

# 사회보장 장기재정분석 I

## : 사회보장 재정지표 탐색 및 DB 구축

송창길

나원희·최현수·신정우·송창용·한신실·최광성·고경표·조용찬



사람을  
생각하는  
사람들



KOREA INSTITUTE FOR HEALTH AND SOCIAL AFFAIRS



한국보건사회연구원  
KOREA INSTITUTE FOR HEALTH AND SOCIAL AFFAIRS



## ■ 연구진

연구책임자	송창길	한국보건사회연구원 부연구위원
공동연구진	나원희	한국보건사회연구원 부연구위원
	최현수	한국보건사회연구원 연구위원
	신정우	한국보건사회연구원 선임연구위원
	송창용	한국직업능력연구원 선임연구위원
	한신실	국민연금연구원 부연구위원
	최광성	한국직업능력연구원 부연구위원
	고경표	한국보건사회연구원 전문연구원
	조용찬	한국보건사회연구원 연구원

연구보고서 2025-37

### 사회보장 장기재정분석 I

: 사회보장 재정지표 탐색 및 DB 구축

발 행 일 2025년 12월  
발 행 인 신 영 석  
발 행 처 한국보건사회연구원  
주 소 [30147]세종특별자치시 시청대로 370  
세종국책연구단지 사회정책동(1~5층)  
전 화 대표전화: 044)287-8000  
홈페이지 <http://www.kihasa.re.kr>  
등 록 1999년 4월 27일(제2015-000007호)  
인 쇄 처 고려씨엔피

© 한국보건사회연구원 2025

ISBN 979-11-7252-112-7 [93510]

<https://doi.org/10.23060/kihasa.a.2025.37>

## 발|간|사

사회보장 재정분석은 중장기 정책 수립에 핵심적인 기초자료를 제공하는 도구로서 정책 연구에서 매우 큰 가치를 지닌다. 특히 급격한 인구구조 변화와 복지 수요 확대라는 전례 없는 환경에서, 지출 규모와 제도적 특성을 동시에 파악할 수 있는 통합적 분석 체계의 필요성이 더욱 높아지고 있다.

본 연구는 OECD SOCX의 양적 데이터와 제도적 질적 정보를 대규모 언어모델(LLM) 기술을 활용하여 결합하는 새로운 접근을 시도하였다. 3차원 분류체계를 설계하고 전체 사회보장 영역의 6,637개 프로그램을 포괄하는 통합 데이터베이스를 구축하였으며, 노령 영역 시범분석을 통해 국제비교와 장기재정분석에의 적용 가능성을 확인하였다.

이 연구는 송창길 부연구위원 책임하에, 나원희 부연구위원, 최현수 연구위원, 신정우 선임연구위원, 고경표, 조용찬 연구원이 참여하였고, 외부 연구진으로는 한국직업능력연구원의 송창용, 최광성 박사, 국민연금연구원의 한신실 박사가 참여하였다. 또한, 함영진 선임연구위원과 류재린 연구위원, 그리고 국민연금연구원의 김혜진 박사의 자문으로 연구의 질을 한층 높일 수 있었다. 연구를 수행하는 과정에서 조언을 아끼지 않은 재정통계연구실 박사님들께 깊은 감사를 드린다. 마지막으로 이 보고서의 내용은 한국보건사회연구원의 공식 견해가 아님을 밝힌다.

2025년 12월

한국보건사회연구원 원장

신 영 석



# 목 차

KOREA INSTITUTE FOR HEALTH AND SOCIAL AFFAIRS



요약 .....	1
제1장 서론 .....	5
제1절 연구 배경 및 목적 .....	7
제2절 연구 범위 및 구성 .....	11
제2장 사회보장 재정분석을 위한 이론적 검토 .....	13
제1절 사회보장 재정분석 현황 및 특성 .....	15
제2절 사회보장 재정분석을 위한 분류체계 검토 .....	34
제3절 사회보장 재정분석 체계 구축 방안 .....	63
제3장 사회보장 재정분석을 위한 국제 DB 분석 .....	75
제1절 OECD SOCX DB 분석 .....	77
제2절 주요 사회보장제도 DB 분석 .....	101
제3절 사회보장 재정분석을 위한 DB 구축 방향 .....	131
제4장 사회보장 재정분석을 위한 DB 구축 .....	145
제1절 데이터 수집 및 전처리 .....	147
제2절 SOCX, ISSA 데이터 연계 .....	160
제3절 LLM 기반 제도 정보 구조화 .....	171
제5장 노령 정책영역 파일럿 분석 .....	185
제1절 제도 유형별 분석 .....	187

---

제2절 대상 및 급여 유형별 특성 .....	194
제3절 파일럿 분석의 의의 .....	199
<b>제6장 결론 .....</b>	<b>203</b>
제1절 연구 결과 요약 .....	205
제2절 주요 시사점 .....	209
제3절 연구의 한계 및 향후 과제 .....	212
<b>참고문헌 .....</b>	<b>217</b>
<b>부록: LLM 기반 비정형 텍스트 DB 구축 방법론 .....</b>	<b>225</b>
<b>Abstract .....</b>	<b>259</b>

# 표 목차

KOREA INSTITUTE FOR HEALTH AND SOCIAL AFFAIRS



〈표 1-1〉 3개년 연구의 단계별 목표 및 주요 내용 .....	9
〈표 2-1〉 국내 사회보장 재정분석 사례 비교 .....	29
〈표 2-2〉 SOCX 사회정책 기능별 분류 .....	42
〈표 2-3〉 ESSPROS의 사회보장제도 기능별 정의 .....	46
〈표 2-4〉 ESSPROS의 재원 조달 분류체계 .....	47
〈표 2-5〉 MISSOC 제도별 카테고리 .....	49
〈표 2-6〉 SSI의 사회보장재정 자료 체계 .....	51
〈표 2-7〉 ISSA 데이터 구조 .....	54
〈표 2-8〉 제도유형별 핵심 특성 .....	62
〈표 2-9〉 제도유형×정책영역 매트릭스에 따른 사회보장 재정분류 체계 .....	66
〈표 3-1〉 OECD SOCX 데이터 수집 체계 .....	78
〈표 3-2〉 SOCX 데이터 제공 형태 .....	82
〈표 3-3〉 SOCX Country Notes 구성 .....	83
〈표 3-4〉 주요 사회보장제도 SOCX 코딩 사례 .....	85
〈표 3-5〉 SOCX 분류 일관성 관리 사례 .....	86
〈표 3-6〉 SHA 2011의 재원조달 분류 .....	88
〈표 3-7〉 OECD Employment Database의 ALMP 분류 .....	90
〈표 3-8〉 SOCX의 가족 영역 중 ECEC 분류 예시(한국) .....	92
〈표 3-9〉 ISCED 2011 분류에 따른 취학전 교육단계(ISCED 0) .....	92
〈표 3-10〉 ESSPROS Core System-지출(Expenditure) .....	102
〈표 3-11〉 ESSPROS Core System-수입(Receipts) .....	103
〈표 3-12〉 Detailed datasets와 Selected datasets의 비교 .....	109
〈표 3-13〉 MISSOC-SOCX 카테고리 매핑 .....	111
〈표 3-14〉 SSI 기능별 3차원 데이터 수집 구조(예시: 노령, 실업, 기초교육) .....	119
〈표 3-15〉 SSI 3차원 데이터 구조 .....	121
〈표 3-16〉 ISSA Country Profiles 지역별 포괄 범위 .....	125
〈표 3-17〉 ISSA 5개 주요 영역별 데이터 구성 .....	126

〈표 3-18〉 표준화 항목과 장기재정분석 활용도 .....	128
〈표 3-19〉 정량적 DB 특성 종합 비교 .....	134
〈표 3-20〉 정성적 DB 특성 종합 비교 .....	138
〈표 3-21〉 SOCX와 ISSA의 상호 보완적 특성 .....	141
〈표 3-22〉 SOCX-ISSA 분류체계 매핑 방안 .....	142
〈표 4-1〉 통화 변환 및 비율 계산 절차 .....	154
〈표 4-2〉 SOCX 데이터 구조 예시: 한국 노령 정책영역(2021년) .....	155
〈표 4-3〉 최종 통합 데이터의 구조(JSON 스키마) .....	181
〈표 4-4〉 최종 통합 데이터베이스 구조 .....	182
〈표 5-1〉 노령 지출 구성 비교(2021년) .....	188
〈표 5-2〉 공공 수당/서비스형의 대상 유형별 구성비 .....	195
〈표 5-3〉 사민주의 5개국 공공 수당/서비스형의 대상 유형별 구성비 .....	196
〈표 5-4〉 공공 사회보험의 급여 유형별 지출(체제 평균, GDP 대비 %) .....	197
〈부표 1〉 Open LLM 비교 .....	232
〈부표 2〉 상용 LLM 비교 .....	234
〈부표 3〉 프롬프트 엔지니어링 기법별 특성 및 적용 방안 .....	237
〈부표 4〉 데이터 수집 방법별 특징 비교 .....	240
〈부표 5〉 문서 형식별 처리 도구 및 기법 .....	243
〈부표 6〉 전처리 단계별 처리 내용 및 도구 .....	245

# 그림 목차

KOREA INSTITUTE FOR HEALTH AND SOCIAL AFFAIRS



[그림 2-1] 2020-2060년 장기재정전망 국가 지출 분석 범위 및 추계 방식 .....	18
[그림 2-2] NABO 장기 재정전망 흐름도 .....	19
[그림 2-3] 사회보장 재정추계 분류 및 추계 방법 .....	20
[그림 2-4] 국민연금 장기재정추계모형 구조 .....	22
[그림 2-5] Ageing Report 재정분석 절차 .....	27
[그림 3-1] 국가별, 연도별 데이터 가용 현황 및 다운로드 옵션 .....	81
[그림 3-2] SOCX country notes와 metadata 접근 .....	83
[그림 3-3] SOCX country notes 세부프로그램 설명 예시(한국) .....	95
[그림 3-4] ESSPROS의 구성 .....	104
[그림 3-5] ESSPROS에서 SOCX로의 전환 .....	105
[그림 3-6] EUROSTAT 홈페이지 Database 화면 예시 .....	107
[그림 3-7] EUROSTAT 홈페이지 Data visualisations 화면 예시 .....	108
[그림 3-8] MISSOC Comparative Table .....	112
[그림 3-9] MISSOC Comparative Table Result .....	113
[그림 3-10] ILO World Social Protection Database .....	118
[그림 3-11] ISSA Country Profiles 웹사이트 화면 .....	126
[그림 3-12] ISSA Country Profiles 데이터 구조 예시 .....	127
[그림 4-1] 데이터 수집 구조 및 통합 흐름 .....	148
[그림 4-2] 데이터 연계 프롬프팅 구조 .....	164
[그림 4-3] 데이터 연계 프로세스: 한국 기초연금 예시 .....	167
[그림 4-4] 데이터 구조화 프롬프팅 구조 .....	175
[그림 4-5] 데이터 구조화 및 분류 예시(한국-기초연금) .....	177
[그림 4-6] 데이터 구조화 및 분류 예시(독일-법정연금보험) .....	177
[그림 5-1] 노령 지출 구성 비교(2021년) .....	189
[그림 5-2] 복지체제별 평균 노령 지출 4대 구성요소 분해 .....	192
[그림 5-3] 공공 수당/서비스형의 대상 유형별 구성(2021년) .....	195



[부도 1] 임베딩 연산 예시 .....	226
[부도 2] 트랜스포머 아키텍처 .....	227
[부도 3] 셀프 어텐션 매커니즘 예시 .....	228
[부도 4] UMAP 시각화 및 클러스터 분석 예시 .....	252



## 요약

### 1. 연구의 배경 및 목적

현행 사회보장 재정분석은 개별 제도별로 분절적으로 수행되어 전체 재정 부담을 입체적으로 파악하기 어렵다. 국민연금, 건강보험 등 주요 사회보험을 중심으로 제도별 재정추계가 수행되고 있으나, 사회보장 전체의 통합적 재정 부담 수준과 국제적 위치를 체계적으로 진단할 수 있는 분석 도구는 부재한 실정이다. 특히 제도의 정량적 지출 규모뿐만 아니라 제도유형, 대상범위, 급여방식 등 정성적 제도 특성에 대한 구조화된 정보가 부족하여, 타 국가와의 체계적 비교를 통한 정책 방향 설정이 어렵다는 한계가 있다.

본 연구는 이러한 문제의식에서 출발하여 OECD SOCX 기반 국제 비교 가능한 재정분석 체계를 구축하는 3개년 과제의 1차년도 연구이다. 1차년도의 핵심 목적은 LLM(대규모 언어모델)을 활용하여 정량 데이터(지출 규모)와 정성 데이터(제도 특성)를 통합한 데이터베이스를 개발하고, 노령 영역 파일럿 분석을 통해 구축된 DB의 재정분석 활용 가능성을 검증하는 것이다. 2차년도에는 9개 정책영역 전체로 확대하고, 3차년도에는 장기재정 시뮬레이션 모형을 개발하여 정책 실무에 활용 가능한 체계를 완성할 계획이다.

### 2. 주요 연구 내용

본 연구는 크게 세 단계로 진행되었다. 첫째, 제도 특성 분류 체계 설계 단계에서는 사회보장 프로그램을 제도유형(사회보험 vs 사회수당/서비스), 대상유형(보편 vs 선별 vs 기여), 급여유형(정액 vs 소득비례 vs 적립식 등)의 3차원으로 분류하는 체계를 설계하였다. 이는 기존 OECD SOCX의 정량 분류(정책영역, 제공주체, 급여형태)에 제도 작동 원리를

## 2 사회보장 장기재정분석 I: 사회보장 재정지표 탐색 및 DB 구축

반영한 정성 분류를 추가한 것으로, 재정 경로와 제도 특성을 동시에 파악할 수 있도록 한 것이다.

둘째, 정량·정성 통합 DB 구축 단계에서는 41개국 × 7개 정책영역(보건·ALMP 제외)의 6,637개 Code4 프로그램에 대해 Country Notes와 ISSA 텍스트를 LLM(Google Gemma-3-27B-IT) 기반으로 분석하여 3차원 분류를 자동 수행하였다. 5단계 프로세스(입력 준비 → Country Notes 연계 → ISSA 연계 → 정보 통합 → 검증)를 통해 각 프로그램의 서술형 제도 설명을 구조화된 분류값으로 변환하고, 이를 정량 지출 데이터와 결합하여 PostgreSQL 데이터베이스에 JSONB 형식으로 저장하였다. 최종 DB는 Code4를 키로 하여 “프로그램명-지출규모-제도유형-대상유형-급여유형”을 통합 조회할 수 있는 구조이다.

셋째, 노령 영역 파일럿 분석 단계에서는 구축된 DB가 재정분석에 실제로 유용한지를 검증하였다. 노령 영역은 제도 다양성(사회보험·사회수당·서비스 공존), 국제 비교 가능성(41개국 데이터 충실), 재정 비중(복지지출의 최대 비중)이 높아 파일럿 대상으로 적합하다. 분석 결과, 본 연구의 DB는 (1) 정량·정성 통합 재정지표 산출(한국 GDP 4.8% = 공공-보험 2.4% + 공공-수당 1.2% + 민간-보험 1.2%로 세 경로 구체적 식별), (2) 국제 비교를 통한 한국 위치 진단(41개국 중 최하위, ‘이중 저발전’ 상태 확인), (3) 장기재정분석 활용 가능성(복지체제별 비교를 통해 보수주의 8.7% vs 사민주의 4.8%+4.0% vs 자유주의 3.4%+3.0%의 상이한 재정 경로 파악)의 세 가지 측면에서 유효함을 확인하였다.

## 3. 결론 및 시사점

본 연구는 3차원 분류 체계 기반 정량·정성 통합 DB가 의미 있는 재정

지표 산출, 국제 비교를 통한 위치 진단, 장기재정분석 활용이라는 세 가지 목적에 유효함을 검증하였다. 이는 기존 총량 중심 재정분석이 파악하지 못하던 제도 작동 원리와 재정 경로를 구체적으로 식별함으로써, 제도 특성과 재정 규모를 동시에 분석 가능한 재정분석 인프라를 구축하였다는 점에서 의의를 갖는다.

파일럿 분석을 통해 확인된 한국의 노령 제도 특성은 다음과 같다. 한국은 공공-보험(국민연금 2.4%), 공공-수당(기초연금 1.2%), 민간-보험(퇴직연금 1.2%)이라는 세 경로가 공존하나 각각의 성숙도가 낮아 총지출 4.8%로 41개국 중 최하위이며, 이는 '이중 저발전'(공공 저발전 + 민간 저발전) 상태로 진단된다. 기초연금은 선별형 95.0%로 스웨덴·노르웨이의 보편형(100%)과 상이한 경로에 있으며, 국민연금은 성숙도 약 35% 수준으로 향후 지속 증가가 예상된다. 복지체제별 비교를 통해 보수주의(공공-보험 중심), 사민주의(보험-수당 균형), 자유주의(공공-민간 혼합)의 상이한 재정 구조를 파악하고, 한국의 향후 정책 경로 선택지를 탐색하였다.

본 연구의 방법론은 향후 2~3년차 개선을 통해 9개 정책영역 전체로 확대되고 수요측 지표(인구구조, 사회적 위험, 제도 커버리지)와 결합될 것이다. 최종적으로 한국 사회보장제도의 종합적 진단과 정책 시나리오별 재정 영향 평가가 가능한 장기재정분석 체계로 발전할 수 있을 것으로 기대된다. 이는 개별 제도별로 분절되어 있던 기존 재정추계의 한계를 극복하고, 국제 비교를 통한 정책 방향 설정이 가능한 통합적 분석 도구를 제공한다는 점에서 정책적·학술적 기여를 할 것으로 판단된다.



사람을  
생각하는  
사람들



KOREA INSTITUTE FOR HEALTH AND SOCIAL AFFAIRS



# 제1장

## 서론

제1절 연구 배경 및 목적

제2절 연구 범위 및 구성



# 제 1 장 서론

## 제1절 연구 배경 및 목적

### 1. 연구 배경

한국의 사회보장제도는 지난 수십 년간 압축적 성장을 통해 선진국 수준의 제도적 틀을 갖추어 왔다. 1988년 국민연금제도 도입과 더불어 건강보험, 고용보험, 산재보험 등 주요 사회보험제도가 확대되었고, 기초생활보장제도, 기초연금, 아동수당 등 사회수당제도와 각종 사회서비스도 지속적으로 발전해 왔다. 이러한 제도 확장은 주로 선진국의 모델을 참고하여 국내 상황에 맞게 발전시키는 방식으로 이루어졌으나 개별 제도별로 분절적인 접근이 주를 이루었다.

그러나 현재 한국은 전례 없는 고령화 위기에 직면해 있다. 세계에서 가장 빠른 고령화 속도와 지속적인 저출산으로 인해 2025년에는 초고령 사회에 진입하였고, 이는 사회보장 재정에 직접적인 압박을 가하고 있는 상황이다. 더욱이 한국이 경험하고 있는 인구구조 변화의 속도와 규모는 기존 선진국들이 경험했던 것을 훨씬 상회하여, 기존의 선진국 모델 참조 방식으로는 한계가 있는 상황이다.

이러한 상황에서 기존의 사회보장 재정분석 체계의 한계가 명확히 드러나고 있다. 현재 국내 사회보장 재정분석은 주로 국민연금, 건강보험 등 주요 사회보험을 중심으로 개별 제도별로 수행되고 있으며, 중앙정부와 지방정부 간 역할 분담도 분리되어 분석되고 있다. 이러한 분절적 접

근은 개별 제도의 지속가능성은 평가할 수 있으나, 사회보장제도 전체의 통합적 방향성을 제시하는 데는 한계가 있다.

특히 “한국은 어디에 있고, 어디로 가야 하는가?”라는 근본적 질문에 대한 답을 구하기 어렵다. 기존의 재정분석들은 주로 국내 제도의 내재적 지속가능성에 초점을 맞추고 있어, 국제적 기준에서 본 한국 사회보장제도의 위치나 지향해야 할 방향에 대한 종합적 판단 근거를 제공하지 못하고 있다. OECD 국가들과 비교할 때 한국의 사회보장 지출 수준은 어느 정도인가? 제도의 구성은 어떻게 다른가? 공공과 민간의 역할 분담은 적절한가? 이러한 기초적인 재정지표조차 체계적으로 파악되지 않은 상황에서 정책 방향을 설정하는 데 어려움이 크다.

또한 선진국들이 고령화 과정에서 경험한 다양한 정책 실험과 그 결과를 체계적으로 분석하여 한국적 상황에 맞는 정책 방향을 도출할 수 있는 분석 도구가 부재한 실정이다. 더욱이 사회보장제도 연구에서는 제도가 지향하는 가치와 운영 방식 등 정성적 특징에 대한 분석이 매우 중요하나, 이러한 정보의 수집과 분석에는 상당한 어려움이 따라 기존 국제 비교 연구들은 주로 지출 규모 등 정량적 지표 중심의 제한적 비교에 그치는 경우가 많았다.

이러한 한계들을 극복하기 위해서는 새로운 접근이 필요하다. 최근 대규모 언어모델(LLM) 기술의 발전으로 제도 설명서, 법령 등에 담긴 정성적 정보를 체계적으로 분석하고 정량화할 수 있는 가능성이 열리고 있으며, 이를 활용한다면 기존의 정량적 지표와 정성적 제도 특성을 통합한 포괄적 재정지표 분석이 가능해질 것으로 기대된다. 특히 세계 최고 수준의 고령화 속도를 경험하고 있는 한국의 현 상황에서 이러한 새로운 분석 도구의 신속한 개발과 적용이 요구되며, 지속가능한 사회보장체계 구축을 위한 시급한 과제라 할 수 있다.

## 2. 연구 목적

### 가. 3개년 연구의 전체 목표

본 연구는 한국 사회보장제도의 공급 측면(정부 지출)과 수요 측면(국민 생애주기별 사회적 위험)을 통합하여 장기재정 시뮬레이션이 가능한 분석 체계를 구축하는 것을 최종 목표로 하는 3개년 과제이다. 이는 개별 제도별로 분절되어 있던 기존 재정분석의 한계를 극복하고, 국제 비교를 통한 한국의 위치 파악과 정책 방향 설정이 가능한 통합적 분석 도구를 개발하기 위한 것이다.

3개년 연구의 단계별 목표와 주요 내용은 <표 1-1>과 같다.

<표 1-1> 3개년 연구의 단계별 목표 및 주요 내용

구분	연구 목표	주요 내용
1차년도 (본 연구)	공급 측면 재정지표 탐색 및 DB 구축	<ul style="list-style-type: none"> <li>- OECD 38개국 사회보장 지출 데이터 수집</li> <li>- LLM 기반 제도 특성 정보 구조화(20개국)</li> <li>- 국제 비교를 통한 한국 재정지표 위치 파악</li> <li>- 노령 영역 파일럿 분석 및 방법론 검증</li> </ul>
2차년도	공급 측면 확대 및 수요 측면 전망	<ul style="list-style-type: none"> <li>- 전체 9개 정책영역 확대(영역별 전문가 분석)</li> <li>- 생애주기별 사회적 위험 전망 체계 구축               <ul style="list-style-type: none"> <li>· 인구구조 변화 전망</li> <li>· 생애주기별 위험 발생 전망</li> <li>· 신사회적 위험 전망</li> </ul> </li> <li>- 공급-수요 통합 DB 완성</li> </ul>
3차년도	공급-수요 통합 시뮬레이션	<ul style="list-style-type: none"> <li>- 공급-수요 매칭 모형 개발</li> <li>- 정책 시나리오별 재정 영향 추계               <ul style="list-style-type: none"> <li>· 지출 확대 시나리오</li> <li>· 제도 개편 시나리오</li> <li>· 인구구조 변화 시나리오</li> </ul> </li> <li>- 지속가능한 사회보장 체계 구축 방안 제시</li> </ul>

출처: 저자 작성.

1차년도에는 공급 측면 재정지표 탐색 및 DB 구축에 집중하며, 2차년도에는 전체 정책 영역으로의 확대와 함께 수요 측면 전망 체계를 추가 구축한다. 3차년도에는 공급과 수요를 통합한 시뮬레이션 모형을 완성하여 정책 시나리오별 재정 영향을 추계하고 지속가능한 사회보장 체계 구축을 위한 정책 제언을 도출한다.

## 나. 1차년도(본 연구)의 세부 목적

본 연구는 3개년 연구의 첫해로서, 사회보장 재정분석을 위한 지표 탐색 및 데이터베이스(DB) 구축에 중점을 둔다. 구체적으로는 세 가지 핵심 질문에 답하고자 한다.

첫째, “어떻게 분석할 것인가?”라는 질문에 답하기 위해 사회보장 재정 분석 체계의 이론적·방법론적 기반을 마련한다. 기존 재정분석의 한계를 검토하고, 정량적 지표(지출 규모)와 정성적 지표(제도 특성)를 통합한 새로운 분석 체계를 설계하는 것이 목적이다. 이를 위해 OECD 및 국제기구의 분류체계를 검토하고, 한국 사회보장 재정분석에 적합한 지표 체계를 탐색한다.

둘째, “어떤 데이터를 활용할 것인가?”라는 질문에 답하기 위해 국제 비교 가능한 DB를 체계적으로 구축한다. OECD SOCX 프로그램 레벨 데이터를 수집 및 전처리하고, LLM 기술을 활용하여 제도 특성 정보(Country Notes, ISSA 정보 등)를 구조화하며, 정량 데이터(지출 규모)와 정성 데이터(제도 유형, 적용 대상, 급여 형태)를 통합한 DB를 구축한다.

셋째, “구축된 체계가 실제 분석에 유용한가?”라는 질문에 답하기 위해 파일럿 분석을 수행한다. 노령(Old-age) 정책 영역을 대상으로 정량·정성 통합 분석 체계의 유용성을 검증한다.

## 제2절 연구 범위 및 구성

### 1. 연구 범위

본 연구의 공간적 범위는 OECD 국가들을 대상으로 한다. OECD Social Expenditure Database(SOCX)에 포함된 38개 회원국과 SOCX DB에 추가되어 있는 비회원국 3개 국을 포함한 총 41개 국가의 세부 프로그램(SOCX의 Code4 기준 프로그램) 사회보장 지출 데이터를 수집하여 정량적 재정지표 데이터베이스를 구축하였다.

본 연구에서는 크게 정량 데이터와 정성 데이터를 구분하는데, 정량 데이터는 자국통화 기준 및 US \$ PPP 기준의 지출금액, GDP 및 정부지출 대비 비율 값을 의미한다. 정성 데이터는 OECD SOCX에서 제공하는 각 국가의 Country-Notes의 프로그램 설명 정보와 ISSA(국제사회보장협회) 데이터베이스 등에서 제공하는 제도의 설명 텍스트 정보를 의미한다.

DB 구축은 OECD SOCX의 7개 정책영역(노령, 유족, 근로무능력, 가족, 실업, 주거, 기타)으로 한하며, 파일럿 분석에서는 노령(Old-age) 영역을 분석 대상으로 선정하여 정량·정성 통합 분석 체계의 유용성과 장기 재정분석 적합성을 검증한다.

### 2. 연구 구성

본 보고서는 총 6개 장으로 구성되어 있으며, 각 장은 “사회보장 재정 분석을 위한 체계 구축”이라는 핵심 주제를 중심으로 유기적으로 연결되어 있다.

제2장 ‘사회보장 재정분석을 위한 재정지표 탐색 및 이론적 검토’에서는 재정분석 체계 구축의 이론적·방법론적 기반을 마련한다. 국내외 사회

보장 재정분석의 현황과 한계를 검토하고, OECD SOCX, ESSPROS 등 국제 분류체계가 제공하는 재정지표 유형과 특성을 분석한다. 또한 복지 체제론을 바탕으로 국제 비교 분석의 준거틀을 검토하고, 정량·정성 통합 분석의 필요성과 방향을 제시한다.

제3장 ‘사회보장 재정분석을 위한 국제 DB 분석’에서는 본 연구에서 활용할 데이터의 구조와 특성을 상세히 검토한다. OECD SOCX 데이터 베이스의 구조, 데이터 수집 체계, 프로그램 레벨 데이터의 장점을 분석하고, Country Notes가 제공하는 정성적 제도 정보의 활용 가능성을 검토한다. 또한 ISSA 등 주요 국제기구의 사회보장제도 DB를 검토하여, 정량·정성 통합 DB 구축을 위한 데이터 원천을 확정한다.

제4장 ‘사회보장 재정분석을 위한 DB 구축’에서는 실제 분석용 DB 구축 과정을 상세히 기술한다. OECD 38개국의 정량 데이터 수집 및 전처리, SOURCE별(공공/민간) 집계 로직 분석, LLM 기반 제도 특성 정보의 구조화 방법, 정량·정성 통합 데이터베이스 스키마 설계 등을 다룬다.

제5장 ‘노령 정책영역 파일럿 분석’에서는 구축된 재정분석 체계의 실제 적용 가능성을 검증한다. 특히 파일럿 분석을 통해 정량·정성 통합 DB를 활용한 재정 분석 체계의 유용성을 평가한다.

제6장 결론에서는 1차년도 연구의 주요 성과를 정리하고 향후 과제를 제시한다. 재정분석 체계 구축의 의의, 정량·정성 통합 DB 구축 성과, 파일럿 분석을 통한 시사점 도출 등을 요약하며, 연구의 한계를 명확히 한다. 나아가 2차년도 전체 정책영역 확대 분석 및 3차년도 장기재정 전망 모델 개발을 위한 연계 방안을 제시한다.



## 제2장

### 사회보장 재정분석을 위한 이론적 검토

제1절 사회보장 재정분석 현황 및 특성

제2절 사회보장 재정분석을 위한 분류체계 검토

제3절 사회보장 재정분석 체계 구축 방안



## 제 2 장

# 사회보장 재정분석을 위한 이론적 검토

### 제1절 사회보장 재정분석 현황 및 특성

#### 1. 국내 사회보장 재정분석 현황

그간 한국은 다양한 사회보장제도를 도입 및 시행하면서 복지국가로 자리매김하였다(한신실, 2020). 특히 OECD SOCX data 기준 GDP 대비 사회보장지출 비중은 지속적이면서도 가파른 증가 궤적을 그리고 있으며, 앞으로 이러한 추세는 한동안 유지될 것으로 예상된다. 이러한 사회보장제도에 대한 지출 증가 궤적은 타 OECD 회원국과 비교할 때, 현저한 차이를 보인다. 일례로 2000년부터 2022년까지 OECD 회원국의 GDP 대비 사회지출 비중의 연평균 증가율을 분석한 결과(OECD, 2025), OECD 회원국은 2000년 평균 17.5%에서 2022년 평균 20.5%로 0.8%의 증가율을 기록한 데 반해 동기간 한국은 4.2%에서 16.2%로 4배가량 증가하면서 연평균 6.6%의 증가율로 OECD 회원국 중 가장 높은 증가율을 기록하였다.

문제는 한국 사회보장제도가 급속한 인구구조 변화와 사회경제적 환경 변화에 직면하여 전례 없는 재정적 도전에 마주하고 있다는 것이다. 통계청(2023) 인구추계에 따르면, 한국은 2025년 초고령사회 진입이 확실시되며,<sup>1)</sup> 2030년대 이후에는 생산가능인구의 급격한 감소가 예상된다. 이

1) 실제 2024년 12월 기준, 전체 주민등록인구는 총 51,217,221명이며, 65세 이상 주민등록인구는 10,117,221명이다.

러한 인구구조 변화는 사회보장제도의 지속가능성에 근본적인 위협 요인으로 작용하고 있다. 일례로 국민연금은 저출생·고령화 추세로 가입자는 감소하고 수급자는 증가하면서, 2057년 기금 소진 이후 수지 적자 구조가 심각해질 것으로 전망된다. 건강보험 또한 의료 기술 발전과 고령화로 의료비 증가 압력이 가중되고 있다. 이외에도 다양한 사회보장제도들의 재정적 지속가능성에 대한 부정적인 평가는 더 이상 새로운 위험인자로 볼 수 없는 상황이다.

이처럼 사회보장제도의 재정적 지속가능성에 대한 문제의식은 사회보장 재정분석의 정확성과 신뢰성이 정책 의사결정의 핵심 요소임을 시사한다. 특히, 증거 기반 정책 수립(evidence-based policy making)의 중요성이 강조되는 현시점에서, 정확한 장기재정분석은 단순한 예측을 넘어 정책 대안의 효과성을 사전 평가하는 핵심 도구로 기능해야 한다. 이는 제한된 재원의 효율적 배분과 세대 간 형평성 확보의 필수 조건이기 때문이다.

그러나 현재 한국의 사회보장 재정분석은 제도별 분절적 접근, 기관 간 방법론 차이, 통합적 관점 미흡 등 개선 과제를 안고 있다. 이는 1990년대 이후 사회보장제도 확대 과정에서 각 제도가 개별적으로 발전한 데서 비롯된 것으로 볼 수 있다. 현재도 제도별로 소관 부처·기관이 서로 다른 자료와 방법론을 적용하고 있어, 제도 간 상호작용과 총체적 재정 영향을 종합적으로 파악하려면 추가적인 노력이 필요한 상황이다.

예를 들어, 동일 제도라도 기관별 거시경제 가정과 방법론 차이로 인해 전망 결과가 달라지는 경우가 있다. 이는 각 기관의 목표와 관점 차이에서 비롯된 것으로 이해할 수 있으나, 정책 결정에 일관된 증거를 제공

---

록인구는 10,256,782명으로 20.0%를 기록하였으며, 2025년 6월 기준, 전자는 51,164,582명, 후자는 10,558,842명으로 20.6%이다.

하려면 기관 간 방법론 조율과 가정의 표준화가 필요하다.

여기에서는 국내 사회보장 재정분석의 실제 사례를 검토함으로써, 현행 재정분석 체계의 특성과 한계를 파악하고, 기관별 재정분석 간 비교분석을 통해 향후 개선 방향을 도출하고자 한다. 이에 대상이 되는 사례는 기획재정부(2020)의 “2020-2060년 장기재정전망”과 국회예산정책처(2020)의 “2020 NABO 장기 재정전망”, 사회보장위원회의 “제5차 중장기 사회보장 재정추계”(보건복지부, 2024b), 국민연금재정추계전문위원회(2023)의 “2023 국민연금재정계산”, 보건복지부(2024a)의 “제2차 국민건강보험 종합계획(안)”이다.

## 가. 정부 재정전망

### 1) 기재부 2020-2060년 장기재정전망

기획재정부는 국가재정운용계획을 통해 40회계연도 이상의 기간에 대해 수행한 장기재정전망 결과를 공개하고 있다. 이는 「국가재정법」 제7조에 근거하여 수행되는 것으로 ‘적어도’ 5년마다 실시하고 있으며 2014년과 2020년 두 차례에 걸쳐 수행되었다. 2020년에 수행된 장기재정전망은 인구구조 변화와 경제성장을 둔화 등 구조적 위험 요인에 대응하여 재정의 지속가능성을 점검하는 것을 목적으로 했다. 전망 체계는 민관합동 장기재정전망협의회를 중심으로 8개 전문기관의 분야별 전망을 종합하여 실시되며, 통계청과 KDI의 거시전망지표를 공통지표로 활용하여 전망의 신뢰성을 담보하였다.

전망은 인구 및 거시 변수 전망의 불확실성을 고려하여 현상 유지(정책 무대응), 성장 대응, 인구 대응 등 3개 시나리오를 구성하여 수행된다. 분석 범위는 국가재정 전체를 포괄하되, 의무지출을 중심으로 한 재정건전

성 관점에서 접근한다. 이때 개별 제도보다는 국가 재정 총량에 중점을 두어 의무지출과 재량지출을 구분하고([그림 2-1]), 국가 채무비율의 장기 추이를 분석하는 총량적 접근 방법을 채택하고 있다. 즉, GDP 대비 지출 비중과 국가채무비율을 핵심 지표로 활용하며, 재정준칙 준수 여부를 판단하는 기준으로 활용된다. 이때 재정준칙은 의무지출 증가에 대응하여 재량지출 증가 속도를 낮추어 중장기 총지출 증가율을 경상성장률 수준과 같도록 관리하는 것이다.

[그림 2-1] 2020-2060년 장기재정전망 국가 지출 분석 범위 및 추계 방식

		총 지 출					
		의무지출					재량지출
분야	① 사회연금보험	② 기타 복지지출	③ 지방이전		④ 이자지출	⑤ 기타 의무지출	
종류	(8개) 국민연금 사학연금 공무원연금 군인연금  건강보험 노인장기요양보험 고용보험 산재보험	기초연금 기초생활보장 보육(아동수당) 장애인 적극적노동시장 보훈 공공임대 등	지방 교부세 (내국세 19.24%)  + 중부세	지방교육 교부금 (내국세 20.79%)  + 교육세	국고채 주택채		
추계 방식	추계	추계	추계		비추계	비추계	비추계
	각 기관별 추계	사회보장위원회 추계	법정 교부율				

\* 사회보험, 기타복지지출은 대부분 의무지출이나, 이에 속한 재량지출은 재량지출로 구분

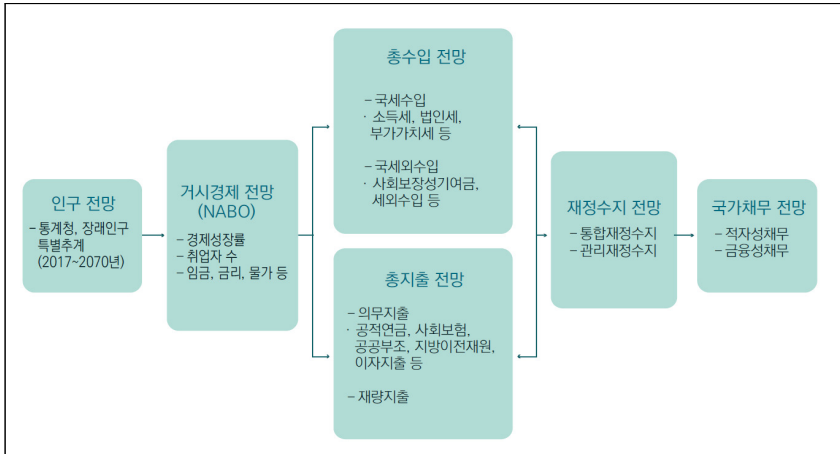
출처: 기획재정부(2020, p.15).

## 2) NABO: 2020 장기재정전망

국회예산정책처의 「2020 NABO 장기 재정전망」은 「국회법」 제22조의 2에 근거하여 국회의 재정통제 기능을 지원하기 위해 실시되는 독립적 재정분석이다. 행정부와 독립된 입법부 산하 기관으로서 객관적이고 중립적인 재정 전망을 제공하는 것을 목적으로 한다. NABO의 전망은 재정준칙을 전제로 수행되는 기획재정부 전망과 달리 기준선(baseline) 전망 방식을 채택하여 현행 제도가 그대로 유지될 경우의 재정 경로를 제시한다.

전망 방법은 총수입, 총지출, 재정수지, 국가채무 등 주요 재정 총량 지표를 중심으로 분석하며([그림 2-2] 참조), 의무지출과 재량지출을 구분하여 지출구조 변화를 전망한다. 개별 복지제도의 경우, 수급자 수와 급여 수준을 추계하여 적산하는 상향식(bottom-up) 방법론을 채택하고 있다.

[그림 2-2] NABO 장기 재정전망 흐름도



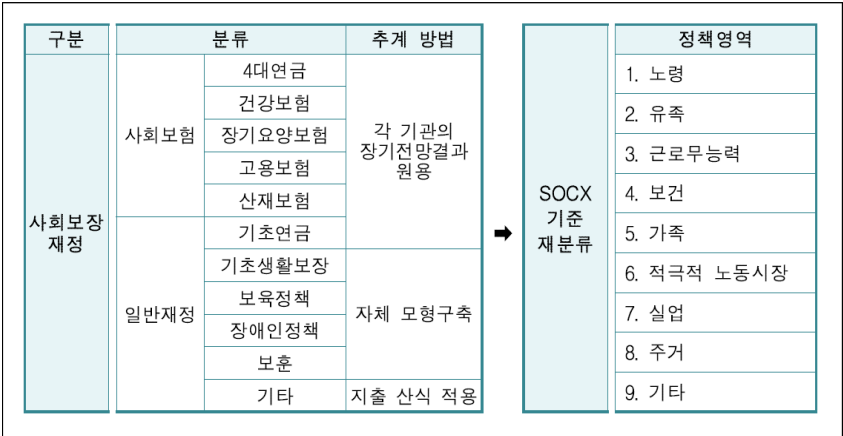
출처: 국회예산정책처(2020, p.6)의 [그림 1] 인용.

나. SOCX 추계

1) 사회보장위원회: 제5차 중장기 사회보장 재정추계

보건복지부는 사회보장위원회와 협력하여 「사회보장기본법」 제5조 제4항에 근거하여 중장기 사회보장 재정추계를 격년으로 수행하고 있다. 추계 기간은 독일이나 스위스 등 해외 사례를 참고하여 약 40년이며, 추계 범위는 OECD 공공사회복지지출(SOCX) 9대 정책영역을 기준으로 하며, 사회보험과 일반재정으로 구분하여 전망하였다([그림 2-3] 참조). 사회보험은 4대 연금과 건강보험, 장기요양보험, 고용보험 및 산재보험 등 8개 제도를, 일반재정은 중앙·지방정부·공기업의 공공부조 및 사회서비스를 포괄한다.

[그림 2-3] 사회보장 재정추계 분류 및 추계 방법



출처: 보건복지부(2024b, p.9).

이 접근법의 핵심은 사회보장 각 제도의 재정 안정성과 급여 적절성을 동시에 고려하려는 시도에 있다. 추계는 사회보험과 일반재정으로 구분

하여 수행되며, OECD SOCX 9대 정책영역별 재정지표를 제공한다. 다만 각 사회보험과 기초연금의 경우 관련 기관에서 수행한 장기전망 결과를 원용하고, 일부 제도에 대해서는 자체 모형을 구축하여 전망함에 따라 각 전망 결과에 사용된 가정이나 방법론의 표준화가 과제로 남아 있다.

## 다. SOCX 추계

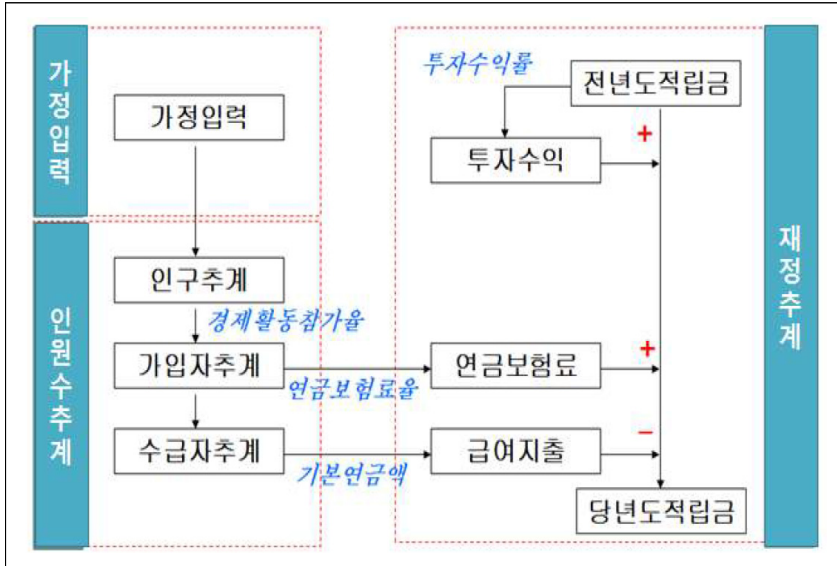
### 1) 국민연금 5차 재정계산

국민연금 재정계산은 1998년 국민연금법 개정 시 도입된 제도로써, 국민연금법 제4조에 근거하여 2003년 이후 5년마다 70년 장기재정추계를 실시하는 법정 재정추계이다. 재정계산의 주요 목적은 인구 및 경제·환경 변화가 국민연금의 보험료 기여 기반과 장기적 재정 안정성에 미치는 영향을 점검하고, 그 결과를 바탕으로 재정안정화 및 기금운용 계획 등 제도 운영 전반에 대한 정책을 선제적으로 수립하는 것이다.

국민연금재정계산은 초장기 전망으로서 보험수리적 균형과 세대 간 형평성 분석에 특화되어 있으며, 확률론적 방법론을 일부 도입하여 불확실성 분석을 강화하고 있다. 전망 결과는 모형 내 적용되는 가정 또는 기초율에 의해 단일 값으로 산출되며, 재정추계모형은 [그림 2-4]에서와 같이 모든 연산 과정이 순차적으로 이루어지는 부분 모듈의 합으로 구조화되어 있다.

다만 단일 제도에 집중된 전망으로 다층연금체계 내에서의 국민연금 역할이나 기초연금과의 상호작용 효과에 대한 분석이 보완된다면, 노령 영역 전반에 대한 통합적 재정지표 제공이 가능할 것이다.

[그림 2-4] 국민연금 장기재정추계모형 구조



출처: 보건복지부(2024b, 9).

## 2) 제2차 국민건강보험 종합계획(안)<sup>2)</sup>

건강보험공단은 「국민건강보험법」 제3조의2에 따라 건강보험의 건전한 운영을 위하여 5년마다 수립하는 정부 법정계획인 「국민건강보험 종합계획」<sup>3)</sup> 일환으로 건강보험의 중장기(5년) 재정 전망(이후, 건강보험 재정전망)을 수행하고 있다.<sup>4)</sup> 지난 2024년 발표된 「제2차 국민건강보험

2) 이하 내용은 보건복지부(2024a)를 참고하였다.

3) 「건강보험 종합계획」에 포함되어야 할 사항은 법률(「국민건강보험법」 제3조의2 제2항)로 규정되어 있으며, 총 11개 영역으로서 건강보험의 중장기 재정 전망 및 운영이 포함되어 있다.

4) 건강보험에 대한 공식적인 재정전망은 국가재정법에 따른 전망과 국민건강보험법에 따른 전망으로 이원화되어 있다(임슬기, 2023). 이외에도 국회예산정책처나 개별 연구자들에 의해 건강보험에 대한 장기재정전망이 수행되고 있다. 문제는 이러한 전망이 반복해서 기술하고 있다시피 수행 주체마다 상이한 전망 결과를 제시하고 있어, 제도 개선 및 운영에 혼란된 신호를 주는 것이다.

종합계획」에 포함된 건강보험 재정전망은 구체적인 수치보다는 재정 지속가능성 제고를 위한 정책 방향에 초점을 두고 있다.

재정 분석은 전망 직전년도 수입 및 지출 변수를 토대로 하며,<sup>5)</sup> 전망 결과 2026년부터 당기수지 적자로 전환할 것으로 추계할 수 있으나 수입·지출의 변수 변화에 따라 매우 다양한 전망이 가능하다는 점에서 전망 결과 해석에 주의할 것을 명시하고 있다. 이는 의료비 증가의 장기 추세 예측에는 근본적 불확실성이 존재하기 때문인 것으로 해석된다. 즉, 의료 기술의 급속한 발전과 새로운 치료법의 도입이 재정에 미치는 영향을 정확히 예측하기 어려우며, 건강보험과 의료급여 간의 역할 분담 변화나 민간의료보험과의 관계 등에 대한 종합적 분석과 체계적인 전망 방법론이 부재한 데서 기인한다.

## 2. 국외 사회보장 재정분석 현황

한국보다 일찍 인구 고령화를 경험한 서구 복지국가들에서도 사회보장 재정분석이 수행되고 있다. 특히, OECD 주요국과 EU는 단순한 개별 제도 분석을 넘어 제도 간 상호작용을 고려한 통합적 접근 방법론을 개발하여 활용하고 있다. 통합적 접근 방법 이외의 제도 특화 재정분석 사례는 따로 검토하지 않았다.<sup>6)</sup> 여기에서는 통합적 접근 방법론을 채택하고 있는 해외 선진국 사례를 체계적으로 검토하여 사회보장 재정분석에 대한 시사점을 도출하고자 한다.

5) 재정전망과 관련해서는 향후 5년간 제도 수지 관련 결과만을 공개하며, 보험료율 인상률, 정부지원금, 수가인상률 등 간단한 가정만 밝히고 있다.

6) 통합적 접근 방법 이외의 제도 특화 재정분석 사례는 따로 검토하지 않았다.

## 가. 캐나다 OCA의 CPP 및 OAS Actuarial Report

캐나다의 Office of the Chief Actuary(이후, OCA)의 Actuarial Report(이후, 재정계산 보고서)는 다층으로 구성된 공적연금 재정분석에서 방법론적 일관성을 확보한 사례이다. OCA는 Canada Pension Plan(이후, CPP)과 Old Age Security Program(이후, OAS)에 대해 각각 독립적인 재정전망을 수행하지만, 동일한 인구·경제 가정을 적용함으로써 분석의 정합성을 확보한다.

OCA의 분석 체계는 다음과 같은 특징을 가진다(Office of the Chief Actuary, 2022; 2023). 첫째, CPP는 75년, OAS는 60년의 장기 전망을 수행하여 각 제도의 특성에 맞는 시계를 설정하고 있다. 둘째, 두 제도의 재정전망에 동일한 인구추계와 경제전망을 적용하여 일관성을 유지한다. 셋째, 각 제도의 독립적 분석을 통해 기여형 연금인 CPP와 조세 기반 연금인 OAS의 상이한 재정 동학을 개별적으로 분석한다.

두 제도에 대한 재정분석은 독립적으로 수행되지만 공통 가정을 사용함으로써 정책 결정자는 공적연금체계 전체의 재정지표(GDP 대비 지출 비중, 기여율, 급여 수준 등)를 일관되게 평가할 수 있다.

## 나. 호주의 Intergenerational Report<sup>7)</sup>

호주 재무부(The Treasury)는 2002년 이후 총 6차례에 걸쳐 Intergenerational Report(이하 세대 간 보고서)를 발간해 오고 있다. 세대 간 보고서는 향후 40년간 호주의 경제 및 정부의 예산, 지출 등에 대한 전망 결과를 수록하고 있다. 정부 지출 전망은 보건, 장애, 장기요양,

---

7) 이하 내용은 Commonwealth of Australia(2023)을 참고했다.

공공부조, 연금제도, 교육·훈련 등 필수 공공서비스와 사회 인프라 및 국가 안보 등 총 9개 영역을 포괄한다.

호주의 세대 간 보고서의 독특한 특징은 세대 간 형평성 관점을 명시적으로 도입하여 정부 지출의 재정적 지속가능성을 평가한다는 데 있다. 사회보장제도를 비롯한 정부 재정 지출 전반을 통합적으로 분석하여, 현재 정책이 미래 세대에 전가하는 재정 부담을 정량화하고 이를 완화하기 위한 정책 선택지를 제시한다. 이를 통해 GDP 대비 지출 비중, 정책영역별 지출 구성 등의 재정지표가 장기적으로 어떻게 변화하는지를 보여준다.

#### 다. 영국 Office for Budget Responsibility의 재정전망

영국의 Office for Budget Responsibility(이하 OBR)는 정치적 독립성을 확보한 재정분석 기관으로서 2011년부터 공공재정 장기전망을 다룬 재정 지속가능성 보고서를, 2017년부터 재정위험보고서를 발간해 왔다. 2022년부터는 두 보고서를 통합한 Fiscal Risks and Sustainability(이하 FRS)를 매해 발간하고 있으며, FRS에서는 공공재정 장기전망과 주요 재정 위험 요인에 대한 전망을 수행하고 있다.<sup>8)</sup>

FRS의 특징은 인구고령화, 기후 변화, 국방 예산 증가 등 다양한 재정 위험요인을 종합적으로 고려한 50년 장기전망을 수행한다는 점이다. 공공지출 대비 GDP 비중, 국가부채 비율 등의 재정지표를 통해 재정 지속

8) FRS에서 검토한 연도별 공공재정 위험요인은 다음과 같다. 2023년 보고서에서는 팬데믹 이후에 확인되는 경제활동연령기 건강 관련 비경활자의 영향, 탄소중립(the net zero emissions)으로 인한 재정비용, 금리 변동의 영향을 다뤘다. 2024년 보고서에서는 기후 변화에 따른 잠재적 재정비용, 건강 지출 관련 재정분석, 이민으로 인한 재정분석을 다뤘다. 2025년 보고서에서는 연금 관련 재정분석, 공공 부문 재정 상태, 탄소중립목표 달성에 따른 재정 비용 등을 다뤘다(Office for Budget Responsibility 2023; 2024; 2025).

가능성을 평가하며, 정책 변화의 재정적 영향을 정량화한다.

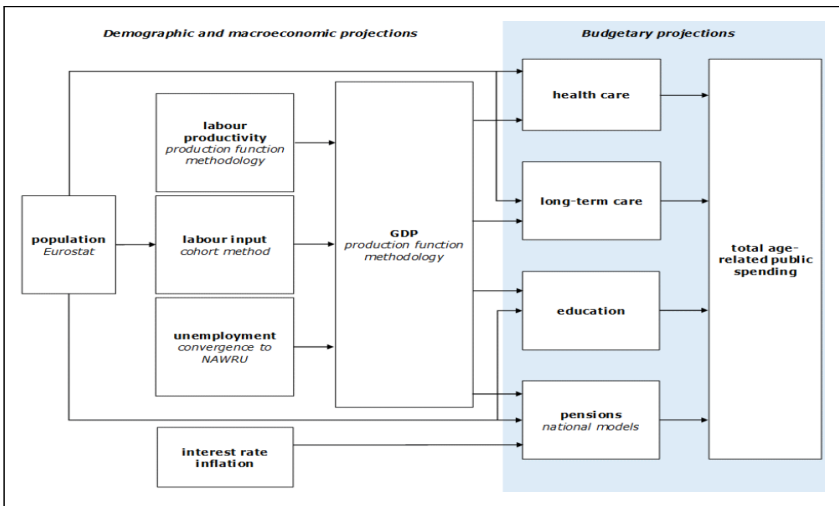
## 라. EU의 Ageing Report

EU 집행위원회(European Commission)와 경제정책위원회(Economic Policy Committee)가 공동으로 발간하는 Ageing Report는 통합적 사회보장 재정분석 모형이다(European Commission, 2024). 분석 대상 국가는 27개 EU 회원국과 노르웨이 등 총 28개국이며, 이들 국가를 대상으로 표준화된 방법론에 공통 가정을 적용하여 고령화와 관련한 사회보장 재정분석을 수행함으로써 국가 간 비교가능성을 담보한다. 이때 공통 가정의 핵심 요소는 다음과 같다. 첫째, Eurostat의 EUROPOP 2023 인구추계를 모든 국가에 동일하게 적용한다. 둘째, 잠재성장률은 생산함수 접근법(production function approach)을 통해 추정하며, 총요소생산성 증가율은 장기적으로 1.0%로 수렴한다고 가정한다. 셋째, 실업률은 구조적 실업률(NAWRU)로 점진적으로 수렴한다고 가정한다. 마지막으로 각국의 최신 정보와 법제화된 개혁을 반영한다.

Ageing Report의 핵심 강점은 고령화에 관련된 사회보장제도 간 상호작용을 명시적으로 고려하여 재정분석을 한다는 점이다. 주요 상호작용 메커니즘은 다음과 같다. 첫째, 연금제도와 노동시장 간 상호작용으로서는 연금 수급연령 상향이 고령자 고용률에 미치는 영향을 모형화하고, 이것이 다시 연금재정과 조세수입에 미치는 효과를 추정한다. 둘째, 건강보험과 장기요양보험제도 간 상호작용으로서 건강수명 연장이 장기요양 수요를 감소시키는 효과를 ‘건강한 고령화 healthy ageing’ 시나리오를 통해 정량화하여 재정분석을 수행한다. 가령, 2024년 보고서는 건강수명이 기대수명의 50% 속도로 증가할 경우 장기요양 지출이 2070년 GDP

대비 0.7%p 감소할 것으로 전망했다. 셋째, 교육과 생산성 간 상호작용의 경우, 교육 투자가 미래 생산성과 성장 잠재력을 증가시키는 효과가 있는 것을 가정하고 재정분석을 수행한다.

[그림 2-5] Ageing Report 재정분석 절차



출처: European Commission(2024, 2)의 Graph 1 인용.

Ageing Report의 재정분석은 인구 고령화의 시점, 경제적 영향 및 관련 예산에 대한 장기 전망 결과를 제공한다. Ageing Report는 3년을 주기로 발간되며, 기본 가정과 전망 방법론([그림 2-5] 참조)에 대해 합의한 후 연령별 지출 전망을 작성한다. 재정 분석은 공적연금, 보건의료, 장기요양, 교육에 대한 공공지출 등 4가지 예산 항목에 대해 수행된다. 개별 국가의 제도 특성을 반영하면서도 공통된 가정과 방법론을 기반으로 하기 때문에 국가 간 비교가능성이 확보되며, ‘정책 변경 없음(no-policy-change)’ 가정을 기반으로 현행 제도 유지 시 미래의 사회보장재정을 전망한다.

Ageing Report를 통해 표준화된 지표 체계는 국가 간 정책 학습을 촉진하고 제도 개선 방향을 도출하는 근거자료로 활용된다. 주요 지표로는 고령화 비용 지표(GDP 대비 연금, 보건의료, 장기요양, 교육, 실업급여 지출의 합), 지속가능성 격차(장기 재정건전성 확보를 위한 구조적 기초 재정수지 조정 규모), 재정 스트레스 지표(인구고령화로 인한 재정 압박 정도) 등이 있다.

Ageing Report의 사회보장 재정분석 결과는 EU의 경제정책 조율 메커니즘인 European Semester에 직접 활용된다. 즉, 매년 각국이 제출하는 안정화프로그램(Stability Programme) 또는 수렴프로그램(Convergence Programme)의 장기 재정전망은 Ageing Report의 방법론과 가정을 따라야 한다. 이에 집행위원회는 Ageing Report를 기반으로 각국의 중기재정계획을 평가하고, 회원국별 정책 권고안(Country-Specific Recommendations)을 도출한다. 예를 들어, 2024년 Ageing Report에서 고령화 비용 급증 국가로 전망된 룩셈부르크나 벨기에 등을 대상으로 한 주요 권고 사항은 연금 수급 연령과 기대수명의 자동 연계, 조기퇴직 경로 차단, 연금 크레딧 축소 등이었으며, 의료비 증가 우려 국가로 전망된 아일랜드나 몰타 등을 대상으로는 일차의료 강화, 예방 투자 확대, 의료 ICT 투자 등이었다.

### 3. 사회보장 재정분석의 주요 쟁점

이상에서 검토한 국내외 사회보장 재정분석 현황을 종합하면, 한국의 사회보장 재정분석은 급속한 제도 확장 과정에서 형성된 구조적 한계로 인해 개선이 필요한 상황이다. 특히 분절적 접근, 방법론적 차이, 통합적 관점 보완 필요 등의 과제는 1990년대 이후 각 제도가 개별적으로 발전

하면서 형성된 것으로 이해할 수 있다.

## 가. 분절적 분석 체계와 방법론적 불일치 문제

현재 한국의 사회보장 재정분석은 기관별, 제도별로 분절적으로 수행되고 있어 제도 간 상호작용과 총체적 재정 영향을 파악하기 어려운 상황이다(〈표 2-1〉 참조).

〈표 2-1〉 국내 사회보장 재정분석 사례 비교

구분	장기재정전망	NABO 장기재정전망	사회보장 재정추계	국민연금 재정계산	건강보험 재정전망
담당 부처	기획재정부	국회예산 정책처	사회보장 위원회	국민연금공단	보건복지부
관련 근거	국가재정법 제7조	국회법 제22조의2	사회보장 기본법 제5조 제4항	국민연금법 제4조	국민건강 보험법 제3조의2
주기	5년	2년	5년	5년	5년
분석 범위	국가재정전체, 의무지출 중심	국가재정전체	사회보장제도 전반	국민연금	건강보험 및 장기요양보험
시계	장기(40년)	장기(50년)	장기(약40년)	초장기(70년)	중단기(5-10년)
특징	재정건전성 관 점, GDP 대비 지출 비중 중심	전망 시점 기준 예산에 반영된 정책이나 제도 유지 가정	제도 지속가능 성 중심, 급여- 부담 균형 분석	보험수리적 균 형, 세대 간 형 평성 분석	의료비 증가 요 인 분석, 보험 료를 조정 시나 리오
한계	제도별 세부 분 석 부족, 급여 적절성 미고려	자체 경제지표 전망치 및 추계 모형 사용으로 외적 일관성 부 족	부처 소관 제도 중심, 타 부처 제도와 연계 부족	다층연금 체계 내 역할 분석 미흡, 타제도 미고려	인구학적 요인 의 장기 전망 불확실성, 의료 기술 변화 반영 곤란

출처: 기획재정부(2020), 국회예산정책처(2020), 보건복지부(2024a; 2024b), 국민연금재정추계전문위원회(2023)을 참고하여 저자 작성.

분석 대상 5개 사례의 추계 시계는 5년에서 70년까지 각기 다르며, 추계 범위도 총량-제도-특정 제도의 3계층 구조로 구분된다. 기획재정부와

NABO는 국가재정 총량을, 사회보장위원회는 사회보장제도 전반을, 개별 제도 분석은 특정 제도만을 다루어 분석 깊이와 정책 활용도 간 상충 관계가 발생한다.

또한 제도 간 상호작용을 충분히 고려하지 못하는 한계도 있다. 국민기초생활보장제도와 기초연금의 대체관계, 기초연금과 국민연금의 연계 구조, 현금급여와 사회서비스의 보완관계 등 제도 간 동태적 상호작용에 대한 분석이 보완된다면, 보다 정확한 재정 영향 파악이 가능할 것이다.

각 기관이 서로 다른 분석 방법론과 가정을 사용함으로써 동일 제도에 대한 전망 결과에 차이가 발생하기도 한다. 이는 각 기관의 분석 목적과 관점이 다르기 때문으로 이해할 수 있으나, 정책 결정 과정에서 일관된 증거를 제공하기 위해서는 방법론적 조율이 필요한 부분이다.

재정적 지속가능성 개념도 다층화되어 있어 조율이 필요한 상황이다. 기획재정부는 재정규율적 지속가능성을, 국회예산정책처는 거시경제적 지속가능성을, 사회보장위원회는 국제비교적 지속가능성을, 개별 제도 분석은 제도적/운영적 지속가능성을 각각 강조한다. 이처럼 상이한 지속가능성 개념은 정책 우선순위 설정과 자원배분에서 조율이 필요함을 시사한다.

## 나. 국제 비교의 한계와 표준화 부재

한국의 사회보장 재정분석은 국제 표준과의 정합성을 높일 필요가 있다. 각 기관의 전문성과 독립성을 존중하면서도 데이터 표준화와 방법론 일관성을 확보하는 균형점을 모색해야 하며, OECD 등 국제 표준과의 정합성 제고는 한국 사회보장제도의 국제 비교 가능성을 높이는 중요한 과제이다.

데이터 표준화와 분류체계의 일관성 확보는 정책적 일관성과 국제 비교 가능성을 높이는 데 기여할 수 있으며, 이는 사회보장제도의 발전과 국민의 복지 향상을 위한 기초가 될 것이다.

#### 다. 통합적 분석 체계 구축의 필요성과 방향

해외 선진사례는 이러한 한계를 극복할 수 있는 구체적인 방향을 제시한다.

캐나다 OCA는 다층연금체계 내에서 CPP와 OAS에 대해 각각 독립적인 재정전망을 수행하지만, 동일한 인구·경제 가정을 적용함으로써 분석의 정합성을 확보하고 있다. 이를 통해 정책 결정자는 공적연금체계 전체의 재정 부담 및 급여 수준을 일관되게 평가할 수 있다.

호주의 세대 간 보고서는 세대 간 형평성 관점을 명시적으로 도입하여 정부 지출의 재정적 지속가능성을 평가한다. 보건, 장애, 장기요양, 공공부조, 연금제도 등 9개 영역을 통합적으로 분석하여 현재 정책이 미래 세대에 전가하는 재정 부담을 정량화하고, 이를 완화하기 위한 정책 선택지를 제시한다.

영국 OBR은 정치적 독립성을 확보한 재정분석 기관으로서 공공재정 장기전망과 재정위험 분석을 통합한 FRS(Fiscal Risks and Sustainability)를 발간하고 있다. 특히 인구 고령화, 기후 변화, 국방 예산 증가 등 다양한 재정 위험 요인을 종합적으로 고려한 50년 장기전망을 수행한다.

EU의 Ageing Report는 가장 주목할 만한 사례로, 28개국을 대상으로 표준화된 방법론과 공통 가정을 적용하여 국가 간 비교 가능성을 담보하면서도 제도 간 상호작용을 명시적으로 고려한 통합적 분석을 수행하

고 있다. 연금제도와 노동시장 간 상호작용, 건강보험과 장기요양보험제도 간 상호작용, 교육과 생산성 간 상호작용 등을 모형화하여 재정분석에 반영하고 있으며, 분석 결과는 EU의 경제정책 조율 메커니즘인 European Semester에 직접 활용되어 정책 실효성을 확보하고 있다.

현행 한국 사회보장 재정분석의 가장 큰 제약과 이로 인한 과제는 기관별 분절적 접근에서 벗어난 통합적 체계 구축이다. 이를 위해서는 다음과 같은 단계적 개선이 필요하다.

첫째, 통합적 분류체계 개발이 우선되어야 한다. 상호복합적으로 연계되어 있는 각종 사회보장제도를 배타적으로 분류할 수 있는 분류틀의 개발이 필요하며, 이 과정에서 제도의 재원 조달 방식이나 급여 방식 등에 대한 분류법이 예각화되어야 한다. 개발된 제도 분류 틀을 통해 제도 간 상호작용 효과를 포착할 수 있는 정량적 측정 방법론에 대한 연구가 동반되어야 한다.

둘째, 국제 표준과의 정합성 제고를 통해 보다 정확하고 신뢰할 수 있는 재정분석 체계를 구축해야 한다. OECD, EU 등 국제기구와의 협력을 강화하여 방법론적 선진화를 추진하고, 국제 비교 가능성을 확보하기 위해 OECD SOCX, EU ESSPROS 등 국제 분류체계와의 정합성을 높여야 한다.

셋째, 법적·제도적 기반 구축이 필요하다. 「국가재정법」, 「사회보장기본법」 등 관련 법령을 개정하여 통합적 장기재정분석의 법적 근거를 명확히 하고, 분석 주기, 시계, 방법론 등을 법령에 구체적으로 명시하여 경제나 정치 상황에 관계없이 주기적으로 일관된 분석이 이루어지도록 보장해야 한다.

넷째, 통합 거버넌스 체계 구축이 필요하다. 사회보장위원회를 중심으로 한 범정부 차원의 통합 재정분석 체계를 구축하여, 해외 사례와 같이 각 부처와 기관의 개별 분석을 조율하고 통합하는 명확한 거버넌스 구조

를 마련해야 한다. 분석 결과의 국회 보고 의무화, 정책 반영 절차 등을 법제화하여 실효성을 확보해야 한다.

이상의 4가지 개선 방향을 실현하기 위해서는 무엇보다 첫 번째 단계인 통합적 분류체계 개발이 체계적으로 이루어져야 한다. 해외 사례에서 확인되듯이, 효과적인 통합적 재정분석은 모두 견고한 이론적 기반 위에 구축된 분류체계에서 출발한다.

따라서 사회보장제도의 분류에 대한 이론적 발전 과정과 현재 국제적으로 활용되고 있는 주요 분류체계들에 대한 체계적 검토가 선행되어야 한다. 특히 제도유형별 접근과 정책영역별 접근을 결합한 다차원적 분석틀을 구축하기 위해서는 기존 분류 이론의 한계와 개선 방향을 명확히 파악하고, 이를 바탕으로 한국적 상황에 적합한 새로운 분류체계와 분석 방법론을 개발해야 한다.

특히 통합적 재정분석 체계 구축에서 핵심은 한국 사회보장제도의 현재 위치를 정확히 파악할 수 있는 재정지표 탐색이다. “한국은 어디에 있는가?”라는 질문에 답하기 위해서는 OECD 국가들과 비교 가능한 표준화된 재정지표가 필요하며, 이는 단순한 지출 규모뿐만 아니라 제도의 구성, 자원 구조, 급여 형태 등 다차원적 측면을 포괄해야 한다. 기존 재정분석이 주로 정량적 지표(지출 규모, GDP 대비 비중 등)에 집중한 반면, 제도의 성격을 결정하는 정성적 특성(제도 유형, 적용 대상, 급여 산정 방식 등)은 체계적으로 분석되지 못했다. 이러한 정량·정성 정보의 통합은 한국의 상대적 위치를 정확히 진단하고, 지향해야 할 정책 방향을 설정하는 데 필수적이다. 따라서 다음 절에서는 국제적으로 활용되는 주요 사회보장 재정 분류 체계를 검토하여, 이들이 제공하는 재정지표의 종류와 특성을 파악하고, 정량·정성 통합 재정지표 체계 구축의 방향을 도출하고자 한다.

## 제2절 사회보장 재정분석을 위한 분류체계 검토

사회보장 재정지표 탐색을 위해서는 제도의 본질적 특성을 반영한 분류체계의 구축이 선행되어야 한다. 이를 위해서는 상술한 바와 같이 기존의 단편적·개별적 접근에서 벗어나 제도 간 구조적 차이와 상호작용을 체계적으로 분석할 수 있는 이론적 틀의 정립이 필요하다. 본 절에서는 사회보장제도 분류의 이론적 발전과정과 국제적으로 활용되는 사회지출 분류체계를 체계적으로 검토하여, 이들이 제공하는 재정지표의 종류와 특성을 파악하고, 사회보장 재정지표 체계의 개선 방안을 제시하였다.

### 1. 사회보장제도 분류의 이론적 기초와 발전

사회보장제도 분류 체계는 전통적 분류체계와 현대적 분류체계의 관점에서 정리할 수 있다. 전통적 분류체계가 주로 단일 차원에서 제도의 구조적 특성과 급여 방식에 초점을 맞춘 이분법적 구분에서 기초적 분류틀을 제공했다면(Titmuss, 1958; Baldwin, 1990), 현대적 분류 체계는 개별 국가에 시행되고 있는 다양한 사회보장제도를 복지국가의 기능적 특성과 정치경제적 맥락을 통합한 일련의 체제로서 이해하는 분류틀을 제공한다(Esping-Andersen, 1990; Bonoli, 1997). 이러한 이론적 진화는 복지국가 연구를 단순한 총량 차원의 지출 비교에서 제도적 배치와 결과를 포괄하는 체계적 비교 분석으로 전환시켰다. 그리고 사회보장제도 분류법의 발전은 방법론적 정교화와 새로운 사회적 위험에 대한 대응을 통해 사회보장 재정분석에 대한 이론적 토대를 구축하였다(Arts & Gelissen, 2002; Hemerijck, 2013). 여기에서는 사회보장제도의 전통적 분류체계와 현대적 분류체계를 개괄적으로 살펴보았다.

### 가. 사회보장제도의 전통적 분류체계

사회보장제도에 대한 전통적 분류 방식은 사회연대적 차원에서 수급권 부여 원리와 그에 따른 제도의 목표를 기준으로 채택하고 있다. 전통적 사회보장제도 분류의 핵심은 베버리지(Beveridge) 방식 대 비스마르크(Bismarck) 방식의 이분법적 구분이다. 전자는 1942년 영국의 사회 재건 과정에서 형성되어 시민권 기반의 보편적 보장, 정액 혹은 균등 급여(flat-rate benefit), 조세 재원을 통한 재분배적 접근을 추구한다(Beveridge, 1942; Marshall, 1950). 후자는 1880년대 독일에서 사회주의 압력에 대한 정치적 대응책으로 등장하여 고용 기반 사회보험, 기여-급여 연계, 직업별 계층화를 특징으로 하는 보수적 접근을 핵심 특징으로 한다(Manow, 2001). 이러한 분류 방식이 갖는 이론적 함의는 사회연대의 서로 다른 개념화에서 찾을 수 있다. 전자가 시민권을 기반으로 한 사회민주적 연대를 구현한다면, 후자는 노동 계급과 지위를 유지하면서 사회적 안전을 제공하는 보수적 연대를 구현하기 때문이다. 하지만 현실의 사회보장제도는 개별 방식의 순수한 형태보다는 두 방식의 요소가 혼재된 특징을 보여준다는 데서 현실의 복잡성을 과도하게 단순화한다는 한계를 내재한다(Palier, 2010).

또 다른 분류는 보편주의와 선별주의와 같은 사회정책의 철학에 기초한 구분이 있다. 보편주의는 사회적 시민권 이론에 기반하여 ‘전체 시민에 대한 권리로서의 급여’를 강조한다(Marshall, 1950). 이는 행정적 효율성, 정치적 지속가능성, 사회통합 효과를 통해 빈곤 함정과 낙인화를 방지하는 장점을 가진다(Korpi & Palme, 1998). 선별주의는 대상 선정 과정에서의 표적화를 통한 효율성과 비용 효과성을 추구한다. 하지만 이에 대한 반대급부로 복잡한 자산조사와 사회적 배제 위험을 수반한다

(van Oorschot, 2002). 전통적 접근에서는 보편주의와 선별주의를 상호 배타적인 개념으로 이해하였으나, 현대적 접근에서는 보편적 포괄성과 욕구별 차등화를 동시에 추구하는 ‘차별화된 보편주의(differentiated universalism)’로 개념을 통합하는 방향으로 발전하고 있다(Anttonen & Sipilä, 1996; Palier, 2010). 또한, Korpi & Palme(1998)는 실증분석 결과, 보편주의적 제도가 오히려 더 큰 재분배 효과를 갖는다는 발견을 통해 보편주의와 선별주의에 대한 직관적 이해에 대한 반론을 제기했다. 즉, 보편주의 제도가 중산층의 정치적 지지를 확보함으로써 전체적으로 더 관대한 급여 수준을 달성할 수 있는 “재분배의 역설(paradox of redistribution)”을 가진다는 것이다. 이러한 발견은 단순한 이분법적 분류의 한계를 명확히 보여주었다.

전통적인 사회보장제도 분류체계는 복지국가 논의 과정에서 예각화되었다. 특히, Titmuss의 분류체계는 사회보장제도 분류 체계에 대한 근본적인 전환을 야기하였다. Titmuss는 사회정책의 근본 논리를 분석해 시장과 가족의 실패 시에만 국가가 개입하는 잔여주의 모형과 성과와 공로에 따른 배분으로 국가가 개입하는 산업적 성취 모형, 보편적 시민권에 기반해 재분배를 하는 제도적 재분배 모형으로 사회보장제도를 분류하였다(Titmuss, 1958; Rein, 1982). Titmuss의 분류체계는 ‘사회복지의 빙산’ 개념을 통해 정립되었다. 직접적인 복지 공급 외에도 조세 지출과 같은 재정 복지와 시장에서 민간 기업 차원으로 제공되는 기업 복지를 포함한 포괄적 복지 개념을 제시하여 사회보장제도 분류체계의 범위를 확장시킨 것이다. 이는 Esping-Andersen(1990)의 복지체제론에 직접적인 영향을 미쳤을 뿐만 아니라 향후 사회보장제도 분류체계의 이론적 토대가 되었다.

## 나. 분류 체계의 발전과 방법론적 진화

사회보장제도의 현대적 분류체계는 복지국가 체제 논의와 그 결을 같이 한다. Esping-Andersen이 1990년 발간한 『복지 자본주의의 세 가지 세계(The Three Worlds of Welfare Capitalism)』(Esping Andersen, 1990)는 지출 중심 분석에서 제도적 배치 분석으로의 근본적 전환을 가져왔다. ‘탈상품화(decommodification)’와 ‘계층화(stratification)’ 개념을 통해 자유주의, 보수주의, 사회민주주의 3가지 체제로 복지국가를 유형화한 이 이론은 복지국가를 일관된 제도적 배치로 이해하는 새로운 관점을 제시하였다(Goodin et al., 1999). 이후, Esping-Andersen은 젠더 비판에 대한 응답으로 ‘탈가족화(defamiliaisation)’ 개념을 추가하여 여성의 독립을 지원하는 복지국가 역할을 분석에 포함시켰다(Esping-Andersen, 1999). 그리고 Esping-Andersen의 복지체제론은 사회보장제도를 통해 복지국가를 설명하려는 수많은 연구자에게 지대한 영향을 미쳤으며 사회보장제도 분류체계의 이론적 진화 및 방법론적 발전의 변곡점이 되었다.

Bonoli(1997)는 전통적 분류체계를 활용하여 현대적 분류체계의 집합적 결과물인 복지체제를 실증하였다. Bonoli는 베버리지형과 비스마르크형으로 구분되는 역사적 제도 유산을 한 축으로 하고, 사회지출에 대한 국가 개입의 수준, 즉 GDP 대비 사회지출 비중(저지출-고지출)을 한 축으로 하는 이차원 분류체계를 제시하였다. 이를 통해 저지출-비버리지형의 자유주의 체제와 고지출-비스마르크형의 보수주의 체제, 고지출-비버리지형의 사회민주주의체제, 저지출-비스마르크형의 남유럽 체제로 4가지 유형의 복지체제를 도출하였다. 또한, 경로의존적 제도 발전과 정책 일관성을 동시에 고려하여 유럽 복지국가의 수렴 및 발산 과정 분석에 효

과적으로 적용되었다(Leibfried, 1992). 이러한 분류체계는 전통적인 이분법적 사회보장 분류체계를 연계한 다차원적 분석을 통해 다양한 사회보장제도로 구성된 복지국가의 복잡성을 포착하면서도 개념적 명확성이 유지된다는 점에서 중요한 함의를 가진다(Powell & Barrientos, 2004).

Ferrera는 탈상품화와 계층화만으로는 사회보장제도를 온전히 분류할 수 없음을 실증하였다(Ferrera, 1996). Ferrera의 비판은 민주주의 이행 과정이 복지국가 발전에 미치는 영향을 부각시키고 유럽 통합이 국가별 복지국가 발전에 미치는 효과가 차별적이라는 것이었다(Guillén, 2010; Moreno, 2006). 이를 위해 이탈리아, 스페인, 포르투갈, 그리스 등 남유럽 국가 사례에서 확인되는 공통적 제도적 특성을 규명하여 기존 복지체제 분류의 한계를 보완하였다. 이들 국가에서 확인되는 차별점은 제도적 분절화, 후견주의적 관행, 가족주의적 복지 제공이라는 특징을 통해 현금 중심 제도, 보편적 의료서비스, 취약한 사회부조, 지역적 편차로 기존 사회보장제도 분류체계에서 포착되지 않는 정책적 특성이었다(Rhodes, 1997). 즉, Ferrera의 남유럽 모델은 지역적 맥락과 제도적 특수성의 중요성을 인식시켜 복지국가 이론의 글로벌 적용 가능성에 대한 성찰을 제공하였으며(Holliday, 2000), 이는 이후 동아시아 복지국가들의 복지체제 논의로까지 영향을 미쳤다(한신실, 2020).

한편, Korpi와 Palme(1998)는 베버리지형과 선별주의에 입각해 ‘빈곤층을 표적화할수록 빈곤과 불평등 감소 효과가 줄어든다’는 역설적 명제를 실증한 ‘재분배의 역설’을 통해 사회보장제도의 재분배 기제를 재분류하였다. 이때 분류된 사회보장제도 분류는 기초보장형(Basic security), 조합주의형(Corporatist), 포괄주의형(Encompassing), 선별주의형(Targeted), 자발적 국가보조형(Voluntary state-subsidized) 등 총 5가지 이념형으로 분류되었으며, 이중 보편적 소득비례제도가 가

장 효과적인 재분배를 달성하였음을 실증하였다. Korpi와 Palme는 이러한 결과의 핵심이 연합정치 메커니즘에 있다고 보았다. 보편적 제도는 중산층 지지를 확보하여 더 관대한 예산과 강력한 정치적 지속가능성을 얻으며, 이것이 결과적으로 더 큰 재분배 효과를 가져올 수 있기 때문이다. 이는 전통적인 분류체계에서 선별주의가 표방하는 표적화 효율성 논리에 반하는 발견으로 제도 설계가 재분배 결과에 미치는 결정적 영향을 입증하였다(Brady, 2009).

특히, 사회보장제도 분류체계에 대한 방법론적 발전은 분석자료와 분석 방법 등 다양한 층위에서 진행되었다. Scruggs & Allan(2006)는 Esping-Andersen의 연구를 비판적으로 검토하는 과정에서 원 탈상품화 지수에서 계산 오류를 발견하였다. 그리고 이러한 오류를 수정하면, 기존의 세 개로 유형화된 복지체제의 군집이 유효하지 않을 수 있음을 실증하였다. Arts & Gelissen(2002)은 실제 복지국가들이 일관된 성격의 사회보장제도로 구성된 순수 사례가 아니라 다양한 성격의 사회보장제도가 혼합된 사례임을 주장하면서 이산적이면서 상호 배타적인 유형 분류는 근본적으로 한계를 내포하고 있음을 주장한 것이다. 이러한 복지국가 유형론, 즉 사회보장제도 분류체계의 결정적 한계로 형식적 이론화 부족, 분류 타당성 문제, 방법론적 비일관성 등이 지적되었다(Powell & Barrientos, 2004). 이는 사회보장제도의 유형화를 통한 복지국가 논의에 있어서 방법론적 정교화의 필요성을 여실히 보여주는 사례이다.

이처럼 복지국가를 유형화하고 분류하기 위한 연구는 사회보장제도를 분류하는 과정에서 여러 기술적 문제를 해결해야 하는 과업을 가진다. 주요 쟁점은 분석 자료나 지표의 선택, 그리고 지표 산정 과정에서의 가중치 부여의 자의성, 분석 시점에 대한 안정성과 문화적 동등성, 행정적 차이와 조작정 정의의 비일관성 등과 관련된 종속변수의 문제가 해당한다

(Green-Pedersen, 2004). 또한, 사회보장제도 분류체계의 접근법이 양적 방법론과 질적 방법론으로 이분되어 있는 상황도 비판의 대상이 되었다(Clasen & Siegel, 2007). 즉, 양적 접근법은 체계적인 비교 가능성을 제공하지만 제도의 특성이 배제되며, 질적 접근법은 사회보장제도에 대해 풍부한 기술을 통해 높은 설명력을 지니지만 일반화 가능성과 비교 가능성 측면에서 제약이 있다는 것이다.

## 2. 사회보장 재정분석을 위한 분류 체계

사회지출 분류체계는 상술한 복지국가 연구의 핵심 방법론적 기반을 제공해 왔다. 지난 수십 년간 OECD, EU, ILO 등 국제기구들은 각자의 목적에 부합하는 사회보장 재정분류체계를 발전시켜 왔으며, 이는 국가 간 비교연구의 기초를 제공했다(Adema & Whiteford, 2010; Castles, 2004). 이에 여기에서는 OECD의 사회지출데이터베이스(SOCX)와 EU의 ESSPROS, MISSOC, ILO의 Social Security Inquiry, ISSA의 Country Profiles에 대해 살펴보고, 사회보장 장기재정분석을 위한 기존 사회보장 재정분류체계의 한계를 도출하고 개선 방안을 모색하였다.

### 가. OECD의 Social Expenditure Database<sup>9)</sup>

OECD 사회지출데이터베이스(Social Expenditure Database, 이하 SOCX)는 1996년 최초 구축된 이후 현재까지 국가 간 사회정책 비교연구의 핵심 도구로 기능을 수행해 왔다. 초기 6개국 순공공사회지출 추계에서 시작된 SOCX는 현재 36개 OECD 회원국 및 일부 가입후보국을 포

---

9) 여기에서 소개되지 않은 보다 구체적인 내용은 OECD(2019)를 참조하면 된다.

팔하는 포괄적 사회지출 분류체계로 확장되었다. SOCX는 정기적인 자료 최신화 및 수정 작업을 통해 최신 동향을 반영하고 있다.

SOCX 구축의 근본 목적은 OECD 국가 간 사회정책 지출의 추세와 구성 변화를 세부 프로그램 수준에서 분석할 수 있는 신뢰성 있고 국제적으로 비교 가능한 통계를 제공하는 것이다. 이를 위해 OECD는 국민계정체계(System of National Accounts 2008)와 완전한 호환성을 확보하였으며, 건강계정체계(System of Health Accounts)와도 통합적 연계를 유지하고 있다.

SOCX는 사회정책의 기능, 제공 주체, 급여 형태 등 3가지 차원으로 사회보장 재정을 분류하고 있다.

첫 번째 차원에서는 사회정책의 기능을 사회적 위험과 욕구에 따라 9개 영역으로 분류하고 있다(〈표 2-2〉 참조). 노령(old age), 유족(survivors), 근로무능력(incapacity-related benefits), 보건(health), 가족(family), 적극적노동시장정책(active labour market programmes), 실업(unemployment), 주거(housing), 기타 사회정책영역(other social policy areas)으로 구분되는 이 체계는 영역별로 제공 주체와 급여 형태에 따른 세분화를 통해 다차원적 분석을 가능하게 한다.

두 번째 차원에서는 사회정책의 제공 주체에 따라 사회보장 재정지출을 공공사회지출(Public social expenditure), 의무민간사회지출(Mandatory private social expenditure), 자발적민간사회지출(Voluntary private social expenditure)로 구분한다. 공공사회지출은 정부가 직접 제공하거나 통제하는 사회보장급여가 해당한다. 의무민간사회지출은 법적 의무에 의한 민간 사회보장급여가 해당한다. 자발적 민간사회지출은 민간에서 자발적으로 제공하는 사회보장 목적의 급여가 해당한다. 이는 복지국가의 공공-민간 역할 분담, 즉 복지혼합을 분석하는 중

요한 기준이 된다. 세 번째 차원에서는 사회정책을 급여 형태에 따라 구분한다. 급여 형태는 현금(cash benefits)과 현물(benefits in kind)로 구분된다. 이는 사회보장급여의 경제적 효과와 재분배 메커니즘 분석에 활용된다.

〈표 2-2〉 SOCX 사회정책 기능별 분류

기능	관련제도
노령	연금급여, 조기퇴직급여, 재가급여, 기타 현금 및 현물 급여
유족	유족연금, 사망일시금, 장제비, 기타 현금 및 현물 급여
근로무능력	장애급여, 업무상 재해급여, 상병급여, 재가급여, 재활급여, 기타 현금 및 현물 급여
보건	의료서비스, 의료용품 및 예방적 돌봄 용품 등
가족	가족수당, 육아휴직급여, 보육서비스, 기타 현금 및 현물 급여
적극적노동시장	공공고용서비스, 직업훈련, 직접적 일자리 창출 관련 서비스, 고용장려금, 창업지원금 등
실업	실업급여, 조기퇴직 및 실업시 소득지원 급여 등
주거	주거급여 및 기타 주거지원
기타	타 영역에 분류되지 않은 사회부조 및 기타 사회보장제도

출처: OECD(2019, 16)의 Table 2 재구성.

이러한 3차원 분류체계는 사회보장 재정지표 탐색에 필수적인 다층적 지표 체계를 제공한다. SOCX가 제공하는 주요 재정지표는 다음과 같다. 첫째, 지출 수준 지표로서 GDP 대비 사회지출 비중, 1인당 사회지출(USD PPP 기준), 정부지출 대비 사회지출 비중 등을 통해 국가별 사회보장 지출의 절대적·상대적 규모를 파악할 수 있다. 둘째, 재원 구성 지표로서 공공·민간 비중, 의무적·자발적 민간 비중 등을 통해 사회보장 재원조달의 공공·민간 역할 분담 구조를 분석할 수 있다. 셋째, 급여 구성 지표로서 현금급여와 현물급여의 비중을 통해 사회보장제도의 급여 제공 방식과 재분배 메커니즘을 파악할 수 있다. 넷째, 정책영역별 지출 구성 지

표로서 9개 정책영역별 지출 비중을 통해 국가별 사회보장 우선순위와 제도 성숙도를 비교할 수 있다. 다섯째, 성장 추세 지표로서 시계열 변화율, 연평균 증가율 등을 통해 사회보장 지출의 확대 속도와 방향을 분석할 수 있다. 이러한 다차원 재정지표는 한국의 현재 위치를 OECD 국가들과 비교하고, 지향해야 할 정책 방향을 설정하는 데 핵심적인 준거들을 제공한다.

SOCX의 또 다른 중요한 방법론적 기여는 조세제도의 영향을 반영한 순사회지출 개념의 도입이다. 이는 국가별로 상이한 조세-급여 연계구조를 고려하여 실질적 사회지출 수준을 정확히 측정하고자 하는 시도이다. 순총사회지출은 13단계의 복잡한 계산과정을 거쳐 산출되는데, 총직접공공사회지출에서 공공현금급여 관련 직접세와 사회기여금을 차감하고, 간접세 효과를 고려한 후, 사회목적 조세지출을 가산하는 방식으로 계산된다. 이때, 사회목적 조세지출은 수입포기법을 통해 평가되며, 현금급여 유사 조세지출과 현행민간급여 관련 조세지출로 구분된다. 이러한 정교한 계산을 통해 미국과 같이 총공공사회지출은 OECD 평균을 하회하지만 조세지출을 포함하면 중위 수준에 근접하는 국가의 실질적 사회보장 수준을 정확히 파악할 수 있다. 국민계정 기준과의 정합성 확보를 통해 거시경제 분석과의 일관성을 보장하며, 사회정책영역과 제공 주체를 축으로 하는 9×3 매트릭스 구조를 통한 교차분석으로 복지국가 유형론, 공공-민간 역할 분담, 기능별 지출 추세 등 다차원적 정책 분석을 지원한다.

다만, 이러한 여러 장점에도 불구하고 SOCX는 다음과 같은 한계를 가지고 있다. 첫째, SOCX는 사회지출 정보에만 집중하고 있어 사회보장제도의 재원조달 방식에 대한 체계적 정보를 제공하지 않는다는 한계가 있다. 이는 특히 사회보장제도의 재정지속가능성 분석 시 중대한 제약요인으로 작용한다. 둘째, 총량적 지출 정보 중심으로 구축되어 개별 급여의

적절성이나 소득대체율 등 급여 내용의 질적 측면에 대한 분석 기능이 제한적이다. 따라서 사회지출의 양적 증가가 반드시 사회보장제도의 질적인 개선을 의미하지 않는다는 점은 결정적인 한계로 남을 수밖에 없다. 셋째, 개별 제도의 독립적 분류에 치중하여 제도 간 상호작용 효과나 정책조합의 총체적 효과 분석이 어렵다는 한계를 내포한다. 예를 들어, 연금제도와 노인장기요양제도, 또는 실업급여와 적극적 노동시장정책 간의 상호보완 또는 대체관계에 대한 분석은 현재의 분류체계 하에서는 제약이 있을 수밖에 없다.

## 나. EU의 ESSPROS 및 MISSOC

### 1) ESSPROS<sup>10)</sup>

유럽 통합사회보장통계 체계(European System of integrated Social Protection Statistics, 이하 ESSPROS)는 1981년 도입된 이래 EU 사회정책 조정의 핵심 통계 기반으로 발전해 왔다. ESSPROS는 단순한 통계 체계를 넘어서 유럽 통합 과정에서 회원국 간 사회정책 수렴을 위한 정책 조정 도구로 기능하며, European Semester, Ageing Report, Social Protection Committee의 연례 보고서 등 EU 거버넌스와 직접 연계된다.

ESSPROS의 가장 중요한 특징은 EU 회원국의 사회보장제도에 대한 지출 및 재원 조달과 관련된 정보를 통합적으로 제공한다는 점이다. 이는 OECD SOCX가 지출 정보에 집중하는 것과 대조적으로, 사회보장제도의 재정구조와 지속가능성을 종합적으로 분석할 수 있는 정보를 제공한

---

10) 여기에서 소개되지 않은 보다 구체적인 내용은 EUROSTAT(2016)을 참조하면 된다.

다는 데서 강점을 지닌다. 따라서 ESSPROS의 통계단위는 사회보장제도 (Social Protection)로 정의되는데, 이때 사회보장제도는 “공공 또는 민간 기관이 상호 또는 개별 계약이 존재하지 않는 상황에서 가구 및 개인이 정의된 위험 또는 필요의 부담을 완화하기 위해 실시하는 모든 개입”으로 규정된다. 또한, 사회보장제도는 하나 또는 그 이상의 기관 단위가 지원하는 별개의 규칙 체계로서 사회보장급여의 제공과 그 재원 조달을 관장하여야 한다. 따라서 사회보장제도는 항상 별도의 수입-지출 계정을 작성할 수 있어야 하며, 가급적 단일한 위험이나 욕구에 대한 보호를 제공하고 특정한 수급자 집단을 대상으로 해야 한다. 그리고 ESSPROS는 의사결정권자, 정부의 가입 강제성, 기여 조건부 수급권, 보장 범위, 보호 수준의 5개 기준에 따라 개별사회보장제도의 특성을 정의한다. 이러한 다차원적 특성 정의를 통해 EU 회원국의 다양한 사회보장제도를 체계적으로 비교 분석할 수 있는 기반을 마련한 것이다.

ESSPROS도 SOCX와 유사하게 사회보장제도를 질병/의료(sickness/health care), 장애(disability), 노령(old age), 유족(survivors), 가족/아동(family/childrenn), 실업(unemployment), 주거(housing), 기타 분류되지 않은 사회적 배제(social exclusion) 등 사회적 위험과 욕구에 따른 8개 기능으로 분류하였다(〈표 2-3〉 참조).

〈표 2-3〉 ESSPROS의 사회보장제도 기능별 정의

기능	설명
질병/의료	- 신체적 또는 정신적 질병과 관련된 소득 유지 및 현금 지원(장애 제외) - 보호 대상자의 건강을 유지, 회복 또는 개선하기 위한 의료 서비스(질병의 원인과 무관)
장애	- 신체적 또는 정신적 장애로 인해 경제적 및 사회적 활동에 참여할 수 없는 사람들에 대한 소득 유지 및 현금 또는 현물 지원(의료 서비스 제외).
노령	- 노령과 관련된 현금 또는 현물 지원(의료 서비스 제외).
유족	- 가족 구성원의 사망과 관련된 현금 또는 현물 지원.
가족/아동	- 임신, 출산, 입양, 자녀 양육 및 기타 가족 구성원 돌봄 비용과 관련된 현금 또는 현물 지원(의료 서비스 제외).
실업	- 실업과 관련된 현금 또는 현물로 제공되는 소득 유지 및 지원.
주거	- 주택 비용 지원.
사회적 배제	- 기타 기능에 해당되지 않는 사회적 배제를 방지하기 위해 특별히 마련된 현 금 또는 현물로 제공되는 혜택(의료 서비스 제외).

출처: EUROSTAT(2016, 36)의 Table F 인용.

ESSPROS의 기능별 분류는 사회보장의 목적과 대상에 따른 명확한 기준을 제시하여 회원국 간 일관성 있는 비교 분석을 가능하게 한다. 차별점이 있다면 사회적 배제 기능과 같이 전통적 사회보장제도로 포괄되지 않는 새로운 사회적 위험에 대한 대응을 포함하여 현대적 사회보장 개념을 반영한다는 점이다. 그리고 각 기능은 현금급여와 현물급여로 세분화되어 급여 형태별 분석을 지원한다. 예를 들어 질병·의료 기능에서는 상병급여, 유급병가 등의 현금급여와 외래의료, 입원치료 등의 현물급여가 구분되어 집계된다.

타 사회보장 재정분류 체계와 비교할 때, ESSPROS가 갖는 추가적인 장점은 사회보장제도의 재원 조달 정보를 체계적으로 분류하고 있다는 점이다. 재원 조달은 크게 사회기여금(social contributions), 정부기여금(general government contribution), 타제도 이전수입(transfers from other schemes), 기타 수입(other receipts)으로 구분되며, 각각

에 대해 세부적인 하위 분류를 제공한다(〈표 2-4〉 참조). 이러한 재원 조달 정보는 사회보장제도의 재정 구조를 이해하고 지속가능성을 평가하는데 필수적이다. 특히 인구 고령화에 따른 사회보험료 부담 증가, 정부 재정 압박 등의 문제를 분석하기 위해서는 재원 조달원별 구성과 변화 추이에 대한 정보가 중요한 역할을 한다.

〈표 2-4〉 ESSPROS의 재원 조달 분류체계

대분류	중분류	소분류
사회기여금(1)	고용주의 사회기여금(11)	실제 고용주 기여금(111)
		귀속 고용주 기여금(112)
	피보험자의 기여금(12)	직장가입자 기여금(121)
		지역가입자 기여금(122)
		연금수급자 및 기타(123)
일반 정부기여금(2)	목적세(21)	
	일반조세(22)	
타제도 이전수입(3)	타제도로부터의 재이전된 기여금(31)	
	타제도로부터의 기타 이전금(33)	
기타수입(4)	자본수입(41)	
	기타(42)	

출처: EUROSTAT(2016, 25) Table C 인용.

이처럼 사회보장제도의 지출과 재원 조달 정보를 균형있게 제공하여 제도별 재정 구조를 체계적으로 분석 및 전망할 수 있다는 점은 ESSPROS만의 강점이다. 이는 인구 고령화에 따른 재정 압박이 예측되는 상황에서 고용주 기여금, 피보험자 기여금, 정부 기여금 등의 구성비와 변화 추이를 통해 사회보장제도의 재정적 지속가능성을 평가할 수 있다는 점에서 특히 중요한 역할을 수행한다. 즉, 제도별 재정 균형과 지속가능성을 분석하는 기초가 되며, EU 차원의 정책 조정에서 중요한 근거자료로 활용될 수 있다.

반대로 ESSPROS가 EU 회원국만을 대상으로 하는 분류체계이기 때문에 확장성 측면에 극복하기 어려운 한계를 내포한다. 이외에도 OECD SOCX와 달리 조세제도의 영향을 체계적으로 고려하지 않았을 뿐만 아니라 EU 회원국 간에도 조세-급여 연계 구조가 상이함에도 이를 반영하지 않아 사회보장 재정분석의 정확성을 제약하는 한계를 지닌다.

## 2) MISSOC<sup>11)</sup>

유럽 상호사회보장정보시스템(Mutual Information System on Social Protection, 이하 MISSOC)은 1990년 유럽연합 초기 회원국들의 사회보장 정보에 대한 상호 교류를 촉진하기 위해 설립된 정보 공유 플랫폼이다. 현재 EU 27개 회원국뿐만 아니라 유럽경제지역(EEA) 소속 아이슬란드, 리히텐슈타인, 노르웨이와 스위스 등 총 31개국의 사회보장 제도에 대한 상세 정보를 제공하고 있다.

MISSOC는 ESSPROS와 달리 정량적 지출 데이터보다는 각국 사회보장제도의 구조적 특성과 운영 방식에 대한 정성적 정보 제공에 특화되어 있다. 각국 정부가 직접 정보를 제공하고 검토함으로써 공식성과 신뢰성을 확보하고 있으며, 비교표(Comparative Tables) 기능을 통해 여러 국가의 사회보장제도를 동일한 구조하에서 비교 분석할 수 있는 도구를 제공한다.

MISSOC의 분류체계는 사회보장제도의 철학적 기반에서부터 구체적인 운영 방식까지를 포괄한다. 철학적 기반으로는 거주 기반(Beveridge형)과 직업 기반(Bismarck형)으로, 재정 구조로는 조세 기반과 사회보험 기반으로, 그리고 관리 체계로는 중앙집중형과 지방 분산형으로 분류한다.

11) 여기에서 소개되지 않은 보다 구체적인 내용은 EUROSTAT(2016)을 참조하면 된다.

〈표 2-5〉 MISSOC 제도별 카테고리

구분	분류 카테고리
1	Financing
2	Health Care
3	Sickness - Cash Benefits
4	Maternity/Paternity
5	Invalidity
6	Old-Age
7	Survivors
8	Accidents at Work and Occupational Diseases
9	Family Benefits
10	Unemployment
11	Guaranteed Minimum Resources
12	Long-term Care

출처: MISSOC(2025).

제도별 카테고리는 재원조달(Financing), 보건의료(Health Care), 상병 급여(Sickness), 모성/부성급여(Maternity/Paternity), 장애(Invalidity), 노령(Old-Age), 유족(Survivors), 산재 및 직업병(Accidents at Work and Occupational Diseases), 가족급여(Family Benefits), 실업(Unemployment), 최저생계보장(Guaranteed Minimum Resources) 등 12개 영역으로 구성되어 있다(〈표 2-5〉 참조).

MISSOC의 핵심 강점은 제도의 구조적 특성에 대한 상세하고 체계적인 정보 제공에 있다. 급여 수급 요건, 급여 산정 방식, 관리기관, 재원 조달 방식 등 제도 운영의 핵심 요소들을 표준화된 형식으로 기술하여 국가 간 제도적 차이를 명확히 파악할 수 있게 한다. 또한 영어, 프랑스어, 독일어로 제공되어 접근성이 높으며, 웹 기반 인터페이스를 통해 실시간 국가 간 비교가 가능하다.

그러나 MISSOC는 정성적 제도 설명에 집중되어 있어 정량적 재정 분

석에는 한계가 있다. 지출 규모, 수급자 수, 급여 수준 등에 대한 시계열 데이터가 제한적이어서 재정 지속가능성 분석이나 장기 전망에 직접 활용하기 어렵다. 또한 텍스트 기반의 정보 구조로 인해 자동화된 데이터 처리나 통계적 분석을 위해서는 별도의 가공 과정이 필요하다는 제약이 있다.

ESSPROS와 MISSOC는 상호보완적 관계에 있다. ESSPROS가 유럽 사회보장제도의 재정적 측면을 정량적으로 분석하는 도구라면, MISSOC는 제도의 구조적·운영적 특성을 정성적으로 이해하는 도구로서 기능한다. 두 체계를 결합하여 활용할 경우 유럽 사회보장제도에 대한 포괄적이고 심층적인 분석이 가능하다. 특히 제도 개혁의 효과를 분석하거나 정책 이전(policy transfer)을 위한 비교연구에서 MISSOC의 정성적 정보는 ESSPROS의 정량적 데이터와 함께 중요한 분석 기반을 제공한다.

#### 다. ILO의 Social Security Inquiry<sup>12)</sup>

국제노동기구(ILO)의 사회보장조사(Social Security Inquiry, 이하 SSI)는 1949년 이래 전 세계 사회보장제도의 발전을 추적해 온 가장 오랜 역사를 지닌 국제 사회보장 재정 데이터베이스이다. ILO는 1949년부터 1997년까지 18차례에 걸쳐 사회보장비용조사를 실시하였으며, 이는 ILO 협약 제102호와 권고 제67호, 제69호의 틀 내에서 수행되었다. 한편 자원 부족으로 1999년 조사가 중단되었다가, 2005년부터 사회보장 분야의 지식기반 개선을 위한 일환으로 사회보장조사를 재개하였다. 이때, 기존 사회보장비용조사에서 사용된 개념을 적극 활용하는 동시에 보다 포괄적인 접근법을 취하여 지출과 자원 조달뿐만 아니라 수급자 수와

12) 여기에서 소개되지 않은 보다 구체적인 내용은 ILO(2019)를 참조하면 된다.

보호인구 수, 평균 급여 수준에 대한 자료도 구축하였다.

SSI도 ILO 협약 제102호와 권고 제67호, 제69호를 근거로 사회보장제도를 우발 사태(contingencies), 위험(risks), 욕구(needs) 등을 기준으로 제도를 기능별로 분류하고 있다. 이때 사회보장제도의 영역은 노령(old age), 장애(disability), 유족(survivors), 질병 및 의료(sickness & health), 실업(unemployment), 고용 상해 및 산업 재해(employment injury & occupational disease), 가족/아동(family/children), 모성(maternity) 등 8개 영역에 추가적인 사회보장제도로써 주거(housing), 기초교육(basic education), 기타 소득 보장 및 공공부조/사회적 배제(other income support & assistance(n.e.c.)/social exclusion) 등 3개 영역을 보완하여 총 11개 영역으로 분류하였다.

〈표 2-6〉 SSI의 사회보장재정 자료 체계

차원	자료	소개
지출	총 사회보장 지출	모든 사회보장제도의 지출 합계
	급여 지출	수급자에게 직접 지급되는 현금 및 현물 급여
	관리 비용	제도 운영을 위한 행정 비용
	기타 지출	기타 지출
재원 조달	기여금	고용주 기여금, 피보험자 기여금
	정부기여금	중앙정부 및 지방정부 기여금
	투자 수익	기금 운용 수익
	기타 수입	기타 재원조달원
보장성	적용대상 인구	제도별 보장 대상 인구
	수급자 수	실제 급여를 받는 사람의 수
	평균 급여 수준	수급자 1인당 평균 급여액
	보장률	보허대상 대비 실제 보장인구 비율

출처: ILO(2019)를 활용하여 저자 작성.

SSI의 독특한 특징은 개발도상국의 비공식부문 사회보장까지 포괄한다는 점이다. 전통적인 사회보험 중심 접근을 넘어서 사회부조, 사회수당, 고용보장사업, 기초교육 지원, 비공식부문 보호까지 포함하여 사회보장체계의 전체상을 총체적으로 파악할 수 있게 한다. 이는 사회보장제도 발전 초기 단계에 있는 국가들의 정책 설계에 중요한 시사점을 제공하는 데 목적을 두기 때문이다.

SSI는 사회보장 재정분석과 관련하여 개별 사회보장제도에 대한 자료를 지출, 재원 조달, 보장성의 측면에서 다차원적으로 구축한다(〈표 2-6〉 참조). 먼저, 지출 자료는 총 사회보장 지출, 급여 지출, 관리 비용, 기타 지출로 구분되며, 재원 조달 자료는 기여금, 정부 기여금, 투자수익, 기타 수입으로 분류된다. 보장성 자료는 보호인구 수, 수급자 수, 평균 급여 수준, 보장률 등을 포함하여 사회보장제도의 양적 확장과 질적 개선을 균형 있게 평가할 수 있다. 이러한 다차원적 접근은 재정적 규모뿐만 아니라 인구집단별 보호수준, 급여적절성, 보장률 등을 종합적으로 평가하여 사회보장제도의 실질적 성과를 측정할 수 있게 한다. 이는 “사회보장 최저선(Social Protection Floors)” 개념의 실현 정도를 측정하는 데 유용하며, 유엔 지속가능발전목표 중 사회보장 관련 목표의 모니터링과 직접 연계된다.

SSI의 가장 중요한 장점은 선진국뿐만 아니라 개발도상국까지 포괄하는 유일한 사회보장 재정분류 체계라는 점이다. 공식 사회보험에서부터 비공식 부문 사회 보호까지 포괄하여 사회보장체계의 전체상을 파악할 수 있으며, 다차원적 성과 평가를 통해 사회보장제도의 발전 수준을 종합적으로 측정할 수 있다. 하지만 개발도상국의 비공식부문 포함으로 인한 데이터 수집과 검증상 현실적 어려움이 상당하며, 통계 인프라가 부족한

국가에서는 데이터의 완전성과 정확성을 확보하기 어려워 국제 비교 분석의 신뢰성을 담보하기 어렵다. 또한, 재정분석을 위한 자료체계의 정기적 최신화에 한계가 있어 최신 동향 반영에 제약이 있으며, 국가별로 상이한 제도 특성과 발전 수준으로 인한 방법론적 표준화의 어려움이 존재한다.

## 라. ISSA의 Country Profiles<sup>13)</sup>

국제사회보장협회(International Social Security Association, 이하 ISSA)의 Country Profiles는 1990년대부터 구축된 전 세계 사회보장 제도에 대한 포괄적 정보 데이터베이스이다. 2025년 현재 190개 국가와 지역의 사회보장제도를 체계적으로 기술하고 있으며, 미국 사회보장청(SSA)과의 협력을 통해 발간되는 “Social Security Programs Throughout the World”의 핵심 기반 자료로 활용되고 있다.

ISSA Country Profiles의 가장 중요한 특징은 정성적 제도 정보의 체계적 제공에 있다. 기존의 OECD SOCX, EU ESSPROS, ILO SSI 등이 주로 정량적 지출 및 재정 데이터에 집중하는 것과 달리, Country Profiles는 각국 사회보장제도의 제도적 특성, 수급 요건, 급여 산정 방식, 관리 운영 체계 등에 대한 상세한 정성적 정보를 제공한다.

Country Profiles는 사회보장제도를 크게 5개 주요 영역으로 분류한다. 노령·장애·유족연금(Old-age, disability, and survivors), 상병·모성급여(Sickness and maternity), 고용상해(Work injury), 실업급여(Unemployment), 가족수당(Family allowances)이 핵심 영역이며, 영역별로 제도의 유형, 적용 범위, 재원 조달, 수급 요건, 급여 산정, 관리기

13) 여기에서 소개되지 않은 보다 구체적인 내용은 ISSA(2025)를 참조하면 된다.

관 등에 대한 상세 정보를 체계적으로 정리하고 있다.

국가별 정보는 표준화된 양식에 따라 구성되어 국가 간 비교가능성을 확보하고 있다. 제도 유형(사회보험, 사회수당, 공공부조 등), 적용 대상(강제가입, 임의가입, 적용 제외), 재원 조달 방식(보험료율, 정부 부담, 기여 상한 등), 수급 요건(가입 기간, 연령, 소득 기준 등), 급여 내용(급여율, 급여 기간, 최저·최고 보장 등), 관리기관(정부 부처, 공단, 민간 기관 등)으로 구성된 6개 핵심 항목을 통해 각국 제도의 세부 특성을 비교 분석할 수 있다.

Country Profiles의 핵심 강점은 제도의 정성적 특성에 대한 체계적 정보 제공이다. 예를 들어, 연금제도의 경우 단순한 지출 규모나 수급자 수를 넘어서 급여 산정 공식, 물가 연동 방식, 조기퇴직 조건, 크레딧 인정 범위 등 제도 운영의 핵심 요소들을 상세히 기술한다. 이는 제도 간 실질적 비교와 정책 효과 분석을 위한 필수적 정보를 제공한다.

〈표 2-7〉 ISSA 데이터 구조

항목	세부 내용	분석 가능 요소
제도 유형	사회보험, 사회수당, 공공부조	제도 성격 비교
적용 대상	강제가입, 임의가입, 적용 제외	보장 범위 분석
재원 조달	보험료율, 정부 부담, 기여 상한	재정 구조 파악
수급 요건	가입 기간, 연령, 소득 기준	수급 자격 비교
급여 내용	급여율, 급여 기간, 보장한도	급여 적절성 평가
관리 기관	정부 부처, 공단, 민간 기관	거버넌스 분석

출처: ILO(2019)를 활용하여 저자 작성.

또한 정기적 업데이트를 통한 최신성 확보도 중요한 장점이다. 각국의 사회보장제도 개혁과 정책 변화를 반영하여 지속적으로 정보를 갱신하고 있으며, 특히 주요 제도 개혁이 있는 경우 신속한 업데이트를 통해 최신

동향을 반영한다.

전 세계적 포괄성도 주목할 만한 특징이다. 선진국뿐만 아니라 개발도상국, 최빈개도국까지 포괄하여 사회보장제도의 글로벌 현황을 종합적으로 파악할 수 있다. 이는 국제기구나 개발 협력 기관의 정책 설계와 기술 지원에 중요한 기초자료를 제공한다.

Country Profiles의 주요 한계는 정량적 재정 데이터의 부족이다. 제도의 특성에 대한 상세한 정보는 제공하지만, 지출 규모, 재정수지, 적립 기금 현황 등 재정분석에 필수적인 정량적 데이터는 제한적으로만 제공된다. 이로 인해 재정 지속가능성 분석이나 장기 전망 등에 직접 활용하기 어려운 한계가 있다.

또한 정보의 표준화와 비교가능성 한계도 지적될 수 있다. 국가별로 상이한 제도 구조와 용어 사용으로 인해 완전한 표준화가 어려우며, 특히 복합적 제도나 과도기적 제도의 경우 명확한 분류가 곤란한 경우가 있다.

업데이트의 불규칙성도 제약요인이다. 국가별로 정보 제공 시기와 상세 수준이 다르며, 일부 국가의 경우 상당 기간 업데이트가 지연되는 경우가 있어 정보의 최신성과 완전성에 한계가 있다.

그럼에도 불구하고 Country Profiles는 사회보장 장기재정분석을 위한 정성적 기초자료로서 중요한 가치를 지닌다. 제도의 구조적 특성과 운영 방식에 대한 상세 정보는 재정전망 모형의 가정 설정과 시나리오 구성에 필수적이며, 특히 제도 간 상호작용 효과나 정책 변화의 영향을 분석하는 데 중요한 기초를 제공한다.

### 3. 제도유형별 접근을 통한 사회보장 재정분석 체계 개선 방안

#### 가. 기존 분류체계의 쟁점

국제적으로 사용되고 있는 5가지 주요 사회보장 재정분류 체계를 살펴본 결과, 각각 고유한 방법론적 장단점을 가지고 있으며 상호보완적 활용을 통해 사회보장 장기재정분석의 정교화가 가능하다. OECD SOCX의 조세 효과 고려 방법론은 실질적 사회지출 수준 측정에서, EU ESSPROS의 재원 조달 정보와 MISSOC의 제도 구조 비교 기능은 재정 지속가능성 분석에서, ILO SSI의 포괄적 사회보장 접근은 개발도상국을 포함한 글로벌 비교 분석에서, ISSA Country Profiles의 정성적 제도 특성 정보는 제도 간 구조적 차이 분석에서 각각 중요한 역할을 한다.

그러나 기존 분류체계는 복지국가의 제도적 구조와 작동 방식을 포착하는 데 일정한 한계를 보인다. 첫째, 정책목표에 따른 분류는 동일한 목표를 가진 프로그램들이 상이한 제도적 설계와 작동 원리를 가질 수 있다는 점을 간과한다. 예컨대, 노령이라는 동일한 영역에서도 기여 기반 연금, 보편적 기초연금, 자산조사 기반 노인 지원금, 노인돌봄서비스는 각기 다른 제도적 논리와 재분배 메커니즘을 가진다. 둘째, 지출 규모 중심의 접근은 복지프로그램의 구조적 특성과 제도적 효과를 충분히 반영하지 못한다. 셋째, 기존 분류체계는 복지국가 개혁의 질적 변화를 포착하는 데 제한적이다. 동일한 지출 규모 내에서도 제도적 구성의 변화가 발생할 수 있으며, 이는 복지국가의 본질적 변화를 의미할 수 있다.

특히 한국과 같이 압축적 사회보장제도 발전을 경험하는 국가에서는 제도유형별 재정 동학의 차이가 장기재정전망의 정확성을 좌우하는 핵심 변수로 작용한다. 따라서 제도유형별 분류와 정책영역별 분류를 결합한

이차원적 접근이 불가피하다.

## 나. 제도유형을 고려한 분류체계의 설계

복지국가 발전의 제도주의적 접근은 제도의 형태적 특성과 기능적 목적이 상호 독립적이면서도 보완적 차원임을 강조한다. Esping-Andersen의 복지체제론이 탈상품화와 계층화라는 이중적 차원을 통해 복지국가를 분석한 것처럼, 본 연구에서 제안하는 사회보장 재정분류 체계는 사회적 위험에 대응하는 정책 영역(수직축)과 재원 조달과 급여 형태를 기준으로 한 제도유형(수평축)의 교차를 통해 사회보장제도의 전체 지형을 포괄한다.

정책 영역 차원은 사회적 위험의 특성과 정책 대상의 차이를 구조화한다. 본 연구에서는 OECD SOCX의 9개 정책 영역(노령, 유족, 장애, 보건, 가족, 적극적 노동시장, 실업, 주거, 기타)을 채택한다. SOCX의 사회보장제도 기능별 분류체계는 ESSPROS의 사회보장 기능별 분류체계와 서로 변환 가능하며, 국가 포괄성이 더 큰 SOCX의 분류체계가 국제 비교 분석에 유리하다. 각 정책 영역은 고유한 인구학적 특성, 경제적 민감도, 정책 수요 탄력성을 갖는다.

제도유형 차원은 사회보장제도의 재원 조달과 급여 형태의 근본적 차이를 반영한다. 본 연구에서는 급여 형태와 재원 조달 방식을 교차한 수당형, 사회보험형, 사회서비스형 등 3개의 제도유형별 분류체계를 제시한다. 이러한 분류는 국가별 사회보장제도의 현실적 다양성을 체계적으로 포착하면서도 국제 비교 가능성과 재정 분석의 정확성을 동시에 확보할 수 있다는 장점을 갖는다.

#### 다. 재정분석을 위한 제도유형별 주요 특성

본 연구는 사회보장제도를 재원 조달 방식과 급여 형태에 따라 수당형, 사회보험형, 사회서비스형의 세 가지 유형으로 구분한다.

수당형은 주로 일반 조세를 주요 재원으로 하여 특정 인구 집단에 정액 또는 차등 급여를 제공하는 제도로 정의된다. 이 유형은 베버리지형의 핵심 특징인 시민권 기반 수급권과 재분배 기능 극대화를 구현한다. Marshall(1950)의 사회적 시민권 이론과 직접적으로 연결되며, Rawls(1971)의 정의론에서 강조한 분배정의 원칙을 실현하는 주요 수단이다. 이론적으로 이 유형은 시장실패에 대한 정부개입의 전형적 사례이다. Arrow(1963)의 불완전정보 이론과 Akerlof(1970)의 역선택 이론에 따르면, 시장 기제의 역할만으로는 효율적 위험 분산이 불가능한 영역에서 정부의 강제적 개입이 정당화된다. 특히 노령, 장애, 아동양육 등의 사회적 위험은 개인의 선택이나 노력과 무관한 경우가 많아 사회적 연대의 원칙에 따른 집합적 대응이 필요하기 때문이다. 또한, 기여 이력과 무관한 급여 자격은 이 유형의 핵심 특징이다. 수급 자격은 연령, 소득, 장애, 가족구조 등 객관적 조건에 의해 결정되며, 과거 기여 여부는 고려되지 않는다. 이는 Sen(1999)의 역량접근법(capability approach)과 연결되어, 개인의 기본적 역량 확보를 사회의 책임으로 인식한다. 즉, 시민권 또는 거주권 기반 수급권은 사회구성원으로서의 지위 자체가 급여의 근거가 됨을 의미한다. 그리고 이 유형의 기대되는 성과로서 재분배 효과 극대화는 고소득층의 조세부담을 통해 저소득층에게 급여를 제공하는 구조적 특성에서 비롯된다. Musgrave(1959)의 재정학 이론에 따르면, 이는 정부의 분배 기능(distribution function)의 전형적 실현이다.

일반 조세를 기반하는 수당형의 재정적 특성은 다음과 같다. 첫째, 조

세 의존성은 이 제도유형의 가장 중요한 재정적 특성이다. 제도의 지속가능성이 국가의 전반적 재정 상황과 직결되며, 경기 침체 시 재정 압박이 직접적으로 급여 삭감 압력으로 작용할 수 있다. 둘째, 정치적 지속가능성은 납세자의 정치적 지지에 크게 의존한다. 선행 연구들에 따르면 (Pierson, 1994; Korpi & Palme, 1998), 조세 기반 제도라 하더라도 목표집단(target group)이 명확하고 수혜자와 부담자가 구분되는 경우 정치적 취약성이 크다. 이에 반해 광범위한 수혜 계층을 갖는 수당 제도의 경우 상대적으로 안정적인 정치적 지지를 얻을 수 있다. 셋째, 인구구조 민감성은 수급 대상 인구의 변화가 재정 소요에 직접적 영향을 미친다는 특성이다. 예컨대, 전 세계적인 인구 고령화 기조에 따른 노인인구의 자연 증가는 연금 관련 지출 급증을 초래한다. 넷째, 경기변동 대응성은 제한적이다. 경기 침체 시 저소득층이 증가하여 수급자는 늘어나지만 조세 수입은 감소하여 자동안정화 기능이 떨어진다.

두 번째 유형은 사회보험형이다. 사회보험제도는 가입자의 기여금을 주요 재원으로 하여 특정 사회적 위험 발생 시 급여를 제공하는 제도이다. 이는 비스마르크형의 전형으로, 보험원리와 사회연대 원리를 결합한 독특한 제도적 특성을 갖는다. 경제학적으로는 시장보험의 한계를 극복하기 위한 사회보험(social insurance)의 원리에 기초한다. Diamond(1977)에 따르면, 사회보험은 시장보험이 제공할 수 없는 위험 분산과 재분배 기능을 동시에 수행한다. 특히 장기간에 걸친 위험(장수 위험, 장애 위험 등)과 상호 관련성이 높은 위험(경기변동에 따른 대량 실업 등)에 대해서는 시장 기제보다 사회보험이 더 효율적이다. 그리고 기여-급여 연계 원칙이 이 유형의 핵심이다. 기여 수준과 기간에 따라 급여 수준이 결정되는 구조로, 이는 개인의 권리와 의무를 명확히 연계시킨다. 그러나 순수한 보험수리적 등가성이 아닌 사회적 적절성을 고려한 수정

된 연계성을 갖는다. 사회보험형의 위험분산(risk pooling) 기능은 대수의 법칙에 기반하여 개별 위험을 집단적으로 분산하는 보험의 기본 원리를 준용한다. 하지만 공적 제도로써 사회보험은 순수한 위험 분산을 넘어 소득재분배 기능도 수행한다. 이는 Barr(2001)가 강조한 사회보험의 이중적 성격이다. 또한, 강제가입을 통한 역선택 방지는 고위험군의 선별적 가입을 방지하여 제도의 재정안정성을 확보한다. Akerlof(1970)의 역선택 이론에 따르면, 자발적 가입 시 고위험군의 집중으로 보험시장이 붕괴할 수 있으나, 강제가입은 이를 방지한다.

기여를 전제한 사회보험형의 재정 원칙과 특성은 다음과 같다. 첫째, 보험수리적 균형은 장기적으로 수입과 지출이 균형을 이루어야 한다는 사회보험의 기본 원칙이다. 하지만 인구구조 변화로 인해 대다수 국가가 장기적으로 수지 불균형 상태에 직면할 것으로 보인다. 둘째, 세대 간 부양 구조는 부과방식 운영으로 인해 현재 근로 세대가 현재 수급 세대를 부양하는 구조적 특성이다. 급속한 고령화로 인해 부양비(가입자 대비 수급자 비율)가 빠르게 악화되고 있는 상황에서 이러한 특성이 세대 간 형평성 문제가 제기되는 근원이 되고 있다. 셋째, 제도 성숙도 영향은 제도 도입 초기에는 기여자가 많고 수급자가 적다가, 제도 성숙과 함께 수급자가 급증하는 현상이다. 넷째, 기금운용 수익은 적립 기금의 투자수익이 제도의 재정 안정성에 미치는 영향을 의미한다.

세 번째 유형은 사회서비스형이다. 이 유형은 급여가 현금이 아닌 현물 형태로 제공되는 사회보장 급여를 의미한다. 이는 OECD SOCX의 ‘현물 급여(benefits in kind)’ 범주에 해당하며, 탈상품화된 사회서비스 제공을 통한 시민의 실질적 복지 향상을 목표로 한다. 이론적 기반은 Baumol(1967)의 서비스 경제학 이론과 Stiglitz(2000)의 공공재 이론에 근거한다. 예컨대, 서비스의 경제학적 특성은 무형성과 이질성, 비분리

성, 소멸성으로 설명 가능하다. 무형성(intangibility)으로 인해 품질 측정과 표준화가 어렵고, 이질성(heterogeneity)으로 인해 동일한 서비스라도 제공자와 이용자에 따라 품질이 달라진다. 비분리성(inseparability)은 생산과 소비가 동시에 일어나며, 소멸성(perishability)은 저장 불가능하다는 특성이다. 이러한 특성으로 인해 서비스를 통해 제공되는 사회보장급여는 시장실패가 빈번하게 발생한다. 정보 비대칭, 외부효과, 공공재적 성격 등이 복합적으로 작용하여 시장 기제만으로는 효율적 공급이 어렵다. 따라서 정부의 직접 공급이나 규제를 통한 시장 개입이 필요하다.

현물급여 중심의 사회서비스형은 다음과 같은 재정적 특성을 가진다. 첫째, 공급 체계 의존성은 서비스 제공을 위한 인프라 구축 비용이 선행되어야 한다는 특성이다. 시설 건립, 장비 구입, 인력 양성 등 초기 투자 비용이 크고, 이후에도 지속적인 유지관리비가 필요하다. 둘째, 질적 수준과 비용의 상충관계(trade-off)는 서비스 품질 향상이 비용 증가를 수반한다는 구조적 특성이다. 보육교사 대 아동 비율 개선, 요양보호사의 처우 개선 등은 모두 비용 증가 요인이다. 둘째, 인건비 비중이 총비용의 절 비중을 차지하는 노동집약적 특성을 갖는다. Baumol의 비용 질병(cost disease) 이론에 따르면, 생산성 향상이 제한적인 서비스업에서 임금 상승은 비용 상승으로 직결되기 때문에 시장에서 효율적인 분배가 어려운 제약이 있다. 넷째, 규모의 경제는 이용자 밀도가 높은 지역일수록 서비스 제공의 효율성이 증대되는 현상이다. 이는 도시와 농촌 간 서비스 공급 비용의 차이를 설명하며, 지역 간 형평성 확보의 과제를 제기한다.

〈표 2-8〉 제도유형별 핵심 특성

구분	수당형	사회보험형	사회서비스형
재원	일반조세(국세, 지방세)	사회보험료(+국고지원)	조세, 사회보험료, 이용자 부담
급여 형태	현금급여(정액/차등)	현금급여(소득비례, 정액)	현물급여
수급 자격	시민권/거주기간/육구	기여 이력+사회적 위험	육구+소득수준
재분배성	수직적 재배분 강함	수평적 재배분 중심	현물 재분배
정치적 지지	납세자 저항 가능성 존재	기여자-수급자 연대	이용자-공급자 연합
재정 예측성	정치적 결정에 의존	인구 및 경제변수 의존	수급 균형에 의존
정기 민감도	높음(조세수입 변동)	중간(고용 및 임금 연동)	낮음(필수서비스 특성)
SOCX 연계성	공적 현금 급여	사회보험	공적 현물급여

출처: 저자 작성.

본 연구에서 제안하는 정책영역과 제도유형을 교차한 9×3 매트릭스 구조의 사회보장 재정분류 체계는 다음과 같은 분석적 장점을 갖는다.

첫째, 정책 대체성과 보완성의 체계적 분석이 가능하다. 동일 정책영역 내에서 제도유형 간 대체관계(예: 기초연금과 국민연금의 노령소득보장 기능)와 상이한 정책영역 간 보완관계(예: 고용보험과 직업훈련의 노동시장 통합 기능)를 동시에 포착할 수 있다. 이는 사회보장제도 간 상승효과와 대체효과를 정량화하는 기초를 제공한다.

둘째, 재정 위험의 다차원적 평가가 실현된다. 제도유형별로 상이한 재정 위험(조세 의존성, 보험수리적 불균형, 비용 상승 압력)과 정책영역별로 차별화된 수요 압력(인구학적, 경제적, 사회문화적)을 교차 분석함으로써 복합적 위험 프로파일을 도출할 수 있다.

셋째, 국제 비교가능성과 국가 특수성의 균형이 확보된다. OECD SOCX 정책영역 분류는 국제 비교의 표준을 제공하면서, 제도유형 분류는 개별 국가가 가진 사회보장제도의 특수한 발전 경로와 제도적 특성을 반영한다. 이러한 이중적 접근은 글로벌 벤치마킹과 맥락적 분석을 동시에 가능케 한다.

넷째, 정성적 정보의 체계적 활용이 가능하다. ISSA Country Profiles에서 제공하는 풍부한 정성적 제도 특성 정보를 체계적으로 분석하여 각 제도를 9×3 매트릭스 내 적절한 위치에 배치할 수 있으며, 최근 LLM 기술의 발전으로 이러한 정성적 정보의 정량화와 분석 자동화도 가능해지고 있다.

이러한 통합적 분류체계는 기존 국제 분류체계들의 한계를 보완하면서 사회보장 장기재정분석의 정확성과 유용성을 높일 수 있을 것으로 기대된다. 특히 OECD SOCX의 국제 비교 기준과 국내 제도의 특수성을 동시에 고려함으로써 보다 실용적인 분석 도구로 활용될 수 있을 것이다. 다음 절에서는 이러한 이론적 토대를 바탕으로 구체적인 분석 체계 구축 방안을 제시하고자 한다.

### 제3절 사회보장 재정분석 체계 구축 방안

앞서 검토한 국내의 사회보장 재정분석의 현황과 한계, 그리고 재정지표 분류체계의 이론적 토대를 바탕으로, 본 절에서는 통합적 사회보장 재정지표 데이터베이스의 구체적 구축 방안을 제시한다. 특히 3개년 연구의 전체 설계와 1차년도 데이터베이스 구축 방안을 중심으로 체계적인 접근 방법을 제시하고자 한다.

#### 1. 사회보장 재정지표 데이터베이스 설계

##### 가. 통합적 분석을 위한 DB 구축 체계

현대 복지국가는 다층적이면서도 다축적으로 상호 연계된 사회보장제

도로 구성되어 있다. 최근 인구 고령화와 경제구조 변화가 가속화되면서 기존 생산 체제를 전제로 설계된 사회보장제도의 기저가 훼손되고 있다. 한국의 사회보장 체계 역시 급격한 인구구조 변화와 저성장 기조가 지속됨에 따라 재정적 지속가능성에 대한 경각심이 커지고 있는 실정이다. 이에 따라 복지국가의 지속가능성을 담보하기 위해서는 사회보장 장기재정 분석이 필수적인 과업이라 할 수 있다.

지금까지의 논의를 종합하면 개별 제도를 중심으로 한 사회보장 재정분석은 개선의 여지가 있다. 즉, 많은 국가의 사회보장 재정분석은 제도별로 수행되는 경향이 있다. 다양한 사회보장제도들이 각기 다른 관리기관, 분석 방법론, 시간 범위로 독립적인 재정 추계를 수행함에 따라 통합적 관점에서의 분석이 보완될 필요가 있다. 일부 사례에서 통합적인 관점을 가지고 장기재정분석을 수행하고 있지만, 일관된 제도 분류 체계하에서 구축된 분석 자료가 확충된다면 보다 체계적인 재정분석이 가능할 것이다.

제도별 접근의 한계는 제도 간 상호작용에 대한 동태적 관점을 충분히 반영하기 어렵다는 점이다. 사회보장제도 분류체계의 현대적 관점에서 강조하는 정책 상호 보완성의 관점에서 볼 때, 사회보장제도는 상호 보완적이면서도 대체 관계를 형성하는 유기적 체계이다. 즉, 사회보장제도의 재정 효율성은 개별 제도의 효율성의 합이 아닌 사회보장체계 전체의 최적화에서 비롯된다(Hemerijck, 2013). 한국의 경우만 보더라도, 국민연금과 기초연금, 그리고 국민기초생활보장제도의 대체관계, 건강보험과 장기요양보험, 고용보험과 직업훈련의 보완관계 등이 복합적으로 작용하고 있다. 따라서 이러한 제도 간 상호작용 효과를 체계적으로 분석할 수 있는 통합적 접근이 필요하다.

수급자 관점에서도 통합적 관점은 매우 중요하다. 개인은 전 생애에 걸쳐 다양한 사회보장제도를 경험하며 하나의 제도 변화가 다른 제도 이용

에 미치는 행동 반응을 고려해야 한다. 예를 들어, 국민연금 수급 연령 상향 조정은 고용보험과 기초연금 재정에 직접적인 영향을 미치며, 이러한 파급효과는 통합적 분석들에서만 정확한 측정이 가능하기 때문이다.

통합적 사회보장 재정분석을 위해 앞서 정책영역과 제도유형을 교차한 사회보장 재정 분류 체계를 제안한 바 있다. 해당 체계는 3개의 제도유형(수당형, 사회보험형, 사회서비스형)과 9개의 정책 영역(노령, 유족, 장애, 보건, 가족, 적극적 노동시장정책, 실업, 주거, 기타)을 교차한 매트릭스 구조로서 총 27개의 분석 영역으로 구조화되어 있다. 각 영역은 고유한 제도적 특성과 재정 동학을 갖는 분석 단위가 된다(〈표 2-9〉 참조).

사회보장 재정의 세분화는 단순한 분류를 넘어 정책 간 상호작용을 체계적으로 분석하는 기초로 작동한다. 예컨대, 정책 영역을 노령 영역으로 고정한다면 동일한 사회보장기능하에서 각 제도유형 간의 상호 관계, 가령 기초연금과 국민연금의 대체관계, 기초연금 및 국민연금과 노인요양서비스의 보완관계 등을 동시에 파악할 수 있다. 그리고 제도 유형을 사회보험형에서 고정한 채 정책 영역을 횡단하여 분석하면 국민연금과 특수직역연금, 건강보험, 장기요양보험, 고용보험, 산재보험 등 8대 사회보험 간 상호작용과 재정 연계성을 평가할 수 있다.

이러한 매트릭스는 사회보장 장기재정분석을 위한 기초 데이터 구조를 제시한다는 점에서 중요한 함의를 가진다. 각 영역에 포함된 제도는 서로 다른 재정 메커니즘과 지속가능성 과제를 가지고 있으므로 셀별로 특화된 재정추계 모형을 개발하고 이를 통합하는 접근이 필요하다. 즉, 단순히 개별 제도를 합산하여 사회보장 장기재정분석을 수행하는 것에서 나아가 영역 간 복잡한 상호작용을 고려한 체계적 분석의 필요성이 요구된다고 할 수 있다.

〈표 2-9〉 제도유형×정책영역 매트릭스에 따른 사회보장 재정분류 체계

구분	수당형	사회보험형	사회서비스형
노령	기초연금	국민연금, 특수직역연금	노인복지서비스, 노인일자리사업
유족	보훈급여	유족연금	유족 대상 프로그램
근로 무능력	장애연금, 장애(아동)수당	장애연금, 장애급여 산재보험	장애인활동지원, 장애인 복지시설
보건	건강보험료 지원	요양급여, 재가 및 시설급여	의료급여, 보건소서비스
가족	아동수당, 한부모가족지원 등	출산전후 휴가급여, 육아휴직 급여	보육서비스, 아이돌봄서비스 등
적극적노 동시장	구직촉진수당	직업능력개발사업, 고용안정사업 등	고용센터 서비스, 직업훈련 등
실업	실업부조	구직급여	취업성공패키지 등
주거	주거급여, 주거복지바우처 등	-	공공임대주택 등
기타	기초생활보장, 긴급복지지원 등	-	지역사회서비스 등

출처: 저자 작성.

본 연구에서 구축하고자 하는 사회보장 장기재정분석 데이터베이스는 다음과 같은 기본 원칙에 따라 설계된다. 첫째, 국제 비교가능성 확보이다. OECD SOCX, EU ESSPROS 등 기존 국제 분류체계와의 정합성을 유지하면서도 한국의 제도적 특수성을 반영할 수 있는 구조를 채택한다. 이를 위해 OECD SOCX의 9개 정책영역 분류를 기본 틀로 활용하되, 제도유형별 세분화를 통해 분석의 정교화를 추구한다. 둘째, 제도 간 상호작용 포착이다. 개별 제도의 독립적 분석을 넘어 제도 간 대체관계와 보완관계, 누수효과와 중복급여 등을 체계적으로 분석할 수 있는 구조를 마련한다. 이를 위해 27개 분석 영역별로 표준화된 데이터 수집 체계를 구축하고, 영역 간 연계성을 추적할 수 있는 메타데이터를 함께 구축한다. 셋째, 정량적·정성적 정보의 통합이다. 기존 데이터베이스들이 주로 정량적 지출 정보에 집중하는 한계를 극복하기 위해 제도의 구조적 특성, 운

영 방식, 개혁 동향 등 정성적 정보를 체계적으로 수집하고 분석한다. 특히 최근 발전하고 있는 대규모 언어모델(LLM) 기술을 활용하여 정성적 정보의 정량화와 분석 자동화를 추진한다. 넷째, 시계열 일관성 확보이다. 제도 변화가 빈번한 사회보장 분야의 특성을 고려하여 제도 개편 전후의 연속성을 확보하고, 구조적 단절점을 명시적으로 처리할 수 있는 체계를 구축한다.

사회보장 재정분석의 기초가 되는 각 사회보장제도에 대한 데이터 수집은 27개 영역별로 표준화된 규정을 따른다. 자료의 표준화는 OECD SOCX 방법론을 준용하되, 각국의 제도적 특성을 반영한 조정이 요구되며, 개별 국가의 특수성을 고려한 자료 통합 규칙을 수립할 필요가 있다. 수집할 자료는 개별 사회보장제도와 관련된 재정 자료와 수급자 자료, 제도 모수(parameter), 성과지표 등이다. 재정 자료는 해당 제도의 수입과 지출, 적립금 규모 등이며, 수급자 자료는 수급자 수, 인구학적 특성, 수급률, 평균 급여액 등이다. 제도 모수는 급여 산식, 수급 요건, 기여율, 정부 지원 비율 등이 해당되며, 성과지표는 소득대체율, 빈곤 감소 효과, 서비스 이용률이 될 수 있다.

또한, 장기재정분석을 위해서는 최대한 장기 시계열 자료 구축이 요구된다. 이를 토대로 구축 가능한 가장 최신까지의 실적 자료를 기초로 장기 전망을 수행할 수 있는 것이다. 이때, 시계열 자료는 다음과 같은 원칙을 일관되게 준수하여야 한다. 첫째, 제도 변화 상태가 반영되어야 한다. 즉, 해당 제도의 도입 및 확대, 축소 등 제도 개편 시점을 명확히 표시하고 구조적 단절 처리가 되어야 한다. 둘째, 비교 가능성과 이해를 돕기 위해 모든 화폐 가치는 기준년도 불변가격으로 환산하는 물가 조정 처리가 수반되어야 한다. 이때, 국가 간 비교 가능성을 제고하는 방안으로 기준 년도 시점의 구매력 평가 지수(Purchasing Power Parity)를 활용할 수 있다.

## 나. 추계 방법론 및 정책분석 활용 체계 개발

사회보장 장기재정분석을 위해서는 재정 분류상 27개 영역별로 특화된 추계 모형을 개발하여야 한다. 각 제도유형별로 상이한 재정 메커니즘과 위험 요인을 고려한 차별화된 접근이 필요하다. 수당형의 경우, 인구 추계에 기반한 수급자 전망과 급여 인상률을 적용하여 재정소요액을 산출하는 방식이다. 대상 인구의 변화가 재정 소요에 직접적 영향을 미치므로 인구학적 요인에 대한 정확한 분석이 핵심이다. 특히 조세 의존성으로 인한 경기변동 효과와 정치적 결정에 따른 급여 수준 변화를 고려한 확률적 접근이 필요하다. 사회보험형의 경우, 코호트 모형 기반 가입자-수급자 전망에 보험 수리 모형을 적용하여 수지 전망을 수행하는 방식이다. 보험수리적 균형 원칙에 따른 장기적 재정 안정성 평가가 핵심이며, 세대 간 부양 구조와 제도 성숙도 효과를 명시적으로 모형화해야 한다. 특히 기여-급여 연계 구조와 인구구조 변화에 따른 부양비 악화 효과를 정교하게 반영해야 한다. 사회서비스형의 경우, 서비스 수요함수와 공급 제약 변인을 반영하여 비용을 전망하는 방식이다. 서비스의 노동집약적 특성으로 인한 비용 질병(cost disease) 효과와 품질 개선 요구에 따른 구조적 비용 상승 압력을 고려해야 한다. 또한 지역별 공급 격차와 시설 투자의 규모의 경제 효과를 반영한 공간적 분석이 필요하다.

각 영역 간 상호 정합성을 확보하기 위한 통합적인 분석틀을 구축할 필요가 있다. 모형 간 연계는 경제성장률과 물가상승률, 임금상승률 등 공통 거시경제 가정과 출생률, 사망률, 이동률 등 인구 가정을 공유하고, 제도 간 이동 및 중복수급을 명시적으로 모형화하는 것이다. 특히 제도 간 상호작용 효과를 포착하기 위해서는 개별 제도의 정책 변화가 다른 제도에 미치는 파급효과를 계량화할 수 있는 연계 모듈을 개발해야 한다. 예

를 들어, 국민연금 수급 연령 상향이 기초연금과 고용보험에 미치는 영향, 건강보험 보장성 강화가 장기요양보험 이용률에 미치는 영향 등을 체계적으로 분석할 수 있는 구조를 마련한다.

다양한 정책 시나리오의 영향을 체계적으로 평가할 수 있는 분석들을 제공한다. 첫째, 제도 개혁과 관련해서는 각 제도의 급여 인상이나 축소 등이 연계된 제도의 재정 상태 변화에 미치는 영향 등 제도 간 상호작용을 정량화할 수 있다. 둘째, 인구 가정과 관련해서, 출생률의 변화가 27개 각 영역에 미치는 차별적 영향을 분석할 수 있다. 예컨대 저출산은 아동이나 가족 관련 지출을 감소시키지만 장기적으로 사회보험 가입자 감소를 초래함으로써 해당 영역들에 대한 재정 압박을 가증시키는 것이다. 셋째, 경제 가정과 관련해서는 저성장 기조에 따른 경제성장률 둔화가 사회보장체계 전반에 미치는 차별적 영향력을 분석할 수 있다.

구축된 데이터베이스와 추계 모형을 활용하여 정책 결정자들이 활용할 수 있는 다양한 분석 도구를 개발한다. 우선 정책 시뮬레이션 도구로서 다양한 정책 옵션의 재정 영향을 사전 평가할 수 있는 시뮬레이션 도구를 개발하며, 특히 제도 간 상호작용을 고려한 통합적 영향 평가가 가능한 도구를 구축한다. 다음으로 국제 비교 분석 체계로서 OECD 국가들과의 체계적 비교를 통해 한국 사회보장제도의 상대적 위치를 진단하고 정책 방향을 도출할 수 있는 분석 체계를 구축한다. 27개 영역별로 OECD 평균, 복지체제 유형별 평균과 비교하여 개별 국가의 상대적 위치를 진단할 수 있다. 마지막으로 위험 평가 도구로서 사회보장 재정분류 체계의 각 영역이 갖는 고유한 위험 수준(risk profile)을 체계적으로 분석할 수 있는 도구를 개발한다. 발생 가능한 다양한 위험 요인을 확률적으로 모형화하여 몬테카를로 시뮬레이션을 수행할 수 있는 방법론을 구축하여 전체 사회보장체계의 취약 지점을 식별하고 우선순위에 따른 개혁 과제를 도출할 수 있다.

## 2. 사회보장 장기재정분석 체계 구축 시 고려사항

### 가. 제도유형별 고려사항

수당형에 대한 장기재정분석 시 인구추계의 정확성과 급여 인상률 가정, 대상 범위, 재원 분담구조 등을 고려하여야 한다. 특히, 인구추계의 정확성은 가장 중요한 고려 사항이다. 인구추계에서 전망한 대상 인구 규모와 구조 변화가 재정 소요를 직접 결정하기 때문이다. 특히, 장래인구 추계의 불확실성이 클수록 재정 전망의 오차도 커질 수 있기 때문에 단일 결과 산출 시 유의하여야 한다. 또한, 급여 인상률 가정의 경우, 물가 연동, 소득 연동, 또는 정치적 결정에 따른 인상률 설정을 어떻게 가정할지에 따라 장기 재정 전망에 결정적인 영향을 미친다. 이에 과거 급여 변동 이력 분석을 통한 확률적 접근 또는 현행 제도 유지를 전제한 접근을 통해 불확실성을 최소화할 필요가 있다. 대상 범위의 변동 가능성도 고려할 필요가 있다. 수당 제도의 경우, 대상 인구의 연령 기준 및 수급자 선정 기준에 대한 정치적 압력에 따라 제도 변경이 빈번하게 발생할 수 있기 때문이다. 마지막으로 지방정부의 재정 부담 증가 추세도 고려 요인 중 하나이다. 중앙-지방 간 비용을 분담하고 있는 일반 조세 기반 수당형 제도의 경우, 지방비 부담 비율이 높아질수록 지방 재정 압박이 가중될 수 밖에 없는 구조이다.

사회보험형에 대한 장기재정분석 시에는 무엇보다도 경제활동 참가율과 임금 및 물가상승률, 기금운용수익률 등 거시경제 가정 설정에 주의할 필요가 있다. 경제활동참가율은 사회보험제도의 가입자 규모에 영향을 미치기 때문에 인구구조 변화와 노동시장 여건을 종합적으로 고려하여 전망하여야 한다. 임금 상승률과 물가상승률도 실질 급여 수준과 기여 부

담 수준을 결정하는 중요한 변인이기 때문에 가정 설정 시 주의하여야 한다. 또한, 국가간 비교 관점에서 이러한 거시경제 가정은 공통된 가정을 사용할 수도 있으나 국가별 경제 수준 및 산업구조, 노동시장 상태, 인구 구조 등도 고려할 수 있다.

사회서비스형에 대한 장기재정분석 시 주의할 사항은 해당 사회보장제도에서 제공하는 서비스와 관련된다. 가령, 서비스 이용률의 경우 잠재 수요와 실제 이용률 간 괴리를 정확히 예측할 필요가 있으며, 사회발전에 따른 품질 기대 수준 상승도 고려할 필요가 있다. 이는 임금 상승률이나 물가상승률에 따라 정량화가 가능한 서비스 비용 증가 요인과 달리 제도의 질적 변화에 따른 비용 증가를 초래하기 때문이다. 이외에도 인구구조 변화에 따른 수요 증가에 대응하는 시설 및 인력 확충 계획 등도 사전에 검토하여 재정 전망 모형 설계 시 변인으로 투입하는 것으로 고려할 수 있다.

## 나. 분석 체계 구축 시 고려 사항

추계 방법론 구축 측면에서는 먼저 데이터 품질 관리가 중요하다. 데이터의 정확성과 일관성을 확보하기 위한 체계적인 품질 관리 방안이 필요하며, 이를 위해 데이터 수집 과정에서의 검증 절차, 이상값 탐지 및 처리 방안, 결측값 보정 방법론 등을 사전에 수립한다. 또한 모형의 확장성을 고려해야 한다. 향후 새로운 제도 도입이나 기존 제도 개편에 대응할 수 있는 유연한 모형 구조를 설계하며, 특히 제도 분류체계의 확장이나 변경이 필요한 경우에도 기존 모형과의 연속성을 유지할 수 있는 방안을 마련한다. 불확실성 처리 방법론 구축도 핵심적이다. 장기 전망의 특성상 인구·경제·정책 변수의 불확실성이 크므로 이를 체계적으로 처리할 수 있

는 확률적 접근법을 개발하여야 하며, 몬테카를로 시뮬레이션, 시나리오 분석, 민감도 분석 등 다양한 방법론을 조합하여 활용한다. 마지막으로 모형 간 정합성 확보가 필요하다. 27개 영역별 개별 모형들이 일관된 가정하에서 작동하도록 통합 관리 시스템을 구축하고, 특히 거시경제 변수와 인구학적 변수의 일관성을 유지하며 제도 간 상호작용을 적절히 반영할 수 있는 연계 메커니즘을 개발한다.

정책 분석 활용 측면에서는 정책 일관성 확보가 중요하다. 개별 제도별 분석과 통합 분석 간의 일관성을 확보하기 위한 조정 메커니즘을 마련하며, 특히 각 부처에서 수행하는 개별 제도 분석과의 정합성을 유지하면서도 통합적 관점에서의 새로운 통찰을 제공할 수 있는 방안을 모색한다. 정책 활용성 제고도 필수적이다. 분석 결과가 실제 정책 결정 과정에서 활용될 수 있도록 정책 결정자들의 수요를 사전에 파악하고 이를 반영한 분석 체계를 구축하며, 특히 정책 대안별 비교 분석과 우선순위 도출이 가능한 도구를 제공한다. 투명성과 재현 가능성 확보도 중요한 고려 사항이다. 분석 방법론과 가정, 데이터 출처 등을 투명하게 공개하여 분석 결과의 신뢰성을 확보하고, 동일한 조건에서 동일한 결과를 재현할 수 있는 체계를 구축한다. 마지막으로 사용자 인터페이스를 고려해야 한다. 정책 결정자와 연구자들이 쉽게 활용할 수 있는 직관적인 사용자 인터페이스를 설계하고, 복잡한 분석 결과를 시각화하여 제공하며 다양한 시나리오 분석을 손쉽게 수행할 수 있는 도구를 제공한다.

### 3. 사회보장 재정지표 데이터베이스 구축 방안

본 연구는 3개년 연구의 첫해로서, 사회보장 재정지표 탐색을 위한 데이터베이스 구축에 중점을 둔다. 이를 통해 구축된 재정지표 DB는 향후

2-3차년도에 수요 측면 전망과 통합되어, 궁극적으로는 공급-수요를 연계한 한국형 사회보장 장기재정분석 모델로 발전될 것이다.

앞서 제시한 27개 매트릭스 기반 통합 분석 체계를 실제로 구현하기 위해서는 체계적인 재정지표 데이터베이스 구축이 선행되어야 한다. 본 연구는 3개년 연구의 첫해로서 이러한 재정지표 DB 구축에 중점을 두며, 특히 국제 비교가능성과 정량·정성 통합 분석이 가능한 재정지표 체계를 구축하고자 한다.

1차년도 재정지표 데이터베이스 구축은 크게 세 단계로 구성된다. 첫 번째 단계에서는 OECD SOCX 등 기존 국제 재정지표 데이터베이스의 체계적 분석을 통해 본 연구에 적합한 설계 방향을 도출한다. 두 번째 단계에서는 OECD 38개국의 사회보장 재정지표 데이터를 정량 데이터(지출 규모)와 정성 데이터(제도 특성)로 구분하여 수집·정리한다. 세 번째 단계에서는 구축된 재정지표 DB의 유용성을 검증하기 위해 노령 영역 파일럿 분석을 수행한다.

이러한 단계적 접근을 통해 국제 표준에 부합하면서도 한국의 위치를 정확히 파악할 수 있는 재정지표 데이터베이스를 구축할 수 있을 것으로 기대된다. 구축된 재정지표 DB는 한국이 어디에 있고, 어디로 가야 하는지에 대한 답을 제공하며, 향후 정책 시나리오 분석과 장기재정 시뮬레이션의 핵심 기초자료가 될 것이다.





## 제3장

### 사회보장 재정분석을 위한 국제 DB 분석

제1절 OECD SOCX DB 분석

제2절 주요 사회보장제도 DB 분석

제3절 사회보장 재정분석을 위한 DB 구축 방향



## 제3장

# 사회보장 재정분석을 위한 국제 DB 분석

### 제1절 OECD SOCX DB 분석

본 절은 본 연구에서 구축하고자 하는 사회보장 장기재정분석 데이터베이스의 설계 방향을 도출하기 위해 먼저 국제적으로 가장 널리 활용되고 있는 OECD 사회지출데이터베이스(SOCX)를 심층 분석한다. SOCX는 앞서 이론적 검토에서 살펴본 바와 같이 OECD 국가들의 사회보장 재정 비교 분석을 위한 대표적 도구로서, 본 연구의 27개 매트릭스 구조 설계에 중요한 기준점을 제공한다.

#### 1. OECD SOCX DB 구조

##### 가. OECD의 데이터 수집 체계

OECD SOCX는 회원국들의 사회지출 데이터를 체계적으로 수집하고 관리하기 위해 이원화된 데이터 수집 체계를 운영하고 있다. 이러한 체계는 각국의 데이터 생산 현황과 국제기구 간 협력 체계를 고려하여 설계되었으며, 데이터의 품질과 국제 비교 가능성을 동시에 확보하려는 노력의 결과이다.

##### 1) 이원화된 데이터 수집 방식

OECD SOCX의 데이터 수집은 크게 두 가지 경로로 구분된다. 첫째는

유럽 국가들을 대상으로 한 EUROSTAT 경유 수집 방식이고, 둘째는 비유럽 국가들을 대상으로 한 직접 수집 방식이다.

유럽 국가 데이터 수집 체계는 25개 유럽 국가(EU 22개국, 아이슬란드, 노르웨이, 스위스)를 대상으로 EUROSTAT의 ESSPROS(European System of Integrated Social Protection Statistics) 데이터를 SOCX 분류체계로 변환하여 활용한다. 이는 중복적인 데이터 제출 부담을 줄이고, 유럽 차원에서 이미 검증된 데이터의 품질을 활용하려는 취지이다. 다만 이러한 방식은 OECD가 개별 국가와 직접적인 상호작용을 할 수 없다는 한계를 가지고 있으며, ESSPROS의 데이터 검증이 완료된 후에야 SOCX에 반영될 수 있어 데이터 공개에 시차가 발생한다.

비유럽 국가 데이터 수집 체계는 10개 비유럽 OECD 국가들이 사회정책 작업반(Working Party on Social Policy)을 통해 SOCX 질문지에 직접 응답하는 방식이다. 이 방식은 국가별 사회보장제도의 특성을 보다 세밀히 반영할 수 있고, 데이터 생산 과정에서 OECD와 각국 간의 직접적인 협의가 가능하다는 장점이 있다.

〈표 3-1〉 OECD SOCX 데이터 수집 체계

구분	데이터 수집 경로	주요 특징
EU 25개국	EUROSTAT ESSPROS → SOCX 변환	중복 보고 부담 감소 OECD와 직접 상호작용 제한 데이터 공개 시차 발생
비유럽국가	OECD 질문지 직접 응답	현재도 특성 세밀한 반영 OECD와 직접 협의 가능 국가별 특징 반영 가능

출처: OECD SOCX(2025) 활용 저자 재작성.

## 2) 다층적 데이터 통합 체계

SOCX는 단순한 사회지출 통계를 넘어 다른 OECD 데이터베이스와의 체계적 연계를 통해 다차원적 분석을 지원한다. 보건 영역의 경우 OECD Health Statistics와 연계하여 의료비 지출 데이터를 통합하고, 적극적인 노동시장정책 분야는 OECD Employment Database의 ALMP 데이터를 활용한다. 또한 아동보육 관련 지출은 OECD Education Database의 조기아동교육 데이터를 연계하여 포괄적인 가족정책 지출을 파악할 수 있도록 구성되어 있다.

이러한 데이터베이스 간 연계는 사회정책의 통합적 분석을 가능하게 하지만, 동시에 데이터베이스별 분류 기준과 측정 방법의 차이로 인한 일관성 문제를 야기하기도 한다. 특히 장기요양서비스의 경우 보건 통계와 사회서비스 통계 간 중복 계상의 가능성이 있어 지속적인 조정 작업이 필요하다.

## 3) 국가별 특화된 수집 체계: 한국 사례

한국의 경우 한국보건사회연구원(KIHASA)이 중심이 되어 복수 부처의 데이터를 통합 수집하는 체계를 구축하고 있다. 이는 한국의 사회보장 제도가 부처별로 분절되어 운영되는 특성을 반영한 것으로, 보건복지부, 고용노동부, 교육부, 인사혁신처 등 여러 부처의 데이터를 KIHASA가 SOCX 기준에 맞춰 재분류하고 통합하는 방식이다.

특히 2018년 이후 한국은 결산 데이터에 기반한 수집 체계로 전환하여 데이터의 정확성을 크게 향상시켰다. 기획재정부의 한국공공재정정보서비스(KPFIS)의 디지털예산회계시스템(dBrain)과 통합보조금관리시스템(IGSMS), 그리고 행정안전부의 지방재정관리시스템(LFMS)을 연계하여

243개 지방정부의 매칭펀드 프로그램까지 완전히 포괄하는 체계를 구축했다.

#### 4) 데이터 품질 관리 및 업데이트 체계

OECD SOCX는 프로그램 단위의 상세한 기록을 통해 높은 수준의 투명성을 확보하고 있다. 각국은 개별 사회보장 프로그램별로 지출 내역을 보고하며, 이는 총량 데이터의 검증과 재분류를 가능하게 한다. 예를 들어, 캐나다, 네덜란드, 미국의 경우 60~65개의 개별 프로그램으로, 리투아니아는 약 80개, 프랑스는 약 350개의 프로그램으로 세분화하여 관리한다.

이러한 상세한 기록 체계는 중복 계상이나 부적절한 분류를 방지하는 중요한 품질 관리 수단으로 작용한다. 또한 국가별 Country Notes를 통해 각 프로그램의 특성, 변화 이력, 측정 방법 등에 대한 메타데이터를 체계적으로 관리하여 시계열 분석과 국제 비교의 신뢰성을 높이고 있다.

SOCX는 2년마다 정기적으로 업데이트되며, 이 과정에서 과거 데이터의 소급 수정도 동시에 이뤄진다. 1996년 초기 버전 이후 1999, 2000, 2001년에는 매년 업데이트되었으나, 이후 2004, 2007, 2009, 2010, 2012, 2014, 2016, 2018년 판으로 2년 주기 업데이트 체계를 확립했다.

데이터의 시의성 확보를 위해 상세한 프로그램별 데이터가 확정되기 전에 국가 총량 데이터를 활용한 “nowcasting” 기법을 도입하여 최신 2~3년 간의 추계치를 제공하고 있다. 이는 OECD Economic Outlook Database와 EU의 AMECO Database의 거시경제 지표를 활용한 것으로, 정책 분석의 시의성을 높이는 데 기여하고 있다.

## 나. 데이터 구성 및 범위

### 1) 정량적 데이터 구성

OECD SOCX는 현재 38개 OECD 회원국과 브라질, 중국, 인도, 인도네시아, 러시아, 남아프리카공화국 등 주요 비회원국을 포함하여 총 42개국의 사회지출 데이터를 제공한다. 시계열 범위는 1980년부터 최신 연도까지이며, 국가별로 데이터 가용성에 차이가 있다. [그림 3-1]에서 보는 바와 같이 OECD Data Explorer를 통해 사용자는 국가별, 연도별 데이터 가용 현황을 직관적으로 확인할 수 있으며, 다양한 형태로 데이터를 다운로드할 수 있다.

[그림 3-1] 국가별, 연도별 데이터 가용 현황 및 다운로드 옵션

OECD Data Explorer

< Back to the search results

Refine your data selection:

Time period: [1] 616 data points selected in this dataset with:

Unit of measure: Percentage of GDP × Expenditure source: Public × Spending type: Total × Programme type: Total × Time period: Start 2010 ×

Clear all ×

Reference area: [1]

Unit of measure: [1]

Expenditure source: [1]

Spending type: [1]

Programme type: [1]

Price base: [1]

Overview Table Chart

Labels Download Developer API Full screen

### Social expenditure aggregates

The OECD Social Expenditure Database (SOCX) has been developed in order to serve a growing need for indicators of social policy. It includes reliable and internationally comparable statistics on public and mandatory and voluntary private social expenditure at programme level. SOCX provides a unique tool for monitoring trends in aggregate social expenditure and analysing changes in its composition. It covers 38 OECD countries and some accession countries for the period 1980-2021/23 and estimates for aggregates for 2022-24. A Social Expenditure Update can be found under [www.oecd.org/data/data/social-expenditure-database/socx.htm](http://www.oecd.org/data/data/social-expenditure-database/socx.htm). The main social policy areas are as follows: Old age, Survivors, Incapacity-related benefits, Health, Family, Active labor market programmes, Unemployment, Housing, and Other social policy areas. This version also includes estimates of [net total social spending](#) for 2021 for 38 OECD countries. SOCX aggregated data as well as sources and methodology are described in [The OECD SOCX Manual – 2019 edition. A guide to the OECD Social Expenditure Database](#).

Measure: Social expenditure

Topic: Society > Social policy

Number of unfiltered data points: 3060448

Last updated: July 29, 2025 at 12:55:39 AM

You might also be interested in these data:

- Net total social expenditure in % GDP
- Public and private social expenditure
- Public expenditure on disability and sickness cash benefits in % GDP
- Public expenditure on family by type of expenditure (cash and in kind) in % GDP
- Public expenditure on old-age and survivors cash benefits in % GDP

© Social expenditure aggregates © OECD

출처: OECD SOCX. (2025). <https://web.archive.oecd.org/2024-06-24/63248-expenditure.htm>. 2025. 4. 18. 접속

또한 <표 3-2>와 같이 데이터는 다양한 형태로 제공되어 분석 목적에 따른 선택적 활용이 가능하다. 명목 지출액은 각국 화폐 단위와 USD PPP 기준으로 제공되어 규모 파악에 유용하며, 실질 지출액은 기준연도 불변가격으로 환산되어 시계열 분석에 적합하다. GDP 대비 비율은 국가 간 비교의 핵심 지표로 활용되며, 1인당 지출액은 개별 국민의 사회보장 수준을 비교하는 데 유용하다.

한국의 경우 1990년부터 데이터가 제공되기 시작하여 상대적으로 짧은 시계열을 가지고 있으나, 2000년 이후 데이터의 완성도는 OECD 평균 수준에 근접하고 있다. 특히 2016년 이후에는 조세지출 항목이 추가되면서 순사회지출 산정이 가능해졌으며, 이는 한국 사회보장제도의 실질적 보호 수준을 보다 정확히 측정할 수 있게 하는 중요한 개선이다.

<표 3-2> SOCX 데이터 제공 형태

지표 유형	단위	기준연도	활용 목적
명목 지출액	각국, 화폐, USD PPP	해당 연도	규모 파악
실질 지출액	각국, 화폐, USD PPP	2015(예)	시계열 분석
GDP대비 비율	%	-	국가간 비교
1인당 지출액	USD PPP	해당 연도	수준 비교

출처: OECD SOCX(2025) 활용 저자 재작성.

## 2) Country Notes 구성

[그림 3-2]에서 확인할 수 있듯이 SOCX는 정량적 데이터와 함께 각국의 사회보장제도 특성을 설명하는 Country Notes와 메타데이터를 체계적으로 제공한다. <표 3-3>과 같이 Country Notes는 표준화된 구성 요소를 따르되, 각국의 제도적 특수성을 반영한 유연한 기술 방식을 채택하고 있다.

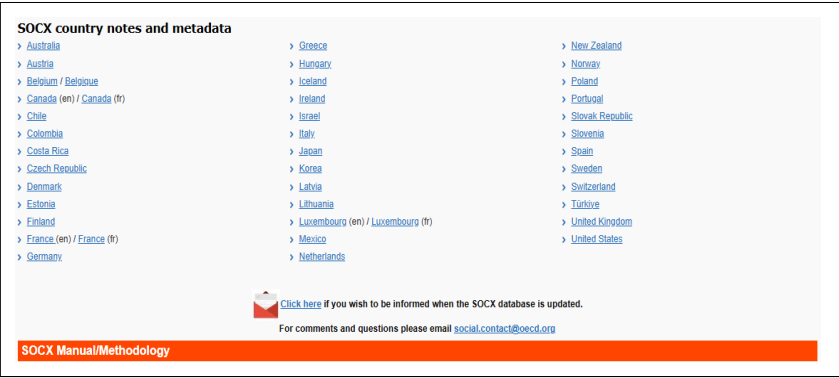
기본 정보 부분에서는 각국의 명명법(Nomenclature), 화폐 단위, 회계 연도 등 데이터 해석에 필요한 기초 정보를 제공한다. 제도 개관에서는 SOCX의 9개 정책영역별로 해당 국가의 주요 제도를 간략히 소개하여 전체적인 사회보장체계의 구조를 파악할 수 있게 한다. 데이터 특이 사항에서는 시계열 단절(Break in series)이나 특별한 추정 방법 등 데이터 활용 시 주의해야 할 사항들을 명시한다. 제도별 상세 분류에서는 SOCX의 세부 프로그램(Code 3, Code 4) 수준에서 개별 제도의 특성을 상세히 설명한다.

〈표 3-3〉 SOCX Country Notes 구성

구성요소	주요 내용
기본 정보	Nomenclature, 화폐단위, 회계연도
제도 개관	9개 정책영역별 주요 제도
데이터 특이사항	Break in series, 추정 방법
제도별 상세분류	SOCX 세부프로그램(Code3, 4별) 제도 설명

출처: OECD SOCX(2025) 활용 저자 재작성.

〔그림 3-2〕 SOCX country notes와 metadata 접근



출처: OECD SOCX. (2025). <https://web-archieve.oecd.org/2024-06-24/63248-expenditure.htm>. 2025. 4. 18. 접속

Country Notes는 3~5년 주기로 업데이트되며, 주요 제도 개혁이 있을 경우 수시로 보완된다. 각국의 SOCX 담당자가 초안을 작성하고 OECD 사무국이 검토하여 최종 확정하는 방식으로 작성된다.

한국의 경우 한국보건사회연구원(KIHASA)이 주관하여 Country Notes를 작성하며, 복수 부처(보건복지부, 고용노동부, 교육부, 인사혁신처 등)의 검토를 거쳐 완성된다. 2018년 결산 기준 데이터 전환 이후 Country Notes의 정확성이 크게 향상되었으며, 243개 지방정부의 매칭 펀드까지 포괄하는 상세한 정보를 제공하고 있다.

## 다. 데이터 분류 체계 및 관리

OECD SOCX는 앞서 2장에서 살펴본 사회정책의 기능, 제공 주체, 급여 형태라는 3차원 분류체계를 실제 데이터베이스에서 체계적인 코딩 구조로 구현하고 있다. 9개 정책 영역, 3개 제공 주체, 2개 급여 형태 등의 이론적 분류가 구체적인 코드 체계로 관리되는 방식은 본 연구의 데이터 베이스 구축에 중요한 시사점을 제공한다.

### 1) 분류 코드 구성

각 사회보장 프로그램은 6개 구성 요소로 이루어진 고유 코드를 통해 체계적으로 관리된다. <표 3-4>에서 보는 바와 같이 한국의 주요 제도들을 예로 들면, 국가코드(410), 제공 주체(10=공공), 정책 영역(1=노령, 5=가족, 8=주거), 급여 형태(1=현금, 2=현물), 세부 프로그램 분류(1.1, 1.2, 1.6 등)로 구성된다.

〈표 3-4〉 주요 사회보장제도 SOCX 코딩 사례

지표 유형	SOCX 코드 한국-410	제공주체	정책영역	급여형태	세부 프로그램
국민연금(노령)	410.10.1.1.1.6	공공(10)	노령(1)	현금(1)	(1,1)
기초연금	410.10.1.1.1.6	공공(10)	노령(1)	현금(1)	(1,6)
영유아보육	410.10.5.2.1.2	공공(10)	가족(5)	현물(2)	(1,2)
주거급여	410.10.8.2.1.1	공공(10)	주거(8)	현물(2)	(1,1)

출처: OECD SOCX(2025) 활용 저자 재작성.

이러한 코딩 체계는 사회적 목적(Social Purpose)과 재분배 또는 강제성(Redistribution or Compulsion)의 두 가지 핵심 기준에 따라 운영된다. 사회적 목적은 개인이나 가계가 직면하는 사회적 위험(노령, 장애, 실업, 질병 등)이나 사회적 필요(아동양육, 주거 등)에 대응하는 것을 의미하며, 재분배 또는 강제성은 개인의 위험 프로필과 무관한 집단 내 자원 재분배나 법률에 의한 참여 의무화를 뜻한다.

코딩 체계의 장점은 데이터베이스의 확장성과 연계성을 보장한다는 점이다. 새로운 제도가 도입되거나 기존 제도가 개편될 경우에도 일관된 분류 원칙을 적용하여 코드를 부여할 수 있으며, 시계열 분석 시 제도 변화의 영향을 추적할 수 있다. 또한 각 코드는 프로그램의 기본 속성뿐만 아니라 시계열 데이터, 관련 법령, 관리기관 정보 등과 연결되어 포괄적 분석을 가능하게 한다.

## 2) 분류 체계 관리

SOCX는 시간의 경과와 제도 변화에도 불구하고 분류의 일관성을 유지하기 위한 체계적 관리 방안을 운영한다. 이는 장기 시계열 분석의 신뢰성을 확보하는 핵심 요소이다. 제도 개편이나 분류 기준 변경이 발생할 경우 ‘Break in series’를 명시하여 연구자들이 시계열 분석 시 주의할

수 있도록 한다.

한국의 경우 대표적인 사례로 2016년 조세지출 항목 추가를 들 수 있다. 이전까지는 직접적인 사회보장 지출만 집계되었으나, 2016년부터 근로장려세제(EITC), 자녀장려세제(CTC) 등 조세지출 형태의 사회보장급여가 포함되면서 순사회지출 산정에 구조적 변화가 발생했다. SOCX는 이를 명시적으로 표시하여 시계열 분석 시 해석상 오류를 방지하고 있다.

또한 제도명이나 관리기관이 변경되는 경우에도 연속성을 유지하기 위한 조치를 취한다. 예를 들어, 한국의 보육료 지원이 ‘무상보육’에서 ‘영유아보육료 지원’으로 명칭이 변경되고 지원 방식이 개편되었을 때도 동일한 코드(410.10.5.2.1.2) 하에서 연속성을 유지하면서 변화 내용을 Country Notes에서 상세히 설명하는 방식을 채택하고 있다.

〈표 3-5〉 SOCX 분류 일관성 관리 사례

관리 영역	방법	적용 사례	목적
제도 개편	Break in series 표시	한국 2016년 조세지출 추가	시계열 단절 명시
신규 제도	기존 체계 내 분류	한국 기초연금 도입 시 노령영역 분류	일관성 유지
정의 변경	소급 적용	사회적 기업 지원 포함/배제	비교 가능성 확보
기준 변경	전체 재산정	SNA 2008 기준 적용	국제 표준 준수

출처: OECD SOCX(2025) 활용 저자 재작성.

이러한 SOCX의 데이터베이스 구조와 관리 체계는 본 연구에서 구축하고자 하는 사회보장 장기재정분석 데이터베이스의 설계에 중요한 시사점을 제공한다. 특히 체계적인 코딩 구조, 다차원적 연계 분석, 시계열 일관성 관리 등은 27개 매트릭스 기반 통합 분석 체계 구축 시 적극 참조해야 할 요소들이다.

## 2. OECD 내부 DB간 연계 구조

SOCX는 단순한 지출 데이터베이스를 넘어 다른 OECD 데이터베이스와의 체계적 연계를 통해 다차원적 분석을 지원한다. OECD는 통합적 데이터 생태계 구축을 위해 SOCX를 중심으로 Health Statistics, Employment Database, Education Database 등과의 연계 분석 기능을 제공하고 있다. 이러한 연계 구조는 본 연구의 통합적 분석 체계 구축에 중요한 시사점을 제공한다.

### 가. Health Statistics와 보건 영역

#### 1) 보건 데이터의 SOCX 연계 구조

SOCX의 보건 정책영역은 의료서비스, 질병 및 장애 급여, 산업재해보상 등을 포괄하는 광범위한 사회보장 지출을 다룬다. 이 영역에서 OECD Health Statistics(보건통계)와의 연계는 특히 의료비 지출의 정확한 측정과 국제 비교에서 핵심적인 역할을 수행한다.

SOCX 보건영역 데이터는 주로 각국의 국가 통계와 행정자료를 통해 수집되지만, 의료비 지출의 세부 분류와 국제 비교 가능성 확보를 위해 OECD Health Statistics의 System of Health Accounts(SHA) 체계와 긴밀히 연계되어 있다. 특히 공공 의료비 지출의 범위 설정과 민간 의료비와의 구분에서 보건통계의 표준화된 분류 체계를 활용한다.

#### 2) 보건통계와의 데이터 연계 방식

OECD 보건통계에서 SOCX로의 데이터 연계는 System of Health Accounts(SHA) 2011 기준을 중심으로 이루어진다. SHA 체계의 재원

조달 분류(HF: Health care Financing schemes)가 SOCX 보건영역의 공공지출 범위 설정에 직접 활용된다.

〈표 3-6〉 SHA 2011의 재원조달 분류

HF 코드	분류명(영문)	SOCX 포함 여부
HF.1	Government financing schemes	포함
HF.1.1	Central government schemes	포함
HF.1.2	State/regional government schemes	포함
HF.1.3	Local/municipal government schemes	포함
HF.2	Compulsory contributory health insurance schemes	포함
HF.2	Social health insurance schemes	포함
HF.2	Compulsory private insurance schemes	포함
HF.3	Voluntary health care payment schemes	제외
HF.3.1	Voluntary health insurance schemes	제외
HF.3.2	Non-profit institutions financing schemes	제외
HF.3.3	Enterprise financing schemes	제외
HF.4	Household out-of-pocket payment	제외

출처: OECD, Eurostat and World Health Organization(2017), Chapter 7 인용 및 재구성.

SHA의 정부 재원조달 방식(HF.1)과 의무사회보험 방식(HF.2)이 SOCX의 공공 의료비 지출 산정 기준이 되며, 자발적 의료보험(HF.3)과 가계 직접부담(HF.4)은 SOCX에서 제외되어 민간 의료비로 분류된다.

### 3) 연계 체계의 의의와 한계

이러한 연계 메커니즘은 보건통계에서 SOCX로의 일방향 데이터 흐름과 상호 검증을 통한 양방향 품질 관리로 구성된다. SHA 체계의 상세한 재원조달 분류를 통해 SOCX 보건영역의 공공지출 범위가 명확히 정의되며, 각국 보건 전문기관에서 수집한 정밀한 의료비 자료를 사회보장 통

계에 활용할 수 있다.

이러한 OECD 내부 데이터베이스 간 연계는 의료비 지출의 복잡한 재원조달 구조를 체계적으로 분류하고, 국제적으로 표준화된 의료비 통계를 사회보장 분석에 활용할 수 있다는 장점이 있다. 또한 보건 분야의 전문성과 사회보장 분야의 정책적 관점을 결합하여 보다 종합적인 분석이 가능하다. 그러나 SHA 체계의 복잡성으로 인한 분류 오류 가능성, 보건 통계 업데이트 주기와 SOCX 작성 일정 간의 불일치, 그리고 국가별 의료제도의 특수성을 충분히 반영하기 어려운 한계 등이 존재한다.

## 나. Employment Database와 ALMP 영역

### 1) ALMP 데이터의 SOCX 연계 구조

SOCX의 적극적 노동시장정책(Active Labour Market Programmes, ALMP) 영역은 실업자와 취업취약계층의 노동시장 참여를 지원하는 다양한 정책적 개입을 포괄한다. 이 영역에서 OECD Employment Database(고용데이터베이스)와의 연계는 특히 ALMP 지출의 정확한 측정과 정책 효과성 분석에 있어 핵심적인 역할을 수행한다.

SOCX ALMP 영역 데이터는 주로 각국의 노동부 및 고용서비스 기관의 행정자료를 통해 수집되지만, ALMP 세부 정책 분류와 국제 비교 가능성 확보를 위해 OECD Employment Database의 Labour Market Policy(LMP) 분류 체계와 긴밀히 연계되어 있다. 특히 공공고용서비스, 직업훈련, 고용보조금 등의 세부 분류에서 고용데이터베이스의 표준화된 분류 체계를 활용한다.

## 2) 고용데이터베이스와의 데이터 연계 방식

OECD 고용데이터베이스에서 SOCX로의 데이터 연계는 Labour Market Policy Database의 분류 체계를 기반으로 이루어진다. 고용데이터베이스의 ALMP 세부 분류가 SOCX ALMP 영역의 정책 유형별 지출 산정에 직접 활용된다.

〈표 3-7〉 OECD Employment Database의 ALMP 분류

번호	분류명(영문)	SOCX 포함 여부
1	Public employment services	직접 연계
2	Training	직접 연계
3	Employment incentives	직접 연계
4	Sheltered and supported employment and rehabilitation	직접 연계
5	Direct job creation	직접 연계
6	Start-up incentives	직접 연계
7	Out-of-work income maintenance and support	SOCX 실업

출처: OECD. (n.d.). OECD Employment Database; European Commission. (n.d.). Labour market policies (LMP)

고용데이터베이스의 LMP 분류 1~6번(공공고용서비스부터 창업 지원까지)이 SOCX의 ALMP 영역에 직접 포함되며, LMP 7번(실업급여)은 SOCX의 실업(Unemployment) 영역으로 별도 분류된다.

## 3) 연계 체계의 의의와 한계

이러한 연계 메커니즘은 고용데이터베이스에서 SOCX로의 일방향 데이터 흐름과 정책 분류의 표준화를 통한 국제 비교성 확보로 구성된다. 고용 전문기관에서 수집한 상세한 ALMP 정책별 지출 자료를 사회보장 통계 체계 내에서 활용할 수 있으며, 각국의 다양한 ALMP 정책들을 표준

화된 분류로 비교 분석할 수 있다.

이러한 OECD 내부 데이터베이스 간 연계는 ALMP 정책의 복잡하고 다양한 프로그램들을 체계적으로 분류하고, 노동시장정책의 전문성을 사회보장 분석에 효과적으로 결합할 수 있다는 장점이 있다. 또한 고용정책과 사회보장정책 간의 연계성을 종합적으로 분석할 수 있는 기반을 제공한다. 그러나 각국의 ALMP 정책 설계와 운영 방식의 차이로 인한 분류의 모호성, 고용데이터베이스 업데이트 주기와 SOCX 작성 일정 간의 시차, 그리고 정책 효과성 측정을 위한 질적 정보의 부족 등의 한계가 존재한다.

## 다. ECEC와 가족 영역

### 1) ECEC 데이터의 SOCX 연계 구조

SOCX는 일반적으로 교육비 지출을 포함하지 않지만, 취학 전 교육 및 보육(Early Childhood Education and Care, ECEC)에 대한 지출은 예외적으로 가족 정책영역에 포함한다. 이는 ECEC가 단순한 교육 서비스를 넘어 부모의 노동시장 참여 지원, 아동 복지 증진, 사회적 형평성 제고 등 복합적인 사회정책 기능을 수행하기 때문이다.

SOCX에서 ECEC 관련 데이터는 두 가지 경로를 통해 수집된다. 첫째, 각국의 국가 통계 및 Eurostat을 통한 직접 수집이며, 둘째는 OECD Education Database(교육DB)를 통한 외삽 방식이다. 특히 후자는 SOCX와 OECD 내부 데이터베이스 간 연계의 대표적 사례로, 교육DB의 ISCED 0 단계(취학 전 교육) 데이터를 활용하여 SOCX의 가족영역 지출을 산출한다.

## 2) 교육DB와의 데이터 연계 방식

OECD 교육DB에서 SOCX로의 데이터 연계는 국제표준교육분류(ISCED) 체계를 기반으로 한다. 교육DB의 ISCED 0단계 데이터가 SOCX 가족영역의 특정 분류 코드와 직접 연결되는 구조이다.

〈표 3-8〉 SOCX의 가족 영역 중 ECEC 분류 예시(한국)

SOCX Code	Title_E	Title_K	정의/설명
410.10.5.2.0.0	Benefits in kind	현물	-
410.10.5.2.1.0	Early childhood education and care(ECEC)	영유아 교육 및 보육	-
410.10.5.2.1.1	Child welfare services	아동복지서비스	방과후, 주말 아동·청소년 돌봄 및 활동 지원
410.10.5.2.1.2	Childcare subsidy (Vouchers for Childcare Service)	영유아보육	어린이집 보육료 및 운영(육아종합지원센터 포함) 지원
410.10.5.2.1.3	Pre-primary education(3-5 years old) (Vouchers for Childcare Service)	보육료 지원바우처	OECD-ECE

출처: 신정우 외. (2023). p.49. 〈표 2-8〉

〈표 3-9〉 ISCED 2011 분류에 따른 취학전 교육단계(ISCED 0)

ISCED 분류	국제분류 명칭		한국 학제
ISCED 0	Early childhood education	영유아교육	유치원, 어린이집, 특수학교(영아과정, 유치원 과정)
ISCED01	Early childhood educational development	유아초기 교육발달 프로그램	0~2세 영유아 대상, 어린이집 등 보육시설(기관)
ISCED02	Pre-primary education	유치과정 교육	3세 이상 초등학교 입학전까지의 아동 대상, 유치원 등

출처: UNESCO Institute for Statistics. (2012)

위 <표 3-8> SOCX의 가족 영역 중 ECEC 관련 분류와 <표 3-9> ISCED 2011 분류에 따른 취학 전 교육단계(ISCED 0)에서 보는 바와 같이, ISCED 01(유아 초기 교육발달, 만 02세)은 SOCX의 아동복지서비스 및 영유아보육 바우처와 연계되며, ISCED 02(취학 전 교육, 만 3세 초등 입학 전)는 SOCX의 취학 전 교육 분류와 직접 연계된다.

### 3) 연계 체계의 의의와 한계

이러한 연계 메커니즘은 교육DB에서 SOCX 방향의 일방향 데이터 흐름으로 이루어지며, ISCED 02 단계 재정지출 데이터가 외삽 방식으로 SOCX에 자동 반영된다. 각국 교육통계 기관에서 수집한 원자료를 OECD가 표준화하여 활용하는 방식이다.

이러한 OECD 내부 데이터베이스 간 효율적 연계는 중복 조사를 방지하고 교육 전문 기관의 정밀한 자료 수집 역량을 사회보장 통계에 활용할 수 있다는 장점이 있다. 또한 국제 비교 가능성 제고를 위한 표준화된 분류 체계를 적용할 수 있다. 그러나 교육DB의 업데이트 지연이 SOCX 작성에 직접적 영향을 미치고, 교육 관점에서 수집된 자료를 사회보장 관점에서 재해석할 때의 개념적 괴리, 그리고 국가별로 상이한 ECEC 제도 특성을 충분히 반영하지 못할 가능성 등의 한계도 존재한다.

## 3. Country Notes 분석

### 가. Country Notes의 구성과 역할

OECD SOCX는 정량적 데이터와 함께 각국의 사회보장제도 특성을 설명하는 Country Notes를 제공한다. Country Notes는 SOCX

Manual에서 언급하는 바와 같이 “정량적 지출 데이터만으로는 파악하기 어려운 각국 사회보장제도의 제도적 특성, 분류 방식, 측정상의 특이 사항”을 기술한 정성적 정보이다.

Country Notes는 OECD.Stat의 메타데이터 정보 항목으로 제공되며, 국가명 옆의 정보 아이콘(i)을 통해 접근할 수 있다. 영어와 프랑스어로 제공되며(국가에 따라 다름), 일반적으로 3~8페이지 분량으로 구성된다.

Country Notes의 표준 구성체계는 다음과 같다. 첫째, 제도 개관(General System Overview)이다. 해당국 사회보장제도의 기본 구조와 특성을 간략히 소개한다. 제도의 역사적 발전 과정, 기본 철학(베버리지형/비스마르크형), 전체적인 적용 범위 등을 포함한다.

둘째, SOCX 분류체계 적용 방식이다. SOCX Manual에서 강조하는 바와 같이, 각국의 제도를 SOCX의 9개 정책영역으로 분류할 때 적용되는 구체적 기준과 방법을 설명한다. 예를 들어, 한 프로그램이 여러 정책 영역에 걸쳐 있을 때의 분류 원칙이나, 국가 특수적 제도의 분류 방식을 명시한다.

셋째, 프로그램별 상세 설명이다. SOCX의 세부 프로그램 코드([그림 3-3] 참조) 수준에서 각 제도의 특성을 기술한다. 법적 근거, 급여 산정 방식, 수급 요건, 재정 방식 등을 포함한다.

[그림 3-3] SOCX country notes 세부프로그램 설명 예시(한국)

COUNTRY NOTE		
Database on Social Expenditure ( <a href="http://www.oecd.org/social/expenditure.htm">www.oecd.org/social/expenditure.htm</a> )		
OECD - Social Policy Division - Directorate of Employment, Labour and Social Affairs		
KOREA		
Code	Title of the programme	Description of the programme and attached notes
I.	OLD AGE	
410.10.1.1.1.1	National pension	Established in 1988, the National Pension is a defined benefit program for the public, combining earnings-related and redistributive components together. - The contribution rate is 9% of the insured person's Standard Monthly Income. The contribution of workplace-based insured persons is equally shared by the employer and the employee, while individually insured persons pay all of their contributions by themselves. - The government's financial support is provided for some portion of contributions paid by farmers, fishermen and low-wage employees. Figures represent monthly NP payments for persons aged 60 and older with more than 10 years of insured period. Pensionable age varies from 60 to 65 depending on the pensioner's year of birth.
410.20.1.1.1.1	Retirement pension	Introduced in 2005 as the second pillar of old-age income security, Korea's retirement pension allows the employer and employee to choose one of the following three pension types: DB (defined benefit), DC (defined contribution), and IRP (individual retirement pension). Insured persons are paid sometime after turning 55 years old.
410.10.1.1.1.3	Government Employees Pension	Introduced in 1960 and funded by the government by legislation, the POP pays monthly benefits to retired public officials aged 60 and over who have fulfilled at least 20 years of service. For officials recruited after 2010, the pensionable age is set at 65.

출처: OECD SOCX. (2025). <https://web-archive.oecd.org/2024-06-24/63248-expenditure.htm>. 2025. 4.18. 접속

넷째, 데이터 출처 및 측정상 특이 사항이다. SOCX Manual에서 설명하는 바와 같이, 각국마다 다른 회계 기준이나 통계 생산 방식으로 인한 측정상의 차이점을 명시한다. 또한 행정 기록, 예산자료, 설문조사 등 데이터 출처별 특성과 한계를 설명한다.

다섯째, 시계열 단절(Break in Series) 설명이다. 주요 제도 개혁이나

측정 방식 변경으로 인해 시계열 연속성이 단절되는 경우, 그 원인과 영향 범위를 상세히 기술한다. SOCX Manual에서 언급하는 바와 같이, 이는 시계열 분석의 정확성을 위해 필수적인 정보이다.

한국의 Country Notes(2018년 기준)를 보면, 급속한 제도 확대 과정에서 나타나는 특성들이 잘 기술되어 있다. 국민연금제도의 성숙 과정, 기초연금 도입(2014년)에 따른 Old-age 영역 내 분류 변화, 건강보험 보장성 강화 정책의 지출 증가 효과 등이 구체적으로 설명되어 있다.

## 나. Country Notes의 정보 활용 가능성

Country Notes는 SOCX의 정량적 데이터 해석에 필수적인 맥락 정보를 제공한다. SOCX Manual에서 강조하는 바와 같이, “동일한 정책영역이라도 국가별로 상이한 제도 설계와 운영 방식”을 이해할 수 있어, 단순한 지출 비교를 넘어선 정책 분석이 가능하다.

구체적인 활용 가능성은 다음과 같다:

첫째, 제도 구조 차이의 정확한 이해이다. 예를 들어, SOCX 데이터에서 독일과 프랑스의 가족정책 지출 수준이 유사하게 나타나더라도, Country Notes를 통해 프랑스는 직접 현금급여(가족수당) 중심인 반면, 독일은 조세감면(child tax allowances)과 서비스 제공이 더 중요한 역할을 한다는 구조적 차이를 파악할 수 있다.

둘째, 정책 개혁의 영향 분석이다. Country Notes는 지출 변화의 배경이 되는 정책 개혁을 구체적으로 설명한다. 스웨덴의 경우 1998년 연금개혁(notional defined contribution system 도입)으로 인한 장기적 지출 궤도 변화를 Country Notes에서 상세히 다루고 있어, 개혁 효과 분석의 기초자료로 활용할 수 있다.

셋째, 측정 방법론상 차이점 보정이다. SOCX Manual에서 지적하는 바와 같이, 각국의 회계 기준 차이나 행정 체계 차이(연방제 vs 단일제)로 인한 측정상 특이 사항을 Country Notes에서 명시한다. 이는 국가 간 비교 분석 시 필수적으로 고려해야 할 요소이다.

넷째, 공사 역할 분담의 복합적 이해이다. SOCX Manual의 Table 1에서 제시하는 public, mandatory private, voluntary private 분류가 각국에서 어떻게 적용되는지를 Country Notes를 통해 파악할 수 있다. 네덜란드의 경우 공적연금 1층(AOW)과 기업연금 2층의 결합 구조, 미국의 경우 사회보장(Social Security)과 401(k) 등 사적연금의 역할 분담 등이 구체적으로 설명되어 있다.

그러나 Country Notes는 비구조화된 서술형 정보로 구성되어 있어 체계적 활용에는 한계가 있다. 국가별로 기술 방식과 상세 수준이 상이하며, 업데이트 주기도 일정하지 않다. SOCX Manual에서도 인정하는 바와 같이, “상세한 프로그램 수준 정보는 약 2~3년의 시차를 갖고 제공되며”, 모든 국가가 동일한 수준의 상세 정보를 제공하지는 않는다.

이러한 한계를 극복하기 위해서는 Country Notes의 정성적 정보를 구조화하고 표준화하는 작업이 필요하다. 본 연구에서 구축하고자 하는 사회보장 장기재정분석 DB에서는 이러한 정성적 정보를 제도 유형별로 체계화하여 정량 분석과 연계할 수 있는 방안을 모색할 필요가 있다.

#### 4. SOCX의 한계 및 정성적 DB 결합 필요성

##### 가. SOCX의 구조적 한계

OECD SOCX는 국제 비교 가능한 사회지출 데이터베이스로서 중요한 역할을 하고 있으나, 장기재정분석을 위한 활용에는 몇 가지 구조적 한계

가 존재한다.

첫째, 지출 중심의 접근법으로 인한 제도 특성 파악의 한계이다. SOCX는 SOCX Manual에서 명시하는 바와 같이 “사회지출(social expenditure)”에 초점을 맞춘 데이터베이스이다. 이는 각국의 사회보장제도가 달성하는 실질적 성과나 제도 설계의 구조적 특성을 직접적으로 보여주지 못한다. 예를 들어, 동일한 수준의 연금 지출이라도 급여 산정 방식(확정급여 vs 확정기여), 재정 방식(부과방식 vs 적립방식), 소득대체율 수준 등에 따라 장기재정 전망은 완전히 달라질 수 있다. 그러나 SOCX 데이터만으로는 이러한 제도 설계상의 차이를 파악하기 어렵다.

둘째, 재원조달 정보의 부재이다. SOCX Manual에서도 인정하는 바와 같이, “SOCX는 사회보장 프로그램의 재원조달(financing)에 대한 정보를 포함하지 않는다.” 이는 장기재정분석에서 치명적 한계로 작용한다. 사회보장제도의 지속가능성은 지출뿐만 아니라 수입 구조와 밀접하게 연관되어 있기 때문이다. 예를 들어, 독일과 프랑스의 연금지출 수준이 유사하더라도, 독일은 기여율과 국고지원의 조합인 반면 프랑스는 상대적으로 높은 기여율에 의존하는 구조적 차이가 있다. 이러한 재원조달 방식의 차이는 인구구조 변화에 대한 제도의 민감도를 결정하는 핵심 요소이다.

셋째, 정책 개혁의 동태적 효과 반영의 어려움이다. SOCX는 특정 시점의 지출 수준을 포착하지만, 정책 개혁이 장기적으로 미치는 영향을 추적하기에는 한계가 있다. SOCX Manual에서 언급하는 “시계열 단절(Break in Series)”이 이를 단적으로 보여준다. 예를 들어, 스웨덴의 1998년 연금개혁이나 독일의 2001년 리스터 연금 도입과 같은 구조적 개혁의 경우, 단기적으로는 지출 감소 효과가 미미하지만 장기적으로는 지출 궤도를 근본적으로 변화시킨다. 그러나 SOCX 데이터만으로는 이러한 개혁의 장기적 파급효과를 예측하기 어렵다.

넷째, 제도 간 상호작용의 복합성 미반영이다. 현대 사회보장제도는 단일 제도가 아닌 다층적 체계로 구성되어 있으며, 제도 간 보완적 또는 대체적 관계를 형성한다. 예를 들어, 네덜란드의 경우 1층 공적연금(AOW) 축소와 2층 기업연금 강화가 동시에 진행되어 전체적인 노후소득보장 수준은 유지되고 있다. 그러나 SOCX의 public과 mandatory private 분류만으로는 이러한 제도 간 역할 분담의 동적 변화를 충분히 포착하기 어렵다.

다섯째, 데이터 가용성과 시의성의 한계이다. SOCX Manual에서 명시하는 바와 같이, “상세한 SOCX 데이터는 사건 발생 후 약 2년의 시차를 갖고 제공된다.” 또한 최근 연도의 데이터는 “nowcasting” 방식으로 추정되어 정확성이 떨어진다. 이는 빠르게 변화하는 정책 환경에서 실시간 분석을 어렵게 만든다.

## 나. 정성적 정보의 중요성

위에서 지적한 SOCX의 구조적 한계를 극복하기 위해서는 정성적 정보와의 결합이 필수적이다. 제1장에서 제기된 “분절적 분석 한계 극복”과 “국제 비교를 통한 정책 방향 설정”이라는 연구 목적을 달성하기 위해서는 정량적 지출 데이터를 보완하는 제도의 구조적 특성 정보가 반드시 필요하다.

첫째, 제도 설계 특성의 체계적 파악이다. 장기재정분석에서 가장 중요한 것은 제도의 구조적 특성이다. 급여 산정 방식, 수급 요건, 재정 방식 등은 인구구조 변화나 경제 여건 변화에 대한 제도의 민감도를 결정한다. 예를 들어, 확정급여형 연금제도와 확정기여형 연금제도는 동일한 기여율이라도 인구 고령화에 대한 재정적 영향이 완전히 다르다. 이러한 제도

설계상의 차이는 정성적 정보를 통해서만 체계적으로 파악할 수 있다.

둘째, 정책 개혁의 맥락과 방향성 이해이다. 사회보장 정책은 단순한 지출 증감이 아니라 사회경제적 변화에 대응하는 구조적 적응 과정이다. 제2장에서 검토한 해외 선진사례들을 보면, 성공적인 장기재정분석은 정책 개혁의 배경과 목표, 예상되는 효과를 종합적으로 고려한다. 예를 들어, 캐나다 OCA의 CPP 분석에서는 단순한 지출 전망을 넘어 제도 개혁이 세대 간 형평성에 미치는 영향까지 분석하고 있다. 이는 정성적 정보에 기반한 정책 맥락 이해가 있어야 가능하다.

셋째, 제도 간 상호작용의 동적 분석이다. 제2장에서 제안한 9×3 매트릭스 분석체계가 효과적으로 작동하기 위해서는 각 제도 유형 특성과 제도 간 연계 메커니즘에 대한 심층적 이해가 필요하다. 예를 들어, 기초연금 도입이 국민연금 가입 인센티브에 미치는 영향이나, 건강보험 보장성 강화가 장기요양보험 이용 패턴에 미치는 영향 등은 정성적 분석을 통해서만 파악할 수 있다.

넷째, 국제 비교의 정확성 제고이다. 제1장에서 지적한 “국제 비교 왜곡” 문제를 해결하기 위해서는 각국 제도의 고유한 특성과 맥락을 정확히 이해해야 한다. 단순한 지출 수준 비교를 넘어, 제도의 목표와 설계 철학, 사회경제적 맥락의 차이를 고려한 비교 분석이 가능해진다.

다섯째, 정책 시나리오 분석의 현실성 제고이다. 장기재정분석의 궁극적 목표는 정책 대안의 효과를 사전에 평가하는 것이다. 이를 위해서는 정책 변화가 제도 전반에 미치는 파급효과를 종합적으로 고려해야 한다. 정성적 정보는 이러한 파급효과의 경로와 규모를 예측하는 데 필수적인 맥락 정보를 제공한다.

따라서 본 연구에서 구축하고자 하는 사회보장 장기재정분석 DB는 SOCX의 정량적 정보와 ISSA, MISSOC 등의 정성적 정보를 체계적으로

결합하여, 단순한 지출 전망을 넘어선 종합적 정책 분석 도구로 발전시켜야 한다. 이를 통해 제1장에서 제기된 기존 분석의 구조적 문제들을 근본적으로 해결할 수 있을 것이다.

## 제2절 주요 사회보장제도 DB 분석

### 1. EU DB

#### 가. ESSPROS

##### 1) 지출 및 재원조달 데이터 구성

ESSPROS의 가장 핵심적인 특징은 사회보장제도의 지출과 재원조달 정보를 통합적으로 제공한다는 점이다. 이는 OECD SOCX가 주로 지출 정보에 집중하는 것과 대조적으로, 사회보장제도의 재정구조와 지속가능성을 종합적으로 분석할 수 있는 정보 기반을 제공한다는 점에서 장기재정분석에 특별한 의미를 갖는다.

지출 측면에서 ESSPROS는 Core System을 통해 제도별 사회보장급여, 행정관리비, 기타 지출로 구분하여 상세한 지출 구조를 파악할 수 있게 한다. <표 3-10>에서 보는 바와 같이, 유형별 지출(spr\_exp\_type), 기능별 지출(spr\_exp\_func), 연금 관련 지출(spr\_exp\_pens) 등으로 세분화되어 장기재정전망에 필요한 기초자료를 체계적으로 제공한다. 특히 급여유형을 현금급여와 현물급여로 구분하고, 소득 조사 여부에 따른 분류를 통해 제도별 재정 특성을 정밀하게 분석할 수 있다.

〈표 3-10〉 ESSPROS Core System - 지출(Expenditure)

분류	세부내용	데이터셋 코드
유형별 지출	사회보장급여, 행정관리비, 기타	spr_exp_type
기능별 지출	8대 기능 분류로 급여 지출 집계, 조기추정 포함	spr_exp_func
연금 관련 지출	기능별로 흩어져 있는 연금 지출(노령·장애·유족·실업)을 전체 연금 지출로 통합 소득조사 여부 구분	spr_exp_pens
교차 분류	기능 × 급여유형(현금/현물) × 소득조사 여부	-
기능별 세부테이블	질병: spr_exp_fsi, 장애: spr_exp_fdi, 노령: spr_exp_fol, 유족: spr_exp_fsu, 가족/아동: spr_exp_ffa, 실업: spr_exp_fun, 주거: spr_exp_fho, 사회적 배제: spr_exp_fex	각 기능별 코드
전체 기능 결합표	모든 기능 통합 분류	spr_exp_ftm
급여유형	현금(Cash) / 현물(In kind)	-
표시 단위	백만 유로, 고정가격(2010/2015 기준), 국가통화, PPS, 1인당, GDP 대비 %, 지출구성비 %	-

출처: ESSPROS reference metadata를 참조하여 작성하였음([https://ec.europa.eu/eurostat/cache/metadata/en/spr\\_esms.htm](https://ec.europa.eu/eurostat/cache/metadata/en/spr_esms.htm)에서 2025.8.9. 인출).

재원조달 측면에서는 〈표 3-11〉에서 보여주듯이 사회기여금, 일반정부 기여금, 기타 수입으로 구분하여 재원 구조를 파악할 수 있다. 더 나아가 기여자 유형(고용주, 보호대상자, 근로자, 자영업자, 연금수급자 등)과 발생 부문(기업, 중앙정부, 주·지방정부, 가계, 비영리, 해외)별 분류를 통해 사회보장제도의 재정적 의존 구조를 상세히 분석할 수 있다. 이러한 정보는 인구 고령화에 따른 재정 압박 상황에서 고용주 기여금, 피보험자 기여금, 정부 기여금 등의 구성비와 변화 추이를 통해 제도별 재정적 지속가능성을 평가하는 데 필수적이다.

〈표 3-11〉 ESSPROS Core System – 수입(Receipts)

분류	세부내용	데이터셋 코드
수입 유형	사회기여금 / 일반정부 기여금 / 기타	spr_rec_sumt
기여자 유형	고용주, 보호대상자(피보험자), 근로자, 자영업자, 연금수급자, 기타 보호대상자	spr_rec_scon
발생 부문	기업, 중앙정부, 주·지방정부, 가계, 비영리, 해외	spr_rec_sums
유형 × 부문	수입 유형과 제도/제도부문 결합 분류	spr_rec_suts
표시 단위	금액(유로/국가통화/고정가격/PPS/1인당), GDP 대비 %, 총수입 대비 %	-

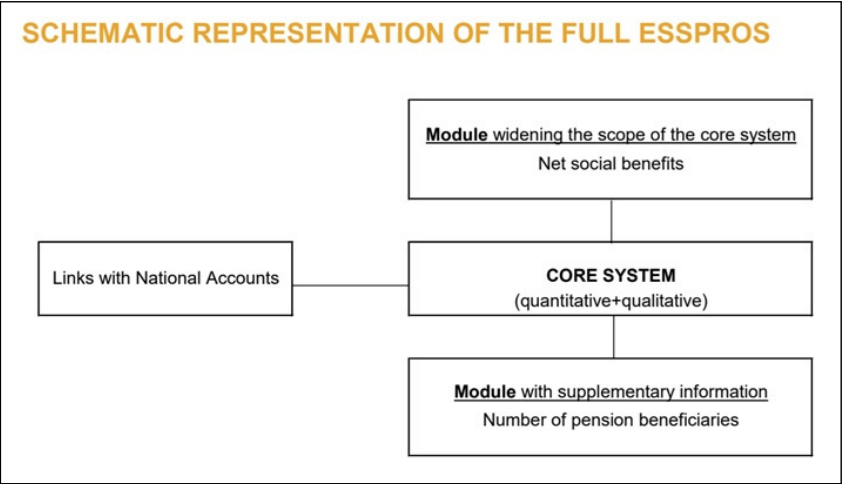
출처: ESSPROS reference metadata를 참조하여 작성하였음([https://ec.europa.eu/eurostat/cache/metadata/en/spr\\_esms.htm](https://ec.europa.eu/eurostat/cache/metadata/en/spr_esms.htm)에서 2025.8.9. 인출).

## 2) 8개 기능별 분류와 세부 항목

ESSPROS의 8개 기능별 분류 체계는 장기재정분석에서 특히 중요한 의미를 갖는다. 질병/의료, 장애, 노령, 유족, 가족/아동, 실업, 주거, 사회적 배제 등 8개 기능으로 구분된 분류는 인구구조 변화와 사회경제적 변화에 따른 기능별 재정 압박 정도를 차별적으로 분석할 수 있게 한다.

[그림 3-4]에서 확인할 수 있듯이, Core System을 중심으로 한 모듈별 세분화는 장기재정분석의 정밀도를 높인다. 연금 수급자 모듈(PENS)은 7가지 연금 유형별 수급자 수를 제공하여 인구 고령화에 따른 연금 지출 전망의 기초자료를 제공하며, 순 사회보장급여 모듈(NET)은 조세와 사회기여금을 차감한 순급여를 통해 실질적인 재정 부담을 평가할 수 있게 한다.

[그림 3-4] ESSPROS의 구성



출처: Eurostat. (2022, p.9).

특히 기능별 교차 분류를 통해 <표 3-7>의 기능별 세부 테이블과 같이 각 기능에 대해 현금급여와 현물급여, 소득 조사 여부별로 세분화된 분석이 가능하다. 이는 노령화로 인한 노령 기능 지출 증가, 만성질환 증가에 따른 장애 기능 지출 변화, 저출산에 따른 가족/아동 기능 정책 효과성 등을 기능별로 차별화하여 장기전망할 수 있는 기반을 제공한다.

### 3) ESSPROS와 SOCX의 관계

ESSPROS와 SOCX 간의 관계는 단순한 병렬적 비교를 넘어서 실제 데이터 연계와 변환 과정을 통해 이루어진다. 특히 유럽 25개국(EU 22개국 + 아이슬란드, 노르웨이, 스위스)의 사회지출 자료는 ESSPROS 데이터베이스를 기반으로 SOCX에 제공되며, 이 과정에서 체계적인 데이터 변환이 이루어진다.

[그림 3-5]에서 보는 바와 같이, ESSPROS의 사회보장급여가 SOCX로

전환되는 과정은 단순한 이관이 아니라 복잡한 재분류 과정을 거친다. ESSPROS는 EU 회원국의 사회보장제도에 대한 포괄적 정보를 제공하지만, SOCX는 OECD 전체 회원국과의 비교 가능성을 위해 일부 항목을 제외하고 다른 OECD 데이터베이스의 정보를 추가하는 방식으로 재구성된다.

[그림 3-5] ESSPROS에서 SOCX로의 전환

**Table A.I.1.1a From ESSPROS to SOCX: a brief overview**

ESSPROS		SOCX
1. Sickness/Health care cash	->	3. Incapacity
services	->	4. Health (OECD Health data) (ILTC overlap)
2. Disability Economic integration of the handicapped	->	
3. Old age	->	1. Old age
4. Survivors	->	2. Survivors
5. Family/children	->	5. Family Child care (pre-primary education from OECD Education database)
6. Unemployment cash	->	7. Unemployment
Vocational training allowance	->	
services	->	
7. Housing	->	8. Housing
8. Social exclusion	->	9. Other social policy areas
	->	6. ALMPs (OECD LMP database)

출처: Adema, W., & Fron, P. (2019, p.42).

구체적인 ESSPROS → SOCX 재분류 공식은 다음과 같다:

- (1) ESSPROS 사회보장급여
  - (2) 장애인의 경제통합 관련 지출
  - (3) 질병 관련 현물급여 (의료서비스)

- (4) 실업 관련 직업훈련 수당 및 서비스
- + (5) 보건서비스(OECD 보건통계에서)
- + (6) 유아보육/유아교육(OECD 교육통계에서, 일부 국가 대상)
- + (7) 적극적 노동시장정책(OECD ALMP 통계에서)
- = (11) SOCX 총 사회지출

이러한 변환 과정은 장기재정분석 관점에서 중요한 의미를 갖는다. ESSPROS의 강점인 자원조달 정보와 제도별 세분화된 지출 정보가 SOCX의 국제 비교 가능성과 결합됨으로써, 유럽 국가들의 장기재정분석에서 보다 정밀하고 포괄적인 분석이 가능해진다. 특히 ESSPROS에서 제외되지만 SOCX에 포함되는 보건서비스, ECEC, ALMP 지출 정보는 1절에서 살펴본 OECD 내부 데이터베이스 간 연계 구조를 통해 보완되어, 유럽 국가들의 사회보장 지출에 대한 완전한 그림을 제공한다.

또한 ESSPROS의 의무적 민간사회지출과 임의적 민간사회지출 정보가 SOCX에 추가됨으로써, 공공지출 중심의 ESSPROS 분석을 민간 영역까지 확장할 수 있다. 이는 특히 유럽 국가들의 혼합형 사회보장제도 하에서 공공-민간 간 역할 분담의 장기적 변화를 분석하고 전망하는 데 필수적인 정보를 제공한다.

#### 4) EU 회원국 사회정책 조정 도구로서의 역할

ESSPROS는 단순한 통계체계를 넘어서 EU 거버넌스와 직접 연계된 정책 조정 도구로 기능한다는 점에서 장기재정분석에 특별한 가치를 지닌다. European Semester, Ageing Report, Social Protection Committee의 연례 보고서 등 EU 차원의 정책조정 과정에서 ESSPROS 데이터가 핵심 근거 자료로 활용된다.

[그림 3-6] EUROSTAT 홈페이지 Database 화면 예시



출처: Eurostat 홈페이지(<https://ec.europa.eu/eurostat/>)에서 2025. 8. 9. 인출).

[그림 3-6]에서 보듯이, ESSPROS는 상세 데이터셋(Detailed datasets)과 선택 데이터셋(Selected datasets)을 구분하여 제공함으로써 심층 분석과 정책 모니터링을 동시에 지원한다. 조기 추정치(Early Estimates) 제공을 통해 최신성을 확보하고, [그림 3-7]과 같은 시각화 도구를 통해 정책 담당자들이 쉽게 접근할 수 있도록 한다.

[그림 3-7] EUROSTAT 홈페이지 Data visualisations 화면 예시



출처: Eurostat 홈페이지(<https://ec.europa.eu/eurostat/>)에서 2025. 8. 9. 인출).

이러한 정책 조정 도구로서의 역할은 장기재정분석의 정책 활용성을 크게 높인다. EU 차원에서 표준화된 분류 체계와 품질 관리 체계를 통해 회원국 간 일관성 있는 비교 분석이 가능하며, 이는 개별 국가의 장기재정분석 결과를 유럽 전체 맥락에서 해석하고 정책적 시사점을 도출하는데 중요한 기준점을 제공한다. 특히 <표 3-9> Detailed datasets와 Selected datasets의 비교에서 확인되듯이, 연구자와 정책 담당자의 서로 다른 요구를 동시에 충족시킴으로써 장기재정분석의 정책 연계성을 강화한다.

〈표 3-12〉 Detailed datasets와 Selected datasets의 비교

구분	Detailed dataset	Selected dataset
목적	심층 분석, 세부 분류별 맞춤 분석	주요 지표 제공, 간단 비교·보고용
차원 수	다차원(기능, 급여형태, 성별, 연령, 소득집사여부 등)	단일 또는 소수 차원
이용 난이도	높음(코드·분류 이해 필요)	낮음(바로 사용 가능)
형태	원자료에 가까운 테이블	가공·집계된 요약 테이블
대상 사용자	연구자·분석가	정책담당자·언론·일반 이용자

출처: 저자 작성.

## 나. MISSOC

### 1) 31개국 제도별 정성적 정보 구조

MISSOC는 현재 EU 27개 회원국과 EEA 소속 아이슬란드, 리히텐슈타인, 노르웨이, 그리고 스위스를 대상으로 하며 12개 주요 사회보장 영역에서 300개 이상의 세부 카테고리를 통해 정성적 정보를 체계화하는 독특한 데이터베이스 구조를 갖고 있다. 이는 단순한 분류체계를 넘어서 장기재정분석에 필요한 제도적 세부 사항을 포착하는 정교한 메타데이터 시스템으로 기능한다.

MISSOC의 정보 구조는 크게 세 가지 계층으로 구성된다. 첫째, 비교표(Comparative Tables)는 법적 근거, 적용 범위, 조건, 급여 수준 등의 카테고리로 세분화되어 제도의 핵심 요소들을 표준화된 형식으로 제공한다. 둘째, 조직 구조(Organisation)는 각국의 사회보장 구조를 조직도와 간략한 텍스트로 설명하며, 사회보장 담당 부처와 주요 기관의 인터넷 주소를 포함한다. 셋째, 사회보장 가이드(Social Security Guides)는 유럽 내 이주하는 시민들을 대상으로 접근 가능하고 이해하기 쉬운 언어로 국가별 사회보장제도를 설명한다.

장기재정분석 관점에서 특히 중요한 것은 MISSOC의 연 2회(1월 1일, 7월 1일) 정기 업데이트 체계이다. 이는 제도 변화의 실시간 추적을 가능하게 하며, 장기재정전망에서 제도 개혁의 영향을 시의적절하게 반영할 수 있는 기반을 제공한다. 또한 자영업자 전용 섹션을 별도로 운영하여 전통적인 임금근로자 중심 사회보장제도에서 포착되지 않는 고용 형태의 변화를 체계적으로 추적한다.

## 2) 표준화된 비교 분석 체계와 SOCX 연계 가능성

MISSOC의 12개 카테고리 체계는 OECD SOCX의 9개 정책영역과 상당한 유사성을 보이면서도 EU 특성을 반영한 차별화된 접근을 취한다. <표 3-13>에서 보는 바와 같이, MISSOC의 Financing, Old-Age, Survivors, Invalidity, Family Benefits, Unemployment 등은 SOCX와 직접적으로 대응되는 반면, Long-term Care, Guaranteed Minimum Resources는 EU의 사회보장 발전 수준을 반영한 고유 카테고리이다.

이러한 매핑 관계는 장기재정분석에서 MISSOC의 정성적 정보와 SOCX의 정량적 데이터를 결합한 통합 분석의 가능성을 제시한다. 특히 MISSOC에서만 제공되는 Financing 정보는 SOCX의 지출 중심 분석을 재원조달 측면에서 보완할 수 있으며, Long-term Care와 Guaranteed Minimum Resources는 미래 사회보장 지출 증가의 새로운 동인을 포착하는 데 중요한 역할을 한다.

〈표 3-13〉 MISSOC-SOCX 카테고리 매핑

MISSOC 카테고리	SOCX 정책영역	비고
Financing	-	SOCX 정보 없음
Health Care	보건	직접 대응
Sickness - Cash Benefits	보건	현금 급여
Maternity/Paternity	가족	가족 하위 범주
Invalidity	근로무능력	직접 대응
Old-Age	노령	직접 대응
Survivors	유족	직접 대응
Accidents at Work and Occupational Diseases	근로무능력	부분 대응
Family Benefits	가족	직접 대응
Unemployment	실업	직접 대응
Guaranteed Minimum Resources	기타	부분 대응
Long-term Care	보건/기타	부분 대응

출처: 저자 작성.

MISSOC에서는 [그림 3-8], [그림 3-9]의 MISSOC의 Comparative Table 기능에서 확인할 수 있듯이, 자유롭게 국가 및 카테고리 선택을 통한 맞춤형 비교 분석이 가능하다. MISSOC은 사회보장 제도의 철학 등과 같은 일반적인 원칙부터 12번 장기 요양에 관한 내용을 카테고리 체계에 맞게 분류하여 정보를 제공하고 있으며, 해당 기능을 통해 각각의 주제별 제도가 국가별로 어떻게 운영되고 있는지 확인할 수 있다. 이는 장기재정 분석에서 특정 제도 영역이나 국가군에 대한 심층 분석을 수행할 때 매우 유용한 기능이다.

[그림 3-8] MISSOC Comparative Table

Update as at 2025-01-01

Please select one or more countries

(check/uncheck to add/remove countries)

All countries

All EU countries

☒ Austria

☒ Belgium

☒ Bulgaria

☒ Croatia

☒ Cyprus

☒ Czech Republic

☒ Denmark

☒ Estonia

☒ Finland

☒ France

☒ Germany

☒ Greece

☒ Hungary

☒ Ireland

☒ Italy

☒ Latvia

☒ Lithuania

☒ Luxembourg

☒ Malta

☒ Poland

☒ Portugal

☒ Romania

☒ Slovakia

☒ Slovenia

☒ Spain

☒ Sweden

☒ The Netherlands

All EFTA countries

☒ Iceland

☒ Liechtenstein

☒ Norway

☒ Switzerland

Please select one or more topics

(check/uncheck to add/remove topics)

All tables

General principles

I. Financing

II. Health care

III. Sickness - cash benefits

IV. Maternity/Paternity

V. Invalidity

VI. Old-age

VII. Survivors

VIII. Accidents at work and occupational diseases

IX. Family benefits

X. Unemployment

XI. Guaranteed minimum resources

XII. Long-term care

Select the format

☒ HTML

☐ Excel

☐ flip Excel

DISPLAY RESULTS

출처: MISSOC. (n.d.). Comparative tables (Retrieved 2025. 7. 31., from <https://www.missoc.org/missoc-database/comparative-tables/>)

[그림 3-9] MISSOC Comparative Table Result

출처: MISSOC. (n.d.). Comparative tables (Retrieved 2025. 7. 31., from <https://www.missoc.org/missoc-database/comparative-tables/>)

### 3) 제도 변화 추적과 장기재정분석 활용성

MISSOC의 가장 큰 강점 중 하나는 1990년 설립 이래 지속적인 정보 교환을 통해 축적된 장기 시계열 정보이다. 각국의 사회보장제도 변천사와 주요 개정 사항을 체계적으로 기록하고 있어, 제도 개혁의 패턴과 방향성을 파악하여 장기전망에 반영할 수 있다.

특히 MISSOC는 공공기관, 전문 사용자, 일반 시민을 모두 대상으로 하는 중앙 데이터베이스로 발전하면서, 다층적 사용자 요구를 충족하는 정보 체계를 구축했다. 이는 장기재정분석에서 정책 입안자, 연구자, 시민 사회 등 다양한 이해관계자 간의 소통과 합의 형성에 중요한 역할을 한다.

MISSOC과 SOCX의 연계 활용은 세 가지 차원에서 장기재정분석의 정확성과 유용성을 높인다. 첫째, MISSOC의 제도 구조 정보는 SOCX의 지출 데이터 해석에 필요한 맥락을 제공한다. 예를 들어, 동일한 노령연

금 지출 수준이라도 기여제와 비기여제의 비중, 급여 산정 방식의 차이에 따라 인구 고령화의 재정 영향이 달라질 수 있다. 둘째, MISSOC의 최신 제도 개혁 정보는 SOCX 기반 장기전망의 가정을 업데이트하는 데 활용할 수 있다. 셋째, MISSOC의 EU 특화 정보(Long-term Care, Guaranteed Minimum Resources 등)는 SOCX에서 포착되지 않는 새로운 사회보장 지출 동인을 식별하고 전망하는 데 기여한다.

이러한 MISSOC-SOCX 연계 분석은 유럽 국가들의 장기재정분석에서 정량적 정확성과 정성적 타당성을 동시에 확보할 수 있는 중요한 방법론적 기반을 제공한다.

## 다. EU DB 특징 및 한계

### 1) SOCX의 자원 조달 정보 보완

ESSPROS와 MISSOC로 대표되는 EU 데이터베이스의 가장 큰 강점은 OECD SOCX가 상대적으로 취약한 자원 조달 정보를 체계적으로 제공한다. 장기재정분석에서 지출 전망만큼 중요한 것이 자원 조달 구조의 지속가능성 분석인데, EU 데이터베이스는 이러한 요구를 충족시키는 핵심적 역할을 한다.

ESSPROS는 사회기여금, 일반정부 기여금, 기타 수입으로 구분된 수입 분류를 통해 사회보장제도의 재정구조를 정밀하게 파악할 수 있게 한다. 특히 고용주, 피보험자, 자영업자, 연금수급자 등 기여자 유형별 분류와 기업, 중앙정부, 지방정부, 가계 등 발생 부문별 분류를 결합하여, 인구구조 변화가 자원 조달에 미치는 영향을 세밀하게 분석할 수 있다. 이는 SOCX의 지출 중심 접근법을 재정균형 관점에서 보완하는 중요한 정보를 제공한다.

MISSOC의 Financing 카테고리는 정성적 관점에서 각국의 재원 조달 철학과 구조를 심층적으로 이해할 수 있게 한다. 조세 기반과 사회보험 기반의 제도적 차이, 중앙집중형과 지방분산형의 관리 체계 차이 등은 동일한 인구 고령화 압박에도 불구하고 국가별로 상이한 재정 대응 능력을 갖게 하는 핵심 요인들이다. 이러한 정보는 장기재정분석에서 단순한 지출 증가율 적용을 넘어서 제도적 맥락을 고려한 차별화된 전망을 가능하게 한다.

## 2) 정량-정성 정보 결합의 분석적 강점

EU 데이터베이스의 두 번째 특징은 ESSPROS의 정량적 데이터와 MISSOC의 정성적 정보가 상호 보완하여 제공하는 분석적 강점을 제공할 수 있다는 점이다. 이는 SOCX와 같은 단일 데이터베이스로는 달성하기 어려운 통합적 분석 기반을 제공한다.

ESSPROS의 기능별, 제도별 지출 데이터는 MISSOC의 제도 구조 정보와 결합될 때 비로소 완전한 의미를 갖는다. 예를 들어, 노령 지출의 증가 추이를 분석할 때 ESSPROS의 수치 데이터만으로는 증가 원인과 향후 지속가능성을 판단하기 어렵다. 그러나 MISSOC의 연금제도 구조 정보(급여 산정 방식, 수급 요건, 재원 조달 방식 등)와 결합하면, 지출 증가의 구조적 원인을 파악하고 제도 개혁의 필요성과 방향을 도출할 수 있다.

특히 장기재정분석에서 중요한 것은 제도 개혁의 가능성과 효과를 예측하는 것인데, MISSOC의 제도 변화 추적 정보는 이러한 분석에 필수적인 기반을 제공한다. 과거 개혁 사례와 현재 개혁 논의 동향을 통해 미래 제도 변화를 예측하고, 이를 ESSPROS 기반 재정 전망에 반영할 수 있다.

### 3) EU 회원국 범위의 제약과 확장성 한계

다만 EU 데이터베이스의 가장 근본적인 한계는 지리적 범위가 EU 회원국과 일부 유럽 국가로 제한된다는 점이다. ESSPROS는 EU 27개국과 EEA 3개국, 그리고 스위스를 포함하여 총 31개국을 대상으로 하며, MISSOC도 동일한 범위를 갖는다. 이는 OECD SOCX의 38개 회원국과 일부 비회원국을 포괄하는 범위에 비해 상당히 제한적이다.

이러한 범위의 제약은 장기재정분석의 국제 비교 가능성을 제한한다. 특히 아시아-태평양 지역의 주요 국가들(일본, 한국, 호주, 뉴질랜드 등)과의 비교 분석이 불가능하며, 북미 지역(미국, 캐나다)과의 비교도 제한된다. 이는 사회보장제도의 다양한 모델과 발전 경로를 종합적으로 비교 분석하는 데 구조적 한계를 초래한다.

또한 EU 회원국들은 European Semester, Open Method of Coordination 등을 통해 사회정책 조정과 수렴을 추진하고 있다. 이러한 EU 차원의 정책 조정 체계는 회원국들의 사회보장제도가 유사한 방향으로 발전하도록 하는 수렴 압력을 형성한다. 예를 들어, 연금 개혁에서 지속가능성 강화, 활성화 정책에서 근로 인센티브 제고 등이 공통된 개혁 방향으로 나타난다. 따라서 EU 데이터베이스가 포착하는 제도 변화 패턴은 이러한 정책 조정의 영향을 받은 것이며, 이는 미국, 일본, 한국 등 다른 지역 국가들의 자율적 제도 발전 경로와는 다른 양상을 보일 수 있다. 이러한 특성은 EU 데이터베이스 기반 분석 결과를 다른 지역에 적용할 때 신중한 해석이 필요함을 의미한다.

더 나아가 Brexit와 같은 정치적 변화는 EU 데이터베이스의 연속성과 일관성에 영향을 미칠 수 있다. 영국의 MISSOC 제외는 장기 시계열 분석에서 구조적 단절을 초래하며, 향후 유사한 정치적 변화가 발생할 경우 데이터베이스의 안정성에 추가적인 위험 요인이 될 수 있다.

이러한 한계에도 불구하고 EU 데이터베이스는 유럽 지역의 사회보장 제도 분석에서는 매우 유용한 정보를 제공하므로, 장기재정분석의 목적과 대상 지역에 따라 적절한 데이터베이스를 선택하거나 여러 데이터베이스를 결합하여 활용하는 접근이 필요하다.

## 2. ILO SSI DB

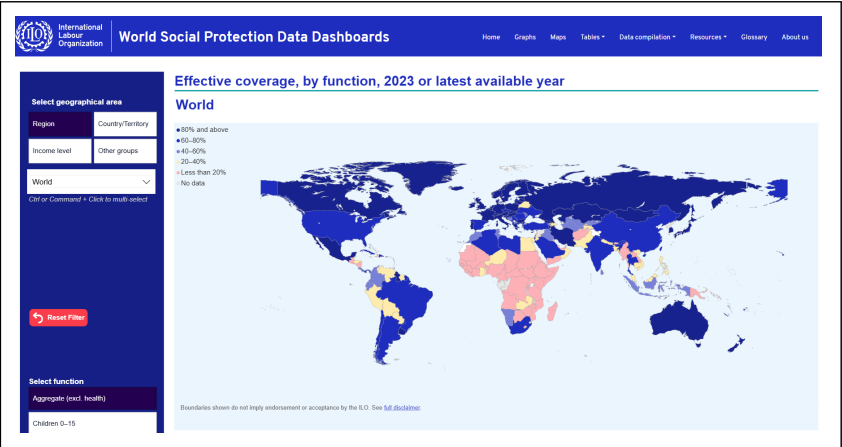
### 가. SSI 데이터 범위 및 구조

#### 1) 전 세계 190여 개국 포괄 범위

ILO Social Security Inquiry(SSI)의 가장 큰 특징은 전 세계 190여 개국을 포괄하는 광범위한 지리적 범위다. 이는 OECD SOCX의 38개국, EU ESSPROS의 31개국과 비교할 때 그 범위가 크게 넓고, 장기재정분석에서 글로벌 차원의 비교 분석과 개발도상국의 사회보장제도 발전 경로 예측을 가능하게 한다.

SSI는 아동, 실업자, 고령자, 장애인, 출산 여성, 산업재해 피해자, 빈곤층과 취약계층에 대한 사회보장 적용 범위 데이터를 수집하는 주요한 종합 도구로 기능한다. 특히 OECD 세계 밖에서의 (비교 가능한) 사회보장 통계 부족 문제를 해결하기 위한 목적으로 설계되어, 선진국 중심의 기존 데이터베이스가 포착하지 못하는 개발도상국의 사회보장제도 현실을 체계적으로 파악할 수 있다.

[그림 3-10] ILO World Social Protection Database



출처: MISSOC. (n.d.). Comparative tables(Retrieved 2025. 7. 31., from <https://www.missoc.org/missoc-database/comparative-tables/>)

장기재정분석 관점에서 이러한 광범위한 포괄성은 세 가지 중요한 의미를 갖는다. 첫째, 사회보장제도의 다양한 발전 모델과 경로를 비교 분석할 수 있어, 개별 국가의 미래 발전 방향을 예측하는 데 유용한 참조점을 제공한다. 둘째, 저소득 국가에서 고소득 국가까지의 발전 단계별 사회보장 지출 패턴을 파악하여, 경제발전에 따른 사회보장 지출 증가의 장기적 궤적을 분석할 수 있다. 셋째, 지역별, 소득수준별 차이를 고려한 차별화된 장기전망이 가능하다.

## 2) 11개 기능별 분류 체계

SSI의 11개 기능별 분류 체계는 SOCX의 9개 정책영역, ESSPROS의 8개 기능보다 더 세분화되고 포괄적인 접근을 취한다. 전통적인 사회보험 8개 영역(노령, 장애, 유족, 질병 및 의료, 실업, 고용상해 및 산업재해, 가족/아동, 모성)에 추가로 주거, 기초교육, 기타 소득보장 및 공공부조/

사회적 배제 등 3개 영역을 포함하여 사회보장의 개념을 확장했다.

〈표 3-14〉 SSI 기능별 3차원 데이터 수집 구조(예시: 노령, 실업, 기초교육)

기능	주요지표		
	지출	재원조달	포괄성
노령	연금급여 지출 조기퇴직 지출 최저연금 지출 관리운영비	고용주 기여율 근로자 기여율 정부보조금 기금운용수익	법정 수급연령 실제 수급연령 소득대체율 연금수급자 수 가입자 대비 수급률
실업	실업급여 지출 적극적 노동시장정책 직업훈련 지출 구직활동 지원	고용보험료율 정부 재정지원 고용안정기금 EU 기금(해당시)	급여 지급기간 소득대체율 적용 근로자 범위 실업급여 수급률 구직 성공률
기초교육	초등교육 지출 급식비 지원 교재비 지원 교통비 지원	정부 예산 지방정부 분담 국제원조 민간 기부	취학연령 인구 실제 취학을 중도탈락률 교육비 부담률 지역별 격차

출처: 저자 작성.

SSI의 독특한 강점은 기능별로 이처럼 세밀한 3차원 데이터를 체계적으로 수집한다는 점이다. 예를 들어 노령 기능의 경우 단순한 연금 지출 총액뿐만 아니라 조기퇴직, 최저연금 등 세부 프로그램별 지출과 고용주-근로자 기여율, 기금운용수익 등 재원조달 구조, 그리고 법정/실제 수급연령, 소득대체율, 수급률 등 보장성 지표까지 포괄적으로 파악한다.

이러한 확장된 분류 체계는 장기재정분석에서 중요한 의미를 갖는다. 기초교육의 포함은 인적자본 투자와 사회보장의 연계를 보여주며, 이는 장기적 경제성장과 사회보장 지출의 상호작용을 분석하는 데 필수적이다. 주거 영역의 독립적 분류는 도시화와 주택 가격 상승이 사회보장 지출에 미치는 영향을 체계적으로 추적할 수 있게 한다.

### 3) 지출-재원조달-보장성 3차원 데이터

SSI의 또 다른 주요 특징은 지출, 재원조달, 포괄성을 통합한 3차원 데이터 구조를 갖고 있다는 것이다. 이는 SOCX의 지출 중심 접근, ESSPROS의 지출-재원조달 2차원 접근을 넘어서 사회보장제도의 성과와 지속가능성을 종합적으로 평가할 수 있는 기반을 제공한다.

지출 차원에서는 총 사회보장지출, 급여지출, 관리비용, 기타 지출로 구분하여 비용 구조를 정밀하게 분석할 수 있다. 재원조달 차원에서는 기여금, 정부 기여금, 투자수익, 기타 수입으로 분류하여 재원 구조의 지속가능성을 평가할 수 있다. 보장성 차원에서는 보호인구 수, 수급자 수, 평균급여 수준, 보장률 등을 통해 사회보장제도의 실질적 효과성을 측정할 수 있다.

장기재정분석에서 이러한 3차원 접근은 특별한 가치를 갖는다. 단순한 지출 증가 전망을 넘어서 보장률 확대, 급여 적절성 개선, 재원조달 구조 변화 등을 종합적으로 고려한 시나리오 분석이 가능하다. 특히 개발도상국의 경우 사회보장제도 확대 과정에서 적용범위 확장과 급여 개선이 동시에 진행되므로, 3차원 데이터는 이러한 복합적 변화의 재정적 영향을 정확히 예측하는 데 필수적이다.

또한 3년 주기로 발간되는 ILO 플래그십 World Social Protection Report는 이러한 3차원 데이터를 기반으로 한 종합적 분석 결과를 제공하여, 장기재정분석의 국제적 맥락과 벤치마킹 기준을 제시한다. 이는 개별 국가의 장기재정분석 결과를 글로벌 트렌드와 비교하고 정책적 시사점을 도출하는 데 중요한 역할을 한다.

〈표 3-15〉 SSI 3차원 데이터 구조

차원	주요 지표	세부 내용
지출	총 사회보장지출	모든 사회보장제도 지출 합계
	급여지출	수급자 직접 지급 급여
	관리비용	제도 운영 행정비용
재원조달	기여금	고용주 및 피보험자 기여금
	정부 기여금	중앙 및 지방 정부 기여금
	투자수익	기금 운용 수익
포괄성	적용대상 인구	제도 보장 대상
	수급자 수	실제 급여 수급자
	평균 급여수준	수급자 1인당 평균 급여
	보장률	보호대상 대비 실제 보장 비율

출처: 저자 작성.

## 나. SSI의 특징 및 한계

### 1) 다양한 제도 발전 경로의 비교 분석 강점

SSI의 가장 독특한 강점은 전 세계 190여 개국을 포괄하여 사회보장제도의 다양한 발전 모델과 경로를 종합적으로 분석할 수 있는 유일한 데이터베이스라는 점이다. 이는 OECD SOCX나 EU 데이터베이스가 제도 성숙 국가 중심으로 구성된 것과 대조적으로, 제도 발전 단계별 특성과 패턴을 비교 분석할 수 있게 한다.

장기재정분석 관점에서 이러한 다양성 포괄은 세 가지 중요한 가치를 제공한다. 첫째, 압축 성장을 경험한 국가들의 사회보장제도 발전 패턴을 분석할 수 있다. 한국, 대만, 싱가포르 등 단기간에 급속한 경제성장과 사회보장제도 확대를 동시에 경험한 국가들의 사례를 통해, 경제발전과 사회보장 지출 확대의 관계를 실증적으로 분석할 수 있다. 이는 제도 도입 초기 단계에서 급속한 적용 범위 확대와 급여 개선이 재정에 미치는 영향

을 예측하는 데 중요한 근거를 제공한다.

둘째, 점진적 제도 발전 국가들과의 비교를 통해 서로 다른 발전 경로의 장단점을 평가할 수 있다. 서유럽의 점진적 발전 모델과 동아시아의 압축 성장 모델 간의 차이를 분석하여, 각국의 역사적·사회적 맥락에 적합한 장기전망 시나리오를 구성할 수 있다. 특히 제도 성숙 속도와 재정 부담 증가 패턴 간의 관계를 파악하는 데 유용하다.

특히 SSI가 비공식부문 사회보장까지 포괄한다는 점은 제도 확대 과정의 재정분석에서 결정적 의미를 갖는다. 사회부조에서 사회보험으로, 비공식부문에서 공식부문으로의 제도 이행 과정에서 발생하는 복잡한 재정 변화를 체계적으로 추적할 수 있다. 이는 제도 성숙 과정에 있는 국가들의 장기재정분석에서 단순한 선형적 증가 가정을 넘어서 제도 전환의 비선형적 효과를 반영할 수 있게 한다.

또한 SSI는 “사회보장 최저선(Social Protection Floors)” 개념과 유엔 지속가능발전목표(SDG) 1.3의 모니터링 도구로 기능한다. 이는 제도 발전 단계에 있는 국가들의 장기재정분석에서 국제적 기준과 목표에 부합하는 정책 시나리오를 설계할 수 있게 한다.

## 2) 3년 주기 데이터 수집의 시의성 한계

SSI의 주요 한계 중 하나는 3년 주기의 데이터 수집으로 인한 시의성 부족이다. ILO World Social Protection Report가 3년마다 발간되고, 국가별 데이터 업데이트도 이와 유사한 주기를 갖는다. 이는 OECD SOCX의 연간 업데이트, EU ESSPROS의 연 2회 업데이트와 비교할 때 상당히 느린 속도이다.

장기재정분석에서 이러한 시의성 제약은 여러 문제를 야기한다. 첫째, 경제위기나 팬데믹과 같은 급격한 변화에 대한 정책 대응을 실시간으로

추적하기 어렵다. COVID-19 팬데믹 기간 동안 각국이 도입한 긴급 사회보장 조치들의 재정적 영향을 분석하고 이를 장기전망에 반영하는 데 상당한 시차가 발생한다. 둘째, 제도 개혁의 효과를 신속하게 평가하기 어렵다. 연금개혁, 건강보험 확대 등 주요 제도 변화의 재정적 영향을 파악하고 장기전망 모형에 반영하는 데 최소 3년의 지연이 발생할 수 있다.

특히 제도 발전 과정에 있는 국가들의 경우 사회보장제도가 빠르게 확대되고 변화하는 단계에 있어, 3년 주기의 데이터로는 제도 발전의 동태적 과정을 정확히 포착하기 어렵다. 예를 들어, 전 국민건강보험 도입, 국민연금 적용범위 확대, 기초연금 신설 등 주요 제도 변화가 단기간에 연속적으로 일어나는 상황에서 3년 주기 데이터로는 이러한 급변하는 제도 환경을 실시간으로 추적하고 분석하는 데 한계가 있다.

이러한 시의성 제약은 SSI 기반 장기재정분석의 정확성과 정책적 유용성을 제한하는 요인으로 작용한다. 따라서 SSI 데이터를 활용할 때는 국가별 행정자료나 다른 국제기구의 최신 통계로 보완하는 종합적 접근이 필요하다.

### 3) 제도 성숙 국가 세부 분석의 정밀도 한계

SSI의 세 번째 한계는 사회보장제도가 고도로 발달한 국가들의 복잡하고 정교한 제도 구조에 대한 세부 분석에서 상대적으로 제약이 있다는 점이다. 190여 개국을 포괄하는 광범위한 접근법은 필연적으로 개별 국가의 제도적 세부 사항을 단순화할 수밖에 없으며, 이는 특히 제도가 고도로 성숙한 국가에서 분석의 정밀도를 제한한다.

제도 성숙 국가의 사회보장제도는 여러 층위(pillar)로 구성된 복합적 구조, 공공-민간 혼합 방식, 세제와의 복잡한 연계 등 매우 정교한 특성을 갖는다. 예를 들어, 네덜란드의 3층 연금제도, 독일의 사회보험-조세 혼

합 방식, 프랑스의 직종별 특수제도 등은 SSI의 표준화된 분류로는 충분히 포착하기 어려운 복잡성을 갖는다. 이러한 제도적 세부 사항은 장기재정분석에서 매우 중요한 의미를 갖는데, 인구 고령화의 재정적 영향이 제도 구조에 따라 크게 달라지기 때문이다.

또한 SSI는 제도 발전 단계가 다양한 국가들을 포괄하기 위한 표준화된 접근법을 채택하고 있어, 제도 성숙 국가의 고도화된 특성을 분석하는 데 상대적으로 부적합한 측면이 있다. 예를 들어, 직업 연금의 복잡한 구조, 세제 혜택의 정교한 설계, 민간 사회보험의 다양한 형태 등은 OECD SOCX나 EU ESSPROS에서 더 정밀하게 다뤄진다.

이러한 한계는 제도 성숙 국가의 장기재정분석에서 SSI를 주 데이터베이스로 활용하기 어려운 이유가 된다. 제도 성숙 국가의 경우 SSI는 글로벌 맥락에서의 위치 파악과 다른 발전 단계 국가들과의 비교 분석에는 유용하지만, 제도별 세부 분석과 정밀한 장기전망에는 OECD SOCX나 EU 데이터베이스를 주로 활용하고 SSI를 보완적으로 사용하는 것이 적절하다.

따라서 SSI의 활용 전략은 분석 대상 국가의 제도 발전 단계와 성숙도에 따라 차별화되어야 한다. 제도 발전 과정에 있는 국가에서는 SSI가 가장 포괄적이고 유용한 데이터베이스로 기능하지만, 제도 성숙 국가에서는 다른 특화된 데이터베이스와의 결합을 통한 보완적 활용이 필요하다.

### 3. ISSA DB

#### 가. ISSA Country Profiles 구성과 범위

##### 1) 190개국 사회보장제도 정성적 정보

ISSA Country Profiles 데이터베이스는 전 세계적 포괄성을 바탕으로

190개국의 사회보장제도에 대한 체계적인 정성적 정보를 제공한다. 이는 장기재정분석 관점에서 특히 중요한 의미를 갖는데, 정량적 재정 데이터 만으로는 파악하기 어려운 제도의 구조적 특성과 운영 메커니즘에 대한 상세한 정보를 제공하기 때문이다.

데이터베이스의 지리적 범위는 <표 3-16>에서 보는 바와 같이 OECD 선진국, EU 회원국뿐만 아니라 아시아, 아프리카, 라틴아메리카의 개발도상국과 최빈개도국까지 포괄한다. [그림 3-11]은 ISSA 웹사이트의 Country Profiles 메인 화면으로, 세계지도를 통해 190개국의 포괄적 범위를 시각적으로 확인할 수 있으며, 국가별, 지역별, 제도 영역별 검색이 가능한 사용자 친화적 인터페이스를 제공한다.

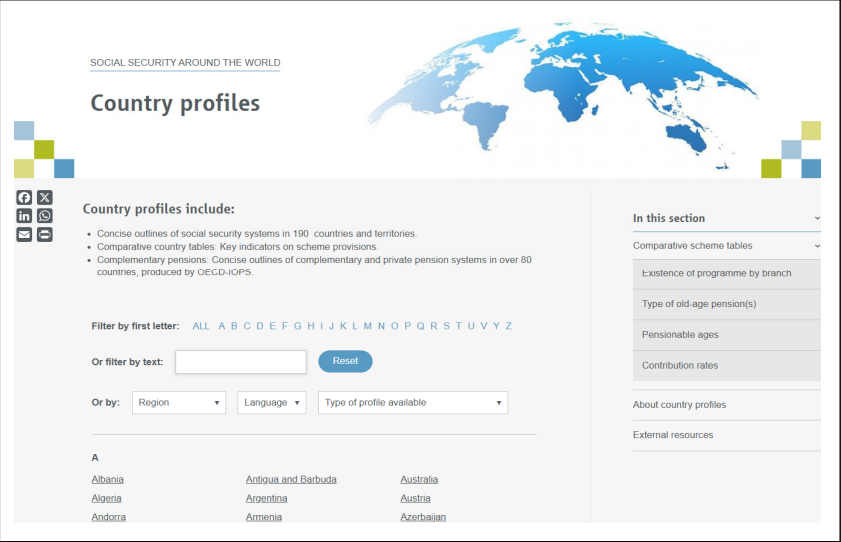
이러한 포괄성은 한국의 사회보장제도를 단순히 선진국과의 비교를 넘어서 다양한 발전 경로와 제도 모형을 참조하여 장기적 제도 발전 방향을 설정하는 데 중요한 기반을 제공한다. 국가별 정보의 깊이와 상세도는 해당 국가의 제도 성숙도와 정보 제공 역량에 따라 차이가 있지만, 모든 국가에 대해 최소한의 표준화된 정보 구조를 유지한다. 이는 장기재정분석에서 국가 간 제도적 차이를 체계적으로 비교하고, 특정 제도 모형의 재정적 함의를 분석하는 데 필수적인 기반을 제공한다.

<표 3-16> ISSA Country Profiles 지역별 포괄 범위

지역	국가 수	주요 특징	활용도
유럽	48개국	EU회원국+비회원국 포괄	제도 성숙도 비교
아시아	52개국	다양한 발전 혼재	아시아 모델 비교
아프리카	54개국	신흥 사회보장제도	제도 도입 사례
아메리카	36개국	남북미 다양한 모델	연금 개혁 사례
전체	190개국	전 발전단계 포괄	글로벌 제도 비교

출처: 저자 작성.

[그림 3-11] ISSA Country Profiles 웹사이트 화면



출처: ISSA. (2025). <https://www.issa.int/databases/country-profiles>. 2025.6.18. 접속

2) 5개 주요 영역별 제도 설명 구조

ISSA Country Profiles는 사회보장제도를 5개 주요 영역으로 체계화 하여 정보를 제공한다. 이러한 영역별 구분은 장기재정분석에서 인구구조 변화와 사회경제적 변화가 각 영역에 미치는 차별적 영향을 분석하는데 핵심적 틀을 제공한다.

〈표 3-17〉 ISSA 5개 주요 영역별 데이터 구성


지역	세부 내용	재정분석 관련성	주요 데이터
노령·장애·유족	연금제도 전반	인구고령화	수급요건, 급여산정, 재정방식
상병·모성급여	의료보장/소득보전	의료비	급여기간, 급여수준, 적용범위
고용상해	산재보험	산업구조 변화	보상범위, 재할급여, 예방조치
실업급여	실업보험 및 지원	노동시장 변화	수급조건, 급여기간, 재취업자
가족수당	아동수당/가족지원	저출산 대응	급여대상, 급여수준, 지급기간

출처: 저자 작성.


영역별로 제도의 기본 구조, 적용 범위, 급여 내용, 재원 조달 방식, 관리 운영 체계에 대한 상세한 정보가 표준화된 형식으로 제공된다. 예를 들어, 노령연금 영역의 경우 [그림 3-12]에서 보는 바와 같이 제도 유형(사회보험/사회수당), 가입 대상, 수급 요건(연령, 가입 기간), 급여 산정 방식, 연동 메커니즘 등이 체계적으로 기술된다.

이러한 영역별 구조화는 장기재정분석에서 특히 중요한 의미를 갖는다. 인구 고령화는 주로 노령연금과 의료보장에 집중적인 재정 압박을 가하는 반면, 저출산은 가족수당 정책의 확대와 노동력 감소에 따른 실업급여 지출 변화를 초래한다. ISSA의 영역별 구분은 이러한 차별적 영향을 체계적으로 분석할 수 있는 정보 체계를 제공한다.

[그림 3-12] ISSA Country Profiles 데이터 구조 예시


ISSA COUNTRY PROFILE: REPUBLIC OF KOREA

ISSA COUNTRY PROFILE: REPUBLIC OF KOREA  
[www.issa.int/country-profiles](https://www.issa.int/country-profiles)


ISSA COUNTRY PROFILE: REPUBLIC OF KOREA

ISSA COUNTRY PROFILE: REPUBLIC OF KOREA  
[www.issa.int/country-profiles](https://www.issa.int/country-profiles)

## Republic of Korea

Policies as of 1 January 2022. Updated via a country survey and desk review.  
 Currency: Won (KRW)

### Profile structure

The profiles are organized in the branches:

- Old age, invalidity and survivors
- Health and long-term care benefits
- Sickness and maternity
- Accidents at work and occupational diseases
- Unemployment
- Family and household benefits

Each branch includes the following categories:

- Branch overview
- Regulatory framework
- Coverage
- Source of funds
- Benefit details
- Administrative organization

## Old age, invalidity and survivors

### Branch overview

Type of programme	Programme name
Social insurance	National Pension
Social assistance	Basic Pension

**Note:** A special system exists for civil servants, private-school employees, military personnel, and employees of the special post office.

### Regulatory framework

Year	Law name and number	Amendments
1988	National Pension Act No. 2392	1999 and 2007
1994	Basic Pension Act No. 4242	

### Coverage

Programme	Mandatory coverage	Voluntary coverage	Exclusions	Citizenship conditions
National Pension (social insurance)	Employed and self-employed persons, including farmers and fishermen	Persons without any income, including persons younger than age 22 and spouses of insured persons with mandatory coverage, and certain needy persons.	Residenceless who are nationals of countries that do not cover Korean nationals under their public pension schemes are excluded from coverage.	Other: Yes, citizens only.
Basic Pension (social assistance)	Low-income citizens of South Korea.			

### Source of funds

Programme name and persons covered (1)	Contributions	Income base	Government contributions
<b>National Pension (social insurance)</b>			
Employees (1)	Covered persons: 6.5% (mandatory insured), 9% of the previous year's medical monthly income of all voluntarily insured persons or more (voluntarily insured) Employer: 4.5% (mandatory insured), none (voluntarily insured)	Gross monthly covered earnings (mandatory insured), via flat-rate amount, voluntarily insured persons: Minimum: 330,000 won a month, Maximum: 5,240,000 won a month	Part of the cost of administration; contributions for some farmers and fishermen, some low-income employees, parents with more than two children, and insured persons with military service.
Self-employed	9%	Gross monthly covered earnings: Minimum: 330,000 won a month, Maximum: 5,240,000 won a month	Part of the cost of administration; contributions for some farmers and fishermen, some low-income employees, parents with more than two children, and insured persons with military service.
<b>Basic Pension (social assistance)</b>			
All covered persons	None	None	The government pays the total cost.

(1) Employees (1) refers to all persons who are covered who are not self-employed employees, pensioners, and other protected groups where applicable, such as students, unemployed persons, etc.

### Old-age benefits

#### Old-age pension – National Pension (social insurance)

Qualifying conditions	Age requirements	Minimum qualifying period	Means test	Employment conditions	Benefit details
	62	With at least 20 years of coverage (flat pension). Reduction based on income or assets	See amount. Reduction based on income (see amount)		The pension is the insured's monthly basic pension amount (BPA). An increment is paid for years of coverage exceeding 20 years. The pension is reduced for any monthly income exceeding 2,682,234 won, up to a monthly reduction of 500,000 won if the pensioner's monthly income exceeds that amount by at least 4,000,000
					Legislated increases: Gradually rising to age 62 and over by 1949. Legislated increases: None foreseen.

출처: ISSA. (2025). <https://www.issa.int/databases/country-profiles>. 2025.6.18. 접속

### 3) 6개 표준화 항목

ISSA Country Profiles의 각 영역은 6개 표준화 항목을 통해 일관성 있는 정보 구조를 유지한다. 이러한 표준화는 국가 간 비교 분석과 장기 재정분석 모형의 입력 자료로 활용하는 데 필수적 기반을 제공한다.

각 표준화 항목은 장기재정분석의 서로 다른 측면에 필수적 정보를 제공한다. 제도 유형 정보는 사회보험 방식의 기여-급여 연계 구조와 조세 방식의 일반재정 의존 구조를 구분하여 인구구조 변화에 따른 재정적 지속가능성을 차별적으로 분석할 수 있게 한다. 적용 대상 정보는 고용 형태 변화, 경제활동 참가율 변화에 따른 피보험자 기반의 변화를 전망하는 데 핵심적 기초를 제공한다.

재원 조달 항목은 보험료율, 정부 지원 비율, 기여 상한선 등의 정보를 통해 재정수입 구조의 특성과 한계를 파악할 수 있게 한다. 특히 기여 상한선이 있는 제도의 경우 소득 불평등 확대가 재정수입에 미치는 영향을 분석하는 데 중요한 정보를 제공한다. 수급 요건과 급여 내용 정보는 수급자 수와 개별 급여 수준을 전망하는 데 직접적으로 활용되며, 관리 기관 정보는 제도 운영의 효율성과 개혁 가능성을 평가하는 데 참고 자료를 제공한다.

〈표 3-18〉 표준화 항목과 장기재정분석 활용도

항목	주요 정보	재정분석 활용도	데이터 특성
제도 유형	사회보험, 사회수당, 공공부조 등	제도 유형화	정성적, 범주형
적용 대상	당연(임의)가입, 적용제외 등	대상자 구조	정성적, 범주형
재원 조달	보험료율, 정부부담, 기여상한 등	재원조달 구조	정량적+정성적
수급 요건	가입기간, 연령, 소득기준	수급자 구조	정량적+정성적
급여 내용	급여율, 급여기간, 보장한도 등	급여 구조	정량적+정성적
관리 기관	정부부처, 공단, 민간 등	운영 구조	정량적+정성적

출처: 저자 작성.

## 나. ISSA의 특징 및 한계

### 1) 제도 구조적 특성의 체계적 정보 제공

ISSA Country Profiles의 가장 중요한 강점은 사회보장제도의 구조적 특성에 대한 체계적이고 상세한 정보를 제공한다는 점이다. 이는 단순한 지출 규모나 수급자 수를 넘어서 제도가 어떻게 설계되고 운영되는지에 대한 깊이 있는 이해를 가능하게 한다.

이러한 구조적 정보는 장기재정분석에서 단순한 추세 연장을 넘어서 제도 내재적 메커니즘을 반영한 정밀한 전망을 가능하게 한다. 예를 들어, 급여 산정 공식과 연동 메커니즘에 대한 정보는 임금 상승률, 물가 상승률 변화가 개별 급여 수준에 미치는 영향을 정확히 산정할 수 있게 한다. 조기퇴직 조건과 크레딧 인정 범위에 대한 정보는 노동시장 변화와 가족 구조 변화가 연금 수급 행태에 미치는 영향을 분석하는 데 필수적이다.

### 2) 전 세계적 포괄성과 높은 최신성

ISSA Country Profiles의 두 번째 핵심 강점은 전 세계 190개국을 포괄하는 광범위성과 지속적인 업데이트를 통한 높은 최신성이다. 이는 다른 국제 데이터베이스와 비교할 때 매우 독특한 특징으로, 장기재정분석의 국제 비교 기반을 크게 확장한다.

전 세계적 포괄성은 한국의 사회보장제도를 다양한 관점에서 비교 분석할 수 있게 한다. OECD 선진국과의 비교를 통해서도 제도 성숙화 과정에서 나타나는 재정적 도전과 개혁 방향을 파악할 수 있고, 아시아 신흥국과의 비교를 통해서도 유사한 사회경제적 배경에서의 제도 설계 대안을 검토할 수 있다. 특히 인구구조 전환이 한국보다 늦게 시작되는 국가들의 사례는 장기재정분석의 시나리오 구성에 중요한 참조점을 제공한다.

높은 최신성은 급변하는 사회보장 환경에서 특히 중요하다. COVID-19 팬데믹과 같은 외부 충격이나 주요 제도 개혁이 있을 때 신속한 정보 업데이트를 통해 최신 동향을 반영할 수 있다.

### 3) 정량적 재정 데이터 부족의 한계

ISSA Country Profiles의 가장 중요한 한계는 정량적 재정 데이터의 부족이다. 제도의 구조적 특성에 대한 상세한 정보는 제공하지만, 장기재정분석에 직접 필요한 지출 규모, 재정 수지, 적립기금 현황 등의 계량적 데이터는 제한적으로만 제공된다.

이러한 한계는 ISSA를 장기재정분석에 활용할 때 반드시 다른 정량적 데이터베이스와의 결합이 필요함을 의미한다. ISSA의 정성적 정보는 제도의 작동 메커니즘을 이해하고 분석 모형의 가정을 설정하는 데 핵심적 역할을 하지만, 실제 재정 전망을 위해서는 SOCX, ESSPROS, SSI 등의 정량적 데이터가 필수적으로 결합되어야 한다.

또한 정보의 표준화 한계도 중요한 제약요인이다. 국가별로 상이한 제도 구조와 용어 사용으로 인해 완전한 표준화가 어려우며, 특히 복합적 제도나 과도기적 제도의 경우 정확한 분류와 비교에 어려움이 있다. 이는 장기재정분석에서 국가 간 비교 결과를 해석할 때 각별한 주의가 필요함을 의미한다.

그러나 이러한 한계에도 불구하고 ISSA는 사회보장 장기재정분석을 위한 정성적 기초자료로서 대체 불가능한 가치를 지닌다. 제도의 구조적 특성과 운영 방식에 대한 상세한 정보는 재정전망 모형의 가정 설정과 시나리오 구성에 필수적이며, 특히 제도 간 상호작용 효과나 정책 변화의 영향을 분석하는 데 중요한 기초를 제공한다.

## 제3절 사회보장 재정분석을 위한 DB 구축 방향

### 1. 주요 국제 DB 비교 분석 결과

본 절에서는 1절과 2절에서 개별적으로 분석한 주요 국제 사회보장 데이터베이스들의 특성을 체계적으로 비교 분석한다. 이러한 비교 분석은 본 연구에서 구축하고자 하는 사회보장 장기재정분석 DB의 설계 방향을 도출하고, 핵심 기반 데이터베이스를 선정하는 데 중요한 근거를 제공한다.

#### 가. 정량적 DB의 특성 비교

##### 1) SOCX: 순지출 중심, 9개 정책영역, 38개국

OECD SOCX는 1절에서 분석한 바와 같이 사회보장 지출 비교 분석의 국제 표준으로 자리 잡은 대표적인 정량적 데이터베이스이다. SOCX의 핵심 특성은 순지출 중심의 접근법에 있다. 총지출에서 조세 환수분을 차감한 순지출(net expenditure) 개념을 통해 각국 사회보장제도의 실질적 보호 수준을 비교할 수 있게 한다. 이는 조세제도와 사회보장제도가 긴밀하게 연계된 현대 복지국가의 특성을 정확히 반영하는 중요한 접근법이다.

9개 정책영역 분류체계는 노령(Old-age), 유족(Survivors), 근로무능력 관련(Incapacity-related), 보건(Health), 가족(Family), 적극적 노동시장정책(ALMP), 실업(Unemployment), 주거(Housing), 기타 사회정책 영역(Other social policy areas)으로 구성된다. 이러한 분류는 사회정책의 기능적 접근과 제도적 접근을 균형있게 반영하여, 장기재정분석에서 인구구조 변화와 사회경제적 변화의 차별적 영향을 체계적으로

분석할 수 있는 틀을 제공한다.

38개 OECD 회원국과 주요 비회원국을 대상으로 하는 지리적 범위는 제도 성숙 국가들 간의 정밀한 비교 분석을 가능하게 한다. 특히 한국을 포함한 OECD 국가들의 1980년부터의 장기 시계열 데이터는 제도 발전 과정과 재정 변화 패턴을 분석하는 데 귀중한 자료를 제공한다. 1절에서 확인한 바와 같이 한국의 경우 1990년부터 데이터가 제공되며, 2018년 이후 결산 기준으로 전환되어 데이터 품질이 크게 향상되었다.

SOCX의 한계는 지출 중심 접근으로 인한 제도 특성 파악의 제약과 재원조달 정보의 부재이다. 동일한 지출 수준이라도 제도 설계와 재정구조에 따라 장기적 지속가능성이 달라질 수 있으나, SOCX만으로는 이러한 구조적 차이를 파악하기 어렵다.

## 2) ESSPROS: 지출+재원조달, 8개 기능, EU 회원국

EU ESSPROS는 2절에서 분석한 바와 같이 지출과 재원조달 정보를 통합적으로 제공하는 독특한 특성을 갖는다. 이는 SOCX의 가장 중요한 한계인 재원조달 정보 부재를 보완하는 핵심적 강점이다. 사회기여금, 일반정부 기여금, 기타 수입으로 구분된 수입 분류와 고용주-피보험자-자영업자 등 기여자 유형별 세분화를 통해 사회보장제도의 재정구조를 정밀하게 분석할 수 있다.

8개 기능별 분류체계는 질병/의료(Sickness/health care), 장애(Disability), 노령(Old age), 유족(Survivors), 가족/아동(Family/children), 실업(Unemployment), 주거(Housing), 사회적 배제(Social exclusion)로 구성된다. SOCX의 9개 영역과 상당한 유사성을 보이면서도, 사회적 배제를 독립 영역으로 설정한 점이 특징적이다. 이는 EU의 사회통합 정책 철학을 반영한 것으로, 장기재정분석에서 사회적 배제 대응

비용의 증가 추이를 체계적으로 추적할 수 있게 한다.

EU 27개 회원국과 EEA 3개국, 스위스를 대상으로 하는 지리적 범위는 유럽 지역의 심층적 분석에는 매우 유용하지만, 글로벌 비교 분석에는 제약이 있다. 특히 아시아-태평양 지역이나 북미 지역과의 비교가 불가능하여, 다양한 사회보장 모델 간의 비교 분석에 한계가 있다.

ESSPROS의 강점은 Core System을 중심으로 한 모듈별 세분화와 EU 거버넌스와의 긴밀한 연계이다. 2절에서 확인한 바와 같이 연금 수급자 모듈(PENS), 순 사회보장급여 모듈(NET) 등을 통해 장기재정분석에 필요한 상세 정보를 제공하며, European Semester, Ageing Report 등 EU 정책 조정 과정에서 핵심 근거 자료로 활용된다.

### 3) SSI: 3차원 통합데이터, 11개 기능, 190개국

ILO SSI는 2절에서 분석한 바와 같이 지출-재원조달-보장성의 3차원 통합데이터를 제공하는 독특한 구조를 갖는다. 이는 SOCX의 지출 중심 접근과 ESSPROS의 지출-재원조달 2차원 접근을 넘어서, 사회보장제도의 성과와 지속가능성을 종합적으로 평가할 수 있는 가장 포괄적인 정량 데이터 구조이다. 지출 차원에서는 총지출, 급여지출, 관리비용을 구분하고, 재원조달 차원에서는 기여금, 정부지원, 투자수익을 분류하며, 보장성 차원에서는 보호인구, 수급자 수, 보장률 등을 포괄한다.

11개 기능별 분류체계는 전통적인 사회보험 8개 영역에 주거, 기초교육, 기타 소득보장을 추가하여 사회보장의 개념을 가장 포괄적으로 확장했다. 노령, 장애, 유족, 질병 및 의료, 실업, 고용상해, 가족/아동, 모성의 전통적 영역과 함께 주거, 기초교육, 기타 소득보장 및 공공부조/사회적 배제를 포함한다. 이는 장기재정분석에서 인적자본 투자(교육)와 사회보장의 연계, 도시화에 따른 주거비 지원 확대 등 새로운 사회적 위험에 대

한 대응을 체계적으로 분석할 수 있게 한다.

전 세계 190여 개국을 포괄하는 지리적 범위는 다른 어떤 데이터베이스도 달성하지 못한 광범위성이다. 이는 사회보장제도의 다양한 발전 모델과 경로를 종합적으로 비교 분석할 수 있는 유일한 데이터베이스로서의 가치를 제공한다. 특히 제도 도입 초기 단계부터 성숙 단계까지의 발전 과정을 분석하고, 경제발전 수준에 따른 사회보장 지출 패턴을 파악하는 데 중요한 기여를 한다.

SSI의 한계는 3년 주기의 데이터 수집으로 인한 시의성 부족과 제도 성숙 국가에 대한 분석의 정밀도 한계이다. 190개국을 포괄하는 광범위한 접근법은 필연적으로 개별 국가의 복잡한 제도 구조를 단순화할 수밖에 없으며, 이는 특히 제도가 고도로 발달한 국가들의 세부 분석에서 제약으로 작용한다.

〈표 3-19〉 정량적 DB 특성 종합 비교

구분	SOCX	ESSPROS	SSI
주요 특성	순지출	지출+재원조달	3차원 통합
분류 체계	9개 정책영역	8개 기능	11개 기능
국가 범위	OECD 38개국	EU 31개국	전세계 190개국
시계열 범위	1980년~	1990년~	3년 주기
업데이트 주기	연간	연간	3년
주요 강점	국제비교 표준	재원조달 정보	포괄적 정보
주요 한계	재원 조달 부재	지역적 제약	시의성 부족

출처: 저자 작성.

## 나. 정성적 데이터베이스의 특성 비교

### 1) SOCX Country Notes: 제도 특성 및 분류 기준 설명

SOCX Country Notes는 1절에서 분석한 바와 같이 제도 특성과 분류 기준에 대한 상세한 설명을 제공하는 정성적 정보의 핵심이다. 각국의 사회보장제도를 SOCX의 9개 정책영역으로 분류할 때 적용되는 구체적인 기준과 방법, 국가 특수적 제도의 분류 방식, 시계열 단절(Break in Series)의 원인과 영향 등을 체계적으로 기술한다.

Country Notes의 가장 중요한 기능은 정량 데이터 해석을 위한 맥락 정보 제공이다. 동일한 정책영역이라도 국가별로 상이한 제도 설계와 운영 방식을 이해할 수 있어, 단순한 지출 비교를 넘어선 정책 분석이 가능하다. 예를 들어, 가족정책 지출이 유사한 독일과 프랑스의 경우에도 Country Notes를 통해 독일은 조세감면 중심, 프랑스는 직접 현금급여 중심이라는 구조적 차이를 파악할 수 있다.

또한 정책 개혁의 배경과 영향 분석에 필수적인 정보를 제공한다. 스웨덴의 1998년 연금개혁, 독일의 2001년 리스터 연금 도입 등 주요 제도 개혁의 내용과 장기적 파급효과를 Country Notes에서 상세히 다루고 있어, 개혁 효과 분석과 미래 개혁 방향 설정에 중요한 참고자료가 된다.

Country Notes의 한계는 비구조화된 서술형 정보로 구성되어 체계적 활용에 제약이 있다는 점이다. 국가별로 기술 방식과 상세 수준이 다르며, 업데이트 주기도 일정하지 않아 정보의 최신성과 완전성에 한계가 있다. 또한 약 2~3년의 시차를 갖고 제공되어 최신 제도 변화를 실시간으로 반영하기 어렵다.

## 2) MISSOC: 표준화된 제도 비교표, EU 31개국

EU MISSOC는 2절에서 분석한 바와 같이 표준화된 제도 비교표를 통해 31개 유럽 국가의 사회보장제도를 체계적으로 비교할 수 있는 독특한 정성적 데이터베이스이다. 12개 주요 카테고리과 300개 이상의 세부 항목을 통해 법적 근거, 적용 범위, 조건, 급여 수준 등 제도의 핵심 요소들을 표준화된 형식으로 제공한다.

MISSOC의 핵심 강점은 높은 표준화 수준과 비교 가능성이다. 동일한 항목에 대해 모든 국가가 일관된 형식으로 정보를 제공하여, 국가 간 제도 차이를 체계적으로 비교 분석할 수 있다. 특히 [그림 3-8]에서 확인한 Comparative Table 기능을 통해 특정 제도 영역이나 국가군에 대한 맞춤형 비교 분석이 가능하다.

또한 연 2회(1월 1일, 7월 1일) 정기 업데이트를 통해 제도 변화를 실시간으로 추적할 수 있어, 다른 정성적 데이터베이스에 비해 상대적으로 높은 최신성을 확보하고 있다. 자영업자 전용 섹션을 별도 운영하여 전통적인 임금근로자 중심 제도에서 포착되지 않는 고용 형태 변화를 체계적으로 반영한다.

MISSOC와 SOCX 간의 카테고리 매핑 가능성도 중요한 특징이다. 2절에서 확인한 바와 같이 MISSOC의 12개 카테고리는 SOCX의 9개 정책영역과 상당한 유사성을 보여, 정량-정성 정보의 결합 분석이 가능하다. 특히 MISSOC에서만 제공되는 Financing 정보는 SOCX의 지출 중심 분석을 재원조달 측면에서 보완하는 중요한 역할을 한다.

MISSOC의 한계는 EU 지역으로 제한된 지리적 범위이다. 31개 유럽 국가만을 대상으로 하여 글로벌 차원의 비교 분석이 불가능하며, 아시아-태평양이나 북미 지역의 제도 모델과의 비교 연구에는 활용할 수 없다. 또한 EU 차원의 정책 조정 압력으로 인해 회원국들의 제도가 유사한 방

향으로 수렴하는 경향이 있어, 제도 다양성 측면에서는 상대적으로 제약이 있다.

### 3) ISSA: 체계적 제도 정보, 190개국

ISSA Country Profiles는 2절에서 분석한 바와 같이 전 세계 190개국의 체계적 제도 정보를 제공하는 가장 포괄적인 정성적 데이터베이스이다. 5개 주요 영역(노령·장애·유족연금, 상병·모성급여, 고용상해, 실업급여, 가족수당)과 6개 표준화 항목(제도 유형, 적용 대상, 재원 조달, 수급요건, 급여 내용, 관리 기관)을 통해 일관성 있는 정보 구조를 유지한다.

ISSA의 가장 중요한 특징은 제도 구조적 특성에 대한 심층적 정보 제공이다. 급여 산정 공식, 연동 메커니즘, 조기퇴직 조건, 크레딧 인정 범위 등 장기재정분석에 필수적인 제도 설계 세부 사항을 체계적으로 기술한다. 이는 단순한 지출 규모나 수급자 수를 넘어서 제도가 어떻게 설계되고 운영되는지에 대한 깊이 있는 이해를 가능하게 한다.

전 세계적 포괄성은 ISSA의 독보적 강점이다. OECD 선진국부터 최빈개도국까지 모든 발전 단계의 국가를 포괄하여, 사회보장제도의 다양한 발전 모델과 경로를 비교 분석할 수 있다. 이는 한국의 사회보장제도를 글로벌 맥락에서 위치 지우고, 다양한 정책 대안을 모색하는 데 중요한 기반을 제공한다.

높은 최신성과 지속적 업데이트도 ISSA의 중요한 장점이다. COVID-19 팬데믹과 같은 급격한 변화나 주요 제도 개혁에 대한 신속한 정보 반영을 통해 변화하는 사회보장 환경을 실시간으로 추적할 수 있다. 또한 미국 사회보장청(SSA)과의 협력을 통한 권위성과 신뢰성을 확보하고 있다.

ISSA의 한계는 정량적 재정 데이터의 부족이다. 제도의 구조적 특성에

대한 상세한 정보는 제공하지만, 지출 규모, 재정 수지, 적립 기금 현황 등 장기재정분석에 직접 필요한 계량적 데이터는 제한적으로만 제공된다. 또한 190개국을 포괄하는 광범위한 접근으로 인해 정보의 표준화 한계가 존재하며, 국가별로 상이한 제도 구조와 용어 사용으로 완전한 비교가 어려운 경우가 있다.

〈표 3-20〉 정성적 DB 특성 종합 비교

구분	SOCX Country Notes	MISSOC	ISSA
주요 특성	분류 기준 설명	표준화된 비교	체계적 제도 정보
정보 구조	서술형 설명	12개 카테고리	5개 영역, 6개 항목
국가 범위	OECD 38개국	EU 31개국	전세계 190개국
업데이트 주기	연간	연 2회	연중 수시
표준화 수준	낮음	높음	중간
주요 강점	SOCX 데이터 해석	유럽 제도 비교	글로벌 포괄성
주요 한계	비구조화 정보	지역적 제약	정량 데이터 부족

출처: 저자 작성.

## 2. DB 구축을 위한 데이터 결합 방안

### 가. 결합 데이터베이스 선정: SOCX + ISSA

앞서 분석한 주요 국제 데이터베이스들의 특성과 한계를 종합적으로 검토한 결과, 본 연구에서 구축하고자 하는 사회보장 장기재정분석 DB의 핵심 기반으로 OECD SOCX와 ISSA Country Profiles의 결합을 선정하였다. 이러한 선정은 정량적 분석의 국제 표준성과 정성적 분석의 포괄성을 동시에 확보하려는 전략적 판단에 기반한다.

### 1) SOCX: 정량적 국제 비교 표준, 한국 포함 OECD 국가 중심

OECD SOCX를 핵심 정량 데이터베이스로 선정한 이유는 다음과 같다. 첫째, 1항에서 확인한 바와 같이 국제 비교 분석의 표준적 지위를 확보하고 있다는 점이다. <표 3-21>에서 보듯이 SOCX는 사회보장 지출 비교 분석에서 가장 널리 활용되는 국제 표준 데이터베이스로, 학술 연구와 정책 분석에서 준거점 역할을 하고 있다. 이는 본 연구의 분석 결과가 국제적 맥락에서 해석되고 비교될 수 있는 기반을 제공한다.

둘째, 한국을 포함한 OECD 국가 중심의 구성으로 한국의 장기재정분석에 직접적 활용도가 높다. 1절에서 살펴본 바와 같이 SOCX는 38개 OECD 회원국과 주요 비회원국을 포괄하여, 한국과 유사한 제도 발전 단계에 있는 국가들과의 비교 분석이 가능하다. 특히 한국의 SOCX 데이터는 1990년부터 제공되어 충분한 시계열을 확보하고 있으며, 2018년 이후 결산 기준으로 전환되어 데이터 정확성이 크게 향상되었다.

셋째, 체계적인 분류 체계와 품질 관리 시스템을 갖추고 있다. 1절에서 분석한 9개 정책영역, 3개 제공 주체, 2개 급여 형태의 3차원 분류체계는 사회보장제도의 복잡성을 체계적으로 정리하여 분석할 수 있는 틀을 제공한다. 또한 프로그램 단위의 상세 기록과 Country Notes를 통한 메타데이터 관리는 데이터의 신뢰성과 해석 가능성을 보장한다.

넷째, 다른 OECD 데이터베이스와의 연계 구조가 확립되어 있다. 1절 2항에서 확인한 바와 같이 Health Statistics, Employment Database, Education Database와의 체계적 연계를 통해 사회보장 지출의 포괄적 분석이 가능하다. 이는 장기재정분석에서 요구되는 다차원적 접근을 지원하는 중요한 기반이다.

## 2) ISSA: 체계적 정성 정보, 전 세계 포괄성, 높은 최신성

ISSA Country Profiles를 핵심 정성 데이터베이스로 선정한 이유는 다음과 같다. 첫째, 사회보장제도의 구조적 특성에 대한 체계적이고 상세한 정보를 제공한다는 점이다. 2절에서 분석한 바와 같이 ISSA는 5개 주요 영역, 6개 표준화 항목을 통해 제도의 설계 원리, 급여 산정 방식, 수급 요건, 재원조달 구조 등을 일관성 있게 기술한다. 이는 SOCX의 지출 데이터를 해석하고 장기전망 모형의 가정을 설정하는 데 필수적인 맥락 정보를 제공한다.

둘째, 전 세계 190개국을 포괄하는 광범위한 지리적 범위를 갖고 있다. 이는 한국의 사회보장제도를 OECD 선진국뿐만 아니라 아시아 신흥국, 개발도상국 등 다양한 발전 단계의 국가들과 비교 분석할 수 있게 한다. 특히 인구구조 전환과 경제발전 과정에서 나타나는 사회보장제도의 변화 패턴을 다각도로 분석하는 데 중요한 기반을 제공한다.

셋째, 높은 최신성과 지속적인 업데이트 체계를 갖추고 있다. COVID-19 팬데믹과 같은 급격한 변화나 주요 제도 개혁에 대한 신속한 정보 업데이트를 통해 장기재정분석의 현실 적합성을 높인다. 이는 빠르게 변화하는 사회보장 환경에서 분석의 시의성을 확보하는 중요한 요소이다.

넷째, 미국 사회보장청(SSA)과의 협력을 통한 신뢰성과 권위를 확보하고 있다. “Social Security Programs Throughout the World”의 핵심 기반 자료로 활용되는 ISSA의 정보는 국제적으로 인정받는 공신력을 갖고 있으며, 이는 본 연구의 분석 결과에 대한 신뢰성을 높이는 요소이다.

3) 두 DB의 상호 보완적 관계

SOCX와 ISSA의 결합은 단순한 데이터 병합을 넘어서 <표 3-21>에서 보는 바와 같이, 두 데이터베이스는 각각의 강점으로 상대방의 한계를 보완하는 구조를 형성한다.

<표 3-21> SOCX와 ISSA의 상호 보완적 특성

항목	SOCX 강점	ISSA 보완	ISSA 강점	SOCX 보완
데이터 성격	정량적 지출	제도 구조 정보	정성적 제도정보	재정 규모
국가 범위	OECD 38개국	190개국 정보	190개국 정보	OECD 표준
분석 깊이	표준화된 비교	제도 세부 특성	구조적 특성	계량적 근거
시계열 범위	1980년 이후	제도 변화 맥락	최신 제도 현황	장기 추이 분석
정책 활용	국제 비교 기준	정책 방향 탐색	제도 설계 참고	재정 영향 검토

출처: 저자 작성.

정량적 측면에서 SOCX의 지출 데이터는 장기재정분석의 기초자료를 제공하지만, 지출 변화의 원인과 메커니즘을 이해하기 위해서는 ISSA의 제도 구조 정보가 필수적이다. 예를 들어, 연금 지출의 증가 추이를 분석할 때 SOCX는 지출 규모와 변화율을 제공하지만, 증가의 원인이 수급자 증가인지, 급여 수준 상승인지, 제도 확대인지는 ISSA의 제도 설명을 통해서만 파악할 수 있다.

정성적 측면에서 ISSA의 제도 특성 정보는 각국의 사회보장 철학과 설계 원리를 이해할 수 있게 하지만, 제도의 실질적 규모와 영향을 평가하기 위해서는 SOCX의 정량 데이터가 필요하다. 제도 설계가 우수하다고 평가되는 국가라도 실제 지출 규모나 보장 수준이 낮을 수 있으며, 이는 SOCX 데이터를 통해서만 확인할 수 있다.

나. SOCX와 ISSA 정성적 결합 체계

1) SOCX, ISSA DB 결합 방안

SOCX와 ISSA의 결합은 4장에서 구체적으로 설계될 예정이지만, 3장의 분석 결과를 바탕으로 기본적인 결합 방향을 제시하면 다음과 같다.

첫째, 분류체계 매핑을 통한 구조적 결합이다. 앞서 분석한 SOCX의 9개 정책영역과 ISSA의 5개 주요 영역 간의 체계적 매핑을 통해 일관성 있는 분석틀을 구축한다. <표 3-22>과 같이 영역별 대응 관계를 설정하여 정량-정성 정보의 연계 분석이 가능하도록 한다.

<표 3-22> SOCX-ISSA 분류체계 매핑 방안

SOCX 정책영역	ISSA 주요영역	결합 분석 요소	활용 방안
노령	노령·장애·유족	연금제도 구조와 지출	인구고령화 영향 분석
근로무능력	노령·장애·유족	장애급여 체계와 지출	장애인구 변화 대응
보건	상병·모성	의료보장 구조와 지출	의료비 영향 분석
가족	가족수당	가족정책 체계와 지출	저출산 대응 효과
실업	실업급여	고용보장 구조와 지출	노동시장 변화 대응

출처: 저자 작성.

둘째, 메타데이터 통합을 통한 해석 기반 구축이다. SOCX의 Country Notes와 ISSA의 제도 설명을 결합하여 각국의 지출 데이터에 대한 종합적 해석 기반을 마련한다. 이는 단순한 수치 비교를 넘어서 제도적 맥락을 고려한 심층 분석을 가능하게 한다.

셋째, 시계열 분석을 위한 제도 변화 추적 체계를 구축한다. SOCX의 장기 시계열 데이터와 ISSA의 제도 개혁 정보를 연계하여, 제도 변화가 지출에 미치는 영향을 체계적으로 분석할 수 있는 체계를 마련한다. 이는

장기재정분석에서 제도 개혁 시나리오의 효과를 예측하는 데 핵심적 도구로 활용될 수 있다.

## 2) 결합 체계의 기대 효과

SOCX와 ISSA의 정성적 결합을 통해 기대되는 효과는 다음과 같다.

첫째, 분석의 정확성 향상이다. 정량 데이터만으로는 파악하기 어려운 지출 변화의 구조적 원인을 제도 정보를 통해 정확히 진단할 수 있다. 이는 장기전망의 신뢰성을 크게 높이는 요소이다.

둘째, 정책 시나리오의 현실성 제고이다. 제도의 실제 작동 메커니즘에 대한 이해를 바탕으로 보다 현실적이고 실현 가능한 정책 대안을 설계할 수 있다. 이는 장기재정분석의 정책적 유용성을 높이는 핵심 요소이다.

셋째, 국제 비교의 타당성 강화이다. 각국의 제도적 맥락을 고려한 비교 분석을 통해 단순한 수치 비교의 한계를 극복하고, 정책적 시사점을 도출할 수 있다.

넷째, 통합적 분석 체계 구축이다. 제2장에서 제안한 9×3 매트릭스 분석체계를 정량-정성 통합 정보로 구현하여, 사회보장제도의 복합적 특성을 종합적으로 분석할 수 있는 체계를 구축한다.

## 3) 향후 보조 데이터베이스 활용 방안: EU, ILO

본 연구에서는 SOCX와 ISSA를 핵심으로 하되, 분석의 완성도를 높이기 위해 EU 데이터베이스와 ILO SSI를 보조적으로 활용하는 방안을 마련한다.

EU 데이터베이스 활용 방안은 다음과 같다. ESSPROS의 재원조달 정보를 활용하여 SOCX의 지출 중심 분석을 보완한다. 특히 유럽 국가들의 사회기여금, 정부기여금, 기타 수입의 구성비 정보는 재원조달 구조의 지

속가능성 분석에 중요한 참고 자료를 제공한다. MISSOC의 제도 비교 정보를 활용하여 ISSA의 글로벌 정보를 보완한다. 특히 제도 성숙 국가들의 복잡한 제도 구조에 대한 상세 정보는 한국의 제도 발전 방향 설정에 유용한 벤치마킹 자료가 된다.

ILO SSI 활용 방안은 다음과 같다. SSI의 3차원 데이터(지출-재원조달-보장성)를 활용하여 SOCX-ISSA 결합 체계를 보완한다. 특히 보장성 지표는 지출 효과성 분석에 중요한 정보를 제공한다. SSI의 개발도상국 정보를 활용하여 OECD 중심의 SOCX를 보완한다. 이는 한국과 유사한 발전 경로를 거친 국가들과의 비교 분석을 가능하게 한다.

이러한 다층적 데이터베이스 활용 체계는 4장에서 구체적인 구축 방법론과 함께 상세히 설계될 예정이다. 특히 각 데이터베이스의 업데이트 주기와 품질 수준을 고려한 효율적 통합 방안과, 데이터 일관성 확보를 위한 검증 체계가 핵심적으로 다뤄져야 할 것이다.



## 제4장

### 사회보장 재정분석을 위한 DB 구축

제1절 데이터 수집 및 전처리

제2절 SOCX, ISSA 데이터 연계

제3절 LLM 기반 제도 정보 구조화



## 제4장

# 사회보장 재정분석을 위한 DB 구축

### 제1절 데이터 수집 및 전처리

사회보장 장기재정분석을 위한 데이터베이스 구축은 정량 데이터와 정성 데이터를 체계적으로 통합하는 과정을 필요로 한다. 정량 데이터는 각국의 사회보장 지출 규모와 변화 추세를 측정하는 기초자료이며, 정성 데이터는 이러한 지출이 어떤 제도적 특성에서 비롯되었는지를 설명하는 맥락 정보이다. 본 절에서는 이러한 정량·정성 통합 데이터베이스 구축을 위한 원천 데이터 수집 및 전처리 과정을 기술한다.

전체 구축 프로세스는 [그림 4-1]과 같이 3단계로 구성된다: ① 정량 및 정성 데이터 수집, ② 데이터 연계(Country Notes, ISSA), ③ 데이터 구조화. 특히 데이터 연계(2단계)와 데이터 구조화(3단계) 단계에서 LLM이 활용된다.

본 연구 DB 구축의 주요 특징은 대규모 언어모델(LLM)을 활용한 정성 정보의 구조화이다. Country Notes와 ISSA는 서술형 텍스트로 작성되어 있어 정량 데이터와 직접 결합이 어렵다. 본 연구에서는 LLM 기반 자연어 처리를 통해 OECD 41개국(38개 회원국 + 3개국) × 7개 정책영역(Branch 1,2,3,5,7,8,9: 보건과 적극적노동시장정책 제외)에 대한 정성 정보를 JSON 형식으로 구조화하였다. 이는 OECD에서 제공하는 약 6600개 Code4 프로그램의 제도 특성(scheme\_type, coverage\_type, benefit\_type)을 체계적으로 분류하고 정량 데이터와 연계 가능하게 한 핵심 방법론이다.

[그림 4-1] 데이터 수집 구조 및 통합 흐름



주: SOCX, Country Notes, ISSA 3개 데이터 소스의 수집 및 통합 과정.  
출처: 저자 작성.

본 절은 3개 소절로 구성되며, 각각 [그림 4-1]의 1단계(데이터 수집) 세부 과정을 다룬다. 첫째, OECD SOCX 정량 데이터의 수집 방법과 SOURCE별 집계 수준의 차이를 분석한다. 이는 Code4 프로그램을 식별하고 분류 대상 범위를 확정하는 기초 작업이다. 둘째, Country Notes의 수집 방법과 Code4별 관련 텍스트 추출 결과를 제시한다. 이는 [그림 4-1]의 2단계(데이터 연계)로 이어지는 입력 자료가 된다. 셋째, ISSA 데이터의 웹 크롤링 방법을 기술한다. 이 세 가지 데이터 소스가 통합되어 제2절의 데이터 연계와 제3절의 데이터 구조화 작업으로 이어진다.

## 1. OECD SOCX 정량 데이터 수집 및 구조

### 가. 수집 대상 및 데이터 구조

OECD SOCX는 OECD 회원국의 사회보장 재정을 체계적으로 비교 분석할 수 있도록 설계된 국제 표준 데이터베이스이다. 3장에서 분석한 바와 같이 SOCX는 1980년대 초반부터 구축되기 시작하여 현재 38개 회원국의 1980~2023년 시계열 데이터를 포함하고 있으며, 매년 업데이트되어 최신 재정 동향을 반영한다. 본 연구에서는 2024년 공개된 최신 데이터셋을 활용하였다.

SOCX의 핵심 강점은 프로그램 수준(program-level) 세분화에 있다. 대부분의 국제 재정통계가 정책영역 합계만 제공하는 반면, SOCX는 각국의 개별 사회보장 프로그램(예: 한국의 국민연금, 기초연금, 공무원연금 등)을 식별하고 지출 규모를 분리하여 집계한다. 이는 제도 개혁의 재정적 영향을 정밀하게 추적할 수 있게 하는 핵심 기반이다. 예를 들어, 2014년 한국의 기초연금 도입으로 인한 노령 지출 증가를 국민연금 지출 증가와 분리하여 파악할 수 있으며, 이는 정책 평가와 미래 재정 전망에 필수적이다.

SOCX 데이터는 6단계 계층 구조로 조직되어 있다. 최상위 수준은 국가(Country)이며, 그 하위로 제공 주체(Source), 정책 영역(Branch), 급여 형태(Type), 프로그램 유형(Typrog), 세부 프로그램(Nprog) 순으로 세분화된다. 이러한 계층 구조는 다양한 분석 목적에 따라 유연하게 집계 수준을 조정할 수 있게 한다. 정책영역 수준 분석을 통해 한국의 노령 지출 전체 규모를 파악할 수 있으며, 프로그램 수준 분석을 통해 국민연금과 기초연금의 재정 기여도를 구분할 수 있다.

제공 주체(Source) 분류는 공공-민간 자원 구성을 파악하는 핵심 차원이다. SOURCE 10(공공)은 중앙정부, 지방정부, 사회보험기금 등 공공 부문이 재원을 조달하고 관리하는 프로그램을 포함한다. SOURCE 20(의무적 민간)은 법률로 가입이 강제되는 민간 제공 프로그램으로, 예를 들어 호주의 Superannuation이나 네덜란드의 직역연금이 여기에 속한다. SOURCE 30(임의적 민간)은 자발적 가입에 기초한 민간 연금 및 보험으로, 세제 혜택을 통해 정부가 간접 지원하는 경우가 많다. 이러한 구분은 복지 제공의 공공-민간 역할 배분을 분석하는 기초가 되며, 2장에서 검토한 복지제제론의 실증적 검증에 핵심 지표로 활용된다.

정책 영역(Branch) 분류는 사회적 위험에 대응하는 제도 영역을 구분한다. SOCX는 노령(Old age), 유족(Survivors), 근로무능력(Incapacity-related benefits), 보건(Health), 가족(Family), 적극적노동시장정책(Active labour market programmes), 실업(Unemployment), 주거(Housing), 기타(Other social policy areas)의 9개 영역으로 분류하며, 이는 ILO 및 Eurostat의 분류 기준과 대체로 일치하여 국제 비교 가능성을 확보하고 있다.

급여 형태(Type) 분류는 현금급여(Cash benefits)와 현물급여(In-kind benefits)를 구분한다. 현금급여는 연금, 수당 등 금전으로 지급되는 급여이며, 현물급여는 의료서비스, 돌봄서비스 등 서비스 형태로 제공되는 급여이다. 이러한 구분은 사회보장의 재분배 방식과 경제적 효과를 분석하는 데 중요하다. 현금급여는 수급자의 선택권을 존중하나 서비스 시장이 미발달한 경우 실효성이 제한될 수 있으며, 현물급여는 서비스 접근성을 직접 보장하나 공급 체계 구축에 높은 행정비용이 소요된다.

프로그램 유형(Typrog)과 세부 프로그램(Nprog)은 각 정책 영역 내에서 유사한 기능을 하는 프로그램을 그룹화하고 개별 프로그램을 식별하

는 수준이다. 예를 들어, 한국 노령 영역에서 Typrog=“Pensions”에는 국민연금, 공무원연금, 사학연금, 군인연금 등이 Nprog로 세분화되며, Typrog=“Old age cash benefits”에는 기초연금이 포함된다. 이러한 세분화는 제도 유형별 재정 부담을 파악하고, 향후 개별 제도 개혁의 영향을 평가하는 기초가 된다.

측정 단위는 국제 비교와 시계열 분석을 모두 지원하도록 다양하게 제공된다. 자국 통화 기준(현행가격/불변가격)은 각국 내부의 시간 흐름에 따른 실질 지출 변화를 파악하는 데 적합하며, USD PPP(구매력 평가) 기준은 국가 간 실질 구매력을 비교하는 데 적합하다. GDP 대비 비율은 경제 규모 대비 사회보장 부담 수준을 나타내며, 정부지출 대비 비율은 재정 우선순위를 나타낸다. 본 연구에서는 주로 GDP 대비 비율을 활용하여 국제 비교를 수행하며, 필요에 따라 1인당 USD PPP 지출을 보조 지표로 활용하였다.

한국 데이터는 1990년부터 제공되어 34년간의 시계열을 확보할 수 있다. 1990년대 초반 데이터는 일부 영역에서 결측이 있으나, 2000년 이후에는 9개 영역 전체에서 완전한 데이터가 제공된다. 특히 2018년 이후 예산 기준에서 결산 기준으로 전환되면서 정확성이 크게 향상되었으며, 이는 3장에서 분석한 바와 같이 2018년 전후 시계열 불연속을 초래하기도 하였다. 본 연구에서는 이러한 기준 변화를 고려하여 시계열 분석을 수행하였다.

## 나. 수집 결과 및 데이터 특성

SOCX 원천 데이터는 Excel 형식으로 제공된다. 데이터는 OECD 38개 회원국과 비회원 3개국<sup>14)</sup>, 44년(1980-2023) 시계열, 3개 제공 주체,

9개 정책영역, 2개 급여 형태, 38개 프로그램 유형 조합 그리고 국가별  $n$  개(한국의 경우 365개)의 프로그램을 포함한다. 예를 들어, 한국 노령 영역은 2021년 기준 7개 세부 프로그램(국민연금, 기초연금, 공무원연금, 사학연금, 군인연금, 장기요양서비스, 기타)으로 구성되며, 프로그램별로 연도별 지출이 별도로 기록된다.

데이터 처리 과정에서 파악한 중요한 구조적 특성은 제공 주체별로 집계 수준이 상이하다는 점이다. 공공 부문(SOURCE 10)은 세부 프로그램(nprog) 수준까지 공개되어 국민연금, 기초연금, 공무원연금 등 개별 제도의 지출을 구분할 수 있다. 의무적 민간(SOURCE 20)은 프로그램 유형(typrog) 수준의 소계만 제공되어 “Mandatory private pensions”와 같은 범주로만 집계된다. 임의적 민간(SOURCE 30)은 정책영역 전체 합계만 제공되어 “Old age - Voluntary private” 수준에서만 지출 규모를 파악할 수 있다.

이러한 차등적 공개 수준은 OECD 회원국 간 데이터 제공 합의의 한계를 반영한다. 공공 부문은 정부 예산 및 결산 자료를 기반으로 프로그램별 지출을 정확히 파악할 수 있으나, 민간 부문은 개별 금융기관 및 보험 회사의 자료를 수집해야 하므로 세분화된 정보 확보가 어렵다. 특히 임의적 민간 부문은 상업적 보험 및 연금 상품이 주를 이루어 공개 범위가 가장 제한적이다.

#### 다. 통화 변환 및 비율 계산

국제 비교 분석에서 핵심 과제는 서로 다른 통화와 경제 규모를 가진 국가들의 사회지출을 비교 가능한 지표로 변환하는 것이다. SOCX 원천

14) 불가리아, 크로아티아, 루마니아

데이터는 자국 통화 기준 지출액을 제공하나, 이를 직접 비교하면 환율 변동과 물가 차이의 영향을 받아 실질적인 지출 수준을 파악하기 어렵다. 본 연구에서는 OECD가 제공하는 거시경제 참조 데이터를 활용하여 3가지 표준화 지표를 계산하였다.

첫째, 자국 통화 지출액을 1인당 USD PPP(Purchasing Power Parity, 구매력 평가)로 변환하였다. PPP는 국가 간 물가 차이를 조정한 구매력 기준 환율로, 예를 들어 한국에서 1만 원으로 구매할 수 있는 재화·서비스와 동일한 양을 미국에서 구매하는 데 필요한 달러를 산출한다. OECD가 제공하는 국가별·연도별 PPP 환율을 적용하여 자국 통화 지출액을 USD PPP로 변환하고, 이를 인구수로 나누어 1인당 지출액을 계산하였다. 이는 국가 간 실질 구매력 기준 사회보장 수준을 비교하는 기본 지표가 된다.

둘째, 사회지출의 GDP 대비 비율을 계산하였다. 이는 사회보장 재정 지표의 가장 널리 활용되는 지표로, 국가 경제 규모 대비 사회보장 부담 수준을 나타낸다. 계산식은  $(\text{사회지출} / \text{GDP}) \times 100$ 이며, SOCX의 자국 통화 지출액과 OECD 국민계정의 GDP 데이터를 연계하여 산출하였다. GDP 대비 비율은 경제성장률과 사회지출 증가율의 상대적 속도를 반영하므로, 시계열 분석 시 재정 부담의 지속가능성을 평가하는 핵심 지표로 활용된다.

셋째, 정부 총지출 대비 사회지출 비율을 계산하였다. 이는 재정 우선순위를 나타내는 지표로, 정부가 전체 지출에서 사회보장에 얼마나 할당하는지를 보여준다. 계산식은  $(\text{사회지출} / \text{정부 총지출}) \times 100$ 이며, OECD 정부재정통계의 정부지출 데이터를 활용하였다. 동일한 GDP 대비 사회지출 비율을 가진 국가라도 정부 총지출 대비 비율이 다르다면, 이는 정부 규모(정부지출/GDP)의 차이를 반영한다. 예를 들어, 사민주의 국가들

은 정부 규모가 커서 GDP 대비 사회지출 비율이 높지만, 정부지출 대비 비율은 자유주의 국가와 유사한 경우가 많다.

이러한 파생 변수 계산을 위해 OECD 거시경제 데이터(GDP, PPP 환율, 인구, 정부지출)를 별도 참조 테이블로 구축하고, 원천 지출 데이터와 국가코드 및 연도 기준으로 결합하여 계산하였다. <표 4-1>은 이 과정을 의사코드(pseudo code) 형태로 보여준다.

<표 4-1> 통화 변환 및 비율 계산 절차

단계	처리 내용	의사코드
1	거시경제 데이터 결합	'결합(원천_지출_데이터, 거시경제_참조, 키=[국가코드, 연도])'
2	1인당 USD PPP 계산	'1인당_PPP = 지출액 / PPP환율 / 인구'
3	GDP 대비 비율 계산	'GDP비율 = (지출액 / GDP) × 100'
4	정부지출 대비 비율 계산	'정부지출비율 = (지출액 / 정부총지출) × 100'
5	결과 저장	'저장(비율_포함_지출_데이터)'

출처: 저자 작성.

라. 데이터 구조 예시

SOCX 데이터의 구조를 구체적으로 이해하기 위해 한국 노령 정책영역의 2021년 데이터를 예시로 살펴본다. <표 4-2>는 주요 프로그램별 지출을 자국 통화, GDP 대비 비율, USD PPP 등 다양한 측정 단위로 보여준다.

프로그램 수준 세분화는 제도 개혁의 재정적 영향을 정밀하게 추적할 수 있게 한다. 예를 들어, 한국의 노령 지출은 2014년 기초연금 도입 이후 급격히 증가하였는데, 프로그램 수준 데이터가 없다면 이 증가가 국민

연금 성숙에 의한 것인지 기초연금 신설에 의한 것인지 구분하기 어렵다. <표 4-2>에서 국민연금 노령연금(1.0%)과 기초연금(0.9%)이 비슷한 수준임을 파악할 수 있으며, 이는 2014년 이전에는 존재하지 않던 기초연금이 단기간에 국민연금과 대등한 재정 규모로 성장했음을 보여준다.

<표 4-2> SOCX 데이터 구조 예시: 한국 노령 정책영역(2021년)

프로그램명	SOURCE	자국통화 (조원)	GDP 대비 (%)	USD PPP (달러/1인)	급여형태
국민연금(노령)	10	21.3	1.02	496.3	현금
기초연금	10	18.5	0.89	432.6	현금
노인돌봄서비스	10	0.6	0.03	14.1	현물
개인연금 등	20	23.8	0.93	451.3	현금

주: SOURCE 10=공공, 20=법적 민간, 국민연금(노령)은 조기노령연금 제외.  
출처: OECD SOCX(2025) 활용 저자 재작성.

또한 급여 형태 구분을 통해 현금급여와 현물급여의 재정 배분을 파악할 수 있다. 노인돌봄서비스(0.03%)는 현물급여로 분류되며, 이는 노령 지출의 대부분이 현금급여(연금)로 구성되고 서비스(돌봄) 비중은 상대적으로 낮음을 보여준다. 이러한 구조는 사민주의 국가들의 현물 서비스 중심 노령보장과 대조되며, 5장에서 한국의 특성을 분석하는 기초가 된다.

법적 민간 부문(SOURCE 20)의 개인연금은 0.93%로 기초연금과 유사한 수준이다. 이는 한국의 노령보장이 공공 부문과 민간연금이 2:1 수준으로 분담하고 있음을 의미한다. 앞서 언급한 바와 같이 민간 부문은 세부 프로그램 수준 정보가 제공되지 않아 전체 합계만 파악 가능하며, 세부 유형(연금저축, IRP 등)은 Country Notes와 ISSA 정보를 통해 보완이 필요한 부분이다.

## 2. Country Notes 정성 데이터 수집

### 가. Country Notes의 역할과 수집 범위

SOCX 정량 데이터는 각국의 사회보장 지출 규모를 측정하는 기초 자료이나, 그러한 지출이 어떤 제도적 특성에서 비롯되었는지에 대한 맥락 정보는 제공하지 않는다. 예를 들어, 앞서 <표 4-2>에서 확인한 한국 기초연금의 GDP 대비 지출 비율 0.89%라는 수치만으로는 이 프로그램이 보험료를 재원으로 하는지 조세를 재원으로 하는지, 전체 노인을 대상으로 하는지 저소득층만을 대상으로 하는지, 정액으로 지급되는지 소득에 비례하여 지급되는지를 알 수 없다. 이러한 제도적 맥락을 제공하는 것이 SOCX Country Notes의 역할이다.

Country Notes는 SOCX 정량 데이터베이스의 보완 문서로서, OECD 회원국별로 작성되어 PDF 형식으로 제공된다. 본 연구에서는 OECD 38개 회원국과 비회원 3개국인 불가리아, 크로아티아, 루마니아를 포함한 총 41개국의 Country Notes를 수집하였다. 다만 본 연구의 분석 범위는 보건 영역과 적극적노동시장정책 영역은 외삽 자료로 제외되므로, 실제 수집 대상은 41개국 × 7개 영역으로 총 287개 국가-영역 조합이다. 이는 약 2,500페이지에 달하는 서술형 텍스트에 해당한다.

### 나. LLM 기반 정보 연계 방법론

Country Notes의 핵심 특성은 표준화와 유연성의 균형에 있다. OECD는 각국에 동일한 템플릿을 제공하여 정책영역별 장 구성과 프로그램별 소절 구성을 표준화하고 있으나, 각 프로그램에 대한 서술 내용은

국가별 제도 특성을 반영하여 자유롭게 작성된다. 이는 국제 비교 가능성을 확보하면서도 각국의 제도적 특수성을 충실히 반영할 수 있다는 장점이 있으나, 정보 추출의 관점에서는 단순한 키워드 매칭이나 정규표현식 만으로는 정확한 구조화가 어렵다는 과제를 제기한다.

예를 들어, 한국 국민연금의 급여 수준은 Country Notes에서 “기본연금액의 1.35배를 적용하되, 2028년까지 연차적으로 1.2배로 하향 조정할 예정”이라고 서술되어 있다. 이 정보를 구조화하기 위해서는 현재 급여율이 1.35배임을 파악하고, 향후 개혁 계획으로 1.2배로 변경될 예정임을 인식하며, 2028년이라는 시점을 추출해야 한다. 단순 키워드 추출로는 “1.35”, “1.2”, “2028” 등의 숫자만 얻을 수 있으나, 이들 간의 관계와 의미를 파악하기는 어렵다. 이러한 맥락적 이해가 필요한 서술형 정보를 체계적으로 처리하기 위해, 본 연구에서는 제2절에서 설명할 대규모 언어모델 기반 자연어 처리 방법론을 개발하였다.

Country Notes의 또 다른 구조적 특성은 Code4 프로그램 코드와의 연계 가능성이다. Country Notes의 각 프로그램을 설명하고 있으며, 해당 프로그램은 SOCX 정량 데이터베이스의 Code4로 식별된다. 예를 들어, 한국 Country Notes의 “Basic Pension”은 SOCX 정량 데이터의 “410.10.1.1.1.6” 코드와 대응된다. 이러한 연계는 본 연구 데이터베이스 구축의 핵심 원리로서, 정량 데이터의 지출 규모 정보와 정성 데이터의 제도 특성 정보를 통합하는 기반이 된다.

### 3. ISSA 데이터 수집

#### 가. ISSA의 역할과 특성

ISSA는 International Social Security Association의 약어로, 제네바에 본부를 둔 국제 사회보장기구이다. ISSA는 전 세계 190개국의 사회보장제도 운영 실태를 수집하여 국가별 프로파일 형태로 공개하고 있다. SOCX Country Notes가 OECD 회원국에 한정되고 SOCX 코드 체계와 직접 연동되어 있다면, ISSA는 비OECD 국가까지 포괄하며 제도영역별로 표준화된 항목을 일관되게 제공한다는 차이가 있다.

ISSA Country Profiles의 구조는 제도영역별로 구성되며, 각 제도영역은 법적 근거, 제도 유형, 적용 대상, 자원 조달, 수급 요건, 급여 수준, 관리 기관 등의 표준 항목을 포함한다. 이러한 표준화된 구조는 국가 간 비교를 용이하게 하는 장점이 있으나, 각 항목의 서술 분량이 Country Notes보다 간결하여 제도의 역사적 맥락이나 최근 개혁 과정에 대한 정보는 상대적으로 제한적이다.

ISSA와 Country Notes의 가장 중요한 차이는 정보의 조직 방식에 있다. Country Notes는 SOCX 정량 데이터의 Code4 프로그램별로 구성하여 일대일 대응이 가능한 반면, ISSA는 제도영역 카테고리로서 서술하므로 개별 Code4 프로그램과의 직접 대응은 어렵다. 예를 들어, 한국 Country Notes의 노령에는 국민연금, 기초연금, 공무원연금과 같은 각각의 프로그램 설명으로 구성되어 있으나, ISSA의 ‘노령, 장애 및 유족’ 영역에는 이들이 혼재되어 서술되어 있다. 따라서 ISSA 정보를 Code4 프로그램별로 연계하기 위해서는 맥락 기반의 별도 연계 작업이 필요하며, 이는 제2절에서 설명할 LLM 기반 데이터 연계 프로세스에 포함된다.

## 나. 웹 크롤링 방법 및 기술적 구현

ISSA 데이터는 [www.issa.int/databases/country-profiles](http://www.issa.int/databases/country-profiles)를 통해 제공되며, 각국 Country Profile 페이지에서 제도영역별 정보를 HTML 형식으로 열람할 수 있다. 다만 SOCX와 달리 일괄 다운로드 기능이 제공되지 않아, 본 연구에서는 Python 기반 웹 크롤링을 통해 데이터를 자동 수집하였다.

크롤링 스크립트는 다음과 같이 구성하였다. 첫째, requests 라이브러리를 통해 각국 URL에 HTTP 요청을 보내고 HTML을 수신한다. 둘째, BeautifulSoup 라이브러리를 통해 HTML 구조를 파싱하고 제도영역별 섹션을 식별한다. ISSA 웹사이트는 제도영역을 <div class="scheme-type"> 태그로 구분하고 있어, CSS Selector를 통해 각 영역을 추출할 수 있다. 셋째, 각 제도영역 내에서 법적 근거(Legal basis), 제도 유형(Type of program), 적용 대상(Coverage), 재원 조달(Financing), 수급 요건(Qualifying conditions), 급여 수준(Benefit amount) 등 항목별 정보를 추출한다. 넷째, 추출된 정보를 국가-제도영역-항목 3차원 구조의 JSON 형식으로 변환하여 저장한다.

크롤링 과정에서 고려한 기술적 사항은 다음과 같다. 첫째, 웹사이트 부하를 최소화하기 위해 요청 간 3초 지연(time.sleep)을 적용하였다. 둘째, 일부 국가는 페이지 구조가 상이하거나 정보 누락이 있어, 예외 처리(try-except)를 통해 크롤링 실패 시에도 다음 국가로 진행되도록 하였다. 셋째, HTML 원본을 별도로 보존하여 향후 재파싱 또는 검증이 필요할 경우 대비하였다. 전체 크롤링은 OECD 38개국 × 5개 주요 제도영역(노령·장애·유족, 질병·출산, 산업재해, 실업, 가족급여)에 대해 수행되었다.

## 제2절 SOCX, ISSA 데이터 연계

제1절에서 수집한 세 가지 원천 데이터는 각기 다른 형식과 구조를 가지고 있다. SOCX 정량 데이터는 Code4 프로그램 코드를 중심으로 정형화된 데이터베이스 구조를 갖추고 있으나, Country Notes와 ISSA는 서술형 텍스트로 작성되어 있어 Code4 프로그램과의 직접적인 연계가 어렵다. 본 절에서는 이러한 서로 다른 형식의 데이터를 Code4 기준으로 통합하기 위한 LLM 기반 데이터 연계 방법론을 제시한다.

데이터 연계의 핵심 과제는 약 9,400개의 Code4 프로그램 각각에 대해 Country Notes와 ISSA의 전체 텍스트에서 관련 정보를 정확히 추출하는 것이다. 이는 단순한 키워드 매칭으로는 불가능하며, 텍스트의 의미론적 맥락을 이해하고 프로그램 간 관계를 파악할 수 있는 고도의 자연어 처리 능력을 요구한다. 본 연구에서는 대규모 언어모델을 활용하여 이러한 과제를 해결하였으며, 그 결과 정량 데이터와 정성 데이터를 프로그램 수준에서 체계적으로 통합할 수 있었다.

본 절은 3개 소절로 구성된다. 첫째, 데이터 연계의 개념과 필요성을 명확히 하고 전통적 방법론의 한계를 제시한다. 둘째, LLM 기반 연계 프로세스의 전체 구조와 각 단계 처리 방법을 상세히 기술한다. 셋째, 연계 프롬프트의 설계 원리와 구체적인 프롬프트 구성 방식을 설명하고 실제 결과를 제시한다.

### 1. 정량·정성 데이터의 연계 개념

#### 가. Code4 기반 데이터 통합의 필요성

제1절에서 수집한 SOCX 정량 데이터는 Code4 프로그램 코드를 통해

각 프로그램을 고유하게 식별한다. 예를 들어, 한국 기초연금은 “410.10.1.1.1.6”이라는 Code4로 식별되며, 이 코드는 국가코드 410(한국), 제공 주체 10(공공), 정책 영역 1(노령), 급여 형태 1(현금), 프로그램 유형 1, 세부 프로그램 6이라는 계층 구조를 내포하고 있다. SOCX 데이터베이스에서 이 Code4는 2021년 기준 GDP 대비 0.89%의 지출 규모, 18.5조 원의 자국 통화 지출액, 1인당 432.6달러의 PPP 지출액 등 정량 정보와 연결되어 있다.

반면 Country Notes는 각국 PDF 문서로 제공되나, Code4의 식별자가 명확히 명시되지 않고 범위로 명시되는 경우도 많다. 또한 프랑스의 경우에는 Country Notes에 프로그램에 대한 정보가 존재하지 않는다.

ISSA는 이보다 더 복잡한 구조를 갖는다. ISSA의 한국 프로파일에서 노령 관련 정보는 “Old age, invalidity and survivors” 제도 영역 전체에 대해 통합적으로 서술되며, 그 안에서 “Basic old-age pension(social assistance, income tested): Age 65. Income test: Monthly income must not exceed 1,310,000 won for a single person”과 같이 기초연금에 대한 정보가 포함되어 있다. 이 텍스트 역시 Code4와 직접 연결되어 있지 않으며, 국민연금, 공무원연금 등 다른 프로그램에 대한 서술과 혼재되어 있다.

따라서 정량 데이터의 Code4와 정성 데이터의 관련 텍스트를 연계하기 위해서는 다음과 같은 단계가 필요하다. 첫째, Country Notes와 ISSA의 전체 텍스트에서 “Basic Pension”이라는 프로그램명과 관련된 모든 문장을 식별해야 한다. 둘째, 식별된 문장들이 실제로 Code4 “410.10.1.1.1.6”에 해당하는 프로그램을 설명하는지 확인해야 한다. 프로그램명이 여러 맥락에서 언급될 수 있기 때문이다. 셋째, 두 원천에서 추출된 텍스트를 원천별로 구분하여 통합해야 한다. 이러한 과정을

9,400개 모든 Code4 프로그램에 대해 반복 수행하는 것이 데이터 연계 작업의 목표이다.

## 나. 전통적 방법론의 한계

서술형 텍스트에서 특정 정보를 추출하는 전통적인 방법론은 크게 세 가지로 분류할 수 있다. 첫째, 키워드 매칭은 미리 정의된 키워드가 포함된 문장을 추출하는 방식이다. 예를 들어, “Basic Pension”이라는 키워드가 포함된 모든 문장을 추출할 수 있다. 이 방법은 구현이 간단하고 처리 속도가 빠르다는 장점이 있으나, 동일한 키워드가 서로 다른 맥락에서 사용되는 경우를 구분하지 못한다. Country Notes에서 “The Basic Pension supplements the National Pension”이라는 문장은 기초연금이 아닌 국민연금을 설명하는 맥락에서 기초연금을 언급한 것이므로, 이를 기초연금의 설명으로 분류하는 것은 부적절하다.

둘째, 정규표현식은 특정 패턴을 가진 텍스트를 추출하는 방식이다. 예를 들어, “Basic Pension: [설명]”과 같이 프로그램명 뒤에 콜론이 오는 패턴을 정의할 수 있다. 이 방법은 구조화된 텍스트에서는 효과적이거나, Country Notes와 ISSA는 서술 방식이 국가별로 상이하여 일관된 패턴을 정의하기 어렵다. 한국 Country Notes는 소절 제목으로 “Basic Pension”을 사용하는 반면, 일부 유럽 국가들은 “Means-tested old-age allowance(기초연금에 해당)”와 같이 제도 유형을 우선 기술하고 괄호 안에 프로그램명을 보충한다.

셋째, 개체명 인식은 텍스트에서 고유명사를 식별하고 분류하는 방식이다. 자연어 처리 분야에서 널리 활용되는 방법으로, 사람 이름, 장소, 조직명 등을 자동으로 추출할 수 있다. 그러나 사회보장 프로그램명은 일

반적인 개체명 인식 모델이 학습한 범주에 포함되지 않으며, 프로그램명과 제도 유형이 혼용되는 경우가 많아 정확한 추출이 어렵다. 예를 들어, “social assistance”는 제도 유형이지만 일부 국가에서는 특정 프로그램의 명칭으로도 사용된다.

이러한 전통적 방법론의 한계는 근본적으로 텍스트의 의미론적 맥락을 이해하지 못하는 데서 비롯된다. 한 문장이 특정 프로그램을 직접 설명하는지, 다른 프로그램과의 관계를 설명하는 과정에서 부수적으로 언급하는지, 역사적 변천 과정을 설명하는지를 구분하려면 문장의 의미를 이해해야 한다. 또한 Country Notes와 ISSA는 동일한 프로그램을 서로 다른 용어로 지칭하는 경우가 있어, 용어 변이를 처리할 수 있는 능력도 필요하다. 본 연구에서는 이러한 과제를 해결하기 위해 대규모 언어모델을 활용하였다.

## 2. LLM 기반 연계 프로세스

### 가. 전체 프로세스 구조

LLM 기반 데이터 연계는 [그림 4-2]와 같이 총 3단계로 구성된다. 1단계는 입력 준비로, SOCX 데이터베이스에서 Code4 프로그램 목록과 메타정보를 추출한다. 2단계는 데이터 원천 정보들의 결합이다. Country Notes에서는 각 프로그램에 대해 Country Notes PDF에서 관련 텍스트를 추출하며, ISSA 데이터에서도 동일한 프로그램에 대해 ISSA 웹 데이터에서 관련 텍스트를 추출한다. 3단계는 두 원천의 정보를 통합하여 하나의 JSON 객체로 저장한다. 3단계는 저장된 JSON을 검증하고 제3절의 분류 작업을 위한 입력으로 제공한다.

[그림 4-2] 데이터 연계 프롬프팅 구조

프롬프팅 (연계)

프롬프트 엔지니어링 5요소를 데이터 연계 전략에 맞게 재구성

Task Format → ①

Context → ②

Rules Examples → ③

① 필수 요구사항

입력: Code4 (국가, dsource, branch, type\_exp, type\_prog, nprog) + 프로그램명  
출력: 통합 JSON (SOCX Country Notes + ISSA 관련 정보)  
원천 구분: SOCX Country Notes / ISSA 명시

② 데이터 원천 정보

SOCX Country Notes: Code4별 프로그램 설명 (PDF, 검색 있음)  
ISSA: 제도영역 전체 정보 → Code4 프로그램과 매칭되는 정보 추출

③ 추가 지시사항

원천 간 정보 통합: Country Notes와 ISSA를 종합하여 최대 정보 확보  
정보 추출 기준: Code4 특성에 부합하는 정보 식별 및 추출  
충분성 검증: 수급 조건, 급여 방식 등 핵심 정보 포함 여부 확인

출처: 저자 작성.

1단계 입력 준비에서는 SOCX 데이터베이스를 조회하여 연계 대상 프로그램을 확정한다. 본 연구의 분석 범위인 41개국 × 7개 정책영역 × SOURCE 10(공공) 조건을 만족하는 Code4 프로그램을 SQL 쿼리로 추출하며, 이는 총 9,400개에 해당한다. 각 프로그램에 대해 Code4 코드, 프로그램명, 국가 코드, 국가명, 정책영역 코드, 정책영역명 등의 메타정보를 수집한다. 예를 들어, 한국 기초연금의 경우 Code4 “410.10.1.1.1.6”, 프로그램명 “Basic Pension”, 국가명 “Korea”, 정책영역명 “Old age”라는 메타정보가 준비된다.

2단계 데이터 원천 정보들을 결합하는 과정이다. 먼저 Country Notes에서는 1단계에서 준비된 메타정보를 바탕으로 해당 국가의

Country Notes PDF를 읽어 들인다. 한국의 경우 전체 Country Notes 문서에서 노령만을 추출한다. 이 전체 텍스트를 LLM에 입력하고, “Basic Pension”과 관련된 단락을 추출하도록 지시하는 프롬프트를 제공한다. LLM은 제공된 전체 정보에서 기초연금을 직접 설명하는 텍스트를 추출한다. 두 번째는 ISSA 데이터에서 동일한 프로그램에 대한 처리를 진행한다. 한국의 경우 “Old age, invalidity and survivors” 제도영역 전체 텍스트를 읽어 들이며 이 텍스트를 LLM에 입력하고, “Basic Pension” 또는 “Basic old-age pension”과 관련된 단락을 추출하도록 지시한다. LLM은 ISSA의 “Old age, invalidity and survivors” 제도영역 전체 텍스트에서 기초연금 관련 텍스트를 추출한다. 3단계는 정보의 통합이다. 1단계에서 정한 프로그램 정보에 2단계에서 추출된 텍스트를 하나의 JSON 객체로 결합하는 과정이다. JSON 구조는 Code4, 프로그램명, 국가 정보 등의 메타데이터 필드와, Country Notes 텍스트 필드, ISSA 텍스트 필드로 구성된다. 원천별 텍스트는 별도 필드에 저장되어, 제3절의 분류 작업에서 두 원천의 정보를 구분하여 활용할 수 있도록 한다. 저장된 JSON 파일은 Code4를 파일명으로 사용하여 “410.10.1.1.1.6.json”과 같이 저장된다.

#### 나. LLM 모델 선정 및 실행 환경

데이터 연계 작업을 위한 LLM 모델 선정에서는 처리 정확도, 처리 속도, 비용 효율성, 데이터 보안의 네 가지 요소를 종합적으로 고려하였다. 먼저 처리 정확도 측면에서는 서술형 텍스트에서 특정 프로그램 관련 단락을 정확히 식별할 수 있는 의미 이해 능력이 필요하다. 처리 속도 측면에서는 9,400개 프로그램에 대해 연계 작업을 수행해야 하므로, 프로그

램당 처리 시간이 짧을수록 유리하다. 비용 효율성 측면에서는 대량 처리를 고려할 때 API 호출 비용이 중요한 요소이다. 데이터 보안 측면에서는 Country Notes와 ISSA 텍스트가 외부 서버로 전송되지 않도록 로컬 실행이 가능한 모델이 선호된다.

이러한 요소를 종합하여 본 연구에서는 Google의 Gemma-3-27B-IT 모델을 선택하였다. Gemma-3-27B-IT는 270억 개의 파라미터를 가진 오픈소스 언어모델로, Instruction Tuning 버전은 사용자의 지시를 따르는 능력이 최적화되어 있다. 이 모델은 로컬 서버에서 실행 가능하여 데이터 보안 문제가 없으며, API 호출 비용이 발생하지 않는다. 처리 정확도는 사전 테스트 결과 OpenAI GPT-4와 유사한 수준을 보였으며, 처리 속도는 NVIDIA A100 GPU 2개를 활용한 환경에서 프로그램당 평균 2분 이내로 적절한 수준이었다.

실행 환경은 다음과 같이 구성하였다. 하드웨어는 NVIDIA A100 80GB GPU 4개를 장착한 리눅스 서버를 사용하였다. 소프트웨어는 vLLM 프레임워크를 활용하여 모델 추론을 최적화하였으며, 데이터 입출력은 Pandas 및 JSON 표준 라이브러리를 활용하였다.

### 3. 원천별 텍스트 추출 및 통합

[그림 4-3]은 한국 기초연금을 예시로 데이터 연계 프로세스의 3단계를 보여준다. 1단계에서는 SOCX 정량 데이터베이스에서 Code4 “410.10.1.1.1.6”과 프로그램명 “Basic Pension”을 입력 정보로 준비한다. 2단계에서는 Country Notes와 ISSA에서 이 프로그램과 관련된 텍스트를 추출한다. 3단계에서는 두 원천의 정보를 Code4 기준으로 통합하여 저장한다.

[그림 4-3] 데이터 연계 프로세스: 한국 기초연금 예시



출처: 저자 작성.

## 가. Country Notes에서의 텍스트 추출

Country Notes는 각국별 PDF 문서로 제공되며, 9개 정책영역이 표로 구분되어 있다. 한국 Country Notes의 노령은 국민연금, 기초연금, 공무원연금, 사학연금, 군인연금, 노인돌봄서비스 등 여러 프로그램에 대한 설명을 포함하고 있다. [그림 4-2]의 2단계에서 볼 수 있듯이, 컨트리 노트 전체 텍스트에서 기초연금과 관련된 단락을 추출하는 것이 Country Notes 텍스트 추출의 목표이다.

한국 기초연금의 경우, Country Notes에서 추출된 핵심 텍스트는 다음과 같다. “The Basic Pension was introduced in 2008 to provide income assistance to the elderly as a complement of the National Pension. Senior citizens aged 65 and over in the lowest 70% income bracket receive the non-contributory, non-taxable monthly benefits.” 이 텍스트는 제도 도입 연도, 정책 목표, 수급 대상, 재원 조달 방식, 급여 형태 등 제3절의 분류 작업에서 필요한 핵심 정보를 포함하고 있다.

Country Notes 텍스트 추출에서 중요한 점은 프로그램명이 여러 맥락에서 언급되는 경우를 구분하는 것이다. 예를 들어, 국민연금을 설명하는 단락에서 “The Basic Pension supplements the National Pension”과 같이 기초연금이 부수적으로 언급될 수 있다. 이러한 문장은 기초연금의 특성을 직접 설명하는 것이 아니므로 추출 대상에서 제외해야 한다. LLM은 문장의 의미론적 맥락을 이해하여 이러한 구분을 수행한다.

## 나. ISSA에서의 텍스트 추출

ISSA는 제도영역별로 정보를 제공하며, 한국의 경우 “Old age, invalidity and survivors” 제도영역에서 노령, 장애, 유족 관련 모든 프로그램을 통합하여 서술한다. 이 제도영역의 텍스트는 약 8페이지 분량이며, 국민연금, 기초연금, 공무원연금뿐만 아니라 장애연금, 유족연금 등도 함께 포함되어 있다. Country Notes와 달리 프로그램별 소절 구분이 명확하지 않아, 전체 텍스트에서 기초연금 관련 정보를 찾아내는 것이 더욱 중요하다.

한국 기초연금의 경우, ISSA에서 추출된 핵심 텍스트는 다음과 같다. “Basic old-age pension(social assistance, income tested): Age 65. Income test: Monthly income must not exceed 1,310,000 won for a single person or 2,096,000 won for a couple. Basic old-age pension(social assistance): Up to 209,960 won a month is paid, depending on the beneficiary's income.” 이 텍스트는 Country Notes보다 정량적 파라미터가 구체적으로 제시되어 있다는 특징이 있다.

ISSA 텍스트 추출에서 주의할 점은 프로그램명이 원천에서 다른 형태로 표현될 수 있다는 것이다. 한국 기초연금은 Country Notes에서 “Basic Pension”으로 지칭되지만, ISSA에서는 “Basic old-age pension”으로 지칭된다. 또한 제도 유형을 함께 표기하여 “Basic old-age pension(social assistance)”와 같이 괄호 안에 추가 정보를 제공하는 경우도 있다. LLM은 이러한 명칭 변이를 인식하고 의미적으로 동일한 프로그램을 지칭하는 텍스트를 추출한다.

## 다. 두 원천 정보의 통합 및 상호보완

[그림 4-3]에서 보여주는 통합 출력은 Country Notes와 ISSA에서 추출된 텍스트를 원천별로 구분하여 저장한 결과이다. 한국 기초연금의 경우, Country Notes에서는 “2008년 도입, 국민연금 보완”, “65세 이상 하위 70% 소득계층 대상”, “비기여, 비과세 월 급여”, “소득 수준에 따라 급여액 차등”이라는 정보가 추출되었다. ISSA에서는 “사회부조 방식, 소득심사”, “65세 수급 연령”, “소득기준: 단독 1,310,000원, 부부 2,096,000원”, “월 최대 209,960원 지급”이라는 정보가 추출되었다.

두 원천의 정보를 비교하면 상호보완적 관계가 명확히 드러난다. Country Notes는 제도 도입 배경과 정책 목표를 서술하여 “국민연금 보완”이라는 정책 맥락을 제공한다. 반면 ISSA는 구체적인 수치를 제공하여 “단독 1,310,000원”이라는 정확한 소득 기준과 “최대 209,960원”이라는 급여액을 명시한다. 재원 조달 방식은 Country Notes에서 “non-contributory”로, ISSA에서 “social assistance”로 표현되어 동일한 의미를 서로 다른 용어로 설명한다. 수급 대상은 Country Notes에서 “lowest 70% income bracket”으로, ISSA에서 “income tested”로 표현되어 선별적 구조임을 공통적으로 설명한다.

이러한 통합 정보는 제3절의 분류 작업에서 입력으로 사용된다. 제3절에서는 이 텍스트를 분석하여 scheme\_type을 “social\_allowance”로, coverage\_type을 “means\_tested”로, benefit\_type을 “flat\_rate”로 분류한다. Country Notes의 “non-contributory”와 ISSA의 “social assistance”를 종합하여 조세 기반의 사회수당임을 판단하고, 두 원천의 “lowest 70%”와 “income tested”를 종합하여 선별적 대상임을 판단하며, ISSA의 “월 최대 209,960원”이라는 정액 급여를 근거로 정액형임을

판단한다. 이처럼 두 원천의 통합은 단순한 정보 누적이 아니라, 상호 검증과 보완을 통해 보다 정확한 분류를 가능하게 한다.

## 제3절 LLM 기반 제도 정보 구조화

제2절에서 구축한 연계 데이터는 Code4 프로그램별로 Country Notes와 ISSA의 텍스트를 포함하지만, 이는 여전히 비구조화된 자연어 형태이다. 제3절에서는 이러한 텍스트를 제도 유형, 대상 유형, 급여 유형의 3차원으로 분류하여 구조화된 데이터로 변환하는 과정을 다룬다. 이 작업 역시 LLM을 활용하여 수행하며, 분류 체계 설계, 프롬프트 구조, 실제 분류 결과를 순서대로 제시한다.

### 1. 분류 체계의 적용

#### 가. 분류체계 정의

SOCX 데이터베이스는 Code4를 통해 프로그램의 정책영역, 급여 형태 등 일부 특성을 표현하지만, 제도유형이나 대상 유형과 같은 질적 특성은 포함하지 않는다. 예를 들어, 한국 기초연금과 독일 법정연금보험은 모두 Code4에서 노령 영역의 현금 급여로 분류되지만, 재원 조달 방식(조세 vs 보험료)과 수급 대상(선별 vs 기여자)이 근본적으로 다르다. 이러한 질적 차이를 체계적으로 표현하기 위해 3차원 분류 체계를 설계하였다.

첫 번째 차원은 제도 유형(scheme\_type)으로, 재원 조달 방식을 기준으로 한다. 사회보험(social\_insurance)은 보험료 기여를 기반으로 하

며, 사회수당(social\_allowance)은 조세 재원을 기반으로 하고, 사회서비스(social\_service)는 현물 서비스를 직접 제공한다. 이 분류는 복지국가 유형론에서 제도 재정의 지속가능성과 재분배 효과를 분석하는 핵심 기준이다.

두 번째 차원은 대상 유형(coverage\_type)으로, 수급 자격 조건을 기준으로 한다. 보편형(universal)은 연령이나 거주 요건만으로 수급 가능하며, 기여형(contributory)은 과거 기여 이력을 요구하고, 자산조사형(means\_tested)은 소득이나 자산 심사를 통과해야 한다. 이 분류는 복지 접근성과 낙인 효과를 분석하는 기준이 된다.

세 번째 차원은 급여 유형(benefit\_type)으로, 급여 산정 방식을 기준으로 한다. 현금 급여는 정액형(flat\_rate), 소득비례형(earnings\_related), 계정기반형(account\_based), 혼합형(mixed)으로 세분화되며, 현물급여는 서비스(service), 바우처(voucher), 보조금(subsidy), 혼합형(mixed)으로 구분된다. 이 분류는 급여 적절성과 형평성을 평가하는 기준이 된다.

## 나. 분류 기준 상세

제도유형 분류에서 가장 중요한 기준은 보험료 납부 여부이다. Country Notes나 ISSA에서 “contribution”, “premium”, “payroll tax”, “18.6% of earnings” 등의 표현이 나타나면 사회보험으로 분류한다. 이러한 표현이 없고 “tax-funded”, “general revenue”, “non-contributory” 등이 나타나면서 현금 급여를 제공하면 사회수당으로 분류한다. 현물 서비스를 제공하는 경우 “care services”, “day care center”, “home help” 등의 표현과 함께 사회서비스로 분류한다.

대상 유형 분류에서는 수급 조건의 엄격성을 평가한다. “all residents aged 65 and over”, “every citizen”과 같이 연령이나 시민권만을 요구하면 보편형으로 분류한다. “at least 5 years contribution”, “contribution period required”, “insured workers”와 같이 과거 기여 이력을 요구하면 기여형으로 분류한다. “lowest 70% income bracket”, “income tested”, “means-tested”, “asset test”와 같이 소득이나 자산 심사를 요구하면 자산조사형으로 분류한다.

급여유형 분류에서는 급여 산정 방식을 판단한다. “fixed amount”, “flat-rate”, “209,960 won a month”와 같이 정액을 지급하면 정액형으로 분류한다. “based on earnings”, “earnings-related”, “1.5% of average earnings”와 같이 소득에 비례하면 소득비례형으로 분류한다. “individual account”, “accumulated contributions”와 같이 개인 계정 누적액을 기반으로 하면 계정기반형으로 분류한다. 정액과 소득비례를 결합하면 혼합형으로 분류한다.

## 다. 분류 규칙의 논리적 일관성

분류 체계의 신뢰성을 확보하기 위해 논리적으로 모순되는 조합을 금지 규칙으로 명시하였다. 첫째, 사회수당과 기여형의 조합을 금지한다. 사회수당은 비기여 기반 조세 재원을 전제로 하는데, 기여형은 과거 기여 이력을 요구하므로 논리적으로 양립할 수 없다. 이러한 조합이 나타나면 제도유형을 사회보험으로 재평가해야 한다.

둘째, 사회서비스와 소득비례형의 조합을 금지한다. 사회서비스는 현물을 제공하는 것이므로, 소득비례형이라는 현금급여 산정 방식이 적용될 수 없다. 이러한 조합이 나타나면 급여 유형을 서비스, 바우처, 보조금

중 하나로 재평가해야 한다.

이러한 금지 규칙은 LLM이 텍스트 분석 과정에서 논리적 오류를 범하는 것을 방지한다. 프롬프트에 금지 조합을 명시하고, 출력된 분류 결과를 검증하는 단계에서 이러한 조합이 발견되면 프로그램을 재처리하도록 설계하였다.

## 2. 데이터 구조화 프롬프트 설계

### 가. 데이터 구조화 프롬프트 구조

[그림 4-4]는 분류 작업을 위한 LLM 프롬프트의 구조를 보여준다. 프롬프트는 세 개의 핵심 요소로 구성된다. ① 필수 요구사항은 분류 대상, 출력 형식, 판단 근거를 명시한다. ② 분류 기준은 제도 유형, 대상 유형, 급여 유형 각각의 판단 기준을 상세히 제시한다. ③ 추가 지시 사항은 금지 조합, 판단 순서, 검증 체크리스트를 제공한다.

첫 번째, 필수 요구사항 부분은 LLM이 수행해야 할 작업의 범위와 출력 형식을 정의한다. 분류 대상으로 `scheme_type`, `coverage_type`, `benefit_type` 세 개 필드를 명시하고, 각 필드는 사전 정의된 값 중 하나를 선택해야 함을 지시한다. 출력 형식은 JSON 구조로 정의하며, 각 분류 필드와 함께 `reasoning` 필드를 반드시 포함하도록 한다. `reasoning` 필드에는 재원 방식, 수급 조건, 급여 산정 방식에 대한 판단 근거를 텍스트에서 추출한 키워드와 함께 기술해야 한다.

두 번째, 분류 기준 부분은 차원별 구체적인 판단 기준을 제시한다. 제도 유형 판단을 위해 보험료 납부 관련 키워드 목록과 조세 재원 관련 키워드 목록을 제공한다. 대상 유형 판단을 위해 보편형, 기여형, 자산조사

형을 구분하는 키워드 목록을 제공한다. 급여 유형 판단을 위해 정액형, 소득비례형, 계정기반형, 혼합형을 구분하는 키워드와 현물급여의 세부 유형을 제공한다. 이러한 키워드 목록은 사전 테스트 과정에서 실제 Country Notes와 ISSA 텍스트에서 빈번하게 나타나는 표현을 수집하여 구성하였다.

세 번째, 추가 지시 사항 부분은 분류 과정에서 발생할 수 있는 오류를 방지하는 규칙을 명시한다. 금지 조합을 구체적으로 나열하고, 이러한 조합이 나타나는 경우 어떻게 재평가해야 하는지 지시한다. 판단 순서를 "보험료 납부 여부 → 현금/현물 구분 → 급여 산정 방식"으로 제시하여 체계적인 분류를 유도한다. 검증 체크리스트를 제공하여 LLM이 자체적으로 출력 결과를 점검하도록 한다.

[그림 4-4] 데이터 구조화 프롬프팅 구조

**프롬프팅**

프롬프트 엔지니어링 5요소를 데이터 구조화 전략에 맞게 재구성  
Task Format → ① | Context → ② | Rules Examples → ③

① 필수 요구사항

② 분류 기준

③ 추가 지시사항

**① 필수 요구사항**

- 분류 대상: scheme\_type (제도유형), coverage\_type (대상유형), benefit\_type (급여유형)
- 출력 형식: JSON 구조로 3개 자질 반환
- 판단 근거: reasoning 필드에 재원수급급여 방식 명시

**② 분류 기준**

- 제도유형: social\_insurance / social\_allowance / social\_service (재원 방식)
- 대상유형: universal / contributory / means\_tested (수급 조건)
- 급여유형: Cash (flat\_rate/earnings\_related 등) / In-kind (service/voucher 등)

**③ 추가 지시사항**

- 금지 조합: allowance+contributory, service+earnings\_related (논리 모순 방지)
- 판단 순서: 보험료 납부? → 현금/현물? → 급여 산정 방식
- 사례 검증: 프랑스 ASPA 등 애매한 사례로 규칙 적용 확인
- 체크리스트: 금지 조합 위반 여부, reasoning 충분성 검증

출처: 저자 작성.

## 나. 경계 사례를 통한 규칙 명확화

프롬프트 설계에서 가장 중요한 것은 분류가 애매한 경계 사례(edge case)를 어떻게 처리할지 명확히 하는 것이다. 프랑스 ASPA(Allocation de solidarité aux personnes âgées)는 이러한 경계 사례의 대표적 예이다. ASPA는 과거 연금제도에 가입했으나 현재 급여가 낮은 노인을 대상으로 보충 급여를 제공한다. 이 프로그램은 과거 기여 이력을 전제로 하지만, 현재 소득수준에 따라 급여 수급 여부가 결정된다.

초기 프롬프트에서는 ASPA를 사회수당(tax-funded)과 기여형(requires pension membership)의 조합으로 분류하는 오류가 발생하였다. 이는 금지 조합에 해당하므로, 프롬프트에 다음과 같은 지시를 추가하였다. “프랑스 ASPA와 같이 과거 연금 가입을 전제로 하되 현재 자산조사를 통과해야 하는 경우, social\_insurance + means\_tested로 분류하십시오. 이는 연금 체계의 일부로서 과거 기여 이력을 기반으로 하며, 현재 자산조사는 수급 조건일 뿐 제도유형을 변경하지 않습니다.”

이러한 구체적 사례를 프롬프트에 포함함으로써, LLM은 유사한 구조를 가진 다른 프로그램을 분류할 때 일관된 기준을 적용할 수 있게 되었다. 독일의 Grundsicherung im Alter, 영국의 Pension Credit, 일본의 老齡基礎年金 등 유사한 보충적 연금제도들이 모두 social\_insurance + means\_tested 조합으로 일관되게 분류되었다.

## 다. 분류 프로세스

[그림 4-5], [그림 4-6]는 한국 기초연금과 독일 법정연금보험을 예시로 실제 분류 프로세스를 보여준다.

[그림 4-5] 데이터 구조화 및 분류 예시(한국 - 기초연금)

**예시 1: 한국 기초연금 (Korea - Basic Pension)**

① 분류를 위한 텍스트 분석

→ 데이터 연계 과정에서 추출된 텍스트

"Senior citizens aged 65 and over in the lowest 70% income bracket receive the non-contributory, non-taxable monthly benefits."

"Basic old-age pension (social assistance, income tested): Age 65."

"Up to 209,960 won a month is paid."

제도유형: social\_allowance → "non-contributory, social assistance"

커버리지: means\_tested → "lowest 70%, income tested"

급여유형: flat\_rate → "monthly benefits, 209,960 won"

↓ 구조화

② 구조화된 출력 (JSON)

제도유형	social_allowance	포세 지원, 비기여 기반 무조
커버리지	means_tested	65세 이상 하위 70% 소득계층 대상
급여유형	flat_rate	매월 정액 (최대 209,960원)

■ 제도유형 (Scheme Type)
 ■ 커버리지 (Coverage Type)
 ■ 급여유형 (Benefit Type)

출처: 저자 작성.

[그림 4-6] 데이터 구조화 및 분류 예시(독일 - 법정연금보험)

**예시 2: 독일 법정연금보험 (Germany - Statutory Pension Insurance)**

① 분류를 위한 텍스트 분석

→ 데이터 연계 과정에서 추출된 텍스트

"To qualify for a standard pension, a person must be 65 years and 8 months of age and have contributed for at least five years."

"Old-age pension: 18.6% of insured's monthly earnings. The pension is based on earnings points."

"Type of program: Social insurance system."

제도유형: social\_insurance → "social insurance system, 18.6% of insured's monthly earnings"

커버리지: contributory → "have contributed for at least five years"

급여유형: earnings\_related → "based on earnings points"

↓ 구조화

② 구조화된 출력 (JSON)

제도유형	social_insurance	의무 보험료 기여 (18.6%)
커버리지	contributory	5년 이상 기여 필요
급여유형	earnings_related	소득함수 기반 소득비례

■ 제도유형 (Scheme Type)
 ■ 커버리지 (Coverage Type)
 ■ 급여유형 (Benefit Type)

출처: 저자 작성.

각 프로그램에 대해 ① 연계된 텍스트를 분석하여 핵심 키워드를 식별하고, ② 이를 바탕으로 3차원 분류를 수행하며, ③ 구조화된 JSON 형식으로 출력한다.

한국 기초연금의 경우, 입력 텍스트에서 “non-contributory”, “social assistance”, “lowest 70% income bracket”, “income tested”, “monthly benefits”, “209,960 won a month”라는 키워드가 추출되었다. “non-contributory”와 “social assistance”를 근거로 scheme\_type을 social\_allowance로 분류하였다. “lowest 70%”와 “income tested”를 근거로 coverage\_type을 means\_tested로 분류하였다. “monthly benefits”와 정액인 “209,960 won”을 근거로 benefit\_type을 flat\_rate로 분류하였다.

독일 법정연금보험의 경우, 입력 텍스트에서 “social insurance system”, “18.6% of insured’s monthly earnings”, “have contributed for at least five years”, “based on earnings points”라는 키워드가 추출되었다. “social insurance system”과 “18.6% of insured’s monthly earnings”를 근거로 scheme\_type을 social\_insurance로 분류하였다. “have contributed for at least five years”를 근거로 coverage\_type을 contributory로 분류하였다. “based on earnings points”를 근거로 benefit\_type을 earnings\_related로 분류하였다.

출력된 JSON 객체는 세 개의 분류 필드와 reasoning 필드로 구성된다. 한국 기초연금의 reasoning은 “조세 재원, 비기여 기반 부조 / 65세 이상 하위 70% 소득계층 대상 / 매월 정액(최대 209,960원)”으로 기술되었다. 독일 법정연금보험의 reasoning은 “의무 보험료 기여(18.6%) / 5년 이상 기여 필요 / 소득점수 기반 소득비례”로 기술되었다.

### 3. 데이터 통합의 전체 프로세스와 성과

#### 가. 정량-정성 통합 프로세스 요약

본 장에서는 SOCX 정량 데이터베이스와 Country Notes, ISSA 정성 정보를 구조화하여 통합하는 과정을 다루었다. 데이터 통합의 전체 흐름은 다음과 같이 요약된다. SOCX 데이터베이스에서 41개국 × 7개 정책 영역의 6,637개 Code4 프로그램을 추출한다(제1절). 각 프로그램에 대해 Country Notes와 ISSA의 전체 텍스트에서 관련 단락을 LLM으로 추출하여 프로그램별로 텍스트를 연계한다(제2절). 연계된 텍스트를 LLM으로 분석 및 구조화하여 제도 유형, 대상 유형, 급여 유형을 분류하였다(제3절). 최종 데이터는 Code4를 키로 하는 JSONB 형식으로 PostgreSQL 데이터베이스에 저장되며, 정량 정보, 연계된 텍스트, 구조화된 분류 결과를 모두 포함한다.

이러한 통합 프로세스의 핵심 기여는 LLM의 자연어 이해 능력을 활용하여 전통적 키워드 매칭 방법론의 한계를 극복한 것이다. Country Notes와 ISSA는 프로그램명이 다르고, 제도영역 구분이 다르며, 서술 방식이 다른 이질적 자료이다. 키워드 매칭으로는 이러한 변이를 포착하기 어려우나, LLM은 의미론적 동일성을 인식하여 “Basic Pension”과 “Basic old-age pension”을 동일한 프로그램으로 식별하고, “non-contributory”와 “social assistance”를 동일한 재원 방식으로 이해한다.

## 나. 통합 데이터 구조

4장 전체 프로세스를 통해 생성된 최종 데이터의 구조는 <표 4-3>과 같다. 최종 데이터는 PostgreSQL 데이터베이스 내에 Code4를 키(key)로 하는 지출 정보와 함께 JSONB 형식의 변수로 저장되며, 각 JSON 객체는 네 개의 주요 섹션을 포함한다.

metadata 섹션은 SOCX 데이터베이스에서 추출한 프로그램의 기본 정보를 포함한다. Code4, 프로그램명, 국가 코드와 국가명, 정책영역 코드와 정책영역명, 재원 유형(공공/민간) 등이 포함되며, 이는 제1절에서 확정한 공통 식별자 역할을 한다.

linked\_text 섹션은 제2절에서 추출한 Country Notes와 ISSA 텍스트를 포함한다. 원천별로 텍스트 추출 성공 여부(extracted), 추출된 텍스트 전체(text), 단어 수(length\_words)를 기록한다. 이 섹션은 제3절의 분류 작업뿐만 아니라, 이후 연구자가 원문을 직접 검토하거나 다른 분류 체계를 적용할 때도 활용될 수 있다.

classification 섹션은 제3절에서 수행한 3차원 분류 결과를 포함한다. 제도유형(scheme\_type), 대상유형(coverage\_type), 급여유형(benefit\_type) 각각의 분류 값과, 각 판단의 근거를 담은 reasoning 객체를 포함한다. reasoning 객체는 재원 방식(scheme), 수급 조건(coverage), 급여 산정(benefit)으로 구성되며, 각 필드에는 판단 근거와 함께 텍스트에서 추출한 핵심 키워드를 괄호 안에 명시한다. 분류 수행 날짜와 사용된 LLM 모델 버전도 기록하여 재현성을 확보한다.

validation 섹션은 분류 결과의 품질 검증 정보를 포함한다. 논리적 일관성 여부(logical\_consistency), 금지 조합 규칙 통과 여부(forbidden\_combination\_check), reasoning 충분성 평가

(reasoning\_sufficiency)를 기록한다. 이 정보는 이후 데이터 사용자가 분류 결과의 신뢰성을 판단하는 데 활용될 수 있다.

〈표 4-3〉 최종 통합 데이터의 구조(JSON 스키마)

통합 데이터의 구조(JSON 스키마)

```
{
  "metadata": {
    "code4": "410.10.1.1.1.6",
    "program_name": "Basic Pension",
    "country_code": 410,
    "country_name": "Korea",
    "policy_area_code": 1,
    "policy_area_name": "Old age",
    "source_type": 10
  },
  "linked_text": {
    "country_notes": {
      "extracted": true,
      "text": "The Basic Pension was introduced in 2008 to provide ...",
      "length_words": 1150
    },
    "issa": {
      "extracted": true,
      "text": "Basic old-age pension(social assistance, income tested): ...",
      "length_words": 780
    }
  },
  "classification": {
    "scheme_type": "social_allowance",
    "coverage_type": "means_tested",
    "benefit_type": "flat_rate",
    "reasoning": {
      "scheme": "비기여 기반, 조세 재원(non-contributory, social assistance)",
      "coverage": "65세 이상 하위 70% 소득계층 대상(lowest 70% income ...)",
      "benefit": "매월 정액 급여(monthly benefits, 최대 209,960원)"
    },
    "classified_date": "2024-11-15",
    "model_version": "gemma-3-27b-it"
  },
  "validation": {
    "logical_consistency": true,
    "forbidden_combination_check": "passed",
    "reasoning_sufficiency": "sufficient"
  }
}
```

출처: 저자 작성.

다. 주요 DB구축 결과

〈표 4-4〉는 본 장에서 구축한 최종 통합 데이터베이스의 구조를 요약한다.

〈표 4-4〉 최종 통합 데이터베이스 구조

구분	항목	결과
대상 범위	국가 수	41개국(OECD + 일부 비OECD)
	정책영역	보건, ALMP를 제외한 7개 영역
	Code4 프로그램 총계	6637
	노령(Branch:01)	1,195
	유족(Branch:02)	7,64
	근로무능력(Branch:03)	1,710
	가족(Branch:05)	1,363
	실업(Branch:07)	390
	주거(Branch:08)	278
	기타(Branch:09)	937
데이터 원천	Country-Notes 원천	각국 정책영역별 PDF
	ISSA 원천	각국 제도영역별 웹 HTML
	연계 방법	LLM(Google Gemma-3-27B-IT) 활용 프로그램(Code4) 단위 정리
데이터 구조화	대상 유형	universal / contributory / means_tested
	급여 유형	현금(flat_rate, earnings_related, mixed 등) / 현물(service, voucher 등)
최종 산출물	DB 형식	PostgreSQL (프로그램별 지출 정보 + JSONB)
	Code4별 JSONB 구조	metadata + linked_text + classification + validation

주: 41개국 × 7개 정책영역 × 공공 부문의 6,637개 Code4 프로그램에 대한 통합 데이터 구축.  
출처: 저자 작성.

제1절에서 확정한 Code4 기반 통합 전략에 따라, 제2절에서는 Country Notes와 ISSA 텍스트를 Code4 프로그램별로 연계하였고, 제

3절에서는 연계된 텍스트를 3차원으로 분류하였다. 최종 산출물은 PostgreSQL 데이터베이스에 프로그램별 지출 정보와 함께 JSONB 형식으로 저장되었다.

본 장에서 구축한 데이터베이스는 SOCX 정량 데이터에 Country Notes와 ISSA의 정성 정보를 Code4 기준으로 연계하고, 이를 제도유형·대상유형·급여유형의 3차원으로 구조화한 것이다. PostgreSQL 데이터베이스에는 각 프로그램의 지출 정보와 함께 JSONB 형식으로 네 개의 섹션(metadata, linked\_text, classification, validation)이 저장되며, Code4를 키로 하여 정량·정성 정보를 통합 조회할 수 있다. 예를 들어, “한국 국민연금(Code4: 410.10.1.1.1.1)의 2021년 지출은 GDP 1.02%이며, 제도유형은 social\_insurance, 대상유형은 contributory, 급여유형은 earnings\_related”와 같은 통합 정보를 단일 쿼리로 추출할 수 있다.

이어서 5장에서는 이렇게 구축된 DB의 활용 가능성과 분석 잠재력을 노령 영역의 파일럿 분석을 통해 구체적으로 검증해 본다. 노령 영역을 파일럿 대상으로 하여 분석 잠재력을 검증하는 이유는 다음과 같다. 먼저, 타 영역에 비해 제도의 다양성이 높다고 할 수 있다. 사회보험(연금), 사회수당(기초연금), 서비스(돌봄) 등 다양한 제도유형이 공존하여 본 연구의 3차원 분류 체계의 분석적 유용성을 검증하기에 적합한 영역이다. 둘째, 비교 가능한 대상 국가들이 많다. OECD SOCX에서 가장 데이터가 충실하게 구축된 영역으로 41개국 간 체계적 비교가 가능하다. 셋째, 정책 영역 중 노령 영역의 재정 부담 비중이 타 영역에 비해서 크다는 점이다. 이러한 이유로 본 연구의 DB 구축 방법론과 분석 체계의 유효성을 검증하기에 적합한 파일럿 대상을 노령 영역으로 선택하였고, 5장에서는 이를 심도있게 다루도록 한다.





## 제5장

### 노령 정책영역 파일럿 분석

제1절 제도 유형별 분석

제2절 대상 및 급여 유형별 특성

제3절 파일럿 분석의 의의



## 제5장 노령 정책영역 파일럿 분석

### 제1절 제도 유형별 분석

기존의 OECD SOCX 데이터는 총지출(GDP 대비 %)과 재원 주체(공공/민간)라는 2차원적 정보를 제공하여 국가 간 지출 규모 비교에 유용하게 활용되었다. 그러나 장기 재정 전망 및 정책 시뮬레이션을 위해서는 그 지출이 구체적으로 어떻게(How) 구성되어 있는가에 대한 제도적 특성 정보가 필수적이다. 예를 들어, 동일한 GDP 10%의 노령 지출이라도 그것이 기여금 기반의 사회보험인지, 조세 기반의 사회수당인지에 따라 재정 부담의 성격과 향후 추계 방식이 근본적으로 달라지기 때문이다.

#### 1. 기존 분석의 한계: 제도 유형 정보의 부재

본 연구는 이러한 한계를 보완하기 위해, 첫 번째 새로운 분류 축으로 제도 유형(Scheme Type: 사회보험 vs 사회수당/서비스)을 추가하였다. 이는 지출의 재정 경로를 구분하는 주된 기준이 된다. 사회보험은 기여금을 통해 재원을 조달하며 기여 이력에 기반한 급여를 제공하는 반면, 사회수당 및 서비스는 주로 조세를 통해 재원을 조달하며 자산조사나 보편적 자격 기준에 따라 급여를 제공한다. 이러한 차이는 재정 추계 시 기여율 인상, 조세 인상, 민간 시장 활성화 등 서로 다른 정책 수단과 직접 연결된다.

본 연구는 이러한 제도 유형 분류를 41개국 노령 프로그램에 적용하

고, 기존의 총량 및 재원 정보와 결합하여 각국의 지출을 4대 구성 요소로 분해하였다. 즉, 공공-보험(public\_insurance), 공공-수당/서비스(public\_non\_insurance), 민간-보험(private\_insurance), 민간-수당/서비스(private\_non\_insurance)라는 네 가지 경로로 지출을 세분화한 것이다. 이를 7개 복지체제별로 평균한 결과는 <표 5-1>과 같다.

<표 5-1> 노령 지출 구성 비교(2021년)

(단위: %)

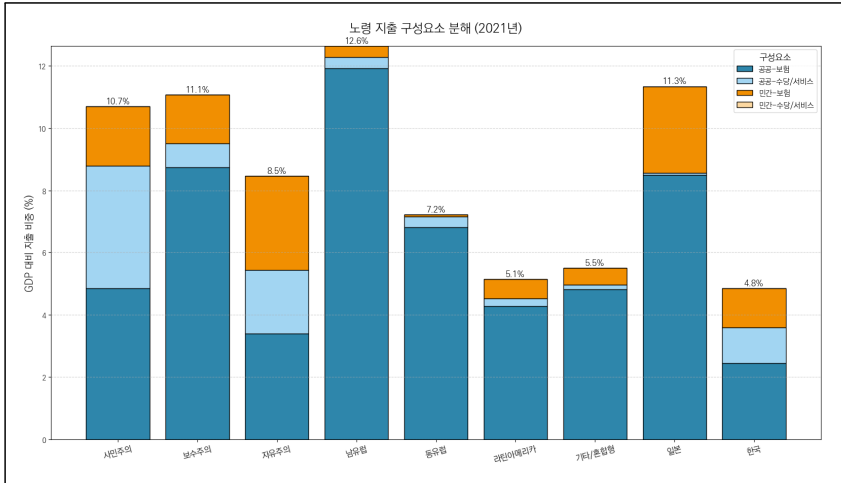
체제	총지출	공공보험	공공-수당 /서비스	민간보험	민간-수당 /서비스	보험비중	민간비중
(GDP 대비 %)				(구성비, %)			
자유주의	8.5	3.4	2.0	3.0	0.0	69.7	31.2
보수주의	11.1	8.7	0.8	1.6	0.0	93.3	14.0
사민주의	10.7	4.8	4.0	1.9	0.0	61.5	20.5
남부유럽	12.6	11.9	0.4	0.4	0.0	96.9	2.8
동부유럽	7.2	6.8	0.3	0.1	0.0	86.4	1.0
라틴아메리카	5.1	4.3	0.2	0.6	0.0	95.2	13.1
기타/혼합형	5.5	4.8	0.1	0.5	0.0	97.4	8.9
일본	11.3	8.5	0.1	2.8	0.0	99.4	24.4
한국	4.8	2.4	1.2	1.2	0.0	76.1	25.8

주: 기타/혼합형 혼합형 국가는 이스라엘과 터키임.  
출처: 저자 작성.

남유럽(96.9%)과 동아시아(97.4%)가 모두 매우 높아 총량만으로는 알 수 없는 제도적 유사성이 드러난다. 이처럼 제도 유형 정보는 기존 분석에서는 확인하지 못하던 국가 간 차이와 유사성을 동시에 보여준다.

[그림 5-1] 노령 지출 구성 비교(2021년)

(단위: %)



출처: 저자 작성

## 2. 제도 유형 분류가 밝히는 차이

### 가. 재정 경로의 차이: 보수주의 vs 사민주의

보수주의(11.1%)와 사민주의(10.7%)는 총지출 수준에서는 거의 같다. 그러나 insurance\_ratio는 보수주의(93.3%)가 사민주의(61.5%)보다 1.5배 이상 높다. 이는 <표 5-1>의 4대 구성요소에서 더욱 명확히 드러난다. 보수주의는 public\_insurance(8.7%)가 지출의 대부분을 차지하는 반면, 사민주의는 public\_insurance(4.8%)와 public\_non\_insurance(4.0%)가 거의 1대 1로 균형을 이룬다.

제도유형 정보가 없다면 이러한 구조적 차이를 재정 분석에 반영할 수 없다. 예를 들어 한국의 노령 소득보장 개편을 위해 타 국가 체제를 참조한 시나리오를 설정한다고 가정해 보자. 만약 보수주의형 구조(사회보험

중심)를 벤치마킹하는 시나리오를 구성하려면, 국민연금 중심의 확대 경로에서 기여율과 급여 산식이 어떻게 조정되어야 하는지를 추정해야 한다. 이 경우 재정 추계는 가입자 수, 평균 기여 소득, 기여율 변화를 중심으로 이뤄진다. 반면 사민주의형 구조(보험-수당 혼합)를 벤치마킹하는 시나리오를 구성하려면, 국민연금과 기초연금의 역할 배분, 각각의 재원 조달 방식(기여금 vs 조세), 대상 범위가 어떻게 달라져야 하는지를 검토해야 한다. 이 경우 재정 추계는 보험 부문과 수당 부문을 종합해야 한다. 그런데 이러한 시나리오 설정은 제도유형별 지출 구성 정보 없이는 불가능하다. 보수주의가 public\_insurance 중심(8.7%)이고 사민주의가 public\_insurance(4.8%)와 public\_non\_insurance(4.0%)의 균형 구조라는 정보가 있어야, 한국의 현재 구조(2.4% : 1.2%)를 어느 방향으로 조정할지 판단할 수 있기 때문이다.

#### 나. 재원 주체의 차이: 보수주의 vs 자유주의

보수주의(93.3%)와 자유주의(69.7%)는 모두 insurance\_ratio가 높아 보험 중심 시스템이라는 점에서 유사하다. 그러나 private\_ratio는 자유주의(31.2%)가 보수주의(14.0%)의 2배가 넘는다. <표 5-1>의 4대 구성요소를 보면, 자유주의는 private\_insurance(3.0%)가 보수주의(1.6%)보다 거의 2배 크며, 이는 민간보험이 노령 소득보장에서 상당한 역할을 함을 의미한다. 즉, 두 체제 모두 보험이라는 동일한 제도유형에 의존하지만, 그 재원 주체(공공 vs 민간)는 근본적으로 다르다.

이러한 차이는 위험 분담 주체와 재정 책임의 성격을 완전히 바꾼다. 보수주의형 구조에서는 공공보험이 중심이므로 국가가 장기 재정 책임을 지며, 재정 추계는 공공 재정의 지속가능성에 초점을 맞춘다. 반면 자유

주의형 구조에서는 민간보험 의존도가 높아 개인과 시장이 상당 부분 책임을 지며, 재정 추계는 공공 재정 부담뿐 아니라 민간 시장의 접근성, 수급 격차, 시장 안정성까지 고려해야 한다.

제도유형 정보만으로는 이러한 차이를 구분할 수 없다. 예를 들어 한국의 노령 소득보장 개편 시나리오를 설정할 때, "보험 중심 경로"를 선택한다고 해도 그것이 공공보험 확대인지 민간보험 활성화인지에 따라 정책 수단과 재정 분석 대상이 완전히 달라진다. 보수주의형을 벤치마킹한다면 국민연금의 보장성 강화와 사각지대 해소가 핵심 과제가 되지만, 자유주의형을 벤치마킹한다면 퇴직연금 활성화와 세제 지원이 핵심 과제가 된다. 이러한 시나리오 설정은 재원 주체별 지출 구성 정보가 있어야만 가능하다. 보수주의가 public\_insurance 중심(8.7%)이고 자유주의가 public\_insurance(3.4%)와 private\_insurance(3.0%)의 혼합 구조라는 정보가 있어야, 한국의 현재 구조(public 2.4%, private 1.2%)를 어떻게 조정할지 판단할 수 있기 때문이다.

이처럼 제도유형과 재원 주체라는 두 가지 분류 축은 국가 간 재정 경로와 위험 분담 구조의 차이를 명확히 보여준다고 할 수 있다. 그렇다면 한국의 현재 구조는 이러한 국제 비교 속에서 어떤 위치에 있으며, 어떤 특징을 갖는가? 이를 다음 항에서 살펴본다.

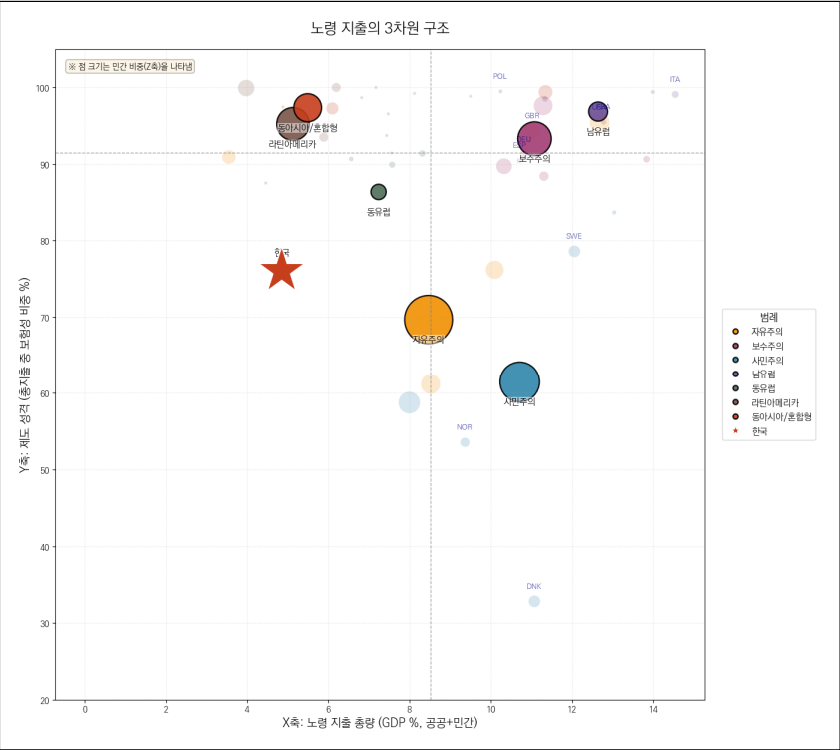
### 3. 한국의 노령 지출 구조: 복합적 제도유형

한국의 노령 지출 구조는 앞서 분석한 <표 5-1>에서 나타났듯이 매우 독특하다. 총지출(4.8%)은 41개국 중 하위권이지만, public\_insurance(2.4%), public\_non\_insurance(1.2%), private\_insurance(1.2%)가 2대 1대 1의 비율로 제도 유형이 혼재되어 있다. in-

surance\_ratio(76.1%)는 중간 수준이며, private\_ratio(25.8%)는 일본(24.4%)과 함께 상대적으로 높은 수준이라 할 수 있다.

[그림 5-2] 복지체제별 평균 노령 지출 4대 구성요소 분해

(단위: %)



출처: 저자 작성.

위 [그림 5-2]는 이러한 구조적 특징을 더욱 명확히 드러낸다. 보수주의와 남유럽은 public\_insurance가 압도적이고, 사민주의는 public\_insurance와 public\_non\_insurance가 균형을 이루며, 자유주의는 세 경로가 모두 상당한 비중을 차지한다. 반면 한국은 국민연금(2.4%), 기초연금 등(1.2%), 퇴직연금 등(1.2%)이라는 세 경로가 공존하

지만 각각의 성숙도가 낮아, 총량 자체도 낮게 나타난다.

이는 재정 분석에서 중요한 시사점을 준다. 한국은 OECD 최고 수준의 고령화 속도에 직면해 있다. 65세 이상 인구 비율이 2025년 25%, 2050년 40%에 육박할 것으로 전망되는 상황에서, 한국의 노령 지출은 크게 증가할 수밖에 없다. 그러나 한국의 노후소득보장은 단일 경로가 아닌 세 경로로 분산되어 있어, 증가하는 지출이 어느 경로에서 주로 발생할 것인지에 따라 재정 부담 주체(기여금, 조세, 민간부담)와 정책 수단이 완전히 달라진다. 더욱이 각 경로는 제도 성숙도가 낮고 발전 방향이 불확실하다. 국민연금만 아직 성숙 단계에 진입하지 않았으며, 제도 개혁 논의는 보장성 확대와 재정 안정화라는 상반된 방향으로 진행 중이다. 기초연금은 도입 이후 급여 수준이 정치적 국면에 따라 빈번히 조정되어 왔으며, 제도의 장기 방향성이 불명확하다. 퇴직연금은 의무화 이후에도 가입률과 적립 수준이 불안정하며, 최근 퇴직연금기금 도입 논의가 진행 중이나 제도 설계 방향은 아직 정해지지 않았다. 이러한 불확실성 속에서 각 경로의 향후 지출 규모와 재정 부담을 예측하려면, 세 경로를 종합적으로 분석하는 것이 필수적이다. 또한 세 경로 간 역할 분담과 조정도 필수적이다. 예를 들어 기초연금을 확대할 경우 국민연금의 재분배 기능과 중복되거나 상쇄될 가능성이 있으며, 퇴직연금을 활성화할 경우 공적 연금과의 정합성을 고려해야 한다.

결국 한국은 세 경로 중 어디에 집중할 것인지, 각 경로를 어떻게 조합할 것인지에 대한 정책 선택에 직면해 있다. 이러한 역할 분담과 조정 방향을 설정하는 데 있어, 이미 성숙한 노후소득보장체계를 운영하고 있는 타 국가들의 제도 구성은 중요한 참조점이 된다. 본 연구가 구축한 제도 유형별 지출 구성 정보는 이러한 국제 비교를 위한 기초 자료로서 의의를 갖는다고 하겠다. 다만 제도 유형(사회보험 vs 사회수당) 및 재원 주체(공

공 vs 민간)라는 분류만으로는 재정 분석에 충분하지 않다. 예를 들어 동일하게 ‘공공-수당’으로 분류되더라도, 그것이 전체 노인을 대상으로 하는 보편적 수당인지, 저소득층만을 대상으로 하는 선별적 수당인지에 따라 재정 소요 규모가 완전히 달라진다. 또한 동일하게 ‘공공-보험’으로 분류되더라도, 급여가 정액 방식인지 소득비례 방식인지에 따라 제도 내 재분배 효과와 재정 부담 구조가 달라진다. 이러한 세부 특성을 파악하기 위해서는 각 경로 내에서 누구에게(대상유형), 무엇을(급여유형) 제공하는지에 대한 추가 분류가 필요하며, 제2절에서는 41개국 노령 프로그램의 대상유형과 급여유형을 분석한다.

## 제2절 대상 및 급여 유형별 특성

본 절에서는 ① 대상유형(Coverage Type: 보편적 vs 선별적 vs 기여 기반), ② 급여유형(Benefit Type: 정액 vs 소득비례 vs 적립식)이라는 세부 특성이 재정 분석에서 필요한 이유를 41개국 사례를 통해 살펴본다.

### 1. 공공수당의 대상유형: 보편 vs 선별

제1절 <표 5-1>에서 사민주의는 public\_non\_insurance(공공-수당/서비스)가 4.0%로 상당한 비중을 차지하였다. 그러나 이 4.0%가 전체 노인을 대상으로 하는지, 저소득층 노인만을 대상으로 하는지에 따라 재정 분석의 전제와 정책 설계가 완전히 달라진다. <표 5-2>와 <표 5-3>은 41개국의 공공-수당/서비스 지출을 대상유형별로 분해한 결과와 사민주의 5개국의 개별 수치를 보여준다.

〈표 5-2〉 공공 수당/서비스형의 대상 유형별 구성비

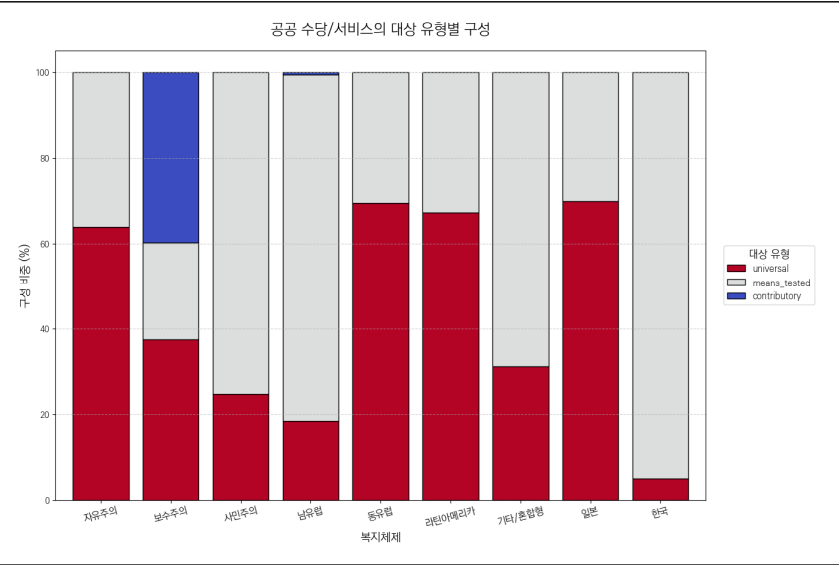
(단위: %)

체제	보편형	선별형	기여형
자유주의	63.9	36.1	0.0
보수주의	37.4	22.9	39.7
사민주의	56.7	43.3	0.0
남부유럽	18.4	81.0	0.6
동부유럽	69.4	30.6	0.0
라틴아메리카	67.2	32.8	0.0
기타/혼합형	31.1	68.9	0.0
일본	69.8	30.2	0.0
한국	5.0	95.0	0.0

주: 기타/혼합형 혼합형 국가는 이스라엘과 터키임. 구성비는 각 체제 공공-수당 지출 총합 내 비중.  
출처: 저자 작성.

〔그림 5-3〕 공공 수당/서비스형의 대상 유형별 구성(2021년)

(단위: %)



출처: 저자 작성.

〈표 5-3〉 사민주의 5개국 공공 수당/서비스형의 대상 유형별 구성비

(단위: %)

국가	보편형	선별형
핀란드	100.0	0.0
노르웨이	100.0	0.0
스웨덴	99.1	0.9
덴마크	23.7	76.3
아이슬란드	12.8	87.2

주: 사민주의 5개국의 공공 수당/서비스 지출 내 대상 유형별 구성비.  
출처: 저자 작성.

위의 〈표 5-3〉 결과는 동일한 사민주의 내부에도 서로 다른 두 가지 경로가 공존함을 보여준다. 핀란드, 노르웨이, 스웨덴은 거의 100% 보편적 수당이지만, 덴마크와 아이슬란드는 70% 이상이 선별적 수당이다. 대상 유형 정보가 있으면 이러한 체계적 차이를 파악할 수 있으며, 재정 분석 시 적절한 참조 모델을 선택할 수 있다.

예를 들어, 한국이 기초연금을 보편적 방향(전체 노인 대상)으로 개혁하여 사민주의 수준의 보장을 목표로 설정했다고 가정해 보자. 이때 급여 수준을 사민주의 평균으로 가정하고 재정 추계를 수행하면, 공공-수당 지출이 4.0%보다 훨씬 크게 나올 수 있다. 사민주의 평균(4.0%)에는 선별형(덴마크)이 포함되어 있어, 대상이 제한된 상태에서의 1인당 급여 수준이 반영되어 있기 때문이다. 이를 전체 노인에게 적용하면, 대상 규모 차이로 인해 총지출이 4.0%를 크게 초과하게 된다. 이러한 경우 대상유형 정보가 있다면, 보편형(스웨덴)과 선별형(덴마크)을 구분하여 보다 정교한 재정 분석을 수행할 수 있다.

## 2. 공공보험의 급여유형: 정액 vs 소득비례

제1절 <표 5-1>에서 보수주의와 일본은 public\_insurance(공공-보험)가 각각 8.7%, 8.5%로 비슷한 수준이다. 그러나 동일한 공공보험 지출이라도 급여가 정액 방식인지 소득비례 방식인지에 따라 제도 내 재분배 효과와 재정 부담 구조가 완전히 달라진다. <표 5-4>는 41개국의 공공-보험 지출을 급여유형별로 분해한 결과이다.

<표 5-4> 공공 사회보험의 급여 유형별 지출(체제 평균, GDP 대비 %)

(단위: %)

체제	정액	소득비례	적립식	혼합
자유주의	1.79	1.17	0.36	0.00
보수주의	0.30	1.15	0.00	3.22
사민주의	0.00	0.74	0.32	0.06
남부유럽	0.07	1.18	0.00	0.00
동부유럽	0.12	2.03	1.34	1.54
라틴아메리카	0.00	1.27	0.53	0.00
기타/혼합형	0.75	1.64	0.00	0.00
일본	4.12	0.88	0.00	3.36
한국	0.00	1.19	0.00	1.25

주: 체제 평균은 41개국 전체 평균(KOR, JPN 제외).

출처: 저자 작성.

위의 <표 5-4>는 제도 내부의 작동 원리를 명확히 보여준다. 일본은 flat\_rate(정액) 지출이 4.12%로, earnings\_related(소득비례) 0.88%를 압도한다. 이는 일본의 공적연금 1층(기초연금)이 기여 기반의 정액 급여로 작동하며, 제도 내 재분배 기능이 강함을 시사한다. 반면, 보수주의는 flat\_rate이 0.30%에 불과하며 earnings\_related(1.15%)와 mixed(3.22%)가 중심이다. 남유럽 역시 earnings\_related(1.18%)가 지배적이다. 이는 이들 체제가 기여금과 급여가 연동되는 소득비례 원칙

(비스마르크 방식)을 따르고 있음을 보여준다.

즉, 급여유형 정보를 통해 재정분석에서 요구되는 제도 내 재분배 효과를 보다 체계적으로 분석할 수 있다. 예를 들어, 한국이 공공보험을 확대하는 시나리오를 타 국가와 비교한다고 가정해 보자. 일본 방식(정액 기반)으로 참조한다면, 저소득층과 고소득층이 동일한 급여를 받으므로 재분배 효과가 강하다. 반면 보수주의 방식(소득비례 기반)으로 참조한다면, 기여 인센티브는 높으나 재분배 효과는 제한적이다. 이처럼 급여유형 정보가 있다면, 어느 방식이 한국의 현재 구조와 정책 목표에 더 적합한지 판단할 수 있다.

### 3. 한국 노령 지출의 재정 분석 함의

제1절과 제2절의 분석은 본 연구의 분류 체계 추가를 통해 재정 구조의 입체적인 특징을 분석했고, 보다 정교한 재정 분석이 가능할 수 있음을 보였다.

첫째, 다중 경로 식별을 통한 재정 추계의 정교화가 가능하다. 제1절 <표 5-1>의 4대 구성요소 분석은 한국의 노령 지출 4.8%가 public\_insurance(2.4%, 국민연금), public\_non\_insurance(1.2%, 기초연금 등), private\_insurance(1.2%, 퇴직연금 등)라는 서로 다른 세 개의 재정 경로로 구성되어 있음을 보여준다. 이 세 경로는 재원 조달 방식, 제도 성숙도, 정책 통제 변수가 모두 다르다. 국민연금의 미래 지출은 기여율, 가입 기간, 급여 산식 등 제도적 성숙에 따라 결정되며, 기초연금의 지출은 대상 범위(현행 70%)와 급여 수준에 대한 정책 결정에 따라 좌우되고, 퇴직연금의 지출은 시장 상황과 가입자의 선택에 영향을 받는다. 제1절 3항에서 확인했듯이 한국은 세 경로 모두 제도 성숙도가 낮고 발전 방향이 불확실한 상황

이다. 이러한 불확실성 속에서 장기 재정 추계는 세 경로를 분리하여, 각기 다른 변수와 시나리오를 적용해야 보다 현실적인 예측이 가능하다. 본 연구의 경로별 분류는 이러한 정교한 재정 분석을 위한 기초 자료를 제공한다.

둘째, 타 국가 참조를 통한 구체적인 정책 시나리오 구성이 가능하다. 공공수당 강화 시나리오의 경우, 제2절 <표 5-3>에서 확인했듯이 스웨덴 방식의 보편적 경로와 덴마크 방식의 선별적 경로는 동일한 지출 수준이라도 정책 대상과 1인당 급여 수준이 완전히 다르다. 한국의 기초연금에 현재 선별적 방식(<표 5-2> 95.0%)임을 감안할 때, 향후 보편 방식으로 전환할 경우의 추가 재정 소요를 추계하는 데 대상유형 정보는 핵심적인 비교 근거가 된다. 공공보험 강화 시나리오의 경우, <표 5-4>에서 보듯이 일본과 같은 정책형 강화는 제도 내 재분배 효과가 큰 반면, 보수주의 국가와 같은 소득비례형 강화는 기여 인센티브는 높으나 재분배 효과는 제한적이다. 한국의 공공보험이 소득비례 및 혼합형 중심임을 고려할 때, 급여유형 정보는 기초연금의 재분배 역할과 두 제도의 역할 조정을 설계하는 데 필요한 재정적 판단 근거를 제공한다.

## 제3절 파일럿 분석의 의의

### 1. 파일럿 분석의 의의

본 5장의 노령 영역 파일럿 분석은 기존의 총량 중심 사회지출 분석이 가진 한계를 극복하기 위한 새로운 다차원적 분석틀의 유효성을 검증했다는 점에서 의의를 찾을 수 있다.

첫째, 본 연구가 제시한 3개의 새로운 분류 축(제도유형, 대상유형, 급

여유형)은 기존 SOCX 데이터의 블랙박스였던 제도적 구성을 구체적으로 식별해냈다. 제도유형은 보험(기여금)과 수당(조세)이라는 재정 경로의 차이를, 대상유형은 보편과 선별이라는 정책 철학의 차이를, 급여유형은 정액과 소득비례라는 재분배 구조의 차이를 명확히 구분하였다. 이러한 분류는 "제도의 어떤 특성이 재정에 영향을 미치는가"라는 질문에 보다 명확한 답을 제시할 수 있는 기반이 될 수 있다.

둘째, 이 분석틀은 기존의 복지체제론을 데이터 기반으로 확인하고 심화하였다. 제1절에서 보수주의(공공보험 중심 8.7%)와 사민주의(보험-수당 균형 4.8% : 4.0%)의 상이한 재정 구조를 확인했으며, 제2절에서는 사민주의 내부에도 보편형(스웨덴, 노르웨이)과 선별형(덴마크, 아이슬란드)이 공존함을 보였다. 이는 기존 총량 지출 집계로는 파악할 수 없었던 국가 간 차이를 드러낸 것이라 하겠다.

셋째, 이러한 분류가 단순한 현상 기술을 넘어 정교한 재정 분석의 필수 전제조건임을 확인하였다. 한국의 노령 지출을 세 개의 재정 경로(국민연금 2.4%, 기초연금 1.2%, 퇴직연금 1.2%)로 식별하고, 각 경로의 정책 시나리오를 타 국가와 비교할 수 있었던 것은 이러한 분석 체계를 구성하였기에 가능했다. 특히 한국은 세 경로 모두 제도 성숙도가 낮고 발전 방향이 불확실하여, OECD 최고 수준의 고령화 속도를 고려할 때 재정 추계의 불확실성이 매우 크다. 본 연구의 분류 체계는 이러한 한국의 재정 구조를 보다 명확히 진단하고, 타 국가와의 체계적 비교를 통해 정책 방향을 설정하는 데 기여할 수 있다.

## 2. 향후 확장 방안: 전 영역 포트폴리오 분석의 필요성

본 5장의 파일럿 분석은 방법론의 유효성을 노령이라는 단일 영역에

한정하여 검증하였다. 그러나 한 국가의 사회보장 시스템은 9개 정책 영역(노령, 유족, 장애, 보건, 가족, 적극적노동시장, 실업, 주거, 기타)이 상호작용하는 유기적인 포트폴리오이다.

본 분석에서 확인된 한국의 노령 영역 특징(4.8%의 낮은 지출과 세 경로 공존)은 그 자체로도 중요하지만, 보건 영역의 고지출-공공보험 구조나 가족 영역의 저지출-공공수당 구조와 함께 볼 때 비로소 완전한 정책적 함의를 갖는다. 한정된 재원을 두고 각 영역이 경합하는 과정에서, 특정 영역의 제도 설계는 필연적으로 타 영역의 자원 배분과 제도적 정합성 문제에 영향을 미치기 때문이다. 예를 들어, 한국의 노령 영역에서 보편적 공공수당을 강화하는 정책은, 이미 보편적 공공보험으로 운영되는 보건 영역과의 재정 부담(조세 vs 보험료) 배분 문제를 야기할 수 있다.

또한 본 연구는 2021년 단년도 데이터를 기반으로 하여, 제도 변화의 동태적 과정을 포착하지 못한다는 한계가 있다. 한국의 국민연금과 기초연금은 도입 이후 지속적으로 제도가 변화해 왔으며, 향후에도 정책 논의가 활발하다. 이러한 변화를 추적하려면 시계열 데이터가 필요하다.

따라서 2~3차년도 연구에서는 본 5장에서 유효성을 검증한 4대 구성요소 및 제도 특성 분석틀을 9개 전 영역으로 확장하는 것이 필수적이다. 이를 통해 41개국의 국가 전체 사회보장 포트폴리오 맵을 완성하고, 영역 간 재정 배분의 우선순위와 제도적 정합성을 종합적으로 분석함으로써, 한국 사회보장제도의 지속가능한 재정 운용 방안을 모색하는 것이 본 연구의 궁극적인 목표이다. 또한 2015년, 2019년, 2021년 등 주요 시점의 데이터를 추가하여 제도 변화의 동태적 과정을 분석함으로써, 각국의 복지 개혁 방향과 한국의 정책 궤적을 보다 입체적으로 이해할 수 있다. 본 연구의 노령 파일럿 분석은 이러한 확장 연구를 위한 방법론적 토대와 실증적 근거를 제공한다고 하겠다.





## 제6장

### 결론

제1절 연구 결과 요약

제2절 주요 시사점

제3절 연구의 한계 및 향후 과제



## 제6장 결론

### 제1절 연구 결과 요약

본 연구는 “사회보장 장기재정분석 I: 사회보장 재정지표 탐색 및 DB 구축”이라는 주제로 3개년 연구의 1차년도 성과를 정리한 것이다. 3개년 연구 로드맵에서 1차년도는 공급측 재정지표 탐색과 데이터베이스 구축을 목표로 하였으며, 2차년도는 9개 정책영역 전체로 확대하고 수요측 지표를 추가하며, 3차년도는 통합 시뮬레이션 모형 구축을 계획하고 있다. 1차년도 연구는 이러한 3년 로드맵의 기초를 확립하는 단계로서, OECD SOCX 기반 재정지표 체계를 탐색하고 LLM을 활용한 정량·정성 통합 데이터베이스 구축 방법론을 개발하였다.

#### 1. 재정지표 체계 탐색

2장에서는 사회보장 재정분석을 위한 이론적 기반을 검토하였다. 국내 재정분석 현황(기획재정부, NABO, 사회보장위원회)을 살펴본 결과, 각 기관이 서로 다른 방법론과 가정으로 재정추계를 수행하여 결과 간 정합성이 제한적이며, 개별 제도별 분석이 분절적으로 이루어져 사회보장 전체의 재정 부담을 통합적으로 파악하기 어렵다는 한계를 확인하였다. 이러한 한계를 개선하기 위해서는 국제 비교 가능한 표준화된 재정지표 체계가 필요하며, OECD SOCX가 이러한 체계의 기준점을 제공함을 제시하였다.

SOCX 분류체계는 9개 정책영역(노령, 유족, 근로무능력, 보건, 가족, 적극적노동시장정책, 실업, 주거, 기타), 3개 제공주체(공공, 의무적 민간, 임의적 민간), 2개 급여형태(현금, 현물)로 구성되어 있으며, 이는 정량 재정지표의 기본 분류틀을 제공한다. 또한 복지체제론(Esping-Andersen의 3개 체제 및 확장된 5개 체제)은 국제 비교의 준거틀로서, 복지체제별 재정지표의 특성을 이해하고 한국의 위치를 가늠하는 좌표를 제공한다.

3장에서는 OECD SOCX 데이터베이스의 구조와 특성을 심층 분석하였다. SOCX는 OECD 38개국의 1980-2023년 사회지출 데이터를 프로그램 수준까지 세분화하여 제공하며, 자국 통화, PPP 달러, GDP 대비 비율, 정부지출 대비 비율 등 다양한 측정 단위로 국제 비교와 시계열 분석을 지원한다. 한국 데이터는 1990년부터 제공되며, 2018년 이후 결산 기준 전환으로 정확성이 크게 향상되었다. Country Notes는 SOCX 정량 데이터의 해석에 필수적인 정성적 맥락을 제공하며, 각국의 개별 프로그램별 제도 특성(급여 산정 방식, 수급 요건, 재원 조달 방식 등)을 서술형 텍스트로 기술하고 있다.

## 2. 정량·정성 통합 DB 구축

4장에서는 SOCX 정량 데이터, Country Notes 정성 정보, ISSA 제도 특성 정보를 통합한 데이터베이스를 구축하였다. 본 연구의 핵심은 41개국 × 7개 정책영역(보건, ALMP 제외) = 287개 국가-정책영역 조합 전체에 대해 LLM(Google Gemma-3-27B-IT) 기반 자연어 처리를 통해 정성 정보를 구조화한 점이다.

데이터 연계 단계에서는 6,637개 Code4 프로그램에 대해 Country

Notes와 ISSA 텍스트를 추출하였다. Country Notes는 각국 정책영역별 PDF에서, ISSA는 각국 제도영역별 웹 HTML에서 프로그램별로 관련 텍스트를 연계하였다. 5단계 프로세스(입력 준비 → Country Notes 연계 → ISSA 연계 → 정보 통합 → 검증)를 통해 체계적으로 수행되었다.

데이터 구조화 단계에서는 연계된 텍스트를 3차원으로 분류하였다. 제도유형(scheme\_type: social\_insurance / social\_allowance / social\_service), 대상유형(coverage\_type: universal / contributory / means\_tested), 급여유형(benefit\_type: flat\_rate, earnings\_related, mixed 등 현금 유형 및 service, voucher 등 현물 유형)으로 분류하였다. 3요소 프롬프트(필수 요구사항 + 분류 기준 + 추가 지시사항)를 설계하고, 금지 조합 규칙(allowance+contributory, service+earnings\_related)을 적용하여 논리적 일관성을 확보하였다. 최종 산출물은 PostgreSQL 데이터베이스에 저장되었다. 각 프로그램별 지출 정보와 함께 JSONB 형식으로 4개 섹션(metadata + linked\_text + classification + validation)이 통합 저장되어, Code4를 키로 정량·정성 정보를 통합 조회할 수 있다. 정책영역별 구성을 보면, 노령 1,195개, 유족 764개, 근로무능력 1,710개, 가족 1,363개, 실업 390개, 주거 278개, 기타 937개로 총 6,637개 프로그램이 구축되었다. 이는 41개국 × 7개 정책영역 × 공공 부문에 대한 체계적 데이터 통합의 성과이다.

### 3. 노령 정책영역 파일럿 분석

5장에서는 구축된 재정분석 체계의 실제 적용 가능성을 노령 영역 파일럿 분석을 통해 검증하였다. 분석 대상은 41개국 노령 프로그램이며, 2021년 시점 데이터를 기준으로 하였다.

제도유형 기반 분석에서는 기존 SOCX의 2차원 정보(지출 규모 + 재원 주체)에 제도유형을 추가하여 4대 구성요소(공공-보험, 공공-수당/서비스, 민간-보험, 민간-수당/서비스)로 분해하였다. 복지체제별 분석 결과, 보수주의는 공공-보험이 8.7%로 압도적이며 보험비중 93.3%, 사회민주주의는 공공-보험 4.8%와 공공-수당 4.0%가 균형을 이루며 보험비중 61.5%, 자유주의는 공공-보험 3.4%와 민간-보험 3.0%가 혼합되며 민간 비중 31.2%로 나타났다. 한국은 총지출 4.8%(공공-보험 2.4%, 공공-수당 1.2%, 민간-보험 1.2%)로 41개국 중 최하위이며, 보험비중 76.1%, 민간비중 25.8%로 세 경로가 공존하지만 각각의 성숙도가 낮은 '이중 저발전' 상태임을 확인하였다.

제도 특성별 비교에서는 대상유형과 급여유형을 추가 분석하였다. 공공-수당의 대상유형 분석 결과, 사회민주주의 내에서도 스웨덴·노르웨이는 보편형 100%인 반면 덴마크·아이슬란드는 선별형 70% 이상으로 상이한 경로가 공존함을 확인하였다. 한국은 선별형 95.0%로 기초연금이 선별적 사회수당임을 보여준다. 공공-보험의 급여유형 분석 결과, 일본은 정액형(flat\_rate) 4.12%가 소득비례형 0.88%를 압도하는 반면, 보수주의는 소득비례형(1.15%)과 혼합형(3.22%)이 중심이다. 한국은 소득비례형 1.19%와 혼합형 1.25%로 구성되어 있다.

한국 노령 지출의 재정 분석 함의로는 첫째, 다중 경로 식별을 통한 재정 추계의 정교화가 가능하다. 한국의 세 경로(국민연금 2.4%, 기초연금 1.2%, 퇴직연금 1.2%)는 재원 조달 방식, 제도 성숙도, 정책 통제 변수가 모두 다르므로 각 경로별 시나리오가 필요하다. 둘째, 타 국가 참조를 통한 구체적인 정책 시나리오 구성이 가능하다. 기초연금 보편화 시나리오는 스웨덴형(보편)과 덴마크형(선별)을 비교 가능하며, 공공보험 강화 시나리오는 일본형(정액 기반)과 보수주의형(소득비례 기반)을 비교 가능하다.

파일럿 분석의 의의는 첫째, 3개의 새로운 분류 축(제도유형, 대상유형, 급여유형)이 기존 SOCX 데이터의 블랙박스였던 제도적 구성을 구체적으로 식별해냈다. 둘째, 복지체제론을 데이터 기반으로 확인하고 심화하였다. 셋째, 한국의 노령 지출을 세 개의 재정 경로로 식별하고 각 경로의 정책 시나리오를 타 국가와 비교할 수 있었다.

## 제2절 주요 시사점

5장의 노령 정책영역 파일럿 분석은 정량·정성 통합 DB 구축 방법론의 타당성을 검증하기 위한 것이었다. 파일럿 분석에서 설정한 3가지 검증 목적의 달성 여부를 평가하고, 이를 토대로 2~3차년도 연구 확대 방향과 시사점을 도출한다.

### 1. 정량·정성 통합 재정지표의 유용성

첫 번째 검증의 목적은 정량 데이터(지출 규모)와 정성 데이터(제도 특성)가 효과적으로 결합되어 의미 있는 재정지표를 산출하는지 확인하는 것이었다. 5장 분석 결과, 정량 지표만으로는 한국이 41개국 중 최하위 지출 수준(GDP 4.8%)이라는 사실을 파악할 수 있으나, 왜 낮은지, 어떤 제도 구조적 특성 때문인지는 알 수 없었다.

정성 정보를 결합한 결과, 한국의 낮은 지출 수준이 (1) 국민연금의 낮은 제도 성숙도(약 35%), (2) 기초연금의 선별적 구조(자산조사형 95.0%), (3) 민간연금의 미발달에 기인함을 구체적으로 진단할 수 있었다. LLM 기반 제도 특성 분류(scheme\_type, coverage\_type, bene-

fit\_type)는 6,637개 프로그램 중 노령 1,195개 프로그램을 체계적으로 비교 가능하게 하였으며, 이를 통해 한국 국민연금의 독일·일본과 같은 기여형 소득비례 사회보험 경로에 속하나 지출 규모는 훨씬 낮음을 파악하였다. 또한 한국 기초연금이 선별적 사회수당(선별형 95.0%)으로 분류되어 스웨덴·노르웨이의 보편적 기초연금(보편형 100%)과는 다른 경로에 있음을 파악하였다.

이러한 정량·정성 통합 분석을 통해 단순 지출 규모 비교를 넘어, 제도 설계 원리와 재정적 함의를 동시에 이해할 수 있었다. 이러한 분석 결과는 한국 사회보장제도가 어떤 방향성을 선택할 것인지 파악하기 위한 기초 정보를 제공하며, 정량·정성 통합 재정지표 접근이 현 위치와 특성을 입체적으로 진단하는 데 유용함을 시사한다.

## 2. 한국 위치 진단 및 정책 방향 탐색 가능성

두 번째 검증의 목적은 통합 재정지표가 국제 비교를 통한 한국의 위치 진단에 유용한지 확인하는 것이었다. 5장 분석 결과, 한국은 지출 규모 측면에서 41개국 중 최하위이면서 동시에 민간 재원 비중도 낮아 '이중 저발전' 상태에 있음을 진단할 수 있었다.

복지체제별 비교를 통해 한국이 동아시아 복지체제에 속하나 같은 체제 내 일본(GDP 11.3%)보다 6.5%p 낮은 수준임을 파악하였으며, 향후 경로 선택지로 (1) 공공연금 확대(보수주의·사민주의형), (2) 민간연금 활성화(자유주의형), (3) 다층보장체계 정비(혼합형) 등을 고려할 수 있다.

4대 구성요소 분석을 통해 한국의 현재 구조(공공-보험 2.4%, 공공-수당 1.2%, 민간-보험 1.2%)를 명확히 파악하고, 보수주의형(공공-보험 중심 8.7%), 사민주의형(공공-보험 4.8% + 공공-수당 4.0% 균형), 자유주

의형(공공-보험 3.4% + 민간-보험 3.0% 혼합)과 비교하여 정책 방향을 탐색할 수 있었다.

이러한 분석 결과는 본 연구의 정량·정성 통합 재정지표 접근이 한국의 현 위치를 진단하고 정책 방향을 종합적으로 탐색하는 데 유용한 증거들을 제공할 수 있음을 시사한다.

### 3. 장기재정분석 활용 적합성

세 번째 검증의 목적은 통합 DB와 분석 체계가 장기재정분석에 활용 가능한지 평가하는 것이었다. 5장 분석 결과, 제도유형·대상유형·급여유형 분류를 통해 한국과 유사한 제도 특성을 가진 국가군을 식별하고, 이들의 재정 경로를 참조하여 한국의 향후 지출 전망을 정교화할 수 있는 가능성을 파악하였다.

제도 개혁 시나리오 분석 가능성도 파악되었다. 기초연금 보편화 시나리오는 스웨덴·노르웨이의 보편적 기초연금(보편형 100%)과 비교하여 재정적 영향을 사전 평가할 수 있으며, 국민연금 기여율 인상 시나리오는 독일·일본의 소득비례형 사회보험과 비교하여 재정 지속가능성을 검토할 수 있다.

다중 경로 식별을 통한 재정 추계의 정교화가 가능하다. 한국의 노령 지출 4.8%가 세 개의 재정 경로(국민연금 2.4%, 기초연금 1.2%, 퇴직연금 1.2%)로 구성되어 있음을 확인하였으며, 각 경로는 재원 조달 방식, 제도 성숙도, 정책 통제 변수가 모두 다르므로 장기 재정 추계는 세 경로를 분리하여 각기 다른 변수와 시나리오를 적용해야 보다 현실적인 예측이 가능함을 제시하였다.

이러한 분석 결과는 본 연구에서 구축한 정량·정성 통합 DB와 분석 체

제가 장기재정분석의 기초자료 및 분석틀로서 활용 가능성을 보여주었다.

### 제3절 연구의 한계 및 향후 과제

본 연구는 사회보장 장기재정분석을 위한 재정지표 체계를 탐색하고 정량·정성 통합 데이터베이스를 구축하였으며, 파일럿 검증을 통해 연구 방향에 대한 긍정적 결과를 얻었다. 다만, 1차년도 연구의 범위와 방법론상 한계가 존재한다. 이러한 한계를 인식하고 2-3차년도 연구에서 보완·확장할 과제를 제시한다.

#### 1. 1차년도 연구의 한계

첫째, 1차년도 파일럿 분석은 노령 정책영역 1개에 국한되어 사회보장 전체의 재정지표를 제시하지 못하였다. 노령 영역은 한국의 고령화와 연금제도 발전 과제를 고려하여 우선 분석 대상으로 선정하였으나, 보건, 가족, 실업 등 다른 정책영역의 재정 특성과 한국의 위치는 파악하지 못하였다. 사회보장 전체의 재정 부담을 통합적으로 진단하기 위해서는 9개 정책영역 전체에 대한 분석이 필요하다.

둘째, 공급측 재정지표(지출 규모, 제도 특성)만 다루었으며, 수요측 지표는 포함하지 못하였다. 인구구조 변화, 사회적 위험 분포, 제도 커버리지, 급여 수급률 등 수요측 요인이 재정 부담에 미치는 영향을 분석하지 못하여, 지출 증가의 원인을 공급측 제도 특성과 수요측 인구·사회 변화로 분해하는 데 한계가 있었다. 장기재정분석에서는 인구구조 변화에 따

른 수요 증가와 제도 확대에 따른 공급 증가를 구분하여 추계하는 것이 필수적이다.

셋째, LLM 기반 제도 특성 분류의 정확도 검증이 제한적이었다. 5장 파일럿 분석에서 6,637개 프로그램 중 노령 1,195개 프로그램의 LLM 분류 결과를 전수 검증하지 못하고, 주요 프로그램 일부만 샘플 검증하였다. LLM 분류는 Country Notes와 ISSA의 서술형 텍스트를 자동 해석하여 scheme\_type, coverage\_type, benefit\_type을 추론하는 과정으로, 해석의 모호성과 분류 오류 가능성이 존재한다. 2차년도 연구에서는 제도 전문가의 교차 검증을 통해 LLM 분류 정확도를 체계적으로 평가하고, 프롬프트 개선을 통해 정확도를 향상시킬 필요가 있다.

넷째, 한국 데이터의 시계열 분석이 제한적이었다. 5장 분석은 2021년 단년도 데이터를 중심으로 하였으며, 1990~2021년 시계열 추이 분석은 일부만 수행하였다. 한국의 빠른 제도 확대 속도와 2018년 결산 기준 전환에 따른 불연속 등을 시계열 분석을 통해 심층적으로 파악하지 못하였다. 장기재정분석에서는 과거 추세 분석이 미래 전망의 기초가 되므로, 시계열 분석 강화가 필요하다.

## 2. 향후 연구 과제

향후 2~3차년도 연구는 1차년도에서 구축한 정량·정성 통합 DB와 분석 체계를 기반으로, 9개 정책영역 전체로의 확대와 장기재정 시뮬레이션 모형 개발을 목표로 한다.

2차년도에서는 노령 외 8개 정책영역에 대한 복지체제별 비교분석을 수행하여 한국 복지국가의 전체 상을 도출하고, 인구구조·사회적 위험·제도 커버리지 등 수요측 지표를 추가하여 공급-수요 통합 분석 체계를

구축한다. 또한 LLM 분류 정확도를 제도 전문가 교차 검증을 통해 체계적으로 평가하고, 1990~2023년 시계열 데이터를 활용한 요인분해 분석을 통해 지출 증가 패턴을 심층 분석한다.

3차년도에서는 제도 특성 기반 유사 국가 참조 방법론을 구체적인 추세 모형으로 구현하고, 제도 개혁 시나리오(기초연금 보편화, 국민연금 기여율 인상, 건강보험 보장성 강화 등)의 재정적 영향을 사전 평가하는 도구를 개발한다. 인구구조 변화, 거시경제 변수, 제도 파라미터를 입력하면 영역별·제도별 지출 전망을 산출하는 Python 기반 통합 시뮬레이션 모형을 구축하며, 2050년 한국의 사회지출 전망을 OECD 주요국과 비교하여 재정 지속가능성을 평가하는 벤치마킹 체계를 완성한다. 최종적으로 정책 브리프, 웹 대시보드 등 정책 실무 활용 도구를 제공하여 기획재정부, NABO, 보건복지부, 사회보장위원회 등에서 장기재정분석 및 제도 개혁 논의에 실질적으로 활용될 수 있도록 한다.

### 3. 맺음말

본 연구는 사회보장 장기재정분석을 위한 새로운 접근법을 제시하였다. 기존 재정추계가 개별 제도별로 분절적으로 수행되어 사회보장 전체의 재정 부담을 통합적으로 파악하기 어렵다는 한계를 인식하고, OECD SOCX 기반 국제 비교 가능한 재정지표 체계를 탐색하였다. 또한 LLM을 활용하여 Country Notes와 ISSA의 정성 정보를 구조화하고, 정량 데이터와 통합하여 제도 특성과 재정 규모를 동시에 비교 가능한 데이터베이스를 구축하였다.

구축된 데이터베이스는 41개국 × 7개 정책영역 × 공공 부문의 6,637개 Code4 프로그램을 포괄하며, 각 프로그램은 제도유형(scheme\_type),

대상유형(coverage\_type), 급여유형(benefit\_type)의 3차원으로 분류되어 PostgreSQL 데이터베이스에 JSONB 형식으로 저장되었다. 이는 정량·정성 정보를 통합 조회 가능한 체계적 인프라이다.

1차년도 노령 정책영역 파일럿 분석을 통해 정량·정성 통합 재정지표가 한국의 현 위치를 진단하고 정책 방향을 탐색하는 데 유용함을 확인하였다. 또한 본 연구의 데이터를 통해 한국의 노령 제도의 상태(‘이중 저발전’ 상태(총지출 4.8%, 공공-보험 2.4%, 공공-수당 1.2%, 민간-보험 1.2%))와 같이 진단할 수 있음을 보였으며, 나아가 개혁에 필요한 벤치마크 국가들의 현황도 구체적으로 파악할 수 있음을 보였다.

2~3차년도 연구에서는 9개 정책영역 전체로 확대하고, 수요측 지표를 추가하며, 통합 시뮬레이션 모형을 개발하여 1차년도의 한계를 보완할 것이다. 3개년 연구가 완료되면, 한국 사회보장제도의 전체 재정 부담을 국제 비교 가능한 지표로 진단하고, 제도 개혁 시나리오의 재정적 영향을 사전에 평가하는 체계적 도구를 정책 실무에 제공할 수 있을 것으로 기대한다.





- 기획재정부. (2020). 2020~2060년 장기재정전망. 기획재정부.
- 국민연금재정추계전문위원회. (2023). 2023 국민연금 재정계산: 국민연금 장기 재정추계. 국민연금재정추계위원회.
- 국회예산정책처. (2020). 2020 NABO 장기 재정전망. 서울: 국회예산정책처.
- 보건복지부. (2024a). 제2차 국민건강보험 종합계획(안). 세종: 보건복지부.
- 보건복지부. (2024b). 제5차 사회보장 재정추계(안). 세종: 보건복지부.
- 신정우, 안영, 하솔잎, 이아영, 강은나, 황주희, 변수정, 이기호, 박소은, 고제이, 최경덕, 천미경, 황안나, 고경환, 정소라, 박희진, 심현보, 김도희. (2023). 한국의 SOCX 데이터베이스 작성 기준: 2023년. 보건복지부; 한국보건사회연구원.
- 임슬기. (2023). 2023~2032년 건강보험 재정전망. 서울: 국회예산정책처.
- 통계청. (2023). 장래인구추계: 2022~2072년. 통계청.
- 한신실. (2020). 한국은 어떤 복지국가로 성장해왔는가? 한국사회정책, 27(1), 153-185.
- Adema, W., & Whiteford, P. (2010). Public and private social welfare. In F. G. Castles, S. Leibfried, J. Lewis, H. Obinger, & C. Pierson (Eds.), *The Oxford handbook of the welfare state* (pp. 121-138). Oxford University Press.
- Adema, W., & Fron, P. (2019). *The OECD SOCX Manual-2019 Edition. A Guide to the OECD Social Expenditure Database*. Paris: Organisation for Economic Co-operation and Development.
- Akerlof, G. A. (1970). The market for "lemons": Quality uncertainty and the market mechanism. *The Quarterly Journal of Economics*, 84(3), 488-500.
- Anttonen, A., & Sipilä, J. (1996). European social care services: Is it possible to identify models? *Journal of European Social Policy*,

- 6(2), 87-100. <https://doi.org/10.1177/095892879600600201>
- Arrow, K. J. (1963). Uncertainty and the welfare economics of medical care. *The American Economic Review*, 53(5), 941-973.
- Arts, W., & Gelissen, J. (2002). Three worlds of welfare capitalism or more? A state-of-the-art report. *Journal of European Social Policy*, 12(2), 137-158.
- Baldwin, P. (1990). *The politics of social solidarity: Class bases in the European welfare state 1875-1975*. Cambridge: Cambridge University Press.
- Barr, N. (2001). *The Welfare State as Piggy Bank: Information, risk, uncertainty, and the role of the state*. Oxford University Press.
- Baumol, W. J. (1967). Macroeconomics of unbalanced growth: The anatomy of urban crisis. *The American Economic Review*, 57(3), 415-426.
- Beveridge, W. (1942). *Social insurance and allied services*. London: His Majesty's Stationery Office.
- Bonoli, G. (1997). Classifying welfare states: A two-dimension approach. *Journal of Social Policy*, 26(3), 351-372.
- Brady, D. (2009). *Rich democracies, poor people: How politics explain poverty*. Oxford University Press.
- Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. *Advances in neural information processing systems*, 33, 1877-1901.
- Castles, F. G. (2004). *The Future of the Welfare State: Crisis Myths and Crisis Realities*. Oxford University Press.
- Clasen, J., & Siegel, N. A. (Eds.). (2007). *Investigating welfare state change: the 'dependent variable problem' in comparative*

- analysis. Edward Elgar Publishing.
- Commonwealth of Australia. (2023). Intergenerational Report 2023: Australia's future to 2063. The Treasury. <https://treasury.gov.au/publication/2023-intergenerational-report>
- Diamond, P. A. (1977). A framework for social security analysis. *Journal of Public Economics*, 8(3), 275-298.
- Esping-Andersen, G. (1990). *The Three Worlds of Welfare Capitalism*. Princeton: Princeton University Press.
- Esping-Andersen, G. (1999). *Social Foundations of Postindustrial Economies*. Oxford: Oxford University Press.
- European Commission (2024). 2024 Ageing Report : Economic & Budgetary Projections for the EU Member States(2022-2070). Luxembourg: Publications Office of the European Union.
- European Commission. (n.d.). Labour market policies (LMP) database. European Commission. Retrieved October 16, 2025, from [https://employment-social-affairs.ec.europa.eu/policies-and-activities/analysis-evaluation-impact-assessment-and-databases/databases-and-indicators/labour-market-policies-lmp-database\\_en](https://employment-social-affairs.ec.europa.eu/policies-and-activities/analysis-evaluation-impact-assessment-and-databases/databases-and-indicators/labour-market-policies-lmp-database_en)
- European Commission. (n.d.). Social protection (spr) (Reference meta-data in Euro SDMX Metadata Structure [ESMS]). Retrieved August 9, 2025, from [https://ec.europa.eu/eurostat/cache/metadata/en/spr\\_esms.htm](https://ec.europa.eu/eurostat/cache/metadata/en/spr_esms.htm)
- Eurostat. (2016). *European System of Integrated Social Protection Statistics-ESSPROS: Manual and user guidelines*.
- Eurostat. (2022). *ESSPROS Manual and user guidelines – 2022 edition*. Luxembourg: Publications Office of the European Union.
- Ferrera, M. (1996). The 'Southern Model' of welfare in social Europe. *Journal of European Social Policy*, 6(1), 17-37.

- Goodin, R. E., Headey, B., Muffels, R., & Dirven, H. J. (1999). *The Real Worlds of Welfare Capitalism*. Cambridge: Cambridge University Press.
- Green-Pedersen, C. (2004). The dependent variable problem within the study of welfare state retrenchment: Defining the problem and looking for solutions. *Journal of Comparative Policy Analysis*, 6(1), 3-14.
- Guillén, A. M. (2010). Defrosting the Spanish welfare state: The weight of conservative components. In B. Palier (Ed.), *A long goodbye to Bismarck? The politics of welfare reform in continental Europe*. Amsterdam: Amsterdam University Press. 183-206.
- Hemerijck, A. (2013). *Changing Welfare States*. Oxford: Oxford University Press.
- Holliday, I. (2000). Productivist welfare capitalism: Social policy in East Asia. *Political Studies*, 48(4), 706-723.
- ILO. (2005). *ILO Social Security Inquiry*. International Labour Office Social Security Department.
- ILO. (2019). *Social Security Inquiry (SSI) Getting Started*. International Labour Office Social Security Department.
- Kim, S., Kim, D., Park, C., Lee, W., Song, W., Kim, Y., ..., & Kim, S. (2024). SOLAR 10.7B: Scaling large language models with simple yet effective depth up-scaling. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 6: Industry Track)* (pp. 23-35). arXiv preprint arXiv:2312.15166.
- Korpi, W., & Palme, J. (1998). The paradox of redistribution and strategies of equality: Welfare state institutions, inequality, and

- poverty in the Western countries. *American Sociological Review*, 63(5), 661-687.
- Leibfried, S. (1992). Towards a European welfare state? On integrating poverty regimes in the European Community. In Z. Ferge & J. E. Kolberg (Eds.), *Social policy in a changing Europe*. Frankfurt am Main: Campus Verlag. 245-279.
- Manow, P. (2001). Social protection, capitalist production: The Bismarckian welfare state in the German political economy. In B. Ebbinghaus & P. Manow (Eds.), *Comparing welfare capitalism: Social policy and political economy in Europe, Japan and the USA*. London & New York: Routledge. 138-164.
- Marshall, T. H. (1950). *Citizenship and Social Class*. Cambridge: Cambridge University Press.
- Moreno, L. (2006). The model of social protection in Southern Europe. *Revue française des Affaires sociales*, (5), 73-95.
- Musgrave, R. A. (1959). *The Theory of Public Finance: A study in public economy*. New York: McGraw-Hill.
- OECD. (n.d.). OECD Employment Database [Dataset]. OECD. Retrieved October 16, 2025, from <https://www.oecd.org/en/data/datasets/oecd-employment-database.html>
- OECD. (2019). *The OECD SOCX Manual 2019 Edition: A guide to the OECD Social Expenditure Database*. OECD Publishing.
- OECD. (2025). Social Expenditure Database (SOCX) [Data set]. OECD. Retrieved October 31, 2025, from <https://www.oecd.org/en/data/datasets/social-expenditure-database-socx.html>
- OECD, Eurostat and World Health Organization (2017), *A System of Health Accounts 2011: Revised edition*, OECD Publishing, Paris. <http://dx.doi.org/10.1787/9789264270985-en>

- Office for Budget Responsibility. (2023). Fiscal risks and sustainability (CP 870). His Majesty's Stationery Office.
- Office for Budget Responsibility. (2024). Fiscal risks and sustainability (CP 1142). His Majesty's Stationery Office.
- Office for Budget Responsibility. (2025). Fiscal risks and sustainability (CP 1343). His Majesty's Stationery Office.
- Office of the Chief Actuary. (2022). 31th actuarial report on the Canada Pension Plan as at 31 December 2021. Office of the Superintendent of Financial Institutions Canada.
- Office of the Chief Actuary. (2023). 18th actuarial report on the Old Age Security program as at 31 December 2021. Office of the Superintendent of Financial Institutions Canada.
- Palier, B. (2010). A long goodbye to Bismarck? The politics of welfare reform in continental Europe. Amsterdam: Amsterdam University Press.
- Pierson, P. (1994). Dismantling the welfare state?: Reagan, Thatcher, and the politics of retrenchment. Cambridge University Press.
- Powell, M., & Barrientos, A. (2004). Welfare regimes and the welfare mix. *European Journal of Political Research*, 43(1), 83-105.
- Rawls, J. (1971). A theory of justice. Harvard University Press.
- Rein, M. (1982). Dilemmas of welfare policy: Why work strategies haven't worked. New York: Praeger.
- Rhodes, M. (1997). Southern European welfare states: Identity, problems and prospects for reform. *South European Society and Politics*, 2(3), 1-22.
- Scruggs, L., & Allan, J. (2006). Welfare-state decommodification in 18 OECD countries: A replication and revision. *Journal of European Social Policy*, 16(1), 55-72.

- Sen, A. (1999). Development as freedom. Oxford University Press.
- Stiglitz, J. E. (2000). Economics of the Public Sector (3rd ed.). W. W. Norton & Company.
- Titmuss, R. M. (1958). Essays on 'The Welfare State'. London: Allen & Unwin.
- van Oorschot, W. (2002). Targeting welfare: On the functions and dysfunctions of means testing in social policy. In P. Townsend & D. Gordon (Eds.), World Poverty: New policies to defeat an old enemy. Bristol: Policy Press. 171-193.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30, 5998-6008.
- Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E., ... & Zhou, D. (2022). Chain-of-thought prompting elicits reasoning in large language models. Advances in Neural Information Processing Systems, 35, 24824-24837.

[웹사이트]

- 3Blue1Brown. (2025). 3Blue1Brown. <https://www.3blue1brown.com>, 2025.7.2. 접속
- Eurostat. (n.d.). Eurostat. European Commission. <https://ec.europa.eu/eurostat/>, 2025.8.9. 접속
- Hugging Face. (2025). Hugging Face. <https://huggingface.co>, 2025.7.2. 접속
- ISSA. (2025). <https://www.issa.int/databases/country-profiles>. 2025. 6. 18. 접속
- LMSYS Org. (2025). Chatbot Arena. <https://lmarena.ai>, 2025.7.2. 접속

- Meta AI. (2024). Introducing Llama 3.1: Our most capable models to date. <https://ai.meta.com/blog/meta-llama-3-1/>, 2025.7.2. 접속
- MISSOC. (2025). <https://www.missoc.org/>. 2025. 6.24. 접속
- MISSOC. (n.d.). Comparative tables. Retrieved July 31, 2025, from <https://www.missoc.org/missoc-database/comparative-tables/>
- OECD SOCX. (2025). <https://web-archive.oecd.org/2024-06-24/63248-expenditure.htm>. 2025. 4.18. 접속
- UNESCO Institute for Statistics. (2012). International Standard Classification of Education: ISCED 2011 [PDF]. UNESCO. <https://uis.unesco.org/sites/default/files/documents/international-standard-classification-of-education-isced-2011-en.pdf>. 2025. 8.1. 접속



## 제1절 LLM 기반 비정형 텍스트 처리 이론

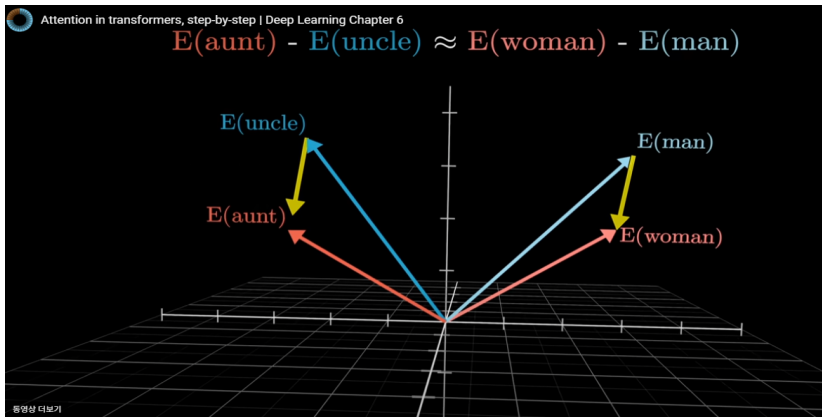
### 1. LLM의 기술적 메커니즘

#### 가. 텍스트 임베딩의 개념과 의미 표현

- 텍스트를 고차원 숫자 벡터로 변환하는 핵심 기술
  - 단어와 문장의 수치적 표현을 통한 컴퓨터 처리 가능화
    - 텍스트의 의미를 보존하면서 수학적 연산이 가능한 형태로 변환
  - 벡터 공간에서의 의미론적 관계 표현
    - 유사한 의미의 단어들이 벡터 공간에서 가까운 위치에 배치
    - ‘aunt’ - ‘uncle’  $\approx$  ‘woman’ - ‘man’과 같은 벡터 연산을 통한 의미 관계 파악 가능([부도 1] 참조)
- 의미적 유사성 기반 공간적 매핑 구조
  - 고차원 벡터 공간에서의 의미 클러스터링
    - 비슷한 맥락에서 사용되는 단어들의 자연스러운 군집 형성
    - 정책 용어들(‘복지’, ‘지원’, ‘급여’ 등)의 의미적 근접성 확보
  - 코사인 유사도를 통한 텍스트 간 관련성 측정
    - 벡터 간 각도를 통한 의미적 유사성 정량화

- 문서 검색, 분류, 클러스터링 등 다양한 NLP 과업의 기초 제공

[부도 1] 임베딩 연산 예시



출처: 3blue1brown. (2025).

## 나. Transformer 아키텍처와 Self-Attention 메커니즘

□ Transformer 모델: “Attention Is All You Need”의 혁신

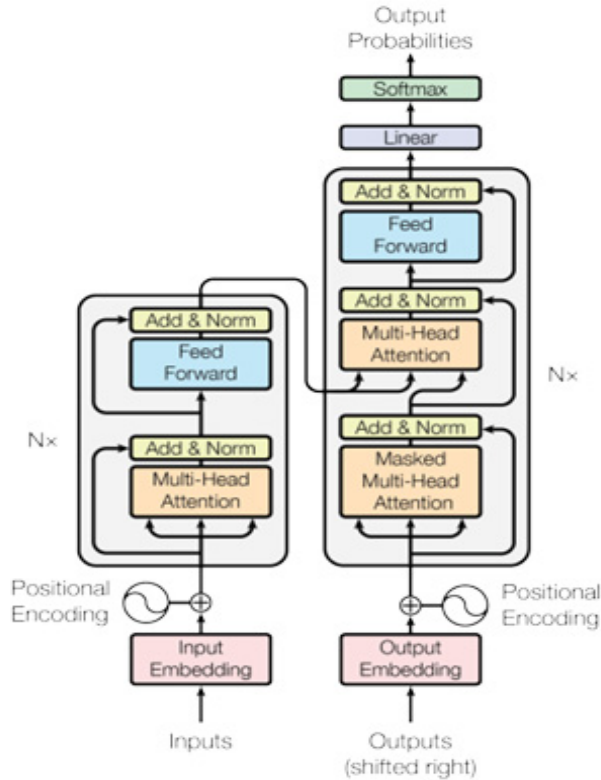
○ Vaswani et al.(2017)의 “Attention Is All You Need”에서 순환 신경망(RNN)의 순차 처리 한계 극복

- 긴 시퀀스에서의 정보 손실 문제 해결
- 병렬 처리를 통한 학습 및 추론 속도 대폭 향상

○ 어텐션 메커니즘을 통한 전역적 정보 활용

- 문장의 모든 위치 간 직접적인 연결을 통한 정보 흐름 최적화
- 장거리 의존성(long-range dependency) 문제 효과적 해결

## [부도 2] 트랜스포머 아키텍처



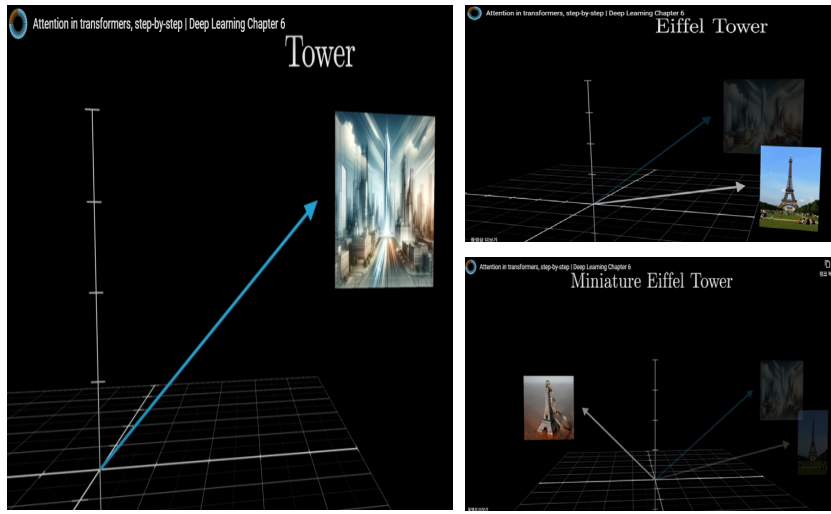
출처: Vaswani et al.(2017). p.3. Figure 1.

## □ 문맥 의존적 동적 가중치 계산

### ○ 셀프 어텐션을 통한 단어 간 관계 가중치 동적 계산

- “배가 고파서 밥을 먹었다”와 “배가 항구에 도착했다”에서 ‘배’의 다른 의미 구분([부도 3] 참조)
- 문맥에 따라 동일한 단어라도 다른 가중치를 부여하여 정확한 의미 파악

[부도 3] 셀프 어텐션 매커니즘 예시



출처: 3blue1brown. (2025).

○ Query, Key, Value 행렬을 통한 어텐션 점수 계산

- 각 토큰이 다른 토큰들과 얼마나 관련이 있는지를 수치적으로 계산(Vaswani et al., 2017)
- 소프트맥스 함수를 통한 정규화로 전체 어텐션 가중치 합이 1 이 되도록 조정

□ 병렬 처리를 통한 효율적 연산 구조

○ 멀티헤드 어텐션을 통한 다양한 관점의 정보 포착

- 여러 개의 어텐션 헤드가 서로 다른 종류의 관계를 학습 (Vaswani et al., 2017)
- 문법적 관계, 의미적 관계, 화용적 관계 등을 동시에 모델링

## ○ GPU 병렬 연산에 최적화된 행렬 연산 구조

- RNN의 순차적 처리와 달리 모든 위치를 동시에 계산
- 대규모 모델의 효율적 학습을 위한 하드웨어 활용 극대화

**다. 언어모델 유형과 최근 동향: Causal LLM 중심**

## □ Causal LLM(GPT 계열)의 순차적 텍스트 생성 메커니즘

## ○ 자기회귀적(Autoregressive) 생성 방식의 작동 원리

- 이전 토큰들을 조건으로 다음 토큰의 확률 분포 예측
- 좌측 컨텍스트만을 참조하는 인과적(Causal) 마스킹 적용

## ○ 확률 기반 텍스트 생성의 제어 가능성

- Temperature, Top-p, Top-k 등의 샘플링 기법을 통한 생성 다양성 조절
- 창의성과 일관성 간의 균형 조절을 통한 과업별 최적화

## □ 지시(Instruction) 기반 복잡 과업 수행 능력

## ○ 인-컨텍스트 러닝(In-Context Learning)을 통한 적응성

- 별도의 모델 수정 없이 프롬프트 내 예시를 통한 과업 학습 (Brown et al., 2020)
- Few-shot 학습을 통한 새로운 도메인으로의 빠른 적응

## ○ 체인 오브 쏫트(Chain-of-Thought) 추론 능력

- 복잡한 문제를 단계별로 분해하여 해결하는 추론 과정(Wei et al., 2022)

- “단계별로 생각해보자”와 같은 프롬프트를 통한 추론 품질 향상

□ 방대한 데이터 학습을 통한 창발적 능력 발현

○ 스케일링 법칙(Scaling Laws)에 따른 성능 향상

- 모델 크기, 데이터 양, 컴퓨팅 자원 증가에 따른 예측 가능한 성능 개선
- GPT-3(175B) → GPT-4(추정 1.7T) 등 대규모화를 통한 질적 변화

○ 다양한 도메인 지식의 통합적 활용

- 단순한 언어 모델링을 넘어선 추론, 요약, 번역, 코딩 등 다중 능력
- 명시적으로 학습하지 않은 과업에서도 뛰어난 성능 발휘

## 2. 최근 LLM 동향 및 특징

### 가. 오픈소스 LLM 생태계

□ Meta Llama 시리즈의 성능 진화 및 특징

○ Llama 3.1 시리즈의 파라미터별 성능 차별화

- 8B, 70B, 405B 모델로 구성된 계층적 접근(Meta AI, 2024)
- 405B 모델의 경우 GPT-4 수준의 성능을 오픈소스로 제공

○ 상용 라이선스와 연구용 라이선스의 이중 구조

- 월 7억 활성 사용자 미만 서비스에 대해서는 무료 상업적 이용 허용
- 대규모 서비스 운영 시 별도 라이선스 협상 필요

#### □ Google Gemma 모델의 효율성 특화 요소

- 경량화된 모델 구조를 통한 효율적 추론 성능
  - Gemma 3 4B, 12B, 27B 모델의 컴팩트한 아키텍처 설계
  - Gemini 기술을 기반으로 한 성능 대비 효율성 최적화
- 안전성과 책임감 있는 AI 개발에 중점
  - 모델 카드와 상세한 안전성 평가 결과 공개
  - 편향성 완화 및 유해 콘텐츠 생성 방지 메커니즘 내장

#### □ Mistral AI 모델의 다국어 처리 능력

- 유럽 중심의 다국어 학습 데이터 활용
  - 프랑스어, 독일어, 스페인어 등 유럽 언어에 특화된 성능
  - 영어 중심 모델 대비 비영어권 언어 처리 능력 우수
- 코딩 및 수학적 추론 능력 강화
  - Mistral Large의 경우 복잡한 논리적 추론 과업에서 높은 성능
  - 함수 호출(Function Calling) 기능을 통한 도구 활용 능력

#### □ 한국의 Open LLM

- LG AI Research - EXAONE 시리즈

- 한국어와 영어 이중 언어 모델로 설계
- 기업용 특화 기능 및 한국 문화 맥락 이해 능력 강화

○ KAKAO - Kanana

- 한국어 자연어 처리에 특화된 사전 학습 모델
- 대화형 AI 서비스에 최적화된 성능 튜닝

○ Upstage Solar

- DUS와 같은 효율적인 학습 방법론을 통한 성능 최적화(Kim et al., 2024)
- 오픈소스 커뮤니티와의 적극적 협력을 통한 생태계 구축

〈부표 1〉 Open LLM 비교

모델	Arena Score	파라미터	요구 메모리	라이선스
google/gemma-3-27b-it	1365	27B	~54GB	Gemma
Qwen/Qwen3-32B	1347	32B	~64GB	Apache2.0
mistralai/Mistral-Small-3.1-24B-Instruct-2503	1343	24B	~48GB	Apache2.0
LGAI-EXAONE/EXAONE-3.5-32B-Instruct	1338 <sup>1)</sup>	32B	~128GB	Exaone
meta-llama/Llama-4-Scout-17B-16E-Instruct	1233	17B-16E <sup>2)</sup>	~200GB+	Llama-4

주: 1) Hugging Face. (2025), 활용 추정  
2) MoE(Mixture of Experts): 17B 파라미터 × 16개 전문가  
출처: Hugging Face. (2025), LMSYS Org. (2025).

나. 상용 LLM 서비스 특성

□ OpenAI GPT 시리즈의 멀티모달 기능

○ GPT-4o의 통합적 멀티모달 처리 능력

- 텍스트, 이미지, 음성을 하나의 모델에서 통합 처리
- 실시간 음성 대화 및 시각적 이해 능력 제공
- API 생태계 구축을 통한 개발자 친화적 환경
  - 다양한 프로그래밍 언어 지원 및 상세한 문서화
  - 배치 API를 통한 대량 처리 비용 절감 옵션 제공
- Anthropic Claude의 MCP 및 에이전트 오케스트레이션 강점
  - Model Context Protocol(MCP)을 통한 외부 도구 연동
    - 데이터베이스, API, 파일 시스템 등과의 표준화된 연결 인터페이스 제공
    - 복잡한 워크플로우에서 여러 도구를 순차적으로 활용하는 능력
  - 에이전트 기반 복합 과업 처리 능력
    - 단일 요청을 여러 하위 작업으로 분해하여 체계적 처리
    - 계획 수립, 실행, 검증의 순환적 과정을 통한 복잡한 문제 해결
- Google Gemini의 대용량 컨텍스트 윈도우 지원
  - 최대 2M 토큰의 초대용량 컨텍스트 처리
    - 전체 코드베이스나 긴 연구 논문을 한 번에 분석 가능
    - 복잡한 문서 간 관계 분석 및 종합적 인사이트 도출
  - Google 검색 및 기타 서비스와의 통합 활용
    - 실시간 정보 검색을 통한 최신성 확보
    - YouTube, Gmail 등 Google 생태계와의 연동 기능

〈부표 2〉 상용 LLM 비교

모델	Arena Score <sup>1)</sup>	비용	특징
Gemini 2.5 Pro	1465	input: \$1.25 output: \$5.00	긴 문서처리, Google서비스 연동
ChatGPT-4o	1443	input: \$2.50 output: \$10.00	멀티모달 통합, 이미지생성
Claude Opus 4	1417	input: \$15.00 output: \$75.00	개발자도구(MCP), 코딩

주: 1) LMSYS Org. (2025)의 Text Overall Score  
출처: LMSYS Org. (2025).

다. 모델 선택 전략 및 기준

□ DB 구축 과업 특성에 따른 모델 선택

○ 대량 문서 처리 요구사항

- 한국 중앙정부 사업설명자료: 연간 약 1만 건 (PDF 2-3페이지 분량)
- 해외(OECD) 정책자료: 웹 크롤링 기반 텍스트 정보
- 한국자료 기준 총 예상 처리량: 연 1만 건 × 5년 = 약 5만 건
- 반복 처리: 분류 기준 개선에 따른 재처리 및 연도별 업데이트

○ 토큰당 비용 산정 예시

- 한국 사업설명자료 1만 건: 약 5천만 토큰 (건당 ~5천 토큰)
- 5천만 토큰 기준 상용 API 비용: \$62,500~375,000 (Gemini~Claude)
- 다년도 누적 시 오픈소스 모델의 경제적 우위 확실

○ 비용 효율성 중심의 오픈소스 모델 우선 검토

- 다년도 누적 + 반복 처리 시 오픈소스 압도적 경제성
- 한 번 구축 후 반복 실험 및 업그레이드 가능 (분류 기준 변경, A/B 테스트 등)

#### □ 도메인 특화 및 연구 유연성

##### ○ 언어 및 정책 도메인 특화 성능 고려 가능

- 사회보장 분야 전문 용어: “급여체계”, “수급요건”, “소득대체율” 등
- 각국 제도 분류 기준: 사회보험/사회수당/사회부조 정확한 구분
- 오픈소스 모델: 도메인 특화 Tuning 및 프롬프트 최적화 자유
- 상용 API: 일반목적 모델로 제한, 커스터마이징 불가

##### ○ 연구 유연성 및 지속가능성

- 반복 실험: 분류 기준 변경 시 무제한 처리
- 알고리즘 투명성: 분류 결과에 대한 완전한 추적 가능성
- 기술 독립성: 특정 모델 및 업체 의존도 최소화
- 예산 예측 가능성: 고정 인프라 비용으로 안정적 운영

#### □ 보안 및 프라이버시

- 자체 서버 활용: 외부 API 전송 시 정보 유출 위험 감소
- 보안 네트워크 운영: 네트워크 통제 가능

### 3. 대량 비정형 텍스트 처리를 위한 LLM 활용 방안

#### 가. 프롬프트 엔지니어링과 In-Context Learning 기법

##### □ Few-shot 프롬프팅을 통한 도메인 적응

###### ○ 예시 기반 학습을 통한 과업 수행 능력 향상

- 3-5개의 고품질 예시를 통한 원하는 출력 패턴 학습
- 예시 선택의 중요성: 대표성, 다양성, 품질을 고려한 선별
- 예시 순서 및 구성이 성능에 미치는 영향 분석

###### ○ 도메인 적응을 위한 예시 설계 원칙

- 타겟 도메인의 전문 용어 및 문체 특성 반영
- 예외 상황 및 경계 사례를 포함한 다양한 시나리오 제시
- 일관된 출력 형식 유도를 위한 템플릿 활용

##### □ Chain-of-Thought 프롬프팅 기법

###### ○ 단계적 추론을 통한 복잡한 문제 해결

- 단계적 지시를 통한 체계적 접근
- 중간 추론 단계를 명시적으로 표현하여 최종 결과 품질 향상
- 복잡한 논리적 관계나 다단계 분석이 필요한 과업에 효과적

###### ○ 추론 품질 향상을 위한 프롬프트 설계

- 명확한 추론 구조 제시: 문제 파악 → 분석 → 결론 도출
- 각 단계별 판단 근거 요구를 통한 투명성 확보
- 오류 발생 가능 지점에 대한 자체 검증 메커니즘 포함

□ 프롬프트 최적화 및 성능 향상 기법

- 체계적 프롬프트 개선 방법론
  - A/B 테스트를 통한 프롬프트 버전 간 성능 비교
  - 실패 사례 분석을 통한 프롬프트 약점 파악 및 보완
- 일관성 및 신뢰성 확보 방안
  - 명확하고 구체적인 지시문을 통한 모호성 제거
  - 출력 형식 제약 조건 명시를 통한 표준화된 결과 유도
  - 불확실한 경우의 처리 방법 사전 정의

〈부표 3〉 프롬프트 엔지니어링 기법별 특성 및 적용 방안

방법론	특징	적용상황	방법
Few-shot Learning	예시 기반 빠른 적응성 템플릿 활용	도메인 특화 분류 새로운 과업 일관된 출력	3~5개 고품질 샘플 제시 대표성, 다양성 고려한 예시 JSON 스키마 기반 구조화
Chain-of-Thought	단계적 추론 투명성 확보 자체 검증	복잡한 과업 판단 근거 필요 시 오류 방지 중요	추론 과정 명시적 표현 문제-분석-결론 구조 각 단계별 점검 메커니즘
최적화 & 성능 향상	반복적 개선 표준화 지속적 발전	성능 부족시 일관성 확보 장기 운영시	A/B 테스트 및 실패 분석 모호성 제거, 형식 통일 오류 분석 후 보완

출처: 저자 작성.

나. 오픈소스 LLM 기반 로컬 처리 환경 구축

- GPU 메모리 효율성을 고려한 모델 선택
  - VRAM 수준을 고려한 모델 선택

- 8GB VRAM: 8B 수준 모델 양자화(4-bit) 활용
- 24GB VRAM: 20B 수준 이상 모델의 양자화 또는 8B 수준 모델 전체 정밀도
- 80GB+ VRAM: 20B 수준 이상 모델 전체 정밀도 또는 여러 모델 동시 로딩

□ 추론 최적화를 위한 기술적 접근

○ vLLM 기반 고성능 추론 환경

- PagedAttention을 통한 메모리 효율적 어텐션 계산
- 동적 배치 구성을 통한 GPU 활용률 극대화
- 연속 배치 처리를 통한 대기 시간 최소화, 처리량 극대화

○ 배치처리를 통한 처리량(throughput) 극대화

- 여러 문서를 동시에 처리하여 GPU 유휴 시간 최소화
- 적응적 배치 크기 조정을 통한 메모리 사용량 최적화
- 비동기 처리를 적용하여 I/O 대기 시간 동안 GPU 연산 지속

□ 구조화된 출력 처리 및 데이터 관리

○ JSON 스키마 기반 구조화 출력

- Pydantic 모델 활용한 데이터 검증
  - Pydantic은 Python에서 데이터 검증(validation)을 위한 라이브러리임.
- 잘못된 형식 출력 자동 재시도 메커니즘
- 스키마 준수율 모니터링 및 품질 관리

## ○ 벡터 데이터베이스 및 RAG 활용

- 텍스트 임베딩 생성 및 저장: ChromaDB, Pinecone 등
- RAG(Retrieval-Augmented Generation)를 통한 분류 정확도 향상
  - 유사 사례 검색 후 참조하여 일관성 있는 분류 수행
- 임베딩 기반 중복 문서 탐지 및 제거
- 기존 분류 사례 학습을 통한 점진적 성능 개선

## 제2절 비정형 텍스트의 구조화 및 DB화 과정

### 1. 비정형 데이터 종류와 수집 방법

#### 가. 웹 기반 데이터 수집

## □ 크롤링

## ○ 정적 웹페이지 크롤링 기법

- BeautifulSoup을 활용한 HTML 파싱 및 텍스트 추출
- CSS 선택자를 이용한 특정 영역 데이터 추출
- 페이지네이션 처리를 통한 대량 문서 자동 수집

## ○ 동적 웹페이지 처리 전략

- Selenium을 활용한 JavaScript 렌더링 페이지 처리
- Playwright를 이용한 현대적 SPA(Single Page Application)

## 대응

- 비동기 로딩 콘텐츠의 완전 로딩 대기 메커니즘

### ○ 크롤링 윤리 및 기술적 고려사항

- robots.txt 준수 및 사이트 정책 확인
- 요청 간격 조절을 통한 서버 부하 최소화
- User-Agent 설정 및 세션 관리를 통한 차단 방지

〈부표 4〉 데이터 수집 방법별 특징 비교

수집방법	장점	단점	기술적 요구사항
정적 웹 크롤링	다양한 웹사이트 접근 비용 효율적 대량 수집 가능	robots.txt 제약 IP 차단 위험 사이트 변경 시 수정 필요	BeautifulSoup CSS 선택자 요청 간격 조절
동적 웹 크롤링	JavaScript 페이지 처리 사용자 상호작용 모방	높은 리소스 소모 복잡한 설정 느린 처리 속도	Selenium Playwright 헤드리스 브라우저
공개 API	안정적 데이터 제공 표준화된 형식 실시간 업데이트	API 키 필요 요청 제한 존재 제공 데이터 한정	API 키 관리 Rate Limit 준수 JSON/XML 파싱
문서파일 처리	공식 문서 직접 처리 완전한 정보 보존 메타데이터 활용	형식별 파서 필요 OCR 품질 의존 수동 업로드 필요	PyMuPDF python-hwp OCR 엔진

출처: 저자 작성.

## □ 공공데이터포털 등 공개 API 활용

### ○ 정부 기관별 API 연동 전략

- API 키 관리 및 요청 제한(Rate Limit) 준수 방안
- 실시간 데이터 업데이트를 위한 스케줄링 시스템 구축

### ○ 표준화된 데이터 형식 처리

- XML, JSON 등 다양한 응답 형식의 통합 처리
- 페이징 처리를 통한 대용량 데이터 분할 수집
- API 응답 오류 처리 및 재시도 로직 구현

○ 메타데이터 수집 및 관리

- 문서 게시일, 수정일, 작성기관 등 부가 정보 확보
- 데이터 출처 추적을 위한 URL 및 수집 시점 기록
- 라이선스 정보 및 이용 제한 사항 확인

## 나. 문서 파일 텍스트 추출

□ PDF 및 한글(HWP), MS Office 문서 처리 기법

○ PDF 문서 텍스트 추출 최적화

- PyMuPDF를 활용한 고품질 텍스트 추출
- 스캔된 PDF 자동 감지 및 OCR 처리 분기
- 페이지별 레이아웃 분석을 통한 읽기 순서 최적화

○ 한글 문서(HWP) 처리 전문화

- python-hwp 라이브러리를 통한 네이티브 파싱
- 한글 고유 서식(각주, 표, 도형) 정보 보존
- 버전별 호환성 처리 및 오류 복구 메커니즘

○ MS Office 문서 통합 처리

- python-docx, openpyxl을 통한 Word/Excel 파일 처리
- 문서 내 이미지, 표, 차트 등 멀티미디어 요소 식별

- 문서 속성 및 변경 이력 추적 정보 활용

□ OCR 및 멀티모달 모델을 활용한 복합 정보 추출

○ 전통적 OCR 기술의 텍스트 인식

- Tesseract를 활용한 무료 OCR 처리
- Google Cloud Vision API의 고정밀도 텍스트 인식

○ 멀티모달 LLM을 활용한 통합적 문서 이해

- GPT, Claude 등 상용 멀티모달 모델 또는 Gemma, Mistral 등 오픈소스 멀티모달 모델을 활용한 문서 구조 및 내용 동시 분석
- 표, 차트, 다이어그램을 포함한 시각적 요소의 의미 해석
- 이미지와 텍스트 간의 관계를 이해한 맥락적 정보 추출

○ 구조화 정보 보존 및 통합 처리

- 멀티모달 모델을 통한 표 구조 자동 인식 및 데이터 추출
- 문서 계층 구조(제목, 소제목, 본문) 및 레이아웃 분석
- 이미지 캡션, 차트 설명, 각주 등과 본문 간의 연관 관계 파악

□ 문서 형식별 특화 파서 개발

○ 문서 템플릿 대응

- 국회 법률안 등의 표준 구조(제안이유, 주요내용, 조문)
- 정부 부처 보도자료의 일반적 형식(헤드라인, 본문, 담당자)
- 연구보고서의 학술적 구조(요약, 서론, 본론, 결론)

○ 예외 상황 처리 및 오류 복구

- 비표준 형식 문서에 대한 fallback 파서
- 파싱 실패 시 원시 텍스트 백업 저장
- 수동 검토가 필요한 문서 자동 플래깅

〈부표 5〉 문서 형식별 처리 도구 및 기법

문서 형식	주요 도구	OCR 필요성	기술적 고려사항
PDF (텍스트)	PyMuPDF PDFplumber pdfminer	불필요	레이아웃 분석 표 구조 인식 직접 텍스트 추출
PDF (이미지)	Tesseract 멀티모달 LLM 등	필수	이미지 품질 OCR → 후처리
HWP	python-hwp pyhwp olefile	경우에 따라	한글 고유 서식 버전 호환성 파서 필요
Word	python-docx python-docx2txt	불필요	표/이미지 분리 구조화 파싱 필요
Excel	openpyxl pandas xlrd	불필요	시트별 처리 셀 병합 인식
이미지 문서	멀티모달 LLM OCR + NLP	필수	레이아웃 이해 맥락 이해 LLM 기반 통합 처리

출처: 저자 작성.

2. 텍스트 전처리 및 구조화

가. 데이터 정제 및 정규화

- 노이즈 제거 및 불필요 요소 필터링
  - 웹 페이지 노이즈 제거
    - HTML 태그, CSS 스타일, JavaScript 코드 완전 제거

- 네비게이션 메뉴, 광고, 푸터 등 콘텐츠 외 요소 식별 및 삭제
- 중복된 헤더/푸터가 포함된 다중 페이지 문서의 반복 요소 제거

○ 문서 형식별 특수 요소 처리

- PDF 변환 시 발생하는 페이지 번호, 워터마크 제거
- 표의 셀 구분자 및 줄바꿈 정규화
- 각주, 참고문헌 영역 별도 식별 및 분리 보관

□ 인코딩 문제 해결 및 텍스트 정규화

○ 문자 인코딩 통일화

- UTF-8 기준 인코딩 변환 및 깨진 문자 복구
- 한글 완성형/조합형 혼재 문제 해결
- 특수문자 및 이모지의 표준 유니코드 변환

○ 텍스트 정규화 처리

- 연속된 공백, 탭, 줄바꿈의 단일 공백 변환
- 전각/반각 문자 통일 및 대소문자 정규화
- 괄호, 따옴표 등 문장부호의 표준화

□ LLM 기반 논리적 구조 인식 파싱

○ 문서 구조 자동 분석

- 제목 계층 구조(H1, H2, H3) 자동 식별 및 태깅
- 본문, 목록, 표, 인용구 등 콘텐츠 유형별 분류
- 문서 내 섹션 간 논리적 연결 관계 파악

- 정책 문서 특화 구조 인식
  - 법령의 조, 항, 호 구조 자동 파싱
  - 정책 보고서의 요약-배경-내용-결론 구조 식별
  - 예산 관련 수치 및 테이블 정보 별도 추출

〈부표 6〉 전처리 단계별 처리 내용 및 도구

전처리 단계	주요 작업	사용 도구	처리 대상
노이즈 제거	HTML 태그 제거 네비게이션 메뉴 삭제 광고/푸터 제거 중복 요소 정리	BeautifulSoup 정규표현식 사용자 정의 필터	웹 크롤링 데이터 PDF 변환 텍스트
인코딩 정규화	UTF-8 변환 깨진 문자 복구 특수문자 표준화 이모지 처리	chardet ftfy unicodedata	모든 텍스트 데이터
텍스트 정규화	공백 정리 전각/반각 통일 대소문자 정규화 문장부호 표준화	re (정규표현식) NLTK 사용자 정의 함수	정제된 텍스트
구조인식	제목 계층 분석 섹션 구분 콘텐츠 유형 분류 논리적 관계 파악	LLM API 사용자 정의 파서	구조화된 문서
청킹	의미 단위 분할 중첩 구간 설정 메타데이터 첨부 순서 관리	LangChain 사용자 정의 분할	전처리 완료 텍스트

출처: 저자 작성.

나. 문서 청킹

□ 의미 기반 분할을 통한 문맥 보존

○ 문단 및 섹션 기반 자연스러운 분할

- 문단 구분 및 의미적 완결성을 고려한 청크 크기 결정
- 한 주제 내용이 분리되지 않도록 섹션 경계 우선 고려
- 문장 중간 절단 방지를 위한 문장 단위 완결성 보장

○ 문서 유형별 최적화된 청킹 전략

- 긴 법령: 조문 단위 분할 및 관련 조항 그룹핑
- 연구보고서: 장/절 구조를 따른 논리적 분할
- 정책 브리프: 핵심 정보 밀도를 고려한 균등 분할

□ 중첩 청킹을 통한 연속성 확보

○ 컨텍스트 연결을 위한 오버랩 설정

- 청크 간 10-20% 중첩을 통한 문맥 연속성 유지
- 이전 청크의 마지막 문단과 다음 청크의 첫 문단 부분 중복
- 중요한 정보가 청크 경계에서 손실되지 않도록 안전장치 구성

○ 메타데이터 보존 및 추적 체계

- 원본 문서명, 페이지 번호, 섹션 정보 각 청크에 첨부
- 청크 간 순서 및 상하위 관계 정보 보존
- 수정 이력 추적을 위한 청크별 버전 관리

## 다. LLM 기반 구조화

□ JSON 스키마 기반 정보 필드 정의

○ 사회보장 정책 특화 스키마 설계

- 정책명, 시행기관, 대상자, 급여내용, 예산규모 등 필수 필드

- 시행일, 근거법령, 신청방법 등 선택적 필드 구성

○ 계층적 정보 구조 반영

- 상위 정책과 하위 세부 사업 간의 관계 설정
- 지역별 시행 차이 및 예외 사항 구조화
- 시계열 변화 추적을 위한 버전 정보 포함

□ Few-shot 프롬프팅을 통한 자동 태깅

○ 도메인 특화 예시 기반 학습

- 3-5개의 고품질 정책 분석 예시를 프롬프트에 포함
- 애매한 경계 사례에 대한 판단 기준 명시
- 전문 용어 및 약어 처리 방법 예시 제공

○ 일관성 보장을 위한 프롬프트 표준화

- 동일한 프롬프트 템플릿을 모든 문서에 적용
- 출력 형식의 엄격한 준수를 위한 제약 조건 명시
- 불확실한 정보에 대한 표기 방법 통일

□ 단계적 정보 추출 및 검증 메커니즘

○ 단계적 정보 추출 체계

- 1단계: 문서 유형 및 주제 분류
- 2단계: 핵심 정보 요소 식별 및 추출
- 3단계: 세부 정보 보완 및 관계 정보 파악

○ 검증 및 품질 관리

- 필수 필드 누락 여부 자동 체크
- 숫자 정보의 합리성 검증 (예: 예산 규모의 단위 일치성)
- 이전 유사 정책과의 일관성 비교 검증

### 3. 구조화된 데이터 저장 체계

#### 가. LLM의 JSON 추론

##### □ 구조화된 정보 추출을 위한 LLM 활용

###### ○ 프롬프트 엔지니어링을 통한 정확한 JSON 생성

- 명확한 JSON 스키마 제시 및 출력 형식 강제
- 각 필드별 추출 방법 및 판단 기준 상세 설명
- 정보 부재 시 처리 방법 (“정보없음”, null 등) 명시

###### ○ 복잡한 정책 정보의 구조화 처리

- 중첩된 정책 구조의 계층적 JSON 표현
- 다중 대상자, 다양한 급여 형태의 배열 형태 처리
- 조건부 적용 규칙의 논리적 구조화

##### □ JSON 품질 검증 및 오류 처리

###### ○ 문법적 검증 및 스키마 준수 확인

- JSON 파싱 오류 자동 감지 및 재처리 요청
- 스키마 위반 항목에 대한 자동 수정 또는 플래깅
- 데이터 타입 불일치 문제 해결 방안 마련 필요

## ○ 의미적 일관성 검증

- 정책 분야별 일반적 패턴과의 일치성 검토
- 상식적 범위를 벗어나는 수치 정보 플래깅
- 전후 맥락과 모순되는 정보 식별

**나. JSON 파일 저장 및 DB화**

## □ 파일 시스템 기반 JSON 저장 관리

## ○ 계층적 디렉토리 구조 설계

- 회계연도/소관부처/정책분야별 구조 구성
- 파일명 규칙 통일 (날짜\_기관\_문서유형\_일련번호.json)
- 원본 문서와 추출된 JSON 간의 매핑 정보 유지

## ○ 버전 관리 및 백업 체계

- Git 기반 변경 이력 추적 및 롤백 기능
- 정기적 백업 및 원격 저장소 동기화

## □ PostgreSQL 기반 구조화 데이터 저장 및 관리

## ○ JSON 데이터의 관계형 테이블 설계

- JSONB 컬럼을 통한 반구조화 데이터 효율적 저장
- JSON 경로 쿼리(->>, #>)를 활용한 중첩 데이터 접근
- GIN 인덱스를 통한 JSON 필드 고속 검색 지원

## ○ JSON 데이터의 관계형 테이블 설계

- 정책 기본 정보 테이블(policies): 정책명, 시행기관, 시행일

등 핵심 필드

- 정책 상세 정보 테이블(policy\_details): 대상자, 급여내용, 예산 등 가변 정보
- 정책 관계 테이블(policy\_relations): 상위-하위 정책, 관련 정책 간 연관성

□ 실시간 검색 및 분석을 위한 Elasticsearch 연동

○ 전문 검색 및 필터링 기능

- 정책명, 대상자, 키워드 기반 실시간 검색
- 다차원 필터링 지원
- 유사 정책 추천을 위한 More Like This 기능 등

○ 대시보드 및 시각화 지원

- 기관별, 분야별 정책 현황 실시간 모니터링
- 예산 규모, 수혜자 수 등 주요 지표의 시계열 변화 추적

## 제3절 텍스트 DB의 품질관리 및 검증 체계

### 1. UMAP을 활용한 시각적 검증

#### 가. 임베딩 기반 차원 축소 시각화

□ 고차원 벡터의 2D/3D 시각화

○ 텍스트 임베딩 벡터의 차원 축소 과정

- 정책 문서별 의미적 유사성을 공간적 거리로 표현
- 고차원 공간의 복잡한 관계를 인간이 직관적으로 이해 가능한 형태로 변환

#### ○ 시각화를 통한 데이터 분포 파악

- 전체 정책 집합의 주제별 클러스터 분포 확인
- 특정 정책 분야의 밀집도 및 다양성 정도 측정
- 시계열 변화에 따른 정책 트렌드의 공간적 이동 추적

#### ○ 품질 검증을 위한 시각적 단서 제공

- 예상과 다른 위치에 배치된 문서들의 이상 패턴 식별
- 동일 주제 문서들이 분산된 경우 전처리 오류 가능성 진단
- 고립된 문서들의 특수성 여부 판단을 위한 시각적 근거 제공

### □ UMAP 알고리즘의 특성 및 장점

#### ○ t-SNE 대비 UMAP의 우수성

- 대규모 데이터셋에서 10-100배 빠른 연산 속도 제공
- 전역적 구조 보존으로 클러스터 간 거리의 의미적 해석 가능
- 하이퍼파라미터에 대한 robust성으로 안정적 결과 도출

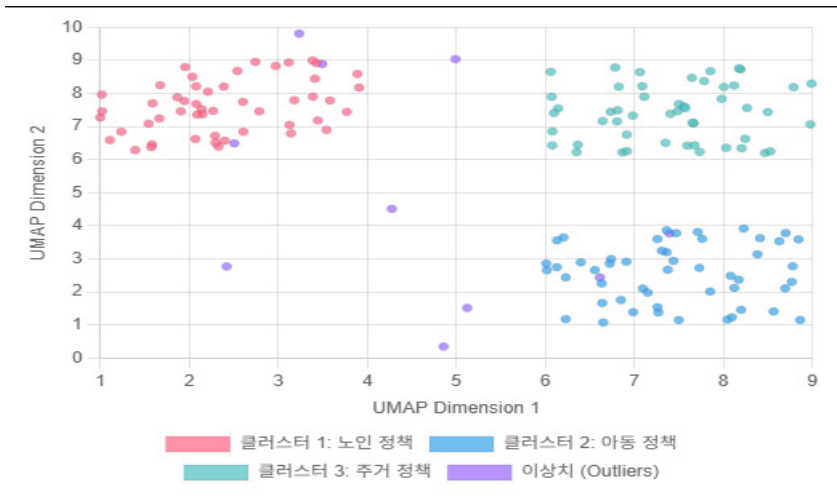
#### ○ 도메인 특화 분석에 UMAP 활용

- n\_neighbors 파라미터 조정을 통한 지역적/전역적 구조 균형
- min\_dist 설정으로 정책 세부 분야 간 적절한 분리도 확보
- metric 선택(cosine, euclidean)에 따른 의미적 유사성 측정 최적화

### ○ 재현 가능한 결과를 위한 설정 관리

- random\_state 고정을 통한 일관된 시각화 결과 보장
- 동일 데이터셋에 대한 반복 실행 시 동일한 레이아웃 유지
- 버전 관리 시스템과 연동한 파라미터 설정 이력 추적

[부도 4] UMAP 시각화 및 클러스터 분석 예시



출처: 저자 작성.

## 나. 클러스터 분석 및 이상치 탐지

### □ 클러스터 분석

#### ○ HDBSCAN을 활용한 밀도 기반 클러스터링

- 정책영역별 자연스러운 군집 형성 및 경계 식별
- 클러스터 계층 구조를 통한 상위-하위 정책 분야 관계 파악
- 노이즈 포인트 식별로 분류 애매한 문서 별도 관리

## ○ 도메인 지식과의 일치성 검증

- 사회보장 정책 분류 체계와 클러스터 결과 비교
- 예상 외 클러스터 구성에 대한 정책학적 해석 및 검증
- 전문가 의견 수렴을 통한 클러스터 타당성 평가

## □ 이상치 탐지

## ○ 통계적 이상치 탐지 기법

- Local Outlier Factor(LOF)를 통한 지역적 밀도 기반 이상치 식별
- Z-score 및 IQR 방법을 통한 임베딩 차원별 이상값 검출

## ○ 도메인 특화 이상치 분석

- 특정 정책 분야에서 크게 벗어난 내용을 포함한 문서 식별
- 분류 오류 또는 전처리 실패로 인한 노이즈 데이터 탐지
- 새로운 정책 영역 또는 융합 정책의 특성 구분

## ○ 이상치 처리 전략 수립

- 재처리가 필요한 문서와 별도 분석이 필요한 문서 분류
- 이상치 처리 결과에 대한 이력 관리 및 개선 체계 마련

**다. 대화형 시각화 도구 구축**

## □ 동적 시각화 도구 설계 및 구현

## ○ 웹 기반 대화형 인터페이스 구축

- Plotly Dash를 활용한 반응형 시각화 대시보드 개발

- Streamlit을 통한 빠른 프로토타입 및 사용자 피드백 수렴
- D3.js 기반 커스텀 시각화로 정책 문서 특화 인터랙션 구현

○ 다양한 시각화 모드 제공

- 2D/3D 전환 기능으로 다각도 데이터 관찰
- 색상 코딩을 통한 시간, 기관, 주제별 문서 구분

○ 업데이트 및 필터링

- 업데이트된 데이터에 대한 임베딩 및 시각화 갱신

□ 탐색적 데이터 분석 기능 제공

○ 인터랙티브 데이터 탐색 기능

- 마우스 호버를 통한 개별 문서 정보 즉시 표시
- 클릭을 통한 원본 문서 및 추출된 정보 상세 보기
- 드래그 앤 드롭으로 관심 영역 선택 및 일괄 분석

○ 비교 분석 및 패턴 발견 도구

- 두 개 이상 정책 문서 간 유사성 및 차이점 시각적 비교
- 특정 키워드 검색 결과의 공간적 분포 패턴 분석

## 2. 시각적 검증 기반 전문가 검수 체계

### 가. 전략적 샘플링

□ 클러스터 중심부 대표 문서 선별

○ 클러스터 중심성 기반 문서 선택

- 각 클러스터의 무게중심(centroid)에 가장 가까운 문서들 우선 선별
- 클러스터 내 평균적 특성을 대표하는 전형적 사례 확보
- 클러스터 크기에 비례한 샘플 수 할당으로 전체 분포 반영

○ 의미적 대표성 확보 방안

- 클러스터 내 문서 간 유사도 분포를 고려한 다양성 샘플링
- 정책 유형별 균형적 표본 구성

○ 효율적 검수를 위한 우선순위 설정

- 중요도가 높은 정책 분야(예: 기초생활보장, 국민연금) 클러스터 우선 처리
- 신규 사업에 대한 집중적 검토

□ 경계 지역 모호한 분류 사례 샘플링

○ 클러스터 경계 문서 식별 및 선별

- 두 개 이상 클러스터의 경계에 위치한 애매한 분류 사례 우선 검토
- 클러스터 할당 확률이 낮은 문서들의 분류 타당성 검증
- HDBSCAN의 노이즈 포인트로 분류된 문서들의 재검토

○ 분류 불확실성이 높은 사례 처리

- 여러 정책 분야에 걸친 중첩 정책의 주 분야 결정
- 시기별로 정책 성격이 변화한 사업의 분류 기준 설정

○ 도메인 전문 지식 기반 경계 사례 해석

- 정책학 관점에서의 분류 타당성 평가
- 예산체계상의 분류와 정책 내용 분류 간 불일치 해소를 위한 기준 설정

## 나. 전문가 검수 및 품질 관리 체계

### □ 전문가 간 협업 및 의견 조율 체계

#### ○ 다분야 전문가 협업 체계 구축

- 정책학, 사회복지학, 행정학 등 관련 분야 전문가 풀 구성
- 분야별 전문성을 고려한 문서 할당 및 교차 검증 체계 구성

#### ○ 체계적 검수 프로세스 설계

- 1차 개별 검수 → 2차 교차 검수 → 3차 합의 도출의 단계적 접근
- 검수 기준 및 가이드라인 표준화를 통한 일관성 확보
- 전문가 의견 불일치, 복잡한 사례 등에 대한 의사결정 체계 설정

### □ 검수 이력 관리 체계

#### ○ 검수 기록 및 이력 관리 체계

- 검수자, 검수 일시, 수정 내용, 근거 등 상세 이력 기록
- 사업별 검수 단계 및 진행 상태 기록
- 검수 결과의 승인/반려 사유 및 후속 조치 사항 관리

#### ○ 개선을 위한 피드백 데이터 축적

- 검수 과정에서 발견된 공통 오류 패턴 데이터베이스 구축
- 검수 결과를 활용한 LLM 프롬프트 개선 방향 도출

### 3. 지속적 개선 및 시스템 고도화 방안

#### □ 피드백 루프 기반 시스템 개선 체계

##### ○ 오류 패턴 분석 및 근본 원인 파악

- 정보 추출 오류, 구조화 오류, 의미적 오류 등 체계적 오류 분류 및 분석
- 빈발 오류 패턴의 근본 원인 추적
- 정량적 오류 분석을 통한 개선 우선순위 설정

##### ○ 프롬프트 최적화 및 시스템 보완

- 데이터 기반 프롬프트 엔지니어링
- 정책 용어 사전 및 분류 체계 및 전문가 의견 등 정확한 정보를 참조 데이터로 축적하고, 이를 RAG 시스템 지식베이스로 확장

#### □ 시스템 개선 기반 전체 데이터 재처리 방안

##### ○ 우선순위 기반 단계적 재처리

- 중요도가 높은 사업부터 개선된 시스템으로 우선 재처리
- 오류율이 높았던 그룹에 대한 집중적 재처리
- 최신순으로 재처리하여 현재성 높은 정보 우선 보완

##### ○ 버전 관리 및 롤백 체계

- 재처리 결과의 버전별 관리 및 이전 버전과의 차이점 추적
- 재처리 결과가 부정적인 경우 이전 버전으로의 신속한 롤백
- 재처리 과정의 모든 단계에 대한 로그 기록 및 추적 가능성 확보

○ 품질 관리 위한 검증 체계

- 재처리 결과 품질 기준 설정
- 이전 버전 대비 급격한 변화가 있는 경우 자동 플래깅
- 통계적 이상치 탐지를 통한 재처리 오류 조기 발견



## Abstract

### **Long-term Social Security Financial Analysis I** : Database Development for Social Security Financial Analysis

Project Head: Song, Changgil

Current social security fiscal analysis in Korea is conducted separately by program, making it difficult to comprehensively assess the overall fiscal burden. This study is the first-year component of a three-year project aimed at establishing an internationally comparable fiscal analysis framework based on OECD SOCX. The primary objective is to develop an integrated database that combines quantitative data (expenditure levels) and qualitative data (institutional characteristics) using Large Language Models (LLMs), and to validate this methodology through a pilot analysis of the old-age policy domain.

We designed a three-dimensional classification framework—consisting of scheme type, coverage type, and benefit type—and applied it to 6,637 programs across all SOCX domains. Using LLM-based text processing, we constructed an integrated quantitative-qualitative database. The pilot analysis of the old-age domain demonstrated that the database enables (1) generation of integrated quantitative-qualitative fiscal indicators, (2) assessment of Korea's position through interna-

---

Co-Researchers: Na, Wonhee · Choi, Hyunsoo · Shin, Jeongwoo · Song, Changyong · Han, Shinsil · Choi, Koangsung · Ko, Kyungpyo · Jo, Yongchan

tional comparison, and (3) application to long-term fiscal analysis.

This study is significant in that it establishes a fiscal analysis infrastructure capable of overcoming the limitations of expenditure-centered approaches and enabling simultaneous analysis of institutional characteristics and expenditure levels. With further improvements over the next two years, this framework is expected to evolve into a comprehensive long-term fiscal analysis system that can provide holistic assessments of Korea's social security system and evaluate the fiscal impacts of alternative policy scenarios.

Key words: Social Security Finance, Long-term Fiscal Analysis,  
OECD SOCX, LLM