

고용관련 빅데이터 구축 전략 및 활용방안 연구

A Study on the Development Strategy and Utilization of Big Data Related to Employment

최기성
한국고용정보원

Ki-Sung Choi(cksung92@keis.or.kr)

요약

본 논문은 맞춤형 고용서비스 개발을 지원하기 위한 '고용관련 빅데이터 센터(가칭)'설립에 앞서 우리나라의 고용관련 DB 중에서 워크넷, 고용보험, 직업훈련 DB 등과 함께 교육과 복지 DB를 중심으로 DB별 개요, 구축 과정, 주요내용, 활용, 한계점 등을 살펴보았다. 이어서, 고용관련 빅데이터를 구축·운영하고 있는 독일과 영국의 사례를 분석하였다.

이상의 분석을 통해 개인이 생애에 경험하게 되는 다양한 노동력 상태간의 이행에서 발생할 수 있는 사회적 위험을 체계적으로 관리함으로써 이행을 통한 소득확보를 강조하는 이행노동시장론에 입각하여, 고용보험 DB를 중심으로 개인단위에서 교육/훈련/경력/자격, 구직, 고용, 실업, 비경활/복지 등의 DB를 연계하는 체계로 고용관련 빅데이터 구축방안과 활용방안을 제시하였다.

마지막으로 고용관련 DB들에 대한 대대적인 정비, 고용보험 DB를 중심으로 구축할 빅데이터의 대표성 제고, DB의 신뢰도 제고 등 향후 과제들을 제시하였다. 본 논문을 계기로 고용관련 빅데이터 구축 및 활용에 대한 논의가 더욱 활발해지기를 기대한다.

■ 중심어 : | 고용관련 빅데이터 | 행정DB | 고용보험 | 이행노동시장론 | 고용정책 평가 |

Abstract

Prior to the establishment of 'Employment-Related Big Data Center (tentative name)' to support the development of customized employment services. This Paper examines the current status and limitation of employment-related data in Korea. Then, the implications were derived through foreign employment-related big data construction cases.

Through the above analysis, I proposed measures to build and utilize employment-related big data at the individual level, focusing on the Transitional Labour Markets theory that emphasizes the implementation of individual labor force states.

Finally, we presented future challenges such as massive maintenance of employment-related DB, increased representation of big data to be built around employment insurance DB, and increased reliability of DB presented.

■ keyword : | Big Data Related to Employment | Administrative DataBase | Employment Insurance | Transitional Labour Markets | Employment Policy Assessment |

I. 서론

2020년 8월 5일부터 개인정보 보호법 등 소위 데이터 3법 개정안이 시행되면서 4차 산업혁명에 대응하기 위해 기업 및 개인들의 정보 활용이 대폭 증가할 것으로 기대된다. 오래전부터 정부와 국회에서도 고용 및 노동, 교육, 건강, 복지, 창업지원 서비스 등에 관한 정책 수립 및 운영에서 데이터에 기반한(Evidence-based) 의사 결정을 통한 효율적 자원 배분과 이를 통한 정책 효과성 제고를 꾸준히 강조해오고 있다.

데이터에 기반한 정책 수립 및 집행을 위해서는 무엇보다도 신뢰도가 높고 시의성 있는 자료가 필요하다. 현재까지는 이러한 자료들이 대부분 표본조사를 통해 수집되어 왔다. 하지만, 최근에 빈번하게 발생하고 있는 개인정보 유출 및 오·남용 등으로 조사를 통한 자료 수집은 응답률 저하, 조사기간 증가, 측정오차(measurement error) 증가 등으로 인해 조사 비용이 증가할 뿐만 아니라 자료의 신뢰도도 낮아지고 있는 상황이다.

이와 달리, 정부, 공공 및 교육 기관 등에서 행정적으로 보고·집계하는 행정자료는 사건의 발생과 거의 동시에 보고됨에 따라 시의성이 높고, 다른 자료와의 교차검증을 거친 경우에는 자료 신뢰도가 높다. 행정자료가 갖는 이러한 장점에도 불구하고 크게 2가지 이유로 인해 지금까지는 폭넓게 활용되지 못하고 있었다.

첫 번째는 행정자료의 저장, 처리 그리고 분석 관련 기술 부족과 높은 비용 때문이다. 하지만, 최근 들어 빅데이터 관련 기술의 비약적인 발전으로 이러한 문제는 어느 정도 완화되었다. 두 번째는 행정자료가 주제별로 독립적으로 관리됨에 따라 개인/사업체 단위에서 다양한 자료를 연계하여 새로운 가치정보를 생산하기 쉽지 않았기 때문이다. 하지만 이 또한 자료 연계의 필요성, 활용 방안 그리고 법적 근거가 명확한 경우에는 활용이 가능할 것이다.

한편, 성욱준(2016)은 향후 공공부문 빅데이터의 분석 및 활용 확산을 위한 과제로 관련 법제 정비, 행정자료 활용에 대한 인식 개선 그리고 담당 기관 내 추진체계의 정비 등을 제시하였다[1].

Xi et al(2020)는 미국 행정 빅데이터의 현주소, 정

책 연구를 발전시킬 수 있는 빅데이터의 잠재성 그리고 데이터 사용의 장점과 해결해야 할 과제에 대해 논의하고, 현재 미국에서 진행 중인 대규모 행정 데이터 프로젝트를 소개하고 있다[2].

최근 들어 행정자료가 갖는 한계점을 극복하고 이를 활용하여 개발한 대국민 서비스들이 다양한 분야에서 제공되고 있다. 대표적인 사례로는 보건복지부 산하 국민건강보험공단이 2012년부터 시행 중인 개인별 맞춤형 사전예방 건강 서비스 등을 들 수 있다. 국민건강보험공단은 전국민의 출생에서 사망까지 병·의원 이용내역과 건강검진 결과, 질병 정보 등 10년 동안 축적된 약 1조 3천억 건의 행정자료를 토대로 '건강보험 빅데이터 운영센터'를 설치·운영하고 있다. 나아가, 정영철 외(2021)는 지난 2019년 '보건의료 빅데이터 플랫폼(hcdl.mohw.go.kr)'을 개통하여 보건의료 빅데이터 신청, 심의 결과 확인, 연계, 비식별화 기능 등을 시범적으로 수행하고 있으며, 보건의료 빅데이터 플랫폼 구축 운영의 법적 근거를 마련하기 위해 '보건의료기술진흥법' 일부 개정을 추진하고 있다고 소개하였다[3].

이외에도 박과영(2018)은 부동산공시제도에서 이용되는 빅데이터 구조 및 현황을 설명하고, 해당 데이터가 지닌 문제점과 개선방안을 제시함으로써 부동산공시제도의 발전방안을 모색하였다[4].

고용 및 노동 분야의 경우 주로 한국고용정보원이 관리하는 행정자료를 주제간 연계없이 독립적으로 분석한 연구들은 다수 존재하나, 고용관련 주제간 연계 분석을 실시한 경우는 매우 제한적이며 이마저도 대부분 2가지 주제만을 연계하여 분석하였다. 먼저, 윤정혜(2019)는 고용보험 자료만을 사용하여 가입자의 이력을 연계하여 근로자의 산업 및 지역 간 노동이동 현황을 분석하였지만, 근로자의 학력 정보의 부재로 학력수준에 따른 세부적인 분석은 수행하지 못했다[5]. 한편, 이재성 외(2008)는 훈련자료와 고용보험 자료를 연계하여 분석하였다. 연구자들은 재정지원 일자리사업에 포함된 구직자훈련 참여자의 동종분야 취업 현황을 분석하였지만, 앞서와 같이 학력수준과 전공 등에 따른 세부적인 분석은 자료의 부재로 수행하지 못했다[6].

이처럼 고용분야 행정자료를 활용한 주제간 연계분석이 미진한 이유는 앞서 밝혔듯이 고용분야 행정자료

들이 주제별(고용보험, Work-Net, HRD-Net, Q-Net 등)로 관리하는 기관이 상이하고, 고용관련 행정자료(교육, 복지, 건강 등)와의 연계도 쉽지 않기 때문이다.

그럼에도 불구하고 우리사회가 겪고 있는 고실업·저성장·저출산·고령화에 효율적으로 대응하기 위해서는 '고용관련 빅데이터 센터(가칭)'를 설립하여 고용분야 행정자료의 주제간 연계분석(고용보험-Worknet, 고용보험-HRD-Net 등)뿐 아니라 고용-복지, 고용-교육 등과 같은 관련 주제들과의 연계분석을 통해 고용관련 개인/사업체 맞춤형 서비스(원스톱 서비스) 개발 및 제공이 반드시 필요하다.

이러한 배경 하에 본 논문은 '고용관련 빅데이터 센터' 설립을 위해 현재 우리나라가 보유하고 있는 고용관련 DB에 대한 현황과 활용 상의 한계점(II)을 살펴보고, 외국의 고용관련 빅데이터 구축 사례를 통해 시사점을 도출(III)한 후 우리나라의 고용관련 빅데이터 구축 및 활용 방안을 제시(IV)하고, 마지막으로 고용관련 빅데이터 구축 및 운영을 위해 해결되어야 할 과제(V)를 제시하였다.

본 논문은 다양하게 존재하는 고용관련 행정 DB를 연계하여 빅데이터로 구축하고, 이를 고용정책 수립 및 평가에 어떻게 활용할 수 있는가를 처음으로 제시했다는 측면에서 의미를 지니며, 이를 계기로 관련 논의가 더욱 활발해지기를 기대한다.

II. 고용관련 DB 현황 및 한계점

1. 고용관련 DB 현황

고용관련 DB는 내용에 따라 고용 영역 DB와 관련 영역 DB로 구분할 수 있다. 먼저, 고용 영역 DB에는 고용과 직접적으로 관련된 내용으로 취업알선(Work-Net), 고용보험, 직업훈련(HRD-Net), 일모아 DB, 자격(Q-Net), 산업재해(근로복지공단), 장애인고용(한국장애인고용공단), 외국인고용관리 전산망 등이 포함된다. 이에 비해 관련 영역 DB에는 고용과 간접적으로 연결되는 내용으로 교육(한국교육개발원 교육통계서비스, 한국교육학술정보원), 복지(사회보장정보원), 건강(건강보험공단 건강보험 빅데이터), 연금(국민연금,

특수연금) 등이 포함된다. 이외에도 통계청, 중앙정부, 노동연구원 등과 같은 기관에서 정기적 또는 부정기적으로 조사를 통해 수집하는 다양한 고용관련 DB들이 존재한다.

본 논문에서는 고용 영역 DB 중에서 워크넷(Work-Net), 고용보험, 직업훈련(HRD-Net), 외국인고용관리 전산망, 일모아 DB 등과 함께 관련 영역 DB 중에서는 교육과 복지 DB를 중심으로 DB별 개요, 구축과정, 주요내용, 활용, 한계점 등을 살펴보겠다.

1.1 워크넷(Work-Net) DB

워크넷은 1998년부터 인터넷 기반으로 개인을 대상으로는 근무지역별, 직종별, 기업형태 등 다양한 일자리 정보를 비롯해 온라인 구직신청, 입사지원관리, 맞춤형 채용정보, 방문상담신청, 이메일/SMS 알림서비스 등을 제공하고, 기업에는 다양한 인재정보 검색과 고용센터를 통하여 온라인 e-채용마당(채용대행) 서비스 제공하고 있다. 또한, 취업지원 기관(고용센터, 지자체)의 직업상담원 등에게는 취업알선, 구인구직 인증, 취업지원프로그램 운영 및 상담 등의 업무를 지원하고, 취업성공패키지, 중소기업 청년인턴, 청년 강소기업체협, 장년 취업인턴제 등 고용안정을 위한 사업 수행을 위한 시스템을 구축하여 제공하고 있다. 나아가, 2011년부터는 사람인과 같은 민간취업포털과 지자체의 구인 정보를 한번에 검색하는 일자리 정보 통합서비스 및 모바일 웹 서비스 구축·제공하고 있다.

워크넷 DB의 구축과정을 살펴보면 먼저, 취업알선 DB는 구인자와 구직자가 워크넷(www.work.go.kr)에 접속하여 회원가입 후 요구되는 정보를 입력하거나 취업알선 담당자가 수행하는 취업알선 과정을 통해 생성된다. 이렇게 생성된 구직자 DB는 1년에 2번 정비되는데, 구직 등록 후 3년이 경과하도록 구직활동을 전혀 하지 않은 구직자의 개인정보를 DB에서 삭제하게 된다.

취업성공패키지 DB는 취업성공패키지에 참여하기 위해 제출한 서류(주민등록등본, 건강보험내역서 등), 취업계획서 상의 자료를 고용센터 상담사가 고용안정 정보망에 입력하거나, 고용관련 DB(고용보험, HRD-Net 등)와의 연계를 통해 생성된다. 아래 [표 1]은 워크넷을 통해 구축되어 있는 대표적인 DB 중 취업

알선 DW(Data Warehouse)의 주요 내용을 정리한 것이다.

표 1. 워크넷 DB의 DW별 주요 내용

DW	주요 TABLE	TABLE 내 주요 정보
취업 알선 DW	구직 현황	구직신청 시 입력된 구직 관련 데이터 -DW고객번호(주민번호 대체번호로 다른 DB 연계) -구직 사정, 구직자 인적정보(성, 연령, 학력, 전공) -희망하는 직종(KECO4자리), 지역, 임금 -구직자 사회적 특성(장애, 새터민, 여성가장 등)
	구인 현황	구인신청 시 입력된 구인 관련 데이터 -사업자등록번호(고용보험,HRD-Net 연계 시 활용) -구인 사정, 사업체 정보(산업, 지역, 규모 등) -구인 요건(학력, 연령, 직종(KECO4자리), 제시임금)
	알선 현황	알선 시 입력된 알선 관련 데이터 -DW고객번호 및 구직자 정보 -사업자등록번호 및 구인기업 정보 -구인/구직 직종, 희망/제시 임금, 취업여부 및 일자
	본인취업	구직신청에서 취업 데이터(취업 사업장 정보 없음)
	알선취업	알선 취업(취업 사업장 정보 있음)

*출처: 한국고용정보원 내부자료(2020년 연구자료 DB 설명자료) 중 일부 발췌

이와 같이 구축된 워크넷 DB는 주제별로 주기에 맞춰 집계수준에서의 노동시장정보를 생산하여 고용정보 통합분석시스템(EIS: <https://eis.work.go.kr>)을 통해 제공해오고 있으나, 다음과 같은 한계점을 지닌다.

첫 번째는 다른 취업알선 사이트와의 제한적 정보공유로 인해, 워크넷 취업알선 DB만으로는 우리나라 전체의 취업알선 실태를 분석할 수는 없다. 즉, 현재 워크넷에서 구인 정보는 민간, 지자체 등과의 정보공유가 이뤄지고 있지만, 구직정보의 경우 정보공유에 대한 구직자 동의와 공유조건이 만족되는 일부만이 공유되고 있기 때문이다.

두 번째는 워크넷 DB의 상당 부분이 구직자, 구인기업 그리고 취업성공패키지 참여자가 직접 입력한 자료로 DB의 신뢰도가 확인되지 않았다는 점이다. 예컨대, 구직자의 인구사회학적 특성에 해당하는 학력, 전공, 장애여부, 여성가장 여부 등과 같은 자료에 대한 별도의 검증이 이뤄지지 않고 있다. 세 번째는 일자리 매칭을 촉진할 수 있는 중요 정보의 부재이다. 구체적으로 구직현황의 경우 의중임금(reservation wage), 희망산업, 희망 고용형태, 자격증 등의 정보가 추가적으로 요구되며, 구인현황의 경우에는 사업체 창업년도, 작년 매출액, 근로자 평균 임금, 4대 사회보험 가입률 등의 정보가 추가될 필요가 있다.

1.2 고용보험 DB

고용정책기본법에 따라 운영되는 고용보험 전산망은 기업에는 근로자의 피보험자격의 취득/상실/변경/전근/이직확인서 등 각종 신고 업무와 지원금/고용창출 참여 신청 서비스를 제공하고 있다. 개인 가입자에게는 육아휴직급여/실업인정/심사/재심사의 온라인 신청 및 고용보험 가입이력조회 등의 서비스를 제공하고 있다. 아울러, 일반국민들에게는 고용보험제도 설명, 실업급여 모의계산, 고용보험관련 서식자료 등을 제공하고 있다. 내부 사용자 서비스로는 고용보험 행정업무(피보험자 관리, 실업급여지급 등)를 처리하기 위한 내부 사용자용(고용센터 등) 서비스를 제공하고 있다.

고용보험 전산망은 1995년에 시스템을 구축하여 현재까지 운영 중이며 2012년에는 자영업자의 고용보험 임의 가입이 도입됨에 따라 이에 대한 서비스까지 확대 시행해 오고 있다.

고용보험 피보험자 DB는 상용직의 경우 사업체에서 근로자를 채용하거나 이직이 발생하게 되면 고용센터 직접방문 또는 우편이나 팩스, 인터넷(EDI, 4대 사회보험 정보연계센터)으로 고용보험 취득 또는 상실에 관한 정보를 입력하는 과정을 통해 생성된다. 한편, 일용직의 경우는 근로내용 확인신고서를 고용센터에 제출하면 이를 고용센터 담당자가 EDI에 입력하는 과정을 통해 생성된다. 한편, 고용보험 사업장 DB는 신규 사업체가 근로복지공단에 사업장 성립을 신청할 때 제출한 자료 중 고용 관련 사업장 정보만을 제공 받아 구축하고 있다.

계속해서, 실업급여 DB는 고용보험 자격 상실자가 고용센터에 방문하여 제출한 실업급여 수급자격 인정 신청서 내용과 실업급여 지급 자료를 고용센터 담당자가 고용보험전산망에 입력하는 과정을 통해 생성된다. 아래 [표 2]는 고용보험 DB 중 피보험자 등의 주요 DW(Data Warehouse)의 핵심 내용을 정리한 것이다.

표 2. 고용보험 DB의 DW별 주요 내용

DW	주요 TABLE	TABLE 내 주요 정보
피보험자	피보험자 종합통계	상용·일용·자영업 피보험자와 사업장 자료 통합 -DW 고객번호 및 사업장관리번호 -자격취득일자, 성, 연령, 학력(2012년부터 미신고) -직종(KECO 3자리), 산업(KSIC 5자리), 상시근로자수, -초임, 주당 근로시간, 고용형태(상용, 일용, 자영업자)

	상실자 종합통계	상용직 피보험자격 상실자 정보 -DW 고객번호 및 사업장관리번호 -자격 취득 및 상실일자, 상실사유 -취득 시 직종(KECO 3자리)과 산업(KSIC 5자리)
실업 급여	수급 신청자	실업급여 수급 신청자 정보 -DW 고객번호 및 사업장관리번호 -기본특성(성, 연령) -이직사유 및 거주지역 -급여기초임금일액 -피보험기간코드, 소정급여일수
	수급 자격자	실업급여 수급자 정보 -DW 고객번호 및 사업장관리번호 -수급자격신청일자 및 승인일자 -이직사유 및 거주지역 -급여기초임금일액 등
	급여 지급	실업급여 지급정보 -DW 고객번호 및 사업장관리번호 -실업급여지급금액 -급여기초임금일액 -소정급여일수

*출처: 한국고용정보원 내부자료(2020년 연구자료 DB 설명자료) 중 일부 발췌

위의 과정으로 구축된 고용보험 DB의 경우 주제별로 집계수준에서의 정보 제공(EIS)뿐만이 아니라 고용보험 DB 내 주요 table 간 연계성을 통해 추가적인 가치정보를 생산할 수 있도록 새롭게 구축된 DB도 고용부의 승인 시 연구자를 위해 제공하고 있는데, 대표적으로 고용보험 피보험자 이력 DB를 들 수 있다.

그럼에도 불구하고 고용보험 DB가 갖는 첫 번째 한계점은 취업자(임금근로자와 비임금근로자 포함) 중 고용보험 미가입 비율이 높아, 고용보험 DB가 우리나라 노동시장 전체를 대표한다고 보기 어렵다는 점이다. 통계청과 한국고용정보원 자료에 따르면, 2020년 3월 기준으로 취업자 수는 2,670만여명이고 고용보험 가입자 수는 1,376만여명으로 전체 취업자 중 약 52%만이 고용보험에 가입하고 있다. 다만, 2020년 5월 정부가 고용안전을 강화하기 위해 전국민 고용보험 가입을 추진하겠다고 밝혔다. 이는 현재 고용보험 임의 가입대상인 영세 자영업자와 가입 대상에서 제외되어 있는 특수 고용형태종사자, 프리랜서, 플랫폼근로자 등을 고용보험 의무가입대상으로 변경하는 것으로 정부 목표대로 2025년까지 취업자 2,100만명이 고용보험에 가입된다면 대표성 문제는 어느 정도 완화될 수 있을 것이다.

두 번째 한계점은 대부분의 자료가 사업체(또는 사무대행 업체) 담당자가 자기기입식(self-reporting)으로 입력한 자료를 별도의 검증과정 없이 집적함으로써 입력 오차의 발생 가능성을 배제할 수 없다는 점이다. 세 번째 한계점으로는 중요 정보의 부재와 최신화

(update)가 이뤄지지 않는다는 점이다. 정보 부재의 대표적인 경우로는 2012년부터 고용보험 취득신고서에서 제외된 자격취득자의 학력과 전공이 해당된다. 취업자의 학력과 전공은 노동시장 진입, 성과, 퇴장 등과 밀접하게 관련되어 있다는 점을 고려하면, 현재의 고용보험 DB로는 노동시장을 세부적으로 분석하는 데는 한계가 있다. 정보의 최신화가 이뤄지지 않는 대표적인 정보로는 취업자의 직업 정보이다. 현재 고용보험 DB의 직업정보는 취업자가 해당 사업장에 취업했을 때의 직업으로 직장 내에서 승급과 전환배치 등에 따른 직업 변동에 관한 정보는 최신화가 이뤄지지 않고 있다.

1.3 HRD-Net DB

2003년부터 운영 중인 HRD-Net을 통해 생성되는 직업훈련 관련 DB는 국민들에게 훈련대상별(구직자, 재직근로자, 사업주 등) 훈련과정·취업률·평가등급·만족도 정보를 제공할 뿐만이 아니라 평생학습기반 무료학습 콘텐츠 및 직업능력개발 관련 정보를 제공하고 있다.

HRD-Net DB의 구축과정은 내일배움카드제(실업자류 훈련 포함) 훈련 DB는 구직자가 워크넷에 구직등록 후 고용센터에 방문하여 고용센터 담당자와 훈련 상담을 하거나 훈련에 참여하는 과정을 통해 생성된다.

사업주 훈련 DB는 사업주가 HRD-Net에 훈련과정 인증신청, 훈련실시, 훈련수료자 등을 보고하면, 산업인력공단 담당자가 이를 검증·승인하는 과정으로 생성된다. 훈련 비용과 각종 수당과 관련해서 살펴보면, 실업자 훈련비 DB는 훈련기관이 훈련 종료 후 고용센터에 훈련비용을 신청하면, 고용센터 담당자가 이를 검증 후 훈련비를 지급하는 과정으로 생성된다. 사업주 훈련비 DB의 경우에는 사업주가 훈련 종료 후 3년 이내에 산업인력공단에 훈련비를 신청하는 과정으로 생성된다. 훈련기관 DB는 훈련기관 및 과정 인증을 위해 훈련기관들이 직업능력심사평가원에 제출한 서류를 심사평가원에서 입력하는 과정을 통해 생성된다. 아래 [표 3]은 HRD-Net DB 중 실업자 훈련 DW와 내일배움카드제 DW에 포함된 주요 정보가 제시되어 있다.

표 3. HRD-Net DB의 DW별 주요 내용

DW	주요 TABLE	TABLE 내 주요 정보
실업자 훈련	실업자 훈련실시	실업자, 국가기전전략, 자치단체훈련 참여자 및 과정 정보 -훈련시작일자 및 훈련과정 ID -훈련생 ID 및 DW 고객번호 -훈련직종코드 및 훈련유형코드 -훈련생 인구사회학적 특성(성, 연령, 지역, 자녀) -실업급여 지급여부 및 각종 훈련수당 지급여부 -취업여부 및 일자 -훈련생 정원 및 참여자수, 중도탈락자, 자격취득자 -조기취업자수, 수료 후 취업자수
내일배움 카드제	내일배움 카드제 마스터	내일배움카드제 참여자 및 과정 정보 -훈련과정 ID 및 회차, 훈련기관번호 및 지역 -훈련의 KECO 코드(3자리) -훈련과정 정원수, 정부지원금액, 자부담금액 -총훈련 시간, 일수, 월수, 1일평균훈련시간수 -1인당훈련비 및 시간당 단가
	내일배움 카드제 훈련결과	내일배움카드제 참여자 훈련결과 및 자격취득정보 -훈련과정 ID 및 회차 -훈련생 ID 및 인구사회학적 특성(성, 연령, 학력) -취업여부, KECO코드(3자리) 및 취업일자 -훈련기관번호 및 지역

*출처: 한국고용정보원 내부자료(2020년 연구자료 DB 설명자료) 중 일부 발췌

이러한 HRD-Net DB를 통해 훈련별로 실시 및 수료 인원, 중도탈락 인원, 훈련 중 취업(조기취업) 및 수료 후 취업 인원, 훈련비 지원 금액 및 수당 등의 정보를 집계수준에서 생산하여 제공하고 있다.

하지만 HRD-Net DB를 활용하는데는 다음과 같은 한계점을 지닌다. 먼저, 현재의 HRD-Net DB로는 훈련의 취업 효과를 정확하게 분석하기 어렵다는 점이다. 현재 실업자 훈련의 취업효과는 참여한 훈련과정과 취업한 직무와의 관련성을 고려하지 않고, 단지 훈련 종료 후 6개월 이내 취업여부만으로 분석하고 있다. 이에 따라 어떤 구직자가 금형관련 훈련과정을 종료 후 버스 운전원으로 취업하여도 훈련 후 취업자에 포함된다. 이는 훈련의 관리단위인 훈련과정과 직업(한국고용직업 분류)이 1대 1일로 연계되지 않기 때문이다.

두 번째는 훈련기관에서 직업훈련을 담당하는 훈련 교·강사에 대한 정보가 부족하다는 점이다. 직업훈련의 질적 향상을 위해서는 훈련 교·강사의 역할이 매우 중요함에도 불구하고 현재의 DB에는 이들의 성, 출생년월, 학력수준, 훈련교사 자격여부 및 분야 정도만이 존재한다. 향후 이들의 임금수준, 직업력, 향상교육 참여, 자격증과 같은 정보 등이 추가될 필요가 있다.

1.4 외국인 관리 DB

외국인 관리 DB는 2004년 외국인근로자 고용허가제 시행에 따라 시스템을 구축한 후 사업주를 대상으로 고용허가제 소개, 관련 서식 신청 및 처리현황 조회 서비스 등을 제공하고, 외국인 근로자들에게는 15개국 언어로 고용허가제 소개 및 한국어능력시험 결과조회 등의 서비스를 제공하고 있다.

외국인 관리 DB의 구축과정으로 사업장 및 고용허가 신청 DB는 외국인 고용을 희망하는 사업장에서 고용센터에 방문하여 외국인 고용허가발급 신청서를 제출하면, 고용센터담당자가 적격 여부를 심사 후 외국인고용관리시스템(EPS)에 입력하는 과정으로 생성된다.

외국인근로자와 구직신청 DB는 외국인 근로자(E-9 비자)가 자국 송출기관에서 운영하는 산업인력공단이 제공한 시스템에 개인정보와 구직정보를 입력하면, 산업인력공단에서 해당 자료에 대한 검증을 거쳐 생성된다. 또, 입국한 외국인 근로자가 이직을 위한 구직신청서를 고용센터에 제출하면 고용센터 담당자가 입력하는 과정으로 생성된다. 끝으로, 고용허가서발급 DB는 고용센터 담당자가 외국인 근로자와 사업장에 대한 취업알선 과정과 결과를 통해 생성되며, 고용허가서발급 변동내역 DB는 외국인근로자의 고용허가가 변경됨에 따라 생성된다.

아래 [표 4]는 외국인관리 DB 중 사업장 및 고용허가 신청 DW와 외국인 근로자 및 구인신청 DW에 포함되어 있는 주요 내용이 제시되어 있다.

표 4. 외국인 관리 DB의 DW별 주요 내용

DW	주요 TABLE	TABLE 내 주요 내용
사업장 및 고용허가 신청	사업장	고용허가제 참여 또는 예정 사업장 기본정보 -생성일시 및 사업장관리번호 -산업 및 종사자수 -상시근로자수 및 외국인 근로자 수
	고용허가 신청	외국인근로자 구인 위해 작성하는 외국인고용신청서 -생성일시 및 사업장관리번호 -구인외국인코드(일반, 특례외국인, 무관) -고용허가승인여부 -구인 인원수, 국가, 성/연령별 구인 수 -구인 직종 및 임금
외국인 근로자 및 구직 신청	외국인 근로자	고용허가제도로 취업 희망하는 외국인 정보 -외국인근로자접수번호 -외국인특성(성, 연령, 최종학력, 출신국가) -채류자격내용 및 출입국날짜
	구직신청	외국인근로자의 구직신청 정보 -외국인근로자접수번호 -구직신청날짜 및 사유 -구직 산업 및 신청 승인 여부

*출처: 한국고용정보원 내부자료(2020년 연구자료 DB 설명자료) 중 일부 발췌

외국인 관리 DB를 활용하여 고용허가서 발급 건수, 구인 구직 건수, 외국인 고용 사업체 수 등 주제별로 집계수준에서 정보를 생산하여 제공하고 있으나, 다음과 같은 한계점을 지닌다.

외국인 관리 DB의 첫 번째 한계점은 특례고용허가제(H-2 비자)로 입국한 외국국적동포에 대한 체계적 관리가 이뤄지지 못하고 있다는 점이다. 이는 이들의 취업활동이 정부에 의해 승인되는 것이 아니라 자발적 신고에 의존하고 있기 때문이다. 두 번째는 외국인근로자의 여권 및 비자 변경 정보가 반영되지 못한다는 점이다. 예컨대, 고용허가제로 입국한 외국인근로자가 내국인과의 혼인을 통해 비자가 H-4로 변경되어도 현재의 DB에는 이러한 정보를 포함하고 있지 않다. 이로 인해 현재의 DB만으로는 불법체류외국인 근로자¹를 정확하게 식별할 수 없다.

세 번째 한계점은 외국인근로자에 대한 정보 부족이다. 대표적으로 외국인근로자의 자녀가 자국에 존재하는가를 들 수 있는데, 이는 자녀유무와 불법체류 가능성이 강하게 상관되어 있다면 외국인 근로자 선발기준으로 이를 활용하여 불법체류 문제를 완화할 수 있기 때문이다.

1.5 일모아 DB

일모아 DB는 2009년부터 일모아시스템(www.ilmoa.go.kr)을 통해 구축·운영되고 있다. 일모아 시스템은 국민들에게 정부의 일자리 사업을 소개하고 일자리 정보를 제공할 뿐만 아니라 복지사업(복지로)에 대한 정보도 함께 제공하고 있다. 고용센터 등 내부 사용자들은 인터넷 서비스를 활용하여 사회보장정보시스템과의 정보 연계를 통한 참여자 선발, 출결 정보 관리 및 급여 자동 계산 등 일자리사업 업무에 활용하고 있다.

김병섭 외(2012)는 일모아 시스템DB 문제점으로 직종분류 불명확성, 일자리 관련 정보의 업데이트 제약, 취약계층 식별 및 관리 정보 부재 등을 지적하고 있다 [7].

1 불법체류 외국인 근로자를 출국예정일자까지 출국하지 않은 외국인으로 정의한다면, 현재와 같이 비자변경 정보를 고려하지 않으면 혼인비자 취득으로 불법체류가 아닌 외국인도 불법체류자로 식별되는 문제가 발생한다.

1.6 교육 DB

현재까지 우리나라의 경우 개인의 교육 이력 및 교육 기관 DB는 학교급에 따라 이원화되어 관리되고 있다. 즉, 유치원부터 고등학교까지 개인의 교육 이력 및 교육기관 DB는 전국의 17개 시도교육청에서 관리²하고 있다. 한편, 대학 이상에서의 교육 이력 및 교육기관 DB는 주제에 따라 관리기관이 달라진다. 대학생 개인에 대한 교육 이력(입학 및 졸업 시점, 대학명, 전공 등)은 개별 대학에서 교육부에 제출한 자료를 한국교육개발원에서 관리하고 있다. 대학알리미(www.academyinfo.go.kr)를 통해 제공되는 대학정보는 개별 대학이 공식 주체지만, 한국대학교육협의회에서 총괄관리하고 있다.

이러한 교육 DB를 분석하여 정보를 제공하는 대표적인 기관은 한국교육개발원(교육통계서비스)과 한국교육학술정보원을 들 수 있다. 이들 기관들은 개인 및 학교에 대한 다양한 주제에 대해 집계수준뿐만 아니라 개인단위 연계정보를 제공해오고 있다.

대표적으로 교육에 대한 이해관계자들의 합리적 선택을 돕기 위해 2001년부터 교육부와 한국교육개발원이 주관하여 대졸자 취업실태 조사에서 졸업생 취업률을 파악하여 공표하고 있다. 해당 정보는 개인들이 대학교육 투자에 대한 편익을 판단하는 정보로 활용하고, 정부는 부실대학 선정 및 대학역량강화사업의 선정 기준으로 활용하고 있다. 하지만, 해당 조사는 건강보험 DB, 국세청 DB, 고용보험 DB를 통해 취업여부만을 확인·검증하고 취업유지율도 6개월로 한정함으로써 졸업생의 취업 질(임금, 계약직 여부, 기업규모 등)을 판단할 수 없다. 최기성 외(2014)는 이로 인해 학생 및 학부모들이 해당 자료만으로 대학교육에 따른 편익을 충분히 파악하기 어렵고, 대학들로부터는 평가기준으로 취업률만을 고려함으로써 평가기준과 관련된 공정성이 부족하다는 문제로부터 자유롭지 못한 상황이다[8].

1.7 복지 DB

사회보장정보시스템(행복e음)은 각종 사회복지 급여

2 한국교육학술정보원은 우리나라 초·중등 교육정보를 구축·운영하고 있다. 해당 기관에서 운영하는 나이스(NEIS)에서 개인의 교육 정보를 확인할 수 있으나, 이는 전국의 17개 시도교육청과 온라인으로 연계하여 제공하는 것으로 실질적인 관리기관은 교육청이다.

및 서비스 지원 대상자의 자격과 이력에 관한 정보를 통합 관리하고, 지자체의 복지업무 처리를 지원하기 위해 개인별 가구별 DB를 구축한 정보시스템이다. 이러한 사회보장정보시스템 DB는 국세청, 건강보험공단, 연금공단 등과 같이 원천정보 보유기관에서 개인의 인적정보, 소득재산 정보, 복지서비스 수급이력 정보 등을 수집·연계하여 생성된다. 이렇게 구축된 DB는 시군구에서 복지수급자 선정 및 관리 등에 활용되고, 한국전력에서는 전기료감면 대상 선정 등과 같이 폭넓게 활용되고 있다.

2. 고용관련 DB 활용의 한계점

현재 한국고용정보원은 원내·외 고용관련 정보를 통합한 고용정보통합분석시스템(Employment Information System)을 활용하여 고용 및 노동과 관련된 각종 통계를 제공함으로써 정책 결정자의 의사결정을 지원하며, 내부연구자 및 현업 사용자가 필요한 노동시장 관련 정보를 직접 추출할 수 있도록 지원하고 있다.

또한, 고용보험, 워크넷, 직업훈련 등 3대 DB에 대한 분석 보고서(월간, 분기, 년보)를 매년 주기적으로 발간하여 노동시장에 관한 정보를 제공해오고 있다. 아울러, 고용부와 고용관련 연구기관(노동연구원, 직업능력개발원 등)에는 부정기적으로 고용관련 DB와 심층 분석 결과를 제공하여 정책 수립에 기초자료로 활용되고 있다.

이와 같이 고용관련 DB에 대한 폭넓은 활용에도 불구하고, 현재의 고용관련 DB에는 다음과 같은 5가지 한계점이 존재한다. 먼저, 노동시장 관련 가치정보를 생산하는데 필요한 DB들의 부재이다. 대표적으로 고용보험 DB에서 피보험자 학력 수준 및 전공 DB의 부재를 들 수 있다. 두 번째로는 시간변동(time-varying) DB들에 대한 최신화가 이뤄지지 않는다는 점으로 고용보험 DB의 직업 정보가 최신화되지 않는다는 것이 대표적이다.

세 번째는 생성 DB에 대한 사전(또는 사후) 검증절차 미흡으로 자료 신뢰도를 확신하기 어렵다는 점이다. 예를 들면, 워크넷에 구인기업 및 구직자가 직접 입력한 DB가 여기에 해당된다. 네 번째는 DB 생성 및 구축 업무 담당자와 이를 활용하는 담당자간의 업무 연계의 부족으로 DB에 대한 접근과 사용에 제약

(organizational silos effect)이 따른다는 점이다. 마지막 다섯 번째는 고용관련 DB간 연계를 통한 가치정보의 생산이 어렵다는 점이다. 현재 고용 영역 DB들은 개인 또는 사업체 단위에서 개체를 고유하게 식별할 수 있는 고유식별정보(개인=주민등록번호 또는 DW고객번호, 사업체=사업장관리번호 또는 사업자등록번호)를 모든 DB들이 포함하고 있어, 연계를 위한 수단은 확보된 상태이다. 하지만, 고용영역 DB와 관련영역 DB간의 연계를 위해서는 개인정보보호법 시행령에 명시된 고유식별정보의 처리가 가능한가에 관한 법률적 검토와 해당 DB를 구축·운영하는 부서 또는 기관의 협조가 필요하다. 이는 많은 시간과 예산을 고용관련 DB간 연계 작업에 투입하여도 연계를 통한 가치정보의 생산을 보장받기 어렵다는 것이다.

III. 고용관련 국외 빅데이터 활용사례 및 시사점

1. 독일³

독일의 고용관련 빅데이터인 IEB(integrated Employment Biographies)는 고지 의무(Notification Duty)에 의해 생성되는 행정 기록을 기반으로 구축되었으며 연방고용청(BA) 산하 고용연구소(IAB)가 DB를 구축·운영하고 있다. 2020년 기준으로 IAB에는 사회과학자 및 DB 전문가 등 50-60명이 참여하고 있으며, IEB 데이터베이스의 첫 번째 버전은 2004년에 생성되었으며, 지난 몇 년간 경제학, 사회과학, 통계학뿐 아니라 정책평가에서도 실증 연구의 중요한 자원이 되고 있다⁴.

IEB 데이터베이스는 독일 연방고용청의 운영과정에서 발생한 DB, 사회보장기관들에게 고지하는 과정(notification process)에서 발생한 DB 그리고 사업

3 해당 내용은 한국고용정보원 국제세미나에서 독일의 Joerg Heining이 발표한 내용을 재구성한 것이다[9].

4 본 데이터베이스를 이용하여 진행된 학술연구가 American Economic Review(AER), Quarterly Journal of Economics(QJE), Journal of Labor Economics(JLE) 등 해외 유명 학술지에 게재되었고 하르츠 개혁과 관련된 독일 노동정책의 효과성 및 효율성을 평가하는데 사용되는 등 학술적 및 정책개발과 관련해서 광범위하게 이용되고 있다.

체 대상 조사를 통해 확보한 DB를 포함하고 있다. 이러한 과정에서 추출된 모든 DB는 연방고용청(BA) 통계국에서 통합되고, 관리되고 있다. IEB 데이터베이스를 구성하는 주된 출처는 고용 이력정보, 보험수급자 이력정보, 실업보험수급자 이력정보, 구직자 이력정보 등의 행정 정보이고 위와 같은 정보는 사회보장번호 및 사업체 ID를 이용하여 서로 연계된 데이터베이스로 구축되어 있다.

아래 [그림 1]은 독일 고용관련 DB의 구축과정을 제시하고 있다. 아래 그림과 같이 사회보장번호 및 사업체ID로 각기 다른 DB를 연결할 수 있으며, 이를 이용하여 다양한 분야의 연구에 사용되는 것이 독일 고용관련 DB의 또 다른 강점으로 생각된다.

Internal Data(AB)			External Data
Social Security Records	Operative systems of BA	Survey data	행정/조사 data
<ul style="list-style-type: none"> ● 사회보장번호 ● 고용기간/형태 ● 성, 연령, 학력 등 ● 직업 ● 임금 ● 고용된 사업체 특성 *1년 주기로 Update 	<ul style="list-style-type: none"> ● 실업급여 수급이력 ● 훈련/프로그램 참여이력 ● 구직이력 ● 사회보장번호 	<ul style="list-style-type: none"> ● Establishment Panel Data 	<ul style="list-style-type: none"> ● 특허정보 ● 기업재무 및 관리자 정보 <p>*학력정보 연계 중</p>
사회보장번호/사업체 ID로 DB 연계			

그림 1. 독일 고용관련 DB 구축

IEB 데이터베이스의 전체 정보는 개인정보보호 및 데이터의 대용량 측면에서 접근이 제한되어 있으나, Research Data Center(FDZ)에서 학술적인 목적으로 이용 가능한 표준화된 데이터 셋을 구축하고 있으므로, 무료로 사용할 수 있는 방법이 존재한다.

The Research Data Center에서 제공하는 다양한 데이터 셋 중 Sample of Integrated Labor Market Biographies(SIAB)의 경우 IEB 데이터베이스의 2% 샘플로서 취업자의 고용기간, 구직 및 실업급여 이력, 성, 연령, 국적, 학력 등의 개인정보 등이 포함되어 있는 것이 특징이다. 또 다른 데이터셋인 Linked Employer-Employee Data의 경우 독일 노동시장의 수요-공급 측면을 동시에 분석하는 것이 가능하다.

Combined IEB-Working and Learning in a Changing World (ALWA-ADIAB) data의 경우 근로자 개인 특성변수와 사업체 개별 특성변수가 포함된 데이터 셋으로서 지역 간 이동 노동시장 진입, 재진입 등의 분석을 할 수 있는 특징이 있다.

한편 FDZ는 IEB에 축적된 행정DB를 기반으로 새로운 데이터 셋을 구축 중인데, 대표적으로 노동시장의 사회연결망, 노동시장과 거래, 기술, 인구학적인 관계를 규명하기 위한 Geocoded Data과 독일 노동시장의 사회보장 수급을 분석하기 위한 the Panel Labor Market and Social Security Data를 들 수 있다.

이와 같이 독일의 경우 다양한 고용관련 DB를 오랜 기간 동안 잘 축적 및 관리하고 있으며, 이를 이용하여 다른 데이터와 연계하는 작업을 꾸준히 하고 있으며, 학술적 목적 및 노동시장 정책적 목적을 위해 다양한 방식으로 고용관련 DB를 이용하고 있다.

2. 영국

영국은 the Administrative Data Research Network(ADRN)에서 통계청, 고용연금부, 교육부, 법무장관부, 내각부, 외무부 등 정부기관 및 유관 연구기관에서 생산 및 관리되는 모든 데이터를 통합하여 행정 데이터와 관련된 정보를 윈스톱으로 제공하는 것이 특징이다. ADRN에는 개별 연구자들이 사회 및 경제적 연구를 위해 다양한 주제의 행정 데이터를 이용할 수 있도록 행정데이터의 출처, 수집과정, 행정데이터 이용 방법, 관련 연구 등의 서비스를 제공하고 있는 것이 특징이다.

ADRN에서 제공하고 있는 행정데이터의 분야는 Economy & Employment, Education & Learning, Business & Third Sector, Crime & Justice, Health & Wellbeing, Population, Housing & Environment 등과 같이 다양하다.

아래 [표 5]는 ADRN에서 제공하는 Economy & Employment와 Education & Learning 분야의 주요 내용으로 2002년부터 축적되어 온 영국의 각 학제별 학생과 관련된 정보(성별, 성적, 종교, 언어 등)를 제공하고 있는 데이터(National Pupil Database), 1999년부터 축적되어 온 고용 및 연금과 관련된 정보를 제

공하고 있는 데이터(Work and Pensions Longitudinal Study) 등이 포함되어 있는 것이 특징이다. 'Education & Learning' 분야에서는 1996년부터 축적되어 온 대학입학정보와 관련된 데이터(Universities and Colleges Admissions Service Higher Education), 1994년부터 축적되어 온 대학생의 취업과 관련된 데이터(Destinations of Leavers from Higher Education) 등을 제공하고 있다.

표 5. Administrative Data Research Network 제공 데이터[10]

주제	데이터 명
Economy & Employment	National Pupil Database(2002-)
	Work and Pensions Longitudinal Study(1999-)
	Child Benefit(2003-)
	Tax Credit Claimants(2003/04-)
	Continuous Recording of Social Housing Lettings and Sales(1989-)
	Scottish Continuous REcording System(1992-)
Education & Learning	Universities and Colleges Admissions Service Higher Education(1996-)
	Destinations of Leavers from Higher Education(1994/95-)
	Destinations of Leavers from Higher Education (longitudinal)(2006-)
	Follow Up School Leaver Destinations Survey (Scotland)(1992/93-)
	Individualised Learner Record(2003/04-)
	National Pupil Database(2002-)

3. 시사점

위에서 살펴본 독일 및 영국의 고용관련 DB의 구축 및 활용사례를 통해 다음과 같은 시사점을 얻을 수 있다. 먼저, 행정 DB의 강점은 실시간으로 축적되는 정보이기 때문에 다른 조사데이터에 비해 정보의 양이 많고, 시의성이 높은 것이 특징이다. 이는 조사데이터가 지나기 힘든 강점으로서 각국에서는 시간이 지날수록 행정 DB의 중요성을 인식하고 있으며, 이를 다양한 방면으로 적극적으로 활용하려는 노력을 하고 있음을 알 수 있다.

다음으로 우리나라도 기관별로 다양한 주제의 행정 DB 및 조사 DB를 생성운영하고 있지만 독일의 사례에서 살펴본 바와 같이 방대한 양의 행정 DB를 보다 효율적으로 구축하고 활용하기 위해서는 연구목적에 맞는 맞춤형 DB를 생성하고, 제공하는 것이 중요하다. 또한 우리나라의 현재 행정 DB는 개인/사업체 단위에서 주제별로 데이터가 생성되고 있으므로, 각 기관별 연계성이 떨어지는 것이 사실이다. 이를 보완하기 위해 개

인 및 사업체를 연계 단위로 하여 유사한 주제를 연결하여 데이터를 생성하고 관리하는 것이 필요하고, 학술적 및 노동시장 정책적 목적을 위해 데이터를 제공하는 것이 중요하다고 판단된다. 마지막으로 무엇보다도 독일 IEB의 사례(고지의무)에서 알 수 있듯이 고용관련 DB들간의 연계를 위해서는 이를 가능하게 하는 법적 근거가 중요함을 알 수 있다.

IV. 고용관련 빅데이터 구축 및 활용 방안

1. 고용정책 방향

1.1 적극적 노동시장 정책(ALMP)

현재 우리나라는 고용없는 성장, 고용형태 다양화, 근로 빈곤, 일과 가정의 양립, 고령화 등의 사회경제적 환경에 효율적으로 대응할 수 있는 고용정책이 요구되고 있다.

초창기 우리나라 고용정책은 실업자에게 구직(실업) 기간 동안의 소득보전을 목적으로 구직급여를 지원하는 다소 소극적이고 수동적인 정책으로 시작하였다. 하지만, 1990년대 후반부터는 이러한 소극적인 급여 중심의 정책으로는 실업률을 낮추기 어렵고 장기실업자를 양산하게 된다는 우려 속에 보다 적극적인 고용정책으로 전환하는데, 이를 적극적 노동시장정책(Active Labor Market Policies: ALMP)이라 부른다.

적극적 노동시장정책의 핵심 과제는 실업자가 직업 훈련, 취업상담 등과 같은 다양한 취업지원 프로그램 참여를 통해 취업동기를 제고시켜 조기에 취업(취업을 통한 소득확보)의미로 Making Work Pay)하게 하는 것이다.

Konle-Seidl(2009)는 적극적 노동시장정책이 효과를 갖기 위해서는 실업자와 정부 간에 체결된 사회적 계약에 따라 상호 책임과 의무를 충실하게 수행해야 하나, 현실에서 이를 기대하는 것은 쉽지 않아 해당 정책으로는 고용문제를 해결하는데 한계가 존재한다고 주장하였다[11].

한편, 독일 등 유럽대륙을 중심으로 적극적 노동시장 정책에서 형식적으로 운영되던 구직자의 권리와 의무

를 더욱 강화함으로써 구직자의 취업능력을 향상시키는 활성화 전략(Activation Strategy)이 수행되고 있으나(Eichhorst, 2008), 취업과 이를 위한 취업능력 제고를 강조했다라는 점에서 이전의 적극적 노동시장정책과의 차별점은 크지 않다[12].

1.2 이행노동시장(TLM)

이행노동시장론(Transitional Labour Markets)은 1990년대 중반에 독일의 Schmid 등에 의해 제기된 노동시장을 바라보는 접근법이다. 이행노동시장론은 개인이 생애에 경험하게 되는 다양한 노동력 상태 간의 이행에서 발생할 수 있는 사회적 위험을 체계적으로 관리 하자는 것으로 '이행을 통한 소득 확보(Making Transitions Pay)'를 실현하려는 목표지향적인 성격을 지닌다.

Schmid(1998)는 개인이 생애 과정에서 직면하게 되는 여러 노동시장 상태 간의 이행을 5개 유형으로 구분하였다[13]. [그림 2]에서 유형-I은 취업상태에서 종사상 지위나 고용형태 간의 이행을 의미한다. 예를 들면, 단시간 근로에서 전일제 근로로 이행, 임금근로에서 비임금근로로 이행, 비정규직에서 정규직으로의 이행 등이 여기에 속한다. 다음으로 유형-II는 취업과 실업 간의 이행을 의미하며, 유형-III은 교육·훈련과 취업 간의 이행을 의미한다. 계속해서, 유형-IV는 가사·육아와 취업 간의 이행이며, 마지막 유형-V는 취업과 은퇴로의 이행을 의미한다.

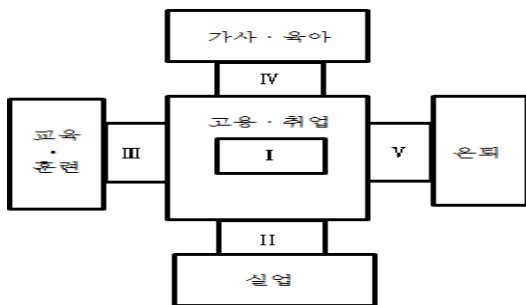


그림 2. Schmid(1998)의 노동시장 이행 5개 유형

Schmid(2006)는 5가지 유형의 이행에는 고유한 사회적 위험이 수반됨으로 이에 대응할 수 있도록 5가지

의 제도적 장치가 마련되어야 한다고 강조했다[14]. 첫 번째로는 여러 고용형태 간의 이행은 개인의 소득 감소와 함께 저임금으로 인한 취약성과 근로 빈곤층으로의 추락 위험을 안고 있어 소득 안정을 보장해주는 제도가 필요하다.

두 번째는 고용과 실업 간의 이행에는 실업으로 인한 소득 불확실성의 위험을 줄이고 구직활동에 전념할 수 있도록 실업기간 중 소득유지 장치가 필요하다. 세 번째 교육·훈련과 취업 간의 이행에는 생애에 걸친 소득 확보 능력의 감식 또는 부족에 따른 위험이 존재한다. 따라서, 취업능력을 개발하고 향상시켜 소득확보 능력을 유지할 수 있게 지원하는 제도적 장치가 필요하다. 네 번째는 생애 과정에서 자녀나 부양가족의 보호라는 사회적 책무수행으로 소득확보 능력이 제한될 위험이 존재한다. 따라서, 이 경우에는 소득지지 장치가 필요하다. 마지막 다섯 번째로는 취업에서 은퇴로의 이행은 소득확보 능력이 심각하게 감소하거나 완전히 상실될 위험이 존재함에 따라 소득대체 장치가 필요하다고 주장하였다.

정병석(2010)은 지금까지의 우리나라 노동시장정책에 대한 비교분석을 통해 현재 직면하고 있는 고용문제를 완화시키기 위해서는 이행노동시장론에 입각한 고용정책이 필요하다고 주장하였다[15]. 연구자는 특히 사회통합적 고용정책을 위해서는 통상적인 노동시장정책뿐만 아니라 직업능력개발 및 교육 정책, 사회보장 정책, 세제 및 연금 정책 등이 고용정책과 연계되어 종합적으로 검토되어야 한다고 주장하였다.

이상의 논의는 고용문제를 비롯하여 현재 우리가 겪고 있는 다양한 사회문제들을 완화시키기 위한 정책개발 및 집행에는 무엇보다도 사회통합적 개인 맞춤형 접근이 필요함을 시사하고 있다. 따라서, 정책개발을 지원하는 고용관련 빅데이터 역시 고용관련 DB뿐만 아니라 교육·훈련, 건강, 복지 등과 같은 관련 주제와의 연계 구축이 필요하다.

2. 고용관련 빅데이터 구축 방안

고용관련 빅데이터는 이행노동시장 관점 하에서 고용보험 DB를 중심으로 개인 단위에서 교육/훈련/경력/자격, 구직, 고용, 실업, 비정활/복지 주제별로 관련

DB를 연계하여 구축한다.

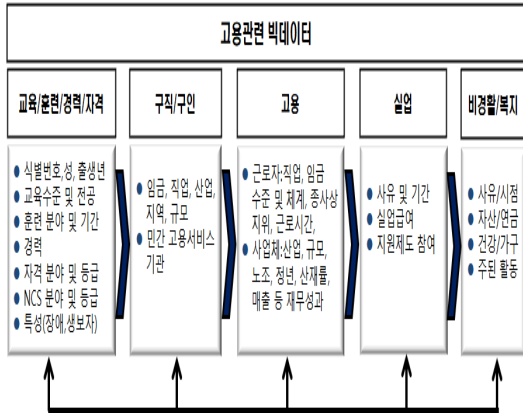


그림 3. 고용관련 빅데이터 구축 체계

빅데이터 구축에 활용하는 DB는 행정 DB를 사용함으로써 자료의 신뢰도와 시의성을 확보하고, 관련 주제에 대한 조사자료 결과와의 비교를 통해 자료 신뢰도를 검증한다. 또한, DB간 연계분석을 위해 직업(한국고용 직업분류), 산업, 훈련 등과 같은 분류체계를 최대한 통일하고 가능하면 세분류 수준에서 구축한다. 아울러, 고용관련 빅데이터에 추가되어야 할 주제(고용서비스 민간사업체, 직업상담사, 훈련강사 등)를 추가 발굴하고, 관련 DB의 확보 및 연계 방안을 마련하고, 미래 고용시장에서 중요한 이슈(ex, 외국인근로자)로 부각할 주제에 대한 연계 방안도 함께 마련할 필요가 있다. 무엇보다도 고용관련 빅데이터 발굴, 연계, 활용 그리고 개선을 위한 관련 기관간의 네트워크 구축 및 운영도 중요한 체계의 일부이다. 아래 [표 6]에는 고용관련 빅데이터 구축에 필요한 주제별 DB 관리 기관이 제시되어 있다.

표 6. 고용관련 빅데이터 주제별 DB 관리 기관

주제	관리 기관
교육 DB	-교육부 한국교육개발원: 대학 이상 -한국교육학술정보원: 고등학교 이하
훈련 DB	-고용센터 및 산업인력공단: 실업자 및 사업주훈련 DB -한국기술교육대학: 직업훈련기관 및 교강사, 훈련과정 및 이수자 평가, 훈련비 -타 부처 및 지자체: 실시 훈련 정보
자격 DB	-한국산업인력공단: 국가기술 및 전문자격 취득자 DB -한국직업능력개발원: 민간자격

임금 DB	-근로복지공단: 고용-신제 보험 보수총액 신고 DB
사업체 재무 DB	-국세청/한국거래소: 재무 DB
복지/가구 DB	-사회보장정보시스템(행복e음): 각종 사회복지 급여 및 서비스 지원 대상자의 자격과 이력 DB
이민/고용허가제	-행정자치부, 법무부 출입국·외국인정책본부 DB

3. 고용관련 빅데이터 활용 방안

위와 같이 구축된 고용관련 빅데이터는 다양한 분야에서 활용될 수 있다. 먼저, 각종 고용관련 주제에 대한 조사 설계 시 모집단과 표집틀로 활용가능하다. 두 번째로는 현재 운영되고 있는 한국직업정보시스템(KNOW)과 바로원(one) 시스템에 직업별 종사자의 임금, 전공, 연령 등과 같은 직업관련 추가 정보가 제공 가능할 것이다.

세 번째는 인력수급전망 시 전공-직업 연관계수 산출, 성장수요에 대체수요 반영, 특정 산업/직업 전망 모형 개발에 필요한 기초자료 제공 등에 활용될 수 있다. 네 번째는 생애진로개발과 관련하여, 특정 전공 후 종사 산업 및 직업을 기반으로 전공별 Career-Map을 작성하고, 다양한 진로 및 취업지원 프로그램에 대한 효과 평가 자료로 활용가능하다. 다섯 번째로는 구직자 유형분류(profiling) 및 서비스 결정, 고용서비스 제공 기관에 대한 성과 평가 등과 같은 고용서비스 효율화에 활용가능하다.

여섯 번째로는 고용보험 등을 포함한 고용관련 정책의 효과 평가에 활용가능하다. 예를 들어 실업자 훈련(ex, 내일배움카드제)의 취업효과 분석방안을 설명하겠다. 지금까지 훈련의 취업효과를 분석하기 위한 자료는 주로 표본조사를 통해 확보하였다. 표본조사를 수행 할 때, 훈련 참여가 무작위적으로 결정되지 않음에 따라 발생가능한 자기선택 편의(Self-Selection Bias)를 교정가능하게 하는 성향점수매칭법(Propensity Score Method)을 사용하기 위하여 실험집단(훈련 참여집단) 뿐만 아니라 비교집단(훈련 비참여집단)도 함께 조사한 후 이중차분법(Difference In Difference) 등을 사용하여 취업효과를 분석하였다.

이러한 방식의 효과분석은 조사를 통한 자료수집 시 발생가능한 비표본오차(회고적 응답에 따른 측정오차 등)로 인한 자료 신뢰도 문제, 대표본 조사에 따른 고비용 그리고 연구 설계, 자료수집, 분석 등에 상당한 시간

이 소요됨에 따른 연구결과의 시의성 제약 문제를 지니고 있다. 이에 견주어, 고용관련 빅데이터를 사용할 경우에는 사전에 분석 모형을 설정하기만 하면 언제라도 양질의 자료로 적은 비용을 투입하여 훈련의 취업효과를 단기에 분석할 수 있다.

이와 관련하여, 유경원(2020)은 행정자료가 기존의 조사자료와 달리 샘플링 편이의 문제가 적고 패널화되어 있어 특정한 정책이 경제 및 사회현상에 미친 인과관계를 사후적으로 엄밀히 검증하기에 적합하다고 밝혔다[16].

표 7. 고용관련 빅데이터 활용 방안

활용분야	활용내용
모집단/표집틀	-고용시장 정보 수집을 위한 조사 설계시 모집단/표집틀 활용
고용정보	-한국직업정보시스템과 바로원 시스템에 직업별 임금, 종사자수, 연령 등과 같은 추가 정보 제공
인력수급전망	-노동공급 전망 시 전공-직업 연관계수 시계열 제공 -노동수요 전망시 성장수요에 대해수요(산업/직업이동) 반영 가능
생애진로개발	-School to Work을 기반으로 Career-Map(전공별 취업 취업 및 산업 등) -진로 및 취업지원 프로그램 효과 평가
고용서비스	-구직자 유형분류(profiling) 및 서비스 결정 -고용서비스 제공기관(고용센터 등) 성과 평가
고용동향분석	-Work-net기반 실업원인 분해, 지역간 노동력 이동, 주제별 고용시장 특성 분석
고용보험평가	-개인(훈련 등의 고용서비스)/기업(임금피크제 등) 지원 프로그램 효과 평가
이민/고용허가제	불법체류자 특성 분석을 통해 쿼터 및 자격요건 정비
DB Product	일(고용)-교육/훈련/자격-복지 포함 공개용 DB

V. 향후 과제

본 연구는 개인별 맞춤형 고용 서비스를 지원하기 위한 '고용관련 빅데이터 센터(가칭)'설립에 앞서 우리나라의 고용관련 행정DB 중에서 워크넷, 고용보험, 직업훈련 DB 등과 함께 교육과 복지 DB를 중심으로 DB별 개요, 구축과정, 주요 내용, 활용실태 그리고 한계점 등을 살펴보았다. 이어서, 고용관련 빅데이터를 구축·운영하고 있는 독일과 영국의 사례 분석을 통해 시사점을 도출하였다.

이상의 분석에 기반하여 개인의 다양한 노동력 상태 간의 이행에서 발생할 수 있는 사회적 위험을 체계적으로 관리해야함을 강조하는 이행노동시장론에 입각하여, 고용보험 DB를 중심으로 개인 단위에서 교육/훈련/경력/자격, 구직, 고용, 실업, 비경활/복지 등의 DB를 연계하는 체계로 고용관련 빅데이터 구축 및 활용 방안을 제시하였다.

지금부터는 고용관련 빅데이터를 구축하기 위해 남겨진 과제 및 연구의 한계점을 제시하겠다.

먼저, 남겨진 과제들은 다음과 같다. 가장 먼저, 고용보험 DB를 비롯한 고용 영역 DB들에 대한 정비가 우선 요구된다. 특히 구축되어 있는 DB 내 개별 변수의 조작적 정의와 값 설명에 대한 정리가 시급하다.

다음으로 고용 영역 DB와 관련 영역 DB간의 연계 활용 우수사례(Best-Practice)를 발굴하여 유관 기관의 협조를 제고함과 동시에 향후 새롭게 생성될 DB에 대한 기관 간 사전 협업으로 연계 시 문제(ex, 각종 분류 체계 불일치 문제)를 최소화할 필요가 있다.

이어서 고용보험 DB를 중심으로 구축된 빅데이터의 대표성 제고가 필요하다. 이를 위해서는 정부가 밝힌 바와 같이 영세 자영업자, 특수형태근로종사자, 프리랜서 등의 고용보험 가입 확대 노력이 필요하다. 또한, 조사로 수집된 인식, 태도, 의향 등에 관한 DB를 연계할 필요가 있다. 계속해서, 고용관련 빅데이터 구축에 사용되는 DB의 신뢰도 제고가 필요하다. 이를 위해서는 유관 기관과의 협조를 통해 DB 최신화주기를 단축하고, 통계청 등에서 제공되는 정보와의 교차검증(Cross-Check) 및 생성 DB에 대한 정제(Cleaning) 작업 강화가 요구된다.

다음으로는 구축된 빅데이터에 대한 개방 수준 및 제공 방식이 결정되어야 한다. 이는 빅데이터 사용에 대한 비용청구 여부를 포함하여, 빅데이터 제공 방식(직접 접속 후 다운로드, 센터 요청 후 파일 수령 등) 및 수준(구축 DB 전체 또는 일부 등) 등에 관한 것으로 향후 고용관련 빅데이터 운영 인력 및 예산 수준과 정보 보안 등을 고려하여 결정해야 할 것이다.

이러한 고용관련 빅데이터를 구축하고 운영하기 위해서는 빅데이터 전문인력 및 인프라(전산센터) 지원이 필요하다. 빅데이터 전문인력(Data-Scientist)은 DB 구축에 관한 기획뿐만이 아니라 DB를 활용한 분석 및

해석이 가능한 인력으로, 우리나라뿐만 아니라 세계적으로도 그 수가 많지 않다. 현재, 우리나라의 경우 한국 데이터베이스진흥원을 중심으로 빅데이터 전문인력의 양성 및 향상 교육이 진행되고 있다. 인프라는 고용관련 빅데이터를 집적·추출·변환하는데 어려움이 없어야 할 뿐만 아니라 향후 일부 주제(DB)에 대한 종단화(logitudinal)에도 문제가 발생하지 않도록 지원되어야 한다.

또한, 고용관련 빅데이터 구축 및 활용을 위해서는 무엇보다도 고유식별정보를 사용한 DB 간 연계가 필수적이다. 현재 개인정보보호법 등에서는 고유식별정보의 처리를 매우 엄격하게 제한하고 있는 상황이다. 최근 행정자치부가 중심이 되어 빅데이터 활용확산을 위해 개인정보 비식별 조치 가이드라인을 제시하였으나, 적법성에 관한 논란이 지속되고 있는 상황이다. 따라서, 개인정보보호법이 추구하는 목적을 훼손하지 않는 범위에서 정부 및 공공기관이 공익을 추구하는 경우에 한하여 고유식별정보를 처리할 수 있도록 지속적인 법 정비가 필요하다.

마지막으로 본 연구가 고용관련 행정DB를 연계하여 고용관련 빅데이터를 구축하고 활용하는 방안을 제시하고는 있으나, 해당 DB들을 어떻게 확보하고, 어떤 변수로 어떻게 서로 연계하고, 이러한 과정에서 예상되는 문제들과 대처방안 등과 같은 구체적인 사안들을 제시하지는 못했다는 한계점을 갖는다.

참 고 문 헌

- [1] 성옥준, “공공부문 빅데이터 정책 활성화 연구,” 한국정책학회보, 제25권, 제2호, pp.125-149, 2016.
- [2] Xi Song and Thomas S. Coleman, *Using Administrative Big Data to Solve Problems in Social Science and Policy Research*, Global Social Security Review, 한국보건사회연구원, 2020.
- [3] 정영철, 오미애, 신정우, *보건복지 정보통계 정책 전망과 과제*, 보건복지포럼, 2021.
- [4] 박과영, “부동산가격공시제도의 공공빅데이터 활용현황과 발전방향,” 감정평가학 논집, 제17권, 제3호, pp.43-55, 2018.
- [5] 윤정혜, *고용보험통계를 이용한 노동이동 분석*, 한국

고용정보원, 2019.

- [6] 이재성, 박승훈, 김지환, *직업훈련 사업 심층평가-구직자훈련사업 동종분야 취업에 관한 연구*, 한국고용정보원, 2018.
- [7] 김병섭, 이민창, 고길곤, 오시영, *자치단체의 재정지원 일자리사업 운영실태 분석*, 서울대학교 행정대학원, 2012.
- [8] 최기성, 어수봉, 조세형, “대학 교육성과의 새로운 평가 방안 연구: 취업통계조사의 취업률을 중심으로,” 조사연구, 제16권, 제2호, pp.59-88, 2015.
- [9] Joerg Heining, “The Integrated Employment Biographies: Origins, Data Access and Further Development,” 한국고용정보원 국제세미나「행정DB를 활용한 노동시장 분석」, 2014.
- [10] <https://adrn.ac.uk/>
- [11] R. Konle-Seidl, “Activation Policies in Germany,” 한국노동연구원 국제세미나 「활성화 정책과 고용서비스 성과」, 2009.
- [12] Eichhorst, *Bringing the Jobless into Work*, IZA, Germany, 2008.
- [13] G. Schmid, *Transitional Labour Markets: A New European employment strategy*, WZB Discussion Paper, 1998.
- [14] G. Schmid, “Transitional Labour Markets: Experience from Europe and Germany,” *Australian Bulletin of Labour*, Vol.32, No.2, pp.114-138, 2006.
- [15] 정병석, “한국노동시장정책의 평가와 발전방안: 이행노동시장 이론의 활용,” *노동정책연구*, 제10권, 제2호, pp.155-185, 2010.
- [16] 유경원, “빅데이터를 이용한 정책분석 방법론에 관한 연구,” *텔코저널*, 제8권, pp.39-75, 2020.

저 자 소 개

최 기 성(Ki-Sung Choi)

정희원



- 2001년 8월 : 고려대학교 신문방송학과(석사)
- 2012년 8월 : 한국기술교육대학교 인력경영학(박사)
- 2015년 2월 ~ 현재 : 한국고용정보원 연구위원

〈관심분야〉 : 직업 및 고용 정보, 인력경영