한다. 이때 전문기관은 신청기업이 제출한 결합대상 정보 집합물에서 임시 대체키가 같은 레코드의 속성들을 결합한 후 임시 대체키는 삭제하고 결합된 결과물만 신청기업에 제공한다. 결합되지 않은 정보는 삭제된다.

전문기관이 결합 처리를 할 때는 내부망에 위치한 시스템에서 안전하게 결합하며, 네트워크 접근을 통제하고 계정 및 권한을 관리한다. 일부 전문기관은 복수의 비식별 정보 집합물을 병합하여 하나의 정보 집합으로 만드는 소프트웨어가 설치된 서버로 구성된 정보집합물결합시스템을 구축하여 정보 집합물 결합을 지원한다.

결합 과정 중 개인을 식별할 수 있는 정보 발견 시 전문기관은 결합 처리를 중단하고 관련 데이터를 일체 삭제 후 신청기업에 통지한다.

전문기관은 정보 집합물의 결합지원을 신청하고자 하는 기업에 결합 업무수행에 필요한 실비의 범위 내에서 수수료를 부과할 수 있으며(지원기관 운영지침 제13조), 중

소기업 등 자체적인 비식별 조치가 어려운 기업을 대상으로 비식별 조치를 대행하거나[334] 적정성 평가를 대행하기도 한다(지원기관 운영지침 제5조 제4호).

전문기관을 통해 정보 집합물을 결합하고 '완료' '중단' '실패' 등 그 결과를 확인한 신청기업은, '결합정보에 대한 안전조치 이행 확약서'(2017.3. '결합정보에 대한 필수조치 이행 확약서'로 개정)를 전문기관에 제출한다.

전문기관은 결합데이터를 신청기업에 전달할 때는 저장 매체를 이용하여 직접 전달하거나, 정보통신망을 통한 암호화 전송을 한다. 결합데이터 전달 후에 전문기관은 즉시 관련 데이터를 복구 불가능한 방법으로 일체 파기한다.

신청기업은 결합DB를 이용하기 전에 적정성 평가를 수행한다. 평가 결과가 '적정'인 경우 그대로 활용 가능하며 '부적정'인 경우 평가 결과가 '적정'이 나올 때까지 추가적인 비식별 조치를 취해야 한다. 신청기업은 평가가 완료된 데이터를 결합 및 평가 단계에서 기재한 이용 목적에 한하여 활용해야 하고, 결합DB를 수신한 직후부터 비식별 정보 안전조치, 재식별 가능성 모니터링 등 사후관리도 수행해야 한다.

비식별 정보 안전조치 항목과 재식별 가능성 모니터링 점검 항목은 각각 <그림 4-52> 및 <그림 4-53>과 같다.

334) 금융보안원·신용정보원, "금융권빅데이터 지원전문기관운영방안(2016. 9. 1)", p7.

그림 4-52 비식별 정보 안전조치 항목

구분	비식별 정보 보호 조치
관리적 보호조치	① 비식별 정보파일 관리담당자 지정 ② 비식별 정보파일 대장관리 ③ 원본정보 관리부서(기관)와 비식별 정보 관리부서(기관) 간 비식별 조치 관련 정보공유 금지 ④ 이용목적 달성 시 지체 없이 파기 ⑤ 비식별 정보파일 유출시 대응계획 수립
기술적 보호조치	⑥ 비식별 정보파일에 대한 접근권한 관리 및 접근통제 ⑦ 비식별 정보 보관시스템에 대한 접속기록 관리 ⑧ 악성 코드 방지 등을 위한 보안프로그램 설치·운영

* 출처: 한국신용정보원·금융보안원, "정보집합물 결합 안내서(2017. 3)", p7.

그림 4-53 재식별 가능성 모니터링 점검 항목

구분	점검 항목
내부 요인의 변화	비식별 조치된 정보와 연계하여 재식별 우려가 있는 추가적인 정보를 수집하였거나 제공받은 경우
	데이터 이용과정에서 생성되는 정보가 비식별 정보와 결합해서 새로운 정보가 생성되는 경우
	이용부서에서 비식별 정보에 대한 비식별 수준을 당초보다 낮추어 달라고 하는 요구가 있는 경우
	신규 또는 추가로 구축되는 시스템이 비식별 정보에 대한 접근을 관리·통제하는 보안체계에 중대한 변화를 초래하는 경우
외부 환경의 변화	이용 중인 데이터에 적용된 비식별 조치 기법과 유사한 방법으로 비식별 조치한 사례가 재식별 되었다고 알려진 경우
	이용 중인 데이터에 적용된 비식별 기법과 기술을 무력화 하는 새로운 기술이 등장하거나 공개된 경우
	이용 중인 데이터와 새롭게 연계 가능한 정보가 출현하거나, 공개된 것으로 알려진 경우

* 출처: 한국신용정보원·금융보안원, "정보집합물 결합 안내서(2017. 3)", p8.

그림 4-54 정보집합물 결합 신청서

정보집합물 결합 신청서

접수번호		접수일	처리기간	
신청 기관명	기관 명칭		담당자(성명,직위)	
	신청부서		전화번호	
	소재지		E-mail	
결합 상대 기관명	기관 명칭			
	신청부서			
	소재지			
신청 내용 (신청기관)	결합 목적			
	결합 희망 정보집합물 주요내용	*간략히 서술 (예시: 16년 9월 A병원의 진료내역)*		
	결합정보 수령 요청일자	*0000년 00월 00일*		
	기타 특이사항 (결합대상 정보집합물 구조, 특성 등 가능한 상세히 기술)	*후면에 첨부 가능*		
비식별 조치 적정성 평가 수행여부	예 □	※ '개인정보 비식별 조치 적정성 평가 결과 통지서'를 첨부하여 본 신청서 제출하여야 함		
	아니오 □	(사유:) ※ 데이터의 이용목적이 개인정보 보호법 제18조제2항제4호의 통계작성 및 학술연구인 경우에는 비식별 조치 후에 적정성 평가단계를 생략할 수 있음		

※ 개인정보 수집.이용 동의
ㅇ 수집하는 개인정보 항목 : 성명, 부서, 직위, 전화번호, 이메일주소
ㅇ 개인정보의 수집.이용 목적 : 업무 관련 정보 요청
ㅇ 개인정보 보유 및 이용기간 : 신청 접수 후 5년까지
위 내용을 이해하였으며 그 내용에 동의합니다.

□ 동의합니다. □ 동의하지 않습니다.

「개인정보 비식별 조치 지원기관 운영지침」 제9조제2항에 따라 위와 같이 결합을 신청합니다.

년 월 일

신청인(기관명) (서명 또는 인)

*(전문기관명)*장 귀하

첨부서류	결합 대상 정보집합물의 데이터 명세서

처리절차

신청서 작성	→	접 수 (검토 및 결재)	→	결합 일정 안내	→	정보집합물 전달	→	결합 실시	→	결합결과 통보
신청기관		전문기관		전문기관		신청기관		전문기관		전문기관

* 출처: 한국신용정보원·금융보안원, "정보집합물 결합 안내서(2017. 3)", p12.

3. 분야별 전문기관

2017년 9월 현재, 개인정보 비식별 조치 전문기관을 분야별로 살펴본 현황은 <표 4-51>과 같다. 이 절에서는 전문기관별로 실제 결합이 이루어진 사례를 살펴본다.

표 4-51 분야별 개인정보 비식별 조치 전문기관 지정 현황 (2017년 9월 현재)

분야	전문기관	소관부처
공공기관 (정부부처, 지자체)	국가정보자원관리원 (구 정부통합전산센터)	행정안전부
공공기관 (정부부처, 지자체 제외)	한국인터넷진흥원	행정안전부
통신	한국인터넷진흥원	방송통신위원회
통신	한국정보화진흥원	과학기술정보통신부 (구 미래창조과학부)
금융	금융보안원	금융위원회
금융	한국신용정보원	금융위원회
보건・복지	사회보장정보원	보건복지부
교육	한국교육학술정보원	교육부

(1) 한국인터넷진흥원

한국인터넷진흥원은 정보통신망 이용촉진 및 정보보호 등에 관한 법률 등에 근거하여 운영되고 있는 위탁집행형 준정부기관이다. 2009년 한국정보보호진흥원(KISA), 한국인터넷진흥원(NIDA), 정보통신국제협력진흥원(KIICA)이 통합되어 출범하였다.

한국인터넷진흥원은 2016년 8월 방송통신위원회와 행정안전부로부터 정보통신사업자(정보통신망법 적용대상) 및 공공기관(중앙부처 및 지자체 제외)의 정보 집합물 결합을 지원하는 전문기관으로 지정되었다. 더불어 2016년 9월 1일부터 '개인정보 비식별 조치 지원센터'를 설치하여 전문기관 간의 비식별 조치 지원 업무를 전반적으로 점검 및 조율하고 있다.

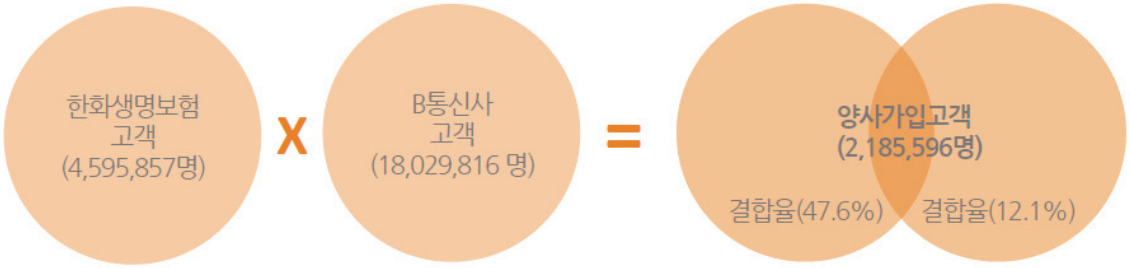
한국인터넷진흥원은 전문기관으로 지정된 2016년 8월부터 2017년 9월까지 정보 집합물 결합을 1건 실시하였다.

표 4-52 한국인터넷진흥원 결합사례

	신청기관	상대기관
명칭	SK텔레콤	한화생명
결합 목적	결합정보의 유효성 검증 중금리 대출 대상 확장 가능성 검증	
데이터수	18,029,816	4,595,857
결합건수 (결합률)	2,185,596건(12.1%)	2,185,596(47.6%)
결합항목	나이, 성별, 사용개월수, 멤버쉽등급, 월평균통화시간, 통화빈도, ARPU(가입자당 평균매출), 결합상품가입여부, 단말기출고가, 이용정기기간, 연체금액, 최대연체금액, 납부방법, 회선상태, 남은할부원금, 가입회선수, tablet보유여부, smart watch보유여부, 멤버쉽 사용금액(월), 멤버쉽사용금액(년), 미납횟수	직업, 신용대출건수, 최초계약날짜, 최초연체등록날짜, 총신용대출금액, 총상환금액, 신용대출연체율, 신용대출연체률2, 신용대출연체율3, 최초신용등급, 최근신용등급, 보험료연체율, 최근1년 보험연체율, 실효해지건수, 기납입보험료, 납입보험료, 추정소득, 가구추정소득, 평균약관대출율, 약간대출금액, 자동이체실패월수
임시 대체키	이름, 주민번호 앞7자리	
비식별조치	나이(3-익명성), 당월연체금액(2-다양성) 적용하고 나머지 항목에 대해 범주화를 통한 비식별 조치 수행 * 다만 성별(2개 범주), 결합상품가입여부(2개 범주), 납부방법(4개 범주), 회선상태(2개 범주), Tablet 보유여부(2개 범주), Smart Watch(2개 범주) 보유여부는 비식별 조치 미적용하여 원본 동일	직업(10-익명성), 그 외 모든항목에(2-다양성) 적용 * 비식별 조치를 수행한 모든 항목에 범주화를 통한 비식별 조치 수행

* 출처: 권은희 의원(2017년 국정감사), 추혜선 의원(2017년 국정감사) 종합.

그림 4-55 한화생명-SK텔레콤 간의 데이터 결합 건수



* 출처: 전도현, "한화생명 개인정보 비식별 결합 사례", 개인정보 비식별 조치 및 결합지원 서비스 설명회(2017. 4. 11), p18.

** B통신사는 SK텔레콤을 의미함

한국정보화진흥원이 수행한 한화생명-SK텔레콤 간의 결합 목적은 보험과 통신 간 상관관계를 분석함으로써 통신데이터의 연계 활용 가능성을 검증하고 중금리 대출 대상 확장 가능성을 검증하는 데 있었다.

이 사례는 한화생명보험 고객 4,595,857명의 고객정보, 거래정보, 신용정보 등 21개 항목과 SK텔레콤 고객 18,029,816명의 고객정보, 거래정보, 신용정보 21개 항목 등 총 42개 항목에 대하여 결합하였고, 결합결과 확장된 속성자 42개 항목 전체와 각각의 특성은 <그림 4-56>과 같다. 결합에 사용된 임시 대체키는 이름, 주민등록번호 앞 7자리에 기반하여 생성하였으며, 결합결과 양사 가입 고객 2,185,596명이 결합되었다.

그림 4-56 한화생명-SK텔레콤 간의 데이터 결합 항목

한화생명	직업	신용대출건수	최초대출날짜	최초연체날짜	총신용대출금액	총상환금액	신용대출연체율
	최근1년 신용대출연체율	30일이내 신용대출연체율	최초신용등급	최근신용등급	보험료연체율	최근1년 보험료연체율	실효해지건수
	기납입보험료	월납입보험료	직업기반 추정소득금액	가구단위 추정소득금액	평균약관대출율	약관대출금액	자동이체 실패월수
B통신사	나이	성별	사용개월수	멤버쉽등급	월평균통화시간	월평균통화빈도	ARPU
	결합상품가입여부	단말기출고가	이용정지기간	당월 통신료연체금액	최근1년 최대 통신료연체금액	납부방법	회선상태
	남은단말기 할부원금	가입회선수	태블릿PC 보유여부	스마트워치 보유여부	멤버쉽 당월사용금액	멤버쉽 당년사용금액	통신료 미납횟수

준식별자 민감정보 신용정보

* 출처: 전도현, 앞의 자료, p21.
** B통신사는 SK텔레콤을 의미함

각 결합 항목에 적용된 비식별 조치는 <표 4-53>과 같다. SK텔레콤은 나이 항목에는 3-익명성, 당월연체금액에는 2-다양성을 적용하고 나머지 항목에 대해 4~303개 구간에 이르는 범주화를 시행하였다. 다만 성별, 결합상품가입 여부, 납부방법, 회선상태, Tablet 보유 여부, Smart Watch 보유 여부 항목은 비식별 조치를 미적용하여 원본과 동일하였다. 한화생명은 직업에 10-익명성을 적용하고 그 외 모든 항목에는 2-다양성을 적용하였으며 비식별 조치로는 항목별로 소수점 단위~210개 구간에 이르는 범주화를 시행하였다.

표 4-53 한화생명-SK텔레콤 결합 항목 비식별 조치

SK텔레콤		
대상	비식별 조치 방법	비고
나이	44개 구간 범주화	준식별자 (k-익명성[3] 적용)
성별	2개 구간 범주화 (미적용)	
사용개월수	9개 구간 범주화	
멤버쉽등급	5개 구간 범주화	
통화시간 (월평균)	89개 구간 범주화	
통화빈도 (월평균)	72개 구간 범주화	
ARPU (가입자당 평균매출)	97개 구간 범주화	
결합상품가입여부	2개 구간 범주화 (미적용)	
단말기출고가	7개 구간 범주화	
이용정지기간 (월)	9개 구간 범주화	
연체금액 (당월)	303개 구간 범주화	민감정보 (l-다양성[2] 적용)
최대연체금액 (최근1년)	12개 구간 범주화	
납부방법	4개 구간 범주화 (미적용)	
회선상태	2개 구간 범주화 (미적용)	
남은할부원금	14개 구간 범주화	
가입회선수	23개 구간 범주화	
tablet보유여부	2개 구간 범주화 (미적용)	
smart watch보유여부	2개 구간 범주화 (미적용)	
멤버쉽 사용금액 (월)	56개 구간 범주화	
멤버쉽 사용금액 (년)	46개 구간 범주화	
미납횟수	4개 구간 범주화	
한화생명		
대상	비식별 조치 방법	비고
직업	26개 구간 범주화	준식별자 (k-익명성[10] 적용)
신용대출건수	16개 구간 범주화 및 교환	민감정보 (l-다양성[2] 적용)
최초계약날짜	210개 구간 범주화 (미적용)	
최초연체등록날짜	210개 구간 범주화 (미적용)	
총신용대출금액	159개 구간 범주화 및 코드화	
총상환금액	139개 구간 범주화 및 코드화	
신용대출연체율	소수점 범주화 (미적용)	
신용대출연체률(30일이상)	소수점 범주화 (미적용)	
신용대출연체율(최근1년)	소수점 범주화 (미적용)	
최초신용등급	항목명칭변경 및 교환 (미적용)	
최근신용등급	항목명칭변경 및 교환 (미적용)	
보험료연체율	소수점 범주화 (미적용)	
최근1년 보험연체율	소수점 범주화 (미적용)	
실효해지건수	15개 구간 범주화	
기납입보험료	77개 구간 범주화 및 코드화	
납입보험료	38개 구간 범주화 및 코드화	
추정소득	53개 구간 범주화 및 코드화	
가구추정소득	103개 구간 범주화 및 코드화	
평균약관대출율	소수점 범주화 (미적용)	
약간대출금액	151개 구간 범주화 및 코드화	
자동이체실패월수	-	

* 출처: 권은희 의원, 추혜선 의원(이상 2017년 국정감사), SK텔레콤, "개인정보 비식별 자료 생성·유통의 현장적용을 위한 실증 최종 보고서", 미래창조과학부 미래 성장 동력 플래그십 프로젝트 사업 제2016연도 최종보고서(2017. 4. 30.) 종합.

SK텔레콤은 한화생명과 결합하여 항목이 42개로 확장된 데이터셋 결과물을 돌려받아 비식별 조치를 시행하였다. 준식별자인 나이, 성별, 직업 등 3개 컬럼은 범주화를 하고 민감정보인 24개 컬럼은 범주화를, 15개 컬럼은 결합 원본값과 동일하였다. 전체 k-익명성 값은 5를, 당월 연체금액에 적용한 l-다양성 값은 2였다.

한화생명은 이런 데이터 결합을 통해 신용등급을 세분화하여 고객의 재정심사, 가입 여력 산출, 소득추정, 대출심사 등에 활용할 수 있을 것으로 보았다. 또한, 신용대출 연체 고객의 특성을 분석하면 연체율, 연체금액, 연체 시점 예측 및 보험계약 해지 가능성을 발견하고 통신사 보유 신용 관련 정보(연체, 미납, 이용정지, 연체금액 등)를 활용한 연체 고객 관리가 가능할 것으로 보았다. 나아가 통신사와 중복된 고객의 규모 및 가입성향을 파악하기 위해 통신사 정보(멤버쉽 등)와의 연관성도 분석하고 향후 각사 제휴 하에 고객 동의에 기반한 식별정보도 송수신함으로써 보험 업무에 활용할 수 있기를 기대하였다.

(2) 한국정보화진흥원

한국정보화진흥원은 정보화 촉진 기본법(현 국가정보화 기본법)에 의해 2009년 설립된 과학기술정보통신부 산하 정부출연기관이다. 한국정보화진흥원은 2016년 8월 미래창조과학부(현 과학기술정보통신부)로부터 통신 등 과학기술정보통신부 소관 산업계의 정보 집합물 결합을 지원하는 전문기관으로 지정되었다.

한국정보화진흥원은 전문기관으로 지정된 2016년 8월부터 2017년 9월까지 정보 집합물 결합을 3건 실시하였다.

우선 사례1 정보통신 분야 LG CNS-LG 유플러스 간 결합 사례의 경우, LG CNS의 약 109만 건 데이터와 LG 유플러스의 약 97만 건 데이터를 결합하여, 그 결과 LG CNS 데이터의 98.97%에 해당하는 약 96만 건에 대한 결합이 이루어졌다. 결합데이터 활용 계획은 개인별 소비 성향(통신데이터, 지역 데이터 + 소비데이터)을 파악하고, 가구 특정 지수를 교차 검증하여 분석 정확도를 향상시키는 데 있었다. 1인 가구 고객 대상 프로모션 등 신규 인사이트도 발굴할 예정이다.[335] 비식별 조치는 다음과 같이 이루어졌다. LG 유플러스의 경우 원본 데이터에서 식별자에 해당하는 휴대전화 번호는 결합 항목에서 삭제하고 임시 대체키로 사용하였으며, 준식별자인 주소도 삭제하였다. 그러나 나머지 항목인 가구 Stage, 영유아 보육지수, 학생 보육지수, 1인 가구 지수는 통계값이기 때문에 비식별 조치를 적용하지 않았다. LG CNS의 경우 임

335) 한국정보화진흥원 발표자료, 개인정보 보호위원회 답변자료, 추혜선 의원 국정감사 자료 종합.

시 대체키인 휴대전화번호는 항목에서 삭제하였으며, 나이와 성별은 범주화하고 6-익명성을 적용하였다. 나머지 항목인 1인 가구 지수, 혼밥지수, 생필품지수, 유아동지수, 구매력지수는 속성자이지만 통계값이기 때문에 비식별 조치를 적용하지 않았다.

표 4-54 한국정보화진흥원 결합 사례

연번	결합일시	신청기관명	내용	건수	결합건수 (결합률)
1	2016년 11월	결합목적	지역별, 연령대별 등 소비성향 파악과 각 기업이 추정하는 가구 특성 지수의 정확도 향상		
		LG CNS	온라인결제데이터 (온라인 상품 구매 및 결제 데이터, 추정 1인 가구 지수 등)	약 97만건	96만건 (약 98.97%)
		LG 유플러스	통신사 고객데이터 (주소, 통신이력, 보육지수, 1인가구지수 등)	약 109만건	96만건 (약 88.07%)
		임시 대체키	각사 고객의 휴대폰번호		
2	2017년 1월	결합목적	소비성향 파악과 소비성향에 따른 구매 상품과의 연관성 연구		
		BC카드	카드결제데이터 (매출, 소득, 결제이력, 소비성향 등 추정 지수 등)	약 5만건	5만 2천여 건 (약 100%)
		W홈쇼핑	구매데이터 (회원정보, 구매 상품이력, 구매 경로 등)	약 7만건	5만 2천여 건 (약 71.43%)
		임시 대체키	승인번호, 승인일시		
3	2017년 7월	결합목적	2017 빅콘테스트 대회 개최를 위한 데이터 제공 및 3사간 공통고객 성향분석		
		SK텔레콤	통신데이터 (이동전화 가입자 통신, 미수납 이력 등)	29,000,573건	248만건 (약 8.55%)
		한화생명	수납데이터 (보험 수납 및 대출 이력 등)	약 917만건	248만건 (약 27.04%)
		SCI평가정보	신용데이터 (대출 및 연체 등 금융정보)	약 3,700만건	248만건 (약 6.70%)
		임시 대체키	생년월일, 성별, 이름		

* 출처: 김성수 의원(2017년 국정감사), 진선미 의원(2017년 국정감사), 개인정보 보호위원회 답변자료 종합[336]

336) 2017년 7월 개인정보 보호위원회는 분야별 전문기관에 현황관련 질의를 보내 답변자료를 받았다.

표 4-55 LG 유플러스 - LG CNS 결합 항목 비식별 조치

LG 유플러스		
대상	비식별 조치 방법	원본 항목
가구 Stage	미적용	1depth: 01, 02, 03, 04, 05 2depth: 01XX, 02XX, etc.
영유아 보육지수	미적용	1depth: 1000 단위 2depth: 100 단위 3depth: 10 단위
학생 보육지수	미적용	1depth: 1000 단위 2depth: 100 단위 3depth: 10 단위
1인가구지수	미적용	1depth: 1000 단위 2depth: 100 단위 3depth: 10 단위
LG CNS		
대상	비식별 조치 방법	원본 항목
나이	k-익명성[6] 적용	10중반 ~ OO 후반
성별	k-익명성[6] 적용	0, 1, 2
1인가구지수	미적용	1000~11000
혼밥지수	미적용	1000~11000
생필품지수	미적용	1000~11000
유아동지수	미적용	1000~11000
구매력지수	미적용	1000~11000

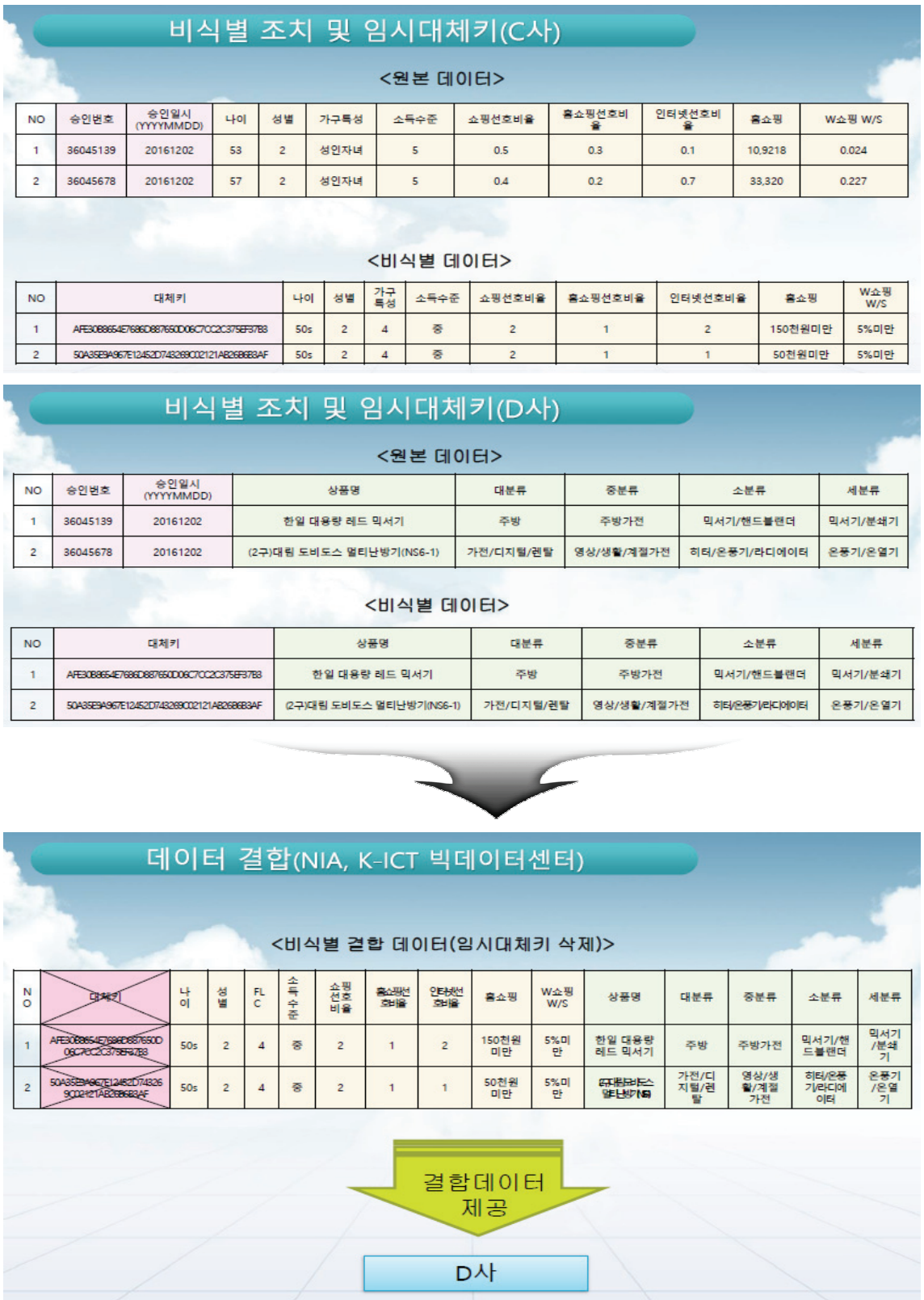
* 출처: 추혜선 의원(2017년 국정감사).

표 4-56 W홈쇼핑 - BC카드 결합 항목 비식별 조치

W홈쇼핑		
대상	비식별 조치 방법	원본 항목
상품명	k-익명성[3] 적용	고유상품명
대분류	k-익명성[3] 적용	총9개
중분류	k-익명성[3] 적용	총30개
소분류	k-익명성[3] 적용	총125개
세분류	k-익명성[3] 적용	총224개
BC카드		
대상	비식별 조치 방법	원본 항목
나이	k-익명성[3] 적용	6가지 (30대이하~70대이상)
성별	k-익명성[3] 적용	2가지 (1: 남 / 2: 여)
FLC	k-익명성[3] 적용	5가지 (1: 1인가구 / 2: 신혼영유아 / 3: 초중고자녀 / 4: 성인자녀 / 5: 노인가구)
예상소득	k-익명성[3] 적용	3가지 (고/중/저)
쇼핑선호비율	k-익명성[3] 적용	3가지 (1: 상위 20% / 2: 나머지 80% / 3: 값없음)
홈쇼핑선호비율	k-익명성[3] 적용	3가지 (1: 상위 20% / 2: 나머지 80% / 3: 값없음)
인터넷선호비율	k-익명성[3] 적용	3가지 (1: 상위 20% / 2: 나머지 80% / 3: 값없음)
홈쇼핑건당이용액	k-익명성[3] 적용	7가지 (0 / 50천원미만 / 100천원미만 / 150천원미만 / 200천원미만 / 500천원미만 / 50만원이상)
W쇼핑 Wallet Share	k-익명성[3] 적용	20가지 (0% / 1% / ~10% ... 90% 이상)

* 출처: 추혜선 의원(2017년 국정감사).

그림 4-57 BC카드-W홈쇼핑 결합 사례 비식별 조치 및 임시 대체키 생성



* 출처: 김배현, 앞의 자료, p8~p10 종합.

** C사는 BC카드, D사는 W홈쇼핑을 의미함

사례2 유통 분야 BC카드와 W홈쇼핑 간의 결합 사례는 다음과 같이 이루어졌다. BC카드의 카드사용 데이터를 비식별 조치하고 W홈쇼핑의 구매고객 데이터도 비식별 조치한 후, 승인번호와 승인 일자에 기반하여 임시 대체키를 생성하여 결합시켰다. 이때 BC카드의 결합대상인 카드사용 회원 데이터 항목은 나이, 성별, 가구유형, 예상소득, 쇼핑·홈쇼핑·인터넷 선호비율, 홈쇼핑건당이용액 등이었고, W홈쇼핑의 결합대상인 고객 구매데이터 항목은 상품명과 상품분류 등이었다. BC카드의 약 5만 건 데이터와 W홈쇼핑의 약 7만 건 데이터를 결합하여, 그 결과 BC카드 데이터의 100%에 해당하는 약 5만 2천여 건에 대한 결합이 이루어졌다. 결합데이터 활용 계획은 특정 성향 고객 대상으로 양사가 공동으로 프로모션을 진행하고, 신규고객 유치 및 상품구성 등 마케팅 전략에 활용하는 데 있었다.

이 사례에서 BC카드와 W홈쇼핑 데이터의 비식별 조치는 <표 4-56> 및 <그림 4-57>과 같이 이루어졌다. BC카드의 경우 식별자인 주민등록번호, 성명 등을 삭제하고, 속성자인 나머지 항목은 범주화하였다. W홈쇼핑의 경우 식별자인 주민등록번호, 성명 등은 삭제하고, 상품명과 상품의 대/중/소/세분류는 비식별 조치를 사실상 미적용하였다.

사례3 SK텔레콤-한화생명-SCI평가정보 간의 결합 사례는 <표 4-57>과 같이 이루어졌다. SK텔레콤의 통신데이터(이동전화 가입자 통신, 미수납 이력 등 20개 항목을 비식별 조치하고 한화생명의 수납데이터(보험 수납 및 대출 이력 등 35개 항목) 및 SCI평가정보 신용데이터(대출 및 연체 등 금융정보 등 20개 항목)도 비식별 조치한 후, 생년월일, 성별, 이름에 기반하여 임시 대체키를 생성하여 결합시켰다. 이때 SK텔레콤의 약 2,900만 건 데이터와 한화생명의 약 917만 건 및 SCI평가정보의 약 3,700만 건 데이터가 결합하여, 그 결과 248만 건에 대한 결합이 이루어졌다.

이 결합의 목적은 2017 빅콘테스트 대회 개최를 위한 데이터 제공 및 3사 간 공통 고객 성향을 분석하는 데 있었다. 특히 SK텔레콤은 2017년 4월 11일 이 사례를 발표하면서 이 데이터 결합으로 통신데이터의 보험사 활용의 의미와 신용등급 상관 관계성을 검증할 수 있었다고 평가하였다. 제공된 데이터 중 멤버십의 등급, 1년간 누적 연체금액, 통화 빈도 등이 유효 변수로 나타났으며 신용등급 산정에 기존 데이터 외 통신데이터가 반영될 때 유의미성이 발견되었다는 것이다. 향후에도 결합데이터로 3사 간 공통고객 성향분석에서 유의미한 값 도출이 가능할 경우 추가 비식별 조치 및 결합 등 통해 빅데이터를 분석하겠다는 계획도 있다.[337]

337) 개인정보 보호위원회 답변자료.

표 4-57 SK텔레콤-한화생명-SCI평가정보 결합 항목 비식별 조치

SK텔레콤		
대상	비식별 조치 방법	원본 항목
생년월일	범주화 (나이) k-익명성 (18세미만, 70세이상)	19000204~20151024
성별코드	k-익명성	1(남), 2(여)
월평균통화시간		0 ~ 55,979분
월평균통화빈도		0 ~ 7,479회
멤버십등급		G(Gold), O(일반), S(실버), V(VIP), 정보없음
태블릿보유여부		Y, N
스마트와치보유여부		Y, N
가입자평균매출		-5,157,175원~7,521,407원
당월납부요금		0~8,315,310원
결합상품가입여부		Y, N
단말기출고가		0~1,793,000원
서비스가입일자	범주화 (서비스가입월)	19840402~20160430
정지일수		0일~5,881일
당월연체유무	l-다양성	Y, N
당월연체금액	l-다양성	0~20,609,060원
최근1년간납부일미준수횟수		0회~12회
최근1년간최대연체금액	l-다양성 (300만원 이상 고액연체자)	0원~20,609,060원
납부방법		은행자동납부, 입금전용계좌, 지로납부, 카드자동납부, 정보없음
회선상태		사용중, 정지, 정보없음
남은할부원금		0~1,308,976
한화생명		
대상	비식별 조치 방법	원본 항목
직업	범주화, k-익명성[5]	대분류 범주화(26가지)
추정소득	라운딩, 범주화	백만 단위 범주화(53개 구간)
가구추정소득	라운딩, 범주화	백만 단위 범주화(103개 구간)
실가족원수	라운딩, 범주화	이상치 범주화(8명 이상)
보험가입가족원수	라운딩, 범주화	이상치 범주화(5명 이상)
막내자녀나이	범주화, k-익명성[5]	5세 단위 범주화(14개 구간)
배우자직업	범주화, k-익명성[5]	대분류 범주화(26가지)
배우자추정소득	라운딩, 범주화	백만 단위 범주화(52개 구간)
신용대출건수	범주화	이상치 범주화(11건 이상)
최초대출날짜	범주화	월단위 범주화(201개 구간)
총신용대출금액	라운딩, 범주화	백만 단위 범주화(50개 구간)
총상환금액	라운딩, 범주화	백만 단위 범주화(50개 구간)
신용대출연체율	라운딩	정수 첫째자리 라운딩, 범주화(99개 구간)
30일이내신용대출연체율	라운딩	정수 첫째자리 라운딩, 범주화(35개 구간)
최근1년신용대출연체율	라운딩	정수 첫째자리 라운딩, 범주화(11개 구간)
최초신용등급	미적용, l-다양성[2]	0~7 (동일)
최근신용등급	미적용, l-다양성[2]	0~10 (동일)
보험료연체율	라운딩	정수 첫째자리 라운딩, 범주화(93개 구간)
최근1년보험료연체율	라운딩	정수 첫째자리 라운딩, 범주화(91개 구간)
평균약대율	라운딩	정수 첫째자리 라운딩, 범주화(101개 구간)

약관대출가능잔액	라운딩, 범주화	백만 단위 범주화(54개 구간)
최근1년약대금액	라운딩, 범주화	백만 단위 범주화(58개 구간)
최근1년약대연체율	라운딩	정수 첫째자리 라운딩, 범주화(10개 구간)
연금저축상품월납입보험료	라운딩, 범주화	만 단위 범주화(34개 구간)
非연금저출상품월납입보험료	라운딩, 범주화	만 단위 범주화(34개 구간)
가구연금저축상품월납입보험료	라운딩, 범주화	만 단위 범주화(34개 구간)
가구非연금저축상품월납입보험료	라운딩, 범주화	만 단위 범주화(36개 구간)
최대월납입보험료	라운딩, 범주화	만 단위 범주화(37개 구간)
기납입보험료	라운딩, 범주화	십만 단위 범주화(78개 구간)
가구기납입보험료	라운딩, 범주화	십만 단위 범주화(78개 구간)
실효해지건수	범주화	이상치 범주화(7건 이상)
최근1년 실효해지건수	범주화	이상치 범주화(7건 이상)
자동이체실패월수	범주화	이상치 범주화(61개월 이상)
가구총지급보험금액	라운딩, 범주화	십만 단위 범주화(463개 구간)
가구총보험금청구건수	범주화	범주화(76개 구간)

SCI평가정보		
대상	비식별 조치 방법	비고
성별	남 1, 여 2	k-익명성
나이	총10가지(범주화) : 5세 범주 * <18, >70 삭제	k-익명성
평점	총69가지(범주화) : 10점 범주	l-다양성
연체등정보 현재총건수	총20가지(범주화)	
연체등정보 현재총금액	총209가지(범주화) : 100만권 범주, 300만원 범주, 500만원 범주	
대출정보 현재총건수	총11가지(범주화)	
대출정보 현재총건수 [은행]	총6가지(범주화)	
대출정보 현재총건수 [카드사/할부사/캐피탈]	총6가지(범주화)	
대출정보 현재총건수 [2산업분류]	총8가지(범주화)	
대출정보 현재총건수 [기타]	총8가지(범주화)	
대출정보 현재총금액	총219가지(범주화) : 300만원 범주, 1000만원 범주	
대출정보 현재총금액 [신용대출]	총219가지(범주화) : 300만원 범주, 1000만원 범주	
대출정보 현재총금액 [은행]	총219가지(범주화) : 300만원 범주, 1000만원 범주	
대출정보 현재총금액 [카드사/할부사/캐피탈]	총79가지(범주화) : 300만원 범주, 1000만원 범주	
대출정보 최근개설일로부터 현재까지 유지기간 [신용대출]	총12가지(범주화) : 12개월 범주	
대출정보 최근개설일로부터 현재까지 유지기간 [2산업분류-신용대출]	총12가지(범주화) : 12개월 범주	
개설정보 현재신용개설총건수 [신용카드]	총12가지(범주화)	
개설정보 최초개설일로부터 현재까지유지기간 [신용카드]	총12가지(범주화) : 12개월 범주	
보증정보 현재보증총건수	총11가지(범주화)	
보증정보 현재보증총금액	총171가지(범주화) : 300만원 범주, 1000만원 범주	

* 출처: 추혜선 의원(2017년 국정감사).

(3) 금융보안원

금융보안원은 카드 3사 개인정보 유출 사고 이후 금융보안 전담기구 설립에 관한 금융위원회 업무계획에 의해 설립된 비영리 사단법인으로, 금융결제원 금융ISAC, 코스콤 증권ISAC, 금융보안연구원을 통합하여 2015년 설립되었다. 2017년 8월 현재 190개 기관(은행, 금융투자회사, 보험사, 카드사, 금융공공기관 등)이 사원기관으로 참여하고 있다. 정보통신기반 보호법에 따른 금융정보공유·분석센터(ISAC)를 운영하고 있으며 전자금융감독규정에 따른 침해사고 대응기관이자 정보통신망법에 따른 정보보호 관리체계(ISMS) 인증기관이다. 금융권 개인정보 비식별 조치 및 빅데이터 분석·활용 활성화를 목적으로 금융빅데이터협의회를 운영하고 있다.

금융보안원은 2016년 8월 금융위원회로부터 금융회사(은행, 금융투자, 신용카드, 보험, 신용정보회사 등), 전자금융업자, 핀테크기업 등 금융권 기업을 대상으로 정보 집합물 결합을 지원하는 전문기관으로 지정되었다.

금융보안원은 전문기관으로 지정된 2016년 8월부터 2017년 9월까지 정보 집합물 결합을 7건 실시하였다.

표 4-58 금융보안원 결합 사례

연번	구분	신청기관	상대기관
1	명칭	한국주택금융공사	주택도시보증공사
	데이터 수(건)	150,036건	364,248건
	결합건(결합률)	6,680건(4.45%)	6,680건(1.83%)
2	명칭	NICE평가정보	그릿연구소
	데이터 수(건)	5,000건	4,668건
	결합건(결합률)	3,900건(78%)	3,900건(83.5%)
3	명칭	신한카드	코리아크레딧뷰로
	데이터 수(건)	106,562건	125,192건
	결합건(결합률)	46,157건(43.3%)	46,157건(43.3%)
4	명칭	신한카드	코리아크레딧뷰로
	데이터 수(건)	28,862건	106,680건
	결합건(결합률)	20,519건(71.1%)	20,519건(19.2%)
5	명칭	KB국민카드	LG유플러스
	데이터 수(건)	18,267,641건	6,608,917건
	결합건(결합률)	2,497,714건(7.31%)	2,497,714건(37.8%)
6	명칭	보험개발원	현대자동차
	데이터 수(건)	154,640,520건	5,545,708건
	결합건(결합률)	154,640,520건(100%)	154,640,520건(2,788.5%)
7	명칭	보험개발원	현대자동차
	데이터 수(건)	154,640,520건	5,545,708건
	결합건(결합률)	154,640,520건(100%)	154,640,520건(2,788.5%)

* 출처: 추혜선 의원(2017년 국정감사), 진선미 의원(2017년 국정감사) 종합.

(4) 한국신용정보원

한국신용정보원은 전국은행연합회, 한국금융투자협회, 생명보험협회, 손해보험협회, 한국여신전문금융업협회 등 5개 금융협회 및 보험개발원에서 분산하여 관리하던 신용정보를 집중관리하고, 이를 효율적으로 활용하기 위해 '신용정보의 이용 및 보호에 관한 법률'에 따라 2016년 1월에 설립된 종합신용정보집중기관이며 사단법인이다. 특히 금융분야 전문기관 두 곳 중 신용정보원의 경우, 신용정보원이 보유한 전 업권의 신용정보를 빅데이터 분석에 활용하는 데 주안점을 두고 있다.[338]

한국신용정보원은 2016년 8월 금융위원회로부터 금융회사를 대상으로 정보 집합물 결합을 지원하는 전문기관으로 지정되었다.

한국신용정보원은 전문기관으로 지정된 2016년 8월부터 2017년 9월까지 정보 집합물 결합을 15건 실시하였다.

표 4-59 한국신용정보원 결합 사례

연번	결합일시	구분	신청기관	상대기관
1	2017년 1월	결합목적	신용평가 모델개발 (이종산업간)	
		명칭	NICE평가정보	KT
		데이터 수(건)	2,903,595건	13,961,710건
		결합건(결합률)	712,842건(24.6%)	712,842건(5.1%)
2	2017년 1월	결합목적	공통가입고객 성향분석 (금융분야)	
		명칭	한화손해보험	한화생명보험
		데이터 수(건)	3,955,524건	7,108,800건
		결합건(결합률)	878,749건(22.2%)	878,749건(12.4%)
3	2017년 2월	결합목적	공통가입고객 성향분석 (금융분야)	
		명칭	삼성생명	삼성카드
		데이터 수(건)	7,630,803건	8,466,576건
		결합건(결합률)	2,345,867건(30.7%)	2,345,867건(27.7%)
4	2017년 2월	결합목적	공통가입고객 성향분석 (금융분야)	
		명칭	삼성생명	삼성카드
		데이터 수(건)	7,579,973건	8,466,576건
		결합건(결합률)	2,321,835건(30.6%)	2,321,835건(27.4%)
5	2017년 2월	결합목적	공통가입고객 성향분석 (금융분야)	
		명칭	삼성생명	삼성카드
		데이터 수(건)	7,644,495건	8,466,576건
		결합건(결합률)	2,349,649건(30.7%)	2,349,649건(27.8%)
6	2017년 2월	결합목적	공통가입고객 성향분석 (금융분야)	
		명칭	삼성생명	삼성카드
		데이터 수(건)	7,555,249건	8,466,576건
		결합건(결합률)	2,317,027건(30.7%)	2,317,027건(27.4%)

338) "임종룡 금융위원장, 금융권 빅데이터 지원 전문기관 지정 등 빅데이터 활성화를 위한 조찬간담회 개최", 금융위원회 보도자료(2016. 8. 31).

연번	결합일시	구분	신청기관	상대기관
7	2017년 2월	결합목적	공통가입고객 성향분석 (금융분야)	
		명칭	삼성생명	삼성카드
		데이터 수(건)	8,019,019건	428,293건
		결합건(결합률)	107,073건(1.3%)	107,073건(25.0%)
8	2017년 2월	결합목적	공통가입고객 성향분석 (금융분야)	
		명칭	삼성생명	삼성카드
		데이터 수(건)	8,014,487건	8,106,386건
		결합건(결합률)	2,313,824건(28.9%)	2,313,824건(28.5%)
9	2017년 2월	결합목적	공통가입고객 성향분석 (금융분야)	
		명칭	삼성생명	삼성카드
		데이터 수(건)	8,014,487건	8,284,128건
		결합건(결합률)	2,356,213건(29.4%)	2,356,213건(28.4%)
10	2017년 2월	결합목적	공통가입고객 성향분석 (금융분야)	
		명칭	삼성생명	삼성카드
		데이터 수(건)	8,014,487건	8,466,263건
		결합건(결합률)	2,409,998건(30.1%)	2,409,998건(28.5%)
11	2017년 2월	결합목적	공통가입고객 성향분석 (금융분야)	
		명칭	삼성생명	삼성카드
		데이터 수(건)	8,014,487건	8,465,547건
		결합건(결합률)	2,409,992건(30.1%)	2,409,992건(28.5%)
12	2017년 2월	결합목적	공통가입고객 성향분석 (금융분야)	
		명칭	삼성생명	삼성카드
		데이터 수(건)	8,014,487건	7,855,900건
		결합건(결합률)	2,237,917건(30.1%)	2,237,917건(28.5%)
13	2017년 2월	결합목적	공통가입고객 성향분석 (금융분야)	
		명칭	삼성생명	삼성카드
		데이터 수(건)	8,014,487건	8,466,581건
		결합건(결합률)	2,410,101건(30.1%)	2,410,101건(28.5%)
14	2017년 2월	결합목적	공통가입고객 성향분석 (금융분야)	
		명칭	삼성생명	삼성카드
		데이터 수(건)	8,014,487건	8,466,581건
		결합건(결합률)	2,410,101건(30.1%)	2,410,101건(28.5%)
15	2017년 2월	결합목적	공통가입고객 성향분석 (금융분야)	
		명칭	삼성생명	삼성카드
		데이터 수(건)	8,014,487건	8,462,743건
		결합건(결합률)	2,409,211건(30.1%)	2,409,211건(28.5%)

* 출처: 추혜선 의원(2017년 국정감사)

이에 대한 보다 자세한 사항은 2017년 4월 11일 개인정보 비식별 조치 및 결합지원 서비스 설명회에서 발표된 사례에서 다음과 같이 살펴볼 수 있다.

우선 결합 사례 1은 이종산업인 금융정보와 통신정보를 결합하여 신용평가 모델을 개발하기 위한 목적으로 수행되었다. 통신사인 KT와 신용정보사인 NICE평가정보의 보유 정보를 비식별 처리 후 결합시킨 데이터를 □□ 인터넷은행에 제공 및 분석하였다. 결합에 사용된 금융정보의 주요 속성 정보로는 신용 스코어, 연체 정보, 대출 정

보, 카드이용 정보 등이 있었으며 범주화를 활용하여 비식별화를 진행하였다. 결합에 사용된 통신정보의 주요 속성 정보로는 통신 등급, 미납/수납 정보 등이 있었으며, 범주화를 활용하여 비식별화를 진행하였다.

사례 2는 한화생명 및 한화손해보험 공통가입 고객의 보험가입 성향분석을 위한 목적에서 이루어졌다. 결합에 사용된 손해보험 주요 속성 정보로는 장기보험 피보험자 고객정보 및 자동차 배기량, 외제 차 여부, 차량 연식, 차량가액, 보유 대수 등 자동차 보험 피보험자 고객정보 등이 있었으며 <그림 4-58>과 같이 범주화를 활용하여 비식별화를 진행하였다. 결합에 사용된 생명보험 정보 주요 속성 정보로는 가입 고객의 인구통계 정보 및 보험가입 현황 정보, 공통가입 고객의 보험가입 성향 등이 있었다.

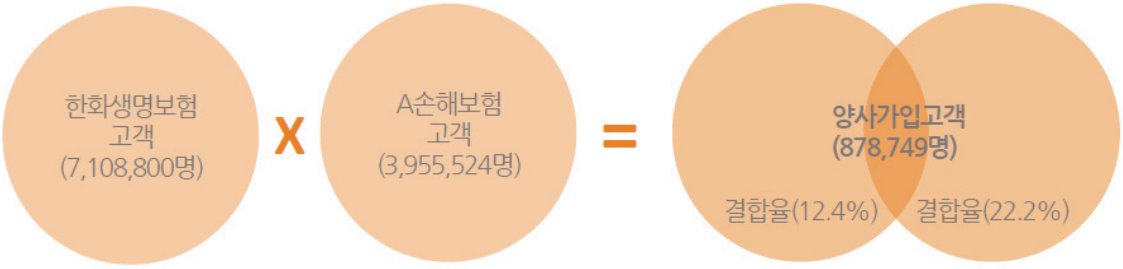
그림 4-58 한화생명보험-한화손해보험 결합을 위한 데이터 비식별 조치 사례

구분	연령	성별	직업코드	주소정보	장기월납환산보험료	장기계약가입담보	차량가액	.....
식별자 구분	준식별자				민감정보			
비식별 조치 기준치	K-익명성 (50이상)				L-다양성 (40이상)			
비식별 조치 기법	범주화 (연령 단위별)		범주화 (직업 구분별)	범주화 (지역별)	범주화 (만원 단위)	범주화 (금액 단위)	범주화 (백만 단위)	.....

* 한국신용정보원, "금융 개인정보 비식별 조치 현황 및 데이터 분석 사례", 개인정보 비식별 조치 및 결합지원 서비스 설명회(2017. 4. 11), p10.

이에 대한 보다 자세한 사항은 2017년 4월 11일 개인정보 비식별 조치 및 결합지원 서비스 설명회에서 한화생명이 발표한 자료에서 다음과 같이 살펴볼 수 있다.

그림 4-59 한화생명보험-한화손해보험 데이터 결합 건수



* 출처: 전도현, 앞의 자료, p11.
** A손해보험은 한화손해보험을 의미함

한화생명보험 고객 7,108,800명과 한화손해보험 고객 3,955,524명의 데이터를 결합한 결과, 양사 가입 고객 878,749명이 결합되었다. 즉, 한화생명보험 고객 12.4%의 데이터가 결합되고 한화손해보험 고객 22.2%의 데이터가 결합되었다.

임시 대체키는 생년월일+성+이름을 이용하여 생성되었고, 데이터 항목으로는 보험 종류별 가입 건수, 가입금액, 가입채널 등 양사 공통항목 외에 양사가 필요한 고유 항목을 더하였다. 그 결과 한화생명 19개 항목과 한화손해보험 41개 항목, 총 60개 항목이 결합되었다. 그 자세한 목록은 <그림 4-60>과 같다.

그림 4-60 한화생명보험-한화손해보험 데이터 결합 항목

한화생명	연령	직업	종신보험 가입건수	보장보험 가입건수	연금보험 가입건수	저축보험 가입건수	합계 월납입보험료
	가구단위합계 월납입보험료	최근가입채널	일반사망 보장금액	일반암진단 보장금액	뇌출혈진단 보장금액	급성심근경색진단 보장금액	실손질병입통원 가입여부
	실손상해입통원 가입여부	질병입원 보장금금액	**직업기반 추정소득금액**	**주소기반 추정소득금액**	**추정주택가격**		
A손해보험	성별	거주지	거주지 등록년도	장기월납 환산보험료	장기_종합형 가입건수	장기_운전자 가입건수	장기_어린이 가입건수
	장기_질병 가입건수	장기_암 가입건수	장기_치아 가입건수	장기_간병 가입건수	장기_상해 가입건수	장기_상조 가입건수	장기_재물 가입건수
	장기_저축 가입건수	장기_연금 가입건수	상해사망담보 가입금액	암진단담보 가입금액	질병입원담보 가입금액	질병통원담보 가입금액	상해입원담보 가입금액
	상해통원담보 가입금액	일반상해의료비 가입금액	질병일당담보 가입금액	질병사망담보 가입금액	상해후유담보 가입금액	질병후유담보 가입금액	뇌진단담보 가입금액
	심장진단담보 가입금액	장기요양진단담보 가입금액	재물보험 가입금액	장기보험 최근가입년도	장기보험 최근가입채널	**자동차 배기량**	**자동차 외제차여부**
	자동차 차량연식	**자동차 차량가액**	**자동차 차량용도**	**자동차 보유차량대수**	**자동차 최근가입년도**	**손보 핵심고객여부**	

준식별자 민감정보 각사별 고유항목

* 출처: 전도현, 앞의 자료, p14.
** A손해보험은 한화손해보험을 의미함

이 결합을 통해 한화생명은 생명보험과 손해보험 동시 가입 고객의 분포 비교 및 특성도출을 통한 마케팅 활용 가능성을 기대하였다. 즉, 성, 연령, 지역, 직업별 고객 분포 비교 분석을 통한 통찰력을 도출하고 추후 해당 고객들에 대한 개인정보 동의를 통해 개인 단위 데이터 교환 시 양사 간 시너지 창출도 기대하였다. 또한, 생명보험과 손해보험 동시 가입 고객에 대해 교차 판매 채널 관련 분석도 시행할 수 있을 것으로 보았다. 더불어 한화생명이 미보유했던 데이터 항목에 대해서도 결합을 통해 분석이 가능할 것으로 기대하였다. 예를 들어 손해보험에서 자동차 가액, 외제차 보유 여부, 자동차 배기량 등 자동차 관련 변수를 생명보험 고객 세그먼트 마케팅에 활용할 수 있을 것으로 보았다.

한편, 사례 3부터 15를 차지하는 삼성생명-삼성카드 결합은 하루 동안 13차례에 걸

쳐 삼성생명과 삼성카드 고객정보가 결합되었으며 결합성공건수는 28,398,808건에 달한다. 그 결합 목적으로는 보험고객에 대한 카드이용특성 분석(삼성생명), 카드고객에 대한 보험고객 이용고객 분석 및 통계모델 개발(삼성카드)을 들고 있어, 전반적으로 공통가입 고객 분석을 목표로 하고 있다. 이를 위하여 삼성생명 보험고객의 보험가입 속성 정보(가입 건수, 보험료, 가입 기간, 가입상품, 연령 등)와 삼성카드 회원의 내부 전략변수 항목 및 카드이용실적정보에 대하여 결합이 이루어졌다.

(5) 사회보장정보원

사회보장정보원은 2009년 한국보건복지정보개발원을 전신으로 하고 2015년 사회보장급여의 이용·제공 및 수급권자 발굴에 관한 법률에 근거하여 출범하였다. 보건복지부 산하 위탁집행형 준정부기관으로서 보건복지 정보의 수집·제공과 보건복지 관련 정보시스템 개발 및 운영 등 보건복지 정보화사업을 수행하고 있다.

사회보장정보원은 2016년 8월 보건복지부로부터 보건복지 분야 민간 및 공공기관 대상으로 정보 집합물 결합을 지원하는 전문기관으로 지정되었다. 다만 발족 당시에는 2017년 결합지원시스템 구축 시까지 타 전문기관 의뢰 방침을 밝혔다. 사회보장정보원이 전문기관으로 지정된 2016년 8월부터 2017년 9월까지 실제 정보 집합물을 결합한 사례는 없다.

(6) 한국교육학술정보원

한국교육학술정보원은 교육부 산하 정부출연기관으로, 1999년 한국교육학술정보원법에 의해 설립되어 에듀넷, 학술연구정보서비스 RISS, 교육행정정보시스템 NEIS 등 교육 및 학술연구 정보 서비스를 운영하고 있다. 한국교육학술정보원은 2016년 12월 교육부로부터 교육 분야에 대한 기관 및 기업 간 정보 집합물 결합을 지원하는 전문기관으로 지정되었다.

한국교육학술정보원은 전문기관 지정 이후 2017년 4월부터 전문기관 운영이 시작되어 2017년 9월까지 실제 정보 집합물을 결합한 사례는 없다. 개인정보 보호위원회 답변자료에서 한국인터넷진흥원, 한국정보화진흥원의 정보 집합물 결합 시스템을 공동으로 활용할 예정임을 밝히고 있다.

(7) 국가정보자원관리원

국가정보자원관리원(구 정부 통합전산센터)은 2005년 발족 이후 행정안전부 소속기관으로 중앙행정기관, 지방자치단체 및 공공기관의 정보시스템과 국가정보통신망 등의 안정적인 운영, 효율적 통합·구축관리와 보호·보안 등에 관한 사항을 관장해 왔고, 2017년 7월 26일 현재 명칭으로 변경되었다.

국가정보자원관리원은 2016년 8월 부처와 지자체 대상의 개인정보 비식별 조치 전문기관으로 지정되었고, 2017년 5월 1일 지자체 담당자 대상 개인정보 비식별 조치 설명회를 개최하고 부처와 지자체에 대한 데이터 결합 및 비식별 조치 지원 계획을 밝혔다.

국가정보자원관리원은 2016년 8월 전문기관 지정 이후 2017년 9월까지 실제 정보 집합물을 결합한 사례가 없다.

4. 평가

이상에서 살펴본 바에 따르면, 지원기관 제도가 시작된 2016년 8월부터 2017년 9월까지 26차례에 걸쳐 총 347,522,005건의 민간 기업의 데이터가 결합된 것으로 나타났다. 국내 전문기관을 통한 데이터 연계·결합 지원제도의 현황과 문제점은 다음과 같이 요약할 수 있다.

첫째, 결합 목적에 있어서 제한을 두고 있지 않다. 결합 목적에 대한 별도의 심사도 이루어지고 있지 않으며, 원칙적으로 모든 신청에 대해 제한 없이 결합이 이루어지고 있었다. 앞서 살펴본 해외 데이터 연계 제도의 경우 공익적 연구 및 통계 목적으로 명확하게 한정하고 그에 대한 심사도 엄격한 데 비하여, 국내 전문기관이 수행한 대부분 사례에서 데이터 결합의 목적은 대출심사나 마케팅 등 민간 기업의 영리적 목적에 할애되어 있다. 이처럼 국가기관 혹은 국가로부터 업무를 위탁받은 전문기관이 민간 기업의 영리적 데이터 결합을 수행하는 데 대한 타당성 문제도 있다.

둘째, 데이터 결합 단계별로 기능 분리가 이루어지고 있지 못하다. 개인정보 보호위원회는 결정(제2017-15-125호)에서 개인의 프라이버시 보호를 위해서는 데이터 연계 과정에서 누구도 관련 데이터에 포함된 정보주체를 알아볼 수 없도록 하여야 하고, 이를 위해서 데이터 연계 절차에 관여하는 각 기관의 기능이 분리되어야 하며 연계되는 데이터의 일부를 보유하고 있는 데이터 제공기관은 해당 과정에 참여할 수 없도록 해야 한다고 지적한 바 있다.

표 4-60 전문기관을 통한 민간기업 결합사례 (2016. 8. ~ 2017. 9.)

연번	전문기관	신청기관 (신청건수)	상대기관 (상대건수)	결합건수
1	한국인터넷진흥원	SK텔레콤 (18,029,816)	한화생명 (4,595,857)	2,185,596
2	한국정보화진흥원	LG CNS (약 970,000)	LG 유플러스 (약 1,090,000)	약 960,000
3		W홈쇼핑 (약 70,000)	BC카드 (약 50,000)	약 50,000
4		SK텔레콤 (29,000,573)	한화생명 (약 9,170,000) SCI평가정보 (약 37,000,000)	약 2,480,000
5	금융보안원	한국주택금융공사 (150,036)	주택도시보증공사 (364,248)	6,680
6		NICE평가정보 (5,000)	그릿연구소 (4,668)	3,900
7		신한카드 (106,562)	코리아크레딧뷰로 (125,192)	46,157
8		신한카드 (28,862)	코리아크레딧뷰로 (106,680)	20,519
9		KB국민카드 (18,267,641)	LG유플러스 (6,608,917)	2,497,714
10		보험개발원 (154,640,520)	현대자동차 (5,545,708)	154,640,520
11		보험개발원 (154,640,520)	현대자동차 (5,545,708)	154,640,520
12	한국신용정보원	NICE평가정보 (2,903,595)	KT (13,961,710)	712,842
13		한화손해보험 (3,955,524)	한화생명보험 (7,108,800)	878,749
14		삼성생명 (7,630,803)	삼성카드 (8,466,576)	2,345,867
15		삼성생명 (7,579,973)	삼성카드 (8,466,576)	2,321,835
16		삼성생명 (7,644,495)	삼성카드 (8,466,576)	2,349,649
17		삼성생명 (7,555,249)	삼성카드 (8,466,576)	2,317,027
18		삼성생명 (8,019,019)	삼성카드 (428,293)	107,073
19		삼성생명 (8,014,487)	삼성카드 (8,106,386)	2,313,824
20		삼성생명 (8,014,487)	삼성카드 (8,284,128)	2,356,213
21		삼성생명 (8,014,487)	삼성카드 (8,466,263)	2,409,998
22		삼성생명 (8,014,487)	삼성카드 (8,465,547)	2,409,992
23		삼성생명 (8,014,487)	삼성카드 (7,855,900)	2,237,917
24		삼성생명 (8,014,487)	삼성카드 (8,466,581)	2,410,101
25		삼성생명 (8,014,487)	삼성카드 (8,466,581)	2,410,101
26		삼성생명 (8,014,487)	삼성카드 (8,462,743)	2,409,211

해외에서도 결합 목적을 충족하기 위하여 이해관계로부터 독립적인 '신뢰받는 제3의 기관(Trusted Third Party, TTP)' 제도를 활용하고 있으며, 데이터 보유기관, 데이터 이용자, 제3의 기관이 기능상 모두 분리되어 있다. 영국 정부는 이 모델의 장점으로 절차 중에 어떤 참여 기관도 전체 데이터셋의 내용이나 식별 데이터를 다룰 수 없다는 점을 들었다. 특히 TTP는 연계키를 생성하는 인덱서(indexer)로서, 공공기관이나 민간기관 중에서 법률에 따라 승인되지만, 데이터 보유기관이 TTP가 될 수는 없다는 것이 원칙이다.[339] 또 해외에서는 연계된 데이터를 데이터 보유기관에 반출하

339) CABINET OFFICE, 2014. "CABINET OFFICE INITIAL DISCUSSION DOCUMENT ON DATA SHARING POLICY FOR PUBLICATION ON http://datasharing.org.uk/"; "Conclusions of civil society and public sector policy discussions on data use in government", http://datasharing.org.uk/wp-content/uploads/sites/2/2015/03/20150327_Conclusions_OPM_paper_Data_final.pdf 종합.

는 경우도 찾아볼 수 없었다. 대개는 연구목적으로 승인된 연계데이터를 반출이 제한된 보안시설에서 해당 연구자에게 열람토록 할 뿐이다.

반면 국내 전문기관 제도의 경우 결합 자체가 데이터 보유기관이 곧 데이터 이용기관으로서 그 이해관계와 매우 밀착되어 있다. 그럼에도 데이터를 보호하기 위한 임시 대체키를 데이터 보유기관이 스스로 생성하여 그 메커니즘을 알도록 하였으며, 데이터 일부를 보유하고 있는 보유기관에 대규모로 결합된 데이터를 반출하고 해당 데이터의 재식별 검사도 데이터 보유기관이 스스로 수행하는 데 맡기고 있었다.

셋째, 투명성이 부족하다. 영국의 경우 승인된 모든 개인과 연구 프로젝트에 대한 등록사항은 공개되는 것이 원칙이다. 이는 심사기준에도 포함되어 있다.[340] 영국은 보건의료 빅데이터 서비스인 care.data가 투명성 부족에 따른 국민적 논란 끝에 2016년 운영이 중단되었다. 영국 정보위원회(ICO)는 이 문제를 언급하며 빅데이터 처리의 투명성 부족이 대중적 신뢰 부족으로 이어져 공공데이터 공유에도 장벽이 될 수 있다고 지적하였다(ICO, 2017: 52.).

반면 국내 전문기관 제도의 경우 별도의 승인 요건이나 절차를 두고 있지 않음에도 데이터 결합 목적이나 결합 기업에 대한 정보가 공개되어 있지 않다.

넷째, 개인정보 보호법을 준수하고 있지 않다. 비식별화 가이드라인의 경우 개인정보 보호법의 보호대상이 되는 '다른 정보와 쉽게 결합하여' 알아볼 수 있는 개인정보를 매우 좁게 해석하였다. 즉, '알아볼 수 있는' 자의 주체를 해당 '정보를 처리하는 자'로, '다른 정보와 쉽게 결합하여'의 의미를 '결합대상이 될 다른 정보의 입수 가능성이 있어야 하고 또 결합 가능성도 높아야 한다'는 것으로 해석하였고 이에 따라 전문기관 또한 기업들이 보유하여 비식별 조치를 취한 데이터셋에 대하여 개인정보가 아닌 것으로 추정하였다. 그러나 법원은 결합의 용이성과 관련하여 "쉽게 다른 정보를 구한다는 의미이기보다는, 구하기 쉬운지 어려운지와는 상관없이 해당 정보와 다른 정보가 특별한 어려움 없이 쉽게 결합하여 특정 개인을 알아볼 수 있게 되는 것"이라 판시하며, 결합 대상인 정보를 쉽게 구하지 못하더라도 결합대상인 정보와 결합 자체에 어려움이 없다면 보호의 대상이 되는 개인정보가 된다는 해석기준을 제시한 바 있다.[341] 자사 보유 데이터셋을 전문기관을 통해 다른 기업의 데이터셋과 결합시켜 돌려받은 기업으로서는 비식별 조치 이전 상태의 데이터셋 원본 또한 가지고 있으며, 임시 대체키를 직접 생성한 상태이므로 해당 데이터셋에서 개인을 알아볼 가능성이 있다고 볼 수 있다. 이런 상태에서 특정한 비식별 조치를 기술적으로 취했다는 이

340) CABINET OFFICE, 2014. "CABINET OFFICE INITIAL DISCUSSION DOCUMENT ON DATA SHARING POLICY FOR PUBLICATION ON http://datasharing.org.uk/".

341) 서울중앙지방법원, 2011.2.23. 선고, 2010고단5343 판결.

유만으로 해당 데이터셋 모두를 개인정보가 아닌 것으로 추정한 것은 개인정보 보호 법제의 적용을 자의적으로 면제한 것이라는 지적을 받는 결과를 낳았다.[342]

영국 디지털경제법은 데이터 공유에 있어서도 개인정보 처리와 관련하여서는 개인정보 보호법 및 관련 지침을 준수하도록 하고 시행지침 작성 시 정보감독위원장과 협의하도록 규정하였다. 나아가 데이터 연계 거버넌스에 있어서 개인정보 감독기구의 역할 강화가 필요하다는 인식이 국제적으로 확대되고 있다. 아이슬란드, 덴마크 등 일부 국가에서는 독립적인 연구윤리위원회의 심사와 더불어, 독립적인 개인정보감독기구가 데이터 열람이나 연계를 승인하고 있다.[343]

다섯째, 앞서의 문제점들이 종합적으로 작용한다면 비식별 정보의 재식별 위험성도 매우 높아진다. 즉, 현행 전문기관 제도의 경우, 결합 목적에 대한 제한이나 승인 절차를 두고 있지 않고 데이터 보유기관에는 연계키 생성 등 결합 절차에서 자사의 이해관계를 관철시킬 수 있는 위치가 부여되어 있다. 그럼에도 국가기관이거나 국가로부터 업무를 위탁받은 전문기관은 범주화 등의 비식별 조치가 절차적으로 이루어졌다는 이유로 결합에 관여된 데이터를 모두 개인정보가 아니라고 추정하면서 결합데이터를 모두 데이터 보유기업에 반출하고 있었다.

결합 절차에서 사용된 비식별 조치가 처음부터 개인을 더 이상 식별하기 불가능한 수준에 이르렀다고 보기 어려운 경우도 있었다. 성별 항목을 남성은 1, 여성은 2로 가명화하도록 권장하거나, 신용등급 등 일부 속성자는 민감하지만 이용 목적에 필요하다는 이유에서 비식별 조치를 미적용하기도 하였다. 일부 전문기관의 경우 데이터 결합을 원하는 신청기업에 “데이터 활용도를 높일 수 있는 비식별 조치 기법”을 권장하기도 한다. 특히 데이터셋 일부를 보유한 기관에 결합데이터를 활용하려는 동기와 이해관계가 크게 작용하고 있을 때 비식별 조치의 적정성 평가에 대한 수용 여부가 자기관 재량사항으로 주어져 있다면 비식별 조치가 완화되고 재식별될 가능성이 커질 수밖에 없다. 이 문제는 개별 결합 사례를 통해 좀 더 구체적으로 살펴볼 수 있다.

한국인터넷진흥원이 수행한 한화생명-SK텔레콤 간 데이터 결합의 경우 한화생명 고객의 47.6%가 SK텔레콤 고객인 것으로 나타났으며, 이들을 식별하는 데 사용될 수 있는 자사 보유 항목은 각 21개에 이르렀다. 각 항목에 대해서는 범주화 등 비식별 조치가 이루어졌지만, 나이, 성별 등의 항목은 결합 이전에 양사가 공동으로 보유하고 있어 개인 식별에 사용될 수 있었음에도 비식별화 수준이 낮은 수준에 그치거나 사실상 재식별이 손쉬운 방식으로 비식별화되었다(p402, <그림 4-55> 한국인터넷진흥원 결합 사례 참조). 이처럼 비식별 조치 수준이 낮은 까닭은 처음부터 그 결합 목적이

342) 국회입법조사처, 2017, “개인정보 비식별 조치에 관한 입법정책적 대응과제”, p48.
343) OECD, 2015, "Health Data Governance ：Privacy, Monitoring and Research", p141.

결합 회사 간에 중복된 고객의 규모 및 가입성향을 파악하고 추후 마케팅 업무에 활용하기 위함에 있었던 데서 찾아볼 수 있다.

한국신용정보원이 수행한 한화생명-한화손해보험 결합 사례에서는 비식별조치 적정성 평가단이 본래 데이터 결합 신청 이전과 결합 이후 모두 구간화/범주화의 수준을 높일 것을 요구한 것으로 나타났다. 그러나 자사 이해관계에 의해 데이터 결합을 추진하고 있는 데이터 보유기관으로서 한화생명은, 이 요구가 '과도'하다고 여기고 추가 회의 등을 통해 기준 완화에 도달한 것으로 나타났다. 이처럼 평가 결과 수용이 보유기업의 재량에 맡겨져 있으면 적정성 평가는 한계를 노정할 수밖에 없다.

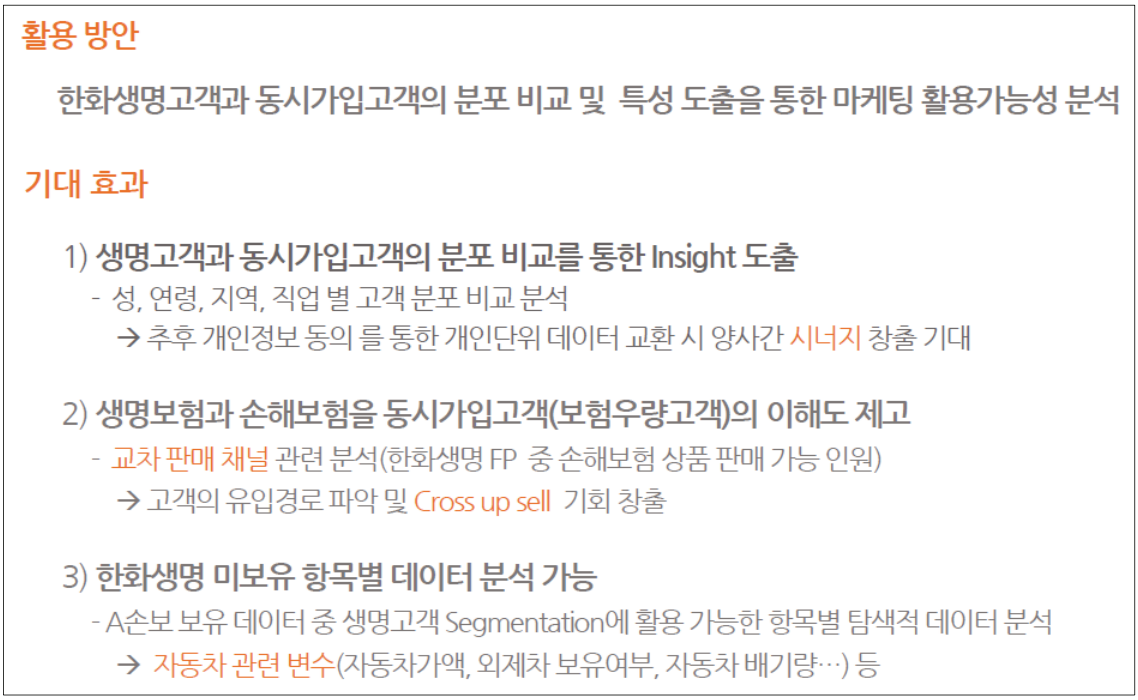
그림 4-61 한화생명-한화손해보험 데이터 결합 비식별조치 적정성 평가

3.평가단 구성 및 적성성평가	Situation & Problem	Decision & Solution
① 평가단 구성	• 변호사, 교수, 비식별 전문가間 평가기준, 주요평가 사항, 평가 일정 등 상이 • 보험 데이터의 이해도 부족	✓ 有경험자 및 일정조율이 가능한 평가위원 선정 ✓ 추가 시간 소요 및 보험 데이터 관련 자료 제공
② 적정성 평가	• 모든 항목을 속성자로 K, L 기준 적용 요구 • 과도한 구간화/범주화 요구 • 평가단과 요청기관간 의견차이 등	✓ 추가 적정성 평가 회의(총3회) ✓ 항목별 구간/범주 기준점 합의 ✓ 양사간 중복 QI 삭제
③ 결합 신청 접수	• 전문기관/요청기관의 DB시스템 차이	✓ 통계프로그램(R 등)을 통한 UTF-8 인코딩 변경
4.결합완료 및 데이터 분석	**Situation & Problem**	**Decision & Solution**
① 결합물 수신 및 적정성 평가	• 결합 후 추가된 항목들에 대한 동일 K, L 기준 적용 요구 • 추가된 항목의 과도한 구간화/범주화 요구	✓ KISA등 상위전문기관 질의 및 답변을 통한 일부 항목만 K, L기준 적용 합의 ✓ 항목 코드화 및 완화된 구간/범주화 기준 적용
② 모니터링 및 사후 평가	• 비식별 데이터 구체적 관리 방안 요구 • 일반 내부데이터와 동일한 관리 기준 적용	✓ 비식별 데이터 관련 내부관리기준 및 규정수립 ✓ 물리적 독립된 DB(Hadoop) 적재 및 양기관간 주기적 모니터링 계획

* 출처: 전도현, 앞의 자료, p13.

나아가 한화생명-한화손해보험 결합 사례의 경우 데이터 보유기관 한측인 한화생명이 데이터 결합 이후 개인 혹은 그룹 재식별화에 대한 동기를 드러냈다. 처음부터 결합 목적을 동시 가입 고객 성향분석에 두고 있었고 결합대상 데이터 항목에도 보험종류별 가입 건수, 가입금액, 가입채널 등 양사 공통항목이 많이 포함되어 있었다. 더불어 한화생명은 타사와 결합된 고객에 대하여 추후 개인정보 동의를 거쳐 마케팅을 수행하거나 상대기업과 데이터 교환을 추진하겠다는 계획을 밝히고, 데이터 결합을 통해 자사가 보유하지 못했던 데이터 항목을 입수하고 분석하겠다는 기대감을 공개적으로 밝혔다.

그림 4-62 한화생명보험-한화손해보험 결합 데이터 활용 방안 및 기대 효과



* 출처: 전도현, 앞의 자료, p16.

다른 사례에서도 비슷한 위험성을 발견할 수 있다. 한국정보화진흥원이 수행한 LG CNS-LG 유플러스 결합데이터 활용 계획 중 하나는 '개인별' 소비 성향 분석 연구에 활용한다는 데 있었고, BC카드-W홈쇼핑 데이터 결합은 양사가 공동으로 공통고객을 대상으로 프로모션을 진행하려는 계획 하에 이루어졌으며, SK텔레콤-한화생명-SCI평가정보 간의 데이터 결합은 그 활용 계획이 '3사 간 공통고객' 성향분석을 한다는 데 두고 있었다. 이는 적어도 각사가 데이터 결합을 통하여 고객 중 누가 상대회사와 공통고객이라는 사실을 이미 식별했거나 식별할 동기를 가지고 있었다는 사실을 보여준다.

이와 같은 목적 하에 비식별 조치는 매우 소극적으로 이루어질 수밖에 없다. XX표에서 성별에서 비식별 조치를 미적용하거나 303개로 범주를 매우 세분화하여 사실상 비식별 조치를 미적용한 효과를 내기도 했다.

이와 같은 사례는 현행 전문기관 제도로는 해당 기업의 부적절한 재식별화를 억지할 수 있는 효과가 크지 않다는 사실을 암시한다. 해외 사례에서처럼 절차적으로 어떤 참여 기관도 전체 데이터셋의 내용이나 식별 데이터를 다룰 수 없도록 기능을 분리하지 않은 상태에서, 보유기업 스스로의 의지만으로 재식별화에 대한 억지 효과가 달성될 것이라고 기대하는 것은 무리이다.

그럼에도 전문기관 제도는 국민 앞의 투명성이 결여되어 있고 개인정보 보호법제나 감독 체계마저 배제되어 있다는 점에서 개인정보에 대한 정보주체의 권리 보호가 위태로운 상태라는 우려가 제기된다.

제5절 기타 정부부처별 민간데이터 연계·결합

1. 미래창조과학부

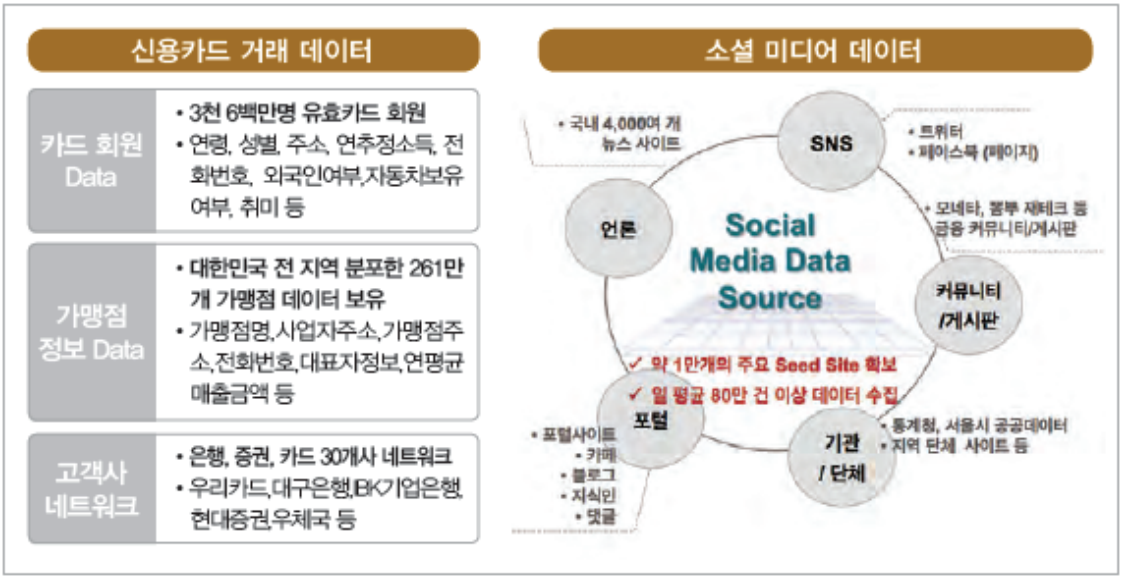
(구)미래창조과학부와 한국정보화진흥원은 비즈니스 중심의 빅데이터 선도 서비스 모델을 발굴하고 개발을 지원하여 빅데이터 서비스 수요를 창출하고 관련 산업을 활성화하는 한편, 빅데이터 서비스 확산을 통해 새로운 고부가가치를 창출하고 창조경제 기반의 신성장동력을 발굴하려는 목적으로, 2013년부터 매년 '빅데이터 시범사업'을 실시해 왔다. (구)미래창조과학부는 미래성장동력을 발굴하기 위한 목적으로 2014년부터 '플래그십 프로젝트' 또한 공모해 왔다. 관련 서비스를 개발하고자 하는 기관 혹은 기업을 대상으로 공모하는 방식으로 이루어진 이들 시범사업 가운데 데이터 연계·결합 사례로는 2015년 BC카드 컨소시움의 '빅데이터 시범사업'과 2016년 SK텔레콤의 '플래그십 프로젝트'를 들 수 있다.

(1) BC카드 컨소시움 연계·결합 사례

2015년 BC카드 컨소시움의 빅데이터 시범사업은 소셜 빅데이터와 카드 결제정보를 연계하여 소비 트렌드를 추출하고 트렌드 프로파일링 작업을 통해 신용카드 분야에서 타켓 마케팅을 실시하는 데 목표가 있었다. 컨소시움 주관 및 참여 기관은 BC카드, LG CNS, 소상공인시장진흥공단으로, BC카드는 연간 약 30억 건, 일평균 9백만 건의 데이터를 데이터베이스에 보관하고 있고, LG CNS는 포털, 커뮤니티, 블로그 및 SNS 등 약 1만 개 이상의 다양한 시드 웹사이트를 확보하여 일 평균 80만 건 이상의 데이터를 모니터링하고 있다.

시범사업 결합에 사용된 BC카드의 신용카드 거래 데이터로는, 3천6백만 명 유효카드 회원의 연령·성별·주소·연 추정소득·전화번호·외국인 여부·자동차보유 여부·취미 등의 카드 회원 데이터, 261만 개 가맹점의 가맹점명·사업자주소·가맹점 주소·전화번호·대표자정보·연평균 매출금액 등의 가맹점 정보 데이터 및 은행, 증권, 카드 등 30개 고객사 네트워크 데이터 등이었다. 결합에 사용된 소셜미디어 데이터로는, 국내 4,000개 뉴스사이트, 포털사이트 카페·블로그·지식인·댓글, 트위터·페이스북 페이지, 모네타·뿜뿌재테크 등 금융 커뮤니티/게시판, 통계청·서울시 공공데이터 및 지역단체 사이트 등 시드 웹사이트에서 수집한 정보였다.

그림 4-63 BC카드 컨소시움 시범사업 활용 데이터



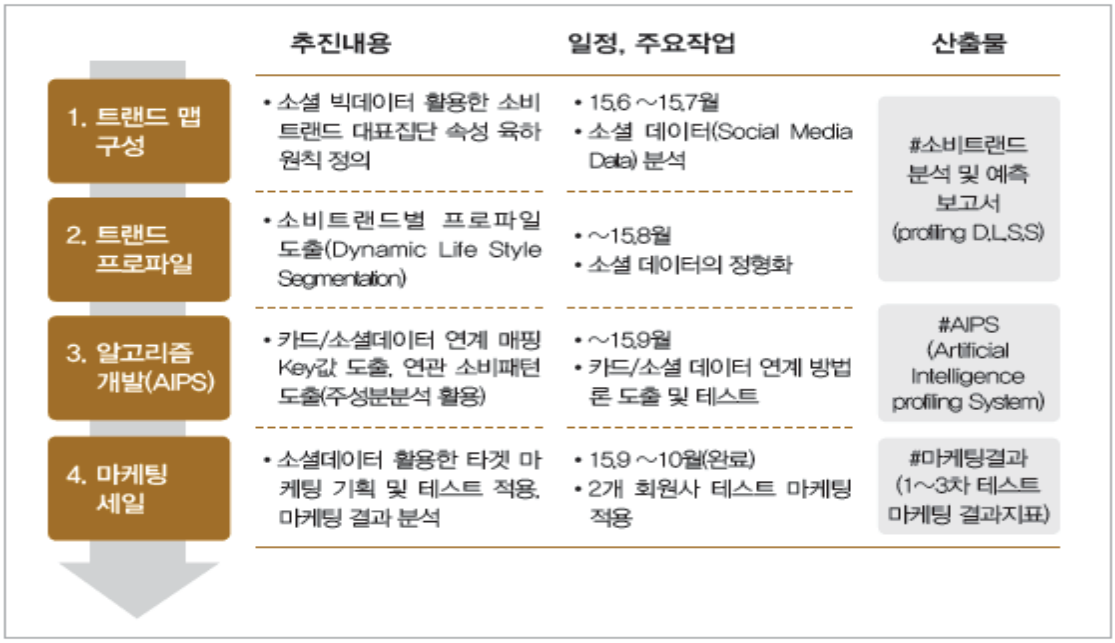
* 출처: 미래창조과학부·한국정보화진흥원·K-ICT 빅데이터센터, "비씨카드 컨소시움", '2016 글로벌 빅데이터 융합 사례집 : 2015 빅데이터 시범사업 및 국내외 사례를 중심으로', 2016.

BC카드 컨소시움은 소셜 데이터의 개인정보 보호법 적용 문제와 관련하여, 데이터 수집이 허가된 도메인만 수집하므로 기관과의 별도 협의는 불요하고, 소셜 데이터의 특성상 개인식별정보에 대한 정확한 식별은 불가능하다고 판단하였다. 개인정보 수집 금지 사이트의 경우 크롤링 단계에서 미리 인지하여 대응하므로 개인정보 수집 이슈는 없다는 것이다. 한편 BC카드 결제정보는 분석 단계에서 개인정보가 포함되지 않은 통계적 데이터를 활용하였고, 마케팅 시행의 경우에도 마케팅에 동의한 고객을 대상으로 하므로 개인정보 이슈는 없다고 판단하였다. 고객의 소비 패턴에 따라 유사한 소비 특성을 보이는 집단을 타겟으로 한 마케팅을 적용하였다는 것이다.

이 사업의 연계 프로세스는 다음과 같았다.

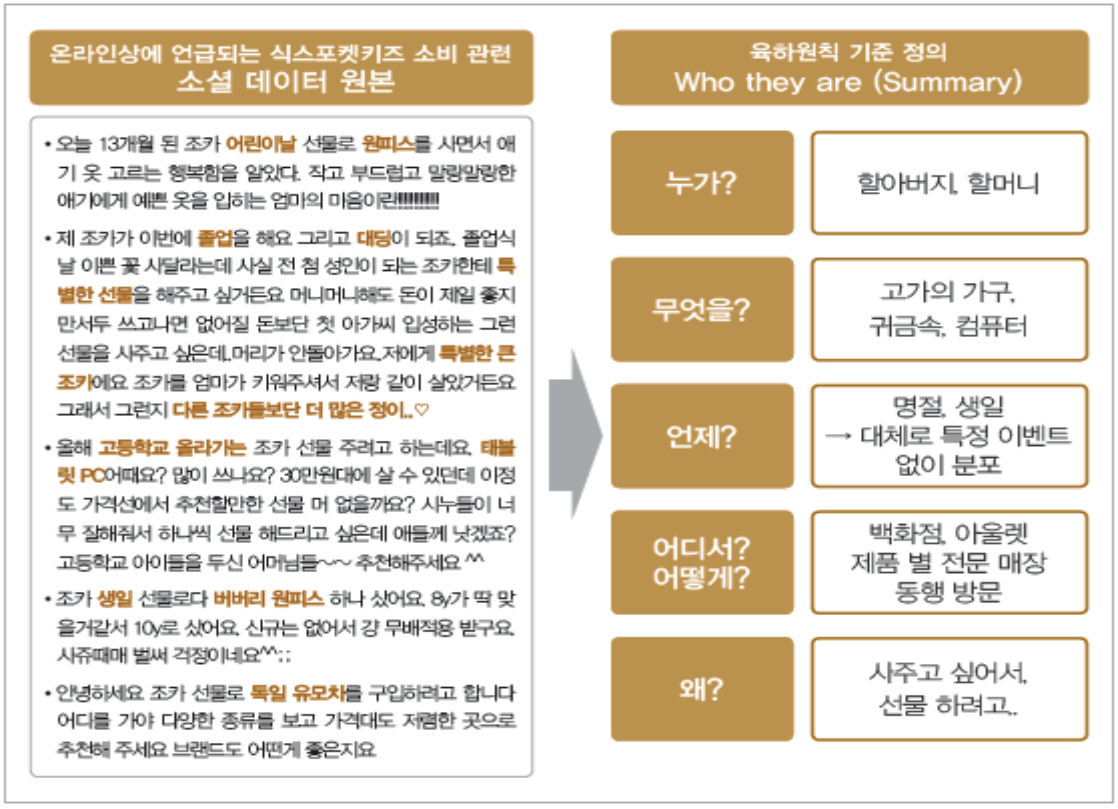
첫째, 소셜 데이터를 분석하여 소비트렌드 맵을 구성하고 소비트렌드별 프로파일을 도출하였다. 트렌드 프로파일링은 각 소비트렌드를 대표하는 집단의 속성을 추론하여 유형화하는 작업으로서, 비식별정보의 식별화, 비정형데이터의 정형화 작업을 포함한다.

그림 4-64 BC카드 컨소시움 시범사업 연계 프로세스



* 출처: 미래창조과학부 등, 앞의 자료.

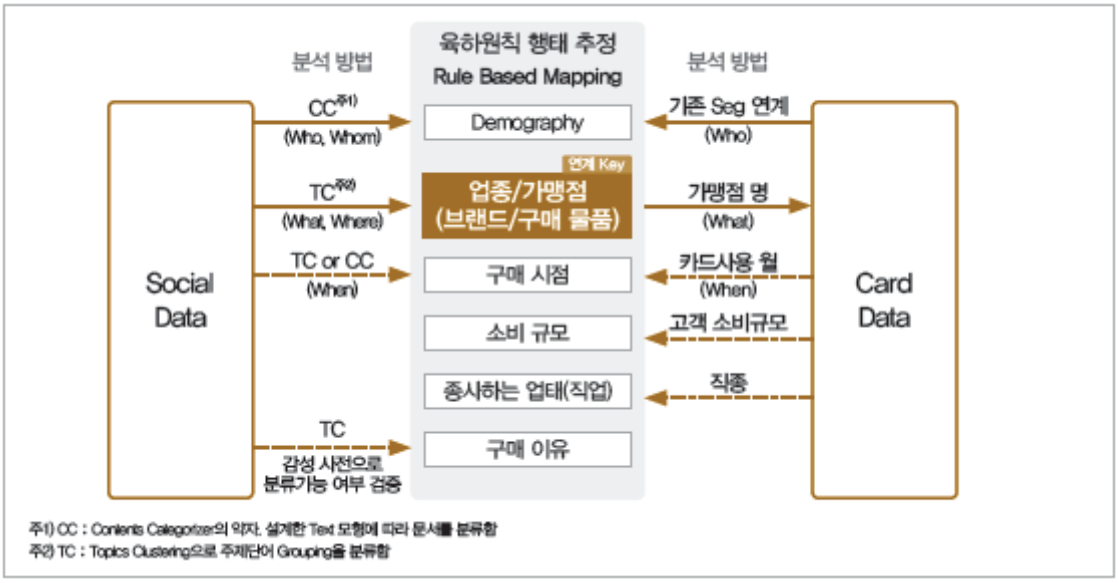
그림 4-65 BC카드 컨소시움 시범사업 프로파일링 작업 예시



* 출처: 미래창조과학부 등, 앞의 자료.

둘째, 카드/소셜 데이터 연계를 위한 알고리즘을 개발하여 매핑 키값을 도출하고 연관 소비패턴을 분석하였다. BC카드 컨소시움은 형태가 다른 정형, 비정형 데이터 연계를 위해 인공지능 마케팅 프로파일링 시스템(AIPS, Artificial Intelligence Marketing Profiling System)을 개발하였고, 소셜 데이터에서 무엇을(what)에 해당하는 업종 정보를 연계 키값으로 선정하였다. 이를 BC카드 데이터와 연계하기 위하여 BC카드 업종을 새롭게 분류하여 매핑에 활용하였다.

그림 4-66 BC카드 컨소시움 시범사업 연계 분석 및 업종코드 매핑



* 출처: 미래창조과학부 등, 앞의 자료.

셋째, 소셜 데이터를 활용한 타겟 마케팅을 기획하고 회원사에 적용하였다. 1차 테스트 마케팅은 BC카드 기존 고객 세그멘테이션 중 소셜 및 카드 연관업종 소비를 많이 보이는 것으로 나타난 4개 세그멘테이션 고객 1만 명을 대상으로 진행하였고, 2차 테스트 마케팅은 약 2만 명을 대상으로 카드/소셜 연관업종 소비 패턴을 보이는 고객 집단을 타겟으로 마케팅을 적용하고 마케팅을 적용하지 않은 일반 그룹과 행사 기간 내 매출 변화를 비교하였다. 3차 테스트 마케팅은 약 7만 명을 대상으로 진행하였고, 12일간 진행한 단기 마케팅임에도 불구하고 오퍼 반응률 6.3% 달성, 약 7천 6백만 원의 수익 발생 효과를 얻었다.

컨소시움은 위와 같은 테스트 마케팅을 2015년 IBK기업은행 고객 대상으로 2차례, 대구은행 고객 대상으로 1차례 진행하였고, 두 곳 은행은 이후에도 협력하여 소셜 데이터를 활용한 트렌드 마케팅을 진행하기로 협의하였다. 컨소시움은 나아가 카드사, 시중은행 및 지방은행, 국가기관 등에 AIPS 관련 설명회를 시행하고 내용 전파를 하였고, 향후에는 금융산업 이외 타 산업으로 전파할 계획을 밝혔다.

BC카드 컨소시움은 이 사업으로 새롭고 정교한 소비트렌드 기반 빅데이터 마케팅이 가능하다는 점을 확인하였고, 개인정보 보호법 등 규제로 인한 산업간 데이터 제휴 어려움을 극복할 방안을 제시할 수 있었으며, 각 산업에 제공 가능한 모델을 제시하였다고 그 효과를 평가하였다. 특히 많은 카드사가 빅데이터를 도입하였지만 대부분 카드사 내부의 매출데이터 분석에 중심을 두고 있는 데 반해, 이 사업은 소셜 데이터와 카드사 내부의 빅데이터를 업계 최초로 결합하여 소셜 데이터 안에 숨겨진 소비 트렌드를 발굴하고 타켓 마케팅을 수행하여 성과를 증명하였다고 자평하였다.

이 사례의 경우 비록 개인 단위로 결합이 이루어지지는 않았으나 이와 같은 방식의 데이터 결합 사례는 소셜 데이터 수집 및 이용의 문제가 쟁점이 될 수 있다. 비록 정보주체인 소셜 계정 이용자가 스스로 공개한 정보이기는 하지만, 이를 통해 개인을 알아보거나 사상·신념, 노동조합·정당의 소속, 정치적 견해, 건강 등 정보주체의 사생활을 현저히 침해할 우려가 있는 개인정보를 유추할 수 있는 경우 이는 개인정보이자 민감정보에 해당할 수 있다. 특히 자동화된 처리로 소셜 데이터를 수집하고 성향이나 관심사, 행동을 파악하거나 다른 데이터셋과 결합시키는 것은 개인에 대한 평가 및 추적을 위한 프로파일링(profiling)에 해당할 수 있다. 2018년 시행을 앞두고 있는 유럽 일반개인정보 보호규정(GDPR)은 '프로파일링'에 대하여 "개인에 관한 특정한 개인적 측면을 평가하기 위해, 특히 개인의 업무능력, 경제 상황, 건강, 개인의 성향이나 관심사, 신뢰도, 행동, 위치, 이동에 관한 측면을 분석 및 예측하기 위해 개인정보를 사용하는 모든 개인정보의 자동처리 형태를 의미한다"고 정의하고 거부권 등 정보주체의 권리를 규정하였다. 현행 우리나라 개인정보 보호법제의 경우 프로파일링 처리

로부터 정보주체의 권리를 보호하는 규율이 미흡하다고 볼 수 있다.

(2) SK텔레콤 연계·결합 사례

SK텔레콤은 2016년도 미래성장동력 플래그십 프로젝트 사업으로 '개인정보 비식별 자료 생성·유통의 현장적용을 위한 실증 적용과제'를 수행하였다. 이 사업은 기업이 보유한 다양한 원시 빅데이터에서 민감한 개인정보를 다양한 비식별화 방법을 적용하여 제거한 후 자유로운 유통 플랫폼을 통해서 안전하게 교환할 수 있는 체계를 실제 기업 환경에 적용하는 것을 목표로 하였다. SK텔레콤은 사업 수행을 위해 이동통신 비식별 데이터를 유통환경에 적용하면서 비식별 조치 가이드라인을 이행하고 그 솔루션의 검증 및 재식별 안정성을 검토하였으며 이종 데이터 간의 연결을 통해 빅데이터 거래를 실증하였다. 더불어 이 사업으로 신규 데이터 거래유통시장을 창출하고 개인 및 중소업체의 데이터 기반 서비스 및 사업 활성화에 기여할 것을 기대하였다.

표 4-61 SK텔레콤 플래그십 시범사업 중 SCI와 결합 데이터 항목

SKT 통신 데이터 - 비식별 결과 데이터 스키마					
순번	한글명	순번	한글명	순번	한글명
1	ABST_ID	16	당월연체유무_1609	31	최근1년간최대연체금액_1603
2	가입자 성	17	납부방법_1609	32	남은할부원금_1603
3	가입자성별	18	회선상태_1609	33	남은할부잔여기간_1603
4	거주지역_시군구	19	연령	34	월평균통화시간_1609
5	멤버쉽등급_1603	20	연령대	35	월평균통화빈도_1609
6	Tablet 보유여부_2016-3	21	가입자생년월일_일수	36	ARPU_1609 (가입자당 평균매출)
7	Smartwatch보유여부_2016-3	22	월평균통화시간_1603	37	당월납부요금_1609
8	결합상품가입여부_1603	23	월평균통화빈도_1603	38	단말기출고가_1609
9	당월연체유무_1603	24	ARPU_1603 (가입자당 평균매출)	39	서비스가입일자_1603_일수_1609
10	납부방법_1603	25	당월납부요금_1603	40	정지일수_1609
11	회선상태_1603	26	단말기출고가_1603	41	당월연체금액_1609
12	멤버쉽등급_1609	27	서비스가입일자_1603_일수	42	최근1년간납부일미준수횟수_1609
13	Tablet 보유여부_2016-9	28	정지일수_1603	43	최근1년간최대연체금액_1609
14	Smartwatch보유여부_2016-9	29	당월연체금액_1603	44	남은할부원금_1609
15	결합상품가입여부_1609	30	최근1년간납부일미준수횟수_1603	45	남은할부잔여기간_1609

플래그십 사업 중 SK텔레콤 가입자 데이터셋과 SCI 평가정보 데이터셋의 결합 사례를 검토해 보면 다음과 같다.

우선 2017년 3월 SK텔레콤은 중금리 대출 이용자의 신용도 향상 가능성을 검증하기 위한 목적으로 SK텔레콤 가입자 중 제3자 데이터 제공동의를 한 서울 지역 가입자를 대상으로 SCI와 직접 데이터 결합을 진행하였다. 이때 SK텔레콤의 결합대상은 45개 항목이었으며, 이때 가입자 성(姓), 연령대, 가입자 성별, 거주지역_시군구의 경우 K-익명성[k=4] 기준을 적용하고, 당월연체금액_2016-3, 당월연체금액_2016-9의 경우 l-다양성[l=2] 기준을 적용하여 비식별 조치가 이루어졌다.

비식별 조치 후 SK텔레콤 결합 대상자 3,754,040건과 SCI 결합 대상자 15,555,049건을 연계한 결과 970,553건의 결합(동기화)이 성공하였다. 이 결합은 별도의 임시대체키를 생성하지 않고 가입자 성, 연령대, 가입자 성별을 키값으로 하여 결합을 수행하였으며 전문기관 개입 없이 자체적으로 결합을 수행하였다.

이와 같은 방식의 데이터 결합 사례의 문제점은 개인정보 보호법의 준수 여부가 불명확하다는 것이다. 우선 이 결합 사례가 근거로 삼은 비식별화 가이드라인의 경우 개인정보 보호법의 보호대상이 되는 개인정보를 매우 좁게 해석하여 '알아볼 수 있는' 자의 주체를 해당 '정보를 처리하는 자'로, '다른 정보와 쉽게 결합하여'의 의미를 '결합대상이 될 다른 정보의 입수 가능성이 있어야 하고, 또 결합 가능성도 높아야 한다'는 것으로 해석하였다. 이에 따라 사업 수행자는 자사 보유 고객 데이터셋에 대하여 비식별 조치를 취한 이후에는 개인정보가 아닌 것으로 추정하고 타사 데이터셋과 결합을 수행하였다. 그러나 법원은 이미 "구하기 쉬운지 어려운지와는 상관없이 해당 정보와 다른 정보가 특별한 어려움 없이 쉽게 결합하여 특정 개인을 알아볼 수 있게 되는 것"은 개인정보 보호 법제에 의해 보호를 받는 개인정보라고 판시한 바 있다. 또한, 데이터셋을 비식별화하거나 다른 기업의 데이터셋과 결합시키기 전의 데이터셋 원본을 가지고 있고 결합 과정에서 연계키를 스스로 생성하고 비식별 조치 또한 스스로 취한 기업으로서는, 결합 이후 데이터셋에서도 개인을 알아볼 가능성이 크다. 특히 이 사례와 같이 별도의 대체키 없이 데이터셋에 포함된 가입자 성, 연령대, 가입자 성별을 키값으로 하여 각 데이터셋에 속한 개인 1명을 직접 연계한 경우, 이 유일한 키값을 통해서도 원본 데이터셋에 포함된 개인을 알아볼 수 있다. 이와 같이 개인정보일 가능성이 있는 데이터 처리에 있어서 개인정보처리자는 개인정보 보호법상의 의무를 준수할 책임이 있다.

다른 한편 플래그십 사업을 수행한 SK텔레콤은 이 플래그십 사업에서 제3자 정보제공에 동의한 고객을 대상으로 연계가 이루어졌다고 밝혔다. 현행 개인정보 보호 법제상 제3자에게 개인정보를 제공할 때 동의를 받으려면, 개인정보를 제공받는 자, 개인정보를 제공받는 자의 개인정보 이용 목적, 제공하는 개인정보의 항목, 개인정보를 제공받는 자의 개인정보 보유 및 이용 기간을 모두 알리고 별도의 동의를 받아야 한

다(정보통신망법 제24조의2 제1항 및 개인정보 보호법 제18조 제3항). SK텔레콤 결합 대상자 3,754,040건뿐 아니라 SCI 결합 대상자 15,555,049건에 대해서 모두 상대 기업명 및 중금리 대출 이용자의 신용도 향상 가능성 검증이라는 목적을 밝히고 별도 동의를 받아야 적법하다고 볼 수 있다.

2. 국토교통부

국토교통부는 교통 분야 민간데이터에 대한 접근을 위하여 법 제도적 장치 마련을 추진해 왔다.

특히 교통안전공단이 2006년부터 국토교통부의 위탁을 받아 실시 중인 대중교통 현황조사 차원에서 교통카드데이터를 수집 및 활용해야 한다는 취지에서, 2015년 12월 29일 '대중교통의 육성 및 이용촉진에 관한 법률'(이하 '대중교통법') 개정이 이루어졌다. 공공정보로써 활용되는 교통카드데이터를 정의하고, 표준수집항목을 규정하는 한편, 교통카드데이터 수집 및 제공 체계를 규정하고 지속적 활용을 위한 시스템 구축 방안을 마련하는 내용이었다.[344]

우선 대중교통법에 '교통카드데이터'에 대하여 "교통카드를 사용하여 대중교통수단을 이용한 전산 자료 중 이용자의 통행실태 파악에 필요한 자료로서 이용자를 알아볼 수 없는 형태로 가공한 자료를 말한다"고 정의하고(법 제2조 제7호), 국토교통부장관이 대중교통수단 이용자의 통행실태를 파악하기 위하여 교통카드데이터를 수집하고 관리하도록 하였다(법 제10조의8). 즉, 국토교통부장관은 대중교통운영자, 대중교통운영자, 전자금융업자, 교통카드정산사업자 등으로부터 △대중교통수단의 이용지역 △대중교통수단의 명칭 △대중교통수단의 승하차 시간 △대중교통수단의 환승에 관한 정보 △대중교통수단의 정류장에 관한 정보 △그 밖에 국토교통부장관이 법 제10조의10 제1항에 따른 교통카드데이터 통합정보시스템의 구축 및 운영을 위하여 필요하다고 인정하여 고시하는 정보 등 교통카드데이터 제출을 요청할 수 있다(대중교통법 시행령 제11조의5). 이에 따라 교통카드데이터 제출을 요청받은 자는 특별한 사유가 없으면 이에 따라야 하고, 요청받은 교통카드데이터를 이동식 저장장치 또는 광디스크 등 전산기록장치를 활용하여 제출하거나, 정보통신망 또는 통합정보시스템을 이용하여 제출할 수 있다.

한편, 교통카드데이터를 이용하려는 자는 국토교통부장관에게 수집된 교통카드데이

344) 한국조사연구학회, '빅데이터 활용 통계생산방법론 연구용역 결과보고서', 통계청 연구용역보고서, 2016 ; 이인묵·박선영·민재홍, "교통카드데이터 공공이용 활성화를 위한 정책방안", 2014년도 한국철도학회 추계학술대회 논문집, 2014 종합.

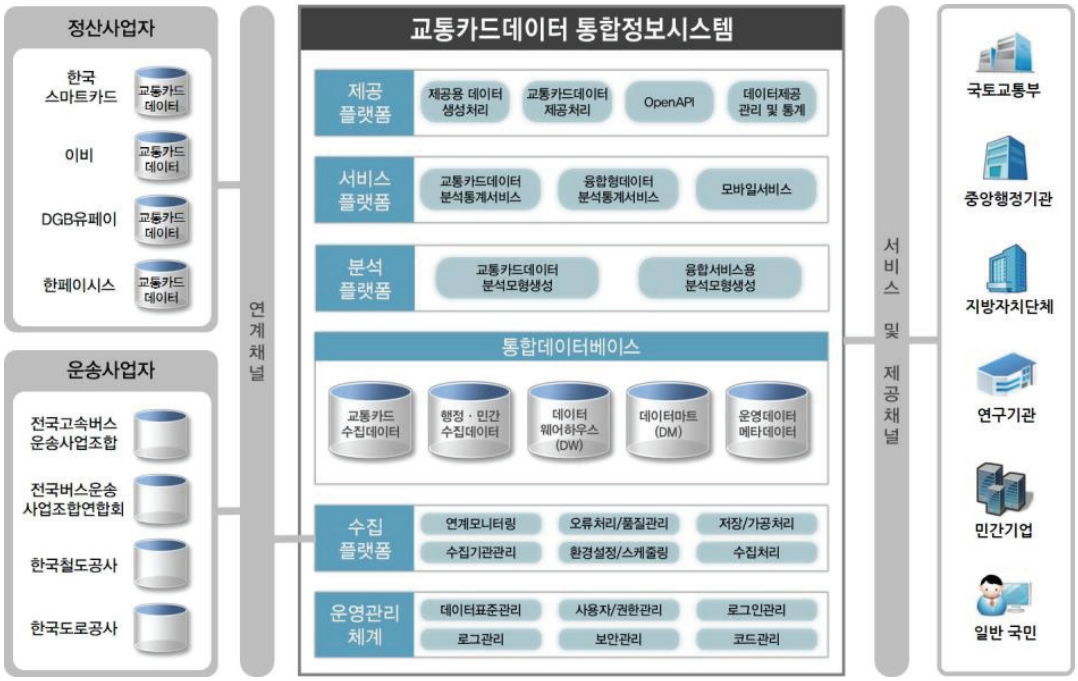
터의 제공을 요청할 수 있다(법 제10조의9). 이때 제공되는 교통카드데이터는 집계자료 형태로 제공되며, 다만 국가, 지방자치단체 및 교통 관련 연구기관이나 공공기관이 교통 관련 정책수립, 업무수행, 통계 작성 및 학술연구 등의 목적으로 요청하는 경우에는 그러하지 아니하다. 교통카드데이터를 제공받은 자는 제3자의 권리를 침해하거나 범죄 등의 불법행위를 할 목적으로 교통카드데이터를 이용하거나, 이를 제3자에게 임의로 제공 및 유출하거나, 위조 및 변조해서는 안 되며, 이를 위반할 경우 2년 이하의 징역 또는 2천만 원 이하의 벌금에 처한다. 또한 교통카드데이터를 제공받은 자는 교통카드데이터가 분실되거나 도난되지 아니하도록 대통령령으로 정하는 바에 따라 안전성 확보에 필요한 조치를 해야 하며, 위반 시 1년 이하의 징역 또는 1천만 원 이하의 벌금에 처한다.

나아가 국토교통부장관은 위 교통카드데이터의 수집·관리·제출 및 제공을 위하여 교통카드데이터 통합정보시스템을 구축·운영할 수 있다(법 제10조의10). 2016년 5월 국토교통부는 교통카드 빅데이터 통합정보시스템 1단계 구축을 발표하였다. 이에 따르면 이전에는 교통카드데이터 수집 및 활용을 위한 법적 근거가 없었고 교통카드 정산사업자(한국스마트카드, 이비카드, 코레일 등 8개사)별로 정보 체계가 달라 효율적인 사용에 제약이 있었으나, 대중교통법 개정으로 법적 근거가 마련된 만큼 2016년 1개 교통카드 정산사업자를 대상으로 1단계 사업을 추진하고, 2017년에 전체 정산사업자로 확대하여 시스템구축을 완료할 계획이다.

대중교통법에서는 이 시스템에서 수집 및 활용하는 교통카드데이터에 대하여 '이용자를 알아볼 수 없는 형태로 가공한 자료'로 규정하고 있으며, 개별 교통카드 정산사업자는 교통카드 번호를 암호화하여 16~64자리의 가상번호로 변환한 뒤 이를 통합정보시스템에 제공한다.

국토교통부는 '교통카드빅데이터 통합정보시스템'의 일차적인 기대 효과로서, 기존에 우리나라 전체 대중교통 9천여 개 노선에 대한 조사비용을 약 97% 절감할 수 있으며 데이터 요청 시 결과 제공까지 걸리던 기간도 기존 45일~ 90일에서 10일 이내로 대폭 단축할 수 있다는 점을 들었다. 또한, 노선 신설·조정, 정차 지점 및 배차 간격 최적화 등 정부·지자체·사업자별로 보다 편리하고 정밀한 교통체계를 만들어 대중교통 이용이 활성화되는 효과가 생기게 된다고 밝혔다. 다른 한편으로 또한 민간에도 관련 데이터를 제공하여 부동산, 통신, 재해・재난, 기상 등 다양한 분야와 연계, 광고 입지 분석, 창업 등에 폭넓게 활용될 것 또한 기대하였다.

그림 4-67 교통카드데이터 통합정보시스템



* 출처: 한국조사연구학회, 앞의 자료, p170.

국토교통부는 2016년 11월 1일부터 대중교통법 제10조의8에 따라 매일 한국스마트카드로부터 시내버스 및 도시철도 이용 교통카드의 가상카드번호, 이용지역, 수단, 승하차 시간, 요금, 이용 거리, 환승에 관한 정보, 정류장에 관한 정보, 노선정보 등을 연계하여 수집하고 있으며, 시내버스, 도시철도, 시외버스, 고속버스, 간선철도에서 운영 중인 다른 교통카드 정산사업자 등과도 교통카드데이터를 연계하고 수집할 계획이다. 이렇게 제공받은 교통카드데이터는 대중교통법 제10조의10에 따른 교통카드빅데이터 통합정보시스템에 연계된다. 또한, 국토교통부는 '16년 1단계 사업을 통해서 교통카드빅데이터 통합정보시스템에 연계·구축된 일부지역(수도권, 세종, 대전, 경상남·북도)의 교통카드데이터를 대중교통법 제10조의9에 따라 한국교통연구원 등 교통관련 기관, 국가기관 및 지방자치단체 등에 제공하고 있다. 이때 국가 및 지방자치단체에 대해서는 가상카드번호, 이용지역, 수단, 승하차 시간, 요금, 이용 거리, 환승에 관한 정보, 정류장에 관한 정보, 노선정보 등의 교통 데이터를 제공하고, 교통관련기관에는 요금정보를 제외한 요청 정보를 제공한다. 다만 요금정보가 필요한 연구의 경우(광역 알뜰교통카드 도입을 위한 현황조사) 요금정보를 제한적으로 제공한다.

대중교통법 제10조의10에 따라 구축되고 있는 교통카드빅데이터 통합정보시스템은, 2016년 12월까지 이루어진 1단계 구축사업으로 교통카드 정산사업자인 한국스마트카드 수도권 교통카드데이터 및 기반데이터가 연계되었다. 2017년 2단계 구축사업으로

한국스마트카드의 수도권 이외 지역 교통카드데이터 및 그 외 교통카드 정산사업자인 이비카드, DGB유페이, 마이비, 한페이시스, 글로벌캐시의 교통카드데이터와 기반데이터를 연계하고, 특정부문사업자인 한국철도공사, SR, 시외버스(전국버스운송사업조합연합회), 고속버스조합(전국고속버스운송사업조합)의 교통카드데이터, 기반데이터와 발권데이터에 대한 연계를 추진하고 있다.[345]

나아가 한국교통연구원은 이와 같은 과정을 통해 수집한 교통 데이터를 업무협약(MOU)을 통해 삼성카드가 보유한 소비데이터와 연계하는 사업을 추진 중이다.[346] 한국교통연구원은 이러한 공공과 민간데이터의 연계를 통하여 소비데이터 중 교통비 지출구조의 특성을 분석하고, 화물차의 이동 경로 및 화물차 운전자의 소비 특성 등에 대하여 분석이 이루어질 것으로 기대하였다.

이와 같이 국가 차원에서 교통카드데이터를 수집하고 이를 집적하여 빅데이터 통합정보시스템을 구축하는 정책은 본래 대중교통 현황조사라는 공익 목적으로 추진되어 왔다. 그러나 대중교통법은 교통카드 번호를 암호화하는 방식으로 안전조치를 취한 후에는 아무런 제한 없이 영리적 기업을 비롯하여 “교통카드데이터를 이용하려는 자” 모두에게 제공할 것을 규정하고 있고 공공과 민간이 보유한 교통카드데이터의 광범위한 연계 또한 예상되고 있다.

그러나 암호화의 경우 암호화 수준이나 컴퓨터 성능 발달에 따라 재식별 위험이 크다. 따라서 암호화 조치만으로 교통카드데이터를 공개하는 것은 해당 데이터셋 단독 혹은 연계를 통해 교통수단을 이용한 국민의 개인정보가 공개되는 결과를 낳을 수도 있다. 따라서 공공정책을 위해 교통 빅데이터를 구축 및 이용하고자 한다면 그 목적에 부합하도록 보관 기간은 물론 수집 및 제공을 제한하는 등 데이터 생애주기별로 세심하게 설계할 필요가 있다. 간헐적으로 이루어지는 현황조사나 노선 정책을 목적으로 국민의 교통카드데이터를 대량으로 장기간 집적할 필요성과 타당성이 충분한 것인지, 해당 정책이 비례성 및 개인정보 최소수집의 원칙을 충족하는지에 대한 검토 또한 필요해 보인다. 무엇보다 개인이 식별될 가능성이 큰 교통 데이터를 임의적으로 목적 외 이용하거나 제공하는 일이 없도록 유의할 필요가 있다. 전반적으로 교통 빅데이터 정책은 개별적인 입법 보다 개인정보 보호법은 물론 통계법상 지정통계 제도 등 기존의 보호 규범 및 관련 제도를 원칙적으로 준수하는 선에서 이루어지는 것이 바람직할 것이다. 데이터 거버넌스 체계 또한 기능 분리 원칙을 준수하여 공익 연구 목적 외 이용을 최소화하고 재식별 위험을 최소화할 필요가 있다.

---

345) 추혜선 의원 자료.

346) “삼성카드, 빅데이터 활용 교통정책 연구 위한 MOU 체결”, 서울경제 2017. 9. 1.; 추혜선 의원 자료.

제5장 정책 제안

제1절 관련 법제의 정비

1. 데이터 연계·결합 관련 법제의 정합성 유지

현재 개인정보 보호법, 전자정부법, 공공데이터법에는 개인정보의 제공, 공개, 연계, 결합 등에 관하여 동일한 영역을 법률마다 상호 모순되게 규정하고 있다. 이와 같은 모순은 우리나라 개인정보 보호 법제에서 의도된 것으로 보기도 어렵고 그로 인해 개인정보 보호의 기본원칙이 심각하게 훼손될 수 있으므로 세 법의 관계를 명확하게 정리할 필요가 있다.

전자정부법은 행정정보 시스템의 연계·통합 등을 주로 효율성의 측면에서 평가하고 추진하는 법률이고 공공데이터법은 공공데이터 이용 활성화를 목적으로 하고 있으므로, 이로 인해서 개인정보 보호 원칙이 훼손되거나 정보주체의 개인정보 자기결정권이 침해될 수 없다는 점을 분명히 해야 한다. 이런 관점에서 세 법의 관계를 설정하고, 정합성이 있도록 개선해 나가야 한다.

전자정부법은 특정한 행정기관의 행정 작용에 대한 법률이 아니고 행정업무의 전자적 처리를 위한 기본원칙, 절차 및 추진방법 등을 규정함으로써 전자정부를 효율적으로 구현하고 행정의 생산성, 투명성 및 민주성을 높여 국민의 삶의 질을 향상시키는 것을 목적으로 하는 법률인데, 실제로는 공공부문에서 데이터의 연계·결합이 이루어지게 하는 근거 법률로 기능하고 있다. 이런 점에서 전자정부법을 본래의 목적에 부합하도록 해석·적용할 필요가 있다.

공공데이터법은 그 대상이 되는 정보가 공공기관이 보유·관리하는 데이터로서 매우 범위가 넓은데, 그 안에는 민감한 개인정보도 포함되어 있다. 공공데이터법은 국민의 공공데이터에 대한 이용권을 보장하고 공공데이터의 민간 활용을 통한 삶의 질 향상과 국민경제 발전에 이바지함을 목적으로 하는 것이므로, 국민의 이용권이 개인정보 자기결정권을 초월하는 가치를 갖는 경우로 본래의 목적에 부합하도록 해석·적용할 필요가 있다. 공공데이터법은 기술적 분리 가능성을 근거로 기술적 분리 후의 공공데이터를 제공하도록 하고 있는데, 기술적 분리가 이루어진 공공데이터도 여전히 연계·결합을 통해 개인 식별의 가능성이 매우 높기 때문에 이런 위험에 대한 대응이 필요하다.

2. 개인정보 보호법 정비: 연구 및 통계 목적 개인정보 처리

연구 및 통계 목적의 개인정보 처리를 위해 보다 구체적인 원칙과 조건을 포함하는 방향으로 현행 개인정보 보호법을 개정할 필요가 있다. 개정 방향은 다음과 같은 내용을 포함할 것을 제안한다.

① 공익을 위한 유지보존의 목적, 학술적·역사적 연구의 목적 또는 통계 목적의 개인정보 처리 시 정보주체의 권리를 보호하기 위해 아래와 같은 적절한 안전조치를 취해야 한다. 이를 조건으로 최초 개인정보 수집 시 이에 대한 동의를 별도로 받지 않았어도 이와 같은 목적의 개인정보 처리를 허용하는 규정을 마련할 필요도 있다. 단, 이는 동의를 얻는 것이 현실적으로 불가능하거나, 지나치게 비용이 많이 들거나 기술적으로 어려운 경우로 한정하는 것이 바람직할 것이다.[347]

나아가 독일 연방정보보호법과 같이 학술적 이익이 개인정보 침해의 위험보다 훨씬 커야 한다는 점을 명시하는 것이 좋을 것이다.

가. 개인정보 최소수집, 처리의 원칙을 보장하기 위한 기술적·조직적 조치

나. 특정 목적의 달성에 가명처리로도 가능하다면 가명처리가 이루어져야 한다.

다. 익명화 처리(정보주체의 식별을 할 수 없거나 더 이상 허용하지 않는 방식의 추가 처리)를 통해서도 연구나 통계 목적의 정보처리 목적이 달성될 수 있다면 익명화 처리를 하여야 한다.

② 개인정보가 학술적·역사적 연구목적, 또는 통계 목적으로 처리되는 경우, 열람권, 수정권, 처리제한권, 처리 거부권, 처리에 대한 안전조치는 일부 제한될 수 있다.

가. 이러한 권리로 인해 특정 목적의 달성이 불가능하거나 심각하게 저해될 가능성이 있고 그 적용을 일부 제한함으로써 특정 목적의 충족이 가능해지는 경우

나. 이 경우 필요한 특정 권리만을 제한해야 한다.

관련하여 개인정보 보호법 제58조에서 통계법 등에 따라 수집되는 개인정보에 대해 제3장부터 7장까지 일률적으로 적용 배제하도록 한 것은 수정될 필요가 있다. 통계의 작성 및 활용도 공공기관의 법령상 의무를 수행하는 과정에서 개인정보를 수집, 처리,

347) GDPR 제89조의 2항 및 3항, 영국 NHS법 2006의 Section 251 지원 조항, 독일 연방정보보호법 제28조, 미국 HIPAA에서 기관평가위원회의 이용 승인 조건 등이 그렇다.

활용하는 경우와 마찬가지로 규율해도 충분하다. 통계법에 의한 통계에 대해서는 정보주체의 동의가 없어도 개인정보를 수집할 수 있고, 목적 외 이용도 통계법의 규정이 있는 경우에는 허용되고, 제3자 제공도 통계법의 규정이 있는 경우는 허용되는 것으로 규율하는 것이 바람직하다. 다만, 통계법에 의하여 처리되는 개인정보에 대해서는 정보주체에게 고지할 의무의 완화(제20조), 개인정보의 정정·삭제(제36조)에 대한 예외, 개인정보의 처리정지(제37조)에 대한 예외 등을 규정할 수는 있을 것이다.

이와 같은 예외적인 규정 외에는 개인정보 보호법의 개인정보 보호원칙이 모두 적용되어야 한다.[348] 연구 및 통계 목적으로 제공된 정보는 해당 목적으로만 사용되어야 한다.

3. 보건의료 관련 법률 및 통계법 등의 정비

(1) 보건의료 관련 법률의 정비

앞서 제4장 제2절에서 상세하게 살펴본 바와 같이 국민의 건강정보가 방대하게 수집·활용·연계되고 있음에도 불구하고, 국민건강보험법, 암관리법, 건강검진기본법 등 현행 보건의료 관련 법률에서는 건강정보의 수집·활용·제공 등의 근거가 부재하거나 모호한 경우가 많았다. 또한, 수집된 건강정보를 (준)영구적으로 보유하고 있는 등 개인정보 보호원칙에서 벗어나는 경우도 많았다. 건강정보가 매우 민감한 개인정보라고 할 때 이러한 법적 근거의 부족은 심각한 문제라고 할 수 있다.

보건의료 분야는 세계적으로도 데이터 연계·결합을 통한 학술연구 및 정책 형성의 요구가 높은 분야이다. 그러나 건강정보 데이터의 연계·결합이 활성화되기 위해서라도 건강정보의 수집부터 활용·제공 등 전반적인 데이터 거버넌스 체제가 정비될 필요가 있다. 무엇보다 현행 보건의료 관련 법률을 재검토하여 법적 근거를 명확하게 하고, 개인정보 보호원칙을 위반하는 건강정보의 수집 관행을 바로잡을 필요가 있다.

(2) 통계법의 정비

앞서 제4장 제3절에서 살펴본 바와 같이 현재 통계법의 규정은 통계의 비밀보호가 충분하지 못한 상황이다. 특히 현행 개인정보 보호법이 통계와 관련한 정보에 대한

348) GDPR은 공익적인 기록 보존, 과학 및 역사 연구 또는 통계 목적을 위해 필요한 경우 민감정보를 처리할 수 있도록 하면서도 법률에 근거하고, 이 법률은 추구하는 목적에 비례하고 개인정보 보호권의 본질을 존중하며 정보주체의 기본권 및 이익을 보호하기 위해 적절하고 구체적인 조치를 제공하는 것이어야 한다는 요건을 두고 있다.

개인정보 보호법의 적용을 배제하고 있는 것을 고려하면, 더더욱 통계법에 그에 대한 규정을 둘 필요가 있다. 통계자료의 목적 달성 후 폐기 의무 및 통계자료를 보유하는 동안에도 통계자료의 개인식별자를 대체번호 등으로 치환하고 연계정보는 별도로 보관하면서 대체식별자로 처리된 통계자료를 활용하는 등의 안전조치를 취하도록 해야 한다. 통계자료를 보유하는 동안 물리적, 조직적, 관리적 안전조치를 유지하도록 하고, 특히 통계의 기밀유지를 위한 여러 규정을 마련할 필요가 있다.

비밀보호의 예외적인 경우로 학술적 연구와 통계 목적의 통계자료 활용에 대해서는 그 절차와 기준에 대한 엄격한 법적 근거를 갖추는 것이 필요하다.

(3) 연구 및 통계 목적의 데이터 활용 및 연계에 대한 법적 규율 마련

연구 및 통계 목적의 데이터 활용 및 연계에 대해서 그 활용을 허용하면서도 분명한 규율을 마련하여 엄격한 요건 하에서 이루어질 수 있도록 할 필요가 있다. 공익적 필요가 큰 반면에 당사자의 동의를 받는 것이 극히 곤란한 부득이한 경우에는 당사자의 동의 없는 데이터 활용 및 연계를 허용할 수 있겠지만, 비밀보호와 개인정보 보호를 위한 다양한 안전장치를 마련해야 한다.

특히 학술 목적으로 개인정보를 활용하고 연계하는 것을 허용하는 요건으로 다음과 같은 기준을 마련할 필요가 있다.

① 접근이 승인된 연구기관에 의해 이루어질 것

② 적절한 연구 제안서가 제출될 것

③ 학술목적으로 요청된 기밀정보의 유형이 적시될 것

④ (통계작성기관 등이) 인가한 접근 시설에서 접근이 제공될 것

⑤ (학술 목적으로 통계자료에 대한 접근을 허용하는 경우) 해당 정보를 제공한 해당 통계작성기관이 승인할 것

연구기관은 최소한 기관의 목적, 양질의 연구 이력과 명성, 연구를 위한 내부 조직체계, 정보보안을 만족할 안전조치 구비를 요건으로 해야 한다. 연구기관의 안전조치와 절차, 연구기관의 공개, 연구제안서의 요건, 개인정보(건강정보, 통계자료 등)에의 접근 조건, 접근 시설 등에 대한 상세한 기준을 마련해야 한다.

이와 관련하여 법령 또는 최소한 고시를 통해서 연구 및 통계 목적으로 개인정보를 제공할 경우에 정보를 제공받는 목적(공익 목적의 학술연구 및 통계), 제안서의 요건, 이용 주체의 자격(예를 들어, 승인된 연구기관 혹은 연구자 요건), 이용의 조건(안전

시설 내에서의 이용, 계약의 체결, 연구 결과물의 검토), 데이터 활용의 조건(안전조치, 계약, 연구 결과물의 공개 전 검토 등), 제반 정책 및 승인 절차 등을 보다 구체적으로 규정하는 것이 바람직하다.

데이터 연계와 관련해서는 유엔의 <통계 및 관련 연구목적을 위해 수행되는 데이터 통합의 기밀성 관련 원칙과 가이드라인>을 각국의 법제 및 가이드라인에 반영할 수 있을 것이다. 뉴질랜드에서 2006년에 <데이터 통합 매뉴얼(Data Integration Manual)>의 두 번째 버전을 만들 때, 이 원칙과 가이드라인을 반영한 것을 참고할 필요가 있다.

또한 대다수 국가의 법률들이 투명성을 강조하고 있는데 이에 대한 규정도 마련하는 것이 바람직하다. 즉, 데이터 제공기관의 정책, 승인이 필요할 경우 그 기준 및 승인 목록 등을 공개하도록 하고 있다. 또한 대부분의 국가 통계법에서는 개별 정보에 접근할 수 있는 사람을 제한하고 있다.

감독기구의 감독 규정도 마련하는 것을 고려해야 한다. 예를 들어 프랑스의 경우, 보건 분야의 공익적인 연구, 조사 혹은 평가를 목적으로 한 개인정보의 처리, 그 목적이 서로 다른 공익을 위한 파일들의 연계 등은 개인정보 감독기구인 CNIL의 허가를 받도록 하고 있어 공익 목적의 연구 승인 기관으로서 감독기구가 역할을 하는 경우도 있다. 국내에서도 개인정보 보호위원회가 데이터 거버넌스와 관련한 원칙 및 정책수립, 다른 거버넌스 기구에 대한 자문 등 국내 데이터 거버넌스 체제 수립에 적극적인 역할을 할 필요가 있다.

제2절 데이터 거버넌스 체제의 구축

지금까지 국내외 현황 검토를 통해 살펴보았듯이, 데이터의 안전한 활용을 위해서는 관련 법제의 정비는 필수적이기는 하지만 충분하지는 않다. 법에서는 데이터 거버넌스를 위한 큰 틀에서의 원칙만을 규정할 뿐, 구체적으로 특정 연구가 얼마나 공익적인지, 연구기관이나 연구자가 자격을 갖추었는지, 기술적·조직적 안전조치는 제대로 갖추어졌는지 등에 대해서 구체적인 사례별로 판단이 이루어질 필요가 있기 때문이다. 또한, 데이터의 수집·이용·제공·연계 등 전반에 걸쳐서 데이터의 활용과 보호를 위한 체제가 갖추어져있지 않으면 법에서 허용하고 있는 데이터의 활용도 행정적인 이유로 제약되거나, 반대로 법에서 규정한 데이터 보호조치가 현실에서 제대로 이행되지 않아 개인정보의 권리가 침해될 수도 있기 때문이다.

국내에서도 인간 관련 연구에 '생명윤리 및 안전에 관한 법률'에 따른 심의를 받도록 하거나 데이터에 대한 접근을 일정한 안전시설에서 하게 하는 등 해외 사례에서 볼 수 있었던 거버넌스 체제나 안전조치를 일부 취하고는 있지만, 아직 보건의료 등 각 영역에서 견고한 데이터 거버넌스 체제가 갖추어져 있다고 보기는 힘들다. 국내에서는 개인정보의 비식별화 등 데이터의 개인 식별성을 어떻게 최소화할 것인지에 대해서만 초점을 맞추는 경향이 있다. 그러나 비식별화는 전반적인 거버넌스의 하나의 요소일 뿐이다. 데이터의 이용과 보호에 관련된 법제, 데이터 접근·연계 정책, 연구기관 혹은 데이터 연계기관의 인증, 심사절차, 데이터 접근 절차 등에 이르는 전반적인 거버넌스 체제를 갖추도록 노력할 필요가 있다. 제1절에서 다루었던 법제 개선과 함께, 전반적인 데이터 거버넌스 체제를 국가적인 차원에서 고민할 필요가 있다.

이미 제2장에서 데이터 거버넌스의 원칙과 모델을 살펴보았고, 제3장 제5절에서 해외 사례의 시사점을 통해 이를 확인하였기 때문에, 여기서는 간략하게 우리가 갖추어야 할 데이터 거버넌스 체제의 주요 요소를 정리해보고자 한다.

① 데이터 거버넌스 프레임워크

보건의료, 통계 등 주요 분야에서 데이터의 수집, 이용, 제공, 연계 등 전반에 걸쳐 개인정보의 활용과 보호의 균형을 맞출 수 있는 데이터 거버넌스 프레임워크를 도입할 필요가 있다. 관련하여 OECD 데이터 거버넌스 프레임워크의 8가지 핵심요소를 참고할 수 있다.

첫째, 건강 정보시스템은 더 나은 보건의료 및 결과를 위한 연구 혁신과 함께 보건의료의 질 및 시스템 성능의 모니터링과 증진을 지원해야 한다.

둘째, 공공보건, 연구 및 통계적 목적의 데이터 처리 및 2차적 이용은 개인정보 보호를 위한 법제에서 명시하고 있는 안전조치를 조건으로 허용된다.

셋째, 개인 건강 데이터의 수집 및 처리와 관련하여 공중에게 정보를 제공하고 협의한다.

넷째, 연구 및 통계 목적의 건강 데이터의 처리를 위한 인증/승인 절차를 구현한다.

다섯째, 프로젝트 승인 절차는 공정하고 투명하며, 의사결정은 독립적이고 다학제적인 프로젝트 심의 기구의 지원을 받는다.

여섯째, 환자 데이터 프라이버시 보호를 위해 데이터 비식별화 모범 관행이 적용된다.

일곱째, 재식별 및 위반 위험을 줄이기 위해 데이터 보안 및 관리의 모범 관행이 적용된다.

여덟째, 새로운 데이터 소스와 기술이 도입됨에 따라 사회적 이익을 극대화하고 사회적 위험을 최소화하기 위해 거버넌스 메커니즘은 국제적인 수준에서 정기적으로 검토된다.

② 데이터 거버넌스 기구

해외 사례에서 볼 수 있다시피, 데이터 거버넌스 기구로는 정보 거버넌스 기구, 프로젝트 승인 기구, 연구윤리위원회 등이 존재한다. 정보 거버넌스 기구는 개인정보의 활용 및 보호와 관련된 전반적인 원칙과 정책을 관장한다. 프로젝트 승인 기구는 연구의 학술적 가치가 프라이버시 침해 위험성보다 큰지 등과 같은 심사기준에 따라 신청서를 검토하여 승인 여부를 결정한다. 연구윤리위원회는 개인정보 문제를 넘어 연구의 윤리적 이슈를 검토한다. 국내에서도 이러한 거버넌스 기구를 분야별, 기능별로 도입할 필요가 있다.

스코틀랜드의 '공익과 프라이버시 패널(Public Benefit and Privacy Panel for Health and Social Care, PBPP)'의 경우처럼 첫 번째, 두 번째 역할을 함께 하는 기구도 있을 수 있다. 영국 ADRN에서 독립적 전문가 및 비전문가로 구성된 승인 패널이 연구 제안서의 심사를 담당하는 것처럼 이러한 거버넌스 기구에는 시민사회의 참여를 고려할 수 있을 것이다.

만일 이러한 거버넌스 기구가 지역별, 분야별로 다수 존재한다면, 영국의 '행정데이터 작업반'이 권고한 바와 같이 인증 혹은 승인의 기준이나 절차를 전국적으로 일관성있게 이루어질 수 있도록 하기 위한 조치도 필요할 것이다. OECD 역시 검토와 승

인 절차가 '증거기반' 원칙에 의해 평가가 이루어져야 하고 객관적이며 공정해야 하며, 적시에 일관성을 촉진하는 방식으로 이루어져야 한다고 권고한 바 있다.

또한, 데이터 거버넌스 기구는 데이터의 수집·이용·제공과 관련한 구체적인 원칙·정책·절차 등을 가이드라인이나 매뉴얼로 정리하고 공개해야 한다.

앞서 지적했듯이, 개인정보 보호위원회는 데이터 거버넌스와 관련한 원칙 및 정책 수립, 다른 거버넌스 기구에 대한 자문 등 국내 데이터 거버넌스 체제 수립에 적극적인 역할을 할 필요가 있다.

③ 연구 데이터 허브

데이터 연계·결합을 통한 데이터의 활용도를 높이고자 한다면 연구 데이터 허브의 장점은 충분히 고려될 만하다. 데이터 허브의 역할이 없다면 연구자들은 데이터에 접근하기 위하여 해당 데이터 보유기관을 개별적으로 접촉해야 하고, 이에 필요한 시간과 비용은 현실적으로 연구를 제약하는 요인으로 작용할 수 있기 때문이다. 연구 데이터 허브는 연구자를 대신해서 데이터 보유기관과 데이터 접근에 대해 협의하고 데이터 보유기관에 법적 자문을 제공할 수 있다. 또한, 데이터 보유기관으로부터 데이터를 제공받아 안전하게 보유·관리하고 연구자가 안전한 환경에서 데이터에 접근할 수 있도록 하는 역할을 한다. 연구 데이터 허브는 데이터 보유기관 사이의 조정, 데이터 표준의 수립이나 품질의 관리, 데이터에 대한 보안, 데이터 연계 방법의 개발 등의 역할을 맡을 수도 있다.

앞서 해외 사례에서 본 바와 같이, 연구 데이터 허브는 직접 데이터를 보유하지는 않고 연구 프로젝트별로 데이터 보유기관으로부터 데이터를 받아 접근을 매개하는 연합형(federated type)과 데이터 제공기관으로부터 일상적으로 데이터를 제공받아 보유, 관리하는 중앙형(centralized type) 모델로 나뉜다.

통계청과 같이 이미 일정한 데이터를 보유하고 있다면 중앙형 모델을 채택할 수 있겠지만, 그렇지 않다면 연합형 모델이 과도한 개인정보 집적을 피할 수 있다는 점에서 바람직할 것이다.

④ 데이터 연계 모델

해외의 많은 사례에서 '신뢰할 수 있는 제3자(TTP)' 모델, 혹은 '방화벽 단일 센터' 모델을 채택하고 있음을 알 수 있었다. TTP 모델은 연계키를 생성하는 기관을 데이터 연계 및 접근을 제공하는 기관과 분리하고 있다. 그러나 '방화벽 단일 센터' 모델

역시 내부적으로 각 기능을 담당하는 부서의 엄격한 분리를 원칙으로 하고 있다.

반면, 국내의 데이터 연계 전문기관의 경우에는 데이터 보유기관이 연계키를 생성하고 연계된 데이터를 다시 데이터 보유기관에 제공한다는 점에서 문제가 있다.

국내에서도 어떠한 방식을 채택하든, 콘텐츠 데이터의 보유, 연계키 생성, 데이터의 연계, 데이터의 제공 등을 담당하는 사람 혹은 기관의 분리를 통해 개인정보 침해의 위험성을 최소화하는 데이터 연계 모델을 도입할 필요가 있다. 해외의 많은 사례에서 '신뢰할 수 있는 제3자(TTP)' 모델은 연계키를 생성하는 기관을 데이터 연계 및 접근을 제공하는 기관과 분리하고 있다는 점을 참고할 수 있다.

⑤ 안전조치

데이터의 수집, 저장, 연계, 제공 등 전 과정에 걸쳐서 개인정보 보호 및 보안을 위한 조치들이 취해져야 한다. 예를 들어, 영국 ADRN의 경우처럼, 연구자(safe people), 연구 프로젝트(safe project), 환경(safe environment), 데이터(safe data), 결과물(safe results) 등에 대한 안전조치가 취해져야 한다.

우선 연구자 혹은 연구기관에 대한 자격요건이 정해져야 한다. 스코틀랜드의 eDRIS의 경우처럼 승인된 연구기관에만 자격을 부여할 수도 있다. 최소한 연구자가 데이터를 적절하게 다룰 수 있는 능력을 갖추었는지 평가해야 하며 이를 위한 훈련을 마련하고 이수하도록 요구해야 한다. 또한, 연구자 혹은 연구기관과의 계약 체결을 통해 연구자들이 개인 식별을 시도하거나 연구목적 외로 이용하는 등 이용규칙을 위반한 경우에 제재할 수 있어야 한다. 연구자뿐만 아니라 데이터 보유기관 및 데이터 허브에 종사하는 직원들에 대한 접근권한 제한, 교육과 훈련, 기밀성 계약 역시 필요하다.

연구 프로젝트가 개인정보 침해 위험성보다 큰 공익성이나 학술적 가치가 있는지에 대한 평가도 엄격하게 이루어질 필요가 있다. 연구의 내용에 따라 프라이버시에 미치는 영향이나 공익적 가치가 달라질 것이다. 요구되는 데이터가 연구에 필요한 최소한의 것인지, 연구 과정에서 개인정보 보호를 위한 제반 조치가 취해졌는지 등에 대한 평가도 포함되어야 한다.

데이터에 대한 접근은 엄격한 안전시설 내에서 이루어져야 한다. 안전시설에서는 데이터 보안을 위한 엄격한 보안 조치가 취해지며 데이터의 이전도 암호화된 보안 전송을 이용하고, 보통 보안시설 내 전자기기 등 기록 가능한 매체의 소지가 제한되며 보안시설 외부에서 데이터의 반입이나 반출도 제한될 수 있다. 또한, 시설 내에서 연구자들의 행위는 향후에 감사를 받을 수 있도록 모두 기록된다. 원격접근을 허용하는

경우에도 이용자에 대한 인증, 보안 접근, 로그 기록의 모니터링 등 보안 조치가 이루어져야 한다. 데이터의 성격이나 위험성에 따라 원격접근 허용 여부 등 접근방식을 달리할 수도 있을 것이다. 이에 대한 규율과 기준 마련이 필요하다.

연구 결과물은 개인정보 침해가 없도록 공개되기 전에 통계 전문가에 의해 철저하게 검토되어야 한다. 이미 통계적 노출 제어를 위한 국제적인 가이드가 마련되어 있다.

식별 데이터의 암호화나 익명화, 안전한 데이터 전송, 접근 통제 등 데이터의 수집, 보관, 이전, 제공 등 전 과정에 걸쳐 적절한 보안 조치가 취해져야 한다. 이에 대한 외부 인증을 받는 것은 관련 기관의 신뢰성을 높이는 데 도움이 될 것이다. 다만 비식별 조치 등 안전조치를 취했다는 이유로 개인정보 보호법이 규정한 관련 책임이 일률적으로 면제되는 것은 아닐 것이다.

또한, 이러한 개인정보 보호 및 보안 조치의 적절성을 검토하기 위해 프라이버시 영향평가를 수행할 수 있을 것이다. 이미 해외의 많은 기관이 프라이버시 영향평가를 수행하고, 이를 공개하고 있었다. 우리의 경우도 프라이버시영향평가를 수행하도록 할 필요가 있다.

⑥ 연구지원단

연구의 설계, 승인, 안전한 환경에서의 데이터 접근에 있어서 연구자들을 돕는 창구 역할을 하는 연구지원단의 설치를 고려할 수 있다. 예를 들어 영국 스코틀랜드의 eDRIS는 전 과정에 걸쳐 연구 코디네이터를 지정하여, 연구 설계 지원, 연구에 대한 전문가 자문, 데이터 접근에 필요한 허가 획득 지원, 데이터 제공기관과의 연결, 보안 시설을 통한 데이터 접근 등 제반 서비스를 제공하고 있다. 앞서 언급한 데이터 허브가 구축된다면, 해당 기관이 이러한 연구지원단의 역할을 함께 수행할 수 있을 것이다.

⑦ 투명성과 대중 참여

OECD나 UN 등 국제기구는 물론이고 해외 기관이나 주요 문서에서 공통적으로 강조하고 있는 것이 데이터 제공 및 이용 과정 전반에 걸쳐서 개인정보 보호 및 보안이 지켜지고 있다는 것에 대한 일반 공중의 신뢰이다. 그리고 이러한 대중의 신뢰를 얻기 위해서는 높은 수준의 투명성과 참여가 필수적이다. 즉, 원칙과 절차, 진행된 사업 내용에 대한 정보를 투명하게 공개해야 하며, 정책 결정 과정에 시민들과 다양한 이해당사자들이 참여할 수 있도록 해야 한다.

국내에서도 개인정보를 둘러싼 논란이 발생하는 가장 큰 원인 중 하나가 그 운영 과정의 불투명과 충분한 사회적 공론화 과정을 거치지 않은 일방적인 정책 추진인 경우가 많다. 이해당사자의 협의를 위한 위원회를 구성하거나 공청회 등을 하기도 하지만, 담당 기관이 선정한 패널로만 구성되거나 자료 및 회의 내용을 공개하지 않는 경우가 많다. 물론 이는 비단 개인정보와 관련된 문제는 아닐 것이다.

국내에서도 현재 빅데이터 환경에서 데이터의 활용도를 높이기 위한 노력이 진행되고 있다. 그러나 지금까지 얘기한 전반적인 데이터 거버넌스 체제의 구축과 이를 위한 사회적인 논의가 충분히 이루어지지 않는다면, 오히려 데이터의 활용에 대한 대중들의 불신을 불러일으킬 수 있다. 정책 추진 과정에서 투명성과 대중 참여의 원칙을 견지할 필요가 있다.

[부록1]

건강 데이터 거버넌스에 대한 OECD 이사회의 권고

Recommendation of the OECD Council on Health Data Governance[349)]

도입

보건의료 데이터는 보건의료 시스템의 성능과 의료 서비스의 품질을 향상시키는 데 점점 더 중요한 요소가 되고 있으며, 진료를 개선하고 생명을 구하는 과학적 발견에 기여한다. 데이터의 규모와 다양성이 증가함에 따라 이 데이터들로부터 추가 정보가 추출될 가능성도 높아지고 있다. 특히 병원, 의사, 약사, 연구자, 바이오뱅크, 통계부서, 의료기기 및 어플리케이션 산업 등 데이터를 수집하는 여러 기관 간에 데이터가 연계되고 결합될 때 더욱 그렇다.

보건의료 데이터는 속성상 민감하고, 그 데이터 공유와 사용 확대는 데이터 유출과 오남용 위험을 야기해서 개인에게 개인적, 사회적, 재정적인 위해를 끼칠 수 있고, 보건의료서비스 제공자와 정부에 대한 대중의 신뢰를 떨어뜨릴 수 있다. 이러한 위험성을 적절하게 감소시키고 관리해야 한다.

데이터 수집 관행의 범위와 규모 또한, 개인 정보 수집 및 이용에 대한 동의 등 고심해야 마땅한 기존 개인정보 보호 기준 및 절차의 구현에 있어 새로운 문제를 야기한다. 또한 새로운 분석 기술의 잠재적인 편익에 도달할 수 있도록 교육 및 인식 제고, 업무능력 개발, 기술적 수단의 홍보 등을 통해 법적 개인정보 보호를 보완해야 할 중요성이 강조되고 있다.

이 OECD 권고는 각국에 보건의료 데이터 거버넌스 체계를 개발하고 실행해서 공익을 위해 보건의료 데이터의 사용을 가능케 하면서 동시에 프라이버시를 보장할 것을 촉구한다. 이 권고는 12가지 고수준 원칙에 따라 구성되었으며, 폭넓은 이해당사자의 참여로부터 개인건강정보 수집과 이용에 대한 효과적인 동의와 선택 체계 및 그 감독과 평가에 이르기까지 포괄하고 있다. 이 원칙들은 보다 조화로운 국내 데이터 거버넌스 체계를 장려하는 환경을 갖춤으로써, 보다 많은 국가들이 연구, 통계, 진료 품질 증진 뿐 아니라 국제 비교 목적으로 건강정보를 사용할 수 있도록 한다.

---

349) https://www.oecd.org/health/health-systems/Recommendation-of-OECD-Council-on-Health-Data-Governance-Booklet.pdf

이 권고는 2016년 12월 13일 OECD 이사회에서 채택되었으며 2017년 1월 17일 파리에서 개최된 OECD 보건 장관 회의에서 환영받았다.

권고

Ⅰ. 이 권고가 건강관련 공익 목적으로 개인 건강정보의 접근과 처리에 적용된다는 점에 동의한다. 이는 보건의료의 품질, 안전, 책임성을 향상시키고, 공중보건 위기를 감소시키며, 건강 경과를 개선하기 위해 새로운 진단 도구와 치료법을 발견 및 평가하고, 보건의료자원을 효율적으로 관리하며, 과학과 의학의 진보에 기여하면서, 공공정책 기획과 평가를 증진하는 한편, 보건의료에 대한 환자들의 참여경험을 향상시키기 위함이다.

Ⅱ. 권고 목적상, 다음 기술적 용어들에 대한 이해를 돕기 위해 간단히 기술할 필요가 있다.

- “개인 건강정보”란 건강과 관련하여 개인을 식별하거나 식별할 수 있는 모든 정보를 의미하며 관련된 여타의 개인정보를 포함한다.

- “개인 건강정보 처리”란 개인 건강정보에 관여하는 모든 정보 관련 작업으로 정보 수집, 이용, 제공, 보관, 기록, 수정, 검색, 이전, 공유, 연계 및 결합, 분석, 삭제를 의미한다.

- “비식별화”(de-identification)란 일련의 개인 건강정보를 대체하여 그 결과적인 정보가 특정 개인과 쉽게 관련될 수 없도록 처리하는 것을 의미한다. 비식별화는 익명화가 아니다. “재식별화”란 비식별화된 정보에 연관된 사람을 식별하기 위하여 비식별화된 정보에서 추정되는 정보를 처리하는 것을 의미한다.

Ⅲ. 국가적인 보건의료 데이터 거버넌스 체계를 수립하고 실행할 것을 정부에 권고한다. 이는 프라이버시, 개인 건강정보의 보호와 정보 보안을 증진함과 동시에, 건강관련 공익 목적으로 개인 건강정보가 사용되는 것을 장려하고 가용성을 높이기 위함이다. 이러한 보건의료 데이터 거버넌스 체계는 다음을 포함해야 한다.

1. 관여와 참여.

특히 공공적인 협의 과정을 통해서 광범위한 이해당사자가 관여하고 참여해야 한

다. 이러한 관여와 참여는 이 체계 하에서 이루어지는 개인 건강정보의 처리가, 공익에 기여하는 한편, 자신의 개인정보가 보건의료 시스템관리, 연구, 통계, 기타 공익에 기여하는 보건의료 관련 목적을 위해 사용되고 또한 보호될 것이라는 개인의 합리적인 기대는 물론 사회적 가치와 부합해야 한다는 관점 속에 이루어져야 한다.

2. 개인 건강정보 처리에 있어 정부내 의견조정 및 공공/민간 부문 모두에서 기관간 협업 증진.

이러한 협업은 다음과 같아야 한다.

(1) 공통 데이터 요소 및 형식, 품질 보증, 데이터 상호운용성 기준을 장려할 것.

(2) 프라이버시를 보호하고 정보 보안을 보장하는 한편 보건의료 시스템관리, 통계, 연구 및 기타 공익에 기여하는 보건의료 관련 목적을 위해 데이터를 공유하는 데 있어, 장애물을 최소화하는 공통 정책과 절차를 장려할 것.

3. 개인 건강정보 처리 및 공중보건 대응을 위해 사용되는 공공부문 보건의료 정보시스템의 성능 검토.

이러한 검토는 다음을 포함해야 한다.

(1) 데이터 가용성, 품질, 사용적합성, 접근성은 물론 프라이버시 보호 및 정보 보안을 포함할 것.

(2) 보건의료 시스템관리, 연구, 통계 및 기타 공익에 기여하는 보건의료 관련 목적을 위해 허용되는 정보처리 요소. 특히 데이터셋 이전 및 데이터셋 기록 연계에는 적절한 보호조치가 뒤따를 것.

4. 개인에 대한 명확한 정보 제공.

이러한 정보제공은 다음을 보장해야 한다.

(1) 개인 건강정보가 개인으로부터 수집될 때, 적법한 제3자의 접근 가능성, 처리 이면의 기본 목표, 처리의 혜택, 제공의 법적 근거 등 개인 건강정보의 처리에 대한 정보를 명확하고, 정확하고, 쉽게 이해가능하고, 눈에 띄는 용어로 제공할 것

(2) 개인 건강정보에 대한 중대한 유출이나 여타 오남용에 대하여 개인들에게 시기적절하게 고지할 것. 개인 고지가 현실적이지 않을 경우 효과적인 공중 통신수단을 통한 고지가 가능함.

5. 설명후 동의(informed consent) 및 적절한 대안.

(1) 동의 체계는 다음과 같아야 한다.

a) 자신의 개인 건강정보 처리에 대한 개인의 동의가 필요한지 여부가 명확해야 하고, 동의가 필요한 경우 그 결정을 위해 사용되는 기준이 명확해야 함. 유효한 동의를 구성하는 요소가 무엇이고 어떻게 동의를 철회할 수 있는지 명확해야 함. 동의를 구하는 것이 불가능하거나 비현실적이거나 보건의료 관련 공익적 목적 달성과 양립하지 않는 환경 등에서는, 동의 요구를 대신할 수 있는 적법한 대안 및 예외가 명확해야 하고, 이러한 처리에는 이 권고에 부합하는 보호조치가 뒤따라야 함.

b) 개인 건강정보가 동의에 기반한 경우, 이 동의는 충분한 설명 후 자유롭게 이루어진 동의일 때 유효함. 개인이 장래의 정보 사용에 대해 동의하거나 철회할 때 그 방법이 명확하고 눈에 띄고 사용하기 쉬웠을 때 유효함.

(2) 개인 건강정보의 처리가 동의에 기반하지 않은 경우, 다음과 같은 체계가 실현 가능한 한 보장되어야 한다.

a) 개인들은 자신의 개인 건강정보 처리에 대한 선호를 표현할 수 있어야 함. 특정 환경에서의 처리에 반대할 수 있어야 할 뿐 아니라, 자신의 개인 건강정보를 연구 및 기타 보건의료 관련 공익적 목적 하에 공유할 것을 적극적으로 요청할 수 있어야 함.

b) 정보 처리 목표와 요청이 지켜질 수 없을 때, 개인은 관련 법적 근거 등 그 사유를 알 수 있어야 한다.

6. 적절한 경우 개인 건강정보를 연구 및 기타 보건의료 관련 공익적 목적으로 사용하는 데 대한 검토와 승인 절차.

이러한 검토와 승인 절차는 다음과 같아야 한다.

(1) 데이터 사용 제안이 공익적인지 여부에 대하여 증거기반 평가가 이루어져야 함

(2) 강력하고 객관적이고 공정해야 함

(3) 시기적절하고 결과의 일관성을 촉진하는 방식으로 이루어져야 함

(4) 정당한 이익을 보장하면서 투명하게 이루어져야 함

(5) 정보 처리가 개인 및 사회에 미치는 편익과 위험성, 그리고 위험성 경감을 평가하는 데 필요한 전문성을 가진 이들에 의해 수행되는 독립적이고 학제적인 검토가 이루어져야 함

7. 건강정보 프라이버시 보호와 정보보안 및 기관의 상업적, 기타 정당한 이익을 침해하지 않는 공공 정보 체계를 통한 투명성.

공공 정보 체계는 다음 요소들을 포함해야 한다.

(1) 개인 건강정보 처리 목적과 이것이 기여하는 보건의료 관련 공익적 목적 뿐 아니라 그 법적 기반.

(2) 개인 건강정보 처리를 승인하는 데 사용되는 절차와 기준. 승인 데이터를 수령하는 자 목록 등 승인 결정의 요지.

(3) 보건의료 데이터 거버넌스 체계의 실행 및 그 효과와 관련한 정보.

8. 프라이버시 보호와 정보 보안을 보장하고 자신의 개인정보 이용에 대한 개인의 통제권을 촉진하는 동시에 개인 건강정보 재사용 및 분석 가용성을 가능케 하는 기술적 수단의 잠재성 최대화와 발전 증진.

9. 감독과 평가 체계.

(1) 개인 건강정보의 사용이 보건의료 관련 공익적 목적 취지에 부합하고 혜택을 가져올 것으로 기대되는지, 개인 건강정보에 대한 프라이버시 보호 및 정보 보안, 정보 유출 및 오남용에 대한 국가적 요구사항에 대한 불충분 이행 등 그 사용이 부정적인 결과를 야기하지 않는지 여부에 대해 평가해야 함. 또한 그 평가 결과를 다음과 같이 향후 개선 과정에 반영했는지 여부를 평가해야 함.

a) 개인 건강정보 가용성의 발전에 대한 정기적인 검토. 보건의료 연구 및 관련 활동에 대한 수요. 공공정책 수요.

b) 보건의료 데이터 거버넌스와 관련하여 프라이버시, 개인 건강정보 보호 및 정보 보안에 대한 위험성 관리 정책 및 실행에 대한 정기적인 평가와 갱신.

(2) 위와 같은 개인 건강정보 처리에 사용되는 기술의 성능, 신뢰성, 취약점에 대해 정기적인 검토 및 평가를 장려함.

10. 위와 같은 개인 건강정보 처리에서 사용되는 프라이버시 및 보안 수단에 대하여, 일반적인 기준과 정보처리기술에 부합하는 적절한 교육훈련 및 기능 개발 방안 수립.

11. 통제권과 보호조치의 실행.

(1) 적절한 감사체계가 동반되어, 개인 건강정보 처리에 대한 명확하고 확고한 책임성 영역을 보장해야 함

(2) 개인 건강정보는, 개인 건강정보 처리와 관련한 역할과 책임에 상응하고 관련 직업 윤리 강령에 부합하는 적절한 정보 프라이버시와 정보보안 훈련을 이수한 직원들의 책임 하에서만 처리된다는 조건을 갖추어야 함

(3) 개인 건강정보 처리 기관들로 하여금, 조직과 직원들에게 프라이버시 보호와 정보 보안에 대한 법적 의무를 고지하는 등 조직 정보 보안 프로그램과 협업하고 책임지는 직원을 지정하도록 장려해야 함

(4) 특히 새로운 계획을 수립하고 신규 실행과정을 도입할 때 공식적인 위험 관리 절차를 포함해야 함. 이는 원치 않은 데이터 삭제, 재식별화, 유출과 기타 오남용 등 위험요소에 대하여 정기적으로 평가하고 대응하며, 정기적으로 갱신되어야 함

(5) 보건의료 관련 공익적 목적에서 개인 건강정보의 효용성을 유지하는 동안에는, 실현가능한 한 프라이버시를 보호하고 정보 보안을 위해 설계된 기술적, 물리적, 조직적 수단을 포함해야 함. 이러한 수단은 다음을 포함해야 함

a) 개인 건강정보에 대한 비식별화 등 개인의 식별을 한정하는 체계. 한편 재식별화를 승인받는 경우 그 허용을 감안하여 데이터 사용 제안을 고려하는 체계. 재식별화는 보건의료 시스템관리, 연구, 통계, 기타 보건의료 관련 공익적 목적으로 장래의 데이터 분석을 수행하기 위해 승인될 수 있으며, 적절한 경우 개인들에게 특정 여건이나 연구 결과를 고지하기 위해서 승인될 수 있음.

b) 개인 건강정보의 효용성을 관리하는 동안 혜택 최대화 및 위험 관리에 기여하는 처리를 위해서, 제3자와 개인 건강정보를 공유할 때의 계약. 이러한 계약은 안전한 데이터 이전을 위한 협약사항을 명기해야 하며, 불이행시 효과적으로 제재할 수 있는 적절한 수단을 포함해야 함

c) 실현가능하고 적절한 경우 데이터 접근 보안 센터 및 원격 데이터 접근 설비 등 제3자 데이터 이전에 대한 대안을 고려할 것.

d) 개인 건강정보에 대한 개인적인 접근에 대해서는 강력하게 신원을 확인하고 인증할 것.

12. 개인 건강정보를 처리하는 기관들에게 국내 보건의료 데이터 거버넌스의 기대 수준에 부합하다는 사실을 입증할 것을 요구.

여기에는 기관들의 개인 건강정보처리에 대한 인증 또는 인가 체계 설립이 포함될 수 있다. 다만 이러한 인증 또는 인가가 개인 건강정보 처리 기준 실행에 기여하고 거버넌스 인정 기준에 부합하는 성능을 입증하는 경우에 한한다.

[부록 2]

데이터 2차 분석 모범 사례 (독일)

Good Practice in Secondary Data Analysis (GPS) (AGNES et al., 2008)[350]

가이드라인 8: 개인정보 보호

개인정보 자기결정권 보호를 위한 현행 개인정보 보호조항은 2차 데이터 분석을 계획하고 실행할 때 존중되어야 한다. 전적으로 필요한 개인정보만 수집하고 보관하도록 요구하는 (독일 연방개인정보보호법 제3a조) 개인정보 회피 및 최소화 원칙 등 현행 개인정보 보호조항 및 개인정보 처리자와 관련된 여타의 규제가 적용되는 경우 이를 존중해야 한다. 연구 프로젝트와 관련하여 개인정보를 다루는 누구나 관련 법률 조항의 내용, 범위, 역할에 대한 정보를 제공받아야 한다. 개인정보와 관련한 연구의 경우, 학술 및 연구의 자유는 물론 개인정보 자기결정권에 대한 개인의 권리가 존중되어야 한다.

권고 8.1 데이터 제공의 목적

(개인정보 보호와 관련하여) 데이터 제공의 목적은 연구 과제에 부합해야 하고 (가이드라인 2 참조) 서면으로 작성되어야 한다.

권고 8.2 가명화와 익명화

데이터 사용은 독일연방개인정보보호법(제3a항 개인정보 회피 및 최소화 원칙)에 명시된 가명화 및 익명화 수단을 통해 이루어져야 한다. 이 단계에서 데이터 보유자의 참여가 고려되어야 한다.

권고 8.3 탈가명화와 재식별

탈가명화를 의도하는지 여부, 만일 그렇다면 어느 경우에 그러한지, 계약일반조건을

350) https://dgepi.de/fileadmin/pdf/leitlinien/gps-version2-final_ENG.pdf

서면으로 규정하는 것이 중요하다. 분석에 있어, 무책임한 재식별화를 방지하기 위해 (기술적 및 계약적으로) 적절한 수단이 채택되어야 한다.

권고 8.4 개인정보 제3자 제공

원칙적으로, 어떠한 개인정보 제공도 데이터 소지자(data owner)만이 할 수 있다.

권고 8.5 외부 데이터 소스와 개인정보 연계

명문화되지 않은 외부 데이터 소스와의 모든 개인정보 연계는 개인정보 보호조항을 준수해야 한다.

권고 8.6 개인정보 보호의 책임자

모든 2차적 분석은 정보보안 및 개인정보 보호에 대한 국내적, 국제적 기준을 준수해야 한다. 연구 부서 내에서 개인정보 처리의 책임을 지는 사람이 지정되어야 하며, 그는 이러한 기준에 대한 준수를 감독해야 한다. 그는 이러한 의무에 적절한 자격을 갖추고 있어야 한다.

권고 8.7 삭제 기한

2차적 데이터 분석을 위해 제공된 데이터가 개인정보 보호를 위해서 연구 목적 달성 후 삭제되거나 익명화되어야 한다면, 이는 권고 제6.2조 및 제6.7조에서 규정한 기준 데이터셋 및 분석 데이터셋의 보관 요건에 부합하게 이루어져야 한다. 마찬가지로, 삭제 기한을 설정할 때, 가이드라인 7에서 규정한 바와 같이 2차적 이용으로 획득된 결과를 점검할 수 있는 기회가 제공되어야 한다.

권고 8.8 개인정보 보호 책임자와 협력

2차적 데이터 분석 기획 단계서부터 가능한 빨리 개인정보 보호의 정당한 책임자와 접촉해야 할 필요성을 염두에 두어야 한다.

[부록 3]

CPRD 접근 라이선스(이용조건) 표준안: 이용 허가와 제한 세부사항

CPRD Access License Template: Details of permitted and restricted use[351)]

3. 허용되는 이용

3.1 제2.1조에서 부여된 라이선스는 다음과 같아야 한다.

(A) 라이선스 비용 지불은 제4조를 따라야 한다.

(B) 사용자 혹은 그 계열사가 이 라이선스에 따라 데이터나 기타 정보를 이용할 때 비영리적인 보건의료 및 건강 연구 목적으로 제한된다. (혼선을 피하기 위해 언급하자면) 이는 사용자에게 다음을 금지하는 것은 아니다.

(1) 해당 연구와 관련된 합리적인 수준의 직접적인 운영 비용의 회수

(2) 수익이 데이터의 분석이나 해석을 통해 사용자가 부가한 가치에 전적으로 기인하는 한, 사용자의 연구 결과의 응용으로 발생한 수익의 회수

(C) 지정된 이용자는 제6조에 따른 교육훈련을 받아야 한다는 요구 사항을 준수해야 한다.

(D) 사용자는 제9조에 명시된 제한 사항을 준수해야 한다.

3.2 이 사용 제한은 허가권의 종료나 만료 이후로도 적용된다.

3.3 제2.1조에 따라 부여된 라이선스의 범위에 대해 사용자가 의문이 있을 경우에는, 이를 명확히 하기 위해 라이선스 허가자에게 연락을 취해야 한다.

(중략)

---

351) https://www.cprd.com/_docs/CPRD%20Access%20Licence%20Template.pdf

9. 데이터 이용에 대한 제한

9.1 해당 데이터 자체든 혹은 다른 데이터와 어떠한 형태로 결합해서든, 사용자는 서비스(CPRD GOLD 서비스)의 제공에 따라 사용자가 획득한 데이터나 다른 정보를 다음과 같은 목적으로 사용하거나 사용하려고 시도해서는 안된다.

(A) 환자의 식별, 접촉, 타케팅

(B) 일반의 혹은 일반의료행위의 식별, 프로파일링, 접촉, 타게팅

(C) 광고 및 영업 효과 연구

또한 사용자는 데이터 사용 결과로서 공개되거나 제3자에게 제공된 보고서, 논문, 통계표가 환자, 그를 진료한 일반의, 혹은 처치된 일반의료행위를 식별하거나 다른 이가 식별하는 데 사용되지 않도록 보장해야 한다. 사용자는 이 서비스를 통해 접근 가능한 데이터베이스에서 어떤 개인, 일반의, 일반의료행위의 식별에 사용될 수 있는 정보가 있다는 것을 발견했을 때는, 23조에 따라 전달된 고지의 방식을 통해 서면으로 즉시 라이선스 허가자에게 알려야 한다.

9.2 사용자는 라이선스 허가자의 서면 허가를 받고 감사(audit)의 목적으로 사용자의 규제 당국에 데이터를 제공하는 경우를 제외하고는, 데이터베이스에서 자신이 다운로드한 어떠한 데이터도 판매, 양도(라이선스에 따라 계열사에 양도하는 것이 허가된 경우는 제외), 거래, 혹은 달리 처분해서는 안된다. 명확히 하자면, 이는 사용자나 그 계열사가 의료 및 학술 저널에 출판한 논문에 혹은 의료 및 학술적 성격의 발표문에 데이터를 포함하는 것을 배제하는 것은 아니다. 다만 그 데이터가 관련 논문과 발표를 뒷받침하기 위해 꼭 필요한 것으로만 제한되어야 한다.

9.3 제7.5조, 제9.4조 및 제9.5조에 의거하여, 사용자는 제3자(계열사는 제외한다)에게 데이터에 대한 접근, 연구, 분석, 참조, 기타 사용을 허용해서는 안 되며, 제3자가 사용자가 데이터베이스로부터 다운로드 한 어떠한 데이터를 복제하도록 허용해서는 안 된다.

9.4 사용자는 라이선스 허가자의 서면 허가가 있는 경우를 제외하고는, 어떤 계약상대방에게도 데이터에 대한 접근, 연구, 분석, 참조, 기타 사용을 허용해서는 안 된다. 허락을 받은 계약자는 데이터에 접근하기 전에 라이선스 허가자가 제공한 형식의 비밀유지협약서에 서명해서 라이선스 허가자에게 돌려주어야 한다. 사용자는 이 조항

에 따른 계약자에게 이전하는 어떠한 데이터도 사용자와 계약자 간의 계약 만료시 사용자에게 회수되도록 보장해야 한다.

9.5 라이선스 허가자가 사용자에게 연구 설명 계획안 제출이 불요하다고 서면으로 설명하지 않은 한, 사용자는 프로젝트의 결과물이 제3자에게 공유될 수 있는 프로젝트 수행을 위해 데이터를 사용하기 위해서는 먼저 허가자의 승인을 얻어야 한다. 허가자는 사용자가 제출한 계획안이 적절한 경우 독립적 학술자문위원회에 자문을 위해 전달하고, 이후 허가자는 계획안의 수용 여부를 사용자에게 확인해주기 위해 다시 연락한다. 계획안 제출 절차는 웹사이트로도 가능하지만, 웹사이트 사용이 불가한 경우 사용자 요청에 응할 수 있어야 한다. 이 조항에 따른 허가자의 결정은 허가자를 통한 영국 면허당국의 견해로 해석되어서는 안 된다.

9.6 사용자는 다음과 같은 경우 검토를 위해 허가자가 승인한 외부 전문가에게 최대 3백 건(혹은 허가자가 서면으로 동의한 경우에는 그 정도의 건수)의 사례 기록(case history)을 보내야 한다.

(A) 사례 기록과 그 사례수는 관련 프로젝트의 목적에 필요한 경우로 엄격하게 제한해야 한다.

(B) 전문가는 제공된 정보의 기밀성을 인지해야 한다.

(C) 허가자로부터 달리 동의받지 않는 한, 프로젝트 당 이러한 검토를 수행하는 전문가의 최대 수치는 10명이어야 한다.

(D) 검토 목적으로 확실히 필요한 경우 외에는, 전문가들이 사례 기록을 복사하는 것은 허용되지 않는다.

(E) 검토가 끝나는 대로, 허가자는 각 전문가에게 보내진 사례 기록과 복사본이 확실히 허가자에게 회수되거나 파기되도록 하고, 사용자는 이것이 완료되는 대로 허가자에게 알려야 한다.

참 고 문 헌

<국내 문헌>

- 건강보험심사평가원, “보건의료빅데이터센터 이용 가이드”, 2017.
- 김두만, “국가통계작성 기획 및 승인관리”, 제5회 국가통계방법론 심포지엄, 2015.
- 김종태, ″공공데이터 개방(Open Data)의 활성화 방안″.KIPA 조사포럼 2016.06(Vol.17)
- 관계부처 합동, 「개인정보 비식별 조치 가이드라인」, 2016. 6. 30.
- 광운대학교 산학협력단, “데이터 관리체계 개선방안 연구”, 2015.12.
- 도세록 등, “의료이용 통계생산 개선에 관한 연구”, 한국보건사회연구원, 2012.10
- 미래창조과학부·한국정보화진흥원·K-ICT 빅데이터센터, “2016 글로벌 빅데이터 융합 사례집 : 2015 빅데이터 시범사업 및 국내외 사례를 중심으로”, 한국정보화진흥원, 2016.
- 박숙희, “표본연구DB 개방, 성과 및 향후 비전″, 건강보험 빅데이터 개방, 2차년도 성과 공유 심포지엄, 2015.
- 배종면, “국가 암등록사업 현황 및 향후 발전 방향″, 국립암센터 제4회 심포지움, 2002.2.
- 서울아산병원, “개인정보의 연구 목적 처리를 위한 법・제도 개선 방안 연구-보건의료연구 분야를 중심으로 -”, 개인정보보호위원회 정책연구용역 결과보고서, 2016.12.
- 송대섭·김연정, “‘한국인 유전체역학조사사업’ 수집자료 통합·정제 지침서 공개″, 주간 건강과 질병 제7권 제10호, 2014.
- 심우민, “개인정보 비식별 조치에 관한 입법정책적 대응과제”, 국회입법조사처 현안보고서 Vol 305, 2017.5.24.
- 알렉산더 로스나겔, “독일의 개인정보보호법″, 개인정보 보호제도의 개선을 위한 한독 국제 심포지엄 발표문, 2004.11.1.
- 안동대학교 산학협력단, “의료정보안내서”,「보건의료빅데이터 활용 고도화 방안 연구」부록
- 오미애, ″보건복지분야 데이터 연계 필요성 및 활용방안″, 보건복지포럼, 2015.9
- 오미애 등, “보건복지통계정보 생산 및 활용 촉진을 위한 마이크로데이터 통합 연계 방안″, 한국보건사회연구원 연구보고서, 2014.2.29.
- 이덕형, “암 빅데이터 플랫폼 구축 사업 기획 연구”, 보건복지부, 2014.10.
- 이인묵·박선영·민재홍, “교통카드데이터 공공이용 활성화를 위한 정책방안”, 2014년도 한국철도학회 추계학술대회 논문집, 2014.
- 임찬수·최재혁·김경미, “표본코호트DB(국민건강보험)와 조사자료(통계청)의 자료연계에 따른 심층분석″, 통계청 2016년 하반기 연구보고서, 2016.
- 전남대산학협력단, “EU 회원국 및 주요국가의 통계목적의 개인정보 처리규정 연구”, 개인정보보호위원회 정책연구용역 결과보고서, 2016.12
- 정영호·고숙자·이용갑·서남규·태윤희·이원영·이경용·김범수·강영호, “한국의료패널의 활용과 기대효과”. 한국보건사회연구원·국민건강보험공단, 2009
- 통계교육원, “국가통계의 이해″. 2015.12.31.
- 한국조사연구학회, “빅데이터 활용 통계생산방법론 연구용역 결과보고서”. 통계청 연구용역보고서, 2016.
- 행정안전부, “전자정부법의 이해와 해설”. 2010. 8.
- 행정자치부, “개인정보보호 법령 및 지침·고시 해설”. 2016. 12.

<국외 문헌>

- Administrative Data Taskforce, “The UK Administrative Data Research Network : Improving Access for Research and Policy”, 2012.12.
- Antoni, M., Schnell, R., “The Past, Present and Future of the German Record Linkage Center (GRLC)”, Journal of Economics and Statistics 2017.
- ARTICLE 29 DATA PROTECTION WORKING PARTY, “Opinion 05/2014 on Anonymisation Techniques″. 2014.4.10
- B.F.M. Bakker et al., “The System of social statistical datasets of Statistics Netherlands: An integral approach to the production of register-based social statistics”, statistical Journal of the IAOS 30 (2014) 411 - 424. IOS Press.
- Deborah Wagner, Mary Layne, “The Person Identification Validation System (PVS): Applying the Center for Administrative Records Research and Applications’ (CARRA) Record Linkage Software″, Center for Administrative Records Research and Applications, U.S. Census Bureau , 2014.7.1.
- HSC, “Honest Broker Service - Annual Report June 2014 - May 2015″. 2015.6.
- Jerry Fishenden, “Submission on Part 5 of the Digital Economy Bill: “Digital Government”″. 2016.10.5.
- Johanna Eberle, “German Record Linkage Center″, Microdata Computation Centre (MiCoCe) Workshop, 2014.4.29.
- Katie Harron, “Introduction to Data Linkage”, ADRN Publication, 2016.6.
- Milieu, “Overview of the national laws on electronic health records in the EU Member States - National Report for France″. 2014.1.
- NRS, “Scottish Information and Linkage Collaboration (SILC)″. Paper 9 PAMS (15) 23, 2015.11.
- OECD, “Health Data Governance - PRIVACY, MONITORING AND RESEARCH″. OECD Health Policy Studies, 2015.
- OECD, “Recommendation of the OECD Council on Health Data Governance”. OECD Health ministerial Meeting. 2017.1.17
- OHE Consulting Report, “Data Governance Arrangements for Real-World Evidence″. 2015.11
- Robert H McLaughlin, Christina A Clarke, LaVera M Crawley and Sally L Glaser, “Are Cancer Registries Unconstitutional?″, Social Science & Medicine Volume 70, Issue 9, 2010.5.
- Rosalyn Moran, “Proposals for an Enabling Data Environment for Health and Related Research in Ireland″. Health Research Board, 2016.5.
- Sharyl J Nass, Laura A Levit, and Lawrence O Gostin, “Beyond the HIPAA Privacy Rule - Enhancing Privacy, Improving Health Through Research″. National Academies Press (US); 2009.
- Siobhán Morgan, “HSCNI Honest Broker Service″. 2016.5.25.
- Statistics New Zealand, “Data integration manual: 2nd edition″. Available from www.stats.govt.nz. 2015.3

• The Scottish Government, “A Charter for Safe Havens in Scotland - Handling Unconsented Data from National Health Service Patient Records to Support Research and Statistics". 2015.11
• Wellcome trust, “Enabling Data Linkage to Maximise the Value of Public Health Research Data : full report”, 2015.3.

<웹사이트>

• 국립암센터 http://www.ncc.re.kr/
• 나라통계 https://www.narastat.kr/
• 보건의료빅데이터개방시스템 http://opendata.hira.or.kr
• 질병관리본부 국립보건연구원 http://www.nih.go.kr/NIH_NEW/main.jsp
• MDIS 홈페이지 https://mdis.kostat.go.kr
• Administrative Data Research Network(ADRN) https://www.adrn.ac.uk/
• Clinical Practice Research Datalink (CPRD) https://www.cprd.com
• German Record Linkage Center http://www.record-linkage.de/
• healthdata.gov http://healthdata.gov/
• ISD scotland, eRIDS http://www.isdscotland.org/Products-and-Services/eDRIS/
• NCHS Data Linkage https://www.cdc.gov/nchs/data-linkage/index.htm
• Population Health Research Network(PHRN) http://www.phrn.org.au
• Research Data Center (FDZ) http://fdz.iab.de/en.aspx
• SAIL Databank https://saildatabank.com/
• Statistics Canada http://www.statcan.gc.ca
• Statistics Netherlands (CBS) https://www.cbs.nl/en-gb
• Statistics NZ http://www.stats.govt.nz
• United States Census Bureau, Data Linkage Infrastructure https://www.census.gov/about/adrm/linkage.html

데이터 연계·결합 지원제도
도입방안 연구

2017년 12월 발행

발행처 개인정보보호위원회
서울특별시 종로구 세종대로 209
정부서울청사
Homepage : www.pipc.go.kr