

캐나다의 보건·복지·사회보장 행정 데이터 및 공공 데이터 연계

Canadian Administrative Data and Public Data Linkage in Health,
Welfare and Social Security

오미애(한국보건사회연구원 연구위원)

4차 산업혁명에서는 데이터가 자산인 만큼 데이터를 얼마나 잘 활용하느냐가 국가 경쟁력 확보의 중요한 요소이다. 데이터 기반 정책 지원을 위해 각국은 데이터 경제를 선도하고자 범국가적 차원의 대책을 마련하고 있는데, 이를 위해서는 행정 데이터 및 공공 데이터 현황을 살펴볼 필요가 있다. 캐나다의 경우, 통계청이 범캐나다 사회경제적 통계 연구를 촉진하기 위해 사회적 데이터연계 환경(SDLE: Social Data Linkage Environment) 프로그램을 개발하였다. 브리티시컬럼비아주는 인구 데이터(PopData: Population Data BC)를 통해 의료서비스, 인구 건강 및 여러 데이터를 기반으로 주민에 대한 개인 수준의 익명화된 종단 데이터를 제공하고 있다. 이 데이터 세트는 주로 행정 데이터이지만 조사 데이터 및 지역사회, 환경을 나타내는 데이터도 포함된다. 온타리오주는 지역 기반의 건강서비스 분석, 보건 정책, 건강의 사회적 결정요인, 데이터 과학 등의 연구를 위해 온타리오주에 거주하는 1340만 명의 정보를 포함한 임상평가과학연구소(ICES: Institute for Clinical Evaluative Sciences)를 통해 데이터를 제공하고 있다. 의미 있는 데이터 분석 및 연구 결과를 위해 행정 데이터 및 공공 데이터의 연계와 활용이 점점 중요해지고 있다. 행정 데이터 및 공공 데이터 연계는 데이터 기반의 가치 창출 생태계를 조성하기 위한 필수 요소로, 연계 데이터의 활용 및 데이터 제공 프로세스가 체계적으로 이루어질 수 있도록 모두가 노력해야 할 것이다.

1. 들어가며

데이터는 의료, 금융, 복지, 경제 등 모든 분야의 핵심 전략으로, 적극적으로 활용되고 있는 추세이다. 정부는 데이터가 새로운 가치 창출 수단을 넘어 경제활동의 중요한 생산요소가 되고 있다고 본다. 4차 산업혁명에서는 데이터가 자산인 만큼 데이터를 얼마나 잘 활용하느냐

나 국가 경쟁력 확보의 중요한 요소라고 하겠다. 데이터 기반 정책 지원을 위해 각국은 데이터 경제를 선도하고자 범국가적 차원의 대책을 마련하고 있다. 이를 위해서는 각 국가의 행정 데이터 및 공공 데이터 연계 자료, 현황을 살펴볼 필요가 있다. 이 글에서는 캐나다의 보건·복지·사회보장 분야에 초점을 맞추어 캐나다 통계청의 사회적 데이터연계 환경(SDLE: Social Data Linkage Environment), 브리티시컬럼비아주의 인구 데이터(PopData: Population Data BC), 온타리오 주의 임상평가과학연구소(ICES: Institute for Clinical Evaluative Sciences) 데이터에 대해 소개하고자 한다.

2. 캐나다 통계청의 사회적 데이터연계 환경(SDLE)¹⁾

가. 사회적 데이터연계 환경(SDLE) 소개

캐나다 통계청은 범캐나다의 사회경제적 통계 연구를 촉진하기 위해 사회적 데이터연계 환경(SDLE) 프로그램을 개발하였다. 사회적 데이터연계 환경(SDLE)은 대규모 통합 데이터베이스(DB)가 아니라 데이터 연계 환경을 의미한다. 이는 기존 캐나다 통계청 조사(survey) 관련성 증가와 행정 데이터 활용 증가의 환경적 변화와 개인정보 보호 및 데이터 보안 표준 준수, 연계 과정과 방법의 표준화된 접근 개발의 기술적 요소를 담고 있다. 사회적 데이터연계 환경(SDLE)은 캐나다인들의 재정, 사회, 경제 관련 정보 격차를 해소할 수 있는 연구들을 가능하게 한다. 데이터 연계 방식이라 새로운 데이터 수집에 따른 비용 및 응답 부담이 없어 사회 데이터 분석에서 중요한 연구자의 욕구(needs)를 충족시킬 수 있다. 또한, 사회적 데이터연계 환경(SDLE)은 처리 시스템과 데이터 연계 결과를 보존하기 때문에 데이터 연계 비용을 절감할 수 있어 효율적이다.

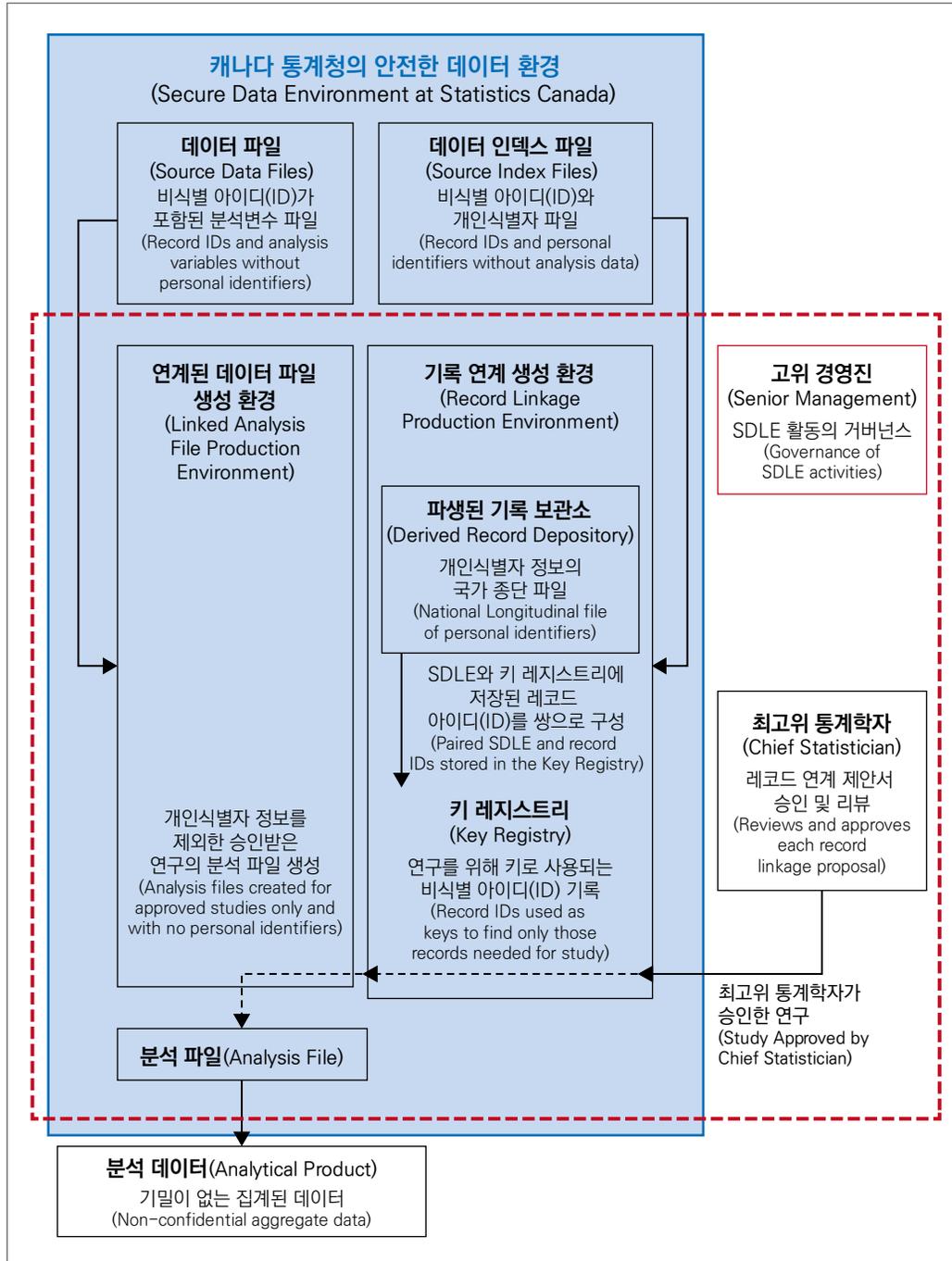
나. 사회적 데이터연계 환경(SDLE) 프로세스

사회적 데이터연계 환경(SDLE)의 핵심은 캐나다 통계청 데이터 파일에 있는 개인의 종단 데이터베이스(DB)인 파생된 기록 보관소(DRD: Derived Record Depository)이며, 기본적인 개인식별자만 포함하고 있다. 파생된 기록 보관소(DRD)는 고유한 개인 목록을 생성하기 위해 개인식별자가

1) 캐나다 통계청의 사회적 데이터연계 환경(SDLE)에 대한 설명은 캐나다 통계청 홈페이지의 사회적 데이터연계 환경(SDLE) 내용 (<https://www.statcan.gc.ca/eng/sdle/overview/>)을 번역 의역하여 요약 정리한 것이다.

있는 인덱스 파일(source index files)을 연결하여 만든다. 파생된 기록 보관소(DRD) 구축에 사용되는 인덱스 파일로는 세금 정보, 아동 세제 혜택 정보, 출생 및 사망 정보, 이민자 정보 등이 있다. 파생된 기록 보관소(DRD)에 저장된 개인식별자는 성, 이름, 생년월일, 성별, 보험 번호, 부모 이름, 결혼 여부, 주소(우편번호 포함), 전화번호, 이민 날짜, 사망 날짜 정보이다. 데이터 연계 방법은 동일한 개체를 찾는 정확매칭(exact matching)과 유사한 개체를 찾는 통계적 매칭(statistical matching)으로 나뉘며, 정확매칭에는 결정 연계(deterministic linkage)(모든 공통 식별변수들의 값이 완전히 같을 경우에 연계하는 방법)와 확률 연계(probabilistic linkage)(각 식별변수 값의 일치와 불일치에 대한 가중치를 계산하여 두 레코드의 연계를 결정하는 방법)가 있다. 사회적 데이터연계 환경(SDLE)의 데이터 연계 방법은 결정 연계를 기본으로 하며 이름, 성별, 생년월일 및 우편번호 정보를 활용하여 확률 연계로 동일인을 추정한다. 활용되는 데이터 파일로는 캐나다 암 등록 자료(Canadian cancer registry), 온타리오 정신건강 보고 시스템(Ontario Mental Health Reporting System), 노동 및 소득 동태 조사(Survey of Labour and Income Dynamics), 전국 인구 건강 조사(National Population Health Survey: Household Component, Longitudinal), 중고등학생 정보 시스템(Postsecondary Student Information System) 등 행정 자료 및 조사 데이터가 포함되어 있다. 캐나다 통계청은 데이터 세트를 안전하게 저장 및 처리하고 승인된 연구를 수행하는 데 필요한 분석 파일을 제공할 책임이 있다. 사회적 데이터연계 환경(SDLE) 연구 프로젝트는 데이터 연계를 포함하기 때문에 데이터 연계에 관한 캐나다 통계청의 지침에 따라 통계청 고위 경영진의 연구별 승인도 필요하다. 실제 분석 환경에서는 개인정보 식별 위험을 줄이고 연계의 효율성과 품질을 향상시키기 위해 별도의 키 정보를 관리하며, 분석 데이터 파일은 개인식별자 없이 제공된다.

그림 1. 사회적 데이터연계 환경(SDLE) 프로세스



자료: Statistics Canada, Social Data Linkage Environment. (SDLE). Retrieved from <https://www.statcan.gc.ca/eng/sdle/overview/> 에서 Figure 1. Social Data Linkage Environment overview diagram 발췌.

3. 캐나다 브리티시컬럼비아주의 인구 데이터(PopData BC)²⁾

가. 인구 데이터 소개

캐나다 브리티시컬럼비아주의 인구 데이터 BC(PopData: Population Data BC)는 건강, 웰빙 등의 연구를 위해 만들어진 개인 단위의 비식별화된 데이터로서 교육, 데이터 연계 등을 지원하는 여러 대학 데이터 및 교육 자원이다. 인구 데이터 BC는 2009년에 만들어졌지만 1980년대에 브리티시컬럼비아주 보건부와 협력으로 데이터 연계 체계 및 정책을 만든 'BC 연계된 건강 데이터 세트(BCLHD: BC Linked Health Data set)'가 전신이다. 행정 데이터의 활용성을 높이고 많은 연구 요구를 충족시킬 수 있는 데이터 인프라를 구축하기 위해 1980년대 후반부터 90년대 중반까지 BC주 연계된 건강 데이터 세트(BCLHD)는 BC주 의료서비스, BC주 암협회(BC Cancer) 및 산재보험 기관인 워크세이프BC(WorkSafeBC)의 데이터를 통합하였고, 1996년부터 연구자들이 데이터를 활용할 수 있도록 하였다. 이 서비스를 시작하기 전에는 공식적인 연계 데이터 제공 절차가 없었다. 이 초기 서비스 모델은 연구에 개인 건강 정보를 사용하는 것에 대한 허용을 포함하는 「BC 정보의 자유 및 개인정보 보호법(FIPPA: BC Freedom of Information and Protection of Privacy Act)」의 선언 시기에도 일치한다. 인구 데이터(PopData)는 정보의 자유 및 개인정보 보호법(FIPPA) 및 BC E-Health(Personal Health Information Access and Protection of Privacy) Act³⁾에 있는 개인정보의 수집, 사용 및 공개에 관한 BC주 입법 요구 사항을 반영하였고 이에 적합한 정책과 절차를 구현하였다.

인구 데이터(PopData)는 의료서비스, 인구 건강 및 여러 데이터를 기반으로 브리티시컬럼비아주 주민에 대한 개인 수준의 익명화된 종단 데이터를 제공한다. 이 데이터 세트는 주로 행정 데이터이지만 조사 데이터 및 지역사회, 환경을 나타내는 데이터도 포함된다. 인구 데이터(PopData)는 개인정보의 일차적인 수집을 하지 않으며, 다른 공공기관이나 개인이 처음에 수집한 개인정보의 2차 이용이나 2차 공개에만 관여한다. 정보의 자유 및 개인정보 보호법(FIPPA) 제33조에 따라 지정된 공공단체는 인구 데이터(PopData)에 개인정보를 공개할 수 있다. 그리고 정보의 자유 및 개인정보 보호법(FIPPA) 35조에 따라, 지정된 공공기관이 수집한 데이

2) 캐나다 브리티시컬럼비아주의 인구 데이터(PopData BC)에 대한 설명은 Ark, Kesselring, Hills와 McGrail(2019), 「Population Data BC: Supporting population data science in British Columbia」 논문 내용을 번역·의역하여 요약 정리한 것이다.

터에 대한 접근 승인 권한을 가진 인구 데이터(PopData) 및 연구자는 연구와 통계 목적을 위해 해당 데이터를 사용할 때 개별적인 동의를 구하지 않아도 된다.

인구 데이터(PopData) 서버의 물리적인 공간은 브리티시컬럼비아대학에 있으며, 연구자 및 데이터 관계자가 포함된 자문위원회의 지원으로 운영된다. 운영 예산은 BC주 압협회(BC Cancer), BC주 보건부(BC MoH) 및 산재보험 기관인 워크세이프BC(WorkSafeBC) 등의 데이터 제공 기관으로부터 지원받고 있다.

나. 연계 데이터

인구 데이터(PopData)의 데이터 연계 방식은 주 및 연방 개인정보 관련 입법 정책을 따르며, 데이터 제공 기관으로부터 주로 연간 데이터를 제공받아 한번에 데이터 연계를 실시한다. 외부 데이터와의 연계를 원하는 연구자는 모든 관련 데이터 제공자들의 승인 후에 직접 식별자를 조합하여 연계할 수 있다. 인구 디렉토리(Population Directory)는 인구 데이터(PopData)가 관리하는 핵심 연결 파일로 이름, 주소, 생년월일, 기타 관련 식별 정보 등 개인정보뿐만 아니라 암호화된 식별자(Linkage ID) 정보도 포함하고 있다. 인구 디렉토리는 의료서비스 계획 등록 및 프리미엄 청구 데이터[Medical Services Plan(MSP) Registration & Premium Billing (R&PB)]를 활용하여 구축되는데, 의료서비스 계획(MSP)은 브리티시컬럼비아주의 건강보험 계획이고 브리티시컬럼비아주의 모든 주민은 개인건강번호(PHN: Personal Health Number)의 유일한 식별번호를 발급받기 때문에 인구 데이터(PopData)는 개인건강번호(PHN), 성, 이름, 생년월일, 성별, 우편번호가 포함된, 거주자의 가장 큰 대표 커버리지 데이터 세트를 제공한다고 볼 수 있다.

인구 데이터(PopData)는 결정 연계와 확률 연계 방법을 이용하여 동일한 개인의 정보를 매칭한다. 데이터 연계를 수행하는 데 사용되는 변수는 데이터 세트마다 다르며 그 변수들은 데이터의 목적에 따라 다양하다. 데이터 연계를 실시할 때는 유일한, 인구학적 식별자 변수만을 사용한다.

표 1. 인구 데이터(PopData)의 데이터 목록

분야	활용 가능한 데이터 세트	데이터 포함 시작 연도	데이터 볼륨 (2019 현재)	업데이트
보건서비스	의료서비스 계획 지불 정보(BC 보건부)	1985	25억 명	연간
	의약보험(BC 보건부)	1986	6억 2천만	PNET의 일부
	퇴원 DB(병원 분리)(BC 보건부)	1985	2270만	연간
	가정 및 지역사회 케어(BC 보건부)	1990	1960만	연간
	정신건강(BC 보건부)	1986	1100만	마지막 업데이트 2011년
	BC 암 등록부(BC 암)	1985	67만 2000명	연간
	환자 중심 측정(BC부), 각종 조사	각종 조사	3만	없음
	출산 전후의 데이터 등록(BC 출산 전후 서비스)	2000	89만 8000	연간
	파마넷(pharmanet)(BC 보건부)	1995	5500만	프로젝트별 기준
	약리처(BC 보건부)	1985	66만	없음
전국 입원관리 보고 시스템(BC 보건부)	2011	720만	프로젝트별 기준	
인구 및 동태 통계	통합 파일(MSP & 프리미엄 청구 데이터)(BC 보건부)	1986	7억 9400만	분기
	인구 출생 통계(BC 인구통계과)	1985	270만	연간
	인구 사망 통계(BC 인구통계과)	1985	87만 3천	연간
	인구 결혼 통계(BC 인구통계과)	1985	70만 6천	연간
인구학 및 생활 과정	인구 생존 통계(Vital Statistics Still Births)(BC 인구통계과)	1985	250만	연간
	BC 제너레이션 프로젝트(BC 암)	2009	2만 9천	연간
	영주권자(이민자, 난민 및 시민권)	1985	100만	연간
직업	소득 구간(캐나다 통계청)	1992, 2002, 2006	59만 9천	간헐적
	WorkSafeBC 청구	1981	1억 8천만	연간
유아	WorkSafeBC 회사 수준 파일			
	초기 발전 도구(Early Development Instrument)(인간 조기 학습 파트너십)	1999	25만 8천	연간
교육	교육부(MED)	1991	260만	연간
환경	캐나다 도시환경 컨소시엄(CANUE)	1983	128만	연간

자료: Ark, T. K., Kesselring, S., Hills, B., & McGrail, K. (2019). Population Data BC: Supporting population data science in British Columbia. International Journal of Population Data Science, 4(2), p. 5 표 발췌.

다. 연구 데이터 제공 프레임워크

인구 데이터(PopData)의 데이터 제공 요건은 인구 데이터(PopData)와 정보의 자유 및 개인정보 보호법(FIPPA)을 준수하는 모든 데이터 관계자 파트너가 개발하고 승인한 문서인 연구 데이터

활용 프레임워크(RDAF: Research Data Access Framework)에 의해 관리된다.

연구 데이터 활용 프레임워크(RDAF)에서 데이터 요청이 승인되기 위해 충족되어야 하는 기준은 다음과 같다.

- 구체적인 연구 질문에 대한 시간 제한적(time-limited) 목적이 있을 것
- 특별히 승인되지 않은 한 행정적 또는 기타 비연구적 목적으로 사용하지 않을 것
- 공공의 이익을 추구할 것
- 상업적 마케팅 목적으로 수행되는 연구와 같은 독점적 연구가 아닐 것
- 과학적인 이점이 있을 것
- 연구윤리위원회의 승인을 받을 것

연구 데이터 활용 프레임워크(RDAF)에 기재된 데이터 요청 승인을 위한 연구과제 요건 외에 연구자가 반드시 충족해야 하는 기준도 있다.

- 연구자는 학생, 교사 또는 BC 대학, BC 법률에서 정의된 대학 또는 지방 기관에 등록, 임명 또는 고용된 다른 개인이어야 함.
- 공공기구(Public Body)의 관련 데이터 관리자에 의해 동의된 다른 개인이어야 함.

위 조건을 만족한 연구자만이 캐나다 안에서 인구 데이터(PopData)의 데이터를 분석할 수 있는 자격이 주어진다.

4. 임상평가과학연구소(ICES) 데이터³⁾

가. 임상평가과학연구소(ICES) 소개

임상평가과학연구소(ICES)는 온타리오주에서 헬스케어 시스템 연구를 위해 1992년에 설립된, 독립적인 비영리 연구소이다. 임상평가과학연구소(ICES)는 온타리오주의 헬스케어 행정 데이터 자료에 대한 접근을 가능하게 하였다. 임상평가과학연구소(ICES) 초기에는 병원 서비스

3) ICES 데이터에 대한 설명은 Schull 외(2019), 「ICES: Data, Discovery, Better Health」 논문 내용을 번역 의역하여 요약 정리한 것이다.

체계, 보건 정책 개발을 목적으로 하였으나, 온타리오주의 7개 물리적 사이트 네트워크를 통해 지역 기반의 건강서비스 분석, 보건 정책, 건강의 사회적 결정요인, 데이터 과학 등의 연구로 확장되었다. 임상평가과학연구소(ICES) 데이터는 캐나다에서 가장 큰 주인 온타리오주에 거주하는 1340만 명의 정보를 포함하며 온타리오주 주민이 주 이외의 곳에서 의료서비스를 받을 때 관련 정보를 제공받는다.

나. 연계 데이터

임상평가과학연구소(ICES) 데이터는 현재 90여 개의 서로 다른 데이터를 보유하고 있다. 온타리오 건강보험 플랜(OHIP: The Ontario Health Insurance Plan)은 온타리오주 주민 대부분에게 제공되는 건강보험 시스템으로 10자리의 고유식별번호가 있는 온타리오 건강보험 플랜(OHIP) 카드가 있다. 이 카드는 의료서비스를 받을 때 활용되며 사망한 경우나 이민자인 경우 폐기된다.

RPDB(Registered Persons Database)에는 온타리오 건강보험 플랜(OHIP)을 발급받은 모든 개인의 정보가 포함되어 있다. 원자료는 온타리오 보건장기요양부(MOHLTC: Ontario Ministry of Health and Long-Term Care)의 특정 데이터 공유 계약에 의해 임상평가과학연구소(ICES)에 제공된다. RPDB에는 개별 건강 카드 번호뿐만 아니라 성, 이름, 성별, 생년월일, 거주지 우편번호, 최근 헬스케어 시스템 자료 등의 정보가 포함되어 있다. RPDB는 임상평가과학연구소(ICES)의 데이터 연계를 위한 연계 키 변수로, 데이터베이스(DB)로 매달 업데이트된다. 온타리오 건강보험 플랜(OHIP) 번호 자체는 수집되거나 제공되지 않으며, 임상평가과학연구소(ICES)에서 사용되는 데이터 연계는 Fellegi-Sunter 방법으로 확률적 연계를 실시한다. 임상평가과학연구소(ICES) 데이터는 90여 개의 데이터에 1407개의 데이터 테이블을 보유하고 있다. 1986년부터 온타리오주 주민의 건강 기록 행정 자료와 임상 데이터, 의료서비스 및 비의료 데이터에 저장된 200억 개의 기록이 있어 코호트 연구가 가능하다.

표 2. 임상평가과학연구소(ICES)의 데이터 목록

데이터 집합 범주	주요 예
보건서비스 (Health Services)	<ul style="list-style-type: none"> • 입원 환자 병원 기록(1988년 이후) • 응급 부서 및 보행 케어(2000년 이후) • 의사 청구서(1991년 이후) • 전자의무기록(2010년 이후) • 홈케어(1994년 이후) • 실험실 테스트 결과(2007년 이후) • 정신건강 및 중독 서비스(2005년 이후) • 당일 수술(1991년 이후) • 장기요양, 복합요양 및 재활서비스(1996년 이후) • 공적 지원 처방 및 모든 마약/통제 물질(1990년 이후)
후천 코호트 및 레지스트리 (Acquired Cohorts and Registries)	<ul style="list-style-type: none"> • 암(1990년 이후) • 비만 수술(2010년 이후) • HIV(1995년 이후) • 치사율(2006년 이후) • 장기기증(1995년 이후)
의료 제공자 (Care Providers)	<ul style="list-style-type: none"> • 의사 인구통계, 전문성 및 기타 교육 정보(1992년 이후) • 의사 진료비 지급 모델(2005년 이후)
코딩 및 지리(Coding and Geography)	<ul style="list-style-type: none"> • 약물, 진단 및 지리 변환 및 참조 표
시설(Facilities)	<ul style="list-style-type: none"> • 보건서비스 기관 정보(1988년 이후)
금융(Financial)	<ul style="list-style-type: none"> • 의료서비스 비용(1992년 이후)
인구 및 인구통계 (Population and Demography)	<ul style="list-style-type: none"> • 중요 통계(사망만 해당)(1990년 이후) • 난민 및 이민 상태(1985년 이후) • 파생 민족 데이터(알고리즘 기반) (1990년 이후) • 인구 조사 프로필(1991년 이후) • 인구 추정 및 추계(1981년 이후)
설문 조사 (Surveyst)	<ul style="list-style-type: none"> • 캐나다 지역사회건강조사(2001년 이후) • 온타리오 건강 연구(2009년 이후) • 건강관리 체험 조사(2006년 이후)
ICES-파생된 코호트 (ICES-Derived Cohorts)	<ul style="list-style-type: none"> • 천식(1993년 이후) • 심장(1991년 이후) • 울혈성 심부전(1991년 이후) • 치매(1996년 이후) • 당뇨병(1991년 이후) • 염증성 장질환(1991년 이후) • 엄마와 아기 쌍(1988년 이후) • 환자-의사-병원 소개 네트워크(2005년 이후) • 류머티즘성 관절염(1993년 이후)
비보건 부문 데이터 (Non-Health Sector Data)	<ul style="list-style-type: none"> • 환경 • 교통편 • 교육 • 정부 장애 지원

자료: Schull, M. J., Azimae, M., Marra, M., Cartagena, R. G., Vermeulen, M. J., Ho, M. M., & Guttmann, A. (2019). ICES: Data, Discovery, Better Health. International Journal of Population Data Science, 4(2), p. 3 표 발췌.

다. 임상평가과학연구소(ICES) 데이터 접근 유형

임상평가과학연구소(ICES)는 현재 두 가지 유형의 데이터와 분석 서비스를 제공하고 있다. 하나는 과학자와 교육생이 수행하는 내부 접근권을 가지는 임상평가과학연구소(ICES) 프로젝트, 다른 하나는 외부 연구자가 수행한 프로젝트에 대한 외부 접근권(임상평가과학연구소(ICES) 데이터 및 분석 서비스) 프로젝트이다. 임상평가과학연구소(ICES) 프로젝트는 연구윤리위원회 승인 없이 보건 시스템의 평가, 계획 및 모니터링을 지원하기 위한 분석 목적으로 데이터에 접근할 수 있다. 모든 임상평가과학연구소(ICES) 연구자는 보안 교육을 받고 기밀 유지 계약에 서명한다. 각 임상평가과학연구소(ICES) 프로젝트는 개인정보보호법(PIIPA: Personal Health Information Protection Act), 임상평가과학연구소(ICES) 정책 및 절차 또는 데이터 파트너와의 계약 의무에 따라 데이터 요청이 허용되는지 확인하기 위해 임상평가과학연구소(ICES) 개인정보 보호 규정 및 법무 사무소가 수행하는 개인정보 영향 평가를 받는다.

개인정보 영향 평가(PIA: Privacy Impact Assessment)는 데이터 보유 근거에 기반하여 프로젝트 목표와 요청한 데이터 세트를 검토하고, 임상평가과학연구소(ICES) 연구자들은 민감성을 제하고 연구 고유 목적으로 활용되는 것으로 결정된 이후에 데이터에 접근할 수 있다. 프로젝트 결과는 외부로 공개되기 전에 임상평가과학연구소(ICES) 책임연구자가 재식별 위험 평가를 실시하여 문서화해야 한다. 재식별 위험 평가는 적은 수의 셀 병합과 재식별할 수 없음을 확인하는 작업으로 이루어진다. 임상평가과학연구소(ICES)에 있는 데이터와 외부 데이터를 연계하여 프로젝트를 수행하는 경우, 개인식별정보를 포함하지 않은 데이터에 대해 임상평가과학연구소(ICES) 개인정보 보호 규정 및 법률 부서 검토 후 데이터 공유 요청 양식을 작성하여 사용 약관을 준수하도록 한다. 개인식별정보가 포함된 데이터에 대해서는 임상평가과학연구소(ICES)와 데이터 관리자 간에 데이터 공유 계약(DSA: Data Sharing Agreement)을 수립해야 한다. 데이터 통합은 임상평가과학연구소(ICES) 직원이 직접 수행한다. 임상평가과학연구소(ICES) 데이터 및 분석 서비스(DAS: Data & Analytic Services)는 외부 연구자가 수행하는 프로젝트를 지원한다. 데이터 및 분석 서비스(DAS)는 임상평가과학연구소(ICES)의 데이터 접근, 분석 지원, 데이터 분석 및 보고서 작성 서비스 등을 제공한다. 연구자는 데이터 및 분석 서비스(DAS)의 서비스 접근을 위해 임상평가과학연구소(ICES) 웹사이트에서 이용할 수 있는 프로젝트 요청서를 제출한다. 데이터 및 분석 서비스(DAS)는 적격성 및 사업 타당성을 평가한 후 사업 목표와 일정 등을 검토하기 위한

협의를 거치고, 연구자는 프로젝트에 대한 연구윤리위원회 승인을 받아야 한다. 데이터 분석 후에 데이터 및 분석 서비스(DAS)는 재식별 위험을 평가하여 결과값을 연구자에게 이메일로 전송한다. 프로젝트가 완료되면 연구자는 데이터 및 분석 서비스(DAS)에 프로젝트 완료를 알려야 하며, 데이터 분석을 위한 접속 권한은 종료된다.

5. 나가며

이 글에서는 캐나다 보건·복지·사회보장 분야의 행정 데이터 및 공공 데이터 연계에 초점을 맞추어 캐나다 통계청의 사회적 데이터연계 환경(SDLE), 브리티시컬럼비아주의 인구 데이터(PopData BC), 온타리오주의 임상평가과학연구소(ICES) 데이터를 살펴보았다. 의미 있는 데이터 분석 및 연구 결과를 위해 행정 데이터 및 공공 데이터의 연계와 활용이 점점 중요해지고 있다. 이를 위해 데이터에 대한 전향적이고 개방적인 접근 방식은 공공서비스 고도화에 필수적이다. 캐나다 정부에서 발표한 2020년 데이터 전략을 보면, 데이터는 정부가 더 나은 결정을 내리고, 더 나은 프로그램을 설계하고, 더 효과적인 서비스를 제공할 수 있도록 하는 힘이다. 캐나다는 연방 공공서비스를 위해 데이터의 전략적인 활용에 대한 로드맵을 제시하고자 4가지 주제로 거버넌스, 데이터 사용 능력 및 기술 향상, 인프라 및 법률 활성화, 가치 있는 자산으로서의 데이터를 꼽았다. 이런 데이터 전략은 캐나다의 디지털 정부 전환을 지원하며 공공서비스가 잘 활용될 수 있도록 한다. 우리나라 역시 건강보험공단, 건강보험심사평가원, 질병관리청, 국립암센터 4개 기관의 정보를 연계하여 연구를 지원하는 보건의료 빅데이터 사업을 진행 중이다. 이러한 행정 데이터 및 공공 데이터 연계는 데이터 기반의 가치 창출 생태계를 조성하기 위한 필수 요소로, 연계 데이터의 활용 및 데이터 제공 프로세스가 체계적으로 이루어질 수 있도록 모두가 노력해야 할 것이다.

참고문헌

- Ark, T. K., Kesselring, S., Hills, B., & McGrail, K. (2019). Population Data BC: Supporting population data science in British Columbia. *International Journal of Population Data Science*, 4(2).
- Government of Canada. (2020). *A Data Strategy Roadmap for the federal public service* (Canada).
- Schull, M. J., Azimae, M., Marra, M., Cartagena, R. G., Vermeulen, M. J., Ho, M. M., & Guttman, A. (2019). ICES: Data, Discovery, Better Health. *International Journal of Population Data Science*, 4(2).
- Statistics Canada, Social Data Linkage Environment. (SDLE). Retrieved from <https://www.statcan.gc.ca/eng/sdle/overview/>